Comparing Spatial-Temporal Knowledge Graph on Spatial Downstream Tasks

Martin Böckling*
University of Mannheim
Mannheim, Germany
martin.boeckling@students.uni-mannheim.de

Heiko Paulheim
University of Mannheim, Data and Web Science Group
Mannheim, Germany
heiko@informatik.uni-mannheim.de

Sarah Detzler
Corporate State University of Mannheim
Mannheim, Germany
sarah.detzler@dhbw-mannheim.de

ABSTRACT

Knowledge graphs have become a universal data representation and integration mechanism. They recently gained interest in the spatial area. Spatial-Temporal Knowledge Graphs (STKGs) in particular have been created to integrate diverse sets of spatial data and model the relationships of spatial entities. Public knowledge graphs, such as KnowWhereGraph and WorldKG, have received a high traction in the domain of STKGs. In this paper, we compare three STKGs using downstream tasks within the Spatio-Temporal domain, and also discuss the underlying modeling decisions. We conduct an evaluation on a wildfire dataset and a housing dataset, comparing different embedding methodologies for the different knowledge graphs. We show that modeling paradigms in STKGs as well as algorithmic choices can have an impact on the downstream performance, and discuss challenges in both areas.

CCS CONCEPTS

• **Information systems** → **Spatial-temporal systems**; • **Applied computing** → *Earth and atmospheric sciences*; • **Computing methodologies** → *Machine learning*.

KEYWORDS

Spatial-Temportal Knowledge Graph, Knowledge Graph embeddings, Wildfire prediction, Price prediction

ACM Reference Format:
Martin Böckling, Heiko Paulheim, and Sarah Detzler. 2024. Comparing Spatial-Temporal Knowledge Graph on Spatial Downstream Tasks. In *The 32nd ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL '24), October 29-November 1, 2024, Atlanta, GA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3678717.3691321

1 INTRODUCTION

STKGs are powerful tools for analyzing complex relationships in spatial-temporal data, integrating time and space to better understand the evolution of entities and events. Their ability to capture intricate patterns makes them essential for applications like urban planning, environmental monitoring, navigation, and emergency response. With the growing volume of spatial-temporal data, effective methods for leveraging spatial-temporal data become increasingly important [9, 12].

This study evaluates the performance of STKGs in spatial-temporal predictions, highlighting the strengths and limitations of each. By exploring different representation paradigms, we provide insights into their real-world applicability and contribute to a better understanding of how STKGs can enhance spatial data modeling and prediction tasks [10].

The choice of representation in STKGs significantly impacts their effectiveness. This work examines KnowWhereGraph [12], WorldKG [9], and OSMh3KG [5], each using a different approach to model spatial-temporal data. Our comparative analysis identifies key factors influencing performance on spatial downstream tasks, offering guidance for future developments in the field.

Our research extends beyond evaluating the three STKGs. We aim to analyze the practical benefits and limitations of each, while also highlighting potential future enhancements in STKG methodologies. We provide the first comparison of different preexisting STKGs on spatial tasks using existing embedding methodologies, offering an overview of existing frameworks in spatial-temporal data analysis. We demonstrate that different Knowledge Graphs (KGs) and embedding models lead to varying performance in downstream tasks, sometimes not exceeding the baseline.

2 PRELIMINARIES

2.1 Problem Definition

In this paper, we analyze how the selected KGs perform overall on different prediction tasks. To provide an overview of our research objectives, we outline in this subsections our research hypotheses together with the associated research questions.

In order to measure the improvement of the KG involvement, we establish a baseline only containing the tabular dataset without KG information. We hypothesize that (H1) enriching tabular datasets with embeddings from KGs will improve predictions in spatial tasks, outperforming the defined baseline. We also expect that (H2) KGs that capture a complete set of spatial data will enhance prediction results in hybrid scenarios.

Additionally, we hypothesize that (H3) different embedding methods will yield varying performances on downstream tasks. By applying various embedding methods to different KGs, we aim to identify which structural patterns are most effective, allowing us to compare our findings with existing research.

The following sections of our paper will elaborate on the above hypotheses and will be examined against our research findings.

SIGSPATIAL '24, October 29-November 1, 2024, Atlanta, GA, USA

ACM ISBN 979-8-4007-1107-7/24/10
https://doi.org/10.1145/3678717.3691321

2.2 Knowledge Graph Embedding Approaches

To utilize KGs in downstream tasks, we compute vector representations, i.e., *embeddings* with a number of established embedding methodologies used in the area of KGs. Different research papers within the area have analyzed extensively the performance of KG embedding approaches on non-spatial KGs [8, 13]. Based on the papers we select overall three different embedding methodologies that allow the creation of embeddings for KG vertices: RDF2Vec, TransE and ComplEx.

RDF2Vec is a walk based algorithm, which combines the extracted random walks with a Word2Vec model in order to generate embeddings. In total two different hyperparameters are used for the RDF2Vec walk extraction: The variable d describes the maximum depth for random walks starting from a single vertex v. Furthermore, with the parameter h the maximum number of extracted walks per vertex is determined. For a specified depth $d \in \mathbb{N}$, graph walks P_v are extracted for a single vertex v using the breadth-first search algorithm. All elements within P_v are treated as single words within Word2Vec. Within our research, we use the CBOW model for training the embeddings. [14].

With TransE, a translational embedding methodology is used. TransE interprets the projection of entities into a vector space as translations in the vector space. The scoring function of TransE uses either the l1 or l2 normalization. The goal of the scoring function is to minimize the score for valid triples and to maximize the score for invalid triples which are not present in the KG. In order to achieve this, TransE uses a margin-based ranking loss function. [3]

The third embedding approach ComplEx allows the modeling of not only symmetric relations but also asymmetric relations. ComplEx makes use of the Hermitian Dot Product for modeling relationships within KGs. It has been shown that ComplEx scales well for larger KGs. [8, 18]

3 METHODS

3.1 Datasets

In our research, we compare the performance of STKGs using two datasets on spatial downstream tasks.

The first dataset focuses on wildfire prediction in California, covering monthly wildfire occurrences from 2014 to 2021. This dataset, based on a GitHub repository [4], includes weather, landscape (elevation, land use), and wildfire data. The original hexagonal grid structure was replaced with the h3 grid at level 7 for California. A binary label indicates whether a wildfire occurred (1) or not (0) for each timestamp and grid cell. The dataset is highly unbalanced, with 0.7313% wildfire cases in the training set and 2.699% in the test set. The spatial distribution of wildfires from 2005 to 2021 shows that some regions in California experienced no wildfires during this period.

The second dataset uses Airbnb data from Boston, comprising three separate datasets: the listing dataset (attributes of Airbnb sites), the calendar dataset (listing Identifier (ID), availability, and price), and the review dataset. For our analysis, we focus on the listing and calendar datasets to predict the price of an Airbnb rental. Since only unbooked places during the data crawl show a price, we drop records with *Null* values for price. The dataset contains 509,331 rows and 19 columns, covering the period from 2016-09-01 to 2017-09-30. The dataset includes price outliers, such as a listing offered at over 6,000$ per night in October 2016, which are retained to evaluate model performance on outliers [1]. Within our research we do not conduct a removal of those outliers in order to measure the overall model performance also on those outliers. A detailed overview of the data statistics of the used KGs can be found under the following GitHub repository page.

3.2 Evaluation Framework

In order to compare the performance of STKGs, we differentiate between two datasets: the BaseCase, consisting of classical spatial data in tabular form without any KG, and the HybridCase, which integrates embeddings from STKGs with the BaseCase dataset.

The BaseCase serves as a baseline for analyzing the performance of the enriched HybridCase datasets. We split the datasets into training and testing sets based on the temporal dimension T to evaluate prediction accuracy. The HybridCase datasets are enriched by joining the h3 grid ID and date with trained embeddings from the KGs. Even if from the start the KG does not use the h3 grid, we assign it to each individual node of the KG.

For classification task, we evaluate performance using F1-score and Area under the receiver operating characteristic curve (AUC) score. For the regression task, we use Mean Absolute Percentage Error (MAPE), and R^2. We employ the XGBoost algorithm for both prediction tasks, as it has proven to be robust in structured data predictions [16]. To assess the significance of our results, we calculate confidence intervals for the different performance metrics.

3.3 Temporal Alignment for Knowledge Graph embedding

In our research, we train over the temporal dimension T independent embedding models which result in a different vector space in order to prevent temporal leakage. In our case, we use the Orthogonal Procrustes alignment to align the different vector spaces for each temporal step t. Research papers have shown that the orthogonal Procrustes alignment have shown robust results in the task of entity alignment between different vector spaces [17]. For our research we therefore make use of the orthogonal Procrustes alignment for the vector alignment. In the following sections we present our results of our paper on the two datasets together with the constraint of hardware resources used in the experiments.

4 EXPERIMENTS AND RESULT

4.1 Experimental Settings

For the RDF2Vec embedding approach, we implemented our own solution to handle large-scale processing, using the C-based igraph library, which showed faster performance compared to other RDF2Vec implementations [7]. We set the walk depth d to 4 and generated 500 random walks per entity. Each generated embedding contains in total 100 dimensions.

For TransE and ComplEx embeddings, we used the pykeen library, loading the KG and splitting it into 80% train, 10% validation, and 10% test datasets. We trained both models for up to 100 epochs, with validation checks every 10 epochs. If validation results worsened, training was stopped early [2]. For the XGBoost training, we used default parameters on the downstream datasets [6]. For classification tasks, we applied random oversampling to balance the training data and one-hot encoding for categorical values. Similarly to RDf2Vec for both models we generated embedding containing in total 100 dimensions. Our computational setup included a virtual machine with 64 vCPUs, 512 GB RAM, and an Nvidia RTX A6000 GPU with 47 GB VRAM. We set a 10-day limit for experiments; if an embedding generation or model training exceeded this, results were marked with ▲. If the VM ran out of GPU vRAM or RAM, results were marked with †.

4.2 Result Overview

We present the test results for both datasets in tables 1 and 2. As outlined in subsection 2.1, a temporal split is applied between the train and test datasets to evaluate model performance. The wildfire test dataset starts from 2020-01-01, while the Airbnb test dataset starts from 2017-04-01.

Table 1: Results for the Wildfire dataset for different KGs and embedding approaches

KG	Embedding approach	F1	AUC
/	/	0.1760 ± 0.0038	0.8519 ± 0.0012
WorldKG	RDF2Vec	0.1183 ± 0.0071	0.8269 ± 0.0027
WorldKG	TransE	0.1328 ± 0.0074	0.8363 ± 0.0025
WorldKG	ComplEx	▲	▲
KWG	RDF2Vec	0.0287 ± 0.0043	0.7849 ± 0.0025
KWG	TransE	0.1596 ± 0.0043	0.8305 ± 0.0015
KWG	ComplEx	0.1787 ± 0.0044	0.8260 ± 0.0016
OSMh3KG	RDF2Vec	**0.3176** ± 0.0045	**0.8565** ± 0.0013
OSMh3KG	TransE	†	†
OSMh3KG	ComplEx	†	†

Table 2: Results for the Airbnb dataset for the different KGs and embedding approaches.

KG	Embedding approach	MAPE	R^2
/	/	0.1774 ± 0.0012	0.6463 ± 0.0012
WorldKG	RDF2Vec	0.1627 ± 0.0002	0.6617 ± 0.0002
WorldKG	TransE	0.1605 ± 0.0001	**0.6636** ± 0.0001
WorldKG	ComplEx	0.1605 ± 0.0002	**0.6636** ± 0.0020
KWG	RDF2Vec	0.2142 ± 0.0071	0.5954 ± 0.0067
KWG	TransE	0.1587 ± 0.0011	0.6524 ± 0.0012
KWG	ComplEx	0.1502 ± 0.0063	0.6274 ± 0.0060
OSMh3KG	RDF2Vec	**0.1434** ± 0.0071	0.6384 ± 0.0012
OSMh3KG	TransE	0.1941 ± 0.0011	0.6339 ± 0.0012
OSMh3KG	ComplEx	0.2045 ± 0.0011	0.6314 ± 0.0013

Based on the results provided in table 1 and 2, we determine several findings based on our initial research hypotheses:

With RDF2Vec, OSMh3KG significantly outperformed the baseline and all other KGs and embedding methods on the Airbnb dataset. However, when using TransE and ComplEx, OSMh3KG performed worse than the baseline. Similarly, in the wildfire dataset, RDF2Vec with OSMh3KG outperformed other KG or embedding combinations, though we couldn't evaluate ComplEx and TransE due to hardware constraints during our research.

For WorldKG, ComplEx outperformed TransE and RDF2Vec on both datasets. While WorldKG surpassed the baseline on the Airbnb dataset, it failed to outperform the baseline on the wildfire dataset. Due to the 10-day processing limit, we couldn't obtain results for the wildfire dataset and the ComplEx dataset.

KnowWhereGraph, which uses thematic datasets rather than large data sources like OpenStreetMap (OSM), outperformed the baseline on the Airbnb dataset with both ComplEx and TransE. ComplEx performed better than TransE for MAPE, though not for R^2. On the wildfire dataset, KnowWhereGraph failed to outperform the baseline with RDF2Vec and TransE, although it achieved a higher F1-score with ComplEx, but did not surpass the baseline in AUC. In subsection 4.3, we will reflect on the different results and provide a comparison to the existing research to determine reasons for the difference in the performances seen in our research. Furthermore, we will reflect on the different hypotheses presented in subsection 2.1.

4.3 Discussion

The datasets selected for this study inherently introduce a bias in the Machine Learning (ML) results, as not all potential influencing factors, such as long-term drought conditions or socioeconomic data, are included. This omission means that the findings from the modeling phase should be critically reviewed, acknowledging that not all relevant factors for the spatial events are captured in the data. Additionally, our choice of wildfire and Airbnb datasets limits the scope of the research to specific spatial regions and issues, meaning the findings are closely tied to these particular datasets [11].

As noted in previous studies, TransE struggles with learning complex relations, particularly when entities share multiple relations, such as spatial intersections. This limitation is evident in the OSMh3KG results, where such relations are prevalent [8]. Interestingly, while OSMh3KG performed well with RDF2Vec embeddings, WorldKG and KnowWhereGraph did not. OSMh3KG's advantage lies in its use of grid cells to model spatial relationships via DE-9IM, capturing spatial patterns within and around each grid cell. This structure is not present in WorldKG or fully utilized in KnowWhereGraph, which impacts their performance confirming partially our second hypothesis.

Contrary to our first hypothesis, the choice of KGs and embedding methods significantly affects the likelihood of outperforming the baseline. While other studies have shown that STKGs typically enhance model performance, our results varied. For example, OSMh3KG with ComplEx and TransE underperformed compared to the baseline in the Airbnb dataset. Similarly, RDF2Vec with KnowWhereGraph also failed to surpass the baseline on the same dataset. Overall, 3 out of 9 KG and embedding combinations for the Airbnb dataset did not exceed baseline performance using MAPE, and for the wildfire dataset, only 2 out of 6 combinations showed better results using F1 and AUC scores.

We also observed that the overall algorithmic choice for KG embedding methods has a significant impact on prediction performance, particularly for KnowWhereGraph and OSMh3KG, where the differences between models were more pronounced. This observation confirms our third hypothesis. In contrast, WorldKG showed minimal deviation in results across different embedding methods, suggesting that its lack of a spatial grid in the KG might contribute to this consistency. Both KnowWhereGraph and OSMh3KG incorporate spatial grid relations between grid cells, which seems to influence how well they perform with specific embedding approaches. This aligns with previous studies that have noted similar effects [8, 15]. However, our findings differ from those where ComplEx performed consistently well across various KGs which is a result that was not replicated in our experiments with OSMh3KG [15].

Additionally, we experimented with increasing the maximum walk distance d for RDF2Vec using OSMh3KG, from 4 to 9. This adjustment led to significantly worse results in both datasets. For the wildfire dataset, the F1-Score dropped to 0.0891 (from 0.3176) and the AUC score to 0.8256. Similarly, for the Airbnb dataset, MAPE increased to 0.2123 (from 0.1434). The analysis of the random walks showed that most walks captured grid-based relations, which did not positively contribute to the prediction tasks in our experiments.

5 CONCLUSION AND OUTLOOK

In our paper, we compared the performance of several preexisting KGs across two different datasets. We found that, in some cases, KGs-enriched predictions significantly outperformed baseline datasets. This suggests that our research, is in line with our first hypothesis, supports the idea that spatial data combined with KGs can have a beneficial impact on prediction tasks. Moreover, our study demonstrated that existing KGs can be effectively reused and aligned with specific prediction needs, eliminating the necessity to build custom KGs. Both WorldKG and OSMh3KG, which leverage large data foundations like OSM, provide a more accurate spatial representation. We also observed that the structure of a KG influences the effectiveness of different embedding approaches, emphasizing the importance of selecting the appropriate embedding algorithm based on the KG used.

For future research, expanding the comparison to include other embedding approaches could be valuable. Due to the scale of our KGs, deep learning-based methods like CompGCN were only applicable to a subset of experimental settings. Implementing models on our own could enable scaling to multiple GPUs. Additionally, most existing embedding approaches treat geometries as string literals, but specialized methods that interpret geometries directly could improve embedding quality.

ACKNOWLEDGMENTS

Map data copyrighted OpenStreetMap contributors and available from https://www.openstreetmap.org.

REFERENCES

[1] Airbnb Inc. 2020. Boston Airbnb Open Data. https://kaggle.com/datasets/airbnb/boston/data

[2] Mehdi Ali, Max Berrendorf, Charles Tapley Hoyt, Laurent Vermue, Sahand Sharifzadeh, Volker Tresp, and Jens Lehmann. 2021. PyKEEN 1.0: A Python Library for Training and Evaluating Knowledge Graph Embeddings. *Journal of Machine Learning Research* 22, 82 (2021), 1–6. http://jmlr.org/papers/v22/20-825.html

[3] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating Embeddings for Modeling Multi-relational Data. In *Advances in Neural Information Processing Systems*, C. J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Eds.), Vol. 26. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2013/file/1cecc7a77928ca8133fa24680a88d2f9-Paper.pdf

[4] Martin Böckling. 2023. MartinBoeckling/WildfirePredictionSTKG: Initial release. https://doi.org/10.5281/ZENODO.8383318

[5] Martin Böckling, Heiko Paulheim, and Sarah Detzler. 2024. A Planet Scale Spatial-Temporal Knowledge Graph Based On OpenStreetMap And H3 Grid. arXiv:2405.15375 [cs].

[6] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, San Francisco California USA, 785–794. https://doi.org/10.1145/2939672.2939785

[7] Gabor Csardi and Tamas Nepusz. 2005. The Igraph Software Package for Complex Network Research. *InterJournal* Complex Systems (Nov. 2005), 1695.

[8] Yuanfei Dai, Shiping Wang, Neal N. Xiong, and Wenzhong Guo. 2020. A Survey on Knowledge Graph Embedding: Approaches, Applications and Benchmarks. *Electronics* 9, 5 (May 2020), 750. https://doi.org/10.3390/electronics9050750

[9] Alishiba Dsouza, Nicolas Tempelmeier, Ran Yu, Simon Gottschalk, and Elena Demidova. 2021. WorldKG: A World-Scale Geographic Knowledge Graph. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. ACM, Virtual Event Queensland Australia, 4475–4484. https://doi.org/10.1145/3459637.3482023

[10] Nicolas Heist, Sven Hertling, and Heiko Paulheim. 2023. KGrEaT: A Framework to Evaluate Knowledge Graphs via Downstream Tasks. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. ACM, Birmingham United Kingdom, 3938–3942. https://doi.org/10.1145/3583780.3615241

[11] Simon Hug. 2003. Selection Bias in Comparative Research: The Case of Incomplete Data Sets. *Political Analysis* 11, 3 (2003), 255–274. https://doi.org/10.1093/pan/mpg014

[12] Krzysztof Janowicz, Pascal Hitzler, Wenwen Li, and et al. 2022. Know, Know Where, KnowWhereGraph: A densely connected, cross-domain knowledge graph and geo-enrichment service stack for applications in environmental intelligence. *AI Magazine* 43, 1 (March 2022), 30–39. https://doi.org/10.1002/aaai.12043

[13] Jan Portisch, Nicolas Heist, and Heiko Paulheim. 2022. Knowledge graph embedding for data mining vs. knowledge graph embedding for link prediction – two sides of the same coin? *Semantic Web* 13, 3 (April 2022), 399–422. https://doi.org/10.3233/SW-212892

[14] Petar Ristoski and Heiko Paulheim. 2016. RDF2Vec: RDF Graph Embeddings for Data Mining. In *The Semantic Web – ISWC 2016*, Paul Groth, Elena Simperl, Alasdair Gray, Marta Sabou, Markus Krötzsch, Freddy Lecue, Fabian Flöck, and Yolanda Gil (Eds.). Vol. 9981. Springer International Publishing, Cham, 498–514.

[15] Andrea Rossi, Denilson Barbosa, Donatella Firmani, Antonio Matinata, and Paolo Merialdo. 2021. Knowledge Graph Embedding for Link Prediction: A Comparative Analysis. *ACM Transactions on Knowledge Discovery from Data* 15, 2 (April 2021), 1–49. https://doi.org/10.1145/3424672

[16] Zhenfeng Shao, Muhammad Nasar Ahmad, and Akib Javed. 2024. Comparison of Random Forest and XGBoost Classifiers Using Integrated Optical and SAR Features for Mapping Urban Impervious Surface. *Remote Sensing* 16, 4 (Feb. 2024), 665. https://doi.org/10.3390/rs16040665

[17] Kamil Tagowski, Piotr Bielak, and Tomasz Kajdanowicz. 2021. Embedding Alignment Methods in Dynamic Networks. In *Computational Science – ICCS 2021*, Maciej Paszynski, Dieter Kranzlmüller, Valeria V. Krzhizhanovskaya, Jack J. Dongarra, and Peter M. A. Sloot (Eds.). Springer International Publishing, Cham, 599–613.

[18] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Eric Gaussier, and Guillaume Bouchard. 2016. Complex Embeddings for Simple Link Prediction. In *Proceedings of The 33rd International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 48)*, Maria Florina Balcan and Kilian Q. Weinberger (Eds.). PMLR, New York, New York, USA, 2071–2080. https://proceedings.mlr.press/v48/trouillon16.html

A ONLINE RESOURCES

Our coding of this paper can be found under the following Github repository. Data produced in the intermediate steps is made available up on request by the authors.

Received ; revised ; accepted