dK-Personalization: Publishing Network Statistics with Personalized Differential Privacy

Masooma Iftikhar
Qing Wang
Yang Li

School of Computing
College of Engineering and Computer Science
The Australian National University, Canberra, Australia

May 17, 2022

- Introduction

Outline

- Introduction
- Problem Formulation

Outline

- Introduction
- Problem Formulation
- Sensitivity Analysis

- Introduction
- Problem Formulation
- Sensitivity Analysis
- Proposed Personalized Approaches

Outline

- Introduction
- Problem Formulation
- Sensitivity Analysis
- Proposed Personalized Approaches
- Experiments and Results

Introduction

- Network analysis provides unique insights about social network activities, disease transmission, consumer behaviour, communication patterns, and recommendations.

- Network analysis provides unique insights about social network activities, disease transmission, consumer behaviour, communication patterns, and recommendations.
- However, given the private nature of data about individuals stored in networks, releasing network data raises privacy concerns, thereby requiring privacy-preserving mechanisms.

Motivation

- Network analysis provides unique insights about social network activities, disease transmission, consumer behaviour, communication patterns, and recommendations.
- However, given the private nature of data about individuals stored in networks, releasing network data raises privacy concerns, thereby requiring privacy-preserving mechanisms.
- The current focus of privacy is around differential privacy (DP). However, a uniform privacy level (i.e., ε [1]) is assigned to each individual while guaranteeing DP, which may over or under protect individuals.

[1] Smaller value of ε implies a stronger privacy guarantee.

Personalized differential privacy (PDP) provides freedom to individuals to set their own privacy parameter ε.

Personalized differential privacy (PDP) provides freedom to individuals to set their own privacy parameter ε.

- **Aim:** To publish higher-order network statics such as degree distribution, and joint degree distribution, under (edge or node) PDP.

Personalized differential privacy (PDP) provides freedom to individuals to set their own privacy parameter ε.

- **Aim:** To publish higher-order network statics such as degree distribution, and joint degree distribution, under (edge or node) PDP.
- **Key Challenges:**

Personalized differential privacy (PDP) provides freedom to individuals to set their own privacy parameter ε.

- **Aim:** To publish higher-order network statics such as degree distribution, and joint degree distribution, under (edge or node) PDP.
- **Key Challenges:**
  - Each individual (node) has its own privacy preference whereas each data point in data distribution reflects information about more than one node.

Australian National University

Personalized differential privacy (PDP) provides freedom to individuals to set their own privacy parameter ε.

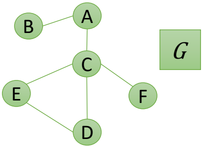
- **Aim:** To publish higher-order network statics such as degree distribution, and joint degree distribution, under (edge or node) PDP.
- **Key Challenges:**
  - Each individual (node) has its own privacy preference whereas each data point in data distribution reflects information about more than one node.
  - Network data is highly sensitive to structural changes under DP.

Problem Formulation

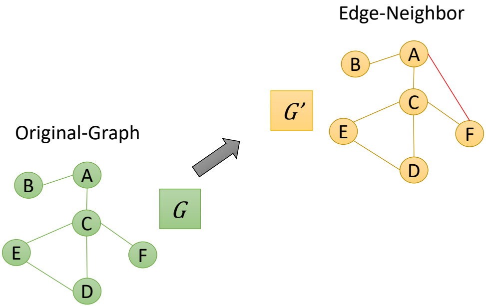
We define the notion of neighboring graphs ($G \sim G'$) under edge and node-DP.

We define the notion of neighboring graphs ($G \sim G'$) under edge and node-DP.

Original-Graph

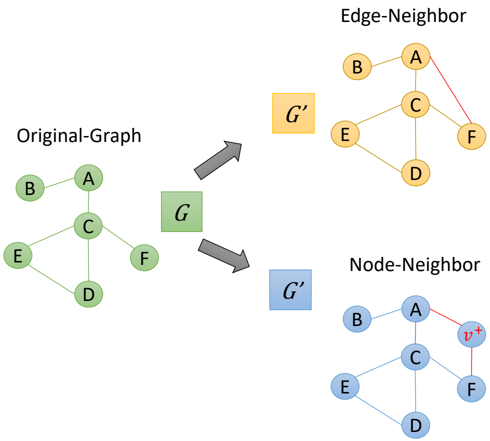


We define the notion of neighboring graphs ($G \sim G'$) under edge and node-DP.



G and G' differ in one edge.

We define the notion of neighboring graphs ($G \sim G'$) under edge and node-DP.



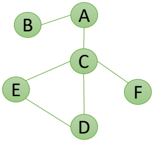
G and *G'* differ in one edge.

G and *G'* differ in one node and the set of edges incident to that node.

The dK-graph model [4] offers a systematized way to extract subgraph degree distributions from a given graph, i.e. dK-distributions.

The dK-graph model [4] offers a systematized way to extract sub-graph degree distributions from a given graph, i.e. dK-distributions.

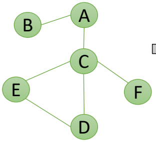
$$G \xrightarrow{1K(G)} \begin{Bmatrix} 1 & 2 \\ 2 & 3 \\ 4 & 1 \end{Bmatrix} \quad \textit{Degree Distribution}$$



DK-DISTRIBUTION

The dK-graph model [4] offers a systematized way to extract sub-graph degree distributions from a given graph, i.e. dK-distributions.

G



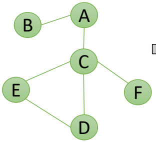
$$\xrightarrow{1K(G)} \left\{ \begin{array}{cc} 1 & 2 \\ 2 & 3 \\ 4 & 1 \end{array} \right\} \quad \textit{Degree Distribution}$$

$$\xrightarrow{2K(G)} \left\{ \begin{array}{l} \langle 1,2 \rangle = 1 \ \ \text{(B-A)} \\ \langle 1,4 \rangle = 1 \ \ \text{(F-C)} \\ \langle 2,2 \rangle = 1 \ \ \text{(E-D)} \\ \langle 2,4 \rangle = 3 \ \ \text{(A-C), (D-C), (E-C)} \end{array} \right\} \quad \textit{Joint Degree Distribution}$$

DK-DISTRIBUTION

The dK-graph model [4] offers a systematized way to extract sub-graph degree distributions from a given graph, i.e. dK-distributions.

G



$1K(G)$ → $\left\{ \begin{array}{cc} 1 & 2 \\ 2 & 3 \\ 4 & 1 \end{array} \right\}$ *Degree Distribution*

$2K(G)$ → *Joint Degree Distribution*

<1,2> = 1 (B-A)
<1,4> = 1 (F-C)
<2,2> = 1 (E-D)
<2,4> = 3 (A-C), (D-C), (E-C)

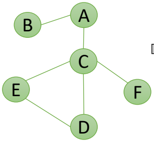
$3K(G)$ → *Clustering Coefficient Distribution*

<1,2,4> = 2 (B-A-C), (F-C-A)
<2,4,2> = 3 (A-C-D), (A-C-E), (D-C-E)

DK-DISTRIBUTION

The dK-graph model [4] offers a systematized way to extract sub-graph degree distributions from a given graph, i.e. dK-distributions.

G



$1K(G)$ → *Degree Distribution*

1	2
2	3
4	1

$2K(G)$ → *Joint Degree Distribution*

<1,2> = 1 (B-A)
<1,4> = 1 (F-C)
<2,2> = 1 (E-D)
<2,4> = 3 (A-C), (D-C), (E-C)

$3K(G)$ → *Clustering Coefficient Distribution*

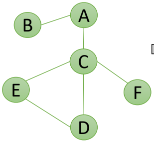
<1,2,4> = 2 (B-A-C), (F-C-A)
<2,4,2> = 3 (A-C-D), (A-C-E), (D-C-E)

- When $d = |V|$, a dK-distribution specifies the entire graph.

DK-DISTRIBUTION

The dK-graph model [4] offers a systematized way to extract sub-graph degree distributions from a given graph, i.e. dK-distributions.

G



$1K(G)$

$$\begin{Bmatrix} 1 & 2 \\ 2 & 3 \\ 4 & 1 \end{Bmatrix}$$

Degree Distribution

$2K(G)$

<1,2> = 1 (B-A)
<1,4> = 1 (F-C)
<2,2> = 1 (E-D)
<2,4> = 3 (A-C), (D-C), (E-C)

Joint Degree Distribution

$3K(G)$

<1,2,4> = 2 (B-A-C), (F-C-A)
<2,4,2> = 3 (A-C-D), (A-C-E), (D-C-E)

Clustering Coefficient Distribution

- When $d = |V|$, a dK-distribution specifies the entire graph.
- $\gamma^{dK}(G)$ queries the dK-distribution of G.

- Given two (edge or node) neighboring graphs $G \sim G'$ where G' is obtained from G by adding (or deleting) an edge (or node) can affect more than one node.

- Given two (edge or node) neighboring graphs $G \sim G'$ where G' is obtained from G by adding (or deleting) an edge (or node) can affect more than one node.
- Thus, PDP should be formalized in terms of all affected nodes to guarantee ε-indistinguishability.

- Given two (edge or node) neighboring graphs $G \sim G'$ where G' is obtained from G by adding (or deleting) an edge (or node) can affect more than one node.
- Thus, PDP should be formalized in terms of all affected nodes to guarantee ε-indistinguishability.
- Given a privacy specification $\Phi = \{\varepsilon_1, \dots \varepsilon_n\}$, denote Φ^v the privacy preference ε of a node v

- **Edge Φ-PDP:** For $G \overset{e}{\sim} G'$, adding (or deleting) an edge affects exactly two nodes u and v.

$$Pr[\mathcal{K}(G) \in \mathcal{O}] \leq e^{min\{\Phi^u, \Phi^v\}} \times Pr[\mathcal{K}(G') \in \mathcal{O}].$$

Australian National University

- **Edge Φ-PDP:** For $G \overset{e}{\sim} G'$, adding (or deleting) an edge affects exactly two nodes u and v.

$$Pr[\mathcal{K}(G) \in \mathcal{O}] \leq e^{min\{\Phi^u, \Phi^v\}} \times Pr[\mathcal{K}(G') \in \mathcal{O}].$$

- **Node Φ-PDP:** For $G \overset{n}{\sim} G'$, adding (or deleting) a node v^+ affects $|E^+|$ nodes incident to v^+ and v^+ itself.

$$Pr[\mathcal{K}(G) \in \mathcal{O}] \leq e^{min\{\Phi^v | (v^+, v) \in E^+\}} \times Pr[\mathcal{K}(G') \in \mathcal{O}] \quad \text{and}$$

$$Pr[\mathcal{K}(G) \in \mathcal{O}] \leq e^{\Phi^{v^+}} \times Pr[\mathcal{K}(G') \in \mathcal{O}]$$

- We want to generate D_Φ that is an anonymized version of D satisfying (edge or node) Φ-PDP.

Australian National University

- We want to generate D_Φ that is an anonymized version of D satisfying (edge or node) Φ-PDP.
- We view the response to γ^{dK} as a collection of responses to degree queries, one for each tuple (entry) in a dK-distribution.

- We want to generate D_Φ that is an anonymized version of D satisfying (edge or node) Φ-PDP.
- We view the response to γ^{dK} as a collection of responses to degree queries, one for each tuple (entry) in a dK-distribution.

(DEGREE QUERY)

A *degree query* $\gamma_q : \gamma^{dK}(G) \rightarrow \mathbb{N}$ maps a degree tuple $d_t \in \gamma^{dK}(G)$ to a frequency value in $\mathbb{N}$ s.t. $(d_t, \gamma_q(G)) \in \gamma^{dK}(G)$.

Sensitivity Analysis

We analyze the sensitivity (Δ) [2] of a single dK-distribution entry, i.e., degree query γ_q rather than the entire dK-distribution γ^{dK}.

[2]The maximum change in γ_q.

We analyze the sensitivity (Δ) [2] of a single dK-distribution entry, i.e., degree query γ_q rather than the entire dK-distribution γ^{dK}.

- $\Delta(\gamma_q)$ of is $|E^+| + 1$ over 1K(G) under node-DP.

[2] The maximum change in γ_q.

We analyze the sensitivity (Δ) [2] of a single dK-distribution entry, i.e., degree query γ_q rather than the entire dK-distribution γ^{dK}.

- $\Delta(\gamma_q)$ of is $|E^+| + 1$ over 1K(G) under node-DP.
- $\Delta(\gamma_q)$ of $(deg(G) + 1) \times |E^+|$ over 2K(G) under node-DP.

[2] The maximum change in γ_q.

We analyze the sensitivity (Δ) [2] of a single dK-distribution entry, i.e., degree query γ_q rather than the entire dK-distribution γ^{dK}.

- $\Delta(\gamma_q)$ of is $|E^+| + 1$ over 1K(G) under node-DP.
- $\Delta(\gamma_q)$ of $(deg(G) + 1) \times |E^+|$ over 2K(G) under node-DP.
- $\Delta(\gamma_q)$ of is 2 over 1K(G) under Edge-DP.

[2] The maximum change in γ_q.

We analyze the sensitivity (Δ) [2] of a single dK-distribution entry, i.e., degree query γ_q rather than the entire dK-distribution γ^{dK}.

- $\Delta(\gamma_q)$ of is $|E^+| + 1$ over 1K(G) under node-DP.
- $\Delta(\gamma_q)$ of $(deg(G) + 1) \times |E^+|$ over 2K(G) under node-DP.
- $\Delta(\gamma_q)$ of is 2 over 1K(G) under Edge-DP.
- $\Delta(\gamma_q)$ of is $2 \times deg(G) + 1$ over 2K(G) under Edge-DP.

[2] The maximum change in γ_q.

We analyze the sensitivity (Δ) [2] of a single dK-distribution entry, i.e., degree query γ_q rather than the entire dK-distribution γ^{dK}.

- $\Delta(\gamma_q)$ of is $|E^+| + 1$ over 1K(G) under node-DP.
- $\Delta(\gamma_q)$ of $(deg(G) + 1) \times |E^+|$ over 2K(G) under node-DP.
- $\Delta(\gamma_q)$ of is 2 over 1K(G) under Edge-DP.
- $\Delta(\gamma_q)$ of is $2 \times deg(G) + 1$ over 2K(G) under Edge-DP.

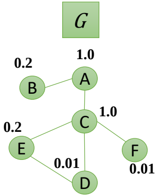
We observe that the sensitivity of γ_q is half as compared to γ^{dK}.

[2] The maximum change in γ_q.

Proposed Personalized Approaches

Local Least Based Personalized Perturbation: LL-dK perturbs entries with the strongest local ε.

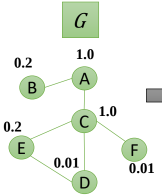
Local Least Based Personalized Perturbation: LL-dK perturbs entries with the strongest local ε.



Personalized Approach I

Australian National University

Local Least Based Personalized Perturbation: LL-dK perturbs entries with the strongest local ε.



$\Upsilon^{1K}(G)$ = 1K(G)

1	2	(B,F)
2	3	(A,D,E)
4	1	(C)

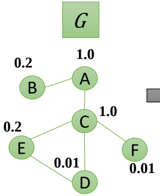
1K–distribution of G

$\Upsilon^{2K}(G)$ = 2K(G)

<1,2> = 1	(B-A)
<1,4> = 1	(F-C)
<2,2> = 1	(E-D)
<2,4> = 3	(A-C), (D-C), (E-C)

2K–distribution of G

Local Least Based Personalized Perturbation: LL-dK perturbs entries with the strongest local ε.



$\Upsilon^{1K}(G) = \text{1K(G)}$

1	2	(B,F)
2	3	(A,D,E)
4	1	(C)

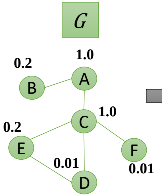
1K–distribution of G

$\Upsilon^{2K}(G) = \text{2K(G)}$

<1,2> = 1	(B-A)
<1,4> = 1	(F-C)
<2,2> = 1	(E-D)
<2,4> = 3	(A-C), (D-C), (E-C)

2K–distribution of G

Local Least Based Personalized Perturbation: LL-dK perturbs entries with the strongest local ε.



$\Upsilon^{1K}(G) = 1\text{K(G)}$

1	2	(B,F)
2	3	(A,D,E)
4	1	(C)

1K–distribution of G

$\Upsilon^{2K}(G) = 2\text{K(G)}$

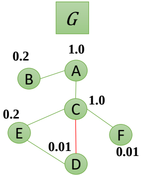
$\langle 1,2 \rangle = 1$	(B-A)
$\langle 1,4 \rangle = 1$	(F-C)
$\langle 2,2 \rangle = 1$	(E-D)
$\langle 2,4 \rangle = 3$	(A-C), (D-C), (E-C)

2K–distribution of G

The frequency value 2 in $1K(G)$, and the frequency value 3 in $2K(G)$ are perturbed with $\varepsilon = min(\Phi^B, \Phi^F)$, and $\varepsilon = min(\Phi^A, \Phi^C, \Phi^D, \Phi^E)$, respectively.

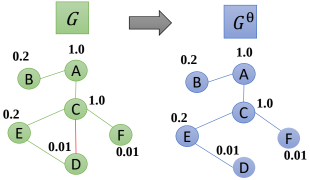
Threshold Projection Based Personalized Perturbation: TP-dK transforms a graph into a θ-bounded graph then removes all nodes with $\varepsilon < \tau$.

Threshold Projection Based Personalized Perturbation: TP-dK transforms a graph into a θ-bounded graph then removes all nodes with $\varepsilon < \tau$.



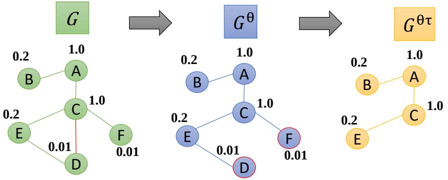
Australian National University

Threshold Projection Based Personalized Perturbation: TP-dK transforms a graph into a θ-bounded graph then removes all nodes with $\varepsilon < \tau$.



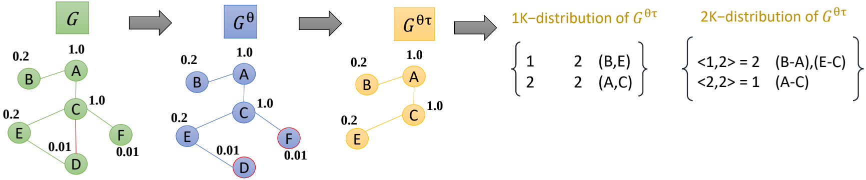
Australian National University

Threshold Projection Based Personalized Perturbation: TP-dK transforms a graph into a θ-bounded graph then removes all nodes with $\varepsilon < \tau$.



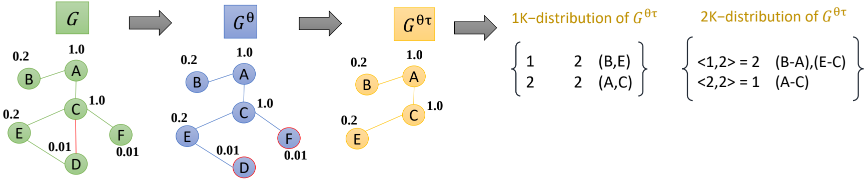
Personalized Approach II

Threshold Projection Based Personalized Perturbation: TP-dK transforms a graph into a θ-bounded graph then removes all nodes with $\varepsilon < \tau$.



Personalized Approach II

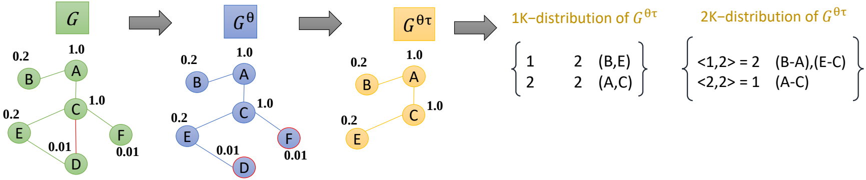
Threshold Projection Based Personalized Perturbation: TP-dK transforms a graph into a θ-bounded graph then removes all nodes with $\varepsilon < \tau$.



Since $deg(G) \leq \theta$, the sensitivity of γ_q is reduced.

PERSONALIZED APPROACH II

Threshold Projection Based Personalized Perturbation: TP-dK transforms a graph into a θ-bounded graph then removes all nodes with $\varepsilon < \tau$.

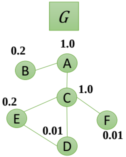


Since $deg(G) \leq \theta$, the sensitivity of γ_q is reduced.

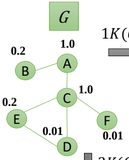
With threshold τ all nodes with high privacy are removed.

Sampling Based Personalized Perturbation: ST-dK first splits entries, and then samples them with non-uniform probabilities.

Sampling Based Personalized Perturbation: ST-dK first splits entries, and then samples them with non-uniform probabilities.



Sampling Based Personalized Perturbation: ST-dK first splits entries, and then samples them with non-uniform probabilities.



$\Upsilon^{1K}(G) = 1\mathrm{K(G)}$

$1K(G) \Rightarrow$

1	2	(B,F)
2	3	(A,D,E)
4	1	(C)

1K–distribution of G

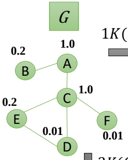
$\Upsilon^{2K}(G) = 2\mathrm{K(G)}$

$2K(G) \Rightarrow$

<1,2> = 1	(B-A)
<1,4> = 1	(F-C)
<2,2> = 1	(E-D)
<2,4> = 3	(A-C), (D-C), (E-C)

2K–distribution of G

Sampling Based Personalized Perturbation: ST-dK first splits entries, and then samples them with non-uniform probabilities.



$1K(G)$ →

$\Upsilon^{1K}(G) = $ 1K(G)

1	2	(B,F)
2	3	(A,D,E)
4	1	(C)

1K–distribution of G

Split →

1	1	(B)
1	1	(F)
2	1	(A)
2	1	(D)
2	1	(E)
4	1	(C)

$2K(G)$ →

$\Upsilon^{2K}(G) = $ 2K(G)

<1,2> = 1	(B-A)
<1,4> = 1	(F-C)
<2,2> = 1	(E-D)
<2,4> = 3	(A-C), (D-C), (E-C)

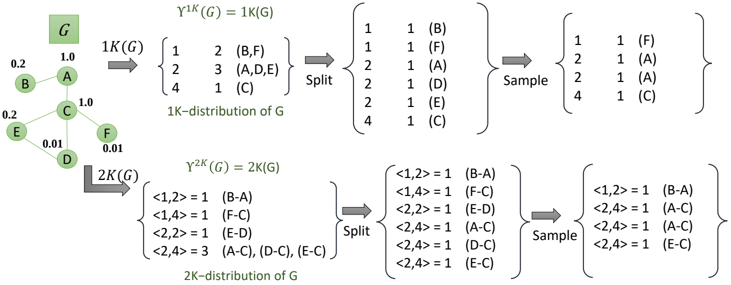
2K–distribution of G

Split →

<1,2> = 1	(B-A)
<1,4> = 1	(F-C)
<2,2> = 1	(E-D)
<2,4> = 1	(A-C)
<2,4> = 1	(D-C)
<2,4> = 1	(E-C)

Personalized Approach III

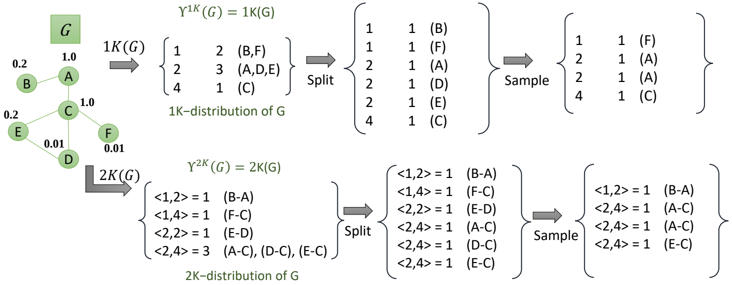
Sampling Based Personalized Perturbation: ST-dK first splits entries, and then samples them with non-uniform probabilities.



Personalized Approach III

Australian National University

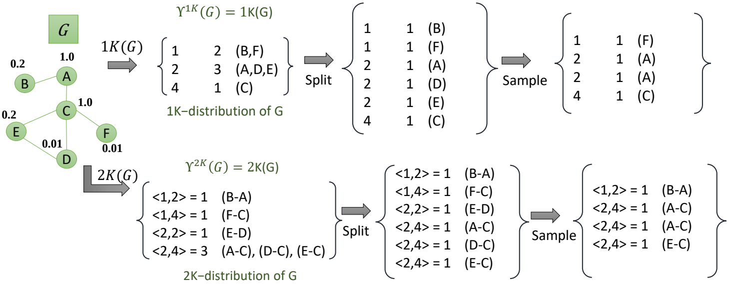
Sampling Based Personalized Perturbation: ST-dK first splits entries, and then samples them with non-uniform probabilities.



Inclusion probability for each entry depends on corresponding ε and global threshold τ.

Personalized Approach III

Sampling Based Personalized Perturbation: ST-dK first splits entries, and then samples them with non-uniform probabilities.



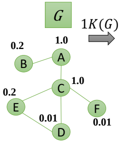
Inclusion probability for each entry depends on corresponding ε and global threshold τ.

Sampled dK-distribution is perturbed with τ.

Aggregation Based Personalized Perturbation: AG-dK computes corresponding ε values to performs aggregation over dK-distribution.

Personalized Approach IV

Aggregation Based Personalized Perturbation: AG-dK computes corresponding ε values to performs aggregation over dK-distribution.



$1K(G)$ → $\Upsilon^{1K}(G) = $ 1K(G)

1	2	(B,F)
2	3	(A,D,E)
4	1	(C)

1K–distribution of G

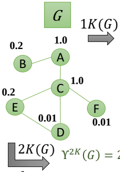
$2K(G)$ → $\Upsilon^{2K}(G) = $ 2K(G)

<1,2> = 1	(B-A)
<1,4> = 1	(F-C)
<2,2> = 1	(E-D)
<2,4> = 3	(A-C), (D-C), (E-C)

2K–distribution of G

Australian National University

Aggregation Based Personalized Perturbation: AG-dK computes corresponding ε values to performs aggregation over dK-distribution.



$\Upsilon^{1K}(G) = 1K(G)$

1	2	(B,F)
2	3	(A,D,E)
4	1	(C)

1K–distribution of G

0.01	0.2	-
0.01	0.2	1.0
1.0	-	-

Corresponding privacy preferences

$\Upsilon^{2K}(G) = 2K(G)$

<1,2> = 1	(B-A)
<1,4> = 1	(F-C)
<2,2> = 1	(E-D)
<2,4> = 3	(A-C), (D-C), (E-C)

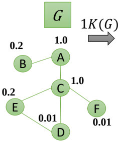
2K–distribution of G

0.2	1.0	-	-
0.01	1.0	-	-
0.01	0.2	-	-
0.01	0.2	1.0	1.0

Corresponding privacy preferences

Australian National University

Aggregation Based Personalized Perturbation: AG-dK computes corresponding ε values to performs aggregation over dK-distribution.



$1K(G)$ → $\Upsilon^{1K}(G) = 1K(G)$

$$\begin{Bmatrix} 1 & 2 & (B,F) \\ 2 & 3 & (A,D,E) \\ 4 & 1 & (C) \end{Bmatrix}$$

1K–distribution of G

0.01	0.2	-
0.01	0.2	1.0
1.0	-	-

Corresponding privacy preferences

Aggregation →

Partition

$$\begin{Bmatrix} 1 & 2 & (B,F) \\ 2 & 3 & (A,D,E) \\ \text{-----} & & \\ 4 & 1 & (C) \end{Bmatrix}$$

Aggregation

$$\begin{Bmatrix} \begin{Bmatrix} 1 \\ 2 \end{Bmatrix} = 5 \ \ (B,F), (A,D,E) \\ \begin{Bmatrix} 4 \end{Bmatrix} = 1 \ \ (C) \end{Bmatrix}$$

$2K(G)$ → $\Upsilon^{2K}(G) = 2K(G)$

$$\begin{Bmatrix} <1,2> = 1 & (B\text{-}A) \\ <1,4> = 1 & (F\text{-}C) \\ <2,2> = 1 & (E\text{-}D) \\ <2,4> = 3 & (A\text{-}C), (D\text{-}C), (E\text{-}C) \end{Bmatrix}$$

2K–distribution of G

0.2	1.0	-	-
0.01	1.0	-	-
0.01	0.2	-	-
0.01	0.2	1.0	1.0

Corresponding privacy preferences

Aggregation →

Partition

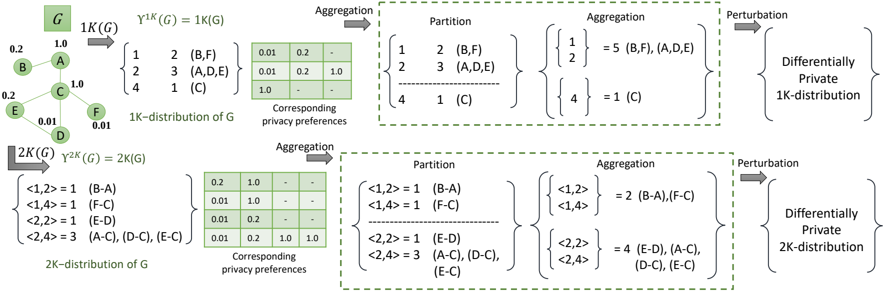
$$\begin{Bmatrix} <1,2> = 1 & (B\text{-}A) \\ <1,4> = 1 & (F\text{-}C) \\ \text{-----} & \\ <2,2> = 1 & (E\text{-}D) \\ <2,4> = 3 & (A\text{-}C), (D\text{-}C), (E\text{-}C) \end{Bmatrix}$$

Aggregation

$$\begin{Bmatrix} \begin{Bmatrix} <1,2> \\ <1,4> \end{Bmatrix} = 2 \ \ (B\text{-}A),(F\text{-}C) \\ \begin{Bmatrix} <2,2> \\ <2,4> \end{Bmatrix} = 4 \ \ (E\text{-}D), (A\text{-}C), (D\text{-}C), (E\text{-}C) \end{Bmatrix}$$

Australian National University

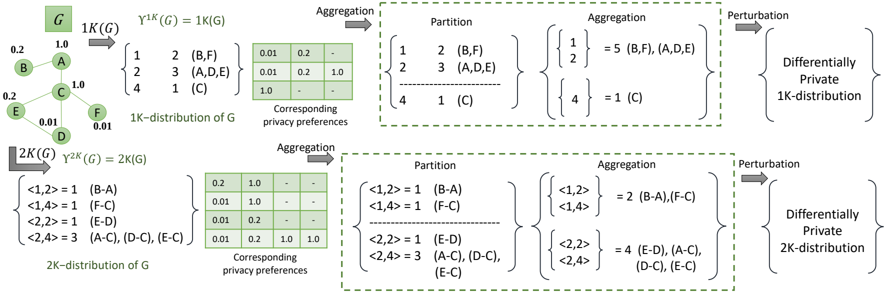
Aggregation Based Personalized Perturbation: AG-dK computes corresponding ε values to performs aggregation over dK-distribution.



Personalized Approach IV

Australian National University

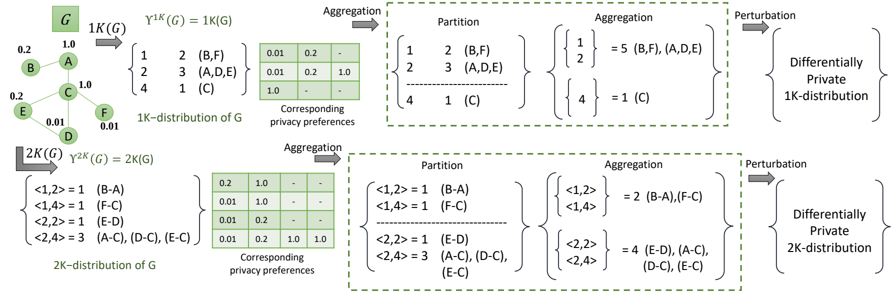
Aggregation Based Personalized Perturbation: AG-dK computes corresponding ε values to performs aggregation over dK-distribution.



Entries are perturbed with the strongest local ε corresponding to each partition.

Personalized Approach IV

Aggregation Based Personalized Perturbation: AG-dK computes corresponding ε values to performs aggregation over dK-distribution.



Entries are perturbed with the strongest local ε corresponding to each partition.

γ_q is approximated to $\gamma_q \circ \mathcal{M}$.

Experiments and Results

- Four network datasets:

- Four network datasets:

  (1) *Facebook* contains 4,039 nodes and 88,234 edges.

  (2) *Wiki-Vote* contains 7,115 nodes and 103,689 edges.

  (3) *Ca-HepPh* contains 12,008 nodes and 118,521 edges.

  (4) *Email-Enron* contains 36,692 nodes and 183,831 edges.

- Four network datasets:
  1. *Facebook* contains 4,039 nodes and 88,234 edges.
  2. *Wiki-Vote* contains 7,115 nodes and 103,689 edges.
  3. *Ca-HepPh* contains 12,008 nodes and 118,521 edges.
  4. *Email-Enron* contains 36,692 nodes and 183,831 edges.
- Two utility metrics [1]:

- Four network datasets:

  (1) *Facebook* contains 4,039 nodes and 88,234 edges.

  (2) *Wiki-Vote* contains 7,115 nodes and 103,689 edges.

  (3) *Ca-HepPh* contains 12,008 nodes and 118,521 edges.

  (4) *Email-Enron* contains 36,692 nodes and 183,831 edges.

- Two utility metrics [1]:

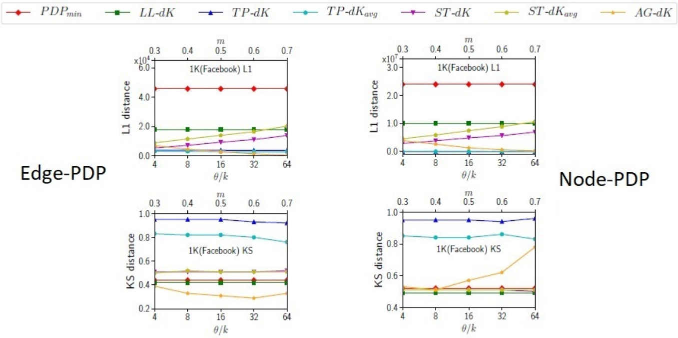
  - *L1 distance* measures the network structural error the original dK-distribution D and its private version D_Φ by calculating $\|D - D_\Phi\|_1 = \sum_{i=1}^{deg(G)} |D_i - D_{\Phi_i}|$.

- Four network datasets:
  1. *Facebook* contains 4,039 nodes and 88,234 edges.
  2. *Wiki-Vote* contains 7,115 nodes and 103,689 edges.
  3. *Ca-HepPh* contains 12,008 nodes and 118,521 edges.
  4. *Email-Enron* contains 36,692 nodes and 183,831 edges.
- Two utility metrics [1]:
  - *L1 distance* measures the network structural error the original dK-distribution D and its private version D_Φ by calculating $\|D - D_\Phi\|_1 = \sum_{i=1}^{deg(G)} |D_i - D_{\Phi_i}|$.
  - *KS distance* measures the closeness between the cumulative distribution functions of D and D_Φ by calculating $KS(D, D_\Phi) = max_i |CDF_{D_i} - CDF_{D_{\Phi_i}}|$.

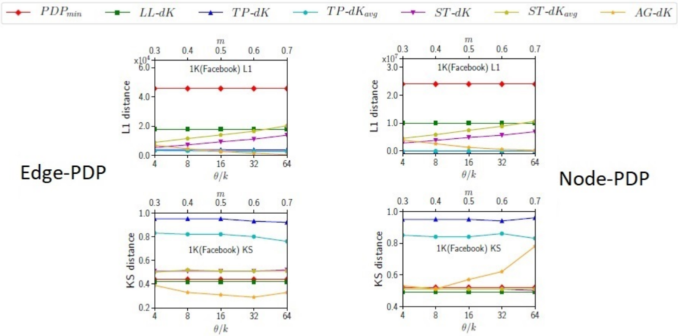
Does the proposed personalized approaches yield more utility in 1K-distribution under edge-PDP and node-PDP?

EXPERIMENTS I

Does the proposed personalized approaches yield more utility in 1K-distribution under edge-PDP and node-PDP?

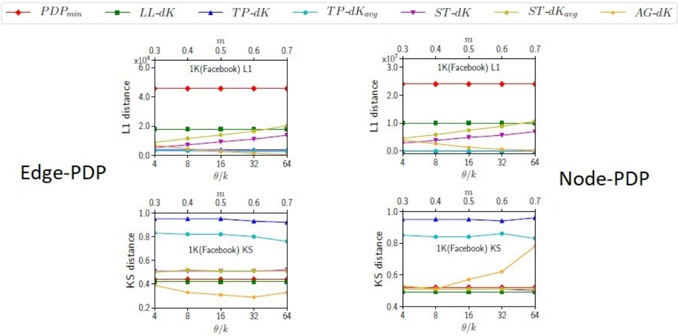


Does the proposed personalized approaches yield more utility in 1K-distribution under edge-PDP and node-PDP?



- Our methods yield less network structural error.

Does the proposed personalized approaches yield more utility in 1K-distribution under edge-PDP and node-PDP?

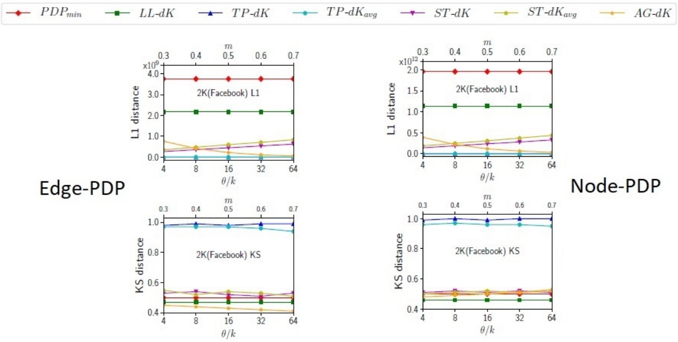


- Our methods yield less network structural error.
- AG-dK outperforms under edge-PDP and LL-dK outperforms under node-PDP by generating more similar 1K-distributions.

Australian National University

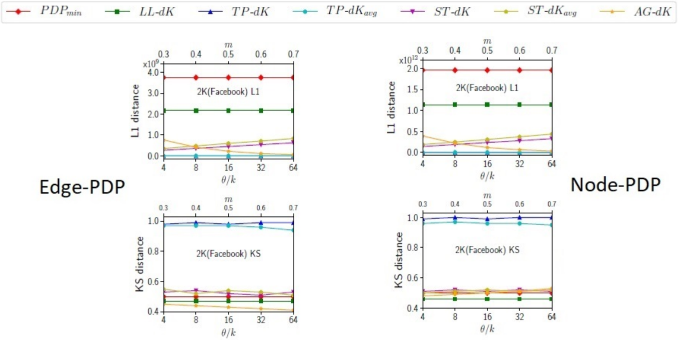
Does the proposed personalized approaches yield yield more utility in 2K-distribution under edge-PDP and node-PDP?

Does the proposed personalized approaches yield yield more utility in 2K-distribution under edge-PDP and node-PDP?



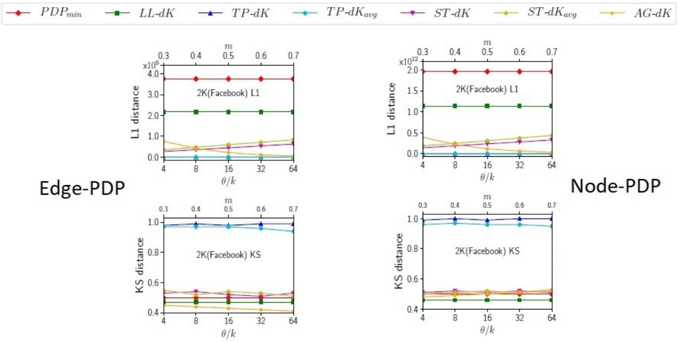
Australian National University

Does the proposed personalized approaches yield yield more utility in 2K-distribution under edge-PDP and node-PDP?



- Our methods yield less network structural error.

Does the proposed personalized approaches yield yield more utility in 2K-distribution under edge-PDP and node-PDP?



- Our methods yield less network structural error.
- AG-dK outperforms under edge-PDP and LL-dK outperforms under node-PDP by generating more similar 2K-distributions.

What kind of trade-off exists between utility and privacy while generating personalized differentially private dK-distributions?

Australian National University

What kind of trade-off exists between utility and privacy while generating personalized differentially private dK-distributions?

- The error caused by sensitivity (Δ) and the privacy preference ε dominates the impact on output utility.

Australian National University

What kind of trade-off exists between utility and privacy while generating personalized differentially private dK-distributions?

- The error caused by sensitivity (Δ) and the privacy preference ε dominates the impact on output utility.

- Increasing ε and decreasing Δ can help to reduce error.

What kind of trade-off exists between utility and privacy while generating personalized differentially private dK-distributions?

- The error caused by sensitivity (Δ) and the privacy preference ε dominates the impact on output utility.
- Increasing ε and decreasing Δ can help to reduce error.
- Reducing sensitivity is more challenging under node-PDP than for edge-PDP as graph data is highly sensitive under node-DP.

Conclusion and Future work

- **Conclusion:**

- **Conclusion:**
  - We have studied the problem of publishing degree distribution and joint degree distribution under PDP.

- **Conclusion:**
  - We have studied the problem of publishing degree distribution and joint degree distribution under PDP.
  - We have theoretically analyzed the sensitivity of these distributions under edge-PDP and node-PDP.

- **Conclusion:**
  - We have studied the problem of publishing degree distribution and joint degree distribution under PDP.
  - We have theoretically analyzed the sensitivity of these distributions under edge-PDP and node-PDP.
  - We have proposed four personalized privacy-preserving mechanisms while enhancing output utility.

Australian National University

- **Conclusion:**
  - We have studied the problem of publishing degree distribution and joint degree distribution under PDP.
  - We have theoretically analyzed the sensitivity of these distributions under edge-PDP and node-PDP.
  - We have proposed four personalized privacy-preserving mechanisms while enhancing output utility.
  - The effectiveness of our proposed work has been empirically verified over four real-world networks.

- **Conclusion:**
  - We have studied the problem of publishing degree distribution and joint degree distribution under PDP.
  - We have theoretically analyzed the sensitivity of these distributions under edge-PDP and node-PDP.
  - We have proposed four personalized privacy-preserving mechanisms while enhancing output utility.
  - The effectiveness of our proposed work has been empirically verified over four real-world networks.
- **Future work:** To this work will consider local differential privacy to release network statistics under personalization.

References

Wei-Yen Day, Ninghui Li, and Min Lyu.
Publishing graph degree distribution with node differential privacy.
In *SIGMOD*, pages 123–138, 2016.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith.
Calibrating noise to sensitivity in private data analysis.
In *TCC*, pages 265–284, 2006.

Priya Mahadevan, Calvin Hubble, Dmitri Krioukov, Bradley Huffaker, and Amin Vahdat.
Orbis: rescaling degree correlations to generate annotated internet topologies.
In *SIGCOMM*, pages 325–336, 2007.

Priya Mahadevan, Dmitri Krioukov, Kevin Fall, and Amin Vahdat.
Systematic topology analysis and generation using degree correlations.
In *SIGCOMM*, pages 135–146, 2006.

Alessandra Sala, Xiaohan Zhao, Christo Wilson, Haitao Zheng, and Ben Y Zhao.
Sharing graphs using differentially private graph models.
In *SIGCOMM*, pages 81–98, 2011.

William E Yancey, William E Winkler, and Robert H Creecy.
Disclosure risk assessment in perturbative microdata protection.
In *Inference control in statistical databases*, pages 135–152. 2002.

Australian National University

Thanks for your attention!

Any **Questions**