A Literature Review of Graph Attention Networks

Author: Aaron Rock Menezes

Neural networks struggle to effectively process data structured as graphs. While conventional approaches either incur high computational costs due to spectral graph decomposition or fail to distinguish between a node's neighbours as seen in GCNs, Velickovic et al. (2017) proposed Graph Attention Networks (GATs). Inspired by the self-attention mechanism introduced in the Transformer architecture (Vaswani et al., 2017), GATs utilize masked self-attention layers to achieve a more fine-grained understanding of node features.

In GATs, nodes selectively focus on the features of their neighbours, assigning greater weight to those that hold more information, i.e. paying more attention to specific nodes. This approach represents a significant departure from prior methods that relied on explicitly calculating weights for each neighbour. The resulting improvement in efficiency allows GATs to well on Transductive as well as Inductive tasks, surpassing the performance of GCNs and GraphSAGE across 4 datasets.

The working of GAT networks is similar to that of the Attention mechanism (Bahdanau et al. 2015). The authors begin by performing a linear transformation on the node features, which is followed by calculation of attention coefficients for all node features. These are parameterized by a learnable weight vector ‘a’. Softmax normalization follows resulting in a set of attention scores for all node features. The user may use a non-linear activation function on these scores for their use. These graph attention layers can be parallelized, resembling the multi-head attention mechanism similar to the one in Vaswani et al. (2017).

In conclusion, GATs address limitations of existing methods for processing graph data. By utilizing self-attention, they assign importance to specific neighbours, reducing computational cost and adapting to unseen graphs. This opens doors for more efficient and generalizable graph-based learning tasks.