Practical work 3 report

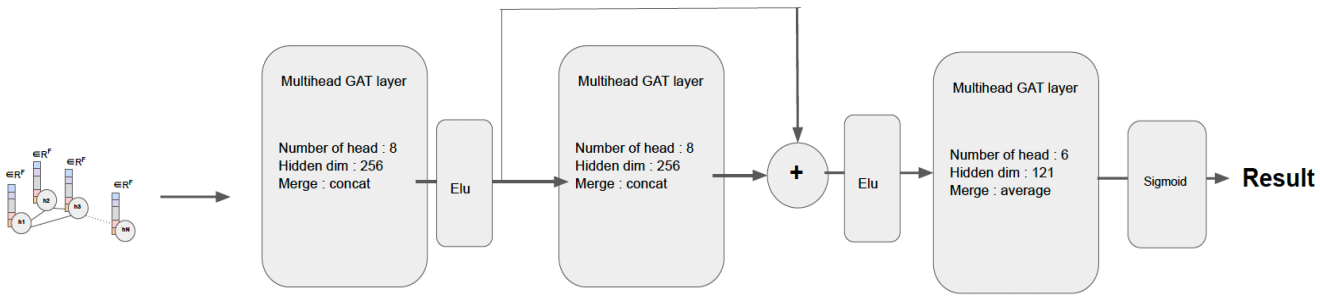
Author: Alassane Watt

Plan:

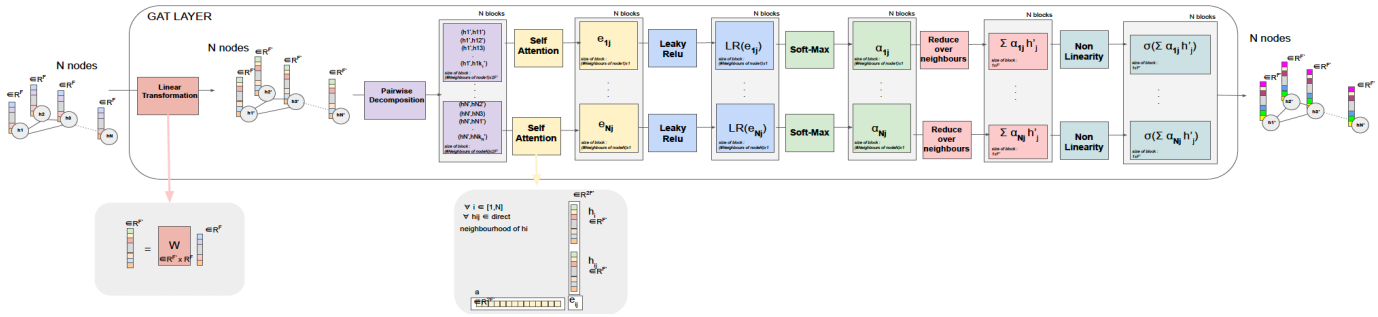
1. Architecture
2. GAT Layer Diagram
3. Multi head GAT Layer Diagram
4. Similarity Attention
5. Utility Attention

Architecture:

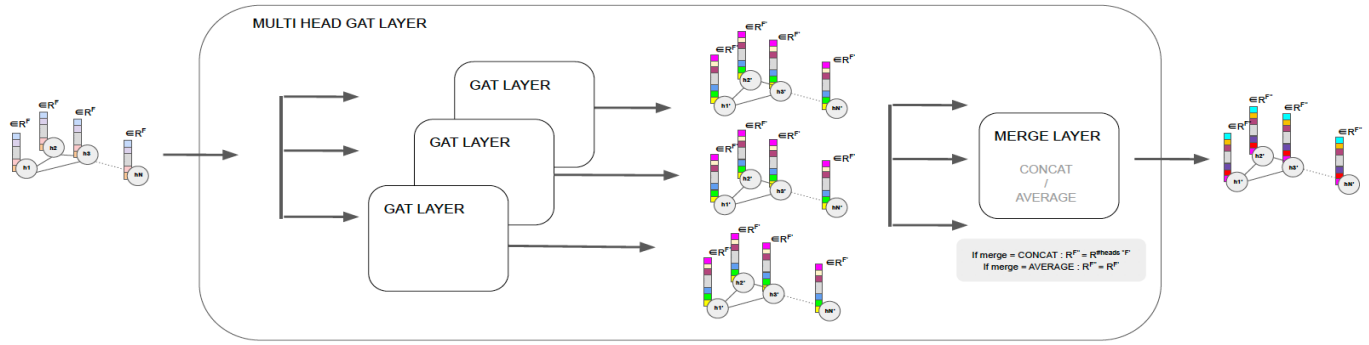
This architecture is inspired by Velic kovic'paper.



GAT Layer Diagram:



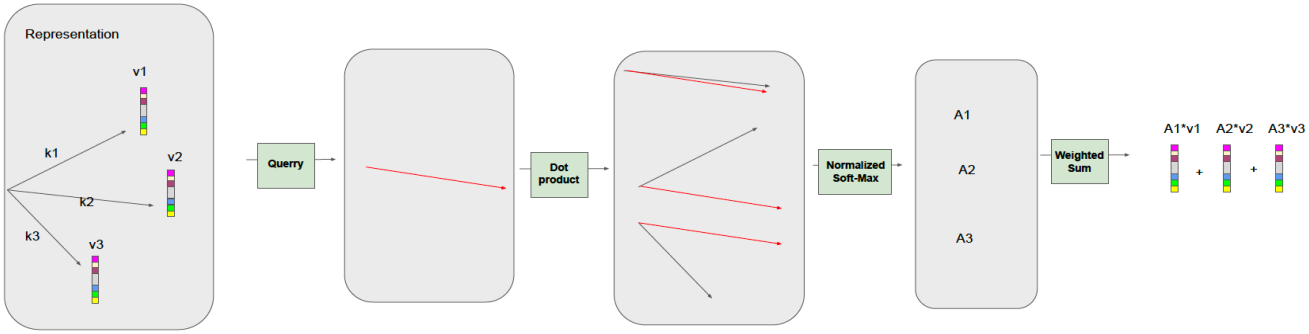
Multi head GAT Layer Diagram:



Similarity Attention:

Similarity attention learns a representation of data (keys of dimension d_k) associated with their features. To compute attention for a piece of data the mechanism is to compute its representation and find the other pieces of data whose representation are most similar to the first one and then perform a weighted sum of their values.
The notion of similarity referred to the fact that attention is computed according to the most similar keys, this notion is expressed via a dot product between key and querry.

$$Attention(Q,K,V) = softmax(Q*K^T/sqrt(d_k))V$$



Utility attention:

The idea of utility attention is to learn a self-attention vector that can predict how important a piece of data is with respect to another piece of data. The mechanism to compute utility attention is to compute the softmax of the relative importance of every neighbor node and to perform the weighted sum of those nodes' features. The notion of utility referred to the fact that attention is computed according to the relative importance of data “how important a node is to another node”, this notion is expressed via a dot product between the hidden representations of the nodes h'_i and h'_j.

$$e_{ij} = a^T * (h'_i \text{ ll } h'_j)$$
$$\alpha_{ij} = \text{softmax}_j (\text{LeakyRely}(e_{ij}))$$
$$h''_i = \text{NonLinearity}(\Sigma_j \alpha_{ij} h'_i)$$