Workshop on Privacy Enhancing Technologies for the Homeland Security Enterprise

Privacy-preserving Graph Analytics: Secure Generation and Federated Learning

Dongqi Fu
(UIUC)

Jingrui He
(UIUC)

Hanghang Tong
(UIUC)

Ross Maciejewski
(ASU)

Presenter: Dongqi Fu
Email: dongqif2@illinois.edu

Homeland Security

CAOE | Center for Accelerating Operational Efficiency
A Department of Homeland Security Center of Excellence

ILLINOIS
University of Illinois at Urbana-Champaign

Roadmap

- **Introduction: Privacy-Enhancing Tech (PETs) with Graphs**
- **Privacy-Preserving Graph Data Generation**
- **Federated Learning with Graphs**
- **Summary and Vision**

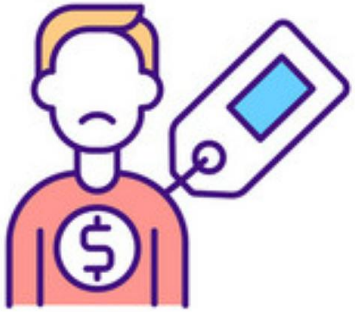
Potential Challenges Threatening HSE

- Homeland Security Enterprise (HSE) is facing unprecedented challenges in multiple critical areas [1]

Sensitive Personal Information

Fabricated Images and Videos

Disinformation



Human Trafficking

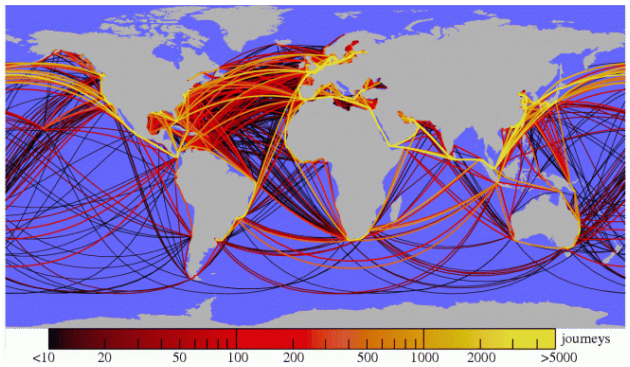
- Privacy-Enhancing Technologies (PETs) in data collection, sharing, and analysis are at the core of many decision-making processes in these areas

[1] Summary of Terrorism Threat to the United States. https://www.dhs.gov/ntas/advisory/national-terrorism-advisory-system-bulletin-june-7-2022

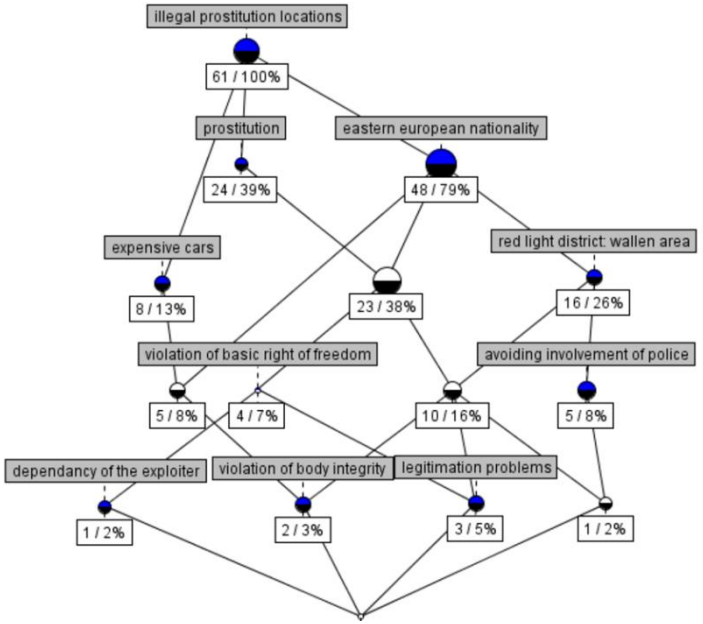
Graphs are Everywhere

- Graph data structure has the capability to represent the complex relationship between entities, in the big data era

Social Network



Cargo-Shipment Network [3]

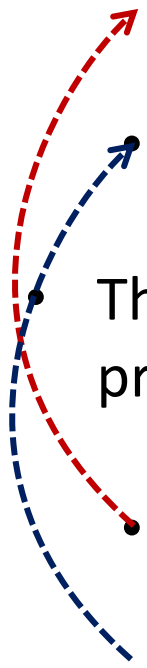


Human-Trafficking Network [2]

[2] https://www.researchgate.net/figure/Concept-lattice-diagram-of-human-trafficking-network_fig3_220271759
[3] https://www.wired.com/2010/01/global-shipping-map/

Privacy-Preserving Analysis of Graphs

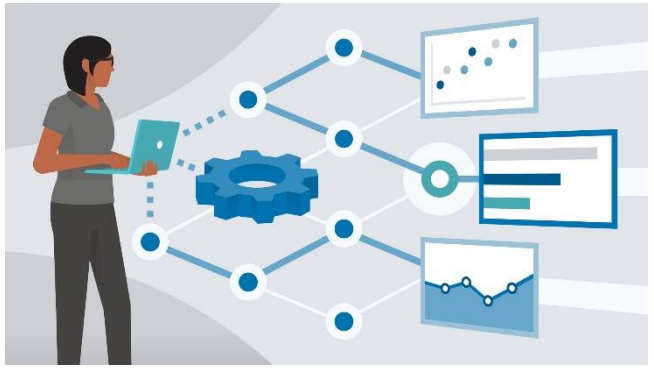
- In this presentation, we focus on **two areas**:
  - Privacy-Preserving Technologies for Linking Identities and Profiles
  - Multiparty Computation with Distributed Private Data Storage
- Then, we introduce the **corresponding AI solutions** w.r.t current progress and future research directions:
  - Privacy-Preserving Graph Generation
  - Federated Learning with Graphs



Roadmap

- Introduction: Privacy-Enhancing Tech (PETs) with Graphs
- **Privacy-Preserving Graph Data Generation**
- **Federated Learning with Graphs**
- **Summary and Vision**

Use Case

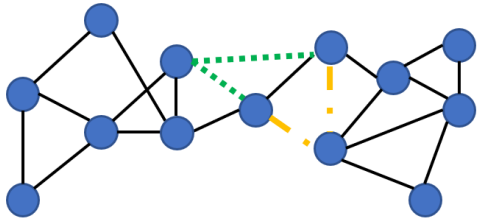


- Description of Need:
  - Privacy Enhanced Information Sharing: CISA@DHS may need to share the entity-interaction data to external scientists [4] or state/local government agencies [5] for
    - the analysis of threat and intelligence gaps
    - or developing new prediction tools
  - Need to create hybrid (real and synthetic) data for privacy
- AI Solution: Privacy-preserving Graph Generation
  - Key idea: Sharing real graphs while hiding sensitive information OR only sharing the realistic generated data

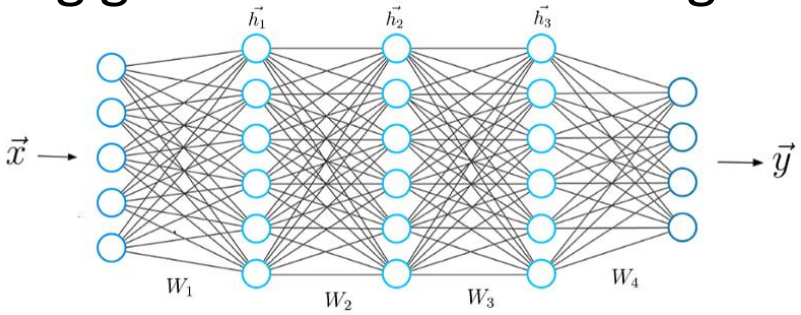
[4] PETS4HSE Use Cases
[5] Law Enforcement Information Sharing Service
https://www.dhs.gov/sites/default/files/publications/privacy-pia-ice-leiss-july2019_0.pdf

Quick Wins

- Current Privacy-Preserving Graph Generation Techs in Two Classes
  - Topology-Guided Structure Perturbation
    - Randomizing the adjacency matrix [6]
    - Injecting the connection uncertainty [7]
    - DP-based permutating connection distribution [8]



  - Deep Learning Generative Models
    - Permutating gradient descent during the training process [9]



[6] Ying et al.: Randomizing Social Networks: a Spectrum Preserving Approach. SDM 2008
[7] Nguyen et al.: Anonymizing Social Graphs via Uncertainty Semantics. CCS 2015
[8] Qin et al.: Generating Synthetic Decentralized Social Graphs with Local Differential Privacy. CCS 2017
[9] Yang et al.: Secure Deep Graph Generation with Link Differential Privacy. IJCAI 2021

Hard Problems

- In the real-world scenario, the networks or graphs are evolving

- Generating privacy-preserving temporal graphs are nascent, and at least three open questions need to be answered

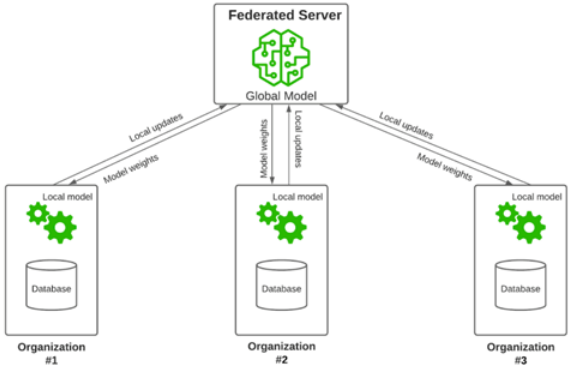
  - Q1. What kind of **time-aware information** is **sensitive** to protect entities' privacy is not clear
  - Q2. When Q1 is determined, the **time-aware protection mechanism** is not yet available
  - Q3. After Q2 is designed, it is challenging to maintain the generation **utility** at the same time with privacy constraints

  - Possible Solution: Dynamic-Preserving Static Graphs + Quick-win

Roadmap

- Introduction: Privacy-Enhancing Tech (PETs) with Graphs
- Privacy-Preserving Graph Data Generation
- **Federated Learning with Graphs**
- **Summary and Vision**

Use Case



- Description of Need:
  - Predict interest for criminal law enforcement [5]: Data from state/local government agencies to CISA@DHS
    - Transmit model parameters instead of real data for privacy
  - Need a decentralized computational framework – Federated Learning (FL) [10]
- AI Solution: Federated Learning on Graphs [11]
  - Key Idea: Robust Learning Models on Graphs, i.e., avoiding the **vulnerability of decentralization**

[10] Kairouz et al.: Advances and Open Problems in Federated Learning. Found. Trends Mach. Learn. 2021
[11] He et al.: FedGraphNN: A Federated Learning System and Benchmark for Graph Neural Networks. CoRR 2021
[12] Figure source: https://sparkd.ai/federated-learning

Quick Wins

- Replacing the gradients averaging with robust estimation of the center, Byzantine-robust method [13]
- But the above adaption is not clear for the non-IID data, two directions of quick wins
  - Observing the performance of existing non-IID Byzantine-robust methods on graphs, to study the impact of IID violation
  - Developing robust federated learning methods for graph data

[13] Blanchard et al.: Machine Learning with Adversaries: Byzantine Tolerant Gradient Descent. NIPS 2017

Hard Problems

- Recent studies [14] show robust FL needs clean and **non-sensitive data on the server** for knowledge distillation
- Recently-proposed data-free distillation method [15] **ignores** the **graph data's non-Euclidean property**
- Both questions are challenging to address nowadays
- Possible Solution: Mapping graphs into grid-like data, and then using the off-the-shelf non-graph methods

[14] Lin et al.: Ensemble Distillation for Robust Model Fusion in Federated Learning. NeurIPS 2020
[15] Zhu et al.: Data-Free Knowledge Distillation for Heterogeneous Federated Learning. ICML 2021

Roadmap

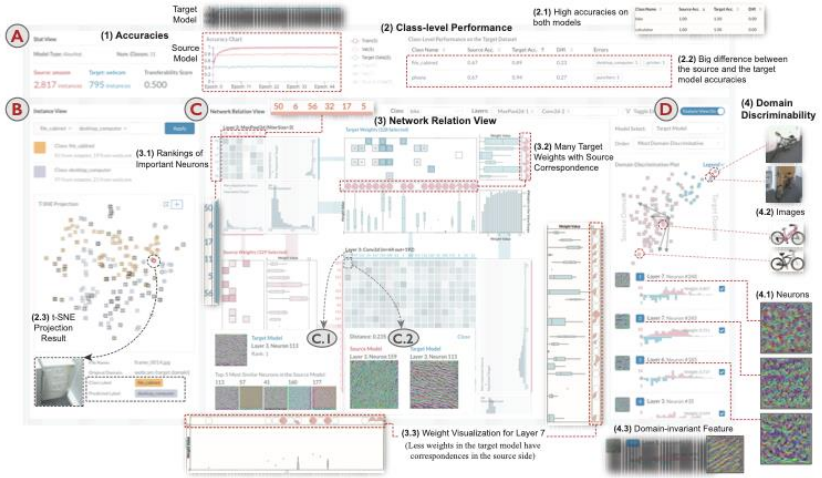
- Introduction: Privacy-Enhancing Tech (PETs) with Graphs
- Privacy-Preserving Graph Data Generation
- Federated Learning with Graphs
- **Summary and Vision**

Promising Directions

- Providing next-generation privacy-enhancing technologies (PETs) for
  - Privacy-Preserving Complex Graph Generation
  - Robust Federated Learning with Graphs

- Aid from visual analytics for providing explanation and interpretation facing various audience

- Ultimately, a unified and integrated system to address critical security problems



A Visual Analytics Interface [16]

[16] Ma et al.: A Visual Analytics Framework for Explaining and Diagnosing Transfer Learning Processes. IEEE Trans. Vis. Comput. Graph. 2021

Workshop on Privacy Enhancing Technologies for the Homeland Security Enterprise

Thanks, and Q&A !

Dongqi Fu
(UIUC)

Jingrui He
(UIUC)

Hanghang Tong
(UIUC)

Ross Maciejewski
(ASU)

Presenter: Dongqi Fu
Email: dongqif2@illinois.edu

Homeland Security

CAOE | CENTER FOR ACCELERATING OPERATIONAL EFFICIENCY
A DEPARTMENT OF HOMELAND SECURITY CENTER OF EXCELLENCE

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN