CROSS-DOCUMENT MISINFORMATION DETECTION BASED ON EVENT GRAPH REASONING

BY

XUEQING WU

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Computer Science
in the Graduate College of the
University of Illinois Urbana-Champaign, 2023

Urbana, Illinois

Adviser:

Professor Heng Ji

ABSTRACT

For emerging events, human readers are often exposed to both real news and fake news. Multiple news articles may contain complementary or contradictory information that readers can leverage to help detect fake news. Inspired by this process, we propose a novel task of **cross-document misinformation detection**. Given a cluster of topically related news documents, we aim to detect misinformation at both *document level* and a more fine-grained level, *event level*. Due to the lack of data, we generate fake news by manipulating real news, and construct 3 new datasets with 422, 276, and $1,413$ clusters of topically related documents, respectively. We further propose a graph-based detector that constructs a cross-document knowledge graph using cross-document event coreference resolution and employs a heterogeneous graph neural network to conduct detection at two levels. We then feed the event-level detection results into the document-level detector. Experimental results show that our proposed method significantly outperforms existing methods by up to 7 F1 points on this new task. Codes and data are at `https://github.com/shirley-wu/cross-doc-misinfo-detection`.

ACKNOWLEDGMENTS

I would like to express my sincerest appreciation and gratitude to those who have contributed to the completion of my master's thesis and have been an integral part of my academic journey.

First and foremost, I would like to thank my advisor, Professor Heng Ji, for her unwavering support and guidance. Her expertise and encouragement have been invaluable not only in my research but also in shaping my future aspirations. Professor Ji has been instrumental in helping me realize my potential and navigate the path towards my professional goals, for which I am truly grateful.

I extend my heartfelt gratitude to the Blender group, a nurturing and friendly academic family that has provided me with a supportive and stimulating environment. I am especially thankful to those who have directly contributed to my projects: Yi Fung, Kung-Hsiang (Steeve) Huang, Sha (Zoey) Li, Pengfei Yu, and Payam Karisani. Their expertise, camaraderie, and encouragement have enriched my academic experience and made this journey all the more enjoyable.

Finally, I would like to acknowledge the delightful Green Street area. The diverse and delicious food offerings have provided much-needed sustenance and moments of joy during the long hours of research and writing. These culinary delights have been a constant source of comfort and inspiration, making Green Street an unforgettable part of my academic journey.

I am truly grateful for the contributions of everyone who has been a part of this incredible adventure, and I look forward to carrying the lessons and memories with me as I embark on the next chapter of my life.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

In recent years, the dissemination of fake news has emerged as a significant social issue, causing confusion, fear, and mistrust among the public. The rapid spread of misleading information through various media channels, particularly on social media platforms, has been amplified by the rise of advanced generative neural network models in natural language processing [1] and computer vision [2]. Using these techniques makes it possible for malicious attackers to generate misinformation at a large scale. These AI-generated news articles and images are realistic, and highly convincing, which can deceive even discerning readers. Studies performed by [1] show that human readers perceive disinformation generated by generative language models as more trustworthy than human-written articles.

The difficulty in discerning truth from falsehoods has grown increasingly challenging due to these developments, leading to profound consequences on public opinion, elections, policy decisions, and even public safety. As a result, there is an urgent need for effective misinformation detection strategies to address this issue. The task of misinformation detection is not only essential but also demands constant innovation to adapt to the evolving tactics employed by those who spread false information. In this context, combating the dissemination of fake news and improving misinformation detection methods are essential steps towards preserving the integrity of our information ecosystem and fostering a more informed society.

For this purpose, many fake news detection methods have been developed, relying on either textual features [1, 3, 4], multi-media features [5, 6], or social context features [7, 8]. However, most of the existing work on fake news detection is limited to judging each document in isolation, which is different from the way human readers distinguish fake news.

For emergent complex events such as COVID-19, Russia-Ukraine war and US election, human readers are usually exposed to multiple news documents, where some are real and others are fake. News documents from different sources naturally form a cluster of topically related documents. We notice that articles about the same topic may contain conflicting or complementary information, which can benefit the task of misinformation detection. An example is shown in Figure 1.1. As shown in the knowledge graph, the death of Rosanne Boyland in 2021 US Capitol attack is a shared event across all four documents. Each document is internally consistent, making it difficult to identify misinformation when judging each news separately. However, the three real news documents complement each other's statements regarding the death of Boyland, while the fake news document contradicts the other stories. Such cross-document connections can be leveraged to help detect misinformation.

Based on such motivation, we propose a novel task of **cross-document misinformation**

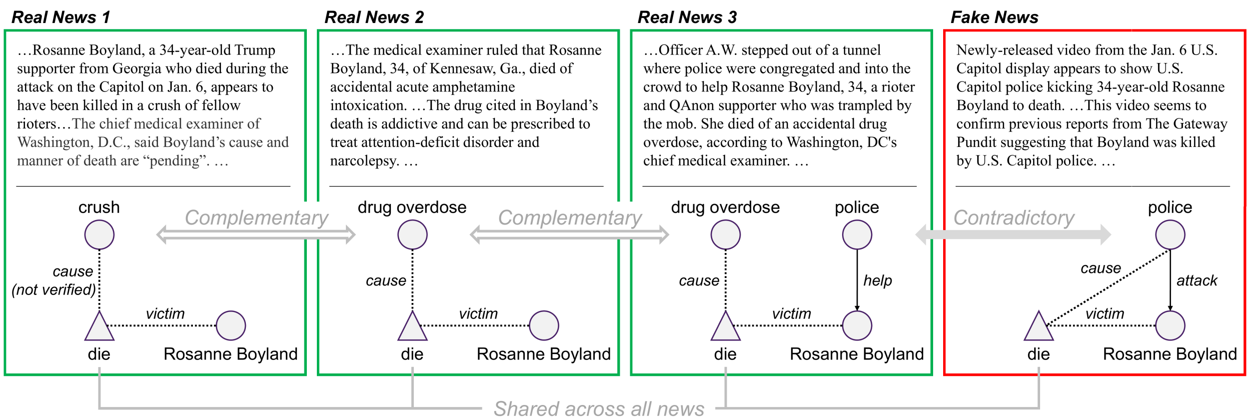

Figure 1.1: An example of cross-document misinformation detection, including the texts and knowledge graphs for four news documents. The three real news documents complement each other, while the fake news contradicts the other news. News 1 falsely speculates that Boyland was crushed to death, but it admits that the cause of death was not yet verified. News 2 and 3 complete the story by reporting that Boyland died of drug overdose. The fake news claims that Boyland was killed by police, which contradicts the other news. Additionally, the fake news states that the police attacked Boyland, which is inconsistent with News 3's claim that the police was trying to help her.

detection that aims to detect fake information from a cluster of topically related news documents. We perform the task at both the document level and event level. Each event describes a specific type of real-world event mentioned in the text (e.g., the death of Boyland in Figure 1.1), and usually involves certain participants to represent different aspects of the event (e.g., the cause of death and the victim of the death event). **Document-level detection** aims to detect fake news documents. **Event-level detection** is a more fine-grained task that aims to detect fake events, thereby pinpointing specific fake information in news documents. The problem of fine-grained misinformation detecton is, though very important, rarely explored in the existing literature. The most relevant work detects triplets of false knowledge [6]. However, we focus on identifying false events instead of relations or entities, because events are more important for storytelling and easier to compare across multiple documents through cross-document coreference resolution.

To effectively conduct research on cross-document misinformation detection, it is crucial to have relevant datasets for training and evaluating misinformation detection systems. However, to the best of our knowledge, there are currently no existing datasets specifically designed for fake news detection that contain clusters of topically related documents. Therefore, we construct 3 new benchmark datasets based on existing real news corpus with such clusters. To construct clusters of documents containing both real news and fake news, we use generative language models to generate the fake news and mix them with the original real news articles. Following [6], we train a generator that generates a document from a knowledge graph (KG), and feed manipulated KGs into the generator to generate fake news documents. By tracking the manipulation operations, we are also able to obtain supervision for event-level detection.

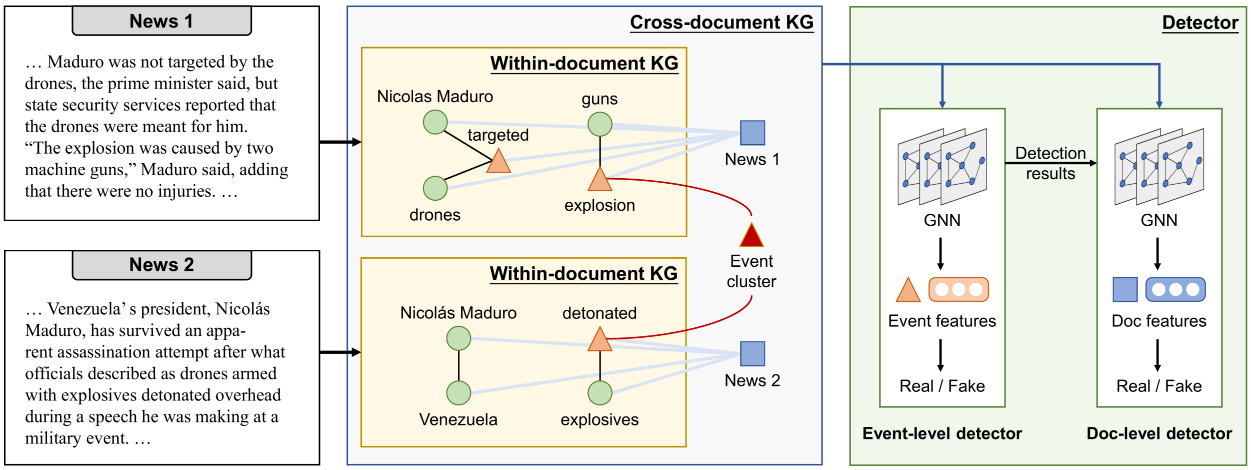

Figure 1.2: An overview of our approach. We first construct a within-document KG for each document based on IE output, where ○ represents an entity and △ represents an event. Then, we construct a cross-document KG by (1) adding a node for each cross-document event cluster and connecting it with events in the cluster, and (2) introducing a document node □ for each document and connecting it with all entities and events in the given document. Finally, we use GNN to encode the cross-document KG, and use the event and document features to conduct misinformation detection at two levels. The two detectors are trained and deployed in a pipeline fashion, where event-level detection results are leveraged to improve document-level detection.

We further propose a detection approach as shown in Figure 1.2. Given a cluster of documents, we first use an information extraction (IE) system [9] to construct a within-document KG for each document. The KG contains entities and nodes as edges, and relations and event arguments as edges. Then, we connect the within-document KGs to form a cross-document KG using cross-document event coreference resolution [10, 11]. Eventually, we use a heterogeneous graph neural network (GNN) to encode the cross-document KG and conduct detection at two levels. The document and event features produced by GNN are further utilized for document-level and event-level misinformation detection respectively. Moreover, we show that event-level detection results serve as useful features for enhancing the document-level detector. Experiments on 3 datasets demonstrate the effectiveness of our proposed method.

Our contributions are summarized as follows:

1. We propose the novel task of cross-document misinformation detection, and conduct

the task at two levels, document level and the more fine-grained event level.

2. We construct 3 new datasets for our proposed task based on existing document clusters categorized by topics.

3. We propose a detector that leverages cross-document information and improve document-level detection by utilizing features produced by the event-level detector. Experiments on three datasets demonstrate that our method significantly outperforms existing methods.

The remaining sections are organized as follows: Chapter 2 presents a literature review covering topics including misinformation detection, fact verification and information extraction. Chapter 3 introduces the formulation of our proposed task, followed by Chapter 4, which outlines the proposed method, including knowledge graph construction, knowledge graph encoder, and the misinformation detector. Chapter 5 discusses our dataset construction approach and presents details about our constructed datasets. In Chapter 6, the results and discussion are presented, including the experimental settings, document-level detection results, event-level detection results, and an in-depth analysis and discussion of the findings. Finally, Chapter 7 offers conclusions drawn from this study and explores possible future directions for this research area.

CHAPTER 2: LITERATURE REVIEW

2.1 MISINFORMATION DETECTION

2.1.1 Methods

Early work on fake news detection uses hand-crafted features such as stylistic features and linguistic features to perform misinformation detection [12, 13, 14, 15, 16, 17]. Later work uses neural networks such as convolutional neural network and recurrent neural networks to encode the document [18, 19]. With the development of Transformer and pre-trained language models, [1] built their system on top of the pre-trained GPT-2 model [20]. They train the model on a large number of news documents using auto-regressive language model loss, and then fine-tune the model to detect the authenticity for each document. The model is able to capture useful features to characterize news documents and achieve impressive detection accuracy.

Some other work aims to capture more complicated internal structure of a news document beyond unstructured text. For example, [21, 22, 23] encode each sentence separately and uses inter-sentence relations to aggregate sentence encodings into a document representation. [21] parses the document into an inter-sentence dependency tree where a parent-child link indicates that the child sentence semantically relies on the parent sentence. [22] constructs a graph-based representation for each document, where edges denote the semantic similarity between sentence pairs, and applies a graph neural network to obtain graph encodings. [23] enriches the graph structure by incorporating topic and entity nodes in addition to sentence nodes. Similar to our approach, they compare the entity node features against an external knowledge base (KB) to check for inconsistencies. However, the correlation between news and KB is not as close as the correlation between related news documents due to the incompleteness of these KBs. [24] performs fake news detection based on information triplets extracted from news documents, while [6] relies on knowledge graph extracted via IE systems.

Since the information contained in plain text is limited, many work has attempted to utilizes additional information. One line of work uses social context information such as user profiles and social relationships. For example, [25, 26] uses the user stances to infer news veracity. [7, 25, 27] propagates the credibility prediction within a social graph based on the assumption that the credibility of a news document is closely related to the credibilities of its relevant social media posts. Another line of work uses multi-media features such as images and captions. [5] encodes text and images using LSTM and convolutional neural network respectively and then fuses the features for fake news detection. To utilize more fine-grained visual information, [28] computes encodings for each object in the image and uses them for multi-modal reasoning. [6] further constructs a cross-modal knowledge graph with objects, entities and events as nodes. However, to the best of our knowledge, no published work has considered using cross-document inference for misinformation detection.

Another shared limitation of most existing work is that they solely predict the authenticity for the overall document. However, the task of fine-grained detection is also important but rarely explored. The most relevant work detects fake knowledge triplets extracted from each individual news article [6].

2.1.2 Resources

Numerous resources have been introduced to support research in fake news detection. The main difficulty in constructing a fake news dataset is to obtain annotations, since obtaining large-scale human-annotated datasets can be highly expensive.

To easily collect document labels, some work obtains labels from the source information. This involves collecting news articles from multiple sources, considering news from reliable sources as real news, and news from unreliable sources as fake news. For example, [29] uses news documents from Gigaword News as real news and gathers fake news from seven unreliable sources such as The Onion. However, this approach may result in noisy labels, as information from reliable sources is not always accurate, and information from unreliable sources is not always false. Another potential issue is that such data collection methods may inadvertently lead the model to focus on stylistic differences across sources rather than learning to effectively distinguish between real and fake information.

Other work relies on human evaluation to obtain authenticity labels. [30] collects real news from trustworthy sources and pairs each real news document with a fake news document. The pairing is achieved either by asking crowd-sourcing annotators to write fake news based on the real news or by manually collecting matching fake news from the web. Their resulting dataset, though of high quality, contains only 340 pairs of fake and real news articles. [31] collects data from fact checking websites, where journalists and domain experts manually review the news articles and provide evaluation results to claim news articles as fake or real. They collect ~1000 documents for political domain and ~20,000 documents for entertainment news.

With the development of generative models, pre-trained generative language models have demonstrated the ability to produce high-quality fake documents that can deceive humans

[1]. Therefore, recent work has started utilizing generative models to generate high-quality fake news in order to construct more challenging fake news detection datasets. For example, [1] trains a fake news generator with the same architecture as GPT-2 [20] on a large-scale news corpus. Using this generator, they build a misinformation detection dataset containing 5000 human-written real news and 5000 generated fake news. [6] achieves finer-grained control of generated content by conditioning the text generator on a knowledge graph. [32] further improves the factual consistency within the generated document by retrieving relevant information from external corpora. [33] employs adversarial reinforcement learning to train the model to stick to a given topic. Our work follows this line of work to construct our dataset using neural text generators.

2.2 FACT VERIFICATION

The task of fact verification aims to verify the veracity of a given statement. Most work in this area assumes the existence of a corpus of articles such as Wikipedia articles that serve as evidence. The fact verification system is supposed to identify the correct sentence-level evidence from the corpus, and then decide if the claim is SUPPORTED or REFUTED by the evidence. In the case that there is not enough information in the evidence corpus, the veracity of the claim should be assessed as NOTENOUGHINFO (NEI). The most influential dataset for the fact verification task is FEVER [34]. However, since FEVER is developed for general domain, many other resources have been proposed for specific domain's requirements such as climate change [35] and COVID-19 [36, 37].

The pipeline approach to fact verification decomposes the task into three subtasks and solves them with three separate modules as follows:

1. **Document retrieval** aims to match the claim against the evidence corpus and selects the most relevant articles. Existing work typically performs document retrieval using hand-crafted features such as TF-IDF [34], named entity matching [38, 39], or keyword matching [40].

2. **Sentence selection** aims to find the most relevant sentences within the selected documents. Similar to document retrieval, many existing work [41, 42, 43] relies on statistical features like TF-IDF. However, recent work usually adopts neural networks to match the claim to candidate evidence sentences. For instance, [39, 44] employ Enhanced LSTM (ESIM), and [45, 46, 47, 48] utilize Transformer-based pre-trained language models like BERT. The neural networks are trained using either binary classification loss or hinge loss. The former trains the network to determine whether an

evidence sentence is relevant to the claim or not, while the latter trains the network to assign higher scores to relevant evidence sentences compared to irrelevant ones.

3. **Claim verification**, the core subtask of fact verification, aims to verify the veracity of a given claim (i.e., SUPPORTED, REFUTED, or NEI) based on the retrieved evidence sentences. For this subtask, most research relies on neural networks. [39] uses ESIM to infer the relevance between evidence and claims. [40] further designed neural semantic matching network (NSMN), a more powerful modification of ESIM. Since pre-trained language models have shown great performance across various NLP tasks, many work also applies BERT-like large language models to claim verification [42, 46, 47]. [49] adopts a graph-based approach where the claim and evidence sentences are represented as nodes in a graph. They then employ a graph neural network to reason over the multiple pieces of evidence and aggregate the results.

In addition to the pipeline approach, a growing body of research is exploring joint approaches that address multiple subtasks using a single model. The motivation is to avoid error propagation among different modules. For instance, [50, 51] trained a single neural network for both sentence selection and claim verification subtasks using a multi-task training strategy. Retrieval-augmented generation (RAG) [52] addresses all three subtasks with one model by first retrieving evidence using a neural retriever and then predicting the veracity based on the retrieved evidence.

The task of fact verification and misinformation detection share a high-level goal of identifying false information. Additionally, fact verification determines the veracity of a claim by comparing its content against an evidence corpus. This approach is similar with our method that predicts the authenticity of a document by comparing its content with other documents. However, fact verification primarily focuses on short, single-sentence statements, limiting its ability to model the complex internal structure of news articles.

2.3 INFORMATION EXTRACTION

Information Extraction (IE) is the task of extracting information (structured data) from a text (unstructured data). For example, named entity recognition (NER) recognizes entities appearing in a text. Relation extraction (RE) identifies the relationships between entities. Event extraction (EE) discovers events occurring in a text.

Traditionally, researchers formalize the task as a language understanding problem and solve it in a pipeline manner. The state-of-the-art methods for NER perform the task on the basis of the pre-trained language model BERT [53]. The pipeline approach to RE divides the problem into NER and relation classification, and conducts the two sub-tasks in a sequential manner [54]. Similarly, the task of EE is typically divided into two sub-tasks, event trigger extraction and event argument extraction, which are performed sequentially using pre-trained BERT [55].

Recently, there is a growing trend of joint training IE models for multiple IE tasks. This approach aims to minimize error propagation and enhance prediction accuracy by allowing interactions between components. For example, [56, 57, 58] aim to address the RE task by jointly optimizing the model for both NER sub-task and relation classification sub-task. The state-of-the-art methods for EE usually jointly train the models with other tasks including NER and RE [59, 60, 61]. Particularly, OneIE [61] performs each sub-task separately with task-specific classifiers, and then employs beam search to obtain the globally optimal IE outputs. In this way, they explicitly model the interdependencies among tasks and instances. In this work, we use OneIE as our IE model to produce knowledge graph.

The outputs of the IE system, including entities, entity-to-entity relations and events, can be arranged into a knowledge graph representing the content of a document. Entities are represented as nodes while entity-to-entity relations are edges between entity nodes. Additionally, event triggers are represented as nodes, and event arguments are represented as edges between event nodes and entity nodes. Such representation can be further utilized to assist downstream NLP tasks such as fake news detection and generation [6].

In the context of cross-document analysis, the task of cross-document IE is necessary. This involves extracting correlated information from multiple documents, such as cross-document event and entity coreference resolution. Existing work [62, 63] in this area directly adapts single-document IE approaches to the cross-document scenario by concatenating multiple documents into one and applying the original single-document IE methods. In this study, we adopt a similar approach for cross-document event extraction, which allows us to construct a cross-document knowledge graph.

CHAPTER 3: TASK FORMULATION

As discussed in Chapter 1, our motivation for proposing the cross-document misinformation detection task is based on the way human readers detect fake news. In the context of complex events, such as the COVID-19 pandemic, political conflicts, or elections, human readers often consult multiple sources to evaluate the reliability of news articles. By analyzing these clusters as a whole, it becomes possible to identify inconsistencies or confirmations in the information, which can be beneficial in detecting misinformation.

In this chapter, we formally define the task of cross-document misinformation detection. This task is performed at both the document and event levels, allowing for a more comprehensive and fine-grained analysis of the information presented in the documents. Each event represents a specific type of real-world occurrence mentioned in the text, typically involving participants that represent different aspects of the event. Given a cluster of topically related news documents, where each document contains a collection of events mentioned within the text, our goal is to identify both fake events and fake documents, thereby pinpointing specific instances of misinformation in the news documents.

Formally, let $\mathbf{S} = \{\mathbf{d}_1, \cdots, \mathbf{d}_N\}$ be the document cluster, and $N = |\mathbf{S}|$ be the size of the cluster. Some documents in $\mathbf{S}$ are real, while others are fake. From each document $\mathbf{d} \in \mathbf{S}$, we extract events $\mathbf{E}(\mathbf{d}) = \{e_1, \cdots, e_m\}$, where $m = |\mathbf{E}(\mathbf{d})|$ is the number of events in document $\mathbf{d}$. In an extracted event set $\mathbf{E}(\mathbf{d})$, some events are real and others are fake. Given $\mathbf{S}$ as input, our task consists of two sub-tasks:

1. **Document-level detection**, as in traditional fake news detection tasks, aims to predict whether each document $\mathbf{d} \in \mathbf{S}$ is real or fake;

2. **Event-level detection** is a more fine-grained task that aims to predict whether each event $e \in \mathbf{E}(\mathbf{d}), \mathbf{d} \in \mathbf{S}$ is real or fake. A fake event means that the textual description regarding the event contains misinformation. In the example in Figure 1.1, the *die* event in the fake news is fake, since it falsely describes Boyland being killed by the police, but she actually died of drug overdose. In real-world applications, document-level detection helps provide an overall judgement for each document, while event-level detection helps pinpoint the misinformation in the fake documents.

CHAPTER 4: PROPOSED METHOD

An overview of our approach is shown in Figure 1.2. Given a cluster of documents, we first construct a within-document KG for each document using an IE system [9], and then connect the within-document KGs into a cross-document KG using cross-document event coreference resolution. Based on the cross-document KG, we use a hetereogeneous GNN [64, 65] to perform detection. We further incorporate the results of event-level detection to help the document-level detector.

4.1 KNOWLEDGE GRAPH CONSTRUCTION

4.1.1 Within-document KG

We first construct a within-document IE-based knowledge graph for each document.

We leverage OneIE [9], a BERT-based end-to-end IE system to extract entities, relations, and events. OneIE conducts IE in four steps: (1) encode a sentence with a pre-trained BERT encoder, (2) identify entity mentions and event triggers using a conditional random fields layer, (3) classify types of entity mentions, events, entity relations, and event arguments using feed-forward networks, and (4) search for a globally optimal IE graph via beam search. In this work, we use the model released by [11]. The model achieves 64.1, 49.7, and 49.5 F1 on trigger extraction, argument extraction and relation extraction respectively on ACE 2005 and ERE [66].

In addition, we use entity linking and entity coreference resolution to identify coreferential entity mentions. For **entity linking**, we use an LSTM-based entity linker to link [67] to link entity mentions to WikiData entries. The entity linker achieves 91.8 F1 and 84.3 accuracy. For **entity coreference resolution**, we use an extension of the e2e-coref model [68] based on XLM-RoBERTa [69]. The model is released by [11] and achieves a 92.4 CoNLL score on OntoNotes [70]. Thus, entity mentions that are linked to the same WikiData entry or identified as coreferences will be considered as the same entity, and their entity nodes in the KG will be merged.

Eventually, we obtain a within-document KG where entities and events are nodes, relations are edges between entities, and arguments are edges between events and entities.

4.1.2 Cross-document KG

We leverage cross-document event coreference resolution to connect the within-document KGs into a cross-document KG as illustrated in Figure 1.2.

We employ a cross-document event coreference resolution system to identify clusters of events from multiple documents that refer to the same real-world events. The system is based on [10], a within-document coreference resolution model. To identify event coreference, [10] utilizes both textual contexts of the event mentions and symbolic features such as the event type information. We extend it to the cross-document scenario following [11]. Given a cluster containing N documents, we concatenate each pair of documents into a "mega-document". The model then conducts coreference resolution on each mega-document. More specifically, for each event mention, the model uses SpanBERT [71] to extract contextualized text embeddings and builds manually designed symbolic features such as event types, attributes, and arguments. Then, the two features are combined selectively using a gated mechanism. Eventually, for each pair of event mentions in a mega-document, the model predicts whether they are coreferential. In this work, we use the model released by [11]. The model achieves 84.8 CoNLL score on ACE 2005.

An example of the detected event cluster is shown in Table 4.1, where the four events of four documents all refer to the same explosion attack on Venezuela's President Nicolas Marduro. These four events contain complementary or contradictory details, which can be used for misinformation detection. For each event cluster, we add a node to represent the overall information of the real-world complex event corresponding to the cluster. Then, an edge is added between each event node and corresponding cluster node to allow reasoning among cross-document coreferential events.

To indicate which document each entity or event belongs to and capture the global information of each document, we further introduce a document node and connect it to the associated entity and event nodes for each document.

The resulting KG contains 4 types of nodes (i.e. entity nodes, event nodes, document nodes, and event cluster nodes) and 5 types of edges (i.e. relation edges, event argument edges, document-to-entity edges, document-to-event edges, and edges connecting event nodes to event cluster nodes). Since all edges are directional, we add an inverse edge for each edge to propagate features along both directions. The final KG contains 10 edge types, accounting for the inverse of existing edge types.

Real	⋯Venezuela' s president, Nicolás Maduro, has survived an apparent assassination attempt after what officials described as drones armed with $\text{explosives}_{\text{arg1}}$ $\textbf{detonated}_{\textbf{trig}}$ overhead during a speech he was making at a military event. ⋯
Real	⋯The BBC quotes anonymous firefighters at the scene who say "the incident was actually a gas tank $\textbf{explosion}_{\textbf{trig}}$ inside an $\text{apartment}_{\text{arg2}}$, but did not provide further details." ⋯
Fake	⋯Maduro was not targeted by the drones, the prime minister said, but state security services reported that the drones were meant for him. "The $\textbf{explosion}_{\textbf{trig}}$ was caused by two machine $\text{guns}_{\text{arg1}}$," Maduro said, adding that there were no injuries. ⋯
Fake	⋯Two drones armed with explosives $\textbf{detonated}_{\textbf{trig}}$ near $\text{PuntoDeCorte}_{\text{arg2}}$, where the Venezuelan Foreign Minister, Jorge Rodríguez, was performing, and near the stage where he was giving a speech. ⋯

Table 4.1: An example of cross-document event cluster from IED dataset, where **trig**, arg1 and arg2 represent the trigger, *ExplosiveDevice* argument and *Place* argument respectively. The four events from four documents all refer to the explosion attack on Nicolas Marduro. The two real news articles complement each other by providing different aspects of the event (*ExplosiveDevice* argument in the first news and *Place* argument in the second news), while the two fake news articles contradict the real news with different details (i.e., different *ExplosiveDevice* and *Place* arguments).

4.1.3 KG representation

We use BERT [53] to initialize the node and edge embeddings in the KG. For a document node, we use BERT to encode the entire document and take the embeddings of `[CLS]` tokens. Similarly, for an entity node, we encode its canonical mention. For an event node, we encode the sentence where the event trigger occurs. For an event cluster node, we take the average of the embeddings of all events in the cluster. For a relation edge or an event argument role edge, we encode the linearized representation of the relation tuple. For example, the *Leadership* relation between "Nicolas Maduro" and "Venezuelan" is represented by "*Nicolas Maduro, Leadership, Venezuelan*" , and "guns" as the *ExplosiveDevice* argument of the *DetonateExplode* event is represented by "*DetonateExplode, ExplosiveDevice, guns*".

4.2 KNOWLEDGE GRAPH ENCODER

4.2.1 Heterogeneous GNN

Given the heterogeneous nature of the cross-document KG, we adopt a heterogeneous GNN to encode the KG.

Formally, let $\mathcal{G}$ denote KG and $\mathcal{V}$ denote the nodes in $\mathcal{G}$. We use $\mathcal{R}$ to denote the 10 types of edges as discussed in the previous section, and for each edge type $r \in \mathcal{R}$, we use $\mathcal{G}_r$ to denote the subgraph of $\mathcal{G}$ that only contains edges of type r. At the l-th layer, the inputs are output features produced by the previous layer denoted as $\mathbf{h}_i^{(l-1)}, i \in \mathcal{V}$. For each edge type $r \in \mathcal{R}$, we apply a separate GNN to encode $\mathcal{G}_r$ and produce a set of features denoted as $\mathbf{h}_{i,r}^{(l)}$. Then, we aggregate the outputs for all edge types into the final output as follows:

$$\mathbf{h}_i^{(l)} = \sum_{r \in \mathcal{R}} \mathbf{h}_{i,r}^{(l)} / |\mathcal{R}| \tag{4.1}$$

For document-to-entity edges, document-to-event edges, and edges connecting event nodes to event cluster nodes, we use standard graph attention network (GAT). For relation edges and event argument edges, we apply edge-aware GAT to leverage the edge features. Here, the edge features refer to the BERT embeddings of text descriptions such as "*Nicolas Maduro, Leadership, Venezuelan*" or "*DetonateExplode, ExplosiveDevice, guns*" as described in Section 4.1. The remainder of Section 4.2 presents details of GAT and edge-aware GAT, i.e., how to produce $\mathbf{h}_{i,r}^{(l)}$ based on $\mathbf{h}_i^{(l-1)}$.

4.2.2 Graph attention network

For each given node, GAT aggregates the node features of its neighbors via attention mechanism [72]. For a given edge type $r \in \mathcal{R}$, let $\mathcal{N}_{i,r}$ denote the neighbors of node i in $\mathcal{G}_r$. At the l-th layer, the attention weights α_{ij} are calculated as follows:

$$e_{ij} = \text{LeakyReLU}\left(\mathbf{a}^\top \left[\mathbf{W}\mathbf{h}_i^{(l-1)} \| \mathbf{W}\mathbf{h}_j^{(l-1)}\right]\right) \tag{4.2}$$

$$\alpha_{ij} = \text{softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}_{i,r}} \exp(e_{ik})} \tag{4.3}$$

where $\mathbf{a}$ and $\mathbf{W}$ are trainable parameters, and $\|$ denotes the feature concatenation. The output features $\mathbf{h}_{i,r}^{(l)}$ for node i in $\mathcal{G}_r$ are calculated as follows:

$$\mathbf{h}_{i,r}^{(l)} = \sum_{j \in \mathcal{N}_{i,r}} \alpha_{ij} \mathbf{W}\mathbf{h}_j^{(l-1)} \tag{4.4}$$

4.2.3 Edge-aware graph attention network

Edge-aware GAT is an extension of GAT that considers edge features in addition to node features [73, 74]. Let $\mathbf{r}_{ij}$ denote the features of the edge between nodes i and j. For a given edge type $r \in \mathcal{R}$, at the l-th layer, the attention weights α_{ij} are computed as follows:

$$\mathbf{r}'_{ij} = \mathbf{W}^r \left[\mathbf{h}_i^{(l-1)} \| \mathbf{h}_j^{(l-1)} \| \mathbf{r}_{ij}\right] \tag{4.5}$$

$$\alpha_{ij} = \mathrm{softmax}_j \left((\mathbf{W}^Q \mathbf{h}_i^{(l-1)})(\mathbf{W}^K \mathbf{r}'_{ij})^\top \right) \tag{4.6}$$

where $\mathbf{W}^r$, $\mathbf{W}^Q$ and $\mathbf{W}^K$ are trainable parameters. The output features $\mathbf{h}_{i,r}^{(l)}$ for node i in $\mathcal{G}_r$ are computed as follows:

$$\mathbf{h}_{i,r}^{(l)} = \sum_{j \in \mathcal{N}_{i,r}} \alpha_{ij} \mathbf{W}^V \mathbf{r}'_{ij} \tag{4.7}$$

where $\mathbf{W}^V$ is a learnable matrix.

4.3 MISINFORMATION DETECTOR

Using the previously described graph encoder, we are able to obtain representations of the document and event nodes. We conduct document-level detection using the document node representations, and event-level detection using the event node representations. We separately train two detectors for these two levels of tasks.

However, these two tasks are not mutually independent. Intuitively, document-level detection can benefit from the results of event-level detection, because the presence of a large number of false events indicates that the document is more likely to be fake. Therefore, we feed the results produced by a well-trained event-level detector into each layer of the document-level detector. Let $\mathbf{e}_i$ denote the representations of node i produced by the event-level detector. At the l-th layer of the document-level detector, instead of using the output features of the previous layer $\mathbf{h}_i^{(l-1)}$ as input features, we use a linear projection of the concatenation of $\mathbf{e}_i$ and $\mathbf{h}_i^{(l-1)}$ calculated as follows:

$$\mathbf{W}_{\text{proj}}^{(l)} \left[\mathbf{e}_i \| \mathbf{h}_i^{(l-1)}\right] \tag{4.8}$$

where $\mathbf{W}_{\text{proj}}^{(l)}$ is a learnable matrix.

CHAPTER 5: DATASET CONSTRUCTION

5.1 METHOD

Currently, there are no existing resources for cross-document misinformation detection. We propose to construct datasets based on real news datasets with clustering information. For each cluster, we randomly sample 50% real news and replace them with manipulated fake news. Figure 5.1 shows an overview of the fake news generation process.

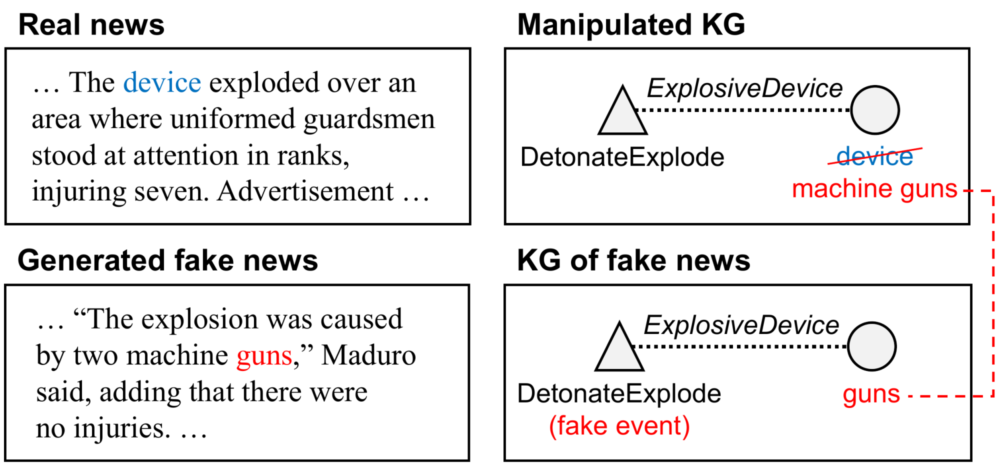

Figure 5.1: An overview of the fake news generation process. Based on the real news and its IE output, we select a high-frequency "*DetonateExplode*" event and replace its argument entity "device" with "machine guns". We then generate the fake news from the manipulated KG. In the KG of generated fake news, the manipulated entity "guns" is an argument of the "*DetonateExplode*" event, so we consider the event as fake.

Following [6], we train a KG-to-text generator from the real news in our datasets, and generate fake news from manipulated KGs. The input to the KG-to-text generator is the linearized representation of the IE-based KG. Since generating the entire document is very challenging, we fine-tune a sentence-level KG-to-text generator from BART [75]. The generator takes the linearized KG and the previous sentence as input and generates the next sentence. Here, the KG only contains information presented in the sentence rather than in the entire document. During inference, the generator generates the entire document sentence-by-sentence in an autoregressive manner.

The linearized representation contains both entity-to-entity relations and events with arguments. Formally, we denote an entity-to-entity relation as (h, r, t), where h is the head entity, t is the tail entity, and r is relation type. The linearized representation would be "$<h$, r, $t>$". For example, the *Leadership* relation between "Nicolas Maduro" and "Venezuelan" is represented by "*<Nicolas Maduro, Leadership, Venezuelan >*". An event can be denoted as $(e, \{(r_1, a_1), ..., (r_n, a_n)\})$ where e is the event type and (r_i, a_i) indicates entity a_i is the argument of role r_i. The event is then linearized into "*[$e \parallel r_1 = a_1, \cdots, r_n = a_n$]*". For example, an *DetonateExplode* event with "drone" as *ExplosiveDevice* argument and "flat" as *Place* argument is represented by "*[DetonateExplode ‖ ExplosiveDevice = drone, Place = flat]*". We represent the entire KG in graph by concatenating the text representations of all relations and events. For example, the KG in Figure 5.2 is: *<juvenile, Physical LocatedNear, home><boys, Physical Resident, home>[Justice ArrestJailDetain Unspecified — Detainee = juvenile]*

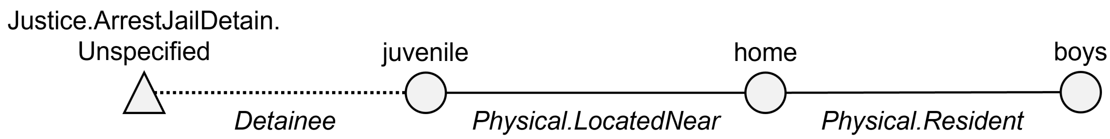

Figure 5.2: An example KG sampled from the training set.

During manipulation, we record the manipulation operations and use a heuristic rule to obtain supervision for event-level detection. Concretely, suppose we replace entity x with $\bar{x}$ during manipulation. We extract the KG from generated fake document. For each event e from the resulted KG, we match its argument entities agianst the entity $\bar{x}$. If any of its argument entities matches $\bar{x}$, we believe that event e contains manipulated information and consider it as fake. Regarding entity matching, we consider two entities as matched as long as one of the two entities is the substring of the other. Though this approach may introduce some noisy labels, we observe that it is generally correct. An example is shown in Figure 5.1, where we replace entity "device" into "machine guns". In the generated fake news, the entity "guns" is the *ExplosiveDevice* argument for event *DetonateExplode*. Since "guns" is a substring of "machine guns", "guns" matches the manipulated entity "machine guns" and the *DetonateExplode* is thus considered as fake.

The main differences between [6]'s method and ours in terms of manipulating KG are: (1) we only conduct entity swapping, and do not adopt other types of manipulation including adding relations or events and subgraph replacement; (2) since we focus on events, we select entities to be replaced that are arguments of high-frequency events, instead of based on entity node degree; (3) we select entities from other documents in the same document cluster to replace the original entities, so that the entities before and after replacement are more similar.

As for implementation details, the KG-to-text generator is based on `bart-large` model containing 24 layers, 1024 hidden dimensions, 16 heads, and 406M parameters. We fine-tune the model on the three datasets respectively. We train the model on a Tesla P100 GPU using the batch size of 1024 tokens, the gradient accumulation step of 16, the learning rate of 3×10^{-5}, the warmup steps of 500 steps, and the total training steps of 12000.

5.2 DATASETS

We constructed three new benchmark datasets based on three datasets that naturally have clusters of topically related documents. **IED** is a complex event corpus, where each complex event refers to a real-world story (e.g., Boston bombing) and is described by multiple documents [76]. Therefore, a complex event can be considered as a document cluster. **TL17** and **Crisis** are two timeline summarization datasets containing multiple news timelines. Each timeline contains multiple documents describing an evolving long-term event such as Influenza H1N1 and Egypt Revolution [77, 78], and thus can be regarded as a document cluster. The detailed statistics of the original datasets are shown in Table 5.1.

	# Cluster	**# Doc**	**# Doc per cluster**
IED	433	7403	17
TL17	17	4650	273
Crisis	4	20463	5116

Table 5.1: Statistics of the original datasets.

		# Cluster	**# Doc**	**# Fake event per doc (%)**
IED	**Train**	422	3865	3.99 (9.91%)
	Dev	140	1297	3.66 (9.14%)
	Test	140	1262	3.68 (9.51%)
TL17	**Train**	276	2610	2.97 (12.70%)
	Dev	92	879	2.69 (12.31%)
	Test	92	892	2.85 (12.13%)
Crisis	**Train**	1413	13337	4.54 (13.95%)
	Dev	177	1648	4.21 (13.29%)
	Test	177	1701	4.38 (13.80%)

Table 5.2: Statistics of the resulting datasets.

However, documents within the same cluster may not be closely related as the story described by a cluster can span up to three years. To obtain smaller and more closely related clusters, we split each timeline into smaller clusters of approximately size 10 based on publication dates. For IED, we randomly split the clusters due to the lack of publication dates. Then, we employ the methods described in Section 5.1 to generate fake documents. The statistics of the constructed datasets are in Table 5.2.

Real News

Amia bombing: Argentina and Iran agree truth commission Published duration 28 January 2013. image caption The two foreign ministers signed the agreement in Addis Ababa on Sunday. Argentina and Iran are to jointly set up a commission to investigate the 1994 bombing of the Israeli-Argentine Mutual Association (Amia) Jewish community centre in Buenos Aires. The commission will be made up of five independent judges, none of whom will be from either Argentina or Iran. Argentine courts have blamed Iran for the attack, which killed 85 people. Iran has always denied any involvement. Israel's foreign ministry said it was "surprised" by news of the commission. Spokesman Yigal Palmor told the AFP news agency that it was waiting "to receive full details" from Argentina. Amia and another key Argentina Jewish organisation were meanwhile reported to be vehemently opposed to the move. High-profile suspect. Iran agreed last July to co-operate with Argentina in the investigation, which it said "was going down the wrong way". Such negotiations have alarmed Israel's government and Argentina's sizeable Jewish community, who fear Argentina is weakening in its resolve to put suspects on trial. "We warned the Argentines from the start that the Iranians would try to set a trap for them and that they should beware," Mr Palmor was quoted by AFP as saying on Monday. The news agency also quoted a joint statement by Amia and the Delegation of Israelite Argentine Associations as saying that the new move would "imply a decline in our sovereignty". image caption The seven-storey Amia building was destroyed in the attack on 18 July 1994. "To ignore everything that Argentine justice has done and to replace it with a commission that, in the best of cases, will issue, without any defined deadline, a 'recommendation' to the parties constitutes, without doubt, a reversal in the common objective of obtaining justice," the statement said. However, Argentine President Cristina Fernandez de Kirchner called the agreement "historic". "It guarantees the right to due process of law, a fundamental principle of international criminal law," Ms Fernandez said. She said Argentine Foreign Minister Hector Timerman and his Iranian counterpart had signed a memorandum of understanding on the sidelines of an African Union summit in Addis Ababa, in Ethiopia. It still has to be ratified by the parliaments of both countries. The commission would "analyse all the documentation presented to date by the judicial authorities of Argentina and Iran", Ms Fernandez wrote on her Twitter account. According to Mr Timerman, the agreement will make it possible for Argentine legal officials to question Iranian suspects in Tehran. Among the suspects named by Argentine prosecutors when they made their case in 2007 is Iran's current Defence Minister, Gen Ahmed Vahidi. At the time of the attack, Gen Vahidi was the commander of a special unit of Iran's Revolutionary Guards. Argentine prosecutors said Iran planned and financed the attack, and that the Lebanese militant group Hezbollah carried it out.

Generated Fake News

Israel's foreign ministry said in a statement that it was “shocked and saddened” by the news. The foreign ministers met in Addis Ababa on Tuesday to discuss the two countries' efforts to promote peace in the Horn of Africa. Argentina and Ethiopia have been close allies in the fight against terrorism since the 1994 bombing of a Jewish community centre in Buenos Aires that left 85 people dead. The commission's judges, none of whom are from Argentina or Iran, were appointed by President Cristina Fernández de Kirchner, who has close ties with Iran. Argentine courts have blamed Kenya for the attack, which killed 85 people, but it has been blamed on Iran. Another bombing in Bolivia, in 1994, was also blamed on Iran, but that has not been proven. Argentina's main Jewish organisation, the Argentina-based Jewish Organisation of Argentina, said it was convinced that Iran was behind the attack. The Jewish community in Argentina has long accused the government of Israel of not doing enough to bring the suspects to trial, and of obstructing efforts to bring them to justice. In an interview with AFP, Palmor said: “We've been asking the Iranians and the Argentines for a long time to do something. We're still waiting for the Iranians to acknowledge that the Amia building was attacked and to take responsibility for it.” Argentine President Cristina Fernandez de Kirchner, who was in office from 2007 to 2015, has accused Iran of being behind the attack, but Tehran has repeatedly denied any involvement. Argentine Foreign Minister Hector Timerman, on his way to an African Union summit in Addis Ababa, Ethiopia, spoke with his Iranian counterpart, Mohammad Javad Zarif, by phone on Monday, according to a statement released by the Iranian Foreign Ministry. Timerman “expressed Argentina's solidarity with the African Union,” the statement said. Iran's Foreign Minister said Iran “condemns in the strongest terms” the decision by Argentine prosecutors to issue an arrest warrant for former intelligence chief Ahmad Vahidi in connection with a case that Argentine prosecutors have been investigating for years. “We call on the authorities of Iran and Argentina to immediately comply with the decision of the judicial authorities of both countries,” Mohammad Javad Zarif said in a statement. Argentine officials met with the Iranian suspects in Tehran last month, and Argentine officials said they had “sufficient evidence” to proceed with the extradition. Both countries' parliaments have to approve the extradition, which is expected to happen in the coming weeks. Vahidi, the commander of an elite unit of Iran's Revolutionary Guards known as the Quds Force, is believed to have masterminded the attack, which Vahidi said was carried out with the help of a “cousin,” Maj. Gen. Mohammad Ali Jafari. Argentine prosecutors have said that Iran paid more than $400 million for the attack, a claim that Iran has repeatedly denied.

KG manipulation

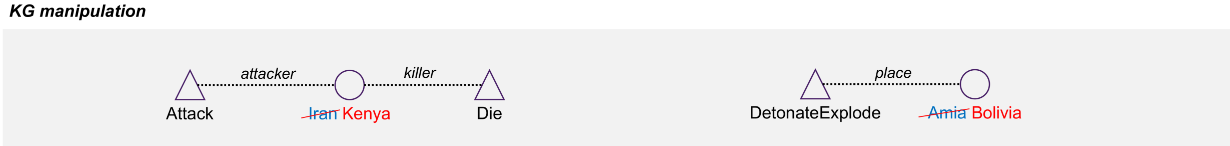

Figure 5.3: An example of generated fake news, including the original real news, manipulated KG, and generated fake news. Real and fake information are marked in blue and red respectively. To save space, we only show some parts of the KG that are manipulated.

We present two examples of generated fake news in Figure 5.3 and 5.4, including the original real news, manipulated KG, and generated fake news. The generated fake news conveys the manipulated misinformation and meanwhile is stylistically similar to real news. To further evaluate the quality of the generated fake news, we conducted a Turing Test by 13 human readers as in [6]. We randomly select 100 documents from the IED dataset, half real and half fake, and ask the human readers to assess the authenticity for each document. The overall accuracy achieved by human readers is 66.88%, with 77.44% accuracy on real documents but only 56.32% accuracy on fake documents. This shows that it is difficult for human readers to detect the generated fake news.

Real News

An IRA member who was part of the group behind the Birmingham pub bombings has apologised. Michael Christopher Hayes – a self-confessed bomb maker – said he was sorry for the killing of innocent people in the 1974 blasts. Speaking to the BBC, Mr Hayes said he hoped his apology on behalf of all active republicans will help grieving relatives find "closure". He said: "My apologies and my heartfelt sympathies to all of you, for the terrible, tragic loss that you've been put through. And all these years you've been trying to find closure. I hope at last God will be merciful and bring you closure. "And I apologise not only for myself ... I apologise for all active republicans who had no intention of hurting anybody and sympathise with you." An eight-minute delay before police were warned of the bombs' location led to the death of 21 people and the injury of 182 others, when they exploded in a pair of city centre pubs. Mr Hayes said the bombs had not been intended to kill people. On the evening of the 21 November 1974, a man with an Irish accent called the Birmingham Post and Evening Mail newspapers to say two bombs were planted in the town centre. He finished the call by giving the official code used by the Provisional IRA to authenticate a warning call and allow civilians to be evacuated. Former IRA officials have since said there was an unintentional delay in issuing the warning. Mr Hayes said the IRA unit in Birmingham had been shocked by the death toll. "That wasn't meant. It wouldn't have been done if that was the case," he said. The wreckage left at the Mulberry Bush pub in Birmingham after a bomb exploded on 21 November 1974 (PA) The Birmingham pub bombings caused the worst single losses of life in the Troubles. Six men were wrongfully convicted for the blasts but no-one has ever been brought to justice. Mr Hayes, 69, who now lives in south Dublin, said he personally defused a third bomb on Birmingham's Hagley Road after he became aware of the death toll in the first two blasts. He refused to say who planted the bombs in the Mulberry Bush and the Tavern in the Town but he said he wanted to speak out to give "the point of view of a participant". But relatives of those who were killed said the apology was "gutless and spineless". Mr Hayes' apology came as an inquest into the bombings was reopened by a coroner stating "a wealth of evidence" had not yet been heard. West Midland Police said their investigation remained open but that no further comments on the suspects' identities will be provided until the proceedings have concluded. Victims' families have repeatedly asked for the names of the suspects to be disclosed. Mr Hayes was arrested and questioned by West Midlands Police about the bombings in 1974 but he was released. He said he had been a member of the IRA for more than 30 years in both Ireland and England and had participated in the group's activities in Birmingham as "an active volunteer". AFP via Getty UK news: Benita Stipp (centre) and Mimi Ferguson (left) react as students at Norwich School receive their A-Level results. PA UK news: A train derailment near Stonehaven has left three people dead.

Generated Fake News

A man with an Irish accent made a telephone call to the Birmingham Post and Evening Mail, claiming to be from the town centre and claiming to have information about the whereabouts of an Irish Republican Army (IRA) bomb which had been planted in the town. The caller claimed that the bomb had been intended to cause an explosion which would have killed all the "scrubbers". A niece was killed and nine others were injured when they were caught up in the bombs which exploded in the two pubs. The Birmingham pub bombings were believed to have been carried out by a member of the IRA's Provisional IRA group. Image copyright Getty Images Image caption Twenty-one people were killed. Only two men have been convicted of the bombing, but no - one has been brought to justice. It has been suggested that IRA officials may have been involved in the planning of the atrocity, but this has been denied. It has been suggested that the IRA unit responsible for the Birmingham killings may have been involved in the planning and execution. The relatives of those killed have been campaigning for a full public inquiry into the killings for more than 20 years. Image copyright Pacemaker Image caption The death toll from the County Antrim pub bombings was the highest in a single day in the Troubles. In a statement, he said: "It is not known who planted the bombs in the Mulberry Bush and the Tavern in the Town but it is clear who placed the bombs at the Tavern. "It is known that Mr Hayes, who was living in south Dublin at the time, defused a bomb in Birmingham's Hagley Road the day before he carried out the Mulberry Bush bombing but it is not known if he was involved in the Tavern in the Town explosion." A number of civilians were evacuated from the scene. Victims' families have expressed their shock and grief. Image copyright Getty Images Image caption The attack has been described as a "heinous act of terrorism" Image copyright Getty Images Image caption Corporation Street in Birmingham was the scene of one of the Birmingham bomb attacks in 1974. In County Antrim, Benita Stipp and Mimi Ferguson, who were students at the time of the bombing, were the two people who died in the blast, as were two other students. The attack was condemned by the Irish Republican Army (IRA) and Sinn Féin. During his time at Elvington Airfield, near York, York, in the UK, Carl Thomas had been involved in the construction of a radar station, which was to be used by the Royal Ulster Constabulary (RUC) and the British Army. The pair, who were born and raised in the UK but grew up in the mountains of the north-west of England, were home-schooled students who had studied in Belgium before going to the Department for Education in London, but after finishing their studies in the Albanian mountains, the pair returned to their home country to work as teaching candidates in the Department of Education in the country, and to work with their teachers. After the attack, James Boyle was arrested by West Midlands Police but he was released after being questioned by the police and released on bail, and was due to appear before a judge on Monday. Police believe he was a member of the IRA in Ireland and in England, and that he was based in Birmingham, but that he had been active in the IRA both in England and in Ireland. 2 of the 3 men convicted of the bombings have been released on appeal. They had their convictions quashed on appeal after serving 16 years in prison and their trial was heard by the Court of Appeal. They were released on compassionate grounds after a judge ruled that they had suffered a miscarriage of justice.

KG manipulation

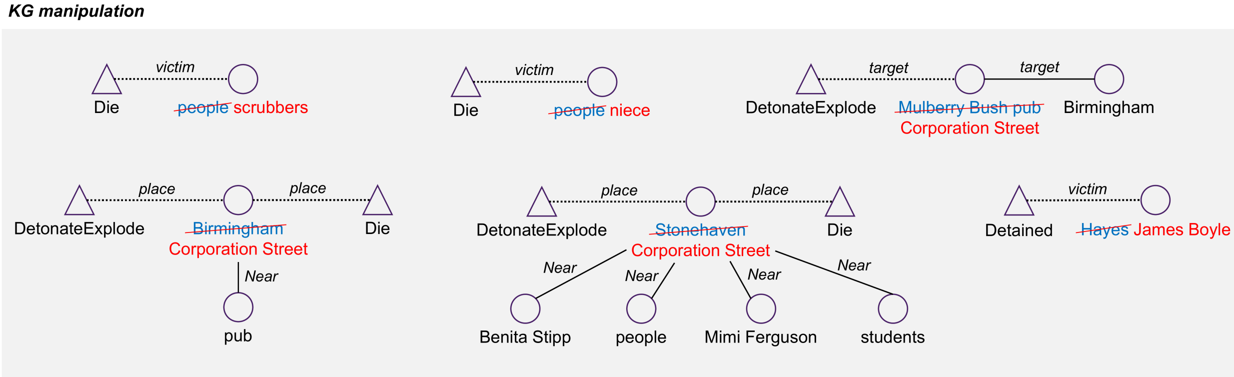

Figure 5.4: An example of generated fake news, including the original real news, manipulated KG, and generated fake news. Real and fake information are marked in blue and red respectively. To save space, we only show some parts of the KG that are manipulated.

CHAPTER 6: RESULTS AND DISCUSSION

6.1 EXPERIMENTAL SETTINGS

For our proposed method, we use a 4-layered heterogeneous GNN, where each GAT layer contains 8 heads. To initialize the node and edge embeddings, we use `bert-base-uncased` model with the feature dimension of 768. Our model contains 233M parameters. To train our method, we use a batch size of 16, and search the learning rate from $\{10^{-3}, 10^{-4}, 10^{-5}\}$ and the number of layers within $\{2, 4, 8\}$. Our best-found hyperparameters are a learning rate of 10^{-5} and a number of layers of 4. We train our model with Adam optimizer until convergence. To reduce computation cost, we freeze BERT's parameters. The training process takes approximately 6 hours on a Tesla P100 GPU.

For comparison, we compare our method against two baselines on the document-level detection task: **HDSF** that models inter-sentence dependency tree [21], and **GROVER** [1], a Transformer-based detector.

Hierarchical **D**iscourse-level **S**tructure for **F**ake (HDSF) aims to model the discourse-level structure within a document by constructing a dependency tree, where each node represents a sentence, and a parent-child link indicates that the child sentence semantically relies on the parent sentence. The model first encodes each sentence separately with an LSTM model, and then propagates the sentence features along the parent-child links to obtain a structurally-aware document representation. We reproduce HDSF on our constructed datasets using the implementation at `https://github.com/hamidkarimi/HDSF/`. We train the model on our datasets using their default hyper-parameters.

GROVER is based on the same architecture of GPT-2 [20]. Based on GPT-2 that is already pre-trained on large general-domain corpus, GROVER is further pre-trained on news-domain documents, and then fine-tuned to perform fake news detection task. We use the implementation at `https://github.com/rowanz/grover` and experiment with two settings, large setting and mega setting. The GROVER-mega model has 48 layers and 1.5 billion parameters, on par with GPT2, while GROVER-large has 24 layers and 355 million parameters on par with BERT-large. Since fine-tuning the GROVER model is computationally expensive, we use GROVER in the zero-shot setting.

For the event-level detection task, since there are no existing methods, we compare our method against three baselines, **random guessing**, **logistic regression**, and **BERT**.

In random guessing, for each event, we randomly draw a value from a uniform distribution between $[0, 1]$ as the probability that the event is false. Since it utilizes no information of the input event, it should serve as a lower bound for any predictive methods. In logistic regression, we use the following features: event type (represented by one-hot feature), number of arguments, and the size of the event cluster that the given event belongs to. This results in a 46-dimensional feature vector, where 43 of them are one-hot features representing the event type. The features are normalized on the training set. We use the implementation of logistic regression and default parameters provided by `sklearn`.

In the BERT baseline, we use the same BERT-based event features as our method, and replace the 4-layer GNN in our model with a feed-forward network. We take the sentence where the event trigger occurs as BERT input, since this sentence usually describes the event. Such represnetation is sufficient since the event arguments usually occur in the same sentence, and the sentence contains more comprehensive descriptions compared to only the argument entity. We use the same hyper-parameters to train the model. The hyper-parameter for training the model is searched on the development set separately for each dataset. The learning rate is searched from $\{10^{-5}, 5 \times 10^{-5}, 2 \times 10^{-4}, 5 \times 10^{-4}\}$ and the number of epochs is searched from $\{40, 60, 120, 240\}$. Eventually, the optimal set of parameters for both IED and TL17 is a learning rate of 5×10^{-4} and the epoch number of 120, while the optimal set of parameters for Crisis is a learning rate of 5×10^{-4} and the epoch number of 40.

Regarding evaluation, we use F1 to evaluate document-level detection. Considering the label imbalance of event-level detection, we use F1 and the area under the ROC curve (AUC) to evaluate event-level detection. For the F1 metric, we select the optimal threshold on the validation set.

6.2 DOCUMENT-LEVEL DETECTION

Table 6.1 shows the results of document-level detection. Compared to the HDSF and GROVER baselines, our method yields consistent improvements on all three datasets and significantly outperforms the baselines that judge the authenticity for each document in isolation. Regarding the two baselines, we see that GROVER consistently outperforms HDSF baseline, since GROVER is based on the more powerful Transformer structure and benefits from the pre-training process over large news corpus. We also observe that GROVER-mega performs significantly better than GROVER-large, which indicates that larger parameter size contributes to better performance.

To further analyze the effectiveness of each component, we conducted an ablation study. The results are presented in Table 6.2. In the ablation study, we examine two key designs in our system: cross-document analysis and feeding event-level detection results to the document-level detector. For the former, since our cross-document analysis mainly relies on

	IED	TL17	Crisis
HDSF	78.42	80.62	82.14
GROVER-large	79.06	79.40	86.84
GROVER-mega	82.90	90.00	87.13
Ours	**86.76**	**90.21**	**93.89**

Table 6.1: F1 results (in %) of document-level detection. We report the F1 scores of HDSF [21], GROVER of two settings [1], and our proposed method.

leveraging cross-document event coreference, we conduct an ablation study by dropping all edges connecting event nodes and event coreference cluster nodes. In this way, the cross-document KG degenerates into isolated components, each representing a single document. For the latter, we conducts ablation study by feeding no additional features to the document-level detector. However, incorporating features naturally introduces extra parameters to the model. To ensure that any performance improvements are not simply due to the extra parameters required for incorporating features, we also conduct an experiment where we feed random features to the model instead of the event-level detection results. Thus, the model has the same number of paramters compared to our system but does not utilize event-level detection results.

Cross-document event coreference	Event-level detection results	IED	TL17	Crisis
✗	✗	80.59	86.55	93.64
✓	✗	84.57	88.99	93.67
✓	Random	83.63	84.86	92.18
✓	✓	**86.76**	**90.21**	**93.89**

Table 6.2: F1 results (in %) of ablation study over document-level detection. We analyze the use of cross-document event coreference resolution and event-level detection results. We further experiment with random features for event-level detection results. Results of our full method are presented in the last row.

From the ablation results, we see that both of our two key designs contribute to improvements in document-level detection, particularly on the two smaller datasets of IED and TL17. More specifically, we have the following two findings:

1. We remove the edges between event nodes and event center nodes to analyze the impact of cross-document event coreference resolution, and find that such information significantly improves the performance on IED and TL17. Without such information, our method performs on par with or slightly worse than the best baselines. We also train our detector with smaller clusters on TL17 and get worse performance (84.53%

and 87.37% on clusters with size 1 and 2 respectively), which verifies that our model benefits from more cross-document information. The benefit of cross-document event coreference resolution is less significant on the large-scale Crisis dataset containing 1.7k documents. This may imply that cross-document misinformation detection is more useful for emerging new events where large-scale training data is not available.

2. Using the event-level detection results consistently improves the performance by 1-3 points on all datasets. Since the projection modules introduce additional parameters, we further train a detector utilizing random features and find that using random features reduces the performance. This verifies that the improvement is brought by utilizing the knowledge learnt by the event-level detector rather than additional parameters. The underlying intuition is that by integrating the event-level detection results of all events, the document-level detector is able to achieve better judgments for the entire document.

6.3 EVENT-LEVEL DETECTION

We track the manipulation operations during the dataset construction process, which allows us to obtain supervision for event-level detection. The results are shown in Table 6.3. We compare our method with random guessing, logistic regression with hand-crafted event features, and BERT.

	IED		**TL17**		**Crisis**	
	F1	AUC	F1	AUC	F1	AUC
Random	16.31	50.44	19.44	49.65	21.70	50.41
LR	31.26	77.87	29.14	68.19	31.67	68.17
BERT	26.43	71.12	31.95	71.42	33.89	71.86
Ours	44.86	**88.46**	**41.56**	**82.59**	**48.48**	**85.60**
Ours$^{\text{(ABLATION)}}$	**45.00**	**88.54**	**41.66**	82.28	47.78	85.17

Table 6.3: Results (in %) of event-level detection. We report the F1 and AUC scores of random guessing (Random), logistic regression (LR), BERT, and our method. We further conduct an ablation study and report the results of our method without cross-document event coreference information, denoted as Ours$^{\text{(ABLATION)}}$.

We find that random guessing performs the worst. This is expected because random guessing do not have access to any informative features of the input event. Logistic regression and BERT perform better than random guessing and achieve satisfactory performance. Surprisingly, though BERT model has much more parameters than LR and has been pre-trained on a large corpus, it performs only on par with LR. This indicates that the event-level detection is a very challenging task even for pre-trained language models. Our method significantly outperforms all baselines by a large margin, which demonstrates the effectiveness of our graph-based approach.

As in document-level detection, we also conduct an ablation study on the use of cross-document event coreference resolution by removing edges between event nodes and event cluster nodes. The results are denoted as Ours$^{\text{(ABLATION)}}$ in Table 6.3. We find that such information brings slight improvements in the AUC metric.

6.4 QUALITATIVE ANALYSIS

To demonstrate the benefits of using cross-document event coreference resolution, we show an example in Figure 6.1, with 4 documents from the same cluster. The documents and corresponding KGs describe a Taliban attack on a university in Afghanistan. The two real news stories agree that the attack was conducted by Taliban. However, each of the two fake news contains false information about the attacker that contradicts all other news. Fake news 1 claims that the attack is performed by a student, while fake news 2 claims that the event is performed by the insurgent. Both information contradicts the true information presented in both real news 1 and real news 2. We report in the tables in Figure 6.1 the detection predictions of our model denoted by "Ours". Additionally, we perform ablation study by dropping without cross-document event coreference resolution information, as in previous sections. The predictions of the ablated model are denoted by "Ours$^{\text{(ABLATION)}}$". We show that the use of cross-document event coreference resolution significantly improves both levels of detection, especially for detecting fake news 1. Compared to the ablated model, our model consistently assigns a higher probability of being fake to fake documents and events, while simultaneously assigning a lower probability of being fake to real documents and events.

We further analyze the remaining errors in our model by manually examining the failing cases. Figure 6.2 shows two representative cases where both document-level and event-level detectors fail to detect misinformation.

In the first example, the entity Abqaiq City is manipulated fake information, and it is supposed to be the *target* argument of the *DetonateExplode* argument. However, the entity is not captured by the IE system. Instead, the IE system identifies the entity "facility" as the textittarget argument, which does not involve any misinformation. As a result, our graph-based detector cannot detect any misinformation from the incomplete knowledge graph. The most natural solution is to employ more advanced IE systems with higher accuracy. Another potential solution is to use an OpenIE system [79] that is able to cover more event and entity

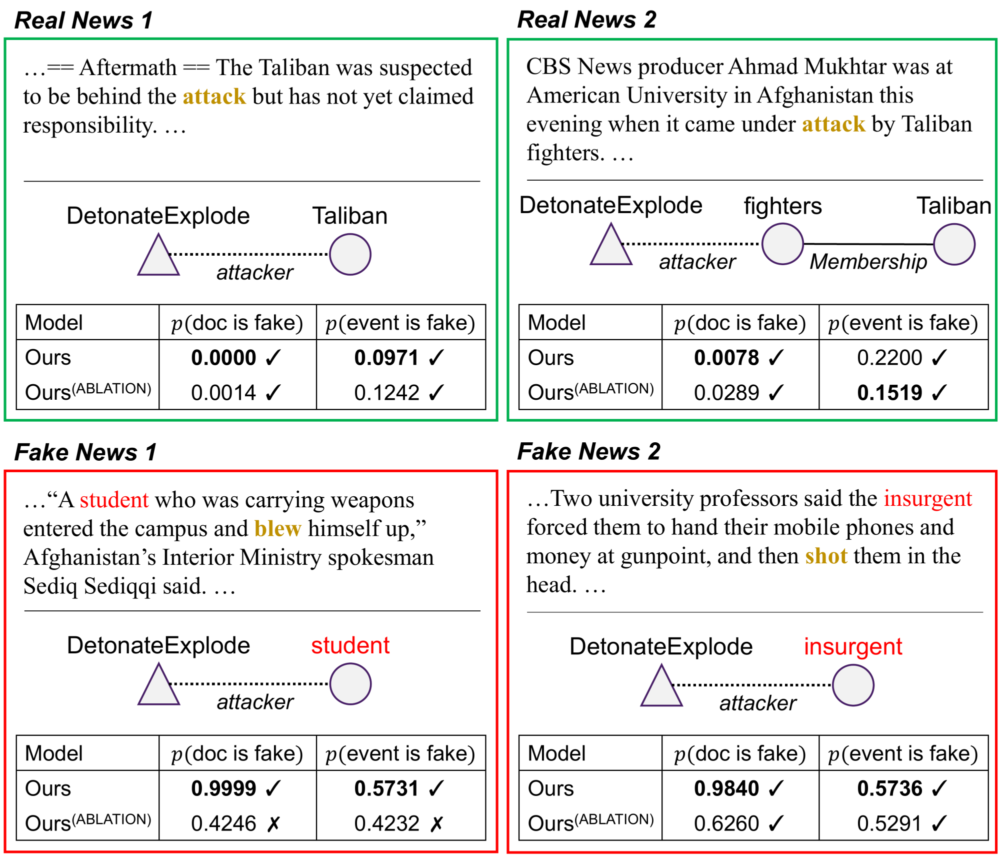

Figure 6.1: An example of four documents from the same cluster in the IED dataset accompanied by their respective KG. We also show the subgraphs of the KGs related to the four documents. Here, the four *DenonateExplode* events are in the same event coreference cluster. Event triggers are bolded and marked in gold, and fake information is marked in red. The two real news agree with each other, while each of the two fake news contains false information that contradicts all other news. The tables show the prediction results of both our model and the ablated model (denoted by "Ours$^{(\text{ABLATION})}$") for each document and event respectively. Better results are bolded.

types. However, due to the open nature of OpenIE systems, they may introduce noisy and unwanted information to the knowledge graph.

The second example is a more challenging case. In this case, the visit of Vajpayee to Mumbai is manipulated fake information. However, such visit is not mentioned by any other documents, and no coreference is detected for the *Transportation* event. This makes it challenging to either confirm or refute this information using cross-document via cross-document reasoning. Thus, our detector cannot identify this misinformation. This indicates that our detector operates on information redundancy across multiple documents. If these documents lack sufficient overlap, becomes ineffective. This introduces complexity in forming document clusters, as the quality of document clustering and the degree of overlapping

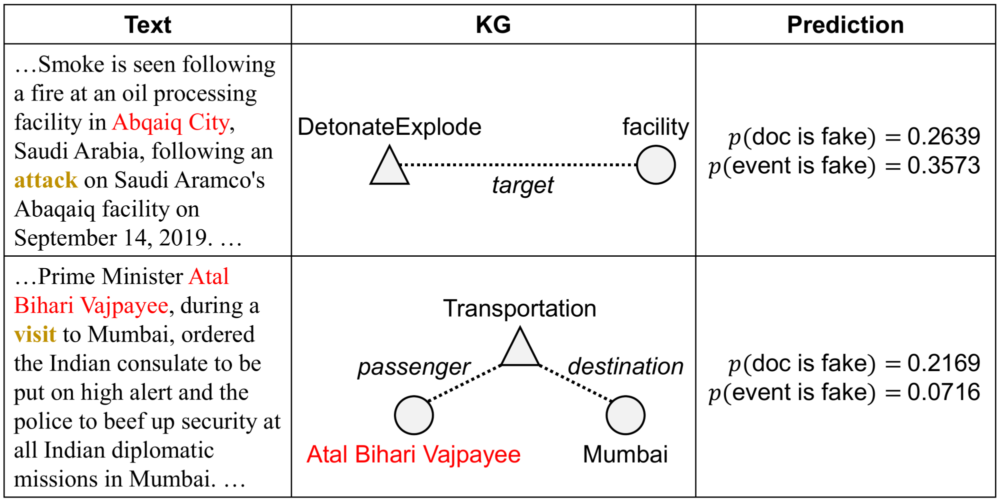

Figure 6.2: Two examples where our detector fails to detect the fake information. **Event triggers** are bolded and marked in gold, and fake information is marked in red. In the first example, the error of IE system is propagated into the detector. In the second example, the event containing fake information is not mentioned in any other document, making it difficult to either verify or disprove via cross-document reasoning.

information among clustered documents impact the final performance. A potential solution is to enable the detection system to actively search for event-related information, increasing the likelihood of accessing overlapping information.

6.5 DISCUSSION

There are several remaining challenges and limitations in our proposed methodology that warrant further exploration, particularly when considering practical fake news detection applications and the complex dynamics of the Internet environment.

First, some cross-document contradictions are difficult to capture by coreference resolution alone. In the example presented in Figure 1.1, discerning that the police are unlikely to both help and attack Boyland at the same time requires commonsense reasoning, a capability not yet incorporated into our current framework. Incorporating commonsense reasoning and other advanced techniques in the future could enhance our method's ability to detect more subtle inconsistencies and contradictions in real-world situations.

Second, an underlying assumption of our framework is that real news articles are consistent and complementary with each other, while fake news often contradicts each other. This assumption holds true for our constructed datasets, as we manipulate the KGs via random entity swapping. However, certain types of human-written fake news documents, such as conspiracy theories, tend to be closely related to each other and convey highly similar information because they share the same biases or aim to manipulate readers in the same way. This may limit the performance of our proposed system in real-world scenarios.

In the context of the constantly evolving Internet environment, where misinformation tactics and strategies are also continuously changing, it is essential to address these limitations in order to develop a more robust and adaptive fake news detection system. Addressing the limitations and challenges of our proposed methodology is a vital step towards enhancing its effectiveness in combating misinformation in today's rapidly changing online landscape. By exploring novel techniques for capturing subtle inconsistencies, incorporating commonsense reasoning, and understanding the complex relationships among various types of misinformation, we can refine our methodology to better tackle the ever-growing problem of misinformation in the digital world.

With this in mind, the primary objective of this work is to advance state-of-the-art research in the field of misinformation detection. However, in the process of developing our approach, we have also proposed a generation method for producing fake news, which inevitably raises ethical concerns. As with any work involving text generation, there is an inherent risk that our methods could be repurposed to create false information with the intention of misleading or manipulating readers. To address these concerns, we have taken some precautionary measures. We commit not to share the source code or checkpoints of our generator, thereby minimizing the potential for misuse. In order to balance reproducibility with these ethical concerns, we provide a high-level overview of our fake news generation approach, alongside crucial details that allow researchers to understand the underlying concepts without facilitating misuse. Furthermore, we emphasize the importance of transparency and accountability in our research and encourage the development of robust and ethical guidelines for future work in the field of misinformation detection and generation.

CHAPTER 7: CONCLUSIONS

In conclusion, our work presents a novel approach to cross-document misinformation detection, a crucial and timely research area that has significant implications for maintaining the accuracy and reliability of information on the Internet. We address this challenge by conducting the task at two levels: document level and the more fine-grained event level. To facilitate the development and evaluation of misinformation detection systems, we construct three new datasets containing clusters of topically related documents, incorporating both real news and AI-generated fake news.

Our proposed method, a graph-based cross-document detector, conducts reasoning over a cross-document knowledge graph that represents the content of a cluster of topically related documents. Additionally, our proposed method feeds the fine-grained event-level detection results to assist document-level detection. By leveraging these results, our method effectively enhances document-level detection capabilities. Our experimental results show that our proposed method performs better than existing techniques, highlighting the importance of using cross-document information and the benefits of combining event-level and document-level detection.

Our work lays the groundwork for further advancements in cross-document misinformation detection and highlights the potential of utilizing event-level information to improve document-level detection. To further boost the detection performance, a promising research direction is to develop more advanced knowledge graph encoders and detectors. Future research can also explore the incorporation of multi-media features including texts, images, audios and videos, which requires the construction of cross-document multi-modal knowledge graphs. Finally, a challenging but important direction is to construct a large-scale fake news detection corpus with human-written fake news containing document clusters and study our method in this scenario. By advancing the state-of-the-art in misinformation detection, we aim to contribute to the development of more reliable and resilient information systems that foster a well-informed society.

REFERENCES

[1] R. Zellers, A. Holtzman, H. Rashkin, Y. Bisk, A. Farhadi, F. Roesner, and Y. Choi, "Defending against neural fake news," in *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. B. Fox, and R. Garnett, Eds., 2019. [Online]. Available: https://proceedings.neurips.cc/paper/2019/hash/3e9f0fc9b2f89e043bc6233994dfcf76-Abstract.html pp. 9051–9062.

[2] Y. Choi, M. Choi, M. Kim, J. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," in *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*. Computer Vision Foundation / IEEE Computer Society, 2018. [Online]. Available: http://openaccess.thecvf.com/content_cvpr_2018/html/Choi_StarGAN_Unified_Generative_CVPR_2018_paper.html pp. 8789–8797.

[3] W. Y. Wang, ""liar, liar pants on fire": A new benchmark dataset for fake news detection," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Vancouver, Canada: Association for Computational Linguistics, July 2017. [Online]. Available: https://aclanthology.org/P17-2067 pp. 422–426.

[4] H. Karimi, P. Roy, S. Saba-Sadiya, and J. Tang, "Multi-source multi-class fake news detection," in *Proceedings of the 27th International Conference on Computational Linguistics*. Santa Fe, New Mexico, USA: Association for Computational Linguistics, Aug. 2018. [Online]. Available: https://aclanthology.org/C18-1131 pp. 1546–1557.

[5] D. Khattar, J. S. Goud, M. Gupta, and V. Varma, "MVAE: multimodal variational autoencoder for fake news detection," in *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*, L. Liu, R. W. White, A. Mantrach, F. Silvestri, J. J. McAuley, R. Baeza-Yates, and L. Zia, Eds. ACM, 2019. [Online]. Available: https://doi.org/10.1145/3308558.3313552 pp. 2915–2921.

[6] Y. Fung, C. Thomas, R. Gangi Reddy, S. Polisetty, H. Ji, S.-F. Chang, K. McKeown, M. Bansal, and A. Sil, "InfoSurgeon: Cross-media fine-grained information consistency checking for fake news detection," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Online: Association for Computational Linguistics, Aug. 2021. [Online]. Available: https://aclanthology.org/2021.acl-long.133 pp. 1683–1698.

[7] K. Shu, S. Wang, and H. Liu, "Beyond news contents: The role of social context for fake news detection," in *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, WSDM 2019, Melbourne, VIC, Australia, February 11-15, 2019*, J. S. Culpepper, A. Moffat, P. N. Bennett, and K. Lerman, Eds. ACM, 2019. [Online]. Available: https://doi.org/10.1145/3289600.3290994 pp. 312–320.

[8] V. Nguyen, K. Sugiyama, P. Nakov, and M. Kan, "FANG: leveraging social context for fake news detection using graph representation," in *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*, M. d'Aquin, S. Dietze, C. Hauff, E. Curry, and P. Cudré-Mauroux, Eds. ACM, 2020. [Online]. Available: https://doi.org/10.1145/3340531.3412046 pp. 1165–1174.

[9] Y. Lin, H. Ji, F. Huang, and L. Wu, "A joint neural model for information extraction with global features," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, D. Jurafsky, J. Chai, N. Schluter, and J. R. Tetreault, Eds. Association for Computational Linguistics, 2020. [Online]. Available: https://doi.org/10.18653/v1/2020.acl-main.713 pp. 7999–8009.

[10] T. M. Lai, H. Ji, T. Bui, Q. H. Tran, F. Dernoncourt, and W. Chang, "A context-dependent gated module for incorporating symbolic semantics into event coreference resolution," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021*, K. Toutanova, A. Rumshisky, L. Zettlemoyer, D. Hakkani-Tür, I. Beltagy, S. Bethard, R. Cotterell, T. Chakraborty, and Y. Zhou, Eds. Association for Computational Linguistics, 2021. [Online]. Available: https://doi.org/10.18653/v1/2021.naacl-main.274 pp. 3491–3499.

[11] H. Wen, Y. Lin, T. Lai, X. Pan, S. Li, X. Lin, B. Zhou, M. Li, H. Wang, H. Zhang, X. Yu, A. Dong, Z. Wang, Y. Fung, P. Mishra, Q. Lyu, D. Surís, B. Chen, S. W. Brown, M. Palmer, C. Callison-Burch, C. Vondrick, J. Han, D. Roth, S. Chang, and H. Ji, "RESIN: A dockerized schema-guided cross-document cross-lingual cross-media information extraction and event tracking system," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Demonstrations, NAACL-HLT 2021, Online, June 6-11, 2021*, A. Sil and X. V. Lin, Eds. Association for Computational Linguistics, 2021. [Online]. Available: https://doi.org/10.18653/v1/2021.naacl-demos.16 pp. 133–143.

[12] V. L. Rubin, N. Conroy, Y. Chen, and S. Cornwell, "Fake news or truth? using satirical cues to detect potentially misleading news," in *Proceedings of the second workshop on computational approaches to deception detection*, 2016, pp. 7–17.

[13] W. Y. Wang, “”liar, liar pants on fire”: A new benchmark dataset for fake news detection,” in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 2: Short Papers*, R. Barzilay and M. Kan, Eds. Association for Computational Linguistics, 2017. [Online]. Available: https://doi.org/10.18653/v1/P17-2067 pp. 422–426.

[14] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi, “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, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, M. Palmer, R. Hwa, and S. Riedel, Eds. Association for Computational Linguistics, 2017. [Online]. Available: https://doi.org/10.18653/v1/d17-1317 pp. 2931–2937.

[15] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, “Automatic detection of fake news,” in *Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, Santa Fe, New Mexico, USA, August 20-26, 2018*, E. M. Bender, L. Derczynski, and P. Isabelle, Eds. Association for Computational Linguistics, 2018. [Online]. Available: https://aclanthology.org/C18-1287/ pp. 3391–3401.

[16] S. D. Sarkar, F. Yang, and A. Mukherjee, “Attending sentences to detect satirical fake news,” in *Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, Santa Fe, New Mexico, USA, August 20-26, 2018*, E. M. Bender, L. Derczynski, and P. Isabelle, Eds. Association for Computational Linguistics, 2018. [Online]. Available: https://aclanthology.org/C18-1285/ pp. 3371–3380.

[17] P. Atanasova, P. Nakov, L. Màrquez, A. Barrón-Cedeño, G. Karadzhov, T. Mihaylova, M. Mohtarami, and J. R. Glass, “Automatic fact-checking using context and discourse information,” *ACM J. Data Inf. Qual.*, vol. 11, no. 3, pp. 12:1–12:27, 2019. [Online]. Available: https://doi.org/10.1145/3297722

[18] H. Karimi, P. Roy, S. Saba-Sadiya, and J. Tang, “Multi-source multi-class fake news detection,” in *Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, Santa Fe, New Mexico, USA, August 20-26, 2018*, E. M. Bender, L. Derczynski, and P. Isabelle, Eds. Association for Computational Linguistics, 2018. [Online]. Available: https://aclanthology.org/C18-1131/ pp. 1546–1557.

[19] J. A. Nasir, O. S. Khan, and I. Varlamis, “Fake news detection: A hybrid cnn-rnn based deep learning approach,” *International Journal of Information Management Data Insights*, vol. 1, no. 1, p. 100007, 2021.

[20] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever et al., “Language models are unsupervised multitask learners,” *OpenAI blog*, vol. 1, no. 8, p. 9, 2019.

[21] H. Karimi and J. Tang, "Learning hierarchical discourse-level structure for fake news detection," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, J. Burstein, C. Doran, and T. Solorio, Eds. Association for Computational Linguistics, 2019. [Online]. Available: https://doi.org/10.18653/v1/n19-1347 pp. 3432–3442.

[22] V. Vaibhav, R. M. Annasamy, and E. H. Hovy, "Do sentence interactions matter? leveraging sentence level representations for fake news classification," in *Proceedings of the Thirteenth Workshop on Graph-Based Methods for Natural Language Processing, TextGraphs@EMNLP 2019, Hong Kong, November 4, 2019*, D. Ustalov, S. Somasundaran, P. Jansen, G. Glavas, M. Riedl, M. Surdeanu, and M. Vazirgiannis, Eds. Association for Computational Linguistics, 2019. [Online]. Available: https://doi.org/10.18653/v1/D19-5316 pp. 134–139.

[23] L. Hu, T. Yang, L. Zhang, W. Zhong, D. Tang, C. Shi, N. Duan, and M. Zhou, "Compare to the knowledge: Graph neural fake news detection with external knowledge," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, C. Zong, F. Xia, W. Li, and R. Navigli, Eds. Association for Computational Linguistics, 2021. [Online]. Available: https://doi.org/10.18653/v1/2021.acl-long.62 pp. 754–763.

[24] J. Z. Pan, S. Pavlova, C. Li, N. Li, Y. Li, and J. Liu, "Content based fake news detection using knowledge graphs," in *The Semantic Web - ISWC 2018 - 17th International Semantic Web Conference, Monterey, CA, USA, October 8-12, 2018, Proceedings, Part I*, ser. Lecture Notes in Computer Science, D. Vrandecic, K. Bontcheva, M. C. Suárez-Figueroa, V. Presutti, I. Celino, M. Sabou, L. Kaffee, and E. Simperl, Eds., vol. 11136. Springer, 2018. [Online]. Available: https://doi.org/10.1007/978-3-030-00671-6_39 pp. 669–683.

[25] Z. Jin, J. Cao, Y. Zhang, and J. Luo, "News verification by exploiting conflicting social viewpoints in microblogs," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA*, D. Schuurmans and M. P. Wellman, Eds. AAAI Press, 2016. [Online]. Available: http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12128 pp. 2972–2978.

[26] L. Wu, Y. Rao, H. Jin, A. Nazir, and L. Sun, "Different absorption from the same sharing: Sifted multi-task learning for fake news detection," 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)*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019. [Online]. Available: https://aclanthology.org/D19-1471 pp. 4644–4653.

[27] J. Zhang, B. Dong, and P. S. Yu, "Fakedetector: Effective fake news detection with deep diffusive neural network," in *36th IEEE International Conference on Data Engineering, ICDE 2020, Dallas, TX, USA, April 20-24, 2020*. IEEE, 2020. [Online]. Available: https://doi.org/10.1109/ICDE48307.2020.00180 pp. 1826–1829.

[28] R. Tan, B. A. Plummer, and K. Saenko, "Detecting cross-modal inconsistency to defend against neural fake news," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, B. Webber, T. Cohn, Y. He, and Y. Liu, Eds. Association for Computational Linguistics, 2020. [Online]. Available: https://doi.org/10.18653/v1/2020.emnlp-main.163 pp. 2081–2106.

[29] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi, "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*. Copenhagen, Denmark: Association for Computational Linguistics, Sep. 2017. [Online]. Available: https://aclanthology.org/D17-1317 pp. 2931–2937.

[30] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic detection of fake news," in *Proceedings of the 27th International Conference on Computational Linguistics*. Santa Fe, New Mexico, USA: Association for Computational Linguistics, Aug. 2018. [Online]. Available: https://aclanthology.org/C18-1287 pp. 3391–3401.

[31] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media," *Big data*, vol. 8, no. 3, pp. 171–188, 2020.

[32] K. Shu, Y. Li, K. Ding, and H. Liu, "Fact-enhanced synthetic news generation," in *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*. AAAI Press, 2021. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/17629 pp. 13 825–13 833.

[33] A. Mosallanezhad, K. Shu, and H. Liu, "Generating topic-preserving synthetic news," in *2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, December 15-18, 2021*, Y. Chen, H. Ludwig, Y. Tu, U. M. Fayyad, X. Zhu, X. Hu, S. Byna, X. Liu, J. Zhang, S. Pan, V. Papalexakis, J. Wang, A. Cuzzocrea, and C. Ordonez, Eds. IEEE, 2021. [Online]. Available: https://doi.org/10.1109/BigData52589.2021.9671623 pp. 490–499.

[34] J. Thorne, A. Vlachos, C. Christodoulopoulos, and A. Mittal, "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, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers)*, M. A. Walker, H. Ji, and A. Stent, Eds. Association for Computational Linguistics, 2018. [Online]. Available: https://doi.org/10.18653/v1/n18-1074 pp. 809–819.

[35] T. Diggelmann, J. L. Boyd-Graber, J. Bulian, M. Ciaramita, and M. Leippold, "CLIMATE-FEVER: A dataset for verification of real-world climate claims," *CoRR*, vol. abs/2012.00614, 2020. [Online]. Available: https://arxiv.org/abs/2012.00614

[36] D. Wadden, S. Lin, K. Lo, L. L. Wang, M. van Zuylen, A. Cohan, and H. Hajishirzi, "Fact or fiction: Verifying scientific claims," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, B. Webber, T. Cohn, Y. He, and Y. Liu, Eds. Association for Computational Linguistics, 2020. [Online]. Available: https://doi.org/10.18653/v1/2020.emnlp-main.609 pp. 7534–7550.

[37] A. Saakyan, T. Chakrabarty, and S. Muresan, "Covid-fact: Fact extraction and verification of real-world claims on COVID-19 pandemic," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, C. Zong, F. Xia, W. Li, and R. Navigli, Eds. Association for Computational Linguistics, 2021. [Online]. Available: https://doi.org/10.18653/v1/2021.acl-long.165 pp. 2116–2129.

[38] T. Chakrabarty, T. Alhindi, and S. Muresan, "Robust document retrieval and individual evidence modeling for fact extraction and verification." in *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*. Brussels, Belgium: Association for Computational Linguistics, Nov. 2018. [Online]. Available: https://aclanthology.org/W18-5521 pp. 127–131.

[39] A. Hanselowski, H. Zhang, Z. Li, D. Sorokin, B. Schiller, C. Schulz, and I. Gurevych, "UKP-athene: Multi-sentence textual entailment for claim verification," in *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*. Brussels, Belgium: Association for Computational Linguistics, Nov. 2018. [Online]. Available: https://aclanthology.org/W18-5516 pp. 103–108.

[40] Y. Nie, H. Chen, and M. Bansal, "Combining fact extraction and verification with neural semantic matching networks," in *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*. AAAI Press, 2019. [Online]. Available: https://doi.org/10.1609/aaai.v33i01.33016859 pp. 6859–6866.

[41] M. Taniguchi, T. Taniguchi, T. Takahashi, Y. Miura, and T. Ohkuma, "Integrating entity linking and evidence ranking for fact extraction and verification," in *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*. Brussels, Belgium: Association for Computational Linguistics, Nov. 2018. [Online]. Available: https://aclanthology.org/W18-5520 pp. 124–126.

[42] A. Chernyavskiy and D. Ilvovsky, "Extract and aggregate: A novel domain-independent approach to factual data verification," in *Proceedings of the Second Workshop on Fact Extraction and VERification (FEVER)*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019. [Online]. Available: https://aclanthology.org/D19-6612 pp. 69–78.

[43] B. Portelli, J. Zhao, T. Schuster, G. Serra, and E. Santus, "Distilling the evidence to augment fact verification models," in *Proceedings of the Third Workshop on Fact Extraction and VERification (FEVER)*. Online: Association for Computational Linguistics, July 2020. [Online]. Available: https://aclanthology.org/2020.fever-1.7 pp. 47–51.

[44] J. Zhou, X. Han, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "GEAR: Graph-based evidence aggregating and reasoning for fact verification," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, July 2019. [Online]. Available: https://aclanthology.org/P19-1085 pp. 892–901.

[45] Z. Liu, C. Xiong, M. Sun, and Z. Liu, "Fine-grained fact verification with kernel graph attention network," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics, July 2020. [Online]. Available: https://aclanthology.org/2020.acl-main.655 pp. 7342–7351.

[46] A. Soleimani, C. Monz, and M. Worring, "BERT for evidence retrieval and claim verification," in *Advances in Information Retrieval - 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14-17, 2020, Proceedings, Part II*, ser. Lecture Notes in Computer Science, J. M. Jose, E. Yilmaz, J. Magalhães, P. Castells, N. Ferro, M. J. Silva, and F. Martins, Eds., vol. 12036. Springer, 2020. [Online]. Available: https://doi.org/10.1007/978-3-030-45442-5_45 pp. 359–366.

[47] S. Subramanian and K. Lee, "Hierarchical Evidence Set Modeling for automated fact extraction and verification," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Online: Association for Computational Linguistics, Nov. 2020. [Online]. Available: https://aclanthology.org/2020.emnlp-main.627 pp. 7798–7809.

[48] W. Zhong, J. Xu, D. Tang, Z. Xu, N. Duan, M. Zhou, J. Wang, and J. Yin, "Reasoning over semantic-level graph for fact checking," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics, July 2020. [Online]. Available: https://aclanthology.org/2020.acl-main.549 pp. 6170–6180.

[49] J. Zhou, X. Han, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "GEAR: graph-based evidence aggregating and reasoning for fact verification," in *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, A. Korhonen, D. R. Traum, and L. Màrquez, Eds. Association for Computational Linguistics, 2019. [Online]. Available: https://doi.org/10.18653/v1/p19-1085 pp. 892–901.

[50] W. Yin and D. Roth, "TwoWingOS: A two-wing optimization strategy for evidential claim verification," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Brussels, Belgium: Association for Computational Linguistics, Oct.-Nov. 2018. [Online]. Available: https://aclanthology.org/D18-1010 pp. 105–114.

[51] C. Hidey and M. Diab, "Team SWEEPer: Joint sentence extraction and fact checking with pointer networks," in *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*. Brussels, Belgium: Association for Computational Linguistics, Nov. 2018. [Online]. Available: https://aclanthology.org/W18-5525 pp. 150–155.

[52] P. S. H. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., 2020. [Online]. Available: https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html

[53] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, J. Burstein, C. Doran, and T. Solorio, Eds. Association for Computational Linguistics, 2019. [Online]. Available: https://doi.org/10.18653/v1/n19-1423 pp. 4171–4186.

[54] Z. Zhong and D. Chen, "A frustratingly easy approach for entity and relation extraction," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Online: Association for Computational Linguistics, June 2021. [Online]. Available: https://aclanthology.org/2021.naacl-main.5 pp. 50–61.

[55] J. Ma, S. Wang, R. Anubhai, M. Ballesteros, and Y. Al-Onaizan, "Resource-enhanced neural model for event argument extraction," in *Findings of the Association for Computational Linguistics: EMNLP 2020*. Online: Association for Computational Linguistics, Nov. 2020. [Online]. Available: https://aclanthology.org/2020.findings-emnlp.318 pp. 3554–3559.

[56] S. Zheng, F. Wang, H. Bao, Y. Hao, P. Zhou, and B. Xu, "Joint extraction of entities and relations based on a novel tagging scheme," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Vancouver, Canada: Association for Computational Linguistics, July 2017. [Online]. Available: https://aclanthology.org/P17-1113 pp. 1227–1236.

[57] X. Zeng, D. Zeng, S. He, K. Liu, and J. Zhao, "Extracting relational facts by an end-to-end neural model with copy mechanism," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia: Association for Computational Linguistics, July 2018. [Online]. Available: https://aclanthology.org/P18-1047 pp. 506–514.

[58] Y. Luan, D. Wadden, L. He, A. Shah, M. Ostendorf, and H. Hajishirzi, "A general framework for information extraction using dynamic span graphs," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, June 2019. [Online]. Available: https://aclanthology.org/N19-1308 pp. 3036–3046.

[59] D. Wadden, U. Wennberg, Y. Luan, and H. Hajishirzi, "Entity, relation, and event extraction with contextualized span representations," 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)*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019. [Online]. Available: https://aclanthology.org/D19-1585 pp. 5784–5789.

[60] T. Zhang, H. Ji, and A. Sil, "Joint entity and event extraction with generative adversarial imitation learning," *Data Intelligence*, 2019.

[61] Y. Lin, H. Ji, F. Huang, and L. Wu, "A joint neural model for information extraction with global features," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics, July 2020. [Online]. Available: https://aclanthology.org/2020.acl-main.713 pp. 7999–8009.

[62] H. Wen, Y. Lin, T. Lai, X. Pan, S. Li, X. Lin, B. Zhou, M. Li, H. Wang, H. Zhang, X. Yu, A. Dong, Z. Wang, Y. Fung, P. Mishra, Q. Lyu, D. Surís, B. Chen, S. W. Brown, M. Palmer, C. Callison-Burch, C. Vondrick, J. Han, D. Roth, S.-F. Chang, and H. Ji, "RESIN: A dockerized schema-guided cross-document cross-lingual cross-media information extraction and event tracking system," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Demonstrations*. Online: Association for Computational Linguistics, June 2021. [Online]. Available: https://aclanthology.org/2021.naacl-demos.16 pp. 133–143.

[63] X. Du, Z. Zhang, S. Li, P. Yu, H. Wang, T. Lai, X. Lin, Z. Wang, I. Liu, B. Zhou, H. Wen, M. Li, D. Hannan, J. Lei, H. Kim, R. Dror, H. Wang, M. Regan, Q. Zeng, Q. Lyu, C. Yu, C. Edwards, X. Jin, Y. Jiao, G. Kazeminejad, Z. Wang, C. Callison-Burch, M. Bansal, C. Vondrick, J. Han, D. Roth, S.-F. Chang, M. Palmer, and H. Ji, "RESIN-11: Schema-guided event prediction for 11 newsworthy scenarios," in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: System Demonstrations*. Hybrid: Seattle, Washington + Online: Association for Computational Linguistics, July 2022. [Online]. Available: https://aclanthology.org/2022.naacl-demo.7 pp. 54–63.

[64] M. S. Schlichtkrull, T. N. Kipf, P. Bloem, R. van den Berg, I. Titov, and M. Welling, "Modeling relational data with graph convolutional networks," in *The Semantic Web - 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings*, ser. Lecture Notes in Computer Science, A. Gangemi, R. Navigli, M. Vidal, P. Hitzler, R. Troncy, L. Hollink, A. Tordai, and M. Alam, Eds., vol. 10843. Springer, 2018. [Online]. Available: https://doi.org/10.1007/978-3-319-93417-4_38 pp. 593–607.

[65] L. Hu, T. Yang, C. Shi, H. Ji, and X. Li, "Heterogeneous graph attention networks for semi-supervised short text classification," 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 2019, Hong Kong, China, November 3-7, 2019*, K. Inui, J. Jiang, V. Ng, and X. Wan, Eds. Association for Computational Linguistics, 2019. [Online]. Available: https://doi.org/10.18653/v1/D19-1488 pp. 4820–4829.

[66] Z. Song, A. Bies, S. M. Strassel, T. Riese, J. Mott, J. Ellis, J. Wright, S. Kulick, N. Ryant, and X. Ma, "From light to rich ERE: annotation of entities, relations, and events," in *Proceedings of the The 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation, EVENTS@HLP-NAACL 2015, Denver, Colorado, USA, June 4, 2015*, E. H. Hovy, T. Mitamura, and M. Palmer, Eds. Association for Computational Linguistics, 2015. [Online]. Available: https://doi.org/10.3115/v1/W15-0812 pp. 89–98.

[67] X. Pan, B. Zhang, J. May, J. Nothman, K. Knight, and H. Ji, "Cross-lingual name tagging and linking for 282 languages," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, R. Barzilay and M. Kan, Eds. Association for Computational Linguistics, 2017. [Online]. Available: https://doi.org/10.18653/v1/P17-1178 pp. 1946–1958.

[68] K. Lee, L. He, M. Lewis, and L. Zettlemoyer, "End-to-end neural coreference resolution," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, M. Palmer, R. Hwa, and S. Riedel, Eds. Association for Computational Linguistics, 2017. [Online]. Available: https://doi.org/10.18653/v1/d17-1018 pp. 188–197.

[69] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov, "Unsupervised cross-lingual representation learning at scale," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, D. Jurafsky, J. Chai, N. Schluter, and J. R. Tetreault, Eds. Association for Computational Linguistics, 2020. [Online]. Available: https://doi.org/10.18653/v1/2020.acl-main.747 pp. 8440–8451.

[70] S. Pradhan, A. Moschitti, N. Xue, O. Uryupina, and Y. Zhang, "Conll-2012 shared task: Modeling multilingual unrestricted coreference in ontonotes," in *Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning - Proceedings of the Shared Task: Modeling Multilingual Unrestricted Coreference in OntoNotes, EMNLP-CoNLL 2012, July 13, 2012, Jeju Island, Korea*, S. Pradhan, A. Moschitti, and N. Xue, Eds. ACL, 2012. [Online]. Available: https://aclanthology.org/W12-4501/ pp. 1–40.

[71] M. Joshi, D. Chen, Y. Liu, D. S. Weld, L. Zettlemoyer, and O. Levy, "Spanbert: Improving pre-training by representing and predicting spans," *Trans. Assoc. Comput. Linguistics*, vol. 8, pp. 64–77, 2020. [Online]. Available: https://transacl.org/ojs/index.php/tacl/article/view/1853

[72] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018. [Online]. Available: https://openreview.net/forum?id=rJXMpikCZ

[73] K. Huang, M. Yang, and N. Peng, "Biomedical event extraction on graph edge-conditioned attention networks with hierarchical knowledge graphs," in *Findings of the Association for Computational Linguistics: EMNLP 2020, Online Event, 16-20 November 2020*, ser. Findings of ACL, T. Cohn, Y. He, and Y. Liu, Eds., vol. EMNLP 2020. Association for Computational Linguistics, 2020. [Online]. Available: https://doi.org/10.18653/v1/2020.findings-emnlp.114 pp. 1277–1285.

[74] M. Yasunaga, H. Ren, A. Bosselut, P. Liang, and J. Leskovec, "QA-GNN: reasoning with language models and knowledge graphs for question answering," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021*, K. Toutanova, A. Rumshisky, L. Zettlemoyer, D. Hakkani-Tür, I. Beltagy, S. Bethard, R. Cotterell, T. Chakraborty, and Y. Zhou, Eds. Association for Computational Linguistics, 2021. [Online]. Available: https://doi.org/10.18653/v1/2021.naacl-main.45 pp. 535–546.

[75] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, "BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, D. Jurafsky, J. Chai, N. Schluter, and J. R. Tetreault, Eds. Association for Computational Linguistics, 2020. [Online]. Available: https://doi.org/10.18653/v1/2020.acl-main.703 pp. 7871–7880.

[76] M. Li, S. Li, Z. Wang, L. Huang, K. Cho, H. Ji, J. Han, and C. R. Voss, "Future is not one-dimensional: Graph modeling based complex event schema induction for event prediction," *CoRR*, vol. abs/2104.06344, 2021. [Online]. Available: https://arxiv.org/abs/2104.06344

[77] G. B. Tran, M. Alrifai, and D. Q. Nguyen, "Predicting relevant news events for timeline summaries," in *22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013, Companion Volume*, L. Carr, A. H. F. Laender, B. F. Lóscio, I. King, M. Fontoura, D. Vrandecic, L. Aroyo, J. P. M. de Oliveira, F. Lima, and E. Wilde, Eds. International World Wide Web Conferences Steering Committee / ACM, 2013. [Online]. Available: https://doi.org/10.1145/2487788.2487829 pp. 91–92.

[78] G. B. Tran, M. Alrifai, and E. Herder, "Timeline summarization from relevant headlines," in *Advances in Information Retrieval - 37th European Conference on IR Research, ECIR 2015, Vienna, Austria, March 29 - April 2, 2015. Proceedings*, ser. Lecture Notes in Computer Science, A. Hanbury, G. Kazai, A. Rauber, and N. Fuhr, Eds., vol. 9022, 2015. [Online]. Available: https://doi.org/10.1007/978-3-319-16354-3_26 pp. 245–256.

[79] G. Stanovsky, J. Michael, L. Zettlemoyer, and I. Dagan, "Supervised open information extraction," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers)*, M. A. Walker, H. Ji, and A. Stent, Eds. Association for Computational Linguistics, 2018. [Online]. Available: https://doi.org/10.18653/v1/n18-1081 pp. 885–895.