Explainable Graph-based Fraud Detection via Neural Meta-graph Search

Zidi Qin, Yang Liu, Qing He, Xiang Ao*

Peering In

What drives the model to make predictions?

How to develop a fraud detector giving high-quality predictions and explanations simultaneously?

Our contributions:

1. Propose NGS to search the optimized message passing graph structure.
2. The meta-graphs offer explanations.

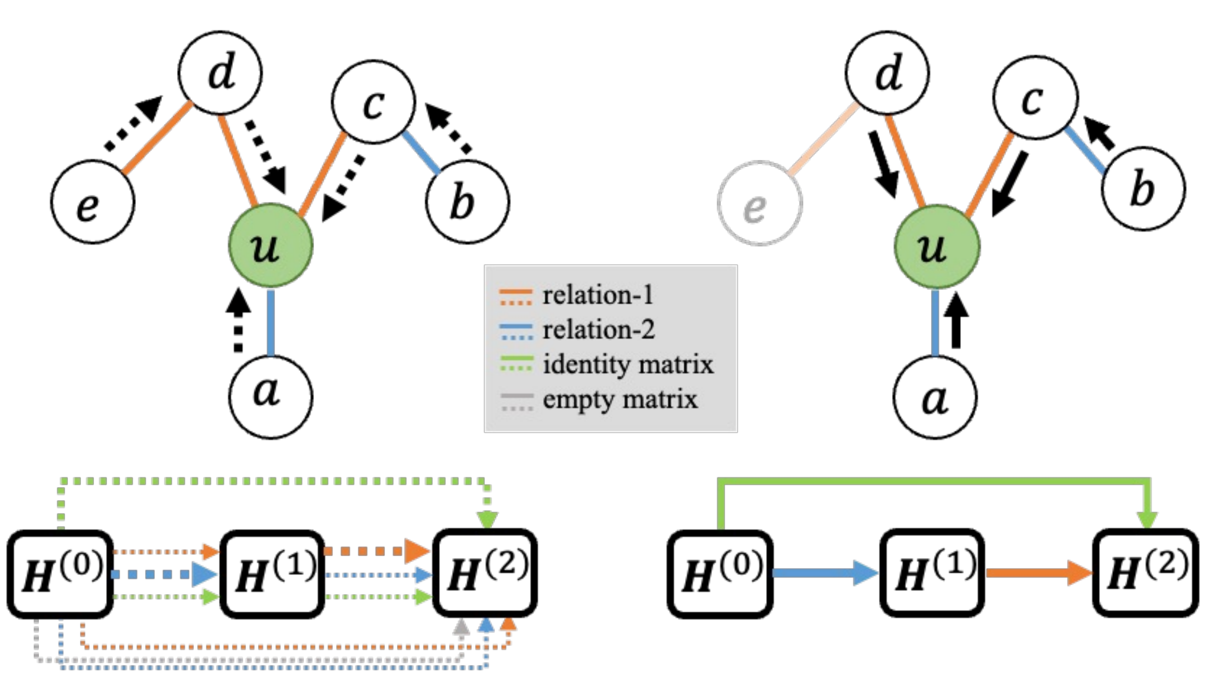
Methodology

The message passing scheme of GNN:

$$\boldsymbol{H}^{(0)} = \mathrm{MLP}\,(\boldsymbol{X}), \quad \boldsymbol{H}^{(l+1)} = \mathrm{Aggr}\left(\boldsymbol{H}^{(l)}; A\right)$$

Use meta-graph $M_{\mathcal{A}}$ to describe which relation the message is passed along, when dealing with multi-relation graph $\mathcal{A} = \{A_r\}|_{r=1}^{R}$

NGS



(a) (b)

We relax the discrete edge type selection to be continuous like DARTS.

$$f_{i,l}\left(\boldsymbol{H}^{(i)}; \mathcal{A}_{i,l}\right) = \sum_{A \in \mathcal{A}_{i,l}} \frac{\exp\left(\alpha_{i,l}^{A}\right)}{\sum_{A' \in \mathcal{A}_{i,l}} \exp\left(\alpha_{i,l}^{A'}\right)} \cdot \mathrm{Aggr}\left(\boldsymbol{H}^{(i)}; A\right)$$

Ref: Yuhui Ding, et al, KDD 2021. DiffMG: Differentiable Meta Graph Search for Heterogeneous Graph Neural Networks.
Hanxiao Liu, et al, ICLR 2019. DARTS: Differentiable Architecture Search.

Search space:

$$\mathcal{A}_{i,l} = \begin{cases} \mathcal{A} \cup \{I\} & l \le L \text{ and } i = l-1 \\ \mathcal{A} \cup \{I\} \cup \{O\} & l \le L \text{ and } i < l-1 \end{cases}$$

Optimization:

$$\min_{\alpha} \mathcal{L}_{\text{val}}\left(\omega^*(\alpha), \alpha\right), \ \text{s.t. } \omega^*(\alpha) = \arg\min_{\omega} \mathcal{L}_{\text{train}}(\omega, \alpha)$$

$$\mathcal{L} = -\sum_{v \in \mathcal{V}} \left[y_v \log p_v + (1 - y_v) \log (1 - p_v)\right]$$

Experiments

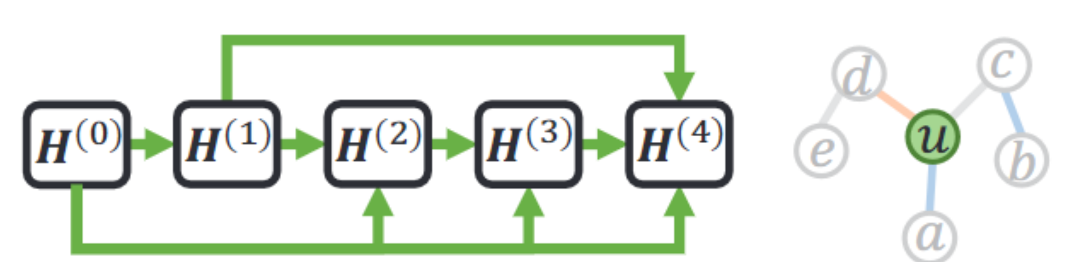
Dataset	#Nodes (Fraud%)	#Edges	Relation	#Relations
Amazon	11, 944 (9.5%)	4, 398, 392	U-P-U U-S-U U-V-U	175, 608 3, 566, 479 1, 036, 737
YelpChi	45, 954 (14.5%)	3, 846, 979	R-U-R R-T-R R-S-R	49, 315 573, 616 3, 402, 743

Two real-world graph-based fraud detection datasets Amazon and YelpChi are adopted to validate NGS's performance.

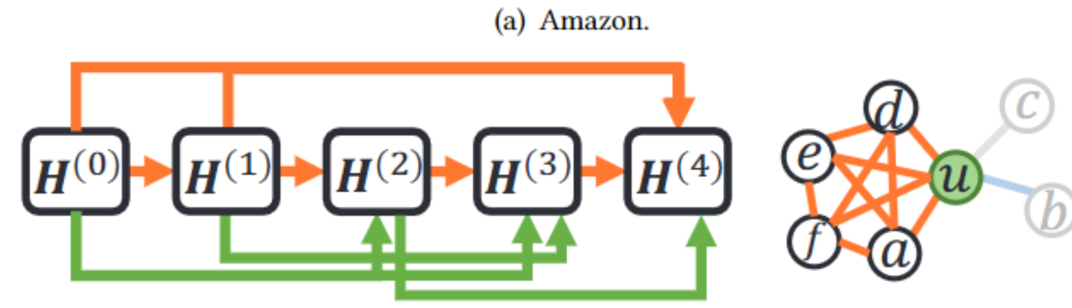
Method	Dataset	Amazon			YelpChi		
	Metric	F1-macro	AUC	GMean	F1-macro	AUC	GMean
Baselines	GCN	0.6571±0.0008	0.8189±0.0008	0.6629±0.0037	0.4963±0.0005	0.5504±0.0001	0.2143±0.0019
	GAT	0.5390±0.0021	0.7426±0.0020	0.3081±0.0173	0.5228±0.0070	0.5519±0.0012	0.2921±0.0193
	GraphSAGE	0.8383±0.0109	0.9149±0.0077	0.8518±0.0077	0.5781±0.0239	0.7409±0.0034	0.6815±0.0049
	CARE-GNN	0.8997±0.0064	0.9482±0.0044	0.8982±0.0015	0.6052±0.0170	0.7748±0.0008	0.7071±0.0035
	PC-GNN	0.8660±0.0164	0.9642±0.0035	0.8986±0.0203	0.6192±0.0479	0.8104±0.0057	0.7225±0.0166
	FRAUDRE	0.8519±0.1055	0.9408±0.0052	0.8847±0.0280	0.6057±0.0381	0.7582±0.0041	0.6862±0.0128
	AO-GNN*	0.8921±0.0045	0.9640±0.0020	0.9096±0.0105	0.7042±0.0051	0.8805±0.0008	0.8134±0.0232
	H^2-FDetector*	0.8392±0.0000	0.9689±0.0000	0.9203±0.0000	0.6944±0.0000	0.8877±0.0000	0.816±0.0000
	ProtGNN	0.7351±0.0112	0.8826±0.0106	0.7785±0.0126	0.5663 ±0.0024	0.6004 ±0.0056	0.4595 ±0.0196
	DiffMG	0.8826±0.0049	0.9290±0.0044	0.8855±0.0057	0.7316±0.0144	0.8799±0.0142	0.7873±0.0147
Ablation	$\mathrm{NGS}_{\backslash A}$	**0.9234±0.0078**	0.9692±0.0136	0.9191±0.0087	0.7604±0.0227	0.9009±0.0215	0.7981±0.0279
Ours	NGS	0.9228±0.0046	**0.9736±0.0035**	**0.9218±0.0042**	**0.7828±0.0055**	**0.9218±0.0032**	**0.8351±0.0056**

Compared with various baselines, NGS exceeding or matching performance across all of them.

Explainability



(a) Amazon.



(b) YelpChi.

Amazon

No relation is involved;
User attributes are the key to identifying fraudsters.

Yelpchi

R-U-R relation is highly relevant to fraud detection;
Suggesting a typical default phenomenon: click farming.

The research work is supported by the National Natural Science Foundation of China under Grant (No.61976204, 92046003, U1811461). Xiang Ao is also supported by the Project of Youth Innovation Promotion Association CAS, Beijing Nova Program Z201100006820062. Yang Liu is also supported by China Scholarship Council.