Assessing Explanations of Graph Neural Networks for Predicting Effects of Gene Perturbations
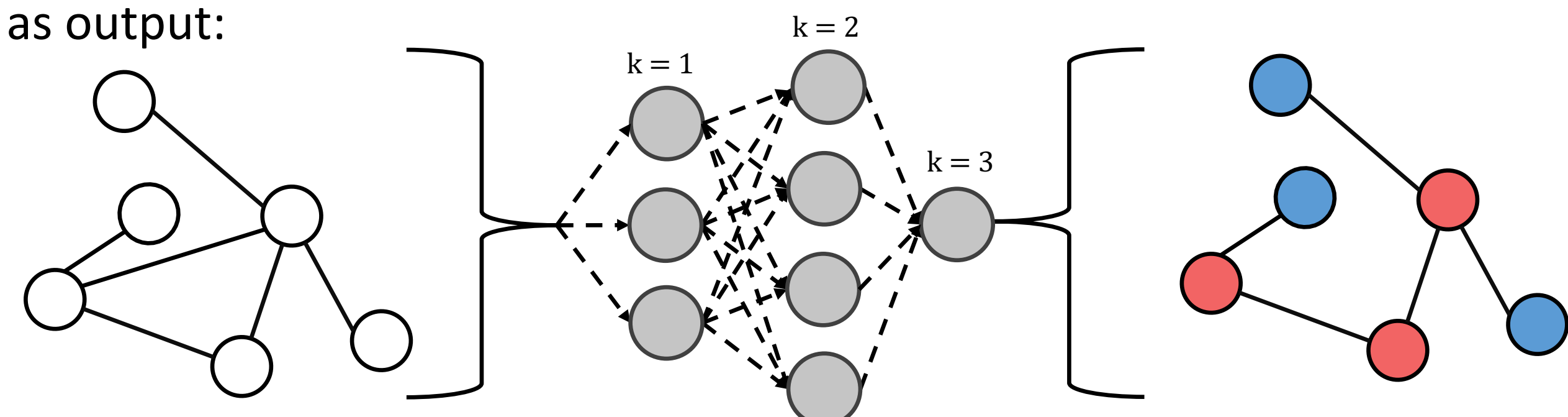
Alyssa Anastasi [1,2], Yiyang (Leo) Jin [2], Mark Craven [2]

[1] University of Illinois at Urbana-Champaign, [2] Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison

Background

- A graph neural network (GNN) is a layered architecture that takes as input a graph, along with node and edge features and produces node classifications as output:

$$m_{u,v}^{(k)} = M^{(k)}(h_u^{(k-1)}, h_v^{(k-1)}, r_{u,v}) \qquad h_v^{(k)} = A^{(k)}(h_v^{(k-1)}, \{m_{u,v}^{(k)}: u \in N(v)\})$$

MOTIVATION: Due to the complexity of GNNs, it is difficult to interpret how a model makes predictions. The ability to interpret and understand GNNs is important in order to:
- Allow the possibility of scientific discovery through the predictions
- Develop trust in the model's predictions
- Debug and improve the model's performance
- Explanation algorithms can be used to obtain an edge and a node mask that represent the subgraph of the most important edges and node features in making the prediction for a specified target node:
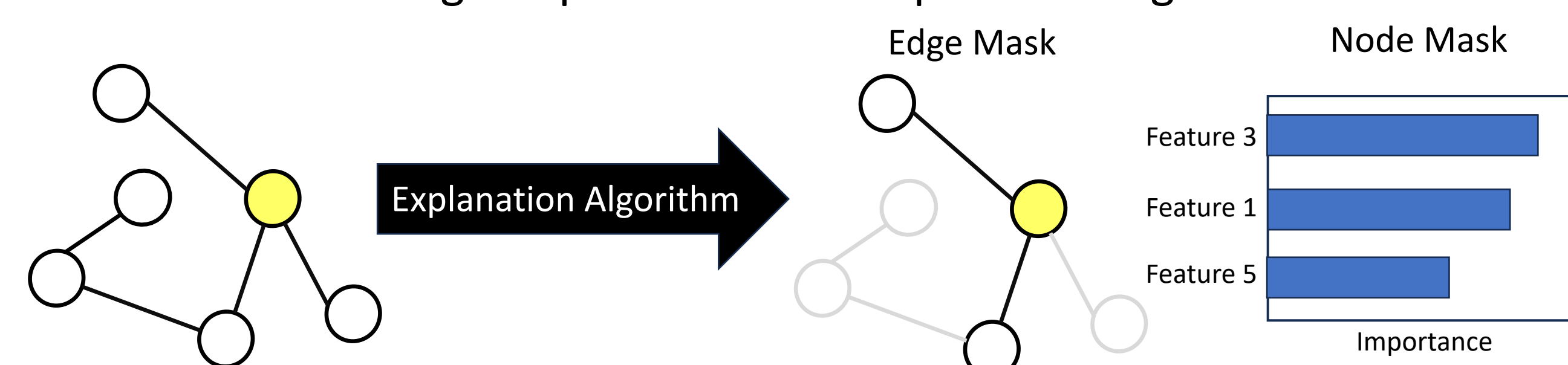
GOAL: Using a GNN trained to predict the effect of gene perturbations on a cellular phenotype, we assess various explanation methods and their produced explanations on genes of interest

Data and Methods

- We analyze a GNN trained to predict the effect of gene perturbations on the phenotype of cholesterol homeostasis
- Data Sources:
  - STRING Protein-Protein Interaction Network [1]
  - Embedded representation of Gene Ontology annotations [2]
  - RNA/protein abundances and subcellular localizations from the Human Protein Atlas [3]
- Graph Size: 18,384 nodes, 165 node features, and 1,991,832 edges
- Training Data: Gene perturbations measured through CRISPR knockouts of over 19,000 genes in HeLa cells [4, 5]
- We generate explanations for set of 10 gene instances

Explainer Methods	Explanation Types		Node Mask	Edge Mask
	Model computes losses using model output	Phenomenon computes losses using target output		
GNN Explainer [6]	✅	✅	✅	✅
PG Explainer [7]	❌	✅	❌	✅
Graph Mask [8]	✅	✅	✅	✅

Results

EXPLANATIONS

Example explanations for gene EXOC1 with edge/node masks limited to size 10

GNN Explainer

Feature importance for top 10 features

PG Explainer


Graph Mask
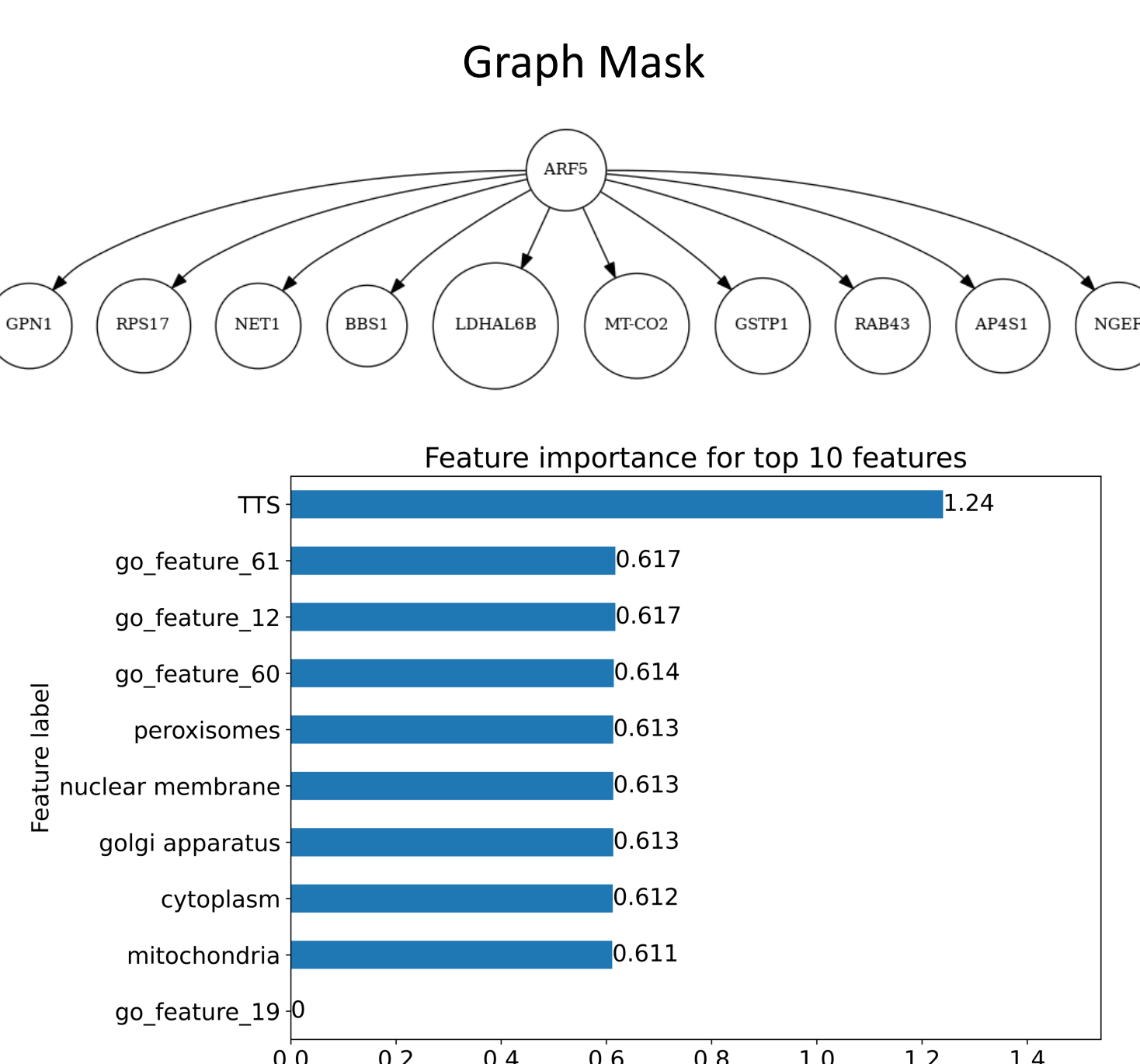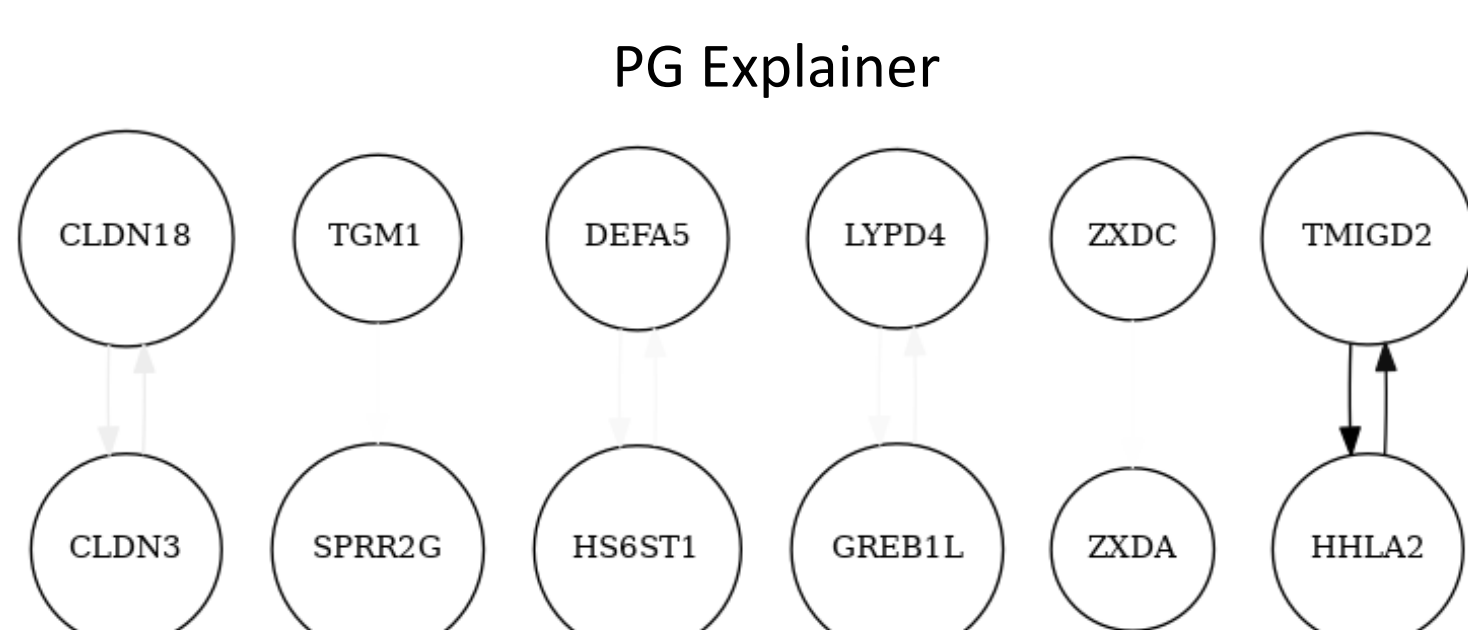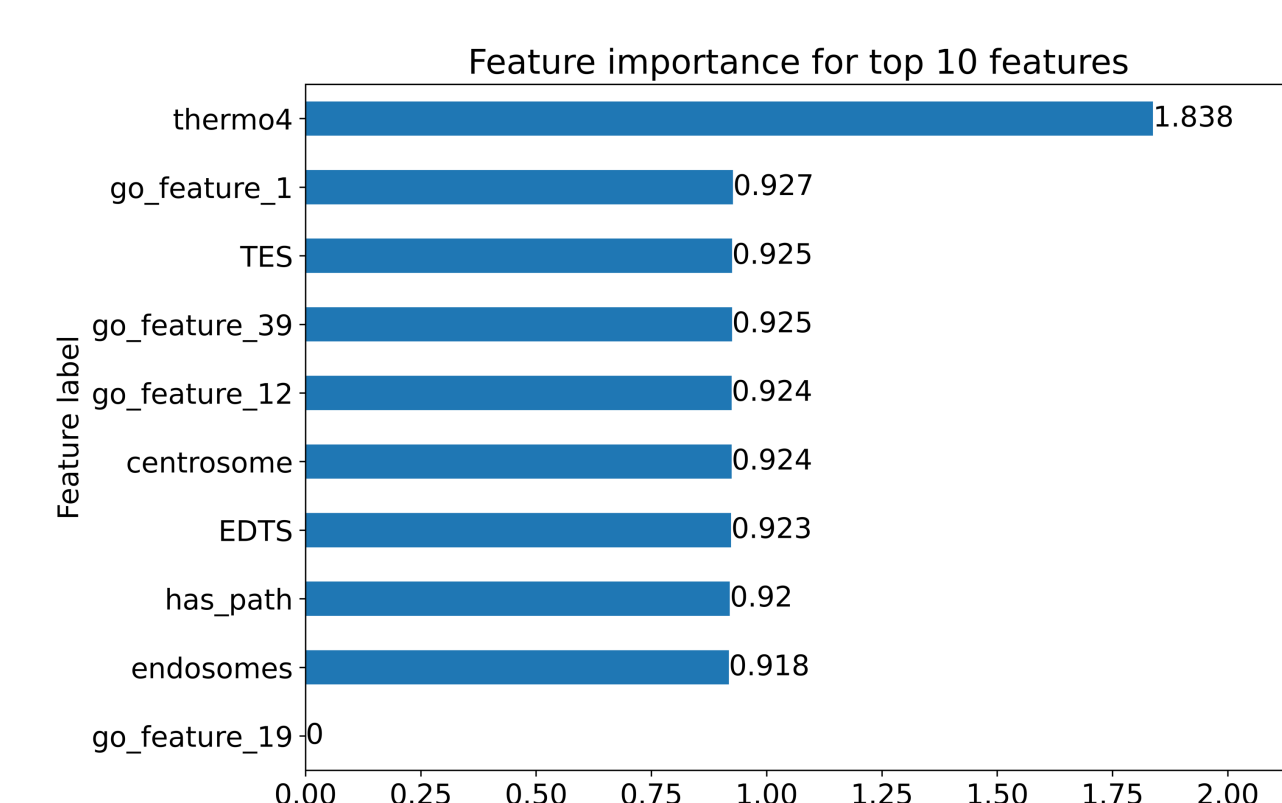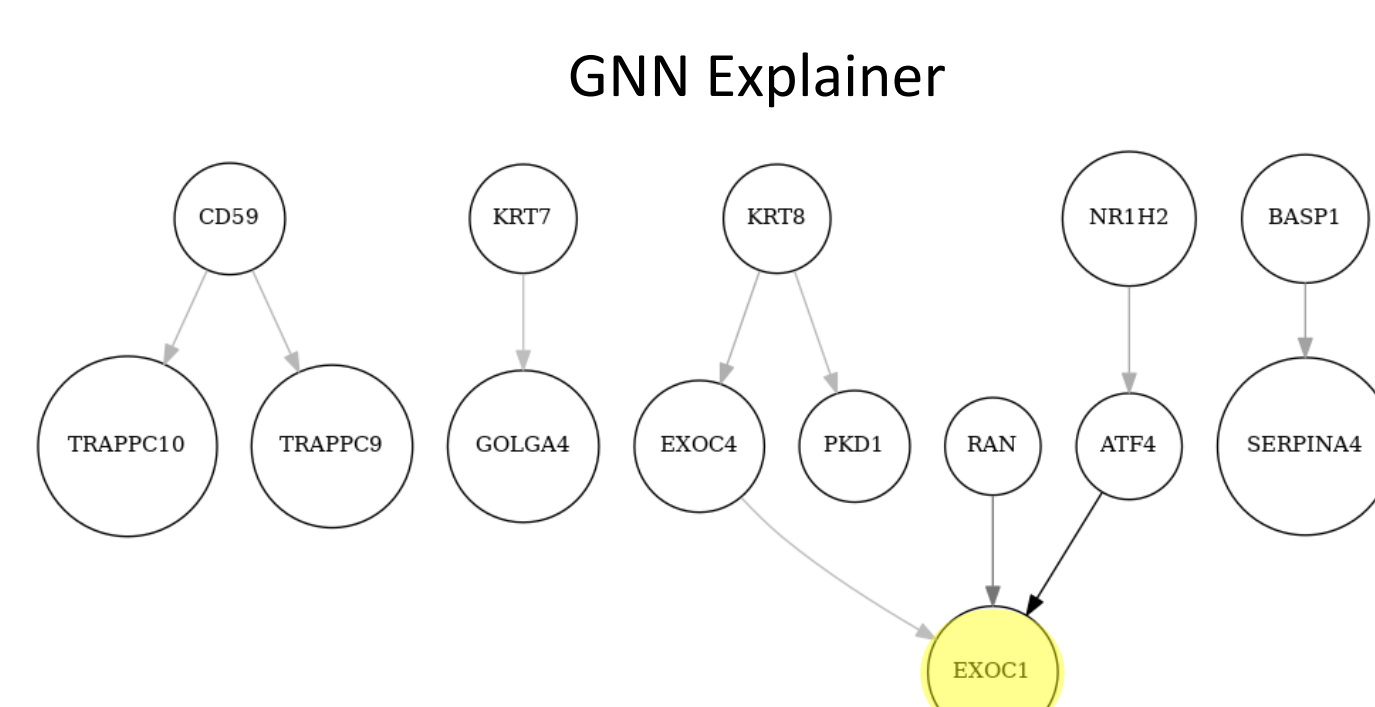
Feature importance for top 10 features

FIDELITY

How does the fidelity of an explanation of the GNN vary as the size of the explanation increases?

- Necessity [9] : $1 - \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(\hat{y}_i^{G_{C/S}} = \hat{y}_i)$
  - An explanation is necessary when the model prediction changes if the subgraph is removed
- Sufficiency [9] : $\frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(\hat{y}_i^{G_S} = \hat{y}_i)$ [3]
  - An explanation is sufficient when the subgraph on its own replicates the model prediction
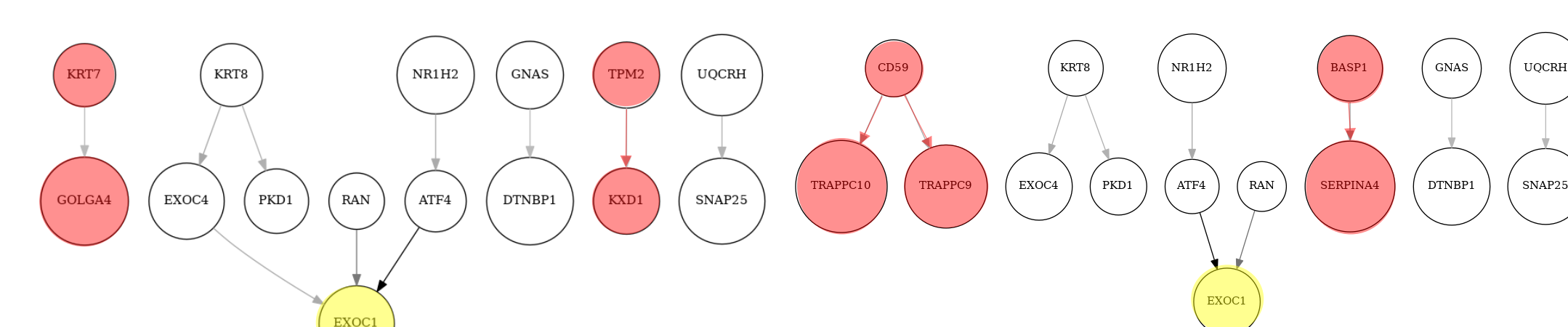
Explainer	Necessity	Sufficiency
GNN Explainer		
PG Explainer		
Graph Mask		
Random Baseline		
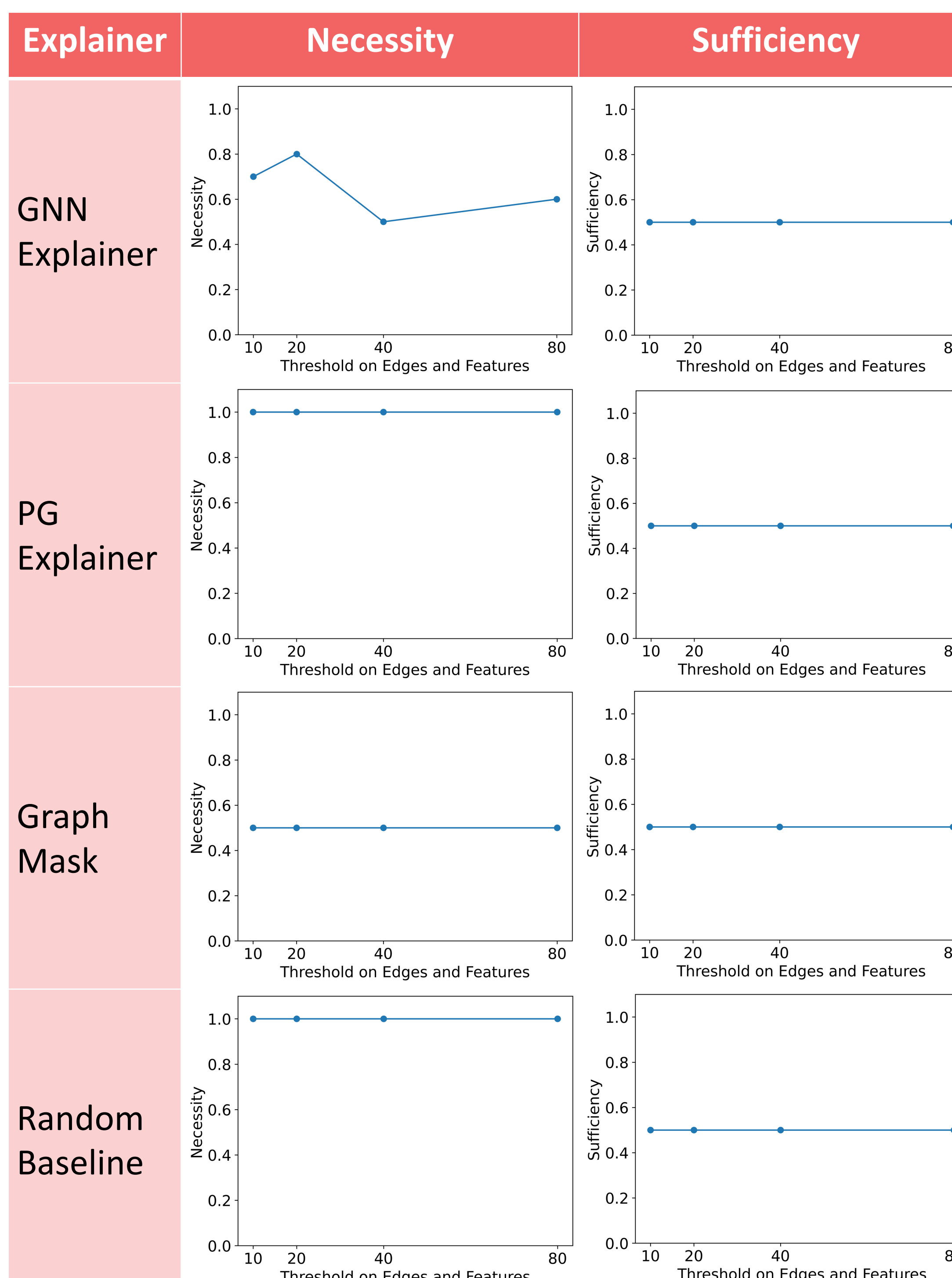
STABILITY

How does the randomness in the explanation algorithms affect the variability of the explanations?

Average Edge Similarity among Instances

Average Feature Similarity among Instances
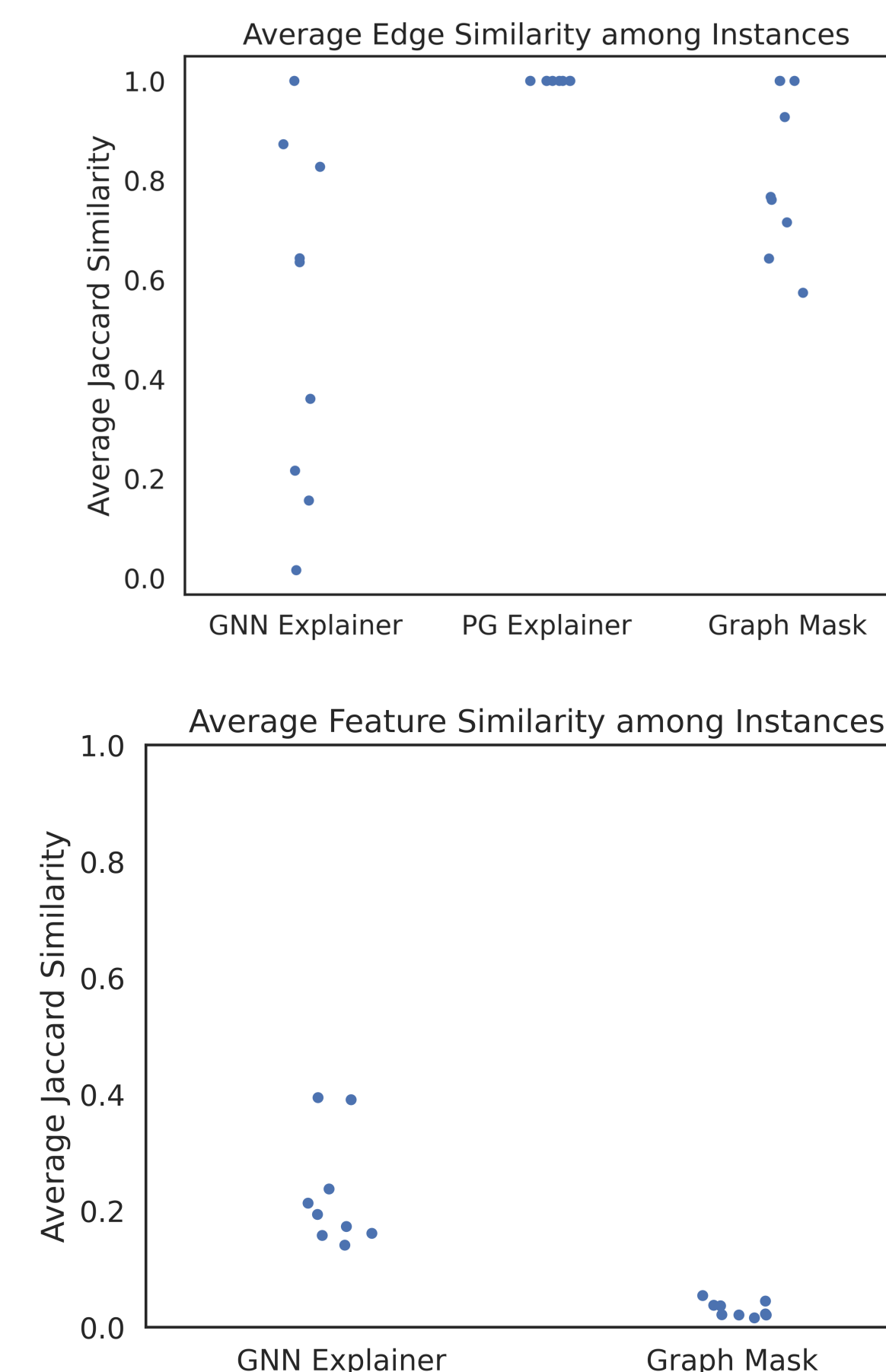
Varying Explanations using GNN Explainer on EXOC1:

RUNTIME

How does the runtime of the explainers compare?

Average Runtime per Instance
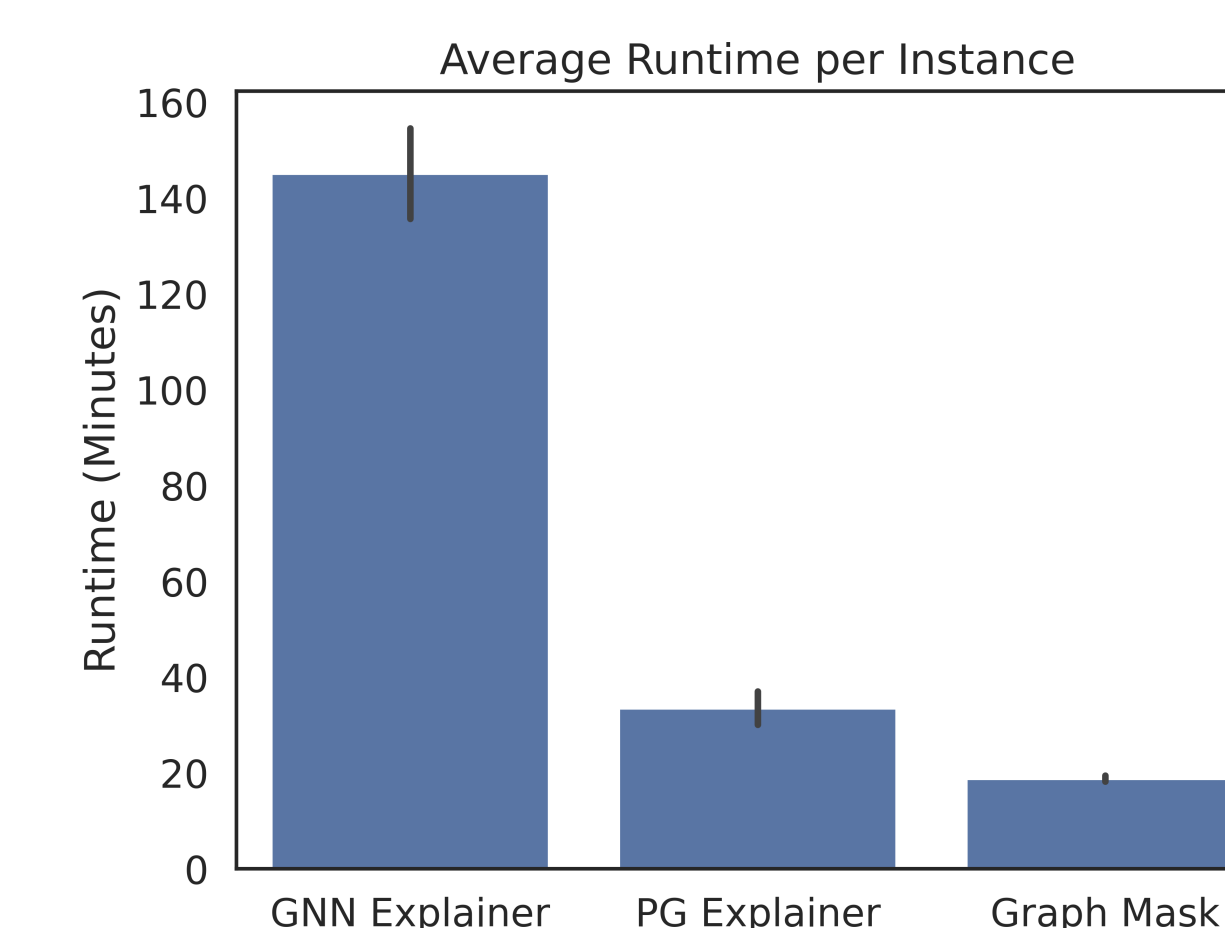
Conclusions

- Explanation methods can identify important aspects of a model:
  - In our network, the structure of the graph is not as important as the node features
- The methods vary in their utility:
  - PG Explainer is incapable of explaining node features
  - Graph Mask lacks stability
- Limiting edges and node features used in an explanation can reduce its fidelity
- The explainers are time consuming which can limit their use in some settings
- Current implementations of these methods lack in-depth documentation and maturity which makes them less trustworthy

References

[1] von Mering, C., et al. (2003) STRING: a database of predicted functional associations between proteins. *Nucleic acids research* 31.1: 258-261.

[2] The Gene Ontology Consortium. (2023) The Gene Ontology knowledgebase in 2023. *Genetics.* 224(1):iyad031

[3] Sjöstedt E, et al. (2020) An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science.*

[4] Xiao, J., et al. (2021) POST1/C12ORF49 regulates the SREBP pathway by promoting site-1 protease maturation. *Protein & cell* 12.4:279-296.

[5] Van den Boomen, D.J.H., et al. (2020) A trimeric Rab7 GEF controls NPC1-dependent lysosomal cholesterol export. *Nature communications* 11.1:5559.

[6] Ying, R., et al. (2019). GNNExplainer: Generating Explanations for Graph Neural Networks. *arXiv [Cs.LG]:* 1903.03894.

[7] Luo, D., et al. (2020). Parameterized Explainer for Graph Neural Network. *arXiv [Cs.LG]:* 2011.04573.

[8] Schlichtkrull, M. S., et al.. (2022). Interpreting Graph Neural Networks for NLP With Differentiable Edge Masking. *arXiv [Cs.CL]:* 2010.00577.

[9] Amara, K., et al. (2024). GraphFramEx: Towards Systematic Evaluation of Explainability Methods for Graph Neural Networks. *arXiv [Cs.LG]:* 2206.09677.