

# A Knowledge Graph Perspective on Supply Chain Resilience

Yushan Liu<sup>1,4,\*</sup>, Bailan He<sup>1,4</sup>, Marcel Hildebrandt<sup>1</sup>, Maximilian Buchner<sup>1</sup>, Daniela Inzko<sup>1</sup>, Roger Wernert<sup>3</sup>, Emanuel Weigel<sup>2</sup>, Dagmar Beyer<sup>1</sup>, Martin Berbalk<sup>1</sup> and Volker Tresp<sup>1,4</sup>

<sup>1</sup>Siemens AG, Otto-Hahn-Ring 6, 81739 Munich, Germany

<sup>2</sup>Siemens AG, Östliche Rheinbrückenstraße 50, 76187 Karlsruhe, Germany

<sup>3</sup>Siemens Schweiz AG, Theilerstraße 1a, 6300 Zug, Switzerland

<sup>4</sup>Ludwig-Maximilians-Universität München, Geschwister-Scholl-Platz 1, 80539 Munich, Germany

## Abstract

Global crises and regulatory developments require increased supply chain transparency and resilience. Companies do not only need to react to a dynamic environment but have to act proactively and implement measures to prevent production delays and reduce risks in the supply chains. However, information about supply chains, especially at the deeper levels, is often intransparent and incomplete, making it difficult to obtain precise predictions about prospective risks. By connecting different data sources, we model the supply network as a knowledge graph and achieve transparency up to tier-3 suppliers. To predict missing information in the graph, we apply state-of-the-art knowledge graph completion methods and attain a mean reciprocal rank of 0.4377 with the best model. Further, we apply graph analysis algorithms to identify critical entities in the supply network, supporting supply chain managers in automated risk identification.

## Keywords

Supply Chain Resilience, Knowledge Graphs, Machine Learning, Graph Analytics

## 1. Introduction

Global crises such as pandemics, natural disasters, and economic events as well as political and regulatory developments lead to increasing requirements regarding supply chain transparency and resilience. To ensure smooth procurement and production processes, it is essential for companies to react timely and flexibly to dynamic conditions and incidents to prevent production delays and bottlenecks within the supply network.

Usually, only direct (tier-1) suppliers of a company are tracked in supply chain management tools. The knowledge of sub-suppliers is often limited and disregarded for decision making. In a survey, almost 80% of the companies cannot even name the number of their tier- $n$  ( $n \geq 2$ ) suppliers [1], let alone their names and locations. Intransparent supply chains make it highly challenging to achieve precise forecasts and react in the best way in case of disruptions.

---

*Second International Workshop on Linked Data-driven Resilience Research (D2R2'23) co-located with ESWC 2023, May 28th, 2023, Hersonissos, Greece*

\*Corresponding author.

✉ yushan.liu@siemens.com (Y. Liu)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

Besides the intransparency of supply chains, another challenge is posed by the decentralized storage of relevant data and their incompleteness. The data come from different sources and are stored in various formats and locations. The disconnectedness makes it difficult to get a good overview of the situation and available information. Some information is also generally hard to retrieve, e. g., the exact production location of a material. Even if the supplier that delivers the material is known, the exact production site is often unknown.

Supply chain management involves monitoring supply chains to ensure their operability. Due to the inherent domain complexity and the high volume of data, significant blind spots at deeper levels of the supply chains remain, which matters because many of today’s most pressing supply shortages (e. g., in the semiconductor industry) happen at these deeper tiers. Therefore, possible risks in the supply chains need to be identified early, i. e., constellations in the supply chains that lack the resistance to withstand disruptive events. For example, constellations can be critical if many suppliers are located in the same region, multiple tier-1 suppliers buy from the same subsupplier, only one supplier is related to a specific business scope, etc. After identifying possible criticalities, strategic decisions and mitigation measures can be derived within the organization. Risk identification is often based on domain knowledge and manual efforts. 75% of the companies in a survey see a need for improvement with respect to risk identification methods, where the potential of machine learning approaches is valued highly [1].

In this paper, we aim at increasing supply chain resilience, based on data from Siemens, by addressing the challenges mentioned above in the following ways:

- Supply chain intransparency and data disconnectedness: We collect and connect supply chain-related data from different sources and create a knowledge graph, which contains information from Siemens suppliers up to tier 3.
- Data incompleteness: We apply state-of-the-art knowledge graph completion methods for link prediction in the knowledge graph to predict missing information.
- Identification of criticalities: We use graph analysis algorithms to identify critical entities in the supply network, where we focus on centrality measures to derive an importance score for each supplier.

The remainder of this paper is organized as follows. Section 2 outlines related work, and Section 3 describes the supply chain knowledge graph. In Section 4, we apply knowledge graph completion methods to the data, while in Section 5, we use graph analytics to find criticalities in the supply chains. The conclusion and further research directions follow in Section 6.

## 2. Related Work

The application of machine learning for supply chain management is becoming an increasingly active field of research [2]. While many supervised machine learning methods (e.g., decision trees, support vector machines, and neural networks) were successfully applied to tasks related to supply chain design, planning, and execution [2, 3], not many works exist in the area of knowledge graphs and graph machine learning.

In 2018, Brintrup et al. [4] published the first work to apply link prediction to a supply network. They modeled the supply network as a homogeneous graph (i. e., containing one

relation type) with handcrafted embeddings and defined a binary classification task to predict new links in the graph. A follow-up work used graph neural networks to predict the supplier relationship between companies [5]. Gopal and Chang [6] also used graph neural networks to predict new supplier relationships, where they included external information about companies, e. g., industry classification and revenue segmentation, as features. Lu and Chen [7] discovered potential partnerships between companies based on graph projections and connectivity patterns. Aziz et al. [8] represented supply networks as heterogeneous graphs (i. e., containing several relation types, also commonly referred to as knowledge graphs) and applied a relational graph convolutional network for link prediction.

### 3. Knowledge Graph Dataset

The supply chain knowledge graph is constructed from both Siemens-internal and external sources to reflect both internal knowledge such as tier-1 suppliers, business scopes, and Siemens parts and external knowledge such as public data about smelters and substances. The information about tier-2 and tier-3 suppliers of Siemens is obtained mainly from public customs data, and a small part is obtained from private customs data and public media. There are in total 16,910 tier-1, 43,759 tier-2, and 49,775 tier-3 suppliers of Siemens, where the suppliers at different tier levels are not mutually exclusive. The graph is modeled via the graph data platform Neo4j.

We define a knowledge graph (KG) as a collection of triples  $\mathcal{G} \subset \mathcal{E} \times \mathcal{R} \times \mathcal{E}$ , where  $\mathcal{E}$  denotes the set of entities and  $\mathcal{R}$  the set of relation types. Elements in  $\mathcal{E}$  correspond to supply chain-related entities, e. g., suppliers, smelters, and components, and are represented as nodes in the graph. Every entity has a unique entity type, which is defined by the mapping  $t : \mathcal{E} \rightarrow \mathcal{T}$ , where  $\mathcal{T}$  stands for the set of entity types. The entities are connected via relation types specified in  $\mathcal{R}$ , represented as directed edges in the graph. All entity and relation types and corresponding

**Table 1**

Entity and relation type statistics. In the graph, there are 8 entity types, where most nodes are suppliers, and 11 relation types, where most edges are from the type *supplies\_to*.

Entity type	Nodes	Relation type	Edges
Supplier	61,234	supplies_to	138,197
Manufacturer Part	1,650	related_to	59,894
Siemens Part	1,295	belongs_to	56,663
Smelter	340	located_in	30,107
Substance	321	includes	10,088
Component	233	produces	7,831
Country	172	produced_in	4,381
Business Scope	32	same_as	1,847
		manufactured_by	1,564
		contains	764
		refines	340
Total	65,277	Total	311,676



**Figure 1:** Knowledge graph schema. There are 8 entity types and 11 relation types.

numbers of nodes and edges are listed in Table 1.

Each relation type  $r \in \mathcal{R}$  connects entities from a fixed set of source entity types  $\mathcal{T}_{source}(r)$  to a fixed set of target entity types  $\mathcal{T}_{target}(r)$ . For example,  $\mathcal{T}_{source}(\text{supplies\_to}) = \{\text{Supplier}, \text{Smelter}\}$  and  $\mathcal{T}_{target}(\text{supplies\_to}) = \{\text{Supplier}\}$ . The schema of the graph depicts the possible connections between the entity types and is shown in Figure 1.

A fact from the graph is represented by a triple  $(\text{subject}, \text{predicate}, \text{object}) \in \mathcal{G}$ , where the subject and object are entities and the predicate is the relation type that connects them, directed at the object. Triples in the graph are assumed to be true facts, while the truth value of non-existing triples could either be wrong or unknown (since the data are highly incomplete).

## 4. Knowledge Graph Completion

### 4.1. Object prediction task

Many KGs suffer from incompleteness, so a common reasoning task in graph machine learning is KG completion or link prediction. We formulate the link prediction problem as an object prediction task. Given a query of the form  $(\text{subject}, \text{predicate}, ?)$ , the goal is to predict a ranked list of entity candidates that are most likely the correct object of the query.

To measure the quality of the predictions, the mean reciprocal rank (MRR) and  $\text{hits}@k$  for  $k \in \mathbb{N}$  are standard metrics used for link prediction on KGs. For a rank  $x \in \mathbb{N}$ , i. e., the position in the ranked list of entity candidates, the reciprocal rank is defined as  $\frac{1}{x}$ , and the MRR is the average of all reciprocal ranks of the correct query objects over all queries. The metric

$\text{hits}@k$  represents the proportion of queries for which the correct object appears under the top  $k$  candidates.

## 4.2. Knowledge graph completion methods

There exists a variety of methods for KG completion [9]. In this paper, we focus on graph representation learning, where the underlying idea is to learn low-dimensional embeddings (i. e., vectors or matrices) for the entities and relation types in the graph that capture their semantic meanings. Based on these embeddings, a score can be calculated for each entity, indicating its likelihood to be the correct object of a query.

We apply the following traditional and state-of-the-art methods to the supply chain KG:

- RESCAL [10] was the first method to be published for learning KG embeddings. It models the KG as a three-way tensor and performs tensor factorization for relational learning tasks such as link prediction.
- ComplEx [11] was the first KG embedding method that learns embeddings in the complex vector space. It is based on tensor factorization and the Hermitian dot product.
- TuckER [12] is a tensor factorization method based on the Tucker decomposition. It can be seen as a generalized version of RESCAL and ComplEx.
- TransE [13] was the first translational method, which models the relations between two entities as translations in the vector space.
- RotatE [14] is a roto-translational method, which models the relations between two entities as rotations in the complex vector space.
- ConvE [15] was the first method that uses convolutional neural networks to model the interactions between entities.
- RGCN [16] consists of a relational graph convolutional network for encoding the entities and a tensor factorization method for scoring.
- CompGCN [17] incorporates composition operators to learn joint embeddings for entities and relation types. It is a generalized version of RGCN.

## 4.3. Experimental setup

For all KG completion methods, we use the implementations provided by the Python library PyKEEN [18]. We split the graph into training, validation, and test dataset, where we operate under the transductive setting, i. e., all entities and relation types from the validation and test set are also included in the training set. The training set consists of 65,277 nodes and 249,340 triples, while both the validation and test set have 31,168 triples. The number of nodes for the validation and test set is 22,212 and 22,213, respectively. All three datasets include all entity and relation types. We use the optimizer Adam and optimize the margin ranking loss with a margin of 1, where one negative triple is sampled for each training triple. We tune the hyperparameters embedding size in the range  $\{16, 32, 64, 256, 512, 1024\}$  and learning rate in the range  $\{0.0001, 0.001, 0.01\}$ . The training of the model is stopped early if there is no improvement regarding the metric  $\text{hits}@10$  on three subsequent evaluations on the validation set, where the evaluation takes place every 10 epochs.

**Table 2**

Results on the test dataset for the object prediction task. The best results are displayed in bold.

Method	MRR	Hits@1	Hits@3	Hits@10
RESICAL	0.1476	0.0684	0.1809	0.2772
ComplEx	0.2535	0.1793	0.2850	0.3949
TuckER	0.1738	0.0749	0.1878	0.4033
TransE	0.1595	0.0873	0.1733	0.3164
RotatE	<b>0.4377</b>	<b>0.3686</b>	<b>0.4733</b>	<b>0.5627</b>
ConvE	0.2289	0.1549	0.2438	0.3875
RGCN	0.2911	0.1784	0.3379	0.5195
CompGCN	0.2223	0.1271	0.2486	0.4229

#### 4.4. Results

Table 2 displays the results of the best models for the selected KG completion methods. All models were able to learn useful embeddings from the training set, while RotatE performed best with respect to all metrics. For more than 36% of the test queries, RotatE predicts the correct object as the highest-ranked entity in the candidates list, and for almost half of the queries, RotatE is able to predict the correct object under the top three entities. Out of the three tensor factorization methods (RESICAL, ComplEx, and TuckER), ComplEx, which learns embeddings in the complex vector space, performs best. Out of the three neural network-based methods (ConvE, RGCN, and CompGCN), RGCN performs best, while ConvE and CompGCN have similar performance. Before materializing the results in the graph, domain experts should check the predicted triples for plausibility.

In Figure 2, the results of the best models are shown for each relation type. For every model, the performance with respect to the MRR is colored from best (green) to worst (red). RotatE performs best for all relation types except for *locate\_in* (where all neural network-based models have higher MRR) and *refines* (where ComplEx and RGCN are better). The models show varying degrees of performance for the different relation types, where most models tend to have good results on *related\_to*, *includes*, and *belongs\_to* and bad performance on *same\_as*, *contains*, and *refines*. Especially *same\_as* is worst for all models. The reason lies in the graph schema and structure, where the *same\_as* relation type connects manufacturer parts and Siemens parts, which do not have any other connections. Without further information, it is difficult to predict the correct Siemens part as query object.

## 5. Graph Analytics

### 5.1. Graph analysis algorithms

Supply chain managers mainly decide based on tier-1 supplier data whether there are risks in the supply chains, and they are usually directly informed by tier-1 suppliers if there are already existing problems. If additional data about tier-*n* suppliers are available, more detailed and precise decisions can be made, and mitigation measures in an earlier phase of the supply chain

MRR	RESCAL	ComplEx	TuckER	TransE	RotatE	ConvE	RGCN	CompGCN	
supplies_to	0.1422	0.2661	0.0539	0.0740	0.3499	0.1574	0.2116	0.1718	Best          Worst
related_to	0.2900	0.2756	0.4317	0.3539	0.7256	0.4876	0.5025	0.3291	
belongs_to	0.0039	0.3411	0.3428	0.6671	0.6675	0.0261	0.0548	0.4000	
located_in	0.0003	0.0653	0.1500	0.0909	0.1526	0.1726	0.2241	0.1973	
includes	0.0004	0.5259	0.4176	0.5084	0.8682	0.0734	0.4607	0.4390	
produces	0.0006	0.2923	0.0341	0.0984	0.3975	0.0113	0.1845	0.2122	
produced_in	0.0003	0.1939	0.2907	0.1131	0.3539	0.0813	0.2043	0.1668	
same_as	0.0002	0.0022	0.0011	0.0003	0.0490	0.0001	0.0049	0.0020	
manufactured_by	0.0005	0.5646	0.3831	0.1420	0.9564	0.0707	0.3648	0.1791	
contains	0.0005	0.0212	0.0180	0.0024	0.2106	0.0015	0.1143	0.1362	
refines	0.0014	0.4593	0.0152	0.0337	0.1501	0.0166	0.4567	0.0402	

**Figure 2:** Results of the best models for each relation type. For each model, the performance with respect to the MRR is colored from best (green) to worst (red).

can be enabled. However, this kind of approach is mainly reactive, and manual decision making is not scalable to complex supply networks. Therefore, we propose to use graph analytics to support supply chain managers by automatically identifying critical suppliers so that they can be monitored more closely and mitigation strategies can be derived together.

For the graph analysis, we use the Neo4j Graph Data Science library and concentrate on the subgraph consisting of supplier entities and the relation type *supplies\_to*. We calculate the following centrality and community detection metrics, which serve as a basis for deriving the importance or criticality of a supplier:

- The **degree centrality** measures the number of incoming and outgoing edges for each node. The number of incoming edges represents the number of suppliers and the number of outgoing edges the number of customers for each company. Companies with high in- or out-degree might be affected by disruptive events more often.
- The **betweenness centrality** for each node is based on the number of shortest paths between all node pairs that the node lies on. A company with high betweenness connects many companies and is more likely to cause a bottleneck.
- The **closeness centrality** measures the average length of the shortest paths between a node and all other nodes. A company with high closeness is a central customer for many suppliers.
- The **triangle count** is a community detection measure that calculates the number of adjacent triangles of a node. A company with a high triangle count is part of an interconnected supply network.

To make the suppliers better comparable, we normalize the metrics in-degree, out-degree, betweenness, closeness, and triangle count to be between 0 and 10 and sum them up to obtain an aggregated importance score for each supplier in the graph.

## 5.2. Results

When comparing the aggregated scores of the suppliers, Siemens is obviously the center of the supply network and has an aggregated score of 37.25. The next supplier has a score of only 16.66, and there are only 3 suppliers with a score above 15. There are in total 988 suppliers with



**Table 3**

Correlation matrix. There is a high correlation between in-degree, betweenness, and triangle count.

Correlation	in-degree	out-degree	betweenness	closeness	triangle count
in-degree	1.0000				
out-degree	0.1969	1.0000			
betweenness	0.8816	0.3928	1.0000		
closeness	0.0686	0.2792	0.0859	1.0000	
triangle count	0.9809	0.2048	0.8774	0.0542	1.0000

a score above 10, which might be critical entities in the supply network and should be examined by domain experts. Since an aggregated score loses information, the suppliers with the highest values for each metric should be analyzed in more detail. Table 3 shows the correlation matrix of the five metrics. There is a high correlation between in-degree, betweenness, and triangle count. That means, companies with many suppliers often lie on a large number of shortest paths (supply chains) and are part of a highly interconnected supply network.

Figure 3 illustrates a possible way to visualize the results in order to identify critical paths in the supply network. The subgraph contains yellow and red nodes, which represent suppliers, and purple nodes, which represent business scopes. The red suppliers have aggregated scores above 10 and might be more critical than the yellow suppliers. Any supply chain containing a critical supplier might have a higher risk. The size of the purple nodes corresponds to the number of suppliers related to the corresponding business scopes. The orange edges represent the edge type *supplies\_to* and the blue edges the edge type *related\_to*. If there are business scopes to which only one supplier is related, then the supplier might be critical since a delay of this supplier would not be compensated easily by another supplier within the same business scope. In the figure, three such business scopes can be identified (purple nodes with blue circles), where two of the corresponding suppliers also have a critical score.

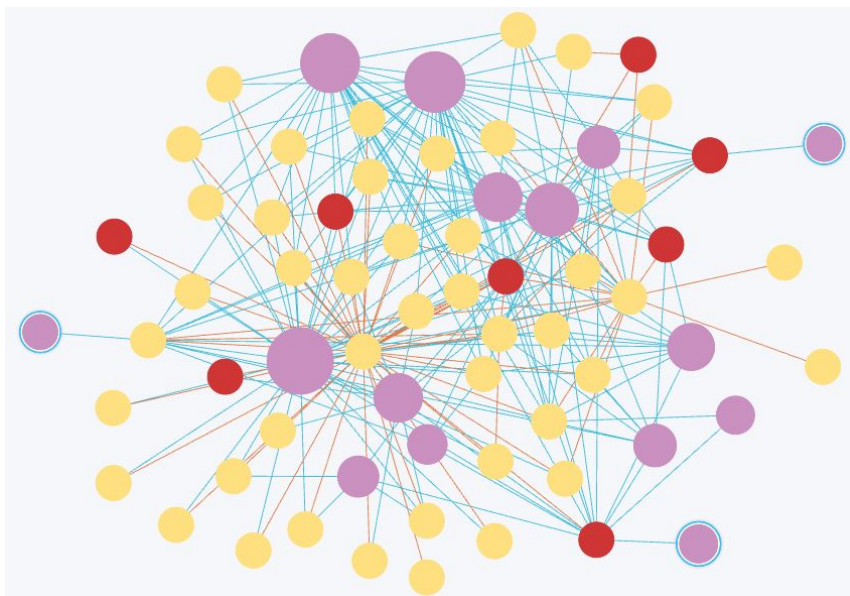
## 6. Conclusion and Further Research Directions

Challenges for supply chain management include supply chain intransparency, data disconnect-edness and incompleteness, and the scalable identification of criticalities in the supply network. We addressed these challenges by modeling supply chain-related information as a knowledge graph. We used state-of-the-art knowledge graph completion methods to predict missing links and applied graph analysis algorithms to compute importance scores for all suppliers. Based on the importance scores and the graph structure, critical supply chains could be identified, which is an essential step towards more resilient supply networks.

For further research, we propose the following possible directions:

- Integration of node and edge properties: In this paper, we only focused on the graph structure for link prediction and graph analysis. For some entity and relation types, however, there exist properties that could be helpful for prediction. For example, the





**Figure 3:** Visualization of a subgraph. The yellow nodes and red nodes represent suppliers, where the red suppliers have aggregated scores above 10. Business scopes are represented as purple nodes. The size of the purple nodes correlates to the number of suppliers related to the corresponding business scopes. The orange edges represent the edge type *supplies\_to* and the blue edges the edge type *related\_to*.

relation type *produced\_in* between a substance and a country has the property Herfindahl-Hirschmann-Index, a measure of market concentration. For the prediction of the relation type *located\_in*, the company name could be a good indicator. These properties could be integrated when learning embeddings or calculating importance scores.

- Node regression or classification: Besides link prediction, node regression or classification are common tasks on knowledge graphs. Given, e.g., risk scores or categories for a subset of companies, one could learn risk scores or categories for companies that are missing this information in the graph.
- Analysis of the complete graph: We conducted the graph analysis based on a subgraph containing the suppliers and the *supplies\_to* relation type. To calculate the importance scores, more information from the graph could be included. For example, a supplier that is located in a country with a high sustainability risk might also have a higher risk, or a supplier that manufactures many Siemens parts would be more critical for Siemens.

## Acknowledgments

This work has been supported by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) as part of the project CoyPu under grant number 01MK21007K.

## References

- [1] M. Brylowski, M. Schröder, W. Kersten, Machine Learning im Supply Chain Risk Management: Studie, Technical Report, Technische Universität Hamburg, 2021.
- [2] M. Brylowski, M. Schröder, S. Lodemann, W. Kersten, Machine learning in supply chain management: A scoping review, in: Proceedings of the Hamburg International Conference of Logistics, volume 31, 2021.
- [3] E. B. Tirkolaee, S. Sadeghi, F. M. Mooseloo, H. R. Vandchali, S. Aeini, Application of machine learning in supply chain management: A comprehensive overview of the main areas, in: Mathematical Problems in Engineering, volume 2021, 2021.
- [4] A. Brintrup, P. Wichmann, P. Woodall, D. McFarlane, E. Nicks, W. Krechel, Predicting hidden links in supply networks, in: Complexity, volume 2018, 2018.
- [5] E. E. Kosasih, A. Brintrup, A machine learning approach for predicting hidden links in supply chain with graph neural networks, in: International Journal of Production Research, volume 60, 2021.
- [6] A. Gopal, C. Chang, Discovering supply chain links with augmented intelligence, in: The Second ACM International Conference on AI in Finance, Workshop Natural Language Processing and Network Analysis in Financial Applications, 2021.
- [7] Z.-G. Lu, Q. Chen, Discovering potential partners via projection-based link prediction in the supply chain network, in: International Journal of Computational Intelligence Systems, volume 13, 2020.
- [8] A. Aziz, E. E. Kosasih, R.-R. Griffiths, A. Brintrup, Data considerations in graph representation learning for supply chain networks, in: The Thirty-Eighth International Conference on Machine Learning, Workshop Machine Learning for Data: Automated Creation, Privacy, Bias, 2021.
- [9] M. Wang, L. Qiu, X. Wang, A survey on knowledge graph embeddings for link prediction, in: Symmetry, volume 13, 2021.
- [10] M. Nickel, V. Tresp, H.-P. Kriegel, A three-way model for collective learning on multi-relational data, in: The Twenty-Eighth International Conference on Machine Learning, 2011.
- [11] T. Trouillon, J. Welbl, S. Riedel, É. Gaussier, G. Bouchard, Complex embeddings for simple link prediction, in: The Thirty-Third International Conference on Machine Learning, 2016.
- [12] I. Balažević, C. Allen, T. M. Hospedales, TuckER: Tensor factorization for knowledge graph completion, in: The 2019 Conference on Empirical Methods in Natural Language Processing and the Ninth International Joint Conference on Natural Language Processing, 2019.
- [13] A. Bordes, N. Usunier, A. Garcia-Durán, Translating embeddings for modeling multi-relational data, in: The Twenty-Seventh Conference on Neural Information Processing Systems, 2013.
- [14] Z. Sun, Z.-H. Deng, J.-Y. Nie, J. Tang, RotatE: Knowledge graph embedding by relational rotation in complex space, in: The Seventh International Conference on Learning Representations, 2019.
- [15] T. Dettmers, P. Minervini, P. Stenetorp, S. Riedel, Convolutional 2D knowledge graph embeddings, in: The Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

- [16] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. van den Berg, I. Titov, M. Welling, Modeling relational data with graph convolutional networks, in: The Fifteenth Extended Semantic Web Conference, 2018.
- [17] S. Vashishth, S. Sanyal, V. Nitin, P. Talukdar, Composition-based multi-relational graph convolutional networks, in: The Eighth International Conference on Learning Representations, 2020.
- [18] M. Ali, M. Berrendorf, C. T. Hoyt, L. Vermue, S. Sharifzadeh, V. Tresp, J. Lehmann, PyKEEN 1.0: A Python library for training and evaluating knowledge graph embeddings, in: Journal of Machine Learning Research, volume 22, 2021.