# From provable security to practical failure

## What cryptography enables - and what it cannot prevent - in distributed graph learning

Aubin Birba - aubin.birba@gmail.com

11/02/2026 - Cybersecurity research presentation @ Hitachi Energy Canada

# whoami

## Aubin Birba

- Security researcher, finishing a PhD at Université du Québec À Montreal (UQÀM)
- Masters in Cyberdefense and Information Security from Université Polytechnique Hauts-de-France
- Interested in secure computation, privacy-preserving and adversarial machine learning
- https://birbaubin.github.io/

# Outline

- Crypto'Graph
  - Context : the modern data sharing landscape
  - Secure distributed graph analytics
  - State-of-the art
  - Our approach
  - Results

# Outline

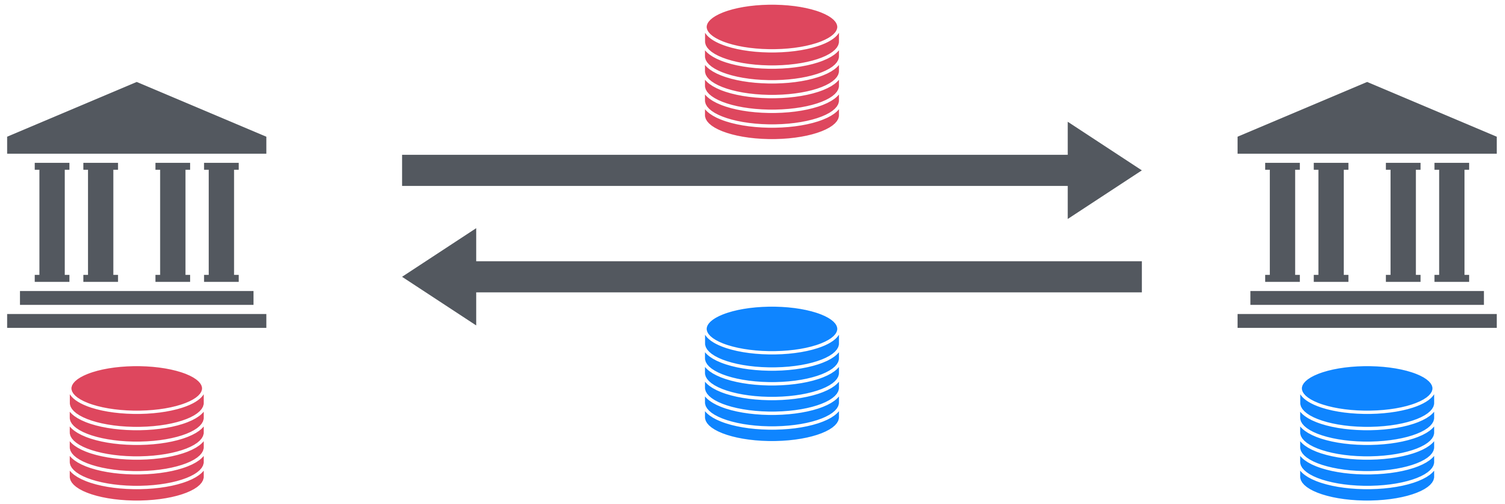
- GRAND
  - The limitations of cryptography
  - Reconstruction in the topological domain
  - Reconstruction in the spectral domain
  - Complete pipeline
  - Results
  - Cross-cutting insights
- Conclusion : relevance to power systems

# Crypto'Graph : Leveraging Privacy-Preserving Link Prediction for Robust Graph Learning

The 14th ACM Conference on Data and Application Security and Privacy, 2024

# Context

## The modern data landscape



Data sharing is desirable for research and industrial advancement...

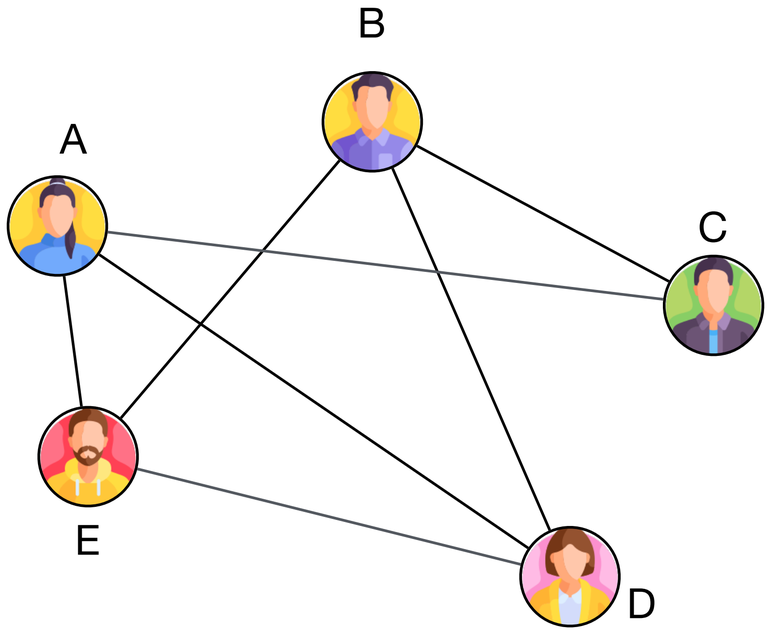
# Context

## The modern data landscape

But it comes at a cost for confidentiality and privacy (intellectual property privacy regulation eg. GDPR)

# Graph analytics

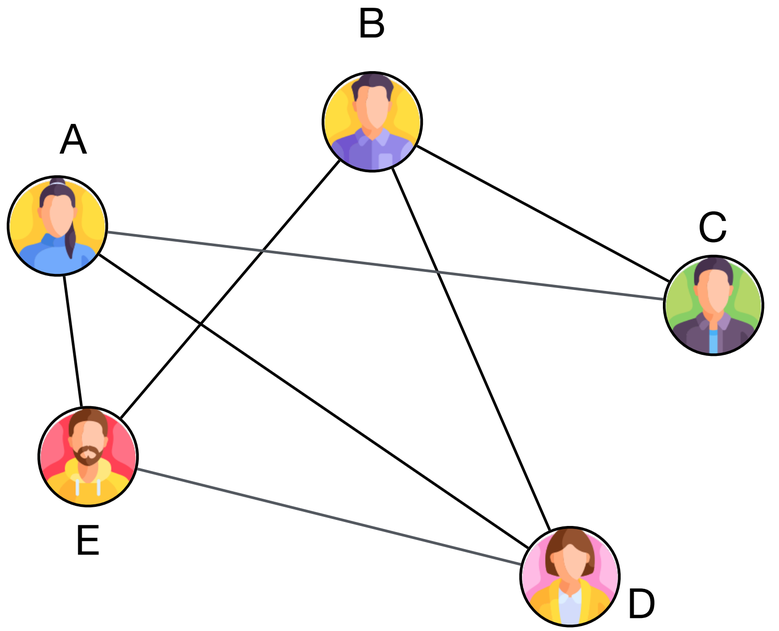
- Graphs are an important data structure that models relationships between entities (sensor networks, power networks)
- Graph analytics and Graph Machine Learning allow deep insights and exploitation of graphs for various purposes (topology growth forecasting, signal processing, etc)



# Graph analytics

## Link prediction

- Important graph analytics task that aims at identifying potential future connexions in the network  (connection forecasting)
- A widespread approach to link prediction : similarity measures (similar nodes tend to form connections in many types of networks)
- Potential applications : network growth forecasting (*Zhang et al*), data curation (*Wu et al.*)



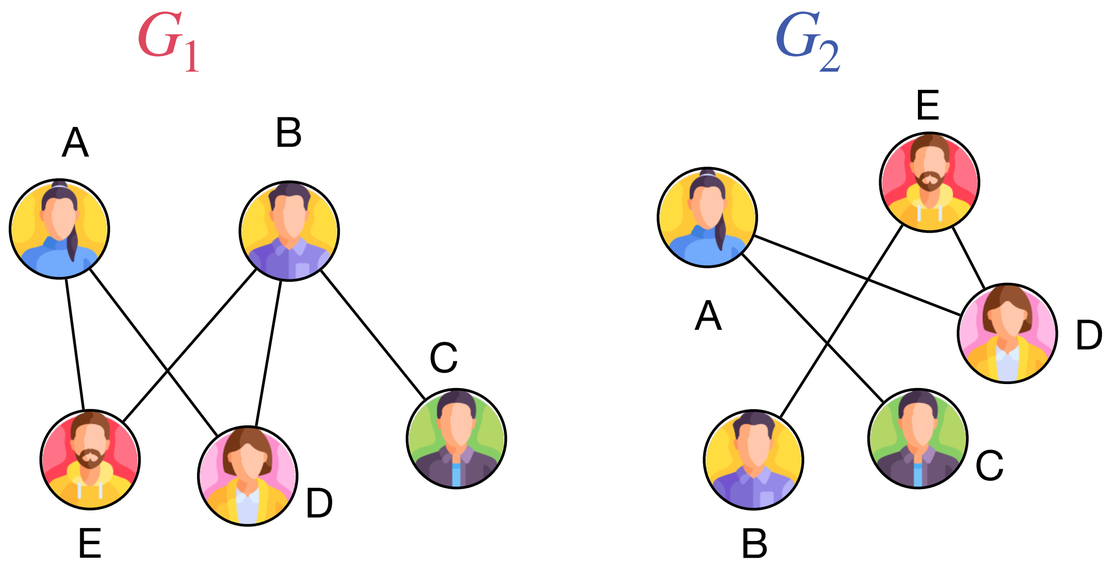



Hai-Feng Zhang, Xiao-Jing Ma, Jing Wang, Xingyi Zhang, Donghui Pan, and Kai Zhong. Privacy-preserving link prediction in multiple private networks. IEEE Transactions on Computational Social Systems
Huijun Wu, Chen Wang, Yuriy Tyshetskiy, Andrew Docherty, Kai Lu, Liming Zhu. Adversarial Examples for Graph Data: Deep Insights into Attack and Defense. Twenty-Eighth International Joint Conference on Artificial Intelligence. 2019

# Secure distributed graph analytics

## Requirements and constraints

- **System model** : Parties who hold graphs collaborate to perform link prediction on their joint network
- **Threat model** : each party tries to break the confidentiality of the other parties network edges while respecting the protocol (honest-but-curious parties)



Can we efficiently and privately,
predict links on a distributed graph ?

# State of the art

| | **Accuracy** | **Speed** | **Cryptographic primitive** | **Similarity measures** |
|---|---|---|---|---|
| Zhang et al. | Approximate | ~1000s | Secret sharing | Common neighbours, Jaccard, Cosine, Adamic Adar |
| Demiral et al. | Exact | ~15000s | PSI | Common neighbours |
| Crypto'Graph | Exact | ~ 200s | PSI | Common neighbours, Jaccard, Cosine |

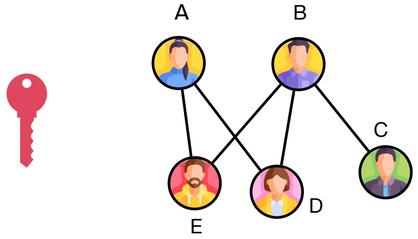
Didem Demirag, Mina Namazi, Erman Ayday, and Jeremy Clark. Privacy preserving link prediction.. Data Privacy Management, Cryptocurrencies and Blockchain Technology. 2023
Hai-Feng Zhang, Xiao-Jing Ma, Jing Wang, Xingyi Zhang, Donghui Pan, and Kai Zhong. Privacy-preserving link prediction in multiple private networks. IEEE Transactions on Computational Social Systems. 2023
Azogagh, Sofiane, Zelma Aubin Birba, Sébastien Gambs, and Marc-Olivier Killijian. "Crypto'Graph: Leveraging Privacy-Preserving Distributed Link Prediction for Robust Graph Learning." In Proceedings of the Fourteenth ACM Conference on Data and Application Security and Privacy,. 2024.

# Preliminaries

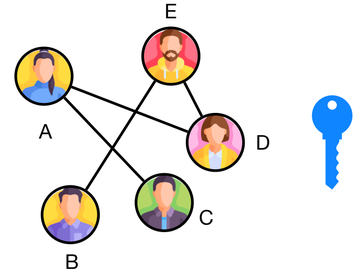
- $G$ the undirected, unweighted distributed graph, $V$ the set of vertices, $E$ the set of edges
- $G_1, G_2$ the two partitions of the graph such that $V$ belongs to both, and $E_1 \cup E_2 = E$
- $\Gamma(x)$ the neighbours of vertex $x$ on graph $G$
- $\Gamma_i(x)$ the neighbours of vertex $x$ on graph $G_i$

# CryptoʼGraph

## Secure distributed link prediction




$\Gamma_1(A) = \{E, D\}$ $\Gamma_1(B) = \{E, D, C\}$




$\Gamma_1(A) = \{E, D\}$ $\Gamma_1(B) = \{E, D, C\}$

$\Gamma_2(A) = \{C, D\}$ $\Gamma_2(B) = \{E\}$

$\Gamma_1(A) = \{E, D\}$ $\Gamma_1(B) = \{E, D, C\}$ $\Gamma_2(A) = \{C, D\}$ $\Gamma_2(B) = \{E\}$

$\Gamma_2(A) = \{C, D\}$ $\Gamma_2(B) = \{E\}$

$\Gamma(A) = \Gamma_1(A) \cup \Gamma_2(A) = \{E, C, D\}$

$\Gamma(B) = \Gamma_1(B) \cup \Gamma_2(B) = \{E, C, D\}$

$\Gamma(A) \cap \Gamma(B) = \{E, C, D\}$

# Crypto'Graph

## The encryption method — Diffie-Hellman-like shared secret

Let $\mathbb{G}$ be a cyclic group over $\mathbb{Z}_p$ ($p$ a big prime number), and $g$ a generator of this group. $\lambda$ the security parameter

- Key generation : $\alpha, \beta \leftarrow \text{KeyGen}(1^\lambda)$
- Encryption : $\text{Enc}_k(x) = g^{kx}, k \in \{\alpha, \beta\}$
- Double encryption : $\text{Enc}_{k_1,k_2}(x) = (g^{k_1 x})^{k_2} = g^{k_1 k_2 x}$
- Discrete logarithm assumption : for $p$ big enough, given $(g, g^e)$, it is hard to get $e$

# Crypto'Graph

## Results

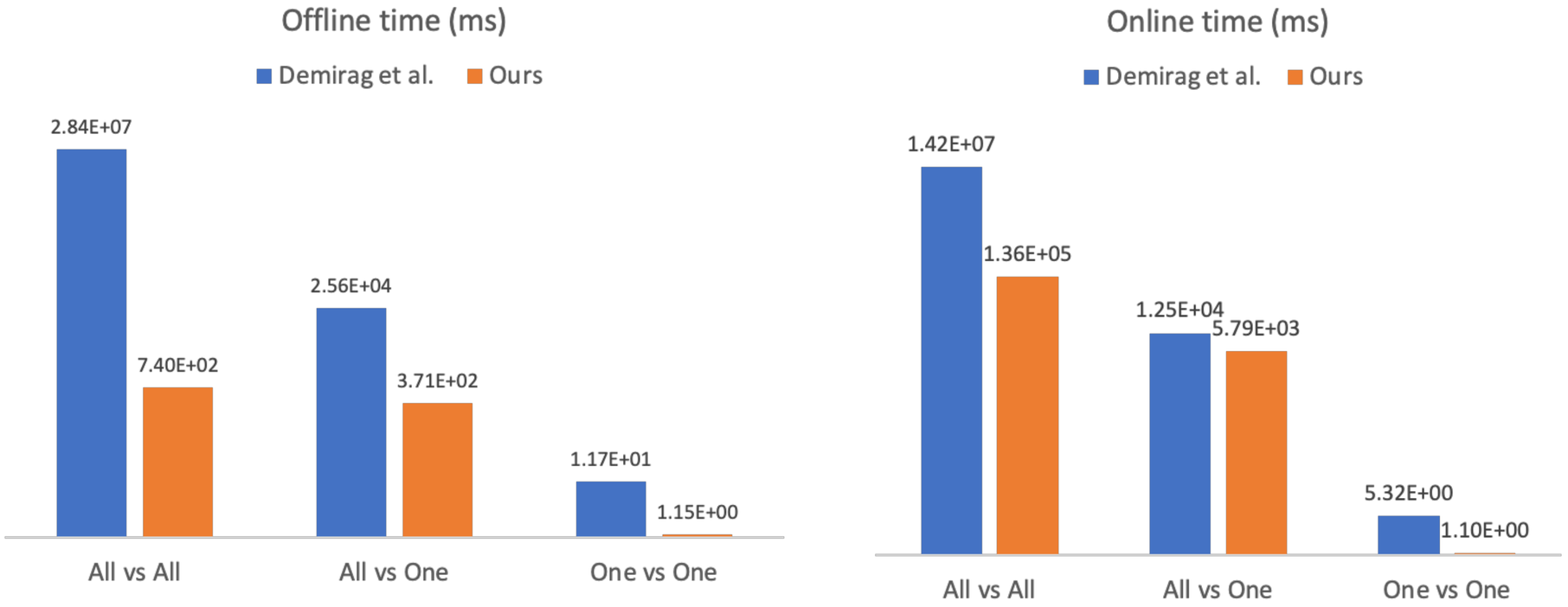
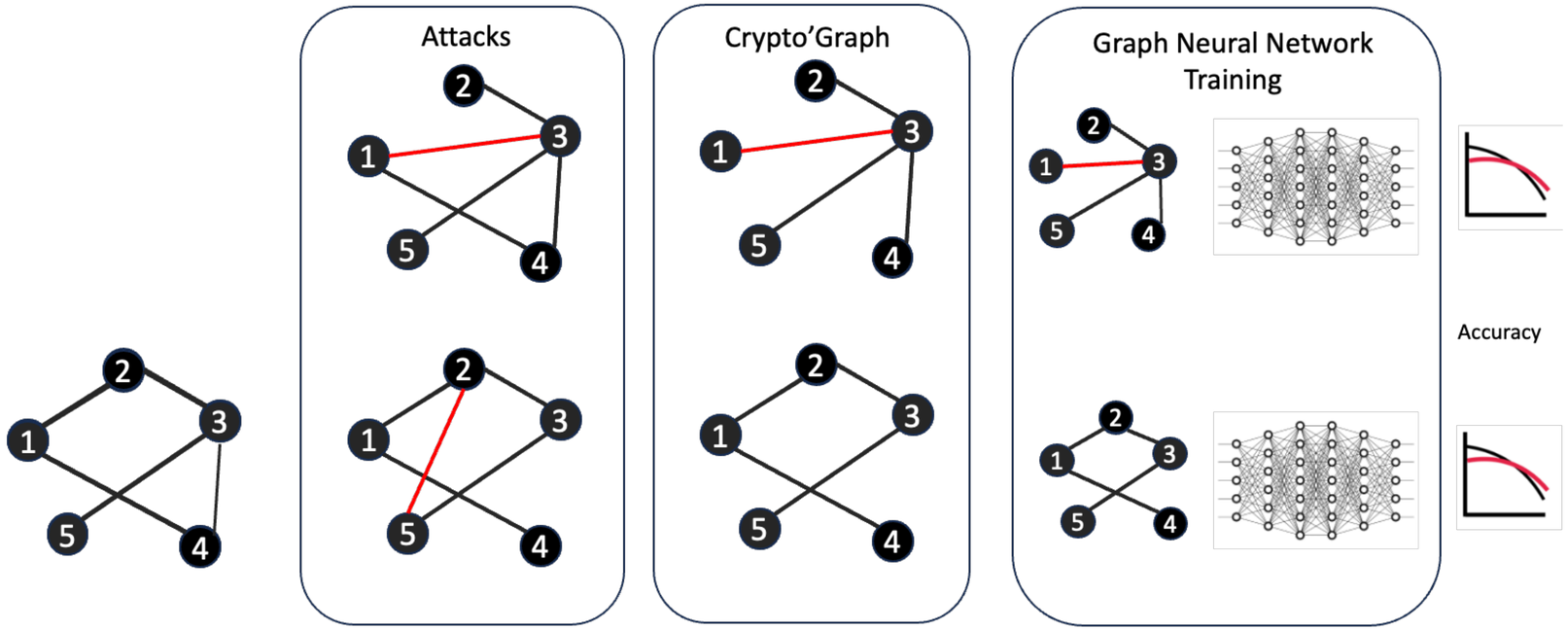



**Figure 1 : Runtime off Crypto'Graph, compared to Demirag et al**

- Link prediction for all node pairs (all vs all), between one node and all others (all vs one), and between two nodes (one vs one)
- We improve time complexity by orders of magnitude

# Crypto'Graph

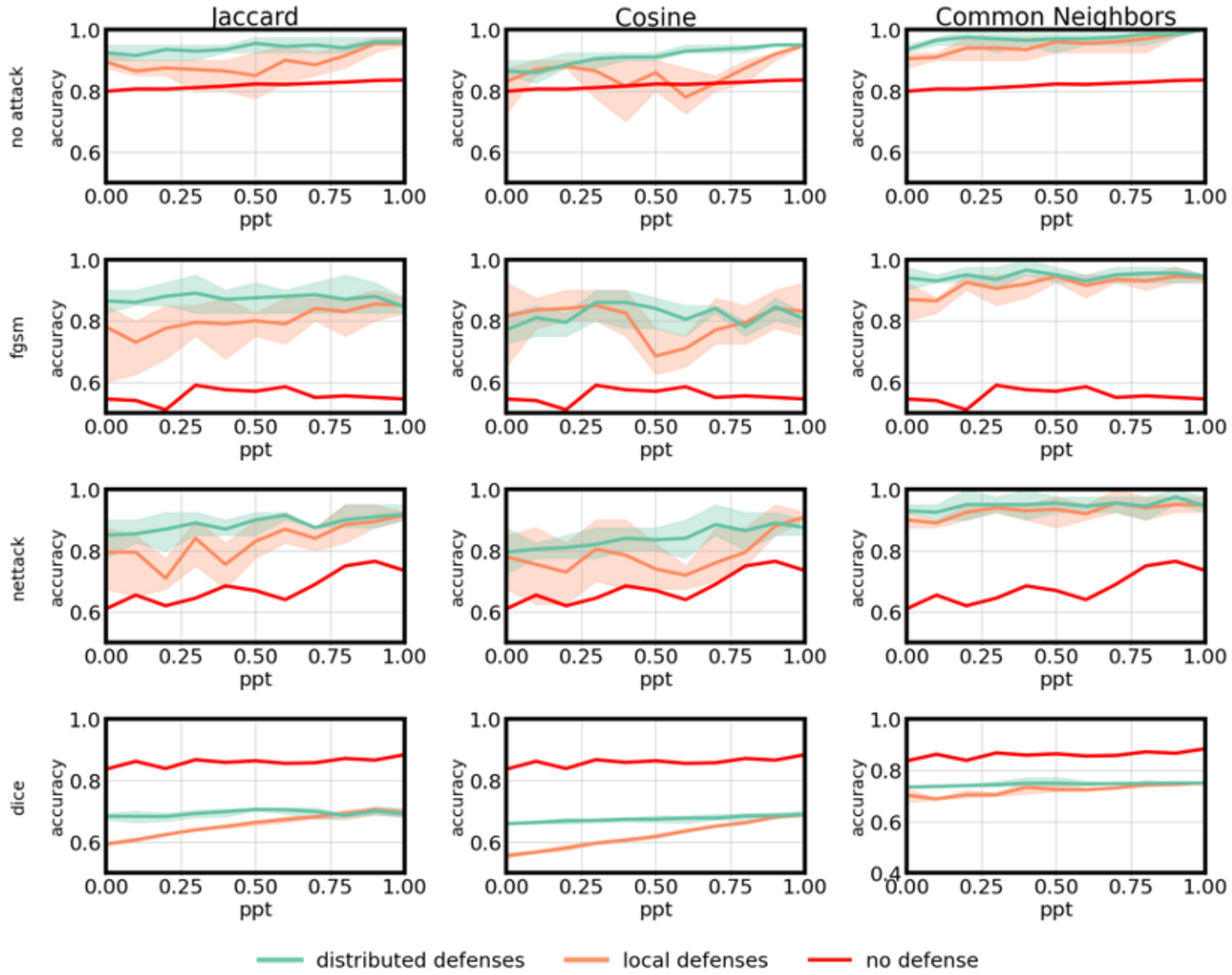
## Utility experiment pipeline




- Crypto'Graph is used as a private preprocessing mechanism
- Links between nodes that have low similarity are considered erroneous and dropped
- Graph neural networks are trained on the resulting graphs to classify nodes
- Crypto'Graph based defense allows better accuracy than local and no defense

# Crypto'Graph

## Results




- Crypto'Graph is used as a private preprocessing mechanism
- Links between nodes that have low similarity are considered erroneous and dropped
- Graph neural networks are trained on the resulting graphs
- Crypto'Graph based defense allows better accuracy than local and no defense

# Crypto'Graph

## What cryptography enables here

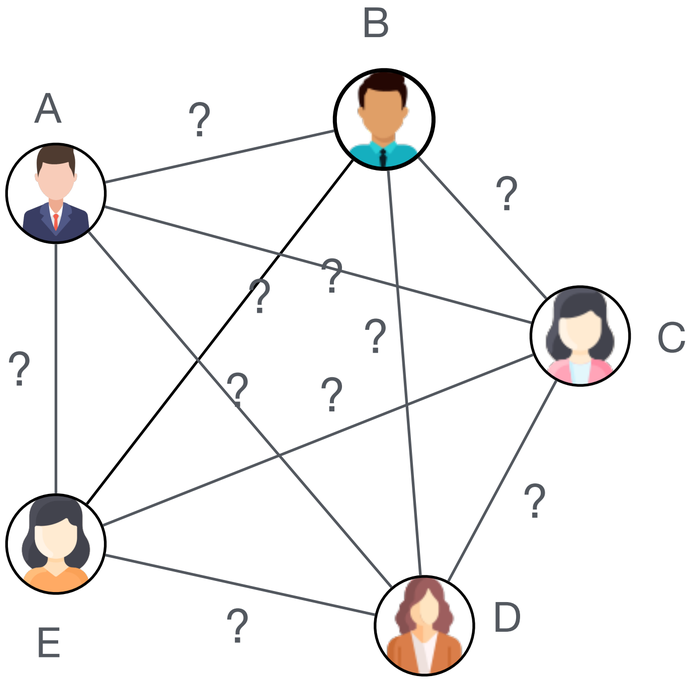
- We can efficiently compute the number of common neighbours on a distributed graph
- The common neighbours metric serves as a partial result for the computation of other similarity metrics (Jaccard Index, cosine similarity, preferential attachment index, etc)
- The structure of the graph is preserved by the Diffie-Hellmann like encryption
- **Proof shows that the protocol is secure as long as the Discrete Logarithm Assumption holds**

# But what if…

| | A | B | C | D | E |
|---|---|---|---|---|---|
| A | 2 | 2 | 0 | 0 | 0 |
| B | 2 | 3 | 0 | 0 | 0 |
| C | 0 | 0 | 1 | 1 | 1 |
| D | 0 | 0 | 1 | 2 | 2 |
| E | 0 | 0 | 1 | 2 | 2 |

$G^2$

Common neighbours matrix




$G$

- RQ1 : Can an attacker guess the underlying graph $G$ that produced $G^2$ ?
- RQ2 : To what extent knowing some existing or non existing edges in $G$ helps the attacker ?

# GRAND : Graph Reconstruction from potential partial Adjacency and Neighborhood Data

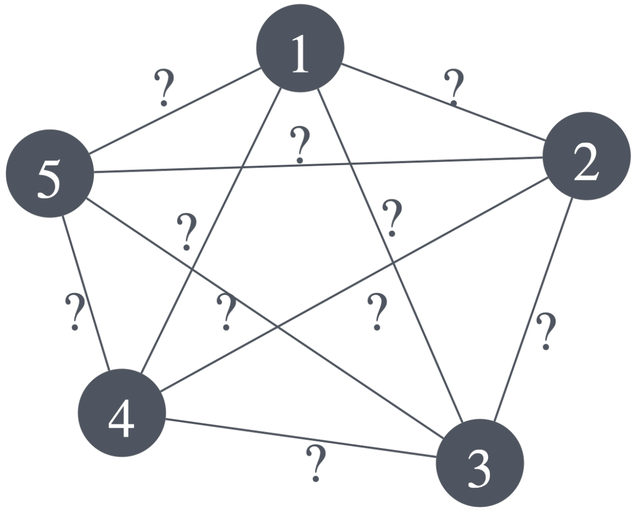
31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining
August 3 - 7, 2025 - Toronto

# Reconstruction in the topological domain

## Initialization

| | | | | |
|---|---|---|---|---|
| 3 | 1 | 2 | 1 | 1 |
| 1 | 2 | 1 | 2 | 1 |
| 2 | 1 | 4 | 1 | 0 |
| 1 | 2 | 1 | 2 | 1 |
| 1 | 1 | 0 | 1 | 1 |

$G^2$




$G^\star$

**Topological attacks:**

- Degree Combination
- Degree Matching
- Neighbour Matching
- Degree Completion
- Neighbour Completion
- Triangles
- BiCliques

# Reconstruction in the spectral domain

## Intuition

If $G$ is the adjacency matrix of an undirected graph :

$G = U\Sigma V$ (Singular Value Decomposition)

$G^2 = U\Lambda V = (U\Sigma V)(U\Sigma V) = U\Sigma^2 V$

- $\forall \sigma_i, \exists \lambda_j$ such that $\sigma_i^2 = \lambda_j$
  - Is $\sigma_i$ positive or negative ?

| | | | |
|---|---|---|---|
| 9 | 0 | 0 | 0 |
| 0 | 4 | 0 | 0 |
| 0 | 0 | 1 | 0 |
| 0 | 0 | 0 | 1 |

$\Lambda$

| | | | |
|---|---|---|---|
| 3 | 0 | 0 | 0 |
| 0 | 2 | 0 | 0 |
| 0 | 0 | 1 | 0 |
| 0 | 0 | 0 | 1 |

$\Sigma = diag(3,2,1,1)$

...

| | | | |
|---|---|---|---|
| 2 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 |
| 0 | 0 | 1 | 0 |
| 0 | 0 | 0 | -3 |

$\Sigma = diag(2,1,1,-3)$

# Reconstruction in the spectral domain

## Solution

A greedy optimization algorithm :

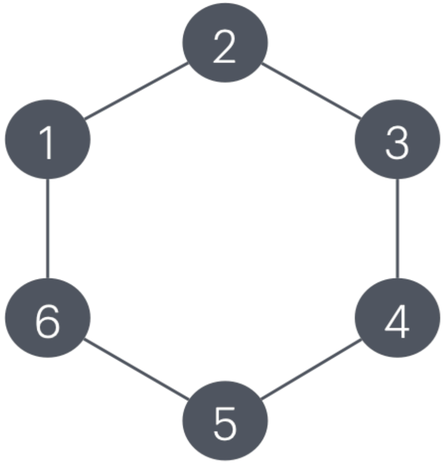
- For each $\sigma_i$, reconstruct the graph using $\sigma_i = \sqrt{\lambda_i}$ and $\sigma_i = -\sqrt{\lambda_i}$
- $\sigma_i$ gets assigned the value that produces a graph that is closest to a binary graph and to the known adjacency information

$$min_{\sigma \in \{\sqrt{\lambda_i}, \sqrt{\lambda_i}\}} \alpha \left\| G_\sigma^\star - BG_\sigma^\star \right\|_F + \beta \left\| G_\sigma^\star - G_t^\star \right\|_F$$

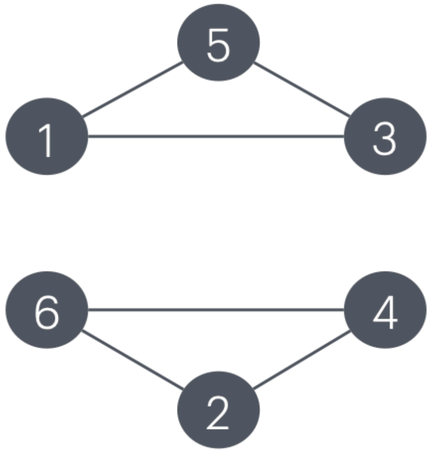
# GRAND

## Cosquare graphs

- Some graphs have the same common neighbour matrix
- We introduce the notion of cosquare graphs to denote non isomorphic graphs that do not have the same common neighbour matrix




$G$



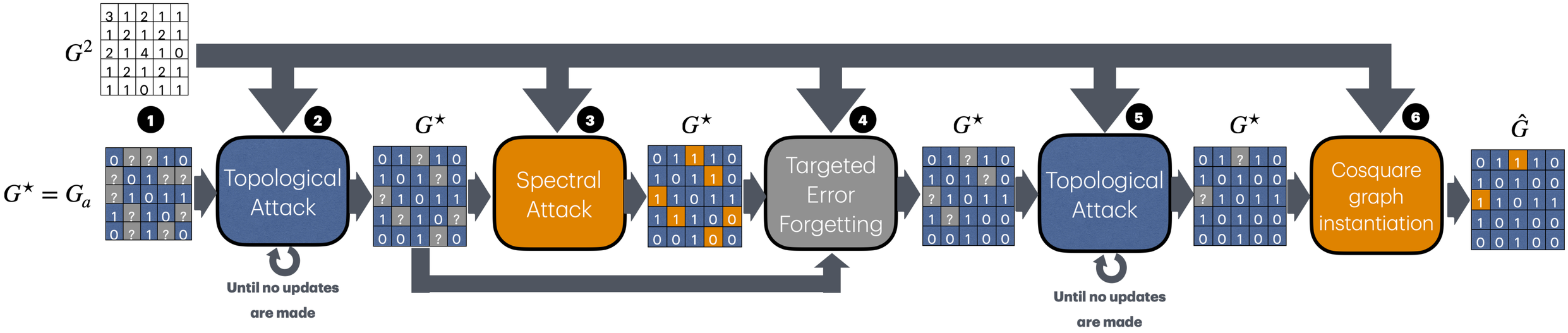

$H$

$$G^2 = H^2 =$$

| | | | | | |
|---|---|---|---|---|---|
| 2 | 0 | 1 | 0 | 1 | 0 |
| 0 | 2 | 0 | 1 | 0 | 1 |
| 1 | 0 | 2 | 0 | 1 | 0 |
| 0 | 1 | 0 | 2 | 0 | 1 |
| 1 | 0 | 1 | 0 | 2 | 0 |
| 0 | 1 | 0 | 1 | 0 | 2 |

# GRAND

## Pipeline



# GRAND

## Results

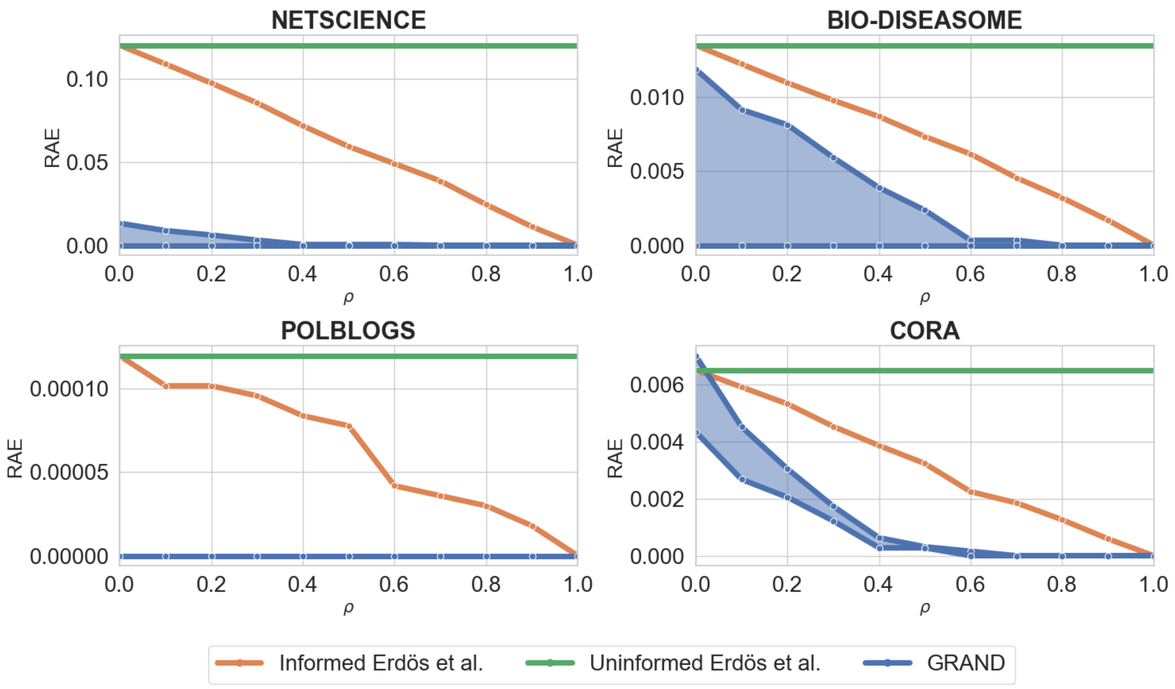



**Fig 1. Relative Absolute Error for various datasets and prior knowledge**

- Prior knowledge is helpful to reconstruct the graph. The more prior knowledge we have, the more we can reconstruct.
- Perfect reconstruction is sometimes possible, even with no prior knowledge

# Cross cutting-insights

- Secure computation is great at what it does : computing on data without directly revealing it
- When the result of the computation is to be disclosed, even if it seems to not reveal the input, special care needs to be taken
- When secure computation needs to be done on the mix of private data from different stakeholders, what each knows about the other matters a lot

# Relevance to power systems

- Machine learning models are increasingly applied to the grid (demand forecasting, expansion forecasting, intrusion detection, etc)
- These models create a new attack surface
  - collaboration between semi-trusted grid operators
  - outsourcing of training in the cloud
  - sensitive data reconstruction from model outputs

# Relevance to power systems

## The GridFM project (https://gridfm.org/)

- Gathers actors of academia and industry to train Foundation Models on grid data for various use cases (demand forecasting, failure prediction, security, etc)
- Adoption of Advanced Privacy-Preserving Federated Learning framework (APPFL) for secure federated learning (Differential privacy, secure aggregation)
    - To what extent the current solutions are vulnerable to reconstruction attacks from the aggregated results ?
    - Can we build efficient privacy-preservation mechanisms for collaborative grid learning ?

gridfm.org
AI FOR THE ELECTRIC GRID

# Thank you !

Happy to take your questions

Link to this presentation:
https://birbaubin.github.io/assets/pdf/hitachi.pdf

Link to Crypto'Graph paper
https://dl.acm.org/doi/abs/10.1145/3626232.3653257

Link to GRAND paper
https://dl.acm.org/doi/10.1145/3711896.3736988