Few-Shot Relation Extraction with Dual Graph Neural Network Interaction

Jing Li, Shanshan Feng and Billy Chiu

***Abstract*—Recent advances in relation extraction with deep neural architectures have achieved excellent performance. However, current models still suffer from two main drawbacks (i) they require enormous volumes of training data to avoid model overfitting and (ii) there is a sharp decrease in performance when the data distribution during training and testing shift from one domain to the other. It is thus vital to reduce the data requirement in training and explicitly model the distribution difference when transferring knowledge from one domain to another. In this work, we concentrate on few-shot relation extraction under domain adaptation settings. Specifically, we propose DUALGRAPH, a novel graph neural network based approach for few-shot relation extraction. DUALGRAPH leverages an edge-labeling dual graph (i.e., an instance graph and a distribution graph) to explicitly model the intra-class similarity and inter-class dissimilarity in each individual graph, as well as the instance-level and distribution-level relations across graphs. A dual graph interaction mechanism is proposed to adequately fuse the information between the two graphs in a cyclic flow manner. We extensively evaluate DUALGRAPH on FewRel1.0 and FewRel2.0 benchmarks under four few-shot configurations. The experimental results demonstrate that DUALGRAPH can match or outperform previously published approaches. We also perform experiments to further investigate the parameter settings and architectural choices, and we offer a qualitative analysis.**

***Index Terms*—Few-Shot Learning, Relation Extraction, Graph Neural Network**

I. INTRODUCTION

RELATION Extraction is an important task in natural language processing (NLP), with the objective of determining the semantic relationship between a head entity and a tail entity in text [1]–[5]. Relation extraction is critical in a variety of NLP applications, such as question answering [6], [7] and knowledge base construction [8], [9].

A substantial amount of effort has gone into developing supervised models for relation extraction, including feature engineering based methods [10], [11], kernel based methods [12], [13] and deep neural network based methods [14]–[16]. Despite their general success, these models are nevertheless limited by the need of massive quantities of annotated corpora to avoid overfitting. To reduce the requirement of annotated training data, distantly supervised methods [17] are proposed to exploit large knowledge bases (e.g., Freebase [18] and DBPedia [19]) to automatically label named entities and their relations, and then utilize the annotated text to produce features and train a model [20]–[22]. However, the automatically labeled training data often includes noise and suffers from the long-tail problem [23], where most relations have few training instances (i.e., samples). This makes it difficult for conventional relation extraction approaches to extract the long-tail relations. Another effective approach to reducing data requirement is domain adaptation [24] (also known as transfer learning), which is a technique that uses a large amount of rich data from a source domain to enhance performance in a low-resource target domain. This has also created an overwhelming demand for novel approaches that can extract relations in the low-resource target domain with very few training instances.

Although rule-base methods [25], [26] can alleviate the problem of data scarcity to some extent, they will only capture the occurrences they explicitly cover. An effective method to relieve the aforementioned data shortage is few-shot learning [27], which seeks to categorize fresh test samples after only seeing a few training examples containing supervised information. Few-shot learning has been extensively studied in computer vision, typically including metric-learning based approaches [28], [29] and meta-learning based approaches [30], [31]. Few-shot learning is less prevalent in natural language processing than it is in computer vision. In the last two years, a few attempts have been made to apply few-shot learning to NLP, including query generation [32] and named entity recognition [33]. In particular, Han et al. [1], [2] introduced the Few-Shot relation extraction Dataset (FewRel1.0), consisting of 100 relations obtained from Wikipedia articles and labeled by crowdworkers. FewRel2.0 [2] was built on the FewRel1.0 dataset by including a new development and test set from a quite different domain. In this study, we focus on the more challenging FewRel2.0 for domain adaptation. Table I shows a data example of the N-way K-shot configuration (that is, K labeled instances/samples for each of N classes).

Several studies have contributed to pushing the boundaries of FewRel2.0. For example, Han et al. [1], [2] investigated prototypical networks [29] in relation extraction through learning a metric space where relation extraction can be accomplished by computing distances between prototype representations of each class. Recently, there has also been a surge in interest in graph neural networks (GNNs) to model rational structures on data [34], [35]. Few-shot GNNs [36] were proposed to propagate label information solely based on node-labeling features. Han et al. [2] also explored few-shot GNNs in relation extraction with a convolutional neural network. However, existing models suffer from a significant degradation in performance when the data distribution during training and testing shift from one domain to the other. Therefore, we

J. Li and S. Feng are with Harbin Institute of Technology, Shenzhen 518055, China. E-mail: jingli.phd@hotmail.com; victor_fengss@foxmail.com. (Corresponding author: Shanshan Feng)

B. Chiu is with Department of Computing and Decision Sciences, Lingnan University, Hong Kong. E-mail: billychiu@ln.edu.hk.

Manuscript received xx, 2023; revised xx , 2023.

TABLE I
AN EXAMPLE OF 3-WAY 1-SHOT RELATION EXTRACTION IN FEWREL2.0. A TASK IS MADE UP OF TWO PARTS: A SUPPORT SET AND A QUERY SET. NOTE THAT THE SAMPLES USED IN TRAINING AND TESTING PHASES COME FROM DIFFERENT DOMAINS.

Training Phase (Wikipedia)	
support	**nominated_for**. #1: [Francis Aston]$_{e1}$ was awarded the 1922 [Nobel Prize in Chemistry]$_{e2}$ for this achievement. **country_of_origin**. #1: [Swedish]$_{e2}$ author John Ajvide Lindqvist released his debut horror novel "[Let the Right One In]$_{e1}$" in 2004. **country_of_citizenship**. #1: [Willy Hofmeister]$_{e1}$ was a [German]$_{e2}$ rugby union player who competed in the 1900 Summer Olympics.
query	**country_of_origin**. #1: Sjofn is an album by [Gjallarhorn]$_{e1}$, a band from [Finland]$_{e2}$.
Testing Phase (Biomedicine)	
support	**gene_Ilays_role_in_Irocess**. #1: the genes mthfr, [mtr]$_{e1}$, mtrr, and tcn2 play key roles in [folate metabolism]$_{e2}$. **gene_found_in_organism**. #1: light-at-night exposure can disrupt the [human]$_{e1}$ circadian rhythm via [clock gene]$_{e2}$ expressions. **inheritance_type_of**. #1: the [ivic syndrome]$_{e1}$ is an [autosomal dominant]$_{e2}$ condition affecting mainly the upper limbs.
query	**gene_Ilays_role_in_Irocess**. #1: [integrins]$_{e1}$ mediate [cell adhesion]$_{e2}$ to the extracellular matrix.

envision that an approach can explicitly model distribution relations among samples to alleviate the domain shift.

Inspired by the success of the edge-labeling GNN [34], [35] in computer vision, we propose DUALGRAPH, a GNN-based approach for few-shot relation extraction in NLP. An instance graph (where each node stands for an instance) is constructed to model the instance-level relation of one instance to another instance. A distribution graph (where each node is generated by pairwise comparison) is built to represent the distribution-level relations between one instance and all other instances. Figure 1 illustrates the instance and distribution graphs. In addition to node features, edge features are utilized to explicitly model both the inter-class dissimilarity and intra-class similarity in each graph. A dual graph interaction mechanism is proposed to iteratively update node and edge features in a cyclic flow manner. More specifically, the node features flow into the edge features in each individual graph. At the same time, the edge features in one graph flow into the node features in another graph. The cyclic flow update leads to adequate fusion not only between node and edge features, but also between instance-level and distribution-level relations. In summary, this study makes four major contributions:

- We propose DUALGRAPH, a novel GNN based approach for few-shot relation extraction. Notably, DUALGRAPH leverages an edge-labeling dual graph to explicitly model the inter-class dissimilarity and intra-class similarity in each individual graph, as well as the instance-level and distribution-level similarities across graphs, simultaneously. A cyclic flow update strategy is proposed to adequately fuse the information between two graphs.
- We extensively evaluate DUALGRAPH on FewRel1.0 and FewRel2.0 benchmarks under four few-shot configurations. The results show that DUALGRAPH can match or outperform previously published approaches. In particular, DUALGRAPH achieves at least comparable results to the current state-of-the-art on FewRel1.0 and delivers the state-of-the-art performance on FewRel2.0 at the time of writing.
- We also did experiments to empirically investigate the rational of architectural choices and parameter settings, and we offer a qualitative analysis.

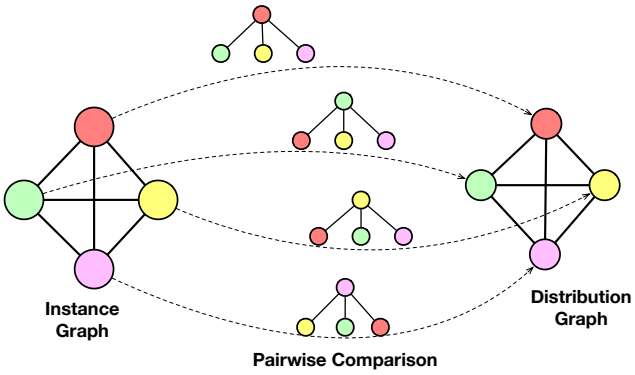


Fig. 1. Illustration of instance and distribution graphs (best viewed in color). The distribution graph is constructed from the instance graph by a pairwise comparison manner.

The rest of this study is structured as follows: Section II examines research on relation extraction, few-shot learning, and graph neural networks. Section III presents the methodology of DUALGRAPH. In Section IV, we evaluate DUALGRAPH on a benchmark and conduct a further analysis study. Section V concludes this work.

II. RELATED WORK

In this section, we discuss related work along three lines: relation extraction, few-shot learning, and graph neural networks.

A. Relation Extraction

The task of extracting semantic relation from text, which generally exist between two or more named entities, is known as relation extraction [37] . Relation extraction can be framed as extracting the triple (e_1, r, e_2), which means that the head entity e_1 has relation r with another tail entity e_2. If the mentions of entities are known or given, relation extraction becomes relation classification where the relation r is classified into one of several predefined classes. In this paper, we fellow this kind of relation extraction. If the mentions and relation classes are both unknown, the paradigm is generally referred to open information extraction [38], [39], where r is often expressed with a free phrase. We concentrate on summarizing relation extraction works in this section. In summary, there are three common paradigms for relation extraction: *rule-based approaches*, *supervised approaches* and *distantly supervised approaches*.

Rule-based approaches usually rely on handcrafted rules, which are designed based on syntactic-lexical patterns. Reiss et al. [25] developed an algebraic method for rule-based IE that

uses query optimization to overcome scalability issues. Bollegala et al. [26] presented a sequential co-clustering method that clusters distinct lexical-syntactic patterns that define a certain semantic relation, as well as diverse entity pairings that share that semantic relation. Rule-based approaches tend to have high precision and can be tailored to specific domains. However, these approaches often suffer from low-recall and a great deal of manual labor is needed to develop all conceivable rules. For *supervised approaches*, feature-based, kernel-based, and neural-based methods are the three mainstream branches. Manually engineered features play an important role in feature-based methods. Kambhatla et al. [10] used maximum entropy approach to integrate various semantic, syntactic, lexical and characteristics gathered from text. Zhou et al. [11] studied the use of SVM to include different semantic, syntactic, and lexical features in relation extraction. Kernel method is a kind of nonparametric density estimation approach to calculating a kernel function (i.e., a similarity measure) between data instances. Some examples include the sequence kernel [12], dependency tree kernel [40], dependency graph path kernel [41] and composite kernel [13]. The key advantage of neural-based methods [14]–[16] is their strong capability for semantic composition and feature learning, which are enabled by both neural processing and distributed vector embedding, without the need of human feature engineering. Finally, the concept behind *distantly supervised approaches* is to leverage huge knowledge bases (e.g., DBPedia and Freebase) to automatically identify named entities in text and then utilize the labeled data to produce latent embeddings and learn a model [20]–[22], [42]. However, automatically labeled training data often includes noise. Some studies have already contributed efforts to noise reduction [43], [44].

B. Few-Shot Learning

The goal of few-shot learning is to categorize fresh test (unseen) samples based on only a few annotated instances with supervised information which have been viewed [27], [45]. Few-shot learning has received a great deal of interest in the field of computer vision. One mainstream branch of few-shot learning lies in using metric learning to learn the distance distributions across classes. Koch et al. [28] provided a method for accomplishing one-shot learning that involves first learning deep convolutional siamese neural networks for verification using the weighted $L1$ distance. Vinyals et al. [46] introduced matching networks, taking use of recent techniques in memory and attention to allow for fast learning. In an embedding space, matching networks may be thought of as a weighted nearest-neighbor classifier. Snell et al. [29] introduced prototypical networks which make image classification by calculating representation distances between class prototypes in the learned metric space. Sung et al. [47] proposed the idea of relation networks, which compares training instances within episodes (i.e., few-shot environment) in a learned distance metric space.

Another flourishing branch of few-shot learning approaches focuses on optimizing model parameters to encourage transferable knowledge between tasks through meta-learning. Simply, meta-learning [48], [49] seeks to build a generic model that is able to fast adopt to unseen tasks given very few annotated instances, without having to be relearned from the ground up. MAML [30], introduced by Finn, tackles the few-shot learning problem by meta-learning a generic parameter initialization. Such an initialization can be fine-tuned at test time with a few gradient updates utilizing a limited number of annotated samples from target domains. Mishra et al. [31] introduced the simple neural attentive learner (SNAIL), which makes use of an innovative mix of causal attention and temporal convolutions.

C. Few-Shot Relation Extraction

Few-shot learning is less prevalent in natural language processing than in computer vision. In particular, few-shot relation extraction aims to extract relations between entities in textual data using a minimal number of annotated relation examples (e.g., 1 or 5) [50], [51]. Few studies were committed to the use of few-shot learning in relation extraction in recent years. Han et al. [1], [2] presented the first few-shot relation extraction dataset (FewRel1.0), which is made up of 100 semantic relations (totally including 70, 000 instances) obtained from Wikipedia articles. REGRAB [52] is a Bayesian meta-learning approach to learn the posterior distribution of the prototype vectors of relations. HCRP [53] is a relation-prototype contrastive learning approach which generates informative prototypes to model small inter-relation variations. Similarly, CP [54] is an entity-masked contrastive pre-training framework for better understanding textual context and entity types. CTEG [55] is proposed to use the entity-guided attention, confusion-aware training based on Transformer encoders. MapRE [56] leverages the label-agnostic and label-aware knowledge in pretraining to improve the model performance in low-resource relation extraction tasks. Some other representative approaches include hybrid attention-based prototypical networks [57], distributional similarity training [58] and the multi-level aggregation and matching networks [3]. Recently, ConceptFERE [59] is proposed to provide clues for relation prediction and boost the relation classification performance by leveraging the external knowledge base, i.e., Concept Graph. KEFDA [60] incorporates general and domain-specific knowledge graphs (i.e., WikiData and UMLS) into the model to improve its domain adaptability. Our approach differs from previous methods in that (1) our approach is a novel graph-based method specifically designed for domain adaptation, rather than a meta-learning based fast algorithm. (2) our approach does not require the computationally expensive pre-training on large-scale corpora. (3) our approach does not rely heavily on external knowledge bases such as Concept Graph and WikiData.

D. Graph Neural Networks

Graph neural networks (GNNs) utilizes the message passing mechanism to capture the graph dependency [61]–[63]. GNNs were first proposed to process the data represented in graph domains with neural networks [61], [64], [65].

Recently, graph autoencoders (GAEs) [66], [67], recurrent graph neural networks (RecGNNs) [68], [69], convolutional

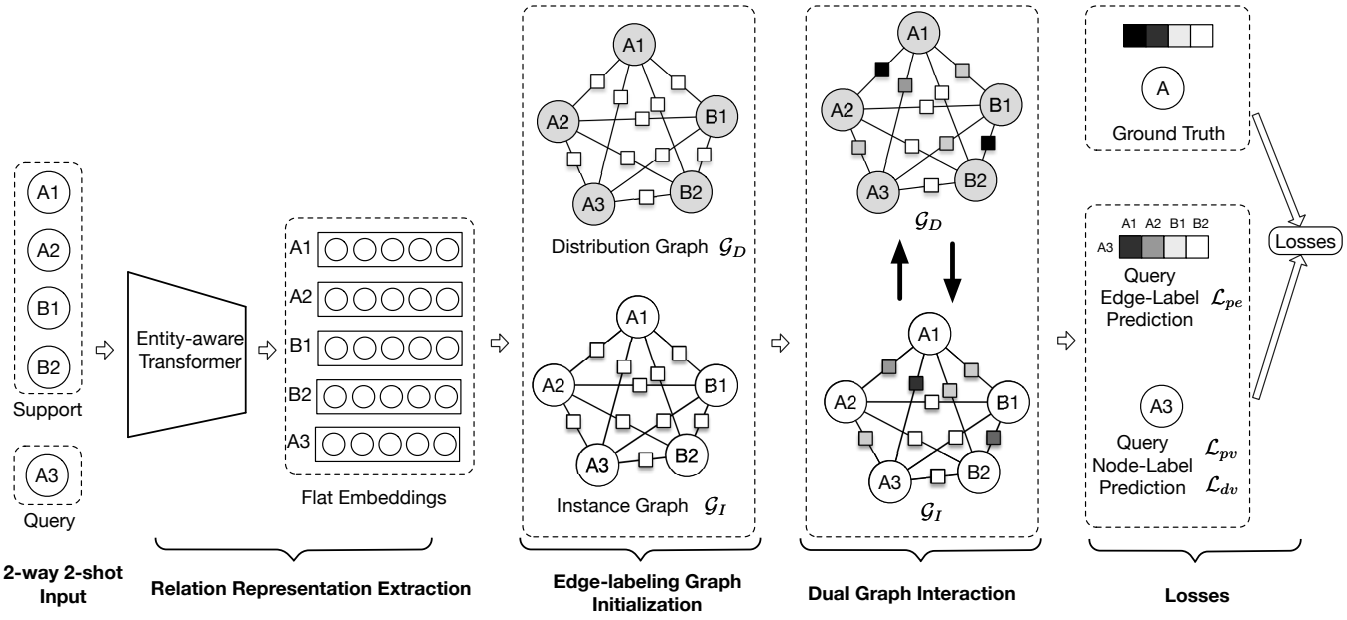


Fig. 2. An overview of our proposed DUALGRAPH. A 2-way 2-shot few-shot learning problem is used as an example in this illustration. "A" and "B" indicate two different relations. "A1" and "A2" represent two different instances of the relation of "A". The proposed approach is composed of four components: relation representation extraction, edge-labeling graph initialization, dual graph interaction and loss generation.

graph neural networks (ConvGNNs) [70], [71] and spatial-temporal graph neural networks (STGNNs) [72], [73] have demonstrated ground-breaking performance in many domains. A few approaches [36], [74] have studied the application of GNNs in few-shot leaning based on the node-labeling approach, in which each node is embedded by vector features. Qu et al. [52] proposed a Bayesian approach with meta-learning for learning the posterior distribution of relations, where the prototype embeddings are parameterized with GNNs on a global graph. Jatin et al. [75] proposed a method for assigning a probability measure to each graph relying on its normalized Laplacian. In addition, Kim et al. [34] introduced the edge-labeling GNN under few-shot settings. In their approach, the node and edge representations are alternatively updated so that both the inter-cluster dissimilarity and intra-cluster similarity can be directly modeled by edge-labeling graph. Yang et al. [35] proposed the distribution propagation graph network (DPGN) for image classification, where the distribution- and instance- level information are explicitly modeled in a dual graph. However, most of existing studies are designed for few-shot image classification and there is no work addressing few-shot relation extraction with edge-labeling GNNs.

III. METHODOLOGY

In this section, we will first formally define the task of few-shot relation extraction. Following that, we will give a step-by-step description of our DUALGRAPH approach.

A. Problem Definition: Few-Shot Relation Extraction

An annotated dataset consisting of three partitions (i.e., a training part $\mathbb{D}^{train}$, a development part $\mathbb{D}^{dev}$ and a test part $\mathbb{D}^{test}$) is available in few-shot relation extraction, . The instance in each partition can be indicated by (s, p, y), where s represents a sentence, p indicates the positions of the head and tail named entities in the sentence, and y indicates the relation between the head and tail named entities. The relation class spaces among these three partitions are often disjoint. In this work, we investigate a more difficult case, i.e., the relation class spaces between the test set and the training set are not only mutually exclusive, but also come from different domains. In few-shot relation extraction, we seek to create a model $f : s \mapsto \hat{y}$ that is able to map a sentence (s, p) with a true label $y \in \mathcal{Y}$ to the prediction $\hat{y} \in \mathcal{Y}$ using very few annotated samples. A task $\mathcal{T}_i$ is a collection of instances, which is composed of a support set $\mathcal{D}^{spt}_{\mathcal{T}_i}$ and a query set $\mathcal{D}^{qry}_{\mathcal{T}_i}$ (note that $\mathcal{D}^{spt} \cap \mathcal{D}^{qry} = \emptyset$). During the training phase, the real (i.e., true) labels of $\mathcal{D}^{spt}_{\mathcal{T}_i}$ and $\mathcal{D}^{qry}_{\mathcal{T}_i}$ are both available for the source task $\mathcal{T}_i \in \mathbb{D}^{train}$. During the testing phase, a new (i.e., unseen during training) target task $\mathcal{T}_j \in \mathbb{D}^{test}$ only comprises a few annotated samples $\mathcal{D}^{spt}_{\mathcal{T}_j}$. The ultimate objective is to predict the labels for $\mathcal{D}^{qry}_{\mathcal{T}_j}$, given a few samples in $\mathcal{D}^{spt}_{\mathcal{T}_j}$.

The N-way K-shot configuration has been extensively adopted in recent studies on few-shot learning, where $\mathcal{D}^{spt}_{\mathcal{T}_i}$ and $\mathcal{D}^{spt}_{\mathcal{T}_j}$ typically both comprise K instances (i.e., K-shot) for each of N randomly selected relations (i.e., N-way) for support sets. $\mathcal{D}^{qry}_{\mathcal{T}_i}$ and $\mathcal{D}^{qry}_{\mathcal{T}_j}$ usually both include one sample for each of the N classes. In summary, we expect that a model trained on the tasks $\mathcal{T}_i \in \mathbb{D}^{train}$ will work well on the tasks $\mathcal{T}_j \in \mathbb{D}^{test}$.

B. The DUALGRAPH Approach

1) Overview of DUALGRAPH*:* As shown in Figure 2, DUALGRAPH is made up of four components: relation representation extraction, edge-labeling graph initialization, dual graph interaction and loss generation. Given a sentence with

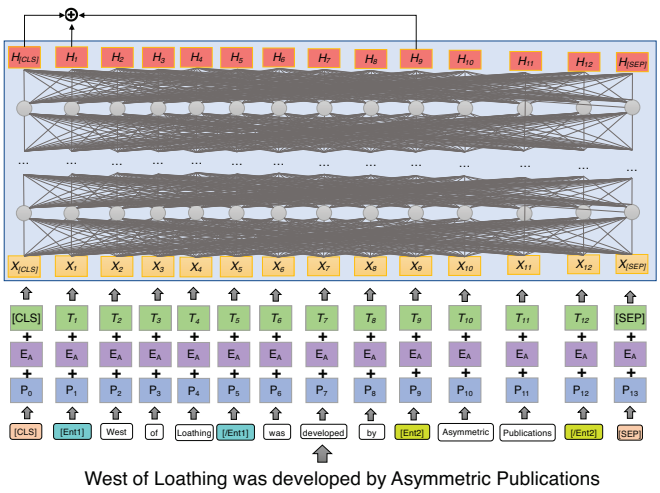


Fig. 3. Architecture of extracting relation representations from entity-aware Transformers. Four entity markers, `[Ent1]`, `[\Ent1]`, `[Ent2]`, `[\Ent2]`, are inserted into the original sentence. The final representation is formulated by concatenating the hidden states of `[CLS]`, `[Ent1]` and `[Ent2]`.

the positions of head and tail entities, the relation representation extraction module aims to produce contextualized relation representations from the text. The edge-labeling graph initialization module aims to construct two graphs (i.e., an instance graph and a distribution graph) and performs the node and edge representation initialization. A node of instance graph stands for an instance. A node of distribution graph is generated by a pairwise comparison (one instance to all other instances) manner (see Figure 1). Notably, the edge-labeling graph is able to explicitly model both the inter-class dissimilarity and intra-class similarity in each individual graph. Next, the dual graph interaction module then aims to propagate label information from annotated samples to unannotated samples, by considering the information at both an instance and distribution level. Specifically, the instance graph and the distribution graph are iteratively updated with each other in a cyclic flow manner (see Algorithm 1), which leads to adequate fusion between instance-level and distribution-level relations across the two graphs. Finally, the loss generation module aims to compute the classification loss based on the node and edge predictions.

2) Relation Representation Extraction: As shown in Figure 3, a relation representation is composed of three parts: a token (T), segment (E) and position (P) embedding. First, a special placeholder token (`[CLS]`) is added as the first token of every sequence and an `[SEP]` is appended as the last token. In addition, the entity markers `[Ent1]` and `[\Ent1]` are inserted into the original sentence and used to indicate the head entity. Similarly, the entity markers `[Ent2]` and `[\Ent2]` are used to indicate the tail entity. These entity markers provide rich entity information when learning language representation models [76]. Token embeddings are based on WordPiece embeddings with a 30,000 token vocabulary. Segment embeddings are learned to identify whether a token comes from sentence A or sentence B. For relation extraction in this study, the input sequence is always a single sentence. Therefore, all tokens in input sequence are from sentence A (i.e., E_A in Figure 3). Position embeddings are learned to capture the location information for each token. The input representation of a token is constructed by summing these three types of embeddings. Then, the input representation is passed into Transformer layers to produce contextualized relation representations.

Formally, given a modified sequence $\boldsymbol{W} = (W_1, W_2, \ldots, W_N)$ of length N (i.e., special tokens and entity markers have already been added into $\boldsymbol{W}$), let W_n represent its n-th token. The input can be denoted as $\mathbf{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N)$ after the input embedding layer. Let H be the hidden dimension of Transformers. Then, Transformer layers are used to encode the sequence context, yielding hidden states $\boldsymbol{h} = \{\boldsymbol{h}_1, \boldsymbol{h}_2, ..., \boldsymbol{h}_N\} \in \mathbb{R}^{N \times H}$. Next, the final relation representation $\boldsymbol{h}_r \in \mathbb{R}^{3 \times H}$ for $\boldsymbol{W}$ is formulated by concatenating the hidden states of `[CLS]`, `[Ent1]` and `[Ent2]`:

$$\boldsymbol{h}_r = \boldsymbol{h}_{[\text{CLS}]} \oplus \boldsymbol{h}_{[\text{Ent1}]} \oplus \boldsymbol{h}_{[\text{Ent2}]} \tag{1}$$

where $\oplus$ stands for a concatenation operation.

3) Edge-Labeling Graph Initialization: For a given N-way K-shot task $\mathcal{T}$, we construct two graphs based on all samples in the support and query sets: an instance graph $\mathcal{G}_I = (\mathcal{V}_I, \mathcal{E}_I)$ where $\mathcal{V}_I$ and $\mathcal{E}_I$ denote nodes and edges in $\mathcal{G}_I$, and a distribution graph $\mathcal{G}_D = (\mathcal{V}_D, \mathcal{E}_D)$ where $\mathcal{V}_D$ and $\mathcal{E}_D$ denote nodes and edges in $\mathcal{G}_D$.

Let v_I^i be the node representation of $\mathcal{V}_I^i$ and e_I^{ij} be the edge representation of $\mathcal{E}_I^{ij}$. Our approach involves L generation layers to propagate the dual graph. For the first generation $l = 0$, the node representation $v_I^i \in \mathbb{R}^{3 \times H}$ is initialized by the output of entity-aware Transformers,

$$v_I^{0,i} = \boldsymbol{h}_r^i \tag{2}$$

The edge $e_I^{ij} \in \mathbb{R}$ indicates the instance similarity between the node $\mathcal{V}_I^i$ and the node $\mathcal{V}_I^j$. It is initialized by

$$e_I^{0,ij} = f_{MLP1}(sim < v_I^{0,i}, v_I^{0,j} >) \tag{3}$$

where $sim < \cdot, \cdot >$ is a similarity function (i.e., we use the Euclidean distance in our implementation) and f_{MLP1} is the feature transformation network that projects the similarity to a scale [35]. The distribution graph $\mathcal{G}_D$ is constructed based on the instance similarities $\mathcal{E}_I^{ij}$ in $\mathcal{G}_I$. The node representation v_D^i aims to aggregate all instance similarities between the sample i and other support samples in the task $\mathcal{T}$. That is, v_D^i works in a pairwise comparison manner (see Figure 1). For the N-way K-shot setting, $v_D^i \in \mathbb{R}^{NK}$ is an $N \times K$ dimensional features, where the j-th element indicates the distribution relation between instances i and j. The node v_D^i is initialized by

$$v_D^{0,i} = \oplus|_{j=0}^{NK} \delta(y_i, y_j) \tag{4}$$

where $\oplus$ stands for a concatenation operation, and $\delta(.)$ is the Kronecker function whose value is equal to 1 when the labels of two nodes are the same, i.e., $y_i = y_j$ and 0 otherwise. The edge representation $e_D^{ij} \in \mathbb{R}$ indicates the distribution

similarity between the node $\mathcal{V}_D^i$ and the node $\mathcal{V}_D^j$. It is initialized by

$$e_D^{0,ij} = f_{MLP2}(sim < v_D^{0,i}, v_D^{0,j} >) \tag{5}$$

4) Dual Graph Interaction: Similar to the existing distribution-based method [35], the instance graph $\mathcal{G}_I$ and the distribution graph $\mathcal{G}_D$ are iteratively updated with each other in a cyclic flow manner. In each respective graph, the node representation flows into the edge representation. At the same time, the edge representation in one graph flows into the node representation in another graph. More specifically, the edge and node representations are updated by the flow $v_I^{l-1} \rightarrow e_I^l \rightarrow v_D^l \rightarrow e_D^l \rightarrow v_I^l$. Note that the update order is fixed because of the way of initialization in Equation 2 and the way of graph construction. This update mechanism leads to adequate fusion between instance-level and distribution-level relations across the two graphs.

In detail, for the l-th generation layer, the edge and node representations in the dual graph are updated by

$$e_I^{l,ij} = f_{MLP3}(sim < v_I^{l-1,i}, v_I^{l-1,j} > \cdot e_I^{l-1,ij}) \tag{6}$$

$$v_D^{l,i} = f_{MLP4}(\oplus|_{j=0}^{NK} e_I^{l,ij}, v_D^{l-1,i}) \tag{7}$$

$$e_D^{l,ij} = f_{MLP5}(sim < v_D^{l,i}, v_D^{l,j} > \cdot e_D^{l-1,ij}) \tag{8}$$

$$v_I^{l,i} = f_{MLP6}(\sum_j (e_D^{l,ij} \cdot v_I^{l-1,j}), v_I^{l-1,i}) \tag{9}$$

where f_{MLP*} are different feature transformation networks, and l-1 denotes the (l-1)-th generation layer. In summary, the dual graph interaction enables our approach to not only fuse the node and edge representations in each individual graph, but also inject the distribution-level information into the instance-level representation across graphs.

5) Loss Generation: First, for a given query sample, we can obtain the relation class prediction based on the edge representation in the instance graph. Second, we can also predict query edge labels in the instance graph to explicitly model both the inter-class dissimilarity and intra-class similarity. Finally, we can obtain the relation class prediction for query sets based on the edge representation in the distribution graph. Therefore, the overall loss consists of the following three parts.

Instance Query Node Loss. The prediction probability of node $\mathcal{V}_I^{l,i}$ can be expressed as follows:

$$P(\hat{y}_i|v_I^{l,i}) = \text{softmax}(\sum_j^{NK} e_I^{l,ij} \cdot y_j) \tag{10}$$

Thus, the instance query node loss can be formulated as:

$$\mathcal{L}_{pv}^l = -\sum_i \log P(\hat{y}_i|v_I^{l,i}) \tag{11}$$

Instance Query Edge Loss. The query edge loss in the instance graph can be formulated as

$$\mathcal{L}_{pe}^l = BCELoss(e_I^{l,ij}, y_{ij}) \tag{12}$$

where $BCELoss$ is the binary cross entropy function, and y_{ij} is the ground-truth edge label between the sample i and the sample j. Note that y_{ij} is equal to 1 for $y_i = y_j$ and 0 otherwise.

Algorithm 1: The inference process of DUALGRAPH

Input: A test N-way K-shot task $\mathcal{T}$, consisting of a support set $\mathcal{D}_\mathcal{T}^{spt}$ and a query set (Q samples) $\mathcal{D}_\mathcal{T}^{qry}$

Output: Relation class predictions for the query set $\mathcal{D}_\mathcal{T}^{qry}$

1 Relation representation extraction by entity-aware Transformers;
2 Dual graph construction and initialization, $\mathcal{G}_I$ and $\mathcal{G}_D$;
/* Update node and edge features via dual graph interaction */
3 **for** $l = 1, ..., L$ **do** // generation layers
4 $e_I^l \leftarrow$ InstanceEdgeUpdate(e_I^{l-1}, v_I^{l-1});
5 $v_D^l \leftarrow$ DistributionNodeUpdate(e_I^l, v_D^{l-1});
6 $e_D^l \leftarrow$ DistributionEdgeUpdate(e_D^{l-1}, v_D^l);
7 $v_I^l \leftarrow$ InstanceNodeUpdate(e_D^l, v_I^{l-1});
/* Query sample prediction */
8 $\{\hat{y}_i\}_i^Q \leftarrow \text{softmax}(\sum_j^{NK} e_I^{L,ij} \cdot y_j)$

Distribution Query Node Loss. Likewise, the prediction probability of node $\mathcal{V}_D^{l,i}$ can be formulated as follows:

$$P(\hat{y}_i|v_D^{l,i}) = \text{softmax}(\sum_j^{NK} e_D^{l,ij} \cdot y_j) \tag{13}$$

Thus, the distribution query node loss can be formulated as:

$$\mathcal{L}_{dv}^l = -\sum_i \log P(\hat{y}_i|v_D^{l,i}) \tag{14}$$

Finally, the overall objective is a weighted sum of these three parts:

$$\mathcal{L} = \sum_l^L (\alpha \mathcal{L}_{pv}^l + \beta \mathcal{L}_{pe}^l + \gamma \mathcal{L}_{dv}^l) \tag{15}$$

where the symbol L represents the total number of generation layers; α, β and γ are weights for controlling the trade-off among the three losses.

6) Algorithm Flow: The inference procedure of DUALGRAPH is detailed in Algorithm 1. Given a few-shot learning task $\mathcal{T}$, DUALGRAPH predicts the Q samples in the query set through the dual graph interaction. Lines 4-7 clearly show the cyclic flow with different font colors. First, the node representations in the instance graph flow into edge representations in this instance graph. Then the edge representations in the instance graph flow into the distribution graph. Next, the node representations in the distribution graph flow into edge representations in this distribution graph. Finally, the edge representations in the distribution graph flow into the instance graph to refine node representations in instance graph. Line 8 shows that predictions are made by the updated dual graph. The query sample prediction (i.e., relation extraction result) can simply be obtained from the edge labels (the last generation layer L) in the instance graph.

TABLE II
STATISTICS OF DATASETS.

	FewFel1.0		FewFel2.0	
	# Relations	# Instances	# Relations	# Instances
Train	64	44,800	64	44,800
Dev	16	11,200	10	1,000
Test	20	14,000	15	1,000

IV. EXPERIMENTS

The experimental setups are presented first in this section. Then, we present our experiments and findings on FewRel2.0. Finally, we present an ablation study, parameters analysis and qualitative analysis.

A. Setups

1) Dataset and Metrics: To the best of our knowledge, FewRel1.0 [1] is only one benchmark for few-shot relation extraction. It comprises 100 semantic relations (totally including 70,000 instances) obtained from Wikipedia articles and manually labeled by crowdworkers. These 100 relations are from the same domain and each relation includes 700 instances. Moreover, they are divided into 20, 16 and 64 relations for test, development and training, respectively. FewRel2.0 [2] was built on the FewRel1.0 dataset by including a new development and test set from a quite different domain. The training set of FewRel2.0 is the same as the original FewRel training set. However, the test set of FewRel2.0 is from the PubMed corpus[1] (biomedical), and consists of 15 newly annotated relations, each of which includes 100 instances. Note that the test of FewRel2.0 is hidden for fair comparison. The development set of FewRel2.0 is constructed based on the SemEval-2010 task 8 dataset [77]. The development set includes 10 relations which are composed of 1000 instances. In this study, we mainly focus on FewRel2.0 because we investigate a more challenging scenario, i.e., the training and test sets are from different domains. Table II summarizes the statistics of datasets.

We conduct experiments on four few-shot configurations: 5-way 1-shot, 5-way 5-shot, 10-way 1-shot and 10-way 5-shot. To preserve the integrity of the test results, the test set of FewRel2.0 is not released to the public. For the test accuracy, we submit our models to the official Leaderboard[2] to obtain the official experimental results on the hidden test set. However, the development set is available for the public. Therefore, we conduct experiments on 1000 randomly selected tasks from the development set for the ablation study and parameters analysis.

2) Baseline Methods: DUALGRAPH is compared to the following competitors:

- **GNN** - GNN [36] considers all query and support instances as nodes to formulate a GNN where relational structures can be leveraged with a task-driven message passing algorithm. As a result, a query instance receives information from its support set in the graph to conduct classification. The instances are represented by convolutional neural networks (CNNs) [15].
- **Prototypical Networks** - Prototypical Networks [29] are founded on the concept that each class has a prototype representation in a latent embedding space. More specifically, prototypical Networks perform relation extraction by calculating the distances between two prototype representations in a learned metric space. We use two encoders for relation representation extraction: CNN and BERT [78]. In addition, we use an adversarial training strategy to learn domain-invariant representations in the prototypical networks. The adversarial methods are denoted as Proto-ADV (CNN) and Proto-ADV (BERT).
- **BERT-PAIR** - BERT-PAIR [2] is a pair-wise method based on BERT. It pairs each query sample with all the support samples in turn, and then concatenates each pair as one sequence which is fed into BERT.
- **HCRP** - HCRP [53] is a hybrid contrastive relation-prototype approach which pulls instances of the same relation class closer in the representation space while pushing dis-similar ones apart.
- **FAEA** - FAEA [79] is a function words adaptively enhanced attention framework for few-shot inverse relation classification, in which a hybrid attention model is designed to attend class related function words based on meta-learning.
- **Anonymous Models on Leaderboard** - We also compare our proposed model with recent anonymous models listed on the official FewRel2.0 Leaderboard.

3) Implementation Details: Our proposed DUALGRAPH is trained with the AdamW optimizer. The weight decay is set to $1e-6$ and the initial learning rate is set to $1e-5$. Note that we do not adopt any learning rate decay strategy. The epsilon for the AdamW optimizer is set to $1e-8$. The number of generation layers (i.e., L in our model) is 6. The max gradient normalization is 1.0. We adopts a fixed L2 regularization of $1e-6$. The dropout is set to 0.1 after all recurrent, convolutional and Transformers layers. We use `bert-base-uncased` to initialize our Transformer layers. For the weights of α, β and γ, a grid search strategy is adopted to determine the optimal values. Our proposed model is written in PyTorch and evaluated on NVIDIA Tesla V100 GPUs.

B. Experimental Results

1) Performance on FewRel2.0 Domain Adaption: Table III shows the accuracies of different models on the test set of FewRel2.0 under four configurations. The upper half of the table is composed of the baselines implemented in the previous work [2]. The bottom half consists of recent anonymous models listed on the official Leaderboard (on the date we submitted our model and wrote up this paper).[3] The following observations are made:

First, DUALGRAPH delivers the state-of-the-art performance, outperforming the six baseline models and four anony-

[1]https://www.ncbi.nlm.nih.gov/pubmed/
[2]https://thunlp.github.io/2/fewrel2_da.html

[3]Note that we do not involve the KEFDA approach [60] in Table III because it leverages huge external resources (i.e., WikiData and UMLS), while our approach operates without the need for any external resource.

TABLE III
ACCURACIES (%) OF DIFFERENT MODELS ON THE FEWREL2.0 TEST SET UNDER FOUR FEW-SHOT CONFIGURATIONS.

Models	5-way 1-shot	5-way 5-shot	10-way 1-shot	10-way 5-shot
GNN (CNN) [36]	27.94	29.33	16.44	18.26
Proto (CNN) [29]	35.09	49.37	22.98	35.22
Proto (BERT) [29]	40.12	51.50	26.45	36.93
Proto-ADV (BERT) [29]	41.90	54.74	27.36	37.40
Proto-ADV (CNN) [29]	42.21	58.71	28.91	44.35
BERT-PAIR [2]	67.41	78.57	54.89	66.85
HCRP (BERT) [53]	76.34	83.03	63.77	72.94
FAEA (BERT) [79]	73.58	90.10	62.98	80.51
Anonymous Groundhog	67.23	82.09	54.32	71.01
Anonymous Python	66.41	83.52	51.85	73.60
Anonymous Pony	76.71	86.69	66.72	78.46
Anonymous PAMN	77.54	90.40	65.98	82.03
DUALGRAPH (ours)	**80.11**	**91.01**	**73.89**	**82.34**

TABLE IV
ACCURACIES (%) OF DIFFERENT MODELS ON THE FEWREL1.0 VALIDATION SET UNDER FOUR FEW-SHOT CONFIGURATIONS. ♣ INDICATES THAT RESULTS ARE FROM [52]. RESULTS WITH ♢ ARE REPORTED IN FEWREL OFFICIAL LEADERBOARD. RESULTS WITH △ ARE REPORTED IN [53].

Models	5-way 1-shot	5-way 5-shot	10-way 1-shot	10-way 5-shot
MAML [30] ♣	82.93	86.21	73.20	76.06
MTB [58] ♣	84.61	88.76	75.22	80.15
BMAML [80] ♣	85.80	89.71	76.66	81.34
BERT-PAIR [1] ♢	85.66	89.48	76.84	81.76
Proto-BERT [29] ♣	82.92	91.32	73.24	83.68
CTEG-BERT [55] △	84.72	92.52	76.01	84.89
REGRAB [52] ♣	87.95	92.54	80.26	86.72
HCRP [53] △	**90.90**	93.22	**84.11**	87.79
DUALGRAPH (ours)	88.71	**93.92**	81.79	**88.05**

mous models by significant margins. More specifically, DUALGRAPH outperforms GNN (CNN) by absolute improvements of 52.17%, 61.68%, 57.45% and 64.08% for the four few-shot configurations. Moreover, DUALGRAPH outperforms BERT-PAIR by absolute improvements of 12.7%, 12.44%, 19.01% and 15.49% for the four few-shot configurations. It is worth mentioning that DUALGRAPH outperforms recent FAEA model by 6.53, 0.91, 10.91, and 1.83 points for the four few-shot configurations. Notably, DUALGRAPH beats the strongest model (Anonymous PAMN) which was recently submitted to the official Leaderborad by absolute improvements of 2.57%, 0.61%, 7.91% and 0.31% for the four few-shot configurations, respectively. This is due to the fact that DUALGRAPH is effective in propagating label information from labeled instances to unlabeled instances with a novel edge-labeling dual graph, where the cyclic flow update mechanism leads to adequate fusion not only between node and edge features in each individual graph, but also between instance-level and distribution-level information across graphs.

Second, the accuracies in the 5-shot settings are significantly better than in the 1-shot settings, i.e., 91.01% vs. 80.11% for the 5-way setting and 82.34% vs. 73.89% for the 10-way setting. This is understandable because the 5-shot setting provides more training instances than the 1-shot setting.

Third, the naive GNN method with CNN encoders (i.e., GNN (CNN)) is not effective for few-shot relation extraction. In contrast, our GNN-based method is more effective because it benefits from not only entity-aware Transformers, but also the dual graph interaction. From the performance of baselines, we can observe that dynamic Transformer embeddings are more effective than static CNN-based embeddings. This is because dynamic embeddings are commonly contextualized and are more informative and helpful for relation extraction.

2) Performance on FewRel1.0: Although our approach specifically focuses on few-shot relation extraction under domain adaption settings, we also investigate the capability of our approach on the normal setting (i.e., no domain adaption). Following previous studies [52], [53], [55], we conduct experiments on the validation set of FewRel1.0. Table IV reports of accuracy scores of different models on the FewRel1.0 validation set. We make the following observations: (1) DUALGRAPH achieves comparable performance compared to the second best baseline, HCRP, on FewRel1.0. DUALGRAPH outperforms HCRP on 5-shot settings and HCRP outperforms DUALGRAPH on 1-shot settings. However, our approach significant outperforms HCRP on all domain adaptation settings as shown in Table III. (2) DUALGRAPH significantly outperforms other baselines (except HCRP) by large margins on all few-shot settings. More specifically, DUALGRAPH achieve relative improvements from 0.9% to 7.0% on 5-way 1-shot, relative improvements from 1.5% to 8.9% on 5-way 5-shot, relative improvements from 1.9% to 11.7% on 10-way 1-shot,

TABLE V
ABLATION STUDY ON THE DEVELOPMENT SET OF FEWREL2.0. THE SYMBOL "-" INDICATES THE ACCURACY DEGRADATION.

Models	5-way 1-shot	5-way 5-shot	10-way 1-shot	10-way 5-shot
Ours	80.54	89.68	73.75	83.16
$\boldsymbol{h}_r = \boldsymbol{h}_{[\text{CLS}]}$ without entity markers	-7.56	-9.34	-8.27	-10.78
$\boldsymbol{h}_r = \boldsymbol{h}_{[\text{CLS}]}$ with entity markers	-3.43	-4.26	-4.23	-5.73
$\boldsymbol{h}_r = \boldsymbol{h}_{[\text{Ent1}]} \oplus \boldsymbol{h}_{[\text{Ent2}]}$	-1.31	-2.92	-1.98	-3.44
Remove edge-labeling (i.e., node-only)	-4.36	-6.34	-5.89	-7.83
Remove $\mathcal{G}_D$ (i.e., instance-only)	-5.74	-6.87	-6.83	-8.56

and relative improvements from 1.6% to 15.8% on 10-way 5-shot. We attribute this to the fact of the effectiveness of our edge-labeling and dual graph interaction strategies in graph learning. (3) On the 1-shot settings, our approach exhibits inferior performance compared to HCRP. The possible reason is that HCRP leverages external knowledge such as relation textual descriptions to enhance its overall performance on the extremely low-resource settings (i.e., 1-shot). In contrast, our approach relies solely on provided training data. Incorporating external knowledge into our approach is the future work.

C. Further Analysis

First, we provide an ablation research to validate several architectural choices in this section. Then we investigate how the total number of generation layers affects the model performance. Finally, we also present a qualitative analysis to showcase positive and negative cases.

1) Ablation Study: An ablation analysis on the development set of FewRel2.0 is presented in Table V. Our full architecture adopts the relation representation of $\boldsymbol{h}_r = \boldsymbol{h}_{[\text{CLS}]} \oplus \boldsymbol{h}_{[\text{Ent1}]} \oplus \boldsymbol{h}_{[\text{Ent2}]}$. In addition, it adopts the edge-labeling strategy for both the instance graph and distribution graph. There are five ablated variations. "$\boldsymbol{h}_r = \boldsymbol{h}_{[\text{CLS}]}$ without entity markers" indicates that this model does not utilize entity markers for the original sequence. "$\boldsymbol{h}_r = \boldsymbol{h}_{[\text{CLS}]}$ with entity markers" indicates that the relation representation is extracted from the hidden state of [CLS] after inserting entity markers. "$\boldsymbol{h}_r = \boldsymbol{h}_{[\text{Ent1}]} \oplus \boldsymbol{h}_{[\text{Ent2}]}$" indicates that the relation representation is extracted from the concatenation between the hidden states of [Ent1] and [Ent2]. "Remove edge-labeling (i.e., node-only)" means that the edge-labeling is removed and only node features are kept. "Remove $\mathcal{G}_D$ (i.e., instance-only)" stands for removing the distribution graph.

Table V shows that the first variation delivers the worst accuracy for the four few-shot configurations (i.e.,-7.56, -9.34, -8.27 and -10.78, respectively). The performance when concatenating $\boldsymbol{h}_{[\text{Ent1}]}$ and $\boldsymbol{h}_{[\text{Ent2}]}$ is much better than when only using $\boldsymbol{h}_{[\text{CLS}]}$. These results clearly indicate that entity markers provide rich and useful information for relation extraction. Therefore, we adopt $\boldsymbol{h}_r = \boldsymbol{h}_{[\text{CLS}]} \oplus \boldsymbol{h}_{[\text{Ent1}]} \oplus \boldsymbol{h}_{[\text{Ent2}]}$ in our model based on the consideration that $\boldsymbol{h}_{[\text{CLS}]}$ provides context representations, while $\boldsymbol{h}_{[\text{Ent1}]}$ and $\boldsymbol{h}_{[\text{Ent2}]}$ provide entity-oriented representations. Removing the edge-labeling strategy results in a significant performance degradation. This is because the edge-labeling strategy in our approach is able to explicitly model both the inter-class dissimilarity and intra-class similarity with edge features. Removing the distribution graphs also results in a significant performance degradation. This indicates that the dual graph interaction in a cyclic flow manner plays an important role in few-shot relation extraction. In summary, this ablation analysis clearly demonstrates the significance of all components in our proposed DUALGRAPH approach.

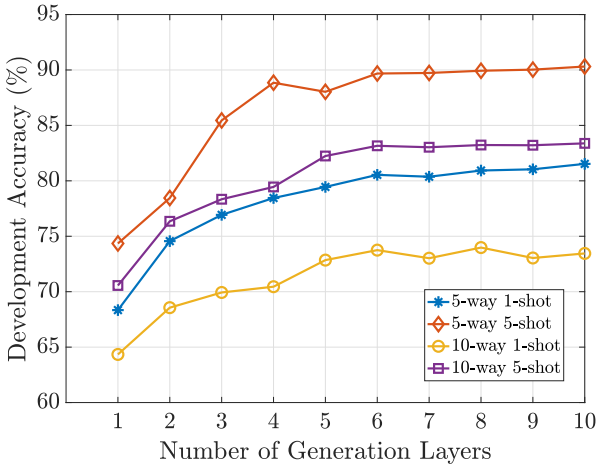
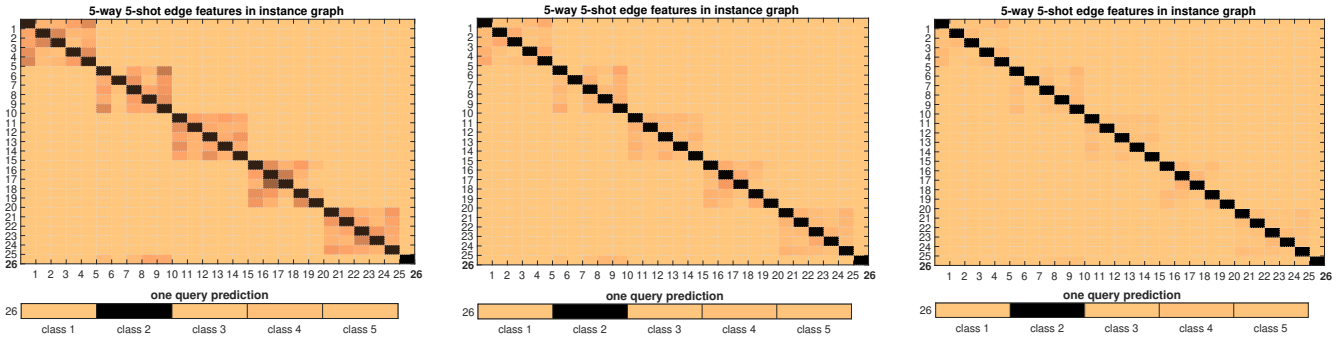


Fig. 4. Impact of different numbers of generation layers.

2) Impact of Number of Generation Layers: Our approach updates the edge and node representations with the dual graph interaction in a cyclic flow manner. As shown in Algorithm 1, we use multiple generation layers to iteratively process the dual graph. The number of generation layers is an important parameter for achieving good relation extraction performance. Thus, we experimentally investigate the impact of using different numbers of generation layers. Figure 4 shows the accuracy scores on the development set of FewRel2.0. We can observe that the performance tends to increase with an increasing number of generation layers, for all four few-shot configurations. In particular, there is a significant improvement from one layer to four layers. However, the performance tends to become stable when the number of steps reaches six. Therefore, our approach adopts six generation layers. This is based on the consideration that more layers require more computations and choosing six layers provides a good balance between computation and accuracy.

Figure 5 shows a visualization of edge feature propagation (upper figures) and query instance prediction (bottom figures) in the testing phase in the 5-way 5-shot configuration. The upper heatmaps show instance edge features, where each value indicates the instance-level similarity between two nodes in the



(a) The 1st generation layer (b) The 3rd generation layer (c) The 6th generation layer

Fig. 5. Visualization of edge feature propagation (upper figures) and query instance prediction (bottom figures). This illustration shows a 5-way 5-shot setting with 26 instances (i.e., 25 instances in the support set and 1 instance in the query set). Note that the first 25 instances are from the support set, and latter 1 (i.e., 26th) is from the query set.

TABLE VI

CASE STUDY UNDER THE 5-WAY 5-SHOT SETTING. OUR DUALGRAPH IS TRAINED ON WIKIPEDIA AND DIRECTLY EVALUATED ON THE BIOMEDICINE DOMAIN. CIRCLED NUMBERS INDICATES RELATION TYPES. RED COLOR (X) INDICATES WRONG PREDICTIONS. "NODE-ONLY" STANDS FOR THE MODEL REMOVING THE EDGE-LABELING STRATEGY. "INSTANCE-ONLY" STANDS FOR THE MODEL REMOVING THE DISTRIBUTION GRAPH.

Support Classes	Query Instance	Prediction
① **inheritance_type_of** ② **ingredient_of** ③ **is_normal_tissue_origin_of_disease** ④ **is_primary_anatomic_site_of_disease** ⑤ **causative_agent_of**	it is clear that the diversity of phenotypes displayed by [breast cancer]$_{e1}$ cells reflects the array of cell types present in the disease-free breast [epithelium]$_{e2}$, including luminal, basal and stem cells. *Ground Truth*: ③	DUALGRAPH: ③ Node-only: ⑤ Instance-only: ④
	[adrenocortical carcinoma]$_{e1}$ (acc) is a rare and aggressive tumor arising from the [adrenal cortex]$_{e2}$ with an incidence of one to two cases per million within the general us population . *Ground Truth*: ③	DUALGRAPH: ③ Node-only: ① Instance-only: ④
	[intravenous leiomyomatosis]$_{e1}$ is considered to be a rare neoplastic disease usually arising from [uterine]$_{e2}$ fibromyomata , but its true incidence may be under-recognized. *Ground Truth*: ④	DUALGRAPH: ④ Node-only: ② Instance-only: ③
	[leukocyte adhesion deficiency type 1]$_{e1}$ (lad-1) is an [autosomal recessive]$_{e2}$ primary immunodeficiency , hallmarked by defective polymorphonuclear transmigration. *Ground Truth*: ①	DUALGRAPH: ① Node-only: ① Instance-only: ②
	spinocerebellar ataxia type 35 ([sca35]$_{e1}$) is an [autosomal dominant]$_{e2}$ neurodegenerative disorder . *Ground Truth*: ①	DUALGRAPH: ① Node-only: ④ Instance-only: ②
	a patient taking [oral risperidone]$_{e1}$ while using cocaine and alcohol presented with priapism shortly after long-acting, injectable [risperidone]$_{e2}$ was prescribed. *Ground Truth*: ②	DUALGRAPH: ② Node-only: ⑤ Instance-only: ①

first, third, and sixth generation layers, respectively. There are a total of 26 instances (i.e., 25 instances from the support set and 1 instance from the query set). Note that the first 25 instances are from the support set, and latter one (i.e., 26th) is from the query set in both the y-axis and x-axis. We can observe that the edge features are gradually refined in the six generation layers. Compared to the first generation layer, the instance-level similarity distances between the same instances are enlarged in the sixth generation layer. In addition, the different instances (from either the same class or different classes) are more discriminative in the sixth generation layer. The bottom heatmaps of Figure 5 show query instance predictions using the equation in Line 8 of Algorithm 1. From this figure, we can observe that the query instances can be correctly predicted in all generation layers (i.e., ground truth class: class 2). In our implementation, we leverage the edge features from the last generation layer to predict query instances. In short, this visualization intuitively shows the effectiveness of using multiple generation layers in the dual graph interaction.

3) Qualitative Analysis: In this subsection, we aim to investigate models' capability of handling the difficult cases

(i.e., specifically those involving closely related concepts). The primary assumption is that the more semantically similar the relations are, the more difficult it becomes to accurately identify them. Consequently, we select qualitative cases based on the extent of word overlap in relation names to assess the models' performance in handling such intricacies.

In the development set of FewRel2.0, the range of overlapping words in relation names varies from 0 to 3. Table VI shows a case study under the 5-way 5-shot setting on FewRel2.0, within the context of maximum number of overlapping words of relation names (i.e., 3). Table VI shows a case study (i.e., the maximum number of overlapping words of relation names is 3) under the 5-way 5-shot setting on FewRel2.0. DUALGRAPH is trained on a Wikipedia corpus and directly adapted to the biomedicine domain. Note that $\mathcal{D}^{qry}_{\mathcal{T}_i}$ and $\mathcal{D}^{qry}_{\mathcal{T}_j}$ both include 5 samples for each of the 5 classes in this experiment. The first column shows the relation classes in the support set, including 5 classes: ① `inheritance_type_of` ② `ingredient_of`, ③ `is_normal_tissue_origin_of_disease`, ④ `is_primary_anatomic_site_of_disease`, and ⑤ `causative_agent_of`. The second column shows some instances in the query set. The third column shows predictions by different models (Node-only and instance-only models are same with the definitions in the Ablation Study Section IV-C1).

We make the following observations: (1) Compared to node-only and instance-only models, DUALGRAPH is more effective in modeling the inter-class dissimilarity. Specifically, the relations of ③ and ④ are very close because both of them are about the description of disease. We observe that DUALGRAPH can correctly distinguish these two relations, while the instance-only model wrongly predicts ③ as ④, and ④ as ③ (see the 1st - 3rd instances). An additional example is the case of relations of ① and ② (see the 4th - 6th instances). The instance-only model wrongly predicts ① as ②, and ② as ①. This qualitative study clearly showcases that the model cannot handle these difficult cases (very similar relations) when removing the distribution graph, demonstrating the importance of edge-labeling distribution graph in modeling inter-class dissimilarity. (2) Compared to node-only and instance-only models, DUALGRAPH is more effective in modeling the intra-class similarity. The 1st and 2nd instances both are belonged to the relation ③. The node-only model wrongly detects them as ⑤ and ①, while the instance-only model wrongly detects them as ④. By contrast, DUALGRAPH can successfully detect these two instances as ③, benefiting from the capability of modeling inter-class dissimilarity. Similar observations hold on the instances of the relation of ① (i.e., the 4th and 5th instances). In short, compared to node-only and instance-only models with a singular fusion mechanism, our distribution graph is constructed from the instance graph by aggregating the information from it, and both of them are the edge-labeling graph. The qualitative study shows the effectiveness of DUALGRAPH in handling difficult cases by benefiting from explicitly modeling inter-class dissimilarity and inter-class dissimilarity.

We further take statistics on the tasks where the number of overlapping words in relation names ranges from 2 to 3. The results show that the accuracy under the 5-way 5-shot setting for these tasks is 85.67%, which is much lower than the overall accuracy of 91.01% (all tasks) in Table III. This significant performance gap clearly reveals that distinguishing two semantically similar relations remains a challenging task for few-shot learning. We anticipate that future research will focus on addressing more challenging cases, such as those involving highly semantically similar relations.

V. CONCLUSION

In this study, we investigated few-shot relation extraction under domain adaptation scenarios. We proposed DUALGRAPH, a novel GNN based approach for few-shot relation extraction. In particular, DUALGRAPH leverages edge-labeling strategies to explicitly model the inter-class dissimilarity and intra-class similarity in each individual graph, and leverages dual graph (i.e., an instance graph and a distribution graph) to explicitly model instance-level and distribution-level relations across graphs. A dual graph interaction mechanism was proposed to adequately fuse the information between two graphs in a cyclic flow manner. We performed extensive experiments on FewRel1.0 and FewRel2.0 benchmarks. The experiment results indicated the effectiveness of DUALGRAPH. We also conducted empirical experiments to further investigate the architectural choices and parameter settings. Finally, we offer a case study for qualitative analysis.

We believe that few-shot relation extraction is a fundamental problem in NLP. Essentially, our few-shot learning approach can be applied in other few-shot tasks and will be beneficial in a wide range of low-resource scenarios. We would want to investigate the applications of our approach in few-shot named entity recognition, entity linking and knowledge graph completion in the future.

REFERENCES

[1] X. Han, H. Zhu, P. Yu, Z. Wang, Y. Yao, Z. Liu, and M. Sun, "Fewrel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation," in *Empirical Methods in Natural Language Processing (EMNLP)*, 2018, pp. 4803–4809.

[2] T. Gao, X. Han, H. Zhu, Z. Liu, P. Li, M. Sun, and J. Zhou, "Fewrel 2.0: Towards more challenging few-shot relation classification," in *Empirical Methods in Natural Language Processing (EMNLP)*, 2019, pp. 6249–6254.

[3] Z. Ye and Z. Ling, "Multi-level matching and aggregation network for few-shot relation classification," in *Association for Computational Linguistics (ACL)*, 2019, pp. 2872–2881.

[4] W. Xu, K. Chen, and T. Zhao, "Discriminative reasoning for document-level relation extraction," in *Findings of Association for Computational Linguistics (ACL)*, 2021, pp. 1653–1663.

[5] S. Zhao, M. Hu, Z. Cai, and F. Liu, "Dynamic modeling cross-modal interactions in two-phase prediction for entity-relation extraction," *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)*, vol. 34, no. 3, pp. 1122–1131, 2023.

[6] X. Li, F. Yin, Z. Sun, X. Li, A. Yuan, D. Chai, M. Zhou, and J. Li, "Entity-relation extraction as multi-turn question answering," in *Association for Computational Linguistics (ACL)*, A. Korhonen, D. R. Traum, and L. Màrquez, Eds., 2019, pp. 1340–1350.

[7] K. Xu, S. Reddy, Y. Feng, S. Huang, and D. Zhao, "Question answering on freebase via relation extraction and textual evidence," in *Association for Computational Linguistics (ACL)*, 2016.

[8] Y. Tang, J. Huang, G. Wang, X. He, and B. Zhou, "Orthogonal relation transforms with graph context modeling for knowledge graph embedding," in *Association for Computational Linguistics (ACL)*, D. Jurafsky, J. Chai, N. Schluter, and J. R. Tetreault, Eds., 2020, pp. 2713–2722.

[9] B. D. Trisedya, G. Weikum, J. Qi, and R. Zhang, "Neural relation extraction for knowledge base enrichment," in *Association for Computational Linguistics (ACL)*, A. Korhonen, D. R. Traum, and L. Màrquez, Eds., 2019, pp. 229–240.

[10] N. Kambhatla, "Combining lexical, syntactic, and semantic features with maximum entropy models for information extraction," in *Association for Computational Linguistics (ACL)*, 2004.

[11] G. Zhou, J. Su, J. Zhang, and M. Zhang, "Exploring various knowledge in relation extraction," in *Association for Computational Linguistics (ACL)*, 2005, pp. 427–434.

[12] R. C. Bunescu and R. J. Mooney, "Subsequence kernels for relation extraction," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2005, pp. 171–178.

[13] S. Zhao and R. Grishman, "Extracting relations with integrated information using kernel methods," in *Association for Computational Linguistics (ACL)*, 2005, pp. 419–426.

[14] C. Liu, W. Sun, W. Chao, and W. Che, "Convolution neural network for relation extraction," in *Advanced Data Mining and Applications (ADMA)*, vol. 8347, 2013, pp. 231–242.

[15] D. Zeng, K. Liu, S. Lai, G. Zhou, and J. Zhao, "Relation classification via convolutional deep neural network," in *International Conference on Computational Linguistics (COLING)*, 2014, pp. 2335–2344.

[16] H. Shahbazi, X. Z. Fern, R. Ghaeini, and P. Tadepalli, "Relation extraction with explanation," in *Association for Computational Linguistics (ACL)*, 2020, pp. 6488–6494.

[17] M. Mintz, S. Bills, R. Snow, and D. Jurafsky, "Distant supervision for relation extraction without labeled data," in *Association for Computational Linguistics (ACL)*, K. Su, J. Su, and J. Wiebe, Eds., 2009, pp. 1003–1011.

[18] K. D. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, "Freebase: a collaboratively created graph database for structuring human knowledge," in *SIGMOD*, J. T. Wang, Ed., 2008, pp. 1247–1250.

[19] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. G. Ives, "Dbpedia: A nucleus for a web of open data," in *ISWC*, vol. 4825, 2007, pp. 722–735.

[20] A. Smirnova and P. Cudré-Mauroux, "Relation extraction using distant supervision: A survey," *ACM Comput. Surv.*, vol. 51, no. 5, pp. 106:1–106:35, 2019.

[21] I. Augenstein, D. Maynard, and F. Ciravegna, "Distantly supervised web relation extraction for knowledge base population," *Semantic Web*, vol. 7, no. 4, pp. 335–349, 2016.

[22] S. Zheng, X. Han, Y. Lin, P. Yu, L. Chen, L. Huang, Z. Liu, and W. Xu, "DIAG-NRE: A neural pattern diagnosis framework for distantly supervised neural relation extraction," in *Association for Computational Linguistics (ACL)*, 2019, pp. 1419–1429.

[23] S. Asmussen, *Applied probability and queues*. Springer Science & Business Media, 2008, vol. 51.

[24] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010.

[25] F. Reiss, S. Raghavan, R. Krishnamurthy, H. Zhu, and S. Vaithyanathan, "An algebraic approach to rule-based information extraction," in *International Conference on Data Engineering (ICDE)*, 2008, pp. 933–942.

[26] D. Bollegala, Y. Matsuo, and M. Ishizuka, "Relational duality: unsupervised extraction of semantic relations between entities on the web," in *The Web Conference (WWW)*, 2010, pp. 151–160.

[27] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a few examples: A survey on few-shot learning," *ACM Comput. Surv.*, vol. 53, no. 3, pp. 63:1–63:34, 2020.

[28] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," in *International Conference on Machine Learning (ICML) deep learning workshop*, vol. 2, 2015.

[29] J. Snell, K. Swersky, and R. S. Zemel, "Prototypical networks for few-shot learning," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 4077–4087.

[30] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *International Conference on Machine Learning (ICML)*, 2017, pp. 1126–1135.

[31] N. Mishra, M. Rohaninejad, X. Chen, and P. Abbeel, "A simple neural attentive meta-learner," in *International Conference on Learning Representations (ICLR)*, 2018.

[32] P. Huang, C. Wang, R. Singh, W. Yih, and X. He, "Natural language to structured query generation via meta-learning," in *North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 2018, pp. 732–738.

[33] J. Li, S. Shang, and L. Shao, "Metaner: Named entity recognition with meta-learning," in *The Web Conference (WWW)*, 2020, pp. 429–440.

[34] J. Kim, T. Kim, S. Kim, and C. D. Yoo, "Edge-labeling graph neural network for few-shot learning," in *Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 11–20.

[35] L. Yang, L. Li, Z. Zhang, X. Zhou, E. Zhou, and Y. Liu, "DPGN: distribution propagation graph network for few-shot learning," in *Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 13 387–13 396.

[36] V. G. Satorras and J. B. Estrach, "Few-shot learning with graph neural networks," in *International Conference on Learning Representations (ICLR)*, 2018.

[37] M. Aydar, O. Bozal, and F. Ozbay, "Neural relation extraction: a survey," *arXiv preprint arXiv:2007.04247*, 2020.

[38] O. Etzioni, M. Banko, S. Soderland, and D. S. Weld, "Open information extraction from the web," *Commun. ACM*, vol. 51, no. 12, pp. 68–74, 2008.

[39] Z. Jiang, P. Yin, and G. Neubig, "Improving open information extraction via iterative rank-aware learning," in *Association for Computational Linguistics (ACL)*, 2019, pp. 5295–5300.

[40] A. Culotta and J. S. Sorensen, "Dependency tree kernels for relation extraction," in *Association for Computational Linguistics (ACL)*, 2004, pp. 423–429.

[41] R. C. Bunescu and R. J. Mooney, "A shortest path dependency kernel for relation extraction," in *Empirical Methods in Natural Language Processing (EMNLP)*, 2005, pp. 724–731.

[42] W. Song, W. Gu, F. Zhu, and S. C. Park, "Interaction-and-response network for distantly supervised relation extraction," *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)*, Early Access, 2023.

[43] W. Jia, D. Dai, X. Xiao, and H. Wu, "ARNOR: attention regularization based noise reduction for distant supervision relation classification," in *Association for Computational Linguistics (ACL)*, 2019, pp. 1399–1408.

[44] B. Roth, T. Barth, M. Wiegand, and D. Klakow, "A survey of noise reduction methods for distant supervision," in *AKBC@CIKM*, 2013, pp. 73–78.

[45] J. Li, P. Han, X. Ren, J. Hu, L. Chen, and S. Shang, "Sequence labeling with meta-learning," *IEEE Transactions on Knowledge and Data Engineering*, 2021.

[46] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching networks for one shot learning," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2016, pp. 3630–3638.

[47] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," in *Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 1199–1208.

[48] J. Schmidhuber, "On learning how to learn learning strategies," 1995.

[49] S. Thrun and L. Y. Pratt, Eds., *Learning to Learn*. Springer, 1998.

[50] Y. Xiao, Y. Jin, and K. Hao, "Adaptive prototypical networks with label words and joint representation learning for few-shot relation classification," *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)*, vol. 34, no. 3, pp. 1406–1417, 2023.

[51] J. Wang, L. Zhang, J. Liu, K. Ma, W. Wu, X. Zhao, Y. Wu, and Y. Huang, "Tgin: Translation-based graph inference network for few-shot relational triplet extraction," *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)*, Early Access, 2022.

[52] M. Qu, T. Gao, L. A. C. Xhonneux, and J. Tang, "Few-shot relation extraction via bayesian meta-learning on relation graphs," in *Proceedings of the 37th International Conference on Machine Learning (ICML)*, vol. 119, 2020, pp. 7867–7876.

[53] J. Han, B. Cheng, and W. Lu, "Exploring task difficulty for few-shot relation extraction," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, M. Moens, X. Huang, L. Specia, and S. W. Yih, Eds., 2021, pp. 2605–2616.

[54] H. Peng, T. Gao, X. Han, Y. Lin, P. Li, Z. Liu, M. Sun, and J. Zhou, "Learning from context or names? an empirical study on neural relation extraction," in *The Conference on Empirical Methods in Natural Language Processing (EMNLP)*, B. Webber, T. Cohn, Y. He, and Y. Liu, Eds., 2020, pp. 3661–3672.

[55] Y. Wang, J. Bao, G. Liu, Y. Wu, X. He, B. Zhou, and T. Zhao, "Learning to decouple relations: Few-shot relation classification with entity-guided attention and confusion-aware training," in *The 28th International Conference on Computational Linguistics (COLING)*, D. Scott, N. Bel, and C. Zong, Eds., 2020, pp. 5799–5809.

[56] M. Dong, C. Pan, and Z. Luo, "Mapre: An effective semantic mapping approach for low-resource relation extraction," in *The 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, M. Moens, X. Huang, L. Specia, and S. W. Yih, Eds., 2021, pp. 2694–2704.

[57] T. Gao, X. Han, Z. Liu, and M. Sun, "Hybrid attention-based prototypical networks for noisy few-shot relation classification," in *The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI)*, 2019, pp. 6407–6414.
[58] L. B. Soares, N. FitzGerald, J. Ling, and T. Kwiatkowski, "Matching the blanks: Distributional similarity for relation learning," in *Association for Computational Linguistics (ACL)*, 2019, pp. 2895–2905.
[59] S. Yang, Y. Zhang, G. Niu, Q. Zhao, and S. Pu, "Entity concept-enhanced few-shot relation extraction," in *The 59th Annual Meeting of the Association for Computational Linguistics(ACL)*, C. Zong, F. Xia, W. Li, and R. Navigli, Eds., 2021, pp. 987–991.
[60] J. Zhang, J. Zhu, Y. Yang, W. Shi, C. Zhang, and H. Wang, "Knowledge-enhanced domain adaptation in few-shot relation classification," in *The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, F. Zhu, B. C. Ooi, and C. Miao, Eds., 2021, pp. 2183–2191.
[61] J. Zhou, G. Cui, Z. Zhang, C. Yang, Z. Liu, and M. Sun, "Graph neural networks: A review of methods and applications," *CoRR*, vol. abs/1812.08434, 2018.
[62] L. Song, Z. Wang, M. Yu, Y. Zhang, R. Florian, and D. Gildea, "Evidence integration for multi-hop reading comprehension with graph neural networks," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 2, pp. 631–639, 2022.
[63] Y. Xiong, Y. Zhang, X. Kong, H. Chen, and Y. Zhu, "Graphinception: Convolutional neural networks for collective classification in heterogeneous information networks," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 5, pp. 1960–1972, 2021.
[64] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Trans. Neural Networks*, vol. 20, no. 1, pp. 61–80, 2009.
[65] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *arXiv preprint arXiv:1901.00596*, 2019.
[66] T. N. Kipf and M. Welling, "Variational graph auto-encoders," *arXiv preprint arXiv:1611.07308*, 2016.
[67] Y. Li, O. Vinyals, C. Dyer, R. Pascanu, and P. W. Battaglia, "Learning deep generative models of graphs," *arXiv preprint arXiv:1803.03324*, 2018.
[68] Y. Li, D. Tarlow, M. Brockschmidt, and R. S. Zemel, "Gated graph sequence neural networks," in *International Conference on Learning Representations (ICLR)*, 2016.
[69] H. Dai, Z. Kozareva, B. Dai, A. J. Smola, and L. Song, "Learning steady-states of iterative algorithms over graphs," in *International Conference on Machine Learning (ICML)*, vol. 80, 2018, pp. 1114–1122.
[70] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," in *International Conference on Learning Representations (ICLR)*, 2014.
[71] J. Atwood and D. Towsley, "Diffusion-convolutional neural networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2016, pp. 1993–2001.
[72] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," in *International Joint Conference on Artificial Intelligence (IJCAI)*, 2018, pp. 3634–3640.
[73] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," in *International Conference on Learning Representations (ICLR)*, 2018.
[74] Y. Liu, J. Lee, M. Park, S. Kim, E. Yang, S. J. Hwang, and Y. Yang, "Learning to propagate labels: Transductive propagation network for few-shot learning," in *International Conference on Learning Representations (ICLR)*, 2019.
[75] J. Chauhan, D. Nathani, and M. Kaul, "Few-shot learning on graphs via super-classes based on graph spectral measures," in *8th International Conference on Learning Representations (ICLR)*, 2020.
[76] Z. Zhang, X. Han, Z. Liu, X. Jiang, M. Sun, and Q. Liu, "ERNIE: enhanced language representation with informative entities," in *Association for Computational Linguistics (ACL)*, A. Korhonen, D. R. Traum, and L. Màrquez, Eds., 2019, pp. 1441–1451.
[77] I. Hendrickx, S. N. Kim, Z. Kozareva, P. Nakov, D. Ó. Séaghdha, S. Padó, M. Pennacchiotti, L. Romano, and S. Szpakowicz, "Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals," in *Workshop on Semantic Evaluations: Recent Achievements and Future Directions (SEW@NAACL-HLT)*, 2009, pp. 94–99.
[78] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 2019, pp. 4171–4186.
[79] C. Dou, S. Wu, X. Zhang, Z. Feng, and K. Wang, "Function-words adaptively enhanced attention networks for few-shot inverse relation classification," in *The Thirty-First International Joint Conference on Artificial Intelligence (IJCAI)*, L. D. Raedt, Ed., 2022, pp. 2937–2943.
[80] J. Yoon, T. Kim, O. Dia, S. Kim, Y. Bengio, and S. Ahn, "Bayesian model-agnostic meta-learning," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2018, pp. 7343–7353.

Jing Li is currently a Full Professor with Harbin Institute of Technology (HIT), Shenzhen, China. Prior to that, he was a Research Scientist at the Inception Institute of Artificial Intelligence (IIAI), Abu Dhabi, United Arab Emirates. He received his PhD degree from Nanyang Technological University (NTU), Singapore, in 2018. His research aims to build up semantic Web systems to support information needs of users via deep text understanding, web information extraction, machine intelligent question answering, knowledge representation, as well as social media analysis.

Shanshan Feng is an Associate Professor with Harbin Institute of Technology, Shenzhen, China. He received the Ph.D. degree in Computer Science from Nanyang Technological University, Singapre in 2017. He was a research scientist at the Inception Institute of Artificial Intelligence, Abu Dhabi, UAE. His research interests include graph learning, recommender systems, spatial data mining.

Billy Chiu is an Assistant Professor with Department of Computing and Decision Sciences, Lingnan University, Hong Kong. He received his PhD in the Language Technology Laboratory, University of Cambridge, United Kingdom, in 2019. Billy specializes in representation learning and lexical semantic in biomedical data. His current research investigates the applications of representation learning models for downstream Natural Language Processing (NLP) tasks. His work has been published in a Journal of BMC bioinformatics, Journal of biomedical semantics, as well as conferences including the International Conference on Language Resources and Evaluation and the Association for Computational Linguistics.

@article{jing23dualgraph,
author = {Jing Li and Shanshan Feng and Billy Chiu},
title = {Few-Shot Relation Extraction With Dual Graph Neural Network Interaction},
journal = {IEEE Transactions on Neural Networks and Learning Systems (TNNLS)},
pages = {Early Access },
year = {2023},
url = {https://doi.org/10.1109/TNNLS.2023.3278938},
doi = {10.1109/TNNLS.2023.3278938},
}