FaGANet: An Evidence-Based Fact-Checking Model with Integrated Encoder Leveraging Contextual Information

Weiyao Luo[1,2], Junfeng Ran[1,2], Zailong Tian[1,2], Sujian Li[1,3], Zhifang Sui[1,3,*]
[1]National Key Laboratory for Multimedia Information Processing, Peking University
[2]School of Software & Microelectronics, Peking University
[3]School of Computer Science, Peking University
wyluo@stu.pku.edu.cn; szf@pku.edu.cn

Abstract

In the face of the rapidly growing spread of false and misleading information in the real world, manual evidence-based fact-checking efforts become increasingly challenging and time-consuming. In order to tackle this issue, we propose FaGANet, an automated and accurate fact-checking model that leverages the power of sentence-level attention and graph attention network to enhance performance. This model adeptly integrates encoder-only models with graph attention network, effectively fusing claims and evidence information for accurate identification of even well-disguised data. Experiment results showcase the significant improvement in accuracy achieved by our FaGANet model, as well as its state-of-the-art performance in the evidence-based fact-checking task. We release our code and data in https://github.com/WeiyaoLuo/FaGANet.

Keywords: Evidence-Based Fact-Checking, Supervised Learning, Graph Attention Network

1. Introduction

The spread of false and misleading information on the internet is accelerating at an alarming rate, posing significant challenges to fact-checking efforts (Botnevik et al., 2020; Pradeep et al., 2021; Vosoughi et al., 2018). However, the verification process for misleading information can be time-consuming and often demands not only a straightforward analysis but also complex reasoning. Therefore, automating and accurately fact-checking false information is crucial. Fact-checking, as a key task in verifying the factuality of claims made in language, is an essential approach to addressing this challenge (Adair et al., 2017; Chen et al., 2022; Graves, 2018; Nielsen and McConville, 2022; Vo and Lee, 2018).

Early efforts in fact-checking focused on verifying the truthfulness of facts based solely on claims (Rashkin et al., 2017; Wang, 2017). However, relying solely on surface patterns of claims makes it difficult to identify subtle connections between claims and evidence (Schuster et al., 2020). Researchers then considered creating synthetic datasets by asking annotators to combine Wikipedia content to create claim and evidence datasets (Thorne et al., 2018; Aly et al., 2021). However, limiting world knowledge to a single source like Wikipedia is different from the diverse knowledge obtained through various media in the real world. To address this issue, some researchers used Google's returned summary snippets as evidence (Augenstein et al., 2019). Just as shown in Table 1, a real-world claim from Chinese social media and corresponding source document are retrieved through Google search engine. But in reality, summary snippets do not provide enough information to verify claims. Consequently, more recent efforts retrieved documents from web pages and selected relevant sentences as evidence (Hu et al., 2022, 2023). However, we believe that data obtained directly from web pages is the closest to the real-world scenario for fact-checking, which is the focus of this work.

Claim	Starting December 18, 2021, the minimum hourly wage for the privately regulated sector in Canada will increase to 15 Canadian dollars.
Evidence	By the end of this year, Canada's federal minimum wage plan will increase to $15 CAD per hour;...60,000 hourly workers in federally regulated private sectors earning less than 15 Canadian dollars will benefit from this minimum wage adjustment.
Label	0(supported); **Source**: Web pages

Table 1: An example of the evidence-based fact-checking task (Chinese text translated into English). For brevity, only the relevant snippet of the document is shown.

Some studies have shown that graph neural networks are helpful in capturing rich relationships between claims and multiple pieces of evidence in fact-checking tasks (Zhou et al., 2019; Velickovic et al., 2017). Specifically, graph attention network (GAT) combines graph neural networks and attention mechanisms to enable adaptive neighbor

*Corresponding author

LREC-COLING 2024, pages 7082–7088
20-25 May, 2024. © 2024 ELRA Language Resource Association: CC BY-NC 4.0

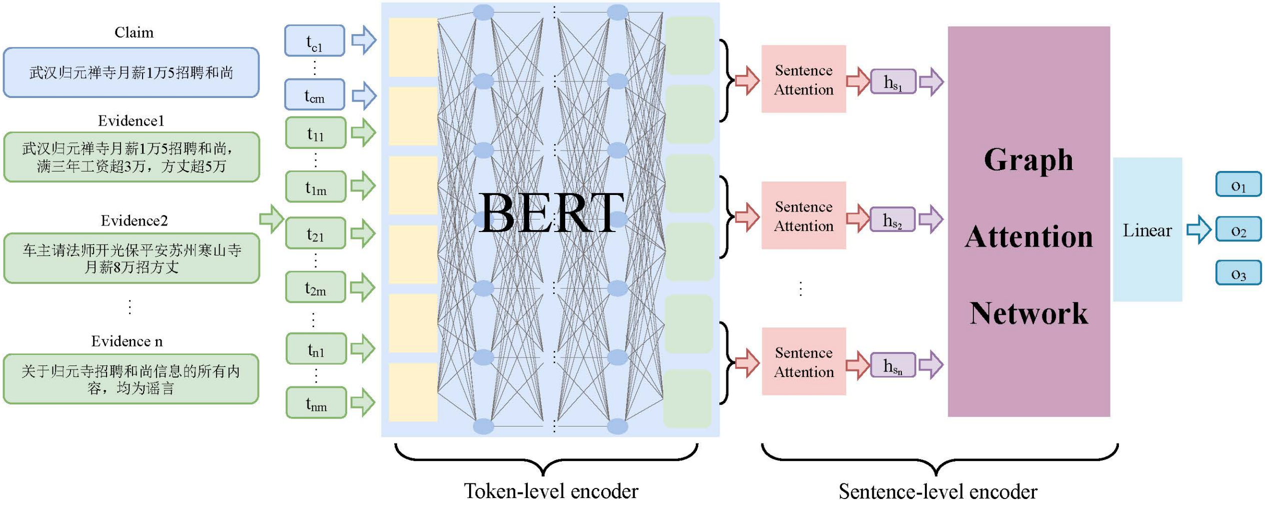

Figure 1: The implementation structure of the encoder which includes the token-level encoder part and the sentence-level encoder part. The token-level encoder generates the hidden representation of each token using the BERT-base model. The sentence-level encoder provides the hidden representation of each sentence using the attention mechanism and graph attention network. Finally, the linear layer is executed on the hidden representations of sentences.

aggregation on graph structures (Velickovic et al., 2017). GAT exhibit high performance in evidence-based fact-checking tasks, as they can flexibly assign different weights to each node, thereby better capturing the interrelationships between evidence. For instance, Liu et al. (2020) proposed a kernel graph attention network, in which node and edge kernels are applied on the graph for fine-grained evidence propagation.

Furthermore, we propose a sentence-level attention layer that calculates the weights of individual tokens within each sentence to obtain a comprehensive representation of the sentence. This provides suitable input for the subsequent GAT, further facilitating the establishment of interrelations among sentences.

Thus, we propose a **Fa**ct-checking **G**raph **A**ttention **Net**work called FaGANet, which combines graph attention networks for claim and evidence fusion.

The main contributions of this study are summarized as follows:

- To the best of our knowledge, we are the first to combine an encoder-only model with a graph attention network for the encoding stage of Chinese fact-checking tasks.

- We innovatively construct a sentence-level attention layer, transforming word-level feature representations into sentence-level features. This approach facilitates the fusion of information between claims and evidence.

- Experimental results demonstrate that our FaGANet achieves SOTA performance. In addition, our method also outperforms ChatGPT-CoT results on the fact-checking task.

In conclusion, our proposed FaGANet model, which incorporates a sentence-level attention layer on top of the BERT-base model and integrates a graph attention network, demonstrates enhanced accuracy and state-of-the-art performance in the evidence-based fact-checking task.

2. PROPOSED FaGANet

2.1. Encoding Overview

In this section, we will describe our proposed model, FaGANet, as illustrated in Figure 1.

To capture the hidden states of claims and evidence, our encoder incorporates a token-level BERT-base encoder, a sentence-level attention layer, and a graph attention layer. During preprocessing, claims and their corresponding evidence pieces are concatenated intelligently. In the preprocessing stage, we concatenate the claim and its corresponding five pieces of evidence in a suitable manner.

For each claim C and its corresponding pieces of evidence E, we combine them into a unified paragraph, which serves as the input for our model. This integration strategy facilitates capturing the contextual relationships between the claim and its associated evidence. In the paragraph, each sentence represents either the claim itself or a segment from one of the pieces of evidence. Due to the model's length limitation, the evidence used is derived from the original evidence employing the

Hybrid Ranker, as described in Hu et al. (2022). Following this step, the hidden states of all tokens within the paragraph are acquired using the BERT-base model. Subsequently, a sentence-level attention layer is employed to construct hidden states for individual sentences. In the final phase, interactions among the elements of the paragraph are established using a graph attention network (GAT) (Velickovic et al., 2017). This process culminates in the derivation of our desired hidden states, denoted as H.

2.2. BERT-base Model

We use the BERT-base model to obtain the hidden states of all tokens in the paragraph, resulting in H_P.

$$P = \{t_{11}, t_{12}, \ldots, t_{n1}, t_{n2}\}$$

$$H_P = \textbf{BERT}(P) = \{ht_{11}, ht_{12} \ldots, ht_{n1}, ht_{n2}\}$$

In this context, t_{ij} denotes the j-th token of the i-th sentence, and ht_{ij} represents the corresponding hidden state encoded by the BERT-base model. $H_P \in \mathbf{R}^{|P| \times d}$ denotes the hidden states of P with a d-dimensional hidden state.

2.3. Sentence-level Attention Layer

To extract sentence-level feature representations from the word-level hidden states, we employ a sentence-level attention layer that calculates the weights of individual words within each sentence, thereby obtaining a comprehensive representation of the sentence. The primary objective of this approach is to provide suitable input for the subsequent graph attention network, facilitating the establishment of interrelations among sentences and capturing the contextual information between claims and evidence.

To compute the hidden state hs_i of a sentence S_i, we take the weighted sum of the hidden states of all its tokens. H_P^{Si} denotes the corresponding sentence within the paragraph, obtained by merging the hidden states ht_{i1}, ht_{i2}, ... from several sentences. This representation encompasses both the claim and its associated evidence, allowing for a more comprehensive understanding of the context and relationships. After applying a Softmax function, we pass all the sentences through an attention layer, ultimately obtaining the hidden states H_S for each sentence. This process allows for a more comprehensive representation of the sentences, capturing the contextual information embedded within the paragraph.

$$hs_i = Softmax(w_i^\top H_P^{Si} + b_i) \cdot H_P^{Si}$$

$$H_S = Attn(H_P) = \{hs_1, hs_2, \ldots, hs_n\}$$

where w_i and b_i represent learnable parameters.

2.4. Graph Attention Network

In the final step, we employ the graph attention network (GAT) (Velickovic et al., 2017) to build up the interaction among paragraphs and capture the contextual relationships between the sentences. GAT is a specialized neural architecture for graph-structured data, to transform claims and evidence texts into interconnected graphs. Sentences are represented as nodes, and edges capture intricate sentence-level relationships. This strategic use of edges enables our model to extract pertinent features from the input sentences, enhancing the depth of analysis and information integration.

We utilize a single-layer graph attention layer to accomplish this. Each layer of the GAT consists of multiple attention mechanisms that compute the weights for the connections between sentences, enabling the model to focus on relevant information while ignoring irrelevant content. The attention mechanisms are learned during the training process, allowing the model to adapt to different input data structures and relationships.

The GAT can be formally represented as:

$$H = \textbf{GAT}(H_S) = h_1, h_2, \ldots, h_n$$

, where H_S denotes the input hidden states of the sentences, and h_i is the final hidden state output of sentence S_i using our encoder.

By incorporating GAT in our model, we can effectively capture the complex relationships between the claims and evidence within the paragraphs. This enables the model to better understand the underlying structure of the text and make more accurate predictions in the fact-checking task. Additionally, the GAT architecture provides a flexible and scalable approach to handling varying input sizes and graph structures, making it suitable for a wide range of applications.

The brief structure of our whole encoder can be summarized as

$$H = Encoder(T) = \mathbf{GAT}(\mathbf{Attn}(\mathbf{BERT}(\mathbf{T})))$$

3. Experiments and Analysis

3.1. Dataset and Settings

3.1.1. Dataset

To investigate the effectiveness of the proposed method, we conduct experiments on the CHEF dataset(Hu et al., 2022). The train/dev/test sets of CHEF comprise 8,002/999/999 samples, respectively. CHEF also offers summaries of source documents from Google Web pages as evidence (Gupta and Srikumar, 2021) and the previous best-performing approach (Hu et al., 2023) made further modifications based on this dataset, which

Table 2: The performance of different models.

System / Evidence		BERT-Based Model		RoBERTa-Based Model	
Condition	Model	Micro F1 Score	Macro F1 Score	Micro F1 Score	Macro F1 Score
(a) Pipeline	No Evidence	54.46	52.49	55.34	53.22
(b) Pipeline	Google Snippets	62.07	60.61	62.53	61.55
(c) Pipeline	Surface Ranker	63.17	61.47	64.21	62.05
(d) Pipeline	Semantic Ranker	63.47	61.94	64.35	62.24
(e) Pipeline	Hybrid Ranker	63.29	61.80	63.98	61.78
(f) Pipeline	ReRead	70.87	68.78	71.24	69.52
(g) Joint	Reinforce	64.37	62.46	65.04	63.05
(h) Joint	Multi-task	65.02	63.12	65.87	63.79
(i) Joint	Latent	66.77	64.65	66.95	65.13
(j) Pipeline	ChatGPT	35.14	33.51	35.14	33.51
(k) Pipeline	GPT-4	68.88	64.88	68.88	64.88
(l) Pipeline	FaGANet(w/o GAT)	72.07	69.88	72.77	70.95
(m) Pipeline	FaGANet(Ours)	**73.37**	**71.71**	**73.27**	**71.64**

led to an improvement in performance. In Gupta and Srikumar (2021), they provide claims and their corresponding original evidence, along with processed evidence obtained through Surface Ranker, Semantic Ranker, and Hybrid Ranker. Due to the constraints on model input length, we opt to utilize the content processed by the Hybrid Ranker as our evidence. Therefore, unlike Hu et al. (2023), we do not need to perform additional processing on the data. Instead, we conduct experiments directly on the original dataset to demonstrate the robust performance of our model. This approach allows us to showcase the effectiveness of our model without introducing extraneous modifications.

Following prior efforts (Gupta and Srikumar, 2021; Hu et al., 2022; Liu et al., 2020), we adopt Micro F1 and Macro F1 as evaluation metrics to assess the performance of our model.

3.1.2. Experimental Setup

For the base encoder, similar to Hu et al. (2023), we adopt BERT-Base-Chinese (Devlin et al., 2018) and RoBERTa-Base-Chinese (Liu et al., 2019). Our model is trained for a maximum of 30 epochs using the AdamW optimizer, which features an initial learning rate of 1e-5, a weight decay of 0.01, and a warm-up rate of 0.1. For regularization, we use the dropout with a dropout rate of 0.1. The dimension of each hidden state is 768, both in the BERT-base model and in the graph attention network. The model was run on NVIDIA RTX-3090 GPUs.

3.2. Results and Analysis

We investigate the performance of the FaGANet and evaluate it on the CHEF test set. We compare our method to Hu et al. (2022, 2023) which are recently proposed models for the Chinese fact verification task as shown in Table 2. From serial number (a) to (i) are the previous works. Cao et al. (2023) obtained results on the GPT-3.5-turbo. We also get our own GPT-4 evaluation results by gpt-4-1106-preview API(OpenAI, 2023) based on Chain-of-Thought(Wei et al., 2022) prompt. The prompt we used for GPT-4 evaluation is provided in Appendix A.

Several observations can be derived from the results: 1) In the absence of evidence input, the upper bound of F1 stands at 55.34, indicating a relatively low performance; 2) The utilization of real-world evidence enhances the effectiveness of claim verification, with source documents providing greater improvements than Google snippets, which can be attributed to the fact that source documents contain more comprehensive information; 3) ChatGPT's results indicate limited zero-shot effectiveness in addressing the fact-checking task, potentially due to inadequate capture of complex inter-sentence relationships. In contrast, GPT-4 achieved better performance, with a Micro F1 score of 68.88. In future research, we will explore the potential of harnessing ChatGPT for tackling the fact-checking task. 4) Compared to the previous SOTA model ReRead(Hu et al., 2023), FaGANet achieves a Micro F1 score of 73.37, which surpasses ReRead (70.87 Micro F1) with a +2.5% relative improvement. ReRead attributes its performance enhancement to the faithful and plausible evidence retrieved from source documents. In contrast, we do not adopt additional measures to process the evidence extraction procedure but instead utilize the raw data as in Hu et al. (2022). To demonstrate the effectiveness of GAT, we present the results of FaGANet without GAT. The comparison highlights the improvement in evaluation metrics with the incorporation of relationships. Moreover, in contrast to ReRead, we enhance the BERT-base model by introducing a sentence-level attention layer. This layer transforms word-level features into sentence-level representations, further harnessing the superiority of

GAT.

This finding demonstrates that the FaGANet model is capable of effectively capturing the contextual relationships between claims and evidence, thereby achieving information fusion between the two. Furthermore, even without utilizing the meticulously curated data from ReRead, FaGANet consistently achieves superior performance in the fact-checking task. This further substantiates the robustness of our proposed model.

4. Conclusions

In this paper, we propose a novel evidence-based fact-checking framework, FaGANet, which effectively integrates information between claims and evidence by incorporating a graph attention network. Contrary to prior models, our approach emphasizes the utilization of both a sentence-level attention mechanism and GAT within the encoder. This not only enables deep integration of claims and evidence during the encoding process but also allows the model to attend to contextual information. As a result, we successfully enhance the model capability and achieve SOTA results on the CHEF dataset without performing additional operations on the dataset.

5. Acknowledgements

This paper is supported by the National Key Research and Development Program of China 2020AAA0106700 and NSFC project U19A2065.

6. References

Bill Adair, Chengkai Li, Jun Yang, and Cong Yu. 2017. Progress toward “the holy grail”: The continued quest to automate fact-checking. In *Computation+ Journalism Symposium,(September)*.

Rami Aly, Zhijiang Guo, Michael Sejr Schlichtkrull, James Thorne, Andreas Vlachos, Christos Christodoulopoulos, Oana Cocarascu, and Arpit Mittal. 2021. The fact extraction and VERification over unstructured and structured information (FEVEROUS) shared task. In *Proceedings of the Fourth Workshop on Fact Extraction and VERification (FEVER)*, pages 1–13, Dominican Republic. Association for Computational Linguistics.

Isabelle Augenstein, Christina Lioma, Dongsheng Wang, Lucas Chaves Lima, Casper Hansen, Christian Hansen, and Jakob Grue Simonsen. 2019. MultiFC: A real-world multi-domain dataset for evidence-based fact checking of claims. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4685–4697, Hong Kong, China. Association for Computational Linguistics.

Bjarte Botnevik, Eirik Sakariassen, and Vinay Setty. 2020. Brenda: Browser extension for fake news detection. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*, pages 2117–2120.

BSI. 1973a. *Natural Fibre Twines*, 3rd edition. British Standards Institution, London. BS 2570.

BSI. 1973b. Natural fibre twines. BS 2570, British Standards Institution, London. 3rd. edn.

Han Cao, Lingwei Wei, Mengyang Chen, Wei Zhou, and Songlin Hu. 2023. Are large language models good fact checkers: A preliminary study. *arXiv preprint arXiv:2311.17355*.

A. Castor and L. E. Pollux. 1992. The use of user modelling to guide inference and learning. *Applied Intelligence*, 2(1):37–53.

Jiangui Chen, Ruqing Zhang, Jiafeng Guo, Yixing Fan, and Xueqi Cheng. 2022. Gere: Generative evidence retrieval for fact verification. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2184–2189.

J.L. Chercheur. 1994. *Case-Based Reasoning*, 2nd edition. Morgan Kaufman Publishers, San Mateo, CA.

N. Chomsky. 1973. Conditions on transformations. In *A festschrift for Morris Halle*, New York. Holt, Rinehart & Winston.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Umberto Eco. 1990. *The Limits of Interpretation*. Indian University Press.

D Graves. 2018. Understanding the promise and limits of automated fact-checking. *Reuters Institute for the Study of Journalism*.

Ashim Gupta and Vivek Srikumar. 2021. X-fact: A new benchmark dataset for multilingual fact checking. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 675–682, Online. Association for Computational Linguistics.

Paul Gerhard Hoel. 1971a. *Elementary Statistics*, 3rd edition. Wiley series in probability and mathematical statistics. Wiley, New York, Chichester. ISBN 0 471 40300.

Paul Gerhard Hoel. 1971b. *Elementary Statistics*, 3rd edition, Wiley series in probability and mathematical statistics, pages 19–33. Wiley, New York, Chichester. ISBN 0 471 40300.

Xuming Hu, Zhijiang Guo, GuanYu Wu, Aiwei Liu, Lijie Wen, and Philip Yu. 2022. CHEF: A pilot Chinese dataset for evidence-based fact-checking. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3362–3376, Seattle, United States. Association for Computational Linguistics.

Xuming Hu, Zhaochen Hong, Zhijiang Guo, Lijie Wen, and Philip Yu. 2023. Read it twice: Towards faithfully interpretable fact verification by revisiting evidence. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '23, page 2319–2323, New York, NY, USA. Association for Computing Machinery.

Otto Jespersen. 1922. *Language: Its Nature, Development, and Origin*. Allen and Unwin.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Zhenghao Liu, Chenyan Xiong, Maosong Sun, and Zhiyuan Liu. 2020. Fine-grained fact verification with kernel graph attention network. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7342–7351, Online. Association for Computational Linguistics.

Dan S Nielsen and Ryan McConville. 2022. Mumin: A large-scale multilingual multimodal fact-checked misinformation social network dataset. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 3141–3153.

OpenAI. 2023. Gpt-4 technical report.

Ronak Pradeep, Xueguang Ma, Rodrigo Nogueira, and Jimmy Lin. 2021. Vera: Prediction techniques for reducing harmful misinformation in consumer health search. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2066–2070.

Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pages 2931–2937.

Tal Schuster, Roei Schuster, Darsh J Shah, and Regina Barzilay. 2020. The limitations of stylometry for detecting machine-generated fake news. *Computational Linguistics*, 46(2):499–510.

Charles Joseph Singer, E. J. Holmyard, and A. R. Hall, editors. 1954–58. *A history of technology*. Oxford University Press, London. 5 vol.

Jannik Strötgen and Michael Gertz. 2012. Temporal tagging on different domains: Challenges, strategies, and gold standards. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 3746–3753, Istanbul, Turkey. European Language Resource Association (ELRA).

S. Superman, B. Batman, C. Catwoman, and S. Spiderman. 2000. *Superheroes experiences with books*, 20th edition. The Phantom Editors Associates, Gotham City.

James Thorne, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal. 2018. FEVER: a large-scale dataset for fact extraction and VERification. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 809–819, New Orleans, Louisiana. Association for Computational Linguistics.

Nicolas Turenne. 2018. The rumour spectrum. *PloS one*, 13(1):e0189080.

Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Bengio, et al. 2017. Graph attention networks. *stat*, 1050(20):10–48550.

Nguyen Vo and Kyumin Lee. 2018. The rise of guardians: Fact-checking url recommendation to combat fake news. In *The 41st international ACM SIGIR conference on research & development in information retrieval*, pages 275–284.

Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *science*, 359(6380):1146–1151.

William Yang Wang. 2017. " liar, liar pants on fire": A new benchmark dataset for fake news detection. *arXiv preprint arXiv:1705.00648*.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Ed Huai hsin Chi, F. Xia, Quoc Le, and Denny Zhou. 2022. Chain of thought prompting elicits reasoning in large language models. *ArXiv*, abs/2201.11903.

Zhi-Xiu Ye and Zhen-Hua Ling. 2019. Multi-level matching and aggregation network for few-shot relation classification. *arXiv preprint arXiv:1906.06678*.

Jie Zhou, Xu Han, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. 2019. GEAR: Graph-based evidence aggregating and reasoning for fact verification. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 892–901, Florence, Italy. Association for Computational Linguistics.

A. Prompt for GPT-4 Evaluation

Please use the evidence to determine whether the following claim is supported by it (label 0), refuted by it (label 1), or the evidence does not provide sufficient information to make a judgment (label 2).
Claim: *{claim}*
Evidence 1: *{evidence1}*
Evidence 2: *{evidence2}*
Evidence 3: *{evidence3}*
Evidence 4: *{evidence4}*
Evidence 5: *{evidence5}*
Judging from the evidence, the label is