Privacy and Transparency in Graph Machine Learning

A Unified Perspective

Megha Khosla

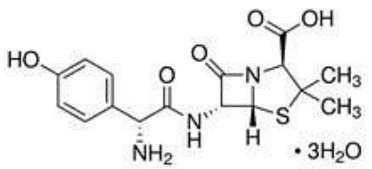
https://khosla.github.io

m.khosla@tudelft.nl

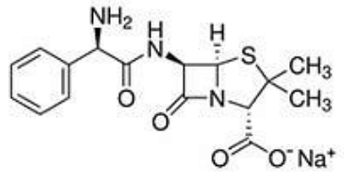
AIMLAI@CIKM2022

TU Delft

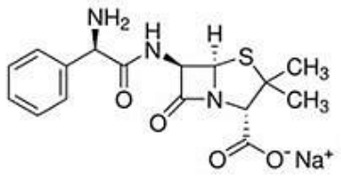
Success (Potential) of Graph Machine Learning



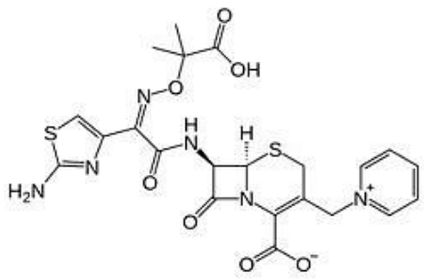
Amoxicillin



Ampicillin



Penicillin G

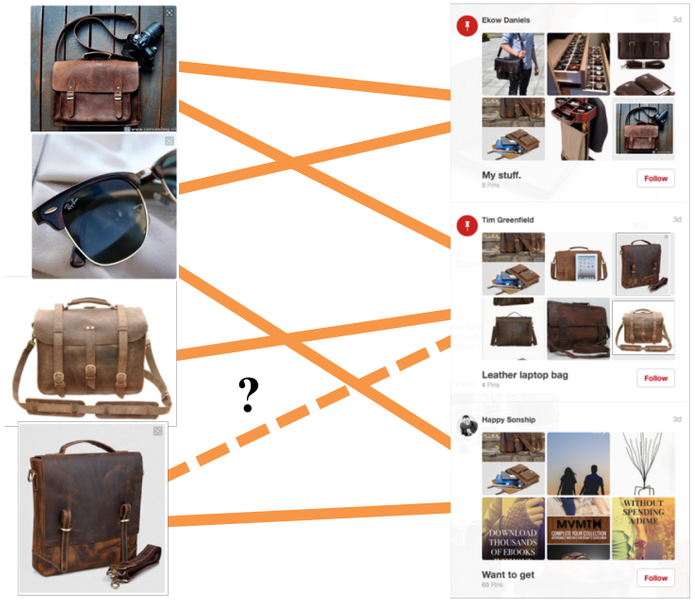


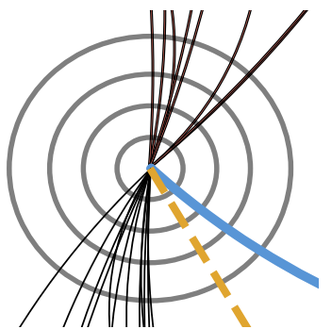
Ceftazidime

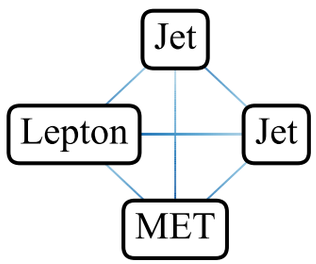
discover **novel antibiotics** (Stokes *et al.*, Cell'20)

Image Source : Coman et al. 2017

power **web-scale recommender systems** (Ying *et al.*, KDD'18; Pal *et al.*, KDD'20)

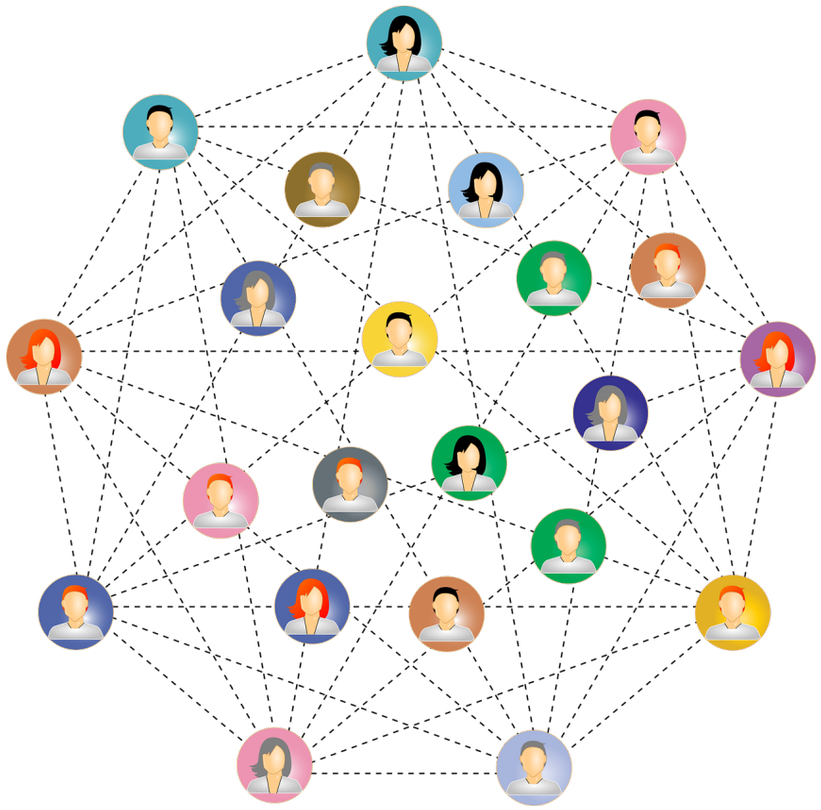






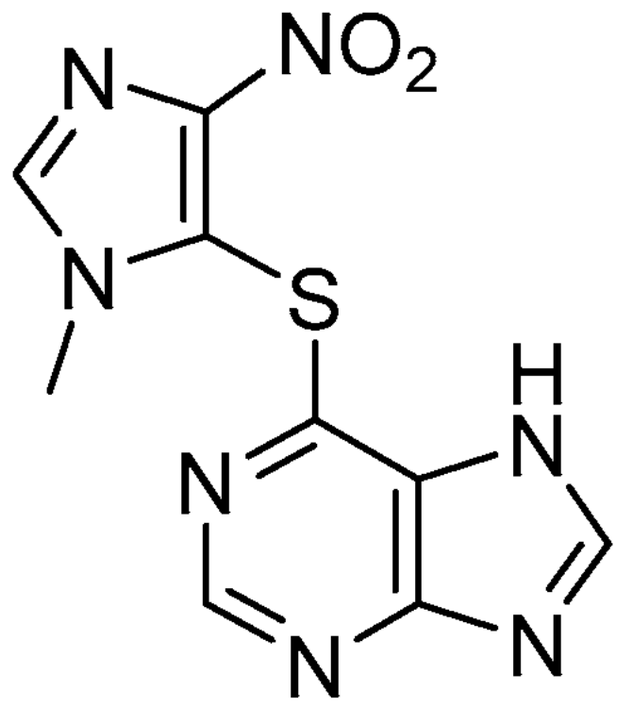
assist **particle physicists** (Shlomi *et al.*, Mach. Learn.: Sci. Technol'21)

Graphs are everywhere



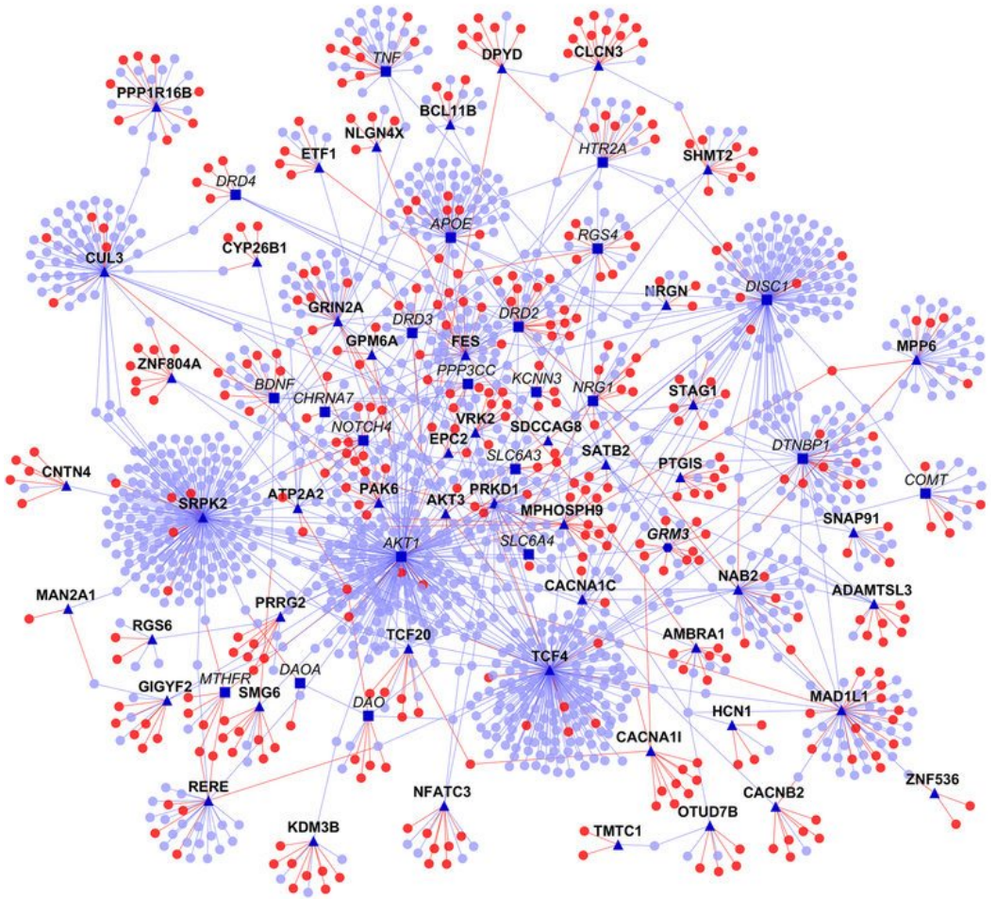
Social Networks

Image Source : Medium



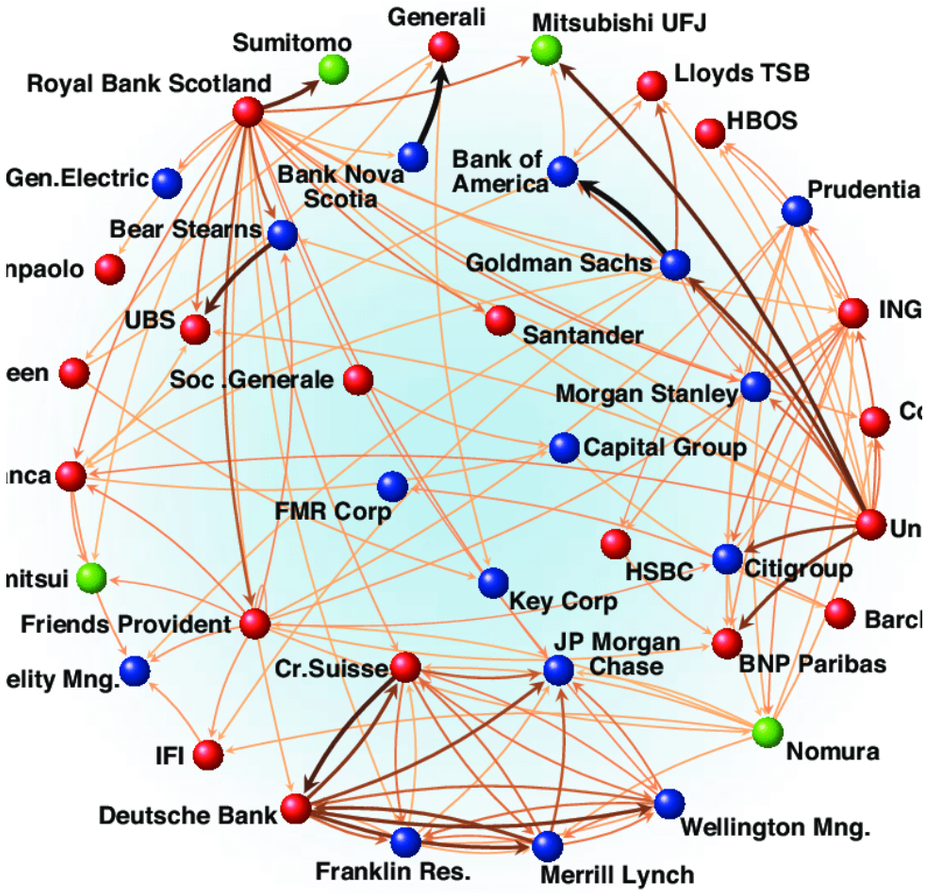
Drug Molecules

Graphs are everywhere



Protein interaction network

Image Source : wikipedia



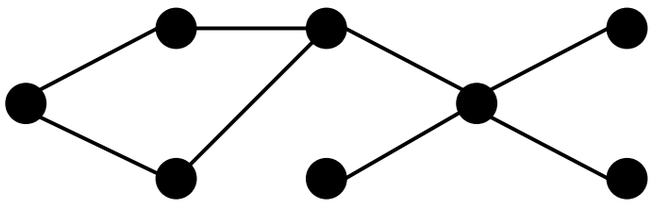
Financial network

Image Source : Schweitzer et al. 2009

Different kinds of graphs

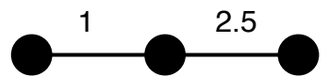
Some of the graph types

Undirected, unweighted, static

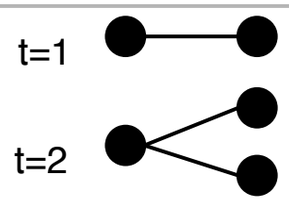


Directed

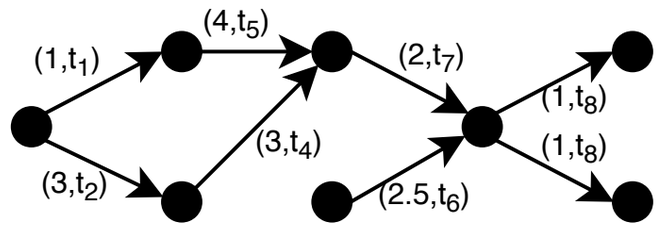
Weighted



Sliced



Directed, weighted, continuous



Feature types:

No features

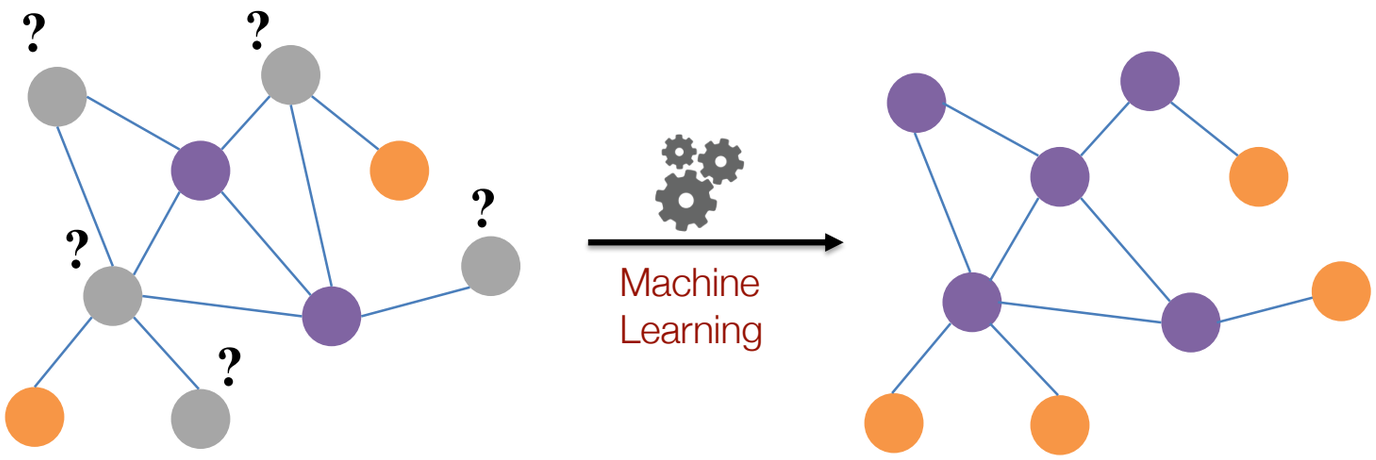
Node features

Edge features

Dense features, e.g. word embeddings

Sparse features, e.g. one-hot encodings

Typical ML Tasks on Graphs



Node classification

Link prediction

Graph classification

Community detection

Graph Machine Learning (GraphML)

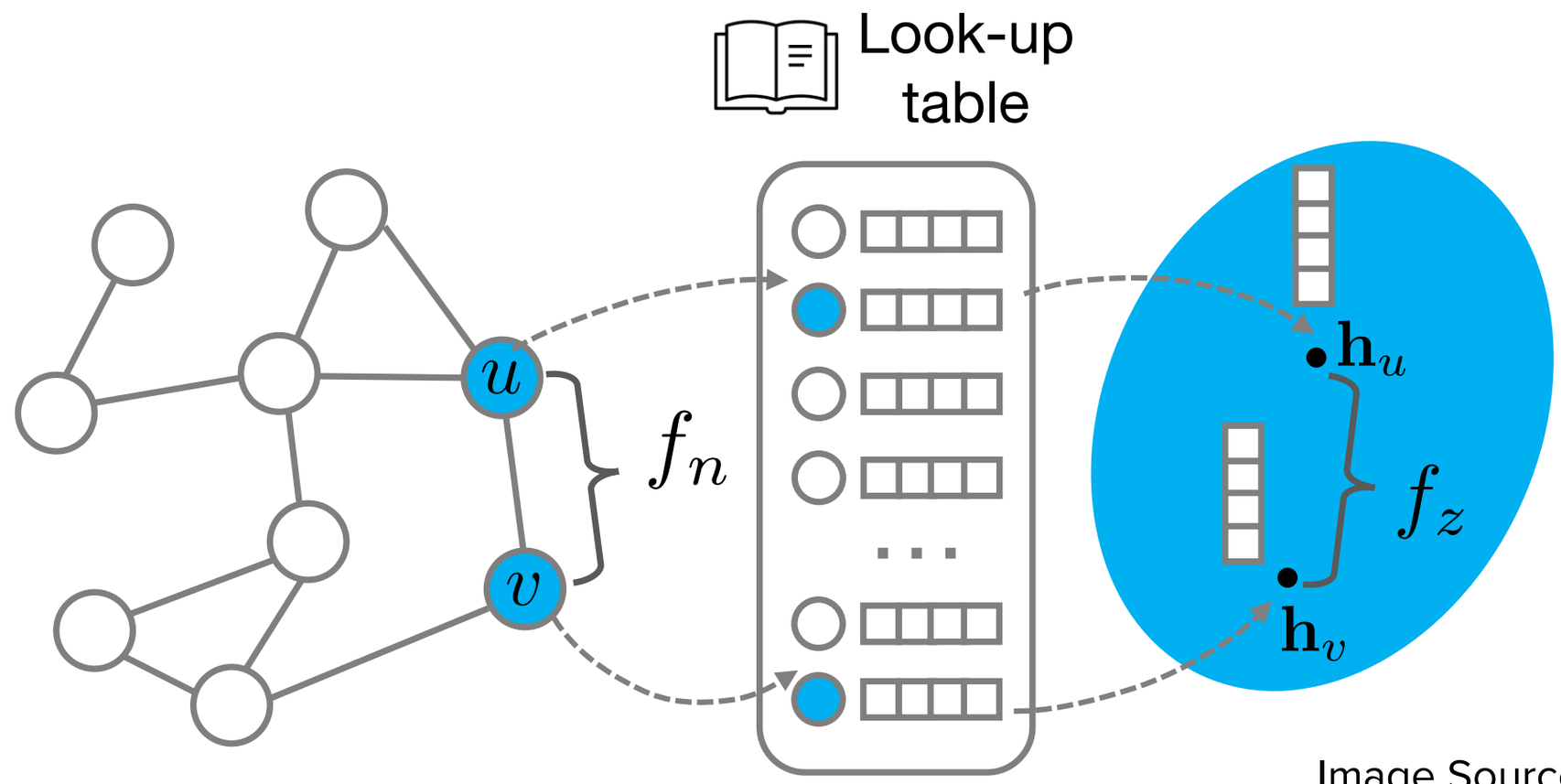
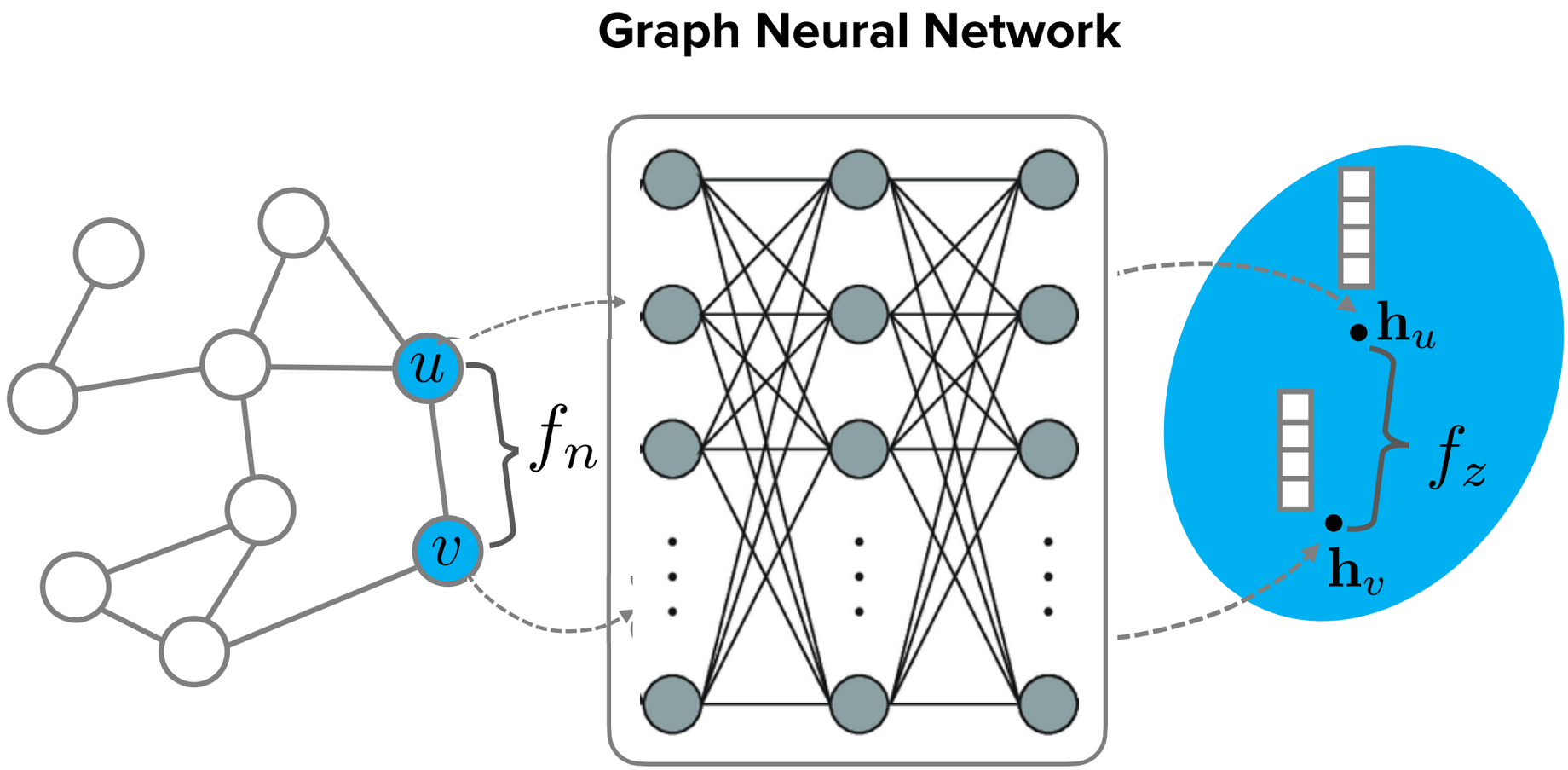


Image Source: [Li et al., 2022]

Shallow Network Embedding Methods

Examples : **DeepWalk, Node2Vec, NERD, HOPE**

Graph Machine Learning

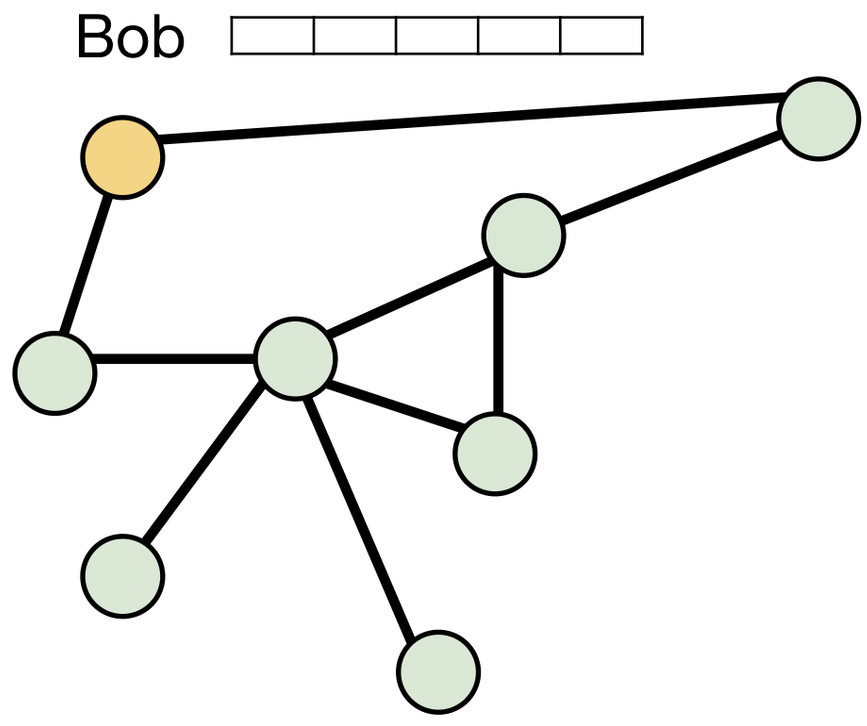


Examples :

GCN, GAT, GIN

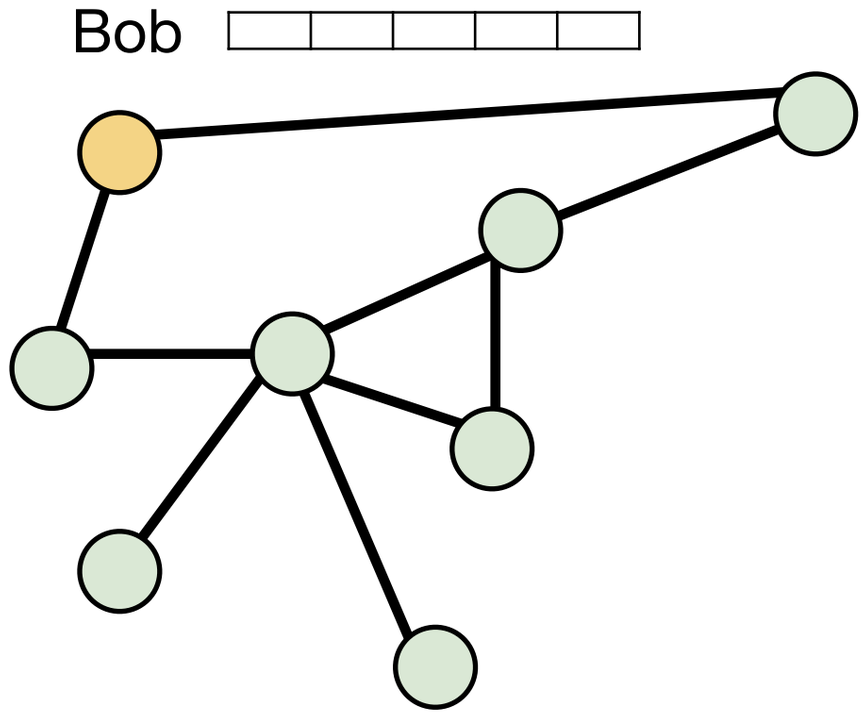
Transparency

Why was Bob's loan denied?

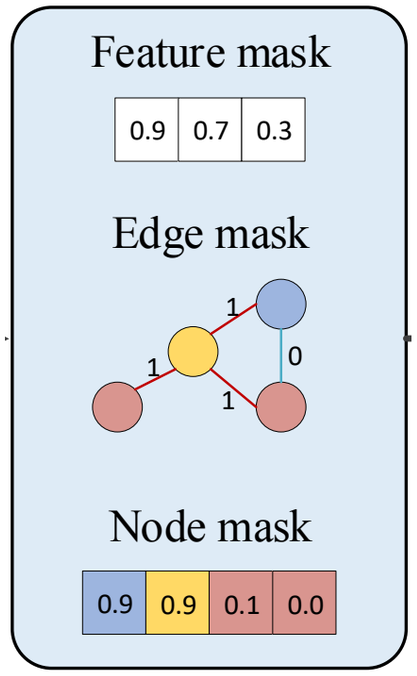


Transparency

Why was Bob's loan denied?

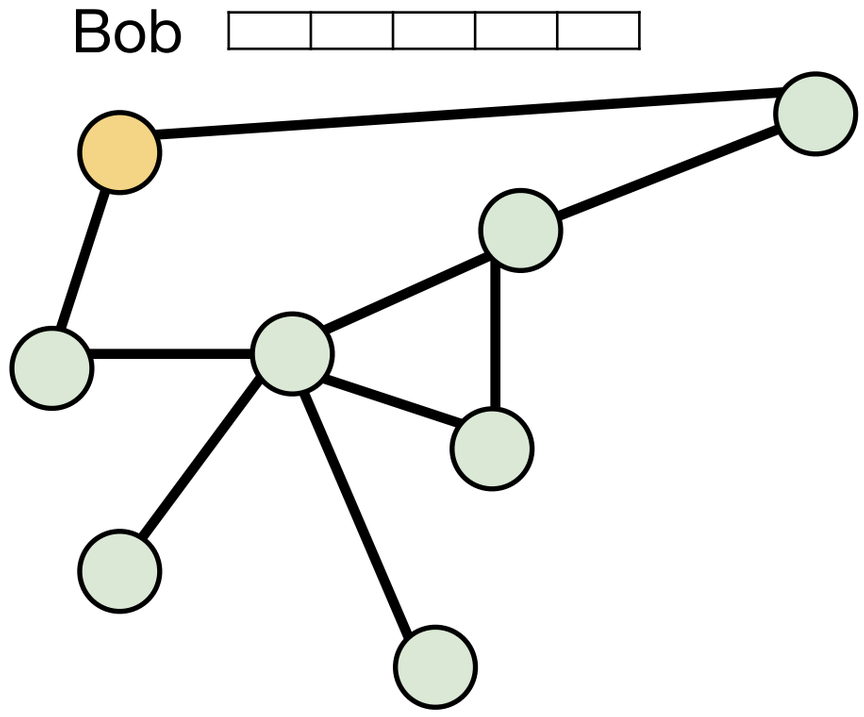


Explanation

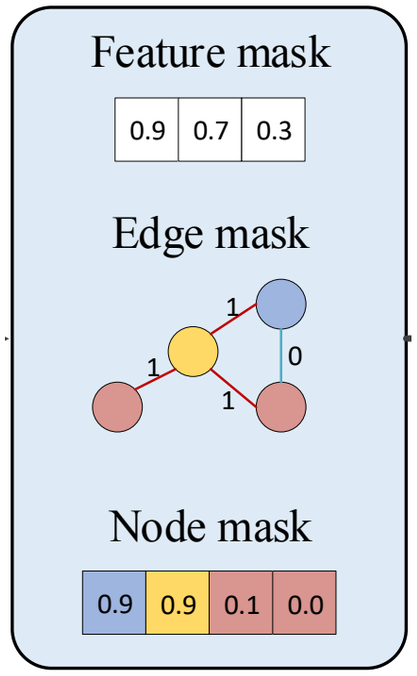


Transparency

Why was Bob's loan denied?



Explanation



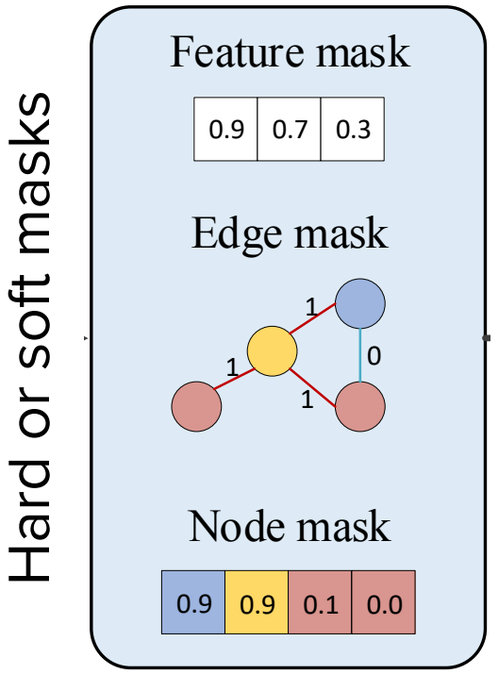
Decision has to be explained not only in terms of features but also graph structure. General explainability methods cannot be trivially applied for graphs.

Post-hoc explanations

Node Set **Edge Set** **Node/Edge Features**

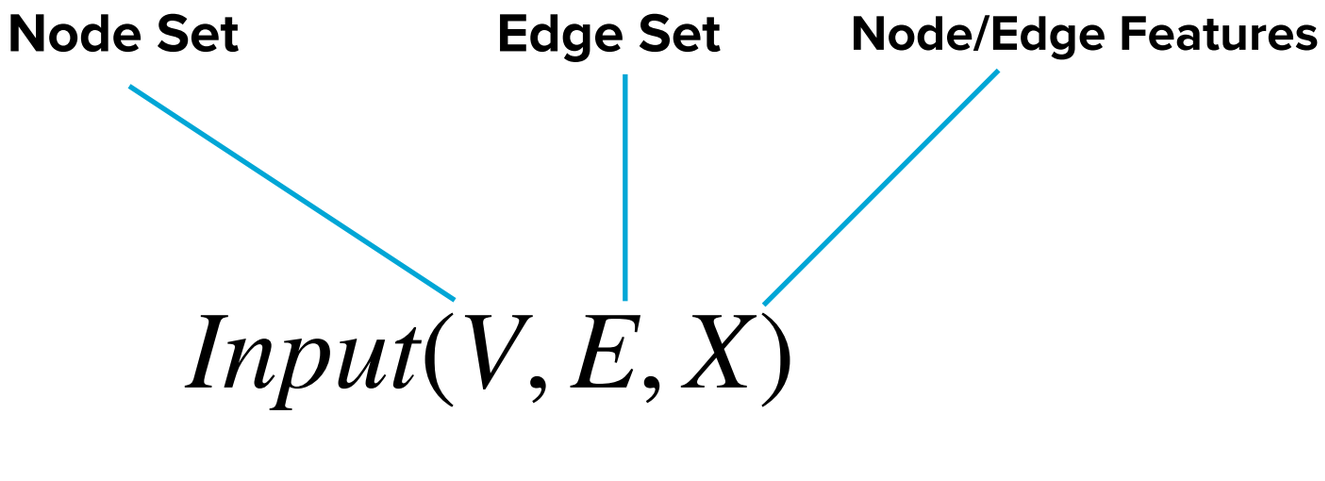
$$Input(V, E, X)$$

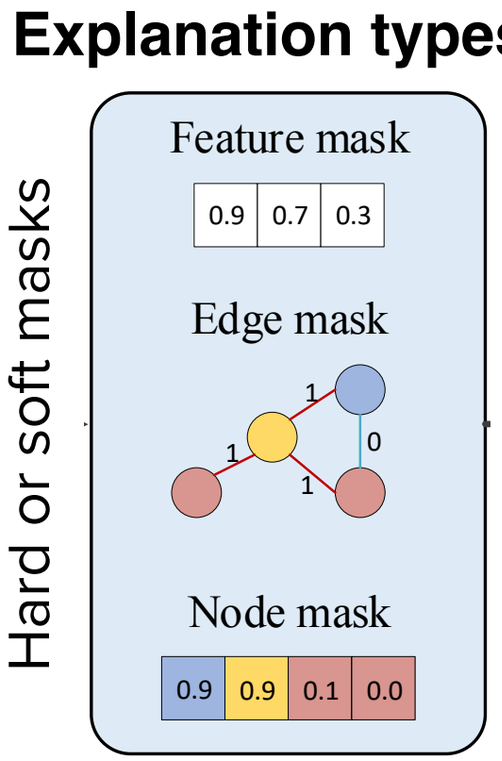
Explanation types



Examples: GNNExplainer, Zorro, PGExplainer

Post-hoc explanations





Examples: GNNExplainer, Zorro, PGExplainer

Explanation types:

Feature explanations in terms of most relevant features $X' \subset X$

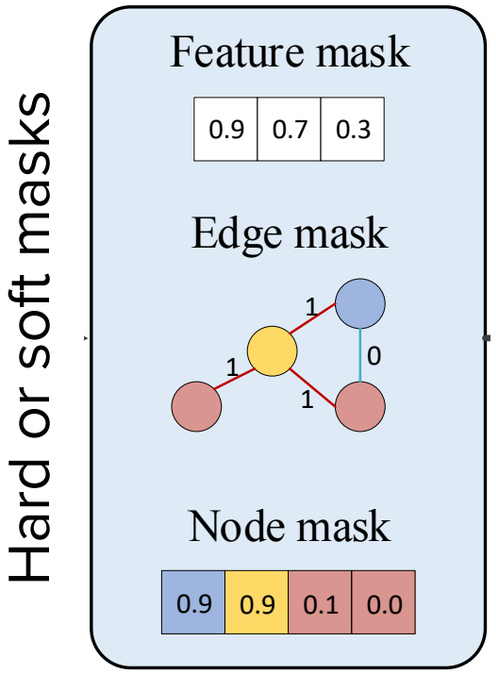
Structure explanations in terms of most relevant nodes $(V' \subset V)$ or edges $(E' \subset E)$

Post-hoc explanations

Node Set **Edge Set** **Node/Edge Features**

$$Input(V, E, X)$$

Explanation types



Examples: GNNExplainer, Zorro, PGExplainer

Explanation types:

Feature explanations in terms of most relevant features $X' \subset X$

Structure explanations in terms of most relevant nodes $(V' \subset V)$ or edges $(E' \subset E)$

We are interested in finding both feature and structure explanations which effectively capture interplay of structure and features in model's decision making.

Privacy

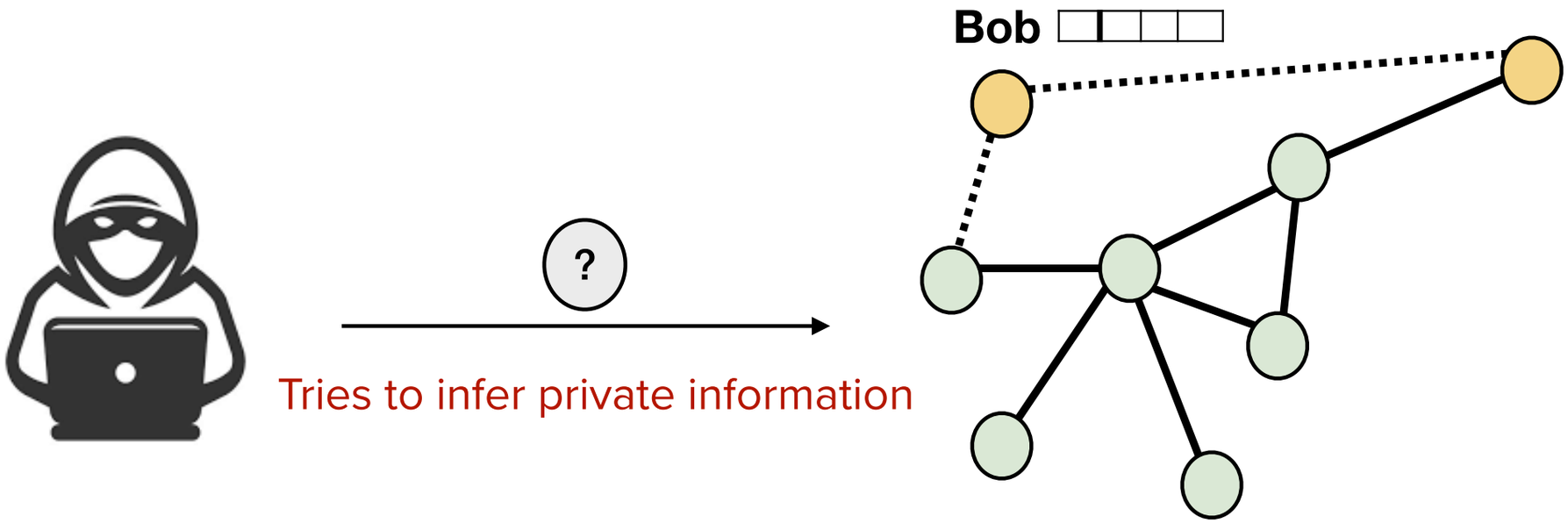
Graphs can contain sensitive information

- User's sensitive attributes
- Sensitive relations

GNNs encode relation information within the model, could memorise such information

- Your identity could be revealed because of your neighbour

Privacy

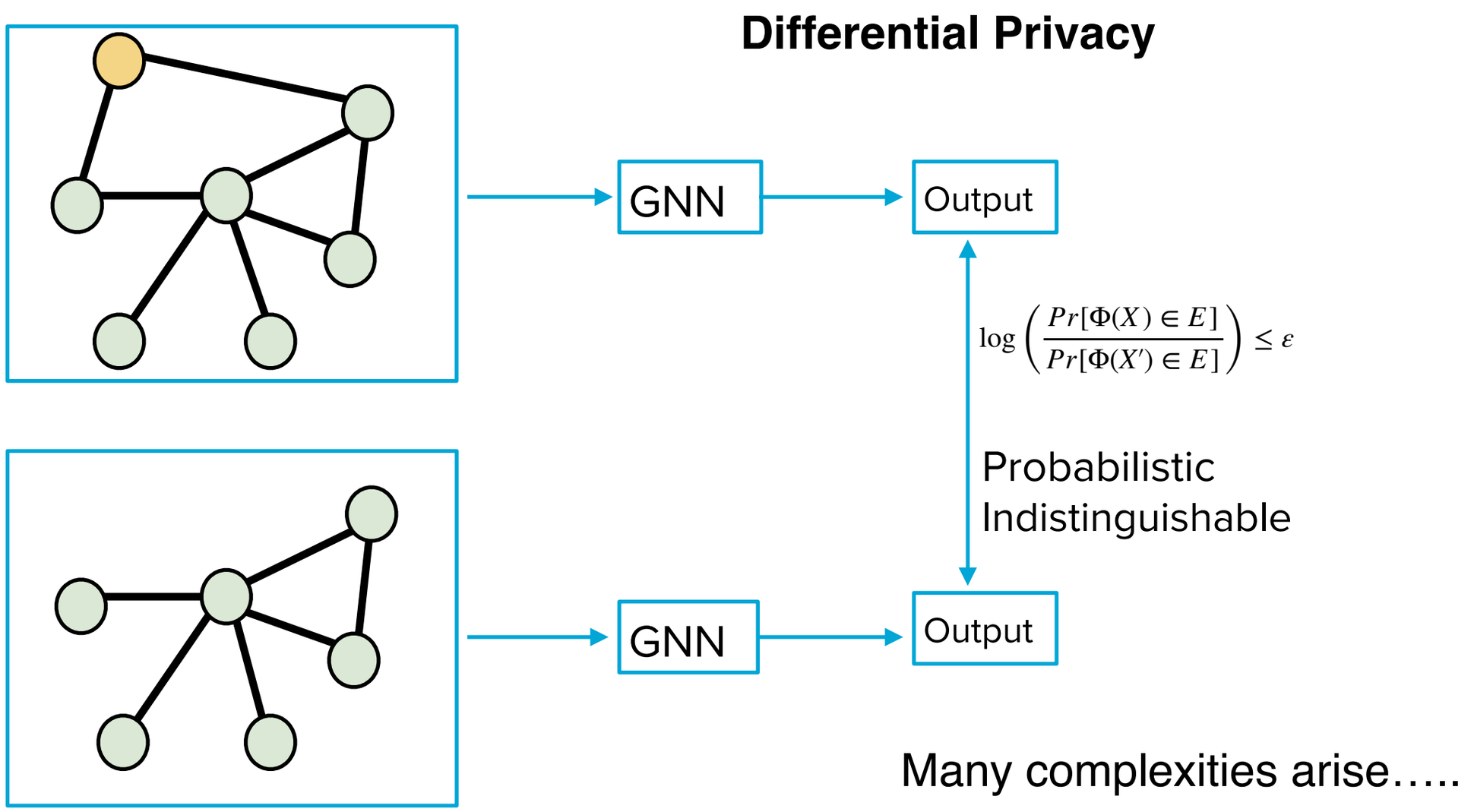


Node Membership Inference : Is Bob a part of training data? [Olatunji et al., ’21] [Duddu et al., ’20]

Relation reconstruction : Who are friends of Bob? [He et al., ’21] [Zhang et al., ’20]

Attribute Inference : Does Bob smoke?

Building Private GNN Models



Building Privacy Preserving models for graphs

A direct application of techniques like DP-SGD is not possible due to

- Unbounded sensitivity (think of the effect of leaving out or adding one node in a graph)

- Violation of i.i.d. assumption

- Need for inference privacy (as training data might be used during inference)

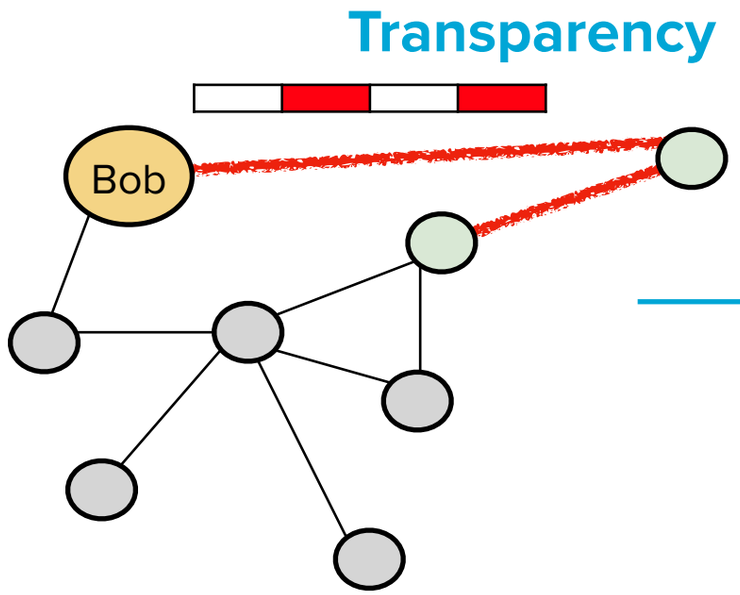
PrivGNN (Olatunji, Funke, Khosla, 2021), GAP (Sajadmanesh, Shamsabadi, A, Bellet, et al. 2022)

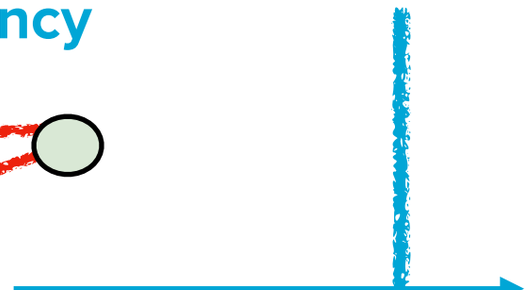
Transparency - Privacy Tradeoffs

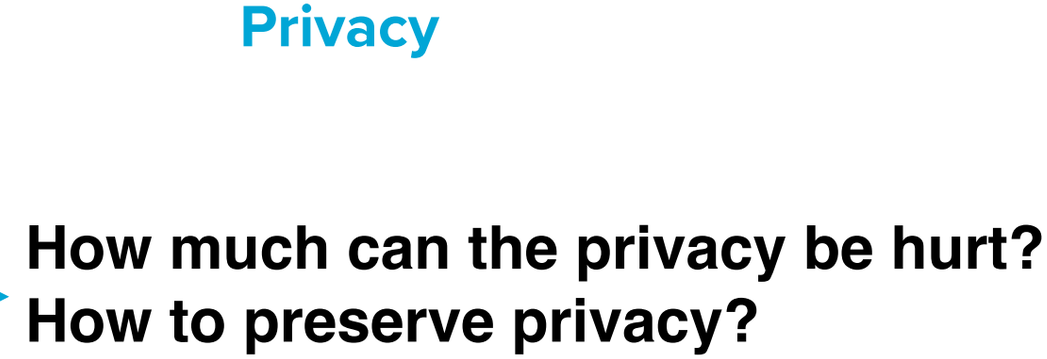
But we want our models to be **transparent** and **private** simultaneously

Transparency - Privacy Tradeoffs

But we want our models to be **transparent** and **private** simultaneously



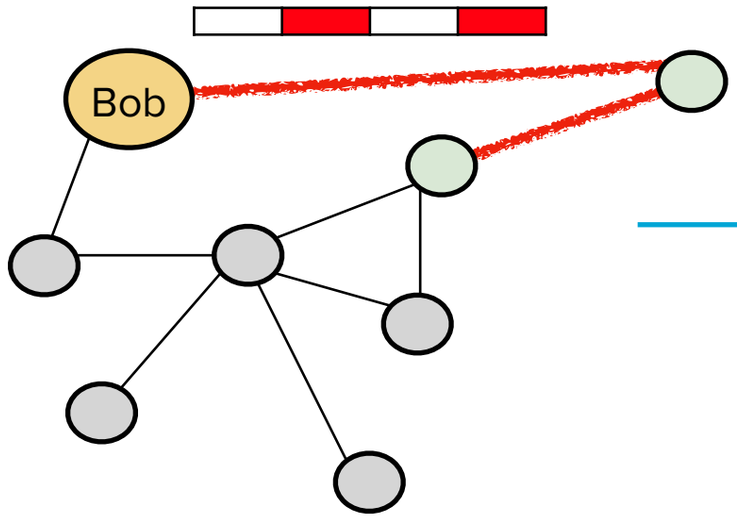




Transparency - Privacy Tradeoffs

But we want our models to be **transparent** and **private** simultaneously

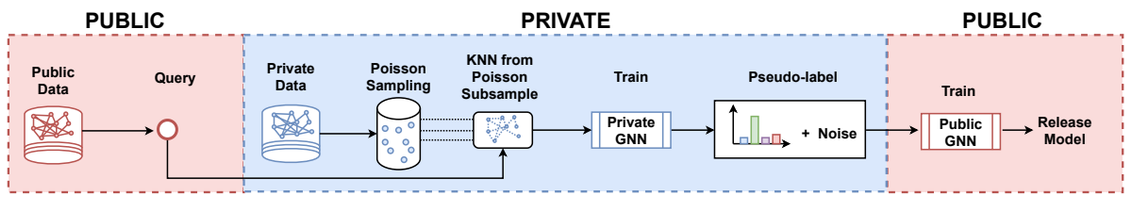
Transparency



Can we explain private models? How?

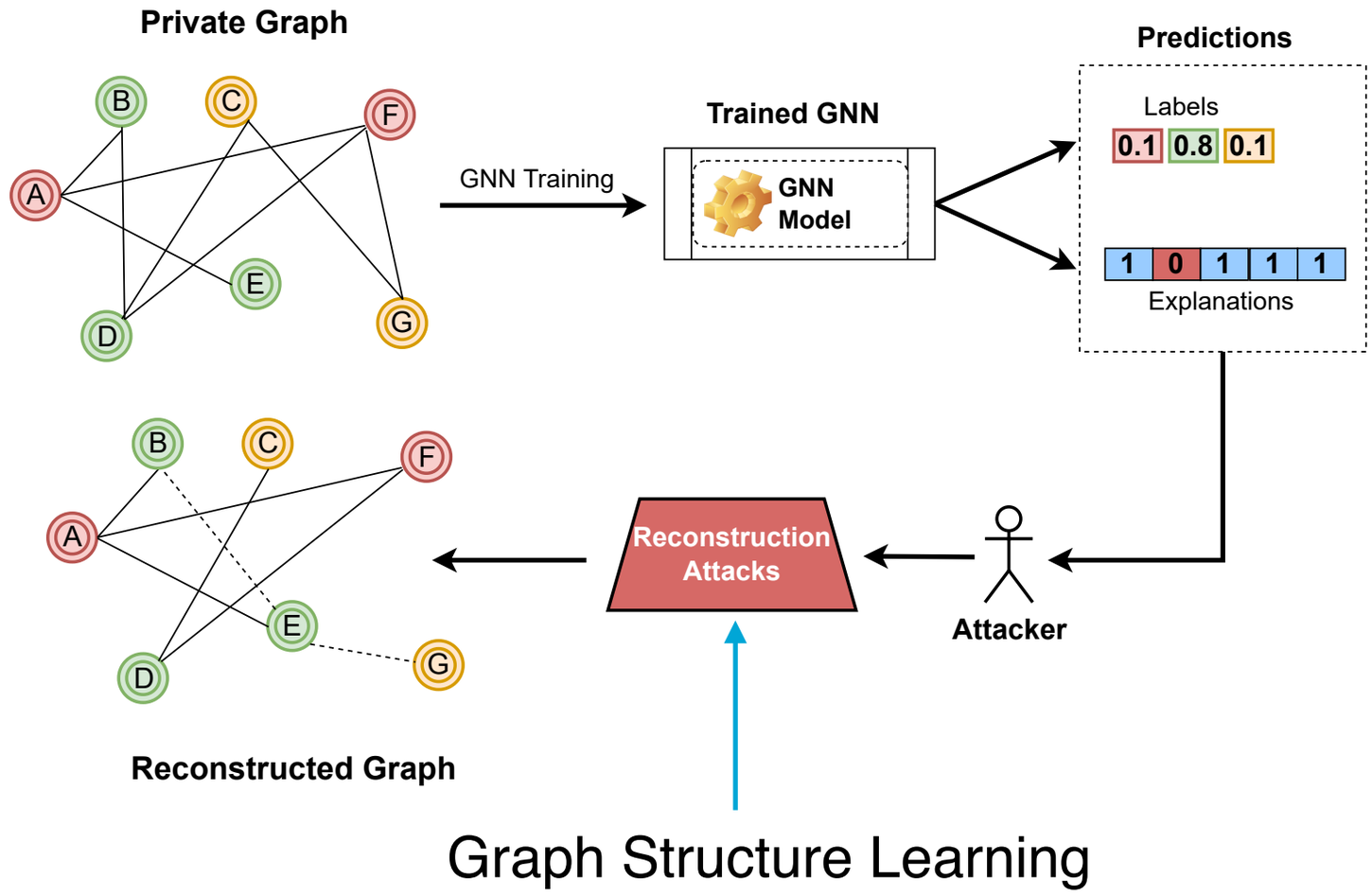
Privacy

How much can the privacy be hurt?
How to preserve privacy?



Complex privacy preserving mechanism

Reconstructing graphs from feature explanations



Private Graph Extraction via Feature Explanations [Olatunji et al. 2022]
https://arxiv.org/abs/2206.14724

Some interesting findings

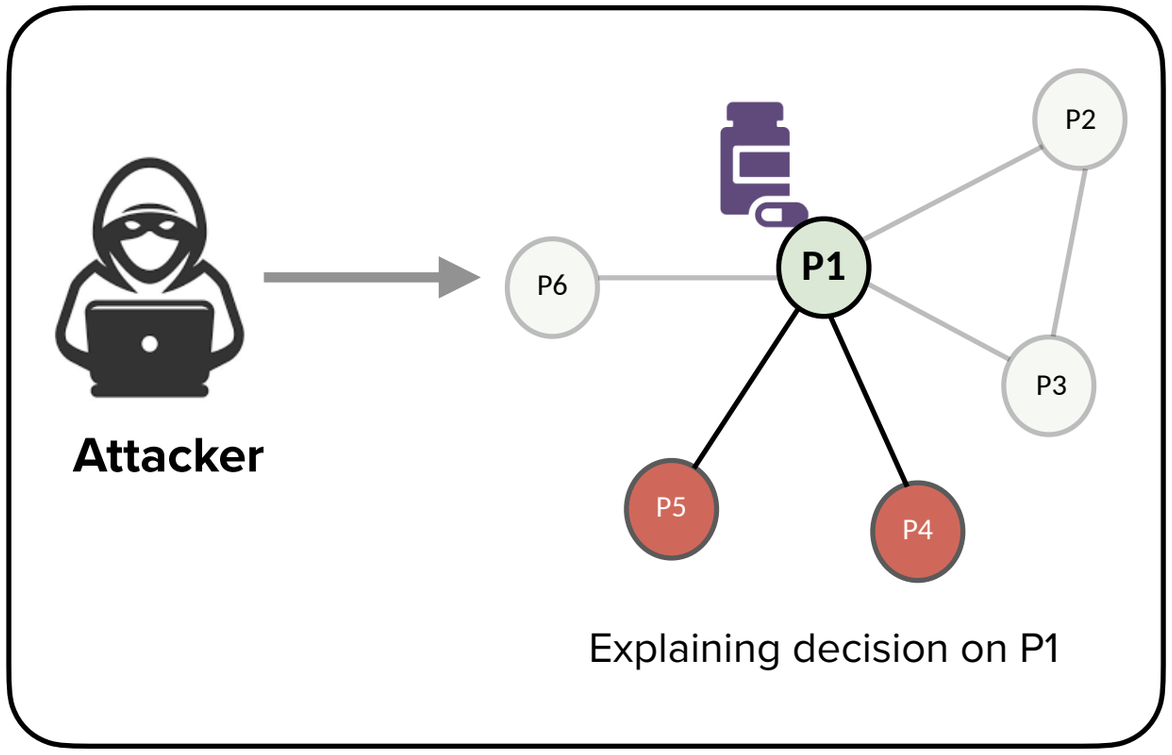
- Training graph could be reconstructed using alone the feature explanations and the labels

- Certain explanations leak more information than others

- Gradient based explanations incur high privacy loss while showing low **utility** (quantified by high faithfulness and sparsity)

Challenges

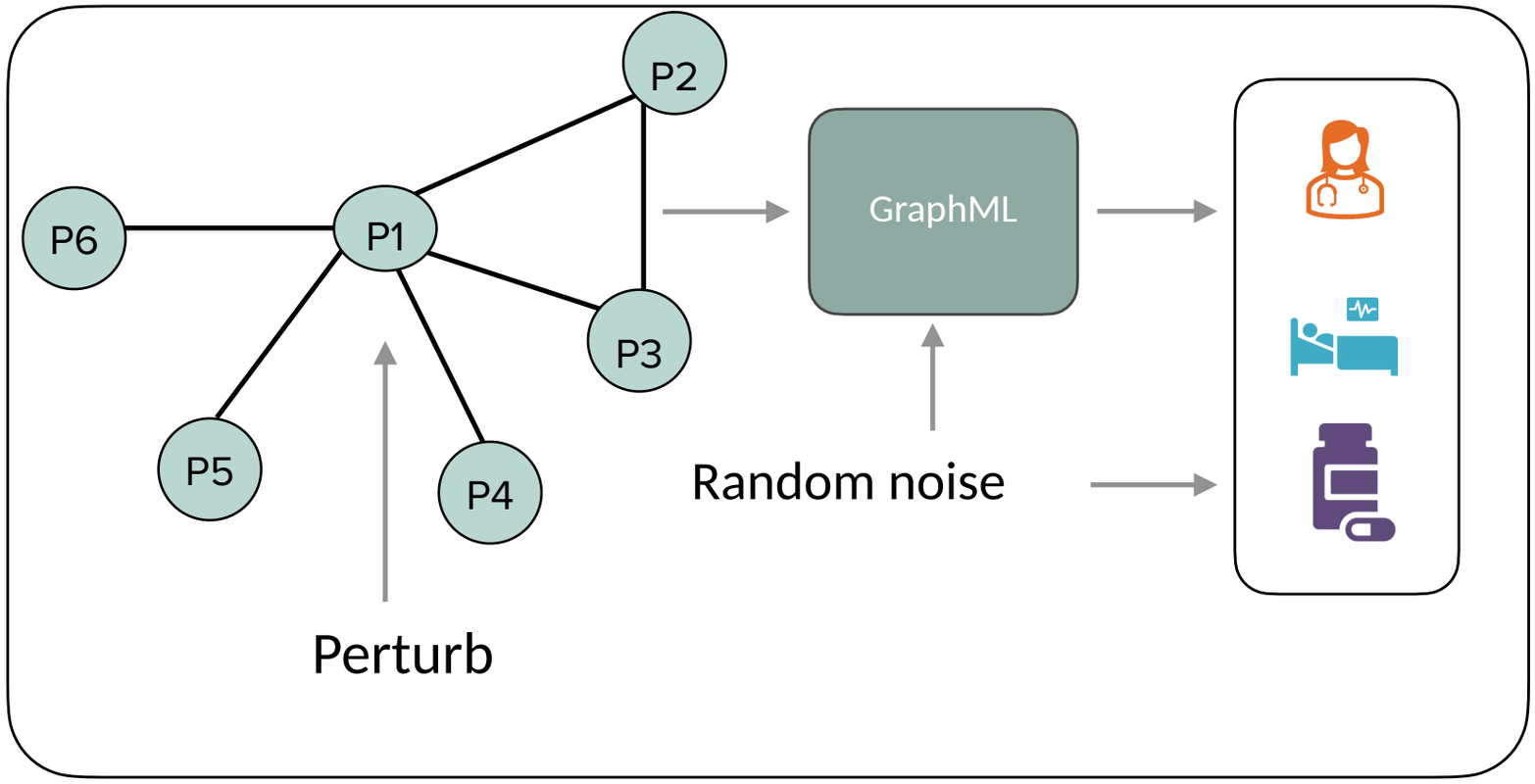
Structure explanations can directly reveal information about neighbours



Explanations of neighbouring datapoints would be correlated

Challenges

Private learning over graphs is more complex than that in standard ML



How to define explanation for a private model?

Research Directions and Open Questions

Quantification of privacy leakage in presence of different explanation types

- How can we measure information leakage due to different explanation types?
- Risk-utility assessment of different explainers/explanations
- Can we release explanations privately while still maintaining their utility?

Research Directions and Open Questions

Quantification of privacy leakage in presence of different explanation types

- How can we measure information leakage due to different explanation types?
- Risk-utility assessment of different explainers/explanations
- Can we release explanations privately while still maintaining their utility?

Explaining the decisions of privacy-preserving models

Research Directions and Open Questions

Quantification of privacy leakage in presence of different explanation types

- How can we measure information leakage due to different explanation types?
- Risk-utility assessment of different explainers/explanations
- Can we release explanations privately while still maintaining their utility?

Explaining the decisions of privacy-preserving models

- What should be the properties of an explanation for a privacy-preserving model?
  - Such properties might need to be defined based on the private learning strategy

Research Directions and Open Questions

Quantification of privacy leakage in presence of different explanation types

- How can we measure information leakage due to different explanation types?
- Risk-utility assessment of different explainers/explanations
- Can we release explanations privately while still maintaining their utility?

Explaining the decisions of privacy-preserving models

- What should be the properties of an explanation for a privacy-preserving model?
  - Such properties might need to be defined based on the private learning strategy
- How to release such explanations in a private manner?

Research Directions and Open Questions

Quantification of privacy leakage in presence of different explanation types

- How can we measure information leakage due to different explanation types?
- Risk-utility assessment of different explainers/explanations
- Can we release explanations privately while still maintaining their utility?

Explaining the decisions of privacy-preserving models

- What should be the properties of an explanation for a privacy-preserving model?
  - Such properties might need to be defined based on the private learning strategy
- How to release such explanations in a private manner?

Joint optimization of privacy and transparency

Research Directions and Open Questions

Quantification of privacy leakage in presence of different explanation types

- How can we measure information leakage due to different explanation types?
- Risk-utility assessment of different explainers/explanations
- Can we release explanations privately while still maintaining their utility?

Explaining the decisions of privacy-preserving models

- What should be the properties of an explanation for a privacy-preserving model?
  - Such properties might need to be defined based on the private learning strategy
- How to release such explanations in a private manner?

Joint optimization of privacy and transparency

- How can we optimise for the combined requirements of privacy and transparency in GraphML?

Research Directions and Open Questions

Quantification of privacy leakage in presence of different explanation types

- How can we measure information leakage due to different explanation types?
- Risk-utility assessment of different explainers/explanations
- Can we release explanations privately while still maintaining their utility?

Explaining the decisions of privacy-preserving models

- What should be the properties of an explanation for a privacy-preserving model?
  - Such properties might need to be defined based on the private learning strategy
- How to release such explanations in a private manner?

Joint optimization of privacy and transparency

- How can we optimise for the combined requirements of privacy and transparency in GraphML?
- Can we use interpretability techniques to overcome model overfitting which is the major cause of privacy leakage?