LREC-COLING 2024

Claim-Centric And Sentiment Guided Graph Attention Network for Rumour Detection

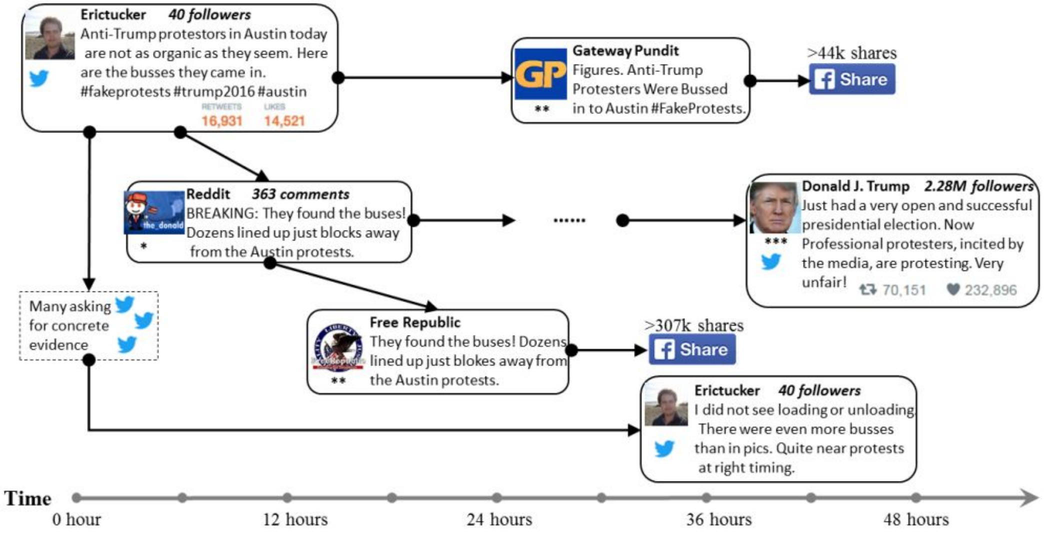
Sajad Ramezani, Mauajama Firdaus, Lili Mou

UNIVERSITY OF ALBERTA

amii

CIFAR

Motivation



Picture from Ma, Jing, Wei Gao, and Kam-Fai Wong. "Detect rumors in microblog posts using propagation structure via kernel learning." Association for Computational Linguistics, 2017.

Problem Definition

- Source Tweet (***S***): The initial tweet containing the claim to be classified as either a rumour or not a rumour.

- Text Document ($\boldsymbol{d_i}$): This represents the text from the source tweet and all related tweets and comments.

- Propagation Graph (***G***): An attributed graph where nodes ($\boldsymbol{v_i}$) are tweets or comments associated with the source tweet. Edges ($\boldsymbol{e_{ij}}$) represent relationships between nodes, such as retweets or comments. Each node feature primarily includes the text of the node.

- Task of Rumour Detection: The objective is to determine whether the source tweet, treated as a node within the propagation graph (***G***), can be classified into discrete categories: rumour or non-rumour. This task involves analyzing the graph structure along with node features to identify misinformation.

Prior Work

- Approach 1: Text-Based Classification
  - Early models focused solely on textual content to classify rumours. These models analyze the text of the source tweet and related discussions to determine the nature of the information.
  - The classification function for this approach is denoted as $\boldsymbol{f(S, d_i)}$, where $\boldsymbol{S}$ is the source tweet and $\boldsymbol{d_i}$ represents the text document comprising the source tweet and its related content.
  - Notable studies employing this method include Liu et al. (2015) and Chen et al. (2018)

- Approach 2: Graph-Based Classification with Text Features
  - More recent advancements utilize both the propagation graph and text features to enhance rumour detection. This approach considers how information spreads through social networks alongside the textual content of the interactions.
  - The classifier in this approach is expressed as $\boldsymbol{f(S, G)}$, indicating that it uses the source tweet $\boldsymbol{S}$ and the propagation graph $\boldsymbol{G}$ to perform classification.
  - Studies such as Bian et al. (2020) and Sun et al. (2022) have explored the integration of graph structures with textual analysis, leading to improved accuracy in identifying rumours.

Rumour Propagation Graph in GNN

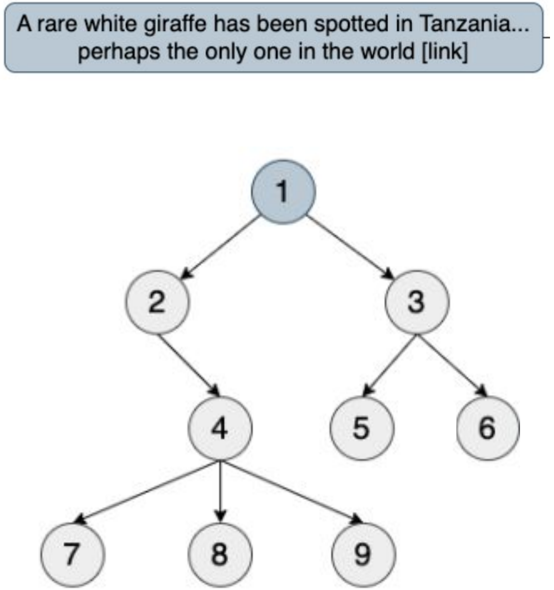
Source Tweet Text

A rare white giraffe has been spotted in Tanzania...
perhaps the only one in the world [link]

Rumour Propagation Graph in GNN

Source Tweet Text

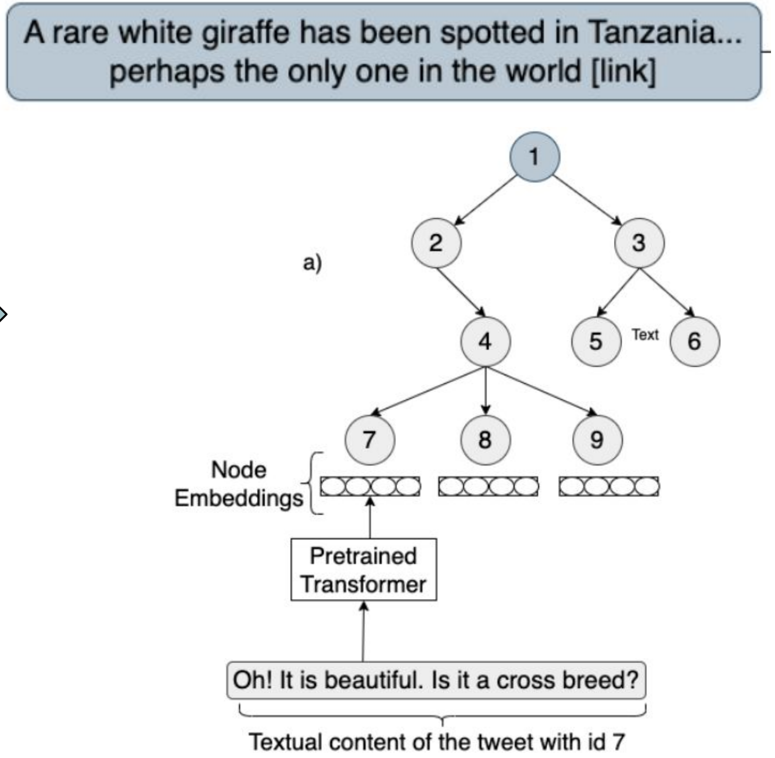
Propagation Graph



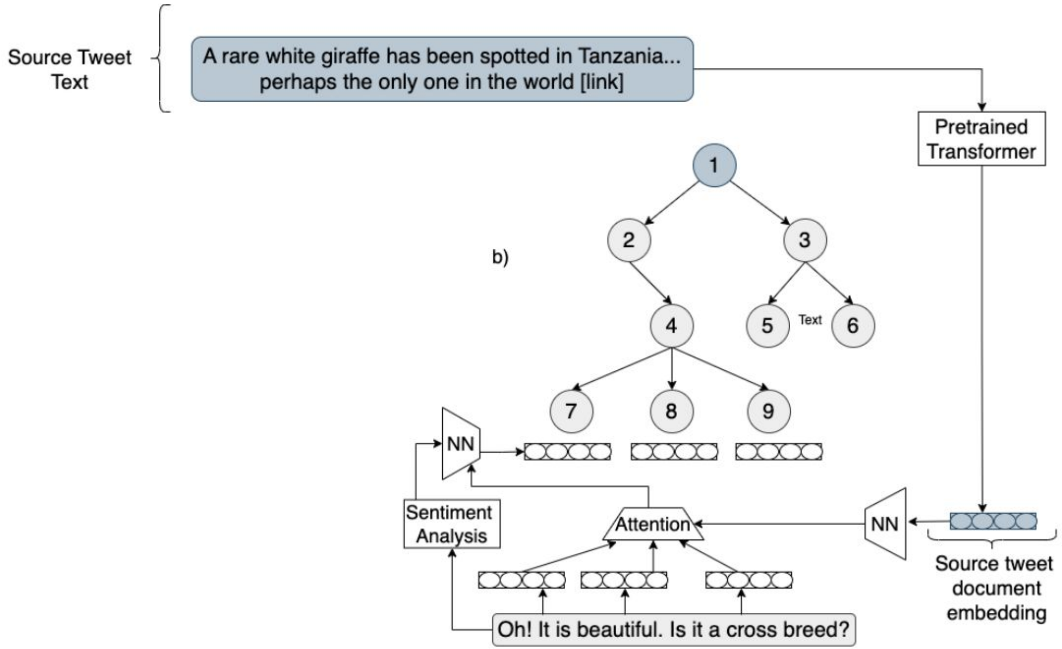
Rumour Propagation Graph in GNN

Source Tweet Text

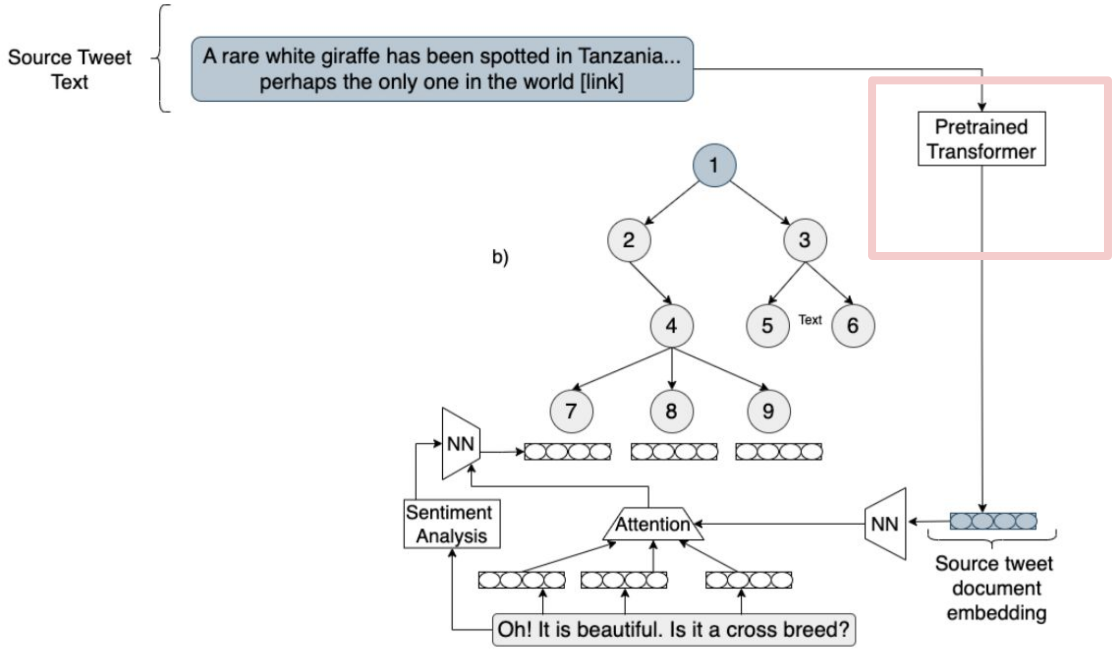
Propagation Graph



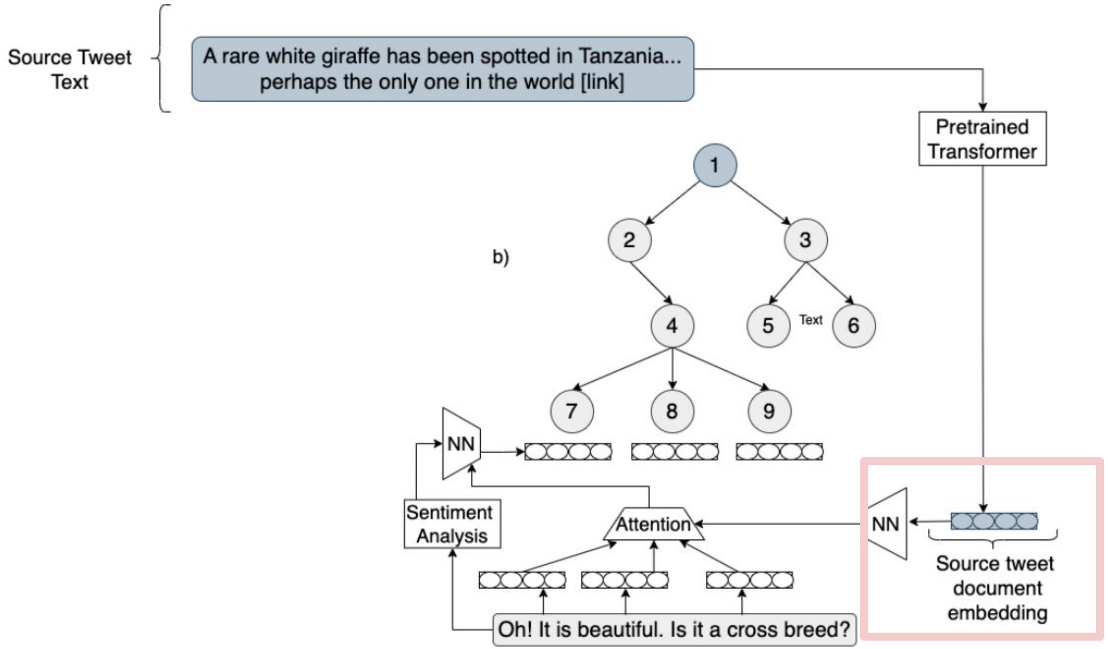
Our Proposed Method



Double Level Attention for Node Embeddings



Double Level Attention for Node Embeddings

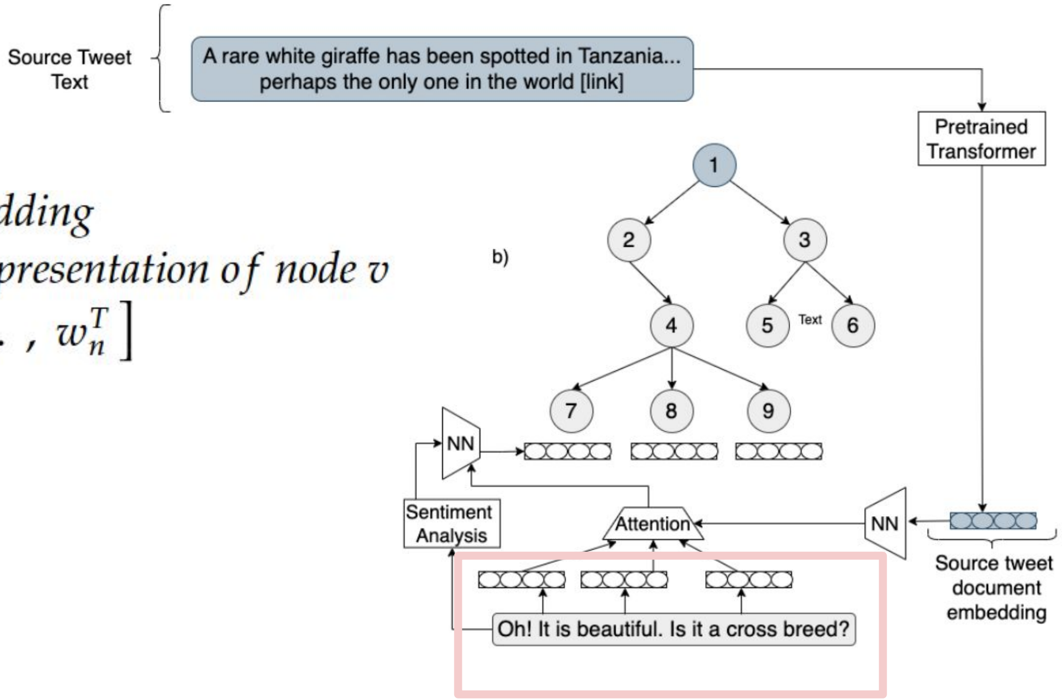


Double Level Attention for Node Embeddings

w_i : *the word embedding*

$D_v :=$ *document representation of node v*

$$D_v = \left[w_1^T, w_2^T, \dots, w_n^T \right]$$

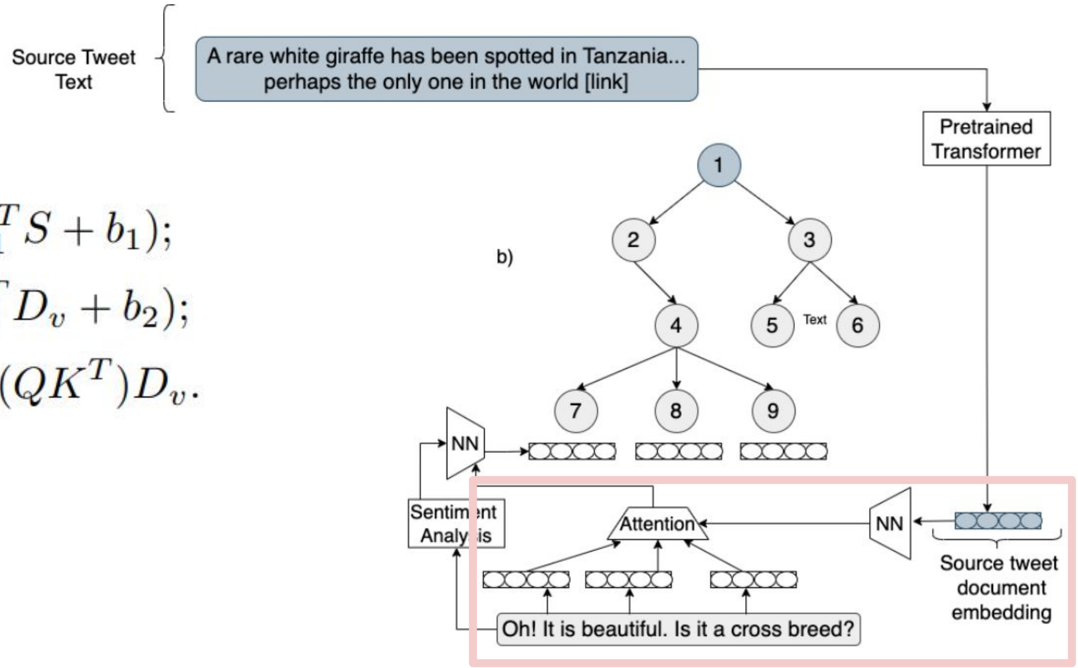


Double Level Attention for Node Embeddings

$$Q = Relu(W_1^T S + b_1);$$

$$K = Relu(W_2^T D_v + b_2);$$

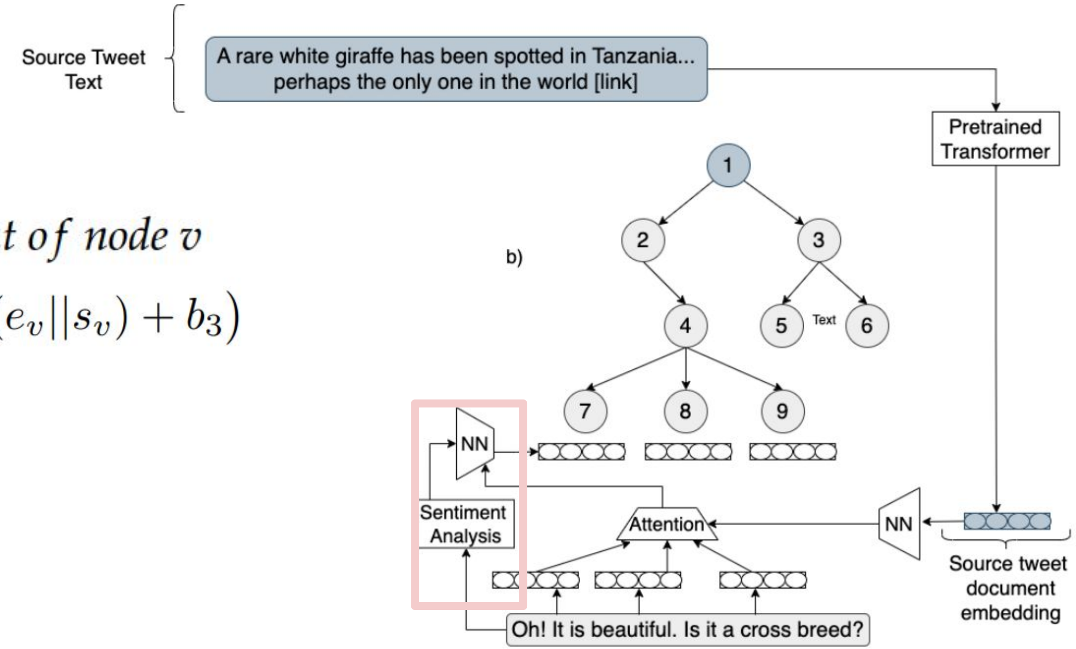
$$e_v = Softmax(QK^T)D_v.$$



Sentiment Features

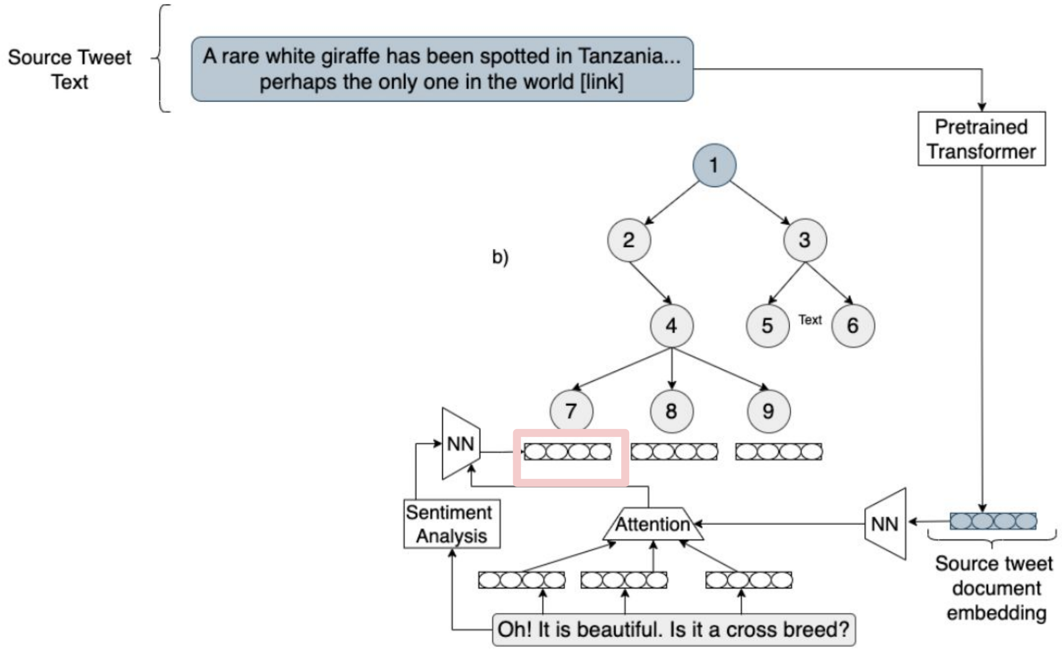
$$s_v := \ sentiment\ of\ node\ v$$

$$r_v = relu\left(W_3^T(e_v||s_v) + b_3\right)$$



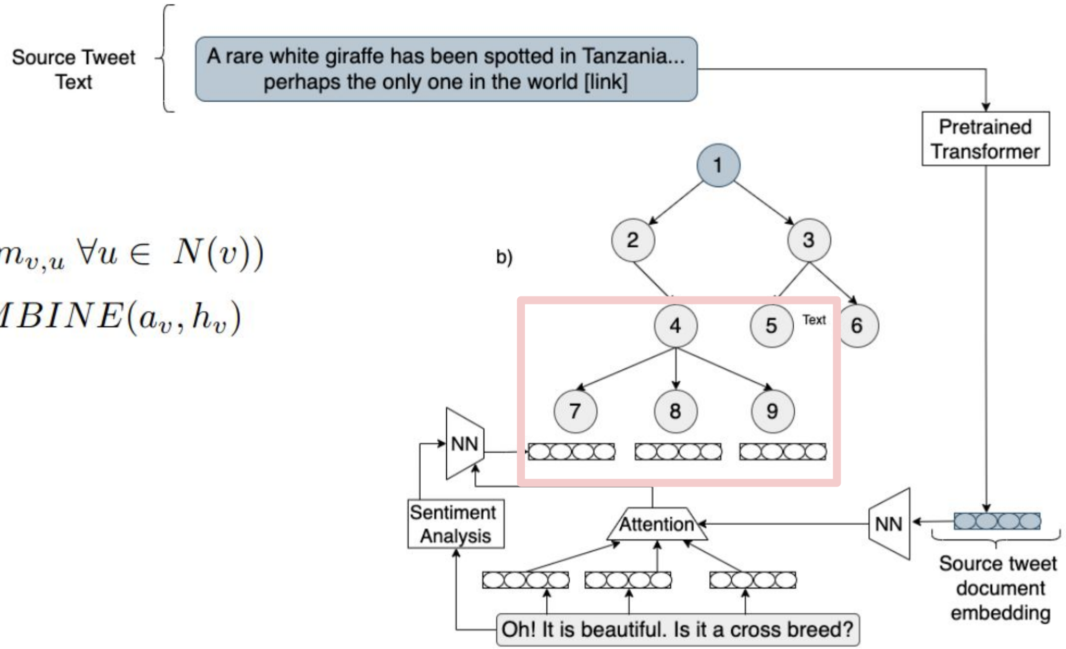
Sentiment Features

$$h_v \;=\; AGG(h_v, r_v)$$



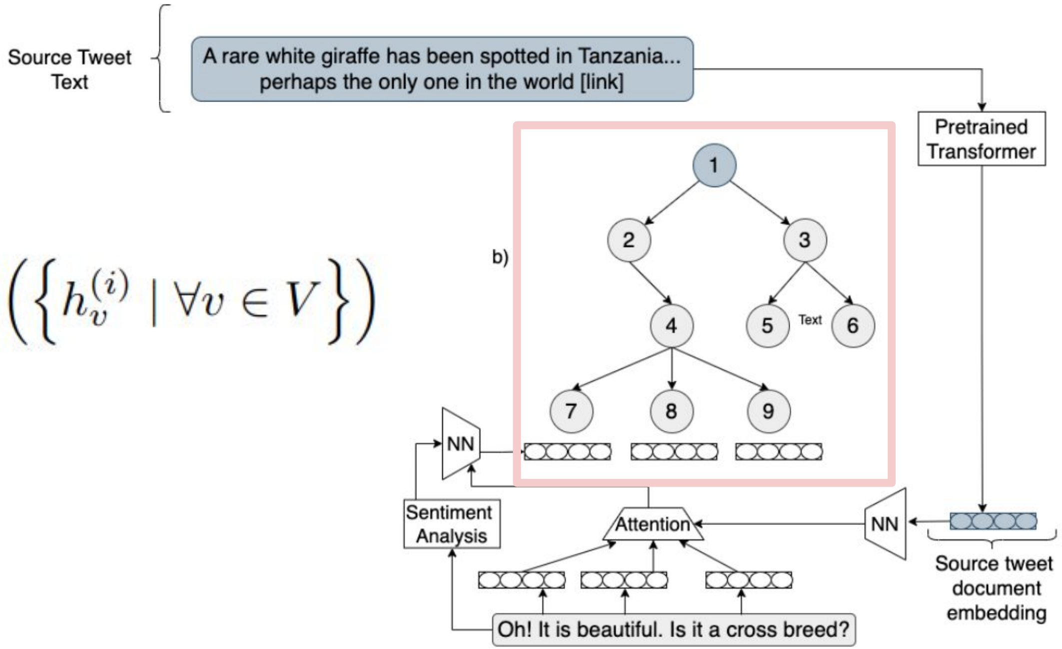
Graph Message Passing

$$a_v = AGG(m_{v,u} \, \forall u \in N(v))$$
$$h_v = COMBINE(a_v, h_v)$$



Propagation Graph Embedding

$$h_g^{(i)} = Max\left(\left\{h_v^{(i)} \mid \forall v \in V\right\}\right)$$



Augmentation

Previous works have been focused on augmenting the propagation graph using random masking in both graph and node features

We instead focus on using generative models to paraphrase the actual claim .From each source tweet (***S***), we generated two paraphrase version, resulting in a a training size three times the original

Experiment Setup

Rumour Datasets

Statistics	Twitter 15	Twitter 16	PHEME
# Users	276,663	173,487	197,852
# Source Tweets	1490	818	6425
# Non Rumours	374	205	4023
# False Rumours	370	205	2402
# True Rumours	372	205	NA
# Unverified Rumours	374	203	NA

Table 1: Dataset Statistics

Result

	PHEME			Twitter 15					Twitter 16				
Method	*Acc*	*R F1*	*N F1*	*Acc*	*U F1*	*N F1*	*T F1*	*F F1*	*Acc*	*U F1*	*N F1*	*T F1*	*F F1*
cPTK (Ma et al., 2017)	-	-	-	0.750	0.733	0.804	0.765	0.698	0.732	0.686	0.740	0.836	0.709
RvNN (Ma et al., 2018)	0.763	0.631	0.825	0.723	0.654	0.682	0.821	0.758	0.737	0.708	0.662	0.835	0.743
Bi-GCN (Bian et al., 2020)	0.824	0.741	0.865	0.886	0.864	0.891	0.930	0.860	0.880	0.865	0.847	0.937	0.869
GACL (Sun et al., 2022)	0.850	0.771	0.885	0.901	**0.876**	**0.958**	0.903	0.851	0.920	**0.907**	**0.934**	**0.959**	0.869
Our Method	**0.891**	**0.843**	**0.914**	**0.908**	0.873	0.896	**0.946**	**0.873**	**0.925**	0.883	0.918	**0.959**	**0.883**

Table 2: Main Results. Here R F1:Rumour F1, N F1:Non-Rumour F1, U F1: Unverified Rumour F1, T F1: True Rumour, F F1: False Rumour F1

Discussion

1. Without Word level and Document level Attention (WWDA)
2. Document Level Attention (DLA)
3. Document Level Attention and Sentiment(DLAS)
4. Our Method Containing all the featured described

Model Variation	Accuracy
WWDA	0.842
DLA	0.858
DLAS	0.875
Our Method	0.891

Table 4: Ablation Study on the PHEME dataset

Conclusion

- **Enhanced Node Representation**: Our two-level attention mechanism effectively captures the nuances of source tweet claims, significantly improving rumour classification accuracy.
- **Sentiment Integration**: Incorporating sentiment analysis enhances detection capabilities, reflecting how emotions play a crucial role in the spread and perception of rumours.
- **Robust Model Training**: Augmentation through paraphrasing expands our dataset, increasing the model's adaptability and performance in diverse scenarios.