RESEARCH-ARTICLE

A Spectral Heterogeneous Diffusion Framework for Knowledge-aware Recommendation

XIANGMING LI, Xidian University, Xi'an, Shaanxi, China

HUA CHU, Xidian University, Xi'an, Shaanxi, China

CHENGYU FENG, Xidian University, Xi'an, Shaanxi, China

JIANAN LI, Xidian University, Xi'an, Shaanxi, China

YANGTAO ZHOU, Xidian University, Xi'an, Shaanxi, China

QINGSHAN LI, Xidian University, Xi'an, Shaanxi, China

View all

Published: 21 February 2026

WSDM '26:The Nineteenth ACM International Conference on Web Search and Data Mining
February 22 - 26, 2026
ID, Boise, USA

Conference Sponsors:
SIGKDD
SIGWEB
SIGIR
SIGMOD

WSDM '26: Proceedings of the Nineteenth ACM International Conference on Web Search and Data Mining (February 2026)
https://doi.org/10.1145/3773966.3777946
ISBN: 9798400722929

A Spectral Heterogeneous Diffusion Framework for Knowledge-aware Recommendation

Xiangming Li
Xidian University
School of Computer Science and Technology
Xi'an, Shaanxi, China
xiangmli@stu.xidian.edu.cn

Hua Chu
Xidian University
School of Computer Science and Technology
Xi'an, Shaanxi, China
hchu@mail.xidian.edu.cn

Chengyu Feng
Xidian University
School of Computer Science and Technology
Xi'an, Shaanxi, China
cyfeng@stu.xidian.edu.cn

Jianan Li
Xidian University
School of Computer Science and Technology
Xi'an, Shaanxi, China
lijianan@xidian.edu.cn

Yangtao Zhou*
Xidian University
School of Computer Science and Technology
Xi'an, Shaanxi, China
zhou_yt@stu.xidian.edu.cn

Qingshan Li
Xidian University
School of Computer Science and Technology
Xi'an, Shaanxi, China
qshli@mail.xidian.edu.cn

Wanqiang Yang
Shanghai Fairyland Software Co.,Ltd.
Shanghai, China
yangwan@fulan.com.cn

Abstract

Knowledge-aware recommendation leverages rich item-related factual information in Knowledge Graphs (KGs) to enhance recommendation systems. However, most existing methods focus on developing complex models to extract information from a given KG. They essentially follow a model-centric paradigm, overlooking data quality problems. In practice, KG data exhibits two principal quality problems, namely the **noisy knowledge problem** and the **incomplete knowledge problem**, which severely impair the performance of downstream models. To address these problems, we adopt a data-centric paradigm to improve the quality of KG data. Inspired by diffusion models' superior denoising and generation ability by fitting true data distributions, we propose a novel spectral heterogeneous diffusion framework for knowledge-aware recommendation. This framework tailors a diffusion model to capture the recommendation-oriented heterogeneous distribution in the original KG and then converts the fitted distribution into a high-quality KG. Specifically, we design a spectral heterogeneous diffusion model that integrates recommendation prior knowledge to capture task-relevant distribution and aligns its diffusion process with the features of heterogeneous graphs to model heterogeneity. Furthermore, we propose a continuous-discrete mode adapter that transforms the learned continuous distribution into a high-quality discrete KG. The resulting KG is denoised and enriched with task-relevant triples, mitigating noisy and incomplete knowledge problems. Experiments show that our plug-and-play framework can be integrated with any knowledge-aware recommendation model and boost their performance by improving KG quality. The code and theoretical analyses are available at https://github.com/xiangmli/SHGD.

CCS Concepts

• **Information systems → Recommender systems**.

Keywords

Knowledge-aware Recommendation; Data-centric Paradigm; Spectral Diffusion Model; Heterogeneity Modeling

ACM Reference Format:
Xiangming Li, Hua Chu, Chengyu Feng, Jianan Li, Yangtao Zhou, Qingshan Li, and Wanqiang Yang. 2026. A Spectral Heterogeneous Diffusion Framework for Knowledge-aware Recommendation. In *Proceedings of the Nineteenth ACM International Conference on Web Search and Data Mining (WSDM '26), February 22–26, 2026, Boise, ID, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3773966.3777946

1 Introduction

In the era of information abundance, recommendation systems have seen widespread adoption in e-commerce, online advertising, and social media to deliver personalized information services[11, 50]. Collaborative filtering is a prominent approach in recommendation systems, aiming to uncover hidden user preferences based on user feedback and user similarities [14, 18, 31, 53]. However, it relies on limited historical user-item interaction data, making it vulnerable to the data sparsity and cold-start problems [1, 13, 42, 48]. To alleviate this limitation, researchers enhance collaborative filtering by adopting knowledge graphs as prior knowledge on items. This approach is known as knowledge-aware recommendation [5, 57].

*Corresponding Author.

WSDM '26, Boise, ID, USA

ACM ISBN 979-8-4007-2292-9/2026/02
https://doi.org/10.1145/3773966.3777946

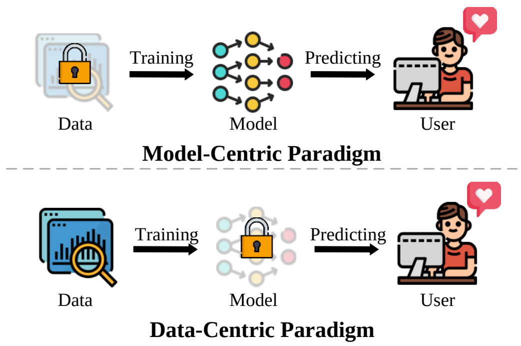


Figure 1: Differences between the model-centric paradigm and the data-centric paradigm

Existing knowledge-aware recommendation methods can be categorized into three groups: embedding-based, path-based, and propagation-based methods [16, 37, 55]. Embedding-based methods [3, 35, 47] employ graph embedding techniques to learn embeddings of entities and relations, which are then used to augment user or item representations. Path-based methods [12, 40, 45] extract semantically meaningful meta-paths from knowledge graphs and leverage them to model high-order connectivity patterns between users and items. Propagation-based methods [30, 36, 38] perform multi-hop information propagation on knowledge graphs to capture high-order information and fuse it with collaborative filtering signals, thereby improving recommendation performance. However, most existing methods prioritize developing sophisticated deep learning models to extract information from a given knowledge graph, essentially adopting a model-centric paradigm, overlooking the quality problems inherent in knowledge graph data.

In practice, knowledge graphs often exhibit quality problems in the following two aspects: (1) **Noisy knowledge problem**. Due to the scale and generality of knowledge graphs, they often contain a large amount of noisy information that is irrelevant to the recommendation task. The introduction of such noise obscures relevant information, thereby degrading the quality of both user and item representations and ultimately impairing recommendation performance [37, 43]. (2) **Incomplete knowledge problem**. Knowledge graphs usually have a large number of missing facts, and many latent relationships between entities have not been fully explored, leading to incomplete graph structure [16, 32]. This deficiency in item-related knowledge reduces the richness of item representations and leads to suboptimal recommendation results.

To address these quality problems in knowledge graphs, we draw inspiration from the data-centric paradigm [19, 44, 46] which has gained prominence in recent years. As shown in Figure 1, unlike the model-centric paradigm that focuses solely on developing advanced models, the data-centric paradigm emphasizes acquiring high-quality data, which not only elevates the upper limits of model capabilities but is also inherently applicable across various downstream models [19, 44]. In recent years, leveraging generative models to synthesize high-quality training samples has emerged as a prominent research direction within the data-centric paradigm [2, 17, 54]. Notably, diffusion models have shown exceptional denoising and generation ability by capturing true data distributions in data generation tasks such as image synthesis [8, 26] and speech synthesis [7, 15]. Building on this strength, we propose to apply diffusion models to learn the underlying distributions of knowledge graphs, thereby producing high-quality knowledge graph data.

However, two key challenges arise when using diffusion models to capture the distributions of knowledge graphs and constructing high-quality knowledge graphs:

- **How to model the recommendation-oriented heterogeneous knowledge distribution.** The conventional Gaussian diffusion process struggles to capture the task-relevant information in heterogeneous knowledge graphs. On the one hand, the abundance of knowledge that is irrelevant to the recommendation task leads the Gaussian diffusion process to predominantly fit a task-irrelevant distribution. On the other hand, the heterogeneity of the knowledge graphs makes it challenging to model their distribution by simply using homogeneous Gaussian noise.
- **How to construct a high-quality knowledge graph based on the learned distribution.** The forward and reverse processes in diffusion models are both carried out in a continuous space, so the learned distribution exhibits continuity, whereas knowledge graphs are essentially discrete. This discrepancy makes it challenging to directly utilize the distribution fitted by the diffusion model to construct a discrete knowledge graph.

To tackle these challenges, we propose a **S**pectral **H**eterogeneous **D**iffusion **F**ramework for Knowledge-aware Recommendation (**SHDF**). This framework consists of two components: a spectral heterogeneous diffusion model and a continuous-discrete mode adapter. The spectral heterogeneous diffusion model, enhanced with recommendation prior knowledge, captures the recommendation-oriented heterogeneous distribution in the knowledge graph to overcome the first challenge. The continuous-discrete mode adapter then transforms this continuous distribution into a high-quality discrete knowledge graph adaptively, addressing the second challenge. Specifically, first, the spectral heterogeneous diffusion model transforms the original knowledge graph into the spectral domain to separate information of varying frequencies. It then utilizes a series of graph filters that incorporate recommendation priors to blur high-frequency noisy components, smoothing the knowledge graph while preserving task-relevant distributions. This diffusion model further assigns differentiated smoothing velocities based on node heterogeneity, ensuring the retention of critical heterogeneous information. This meticulously designed diffusion model learns recommendation-oriented heterogeneous distribution in the original knowledge graph. Second, we introduce a continuous-discrete mode adapter that maps the continuous distribution into a discrete graph structure guided by recommendation signals. The resulting high-quality knowledge graph eliminates noise and is enriched with task-relevant triples, thereby mitigating both the noisy and incomplete knowledge problems while remaining compatible with any downstream knowledge-aware recommendation model.

In summary, the main contributions of this paper are as follows:

- We propose a spectral heterogeneous diffusion framework called SHDF that mitigates noisy and incomplete knowledge problems from a data-centric perspective.

- We propose a diffusion model that incorporates recommendation priors and considers heterogeneity, capturing recommendation-oriented heterogeneous distributions in knowledge graphs.
- We design a continuous-discrete mode adapter that converts the fitted continuous distribution into a discrete knowledge graph.
- Comprehensive evaluations demonstrate that SHDF is a model-agnostic plugin compatible with any downstream knowledge-aware recommendation model, boosting downstream recommendation performance by enhancing knowledge graph quality.

2 Related Work

2.1 Knowledge-aware Recommendation

Knowledge-aware recommendation systems aim to leverage knowledge graphs to enhance item representations and user preference modeling [5, 57]. Existing knowledge-aware recommendation methods can be divided into three categories: embedding-based, path-based, and propagation-based methods [16, 37, 55]. Embedding-based methods employ graph embedding techniques to derive entity and relation representations for refining user and item profiles [3, 35, 47]. For example, CKE [47] employs TransR [20] to learn entity representations and uses the resulting semantic information to enhance collaborative filtering. While these methods leverage rich knowledge graph facts, they inadequately capture long-range semantics and complex user–item dependencies. Path-based methods address this by constructing diverse meta-paths to reveal extended association patterns [12, 40, 45]. For example, MCRec [12] uses manually designed meta-paths to learn representations that capture the context of user-item interactions. Although leveraging semantically meaningful meta-paths, these methods depend on manual path construction, which is labor-intensive and requires considerable domain expertise. Propagation-based methods overcome the limitations of the two aforementioned methods by propagating information over knowledge graphs to capture high-order semantic representations [30, 34, 36]. For instance, KGAT [38] constructs a collaborative knowledge graph based on the knowledge graph and collaborative filtering signals, and then propagates weighted neighbor signals to learn high-order item representations.

Although promising, most methods remain model-centric, overlooking knowledge graph quality problems. Recent studies [16, 42, 43] apply contrastive learning to alleviate these problems. For example, KGCL [43] generates graph views via random augmentations for contrastive learning to mitigate noise, whereas DiffKG [16] employs a Gaussian diffusion model to create knowledge graph views and contrasts them with the original knowledge graph to bolster user and item representations. While these methods enhance model robustness through contrastive learning, they remain model-centric and fail to address underlying knowledge graph quality problems, thus undermining the reliability of the derived knowledge graph views and exhibiting significant limitations. Additionally, several methods preprocess the knowledge graph to improve its quality. For instance, KRDN [55] treats knowledge triples that increase downstream recommendation loss as noise and improves model robustness by removing them. EditKG [32] assumes that similar items share attributes, transferring attributes from the most similar items while pruning those with low relevance. However, since task relevance and noise lack explicit labels, these methods rely on heuristic assumptions. Although they perform well in specific scenarios, the reliance on assumptions that do not always hold true limits their generalizability [44].

To overcome these limitations, we propose a data-centric spectral heterogeneous diffusion framework for knowledge-aware recommendation. Our framework learns the recommendation-oriented heterogeneous distribution of the knowledge graph and directly generates a high-quality graph, addressing knowledge graph quality problems without relying on heuristic assumptions.

2.2 Diffusion Model

Diffusion models, as an emerging class of generative models, have garnered significant attention for their superior ability to approximate true data distributions. Recently, there is a growing research focus on using diffusion models to fit true data distributions and generate high-quality training samples [9, 17, 21, 54]. For example, DistDiff [54] introduces a training-free augmentation framework using diffusion-based hierarchical prototypes to capture true data distributions. In class-incremental object detection, SDDGR [17] uses a diffusion model to generate high-quality samples of previously learned classes and integrates them with new data to prevent catastrophic forgetting. DiffuASR [21] applies diffusion models to create pseudo-sequences for sequential recommendation, matching user preference distributions and mitigating data sparsity issue.

However, in the field of knowledge-aware recommendation, the exploration of generating high-quality training samples using diffusion models remains inadequate. Currently, only DiffKG [16] employs a diffusion model to generate knowledge graph views that mirror the distribution of the original knowledge graph and uses contrastive learning to bolster the robustness of the recommendation system. However, a major limitation of this model-centric approach is its strong coupling with specific data and model architectures, which limits its generalizability. Besides, by adopting a standard Gaussian diffusion process, the method fails to capture task relevance and heterogeneous distributions within knowledge graphs, thereby impairing the quality of generated views.

To bridge this gap, we propose SHDF, a data-centric framework that can be seamlessly integrated into any knowledge-aware recommendation model, achieving strong generalizability. This framework employs a spectral heterogeneous diffusion model that incorporates task prior knowledge, modeling the recommendation-oriented heterogeneous distribution of the knowledge graph accurately. Guided by recommendation signals, it generates high-quality knowledge graphs that ultimately enhance the performance of downstream knowledge-aware recommendation models.

3 Problem Formulation

This section introduces relevant symbols and defines our task.

User-Item Interaction Matrix. In recommendation systems, there exists a user set U and an item set I, where $u \in U$ and $i \in I$ represent a specific user and item. The total number of users and items are denoted as $|U|$ and $|I|$, respectively. User-item interaction histories, such as clicks or purchases, can be used to construct an interaction matrix $Y \in \mathbb{R}^{|U| \times |I|}$, where each element $y_{ui} \in Y$ is 1 if user u interacted with item i, and 0 otherwise.

Knowledge Graphs. A knowledge graph is a structured representation of external information, composed of triplets (h, r, t), where h and t are the head and tail entities, respectively, and r describes the relationship between them. Formally, a knowledge graph is defined as $G_k = \{(h, r, t)|h, t \in \mathcal{E}, r \in \mathcal{R}\}$, where $\mathcal{E}$ and $\mathcal{R}$ are the sets of entities and relations. Items constitute an entity subset, i.e., $I \subseteq \mathcal{E}$, with related entities serving as attributes.

Knowledge-aware Recommendation. Given the interaction matrix Y and the knowledge graph G_k, the goal of knowledge-aware recommendation is to train a model $\hat{y}_{uv} = F(u, v|Y, G_k, \Theta)$ that predicts the interaction probability $\hat{y}_{uv}$ for user u and item v, where Θ represents the learnable parameters of the model.

Task Formulation. Inspired by the data-centric paradigm, we aim to improve the quality of knowledge graph data in knowledge-aware recommendation. Specifically, given the interaction matrix Y and the knowledge graph G_k, we train a high-quality knowledge graph generator $\tilde{G_k} = G(G_k|Y, \Phi)$ parameterized by Φ to produce a refined knowledge graph $\tilde{G_k}$ tailored to the recommendation task.

4 Methodology

In this section, we present the proposed SHDF in detail. As shown in Figure 2, the SHDF consists of a spectral heterogeneous diffusion model and a continuous-discrete mode adapter. First, the tailored diffusion model iteratively smooths and refines the knowledge graph to learn a recommendation-oriented heterogeneous distribution. The continuous-discrete mode adapter then leverages this distribution to construct a high-quality knowledge graph. To model the recommendation-oriented heterogeneous distribution, we first introduce a series of graph filters based on the user-item adjacency matrix to smooth the knowledge graph in the spectral domain with guidance from recommendation signals. Moreover, we further align the smoothing procedure of the diffusion model with the features of heterogeneous graphs, ensuring that vital heterogeneous information is retained. Next, a parameterized reverse process refines the smoothed knowledge graph iteratively, approximating the recommendation-oriented heterogeneous distribution in a stepwise manner. To construct a high-quality discrete knowledge graph based on the fitted continuous distribution, we design a continuous–discrete mode adapter that maps the distribution into a high-quality discrete knowledge graph, which is then used to train downstream knowledge-aware recommendation models.

4.1 Spectral Heterogeneous Diffusion Model

To model the recommendation-oriented heterogeneous distribution, we design a spectral heterogeneous diffusion model tailored to the knowledge-aware recommendation. This model consists of a forward smoothing process and a reverse refining process. The forward smoothing process consists of two subprocesses: diffusion and convection, as shown in Figure 3. The diffusion subprocess first transforms the knowledge graph into the spectral domain, separating information across different frequencies. It then assigns varying smoothing rates according to frequency, filtering out high-frequency noise while preserving low-frequency knowledge that captures task-relevant information [41, 56]. The convection subprocess exploits the fact that variations in node heterogeneity induce different diffusion velocities [51]. It allocates distinct smoothing velocities to different neighboring nodes based on node heterogeneity, retaining the heterogeneous knowledge distribution. In the reverse refining process, the smoothing effect is gradually removed, yielding successive approximations to the recommendation-oriented heterogeneous distribution in the knowledge graph.

4.1.1 *Forward Diffusion Subprocess with Recommendation-Oriented Knowledge Modeling.*

In the Gaussian diffusion models, the forward process incrementally blurs the original data until it becomes pure noise, causing a gradual degradation of local features. Similarly, in graph neural networks, incessant information aggregation from neighboring nodes leads to increasingly blurred node features, eventually resulting in over-smoothing [27, 49]. Inspired by this analogy, we reinterpret the forward process of conventional diffusion models as an iterative procedure of local smoothing that leads to global uniformity. This process can be described by the following Partial Differential Equation (PDE) [56]:

$$\frac{\partial \mathbf{x}}{\partial \tau} = \alpha \nabla^2 \mathbf{x}, \tag{1}$$

where $\mathbf{x}$ is the adjacency matrix representation of the knowledge graph G_k. For any entity e in G_k, $\mathbf{x_e}$ refers to the row of $\mathbf{x}$ corresponding to e, written as $\mathbf{x_e} = [x_e^1, x_e^2, ..., x_e^i]$. Each entry x_e^i equals 1 if there exists a relationship between entity e and item i, and 0 otherwise. α is a hyperparameter controlling the rate of propagation over time τ. $\nabla^2 \mathbf{x}$ is the Laplacian operator, which measures the difference between a node and the average of its neighbors. Equation (1), known as the *heat equation*, draws an analogy between the information propagation in graph neural networks and the diffusion of a substance from a high-density region to a low-density region over time [4], ultimately reaching a steady state. This aligns with the goal of our forward process. To solve Equation (1), we define the Laplacian operator as $\nabla^2 = \mathbf{A} - \mathbf{I}$ by convention [28, 56], where $\mathbf{A}$ is the item-item similarity matrix, which is given by $\mathbf{A} = \mathbf{D_I}^{-\frac{1}{2}} \mathbf{Y}^\top \mathbf{D_U}^{-\frac{1}{2}} \mathbf{Y} \mathbf{D_I}^{-\frac{1}{2}}$ following [56]. $\mathbf{Y}$ represents the user-item interaction matrix. $\mathbf{D_U}$ and $\mathbf{D_I}$ denote the degree matrices for users and items, respectively. Under these settings, the analytical solution of Equation (1) can be written as:

$$\begin{aligned} \mathbf{x}(\tau) &= \exp\{-\tau\alpha(\mathbf{I} - \mathbf{A})\}\, \mathbf{x}(0) \\ &= \mathbf{U}\left[\exp\{-\tau\alpha(\mathbf{1} - \boldsymbol{\lambda})\} \odot [\mathbf{U}^T \mathbf{x}(0)]\right], \quad \tau \geq 0. \end{aligned} \tag{2}$$

where $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_l]$ represents the unit eigenvectors of $\mathbf{A}$, and $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_l]$ denotes the corresponding eigenvalues, which are sorted in descending order. The symbol $\odot$ represents Hadamard product, and $\mathbf{x}(0) = \mathbf{x}$ denotes the initial state of the knowledge graph. $\mathbf{U}^T$, which is commonly referred to as the graph Fourier transform matrix, is the transpose of $\mathbf{U}$. The $\mathbf{1} - \boldsymbol{\lambda}$, which are the eigenvalues of $-\nabla^2$, i.e. $\mathbf{I} - \mathbf{A}$, are called graph frequencies [6]. Equation (2) first transforms the knowledge graph to the spectral domain via the graph Fourier transform. It then applies exponential attenuation to each frequency component of the knowledge graph, ensuring that higher frequencies decay more rapidly. In practice, since $\mathbf{A}$ encodes similarity for a large number of items, the computational cost is extremely high. We solve this by approximating the filters in Equation (2) using a first-order Taylor expansion with respect to τ, formally:

$$\exp\{-\tau\alpha(\mathbf{I} - \mathbf{A})\} = (1 - \tau\alpha)\mathbf{I} + \tau\alpha\mathbf{A} + O(\tau). \tag{3}$$

Based on the simplified filter in Equation (3), we linearly interpolate τ such that $0 = \tau_0 < \tau_1 < \cdots < \tau_T = 1$. This interpolation defines a

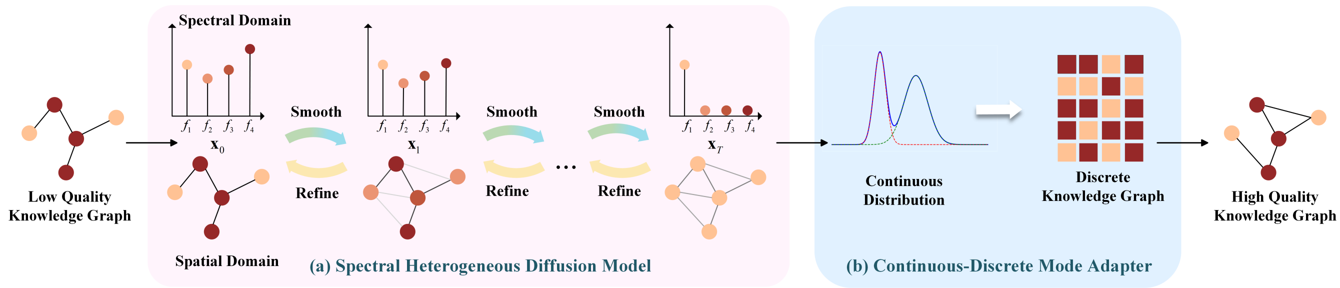


Figure 2: The SHDF comprises (a) a spectral heterogeneous diffusion model that fits the recommendation-oriented heterogeneous distribution in KG and (b) a continuous–discrete mode adapter that converts the distribution into a high-quality discrete KG.

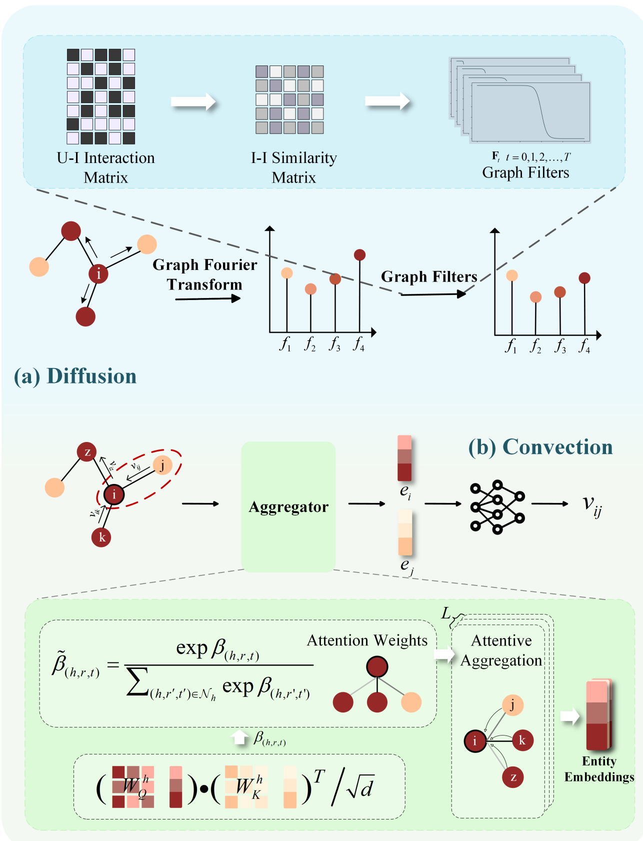


Figure 3: The forward smoothing process in SHDF comprises two subprocesses: (a) Diffusion and (b) Convection.

series of filters, which can be expressed as:

$$\mathbf{F_t} = (1 - \tau_t \alpha)\mathbf{I} + \tau_t \alpha \mathbf{A}, \quad t = 0, 1, 2, \ldots, T. \tag{4}$$

These filters are designed under the guidance of recommendation signals to smooth the knowledge graph while preserving task-relevant information. Based on the deterministic filters in Equation (4), we further incorporate Gaussian noise to introduce probabilistic modeling, yielding latent knowledge graphs:

$$\begin{aligned} \mathbf{x_t} &= \mathbf{F_t}\mathbf{x_0} + \sigma_t \boldsymbol{\epsilon_t}, \\ \boldsymbol{\epsilon_t} &\sim \mathcal{N}(0, I), \quad t = 0, 1, 2, \ldots, T, \end{aligned} \tag{5}$$

where $\mathbf{x_0}$ denotes the adjacency matrix of the original knowledge graph, and $(\sigma_t)_{t=1}^{T}$ controls the noise level at timestep t.

4.1.2 *Forward Convection Subprocess with Heterogeneous Knowledge Distribution Modeling.*

The smoothing process described above, which simulates the heat equation, calculates the diffusion velocity solely based on the average difference between a target node and its neighbors, overlooking node heterogeneity. In heterogeneous graphs, where there can be significant disparities in node features, relying solely on the mean difference obscures critical details of heterogeneous diffusion. Consequently, it is essential to assign distinct diffusion velocities to each neighboring node based on the heterogeneity between it and the target node.

Specifically, adjacent nodes in heterogeneous graphs often exhibit markedly different properties, similar to the presence of different types of particles in gases and fluids. This disparity destabilizes their connections, permitting relatively unrestricted movement. As a result, the diffusion process within gases and fluids is also influenced by the motion of the particles themselves (i.e., convection) [51]. Consequently, the complete diffusion process in heterogeneous graphs can be analogized to that in gases or fluids, where convection effects should also be taken into account, as described by Convection-Diffusion Equation (CDE) [51]:

$$\frac{\partial \mathbf{x}}{\partial \tau} = \underbrace{\alpha \nabla^2 \mathbf{x}}_{\text{Diffusion term}} - \underbrace{\mathbf{div}(\mathbf{vx})}_{\text{Convection term}} \tag{6}$$

This equation introduces a convection term, $\mathrm{div}(\mathbf{vx})$, to the heat equation, where $\mathrm{div}(\cdot)$ represents the divergence operator and $\mathbf{v}$ denotes the velocity field, describing the movement patterns of the convected quantity. By assigning different convection velocities to neighboring nodes based on the discrepancies between the target node and each neighbor, this term enables the forward process to capture the intrinsic characteristics of a heterogeneous graph [51]. To flexibly model the heterogeneity of knowledge graphs, we propose a learnable convection term. To begin with, we quantify heterogeneity between a target node and its neighbors by learning node representations via an attention-based graph convolution that integrates heterogeneous relations, as formalized below:

$$e_h^{(l)} = \frac{1}{|\mathcal{N}_h|} \sum_{(h,r,t) \in \mathcal{N}_h} \tilde{\beta}_{(h,r,t)} e_r \odot e_t^{(l-1)}, \tag{7}$$

where l represents the aggregation layer (L layers in total), and $\mathcal{N}_h$ denotes the set of all triples with h as the head entity. The representations of the head entity, relation, and tail entity in the triplet (h, r, t) are denoted by e_h, e_r and e_t, respectively, while $\tilde{\beta}(h, r, t)$ is the normalized exponential weight for this triplet, defined as:

$$\tilde{\beta}_{(h,r,t)} = \frac{\exp \beta_{(h,r,t)}}{\sum_{(h',r',t') \in \mathcal{N}_h} \exp \beta_{(h',r',t')}}, \tag{8}$$

$$\beta_{(h,r,t)} = \frac{e_h \mathbf{W_Q} \cdot (e_t \mathbf{W_K} \odot e_r)^T}{\sqrt{d}}, \tag{9}$$

where $\mathbf{W_Q}$ and $\mathbf{W_K}$ are the learnable weights used for computing attention. Leveraging the learned entity representations, we assign differentiated convection velocities to each neighboring node. Specifically, for node i, the convection term is instantiated as:

$$\mathrm{div}(\mathbf{vx})_i = \mathbf{W_1} \left(\sum_{j \in \mathcal{N}_i} v_{ij} \odot e_j \right), \tag{10}$$

$$v_{ij} = \tanh \left(\mathbf{W_2} \left(e_j - e_i \right) \right), \tag{11}$$

where $\mathbf{W_1}$ and $\mathbf{W_2}$ are learnable weights controlling the velocity of convection at the neighborhood and edge levels, respectively. $\mathcal{N}_i$ is the neighbor entity set of i. e_i and e_j represent the features of node i and node j, while v_{ij} is the velocity vector between the node pair.

The final forward smoothing process of the SHDF comprises two subprocesses: diffusion and convection, formally:

$$\mathbf{x_t} = \mathbf{F_t x} - \mathbf{div}(\mathbf{vx}) + \sigma_t \epsilon_t. \tag{12}$$

4.1.3 ***Parameterized Reverse Process.*** After smoothing the knowledge graph in the forward process, our framework employs a parameterized model in the reverse process to iteratively remove the smoothing effect. By progressively refining the smoothed graph, the model gradually approximates the recommendation-oriented heterogeneous knowledge distribution present in the original knowledge graph, as represented by the following equation:

$$q(\mathbf{x_{t-1}}|\mathbf{x_t}, \mathbf{x_0}) = \mathcal{N} \left(\mathbf{x_{t-1}}; \mathbf{F_{t-1} x_0} + \sqrt{\sigma_{t-1}^2 - \sigma_{t-1}^2}\, \epsilon_{t-1}, \sigma_{t-1|t}^2 \mathbf{I} \right). \tag{13}$$

Following [29], we further set $\sigma_{t-1|t} = 0$, making the transition from $\mathbf{x_t}$ to $\mathbf{x_{t-1}}$ deterministic, formally:

$$\mathbf{x_{t-1}} = \mathbf{F_{t-1}} \hat{\mathbf{x}}_\theta + \frac{\sigma_{t-1}}{\sigma_t} (\mathbf{x_t} - \mathbf{F_t} \hat{\mathbf{x}}_\theta), \quad t = 1, \ldots, T, \tag{14}$$

where $\hat{\mathbf{x}}_\theta$ is the model's estimation of the ideal knowledge graph distribution. We instantiate this model using a straightforward yet effective Multi-Layer Perceptron (MLP) [10]. The MLP takes the latent knowledge graph $\mathbf{x_t}$ and the time step information t as inputs, learning to fit the optimal distribution within the original knowledge graph. The model is trained with the following loss function:

$$\mathcal{L} = \mathbb{E}_{t \sim U(1,T), q(\mathbf{x_t}|\mathbf{x_0})} \left[\| \hat{\mathbf{x}}_\theta(\mathbf{x_t}, t) - \mathbf{x_0} \|^2 \right], \tag{15}$$

where $t \sim U(1,T)$ indicates that t is uniformly sampled from the set $\{1, 2, \ldots, T\}$ during the training.

4.2 Continuous-Discrete Mode Adapter

After training the spectral heterogeneous diffusion model and yielding the recommendation-oriented heterogeneous distribution fitted by the model, our framework then builds a high-quality knowledge graph from the distribution. However, directly converting this continuous distribution into a discrete knowledge graph is nontrivial. A straightforward strategy is to apply a uniform threshold S, retaining each item's top-S highest probability triples and establishing new connections to the top-S previously unlinked entities. However, varying noise and incompleteness across items make a single threshold overly coarse, while item-specific thresholds lead to an intractably large search space.

To address this, we propose a continuous–discrete mode adapter. By exploiting the learned continuous distribution, this adapter filters out recommendation-irrelevant triples and generates new connections among task-relevant entities, solving noisy and incomplete knowledge problems. Specifically, for each item i, we introduce a retention rate ρ_i^- and a completion rate ρ_i^+. According to the fitted distribution, we retain the top ρ_i^- fraction of existing triples and select the top ρ_i^+ fraction from previously unconnected entities as new knowledge. These rates are parameterized in a trade-off matrix $\mathbf{W_3} \in \mathbb{R}^{|I| \times 2}$, where each row corresponds to an item, and its two columns encode the retention and completion rates. We then learn $\mathbf{W_3}$ through a differentiable knowledge graph construction process. For instance, to denoise the knowledge for item i, we use ρ_i^- to construct the denoising mask m_i^-, formally:

$$m_i^- = \sigma \left(\frac{s_i + \log(uni) - \log(1 - uni)}{temp} \right), \tag{16}$$

$$s_i = \omega \hat{x}_{\text{stand}}^i + b_i; \quad b_i = \log \left(\frac{\rho_i^-}{1 - \rho_i^-} \right), \tag{17}$$

where, $uni \sim \mathcal{U}(0,1)$ denotes uniform noise, $temp$ is the temperature parameter, and ω is the scaling factor. s_i is the retention probability for each candidate link of item i. $\hat{x}_{\text{stand}}^i$ is the fitted distribution after standardizing, which has a zero mean. To ensure that $\mathbb{E}[s_i] = \rho_i^-$, maintaining consistency with the physical meaning of ρ_i^-, we introduce a bias satisfying $\sigma(b_i) = \rho_i^-$. Therefore, b_i is defined as in Equation (17) using the sigmoid activation function. Finally, we apply the Gumbel-Max trick [22, 23] to ensure differentiability, yielding the mask matrix m_i^- for item i. Based on the total mask matrix m^-, we construct the denoised adjacency matrix $\mathbf{x}^-$ of the knowledge graph by performing element-wise multiplication of the mask matrix m_1 with the original adjacency matrix $\mathbf{x}$. Similarly, we then construct $\mathbf{x}^+$ to capture the completed entities, guided by the completion rate. We finally obtain $\tilde{\mathbf{x}} = \mathbf{x}^- \cup \mathbf{x}^+$, where $\tilde{\mathbf{x}}$ is the adjacency matrix of the generated high-quality knowledge graph $\tilde{G_k}$. Based on $\tilde{\mathbf{x}}$, we first obtain entity representations by applying the graph convolution defined in Equation (7). The item representations, which are a subset of the entity representations, are then extracted and used to compute the user representations:

$$e_u^{(l)} = \frac{1}{|\mathcal{N}_u|} \sum_{v \in \mathcal{N}_u} e_v^{(l-1)}, \tag{18}$$

where e_u and e_v denote the representations of user and item , respectively, and $\mathcal{N}_u$ is the set of items with which the user has interacted in the training data. Finally, we optimize the representations and $\mathbf{W_3}$ using the BPR loss [25]:

$$L_{bpr} = \sum_{(u,m,n) \in D} - \log \sigma(\hat{y}_{um} - \hat{y}_{un}), \tag{19}$$

where $D = (u, m, n)$ is a triplet sampled from the training set, with m denoting an item that user u has interacted with and n representing a negative item that the user has not interacted with. $\hat{y}_{um}$ and $\hat{y}_{un}$

Table 1: Statistics of the experimental datasets.

	#Users	#Items	#Interactions	#Entities	#Triples
Last-FM	23,566	48,123	3,034,796	58,266	464,567
MIND	100,000	30,577	2,975,319	24,733	148,568
Alibaba-iFashion	114,737	30,040	1,781,093	59,156	279,155

are the predicted scores for the positive and negative, respectively, which are computed via the cosine similarity between their representations [24]. Guided by recommendation signals, the continuous-discrete mode adapter in our framework adaptively learns the retention and completion rates for each item, aligning with the recommendation task. Finally, based on the recommendation-oriented heterogeneous knowledge distribution approximated by the model, the learned retention rate is utilized to remove task-irrelevant entity connections for items, mitigating the noisy knowledge problem in the original knowledge graph. Simultaneously, the learned completion rate is applied to generate new task-relevant entity connections, effectively addressing the problem of incomplete knowledge. As a result, our framework produces a high-quality knowledge graph that can be directly utilized for the training of any downstream knowledge-aware recommendation model.

5 Experiments

We evaluate the effectiveness of the SHDF via extensive experiments, aiming to address the following Research Questions (RQs):

- **RQ1**: How does integrating SHDF affect the performance of mainstream knowledge-aware recommendation models?
- **RQ2**: How do the various innovative components of SHDF influence its overall effectiveness?
- **RQ3**: How do the key hyperparameters of SHDF affect the overall performance of recommendation?
- **RQ4**: Can SHDF mitigate the noisy and incompelte knowledge problems?
- **RQ5**: Can SHDF provide an intuitive impression for eliminating noise and generating task-relevant triples?

5.1 Experimental Settings

5.1.1 ***Dataset Description.*** We conduct experiments on three public datasets: LastFM, MIND, and Alibaba-iFashion. LastFM contains user listening records from the Last.fm music website. Following [38, 52], we link items to Freebase and extract their knowledge triples. MIND is built from Microsoft News click logs, and we construct a KG from Wikidata following [33]. Alibaba-iFashion comprises clothing click records from Alibaba's shopping platform. We manually build a KG using categorical item information following the approach proposed in [39]. Table 1 summarizes their statistics.

5.1.2 ***Evaluation Metrics.*** We evaluate the effectiveness of SHDF using two common metrics: Recall and Normalized Discounted Cumulative Gain (NDCG). Recall@K measures the proportion of relevant items correctly recommended in the top K list, while NDCG@K assesses the ranking quality of these top-K recommendations. Higher values of both metrics indicate superior performance. In the experiments conducted in this study, we set K to {10, 20}. Hereafter, we abbreviate Recall@K as R@K and NDCG@K as N@K.

5.1.3 ***Baselines.*** We evaluate the effectiveness of SHDF by comparing it with a range of representative knowledge-aware recommendation models. These include methods that directly utilize the original knowledge graph, such as CKE [47], KGIL [37], KGAT [38], KGIN [39], KGCL [43], KGRec [42], and DiffKG [16]. Additionally, we consider models that preprocess the knowledge graph to improve its utility, such as KDRN [55] and EditKG [32].

5.2 Performance Comparison (RQ1)

Table 2 presents the performance of all methods on three datasets. Based on these results, we make the following observations:

Firstly, integrating the proposed SHDF remarkably improves the performance of mainstream knowledge-aware recommendation methods across all three datasets. This indicates that SHDF accurately captures the recommendation-oriented heterogeneous distributions and adaptively constructs high-quality knowledge graphs. These generated graphs are better tailored to recommendation tasks, boosting downstream model performance. Furthermore, SHDF can be seamlessly integrated into any knowledge-aware recommendation model, regardless of the underlying model architecture or dataset, demonstrating strong generalization ability.

Secondly, in Last-FM and Alibaba-iFashion, methods with knowledge graph preprocessing (KRDN and EditKG) outperform methods without preprocessing, highlighting significant quality issues in the original knowledge graphs, such as noisy and incomplete knowledge problems. Therefore, improving the quality of knowledge graphs is crucial for enhancing the effectiveness of knowledge-aware recommendation systems. However, in MIND, knowledge-aware recommendation methods with preprocessing perform worse than methods without preprocessing. This is due to their reliance on heuristic assumptions. For instance, KRDN assumes that triples leading to high downstream loss are noise, while EditKG assumes that similar items should share identical attributes. These assumptions do not always hold true, causing these methods to perform optimally only on specific datasets and restricting their effectiveness and generalizability, thus revealing the significant drawbacks of the model-centric paradigm.

Thirdly, although methods with knowledge graph preprocessing generally achieve higher performance, integrating our framework into unpreprocessed methods allows these methods to surpass the preprocessing-based methods on some datasets. We attribute this improvement to SHDF's ability to learn recommendation-oriented heterogeneous knowledge distributions without relying on any heuristic assumptions. By enhancing knowledge graph quality from a data-centric perspective, SHDF significantly boosts downstream model performance and demonstrates strong generalizability.

5.3 Ablation Study (RQ2)

To assess the contribution of each component in SHDF to the overall performance, we conduct a series of ablation experiments. Specifically, we remove each component in turn while keeping the others unchanged, resulting in the following five variants of SHDF:

- **w/o. SHD:** Replace the Spectral Heterogeneous Diffusion process with a standard Gaussian diffusion process.
- **w/o. Den:** Exclude the Denoising operation for knowledge graphs.
- **w/o. Gen:** Remove the task-related triple Generation operation.

Table 2: Overall performance comparison of various methods on three datasets. "+SHDF " indicates the integration of SHDF into the respective backbone model. Bold represent improvements achieved by incorporating SHDF.

	Last-FM				MIND				Alibaba-iFashion			
Methods	R@10	R@20	N@10	N@20	R@10	R@20	N@10	N@20	R@10	R@20	N@10	N@20
CKE	0.0571	0.0825	0.0633	0.0706	0.0246	0.0379	0.0207	0.0250	0.0520	0.0811	0.0390	0.0495
+SHDF	**0.0593**	**0.0849**	**0.0644**	**0.0716**	**0.0270**	**0.0402**	**0.0219**	**0.0261**	**0.0538**	**0.0827**	**0.0399**	**0.0506**
KGAT	0.0571	0.0825	0.0633	0.0706	0.0260	0.0326	0.0228	0.0282	0.0524	0.0837	0.0385	0.0498
+SHDF	**0.0583**	**0.0842**	**0.0634**	**0.0710**	**0.0277**	**0.0341**	**0.0236**	**0.0291**	**0.0535**	**0.0848**	**0.0389**	**0.0504**
KGIN	0.0686	0.0950	0.0764	0.0836	0.0196	0.0376	0.0173	0.0241	0.0756	0.1146	0.0577	0.0717
+SHDF	**0.0708**	**0.0973**	**0.0776**	**0.0856**	**0.0229**	**0.0414**	**0.0192**	**0.0261**	**0.0781**	**0.1189**	**0.0597**	**0.0748**
KGCL	0.0422	0.0647	0.0523	0.0579	0.0226	0.0395	0.0175	0.0239	0.0755	0.1149	0.0576	0.0717
+SHDF	**0.0483**	**0.0703**	**0.0552**	**0.0616**	**0.0251**	**0.0421**	**0.0208**	**0.0279**	**0.0817**	**0.1212**	**0.0614**	**0.0762**
KGRec	0.0647	0.0915	0.0722	0.0796	0.0334	0.0422	0.0368	0.0373	0.0764	0.1163	0.0586	0.0729
+SHDF	**0.0681**	**0.0951**	**0.0742**	**0.0824**	**0.0352**	**0.0450**	**0.0376**	**0.0388**	**0.0782**	**0.1188**	**0.0599**	**0.0746**
KGIL	0.0696	0.0969	0.0768	0.0844	0.0204	0.0222	0.0225	0.0257	0.0774	0.1172	0.0591	0.0734
+SHDF	**0.0735**	**0.1003**	**0.0786**	**0.0863**	**0.0249**	**0.0288**	**0.0240**	**0.271**	**0.802**	**0.1203**	**0.0609**	**0.751**
DiffKG	0.0743	0.0985	0.0862	0.0921	0.0389	0.0625	0.0306	0.0397	0.0846	0.1275	0.0645	0.0799
+SHDF	**0.0803**	**0.1051**	**0.0872**	**0.0942**	**0.0425**	**0.0667**	**0.0329**	**0.0421**	**0.0919**	**0.1368**	**0.0685**	**0.0843**
KRDN	0.0732	0.1027	0.0848	0.0937	0.0283	0.0435	0.0245	0.0294	0.0920	0.1369	0.0717	0.0878
EditKG	0.0859	0.1173	0.0981	0.1061	0.0278	0.0423	0.0224	0.0326	0.0818	0.1230	0.0642	0.0786

Table 3: Performance of ablated models on three datasets.

	Last-FM		MIND		Alibaba-iFashion	
	R@20	N@20	R@20	N@20	R@20	N@20
w/o SHD	0.0963	0.0902	0.0602	0.0371	0.1283	0.0797
w/o Den	0.0995	0.0917	0.0611	0.0378	0.1312	0.0817
w/o Gen	0.1013	0.0922	0.0631	0.0382	0.1305	0.0809
w/o Ada	0.1011	0.0924	0.0633	0.0389	0.1326	0.0823
w/o Con	0.1037	0.0931	0.0646	0.0407	0.1338	0.0829
SHDF	**0.1051**	**0.0942**	**0.0667**	**0.0421**	**0.1368**	**0.0843**

- **w/o. Ada:** Disable the continuous-discrete Adapter, setting the same retention and completion rates for all items (extensive experiments show that optimal performance is achieved when the retention rate is 0.95 and the completion rate is 0.05).
- **w/o. Con:** Remove the Convection subprocess from the forward smoothing process.

In addition, we also evaluated the complete SHDF. We summarize the results in Table 3. Through analysis of the experimental results, we draw the following five conclusions. 1) Removing any component degrades overall recommendation performance, confirming that every module in SHDF is indispensable. 2) The w/o. SHD suffers the most significant performance decline among all ablated models, indicating that traditional Gaussian diffusion models tend to capture task-irrelevant distributions within the knowledge graph and disrupt its heterogeneity, thereby severely impairing the quality of the generated knowledge graph. To address this, we propose a recommendation-oriented heterogeneous diffusion model that effectively captures both the task-relevant and heterogeneous distributions, generating high-quality knowledge graphs tailored to the recommendation task. 3) The w/o. Den and w/o. Gen variants both incur significant performance drops, highlighting two major knowledge graph quality problems: the noisy knowledge problem and the incomplete knowledge problem. By filtering out noisy triples and generating task-relevant ones according to the recommendation-oriented heterogeneous knowledge distribution learned by our diffusion model, we markedly improve knowledge-graph quality and recommendation accuracy. 4) The w/o. Ada variant also performs worse, demonstrating that noise and incompleteness levels vary across items and that uniform retention and completion rates cannot meet task-specific requirements. Our continuous–discrete mode adapter addresses this by dynamically assigning retention and completion rates to each item based on the recommendation objective. Then it combines these learned rates with the continuous optimal knowledge distribution to build a high-quality, discrete knowledge graph. 5) The performance decline of the w/o. Con variant indicates that a convection term in the diffusion process is essential to capture the intrinsic properties of heterogeneous knowledge graphs. Incorporating this term effectively models heterogeneity and enhances the quality of the generated graph.

5.4 Hyperparameter Sensitivity (RQ3)

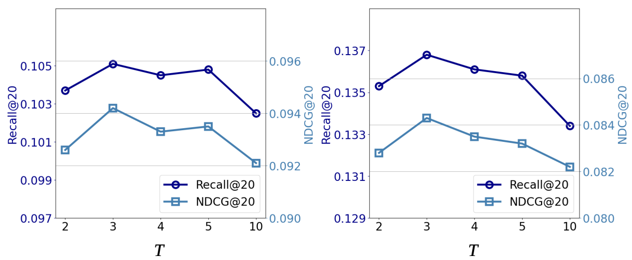
We conduct a sensitivity analysis of the diffusion step T and smoothing strength α to assess their impact on SHDF 's performance.

5.4.1 Effect of T: We vary T over $[2, 3, 4, 5, 10]$ and report the results in Figure 4. Performance peaks at $T = 3$, with larger values resulting in performance degradation. This suggests that an excessive number of diffusion steps distorts the task-relevant distribution in the knowledge graph, thereby hindering denoising and distribution approximation and ultimately degrading overall performance.

5.4.2 Effect of α: We vary α from 0.5 to 2.5 in increments of 0.5 and present the results in Figure 5. Performance peaks at $\alpha = 1.5$ and then declines. The parameter α controls information exchange during forward smoothing. When α is too low, smoothing is inadequate, so the reverse process simply restores the original graph and fails to fit the target distribution. When α is too high, excessive smoothing destroys local structures and impairs the fitted distribution.

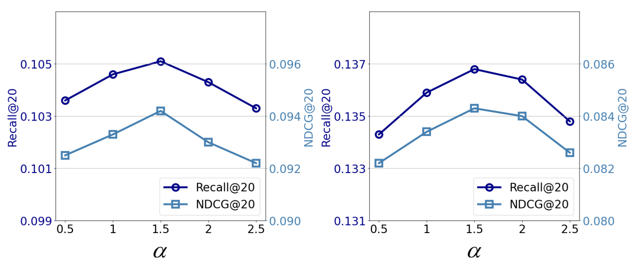
5.5 Denoising and Completion Analysis (RQ4)

To evaluate the effectiveness of SHDF in solving noisy and incomplete knowledge problems, we corrupt the knowledge graphs by random triple addition (to simulate noise) and random triple removal (to simulate incompleteness). We integrate SHDF into two



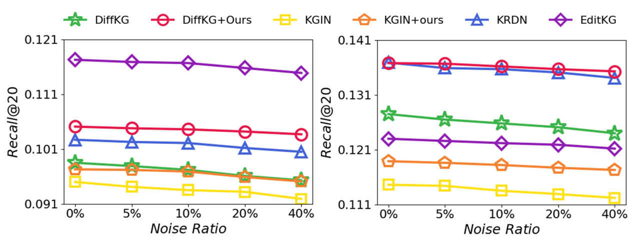
(a) Last-FM (b) Alibaba-iFashion

Figure 4: Effect of the number of diffusion step T.



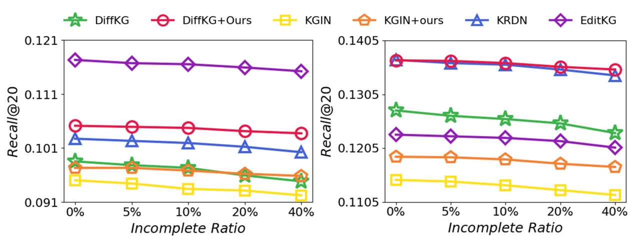
(a) Last-FM (b) Alibaba-iFashion

Figure 5: Effect of the smoothing strength α.



(a) Last-FM (b) Alibaba-iFashion

Figure 6: Performance under varying noise ratio.



(a) Last-FM (b) Alibaba-iFashion

Figure 7: Performance under varying incomplete ratio.

baselines without knowledge graph preprocessing, and then compare their decline trends against the original versions and two baselines with preprocessing. As shown in Figure 6 and Figure 7, all models' performance declines as the noise and incomplete ratio increases, however, those augmented with SHDF exhibit a markedly slower drop. This demonstrates that SHDF not only effectively mitigates noise and incompleteness but also enhances the robustness of downstream models. Moreover, although methods with preprocessing achieve state-of-the-art results on certain datasets, integrating SHDF enables models without any preprocessing to attain competitive performance, especially on the Alibaba-Fashion dataset,

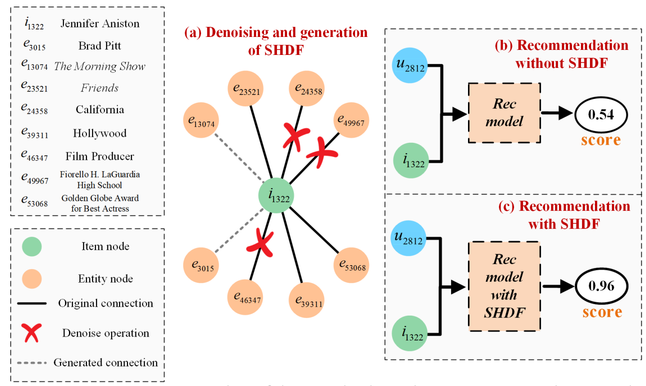


Figure 8: An example of knowledge denoising and completion in MIND and its impact on recommendation.

highlighting SHDF 's superior generalizability beyond heuristic preprocessing assumptions.

5.6 Case Study (RQ5)

As a case study, we randomly select an item from MIND and visualize its original entity connections and those obtained after processing by SHDF, demonstrating our method's effectiveness in handling noisy and incomplete knowledge for downstream recommendation. As shown in Figure 8, i_{1322} represents Jennifer Aniston. SHDF removes irrelevant facts (e.g., birthplace, her high school) while preserving knowledge relevant to user preference (e.g., her role in the TV series *Friends*). Additionally, SHDF generates new relevant facts—such as her role in the recent series *The Morning Show*—that the original graph omits. Users may also be interested in news about her because of her relationship with Brad Pitt. Moreover, we also show how this enhancement of knowledge graph quality bolsters downstream recommendation, using KGIN as the example model. For u_{2812} who interacted with i_{1322} in the test set, KGIN predicts a score of 0.54 for this ground-truth pair using the original graph versus 0.96 using the refined graph. This result shows that pruning irrelevant facts and adding key item knowledge significantly improves recommendation accuracy, validating our motivation and demonstrating the effectiveness of our framework.

6 Conclusion

In this paper, we introduce SHDF, a model-agnostic framework that leverages a tailored diffusion model to generate high-quality knowledge-graph data, thereby mitigating the noisy and incomplete knowledge problems. In the future, we plan to apply the proposed recommendation-oriented heterogeneous forward process in latent spaces and instantiate more diverse, complex reverse models within our framework.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (U21B2015 and 62472341), Natural Science Foundation of Jiangsu Province (BK20232028), the Young Scientists Fund of the National Natural Science Foundation of China (62202356 and 62302373), Intelligent Financial Software Engineering New Technology Joint Laboratory Project (99901220858).

7 Ethical Considerations

The proposed research does not raise any ethical concerns. It does not involve human participants, personal or sensitive data, or interactions with vulnerable populations. The methods and results are unlikely to cause harm under either intended or foreseeable misuse scenarios. Furthermore, the work does not pose risks related to fairness, privacy, security, or safety, and no adverse societal impact is anticipated.

References

[1] Manal A Alshehri and Xiangliang Zhang. 2022. Generative adversarial zero-shot learning for cold-start news recommendation. In *Proceedings of the 31st ACM international conference on information & knowledge management*. 26–36.

[2] Hritik Bansal and Aditya Grover. [n. d.]. Leaving Reality to Imagination: Robust Classification via Generated Datasets. In *ICLR 2023 Workshop on Trustworthy and Reliable Large-Scale Machine Learning Models*.

[3] Yixin Cao, Xiang Wang, Xiangnan He, Zikun Hu, and Tat-Seng Chua. 2019. Unifying knowledge graph learning and recommendation: Towards a better understanding of user preferences. In *The world wide web conference*. 151–161.

[4] Ben Chamberlain, James Rowbottom, Maria I Gorinova, Michael Bronstein, Stefan Webb, and Emanuele Rossi. 2021. Grand: Graph neural diffusion. In *International conference on machine learning*. PMLR, 1407–1418.

[5] Yankai Chen, Menglin Yang, Yingxue Zhang, Mengchen Zhao, Ziqiao Meng, Jianye Hao, and Irwin King. 2022. Modeling scale-free graphs with hyperbolic geometry for knowledge-aware recommendation. In *Proceedings of the fifteenth ACM international conference on web search and data mining*. 94–102.

[6] Fan RK Chung. 1997. *Spectral graph theory*. Vol. 92. American Mathematical Soc.

[7] Yan Deng, Ning Wu, Chengjun Qiu, Yangyang Luo, and Yan Chen. 2023. MixGAN-TTS: Efficient and stable speech synthesis based on diffusion model. *IEEE Access* 11 (2023), 57674–57682.

[8] Prafulla Dhariwal and Alexander Nichol. 2021. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems* 34 (2021), 8780–8794.

[9] Yupeng Fang, Ruirui Liu, Huichou Huang, Peilin Zhao, and Qingyao Wu. 2024. A Spatio-Temporal Diffusion Model for Missing and Real-Time Financial Data Inference. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. 602–611.

[10] Matt W Gardner and Stephen R Dorling. 1998. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric environment* 32, 14-15 (1998), 2627–2636.

[11] Shuyu Guo, Shuo Zhang, Weiwei Sun, Pengjie Ren, Zhumin Chen, and Zhaochun Ren. 2023. Towards explainable conversational recommender systems. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2786–2795.

[12] Binbin Hu, Chuan Shi, Wayne Xin Zhao, and Philip S Yu. 2018. Leveraging meta-path based context for top-n recommendation with a neural co-attention model. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 1531–1540.

[13] Binbin Hu, Weifan Wang, Shuhan Wang, Ziqi Liu, Bin Shen, Yong He, and Jiawei Chen. 2024. The Devil is in the Sources! Knowledge Enhanced Cross-Domain Recommendation in an Information Bottleneck Perspective. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. 880–889.

[14] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative filtering for implicit feedback datasets. In *2008 Eighth IEEE international conference on data mining*. Ieee, 263–272.

[15] Rongjie Huang, Max WY Lam, Jun Wang, Dan Su, Dong Yu, Yi Ren, and Zhou Zhao. 2022. Fastdiff: A fast conditional diffusion model for high-quality speech synthesis. *arXiv preprint arXiv:2204.09934* (2022).

[16] Yangqin Jiang, Yuhao Yang, Lianghao Xia, and Chao Huang. 2024. Diffkg: Knowledge graph diffusion model for recommendation. In *Proceedings of the 17th ACM international conference on web search and data mining*. 313–321.

[17] Junsu Kim, Hoseong Cho, Jihyeon Kim, Yihalem Yimolal Tiruneh, and Seungryul Baek. 2024. Sddgr: Stable diffusion-based deep generative replay for class incremental object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 28772–28781.

[18] Yehuda Koren. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. 426–434.

[19] Riwei Lai, Rui Chen, and Chi Zhang. 2024. A survey on data-centric recommender systems. *arXiv preprint arXiv:2401.17878* (2024).

[20] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 29.

[21] Qidong Liu, Fan Yan, Xiangyu Zhao, Zhaocheng Du, Huifeng Guo, Ruiming Tang, and Feng Tian. 2023. Diffusion augmentation for sequential recommendation. In *Proceedings of the 32nd ACM International conference on information and knowledge management*. 1576–1586.

[22] Dongsheng Luo, Wei Cheng, Dongkuan Xu, Wenchao Yu, Bo Zong, Haifeng Chen, and Xiang Zhang. 2020. Parameterized explainer for graph neural network. *Advances in neural information processing systems* 33 (2020), 19620–19631.

[23] Chris J Maddison, Andriy Mnih, and Yee Whye Teh. 2016. The concrete distribution: A continuous relaxation of discrete random variables. *arXiv preprint arXiv:1611.00712* (2016).

[24] Suja Cherukullapurath Mana and T Sasipraba. 2021. Research on cosine similarity and pearson correlation based recommendation models. In *Journal of Physics: Conference Series*, Vol. 1770. IOP Publishing, 012014.

[25] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2012. BPR: Bayesian personalized ranking from implicit feedback. *arXiv preprint arXiv:1205.2618* (2012).

[26] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10684–10695.

[27] T Konstantin Rusch, Michael M Bronstein, and Siddhartha Mishra. 2023. A survey on oversmoothing in graph neural networks. *arXiv preprint arXiv:2303.10993* (2023).

[28] Yifei Shen, Yongji Wu, Yao Zhang, Caihua Shan, Jun Zhang, B Khaled Letaief, and Dongsheng Li. 2021. How powerful is graph convolution for recommendation?. In *Proceedings of the 30th ACM international conference on information & knowledge management*. 1619–1629.

[29] Jiaming Song, Chenlin Meng, and Stefano Ermon. 2020. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502* (2020).

[30] Rui Sun, Xuezhi Cao, Yan Zhao, Junchen Wan, Kun Zhou, Fuzheng Zhang, Zhongyuan Wang, and Kai Zheng. 2020. Multi-modal knowledge graphs for recommender systems. In *Proceedings of the 29th ACM international conference on information & knowledge management*. 1405–1414.

[31] Zhongxiang Sun, Zihua Si, Xiaoxue Zang, Kai Zheng, Yang Song, Xiao Zhang, and Jun Xu. 2024. Large language models enhanced collaborative filtering. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. 2178–2188.

[32] Gu Tang, Xiaoying Gan, Jinghe Wang, Bin Lu, Lyuwen Wu, Luoyi Fu, and Chenghu Zhou. 2024. EditKG: Editing knowledge graph for recommendation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 112–122.

[33] Yu Tian, Yuhao Yang, Xudong Ren, Pengfei Wang, Fangzhao Wu, Qian Wang, and Chenliang Li. 2021. Joint knowledge pruning and recurrent graph convolution for news recommendation. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*. 51–60.

[34] Ke Tu, Peng Cui, Daixin Wang, Zhiqiang Zhang, Jun Zhou, Yuan Qi, and Wenwu Zhu. 2021. Conditional graph attention networks for distilling and refining knowledge graphs in recommendation. In *Proceedings of the 30th ACM international conference on information & knowledge management*. 1834–1843.

[35] Hongwei Wang, Fuzheng Zhang, Xing Xie, and Minyi Guo. 2018. DKN: Deep knowledge-aware network for news recommendation. In *Proceedings of the 2018 world wide web conference*. 1835–1844.

[36] Hongwei Wang, Miao Zhao, Xing Xie, Wenjie Li, and Minyi Guo. 2019. Knowledge graph convolutional networks for recommender systems. In *The world wide web conference*. 3307–3313.

[37] Shuyao Wang, Yongduo Sui, Chao Wang, and Hui Xiong. 2024. Unleashing the power of knowledge graph for recommendation via invariant learning. In *Proceedings of the ACM Web Conference 2024*. 3745–3755.

[38] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. 2019. Kgat: Knowledge graph attention network for recommendation. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 950–958.

[39] Xiang Wang, Tinglin Huang, Dingxian Wang, Yancheng Yuan, Zhenguang Liu, Xiangnan He, and Tat-Seng Chua. 2021. Learning intents behind interactions with knowledge graph for recommendation. In *Proceedings of the web conference 2021*. 878–887.

[40] Xiang Wang, Dingxian Wang, Canran Xu, Xiangnan He, Yixin Cao, and Tat-Seng Chua. 2019. Explainable reasoning over knowledge graphs for recommendation. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 5329–5336.

[41] Rui Xia, Yanhua Cheng, Yongxiang Tang, Xiaocheng Liu, Xialong Liu, Lisong Wang, and Peng Jiang. 2025. S-Diff: An Anisotropic Diffusion Model for Collaborative Filtering in Spectral Domain. In *Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining*. 70–78.

[42] Yuhao Yang, Chao Huang, Lianghao Xia, and Chunzhen Huang. 2023. Knowledge graph self-supervised rationalization for recommendation. In *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining*. 3046–3056.

[43] Yuhao Yang, Chao Huang, Lianghao Xia, and Chenliang Li. 2022. Knowledge graph contrastive learning for recommendation. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*. 1434–1443.

[44] Mingjia Yin, Hao Wang, Wei Guo, Yong Liu, Suojuan Zhang, Sirui Zhao, Defu Lian, and Enhong Chen. 2024. Dataset regeneration for sequential recommendation. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 3954–3965.

[45] Xiao Yu, Xiang Ren, Yizhou Sun, Quanquan Gu, Bradley Sturt, Urvashi Khandelwal, Brandon Norick, and Jiawei Han. 2014. Personalized entity recommendation: A heterogeneous information network approach. In *Proceedings of the 7th ACM international conference on Web search and data mining*. 283–292.

[46] Daochen Zha, Zaid Pervaiz Bhat, Kwei-Herng Lai, Fan Yang, Zhimeng Jiang, Shaochen Zhong, and Xia Hu. 2025. Data-centric artificial intelligence: A survey. *Comput. Surveys* 57, 5 (2025), 1–42.

[47] Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. 2016. Collaborative knowledge base embedding for recommender systems. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 353–362.

[48] Shuai Zhang, Hua Chu, Jianan Li, Yangtao Zhou, Shirong Wang, and Qiaofei Sun. 2025. DeMBR: Denoising Model with Memory Pruning and Semantic Guidance for Multi-Behavior Recommendation. In *Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining*. 521–529.

[49] Xu Zhang, Yonghui Xu, Wei He, Wei Guo, and Lizhen Cui. 2023. A comprehensive review of the oversmoothing in graph neural networks. In *CCF Conference on Computer Supported Cooperative Work and Social Computing*. Springer, 451–465.

[50] Jujia Zhao, Wenjie Wang, Xinyu Lin, Leigang Qu, Jizhi Zhang, and Tat-Seng Chua. 2023. Popularity-aware distributionally robust optimization for recommendation system. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. 4967–4973.

[51] Kai Zhao, Qiyu Kang, Yang Song, Rui She, Sijie Wang, and Wee Peng Tay. 2023. Graph Neural Convection-Diffusion with Heterophily. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*. 4656–4664.

[52] Wayne Xin Zhao, Gaole He, Kunlin Yang, Hongjian Dou, Jin Huang, Siqi Ouyang, and Ji-Rong Wen. 2019. Kb4rec: A data set for linking knowledge bases with recommender systems. *Data Intelligence* 1, 2 (2019), 121–136.

[53] Yangtao Zhou, Hua Chu, Qingshan Li, Jianan Li, Shuai Zhang, Feifei Zhu, Jingzhao Hu, Luqiao Wang, and Wanqiang Yang. 2025. Dual-tower model with semantic perception and timespan-coupled hypergraph for next-basket recommendation. *Neural Networks* 184 (2025), 107001.

[54] Haowei Zhu, Ling Yang, Jun-Hai Yong, Hongzhi Yin, Jiawei Jiang, Meng Xiao, Wentao Zhang, and Bin Wang. 2024. Distribution-aware data expansion with diffusion models. *arXiv preprint arXiv:2403.06741* (2024).

[55] Xinjun Zhu, Yuntao Du, Yuren Mao, Lu Chen, Yujia Hu, and Yunjun Gao. 2023. Knowledge-refined denoising network for robust recommendation. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 362–371.

[56] Yunqin Zhu, Chao Wang, Qi Zhang, and Hui Xiong. 2024. Graph signal diffusion model for collaborative filtering. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1380–1390.

[57] Ding Zou, Wei Wei, Ziyang Wang, Xian-Ling Mao, Feida Zhu, Rui Fang, and Dangyang Chen. 2022. Improving knowledge-aware recommendation with multi-level interactive contrastive learning. In *Proceedings of the 31st ACM international conference on information & knowledge management*. 2817–2826.