



# UrbanAgriKG: A knowledge graph on urban agriculture and its embeddings

Bikram Pratim Bhuyan, Ravi Tomar, Thipendra Singh, Amar Ramdane-Cherif

## ► To cite this version:

Bikram Pratim Bhuyan, Ravi Tomar, Thipendra Singh, Amar Ramdane-Cherif. UrbanAgriKG: A knowledge graph on urban agriculture and its embeddings. Open Agriculture, 2024, 9 (1), <10.1515/opag-2022-0265>. <hal-04494487>

**HAL Id: hal-04494487**

**<https://hal.science/hal-04494487v1>**

Submitted on 15 May 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

## Research Article

Bikram Pratim Bhuyan\*, Ravi Tomar, Thipendra P. Singh, Amar Ramdane-Cherif

# UrbanAgriKG: A knowledge graph on urban agriculture and its embeddings

<https://doi.org/10.1515/opag-2022-0265>

received October 4, 2023; accepted January 30, 2024

**Abstract:** This research article outlines a study that examines the creation of a comprehensive knowledge graph specifically designed for the domain of urban agriculture. The research centers on the acquisition, synthesis, and arrangement of pertinent information from various origins in order to establish a specialized knowledge graph tailored for urban agricultural systems. The graph depicts the interrelationships and attributes of various entities, including urban farms, crops, farming methods, environmental factors, and economic elements. Moreover, this study investigates the efficacy of different graph embedding methodologies in the domain of urban agriculture. The aforementioned techniques are utilized in the context of the urban agriculture knowledge graph in order to extract significant representations of entities and their relationships. The primary objective of the experimental study is to investigate and reveal semantic relationships, patterns, and predictions that have the potential to improve decision-making processes and optimize practices in the field of urban agriculture. The results of this study make a significant contribution to the existing body of knowledge in the area of urban agriculture. Additionally, they offer valuable insights into the potential uses of graph embedding techniques within this field.

**Keywords:** urban agriculture, knowledge graph, graph embedding methods, link prediction, graph representation learning, sustainability

\* **Corresponding author: Bikram Pratim Bhuyan**, LISV Laboratory, University of Paris Saclay, 10-12 Avenue of Europe, Velizy, France; School of Computer Science, University of Petroleum and Energy Studies, Dehradun, India, e-mail: [bikram-pratim.bhuyan@universite-paris-saclay.fr](mailto:bikram-pratim.bhuyan@universite-paris-saclay.fr)

**Ravi Tomar:** Persistent Systems, Pune, India

**Thipendra P. Singh:** School of Computer Science, University of Petroleum and Energy Studies, Dehradun, India; School of Computer Science Engineering & Technology, Bennett University, Greater Noida, India

**Amar Ramdane-Cherif:** LISV Laboratory, University of Paris Saclay, 10-12 Avenue of Europe, Velizy, France

## 1 Introduction

Urban agriculture is the act of engaging in the cultivation of crops, the rearing of animals, and the production of food within areas classified as urban or peri-urban [1]. This response addresses the challenges arising from the rapid urbanization process, the limited availability of agricultural land, and the growing demand for local food production within densely populated urban areas [2].

Urban agriculture plays a pivotal role in effectively addressing food security, sustainability, and community resilience concerns [3]. The integration of agriculture into urban areas yields various advantages, including enhanced availability of fresh and nutritious food, decreased carbon emissions resulting from shorter food distribution networks, improved urban microclimates, and strengthened social cohesion through community engagement [4].

A comprehensive comprehension of the dynamics of urban agricultural systems is imperative due to their intricate nature. This entails understanding the interplay between different elements, including farms, markets, distribution networks, and social networks [5]. The utilization of graph representation learning or graph embeddings [6] offers a robust framework for examining and modeling complex relationships, effectively capturing the inherent interdependencies and patterns within urban agriculture [7].

Graph representation learning [8] captures and encodes the structural and semantic information inherent in graphs into continuous vector representations, commonly called embeddings [9]. Graph-based data analysis is a specialized area within machine learning and data mining, which centers on the comprehension and acquisition of knowledge from data organized in a graph structure [10].

Graph representation learning techniques aim to convert the nodes and edges within a graph into low-dimensional vectors, typically situated in a continuous vector space [11]. The embeddings effectively retain crucial characteristics and associations of the initial graph, facilitating the practical execution of analysis, visualization, and prediction endeavors [12].

The primary goal of graph representation learning is to facilitate machines in effectively engaging in reasoning,

classification, and prediction tasks by leveraging on the graph's structure and inherent characteristics. By acquiring vector representations that effectively capture the graph's local and global information, these techniques enable a range of subsequent tasks, including link prediction, node classification, recommendation systems, and anomaly detection [13]. Using graph representation learning principles, scholars can acquire significant knowledge regarding urban agricultural systems' dynamics and long-term viability [14]. This, in turn, can result in enhanced methodologies, allocation of resources, and decision-making processes within urban agriculture.

Recognizing the need for comprehensive and well-organized data within the realm of urban agriculture holds significant importance [15]. The scarcity of standardized data sources in urban agriculture research can be attributed to the emerging state of the field and the decentralized and diverse nature of urban farming practices [16]. The limited availability of resources presents a significant obstacle when attempting to construct a comprehensive and precise knowledge graph for urban agriculture. Nevertheless, notwithstanding these obstacles, the primary objective of this study is to surmount the constraints associated with data and advance the development of a specialized knowledge graph pertaining to urban agriculture.

Given the nascent nature of urban agriculture, there is a noticeable void in utilizing knowledge graph techniques within this domain. The absence of established knowledge graphs pertaining to urban agriculture highlights the originality and innovative nature of this research. Acknowledging that this research represents one of the initial endeavors to construct a knowledge graph within the field of urban agriculture, it underscores the possibility of generating innovative perspectives and advancements in this particular domain.

The research aims to achieve the following:

a. The primary aim of this study is to develop a comprehensive knowledge graph "UrbanAgriKG" pertaining to urban agriculture. This entails collecting, integrating, and organizing pertinent data from various sources, with the specific purpose of constructing a knowledge graph specifically designed for urban agriculture. This encompasses the documentation of the interconnections and characteristics of various entities, such as urban farms, crops, farming techniques, environmental elements, and socioeconomic factors.

b. The objective of this research is to assess the effectiveness of different graph embedding techniques in the domain of urban agriculture. The study seeks to apply these techniques to the urban agriculture knowledge graph, with the goal of acquiring meaningful representations of entities

and relationships. This experimental study aims to facilitate the identification of semantic relationships, patterns, and predictions.

c. Supplement future goals: These potential research areas encompass a range of topics related to urban agriculture. First, one could investigate the effects of urban agriculture on food security, sustainability, and resilience within urban areas. Second, the potential of the knowledge graph to facilitate decision-making and policy development in urban agriculture could be explored. Lastly, the socioeconomic implications and community dynamics associated with urban farming practices could be examined.

## 2 Materials and methods

### 2.1 Dataset

The knowledge graph UrbanAgriKG was primarily constructed by utilizing data acquired from the research project known as "FEW-Meter [17]." The data collection for the project was carried out throughout the 2019 growing season, commencing on March 1st and concluding on October 31st. The data collection methodology utilized a citizen science approach, encompassing case studies conducted in five countries, namely France, Germany, Poland, the United Kingdom, and the United States of America [18]. The study comprised a total of 74 sites, which encompassed a range of urban agricultural spaces. These sites can be classified into three primary categories:

The individual urban garden category encompasses various types found in urban areas, including allotment and home gardens. These spaces are owned or leased by individual gardeners who cultivate crops and plants for personal use or household consumption. The primary emphasis lies on small-scale land parcels that are privately owned or leased.

The urban collective/community garden category encompasses various types of gardens, including community gardens, community farms, and school gardens. Within these designated areas, the cultivated food is distributed among the individuals involved, and the organization operates on a non-profit basis. These gardens frequently function as communal areas where members of the community convene to engage in the cultivation and distribution of agricultural produce. Urban farms fall into the third category, whose primary objective is to generate profit by selling agricultural produce. These agricultural enterprises function on a larger magnitude in contrast to individual and collective gardens and are

distinguished by implementing commercial farming methods. Including these 74 sites enabled us to encompass the spatial range of urban agriculture in our study. In addition, the precipitation and temperature levels were collected from [19,20].

## 2.2 UrbanAgriKG – proposed knowledge graph

The construction of the knowledge graph involved the identification of entities and relationships that are pertinent to the field of urban agriculture, utilizing the data that gathered (Figure 1).

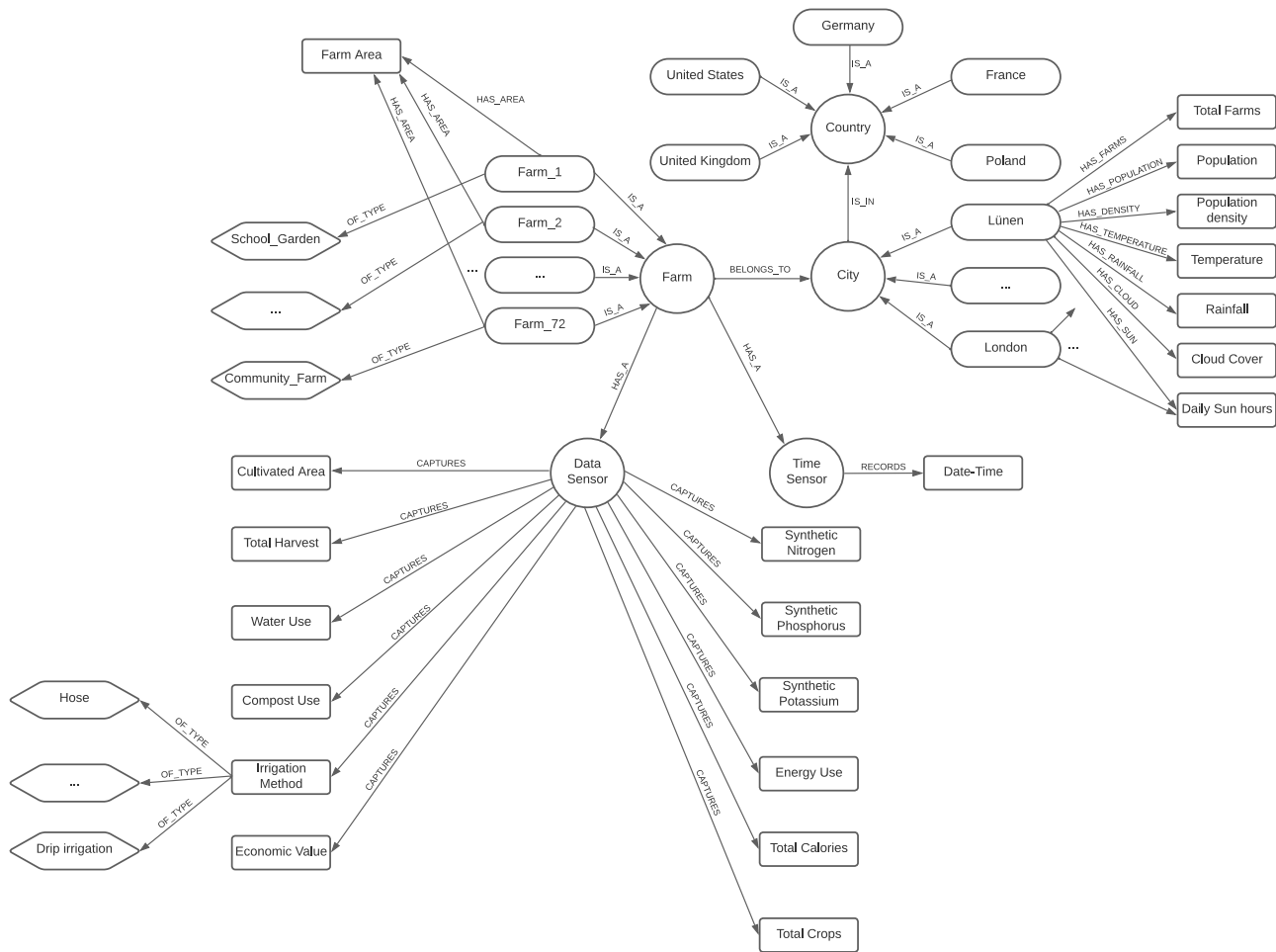
**Definition 1.** We define a knowledge graph “G” by a quadruple  $(\Sigma, \mathfrak{R}, \Psi, \delta)$  where  $\Sigma$  is the set of entities,  $\mathfrak{R} \subseteq \Sigma \times \Sigma$  is the set of relationships,  $\Psi$  is the set of labels, and  $\delta: \mathfrak{R} \rightarrow \Psi$  is the function from relationships to labels.

An assignment of this graph is defined by a triple “ $\Delta$ ,”  $(\alpha, \mathfrak{R}, \beta)$ ;  $\alpha, \beta \in \Sigma$ ; where  $\alpha$  is the head entity,  $\mathfrak{R}$  is the relationship, and  $\beta$  is the tail entity.

We use the YAMO methodology [21] for the creation of the knowledge graph. The following procedures were executed:

**1. Identification of entities:** The critical entities within the domain of urban agriculture were identified through analysis of the collected data. The entities encompassed within this category consist of urban individual gardens, urban collective/community gardens, and urban farms. Furthermore, the knowledge graph was enhanced by incorporating other pertinent entities such as crops, environmental factors (such as soil quality and sunlight exposure), and socio-economic variables (including ownership models and profit-sharing).

**2. Identification of relationships:** Subsequently, we ascertained the relationships that are present among the entities that have been identified. For example,



**Figure 1:** A section of the urban agriculture knowledge graph.

relationships encompass terms such as “HAS\_RAINFALL,” “CAPTURES” “IS\_A” “BELONGS\_TO” and “OF\_TYPE.” The aforementioned relationships encompass the various connections, dependencies, and interactions that exist within the urban agricultural system.

**3. Identification of attributes:** In addition to identifying entities and relationships, we have also identified pertinent attributes that are associated with these entities. The aforementioned attributes encompass crop varieties, irrigation methods, land area, yield, levels of different fertilizers, and economic indicators. The incorporation of attributes enhances the knowledge graph by offering supplementary information and contextual details for every entity.

**4. Construction of the graph:** The knowledge graph was constructed based on the identified entities, relationships, and attributes. In this model, every entity was depicted as a node, while the connections between the nodes represented the relationships and attributes. As a consequence, a graph structure was created to represent the interconnections and semantic associations within the field of urban agriculture.

We find the following characteristics of the knowledge graph as shown in Table 1.

## 2.3 Graph embedding technique on UrbanAgriKG

In the UrbanAgriKG framework, which focuses on entities and relationships relevant to urban agriculture, graph embedding is integral [22]. This technique facilitates the discovery of potential connections among various entities, such as urban gardens, farms, and socio-economic factors, by mapping these entities and their relationships into a vector space. Such a representation allows machine learning algorithms to predict missing links, reveal hidden patterns, and offer deeper insights into the urban agricultural ecosystem.

Key steps in the graph embedding process for UrbanAgriKG are as follows:

**1. Representation of entities and relationships:** In this step, each entity (like urban gardens, crops, environmental factors) and relationship (for instance, “HAS\_RAINFALL,” “IS\_A”) within the knowledge graph is transformed into a vector representation in the embedding space based on a

scoring function  $F$ . This encoding captures the unique characteristics and connections of each element in the graph. Formally,

**Definition 2.** The scoring function  $F$  typically relies on algebraic theory and returns a score based on the vector embedding of the triple “ $\Delta$ ,” ( $\alpha$ ,  $\mathfrak{R}$ , and  $\beta$ ). The vector embedding is represented as  $(\vec{\alpha}, \vec{\mathfrak{R}}, \text{ and } \vec{\beta})$ .

The function depends on proximity of the embedded entities  $(\vec{\alpha}, \vec{\beta})$ ; computed by

$$p_{\alpha,\beta} = F(\vec{\alpha}, \vec{\beta}) \rightarrow \mathbb{R}. \quad (1)$$

The larger value of  $F$  denotes the higher possibility of a relationship between the entities. Thus, the optimization goal is to maximize  $F$  for observable triple “ $\Delta$ ,” ( $\alpha$ ,  $\mathfrak{R}$ , and  $\beta$ ) and minimize for non-observable (negative) ones. This can be obtained using concepts from machine learning like loss function “ $L$ ” and hyper-parameter set “ $\Omega$ .” Thus, the optimization equation with  $\phi$  being the objective function can be written as:

$$\arg \min_{\theta} \phi(L(F(\vec{\alpha}, \vec{\beta}), p_{\alpha,\beta}), \Omega) \quad (2)$$

**2. Dimensionality reduction:** The inherently high-dimensional data of the knowledge graph condensed into a lower-dimensional space. This reduction simplifies computational requirements and assists in uncovering latent patterns while preserving the graph’s essential topological and semantic features. Formally,

**Definition 3.** We define dimensionality reduction of UrbanAgriKG as a function “ $\Theta$ ” which projects a knowledge graph “ $G$ ” to a Euclidean vector space  $\mathbb{R}^{|\Sigma| \times d}$ , where  $|\Sigma|$  is the total number of entities and  $d \ll |\Sigma|$  is the dimension of the Euclidean space  $\mathbb{R}$ . Mathematically,

$$\Theta : G(\Sigma, \mathfrak{R}, \Psi, \delta) \rightarrow \mathbb{R}^{|\Sigma| \times d}. \quad (3)$$

**3. Link prediction:** Utilizing the embeddings in UrbanAgriKG, link prediction is particularly focused on inferring missing entities in the relationships within the urban agricultural context. Link prediction here involves determining either the head or tail entity in a relationship. Formally,

**Definition 4.** The link prediction is the prediction of the head or tail entity. In other words, predict: ( $?$ ,  $\mathfrak{R}$ , and  $\beta$ ) or ( $\alpha$ ,  $\mathfrak{R}$ , and  $?$ ).

In UrbanAgriKG’s context, link prediction plays a pivotal role in uncovering hidden connections within the urban

**Table 1:** Characteristics table of the knowledge graph

$ \Sigma $	$ \mathfrak{R} $	$ \Delta $
730	27	693



agricultural network, especially focusing on relationships like “HAS\_RAINFALL,” “CAPTURES,” “IS\_A,” “BELONGS\_TO,” and “OF\_TYPE.” For instance, under “HAS\_RAINFALL,” the model could predict the typical precipitation level for an urban farm (e.g., predicting the rainfall amount for “Urban Farm A”). In the “CAPTURES” relationship, it might identify the specific environmental benefits or resources that a community garden captures (e.g., “Community Garden B captures high levels of atmospheric carbon”). The “IS\_A” relationship could be used to classify a newly introduced farming method under a broader category (e.g., classifying “Hydroponic Technique C” as a sustainable farming method). For “BELONGS\_TO,” the prediction might link an urban garden to a specific community or collective (e.g., determining which local community initiative “Urban Garden D” is part of). Finally, under “OF\_TYPE,” the algorithm could specify the kind of crop best suited for a particular type of urban farm (e.g., identifying “Crop E” as ideal for rooftop farming). These predictions, leveraging graph embedding techniques, are crucial in enhancing the UrbanAgriKG by revealing potential linkages and fostering a more comprehensive understanding of urban agricultural systems.

The various popular graph embedding methods in the literature with their respective scoring function and power of expressiveness are listed in Table A1.

## 2.4 Link prediction evaluation

The graph embedding-based link prediction techniques discussed in the previous section are evaluated by assessing their performance in ranking a holdout set of triples derived from the original graph. The implementation of this evaluation protocol holds great importance in ensuring reproducibility, as the decisions made during the evaluation process can have a substantial influence on the comparative performance outcomes. The evaluation process entails assessing the model’s capacity to forecast absent entities within a triple accurately. A specific procedure is followed to generate two corrupted sets for each triple in the test set  $\Delta_{\text{Test}} \in \Delta$ . One is the set where the head is corrupted  $\bar{\kappa}_\alpha = \{\bar{\alpha}, \mathbb{R}, \beta\}$  and the other is the set where the tail is corrupted  $\bar{\kappa}_\beta = \{\alpha, \mathbb{R}, \bar{\beta}\}$ . This procedure involves replacing the head entity with every possible entity, as well as replacing the tail entity in a similar manner. This study aims to ensure that the model assigns the original true triple a higher score than the corrupted triples. Nevertheless, it is imperative to make a determination as to whether any authentic triples that are

already included in the compromised sets ought to be eliminated prior to scoring in order to mitigate potential bias.

In certain instances, it is possible for the model to assign identical scores to multiple triples in the test set when they are being ranked. The manner in which this situation is managed can potentially influence the outcomes of the evaluation. One potential strategy involves considering two contrasting scenarios: hypothesizing that the actual triple is located at either the beginning or the conclusion of the ordered list. In this study, the mean rank is determined by computing the average of these two assumptions.

In order to evaluate the efficacy of graph embedding techniques, we employ widely accepted metrics commonly utilized in the domain of knowledge graphs. The metrics that are commonly employed in this context encompass Mean Reciprocal Rank (MRR) and Hits@k [23]. MRR calculates the average of the inverse ranks of the true triples, offering a comprehensive assessment of the model’s capacity to prioritize the correct entities. The Hits@k metric quantifies the proportion of test triples in which the accurate entity is ranked among the top k positions, thereby serving as an indicator of the predictive precision of the model. Formally,

**Definition 5.** A hit is defined as the occurrence of the correct entity within the top-k-ranked entities. The Hits@k score is determined by dividing the number of hits by the total number of test triples. Formally:

$$\text{Hits@k} = \frac{|\{t \in \Delta_{\text{Test}} | \text{rank}(t) \leq k\}|}{|\Delta_{\text{Test}}|}. \quad (4)$$

For each triple in a given set of test triples, the process involves removing either the head or tail entity and substituting it with all entities present in the knowledge graph. This results in a collection of candidate triples. The plausibility scores of these candidate triples are subsequently determined by the scoring function of the knowledge graph embedding model, leading to their ranking.

**Definition 6.** MRR is formally defined as:

$$\text{MRR} = \frac{1}{|\Delta_{\text{Test}}|} \sum_{i=1}^{\Delta_{\text{Test}}} \frac{1}{\text{rank}_i}. \quad (5)$$

The experiments were carried out on computer systems featuring Intel(R) Core(TM) i7-10700 CPU @ 2.90GHz processors, excluding the utilization of graphics processing units (GPUs). Python 3.9.13 was consistently employed as the software environment throughout all experiments. The

experiments on identical hardware and software setups guarantee a standardized and controlled setting, thereby facilitating the comparison and reproducibility of the results.

### 3 Results

We now present the results of the embedding methods for link prediction on UrbanAgriKG. The tuples were initially partitioned into training, testing, and validation sets using an 8:1:1 ratio. We implemented different embedding methods using a stochastic local closed-world assumption training approach, with an embedding dimension of 128 and a random seed of 42.

#### 3.1 Graph embedding without hyper-parameter optimization

Table 2 provides an analysis of various graph embedding functions, designated by  $F$ , evaluated on specific metrics before the optimization of hyper-parameters. These metrics, namely Hits@ $N$  (for  $N = 1, 3, 5, 10$ ) and MRR, are crucial in assessing the quality of embeddings, especially in link prediction tasks. The Hits@ $N$  metric measures the fraction of

times the true positive is within the top  $N$  predictions, while MRR gives an average ranking for the positive examples.

TransH [13] emerges as a dominant performer, leading in Hits@1, Hits@5, Hits@10, and MRR. This indicates that, even without optimization, TransH consistently ranks the true positives at the top of its predictions and maintains a higher average rank across the board. TransR, while not the absolute best in any metric, remains competitive, particularly in Hits@3, suggesting that it often ranks the true positives within the top three predictions.

On the other end of the spectrum, models like TransE [24], TransD [26], and TransF [27] exhibit zero or extremely low performance in Hits@1, hinting that these models, in their current non-optimized state, struggle to rank the true positive as the top prediction. Furthermore, models like TransF [27], ComplEx [30], and SimpleE [34] have particularly low scores across most metrics, indicating potential challenges in capturing the nuances of the data or perhaps requiring significant hyper-parameter tuning.

An anomalous observation can be seen in the performance of BoxE [32]. Its Hits@10 score is a mere 0.005, which contrasts sharply with its decent performance in Hits@1, Hits@3, and Hits@5. This divergence is unusual because a model that performs well at lower Hits@ $N$  values generally maintains or even improves its performance as  $N$  increases. Such an inconsistency suggests potential issues in the model's behavior or evaluation process and would warrant further investigation.

Another noteworthy point is the evident disparity in the performance of different models, with some models like TransH and TransR showcasing robust performance, while others like SimpleE and ComplEx lag behind. This disparity underscores the importance of hyper-parameter optimization, as the non-optimized state can lead to underutilization of a model's potential.

Table 3 showcases selected hyper-parameters for various graph embedding functions, which are pivotal in determining the performance and characteristics of these models. These hyper-parameters have been identified through a rigorous process that is well-documented. For training, the study employs the Margin Ranking Loss combined with the Adam optimizer, utilizing the stochastic local closed-world assumption training approach.

#### 3.2 Graph embedding after hyper-parameter optimization

After the optimization process, the performance of the models is presented in Table 4.

**Table 2:** Evaluation matrices without optimization of hyper-parameters with embedding dimension = 128, random seed = 42, and stochastic local closed world assumption training approach

$F$	Hits@1	Hits@3	Hits@5	Hits@10	MRR
TransE [24]	0.0	0.080	0.135	0.221	0.070
TransH [13]	<b>0.345</b>	0.451	<b>0.489</b>	<b>0.517</b>	<b>0.422</b>
TransR [25]	0.287	<b>0.454</b>	0.479	0.503	0.379
TransD [26]	0.0	0.112	0.209	0.327	0.089
TransF [27]	0.0	0.008	0.017	0.032	0.016
RotatE [28]	0.034	0.063	0.077	0.112	0.062
DistMult [29]	0.083	0.117	0.129	0.178	0.121
ComplEx [30]	0.002	0.005	0.011	0.020	0.012
AutoSF [31]	0.005	0.005	0.005	0.005	0.011
BoxE [32]	0.137	0.275	0.396	0.005	0.251
HolE [33]	0.008	0.011	0.017	0.031	0.019
SimpleE [34]	0.0	0.002	0.002	0.011	0.006
CP [35]	0.002	0.005	0.005	0.014	0.010
QuatE [36]	0.045	0.109	0.137	0.192	0.098
PairRE [37]	0.238	0.316	0.370	0.413	0.302
MuRP [38]	0.189	0.261	0.339	0.479	0.271
CrossE [39]	0.135	0.215	0.241	0.281	0.190
ConvKB [40]	0.054	0.135	0.189	0.281	0.134
ConvE [41]	0.017	0.054	0.074	0.129	0.055

Bold values represent the highest among all the values of the function.

**Table 3:** Hyper-parameters used for the scoring functions, with Margin Ranking Loss, Adam optimizer, and stochastic local closed world assumption training approach

F	Embedding_dim	Loss_margin	Learning_rate	Num_epochs	Batch_size	Num_negatives
TransE	300	12	0.0177	200	1,024	71
TransH	400	8	0.0329	600	512	18
TransR	400	4	0.0245	800	4,096	4
TransD	100	7	0.0594	500	256	57
TransF	500	21	0.0317	500	64	86
RotatE	200	16	0.0086	500	256	10
DistMult	400	1	0.0285	1,000	256	7
ComplEx	300	11	0.0456	800	128	23
AutoSF	500	16	0.0939	600	32	11
BoxE	100	19	0.0870	600	4,096	43
HolE	100	12	0.0155	1,000	4,096	2
SimpleE	100	9	0.0909	700	4,096	72
CP	100	21	0.0941	200	256	6
QuatE	300	22	0.0448	600	16	67
PairRE	300	11	0.0379	900	16	4
MuRP	200	9	0.0417	900	4,096	87
CrossE	100	16	0.0902	200	1,024	17
ConvKB	300	12	0.0177	200	1,024	71
ConvE	300	12	0.0177	200	1,024	71

The search is within the Embedding\_dim  $\in \{100, 200, 300, 400, 500\}$  and Loss\_margin  $\in [1, 24]$ .

TransH emerges as a standout performer in Hits@1 and Hits@3, indicating that it frequently ranks the true positive links as the top prediction or within the top three predictions. TransR, while not the absolute leader in Hits@1 or Hits@3, shows remarkable performance in Hits@5 and Hits@10, and

leads in MRR. This suggests that TransR consistently ranks the true positive links high in its predictions.

Conversely, models like TransE, TransD, and RotatE exhibit lower scores in Hits@1, hinting that these models might face challenges in ranking the true positive as the top prediction, especially when compared to models like TransH and TransR. The performance of TransF and AutoSF is particularly noteworthy, as these models show relatively low scores across all metrics, indicating potential challenges in capturing the intricacies of the data or that their optimal hyper-parameter configurations might still not be fully realized.

A few models, such as BoxE, showcase discrepancies in their performance across different metrics. For instance, BoxE's Hits@10 score is remarkably lower than its performance in Hits@1, Hits@3, and Hits@5. Such inconsistencies might indicate specific nuances in the model's behavior or the nature of the dataset it was evaluated.

This study compares Margin Ranking loss and the number of epochs in different embedding methods. Margin Ranking loss, a ranking optimization technique, aims to improve the ordering of positive and negative examples in a knowledge graph. Initially, all methods show high loss values due to random initial embeddings and lack of meaningful representations. As training progresses, a decline in loss values is expected, indicating better comprehension of the knowledge graph. Variability in the rate of loss reduction is observed

**Table 4:** Evaluation matrices after optimization of hyper-parameters

F	Hits@1	Hits@3	Hits@5	Hits@10	MRR
TransE	0.316	0.442	0.465	0.517	0.398
TransH	<b>0.435</b>	<b>0.542</b>	0.599	0.617	0.422
TransR	0.289	0.517	<b>0.647</b>	<b>0.662</b>	<b>0.490</b>
TransD	0.057	0.357	0.562	0.660	0.458
TransF	0.114	0.436	0.459	0.488	0.286
RotatE	0.103	0.120	0.121	0.172	0.128
DistMult	0.232	0.272	0.289	0.306	0.265
ComplEx	0.054	0.093	0.134	0.159	0.121
AutoSF	0.016	0.027	0.032	0.138	0.103
BoxE	0.254	0.385	0.401	0.012	0.464
HolE	0.016	0.025	0.033	0.042	0.098
SimpleE	0.011	0.014	0.015	0.016	0.019
CP	0.005	0.016	0.015	0.031	0.086
QuatE	0.057	0.198	0.214	0.215	0.147
PairRE	0.257	0.397	0.412	0.455	0.259
MuRP	0.199	0.287	0.367	0.498	0.422
CrossE	0.210	0.296	0.311	0.316	0.147
ConvKB	0.145	0.155	0.198	0.298	0.134
ConvE	0.021	0.061	0.121	0.201	0.101

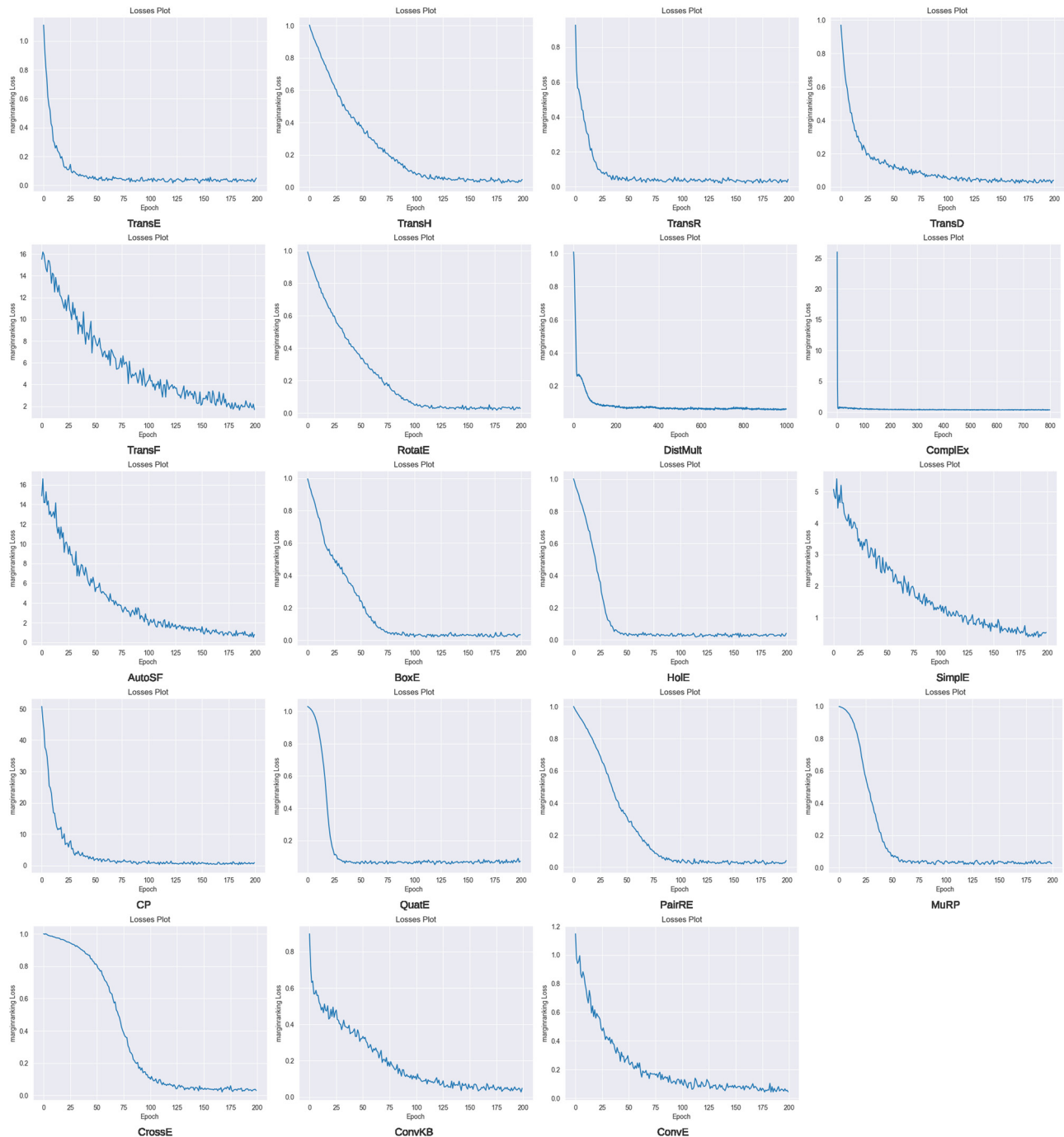
Bold values represent the highest among all the values of the function.



across methods, with some converging faster and others requiring more iterations for optimal performance. The results are illustrated in Figure 2, while Hit ratios and MRR are shown in Figures 3 and 4.

As an overview, TransE, initially struggling with accurate predictions, showed remarkable improvement post-optimization, illustrating its potential in modeling urban

agricultural data. TransH, with its inherent strength in handling complex relationships, displayed robust initial performance, further enhanced after fine-tuning. TransR's unique approach to separate semantic spaces for entities and relations led to commendable baseline results, significantly improved with optimization. Similarly, models like TransD, TransF, and RotatE, each with their



**Figure 2:** Performance (loss functions) of various embedding methods in accordance with the number of epochs against the Margin Ranking loss on UrbanAgriKG.

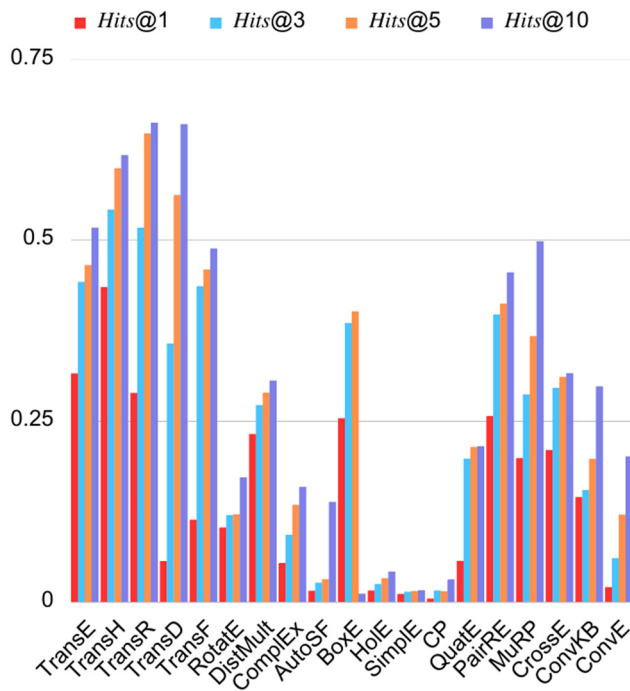


Figure 3: Hit ratios of various embedding methods on UrbanAgriKG.

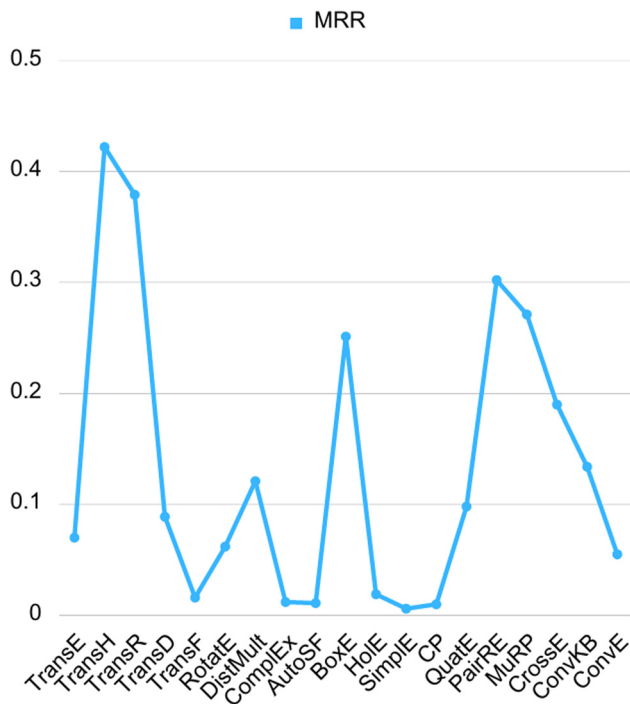


Figure 4: MRR ratios of various embedding methods on UrbanAgriKG.

distinctive mechanisms, demonstrated varying initial performances but showed notable improvements post-optimization, indicating their utility in diverse urban agricultural contexts. The bilinear and complex models like DistMult and

ComplEx initially faced challenges but improved significantly, suggesting their potential to capture complex urban agriculture relationships. Though starting modestly, innovative models such as AutoSF, BoxE, and HolE exhibited substantial post-optimization gains, emphasizing their adaptability. Simple, CP, QuatE, MuRP, PairRE, CrossE, ConvKB, and ConvE each displayed unique strengths and limitations, but all showed enhanced performance post-optimization, underscoring the importance of tailored hyperparameter tuning for each model to capture the nuances of urban agriculture effectively.

## 4 Discussion

### 4.1 Impact of UrbanAgriKG on decision-making and policy development

The UrbanAgriKG model, through its advanced data analysis and graph embedding techniques, offers a profound understanding of the interactions between various stakeholders in urban agriculture, such as farmers, community groups, and policymakers. By utilizing techniques like TransE, TransH, and TransR, the model effectively captures and represents the complex relationships within the urban agricultural system. For example, the high accuracy of TransH in predicting top relevant links (as shown in Table 4) aids in precisely identifying the relationships and dependencies between urban farms and environmental factors like soil quality and rainfall. This granular level of understanding is crucial for stakeholders to appreciate the dynamics of urban agriculture and how their actions or decisions might influence the system. For example, by analyzing the “HAS\_RAINFALL” relationship, stakeholders can discern how different levels of rainfall impact various crops and urban farming practices. This is particularly useful for urban farmers in planning irrigation and crop selection strategies. Moreover, the “IS\_A” relationship helps in classifying urban agricultural entities, such as distinguishing between different types of urban gardens (e.g., individual vs community gardens) or identifying various farming methods (e.g., organic farming, hydroponics). This classification is vital for understanding the diversity within urban agricultural practices.

Insights derived from UrbanAgriKG can significantly inform policy decisions in urban agricultural contexts. The knowledge graph’s ability to reveal patterns and connections, such as the ones indicated by the high performance of TransR in broader predictive accuracy, provides a data-driven foundation for policy formulation. These insights

can guide resource allocation, helping policymakers decide where to invest in urban agriculture infrastructure, such as community gardens or urban farms. Additionally, understanding the impacts of different socio-economic variables and environmental factors on urban agriculture can lead to more effective urban planning and targeted agricultural support programs. For instance, the “BELONGS\_TO” relationship can inform policymakers about the affiliations of urban gardens with specific community groups or NGOs, aiding in targeted support and resource allocation. Additionally, the “OF\_TYPE” relationship can be used to categorize urban agricultural practices, helping policymakers to tailor support programs to specific types of urban farms or gardens, such as rooftop gardens or vertical farms.

## 4.2 UrbanAgriKG’s role in enhancing food security and sustainability

UrbanAgriKG plays a pivotal role in analyzing and predicting the impact of urban agricultural practices on local food security. The knowledge graph, enriched with data on various urban agricultural entities and their interrelations, enables a comprehensive analysis of how these practices contribute to the availability and accessibility of food in urban areas. For instance, the relationship “CAPTURES,” when applied to different urban farms, can reveal which practices are most effective in producing a stable food supply, thus enhancing food security. Additionally, the “IS\_A” and “OF\_TYPE” relationships can help in identifying which types of urban gardens or farms (e.g., community gardens, rooftop farms) are most productive or provide the most nutritional value, enabling targeted interventions to bolster food security in specific urban communities.

Moreover, by predicting potential links, UrbanAgriKG can forecast the impacts of emerging urban agricultural trends on food availability. For example, the model can predict how introducing new farming techniques or expanding urban gardens in a particular area might enhance or diminish food security. This predictive capability is crucial for planning and implementing strategies to ensure a consistent and reliable food supply in urban settings.

UrbanAgriKG also plays a significant role in identifying and promoting sustainable urban farming practices. By analyzing relationships like “BELONGS\_TO,” the knowledge graph can help in understanding which community groups or initiatives are engaged in sustainable farming practices, thereby facilitating the sharing of best practices across different urban settings. The “HAS\_RAINFALL”

relationship, for instance, can be used to analyze how different levels of rainfall affect various sustainable farming practices, enabling adjustments and improvements to be made for greater efficiency and sustainability.

Furthermore, UrbanAgriKG can assist in evaluating the environmental impact of different urban agricultural practices by analyzing their carbon footprint, water usage, and biodiversity impact, among other factors. This evaluation is essential for promoting practices that contribute to food security and align with broader sustainability goals, such as reducing urban carbon emissions and conserving water resources.

## 4.3 Socio-economic implications and community dynamics

UrbanAgriKG serves as a vital tool for analyzing the economic impact of urban agriculture, providing insights into aspects such as profitability, job creation, and community engagement. By leveraging relationships such as “BELONGS\_TO” and “OF\_TYPE,” the knowledge graph can elucidate the economic dynamics of different urban agricultural practices. For example, by assessing which types of urban farms or gardens (“OF\_TYPE”) are the most profitable or which community groups (“BELONGS\_TO”) have the most significant economic impact, policymakers and investors can make more informed decisions regarding funding and support.

Additionally, UrbanAgriKG’s ability to predict and analyze trends in urban agriculture aids in forecasting future economic outcomes. This predictive analysis is crucial for long-term economic planning and ensuring the sustainability of urban agricultural initiatives from an economic standpoint.

UrbanAgriKG is also crucial in understanding and enhancing community involvement in urban agriculture. By analyzing the “CAPTURES” relationship, for instance, the model can identify urban agricultural practices that capture community interest or engagement, helping to promote more inclusive and community-oriented urban farming initiatives. The “BELONGS\_TO” relationship provides insights into how different urban gardens or farms are integrated into community structures, revealing the dynamics of community participation and collaboration.

Furthermore, the knowledge graph can help identify gaps or opportunities for increased community involvement in urban agriculture. This could involve predicting which types of urban gardens or farming practices might be more appealing to different community groups, thereby

fostering collaborative farming efforts and stronger community bonds. By facilitating a better understanding of these community dynamics, UrbanAgriKG can contribute significantly to the development of more community-centric urban agriculture models that not only support food production but also enhance social cohesion and community well-being.

#### 4.4 Technical challenges and future directions

One of the primary technical challenges encountered in the development of UrbanAgriKG was managing the heterogeneity of data and relationships with multiple arity, which resulted in less-than-optimal results in some instances. The diverse nature of data in urban agriculture, ranging from environmental factors to socio-economic metrics, posed a significant challenge in maintaining consistency and accuracy in the knowledge graph. This heterogeneity often led to difficulties in effectively embedding and analyzing the data, impacting the overall predictive performance of the model. To address this, future iterations of UrbanAgriKG could explore the use of higher-dimensional knowledge graphs. These would provide a more nuanced and detailed representation of the complex, multi-faceted data involved in urban agriculture, potentially leading to more accurate and reliable results.

## 5 Conclusion

In conclusion, this study introduces UrbanAgriKG as a method for representing urban agricultural concepts. Our investigation into creating an Urban Agricultural knowledge graph and applying embedding techniques for link prediction highlights their potential to aid urban farming practices. This lays a foundational framework for future research and potential applications, aiming to progressively enhance urban food systems and contribute to their sustainability through improved computational models.

**Acknowledgments:** This work is supported by the “ADI 2022” project funded by the IDEX Paris-Saclay, ANR-11-IDEX-0003-02.

**Funding information:** This research received no external funding.

**Author contributions:** All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

**Conflict of interest:** The authors state no conflict of interest.

**Data availability statement:** All data generated or analysed during this study are included in this published article [and its supplementary information files].

## References

- [1] Azunre G, Amponsah O, Peprah C, Takyi S, Braimah I. A review of the role of urban agriculture in the sustainable city discourse. *Cities*. 2019;93:104–19.
- [2] Langemeyer J, Madrid-Lopez C, Beltran A, Mendez G. Urban agriculture—A necessary pathway towards urban resilience and global sustainability? *Landscape Urban Plan*. 2021;210:104055.
- [3] Oliveira Alves D, Oliveira L. Commercial urban agriculture: a review for sustainable development. *Sustainable Cities and Society*. 2022;87:104185.
- [4] Skar S, Pineda-Martos R, Timpe A, Pölling B, Bohn K, Kylvik M, et al., Urban agriculture as a keystone contribution towards securing sustainable and healthy development for cities in the future. *Blue-Green Systems*. 2020;2(1):1–27.
- [5] Bhuyan B, Tomar R, Gupta M, Ramdane-Cherif A. An ontological knowledge representation for smart agriculture. In: 2021 IEEE International Conference on Big Data (Big Data). IEEE; 2021. p. 3400–6.
- [6] Chen F, Wang Y, Wang B, Kuo C. Graph representation learning: a survey. *APSIPA Transactions on Signal and Information Processing*. 2020;9:e15.
- [7] Bhuyan B, Tomar R, Cherif A. A Systematic Review of Knowledge Representation Techniques in Smart Agriculture (Urban). *Sustainability*. 2022;14(22):15249.
- [8] Hamilton W. Graph representation learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*. 2020;14(3):1–159.
- [9] Wang M, Qiu L, Wang X. A survey on knowledge graph embeddings for link prediction. *Symmetry*. 2021;13(3):485.
- [10] Ji S, Pan S, Cambria E, Marttinen P, Philip S. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Trans Neural Netw Learn Syst*. 2021;33(2):494–514.
- [11] Veličković P. Everything is connected: Graph neural networks. *Current Opinion in Structural Biology*. 2023;79:102538.
- [12] Wang Q, Mao Z, Wang B, Guo L. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*. 2017;29(12):2724–43.
- [13] Wang Z, Zhang J, Feng J, Chen Z. Knowledge graph embedding by translating on hyperplanes. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 28; 2014.
- [14] Bhuyan B, Tomar R, Yadav A. Edge computing in smart agriculture. *Future Connected Technologies: Growing Convergence and Security Implications*. 2023:200.
- [15] Brown K, Jameton A. Public health implications of urban agriculture. *Journal of Public Health Policy*. 2000;21:20–39.

- [16] Sharma R, Kamble S, Gunasekaran A, Kumar V, Kumar A. A systematic literature review on machine learning applications for sustainable agriculture supply chain performance. *Computers & Operations Research*. 2020;119:104926.
- [17] Available from: <http://www.fewmeter.org/>.
- [18] Dorr E, Hawes J, Goldstein B, Fargue-Lelièvre A, Fox-Kämper R, Specht K, et al. Food production and resource use of urban farms and gardens: a five-country study. *Agronomy for Sustainable Development*. 2023;43(1):18.
- [19] Available from: <https://data.europa.eu/>, 2028.
- [20] Available from: <https://www.climate.gov/>.
- [21] Dutta B, Chatterjee U, Madalli D. YAMO: yet another methodology for large-scale faceted ontology construction. *Journal of Knowledge Management*. 2015;19(1):6–24.
- [22] Cai H, Zheng V, Chang K. A comprehensive survey of graph embedding: Problems, techniques, and applications. *IEEE transactions on knowledge and data engineering*. 2018;30(9):1616–37.
- [23] Guan N, Song D, Liao L. Knowledge graph embedding with concepts. *Knowledge-Based Systems*. 2019;164:38–44.
- [24] Bordes A, Usunier N, Garcia-Duran A, Weston J, Yakhnenko O. Translating embeddings for modeling multi-relational data. *Advances in Neural Information Processing Systems*. 2013;26:2787–95.
- [25] Lin Y, Liu Z, Sun M, Liu Y, Zhu X. Learning entity and relation embeddings for knowledge graph completion. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 29; 2015.
- [26] Ji G, He S, Xu L, Liu K, Zhao J. Knowledge graph embedding via dynamic mapping matrix. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (vol. 1:long papers)*; 2015. p. 687–96.
- [27] Feng J, Huang M, Wang M, Zhou M, Hao Y, Zhu X. Knowledge graph embedding by flexible translation. In: *Fifteenth International Conference on the Principles of Knowledge Representation and Reasoning*; 2016.
- [28] Sun Z, Deng Z, Nie J, Tang J. Rotate: Knowledge graph embedding by relational rotation in complex space. *arXiv:1902.10197*; 2019.
- [29] Yang B, Yih W, He X, Gao J, Deng L. Embedding entities and relations for learning and inference in knowledge bases. *arXiv:1412.6575*. 2014.
- [30] Trouillon T, Welbl J, Riedel S, Gaussier É, Bouchard G. Complex embeddings for simple link prediction. In: *International Conference on Machine Learning*. PMLR; 2016. p. 2071–80.
- [31] Zhang Y, Yao Q, Dai W, Chen L. AutoSF: Searching scoring functions for knowledge graph embedding. In: *2020 IEEE 36th International Conference on Data Engineering (ICDE)*. IEEE; 2020. p. 433–44.
- [32] Abboud R, Ceylan I, Lukasiewicz T, Salvatori T. Boxe: A box embedding model for knowledge base completion. *Advances in Neural Information Processing Systems*. 2020;33:9649–61.
- [33] Nickel M, Rosasco L, Poggio T. Holographic embeddings of knowledge graphs. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 30; 2016.
- [34] Kazemi S, Poole D. Simple embedding for link prediction in knowledge graphs. *Advances in neural information processing systems*. 2018;31:4289–300.
- [35] Lacroix T, Usunier N, Obozinski G. Canonical tensor decomposition for knowledge base completion. In: *International Conference on Machine Learning*. PMLR; 2018. p. 2863–72.
- [36] Tran H, Takasu A. MEIM: Multi-partition embedding interaction beyond block term format for efficient and expressive link prediction. 2022. *arXiv:1902.10197*.
- [37] Chao L, He J, Wang T, Chu W. Pairre: Knowledge graph embeddings via paired relation vectors. *arXiv:2011.03798*. 2020.
- [38] Balazevic I, Allen C, Hospedales T. Multi-relational poincaré graph embeddings. *Advances in Neural Information Processing Systems*. 2019;32:4463–73.
- [39] Zhang W, Paudel B, Zhang W, Bernstein A, Chen H. Interaction embeddings for prediction and explanation in knowledge graphs. In: *Proceedings of The Twelfth ACM International Conference on Websearch and Data Mining*; 2019. p. 96–104.
- [40] Nguyen D, Nguyen T, Nguyen D, Phung D. A novel embedding model for knowledge base completion based on convolutional neural network. *arXiv:1712.02121*; 2017.
- [41] Dettmers T, Minervini P, Stenetorp P, Riedel S. Convolutional 2d knowledge graph embeddings. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 32; 2018.



## Appendix

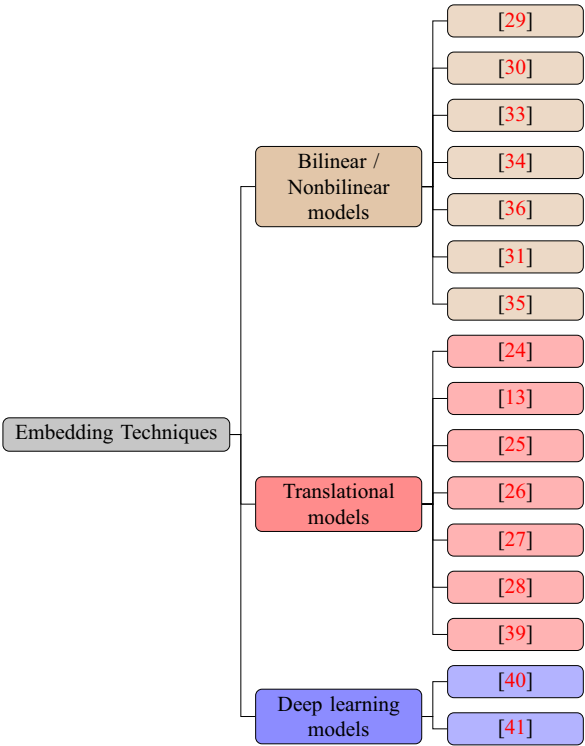
There are three basic types of graph embedding techniques, as shown in Figure A1: bilinear/non-bilinear techniques, translational techniques, and deep learning approaches. Using bilinear or other non-linear operations to describe the complex connections in the graph, bilinear/non-bilinear approaches concentrate on capturing interactions between entities and relations. To highlight the geometric structure of the network, translational techniques like TransE [24], TransH [13], and TransR [25] express entities and relations

as translations in the embedding space. Neural networks are used in deep learning techniques, such as convolutional neural networks and recurrent neural networks [40,41], to learn complicated representations of items and interactions. These techniques use deep learning architectures, which are able to capture local and global information simultaneously to generate more expressive embeddings for use in a wide range of graph analytic applications.

**Table A1:** Graph embedding functions with their parameters

F	Reference	Embedding	Scoring function	Expressiveness
TransE	[24]	$\alpha, \mathfrak{R}, \beta \in \mathbb{R}^d$	$\ \alpha + \mathfrak{R} - \beta\ _2^2$	Antisymmetric, inverse, composite
TransH	[13]	$\alpha, \mathfrak{R}, \beta \in \mathbb{R}^d$	$\ \vec{\alpha} + \mathfrak{R} - \vec{\beta}\ _2^2$	Symmetric, antisymmetric
TransR	[25]	$\alpha, \beta \in \mathbb{R}^d, \mathfrak{R} \in \mathbb{R}^k, M_{\mathfrak{R}} \in \mathbb{R}^{k \times d}$	$\ \alpha M_{\mathfrak{R}} + \mathfrak{R} - \beta M_{\mathfrak{R}}\ _2^2$	Symmetric, antisymmetric
TransD	[26]	$\alpha, \beta \in \mathbb{R}^d, \mathfrak{R} \in \mathbb{R}^k$	$-\ \vec{\alpha} + \mathfrak{R} - \vec{\beta}\ _2^2$	Symmetric, antisymmetric
TransF	[27]	$\alpha, \mathfrak{R}, \beta \in \mathbb{R}^d$	$(\alpha + \mathfrak{R})^T \beta + (\beta - \mathfrak{R})^T \alpha$	Symmetric, antisymmetric
RotatE	[28]	$\alpha, \mathfrak{R}, \beta \in \mathbb{I}^d$	$\ \alpha \circ \mathfrak{R} - \beta\ $	Symmetric, antisymmetric, inverse, composite
DistMult	[29]	$\alpha, \mathfrak{R}, \beta \in \mathbb{R}^d$	$\langle \alpha, \mathfrak{R}, \beta \rangle$	Symmetric
ComplEx	[30]	$\alpha, \mathfrak{R}, \beta \in \mathbb{I}^d$	$\text{Re}(\langle \alpha, \mathfrak{R}, \text{conj}(\beta) \rangle)$	Symmetric, antisymmetric, inverse
AutoSF	[31]	$\alpha, \mathfrak{R}, \beta \in \mathbb{R}^d$	$\langle \alpha^T, \mathfrak{R}, \beta \rangle$	Symmetric, antisymmetric, inverse
BoxE	[32]	$\alpha, \mathfrak{R}, \beta \in \mathbb{R}^d$	$-\sum_i \ \text{dist}(\Sigma_i^{\mathfrak{R}(\Sigma_1, \dots, \Sigma_n)}, \mathfrak{R}^i)\ $	Symmetric, antisymmetric, inverse
HolE	[33]	$\alpha, \mathfrak{R}, \beta \in \mathbb{I}^d$	$\text{Re}(\langle \alpha, \mathfrak{R}, \text{conj}(\beta) \rangle)$	Symmetric, antisymmetric, inverse
SimplE	[34]	$\alpha^{\sim}, \mathfrak{R}^{\sim}, \beta^{\sim} \in \mathbb{R}^d; \alpha^*, \mathfrak{R}^*, \beta^* \in \mathbb{R}^d$	$\langle \alpha^{\sim}, \mathfrak{R}^{\sim}, \beta^* \rangle + \langle \alpha^*, \mathfrak{R}^*, \beta^{\sim} \rangle$	Symmetric, antisymmetric, inverse
CP	[35]	$\alpha^{\sim}, \mathfrak{R}^{\sim}, \beta^{\sim} \in \mathbb{R}^d; \alpha^*, \mathfrak{R}^*, \beta^* \in \mathbb{R}^d$	$\langle \alpha^{\sim}, \mathfrak{R}^{\sim}, \beta^* \rangle + \langle \alpha^*, \mathfrak{R}^*, \beta^{\sim} \rangle$	Symmetric, antisymmetric, inverse
QuatE	[36]	$\alpha, \mathfrak{R}, \beta \in \mathbb{Q}^d$	$\alpha \otimes \mathfrak{R} \cdot \beta$	Symmetric, antisymmetric, inverse
PairRE	[37]	$\alpha, \mathfrak{R}^a, \mathfrak{R}^b, \beta \in \mathbb{R}^d$	$\ \alpha \circ \mathfrak{R}^a - \beta \circ \mathfrak{R}^b\ $	Symmetric, antisymmetric, inverse, composite
MuRP	[38]	$\alpha, \mathfrak{R}, \beta \in \mathbb{B}^d$	see equation (5) [38]	Antisymmetric, inverse, composite
CrossE	[39]	$\alpha, \mathfrak{R}, \beta \in \mathbb{R}^d$	$\sigma(\tanh(\alpha + \mathbb{R} + G)\beta^T)$	—
ConvKB	[40]	$\alpha, \mathfrak{R}, \beta \in \mathbb{R}^d$	$\text{concat}(g([\alpha, \mathbb{R}, \beta]^* \omega)W)$	Symmetric, antisymmetric, inverse, composite
ConvE	[41]	$\alpha, \mathfrak{R}, \beta \in \mathbb{R}^d$	$g(\text{vec}(g(\text{concat}(\alpha, \mathbb{R})^* \omega))W) \cdot \beta$	Symmetric, antisymmetric, inverse, composite

$M_{\mathfrak{R}}$  represents the projection matrix from a head-tail pair to a different space of specific relation, and  $M_{\mathfrak{R}}^T$  represents the transpose of the matrix.  $\circ$  represents the element-wise product (Hadamard product) of the vectors.  $\langle \alpha, \mathfrak{R}, \beta \rangle$  is the dot-product of the vectors.  $\mathbb{I}^d$  represents the imaginary complex space of dimension ‘d’ for a complex vector. For SimplE the embedding and dimensions are split into parts; i.e.  $\alpha = \alpha^{\sim} + \alpha^*$ ;  $\mathfrak{R} = \mathfrak{R}^{\sim} + \mathfrak{R}^*$ ;  $\beta = \beta^{\sim} + \beta^*$  and  $d = d^{\sim} + d^*$ .  $\mathbb{Q}^d$  represents Hamilton’s Quaternion hypercomplex space and  $\otimes$  is the Hamiltonian product of the vectors. PairRE takes  $\mathfrak{R}^a$  and  $\mathfrak{R}^b$  as relation projections for the head and tail entities, respectively. In BoxE method,  $\mathbb{B}^d$  is the box dimension.  $G \in \mathbb{R}^d$  represents the global bias vector in CrossE. In the deep learning models i.e. ConvE and ConvKB,  $g$  represents a non-linear function,  $*$  is the convolution operator and  $\omega$  is the set of filters.



**Figure A1:** Classification of knowledge graph embedding methods.