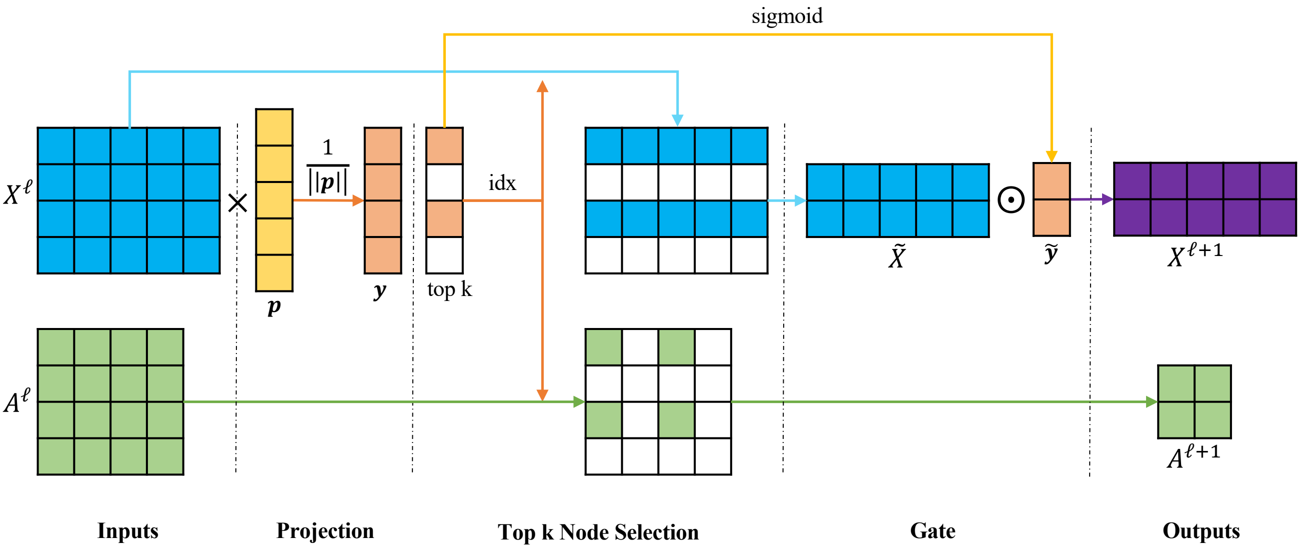

Figure 1. An illustration of the proposed graph pooling layer with $k = 2$. $\times$ and $\odot$ denote matrix multiplication and element-wise product, respectively. We consider a graph with 4 nodes, and each node has 5 features. By processing this graph, we obtain the adjacency matrix $A^\ell \in \mathbb{R}^{4\times4}$ and the input feature matrix $X^\ell \in \mathbb{R}^{4\times5}$ of layer ℓ. In the projection stage, $\mathbf{p} \in \mathbb{R}^5$ is a trainable projection vector. By matrix multiplication and sigmoid$(\cdot)$, we obtain $\mathbf{y}$ that are scores estimating scalar projection values of each node to the projection vector. By using $k = 2$, we select two nodes with the highest scores and record their indices in the top-k-node selection stage. We use the indices to extract the corresponding nodes to form a new graph, resulting in the pooled feature map $\tilde{X}^\ell$ and new corresponding adjacency matrix $A^{\ell+1}$. At the gate stage, we perform element-wise multiplication between $\tilde{X}^\ell$ and the selected node scores vector $\tilde{\mathbf{y}}$, resulting in $X^{\ell+1}$. This graph pooling layer outputs $A^{\ell+1}$ and $X^{\ell+1}$.