Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

FS-KEN: Few-shot Knowledge Graph Reasoning by Adversarial Negative Enhancing

Lingyuan Meng[1] , **Ke Liang**[1†] , **Zeyu Zhu**[1] , **Xinwang Liu**[1†] and **Wenpeng Lu**[2]
[1]National University of Defense Technology
[2]Shandong Computer Science Center(National Supercomputer Center in Jinan)
{mly_edu, zhuzeyu_nudt}@163.com, liangke200694@126.com, xinwangliu@nudt.edu.cn, luwpeng@163.com

Abstract

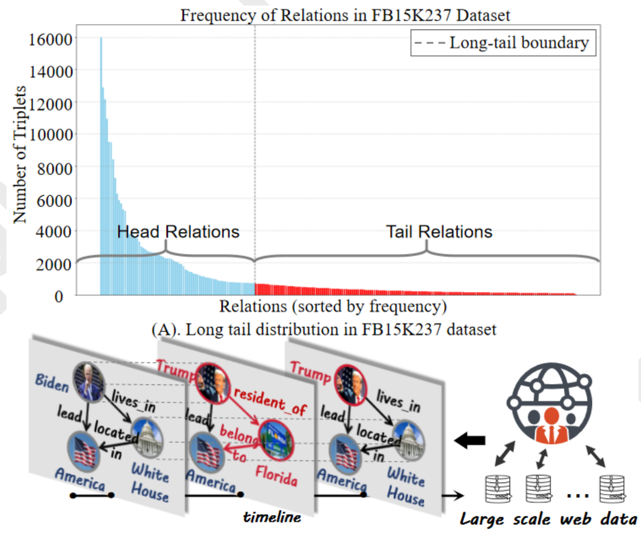
Few-shot knowledge graph reasoning (FS-KGR) aims to infer missing facts in knowledge graphs using limited data (such as only 3/5 samples). Existing strategies have shown good performance by mining more supervised information for few-shot learning through meta-learning and self-supervised learning. However, the problem of insufficient samples has not been fundamentally solved. In this paper, we propose a novel algorithm based on adversarial learning for **E**nhancing **N**egative samples in few-shot scenarios of **FS**-**K**GR, termed FS-KEN. Specifically, we are the first to use GAN to conduct data augmentation on FS-KGR scenario. FS-KEN uses policy gradient GANs for negative sample augmentation, solving the gradient back-propagation issue in traditional GANs. The generator aims to produce high-quality negative entities. while the objective of the discriminator is to distinguish between generated entities and real entities. Comprehensive experiments conducted on two few-shot knowledge graph completion datasets reveal that FS-KEN surpasses other baseline models, achieving state-of-the-art results.

1 Introduction

Knowledge graphs (KGs) structure extensive, multi-relational real-world human knowledge using heterogeneous graph formats [Liang *et al.*, 2024; Zhang *et al.*, 2023; Chen *et al.*, 2024]. Such information is stored as triplets *(head entity, relation, tail entity)*, where $relation$ signifies relations. KGs are instrumental in numerous applications, including recommendation systems [Wu *et al.*, 2024b; Wu *et al.*, 2024a], medical recognition [Jiang *et al.*, 2023], financial analysis [Viswanathan and Singh, 2023], and question answering [Saxena *et al.*, 2020].

However, due to the complexity and diversity of relationships in the human world, KGs have the inherent property of being incomplete [Wu *et al.*, 2023; Chen *et al.*, 2023; Chen *et al.*, 2025]. To improve the quality and usability of KGs, it is necessary to complete the missing facts based on

† Corresponding Authors.



(A). Long tail distribution in FB15K237 dataset

(B). Example of a knowledge graph change and update process

Figure 1: Illustration of the long-tail distribution of KGs.

existing data. This task is termed knowledge graph reasoning (KGR) [Luo *et al.*, 2025; Li *et al.*, 2025]. Although various methods have achieved good performance in solving KGR tasks, the performance of these methods is heavily dependent on sufficient labeled data. However, real-world knowledge graph data often has a long-tail distribution, that is, a few relations contain sufficient triples, while most relations contain too few triples. As shown in Figure 1.A, in the FB15K237 dataset, the number of triples corresponding to more than 60% of the relations only accounts for 10% of the total number of triples. In addition, the knowledge graph is dynamically changing and constantly updated. As shown in Figure 1.b, new entities and relations may be continuously added to the knowledge graph over time, and the sample size is small. When the data distribution is imbalanced, traditional methods will produce performance deviation, overfitting the head class data and ignoring the learning of the tail class, which will hinder performance improvement. Therefore, investigating the application of few-shot learning techniques in knowledge graph reasoning holds substantial practical value and relevance in contemporary research.

Current approaches of FS-KGR are predominantly divided into two main groups, *i.e.*, techniques based on meta-learning and those utilizing self-supervised learning. Specifically, meta-learning-based approaches like GMatching [Xiong *et al.*, 2018] and FSRL [Zhang *et al.*, 2020] seek to develop differentiable metrics that expand the distance between support sets and negative query triplets while reducing the proximity from positive query triplets. Furthermore, Att-FMetric [Sun *et al.*, 2021] incorporates attention-based neighborhood aggregation and path encoding to predict long-tail relationships, while MetaR [Chen *et al.*, 2019] trains a meta-learner with limited instances to optimize the task-specific learner. However, these approaches rely heavily on manually crafted meta-training tasks, incurring significant human labor costs. As a more effective learning paradigm, self-supervised pretraining has demonstrated powerful learning with few-shot samples in other domains [Li *et al.*, 2024; Li *et al.*, 2023] but has been less applied in FS-KGR tasks. The most representative method is CSR [Huang *et al.*, 2022], which learns structural information through subgraph mining, achieving satisfactory performance. Although both types of methods have achieved satisfactory performance, they essentially mine richer supervised information from a small amount of labeled data, and do not fundamentally solve the class imbalance problem caused by a long-tailed distribution. In addition, although the self-supervised-based paradigm does not require the construction of meta-training tasks, it still requires the manual construction of meta-testing tasks, which is time-consuming and laborious in real-world application scenarios.

To address these challenges of insufficient sample data, various data augmentation strategies have been proposed in other fields such as computer vision [Kim and Hwang, 2022] and natural language processing [Liao *et al.*, 2022], among which the Generative Adversarial Network (GAN) has been proven to be an effective method. However, to our best knowledge, GAN has not been extended to FS-KGR tasks. To fill this gap, we introduce GAN to increase the number of samples in FS-KGR tasks. However, triplets reflect real-world facts in KGs, thus augmenting data for a specific relation involves fabricating entities and facts, which rules out direct positive sample enhancement within the knowledge graph. To tackle this challenge, we focus on the negative augmentation, which is crucial to the hinge loss in most KGR models. Therefore, we proposes a FS-KGR method based on adversarial learning for negative sampling. Specifically, the generator and discriminator aims to generate high-quality negative entities jointly. Furthermore, to address the discrete sampling issue, we combine policy gradient theory with GAN. Finally, we integrated FS-KEN into the latest baselines to verify its effectiveness on two datasets. Specifically, the main contributions of our paper include three aspects:

- We propose a novel FS-KGR method, termed FS-KNE, which leverages adversarial learning for negative augmentation. To the best of our knowledge, we are the first to address the class imbalance issue arising from long-tail distributions through data augmentation within the FS-KGR scenario.
- To overcome the limitation of traditional GANs that cannot backpropagate gradients to the generator during discrete sampling, we introduce a reinforcement learning mechanism, designing a policy gradient GAN for discrete data production, ensuring the creation of high-quality negative entities.
- Extensive experiments on two FS-KGR datasets demonstrate the superior performance of FS-KEN in the task of FS-KGR task.

2 Related Work

2.1 Few-shot Knowledge Graph Reasoning

Meta-learning Based Knowledge Graph Reasoning

As an efficient learning paradigm, the meta-learning paradigm achieve knowledge transfer to new tasks through carefully designed meta-tasks. To tackle the issue of limited data availability in few-shot learning scenarios, a significant number of FS-KGR approaches have embraced the meta-learning framework as their primary methodology. As the first model of FS-KGR, GMatching [Xiong *et al.*, 2018]based on metric learning, which obtains embedding representations through neighborhood structures and uses an LSTM to match the embeddings with the target, subsequently producing similarity metrics that quantify the resemblance between the query triplet and the reference dataset. FSRL [Zhang *et al.*, 2020] demonstrates proficiency in extracting insights from diverse graph configurations and synthesizing the feature representations of few-shot instances, assigning different weights to neighborhood information through a heterogeneous neighbor decoder. FAAN [Sheng *et al.*, 2020] incorporates an advanced approach to analyze the evolving characteristics of both entities and relations. By employing an attention-based framework, it effectively identifies and adapts to the shifting attributes that vary with distinct operational contexts. Furthermore, MetaR [Chen *et al.*, 2019] enhances entity embeddings using neighborhood information by adopting a meta-learning framework, which includes gradient meta-learning and relation meta-learning, to perform FKGR tasks, effectively improving the performance. However, the performance of meta-learning-based methods relies heavily on hand-designed meta-tasks. The construction process of meta-tasks is time-consuming and laborious, which hinders further generalization of the model.

Self-supervised Based Knowledge Graph Reasoning

In recent years, self-supervised learning has attracted significant attention in various fields such as computer vision [Dosovitskiy *et al.*, 2020], natural language processing [Brown *et al.*, 2020], graph learning [Yu *et al.*, 2024; Liu *et al.*, 2024], and knowledge graph representation learning [Meng *et al.*, 2024b]. Inspired by their success, some FS-KGR methods have been designed powerful self-supervised pre-training objectives without the need to construct meta-tasks, significantly improving the performance of FS-KGR tasks. Concretely, CSR [Huang *et al.*, 2022] is based on the theory of hypothetical induction, which represents the few-shot relations through the maximal common subgraph, and ultimately transforms the few-sample learning reasoning into an inductive relational reasoning task. SARF [Meng *et al.*, 2024a] proposed an aliasing relation-assisted mechanism,

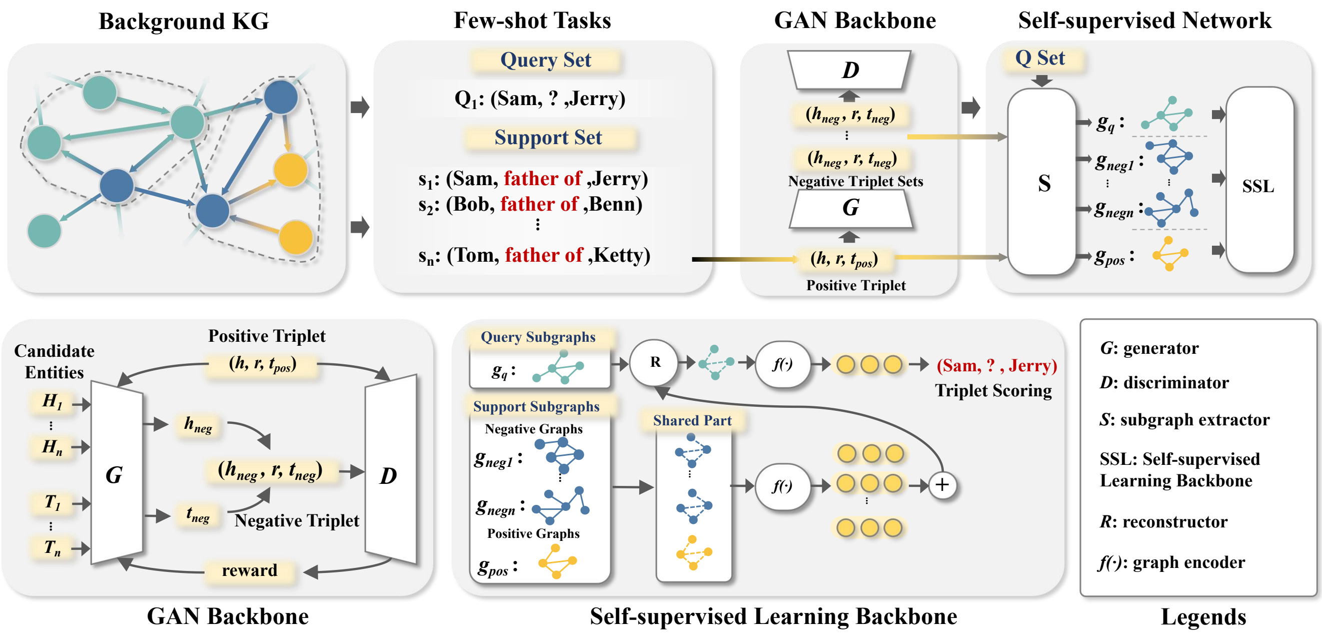


Figure 2: An illustration of the FS-KEN. The FS-KEN framework includes two modules: a GAN backbone network and a self-supervised learning network.

which mines aliasing relations from head relations that are semantically similar to the target long-tail relations. Although these approaches have achieved satisfactory performance due to the mining of more supervision information, the class imbalance problem caused by the long-tail distribution has not been fully solved.

2.2 Few-shot Data Augmentation Strategy

The few-shot data augmentation strategy aims to solve the class imbalance challenge caused by the long-tail distribution and provide sufficient and high-quality samples for training. A large number of methods have been proposed recently. Specifically, H-SMOTE [Chao and Zhang, 2023] proposed a new paradigm based on the reasonable and stable data augmentation. ALP [Kim *et al.*, 2022] proposes a novel text augmentation method that increases the diversity of sentence structure and word choice in sentences while preserving semantic content. CDE-GAN [Chen *et al.*, 2021] utilized generative adversarial networks to try to generate high-quality negative samples for novelty detection. Although these strategies are considered effective, It is difficult to perform data augmentation for the KGs since the data in the KGs is factual.

3 Methodology

3.1 Preliminaries

A knowledge graph can be conceptualized as a directed network structure consisting of three fundamental components: a collection of entities denoted as $E = \{e_1, e_2, ..., e_n\}$, a set of relationships represented by $R = \{r_1, r_2, ..., r_m\}$, and a series of triplets T. Each triplet within this framework is structured as an ordered tuple (h, r, t), where h and t are elements of E, and r belongs to R, signifying the head entity, tail entity, and their relations, respectively. For a incompleteness triplet $t_i \in T'$, where $t_i = (h, r, ?)$ or $t_i = (?, r, t)$ or $t_i = (h, ?, t)$, which has a missing entity/relation, knowledge graph reasoning aims to infer the missing entity/relation based on existing facts T.

Compared with the conventional knowledge graph reasoning, few-shot knowledge graph reasoning refers to the task of reasoning over a knowledge graph where the number of labeled triplets associated with a particular relation is limited. In such scenarios, the goal is to leverage the limited number of labeled examples to generalize to other unseen triplets involving the same or similar relations. Formally, let $S = \{(h_k, t_k) \mid (h_k, r, t_k) \in T, r \in R\}$ denote a set of entity pairs (h_k, t_k) associated with a specific relation r within the triplet set T, the goal of the few-shot knowledge graph completion is to predict triplets for the under-sampled relation r.

3.2 FS-KGR Framework

The FS-KGR based on adversarial learning for negative sampling is shown in Figure 2. Specifically, the FS-KEN algorithm consists of two modules, *i.e.*, a GAN backbone network and a self-supervised learning network. The GAN backbone network aims to generate high quality negative samples for the training process. Specifically, it takes instances in the support set as positive samples and generates high-quality negative samples through a policy gradient GAN. The self-supervised learning network try to learn the representation of few-shot relations. Concretely, it inputs positive and negative samples to extract subgraphs and then reconstructs the maximum common subgraph in these subgraphs using a generative self-supervised method. The detailed process of each module of FS-KEN is introduced below.

GAN Backbone Network

In our FS-KEN framework, both the generator and discriminator are built upon knowledge graph embedding (KGE) models, utilizing softmax probabilities to represent the likelihood of each entity. Specifically, the generator is designed to produce high-quality negative entities by leveraging feedback from the discriminator. These rewards guide the generator in generating negative samples that align more closely with the underlying structure of the knowledge graph. Conversely, the discriminator is tasked with encoding the graph, integrating both true and generated negative entities, and evaluating the quality of the generated samples.

The generator aims to minimize the discrepancy between the negative entities it generates and the true entities in the knowledge graph to ensure that the generated negative samples resemble realistic entities, rather than easily distinguishable false ones. Meanwhile, the goal of the discriminator is to minimize the hinge loss between the target triplets and the generated negative triplets, which can ensure the discriminator can effectively differentiate between true and negative samples, thereby motivating the generator to produce negative samples that are increasingly difficult to distinguish from true ones. During the training process, the generator and discriminator undergo alternating training.

For a positive triplet $T_p \in T$, we first define the probability distribution of the generator to produce a negative triplet as $p_G(T_n|T_p)$, where T_n denotes the negative triplet. After that, a set of negative triplets $\{T_n\}$ will be generated by sampling. Furthermore, the scoring function for the discriminator is denoted by $f_D(T)$. We employ the knowledge graph embedding model as $f_D(T)$ to learn embeddings from few-shot data. Finally, the discriminator tries to minimize the margin-based loss function, which is defined below:

$$\mathcal{L}_D = \sum_{(h,r,t)\in T} \left[f_D(T_p) - f_D(T_n) + \gamma\right]_+ \quad T_n \sim p_G(T_n \mid T_p), \tag{1}$$

where $[\cdot]^+$ denotes the hinge loss function, while γ is a hyperparameter representing a fixed margin. The triplet T_n represents a negative triplet, generated by replacing one of the head or tail entities of a positive triplet with an entity randomly selected from the KG. Formally, $T_n \in \{(h', r, t)|h' \in E\} \cup \{(h, r, t')|t' \in E\}$.

The goal of the generator is to optimize the anticipated value of the negative distance below:

$$\mathcal{L}_G = \sum_{(h,r,t)\in \mathcal{T}} \mathbb{E}\left[-f_D(T_n)\right] \tag{2}$$

$$(T_n) \sim p_G(T_n \mid T_p). \tag{3}$$

However, this process involves discrete sampling steps, making it difficult to compute gradients through direct differentiation. Drawing inspiration from reinforcement learning, which selects optimal actions from a discrete action space according to a policy, we conceptualize the Generative Adversarial Network (GAN) as an agent. In this framework, the entity sampling process is modeled as the agent's behavior of sampling from the discrete action space, effectively addressing the challenges associated with gradient computation.

Specifically, a policy gradient [Sutton *et al.*, 1999] generative adversarial network (PG-GAN) is employed to derive the gradient of $\mathcal{L}_G$ with respect to the generator parameters:

$$\begin{aligned}\nabla_G \mathcal{L}_G = &\sum_{(h,r,t)\in T} \mathbb{E}_{(T_n)\sim p_G(T_n|T_p)} \\ &\left[-f_s(T_n)\nabla_G \log p_G(T_n \mid T_p)\right] \\ \simeq &\sum_{T_p\in T} \frac{1}{N} \sum_{(T_n^i)\sim \pi_G(T_n|T_p), i=1\ldots N} \\ &\left[-f_s(T_n)\nabla_G \log p_G(T_n \mid T_p)\right].\end{aligned} \tag{4}$$

Where the approximate sign indicates that the expectation has been feasibly approximated through sampling. Therefore, the gradient of $\mathcal{L}_G$ can be calculated, and optimization can be performed using gradient-based methods.

For the PG-GAN framework, we further explain the connotation by analogy with RL. In the context of RL terminology, the generator is analogous to the agent, while the discriminator corresponds to the environment. The set of positive triples T_p defines the state space, and the sampling probability $\pi_G(T_n|T_p)$ represents the policy. The set of negative triples T_n constitutes the action space, and the final score assigned to the negative sample serves as the reward. In our model, the generator interacts with the environment, taking actions and evolving to maximize the cumulative reward. Note that in contrast to the conventional RL framework where an agent may take multiple actions within a single episode, our agent takes only one action in each episode. This distinction arises because the agent's action does not influence the state, thereby limiting the interaction to a single action at each step.

Self-Supervised Learning Network

The self-supervised learning framework is designed to identify the subgraph connection patterns associated with specific few-shot relations for the purpose of few-shot relation reasoning. Note that the subgraph connection pattern refers to the largest common subgraph within a set of neighborhood-closed subgraphs, which are sampled based on triplets containing the same relation. Specifically, the self-supervised learning framework consists of two main components, *i.e.*, the enclosing subgraph construction module and the self-supervised backbone network.

Enclosing Subgraph Construction Module: Following the existing work [Teru *et al.*, 2020], we hypothesize that crucial evidence for the target relation is located in the connections between the two target entities. Numerous prior studies have extracted closed subgraphs around triplets to facilitate relational reasoning [Wang *et al.*, 2021]. In our paper, we adopt the GraIL-based model [Teru *et al.*, 2020] to sample the enclosing subgraphs. Concretely, entities of each triplet within the support and query sets are denoted as N_h and N_t. Subsequently, a closed subgraph is formed by intersecting these neighbor sets, after which any isolated nodes that are more than k hops away from either target entity are removed.

Self-Supervised Backbone Network: The self-supervised backbone network learns the maximal common subgraph features for few-shot relations within the framework of a generative self-supervised paradigm, which comprises three distinct

Dataset	$\#\mathcal{E}$	$\#\mathcal{R}$	$\#\mathcal{T}$	Tasks
NELL-One	68544	291	181109	11
FB15K-237	14543	200	268039	30

Table 1: Statistics of all two datasets

stages. Firstly, the module applies a soft edge mask in order to extract the maximal common subgraph from all subgraphs in the support set as the representative subgraph of the target few-shot relation. After that, the representative subgraph is encoded into the vector space over the graph encoder $f_e(\cdot)$. Finally, the features of the representative subgraph are decoded and reconstructed by integrating them with the query subgraph. Consequently, the overarching objective of the network is to train a reconstructor R_G that ensures the reconstructed representative subgraph closely resembles the original subgraph before encoding. The objective function for this process is formalized as follows:

$$\mathcal{L}_r = \mathrm{CE}\left(\mathbf{M}, R_G\left(G, f_e(G, \mathbf{M})\right)\right), \tag{5}$$

where $\mathbf{M} : [0, 1]^{\mathbb{R}}$ represents the mask matrix of the inductive hypothesis subgraph, C_R represents the graph reconstructor, which has the same form as the graph encoder $f_e(\cdot)$, G represents the target subgraph. We expects the distance between the mask matrix of the hypothesis subgraph before and after reconstruction to be minimized in the vector space.

To enable collaborative training with f_e, we introduce a contrastive loss function, denoted as $\mathcal{L}_e$:

$$\mathcal{L}_e = \max\Bigg[\frac{\sum_{\mathrm{j}=1}^{\mathrm{n}}\left(\mathbf{e}_{\mathrm{j}} \times \mathbf{e}_{\mathrm{j}}^{\mathrm{p}}\right)}{\sqrt{\sum_{\mathrm{j}=1}^{\mathrm{n}}\mathbf{e}_{\mathrm{j}}^{2}} \times \sqrt{\sum_{\mathrm{j}=1}^{\mathrm{n}}\mathbf{e}_{\mathrm{j}}^{\mathrm{p}2}}} - \tag{6}$$

$$\frac{\sum_{j=1}^{n}\left(\mathbf{e}_j \times \mathbf{e}_j^{n}\right)}{\sqrt{\sum_{j=1}^{n}\mathbf{e}_j^{2}} \times \sqrt{\sum_{j=1}^{n}\mathbf{e}_j^{n2}}} + \gamma\Bigg]. \tag{7}$$

Here, $\mathbf{e}_j = f_e(G, \mathbf{M})$ represents the embedding of the graph G masked by $\mathbf{M}$, while $\mathbf{e}^p$ and $\mathbf{e}^n$ correspond to the embeddings of the positive and negative graphs, respectively.

Therefore, combining the discriminator loss from Section 3.2.1, the overall optimization objective of FS-KEN is:

$$\mathcal{L} = \lambda_1 \cdot \mathcal{L}_r + \lambda_2 \cdot \mathcal{L}_e, \tag{8}$$

where λ_1 and λ_2 are hyper-parameters. Note that in our target function, two hyper-parameters are utilized due to the different dimensions of the loss terms on different datasets. Furthermore, the proposed generative adversarial negative sample enhancement strategy is a plug-and-play module, and the choice of the self-supervised backbone network is robust.

4 Experiments and Discussion

This chapter elaborates on the experimental framework through three critical dimensions: datasets, assessment criteria, implementation details. Subsequently, an extensive sequence of empirical investigations is undertaken to substantiate the efficacy of the newly developed FS-KEN framework by addressing the following research questions.

- **Q1: Superiority.** Does FS-KEN demonstrate superior performance when integrated with existing few-shot knowledge graph reasoning models?
- **Q2: Effectiveness.** Does the proposed model exhibit acceptable training efficiency? Is FS-KEN robust to variations in its components? How do these individual modules contribute to the overall performance?
- **Q3: Sensitivity.** How sensitive is the performance of FS-KEN to hyper-parameters, especially the graph encoder loss weight (λ_1) and the reconstructor weight (λ_2)?
- **Q4: Case Study.** Does the proposed FS-KEN generate high-quality negative triplets during the experimental evaluation?

In addressing the aforementioned questions, we conduct a range of comprehensive experiments. Specifically, answers of **Q1** to **Q4** are provided in Section 4.2 to Section 4.4.

4.1 Experiment Settings

Datasets and Evaluation Metrics

In order to show the effectiveness of the proposed FS-KEN, we assessed the FS-KEN on two real-world few-shot datasets, *i.e.*, NELL-One [Mitchell *et al.*, 2018] and FB15K-237 [Bollacker *et al.*, 2008]. For the NELL-One, the meta-evaluation and meta-test splits provided in the dataset were used for evaluating and testing few-shot tasks. We also focused on relations containing between 50 and 500 triplets as few-shot tasks. In the case of FB15K-237 [Bollacker *et al.*, 2008], a minority ratio of 7:30 was selected for the target few-shot evaluation and test tasks. Table 1 presents the statistical details of all three datasets.

The performance of the models is evaluated using traditional ranking metrics, namely mean reciprocal rank (MRR) and Hits@h. In this paper, the value of h are set to 1, 5, and 10. Moreover, each test triplet is compared with 50 possible negative triplets.

Implementation Details

The FS-KEN experiments were primarily executed using the PyTorch [Paszke *et al.*, 2019] library and were performed on a single NVIDIA GeForce 3090Ti. Furthermore, the few-shot instance count K was configured to 3. In practice, the methods presented in we can be generalized to any K. For the step of generating closed subgraphs, we generate 2-hop subgraphs in NELL-One and 1-hop subgraphs in FB15K-237. We employed AdamW with a learning rate of 1×10^{-5}. Moreover, the training epochs of model were set to 5000, and the training batch size was set to 8. Various knowledge graph embedding models were employed for the generator to validate its robustness (refer to Table 3).

4.2 Performance Comparison (RQ1)

To highlight the effectiveness of the model in the FS-KGR task, this chapter compares the proposed model with 13 baseline models on two datasets. These include 7 traditional KGR models, 3 meta-learning-based FS-KGR models, and 3 generative self-supervised FS-KGR methods. As shown in Table 2, FS-KEN outperforms all other models across all metrics. For instance, compared to the state-of-the-art baseline

Methods	NELL-One				FB15K-237			
	MRR	Hits@1	Hits@5	Hits@10	MRR	Hits@1	Hits@5	Hits@10
Tranditional KGR Models								
TransE	0.118	0.061	0.132	0.223	0.120	0.570	0.137	0.238
DistMult	0.134	0.083	0.143	0.233	0.094	0.053	0.101	0.172
ComplEx	0.124	0.077	0.134	0.213	0.104	0.058	0.114	0.188
RotatE	0.112	0.060	0.131	0.209	0.115	0.069	0.131	0.200
R-GCN	0.199	0.141	0.219	0.307	0.140	0.082	0.154	0.255
MEAN	0.180	0.124	0.189	0.296	0.114	0.058	0.119	0.217
LAN	0.172	0.116	0.181	0.286	0.112	0.055	0.119	0.218
Meta learning-based FS-KGR Models								
GMatching	0.322	0.225	0.432	0.510	-	-	-	-
FIRE	0.273	0.225	0.364	0.497	0.478	0.413	0.502	0.577
FSRL	0.490	0.327	0.695	0.853	0.684	0.573	0.817	0.902
MetaR	0.471	0.322	0.647	0.763	**0.805**	**0.740**	**0.881**	**0.937**
Self Supervised-based FS-KGR Models								
CSR-OPT	0.463	0.321	0.629	0.760	0.619	0.512	0.747	0.824
CSR-GNN	0.560	0.435	0.703	0.821	0.678	0.612	0.746	0.792
SARF	0.626	**0.493**	0.797	0.875	0.753	0.688	0.814	0.884
FS-KEN	**0.631**↑	0.490	**0.812**↑	**0.887**↑	0.773	0.699	0.856	0.915

Table 2: Performance comparison of our FS-KEN with other state-of-the-art models on the few-shot knowledge graph link prediction tasks.

Datasets	Methods	MRR	Hit@1	Hit@5	Hit@10
FB15K-237	FS-KEN	**0.773**	**0.699**	**0.856**	**0.915**
	FS-KEN w.o. GAN	0.678	0.612	0.746	0.792
	FS-KEN w.o. SSL	0.635	0.571	0.657	0.693
	FS-KEN w.o. GAN+SSL	0.478	0.413	0.502	0.577
NELL-One	FS-KEN	**0.631**	**0.490**	**0.812**	**0.887**
	FS-KEN w.o. GAN	0.560	0.435	0.703	0.821
	FS-KEN w.o. SSL	0.528	0.419	0.657	0.788
	FS-KEN w.o. GAN+SSL	0.273	0.225	0.367	0.497

Table 3: Ablation study of two main components of FS-KEN.

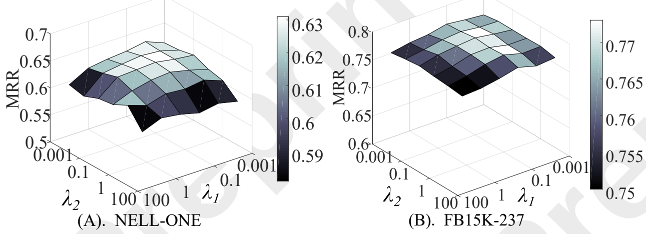


Figure 3: Sensitive analysis result of hyper-parameters λ_1 and λ_2.

model SARF [Meng *et al.*, 2024a], FS-KEN improves the testing MRR by 0.7% and Hits@5 by 1.8% on the NELL-ONE dataset. On the FB15K-237 dataset, FS-KEN further enhances Hits@5 and Hits@10 by 5.16% and 3.5%, respectively. The experimental results indicate that FS-KGR provides high-quality negative samples for model training, effectively enhancing the model's reasoning ability. Specifically, the model achieves greater performance improvement on the FB15K-237 dataset compared to NELL-ONE. This is because FB15K-237 contains a larger set of entities, providing a broader sampling space. Additionally, the entities in the FB15K-237 exhibit higher homogeneity, *i.e.*, the direct neighboring entities of an entity are closely related, making it easier to sample high-quality negative entity samples.

4.3 Ablation Study (RQ2)

To demonstrate the robustness of the model to different generators in the GAN backbone, we verified and analyzed the performance using three different KGE models as generators on the NELL-ONE dataset, namely: TransE [Bordes *et al.*, 2013], TransD [Ji *et al.*, 2015], and DistMult [Yang *et al.*, 2014]. For the discriminator, we selected TransE. The ablation study results are shown in Table 5.

The results demonstrate that, on both the NELL-ONE and FB15K-237 datasets, all three generators achieved outstanding performance, showing the robustness of the FS-KEN model in selecting GAN generators. Specifically, on the NELL-ONE dataset, the Hit@1 metric for FS-KEN under the three generators was 62.9%, 62.4%, and 63.1%, with a performance variance of only 1.12%. On the FB15K-237 dataset, the Hit@1 metric for FS-KEN under the three generators was 76.2%, 77.3%, and 76.5%, with a performance variance of 1.44%. Our future research will explore the application of a broader range of knowledge graph embedding models.

Furthermore, we conducted further ablation experiments to demonstrate the effectiveness of the two main modules of FS-KEN on two datasets, *i.e.*, GAN backbone and the self-supervised learning module. Note that in order to demonstrate the effectiveness of self-supervised learning, we replaced it with the meta-learning method FIRE. The results in Table 3 show that both of the GAN backbone and the self-supervised network are effective in our model.

4.4 Efficiency Study (RQ2)

Considering that the introduction of GANs provides additional negative samples during training, which may increase computational cost, we further measure the inference time of our method compared to three other FS-KGR baselines, *i.e.*, FSRL, CSR-GNN, and SARF. As shown in Figure 4, we observe that the inference runtime of our FS-KEN is reduced compared to the state-of-the-art model SARF. Additionally,

Target Triplet	Low-quality Entities	Similarity Score	High-quality Entities	Similarity Score
(person:Brandon, plays_sport, sport:baseball)	disease:diabetes	0.17	sport:golf	0.44
	artery:vessels	0.21	sport:football	0.78
	clothing:pants	0.15	sport:basketball	0.93
(person:Mario, person_born_in, city:york)	clothing:white	0.07	city:petoskey	0.47
	river:chena	0.24	city:riyadh	0.52
	city:number	0.35	county:york_city	0.63

Table 4: Case study of FS-KEN on real-world toy cases derived from NELL-One.

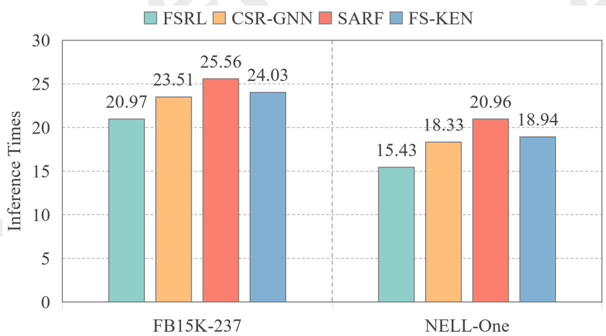


Figure 4: Comparison of inference time of our FS-KEN with other state-of-the-art baselines.

	Methods	MRR	Hit@1	Hit@5	Hit@10
FB15K-237	TransE	0.762	0.706	0.831	0.898
	TransD	**0.773**	**0.699**	**0.856**	**0.915**
	DistMult	0.765	0.695	0.823	0.887
NELL-One	TransE	0.629	0.495	0.81	0.877
	TransD	0.624	0.481	0.804	0.883
	DistMult	**0.631**	**0.490**	**0.812**	**0.887**

Table 5: Ablation study on FB15k-237 of different generators.

compared to CSR-GNN, the inference time of FS-KEN on the FB15K-237 dataset is only 0.5 seconds longer per epoch, which remains acceptable. Since the GAN backbone is not involved in the inference process, the introduction of GANs does not significantly affect the model's inference time. Overall, FS-KEN demonstrates both acceptable time consumption and significant performance improvements.

4.5 Sensitivity Analysis (RQ3)

We analyze the sensitivity of the reconstruction loss weight λ_1 and the discriminator loss weight λ_2 in the FS-KEN algorithm on the MRR metric for both datasets. Specifically, We perform a grid search on both hyper-parameters and report the optimal hyperparameter combination. Both values of the two parameters are chosen from $\{0.01, 0.1, 1, 10, 100\}$, which as depicted in Figure 3. The experimental results show that when λ_1 and λ_2 both take the value of 0.1, the model achieves the best performance on NELL-One dataset. For the FB15k-237 dataset, our model achieves best performance when $\lambda_1 = 1$ and $\lambda_2 = 0.1$. Furthermore, the experiment results demonstrated that our proposed FS-KEN is robust to the hyper-parameters.

4.6 Case Study (RQ4)

In this section, we show several cases from NELL-One dataset in order to qualitatively show the effectiveness of the proposed negative augmentation strategy, which is shown in Table 4. Specifically, for the target triplet, we sample negative tail entities over the traditional random sampling mechanism and the proposed method as low-quality entities and high-quality entities, respectively. Finally, we calculate the cosine similarity score of sampled negative entities and the tail entity of the target triplet. Table 4 shows that FS-KEN can generate high-quality negative samples that are closer to the target triples in the implicit vector space. Taking the target triple (person:Brandon, plays_sport, sport:baseball) as an example, the negative entities generated by the random sampling strategy may be disease, artery, and clothing, which have much smaller semantic similarity with the tail entities of the target triple than those generated by FS-KEN.

5 Conclusion

In this paper, we proposed a novel adversarial learning strategy for few-shot knowledge graph reasoning, termed FS-KEN. Specifically, FS-KEN uses policy gradient GAN for negative sample enhancement on the data in the support set, solving the problem of traditional GANs in preventing gradient back-propagation to the generator during discrete sampling to obtain high-quality and high-quantity negative samples. The generator aims to generate high-quality negative entities, while the discriminator aims to distinguish between generated entities and real entities. To show the superiority of our FS-KEN, various experiments on two few-shot benchmark datasets are conducted. The results indicate that FS-KEN generally outperformed other methods.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (no. 62325604, 62441618, 62276271).

References

[Bollacker *et al.*, 2008] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 1247–1250, 2008.

[Bordes *et al.*, 2013] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana

Yakhnenko. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26, 2013.

[Brown *et al.*, 2020] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[Chao and Zhang, 2023] Xuewei Chao and Lixin Zhang. Few-shot imbalanced classification based on data augmentation. *Multimedia Systems*, 29(5):2843–2851, 2023.

[Chen *et al.*, 2019] Mingyang Chen, Wen Zhang, Wei Zhang, Qiang Chen, and Huajun Chen. Meta relational learning for few-shot link prediction in knowledge graphs. *arXiv preprint arXiv:1909.01515*, 2019.

[Chen *et al.*, 2021] Shiming Chen, Wenjie Wang, Beihao Xia, Xinge You, Qinmu Peng, Zehong Cao, and Weiping Ding. Cde-gan: Cooperative dual evolution-based generative adversarial network. *IEEE Transactions on Evolutionary Computation*, 25(5):986–1000, 2021.

[Chen *et al.*, 2023] Man-Sheng Chen, Chang-Dong Wang, and Jian-Huang Lai. Low-rank tensor based proximity learning for multi-view clustering. *IEEE Trans. Knowl. Data Eng.*, 35(5):5076–5090, 2023.

[Chen *et al.*, 2024] Man-Sheng Chen, Xi-Ran Zhu, Jia-Qi Lin, and Chang-Dong Wang. Contrastive multiview attribute graph clustering with adaptive encoders. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–12, 2024.

[Chen *et al.*, 2025] Man-Sheng Chen, Jia-Qi Lin, Chang-Dong Wang, Dong Huang, and Jian-Huang Lai. Contrastive ensemble clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 2025.

[Dosovitskiy *et al.*, 2020] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[Huang *et al.*, 2022] Qian Huang, Hongyu Ren, and Jure Leskovec. Few-shot relational reasoning via connection subgraph pretraining. *arXiv preprint arXiv:2210.06722*, 2022.

[Ji *et al.*, 2015] Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. 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 (volume 1: Long papers)*, pages 687–696, 2015.

[Jiang *et al.*, 2023] Pengcheng Jiang, Cao Xiao, Adam Cross, and Jimeng Sun. Graphcare: Enhancing healthcare predictions with personalized knowledge graphs. *arXiv preprint arXiv:2305.12788*, 2023.

[Kim and Hwang, 2022] Jun-Hyung Kim and Youngbae Hwang. Gan-based synthetic data augmentation for infrared small target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–12, 2022.

[Kim *et al.*, 2022] Hazel H Kim, Daecheol Woo, Seong Joon Oh, Jeong-Won Cha, and Yo-Sub Han. Alp: Data augmentation using lexicalized pcfgs for few-shot text classification. In *Proceedings of the aaai conference on artificial intelligence*, volume 36, pages 10894–10902, 2022.

[Li *et al.*, 2023] Qian Li, Shu Guo, Jia Wu, Jianxin Li, Jiawei Sheng, Hao Peng, and Lihong Wang. Event extraction by associating event types and argument roles. *IEEE Transactions on Big Data*, 9(6):1549–1560, 2023.

[Li *et al.*, 2024] Qian Li, Jianxin Li, Jia Wu, Xutan Peng, Cheng Ji, Hao Peng, Lihong Wang, and Philip S Yu. Triplet-aware graph neural networks for factorized multimodal knowledge graph entity alignment. *Neural Networks*, 179:106479, 2024.

[Li *et al.*, 2025] Yuhan Li, Xinni Zhang, Linhao Luo, Heng Chang, Yuxiang Ren, Irwin King, and Jia Li. G-refer: Graph retrieval-augmented large language model for explainable recommendation. In *Proceedings of the ACM on Web Conference 2025*, pages 240–251, 2025.

[Liang *et al.*, 2024] Ke Liang, Lingyuan Meng, Meng Liu, Yue Liu, Wenxuan Tu, Siwei Wang, Sihang Zhou, Xinwang Liu, Fuchun Sun, and Kunlun He. A survey of knowledge graph reasoning on graph types: Static, dynamic, and multi-modal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):9456–9478, 2024.

[Liao *et al.*, 2022] Wentong Liao, Kai Hu, Michael Ying Yang, and Bodo Rosenhahn. Text to image generation with semantic-spatial aware gan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 18187–18196, 2022.

[Liu *et al.*, 2024] Meng Liu, Ke Liang, Yawei Zhao, Wenxuan Tu, Sihang Zhou, Xinbiao Gan, Xinwang Liu, and He Kunlun. Self-supervised temporal graph learning with temporal and structural intensity alignment. *IEEE Transactions on Neural Networks and Learning Systems*, 2024.

[Luo *et al.*, 2025] Linhao Luo, Zicheng Zhao, Chen Gong, Gholamreza Haffari, and Shirui Pan. Graph-constrained reasoning: Faithful reasoning on knowledge graphs with large language models. In *Forty-second International Conference on Machine Learning*, 2025.

[Meng *et al.*, 2024a] Lingyuan Meng, Ke Liang, Bin Xiao, Sihang Zhou, Yue Liu, Meng Liu, Xihong Yang, Xinwang Liu, and Jinyan Li. Sarf: Aliasing relation–assisted self-supervised learning for few-shot relation reasoning. *IEEE Transactions on Neural Networks and Learning Systems*, 2024.

[Meng *et al.*, 2024b] Lingyuan Meng, Ke Liang, Hao Yu, Yue Liu, Sihang Zhou, Meng Liu, and Xinwang Liu. Fedean: Entity-aware adversarial negative sampling for federated knowledge graph reasoning. *IEEE Transactions on Knowledge and Data Engineering*, 2024.

[Mitchell *et al.*, 2018] T. Mitchell, W. Cohen, E. Hruschka, P. Talukdar, B. Yang, J. Betteridge, A. Carlson, B. Dalvi, M. Gardner, B. Kisiel, J. Krishnamurthy, N. Lao, K. Mazaitis, T. Mohamed, N. Nakashole, E. Platanios, A. Ritter, M. Samadi, B. Settles, R. Wang, D. Wijaya, A. Gupta, X. Chen, A. Saparov, M. Greaves, and J. Welling. Never-ending learning. *Commun. ACM*, 61(5):103–115, apr 2018.

[Paszke *et al.*, 2019] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[Saxena *et al.*, 2020] Apoorv Saxena, Aditay Tripathi, and Partha Talukdar. Improving multi-hop question answering over knowledge graphs using knowledge base embeddings. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 4498–4507, 2020.

[Sheng *et al.*, 2020] Jiawei Sheng, Shu Guo, Zhenyu Chen, Juwei Yue, Lihong Wang, Tingwen Liu, and Hongbo Xu. Adaptive attentional network for few-shot knowledge graph completion. *arXiv preprint arXiv:2010.09638*, 2020.

[Sun *et al.*, 2021] Jian Sun, Yu Zhou, and Chengqing Zong. One-shot relation learning for knowledge graphs via neighborhood aggregation and paths encoding. *Transactions on Asian and Low-Resource Language Information Processing*, 21(3):1–19, 2021.

[Sutton *et al.*, 1999] Richard S Sutton, David McAllester, Satinder Singh, and Yishay Mansour. Policy gradient methods for reinforcement learning with function approximation. *Advances in neural information processing systems*, 12, 1999.

[Teru *et al.*, 2020] Komal Teru, Etienne Denis, and Will Hamilton. Inductive relation prediction by subgraph reasoning. In *International Conference on Machine Learning*, pages 9448–9457. PMLR, 2020.

[Viswanathan and Singh, 2023] Shyam Balagurumurthy Viswanathan and Gaurav Singh. Advancing financial operations: leveraging knowledge graph for innovation. *International Journal of Computer Trends and Technology*, 71(10):51–60, 2023.

[Wang *et al.*, 2021] Hongwei Wang, Hongyu Ren, and Jure Leskovec. Relational message passing for knowledge graph completion. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 1697–1707, 2021.

[Wu *et al.*, 2023] Likang Wu, Zhi Li, Hongke Zhao, Zhefeng Wang, Qi Liu, Baoxing Huai, Nicholas Jing Yuan, and Enhong Chen. Recognizing unseen objects via multimodal intensive knowledge graph propagation. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2618–2628, 2023.

[Wu *et al.*, 2024a] Likang Wu, Zhi Li, Hongke Zhao, Zhenya Huang, Yongqiang Han, Junji Jiang, and Enhong Chen. Supporting your idea reasonably: A knowledge-aware topic reasoning strategy for citation recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 2024.

[Wu *et al.*, 2024b] Likang Wu, Zhaopeng Qiu, Zhi Zheng, Hengshu Zhu, and Enhong Chen. Exploring large language model for graph data understanding in online job recommendations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 9178–9186, 2024.

[Xiong *et al.*, 2018] Wenhan Xiong, Mo Yu, Shiyu Chang, Xiaoxiao Guo, and William Yang Wang. One-shot relational learning for knowledge graphs. *arXiv preprint arXiv:1808.09040*, 2018.

[Yang *et al.*, 2014] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575*, 2014.

[Yu *et al.*, 2024] Hao Yu, Ke Liang, Dayu Hu, Wenxuan Tu, Chuan Ma, Sihang Zhou, and Xinwang Liu. Gzoo: Black-box node injection attack on graph neural networks via zeroth-order optimization. *IEEE Transactions on Knowledge and Data Engineering*, 2024.

[Zhang *et al.*, 2020] Chuxu Zhang, Huaxiu Yao, Chao Huang, Meng Jiang, Zhenhui Li, and Nitesh V Chawla. Few-shot knowledge graph completion. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 3041–3048, 2020.

[Zhang *et al.*, 2023] Lei Zhang, Wuji Zhang, Likang Wu, Ming He, and Hongke Zhao. Shgcn: Socially enhanced heterogeneous graph convolutional network for multi-behavior prediction. *ACM Transactions on the Web*, 18(1):1–27, 2023.