Semi-supervised graph embedding for multi-label graph node classification

In this paper Kaisheng et al. proposes a Graph convolution network (GCN) based semi-supervised learning approach for multi-label classification called Multi-label graph convolution network (ML-GCN) that tries to improve on where GCN falls short.

GCN is a deep neural network model used to catch the structural information in a graph. GCN uses a graph convolution operation to integrate each node and its 1-hop neighbor information in each layer. After having conducted several layers of convolution, each node can gather its k-hop information in the final layer which is then the embedded feature presentation of the node. Afterwards some supervised information can be used to train a classifier based on the embedded features.
Multi-label classification models are usually trained semi-supervised as not all labels on all instances are obtained values. For example in a multi-label graph dataset, one node can have several labels, which means the correlation between the node and the labels it have are high, they call this the node-label correlation. If two labels are highly correlated the nodes that have these labels should be closer to each other in the embedding space. Example in their dataset with movie genres the genre labels “Western” and “Adventure” always appear together, therefor movies the genres Western and Adventure should also be closer to each other in the embedding space. This they call the label-label correlation. But because these type of correlation is not reflected in the graph structure it cannot be captured by GCN. They therefore try to address this issue in the paper, proposing their Multi-label Graph convolution network (ML-GCN).