Unlearnable Graph Examples

Research Background: The volume of “free” data has been the key to the current success of deep learning. However, it also raises privacy concerns about the unauthorized exploitation of personal data. Many commercial AI models are trained on data collected from the internet without individuals' knowledge, which poses significant threats to privacy, security, and copyright. Thus, it is crucial to develop methods to prevent unauthorized data exploitation. Recent studies have focused on the development of Unlearnable Examples which aim to make the original data unlearnable by adding imperceptible but deceptive perturbations to data samples.

Research Problem: We seek to answer the question of how to create unlearnable graphs, in order to enhance the robustness and security of GNN models against unauthorized graph exploitation.

Methods:

Experimental Goal: GNNs trained on unlearnable examples will have a performance equivalent to random guessing on normal test examples.

- Making an example unlearnable should not affect its quality for normal usage.
- The development of unlearnable examples should use the weaknesses of GNNs.

Node Feature:

- Noise is learnable
- Noise is close to 0
- Introduce specific noise to each individual node based on node features
- Noise is zero mean

Topology Information:

- Perturb only one edge for each graph (mask the edge)
- Initialize a learnable query vector
- Generate edge features by combining the features of the two nodes connected by the edge
- Calculate the similarity between each edge in a graph with the query vector
- Select the edge with the highest score
- Mask the edge

Experiments: Our experiments were conducted based on three datasets: Mutagenicity, AIDS, and PROTEINS. We used the Adam optimizer function and picked 0.01 as the learning rate. For the data split, we used 80% of the data as training data, and the rest of them as testing data.

Result:

	Mutagenicity		AIDS		PROTEINS	
	Train	Test	Train	Test	Train	Test
GCN	0.7760	0.7539	0.8565	0.8633	0.7044	0.7136
GCN + (feature noise)	0.7611	0.5663	0.7841	0.1700	0.7067	0.6808
GCN + (structure noise)	0.7783	0.7145	0.8659	0.8267	0.7022	0.6901
GCN + (Feature & structure)	0.7580	0.4564	0.8200	0.2000	0.7156	0.6432