Revisiting Score Propagation in Graph Out-of-Distribution Detection

Longfei Ma Yiyou Sun Kaize Ding Zemin Liu Fei Wu

浙江大学 ZHEJIANG UNIVERSITY

WISCONSIN UNIVERSITY OF WISCONSIN-MADISON

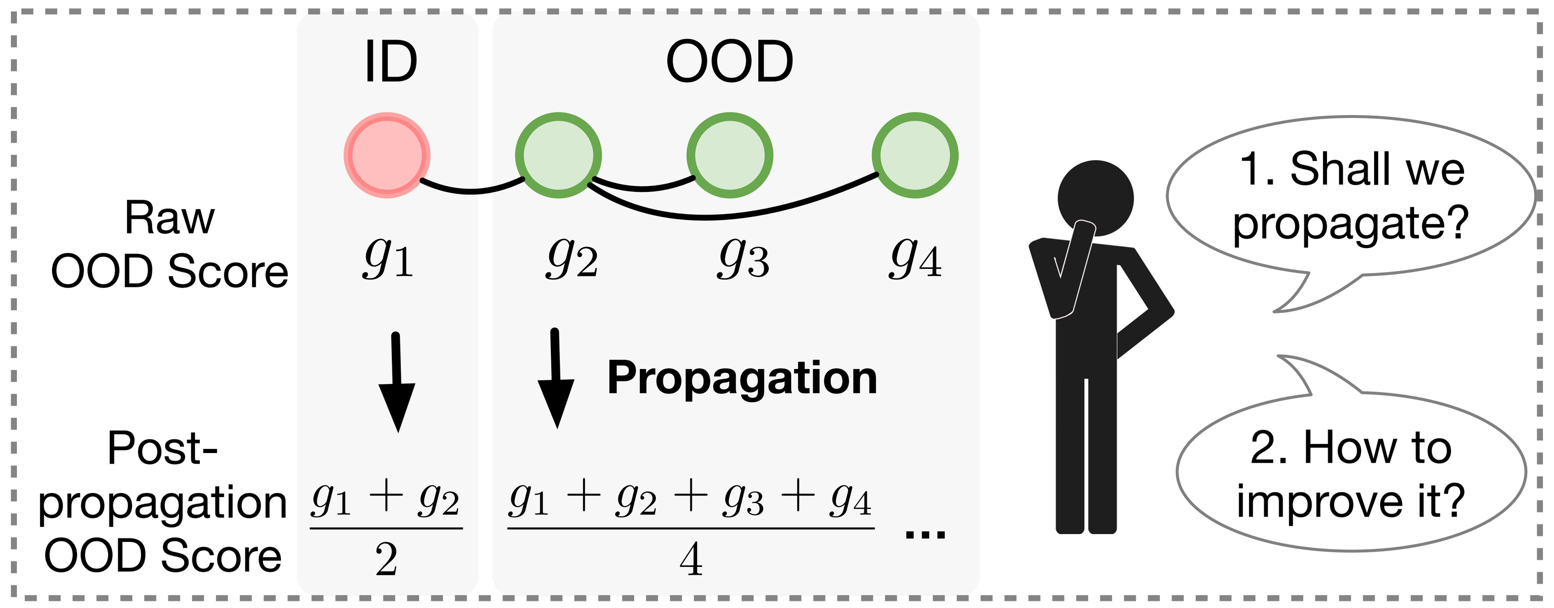
Northwestern University

NEURAL INFORMATION PROCESSING SYSTEMS

Paper Link: https://openreview.net/forum?id=jb5qN3212b Code Link: https://github.com/longfei-ma/GRASP

Introduction

In this paper, we study **Score Propagation** in the Graph Out-of-Distribution problem.

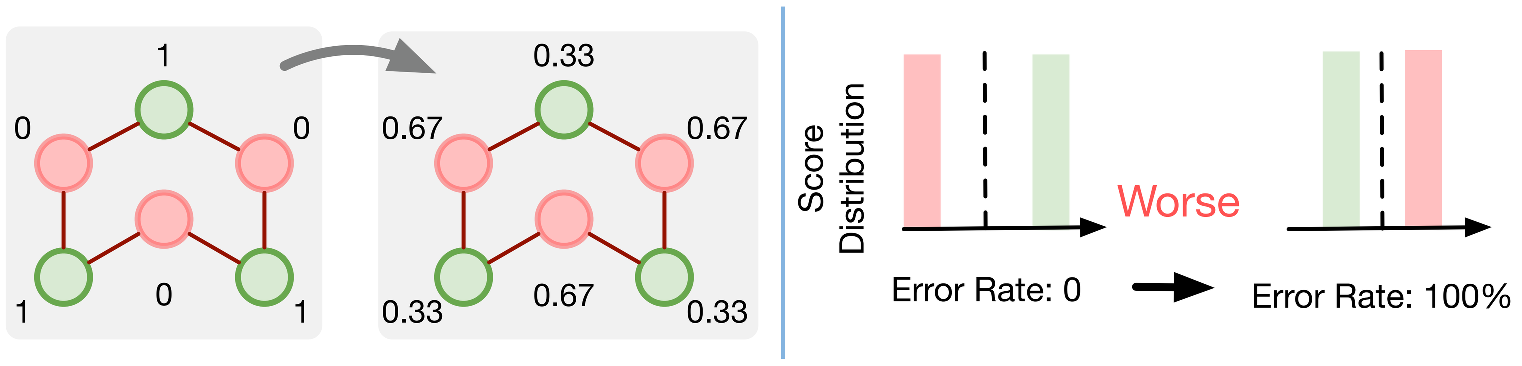


We present a theoretical framework and provide new insight on the research question:

"when does Score Propagation help detect graph OOD nodes and how to improve it?"

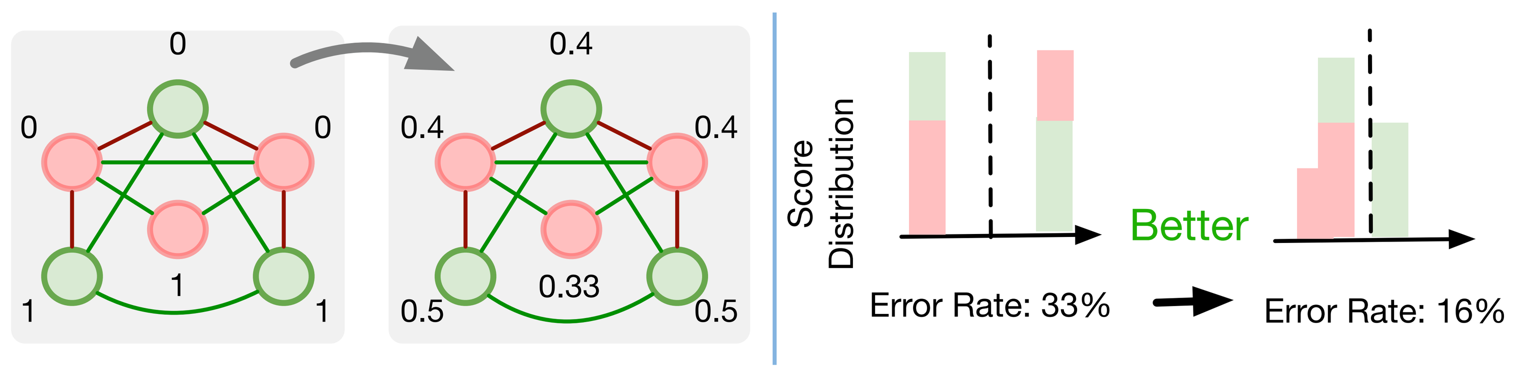
An Illustrative Example

Case 1: Score Propagation (with More Inter-edges)



(a) The case when propagation is harmful.

Case 2: Score Propagation (with More Intra-edges)



(b) The case when propagation is helpful.

Observation: Score Propagation mail fail when inter-edges dominate!

Theoretical Findings

Theorem 3.2. *(Formal) For any two test ID/OOD node set $S_{id} \subset \mathcal{V}_{uid}, S_{ood} \subset \mathcal{V}_{uood}$ with equal size N_s, let the ID-vs-OOD separability $\mathcal{M}_{sep}$ defined on an OOD scoring vector $\hat{\mathbf{g}} \in \mathbb{R}^N$ as*

$$\mathcal{M}_{sep}(\hat{\mathbf{g}}) \triangleq \mathbb{E}_{i \in S_{id}} \hat{\mathbf{g}}_i - \mathbb{E}_{j \in S_{ood}} \hat{\mathbf{g}}_j.$$

If $\mathcal{M}_{sep}(\hat{\mathbf{g}}) > 0$ and $\eta_{intra} - \eta_{inter} > 1/N_s$, for some $\epsilon > 0$ and constant c, we have $\mathbb{P}\left(\mathcal{M}_{sep}(A\hat{\mathbf{g}}) \geq \mathcal{M}_{sep}(\hat{\mathbf{g}}) - \epsilon\right) \geq 1 - exp(-\frac{c\epsilon^2}{\|\hat{\mathbf{g}}\|_2^2})$.

Finding 1: Score Propagation only works when intra-edges dominate

Theorem 4.2. *(Formal) For any two test ID/OOD node set $S_{id} \subset \mathcal{V}_{uid}, S_{ood} \subset \mathcal{V}_{uood}$ with size N_s, let the ID-vs-OOD separability $\mathcal{M}_{sep}$ defined on a non-negative OOD scoring vector $\hat{\mathbf{g}} \in \mathbb{R}^N$ as*

$$\mathcal{M}_{sep}(\hat{\mathbf{g}}) \triangleq \mathbb{E}_{i \in S_{id}} \hat{\mathbf{g}}_i - \mathbb{E}_{j \in S_{ood}} \hat{\mathbf{g}}_j.$$

Let $\mathcal{E}_{S \leftrightarrow S'} \subset \mathcal{E}$ to denote the edge set of edges between two node sets S and S', where $S, S' \subset \mathcal{V}$. If we can find a node set $G \subset \mathcal{V}_l$ such that $|\mathcal{E}_{G \leftrightarrow S_{id}}| > |\mathcal{E}_{G \leftrightarrow S_{ood}}|$, we have

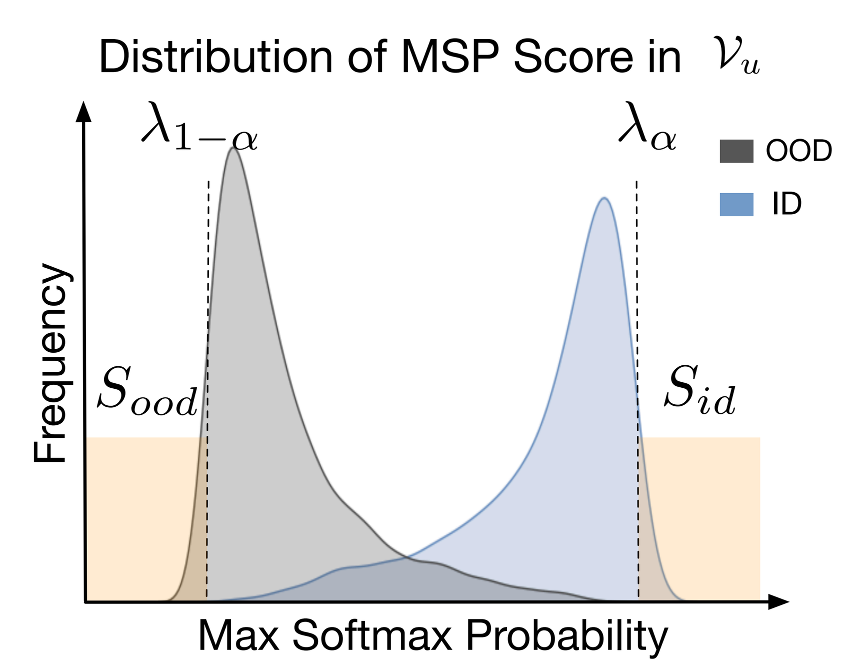
$$\mathcal{M}_{sep}((A + \delta E)^2 \hat{\mathbf{g}}) > \mathcal{M}_{sep}(A^2 \hat{\mathbf{g}}),$$

where $E = \mathbf{e}_G \mathbf{e}_G^\top$ and $\delta > 0$.

Finding 2: Score Propagation along augmentated graph by adding edges within *G* can boost OOD detection performance

Methodology

Selection of *G*



Step 1: Selection of *Sid* and *Sood*

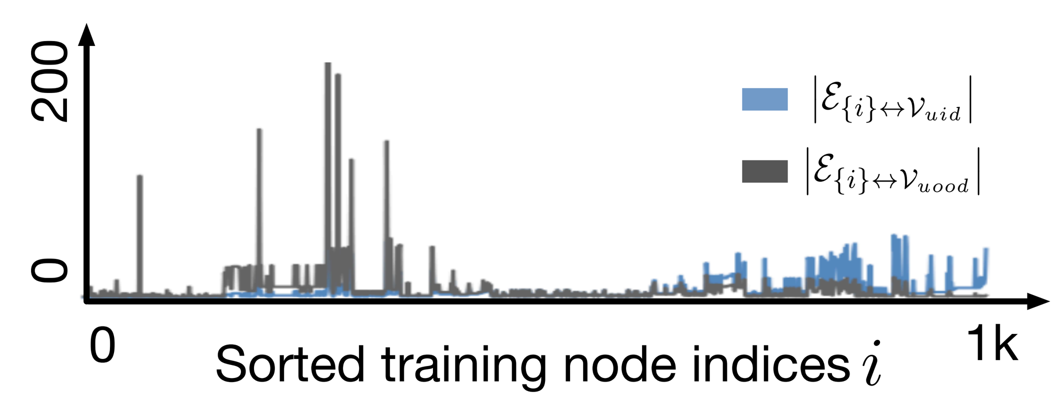
$$S_{id} = \{i \in \mathcal{V}_u | \max_{c \in [C]} f_c(i) > \lambda_\alpha\}$$

$$S_{ood} = \{j \in \mathcal{V}_u | \max_{c \in [C]} f_c(j) < \lambda_{100-\alpha}$$

Step 2: Selection of *G*

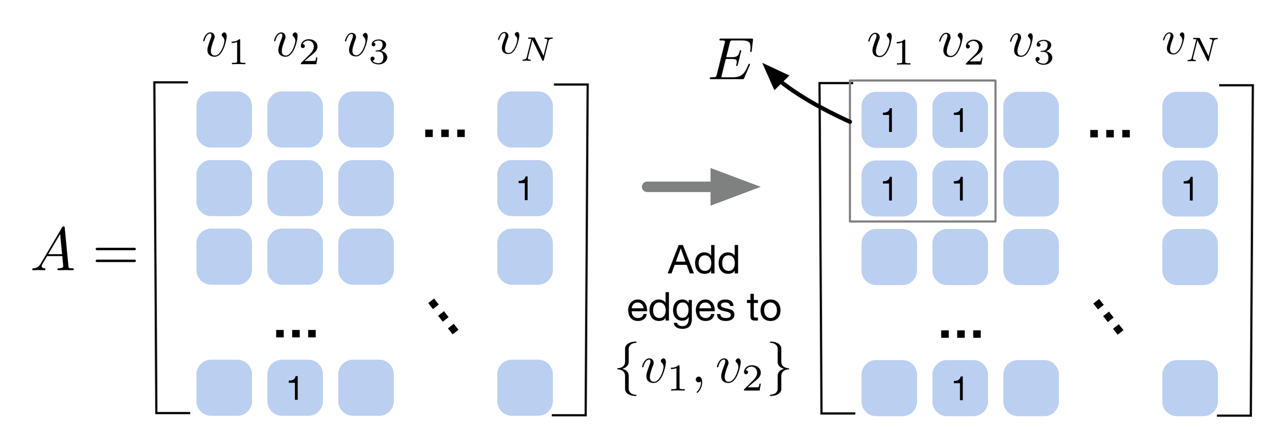
$$h(i) = |\mathcal{E}_{\{i\} \leftrightarrow S_{id}}| / (|\mathcal{E}_{\{i\} \leftrightarrow S_{ood}}| + 1)$$

$$G = \{i \in \mathcal{V}_l | h(i) > \tau_\beta\}$$



***G* are the nodes on the right of the axis**

Augmentated Score Propagation



Step 3: Add edges within G

Step 4: Score Propagation along augmentated graph

$$g_{GRASP} = (\bar{A}_+)^k \hat{g}$$

Experiments

Main Results

Method	Cora		Amazon		Coauthor		Chameleon		Squirrel		Average	
	FPR↓	AUROC↑	FPR↓	AUROC↑	FPR↓	AUROC↑	FPR↓	AUROC↑	FPR↓	AUROC↑	FPR↓	AUROC↑
MSP	70.86	84.56	49.26	89.34	28.82	94.34	85.70	57.96	94.68	48.51	65.86	74.94
Energy	67.54	85.47	42.13	90.28	20.29	95.67	88.06	59.20	93.98	45.07	62.40	75.14
KNN	90.20	70.94	65.19	84.71	51.24	90.13	93.38	57.90	94.72	54.68	78.95	71.67
ODIN	68.41	84.98	44.06	89.90	22.59	95.27	85.31	57.94	94.17	44.08	62.91	74.43
Mahalanobis	69.68	85.48	96.49	75.58	85.71	84.98	95.55	53.19	94.90	54.99	88.47	70.84
GKDE	63.71	86.27	81.29	77.26	25.48	95.13	92.93	50.14	96.71	49.38	72.02	71.64
GPN	58.45	82.93	72.95	82.63	34.11	93.82	82.25	68.20	95.58	48.38	68.67	75.19
OODGAT	94.59	53.63	71.34	66.95	96.53	52.18	94.43	59.67	95.27	46.13	90.43	55.71
GNNSafe	54.71	87.52	22.39	96.27	16.64	95.82	100.00	50.42	100.00	35.88	58.75	73.18
GRASP (Ours)	**29.70**	**93.50**	**14.38**	**96.68**	**7.84**	**97.75**	**66.88**	**76.93**	**85.59**	**61.09**	**40.88**	**85.19**

Method	reddit2		ogbn-products		arxiv-year		snap-patents		wiki		Average	
	FPR↓	AUROC↑	FPR↓	AUROC↑	FPR↓	AUROC↑	FPR↓	AUROC↑	FPR↓	AUROC↑	FPR↓	AUROC↑
MSP	96.59	46.61	86.87	70.19	95.03	47.24	94.31	46.99	95.46	54.70	93.65	53.15
Energy	96.77	44.13	85.09	68.13	94.10	51.35	96.82	46.03	97.31	29.02	94.02	47.73
KNN	90.78	66.74	84.22	73.58	95.35	57.96	90.54	53.45	93.43	43.69	90.86	59.08
ODIN	96.74	44.69	85.65	68.95	95.06	47.36	94.27	45.20	97.88	29.91	93.92	47.22
Mahalanobis	71.73	74.89	OOM	OOM	88.60	59.57	96.03	58.50	72.33	67.95	82.17	65.23
GKDE	OOT	OOT	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	-	-
GPN	OOM	OOM	OOM	OOM	95.62	50.97	OOM	OOM	OOM	OOM	95.62	50.97
OODGAT	OOM	OOM	OOM	OOM	92.90	59.38	OOM	OOM	OOM	OOM	92.90	59.38
GNNSafe	99.49	31.99	77.86	85.66	100.00	35.30	99.92	27.35	72.63	60.32	89.98	48.12
GRASP (Ours)	**2.41**	**98.50**	**39.77**	**93.79**	**73.93**	**81.24**	**75.22**	**72.13**	**58.49**	**77.97**	**49.96**	**84.73**

GRASP also applys to other score functions

Method	Cora	Amazon	Coauth	Chamel	Squirr
MSP	84.56	89.34	94.34	57.96	48.51
MSP+prop	88.02	95.32	97.15	50.35	36.21
MSP+GRASP	**93.50**	**96.68**	**97.75**	**76.93**	**61.09**
Energy	85.47	90.28	95.67	59.20	45.07
Energy+prop	87.52	96.27	95.82	50.42	36.49
Energy+GRASP	**88.34**	**96.35**	**96.64**	**62.04**	**60.66**
KNN	70.94	84.71	90.13	57.90	54.68
KNN+prop	73.70	92.36	95.47	49.76	53.99
KNN+GRASP	**91.48**	**97.43**	**96.52**	**76.32**	**60.24**