CPSC 490 Project Proposal

December 15, 2022

Abhijit Gupta, advised by Prof. Rex Ying

1 Background

1.1 Graph Neural Networks

Recently, deep learning techniques have achieved strong performance on several artificial intelligence tasks, from natural language generation to image and speech recognition. Graph data emerges in many contexts (social networks, molecules, knowledge graphs) but creates unique challenges for traditional ML methods. Graph Neural Networks (GNNs) are a type of neural network that operates directly on the graph structure to perform node and graph classification, link prediction, and more [1].

Several GNN architectures have been developed to perform multiple tasks on varied synthetic and real-world datasets. A key similarity across many methods is message passing, where features of neighboring nodes are aggregated into the central node [2]. Convolutional GNNs generalize the convolution operation from grid data to graph data, using either spectral and spatial methods. Some of the most impactful and effective methods are the Graph Convolutional Network (GCN) [3], GraphSAGE [4], Graph Attention Network (GAT) [5], and Graph Isomorphism Network (GIN) [6]. These and other methods are often evaluated on citation (CiteSeer, CORA, PubMed) [7], bioinformatics (MUTAG, PROTEINS, NCI1) [8, 9], and social network datasets (REDDIT, IMDB) [10].

1.2 GNN Explainability

In order to facilitate real-world applications, machine learning models must be made explainable, either during initial computation or post hoc. Explainability is important in increasing trust in model output and increasing transparency regarding fairness and safety. Explanations can also help end-users identify model shortcomings and incorporate their own domain-knowledge for improved performance. Demystifying black-box models is crucial to applying artificial intelligence to critical roles including healthcare, robotics, and finance.

Just as graph structured data poses challenges to traditional deep learning methods, new explainability methods are required to interpret GNN outputs. Figure 1 helps categorize several promising methods. Most recent work has gone into instance-level explanation, which provides input-dependent predictions for each input graph. Gradient and decomposition based approaches require access to the prediction model backwards computation while perturbation and surrogate models only require the prediction model input and output [11].

Among permutation instance-level explanations, GNNExplainer [12], PGExplainer [13] and SubgraphX [14] are some of the most popular approaches. GNNExplainer and PGExplainer initializes edge and node feature masks that define a smaller explanation subgraph. The masks are optimized to maximize the mutual information between the explanation and original graph. Meanwhile SubgraphX uses Monte Carlo Tree Search to explore potential subgraphs and Shapley values to select the optimal explanation. GraphLIME is a surrogate method that extends the LIME algorithm to graph structured data [15].

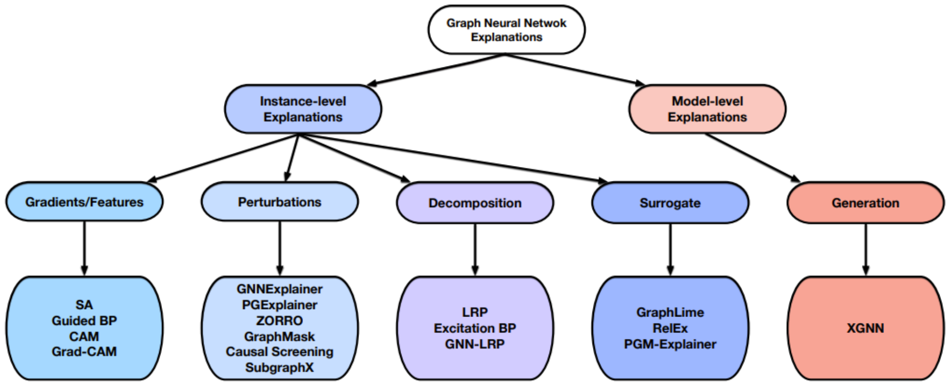

Figure 1: Taxonomy of GNN Explanation Methods, adapted from Yuan et. al [11]

GNN explanation models must balance several goals including fidelity, sparsity, stability, and accuracy [11]. Several metrics have been proposed for each goal individually [16, 17], but many models utilize their own metrics. Recently, the Deconfounded Subgraph Evaluation (DSE) metric attempts to correct the out-of-distribution problem and GraphFramEx attempts to combine fidelity measurements to classify node classification explanations as necessary and/or sufficient [18, 19]. The combination of a graph dataset, GNN prediction model, explanation model, and metrics define an explainability experiment.

1.3 Heterogeneous and Hyper Graphs

Heterogeneous graphs contain multiple types of nodes and/or edges. Meanwhile, hypergraphs are a generalization of graphs where edges can connect more than two nodes. While multiple datasets (IMDB, citation networks, e-commerce, ModelNet40 [20]) have been formulated as heterogeneous or hypergraphs and GNNs (HAN [21], HGNN [22], Hyper-Conv/Att [23], RGCN [24]) exist for both, explainability methods have not been applied to these structures.

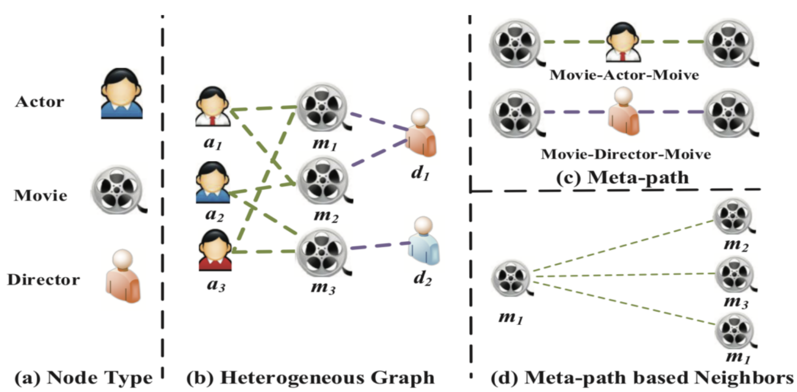

Figure 2: Heterogeneous IMDB Graph Illustration, adapted from Wang et. al [21]

2 Proposed Project

Although GNNs have been applied to a variety of graph datasets and tasks, explanation methods have lagged behind. The proposed project is to investigate applying existing GNN evaluation methods to new contexts including link prediction tasks, heterogeneous graphs, and hypergraphs. While Yuan et. al tabulated standardized results on explanations for graph and node classification tasks, a comprehensive analysis for link prediction does not yet exist [11]. In addition to link prediction, heterogeneous and hypergraph explanations are underexplored.

- **Link Prediction**: Graph datasets and GNNs have often been applied to link prediction tasks. Meanwhile, GNNExplainer, PGExplainer, and SubgraphX all describe how to apply their algorithms to link prediction but omit quantitative results. Although GraphLIME does not mention link prediction, it can be extended in a similar manner as the first three models. Depending on performance, simple optimizations can be explored to best apply these methods to link prediction. Fidelity, sparsity, stability, and accuracy can all be applied to link prediction.
- **Heterogeneous Graphs and Hypergraphs**: Unlike link prediction, GNN explainability methods and metrics will need to be more substantially updated to handle the new graph structure. Node classification, link prediction, and graph classification are all tasks that can be explored on heterogeneous graphs and hypergraphs.

I plan on quickly implementing a single explainability model (ex. GNNExplainer) for node and graph classification, and then focus on brainstorming potential ways to adapt/improve the model on these new tasks. Knowledge of the other methods will provide inspiration for modifications and can be implemented for comparison after getting good results with the first explanation model. Additional work can be done to formulate new types of explanations beyond subgraphs that are more understandable and applicable to these domains.

3 Timeline

Target deliverables <u>underlined</u>.

1. **Weeks 5-6**: Download graph datasets. Train GCN and GNNExplainer in PyG. <u>Replicate basic results for node/graph classification</u>.
2. **Weeks 7-8**: <u>Modify GNNExplainer for link prediction</u>. Modify dataset, prediction model, explanation metrics. Explore optimizations to improve explanations.
3. **Weeks 9-10**: Begin exploring heterogeneous graph and/or hypergraph modifications. Implement additional explanation metrics and/or explanation models.
4. **Weeks 11-12**: <u>Adapt GNNExplainer for heterogeneous graphs and/or hypergraphs</u>. Evaluate and optimize scalability of explanations on large graph datasets.
5. **Weeks 13-14**: <u>Compile results, write final report</u>. If time permits, evaluate additional explanation models alongside GNNExplainer.

References

[1] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1):4–24, 2020.

[2] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pages 1263–1272. PMLR, 2017.

[3] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[4] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems*, 30, 2017.

[5] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

[6] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018.

[7] Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. Collective classification in network data. *AI magazine*, 29(3):93–93, 2008.

[8] Asim Kumar Debnath, Rosa L Lopez de Compadre, Gargi Debnath, Alan J Shusterman, and Corwin Hansch. Structure-activity relationship of mutagenic aromatic and heteroaromatic nitro compounds. correlation with molecular orbital energies and hydrophobicity. *Journal of medicinal chemistry*, 34(2):786–797, 1991.

[9] Karsten M Borgwardt, Cheng Soon Ong, Stefan Schönauer, S V N Vishwanathan, Alex J Smola, and Hans-Peter Kriegel. Protein function prediction via graph kernels. *Bioinformatics (Oxford, England)*, 21 Suppl 1:i47—56, June 2005.

[10] Pinar Yanardag and SVN Vishwanathan. Deep graph kernels. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1365–1374, 2015.

[11] Hao Yuan, Haiyang Yu, Shurui Gui, and Shuiwang Ji. Explainability in graph neural networks: A taxonomic survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–19, 2022.

[12] Zhitao Ying, Dylan Bourgeois, Jiaxuan You, Marinka Zitnik, and Jure Leskovec. Gnnexplainer: Generating explanations for graph neural networks. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in neural information processing systems*, volume 32. Curran Associates, Inc., 2019.

[13] Dongsheng Luo, Wei Cheng, Dongkuan Xu, Wenchao Yu, Bo Zong, Haifeng Chen, and Xiang Zhang. Parameterized explainer for graph neural network. *Advances in Neural Information Processing Systems*, 33:19620–19631, 2020.

[14] Hao Yuan, Haiyang Yu, Jie Wang, Kang Li, and Shuiwang Ji. On explainability of graph neural networks via subgraph explorations. In *International Conference on Machine Learning*, pages 12241–12252. PMLR, 2021.

[15] Qiang Huang, Makoto Yamada, Yuan Tian, Dinesh Singh, and Yi Chang. Graphlime: Local interpretable model explanations for graph neural networks. *IEEE Transactions on Knowledge and Data Engineering*, 2022.

[16] Phillip E Pope, Soheil Kolouri, Mohammad Rostami, Charles E Martin, and Heiko Hoffmann. Explainability methods for graph convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10772–10781, 2019.

[17] Benjamin Sanchez-Lengeling, Jennifer Wei, Brian Lee, Emily Reif, Peter Wang, Wesley Qian, Kevin McCloskey, Lucy Colwell, and Alexander Wiltschko. Evaluating attribution for graph neural networks. *Advances in neural information processing systems*, 33:5898–5910, 2020.

[18] Xiang Wang, An Zhang, Xia Hu, Fuli Feng, Xiangnan He, Tat-Seng Chua, et al. Deconfounding to explanation evaluation in graph neural networks. *arXiv preprint arXiv:2201.08802*, 2022.

[19] Kenza Amara, Rex Ying, Zitao Zhang, Zhihao Han, Yinan Shan, Ulrik Brandes, Sebastian Schemm, and Ce Zhang. Graphframex: Towards systematic evaluation of explainability methods for graph neural networks. *arXiv preprint arXiv:2206.09677*, 2022.

[20] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1912–1920, 2015.

[21] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. Heterogeneous graph attention network. In *The world wide web conference*, pages 2022–2032, 2019.

[22] Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 3558–3565, 2019.

[23] Song Bai, Feihu Zhang, and Philip HS Torr. Hypergraph convolution and hypergraph attention. *Pattern Recognition*, 110:107637, 2021.

[24] Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *European semantic web conference*, pages 593–607. Springer, 2018.