RESEARCH ARTICLE

Integrating Relationship Path and Entity Neighbourhood Information for Knowledge Graph Intelligence of Social Things

Mohammad Shabaz[1,*] **and Mukesh Soni**[2]

[1] Department of Computer Science Engineering, Model Institute of Engineering and Technology, Jammu, J&K 181123, India

[2] Division of Research and Development, Lovely Professional University, Phagwara 144411, India

Abstract

In the evolving framework of the Intelligence of Social Things (IoST), which amalgamates social networks and IoT ecosystems, knowledge graphs are essential for facilitating networked systems to efficiently process and leverage intricate relational data. Knowledge graphs offer essential technical assistance for various artificial intelligence applications, such as e-commerce, intelligent navigation, healthcare, and social media. Nonetheless, current knowledge graphs frequently lack completeness, harboring a considerable quantity of implicit knowledge that remains to be revealed. Consequently, tackling the difficulty of finalising knowledge graphs has emerged as a pressing research priority. Most contemporary methods separately analyse entity neighbourhood information or connection routes, neglecting the significance of entity neighbourhood information in the investigation of relationship paths. A novel approach, RPEN-KGC (Relationship Path and Entity Neighbourhood Knowledge Graph Completion), is suggested to enable the fusion of relationship paths and entity neighbourhood information for knowledge graph completion. RPEN-KGC comprises a sampler and an inferencer. The sampler conducts random walks between entity pairs to furnish dependable inference methods for the inferencer. The sampler utilises a contrastive method grounded in entity neighbourhood similarity to steer random walks, hence enhancing sampling efficiency and augmenting inference strategies. The inferencer derives semantic characteristics of relationship paths and deduces a greater variety of relationship paths within the semantic domain. Experiments performed on the public NELL-995 and FB15K-237 datasets for the link prediction task indicate that RPEN-KGC significantly enhances most metrics relative to baseline approaches. These findings demonstrate that RPEN-KGC proficiently forecasts absent information in knowledge graphs.

Keywords: intelligence of social things, artificial intelligence, knowledge graph, relationship path, entity neighborhood information, graph completion.

1 Introduction

Google introduced the concept of Knowledge Graphs in 2012, significantly enhancing the ability to manage, understand, and organize the vast amounts of data in the world. This advancement is attributed to its structured triplet form, which effectively represents entities, their attributes, and the relationships between them [1]. Over the last many years, knowledge graphs have played an increasingly important role in various perception tasks [2], especially bringing vitality to areas like intelligent question answering [3], software development [4], knowledge alignment [5], and intelligent recommendation [6]. There are many large-scale knowledge graphs, such as DBPedia [7], Freebase [8], NELL [9], and WordNet [10], which are widely used in numerous knowledge-aware artificial intelligence applications. Although knowledge graphs have high application value, they suffer from incompleteness due to the large amount of valuable knowledge that is implicit or missing. Data shows that in current large knowledge graphs, some common basic relationships (such as birth rate, nationality, parents, etc.) are missing over 70% of the time, and there is considerably little evidence of connections with reduced usage rates [11]. To increase the knowledge graph's scalability, Knowledge Graph Completion (KGC) attempts to predict and complete the missing triplets [12].

The main difficulties and challenges currently faced include: first, in the face of complex knowledge graph structures, existing methods cannot effectively reflect the neighborhood information of entities in the entity embedding representations. Second, for the task of knowledge graph completion, the close relationship between the entity neighbourhood information and relationship paths in the knowledge graph plays a crucial role, but current methods fail to fully integrate them. Finally, due to the differences in the training dataset distribution, which is significantly different from the real world, existing completion methods have poor generalization ability and cannot cope with the rapidly changing real world. To solve the above challenges, this paper proposes a knowledge graph completion model that integrates relationship paths and entity neighbourhood information (RPEN-KGC). The specific contributions of this paper are as follows:

1. The proposed RPEN-KGC model integrates entity neighbourhood information and relationship paths in the knowledge graph, and uses semantic feature information from the relationship paths to infer more potential semantic information in the semantic space.

2. The RPEN-KGC model uses a generator-adversarial network-based reasoning engine to train adversarial on the generated samples and expert samples formed by random walks, ultimately obtaining reasoning paths for the completion task and effectively improving the model's robustness to noisy data.

3. To provide more diverse expert paths to the generator-adversarial network, an unsupervised sampler is used to collect richer relationship paths as expert paths based on the feature environment of each entity.

Section 2, review existing studies, highlight contributions, identify research gaps, and in section 3, explain framework addressing gaps with models, diagrams, and algorithms. In section 4, present experiments, evaluate performance, compare results, analyze findings, and in section 5, conclude findings, discuss implications, limitations, and suggest future work.

2 Related Work

To solve the data sparsity issue in knowledge graphs of artificial intelligence of social things, knowledge graph completion uses the existing information to fill in missing facts. This study summarizes previous studies and classifies the known approaches for completing knowledge graphs into three groups: those that rely on translation, those that rely on semantic matching, and those that rely on relationship paths.

2.1 Approaches Relying on Translation

By encoding things into low-dimensional embeddings and the relationships between them as translation vectors, translation-based algorithms provide great performance with low complexity in knowledge graph of artificial intelligence of social things. To determine whether a triplet is accurate, these methods usually use distance-based scoring and establish a translation scoring system based on dependencies. As the foundation of translation models, the TransE model [13] can outperform most conventional methods while maintaining a reasonable level of efficiency by translating entities and their relationships into a continuous low-dimensional vector space. In this space, the tail entity is thought of as the product of translating the head entity by the relationship vector. Nevertheless, when it comes to modeling complicated relationships, the TransE model

falls short due to its oversimplification, which restricts it to 1-to-1 relationships. One example is the TransH model [14], which uses an extra mapping vector for every relationship. This allows it to translate things by assigning them to hyperplanes of the relationship. Convolutional neural networks (CNNs) are used to build the discriminator because of their great performance in semantic feature extraction. This helps to improve the semantic distinction between the expert paths and the generated paths. Semantic feature extraction is carried out after the generated paths and expert paths are packaged as path bundles for every entity pair. The PKICLA model determines the probability score for the validity of the candidate triplets by extracting semantic features from several routes using CNN and Bi-LSTM. By first building entity and relationship embeddings in their own vector spaces, the TransR model [15] can then translate these into their respective relationship spaces. To improve computational performance, the TransD model [16] replaces matrix multiplication with vector multiplication, embeds two vectors for each entity, and assigns a new relationship mapping matrix to each entity, all of which are advancements over TransR. In general, approaches based on translation, which depend on internal structural information, are straightforward, efficient, and frequently utilized as a base for expanding models. These methods are employed to solve knowledge graph completion challenges.

2.2 Techniques Reinforced by Semantic Matching

The conventional approach to knowledge prediction using semantic matching involves calculating a scoring function that estimates the probability of new facts emerging from existing data by measuring the possible semantic similarity of entities and relationships in vector space. Binary relationship data is intrinsically structured, as the RESCAL model [17] illustrates via a third-order tensor decomposition. An entity's possible semantics are represented by a vector, and the interactions between these vectors are simulated using a relation matrix. By substituting a diagonal matrix for the dense one in RESCAL, the DistMult model drastically lowers the model's computing parameters. While the DistMult model is capable of handling asymmetric relations in the knowledge graph, this is because the product operator on real numbers has the property of being balanced. Recognized as the pioneer in handling symmetric and antisymmetric relations using complex-valued embeddings in complex space, the ComplEx model [18] represents

each embedding with two vectors, one for real and one for imaginary components. When it comes to representing entities and relationships, traditional semantic matching-based models have limitations due to their exclusive focus on the intrinsic feature information of the triplet and their disregard of the rich According to research on semantic matching approaches, neural networks can accurately model the semantic links between individual items and intelligently capture their semantic properties, which in turn improves the reasoning accuracy of the model. A two-dimensional convolutional network is employed by the ConvE model [19] to anticipate new information in the knowledge graph of artificial intelligence of social things by combining local spatial aspects of entities and relationships from various viewpoints. The main difference between ConvE and the ConvKB model [20], which was presented later, is that the latter employs a 1D convolutional network to extract global associations along the same dimension of the input triplet matrix, rather than a 2D one. This allows ConvKB to represent each triplet as a 3D matrix. To characterize the possible semantic linkages along routes of any length, the PATH-RNN model [21] applies composite functions recursively using a recurrent neural network., generating corresponding path vectors to infer missing relations in the knowledge graph of artificial intelligence of social things. The Single-Model model, based on PATH-RNN, combines relationships, entities, and entity types to effectively improve the model's reasoning performance.

2.3 Relationship Path-Based Methods

In knowledge graphs, there are many relationship paths between entity pairs that represent their semantic relationships. One reason knowledge graph completion has progressed is because relationship path-based methods have become more popular; these pathways reflect the intricate reasoning patterns between relations in the knowledge graph of artificial intelligence of social things. Learning relationship paths for reasoning in knowledge graphs were initially introduced by the PRA model. It takes a pair of entities as input, uses a classifier to find out if the relationship you're asking about actually exists, and then builds relationship pathways between them using random walks. By including multi-path relationship reasoning and utilizing a resource allocation technique to apply weights to the many routes, the PTransE model improves the efficiency of the translation-based model's predictions. In contrast to the PRA algorithm's random walk technique, the DeepPath model searches

for reasoning paths in the knowledge graph using reinforcement learning, enabling it to regulate the nature of the search. The MINERVA model [22] is very similar to DeepPath in that it uses LSTM networks to define relationship paths and uses path validity as the only measure of success for rewards. By encoding the historical information of the searched paths in an RNN and mapping it into the reinforcement learning state and action space, the M-Walk model [23] improves reasoning accuracy through inferring paths and generating more positive rewards through an improved Monte Carlo Tree Search (MCTS). The DIVINE model can adaptively adjust the reward function to optimize model performance based on different environments. These models only consider single-path reasoning and ignore the semantic similarity between multiple paths. Given the scenario described above, an emergent entity's current neighbors are taken into consideration as an additional kind of input for inductive models. Hamaguchi et al. [24] suggest using a Graph Neural Network (GNN) on the KG, which creates an embedding of a new item by combining all of its neighbors that are known to exist. Hao et al. [25] the neighbors' differences are ignored by their model, which aggregates them using basic pooling techniques.

The APCM model [26] introduces an attention mechanism into relationship paths, assigning different weights to each generated path based on its importance by comparing the semantic similarity between each path and the candidate relationship. The PKICLA model uses CNN and Bi-LSTM to extract semantic features of multiple paths and calculates the probability score for the validity of the candidate triplets.

3 Proposed Method

An extensive overview of the RPEN-KGC model's architecture and training procedure is given in this section. The model's relationship path reasoning engine and expert path sampler make up the bulk of its structure, as seen in Figure 1. The reasoning engine is based on a generative adversarial network (GAN) framework. It uses input expert paths as reasoning strategies, trains the generator in the knowledge graph to generate more reliable paths as completion results. The sampler performs local and global sampling on the relationship paths between entity pairs, providing more diverse expert paths for the reasoning engine to use as reasoning evidence.

3.1 Reasoning Engine

In the RPEN-KGC model, the reasoning engine consists of a generator and a discriminator. The generator performs random walks in the knowledge graph to generate many relationship paths. These paths are then input, along with the expert paths generated by the sampler, into the discriminator. The discriminator evaluates the semantic similarity between the generated paths and the expert paths as reasoning evidence, providing feedback to the generator for parameter adjustment. Once the discriminator and generator have undergone enough adversarial training, reliable paths that match the expert path distribution can be found, which can then be used for knowledge graph of artificial intelligence of social things completion.
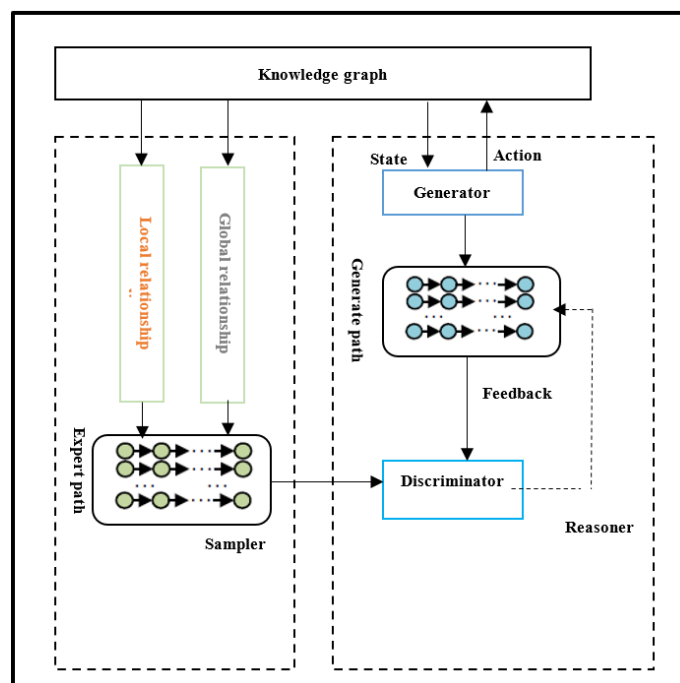

Figure 1. Overall framework of the RPEN-KGC model.

To find reliable paths that match the expert paths in the semantic space, the generator is constructed using a policy-based agent in reinforcement learning. Since convolutional neural networks (CNNs) have high performance in semantic feature extraction, the discriminator is constructed using a CNN, which helps better distinguish between the generated paths and the expert paths semantically. The generated paths and expert paths are packaged for each entity pair in the form of path bundles, and semantic feature extraction is performed. Figure 2 provides a basic schematic diagram of the path bundle embedding representation.
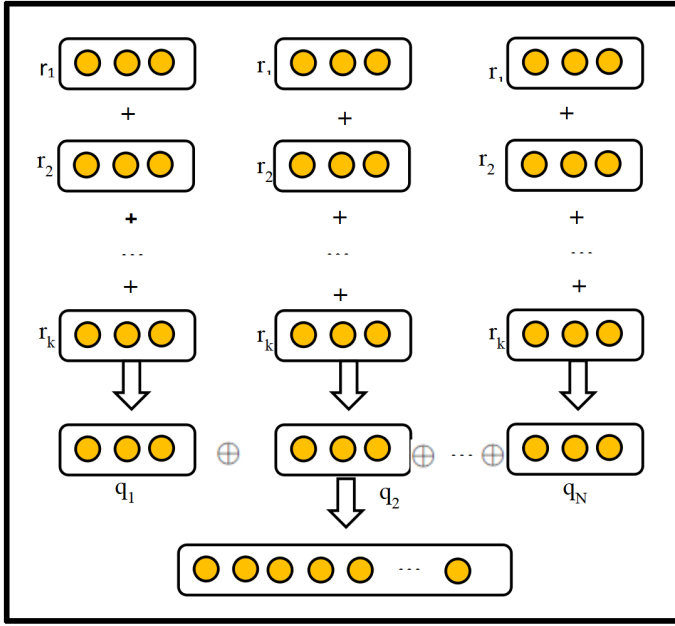
Figure 2. Schematic diagram of path bag embedding representation.

The path $q = \{r_1, r_2, \cdots, r_k, \cdots\}$ is embedded as:

$$q = \sum_{r_k \in q} r_k \tag{1}$$

Each path is composed of **k** relations, and each relation is embedded using the TransE method, mapping it into a vector $r_k \in \mathbb{R}^d$.

Each path bundle **t** contains **N** paths. By embedding and concatenating all paths in the bundle, a real-valued matrix is obtained.

$$t = q_1 \oplus q_2 \oplus \cdots \oplus q_N \tag{2}$$

where, $\oplus$ represents the concatenation operator for the path bundle $t \in \mathbb{R}^{Nd}$.

After encoding the path bundle, it is input into a discriminator based on a convolutional neural network to extract its semantic features. First, local features are extracted through a convolutional layer activated by ReLU non-linear activation.

$$f = \text{ReLU}\left(\text{Conv}(t, \omega) + b_f\right) \tag{3}$$

where, $\omega \in \mathbb{R}^{h \times l}$ is the convolution kernel, and b_f is the bias coefficient.

Then, semantic features are further extracted through two fully connected layers:

$$g = W_2 \text{ReLU}(\mathbf{Z}_1 f) \tag{4}$$

For simplicity, the corresponding bias terms are omitted in the equations. $\mathbf{Z}_1$ and $\mathbf{Z}_2$ represent the linear transformation matrices of two different fully connected layers. The output of Equation (4) is normalized using the sigmoid function, while other layers employ the ReLU function for non-linear transformation.

The probability of generating a path as an expert path is calculated using the sigmoid function, and the knowledge graph of artificial intelligence of social things is completed based on the probability results.

3.2 Sampler

Local Relation Sampling: The knowledge graph uses a bidirectional breadth-first search strategy to discover the shortest path between each entity pair. The correlations between entity pairs are more directly reflected in shorter paths. Although longer paths may hold hidden value, they are more likely to introduce noise, which can reduce the quality of expert paths. By performing local relation sampling for each entity pair, an initial expert path set ΩE is obtained.

Global Relation Sampling: Static sampling techniques are easy to use, but they produce fewer expert paths and ignore each entity's local context in the artificial intelligence of social objects knowledge network. As a result, this research suggests a sampling strategy based on entity neighborhood similarity. For each query triple, the method extracts its relation r and the entity pairs directly connected to it in the knowledge graph as the sampling set. Using the distribution characteristics of word vectors in the semantic space (i.e., the more similar the attribute information or descriptions of entities, the closer they are in the embedding space), the similarity between the entity pairs in the given triple and those in the sampling set is measured via vector dot product. A higher result indicates greater similarity between the entities. To better transfer relation r to other entity pairs, entity pairs with high similarity scores for both head and tail entities are selected for subsequent path exploration.

For the relation $r_i \in R$ in a given triple, all entities e connected to r_i form the set E_j, and the tail entities among them form the set U_j. In the knowledge graph of artificial intelligence of social things, through random walk, if any entity h_j in the set E_j reaches $u_j \in U_j$ after several steps, the path is recorded as τ, and all τ obtained from the relationship r_i are taken as its characteristic path set Γ_{r_i}. $|\Gamma_{r_i}|$ represents the number of characteristic paths in Γ_{r_i}. After the path exploration

is completed, the symbol $\tau_j \in \Gamma_{r_i}(j = 1, 2, \cdots, |\Gamma_{r_i}|)$ represents the j characteristic path obtained by the relationship r_i. In order to ensure the quality of the expert path, the number of random walk steps is set to 3. The possibility of the head entity h_j reaching u_j through each path τ_j is:

$$m_j = \sum_{e \in \text{range}(\tau_j)} q(h_j, e; \tau'_j) Q(u|e; r'_i) \quad (5)$$

where, τ'_j is the sub-part of the path τ_j without the tail relation r'_i, and $Q(u|e; r'_i)$ represents the probability that e reaches u through the relation:

$$Q(u|e; r_i) = d(E_j^c \cdot e) \frac{r'_i(e, u_j)}{|r'_i(e, \cdot)|} \quad (6)$$

where, $d(E_j^c \cdot e)$ represents the degree of association between entities obtained through cosine similarity and is used in the result calculation. Cosine similarity is used to determine the degree of link between items, and this information is used to calculate the result. The dynamic sampling model's parameters are trained in this study using a logistic regression approach, which also maps the computation. Using the sampler and generator, sampling and learning are carried out for every entity pair throughout the training phase. Some baseline models find this difficult, requiring manual hyperparameter modifications to fit various datasets. Creating appropriate hyperparameters by hand might be challenging. This paper employs a logistic regression algorithm to train the parameters of the dynamic sampling model and maps the computation results to the range $(0, 1]$ using the sigmoid function.

$$Y = \frac{1}{1 + e^{-m_j}} \quad (7)$$

where, Y represents the probability of successfully transferring the relation r_i from a given triple to other entity pairs. A higher Y value indicates a greater likelihood of successful transfer.

3.3 Training

During the training phase, for each entity pair, sampling and learning are performed using the sampler and generator. Based on whether the agent can reach the target entity along the currently collected paths, corresponding expert path sets Ω_E and generated path sets Ω_{G0} are created. The collected paths are then grouped into different path bundles: expert path bundles $\{u_e | u_e \subseteq \Omega_E\}$ and generated path bundles $\{u_g | u_g \subseteq \Omega_G\}$.

For each path bundle pair $\langle u_e, u_g \rangle$, the discriminator aims to clearly distinguish between u_e and u_g. To make the adversarial training process more stable and effective, the RPEN-KGC model updates the discriminator using the loss function proposed in the WGAN-GP model.

$$A_C = \mathbb{N}[D(u_g)] - \mathbb{N}[D(u_e)] \quad (8)$$

$$A_P = \lambda \mathbb{N}\left((\|\nabla_{\tilde{u}} D(\tilde{u})\|_2 - 1)^2\right) \quad (9)$$

$$A_D = A_C + A_P \quad (10)$$

where A_C, A_P, and A_D represent the critic function, gradient penalty function, and discriminator loss function, respectively. λ is the gradient penalty coefficient, and $\tilde{u}$ is uniformly sampled between u_g and u_e. Based on the feedback from the discriminator, the reward $\mathfrak{R}_G$ is calculated as:

$$\mathfrak{R}_G = H_g \max\{\mathbb{N}[D(u_g)] - \mathbb{N}[D(u_n)], 0\} \quad (11)$$

$$H_g = \begin{cases} 1, & u_g \subseteq \Omega_G^+ \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

where, u_n represents noise embeddings composed of continuous uniformly distributed random noise, H_g denotes the characteristic function indicating the validity of the bundle u_g, and Ω_G^+ represents the set of all valid generated paths. By rewarding valid paths with higher expectations compared to noise embedding u_n, poor-quality paths are filtered out, improving the efficiency of convergence during training.

Finally, the loss function of the discriminator is optimized using the mini-batch stochastic gradient descent algorithm, while the generator is updated using the Adam algorithm [27].

4 Experiments and Analysis

4.1 Databases and Assessment Criteria

In this study, two commonly used datasets for knowledge graph of artificial intelligence of social things completion experiments: NELL-995 [28] and FB15K-237 [29], for link prediction. Detailed information is shown in Table 1. The link prediction task's NELL-995 and FB15K-237 datasets demonstrate that RPEN-KGC significantly outperforms baseline methods in the majority of metrics. These results demonstrate that RPEN-KGC is capable to accurately predicting knowledge graphs' missing data.

Table 1. Dataset statistics.

Datasets	#entities	#relations	#Train Set	#Valid Set	#Test Set	#Tasks
NELL-995	75492	200	154213	5000	5000	12
FB15K-237	14541	237	272115	17535	20466	20

- NELL-995: This dataset is derived from the NELL system [30] after the 995th iteration. It selects triples corresponding to the top 200 most frequently occurring relations.

- FB15K-237: This is a subset of FB15K [31], where many reversible relation triples are removed. This adjustment was made because models relying on reversible relation principles in certain triples can achieve optimal results, making it difficult to distinguish the true performance of different models.

To evaluate the performance of the model, this study adopts commonly used metrics in the task of knowledge graph completion: MAP (mean average precision), MRR (mean reciprocal rank), and Hits@k.

- Hits@k: Indicates the probability that the correct triple is ranked among the top k. A higher value suggests better performance of the knowledge graph completion method.

- MRR (Mean Reciprocal Rank): Represents the average reciprocal rank of the correct triples. A higher value indicates better model performance.

- MAP (Mean Average Precision): Refers to the mean of the prediction rankings for head or tail entities. A higher value reflects better model performance. The definitions are as follows:

$$MAP = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{|u_i|} \sum_{y \in u} \frac{|\{\text{rank}(y') \leq \text{rank}(y); y' \in u_i\}|}{\text{rank}(y)} \tag{13}$$

In this case, the ranks of positive samples are represented by rank(y) while those of negative samples are represented by rank(y). Here, u_i stands for the test set and m for the overall sample size.

4.2 Comparison Methods and Parameter Settings

The model is tested against multiple baseline models to ensure its efficacy:

- Translation-based models: TransE [32], TransR [33].

- Semantic matching-based models: DistMult [34], ConvE [35], Single-Model [36], and DRGI [37].

- Relation path-based models: PTransE [38], DeepPath [39], JGAN [40], APCM [41], and PKICLA [42].

In the experiments, every path bundle u has N=5 paths, and the embedding dimension is set at 200. In this convolutional neural network, the parameters 3×N for the convolution kernel, λ for the gradient penalty, 1024 for the hidden layer, and a size for the output layer that corresponds to the path embedding dimension are all set. L2 regularization is used to avoid overfitting.

During testing, invalid triples are scored alongside

Table 2. Dataset statistics.

Dataset Model	NELL-995				FB15K-237		
	MAP	MRR	Hits@1	Hits@3	MAP	MRR	Hits@1
TransE	0.737	0.715	0.608	0.702	0.532	0.286	0.190
TransR	0.789	0.727	0.631	0.714	0.54	0.315	0.216
DistMult	0.649	0.860	0.752	0.865	0.534	0.241	0.155
Single-Model	0.827	0.833	0.765	0.903	0.525	0.512	0.496
ConvE	0.812	0.862	0.826	0.919	0.536	0.509	0.430
DRGI	0.823	0.848	0.821	0.923	0.575	0.567	0.492
PTransE	0.793	0.838	0.723	0.793	0.547	0.381	0.361
DeepPath	0.796	0.852	0.808	0.884	0.572	0.495	0.449
APCM	0.838	0.847	0.783	0.905	0.558	0.556	0.513
JGAN	0.853	0.851	0.779	0.874	0.561	0.564	0.517
PKICLA	0.846	0.856	0.829	0.941	0.564	0.589	0.528
RPEN-KGC	0.887	0.859	0.827	0.953	0.659	0.686	0.531

valid triples at a ratio of approximately 1:10, with invalid triples generated by randomly replacing entities.

4.3 Results and Analysis

Table 2 displays the link prediction experiment outcomes in comparison to baseline models. Compared to NELL-995, RPEN-KGC performs better than the majority of baseline models, especially on larger datasets like FB15K-237, which has a more complex distribution of relation types. Specifically:

- NELL-995: The proposed model outperforms translation-based models significantly. While semantic matching-based models and relation path-based models score higher on NELL-995, their performance drops on FB15K-237, likely due to the interference from reversible relations. On NELL-995, RPEN-KGC is slightly behind some models in MRR and Hits@1 metrics. However, it improves over the best-performing models by 0.034 and 0.012 in MAP and Hits@3, respectively.

- FB15K-237: RPEN-KGC achieves the best results across all metrics. Compared to the best-performing fusion-based models, it improves MAP, MRR, Hits@1, and Hits@3 by at least 0.084, 0.097, 0.003, and 0.009, respectively.

Comparison Insights:

- On NELL-995, compared to ConvE and DistMult, RPEN-KGC shows a slight drop in MRR (by 0.003 and 0.001, respectively). This is because NELL-995 contains fewer paths for certain entity pairs, whereas RPEN-KGC requires a large number of paths for expert and generated paths. DistMult uses a bilinear scoring function for triples, and ConvE employs 2D convolution to process entity and relation embeddings, effectively extracting entity similarity features and enhancing performance.

- Compared to PKICLA, RPEN-KGC's Hits@1 is slightly lower because PKICLA not only extracts local features of relation paths but also uses an attention mechanism to assign weights to multiple paths, extracting global features and thus achieving prominent performance.

- An entity neighbourhood similarity comparison mechanism is also included in the model, which offers the reasoner a large number of trustworthy expert pathways and successfully facilitates semantic interaction between relation paths

and entity neighbourhood data. Experiments on two publicly accessible knowledge graph completion datasets show that the RPEN-KGC model performs better than baseline methods on the majority of metrics.

On FB15K-237, RPEN-KGC consistently achieves the best performance across all evaluation metrics. This is due to two key reasons:

1. Compared to other models, RPEN-KGC fully extracts relational path features through adversarial training between expert and generated paths, while also leveraging entity similarity to improve the success rate of expert path mining and capturing entity feature information.

2. FB15K-237's complex relation types make it challenging for some baseline models, which require manual hyper parameter adjustments to fit different datasets. Designing suitable hyper parameters manually can be complex. However, RPEN-KGC automatically optimizes parameters through generative adversarial networks and uses a sampler to unsupervised extract many reliable expert paths from the knowledge graph of artificial intelligence of social things.

As a result, RPEN-KGC demonstrates more stable performance and better real-world applicability compared to other models.

Relationship paths and entity neighbourhood data are combined for knowledge graph fused using a novel method called RPEN-KGC (Relationship Path and Entity Neighbourhood Knowledge Graph Completion). The sampler provides the inferencer with reliable inference methods by randomly walking between entity pairs. To increase sampling efficiency and support inference methodologies, the sampler steers random walks using a contrastive method based on entity neighborhood similarity. The inferencer infers a wider range of relationship paths within the semantic domain and extracts semantic properties of relationship paths. RPEN-KGC greatly improves most metrics compared to baseline methods, according to experiments conducted on the publicly available NELL-995 and FB15K-237 datasets for the link prediction task. These results show that RPEN-KGC effectively predicts knowledge graphs with missing information.

4.4 Ablation Study

To validate the effectiveness of the model's components, an ablation study was conducted by removing certain components of the model and retraining on the FB15K-237 and NELL-995 datasets. This quantifies the contributions of each component. The three model variants in this study are as follows:

1. Removal of the Global Relation Sampling Module: The model uses only the local relation sampling method to obtain expert paths during training.

2. Removal of the Local Relation Sampling Module: The model uses only the global relation sampling method to obtain expert paths during training.

3. Removal of the Sampler Module: The model conducts experiments using only the reinforcement learning approach.

On the NELL-995 and FB15K-237 datasets, the MAP metric is used to evaluate the link prediction performance of the model. The experimental results are shown in Table 3.

Table 3. Ablation experiment results.

Model	MAP	
	NELL-995	FB15K-237
Model in this paper	0.887	0.659
w/o local relation sampling	0.817	0.651
w/o global relation sampling	0.802	0.643
w/o sampler	0.796	0.572

After removing the local relation sampling module, global relation sampling module, and sampler module, the MAP metric of the model shows varying degrees of decline.

- After removing the local relation sampling module, the model only uses global relation sampling to obtain multi-hop paths between entities as expert paths, ignoring the most direct path relationships between entities. The MAP score is close to but lower than that of the original model.

- After removing the global relation sampling module, the model cannot dynamically adjust hyper parameters and lacks attention to entity environmental features. It only uses a small number of single-hop paths between entities as expert paths, resulting in a MAP score significantly lower than the original model.

- After removing the sampler module, the model

degenerates to DeepPath, and it performs the worst compared to the other ablation models.

The ablation study results demonstrate the rationality of each component in RPEN-KGC.

In order to develop reasoning paths for the completion task and hence increase the model's robustness to noisy data, the RPEN-KGC model employs a generator-adversarial network-based reasoning engine to train adversarial on the generated samples and expert samples created by random walks.

5 Conclusion

This paper proposes a knowledge graph completion method, RPEN-KGC, that integrates relation paths and entity domain information. By using a generative adversarial network, it captures many relation paths from the knowledge graph of intelligence of social things. Additionally, by extracting semantic features from these relation paths, it generates more diverse reasoning paths in the semantic space. The model also incorporates an entity neighbourhood similarity comparison mechanism, providing many reliable expert paths for the reasoner and effectively enabling semantic interaction between relation paths and entity neighbourhood information.

Experiments on two publicly available datasets for knowledge graph completion demonstrate that the RPEN-KGC model outperforms baseline models on most metrics. In future research, this method will be enhanced by incorporating multimodal information to further improve the effectiveness of knowledge graph completion.

Data Availability Statement

Data will be made available on request.

Funding

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Khelloufi, A., Ning, H., Naouri, A., Sada, A.B., Qammar, A., Khalil, A., Mao, L., & Dhelim, S.

(2024). A multimodal latent-features-based service recommendation system for the social Internet of Things. *IEEE Transactions on Computational Social Systems*, 11(4), 5388-5403. [CrossRef]

[2] Li, Q., Zhang, Y., & Qiao, G. (2022). GAFM: A Knowledge Graph Completion Method Based on Graph Attention Faded Mechanism. *Information Processing & Management*, 59(5), 103004. [CrossRef]

[3] Xue, X., Yu, X., Zhou, D., Peng, C., Wang, X., Liu, D., & Wang, F.Y. (2023). Computational experiments for complex social systems—Part III: the docking of domain models. *IEEE Transactions on Computational Social Systems*, 11(2), 1766-1780. [CrossRef]

[4] Wan, Y., Jiang, N., & Liu, Z. (2024). Dynamic Hypergraph Modeling and Robustness Analysis for SIoT. *CMES - Computer Modeling in Engineering and Sciences*, 140(3), 3017-3034. [CrossRef]

[5] Gao, H., Xu, K., Cao, M., Xiao, J., Xu, Q., & Yin, Y. (2021). The deep features and attention mechanism-based method to dish healthcare under social IoT systems: An empirical study with a hand-deep local–global net. *IEEE Transactions on Computational Social Systems*, 9(1), 336-347. [CrossRef]

[6] Dong, H., Wang, P., Xiao, M., Ning, Z., Wang, P., & Zhou, Y. (2024). Temporal inductive path neural network for temporal knowledge graph reasoning. *Artificial Intelligence*, 329, 104085. [CrossRef]

[7] Farhadi, B., Rahmani, A.M., Asghari, P., & Hosseinzadeh, M. (2021). Friendship selection and management in social internet of things: A systematic review. *Computer Networks*, 201, 108568. [CrossRef]

[8] Yang, J., Yang, L.T., Wang, H., Gao, Y., Zhao, Y., Xie, X., & Lu, Y. (2023). Representation learning for knowledge fusion and reasoning in Cyber–Physical–Social Systems: Survey and perspectives. *Information Fusion*, 90, 59-73. [CrossRef]

[9] Li, F., Ma, G., Chen, F., Lyu, Q., Wang, Z., & Zhang, J. (2024). Enhanced enterprise-student matching with meta-path based graph neural network. *Journal of King Saud University - Computer and Information Sciences*, 36(6), 102116. [CrossRef]

[10] Singh, I., Agarwal, K., & Ganapathy, S. (2024). Advancing allergy source mapping: A comprehensive multidisciplinary framework integrating machine learning, graph theory and game theory. *Applied Soft Computing*, 166, 112147. [CrossRef]

[11] Termos, M., Ghalmane, Z., Brahmia, M.A., Fadlallah, A., Jaber, A., & Zghal, M. (2024). GDLC: A new Graph Deep Learning framework based on centrality measures for intrusion detection in IoT networks. *Internet of Things*, 26, 101214. [CrossRef]

[12] Wang, X., Wang, Y., Yang, J., Jia, X., Li, L., Ding, W., & Wang, F.Y. (2024). The survey on multi-source data fusion in cyber-physical-social systems: Foundational infrastructure for industrial metaverses and industries

5.0. *Information Fusion*, 107, 102321. [CrossRef]

[13] Adhikari, D., Jiang, W., Zhan, J., Rawat, D.B., & Bhattarai, A. (2024). Recent advances in anomaly detection in Internet of Things: Status, challenges, and perspectives. *Computer Science Review*, 54, 100665. [CrossRef]

[14] Priyadarshini, K., Sikkandar, M.Y., AlDuraywish, A., & Alqahtani, T.M. (2024). Integrating relational and sequential information for enhanced detection of autoimmune disorders with relational Neural Networks and Long Short-Term Memory networks. *Biomedical Signal Processing and Control*, 96(Part A), 106495. [CrossRef]

[15] Zhang, Y., Li, Y., & Zhang, F. (2024). Multi-level urban street representation with street-view imagery and hybrid semantic graph. *ISPRS Journal of Photogrammetry and Remote Sensing*, 218(Part B), 19-32. [CrossRef]

[16] Zavarella, V., Consoli, S., Recupero, D.R., Fenu, G., Angioni, S., Buscaldi, D., Dessí, D., & Osborne, F. (2024). Triplétoile: Extraction of knowledge from microblogging text. *Heliyon*, 10(12), e32479. [CrossRef]

[17] Ammar, K., Inoubli, W., Zghal, S., Borji, A., & Nguifo, E.M. (2023). Trans-Trip: Translation-based embedding with Triplets for Heterogeneous Graphs. *Procedia Computer Science*, 225, 1104-1113. [CrossRef]

[18] Tian, L., Zhou, X., Wu, Y.P., Zhou, W.T., Zhang, J.H., & Zhang, T.S. (2022). Knowledge graph and knowledge reasoning: A systematic review. *Journal of Electronic Science and Technology*, 20(2), 100159. [CrossRef]

[19] Basole, R.C., Park, H., & Seuss, C.D. (2024). Complex business ecosystem intelligence using AI-powered visual analytics. *Decision Support Systems*, 178, 114133. [CrossRef]

[20] Das, R., & Soylu, M. (2023). A key review on graph data science: The power of graphs in scientific studies. *Chemometrics and Intelligent Laboratory Systems*, 240, 104896. [CrossRef]

[21] Roopa, M.S., Pattar, S., Buyya, R., Venugopal, K.R., Iyengar, S.S., & Patnaik, L.M. (2019). Social Internet of Things (SIoT): Foundations, thrust areas, systematic review and future directions. *Computer Communications*, 139, 32-57. [CrossRef]

[22] Villanueva-Merino, A., Urra-Uriarte, S., Izkara, J.L., Campos-Cordobes, S., Aranguren, A., & Molina-Costa, P. (2024). Leveraging Local Digital Twins for planning age-friendly urban environments. *Cities*, 155, 105458. [CrossRef]

[23] Cheng, D., Dong, C., He, W., Chen, Z., Liu, X., & Zhang, H. (2023). A fine-grained detection method for gate-level hardware Trojan based on bidirectional Graph Neural Networks. *Journal of King Saud University - Computer and Information Sciences*, 35(10), 101822. [CrossRef]

[24] Hamaguchi, T., Oiwa, H., Shimbo, M., & Matsumoto, Y.

(2017). Knowledge transfer for out-of-knowledge-base entities: A graph neural network approach. *arXiv preprint arXiv:1706.05674*.

[25] Hao, Y., Zhang, Y., Liu, K., He, S., Liu, Z., Wu, H., & Zhao, J. (2017). An end-to-end model for question answering over knowledge base with cross-attention combining global knowledge. In *ACL*, volume 1, 221–231. [CrossRef]

[26] Nguyen, H.L., Vu, D.T., & Jung, J.J. (2020). Knowledge graph fusion for smart systems: A Survey. *Information Fusion*, 61, 56-70. [CrossRef]

[27] Tiddi, I., Schlobach, S. (2022). Knowledge graphs as tools for explainable machine learning: A survey. *Artificial Intelligence*, 302, 103627. [CrossRef]

[28] Cherif, A., Ammar, H., Kalkatawi, M., Alshehri, S., & Imine, A. (2024). Encoder–decoder graph neural network for credit card fraud detection. *Journal of King Saud University - Computer and Information Sciences*, 36(3), 102003. [CrossRef]

[29] Guo, Z., Liu, Q., & Zou, B. (2022). Research on knowledge reasoning of TCM based on knowledge graphs. *Digital Chinese Medicine*, 5(4), 386-393. [CrossRef]

[30] Bhardwaj, K.K., Banyal, S., Sharma, D.K., & Al-Numay, W. (2022). Internet of things based smart city design using fog computing and fuzzy logic. *Sustainable Cities and Society*, 79, 103712. [CrossRef]

[31] Boy, G.A. (2023). An epistemological approach to human systems integration. *Technology in Society*, 74, 102298. [CrossRef]

[32] Wu, T., You, X., Xian, X., Pu, X., Qiao, S., & Wang, C. (2024). Towards deep understanding of graph convolutional networks for relation extraction. *Data & Knowledge Engineering*, 149, 102265. [CrossRef]

[33] Khlifi, M.K., Boulila, W., & Farah, I.R. (2023). Graph-based deep learning techniques for remote sensing applications: Techniques, taxonomy, and applications — A comprehensive review. *Computer Science Review*, 50, 100596. [CrossRef]

[34] Ju, W., Fang, Z., Gu, Y., Liu, Z., Long, Q., Qiao, Z., Qin, Y., Shen, J., Sun, F., Xiao, Z., Yang, J., Yuan, J., Zhao, Y., Wang, Y., Luo, X., & Zhang, M. (2024). A Comprehensive Survey on Deep Graph Representation Learning. *Neural Networks*, 173, 106207. [CrossRef]

[35] Yuan, B., Liu, L., & Antonopoulos, N. (2018). Efficient service discovery in decentralized online social networks. *Future Generation Computer Systems*, 86, 775-791. [CrossRef]

[36] Jia, Y., Wang, J., Shou, W., Hosseini, M.R., & Bai, Y. (2023). Graph neural networks for construction applications. *Automation in Construction*, 154, 104984. [CrossRef]

[37] Sun, R. (2024). Chapter 1 - History of graph computing and graph databases. In *The Essential Criteria of Graph Databases* (pp. 1-32). Elsevier. [CrossRef]

[38] Wani, N.A., Kumar, R., Mamta, Bedi, J., & Rida, I. (2024). Explainable AI-driven IoMT fusion: Unravelling techniques, opportunities, and challenges with Explainable AI in healthcare. *Information Fusion*, 110, 102472. [CrossRef]

[39] Mahboubi, A., Luong, K., Aboutorab, H., Bui, H.T., Jarrad, G., Bahutair, M., Camtepe, S., Pogrebna, G., Ahmed, E., Barry, B., & Gately, H. (2024). Evolving techniques in cyber threat hunting: A systematic review. *Journal of Network and Computer Applications*, 232, 104004. [CrossRef]

[40] Wei, S., Meng, S., Li, Q., Zhou, X., Qi, L., & Xu, X. (2023). Edge-enabled federated sequential recommendation with knowledge-aware Transformer. *Future Generation Computer Systems*, 148, 610-622. [CrossRef]

[41] Iovanella, A. (2024). Exploiting network science in business process management: A conceptual framework. *Chaos, Solitons & Fractals*, 178, 114344. [CrossRef]

[42] Scaffidi, F. (2024). Average social and territorial innovation impacts of industrial heritage regeneration. *Cities*, 148, 104907. [CrossRef]

Mohammad Shabaz is working as an Assistant Professor with the Model Institute of Engineering and Technology, Jammu, India. He has published over 200+ research papers in various journals indexed in Scopus/Web of Science, five Indian Patents, and three Australian Patents. His area of interest is application of computer science in interdisciplinary domains. (Email: shabaz.cse@mietjammu.in)

Mukesh Soni is working as an Assistant Professor with the Division of Research and Development, Lovely Professional University, Phagwara, India. He is also a member of many international and national professional bodies like IEEE, Asia Society of Researcher, Scientific and Technical Research Association, International Association of Engineers Institute for Engineering Research and Publication, and Scholars Academic and Scientific Society. (Email: soni.mukesh15@gmail.com)