Using GAT Structure to Analyze the Propagation Efficiency of Technology Skills Transfer Networks Among Students in Industry-Academia Collaboration

Shaobin Zhang[1], Shuzhen Wang[2]
[1]Institute of Higher Education, Luoyang Institute of Science and Technology, Luoyang 471023, Henan, China
[2]Development and Planning Department, Luoyang Institute of Science and Technology, Luoyang 471023, Henan, China
Corresponding author: Shuzhen Wang (e-mail: wang65928321@126.com)

ABSTRACT Current studies on technology skills transfer mainly rely on static network models, which have limited capability to characterize the dynamic influence of student heterogeneity on transfer pathways, especially in highly specialized and practice-oriented disciplines. This limitation reduces the accuracy of identifying key skills and efficient propagation paths, thereby constraining precision-oriented industry–academia collaborative training. To address this issue, this paper proposes a Graph Attention Network (GAT)-based framework for analyzing technology skill transfer efficiency. Based on the real learning trajectories of 328 intelligent manufacturing students, a student–skill bipartite network is constructed and projected into a temporally constrained directed transfer graph to preserve the sequential and directional characteristics of skill acquisition. A two-layer GAT with a multi-head attention mechanism is then employed to dynamically learn transfer weights among skills, enabling the identification of asymmetric transfer patterns and heterogeneous learning preferences across different student backgrounds. Experimental results demonstrate a Hit@3 prediction accuracy of 78.4%, while high-propagation skills are primarily concentrated in composite competencies such as system integration and fault diagnosis. The proposed framework establishes an interpretable closed loop integrating dynamic attention modeling, propagation efficiency quantification, and educational intervention, providing measurable and operational support for curriculum optimization and personalized training. Moreover, the graph-based propagation modeling strategy offers methodological insights for intelligent information propagation and graph signal representation in complex engineering systems, with potential relevance to advanced electromagnetic information processing and interdisciplinary learning networks.

INDEX TERMS graph attention network, skill transfer, propagation efficiency, graph signal processing, intelligent information propagation

I. INTRODUCTION

UNDER the national strategy of deepening the integration of industry and education and building modern industrial colleges [1], the process of students' acquisition and transfer of technical skills is increasingly showing complex characteristics of networking, dynamism, and high heterogeneity [2]. Skills no longer exist in isolation, but form a cross-domain transfer chain with clear directionality and time dependence under the multiple coupling effects of real project drive, curriculum system connection and job ability requirements [3]. Whether students can efficiently complete key leaps such as programming of programmable logic controllers (PLC) to industrial robot integration [4] and programming of Python to digital twin modeling directly determines their competence in high-level positions such as intelligent manufacturing production line debugging and system integration [5]. Therefore, accurately depicting the transmission mechanism of skill transfer [6] has become the core premise for optimizing the curriculum dependency structure, dynamically configuring dual-teacher resources and realizing personalized training [7]. Current educational data mining research mainly adopts three paradigms to model skill transmission: Social Network Analysis (SNA) [8], centrality index (such as PageRank) and Graph Convolutional Network (GCN). Early SNA methods construct undirected skill networks by co-occurrence frequency [9].

Although they can identify high-frequency skills, they completely ignore the direction of transfer and temporal logic. Although PageRank-type methods introduce steady-state importance assessment, they still regard students as a homogeneous group [10]. They cannot reflect the individual transfer preferences of mechanical background students who tend to transfer from mechanical design to CAE (Computer Aided Engineering) simulation, while software background students prefer Python to data visualization [11]. Recently, although GCN models generate skill embeddings through neighborhood aggregation to predict learning paths, their homogeneous aggregation mechanism treats all neighbor skills equally and cannot distinguish the semantic asymmetry between basic to application and application to basic [12-14]. This leads to the failure of modeling the importance of transfer direction [15]. To improve modeling accuracy, some studies have attempted to introduce graph structure priors. For example, Huang Q et al. used graph neural networks to predict MOOC learning paths and generated node embeddings through neighborhood aggregation [16]; Saqr M et al. introduced random walks to assess skill centrality and identify high-frequency skill nodes [17]. However, the fundamental limitation of these methods is that they lack the ability to dynamically learn the weights of transfer edges [18]. Their attention mechanisms are either static co-occurrence statistics or homogeneous neighborhood aggregation, which makes it difficult to dynamically characterize the differential dependence of target skills on specific source skills [19]. At the same time, most models do not introduce counterfactual reasoning mechanisms [20], which makes it impossible to quantify the causal contributions between skills [21], resulting in the critical paths identified being severely affected by co-occurrence bias. As a result, the path predictions and centrality indicators output by the models deviate from real educational behavior and are difficult to support precise teaching interventions [22]. To address the aforementioned challenges, this paper proposes a dynamic analysis framework for skill transfer and propagation efficiency based on GAT (Ground-Aspect-Based Attention). The core design logic lies in using educational causality as a constraint and a dynamic attention mechanism to jointly model the directionality, heterogeneity, and contextual dependence of transfer paths. Specifically, firstly, the temporal sequence and direction of transfer are preserved to construct a directed transfer graph; secondly, a cross-attention mechanism is introduced, using the target skill as the query and the source skill as the key/value, to dynamically calculate transfer weights, enabling the model to characterize the likelihood of skill transfer from students with different backgrounds; furthermore, by combining counterfactual reasoning, two quantifiable indicators—local propagation efficiency and skill dissemination potential—are defined, transforming the model output into educational indicators with behavioral interpretability. This framework aims to construct an interpretable closed loop of data modeling—indicator quantification—educational intervention, providing interpretable, measurable, and operable decision support for curriculum optimization and personalized teaching intervention in the context of industry-academia integration.

II. RESEARCH METHODS

Construction of Student-Skill Dynamic Network

To support subsequent dynamic modeling of the efficiency of skills transfer and dissemination, this paper first constructs a student-skill network structure that combines educational authenticity, temporal rationality, and directional clarity. The design logic of this structure stems from the essential characteristics of skills transfer: transfer has a clear sequence and individual selectivity (students with different professional backgrounds tend to take different paths), thus it must transcend the static and undirected assumption of traditional co-occurrence networks. Specifically, 328 full-time higher vocational students from a national-level intelligent manufacturing industry college in 2022–2024 were selected as the research subjects. Their grades in 23 core professional courses and logs of 12 school-enterprise joint projects were collected during the training period. This study was approved by the Ethics Committee of Luoyang Institute of Science and Technology, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form. The skills system was based on the "Vocational Education Professional Catalog (2021)", "National Vocational Technical Skills Standards for Intelligent Manufacturing Engineering Technicians" and the job competency lists of cooperative enterprises such as ABB, Siemens, and Haier Smart Research Institute. After three rounds of expert review by the Delphi method [23], 42 core technical skills were finally determined, covering four major categories: intelligent manufacturing (18 items), information technology (12 items), electromechanical control (8 items) and project management (4 items), as shown in Table 1.

TABLE 1. Distribution of Skill Categories (42 items)

Category	Number of Skills	Representative Skills (ID)
Intelligent Manufacturing	18	Industrial Robot Operation (S03), Digital Twin Modeling (S06), Fault Diagnosis (S07), SCADA (Supervisory Control and Data Acquisition) (S09), System Integration (S11)
Information Technology	12	Edge Computing (S25), Machine Vision (S28)
Electromechanical Control	8	Sensor Integration (S34), PID (Proportional–Integral–Derivative) Tuning (S36)
Management	4	Project Management (S39)

Skill numbers S01–S42 are unique identifiers, assigned in logical order within the category (Intelligent Manufacturing S01–S18, Information Technology S19–S30, Electromechanical Control S31–S38, Project Management S39–S42), used to support graph network modeling and result traceability.

Table 1 shows the category distribution and representative skills of the 42 skills. The skill system is based on the "Vocational Education Professional Catalog (2021)", the "National Vocational Technical Skill Standards for Intelligent Manufacturing Engineering Technicians", and the job competency list of cooperating enterprises. It was determined through three rounds of Delphi expert review and covers four major categories: intelligent manufacturing (18 items, S01–S18), information technology (12 items, S19–S30), electromechanical control (8 items, S31–S38), and project management (4 items, S39–S42). The selection principles for representative skills are:

(1) high relevance in the curriculum system;
(2) high frequency of requirements in enterprise positions;
(3) typicality and educational interpretability.

For example, the intelligent manufacturing category selects core skills that reflect production line integration capabilities, such as industrial robot operation (S03), digital twin modeling (S06), and system integration (S11); the information technology category includes data intelligence basic capabilities such as edge computing (S25) and machine vision (S28). All skill numbers S01–S42 are assigned sequentially by category, serving as unique identifiers for nodes in the graph network and supporting subsequent modeling and analysis. Data fusion employs a multi-source strategy: course scores $\geq$80 (out of 100) are considered mastery; skill proficiency levels in project logs are assessed by corporate mentors using a 5-level Likert scale (1 = not mastered, 5 = proficient application), and after double-blind review, scores $\geq$3 are considered mastery. The two types of data are weighted and fused after Min-Max normalization to generate an initial mastery confidence level (range 0–1). To support the binary determination of skill mastery status, this paper introduces an independent validation set for threshold optimization. The overall data is divided into a training set (230 students), a validation set (49 students), and a test set (49 students) in a 7:1.5:1.5 ratio based on student IDs. The validation set is used for hyperparameter tuning and mastery threshold selection, while the test set is used only for final evaluation to prevent data leakage. On the validation set, a binary classification task was constructed using the enterprise mentor score of $\geq$4 in the student's subsequent project as the true label (positive sample) of skill mastery, and the optimal decision threshold was determined to achieve the best balance between sensitivity and specificity. The threshold selection method refers to the general binary classification model evaluation criteria [24]. To preserve the causal time sequence of the transfer, a time constraint is further introduced: only the skill records mastered by students within 12 months before graduation are retained [25] to avoid interference from long-term skills on the recent transfer path. Next, this sequence is represented as directed edges in chronological order. The directed edges are referred to as a skill transfer graphs and include a total of 1,247 edges, which are used as the input into the graph attention network. The 12-month time window was selected to provide a complete representation of student records related to the time period that they are currently learning; thus, the directed edges are filtered down to represent only those items that were more recently mastered and therefore exhibit causal and temporal sequential relationships. This allows the constructed skill transfer graph to align with the temporal assumptions underlying education-related events.. Subsequently, the skill mastery sequence of each student is converted into directed edges in chronological order, and finally projected into a skill transfer graph containing 1,247 directed edges [26]. This directed graph serves as the input structure of the GAT model, and its directionality ensures the semantic distinguishability of the basic to application class path, laying the topological foundation for the subsequent attention mechanism to learn asymmetric transfer weights. The topological structure of the network and the distribution of high-weight edges are shown in Figure 1.

Figure 1 shows the topology and edge weight distribution of the skill transfer network selected from the original 1,247 directed migration edges, comprising 42 skill nodes and 226 significant migration paths. The two sets of edges – a set of all 1,247 edges, and a filtered subset of the 226 most significant edges – fulfill different functions within the analysis process. The complete edge dataset will be fed into the Graph Attention Network, allowing the model to determine full transfer weights. Meanwhile, the Filtered Edge data set has been filtered to highlight significant migration trends from the Global Perspective, and used to both visually represent the macro-topological characteristics associated with these trends, and to analyse them as well. Nodes are colored according to four main categories. The network structure reveals numerous cross-domain connections between intelligent manufacturing and information technology skills, reflecting the trend of knowledge integration in the industry-academia collaboration curriculum. The weight distribution of these 226 edges has a mean of 0.54 and a median of 0.54, indicating that the overall migration intensity is at a medium-to-high level and the distribution is relatively uniform without extreme skewness.

Transfer Weight Learning Model Based on GAT

After obtaining the directional skill transfer graph, this paper needs to learn a dynamic transfer weight for each directed edge i to j to reflect the transfer influence of skill i on skill j. The weight must meet two educational scenario requirements: first, directional sensitivity (such as PLC programming to human-machine interface (HMI) configuration should be stronger than the reverse path), and second, semantic rationality (transfer between similar skills should be stronger than weak association path across categories). Traditional graph convolutional networks (GCNs)

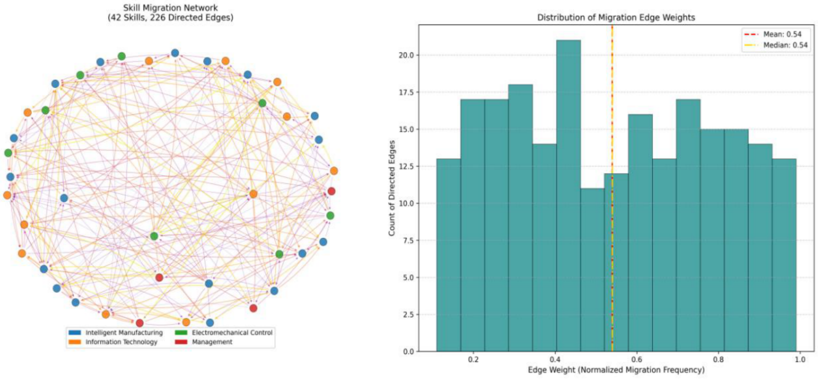

FIGURE 1. Skill transfer network topology and edge weight distribution

use homogeneous neighborhood aggregation, which cannot distinguish the direction and semantic differences of edges. Therefore, this paper selects GAT [27], which independently assigns weights to each edge through a learnable attention mechanism, naturally supporting asymmetric modeling. To make attention computation educational, the initial features of nodes are integrated with two types of prior information:

(1) one-hot encoding of skill categories (4-dimensional, corresponding to the four categories in Table 1) [28], to ensure that similar skills are close in the embedding space;
(2) mastery rate of the whole sample (1-dimensional) [29], to reflect the degree of skill popularization and avoid the model from paying too much attention to low-frequency but occasionally co-occurring skill pairs.

Differences in students' professional background lead to different transfer preferences, which are reflected in the way the model is constructed. Specifically, the input data used to create the model included this heterogeneous information implicitly. While the model was being trained, the Graph Attention Network (GAT) utilized a multi-head attention mechanism to learn not just isolated skill node features, but the entire sequence of skill transfer edges between students who share a professional background. Thus, when students from a certain professional background exhibit higher rates of transfer along certain paths compared to other students in the same background, these differences are recorded as part of the structure of the training graphs and as patterns of temporal co-occurrence between edges. With its multi-head attention mechanism, the GAT model can learn which weights should be applied to each edge in relation to each of the various subgraph structures present within the overall graph of data. Instead of explicitly entering professional background as a categorical variable, the GAT model dynamically captures and differentiates between various substructure differences present in the graph structure based on evidence of those differences obtained through attention weights. The two are concatenated and mapped to a 16-dimensional embedding by Linear(5,16)+ReLU, which is used as the input of GAT. To select the optimal number of input dimensions among the possible combinations of dimension sizes ranging from low (8, 10, and 12) to high (16, 20, 24, and 32), hyperparameter grid search was performed. The results indicated that dimensionalities less than 12 led to limited performance due to insufficient representation of features in the input space. Conversely, dimensionalities greater than 20 did not significantly benefit model performance and resulted in increased validation set variability (i.e., overfitting). Based on the conditions of the current study, 16-dimensional embedding spaces were the best trade-off between model complexity and performance. The model adopts a two-layer architecture. The first layer contains 8 attention heads, each of which independently learns different transfer modes (such as control chain and data chain) and enhances the coverage of heterogeneous paths through the multi-head mechanism [30]. The output is concatenated into a 64-dimensional high-order representation. The second layer is a single-head attention, which aggregates the high-order representation into scalar attention coefficients α_{ij} to directly represent the transfer strength from edge i to j. This design avoids the problem of insufficient expressive power of single-head attention and prevents redundant noise from being introduced into the final layer by multi-head output. The training adopts binary cross-entropy loss, Adam optimizer (learning rate 0.005), Dropout=0.6 [31]. Finally, the output of the model α_{ij} is the dynamic transfer weight required for subsequent propagation efficiency calculation, as shown in Figure 2.

Figure 2 fully illustrates the skill transfer modeling process constructed in this paper: starting from a student-skill binary network, a directed transfer graph is projected through time constraints, serving as the input structure for the GAT model. The GAT model first embeds and maps the initial features (including category encoding and mastery rate) of each skill node. Then, it learns multi-perspective transfer relationships in parallel through an 8-head attention mechanism in the first layer. After concatenation and non-linear activation, the information is aggregated into a single-head attention module in the second layer, ultimately outputting the skill mastery probability. During this process, the model adaptively generates attention weights for each

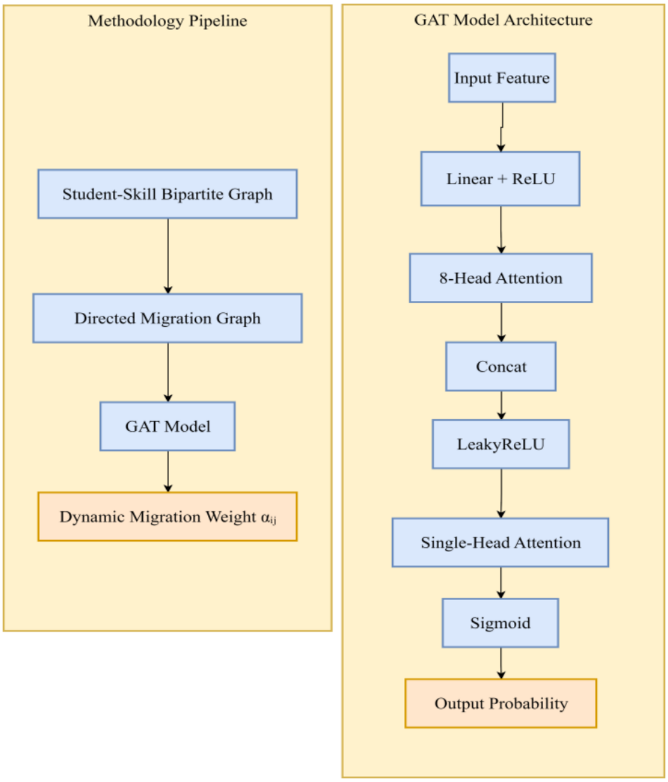

FIGURE 2. Skill transfer modeling architecture based on two-layer GAT

directed edge; these weights serve as a quantitative basis for reflecting the dynamic transfer influence between skills. The entire process demonstrates end-to-end modeling capabilities from raw educational data to interpretable transfer weights, laying the foundation for the subsequent definition and analysis of propagation efficiency. The experiment is divided into training, validation, and test sets (ratio 7:1.5:1.5) based on student IDs to prevent data leakage. Ultimately, the model output α_{ij} represents the dynamic transfer weights required for subsequent propagation efficiency calculations. Table 2 shows GAT hyperparameter settings.

TABLE 2. GAT Hyperparameter Settings

Hyperparameter	Value
Network architecture	2-layer GAT (8 heads to 1 head)
Input / Hidden / Output dims	16 / 64 / 1
Learning rate	0.005
Dropout	0.6
Optimizer	Adam (weight decay = $1e^{-4}$)
Activation functions	LeakyReLU / Sigmoid
Training strategy	Full-graph training, early stopping (patience = 15)

The GAT hyperparameters listed in Table 2 are not empirical settings, but rather a systematic design closely aligned with the core objectives of dynamically modeling the directionality, individual heterogeneity, and educational interpretability of skill transfer. A two-layer GAT (from 8 heads to 1 head) structure is employed, capturing diverse transfer semantics such as control chains and data chains through a multi-head mechanism, while ensuring the stability and quantifiability of transfer weights through single-head output. The input/hidden/output dimensions (16/64/1) integrate skill category priors and mastery statistics, balancing semantic rationality and generalization ability. A high Dropout (0.6) and early stopping strategy (patience=15) effectively suppress overfitting under small sample sizes. The learning rate, optimizer, and activation functions (LeakyReLU/Sigmoid) are all optimized on a validation set, ensuring the probabilistic interpretability of attention weights. The overall configuration reflects a modeling paradigm driven by educational logic and decision support, ensuring that the model output truly reflects the intrinsic mechanism of skill transfer.

Quantitative Indicators of Transmission Efficiency

While the attention weight α_{ij} output by the GAT model can reflect the dynamic attention given to skill i to skill j, relying solely on it to measure transfer value is susceptible to two types of biases: first, high-frequency co-occurrence bias (e.g., PLC programming and electrical drawing interpretation co-occur frequently due to course binding, but without substantial transfer gain); second, static influence bias (high α_{ij} may stem from high skill prevalence rather than genuine facilitation). To overcome these problems, this paper introduces a counterfactual mastery probability gain Δp_j to quantify the causal contribution of skill i to skill j. Specifically, for any directed transfer path i to j, its local propagation effect formula is defined as:

$$\eta_{ij} = \alpha_{ij} \cdot \Delta p_j \tag{1}$$

In formula (1), α_{ij} is the attention weight output by the second layer of GAT, reflecting the dynamic influence of skill s_i on s_j; Δp_j is the gain of the probability of s_j mastering after mastering s_i, calculated as shown in formula (2):

$$\Delta p_j = \hat{y}_j^{(i)} - \hat{y}_j^{(0)}, \tag{2}$$

In formula (2), $\hat{y}_j^{(0)}$ is the predicted mastery probability of s_j in the initial state of the student (GAT output when s_j mastery rate is 0 in the input feature), and $\hat{y}_j^{(i)}$ is the predicted mastery probability of s_j obtained by forward propagation after forcibly setting the s_i mastery state (i.e. setting the mastery rate to 1 in $h_i^{(0)}$). The counterfactual mastery probability $\hat{y}_j^{(0)}$ for skill s_j without mastering skill s_i is obtained by setting the mastery rate feature of node s_i to 0 in the initial node feature vector while keeping all other node features unchanged, and then performing a forward pass through the trained GAT model to obtain the predicted probability. Conversely, $\hat{y}_j^{(i)}$ is obtained by setting the mastery rate feature of node s_i to 1 in $h_i^{(0)}$, with all other features unchanged, and performing the same forward pass. The difference between these two outputs quantifies the causal contribution of mastering s_i to the predicted mastery of s_j, isolating the effect of s_i from other factors. The initial node feature vector $h_i^{(0)}$ for each skill includes its one-hot category encoding and its global mastery rate across all students. For the counterfactual calculation, only the mastery rate component corresponding to skill s_i is manipulated (set to 0 or 1), while the mastery rate features for all other skills remain at their original empirical values from the dataset. This ensures that the computed Δp_j reflects the specific change attributable to s_i, controlling for the baseline influence of other skills and their prevalence. The difference is quantified through counterfactual inference [32] to quantify the causal contribution of s_i to s_j, avoiding co-occurrence bias. To suppress noise interference from low-frequency migration edges and introduce prior knowledge of real behavior, this paper introduces normalized migration frequency as a prior adjustment factor for edge weights [33]. The eligibility period for skill mastery is determined by the 12-month timeframe used to build the transfer graph, so that the skills being transferred are comparable to recent experiences of the student, as well as within the timeframe of their later years in school, when they are likely to have greater opportunities to master these skills. Conversely, the 180-day period used to derive the frequency of skill migrations $\text{count}(i \to j)$ imposes a more stringent definition of the cause-effect relationship of skill migrations. The definition of a valid skill migration is confined to the skill i that was mastered within 180 days of mastering the preceding skill j; therefore, the cause and effect are represented in the equation i and j. For any directed migration edge $i \to j$, its normalized migration frequency is defined as:

$$\tilde{f}_{ij} = \frac{\text{count}(i \to j)}{\max_{u,v} \text{count}(u \to v)} \in [0, 1] \tag{3}$$

In formula (3), $\text{count}(i \to j)$ represents the total number of migration events in which skill i is mastered within 180 days after skill j is mastered in the training set, and the denominator is the maximum frequency among all migration edges. This normalization ensures $\tilde{f}_{ij}$ comparability while preserving the relative distribution of the original migration strength. For each skill node j, its propagation potential Φ_j is defined as the weighted average of the local efficiencies of all outgoing edges, as shown in formula (4):

$$\Phi_j = \frac{1}{|N_j^{\text{out}}|} \sum_{k \in N_j^{\text{out}}} \bar{f}_{jk} \cdot \eta_{jk} \tag{4}$$

In formula (4), N_j^{out} represents the set of out-neighbors of skill j (i.e., the skills that can be transferred from j), η_{jk} represents the local propagation efficiency from edge j to k, and $\bar{f}_{jk}$ represents the normalized transfer frequency, which serves as a priori weight to suppress the influence of low-frequency noise edges. This index combines the dynamic attention and real transfer strength of GAT learning, reflecting the comprehensive ability of skill j to diffuse knowledge outward. To verify the effectiveness of the proposed local propagation efficiency index, this paper conducts a correlation analysis with empirical conditional probability $P(s_j|s_i)$. Here, $P(s_j|s_i)$ is defined as the proportion of students who actually mastered the skill within 180 days in the entire sample s_i. This value can be directly obtained from the time-series data of student skill mastery and reflects the true transfer intensity. Since both η_{ij} and $P(s_j|s_i)$ are non-normally distributed, Spearman's rank correlation coefficient is used to assess their consistency and to verify whether the proposed index can more accurately characterize the true transfer intensity between skills, providing a basis for subsequent high-potential skill identification and pivotal student analysis.

III. EXPERIMENTAL DESIGN AND EVALUATION

A. EXPERIMENTAL SETUP

To verify the effectiveness of the GAT dynamic modeling framework, this paper conducts experiments with skills transfer prediction as the core task. This task directly corresponds to educational needs: given the set of skills a student currently possesses, the model needs to predict the top-3 skills the student is most likely to master in the next 180 days. Data was divided into training (230 students), validation (49 students), and test (49 students) sets in a 7:1.5:1.5 ratio based on student ID, ensuring no significant difference in skill coverage distribution. All models were trained and evaluated using the same directed transfer graph and node features (one-hot encoding of skill categories +

full sample mastery rate) to ensure fair comparison. Model selection followed a progressive control principle:

1) GAT is the two-layer architecture (8 heads to 1 head) proposed in this article, used to verify the effectiveness of dynamic attention mechanisms;
2) GCN uses the same number of layers and dimensions, but employs homogeneous neighborhood aggregation to verify the necessity of directional modeling;
3) PageRank and Restart Random Walk (RWR): It uses migration frequency as edge weight, representing static centrality method, to verify the difference between dynamic and static modeling;
4) Logistic Regression (LR) is characterized by the co-occurrence frequency of skills and represents traditional statistical methods.

Evaluation Indicators

To verify the model performance, this paper constructs an evaluation logic from four dimensions: predictive ability, interpretability, decision support, and robustness. Predictive ability: Hit@3 (i.e., the accuracy of predicting at least one of the Top-3 skills) [34] is used as the core prediction indicator.

$$\text{Hit@3} = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|\hat{S}_i^{(3)} \cap S_i^{\text{true}}\right|}{3} \tag{5}$$

In formula (5), N represents the number of test students, $\hat{S}_i^{(3)}$ represents the set of Top-3 skills predicted by the model, and S_i^{true} represents the set of skills actually mastered by the students. This indicator directly reflects the accuracy of the model in capturing the transfer path and is the basis for verifying the effectiveness of GAT modeling. Interpretability is judged by calculating the Spearman correlation coefficient between the attention weights of the GAT output α_{ij} and the empirical conditional probability $P(s_j|s_i)$. This method has been used to verify the behavioral consistency of the attention mechanism in educational models [35]. If the two are highly correlated, it indicates that the attention weights have behavioral interpretability and can serve as a reliable basis for calculating local propagation efficiency. Decision support capability is defined as follows:

1. Average propagation potential energy of high-potential skills: Calculate the average propagation potential energy η of the top 10% of skills to verify whether this indicator can effectively distinguish core skills;
2. Identification Overlap Rate (IOR) of hub students:

$$\text{IOR} = \frac{|H_{\text{pred}} \cap H_{\text{true}}|}{|H_{\text{true}}|} \tag{6}$$

In formula (6), H_{true} represents the real hub set of 8 high impact students who have been confirmed by both enterprise mentors and academic records in the test set, including 5 who have served as leaders of school enterprise joint projects; 3 winners of provincial and above professional skills competition awards. H_{pred} is the hub student set for model recognition. This indicator measures the model's ability to identify truly high-impact individuals. Robustness was evaluated by measuring the attenuation of Hit@3 under random occlusion of 10%–50% of skill records. The performance attenuation rate was defined as:

$$\Delta = \text{Hit@3}_{\text{full}} - \text{Hit@3}_{\text{masked}} \tag{7}$$

The smaller the decay rate, the higher the model's tolerance for incomplete data in real-world educational scenarios.

IV. RESULTS

Migration Prediction Accuracy is Significantly Better than the Baseline Model

To verify the predictive ability assessment logic, this paper uses a skills transfer prediction task as a benchmark to quantify the accuracy of each model in capturing the propagation path. The prediction task is defined as follows: given a student's current set of skills, the model needs to predict the top-3 skills that the student is most likely to master in the next 180 days. The evaluation metric is Hit@3, which is the intersection ratio between the predicted set and the actual mastered set, as shown in Figure 3.

Figure 3 shows a performance comparison of the five models in the skill transfer prediction task. Subfigure (a) shows that GAT converges quickly thanks to its multi-head attention mechanism, with Hit@3 increasing from 61.5% to 78.4%, outperforming GCN (65.8%). Meanwhile, static centrality methods such as PageRank and RWR, as well as traditional statistical methods like logistic regression (LR), all have lower performance (52.1%, 48.7%, and 41.3%, respectively). Subfigure (b) presents the final performance differences. These performance differences stem from the inherent nature of the model architecture: GAT can dynamically capture high-potential transfer paths such as "PLC to digital twin," effectively modeling directionality and temporal dependencies; GCN, due to homogeneous aggregation, struggles to distinguish transfer directions, resulting in slow convergence and limited accuracy; while static methods like PageRank and RWR do not involve parameter training, resulting in flat performance curves, ignoring individual heterogeneity and causal temporal relationships, relying only on co-occurrence frequency, leading to performance stagnation; LR, as a traditional statistical model, has the lowest performance because it does not utilize graph structure information. The GAT stabilized after 80 rounds, indicating that it had fully explored dynamic propagation signals and verified the dual value of the attention-driven propagation efficiency framework in improving prediction accuracy and enhancing educational interpretability.

Attention Mechanisms Effectively Capture Real Transfer Patterns

To examine whether the attention weights output by the GAT model possess behavioral interpretability, this paper

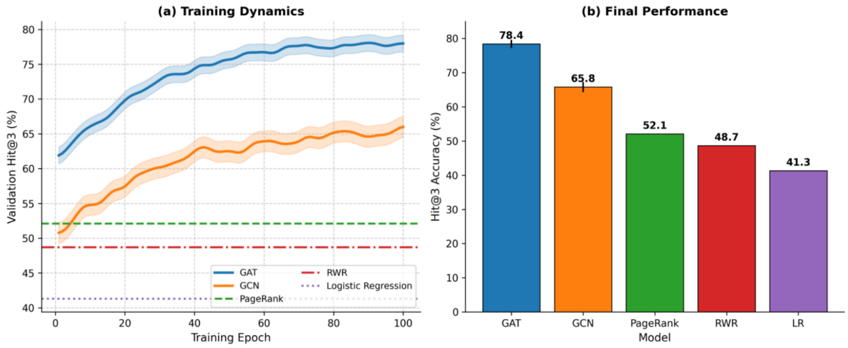

FIGURE 3. Comparison of training dynamics and final performance of five models in skill transfer prediction tasks

performs a correlation analysis with the actual frequency of skill transfer. Specifically, for each directed transfer edge i to j, its normalized transfer frequency is calculated $\tilde{f}_{ij}$—that is, the number of times the edge appears in the training set divided by the maximum frequency of all edges, with a value range of [0,1], serving as a benchmark for measuring the strength of actual transfer. Subsequently, the consistency between the attention weights α_{ij} output by the second layer of GAT is evaluated by calculating their correlation with $\tilde{f}_{ij}$, the actual transfer frequency. This analysis covers all 1,247 valid transfer edges, verifying whether the attention mechanism can effectively reflect the actual transfer patterns between skills, as shown in Figure 4.

Figure 4 shows the scatter distribution between GAT attention weights α_{ij} and normalized transfer frequencies $\tilde{f}_{ij}$. The blue scatter points represent data points from 1,247 transfer edges, and the red dashed line is the linear fitting trend line, whose slope and intercept intuitively reflect the positive correlation between the two. The Pearson correlation coefficient (0.67) and p-value are used to quantify the strength and significance of this association. The data points are clustered roughly along the diagonal, indicating that transfer paths with high attention weights often correspond to higher actual transfer frequencies, demonstrating that the attention mechanism learned by GAT can effectively capture real-world behavioral patterns in skill transfer.

High-Potential Skills are Concentrated in Multi-Faceted Abilities

To identify core skills that strongly drive skill transfer networks, this paper quantifies and ranks 42 skills based on skill propagation potential Φ_j (GAT) and compares their categories. This indicator integrates GAT dynamic attention weights and actual transfer frequency, reflecting the comprehensive ability of skills to diffuse knowledge outwards. The analysis focuses on the distribution of skill categories, the trend of potential concentration, and their educational implications, aiming to reveal which skills play a pivotal role in real learning behaviors, as shown in Figure 5.

Figure 5 shows the top 10 skills in terms of diffusion potential, colored according to the four categories shown in Table 1. The top five skills are: system integration (S11, $\Phi = 0.83$), digital twin modeling (S06, $\Phi = 0.81$), fault diagnosis (S07, $\Phi = 0.79$), industrial robot operation (S03, $\Phi = 0.77$), and SCADA configuration (S09, $\Phi = 0.75$), all of which are composite capabilities integrating knowledge from multiple domains. Notably, five of the top 10 skills belong to the intelligent manufacturing category (S03–S11) and involve cross-domain integration, indicating that this type of skill has a significant diffusion advantage in the transfer network. This result verifies that the design logic of "system integration capability as the core" in the industry-university curriculum system is highly consistent with actual transfer behavior. The top 10 skills in terms of propagation potential (Figure 5) are highly concentrated in the representative skills of intelligent manufacturing and information technology listed in Table 1, indicating that the design of this skill system is highly consistent with actual migration behavior. Figure 6 shows the differences in the distribution of propagation potential among different skill categories.

Figure 6 illustrates the distribution of propagation potential energy grouped by the four major skill categories. The median propagation potential energy for intelligent manufacturing skills is 0.71, higher than the other three categories; its data points are densely distributed in the higher potential energy range, and the top five high-potential-energy skills (marked with gold stars) all belong to this category. In contrast, the potential energy distribution for information technology, electromechanical control, and project management skills is generally lower, and the data dispersion is smaller. High-potential-energy skills are mainly concentrated in cross-domain composite capabilities such as system integration (S11), digital twin modeling (S06), and fault diagnosis (S07), reflecting the structural centrality of these skills in the migration network.

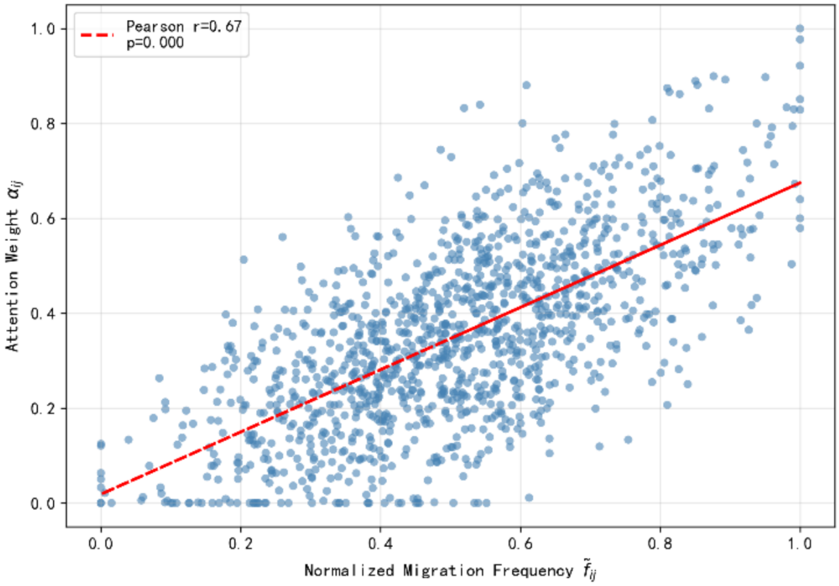

FIGURE 4. Scatter plot of attention weights versus actual transfer frequency

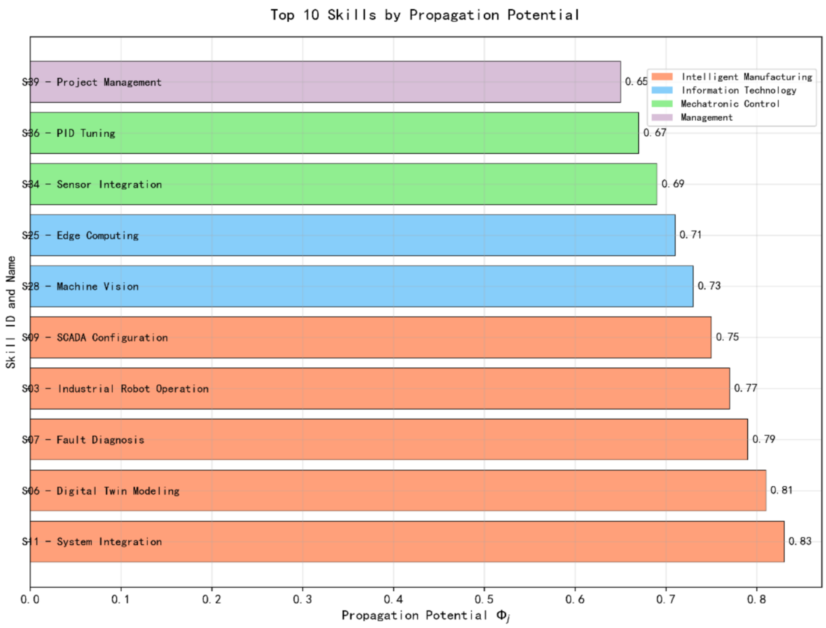

FIGURE 5. Bar chart of the top 10 skills in terms of propagation potential

Accurately Identify Pivotal Students Who Dominate the Transfer Process

To verify the model's ability to identify key minority students, this paper uses 8 real hub students in the test set as ground-truth labels for evaluation. This set is strictly defined by dual criteria:

(1) having served as the leader of a joint school-enterprise project (term $\geq$ 3 months), or

(2) having won provincial or higher-level professional skills competition awards and having played a leading or coordinating technical role in the project as confirmed by the enterprise mentor.

Accordingly, the pivotal student identified by the model is defined as one of the top 15% of students in the test set who, as a source node (i.e., migration initiator), triggers the most high-weight migration edges ($\alpha_{ij} > 0.7$) and possesses at least 5 high-impact skills with a propagation potential ≥ 0.58. A student node in the test dataset that qualifies as a model-identified hub student is defined as having the highest 15% of its outgoing edges, weighted by local propagation efficiency η_{ij}. The weighted out-degree is computed as denoted by the equation, where all outgoing edges for that student node are counted if η_{ij} is greater than or equal to 0.05. Thus, the sum of η_{ij} provides quantitative representation of the total triggering strength associated with high-efficiency paths initiated from this student node

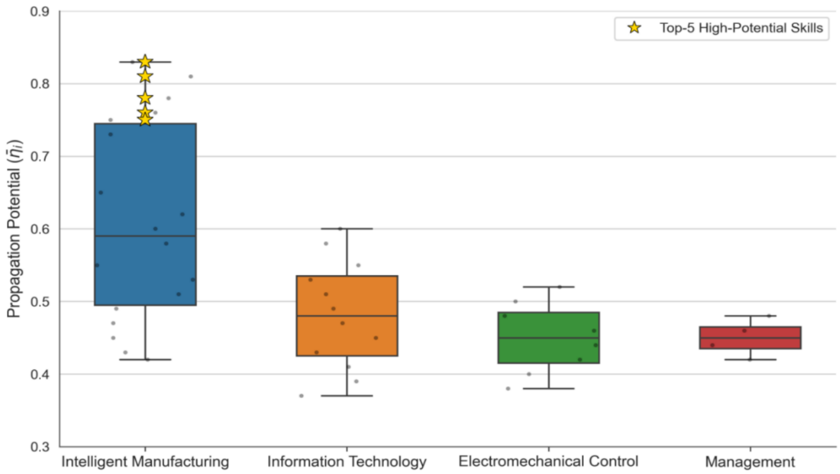

FIGURE 6. Box Plot of Skill Propagation Potential Distribution (grouped by skill category)

to other nodes. Additionally, at least five skills must be mastered by the student, and the skill propagation potentials (Φ_j) should equal or exceed 0.58. The first dimension, "migration initiation activity", is quantified by the weighted out-degree. The second dimension, "breadth of high-impact skills," is quantified by the count of mastered skills satisfying $\Phi_j \geq 0.58$. This two-dimensional quantitative definition ensures the identification results reflect both the individual's dynamic influence as a propagation source in the transfer network and their possession of skills with high dissemination potential, thereby maintaining educational behavior interpretability.

Table 3 shows that GAT identified 9 predicted hub students, 7 of whom overlapped with the actual 8. The predicted recall was 87.5%, the precision was 77.8%, and the F1-score was 82.4%, outperforming GCN and other baseline methods. This indicates that GAT can effectively combine dynamic attention and propagation potential to accurately locate truly high-influence individuals.

TABLE 3. Comparison of student recognition performance at hubs

Model	Predicted Hub Students	Ground-Truth Hubs	Overlap Count	Recall (%)	Precision (%)	F1-score
GAT	9	8	7	87.5	77.8	82.4
GCN	12	8	5	62.5	41.7	50
PageRank	15	8	4	50	26.7	34.8
RWR	14	8	4	50	28.6	36.4
LR	13	8	3	37.5	23.1	28.6

Further analysis revealed that only 8 key students (16.3% of the test set) possessed highly concentrated skills in high-potential composite abilities such as system integration (S11), digital twin modeling (S06), and fault diagnosis (S07). This result indicates that high-impact transfer behavior is highly concentrated among a small core student group, reflecting the leverage effect of the 'critical few' in educational communication.

Model is Highly Robust to Missing Information

To test the robustness evaluation logic, this paper randomly occludes 10%–50% of the skill records on the test set, as shown in Table 4.

TABLE 4. Accuracy variation under different missing rates

Missing Rate	GAT	GCN	PageRank	RWR	LR
10%	76.1	63.2	48.3	44.6	36.8
20%	73.8	60.1	45	40.9	33.2
30%	71.5	56.3	41.2	37	29.5
40%	70	51.6	37.5	33.1	26
50%	69.2	47.1	29.7	24.6	22.5

Table 4 shows that GAT's Hit@3 score decreased by only 9.2 percentage points (to 69.2%) under 50% occlusion, outperforming GCN (Δ=18.7 %) and PageRank (Δ=22.4 %). This robustness stems from the adaptive compensation capability of the attention mechanism: when a skill is occluded, the model can indirectly infer missing information through strong transfer paths of high-potential skills (e.g., using digital twin modeling to infer machine vision). This characteristic ensures the reliability of the method in real-world educational scenarios (e.g., incomplete records), meeting the practical needs of industry academies for model stability.

Successfully Captured Individual Transfer Preferences Based on Professional Background

To verify the GAT model's ability to model individual heterogeneity, this paper groups students by their professional background and analyzes the distribution of their attention weights on typical transfer paths. Specifically, the test set students are divided into three categories:

(1) Mechanical/Electrical Engineering (e.g., Mechatronics, Electrical Automation, 22 students);
(2) Software Engineering (e.g., Computer Applications, Big Data Technology, 18 students);
(3) Interdisciplinary Engineering (e.g., Intelligent Manufacturing Engineering, with both hardware and software backgrounds, 9 students).

Six typical transfer paths are selected as the analysis dimensions. The first three belong to the industrial control chain (e.g., PLC to HMI, HMI to SCADA), and the last three belong to the data intelligence chain (e.g., Python to Machine Vision, Machine Vision to Digital Twin Modeling), as shown in Figure 7.

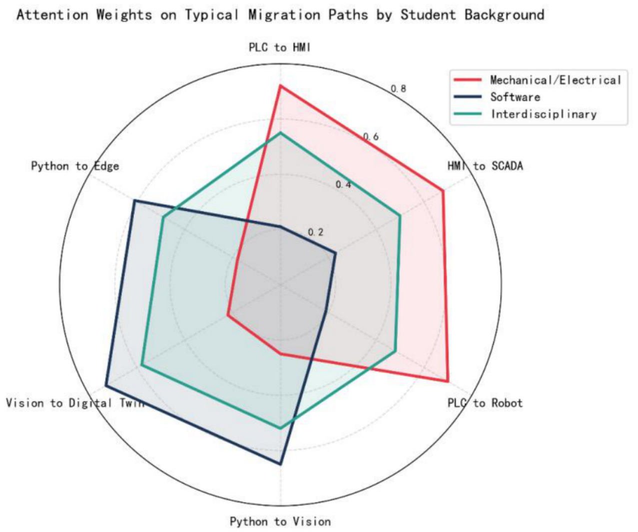

FIGURE 7. Attention weight distribution of students from different academic backgrounds on typical transfer paths

Figure 7 illustrates the attention weight distribution of the three types of students along typical transfer paths. Students with mechanical and electrical engineering backgrounds show higher attention weights (mean 0.68–0.72) on industrial control chains (e.g., PLC to HMI, HMI to SCADA), while students with software backgrounds dominate on data intelligence chains (e.g., Python to machine vision, machine vision to digital twin modeling) (mean 0.65–0.73). Students with interdisciplinary backgrounds exhibit relatively balanced transfer preferences. The attention distributions of the two types of paths have almost no overlap (e.g., the PLC to industrial robot operation edge has $\alpha = 0.70$ in the mechanical category, but only 0.19 in the software category), indicating that GAT can adaptively adjust transfer weights based on students' prior knowledge structure.

V. CONCLUSION

This paper constructs a technology skills transfer network for students in industry-academia collaboration based on the GAT framework. Through directed transfer graph modeling and a dynamic attention mechanism, it effectively overcomes the limitations of traditional static models in characterizing the directionality of transfer and individual heterogeneity. Experiments show that GAT achieves a Hit@3 accuracy of 78.4% in skills transfer prediction tasks, outperforming baseline methods such as GCN. The defined local propagation efficiency and skills propagation potential indicators have strong behavioral interpretability and can accurately identify high-impact skills and core students. The study finds that composite skills such as system integration and fault diagnosis have the highest propagation potential, exceeding that of basic skills. Hub student analysis further reveals the 'critical few' leverage effect in educational propagation: a small number of students possessing high propagation potential skills play a core initiator role in the transfer network. This framework not only maintains high robustness even with 50% skill record loss (Hit@3 decay of only 9.2%), but more importantly, by combining dynamic attention with counterfactual causal reasoning, it constructs an interpretable, measurable, and operable propagation efficiency indicator system. This effectively transforms the algorithm output into a decision-making basis for curriculum optimization and personalized teaching intervention, promoting industry-academia integration from experience-driven to data-driven approaches.

AUTHOR CONTRIBUTIONS

Conceptualization –Shaobin Zhang and Shuzhen Wang; methodology – Shaobin Zhang and Shuzhen Wang; investigation – Shaobin Zhang and Shuzhen Wang; writing-original draft preparation – Shaobin Zhang and Shuzhen Wang. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research was supported by, Teaching Reform Research and Practice Major Project of Henan Province in 2024 "Research and Practice on High Quality Development of Modern Industry College in Local Applied Undergraduate Universities" (2024SJGLX0028).

HUMAN RESEARCH SUBJECTS

This study was approved by the Ethics Committee of Luoyang Institute of Science and Technology, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] H. Zhang, "The Intrinsic Logic and Practical Pathways of Empowering Vocational Education Through Industry-Education Integration: Summary of the Session on Industry-Education Integration for Advancing High-Quality Vocational Education," World Vocational and Technical Education, vol. 1, no. 1, pp. 38-54, 2025.

[2] K. Hötte, "Skill transferability and the stability of transition pathways-A learning-based explanation for patterns of diffusion," Journal of Evolutionary Economics, vol. 31, no. 3, pp. 959-993, 2021.

[3] C. D. Driver and M. J. Tomasik, "Formalizing complex developmental phenomena as continuous-time systems: Learning gains in multiple domains," Child Development, vol. 94, no. 6, pp. 1454-1471, 2023.

[4] P. Bilancia, J. Schmidt, R. Raffaeli, M. Peruzzini, and M. Pellicciari, "An overview of industrial robots control and programming approaches," Applied Sciences, vol. 13, no. 4, pp. 2582, 2023, doi: 10.3390/app13042582.

[5] R. Mahajan, P. Gupta, and R. Misra, “Employability skills framework: A tripartite approach,” Education+ Training, vol. 64, no. 3, pp. 360-379, 2022, doi: 10.1108/ET-12-2020-0367.
[6] P. Wan, X. Wang, Y. Lin, and G. Pang, “A knowledge diffusion model in autonomous learning under multiple networks for personalized educational resource allocation,” IEEE Transactions on Learning Technologies, vol. 14, no. 4, pp. 430-444, 2021, doi: 10.1109/TLT.2021.3103006.
[7] F. Peng and L. Guo, “Personalized learning path planning and optimization methods of vocational education combined with DQN,” Journal of Computational Methods in Sciences and Engineering, vol. 25, no. 4, pp. 3780-3792, 2025, doi: 10.1177/14727978251323112.
[8] D. C. Kavargyris, K. Georgiou, N. Mittas, and L. Angelis, “Extracting knowledge and highly demanded skills from European Union educational policies: A network analysis approach,” Social Network Analysis and Mining, vol. 15, no. 1, pp. 1-41, 2025.
[9] C. Cao, Y. Wang, Y. Zhang, Y. Lu, X. Zhang, and Y. Zhang, “Co-occurrence matters: Learning action relation for temporal action localization,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 34, no. 5, pp. 3327-3339, 2023, doi: 10.1109/TCSVT.2023.3321508.
[10] A. Usta, I. S. Altingovde, R. Ozcan, and Ö. Ulusoy, “Learning to rank for educational search engines,” IEEE Transactions on Learning Technologies, vol. 14, no. 2, pp. 211-225, 2021, doi: 10.1109/TLT.2021.3075196.
[11] H. Li, F. Cao, and W. Dai, “Major-based undergraduate curriculum as an obstacle to graduate employability development,” Higher Education Research & Development, vol. 43, no. 3, pp. 705-719, 2024, doi: 10.1080/07294360.2023.2258844.
[12] A. A. Mubarak, H. Cao, I. M. Hezam, and F. Hao, “Modeling students’ performance using graph convolutional networks,” Complex & Intelligent Systems, vol. 8, no. 3, pp. 2183-2201, 2022.
[13] X. Li, G. Xiao, Y. Chen, Z. Tang, W. Jiang, and K. Li, “An explicitly weighted gcn aggregator based on temporal and popularity features for recommendation,” ACM Transactions on Recommender Systems, vol. 1, no. 2, pp. 1-23, 2023, doi: 10.1145/3587272.
[14] Q. Zhang, K. Hua, Z. Zhang, Y. Zhao, and P. Chen, “Acnet: An attention–convolution collaborative semantic segmentation network on sensor-derived datasets for autonomous driving,” Sensors, vol. 25, no. 15, pp. 4776, 2025, doi: 10.3390/s25154776.
[15] X. Ouyang, L. Liu, W. Chen, C. Wang, X. Sun, C. He, and G. Liu, “Systematic risks of the global lithium supply chain network: From static topological structures to cascading failure dynamics,” Environmental Science & Technology, vol. 58, no. 50, pp. 22135-22147, 2024, doi: 10.1021/acs.est.4c10523.
[16] Q. Huang and J. Chen, “Enhancing academic performance prediction with temporal graph networks for massive open online courses,” Journal of Big Data, vol. 11, no. 1, pp. 52, 2024.
[17] M. Saqr and S. López-Pernas, “The curious case of centrality measures: A large-scale empirical investigation,” Journal of Learning Analytics, vol. 9, no. 1, pp. 13-31, 2022, doi: 10.18608/jla.2022.7415.
[18] G. Kaur, B. Singh, R. S. Batth, and R. Garg, “BATFE: Design of a hybrid bioinspired model for adaptive traffic flow control in edge devices,” Microsystem Technologies, vol. 31, no. 8, pp. 1987-2002, 2025.
[19] Y. Zhang, R. An, S. Liu, J. Cui, and X. Shang, “Predicting and understanding student learning performance using multi-source sparse attention convolutional neural networks,” IEEE Transactions on Big Data, vol. 9, no. 1, pp. 118-132, 2021, doi: 10.1109/TBDATA.2021.3125204.
[20] J. Engle and C. M. Walker, “Thinking counterfactually supports children’s evidence evaluation in causal learning,” Child Development, vol. 92, no. 4, pp. 1636-1651, 2021, doi: 10.1111/cdev.13518.
[21] R. van der Velden, I. Bijlsma, M. C. Fregin, and M. Levels, “Are general skills important for vocationally educated?,” Acta Sociologica, vol. 67, no. 3, pp. 387-405, 2024, doi: 10.1177/00016993231219135.
[22] Y. J. Wang, C. L. Gao, and X. D. Ye, “A data-driven precision teaching intervention mechanism to improve secondary school students’ learning effectiveness,” Education and Information Technologies, vol. 29, no. 9, pp. 11645-11673, 2024.
[23] J. G. Karstensen, L. J. Nayahangan, L. Konge, and P. Vilmann, “A core curriculum for basic EUS skills: An international consensus using the Delphi methodology,” Endoscopic Ultrasound, vol. 11, no. 2, pp. 122-132, 2022, doi: 10.4103/EUS-D-21-00125.
[24] Ó. Cuéllar, M. Contero, and M. Hincapié, “Personalized and Timely Feedback in Online Education: Enhancing Learning with Deep Learning and Large Language Models,” Multimodal Technologies and Interaction, vol. 9, no. 5, pp. 45, 2025, doi: 10.3390/mti9050045.
[25] Y. Xu, S. Sun, H. Zhang, C. Yi, Y. Miao, D. Yang, et al., “Time-aware graph embedding: A temporal smoothness and task-oriented approach,” ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 16, no. 3, pp. 1-23, 2021, doi: 10.1145/3480243.
[26] G. Zhu, Y. Chen, and S. Wang, “Graph-community-enabled personalized course-job recommendations with cross-domain data integration,” Sustainability, vol. 14, no. 12, pp. 7439, 2022, doi: 10.3390/su14127439.
[27] M. Yang, Z. Li, Y. Gao, C. He, F. Huang, and W. Chen, “Heterogeneous graph attention networks for depression identification by campus cyber-activity patterns,” IEEE Transactions on Computational Social Systems, vol. 11, no. 3, pp. 3493-3503, 2024, doi: 10.1109/TCSS.2023.3343689.
[28] T. Dash, A. Srinivasan, and A. Baskar, “Inclusion of domain-knowledge into GNNs using mode-directed inverse entailment,” Machine Learning, vol. 111, no. 2, pp. 575-623, 2022.
[29] D. Yan, S. Bao, Z. Zhang, J. Sun, and M. Zhou, “Leveraging pharmacovigilance data to predict population-scale toxicity profiles of checkpoint inhibitor immunotherapy,” Nature Computational Science, vol. 5, no. 3, pp. 207-220, 2025.
[30] X. Mo, Z. Huang, Y. Xing, and C. Lv, “Multi-agent trajectory prediction with heterogeneous edge-enhanced graph attention network,” IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 7, pp. 9554-9567, 2022, doi: 10.1109/TITS.2022.3146300.
[31] F. Yang, H. Zhang, and S. Tao, “Simplified multilayer graph convolutional networks with dropout,” Applied Intelligence, vol. 52, no. 5, pp. 4776-4791, 2022.
[32] R. W. Runhardt, “Concrete counterfactual tests for process tracing: Defending an interventionist potential outcomes framework,” Sociological Methods & Research, vol. 53, no. 4, pp. 1591-1628, 2024, doi: 10.1177/00491241221134523.
[33] I. W. Jones, J. S. Bersson, J. Liu, K. Sharma, O. A. Vasilyev, T. A. Miller, et al., “Calculated and empirical values of vibronic transition dipole moments of reactive chemical intermediates for determination of concentrations,” The Journal of Physical Chemistry A, vol. 127, no. 21, pp. 4670-4681, 2023, doi: 10.1021/acs.jpca.3c01584.
[34] G. Zare, N. Jafari, M. Hosseinzadeh, and A. Sahafi, “DAC-GCN: A Dual Actor-Critic Graph Convolutional Network with Multi-Hop Aggregation for Enhanced Recommender Systems,” Acta Informatica Pragensia, vol. 2025, no. 3, pp. 340-364, 2025, doi: 10.18267/j.aip.261.
[35] F. Essam, H. El, and S. R. H. Ali, “A comparison of the pearson, spearman rank and kendall tau correlation coefficients using quantitative variables,” Asian Journal of Probability and Statistics, vol. 20, no. 3, pp. 36-48, 2022, doi: 10.9734/AJPAS/2022/v20i3425.