Proceedings of the 2025 Winter Simulation Conference
E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.

BEYOND CO-AUTHORSHIP: DISCOVERING NOVEL COLLABORATORS WITH MULTILAYER RANDOM-WALK-BASED SIMULATION IN ACADEMIC NETWORKS

Siyu Chen[1], Keng Hou Leong[1], Jiadong Liu[1], Wei Chen[1,2],and Wai Kin (Victor) Chan[1]

[1]Institute of Data and Information, Tsinghua Shenzhen International Graduate School (SIGS),
Tsinghua University, Shenzhen, P.R.CHINA
[2]Tencent Technology (Shenzhen) Co. LTD., Shenzhen, P.R.CHINA

ABSTRACT

Academic collaboration is vital for enhancing research impact and interdisciplinary exploration, yet finding suitable collaborators remains challenging. Conventional single-layer random walk methods often struggle with the heterogeneity of academic networks and limited recommendation novelty. To overcome these limitations, we propose a novel Multilayer Random Walk simulation framework (MLRW) that simulates scholarly interactions across cooperation, institutional affiliation, and conference attendance, enabling inter-layer transitions to capture multifaceted scholarly relationships. Tested on the large-scale SciSciNet dataset, our MLRW simulation framework significantly outperforms conventional random walk methods in accuracy and novelty, successfully identifying potential collaborators beyond immediate co-authorship. Our analysis further confirms the significance of institutional affiliation as a collaborative predictor, validating its inclusion. This research contributes a more comprehensive simulation approach to scholar recommendations, enhancing the discovery of latent practical collaborations. Future research will focus on integrating additional interaction dimensions and optimizing weighting strategies to further improve diversity and relevance.

1 INTRODUCTION

Academic collaboration enables the synergistic integration of complementary expertise, enhances research impact, and fosters interdisciplinary exploration. The identification of optimal collaborators who demonstrate strong alignment with a researcher's academic specialization and interests remains a pivotal challenge in scholar recommendation systems. To optimize recommendations, these systems must comprehensively analyze multidimensional scholar attributes, including but not limited to publication performance (Kong et al. 2018), historical collaborations (Yang et al. 2018), institutional affiliations (Dong et al. 2022), and participation in scholarly events (Desai et al. 2023). Furthermore, they should account for the dynamic evolution of scholarly networks over time (Nie et al. 2020).

Random walk as a stochastic simulation method is widely used in scholar recommendation systems because of its effectiveness in exploring multifaceted academic connections (Xia et al. 2019). Traditional single-layer random walk approaches exhibit inherent limitations (Cai et al. 2024), where some permanently isolated scholars never get recommended. Two main strategies have been proposed to address these issues. The first strategy incorporates random restart simulations, allowing jumps to similar or arbitrary nodes with predefined probabilities (Pan et al. 2004), yet such heuristic solutions often lack strong theoretical support and yield suboptimal recommendation outcomes. The second method employs layer compression through weighted aggregation of multiple attribute networks (Xu et al. 2019), but this process fundamentally struggles to fully explore the essential heterogeneity across distinct scholarly interaction dimensions.

We propose a Multilayer Random Walk simulation framework (MLRW) for scholar recommendation systems, where each layer simulates distinct collaboration patterns within heterogeneous academic networks. Through inter-layer transition mechanisms, the model enables systematic exploration of cross-dimensional scholarly interactions while preserving the dynamic nature of academic collaborations in each layer. The

Multilayer Random Walk surpasses the single-layer Random Walk not only in recommendation coverage but also in discovering more novel collaborators with higher precision.

Beyond historical collaboration networks, which is the conventional foundation for random walk approaches, our quantitative analysis identifies institutional affiliation as another statistically significant predictor of scholarly collaboration patterns. Furthermore, as demonstrated by Wang et al. (2019), conference co-attendance serves as another reliable predictor of future collaborations. To efficiently synthesize these multidimensional interaction modalities, we implement a tri-layer network architecture, integrating institutional affiliation, historical co-authorship, and conference co-attendance layers. Besides, this framework employs a weight calibration mechanism via softmax normalization to dynamically prioritize network influences during cross-layer transitions. Empirical evaluations demonstrate the model's efficacy in concurrently improving both recommendation accuracy and novel collaborators discovery, successfully identifying a substantially greater number of potential collaborators beyond existing co-authorship relationships.

This study makes three key contributions to scholarly collaborator recommendation systems:

1. We present a novel simulation method using Multilayer Random Walk framework that systematically captures the heterogeneous nature of academic interactions through cross-layer exploration, overcoming the limitations of conventional homogeneous network representations.
2. We establish institutional affiliation as a statistically significant collaboration patterns, which motivates our inclusion of Affiliation networks as a distinct layer in the recommendation framework.
3. Our comprehensive evaluation on the *SciSciNet* dataset, a large-scale scholarly knowledge base containing rich collaboration metadata, demonstrates that our method significantly outperforms conventional random walk baselines across multiple academic disciplines, achieving substantial improvements in both accuracy and novelty metrics.

2 RELATED WORKS

2.1 Collaborator Recommendation

Collaborator recommendation systems have received considerable attention in the academic research community, with various methodologies to improve the effectiveness of these systems. The methods can be divided into three types: Content-Based Filtering (CBF), Collaborative Filtering (CF), and Hybrid.

In the context of scholar collaborator recommendation, **Content-Based Filtering (CBF)** evaluates semantic similarity among researchers' academic profiles, disciplinary domains, and research interests. This approach models user preferences through their interaction patterns, derived from authorship (Sugiyama and Kan 2010), paper collections (Jack 2012), social tags (Ferrara et al. 2011), or activities like downloading (Pennock et al. 2013), reading (Yang et al. 2009), and browsing papers (Bollen and Van de Sompel 2006). Typical implementations employ the Vector Space Model (VSM) (Yukawa et al. 2001), topic clustering models (Afzal and Maurer 2011), and heuristic greedy algorithm (Yang et al. 2014), which enable personalized recommendations with little upfront classification effort. However, this approach exhibits limitations in adequately evaluating item quality assessment and popularity, demonstrating an overspecialization tendency that favors items highly similar to known preferences (Ricci et al. 2011).

Collaborative filtering (CF) techniques identify potential collaborators who have worked with a target author's co-author but lack direct collaboration with the target author himself. This approach builds upon co-authorship network analysis derived from publication records, subsequently driving significant advancements in link prediction algorithms and edge weighting techniques. Koh and Dobbie (2012) introduce a sociability-based weighted association rule for co-authorship networks in academic collaborator recommendation systems. Besides, the random walk model (Li et al. 2014) is also a popular approach in network-based recommendations due to its ability to quantify the recommendation confidence.

Compared to CBF methods, CF offers advantages such as content independence, reliance on human quality assessments (Torres et al. 2004; Dong et al. 2009), and serendipitous recommendations due to user-based rather than item-based similarity (McNee et al. 2006). However, CF also faces challenges, including user participation dependence and data sparsity, which can hinder recommendation accuracy.

Recognizing these limitations, **Hybrid models**, which combine the strengths of both CBF and CF, have taken up over half of studies in the collaborative cooperation area (Zhang et al. 2023). These hybrid approaches typically operate on heterogeneous networks, integrating multiple data dimensions, such as the profiles of researchers, the results of topic modeling or clustering, and the citation relationship between researchers and their published papers.

The following section reviews the random-walk-based simulation methods and their integration with other methods, such as Topic Clustering or weighting indices in the collaborator recommendation field. This combination enables the simultaneous consideration of user-item interactions and item quality metrics.

2.2 Random Walk in the Collaborator Recommendation

Random walks, particularly Random Walk with Restart (RWR) and Personalized PageRank (PPR), have been used to improve collaborator recommendation by simulating the cooperation pattern and exploring the underlying academic network. These algorithms simulate a researcher's exploration of the network by randomly traversing the graph, where the walk is biased towards nodes (researchers) with similar attributes, such as common publications, research interests or citations.

Xia et al. (2014) presented MVCWalker, a random walk-based model which defines the importance of collaboration links based on academic factors such as co-author order, latest collaboration time, and times of collaboration. Kong et al. (2016) combined the topic clustering model on researchers' publications with a scholar collaboration network using the RWR model. While the method is effective in recommending similar scholars, it may not fully address cross-community recommendations, which could be important for enhancing the diversity of recommendations. Wang et al. (2019) attempted to address this gap by introducing the SCORE method based on RWR, which leverages the concept of *conference closure* to identify potential collaborators based on weak ties formed through shared conference attendance. Zhou et al. (2018) integrated various relationships—researcher-researcher, researcher-article, and article-article—while considering both collaboration history and the academic impact of researchers' work. Although effective in identifying researchers who do not know each other, this model may blur the importance of collaboration and citation relationships.

Despite their effectiveness, the existing works mainly focus on the existing cooperation records and academic features, but ignoring the latent correlation among scholars or struggling with designing weighting mechanism for diverse features. Therefore, we aim to propose the Multilayer Random-Walk-based simulation model.

3 DATA

3.1 Dataset

In order to test the ability of Multilayer Random Walk simulation framework on diverse fields, we choose *SciSciNet* (Lin et al. 2023), a large-scale real-world open data lake for the science of science research as the preliminary dataset. *SciSciNet* includes 134,129,188 publications, 134,197,162 authors, 26,998 institutions, 49,066 journals, 4,551 conference series, 311 fields of study, and the internal links between them. We selected three field datasets among the 311 fields to conduct the further experiments. The detailed information of the datasets is shown in Table 1.

Table 1: Information of the three fields datasets.

	Nodes	Edges	Affiliations	Conferences
Regional Science	13095	15488	1756	357
Industrial Engineering	19347	23807	2013	5316
Economic Geography	14275	19605	1680	283

3.2 Analysis on Same Affiliation Cooperation

Wuchty et al. (2007) noted that geographical proximity for resource sharing and administrative barriers jointly determine the topological structure of collaboration networks. We confirm this perspective by conducting quantitative analysis of 32,603,511 papers across 311 disciplines using the *SciSciNet* dataset. For each paper, when above 50 percent of the authors belong to one affiliation, we assume it as a paper cooperated inside the same affiliation.

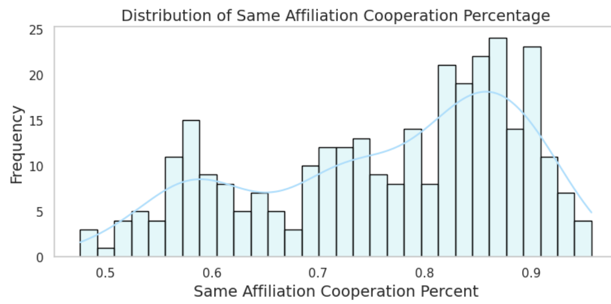

Figure 1: The distribution of the percentage of collaborators affiliated to the same affiliation across 311 fields.

The results in Figure 1 show that the average same-affiliation collaboration percentage accounts for as high as 76%, with only 3 out of 311 disciplines having same-affiliation collaboration percentage below 50%. In particular, experiment-intensive disciplines such as medicinal chemistry tends to show a prominent trend with a same-affiliation percentage of 0.95. This localization tendency aligns with the research by Agrawal & Henderson Zilber (2002) on the efficiency of tacit knowledge transfer, indicating that the physical proximity of laboratory equipment sharing and team collaboration remains a core driver of academic cooperation.

The above quantitative analysis on 311 fields provides the insight that scholars in the same affiliation tend to cooperate with each other, suggesting to take the institutional affiliation into consideration in collaborator recommendation.

3.3 Analysis on the H-index Differences Between Collaborators

In academic collaboration networks, high-impact researchers often occupy core positions, attracting collaboration from other scholars (Leydesdorff and Wagner 2008). Scholars exhibit a *path dependence* tendency when selecting collaborators, favoring those with whom they have recently collaborated—particularly high-performing scholars who have highly cited papers (Ying et al. 2024). We examine this pattern by analyzing the distribution of pairwise H-index for collaborators in the above three fields as shown in Figure 2.

Across all three fields, we observe a general trend where authors with higher H-indices tend to collaborate with co-authors who also have higher H-indices. Specifically, as the author's H-index increases, the median H-index of their co-authors generally rises. Moreover, the inter-quartile range also tends to shift upwards, indicating that authors with higher impact tend to collaborate with peers who also possess higher H-indices on average. Scattered high H-index co-authors for low H-index authors may reflect mentorship relationships or cross-career-stage collaborations, while low H-index outliers for high H-index authors could indicate the inclusion of early-career researchers in teams.

Therefore, the pattern of assortative mixing by academic impact strongly suggests that high-impact researchers preferentially collaborate with other high-impact researchers, aligning with the observation that scholars are drawn to collaborate with high-performing peers. This points out the importance of connecting to high H-index scholar in real-world collaborator recommendation.

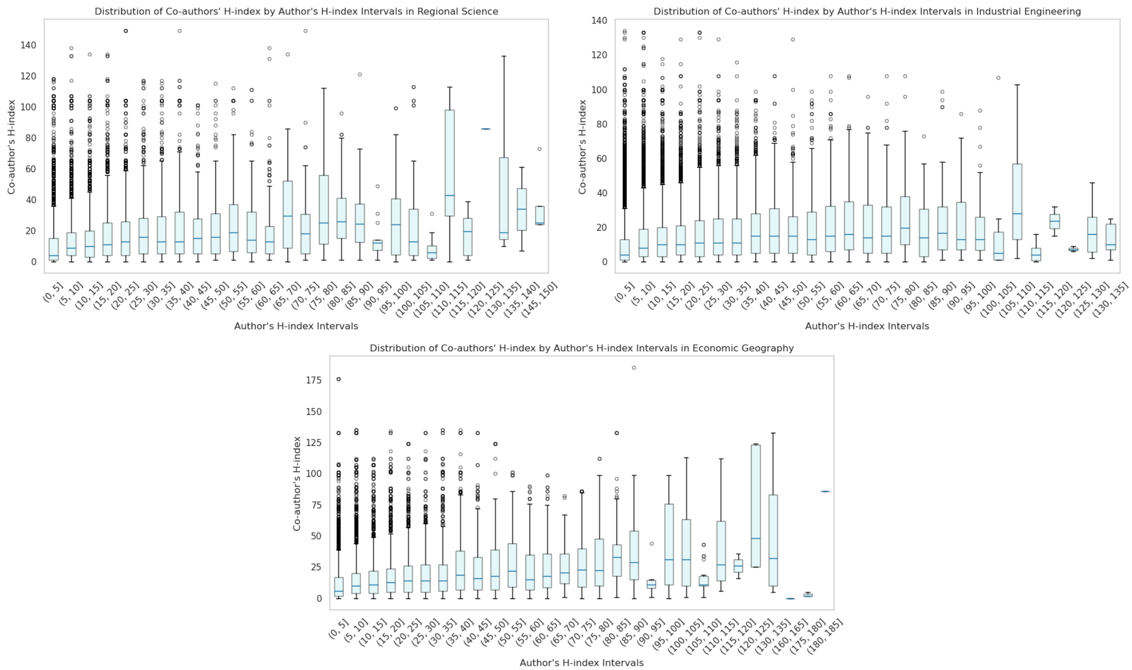

Figure 2: The distribution of pairwise H-index across three Fields.

4 METHODOLOGY

4.1 Multilayer Random Walk

Multilayer Random Walk is an advanced simulation method designed for exploring and analyzing the intricate structures of multilayer networks, which consist of multiple interconnected single-layer networks that may contain different types of nodes and edges. Besides intra-layer edges that connect nodes within the same layer, there can also be inter-layer edges that connect nodes across different layers. During a random walk, the multilayer walker can move between nodes within the same layer or jump to a node in a different layer, following specific transition probabilities. The fundamental idea is to perform random walks simultaneously across these multiple layers, while simulating different types of interactions in each layer. The transition probability of moving from one node to another is a weighted combination of the probabilities across all layers, which allows the model to reflect the multifaceted nature of academic collaboration.

Denote a multilayer network as $\boldsymbol{G} = \left\{G^{(1)}, G^{(2)}, \ldots, G^{(K)}\right\}$, where each layer $G^{(\alpha)}$ is represented by $G^{(\alpha)} = (V^{(\alpha)}, E^{(\alpha)})$, where K denotes the total number of layers. The node set is $V^{(\alpha)} = \{v_i^{(\alpha)} \mid i = 1, 2, \ldots, N\}$, and the link set is $E^{(\alpha)} = \{e_j^{(\alpha)} \mid j = 1, 2, \ldots, M\}$. Since users vary across different social networks, G comprises the same users across all layers. Formally, we can express multilayer networks as $\boldsymbol{G} = \{V, E^{(1)}, E^{(2)}, \ldots, E^{(K)}\}$, where $V = V^{(1)} = V^{(2)} = \cdots = V^{(K)}$.

The adjacency matrix of layer $G^{(\alpha)}$ is denoted as $A^{(\alpha)}$ in binary form, and $a_{i,j}^{(\alpha)} \in A^{(\alpha)}$ provides the neighborhood status of nodes v_i and v_j in $G^{(\alpha)}$. The intra-layer probability of jumping from the current v_i to a neighboring v_j within the same layer $G^{(\alpha)}$ is:

$$p_{i,j}^{(\alpha)} = w_{i,j}^{(\alpha)} a_{i,j}^{(\alpha)}, \tag{1}$$

where $w_{i,j}^{(\alpha)}$ denotes the edge weight between v_i and v_j in $G^{(\alpha)}$. The inter-layer probability of jumping from the current v_i in layer G^{α} to the neighboring v_j in G^{β} is:

$$p_{i,j}^{(\alpha,\beta)} = p_{i,j}^{(\alpha)}\ q^{(\alpha,\beta)}, \tag{2}$$

where $q^{(\alpha,\beta)}$ denotes the jumping probability from layer $G^{(\alpha)}$ to layer $G^{(\beta)}$. Let $r(t)_i^{(\alpha)}$ be the probability of the walker at state (node) i and layer $G^{(\alpha)}$ at time t. The state and layer probability is updated with:

$$r(t+1)_j^{(\beta)} = p_{i,j}^{(\alpha,\beta)} r(t)_i^{(\alpha)} = q^{(\alpha,\beta)}\ p_{i,j}^{(\alpha)}\ r(t)_i^{(\alpha)}, \tag{3}$$

where $r(0)_l^{(1)} = 1$ and 0 elsewhere, in which the walker starts at the l-th position (the starting node) and the first layer.

The model will finally select the top s nodes with highest probability (or occurrence) in $\mathbf{r}(t)$ as the recommended collaborators for scholar l.

4.2 Transition Probability Calculation

In the random walk model, scholars are represented as nodes and the cooperation relationship between scholars is referred to as edges. For Scholar i, we have indicators including normalized H-index H_i, Productivity P_i, Average Log10C (Log for ten-year citations) C_i, and Collaboration counts CO_{ij}. To simulate the real-world cooperation trend with high academic performance scholars, we design the walker tends to select the next node with higher edge weight. For Node i,

$$\text{Node\_weight}_i = \frac{H_i + P_i + C_i}{3}, \tag{4}$$

In the Cooperation Layer ($G^{(1)}$), only the scholars who have once cooperated have an edge between each other, and the walker selects the next node based on the edge weight defined as:

$$\text{Edge\_weight}_{i,j} = \alpha \cdot \frac{\text{Node\_weight}_i + \text{Node\_weight}_j}{2} + (1-\alpha) \cdot CO_{ij}, \tag{5}$$

where α is the balancing hyperparameter. To enhance the preference for high-weight edges and normalize the transition probability distribution within the framework of multilayer networks $\boldsymbol{G}$, we reweigh the transition probability with Softmax weighting mechanism:

$$w_{i,j}^{(1)} = \frac{\exp\left(\beta \cdot \text{Edge\_weight}_{i,j} - \phi_i^{(1)}\right)}{\sum_{v_k \in N(v_i^{(1)})} \exp\left(\beta \cdot \text{Edge\_weight}_{i,j} - \phi_i^{(1)}\right)}, \tag{6}$$

where β is the bias factor ($\beta > 0$), $N(v_i^{(1)})$ denotes the neighbor set of node $v_i^{(1)}$ in $G^{(1)}$, and $\phi_i^{(1)} = \max_{v_k \in N(v_i^{(1)})} \left(\beta \cdot \text{Edge\_weight}_{i,j}\right)$ is the maximum scaling weight of all neighbors of the current node $v_i^{(1)}$ in $G^{(1)}$.

In the Affiliation Layer ($G^{(2)}$), we build edges between scholars from the same affiliation. Denote aff_i as the affiliation of node i, the transition probability is defined as:

$$w_{i,j}^{(2)} = I\left(\text{aff}_i = \text{aff}_j\right) \tag{7}$$

In the Conference Attendance Layer ($G^{(3)}$), we define the edge between the scholars i and j attending the same conference. Denote $\boldsymbol{C} = \{C_1, ..., C_D\}$, and $c_{ik} = 1$ if scholar i attends conference k and 0 if else. Then, the transition probability is defined as:

$$w_{i,j}^{(3)} = \begin{cases} 1, \text{if } \exists\ k \in \{1, ..., D\}, c_{ik} + c_{jk} = 2 \\ 0, \text{if else} \end{cases} \tag{8}$$

5 EXPERIMENTS

5.1 Evaluation Metrics

We use Precision, Recall and F1-score to evaluate the performance of our proposed methods, which are defined as the below formula. Besides these three classical evaluation metrics in collaborator recommendation (Zhang et al. 2023), we introduce two novel metrics, NewRec and NonRec to evaluate the ability of the methods on discovering novel cooperation instead of only recommending the familiar old collaborators.

$$\text{Recall} = \frac{TP}{TP+FN}, \quad \text{Precision} = \frac{TP}{TP+FP}, \quad \text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$
$$\text{NewRec} = TP - \#\text{ old collaborators}, \quad \text{NonRec} = N_{nodes} - N(TP+FP)_{nodes} \tag{9}$$

where TP denotes the number of recommended pair that actually cooperate in the test dataset, FP denotes the number of recommended pair that do not cooperate, FN denotes the number of non-recommended pair but actually cooperate, while *NewRec* evaluates the number of novel true cooperation compared with the original train dataset and *NonRec* evaluates the number of scholars who do not obtain a recommendation, remaining as an isolated node from the scholar network.

5.2 Comparison with Other Methods

We conducted experiments on three datasets of different fields extracted from the *SciSciNet* Dataset: *Regional Science*, *Industrial Engineering*, and *Economic Geography*. Each dataset is separated into training data (papers published in year 2000-2015) and test data (papers published in year 2016-2022).

In the experiments, we compare our proposed Multilayer Random Walk simulation model (MLRW) with six random-walk-based methods, including two advanced single-layer random-walk model MVCWalker (Xia et al. 2014) and SCORE (Wang et al. 2019). *Di-softmax* refers to two-layer random walk model with Cooperation Layer and Affiliation Layer, while *Tri-softmax* adds the Conference Attendance Layer, referring to three-layer random walk model. The results are shown in Table 2. All the methods have been selected the result with the highest Recall by conducting grid search on parameters.

5.2.1 Multilayer Random Walk Achieves Better Performance than Single-layer

A critical advantage of the MLRW simulation framework lies in its capacity to recommend novel collaborators, which is unachievable by single-layer baselines only based on existing cooperation records. Particularly, the substantial number of new recommendations (NewRec) generated by our models, which are largely absent in the single-layer baselines, directly contribute to the high true positive (TP) counts. This ability to recommend truly novel collaborators is crucial for practical recommendation systems, as it moves beyond simply suggesting already popular or well-connected authors. By allowing the random walk to transition between layers—moving from an author to his or her institutional colleague or conference co-attendees, it surpasses the single-layer's restriction on ordinary cooperation network. Moreover, the significant reduction in NonRec compared with RW and RWR demonstrates the ability of MLRW to connect the isolated scholars with the community. The competitive performance of both the Di-softmax and Tri-softmax further validates the scalability of our simulation framework under reduced computational complexity.

5.2.2 Multilayer Random Walk Outperforms All the Baseline Methods

As demonstrated in Table 2, the proposed MLRW simulation models significantly outperform traditional random-walk-based methods across most evaluation metrics. Specifically, the Di-softmax and Tri-softmax models respectively achieves the highest performance in three datasets. While MVCWalker and SCORE introduce valuable domain-specific heuristics, their localized and single-layer designs limit scalability and discovery. The lower performance of SCORE shows the limitation of combining the weak-tie information through edge weighting methods. Multilayer approaches transcend these constraints through cross-domain

Table 2: Performance comparison of different methods.

	Method	TP	Precision	Recall	F1score	NewRec	NonRec
Regional Science	RW	60	0.0214	0.0081	0.0117	0	7006
	RW-softmax	64	0.0228	0.0086	0.0125	0	7006
	RWR	66	0.0235	0.0089	0.0129	0	7006
	RWR-softmax	65	0.0232	0.0088	0.0127	0	7006
	MVCWalker	47	0.0048	0.0032	0.0038	0	0
	SCORE	44	0.0045	0.0059	0.0051	0	0
	Di-softmax (OURS)	**679**	**0.0731**	**0.0902**	**0.0807**	**615**	523
	Tri-softmax (OURS)	663	0.0714	0.0882	0.0789	599	523
Industrial Engineering	RW	96	0.0144	0.0095	0.0115	2	7922
	RW-softmax	94	0.0141	0.0093	0.0112	0	7922
	RWR	108	0.0162	0.0107	0.0129	0	7922
	RWR-softmax	104	0.0156	0.0103	0.0124	0	7922
	MVCWalker	74	0.0051	0.0073	0.0060	0	0
	SCORE	78	0.0053	0.0077	0.0063	0	0
	Di-softmax (OURS)	912	0.0666	0.0894	0.0763	818	882
	Tri-softmax (OURS)	**956**	**0.071**	**0.0937**	**0.0808**	**863**	1116
Economic Geography	RW	104	0.0298	0.0184	0.0227	2	8159
	RW-softmax	114	0.0326	0.0202	0.0249	0	8159
	RWR	127	0.0363	0.0225	0.0278	0	8159
	RWR-softmax	122	0.0349	0.0216	0.0267	0	8159
	MVCWalker	109	0.0094	0.0192	0.0126	0	0
	SCORE	96	0.0082	0.0170	0.0111	0	0
	Di-softmax (OURS)	**523**	**0.0468**	**0.0914**	**0.0619**	**409**	484
	Tri-softmax (OURS)	475	0.0456	0.0832	0.0589	361	1237

integration and adaptive softmax weighting, by simulating multiple distinct types of academic relationships as separate layers. The resulting inter-layer random walk captures richer interaction patterns, leading to superior accuracy. However, in terms of recommendation coverage, MLRW models are slightly inferior to MVCWalker and SCORE which report a NonRec of 0, indicating they provide a recommendation list for every scholar. This universal coverage may come at the cost of low precision and recall, while MLRW strikes a better balance, providing high-quality, novel recommendations to a much broader audience than basic methods, without sacrificing relevance. This highlights our approach's enhanced ability to identify potential, previously unobserved collaborators, fulfilling a key goal of recommendation systems in discovering truly novel and valuable research connections.

5.2.3 Softmax Shows Advantages in Original Random Walk but Limited

Softmax slightly improves the F1-score over the standard RW on two of the three datasets. Furthermore, when applied to RWR, the addition of Softmax consistently results in slightly lower F1-scores. This indicates that while Softmax can offer marginal benefits for basic RW in some cases, its effect is inconsistent and even detrimental when combined with RWR, highlighting the limitation of only adjusting edge-weighting methods compared to the significant improvements shown by our proposed methods Di-softmax and Tri-softmax.

5.3 Sensitive Experiments

5.3.1 Dilayer Random Walk

We implement two variants of the Dilayer Random Walk to conduct its sensitive experiments.

In the first implementation, both layers have the same jump probability, meaning that at each step, the likelihood of transitioning from the Collaboration Layer to the Affiliation Layer is equal to that of remaining in the current layer. Figure 3 shows the model's performance in Regional Science as *jump prob* varies from 0 to 1. These results strongly indicate that facilitating interaction between the Collaboration and Affiliation layers is highly beneficial. Increasing the frequency of jumps between layers (*jump prob*) allows the simulation model to leverage institutional information, leading to more relevant and novel recommendations, better overall performance (F1-Score), and significantly improved coverage (low NonRec). Optimal performance is achieved when the probability of switching layer is high.

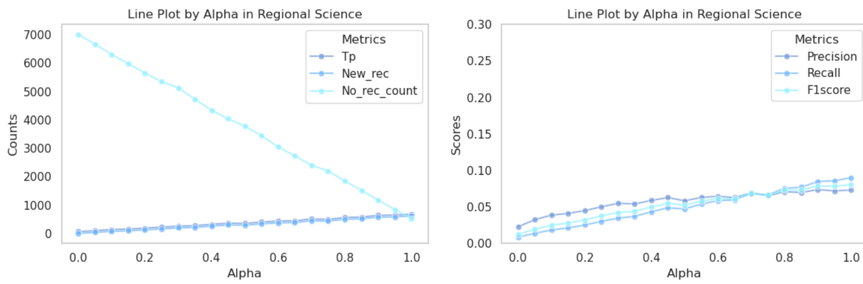

Figure 3: The sensitive experiments of Dilayer Random Walk with same layer jump probability.

The second implementation allows for different jump probabilities between layers, with α dictating the likelihood of jumping from the Collaboration Layer to the Affiliation Layer and $1-\alpha$ to jump from the Affiliation Layer to Collaboration Layer. Figure 4 plots the performance metrics as α varies from 0 to 1. Similar to the first variant, this configuration highlights the importance of utilizing the Affiliation Layer. Increasing the probability of transitioning from the collaboration network to the institutional context significantly boosts performance, novelty, and coverage.

Both sensitivity analyses for the Dilayer Random Walk consistently demonstrate the value of integrating the Affiliation Layer with the Collaboration Layer. Whether through symmetric (*jump prob*) or asymmetric (α) control, increasing the interaction with the Affiliation Layer leads to substantial improvements in recommendation relevance (TP, Precision, Recall, F1), novelty (NewRec), and coverage (NonRec). The optimal strategy involves frequent transitions that leverage institutional information, particularly prioritizing jumps into the institutional context.

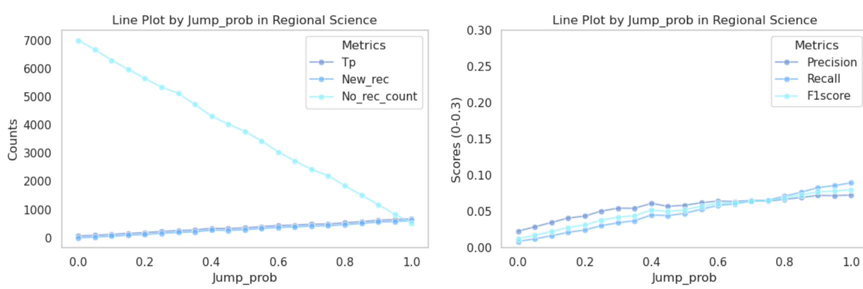

Figure 4: The sensitive experiments of Dilayer Random Walk with different layer jump probability.

5.3.2 Trilayer Random Walk

We investigate the sensitivity of the Trilayer Random Walk simulation model to two key hyperparameters: the overall probability of jumping between layers (*jump prob*) and the parameter α, which controls the directional bias of jumps involving the Affiliation Layer. Specifically, given a jump occurs α represents the probability of jumping to the Affiliation Layer from either the Conference or Cooperation Layer, while $1-\alpha$ is the probability of jumping from the Affiliation Layer to one of the other two layers. The results across various metrics are visualized as heatmaps in Figure 5. Across nearly all performance metrics, higher

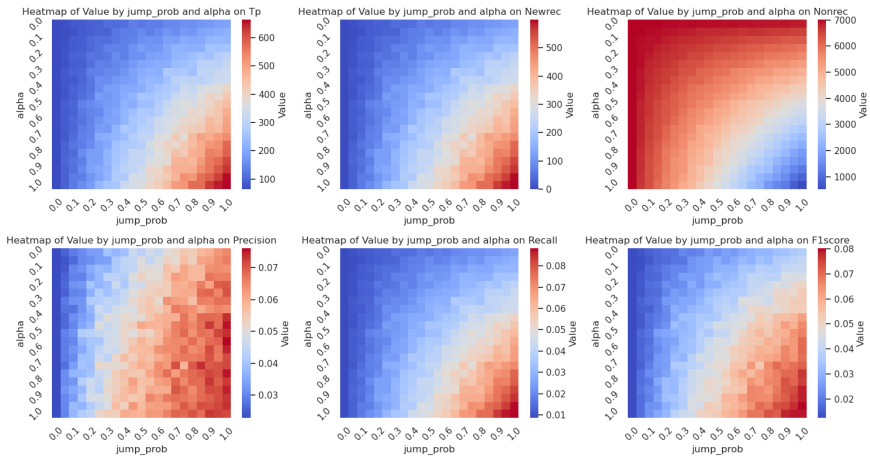

Figure 5: The sensitive experiments of Trilayer Random Walk in Regional Science dataset.

values of *jump prob* generally lead to better performance. Conversely, the number of authors receiving non recommendations (NonRec) is highest when *jump prob* is low. This indicates that enabling frequent transitions between the Conference, Cooperation, and Affiliation Layers is essential not only for improving recommendation quality and novelty but also for enhancing the model's coverage, ensuring more authors receive potentially relevant suggestions. The parameter α significantly influences performance, highlighting the role of the Affiliation Layer. The optimal regions (darkest red) for these key metrics are consistently found where α is high. Similarly, higher α leads to a marked decrease in NonRec, meaning that prioritizing jumps to the Affiliation Layer helps the model generate recommendations for a larger fraction of authors. This suggests that leveraging the institutional context as a frequent intermediary or target during the random walk is highly beneficial for uncovering relevant connections, boosting novelty, and improving overall recommendation coverage.

The heatmaps consistently demonstrate that the best overall performance is achieved when both *jump prob* and α are high. This configuration corresponds to a random walk that frequently decides to switch layers and, when transitioning from the Conference or Cooperation Layer, strongly prefers jumping to the Affiliation Layer, likely simulating the real-world cooperation pattern.

In summary, the Trilayer Random Walk's effectiveness and coverage are sensitive to both the frequency of layer switching and the specific bias towards the Affiliation Layer. The results strongly advocate for enabling frequent exploration across layers, with a particular emphasis on utilizing transitions to the institutional context as a mechanism to discover practical potential collaborators and provide recommendations for a broader set of authors, matching with the real-world cooperation pattern we found in the quantitative analysis on the datasets.

6 CONCLUSION

In this paper, we proposed a novel Multilayer Random Walk simulation framework (MLRW) for scholar collaborator recommendation systems. This framework systematically simulates the heterogeneous nature of academic interactions by integrating multiple layers of scholarly networks, including cooperation, institutional affiliation, and conference attendance. We demonstrated the statistical significance of institutional affiliation as a collaborative pattern, motivating its inclusion. Comprehensive empirical evaluations on the *SciSciNet* dataset confirmed that our proposed multilayer random-walk-based simulation method significantly outperforms conventional random walk baselines across multiple disciplines, yielding substantial gains in both accuracy and novelty metrics. These findings highlight the efficacy of the multilayer approach in identifying relevant yet potentially non-obvious collaborators, thereby enhancing recommendation coverage and practical value. As future work, we suggest incorporating further interaction layers and refining the transition weighting strategies to potentially boost recommendation diversity and better simulate real-world cooperation pattern.

ACKNOWLEDGMENTS

This research was funded by the Tencent (Shenzhen) Co., Ltd. (Tencent University Cooperation Project No. 20239680152) and the Guangdong Pearl River Plan (2019QN01X890).

REFERENCES

Afzal, M. T., and H. A. Maurer. 2011. "Expertise Recommender System for Scientific Community.". *J. Univers. Comput. Sci.* 17(11):1529–1549.

Bollen, J., and H. Van de Sompel. 2006. "An Architecture for the Aggregation and Analysis of Scholarly Usage Data". In *Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries*, 298–307.

Cai, W., X. Chang, and P. Yang. 2024. "Link Prediction in Multilayer Social Networks Using Reliable Local Random Walk and Boosting Ensemble Classifier". *Chaos, Solitons & Fractals* 188:115530.

Desai, M., R. G. Mehta, and D. P. Rana. 2023. "ScholarRec: A Scholars' Recommender System That Combines Scholastic Influence and Social Collaborations in Academic Social Networks". *International Journal of Data Science and Analytics* 16(2):203–216.

Dong, R., L. Tokarchuk, and A. Ma. 2009. "Digging Friendship: Paper Recommendation in Social Network". In *Proceedings of Networking & Electronic Commerce Research Conference (NAEC 2009)*, 21–28.

Dong, W., H. Xiong, J. Du, and N. Wang. 2022. "Recommending Research Collaborators Based on Scholar Profiling". *Data Analysis and Knowledge Discovery* 6(10):20–34.

Ferrara, F., N. Pudota, and C. Tasso. 2011. "A Keyphrase-based Paper Recommender System". In *Digital Libraries and Archives: 7th Italian Research Conference, IRCDL 2011, Pisa, Italy, January 20-21, 2011. Revised Papers 7*, 14–25. Springer.

Jack, K. 2012. "Mahout Becomes a Researcher: Large Scale Recommendations at Mendeley". In *Presentation at big data week conferences*. https://www.slideshare.net/slideshow/mahout-becomes-a-researcher-large-scale-recommendations-at-mendeley/12697911, accessed 5$^{\text{th}}$ May 2025.

Koh, Y. S., and G. Dobbie. 2012. "Indirect Weighted Association Rules Mining for Academic Network Collaboration Recommendations". In *Proceedings of the Tenth Australasian Data Mining Conference-Volume 134*, 167–173.

Kong, X., H. Jiang, Z. Yang, Z. Xu, F. Xia, and A. Tolba. 2016. "Exploiting Publication Contents and Collaboration Networks for Collaborator Recommendation". *PloS one* 11(2):e0148492.

Kong, X., M. Mao, J. Liu, B. Xu, R. Huang, and Q. Jin. 2018. "TNERec: Topic-aware Network Embedding for Scientific Collaborator Recommendation". In *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, 1007–1014. IEEE.

Leydesdorff, L., and C. S. Wagner. 2008. "International Collaboration in Science and the Formation of a Core Group". *Journal of informetrics* 2(4):317–325.

Li, J., F. Xia, W. Wang, Z. Chen, N. Y. Asabere, and H. Jiang. 2014. "Acrec: A Co-authorship Based Random Walk Model for Academic Collaboration Recommendation". In *Proceedings of the 23rd international conference on world wide web*, 1209–1214.

Lin, Z., Y. Yin, L. Liu, and D. Wang. 2023. "SciSciNet: A large-scale Open Data Lake for the Science of Science Research". *Scientific Data* 10(1):315.

McNee, S. M., N. Kapoor, and J. A. Konstan. 2006. "Don't Look Stupid: Avoiding Pitfalls when Recommending Research Papers". In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, 171–180.

Nie, H., X. Chen, X. Chu, W. Wang, Z. Xu, and F. Xia. 2020. "Collaborator Recommendation Based on Dynamic Attribute Network Representation Learning". In *2020 7th International Conference on Behavioural and Social Computing (BESC)*, 1–6. IEEE.

Pan, J.-Y., H.-J. Yang, C. Faloutsos, and P. Duygulu. 2004. "Automatic Multimedia Cross-modal Correlation Discovery". In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 653–658.

Pennock, D. M., E. J. Horvitz, S. Lawrence, and C. L. Giles. 2013. "Collaborative Filtering by Personality Diagnosis: A Hybrid Memory and Model-based Approach". *arXiv preprint arXiv:1301.3885*.

Ricci, F., L. Rokach, B. Shapira, and P. B. Kantor. (Eds.) 2011. *Recommender Systems Handbook*. Berlin: Springer.

Sugiyama, K., and M.-Y. Kan. 2010. "Scholarly Paper Recommendation Via User's Recent Research Interests". In *Proceedings of the 10th annual joint conference on Digital libraries*, 29–38.

Torres, R., S. M. McNee, M. Abel, J. A. Konstan, and J. Riedl. 2004. "Enhancing Digital Libraries with TechLens+". In *Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries*, 228–236.

Wang, W., J. Liu, Z. Yang, X. Kong, and F. Xia. 2019. "Sustainable Collaborator Recommendation Based on Conference closure". *IEEE Transactions on Computational Social Systems* 6(2):311–322.

Wuchty, S., B. F. Jones, and B. Uzzi. 2007. "The Increasing Dominance of Teams in Production of Knowledge". *Science* 316(5827):1036–1039.

Xia, F., Z. Chen, W. Wang, J. Li, and L. T. Yang. 2014. "MVCWalker: Random Walk-based Most Valuable Collaborators Recommendation Exploiting Academic Factors". *IEEE Transactions on Emerging Topics in Computing* 2(3):364–375.

Xia, F., J. Liu, H. Nie, Y. Fu, L. Wan, and X. Kong. 2019. "Random Walks: A Review of Algorithms and Applications". *IEEE Transactions on Emerging Topics in Computational Intelligence* 4(2):95–107.

Xu, Z., Y. Yuan, H. Wei, and L. Wan. 2019. "A Serendipity-biased Deepwalk for Collaborators Recommendation". *PeerJ Computer Science* 5:e178.

Yang, C., T. Liu, L. Liu, and X. Chen. 2018. "A Nearest Neighbor Based Personal Rank Algorithm for Collaborator Recommendation". In *2018 15th international conference on service systems and service management (ICSSSM)*, 1–5. IEEE.

Yang, C., J. Ma, J. Sun, T. Silva, X. Liu, and Z. Hua. 2014. "A Weighted Topic Model Enhanced Approach for Complementary Collaborator Recommendation". In *PACIS 2014 Proceedings*, 297.

Yang, C., B. Wei, J. Wu, Y. Zhang, and L. Zhang. 2009. "CARES: A Ranking-oriented CADAL Recommender System". In *Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries*, 203–212.

Ying, L., C. Huang, and X. Liu. 2024. "Comparative Analysis of Helicobacter Pylori Infection Status in High Altitude and Plain Areas of China". *Journal of Biosciences and Medicines* 12(12):109–123.

Yukawa, T., K. Kasahara, T. Kato, and T. Kita. 2001. "An Expert Recommendation System Using Concept-based Relevance Discernment". In *Proceedings 13th IEEE International Conference on Tools with Artificial Intelligence. ICTAI 2001*, 257–264. IEEE.

Zhang, Z., B. G. Patra, A. Yaseen, J. Zhu, R. Sabharwal, K. Roberts, *et al*. 2023. "Scholarly Recommendation Systems: A Literature survey". *Knowledge and Information Systems* 65(11):4433–4478.

Zhou, X., W. Liang, I. Kevin, K. Wang, R. Huang, and Q. Jin. 2018. "Academic Influence Aware and Multidimensional Network Analysis for Research Collaboration Navigation based on Scholarly Big Data". *IEEE Transactions on Emerging Topics in Computing* 9(1):246–257.

Zilber, T. B. 2002. "Institutionalization as an Interplay Between Actions, Meanings, and Actors: The Case of a Rape Crisis Center in Israel". *Academy of management journal* 45(1):234–254.

AUTHOR BIOGRAPHIES

SIYU CHEN is a master student of the Shenzhen International Graduation School (SIGS), Tsinghua University, China. Her email address is chen-sy23@mails.tsinghua.edu.cn.

KENG HOU LEONG is a master student of the Shenzhen International Graduation School (SIGS), Tsinghua University, China. His email address is liangjh22@mails.tsinghua.edu.cn.

JIADONG LIU is a master student at the Shenzhen international Graduate Schoolol (SIGS), Tsinghua University, China. His email address is ljd23@mails.tsinghua.edu.cn.

WEI CHEN is a D.Eng candidate at Shenzhen International Graduate School (SIGS), Tsinghua University, China. He currently serves as a Director at Tencent. Additionally, Wei is the founder and director of Wizsci (wizsci.com), a scientific research platform dedicated to advancing academic-industry collaboration. His email address is cw.chen@163.com.

WAI KIN (VICTOR) CHAN is Professor of the Shenzhen International Graduation School (SIGS), Tsinghua University, China. His research interests include discrete-event simulation, agent-based simulation, big-data analytics and their applications in social networks, business big-data, service systems, healthcare, transportation, energy markets, and manufacturing. His email address is chanw@sz.tsinghua.edu.cn.