Supplementary Material: Faithful Explanations for Graph Classification using Logic

Alessio Ragno[1], Marc Plantevit[2], and Céline Robardet[1]

[1] INSA Lyon, CNRS, LIRIS UMR 5205, F-69621 Villeurbanne, France
{alessio.ragno,celine.robardet}@insa-lyon.fr
[2] EPITA Research Laboratory (LRE), FR-94276, Le Kremlin-Bicêtre, France
marc.plantevit@epita.fr

1 Models hyperparameters and accuracy

Dataset	Accuracy	Batch size	Dropout	Epochs	Hid. Dim	L2	LR
Ba2Motifs	100.00 ± 0.00	128	0	3000	16	1e-4	0.001
MUTAG	84.74 ± 8.02	128	0.5	3000	32	1e-4	0.001
Mutagenicity	82.21 ± 1.87	32	0.5	3000	32	1e-4	0.001
NCI1	81.87 ± 1.39	128	0	3000	32	1e-4	0.001
BBBP	87.95 ± 2.07	32	0.5	3000	16	1e-4	0.001
PROTEINS	72.59 ± 3.04	32	0	3000	32	1e-4	0.001

Table 1: Performance and hyperparameters for selected datasets

2   Explanation Performances

Table 2: Fidelity Scores when using the top 1, top 3 and top 5 nodes as hard mask. Values are reported as $\mu \pm \sigma$ over 5 seeds. The highest result is highlighted in bold, while values within one σ of the best result are underlined.

	BA2Motifs			MUTAG		
Top Nodes	1	3	5	1	3	5
GNNExpl.	5.0 ± 3.0	13.2 ± 5.5	21.0 ± 5.8	18.9 ± 10.8	42.1 ± 9.4	$\underline{57.9 \pm 13.3}$
PGExpl.	24.8 ± 3.3	24.6 ± 3.8	24.0 ± 5.5	23.2 ± 15.8	34.7 ± 17.5	40.0 ± 19.6
IG	16.4 ± 13.2	17.2 ± 15.4	21.0 ± 13.8	17.9 ± 20.1	27.4 ± 12.2	37.9 ± 13.1
SubGraphX	$\underline{47.6 \pm 30.0}$	$\underline{48.4 \pm 13.5}$	$\underline{48.0 \pm 12.8}$	27.4 ± 11.7	48.4 ± 6.1	$\underline{58.9 \pm 10.7}$
GStarX	14.2 ± 7.2	$\underline{46.8 \pm 2.9}$	$\underline{50.0 \pm 1.3}$	25.3 ± 11.7	49.5 ± 16.5	$\underline{62.1 \pm 15.8}$
TELL (ours)	$\mathbf{57.4 \pm 21.4}$	$\mathbf{64.0 \pm 20.3}$	$\mathbf{58.0 \pm 21.2}$	$\mathbf{42.1 \pm 11.0}$	$\mathbf{66.3 \pm 14.0}$	$\mathbf{65.3 \pm 12.3}$

	Mutagenicity			NCI1		
Top Nodes	1	3	5	1	3	5
GNNExpl.	7.6 ± 1.9	16.0 ± 3.4	21.9 ± 3.6	9.0 ± 1.3	19.3 ± 1.6	$\underline{26.7 \pm 2.5}$
PGExpl.	12.6 ± 1.8	22.0 ± 2.6	26.5 ± 2.0	14.5 ± 1.7	20.0 ± 2.8	$\underline{26.3 \pm 2.1}$
IG	6.9 ± 2.4	14.3 ± 5.1	17.7 ± 7.3	10.3 ± 2.7	16.5 ± 3.2	21.6 ± 2.8
SubGraphX	12.4 ± 3.8	21.9 ± 4.3	28.2 ± 3.0	14.7 ± 1.9	23.6 ± 2.4	$\underline{29.6 \pm 3.2}$
GStarX	$\underline{17.6 \pm 1.2}$	$\mathbf{36.5 \pm 4.1}$	$\mathbf{49.4 \pm 5.7}$	12.9 ± 2.6	23.3 ± 4.4	$\underline{29.2 \pm 5.5}$
TELL (ours)	$\mathbf{20.0 \pm 5.8}$	$\underline{34.0 \pm 6.9}$	40.8 ± 6.0	$\mathbf{24.4 \pm 6.5}$	$\mathbf{33.1 \pm 8.4}$	$\mathbf{36.6 \pm 10.9}$

	BBBP			PROTEINS		
Top Nodes	1	3	5	1	3	5
GNNExpl.	$\underline{5.0 \pm 1.6}$	$\underline{9.3 \pm 2.0}$	$\underline{12.8 \pm 2.1}$	12.7 ± 4.3	21.7 ± 3.7	28.8 ± 5.4
PGExpl.	$\underline{5.1 \pm 1.5}$	$\underline{10.3 \pm 0.6}$	$\underline{13.3 \pm 1.0}$	12.9 ± 3.4	20.4 ± 3.6	26.4 ± 5.4
IG	$\underline{3.9 \pm 1.6}$	8.1 ± 2.7	$\underline{12.8 \pm 4.9}$	8.7 ± 3.5	20.8 ± 3.9	28.3 ± 3.6
SubGraphX	$\underline{5.8 \pm 1.3}$	$\underline{10.7 \pm 3.3}$	$\underline{14.6 \pm 3.6}$	12.2 ± 2.8	24.9 ± 4.8	32.0 ± 3.2
GStarX	$\underline{8.0 \pm 1.6}$	$\underline{15.5 \pm 3.9}$	20.7 ± 6.3	$\mathbf{16.2 \pm 6.1}$	27.7 ± 4.9	$\underline{41.5 \pm 5.5}$
TELL (ours)	$\mathbf{15.5 \pm 13.8}$	$\mathbf{25.1 \pm 15.9}$	$\mathbf{29.8 \pm 18.3}$	12.5 ± 4.0	$\mathbf{33.7 \pm 3.6}$	$\mathbf{41.6 \pm 2.8}$

Table 3: InvFidelity Scores when using the top $N-1$, top $N-3$ and $N-5$ nodes as hard mask. Values are reported as $\mu \pm \sigma$ over 5 seeds. The highest result is highlighted in bold, while values within one σ of the best result are underlined.

	BA2Motifs			MUTAG		
Top Nodes	$N-1$	$N-3$	$N-5$	$N-1$	$N-3$	$N-5$
GNNExpl.	3.6 ± 2.1	16.4 ± 6.7	24.0 ± 6.8	$\underline{16.8 \pm 12.2}$	40.0 ± 20.7	56.8 ± 14.7
PGExpl.	24.6 ± 5.0	25.0 ± 3.8	24.4 ± 2.2	26.3 ± 8.8	49.5 ± 7.9	52.6 ± 11.0
IG	35.0 ± 22.4	40.6 ± 24.2	38.4 ± 18.8	37.9 ± 16.1	55.8 ± 19.6	61.1 ± 16.8
SubGraphX	$\mathbf{0.0 \pm 0.0}$	$\mathbf{0.0 \pm 0.0}$	$\mathbf{0.0 \pm 0.0}$	30.5 ± 3.9	56.8 ± 14.3	61.1 ± 12.7
GStarX	0.2 ± 0.4	0.2 ± 0.4	$\mathbf{0.0 \pm 0.0}$	$\underline{16.8 \pm 8.4}$	45.3 ± 13.6	57.9 ± 11.0
TELL (ours)	$\mathbf{0.0 \pm 0.0}$	$\mathbf{0.0 \pm 0.0}$	$\mathbf{0.0 \pm 0.0}$	$\mathbf{12.6 \pm 10.8}$	$\mathbf{22.1 \pm 13.1}$	$\mathbf{28.4 \pm 14.4}$

	Mutagenicity			NCI1		
Top Nodes	$N-1$	$N-3$	$N-5$	$N-1$	$N-3$	$N-5$
GNNExpl.	7.5 ± 0.7	14.7 ± 1.3	18.7 ± 2.1	10.5 ± 2.0	20.6 ± 0.1	27.3 ± 2.1
PGExpl.	11.4 ± 1.3	21.8 ± 1.3	27.3 ± 1.9	13.4 ± 1.5	23.4 ± 2.9	27.9 ± 2.9
IG	23.6 ± 4.6	37.6 ± 6.9	44.5 ± 9.2	23.4 ± 2.3	34.2 ± 4.1	39.3 ± 3.5
SubGraphX	8.5 ± 2.9	13.8 ± 3.2	19.6 ± 5.0	11.1 ± 0.9	23.4 ± 2.1	28.6 ± 3.0
GStarX	$\underline{6.5 \pm 2.1}$	$\underline{8.8 \pm 3.2}$	$\mathbf{9.3 \pm 3.1}$	16.1 ± 1.3	31.3 ± 5.1	37.5 ± 6.1
TELL (ours)	$\mathbf{4.5 \pm 2.2}$	$\mathbf{7.6 \pm 2.8}$	$\underline{10.9 \pm 3.3}$	$\mathbf{5.2 \pm 2.0}$	$\mathbf{11.4 \pm 3.3}$	$\mathbf{17.3 \pm 4.5}$

	BBBP			PROTEINS		
Top Nodes	$N-1$	$N-3$	$N-5$	$N-1$	$N-3$	$N-5$
GNNExpl.	$\underline{4.4 \pm 1.8}$	9.0 ± 2.4	12.2 ± 2.5	9.5 ± 2.7	20.6 ± 6.9	29.0 ± 5.0
PGExpl.	6.3 ± 1.5	11.7 ± 2.0	$\underline{13.6 \pm 2.3}$	11.3 ± 4.3	23.9 ± 4.1	$\underline{26.6 \pm 4.0}$
IG	10.5 ± 2.4	19.0 ± 3.3	23.7 ± 1.9	19.7 ± 5.8	30.4 ± 6.4	39.0 ± 6.5
SubGraphX	$\underline{4.1 \pm 1.9}$	$\mathbf{8.2 \pm 2.2}$	$\mathbf{11.4 \pm 2.6}$	$\underline{10.5 \pm 4.9}$	22.0 ± 5.5	28.8 ± 5.7
GStarX	$\underline{5.8 \pm 3.1}$	12.2 ± 9.0	18.2 ± 15.9	$\underline{10.0 \pm 4.3}$	$\mathbf{15.3 \pm 2.5}$	$\mathbf{23.1 \pm 6.3}$
TELL (ours)	$\mathbf{3.4 \pm 2.7}$	$\underline{8.8 \pm 8.2}$	$\underline{12.4 \pm 11.7}$	$\mathbf{7.9 \pm 3.3}$	19.1 ± 4.1	28.9 ± 6.4

Table 4: Stability Scores when using the top $N-1$, top $N-3$ and $N-5$ nodes as hard mask. Values are reported as $\mu \pm \sigma$ over 5 seeds. The highest result is highlighted in bold, while values within one σ of the best result are underlined.

	BA2Motifs			**MUTAG**		
Top Nodes	$N-1$	$N-3$	$N-5$	$N-1$	$N-3$	$N-5$
GNNExpl.	62.0 ± 0.8	64.2 ± 1.1	67.6 ± 1.0	75.8 ± 2.2	77.5 ± 2.2	81.5 ± 1.4
PGExpl.	98.5 ± 0.5	$\mathbf{97.8 \pm 0.6}$	$\mathbf{97.6 \pm 0.4}$	$\mathbf{99.4 \pm 0.3}$	$\mathbf{99.2 \pm 0.5}$	$\mathbf{99.0 \pm 0.2}$
IG	30.9 ± 24.7	28.7 ± 25.8	37.2 ± 37.9	58.7 ± 9.4	34.9 ± 20.1	14.2 ± 34.2
SubGraphX	76.1 ± 8.4	11.2 ± 6.3	8.6 ± 4.8	42.4 ± 6.4	31.1 ± 9.0	35.5 ± 7.0
GStarX	96.1 ± 3.0	94.1 ± 3.1	93.9 ± 4.1	55.9 ± 12.8	8.2 ± 24.4	-6.1 ± 20.7
TELL (ours)	$\mathbf{99.5 \pm 0.7}$	92.4 ± 6.6	92.2 ± 6.4	97.3 ± 1.7	94.4 ± 3.2	93.3 ± 1.7

	Mutagenicity			**NCI1**		
Top Nodes	$N-1$	$N-3$	$N-5$	$N-1$	$N-3$	$N-5$
GNNExpl.	70.5 ± 0.8	72.5 ± 0.5	72.5 ± 1.0	74.0 ± 0.8	74.9 ± 0.6	75.6 ± 0.6
PGExpl.	$\mathbf{99.2 \pm 0.1}$	$\mathbf{98.0 \pm 0.2}$	$\mathbf{96.5 \pm 0.2}$	$\mathbf{99.7 \pm 0.1}$	$\mathbf{99.4 \pm 0.1}$	$\mathbf{98.9 \pm 0.2}$
IG	53.5 ± 10.8	31.1 ± 11.8	21.6 ± 11.2	49.7 ± 10.6	30.0 ± 10.4	20.6 ± 8.8
SubGraphX	63.8 ± 6.1	59.7 ± 7.5	57.5 ± 7.4	40.3 ± 7.7	36.7 ± 7.4	37.5 ± 6.0
GStarX	82.4 ± 5.4	73.0 ± 10.1	67.7 ± 12.3	64.2 ± 2.7	34.1 ± 9.1	23.5 ± 10.8
TELL (ours)	94.5 ± 2.4	89.5 ± 5.3	85.7 ± 6.2	93.8 ± 1.6	88.7 ± 3.6	85.3 ± 3.9

	BBBP			**PROTEINS**		
Top Nodes	$N-1$	$N-3$	$N-5$	$N-1$	$N-3$	$N-5$
GNNExpl.	72.1 ± 0.2	74.0 ± 0.4	74.4 ± 0.7	90.4 ± 0.5	91.0 ± 0.9	88.5 ± 1.3
PGExpl.	$\mathbf{99.6 \pm 0.1}$	$\mathbf{99.0 \pm 0.4}$	$\mathbf{98.1 \pm 0.3}$	$\mathbf{99.8 \pm 0.1}$	$\mathbf{99.5 \pm 0.2}$	$\mathbf{97.3 \pm 0.7}$
IG	53.8 ± 5.9	28.1 ± 4.3	19.4 ± 3.3	49.7 ± 17.2	28.9 ± 19.4	21.0 ± 16.5
SubGraphX	60.0 ± 1.7	48.9 ± 8.7	47.8 ± 7.8	52.5 ± 10.0	46.5 ± 9.9	45.2 ± 9.6
GStarX	81.9 ± 5.7	66.0 ± 15.0	53.2 ± 27.8	75.7 ± 8.3	63.3 ± 4.8	49.5 ± 10.0
TELL (ours)	95.3 ± 6.6	90.8 ± 10.8	88.9 ± 10.8	99.0 ± 1.0	97.7 ± 2.3	$\underline{96.7 \pm 3.6}$