Paper 1 – **Graphormer**, Do Transformers Really Perform Bad for Graph Representation?

Centrality Encoding –

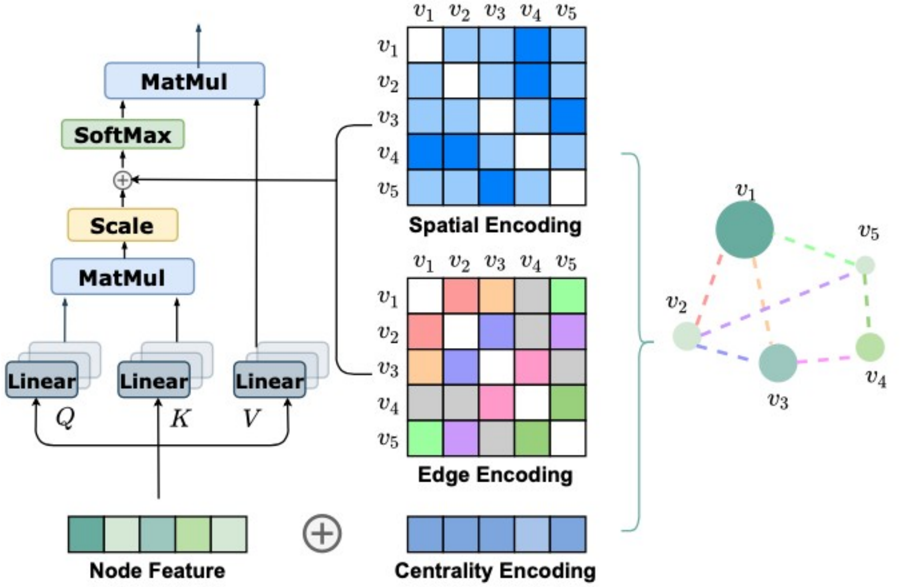
$$h_i^{(0)} = x_i + z^{-}_{\deg^{-}(v_i)} + z^{+}_{\deg^{+}(v_i)},$$

Spatial Encoding –

$$A_{ij} = \frac{(h_i W_Q)(h_j W_K)^T}{\sqrt{d}} + b_{\phi(v_i,v_j)},$$

Edge Encoding –

$$A_{ij} = \frac{(h_i W_Q)(h_j W_K)^T}{\sqrt{d}} + b_{\phi(v_i,v_j)} + c_{ij}, \text{ where } c_{ij} = \frac{1}{N}\sum_{n=1}^{N} x_{e_n}(w_n^E)^T,$$



More expressive power than GIN!
Can distinguish graphs where 1-WL test fails.

Very Innovative; **addition of a special node [VNode]** in analogy with [CLS] token in BERT.

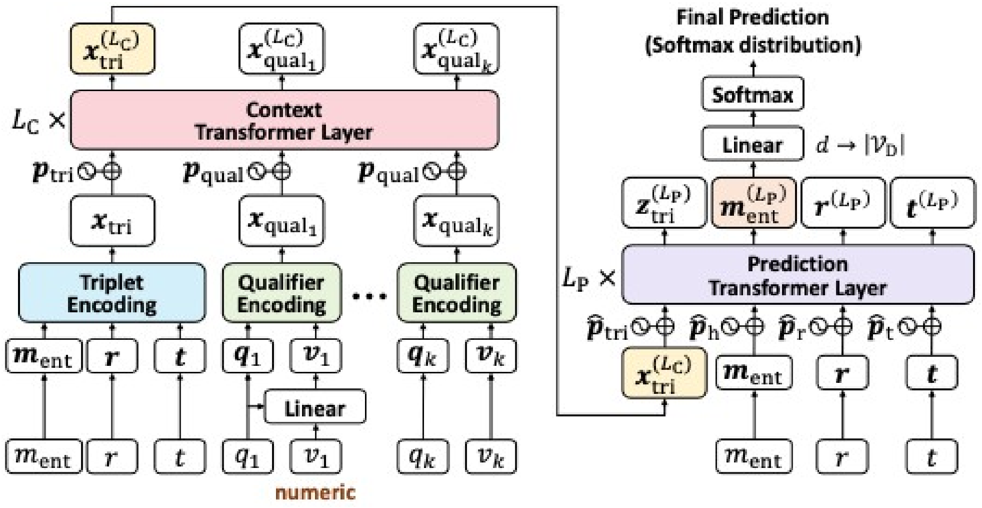
- There is an edge between [VNode] and each node.
- Each step updates [VNode] like a normal node.
- Finally, [VNode] is used as a representation of the Graph!

- *”While the [VNode] is connected to all other nodes in graph, which means the distance of the shortest path is 1 for any φ([VNode], vj) and φ(vi , [VNode]),* ***the connection is not physical****. To distinguish the connection of physical and virtual, we reset all spatial encodings for bφ([VNode],vj) and bφ(vi ,[VNode]) to a* ***distinct learnable scalar****”.*

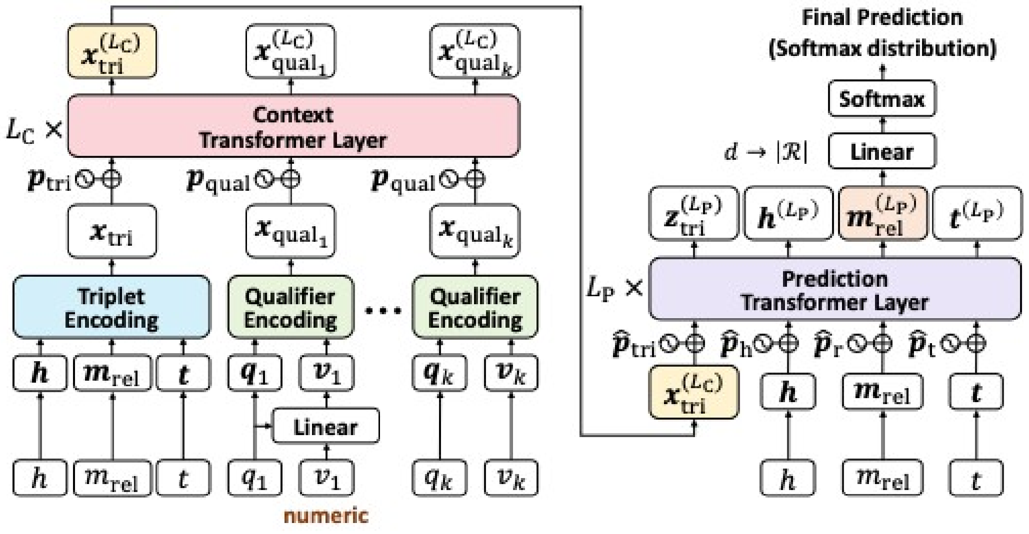
- *“Conceptually, the benefit of the virtual node is that it can* ***aggregate the information of the whole graph*** *(like the READOUT function) and then* ***propagate it to each node****”.*

- Took special care to avoid over-smoothing.

Paper 2 – **HyNT**, Representation Learning on Hyper-Relational and Numeric Knowledge Graphs with Transformers

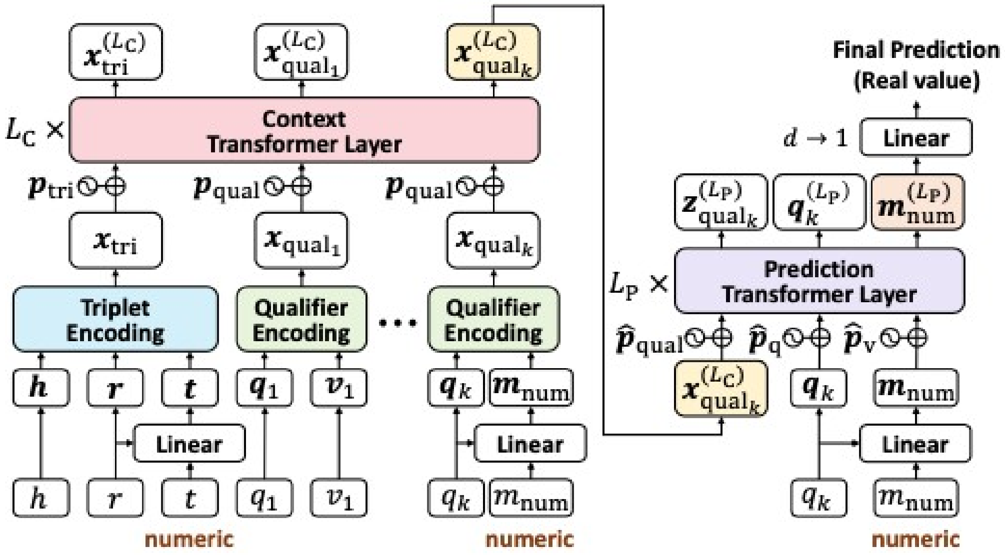


(a) Predicting a discrete head entity in a primary triplet.



(b) Predicting a relation in a primary triplet.

Rest from the paper!



(c) Predicting a numeric value in a qualifier.