

If leaf node is masked, it contributes to loss

Every leaf token embedding is fused with relation embedding and entire head token embeddings

The gradient will backpropagate from loss to the masked t1.

 $\underline{Backpropagation\ path} : loss \rightarrow transformer\ layers \rightarrow H\text{-}GAT \rightarrow input\ token\ embeddings}.$