xAI-based Regularizers for Graph Neural Networks

Vincenzo Marco De Luca[1], Antonio Longa[1], Pietro Liò[2] and Andrea Passerini[1]

[1]University of Trento [2]University of Cambridge

vincenzomarco.deluca@unitn.it

Introduction and Background

GNN Limitations

Overfitting
Out-of-distribution generalization
Oversmoothing
Oversquashing
Noise propagation

GNN and xAI

IntegratedGradients integrates the gradient along a path. Specifically, given $x' \in \mathbb{R}^d$ a baseline input which represents a neutral input, the resulting explanation is computed as:

$$\mathrm{H}^{c}_{\text{IntegratedGradients}}[n] = (\boldsymbol{X}_n - x') \int_0^1 \frac{\partial f(x' + \alpha(\boldsymbol{X}_n - x'))}{\partial \boldsymbol{X}_n} \, d\alpha$$

Sufficiency

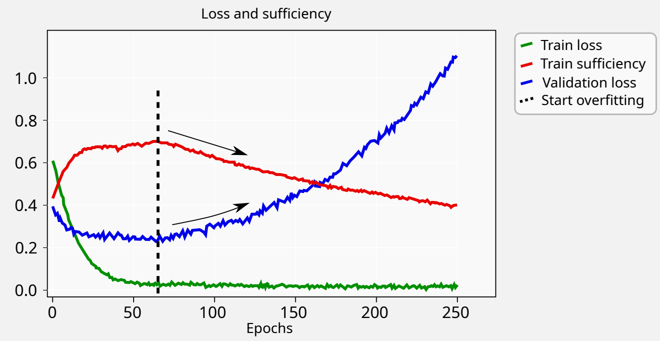
The fidelity sufficiency F_{suf} is the difference in the predicted probability when computed on the graph G and on the explanation. Since the explanation is a soft mask, we fix a number of levels $N_t \in \mathbb{N}$ and apply an incremental thresholding with N_t+1 threshold levels $t_k = k/N_t, k = 0, \ldots, N_t$
Where we define $G_{\text{exp}}(t_k)$ to be the hard mask explanation derived from G_{exp} with threshold t_k

$$F_{suf} = \frac{1}{N_t - 1} \sum_{k=1}^{N_t - 1} \left(g(G) - g(G_{\text{exp}}(t_k))\right)$$

Intuition

Experimental evidence

In the initial stages of experiments with synthetic datasets, it became evident that an increase in validation loss leads to a decrease in sufficiency



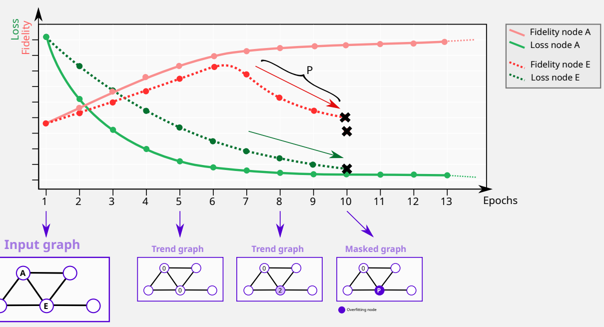
METHODS

Measure – Noise localization

Compute the loss L_n and the sufficiency F_{suf_n} $\forall n \in N$

xReg-Trend

Given a patient hyperparameter (P). The node n is overfitting if its loss is decreasing and its sufficiency is decreasing over P consecutive epochs

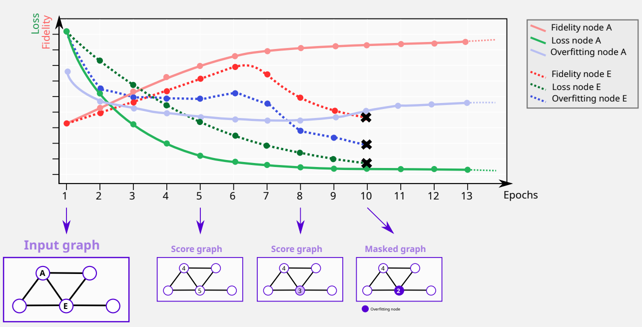


xReg-Score

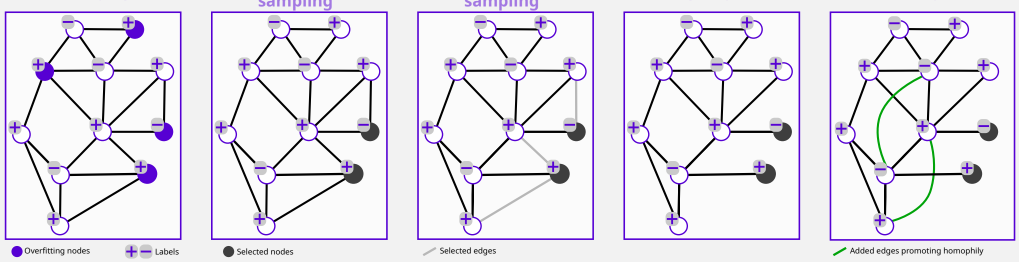
Given γ and α , compute

$$O_n = L_n + \gamma \cdot F_{suf_n} \quad \forall n \in N$$

remove α nodes with the smallest overfitting scores



xReg



PRELIMINARY RESULTS

Dataset statistics

	Cora	Citeseer
Nb. nodes	2708	3327
Nb. edges	5429	4732
Nb. features	1433	3703
Nb. classes	7	6

GCN

Model	Cora	Citeseer
Base	81.5 ±(0.3)	70.3 ±(0.9)
LP	70.4 ±(0.0)	50.4 ±(0.0)
MixHOP	81.9 ±(0.2)	71.4 ±(0.4)
GAUG	83.6 ±(0.5)	73.3 ±(1.1)
DropEdge	82.8 ±(0.9)	72.3 ±(1.3)
GraphMix	84.5 ±(0.6)	74.7 ±(0.6)
GRAND	84.3 ±(0.3)	74.2 ±(0.3)
NodeAug	85.1 ±(0.4)	74.9 ±(0.5)
Nasa	84.7 ±(0.3)	75.5 ±(0.4)
xReg-S	84.6 ±(0.4)	75.2 ±(0.4)
xReg-T	**85.3 ±(0.5)**	**75.9 ±(0.4)**

GAT

Model	Cora	Citeseer
Base	83.0 ±(0.7)	72.5 ±(0.7)
xReg-S	84.4 ±(0.5)	75.6 ±(0.4)
xReg-T	**85.5 ±(0.8)**	**76.2 ±(0.6)**

GraphSAGE

Model	Cora	Citeseer
Base	81.6 ±(0.4)	70.4 ±(1.1)
xReg-S	83.3 ±(0.5)	73.8 ±(0.9)
xReg-T	**83.5 ±(0.7)**	**74.1 ±(1.0)**

FUTURE WORKS

Additional datasets

Additional explainers

Regularizer hyperparameter tuning

Explainability as penalty

Ablation study

xAI for robust learning

UNIVERSITY OF CAMBRIDGE

UNIVERSITÀ DI TRENTO