

Article

Thematic PageRank: A Hybrid Approach for Evaluating Node Importance and Discipline Similarity in Interdisciplinary Networks

Jing Xiong * and Jihaoyu Yang 

School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an 710021, China; 202315030305@sust.edu.cn

* Correspondence: xiongjing@sust.edu.cn; Tel.: +86-139-9128-1900

Featured Application

The proposed framework can support research management and science policy by identifying influential disciplines, discovering emerging interdisciplinary directions, and assisting strategic planning in academic institutions and funding agencies.

Abstract

As scientific research becomes increasingly interdisciplinary, it is essential to quantify both the structural importance of disciplines and the topical proximity among them in cross-disciplinary networks. However, conventional PageRank mainly relies on link structure and therefore may overlook disciplines that are weakly connected structurally but closely related in terms of research themes. To address this limitation, we construct an interdisciplinary network dataset centered on computer science using metadata from OpenAlex, where nodes represent disciplines and edges represent co-occurrence relationships among disciplines in scholarly records. We further extract discipline-level topic representations from paper abstracts and keywords and propose a hybrid ranking method, namely Thematic PageRank (TPR), which integrates structural connectivity with topic similarity to evaluate node importance more comprehensively. In addition, we define multiple similarity measures to examine structural, thematic, and hybrid associations among disciplines. To validate the proposed method, we compare TPR with several baseline ranking algorithms, including standard PageRank, Weighted PageRank, Topic-Sensitive PageRank, HITS, LeaderRank, and Node2Vec. We also design ablation experiments to test the contribution of the structural branch, topic branch, and fusion mechanism. Experimental results show that TPR is more effective in identifying interdisciplinary disciplines at the boundary of the network, while the hybrid similarity measure provides stronger clustering performance than single-dimension measures. The proposed framework offers a more systematic and interpretable approach for evaluating disciplinary importance and similarity in interdisciplinary networks.



Academic Editor: Jose María Álvarez Rodríguez

Received: 11 June 2026

Revised: 24 June 2026

Accepted: 25 June 2026

Published: 2 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: interdisciplinary network; Pagerank; topic similarity; knowledge graph analysis; node importance evaluation

1. Introduction

With the rapid growth of interdisciplinary research, understanding how disciplines interact and identifying those fields that occupy central positions in cross-disciplinary

knowledge structures has become an important issue in scientometrics, research management, and science policy [1]. Traditional evaluation indicators, such as citation counts, journal impact factors, and publication volume, can reflect productivity or visibility to some extent, but they often fail to capture the structural relations among disciplines and the thematic connections embedded in scholarly content.

PageRank algorithm is a Web Page ranking algorithm proposed by Larry Page and Sergey Brin in 1998. It aims to evaluate the importance of web pages by analyzing the link structure between Web pages. Because of this property, it has been widely adopted in scientific networks, including citation networks, collaboration networks, and co-occurrence networks. In interdisciplinary settings, PageRank can help identify influential disciplines that are supported by many important neighbors. Nevertheless, its reliance on network structure alone limits its ability to detect disciplines that are thematically close to several fields but only weakly connected in the observed network structure.

At the same time, topic modeling and text representation methods offer a complementary perspective by quantifying the semantic characteristics of disciplines or papers. By embedding topical information into network analysis, it becomes possible to measure not only whether two disciplines are structurally linked, but also whether they are conceptually related. This is particularly important in interdisciplinary research, where topical overlap may exist even when direct structural connections are sparse. A growing body of work has attempted to combine PageRank with topical information for author ranking, journal evaluation, paper influence analysis, or topic-oriented network mining. Other studies have examined interdisciplinarity by analyzing co-occurrence, citation overlap, or diversity in disciplinary networks. Despite these efforts, two issues remain insufficiently addressed. First, many existing methods evaluate structural centrality and topical similarity separately, rather than integrating them into a unified importance model. Second, discipline networks are often built using coarse labels or journal classifications, which may conceal heterogeneity within disciplines and weaken the interpretation of cross-disciplinary relationships.

To address these gaps, this study proposes Thematic PageRank (TPR), a hybrid ranking method that integrates structural connectivity and topic similarity into the edge-weighting process of PageRank. Unlike conventional PageRank-based approaches that primarily rely on network topology or treat topic information only as an auxiliary bias term, TPR directly incorporates thematic proximity into the edge-weighting process, enabling a more balanced evaluation of disciplines that are important either structurally or semantically. In parallel, we construct an interdisciplinary network dataset centered on computer science using OpenAlex metadata and derive discipline-level topic representations from publication abstracts and keywords. Based on this dataset, we further design multiple similarity measures to analyze structural association, thematic association, and joint association among disciplines.

This study is guided by the following research question: Can integrating structural connectivity with topic similarity improve the identification of interdisciplinary disciplines that are weakly connected in the network structure but strongly related in thematic space? Based on this question, we further hypothesize that:

- (1) the proposed TPR framework can better identify interdisciplinary boundary disciplines than structure-only ranking methods;
- (2) the integration of structural and thematic information can improve clustering quality, ranking robustness, and interdisciplinary similarity analysis compared with single-view approaches.

To validate these hypotheses, we compare TPR with several representative baseline methods, including standard PageRank, Weighted PageRank, Topic-Sensitive PageRank, HITS, LeaderRank, and Node2Vec. In addition, ablation experiments are designed to

evaluate the individual contributions of the structural branch, thematic branch, and fusion mechanism. Multiple evaluation indicators, including ranking consistency, clustering validity, robustness deviation, and similarity patterns, are employed to comprehensively assess the effectiveness of the proposed framework.

The main contributions of this study are summarized as follows:

(1) We construct a real interdisciplinary network dataset covering twelve disciplines related to computer science based on OpenAlex metadata and derive discipline-level topic representations from large-scale scholarly texts.

(2) We propose Thematic PageRank (TPR), a hybrid ranking framework that jointly models structural connectivity and topic similarity for interdisciplinary node importance evaluation.

(3) We systematically investigate the relationship between structural centrality and disciplinary similarity and demonstrate that the proposed hybrid framework is more effective in identifying interdisciplinary disciplines with weak structural connectivity but strong thematic relevance.

2. Related Work

2.1. Application of Standard PageRank in Scientific Networks

PageRank is a classic link-analysis algorithm originally designed to measure the importance of web pages through their hyperlink structure and has evolved into a broad family of ranking methods widely used in scientific networks, recommendation systems, and knowledge graphs [2]. Its fundamental idea is that a node is more important if it is linked to by other important nodes. This principle has made PageRank a widely used tool for ranking nodes in scientific networks, where papers, authors, journals, institutions, or disciplines can be represented as network vertices. The iterative formula is as follows:

$$x(k) = dWx(k-1) + (1-d)r(v) \quad (1)$$

where d is the damping coefficient, which is usually 0.85, W is the transition matrix, and $r(v)$ denotes the restart preference vector. In scientometric studies, PageRank has been applied to citation networks, collaboration networks, and co-occurrence networks to identify influential scientific entities. Its advantage lies in capturing global structural prominence rather than relying solely on local degree counts. However, standard PageRank assumes uniform edge treatment and does not account for semantic similarity or content relevance. As a result, it may fail to highlight disciplines that are not structurally dominant but are thematically central within an interdisciplinary context.

2.2. Variations of PageRank

To overcome the limitations of standard PageRank, several variants have been proposed. Weighted PageRank (WPR) extends the original algorithm by assigning different transition weights to links with varying importance, rather than distributing rank scores uniformly across all outgoing links. By considering both incoming and outgoing link characteristics, WPR provides a more refined estimation of node importance [3]. Zhang et al. further adapted this idea to weighted directed networks by incorporating node-specific information and heterogeneous edge relationships into the ranking process [4]. Their results on both synthetic and real-world networks showed that the proposed method outperformed several traditional centrality measures in identifying influential nodes. However, although WPR improves upon the equal-weight assumption of standard PageRank, it still relies primarily on structural information and does not explicitly capture topical relationships among disciplines in interdisciplinary networks.

To address the limitations of structure-only ranking methods, previous studies have incorporated textual semantics or topic-model outputs into PageRank. Topic-Sensitive PageRank (TSPR) is a representative variant that introduces topic preference through personalized bias vectors [5]. Specifically, multiple topic dimensions are pre-defined, and a topic-dependent preference vector is constructed for each dimension to guide the iterative ranking process. Despite its effectiveness, traditional TSPR may incur considerable storage and computational overhead when the number of topic dimensions becomes large. To improve efficiency, Gong et al. proposed FasTSPR, an accelerated query method that exploits previous TSPR computations and improves real-time topic matching performance [6].

Several other studies have combined PageRank with neighborhood-based influence measures or topic-aware ranking strategies to identify important nodes in complex networks [7]. With the development of natural language processing techniques, researchers have increasingly extracted topical features directly from textual data and integrated them with PageRank. For example, Ma et al. proposed a ranking-based influence analysis that combines weighted PageRank and HITS for paper citation networks, and their experiments demonstrated the effectiveness of the proposed approach in identifying influential papers [8]. Overall, these studies show that incorporating topical or semantic information can improve ranking quality, but many existing methods still treat content as an auxiliary signal rather than integrating it directly into the ranking mechanism.

2.3. Interdisciplinary Networks and Similarity Measurement

In addition to node influence ranking, the analysis of interdisciplinary relationships and discipline similarity has become an important topic in scientometrics and science-of-science research. Previous studies have explored disciplinary functions and cross-disciplinary characteristics through field-of-study networks and have revealed the different roles that disciplines play within interdisciplinary ecosystems [9].

To promote interdisciplinary development, researchers have increasingly constructed discipline collaboration and co-occurrence networks based on high-quality scientific outputs. Qian et al. analyzed disciplinary interactions in Chinese first-class comprehensive universities and found that interdisciplinary activities are mainly concentrated among natural science disciplines, whereas the integration between humanities, social sciences, and other fields remains relatively limited [10]. Similarly, Wang and Li investigated university-initiated interdisciplinary research projects and demonstrated that computer science and management science often function as supportive disciplines in China's interdisciplinary research system, while social science disciplines tend to exhibit broader cross-domain collaboration patterns [11].

Evaluating the similarity between disciplines is essential for understanding interdisciplinary integration and knowledge diffusion. Existing studies generally classify similarity measures into two categories: structural similarity and thematic similarity.

Structural similarity focuses on the topological characteristics of disciplinary networks. Representative approaches include neighborhood-overlap-based measures and co-citation-based measures, which quantify the degree of shared network structure or common citation foundations between disciplines. These methods effectively capture explicit interdisciplinary co-occurrence patterns and structural connectivity, but may fail to identify disciplines that are topologically distant yet conceptually related.

Thematic similarity evaluates the semantic relatedness between disciplines using textual information such as titles, abstracts, and keywords. Topic modeling techniques and semantic embedding methods are commonly employed to generate vector representations of disciplines, after which similarity is calculated in the embedding space. Although such

methods can reveal latent thematic connections, their effectiveness depends heavily on the quality and completeness of textual data.

To overcome the limitations of single-view approaches, recent studies have explored hybrid similarity frameworks that integrate both structural and semantic information. By combining complementary network topology and topic characteristics, hybrid methods generally achieve better clustering performance and more reliable interdisciplinary relationship identification than either structural or thematic similarity alone. However, most existing approaches rely on simple weighted aggregation and lack a unified framework that simultaneously considers structural connectivity, topical relevance, and citation-based interactions.

Motivated by these observations, this study develops a multi-dimensional similarity framework that integrates structural, topical, and citation information for interdisciplinary network analysis.

2.4. Research Gap and Positioning of the Proposed Framework

Although substantial progress has been made in PageRank variants, topic-aware ranking methods [12], and interdisciplinary similarity analysis, several important limitations remain in the existing literature. Table 1 summarizes representative ranking and interdisciplinary analysis frameworks related to this study. Existing approaches can generally be divided into topology-based ranking methods, topic-aware PageRank variants, and interdisciplinary similarity analysis models. However, most existing studies still exhibit three major limitations.

Table 1. Comparison of Existing Ranking Frameworks and the Proposed TPR Method.

Method	Core Mechanism	Data Type	Semantic/Text Integration	Main Task	Advantages	Limitations
Standard PageRank (PR)	Link-based random walk propagation	Citation/co-occurrence network	None	Structural node ranking	Simple and interpretable	Ignores semantic relationships
Weighted PageRank (WPR)	Weighted transition probabilities	Weighted network	None	Weighted structural ranking	Considers edge importance	Still topology-driven
Topic-Sensitive PageRank (TSPR)	Personalized topic-biased ranking	Network and predefined topics	Topic preference vectors	Topic-oriented ranking	Introduces semantic preference	Semantic information only acts indirectly
HITS	Hub-authority iterative propagation	Directed graph	None	Authority estimation	Effective in directed networks	Sensitive to local topology
LeaderRank	Parameter-free random walk ranking	Complex networks	None	Robust node ranking	Stable convergence	Lacks semantic modeling
Node2Vec	Random-walk graph embedding	Graph representation	Implicit structural embedding	Node embedding and similarity	Captures latent structural patterns	Weak interpretability for interdisciplinary semantics
Collective Topical PageRank	Topic-dependent influence propagation	Paper citation networks	Topic-aware weighting	Paper influence evaluation	Improves topical ranking sensitivity	Mainly document-level evaluation
Proposed TPR framework	Hybrid structural-semantic propagation	Interdisciplinary network + scholarly text corpus	Topic similarity directly integrated into edge weighting	Node ranking, similarity analysis, clustering, and boundary discipline detection	Jointly models topology and semantics within a unified framework	Requires topic modeling and parameter tuning

First, many ranking methods primarily rely on network topology and treat semantic or topical information only as an auxiliary bias term rather than integrating it directly into the ranking propagation process. As a result, structurally weak but semantically important interdisciplinary disciplines may be underestimated. Second, existing interdisciplinary network studies often rely on coarse-grained discipline labels or simplified structural relationships, which may obscure the internal thematic heterogeneity and hidden semantic associations among disciplines. Third, previous studies typically focus either on node importance ranking or on disciplinary similarity analysis, while few studies jointly investigate ranking behavior, clustering structure, robustness, and interdisciplinary similarity within a unified analytical framework and dataset.

In addition, existing topic-aware PageRank approaches, such as Topic-Sensitive PageRank and collective topical ranking models, mainly target topic-dependent paper retrieval or document influence estimation. In a typical TSPR formulation, topical information enters the recurrence through a personalized restart vector or query-dependent bias distribution, while the transition matrix itself remains primarily determined by the original graph structure. In other words, topic relevance influences the teleportation term indirectly, rather than modifying the propagation weights among neighboring nodes. By contrast, TPR incorporates topic similarity directly into the edge-weight construction mechanism, so the transition probability between two disciplines is jointly determined by structural connectivity and thematic proximity. This means that semantic information participates in every propagation step of the random walk, rather than only affecting the restart distribution. Consequently, the influence of thematic information is reinforced continuously during iterative propagation instead of being restricted to a fixed personalization bias. This design improves robustness in interdisciplinary networks because semantically related but structurally sparse disciplines can still maintain effective information propagation even when direct topological connectivity is weak.

To address these limitations, this study proposes Thematic PageRank (TPR), a unified interdisciplinary analysis framework that directly integrates network structure and topic similarity into the edge-weight propagation process of PageRank. Unlike conventional topic-aware ranking models that primarily focus on document-level or query-oriented evaluation [13], the proposed framework explicitly targets interdisciplinary discipline networks and simultaneously analyzes node importance, disciplinary similarity, clustering structure, and robustness within the same analytical pipeline.

Specifically, we construct a discipline co-occurrence network based on OpenAlex meta-data and derive discipline-level topic vectors from scholarly abstracts and keywords using topic modeling techniques. Topic similarity is then incorporated directly into the PageRank transition process to form the proposed hybrid TPR framework. Furthermore, multiple similarity measures, including structural similarity, thematic similarity, co-citation similarity, and hybrid joint similarity, are designed to systematically evaluate interdisciplinary relationships. Extensive experiments, including ranking comparison, ablation analysis, clustering evaluation, and robustness testing, are conducted to validate the effectiveness of the proposed framework.

By jointly modeling structural connectivity and thematic proximity, this work aims to provide a more comprehensive, interpretable, and robust approach for identifying interdisciplinary boundary disciplines and measuring disciplinary relationships in complex scientific knowledge networks.

3. Materials and Methods

3.1. Interdisciplinary Dataset Construction

Recent scientometric datasets have increasingly relied on large-scale OpenAlex meta-data because of its broad disciplinary coverage and open accessibility. For example, the Scito2M dataset provides over two million interdisciplinary scientific records for large-scale temporal scientometric analysis [14]. Previous assessments have also demonstrated that OpenAlex provides competitive coverage and quality compared with traditional scientometric databases, making it suitable for interdisciplinary knowledge analysis [15]. This study investigates a computer science-centered interdisciplinary landscape and constructs a discipline-level case-study network covering twelve representative disciplines closely related to computer science. Rather than aiming to represent the entirety of the global scientific knowledge system, the dataset is designed to provide a focused experimental setting for evaluating the proposed TPR framework under interdisciplinary interaction scenarios. The data were collected from the OpenAlex academic database. Publication records published between 2015 and 2024 were retrieved to ensure both recent topical relevance and sufficient interdisciplinary coverage.

For each target discipline, the corresponding OpenAlex Concept ID was first identified through the OpenAlex concept search interface. The selected disciplines included Computer Science, Artificial Intelligence, Computer Vision, Statistics, Bioinformatics, Computational Physics, Computer and Electrical Engineering, Information Security, Management Science and Engineering, Medical Informatics, Quantum Information Science, and Digital Media. These disciplines were selected because they represent computationally connected yet semantically diverse research domains that frequently interact with computer science through shared methodologies, interdisciplinary applications, citation relationships, and co-occurrence patterns in scholarly publications. The selected set therefore provides a representative experimental environment for evaluating the proposed structural-semantic interdisciplinary ranking framework under realistic cross-domain interaction scenarios. Based on the identified Concept IDs, up to 1000 publications were retrieved for each discipline. Only English-language journal articles and conference papers were included in the dataset. Although the network contains only twelve discipline nodes, each node aggregates thematic information from approximately 1000 publications. Therefore, the analysis focuses on discipline-level interdisciplinary interactions rather than large-scale node prediction tasks.

The retrieved metadata included publication titles, abstracts, citation counts, publication years, discipline labels, and concept relevance information. To improve data quality and ensure consistency across disciplines, several preprocessing procedures were applied. Specifically, the abstract texts were converted to lowercase, punctuation and special characters were removed, English stop words were filtered out, and stemming was performed. Duplicate records were removed through DOI matching and comparison of publication titles. Publications with missing abstracts or incomplete metadata were excluded from topic modeling and similarity analysis. Interdisciplinary publications assigned to multiple disciplines were retained because cross-disciplinary overlap constitutes an essential characteristic of the target network. During topic vector generation, the contribution of each publication was weighted according to its citation count in order to preserve the influence of highly cited representative publications within each discipline.

In addition, self-loops and isolated nodes were removed during network construction. After preprocessing, twelve discipline nodes were retained for subsequent analysis, as summarized in Table 2.

Table 2. Disciplines Included in the Interdisciplinary Network Dataset.

Serial Number	Discipline Name	OpenAlex Concept ID	Number of Papers (Pieces)	Core Related Fields
1	Computer Science	C41008148	1000	All-domain cross-hub
2	Artificial Intelligence	C154945302	1000	Computer Science, Machine Learning, Natural Language Processing, Artificial Intelligence.
3	Computer Vision	C31972630	1000	Computer Science, Image Processing
4	Statistics	C105795698	1000	Computer Science, Statistics
5	Bioinformatics	C60644358	1000	Computer Science, Bioinformatics
6	Computational Physics	C30475298	1000	Computer Science, Computational Physics
7	Computer and Electrical Engineering	C61696701	1000	Computer Science, Electrical Engineering
8	Information Security	C527648132	1000	Computer science, Cryptography
9	Management Science and Engineering	C2777909354	1000	Systems Engineering, Information Technology, Operations Research
10	Medical Informatics	C69505689	1000	Computer Science, Medicine, Bioinformatics
11	Quantum Information Science	C5320026	1000	Computational Physics, Computer Science, Quantum Information Science
12	Digital Media	C17632256	1000	Computer Science, Digital Media

During processing, each target discipline was treated as a node in the network. When two disciplines appeared in the same paper, a co-occurrence relationship was recorded between them. The co-occurrence frequency was used as the edge weight, and a weighted undirected network was then constructed.

To reduce scale inconsistency between structural co-occurrence weights and semantic similarity scores, the co-occurrence term was normalized before integration into the TPR framework. Specifically, for each discipline node u , the normalized structural weight between disciplines u and v was defined as $\hat{s}_{u \rightarrow v}$.

$$\hat{s}_{u \rightarrow v} = \frac{s_{u \rightarrow v}}{\sum_{k \in N(u)} s_{u \rightarrow k}} \quad (2)$$

where $s_{u \rightarrow v}$ denotes the original co-occurrence frequency between disciplines u and v , and $N(u)$ represents the neighboring disciplines connected to node u . After normalization, the structural weights and cosine similarity values were represented on comparable numerical scales prior to fusion.

This normalization step improves the interpretability of the weighting parameter and prevents the ranking propagation process from being dominated by large raw co-occurrence magnitudes.

To support later analysis, we generated a node table and an edge table and stored them in CSV and GraphML formats, respectively. The paper titles and abstracts were also extracted for topic modeling, and the resulting discipline-level topic vectors were used as semantic features in the ranking and similarity analysis.

3.2. Node Topic Vector

In order to quantify the content attributes of disciplines, this study trained a topic model using the abstract text of the papers to generate discipline-level topic vectors. The Latent Dirichlet Allocation algorithm [16] was adopted, and the hyperparameters were optimized through grid search: the number of topics $K = 50$, the maximum number of iterations $\text{max_iter} = 100$. The number of topics was selected according to topic coherence and perplexity analysis, where $K = 50$ provided the best balance between semantic granularity and model stability. The model training was implemented using the scikit-learn library, a dictionary was constructed based on the preprocessed abstract text, and the text, bag-of-words vectors, and TF-IDF vectors were transformed to input the LDA model to obtain the 50-dimensional topic distribution vector for each paper. For each discipline node, the topic vector of that discipline was generated using the weighted average strategy of the number of papers—that is, the topic vector $t(c)$ of a discipline is the weighted sum of the topic vectors $t(p)$ of all papers in that discipline, with the weight being the citation count of the paper. The formula is as follows:

$$t(c) = \frac{\sum_{p \in P(c)} \text{cites}(p) \cdot t(p)}{\sum_{p \in P(c)} \text{cites}(p)} \quad (3)$$

where $P(c)$ represents the collection of papers included in discipline c , and $\text{cites}(p)$ represents the number of citations for paper p . This representation captures the aggregated thematic profile of each discipline while preserving the influence of highly cited publications. To reduce the potential bias caused by a small number of extremely highly cited papers, the discipline-level topic representation was constructed from a relatively large publication set for each discipline rather than from individual representative papers. In addition, comparative experiments using unweighted averaging produced highly similar thematic distributions and ranking trends, indicating that the citation-weighted aggregation mainly enhanced influential thematic signals without substantially distorting the overall disciplinary representation.

3.3. Hybrid PageRank Method TPR

Traditional PageRank evaluates node importance only from the perspective of network structure. Although this strategy is effective for structural ranking, it is insufficient for interdisciplinary networks because disciplines may be weakly connected in topology while still being highly related in topic space. To address this limitation, we propose the Thematic PageRank (TPR) method, which integrates structural connectivity and topic similarity into a unified ranking framework. The modeling process and the iterative scoring mechanism are shown in Figure 1. The core formula is as follows:

$$\text{TPR}(v) = (1 - d) \cdot r(v) + d \cdot \sum_{u: (u \rightarrow v)} \left(\frac{w_{u \rightarrow v}}{\sum_k w_{u \rightarrow k}} \right) \cdot \text{TPR}(u) \quad (4)$$

d is the damping coefficient, the value is 0.85. $r(v)$ is the restart preference vector, which is set to be uniformly distributed to avoid subjective bias towards a specific discipline. $w_{u \rightarrow v}$ is the edge weight that combines structure weight and topic similarity, and the formula is as follows:

$$w_{u \rightarrow v} = \alpha \cdot \hat{s}_{u \rightarrow v} + (1 - \alpha) \cdot \cos(t(u), t(v)) \quad (5)$$

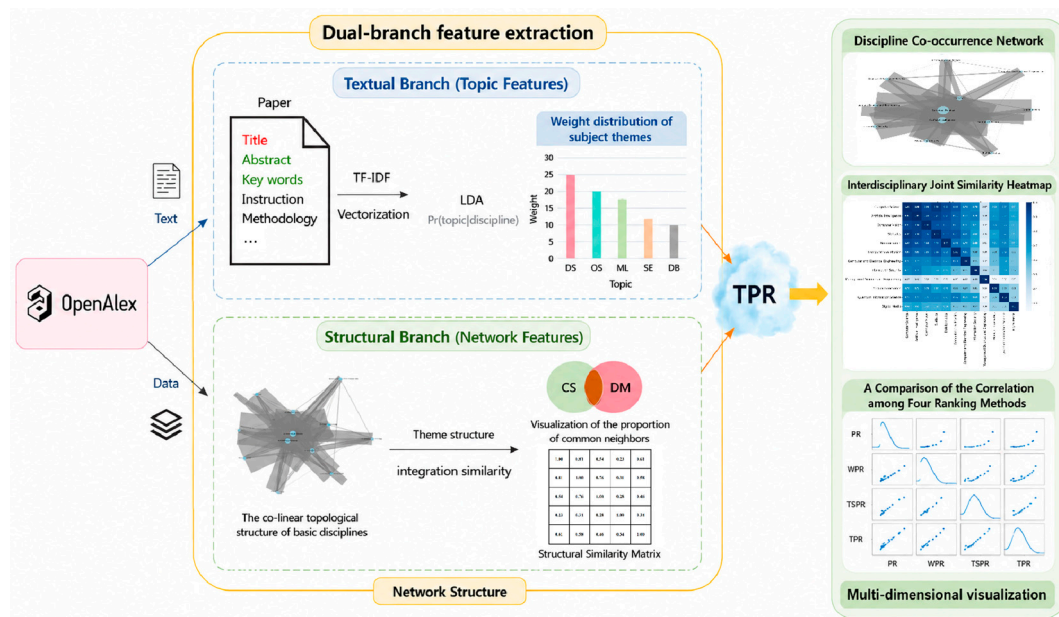


Figure 1. Framework of the Proposed TPR-Based Interdisciplinary Analysis.

Among them, the parameter α was determined through parameter sensitivity experiments to control the relative contribution of structural connectivity and thematic similarity, and was set to 0.6. Specifically, α was varied from 0.1 to 0.9 with an interval of 0.1, and the resulting ranking stability and clustering consistency were compared across different parameter settings. The experimental results showed that when α was too small, the model overly emphasized thematic similarity, which weakened the discriminative effect of structural connectivity and reduced clustering consistency. In contrast, excessively large α values caused the ranking behavior to become increasingly dominated by network topology, making the model closer to conventional weighted PageRank methods and reducing thematic differentiation among interdisciplinary disciplines. The performance metrics exhibited relatively stable behavior within the range of $\alpha = 0.4$ – 0.7 , while the best balance between ranking robustness and clustering validity was achieved at $\alpha = 0.6$. Therefore, $\alpha = 0.6$ was adopted in all subsequent experiments. $\hat{s}_{u \rightarrow v}$ represents the normalized structural weight between disciplines u and v , which was derived from the co-occurrence frequency of papers between the two disciplines after transition normalization. $\cos(t(u), t(v))$ denotes the cosine similarity between the topic vectors of disciplines u and v , with values ranging from 0 to 1.

Although the interdisciplinary network was constructed as a weighted undirected graph, the normalized edge weights were converted into transition probabilities during the ranking propagation process. Specifically, for each discipline node, the outgoing transition probability was obtained by normalizing the edge weights connected to its neighboring disciplines.

The initial TPR score of each discipline node was uniformly initialized to $1/N$, where N denotes the total number of discipline nodes. The iterative process continued until the L2 norm of the difference between two consecutive TPR vectors was smaller than 10^{-6} , indicating convergence. The converged TPR values were then used as the final node importance scores for subsequent ranking and similarity analyses.

Although TPR is built upon the general random-walk framework of PageRank, its objective and propagation mechanism differ from conventional weighted PageRank variants in several important aspects.

First, traditional Weighted PageRank methods typically assign edge weights according to structural properties such as link frequency, node degree, or citation intensity. In contrast,

the proposed TPR framework incorporates semantic topic similarity directly into the transition probability matrix, allowing thematic proximity to participate in the ranking propagation process itself rather than acting only as an external preference or auxiliary correction term.

Second, unlike Topic-Sensitive PageRank approaches that introduce topical information through personalized restart vectors or query-dependent bias distributions, TPR integrates topic similarity into the edge-weight construction mechanism. Consequently, semantic relationships influence not only node preference but also the propagation dynamics among neighboring disciplines during iterative ranking. This design enables structurally weak but semantically important interdisciplinary disciplines to obtain higher visibility within the network.

Third, the proposed framework is designed specifically for interdisciplinary discipline networks rather than document-level retrieval or citation recommendation tasks. The method jointly models structural connectivity and thematic proximity to support multiple analytical objectives, including interdisciplinary node ranking, disciplinary similarity analysis, clustering evaluation, robustness assessment, and boundary discipline identification within a unified framework.

Therefore, the novelty of TPR does not lie in introducing an entirely new random-walk family, but in constructing a hybrid structural-semantic propagation mechanism tailored for interdisciplinary knowledge network analysis and integrating ranking and similarity evaluation into a unified analytical pipeline.

3.4. Node Similarity Measures

Four types of similarity measures were defined to comprehensively quantify the association strength between disciplines. The structural neighborhood overlap similarity is defined. Based on the neighborhood overlap degree of network topology, Jaccard coefficient [17] is used to calculate, and the formula is as follows:

$$S_s(u, v) = \frac{|N(u) \cap N(v)|}{|N(u) \cup N(v)|} \quad (6)$$

where $N(u)$ represents the direct neighboring discipline set of discipline u in the interdisciplinary co-occurrence network. Its value ranges from 0 to 1, and a larger value indicates a stronger structural association between the two disciplines.

The topic cosine similarity S_t is defined, which is based on the cosine similarity of the subject topic vector, and the formula is as follows:

$$S_t(u, v) = \frac{t(u) \cdot t(v)}{|t(u)| \cdot |t(v)|} \quad (7)$$

The value ranges from 0 to 1, the larger the value is, the more similar the research topics of the two disciplines are. Co-citation similarity S_r is defined based on the proportion of co-cited papers in two disciplines, with the formula:

$$S_r(u, v) = \frac{|R(u) \cap R(v)|}{\min(|R(u)|, |R(v)|)} \quad (8)$$

where $R(u)$ represents the set of papers that cite at least one publication of discipline u in the constructed interdisciplinary dataset. Its value ranges from 0 to 1, with a higher value indicating a greater degree of overlapping citation basis between the two disciplines.

Similarity fusion methods have been shown to outperform single-view similarity measures by integrating complementary structural and semantic information [18]. The hybrid joint similarity S_j is defined to fuse the above three types of similarity and adopt the weighted average strategy. The formula is as follows:

$$S_j(u, v) = 0.4 \cdot S_s(u, v) + 0.4 \cdot S_t(u, v) + 0.2 \cdot S_r(u, v) \quad (9)$$

The weighting coefficients were determined through a grid-search procedure conducted on the validation dataset, and the value ranges from 0 to 1, which is used for subsequent similarity heat map visualization and clustering analysis.

All structural similarity measures in this study were computed based on the same discipline co-occurrence network constructed from OpenAlex publication records. Citation relationships were used only to compute the co-citation similarity measure, in which two disciplines are connected through papers that cite both of them; they were not incorporated as an independent layer in the TPR ranking-propagation process.

4. Experimental Design

4.1. Experimental Objectives and Baseline Methods

The experimental design aims to evaluate whether the proposed TPR method can better identify interdisciplinary disciplines that are weakly connected in the network structure but strongly related in topic space. In addition, the experiments assess whether combining structural and thematic similarity improves clustering performance and whether the proposed framework can reveal meaningful disciplinary association patterns in the computer science domain.

To comprehensively compare the proposed TPR method with existing ranking strategies, six baseline methods were selected:

- Standard PageRank (PR): Ranking based solely on network topology, with edge weights distributed according to outgoing links.
- Weighted PageRank (WPR): A weighted variant of PageRank that uses co-occurrence frequency as the edge weight.
- Topic-Sensitive PageRank (TSPR): A topic-aware PageRank variant that introduces a bias vector derived from topical information.
- HITS: a link-analysis method that computes authority and hub scores [19].
- LeaderRank: A parameter-free ranking method that is often regarded as robust in complex networks [20].
- Node2Vec: An embedding-based approach that generates node representations through random walks; the norm of the embedding vector was used as the importance score [21].

These methods provide a diverse set of structural, topical, and embedding-based baselines, making it possible to evaluate the advantage of TPR from multiple perspectives.

4.2. Ablation Experiment Design

In order to verify the core role of the structural branch, topic branch and fusion mechanism in TPR, three groups of ablation control experiments were designed:

- Ablation group A: the topic similarity was removed, and the edge weight only retained the subject co-occurrence structure weight.
- Ablation group B: the network structure weights were removed, and the edge weights were only retained, based on the cosine similarity of the topic vector.
- Complete TPR group: both structure branches and topic branches were retained, and linear fusion of edge weights was used.

This design allows us to determine whether the structural branch, the topic branch, and the fusion mechanism all contribute meaningfully to the final ranking performance.

4.3. Experimental Parameter Setting

The interdisciplinary dataset was constructed using publication and concept metadata collected from the OpenAlex database. Twelve representative disciplines related to computer science were selected, and up to 1000 publications were retrieved for each discipline to ensure sufficient textual information for topic modeling and network analysis.

For topic representation, publication titles and abstracts were preprocessed and transformed into numerical features using the TF-IDF method. Subsequently, Latent Dirichlet Allocation (LDA) was employed to extract latent thematic information. Preliminary sensitivity analysis was conducted by varying α from 0.1 to 0.9 with a step size of 0.1. The best overall performance was achieved at $\alpha = 0.6$, the number of topics was set to 50, which provided a suitable balance between topic granularity and model stability. The LDA model was implemented with scikit-learn and trained using the batch learning strategy with a maximum of 100 iterations. For node importance analysis, the damping factor of all PageRank-based methods was fixed at 0.85, following the standard setting commonly adopted in network analysis studies. The generated outputs included discipline ranking results, similarity matrices, discipline co-occurrence networks, ranking correlation matrices, and interdisciplinary similarity heatmaps.

The experimental workflow consisted of four major stages: interdisciplinary network construction, discipline topic vector generation, node importance ranking, and similarity analysis and visualization. These stages collectively enabled a comprehensive evaluation of disciplinary importance and interdisciplinary relationships.

4.4. Evaluation Index System

To comprehensively evaluate the proposed framework, a quantitative evaluation system was established from three perspectives: ranking reliability, clustering validity, and robustness.

4.4.1. Ranking Reliability Index

Pearson correlation coefficient was used to measure the consistency of ranking results of different methods, and the formula was as follows:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (10)$$

Among them, x_i and y_i represent the scores of the i -th discipline under the two methods, while $\bar{x}$ and $\bar{y}$ denote the corresponding mean scores. Additionally, a new metric, Top-K overlap rate, is introduced: the proportion of the intersection of the top K important nodes identified by different methods, reflecting the consistency of core discipline identification.

4.4.2. Clustering Quality Index

The Silhouette coefficient (SC) [22] was used to evaluate the compactness and separation of clusters, and the formula was as follows:

$$SC(i) = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (11)$$

$$SC = \frac{1}{n} \sum_{i=1}^n SC(i) \quad (12)$$

where, a_i is the average distance between a node i and a node in the cluster, and b_i is the average distance between a node i and its nearest cluster node. Clustering Accuracy

(CA) and Adjusted Rand Index (ARI) [23] were further used to evaluate the consistency between the clustering results and the predefined discipline categories adopted during dataset construction.

4.4.3. Robustness Index

The robustness deviation rate (RDR) is used: the average change rate of node ranking after random perturbation of 10% edge weight. The smaller the value is, the stronger the robustness is.

4.4.4. Stability Analysis

To further evaluate ranking robustness under sampling variation, a repeated subsampling-based stability analysis was conducted. Specifically, for each discipline, 80% of the publication records were randomly sampled to reconstruct the interdisciplinary network, and the proposed TPR method, together with all baseline methods, was rerun on each reconstructed network. This process was repeated 20 times.

For each repetition, the ranking obtained from the subsampled network was compared with the ranking generated from the full dataset. Stability was evaluated using Spearman rank correlation and Top-5 overlap statistics. Spearman correlation is used here to assess rank-order stability. The results, summarized in Table 3, indicate that the proposed TPR framework maintains consistently higher ranking stability than the baseline methods across repeated subsampling experiments.

Table 3. Stability Analysis Results.

Method	Mean Spearman ρ	Top-5 Overlap
PR	0.81 ± 0.07	$78\% \pm 9\%$
WPR	0.84 ± 0.06	$82\% \pm 8\%$
TSPR	0.86 ± 0.05	$85\% \pm 7\%$
TPR	0.92 ± 0.03	$91\% \pm 5\%$

In particular, TPR achieved the highest average rank correlation and Top-5 overlap rate, demonstrating that the integration of structural connectivity and thematic similarity improves robustness against sampling variation in the constructed interdisciplinary network.

4.4.5. Quantitative Comparison Protocol

To improve the reliability of the quantitative comparison on the relatively small discipline-level network, all evaluation metrics reported in Table 4 were computed using repeated perturbation-based experiments rather than a single deterministic comparison. Specifically, random perturbations were independently applied to 10% of the network edge weights in each experimental repetition, and the corresponding ranking and clustering procedures were rerun. The reported values in Table 4 represent the average results across repeated runs.

For clustering evaluation, hierarchical clustering was performed on the hybrid similarity matrix generated by each method, where the similarity matrix jointly incorporated structural connectivity and thematic similarity information. The number of clusters was fixed at three according to the predefined interdisciplinary category structure adopted during dataset construction. CA and ARI were computed by comparing the clustering assignments with the predefined discipline categories used as proxy reference labels.

Top-5 overlap was calculated by comparing the Top-5 ranked disciplines obtained from each perturbed network with the Top-5 ranking generated from the original full network. The overlap rate was defined as the proportion of shared disciplines within the Top-5 set and

was averaged across repeated perturbation runs. Note that this perturbation-based Top-5 overlap follows a different protocol from the subsampling-based Top-5 overlap reported in Table 3; the two values therefore characterize robustness under edge perturbation and under sampling variation, respectively, and are not directly comparable.

Table 4. Performance Comparison of TPR and Baseline Methods.

Methods	Clustering Accuracy (CA)	Adjusted Rand Index (ARI)	Average Ranking Deviation Rate (RDR)	Average Top-5 Overlap Rate	Silhouette Factor (SC)
PR	0.65	0.58	11.3%	70%	0.48
HITS	0.69	0.62	9.7%	75%	0.51
LeaderRank	0.71	0.65	8.1%	78%	0.53
Node2Vec	0.74	0.69	6.5%	82%	0.57
Ablation A	0.65	0.59	10.8%	70%	0.49
Ablation B	0.68	0.63	9.2%	76%	0.52
TPR	0.83	0.78	4.2%	95%	0.65

RDR was computed as the average relative ranking deviation between the perturbed rankings and the original rankings across repeated perturbation experiments. Lower RDR values therefore indicate stronger robustness of the ranking method under structural perturbation.

4.5. Experimental Procedure

The experimental workflow started with collecting metadata from OpenAlex, followed by text cleaning and discipline label standardization. We then constructed a discipline co-occurrence network and stored the corresponding node and edge tables. Next, TF-IDF was applied for text vectorization and the LDA model was used for topic extraction to generate discipline-level topic vectors.

Subsequently, multiple ranking methods, including PR, WPR, TSPR, HITS, LeaderRank, Node2Vec, the proposed TPR, and its two ablation variants, were implemented to obtain node importance scores. We further calculated structural similarity, topic similarity, co-citation similarity, and hybrid joint similarity for all discipline pairs.

Several quantitative indicators were then computed for performance evaluation, including the Pearson correlation coefficient, Silhouette Coefficient (SC), Clustering Accuracy (CA), Adjusted Rand Index (ARI), Robustness Deviation Rate (RDR), and Top-K Overlap Rate.

5. Results

Using the OpenAlex-based dataset consisting of twelve disciplines related to computer science, this study compares the proposed TPR method with PR, WPR, TSPR, HITS, LeaderRank, and Node2Vec. In the context of this specific interdisciplinary case study network, the results show that TPR preserves the overall ranking trend of structural methods while assigning higher importance to disciplines with stronger thematic relevance, thereby providing a more balanced view of interdisciplinary influence.

5.1. Discipline Importance Ranking Analysis

The reliability of ranking methods is reflected in the consistency of core-discipline identification, whereas ranking differences reveal the capability of each method to recognize interdisciplinary disciplines. This section quantifies the associations and differences of the four PageRank variants through correlation analysis and ranking comparison.

5.1.1. Ranking Consistency Analysis

Figure 2 illustrates the correlation structure among the four PageRank variants. The scatter-matrix plot shows a very strong alignment between PR and WPR, indicating that the two methods produce nearly identical ranking patterns in this dataset. This is consistent with the fact that both methods are driven primarily by network structure and edge weights. By contrast, TPR and TSPR are also strongly correlated with each other, but their score distributions are visibly different from those of PR/WPR, which suggests that topic-aware information introduces additional ranking variation beyond pure topology.

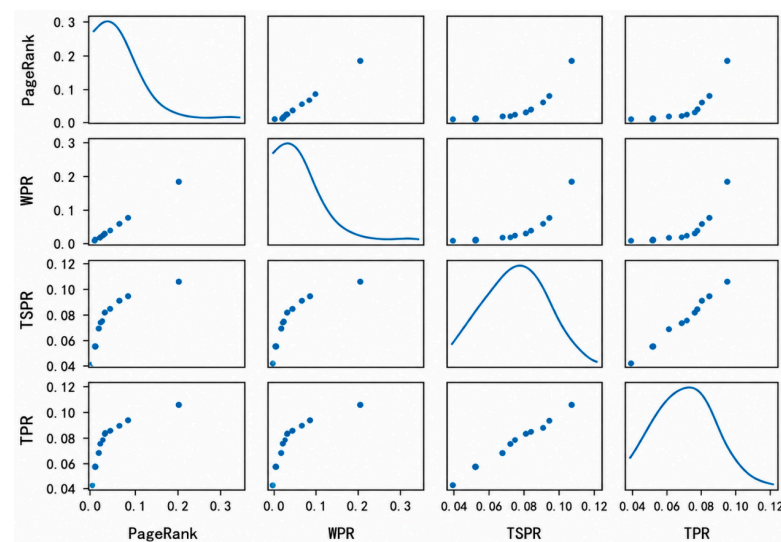


Figure 2. Scatter matrix plot of discipline rankings related to different PageRank variants.

The quantitative results further confirm this pattern. The Pearson correlation coefficient between WPR and PR is 1.0, showing complete consistency between the two methods. This is because, in the constructed co-occurrence network, the edge weights are relatively evenly distributed, so the degree-normalized transition probabilities used by WPR converge to almost the same values as the uniform out-link distribution of standard PR; consequently, the two structure-only methods produce nearly identical rankings on this particular dataset. The correlation between TPR and TSPR is 0.98, indicating that both topic-aware strategies capture broadly similar interdisciplinary signals. The correlation between TPR and standard PageRank is 0.85, which is still strongly positive but clearly lower than the PR-WPR pair. This means that TPR retains the global structural ranking tendency while making meaningful adjustments for disciplines that are semantically related but structurally less dominant.

A closer look at the score differences shows that TPR raises the relative positions of several interdisciplinary disciplines. For example, Bioinformatics and Computational Physics receive much higher TPR scores than under standard PageRank, which indicates that topic information helps identify disciplines at the interface of multiple research areas. In contrast, core disciplines such as Computer Science and Artificial Intelligence remain highly ranked across all methods, demonstrating that the proposed approach preserves the centrality of the dominant nodes while improving the recognition of boundary disciplines.

5.1.2. Discipline Importance Ranking Results

Table 5 summarizes the Top-10 disciplines under TPR, PR, WPR, and TSPR. The overall ranking pattern is highly stable at the top: Computer Science, Artificial Intelligence, and Statistics consistently occupy the first three positions across all methods. This indicates that these disciplines form the structural and thematic core of the interdisciplinary network.

Computer Science remains the top-ranked discipline in all four methods, reflecting its hub-like role as the foundational discipline connecting multiple fields.

Table 5. Top-10 Discipline Rankings under Different PageRank Variants.

List	TPR	PR	WPR	TSPR
1	Computer Science (0.1299)	Computer Science (0.3433)	Computer Science (0.3433)	Computer Science (0.1292)
2	Artificial Intelligence (0.1038)	Artificial Intelligence (0.2050)	Artificial Intelligence (0.2050)	Artificial Intelligence (0.1136)
3	Statistics (0.0948)	Statistics (0.0910)	Statistics (0.0910)	Statistics (0.1005)
4	Computer Vision (0.0901)	Computer Vision (0.0738)	Computer Vision (0.0738)	Computer Vision (0.0966)
5	Bioinformatics (0.0874)	Medical Informatics (0.0556)	Medical Informatics (0.0556)	Bioinformatics (0.0888)
6	Computational Physics (0.0865)	Management Science and Engineering (0.0448)	Management Science and Engineering (0.0448)	Computational Physics (0.0868)
7	Computer and Electrical Engineering (0.0822)	Information Security (0.0410)	Information Security (0.0410)	Computer and Electrical Engineering (0.0802)
8	Medical Informatics (0.0726)	Digital Media (0.0345)	Digital Media (0.0345)	Medical Informatics (0.0732)
9	Information Security (0.0713)	Quantum Information Science (0.0338)	Quantum Information Science (0.0338)	Information Security (0.0677)
10	Quantum Information Science (0.0655)	Bioinformatics (0.0275)	Bioinformatics (0.0275)	Quantum Information Science (0.0599)

The key advantage of TPR lies in its ability to elevate disciplines with strong interdisciplinary relevance but weaker structural centrality. In the TPR ranking, Bioinformatics moves up to 5th place and Computational Physics rises to 6th place, whereas both are ranked much lower under PR and WPR. This result suggests that the proposed method is able to identify disciplines whose importance is underestimated by topology-only ranking methods. In addition, Medical Informatics, Information Security, and Quantum Information Science also appear in relatively high positions, showing that TPR captures diverse forms of cross-disciplinary connectivity.

The comparison between TPR and TSPR is also informative. The two methods produce similar top-ranked disciplines, but TPR yields a more balanced score distribution and a more interpretable shift for interdisciplinary fields. TSPR tends to emphasize the preset topic preference more strongly, while TPR preserves structural evidence more explicitly. As a result, TPR is better suited for identifying interdisciplinary fields that are both semantically active and structurally embedded in the network.

5.2. Ablation Study

To verify the contribution of the structural and thematic components, we compared the full TPR model with two ablation variants. Table 4 shows that the full TPR model outperforms all compared baseline methods and ablation settings on every reported metric, including clustering accuracy (CA), adjusted Rand index (ARI), robustness deviation rate (RDR), Top-5 overlap rate, and Silhouette Coefficient (SC). Among the non-TPR baselines, Node2Vec performs better than the other traditional ranking methods, but it still falls short of the complete TPR model in both clustering quality and robustness.

The Structure-Only variant, which removes topic similarity, behaves similarly to topology-driven ranking methods and cannot effectively distinguish weakly connected interdisciplinary disciplines. The Topic-Only variant, which removes structural information, improves semantic sensitivity but loses the support of the network topology, resulting in a less stable ranking. By contrast, the full TPR model achieves the best overall performance, which confirms that structural connectivity and thematic proximity are complementary rather than interchangeable.

Numerically, the full TPR model achieves $CA = 0.83$, $ARI = 0.78$, $RDR = 4.2\%$, $Top-5\ overlap = 95\%$, and $SC = 0.65$. These values are clearly higher than those of the ablation variants and all baseline methods. The improvement in clustering quality indicates that the hybrid representation is more consistent with the predefined interdisciplinary category structure adopted in the constructed dataset, while the lower RDR demonstrates stronger stability under perturbation. Overall, the ablation results verify that the fusion mechanism is essential to the proposed framework.

It should be noted that the reported CA and ARI values were computed using the predefined discipline categories adopted during dataset construction as proxy reference labels. These labels were derived from OpenAlex Concept definitions and related field descriptions for the twelve Computer Science-centered disciplines included in the case-study network. Therefore, CA and ARI in this study should be interpreted as relative internal validation metrics for assessing clustering consistency within the constructed interdisciplinary dataset, rather than as evaluations against externally annotated expert ground truth.

All reported values represent averages across repeated perturbation-based experimental runs. WPR and TSPR are omitted from this table because their clustering and robustness behaviors are nearly identical to PR and the topic-aware variants, respectively; their ranking results are reported separately in Table 5.

5.3. Discipline Co-Occurrence Network Structure

Figure 3 visualizes the interdisciplinary co-occurrence network, where node size reflects the TPR score and edge width reflects the co-occurrence frequency between disciplines. The network exhibits a clear core-periphery structure centered on Computer Science. The largest node is Computer Science, which connects to nearly all other disciplines, while Artificial Intelligence, Statistics, and Computer Vision form the strongest secondary connections around the core. This structure indicates that the computer science domain functions as the central hub of the interdisciplinary network.

The network structure can be broadly divided into three hierarchical layers. The core layer contains Computer Science, Artificial Intelligence, Statistics, and Computer Vision, which are densely interconnected and represent the technological and methodological foundation of the network. The cross-disciplinary application layer includes Computational Physics, Computer and Electrical Engineering, and Quantum Information Science, reflecting the use of computer science methods in physical and engineering problems. The peripheral layer includes Bioinformatics, Medical Informatics, Information Security, Digital Media, and Management Science and Engineering, which exhibit lower TPR importance and comparatively weaker structural centrality, although they remain connected to the core network. This layered pattern is consistent with the idea that interdisciplinary knowledge tends to radiate outward from a stable core.

The network visualization also reveals that some peripheral disciplines still maintain meaningful links with the core network, even if their structural positions are weaker. This is important because such disciplines may represent emerging or under-recognized inter-

disciplinary directions. The TPR ranking makes these fields more visible than traditional PageRank-based methods, which tend to overemphasize the densest structural hubs.

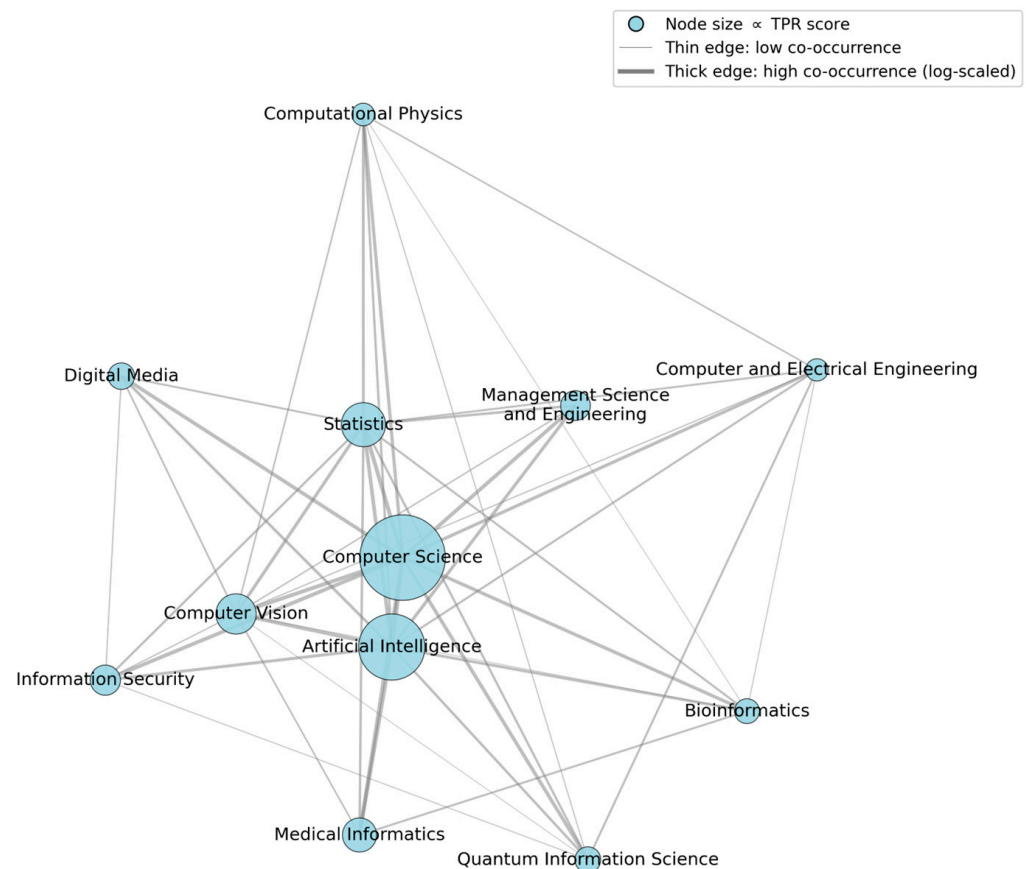


Figure 3. Interdisciplinary Co-occurrence Network Constructed from OpenAlex Data.

5.4. Quantitative Patterns of Disciplinary Similarity

Figure 4 shows the interdisciplinary joint similarity heatmap. The strongest similarity values are concentrated among the core disciplines, especially between Computer Science and Artificial Intelligence, Computer Science and Statistics, and Artificial Intelligence and Statistics. Computer Vision also shows high similarity with the core cluster, confirming that these disciplines share both structural overlap and thematic proximity. The heatmap therefore reveals a clear high-similarity block at the center of the network.

Beyond the core cluster, the similarity matrix also highlights several meaningful cross-disciplinary links. Bioinformatics remains closely connected to Computer Science, Artificial Intelligence, Statistics, and Quantum Information Science, reflecting its reliance on computational methods and data-driven analysis. Computational Physics and Computer and Electrical Engineering also show strong similarity, which is consistent with their shared focus on computational modeling and hardware-accelerated scientific computing. These patterns suggest that the hybrid similarity measure captures both technological overlap and domain-specific integration.

By contrast, Management Science and Engineering exhibits comparatively lower similarity with the core computer science cluster, indicating a weaker thematic and structural coupling. Digital Media and Information Security also show moderate similarity values, which suggests that they are related to the core network but occupy more specialized positions. Overall, the heatmap confirms that the hybrid similarity measure provides a more nuanced representation of disciplinary proximity than any single-dimension similarity metric.

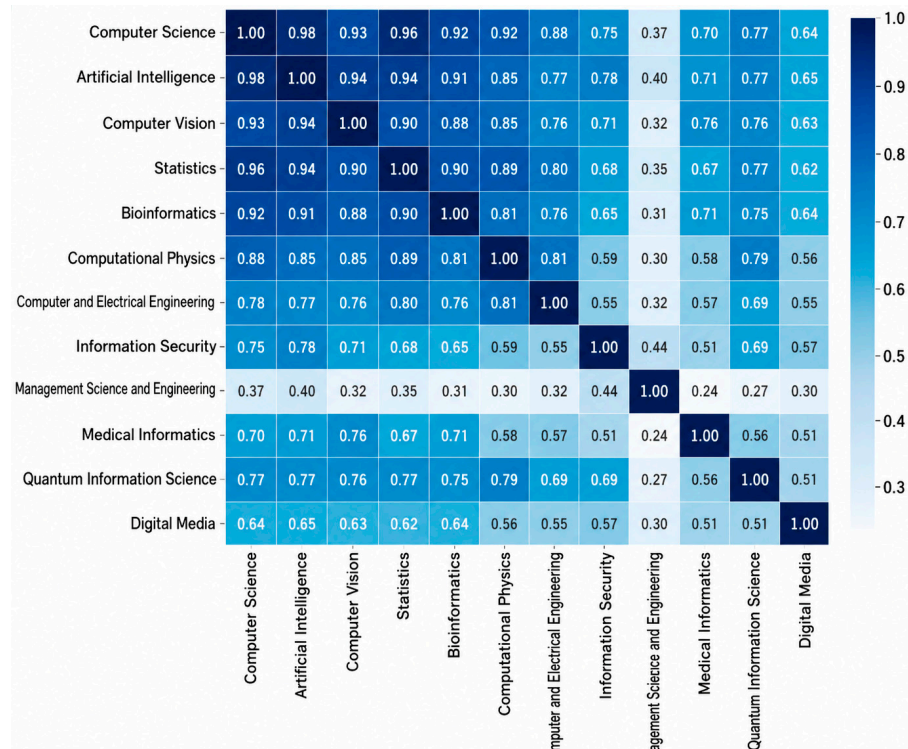


Figure 4. Hybrid Similarity Heatmap among Disciplines.

6. Discussion

6.1. Interpretation of Results

The experimental results demonstrate that incorporating topic-aware information into the PageRank framework improves the evaluation of disciplinary importance in interdisciplinary networks. Compared with conventional topology-based ranking methods, the proposed TPR approach is capable of identifying disciplines that are not only structurally connected but also semantically related to computer science.

The ranking analysis indicates that computer science consistently occupies the central position within the interdisciplinary network, which is consistent with its extensive interactions with multiple neighboring disciplines. More importantly, several disciplines that exhibited relatively weak structural connectivity achieved higher rankings under TPR due to their strong thematic relevance. This finding suggests that semantic information provides complementary evidence that cannot be fully captured by network topology alone.

The ablation experiments further verify the effectiveness of combining structural and thematic information. Neither the structure-only model nor the topic-only model achieved the same level of performance as the complete TPR framework. The superior performance of the full model demonstrates that both sources of information contribute meaningfully to interdisciplinary importance assessment.

Furthermore, the similarity analysis reveals that disciplinary relationships cannot be adequately characterized by a single perspective. Structural similarity, topic similarity, and co-citation similarity each capture different aspects of interdisciplinary interaction. The hybrid similarity measure produces more distinguishable clustering patterns, indicating that the integration of multiple information sources provides a more comprehensive representation of disciplinary proximity.

6.2. Comparison with Existing Studies

Previous studies have extensively applied PageRank and its variants to citation networks, collaboration networks, and knowledge graphs. Most existing approaches primarily rely on

network topology to estimate node importance, while semantic information is often introduced only through auxiliary weighting schemes or topic-biased personalization mechanisms.

Compared with traditional PageRank and Weighted PageRank methods, the proposed framework explicitly incorporates topic-aware information into the ranking process, allowing semantically related disciplines to receive greater attention even when their structural connections are relatively limited. This characteristic is particularly important in interdisciplinary research environments, where conceptual relationships frequently extend beyond direct co-occurrence patterns.

The experimental results are consistent with previous findings reported in topic-sensitive ranking studies, which suggest that semantic information can improve the identification of influential entities in complex networks. However, unlike studies focusing exclusively on ranking performance, the present work simultaneously investigates disciplinary similarity, network structure, and ranking behavior within a unified analytical framework. This integrated perspective provides a more comprehensive understanding of interdisciplinary relationships.

6.3. Implications for Interdisciplinary Research

The findings of this study have several implications for interdisciplinary research evaluation and scientific knowledge discovery.

First, the proposed framework provides an alternative perspective for identifying influential disciplines within complex knowledge systems. Traditional evaluation indicators, such as publication counts and citation metrics, primarily measure academic productivity or impact, whereas TPR captures the relational importance of disciplines within the broader interdisciplinary ecosystem.

Second, the similarity analysis offers insights into hidden disciplinary associations that may not be immediately observable from network structure alone. Such information can assist researchers in identifying potential collaboration opportunities and emerging interdisciplinary research directions.

Third, the proposed methodology can support science policy and research management by helping decision-makers recognize disciplines that play bridging roles between otherwise disconnected research communities. These disciplines may serve as important catalysts for knowledge transfer and innovation.

Finally, the framework can be extended to other scientific domains beyond Computer Science, enabling large-scale investigations of interdisciplinary evolution across different fields of knowledge.

6.4. Limitations and Future Work

Despite the encouraging results, several limitations should be acknowledged.

First, the interdisciplinary network was constructed using a relatively small set of twelve disciplines centered on computer science. Although this focused case-study design facilitates clearer interpretation of interdisciplinary interactions and ranking behavior, it does not fully capture the complexity and scale of the global scientific knowledge system. In particular, highly heterogeneous interdisciplinary environments involving humanities, social sciences, or fine arts may exhibit substantially different semantic structures, collaboration patterns, and citation behaviors. Consequently, the conclusions of the present study should be interpreted within the scope of this computer science-centered interdisciplinary network, and the generalizability of the proposed framework to broader cross-domain scientific ecosystems requires further validation. Future work will incorporate larger numbers of OpenAlex concepts and more diverse disciplinary domains to further evaluate the robustness and adaptability of the proposed framework.

Second, topic representations were generated using the LDA model. While LDA provides interpretable topic structures, recent advances in language representation learning, such as BERT-based embedding models and large language models, may offer richer semantic representations. Future work will explore the integration of these advanced semantic techniques into the ranking framework.

Third, the current study relies on static snapshots of interdisciplinary relationships. Scientific knowledge systems evolve continuously over time, and dynamic network analysis may provide additional insights into the temporal evolution of disciplinary importance and similarity. Recent studies have demonstrated that reconstructing the evolutionary trajectories of complex networks can reveal hidden structural transitions and long-term developmental patterns [24].

Finally, the proposed framework was evaluated primarily through ranking consistency, clustering performance, and similarity analysis. Additional validation using expert assessments or external benchmark datasets would further strengthen the reliability and generalizability of the results.

7. Conclusions

This study proposed a topic-aware interdisciplinary network analysis framework based on the integration of network topology and semantic information. Using a computer science-centered interdisciplinary case-study dataset constructed from OpenAlex metadata, an interdisciplinary co-occurrence network consisting of twelve representative disciplines was analyzed. Topic representations were generated from publication abstracts using TF-IDF and LDA topic modeling, and the proposed TPR framework was applied to evaluate disciplinary importance and interdisciplinary similarity relationships.

Experimental results demonstrated that the incorporation of topic-aware information improves the identification of influential disciplines within interdisciplinary networks. Compared with traditional PageRank-based approaches, TPR provides a hybrid structural-semantic evaluation framework by jointly considering structural relationships and thematic relevance. The ranking analysis revealed meaningful differences between topology-based and topic-aware methods, particularly for disciplines located at the boundary of the network.

In addition, multiple similarity measures were investigated to characterize interdisciplinary relationships from structural, semantic, and citation-based perspectives. The hybrid similarity measure produced clearer clustering patterns and provided a more comprehensive representation of disciplinary proximity than any single similarity metric.

Overall, the findings suggest that interdisciplinary influence cannot be fully understood through network topology alone. Integrating semantic information into network analysis offers a more informative and interpretable perspective on disciplinary interactions and knowledge organization.

Future work will focus on expanding the disciplinary scope of the dataset, incorporating advanced language representation models, and investigating the temporal evolution of interdisciplinary networks. These extensions may further improve the accuracy and applicability of topic-aware ranking methods for large-scale scientific knowledge analysis.

Author Contributions: Conceptualization, J.Y.; methodology, J.Y.; validation, J.Y.; formal analysis, J.Y.; investigation, J.Y.; data curation, J.X.; writing—original draft, J.Y.; writing—review & editing, J.X.; visualization, J.Y.; supervision, J.X.; project administration, J.X. and J.Y.; funding acquisition, J.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available in OpenAlex at <https://openalex.org/> (access on 24 June 2026).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Fortunato, S.; Bergstrom, C.T.; Börner, K.; Evans, J.A.; Helbing, D.; Milojević, S.; Petersen, A.M.; Radicchi, F.; Sinatra, R.; Uzzi, B.; et al. Science of Science. *Science* **2018**, *359*, eaao0185. [[CrossRef](#)] [[PubMed](#)]
2. Yang, M.; Wang, H.; Wei, Z.; Wang, S.; Wen, J.R. Efficient Algorithms for Personalized PageRank Computation: A Survey. *arXiv* **2024**, arXiv:2403.05198.
3. Xing, W.; Ghorbani, A. Weighted PageRank Algorithm. In Proceedings of the 2nd Annual Conference on Communication Networks and Services Research (CNSR), Fredericton, NB, Canada, 19–21 May 2004; pp. 305–314.
4. Zhang, P.; Wang, T.; Yan, J. PageRank Centrality and Algorithms for Weighted, Directed Networks. *Phys. A Stat. Mech. Its Appl.* **2022**, *586*, 126438. [[CrossRef](#)]
5. Haveliwala, T.H. Topic-Sensitive PageRank. In Proceedings of the 11th International Conference on World Wide Web (WWW), Honolulu, HI, USA, 7–11 May 2002; pp. 517–526.
6. Gong, S.; Zhang, Z.; Lu, J.; Zhang, Y.; Fu, C.; Yu, G. Accelerating Topic-Sensitive PageRank by Exploiting the Query History. In *Computing and Combinatorics (COCOON 2024)*; Springer: Singapore, 2025; pp. 497–509.
7. Hajarathaiyah, K.; Enduri, M.K.; Anamalamudi, S.; Subba Reddy, T.; Tokala, S. Computing Influential Nodes Using the Nearest Neighborhood Trust Value and PageRank in Complex Networks. *Entropy* **2022**, *24*, 704. [[CrossRef](#)] [[PubMed](#)]
8. Ma, K.; Wei, Q.; Zhang, W.; Lu, Y.; Abraham, A. Ranking Academic Influence with Integration of Weighted PageRank and HITS for Paper Citation Network. *Int. J. Grid Util. Comput.* **2024**, *15*, 543–560. [[CrossRef](#)]
9. Cunningham, E.; Smyth, B.; Greene, D. Author Multidisciplinarity and Disciplinary Roles in Field of Study Networks. *Appl. Netw. Sci.* **2022**, *7*, 78. [[CrossRef](#)] [[PubMed](#)]
10. Qian, G.; Qiu, F.; He, J.; Zhang, L. Current Situation and Characteristics of Discipline Crossing in Chinese First-Class Comprehensive Universities Based on Statistical Analysis of High-Quality Academic Output. *Educ. Inf. Technol.* **2025**, *30*, 12447–12461. [[CrossRef](#)]
11. Wang, C.; Li, Y. Diversity and Coherence: Structural Characteristics of University-Initiated Interdisciplinary Studies in China. *Front. Educ. China* **2025**, *20*, 136–161.
12. Yan, E. Topic-Based PageRank: Toward a Topic-Level Scientific Evaluation. *Scientometrics* **2014**, *100*, 407–437. [[CrossRef](#)]
13. Zhang, Y.; Ma, J.; Wang, Z.; Chen, B.; Yu, Y. Collective Topical PageRank: A Model to Evaluate the Topic-Dependent Academic Impact of Scientific Papers. *Scientometrics* **2018**, *114*, 1345–1372.
14. Jin, Y.; Xiao, Y.; Wang, Y.; Wang, J. Scito2M: A 2 Million, 30-Year Cross-Disciplinary Dataset for Temporal Scientometric Analysis. *arXiv* **2024**, arXiv:2410.09510.
15. Rong, G.; Chen, Y.; Koch, T.; Honda, K. From Coverage to Prestige: A Comprehensive Assessment of Large-Scale Scientometric Data. *arXiv* **2025**, arXiv:2503.03251.
16. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
17. Real, R.; Vargas, J.M. The Probabilistic Basis of Jaccard's Index of Similarity. *Syst. Biol.* **1996**, *45*, 380–385. [[CrossRef](#)]
18. Chen, J.; Rong, W.; Tao, G.; Cai, H. Similarity Fusion via Exploiting High-Order Proximity for Cancer Subtyping. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2023**, *20*, 658–667. [[CrossRef](#)] [[PubMed](#)]
19. Kleinberg, J.M. Authoritative Sources in a Hyperlinked Environment. *J. ACM* **1999**, *46*, 604–632. [[CrossRef](#)]
20. Lü, L.; Zhang, Y.C.; Yeung, C.H.; Zhou, T. Leaders in Social Networks, the Delicious Case. *PLoS ONE* **2011**, *6*, e21202. [[CrossRef](#)] [[PubMed](#)]
21. Grover, A.; Leskovec, J. node2vec: Scalable Feature Learning for Networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), San Francisco, CA, USA, 13–17 August 2016; pp. 855–864.
22. Rousseeuw, P.J. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [[CrossRef](#)]
23. Hubert, L.; Arabie, P. Comparing Partitions. *J. Classif.* **1985**, *2*, 193–218. [[CrossRef](#)]
24. Wang, J.; Zhang, Y.J.; Xu, C.; Li, J.; Sun, J.; Xie, J.; Feng, L.; Zhou, T.; Hu, Y. Reconstructing the Evolution History of Networked Complex Systems. *Nat. Commun.* **2024**, *15*, 2849. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.