5. (d)

(17) $h_i^{l+1} = \text{Attention}(Q^l h_i^l, K^l h_j^l, V^l h_j^l)$

$$= \sum_{j \in S} (\text{softmax}_j(Q^l h_i^l \cdot K^l h_j^l) V^l h_j^l)$$

(15) can be

$$h_i^{l+1} = \phi(h_i^l, \sum_{j \in N_i} (u^l h_j^l))$$

in this case, we can write w_{ij} as

$$w_{ij} = \text{softmax}_j(Q^l h_i^l \cdot K^l h_j^l)$$

then (17) could be written as

$$h_i^{l+1} = \sum_{j \in N(v_i)} w_{ij} (V^l h_j^l) = W(h_i^l)^T \sum_{j \in N(v_i)} (V^l h_j^l)$$

which is a special case for (15)

(e) This is because for fully-connected graphs, number of edges in the graph scales quadratically with the number of nodes. So for sentence with n words, the Transformer / GNN will compute n^2 pairs of words. Thus for a very-long-term n, the computation will be problematic.