Practical and Accurate Local Edge Differentially Private Graph Algorithms

Pranay Mundra[1], Charalampos Papamanthou (Babis)[1],
Julian Shun[2], Quanquan C. Liu[1]

Research Track 28 : Social Networks

1Yale University, 2 MIT CSAIL

Local Edge $\boxed{\text{Differentially Private}}$

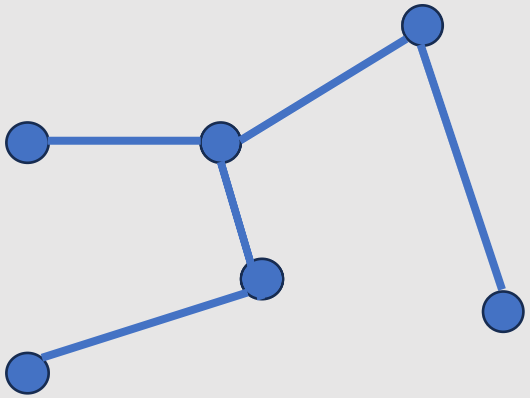
Differential Privacy (DP) [Dwork-McSherry-Nissim-Smith '06]

An algorithm A is ε-differentially private if for all pairs of neighboring inputs G and G' and all sets of possible outputs S, the following holds:

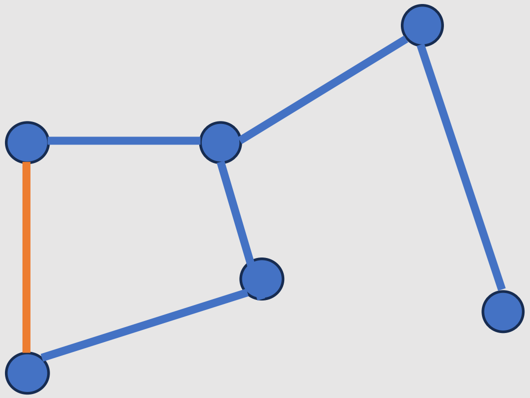
$$Pr[A(G) \in S] \le e^{\varepsilon} \cdot Pr[A(G') \in S]$$

Local $\boxed{\text{Edge}}$ Differentially Private

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



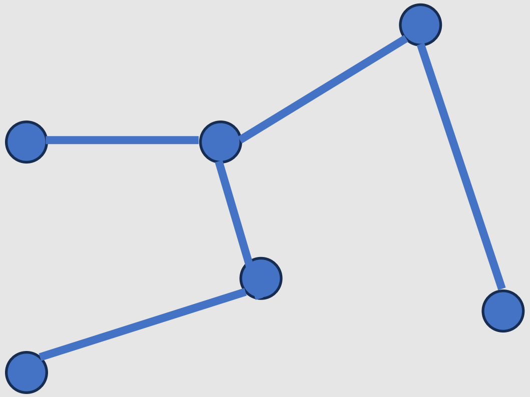
G



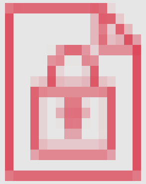
G'

Local Edge Differentially Private

Central Model of DP



Trusted Curator

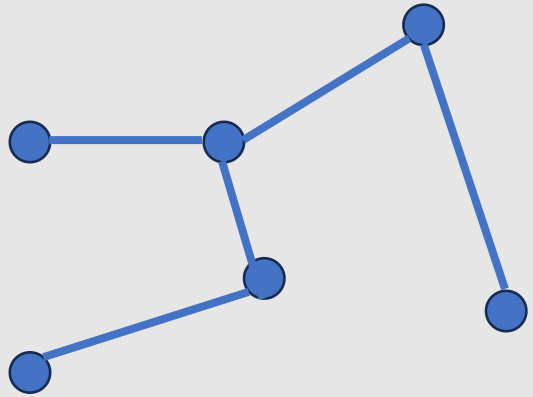


Published DP-Statistics

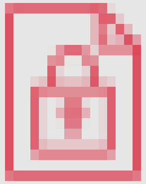
- Trust in a central authority, that collects the raw data
- Runs a DP-mechanism and publishes the privatized statistics

Local Edge Differentially Private

Central Model of DP



Trusted Curator



Published DP-Statistics

TECHNOLOGY

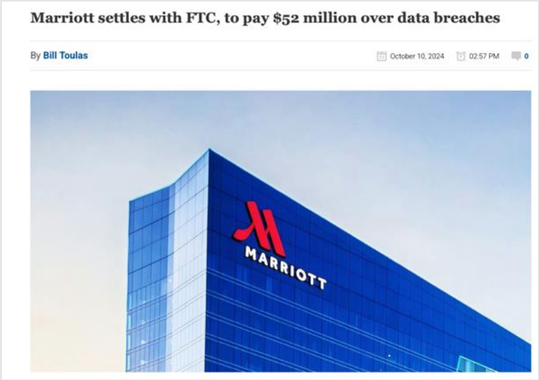
After Data Breach Exposes 530 Million, Facebook Says It Will Not Notify Users

APRIL 9, 2021 · 11:58 PM ET

Emma Bowman

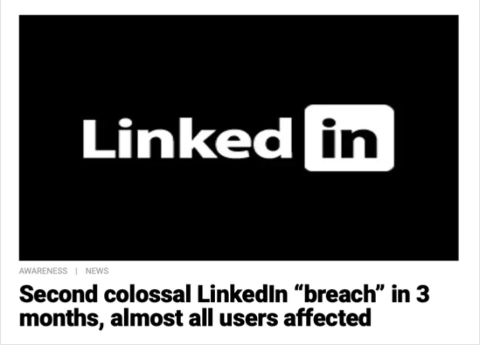
The leaked data includes personal information from 533 million Facebook users in106 countries.

Olivier Douliery/AFP via Getty Images



Marriott settles with FTC, to pay $52 million over data breaches

By Bill Toulas October 10, 2024 02:57 PM 0

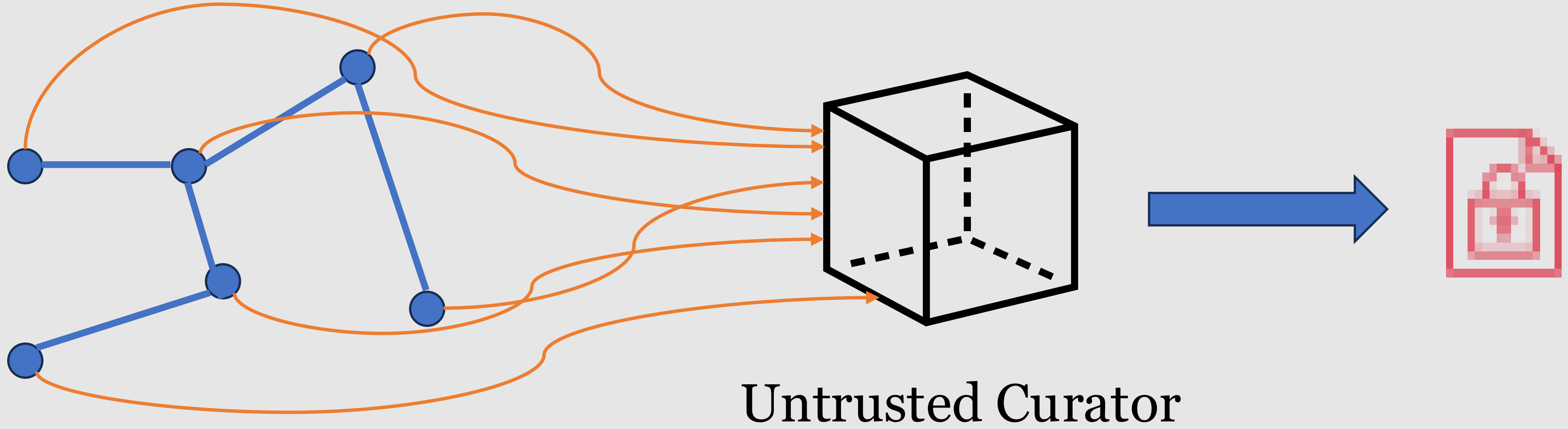


AWARENESS | NEWS

Second colossal LinkedIn "breach" in 3 months, almost all users affected

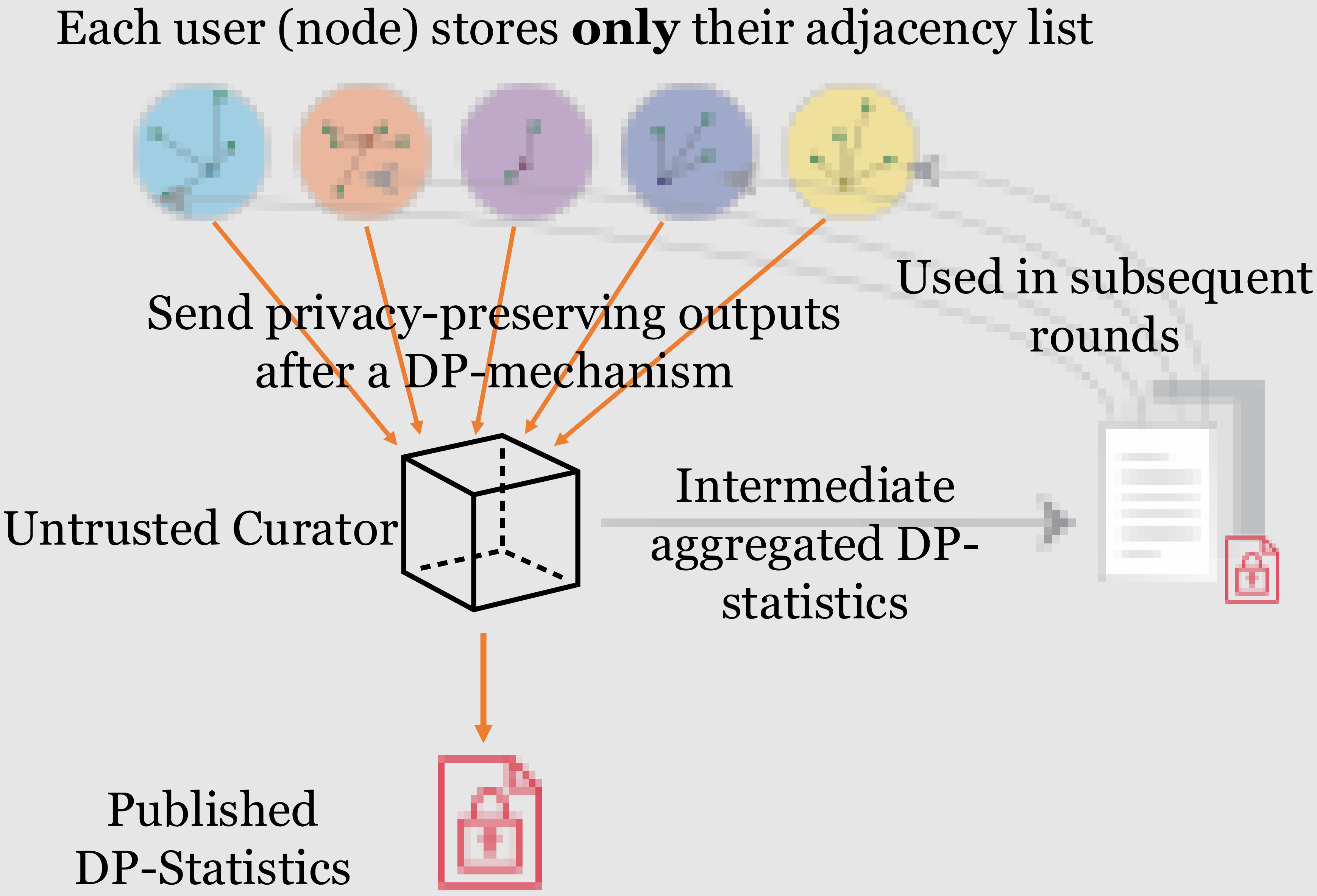
Local Edge Differentially Private

Local Model of DP



- Stronger notion of privacy: Individual trusts no one!
- Each node runs a DP-mechanism independently and sends its privatized output to the curator
- The curator aggregates these outputs to publish the statistics.

LEDP Model



- Round based synchronous algorithm
- Only privacy-preserving outputs are shared
- No raw data is ever shared
- No trust in any individual
- Distributed in nature

Challenges

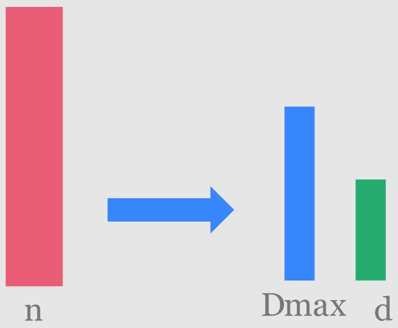
Theory-Practice Gap

Naïve Randomized Response Methods

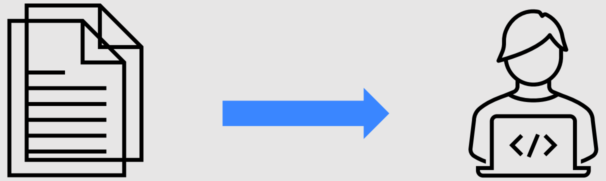
Graph Size Scaled Bounds

No Practical Framework

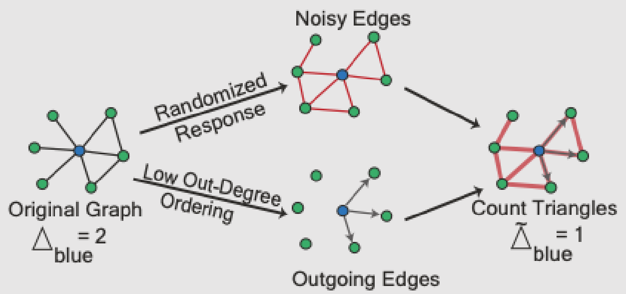
Our Contributions



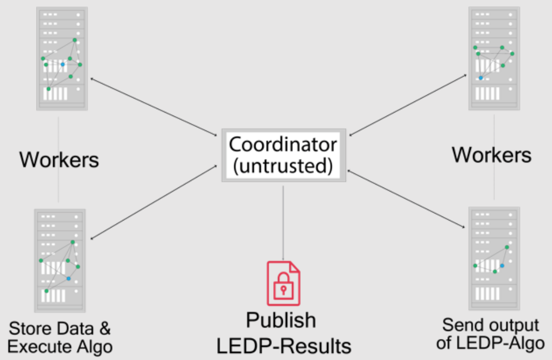
Structure-Scaled Bounds



First Private k-Core Implementation

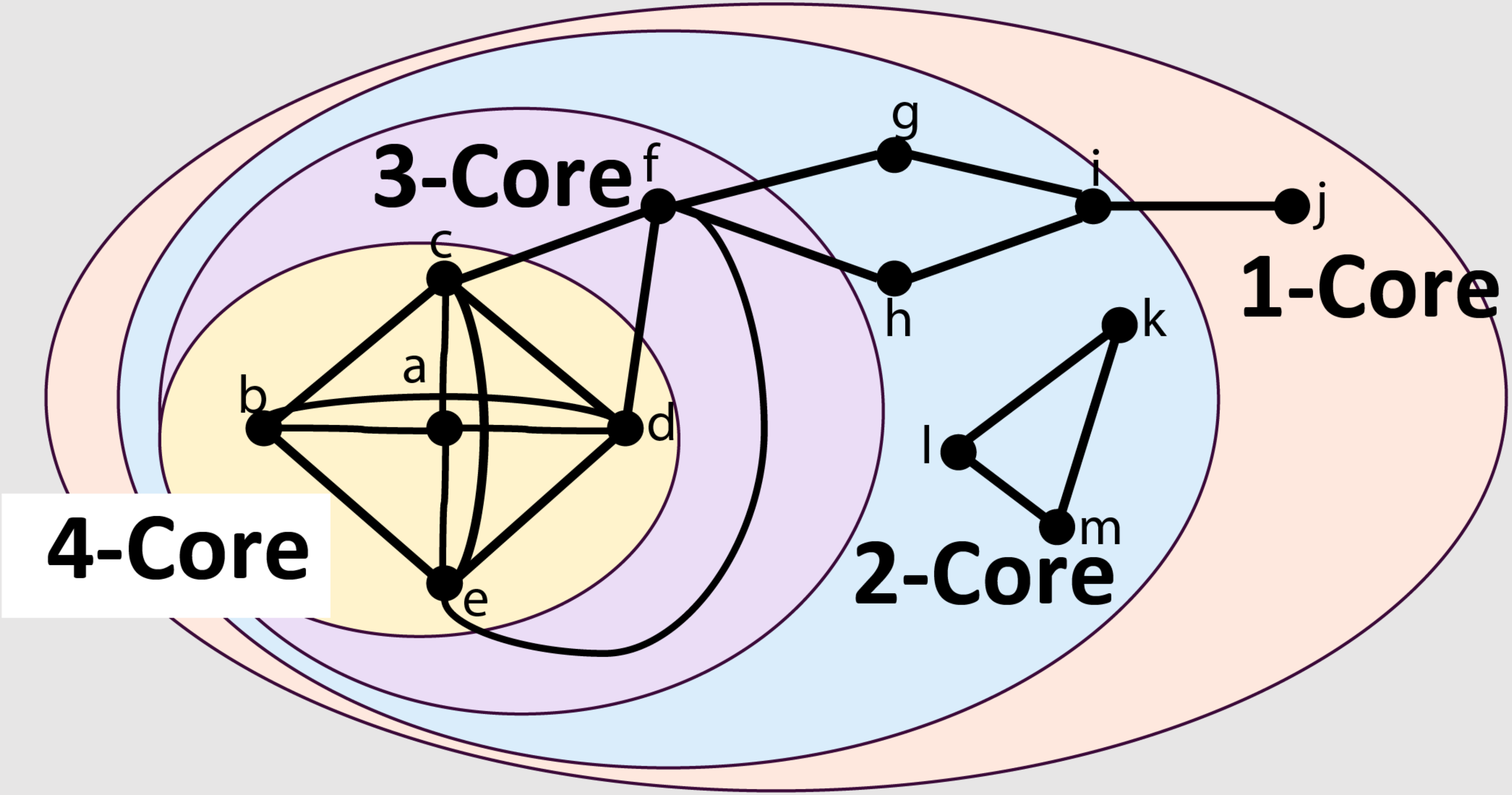


Novel Triangle Counting via k-Core



Distributed LEDP Framework

k-Core Decomposition



LEDP k-Core Decomposition

Sends updated LDS

Drawbacks:

- ➢ Both the approximation error and number of rounds, R, scale with the graph size, i.e., the number of nodes.
- ➢ Large negative noise prevents nodes from moving up in the first few rounds

LDS

re

Numbers

Level Moving update

Untrusted Curator

Individual Users

[1] Dhulipala et. al "Differential Privacy from Locally Adjustable Graph Algorithms: k-Core Decomposition, Low Out-Degree Ordering, and Densest Subgraphs," FOCS `22

k-CoreD Algorithm

Degree Thresholding

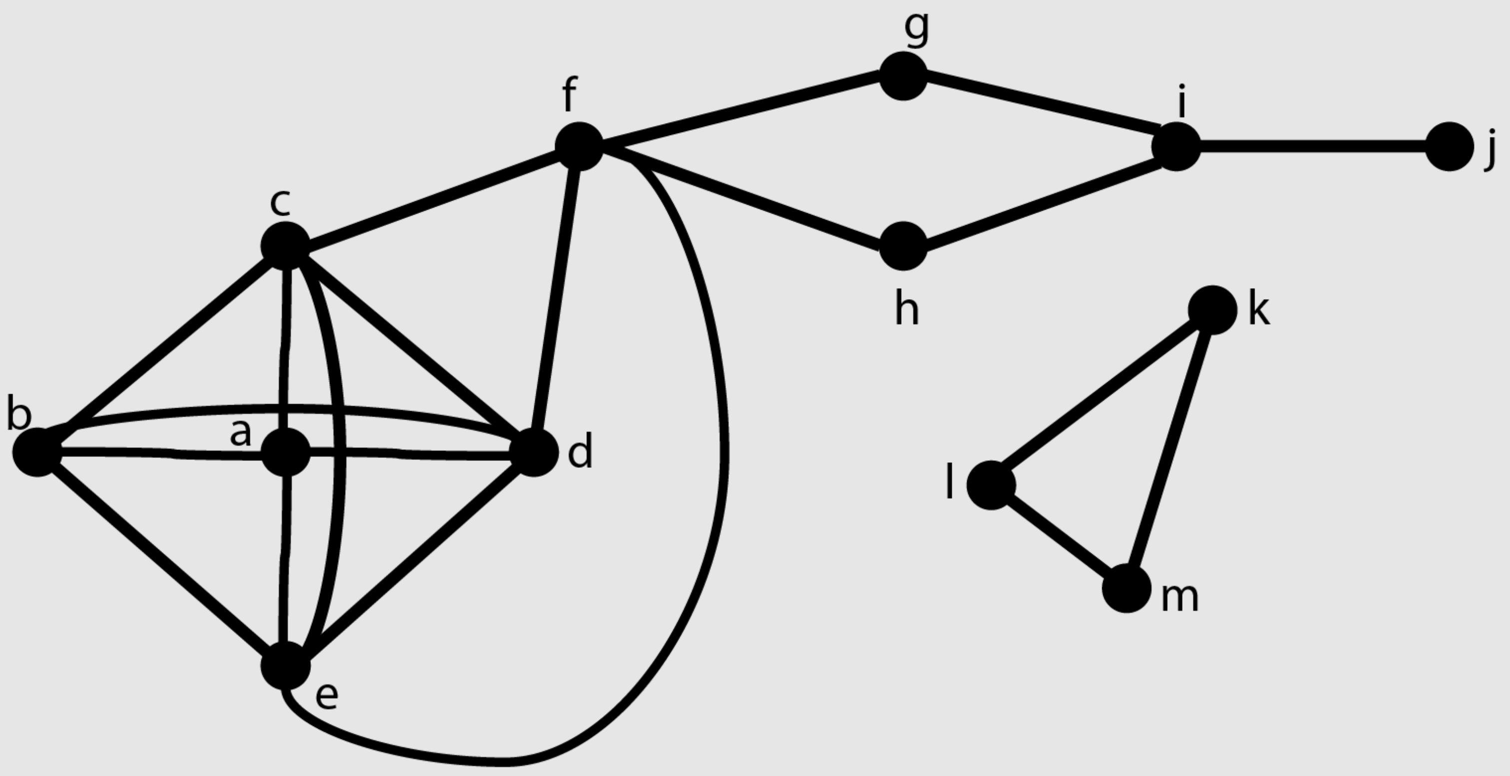
Each node computes a threshold value based on its noisy degree, that determines the max number of rounds the node participates in. Thus, the number of rounds

R ~ max(degree-thresholds)

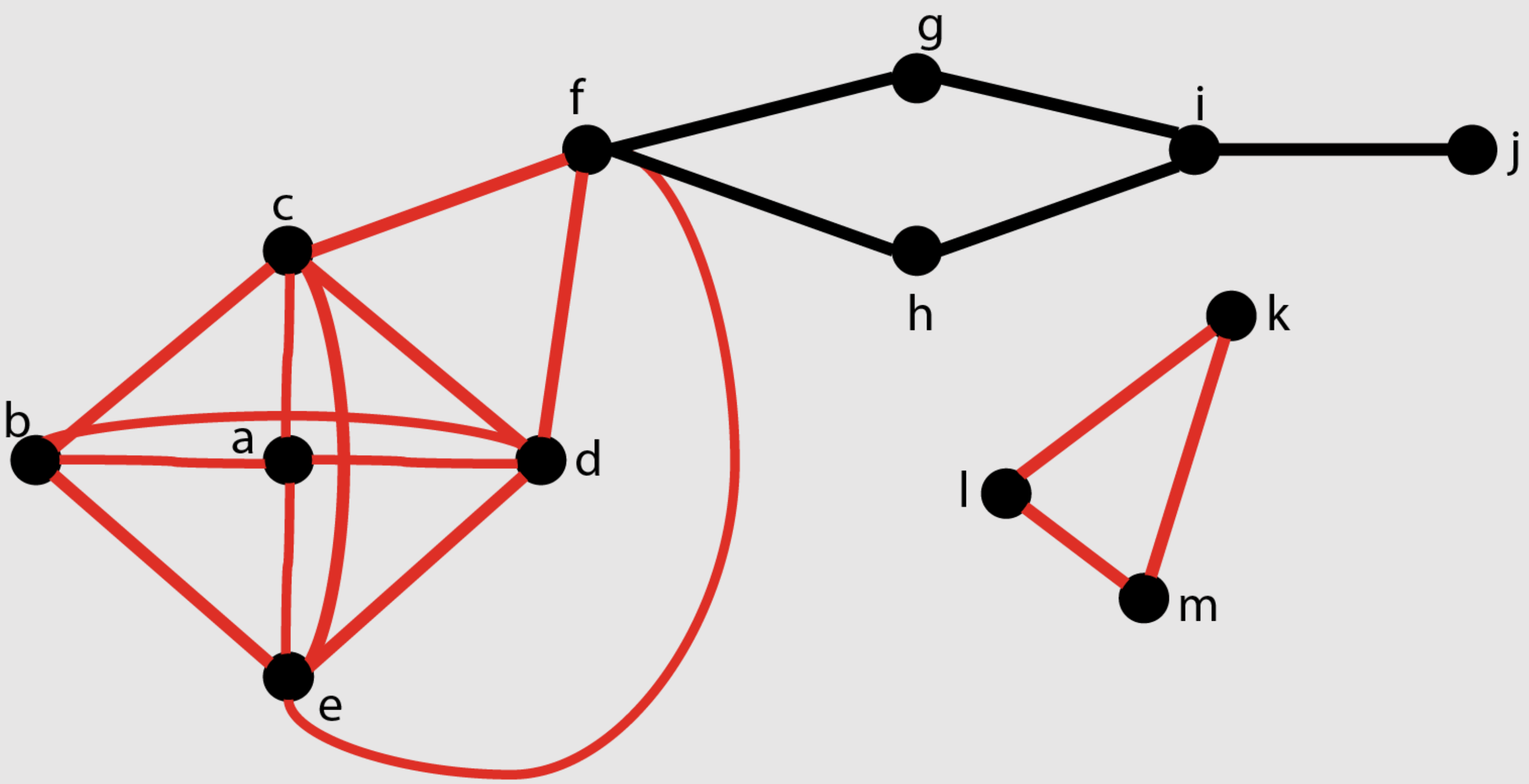
Bias Terms

We use analytically derived bias terms based on the standard deviation of the symmetric geometric distribution to account for the case when large positive or negative noise is drawn

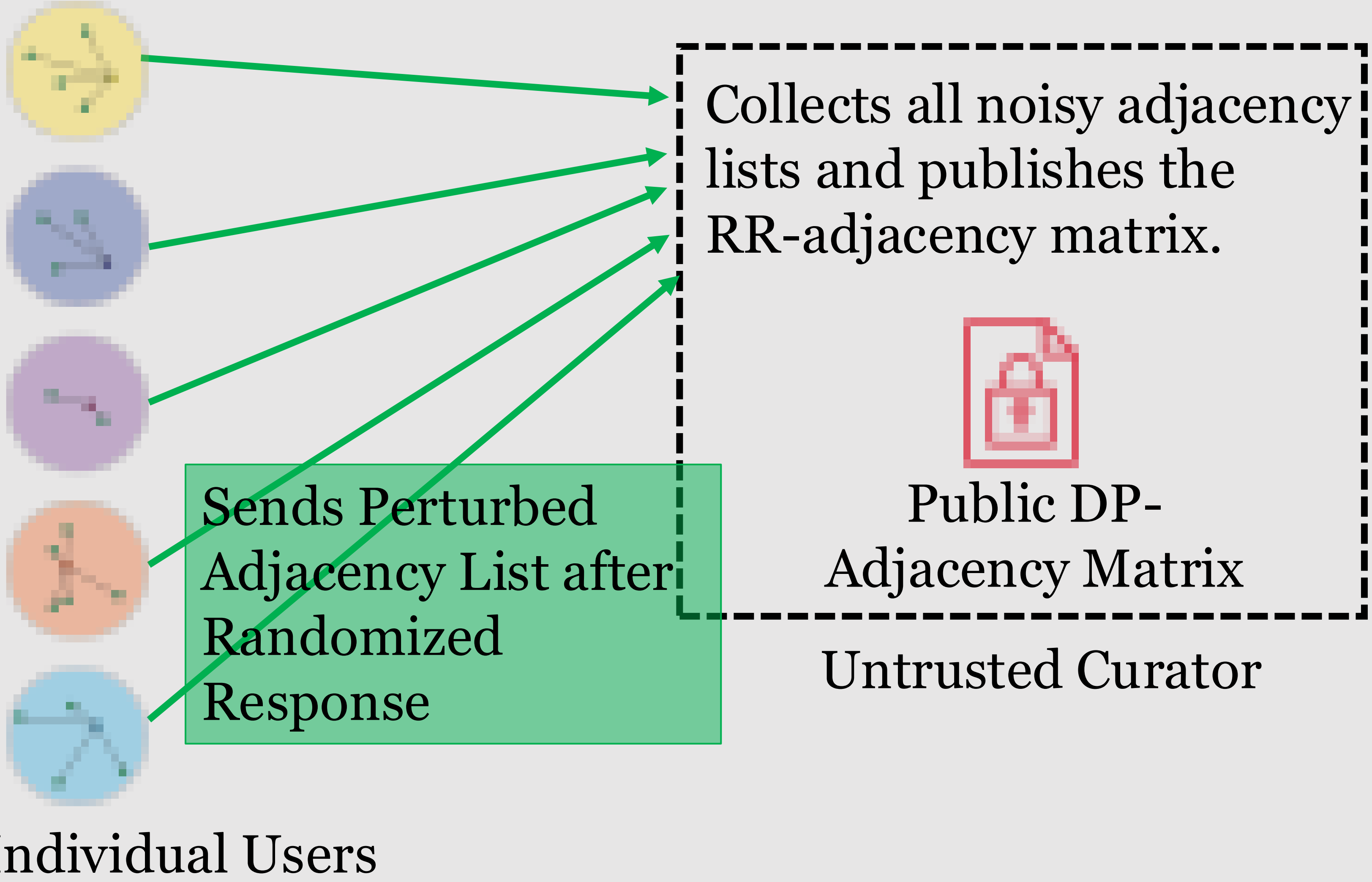
Triangle Counting



Triangle Counting



LEDP Triangle Counting

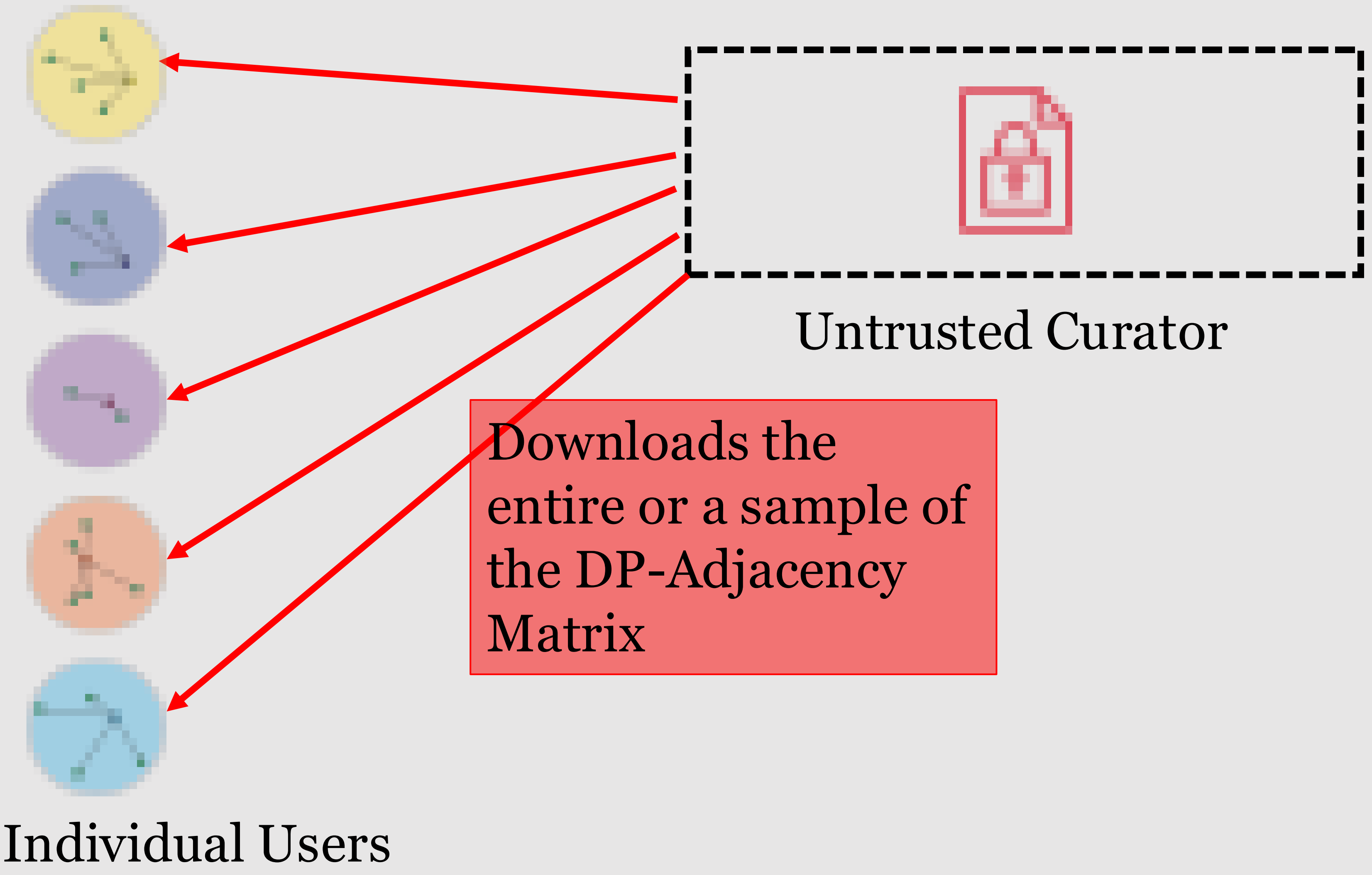


Randomized Response

Independently flips the presence or absence of each edge with a probability based on the privacy parameter

[1] Imola, Jacob, Takao Murakami, and Kamalika Chaudhuri. "{Communication-Efficient} triangle counting under local differential privacy." *USENIX Security `22*
[2] Warner SL. Randomized response: a survey technique for eliminating evasive answer bias. J Am Stat Assoc. 1965 Mar;60(309):63-6. PMID: 12261830.

LEDP Triangle Counting

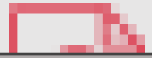


[1] Imola, Jacob, Takao Murakami, and Kamalika Chaudhuri. "{Communication-Efficient} triangle counting under local differential privacy." *USENIX Security `22*
[2] Warner SL. Randomized response: a survey technique for eliminating evasive answer bias. J Am Stat Assoc. 1965 Mar;60(309):63-6. PMID: 12261830.

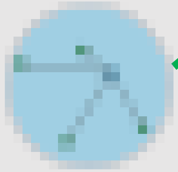
LEDP Triangle Counting

Drawbacks:

- ➢ Relies on either all or a sample of edges after RR
- ➢ Error Bounds scale poorly with graph size



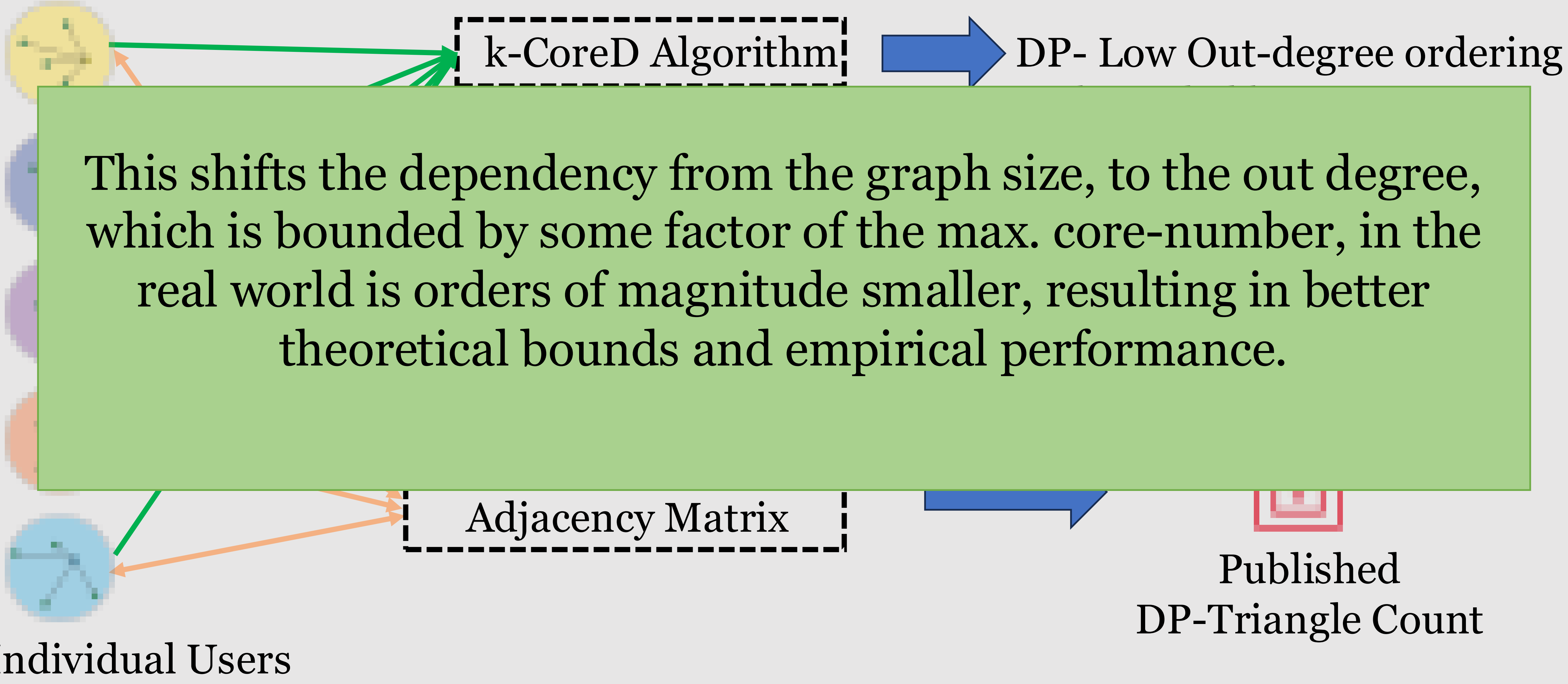
counts



Individual Users

[1] Imola, Jacob, Takao Murakami, and Kamalika Chaudhuri. "{Communication-Efficient} triangle counting under local differential privacy." *USENIX Security `22*
[2] Warner SL. Randomized response: a survey technique for eliminating evasive answer bias. J Am Stat Assoc. 1965 Mar;60(309):63-6. PMID: 12261830.

EdgeOrient$_\Delta$



This shifts the dependency from the graph size, to the out degree, which is bounded by some factor of the max. core-number, in the real world is orders of magnitude smaller, resulting in better theoretical bounds and empirical performance.

Experimental Evaluation

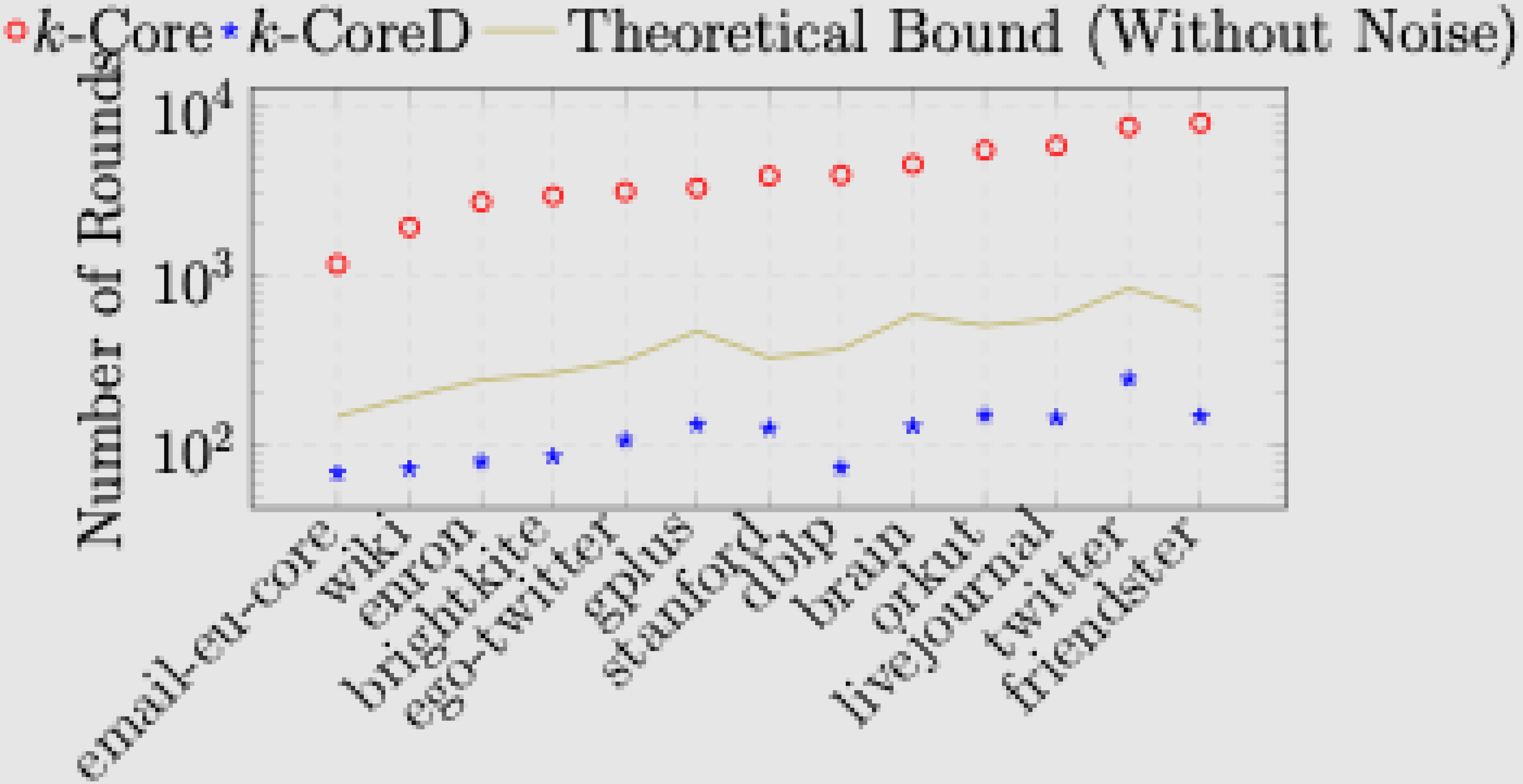
Datasets

Name	Num. Vertices	Num. Edges	Max Degree	Max. Core Num.	Num. Triangles
email-eu-core	986	1,329,336	345	34	105,461
wiki	7115	100,761	1065	35	608,387
enron	36,692	183,830	1383	43	727,044
brightkite	58,228	214,078	1134	52	494,728
ego-twitter	81,306	1,342,296	3383	96	13,082,506
gplus	107,614	12,238,235	20,127	752	1,073,677,742
stanford	281,903	1,992,635	38,625	71	11,329,473
dblp	317,080	1,049,866	343	113	2,224,385
brain	784, 262	267,844,669	21,743	1200	
orkut	3,072,441	117,185,083	33,313	253	
livejournal	4,846,609	42,851,237	20,333	372	
twitter	41,652,230	**1,202,513,046**	2,997,487	2488	
friendster	65,608,366	**1,806,067,135**	5214	304	

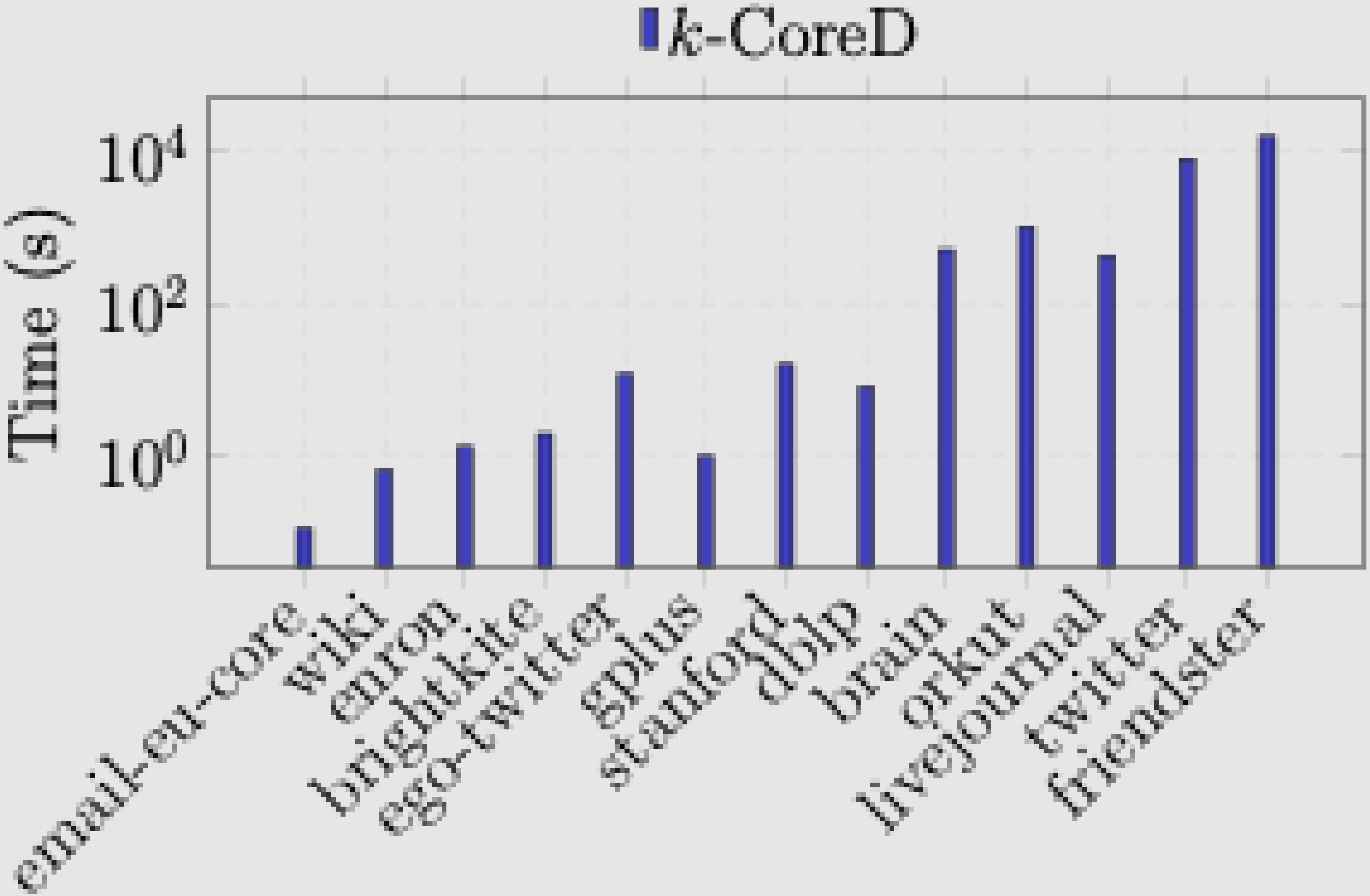
Setup

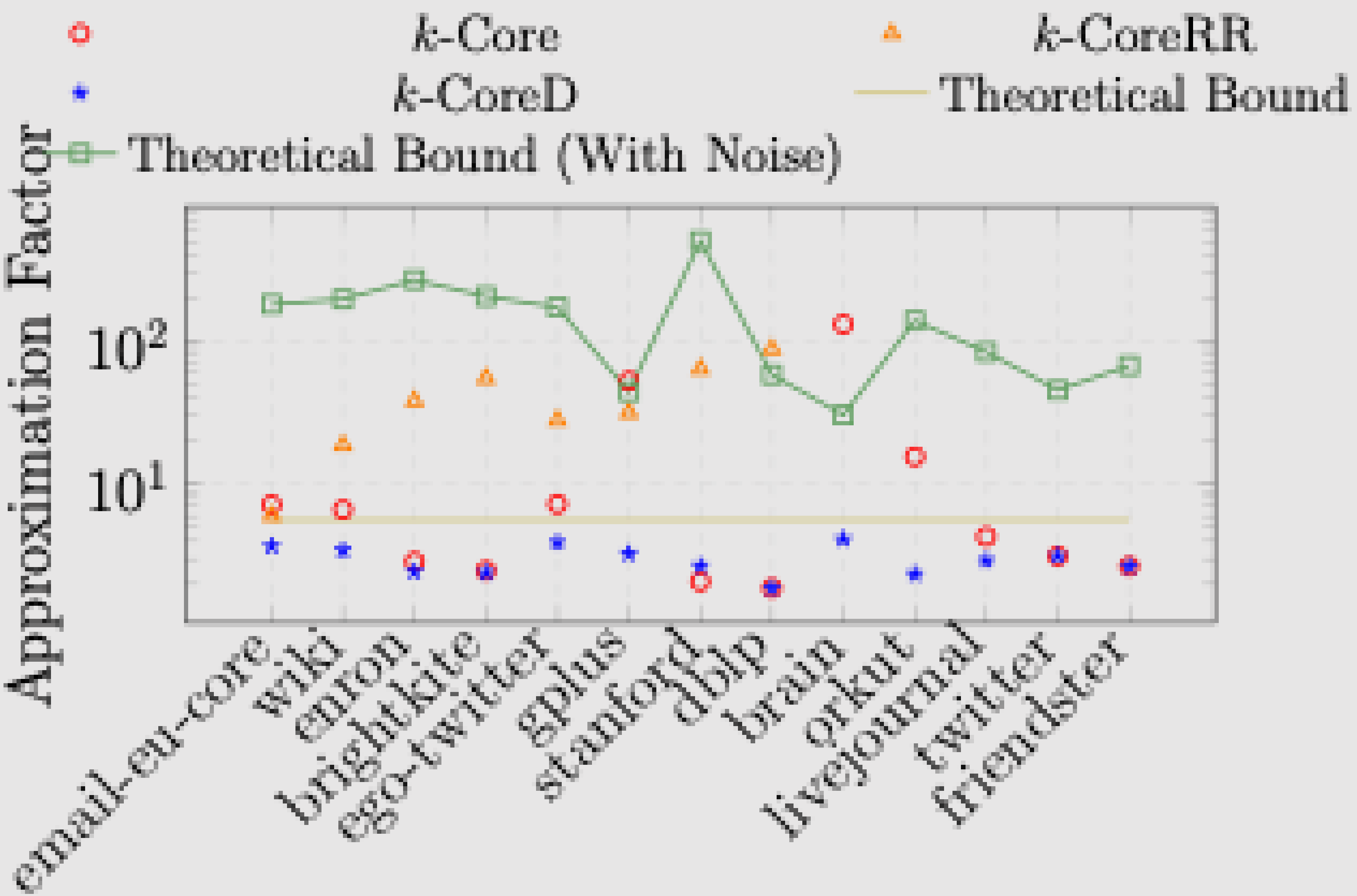
- Compute Resources: We use a Google Cloud c3-standard-176-instance with 88 physical cores, and disable hyperthreading
- Baselines:
  - We implement naïve RR based baselines
  - For k-Core Decomposition we implement the SOTA
  - For Triangle Counting we compare against the SOTA
- Parameters:
  - Epsilon (Privacy Parameter) = 1.0
  - Wall-Clock Limit Per Run = 4 hours
- Error Metric:

$$Approx.Factor = \frac{\max(\hat{k}, k)}{\min(\hat{k}, k)}$$

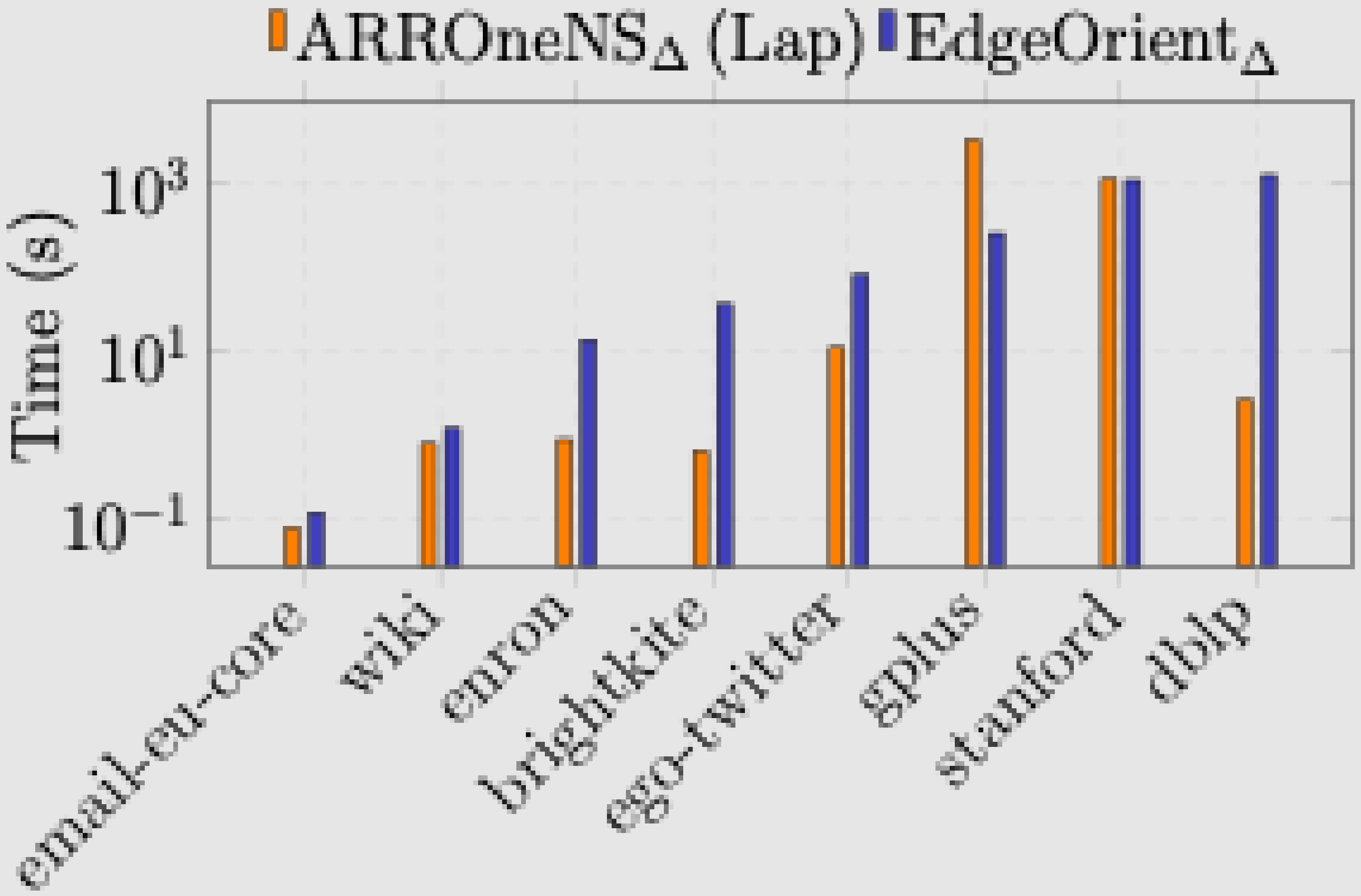


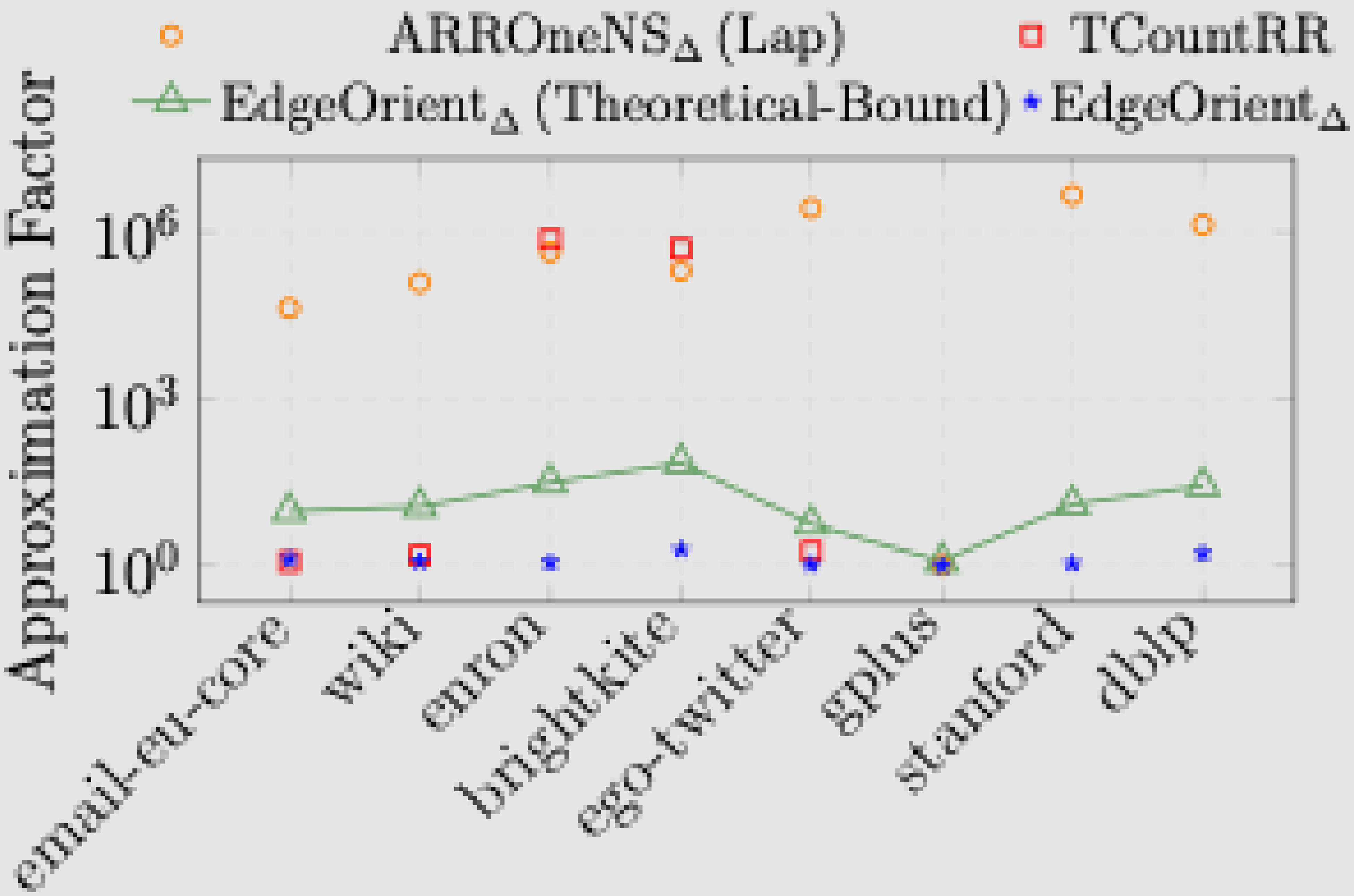
k-Core
k-CoreD
Theoretical Bound (Without Noise)
Number of Rounds
10^4
10^3
10^2
email-eu-core
wiki
enron
brightkite
ego-twitter
gplus
stanford
dblp
brain
orkut
livejournal
twitter
friendster


k-CoreD
Time (s)
10^4
10^2
10^0
email-eu-core
wiki
enron
brightkite
ego-twitter
gplus
stanford
dblp
brain
orkut
livejournal
twitter
friendster



k-Core
k-CoreRR
k-CoreD
Theoretical Bound
Theoretical Bound (With Noise)
Approximation Factor
10^2
10^1
email-eu-core
wiki
enron
brightkite
ego-twitter
gplus
stanford
dblp
brain
orkut
livejournal
twitter
friendster


ARROneNS$_\Delta$ (Lap)
EdgeOrient$_\Delta$
Time (s)
10^3
10^1
10^{-1}
email-eu-core
wiki
enron
brightkite
ego-twitter
gplus
stanford
dblp


ARROneNS$_\Delta$ (Lap)
TCountRR
EdgeOrient$_\Delta$ (Theoretical-Bound)
EdgeOrient$_\Delta$
Approximation Factor
10^6
10^3
10^0
email-eu-core
wiki
enron
brightkite
ego-twitter
gplus
stanford
dblp

Thank You & Questions

Code

Full Paper