



# Heterogeneous deep graph convolutional network with citation relational BERT for COVID-19 inline citation recommendation

Tao Dai<sup>a,\*</sup>, Jie Zhao<sup>b</sup>, Dehong Li<sup>b</sup>, Shun Tian<sup>a</sup>, Xiangmo Zhao<sup>c</sup>, Shirui Pan<sup>d</sup>

<sup>a</sup> School of Future Transportation, Chang'an University, Xi'an, Shaanxi 710064, China

<sup>b</sup> School of Economics and Management, Chang'an University, Xi'an, Shaanxi 710064, China

<sup>c</sup> School of Information Engineering, Chang'an University, Xi'an, Shaanxi 710064, China

<sup>d</sup> Faculty of Information Technology, Monash University, Melbourne, Australia

## ARTICLE INFO

### Keywords:

COVID-19 citation recommendation  
Deep graph convolutional network  
Heterogeneous graph  
Citation enhanced BERT  
Text representation learning

## ABSTRACT

The outbreak of COVID-19 brings almost the biggest explosions of scientific literature ever. Facing such volume literature, it is hard for researchers to find desired citation when carrying out COVID-19 related research, especially for junior researchers. This paper presents a novel neural network based method, called citation relational BERT with heterogeneous deep graph convolutional network (CRB-HDGCN), for COVID-19 inline citation recommendation task. The CRB-HDGCN contains two main stages. The first stage is to enhance the representation learning of BERT model for COVID-19 inline citation recommendation task through CRB. To achieve the above goal, an augmented citation sentence corpus, which replaces the citation placeholder with the title of the cited papers, is used to lightly retrain BERT model. In addition, we extract three types of sentence pair according citation relation, and establish sentence prediction tasks to further fine-tune the BERT model. The second stage is to learn effective dense vector of nodes among COVID-19 bibliographic graph through HDGCN. The HDGCN contains four layers which are essentially all sub neural networks. The first layer is initial embedding layer which generates initial input vectors with fixed size through CRB and a multilayer perceptron. The second layer is a heterogeneous graph convolutional layer. In this layer, we expand traditional homogeneous graph convolutional network into heterogeneous by subtly adding heterogeneous nodes and relations. The third layer is a deep attention layer. This layer uses trainable project vectors to reweight the node importance simultaneously according to both node types and convolution layers, which further promotes the performance of learnt node vectors. The last decoder layer recovers the graph structure and let the whole network trainable. The recommendation is finally achieved by integrating the high performance heterogeneous vectors learnt from CRB-HDGCN with the query vectors. We conduct experiments on the CORD-19 and LitCovid datasets. The results show that compared with the second best method CO-Search, CRB-HDGCN improves MAP, MRR, P@100 and R@100 with 21.8%, 22.7%, 37.6% and 21.2% on CORD-19, and 29.1%, 25.9%, 15.3% and 11.3% on LitCovid, respectively.

## 1. Introduction

The COVID-19 pandemic broken out in early 2020 has brought a worldwide public health crisis, and results in an exponentially growth of COVID-19 research papers. According to previous studies, the volume of COVID-19 literature is doubling increasing every two weeks (Brainard, 2020). Facing such large amount of papers, it is hard for researchers to acquire desired cited COVID-19 scholarly papers which meet their requirements. Therefore, it is necessary to provide an efficient an effective COVID-19 citation recommendation service for researchers, especially for who are new to COVID-19 research field.

Many approaches have been proposed to retrieve COVID-19 papers (Fister, Fister, & Fister, 2020; MacAvaney, Cohan, & Goharian, 2020; Wise et al., 2020), and many of them applied pretrained language models (PLMs) such as BERT (Kenton & Toutanova, 2019) which has been widely used in many other tasks. However, there are still two defects for the above methods. Firstly, the used word vectors in most previous studies lack deep understanding of COVID-19 papers. In COVID-19 literature, there are a lot of frequently used specific words which rarely appear in general fields. Only using common papers to train word vectors will result in deviation in the understanding of

\* Corresponding author.

E-mail addresses: [daitao@chd.edu.cn](mailto:daitao@chd.edu.cn) (T. Dai), [zhaojie2012@chd.edu.cn](mailto:zhaojie2012@chd.edu.cn) (J. Zhao), [lidehong\\_derek@chd.edu.cn](mailto:lidehong_derek@chd.edu.cn) (D. Li), [tianshun@chd.edu.cn](mailto:tianshun@chd.edu.cn) (S. Tian), [xmzhao@chd.edu.cn](mailto:xmzhao@chd.edu.cn) (X. Zhao), [Shirui.Pan@monash.edu](mailto:Shirui.Pan@monash.edu) (S. Pan).

<https://doi.org/10.1016/j.eswa.2022.118841>

Received 12 July 2022; Received in revised form 2 September 2022; Accepted 12 September 2022

Available online 17 September 2022

0957-4174/© 2022 Elsevier Ltd. All rights reserved.

COVID-19 information (Andre et al., 2020; MacAvaney et al., 2020; Mass & Roitman, 2020), which further decrease the performance of COVID-19 inline citation recommendation. Secondly, most previous methods solely obtain paper vectors based on their content, and neglect valuable citation relationship information between COVID-19 papers. Citation link is a special data that widely existing in academic literature, which contains latent semantic co-occurrences between papers. Semantic co-occurrences relations has been proved as a vital factor when training PLMs (Gan et al., 2022). For example, a citing sentence contains “coronavirus cross immunity” usually citing papers that contains “antibody-depend enhancement from coronavirus-like”. On the surface, there is only one common key word “coronavirus” between them, but the two sentences actually express nearly similar meaning and are co-occurred with the form of citation relations. Therefore, how to use these citation relationships to further enhance the semantical understanding of PLMs for COVID-19 papers is an important issue for COVID-19 citation recommendation task. Besides the above defects, existing COVID-19 paper retrieval and recommendation methods are difficult to effectively utilize heterogeneous information. Due to the burstiness and severity, the COVID-19 pandemic has evolved into a cross-border crisis and attracted multiple areas of research, including social management, medicine, biology and computer science (Aris-tovnik, Ravšelj, & Umek, 2020). Comparing with common scientific literature data, the information among COVID-19 papers, such as text content, authors and venues, have more character of heterogeneous. For paper and citation recommendation task, the heterogeneous character will bring more obstacles for finding desired citations (Zhu et al., 2021). Therefore, how to fully mine the association between heterogeneous nodes, and distinguish the importance of them to further refine optimal node representation, is a key issue for improving the performance of COVID-19 citation recommendation.

To address the aforementioned problems, we propose a novel deep learning based method, called citation relational BERT with heterogeneous deep graph convolutional network (CRB-HDGCN), for COVID-19 inline citation recommendation task. The proposed method first uses a citation augmented corpus to lightly retrain and fine-tune BERT model. Then the method uses a heterogeneous deep graph convolutional network to obtain reliable vector representations for different nodes. Specifically, we introduce a deep attention layer for convolution network, which is able to distinguish node importance by considering both the convolutional layers and node types simultaneously. Experimental results on CORD-19 and LitCovid datasets show that our CRB-HDGCN model is more effective than other baseline methods.

The main contributions of this paper are summarized as follows.

(1) We propose a novel neural network based citation recommendation method, called CRB-HDGCN, for COVID-19 related papers. CRB-HDGCN is able to fully utilize citation relations and heterogeneous links to obtain reliable vectors for papers, authors and venues related to COVID-19.

(2) We propose a new solution to retrain and fine-tune PLMs by utilizing citation relations among scientific papers, which is benefit PLMs to better understand citing activities.

(3) We expand homogeneous graph convolutional network into heterogeneous deep graph convolutional network. The new network can fully utilize the heterogeneous relations to learn reliable node vectors, and is able to distinguish node importance by considering both the depth of network and the node types simultaneously.

(4) Thorough experimental studies on CORD-19 and LitCovid datasets are carried out to validate the effectiveness of the proposed method.

The rest of the paper are organized as follows. Section 2 reviews related work. The retrain and fine-tune manner of citation relational BERT is introduced in Section 3. Section 4 presents heterogeneous deep graph convolutional network in detail. Section 5 presents the experimental results and analysis. The paper is concluded in Section 6.

## 2. Related work

### 2.1. COVID-19 paper retrieval

As far as we know, there is no research work on COVID-19 citation recommendation. The most closely research work is COVID-19 paper retrieval. Therefore, we briefly overview COVID-19 paper retrieval methods in this subsection. The related previous works of COVID-19 paper retrieval are summarized in Table 1. Since the outbreak of COVID-19 pandemic, scientists worldwide have been carried many works to deal with this global public health crisis, and the research literature about COVID-19 is increasing with an extraordinary rate. In order to further accelerate the research of COVID-19, Allen AI Institute and other top research groups publicly released the COVID-19 open research dataset (CORD-19) (Wang et al., 2020) in March 2020, which covers all current coronavirus research literature and continuously updated weekly. Meanwhile, National Library of Medicine published a lightweight COVID-19 paper dataset LitCovid (Chen, Allot, & Lu, 2021), which is a collection of recently published PubMed articles which are directly related to the 2019 coronavirus. From that on, a variety of COVID-19 paper retrieval approaches have been proposed in the literature. The research team of Amazon released AWS CORD-19 Search,<sup>1</sup> a literature search engine for the field of COVID-19, by using named entity recognition, knowledge graph and topic model technologies. The research team of Google released COVID-19 Research Explorer<sup>2</sup> which uses biomedical field data to train the BERT model (Kenton & Toutanova, 2019), and combined keywords and deep paper vectors to achieve COVID-19 paper search. However, according to the research of Soni and Roberts (2021), the accuracy of these two commercial COVID-19 search engines are not satisfied. Zhang, Gupta, Nogueira, Cho, and Lin (2020) proposed a coronavirus literature retrieval method Neural Covidex based on BioBERT (Lee et al., 2020). Neural Covidex obtains deep vector representation for the query and candidate COVID-19 papers, and establishes paragraph level index to increase the search granularity. Rohatgi et al. (2020) expanded the CORD-19 dataset, and developed a literature search engine COVIDSeer by using SciBERT (Beltagy, Lo, & Cohan, 2019). Wise et al. (2020) combined SciBERT and knowledge graph embedding to obtain paper vector to achieve COVID-19 paper retrieval. Fister et al. (2020) analyzed papers in COVID-19 field by using word frequency information, and then used association rules based on information cartography to obtain COVID-19 papers. Andre et al. (2020) linearly combined TF-IDF vector, BM25 vector and Sentence-BERT (Reimers & Gurevych, 2019) vector to calculate the similarity between query and candidate papers to generate final retrieval results. Ostendorff, Ruas, Blume, Gipp, and Rehm (2020) used chapter title in cited literature as an aspect to measure the similarity between cited literature. They also compared the performance of different pre-training models when applied to COVID-19 paper retrieval. MacAvaney et al. (2020) used BM25 vectors to obtain initial search list, and then used SciBERT to reorder the list so as to obtain the final COVID-19 retrieve list. Ahamed and Samad (2020) applied a betweenness centrality measurement algorithm in social networks field, and used keywords related to drugs, diseases, pathogens, hosts of pathogens, and biomolecules to construct a keywords graph for COVID-19 field, which can be used for fine-grained COVID-19 paper retrieval. Nguyen, Rybinski, Karimi, and Xing (2020) proposed a hybrid COVID-19 literature retrieval method. The method first combines the BM25 similarity and cosine similarity between the query and the candidate literature. Then it uses divergence from randomness similarity to reorder the retrieval list, and finally applied BioBERT vectors to sort the retrieval list again to obtain the final results. Xiong et al. (2020) mitigated the domain discrepancy and label scarcity problems by the

<sup>1</sup> <https://cord19.aws/>

<sup>2</sup> <https://covid19-research-explorer.appspot.com/biomedexplorer/>

**Table 1**  
Related works summarization for COVID-19 paper retrieval.

	Title	Date	Using techniques
COVID-19 academic datasets	CORD-19: The COVID-19 open research dataset.	2020	
	LitCovid: An open database of COVID-19 literature.	2020	
Commercial searching engines	Amazon AWS CORD-19 search.	2020	Knowledge Graph; Topic model
	Google COVID-19 research explorer.	2020	BERT; Hybrid Neural model
COVID-19 paper retrieve methods	Rapidly deploying a neural search engine for the COVID-19 open research dataset.	2020	BioBERT; BM25; Paragraph-level index
	Covidseer: Extending the CORD-19 dataset.	2020	TF-IDF; SciBERT
	COVID-19 knowledge graph: accelerating information retrieval and discovery for scientific literature.	2020	Knowledge graph; SciBERT
	Discovering associations in COVID-19 related research papers.	2020	Association rule; Information cartography
	Co-search: COVID-19 information retrieval with semantic search, question answering, and abstractive summarization.	2020	TF-IDF; BM25; Sentence-BERT
	Aspect-based document similarity for research papers.	2020	PLMs
	SLEDGE-Z: A zero-shot baseline for COVID-19 literature search.	2020	BM25; SciBERT
	Information mining for COVID-19 research from a large volume of scientific literature.	2020	Betweenness centrality algorithm
	Searching scientific literature for answers on COVID-19 questions.	2020	BM25; BioBERT
	CMT in TREC-COVID round 2: mitigating the generalization gaps from web to special domain search.	2020	Domain-adaptive pretraining; SciBERT
	Ad-hoc document retrieval using weak supervision with BERT and GPT2.	2020	BERT; GPT2
	Covrexlex: A covid-19 retrieval system with relation extraction.	2021	BERT; named entity
	COVID-19 literature mining and retrieval using text mining approaches.	2022	BOW; TF-IDF; Word2Vec; BERT

domain-adaptive pretraining and few-shot learning technologies, and integrated dense retrieval to alleviate vocabulary mismatch obstacle of sparse retrieval. [Mass and Roitman \(2020\)](#) proposed a COVID-19 literature retrieval method based on weakly supervised learning. The method first fine-tunes BERT and GPT ([Radford et al., 2019](#)) models with title-abstract relationship, and then integrates multiple similarity to generate output the final paper retrieval results according to a comprehensive score. [Tran et al. \(2021\)](#) proposed a COVID-19 paper retrieval system from extracted entities and relations. They use BERT to encode each relation, and score them using pointwise mutual information. [Uday, Pavani, Lakshmi, and Chivukula \(2022\)](#) combined vectors from bag of words, Word2Vec, BERT and TF-IDF to find relevant COVID-19 documents. The main defect of existing COVID-19 paper retrieval methods is the retraining process of PLMs totally neglect citation information, which will result in lacking deep understand for citing activities. Moreover, they all neglect valuable heterogeneous information such as author and venue, which are vital for inline citation recommendation task. Our method solves the first defect by expanding original citing sentence with cited titles, which subtly infuses citation relations into PLMs without breaking its originally training procedure. We also use several sentence relations extracted from citation relations to further enhance the PLMs comprehension ability for citing activities. The second defect is solved by introducing HDGCN which is easy to integrate heterogeneous information when learning the high level representations for COVID-19 corpus, which can essentially improve the performance of COVID-19 inline recommendation.

## 2.2. Inline citation recommendation

In the academic landscape, the rapid rise of big scholarly data (BSD) brings about new issues and challenges with respect to data management and analysis ([Xia, Wang, Bekele, & Liu, 2017](#)). Exploring BSD can provide great benefits for various stakeholders, including community detection ([Mercorio, Mezzananza, Moscato, Picariello, & Sperli, 2019](#)), academic social network analysis ([Kong, Shi, Yu, Liu,](#)

[& Xia, 2019](#)), scientific impact evaluation ([Kong et al., 2020](#)), etc. Citation recommendation is one of key application for BSD which aims to provide researchers a list of references satisfying their citing expectations. According to different usages, citation recommendation can be divided into global recommendation and inline recommendation. The global citation recommendation ([Ayala-Gomez, Daróczy, Benczúr, Mathioudakis, & Gionis, 2018](#); [Dai et al., 2021](#); [Zhu et al., 2021](#)) returns the reference that are relevant to the entire query paper. In contrast, inline citation recommendation, also called context-aware citation recommendation, only provides references for a citation placeholder according to its local context. The work of this paper belongs to context-aware citation recommendation. [He, Pei, Kifer, Mitra, and Giles \(2010\)](#) first introduced the concept of context-aware citation recommendation in 2010. From then on, many context-aware citation recommendation methods have been proposed in the literature ([Dai et al., 2021](#); [Duma, Liakata, Clare, Ravenscroft, & Klein, 2016](#); [Ebesu & Fang, 2017](#); [Färber, Thiemann, & Jatowt, 2018](#); [Gu, Gao, & Hahnloser, 2022](#); [He, Kifer, Pei, Mitra, & Giles, 2011](#); [Huang, Wu, Liang, Mitra, & Giles, 2015](#); [Huang, Wu, Mitra, & Giles, 2014](#); [Jiang, Yin, Gao, Lu, & Liu, 2018](#); [Khadka & Knoth, 2018](#); [Peng, Liu, & Lin, 2016](#); [Totti, Mitra, Ouzzani, & Zaki, 2016](#); [Yin & Li, 2017](#); [Zhang & Ma, 2022](#)). Recently, neural networks have been widely used in context-aware citation recommendation field due to its promising performance. [Huang et al. \(2015\)](#) presented the first neural network based context-aware citation recommendation method. The method is formed as a multi-layer neural network to predict a cited scientific paper by given a query citation context. [Ebesu and Fang \(2017\)](#) proposed a context-aware citation recommendation method by using neural encoder-decoder architecture. The encoder part of the model is based on max time delay neural network (TDNN) which encode a query citation context into a hidden vector. The decoder is based on recurrent neural network (RNN) to recover the hidden vector into the title of cited paper. The method also considers author information when generating hidden vectors, which is proved to be effective for recommendation. [Yin and Li \(2017\)](#) proposed a personalized convolutional neural networks (PCNN) to recommend citations when given a citation context. PCNN also considers

**Table 2**  
Related works summarization for citation recommendation.

	Title	Date	Using techniques
Global citation recommendation	Global citation recommendation using knowledge graphs.	2018	Knowledge graph; Learning to rank
	Recommending scientific paper via heterogeneous knowledge embedding based attentive recurrent neural networks.	2021	Graph learning; Bi-LSTM
	Gated relational stacked denoising autoencoder with localized author embedding for global citation recommendation.	2021	SDAE; attention mechanism
Inline citation recommendation	Context-aware citation recommendation.	2010	Non-parametric probabilistic model
	A neural probabilistic model for context based citation recommendation.	2015	Multi-layer neural network
	Neural citation network for context-aware citation recommendation.	2017	Time delay neural network; RNN
	Personalized citation recommendation via convolutional neural networks.	2017	CNN
	Using citation-context to reduce topic drifting on pure citation-based recommendation.	2018	Topic model; Word2Vec
	Attentive stacked denoising autoencoder with Bi-LSTM for personalized context-aware citation recommendation.	2019	SDAE; Bi-LSTM; attention mechanism
	Local citation recommendation with hierarchical-attention text encoder and SciBERT-based reranking.	2022	Sci-BERT; attention mechanism
	Dual attention model for citation recommendation with analyses on explainability of attention mechanisms and qualitative experiments.	2022	Neural model; Dual attention mechanism

author information in the input layer. Khadka and Knoth (2018) tried to avoid topic drifting problem in citation recommendation by combining citation context with citation position. Dai, Zhu, Wang, and Carley (2019) propose a context-aware citation recommendation method ASL by combining stacked denoising autoencoders (SDAE) and Bi-LSTM. For a scientific paper, they use attention mechanism to extend vanilla SDAE architecture into attentive SDAE (ASDAE). They also extended Bi-LSTM with a local attention layer to learn the hidden vectors for citation context, which can not only capture key information for effective embedding, but also benefit for extracting suitable citation context. Gu et al. (2022) used a hierarchical attention network to excavate the text embeddings of cited papers, and use these embedding to prefetch a limited number of relevant documents. The used hierarchical attention network is coupled with a SciBERT reranker fine-tuned on context-aware citation recommendation task. Zhang and Ma (2022) proposed a embedding-based neural network called dual attention model for citation recommendation task. The model comprises two attention mechanisms: self-attention and additive attention. The self-attention aims to capture the relatedness between the contextual words and structural context, and additive attention aims to learn importance of them. The two attention mechanisms is able to interpret “relatedness” and “importance” through the learned attention weights. For readers to quick check the precious works of citation recommendation, we list them in Table 2.

Due to the reason of generality, the existing inline citation recommendation methods are all focus on common scientific literature, which lead to poor vector performance for COVID-19 domain specific inline citation recommendation task. Our method solves the above problem by using citation relational BERT that retrained and fine-tuned on citation relations among COVID-19 corpus, which improves the domain representation of vectors while fitting inline citation recommendation task. Moreover, HDGCN in our method expands the learning ability of conventional GCN. It can be easily expanded with various heterogeneous information. Moreover, the deep attention layer in HDGCN makes it can be applied into large graph when carrying out multiple deep graph convolution, which is also benefit for common citation recommendation task.

### 3. Citation relational BERT on COVID-19 papers

Contextualised representations of pretrained language models (PLMs) like ELMo (Peters, Neumann, Iyyer, Gardner, & Zettlemoyer, 2018),

BERT (Kenton & Toutanova, 2019), and GPT (Radford et al., 2019) have accomplished promising results in many natural language processing (NLP) tasks. Especially, Hugging Face provided a bert-small-cord19 model<sup>3</sup> which uses a subset of CORD-19 dataset to fine-tune vanilla BERT. Chen, Du, Allot, and Lu (2022) presented LitMC-BERT<sup>4</sup> which trains BioBERT (Lee et al., 2020) on LitCovid dataset for multi-label classification task. Nevertheless, these PLMs still neglect vital relations between COVID-19 papers, which may results in lacking understanding citing activities in papers. In this section, we use citation relations among COVID-19 papers to lightly retrain and fine-tune BERT to let PLM is able to generate more reliable word vectors that more suitable for citation recommendation task. We call the new BERT model as citation relational BERT (CRB). For CORD-19 dataset, the retraining and fine-tuning process is carried on bert-small-cord19 model. For LitCovid dataset, this process is carried on the back bone of LitMC-BERT. The brief retraining and fine-tuning process is demonstrated in Fig. 1.

#### 3.1. Retraining language model with augmented citing sentences

Existing BERT based COVID-19 paper retrieval and recommendation methods all simply delete citation information in papers. Inspired by the work of Huang et al. (2012), we believe the appearance of references has its specific meaning which may conceal particular latent information for citation recommendation task. The latent semantic information can be revealed in two aspects. The first one is the particular citing places. When an author want to cite a paper, written a citing sentence is the pre-order action. According to the research of Qazvinian and Radev (2010), citing sentences usually have their own particular syntax compared with non-citing sentences, and different pattern of citing sentences may existed in different citing places. The second one is the deep semantic of citing sentences. Subjected to the length of papers, authors usually briefly introduce the work of cited papers, and present the reference for readers to further understand related works. This means a citing sentence contains much more deep semantic information that beyond the citing sentence itself. Therefore, both from the syntax and semantic level, simply neglecting citations in

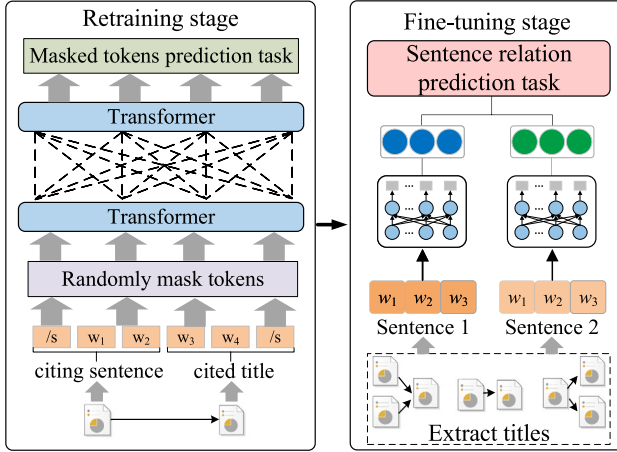
<sup>3</sup> <https://huggingface.co/NeuML/bert-small-cord19>

<sup>4</sup> <https://ftp.ncbi.nlm.nih.gov/pub/lu/LitCovid/litmcbert/>



**Table 3**  
An example construction of citation augmented corpus.

Original citing sentence	Augmented citing sentence
These structures defined a novel family of enzymes, now called sedolins or serine-carboxyl peptidases, that is characterized by the utilization of a fully conserved catalytic triad (Ser, Glu, Asp) and by the presence of an Asp in the oxyanion hole [8].	These structures defined a novel family of enzymes, now called sedolins or serine-carboxyl peptidases, that is characterized by the utilization of a fully conserved catalytic triad (Ser, Glu, Asp) and by the presence of an Asp in the oxyanion hole, <b>Aorsin, a novel serine proteinase with trypsin-like specificity at acidic pH.</b>



**Fig. 1.** The brief retraining and fine-tuning process of CRB.

text body will loss these latent information, and result in deviation in understanding citation activities.

To further explore the latent information in citations, we construct a special citation augmented corpus with COVID-19 related papers. The construction manner is demonstrated in Table 3. As can be seen, we regard the citation placeholder in the original citing sentence as a special fragment, and replace the citation placeholder with the title of the cited papers to form an augmented citation sentence. As can be seen, for one hand, the added title is among the citing sentence, which does not break the structure context of other non-citing sentences. On the other hand, the added title provides a further supplement for the citing sentence, and enriches its meaning. The added title reorganizes the written logical of citing sentence, which makes the original paper presents a new language structure.

We select 1,861,412 and 235,421 citing sentences from CORD-19 and LitCovid datasets respectively, and use them to form augmented corpus. Then we use all these augmented sentences to retrain BERT by applying the Masked Language Model (MLM) objective. The model is retrained for 100 epochs in batches of 64 samples by employing the Adam with learning rate 5e-5.

### 3.2. Fine-tuning language model with additional link information

To enable BERT further effectively understand the citing relations between COVID-19 papers, we further use sentence relation prediction task to lightly fine-tune it. We first extract three types of sentences pair, including: (1) the title of citing and cited papers; (2) the title of papers that cite same paper; (3) the title of papers that are cited by same paper. After obtaining the above three sentence pairs, we use retained BERT in previous subsection to predict whether there is correlation between them to execute fine-tune process. For example, suppose we have sentence  $s_1$  and  $s_2$  which are the titles of citing and cited papers, we send them into retained BERT and obtain all word

vectors of them. Then we sum all word vectors and calculate mean vectors to get sentence vectors  $v_{s_1}$  and  $v_{s_2}$ . After that, we calculate the relevance score of  $s_1$  and  $s_2$  by  $\hat{c}_{s_1, s_2} = \text{sigmod}(v_{s_1} v_{s_2}^T)$ , and use the following cross entropy loss to execute fine-tune process.

$$L = - \sum_{i=1}^{N_s} c_{s_1, s_2} \log(\hat{c}_{s_1, s_2}) \quad (1)$$

where  $N_s$  is the overall number of sentence pairs,  $c_{s_1, s_2}$  is 1 if  $s_1$  and  $s_2$  is a selected sentence pair.

The main novelty of CRB is we infuse the learning of citation relations by a masked tokens prediction task, which can be easily carried on by partial original training process of BERT model. Such a practice is not like common approaches that heavily retrain BERT model with full COVID-19 corpus following whole original training process (Kenton & Toutanova, 2019; Müller, Salathé, & Kummervold, 2020), which totally ignore the citation relations. It is also not like other solutions that consider relation prediction as a downstream task to fine-tune BERT model (Andre et al., 2020; Mass & Roitman, 2020), which cannot acquire fine-grained semantic co-occurrences among citation relations.

## 4. Heterogeneous deep graph convolutional network for COVID-19 paper recommendation

In previous section, we have obtained CRB by excavating latent citation information to obtain more robust word vectors. To achieve promising citation recommendation for COVID-19 papers, we still need to obtain reliable paper vectors. Moreover, other related items, such as authors and venues with their correlations, are also need to be utilized. In this section, we present heterogeneous deep graph convolutional network to achieve the above goal. The detailed architecture of proposed network are shown in Fig. 2, which contains initial embedding layer, heterogeneous graph convolutional layer, deep attention layer and decoder layer.

The main novelties of our HDGCN are reflected in two aspects. The first one is we extend homogeneous GCN (Jeong, Jang, Park, & Choi, 2020) into heterogeneous by subtly expanding the normalized Laplacian with various nodes and relations. The second one is the deep attention layer in HDGCN is able to promote the importance of long distance heterogeneous nodes when stacking plenty convolution layer, thus relieves over-smoothing issue (Chen et al., 2020; Li, Han, & Wu, 2018; Liu, Gao, & Ji, 2020; Xu et al., 2018) of deep GCN in heterogeneous condition.

### 4.1. Initial embedding layer

The initial embedding layer is aim to embed raw words into vectors. For  $n$ th COVID-19 paper  $P_n$ , we send its all sentences into CRB to obtain all word vectors. Then we calculate the mean vectors from obtained word vectors to get paper vector  $v_{P_n} \in \mathbb{R}^{d_B}$ . Then we use a multilayer perceptron (MLP) to embed it into a lower dimension vector  $X_{P_n} \in \mathbb{R}^d$

$$X_{P_n} = L_2(\text{Relu}(L_1(v_{P_n}))) \quad (2)$$

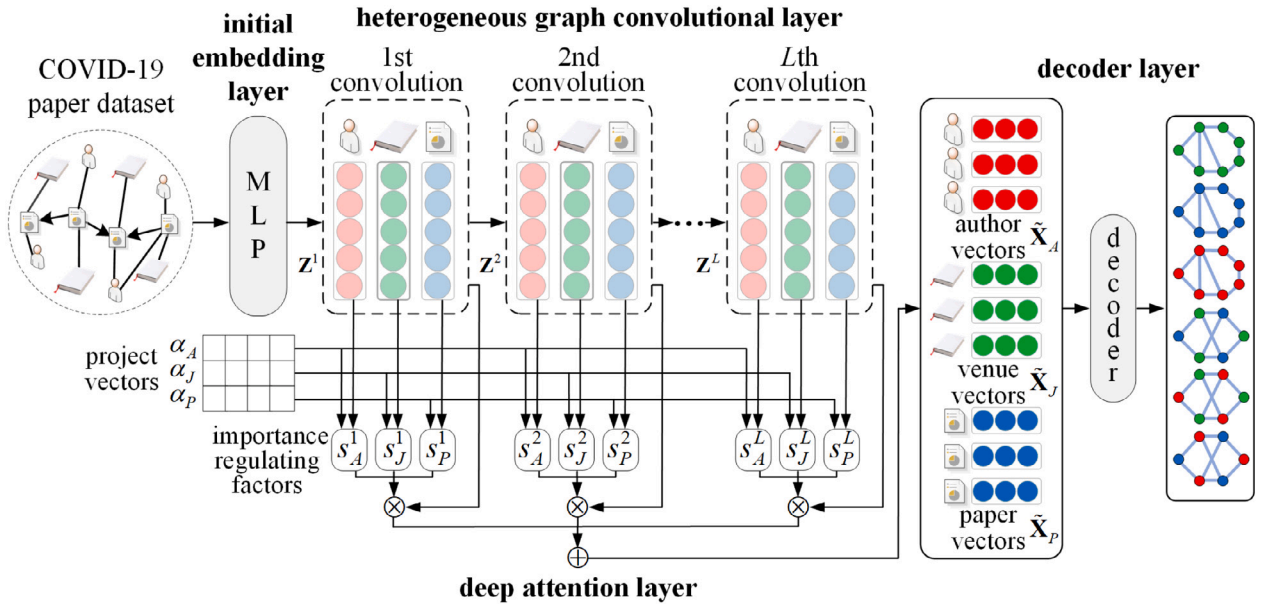


Fig. 2. The detailed architecture of heterogeneous deep graph convolutional network. There are essentially four layers in the network. The initial embedding layer firstly transfers nodes in COVID-19 paper dataset into vectors with fixed size by a multilayer perceptron. Secondly, the heterogeneous graph convolutional layer embeds the initial vectors into  $L$  heterogeneous matrices  $Z^1, Z^2, \dots, Z^L$  by utilizing multiple relations among them. Thirdly, for each  $l$ th vector matrices  $Z^l$ , the deep attention layer utilizes three learnable project vectors  $\alpha_A, \alpha_J$  and  $\alpha_P$  to extract importance regulating factors  $s_A^l, s_J^l$  and  $s_P^l$ , respectively. The final vectors are then generated by combining all reweighted vectors in all convolutional layers. Finally, the decoder layer recovers the network structure from obtained vectors to ensure the trainability of the whole network.

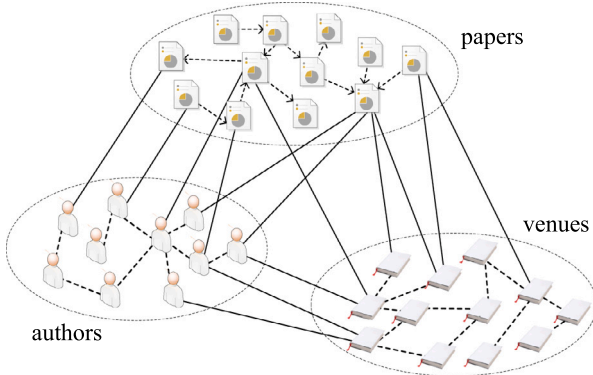


Fig. 3. A demonstration graph which illustrates the involved relationships and nodes in bibliographical heterogeneous graph.

where  $L_2(x) = \mathbf{W}_i x + b_i$  is a fully connected linear function. The layer dimension in MLP is set to  $d_s$ .

After getting initial vectors for all papers, the next step is to obtain initial vectors of authors and venues. For  $m$ th author  $A_m$  and  $r$ th venue  $J_r$ , we obtain their initial vectors  $X_{A_m} \in \mathbb{R}^{d_s}$  and  $X_{J_r} \in \mathbb{R}^{d_s}$  as follows.

$$X_{A_m} = \text{avg}_{P_i \in P_{A_m}} (X_{P_1}, X_{P_2}, \dots, X_{P_{|P_{A_m}|}}) \quad (3)$$

$$X_{J_r} = \text{avg}_{P_i \in P_{J_r}} (X_{P_1}, X_{P_2}, \dots, X_{P_{|P_{J_r}|}}) \quad (4)$$

where  $\text{avg}(\cdot)$  is vector mean function.  $P_{A_m}$  represents all papers published by author  $A_m$ , and  $P_{J_r}$  represents all papers included in venue  $J_r$ .

#### 4.2. Heterogeneous graph convolutional layer

In previous subsection, we have obtained initial vector representations COVID-19 related papers, authors and venues. These vectors

only contain information based on text content, and neglect vital link relationship between them. In this subsection, we will utilize link relationships to build the connection between these vectors, and use heterogeneous graph convolutional layer to further boost the performance of learnt vectors.

We first represent COVID-19 paper dataset as a heterogeneous graph  $G = \langle \mathbf{X}, \mathbf{R} \rangle$ , where  $\mathbf{X} = \{X_P, X_A, X_J\}$  denotes node vectors of papers, authors and venues. The value of  $\mathbf{X}$  is obtained from the initial embedding layer in previous subsection. According to different types of nodes, edge set  $\mathbf{R} = \{R_{PP}, R_{PA}, R_{PJ}, R_{AA}, R_{AJ}, R_{JJ}\}$  contains 6 types of relations, including 3 types of homogeneous relations and heterogeneous relations: paper citation relations  $R_{PP}$ , co-author relations  $R_{AA}$ , venue relations  $R_{JJ}$ , author writing relations  $R_{PA}$ , paper publication relations  $R_{PJ}$ , and author publication relations  $R_{AJ}$ . Besides  $R_{JJ}$ , other relations have explicit meaning and are easily to be established. For example, if author  $A_i$  published a paper in venue  $J_j$ , then the value of  $R_{A_i J_j}$  is 1; otherwise 0. The  $R_{JJ}$  is calculated according to the cosine similarity between venue vectors. We select the top 30% nodes as semantic neighbors for each venue. Then we set the value to 1 if a venue is a semantic neighbor of an other venue. A demonstration graph is presented in Fig. 3 which illustrates the involved relationships and nodes in heterogeneous bibliographical graph  $G$ .

Based on  $\mathbf{R}$ , we construct an augmented matrix  $\mathbf{R}^*$  as follows.

$$\mathbf{R}^* = \begin{bmatrix} \mathbf{R}_{PP} & \mathbf{R}_{PA} & \mathbf{R}_{PJ} \\ \mathbf{R}_{AP} & \mathbf{R}_{AA} & \mathbf{R}_{AJ} \\ \mathbf{R}_{JP} & \mathbf{R}_{JA} & \mathbf{R}_{JJ} \end{bmatrix} \quad (5)$$

The next step is to send node vectors  $\mathbf{X}$  into heterogeneous graph convolutional network. The network contains multiply layers. The  $l$ th layer is executed as follows.

$$\mathbf{Z}^{(l+1)} = \text{Relu}(\tilde{\mathbf{D}}^{(-1/2)} \tilde{\mathbf{R}}^* \tilde{\mathbf{D}}^{(-1/2)} \mathbf{Z}^l \mathbf{W}^l) \quad (6)$$

where  $\text{Relu}(\cdot)$  is the rectifier activation function.  $\mathbf{W}^l \in \mathbb{R}^{d_s \times d_s}$  is a learnable convolution kernel.  $\tilde{\mathbf{R}}^* = \mathbf{R}^* + \mathbf{I}$ , and  $\tilde{\mathbf{D}}_{ii} = \sum_i \tilde{\mathbf{R}}_{ii}^*$ .  $\mathbf{Z}^l \in \mathbb{R}^{(N_P + N_A + N_J) \times d_s}$  represents the node representation in  $l$ th layer.  $N_P, N_A$  and  $N_J$  are the overall number of papers, authors and venues in dataset, respectively. Particular,  $\mathbf{Z}^0$  is formed by the initial node vectors. The

exactly construction of  $\mathbf{Z}^0$  and  $\mathbf{Z}^l$  are as follows.

$$\mathbf{Z}^0 = \begin{bmatrix} \mathbf{X}_P \\ \mathbf{X}_A \\ \mathbf{X}_J \end{bmatrix} \quad \mathbf{Z}^l = \begin{bmatrix} \mathbf{X}_P^l \\ \mathbf{X}_A^l \\ \mathbf{X}_J^l \end{bmatrix} \quad (7)$$

where  $\mathbf{X}_P^l$ ,  $\mathbf{X}_A^l$  and  $\mathbf{X}_J^l$  are the paper, author and venue vectors in  $l$ th layers, respectively.

#### 4.3. Deep attention layer

Through the layered by layered manner, graph convolutional network is able to incorporate heterogeneous correlations between nodes to improve the performance of node vectors. However, there is a serious problem for a deep GCN: each layer more considers immediate neighbors, and the performance decreases when the whole network goes deeper. Several recent studies attribute this performance deterioration to the over-smoothing issue, which states that repeated propagation makes node representations of different classes indistinguishable (Chen et al., 2020; Li et al., 2018; Xu et al., 2018). Liu et al. (2020) propose a solution to address the above problem. Nevertheless, their solution only considered homogeneous graphs. In a scholar network, different types of nodes obviously plays different roles when the author want to cite a paper. For example, one researcher cited a paper because it was published in an influential venue, while another researcher cited the paper because it was the work of a coauthor. This enlightens us that we need to develop a mechanism to distinguish node importance by considering both the depth of network and the type of nodes simultaneously. We use the deep attention layer in this subsection to achieve the above goal.

The core of deep attention layer is three trainable project vectors  $\{\alpha_P, \alpha_A, \alpha_J\} \in \mathbb{R}^{d_\alpha}$ , which are used to learn the importance of papers, authors and venues in different convolutional layers. As shown in Fig. 2, we stack output vectors in all layers according to different type, thereby obtaining three tensors  $\mathbf{X}_P^* \in \mathbb{R}^{N_P \times L \times d_s}$ ,  $\mathbf{X}_A^* \in \mathbb{R}^{N_A \times L \times d_s}$  and  $\mathbf{X}_J^* \in \mathbb{R}^{N_J \times L \times d_s}$  for papers, authors and venues, respectively. Then we use the following process to generate importance regulating factors for all  $L$  layers.

$$\begin{bmatrix} s_P \\ s_A \\ s_J \end{bmatrix} = \text{softmax}(\sigma \begin{bmatrix} \mathbf{X}_P^* \mathbf{W}_\alpha \alpha_P \\ \mathbf{X}_A^* \mathbf{W}_\alpha \alpha_A \\ \mathbf{X}_J^* \mathbf{W}_\alpha \alpha_J \end{bmatrix}) \quad (8)$$

where  $\sigma(\cdot)$  represents sigmoid function.  $\mathbf{W}_\alpha \in \mathbb{R}^{d_s \times d_\alpha}$  is a learnable mapping matrix.  $s_P \in \mathbb{R}^{N_P \times L}$ ,  $s_A \in \mathbb{R}^{N_A \times L}$ ,  $s_J \in \mathbb{R}^{N_J \times L}$  are generated importance regulating factors for paper vectors, author vectors and venue vectors in all layers.

Next, we use the obtained importance regulating factors to reweight all output vectors to obtain final vector representation  $\tilde{\mathbf{X}}$ .

$$\tilde{\mathbf{X}} = \begin{bmatrix} \tilde{\mathbf{X}}_P \\ \tilde{\mathbf{X}}_A \\ \tilde{\mathbf{X}}_J \end{bmatrix} = \begin{bmatrix} (\mathbf{X}_P^1 \otimes s_P^1) \oplus (\mathbf{X}_P^2 \otimes s_P^2) \oplus \dots \oplus (\mathbf{X}_P^L \otimes s_P^L) \\ (\mathbf{X}_A^1 \otimes s_A^1) \oplus (\mathbf{X}_A^2 \otimes s_A^2) \oplus \dots \oplus (\mathbf{X}_A^L \otimes s_A^L) \\ (\mathbf{X}_J^1 \otimes s_J^1) \oplus (\mathbf{X}_J^2 \otimes s_J^2) \oplus \dots \oplus (\mathbf{X}_J^L \otimes s_J^L) \end{bmatrix} \quad (9)$$

where  $\otimes$  represents Hadamard product, and  $\oplus$  denotes obtaining the sum mean vector.

#### 4.4. Decoder layer

In order to ensure the effectiveness of obtained vectors, we put  $\tilde{\mathbf{X}}$  into a decoder layer to recover the link information of heterogeneous graph  $G$  with probability  $P(\mathbf{R}^*|\tilde{\mathbf{X}})$ , which equals to maximum the following likelihood function.

$$\max \prod_{i=1}^{N_s} \prod_{j=1}^{N_s} P(\mathbf{R}_{ij}^*|\tilde{\mathbf{X}}_i, \tilde{\mathbf{X}}_j) \quad (10)$$

where  $N_s = N_P + N_A + N_J$  is the overall number of nodes.  $\tilde{\mathbf{X}}_i$  and  $\tilde{\mathbf{X}}_j$  represent any two node vectors in  $\tilde{\mathbf{X}}$ .  $\mathbf{R}_{ij}^*$  is the entry value of the augmented matrix  $\mathbf{R}^*$  related to  $i$ th node and  $j$ th node.

#### 4.5. COVID-19 paper recommendation

When finishing training, we can use our method to recommend COVID-19 citations. Given a citation context  $q_w$  with query author  $q_A$  and query venue  $q_J$ , we first send them into CRB, and then use initial embedding layer in Section 4.1 to obtain query vectors  $v_q = [v_q^w, v_q^A, v_q^J]$ , where  $v_q^w \in \mathbb{R}^{d_s}$  is the mean vector of  $q_w$ ,  $v_q^A \in \mathbb{R}^{d_s}$  is the mean vector of the papers published by the provided author and  $v_q^J \in \mathbb{R}^{d_s}$  is the mean vector of the papers contained in the provided venue. Then we calculate the score  $s_{q, P_i}$  for each candidate paper  $P_i$  as follows.

$$s_{q, P_i} = v_q^w (\tilde{\mathbf{X}}_{P_i})^T + v_q^A (\tilde{\mathbf{X}}_{A_{P_i}})^T + v_q^J (\tilde{\mathbf{X}}_{J_{P_i}})^T \quad (11)$$

where  $\tilde{\mathbf{X}}_{P_i}$ ,  $\tilde{\mathbf{X}}_{A_{P_i}}$  and  $\tilde{\mathbf{X}}_{J_{P_i}}$  are the paper vector, author vector and venue of candidate paper  $P_i$ . The final recommendation list is generated according to  $s_{q, P_i}$  with the higher score.

#### 4.6. A running example for HDGCN

To let readers understand the process of our HDGCN more clearly, we present a running example of a HDGCN with 2 convolutional layers. As shown in the left part of Fig. 4, the input heterogeneous graph is consist of three types of nodes. We use round, diamond and pentagram to represent paper, author and venue, respectively. According to the degree of nodes, the input graph can be divided into three circles. We only observe the vector changing of nodes in the inner circle.

The initial vectors are shown in the top part of Fig. 4. As can be proved in Li et al. (2018), a single layer convolution means a propagation process which makes nodes gathering neighborhood information, and more layers of convolution will involve neighbors that have longer distance. Therefore, as can be seen in the middle part of Fig. 4, when carrying out 1 layer heterogeneous graph convolution, the propagation mechanism makes the nodes in the inner circle gathering their neighborhood information. The colorless nodes denote the neighbors that have not been gathered, while purple nodes is the neighbors that have been gathered. Because we only carried out convolution for one time now, only neighbors that have 1 degree are involved. For the sake of simplicity, let us assume that a node is more important only when it has more information gathering ability. In this time, we can see that the author node (inner red diamond) contributes more information gathering for it has more neighbors with 1 degree. Therefore, our attention layer preserves the vector of author and decrease the importance of the other two nodes. When we continue carrying out 2 layer convolution, the inner nodes will continue gather the nodes with 2 degree. As can be seen in the bottom of Fig. 4, our deep attention layer preserves the vector of venue node (inner green pentagram), for it has more neighbors with 2 degree. It is worth noting that the color of preserved venue vector is a little lighter, for our deep layer considers venue has less impact for paper retrieval task. The final vectors are then generated according to the reweighted vectors in all convolution layers. In this running example, we can see that our HDGCN is able to reweight node vectors by not only the information of local or global neighbors, but also the contribution of node types, thus generate more reliable vector representations.

### 5. Experiments and evaluations

#### 5.1. Datasets and preprocessing

In order to evaluate the performance of our proposed method, we use two publicly available datasets: CORD-19 and LitCovid. The briefly introduction of used datasets are as follows.

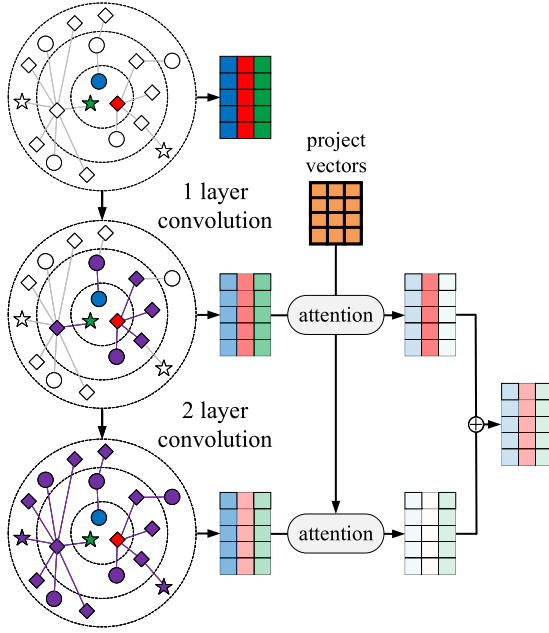


Fig. 4. A running example of a HDGCN with 2 convolutional layers.

· CORD-19 dataset<sup>5</sup>: the COVID-19 Open Research Dataset (CORD-19) (Wang et al., 2020) is established by the U.S. White House, along with the U.S. National Institutes of Health, the Allen Institute for AI, the Chan-Zuckerberg Initiative, Microsoft Research, and Georgetown University. It is one of the earliest scientific paper datasets related to COVID-19 and similar coronaviruses such as SARS, MERS and other earlier coronaviruses since 2002. The sources of these papers including the WHO, and bioRxiv, medRxiv, etc. For evaluation purpose, we only keep the paper that cited over 5 times and related citing paper. We finally obtain a subset that contains 327,619 papers.

· LitCovid dataset<sup>6</sup>: this lightweight dataset is published by National Library of Medicine (Chen et al., 2021). It is a collection of recently published PubMed articles which are directly related to the 2019 novel coronavirus. The dataset contains upwards of 240,000 articles and growing every day, making it a comprehensive resource for keeping researchers up to date with the current COVID-19 crisis. For our recommendation purpose, we select as subset that contains 103,175 articles.

For each papers, we pre-process it by removing stop words and stemming using Porter stemmer. To reduce the impact of short words, we remove the words which consist of less than two characters and appear less than ten times in the dataset. We consider the citing sentence with additional 2 sentences around the citation placeholder as a query, and the corresponding cited papers as ground truth retrieve result. For evaluating the performance of citation recommendation, we select 37,289 and 6613 citation context in CORD-19 and LitCovid as test set, respectively. Table 4 summarizes the statistics of used two datasets.

## 5.2. Evaluation metrics

To thoroughly evaluate the recommendation of our method, we choose the following four types of metric.

<sup>5</sup> <https://www.kaggle.com/datasets/allen-institute-for-ai/CORD-19-research-challenge>

<sup>6</sup> <https://www.ncbi.nlm.nih.gov/research/coronavirus/#data-download>

Table 4  
Statistics of used datasets.

		Papers	Authors	Venues
CORD-19	Train	327,619	284,581	15,924
	Test	37,289	35,417	7,423
LitCovid	Train	103,175	90,692	6,305
	Test	6,613	5,934	2,081

· Precision and recall: These two types of metric are the most widely used metrics in information retrieval area. For test samples, we evaluate precision and recall according to the top N results using citing score. Precision@N (P@N) measures the ratio of correctly retrieved citations to N. Recall@N (R@N) measures the ratio of correctly retrieved citations to all ground truths. These two metrics are defined as follows.

$$\text{Precision@N} = \frac{\sum_{q \in D_q} |R(q) \cap T(q)|}{R(q)} \quad (12)$$

$$\text{Recall@N} = \frac{\sum_{q \in D_q} |R(q) \cap T(q)|}{T(q)} \quad (13)$$

where  $q$  is a query citation context.  $D_q$  is the testing set.  $R(q)$  represents the top N recommended results for  $q$ .  $T(q)$  denotes all ground truth citations of  $q$ .

· Mean Reciprocal Rank (MRR): This metric is calculated by the reciprocal value of position for the first matched results. MRR is defined as follows.

$$\text{MRR} = \frac{1}{|D_q|} \sum_{q \in D_q} \frac{1}{\text{rank}_q} \quad (14)$$

where  $|D_q|$  is the size of testing set.  $\text{rank}_q$  is the first matched recommend result for query citation context  $q$ . It should be noted that because a query citation context may contains multiple references, we treat them equally and calculate the MRR score for each ground truth citation.

· Mean Average Precision (MAP): Unlike MRR that limits a query only have one matched result, MAP considers the situation of multiple matched results. For a query  $q$ , we calculate average precision(AveP) as follows.

$$\text{AveP} = \frac{1}{|T(q)|} \sum_{m=1}^{|T(q)|} \frac{m}{\text{rank}_q^m} \quad (15)$$

where  $|T(q)|$  is the overall number of ground truths of query  $q$ .  $\text{rank}_q^m$  is the position of  $m$ th matched ground truth in  $T(q)$ . We calculate AveP value for all testing samples, and then average them to obtain the final MAP value.

## 5.3. Comparing with other approaches

We use the following compared methods to evaluate the performance of our method for COVID-19 citation recommendation.

· TF-IDF: We extract TF-IDF vectors for all papers and query citation context, and retrieve the papers that have high cosine similarity with the query. To promote the calculation, we only preserve top 8000 words according to word frequency.

· BM25 (Robertson & Walker, 1994): BM25 is a well-known ranking method for measuring the relevance of matching documents to a query based on the text. We calculate the text similarity between the papers and query citation context by using both TF and IDF for BM25.

· SciBERT (Beltagy et al., 2019): We send all papers and query into SciBERT, and obtain the mean vector according to their output word vectors. Then same with TF-IDF method, we consider the papers with higher cosine similarity with the query as the recommend results.

· Association rules (AR) (Fister et al., 2020): By calculating term frequency and inverse term frequency (TF/ITF) and constructing term



**Table 5**  
Performance comparison between different methods on CORD-19.

Methods	MAP	MRR	P@25	P@50	P@75	P@100	R@25	R@50	R@75	R@100
TF-IDF	0.0541	0.0604	0.0191	0.0148	0.0062	0.0013	0.0719	0.1005	0.1251	0.1694
BM25	0.0562	0.0689	0.0224	0.0161	0.0074	0.0017	0.0943	0.1139	0.1354	0.1825
SciBERT	0.0716	0.0845	0.0314	0.0193	0.0124	0.0048	0.1152	0.1382	0.1584	0.2136
BERT-DVGCN	0.0826	0.0912	0.0408	0.0214	0.0147	0.0085	0.1282	0.1426	0.1671	0.2216
AR	0.1107	0.1169	0.0594	0.0377	0.0272	0.0142	0.1442	0.1681	0.1869	0.2454
CO-Search	0.1227	0.1289	0.0697	0.0493	0.0423	0.0197	0.1569	0.1727	0.2036	0.2563
CRB-HDGCN	0.1495	0.1582	0.0848	0.0608	0.0572	0.0271	0.1953	0.2279	0.2483	0.3107

**Table 6**  
Performance comparison between different methods on LitCovid.

Methods	MAP	MRR	P@25	P@50	P@75	P@100	R@25	R@50	R@75	R@100
TF-IDF	0.0705	0.0753	0.0334	0.0237	0.0102	0.0034	0.1153	0.1406	0.1690	0.1907
BM25	0.0723	0.0831	0.0365	0.0253	0.0179	0.0065	0.1268	0.1512	0.1784	0.2198
SciBERT	0.0923	0.1005	0.0564	0.0301	0.0240	0.0185	0.1706	0.1880	0.2037	0.2610
BERT-DVGCN	0.1081	0.1138	0.0626	0.0459	0.0307	0.0204	0.1792	0.2065	0.2204	0.2709
AR	0.1492	0.1673	0.1051	0.0761	0.0529	0.0306	0.2054	0.2275	0.2373	0.3096
CO-Search	0.1594	0.1717	0.1165	0.0852	0.0621	0.0378	0.2107	0.2388	0.2463	0.3159
CRB-HDGCN	0.2057	0.2161	0.1295	0.0952	0.0784	0.0431	0.2409	0.2770	0.2951	0.3517

graph, AR uses information cartography to extract associate rules to explore knowledge among COVID-19 papers. We identify the terms that occur in the best metro map for queries and candidate papers, and return the results that have higher term matching.

- BERT-DVGCN (Jeong et al., 2020): This method is originally for general citation recommendation task. It uses vanilla BERT to encode query sentences into context embeddings, and uses GCN to embeds paper vectors which is obtained by Doc2Vec (Le & Mikolov, 2014) into graph embedding according to citation relationships. Then Bert-DVGCN concatenates them and uses a softmax output layer to generate recommendation results.

- CO-Search (Esteve et al., 2021): This method linearly combines TF-IDF vector, Siamese-BERT vector and BM25 vector of queries and answers to generate recommendation results. We use same queries as our method, and consider cited papers as answers for CO-Search.

As we can see from Tables 5 and 6, our CRB-HDGCN is superior to other methods in terms of all metrics. For example, compared with the second best method CO-Search, CRB-HDGCN improves MAP, MRR, P@100 and R@100 with 21.8%, 22.7%, 37.6% and 21.2% on CORD-19, and 29.1%, 25.9%, 15.3% and 11.3% on LitCovid. This can be ascribed to our method uses domain related citation relations to let PLM interprets latent citation information between the COVID-19 papers more clearly. Moreover, we excavate a fine-grained paper graph by using multiple relations, and use deep graph neural network to differentiate the importance of nodes according to both type and range, which is able to obtained reliable paper vectors and consequently further improve the performance of citation recommendation. We can find that due to only utilize word frequency information, both TF-IDF and BM25 are obviously worse than all other methods. SciBERT works much better than TF-IDF and BM25, which indicates that the latent information excavated by PLMs is much more useful than word frequency in finding relevant papers. By utilizing link relations, BERT-DVGCN further improves the performance, which demonstrated that link information plays an important role in citation recommendation task. However, both SciBERT and BERT-DVGCN are only trained on common papers, which will results in understanding deviation for domain specific corpus such as COVID-19. AR uses association rules to deeply mine the domain knowledge among COVID-19 papers, and its performance is greatly improved compared with SciBERT and BERT-DVGCN. By employing high performance pre-trained word vector in COVID-19 field, CO-Search further improves performance compared with AR which only utilize word occurrence information. However, the training of word vectors in CO-Search fails to effectively mine citation information among COVID-19 papers, and does not utilize vital author and venue information. Our CRB-HDGCN mitigates above problems by combining citation enhanced PLM with deep heterogeneous graph learning.

#### 5.4. Comparing with variants of our method

Our proposed CRB-HDGCN consists of multiple sub-networks. In order to evaluate the effectiveness of these sub-networks, we conduct experiment with partial CRB-HDGCN on selected datasets. The compared variants of our method are summarized as follows.

- CRB: This variant only uses citation relational BERT part to obtain mean vectors for papers, authors and venues.
- BERT-HDGCN-PA: This variant uses vanilla BERT to obtain initial vectors, and uses HDGCN with paper and author nodes.
- BERT-HDGCN-PJ: This variant is similar with BERT-HDGCN-PA except that it uses paper and venue nodes in HDGCN.
- BERT-HDGCN: This variant extends BERT-HDGCN-PJ by adding author nodes.
- CRB-HDGCN-PA: This variant replaces BERT-HDGCN-PA with CRB.
- CRB-HDGCN-PJ: This variant replaces author nodes in CRB-HDGCN-PA by venue nodes.

The experimental results are shown in Tables 7 and 8. We can see that CRB performs worse than any other variants because it only utilizes content information. Nevertheless, it can be seen that CRB still outperforms than SciBERT and BERT-DVGCN (in Tables 5 and 6). It is should be noticed that CRB does not using the whole COVID-19 literature dataset to heavily train BERT, but only chooses reference sentence relations to slightly retrain and fine-tune it. This reveals that it is necessary to use citation information to further retrain and fine-tune PLMs for citation recommendation task. As can be seen when introducing heterogeneous deep graph convolution, both BERT-HDGCN-PA and BERT-HDGCN-PJ perform better than CRB and BERT-DVGCN (in Tables 5 and 6). Particularly, we can see that BERT-HDGCN-PA is better than BERT-HDGCN-PJ, which shows that author information is more important than venue information. This can be ascribe into author usually has customized citing activity, for an author is tend to cite papers that he/she read or cite before, which result in the citing activities have high time continuity. Nevertheless, when using all types of nodes, BERT-HDGCN is able to automatically differentiate the importance of each node type and to generate more robust node vector representation, and consequently reaches better performance. We can also see that the performance of variants that using CRB is consistently outperform than that only using vanilla BERT.

#### 5.5. Efficiency evaluation

In this subsection, we conduct experiment on CORD-19 and LitCovid to evaluate the efficiency of our method. The compared methods are BERT-DVGCN and CO-Search. We first use 10% of samples in dataset to

**Table 7**

Performance comparison between different variants on CORD-19.

Methods	MAP	MRR	P@25	P@50	P@75	P@100	R@25	R@50	R@75	R@100
CRB	0.0842	0.0929	0.0415	0.0237	0.0151	0.0093	0.1293	0.1442	0.1689	0.2236
BERT-HDGCN-PA	0.1224	0.1335	0.0653	0.0374	0.0349	0.0185	0.1702	0.1983	0.2263	0.2752
BERT-HDGCN-PJ	0.1135	0.1267	0.0606	0.0344	0.0316	0.0169	0.1683	0.1901	0.2185	0.2703
BERT-HDGCN	0.1265	0.1388	0.0717	0.0452	0.0392	0.0215	0.1762	0.2043	0.2308	0.2841
CRB-HDGCN-PA	0.1452	0.1541	0.0803	0.0559	0.0543	0.0246	0.1914	0.2218	0.2411	0.2983
CRB-HDGCN-PJ	0.1385	0.1479	0.0762	0.0512	0.0515	0.0226	0.1892	0.2176	0.2385	0.2925
CRB-HDGCN	0.1495	0.1582	0.0848	0.0608	0.0572	0.0271	0.1953	0.2279	0.2483	0.3107

**Table 8**

Performance comparison between different variants on LitCovid.

Methods	MAP	MRR	P@25	P@50	P@75	P@100	R@25	R@50	R@75	R@100
CRB	0.1403	0.1492	0.0872	0.0593	0.0362	0.0261	0.1813	0.2096	0.2283	0.2856
BERT-HDGCN-PA	0.1782	0.1852	0.1105	0.0762	0.0614	0.0335	0.2203	0.2373	0.2552	0.3169
BERT-HDGCN-PJ	0.1646	0.1791	0.1025	0.0707	0.0585	0.0308	0.2102	0.2285	0.2497	0.3102
BERT-HDGCN	0.1892	0.1915	0.1175	0.0812	0.0693	0.0641	0.2261	0.2496	0.2634	0.3253
CRB-HDGCN-PA	0.1993	0.2093	0.1226	0.0885	0.0724	0.0382	0.2336	0.2698	0.2867	0.3428
CRB-HDGCN-PJ	0.1925	0.2037	0.1176	0.0824	0.0700	0.0360	0.2301	0.2635	0.2839	0.3380
CRB-HDGCN	0.2057	0.2161	0.1295	0.0952	0.0784	0.0431	0.2409	0.2770	0.2951	0.3517

**Table 9**

Retraining and fine-tune time of CRB.

Dataset	Retraining	Fine-tune
CORD-19	10,401 s	6,729 s
LitCovid	6,158 s	4,613 s

carry out training and testing process, and then incrementally increase the volume of samples to check the difference of running time and R@100 between compared methods. Because citation relational BERT part in our method can be considered as a pre-processing step for getting better PLMs model, we separately list the retraining time and fine-tune time of CRB in Table 9, and only present rest time of our method for the fairness of comparison.

The comparison of efficiency is demonstrated in Tables 10 and 11. We can see that our CRB-HDGCN uses competitive time compared with BERT-DVGCN, but generates much better performance. There are two reasons for this result. The first reason is the convolutional part of CRB-HDGCN is similar with BERT-DVGCN, except adding more heterogeneous nodes which can be simultaneously executed by convolution process. The second reason is the added deep attention layer is essentially a shallow network compared with stacked deep convolution, which also brings little burden for training and testing. The training time of CO-Search is less than our CRB-HDGCN, for it only applied paragraphs and citations tuples to training Siamese-BERT. However, in testing stage, CO-Search executes a much more time consuming clustering and re-ranking process, which results in it spends much more time than CRB-HDGCN.

### 5.6. Parameter tuning

There are several manually set hyperparameters in our CRB-HDGCN, which will affect the performance of recommendation. In this subsection, we conduct experiment to investigate the optimal value of them. These essential hyperparameters are:  $L$ ,  $d_s$  and  $d_a$ .  $L$  is the overall number of convolution layers. According previous study on Liu et al. (2020), we set the range of  $L$  from 1 to 25.  $d_s$  is the size of hidden vector dimension in heterogeneous deep graph convolutional neural network. Because  $d_s$  determines the learning ability of HDGCN, we set the value of  $d_s$  from 50 to 500.  $d_a$  is the dimension of trainable project vectors in deep attention layer. Same as  $d_s$ , the range of  $d_a$  is set from 50 to 500. We only demonstrate the tuning results on R@100, for other metrics generate similar results.

We first empirically fix  $d_a = 200$ , and evaluate the effect of  $d_s$  with different  $L$ . As can be seen in Fig. 5(a) and (b), when the

layer of HDGCN is small, the performance of recommendation is not satisfied. As the number of layers increases, the performance gradually improves. When the number of layers reaches 15, the performance reaches to optimum. After that, when the number of layers increases, the performance drops slightly. We believe the reason behind this is that more graph convolution layers will let more long distant neighbors participate the learning process of node representation. In the scientific literature network, more long distant nodes will bring more research drift, which meantime bring more noise, and consequently resulting in performance degradation. It also can be seen that when dimension of  $d_s$  is small ( $= 50$ ), the recommendation performances with different layers are relatively close. As  $d_s$  increases, the performance gradually increases. When the value of  $d_s$  reaches about 300, the performance gradually stabilize. Since a larger  $d_s$  will lead to larger size of parameter scale and will increase training difficulty and time, the optimal value of  $d_s$  is set to 300.

We then fix  $d_s = 300$ , and study the impact of  $d_a$  with different  $L$ . The results are demonstrated in Fig. 5(c) and (d). It can be seen that when the convolutional layer is small ( $= 5$ ), the change of  $d_a$  has little impact on recommendation performance. This is due to limited long distance nodes are involved for node representation learning in this situation. Therefore, the deep attention layer in our model is hard to take full advantage. Nevertheless, as the number of layers  $L$  increases, we can be seen that the performance curve rises steeply at the beginning increasing of  $d_a$ . This is because in this situation, more long distance nodes are involved to participate in node representation learning, and the deep attention layer is able to effectively distinguish the importance of these nodes and greatly improving the recommendation performance. The performance reaches the best when  $d_a = 200$ . After that, the performance declines slightly with the increase of  $d_a$ . This is because a large  $d_a$  will bring more overfitting and decrease the performance. According to the above experimental results, the optimal value of  $d_a$  is set to 200.

### 5.7. Vector visualization and case study

To vividly demonstrate the effectiveness of our method, we show the vector visualization of CRB in our method, and present some recommendation cases.

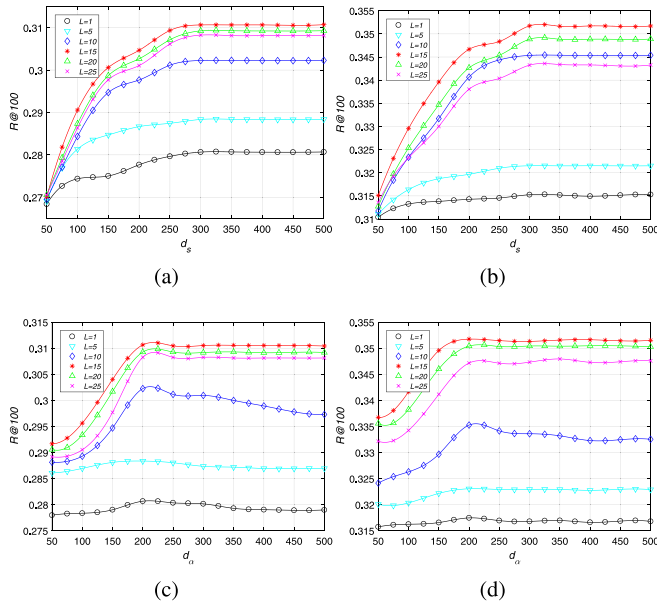
We choose 6 hot keywords in COVID-19 area, including: “coronavirus”, “bacteria”, “antibodies”, “quarantine”, “vaccine”, “respiratory” and “dysgeusia”. Given a sentence, BERT will tokenize it into a word list by using tokenizer. Because the wordpiece vocabulary of BERT contains limited words, some COVID-19 related words will be tokenized into several sub-token. For example, “coronavirus”, “quarantine” and “dysgeusia” will be tokenized into [‘corona’, ‘##virus’], [‘qu’,

**Table 10**  
Efficiency evaluation on CORD-19.

Methods	10% samples			50% samples			100% samples		
	Train time	Test time	R@100	train time	Test time	R@100	Train time	Test time	R@100
BERT-DVGCN	14,052 s	582 s	0.0937	73,269 s	2,353 s	0.1547	143,839 s	4,523 s	0.2216
CO-Search	11,503 s	904 s	0.1037	56,731 s	3,462 s	0.1835	116,493 s	6,620 s	0.2563
CRB-HDGCN	14,937 s	597 s	0.1352	74,820 s	2,495 s	0.2235	149,522 s	4,782 s	0.3107

**Table 11**  
Efficiency evaluation on LitCovid.

Methods	10% samples			50% samples			100% samples		
	Train time	Test time	R@100	train time	Test time	R@100	Train time	Test time	R@100
BERT-DVGCN	4,641 s	117 s	0.1163	21,659 s	450 s	0.1703	47,628 s	953 s	0.2709
CO-Search	3,128 s	182 s	0.1328	16,510 s	738 s	0.2094	42,627 s	1,773 s	0.3159
CRB-HDGCN	5,002 s	128 s	0.1561	23,569 s	497 s	0.2554	49,953 s	1,084 s	0.3517



**Fig. 5.** The impact of hyperparameters with different convolution layers  $L$ : (a)  $d_s$  on CORD-19; (b)  $d_s$  on LitCovid; (c)  $d_a$  on CORD-19; (d)  $d_a$  on LitCovid.

'##aran', '##ntine'] and ['d', '##ys', '##ge', '##usia'], respectively. In order to obtain reliable vectors of these words, we choose 1000 sentences for each words that contain it, and send these sentences to obtain output vector. The final word vector is calculated by the sum mean of all related sub-token in all chose sentences. After that, we compute the cosine similarity between these 6 keywords. We select different BERT model to generate results, including vanilla BERT, SciBERT, COVID-Twitter-BERT (Müller et al., 2020) and CRB. The visualization results are shown in Fig. 6. The diagonal is covered with white color for the cosine similarity with itself is equal to 1. We can see that vanilla BERT generate pretty low correlations between these words, which is because vanilla BERT is trained on common corpus and lacks deep understanding COVID-19 related content. SciBERT seems generates similar result. But SciBERT can still reveal correlation between some words. For example the cosine similarity between “bacteria” and “quarantine” generated by SciBERT is higher than vanilla BERT. This is because SciBERT is trained through scientific papers, which is helpful to discover more deep relation between text content. It can be seen that COVID-Twitter-BERT can excavate more correlations. For example, the cosine similarity between “coronavirus” and “quarantine” of COVID-Twitter-BERT are much higher than vanilla BERT and SciBERT. The reason is COVID-Twitter-BERT is trained on COVID-19 related twitters,

which is benefit for COVID-Twitter-BERT to find co-occurrences of these two words. However, only using twitter data to train BERT may results in understanding drifting problem, which is a serious problem for COVID-19 scientific citation recommendation task. We can see that our CRB is able to reveal more relations between these words, which shows using citation link and text is help for PLM comprehending COVID-19 related content.

To further show the effectiveness our method, we present some recommendation results in this subsection. We choose a query citation context from CORD-19 dataset. The paper ID is “PMC7096991” (Wack, Terczyńska-Dyla, & Hartmann, 2015) and the citing sentence is “This event creates docking sites for STAT1 and STAT2 signaling molecules, which leads to their recruitment and subsequent phosphorylation”. We also list the titles of three ground truth papers. The compared methods are using 5 and 15 layers of heterogeneous deep graph convolution, which are denoted as CRB-HDGCN-5 and CRB-HDGCN-15. As can be seen in Table 12, CRB-HDGCN-15 is able to correctly recommend all ground truths, while CRB-HDGCN-5 miss one ground truth. We can see that the correctly recommended results by CRB-HDGCN-5 are all published on same venue “Nature immunology”, which is also the publication venue of query paper. It is also interesting to further check the wrong results. We can see that CRB-HDGCN-5 still generate papers that published on the venue that very close to “Nature immunology”, while CRB-HDGCN-15 generate a much far wrong result. We checked the wrong result of CRB-HDGCN-15 and find it is the cited paper of ground truth 3, which means this wrong result may still have high possibility that cited by the query. The above findings indicate that CRB-HDGCN-5 can only find results that have more closer distance in paper graph. In contrast, CRB-HDGCN-15 is able to improve the importance of nodes that are far from the query, and consequently generate more effectiveness recommendation results.

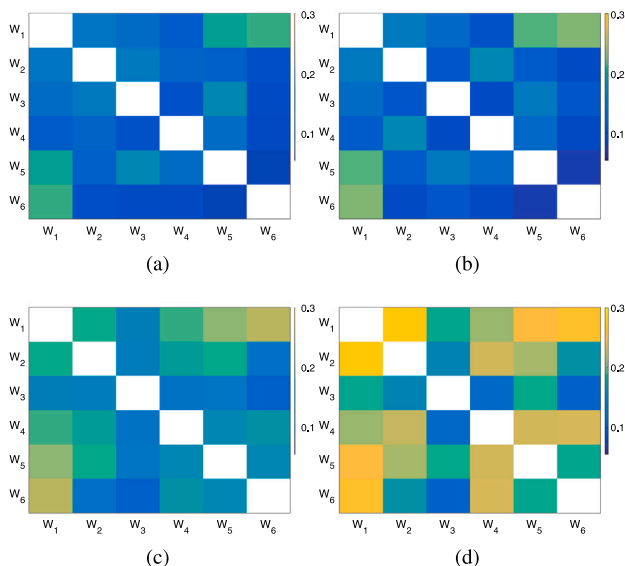
## 6. Conclusion

This paper presented an efficient and effective COVID-19 citation recommendation method CRB-HDGCN. The method first augments citing sentences among COVID-19 papers with the title of cited paper, and uses these augmented citing sentences to lightly retrain BERT model. Then we choose three types of sentence pairs between citing and cited paper to further fine-tune BERT, in order to let PLM better comprehend the citing activities among COVID-19 paper. Secondly, we develop a heterogeneous deep graph convolutional network (HDGCN) to learn vector representations for papers, authors and venues in COVID-19 literature dataset. The network uses an initial embedding layer to transfer the vector of all types of node into fixed length, and uses a heterogeneous graph convolutional layer to obtain hidden vectors for all nodes. Moreover, we add a deep attention layer to let the network is able to automatically differentiate the importance of nodes both according to types and distance, which is benefit for getting reliable

**Table 12**

An example of 4 recommended citations by CRB-HDGCN-5 and CRB-HDGCN-15 for paper “PMC7096991”.

Ground truths:	CRB-HDGCN-5	CRB-HDGCN-15
<b>Query:</b> This event creates docking sites for STAT1 and STAT2 signaling molecules, which leads to their recruitment and subsequent phosphorylation. (venue: Nature Immunology)	<b>[O] result 1:</b> IFN- $\lambda$ s mediate antiviral protection through a distinct class II cytokine receptor complex. (venue: Nature Immunology)	<b>[O] result 1:</b> IFN- $\lambda$ s mediate antiviral protection through a distinct class II cytokine receptor complex. (venue: Nature Immunology)
<b>ground truth 1:</b> IFN- $\lambda$ s mediate antiviral protection through a distinct class II cytokine receptor complex. (venue: Nature Immunology)	<b>[X] result 2:</b> Diverse intracellular pathogens activate type III interferon expression from peroxisomes. (venue: Nature Immunology)	<b>[O] result 2:</b> IL-28, IL-29 and their class II cytokine receptor IL-28R. (venue: Nature Immunology)
<b>ground truth 2:</b> IL-28, IL-29 and their class II cytokine receptor IL-28R. (venue: Nature Immunology)	<b>[O] result 3:</b> IL-28, IL-29 and their class II cytokine receptor IL-28R. (venue: Nature Immunology)	<b>[X] result 3:</b> Cloning of a new type II cytokine receptor activating signal transducer and activator of transcription (STAT) 1, STAT2 and STAT3. (venue: Biochemical Journal)
<b>ground truth 3:</b> Role of the interleukin (IL)-28 receptor tyrosine residues for antiviral and antiproliferative activity of IL-29/interferon- $\lambda$ 1: similarities with type I interferon signaling. (venue: Journal of Biological Chemistry)	<b>[X] result 4:</b> Mechanisms of type-I-and type-II-interferon-mediated signaling. (venue: Nature Reviews Immunology)	<b>[O] result 4:</b> Role of the interleukin (IL)-28 receptor tyrosine residues for antiviral and antiproliferative activity of IL-29/interferon- $\lambda$ 1: similarities with type I interferon signaling. (venue: Journal of Biological Chemistry)



**Fig. 6.** The visualization of cosine similarity between words. The  $w_1$  to  $w_6$  represent “coronavirus”, “bacteria”, “antibodies”, “quarantine”, “vaccine”, “respiratory” and “dysgeusia”, respectively. The sub-graphs are the results by choosing different BERT model, including: (a) vanilla BERT, (b) SciBERT, (c) COVID-Twitter-BERT and (d) CRB.

vectors. The experimental results carried on CORD-19 and LitCovid datasets validated the effectiveness of our method.

In the future, we will try to add more domain knowledge to help PLMs better understanding COVID-19 papers. We will also try to use more available fine-grained correlations among COVID-19 papers to further help graph neural network to get more robust vector representations.

#### CRedit authorship contribution statement

**Tao Dai:** Conceptualization, Methodology, Writing – original draft, Implementation of algorithm, Writing – review & editing. **Jie Zhao:** Conceptualization, Methodology, Writing – reviewing, Editing, Supervision. **Dehong Li:** Data curation, Visualization, Carrying out experiments. **Shun Tian:** Implementation of algorithm, Carrying out experiments. **Xiangmo Zhao:** Supervision. **Shirui Pan:** Validation, Investigation, Editing.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Data availability

Data will be made available on request.

#### Acknowledgments

This work was partially support by the Natural Science Foundation in Shaanxi Province of China (Project No. 2021JQ-289; No. 2022JQ-657; No. 2022JQ-007), China Postdoctoral Science Foundation (Project No. 2021M700533; No. 2021M700532; No. 2021M692737), MOE (Ministry of Education in China) Project of Humanities and Social Sciences (Project No. 21YJC630057) and the Fundamental Research Funds for the Central Universities, CHD (Project No. 300102231302; No. 300102232602; No. 300102231101).

#### References

- Ahamed, S., & Samad, M. (2020). Information mining for covid-19 research from a large volume of scientific literature. <http://dx.doi.org/10.48550/arXiv.2004.02085>, arXiv preprint [arXiv:2004.02085](https://arxiv.org/abs/2004.02085).
- Andre, E., Anuprit, K., Romain, P., Kazuma, H., Wenpeng, Y., & Dragomir, R. (2020). Co-search: Covid-19 information retrieval with semantic search, question answering, and abstractive summarization. *Information Retrieval Repository*, <http://dx.doi.org/10.1038/s41746-021-00437-0>.
- Aristovnik, A., Ravšelj, D., & Umek, L. (2020). A bibliometric analysis of COVID-19 across science and social science research landscape. *Sustainability*, 12(21), 9132. <http://dx.doi.org/10.3390/su12219132>.
- Ayala-Gomez, F., Daróczy, B., Benczúr, A., Mathioudakis, M., & Gionis, A. (2018). Global citation recommendation using knowledge graphs. *Journal of Intelligent & Fuzzy Systems*, 34(5), 3089–3100. <http://dx.doi.org/10.3233/JIFS-169493>.
- Beltagy, I., Lo, K., & Cohan, A. (2019). SciBERT: A pretrained language model for scientific text. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing* (pp. 3615–3620). <http://dx.doi.org/10.18653/v1/D19-1371>.
- Brainard, J. (2020). *New tools aim to tame pandemic paper tsunami*. American Association for the Advancement of Science, <http://dx.doi.org/10.1126/science.368.6494.924>.
- Chen, Q., Allot, A., & Lu, Z. (2021). LitCovid: an open database of COVID-19 literature. *Nucleic Acids Research*, 49(D1), D1534–D1540. <http://dx.doi.org/10.1093/nar/gkaa9522>.
- Chen, Q., Du, J., Allot, A., & Lu, Z. (2022). LitMC-BERT: transformer-based multi-label classification of biomedical literature with an application on COVID-19 literature curation. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, <http://dx.doi.org/10.1109/TCBB.2022.3173562>.



- Chen, D., Lin, Y., Li, W., Li, P., Zhou, J., & Sun, X. (2020). Measuring and relieving the over-smoothing problem for graph neural networks from the topological view. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34 (04), (pp. 3438–3445). <http://dx.doi.org/10.1609/aaai.v34i04.5747>.
- Dai, T., Yan, W., Zhang, K., Qiu, C., Zhao, X., & Pan, S. (2021). Gated relational stacked denoising autoencoder with localized author embedding for global citation recommendation. *Expert Systems with Applications*, 184, Article 115359. <http://dx.doi.org/10.1016/j.eswa.2021.115359>.
- Dai, T., Zhu, L., Wang, Y., & Carley, K. M. (2019). Attentive stacked denoising autoencoder with bi-lstm for personalized context-aware citation recommendation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28, 553–568. <http://dx.doi.org/10.1109/TASLP.2019.2949925>.
- Duma, D., Liakata, M., Clare, A., Ravenscroft, J., & Klein, E. (2016). Applying core scientific concepts to context-based citation recommendation. In *LREC*.
- Ebesu, T., & Fang, Y. (2017). Neural citation network for context-aware citation recommendation. In *Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval* (pp. 1093–1096). ACM, <http://dx.doi.org/10.1145/3077136.3080730>.
- Esteva, A., Kale, A., Paulus, R., Hashimoto, K., Yin, W., Radev, D., et al. (2021). COVID-19 information retrieval with deep-learning based semantic search, question answering, and abstractive summarization. *NPJ Digital Medicine*, 4(1), 1–9. <http://dx.doi.org/10.1038/s41746-021-00437-0>.
- Färber, M., Thiemann, A., & Jatowt, A. (2018). CITEWERTs: A system combining cite-worthiness with citation recommendation. In *European conference on information retrieval* (pp. 815–819). Springer, [http://dx.doi.org/10.1007/978-3-319-76941-7\\_82](http://dx.doi.org/10.1007/978-3-319-76941-7_82).
- Fister, I., Jr., Fister, K., & Fister, I. (2020). Discovering associations in COVID-19 related research papers. <http://dx.doi.org/10.48550/arXiv.2004.03397>, arXiv preprint [arXiv:2004.03397](https://arxiv.org/abs/2004.03397).
- Gan, L., Teng, Z., Zhang, Y., Zhu, L., Wu, F., & Yang, Y. (2022). Semglove: Semantic co-occurrences for glove from bert. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, <http://dx.doi.org/10.1109/TASLP.2022.3197316>.
- Gu, N., Gao, Y., & Hahnloser, R. H. (2022). Local citation recommendation with hierarchical-attention text encoder and SciBERT-based reranking. In *European conference on information retrieval* (pp. 274–288). Springer, [http://dx.doi.org/10.1007/978-3-030-99736-6\\_19](http://dx.doi.org/10.1007/978-3-030-99736-6_19).
- He, Q., Kifer, D., Pei, J., Mitra, P., & Giles, C. L. (2011). Citation recommendation without author supervision. In *Proceedings of the fourth ACM international conference on web search and data mining* (pp. 755–764). ACM, <http://dx.doi.org/10.1145/1935826.1935926>.
- He, Q., Pei, J., Kifer, D., Mitra, P., & Giles, L. (2010). Context-aware citation recommendation. In *Proceedings of the 19th international conference on world wide web* (pp. 421–430). ACM, <http://dx.doi.org/10.1145/1772690.1772734>.
- Huang, W., Kataria, S., Caragea, C., Mitra, P., Giles, C. L., & Rokach, L. (2012). Recommending citations: translating papers into references. In *Proceedings of the 21st ACM international conference on information and knowledge management* (pp. 1910–1914). <http://dx.doi.org/10.1145/2396761.2398542>.
- Huang, W., Wu, Z., Liang, C., Mitra, P., & Giles, C. L. (2015). A neural probabilistic model for context based citation recommendation. In *Twenty-Ninth AAAI conference on artificial intelligence*. <http://dx.doi.org/10.1609/aaai.v29i1.9528>.
- Huang, W., Wu, Z., Mitra, P., & Giles, C. L. (2014). Refseer: A citation recommendation system. In *IEEE/ACM joint conference on digital libraries* (pp. 371–374). IEEE, <http://dx.doi.org/10.1109/JCDL.2014.6970192>.
- Jeong, C., Jang, S., Park, E., & Choi, S. (2020). A context-aware citation recommendation model with BERT and graph convolutional networks. *Scientometrics*, 124(3), 1907–1922. <http://dx.doi.org/10.1007/s11192-020-03561-y>.
- Jiang, Z., Yin, Y., Gao, L., Lu, Y., & Liu, X. (2018). Cross-language citation recommendation via hierarchical representation learning on heterogeneous graph. In *The 41st international ACM SIGIR conference on research & development in information retrieval* (pp. 635–644). ACM, <http://dx.doi.org/10.1145/3209978.3210032>.
- Kenton, J. D. M.-W. C., & Toutanova, L. K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT* (pp. 4171–4186). <http://dx.doi.org/10.18653/v1/N19-1423>.
- Khadka, A., & Knott, P. (2018). Using citation-context to reduce topic drifting on pure citation-based recommendation. In *Proceedings of the 12th ACM conference on recommender systems* (pp. 362–366). ACM, <http://dx.doi.org/10.1145/3240323.3240379>.
- Kong, X., Shi, Y., Yu, S., Liu, J., & Xia, F. (2019). Academic social networks: Modeling, analysis, mining and applications. *Journal of Network and Computer Applications*, 132, 86–103. <http://dx.doi.org/10.1016/j.jnca.2019.01.029>.
- Kong, X., Zhang, J., Zhang, D., Bu, Y., Ding, Y., & Xia, F. (2020). The gene of scientific success. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 14(4), 1–19. <http://dx.doi.org/10.1145/3385530>.
- Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. In *International conference on machine learning* (pp. 1188–1196). PMLR, <http://dx.doi.org/10.48550/arXiv.1405.4053>.
- Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., et al. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4), 1234–1240. <http://dx.doi.org/10.1093/bioinformatics/btz682>.
- Li, Q., Han, Z., & Wu, X.-M. (2018). Deeper insights into graph convolutional networks for semi-supervised learning. In *Thirty-second AAAI conference on artificial intelligence*. <http://dx.doi.org/10.1609/aaai.v32i1.11604>.
- Liu, M., Gao, H., & Ji, S. (2020). Towards deeper graph neural networks. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 338–348). <http://dx.doi.org/10.1145/3394486.3403076>.
- MacAvaney, S., Cohan, A., & Goharian, N. (2020). SLEDGE-Z: A zero-shot baseline for COVID-19 literature search. In *Proceedings of the 2020 conference on empirical methods in natural language processing* (pp. 4171–4179). <http://dx.doi.org/10.18653/v1/2020.emnlp-main.341>.
- Mass, Y., & Roitman, H. (2020). Ad-hoc document retrieval using weak-supervision with BERT and GPT2. In *Proceedings of the 2020 conference on empirical methods in natural language processing* (pp. 4191–4197). <http://dx.doi.org/10.18653/v1/2020.emnlp-main.343>.
- Mercorio, F., Mezzananza, M., Moscato, V., Picariello, A., & Sperli, G. (2019). DICO: A graph-db framework for community detection on big scholarly data. *IEEE Transactions on Emerging Topics in Computing*, 9(4), 1987–2003. <http://dx.doi.org/10.1109/TETC.2019.2952765>.
- Müller, M., Salathé, M., & Kummervold, P. E. (2020). Covid-twitter-bert: A natural language processing model to analyse covid-19 content on twitter. <http://dx.doi.org/10.48550/arXiv.2005.07503>, arXiv preprint [arXiv:2005.07503](https://arxiv.org/abs/2005.07503).
- Nguyen, V., Rybinski, M., Karimi, S., & Xing, Z. (2020). Searching scientific literature for answers on covid-19 questions. <http://dx.doi.org/10.48550/arXiv.2007.02492>, arXiv preprint [arXiv:2007.02492](https://arxiv.org/abs/2007.02492).
- Ostendorf, M., Ruas, T., Blume, T., Gipp, B., & Reh, G. (2020). Aspect-based document similarity for research papers. In *Proceedings of the 28th international conference on computational linguistics* (pp. 6194–6206). <http://dx.doi.org/10.48550/arXiv.2010.06395>.
- Peng, H., Liu, J., & Lin, C.-Y. (2016). News citation recommendation with implicit and explicit semantics. In *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 1: long papers)*, Vol. 1 (pp. 388–398). <http://dx.doi.org/10.18653/v1/P16-1037>.
- Peters, M., Neumann, M., Iyyer, M., Gardner, M., & Zettlemoyer, L. (2018). Deep contextualized word representations. <http://dx.doi.org/10.18653/v1/N18-1202>.
- Qazvinian, V., & Radev, D. (2010). Identifying non-explicit citing sentences for citation-based summarization. In *Proceedings of the 48th annual meeting of the association for computational linguistics* (pp. 555–564).
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8), 9.
- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using siamese BERT-networks. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing* (pp. 3982–3992). <http://dx.doi.org/10.18653/v1/D19-1410>.
- Robertson, S. E., & Walker, S. (1994). Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In *SIGIR'94* (pp. 232–241). Springer, [http://dx.doi.org/10.1007/978-1-4471-2099-5\\_24](http://dx.doi.org/10.1007/978-1-4471-2099-5_24).
- Rohatgi, S., Karishma, Z., Chhay, J., Keesara, S. R. R., Wu, J., Caragea, C., et al. (2020). Covidseer: Extending the COVID-19 dataset. In *Proceedings of the ACM symposium on document engineering 2020* (pp. 1–4). <http://dx.doi.org/10.1145/3395027.3419597>.
- Soni, S., & Roberts, K. (2021). An evaluation of two commercial deep learning-based information retrieval systems for COVID-19 literature. *Journal of the American Medical Informatics Association*, 28(1), 132–137. <http://dx.doi.org/10.1093/jamia/ocaa271>.
- Totti, L. C., Mitra, P., Ouzzani, M., & Zaki, M. J. (2016). A query-oriented approach for relevance in citation networks. In *Proceedings of the 25th international conference companion on world wide web* (pp. 401–406). International World Wide Web Conferences Steering Committee, <http://dx.doi.org/10.1145/2872518.2890518>.
- Tran, V., Tran, V.-H., Nguyen, P., Nguyen, C., Satoh, K., Matsumoto, Y., et al. (2021). CovRelex: A COVID-19 retrieval system with relation extraction. In *Proceedings of the 16th conference of the european chapter of the association for computational linguistics: system demonstrations* (pp. 24–31). <http://dx.doi.org/10.18653/v1/2021.eacl-demos.4>.
- Uday, S. S., Pavani, S. T., Lakshmi, T. J., & Chivukula, R. (2022). COVID-19 literature mining and retrieval using text mining approaches. <http://dx.doi.org/10.48550/arXiv.2205.14781>, arXiv preprint [arXiv:2205.14781](https://arxiv.org/abs/2205.14781).
- Wack, A., Terczyńska-Dyla, E., & Hartmann, R. (2015). Guarding the frontiers: the biology of type III interferons. *Nature Immunology*, 16(8), 802–809. <http://dx.doi.org/10.1038/ni.3212>.
- Wang, L. L., Lo, K., Chandrasekhar, Y., Reas, R., Yang, J., Eide, D., et al. (2020). Cord-19: The covid-19 open research dataset. <http://dx.doi.org/10.48550/arXiv.2004.10706>, arXiv.

- Wise, C., Ioannidis, V. N., Calvo, M. R., Song, X., Price, G., Kulkarni, N., et al. (2020). COVID-19 knowledge graph: accelerating information retrieval and discovery for scientific literature. <http://dx.doi.org/10.48550/arXiv.2007.12731>, arXiv preprint [arXiv:2007.12731](https://arxiv.org/abs/2007.12731).
- Xia, F., Wang, W., Bekele, T. M., & Liu, H. (2017). Big scholarly data: A survey. *IEEE Transactions on Big Data*, 3(1), 18–35. <http://dx.doi.org/10.1109/TBDATA.2016.2641460>.
- Xiong, C., Liu, Z., Sun, S., Dai, Z., Zhang, K., Yu, S., et al. (2020). CMT in TREC-COVID round 2: mitigating the generalization gaps from web to special domain search. <http://dx.doi.org/10.48550/arXiv.2011.01580>, arXiv preprint [arXiv:2011.01580](https://arxiv.org/abs/2011.01580).
- Xu, K., Li, C., Tian, Y., Sonobe, T., Kawarabayashi, K.-i., & Jegelka, S. (2018). Representation learning on graphs with jumping knowledge networks. In *International conference on machine learning* (pp. 5453–5462). PMLR, <http://dx.doi.org/10.48550/arXiv.1806.03536>.
- Yin, J., & Li, X. (2017). Personalized citation recommendation via convolutional neural networks. In *Asia-pacific web (APWeb) and web-age information management (WAIM) joint conference on web and big data* (pp. 285–293). Springer, [http://dx.doi.org/10.1007/978-3-319-63564-4\\_23](http://dx.doi.org/10.1007/978-3-319-63564-4_23).
- Zhang, E., Gupta, N., Nogueira, R., Cho, K., & Lin, J. (2020). Rapidly deploying a neural search engine for the COVID-19 open research dataset. In *Proceedings of the 1st workshop on NLP for COVID-19 At ACL 2020*. <http://dx.doi.org/10.48550/arXiv.2004.05125>.
- Zhang, Y., & Ma, Q. (2022). Dual attention model for citation recommendation with analyses on explainability of attention mechanisms and qualitative experiments. *Computational Linguistics*, 48(2), 403–470. [http://dx.doi.org/10.1162/coli\\_a\\_00438](http://dx.doi.org/10.1162/coli_a_00438).
- Zhu, Y., Lin, Q., Lu, H., Shi, K., Qiu, P., & Niu, Z. (2021). Recommending scientific paper via heterogeneous knowledge embedding based attentive recurrent neural networks. *Knowledge-Based Systems*, 215, Article 106744. <http://dx.doi.org/10.1016/j.knsys.2021.106744>.