
awesome-graph-explainability-papers

Papers about the explainability of GNNs

Surveys

1. [Proceedings of the IEEE 24] **Trustworthy Graph Neural Networks: Aspects, Methods and Trends** paper
2. [Arxiv 23] **A Survey on Explainability of Graph Neural Networks** paper
3. [ACM computing survey] **A Survey on Graph Counterfactual Explanations: Definitions, Methods, Evaluation, and Research Challenges** paper
4. [TPAMI 22] **Explainability in graph neural networks: A taxonomic survey.** *Yuan Hao, Yu Haiyang, Gui Shurui, Ji Shuiwang.* paper
5. [Arxiv 22] **A Survey of Explainable Graph Neural Networks: Taxonomy and Evaluation Metrics** paper
6. [Arxiv 22] **A Survey of Trustworthy Graph Learning: Reliability, Explainability, and Privacy Protection** paper
7. [Big Data 2022] **A Survey of Explainable Graph Neural Networks for Cyber Malware Analysis** paper
8. [Arxiv 23] **A Comprehensive Survey on Trustworthy Graph Neural Networks: Privacy, Robustness, Fairness, and Explainability** paper
9. [Arxiv 22] **Explaining the Explainers in Graph Neural Networks: a Comparative Study** paper
10. [Book 23] **Generative Explanation for Graph Neural Network: Methods and Evaluation** paper

Platforms

1. **PyTorch Geometric** [Document] [Blog]
2. **DIG: A Turnkey Library for Diving into Graph Deep Learning Research** paper Code
3. **GraphXAI: Evaluating Explainability for Graph Neural Networks** paper Code
4. **GraphFrameX: Towards Systematic Evaluation of Explainability Methods for Graph Neural Networks** paper Code
5. **GNNExplainer and PGExplainer** paper Code
6. **BAGEL: A Benchmark for Assessing Graph Neural Network Explanations** [paper] Code

Most Influential Papers selected by

[Cogdl](https://github.com/THUDM/cogdl/blob/master/gnn_papers.md#explainability)

1. **Explainability in graph neural networks: A taxonomic survey.** Yuan Hao, Yu Haiyang, Gui Shurui, Ji Shuiwang. ARXIV 2020. paper
2. **Gnnexplainer: Generating explanations for graph neural networks.** Ying Rex, Bourgeois Dylan, You Jiaxuan, Zitnik Marinka, Leskovec Jure. NeurIPS 2019. paper code
3. **Explainability methods for graph convolutional neural networks.** Pope Phillip E, Kolouri Soheil, Rostami Mohammad, Martin Charles E, Hoffmann Heiko. CVPR 2019.paper
4. **Parameterized Explainer for Graph Neural Network.** Luo Dongsheng, Cheng Wei, Xu Dongkuan, Yu Wenchao, Zong Bo, Chen Haifeng, Zhang Xiang. NeurIPS 2020. paper code
5. **Xggn: Towards model-level explanations of graph neural networks.** Yuan Hao, Tang Jiliang, Hu Xia, Ji Shuiwang. KDD 2020. paper.
6. **Evaluating Attribution for Graph Neural Networks.** Sanchez-Lengeling Benjamin, Wei Jennifer, Lee Brian, Reif Emily, Wang Peter, Qian Wesley, McCloskey Kevin, Colwell Lucy, Wiltschko Alexander. NeurIPS 2020.paper
7. **PGM-Explainer: Probabilistic Graphical Model Explanations for Graph Neural Networks.** Vu Minh, Thai My T.. NeurIPS 2020.paper
8. **Explanation-based Weakly-supervised Learning of Visual Relations with Graph Networks.** Federico Baldassarre and Kevin Smith and Josephine Sullivan and Hossein Azizpour. ECCV 2020.paper
9. **GCAN: Graph-aware Co-Attention Networks for Explainable Fake News Detection on Social Media.** Lu, Yi-Ju and Li, Cheng-Te. ACL 2020.paper
10. **On Explainability of Graph Neural Networks via Subgraph Explorations.** Yuan Hao, Yu Haiyang, Wang Jie, Li Kang, Ji Shuiwang. ICML 2021.paper

Year 2024

1. [ICLR 24] **GraphChef: Decision-Tree Recipes to Explain Graph Neural Networks** [paper]
2. [ICLR 24] **GOAt: Explaining Graph Neural Networks via Graph Output Attribution** [paper]
3. [ICLR 24] **Towards Robust Fidelity for Evaluating Explainability of Graph Neural Networks** [paper]
4. [ICLR 24] **UNR-Explainer: Counterfactual Explanations for Unsupervised Node Representation Learning Models** [paper]
5. [AISTATS 24] **Two Birds with One Stone: Enhancing Uncertainty Quantification and Interpretability with Graph Functional Neural Process** [paper]
6. [WWW 24] **Game-theoretic Counterfactual Explanation for Graph Neural Networks** [paper]
7. [WWW 24] **EXGC: Bridging Efficiency and Explainability in Graph Condensation**[paper]

-
8. [WWW 24] **Adversarial Mask Explainer for Graph Neural Networks**
 9. [WWW 24] **Globally Interpretable Graph Learning via Distribution Matching**[paper]
 10. [SDM 24] **XGExplainer: Robust Evaluation-based Explanation for Graph Neural Networks**[paper]
 11. [TPAMI 24] **Towards Inductive and Efficient Explanations for Graph Neural Networks**[paper]
 12. [SIGMOD 24] **View-based Explanations for Graph Neural Networks** [paper]
 13. [ICSE 24] **Coca: Improving and Explaining Graph Neural Network-Based Vulnerability Detection Systems**[paper]
 14. [AAAI 24] **Factorized Explainer for Graph Neural Networks**[paper]
 15. [AAAI 24] **Self-Interpretable Graph Learning with Sufficient and Necessary Explanations**
 16. [AAAI 24] **Explainable Origin-Destination Crowd Flow Interpolation via Variational Multi-Modal Recurrent Graph Auto-Encoder** [paper]
 17. [AAAI 24] **Stratified GNN Explanations through Sufficient Expansion**[paper]
 18. [AAAI workshop] **Semi-Supervised Graph Representation Learning with Human-centric Explanation for Predicting Fatty Liver Disease**[paper]
 19. [Arxiv 24.04] **Superior Polymeric Gas Separation Membrane Designed by Explainable Graph Machine Learning** [paper]
 20. [Arxiv 24.04] **Improving the interpretability of GNN predictions through conformal-based graph sparsification** [paper]
 21. [Arxiv 24.03] **GreeDy and CoDy: Counterfactual Explainers for Dynamic Graph**[paper]
 22. [Arxiv 24.03] **Explainable Graph Neural Networks for Observation Impact Analysis in Atmospheric State Estimation**[paper]
 23. [Arxiv 24.03] **Securing GNNs: Explanation-Based Identification of Backdoored Training Graphs**[paper]
 24. [Arxiv 24.03] **Iterative Graph Neural Network Enhancement via Frequent Subgraph Mining of Explanations**[paper]
 25. [Arxiv 24.03] **Explainable Graph Neural Networks for Observation Impact Analysis in Atmospheric State Estimation**[paper]
 26. [Arxiv 24.02] **Feature Attribution with Necessity and Sufficiency via Dual-stage Perturbation Test for Causal Explanation**[paper]
 27. [Arxiv 24.02] **Interpreting Graph Neural Networks with In-Distributed Proxies**[paper]
 28. [Arxiv 24.02] **PAC Learnability under Explanation-Preserving Graph Perturbations**[paper]
 29. [Arxiv 24.02] **Explainable Global Wildfire Prediction Models using Graph Neural Networks**[paper]
 30. [Arxiv 24.02] **Incorporating Retrieval-based Causal Learning with Information Bottlenecks for Interpretable Graph Neural Networks**[paper]
 31. [Arxiv 24.01] **SynHIN: Generating Synthetic Heterogeneous Information Network for Explainable AI**[paper]
-

-
32. [Arxiv 24.01] **On Discrepancies between Perturbation Evaluations of Graph Neural Network Attributions**[paper]
 33. [ASP=DAC 24] **LIPSTICK: Corruptibility-Aware and Explainable Graph Neural Network-based Oracle-Less Attack on Logic Locking**[paper]
 34. [Biorxiv 24] **Community-aware explanations in knowledge graphs with XP-GNN**[paper]
 35. [Information Procs. & Mana.] **Towards explaining graph neural networks via preserving prediction ranking and structural dependency**[paper]
 36. [Applied Energy] **Explainable Spatio-Temporal Graph Neural Networks for multi-site photovoltaic energy production** [paper]
 37. [Computational Materials Science] **Graph isomorphism network for materials property prediction along with explainability analysis**[paper]
 38. [Arxiv 24] **GNNShap: Fast and Accurate GNN Explanations using Shapley Values** [paper]
 39. [NN 24] **Explanatory subgraph attacks against Graph Neural Networks**[paper]
 40. [Neural Networks 24] **CI-GNN: A Granger Causality-Inspired Graph Neural Network for Interpretable Brain Network-Based Psychiatric Diagnosis** [paper]
 41. [NeuroImage 24] **BPI-GNN: Interpretable brain network-based psychiatric diagnosis and subtyping**[paper]
 42. [MedRxiv 24] **An Interpretable Population Graph Network to Identify Rapid Progression of Alzheimer's Disease Using UK Biobank**[paper]
 43. [IEEE TDSC 24] **TrustGuard: GNN-based Robust and Explainable Trust Evaluation with Dynamicity Support** [paper]
 44. [IEEE Transactions] **IEEE Transactions on Computational Social Systems**[paper]
 45. [Journal of Physics] **Explainer on GNN-based segmentation networks**[paper]
 46. [Energy and AI] **Electricity demand forecasting at distribution and household levels using explainable causal graph neural network** [paper]

Year 2023

1. [NeurIPS 23] **Interpretable Graph Networks Formulate Universal Algebra Conjectures**[paper]
2. [NeurIPS 23] **SAME: Uncovering GNN Black Box with Structure-aware Shapley-based Multipiece Explanation** [paper]
3. [NeurIPS 23] **Train Once and Explain Everywhere: Pre-training Interpretable Graph Neural Networks**[paper]
4. [NeurIPS 23] **D4Explainer: In-distribution Explanations of Graph Neural Network via Discrete Denoising Diffusion** [paper]
5. [NeurIPS 23] **TempME: Towards the Explainability of Temporal Graph Neural Networks via Motif Discovery** [paper]

-
6. [NeurIPS 23] **V-InFoR: A Robust Graph Neural Networks Explainer for Structurally Corrupted Graphs** [paper]
 7. [NeurIPS 23] **Towards Self-Interpretable Graph-Level Anomaly Detection** [paper]
 8. [NeurIPS 23] **Evaluating Post-hoc Explanations for Graph Neural Networks via Robustness Analysis** [paper]
 9. [NeurIPS 23] **Interpretable Prototype-based Graph Information Bottleneck** [paper]
 10. [ICML 23] **Rethinking Explaining Graph Neural Networks via Non-parametric Subgraph Matching** [paper]
 11. [ICML 23] **Relevant Walk Search for Explaining Graph Neural Networks** [paper]
 12. [ICML 23] **Towards Understanding the Generalization of Graph Neural Networks** [paper]
 13. [ICLR 23] **GNNInterpreter: A Probabilistic Generative Model-Level Explanation for Graph Neural Networks** [paper]
 14. [ICLR 23] **Global Explainability of GNNs via Logic Combination of Learned Concepts** [paper]
 15. [ICLR 23] **Explaining Temporal Graph Models through an Explorer-Navigator Framework** [paper]
 16. [ICLR 23] **DAG Matters! GFlowNets Enhanced Explainer for Graph Neural Networks** [paper]
 17. [ICLR 23] **Interpretable Geometric Deep Learning via Learnable Randomness Injection** [paper]
 18. [ICLR 23] **A Differential Geometric View and Explainability of GNN on Evolving Graphs** [paper]
 19. [KDD 23] **MixupExplainer: Generalizing Explanations for Graph Neural Networks with Data Augmentation** [paper]
 20. [KDD 23] **Counterfactual Learning on Heterogeneous Graphs with Greedy Perturbation** [paper]
 21. [KDD 23] **Empower Post-hoc Graph Explanations with Information Bottleneck: A Pre-training and Fine-tuning Perspective** [paper]
 22. [KDD 23] **Less is More: SlimG for Accurate, Robust, and Interpretable Graph Mining.** [paper]
 23. [KDD 23] **Shift-Robust Molecular Relational Learning with Causal Substructure** [paper]
 24. [AAAI 23] **Global Concept-Based Interpretability for Graph Neural Networks via Neuron Analysis** [paper]
 25. [AAAI 23] **On the Limit of Explaining Black-box Temporal Graph Neural Networks** [paper]
 26. [AAAI 23] **Towards Fine-Grained Explainability for Heterogeneous Graph Neural Network** [paper]
 27. [AAAI 23] **Interpretable Chirality-Aware Graph Neural Network for Quantitative Structure Activity Relationship Modeling in Drug Discovery** [paper]
 28. [VLDB 23] **HENCE-X: Toward Heterogeneity-agnostic Multi-level Explainability for Deep Graph Networks** [paper]
 29. [VLDB 23] **On Data-Aware Global Explainability of Graph Neural Networks** [paper]
-

-
30. [AISTATS 23] **Distill n’ Explain: explaining graph neural networks using simple surrogates** [Paper]
 31. [AISTATS 23] **Probing Graph Representations** [paper]
 32. [ICDE 23] **INGREX: An Interactive Explanation Framework for Graph Neural Networks**[paper]
 33. [ICDE 23] **Jointly Attacking Graph Neural Network and its Explanations** [paper]
 34. [WWW 23]**PaGE-Link: Path-based Graph Neural Network Explanation for Heterogeneous Link Prediction** [paper]
 35. [ICDM 23] **Limitations of Perturbation-based Explanation Methods for Temporal Graph Neural Networks**
 36. [ICDM 23] **Interpretable Subgraph Feature Extraction for Hyperlink Prediction**[paper]
 37. [WSDM 23]**Interpretable Research Interest Shift Detection with Temporal Heterogeneous Graphs** [paper]
 38. [WSDM 23]**Cooperative Explanations of Graph Neural Networks** [paper]
 39. [WSDM 23]**Towards Faithful and Consistent Explanations for Graph Neural Networks** [paper]
 40. [WSDM 23] **Global Counterfactual Explainer for Graph Neural Networks** [paper]
 41. [CIKM 23] **Explainable Spatio-Temporal Graph Neural Networks** [paper]
 42. [CIKM 23] **DuoGAT: Dual Time-oriented Graph Attention Networks for Accurate, Efficient and Explainable Anomaly Detection on Time-series.** [paper]
 43. [CIKM 23] **Heterogeneous Temporal Graph Neural Network Explainer** [paper]
 44. [CIKM 23] **ACGAN-GNNExplainer: Auxiliary Conditional Generative Explainer for Graph Neural Networks**[paper]
 45. [CIKM 23] **KG4Ex: An Explainable Knowledge Graph-Based Approach for Exercise Recommendation** [paper]
 46. [ECML-PKDD 23] **ENGAGE: Explanation Guided Data Augmentation for Graph Representation Learning** [paper]
 47. [TPAMI 23] **FlowX: Towards Explainable Graph Neural Networks via Message Flows** [paper]
 48. [TAI] **Prototype-based interpretable graph neural networks.** [paper]
 49. [TKDE 23] **Counterfactual Graph Learning for Anomaly Detection on Attributed Networks** [paper]
 50. [Scientific Data 23] **Evaluating explainability for graph neural networks** [paper]
 51. [Nature Communications 23] **Chemistry-intuitive explanation of graph neural networks for molecular property prediction with substructure masking** [paper]
 52. [ACM Computing Surveys 23] **A Survey on Graph Counterfactual Explanations: Definitions, Methods, Evaluation** [paper]
 53. [TIST 23] **Faithful and Consistent Graph Neural Network Explanations with Rationale Alignment** [paper]
-

-
54. [Openreview 23] **STExplainer: Global Explainability of GNNs via Frequent SubTree Mining** [paper]
 55. [Openreview 23] **Everybody Needs a Little HELP: Explaining Graphs via Hierarchical Concepts** [paper]
 56. [Openreview 23] **Iterative Graph Neural Network Enhancement Using Explanations** [paper]
 57. [Openreview 23] **Interpretable and Generalizable Graph Neural Networks via Subgraph Multilinear Extension** [paper]
 58. [Openreview 23] **Interpretable and Convergent Graph Neural Network Layers at Scale** [paper]
 59. [Openreview 23] **InduCE: Inductive Counterfactual Explanations for Graph Neural Networks** [paper]
 60. [NeurIPS 2023 Workshop XAIA] **GInX-Eval: Towards In-Distribution Evaluation of Graph Neural Networks Explanations** [paper]
 61. [NeurIPS 2023 Workshop XAIA] **On the Consistency of GNN Explainability Methods** [paper]
 62. [Arxiv 23] **Evaluating Neighbor Explainability for Graph Neural Networks** [paper]
 63. [Arxiv 23] **GNNX-BENCH: Unravelling the Utility of Perturbation-based GNN Explainers through In-depth Benchmarking** [paper]
 64. [Arxiv 23] **DyExplainer: Explainable Dynamic Graph Neural Networks** [paper]
 65. [Arxiv 23] **Explainability-Based Adversarial Attack on Graphs Through Edge Perturbation** [paper]
 66. [AICS 23] **A subgraph interpretation generative model for knowledge graph link prediction based on uni-relation transformation** [paper]
 67. [GUT 23] **Screening of normal endoscopic large bowel biopsies with interpretable graph learning: a retrospective study** [paper]
 68. [PR 2023] **Towards self-explainable graph convolutional neural network with frequency adaptive inception** [paper]
 69. [MLG 2023] **Understanding how explainers work in graph neural networks** [paper]
 70. [MLG 2023] **Graph Model Explainer Tool** [paper]
 71. [Information Science 23] **Robust explanations for graph neural network with neuron explanation component** [paper]
 72. [Recsys 23] **Explainable Graph Neural Network Recommenders; Challenges and Opportunities** [paper]
 73. [xAI 23] **Counterfactual Explanations for Graph Classification Through the Lenses of Density** [paper]
 74. [XAI 23] **Evaluating Link Prediction Explanations for Graph Neural Networks** [[paper]](<https://arxiv.org/abs/2308.01682>)
 75. [xAI 23] **XInsight: Revealing Model Insights for GNNs with Flow-based Explanations** [paper]
 76. [xAI 23] **Quantifying the Intrinsic Usefulness of Attributional Explanations for Graph Neural**

-
- Networks with Artificial Simulatability Studies** [paper]
77. [xAI 23] **MEGAN: Multi Explanation Graph Attention Network** [paper]
 78. [XKDD 23] **Game Theoretic Explanations for Graph Neural Networks** [paper]
 79. [XKDD 23] **From Black Box to Glass Box: Evaluating Faithfulness of Process Predictions with GCNNs** [paper]
 80. [IJCNN 23] **MEGA: Explaining Graph Neural Networks with Network Motifs** [paper]
 81. [LOG Poster 23] **On the Robustness of Post-hoc GNN Explainers to Label Noise** [paper]
 82. [LOG Poster 23] **How Faithful are Self-Explainable GNNs?** [paper]
 83. [LOG Poster 23] **RegExplainer: Generating Explanations for Graph Neural Networks in Regression Task** [paper]
 84. [LOG Poster 23] **Explaining Link Predictions in Knowledge Graph Embedding Models with Influential Examples** [paper]
 85. [Bioriv 23] **Building explainable graph neural network by sparse learning for the drug-protein binding prediction** [paper]
 86. [ICAID 2023] **Explanations for Graph Neural Networks via Layer Analysis.** [paper]
 87. [ECAI 23] **XGBD: Explanation-Guided Graph Backdoor Detection** [paper]
 88. [IEEE Transactions on Consumer Electronics 23] **Human Pose Prediction Using Interpretable Graph Convolutional Network for Smart Home** [paper]
 89. [KBS 23] **KE-X: Towards subgraph explanations of knowledge graph embedding based on knowledge information gain** [paper]
 90. [ICML workshop 23] **Generating Global Factual and Counterfactual Explainer for Molecule under Domain Constraints** [paper]
 91. [Thesis 23] **Developing interpretable graph neural networks for high dimensional feature spaces** [paper]
 92. [Thesis 23] **Evaluation of Explainability Methods on Single-Cell Classification Tasks Using Graph Neural Networks** [paper]
 93. [Arxiv 23] **On the Interplay of Subset Selection and Informed Graph Neural Networks** [paper]
 94. [ISSTA23] **Interpreters for GNN-Based Vulnerability Detection: Are We There Yet?** [paper]
 95. [ICECAI23] **Improved GraphSVX for GNN Explanations Based on Cross Entropy** [paper]
 96. [ICRA Workshop 23] **Towards Semantic Interpretation and Validation of Graph Attention-based Explanations** [paper]
 97. [Arxiv 23] **Graph Neural Network based Log Anomaly Detection and Explanation** [paper]
 98. [Arxiv 23] **Interpreting GNN-based IDS Detections Using Provenance Graph Structural Features** [paper]
 99. [Thesis 23] **Interpretability of Graphical Models** [paper]
 100. [Bioengineering 2023] **Personalized Explanations for Early Diagnosis of Alzheimer’s Disease Using Explainable Graph Neural Networks with Population Graphs** [paper]
-

-
101. [BDSC 2023] **MDC: An Interpretable GNNs Method Based on Node Motif Degree and Graph Diffusion Convolution** [[paper]] (https://link.springer.com/chapter/10.1007/978-981-99-3925-1_24)
 102. [Arxiv 2023] **In-Process Global Interpretation for Graph Learning via Distribution Matching** [paper]
 103. [Information Science 2023] **Explainability techniques applied to road traffic forecasting using Graph Neural Network models** [paper]
 104. [Arxiv 23] **Efficient GNN Explanation via Learning Removal-based Attribution** [paper]
 105. [Arxiv 23] **Empowering Counterfactual Reasoning over Graph Neural Networks through Inductivity** [paper]
 106. [ICLR Tiny 23] **Message-passing selection: Towards interpretable GNNs for graph classification** [paper]
 107. [ICLR Tiny 23] **Revisiting CounterGAN for Counterfactual Explainability of Graphs** [paper]
 108. [MICCAI Workshop 23] **IA-GCN: Interpretable Attention based Graph Convolutional Network for Disease prediction** [paper]
 109. [Arxiv 23] **Robust Ante-hoc Graph Explainer using Bilevel Optimization** [paper]
 110. [GRADES & NDA'23] **A Demonstration of Interpretability Methods for Graph Neural Networks** [paper]
 111. [Arxiv 23] **Self-Explainable Graph Neural Networks for Link Prediction** [paper]
 112. [ChemRxiv 23] **Interpreting Graph Neural Networks with Myerson Values for Cheminformatics Approaches** [paper]
 113. [Neural Networks 23] **Generating Post-hoc Explanations for Skip-gram-based Node Embeddings by Identifying Important Nodes with Bridgeness** [paper]
 114. [ICASSP 23] **Towards a More Stable and General Subgraph Information Bottleneck** [paper]
 115. [ESANN 23] **Combining Stochastic Explainers and Subgraph Neural Networks can Increase Expressivity and Interpretability** [Paper]
 116. [IEEE Access] **Generating Real-Time Explanations for GNNs via Multiple Specialty Learners and Online Knowledge Distillation** [Paper]
 117. [IEEE Access] **Providing Post-Hoc Explanation for Node Representation Learning Models Through Inductive Conformal Predictions** [paper]
 118. [Journal of Software 23] **A Slice-level vulnerability detection and interpretation method based on graph neural network** [paper]
 119. [Automation in Construction 23] **Learning from explainable data-driven tunneling graphs: A spatio-temporal graph convolutional network for clogging detection** [paper]
 120. [Briefings in Bioinformatics] **Predicting molecular properties based on the interpretable graph neural network with multistep focus mechanism** [paper]
 121. [Briefings in Bioinformatics] **Identification of vital chemical information via visualization of graph neural networks** [paper]

-
122. [Bioinformatics 23] **Explainable Multilayer Graph Neural Network for Cancer Gene Prediction** [paper]
 123. [ICLR Workshop 23] **GCI: A Graph Concept Interpretation Framework** [paper]
 124. [Arxiv 23] **Structural Explanations for Graph Neural Networks using HSIC** [paper]
 125. [Internet of Things 23] **XG-BoT: An Explainable Deep Graph Neural Network for Botnet Detection and Forensics** [paper]
 126. [JOS23] **A Generic Explaining & Locating Method for Malware Detection based on Graph Neural Networks** [paper]

Year 2022

1. [NeurIPS 22] **GStarX: Explaining Graph-level Predictions with Communication Structure-Aware Cooperative Games** [paper]
2. [NeurIPS 22] **Debiasing Graph Neural Networks via Learning Disentangled Causal Substructure** [paper]
3. [NeurIPS 22] **Task-Agnostic Graph Neural Explanations** [paper]
4. [NeurIPS 22] **CLEAR: Generative Counterfactual Explanations on Graphs** [paper]
5. [ICML 22] **Interpretable and Generalizable Graph Learning via Stochastic Attention Mechanism** [paper]
6. [ICLR 22] **DEGREE: Decomposition Based Explanation for Graph Neural Networks** [paper]
7. [ICLR 22] **Explainable GNN-Based Models over Knowledge Graphs** [paper]
8. [ICLR 22] **Discovering Invariant Rationales for Graph Neural Networks** [paper]
9. [KDD 22] **On Structural Explanation of Bias in Graph Neural Networks** [paper]
10. [KDD 22] **Causal Attention for Interpretable and Generalizable Graph Classification** [paper]
11. [CVPR 22] **OrphicX: A Causality-Inspired Latent Variable Model for Interpreting Graph Neural Networks** [paper]
12. [CVPR 22] **Improving Subgraph Recognition with Variational Graph Information Bottleneck** [paper]
13. [AISTATS 22] **Probing GNN Explainers: A Rigorous Theoretical and Empirical Analysis of GNN Explanation Methods** [paper]
14. [AISTATS 22] **CF-GNNExplainer: Counterfactual Explanations for Graph Neural Networks** [paper]
15. [TPAMI 22] **Differentially Private Graph Neural Networks for Whole-Graph Classification** [paper]
16. [TPAMI 22] **Reinforced Causal Explainer for Graph Neural Networks** [paper]
17. [VLDB 22] **xFraud: Explainable Fraud Transaction Detection on Heterogeneous Graphs** [paper]

-
18. [LOG 22] **GraphFramEx: Towards Systematic Evaluation of Explainability Methods for Graph Neural Networks** [paper]
 19. [LOG 22] **Towards Training GNNs using Explanation Directed Message Passing** [paper]
 20. [The Webconf 22] **Learning and Evaluating Graph Neural Network Explanations based on Counterfactual and Factual Reasoning** [paper]
 21. [AAAI 22] **Prototype-Based Explanations for Graph Neural Networks** [paper]
 22. [AAAI 22] **KerGNNs: Interpretable Graph Neural Networks with Graph Kernels** [paper]
 23. [AAAI 22] **ProtGNN: Towards Self-Explaining Graph Neural Networks** [paper]
 24. [IEEE Big Data 22] **Trade less Accuracy for Fairness and Trade-off Explanation for GNN** [paper]
 25. [CIKM 22] **GRETEL: A unified framework for Graph Counterfactual Explanation Evaluation** [paper]
 26. [CIKM 22] **GRETEL: Graph Counterfactual Explanation Evaluation Framework** [paper]
 27. [CIKM 22] **A Model-Centric Explainer for Graph Neural Network based Node Classification** [paper]
 28. [IJCAI 22] **What Does My GNN Really Capture? On Exploring Internal GNN Representations** [paper]
 29. [ECML PKDD 22] **Improving the quality of rule-based GNN explanations** [paper]
 30. [MICCAI 22] **Interpretable Graph Neural Networks for Connectome-Based Brain Disorder Analysis** [paper]
 31. [MICCAI 22] **Sparse Interpretation of Graph Convolutional Networks for Multi-modal Diagnosis of Alzheimer's Disease** [paper]
 32. [EuroS&P 22] **Illuminati: Towards Explaining Graph Neural Networks for Cybersecurity Analysis** [paper]
 33. [INFOCOM 22] **Interpretability Evaluation of Botnet Detection Model based on Graph Neural Network** [paper]
 34. [GLOBECOM 22] **Shapley Explainer - An Interpretation Method for GNNs Used in SDN** [paper]
 35. [GLOBECOM 22] **An Explainer for Temporal Graph Neural Networks** [paper]
 36. [TKDE 22] **Zorro: Valid, Sparse, and Stable Explanations in Graph Neural Networks** [paper]
 37. [TNNLS 22] **Interpretable Graph Reservoir Computing With the Temporal Pattern Attention** [paper]
 38. [TNNLS22] **A Meta-Learning Approach for Training Explainable Graph Neural Networks** [paper]
 39. [TNNLS 22] **Explaining Deep Graph Networks via Input Perturbation** [paper]
 40. [TNNLS 22] **A Meta-Learning Approach for Training Explainable Graph Neural Network** [paper]
 41. [DMKD 22] **On GNN explainability with activation patterns** [paper]
 42. [KBS 22] **EGNN: Constructing explainable graph neural networks via knowledge distilla-**

-
- tion [paper]
43. [XKDD 22] **GREASE: Generate Factual and Counterfactual Explanations for GNN-based Recommendations** [paper]
 44. [AI 22] **Are Graph Neural Network Explainers Robust to Graph Noises?** [paper]
 45. [BRACIS 22] **ConveXplainer for Graph Neural Networks** [paper]
 46. [GLB 22] **An Explainable AI Library for Benchmarking Graph Explainers** [paper]
 47. [DASFAA 22] **On Global Explainability of Graph Neural Networks** [paper]
 48. [ISBI 22] **Interpretable Graph Convolutional Network Of Multi-Modality Brain Imaging For Alzheimer’s Disease Diagnosis** [paper]
 49. [Bioinformatics] **GNN-SubNet: disease subnetwork detection with explainable Graph Neural Networks** [paper]
 50. [Medical Imaging 2022] **Phenotype guided interpretable graph convolutional network analysis of fMRI data reveals changing brain connectivity during adolescence** [paper]
 51. [NeuroComputing 22] **Perturb more, trap more: Understanding behaviors of graph neural networks** [paper]
 52. [DSN 22] **CFGExplainer: Explaining Graph Neural Network-Based Malware Classification from Control Flow Graphs** [paper]
 53. [IEEE Access 22] **Providing Node-level Local Explanation for node2vec through Reinforcement Learning** [paper]
 54. [Patterns 22] **Quantitative Evaluation of Explainable Graph Neural Networks for Molecular Property Prediction** [paper]
 55. [Arxiv 22] **GRAPHSHAP: Motif-based Explanations for Black-box Graph Classifiers** [paper]
 56. [IEEE Access 22] **Providing Post-Hoc Explanation for Node Representation Learning Models Through Inductive Conformal Predictions** [paper]
 57. [IEEE 22] **Explaining Graph Neural Networks With Topology-Aware Node Selection: Application in Air Quality Inference** [paper]
 58. [BioRxiv 22] **GNN-SubNet: disease subnetwork detection with explainable Graph Neural Networks** [paper]
 59. [IEEE Robotics and Automation Letters 22] **Efficient and Interpretable Robot Manipulation with Graph Neural Networks** [paper]
 60. [Arxiv 22] **Deconfounding to Explanation Evaluation in Graph Neural Networks** [paper]
 61. [ICCPR 22] **GANExplainer: GAN-based Graph Neural Networks Explainer** [paper]
 62. [Arxiv 22] **On the Probability of Necessity and Sufficiency of Explaining Graph Neural Networks: A Lower Bound Optimization Approach** [paper]
 63. [Arxiv 22] **Exploring Explainability Methods for Graph Neural Networks** [paper]
 64. [Arxiv 22] **PAGE: Prototype-Based Model-Level Explanations for Graph Neural Networks** [paper]
 65. [Arxiv 22] **Toward Multiple Specialty Learners for Explaining GNNs via Online Knowledge**
-

-
- Distillation** [paper]
66. [Openreview 23] **TGP: Explainable Temporal Graph Neural Networks for Personalized Recommendation** [paper]
 67. [Openreview 23] **On Regularization for Explaining Graph Neural Networks: An Information Theory Perspective** [paper]
 68. [Arxiv 22] **L2XGNN: Learning to Explain Graph Neural Networks** [paper]
 69. [Arxiv 22] **Towards Prototype-Based Self-Explainable Graph Neural Network** [paper]
 70. [Arxiv 22] **PGX: A Multi-level GNN Explanation Framework Based on Separate Knowledge Distillation Processes** [paper]
 71. [Arxiv 22] **Explainability in subgraphs-enhanced Graph Neural Networks** [paper]
 72. [Arxiv 22] **Defending Against Backdoor Attack on Graph Neural Network by Explainability** [paper]
 73. [Arxiv 22] **Explaining Dynamic Graph Neural Networks via Relevance Back-propagation** [paper]
 74. [Arxiv 22] **EiX-GNN : Concept-level eigencentality explainer for graph neural networks** [paper]
 75. [Arxiv 22] **MotifExplainer: a Motif-based Graph Neural Network Explainer** [paper]
 76. [Arxiv 22] **Faithful Explanations for Deep Graph Models** [paper]
 77. [Arxiv 22] **Towards Explanation for Unsupervised Graph-Level Representation Learning** [paper]
 78. [Arxiv 22] **BAGEL: A Benchmark for Assessing Graph Neural Network Explanations** [paper]
 79. [Arxiv 22] **BrainIB: Interpretable Brain Network-based Psychiatric Diagnosis with Graph Information Bottleneck** [paper]
 80. [Arxiv 22] **A Comprehensive Survey on Trustworthy Graph Neural Networks: Privacy, Robustness, Fairness, and Explainability** [paper]
 81. [Arxiv 22] **Explainability in Graph Neural Networks: An Experimental Survey** [paper]
 82. [IEEE TSIPN 22] **Explainability and Graph Learning from Social Interactions** [paper]
 83. [Arxiv 22] **Cognitive Explainers of Graph Neural Networks Based on Medical Concepts** [paper]

Year 2021

1. [NeurIPS 21] **SALKG: Learning From Knowledge Graph Explanations for Commonsense Reasoning** [paper]
2. [NeurIPS 2021] **Reinforcement Learning Enhanced Explainer for Graph Neural Networks** [paper]
3. [NeurIPS 2021] **Towards Multi-Grained Explainability for Graph Neural Networks** [paper]
4. [NeurIPS 2021] **Robust Counterfactual Explanations on Graph Neural Networks** [paper]

-
5. [ICML 2021] **On Explainability of Graph Neural Networks via Subgraph Explorations**[paper]
 6. [ICML 2021] **Generative Causal Explanations for Graph Neural Networks**[paper]
 7. [ICML 2021] **Improving Molecular Graph Neural Network Explainability with Orthonormalization and Induced Sparsity**[paper]
 8. [ICML 2021] **Automated Graph Representation Learning with Hyperparameter Importance Explanation**[paper]
 9. [ICLR 2021] **Interpreting Graph Neural Networks for NLP With Differentiable Edge Masking**[paper]
 10. [ICLR 2021] **Graph Information Bottleneck for Subgraph Recognition** [paper]
 11. [KDD 2021] **When Comparing to Ground Truth is Wrong: On Evaluating GNN Explanation Methods**[paper]
 12. [KDD 2021] **Counterfactual Graphs for Explainable Classification of Brain Networks** [paper]
 13. [CVPR 2021] **Quantifying Explainers of Graph Neural Networks in Computational Pathology**. [paper]
 14. [NAACL 2021] **Counterfactual Supporting Facts Extraction for Explainable Medical Record Based Diagnosis with Graph Network**. [paper]
 15. [AAAI 2021] **Motif-Driven Contrastive Learning of Graph Representations** [paper]
 16. [TPAMI 21] **Higher-Order Explanations of Graph Neural Networks via Relevant Walks** [paper]
 17. [WWW 2021] **Interpreting and Unifying Graph Neural Networks with An Optimization Framework** [paper]
 18. [Genome medicine 21] **Explaining decisions of Graph Convolutional Neural Networks: patient-specific molecular subnetworks responsible for metastasis prediction in breast cancer** [paper]
 19. [IJCKG 21] **Knowledge Graph Embedding in E-commerce Applications: Attentive Reasoning, Explanations, and Transferable Rules** [paper]
 20. [RuleML+RR 21] **Combining Sub-Symbolic and Symbolic Methods for Explainability** [paper]
 21. [PAKDD 21] **SCARLET: Explainable Attention based Graph Neural Network for Fake News spreader prediction** [paper]
 22. [J. Chem. Inf. Model] **Coloring Molecules with Explainable Artificial Intelligence for Preclinical Relevance Assessment** [paper]
 23. [BioRxiv 21] **APRILE: Exploring the Molecular Mechanisms of Drug Side Effects with Explainable Graph Neural Networks** [paper]
 24. [ISM 21] **Edge-Level Explanations for Graph Neural Networks by Extending Explainability Methods for Convolutional Neural Networks** [paper]
 25. [Arxiv 21] **Towards the Explanation of Graph Neural Networks in Digital Pathology with Information Flows** [paper]
 26. [Arxiv 21] **SEEN: Sharpening Explanations for Graph Neural Networks using Explanations**

-
- from Neighborhoods** [paper]
27. [Arxiv 21] **Preserve, Promote, or Attack? GNN Explanation via Topology Perturbation** [paper]
 28. [Arxiv 21] **Learnt Sparsification for Interpretable Graph Neural Networks** [paper]
 29. [ICML workshop 21] **GCEExplainer: Human-in-the-Loop Concept-based Explanations for Graph Neural Networks** [paper]
 30. [ICML workshop 21] **Reliable Graph Neural Network Explanations Through Adversarial Training** [paper]
 31. [ICML workshop 21] **Reimagining GNN Explanations with ideas from Tabular Data** [paper]
 32. [ICML workshop 21] **Towards Automated Evaluation of Explanations in Graph Neural Networks** [paper]
 33. [ICDM 2021] **GNEs: Learning to Explain Graph Neural Networks** [paper]
 34. [ICDM 2021] **GCN-SE: Attention as Explainability for Node Classification in Dynamic Graphs** [paper]
 35. [ICDM 2021] **Multi-objective Explanations of GNN Predictions** [paper]
 36. [CIKM 2021] **Towards Self-Explainable Graph Neural Network** [paper]
 37. [ECML PKDD 2021] **GraphSVX: Shapley Value Explanations for Graph Neural Networks** [paper]
 38. [WiseML 2021] **Explainability-based Backdoor Attacks Against Graph Neural Networks** [paper]
 39. [IJCNN 21] **MEG: Generating Molecular Counterfactual Explanations for Deep Graph Networks** [paper]
 40. [ICCSA 2021] **Understanding Drug Abuse Social Network Using Weighted Graph Neural Networks Explainer** [paper]
 41. [NeSy 21] **A New Concept for Explaining Graph Neural Networks** [paper]
 42. [Information Fusion 21] **Towards multi-modal causability with Graph Neural Networks enabling information fusion for explainable AI** [paper]
 43. [Patterns 21] **hcga: Highly Comparative Graph Analysis for network phenotyping** [paper]

Year 2020 and Before

1. [NeurIPS 2020] **Parameterized Explainer for Graph Neural Network**. [paper]
2. [NeurIPS 2020] **PGM-Explainer: Probabilistic Graphical Model Explanations for Graph Neural Networks** [paper]
3. [KDD 2020] **XGNN: Towards Model-Level Explanations of Graph Neural Networks** [paper]
4. [ACL 2020] **GCAN: Graph-aware Co-Attention Networks for Explainable Fake News Detection on Social Media**. paper
5. [Arxiv 2020] **Graph Neural Networks Including Sparse Interpretability** [paper]

-
6. [NeurIPS Workshop 20] **Towards explainable message passing networks for predicting carbon dioxide adsorption in metal-organic frameworks** [paper]
 7. [ICML workstop 2020] **Contrastive Graph Neural Network Explanation** [paper]
 8. [ICML workstop 2020] **Towards Explainable Graph Representations in Digital Pathology** [paper]
 9. [NeurIPS workshop 2020] **Explaining Deep Graph Networks with Molecular Counterfactuals** [paper]
 10. [DataMod 2020] **Exploring Graph-Based Neural Networks for Automatic Brain Tumor Segmentation”** [paper]
 11. [OpenReview 20] **A Framework For Differentiable Discovery Of Graph Algorithms** [paper]
 12. [OpenReview 20] **Causal Screening to Interpret Graph Neural Networks** [paper]
 13. [Arxiv 20] **Understanding Graph Neural Networks from Graph Signal Denoising Perspectives** [paper]
 14. [Arxiv 20] **Understanding the Message Passing in Graph Neural Networks via Power Iteration** [paper]
 15. [Arxiv 20] **xERTE: Explainable Reasoning on Temporal Knowledge Graphs for Forecasting Future Links** [paper]
 16. [IJCNN 20] **GCN-LRP explanation: exploring latent attention of graph convolutional networks** [paper]
 17. [CD-MAKE 20] **Explain Graph Neural Networks to Understand Weighted Graph Features in Node Classification** [paper]
 18. [ICDM 19] **Scalable Explanation of Inferences on Large Graphs**[paper]