ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202605_9(5).0005

Research on The Learning Path of Database Courses Based on Question-oriented Knowledge Graph

Pengfei Song[1], Xinan Yue[2]

[1]Sias University, Zhengzhou, Henan, China

[2]Zhengzhou Railway Vocational & Technical College, Zhengzhou, Henan, China

Abstract

With the rapid development of information technology, database technology has become a central foundational discipline in the field of computer science and technology. However, traditional teaching methods for database courses often follow a linear, fixed knowledge delivery approach, which hardly meets the individual learning needs of students and fails to effectively bridge the gap between theoretical knowledge and the ability to solve practical problems. To address these challenges, this article presents a model for recommending learning paths for database courses based on a problem-knowledge graph (Problem-Knowledge Graph based Learning Path, P-KGLP). First, a Database Problem-Knowledge Graph (DB-PKG) is created by conducting in-depth analysis of textbooks, online courses (MOOCs), technical forums, and project examples in the field of database courses, integrating the three-part relationship of "knowledge point – problem – resource." This graph not only reveals logical dependencies between knowledge points but primarily establishes a mapping between knowledge content and real-world problems. Based on this, a personalized algorithm for generating learning paths is developed, which takes into account the current knowledge level and interests of students. The algorithm employs graph traversal techniques and dynamically plans a problem-oriented learning sequence that both deepens weak knowledge areas and enhances student motivation, while adhering to the prerequisites of the knowledge points. To verify the effectiveness of the model, experiments were conducted using publicly available MOOC datasets (such as MOOCCube). The results show that the proposed P-KGLP model significantly outperforms classical methods like the collaborative filtering algorithm (BPR), knowledge graph embedding-based recommendation approaches (TransE), and modern graph neural network models (LightGCN) in key metrics such as recommendation accuracy (Hits@10) and normalized discounted cumulative gain (NDCG@10). This study thus provides a new and effective approach for intelligent, personalized, and problem-based teaching in database courses.

Keywords

Knowledge graph, Database course, Learning path recommendation, Problem-oriented learning, Personalized learning.

1. Introduction

Database systems are fundamental to data management in today's information-driven society, and a solid grasp of their theory, design, and application is an essential competency for computer science students [1]. Yet, current database instruction confronts three major difficulties. First, teaching content is often organized rigidly by textbook chapters, making it difficult to meet the individualized needs of learners with varying prior knowledge and cognitive styles, which leads to a situation where advanced learners are under-challenged while weaker students struggle to follow. Second, the link between conceptual knowledge and real-

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202605_9(5).0005

world practice is weak; after learning abstract notions like "normal form" and "transaction," students frequently fail to see how these concepts solve concrete problems in software development, resulting in a recurring "learn–forget" pattern and a failure to cultivate higher-order problem-solving skills. Third, although a wealth of online learning materials (e.g., MOOC videos, blog posts, open-source projects) facilitates self-directed study, it also creates "information overload" and "learning disorientation," making it hard for students to identify high-quality, well-structured resources [2].

Problem-Based Learning (PBL) is widely recognized as an advanced instructional philosophy that situates learning within complex, meaningful problem contexts and promotes knowledge construction and application through problem-solving [3]. Incorporating PBL into database education is an effective means of strengthening students' practical abilities. Meanwhile, artificial intelligence, especially Knowledge Graph (KG) technology, has been revolutionizing education. With powerful semantic representation and associative reasoning, knowledge graphs can organize domain knowledge in a structured way, supporting applications such as personalized recommendation and intelligent question answering [4]. Prior work has attempted to build knowledge graphs and recommend learning paths in courses like "Data Structures," yielding positive outcomes [5].

However, most existing studies concentrate on constructing "knowledge point–knowledge point" dependency graphs, and only a limited number incorporate "problems" as core elements in the construction and application of knowledge graphs [6][7]. Therefore, this paper explores a new method that integrates PBL concepts with knowledge graph technology, proposing the construction of a "Database Problem Knowledge Graph" (DB-PKG). This graph not only captures hierarchical relationships among knowledge points, but also explicitly models "which knowledge points are required to solve a given problem" and "which problems a given knowledge point can be applied to." Based on this graph, the paper designs and implements a personalized learning path recommendation model, P-KGLP, with the goal of generating learning paths that combine systematic knowledge with engaging problems for students.

The main contributions of this work are as follows:

A problem-oriented knowledge graph for database courses (DB-PKG) is proposed and built. It innovatively treats "problems" as first-class entities and establishes multidimensional associations among knowledge, problems, and learning resources.

A personalized learning path recommendation algorithm named P-KGLP is designed, which comprehensively considers knowledge dependencies, students' current knowledge states, and their potential interest in problems.

Comparative experiments on real datasets demonstrate that the P-KGLP model significantly outperforms multiple baseline methods in terms of recommendation effectiveness.

2. Research Countent and Methodology

To achieve the aforementioned research objectives, this paper designs an overall architecture comprising a data layer, a knowledge graph layer, a user model layer, and a path recommendation layer.

2.1. Overall Architecture of the Model

The overall structure of the P-KGLP model is depicted in Figure 1. The architecture is structured into five hierarchical layers from bottom to top:

Data Layer: Serving as the foundation of the model, this layer is tasked with gathering raw corpora from multi-source heterogeneous data. The data sources encompass, but are not limited to, classic database textbooks (e.g., Database System Concept)[8], courses offered on mainstream MOOC platforms (e.g., Coursera, database courses on XuetangX), syllabi, high-

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202605_9(5).0005

quality technical blogs, database-related Q&A sessions on Stack Overflow, and course exercise repositories.

Knowledge Graph Layer: This layer constitutes the core of the present study. It is responsible for extracting knowledge from the raw data and constructing a Database Problem Knowledge Graph (DB-PKG). Internally, it comprises a knowledge extraction and fusion module, along with graph storage capabilities. Utilizing natural language processing techniques, the module performs entity recognition, relationship extraction, knowledge alignment, and fusion, ultimately forming a knowledge network with triples (head entity, relationship, tail entity) as the fundamental units, which are stored in a Neo4j graph database[9].

User Model Layer: This layer is dedicated to constructing dynamic, multidimensional profiles for each learner. The profile encompasses not only basic information but also a knowledge state vector, which quantifies students' mastery of each knowledge point within the DB-PKG (updatable through exercises, self-assessments, etc.), as well as a history of question interactions, which aids in inferring students' interest preferences.

Path Recommendation Layer: As the decision-making core of the model, this layer generates personalized learning paths. It takes the DB-PKG from the Knowledge Graph Layer and the user profile from the User Model Layer as inputs, and through its internal candidate path generation module and path ranking and selection module, it outputs an optimal sequence of learning paths.

Application Layer: Functioning as the interface for user interaction, this layer visually presents the recommended learning paths to students, collects behavioral data during the learning process, and feeds this information back to the User Model Layer, thereby forming a closed-loop adaptive learning system.

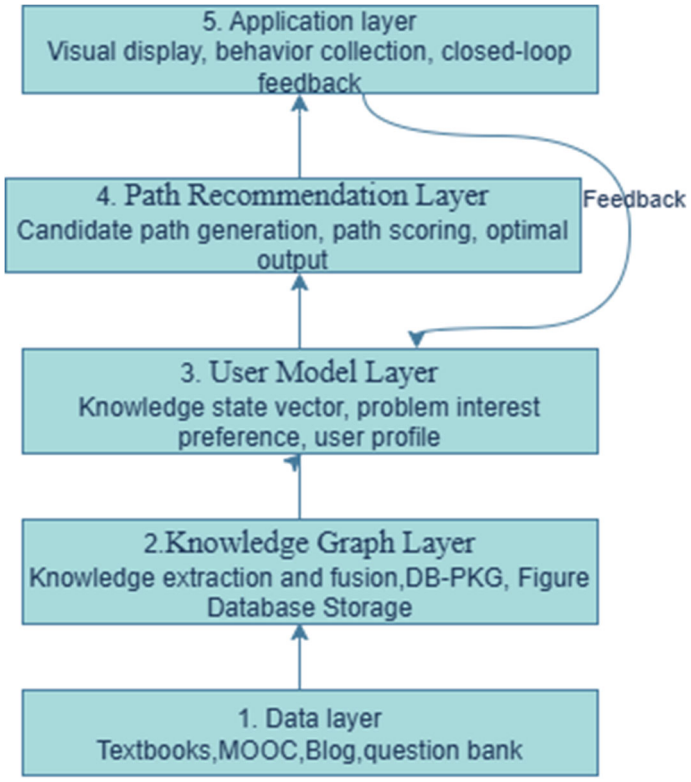

Figure 1. Diagram of the P-KGLP Model

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202605_9(5).0005

2.2. Construction of Database Problem Knowledge Graph (DB-PKG)

The construction of DB-PKG serves as the cornerstone of this research. The construction process adheres to the general technical pathway of knowledge graph construction, yet it is specifically tailored to accommodate educational scenarios and a problem-driven philosophy.

Ontology Design: Initially, the top-level schema of the graph is defined, encompassing entity types and relationship types.

(1) Entity Types:

a. Knowledge Point: Core concepts, theories, or technologies within the curriculum, exemplified by "relational algebra," "B+ tree index," and "database transaction."

b. Problem: A specific scenario or task necessitating the application of database knowledge for resolution. For instance, "How to address the overselling issue in an e-commerce system?", "How to optimize a sluggish query?", and "Design an E-R diagram for a student course selection system."

c. Learning Resource: Specific educational materials pertinent to knowledge points or problems, including video clips, article links, and code examples.

(2) Relationship Types:

a. Hasprerequisite: The dependency relationship between knowledge points, illustrated by (B+ tree index, hasprerequisite, index basis).

b. Solves: The application relationship between knowledge points and problems, demonstrated by (database transaction, solves, how to solve the overselling problem in an e-commerce system?).

c. Related_to: The correlation between entities, such as two analogous problems or knowledge points.

d. Has_resource: The linkage from knowledge points or problems to specific learning resources.

Knowledge Extraction: A hybrid approach is employed to extract triples from the textual data sourced from the data layer. For structured and semi-structured data (e.g., course syllabi and textbook catalogs), parsing scripts are developed to directly extract knowledge points and their has_prerequisite relationships.

For unstructured data (e.g., technical blogs and Q&A communities), a joint extraction model grounded in deep learning is utilized. Specifically, a pre-trained BERT model serves as the encoder, succeeded by a BiLSTM-CRF layer for named entity recognition (identifying knowledge points and problem entities), [10]and subsequently, a Graph Attention Network (GAT) is employed to capture long-range dependencies within sentences and extract relationships (identifying solves and other relationships)[11].

Knowledge Fusion and Storage: Given the potential diversity in entity names and relationship expressions (e.g., synonyms like "transaction" and "Transaction"), entity alignment techniques are leveraged for knowledge fusion. The cleansed and fused triples are then stored in the Neo4j graph database to facilitate efficient graph queries and path analyses.

2.3. Personalized Learning Path Generation Algorithm (P-KGLP)

The learning path can be delineated as a sequence of nodes, denoted as $P = (n_1, n_2, \ldots, n_k)$, within the DB-PKG framework, where each node n_i may represent either a knowledge point or a problem. The objective of the P-KGLP algorithm is to generate an optimal path P^* tailored for a specific user u.

1.Candidate Path Generation: Given a learning objective (e.g., mastering knowledge related to "database index"), starting from the target knowledge points or their associated problems, perform a Depth-First Search (DFS) or Breadth-First Search (BFS) on DB-PKG, while strictly adhering to the has_prerequisite relationship constraint, which stipulates that the precursor

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202605_9(5).0005

node of any knowledge point must precede it in the path. Consequently, a set of candidate paths $\{P_1, P_2, \ldots, P_m\}$ that satisfy the logical structure of knowledge can be generated.

2. A comprehensive scoring function $Score(P, u)$ is designed to evaluate the value of each candidate path P to user u. The final score of a path is weighted by the following three parts

$$Score(P, u) = \alpha \cdot S_{knowledge}(P, u) + \beta \cdot S_{interest}(P, u) + \gamma \cdot S_{coherence}(P) \quad (1)$$

(1) Knowledge coverage and reinforcement score $S_{knowledge}$: This item aims to guide students to learn knowledge they have not mastered or have not mastered firmly. It calculates the sum of the product of the importance of all knowledge nodes k_i in the path and the degree of user u s mastery of them.

$$S_{knowledge}(P, u) = \sum_{k_i \in P} w(k_i) \cdot (1 - M(u, k_i)) \quad (2)$$

Among them, $w(k_i)$ is the importance of knowledge point k_i (which can be calculated on the spectrum through PageRank and other algorithms)[12], and $M(u, k_i) \in [0,1]$ is the user u mastery of k_i recorded in the user model. This score encourages the path to include weak and important knowledge points of users.

(2) Question driven interest score $S_{interest}$: This item aims to stimulate learning motivation through interesting questions. It measures the matching degree between the problem node q_j in the path and the user's historical interest.

$$S_{interest}(P, u) = \sum_{q_j \in P} Sim(Emb(q_j), Emb(H_u)) \quad (3)$$

Where, $Emb(q_j)$is the embedding vector of question q_j (which can be obtained through pre training of knowledge map embedding methods such as TransE) [13], and $Emb(H_u)$ is the aggregation embedding vector of user history interaction question sequence. Sim is cosine similarity. This score encourages the path to include questions that may be of interest to the user.

(3) Path coherence score $S_{coherence}$: this item is used to ensure the logicality and smoothness of the path content. It calculates the semantic correlation between adjacent nodes in the path on the spectrum.

$$S_{coherence}(P) = \frac{1}{|P|-1} \sum_{i=1}^{|P|-1} Rel(n_i, n_{i+1}) \quad (4)$$

Where, $Rel(n_i, n_{i+1})$ represents the relationship strength of node n_i and n_{i+1} in DB-PKG (for example, 1 for direct connection, 0.5 for connection through an intermediate node, etc.).

Finally, the algorithm will select the path $P^* = \arg\max_P Score(P, u)$ with the highest score as the final recommendation result. The weight coefficient α, β, γ can be adjusted according to the teaching objectives (for example, whether to focus on compensation or exploration).

3. Experiment and Analysis

In order to evaluate the performance of P-KGLP model, this paper carried out a series of comparative experiments.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202605_9(5).0005

3.1. Experimental setup

Data set: This paper uses the open large-scale online course data set MOOCCube [14]. This dataset contains a large number of students' course registration, video watching and exercise answer records, as well as the dependencies and concept tags between courses. We selected several course data related to "database", combined with external knowledge sources, and constructed a DB-PKG containing about 5000 knowledge points, 3000 questions, and nearly 50000 relationships according to the method described in section 2.2. The data set is divided into training set, verification set and test set according to the ratio of 8:1:1.

Evaluation index: learning path recommendation is essentially a sequential recommendation task. Top K recommended evaluation indicators recognized by academia [15]:

Hits@K (Hit rate): measure the proportion of the next learning item in the test set that appears in the top K positions of the recommended list.

NDCG@K (Normalized cumulative loss gain): not only consider whether to hit, but also consider the ranking of hit items in the recommended list. The higher the ranking, the higher the score.

Mainly focus on the indicators when K=10, namely Hits@10 and NDCG@10 .

Baselines: The following representative algorithms are selected for comparison:

BPR (Bayesian Personalized Ranking): a classic collaborative filtering recommendation algorithm based on matrix decomposition, which only uses user project interaction history [16].

TransE Rec: a method based on knowledge map embedding. It uses the vector representation of entities and relationships in TransE Learning DB-PKG, and then makes recommendations based on the distance between the embedded vector of user history learning items and the embedded vector of candidate items.

Apply it to the interaction diagram of "user knowledge point" and "user problem" [17].

P-KGLP w/o Problem: An ablation version of P-KGLP, whose knowledge map does not contain problem entities and solutions relationships, is used to verify the effectiveness of the "problem driven" core design.

3.2. Experimental results and analysis

All models are evaluated on the test set, and the results are shown in Table 1 and Figure 2.

Table 1. Performance comparison of different models on MOOCCube dataset

Models	Hits@10	NDCG@10
BPR	0.1582	0.0915
TransE-Rec	0.1895	0.1137
LightGCN	0.2241	0.1426
P-KGLP w/o Problem	0.2178	0.1385
P-KGLP	0.2453	0.1612

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202605_9(5).0005

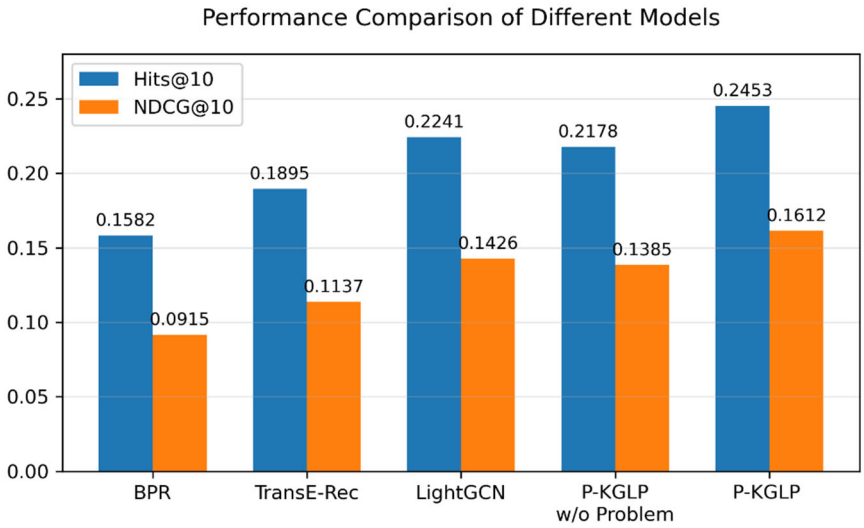

Figure 2. Model Comparison Diagram

From the experimental results, we can draw the following conclusions:

P-KGLP model has the best performance: whether Hits@10 still NDCG@10 The P-KGLP model proposed in this paper is significantly superior to all comparison algorithms. This proves the effectiveness of the path scoring function designed in this paper, which integrates knowledge dependence, user status and problem interest. Compared with the strongest baseline model LightGCN, P-KGLP Hits@10 and NDCG@10 And increased by 9.46% and 13.04% respectively.

The value of knowledge map: TransE Rec and LightGCN outperform BPR, which indicates that the introduction of structured knowledge (such as precursor relationships) contained in the knowledge map can effectively alleviate the problem of data sparsity, capture deeper associations than simple collaborative filtering, and improve the accuracy of recommendations.

Effectiveness of "problem driven" design: By comparing P-KGLP with its ablation version P-KGLP w/o Problem, it is found that the full version of P-KGLP has better performance. This strongly proves that the introduction of "problem" as a core entity into the knowledge map and the inference of user interest can more accurately predict the user's learning intention and generate more attractive and personalized learning paths. The performance of P-KGLP w/o Problem is slightly lower than that of LightGCN, probably because it only uses simple graph traversal, while LightGCN uses a more powerful graph convolution network to aggregate information. But when P-KGLP added the problem dimension, its more explanatory path generation logic exceeded the purely data-driven GNN model.

4. Conclusions

Aiming at the problems of lack of individuation and disconnection of theory and practice in traditional database teaching, this paper proposes a research scheme of database course learning path based on problem knowledge map. A multi-dimensional database problem knowledge map (DB-PKG) integrating knowledge points, problems and resources is designed and constructed, and a personalized learning path recommendation model P-KGLP is proposed on this basis. Experimental results show that the model can effectively generate high-quality learning paths, and its performance is significantly better than many mainstream recommendation algorithms.

The core innovation of this research is to deeply integrate the concept of problem-based learning (PBL) into the construction and application of the knowledge map, and organically link the abstract knowledge system with specific application scenarios through the bridge of

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202605_9(5).0005

"problems", so that the recommended learning path not only conforms to the internal logic of knowledge, but also can stimulate students' learning interest and problem solving ability.

Of course, there are still some improvements in this study. First, the current DB-PKG is built statically. In the future, we can study how to dynamically update and expand the atlas based on the latest online resources and user feedback. Secondly, the user model can be further refined, for example, the introduction of a cognitive diagnostic model to more accurately assess students' knowledge mastery status. Finally, the future work will be devoted to integrating the P-KGLP model into a real online learning platform, testing its application effect in the actual teaching environment through large-scale user experiments, and exploring its portability in other computer core courses.

Acknowledgments

The 2025 Educational Reform Fund Project of Sias University (Project Number: 2025JGYB43)

References

[1] Teaching Guidance Committee for Computer Specialties in Colleges and Universities, Ministry of Education. National Standards for Teaching Quality of Computer Specialties in Colleges and Universities [M]. Beijing: Tsinghua University Press, 2018.

[2] Li Xiaoming. MOOCs and the Teaching Reform of Core Courses in Computer Science [J]. Teaching in Chinese Universities, 2014 (5): 7-11. DOI: 10.16298/j.cnki.1005-0450.2014.05.002.

[3] BARROWS H S. Problem-based learning in medicine and beyond: A brief overview [J]. New Directions for Teaching and Learning, 1996, 1996 (68): 3-12. DOI:10.1002/tl.37219966803.

[4] Liu Zhiyuan, Han Xu, Sun Maosong. Overview and Future Prospects of Educational Knowledge Graph Research [J]. Science China: Information Sciences, 2022, 52(2): 205-246. DOI: 10.1360/SSI-2020-0337

[5] Wu Linjing, Huang Jingxiu, Liu Qingtang, et al. Construction of Knowledge Graph and Learning Path Recommendation for Data Structure Courses [J]. China Educational Technology, 2020 (3): 89-97. DOI: 10.13541/j.cnki.chinade.2020.03.012

[6] Zhang Jinbao, Zhang Sheng, Wang Jijun. Research Status and Trends of Personalized Learning Path Recommendation Based on Knowledge Graphs [J]. Journal of Educational Technology Research, 2021, 42(8): 77-85. DOI: 10.13811/j.cnki.eer.2021.08.010

[7] SHI D, WANG T, XING H, et al. A survey of knowledge graph-based adaptive learning systems for personalized recommendation [J]. IEEE Access, 2020, 8: 200454-200476. DOI:10.1109/ACCESS.2020.3035540.

[8] Silberschatz, Kress, Sudalshan. Database System Concepts (7th Edition)[M]. Translated by Yang Dongqing, Li Hongyan, Zhang Jinbo, et al. Beijing: China Machine Press, 2020

[9] Wang Haofen, Qi Guilin, Chen Huajun. Knowledge Graph: Methods, Practice, and Applications [M]. Beijing: Publishing House of Electronics Industry, 2019

[10] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis: Association for Computational Linguistics, 2019: 4171-4186. DOI:10.18653/v1/N19-1423.

[11] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks [C]//International Conference on Learning Representations. Vancouver: OpenReview, 2018.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202605_9(5).0005

[12] PAGE L, BRIN S, MOTWANI R, et al. The pagerank citation ranking: Bringing order to the web [R]. Stanford: Stanford InfoLab, 1999.

[13] BORDES A, USUNIER N, GARCIA-DURAN A, et al. Translating embeddings for modeling multi-relational data [C]//Advances in Neural Information Processing Systems 26. Lake Tahoe: Curran Associates, 2013: 2787-2795.

[14] YU J, YIN Y, LIU H, et al. MOOCCube: A large-scale data repository for NLP applications in MOOCs [C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop. Online: Association for Computational Linguistics, 2020: 169-175. DOI:10.18653/v1/2020.acl-srw.23.

[15] RICCI F, ROKACH L, SHAPIRA B. Introduction to recommender systems handbook [M]. 2nd ed. Boston: Springer, 2011: 1-35. DOI:10.1007/978-0-387-85820-3_1.

[16] RENDLE S, FREUDENTHALER C, GANTNER Z, et al. BPR: Bayesian personalized ranking from implicit feedback [C]//Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence. Montreal: AUAI Press, 2009: 452-461.

[17] HE X, DENG K, WANG X, et al. LightGCN: Simplifying and powering graph convolution network for recommendation [C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. Virtual Event: ACM, 2020: 639-648. DOI:10.1145/3397271.3401063.