Spreading Out-of-Distribution on Graphs

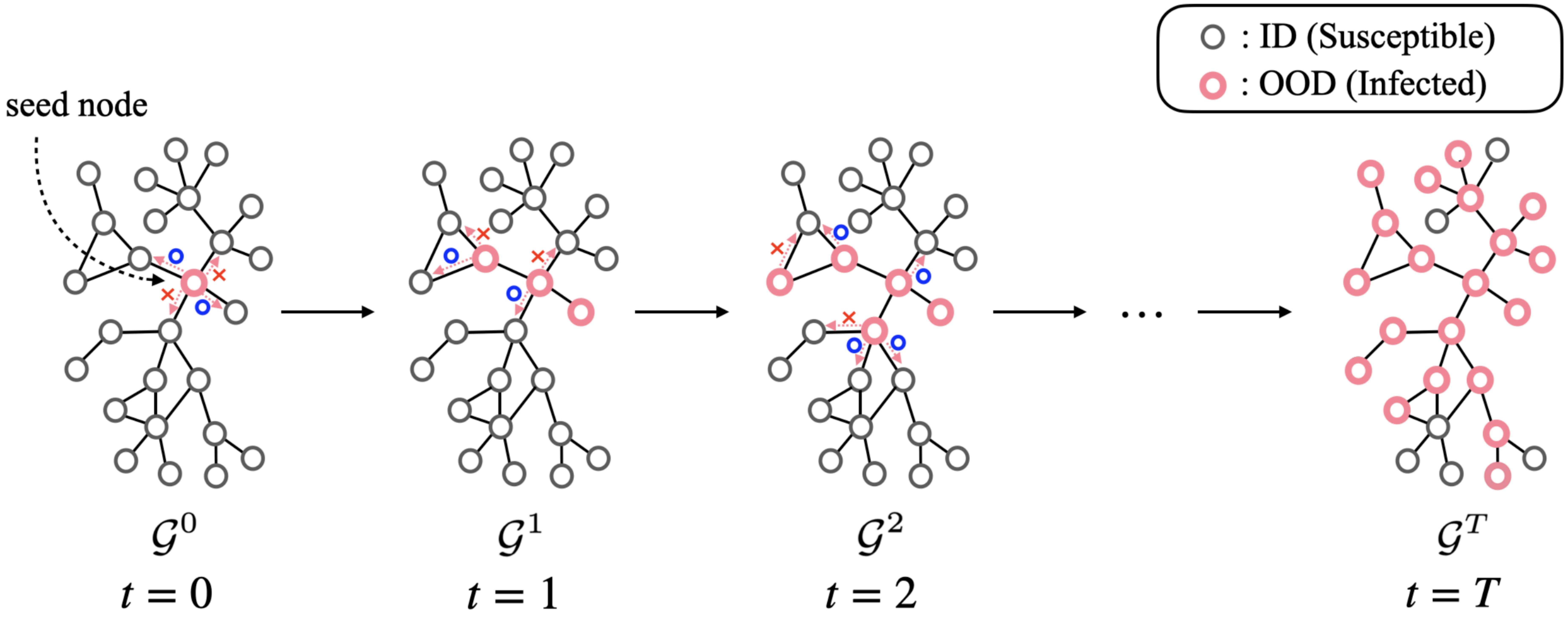
Daeho Um[1], Jongin Lim[1], Sunoh Kim[2], Yuneil Yeo[3], Yoonho Jung[4]

[1]AI Center, Samsung Electronics, [2]Dankook University, [3]UC Berkeley, [4]Seoul National University

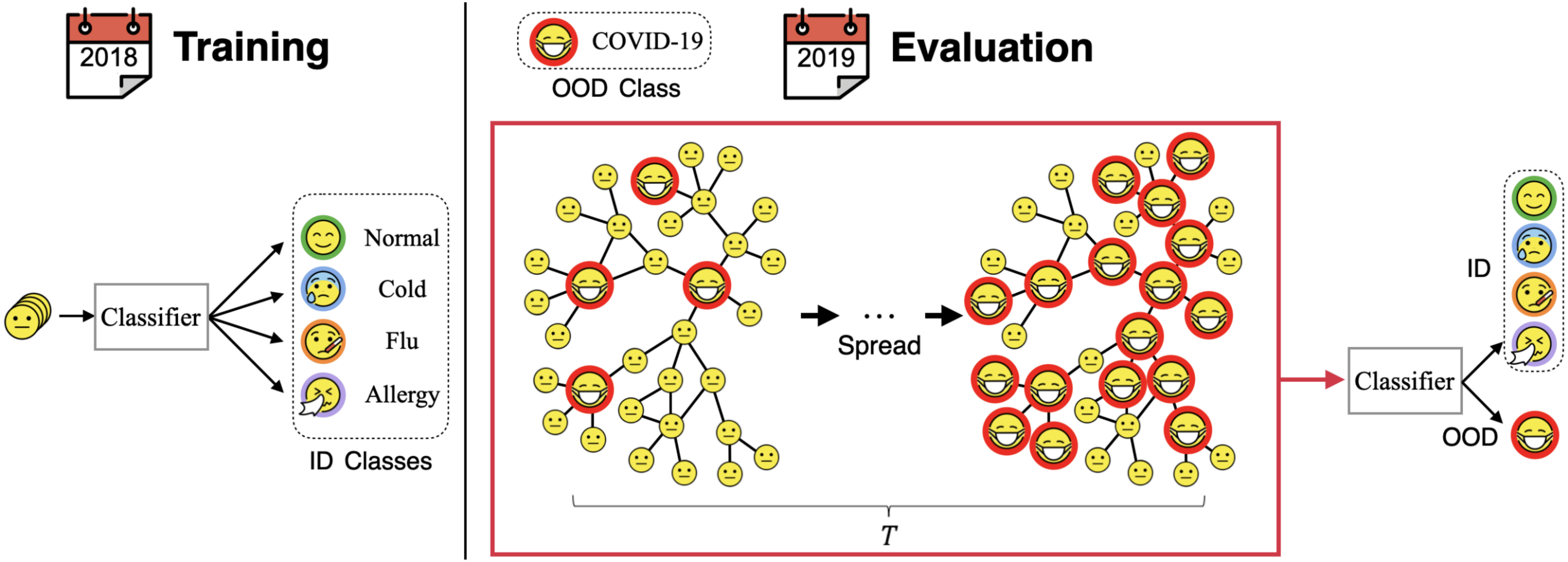
Node-Level OOD Detection Benchmark

- Node-level OOD detection methods are commonly evaluated on synthetic graph datasets where OOD samples are assigned to randomly selected nodes.
- Nodes belonging to randomly selected classes or randomly selected nodes are designated as OOD.
- These previous evaluation scenarios fail to reflect interactions with the various nodes associated with OOD samples.

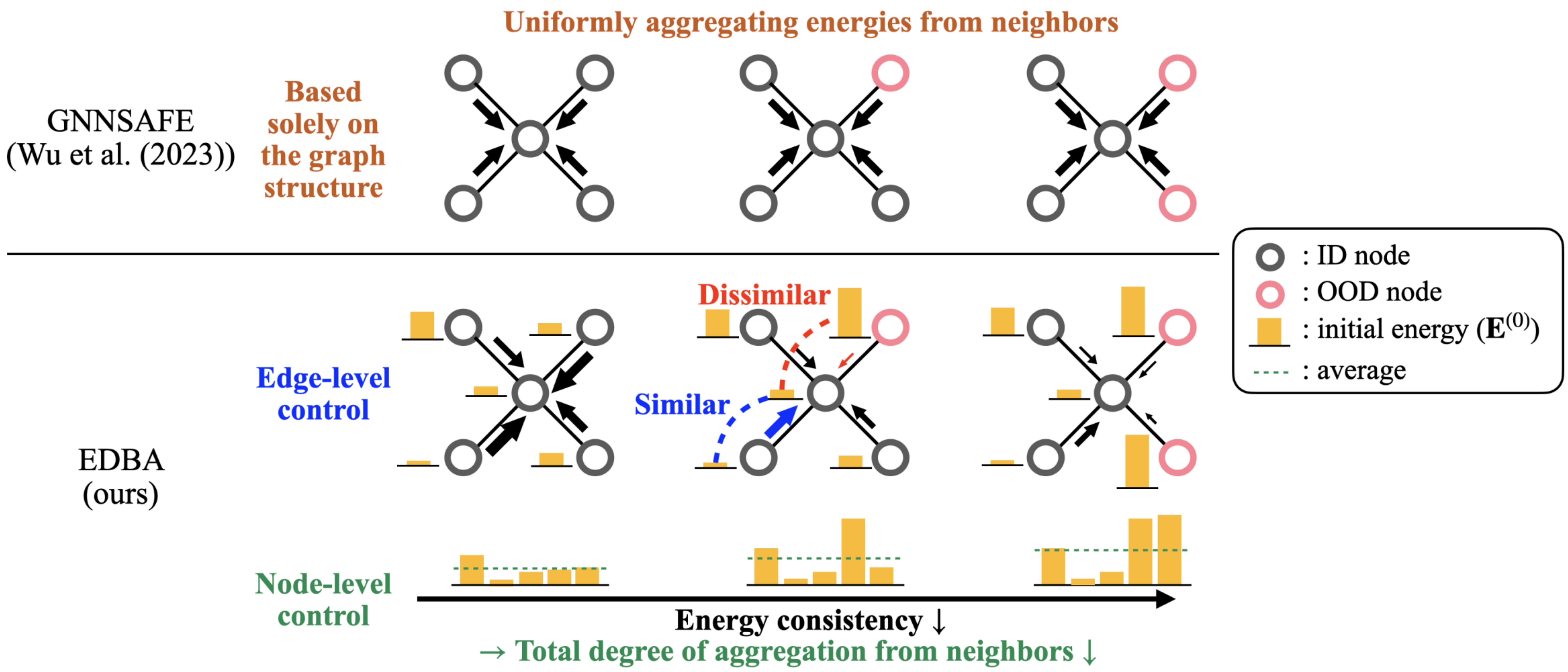
Spreading OOD Detection



Spreading COVID-19 Dataset



Energy Distribution-Based Aggregation (EDBA)



Experiments

Label leave-out

Dataset	Cora			Amazon-Photo		
Method	FPR95(↓)	AUROC(↑)	AUPR(↑)	FPR95(↓)	AUROC(↑)	AUPR(↑)
MSP	40.37 ± 2.19	91.13 ± 0.22	78.16 ± 0.19	28.87 ± 1.65	94.41 ± 0.61	92.44 ± 0.75
ODIN	100.00 ± 0.00	49.05 ± 0.57	24.18 ± 0.08	92.72 ± 8.43	63.30 ± 7.61	51.72 ± 7.43
Mahalanobis	86.11 ± 6.19	66.93 ± 1.95	40.56 ± 3.77	56.11 ± 16.06	82.51 ± 4.83	75.73 ± 7.09
Energy	38.36 ± 3.46	91.46 ± 0.33	78.10 ± 0.29	30.49 ± 3.93	93.96 ± 0.68	91.73 ± 0.75
GKDE	60.88 ± 2.25	87.15 ± 0.60	72.12 ± 1.10	91.60 ± 8.81	60.00 ± 11.43	56.61 ± 12.77
GPN	44.04 ± 5.85	87.48 ± 6.38	81.21 ± 7.40	35.54 ± 11.48	91.48 ± 2.71	88.04 ± 3.41
OODGAT	85.21 ± 1.66	64.81 ± 0.87	62.65 ± 1.01	13.33 ± 0.46	97.27 ± 0.33	95.01 ± 0.60
GNNSAFE	31.31 ± 1.11	92.84 ± 0.38	82.22 ± 0.40	6.57 ± 0.38	97.36 ± 0.04	97.13 ± 0.10
EDBD	$\mathbf{30.48 \pm 1.11}$	$\mathbf{92.95 \pm 0.38}$	$\mathbf{82.31 \pm 0.38}$	$\mathbf{5.82 \pm 0.66}$	$\mathbf{97.48 \pm 0.07}$	$\mathbf{97.60 \pm 0.06}$

Dataset	Amazon-Computers			Coauthor-CS		
Method	FPR95(↓)	AUROC(↑)	AUPR(↑)	FPR95(↓)	AUROC(↑)	AUPR(↑)
MSP	70.77 ± 3.54	76.81 ± 2.31	71.01 ± 2.23	29.07 ± 3.53	94.15 ± 0.73	97.73 ± 0.28
ODIN	98.72 ± 2.56	53.36 ± 1.91	45.93 ± 2.83	100.00 ± 0.00	52.35 ± 4.36	75.26 ± 1.96
Mahalanobis	71.09 ± 1.90	73.14 ± 1.27	62.63 ± 2.43	64.40 ± 14.83	81.73 ± 3.53	84.30 ± 15.80
Energy	58.40 ± 3.41	84.72 ± 1.50	79.36 ± 1.66	18.60 ± 3.87	95.98 ± 0.71	98.39 ± 0.29
GKDE	90.64 ± 7.22	58.59 ± 10.46	49.23 ± 8.10	59.70 ± 7.83	88.02 ± 1.77	95.50 ± 0.65
GPN	80.55 ± 16.98	74.08 ± 15.09	69.27 ± 17.30	26.68 ± 11.56	93.54 ± 3.35	97.40 ± 1.47
OODGAT	86.16 ± 7.35	73.55 ± 5.48	84.17 ± 2.99	13.16 ± 1.13	96.83 ± 0.21	96.58 ± 0.10
GNNSAFE	39.94 ± 6.84	89.75 ± 1.79	85.63 ± 3.36	11.31 ± 1.69	97.44 ± 0.35	99.06 ± 0.13
EDBD	$\mathbf{35.59 \pm 6.94}$	$\mathbf{92.45 \pm 0.98}$	$\mathbf{90.34 \pm 1.45}$	$\mathbf{10.19 \pm 1.63}$	$\mathbf{97.68 \pm 0.35}$	$\mathbf{99.14 \pm 0.13}$

Experiments

Spreading OOD detection in the COVID-19 dataset

Single-seed setting

Epidemic model	SI			SIS		
Method	FPR95-T (↓)	AUROC-T (↑)	AUPR-T (↑)	FPR95-T (↓)	AUROC-T (↑)	AUPR-T (↑)
MSP	94.25 ± 4.56	38.51 ± 5.41	70.17 ± 4.55	97.98 ± 1.38	34.45 ± 5.87	70.01 ± 2.73
ODIN	93.19 ± 5.00	61.82 ± 11.50	78.08 ± 6.05	89.87 ± 5.76	65.86 ± 10.63	80.26 ± 5.54
Mahalanobis	63.64 ± 44.93	47.29 ± 42.24	79.46 ± 15.28	72.71 ± 36.33	46.61 ± 34.07	78.95 ± 12.00
Energy	54.82 ± 17.43	80.46 ± 8.54	88.00 ± 5.52	67.94 ± 18.98	73.94 ± 14.49	85.51 ± 5.70
GPN	95.29 ± 2.36	50.35 ± 13.69	79.55 ± 65.45	87.07 ± 14.50	67.20 ± 12.68	83.03 ± 4.40
GNNSAFE	69.76 ± 15.23	80.65 ± 6.07	87.40 ± 4.12	77.76 ± 14.64	76.02 ± 12.65	85.53 ± 5.30
EDBD	$\mathbf{54.67 \pm 17.58}$	$\mathbf{81.60 \pm 8.35}$	$\mathbf{88.22 \pm 5.59}$	$\mathbf{67.42 \pm 19.75}$	$\mathbf{76.22 \pm 14.58}$	$\mathbf{85.94 \pm 5.94}$

Multi-seed setting

Epidemic model	SI			SIS		
Method	FPR95-T (↓)	AUROC-T (↑)	AUPR-T (↑)	FPR95-T (↓)	AUROC-T (↑)	AUPR-T (↑)
MSP	96.29 ± 1.94	36.41 ± 5.86	58.63 ± 3.18	96.14 ± 1.87	37.49 ± 5.36	55.23 ± 2.98
ODIN	92.61 ± 3.42	63.15 ± 12.07	72.47 ± 6.61	91.93 ± 4.69	64.59 ± 10.37	69.05 ± 6.80
Mahalanobis	90.44 ± 0.53	27.93 ± 0.67	53.78 ± 0.95	90.55 ± 0.50	31.09 ± 0.63	54.08 ± 0.62
Energy	49.78 ± 14.73	82.66 ± 6.79	85.94 ± 4.92	60.40 ± 10.33	80.18 ± 6.24	81.63 ± 4.84
GPN	87.72 ± 6.17	62.81 ± 9.26	75.13 ± 5.38	89.29 ± 5.18	61.85 ± 8.69	71.39 ± 5.29
GNNSAFE	63.42 ± 14.09	82.21 ± 5.91	$\mathbf{86.11 \pm 4.33}$	63.32 ± 1.14	77.80 ± 5.16	80.40 ± 3.96
EDBD	$\mathbf{49.62 \pm 15.22}$	$\mathbf{83.29 \pm 6.68}$	$\mathbf{86.11 \pm 5.01}$	$\mathbf{59.93 \pm 10.47}$	$\mathbf{80.68 \pm 6.16}$	$\mathbf{81.70 \pm 4.92}$

Experiments

Spreading OOD detection in benchmark datasets

Epidemic model	Dataset	Cora			LastFM Asia		
	Method	FPR95-T (↓)	AUROC-T (↑)	AUPR-T (↑)	FPR95-T (↓)	AUROC-T (↑)	AUPR-T (↑)
SI	MSP	24.80 ± 16.38	96.04 ± 2.26	97.73 ± 1.13	52.98 ± 0.41	93.71 ± 1.80	97.96 ± 0.44
	ODIN	100.00 ± 0.00	45.65 ± 10.85	74.03 ± 4.70	100.00 ± 0.00	17.35 ± 6.96	68.61 ± 4.08
	Mahalanobis	65.32 ± 44.90	65.50 ± 32.93	88.21 ± 11.85	59.29 ± 48.43	87.34 ± 15.97	96.66 ± 3.92
	Energy	22.79 ± 18.21	96.36 ± 2.12	98.03 ± 1.07	54.44 ± 16.82	93.56 ± 1.71	97.86 ± 0.45
	GPN	100.00 ± 0.00	53.71 ± 4.38	73.93 ± 2.67	100.00 ± 0.00	19.70 ± 9.19	75.05 ± 18.16
	GNNSAFE	31.15 ± 10.74	93.45 ± 2.16	97.59 ± 0.71	64.83 ± 8.52	83.73 ± 4.22	93.94 ± 1.05
	EDBD	$\mathbf{14.99 \pm 14.53}$	$\mathbf{97.54 \pm 1.81}$	$\mathbf{98.68 \pm 0.85}$	$\mathbf{46.68 \pm 14.83}$	$\mathbf{94.10 \pm 1.71}$	$\mathbf{98.11 \pm 0.41}$
SIS	MSP	40.47 ± 8.78	92.91 ± 1.34	93.22 ± 1.59	57.19 ± 18.25	91.48 ± 2.88	94.26 ± 1.43
	ODIN	95.52 ± 8.97	52.14 ± 11.24	69.25 ± 6.79	100.00 ± 0.00	20.71 ± 9.38	61.54 ± 5.78
	Mahalanobis	74.28 ± 40.81	68.97 ± 25.99	88.02 ± 10.64	86.86 ± 28.32	83.20 ± 6.94	$\mathbf{94.87 \pm 1.96}$
	Energy	38.95 ± 9.16	93.27 ± 1.34	93.59 ± 1.55	58.12 ± 18.99	91.08 ± 3.27	94.07 ± 1.57
	GPN	100.00 ± 0.00	47.38 ± 4.39	69.93 ± 6.55	100.00 ± 0.00	22.41 ± 13.41	87.30 ± 9.84
	GNNSAFE	47.55 ± 11.24	90.91 ± 2.78	93.19 ± 1.54	73.94 ± 8.07	82.73 ± 4.55	90.17 ± 1.64
	EDBD	$\mathbf{32.61 \pm 7.18}$	$\mathbf{94.17 \pm 1.22}$	$\mathbf{94.19 \pm 1.39}$	$\mathbf{52.37 \pm 19.93}$	$\mathbf{91.78 \pm 3.04}$	94.40 ± 1.52

Conclusion

✓ We introduce spreading OOD detection, a problem that facilitates the evaluation of node-level OOD detection under realistic settings.

✓ To highlight the significance of this problem, we then present the Spreading COVID-19 dataset.

✓ We propose EDBD that achieves state-of-the-art performance in both spreading OOD detection and conventional node-level OOD detection tasks.

✓ We hope that our work will inspire the creation of datasets for practical node-level OOD detection research within the machine learning community.