# Learning Joint Relational Co-evolution in Spatial-Temporal Knowledge Graph for SMEs Supply Chain Prediction


Youru Li
Institute of Information Science,
Beijing Jiaotong University
Beijing, China
liyouru@bjtu.edu.cn

Zhenfeng Zhu*
Institute of Information Science,
Beijing Jiaotong University
Beijing, China
zhfzhu@bjtu.edu.cn

Xiaobo Guo
Beijing Jiaotong University
& MYBank, Ant Group
Beijing, China
xb_guo@bjtu.edu.cn

Linxun Chen
MYBank, Ant Group
Beijing, China
linxun.clx@antgroup.com

Zhouyin Wang
MYBank, Ant Group
Hangzhou, China
wangzhouyin.zy@antgroup.com

Yinmeng Wang
MYBank, Ant Group
Shanghai, China
junheng.wym@antgroup.com

Bing Han
MYBank, Ant Group
Beijing, China
hanbing.hanbing@antgroup.com

Yao Zhao
Institute of Information Science,
Beijing Jiaotong University
Beijing, China
yzhao@bjtu.edu.cn



## ABSTRACT

To effectively explore the supply chain relationships among Small and Medium-sized Enterprises (SMEs), some remarkable progress in such a relation modeling problem, especially knowledge graph-based methods have been witnessed during these years. As a typical link prediction task, supply chain prediction can usually predict the unknown future relationship facts between SMEs by utilizing the historical semantic connections between entities in knowledge graphs (KGs). However, it is still a great challenge for existing models as seldom of them can consider both temporal dependency and cooperative correlation of the connectivity pattern along the timeline synergistically. Accordingly, we propose a novel framework to learn joint relational co-evolution in Spatial-Temporal Knowledge Graphs (STKG). Specifically, on the base of the constructed large-scale financial STKG, a multi-view relational sequences mining method is proposed to reveal the semantic information from ontological concepts. Furthermore, a relational co-evolution learning module is also developed to capture the regularity of evolving connectivity patterns from the spatial-temporal view. Meanwhile, a multiple random subspace representation learning layer is also designed to improve both compatibility and complementarity during knowledge aggregation. Experimental results on large-scale SMEs supply chain prediction tasks from four real-world industries in China have illustrated the effectiveness of the proposed model.



*Corresponding author.





*KDD '23, August 6-10, 2023, Long Beach , CA , USA*





ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn



## CCS CONCEPTS

• **Applied computing → Economics**; • **Computing methodologies** → *Knowledge representation and reasoning*; • **Information systems** → Spatial-temporal systems.

## KEYWORDS

Knowledge graph, Deep learning, Supply chain prediction

**ACM Reference Format:**
Youru Li, Zhenfeng Zhu, Xiaobo Guo, Linxun Chen, Zhouyin Wang, Yinmeng Wang, Bing Han, and Yao Zhao. 2023. Learning Joint Relational Co-evolution in Spatial-Temporal Knowledge Graph for SMEs Supply Chain Prediction. In *KDD '23, August 6-10, 2023, Long Beach , CA , USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn


## 1 INTRODUCTION

During these years, supply chain systems have played an impressive part in the digital economy [45], leading to a significant contribution in multiple areas [4, 25, 26]. For instance, with the help of supply chain networks, enterprises can improve their production and operation efficiency [22], and banks can clarify the operational relationship between enterprises, so as to grant credits more accurately and provide financing assistance in a more targeted manner [11]. However, these promising practices are mostly experienced by large enterprises, but rarely reach Small and Medium Enterprises (SMEs). In fact, it is rather challenging to accurately predict the supply chain relationships between SMEs due to their large base, unstable business status, less available information, and many other objective reasons [1].

Fortunately, Graph Neural Networks (GNNs) [13, 16, 31, 42] techniques are constantly improving the performance in modeling relational data, and especially knowledge graph-based methods are delivering excellent results in many real-world scenarios [7, 10, 19, 37], which all provide us with the experience to explore the supply chain relationships among SMEs. However, it is still a

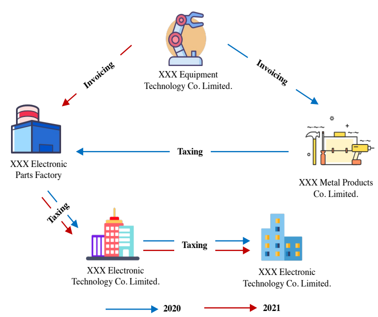


**Figure 1: An example of the supply chain relationships among SMEs in the manufacturing industry.**

great challenge for these existing models as they hardly consider the dynamics of SMEs relationships along the timeline [17, 20]. For example, Fig. 1 highlights the relationship paths for small- and medium-sized manufacturing enterprises in different years with different colors. We can see clearly that as the 'XXX' Metal Products Co. Limited was removed from the original path (in blue), a new supply chain path took shape over time (in red). As a matter of fact, the real supply chain relationships among SMEs can change more profoundly than that, which reflects the necessity of modeling temporal dependency in supply chain relationships.

The cooperative correlation among the supply chain relationships should also be noted. Generally, plural connection paths exist between two given entities in a large-scale interconnected knowledge graph, through which multiple semantic information is conveyed [27]. In practice, these path connectivity patterns can be concluded to make supply chain prediction more accurate [6, 24]. Besides, the plural geographical distribution of supply chains resulting from the uneven regional economic development and diverse industrial structures means that geographical information of SMEs should also be considered in supply chain prediction [29]. For these, a common characteristic is that different factors are often influenced by or correlated with others [39]. Thus, both the temporal dependency along the timeline and the cooperative correlation between the plural connection paths should be considered synergistically. Given these concerns, the following challenges remain to be addressed.

- ***CHALLENGE 1:*** How to explicitly describe the plural relationships in the supply chain so as to capture their inherent connectivity pattern from the spatial-temporal view?
- ***CHALLENGE 2:*** How to jointly model both temporal dependency and cooperative correlation in the supply chain so as to learn the regularity of evolution for such spatial-temporal connectivity patterns?
- ***CHALLENGE 3:*** How to adequately coordinate the coupling between temporal dependency and cooperative correlation so as to aggregate multi-view semantic information?

To this end, we provide a new perspective for SMEs supply chain prediction by learning joint relational co-evolution in the spatial-temporal knowledge graph. In summary, our contributions are highlighted as follows:

- To effectively organize the multiple financial relationships among millions of SMEs, we constructed a large-scale financial Spatial-Temporal Knowledge Graph (STKG).
- To reveal the semantic information embedded in the constructed large-scale financial STKG, we proposed a multi-view relational sequence encoding method for more extensive relation exploring.
- To consistently capture the evolution regularity of supply chain connectivity patterns, we developed a relational co-evolution learning module for jointly modeling the multi-view semantic information from the spatial-temporal view.
- To better disentangle the evolution of connectivity patterns from different views, we designed a multiple random subspaces layer for knowledge aggregation.

Moreover, to evaluate the performance of models, we build four large-scale SMEs supply chain prediction datasets from real-world industries in China and the experimental results demonstrate the effectiveness of the proposed JRCL.

## 2 RELATED WORK

In this section, we will review and discuss the existing work on the general practices of KG-based recommendation systems and temporal knowledge graphs.

### 2.1 KG-based Recommendation Systems.

With knowledge graphs being treated as side information, more and more KG-based recommendation systems have been developed to better address the challenges of data sparsity and cold start in real-world applications. Generally, they are mainly categorized into embedding-based methods [43], propagation-based methods [32], GNNs-based methods [33–35] and connection-based methods [28, 36]. Compared to the existing deep learning-based methods, they can utilize the knowledge graph for not only obtaining accurate recommendations but also providing explanations [7]. As a similar link prediction task, they are also considered a feasible manner to predict supply chain relationships between SMEs. However, unlike the factual relationships between items in recommendation systems, the financial events from SMEs are valid only in certain moments or within a range of time, which brings more challenges.

### 2.2 Temporal Knowledge Graphs.

To effectively capture the dynamics of facts along the timeline in KGs, some temporal knowledge graphs-based modeling methods are raised during the past several years [2]. In general, to better predict the past missing facts or unknown future facts, most prior efforts improved existing knowledge graph embedding technologies designed for static knowledge graphs by embedding the temporal information in various manners [12, 18, 23, 44]. Although they can learn the evolutional representations of entities and relations at each timestamp through the knowledge graph structures with historical dependencies [15, 17], such performance is not sufficient as they ignore learning the co-evolution of connectivity pattern and temporal semantic information. Moreover, it is worth noting that supply chain prediction is a binary relation classification problem, while existing temporal knowledge graphs reasoning methods are almost designed for multi-label prediction.

# 3 PRELIMINARIES

## 3.1 Notations and Definitions

To facilitate the elaboration of the KG-based SMEs supply chain prediction, we give some notations in Table 1.

**SMEs Supply Chain Relationships.** Let $E = \{e_1, e_2, \ldots, e_{|E|}\}$ denote a set of SMEs. For a given historical time window $\tau = [T - \tau + 1, T]$, their supply chain relationships can be defined as $R = \{r_{e_i,e_j} \in \{0, 1, \emptyset\}\}_{i,j=1,2,\ldots,|E|}$, where $|\cdot|$ denotes the volume of a set. Particularly, for two SMEs $e_i$ and $e_j$, $r_{e_i,e_j} = 1$ (positive) or 0 (negative) denotes whether a supply chain relationship between them or not, whereas $r_{e_i,e_j} = \emptyset$ means an unknown relationship. In addition, the basic information $B = \{b_1, b_2, \ldots, b_{|E|}\}$ for each enterprise $e_i \in E$ (e.g., industry, scale, etc.) is also given. Here, we only consider the setting of each attribute in $B$ as a discrete categorical variable, while the continuous real-valued one will be pre-processed into a discrete categorical variable by binning.

**Spatial-Temporal Knowledge Graph.**[1] As a special type of KG, the STKG $\mathcal{G}$ can be defined as a sequence of statics KG slices $\mathcal{G} = \{\mathcal{G}_{T-\tau+1}, \ldots, \mathcal{G}_{T-1}, \mathcal{G}_T\}$. For each slice $\mathcal{G}_t$ at $t$, similar to the statics KG, we let $\mathcal{E}_t = \left\{\overline{e}_1^t, \overline{e}_2^t, \ldots, \overline{e}_{|\mathcal{E}_t|}^t\right\}$ and $\phi : \mathcal{E}_t \rightarrow \mathcal{A}_t$ denote the sets of entities and an entity type mapping function, respectively, where $\mathcal{A}_t = \{a_i^t\}$, $i = 1, \cdots, |\mathcal{A}_t|$, is the entity type set, such as enterprise, natural person, location and industry, etc. It is worth noting that the SMEs set $E$ is indeed a subset of the entity set $\mathcal{E}_t$, i.e., $E \subseteq \mathcal{E}_t$. Similarly, we also denote $\mathcal{R}_t = \{r^t_{\overline{e}_i,\overline{e}_j}\}_{i,j=1,2,\ldots,|\mathcal{E}_t|}$ and $\psi : \mathcal{L}_t \rightarrow \mathcal{R}_t$ as the sets of relations and a relation type mapping function, respectively, where each one-hop link $l^t_{\overline{e}_u,\overline{e}_v} \in \mathcal{L}_t$ connecting two entities $\overline{e}_u, \overline{e}_v$ of $\mathcal{G}_t$ can be mapped to a certain type of relation $\psi(l^t_{\overline{e}_u,\overline{e}_v}) = r^t_{\overline{e}_u,\overline{e}_v}$. Then, the STKG slice can be defined as a directed graph $\mathcal{G}_t = (\mathcal{E}_t, \mathcal{R}_t)$.

**Task Definition.** Given the historical supply chain relationships $R$, the basic information $B$ and the STKG $\mathcal{G}$, we aim to predict whether the enterprise $e_i$ will have any supply chain relationships with $e_j$ at future time $T + 1$:

$$\widehat{y}_{i,j}^{T+1} = \mathcal{F}_\Theta(< e_i, e_j > | R, B, \mathcal{G}), \tag{1}$$

where $\mathcal{F}(\cdot)$ denotes the underlying model with parameters $\Theta$ that we need to learn, and $\widehat{y}_{i,j}^{T+1}$ means the predicted probability of supply chain relationships between enterprise $e_i$ and $e_j$ at $T + 1$.

## 3.2 Model Overview

Fig. 2 has shown a running example of the proposed JRCL, and its framework mainly consists of three modules:

(a) **Multi-view Relation Sequences Mining (MvR)**. It aims to reveal the semantic information encoded in the constructed large-scale financial STKG more extensively;
(b) **Relational Co-evolution Learning (CoEvo)**. To capture the regularity of both temporal dependency and cooperative correlation in connectivity patterns evolving, a relational co-evolution learning module is also developed;
(c) **Multiple Random Subspaces (MRS)**. As a representation refining layer, the MRS is designed for aggregating knowledge and making the final prediction.

[1] More details about the STKG are given in the Appendix A.2

**Table 1: Notation and description**

| Notation | Description |
|---|---|
| $E = \{e_i\}$ | The set of $\|E\|$ SMEs |
| $R = \{r_{e_i,e_j}\}$ | The supply chain relationships set |
| $B = \{b_i\}$ | The set of $\|E\|$ basic information |
| $\mathcal{E}_t = \{\overline{e}_k^t\}$ | The set of $\|\mathcal{E}\|$ entities in STKG slice at $t$ |
| $\mathcal{A}_t = \{a_i^t\}$ | The $\|\mathcal{A}_t\|$ entity types set in STKG slice at $t$ |
| $\phi : \mathcal{E}_t \rightarrow \mathcal{A}_t$ | The entity type mapping function |
| $\mathcal{L}_t = \{l^t_{\overline{e}_u,\overline{e}_v}\}$ | The set of $\|\mathcal{L}_t\|$ links in STKG slice at $t$ |
| $\mathcal{R}_t = \{r^t_{\overline{e}_i,\overline{e}_j}\}$ | The set of $\|\mathcal{R}_t\|$ relations in STKG slice at $t$ |
| $\psi : \mathcal{L}_t \rightarrow \mathcal{R}_t$ | The relation type mapping function |
| $\mathcal{G}_t = (\mathcal{E}_t, \mathcal{R}_t)$ | The STKG slice at time $t$ |
| $\mathcal{G} = \{\mathcal{G}_t\}$ | The spatial-temporal knowledge graph |
| $y_{i,j}^{T+1}, \widehat{y}_{i,j}^{T+1}$ | The ground truth and predicted labels for SMEs pair $< e_i, e_j >$ at $T + 1$ |

# 4 METHODOLOGY

## 4.1 Multi-view Relation Sequences Mining

Multi-view relation sequences in knowledge graph paths jointly represent two components: a semantic view from ontological concepts, and a geographical view of specific entities.

*4.1.1* ***Semantic-view Relation Sequence.*** The ontological concepts in STKG clearly outline direct or indirect connectivity patterns between SMEs, which shall constitute one or several KG paths [38]. To better encode these connectivity patterns, we extract the $K$-shortest paths from STKG by using [41] as the semantic-view relation sequence (SenRS). For the $t$-th slice in STKG, we formally define the set of KG paths (a.k.a. semantic-view relation sequences) connecting $e_i$ and $e_j$ as $P_{i,j}^t$ and the $k$-th path $p_{i,j}^t|_k \in P_{i,j}^t$, $k = 1, \ldots, K$ is a set of sequence of entities and relations. To be specific, taking a given $k$-th path as an example, $p_{i,j}^t|_k = \{e_i, \psi(l^t_{e_i,\overline{e}_u}), \overline{e}_u^t, \ldots, \overline{e}_v^t, \psi(l^t_{\overline{e}_v,e_j}), e_j\}$. By seeing these entities or relations ID in $P_{i,j}^t$ as sequences of words in sentences [21] and following the formulations $F_{\Theta_1}(\cdot)$ given in [14], we can embed each path $p_{i,j}^t|_k$ in $P_{i,j}^t$ and sum them into $E(P_{i,j}^t) \in \mathbb{R}^{K\times d}$.

*4.1.2* ***Geographical-view Relation Sequence.*** The uneven regional economic development and diverse industrial structures mean that the geographical information should also be considered in supply chain prediction. However, it is hard to capture such geographical relationships between enterprises in an explicit manner. Fortunately, in the constructed STKG, each entity has been linked with a unique location entity, and it provides a possible way to address this issue. To effectively reveal the geographical connectivity patterns between the SMEs $e_i$ and $e_j$ in STKG at time $t$, with the extracted KG paths $P_{i,j}^t$, we define geographical relation sequence (GeoRS) as $Loc_{i,j}^t$. To be specific, for each SenRS $p_{i,j}^t|_k \in P_{i,j}^t$, we have the associated $loc_{i,j}^t|_k = \{loc_i^t, loc_u^t, \ldots, loc_v^t, loc_j^t\}$, where the entities such as $loc_u^t, loc_v^t \subseteq \mathcal{E}_t$ denotes the corresponding location entity in the $t$-th slice of STKG $\mathcal{G}_t$ and links with the original entities in SenRS $\overline{e}_u^t$ and $\overline{e}_v^t$. Similarly, with the given formulations in $F_{\Theta_1}(\cdot)$, we can also embed each $loc_{i,j}^t|_k$ in $Loc_{i,j}^t$ into $E(Loc_{i,j}^t) \in \mathbb{R}^{K\times d}$.

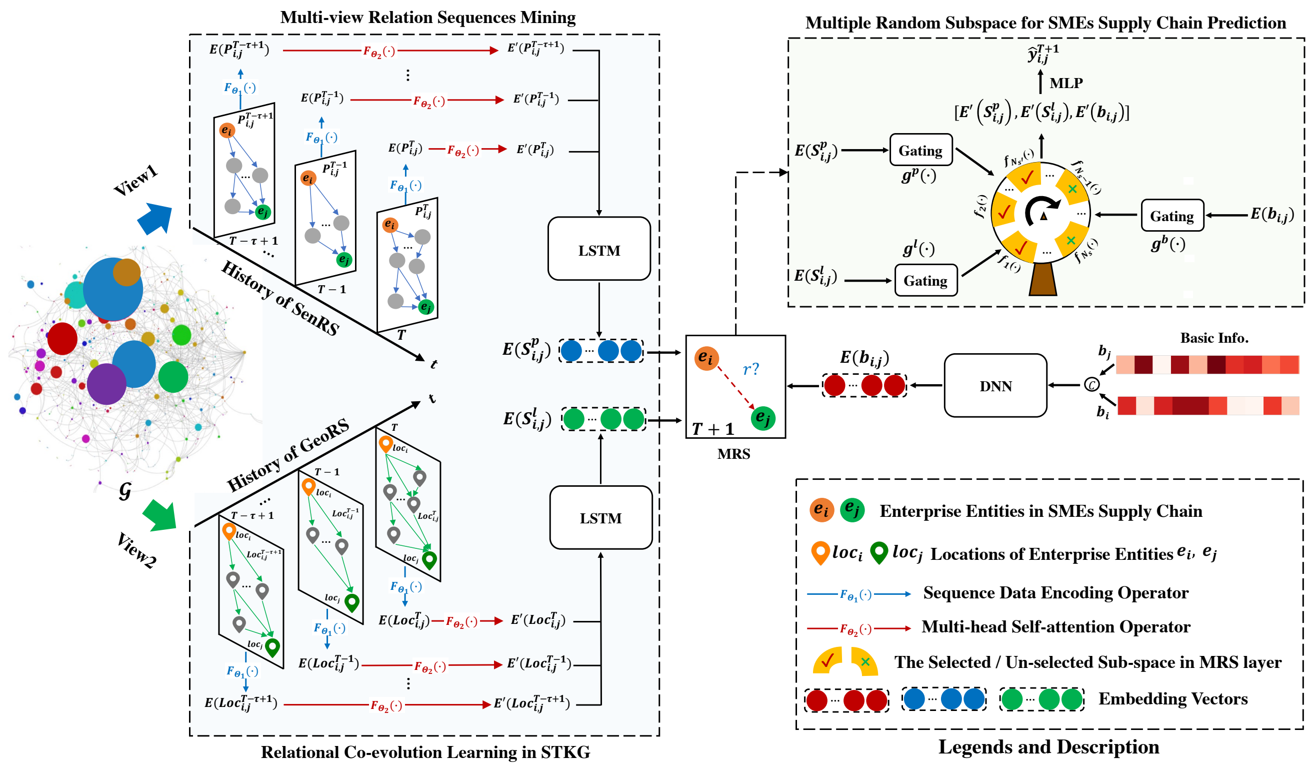


**Figure 2: Graphical illustration of learning joint relational co-evolution in spatial-temporal knowledge graph for SMEs supply chain prediction. It is mainly composed of three modules: (a) Multi-view Relation Sequences Mining (MvR); (b) Relational Co-evolution Learning (CoEvo); (c) Multiple Random Subspaces (MRS).**

## 4.2 Relational Co-evolution Learning

*4.2.1 **Cooperative Correlation Refining**.* The common characteristic for multiple sequences is that different sequences are often inherently inter-connected or correlated with each other [39]. For example, the SMEs in different semantic-view relation sequences are correlated since there exist competition and cooperation relationships among them. To be more specific, for both semantic-view and geographical-view relation sequences $P^t_{i,j}$ and $Loc^t_{i,j}$ at time $t$, the cooperative correlation among each individual sequence $p^t_{i,j}|_k$ and $loc^t_{i,j}|_k$ should be well explored. Furthermore, their embedding $E(P^t_{i,j}) \in \mathbb{R}^{K\times d}$ and $E(Loc^t_{i,j}) \in \mathbb{R}^{K\times d}$ can be fed into a $N_H$-head self-attention network (denoted as $F_{\Theta_2}(\cdot)$) [30] for correlation-wise representation refining, respectively. For $E(P^t_{i,j})$, we have the refined representation $E'(P^t_{i,j}) \in \mathbb{R}^{K\times d_1}$:

$$\begin{aligned} E'(P^t_{i,j}) &= MultiHeadAttention(E(P^t_{i,j})) \\ &= [E'^{(1)}_{(P^t_{i,j})} \oplus E'^{(2)}_{(P^t_{i,j})} \dots \oplus E'^{(N_H)}_{(P^t_{i,j})}] \cdot W^H \end{aligned} \tag{2}$$

where $E'^{(h)}_{(P^t_{i,j})} = head^{(h)}(E(P^t_{i,j})) \in \mathbb{R}^{K\times d_2}$ is the learned refined representations from $h$-th attention head, $\oplus$ is the concatenation operation and $W^H \in \mathbb{R}^{(d_2\times N_H)\times d_1}$ is a linear projection matrix. For the $h$-th $head^{(h)}(E(P^t_{i,j}))$, it is given by:

$$head^{(h)}(E(P^t_{i,j})) = softmax(\frac{Q^{(h)}_{i,j} \cdot K^{(h)\mathrm{T}}_{i,j}}{\sqrt{d_1}}) \cdot V^{(h)}_{i,j}, \tag{3}$$

where $\{Q^{(h)}_{i,j}, K^{(h)}_{i,j}, V^{(h)}_{i,j}\} = \{E(P^t_{i,j}) \cdot W^{(h)}_q, E(P^t_{i,j}) \cdot W^{(h)}_k, E(P^t_{i,j}) \cdot W^{(h)}_v\}$ with $W^{(h)}_q, W^{(h)}_k, W^{(h)}_v \in \mathbb{R}^{d\times d_2}$ being the globally shared projection matrices. Meanwhile, for the GeoRS $Loc^t_{i,j}$ at time $t$, the refined representation $E'(Loc^t_{i,j}) \in \mathbb{R}^{K\times d_1}$ can be also obtained with the same formulations in $F_{\Theta_2}(\cdot)$.

*4.2.2 **Temporal Dependency Modeling**.* To consistently model the temporal dependencies in the evolution of supply chain connectivity patterns from $T-\tau+1$ to $T$, LSTM [8] is employed. In practice, the refined representations of historical SenRS are flattened as $S^p_{i,j} = \{E'(P^{T-\tau+1}_{i,j}), \dots, E'(P^{T-1}_{i,j}), E'(P^T_{i,j})\} \in \mathbb{R}^{(K\times d_1)\times\tau}$ and fed into an LSTM network $LSTM^{(p)}$. Then, we have:

$$E(S^P_{i,j}) = LSTM^{(P)}(E'(P^{T-\tau+1}_{i,j}), \dots, E'(P^{T-1}_{i,j}), E'(P^T_{i,j})), \tag{4}$$

where $E(S^p_{i,j}) \in \mathbb{R}^d$ is the learned representation for $S^p_{i,j}$. Similarly, with another network $LSTM^{(l)}$ the representation $E(S^l_{i,j}) \in \mathbb{R}^d$ for history of GeoRS $S^l_{i,j}$ can also be learned.

**Algorithm 1** Joint Relational Co-evolution Learning

**Require:**
$R$: historical supply chain relationships, $B$: basic information of SMEs, $\mathcal{G}$: spatial-temporal knowledge graph

**Ensure:**
$\hat{y}_{i,j}^{T+1}$: supply chain relationship probability between $e_i$ and $e_j$

**for** $t \in [T-\tau+1, T]$ **do**
  $P_{i,j}^t, Loc_{i,j}^t \leftarrow MvR(\mathcal{G}_t)$
  **for** $k \in (1, 2, \ldots, K)$ **do**
    $E(p_{i,j}^t|_k), E(loc_{i,j}^t|_k) \in \mathbb{R}^d \leftarrow F_{\Theta_1}(p_{i,j}^t|_k, loc_{i,j}^t|_k)$
  **end for**
  $E'(P_{i,j}^t), E'(Loc_{i,j}^t) \in \mathbb{R}^{K \times d_1} \leftarrow F_{\Theta_2}(E(P_{i,j}^t), E(Loc_{i,j}^t))$
**end for**
**for** $t \in [T-\tau+1, T]$ **do**
  $E(S_{i,j}^p), E(S_{i,j}^l) \in \mathbb{R}^d \leftarrow LSTM^{(p,l)}(E'(P_{i,j}^t), E'(Loc_{i,j}^t))$
**end for**
$E(b_{i,j}) \in \mathbb{R}^d \leftarrow DNN(b_{i,j})$
$E'(S_{i,j}^p), E'(S_{i,j}^l), E'(b_{i,j}) \leftarrow MRS(E(S_{i,j}^p), E(S_{i,j}^l), E(b_{i,j}))$
$\hat{y}_{i,j}^{T+1} \leftarrow MLP([E'(S_{i,j}^p), E'(S_{i,j}^l), E'(b_{i,j})])$

## 4.3 Multiple Random Subspaces

In addition to the evolution patterns $E(S_{i,j}^p)$ and $E(S_{i,j}^l)$ learned from different views, the basic information of SMEs should also be considered in supply chain prediction tasks. For a SMEs pair $< e_i, e_j >$, their basic information $b_i$ and $b_j$ can be concatenated as $b_{i,j}$ and mapped into a $d$-dimension embedding vector $E(b_{i,j}) \in \mathbb{R}^d$ using a one-layer MLP. However, the inherent couplings among the multi-information bring challenges to effective knowledge aggregation. Thus, motivated by the original idea of Mixture-of-Experts (MoE) networks, we design a multiple random subspace (MRS) representation learning layer [9] for further disentangling the learned representations $E(S_{i,j}^p)$, $E(S_{i,j}^l)$ and $E(b_{i,j})$ one by one. For $E(S_{i,j}^p)$, we project it into $N_s$ $d_s$-dimensional subspaces as: $\{f_n(E(S_{i,j}^p)) \in \mathbb{R}^{d_s}\}_{n=1,2,\ldots,N_s}$. For these, we expect that they can capture connectivity patterns from different aspects, like the expert networks in the MoE layer. In fact, it is hard to realize this due to the winner-take-all issue in training. To address it, we randomly sample a subset of size $N_s'$ from the $N_s$ subspaces to strengthen the uncertainty during parameter confirmation. Then, we weight the representation from $N_s'$ subspaces with an individual gating network $g^p(\cdot)$ as:

$$E'(S_{i,j}^p) = \sum_{n=1}^{N_s'} g_n^p(E(S_{i,j}^p)) \cdot f_n(E(S_{i,j}^p)), \tag{5}$$

where $g^p(E(S_{i,j}^p)) = softmax(W_g^p \cdot E(S_{i,j}^p))$ with trainable weight matrix $W_g^p \in \mathbb{R}^{N_s' \times d_s}$, and $g_n^p(\cdot) \in (0,1)$ is the $n$-th output of the gating network. Similarly, $E(S_{i,j}^l)$ and $E(b_{i,j})$ can also be further refined into $E'(S_{i,j}^l) \in \mathbb{R}^{d_s}$ and $E'(b_{i,j}) \in \mathbb{R}^{d_s}$, respectively. Finally, the prediction of the supply chain relationship between $e_i$ and $e_j$ at time $T+1$ can be obtained by using an MLP layer:

$$\hat{y}_{i,j}^{T+1} = MLP([E'(S_{i,j}^p), E'(S_{i,j}^l), E'(b_{i,j})]). \tag{6}$$

## 4.4 Model Optimization

Generally, the SMEs supply chain prediction task is defined as a binary classification problem, and the cross entropy is set as the loss function,

$$\mathcal{L}_{Class} = -\left( \sum_{y_{i,j}^{T+1} \in R^+} \log \hat{y}_{i,j}^{T+1} + \sum_{y_{i,j}^{T+1} \in R^-} \log(1 - \hat{y}_{i,j}^{T+1}) \right), \tag{7}$$

in which $R^+$ and $R^-$ are the samples with positive and negative target values. In summary, the implementation details of the proposed JRCL framework are outlined in Algorithm 1.

# 5 EXPERIMENTAL RESULTS AND ANALYSES

In this section, we perform experiments on four real-world supply chain prediction datasets to evaluate our proposed method. We aim to answer the following research questions:

- RQ1: How to set the SMEs supply chain prediction tasks?
- RQ2: How to evaluate the performance of the JRCL?
- RQ3: How does the JRCL perform and why?

## 5.1 Data Description and Settings (RQ1)

The detailed statistics for the large-scale industrial datasets of the SMEs supply chain prediction and the constructed financial STKG collected by MYBank, Ant Group, are given in the appendix A.1-2. Here is a brief summary.

- **SMEs Supply Chain Dataset [2]:** It includes rich basic profiles and facts of supply chain relationships between SMEs across more than 100 real-world industries such as manufacturing, civil engineering, retailing, and wholesaling, etc.
- **Ant-MYBank Financial STKG [3]:** It is a large-scale financial STKG used to depict financial relationships among millions of SMEs, covering more than 20 relation types and 4 entity types, with a total of 90 million entities and 2.8 billion facts.

For SMEs supply chain prediction, more than 15 million ground-truth relationships are used for model training and evaluation in civil engineering, manufacturing, retailing, and wholesaling industries, respectively. We also constructed a large-scale financial STKG to help explore these supply chain relationships more effectively.

## 5.2 Experimental Settings (RQ2)

*5.2.1* ***Metrics.*** We used Accuracy (ACC), the area under the receiver operating characteristic curve (AUROC), the area under the precision-recall curve (AUPRC) and the minimum precision and sensitivity min (Se, P+) to evaluate the performance of models in SMEs supply chain prediction tasks.

*5.2.2* ***Baselines.*** We compared the proposed JRCL with some traditional predictors [3, 5], some recent KG-based recommendation methods [35, 36] and temporal knowledge graph reasoning models [15, 40]. Noted, in order to ensure the fairness of the comparison, we reproduced the results for each baseline with their original open-source implementations and all the reproduced ones have been carefully fine-tuned by the grid-searching strategy.

[2]https://bksupplychain-assets.mybank.cn/
[3]https://tech.antfin.com/products/TuGraph

**Table 2: Performance comparison (the best is in red and the second is in blue).**

| Dataset | Model | SMEs Supply Chain Prediction (Bootstrapping = 1000) | | | |
|---|---|---|---|---|---|
| | | ACC | AUROC | AUPRC | min(Se, P+) |
| Civil Eng. | LR [3] | 0.7219 ± 0.0161 | 0.6014 ± 0.0242 | 0.4027 ± 0.0355 | 0.4154 ± 0.0319 |
| | GBDT [5] | 0.7290 ± 0.0167 | 0.5769 ± 0.0231 | 0.3659 ± 0.0316 | 0.3509 ± 0.0288 |
| | KGAT [35] | 0.7272 ± 0.0163 | 0.6281 ± 0.0235 | 0.3920 ± 0.0324 | 0.4167 ± 0.0298 |
| | KGIN [36] | 0.7358 ± 0.0166 | 0.6302 ± 0.0238 | 0.4043 ± 0.0338 | 0.4132 ± 0.0318 |
| | TiRGN [15] | **0.7437 ± 0.0167** | 0.6403 ± 0.0240 | 0.4242 ± 0.0359 | 0.4593 ± 0.0307 |
| | ST-GNN [40] | 0.7402 ± 0.0162 | **0.7307 ± 0.0204** | **0.4808 ± 0.0371** | **0.5028 ± 0.0310** |
| | **JRCL** | **0.8161 ± 0.0143** | **0.8317 ± 0.0169** | **0.6607 ± 0.0375** | **0.6407 ± 0.0283** |
| Manufacturing | LR [3] | 0.7223 ± 0.0185 | 0.5657 ± 0.0297 | 0.2972 ± 0.0340 | 0.3251 ± 0.0346 |
| | GBDT [5] | 0.7645 ± 0.0180 | 0.5882 ± 0.0273 | 0.3186 ± 0.0386 | 0.3099 ± 0.0288 |
| | KGAT [35] | 0.7650 ± 0.0180 | 0.6160 ± 0.0274 | 0.3262 ± 0.0370 | 0.3440 ± 0.0315 |
| | KGIN [36] | 0.7479 ± 0.0182 | 0.6192 ± 0.0303 | 0.3567 ± 0.0380 | 0.4057 ± 0.0360 |
| | TiRGN [15] | 0.7548 ± 0.0186 | 0.6455 ± 0.0276 | 0.3605 ± 0.0393 | 0.3921 ± 0.0370 |
| | ST-GNN [40] | **0.7658 ± 0.0181** | **0.7328 ± 0.0247** | **0.4470 ± 0.0452** | **0.4375 ± 0.0383** |
| | **JRCL** | **0.7680 ± 0.0175** | **0.7380 ± 0.0256** | **0.4765 ± 0.0493** | **0.4733 ± 0.0378** |
| Wholesaling | LR [3] | 0.8081 ±0.0158 | 0.5835 ± 0.0282 | 0.2311 ± 0.0278 | 0.2654 ± 0.0317 |
| | GBDT [5] | 0.8096 ± 0.0156 | 0.5946 ± 0.0282 | 0.2788 ± 0.0372 | 0.2607 ± 0.0323 |
| | KGAT [35] | **0.8340 ± 0.0148** | 0.6183 ± 0.0304 | 0.2403 ± 0.0304 | 0.2903 ± 0.0326 |
| | KGIN [36] | **0.8417 ± 0.0144** | 0.6059 ± 0.0337 | **0.2926 ± 0.0422** | 0.3206 ± 0.0392 |
| | TiRGN [15] | 0.8111 ± 0.0150 | 0.5992 ± 0.0285 | 0.2870 ± 0.0362 | **0.3352 ± 0.0372** |
| | ST-GNN [40] | 0.8121 ± 0.0152 | **0.6445 ± 0.0269** | 0.2830 ± 0.0346 | 0.3347 ± 0.0388 |
| | **JRCL** | 0.8112 ± 0.0153 | **0.7824 ± 0.0231** | **0.4469 ± 0.0456** | **0.5000 ± 0.0399** |
| Retailing | LR [3] | 0.7897 ± 0.0180 | 0.5191 ± 0.0317 | 0.2137 ± 0.0278 | 0.2529 ± 0.0355 |
| | GBDT [5] | 0.8042 ± 0.0175 | 0.5621 ± 0.0296 | 0.2515 ± 0.0360 | 0.2720 ± 0.0421 |
| | KGAT [35] | **0.8054 ± 0.0180** | 0.6043 ± 0.0334 | 0.3059 ± 0.0445 | 0.3341 ± 0.0408 |
| | KGIN [36] | 0.8013 ± 0.0179 | 0.6134 ± 0.0344 | 0.3166 ± 0.0443 | 0.3374 ± 0.0421 |
| | TiRGN [15] | **0.8071 ± 0.0175** | 0.6557 ± 0.0291 | **0.3283 ± 0.0453** | 0.3484 ± 0.0428 |
| | ST-GNN [40] | 0.8052 ± 0.0169 | **0.6692 ± 0.0293** | 0.3196 ± 0.0426 | **0.3576 ± 0.0404** |
| | **JRCL** | 0.8043 ± 0.0175 | **0.7225 ± 0.0301** | **0.4561 ± 0.0517** | **0.2473 ± 0.0028** |

## 5.3 Experimental Results and Analyses (RQ3)

*5.3.1* ***Performance Comparison.*** Table 2 has summarized the experimental results of performance comparison. Generally, the JRCL can outperform almost all the state-of-the-art competitors in four datasets significantly, demonstrating the effectiveness of our model. To be more specific, we have the following findings:

- As simple predictors, although well-designed features are given, both plain LR [3] and GBDT [5] performed poorly by using only the basic information. It shows that it is necessary to introduce the knowledge graph in order to better handle such a link prediction task towards relational data.
- The proposed MSCL also outperformed the recent KG-based recommendation methods [35, 36]. Overall, the performances of them are better than plain LR and GBDT with the help of the knowledge graph. Specifically, as state-of-the-art GNN-based recommenders, they are developed to model high-order connectivity and recursively integrate the long-range relational information between the focused entities in an end-to-end fashion. However, as the dynamics of facts along the timeline in KG are not considered, the improvements remain not sufficient.
- Compared to those static KG-based recommendation methods, the temporal KG reasoning methods [15] have made further progress by comprehensively considering the sequential independence of historical facts along the timeline in KG. In addition, the spatial-temporal aware graph neural network model [40] also shows a workable performance by aggregating both neighbor and relational information in STKG. However, the performance is also insufficient as temporal dependency and cooperative correlation are not considered synergistically.

To sum up, the proposed JRCL has made a remarkable performance in SMEs supply chain prediction tasks across four real-world industries compared to related state-of-the-art baselines.

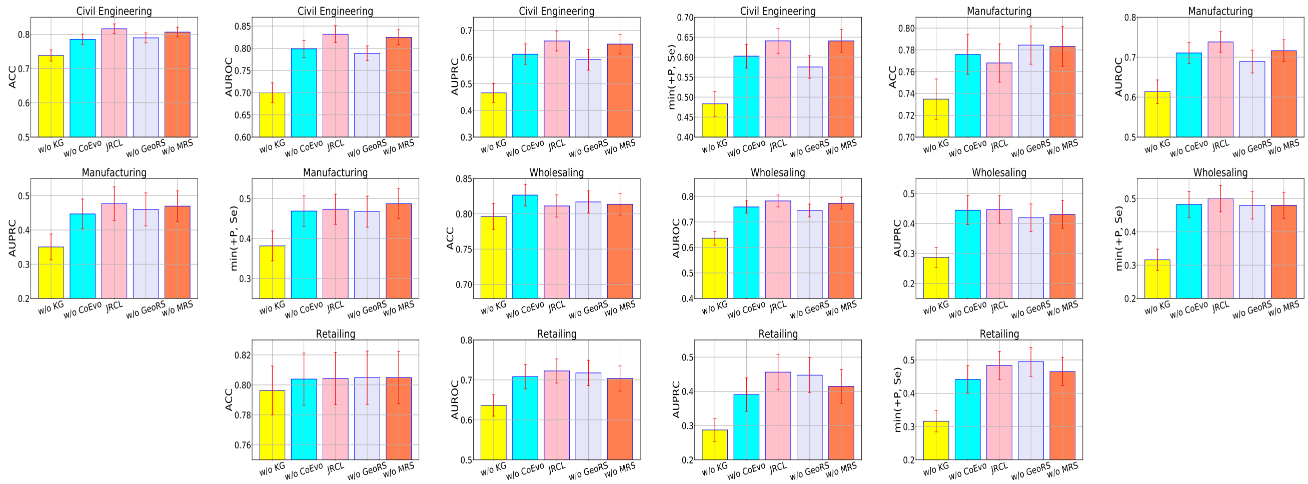


**Figure 3: Results of ablation studies.**

*5.3.2 **Ablation Studies**.* We conducted the ablation studies with the following settings:

- **JRCL (w/o KG)**: To demonstrate the effectiveness of introducing auxiliary information from knowledge graphs, we only applied a plain DNN to make a prediction with the basic information of SMEs.
- **JRCL (w/o CoEvo)**: We also removed the cooperative correlation refining module to demonstrate its usefulness in capturing the inherent correlation among each individual sequence in multi-view relation sequences at time $t$.
- **JRCL (w/o GeoRS)**: We only consider the SenRs to demonstrate the effectiveness of modeling the geographical relationships between enterprises in an explicit manner.
- **JRCL (w/o MRS)**: We removed the MRS module to demonstrate the effectiveness of aggregating multiple representations by coordinating the coupling between temporal dependency and cooperative correlation adequately.

The results given in Fig. 3 have shown that all variants of JRCL perform worse than the original JRCL, proving its effectiveness in each module. From the results, the KG (w/o KG) has the greatest impact on performance, which shows an obvious fact that the auxiliary information encoded in knowledge graphs is crucial for the supply chain relationships prediction. The relational co-evolution learning (w/o CoEvo) also has a consistent impact on all the datasets, which shows the necessity of learning the correlations from the evolution of spatial-temporal patterns in multiple spatial-temporal knowledge graph paths. The consideration of geographical-view relation sequence (w/o GeoRS) is also helpful for performance improvement, proving the fact that the geographical relationships between SMEs should also be considered in the supply chain prediction. In addition, considering both the compatibility and complementarity during knowledge aggregation, the multiple random subspaces (w/o MRS) also lead to positive performance in most cases. These results further show that different variants of the JRCL are all helpful for performance improvement.

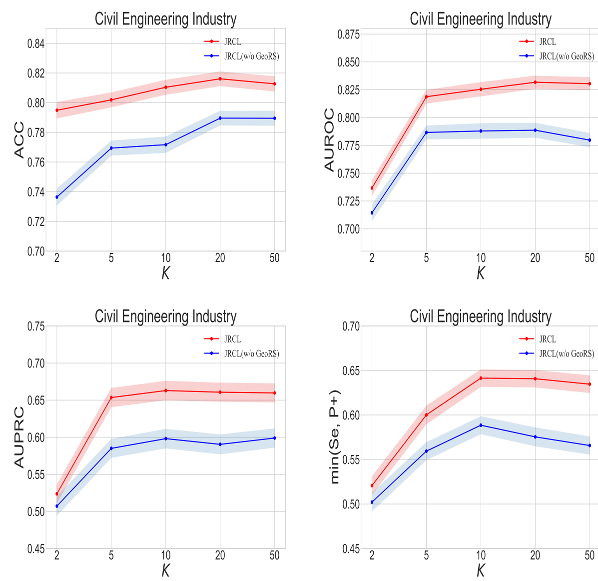


**Figure 4: Analysis of parameter $K$.**

Moreover, we also explore how the main hyper-parameters involved in JRCL affect the model performance for different variants. Specifically, we check the sensitivity of hyper-parameter $K$ and Fig. 4 shows the performance of different variants under different hyper-parameters. We can see that with a consideration of both efficiency and performance, a relatively smaller $K$ (not too small) and a larger (not too large) hyper-parameter lead to the best result. Meanwhile, we can also see that all the performance is insufficient without consideration of geographical relation sequence (GeoRS). It also proves the consideration of geographical-view relation sequence (w/o GeoRS) is helpful for performance improvement. Noted, we take the civil engineering dataset as an example, and other datasets also show a similar trend.

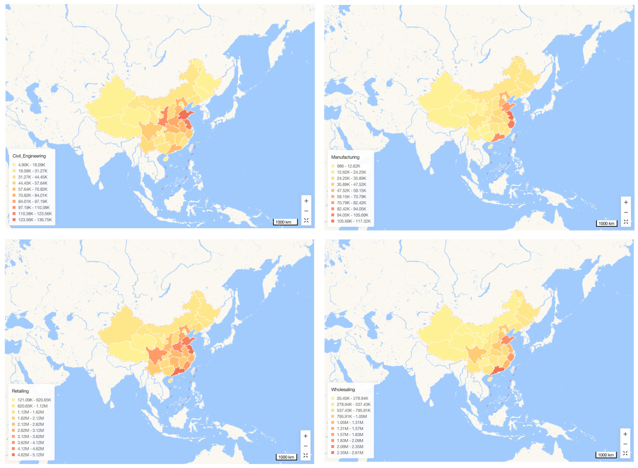

Figure 5: The geographical distribution of SMEs from different industries in China.

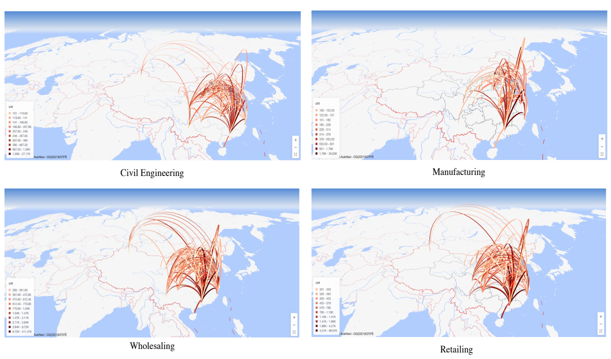


Figure 6: The flow maps of SMEs supply chain from different industries in China.

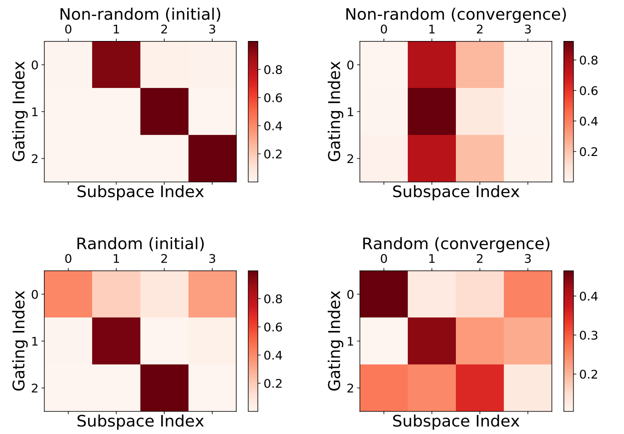


Figure 7: Analysis of Gating Networks in MRS.

5.3.3 ***Case Study***. In order to analyze the experimental results more intuitively, we give a case study with real data.

As we discussed before, geographic information is very useful for supply chain relationship prediction. Here, we first investigate SMEs' geographic distribution and the flow map of their supply chains to help give it a more intuitive interpretation in Fig 5-6 through visualization with a self-developed tool [4]). To be specific, the geographic distribution rules of SMEs and their supply chains are highly related to industries. Taking the samples used in our experiment from China as an example, we can clearly find that the SMEs from manufacturing and wholesale industries are mainly distributed in the southeastern coastal region of China and the SMEs from the civil engineering industry are mainly distributed in several provinces of China, showing a typical spatial aggregation characteristic. From the results of ablation studies, we can also find the GeoRS is more significant for performance improvements in these three industries with a higher geographic distribution correlation, while that is less significant in the retail industry. It conforms to our assumptions.

Meanwhile, to show how the multiple random subspaces address the winner-take-all issue in parameter training, we give an analysis of its gating networks. Specifically, we randomly select a batch of samples from the civil engineering industry dataset and draw the distribution of gating weights for each individual subspace network from the initial to convergence status in Fig. 7. Overall, we can observe that each subspace network in the MRS layer can capture patterns from different aspects of the convergence status, like the expert networks in the MoE layer, while the non-random one has failed. It shows the benefits of heightened uncertainty during parameter confirmation in the MRS layer, which also conforms to our assumptions.

[4]https://deepinsight.alipay.com/

# 6 CONCLUSION AND FUTURE WORK

This paper proposes a novel joint relational co-evolution representation learning framework in spatial-temporal knowledge graph for SMEs supply chain prediction. Specifically, the proposed multi-view relational sequence mining method has been performed to reveal the multiple semantic information encoded in the constructed large-scale financial STKG. Meanwhile, the relational co-evolution learning module and the multiple random subspace representation learning layer have also been developed to aggregate the multi-view semantic information from the spatial-temporal view by considering both compatibility and complementarity. In addition, we experimented on four real-world industrial-level datasets to show the effectiveness of the proposed method.

For future work, we will explore more challenging issues in such knowledge graph enhanced SMEs relationship prediction tasks, such as the prediction of the competitive relationship between them.

# 7 ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers for their insightful comments and valuable suggestions.

# REFERENCES


[1] Ozlem Bak, Sarah Shaw, Claudia Colicchia, and Vikas Kumar. 2023. A Systematic Literature Review of Supply Chain Resilience in Small-Medium Enterprises (SMEs): A Call for Further Research. *IEEE Trans. Engineering Management* 70, 1 (2023), 328–341. https://doi.org/10.1109/TEM.2020.3016988

[2] Borui Cai, Yong Xiang, Longxiang Gao, He Zhang, Yunfeng Li, and Jianxin Li. 2022. Temporal Knowledge Graph Completion: A Survey. *CoRR* abs/2201.08236 (2022). arXiv:2201.08236 https://arxiv.org/abs/2201.08236

[3] Stephan Dreiseitl and Lucila Ohno-Machado. 2002. Logistic regression and artificial neural network classification models: a methodology review. *Journal of biomedical informatics* 35, 5-6 (2002), 352–359.

[4] Pankaj Dutta, Tsan-Ming Choi, Surabhi Somani, and Richa Butala. 2020. Blockchain technology in supply chain operations: Applications, challenges and research opportunities. *Transportation research part e: Logistics and transportation review* 142 (2020), 102067.

[5] Jerome H Friedman. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics* (2001), 1189–1232.

[6] Shijie Geng, Zuohui Fu, Juntao Tan, Yingqiang Ge, Gerard de Melo, and Yongfeng Zhang. 2022. Path Language Modeling over Knowledge Graphsfor Explainable Recommendation. In *WWW '22*. ACM, 946–955. https://doi.org/10.1145/3485447.3511937

[7] Qingyu Guo, Fuzhen Zhuang, Chuan Qin, Hengshu Zhu, Xing Xie, Hui Xiong, and Qing He. 2022. A Survey on Knowledge Graph-Based Recommender Systems. *IEEE Trans. Knowl. Data Eng.* 34, 8 (2022), 3549–3568. https://doi.org/10.1109/TKDE.2020.3028705

[8] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Comput.* 9, 8 (1997), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[9] Robert A. Jacobs, Michael I. Jordan, Steven J. Nowlan, and Geoffrey E. Hinton. 1991. Adaptive Mixtures of Local Experts. *Neural Comput.* 3, 1 (1991), 79–87. https://doi.org/10.1162/neco.1991.3.1.79

[10] Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S. Yu. 2022. A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. *IEEE Trans. Neural Networks Learn. Syst.* 33, 2 (2022), 494–514. https://doi.org/10.1109/TNNLS.2021.3070843

[11] Wen-Hui Jiang, Ling Xu, Zhen-Song Chen, Kannan Govindan, and Kwai-Sang Chin. 2022. Financing equilibrium in a capital constrained supply Chain: The impact of credit rating. *Transportation Research Part E: Logistics and Transportation Review* 157 (2022), 102559.

[12] Jaehun Jung, Jinhong Jung, and U Kang. 2021. Learning to Walk across Time for Interpretable Temporal Knowledge Graph Completion. In *KDD '21*. ACM, 786–795. https://doi.org/10.1145/3447548.3467292

[13] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *5th International Conference on Learning Representations, ICLR 2017*. OpenReview.net. https://openreview.net/forum?id=SJU4ayYgl

[14] Quoc V. Le and Tomás Mikolov. 2014. Distributed Representations of Sentences and Documents. In *ICML 2014 (JMLR Workshop and Conference Proceedings, Vol. 32)*. JMLR.org, 1188–1196. http://proceedings.mlr.press/v32/le14.html

[15] Yujia Li, Shiliang Sun, and Jing Zhao. 2022. TiRGN: Time-Guided Recurrent Graph Network with Local-Global Historical Patterns for Temporal Knowledge Graph Reasoning. In *IJCAI 2022*, Luc De Raedt (Ed.). ijcai.org, 2152–2158. https://doi.org/10.24963/ijcai.2022/299

[16] Yujia Li, Oriol Vinyals, Chris Dyer, Razvan Pascanu, and Peter W. Battaglia. 2018. Learning Deep Generative Models of Graphs. *CoRR* abs/1803.03324 (2018). arXiv:1803.03324 http://arxiv.org/abs/1803.03324

[17] Zixuan Li, Saiping Guan, Xiaolong Jin, Weihua Peng, Yajuan Lyu, Yong Zhu, Long Bai, Wei Li, Jiafeng Guo, and Xueqi Cheng. 2022. Complex Evolutional Pattern Learning for Temporal Knowledge Graph Reasoning. In *ACL 2022*. Association for Computational Linguistics, 290–296. https://doi.org/10.18653/v1/2022.acl-short.32

[18] Siyuan Liao, Shangsong Liang, Zaiqiao Meng, and Qiang Zhang. 2021. Learning Dynamic Embeddings for Temporal Knowledge Graphs. In *WSDM '21*. ACM, 535–543. https://doi.org/10.1145/3437963.3441741

[19] Lihui Liu, Boxin Du, Jiejun Xu, Yinglong Xia, and Hanghang Tong. 2022. Joint Knowledge Graph Completion and Question Answering. In *KDD '22*. ACM, 1098–1108. https://doi.org/10.1145/3534678.3539289

[20] Johannes Messner, Ralph Abboud, and İsmail İlkan Ceylan. 2022. Temporal Knowledge Graph Completion Using Box Embeddings. In *AAAI 2022*. AAAI Press, 7779–7787. https://ojs.aaai.org/index.php/AAAI/article/view/20746

[21] Tomás Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *NIPS 2013*. 3111–3119. https://proceedings.neurips.cc/paper/2013/hash/9aa42b31882ec039965f3c4923ce901b-Abstract.html

[22] Sonia Farhana Nimmy, Omar K Hussain, Ripon K Chakrabortty, Farookh Khadeer Hussain, and Morteza Saberi. 2022. Explainability in supply chain operational risk management: A systematic literature review. *Knowledge-Based Systems* 235 (2022), 107587.

[23] Namyong Park, Fuchen Liu, Purvanshi Mehta, Dana Cristofor, Christos Faloutsos, and Yuxiao Dong. 2022. EvoKG: Jointly Modeling Event Time and Network Structure for Reasoning over Temporal Knowledge Graphs. In *WSDM '22*. ACM, 794–803. https://doi.org/10.1145/3488560.3498451

[24] Zile Qiao, Wei Ye, Tong Zhang, Tong Mo, Weiping Li, and Shikun Zhang. 2022. Exploiting Hybrid Semantics of Relation Paths for Multi-hop Question Answering over Knowledge Graphs. In *COLING 2022*. International Committee on Computational Linguistics, 1813–1822. https://aclanthology.org/2022.coling-1.156

[25] Abderahman Rejeb, John G Keogh, and Karim Rejeb. 2022. Big data in the food supply chain: a literature review. *J. of Data, Info. and Mang.* (2022), 1–15.

[26] Youssra Riahi, Tarik Saikouk, Angappa Gunasekaran, and Ismail Badraoui. 2021. Artificial intelligence applications in supply chain: A descriptive bibliometric analysis and future research directions. *Expert Systems with Applications* 173 (2021), 114702.

[27] Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S. Yu, and Tianyi Wu. 2011. PathSim: Meta Path-Based Top-K Similarity Search in Heterogeneous Information Networks. *Proc. VLDB Endow.* 4, 11 (2011), 992–1003. http://www.vldb.org/pvldb/vol4/p992-sun.pdf

[28] Zhu Sun, Jie Yang, Jie Zhang, Alessandro Bozzon, Long-Kai Huang, and Chi Xu. 2018. Recurrent knowledge graph embedding for effective recommendation. In *RecSys 2018*. ACM, 297–305. https://doi.org/10.1145/3240323.3240361

[29] Philip A Tominac, Weiqi Zhang, and Victor M Zavala. 2022. Spatio-temporal economic properties of multi-product supply chains. *Computers & Chemical Engineering* 159 (2022), 107666.

[30] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017*. 5998–6008. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[31] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2017. Graph Attention Networks. *CoRR* abs/1710.10903 (2017). arXiv:1710.10903 http://arxiv.org/abs/1710.10903

[32] Hongwei Wang, Fuzheng Zhang, Jialin Wang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. 2018. Ripplenet: Propagating user preferences on the knowledge graph for recommender systems. In *CIKM '18*. 417–426.

[33] Hongwei Wang, Fuzheng Zhang, Mengdi Zhang, Jure Leskovec, Miao Zhao, Wenjie Li, and Zhongyuan Wang. 2019. Knowledge-aware Graph Neural Networks with Label Smoothness Regularization for Recommender Systems. In *KDD 2019*. ACM, 968–977. https://doi.org/10.1145/3292500.3330836

[34] Hongwei Wang, Miao Zhao, Xing Xie, Wenjie Li, and Minyi Guo. 2019. Knowledge graph convolutional networks for recommender systems. In *WWW '19*. 3307–3313.

[35] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. 2019. KGAT: Knowledge Graph Attention Network for Recommendation. In *KDD 2019*. ACM, 950–958. https://doi.org/10.1145/3292500.3330989

[36] Xiang Wang, Tinglin Huang, Dingxian Wang, Yancheng Yuan, Zhenguang Liu, Xiangnan He, and Tat-Seng Chua. 2021. Learning Intents behind Interactions with Knowledge Graph for Recommendation. In *WWW '21*. ACM / IW3C2, 878–887. https://doi.org/10.1145/3442381.3450133

[37] Xiting Wang, Kunpeng Liu, Dongjie Wang, Le Wu, Yanjie Fu, and Xing Xie. 2022. Multi-level Recommendation Reasoning over Knowledge Graphs with Reinforcement Learning. In *WWW '22*. ACM, 2098–2108. https://doi.org/10.1145/3485447.3512083

[38] Xiang Wang, Dingxian Wang, Canran Xu, Xiangnan He, Yixin Cao, and Tat-Seng Chua. 2019. Explainable Reasoning over Knowledge Graphs for Recommendation. In *AAAI 2019*. AAAI Press, 5329–5336. https://doi.org/10.1609/aaai.v33i01.33015329

[39] Yaojing Wang, Yuan Yao, Feng Xu, Yada Zhu, and Hanghang Tong. 2022. Structure Meets Sequences: Predicting Network of Co-evolving Sequences. In *WSDM '22*. ACM, 1090–1098. https://doi.org/10.1145/3488560.3498411

[40] Shuo Yang, Zhiqiang Zhang, Jun Zhou, Yang Wang, Wang Sun, Xingyu Zhong, Yanming Fang, Quan Yu, and Yuan Qi. 2020. Financial Risk Analysis for SMEs with Graph-based Supply Chain Mining. In *IJCAI 2020*, Christian Bessiere (Ed.). ijcai.org, 4661–4667. https://doi.org/10.24963/ijcai.2020/643

[41] Jin Y Yen. 1971. Finding the k shortest loopless paths in a network. *management Science* 17, 11 (1971), 712–716.

[42] Bing Yu, Haoteng Yin, and Zhanxing Zhu. 2018. Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. In *IJCAI 2018*. ijcai.org, 3634–3640. https://doi.org/10.24963/ijcai.2018/505

[43] Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. 2016. Collaborative knowledge base embedding for recommender systems. In *KDD '16*. 353–362.

[44] Fuwei Zhang, Zhao Zhang, Xiang Ao, Fuzhen Zhuang, Yongjun Xu, and Qing He. 2022. Along the Time: Timeline-traced Embedding for Temporal Knowledge Graph Completion. In *CIKM '22*. 2529–2538.

[45] Jixin Zhao, Meng Ji, and Bo Feng. 2020. Smarter supply chain: A literature review and practices. *Journal of Data, Information and Management* 2, 2 (2020), 95–110.

# A APPENDIX

## A.1 Supply Chain Prediction Dataset

The detailed statistics for the supply chain prediction datasets across four real-world industries are given in Table 3. To be more specific, the SMEs supply chain prediction dataset[5] includes rich basic profiles and facts of supply chain relation of SMEs from different industries in China such as manufacturing, civil engineering, wholesaling and retailing. In the experiments, millions of ground-truth relationships are used for model training, validating and testing in each dataset, and the detailed settings are given in Table 4.

**Table 3: Detailed statistics of supply chain prediction datasets (sparsity is the proportion of negative samples, indicating the imbalanced and skewed level of each dataset).**

| Dataset | #SMEs | #static info. | #sparsity |
|---|---|---|---|
| Civil Engineering | 513,389 | 34 | 0.761 |
| Manufacturing | 443,594 | 34 | 0.732 |
| Wholesaling | 1,570,749 | 34 | 0.813 |
| Retailing | 1,077,171 | 34 | 0.814 |

**Table 4: Detailed settings of supply chain prediction datasets for model training, validating and testing.**

| Dataset | #train | #valid | #test |
|---|---|---|---|
| Civil Engineering | 817,358 | 102,169 | 102,169 |
| Manufacturing | 701,945 | 87,743 | 87,743 |
| Wholesaling | 9,270,052 | 1,158,756 | 1,158,756 |
| Retailing | 3,493,188 | 436,648 | 436,648 |

## A.2 Large-scale Financial STKG for SMEs

To well organize the facts of financial-related facts among millions of SMEs and billions of natural persons in China, we construct a large-scale financial spatial-temporal knowledge graph. However, it is a very challenging task as the activities of SMEs often take place offline in a highly frequent manner without being reported officially, which results in a seemingly isolated online relationship. Fortunately, as an online payment platform and e-Bank with a wide audience, Alipay[6] and MYBank[7] have recorded numerous transaction facts such as payments, transfers, debts, guarantees and invoicing from millions of SMEs, making it possible to identify the supply chain relationship among them and construct the financial knowledge graph effectively. To the best of our knowledge, it is the first attempt to construct an industrial-scale financial spatial-temporal knowledge graph for SMEs supply chain prediction.

Specifically, we have collected numerous real online and offline transaction records and electronic invoicing records over the years, as well as the publicly available registration information of SMEs. Four types of entities are defined in the STKG, namely enterprises (B), natural persons (C), industries (I) and locations (L). Processing the data with information filtering, entity and relation extraction, relation alignment and semantic disambiguation, we have constructed multiple relationships between enterprise and enterprise (B2B), enterprise and natural persons (B2C), enterprise and industries (B2I), enterprise and locations (B2L), natural persons and natural persons (C2C), natural persons and locations (C2L), industries and industries (I2I), and locations and locations (L2L). As shown in Fig. 8 and Table 5-6, we have collected more than 900 million entities and 2.8 billion triplets. To realize real-time storage and query capability, we store the constructed large-scale STKG in our self-developed graph database: TuGraph[8].

**Table 5: Detailed statistics of STKG.**

| Items | Statistics |
|---|---|
| #STKG slices | 12 |
| #STKG granularity | quarter |
| #STKG time span | 2020-2022 |
| #STKG triplets | 2,873,027,662 |
| #STKG entities | 909,984,654 |
| #STKG entity type | 4 |
| #STKG relations type | 20 |

**Table 6: Distribution of relations in STKG.**

| Relations | #types | #triplets |
|---|---|---|
| B2B | 11 | 130,402,087 |
| B2C | 1 | 177,321,331 |
| B2I | 1 | 173,301,542 |
| B2L | 1 | 173,711,683 |
| C2C | 2 | 1,481,984,738 |
| C2L | 1 | 736,294,304 |
| I2I | 2 | 11,523 |
| L2L | 1 | 454 |

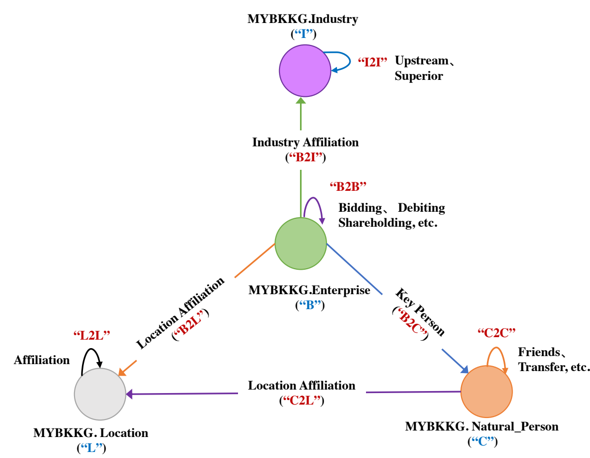


**Figure 8: The schema of STKG.**

[5]https://bksupplychain-assets.mybank.cn/
[6]https://www.alipay.com/
[7]https://www.mybank.cn/
[8]https://tech.antfin.com/products/TuGraph

## A.3 Baselines

We compared the proposed JRCL with some traditional predictors [3, 5], some recent KG-based recommendation methods [35, 36] and temporal knowledge graph reasoning models [15, 40]. The detailed settings for those in our SMEs supply chain prediction task are listed as follows:

- **LR** [3]: a classic logistic regression predictor. We used hand-engineered features extracted from the basic information of SMEs as the raw features.
- **GBDT** [5]: also a classic tree-based ensemble learning predictor. The hand-engineered features of the model input were the same as the LR.
- **KGAT** [35]: an embedding-based link prediction model using static knowledge graph. It can generate enterprise representations with an attentive neighborhood information aggregation module. Specifically, for each enterprise, we concatenated the output of the $l$-th layer and the embedding of its basic information into a single vector, respectively, as the input of the final predictor. In addition, we set the latest KG slice in STKG as the static knowledge graph.
- **KGIN** [36]: an explainable KG-based link prediction model by integrating relational information from multi-hop paths in KG. For the prediction layer, both the representations of enterprise pairs at different layers and the embeddings of basic information are summed up as the final representations. The KG used is also the latest slice in STKG.
- **TiRGN** [15]: a recent representation learning model for temporal knowledge graph reasoning, which simultaneously considers the sequential, repetitive and cyclical patterns of historical facts in knowledge graphs. We follow its settings in the relation prediction tasks.
- **ST-GNN** [40]: a spatial-temporal aware graph neural network model for SMEs supply chain relationship mining. It employed a plain spatial-aware aggregator and temporal-aware aggregator for aggregating both neighbor and relational information in STKG.

## A.4 Parameters

There were some training parameters in MSCL, *i.e.*, learning rate $lr$ and batch size $B$. In addition, there were also some hyperparameters in the process of multi-view relation sequences mining and relational co-evolution *i.e.*, the size of the selected sequence $K$ and the embedding size $d$. With both the efficiency and performance taken into account, the settings are given in Table 7. Moreover, for the hyperparameters in MRS layer *i.e.*, the embedding size $d_s$ for sub-space networks and the number of sub-space $N_s$, the settings were: $d_s = 16$ and $N_s = 4$. Note that, in order to guarantee the optimal parameters in experiments, we conduct grid searches and set the optimal hyperparameters for both our model and other competitors, respectively.

## A.5 Data Protection Statement

We list the data protection statement as follows:

- The data used in this research does not involve any **P**ersonal **I**dentifiable **I**nformation (PII).
- The data used in this research were all processed by data abstraction and data encryption, and the researchers were unable to restore the original data.
- Sufficient data protection was carried out during the experimental process to prevent data leakage and the data was destroyed after the experiments were finished.
- The data is only used for academic research and sampled from the original data, and therefore it does not represent any real business situation in MYBank, Ant Group.

**Table 7: Detailed hyperparameter settings for JRCL.**

| Dataset | $K$ | $d$ | $lr$ | $B$ |
|---|---|---|---|---|
| Civil Engineering | 20 | 64 | 0.0001 | 256 |
| Manufacturing | 20 | 64 | 0.0005 | 256 |
| Wholesaling | 20 | 32 | 0.0005 | 512 |
| Retailing | 10 | 32 | 0.0001 | 512 |