CPSC 768: Scalable and Private Graph Algorithms

Lecture 11 and 12: Differential Privacy Tools and Graphs

Quanquan C. Liu
quanquan.liu@yale.edu

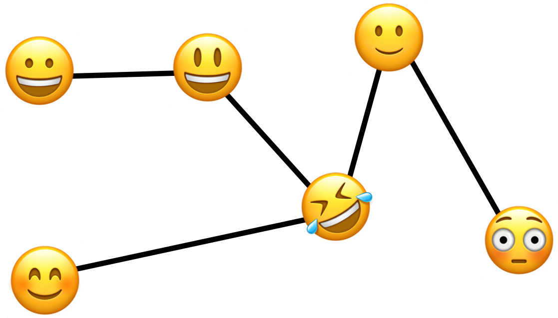
Announcements

- Check the latest announcement on Canvas:
  - Scheduling Lectures survey: due **Feb. 26**
  - Final Project Proposal: due **Feb. 29, one page**
  - Final Project Examples

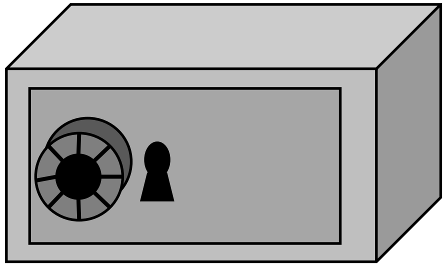
- Open problem sessions:
  - Link for joining CPSC 768 Slack
  - Open Problem Session food orders

Private Analysis of Graph Data

Graph G



Trusted Curator



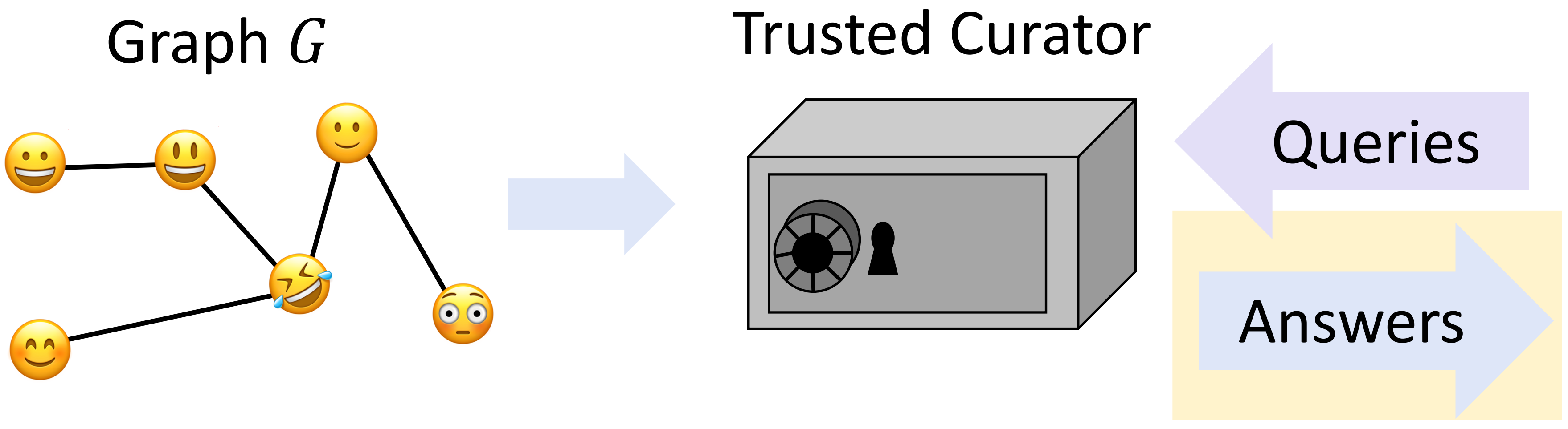
Queries

Answers

Users

Researchers,
Government,
Businesses,
and
Malicious Adversaries

Private Analysis of Graph Data



Users

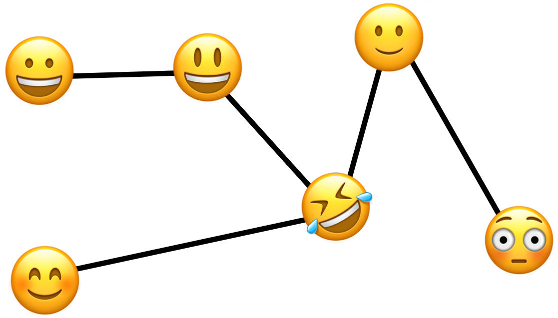
Researchers,
Government,
Businesses,
and
Malicious Adversaries

Two **conflicting** goals:

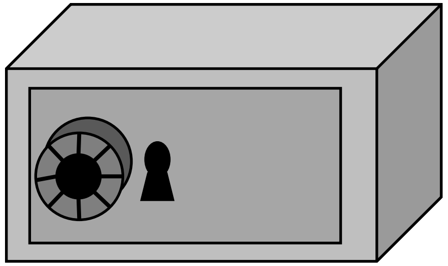
- **Accurate** outputs
- **Data privacy**

Private Analysis of Graph Data

Graph G



Trusted Curator



Queries

Answers

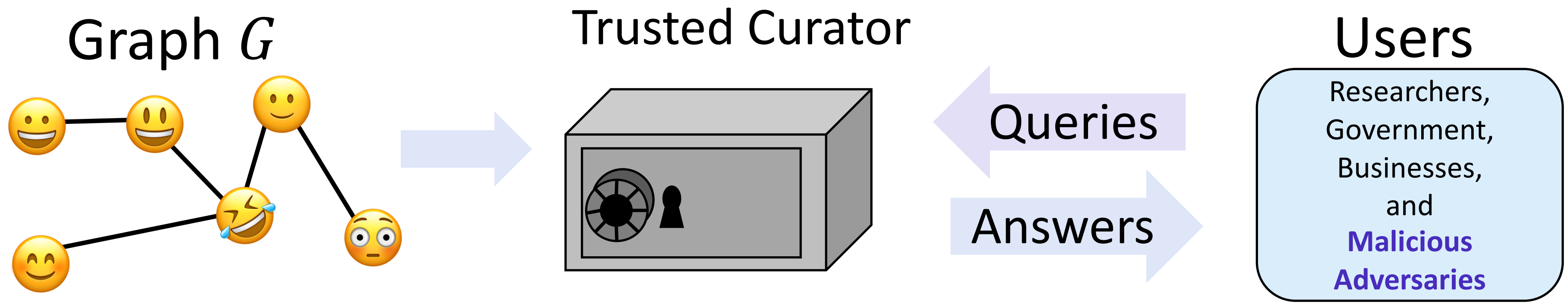
Users

Researchers,
Government,
Businesses,
and
Malicious Adversaries

Two **conflicting** goals:

- **Accurate** outputs
- **Data privacy**

(Central Model of) Differential Privacy



- **Neighboring** inputs differ in some information we'd like to hide

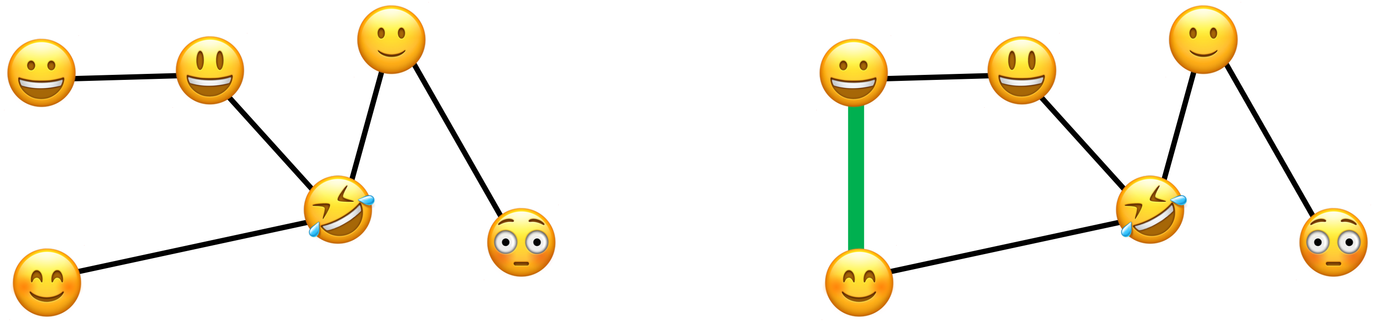
<u>Differential Privacy [Dwork-McSherry-Nissim-Smith '06]</u>

An algorithm $\mathcal{A}$ is **ε-differentially private** if for all pairs of neighbors G and G' and all sets of possible outputs S:

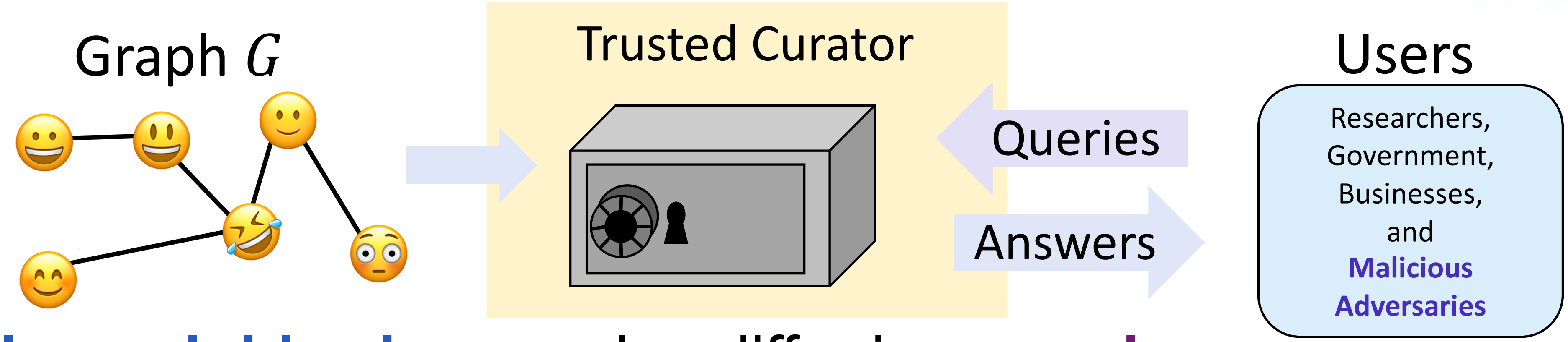
$$\Pr[\mathcal{A}(G) \in S] \leq e^{\varepsilon} \cdot \Pr[\mathcal{A}(G') \in S].$$

Edge-Neighboring Graphs

- **Edge-neighboring** graphs: differ in **one edge**



(Central Model of) Differential Privacy



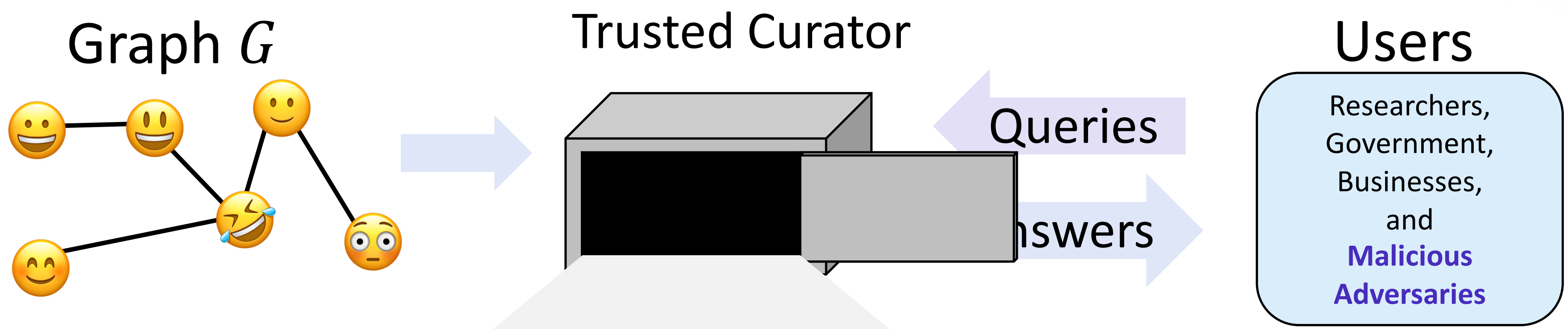
- **Edge-neighboring** graphs: differ in **one edge**

Differential Privacy [Dwork-McSherry-Nissim-Smith '06]

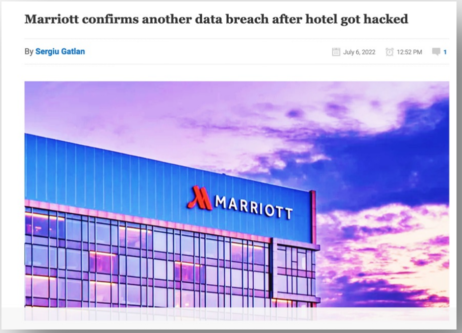
An algorithm $\mathcal{A}$ is **ε-differentially private** if for all pairs of neighbors G and G' and all sets of possible outputs S:

$$\Pr[\mathcal{A}(G) \in S] \leq e^{\varepsilon} \cdot \Pr[\mathcal{A}(G') \in S].$$

(Central Model of) Differential Privacy



https://www.npr.org/2021/04/09/986005820/after-data-breach-exposes-530-million-facebook-says-it-will-not-notify-users

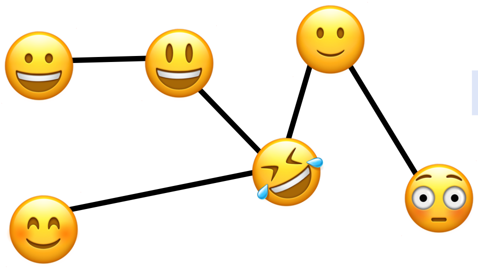


https://www.bleepingcomputer.com/news/security/marriott-confirms-another-data-breach-after-hotel-got-hacked/

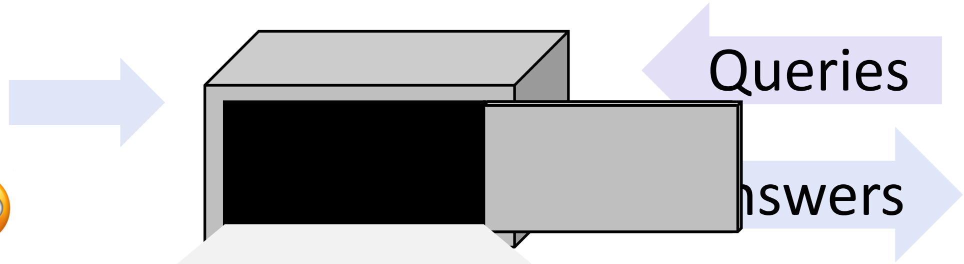
https://www.malwarebytes.com/blog/news/2021/06/second-colossal-linkedin-breach-in-3-months-almost-all-users-affected

(Central Model of) Differential Privacy

Graph G



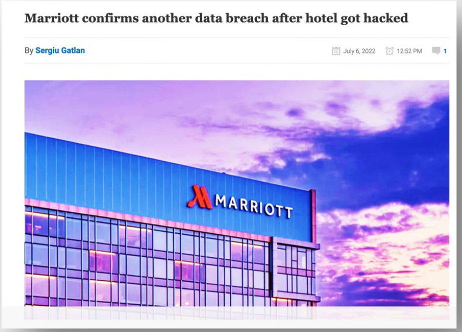
Trusted Curator



Users

Researchers,
Government,
Businesses,
and
Malicious Adversaries

https://www.npr.org/2021/04/09/986005820/after-data-breach-exposes-530-million-facebook-says-it-will-not-notify-users

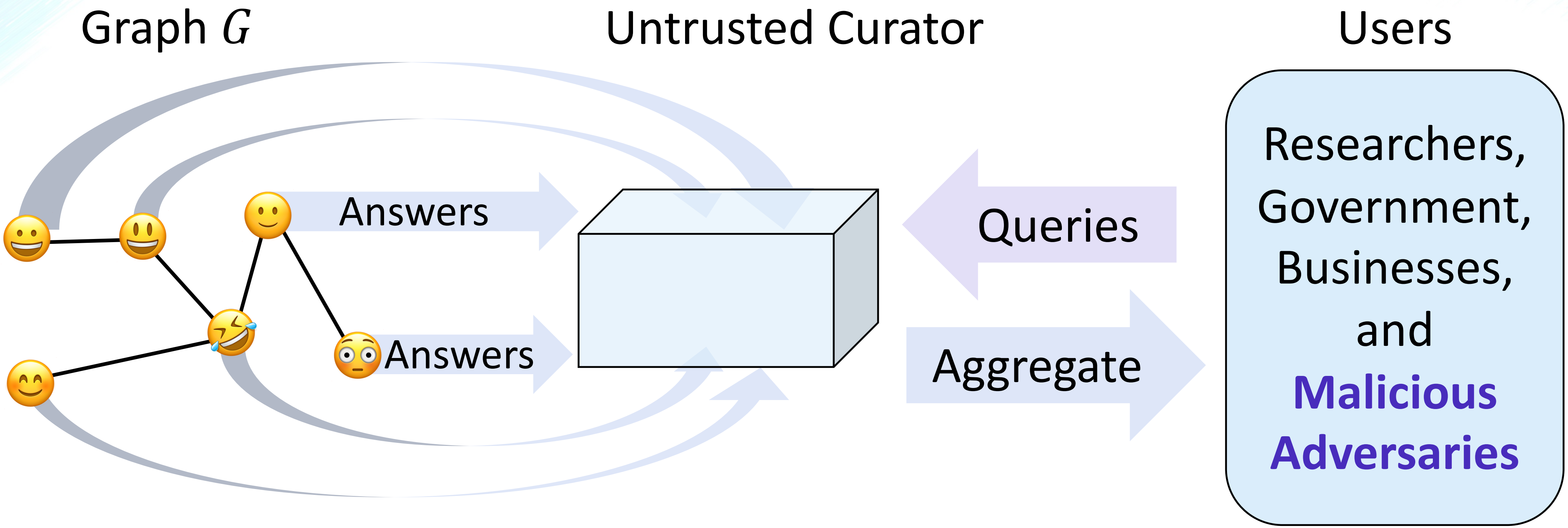


https://www.bleepingcomputer.com/news/security/marriott-confirms-another-data-breach-after-hotel-got-hacked/

https://www.malwarebytes.com/blog/news/2021/06/second-colossal-linkedin-breach-in-3-months-almost-all-users-affected

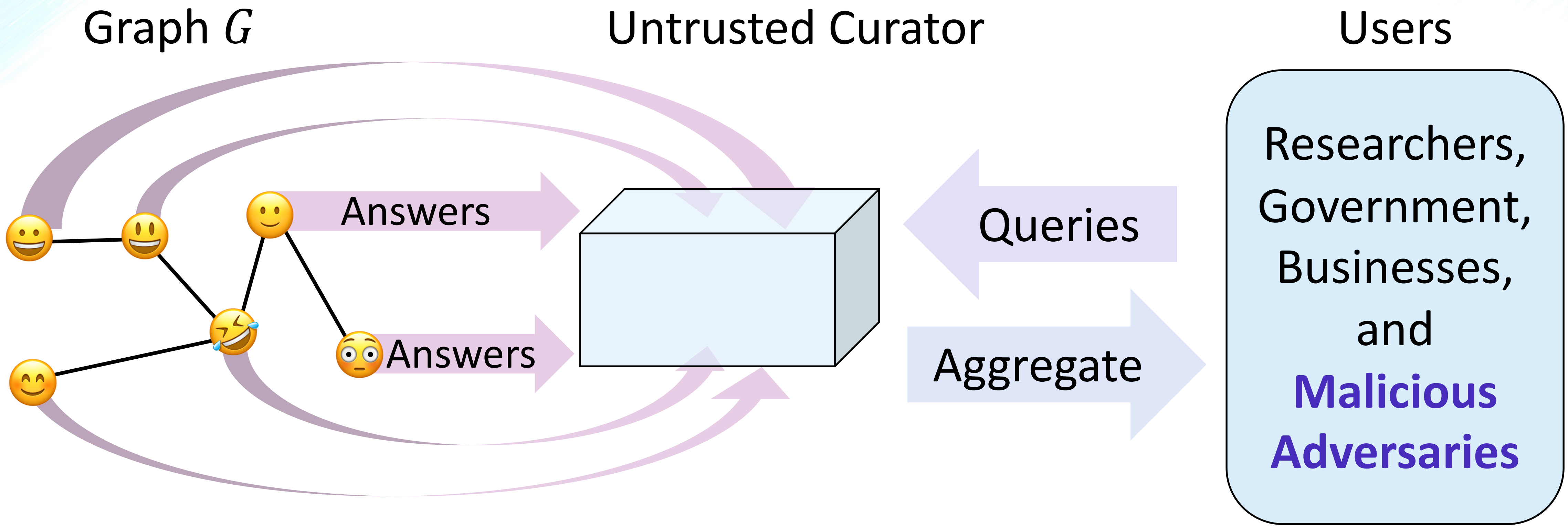
Unrealistic trust in **trusted curator**

Weaker Notion of Trust: Local Model



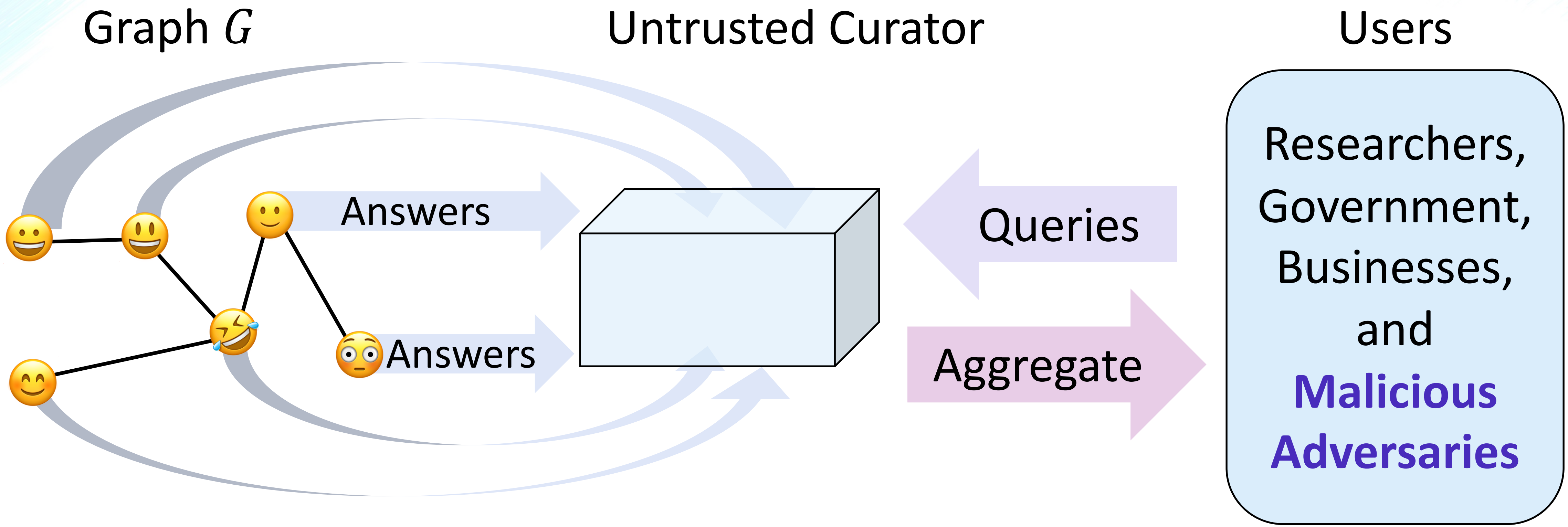
- **Each node publishes privatized** output
- Curator computes aggregated statistics using outputs

Weaker Notion of Trust: **Local Model**



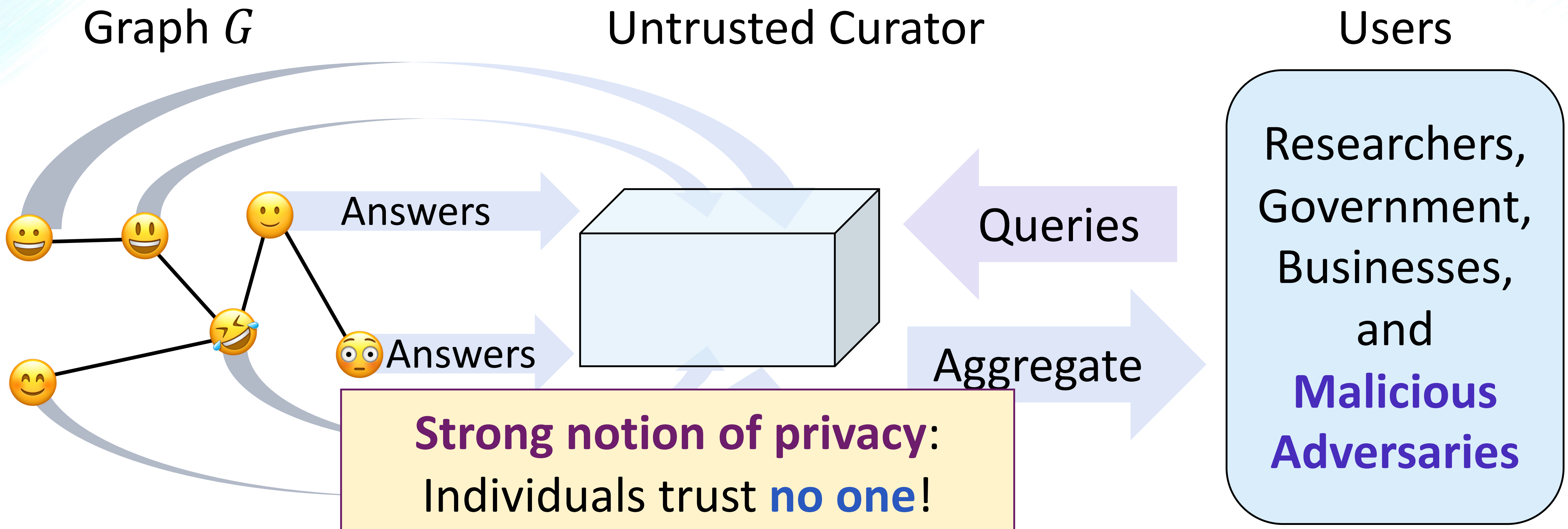
- **Each node publishes privatized** output
- Curator computes aggregated statistics using outputs

Weaker Notion of Trust: **Local Model**



- **Each node publishes privatized** output
- Curator computes aggregated statistics using outputs

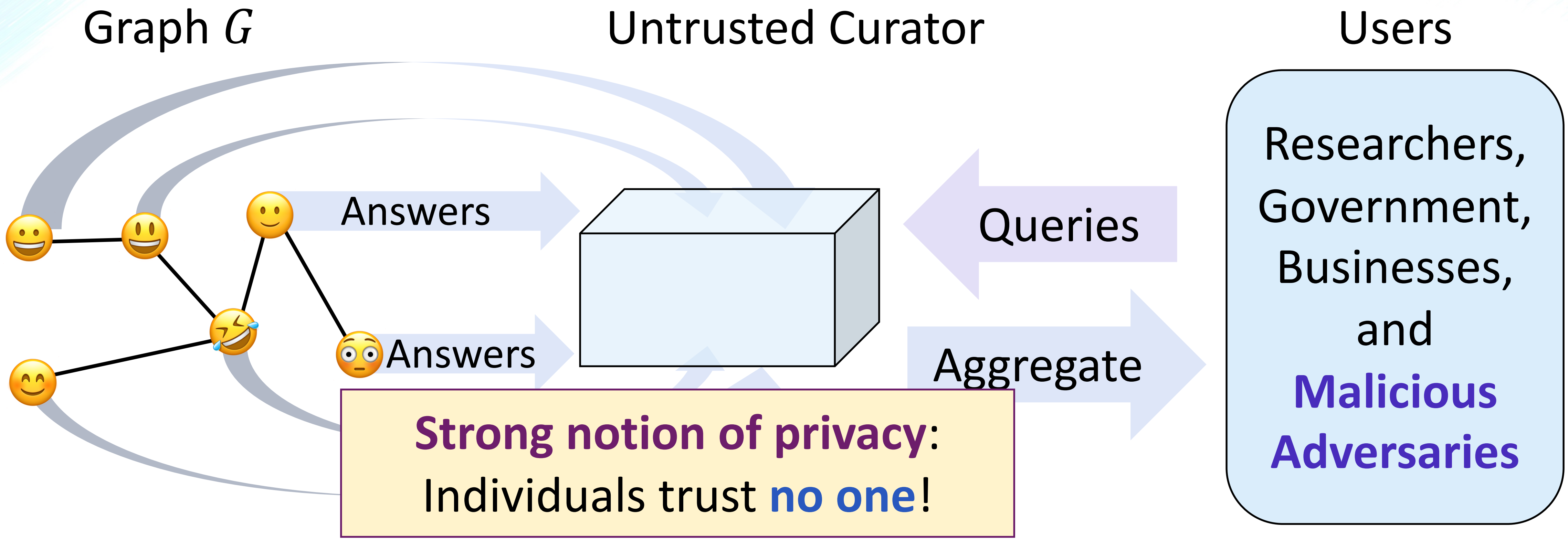
Weaker Notion of Trust: **Local Model**



Strong notion of privacy:
Individuals trust **no one**!

- **Each node publishes privatized** output
- Curator computes aggregated statistics using outputs

Weaker Notion of Trust: **Local Model**



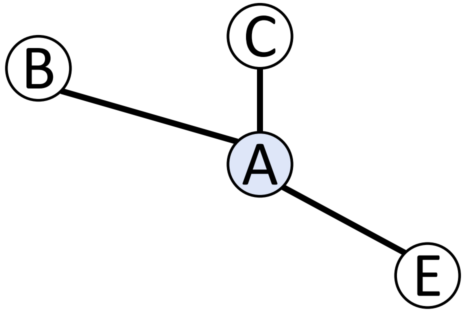
- **Each node publishes privatized** output
- Curator computes aggregated statistics using outputs

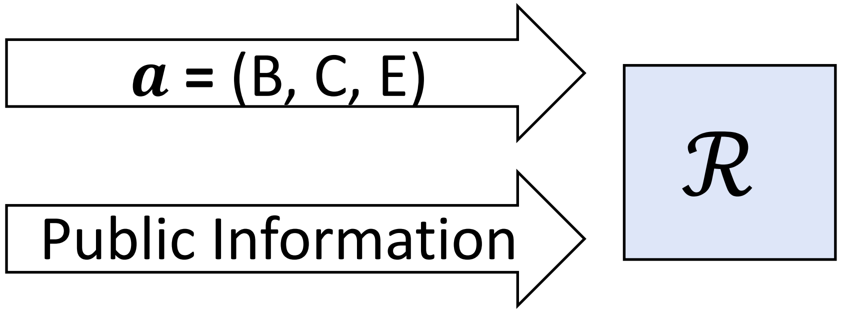
Local Edge Differential Privacy (LEDP)

Local Randomizer

[Adapted from Kasiviswanathan-Lee-Nissim-Raskhodnikova-Smith '11]

An **ε-local randomizer** $\mathcal{R}$ is an ε-differentially private algorithm that takes as input an adjacency list $\boldsymbol{a}$ and public information.





Local Edge Differential Privacy (LEDP)

Untrusted Curator

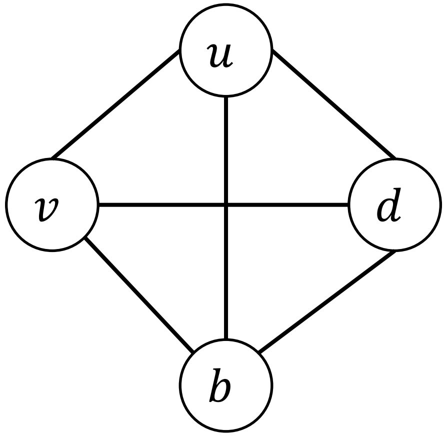
Users

Algorithm proceeds in **rounds** in **distributed graph** using local randomizers

Queries

Aggregate

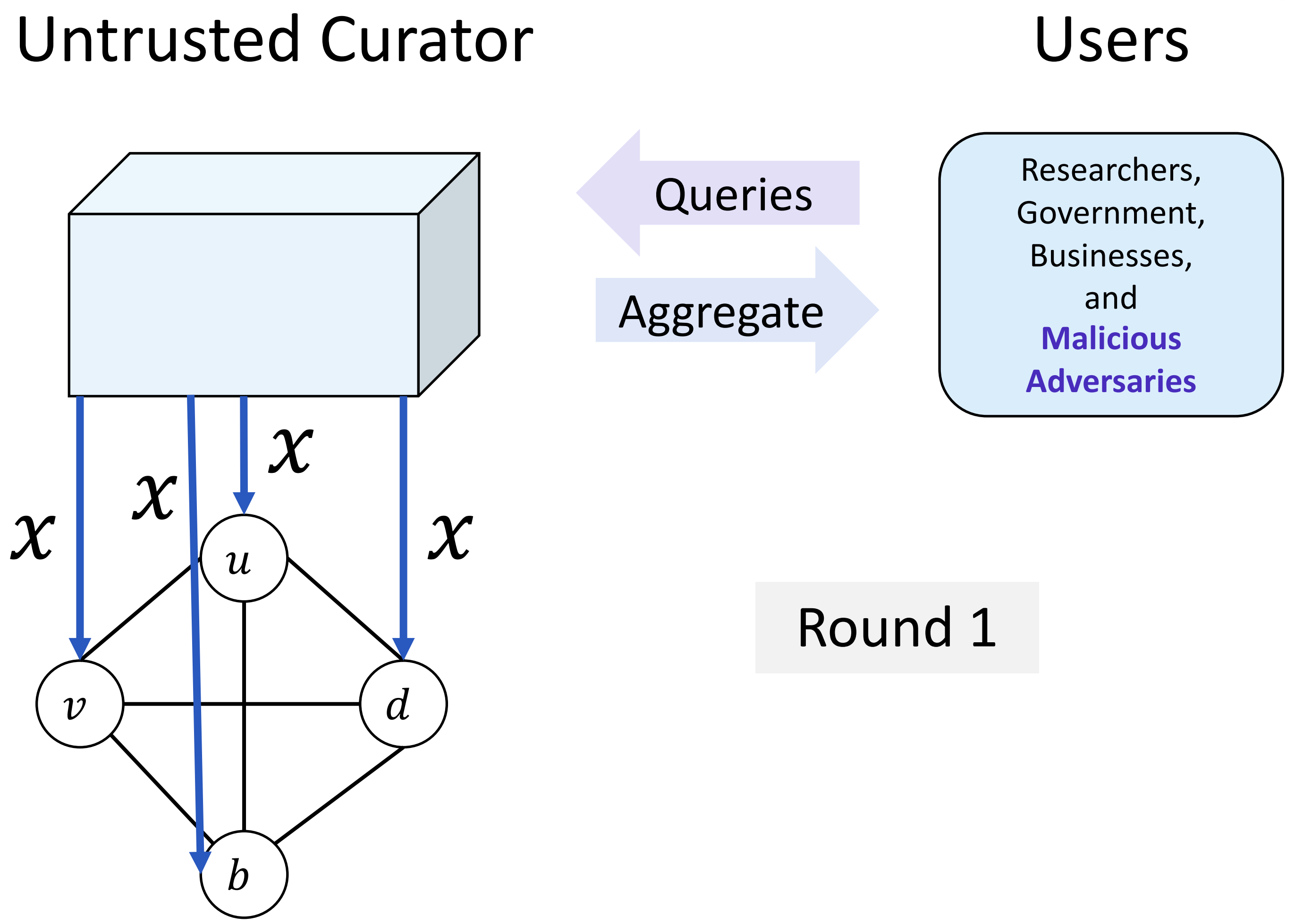
Researchers, Government, Businesses, and **Malicious Adversaries**



Round 1

Local Edge Differential Privacy (LEDP)

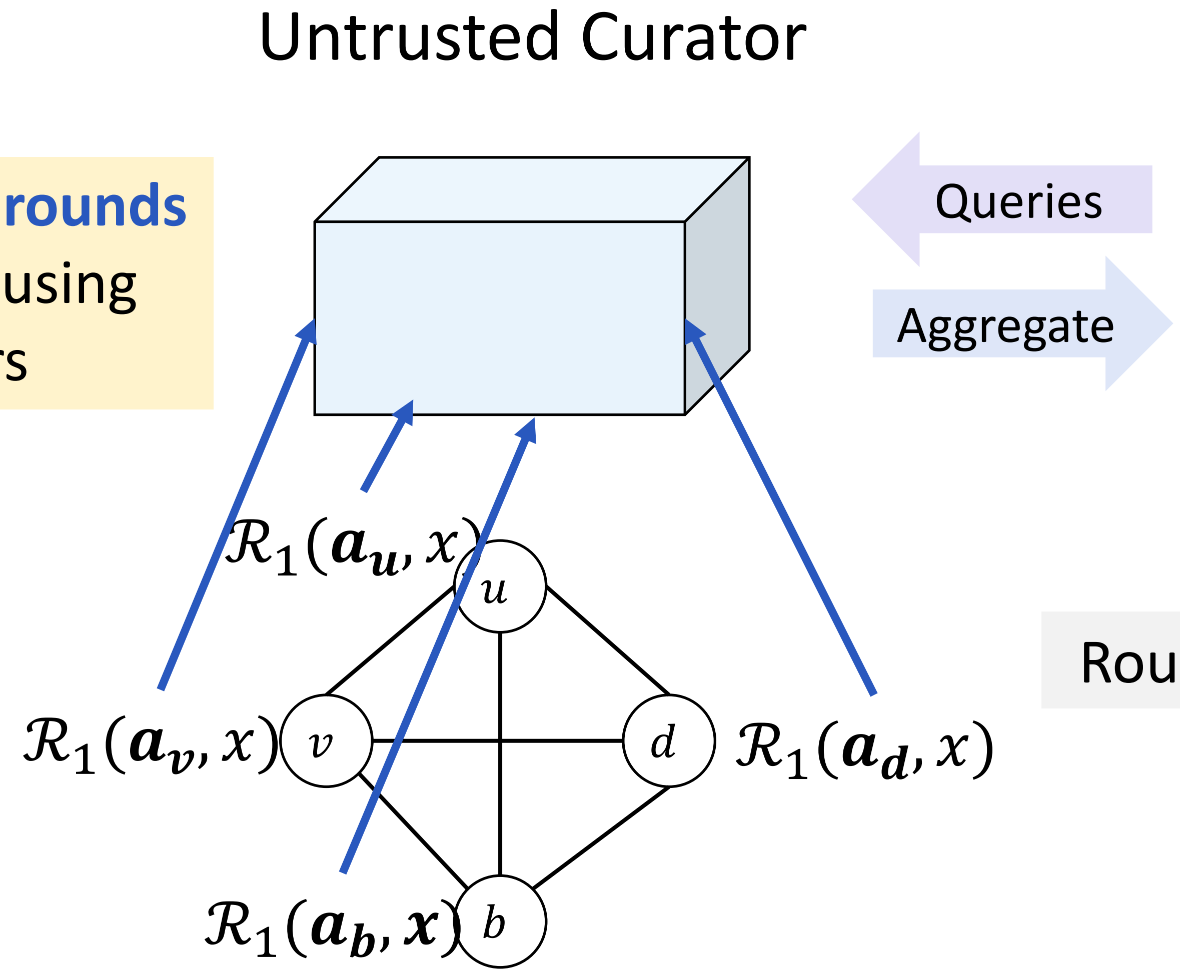
Algorithm proceeds in **rounds** in **distributed graph** using local randomizers



Round 1

Local Edge Differential Privacy (LEDP)

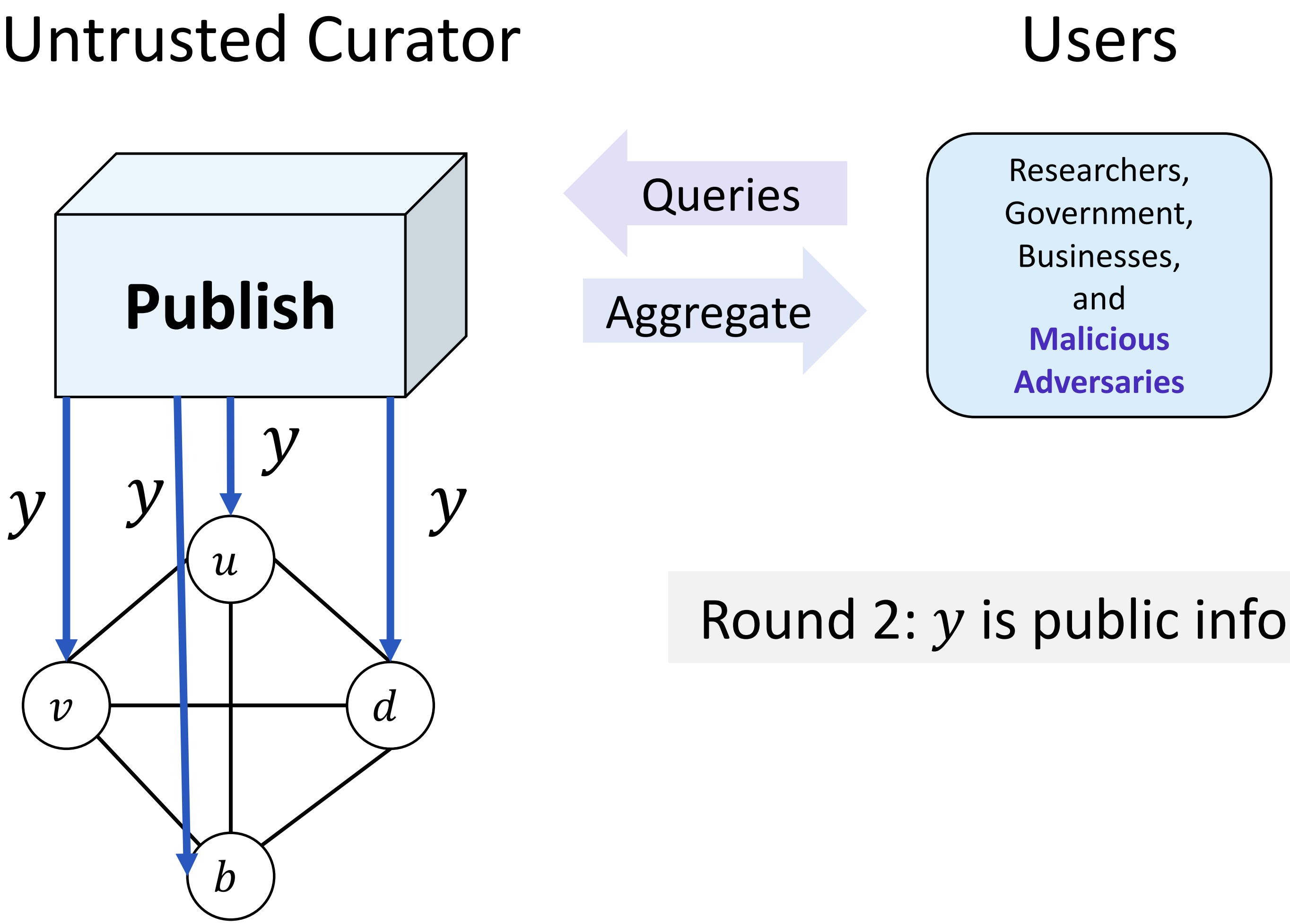
Algorithm proceeds in **rounds** in **distributed graph** using local randomizers



Round 1: x is public info

Local Edge Differential Privacy (LEDP)

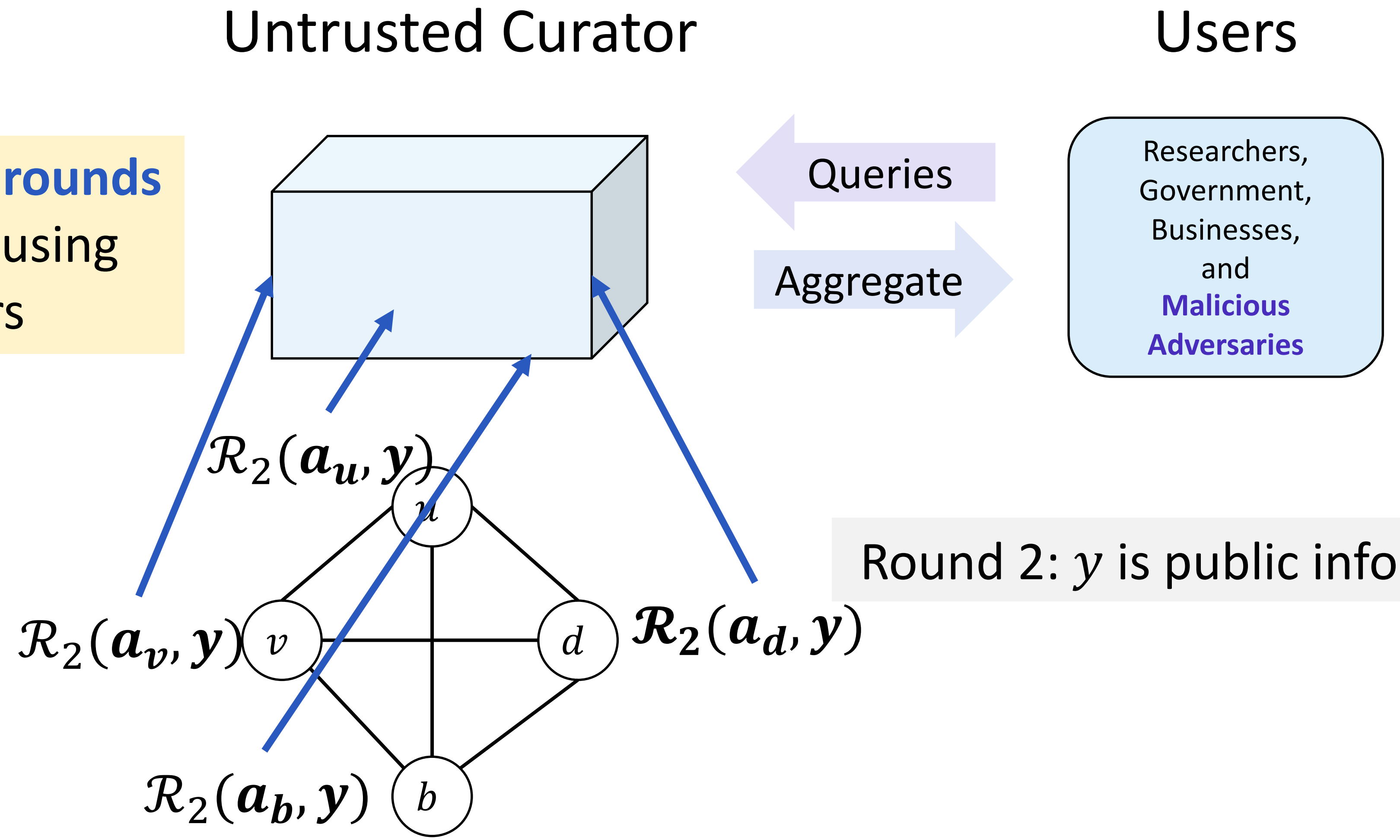
Algorithm proceeds in **rounds** in **distributed graph** using local randomizers



Round 2: y is public info

Local Edge Differential Privacy (LEDP)

Algorithm proceeds in **rounds** in **distributed graph** using local randomizers



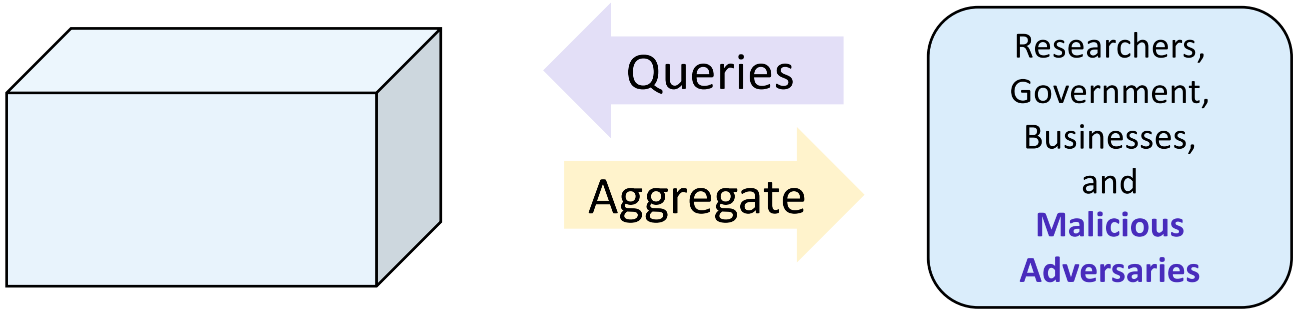
Round 2: y is public info

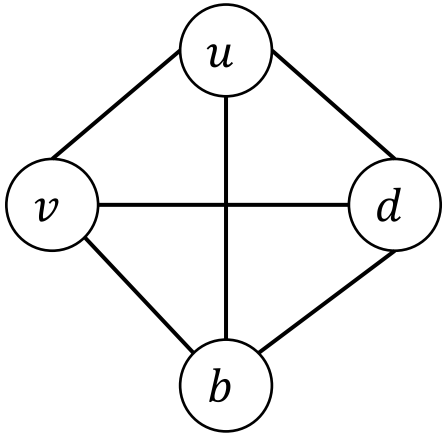
Local Edge Differential Privacy (LEDP)

Untrusted Curator

Users

Algorithm proceeds in **rounds** in **distributed graph** using local randomizers





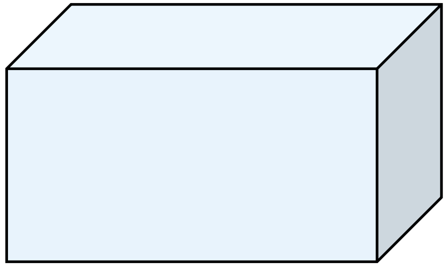
Round 2

Local Edge Differential Privacy (LEDP)

Untrusted Curator

Users

Algorithm proceeds in **rounds** in **distributed graph** using local randomizers

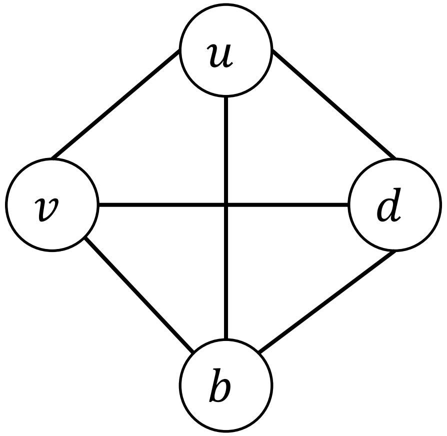


Queries

Aggregate

Researchers, Government, Businesses, and **Malicious Adversaries**

Relevant Complexity Measure:
Number of Rounds of Communication



Round 2

Local Edge Differential Privacy (LEDP)

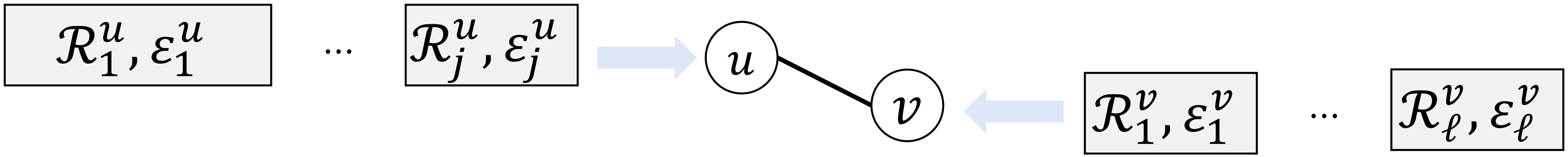
Local Edge Differential Privacy

[DLRSSY '22 Adapted from Kasiviswanathan-Lee-Nissim-Raskhodnikova-Smith '11]

Let algorithm $\mathcal{A}$ use (potentially different) local randomizers $\mathcal{R}_1^u, \ldots, \mathcal{R}_j^u$ and $\mathcal{R}_1^v, \ldots, \mathcal{R}_\ell^v$ on nodes u, v with privacy parameters $\varepsilon_1^u, \ldots, \varepsilon_j^u$ and $\varepsilon_1^v, \ldots, \varepsilon_\ell^v$.

$\mathcal{A}$ is **ε-local edge differentially private (ε-LEDP)** if for every edge $\{u, v\}$,

$$\varepsilon_1^u + \cdots + \varepsilon_j^u + \varepsilon_1^v \cdots + \varepsilon_\ell^v \leq \varepsilon.$$



$$\varepsilon_1^u + \cdots + \varepsilon_j^u + \varepsilon_1^v \cdots + \varepsilon_\ell^v \leq \varepsilon$$

Related Work

- **Local edge differentially private algorithms**:
  - Relatively **new direction**
  - *k*-**Core Decomposition, Densest Subgraphs, Low Out-degree Ordering:** [Dhulipala-Liu-Raskhodnikova-Shi-Shun-Yu ‘22, Dinitz-Kale-Lattanzi-Vassilvitskii ‘23, Dhulipala-Li-Liu ‘23]
  - **Triangle and other subgraph counting:** [Imola-Murakami-Chaudhuri ‘21, ’22; Eden-Liu-Raskhodnikova-Smith ‘23]
  - **Other graph problems** in empirical settings in “decentralized” privacy models [Sun-Xiao-Khalil-Yang-Qin-Wang-Yu ‘19; Qin-Yu-Yang-Khalil-Xiao-Ren ‘17; Gao-Li-Chen-Zou ‘18; Ye-Hu-Au-Meng-Xiao ‘20]

Randomized Response

- Given an input set of bits $X_1, \ldots, X_n \in \{0, 1\}$ for each individual (i.e. has COVID)

Randomized Response

- Given an input set of bits $X_1, \ldots, X_n \in \{0, 1\}$ for each individual (i.e. has COVID)
- Randomly report the same bit or flipped bit:
  - $Y_i = \begin{cases} X_i & \text{w.p. } 1/2 + \varepsilon \\ 1 - X_i & \text{w.p. } 1/2 - \varepsilon \end{cases}$

Randomized Response

- Given an input set of bits $X_1, \ldots, X_n \in \{0, 1\}$ for each individual (i.e. has COVID)
- Randomly report the same bit or flipped bit:
  - $Y_i = \begin{cases} X_i & \text{w.p. } ½ + \varepsilon \\ 1 - X_i & \text{w.p. } ½ - \varepsilon \end{cases}$
- Sum of Y_i bits: $O\left(\frac{\sqrt{n}}{\varepsilon}\right)$ error but locally private

Randomized Response

- Given an input set of bits $X_1, \ldots, X_n \in \{0, 1\}$ for each individual (i.e. has COVID)
- Randomly report the same bit or flipped bit:
  - $Y_i = \begin{cases} X_i & \text{w.p. } ½ + \varepsilon \\ 1 - X_i & \text{w.p. } ½ - \varepsilon \end{cases}$
- Sum of Y_i bits: $O\left(\frac{\sqrt{n}}{\varepsilon}\right)$ error but locally private
- Geometric mechanism: $O\left(\frac{1}{\varepsilon}\right)$ error but not locally private

Locally Private Triangle Counting

Locally Private Triangle Counting

Central DP vs. LEDP

Locally Private Triangle Counting

Central DP vs. LEDP		
	DP Upper Bound	**LEDP Lower Bound**
Triangle Counting	$O\left(\frac{n}{\varepsilon}\right)$ **additive error** (trivial)	$\Omega\left(\frac{n^{3/2}}{\varepsilon}\right)$ **additive error** (multiple rounds) $\Omega\left(\frac{n^{2}}{\varepsilon}\right)$ **additive error** (one round) [Eden-Liu-Raskhodnikova-Smith **ICALP '23**]

One-Round Locally Private Triangle Counting [Eden-Liu-Raskhodnikova-Smith '23]

- **Input:** Graph $G = ([n], E)$ represented by $n \times n$ adjacency matrix A with entries $a_{ij}, \varepsilon > 0$

One-Round Locally Private Triangle Counting [Eden-Liu-Raskhodnikova-Smith '23]

- **Input:** Graph $G = ([n], E)$ represented by $n \times n$ adjacency matrix A with entries $a_{ij}, \varepsilon > 0$

Each node holds their own adjacency list as private info

One-Round Locally Private Triangle Counting [Eden-Liu-Raskhodnikova-Smith '23]

- **Input:** Graph $G = ([n], E)$ represented by $n \times n$ adjacency matrix A with entries $a_{ij}, \varepsilon > 0$
- **Output:** Approximate number of triangles in G

One-Round Locally Private Triangle Counting [Eden-Liu-Raskhodnikova-Smith '23]

- **Input:** Graph $G = ([n], E)$ represented by $n \times n$ adjacency matrix A with entries $a_{ij}, \varepsilon > 0$
- **Output:** Approximate number of triangles in G
  - for $i = 1, \ldots, n$:
    - Release $\left(X_{i,i+1}, \ldots, X_{i,n}\right)$ where $X_{i,j} = 1 - a_{i,j}$ with probability $\frac{1}{e^{\varepsilon}+1}$ and $a_{i,j}$ otherwise

One-Round Locally Private Triangle Counting [Eden-Liu-Raskhodnikova-Smith '23]

- **Input:** Graph $G = ([n], E)$ represented by $n \times n$ adjacency matrix A with entries $a_{ij}, \varepsilon > 0$
- **Output:** Approximate number of triangles in G
  - for $i = 1, \ldots, n$:
    - Release $\left(X_{i,i+1}, \ldots, X_{i,n}\right)$ where $X_{i,j} = 1 - a_{i,j}$ with probability $\frac{1}{e^{\varepsilon}+1}$ and $a_{i,j}$ otherwise

Releasing noisy upper triangular matrix

1	**0**	**1**
0	1	**0**
0	0	1

One-Round Locally Private Triangle Counting [Eden-Liu-Raskhodnikova-Smith '23]

- **Input:** Graph $G = ([n], E)$ represented by $n \times n$ adjacency matrix A with entries $a_{ij}, \varepsilon > 0$
- **Output:** Approximate number of triangles in G
  - for $i = 1, \ldots, n$:
    - Release $\left(X_{i,i+1}, \ldots, X_{i,n}\right)$ where $X_{i,j} = 1 - a_{i,j}$ with probability $\frac{1}{e^{\varepsilon}+1}$ and $a_{i,j}$ otherwise
  - For all $\{i, j\} \in \binom{[n]}{2}$, set $Y_{i,j} \leftarrow \frac{\left(X_{i,j} \cdot (e^{\varepsilon}+1)-1\right)}{e^{\varepsilon}-1}$

One-Round Locally Private Triangle Counting [Eden-Liu-Raskhodnikova-Smith ‘23]

- **Input:** Graph $G = ([n], E)$ represented by $n \times n$ adjacency matrix A with entries $a_{ij}, \varepsilon > 0$
- **Output:** Approximate number of triangles in G
  - for $i = 1, \ldots, n$:
    - Release $\left(X_{i,i+1}, \ldots, X_{i,n}\right)$ where $X_{i,j} = 1 - a_{i,j}$ with probability $\frac{1}{e^{\varepsilon}+1}$ and $a_{i,j}$ otherwise
  - For all $\{i, j\} \in \binom{[n]}{2}$, set $Y_{i,j} \leftarrow \frac{\left(X_{i,j} \cdot (e^{\varepsilon}+1)-1\right)}{e^{\varepsilon}-1}$
  - For all $\{i, j, k\} \in \binom{[n]}{3}$, set $Z_{i,j,k} \leftarrow Y_{i,j} \cdot Y_{j,k} \cdot Y_{i,k}$

Normalized $Y_{i,j}$ so that $E\left[Z_{i,j,k}\right] = 1$ if triangle exists and 0 otherwise

One-Round Locally Private Triangle Counting [Eden-Liu-Raskhodnikova-Smith '23]

- **Input:** Graph $G = ([n], E)$ represented by $n \times n$ adjacency matrix A with entries $a_{ij}, \varepsilon > 0$
- **Output:** Approximate number of triangles in G
  - for $i = 1, \ldots, n$:
    - Release $\left(X_{i,i+1}, \ldots, X_{i,n}\right)$ where $X_{i,j} = 1 - a_{i,j}$ with probability $\frac{1}{e^{\varepsilon}+1}$ and $a_{i,j}$ otherwise
  - For all $\{i, j\} \in \binom{[n]}{2}$, set $Y_{i,j} \leftarrow \frac{\left(X_{i,j} \cdot (e^{\varepsilon}+1) - 1\right)}{e^{\varepsilon}-1}$
  - For all $\{i, j, k\} \in \binom{[n]}{3}$, set $Z_{i,j,k} \leftarrow Y_{i,j} \cdot Y_{j,k} \cdot Y_{i,k}$
  - Return $\boldsymbol{\widehat{T}} \leftarrow \sum_{\{\boldsymbol{i,j,k}\} \in \binom{[n]}{3}} \boldsymbol{Z_{i,j,k}}$

One-Round Locally Private Triangle Counting [Eden-Liu-Raskhodnikova-Smith '23]

- **Input:** Graph $G = ([n], E)$ represented by $n \times n$ adjacency matrix A with entries $a_{ij}, \varepsilon > 0$
- **Output:** Approximate number of triangles in G
  - for $i = 1, \ldots, n$:
    - Release $\left(X_{i,i+1}, \ldots, X_{i,n}\right)$ where $X_{i,j} = 1 - a_{i,j}$ with probability $\frac{1}{e^{\varepsilon}+1}$ and $a_{i,j}$ otherwise
  - For all $\{i, j\} \in \binom{[n]}{2}$, set $Y_{i,j} \leftarrow \frac{\left(X_{i,j} \cdot (e^{\varepsilon}+1) - 1\right)}{e^{\varepsilon}-1}$
  - For all $\{i, j, k\} \in \binom{[n]}{3}$, set $Z_{i,j,k} \leftarrow Y_{i,j} \cdot Y_{j,k} \cdot Y_{i,k}$
  - Return $\boldsymbol{\hat{T}} \leftarrow \sum_{\{\boldsymbol{i,j,k}\} \in \binom{[n]}{3}} \boldsymbol{Z_{i,j,k}}$

Therefore, $E\left[\hat{T}\right] = T$

Analysis of the Expectation and Variance

- **Lemma:** Returns an unbiased estimate of the number of triangles in the input graph

Analysis of the Expectation and Variance

- **Lemma:** Returns an unbiased estimate of the number of triangles in the input graph
- Proof:
  - $X_{i,j}$ is indicator variable for presence of noisy edge $\{i, j\}$

Analysis of the Expectation and Variance

- **Lemma:** Returns an unbiased estimate of the number of triangles in the input graph
- Proof:
  - $X_{i,j}$ is indicator variable for presence of noisy edge $\{i, j\}$
  - If $\{i, j\} \in E$, then $E[X_{i,j}] = \frac{e^{\varepsilon}}{e^{\varepsilon}+1}$; otherwise $E[X_{i,j}] = \frac{1}{e^{\varepsilon}+1}$, **why?**

Analysis of the Expectation and Variance

- **Lemma:** Returns an unbiased estimate of the number of triangles in the input graph
- Proof:
  - $X_{i,j}$ is indicator variable for presence of noisy edge $\{i, j\}$
  - If $\{i, j\} \in E$, then $E[X_{i,j}] = \frac{e^{\varepsilon}}{e^{\varepsilon}+1}$; otherwise $E[X_{i,j}] = \frac{1}{e^{\varepsilon}+1}$, **why?**
    - If $\{i, j\} \in E$, probability **doesn't flip** is $\frac{e^{\varepsilon}}{e^{\varepsilon}+1}$; otherwise, **probability flips** is $\frac{1}{e^{\varepsilon}+1}$

Analysis of the Expectation and Variance

- **Lemma:** Returns an unbiased estimate of the number of triangles in the input graph
- Proof:
  - $X_{i,j}$ is indicator variable for presence of noisy edge $\{i,j\}$
  - If $\{i,j\} \in E$, then $E[X_{i,j}] = \frac{e^{\varepsilon}}{e^{\varepsilon}+1}$; otherwise $E[X_{i,j}] = \frac{1}{e^{\varepsilon}+1}$, **why?**
    - If $\{i,j\} \in E$, probability **doesn't flip** is $\frac{e^{\varepsilon}}{e^{\varepsilon}+1}$; otherwise, **probability flips** is $\frac{1}{e^{\varepsilon}+1}$
  - Then, $E[Y_{i,j}] = E\left[\frac{(X_{i,j}\cdot(e^{\varepsilon}+1)-1)}{e^{\varepsilon}-1}\right] = \frac{(E[X_{i,j}]\cdot(e^{\varepsilon}+1)-1)}{e^{\varepsilon}-1}$

Analysis of the Expectation and Variance

- Proof:
  - If $\{i,j\} \in E$, then $E[X_{i,j}] = \frac{e^{\varepsilon}}{e^{\varepsilon}+1}$; otherwise $E[X_{i,j}] = \frac{1}{e^{\varepsilon}+1}$
  - Then, $E[Y_{i,j}] = E\left[\frac{(X_{i,j}\cdot(e^{\varepsilon}+1)-1)}{e^{\varepsilon}-1}\right] = \frac{(E[X_{i,j}]\cdot(e^{\varepsilon}+1)-1)}{e^{\varepsilon}-1}$

Analysis of the Expectation and Variance

- Proof:
  - If $\{i, j\} \in E$, then $E[X_{i,j}] = \frac{e^{\varepsilon}}{e^{\varepsilon}+1}$; otherwise $E[X_{i,j}] = \frac{1}{e^{\varepsilon}+1}$
  - Then, $E[Y_{i,j}] = E\left[\frac{\left(X_{i,j} \cdot (e^{\varepsilon}+1)-1\right)}{e^{\varepsilon}-1}\right] = \frac{\left(E[X_{i,j}] \cdot (e^{\varepsilon}+1)-1\right)}{e^{\varepsilon}-1}$
  - If $\{i, j\} \in E$, then $\frac{E[X_{i,j}] \cdot (e^{\varepsilon}+1)-1}{e^{\varepsilon}-1} = \frac{\left(\left(\frac{e^{\varepsilon}}{e^{\varepsilon}+1}\right) \cdot (e^{\varepsilon}+1)-1\right)}{e^{\varepsilon}-1} = \frac{(e^{\varepsilon}-1)}{e^{\varepsilon}-1} = 1$

Analysis of the Expectation and Variance

- Proof:
  - If $\{i,j\} \in E$, then $E\left[X_{i,j}\right] = \frac{e^{\varepsilon}}{e^{\varepsilon}+1}$; otherwise $E\left[X_{i,j}\right] = \frac{1}{e^{\varepsilon}+1}$
  - Then, $E\left[Y_{i,j}\right] = E\left[\frac{\left(X_{i,j} \cdot (e^{\varepsilon}+1)-1\right)}{e^{\varepsilon}-1}\right] = \frac{\left(E\left[X_{i,j}\right] \cdot (e^{\varepsilon}+1)-1\right)}{e^{\varepsilon}-1}$
  - If $\{i,j\} \in E$, then $\frac{E\left[X_{i,j}\right] \cdot (e^{\varepsilon}+1)-1}{e^{\varepsilon}-1} = \frac{\left(\left(\frac{e^{\varepsilon}}{e^{\varepsilon}+1}\right) \cdot (e^{\varepsilon}+1)-1\right)}{e^{\varepsilon}-1} = \frac{(e^{\varepsilon}-1)}{e^{\varepsilon}-1} = 1$
  - If $\{i,j\} \notin E$, then $\frac{E\left[X_{i,j}\right] \cdot (e^{\varepsilon}+1)-1}{e^{\varepsilon}-1} = \frac{\left(\left(\frac{1}{e^{\varepsilon}+1}\right) \cdot (e^{\varepsilon}+1)-1\right)}{e^{\varepsilon}-1} = 0$

Analysis of the Expectation and Variance

- Proof:
  - If $\{i,j\} \in E$, then $E[X_{i,j}] = \frac{e^{\varepsilon}}{e^{\varepsilon}+1}$; otherwise $E[X_{i,j}] = \frac{1}{e^{\varepsilon}+1}$
  - Then, $E[Y_{i,j}] = E\left[\frac{(X_{i,j} \cdot (e^{\varepsilon}+1) - 1)}{e^{\varepsilon}-1}\right] = \frac{(E[X_{i,j}] \cdot (e^{\varepsilon}+1) - 1)}{e^{\varepsilon}-1}$
  - If $\{i,j\} \in E$, then $\frac{E[X_{i,j}] \cdot (e^{\varepsilon}+1) - 1}{e^{\varepsilon}-1} = \frac{\left(\left(\frac{e^{\varepsilon}}{e^{\varepsilon}+1}\right) \cdot (e^{\varepsilon}+1) - 1\right)}{e^{\varepsilon}-1} = \frac{(e^{\varepsilon}-1)}{e^{\varepsilon}-1} = 1$
  - If $\{i,j\} \notin E$, then $\frac{E[X_{i,j}] \cdot (e^{\varepsilon}+1) - 1}{e^{\varepsilon}-1} = \frac{\left(\left(\frac{1}{e^{\varepsilon}+1}\right) \cdot (e^{\varepsilon}+1) - 1\right)}{e^{\varepsilon}-1} = 0$
  - Finally, $E[Z_{i,j,k}] = E[Y_{i,j} \cdot Y_{j,k} \cdot Y_{i,k}] = E[Y_{i,j}] \cdot E[Y_{j,k}] \cdot E[Y_{i,k}] = 1_{i,j,k}$

Analysis of the Expectation and Variance

- Proof:
  - If $\{i,j\} \in E$, then $E\left[X_{i,j}\right] = \frac{e^{\varepsilon}}{e^{\varepsilon}+1}$; otherwise $E\left[X_{i,j}\right] = \frac{1}{e^{\varepsilon}+1}$
  - Then, $E\left[Y_{i,j}\right] = E\left[\frac{\left(X_{i,j}\cdot(e^{\varepsilon}+1)-1\right)}{e^{\varepsilon}-1}\right] = \frac{\left(E\left[X_{i,j}\right]\cdot(e^{\varepsilon}+1)-1\right)}{e^{\varepsilon}-1}$
  - If $\{i,j\} \in E$, then $\frac{E\left[X_{i,j}\right]\cdot(e^{\varepsilon}+1)-1}{e^{\varepsilon}-1} = \frac{\left(\left(\frac{e^{\varepsilon}}{e^{\varepsilon}+1}\right)\cdot(e^{\varepsilon}+1)-1\right)}{e^{\varepsilon}-1} = \frac{(e^{\varepsilon}-1)}{e^{\varepsilon}-1} = 1$
  - If $\{i,j\} \notin E$, then $\frac{E\left[X_{i,j}\right]\cdot(e^{\varepsilon}+1)-1}{e^{\varepsilon}-1} = \frac{\left(\left(\frac{1}{e^{\varepsilon}+1}\right)\cdot(e^{\varepsilon}+1)-1\right)}{e^{\varepsilon}-1} = 0$
  - Finally, $E\left[Z_{i,j,k}\right] = E\left[Y_{i,j}\cdot Y_{j,k}\cdot Y_{i,k}\right] = E\left[Y_{i,j}\right]\cdot E\left[Y_{j,k}\right]\cdot E\left[Y_{i,k}\right] = 1_{i,j,k}$
  - Linearity of expectations gives $E\left[\hat{T}\right] = E\left[\sum_{ijk\in[n]^3} Z_{i,j,k}\right] = T$

Analysis of the Expectation and Variance

- **Lemma:** Returns an approximate $\widehat{\boldsymbol{T}}$ where $\mathrm{Var}[\hat{T}] = \Theta\left(\frac{C_4}{\varepsilon^2} + \frac{n^3}{\varepsilon^6}\right)$
- Proof:
  - $\mathrm{Var}[X_{i,j}]$?

Analysis of the Expectation and Variance

- **Lemma:** Returns an approximate $\widehat{\boldsymbol{T}}$ where $\mathrm{Var}\left[\hat{T}\right] = \Theta\left(\frac{C_4}{\varepsilon^2} + \frac{n^3}{\varepsilon^6}\right)$
- Proof:
  - $\mathrm{Var}\left[X_{i,j}\right] = p(1-p) = \frac{e^{\varepsilon}}{e^{\varepsilon}+1} \cdot \frac{1}{e^{\varepsilon}+1} = \frac{e^{\varepsilon}}{(e^{\varepsilon}+1)^2}$

Bernoulli variable

Analysis of the Expectation and Variance

- **Lemma:** Returns an approximate $\hat{\boldsymbol{T}}$ where $\mathrm{Var}[\hat{T}] = \Theta\left(\frac{C_4}{\varepsilon^2} + \frac{n^3}{\varepsilon^6}\right)$
- Proof:
  - $\mathrm{Var}[X_{i,j}] = p(1-p) = \frac{e^{\varepsilon}}{e^{\varepsilon}+1} \cdot \frac{1}{e^{\varepsilon}+1} = \frac{e^{\varepsilon}}{(e^{\varepsilon}+1)^2}$
  - $\mathrm{Var}[Y_{i,j}] = \mathrm{Var}\left[\frac{\left(X_{i,j}\cdot(e^{\varepsilon}+1)-1\right)}{e^{\varepsilon}-1}\right]$?

Analysis of the Expectation and Variance

- **Lemma:** Returns an approximate $\hat{\boldsymbol{T}}$ where $\mathrm{Var}[\hat{T}] = \Theta\left(\frac{C_4}{\varepsilon^2} + \frac{n^3}{\varepsilon^6}\right)$
- Proof:
  - $\mathrm{Var}[X_{i,j}] = p(1-p) = \frac{e^{\varepsilon}}{e^{\varepsilon}+1} \cdot \frac{1}{e^{\varepsilon}+1} = \frac{e^{\varepsilon}}{(e^{\varepsilon}+1)^2}$
  - $\mathrm{Var}[Y_{i,j}] = \mathrm{Var}\left[\frac{(X_{i,j} \cdot (e^{\varepsilon}+1) - 1)}{e^{\varepsilon}-1}\right] = \frac{1}{(e^{\varepsilon}-1)^2} \cdot \mathrm{Var}[X_{i,j} \cdot (e^{\varepsilon} + 1) - 1] = \frac{(e^{\varepsilon}+1)^2}{(e^{\varepsilon}-1)^2} \cdot \mathrm{Var}[X_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}-1)^2}$

Analysis of the Expectation and Variance

- **Lemma:** Returns an approximate $\widehat{\boldsymbol{T}}$ where $\mathrm{Var}\left[\hat{T}\right] = \Theta\left(\frac{C_4}{\varepsilon^2} + \frac{n^3}{\varepsilon^6}\right)$
- Proof:
  - $\mathrm{Var}\left[X_{i,j}\right] = p(1-p) = \frac{e^{\varepsilon}}{e^{\varepsilon}+1} \cdot \frac{1}{e^{\varepsilon}+1} = \frac{e^{\varepsilon}}{(e^{\varepsilon}+1)^2}$
  - $\mathrm{Var}\left[Y_{i,j}\right] = \mathrm{Var}\left[\frac{\left(X_{i,j}\cdot(e^{\varepsilon}+1)-1\right)}{e^{\varepsilon}-1}\right] = \frac{1}{(e^{\varepsilon}-1)^2} \cdot \mathrm{Var}\left[X_{i,j} \cdot (e^{\varepsilon}+1) - 1\right] = \frac{(e^{\varepsilon}+1)^2}{(e^{\varepsilon}-1)^2} \cdot \mathrm{Var}\left[X_{i,j}\right] = \frac{e^{\varepsilon}}{(e^{\varepsilon}-1)^2}$
  - $\mathrm{Var}\left[Z_{i,j,k}\right] = E\left[Z_{i,j,k}^2\right] - E\left[Z_{i,j,k}\right]^2 = E\left[Y_{i,j}^2\right] \cdot E\left[Y_{j,k}^2\right] \cdot E\left[Y_{i,k}^2\right] - 1_{i,j,k}^2$

Analysis of the Expectation and Variance

- **Lemma:** Returns an approximate $\hat{\boldsymbol{T}}$ where $\mathrm{Var}\left[\hat{T}\right] = \Theta\left(\frac{C_4}{\varepsilon^2} + \frac{n^3}{\varepsilon^6}\right)$
- Proof:
  - $\mathrm{Var}\left[X_{i,j}\right] = \frac{e^{\varepsilon}}{(e^{\varepsilon}+1)^2}$; $\mathrm{Var}\left[Y_{i,j}\right] = \frac{e^{\varepsilon}}{(e^{\varepsilon}-1)^2}$
  - $\mathrm{Var}\left[Z_{i,j,k}\right] = E\left[Z_{i,j,k}^2\right] - E\left[Z_{i,j,k}\right]^2 = E\left[Y_{i,j}^2\right] \cdot E\left[Y_{j,k}^2\right] \cdot E\left[Y_{i,k}^2\right] - 1_{i,j,k}^2$
  - $E\left[Y_{i,j}^2\right] = \mathrm{Var}\left[Y_{i,j}\right] + E\left[Y_{i,j}\right]^2 = \frac{e^{\varepsilon}}{(e^{\varepsilon}-1)^2} = \Theta\left(\frac{1}{\varepsilon^2}\right)$

$\boldsymbol{e^{\varepsilon} \in [1,3)}$ for $\boldsymbol{\varepsilon \in (0,1)}$

Analysis of the Expectation and Variance

- **Lemma:** Returns an approximate $\hat{\boldsymbol{T}}$ where $\mathrm{Var}[\hat{T}] = \Theta\left(\frac{C_4}{\varepsilon^2} + \frac{n^3}{\varepsilon^6}\right)$
- Proof:
  - $\mathrm{Var}[X_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}+1)^2}$; $\mathrm{Var}[Y_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}-1)^2}$
  - $\mathrm{Var}[Z_{i,j,k}] = E[Z_{i,j,k}^2] - E[Z_{i,j,k}]^2 = E[Y_{i,j}^2] \cdot E[Y_{j,k}^2] \cdot E[Y_{i,k}^2] - 1_{i,j,k}^2$
  - $E[Y_{i,j}^2] = \mathrm{Var}[Y_{i,j}] + E[Y_{i,j}]^2 = \frac{e^{\varepsilon}}{(e^{\varepsilon}-1)^2} = \Theta\left(\frac{1}{\varepsilon^2}\right)$
  - $E[Y_{i,j}^2] \cdot E[Y_{j,k}^2] \cdot E[Y_{i,k}^2] - 1_{i,j,k}^2 = \Theta\left(\frac{1}{\varepsilon^6}\right)$

Analysis of the Expectation and Variance

- **Lemma:** Returns an approximate $\hat{\boldsymbol{T}}$ where $\mathrm{Var}\left[\hat{T}\right] = \Theta\left(\frac{C_4}{\varepsilon^2} + \frac{n^3}{\varepsilon^6}\right)$
- Proof:
  - $\mathrm{Var}\left[X_{i,j}\right] = \frac{e^{\varepsilon}}{(e^{\varepsilon}+1)^2}$; $\mathrm{Var}\left[Y_{i,j}\right] = \frac{e^{\varepsilon}}{(e^{\varepsilon}-1)^2}$; $\mathrm{Var}\left[Z_{i,j,k}\right] = \Theta\left(\frac{1}{\varepsilon^6}\right)$

Analysis of the Expectation and Variance

- **Lemma:** Returns an approximate $\widehat{\boldsymbol{T}}$ where $\mathrm{Var}\left[\hat{T}\right] = \Theta\left(\frac{C_4}{\varepsilon^2} + \frac{n^3}{\varepsilon^6}\right)$
- Proof:
  - $\mathrm{Var}\left[X_{i,j}\right] = \frac{e^{\varepsilon}}{(e^{\varepsilon}+1)^2}$; $\mathrm{Var}\left[Y_{i,j}\right] = \frac{e^{\varepsilon}}{(e^{\varepsilon}-1)^2}$; $\mathrm{Var}\left[Z_{i,j,k}\right] = \Theta\left(\frac{1}{\varepsilon^6}\right)$
  - $\mathrm{Var}\left[\hat{T}\right] = \sum_{\{i,j,k\} \in \binom{[n]}{3}} \mathrm{Var}\left[Z_{i,j,k}\right]$

Analysis of the Expectation and Variance

- **Lemma:** Returns an approximate $\widehat{\boldsymbol{T}}$ where $\mathrm{Var}[\hat{T}] = \Theta\left(\frac{C_4}{\varepsilon^2} + \frac{n^3}{\varepsilon^6}\right)$
- Proof:
  - $\mathrm{Var}[X_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}+1)^2}$; $\mathrm{Var}[Y_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}-1)^2}$; $\mathrm{Var}[Z_{i,j,k}] = \Theta\left(\frac{1}{\varepsilon^6}\right)$
  - $\mathrm{Var}[\hat{T}] = \mathrm{Var}\left[\sum_{\{i,j,k\}\in\binom{[n]}{3}} Z_{i,j,k}\right]$
    - Are we done?

Analysis of the Expectation and Variance

- **Lemma:** Returns an approximate $\widehat{\boldsymbol{T}}$ where $\mathrm{Var}[\hat{T}] = \Theta\left(\frac{C_4}{\varepsilon^2} + \frac{n^3}{\varepsilon^6}\right)$
- Proof:
  - $\mathrm{Var}[X_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}+1)^2}$; $\mathrm{Var}[Y_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}-1)^2}$; $\mathrm{Var}[Z_{i,j,k}] = \Theta\left(\frac{1}{\varepsilon^6}\right)$
  - $\mathrm{Var}[\hat{T}] = \mathrm{Var}\left[\sum_{\{i,j,k\}\in\binom{[n]}{3}} Z_{i,j,k}\right]$
    - Are we done?

Not true $\mathrm{Var}\left[\sum_{\{i,j,k\}\in\binom{[n]}{3}} Z_{i,j,k}\right] = \sum_{\{i,j,k\}\in\binom{[n]}{3}} \mathrm{Var}[Z_{i,j,k}]$**, why?**

Analysis of the Expectation and Variance

- **Lemma:** Returns an approximate $\hat{\boldsymbol{T}}$ where $\text{Var}[\hat{T}] = \Theta\left(\frac{C_4}{\varepsilon^2} + \frac{n^3}{\varepsilon^6}\right)$
- Proof:
  - $\text{Var}[X_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}+1)^2}$; $\text{Var}[Y_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}-1)^2}$; $\text{Var}[Z_{i,j,k}] = \Theta\left(\frac{1}{\varepsilon^6}\right)$
  - $\text{Var}[\hat{T}] = \text{Var}\left[\sum_{\{i,j,k\}\in\binom{[n]}{3}} Z_{i,j,k}\right]$
    - Are we done?

Share $\boldsymbol{Y_{i,j}}$ variables; triangles share edges!

Not true $\text{Var}\left[\sum_{\{i,j,k\}\in\binom{[n]}{3}} Z_{i,j,k}\right] = \sum_{\{i,j,k\}\in\binom{[n]}{3}} \text{Var}[Z_{i,j,k}]$**, why?**

Analysis of the Expectation and Variance

- **Lemma:** Returns an approximate $\hat{\boldsymbol{T}}$ where $\mathrm{Var}[\hat{T}] = \Theta\left(\frac{C_4}{\varepsilon^2} + \frac{n^3}{\varepsilon^6}\right)$
- Proof:
  - $\mathrm{Var}[X_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}+1)^2}$; $\mathrm{Var}[Y_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}-1)^2}$; $\mathrm{Var}[Z_{i,j,k}] = \Theta\left(\frac{1}{\varepsilon^6}\right)$
  - $\mathrm{Var}[\hat{T}] = \mathrm{Var}\left[\sum_{\{i,j,k\}\in\binom{[n]}{3}} Z_{i,j,k}\right]$
    - Are we done?

Share $Y_{i,j}$ variables; triangles share edges!

Not true $\mathrm{Var}\left[\sum_{\{i,j,k\}\in\binom{[n]}{3}} Z_{i,j,k}\right] = \sum_{\{i,j,k\}\in\binom{[n]}{3}} \mathrm{Var}[Z_{i,j,k}]$, **why?**

Analysis of the Expectation and Variance

- **Lemma:** Returns an approximate $\hat{\boldsymbol{T}}$ where $\mathrm{Var}[\hat{T}] = \Theta\left(\frac{C_4}{\varepsilon^2} + \frac{n^3}{\varepsilon^6}\right)$
- Proof:
  - $\mathrm{Var}[X_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}+1)^2}$; $\mathrm{Var}[Y_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}-1)^2}$; $\mathrm{Var}[Z_{i,j,k}] = \Theta\left(\frac{1}{\varepsilon^6}\right)$
  - $\mathrm{Var}[\hat{T}] = \mathrm{Var}\left[\sum_{\{i,j,k\}\in\binom{[n]}{3}} Z_{i,j,k}\right]$
    - Consider change of variables $U_{i,j,k} = Z_{i,j,k} - 1_{i,j,k}$

Analysis of the Expectation and Variance

- **Lemma:** Returns an approximate $\hat{\boldsymbol{T}}$ where $\mathrm{Var}[\hat{T}] = \Theta\left(\frac{C_4}{\varepsilon^2} + \frac{n^3}{\varepsilon^6}\right)$
- Proof:
  - $\mathrm{Var}[X_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}+1)^2}$; $\mathrm{Var}[Y_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}-1)^2}$; $\mathrm{Var}[Z_{i,j,k}] = \Theta\left(\frac{1}{\varepsilon^6}\right)$
  - $\mathrm{Var}[\hat{T}] = \mathrm{Var}\left[\sum_{\{i,j,k\}\in\binom{[n]}{3}} Z_{i,j,k}\right]$
    - Consider change of variables $U_{i,j,k} = Z_{i,j,k} - 1_{i,j,k}$

$$\boldsymbol{E[U_{i,j,k}] = 0 \text{ and } \mathrm{Var}[U_{i,j,k}] = \mathrm{Var}[Z_{i,j,k}]}$$

Analysis of the Expectation and Variance

- **Lemma:** Returns an approximate $\hat{\boldsymbol{T}}$ where $\mathrm{Var}[\hat{T}] = \Theta\left(\frac{C_4}{\varepsilon^2} + \frac{n^3}{\varepsilon^6}\right)$
- Proof:
  - $\mathrm{Var}[X_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}+1)^2}$; $\mathrm{Var}[Y_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}-1)^2}$; $\mathrm{Var}[Z_{i,j,k}] = \Theta\left(\frac{1}{\varepsilon^6}\right)$
  - $\mathrm{Var}[\hat{T} - T] = \mathrm{Var}\left[\sum_{\{i,j,k\}\in\binom{[n]}{3}}(Z_{i,j,k} - 1_{i,j,k})\right] = \mathrm{Var}\left[\sum_{\{i,j,k\}\in\binom{[n]}{3}} U_{i,j,k}\right]$

Analysis of the Expectation and Variance

- **Lemma:** Returns an approximate $\hat{\boldsymbol{T}}$ where $\mathrm{Var}[\hat{T}] = \Theta\left(\frac{C_4}{\varepsilon^2} + \frac{n^3}{\varepsilon^6}\right)$
- Proof:
  - $\mathrm{Var}[X_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}+1)^2}$; $\mathrm{Var}[Y_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}-1)^2}$; $\mathrm{Var}[Z_{i,j,k}] = \Theta\left(\frac{1}{\varepsilon^6}\right)$
  - $\mathrm{Var}[\hat{T} - T] = \mathrm{Var}\left[\sum_{\{i,j,k\}\in\binom{[n]}{3}}(Z_{i,j,k} - 1_{i,j,k})\right] = \mathrm{Var}\left[\sum_{\{i,j,k\}\in\binom{[n]}{3}} U_{i,j,k}\right]$
    - How do we simplify, what do we observe?

Analysis of the Expectation and Variance

- **Lemma:** Returns an approximate $\widehat{\boldsymbol{T}}$ where $\mathrm{Var}[\hat{T}] = \Theta\left(\frac{C_4}{\varepsilon^2} + \frac{n^3}{\varepsilon^6}\right)$
- Proof:
  - $\mathrm{Var}[X_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}+1)^2}$; $\mathrm{Var}[Y_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}-1)^2}$; $\mathrm{Var}[Z_{i,j,k}] = \Theta\left(\frac{1}{\varepsilon^6}\right)$
  - $\mathrm{Var}[\hat{T} - T] = \mathrm{Var}\left[\sum_{\{i,j,k\}\in\binom{[n]}{3}}(Z_{i,j,k} - 1_{i,j,k})\right] = \mathrm{Var}\left[\sum_{\{i,j,k\}\in\binom{[n]}{3}} U_{i,j,k}\right]$
    - How do we simplify, what do we observe?

Covariance is 0 if share at most one vertex, no edges

Analysis of the Expectation and Variance

- **Lemma:** Returns an approximate $\hat{T}$ where $\mathrm{Var}[\hat{T}] = \Theta\left(\frac{C_4}{\varepsilon^2} + \frac{n^3}{\varepsilon^6}\right)$
- Proof:
  - $\mathrm{Var}[X_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}+1)^2}$; $\mathrm{Var}[Y_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}-1)^2}$; $\mathrm{Var}[Z_{i,j,k}] = \Theta\left(\frac{1}{\varepsilon^6}\right)$
  - $\mathrm{Var}[\hat{T} - T] = \mathrm{Var}\left[\sum_{\{i,j,k\} \in \binom{[n]}{3}} (Z_{i,j,k} - 1_{i,j,k})\right] = \mathrm{Var}\left[\sum_{\{i,j,k\} \in \binom{[n]}{3}} U_{i,j,k}\right]$
    - How do we simplify, what do we observe?

Covariance is 0 if share at most one vertex, no edges

Non-zero covariance: share an edge; how many?

Analysis of the Expectation and Variance

- **Lemma:** Returns an approximate $\boldsymbol{\hat{T}}$ where $\mathrm{Var}[\hat{T}] = \Theta\left(\frac{C_4}{\varepsilon^2} + \frac{n^3}{\varepsilon^6}\right)$
- Proof:
  - $\mathrm{Var}[X_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}+1)^2}$; $\mathrm{Var}[Y_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}-1)^2}$; $\mathrm{Var}[Z_{i,j,k}] = \Theta\left(\frac{1}{\varepsilon^6}\right)$
  - $\mathrm{Var}[\hat{T} - T] = \mathrm{Var}\left[\sum_{\{i,j,k\}\in\binom{[n]}{3}}(Z_{i,j,k} - 1_{i,j,k})\right] = \mathrm{Var}\left[\sum_{\{i,j,k\}\in\binom{[n]}{3}} U_{i,j,k}\right]$
    - How do we simplify, what do we observe?

Covariance is 0 if share at most one vertex, no edges

Non-zero covariance: share an edge; number of 4-cycles

Analysis of the Expectation and Variance

- **Lemma:** Returns an approximate $\widehat{\boldsymbol{T}}$ where $\mathrm{Var}\left[\hat{T}\right] = \Theta\left(\frac{C_4}{\varepsilon^2} + \frac{n^3}{\varepsilon^6}\right)$
- Proof:
  - $\mathrm{Var}\left[X_{i,j}\right] = \frac{e^{\varepsilon}}{(e^{\varepsilon}+1)^2}$; $\mathrm{Var}\left[Y_{i,j}\right] = \frac{e^{\varepsilon}}{(e^{\varepsilon}-1)^2}$; $\mathrm{Var}\left[Z_{i,j,k}\right] = \Theta\left(\frac{1}{\varepsilon^6}\right)$
  - $\mathrm{Var}\left[\hat{T}\right] = \mathrm{Var}\left[\hat{T} - T\right] = \mathrm{Var}\left[\sum_{\{i,j,k\}\in\binom{[n]}{3}}(Z_{i,j,k} - 1_{i,j,k})\right] =$

    $\mathrm{Var}\left[\sum_{\{i,j,k\}\in\binom{[n]}{3}} U_{i,j,k}\right] \le \sum_{\{i,j,k\}\in\binom{[n]}{3}} \Theta\left(\frac{1}{\varepsilon^6}\right)$

    $$+ \sum_{\{i,j,k,l\}\in C_4} E\left[U_{i,j,k} \cdot U_{j,k,l}\right]$$

Analysis of the Expectation and Variance

- **Lemma:** Returns an approximate $\hat{\boldsymbol{T}}$ where $\mathrm{Var}[\hat{T}] = \Theta\left(\frac{C_4}{\varepsilon^2} + \frac{n^3}{\varepsilon^6}\right)$
- Proof:
  - $\mathrm{Var}[X_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}+1)^2}$; $\mathrm{Var}[Y_{i,j}] = \frac{e^{\varepsilon}}{(e^{\varepsilon}-1)^2}$; $\mathrm{Var}[Z_{i,j,k}] = \Theta\left(\frac{1}{\varepsilon^6}\right)$
  - $\mathrm{Var}[\hat{T}] = \mathrm{Var}[\hat{T} - T] = \mathrm{Var}\left[\sum_{\{i,j,k\}\in\binom{[n]}{3}}(Z_{i,j,k} - 1_{i,j,k})\right] =$

$$\mathrm{Var}\left[\sum_{\{i,j,k\}\in\binom{[n]}{3}} U_{i,j,k}\right] \leq \sum_{\{i,j,k\}\in\binom{[n]}{3}} \Theta\left(\frac{1}{\varepsilon^6}\right) + \sum_{\{i,j,k,l\}\in C_4} E[U_{i,j,k} \cdot U_{j,k,l}] \leq \sum_{\{i,j,k,l\}\in C_4} E[Y_{i,j} \cdot Y_{j,k}^2 \cdot Y_{i,k} \cdot Y_{l,j} \cdot Y_{l,k}]$$

$$\leq \sum_{\{i,j,k,l\}\in C_4} E[Y_{j,k}^2]$$

Analysis of the Expectation and Variance

- **Lemma:** Returns an approximate $\hat{\boldsymbol{T}}$ where $\mathrm{Var}\left[\hat{T}\right] = \Theta\left(\frac{C_4}{\varepsilon^2} + \frac{n^3}{\varepsilon^6}\right)$
- Proof:
  - $\mathrm{Var}\left[X_{i,j}\right] = \frac{e^{\varepsilon}}{(e^{\varepsilon}+1)^2}$; $\mathrm{Var}\left[Y_{i,j}\right] = \frac{e^{\varepsilon}}{(e^{\varepsilon}-1)^2}$; $\mathrm{Var}\left[Z_{i,j,k}\right] = \Theta\left(\frac{1}{\varepsilon^6}\right)$
  - $\mathrm{Var}\left[\hat{T}\right] = \mathrm{Var}\left[\hat{T} - T\right] = \mathrm{Var}\left[\sum_{\{i,j,k\}\in\binom{[n]}{3}} (Z_{i,j,k} - 1_{i,j,k})\right] =$

$$\mathrm{Var}\left[\sum_{\{i,j,k\}\in\binom{[n]}{3}} U_{i,j,k}\right] \le \sum_{\{i,j,k\}\in\binom{[n]}{3}} \Theta\left(\frac{1}{\varepsilon^6}\right)$$

$$+ \sum_{\{i,j,k,l\}\in C_4} E\left[U_{i,j,k} \cdot U_{j,k,l}\right] \le \Theta\left(\frac{n^3}{\varepsilon^6}\right) + \sum_{\{i,j,k,l\}\in C_4} E\left[Y_{j,k}^2\right] = \Theta\left(\frac{n^3}{\varepsilon^6} + \frac{|C_4|}{\varepsilon^2}\right)$$

On Wednesday, more DP mechanisms!

- Laplace mechanism
- (Geometric mechanism—already discussed)
- Exponential mechanism
- Gaussian mechanism
- Privacy amplification via subsampling

CPSC 768:
Scalable and Private Graph Algorithms

Lecture 12: Differential Privacy Mechanisms

Quanquan C. Liu
quanquan.liu@yale.edu

Announcements

- Check the latest announcement on Canvas:
  - Scheduling Lectures survey: due **Feb. 26**
  - Final Project Proposal: due **Feb. 29, one page (email to me)**
  - Final Project Examples

- Open problem sessions:
  - Link for joining CPSC 768 Slack
  - Open Problem Session food orders

Global Sensitivity

- **Intuition:** Measure of how different the output of a function is on neighboring input

Global Sensitivity

- **Intuition:** Measure of how different the output of a function is on neighboring input

Definition (Global Sensitivity): The global sensitivity of a function: $f: G \rightarrow R$ is defined as:

$$\Delta_f = \max_{\{G \sim G'\}} (|f(G) - f(G')|)$$

where G and G' are edge-neighboring graphs.

Global Sensitivity Examples

Definition (Global Sensitivity): The global sensitivity of a function: $f: G \rightarrow R$ is defined as:

$$\Delta_f = \max_{\{G \sim G'\}} (|f(G) - f(G')|)$$

where G and G' are edge-neighboring graphs.

- What is the global sensitivity of the following problems (assuming for each we have a function that gives the exact solution):
  - Triangle counting
  - Maximum matching
  - Average Degree

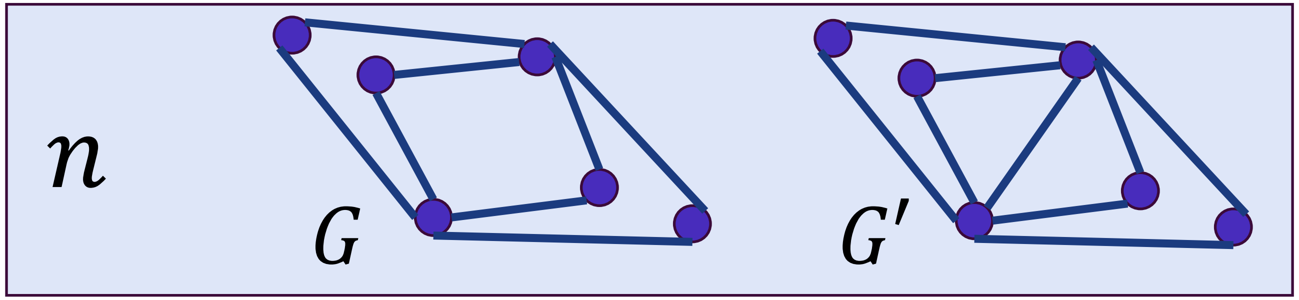
Global Sensitivity Examples

Definition (Global Sensitivity): The global sensitivity of a function: $f: G \rightarrow R$ is defined as:

$$\Delta_f = \max_{\{G \sim G'\}} (|f(G) - f(G')|)$$

where G and G' are edge-neighboring graphs.

- What is the global sensitivity of the following problems (assuming for each we have a function that gives the exact solution):
  - Triangle counting
  - Maximum matching
  - Average Degree



Global Sensitivity Examples

Definition (Global Sensitivity): The global sensitivity of a function: $f: G \rightarrow R$ is defined as:

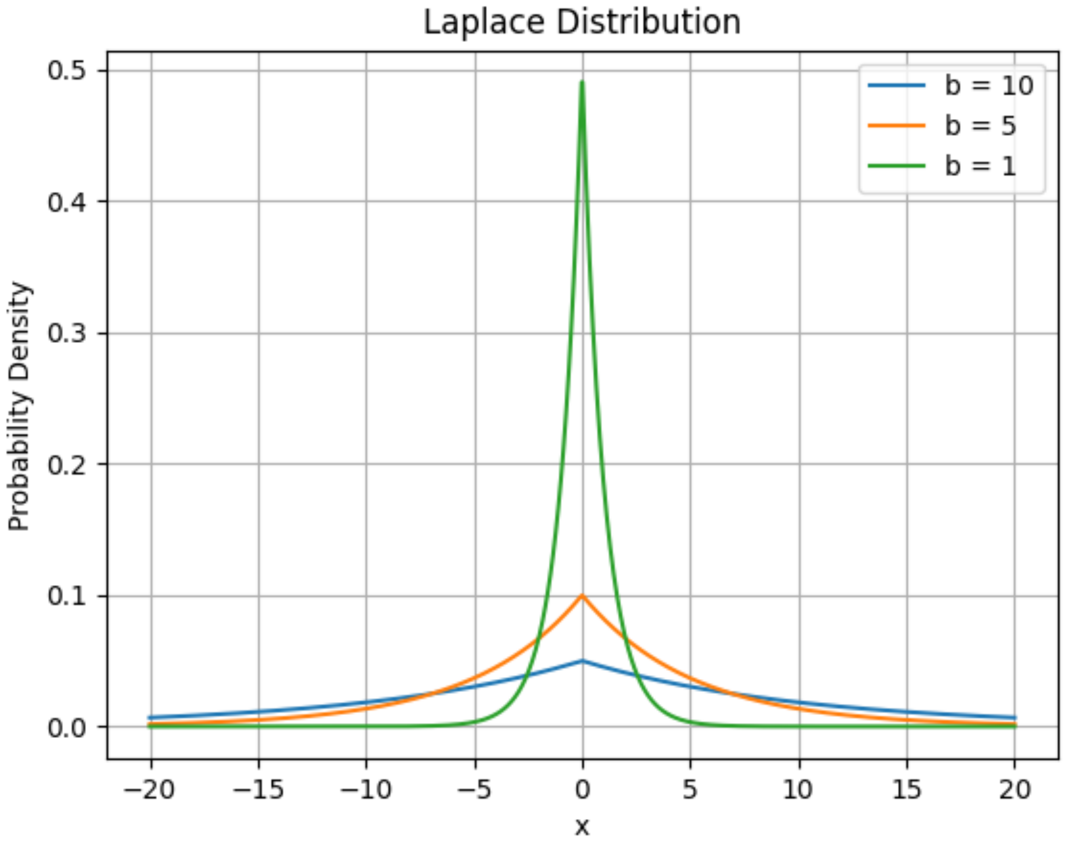
$$\Delta_f = \max_{\{G \sim G'\}} (|f(G) - f(G')|)$$

where G and G' are edge-neighboring graphs.

- What is the global sensitivity of the following problems (assuming for each we have a function that gives the exact solution):
  - Triangle counting: $\boldsymbol{n}$
  - Maximum matching: $\mathbf{1}$
  - Average Degree: $\mathbf{2}/\boldsymbol{n}$

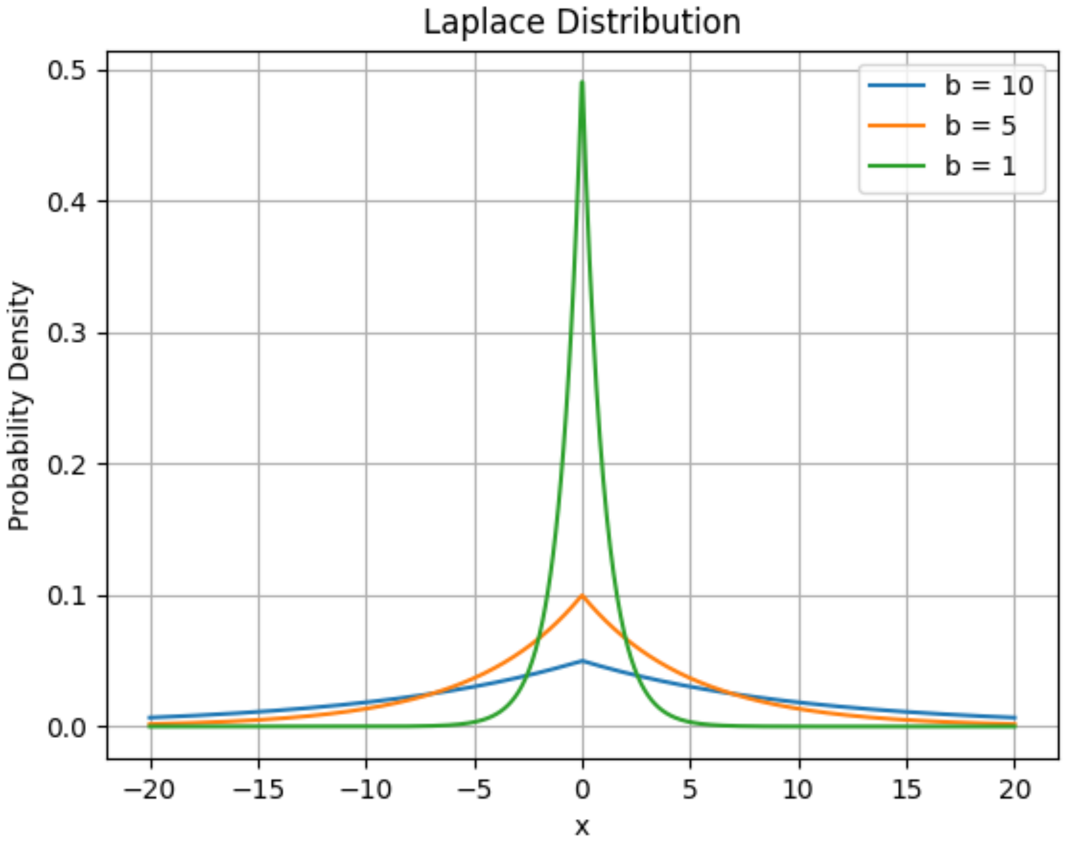
Laplace Distribution

- The PDF of $X \in R$ is: $\mathrm{Lap}(b) = \frac{1}{2b} \cdot \exp\left(-\left(\frac{|X|}{b}\right)\right)$



Laplace Distribution

- The PDF of $X \in R$ is: $\text{Lap}(b) = \frac{1}{2b} \cdot \exp\left(-\left(\frac{|X|}{b}\right)\right)$



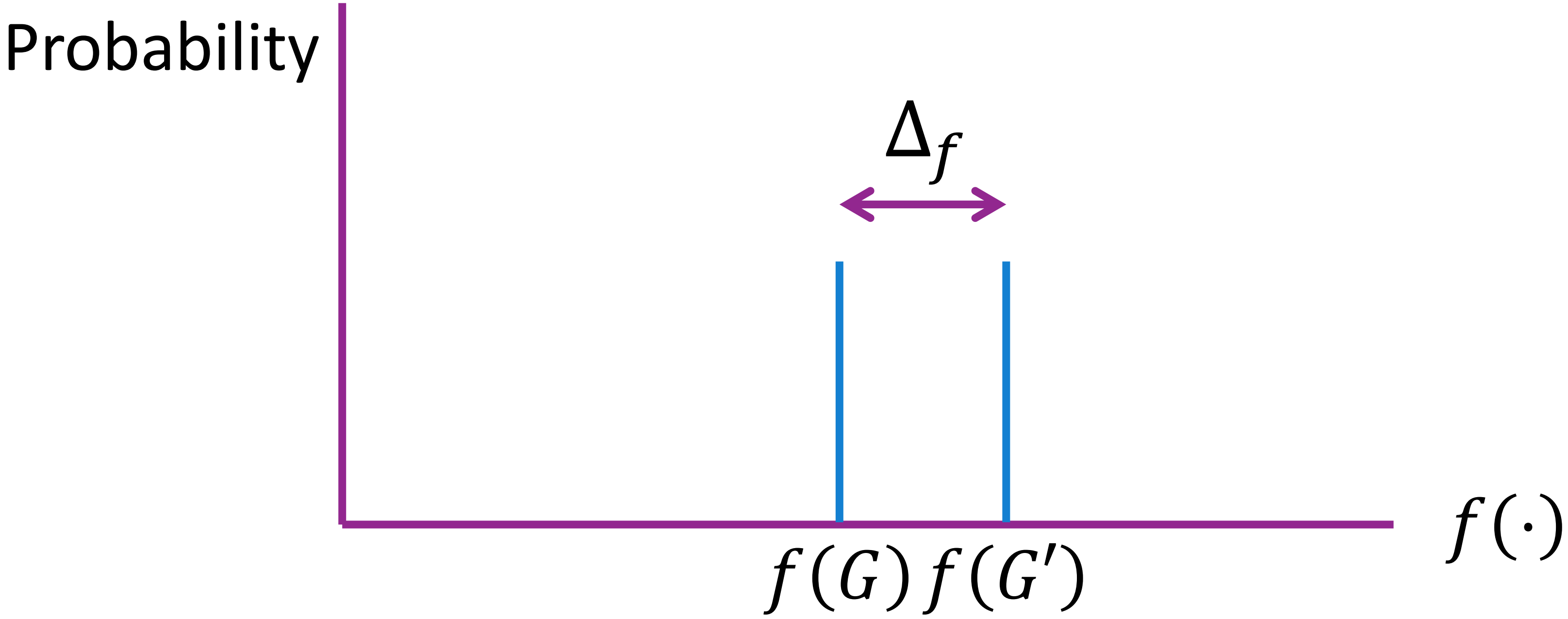
Larger b
heavier tails

Laplace Mechanism

- Lemma: Given a function $f: G \rightarrow R$ with sensitivity Δ_f,

$$\boldsymbol{f(G) + Lap\left(\frac{\Delta_f}{\varepsilon}\right)}$$ is ε-differentially private.

- Intuition: Solution for G and G' differ by Δ_f

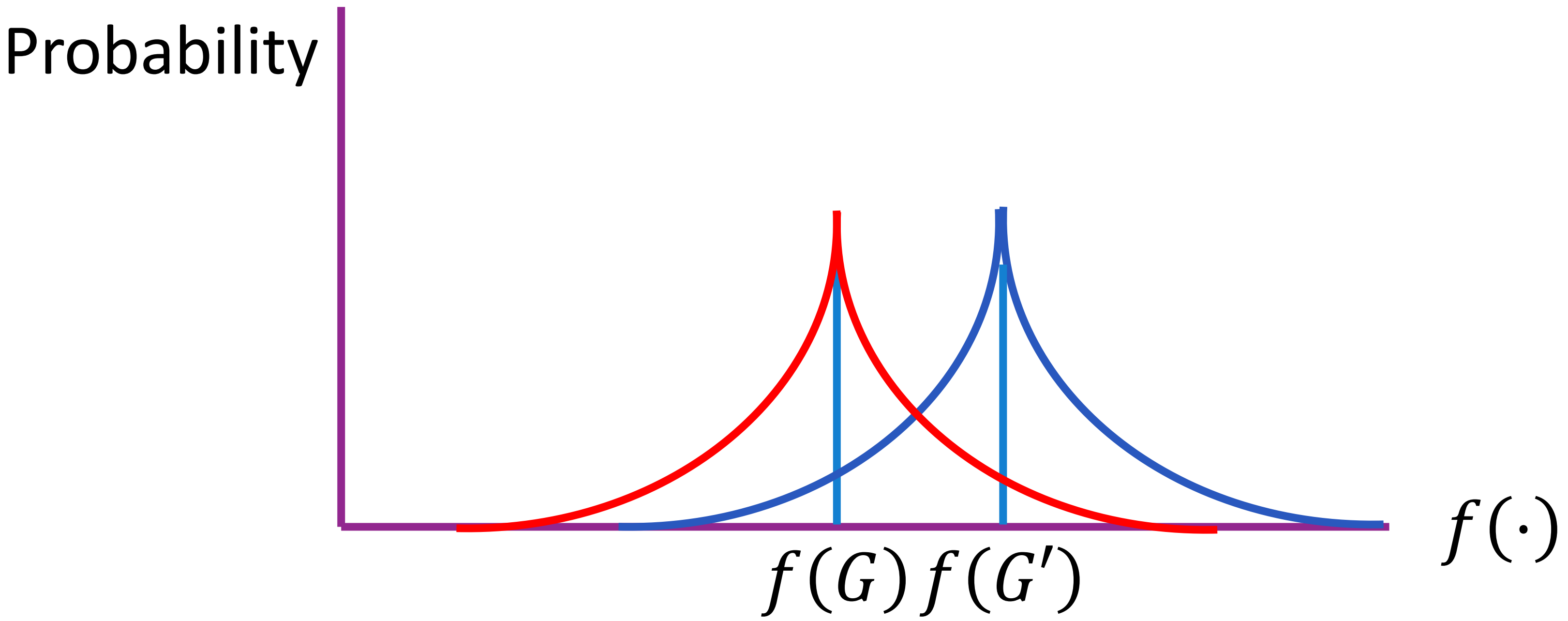


Laplace Mechanism

- Lemma: Given a function $f: G \to R$ with sensitivity Δ_f,

$$\boldsymbol{f(G) + Lap\left(\frac{\Delta_f}{\varepsilon}\right)} \text{ is } \varepsilon\text{-differentially private.}$$

- Intuition: Solution for G and G' differ by Δ_f

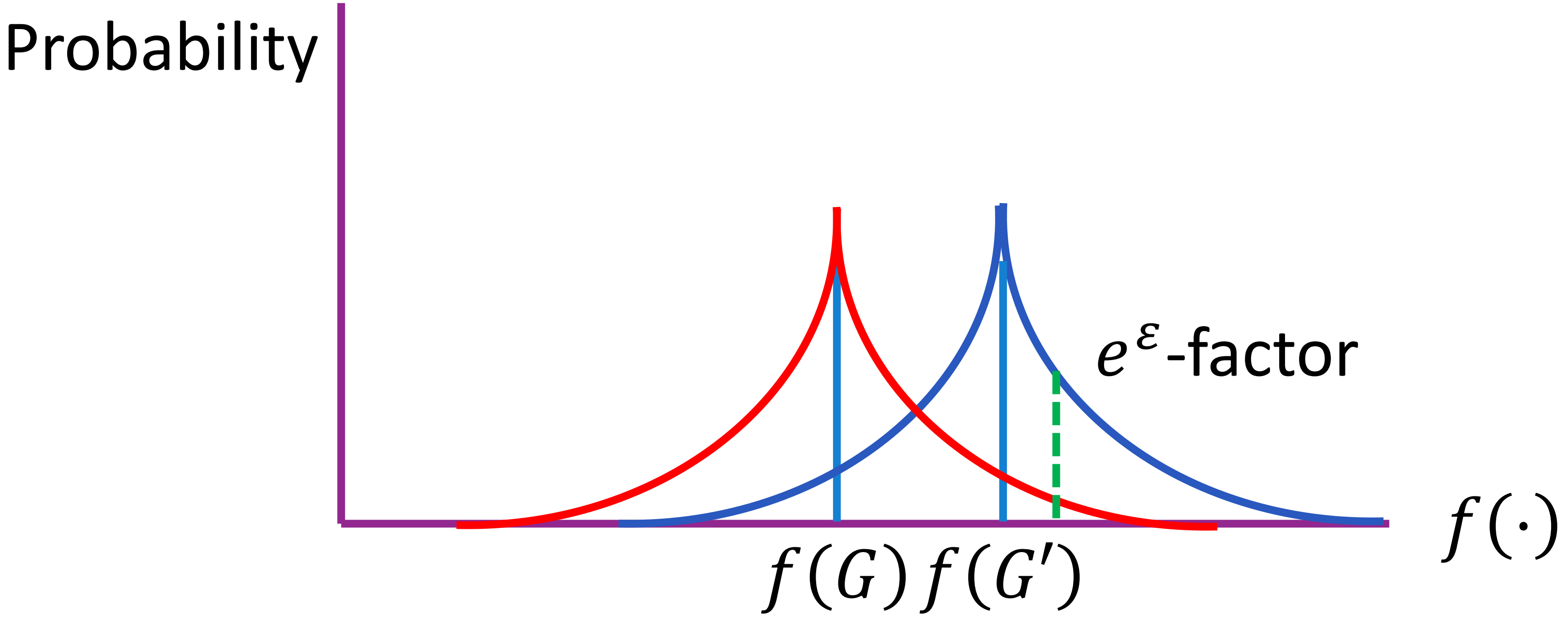


Laplace Mechanism

- Lemma: Given a function $f: G \rightarrow R$ with sensitivity Δ_f,

$$\boldsymbol{f(G) + Lap\left(\frac{\Delta_f}{\varepsilon}\right)} \text{ is } \varepsilon\text{-differentially private.}$$

- Intuition: Solution for G and G' differ by Δ_f

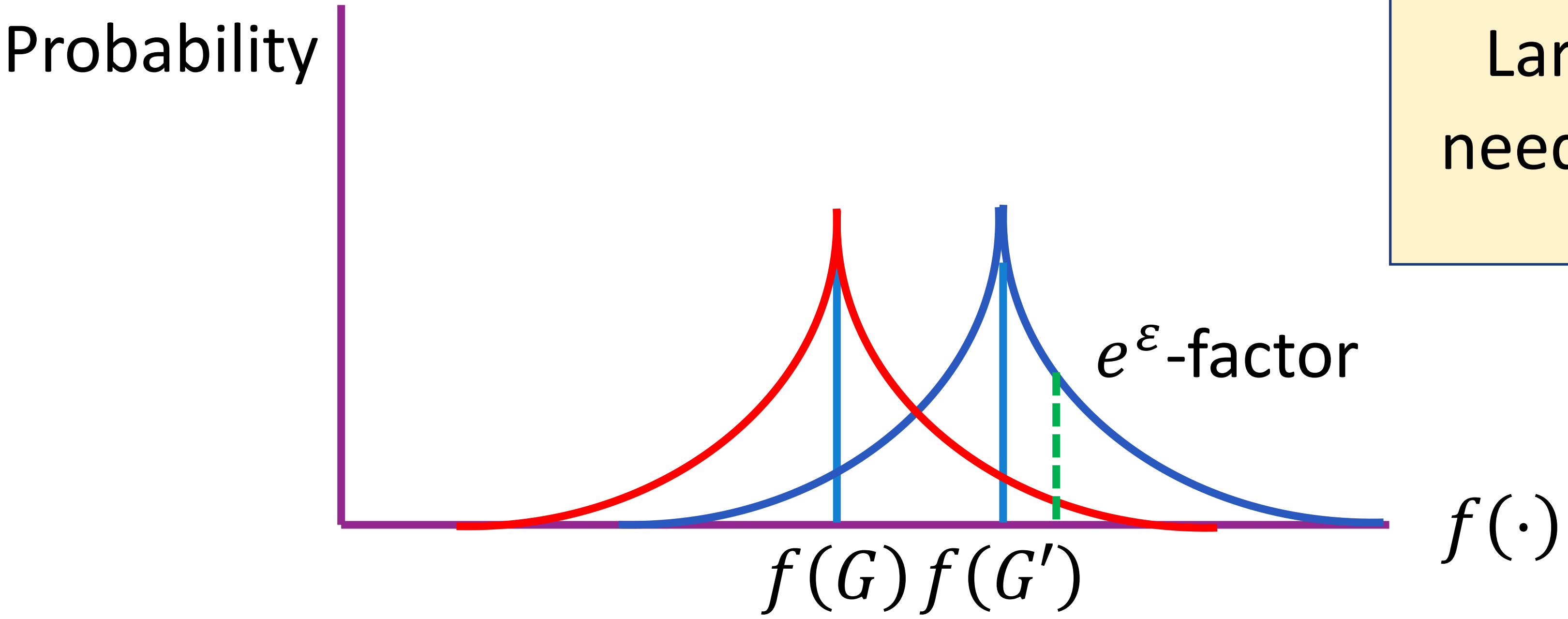


Laplace Mechanism

- Lemma: Given a function $f: G \rightarrow R$ with sensitivity Δ_f,

  $\boldsymbol{f(G) + Lap\left(\frac{\Delta_f}{\varepsilon}\right)}$ is ε-differentially private.

- Intuition: Solution for G and G' differ by Δ_f



Larger Δ_f means need a flatter curve

Laplace Mechanism

- Lemma: Given a function $f: G \rightarrow R$ with sensitivity Δ_f,

$$\boldsymbol{f(G) + Lap\left(\frac{\Delta_f}{\varepsilon}\right)}$$ is ε-differentially private.

- Proof: Consider neighboring graphs G and G' and function $f: G \rightarrow R$. Let $\boldsymbol{p}$ and $\boldsymbol{p'}$ denote the probability density functions of $\boldsymbol{f(G) + Lap\left(\frac{\Delta_f}{\varepsilon}\right)}$ and $\boldsymbol{f(G') + Lap\left(\frac{\Delta_f}{\varepsilon}\right)}$, respectively. Then, for an arbitrary point $z \in R$:

$$\frac{p(z)}{p'(z)} = \frac{\exp\left(-\frac{\varepsilon|f(G) - z|}{\Delta_f}\right)}{\exp\left(-\frac{\varepsilon|f(G') - z|}{\Delta_f}\right)}$$

Laplace Mechanism

- Lemma: Given a function $f: G \rightarrow R$ with sensitivity Δ_f, $\boldsymbol{f(G)} + \boldsymbol{Lap\left(\frac{\Delta_f}{\varepsilon}\right)}$ is ε-differentially private.
- Proof:

$$= \exp\left(-\frac{\varepsilon(|f(G') - z| - |f(G) - z|)}{\Delta_f}\right)$$

$$\leq \exp\left(-\frac{\varepsilon(|f(G') - f(G)|)}{\Delta_f}\right)$$

$$\leq \exp\left(-\frac{\varepsilon\Delta_f}{\Delta_f}\right) = \exp(\varepsilon)$$

By the triangle inequality

Laplace Mechanism Accuracy

- Lemma: Given a function $f: G \rightarrow R$ with sensitivity Δ_f, let $\boldsymbol{M(G) = f(G) + Lap\left(\frac{\Delta_f}{\varepsilon}\right)}$, then $\boldsymbol{P\left(|M(G) - f(G)| \leq \frac{\Delta_f}{\varepsilon} \log\left(\frac{1}{\delta}\right)\right) \geq 1 - \delta}$

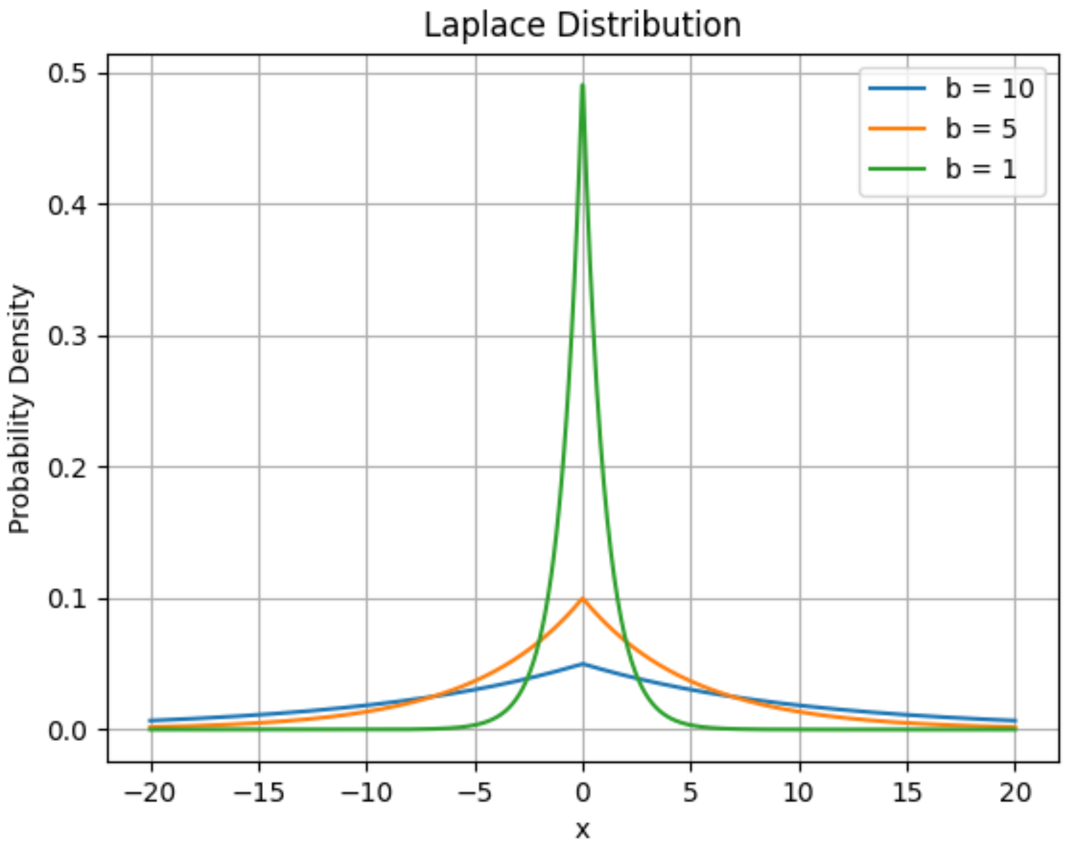
Laplace Mechanism Accuracy

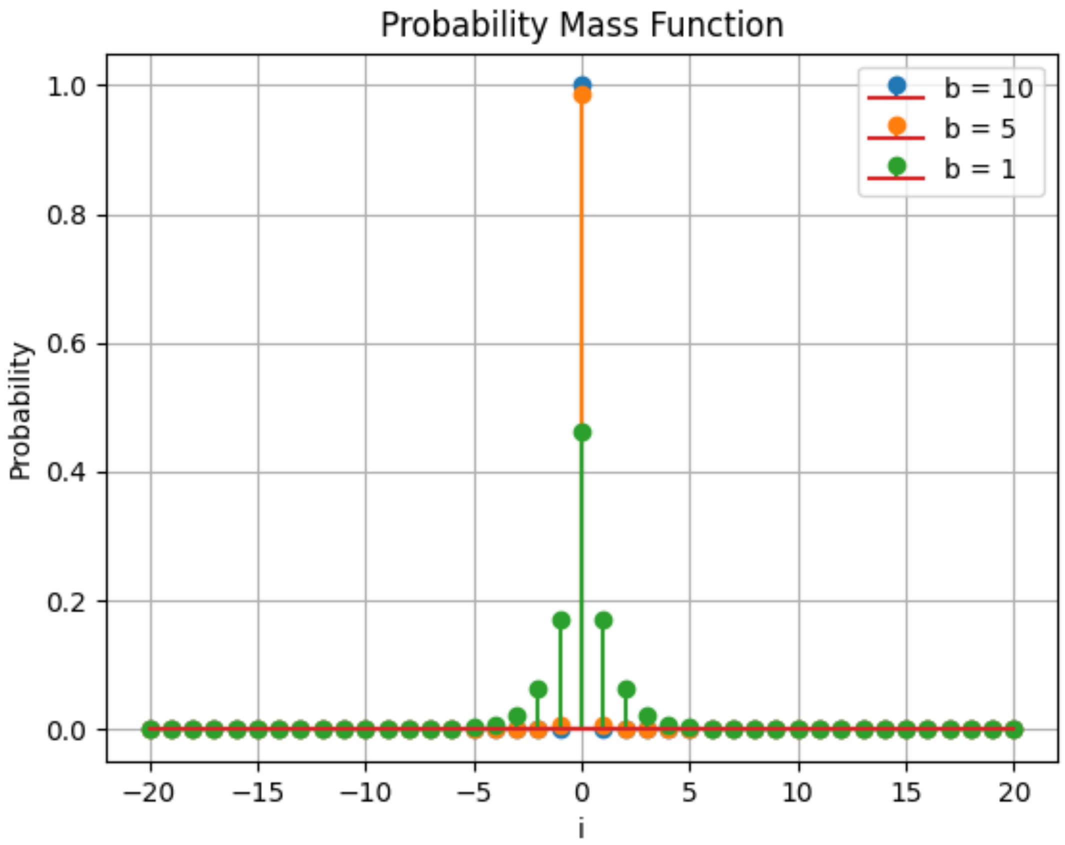
- Lemma: Given a function $f: G \to R$ with sensitivity Δ_f, let $\boldsymbol{M(G) = f(G) + Lap\left(\frac{\Delta_f}{\varepsilon}\right)}$, then $\boldsymbol{P\left(|M(G) - f(G)| \le \frac{\Delta_f}{\varepsilon} \log\left(\frac{1}{\delta}\right)\right) \ge 1 - \delta}$
- Proof: We can simplify $\boldsymbol{P\left(|M(G) - f(G)| \le \frac{\Delta_f}{\varepsilon} \log\left(\frac{1}{\delta}\right)\right) = P\left(Lap\left(\frac{\Delta_f}{\varepsilon}\right) \le \frac{\Delta_f}{\varepsilon} \log\left(\frac{1}{\delta}\right)\right) = 1 - \exp\left(-\ln\left(\frac{1}{\delta}\right)\right) = 1 - \delta}$.

By the Laplace distribution, we have that
$P(|X| \ge bt) = \exp(-t)$

Laplace Mechanism vs. Geometric Mechanism

- Symmetric geometric distribution PMF at $x \in Z$: $\frac{e^{b}-1}{e^{b}+1} \cdot e^{-|i| \cdot b}$





Laplace Mechanism vs. Geometric Mechanism

- Geometric mechanism allows:
  - Integer values

Laplace Mechanism vs. Geometric Mechanism

- Geometric mechanism allows:
  - Integer values
  - Easier analysis for theoretical computer scientists; discrete vs. continuous

Laplace Mechanism vs. Geometric Mechanism

- Geometric mechanism allows:
  - Integer values
  - Easier analysis for theoretical computer scientists; discrete vs. continuous
  - Geometric mechanism is utility-maximizing [Ghosh-Roughgarden-Sundararajan ‘12]

Laplace Mechanism vs. Geometric Mechanism

- Geometric mechanism allows:
  - Integer values
  - Easier analysis for theoretical computer scientists; discrete vs. continuous
  - Geometric mechanism is utility-maximizing [Ghosh-Roughgarden-Sundararajan ‘12]
    - Statistical database and count queries
    - For each fixed count query, there exists a geometric mechanism M^* such that **each** user derives as much utility as a mechanism optimally tailored to that user

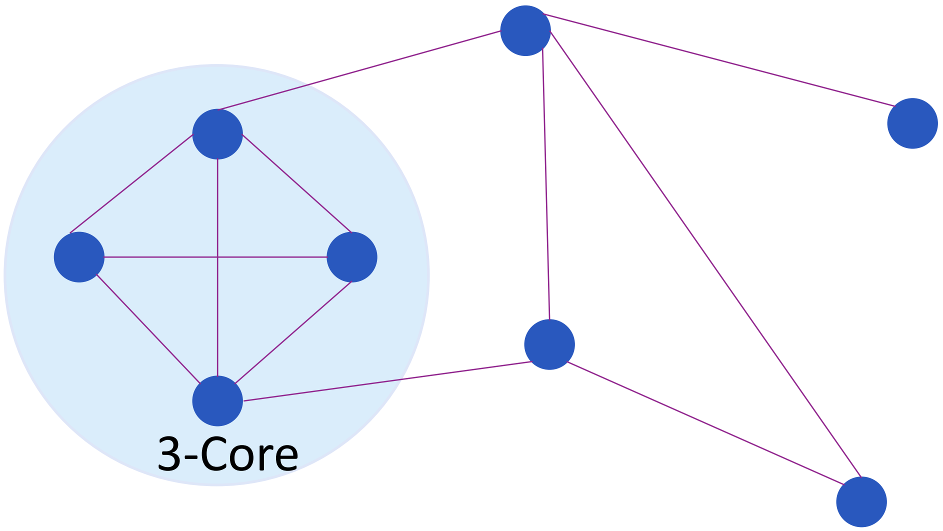
Laplace Mechanism vs. Geometric Mechanism

Implementable on finite-bit computers!
[Balcer and Vadhan 2018]

- Geometric mechanism allows:
  - Integer values
  - Easier analysis for theoretical computer scientists; discrete vs. continuous
  - Geometric mechanism is utility-maximizing [Ghosh-Roughgarden-Sundararajan ‘12]
    - Statistical database and count queries
    - For each fixed count query, there exists a geometric mechanism M^* such that **each** user derives as much utility as a mechanism optimally tailored to that user

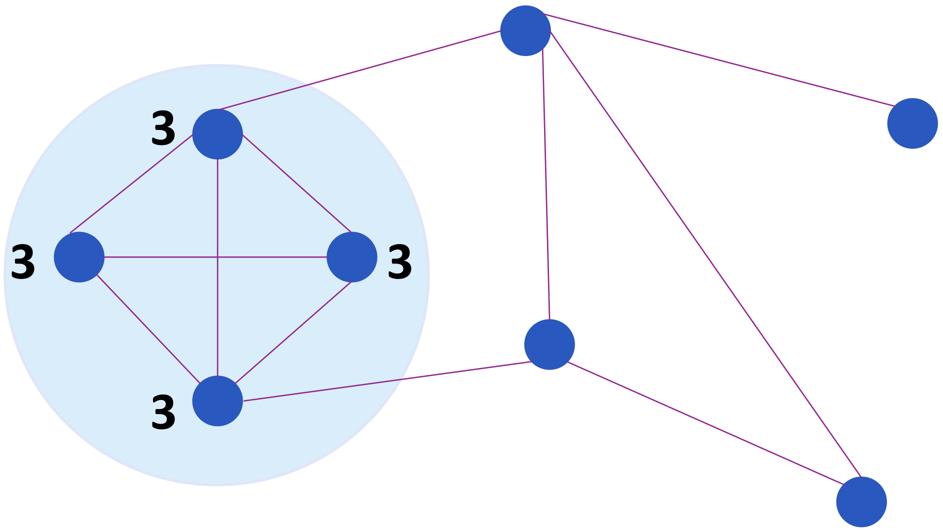
Next: Example of Geometric Mechanism that also gets **local privacy**

k-Core



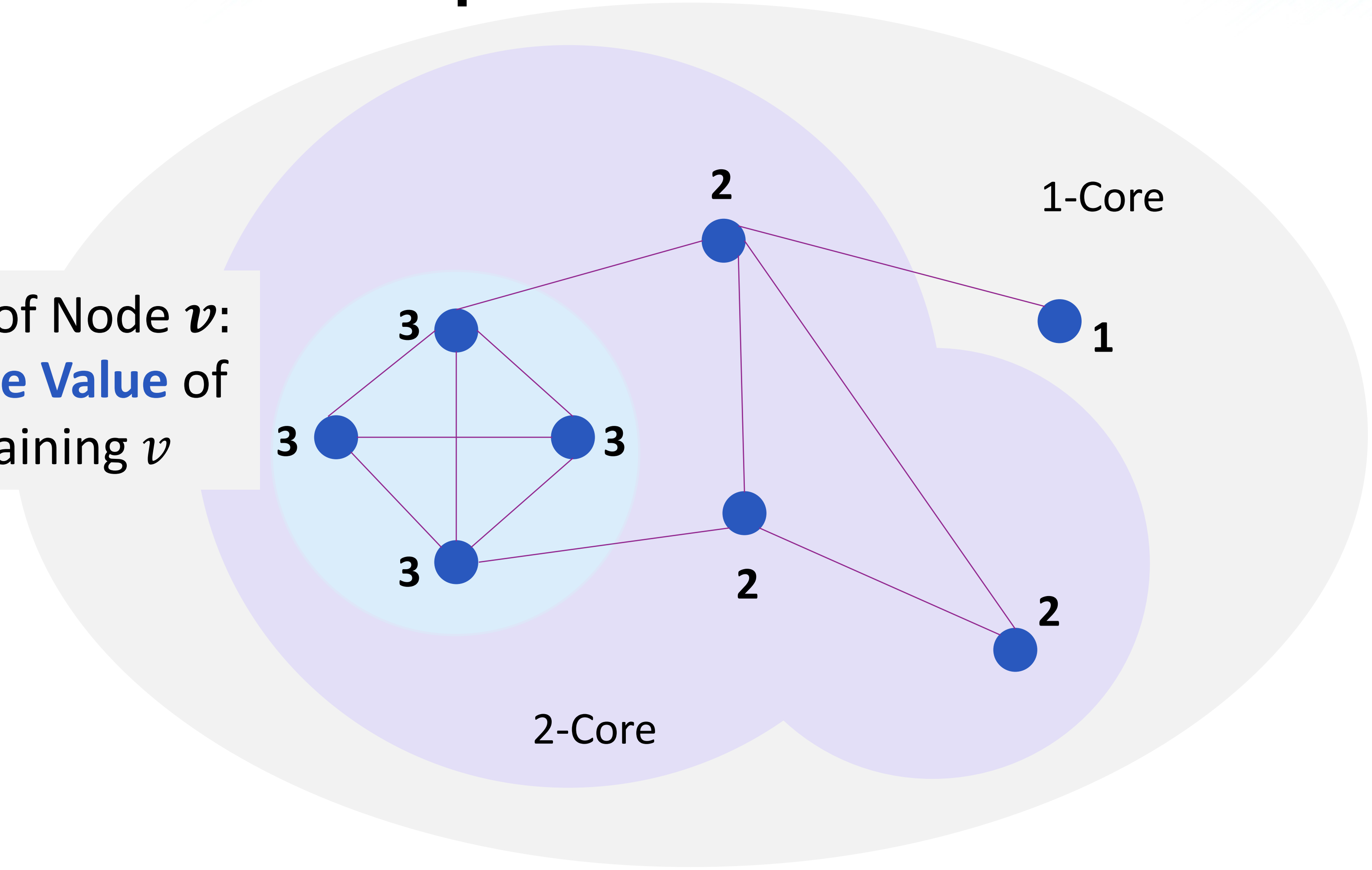
k-Core Decomposition

Core Number of Node v:
Maximum Core Value of a Core Containing v

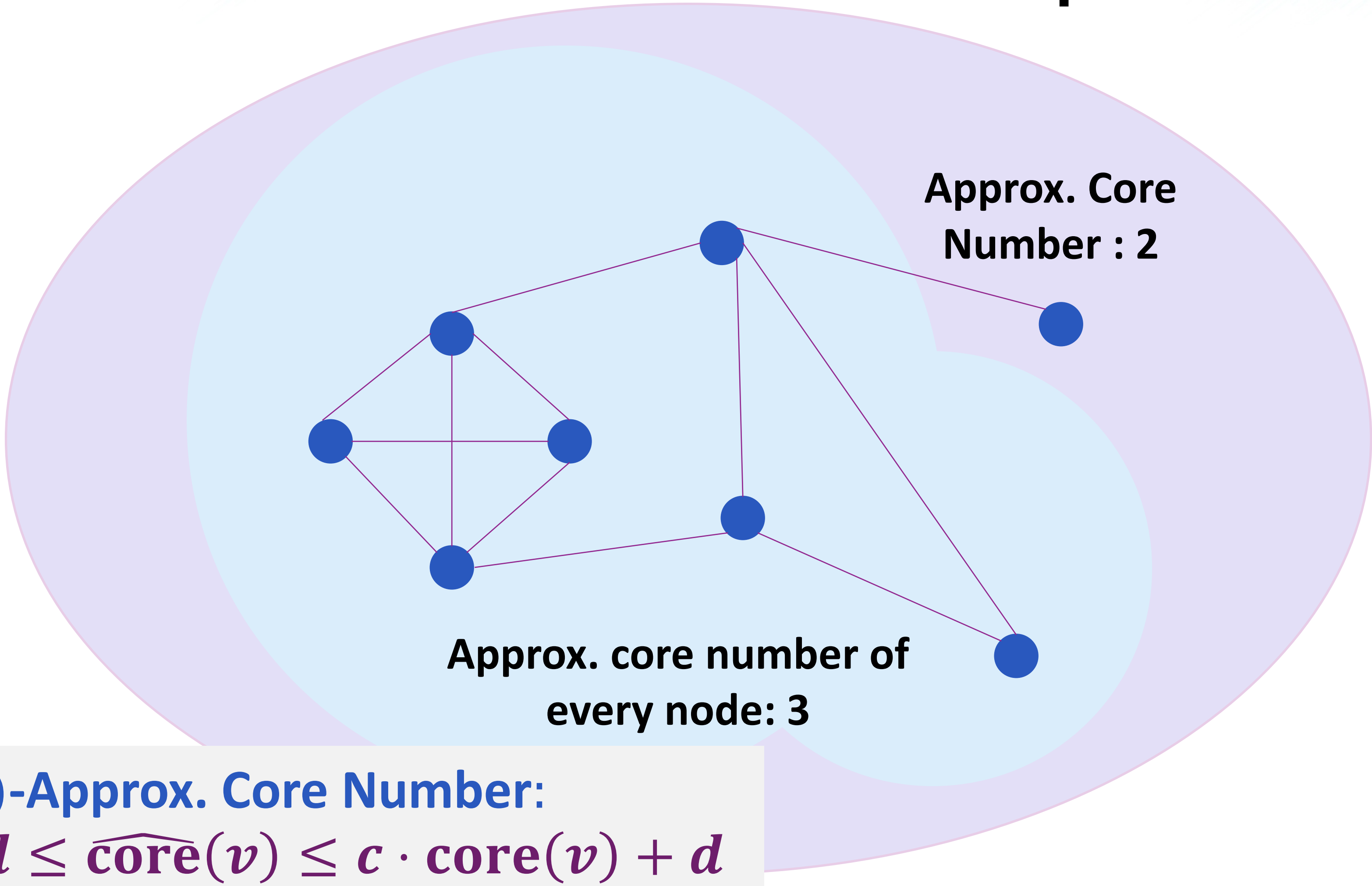


k-Core Decomposition

Core Number of Node v:
Maximum Core Value of
a Core Containing v



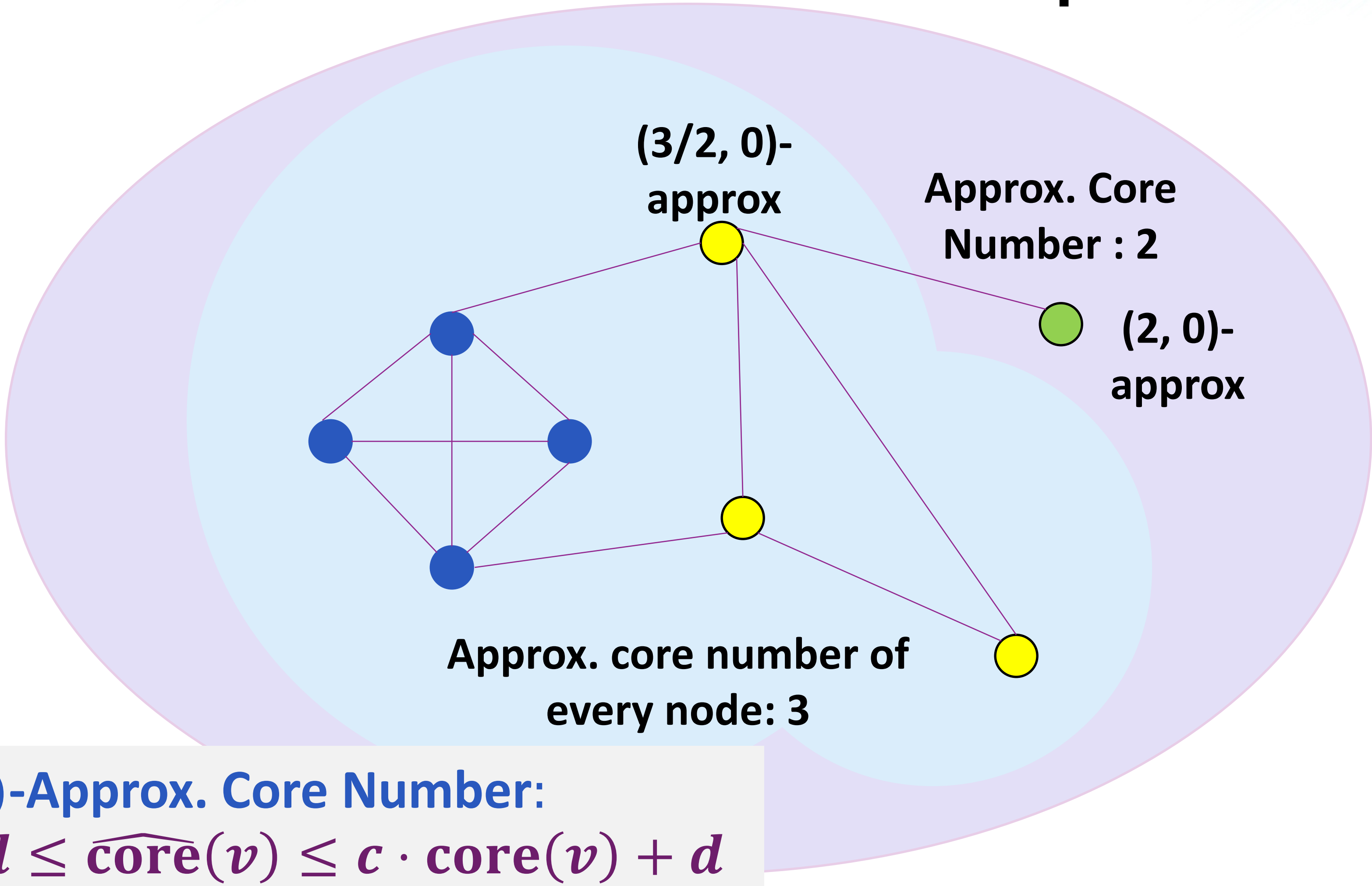
Approximate *k*-Core Decomposition



(c, d)-Approx. Core Number:

$$\mathrm{core}(v) - d \leq \widehat{\mathrm{core}}(v) \leq c \cdot \mathrm{core}(v) + d$$

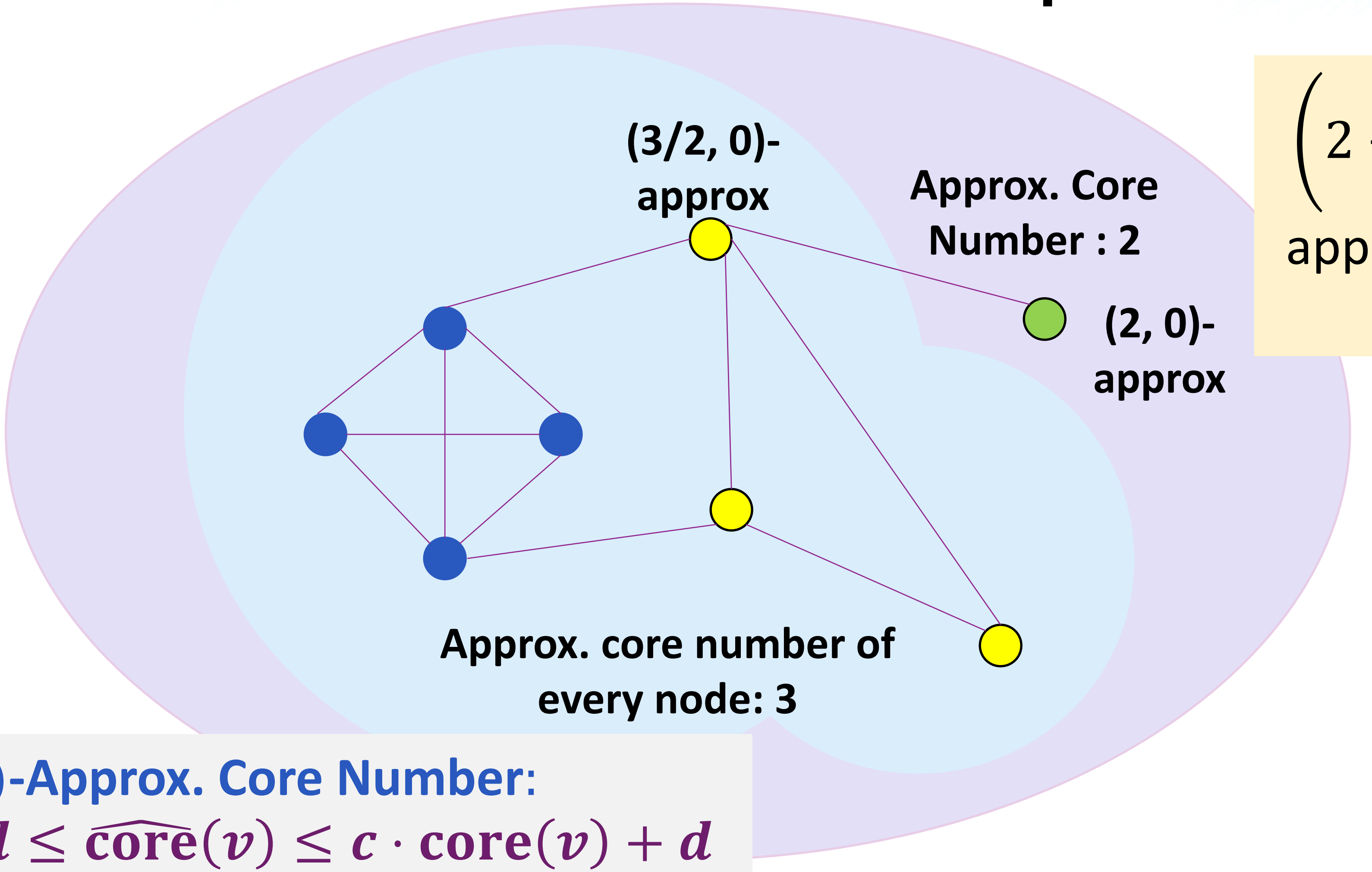
Approximate *k*-Core Decomposition



(c, d)-Approx. Core Number:

$$\mathbf{core}(v) - d \leq \widehat{\mathbf{core}}(v) \leq c \cdot \mathbf{core}(v) + d$$

Approximate *k*-Core Decomposition



$\left(2+\eta, O\left(\frac{\log^3(n)}{\varepsilon}\right)\right)$-approximations in this paper

(c, d)-Approx. Core Number:

$$\mathbf{core}(v) - d \leq \widehat{\mathbf{core}}(v) \leq c \cdot \mathbf{core}(v) + d$$

Level Data Structure and Core Numbers

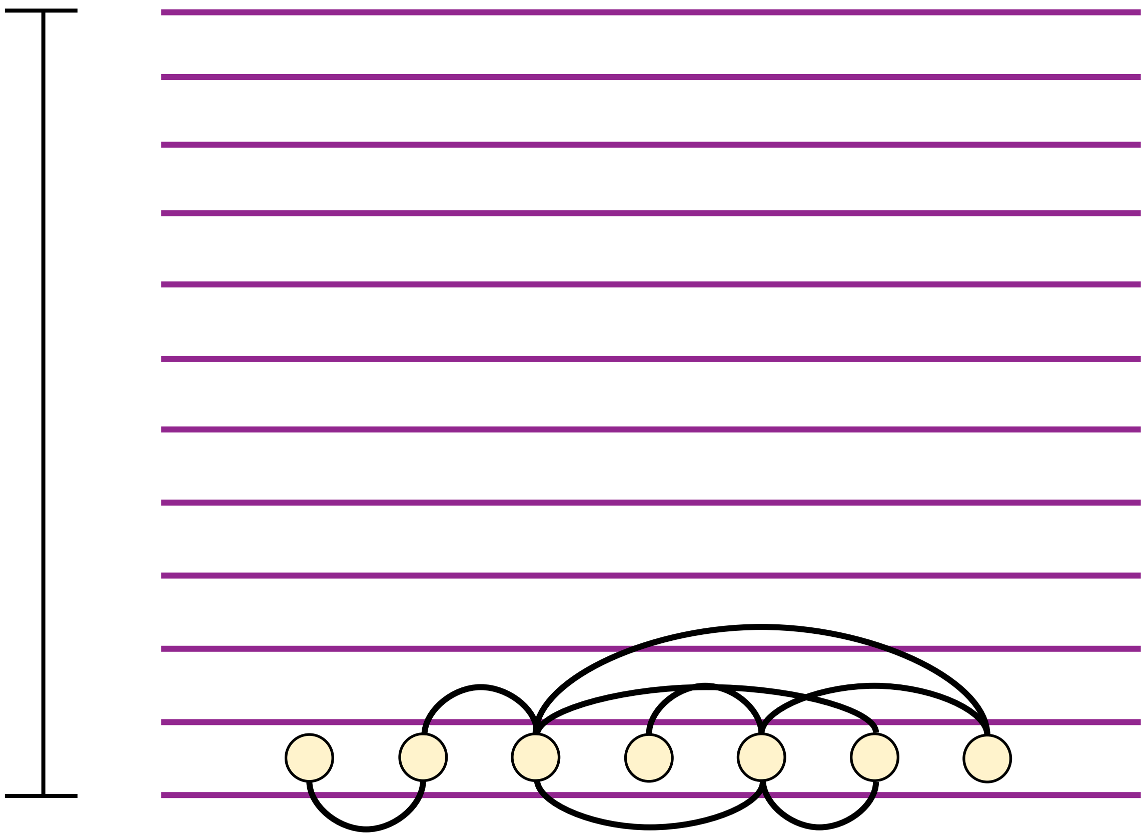
Non-private sequential and parallel level data structures for **dynamic problem**:

[Bhattacharya-Henzinger-Nanongkai-Tsourakakis ‘15;
Henzinger-Neumann-Wiese ‘20;
Liu-Shi-Yu-Dhulipala-Shun ‘22]

Level Data Structure and Core Numbers

In this example:
$\boldsymbol{\eta = 0.1}$

$4\log_{1+\eta}(n)$



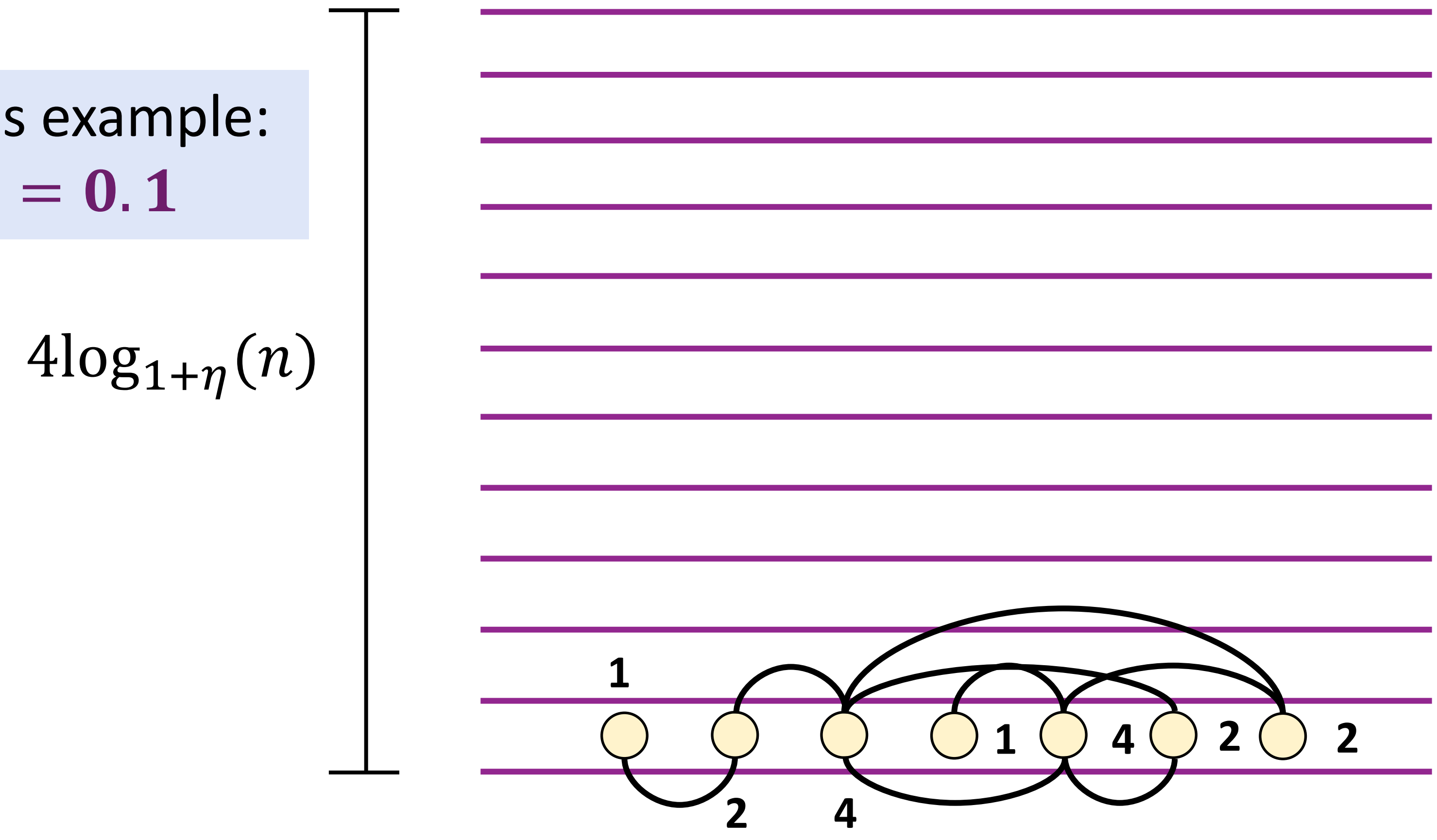
Move up if induced degree in **active** vertices $> (1+\eta)$

Initially all vertices are active

[Bhattacharya-Henzinger-Nanongkai-Tsourakakis '15, Henzinger-Neumann-Wiese '20, Liu-Shi-Yu-Dhulipala-Shun '22]

Level Data Structure and Core Numbers

In this example:
$\boldsymbol{\eta = 0.1}$



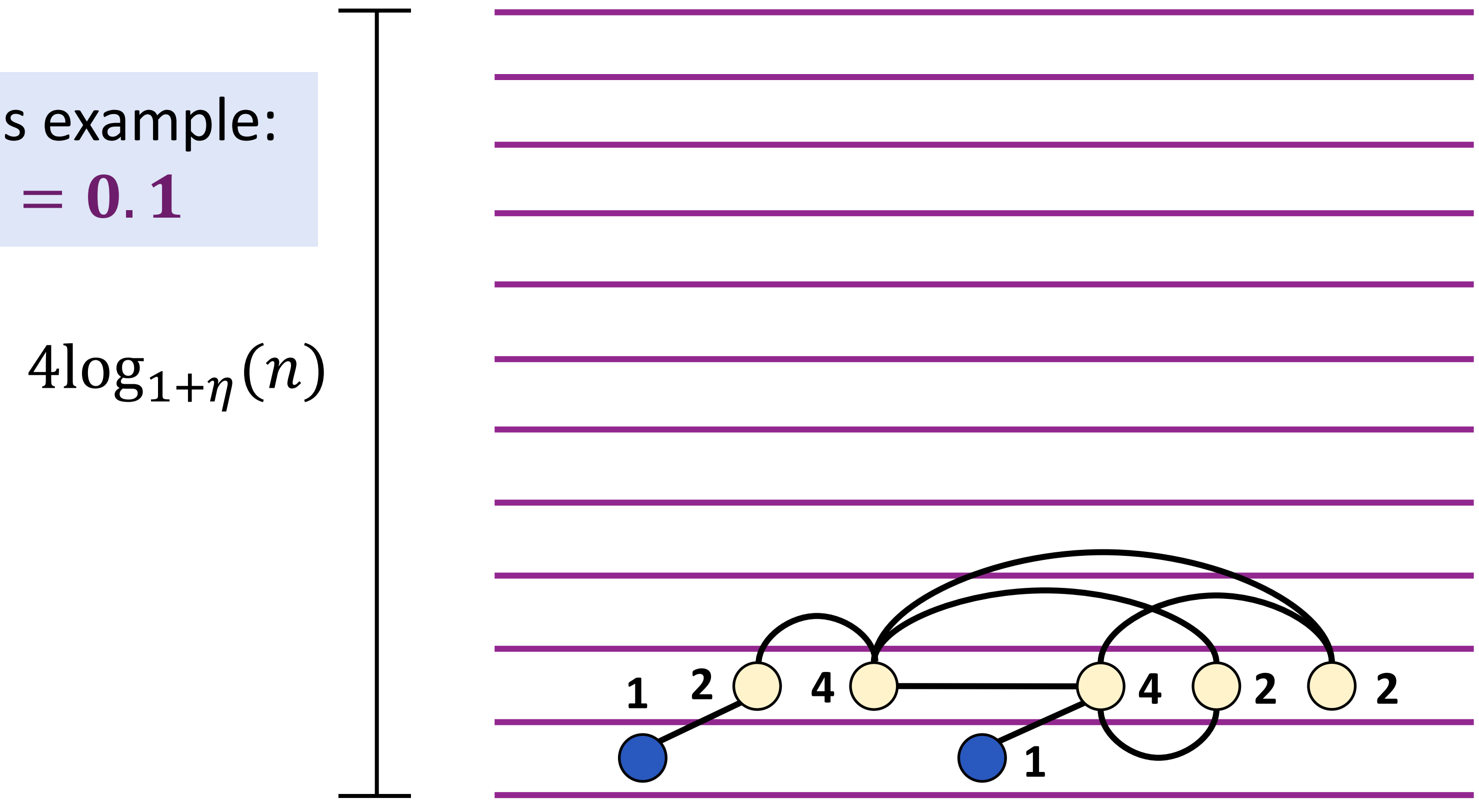
Move up if induced degree in **active** vertices $> (1+\eta)$

Initially all vertices are active

[Bhattacharya-Henzinger-Nanongkai-Tsourakakis ‘15, Henzinger-Neumann-Wiese ‘20, Liu-Shi-Yu-Dhulipala-Shun ‘22]

Level Data Structure and Core Numbers

In this example:
$\boldsymbol{\eta = 0.1}$



Move up if induced degree in **active** vertices $> (1+\eta)$

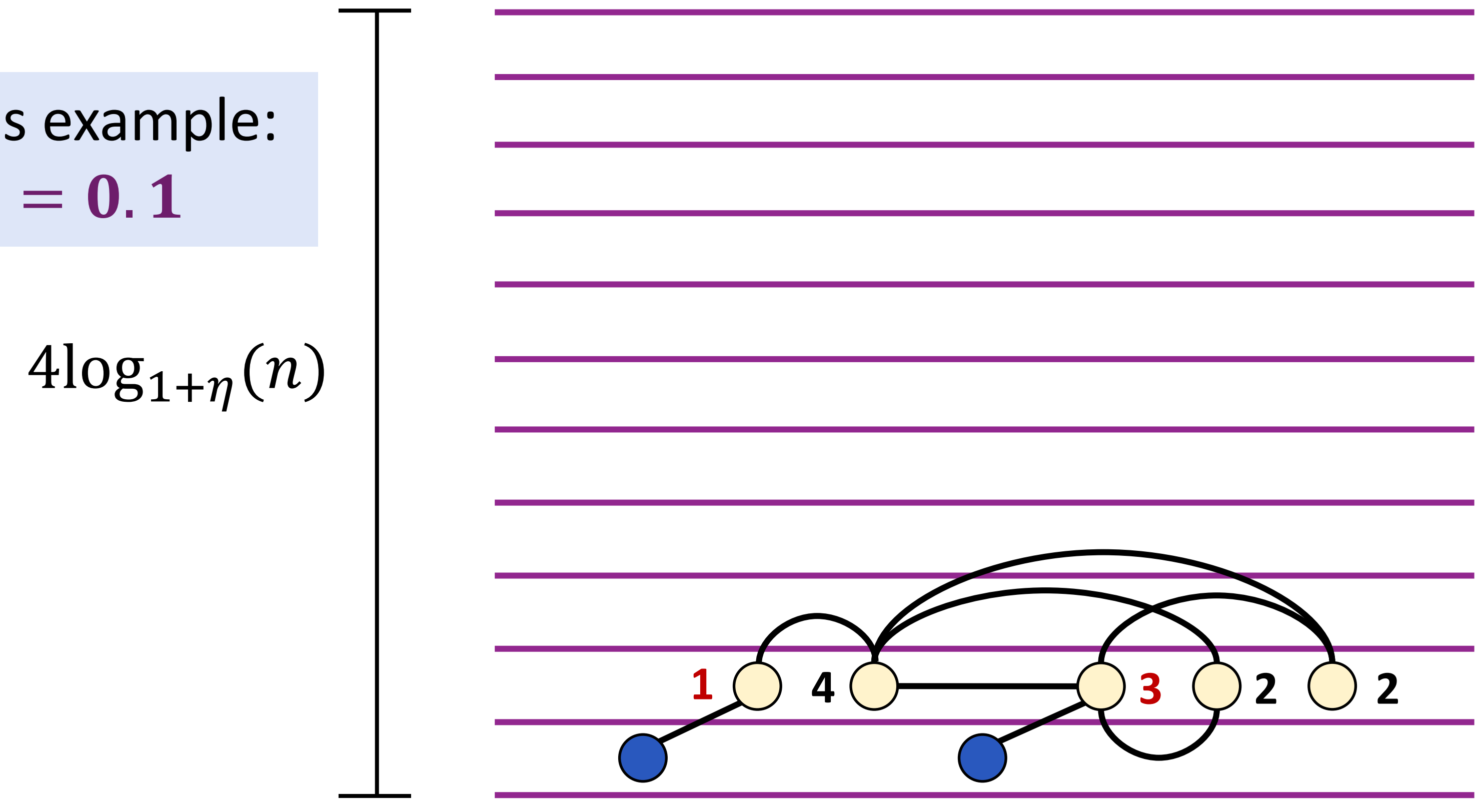
Initially all vertices are active

Vertices which **moved** remain **active**

[Bhattacharya-Henzinger-Nanongkai-Tsourakakis '15, Henzinger-Neumann-Wiese '20, Liu-Shi-Yu-Dhulipala-Shun '22]

Level Data Structure and Core Numbers

In this example:
$\boldsymbol{\eta = 0.1}$



Move up if induced degree in **active** vertices $> (1 + \eta)$

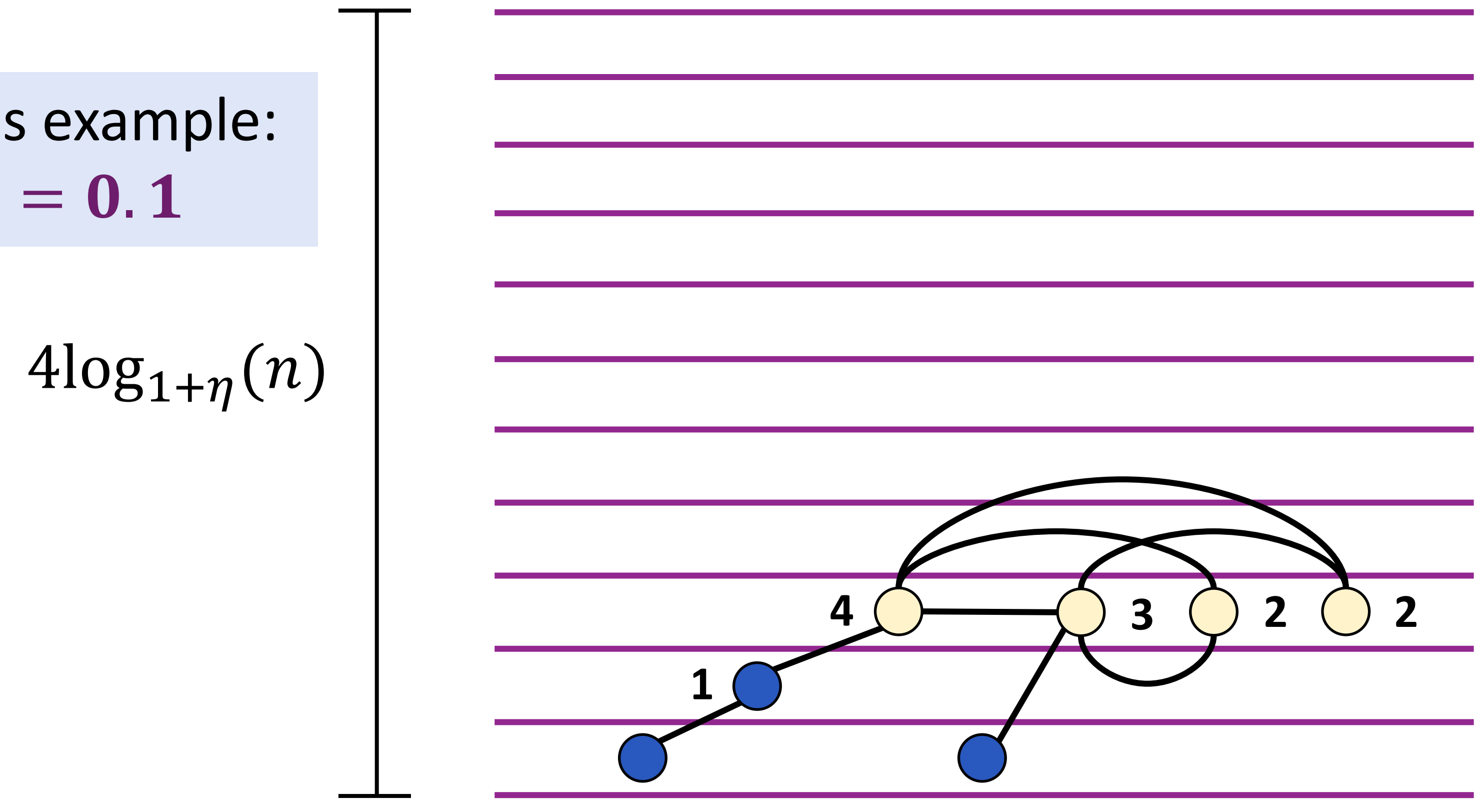
Initially all vertices are active

Vertices which **moved** remain **active**

[Bhattacharya-Henzinger-Nanongkai-Tsourakakis ‘15, Henzinger-Neumann-Wiese ‘20, Liu-Shi-Yu-Dhulipala-Shun ‘22]

Level Data Structure and Core Numbers

In this example:
$\boldsymbol{\eta = 0.1}$



Move up if induced degree in **active** vertices $> (1+\eta)$

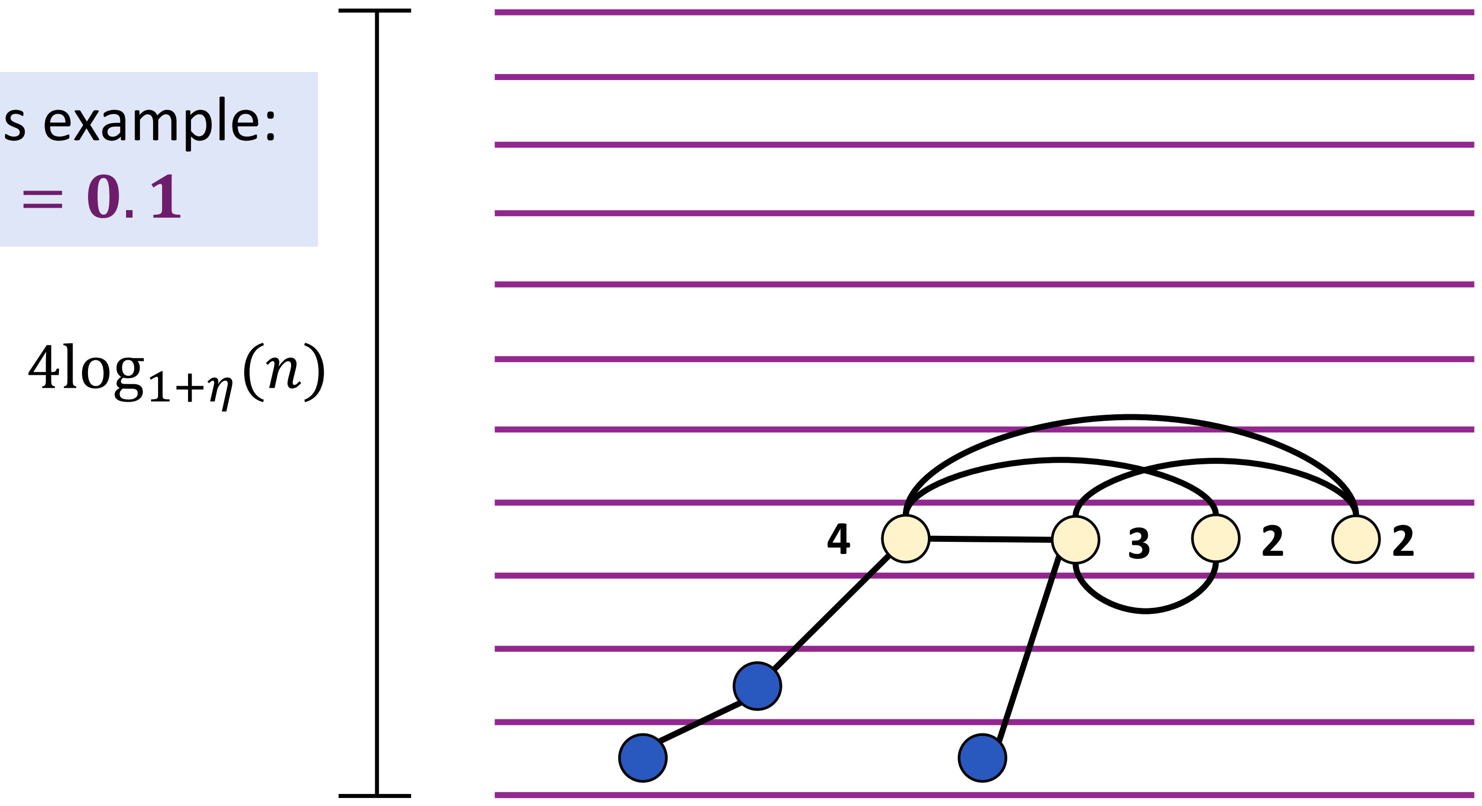
Initially all vertices are active

Vertices which **moved** remain **active**

[Bhattacharya-Henzinger-Nanongkai-Tsourakakis '15, Henzinger-Neumann-Wiese '20, Liu-Shi-Yu-Dhulipala-Shun '22]

Level Data Structure and Core Numbers

In this example:
$\boldsymbol{\eta = 0.1}$



Move up if induced degree in **active** vertices $> (1 + \eta)$

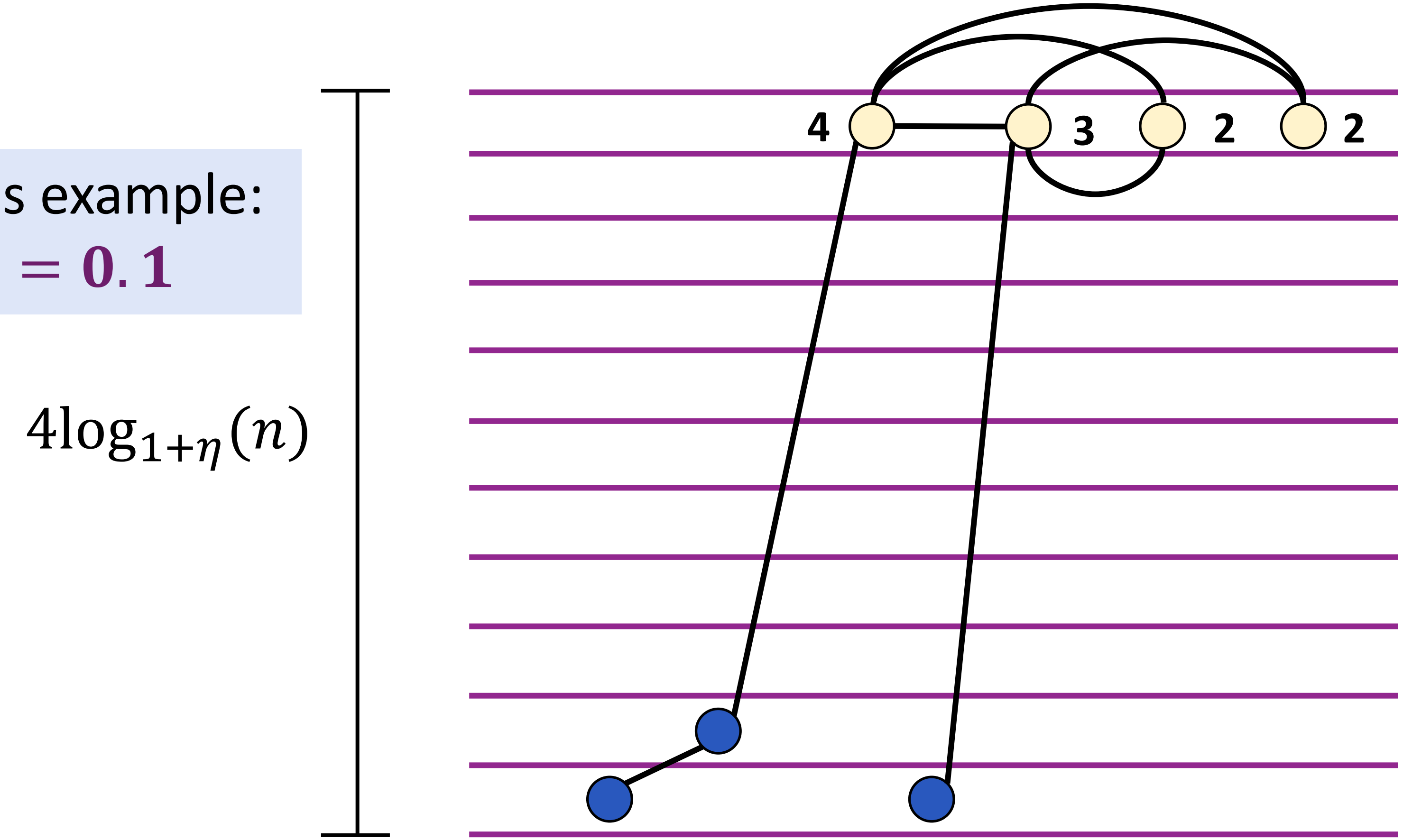
Initially all vertices are active

Vertices which **moved** remain **active**

[Bhattacharya-Henzinger-Nanongkai-Tsourakakis ‘15, Henzinger-Neumann-Wiese ‘20, Liu-Shi-Yu-Dhulipala-Shun ‘22]

Level Data Structure and Core Numbers

In this example:
$\boldsymbol{\eta = 0.1}$



Move up if induced degree in **active** vertices $> (1+\eta)$

Initially all vertices are active

Vertices which **moved** remain **active**

[Bhattacharya-Henzinger-Nanongkai-Tsourakakis '15, Henzinger-Neumann-Wiese '20, Liu-Shi-Yu-Dhulipala-Shun '22]

Non-Private Core Number Approximation

$\boldsymbol{\eta = 0.1}$

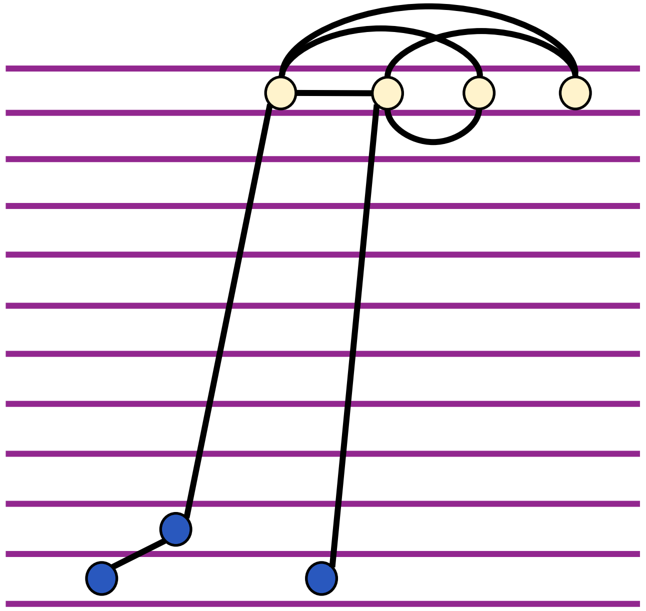
Set cutoffs $(\mathbf{1} + \boldsymbol{\eta})^{\boldsymbol{i}}$ for all $\boldsymbol{i} \in [\mathbf{log}_{\mathbf{1}+\boldsymbol{\eta}}(\boldsymbol{n})]$

[Bhattacharya-Henzinger-Nanongkai-Tsourakakis ‘15, Henzinger-Neumann-Wiese ‘20, Liu-Shi-Yu-Dhulipala-Shun ‘22]

Non-Private Core Number Approximation

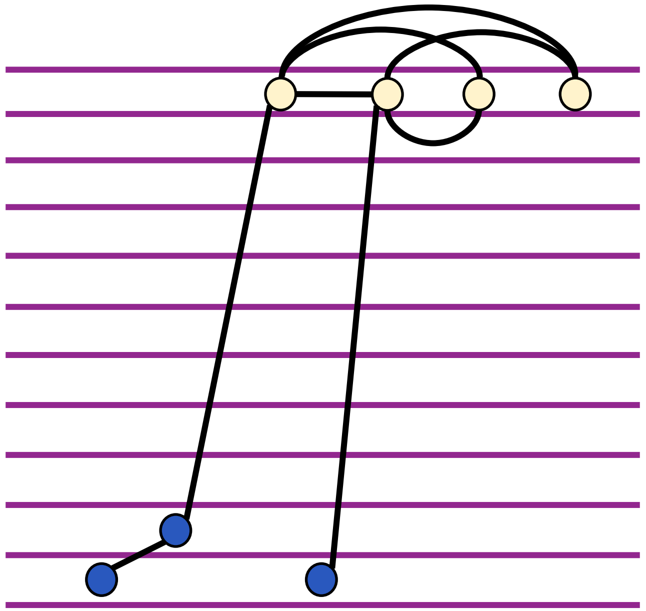
$\eta = 0.1$

Set cutoffs $(1+\eta)^i$ for all $i \in [\log_{1+\eta}(n)]$

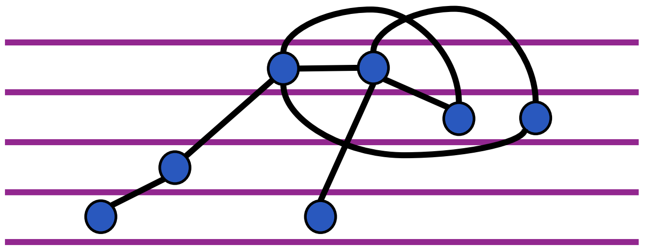


Cutoff: $(1+\eta)$

…



Cutoff: $(1+\eta)^7$



Cutoff: $(1+\eta)^8$

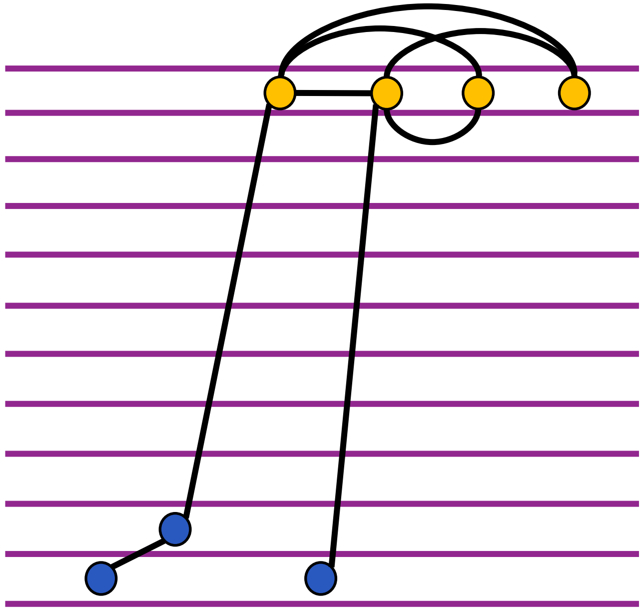
…

Give approx core number $2 \cdot (1+\eta)^i$
using **largest cutoff** where node is on the **topmost level**

Non-Private Core Number Approximation

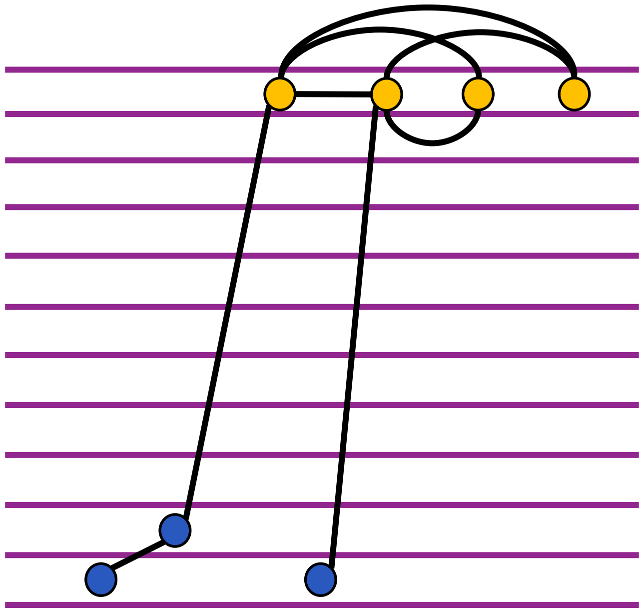
$\eta = 0.1$

Set cutoffs $(1+\eta)^i$ for all $i \in [\log_{1+\eta}(n)]$

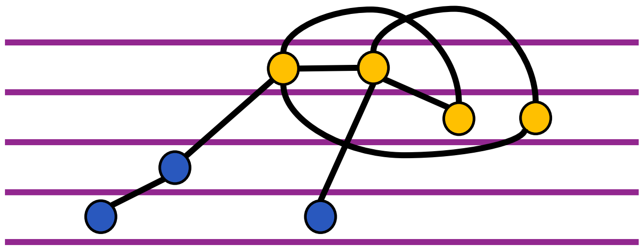


Cutoff: $(1+\eta)$

…



Cutoff: $(1+\eta)^7$



Cutoff: $(1+\eta)^8$

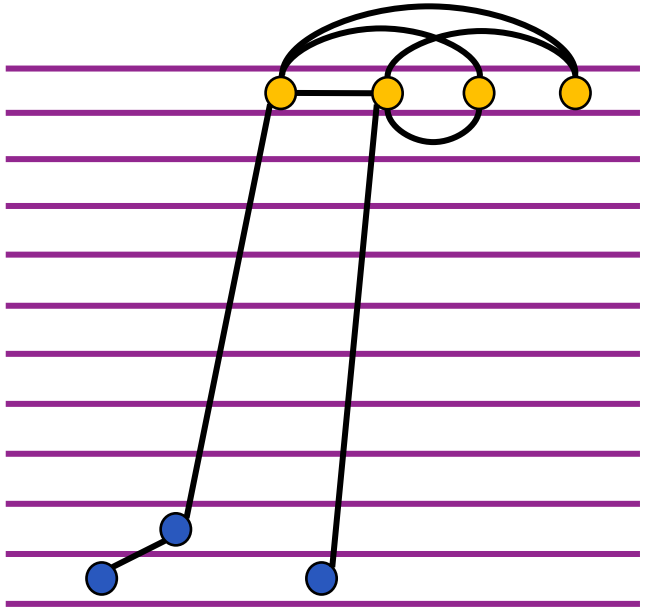
…

Approximation: $2 \cdot (1+\eta)^7 = 2 \cdot 1.1^7 = 2 \cdot 1.95 = 3.9$

Non-Private Core Number Approximation

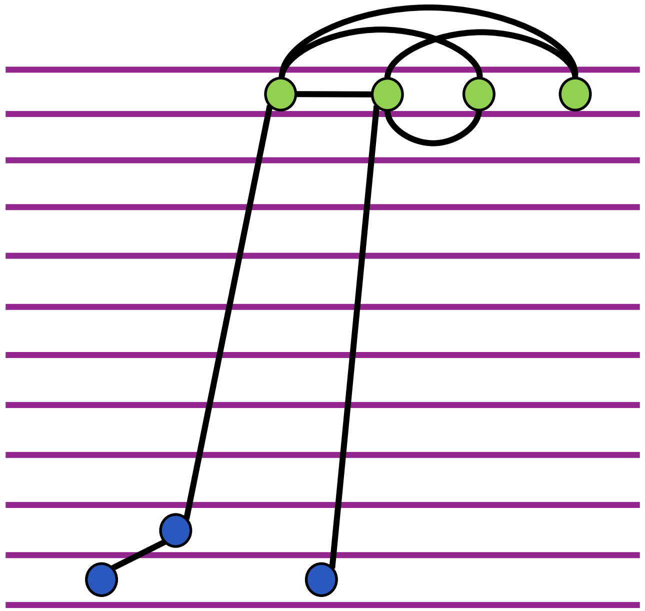
$\eta = 0.1$

Set cutoffs $(1+\eta)^i$ for all $i \in [\log_{1+\eta}(n)]$

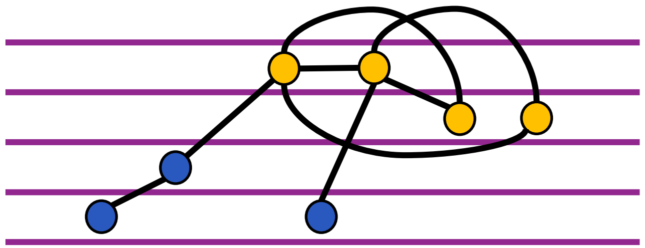


Cutoff: $(1+\eta)$

...



Cutoff: $(1+\eta)^7$



Cutoff: $(1+\eta)^8$

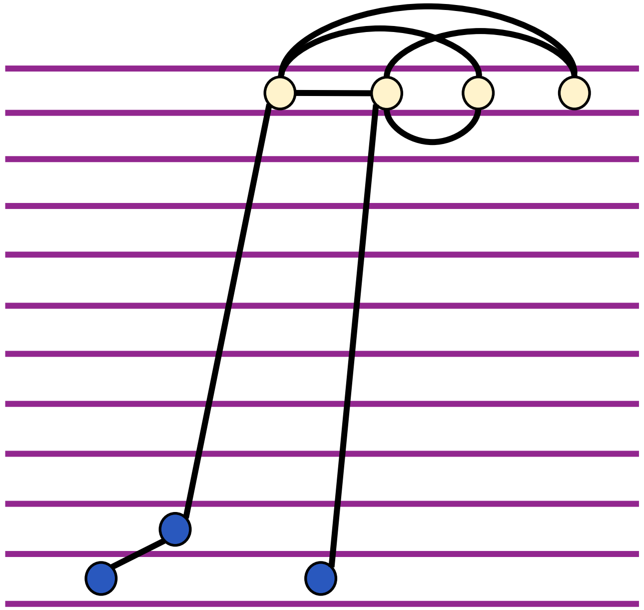
...

Approximation: $2 \cdot (1+\eta)^7 = 2 \cdot 1.1^7 = 2 \cdot 1.95 = 3.9$

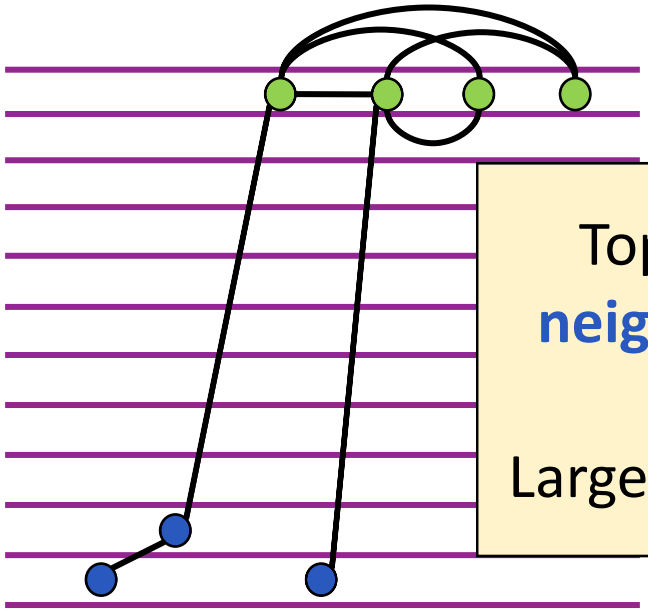
Non-Private Core Number Approximation

$\eta = 0.1$

Set cutoffs $(1+\eta)^i$ for all $i \in [\log_{1+\eta}(n)]$



...



Top level means adjacent to **many neighbors of sufficiently high degree**

Largest cutoff gives **largest such degree**

Cutoff: $(1+\eta)$

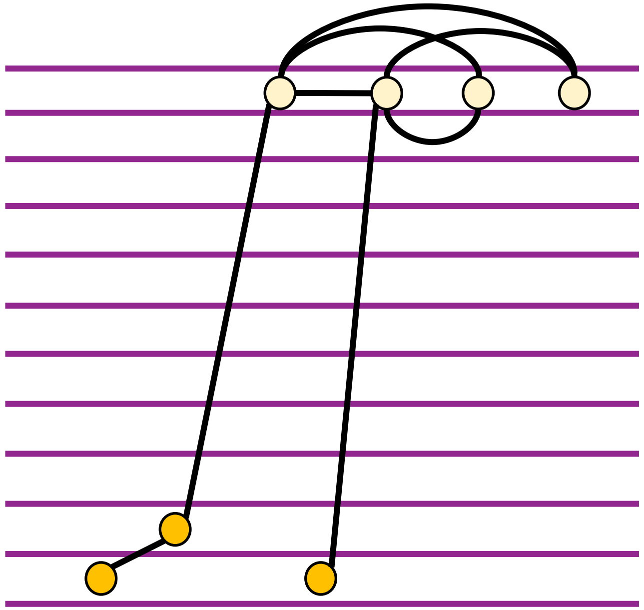
Cutoff: $(1+\eta)^7$

Approximation: $2 \cdot (1+\eta)^7 = 2 \cdot 1.1^7 = 2 \cdot 1.95 = 3.9$

Non-Private Core Number Approximation

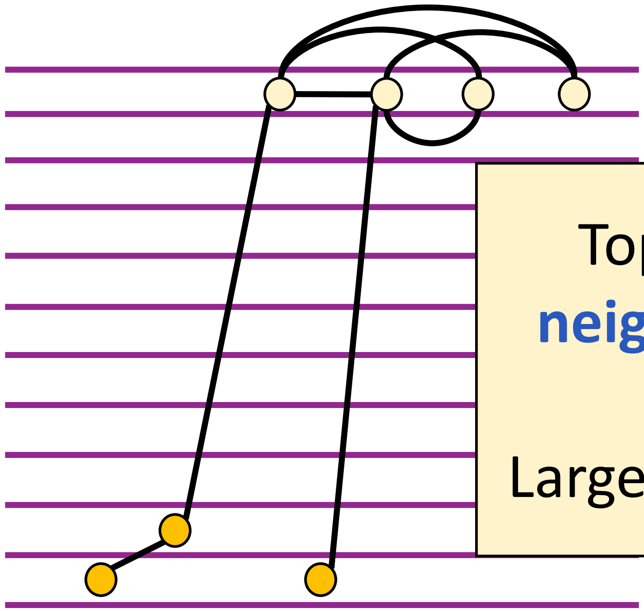
$\boldsymbol{\eta = 0.1}$

Set cutoffs $(\mathbf{1}+\boldsymbol{\eta})^{\boldsymbol{i}}$ for all $\boldsymbol{i} \in [\mathbf{log}_{\mathbf{1}+\boldsymbol{\eta}}(\boldsymbol{n})]$



Cutoff: $(\mathbf{1}+\boldsymbol{\eta})$

…



Cutoff: $(\mathbf{1}+\boldsymbol{\eta})^{\mathbf{7}}$

Top level means adjacent to **many neighbors of sufficiently high degree**

Largest cutoff gives **largest such degree**

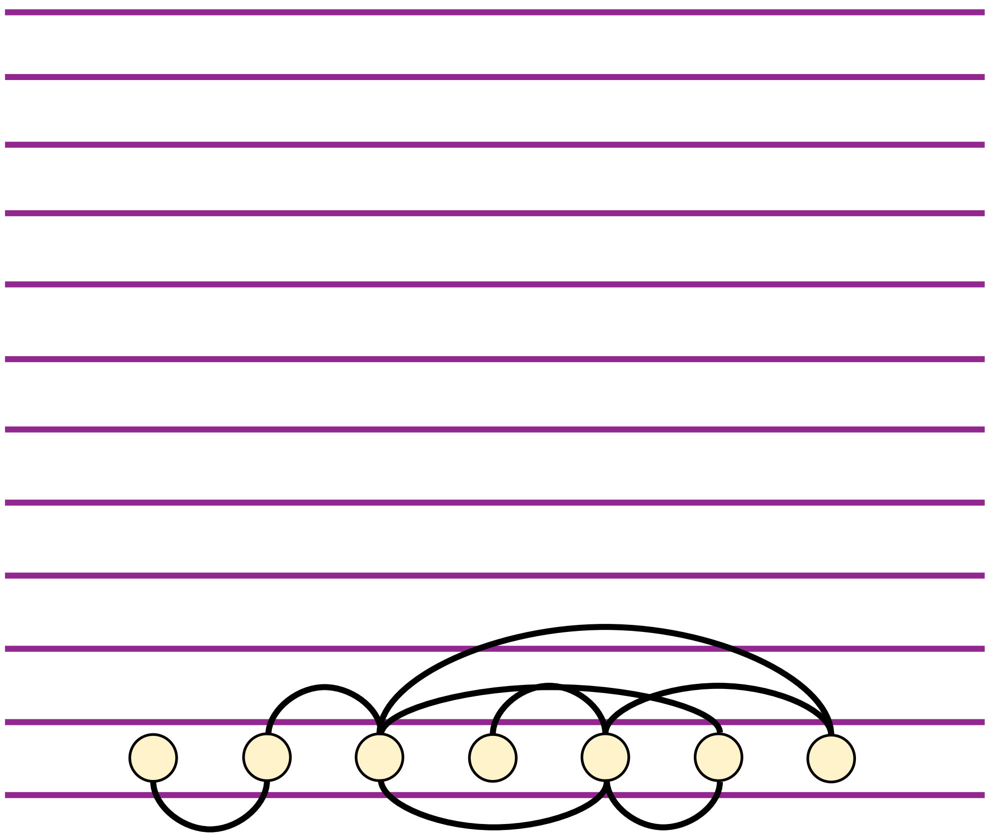
Approximation: $\mathbf{1}$

ε-LEDP Core Numbers

Each **active vertex draws i.i.d. noise** from **symmetric geometric distribution**

Distribution $\boldsymbol{Geom(b)}$

PMF: $\frac{e^b - 1}{e^b + 1} \cdot e^{-|X| \cdot b}$



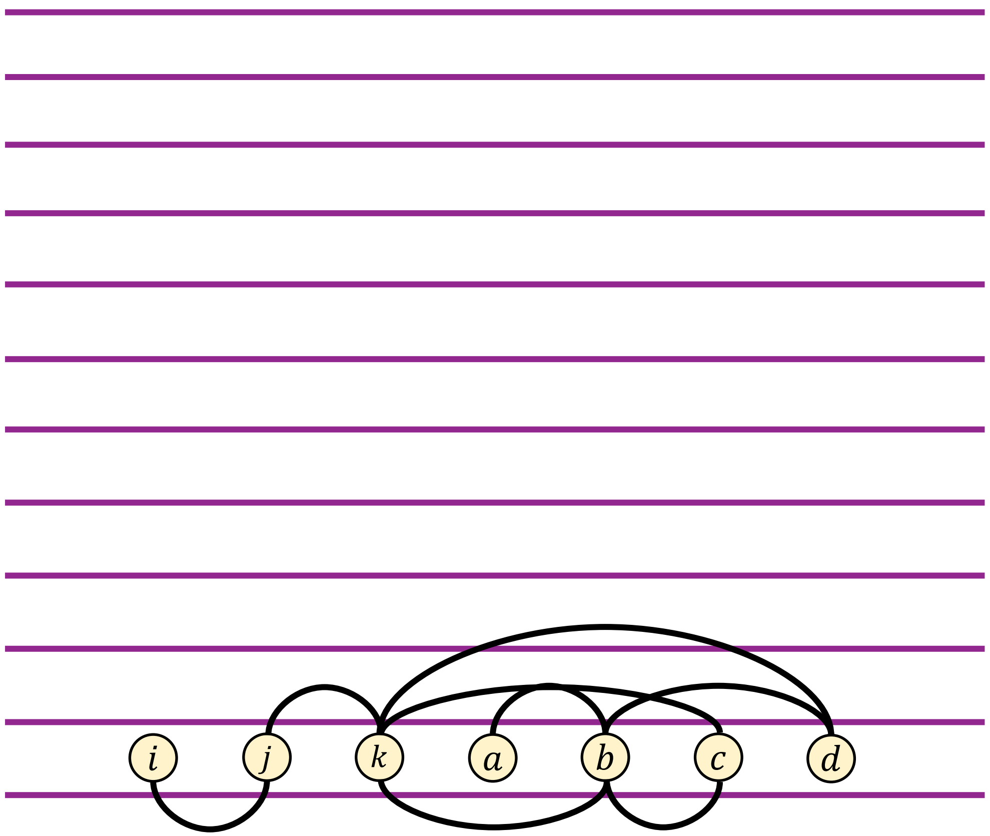
Release and move up degree **+ noise** in **active** vertices $> (1 + \eta)$

ε-LEDP Core Numbers

Each **active vertex draws i.i.d. noise** from **symmetric geometric distribution**

Distribution $\boldsymbol{Geom(b)}$

PMF: $\frac{e^b - 1}{e^b + 1} \cdot e^{-|X| \cdot b}$



Release and move up degree **+ noise** in **active** vertices $> (1 + \eta)$

ε-LEDP Core Numbers

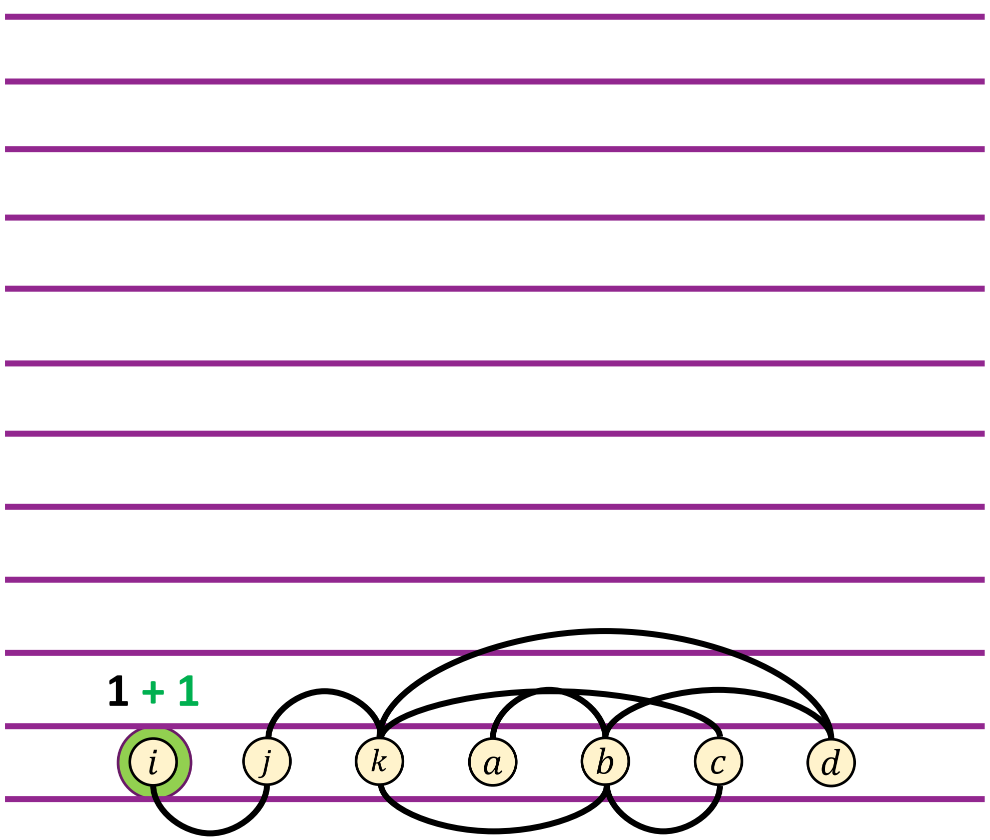
Each **active vertex draws i.i.d. noise** from **symmetric geometric distribution**

Distribution $\boldsymbol{Geom(b)}$

PMF: $\frac{e^b - 1}{e^b + 1} \cdot e^{-|X| \cdot b}$

If $\mathbf{deg}(\boldsymbol{i}) + \boldsymbol{N_i} > (\mathbf{1} + \boldsymbol{\eta})$, **move up**

Where $N_i \sim Geom\left(\frac{\varepsilon}{8\log^2_{1+\eta}(n)}\right)$



Release and move up degree **+ noise** in **active** vertices $> (1 + \eta)$

In this example: $\boldsymbol{\eta = 0.1}$

ε-LEDP Core Numbers

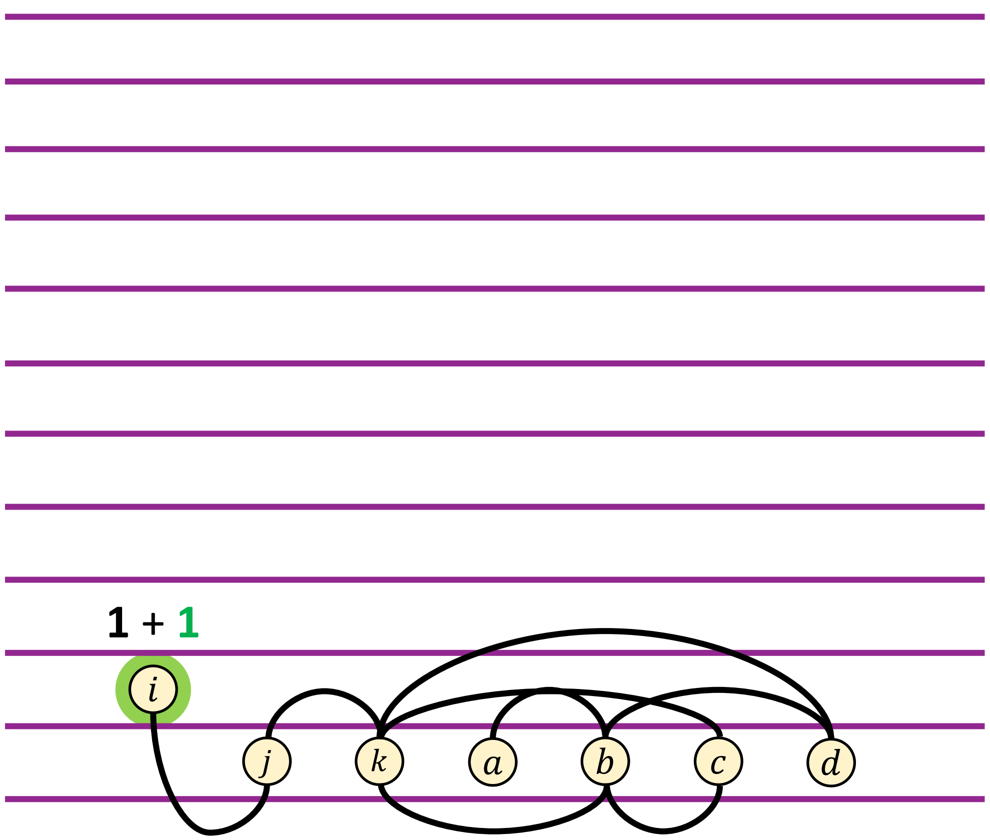
Each **active vertex draws i.i.d. noise** from **symmetric geometric distribution**

Distribution $\boldsymbol{Geom(b)}$

PMF: $\frac{e^b - 1}{e^b + 1} \cdot e^{-|X| \cdot b}$

If $\mathbf{deg}(\boldsymbol{i}) + \boldsymbol{N_i} > (\mathbf{1} + \boldsymbol{\eta})$, **move up**

Where $N_i \sim Geom\left(\frac{\varepsilon}{8\log^2_{1+\eta}(n)}\right)$



Release and move up degree **+ noise** in **active** vertices $> (1 + \eta)$

ε-LEDP Core Numbers

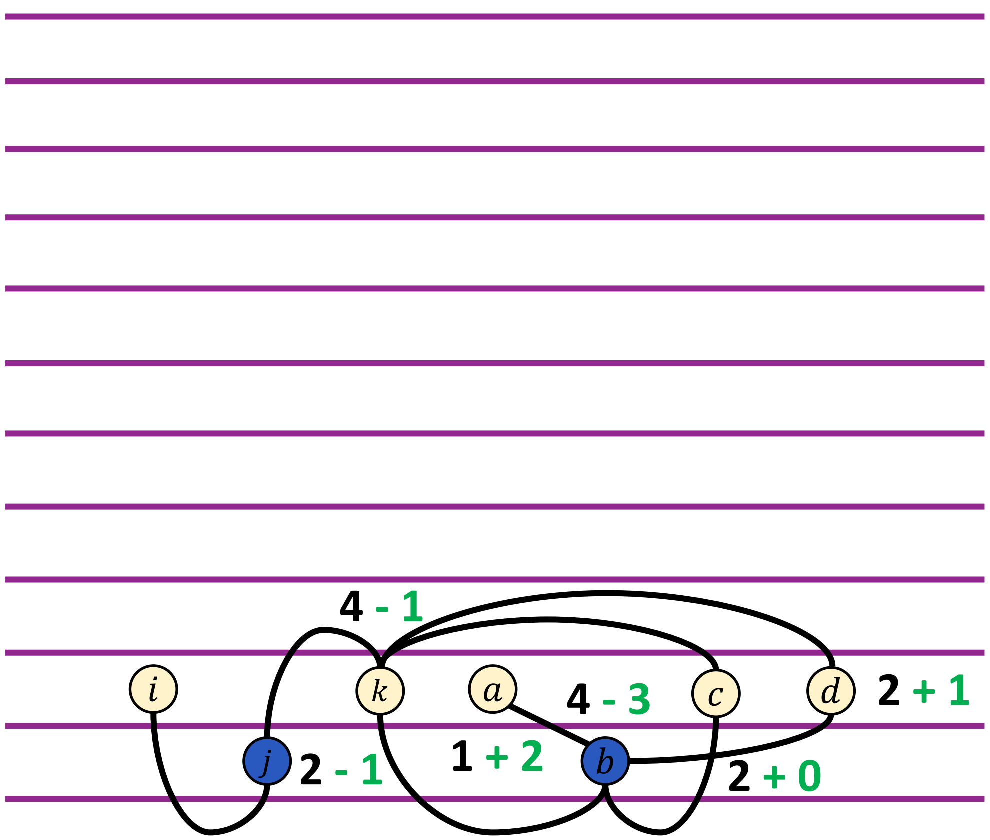
Each **active vertex draws i.i.d. noise** from **symmetric geometric distribution**

Distribution $\boldsymbol{Geom(b)}$

PMF: $\frac{e^b - 1}{e^b + 1} \cdot e^{-|X| \cdot b}$

If $\mathbf{deg(i)} + \boldsymbol{N_i} > (\mathbf{1} + \boldsymbol{\eta})$, **move up**

Where $N_i \sim Geom\left(\frac{\varepsilon}{8\log^2_{1+\eta}(n)}\right)$



Release and move up degree **+ noise** in **active** vertices $> (1 + \eta)$

Redraw new noise each time vertex remains active

ε-LEDP Core Numbers

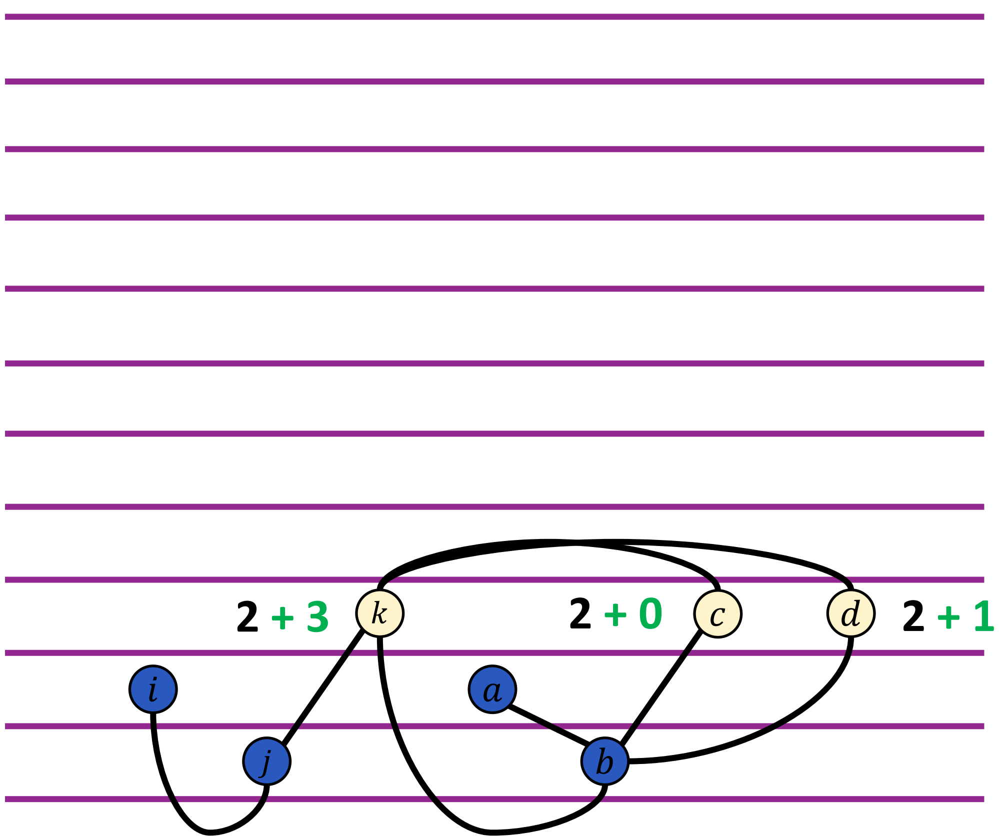
Each **active vertex draws i.i.d. noise** from **symmetric geometric distribution**

Distribution $\boldsymbol{Geom(b)}$

PMF: $\frac{e^b - 1}{e^b + 1} \cdot e^{-|X| \cdot b}$

If $\mathbf{deg}(\boldsymbol{i}) + \boldsymbol{N_i} > (\mathbf{1} + \boldsymbol{\eta})$, **move up**

Where $N_i \sim Geom\left(\frac{\varepsilon}{8\log^2_{1+\eta}(n)}\right)$



Release and move up degree **+ noise** in **active** vertices $> (1 + \eta)$

Redraw new noise each time vertex remains active

Approx. as before $\mathbf{2(1 + \eta)^{i}}$ using topmost level

ε-LEDP Core Numbers

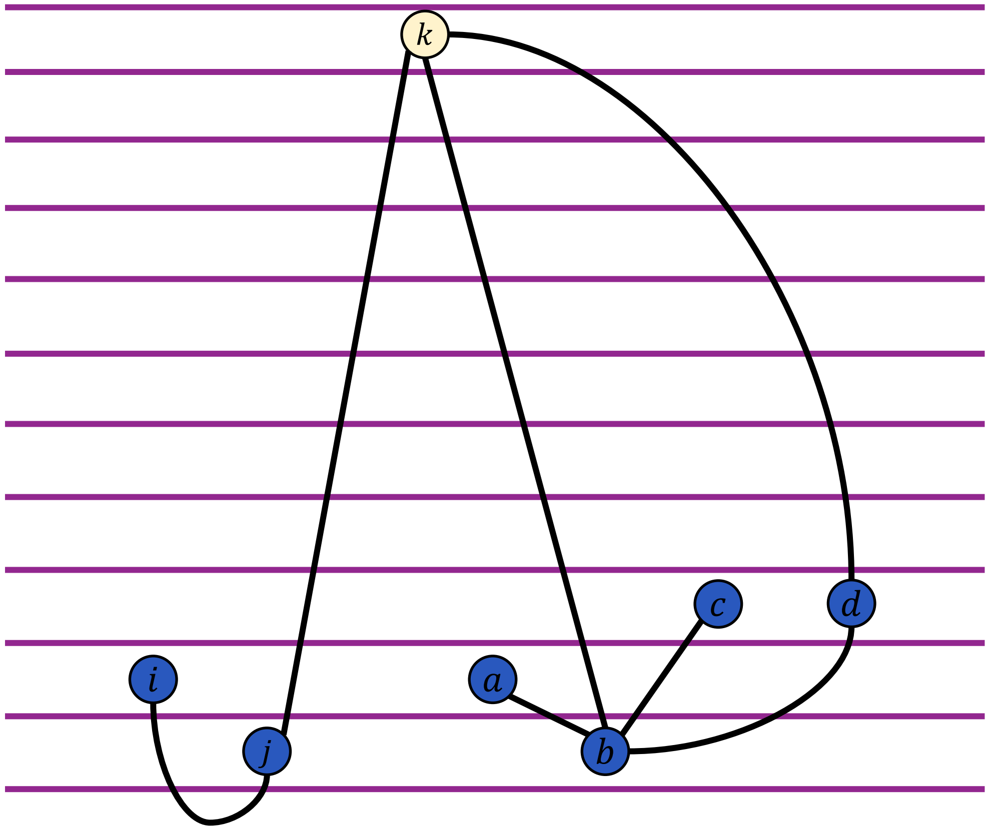
Each **active vertex draws i.i.d. noise** from **symmetric geometric distribution**

Distribution $\boldsymbol{Geom(b)}$

PMF: $\frac{e^b - 1}{e^b + 1} \cdot e^{-|X| \cdot b}$

If $\mathbf{deg}(\boldsymbol{k}) + \boldsymbol{N_k} > (\mathbf{1} + \boldsymbol{\eta})$, **move up**

Where $N_k \sim Geom\left(\frac{\varepsilon}{8\log^2_{1+\eta}(n)}\right)$



Release and move up degree **+ noise** in **active** vertices $> (1 + \eta)$

Redraw new noise each time vertex remains active

Approx. as before $\mathbf{2(1+\eta)^i}$ using topmost level

ε-LEDP Core Numbers

Each **active vertex draws i.i.d. noise** from **symmetric geometric distribution**

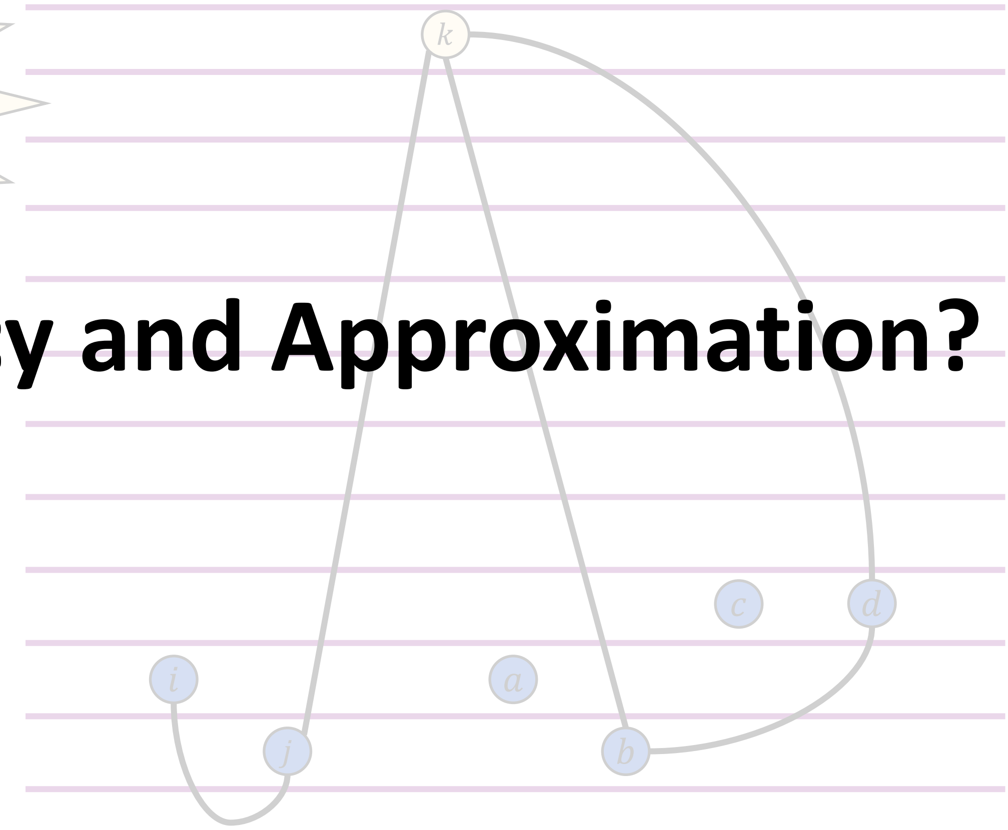
Distribution $Geom(b)$

PMF: $\frac{e^b - 1}{e^b + 1} \cdot e^{-|X| \cdot b}$

If $\mathbf{deg(k) + N_k > (1+\eta)}$, **move up**

Where $N_k \sim Geom\left(\frac{\varepsilon}{8\log_{1+\eta}(n)}\right)$

Privacy and Approximation?



Move up if induced degree **+ noise** in **active** vertices $> (1+\eta)$

Redraw new noise each time vertex remains active and determines whether move up

Approx. as before $\mathbf{2(1+\eta)^i}$ where i largest that vertex is on the topmost level

Privacy Proof

- Can be implemented via **local randomizers R**

Privacy Proof

- Can be implemented via **local randomizers $\boldsymbol{R}$**
- R takes as **input $\boldsymbol{a}$ (adjacency list)** and **public set of active vertices $\boldsymbol{A}$**

Privacy Proof

- Can be implemented via **local randomizers $\boldsymbol{R}$**
- R takes as **input $\boldsymbol{a}$ (adjacency list)** and **public set of active vertices $\boldsymbol{A}$**
  - R computes size of intersection $|\boldsymbol{a} \cap \boldsymbol{A}|$

Privacy Proof

- Can be implemented via **local randomizers $\boldsymbol{R}$**
- R takes as **input $\boldsymbol{a}$ (adjacency list)** and **public set of active vertices $\boldsymbol{A}$**
  - R computes size of intersection $|\boldsymbol{a} \cap \boldsymbol{A}|$

  - Then, add symmetric geometric noise $X \sim Geom\left(\frac{\varepsilon}{8\log_{1+\eta}^2(n)}\right)$

Privacy Proof

- Can be implemented via **local randomizers $\boldsymbol{R}$**
- R takes as **input $\boldsymbol{a}$ (adjacency list)** and **public set of active vertices $\boldsymbol{A}$**
  - R computes size of intersection $|\boldsymbol{a} \cap \boldsymbol{A}|$ **Sensitivity of 1**
  - Then, add symmetric geometric noise $X \sim Geom\left(\frac{\varepsilon}{8\log^2_{1+\eta}(n)}\right)$

Global Sensitivity:

$$\Delta_f = \max_{edge-neighbors\, G\, and\, G'} |f(G) - f(G')|$$

$$f(\boldsymbol{a}, A) = |\boldsymbol{a} \cap A|$$

Privacy Proof

- Can be implemented via **local randomizers $\boldsymbol{R}$**
- R takes as **input $\boldsymbol{a}$ (adjacency list)** and **public set of active vertices $\boldsymbol{A}$**
  - R computes size of intersection $|\boldsymbol{a} \cap \boldsymbol{A}|$ **Sensitivity of 1**
  - Then, add symmetric geometric noise $X \sim Geom\left(\frac{\varepsilon}{8\log^2_{1+\eta}(n)}\right)$

Geometric Mechanism:

[Chan-Shi-Song '11; Balcer-Vadhan '18]

$$M(\boldsymbol{a}, A) = f(\boldsymbol{a}, A) + Geom\left(\frac{\varepsilon}{\Delta_f}\right)$$

$\boldsymbol{M}$ **is $\boldsymbol{\varepsilon}$-DP**

Privacy Proof

- Can be implemented via **local randomizers $\boldsymbol{R}$**
- R takes as **input $\boldsymbol{a}$ (adjacency list)** and **public set of active vertices $\boldsymbol{A}$**
  - R computes size of intersection $|\boldsymbol{a} \cap \boldsymbol{A}|$ **Sensitivity of 1**
  - Then, add symmetric geometric noise $X \sim Geom\left(\frac{\varepsilon}{8\log_{1+\eta}^2(n)}\right)$
- **$\boldsymbol{R}$ is $\boldsymbol{\frac{\varepsilon}{8\log_{1+\eta}^2(n)}}$ - LR** by privacy of **Geometric Mechanism** [Chan-Shi-Song '11; Balcer-Vadhan '18]

Privacy Proof

- Can be implemented via **local randomizers R**
- R takes as **input a (adjacency list)** and **public set of active vertices A**
  - R computes size of intersection $|a \cap A|$ **Sensitivity of 1**
  - Then, add symmetric geometric noise $X \sim Geom\left(\frac{\varepsilon}{8\log_{1+\eta}^2(n)}\right)$
- **R is $\frac{\varepsilon}{8\log_{1+\eta}^2(n)}$ - LR** by privacy of **Geometric Mechanism** [Chan-Shi-Song '11; Balcer-Vadhan '18]
- Same LR called for all vertices $4\log_{1+\eta}^2(n)$ times

Privacy Proof

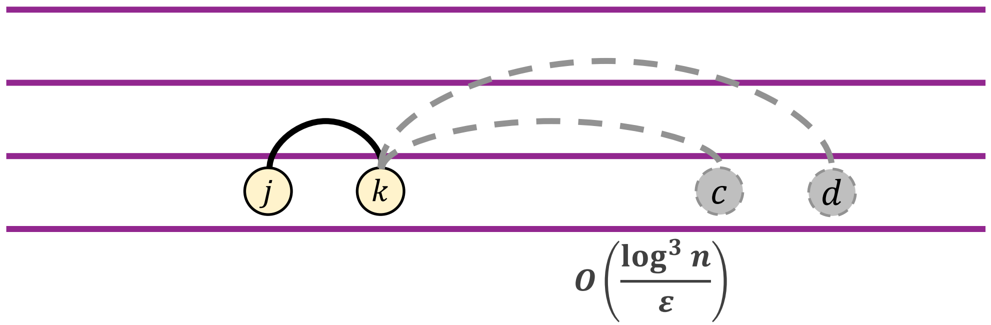
- Can be implemented via **local randomizers R**
- R takes as **input a (adjacency list)** and **public set of active vertices A**
  - R computes size of intersection $|a \cap A|$ **Sensitivity of 1**
  - Then, add symmetric geometric noise $X \sim Geom\left(\frac{\varepsilon}{8\log_{1+\eta}^2(n)}\right)$
- **R is $\frac{\varepsilon}{8\log_{1+\eta}^2(n)}$ - LR** by privacy of **Geometric Mechanism** [Chan-Shi-Song '11; Balcer-Vadhan '18]
- Same LR called for all vertices $4\log_{1+\eta}^2(n)$ times
- For each edge, called $8\log_{1+\eta}^2(n)$; then, $8\log_{1+\eta}^2(n) \cdot \frac{\varepsilon}{8\log_{1+\eta}^2(n)} = \varepsilon$ and so ε-LEDP

Approximation Proof Sketch

- With high probability, magnitude of each drawn noise is **upper bounded by $O\left(\frac{\log^3 n}{\varepsilon}\right)$**

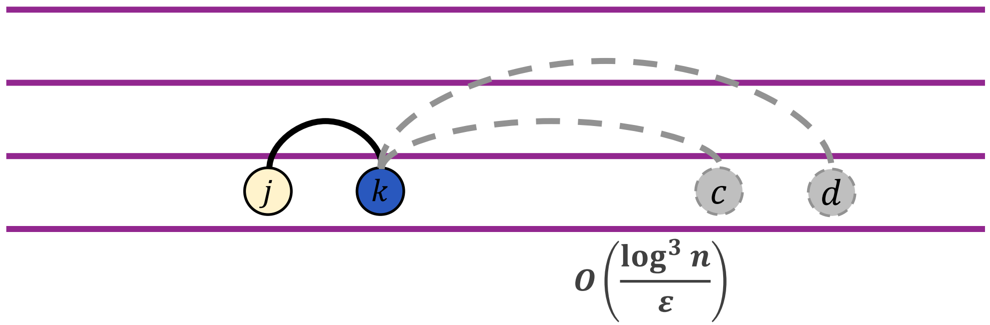
Approximation Proof Sketch

- With high probability, magnitude of each drawn noise is **upper bounded by $O\left(\frac{\log^3 n}{\varepsilon}\right)$**
- **Degree Upper Bound:** If a vertex v is on level $i < 4\log_{1+\eta}(n)$ at end of algorithm, then it has **at most $(1+\eta)^i + O\left(\frac{\log^3 n}{\varepsilon}\right)$ neighbors on levels $\geq i$**



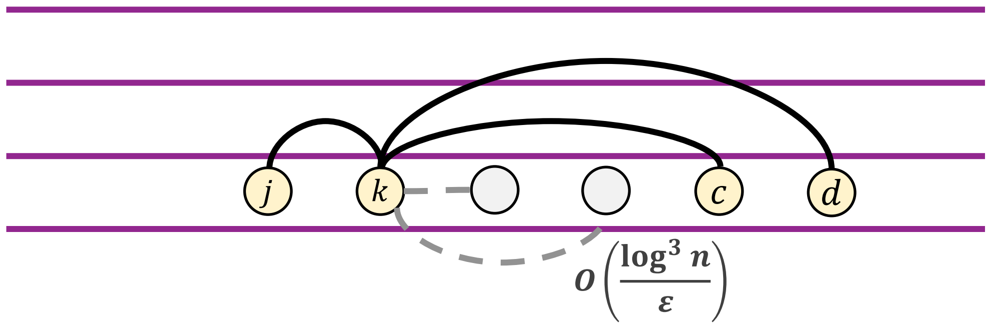
Approximation Proof Sketch

- With high probability, magnitude of each drawn noise is **upper bounded by $O\left(\frac{\log^3 n}{\varepsilon}\right)$**
- **Degree Upper Bound:** If a vertex v is on level $i < 4\log_{1+\eta}(n)$ at end of algorithm, then it has **at most $(1+\eta)^i + O\left(\frac{\log^3 n}{\varepsilon}\right)$ neighbors on levels $\geq i$**



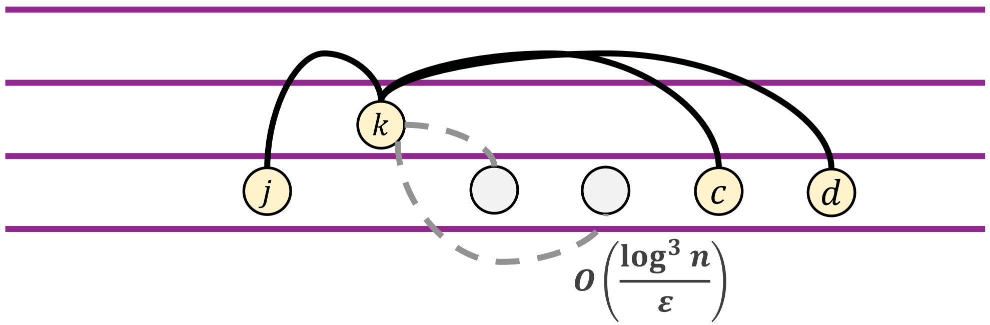
Approximation Proof Sketch

- With high probability, magnitude of each drawn noise is **upper bounded by $O\left(\frac{\log^3 n}{\varepsilon}\right)$**
- **Degree Upper Bound:** If a vertex v is on level $i < 4\log_{1+\eta}(n)$ at end of algorithm, then it has **at most $(1+\eta)^i + O\left(\frac{\log^3 n}{\varepsilon}\right)$ neighbors on levels $\geq i$**
- **Degree Lower Bound:** If a vertex v is on level $i > 0$ at end of algorithm, then it has **at least $(1+\eta)^i - O\left(\frac{\log^3 n}{\varepsilon}\right)$ neighbors on levels $\geq i - 1$**



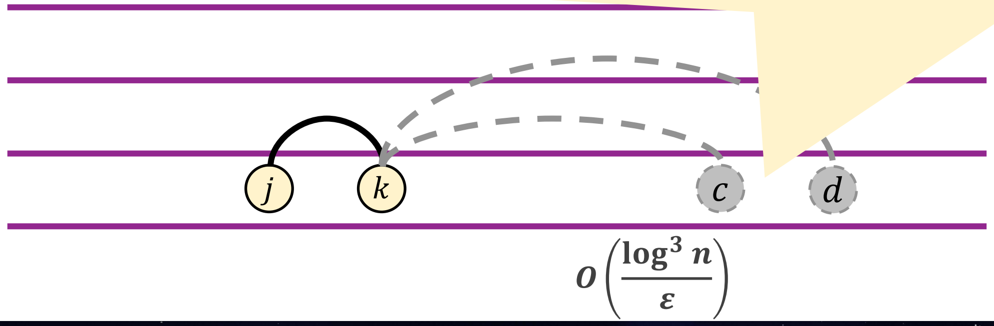
Approximation Proof Sketch

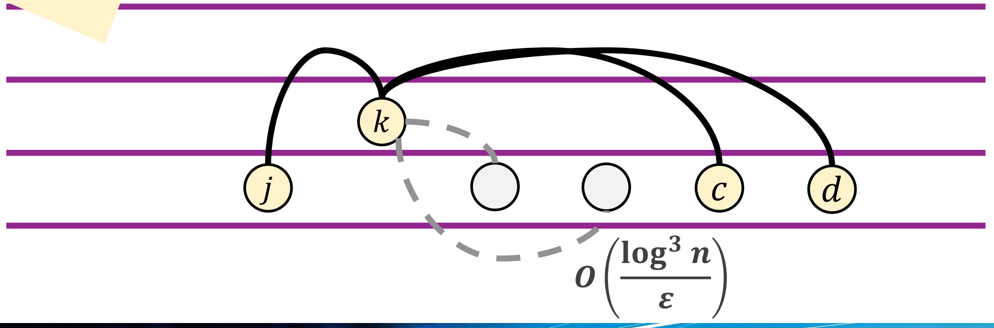
- With high probability, magnitude of each drawn noise is **upper bounded by $O\left(\frac{\log^3 n}{\varepsilon}\right)$**
- **Degree Upper Bound:** If a vertex v is on level $i < 4\log_{1+\eta}(n)$ at end of algorithm, then it has **at most $(1+\eta)^i + O\left(\frac{\log^3 n}{\varepsilon}\right)$ neighbors on levels $\geq i$**
- **Degree Lower Bound:** If a vertex v is on level $i > 0$ at end of algorithm, then it has **at least $(1+\eta)^i - O\left(\frac{\log^3 n}{\varepsilon}\right)$ neighbors on levels $\geq i - 1$**



Approximation Proof Sketch

Key: Largest cutoff increases/decreases by **additive** $O\left(\frac{\log^3 n}{\varepsilon}\right)$





Approximation Proof Sketch

Recently improved to $O\left(\frac{\log(n)}{\varepsilon}\right)$

Key: Largest cutoff increases/decreases by **additive** $O\left(\frac{\log^3 n}{\varepsilon}\right)$