A Hands-on Introduction to Graph Deep Learning, with Examples in PyTorch Geometric - III

Machine Learning and Dynamical Systems Seminar

November 16, 2023

Gabriele Santin (gabrielesantin.github.io)
Antonio Longa (antoniolonga.github.io)
Steve Azzolin (steveazzolin.github.io)
Francesco Ferrini (francescoferrini.github.io)

Introduction
About us

Gabriele Santin
Assistant professor at University of Venice (Venice, Italy)
gabriele.santin@unive.it
https://gabrielesantin.github.io/

Antonio Longa
Assistant professor University of Trento (Trento, Italy)
antonio.longa@unitn.it
https://antoniolonga.github.io/

Steve Azzolin
ELLIS PhD Student at FBK and University of Trento (Trento, Italy)
steve.azzolin@unitn.it
https://steveazzolin.github.io/

Francesco Ferrini
PhD Student at University of Trento, (Trento, Italy)
francesco.ferrini@unitn.it
https://francescoferrini.github.io/

Introduction
Organization and material

Tutorial in four parts (slides + Jupyter notebooks available at github.com/steveazzolin/gdl_tutorial_turinginst):

- **Part I:** November 2, Presenter: **GS**
  Goals: Motivations, Intro of basic concepts, definition of GNNs
- **Part II:** November 9, Presenter: **AL**
  Goals: Implementation of GNNs: How to implement a full GNN pipeline in PyTorch Geometric.
- **Part III:** November 16, Presenter: **SA**
  Goals: Explainability of GNNs: How to shed (a bit of) light into the black box
- **Part IV:** November 23, Presenter: **FF**
  Goals: Heterogeneity in GNNs: How can GNNs effectively model and incorporate a diversity of nodes and edges with different types.

Introduction
Agenda

1. Why **XAI** for deep learning models?
2. What can **XAI** do?
3. **XAI4GNNs**

Why XAI for deep learning models?

Deep learning models achieve great performances in many tasks and they are more and more adopted also in high-stakes applications

- Surveillance
- Crime rate predictions
- Medical Diagnosis
- Autonomous Driving
- …

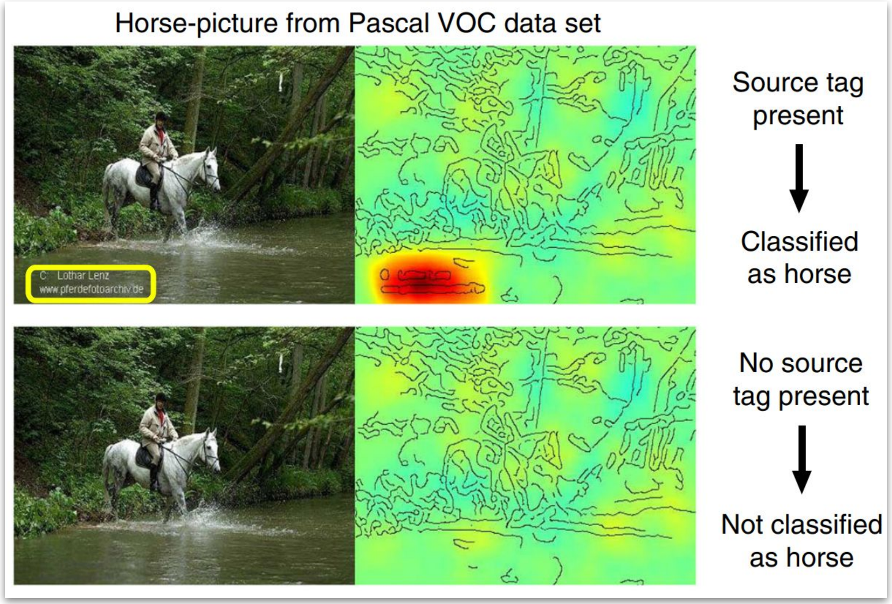
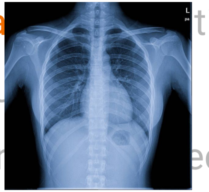


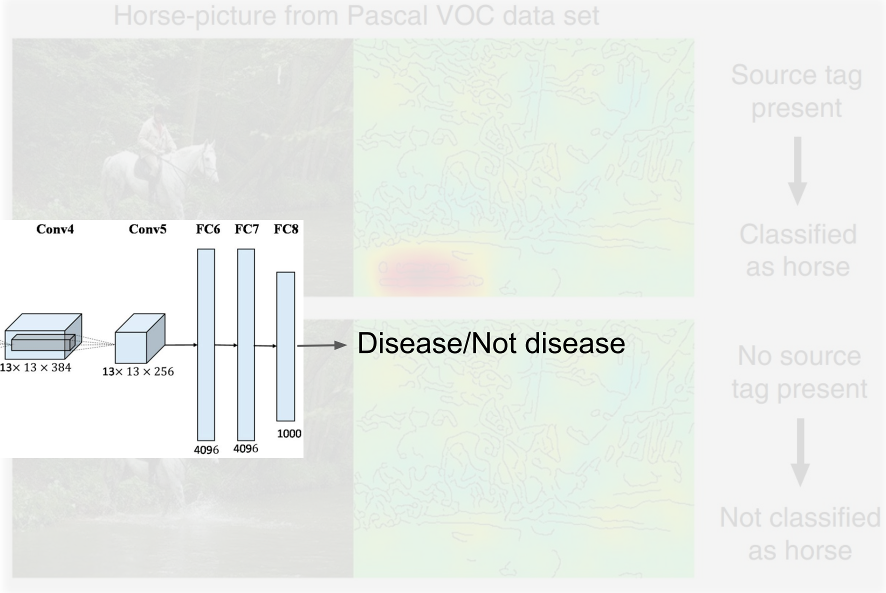
Image from S. Lapuschkin et al., 2019

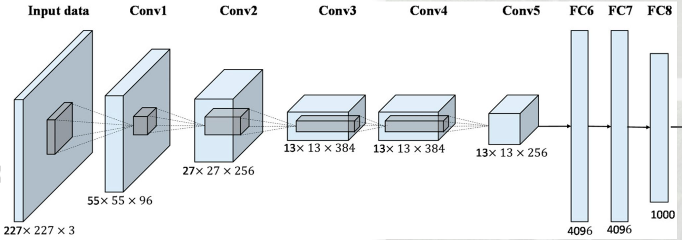
Why XAI for deep learning models?

Deep learning models achieve great performances in many tasks and they are more and more adopted also in high-sta… tions



- Su…
- Cr… edicti…
- Medical Diagnosis
- …





Images from “Pre-Trained AlexNet Architecture with Pyramid Pooling and Supervision for High Spatial Resolution Remote Sensing Image Scene Classification”, Getty Images

Why XAI for deep learning models?

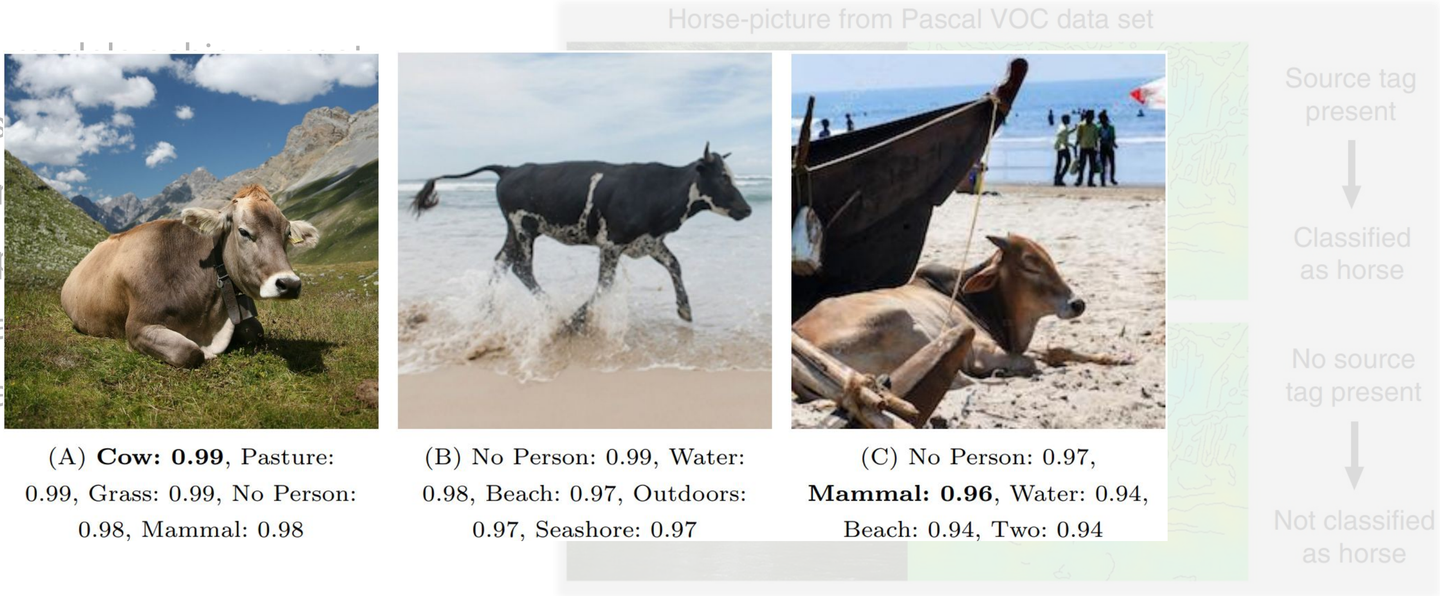
Deep learning

performances

more and mor

high-stakes a

- Surveilla
- Crime ra
- Medical
- ...



(A) **Cow: 0.99**, Pasture: 0.99, Grass: 0.99, No Person: 0.98, Mammal: 0.98

(B) No Person: 0.99, Water: 0.98, Beach: 0.97, Outdoors: 0.97, Seashore: 0.97

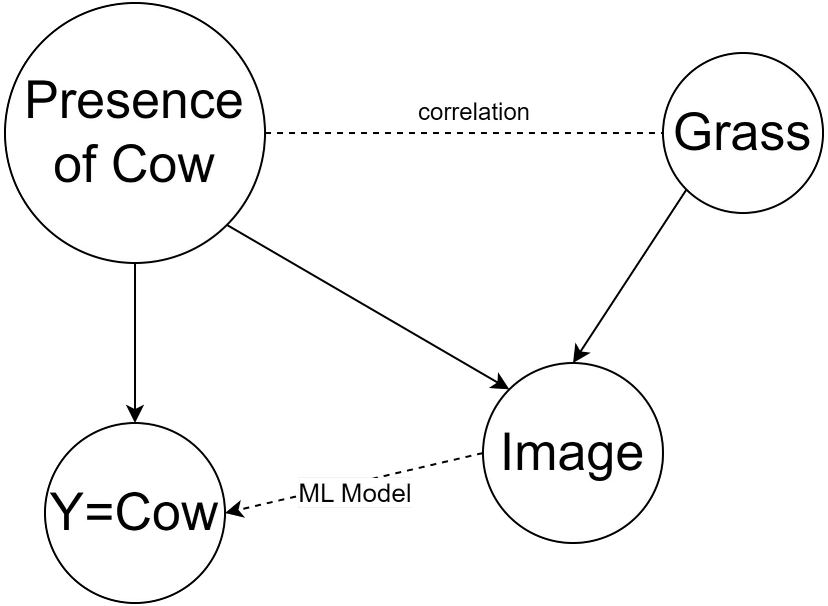
(C) No Person: 0.97, **Mammal: 0.96**, Water: 0.94, Beach: 0.94, Two: 0.94

Images from “Recognition in Terra incognita”, S.Beery, 2018

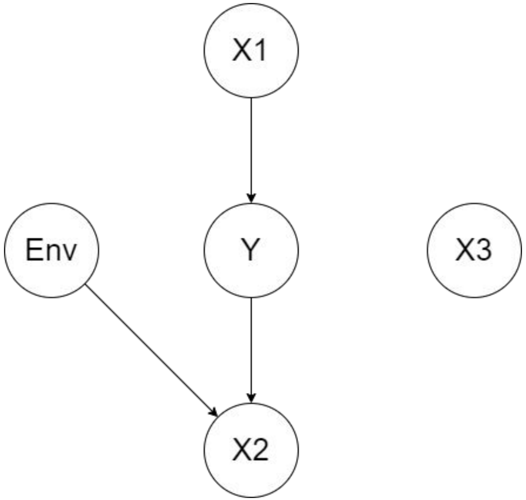
Why XAI for deep learning models?

Deep learning models struggle to capture true causal relationships.

They instead pick up subtle/shortcut features correlated to the target label, but not causally associated to it



Code Session I



Why XAI for deep learning models?
Not all ML models are created equal

White box*

- Linear Models
- Decision Trees

Shallow models, good for tabular data, good with few data

Black box

- CNNs
- Transformers
- GNNs

(Very) Deep models, good for unstructured/high-dimensional data, shine with big data

*caveat: Being simple does not imply being understandable by humans (think of a Decision Tree with thousands of leafs…) (Rudin, C. (2019))

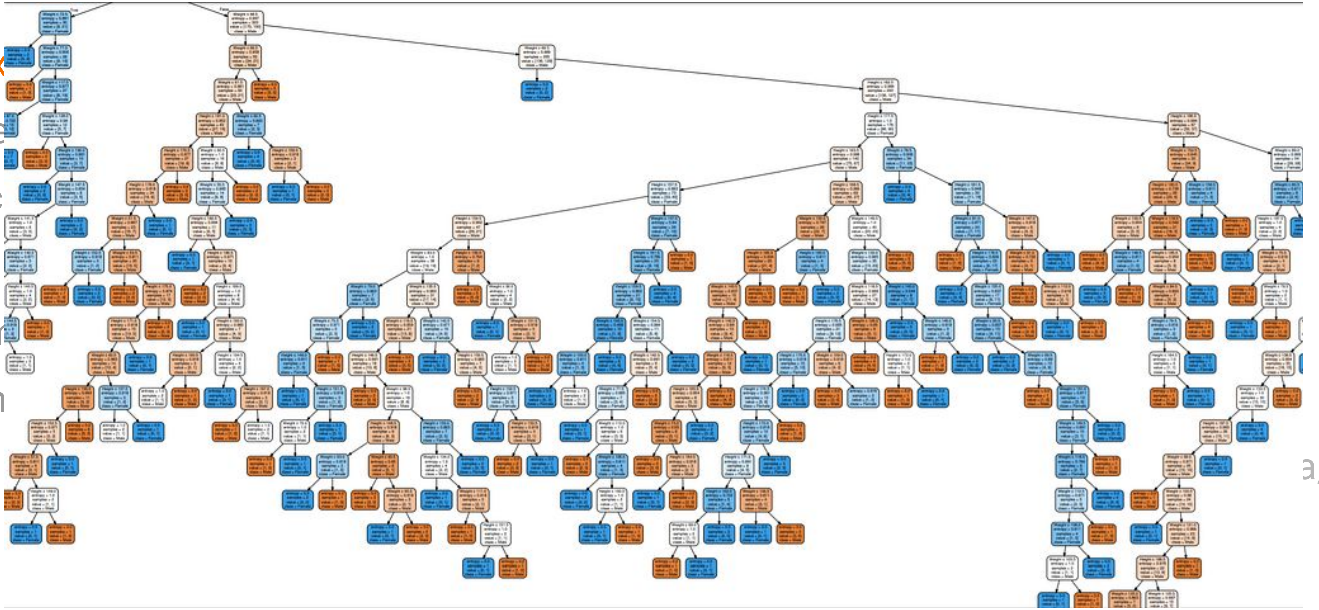
Why XAI for deep learning models?
Not all ML models are created equal

White box

- Line
- Dec

Shallow m

good with



a, shine

*caveat: Being simple does not imply being understandable by humans (think of a Decision Tree with thousands of leafs…) (Rudin, C. (2019))

Image from “Decision Trees Explained With a Practical Example”, Towards AI

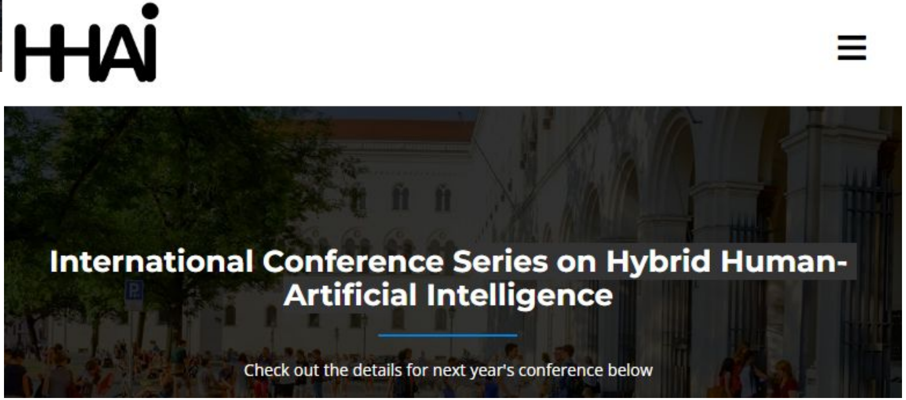
Why XAI for deep learning models?
Not all ML models are created equal

- We need external tools to shed light on the rationales behind deep models' predictions
- We need to design new deep models equipped with better interpretability
  - Concept Bottleneck Models (grey box models) ("Concept Bottleneck Models", Pang Wei Koh, 2020)

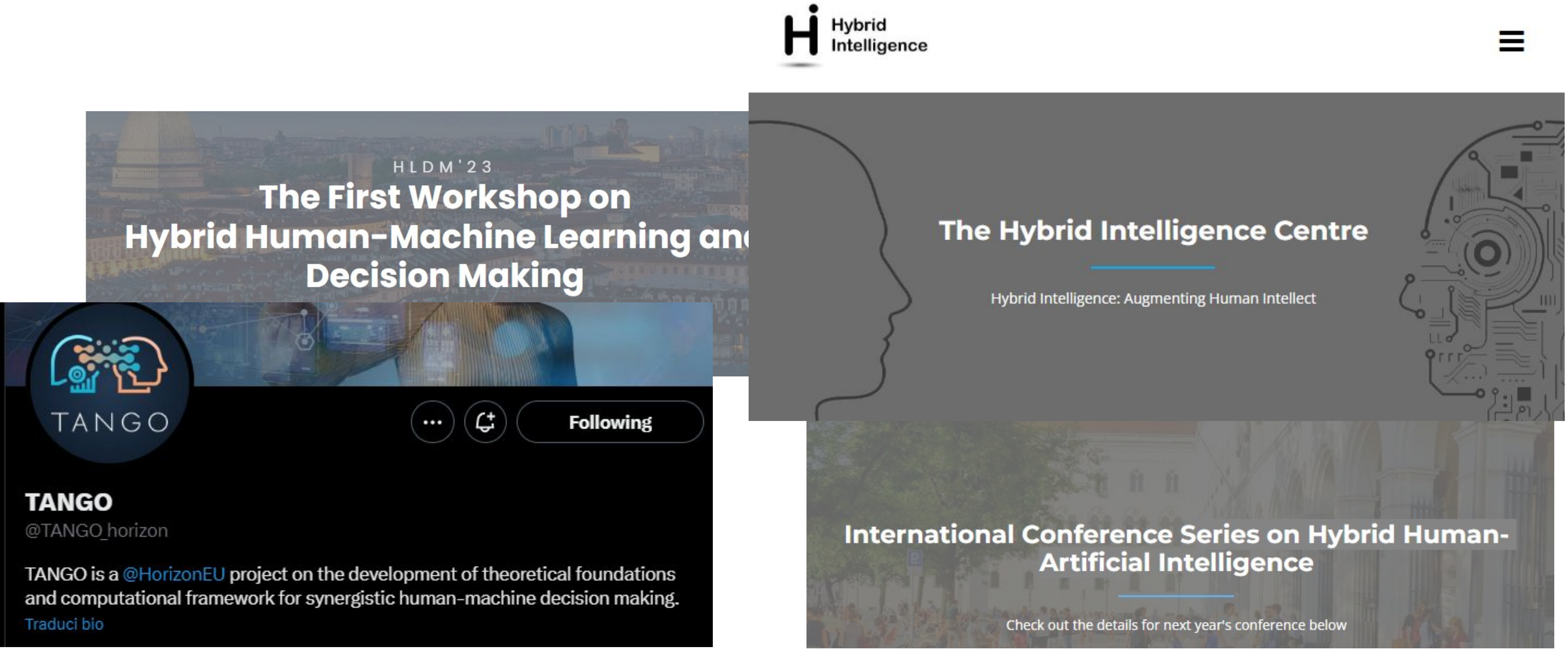
What can XAI do?
Use cases

- Hybrid Decision Making
- Algorithmic Recourse
- Inspect and debug models (*our focus*)

What can XAI do?
Hybrid Decision Making



What can XAI do?
Hybrid Decision Making



What can XAI do?
Algorithmic Recourse

- “the systematic process of reversing unfavourable decisions by algorithms and bureaucracies across a range of counterfactual scenarios” [Venkatasubramanian & Alfano, 2020; Karimi et al., 2021]

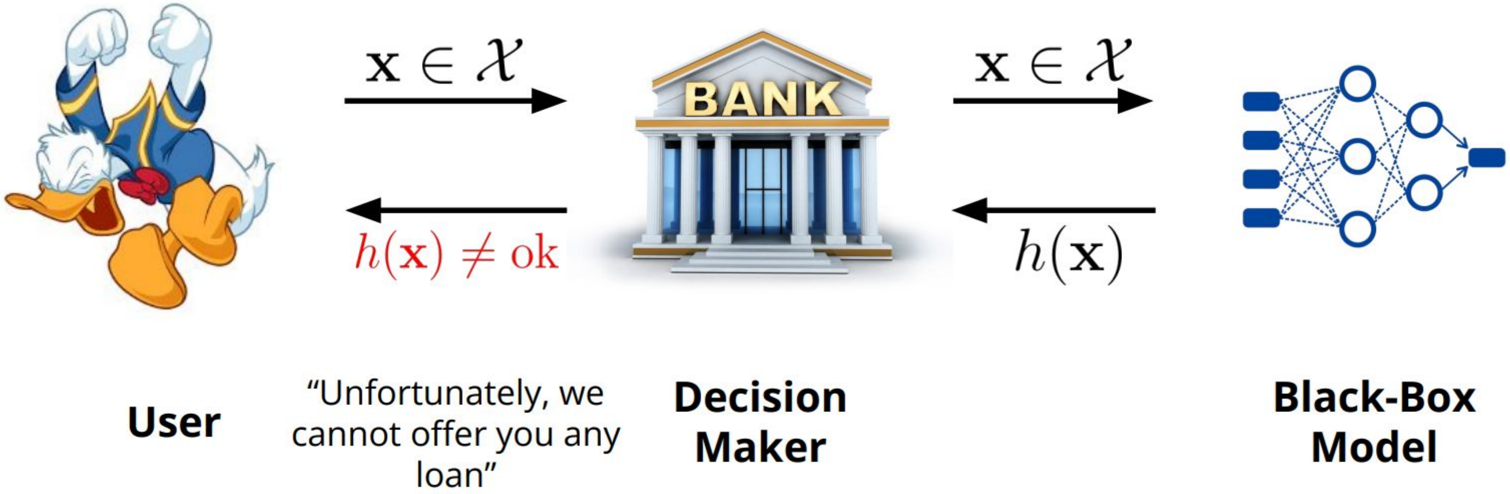


Image from https://aitrentojc.github.io/assets/slides/Talk_2_Algorithmic_Recourse_and_Explainable_Counterfactual_Interventions_Handsout_25-02-2022.pdf

What can XAI do?
Algorithmic Recourse

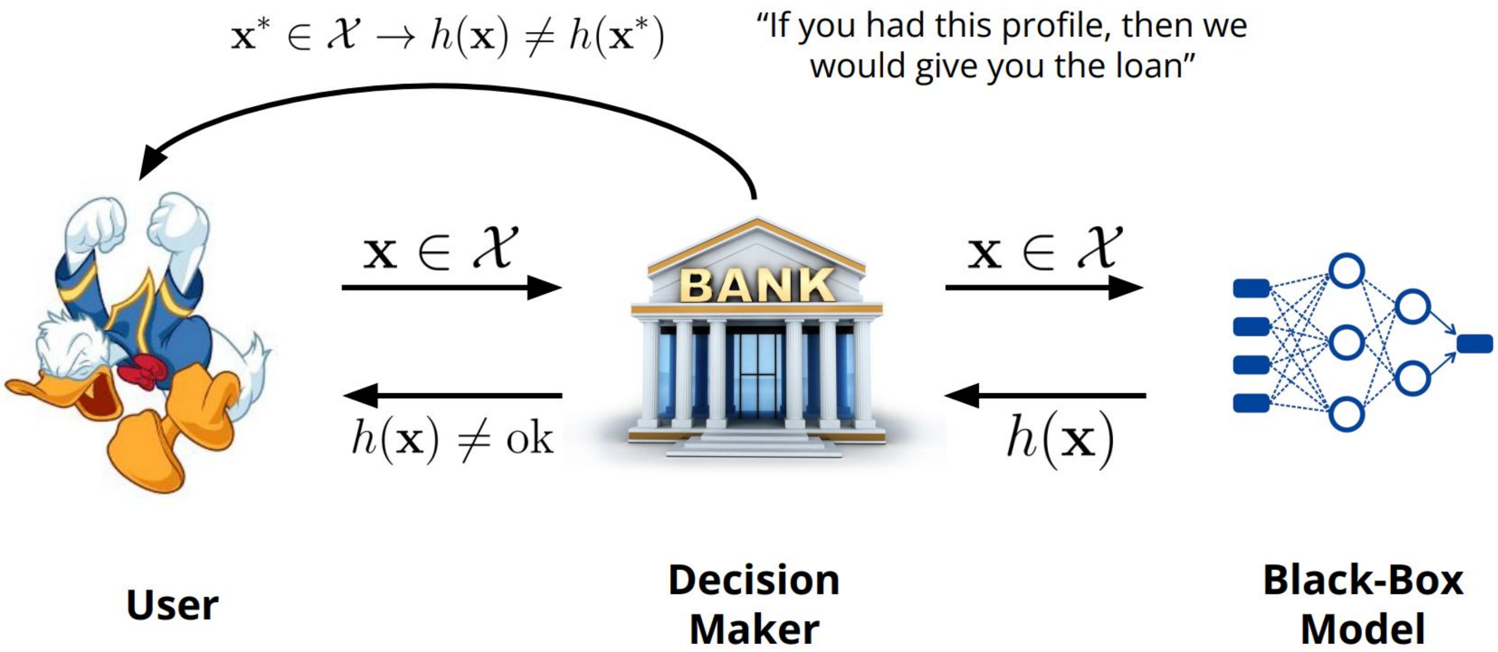


Image from https://aitrentojc.github.io/assets/slides/Talk_2_Algorithmic_Recourse_and_Explainable_Counterfactual_Interventions_Handsout_25-02-2022.pdf

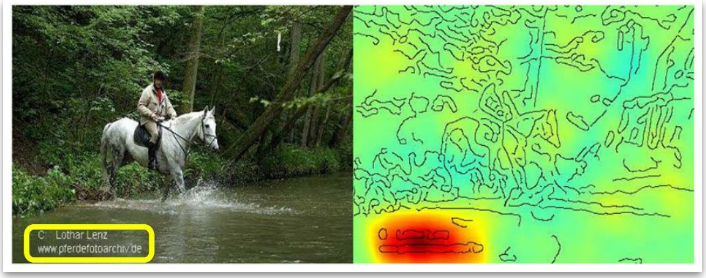
What can XAI do?
Model Debugging

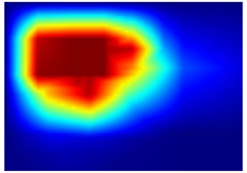
A bunch of methods have been proposed over time:

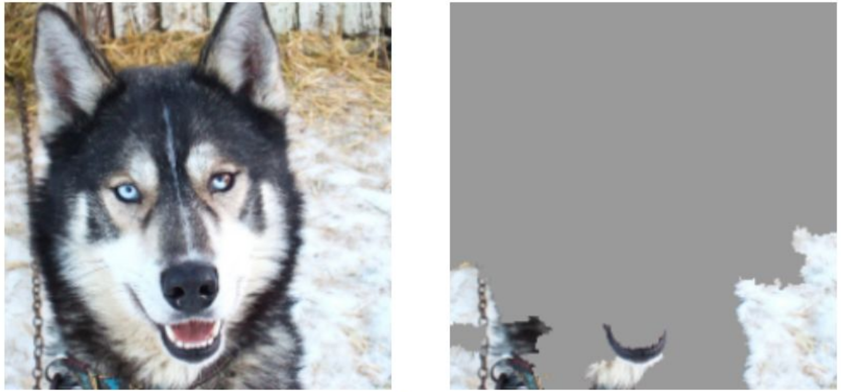
- Shapley values: Explaining prediction models and individual predictions with feature contributions. E. Štrumbelj et al., Knowledge and information systems, 2013
- CAM: Is Object Localization for Free? - Weakly-Supervised Learning With Convolutional Neural Networks. M. Oquab et al., CVPR, 2015
- LIME: “Why Should I Trust You?” Explaining the Predictions of Any Classifier . M. T. Ribeiro et al., ACM SIGKDD, 2016
- Integrated Gradients: Axiomatic Attribution for Deep Networks. M. Sundararajan, ICML, 2017
- …

What can XAI do?
Model Debugging

For image-like data the explanation is an heatmap or as relevance regions







(a) Husky classified as wolf (b) Explanation

Figure 11: Raw data and explanation of a bad model's prediction in the "Husky vs Wolf" task.

XAI4GNNs

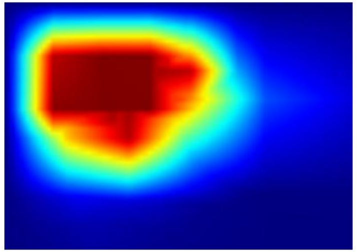
The graph domain poses unique challenges

- Unstructured data type
- Discrete objects
- Node/Edge/Graph attributes
- Different type of explanation

Develop novel techniques for GNNs

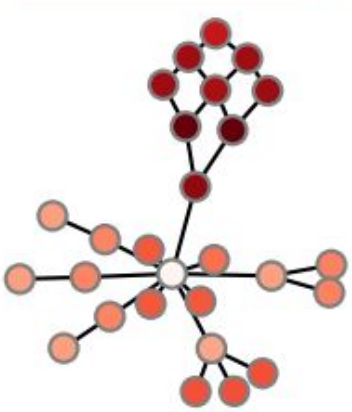
XAI4GNNs

Images

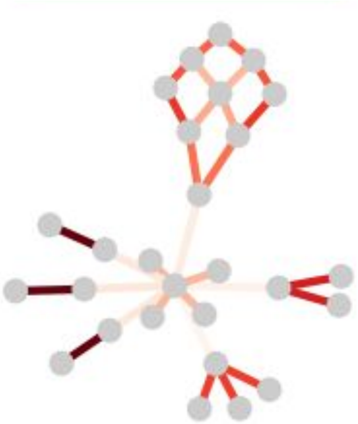


Graph

Node attribution



Edge attribution



Images from M. Oquab CVPR 2015 and A. Longa ArXiV 2022

XAI4GNNs
Taxonomy

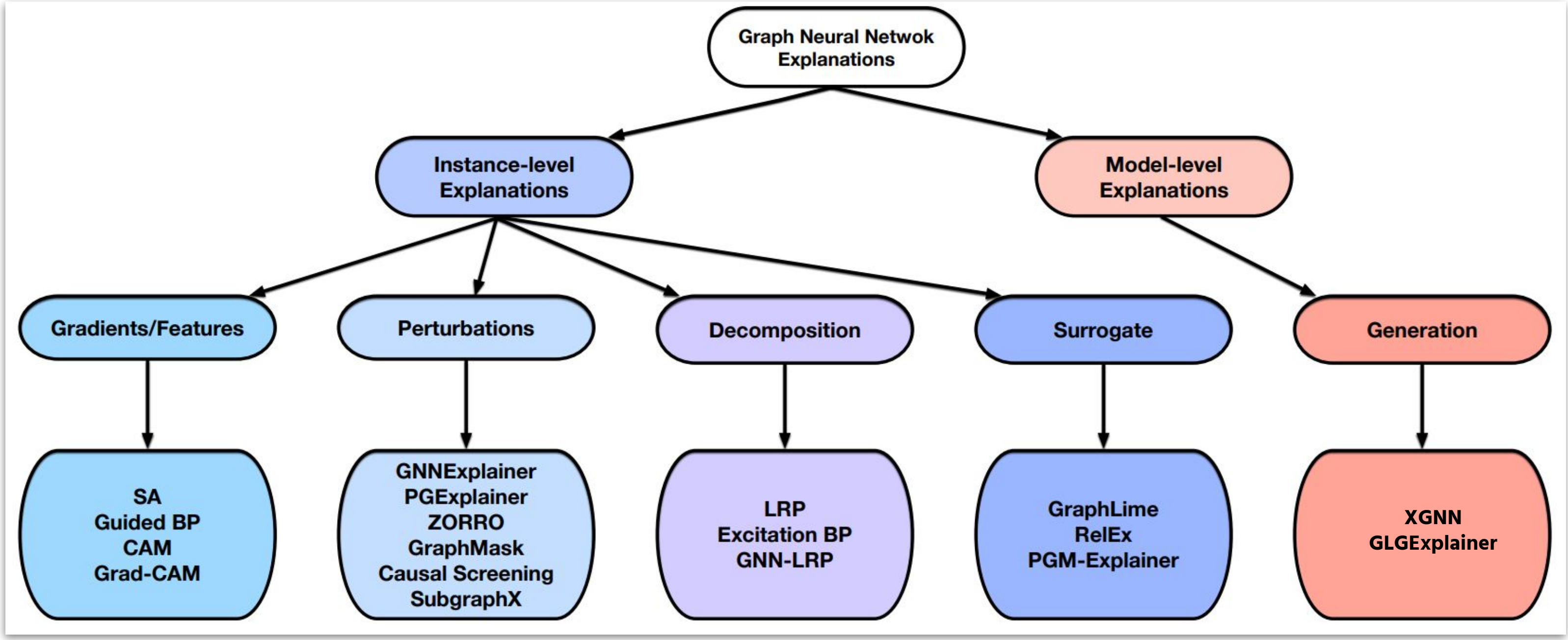
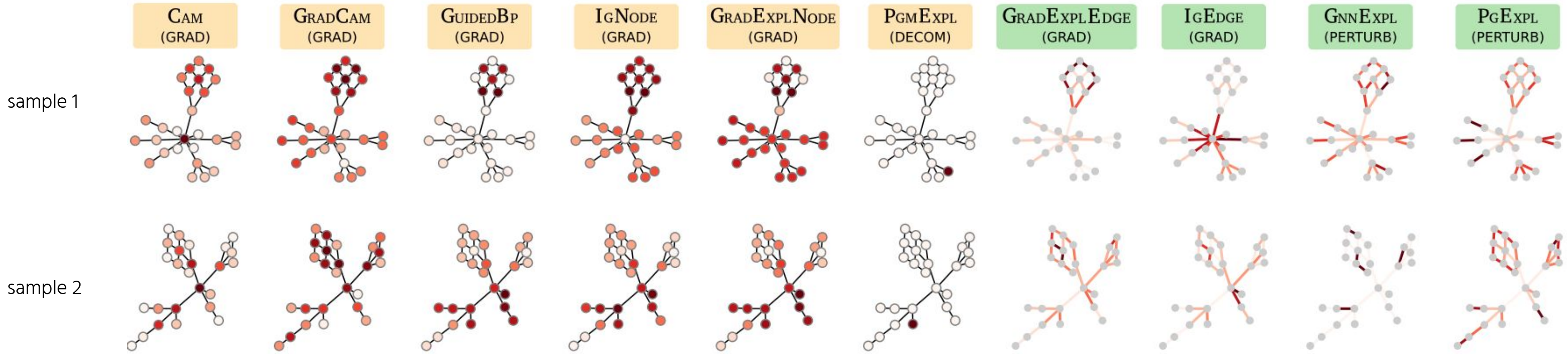


Image from Explainability in Graph Neural Networks: A Taxonomic Survey. H. Yuan et al., 2022 and then revised

XAI4GNNs
Local Explanations

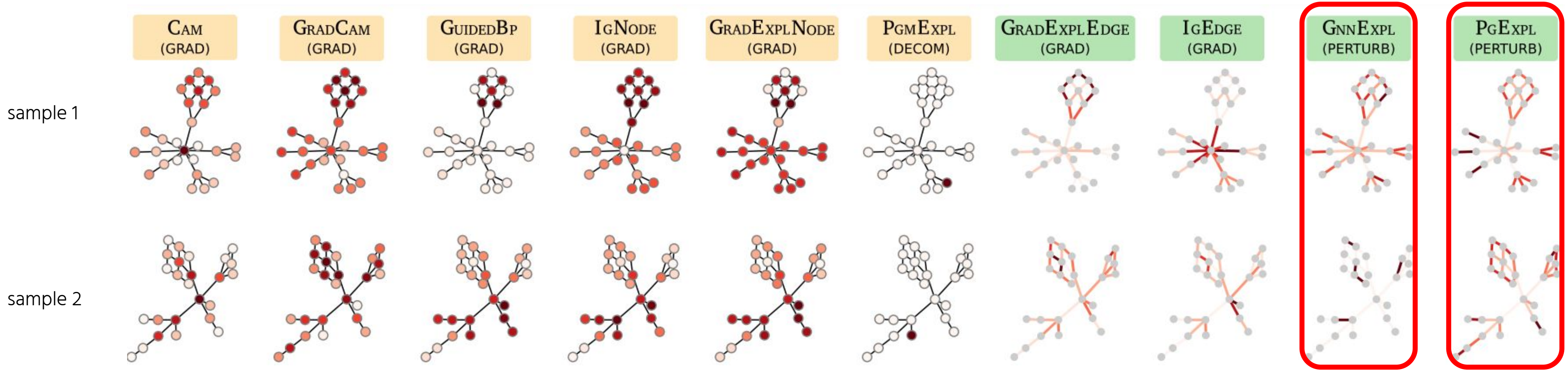
Local (or Instance-level) Explainers highlight the input features most relevant for the prediction of the model to explain



Images from Explaining the Explainers in Graph Neural Networks: a Comparative Study. A. Longa et al., 2022

XAI4GNNs
Local Explanations

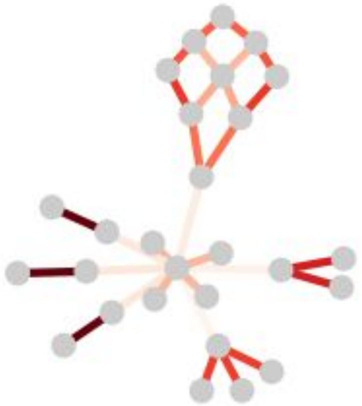
Local (or Instance-level) Explainers highlight the input features most relevant for the prediction of the model to explain



Images from Explaining the Explainers in Graph Neural Networks: a Comparative Study. A. Longa et al., 2022

XAI4GNNs
GNNExplainer

Edge attribution



Intuitively

1. If an edge is relevant, then removing it will decrease the confidence of the prediction
2. So, to find G_S seek for edges whose removal do not impact the prediction of the model, and remove them

Mathematically

$$max_{G_S} MI(Y, G_S) = max_{G_S} H(Y) - H(Y|G_S) = min_{G_S} H(Y|G_S)$$

Operationally, G_S is found by optimizing a mask over the graph

GNNExplainer: Generating Explanations for Graph Neural Networks. R. Ying, NeurIPS 2019

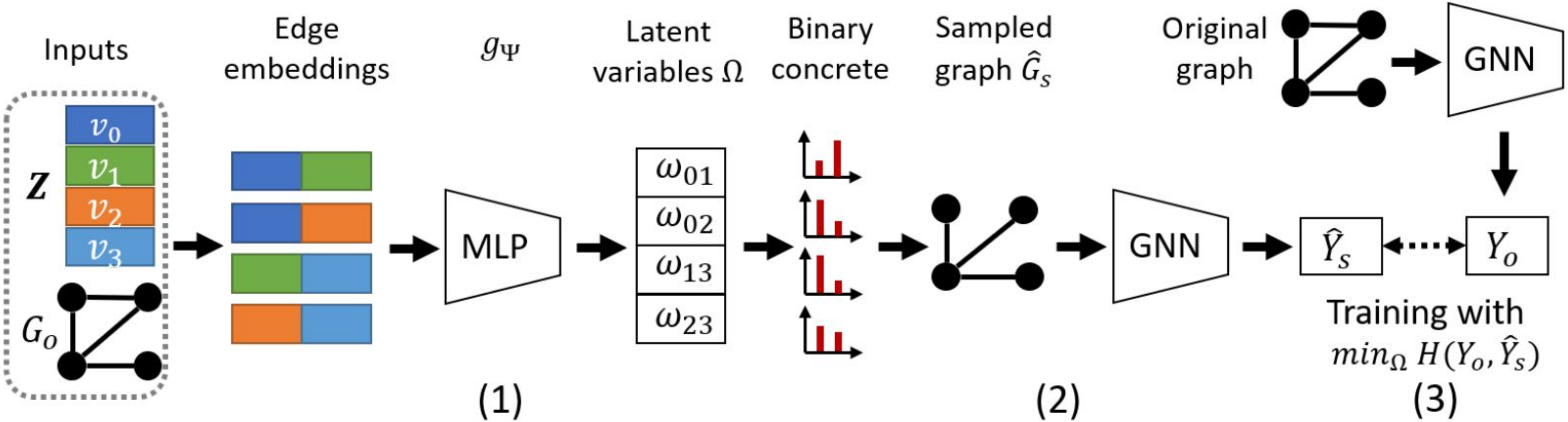
XAI4GNNs
PGExplainer

Intuitively

1. Based on the same principles as **GNNExplainer**
2. Instead of optimizing a mask for each input graph, train a Neural Network that given the features of an edge predicts its importance

Parameterized Explainer for Graph Neural Network. D. Luo, NeurIPS 2020

XAI4GNNs
PGExplainer



Parameterized Explainer for Graph Neural Network. D. Luo, NeurIPS 2020

XAI4GNNs
GNNExplainer vs PGExplainer

GNNExplainer

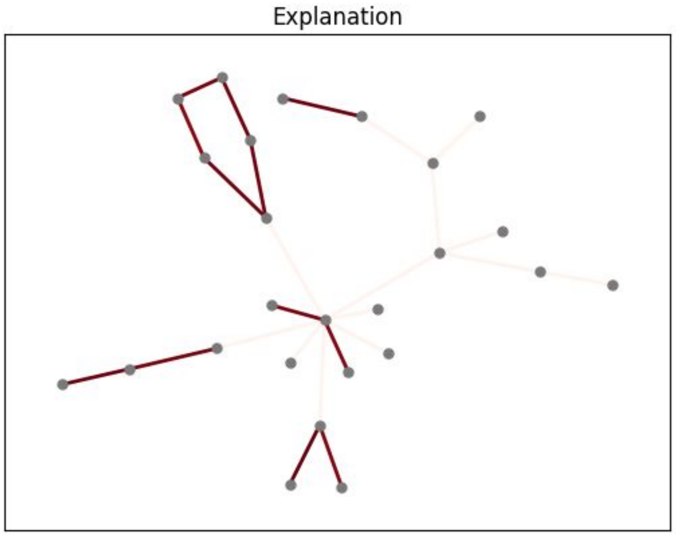
- Train a mask for each graph

PGExplainer

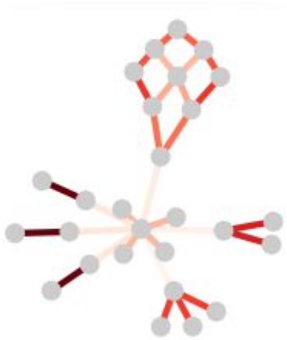
- Train a MLP once for all graphs
- Do inference with the MLP for each explanation

XAI4GNNs
Evaluation Metrics

- If ground truth available
  - accuracy/F1 of the explainer
- If not, unsupervised metrics
  - $Sparsity = \frac{1}{N}\sum_{i=1}^{N}(1 - \frac{|E_i|}{|G_i|})$ $\nearrow$
  - $Fidelity_+ = \frac{1}{N}\sum_{i=1}^{N}(f(G_i)_{\hat{y}} - f(G_i \setminus E_i)_{\hat{y}})$ $\nearrow$
  - $Fidelity_- = \frac{1}{N}\sum_{i=1}^{N}(f(G_i)_{\hat{y}} - f(E_i)_{\hat{y}})$ $\swarrow$



Code Session II



XAI4GNNs
Taxonomy

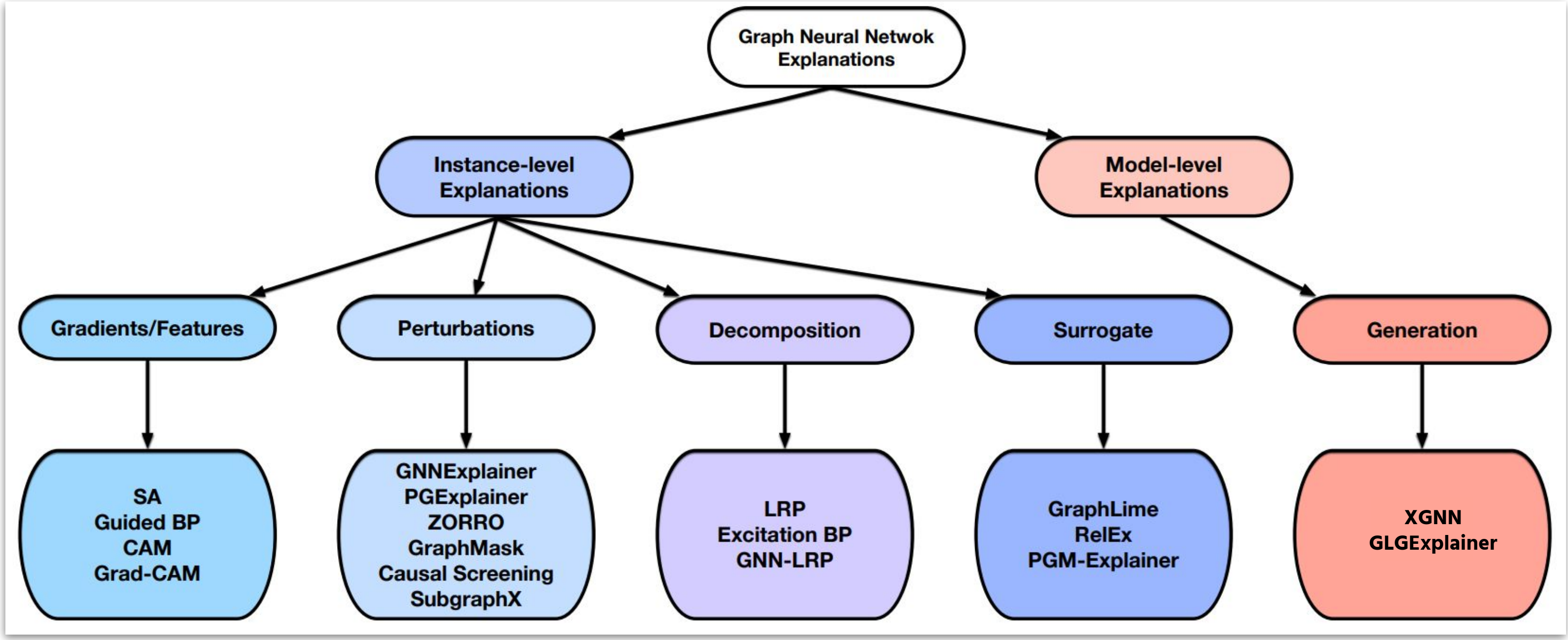


Image from Explainability in Graph Neural Networks: A Taxonomic Survey. H. Yuan et al., 2022 and then revised

XAI4GNNs
Taxonomy

Surveys on XAI4GNNs (not complete):

- Evaluating Explainability for Graph Neural Networks. C. Agarwal et al. 2022
- Probing GNN Explainers: A rigorous Theoretical and Empirical Analysis of GNN Explanation Methods. C. Agarwal et al., 2022
- Explainability in Graph Neural Networks: A Taxonomic Survey, H. Yuan et al., 2022
- Explaining the Explainers in Graph Neural Networks: a Comparative Study, A. Longa, S. Azzolin et al., 2022
- Towards Robust Fidelity for Evaluating Explainability of Graph Neural Networks, Zheng et al., 2023

XAI4GNNs
Limitations

Current limitations:

- Too often random baselines surpass XAI tools
- Non-robustness of XAI tools
- OOD issue during perturbations
- Maybe excessive focus on final metrics, with little attention to whether the explanations actually help the human/the debugging (*personal take*)

XAI4GNNs
Global Explanations

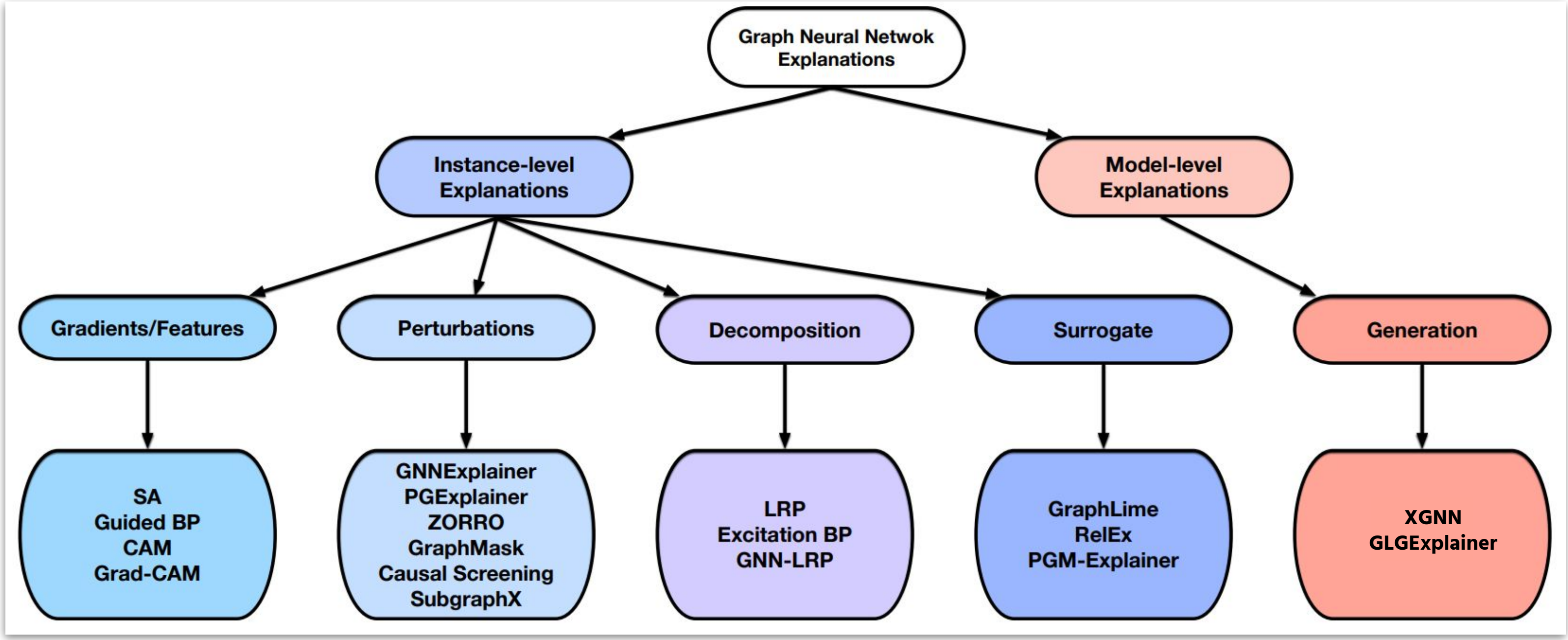


Image from Explainability in Graph Neural Networks: A Taxonomic Survey. H. Yuan et al., 2022 and then revised

XAI4GNNs
Global Explanations

Global (or Model-level) Explainers capture the behaviour of the model as a whole, abstracting individual noisy local explanations

Why global explanations?

Global Explainers are seldom studied + mining local explanations is hard:

1. 1+ for every input sample
2. Often noisy
3. Difficult quality evaluation [1,2]

1. When Comparing to Ground Truth is Wrong: On Evaluating GNN Explanation Methods. L. Faber et al., 2021
2. On Consistency in Graph Neural Network Interpretation. T. Zhao et al., 2022

XAI4GNNs
GLGExplainer

The **Global Logic-based GNN Explainer** (GLGExplainer) extracts logic formulas expressed in terms of learned human-understandable concepts.

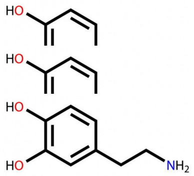
Logic formulas with size constraints can be easily understood by human experts.

XAI4GNNs
GLGExplainer

The **Global Logic-based GNN Explainer** (GLGExplainer) extracts logic formulas expressed in terms of learned human-understandable concepts.

Logic formulas with size constraints can be easily understood by human experts.

Dataset with Mutag/Non Mutag compounds



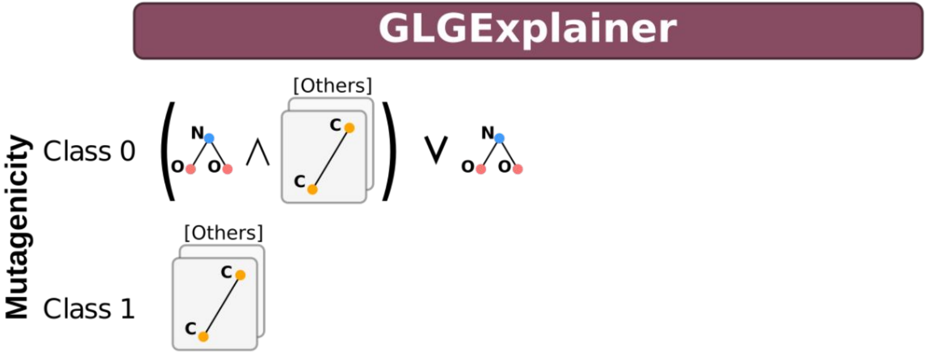


Image from Global Explainability of GNNs via Logic Combination of Learned Concepts. S. Azzolin et al., 2023

XAI4GNNs
GLGExplainer

GLGExplainer in short:

1. Extract local explanations with a local explainer
2. Run GLGExplainer over those local explanations
3. Inspect the generated logic formulas

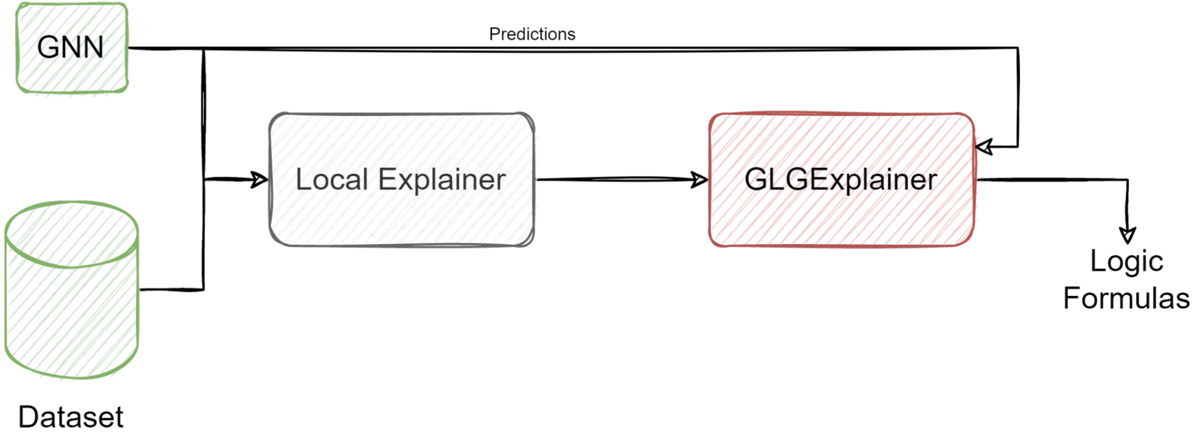


Image from Global Explainability of GNNs via Logic Combination of Learned Concepts. S. Azzolin et al., 2023

XAI4GNNs
GLGExplainer

GLGExplainer in short:

So, GLGExplainer is learning how to combine individual local explanations into a single logic-based formula

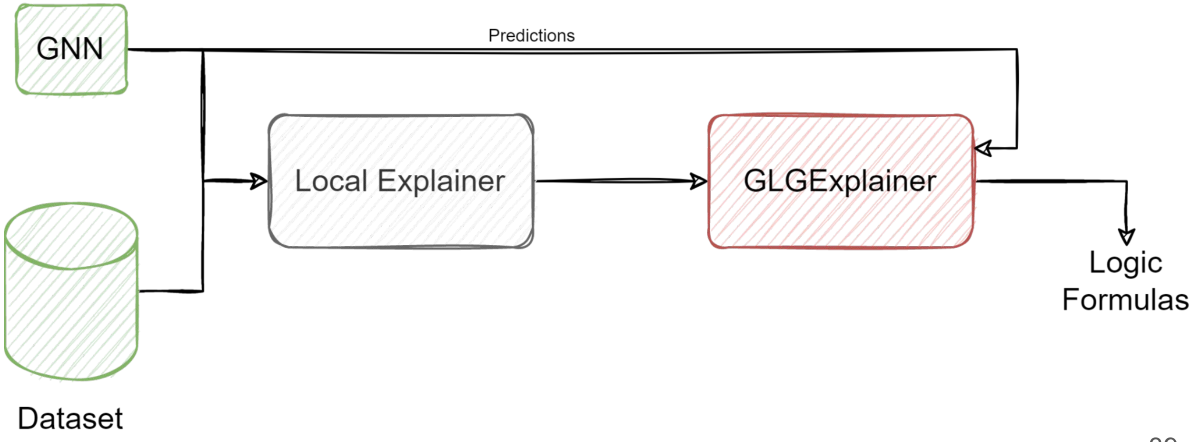
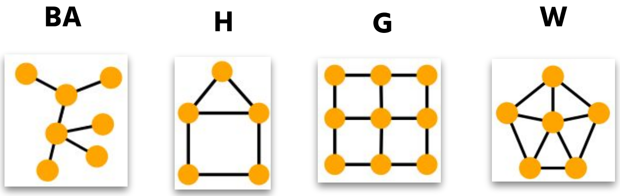


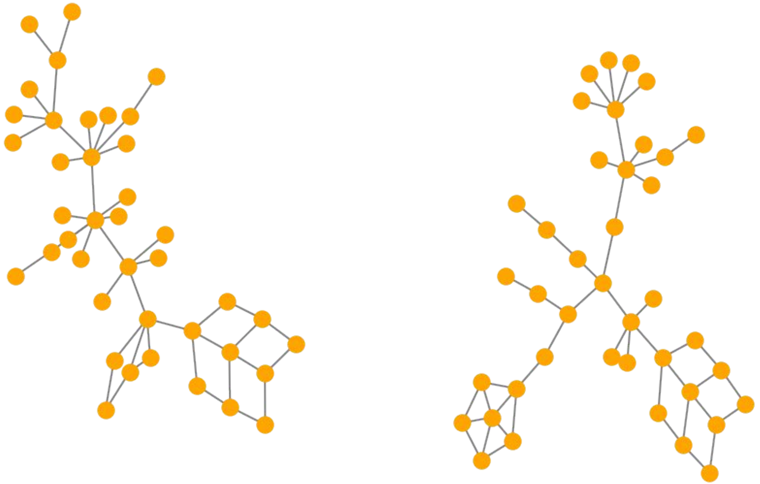
Image from Global Explainability of GNNs via Logic Combination of Learned Concepts. S. Azzolin et al., 2023

XAI4GNNs
GLGExplainer

BAMultiShapes dataset:

- Class0: $\emptyset \vee H \vee G \vee W \vee (H \wedge G \wedge W)$
- Class1: $(H \wedge G) \vee (W \wedge H) \vee (W \wedge G)$





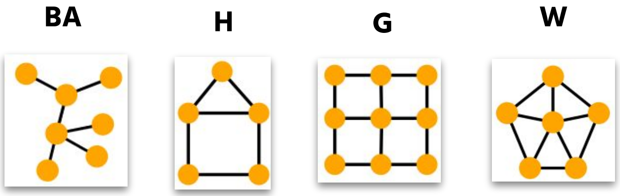
XAI4GNNs
GLGExplainer

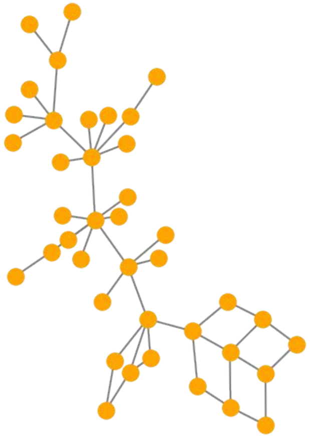
Very few instances and only in train/val data

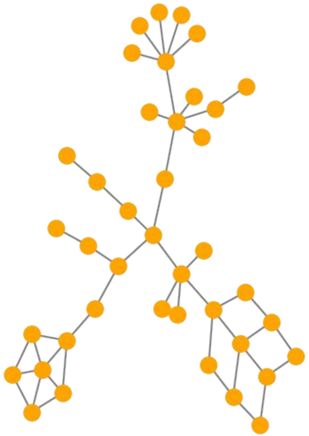
Split	BAMultiShapes
Train	0.94
Val	0.94
Test	0.99

BAMultiShapes dataset:

- Class0: $\emptyset \vee H \vee G \vee W \vee \boxed{(H \wedge G \wedge W)}$
- Class1: $(H \wedge G) \vee (W \wedge H) \vee (W \wedge G)$





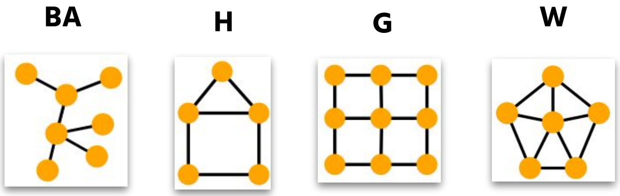


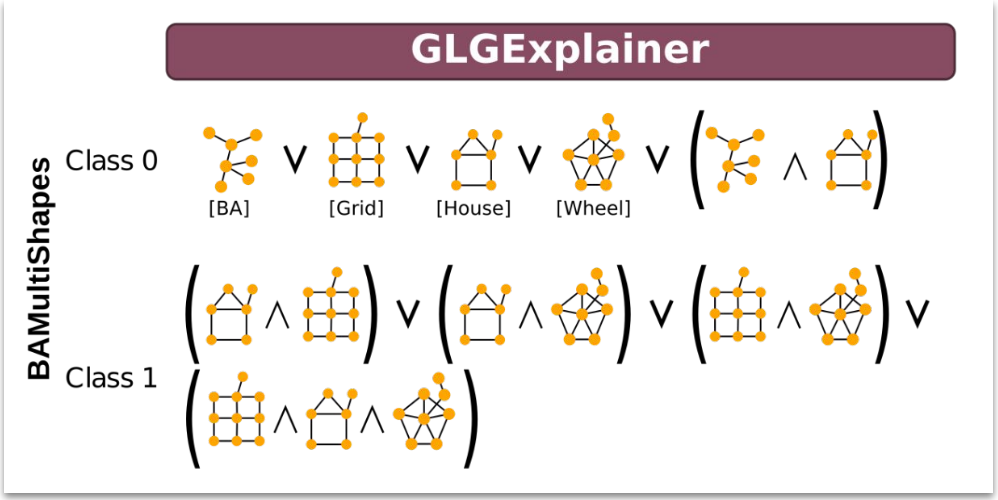
XAI4GNNs
GLGExplainer

Very few instances and only in train/val data

BAMultiShapes dataset:

- Class0: $\emptyset \vee H \vee G \vee W \vee \boxed{(H \wedge G \wedge W)}$
- Class1: $(H \wedge G) \vee (W \wedge H) \vee (W \wedge G)$





XAI4GNNs
GLGExplainer

Current limitations:

- Number of concepts must be defined apriori
- Assumes a nicely working local explainer

Conclusions

- Common shortcomings of standard deep learning models
- Premises and potentials of XAI
- Examples (with some code) of XAI tools for the graph domain
- Limitations of current XAI tools

Conclusions
What's Next

Tutorial in four parts (slides + Jupyter notebooks available):

- ~~**Part I:** November 2, Presenter: **GS**~~
  Goals: Motivations, Intro of basic concepts, definition of GNNs
- ~~**Part II:** November 9, Presenter: **AL**~~
  Goals: Implementation of GNNs: How to implement a full GNN pipeline in PyTorch Geometric.
- ~~**Part III:** November 16, Presenter: **SA**~~
  Goals: Explainability of GNNs: How to shed (a bit of) light into the black box
- **Part IV:** November 23, Presenter: **FF**
  Goals: Heterogeneity in GNNs: How can GNNs effectively model and incorporate a diversity of nodes and edges with different types.

E.O.F.