Link Explanation for Heterogeneous Graphs

Abhijit Gupta, advised by Rex Ying
12/09/2022

GNN Explainability

- Explainability builds trust, promotes fairness, and can improve human-in-the-loop performance

Multiple Tasks

- Why is an item recommended to a user? → Explain Link Prediction
- Why is the molecule mutagenic? → Explain Graph Classification
- Why is the user classified as fraudulent → Explain Node Classification

GNN Explainability

- Explainability builds trust, promotes fairness, and can improve human-in-the-loop performance

Multiple Tasks

- **Why is an item recommended to a user? → Explain Link Prediction**
- Why is the molecule mutagenic? → Explain Graph Classification
- Why is the user classified as fraudulent → Explain Node Classification

GNN Explainability

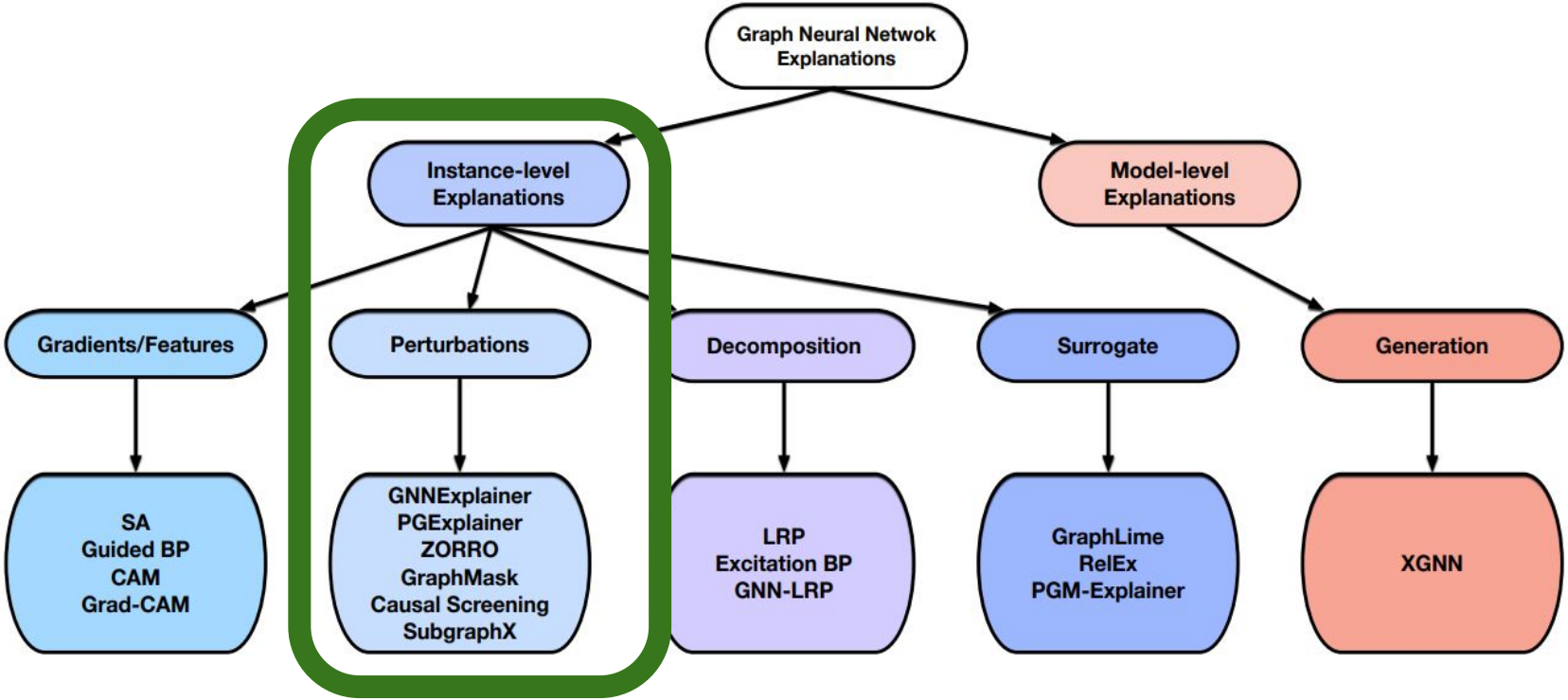
- Explainability builds trust, promotes fairness, and can improve human-in-the-loop performance

Multiple Tasks

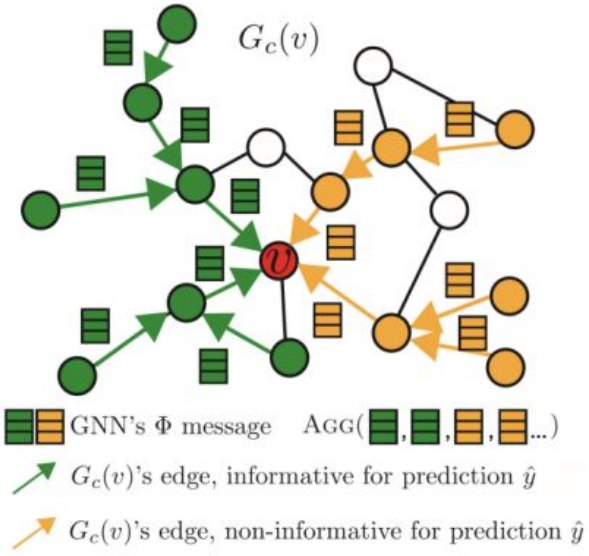
Heterogeneous Graph Explanation

- **Why is an item recommended to a user? → Explain Link Prediction**
- Why is the molecule mutagenic? → Explain Graph Classification
- Why is the user classified as fraudulent → Explain Node Classification

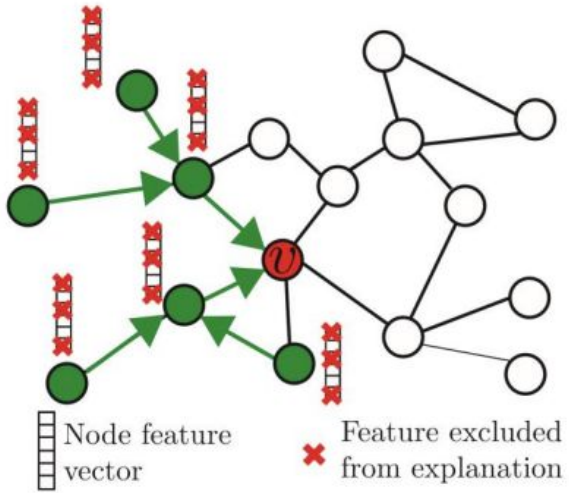
GNN Explainability



Types of Explanations



Structural explanation



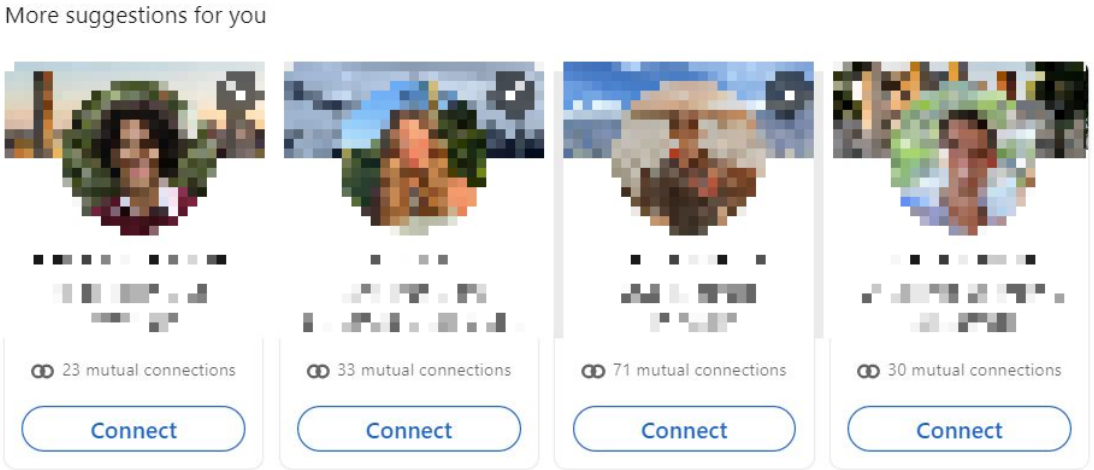
Feature explanation

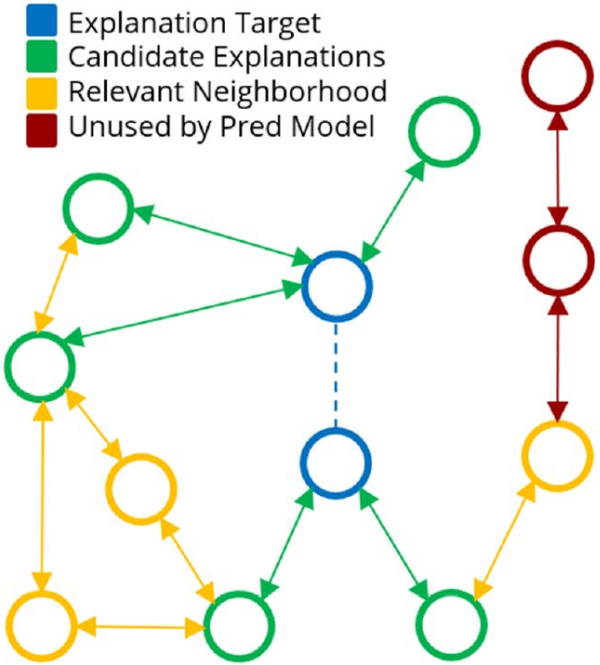
Problem Statement

- Explaining Link Prediction
  - Positive edges only
- Support Heterogeneous Graphs
- Instance-level Perturbation methods
- Focus on Structural explanation, Model explanation

Rethinking Explanation Format

- Explanations are restricted to **immediate neighbors** for increased interpretability real world use cases.





GNNExplainer

- Explain by Mutual Information (MI):

Maximize MI between **label** and **explanation**

$$\max_{G_S} MI\big(Y; (A_S, X_S)\big) = H(Y) - H(Y|A = A_S, X = X_S^F)$$

- Use continuous relaxation, optimize the expected adjacency matrix A_S

- **Modifications**: Do not optimize X_S, only optimize 1-hop neighborhood in A_S

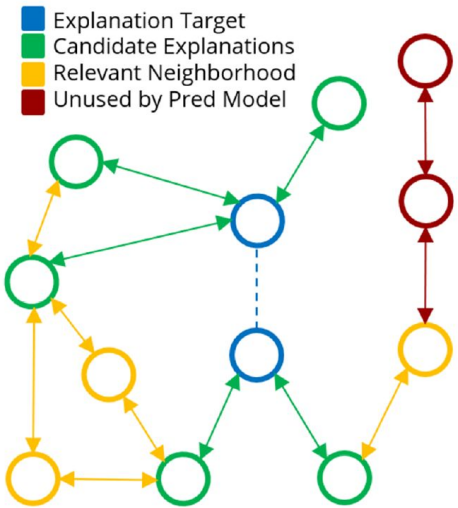
SubgraphX

- Uses **Monte Carlo Tree Search (MCTS)** and **Shapley values** to find subgraph explanations.

Output including explanation

$$\phi(\mathcal{G}_i) = \frac{1}{T}\sum_{t=1}^{T}(f(S_i \cup \{\mathcal{G}_i\}) - f(S_i))$$

Output excluding explanation



- **Modifications**: Remove MCTS component, reduce T from 100 to 5 to improve inference time.

Evaluation Metrics

- Focus on explaining model outputs, not necessarily phenomenon

- Measure fidelity for varying sparsity
  - **Necessary** and **sufficient** explanations, Characterization measures both

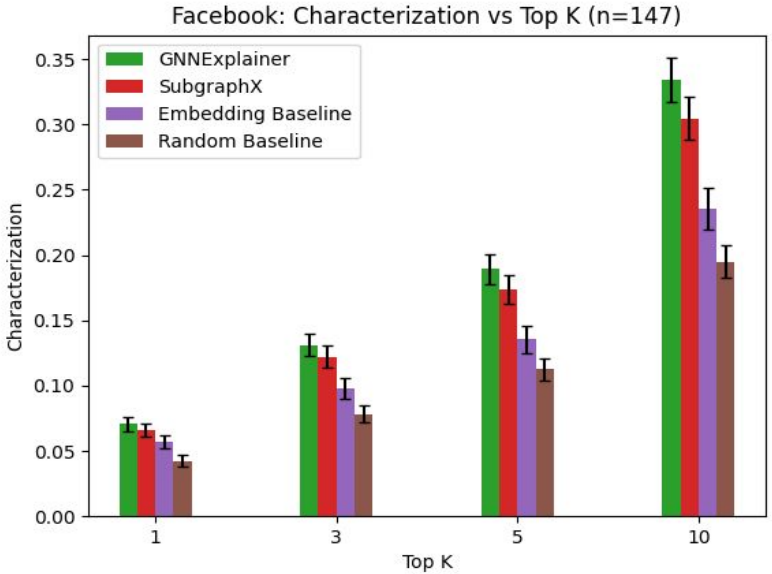
$$fid+^{prob} = \frac{1}{N}\sum_{i=1}^{N}(f(G_C)_{y_i} - f(G_{C\setminus S})_{y_i})$$

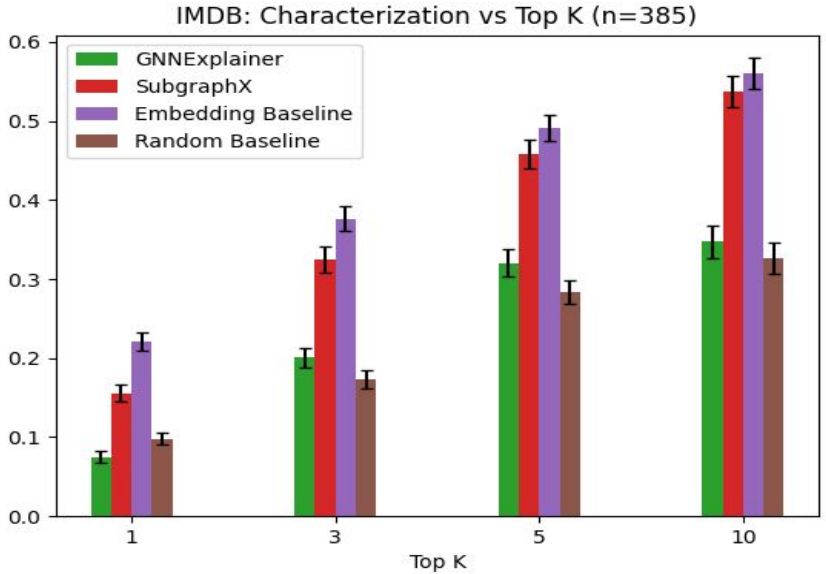
$$fid-^{prob} = \frac{1}{N}\sum_{i=1}^{N}(f(G_C)_{y_i} - f(G_S)_{y_i})$$

$$charact = \frac{w_+ + w_-}{\frac{w_+}{fid_+} + \frac{w_-}{1-fid_-}}$$

Initial Results

- Facebook Ego (Homogeneous) and IMDB (Heterogeneous) datasets





Modified GNNExplainer

- New loss function encourages ordering of candidate nodes, handles varying neighborhood sizes better.

Encourages smaller explanations (in # of nodes)

$$L_{\text{old}} = -H(Y|G = G_S) + \alpha \sum_{e_i \in E_S} e_i + \beta \cdot \text{CrossEntropy}(E_S)$$

Encourages discrete mask

Optimizes explanation towards target

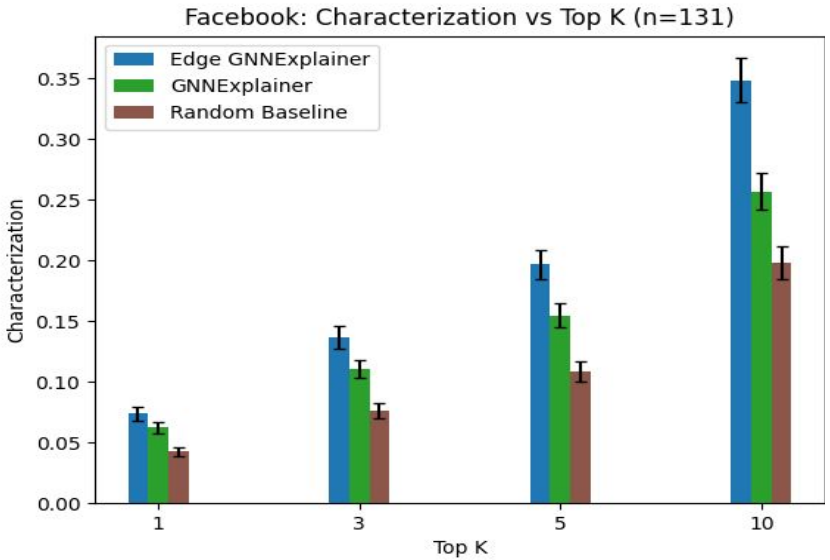
Encourages continuous mask

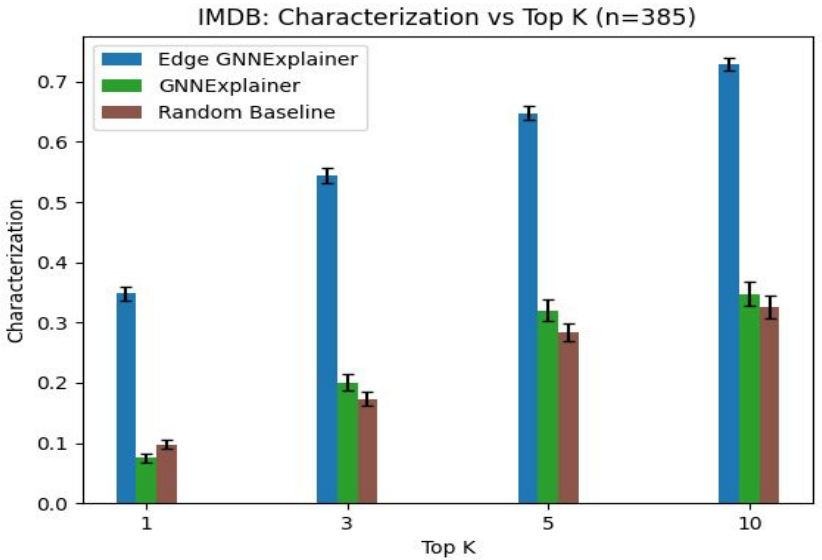
$$L_{\text{new}} = -H(Y|G = G_S) + \alpha \left(\left(\frac{1}{|E_S|} \sum_{e_i \in E_S} e_i \right) - 0.5 \right)^2 - \beta \cdot \text{CrossEntropy}(E_S)$$

Encourages medium explanation (in % of nodes)

Modified GNNExplainer Results

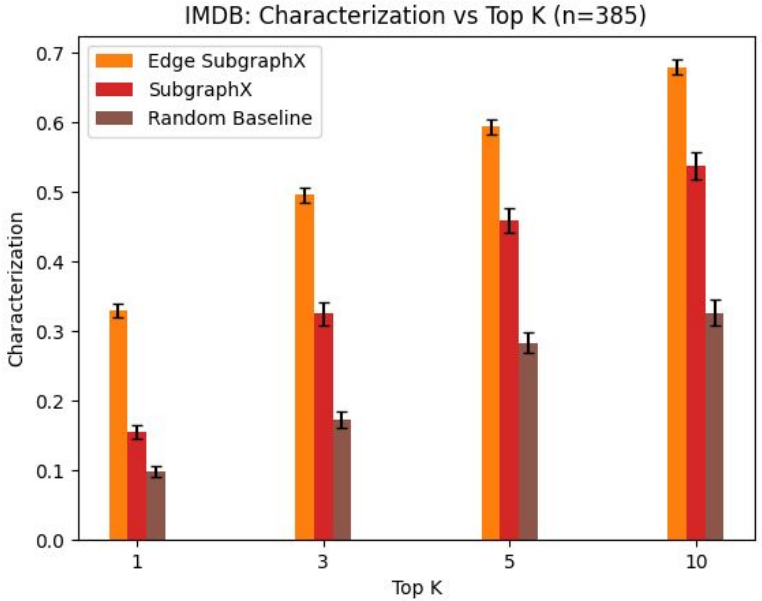
- Moderate improvement on Facebook, substantial improvement on IMDB

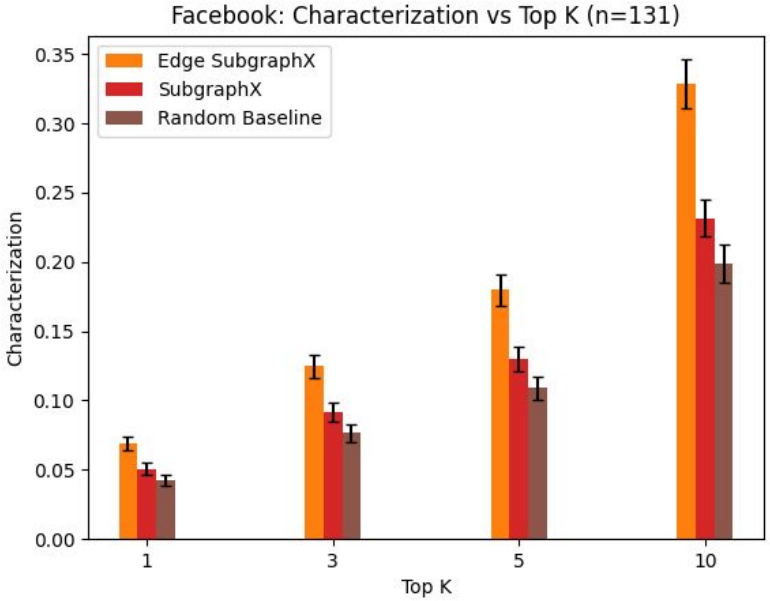




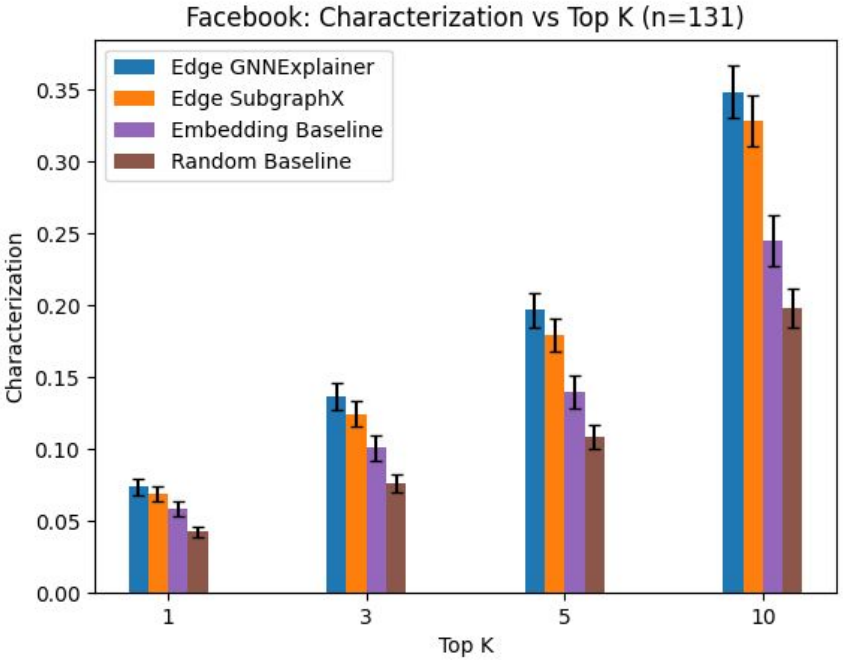
Modified SubgraphX

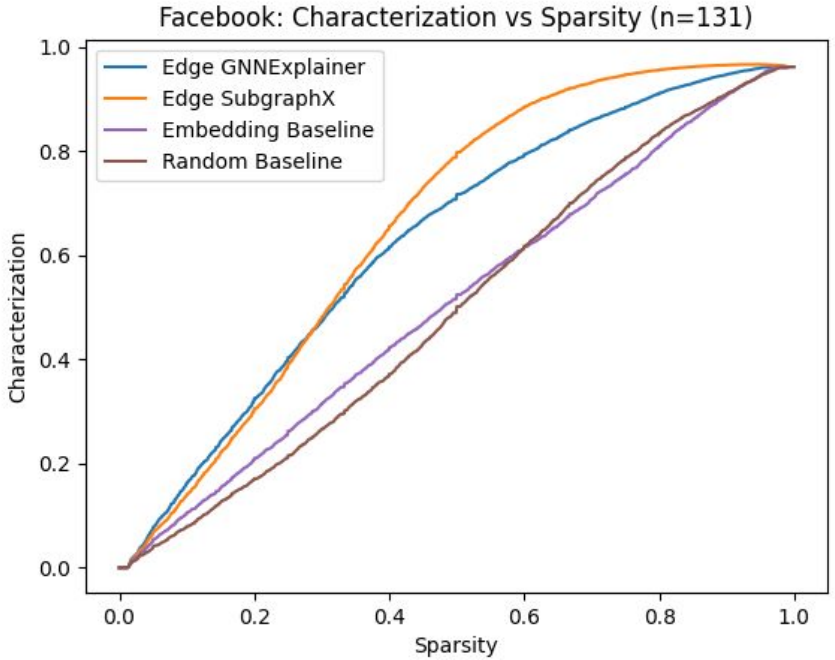
- Normally, SubgraphX masks node by setting all features to 0
- Since every candidate node is adjacent to the target link, only mask the edge between the node and the target endpoint.



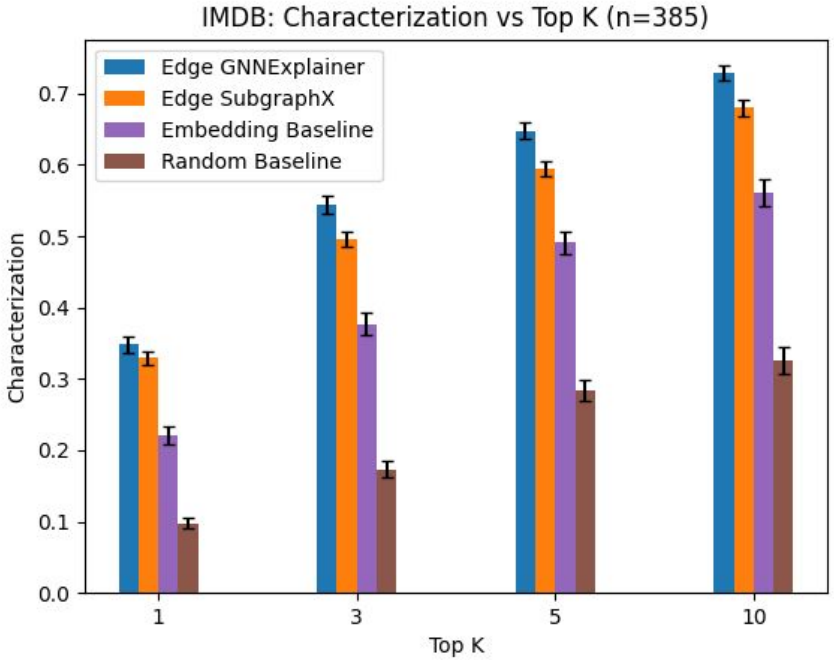


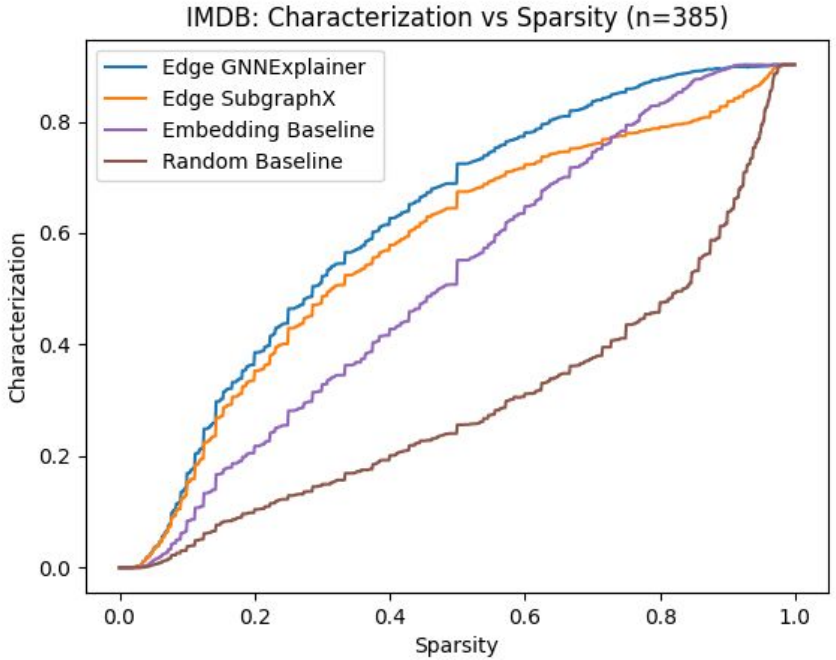
Combined Results: Facebook





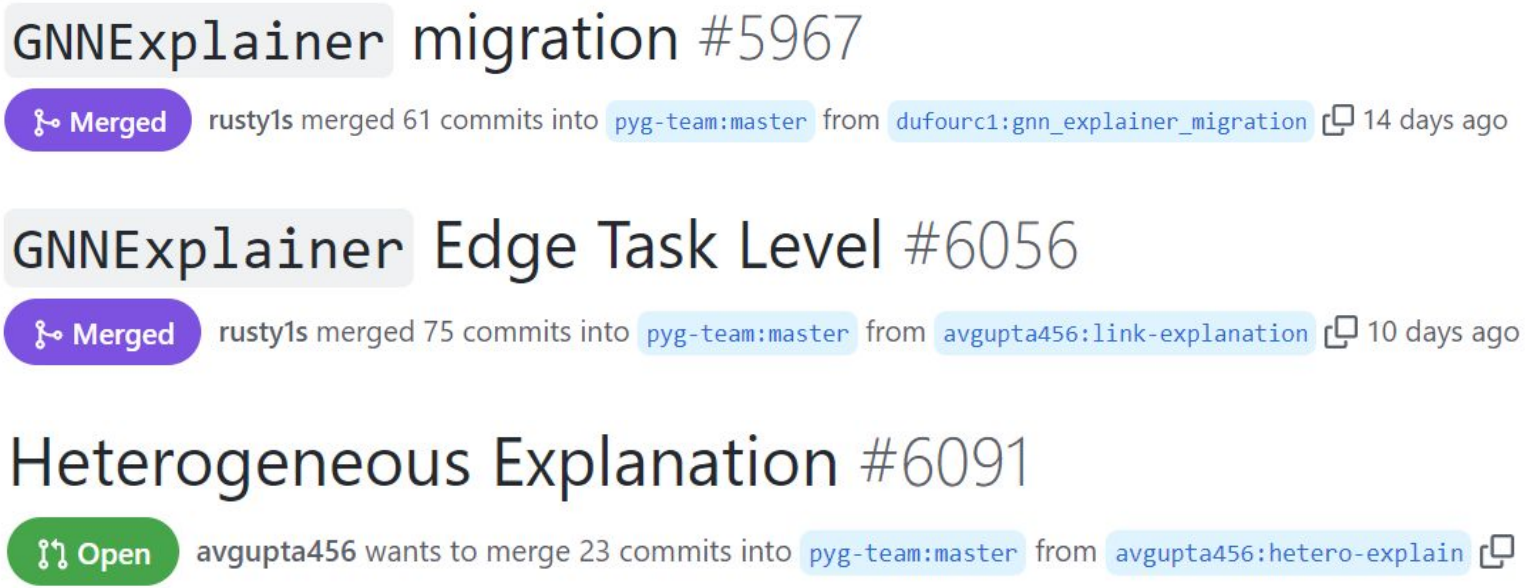
Combined Results: IMDB





Open-Source Contributions

- Contributed to PyTorch Geometric Explainability Sprint
  - New GNNExplainer implementation, Link Explanation support, Heterogeneous Graph support



Next Steps

- Improve scalability of masking implementation, run larger experiments
- Extend to the LastFM heterogeneous dataset for more results and insights
- Develop new explanation formats, methods leveraging heterogeneous graph meta-paths

Questions?