GAT Mathematical Example

Sandesh Bashyal

December 2024

1 Graph

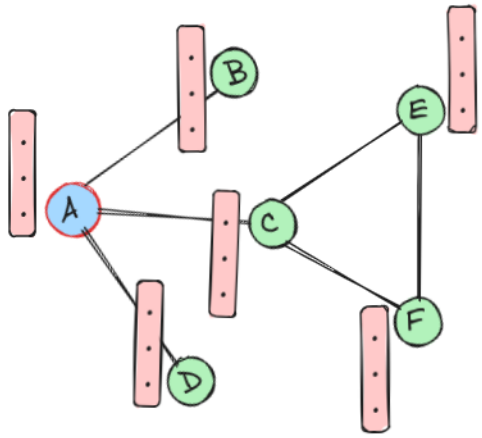

Figure 1: Graph

Let's take a small graph for the discussion. This will help us understand mathematically. There are 6 nodes V = {A, B, C, D, E, F} having undirected edges. The block besides each node with the pink background color is the features of each node which will be assumed in next section. We want to understand the mathematically computation of the node A so is highlighted with the blue color in the figure.

2 Assumptions and Dimensions

- Number of input features per node, F = 5
- Number of output features after applying W, F' = 3
- Number of nodes in the graph: 6
- Attention weight vector a dimension, 2F' = 6

3 Step-by-Step Calculation

3.1 Input Feature Vectors

$$h_A = [1, 2, 3, 4, 5]$$
$$h_B = [2, 3, 4, 5, 6]$$
$$h_C = [3, 4, 5, 6, 7]$$
$$h_D = [4, 5, 6, 7, 8]$$

3.2 Linear Transformation with Weight Matrix W

We apply a shared weight matrix W of size $F' * F = 3 * 5$ to each node's feature vector.
Let's assume:

$$W = \begin{bmatrix} 0.1 & 0.2 & 0.3 & 0.4 & 0.5 \\ 0.5 & 0.4 & 0.3 & 0.2 & 0.1 \\ 0.1 & 0.3 & 0.5 & 0.7 & 0.9 \end{bmatrix}$$

Now, we compute for each neighborhood node of A including itself.
For node A:

$$h'_A = Wh_A = \begin{bmatrix} 0.1 & 0.2 & 0.3 & 0.4 & 0.5 \\ 0.5 & 0.4 & 0.3 & 0.2 & 0.1 \\ 0.1 & 0.3 & 0.5 & 0.7 & 0.9 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix} = \begin{bmatrix} 5.5 \\ 3.5 \\ 11.5 \end{bmatrix}$$

For node B:

$$h'_B = Wh_B = \begin{bmatrix} 6.5 \\ 4.5 \\ 13.5 \end{bmatrix}$$

For node C:

$$h'_C = Wh_C = \begin{bmatrix} 7.5 \\ 5.5 \\ 15.5 \end{bmatrix}$$

For node D:

$$h'_D = Wh_D = \begin{bmatrix} 8.5 \\ 6.5 \\ 17.5 \end{bmatrix}$$

All of these transformed feature vectors has a dimension of $F' = 3$.

3.3 Compute attention score e_{Aj}

This attention mechanism involves a weight vector $a \in \mathbb{R}^2, \quad F' = \mathbb{R}^6$

Let's assume:

$$a = \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \\ 0.4 \\ 0.5 \\ 0.6 \end{bmatrix}$$

Now for e_{AB}:

$$e_{AB} = LeakyReLU(a^T \left[h'_A \parallel h'_B\right]) = LeakyReLU \left(\begin{bmatrix} 0.1 & 0.2 & 0.3 & 0.4 & 0.5 & 0.6 \end{bmatrix} \begin{bmatrix} 5.5 \\ 3.5 \\ 11.5 \\ 6.5 \\ 4.5 \\ 13.5 \end{bmatrix} \right)$$

On computing, $e_{AB} = 17.65$.

Similarly, Compute e_{AC} and e_{AD}.

3.4 Normalize attention scores with Softmax

$$\alpha_{AB} = \frac{\exp(e_{AB})}{\exp(e_{AC}) + \exp(e_{AD}) + \exp(e_{AB})}$$

3.5 Aggregate Functions

$$h''_A = \alpha_{AB} h'_B + \alpha_{AC} h'_C + \alpha_{AD} h'_D$$