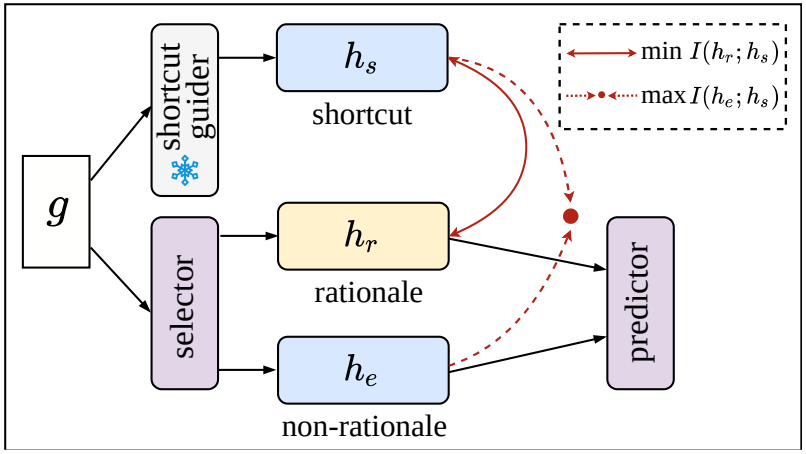
Learning from Shortcut: A Shortcut-guided Approach for Explainable Graph Learning

Linan YUE, Qi LIU, Ye LIU, Weibo GAO, Fangzhou YAO

Frontiers of Computer Science, DOI: 10.1007/s11704-024-40452-4

Problems & Ideas

- Problems of existing eXplainable Graph Learning (XGL):
  - Existing XGL approaches are susceptible to exploiting shortcuts (aka, spurious correlations) in the data to yield task results and compose explanations, undermining the trustworthiness and reliability of XGL.
  - Shortcuts in data are difficult to identify and mitigate.
- Ideas: The core idea is first using shortcut discovery strategies to identify potential shortcut representations and then developing shortcut-conflicted explanations for graph learning.



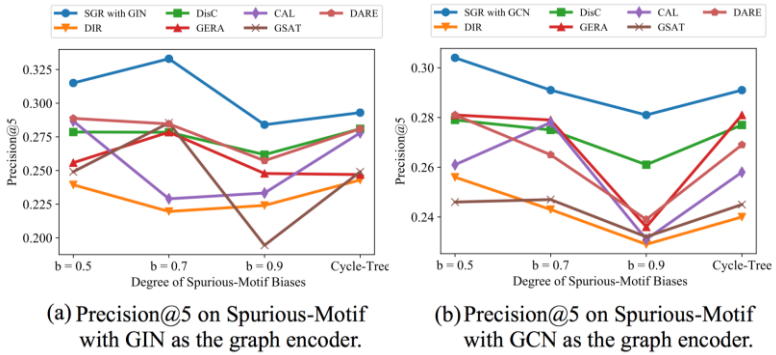
Architecture of Shortcut-guided Graph Rationalization (SGR). SGR use a shortcut discovery strategy to infer shortcut representations and guide the model to compose shortcut-conflicted explanations by mutual information estimation methods.

Main Contributions

- Contributions:
  - A novel Shortcut-guided Graph Rationalization method that explicitly learns from shortcuts in data to improve the robustness and reliability of graph learning models;
  - A shortcut discovery strategy that captures shortcut information with an early stop strategy and a learn-from-shortcuts strategy to ensure the composed explanations are free from shortcut biases;
  - Extensive experiments on synthetic and real-world datasets, demonstrating the effectiveness of SGR in providing faithful and robust explanations, outperforming existing baselines.

		MolHIV	MolToxCast	MolBACE	MolBBBP	MolSIDER
GIN is the backbone	GIN	0.7447 ± 0.0293	0.6521 ± 0.0172	0.8047 ± 0.0172	0.6584 ± 0.0224	0.5977 ± 0.0176
	DIR	0.6303 ± 0.0607	0.5451 ± 0.0092	0.7391 ± 0.0282	0.6460 ± 0.0139	0.4989 ± 0.0115
	DisC	0.7731 ± 0.0101	0.6662 ± 0.0089	0.8293 ± 0.0171	0.6963 ± 0.0206	0.5846 ± 0.0169
	RGDA	0.7714 ± 0.0153	0.6694 ± 0.0043	0.8187 ± 0.0195	0.6953 ± 0.0229	0.5864 ± 0.0052
	CAL	0.7339 ± 0.0077	0.6476 ± 0.0066	0.7848 ± 0.0107	0.6582 ± 0.0397	0.5965 ± 0.0116
	GSAT	0.7524 ± 0.0166	0.6174 ± 0.0069	0.7021 ± 0.0354	0.6722 ± 0.0197	0.6041 ± 0.0096
	DARE	0.7836 ± 0.0015	0.6677 ± 0.0058	0.8239 ± 0.0192	0.6820 ± 0.0246	0.5921 ± 0.0260
	SGR	**0.7945 ± 0.0071**	**0.6723 ± 0.0061**	**0.8305 ± 0.0098**	**0.7021 ± 0.0190**	**0.6092 ± 0.0288**
GCN is the backbone	GCN	0.7128 ± 0.0188	0.6497 ± 0.0114	0.8135 ± 0.0256	0.6665 ± 0.0242	0.6108 ± 0.0075
	DIR	0.4258 ± 0.1084	0.5077 ± 0.0094	0.7002 ± 0.0634	0.5069 ± 0.1099	0.5224 ± 0.0243
	DisC	0.7791 ± 0.0137	0.6626 ± 0.0055	0.8104 ± 0.0202	0.7061 ± 0.0105	0.6110 ± 0.0091
	RGDA	0.7816 ± 0.0079	0.6622 ± 0.0045	0.8044 ± 0.0063	0.6970 ± 0.0089	0.6133 ± 0.0239
	CAL	0.7501 ± 0.0094	0.6006 ± 0.0031	0.7802 ± 0.0207	0.6635 ± 0.0257	0.5559 ± 0.0151
	GSAT	0.7598 ± 0.0085	0.6124 ± 0.0082	0.7141 ± 0.0233	0.6437 ± 0.0082	0.6179 ± 0.0041
	DARE	0.7523 ± 0.0041	0.6618 ± 0.0065	0.8066 ± 0.0178	0.6823 ± 0.0068	0.6192 ± 0.0079
	SGR	**0.7822 ± 0.0079**	**0.6668 ± 0.0026**	**0.8228 ± 0.0283**	**0.7116 ± 0.0169**	**0.6217 ± 0.0291**

The graph classification ROC-AUC on real-world test datasets of OGBG The graph classification ROC-AUC on testing datasets of OGBG, where the test dataset is the out-of-distribution (OOD) test set.



(a) Precision@5 on Spurious-Motif with GIN as the graph encoder.

(b) Precision@5 on Spurious-Motif with GCN as the graph encoder.

The precision results of identifying the ground-truth explainable subgraphs on synthetic datasets.