HC-COVID: A Hierarchical Crowdsource Knowledge Graph Approach to Explainable COVID-19 Misinformation Detection

ZIYI KOU* and LANYU SHANG*, University of Illinois at Urbana-Champaign, USA
YANG ZHANG, University of Notre Dame, USA
DONG WANG, University of Illinois at Urbana-Champaign, USA

The proliferation of social media has promoted the spread of misinformation that raises many concerns in our society. This paper focuses on a critical problem of explainable COVID-19 misinformation detection that aims to accurately identify and explain misleading COVID-19 claims on social media. Motivated by the lack of COVID-19 relevant knowledge in existing solutions, we construct a novel crowdsource knowledge graph based approach to incorporate the COVID-19 knowledge facts by leveraging the collaborative efforts of expert and non-expert crowd workers. Two important challenges exist in developing our solution: i) how to effectively coordinate the crowd efforts from both expert and non-expert workers to generate the relevant knowledge facts for detecting COVID-19 misinformation; ii) How to leverage the knowledge facts from the constructed knowledge graph to accurately explain the detected COVID-19 misinformation. To address the above challenges, we develop HC-COVID, a hierarchical crowdsource knowledge graph based framework that explicitly models the COVID-19 knowledge facts contributed by crowd workers with different levels of expertise and accurately identifies the related knowledge facts to explain the detection results. We evaluate HC-COVID using two public real-world datasets on social media. Evaluation results demonstrate that HC-COVID significantly outperforms state-of-the-art baselines in terms of the detection accuracy of misleading COVID-19 claims and the quality of the explanations.

CCS Concepts: • **Human-centered computing** → **Collaborative and social computing**; • **Computing methodologies** → **Machine learning**; *Artificial intelligence.*

Additional Key Words and Phrases: COVID19; Explainable Misinformation Detection; Human-AI Collaboration

ACM Reference Format:
Ziyi Kou, Lanyu Shang, Yang Zhang, and Dong Wang. 2022. HC-COVID: A Hierarchical Crowdsource Knowledge Graph Approach to Explainable COVID-19 Misinformation Detection. *Proc. ACM Hum.-Comput. Interact.* 6, GROUP, Article 36 (January 2022), 25 pages. https://doi.org/10.1145/3492855

1 INTRODUCTION

Online social media has been a popular channel for news consumption in recent years, where billions of online posts are generated on a daily basis[1]. The proliferation of social media posts also leads to the

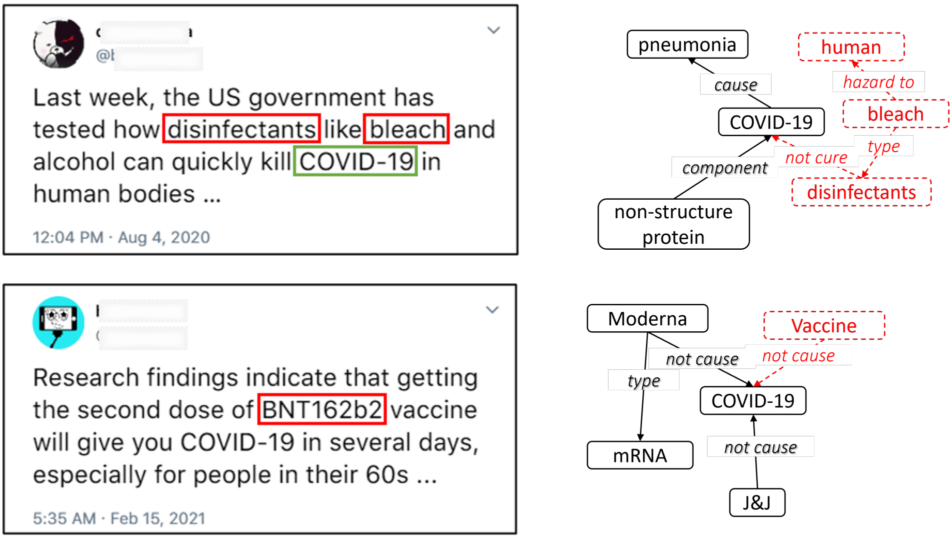
*Both authors contributed equally to this research.

[1] https://www.domo.com/learn/data-never-sleeps-8

Authors' addresses: Ziyi Kou, ziyikou2@illinois.edu; Lanyu Shang, lshang3@illinois.edu, University of Illinois at Urbana-Champaign, Champaign, Illinois, USA, 61820; Yang Zhang, University of Notre Dame, Notre Dame, Indiana, USA, 46556, yzhang42@nd.edu; Dong Wang, University of Illinois at Urbana-Champaign, Champaign, Illinois, USA, 61820, dwang24@illinois.edu.

2573-0142/2022/1-ART36 $15.00
https://doi.org/10.1145/3492855

propagation of misleading information on a variety of topics, such as politics [13], entertainment [7], food safety [48], and human health [17]. Among them, health-related misinformation detection is particularly challenging because it is not a trivial task for non-health researchers who often do not have sufficient medical knowledge to accurately identify false claims about health issues on social media [47]. The rapid spread of health-related misinformation can result in severe negative impacts on social media users and human society [5]. For example, more than 50 cell towers in the UK were attacked by people who believe the misinformation on social media that the COVID-19 virus can spread via 5G networks[2]. In this paper, we focus on a critical problem of the explainable COVID-19 misinformation detection where the goal is to accurately identify and explain misleading COVID-19 claims on social media.



[redacted]

Last week, the US government has tested how disinfectants like bleach and alcohol can quickly kill COVID-19 in human bodies ...

12:04 PM · Aug 4, 2020

[redacted]

Research findings indicate that getting the second dose of BNT162b2 vaccine will give you COVID-19 in several days, especially for people in their 60s ...

5:35 AM · Feb 15, 2021

Fig. 1. COVID-19 Misinformation Detection Problem

A significant amount of efforts have been made to combat the spread of online misinformation on social media [23, 33, 42, 43, 60, 64]. Existing online misinformation detection solutions have been focusing on content-based features (e.g., textual content [33], visual information [23]), context-based factors (e.g., user comments [64], social interaction [43]), and propagation-based characteristics (e.g., news cascade patterns [42]). Recently, a few initial efforts have been made to address the problem of detecting misleading health-related claims by leveraging the medical knowledge graphs [11, 27, 45]. Medical knowledge graphs contain medical or health-related knowledge facts (e.g., relations between two medical concepts) that are usually extracted from medical research articles. These medical knowledge facts are leveraged to detect and explain misleading health-related claims. However, current medical knowledge graph based solutions are not applicable to solve our COVID-19 misinformation detection problem due to some of their inherent limitations [15, 58]. We illustrate two key limitations of current solutions in Figure 1. For both examples (a) and (b) in Figure 1, the social media posts (i.e., tweets) shown on the left contain misleading claims related to COVID-19. The images on the right show two medical knowledge graphs that contain COVID-19 related knowledge facts. For each knowledge graph, the black text denotes existing knowledge

[2] https://www.jmir.org/2020/5/e19458/

facts (e.g., “COVID-19” $\xrightarrow{\text{cause}}$ “pneumonia” in Figure1(a)) in current knowledge graphs. The red text represents the knowledge facts that are necessary to identify misleading claims in the posts but does not exist in current knowledge graphs (e.g., “disinfectants” $\xrightarrow{\text{not cure}}$ “COVID-19”).

The first limitation of current medical knowledge graphs is the *lack of COVID-19 specific knowledge facts* that are critical to detect COVID-19 related misinformation on social media. COVID-19 specific knowledge facts are the knowledge facts that are closely related to COVID-19 and often include both medical and non-medical concepts. For example, the social media post in Figure 1(a) claims the wrong usage of disinfectants to protect humans from COVID-19. However, current knowledge graphs for COVID-19 are usually constructed from professional medical documents that often do not cover the knowledge facts related to non-medical concepts in a timely manner [15]. The second limitation of current medical knowledge graphs lies in the *lack of COVID-19 generalized knowledge facts*. Current medical knowledge graphs about COVID-19 heavily rely on the extraction of COVID-19 related knowledge facts from existing medical research documents and fail to abstract them into generalized knowledge facts for detecting *unseen* misleading posts. For example, the social media post in Figure 1(b) contains a false statement about the new BNT162b2 COVID-19 vaccine that is not promptly covered in the knowledge graph due to the delay of publications [9]. With the above two limitations, the current medical COVID-19 knowledge graphs do not contain enough COVID-19 background knowledge to cover the related misleading content in COVID-19 social media posts. Therefore, if an application aims to detect COVID-19 misinformation based on the limited knowledge graphs, the application is likely to retrieve little useful information from the knowledge graphs as explanations or even makes the wrong prediction results. However, it is well-known that humans are good at summarizing and abstracting key knowledge from different types of text (e.g., news articles, research documents) [53, 54]. For example, most people can conclude that no COVID-19 vaccine gets people infected after reading one or two articles discussing specific vaccines. Therefore, such abstract human knowledge is well complemented with the existing knowledge graphs to collaboratively identify different types of COVID-19 misinformation.

Motivated by the above limitations, we develop a human-AI collaborative approach to accurately identify and explain misleading COVID-19 information on social media. In particular, we construct a novel hierarchical crowdsource knowledge graph by utilizing human intelligence from a group of *expert workers* (i.e., crowd workers with working experience in healthcare) and *non-expert workers* (i.e., crowd workers without such a qualification) from a crowdsourcing platform (e.g., Amazon Mechanical Turks). We leverage the complementary strengths of both types of workers (e.g., reliability vs. availability) to capture both the *specific* and *generalized* knowledge facts from credible COVID-19 articles (i.e., professional medical news and credible fact-checking articles). The captured knowledge facts are then used to construct a hierarchical knowledge graph that can identify misleading COVID-19 claims and explain why the identified claims are false. However, two important challenges remain to be addressed in developing our solution.

Coordination Between Expert and Non-expert Crowd Efforts. The first challenge lies in how to effectively coordinate the diverse crowd efforts from expert and non-expert workers to generate the specific and generalized knowledge facts that are essential elements to construct the hierarchical knowledge graph for COVID-19 misinformation detection. A possible solution is to ask non-expert workers to annotate specific knowledge facts from credible COVID-19 articles and ask expert workers to summarize the generalized knowledge facts based on their existing knowledge about COVID-19. However, the specific knowledge fact annotations from non-expert crowd workers are observed to be noisy due to the unvetted nature of these crowd workers and their lack of COVID-19 related medical knowledge [19, 52]. Moreover, the generalized knowledge facts provided by the expert workers are often either too brief (e.g., COVID-19 is a kind of virus) or too medical-specific (e.g., COVID-19 is detected by using RT-qPCR). Therefore, it remains a challenging problem on how to effectively coordinate and integrate the efforts from both expert and non-expert workers to extract useful knowledge facts for the hierarchical crowdsource knowledge graph construction.

Explainability with Hierarchical Knowledge Graph. The second challenge lies in how to leverage the knowledge facts from the constructed hierarchical knowledge graph to accurately explain the detected COVID-19 misinformation. Unlike traditional medical knowledge graphs that only contain homogeneous knowledge facts [11, 45], the hierarchical knowledge graph consists of both specific and generalized knowledge facts about COVID-19 that can provide comprehensive explanations for the detected COVID-19 misinformation. For example, a reasonable explanation for the post in Figure 1(a) could be "bleach" $\xrightarrow{\text{not cure}}$ "COVID-19" from COVID-19 specific knowledge facts and "chlorine" $\xrightarrow{\text{hazard to}}$ "human body" from the generalized knowledge facts. However, the hierarchical knowledge graph also contains knowledge facts that are not useful to explain the misleading COVID-19 claim in Figure1(a) (e.g., "COVID-19" $\xrightarrow{\text{component}}$ "non-structural proteins"). Therefore, it remains a challenging task to identify the exact knowledge facts of the hierarchical knowledge graph that can effectively contribute to the explanation of the detected COVID-19 misinformation.

To address the above challenges, we propose HC-COVID, a hierarchical crowdsource knowledge graph based framework to address the explainable COVID-19 misinformation detection problem. In particular, to address the first challenge, we explicitly construct a topic-guided hierarchical crowdsource knowledge graph to effectively model both the specific and generalized knowledge facts from credible COVID-19 articles. To address the second challenge, we propose a novel duo hierarchy attention based graph neural network to explore both COVID-19 specific and generalized knowledge facts in the constructed knowledge graph to identify the reasonable explanations for the detection results. To our best knowledge, HC-COVID is the first crowdsourcing framework that constructs a hierarchical knowledge graph to address the explainable COVID-19 misinformation detection problem. We evaluate the HC-COVID framework by utilizing two real-world datasets (i.e., CoAID [10] and CONSTRAINT [32]) from Twitter and Facebook. The results show that HC-COVID not only outperforms all compared baselines by finding the COVID-19 misinformation on social media more accurately but also provides reasonable and well-justified explanations to the detection results. In summary, our contributions are:

- We design a novel crowdsourcing approach for HC-COVID that leverages the collaboration of expert and non-expert crowd workers on COVID-19 to construct a hierarchical topic-guided COVID-19 knowledge graph.
- We propose a novel duo hierarchy attention based graph neural network for HC-COVID to *detect* and *explain* COVID-19 misinformation by exploring the specific and generalized knowledge facts from the constructed knowledge graph.
- We construct both quantitative and user studies to evaluate HC-COVID by comparing it with various state-of-the-art schemes. The experiment results demonstrate the effectiveness of HC-COVID on both COVID-19 misinformation detection and misinformation explanation.

2 RELATED WORK

2.1 Health Misinformation on Social Media

Health misinformation has emerged as a critical issue on online social media, and has drawn much attention in recent years [5, 25, 47]. A significant amount of efforts have been made to combat the spread of health misinformation online [11, 49]. For example, Ghenai *et al.* proposed a user-centric model that identifies users who are prone to spread misleading health-related information by extracting features based on users' attitudes, writing styles, and sentiments from their posts on social media [17]. Zhao *et al.* designed a machine learning based detection framework to detect misleading posts in online health communities by integrating a set of linguistic, topic, sentiment and behavioral features extracted from the post content (e.g., XGBoost) [65]. Safarnejad *et al.* analyzed the propagation patterns of health misinformation on social media by reconstructing the dissemination networks of social media posts to identify misleading health-related posts[36]. However, existing health misinformation detection solutions primarily rely on user behaviors/activities (e.g., post content, user comments, and attitudes) to detect misleading health information on social media. These solutions cannot fully address the problem of detecting health misinformation related to outbreaking diseases (e.g., COVID-19) since common social media users often lack disease-specific knowledge and can easily be misled by such health misinformation. In this paper, we develop a hierarchical crowdsource knowledge graph approach that effectively coordinates the efforts from both expert and non-expert crowd workers to identify critical knowledge facts for detecting COVID-19 misinformation.

2.2 Knowledge Graph

Our work is also related to knowledge graph (KG) that can effectively represent complex and unstructured documents as a structural graph consisting of a collection of relational knowledge facts [21, 56, 62]. Knowledge graph has been widely applied in many application domains, including natural language understanding [46], question answering [14], and recommender systems [18]. For example, Liu *et al.* proposed a knowledge-enabled language representation model K-BERT by incorporating domain knowledge facts from a commonsense knowledge graph with pre-trained language model (e.g., BERT) to effectively learn language representations (e.g., text embeddings) [29]. Huang *et al.* developed a knowledge embedding based question answering framework to answer natural language questions by recovering the embeddings of relevant knowledge facts in a large-scale knowledge graph [20]. Wang *et al.* designed a knowledge graph aware recommender system to learn user interests by propagating the information of user preferences through a knowledge graph [55]. More recently, knowledge graphs have been utilized to model professional medical knowledge for health-related study [28, 39]. Cui *et al.* leveraged knowledge facts extracted from a biomedical knowledge base to support the detection of misleading healthcare information related to well-studied diseases (e.g., diabetes and cancer) [11]. Current knowledge graph based health misinformation detection solutions mainly focus on the specific knowledge facts directly extracted from medical articles. However, they are insufficient to generalize the specific knowledge facts for detecting unseen health-related misinformation, especially for emerging diseases (e.g., COVID-19) about which our prior knowledge is very limited. In our work, we design a topic-guided hierarchical crowdsource knowledge graph that can effectively extract the generalized knowledge facts for identifying unseen misleading information related to COVID-19.

2.3 Explainable Artificial Intelligence

Recent progress in explainable artificial intelligence (XAI) has been made to address the black-box challenge in AI and improve the interpretability of AI-based models [37, 61]. A few solutions have been developed to explain the detection results for misinformation detection. For example, Shu *et al.* developed an explainable fake news detection system to interpret the prediction results by utilizing the co-attention mechanism in deep neural networks for capturing the explainable content in the news articles and user comments [41]. Lu *et al.* proposed a graph-aware co-attention neural network scheme to generate explanations for fake news detection by investigating user comments and retweet patterns on social media [30]. Kou *et al.* designed a graph neural network approach to detect and explain multimodal fauxtography posts on social media [26]. However, existing explainable misinformation detection solutions cannot be applied to solve the COVID-19 misinformation detection problem because none of them explicitly considers the COVID-19 specific knowledge facts. Recently, Ayoub *et al.* proposed an explainable COVID-19 misinformation detection method to detect and explain COVID-19 misinformation by learning semantic representations of COVID-19 posts based on deep natural language processing models [2]. However, the explanations in [2] are only the words extracted from the posts and are not sufficient enough to provide reasons behind the misleading post due to the limited content of the social media posts. Shiao *et al.* developed a COVID-19 misinformation detection framework that detects COVID-19 misinformation based on COVID-19 medical documents and provides most related documents as explanations [40]. However, the explanations from the documents cannot deal with *unseen* COVID-19 misinformation cases because the content of documents is not abstracted to generalized COVID-19 knowledge facts. In contrast, HC-COVID designs a duo hierarchy attention based approach that aims to effectively explain COVID-19 misinformation using the specific and generalized knowledge facts in a hierarchical crowdsource knowledge graph.

3 PROBLEM STATEMENT

In this section, we formally define our explainable COVID-19 misinformation detection problem. We first define a few key terms that will be used in the problem statement.

DEFINITION 1. **COVID-19 Claim:** A COVID-19 claim is a piece of short text (e.g., social media post) that describes a news fact or statement related to COVID-19 (e.g., the social media posts in Figure 1). In particular, we define a set of N COVID-19 claims as $C = \{c_1, c_2, \cdots, c_N\}$. $c_n = \{w_{n,1}, w_{n,2}, \cdots, w_{n,L}\}$ is the n^{th} post in C where w_l is the l^{th} word of c_n.

DEFINITION 2. **COVID-19 Misinformation:** A COVID-19 claim is considered misleading if it contains partial or entire false or unverified information that can mislead its audience. Otherwise, the COVID-19 claim is considered non-misleading. The reason of the definition is to explicitly separate non-misleading claims from the other claims in order to preserve the true COVID-19 information. In particular, for a given post c_n, we define $y_n \in \{1, 0\}$ as the label to indicate if the claim is misleading or not (i.e., 1 for misleading and 0 otherwise).

DEFINITION 3. **COVID-19 Article:** A COVID-19 article is an article related to COVID-19, such as a news article from reliable healthcare news publishers (e.g., Center for Disease Control and Prevention (CDC)) or a debunking article from credible fact-checking websites (e.g., FactCheck.org[3]). In particular, we define a set of M COVID-19 articles as $\mathcal{A} = \{a_1, a_2, \cdots, a_M\}$.

DEFINITION 4. **COVID-19 Specific Knowledge Fact:** The COVID-19 specific knowledge fact refers to the COVID-19 knowledge facts that are extracted from COVID-19 articles and specific to the content of the articles. For example, the knowledge fact "Moderna" $\xrightarrow{\text{not cause}}$ "COVID-19" in Figure 1(b) is extracted from a COVID-19 article that claims no possibility for a person to get infected after receiving Moderna vaccines.

DEFINITION 5. **COVID-19 Generalized Knowledge Fact:** The COVID-19 generalized knowledge fact usually covers a set of COVID-19 articles that belong to similar topics. For example, the knowledge fact "Vaccine" $\xrightarrow{\text{not cause}}$ "COVID-19" in Figure 1(b) is a COVID-19 generalized knowledge fact indicating that no COVID-19 vaccine results in COVID-19 infection.

DEFINITION 6. **Crowd Query (Q):** A crowd query Q is a crowdsourcing task we assign to the crowd workers on the crowdsourcing platform. In HC-COVID, we develop a hierarchical COVID-19 knowledge extraction query that asks crowd workers to extract both COVID-19 specific and

[3] https://www.factcheck.org/fake-news/

generalized knowledge facts from a set of credible COVID-19-articles. We will elaborate on the details of the crowdsourcing tasks in the next section.

Definition 7. **Expert Workers ($\mathcal{W}^+$):** The expert workers are a set of crowd workers who are verified by the crowdsourcing platform (e.g., Amazon MTurk) to have professional healthcare knowledge/experience. Furthermore, we develop a set of COVID-19 related screening questions in the crowd queries to ensure the selected expert workers know COVID-19 related knowledge (e.g., COVID-19 spread, cure, and prevention) in order to provide reliable responses.

Definition 8. **Non-Expert Workers ($\mathcal{W}^-$):** The non-expert workers are a set of crowd workers on the crowdsourcing platform who are not proved to have professional healthcare knowledge/experience and may not be able to understand professional medical terms in COVID-19 news articles. For example, a non-expert worker may not know the meaning of acetaminophen (an active ingredient of Tylenol) and think it as a toxic drug incorrectly.

Given the above definitions, the goal of our explainable COVID-19 misinformation detection problem is to i) classify COVID-19 claims into two categories (i.e., misleading and non-misleading); ii) explain why the COVID-19 claim is misleading or not based on the correlation between the COVID-19 claim and the corresponding knowledge facts in the constructed hierarchical COVID-19 knowledge graph. In particular, if a COVID-19 claim is misleading, we will retrieve the explanation that debunks the claim. Otherwise, we will provide the explanation that supports the claim. Formally, our problem is defined as:

$$\underset{\hat{y}_n \in \mathbb{P}, \hat{\mathcal{T}} \in \mathbb{G}}{\arg\max} \; Pr(\hat{y}_n = y_n | c_n, \mathbb{G}, \mathcal{W}), \forall \; 1 \leq n \leq N \tag{1}$$

where $\mathcal{W} = \mathcal{W}^+ \cup \mathcal{W}^-$ denotes the joint set of expert and non-expert workers. $\hat{y_n}$ and y_n are the estimated and ground truth label of the COVID-19 claim c_n, respectively. $\mathbb{G}$ is the constructed hierarchical COVID-19 knowledge graph and $\hat{\mathcal{T}}$ represents a set of retrieved triples from $\mathbb{G}$ as explanations for $\hat{y_n}$.

4 SOLUTION

In this section, we present the HC-COVID scheme to address the problem of explainable COVID-19 misinformation detection discussed in the previous section. The overview of the HC-COVID scheme is shown in Figure 2. HC-COVID consists of four modules: 1) a Crowdsource Knowledge Graph Constructor (CKGC), 2) a Claim-guided Specific Knowledge Propagator (CSKP), 3) a Topic-based Generalized Knowledge Integrator (TGKI), and 4) a Joint Claim-Graph-based Misinformation Detector (CGMD). First, CKGC constructs the crowdsource hierarchical knowledge graph (CHKG) by leveraging a group of expert and non-expert crowd workers to identify COVID-19 specific knowledge facts and COVID-19 generalized knowledge facts collaboratively from COVID-19 articles. Second, the CSKP module develops a multi-relational graph neural network to encode input COVID-19 claims and integrate the claim information with COVID-19 specific knowledge facts in CHKG. Third, the TGKI module explores the COVID-19 generalized knowledge facts in CHKG that are strongly correlated with input claims by designing a duo hierarchy attention based neural network. The attention outputs are used to retrieve informative graph triples from CHKG as explanations for the COVID-19 misinformation detection results. Finally, the CGMD determines whether the input claim is misleading by jointly exploring the encoded claims and CHKG. We discuss the above modules in detail below.

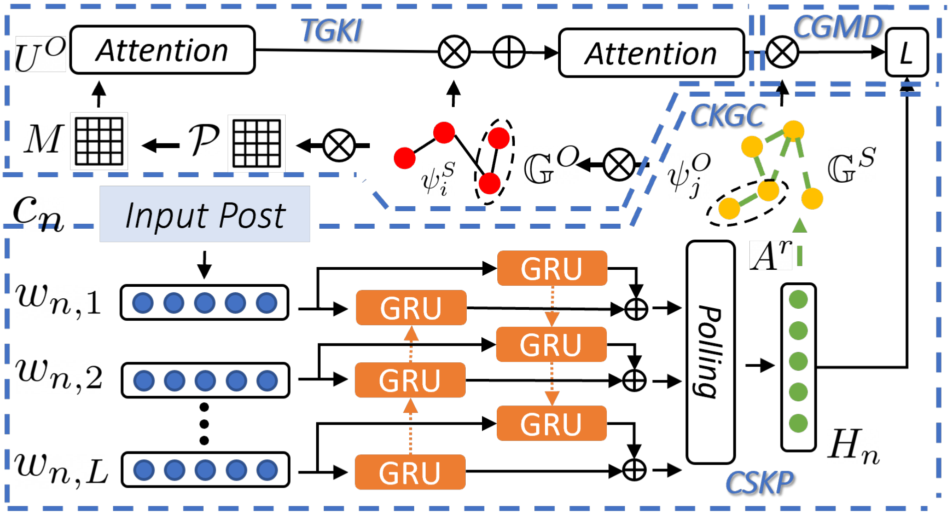


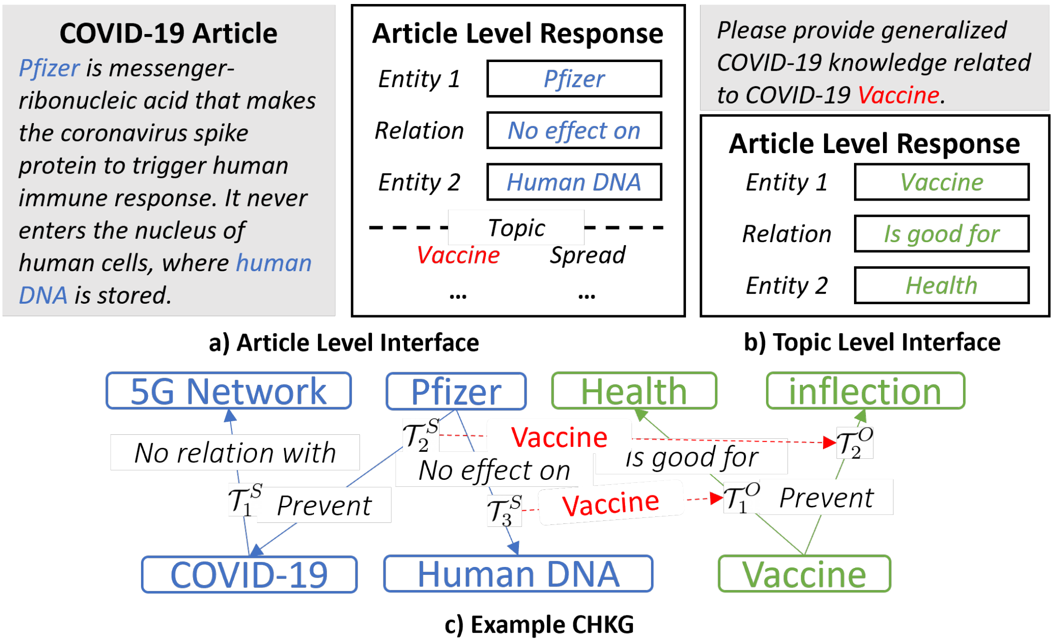
Fig. 2. Overview of HC-COVID Structure

4.1 Crowdsource Knowledge Graph Constructor (CKGC)

CKGC aims to construct CHKG that contains both COVID-19 specific knowledge facts and generalized knowledge facts from COVID-19 articles. In our approach, we design two novel crowdsourcing tasks for a group of expert workers and non-expert workers to analyze a set of COVID-19 articles. Unlike traditional crowdsourcing tasks that only assign workers simple annotation tasks (e.g., image annotation [8, 63], text classification [16, 51]), the CKGC designs a novel crowdsourcing task that expects crowd workers to understand and summarize the content of COVID-19 articles by leveraging their background knowledge. In particular, we design two crowdsourcing task interfaces (i.e., the *article-level* interface and the *topic-level* interface) for the crowd workers. The examples of the two interfaces are shown in Figure 3. The *article-level* interface helps workers to explore COVID-19 specific knowledge facts in COVID-19 articles. The *topic-level* interface lets crowd workers focus on the summarized COVID-19 topics from the article-level interface and propose generalized knowledge facts that can help identify misleading claims with similar topics. We define the responses from crowd workers as *article-level* responses and *topic-level* responses, respectively.

Article-Level Response: The article-level interface requires the non-expert crowd workers to provide COVID-19 specific knowledge facts based on a single COVID-19 article. In order to integrate the crowd responses into CHKG to detect COVID-19 misinformation, we specify the following requirements for the crowd worker's responses.

- A worker needs to provide a 3-tuple statement as shown in Figure 3(a) (e.g., "Entity 1" $\xrightarrow{\text{relation}}$ "Entity 2").
- The input entities should match existing terms in the COVID-19 article. The reason is that non-expert workers usually do not have enough COVID-19 related background knowledge to propose novel concepts. However, they are able to extract key terms from the COVID-19 article to summarize the specific knowledge facts [35].
- The input relation should be selected from the predefined relation pool that includes a set of frequently used relations (e.g., is, close relation to, no effect on) identified by crowd workers in our pilot study. The details of the relation pool is discussed in Section 5.



a) Article Level Interface

b) Topic Level Interface

c) Example CHKG

Fig. 3. Crowdsourcing Interface and Example Result

For example, an article-level response from a non-expert worker could be “Pfizer” $\xrightarrow{\text{no effect on}}$ “human DNA”, which is a COVID-19 specific knowledge fact for the Pfizer vaccine. However, such the COVID-19 knowledge fact summarized from a single COVID-19 article is insufficient to cover the information from other COVID-19 vaccine related articles (e.g., articles discussing Moderna or J&J vaccine). To address this problem, we further ask the non-expert workers to submit the potential COVID-19 topics of the COVID-19 article they read. We create a COVID-19 topic pool for selection and will discuss the selection process in Section 5.

Topic-Level Response: While the article-level responses only focus on the COVID-19 specific knowledge within a single COVID-19 article, the topic-level response serves as a complementary measure to assign expert workers to propose COVID-19 generalized knowledge. There are two key advantages of the topic-level interface: 1) it is both cost and time efficient for expert workers to focus on abstract COVID-19 topics and propose generalized COVID-19 knowledge that can cover the information of different COVID-19 articles; 2) it significantly improves the robustness of the constructed hierarchical knowledge graph because the COVID-19 generalized knowledge can cover unseen COVID-19 concepts that are embedded in COVID-19 articles with similar topics. For example, the generalized COVID-19 knowledge for Figure 3(b) could be “Vaccine” $\xrightarrow{\text{no effect on}}$ “DNA” that covers not only the current main COVID-19 vaccines (e.g., Pfizer, Moderna) but also the unseen and emerging ones (e.g., BNT162b2 in Figure 1(b)). In particular, the topic-level interface firstly collects the COVID-19 topics from responses of the article-level interface. Then the topic-level interface shows the topics to the expert workers and expect the workers to propose COVID-19 generalized knowledge facts related to the given topics. To ensure the quality of both article-level and topic-level responses from crowd workers, we leverage a set of crowd quality control mechanisms (e.g., HITs worker filtering, entity matching) [3, 50] to obtain high-quality responses.

After collecting all responses from both interfaces, the next step is to construct CHKG. The responses from both article-level and topic-level interfaces serve as triples in the knowledge graph. In particular, the entities in the 3-tuple responses are used to construct graph entities and the relations are used to construct the graph edges. We show an example constructed crowdsourcing knowledge graph in Figure 3(c). Formally, we define CHKG and its two sub-graphs as below.

Definition 9. **Crowdsource Article-Level Knowledge Graph ($\mathbb{G}^S$):** the crowdsource article-level knowledge graph $\mathbb{G}^S = \{\mathcal{V}^S, \mathcal{E}^S, \mathcal{T}^S\}$ (e.g., the blue subgraph in Figure 3(c)) contains COVID-19 specific knowledge constructed only by the triples from the article-level responses where $\mathcal{V}^S$, $\mathcal{E}^S$ and $\mathcal{T}^S$ represent the graph entities, graph edges and graph triples, respectively. We further split $\mathcal{E}^S$ as $\mathcal{E}^S = \{\mathcal{E}^{S,r_1}, \cdots, \mathcal{E}^{S,r_Q}\}$ where $\mathcal{R} = \{r_1, \cdots, r_Q\}$ represents all relations in the relation pool and $\mathcal{E}^{S,r_q}$ denotes the graph edges belonging to the relation of r_q.

Definition 10. **Crowdsource Topic-Level Knowledge Graph ($\mathbb{G}^O$):** the crowdsource topic-level knowledge graph $\mathbb{G}^O = \{\mathcal{V}^O, \mathcal{E}^O, \mathcal{T}^O\}$ (e.g., the green subgraph in Figure 3(c)) contains COVID-19 generalized knowledge constructed only by the triples from the topic-level responses. Similarly, $\mathcal{E}^O = \{\mathcal{E}^{O,r_1}, \cdots, \mathcal{E}^{O,r_Q}\}$ and $\mathcal{E}^{O,r_q}$ represents the graph edges belonging to the relation of r_q.

Definition 11. **Hierarchical Knowledge Graph ($\mathbb{G}$):** the hierarchical knowledge graph $\mathbb{G} = \{\mathcal{V}, \mathcal{E}, \mathcal{T}, \mathcal{P}\}$ is constructed by all the triples from both the article-level and topic-level responses where $\mathcal{V} = \{\mathcal{V}^S, \mathcal{V}^O\}$, $\mathcal{E} = \{\mathcal{E}^R, \mathcal{E}^O\}$ and $\mathcal{T} = \{\mathcal{T}^S, \mathcal{T}^O\}$ represents the graph entities, graph relations and graph triples, respectively. $\mathcal{P}$ denotes a binary bipartite adjacent matrix that contains COVID-19 topics to connect triples in $\mathbb{G}^S$ with the triples in $\mathbb{G}^O$. If an expert worker proposes a COVID-19 generalized knowledge fact $\mathcal{T}_i^O \in \mathcal{T}^O$ that is related to the COVID-19 topic "Vaccine", then we set $\mathcal{P}_{ij} = 1$ for all $\mathcal{T}_j^S$ from the COVID-19 articles that belong to the topic "Vaccine" in article-level responses. For example, the binary bipartite adjacent matrix $\mathcal{P} \in \mathbb{R}^{2\times 3}$ for Figure 3(c) is $\mathcal{P}_{i,j} = 1, 1 \le i \le 2, 2 \le j \le 3$ and $\mathcal{P}_{i,j} = 0, 1 \le i \le 2, j = 1$.

In the following sections, we discuss how the constructed CHKG can be utilized to solve the explainable COVID-19 misinformation detection problem.

4.2 Claim-guided Specific Knowledge Propagator (CSKP)

In this subsection, we present the CSKP in HC-COVID that propagates the encoded information of input COVID-19 claim to $\mathbb{G}^S$ for retrieving claim related COVID-19 specific knowledge. The CSKP consists of two specific components: 1) a COVID-19 claim feature encoder, and 2) a multi-relational specific knowledge propagator. We define the two network architectures below.

4.2.1 COVID-19 Claim Feature Encoder. The COVID-19 claim feature encoder aims to encode the input COVID-19 claim and extract high-level semantic features from the claim to propagate the feature of the claim into $\mathbb{G}^S$. We first design a *word-level* feature encoder that converts words in a COVID-19 claim to high-dimensional vectors in order to integrate the semantic information from different words. Given a COVID-19 claim $c_n = \{w_{n,1}, \cdots, w_{n,L}\}$, we convert all words in the claim to *one-hot* vectors and apply an embedding matrix to transform the vectors to a high-dimensional embedding. The embedding can be denoted as $\widetilde{c}_n = \{\widetilde{w}_{n,1}, \cdots, \widetilde{w}_{n,L}\}$ where each word denotes as $\widetilde{w}_{n,l} \in \mathbb{R}^d$.

Using the word embeddings from the word-level feature encoder, we design a bi-directional gated recurrent unit (biGRU) to encode the entire content of the claim. The biGRU strengthens the semantic connection between different words in a claim. In particular, given an embedded claim $\widetilde{c}_n$ with L word embeddings, the biGRU processes the embeddings from both directions of the claim.

The forward biGRU $\overrightarrow{f}_{gru}$ reads from the first word embedding to the last one while the backward biGRU $\overleftarrow{f}_{gru}$ reads them reversely. The process can be formally denoted as:

$$\begin{aligned} \overrightarrow{h}_{n,l} &= \overrightarrow{f}_{gru}(\widetilde{w}_{n,l}),\ l \in \{1, \ldots, L\} \\ \overleftarrow{h}_{n,l} &= \overleftarrow{f}_{gru}(\widetilde{w}_{n,l}),\ l \in \{1, \ldots, L\} \end{aligned} \tag{2}$$

where $\overrightarrow{h}_{n,l} \in \mathbb{R}^d$ and $\overleftarrow{h}_{n,l} \in \mathbb{R}^d$ are hidden states for the l^{th} word of c_n. We then obtain the feature of each word by concatenating its forward and backward hidden states, i.e., $h_{n,l} = [\overrightarrow{h}_{n,l}, \overleftarrow{h}_{n,l}] \in \mathbb{R}^{2d}$. The aggregated feature of c_n can be denoted as $h_n \in \mathbb{R}^{L\times 2d}$. We perform the word-level average pooling operation to integrate h_n into a single *claim-level* feature $H_n \in \mathbb{R}^{1\times 2d}$ that denotes the overall semantic representation of c_n.

4.2.2 Multi-Relational Specific Knowledge Propagator. Given the embedded claim-level feature H_n from Section 4.2.1, the multi-relational specific knowledge propagator aims to propagate the feature into $\mathbb{G}^S$ for retrieving claim related knowledge facts from $\mathbb{G}^S$. In particular, we represent $\mathbb{G}^S$ as a multi-relational graph neural network (RGCN) for the aggregation of COVID-19 specific knowledge. RGCN is a specific type of graph convolutional network that contains multiple types of relations between different graph entities [38]. In our explainable COVID-19 misinformation detection problem, we model $\mathbb{G}^S$ as an RGCN because it can effectively represent different relations in $\mathbb{G}^S$ (e.g., close relation to, no relation with) and aggregate COVID-19 specific knowledge with the information of the input claim. In particular, the entities $\mathcal{E}^S$ in $\mathbb{G}^S$ are represented as high-dimension entity embeddings $\widetilde{\mathcal{E}}^S \in \mathbb{R}^{E^S\times 2d}$ in RGCN where E^S is the number of unique entities in $\mathbb{G}^S$. Similarly, the relations $\mathcal{R}$ in $\mathbb{G}^S$ are represented as relation embeddings $\widetilde{\mathcal{R}} \in \mathbb{R}^{Q\times 2d}$. To learn the latent representations of the entities in $\mathbb{G}^S$, we develop a multi-relation information aggregation strategy defined as:

$$\widetilde{e}_i = \sigma(\sum_{r\in\mathcal{R}} \sum_{(j,r,i)\in\mathcal{T}^*} \frac{1}{z_{i,r}} W^r_{i,j} \widetilde{e}_j A^r_{i,j}) \tag{3}$$

where $\widetilde{e}_i \in \widetilde{\mathcal{E}}^S$ and $\widetilde{e}_j \in \widetilde{\mathcal{E}}^S$ are i^{th} and j^{th} graph entity embeddings in $\mathbb{G}^S$. σ stands for the non-linear activation ReLU function. $\mathcal{R}$ contains all available relations in $\mathbb{G}^S$ and $\mathcal{T}^* \in \mathcal{T}^S$ denotes the set of graph triples consisting of $\widetilde{e}_i$. $z_{i,r}$ is a normalization factor for $\widetilde{e}_i$ and $W^r_{i,j}$ is the learnable parameter. A^r is the adjacent matrix for the relation r and $A^r_{i,j}$ represents the scalar value for $\widetilde{e}_i$ and $\widetilde{e}_j$.

Unlike traditional graph neural networks approaches that simply merge different features together to indicate the relation between the features (e.g., concatenating H_i with entity embeddings in $\mathbb{G}^S$), the CSKP encodes H_n as an additional adjacent matrix for $\mathbb{G}^S$ in our RGCN to perform the claim guided graph convolution. The intuition is that the instance specific knowledge propagation in the RGCN should match the semantic content in the input COVID-19 claim to detect misleading information. For example, a claim that discusses the relation between the Pfizer vaccine and the human DNA can guide the RGCN to retrieve more Pfizer related knowledge facts from $\mathbb{G}^S$ to check the truthfulness of the claim. Formally, given an embedded claim feature $H_n \in \mathbb{R}^{1\times 2d}$, the process for generating the adjacent matrix with relation r in RGCN can be denoted as:

$$A^r = \widetilde{\mathcal{E}}^S \cdot (H_n)^T + H_n \cdot (\widetilde{\mathcal{E}}^S)^T \tag{4}$$

where $A^r \in \mathbb{R}^{E^S\times E^S}$ is the result adjacent matrix corresponding to the relation $r \in \mathcal{R}$. The final output of the multi-relation information aggregation is the updated entity embeddings $\widetilde{e}_i \in \mathbb{R}^{2d}$ given the input graph entity $e_i \in \mathbb{G}^S$.

4.3 Topic-based Generalized Knowledge Integrator (TGKI)

The previous knowledge graph based methods for misinformation detection mainly extract the knowledge from general health-related documents that are not specific to COVID-19. More importantly, the direct knowledge extraction from the documents cannot identify *unseen* COVID-19 misinformation because the knowledge is limited to the content of the documents and not fully generalized. To address the above limitations, the TGKI designs a novel hierarchical co-attention mechanism to retrieve both COVID-19 specific knowledge facts and generalized knowledge facts from CHKG as explanations for the misinformation detection results. We observe that retrieving accurate explanations from CHKG is determined by two correlation factors: 1) the correlation between the input COVID-19 claim and the COVID-19 specific knowledge from $\mathbb{G}^S$ and 2) the correlation between the COVID-19 specific knowledge from $\mathbb{G}^S$ and the generalized knowledge from $\mathbb{G}^O$. The first correlation determines whether the content of the input claim can be matched to any COVID-19 specific knowledge fact extracted from the COVID-19 articles. For example, a misleading COVID-19 claim that makes up a unrealistic side effect (e.g., COVID-19 infection) caused by Pfizer vaccine can be detected by Pfizer specific knowledge (e.g., "Pfizer" $\xrightarrow{\text{not cause}}$ "COVID-19") from $\mathbb{G}^S$. The second correlation determines whether there exists COVID-19 generalized knowledge facts from $\mathbb{G}^O$ that can provide explanations for the input claim based on its topic connections with COVID-19 specific knowledge facts from $\mathbb{G}^S$. For example, if there is no matched Pfizer specific knowledge fact from $\mathbb{G}^S$ for the input claim, TGKI detects the related COVID-19 specific knowledge facts (e.g., Moderna vaccine) and then retrieves their topic-wise connected generalized knowledge facts as explanations ("Vaccine" $\xrightarrow{\text{not cause}}$ "COVID-19").

To retrieve accurate and complementary explanations that explicitly consider the above correlation factors, we propose a novel duo hierarchy attention based neural network for TGKI. The duo hierarchy attention based neural network estimates the possibility of each COVID-19 knowledge fact from both $\mathbb{G}^S$ and $\mathbb{G}^O$ as the explanation for the misinformation detection results. We first define the *triple-level* embedding below for the graph triples from both $\mathbb{G}^S$ and $\mathbb{G}^O$.

Definition 12. **Triple-Level Embedding:** Triple-level embedding represents the semantic features of triples as the joint representations of graph entities and graph edges. Given an embedded triple $\widetilde{\mathcal{T}_k} = \{\widetilde{e_i}, \widetilde{r_q}, \widetilde{e_j}\}$ from CHKG, the triple-level embedding is denoted as $\psi_k = \widetilde{e_i} \odot \widetilde{r_q} \odot \widetilde{e_j} \in \mathbb{R}^{2d}$. For the embedded triples in $\mathbb{G}^S$ that are intergrated with the input claims in the CSKP module, the triple-level embeddings are denoted as $\psi^S \in \mathbb{R}^{N^S \times 2d}$ where N^S is the number of triples. Similarly, for the embedded triples in $\mathbb{G}^O$, the triple-level embeddings are denoted as $\psi^O \in \mathbb{R}^{N^O \times 2d}$ where N^O is the number of triples.

Given the triple-level embeddings ψ^S, our goal is to estimate the possibility of each triple in ψ^S of being the explanation for the input claim. In particular, the duo hierarchy attention based neural network generates the attention scores for ψ^S as $U^S = \text{Softmax}(\psi^S W^S)$ where $U^S \in \mathbb{R}^{N^S \times 1}$ are generated attention scores for all N^S triples from $\mathbb{G}^S$. The higher the score is, the more likely the corresponding triple is correlated with the input claim. In order to explore the complex correlation between ψ^S and ψ^O, the duo hierarchy attention based neural network designs a co-attention mechanism to generate attentions scores as:

$$M = \text{Softmax}_O(\tanh(\psi^S W_M (\psi^O)^T) \odot \mathcal{P}) \tag{5}$$

where $M \in \mathbb{R}^{N^S \times N^O}$ is the generated attention matrix and $M_{i,j}$ is the correlation score for i^{th} triple from $\mathbb{G}^S$ with j^{th} triple from $\mathbb{G}^O$. Softmax_O is the Softmax operation in the N^O dimension. To estimate the possibility for each triple from ψ^O of being the explanation for the input claim, the duo hierarchy attention based network integrate U^S into M, which can be denoted as $U^O = \text{Softmax}(M^T U^S) \in \mathbb{R}^{N^O \times 1}$. We concatenate U^S and U^O as $U = [U^S, U^O]$ as the comprehensive explanations for the input claim. The higher the score is, the more likely the corresponding graph triple can reasonably explain the detection results of HC-COVID.

4.4 Joint Claim-Graph-based Misinformation Detector (CGMD)

Given the feature of the input claim from CSKP and triple-level embeddings from TGKI, the CGMD module aims to determine if the input COVID-19 claim is misleading or not by designing a binary neural network classifier. In particular, given a COVID-19 claim c_n, we output the final prediction as:

$$\hat{y}_n = [H_n, \sum_{i=1}^{N^S} \psi_i^S \times U_i^S, \sum_{j=1}^{N^O} \psi_j^O \times U_j^O] W^b \tag{6}$$

where $[\cdot, \cdot, \cdot]$ denotes the concatenation operation to merge the features of the input COVID-19 claim with all triple-level embeddings from CHKG. $W^b \in \mathbb{R}^{5d \times 2}$ is the learnable parameter and $\hat{y}_n \in \mathbb{R}^2$ is the final prediction. Our loss function is the binary cross-entropy function that minimizes the loss between $\hat{y}_n$ and the ground-truth label y_n for each input COVID-19 claim c_n. The process is denoted as:

$$\mathcal{L} = \sum_{n=1}^{N} -y_n \log(\hat{y}_{n,2}) - (1 - y_n)\log(1 - \hat{y}_{n,1}) \tag{7}$$

where $\hat{y}_{n,1}$ and $\hat{y}_{n,2}$ are 1^{th} and 2^{th} scalar value in $\hat{y}_n$. The loss function measures the difference between two probability distributions (i.e., $\hat{y}_n$ and y_n) that is minimized by HC-COVID. We summarize the HC-COVID scheme in Algorithm 1.

Algorithm 1 HC-COVID Scheme Workflow

```
Objective: Detect and explain the input COVID-19 claim
Input: COVID-19 claim $c_n$, misinformation label $y_n$, COVID-19 article set $\mathcal{A}$, MTurk crowd worker group
    $\mathcal{W} = \{W^+, W^-\}$
Output: COVID-19 misinformation prediction $\hat{y}_n$, explanation list $U$ after $M$ iterations
$\mathbb{G}^S, \mathbb{G}^O, \mathbb{G}$ = CKGC($\mathcal{A}, \mathcal{W}$)
while $i \leq M$ do
    $H_n$ = CSKP $\rightarrow$ biGRU($c_n$)
    $\mathbb{G}^S, \mathbb{G}^O$ = CSKP $\rightarrow$ RGCN($\mathbb{G}^S$), CSKP $\rightarrow$ RGCN($\mathbb{G}^O$)
    $\mathbb{G}^S$ = CSKP $\rightarrow$ RGCN.conv($\mathbb{G}^S, H_n$)
    $\psi^S, \psi^O$ = TGKI $\rightarrow$ Triple-Level-Embed($\mathbb{G}^S$), TGKI $\rightarrow$ Triple-Level-Embed($\mathbb{G}^O$)
    $U^S$ = TGKI.attention($\psi^S$)
    $U^O$ = TGKI.co-attention($\psi^S, \psi^O, U^S$)
    $\hat{y}_n$ = CGMD($H_n, \psi^S, \psi_O, U^S, U^O$)
    AdamOptimization($\mathcal{L}(\hat{y}_n, y_n)$)
end while
$U = [U^S, U^O]$.sort()
return $\hat{y}_n$, $U$
```

5 EVALUATION

In this section, we conduct extensive experiments on two real-world COVID-19 misinformation datasets on social media to answer the following questions:

- **Q1:** Can HC-COVID achieve better misinformation detection performance than the state-of-the-art baselines?
- **Q2:** Given a post, can HC-COVID provide reasonable explanations based on the hierarchical crowdsource knowledge graph?
- **Q3:** How does each component of HC-COVID contribute to its overall performance?

5.1 Dataset and Experiment Setup

Dataset. We use two public COVID-19 misinformation datasets for our experiments. The first dataset is CoAID [10], a COVID-19 health misinformation dataset that consists of *COVID-19 articles* and *COVID-19 claims*. The COVID-19 articles contain reliable COVID-19 medical news and fact-checking articles including both medical and non-medical concepts. The COVID-19 claims contain 1, 000 true COVID-19 related tweets and 1, 000 misleading COVID-19 related tweets as the dataset for evaluating the HC-COVID and state-of-the-art baselines. The second dataset CONSTRAINT [32] is a large-scale COVID-19 fake news dataset that consists of 10, 700 COVID-19 related tweets. In particular, the CONSTRAINT dataset is utilized for evaluating the misleading information detection performance of HC-COVID where its hierarchical crowdsource knowledge graph is constructed from the COVID-19 articles in CoAID. For both datasets, we split the COVID-19 claims with 50% as training set, 20% as validation set, and 30% as testing set[4]. The summary of the two datasets are shown in Table 1.

Table 1. Dataset Summary

Dataset	Type	Data
CoAID [10]	COVID-19 articles	600
	True tweets	1,632
	Misleading tweets	544
CONSTRAINT [32]	True tweets	5,600
	Misleading tweets	5,100

Crowdsourcing Platform. For each COVID-19 article in article-level interface discussed in Section 4.1, we invite five independent Amazon MTurk workers to participate in the construction of our article-level knowledge graph. For topic-level interface, we select crowd workers who are verified by Amazon MTurk as "healthcare worker" and then develop a set of COVID-19 screening questions to select COVID-19 expert workers[5]. There are two types of potential biases in conducting both crowdsourcing tasks: 1) the *demographic bias* of various crowd workers and 2) the *opinion bias* of crowdsourcing responses. To mitigate the demographic bias of crowd workers, we follow the recruiting policy of Amazon MTurk and provide an equal opportunity for each crowd worker. In particular, we design the crowdsourcing interfaces and upload the interfaces to the Amazon MTurk website. The Amazon MTurk displays the interfaces publicly and accepts the interested crowd workers regardless of their demographic attributes (e.g., race, gender, age). To mitigate the opinion bias, we adopt the majority voting mechanism to collect crowdsourcing responses from both article-level and topic-level interfaces. In particular, we accept a submitted article-level or topic-level response only if a submitted 3-tuple statement is same in two or more responses. To

[4] We will make all our codes and datasets publicly available upon the acceptance of the paper.

[5] https://blog.mturk.com/introducing-premium-qualifications-1e473456e7b0

ensure the quality of responses from workers, the workers are selected only if they have a 98% or higher Human Intelligence Task (HIT) rate.

We perform a COVID-19 relation selection pilot study to identify 11 relations as the relation pool. In particular, we randomly select 150 COVID-19 articles from the CoAID dataset [10] and assign 4 non-expert crowd workers and 1 expert worker for each article to summarize the knowledge triples. We allow crowd workers to use free texts to indicate the relations between entities when they accomplish the two crowdsourcing tasks developed in Section 4.1. We then identify the 11 most frequent relations that are used by 15 or more crowd responses. The relation-count summary is as follows: {*is*: 96, *cause*: 81, *close relation with*: 65, *no relation with*: 62, *have*: 49, *is good for*: 39, *no effect on*: 31, *is not*: 27, *is bad for*: 20, *not have*: 17, *prevent*: 15}. Similarly, we carried out a COVID-19 topic selection pilot study to identify 8 unique COVID-19 topics in the article-level interface for non-expert workers to select. In particular, we randomly select 100 COVID-19 articles from the CoAID dataset and asks three COVID-19 expert workers to propose possible COVID-19 topics for each article. We randomly select expert workers from Amazon MTurk to reduce potential opinion bias from the workers. For each COVID-19 article, the selected expert worker needs to create three different COVID-19 topics that can cover the entire or most content of the article. After the study, we collect all proposed COVID-19 topics and select the 8 most frequent topics that are proposed more than 20 times by COVID-19 expert workers. The topic-count summary is: {*Prevention*: 71, *Virus Itself*: 52, *Cure*: 45, *Vaccine*: 44, *Spread*: 37, *Politics*: 32, *Influence*: 25, *Origin*: 21}.

We set the payment to all crowd workers well above the minimum requirement from MTurk[6]. We also closely follow the IRB protocol approved by our institution. The averaged time to complete an article-level and a topic-level task by a crowd worker is 76 and 194 seconds, respectively. Finally, we collect 640 valid triples for the article-level knowledge graph (i.e., nearly 1 knowledge triple for each COVID-19 article) and 80 valid triples for the topic-level graph (nearly 10 knowledge triples for each COVID-19 topic).

Experiment Setup. In our experiments, we pre-select the COVID-19 claims as the independent *testing set* and perform 10-fold cross validation on the *train-validation* set to estimate a more general performance of all schemes. For the implementation details of HC-COVID, the CIKP module holds 2 graph convolutional layers with each layer followed by the *ReLU* activation. We set the hidden state dimensions of the biGRU networks from CIKP as 128. We set the vocabulary size for the COVID-19 claims in the CoAID and CONSTRAINT datasets as $4,500$ and $6,000$, respectively. We set the total number of epochs as 40 and train HC-COVID with an initial learning rate of 0.001 and decay of 0.95 in each epoch. The optimizer is Adam with 5×10^{-4} weight decay. We run our experiments on Ubuntu 16.04 with two NVIDIA 1080Ti.

5.2 Baselines

We conduct experiments with state-of-the-art fake news detection models to evaluate the performance of HC-COVID.

- **HAN [59]:** a hierarchical attention network approach that applies both word-level and sentence-level mechanisms for document classification. In this paper, we use COVID-19 claims in CoAID or CONSTRAINT as the input documents and train the model to classify misleading claims from truthful claims.
- **PLAN [24]:** a multi-head attention network approach to detect rumors in social media by constructing a conversation tree that models the various interactions between the original rumor and the corresponding user replies. In particular, we replace the false and true rumors in PLAN as true and false claims for the classification task, respectively.

[6] https://www.mturk.com/pricing

- **MVAE [23]:** a variational autoencoder neural network approach for fake news detection by learning a hidden representation from the content of social media posts.
- **dEFEND [41]:** a fake news detection model that applies a co-attention strategy to retrieve important sentences from both the text content of fake news and the user comments by analyzing the interaction between them.
- **DETERRENT [11]:** a knowledge-guided graph attention network solution to detect misinformation in health-related articles by incorporating a medical knowledge graph and an article-entity bipartite graph. Specifically, we replace health-related articles in DETERRENT with COVID-19 claims and train the framework for the classification task.
- **COVID19-KG [15]:** a cause-and-effect knowledge model of COVID-19 pathophysiology. In particular, we replace the crowd knowledge graph in HC-COVID with the knowledge graph constructed by COVID19-KG to implement this baseline.
- **KMGCN [57]:** a knowledge-driven and graph-based model to detect fake news in social media posts by exploring the background knowledge hidden in the text content of the posts. In particular, we retrieve the specific graph triples from our hierarchical knowledge graph as background knowledge if the graph triples contain the same word as the one used in the input COVID-19 claim.

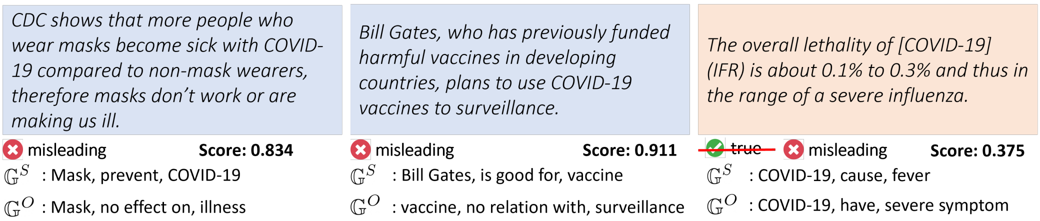
5.3 Evaluation Results

5.3.1 Detection Performance (Q1). To answer question Q1, we first evaluate the misinformation detection performance of the HC-COVID and all the baselines on both the CoAID and CONSTRAINT datasets. The evaluation results are shown in Table 2 and Table 3, respectively. We observe that HC-COVID consistently outperforms all the baseline methods on all evaluation metrics on both the CoAID and CONSTRAINT datasets. In particular, HC-COVID achieves performance gains of 2.2% and 2.7% in terms of F1 score compared to the best-performing baseline (i.e., DETERRENT) on the CoAID and CONSTRAINT datasets, respectively. Such a performance gain can be attributed to the incorporation of the COVID-19 generalized knowledge facts in the hierarchical knowledge graph that can effectively infer the truthfulness of an unseen COVID-19 claim. Moreover, we observe that the medical knowledge graph based baselines (i.e., DETERRENT, COVID19-KG, KMGCN) perform better than other baselines that do not utilize professional medical knowledge. Such an observation further verifies the effectiveness of leveraging medical knowledge facts for the detection of misleading COVID-19 misinformation. However, HC-COVID outperforms these medical knowledge graph based baselines because it develops a crowdsourcing approach to abstract both specific and generalized knowledge facts from COVID-19 articles.

We further visualize several testing cases in Figure 4 to evaluate the detection performance of HC-COVID. The score in each case is the prediction probability that represents the confidence level of HC-COVID. Figure 4(a) shows two testing cases that are correctly identified as misleading by HC-COVID but mis-classified by other baselines. The explanations from our hierarchical knowledge graph demonstrate that HC-COVID can accurately detect and explain the COVID-19 misinformation based on both specific (e.g., “Bill Gates” $\xrightarrow{\text{is good for}}$ “COVID-19”) and generalized knowledge facts (e.g., “Mask” $\xrightarrow{\text{no effect on}}$ “illness”). Moreover, we show one testing case in Figure 4(b) that all methods fail to detect the misinformation in it. The reason of the post being misleading is the actual IFR of COVID-19 is more than 0.65% that cannot be classified as any common influenza. However, the detection of manipulation on real numbers is difficult because it require the algorithms to understand the definition of the numbers (e.g., the definition of IFR) and have the ability to infer the truthfulness of the number associated with the specific concepts (e.g., the possible IFR of COVID-19). We will further explore it in future works.

Table 2. Overall Detection Performance on CoAID

Methods	HAN	PLAN	MVAE	dEFEND	DETERRENT	COVID19-KG	KMGCN	HC-COVID
F1 Score	0.653	0.731	0.688	0.745	0.798	0.761	0.797	**0.820**
Accuracy	0.807	0.846	0.823	0.855	0.885	0.871	0.887	**0.899**
Precision	0.664	0.722	0.685	0.742	0.792	0.802	0.814	**0.826**
Recall	0.642	0.740	0.691	0.748	0.805	0.724	0.780	**0.813**


CDC shows that more people who wear masks become sick with COVID-19 compared to non-mask wearers, therefore masks don't work or are making us ill.

Bill Gates, who has previously funded harmful vaccines in developing countries, plans to use COVID-19 vaccines to surveillance.

The overall lethality of [COVID-19] (IFR) is about 0.1% to 0.3% and thus in the range of a severe influenza.

✗ misleading Score: 0.834
$\mathbb{G}^S$: Mask, prevent, COVID-19
$\mathbb{G}^O$: Mask, no effect on, illness
✗ misleading Score: 0.911
$\mathbb{G}^S$: Bill Gates, is good for, vaccine
$\mathbb{G}^O$: vaccine, no relation with, surveillance
~~✓ true~~ ✗ misleading Score: 0.375
$\mathbb{G}^S$: COVID-19, cause, fever
$\mathbb{G}^O$: COVID-19, have, severe symptom

a) Successful Misinformation Detection by HC-COVID over Other baselines

b) Failure case

Fig. 4. Case Study of COVID-19 Misinformation Detection

Table 3. Overall Detection Performance on CONSTRAINT

Methods	HAN	PLAN	MVAE	dEFEND	DETERRENT	COVID19-KG	KMGCN	HC-COVID
F1 Score	0.750	0.895	0.827	0.868	0.911	0.883	0.910	**0.938**
Accuracy	0.769	0.898	0.830	0.875	0.915	0.886	0.913	**0.939**
Precision	0.788	0.892	0.817	0.887	0.923	0.880	0.912	**0.925**
Recall	0.716	0.897	0.838	0.851	0.899	0.887	0.907	**0.951**

5.3.2 Explainability Performance (Q2). To answer question Q2, we study the explainability performance of the proposed HC-COVID through multiple real-world user studies. In particular, we compare the explainability performance of HC-COVID with the COVID19-KG and DETERRENT baselines which are the only baselines that involve knowledge graphs that can output attention weights to explain the detection results. In the user study, we carry out two sets of experiments using Amazon MTurk. In particular, we randomly select 25 COVID-19 misleading claims and 25 COVID-19 non-misleading claims from the testing set to perform explainability evaluation.

In the first subset of experiments, we study the explainability performance by comparing the quality of the explanations generated from HC-COVID with other schemes. In particular, we define *explainability ranked list* as a list of graph triples retrieved from the knowledge graph based on their attention scores in descending order. For each compared scheme and each COVID-19 claim, we create Top-1, Top-3, and Top-5 explainability ranked list to fully evaluate the explainability performance of each scheme. For each type of explainability ranked list (e.g., Top-1, Top-3, Top-5), we recruit 5 MTurk workers and ask them to select one scheme from all the three compared schemes that can best explain the detection results of each input COVID-19 claim. The explainability performance is evaluated using the following two metrics that are commonly used for quantifying the quality of explanation [41].

- **Percentage of Posts** (% of Posts): the percentage of posts whose explanation picked by the majority workers as their preferred ones belonging to each scheme. For example, given an input COVID-19 claims, if three or more crowd workers believe that the claim is best explained by the knowledge triples from the COVID19-KG scheme. Then we assign COVID19-KG as the best explainable scheme to the claim. If there are totally 10 claims with COVID19-KG, the % of Posts for COVID19-KG is $\frac{10}{50} \times 100\% = 20\%$.
- **Percentage of Workers** (% of Workers): the percentage of workers who select their preferred explanation from the explainability ranked list predicted by each scheme. For example, given an input COVID-19 claim, if there are 3 crowd workers choosing HC-COVID as the best explainable scheme for the claim and 2 crowd workers choosing DETERRENT, we record the number of crowd workers for each scheme. If HC-COVID is finally chosen by 100 crowd works from all 50 claims, the % of Workers for HC-COVID is $\frac{100}{50\times5} = 40\%$.

The above two metrics evaluate the explainability performance of compared schemes from *claim-level* and *worker-level*, respectively. The results are summarized in Figure 6. We observe that HC-COVID significantly outperforms the compared baseline schemes in terms of both metrics. The performance gains demonstrate HC-COVID's capability of generating relevant and accurate explanations by the TGKI module.

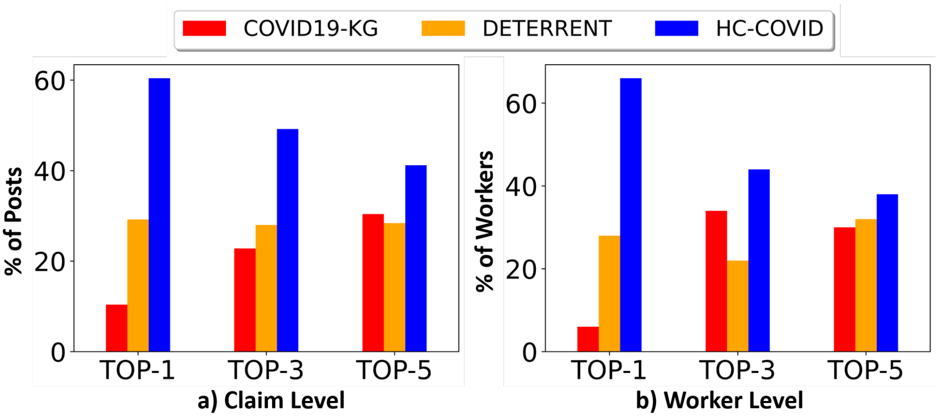


Fig. 5. Explainability Evaluation on CoAID

In the second subset of experiments, we evaluate the explainability performance by investigating the efficiency of the explanations generated by each compared scheme. In particular, we ask 5 MTurk workers to first read through Top-5 explainability ranked list from the first one with the highest attention weights, and stop when a worker thinks the cumulative explanation triples are sufficient to convince the worker on the detection result. The number of explanation triples a worker has read for each post is recorded and is denoted as the *minimum reading index (MRI)*. We them measure the explanation efficiency using the following metrics with respect to MRI:

- **Average Minimum Reading Index** (AvgMRI): The average value of MRI for each compared scheme.
- **Percentage of Posts** (% of Posts): The percentage of posts on which each compared scheme achieves the lowest MRI.

We present the results in Table 4. We observe that HC-COVID achieves the lowest average MRI on both the CoAID and CONSTRAINT datasets compared to the COVID19-KG and DETERRENT

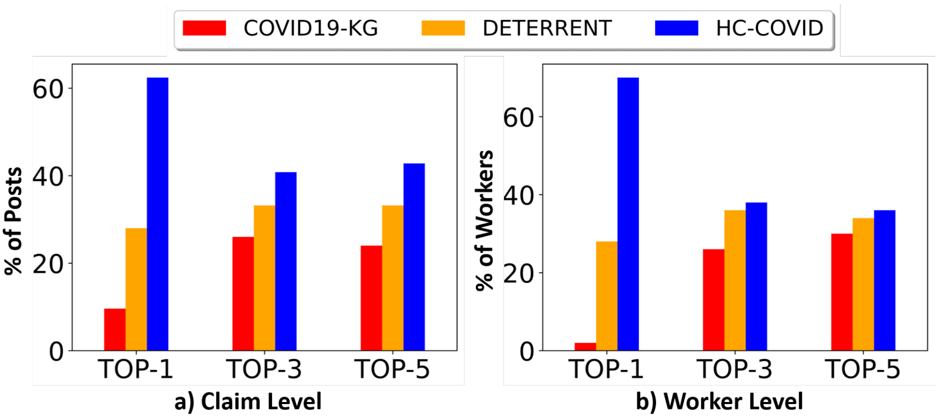


Fig. 6. Explainability Evaluation on CONSTRAINT

schemes. In addition, HC-COVID also explains most of detection results (i.e., 58% on the CoAID dataset and 52% on the DETERRENT dataset) with the lowest number of explanation triples.

Table 4. Evaluation for Minimum Reading Index

Dataset	Metric	COVID19-KG	DETERRENT	HC-COVID
CoAID	AvgMRI	3.32	2.93	**2.25**
	% of Posts	10.0	32.0	**58.0**
CONSTRAINT	AvgMRI	2.94	2.99	**2.48**
	% of Posts	26.0	22.0	**52.0**

5.3.3 Ablation Study (Q3). To answer question Q3, we carry out an ablation study to further investigate the importance of each component in the HC-COVID framework. In particular, we consider three ablations of the HC-COVID framework: i) *HC-COVID\C* that excludes the encoding of claim-level feature into our CHKG by replacing the claim guided adjacency matrix in Equation 4 with binary adjacency matrix; ii) *HC-COVID\T* that excludes the TGKI module from HC-COVID by considering COVID-19 specific and COVID-19 generalized knowledge facts as a homogeneous set of knowledge facts; iii) *HC-COVID\G* that excludes the COVID-19 generalized knowledge facts from the hierarchical knowledge graph. We reported the results of the ablation study in Table 5. We note that HC-COVID achieves the best performance when incorporating all components. In particular, we observe that the COVID-19 generalized knowledge facts significantly contribute to the detection performance of HC-COVID. The reason is that COVID-19 generalized knowledge facts not only contain the summarized knowledge facts about COVID-19 from COVID-19 articles but also allow HC-COVID to identify misleading content that has not appeared in existing COVID-19 articles.

Table 5. Ablation Study for Variants of HC-COVID on CONSTRAINT

	F1 Score	Accuracy	Precision	Recall
HC-COVID	**0.938**	**0.939**	**0.925**	**0.951**
HC-COVID\C	0.920	0.921	0.903	0.938
HC-COVID\T	0.916	0.918	0.905	0.928
HC-COVID\G	0.896	0.895	0.862	0.933

6 DISCUSSION AND FUTURE WORK

We further discuss the possibility of leveraging the HC-COVID framework to solve more general crowdsourcing and machine learning problems. In particular, we elaborate the generalizability of HC-COVID as follows.

- *The generalizability of the hierarchical COVID-19 knowledge graph on COVID-19 related applications:* various COVID-19 related applications can leverage the knowledge facts in our knowledge graph to improve their application-specific performance. In particular, the machine learning based COVID-19 diagnosis approaches [1, 12, 66] can utilize knowledge graph to boost their diagnosis accuracy. For example, the COVID-19 diagnosis approaches usually consider the COVID-19 symptoms (e.g., " fever", "cough") as important features to determine COVID-19 infection of participants. However, it is difficult for the approaches to estimate the complex relations between various COVID-19 symptoms and the COVID-19 infection, especially when the data samples are insufficient. HC-COVID can address the problem by explicitly retrieving COVID-19 knowledge facts from our hierarchical knowledge graph that are relevant to COVID-19 symptoms (e.g., "fever" $\xrightarrow{\text{close relation to}}$ "COVID-19", "cough" $\xrightarrow{\text{cause}}$ "COVID-19"). The COVID-19 symptoms and the corresponding COVID-19 knowledge facts can be integrated into more informative features for more accurate COVID-19 diagnosis.
- *The generalizability of the explainable hierarchical crowdsourcing framework on other knowledge-driven classification problems:* the overall framework of HC-COVID (i.e., explainable graph neural network approach based on the hierarchical crowdsourcing knowledge graph) can be generalized to address different classification problems that require professional knowledge to perform classification and explanation tasks. For example, William *et al.* [6] designed a human-machine system to classify human heart records by assigning expert crowd workers and non-expert workers together to perform the classification task. However, assigning expert workers the same classification task as the non-expert workers is not always effective due to either the lack of available expert workers or the lack of professional medical knowledge of non-expert workers. The framework of HC-COVID can address such a problem by tasking the expert workers to propose generalized medical knowledge facts that are specific to this application (e.g., the characteristics of abnormal heart records) and tasking non-expert workers to identify the abnormal heart records that satisfy the proposed characteristics by the expert workers. A human heart related knowledge graph constructed from the inputs from expert workers can effectively guide the non-expert workers to identify abnormal heart records and provide explicit explanations for the classification results.

Our paper is closely related to the research of crowdsourcing and human-machine systems in CSCW and GROUP communities [4, 6, 17, 22, 31, 44]. For example, Berenberg *et al.* [4] designed a crowdsourcing framework that assigns crowd workers to construct a large-scale topological network. Jiang *et al.* [22] leveraged the uneven difficulty of crowdsourcing tasks and developed a task difficulty identification method to effectively recruit crowd workers. In our paper, we focus on the construction of a hierarchical COVID-19 knowledge graph by exploring the uneven ability and reliability of crowd workers (i.e., expert workers vs. non-expert workers). Park *et al.* [31] proposed a crowdsourcing approach to design AI-based healthcare technologies and outlined the key challenges in future crowd-AI research areas. William *et al.* [6] proposed a hybrid human-machine algorithm to classify human heart records by leveraging both expert and non-expert medical workers. In our paper, we propose a crowdsourcing-based graph neural network approach to detect COVID-19 misinformation on social media, which protects people against misleading COVID-19 claims and helps them make more informed decisions on the topic of COVID-19.

There exist at least two possible strategies for social media users to leverage HC-COVID to digest COVID-19 related information in practice. The first strategy is active misinformation verification where the social media users can actively choose the COVID-19 related posts of their interest and activate HC-COVID to examine the truthfulness of the posts. The second strategy is automatic misinformation monitoring where the social media users can choose to link HC-COVID with their online accounts that allows HC-COVID to automatically check the correctness of all COVID-19 posts in the users' view. For each post in both strategies, the HC-COVID will generate the binary detection results (e.g., misleading or non-misleading) as well as the knowledge triples from the COVID-19 crowdsourcing knowledge graph as explanations of the detection results.

Finally, we identify some limitations and future directions of HC-COVID as below.

- *Misclassification on specific COVID-19 misinformation*: As shown in Figure 4(b), HC-COVID as well as other state-of-the-art schemes all fail to detect the misinformation in the post. The reason is that the algorithms do not have sufficient background knowledge on specific numbers (e.g., 0.1%) related to COVID-19, especially when the definitions of numbers are dynamically changing in different contexts (e.g., the numbers have different semantic meanings if they are associated with different concepts, such as IFR and Infection Rate in social media posts). Therefore, it is an important and challenging task on how to explore the accurate definition of the numeric numbers of a post in our future work of HC-COVID. One possible strategy is to design a crowdsourcing strategy that tasks crowd workers to rephrase numeric content (e.g., 0.3% IFR) to semantic content (e.g., a relatively low IFR). The intuition is that humans are often better than machines at understanding the semantic meanings of numbers by leveraging their background knowledge and the context of the posts [34]. A key challenge in this direction is how to reduce the opinion bias from crowd workers in the numerical content rephrase process (e.g., different crowd workers may propose diverse rephrased semantic contents based on their own background knowledge). Therefore, a novel and effective de-bias method is needed for the crowdsourcing strategy to obtain the accurate rephrased semantic contents.
- *Partial COVID-19 misinformation*: The current HC-COVID is only able to make binary predictions (i.e., misleading or non-misleading) on the input COVID-19 claims. Therefore, it fails to reveal the degree of truthfulness for a given COVID-19 claim. For example, a COVID-19 claim may contain true information in the first sentence but misinformation in the rest of the post. However, we observe that the amount of truthful content in a partially true COVID-19 claim is correlated with the prediction score (e.g., the prediction scores in Figure 4) of our HC-COVID framework. In particular, a higher prediction score is likely to indicate a higher degree of truthfulness for the given COVID-19 claim. In future work, we will further explore the possible quantitative measurements (e.g., correlation coefficients) to accurately estimate the relation between the prediction score and the degree of truthfulness. Moreover, we plan to

explore the possibility of identifying the misleading content from a partially true COVID-19 claim by comparing the degree of the truthfulness of word-level representations with the overall prediction score of the post to identify low-credibility words in the post.

7 CONCLUSION

This paper presents the HC-COVID scheme to address the explainable COVID-19 misinformation detection problem on social media. We design a novel crowdsource hierarchical knowledge graph via the collaboration between expert and non-expert crowd workers to model the COVID-19 specific knowledge and generalized knowledge facts. We also develop a duo hierarchy attention based graph neural network to effectively integrate the COVID-19 knowledge facts contributed by crowd workers to detect misleading COVID-19 claims and explain the detection results. Evaluation results on two real-world datasets of COVID-19 misinformation demonstrate that HC-COVID significantly outperforms the state-of-the-art baselines in terms of both COVID-19 misinformation detection accuracy and explainability.

ACKNOWLEDGMENT

This research is supported in part by the National Science Foundation under Grant No. CHE-2105032, IIS-2008228, CNS-1845639, CNS-1831669. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

REFERENCES

[1] Norah Alballa and Isra Al-Turaiki. 2021. Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: A review. *Informatics in Medicine Unlocked* (2021), 100564.

[2] Jackie Ayoub, X Jessie Yang, and Feng Zhou. 2021. Combat COVID-19 infodemic using explainable natural language processing models. *Information Processing & Management* 58, 4 (2021), 102569.

[3] Michael Azmy, Peng Shi, Jimmy Lin, and Ihab F Ilyas. 2019. Matching entities across different knowledge graphs with graph embeddings. *arXiv preprint arXiv:1903.06607* (2019).

[4] Daniel Berenberg and James P Bagrow. 2018. Efficient crowd exploration of large networks: The case of causal attribution. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–25.

[5] Leticia Bode and Emily K Vraga. 2018. See something, say something: correction of global health misinformation on social media. *Health communication* 33, 9 (2018), 1131–1140.

[6] William Callaghan, Joslin Goh, Michael Mohareb, Andrew Lim, and Edith Law. 2018. Mechanicalheart: A human-machine framework for the classification of phonocardiograms. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–17.

[7] Sonia Castelo, Thais Almeida, Anas Elghafari, Aécio Santos, Kien Pham, Eduardo Nakamura, and Juliana Freire. 2019. A topic-agnostic approach for identifying fake news pages. In *Companion proceedings of the 2019 World Wide Web conference*. 975–980.

[8] Joseph Chee Chang, Saleema Amershi, and Ece Kamar. 2017. Revolt: Collaborative crowdsourcing for labeling machine learning datasets. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2334–2346.

[9] J Lotriet Cornelius. 2012. Reviewing the review process: Identifying sources of delay. *The Australasian medical journal* 5, 1 (2012), 26.

[10] Limeng Cui and Dongwon Lee. 2020. CoAID: COVID-19 Healthcare Misinformation Dataset. arXiv:2006.00885 [cs.SI]

[11] Limeng Cui, Haeseung Seo, Maryam Tabar, Fenglong Ma, Suhang Wang, and Dongwon Lee. 2020. DETERRENT: Knowledge Guided Graph Attention Network for Detecting Healthcare Misinformation. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 492–502.

[12] Andre Filipe de Moraes Batista, Joao Luiz Miraglia, Thiago Henrique Rizzi Donato, and Alexandre Dias Porto Chiavegatto Filho. 2020. COVID-19 diagnosis prediction in emergency care patients: a machine learning approach. *medRxiv* (2020).

[13] Ronald Denaux and Jose Manuel Gomez-Perez. 2020. Linked Credibility Reviews for Explainable Misinformation Detection. In *International Semantic Web Conference*. Springer, 147–163.

[14] Eleftherios Dimitrakis, Konstantinos Sgontzos, and Yannis Tzitzikas. 2020. A survey on question answering systems over linked data and documents. *Journal of Intelligent Information Systems* 55, 2 (2020), 233–259.

[15] Daniel Domingo-Fernández, Shounak Baksi, Bruce Schultz, Yojana Gadiya, Reagon Karki, Tamara Raschka, Christian Ebeling, Martin Hofmann-Apitius, et al. 2020. COVID-19 Knowledge Graph: a computable, multi-modal, cause-and-effect knowledge model of COVID-19 pathophysiology. *BioRxiv* (2020).

[16] Antigoni Founta, Constantinos Djouvas, Despoina Chatzakou, Ilias Leontiadis, Jeremy Blackburn, Gianluca Stringhini, Athena Vakali, Michael Sirivianos, and Nicolas Kourtellis. 2018. Large scale crowdsourcing and characterization of twitter abusive behavior. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 12.

[17] Amira Ghenai and Yelena Mejova. 2018. Fake cures: user-centric modeling of health misinformation in social media. *Proceedings of the ACM on human-computer interaction* 2, CSCW (2018), 1–20.

[18] Qingyu Guo, Fuzhen Zhuang, Chuan Qin, Hengshu Zhu, Xing Xie, Hui Xiong, and Qing He. 2020. A survey on knowledge graph-based recommender systems. *IEEE Transactions on Knowledge and Data Engineering* (2020).

[19] Pei-Yun Hsueh, Prem Melville, and Vikas Sindhwani. 2009. Data quality from crowdsourcing: a study of annotation selection criteria. In *Proceedings of the NAACL HLT 2009 workshop on active learning for natural language processing*. 27–35.

[20] Xiao Huang, Jingyuan Zhang, Dingcheng Li, and Ping Li. 2019. Knowledge graph embedding based question answering. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. 105–113.

[21] Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and S Yu Philip. 2021. A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. *IEEE Transactions on Neural Networks and Learning Systems* (2021).

[22] Yu Jiang, Yuling Sun, Jing Yang, Xin Lin, and Liang He. 2018. Enabling uneven task difficulty in micro-task crowdsourcing. In *Proceedings of the 2018 ACM Conference on Supporting Groupwork*. 12–21.

[23] Dhruv Khattar, Jaipal Singh Goud, Manish Gupta, and Vasudeva Varma. 2019. Mvae: Multimodal variational autoencoder for fake news detection. In *The World Wide Web Conference*. 2915–2921.

[24] Ling Min Serena Khoo, Hai Leong Chieu, Zhong Qian, and Jing Jiang. 2020. Interpretable rumor detection in microblogs by attending to user interactions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 8783–8790.

[25] Ziyi Kou, Lanyu Shang, Yang Zhang, Christina Youn, and Dong Wang. 2021. Fakesens: A social sensing approach to covid-19 misinformation detection on social media. In *2021 17th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE.

[26] Ziyi Kou, Daniel Yue Zhang, Lanyu Shang, and Dong Wang. 2020. ExFaux: A Weakly Supervised Approach to Explainable Fauxtography Detection. In *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 631–636.

[27] Pablo Lara-Navarra, Hervé Falciani, Enrique A Sánchez-Pérez, and Antonia Ferrer-Sapena. 2020. Information management in healthcare and environment: Towards an automatic system for fake news detection. *International journal of environmental research and public health* 17, 3 (2020), 1066.

[28] Linfeng Li, Peng Wang, Jun Yan, Yao Wang, Simin Li, Jinpeng Jiang, Zhe Sun, Buzhou Tang, Tsung-Hui Chang, Shenghui Wang, et al. 2020. Real-world data medical knowledge graph: construction and applications. *Artificial intelligence in medicine* 103 (2020), 101817.

[29] Weijie Liu, Peng Zhou, Zhe Zhao, Zhiruo Wang, Qi Ju, Haotang Deng, and Ping Wang. 2020. K-bert: Enabling language representation with knowledge graph. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 2901–2908.

[30] Yi-Ju Lu and Cheng-Te Li. 2020. GCAN: Graph-aware co-attention networks for explainable fake news detection on social media. *arXiv preprint arXiv:2004.11648* (2020).

[31] Sun Young Park, Pei-Yi Kuo, Andrea Barbarin, Elizabeth Kaziunas, Astrid Chow, Karandeep Singh, Lauren Wilcox, and Walter S Lasecki. 2019. Identifying challenges and opportunities in human-AI collaboration in healthcare. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*. 506–510.

[32] Parth Patwa, Mohit Bhardwaj, Vineeth Guptha, Gitanjali Kumari, Shivam Sharma, Srinivas PYKL, Amitava Das, Asif Ekbal, Shad Akhtar, and Tanmoy Chakraborty. 2021. Overview of CONSTRAINT 2021 Shared Tasks: Detecting English COVID-19 Fake News and Hindi Hostile Posts. In *Proceedings of the First Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situation (CONSTRAINT)*. Springer.

[33] Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2017. Automatic detection of fake news. *arXiv preprint arXiv:1708.07104* (2017).

[34] Martin Potthast, Matthias Hagen, Michael Völske, and Benno Stein. 2013. Crowdsourcing interaction logs to understand text reuse from the web. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1212–1221.

[35] Rohan Ramanath, Florian Schaub, Shomir Wilson, Fei Liu, Norman Sadeh, and Noah Smith. 2014. Identifying relevant text fragments to help crowdsource privacy policy annotations. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 2.

[36] Lida Safarnejad, Qian Xu, Yaorong Ge, Siddharth Krishnan, Arunkumar Bagarvathi, and Shi Chen. 2020. Contrasting Misinformation and Real-Information Dissemination Network Structures on Social Media During a Health Emergency. *American Journal of Public Health* 110, S3 (2020), S340–S347.

[37] Wojciech Samek, Thomas Wiegand, and Klaus-Robert Müller. 2017. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv preprint arXiv:1708.08296* (2017).

[38] Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2017. Modeling Relational Data with Graph Convolutional Networks. arXiv:1703.06103 [stat.ML]

[39] Longxiang Shi, Shijian Li, Xiaoran Yang, Jiaheng Qi, Gang Pan, and Binbin Zhou. 2017. Semantic health knowledge graph: semantic integration of heterogeneous medical knowledge and services. *BioMed research international* 2017 (2017).

[40] William Shiao and Evangelos E Papalexakis. 2021. KI2TE: Knowledge-Infused InterpreTable Embeddings for COVID-19 Misinformation Detection. (2021).

[41] Kai Shu, Limeng Cui, Suhang Wang, Dongwon Lee, and Huan Liu. 2019. defend: Explainable fake news detection. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 395–405.

[42] Kai Shu, Deepak Mahudeswaran, Suhang Wang, and Huan Liu. 2020. Hierarchical propagation networks for fake news detection: Investigation and exploitation. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 14. 626–637.

[43] Kai Shu, Suhang Wang, and Huan Liu. 2019. 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*.

[44] Kinga Skorupska, Manuel Núñez, Wieslaw Kopec, and Radoslaw Nielek. 2018. Older adults and crowdsourcing: Android tv app for evaluating tedx subtitle quality. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–23.

[45] Milan Someswar and Arnab Bhattacharya. 2018. MineAr: using crowd knowledge for mining association rules in the health domain. In *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data*. 108–117.

[46] Shane Storks, Qiaozi Gao, and Joyce Y Chai. 2019. Commonsense reasoning for natural language understanding: A survey of benchmarks, resources, and approaches. *arXiv preprint arXiv:1904.01172* (2019), 1–60.

[47] Victor Suarez-Lledo and Javier Alvarez-Galvez. 2021. Prevalence of health misinformation on social media: systematic review. *Journal of medical Internet research* 23, 1 (2021), e17187.

[48] Catarina Vasconcelos, Renato Lopes Da Costa, Álvaro Lopes Dias, Leandro Pereira, and José Pedro Santos. 2021. Online influencers: healthy food or fake news. *International Journal of Internet Marketing and Advertising* 15, 2 (2021), 149–175.

[49] Emily K Vraga and Leticia Bode. 2017. Using expert sources to correct health misinformation in social media. *Science Communication* 39, 5 (2017), 621–645.

[50] Dong Wang, Tarek Abdelzaher, and Lance Kaplan. 2015. *Social sensing: building reliable systems on unreliable data.* Morgan Kaufmann.

[51] Dong Wang, Tarek Abdelzaher, Lance Kaplan, and Charu C Aggarwal. 2013. Recursive fact-finding: A streaming approach to truth estimation in crowdsourcing applications. In *2013 IEEE 33rd international conference on distributed computing systems*. IEEE, 530–539.

[52] Dong Wang, Tarek Abdelzaher, Lance Kaplan, Raghu Ganti, Shaohan Hu, and Hengchang Liu. 2013. Exploitation of physical constraints for reliable social sensing. In *2013 IEEE 34th Real-Time Systems Symposium*. IEEE, 212–223.

[53] Dong Wang, Md Tanvir Amin, Shen Li, Tarek Abdelzaher, Lance Kaplan, Siyu Gu, Chenji Pan, Hengchang Liu, Charu C Aggarwal, Raghu Ganti, et al. 2014. Using humans as sensors: an estimation-theoretic perspective. In *IPSN-14 proceedings of the 13th international symposium on information processing in sensor networks*. IEEE, 35–46.

[54] Dong Wang, Lance Kaplan, Hieu Le, and Tarek Abdelzaher. 2012. On truth discovery in social sensing: A maximum likelihood estimation approach. In *Proceedings of the 11th international conference on Information Processing in Sensor Networks*. 233–244.

[55] Hongwei Wang, Fuzheng Zhang, Jialin Wang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. 2018. Ripplenet: Propagating user preferences on the knowledge graph for recommender systems. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. 417–426.

[56] Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. 2017. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering* 29, 12 (2017), 2724–2743.

[57] Youze Wang, Shengsheng Qian, Jun Hu, Quan Fang, and Changsheng Xu. 2020. Fake News Detection via Knowledge-driven Multimodal Graph Convolutional Networks. In *Proceedings of the 2020 International Conference on Multimedia Retrieval*. 540–547.

[58] Colby Wise, Vassilis N Ioannidis, Miguel Romero Calvo, Xiang Song, George Price, Ninad Kulkarni, Ryan Brand, Parminder Bhatia, and George Karypis. 2020. COVID-19 knowledge graph: accelerating information retrieval and discovery for scientific literature. *arXiv preprint arXiv:2007.12731* (2020).

[59] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*. 1480–1489.

[60] Daniel Zhang, Dong Wang, Nathan Vance, Yang Zhang, and Steven Mike. 2018. On scalable and robust truth discovery in big data social media sensing applications. *IEEE transactions on big data* 5, 2 (2018), 195–208.

[61] Daniel Zhang, Yang Zhang, Qi Li, Thomas Plummer, and Dong Wang. 2019. Crowdlearn: A crowd-ai hybrid system for deep learning-based damage assessment applications. In *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 1221–1232.

[62] Daniel Zhang, Yang Zhang, Qi Li, and Dong Wang. 2019. Sparse user check-in venue prediction by exploring latent decision contexts from location-based social networks. *IEEE transactions on Big Data* (2019).

[63] Daniel Yue Zhang, Yifeng Huang, Yang Zhang, and Dong Wang. 2020. Crowd-assisted disaster scene assessment with human-ai interactive attention. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 2717–2724.

[64] Daniel Yue Zhang, Lanyu Shang, Biao Geng, Shuyue Lai, Ke Li, Hongmin Zhu, Md Tanvir Amin, and Dong Wang. 2018. Fauxbuster: A content-free fauxtography detector using social media comments. In *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 891–900.

[65] Yuehua Zhao, Jingwei Da, and Jiaqi Yan. 2021. Detecting health misinformation in online health communities: Incorporating behavioral features into machine learning based approaches. *Information Processing & Management* 58, 1 (2021), 102390.

[66] Yazeed Zoabi, Shira Deri-Rozov, and Noam Shomron. 2021. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. *npj digital medicine* 4, 1 (2021), 1–5.

Received July 2021; revised September 2021; accepted October 2021