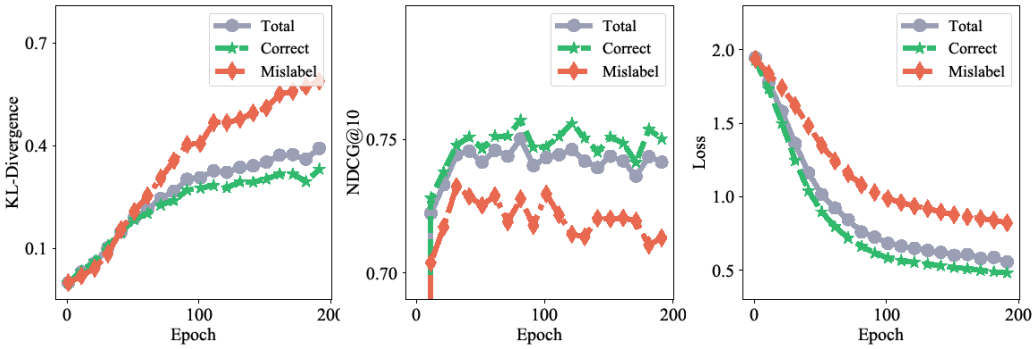
Soft-GNN: Towards Robust Graph Neural Networks via Self-adaptive Data Utilization

Yao WU, Hong HUANG, Yu SONG, Hai JIN

Frontiers of Computer Science, DOI: 10.1007/s11704-024-3575-5

Problems & Ideas

- Problems: the robustness of GNNs trained under label noise
  - The issues of agnostic graph data pollution, such as deliberate adversarial attacks or annotation mistakes by accident.
  - GNNs is confronted with the problem of over-fitting to noisy graph data and suffers from severe performance degradation.
- Ideas: we discover three types of deviations between mislabeled nodes and correct nodes as the training features and select the subset of labeled nodes based on current training state by introducing a sample network.



(a) Local deviation (b) Global deviation (c) Prediction deviation

Figure 1 The training features

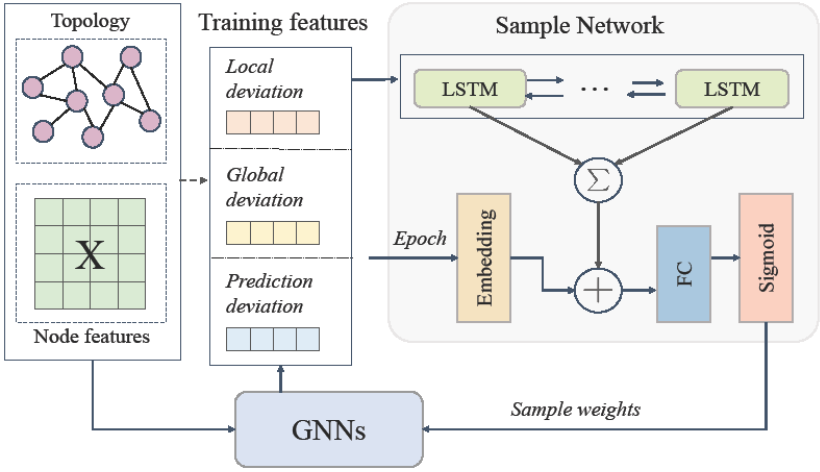


Figure 2 The framework of Soft-GNN

Main Contributions

- Contributions:
  - We analyze the training states of GNNs trained with noisy data and find that there are noticeable deviations from training states for mislabeled nodes.
  - We propose a novel framework to introduce a self-adaptive sample network for dynamic data selection and mitigate the negative impact of label noise, denoted by Soft-GNN.
  - We conduct lots of experiments to verify the effectiveness of the proposed framework and give new insight into the improvements.

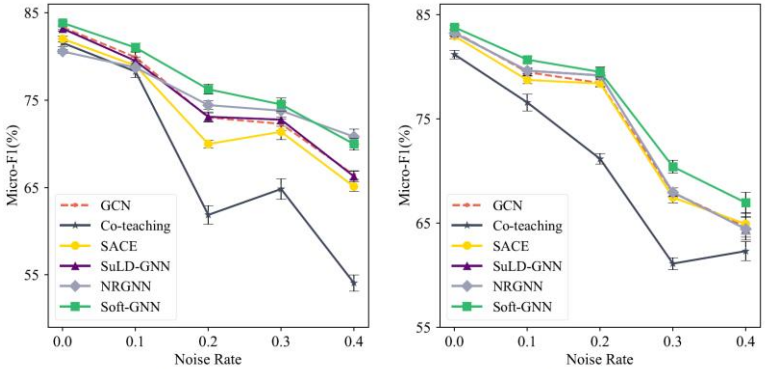


Figure 3 Performance on Cora

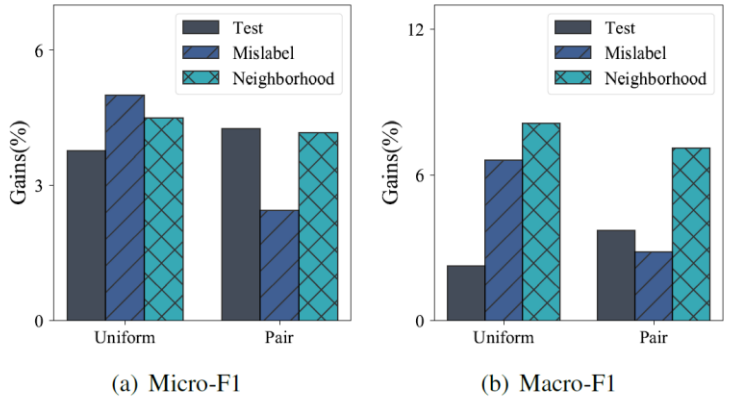


Figure 4 Performance gains over GCN

Soft-GNN performs better than the basic GNNs and the improvements are more clear with a higher noise rate. Left: The results under different noise types and rates on Cora; Right: The performance gains over GCN on Citeseer under noise rate=0.4.