

# Modified GNN-SubNet: Leveraging local versus global Graph Neural Network explanations for disease subnetwork detection

Student: Elena-Oana Milchi (E.O.Milchi@student.tudelft.nl)

Thesis committee: Dr. Megha Khosla, Dr. Jana Weber, Dr. Thomas Abeel

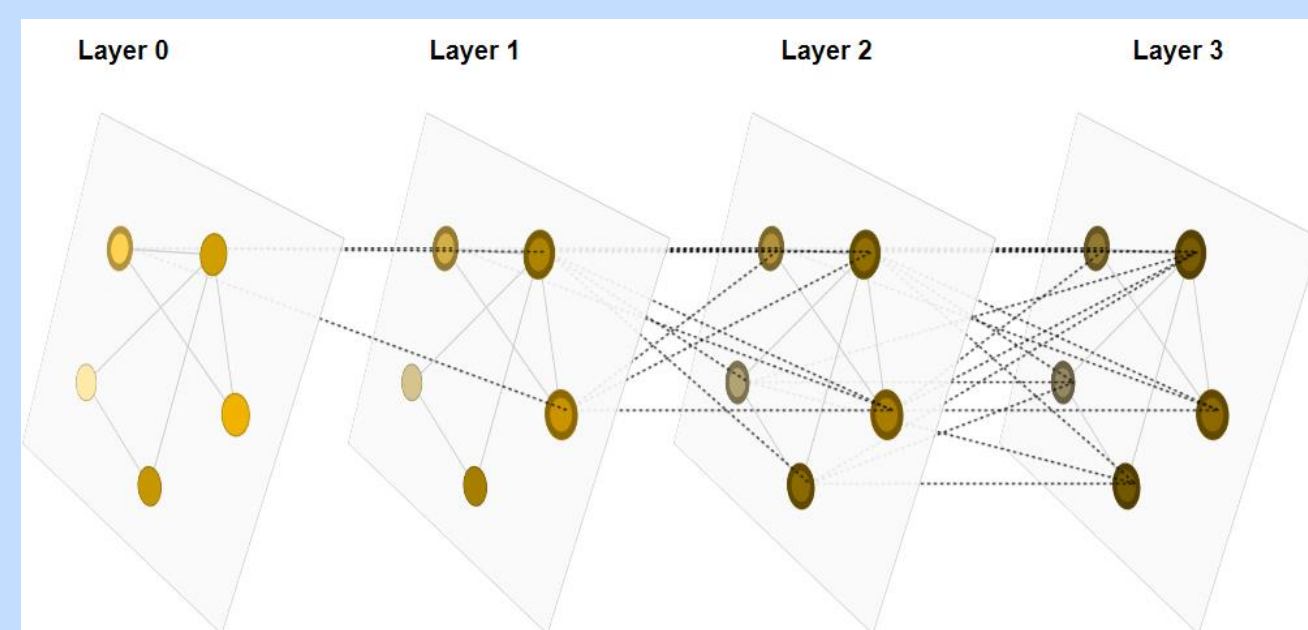


## Background

- Protein Interaction aids in locomotion and metabolism regulation
- Malformed proteins can generate disease phenotypes
  - Analyze PPI Networks to identify disease subnetworks

### GNNs

- Machine learning tools useful in bioinformatics
- Operate on graph data (node features)
- Inherently black-box



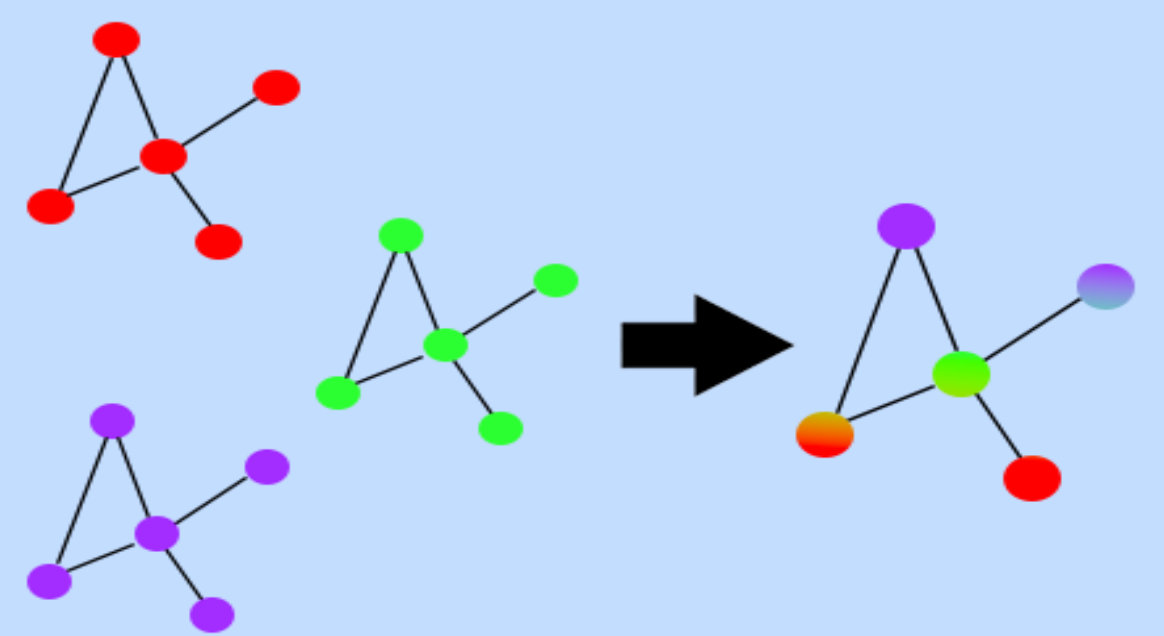
Source: Sanchez-Lengeling, Reif, Pearce, and Wiltchko (2021)

### Explainers for GNNs

- Provide insight on GNNs' predictions:
  - How do certain inputs and outputs correspond?
  - How is data represented in the network?
- Local explanation:** explain how a GNN makes a prediction for a single dataset entry
- Global explanation:** explain a GNN's predictions for an entire dataset

### GNN-SubNet

- Uses a GNN explainer for disease subnetwork detection
- Uses a mask for all nodes optimized with gradient descent => global explanation
- Repeatedly samples input graphs from the dataset



GNN-SubNet optimizes a node mask (right) which explains for each node how important it was in the GNN's predictions of the input graphs (left)

## Research Questions

How does the global-level explanation of a GNN obtained by GNN-SubNet compare with an aggregation of local-level explanations of the same model?

- How can GNN-SubNet be modified to aggregate local-level explanations of a GNN into a global explanation?
- How does GNN-SubNet compare with its proposed modification in terms of the metrics devised to evaluate explainers for GNNs?

## Methods

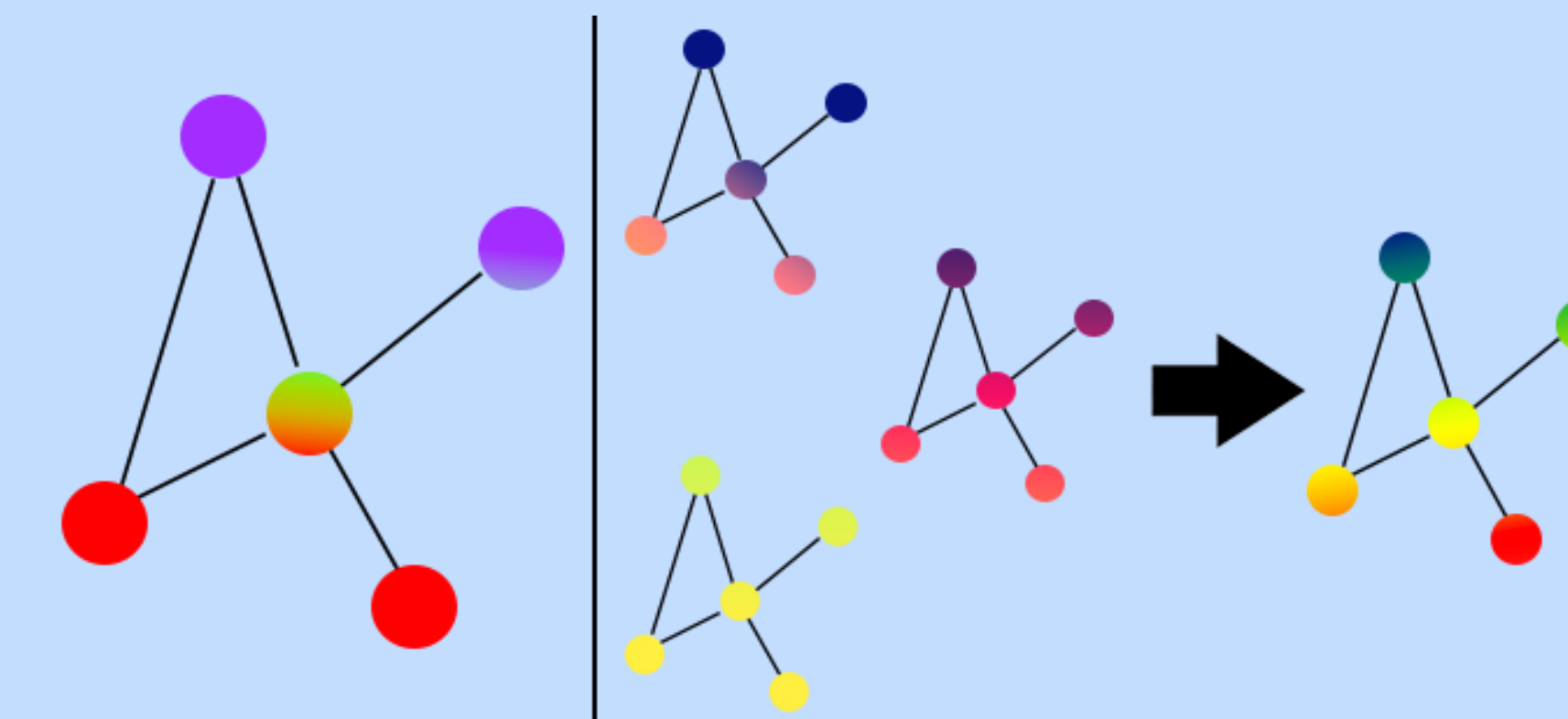
### RQ1: Modifying GNN-SubNet

- GNN-SubNet optimizes a node mask via Gradient Descent:
  - The optimization is done on a restricted sample of the input data (reinitialized every 50 epochs)
  - To obtain a local explanation – remove the sampling scheme

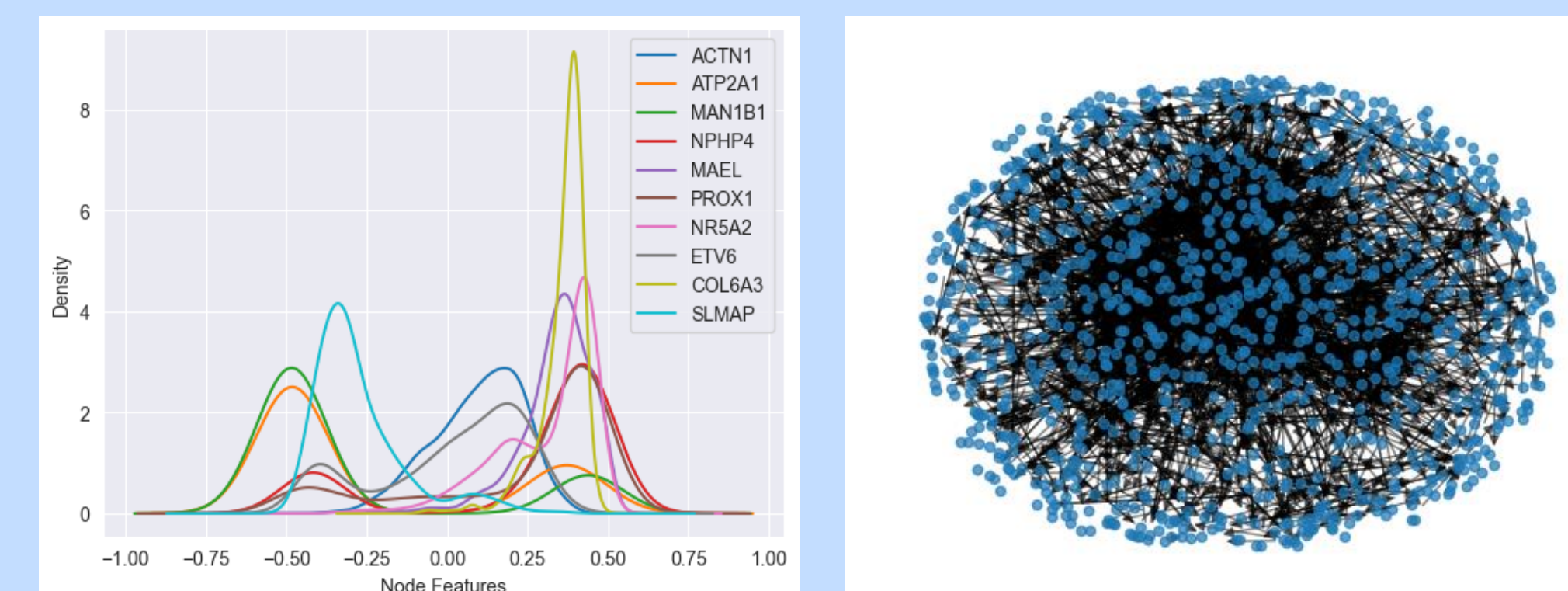
- Then aggregate all local explanations (node masks) per node
  - Mean Aggregation:** takes the **mean** value
  - Median Aggregation:** takes the **median** value

### RQ2: Comparing GNN-SubNet with its modifications

- Input data:** Protein methylation and mRNA features from The Cancer Genome Atlas (KIRC)
- Experimental pipeline:**
  - Run GNN-SubNet and its two modifications on the TCGA dataset 10 times per explainer
  - Obtain most relevant disease subnetworks
  - Obtain metric scores:
    - RDT-Fidelity:** can explanations approximate the model's behavior?
    - Sparsity:** Are the explanations non trivial?
    - Validity+:** does averaging the "important" features change the GNN's prediction?
    - Validity-:** does averaging the "unimportant" features maintain the GNN's prediction?



While GNN-SubNet optimizes a node feature mask for all dataset entries, the proposed modification does this for each graph input

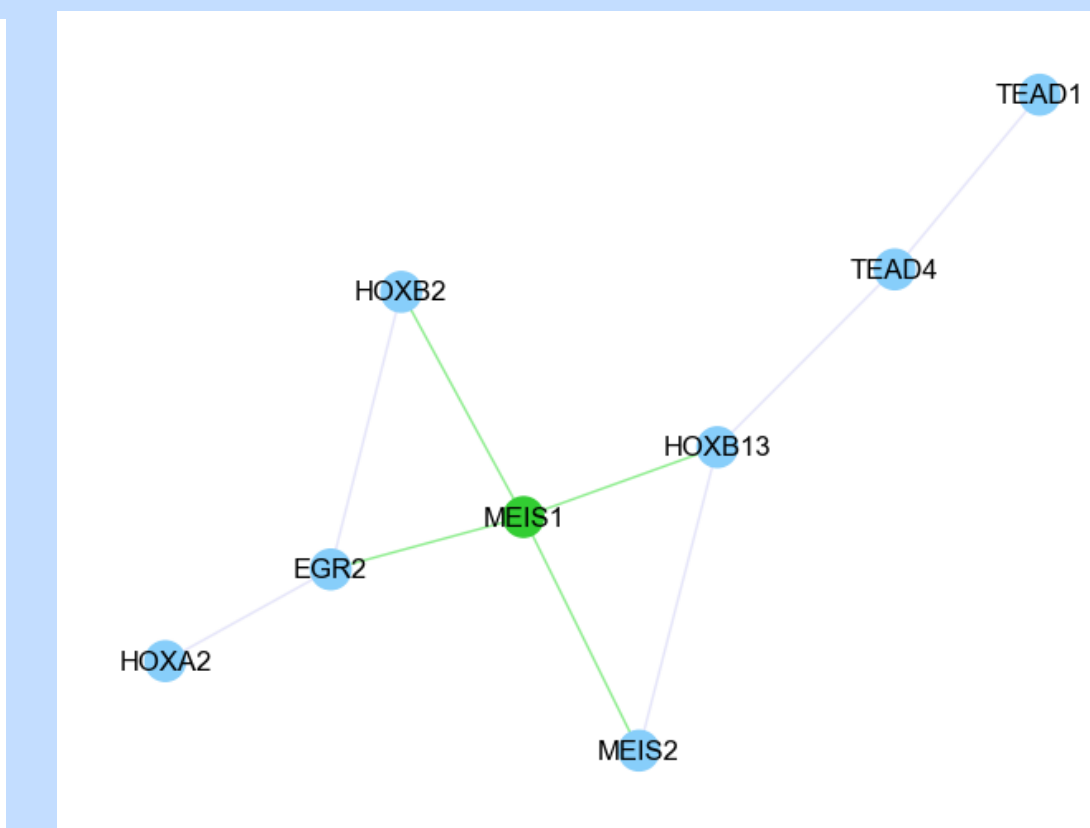
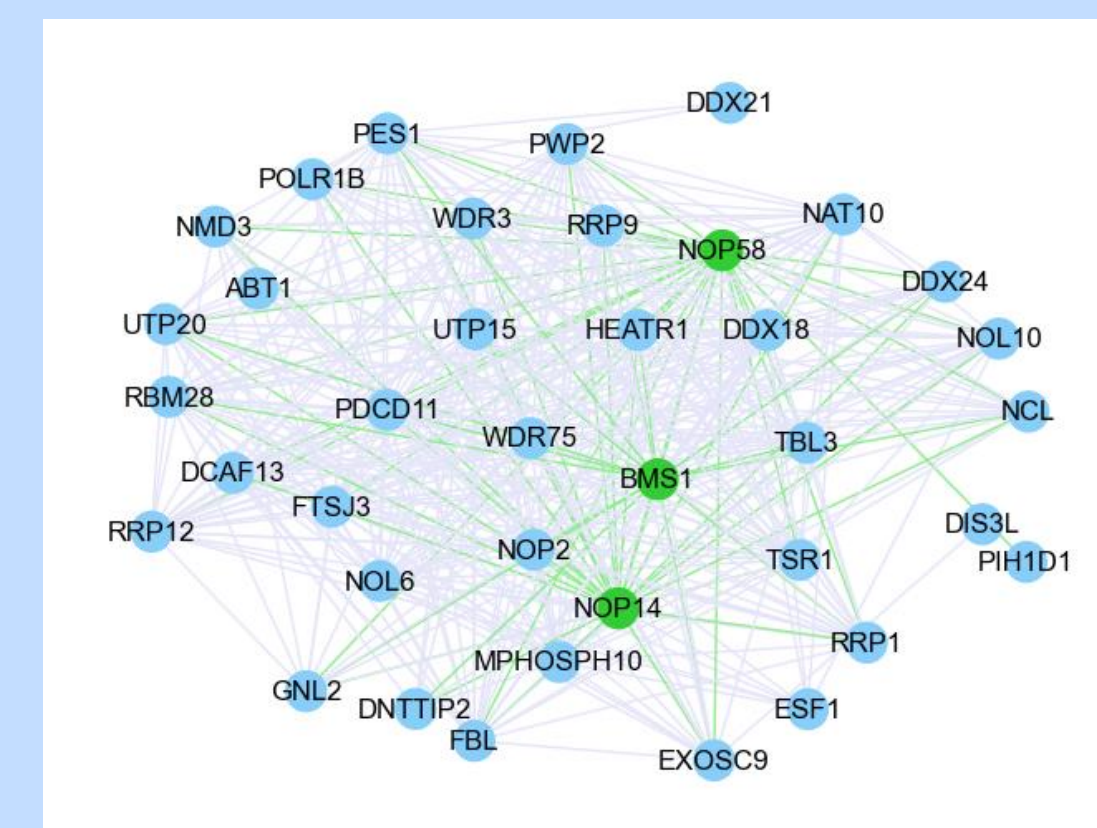


A visualization of the TCGA dataset. On the left, the distribution of node feature values for the first 10 proteins. On the right, a visualization of (part of) the PPI where one node is a protein and one edge is a protein interaction

## Results

### RQ1: Disease subnetwork analysis

- Most frequent among all explainers
- High expression in cancer tumours:
  - NOP14
  - NOP58



- Reported by authors of GNN-SubNet
- Found only by Mean & Median Aggregation in this study
- Relevant for cancer detection:
  - HOXB13 (& interaction with MEIS1, MEIS2)
  - EGR2 (for RCC)

### RQ2: Comparison in terms of metrics

Explainer	GNN Acc.	Mean Fidelity	Mean Sparsity	Thres. Value	Mean Validity+	Mean Validity-
GNN-SubNet	75.0	0.855625	<b>0.042664</b>	30.0	0.25875	0.73625
Mean Aggregation	75.75	<b>0.861375</b>	0.019166	50.0	<b>0.47125</b>	0.85375
				50.0	0.32375	<b>0.8150</b>
Median Aggregation	73.0	0.837375	0.038665	30.0	<b>0.39375</b>	0.80750
				50.0	0.45500	0.83375

#### RDT Fidelity

- Mean Aggregation performs best
- Median Aggregation performs worst

#### Sparsity:

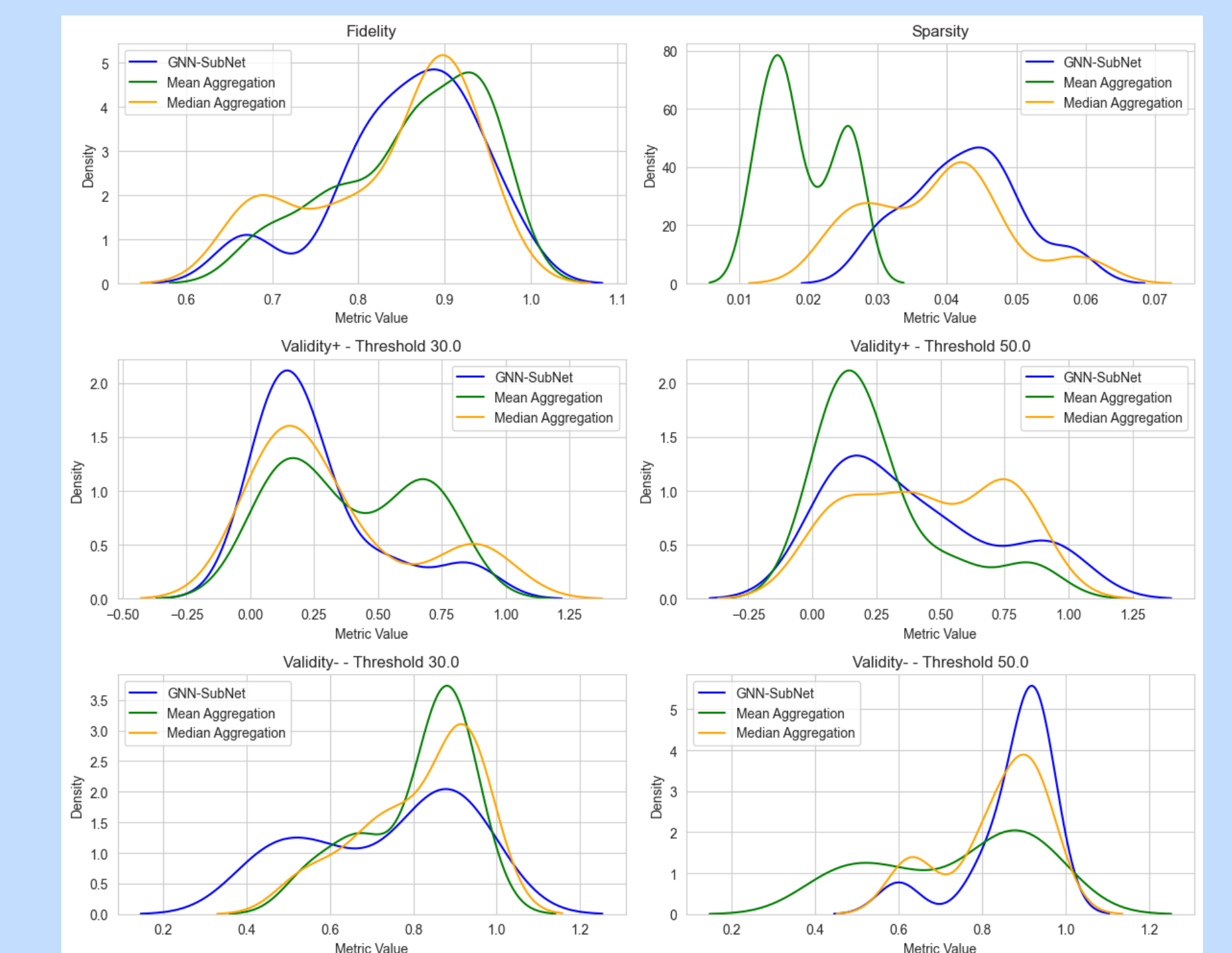
- All explainers obtain dense explanations
- Mean Aggregation performs the worst
- Median Aggregation performs better, but not as good as GNN-SubNet

### Validity+ and Validity-

- No clear winner in terms of Validity+
- Mean Aggregation performs best in terms of Validity-

## Limitations

- GNN-SubNet randomly initializes the node mask in gradient descent
  - This leads to different optimizations and different results per run



A visualization of the metric scores for all explainers over ten runs. Note the skewed distributions and high variance.

## Conclusions

- GNN-SubNet and its two modifications identify subnetworks associated with cancer
- Mean Aggregation shows best RDT Fidelity and Validity-
- No explainer obtains sparse explanations
- Further research into explainable graph models for disease subnetwork detection is needed

## References

- B. Sanchez-Lengeling, E. Reif, A. Pearce, and A. Wiltchko, "A Gentle Introduction to Graph Neural Networks", Distill, vol. 6, no. 8, 10.23915/distill.00033, Aug. 2021, issn: 2476-0757. doi: 10.23915/distill.00033.
- B. Pfeifer, A. Saranti, and A. Holzinger, "GNN-SubNet: Disease subnetwork detection with explainable graph neural networks", Bioinformatics, vol. 38, no. Supplement\_2, pp. ii120-ii126, Sep. 2022, issn: 1367-4803, 1367-4811. doi: 10.1093/bioinformatics/btac478.