

# KGNN: Distributed Framework for Graph Neural Knowledge Representation

Binbin Hu<sup>1</sup> Zhiyang Hu<sup>1</sup> Zhiqiang Zhang<sup>1</sup> Jun Zhou<sup>1</sup> Chuan Shi<sup>2</sup>

## Abstract

Knowledge representation learning has been commonly adopted to incorporate knowledge graph (KG) into various online services. Although existing knowledge representation learning methods have achieved considerable performance improvement, they ignore high-order structure and abundant attribute information, resulting in unsatisfactory performance on semantics-rich KGs. Moreover, they fail to make prediction in an inductive manner and cannot scale to large industrial graphs. To address these issues, we develop a novel framework called KGNN to take full advantage of knowledge data for representation learning in the distributed learning system. KGNN is equipped with GNN based encoder and knowledge aware decoder, which aim to jointly explore high-order structure and attribute information together in a fine-grained fashion and preserve the relation patterns in KGs, respectively. Extensive experiments on three datasets for link prediction and triplet classification task demonstrate the effectiveness and scalability of KGNN framework.

## 1. Introduction

Knowledge graph (KG) represents the heterogeneous structure of entities and their rich relations in triplets of the form  $\langle head\ entity, relation, tail\ entity \rangle$ . For example in Fig. 1, a triplet  $\langle Bob, work\_in, Apple \rangle$  is denoted as a relation *work\_in* connecting two entities: *Bob* and *Apple*. Due to abundant structured information, KG has attracted much attention in many research areas, ranging from information retrieval (Dietz et al., 2018), question answering (Huang et al., 2019) to recommender system (Cao et al., 2019).

To flexibly incorporate such knowledge, knowledge representation learning (Wang et al., 2017) has emerged as

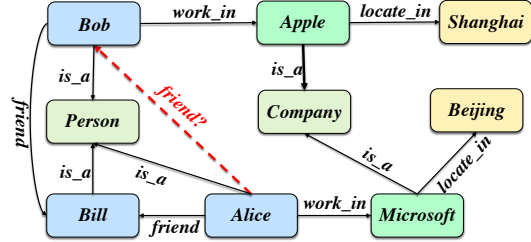


Figure 1. The example of knowledge graph.

a promising direction for knowledge completion (Lacroix et al., 2018), alignment (Wang et al., 2018) and reasoning (Trivedi et al., 2017), which aims to project both entities and relations into a low-dimensional space whilst preserving certain information of the original graph. These methods can be broadly classified as *translational distance models* (Bordes et al., 2013; Wang et al., 2014; Lin et al., 2015; Sun et al., 2019) and *semantic matching models* (Nickel et al., 2011; Jenatton et al., 2012; Yang et al., 2015; Trouillon et al., 2016; Dettmers et al., 2018), which exploit distance-based and similarity-based scoring function for knowledge representation learning, respectively.

Although these methods have yield considerable performance improvements to some extent, they still suffer from several limitations. First, they process each triple independently and abundant attributes in nodes and edges are commonly ignored, resulting in unsatisfactory performance on semantics-rich KGs. Second, they are inherently transductive models, which cannot make prediction for entities unseen in the training set. Third, these methods cannot scale to industrial-scale graphs with hundreds of millions of entities and relations.

To address these issues, in this paper, we aim to build a scalable and distributed knowledge graph representation framework to flexibly distill rich knowledge for downstream applications. Intuitively, the framework is expected to satisfy the following three key properties: (1) **Semantics-rich**: High order structure and attribute information have been already proved effective for preserving properties of original graphs in previous works (Xu et al., 2018; Abu-El-Haija et al., 2019; Luan et al., 2019). Hence, we aim to incorporate such information into knowledge graph representation

<sup>1</sup>Ant Financial Services Group, Hangzhou, China <sup>2</sup>Beijing University of Posts and Telecommunications, Beijing, China. Correspondence to: Binbin Hu <bin.hbb@antfin.com>.

to comprehensively capture rich semantics in KGs. (2) **Inductive**: Current KGs are usually far from complete and thus the new entities will appear everyday in the real-world setting, which prompts the proposal to make prediction for entities unseen in the training set dynamically. (3) **Scalable**: Since KGs in the real-world industrial scenarios are extremely large-scale, a scalable knowledge graph representation framework implemented on distributed learning system is in urgent demand.

To integrate above main idea together, we propose **KGNN**, a distributed framework for graph neural knowledge representation with graph neural network (GNN) based encoder and knowledge aware decoder. With the help recently emerging GNN, KGNN is potential to jointly capture attribute information and high order structure in an inductive, end-to-end framework. Obviously, it is a flexible framework to equip arbitrary GNN based encoder, and in this paper, an attention based GNN is introduced to locate the important and relevant relations or structures for fine-grained semantics. In order to perform model training and inference effectively for real-world KGs, KGNN is implemented on the distributed learning system and the implementation details are uncovered. We make extensive experiments on three real-world datasets on link prediction and triplet classification task, which demonstrates the effectiveness and scalability of the proposed KGNN framework.

## 2. Background

In this section, we give a brief overview of knowledge representation learning and graph neural networks.

**Knowledge representation learning.** A knowledge graph is denoted by  $\mathcal{G} = \{\mathcal{E}, \mathcal{R}\}$ , consisting of the entity set  $\mathcal{E}$  and the relation set  $\mathcal{R}$ . A triplet  $\langle h, r, t \rangle$  is defined as an relation  $r$  between entities  $h$  and  $t$  on  $\mathcal{G}$ , where  $h, r \in \mathcal{G}$ . Learning distributional representations of KGs provides an effective and efficient way for applying structural knowledge in various applications. Hence, a scoring function  $s(e_h, e_r, e_t)$  is defined as the likelihood of triple  $\langle h, r, t \rangle$  being a valid triple, where  $e_h, e_r, e_t$  represent the embeddings of  $h, r, t$ , respectively. A series of scoring functions (Wang et al., 2017) are proposed to preserve different relation patterns of KGs, and here, we introduce the *TransH* based scoring function (Wang et al., 2014), which learns different representations for an entity conditioned on different relations.

$$s(e_h, e_r, e_t) = \|e_h^\perp + e_r - e_t^\perp\|. \quad (1)$$

Here, we have  $e_h^\perp = e_h - w_r^T e_h w_r$  and  $e_t^\perp = e_t - w_r^T e_t w_r$ , in order to project entity embeddings into relation hyperplanes, which allows entities playing different roles under different relations.

**Graph neural network.** Graph neural network (GNN) makes use of the structure of the graph and attributes on

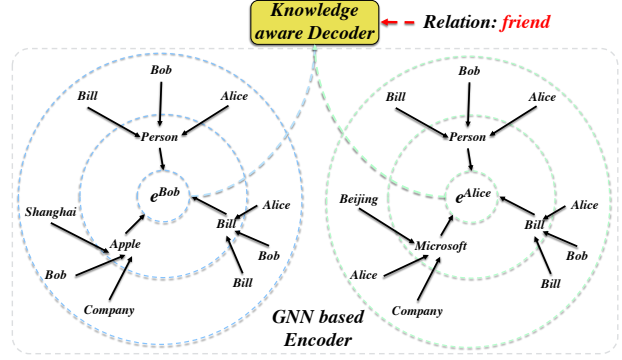


Figure 2. Overview of KGNN model.

nodes for representation learning (Hamilton et al., 2017; Veličković et al., 2018; Kipf & Welling, 2017). Specifically, GNN recursively update a node’s representation by aggregating information from its neighbors. Subsequently, the final representations of the nodes after  $k$  updating capture the structural information as well as the node attributes within  $k$ -hop neighbors. Formally, we can calculate the  $k + 1$ -th representation for node  $v$  with aggregation and updating function as follows,

$$e_v^{k+1} = f^{(U)}(e_v^k, f^{(A)}(\{e_{v'}, v' \in \mathcal{N}_v\}; \Theta^{(A)}); \Theta^{(U)}), \quad (2)$$

where  $f^{(A)}$  and  $f^{(U)}$  denotes the aggregation and updating function parameterized by  $\Theta^{(A)}$  and  $\Theta^{(U)}$ , respectively, and  $\mathcal{N}_v$  is the neighbor set of node  $v$ .

## 3. Methodology

In this section, we present the distributed framework for graph neural knowledge representation, called **KGNN**.

### 3.1. KGNN Model

In this section, we introduce the model part of KGNN to comprehensively distill knowledge graph for representation learning in an inductive manner. We present the architecture of our proposed KGNN in Fig. 2, which intuitively consists of two modules: (1) GNN based encoder and knowledge aware decoder, which flexibly utilizes the local structure information and recursively propagates the embeddings over KGs for expressive representations and (2) knowledge aware decoder, which aims to preserve the relation patterns in KGs through various types of score functions.

**GNN based Encoder.** Different from one-hot representation in previous works, we propose to adopt graph neural network to encode structural knowledge and attributes into entities’ representations. For fine-grained modeling, we introduce an attention based GNN to weighs various underlying preference for each relation. Following the above up-

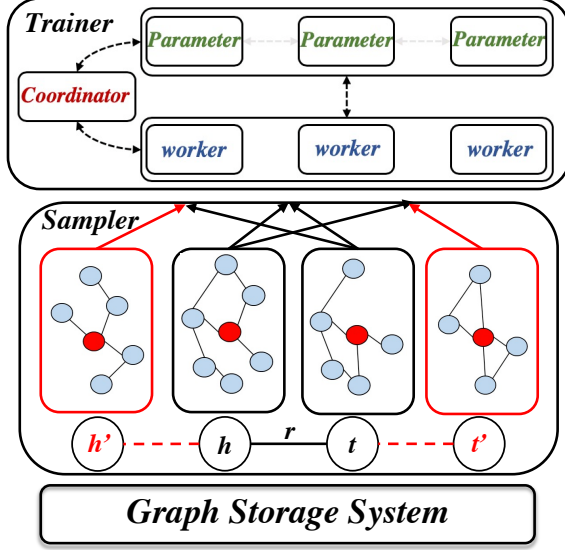


Figure 3. Overview of distributed KGNN framework.

dating principle of entity representations in Eq. 2, we firstly formulate the aggregation function  $f^{(A)}(\cdot)$  as follows:

$$f^{(A)}(\{e_t^k, t \in \mathcal{N}_h^k\}) = \sum_{r, t \in \mathcal{N}_h} \alpha(h, r, t) e_t^k. \quad (3)$$

Here,  $\alpha(h, r, t)$  is the attention value for the triple  $\langle h, r, t \rangle$ , which is implemented as a neural network. And  $\mathcal{N}_h^k = \{(r, t) | (h, r, t) \in \mathcal{G}\}$  is the  $k$ -hop neighbor set for entity  $h$ .

Inspired by the idea jumping knowledge network (Xu et al., 2018), we adopt an adaptive depth function to flexibly multiple hops of neighbors for better structure-aware representation. Here, an LSTM is applied to implement  $f^{(U)}(\cdot)$  for representation updating. Therefore, we can obtain the  $k+1$ -th representation for entity  $h$  as follows:

$$e_h^{k+1} = LSTM(e_h^k, a_h^k), \quad (4)$$

where  $a_h^k$  denotes the aggregated information for entity  $h$ , calculated by Eq. 3.

**Knowledge aware decoder.** The key of link prediction in KGs is to infer the relation patterns *e.g.*, symmetry, inversion and composition with observed triplets (Sun et al., 2019). In order to adaptively preserve different relation patterns on various KGs, KGNN adopts knowledge aware score function as the decoder. Take the *TransH* as an example, we represent the score function for a triple  $\langle h, r, t \rangle$  after  $K$ -hop updating as  $s(e_h^K, e_r, e_t^K)$ . Then, we train KGNN in an end-to-end fashion via the margin based objective with negative sampling:

$$\mathcal{L} = \sum_{\langle h, r, t \rangle \in \mathcal{G}, \langle h', r, t' \rangle \in \mathcal{G}'} [s(e_h^K, e_r, e_t^K) + \lambda - s(e_{h'}^K, e_r, e_{t'}^K)]_+, \quad (5)$$

Table 1. Statistics of data sets.

Dataset	# Ent.	# Rel.	# Trip.	# Attr.
WN18	40, 943	18	151, 442	N.A.
FB15K	14, 951	1, 345	592, 213	N.A.
Alipay	$2.6 \times 10^5$	6	$1.28 \times 10^6$	504

where  $[\cdot] = \max(0, \cdot)$ , and  $\mathcal{G}'$  is the set of incorrect triplets constructed by randomly replacing head entity or tail entity in a valid triplet.

### 3.2. Distributed Implementation

We now zoom into the distribution implementation of KGNN, which provide a complete solution for large-scale knowledge graph representation. As shown in Fig. 3, the distributed KGNN is comprised of three parts:

- **Graph storage system.** It stores the whole knowledge graph as well as corresponding attributes information on nodes under the distributed architecture. With the help of the effective data compression technology, it is capable of serving large-scale industrial graphs.
- **Sampler.** It mainly provides negative sampler and sub-graph sampler for knowledge representation. In particular, the negative sampler randomly replaces head entity or tail entity in a batch of valid triplets for corresponding corrupted triplets. And then, sub-graph sampler will randomly collect  $k$ -hop neighbors set for each entity in batch. It is worth noting that we feed the sub-graph into KGNN instead of the full graph, which helps reduce the time and memory cost.
- **Trainer.** It consisting of several workers and parameter servers, controlled by the coordinator. For effective parameter updating, each work pulls parameters from a parameter server and update them independently during training. In a specific worker, KGNN naturally follows such a work flow: (1) Pre-process the sub-graph and parse the model config. (2) Produce embeddings for entities and relations based on sub-graph with our encoder and decoder introduced in Sec. 3.1. (3) Optimize a certain loss to guide the learning process.

## 4. Experiments

In this section, we evaluate the effectiveness of KGNN for link prediction and triplet classification task.

**Datasets and evaluation metrics.** We evaluate our proposed framework on three datasets (Lin et al., 2015), namely WN18, FB15K and industrial AliPay dataset. The detailed descriptions of the three datasets are summarized in Tab. 1. We perform link prediction on WN18 and FB15K, while

Table 2. Evaluation results on link prediction. (%)

Methods	WN18 (HR@k)			FB15K (HR@k)		
	1	3	10	1	3	10
TransE	70.8	89.6	94.7	64.4	84.0	95.9
TransR	65.5	83.7	92.7	63.6	82.3	95.2
TransH	72.3	90.6	94.9	64.5	84.1	95.9
DistMult	69.3	89.9	94.6	65.2	84.6	96.5
KGNN	<b>78.9</b>	<b>96.9</b>	<b>98.8</b>	<b>67.4</b>	<b>86.4</b>	<b>96.8</b>

apply triple classification on WN18, FB15K and AliPay dataset. Following the same setting in (Bordes et al., 2011; 2013), we adopt hit ratio at rank  $k$  (HR@ $k$ ) and area under ROC curve (AUC) to evaluate the model performance of link prediction and triplet classification, respectively.

Table 3. AUC comparison results on triplet classification. (%)

Methods	WN18	FB15K	Alipay
TransE	91.7	97.5	61.0
TransR	78.6	95.8	74.9
TransH	91.7	97.4	72.6
DistMult	93.5	97.9	65.3
KGNN	<b>94.1</b>	<b>99.0</b>	<b>84.9</b>

**Compared methods and parameter settings.** We consider for representative knowledge representation learning methods for performance comparison, namely TransE (Bordes et al., 2013), TransR (Lin et al., 2015), TransH (Wang et al., 2014) and DistMult (Yang et al., 2015). For fair comparison, we also select one of them as the decoder of KGNN framework. We adopt Adam with learning rate = 0.001 to optimize all models and set the batch size = 256. Moreover, the margin is selected among {1, 2, 5} and the embedding size is searched among {64, 128, 256}.

**Performance Comparison.** We report the comparison results of the proposed KGNN and baselines on link prediction and triplet classification in Tab. 2 and Tab. 3, respectively. We observe that KGNN consistently outperform on three datasets for both tasks, indicating that KGNN is potential to capture high-order structural information for more expressive knowledge representations. It is worthwhile to note that KGNN achieve significant performance improvement over baselines on Alipay datasets. The results may correlated with the characteristics of this dataset: (1) There are 504 attributes on entities, which are ignored by these baselines. (2) The test set contains a part of unseen entities, while these baselines fail to produce proper representations for them. As a comparison, the performance of KGNN demonstrates that KGNN is capable of jointly exploring structure and attribute information together over KGs in an inductive manner.

**Effect of the number of hops.** We analyze the effect of the number of hops on the link prediction task through varying

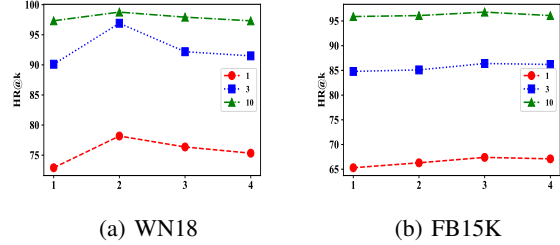


Figure 4. Performance study w.r.t. the number of hops

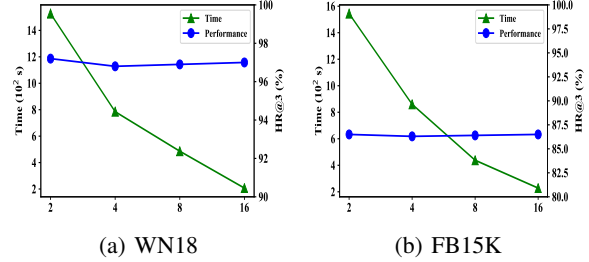


Figure 5. Scalability study w.r.t. the number of workers

it among {1, 2, 3, 4}. As shown in Fig. 4, the proposed KGNN achieve the optimal performance when # hop = 2 on WN18 and # hop = 3 on FB15K. The results indicates high-order structure information exactly help our model learn more powerful representations, while excessive hops of neighbors would harm the performance due to the over-smoothing problem (Chen et al., 2019).

**Scalability study.** To verify the scalability of our proposed distributed KGNN framework, we report the updating time per training epoch w.r.t. the number of workers in Fig. 5. As shown, the speed up in training KGNN on WN18 and FB15K is consistent as we increase the number of workers from 2 to 16. Meanwhile, it also shows that there is almost no loss of predictive performance as the number of workers increases.

## 5. Conclusion

In this paper, we proposed a novel distributed framework called KGNN for graph neural knowledge representation with GNN based encoder and knowledge aware decoder, which jointly exploit high-order structure and attribute information together for powerful knowledge representation as well as preserve relation patterns in KGs. Furthermore, an attention mechanism is introduced to emphasize important information for fine-grained modeling. We implement the proposed KGNN on the distributed learning system and extensive experiments demonstrates its effectiveness and scalability.

## References

- Abu-El-Haija, S., Perozzi, B., Kapoor, A., Alipourfard, N., Lerman, K., Harutyunyan, H., Steeg, G. V., and Galstyan, A. Mixhop: Higher-order graph convolutional architectures via sparsified neighborhood mixing. In *ICML*, pp. 21–29, 2019.
- Bordes, A., Weston, J., Collobert, R., and Bengio, Y. Learning structured embeddings of knowledge bases. In *AAAI*, pp. 301–306, 2011.
- Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., and Yakhnenko, O. Translating embeddings for modeling multi-relational data. In *NIPS*, pp. 2787–2795, 2013.
- Cao, Y., Wang, X., He, X., Hu, Z., and Chua, T.-S. Unifying knowledge graph learning and recommendation: Towards a better understanding of user preferences. In *WWW*, pp. 151–161, 2019.
- Chen, D., Lin, Y., Li, W., Li, P., Zhou, J., and Sun, X. Measuring and relieving the over-smoothing problem for graph neural networks from the topological view. *arXiv preprint arXiv:1909.03211*, 2019.
- Dettmers, T., Minervini, P., Stenetorp, P., and Riedel, S. Convolutional 2d knowledge graph embeddings. In *AAAI*, pp. 1811–1818, 2018.
- Dietz, L., Kotov, A., and Meij, E. Utilizing knowledge graphs for text-centric information retrieval. In *SIGIR*, pp. 1387–1390, 2018.
- Hamilton, W., Ying, Z., and Leskovec, J. Inductive representation learning on large graphs. In *NIPS*, pp. 1024–1034, 2017.
- Huang, X., Zhang, J., Li, D., and Li, P. Knowledge graph embedding based question answering. In *WSDM*, pp. 105–113, 2019.
- Jenatton, R., Roux, N. L., Bordes, A., and Obozinski, G. R. A latent factor model for highly multi-relational data. In *NIPS*, pp. 3167–3175, 2012.
- Kipf, T. N. and Welling, M. Semi-supervised classification with graph convolutional networks. *ICLR*, 2017.
- Lacroix, T., Usunier, N., and Obozinski, G. Canonical tensor decomposition for knowledge base completion. In *ICML*, pp. 2869–2878, 2018.
- Lin, Y., Liu, Z., Sun, M., Liu, Y., and Zhu, X. Learning entity and relation embeddings for knowledge graph completion. In *AAAI*, pp. 2181–2187, 2015.
- Luan, S., Zhao, M., Chang, X.-W., and Precup, D. Break the ceiling: Stronger multi-scale deep graph convolutional networks. In *NIPS*, pp. 10945–10955, 2019.
- Nickel, M., Tresp, V., and Kriegel, H.-P. A three-way model for collective learning on multi-relational data. In *ICML*, pp. 809–816, 2011.
- Sun, Z., Deng, Z.-H., Nie, J.-Y., and Tang, J. Rotate: Knowledge graph embedding by relational rotation in complex space. In *ICLR*, 2019.
- Trivedi, R., Dai, H., Wang, Y., and Song, L. Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. In *ICML*, pp. 3462–3471, 2017.
- Trouillon, T., Welbl, J., Riedel, S., Gaussier, É., and Bouchard, G. Complex embeddings for simple link prediction. In *ICML*, pp. 2071–2080, 2016.
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. Graph attention networks. In *ICLR*, 2018.
- Wang, Q., Mao, Z., Wang, B., and Guo, L. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12):2724–2743, 2017.
- Wang, Z., Zhang, J., Feng, J., and Chen, Z. Knowledge graph embedding by translating on hyperplanes. In *AAAI*, pp. 1112–1119, 2014.
- Wang, Z., Lv, Q., Lan, X., and Zhang, Y. Cross-lingual knowledge graph alignment via graph convolutional networks. In *EMNLP*, pp. 349–357, 2018.
- Xu, K., Li, C., Tian, Y., Sonobe, T., Kawarabayashi, K.-i., and Jegelka, S. Representation learning on graphs with jumping knowledge networks. In *ICML*, pp. 5449–5458, 2018.
- Yang, B., Yih, W.-t., He, X., Gao, J., and Deng, L. Embedding entities and relations for learning and inference in knowledge bases. In *ICLR*, 2015.