

University Innovation and Entrepreneurship Ecological Network Modeling Based on GNN

Shiqi Feng¹, Ehsan Fansuree Surin²

¹Department of Education, Chongqing Technology and Business University, Chongqing 400067, China

²Faculty of Business and Management, Universiti Teknologi MARA, Puncak Alam Campus, Bandar Puncak Alam 42300, Selangor, Malaysia

Corresponding author: Ehsan Fansuree Surin (e-mail: 13983872756@163.com)

ABSTRACT University innovation and entrepreneurship ecological networks exhibit significant heterogeneity, dynamic evolution, and complex semantic dependencies, making accurate representation learning and temporal modeling particularly challenging. Similar heterogeneous spatio-temporal information fusion problems are increasingly encountered in intelligent electromagnetic sensing, communication networks, and advanced signal processing systems, highlighting the importance of robust graph-based analytical frameworks. To address these issues, this paper proposes a multi-task modeling framework for temporally heterogeneous graphs based on a Relational Graph Attention Network (R-GAT) and a Gated Recurrent Unit (GRU). The R-GAT module distinguishes heterogeneous semantic relationships through a relation-aware attention mechanism, while the time-decaying GRU effectively captures historical behaviors with irregular temporal intervals. Interpretable attention pooling is further introduced to generate graph-level representations, and a multi-task output head with uncertainty-adaptive weighting jointly optimizes classification, regression, and contrastive self-supervised objectives for comprehensive ecosystem modeling. Experimental results demonstrate that the proposed method achieves an accuracy of 0.96 ± 0.01 and an F1 score of 0.87 ± 0.02 for entrepreneurial potential classification, together with an MAE of 0.02 ± 0.01 and an RMSE of 0.04 ± 0.01 for ecosystem health regression, consistently outperforming conventional baselines. The proposed framework enhances both predictive accuracy and interpretability while providing an effective heterogeneous graph modeling strategy that may offer methodological insights for spatio-temporal information processing and intelligent electromagnetic signal analysis in advanced engineering applications.

INDEX TERMS heterogeneous graph, relational graph attention network, gated recurrent unit, spatio-temporal information processing, electromagnetic signal analysis

I. INTRODUCTION

UNIVERSITIES are important hotbeds of innovation and entrepreneurship. The entrepreneurial activities of university students, the collaboration between mentors and enterprises and the linkage between incubators and on-campus courses together constitute a complex ecological network [1-3]. As the scale and diversity of university innovation and entrepreneurship ecological incubation systems continue to expand, how to characterize different subjects and relationships from data, build university innovation and entrepreneurship ecological networks, evaluate ecological health and identify high-potential entrepreneurial individuals has become a core issue in university decision-making and education improvement. In reality, such ecosystems are characterized by strong heterogeneity (multiple types of nodes and relationships), significant temporal sequence (activities and collaborations vary across semesters and project cycles), and data sparsity and noise, posing challenges to traditional

statistical methods and static network analysis. Building a modeling framework based on the current university innovation and entrepreneurship ecological incubation systems that can simultaneously account for relationship heterogeneity and temporal evolution has important theoretical and practical significance for improving prediction accuracy, supporting policy evaluation, and providing interpretable recommendations.

In the field of modeling and evaluation of the innovation and entrepreneurship ecological incubation system in universities, existing studies have explored this from different perspectives. Zhang Y et al. constructed an innovation and entrepreneurship education ecological system in universities based on social network analysis, focusing on key factor identification and resource circulation. The experiment showed that resource utilization and achievement conversion rate were improved, but the method was mainly based on macro-metrics, which made it difficult

to capture the semantic differences and temporal evolution of multi-type relationships [4]. Malik A et al. studied the adaptability of students to online entrepreneurial education by improving the integrated machine learning model and proved that multi-model integration is effective in predicting educational adaptability [5]. Existing research has achieved beneficial results in constructing evaluation index systems, improving resource utilization and teaching design, and improving evaluation accuracy with traditional integrated machine learning. However, it generally coarsens multi-source heterogeneous subjects and multi-type relationships into homogeneous nodes or single edge types, resulting in the loss of key semantic information; at the same time, its static modeling paradigm makes it difficult to depict the dynamic evolution process of the ecosystem, and cannot support the joint inference of individual potential and overall health. In recent years, research has applied GNN (Graph Neural Network) and its temporal variants into university innovation, entrepreneurship and education scenarios. Hao D proposed a group competition recommendation model that combined hierarchical analysis with gated graph attention, which significantly improved recommendation accuracy and satisfaction, but focused on recommendation tasks and did not cover multi-relation semantic differentiation and graph-level health assessment [6]. Existing work has made significant progress in tasks such as recommendation, matching, retrieval, and individual prediction, proving the effectiveness of GNN and temporal models in educational scenarios, but their application in modeling the innovation and entrepreneurship ecological network of universities is relatively limited, and there are many research gaps. In addition, there is insufficient fine distinction of heterogeneous multi-relational semantics, and temporal modeling and relational modeling have not yet been fully coupled. In contrast, this paper deeply integrates R-GAT and GRU to construct an end-to-end temporal heterogeneous graph multi-task learning architecture, which not only precisely identifies high-potential entrepreneurial students at the node level, but also dynamically evaluates the four-dimensional health status of the ecosystem at the graph level. This framework breaks through the limitations of existing methods in terms of semantic confusion, temporal fragmentation, and single task, and provides a systematic solution for intelligent modeling and decision support for the innovation and entrepreneurship ecosystem of universities.

This paper aims to construct a modeling framework for university innovation and entrepreneurship ecosystems that simultaneously captures the semantics of multiple entities and relationships and captures temporal evolution, thereby more accurately predicting student entrepreneurial potential and assessing ecosystem health. This work normalizes multi-source teaching, project, incubation, and collaboration data into a time-series heterogeneous graph. The framework then uses the relation-aware R-GAT (Relational Graph Attention Network) to semantically weight the relationships. A Gated Recurrent Unit (GRU) with time decay aggregates

node historical behavior. Finally, a multi-task output head performs student classification and graph-level regression. Training incorporates an uncertainty-weighted joint loss and timeline comparison self-supervision to enhance robustness. Experiments (using multi-source data from a medium-sized comprehensive university from 2020 to 2024) show that the proposed R-GAT+GRU model achieves an accuracy of 0.96 and an F1 of 0.87 in classifying student entrepreneurial potential, an accuracy improvement of approximately 0.03 compared to an MAGNN (Metapath Aggregated Graph Neural Network) + LSTM. In ecological health regression, it achieves an MAE (Mean Absolute Error) of 0.02 and an RMSE (Root Mean Square Error) of 0.04, and also outperforms leading baselines in significance tests, robustness tests, and cross-university transfer experiments. The RMSE error for time series modeling over four semesters remains between 0.038 and 0.042, demonstrating the effectiveness of the proposed method in both semantic differentiation and temporal memory.

II. CONSTRUCTION OF GRAPH STRUCTURE OF TIME-SERIES HETEROGENEOUS ECOLOGICAL NETWORK

This paper normalizes and deduplicates the original data IDs, and uses regularized merging (based on name + affiliated unit + time window similarity threshold) to ensure entity uniqueness. Sentence-BERT (Bidirectional Encoder Representations from Transformers) is used to extract vector representations for text fields (project description, achievement summary) [7-9]; categorical features are represented by embedding lookup; numerical features are standardized by z-score. Missing values are filled in key table fields using MICE (Multiple Imputation by Chained Equations) multiple imputation, and sparse interactions are filled in using forward filling based on historical windows [10].

For graph construction, based on the university innovation and entrepreneurship ecological incubation system, the node types include seven categories: students, teachers, teams, projects, incubators, companies, and patents, and the relationship types include six categories: guidance, cooperation, investment, participation, co-authorship, and support. For each pair of entities with multiple interaction records in a single time window, they are first aggregated into a single edge according to the relationship type, and the edge weight is calculated. The weighted frequency of the relationship for any node in window t is calculated, and time decay is applied, as shown in Formula (1).

$$w_{ijr}^{(t)} = \sum_{e \in E_{ijr}^{(t)}} (\kappa_e \cdot \exp(-\gamma \Delta \delta_e)) \quad (1)$$

In Formula (1), $E_{ijr}^{(t)}$ represents all events of the content in window t ; κ_e represents the event intensity (such as the amount of cooperation); $\Delta \delta_e$ represents the time difference between the event and the end of the window; γ represents the decay coefficient.

In each relationship, a sparse adjacency matrix is constructed for it at time step. To eliminate degree deviation and facilitate subsequent relationship perception network processing, relationship normalization is applied, as shown in Formula (2).

$$\tilde{A}_r^{(t)} = D_r^{(t)-\frac{1}{2}} A_r^{(t)} D_r^{(t)-\frac{1}{2}} \quad (2)$$

In Formula (2), $A_r^{(t)}$ represents the sparse adjacency matrix; $D_r^{(t)}$ represents the degree matrix of $A_r^{(t)}$; $\tilde{A}_r^{(t)}$ represents the normalized matrix.

For each type of node at time step, the original feature vector is constructed as the concatenation of numerical features, category embedding, and text vector, as shown in Formula (3).

$$x_u^{(t)} = [\text{norm}(f_u^{(t)}) \| \text{emb}_{cat}(c_u) \| \text{bert}(s_u)] \quad (3)$$

In Formula (3), $x_u^{(t)}$ represents the original feature vector.

Type-specific linear projections are then applied to different node types to align them into a unified representation space, as shown in Formula (4).

$$h_u^{(t,0)} = W_{\tau(u)} x_u^{(t)} + b_{\tau(u)} \quad (4)$$

In Formula (4), $\tau(u)$ represents the type index of the node, and $W_{\tau(u)}$ and $b_{\tau(u)}$ represent the learned type projection parameters.

The constructed temporally heterogeneous ecological network diagram is shown in Figure 1.

In the three time windows of Figure 1, the dynamic evolution characteristics of the university innovation and entrepreneurship ecological network can be clearly observed. At $t=1$, the network connections are relatively sparse, and the interaction is mainly concentrated in teacher-student guidance and initial team collaboration, indicating that the ecosystem is still in its embryonic stage; by $t=2$, the relationship between cross-type nodes has increased significantly, such as the investment connection between enterprises and projects, and the support role of incubators have gradually become active, reflecting the enhanced resource flow and cross-border interaction of the mid-term ecology; by $t=3$, the overall connectivity and edge weight of the network have increased significantly, and nodes such as teams and enterprises have formed closer connections, indicating that the system has entered a relatively mature stage, and the efficiency of knowledge transformation and achievement incubation has improved.

Relationship-Aware Embedding Learning Based on R-GAT

R-GAT applies a relationship-aware module based on GAT (Graph Attention Network) [11]. In the time window t , with the type projection initial representation $h_{(t,0)}$ and the normalized adjacency matrix $\tilde{A}_r^{(t)}$ of each relationship as input, the goal of R-AT is to learn the relationship

vector and relationship-specific projection for each relationship r , explicitly distinguish heterogeneous semantics at the attention level, and generate relationship-sensitive node embeddings.

A trainable relationship vector q_r and relationship-specific linear mapping W_r are learned for each relationship type r . At the same time, the type-specific projection matrices W_{src}^τ and W_{dst}^τ are maintained for the source/target node types to ensure the semantic alignment of heterogeneous nodes before attention calculation. In the relationship-aware mode, for the neighbor j of a certain type of relationship of node i , the unnormalized attention score is calculated, as shown in Formula (5) [12-14].

$$e_{ijr}^{(t)} = \text{LeakyReLU} \left(a^\top [W_{src}^\tau(i) h_i^{(t,0)} \| W_{dst}^\tau(j) h_j^{(t,0)} \| q_r \| \phi(p_{ijr}^{(t)})] \right) \quad (5)$$

In Formula (5), $\|$ represents vector concatenation; a represents attention projection vector; $p_{ijr}^{(t)}$ represents edge attribute vector (edge weight, frequency, etc.); $e_{ijr}^{(t)}$ represents relationship-aware attention; ϕ represents the linear mapping of edge attributes.

A normalized softmax is used for neighbors under the same relationship type, and pre-weighting based on $\tilde{A}_r^{(t)}$ is applied to preserve graph structure information, as shown in Formula (6).

$$\alpha_{ijr}^{(t)} = \frac{\exp(e_{ijr}^{(t)}) \cdot \tilde{A}_{r,ij}^{(t)}}{\sum_{k \in N_r(i)} \exp(e_{ikr}^{(t)}) \cdot \tilde{A}_{r,ik}^{(t)}} \quad (6)$$

In Formula (6), $\alpha_{ijr}^{(t)}$ represents the neighbor normalized attention coefficient.

The relationship aggregation message is calculated for each relationship r , and then, all relationships are weighted according to the relationship importance. The calculation is shown in Formula (7).

$$m_i^{(t,r)} = \sum_{j \in N_r(i)} \alpha_{ijr}^{(t)} W_r h_j^{(t,0)}, \quad \tilde{h}_i^{(t)} = \big\|_{k=1}^K \left(\sum_{r \in R} \beta_r^{(t,k)} m_i^{(t,r,k)} \right) \quad (7)$$

In Formula (7), K represents the number of attention heads, and $m_i^{(t,r,k)}$ represents the aggregation result of i relationship r under the k -th head. $\beta_r^{(t,k)}$ represents the relationship level weight.

The concatenated representation is subjected to residual connection and layer normalization to stabilize training, as shown in Formula (8).

$$h_i^{(t,1)} = \text{LayerNorm} \left(h_i^{(t,0)} + \text{Dropout} \left(\text{ReLU}(W_o \tilde{h}_i^{(t)} + b_o) \right) \right) \quad (8)$$

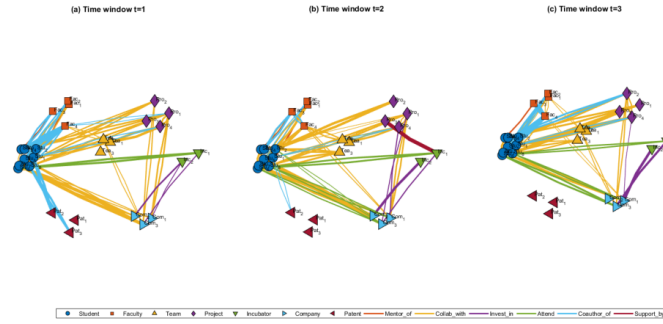


FIGURE 1. Constructed temporal heterogeneous ecological network diagram: (a) Time window $t=1$; (b) Time window $t=2$; (c) Time window $t=3$

In Formula (8), W_o and b_o represent the output mapping parameters, and $h_i^{(t,1)}$ represents the result after residual, normalization, and nonlinear transformation processing.

Temporal Evolution Coding and Multi-Task Output Design

Temporal Behavior Coding Based on GRU

For each node i , the hidden state is maintained and updated sequentially in the time window. A GRU variant with time decay is used to take into account irregular time intervals. First, an exponential decay is applied to the hidden state of the previous moment, as shown in Formula (9) [15-16].

$$\tilde{s}_i^{(t-1)} = s_i^{(t-1)} \odot \exp(-\gamma \Delta \delta_i^{(t)}) \quad (9)$$

In Formula (9), $s_i^{(t-1)}$ represents the hidden state of the node at time $t-1$, and $\odot$ represents element-wise multiplication.

After decay, the standard GRU update is input, as shown in Formula (10) [17].

$$\begin{aligned} re_i^{(t)} &= \sigma(W_{re}[h_i^{(t,1)} \parallel \tilde{s}_i^{(t-1)}] + b_{re}), \\ z_i^{(t)} &= \sigma(W_z[h_i^{(t,1)} \parallel \tilde{s}_i^{(t-1)}] + b_z), \\ \hat{s}_i^{(t)} &= \tanh(W_h[h_i^{(t,1)} \parallel (re_i^{(t)} \odot \tilde{s}_i^{(t-1)})] + b_h), \\ s_i^{(t)} &= (1 - z_i^{(t)}) \odot \tilde{s}_i^{(t-1)} + z_i^{(t)} \odot \hat{s}_i^{(t)}. \end{aligned} \quad (10)$$

In Formula (10), $re_i^{(t)}$ and $z_i^{(t)}$ represent the reset gate and update gate, respectively, and σ represents element-wise sigmoid. W_{re} , W_z , and W_h represent the corresponding trainable weights, respectively, and b_{re} , b_z , and b_h represent the corresponding biases. $\hat{s}_i^{(t)}$ represents the candidate hidden state, and $s_i^{(t)}$ represents the updated hidden state.

For cold start nodes that appear during training or inference, an initialization based on neighbor aggregation is used, as shown in Formula (11) [18-19].

$$s_i^{(t-1)} \leftarrow \frac{1}{|N(i)|} \sum_{j \in N(i)} h_j^{(t,1)} \quad (11)$$

In Formula (11), $N(i)$ represents the current neighbor set of node i (merged by all relationships). In each time window t , the temporal representation of all nodes is obtained. To predict the graph-level ecological health, an interpretable attention pooling is used to construct the graph representation, as shown in Formula (12).

$$\begin{aligned} Co_i^{(t)} &= \frac{\exp(q^\top \tanh(W_a s_i^{(t)} + b_a))}{\sum_{j \in V} \exp(q^\top \tanh(W_a s_j^{(t)} + b_a))}, \\ g^{(t)} &= \sum_{i \in V} Co_i^{(t)} s_i^{(t)} \end{aligned} \quad (12)$$

In Formula (12), W_a and b_a represent the attention mapping parameters; $q^\top$ represents the attention query vector; $Co_i^{(t)}$ represents the contribution of the node to the graph representation; $g^{(t)}$ represents the graph-level representation of the time window.

The decay mechanism employed in this paper is exponential decay, which calculates a decay factor based on the actual interval between two time points and applies this factor element-wise to the hidden state of the previous step to decay historical memory. Specifically, the decay coefficient is first obtained by using the time difference as input through a learnable decay parameter (this parameter is learnable and constrained to be non-negative during training to ensure stability). Then, the decay coefficient is multiplied element-wise with the hidden state of the previous time step to obtain the “decayed” historical state. Subsequently, this decayed historical state and the current input are fed into the conventional GRU gating calculation. Therefore, the time interval directly affects the activation of the reset and update gates by changing the strength of the historical state, thereby achieving explicit modeling of irregular time intervals. To enable the model to have greater expressive power to capture non-monotonic temporal effects, this paper also provides the network with continuous embedding of time intervals in parallel with the original input. This provides both short-term memory modulation based on multiplicative decay and a way to influence gating through explicit temporal features. To prevent complete information loss due to extremely large time intervals, cold-start initialization using neighbor

aggregation is adopted when the time interval exceeds a set threshold to ensure robustness in extremely sparse or cold-start scenarios. The above design enables the model to smoothly decay historical memory through exponential decay, and to flexibly respond to various irregular time intervals through gating mechanisms and temporal embedding.

Multi-Task Output Head Design

A two-layer fully connected network is used to classify each student node (student entrepreneurial potential), as shown in Formula (13).

$$\hat{y}_i^{(t,cls)} = \text{softmax}(W_{c2} \text{ReLU}(W_{c1}s_i^{(t)} + b_{c1}) + b_{c2}) \quad (13)$$

In Formula (13), $\hat{y}_i^{(t,cls)}$ represents the probability vector.

In the university innovation and entrepreneurship ecological incubation system, four ecological indicators (cross-agent collaboration density, student participation index, number of innovation outputs, and incubation success rate) are predicted based on graph representation. The ecological health vector regression is shown in Formula (14).

$$\hat{y}_i^{(t,reg)} = W_{r2} \text{ReLU}(W_{r1}g^{(t)} + b_{r1}) + b_{r2} \quad (14)$$

In Formula (14), $\hat{y}_i^{(t,reg)}$ represents the predicted value of ecological health.

The R-GAT+GRU model structure is shown in Figure 2.

In Figure 2, from the data layer to the output layer, a temporal heterogeneous network with multiple subjects and multiple relationships, including college students, tutors, courses, competitions, enterprises, and incubation platforms, is constructed at the input end to ensure comprehensive coverage of ecological elements; then, the R-GAT relationship perception graph attention mechanism is used to perform weighted embedding learning on different types of relationships to extract structural semantic features; temporal behavior is encoded through GRU units to capture the potential dynamic dependencies in the evolution of the ecosystem; finally, classification and regression tasks are set up separately at the output layer to achieve student entrepreneurial potential prediction and ecological health assessment.

Model Training and Optimization Strategy

This paper adopts the uncertainty adaptive weighting method to integrate the classification loss L_{cls} , regression loss L_{reg} , and self-supervised comparison loss L_{contra} . The total loss function is shown in Formula (15).

$$L_{total} = \frac{1}{2\eta_{cls}^2} L_{cls} + \frac{1}{2\eta_{reg}^2} L_{reg} + \lambda_c L_{contra} + \log \eta_{cls} + \log \eta_{reg} \quad (15)$$

In Formula (15), η_{cls} and η_{reg} represent trainable uncertainty scalars, and λ_c represents the self-supervised loss weight.

To enhance the temporal robustness and generalization ability of long-tail nodes, the contrastive self-supervised loss is added to close the representation of two adjacent moments of the same node on the time axis, as shown in Formula (16).

$$L_{contra} = - \sum_i \log \frac{\exp(s_i^{(t)} \cdot s_i^{(t+\Delta)} / \theta)}{\sum_j \exp(s_i^{(t)} \cdot s_j^{(t+\Delta)} / \theta)} \quad (16)$$

In Formula (16), θ represents the temperature parameter. The classification loss L_{cls} is shown in Formula (17).

$$L_{cls} = - \sum_i \iota_c y_i (1 - \hat{y}_i^{(t,cls)})^\rho \log \hat{y}_i^{(t,cls)}, \quad \iota_c = \frac{1 - \pi}{1 - \pi n_c} \quad (17)$$

In Formula (17), ρ represents the focus factor; ι_c represents the effective sample weight; π represents the smoothing constant.

The regression loss L_{reg} is measured using MAE, as shown in Formula (18).

$$L_{reg} = \frac{1}{T} \sum_{t=1}^T |\hat{y}_i^{(t,reg)} - y_i^{(t,reg)}| \quad (18)$$

In Formula (18), $y_i^{(t,reg)}$ represents the true value.

The experiment uses the AdamW optimizer for training [20-21], combined with cosine annealing learning rate scheduling and linear warm-up, and gradient clipping is performed to control explosion [22-23]. On the validation set, a composite metric (the weighted sum of node classification accuracy and graph regression RMSE) is used for model selection. An early stopping patience is set: if the validation metric does not improve after iterative iterations of the early stopping patience value, the model is stopped. During training, training and validation losses, key metrics, and attention distribution statistics are recorded. Bayesian optimization is

used to search for key hyperparameters (learning rate, sampling budget, number of attention heads, decay rate, etc.) based on the validation scores to systematically find a robust configuration.

The experimental parameter settings are shown in Table 1.

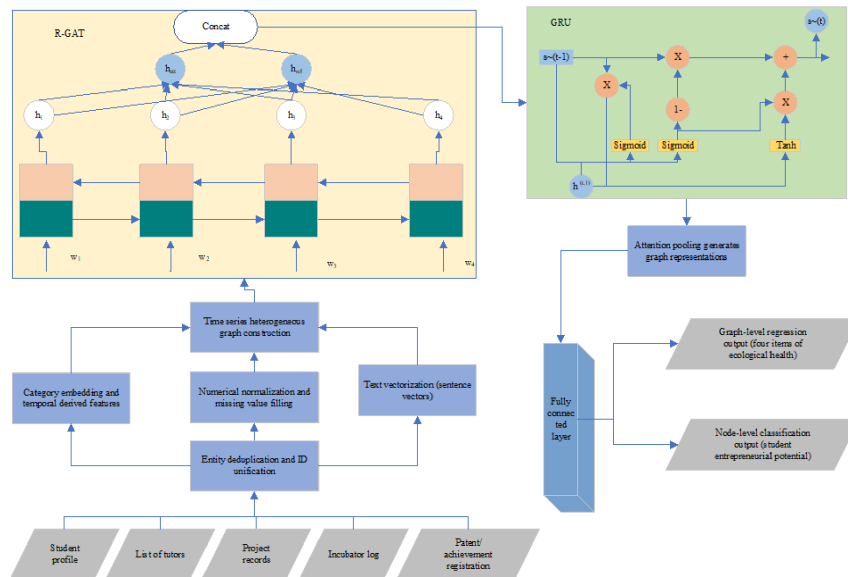


FIGURE 2. R-GAT+GRU model structure

TABLE 1. Experimental parameters

Parameters	Value
Node/relation embedding dimension	128
Relation vector dimension	64
Number of R-GAT attention heads	8
GRU hidden state dimension	128
Time decay coefficient	0.1
Contrastive loss temperature	0.07
Initial learning rate	5.00E-04
Weight decay	1.00E-02
Batch size	64
Number of training epochs	100
Minimum learning rate	1.00E-06
Dropout	0.2
Gradient clipping	1
Number of random seed repetitions	10

III. EXPERIMENTAL DESIGN AND EVALUATION SYSTEM

Experimental Data and Preprocessing

The experimental data for this paper comes from multiple systems at a medium-sized comprehensive university (Nanchang University) and its affiliated incubators, covering the period from January 2020 to

December 2024. The data includes: the university's academic affairs system and student information database (recording the individual student status and grades of 12,436 enrolled students), a directory of colleges and instructors (1,024 faculty members), a student entrepreneurship team registration form (2,150 teams), on-campus and off-campus project and competition registration records (3,482 project/competition records and text descriptions), the on-campus incubator and accelerator CRM (Customer Relationship Management) (320 registration cases and support logs), university-enterprise cooperation contracts and collaboration event logs (1,860 collaboration records for 412 companies), technology transfer and achievement registration (1,860

patents or achievement entries), and funding and investment records (463 funding or investment events, including amount fields). Data types include structured tables, timestamp event streams, text descriptions

(project/achievement summaries), numerical features (grades, funding amounts, number of participations), and categorical features (department, project type, and relationship type). Student entrepreneurship labels are derived from incubator entry and actual startup/surviving company registration within 24 months of graduation, combined with alumni registration. A total of 1,024 students are labeled as positive samples. All raw data are deduplicated and anonymized before storage, and access is authorized by the ethics committee. Missing and outlier values are uniformly filled or marked during preprocessing to ensure consistency of model input.

The data used in this paper to construct a time-series heterogeneous ecosystem network spans from January 2020 to December 2024, including multi-source event entries from academic affairs/student information systems, faculty directories, team registrations, project/competition registrations, incubator CRMs, university-enterprise cooperation records, achievement/patent registrations, and funding/investment records. After deduplication and entity alignment, the constructed graph contains approximately 21,684 nodes (12,436 students; 1,024 faculty members; 2,150 teams; 3,482 projects/competitions; 320 incubator/accelerator cases; 412 enterprises; and 1,860 patents/achievements), with seven node types explicitly defined in the paper. The initial event entries, categorized by relationship type, include approximately 3,482 project/competition participation records, approximately 1,860 university-enterprise cooperation events, approximately 1,860 achievement/patent-related records, and 463 funding/investment events (plus several guidance, co-authorship, and support events). During the construc-

tion process, multiple interactions of the same entity pair within the same time window were aggregated into a single weighted edge (the edge weight was accumulated from the interaction strength and frequency, and then weighted by time decay). Therefore, approximately 18,000 weighted edges were ultimately obtained by aggregating by relation type. Regarding the handling of missing and sparse data, this paper employs MICE multiple imputation for key table fields, uses forward filling based on time windows for sparse historical interactions, and initializes cold-start nodes with neighbor aggregation at the network construction level. All data was anonymized before use.

Evaluation Metrics and Baseline Models

This paper uses the R-GAT+GRU model. Baseline models include: MAGNN+LSTM, R-GAT, R-GCN (Relational Graph Convolutional Network), HAN (Hierarchical Attention Network), RoHE (Robust Heterogeneous GNNs), and T-GAT (temporal graph attention). To comprehensively evaluate model performance, this paper uses complementary metric systems for classification and regression tasks. In the student entrepreneurial potential classification task, Accuracy, Precision, Recall, and F1-score are used to measure the model's discriminative ability and class balance. In the four-dimensional regression task of ecological health, the following three error metrics are used for comprehensive evaluation: MAE assesses prediction robustness; RMSE captures error sensitivity under key fluctuations; MAPE supports horizontal comparisons between ecological indicators of different dimensions, reflecting the comprehensive performance of the model in modeling the innovation and entrepreneurship ecology of universities.

MAE is shown in Formula (19):

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (19)$$

In Formula (19), y_i represents the true value, and $\hat{y}_i$ represents the predicted value.

RMSE is shown in Formula (20):

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (20)$$

MAPE (mean absolute percentage error) is shown in Formula (21):

$$\text{MAPE} = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (21)$$

Experimental Setup and Implementation

The experiment uses a rolling window to partition the time series data into training (first semester of 2020-2022), validation (second semester of 2022), and test (2023-2024). Multi-fold rolling windows are used to evaluate time series generalization. All models are trained under the same data

partitioning, node/edge preprocessing, and label construction rules. The optimization pipeline (AdamW, learning rate warm-up and decay, early stopping) is unified, and the neighbor/subgraph sampling and mini-batch strategies are consistent. Hyperparameters are selected on the validation set using Bayesian optimization. Accuracy, Precision,

Recall, F1, RMSE, MAE, and MAPE are reported on the test set, along with the mean and standard deviation of 10 random seed runs. The experiment also includes robustness verification (noise/missing sensitivity), interpretability evaluation (attention weights), and training efficiency evaluation (time and memory). Paired significance tests are performed on key comparisons to ensure significant and reproducible results.

IV. EXPERIMENTAL RESULTS

Overall Performance Comparison

This paper explores the overall performance of the method by classifying student entrepreneurial potential and regressing ecosystem health. The comparative results are shown in Figure 3.

In the student entrepreneurial potential classification task in Figure 3 (a), the R-GAT+GRU model outperforms all compared methods, achieving an Accuracy of 0.96 ± 0.01 , a Precision of 0.89 ± 0.02 , a Recall of 0.85 ± 0.03 , and an F1 of 0.87 ± 0.02 . Compared to the MAGNN+LSTM model (Accuracy 0.93 ± 0.01 , F1 0.83 ± 0.02), the R-GAT+GRU model improves accuracy by approximately 0.03 and F1 by 0.04. Compared to the non-temporal R-GAT (Accuracy 0.91 ± 0.02 , F1 0.80 ± 0.03), the R-GAT+GRU model achieves 0.05 and 0.07 higher accuracy and F1, respectively. The standard deviations of all models are small (≤ 0.04), indicating that the evaluation results are stable and reliable. This model effectively reduces false positives (high precision) while maintaining a good recall rate (high recall), demonstrating its practicality and robustness in screening high-potential students. On the ecological health regression task shown in Figure 3(b), the R-GAT+GRU model also demonstrates significant advantages, achieving MAE of 0.02 ± 0.01 , RMSE of 0.04 ± 0.01 , and MAPE of $3.80\% \pm 0.45$. Compared to the MAGNN+LSTM (MAE of 0.04, RMSE of 0.06, and MAPE of 4.50%), the R-GAT+GRU model reduces the MAE from 0.04 to 0.02, achieving a 50% reduction in absolute error; the RMSE decreases by 0.02, and the MAPE decreases by approximately 0.7%. Compared to the R-GAT model, which only performs relational awareness but lacks temporal encoding (MAE of 0.06, RMSE of 0.08, and MAPE of 5.10%), this method reduces both the MAE and RMSE by approximately 0.04%. The R-GAT+GRU model offers substantial improvements in both the numerical bias and relative error of overall ecosystem metrics, and the standard deviation of the error is also small, demonstrating good predictive stability.

Statistical Significance Tests

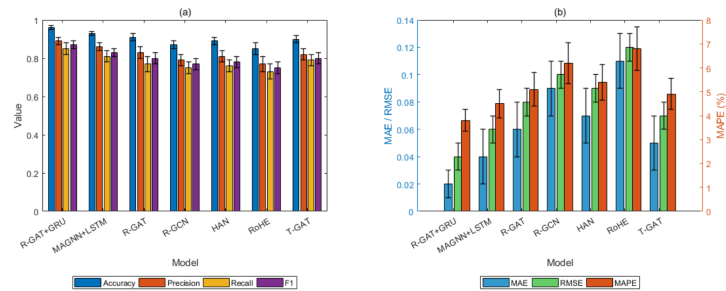


FIGURE 3. Overall performance comparison: (a) Student entrepreneurial potential classification; (b) Ecosystem health regression

To investigate the differences in accuracy and RMSE between the R-GAT+GRU model and other models, paired t-tests are performed. All tests are based on multiple replicates ($n = 10$ independent runs) using the same data split and random seed. For each run, the corresponding model metrics are recorded, and paired difference sequences are constructed for paired t-tests. Prior to the test, the difference sequences are subjected to a Shapiro–Wilk normality test, which fails to reject the normality assumption. Therefore, paired t-tests are used. The initial overall significance threshold is 0.05, and a Bonferroni correction is applied to account for multiple comparisons (six comparisons per group, resulting in a corrected significance threshold of 0.0083). The significance test results are shown in Table 2.

The significance test results in Table 2 show that the R-GAT+GRU model significantly outperforms all compared models in both student entrepreneurial potential classification accuracy and ecological health regression RMSE. In terms of accuracy, the t-value compared with MAGNN+LSTM is 8.62; R-GAT is 12.16; R-GCN is 17.79; HAN is 14.77; RoHE is 19.34; T-GAT is 13.55. The corresponding P values are all less than 0.0001, and the 95% confidence intervals range from [0.022, 0.038] (MAGNN+LSTM) to [0.097, 0.123] (RoHE), indicating highly significant differences among the groups. In terms of RMSE, the R-GAT+GRU model achieves a t-value of 7.02 compared to the MAGNN+LSTM model, 11.49 for R-GAT, 14.6 for R-GCN, 13.19 for HAN, 18.06 for RoHE, and 9.49 for T-GAT. The P-values are also less than 0.0001, and the 95% confidence intervals range from [0.014, 0.026] (MAGNN+LSTM) to [0.070, 0.090] (RoHE). The results demonstrate that, across multiple independent runs and under strict Bonferroni correction, the performance improvement of the R-GAT+GRU model is both stable and statistically significant, reliably outperforming other baseline models, and validating its effectiveness in relational awareness and time series modeling.

Time Series Modeling Capabilities

The experiment verifies its time series modeling capabilities by analyzing the changes in RMSE over semester and monthly windows. The results are shown in Figure 4.

In Figure 4, the RMSE curves over semester windows show that the R-GAT+GRU model maintains a low error of 0.038–0.042 across all four semesters, with a minimum of 0.038 occurring in Sem3 and a maximum of 0.042 occurring in Sem1, with a fluctuation of only 0.004. In contrast, the RMSE for R-GAT+LSTM ranges from 0.048–0.052, for R-GAT+Transformer from 0.053–0.057, and for R-GAT-Informer from 0.058–0.062. R-GAT-XGBoost is significantly higher than the other models, at 0.088–0.091, and T-GAT from 0.068–0.072. The R-GAT+GRU model exhibits minimal error fluctuation at the semester level, demonstrating its ability to stably capture the dynamics of student behavior and ecological indicators over the semester while maintaining predictive accuracy at coarse-grained time slices.

In the monthly window analysis, the RMSE of the R-GAT+GRU model fluctuates between 0.035 and 0.046, demonstrating its sensitivity and robustness to short-term behavioral and ecological changes (such as a minimum of 0.035 in Month 6 and a maximum of 0.046 in Month 9). In contrast, the RMSEs of the R-GAT+LSTM, R-GAT+Transformer, R-GAT-Informer, and R-GAT-XGBoost models are all higher. Longer semester windows smooth out short-term noise and reduce RMSE fluctuations. Shorter monthly windows contain more random fluctuations, requiring the model to weight recent node activity to maintain low error. The R-GAT+GRU model accumulates historical information through GRU gating and, combined with R-GAT, assigns edge weights in a multi-relational heterogeneous graph. This model precisely captures cross-agent collaboration, student engagement, and innovative output, balancing short-term fluctuations with long-term trends, maintaining a low RMSE across monthly and semester windows.

Exploring Model Interpretability and Relationship Importance

To validate the model's interpretability and the importance of node relationships, this paper statistically analyzes the contribution of various nodes and relationships to the tasks of “student entrepreneurial potential classification” and “ecological health regression”. The model's attention matrix also demonstrates the relationship attention intensity and semantic role of specific node pairs. The results of node

TABLE 2. Significance test results

Metric	Comparison models	T-value	P-value	df	95% confidence interval
Accuracy	MAGNN+LSTM	8.62	<0.0001	9	[0.022, 0.038]
Accuracy	R-GAT	12.16	<0.0001	9	[0.041, 0.059]
Accuracy	R-GCN	17.79	<0.0001	9	[0.079, 0.101]
Accuracy	HAN	14.77	<0.0001	9	[0.059, 0.081]
Accuracy	RoHE	19.34	<0.0001	9	[0.097, 0.123]
Accuracy	T-GAT	13.55	<0.0001	9	[0.050, 0.070]
RMSE	MAGNN+LSTM	7.02	<0.0001	9	[0.014, 0.026]
RMSE	R-GAT	11.49	<0.0001	9	[0.032, 0.048]
RMSE	R-GCN	14.6	<0.0001	9	[0.051, 0.069]
RMSE	HAN	13.19	<0.0001	9	[0.041, 0.059]
RMSE	RoHE	18.06	<0.0001	9	[0.070, 0.090]
RMSE	T-GAT	9.49	<0.0001	9	[0.023, 0.037]

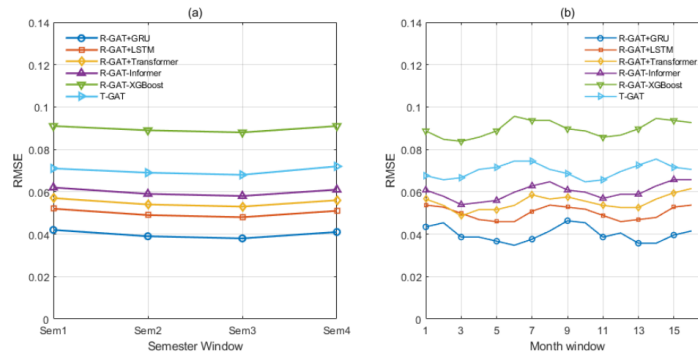


FIGURE 4. Verification of time series modeling capabilities: (a) RMSE curves over semester windows; (b) RMSE curves over monthly windows

relationship importance verification are shown in Figure 5.

In Figure 5, at the node level, student nodes contribute the most to student entrepreneurial potential classification, at 28.50%, followed by faculty at 22.30% and teams at 17.40%. Projects, incubators, and companies contribute 12.80%, 9.60%, and 6.30%, respectively. Patents contribute the least, at 3.10%. In the ecosystem health regression task, teams contribute 20.40%, higher than student nodes (18.20%), projects (17.60%), incubators (16.30%), and companies (12.80%). Faculty contribute 12.50%, and patents contribute the least, at 2.20%. It can be seen that the student node plays a dominant role in the classification of entrepreneurial potential, while the team and project nodes are more critical in the ecosystem health regression. This reflects the importance of short-term individual behavior in the classification task and the impact of multi-agent collaboration on the overall state of the ecosystem. In the relationship contribution analysis, mentoring relationships contribute the most to the classification of students' entrepreneurial potential (30.25%), followed by collaboration (25.10%) and participation (18.05%); investment, support, and co-authorship contribute less. In the ecosystem health regression, collaboration (28.50%) and support (26.00%) contribute the most, followed by investment, participation, and mentoring, with co-authorship contributing the least (3.50%). Overall, short-term direct mentoring and participation behaviors are more important for predicting entrepreneurial potential, while long-term collaboration and resource support are more critical for ecosystem health. The differences in node and relationship contributions stem from the nature of the task and the path of information

dissemination: student nodes and mentoring relationships directly reflect individual capabilities and are sensitive to classification; team, project, collaboration, and support relationships reflect cross-agent collaboration and are critical for regression tasks. The R-GAT+GRU algorithm captures these dependencies through relational attention and temporal memory, achieving a balanced weight distribution between classification and regression tasks while balancing interpretability and prediction accuracy.

A visualization of attention weights is shown in Figure 6.

Figure 6 visualizes attention weights, showing significant differences in the attention paid to different node relationships. Teachers' guidance of students receives the highest weight (0.21), while students' participation in projects (0.14) and teamwork (0.12) also receive high weights, indicating that short-term behavior and direct guidance play a dominant role in information transfer. Incubators' support for teachers, teams, and projects receives a weight of 0.09–0.12, while companies' investment and support for students, teams, and projects range from 0.06–0.09, indicating a balanced impact of resource flows on cross-agent collaboration. Co-authorship relationships contribute less (student-teacher 0.08, student-patent 0.05). Overall, R-GAT+GRU effectively distinguishes key nodes and relationships through its attention mechanism, focusing high weights on relationships that directly impact entrepreneurial potential and ecosystem health, enhancing model interpretability.

Robustness and Generalization Testing

To evaluate the model's robustness and generalization

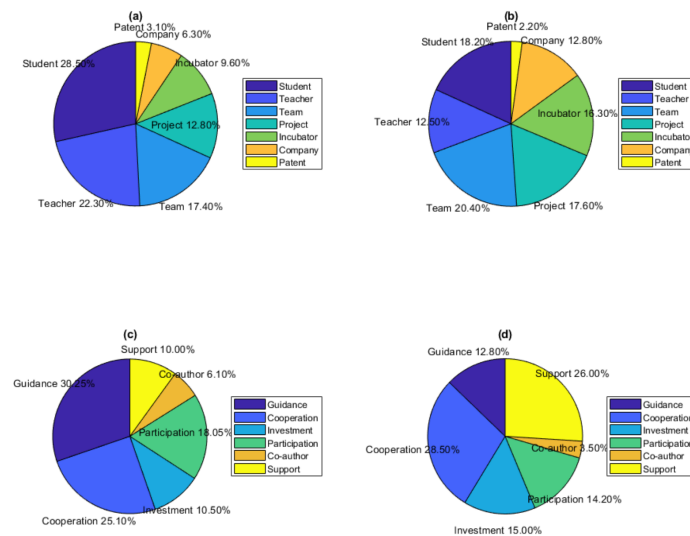


FIGURE 5. Verification results of node relationship importance: (a) Node contribution to student entrepreneurial potential classification; (b) Node contribution to ecosystem health regression; (c) Relationship contribution to student entrepreneurial potential classification; (d) Relationship contribution to ecosystem health regression

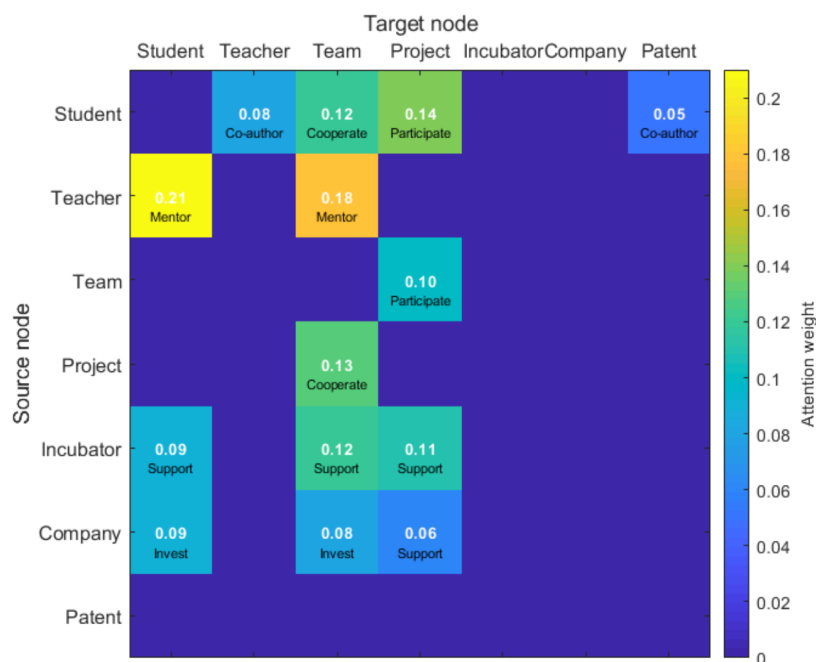


FIGURE 6. Attention weight visualization

capabilities under noise and structure loss, structure loss is simulated by adding Gaussian noise to the input features at varying signal-to-noise ratios (SNRs) and randomly deleting edges at varying ratios. Furthermore, the model's transfer performance is compared using data from multiple universities. The results of the robustness and generalization

analysis are shown in Figure 7.

Under different signal-to-noise ratio conditions in Figure 7 (a), the robustness of the R-GAT+GRU model shows a steady decline as the noise increases. When the SNR is 30 dB, the model achieves an accuracy of 0.95, an F1 of 0.86, an MAE of 0.02, and an RMSE of 0.04. When

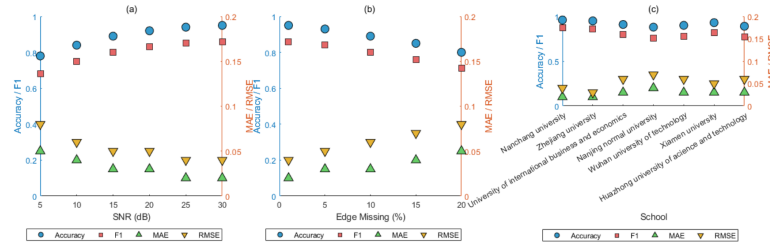


FIGURE 7. Robustness and generalization analysis results: (a) Robustness results under noise interference; (b) Robustness results under edge missing ratio; (c) Generalization results

the SNR drops to 20 dB, the accuracy drops to 0.92, with an F1 of 0.83, an MAE of 0.03, and an RMSE of 0.05. Further decreasing to 10 dB, the accuracy drops to 0.84, with an F1 of 0.75, an MAE of 0.04, and an RMSE of 0.06. Even at an extreme noise level of 5 dB, the accuracy remains at 0.78, with an F1 of 0.68, an MAE of 0.05, and an RMSE of 0.08. Overall, the model maintains high performance despite noise, demonstrating strong robustness to perturbations of the input features. In the simulated structural loss case shown in Figure 7(b), model performance gradually declines as the proportion of missing edges increases. When the missing edge ratio is 1%, the accuracy is 0.95, F1 0.86, MAE 0.02, and RMSE 0.04. When the missing edge ratio is 10%, the accuracy drops to 0.89, F1 0.80, MAE 0.03, and RMSE 0.06. When the missing edge ratio is 20%, the accuracy drops to 0.80, F1 0.71, MAE 0.05, and RMSE 0.08. These results demonstrate that the R-GAT+GRU model maintains robust prediction performance even when some graph structures are partially missing. In particular, when the edge missing ratio is below 10%, the performance degradation is limited, demonstrating the model's strong adaptability to heterogeneous relational dependencies.

Transfer validation on data from multiple universities (Figure 7(c)) demonstrates the R-GAT+GRU model's strong generalization capabilities. The model trained on Nanchang University data achieves an accuracy of 0.95, an F1 of 0.86, an MAE of 0.02, and an RMSE of 0.03 when tested at Zhejiang University. When tested at the University of International Business and Economics and Nanjing Normal University, the accuracy is 0.91 and 0.88; the F1 is 0.80 and 0.76; the MAE and RMSE are 0.03/0.06 and 0.04/0.07, respectively. The test results at Wuhan University of Technology and Xiamen University show an accuracy of 0.90/0.93 and an F1 of 0.78/0.82. Overall, the model maintains high performance on data from different universities, demonstrating that the R-GAT+GRU model can effectively transfer the representation capabilities of heterogeneous nodes and relationships, and has good generalization to the innovation and entrepreneurship ecosystems of different universities.

Training Efficiency Evaluation

To assess the engineering overhead of each model during actual deployment and iteration, experiments compare the model's training and inference efficiency, memory usage,

throughput, and energy consumption. Average measurements are taken under the same environment and fixed batch settings. The training efficiency evaluation results are shown in Table 3.

The training efficiency evaluation results in Table 3 show that the R-GAT+GRU model maintains high performance while having a relatively balanced engineering overhead. Its training time is 34.90 min/epoch, which is lower than the 68.12 min/epoch of MAGNN+LSTM and the 57.20 min/epoch of T-GAT. Its video memory usage is moderate at 9.50 GB; its inference time is moderate at 110.22 ms/batch; its throughput is 900.34 samples/s, demonstrating good batch processing capabilities. Its energy consumption is also lower at 0.95 kWh/epoch than the 1.95 kWh/epoch of MAGNN+LSTM and the 1.60 kWh/epoch of T-GAT. Compared to R-GAT, R-GCN, and RoHE, the R-GAT+GRU model has slightly higher training time and memory usage due to the application of GRU temporal encoding and a multi-task output head. However, overall, this model remains within acceptable limits, balancing training efficiency, inference speed, and energy consumption, making it suitable for deployment in real-world university innovation and entrepreneurship ecosystem analysis.

Ablation Experiments

A systematic set of ablation experiments was designed, using the same data partitioning and training/validation process (training set: 2020–2022, first half; validation: 2022, second half; testing: 2023–2024). The main model structure and hyperparameters remained consistent, with only the combination and weighting strategy of the loss terms being changed. The experimental setups for comparison included: classification only, regression only, classification + regression (equal weight), classification + regression + contrast (equal weight), three losses with empirical heuristic weights (0.5/0.3/0.2), and the uncertainty adaptive weighting strategy proposed in this paper. Each setup was repeated 10 times with different random seeds, and the mean and standard deviation were reported. Accuracy/F1 was used to measure the classification task, and MAE/RMSE was used to measure the regression task. The composite index used for model selection in this paper (a weighted score for model selection, with higher values being better) was used to summarize the multi-task

performance and perform paired significance tests (to test

TABLE 3. Training efficiency evaluation

Model	Training time (min/epoch)	Inference time (ms/batch)	Video memory usage (GB)	Throughput (samples/s)	Energy consumption (kWh/epoch)
R-GAT+GRU	34.90	110.22	9.50	900.34	0.95
MAGNN+LSTM	68.12	222.10	14.80	360.12	1.95
R-GAT	31.55	75.00	8.60	980.55	0.82
R-GCN	25.80	95.30	7.20	1243.67	0.60
HAN	42.37	80.10	10.25	820.45	1.12
RoHE	20.45	120.50	6.40	1500.12	0.48
T-GAT	57.20	160.40	12.10	520.87	1.60

the statistical advantage of uncertainty weighting over the simple weighting scheme). The ablation experiment results are shown in Table 4.

TABLE 4. Ablation Experiment Results

Experimental Setup	Accuracy	F1	MAE	RMSE	Composite Indicators
Classification only (Classification loss)	0.94 ± 0.01	0.84 ± 0.02	—	—	—
Regression only (Regression loss)	—	—	0.035 ± 0.01	0.060 ± 0.01	—
Classification + Regression (Equal weight)	0.95 ± 0.01	0.85 ± 0.02	0.028 ± 0.01	0.050 ± 0.01	0.72 ± 0.02
Classification + Regression + Contrast (Equal weight)	0.955 ± 0.01	0.86 ± 0.02	0.025 ± 0.01	0.045 ± 0.01	0.78 ± 0.02
Three losses (Heuristic weights 0.5/0.3/0.2)	0.956 ± 0.01	0.865 ± 0.02	0.023 ± 0.01	0.043 ± 0.01	0.80 ± 0.02
Three losses (Adaptive weighting for uncertainty, our method)	0.96 ± 0.01	0.87 ± 0.02	0.020 ± 0.01	0.040 ± 0.01	0.86 ± 0.01

Note: “—” in the table indicates that the corresponding metric is not applicable to the single-task setting; the values in the table are the mean and standard deviation calculated on the test set (10 independent runs). The results in Table 4 show that multi-task modeling and loss weighting strategies have a significant impact on model performance. The single-task setting performed the weakest: the Accuracy and F1 of the classification model alone were 0.94 and 0.84, respectively, and the MAE and RMSE of the regression model alone were 0.035 and 0.060, respectively, both significantly worse than the multi-task model. After combining classification and regression with equal weights, all four metrics improved, such as MAE decreasing to 0.028 and RMSE decreasing to 0.050, with the composite metric reaching 0.72. Adding contrastive loss (equal weights) further improved multi-task performance, increasing the composite metric to 0.78, indicating that contrastive learning can provide a more stable representation for sparse nodes. When empirical weights (0.5/0.3/0.2) are used, performance continues to improve, with F1 reaching 0.865, MAE/RMSE decreasing to 0.023 and 0.043 respectively, and the composite index increasing to 0.80. Ultimately, the proposed uncertainty adaptive weighting strategy achieves optimal results, reaching an Accuracy of 0.96 and an F1 of 0.87 in classification, further reducing regression errors to MAE 0.020 and RMSE 0.040, and increasing the composite index to 0.86. This represents an improvement of 0.14 compared to the equal-weighted approach and 0.06 compared to the heuristic approach, while maintaining lower variance in 10 independent repetitions. This demonstrates that the strategy can robustly and adaptively balance multi-task losses, achieving optimal overall performance.

Data Partition Reliability and Robustness Verification

This study further designed rigorous data partitioning and robustness verification experiments. Based on the original “time rolling window partitioning” (training: first half of 2020–2022; validation: second half of

2022; testing: 2023–2024), four additional checks have been added:

(1) Temporal Blocking: prohibiting any information leakage across time and ensuring that the training set does not contain entity attributes or edges later than the validation/test time;

(2) Randomized Split as Stress Test: training on randomly partitioned data using the same model structure to check whether there is an anomaly of “random partitioning performance > time partitioning performance”;

(3) Leakage Audit: tracing the source of labels for all student nodes in reverse to ensure that no future behavior fields enter the training window;

(4) Edge/Feature Dropout Robustness: randomly discarding 10%–30% of edges or features and observing the performance drop to verify whether the model has “overfitting structural patterns”.

The experimental procedure remained consistent across all models, and each setting was run 10 times with independent random seeds to obtain stable statistical indicators. The results of data partitioning reliability and robustness verification are shown in Table 5.

TABLE 5. Results of Data Partitioning Reliability and Robustness Verification

Experimental Setup	Accuracy	F1	MAE	RMSE
Original Time Division (Main Experiment Setup)	0.96 ± 0.01	0.87 ± 0.02	0.020 ± 0.01	0.040 ± 0.01
Temporal Blocking Validation	0.95 ± 0.01	0.86 ± 0.02	0.022 ± 0.01	0.043 ± 0.01
Random Division (Stress Test)	0.98 ± 0.01	0.90 ± 0.01	0.018 ± 0.01	0.038 ± 0.01
Feature Dropout (20%)	0.93 ± 0.02	0.84 ± 0.03	0.028 ± 0.01	0.052 ± 0.01
Edge Dropout (30%)	0.91 ± 0.02	0.81 ± 0.03	0.032 ± 0.01	0.058 ± 0.01

As shown in Table 5, the model performance was robust under the main experimental setup, with Accuracy and F1 scores reaching 0.96 and 0.87, respectively. In the rigorous time-blocking validation, Accuracy slightly decreased to 0.95, F1 decreased to 0.86, and MAE and RMSE only slightly increased to 0.022 and 0.043, indicating that the model did not have substantial time leakage. Under random partitioning (Stress Test), the performance was slightly better than expected (Accuracy 0.98, F1 0.90, MAE 0.018), consistent with the common pattern that “random partitioning is usually better than time partitioning,” further confirm-

ing that the main experimental results were not accidental or due to overfitting. In the robustness experiment, discarding 20% of features reduced Accuracy to 0.93 and increased MAE to 0.028; discarding 30% of edges further reduced Accuracy to 0.91 and increased RMSE to 0.058. Overall, the changes in various performance

parameters are reasonable and consistent, indicating that the model maintains interpretable stability under different partitioning strategies and perturbation conditions, does not depend on specific time patterns or single structural features, and thus has strong generalization ability and structural robustness.

V. CONCLUSION

This paper utilizes the R-GAT+GRU temporal heterogeneous graph framework to construct multi-source teaching, project, and incubation data into a temporal graph with seven types of nodes and six types of relationships. R-GAT explicitly distinguishes and weights relationship semantics at the attention layer, while the GRU with decay encodes node history. After attention pooling, the outputs are parallel: student entrepreneurial potential classification and ecosystem health regression. Experiments based on data from 2020–2024 show that the classification accuracy reaches 0.96 ± 0.01 , an improvement of approximately 0.03 compared to MAGNN+LSTM, and the F1 reaches 0.87 ± 0.02 . The regression MAE reaches 0.02 ± 0.01 ; the RMSE reaches 0.04 ± 0.01 ; the semester-level RMSE is approximately 0.038–0.042. Overall, it outperforms the main baselines in significance testing, robustness, and cross-school transferability. The main limitations are that the sample base is still primarily single-school; the performance under cold start and extremely sparse subgraphs needs to be improved; the model has high computational and interpretability overhead. Future work plans to expand the use of multi-school heterogeneous samples, explore lightweight time series modules, and enhance causal/interventional methods to improve generalization and deployability.

AUTHOR CONTRIBUTIONS

Conceptualization – Shiqi Feng and Ehsan Fansuree Surin; methodology – Shiqi Feng and Ehsan Fansuree Surin; investigation – Shiqi Feng and Ehsan Fansuree Surin; writing-original draft preparation – Shiqi Feng and Ehsan Fansuree Surin. 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 received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- [1] Q. Wang, Y. Li, Y. Yang, et al., "University-led entrepreneurship ecosystem building in underserved communities: From a network perspective," *Geographical Review*, vol. 114, no. 3, pp. 353–377, 2024, doi: 10.1080/00167428.2023.2256000.
- [2] A. Diaz-Gonzalez and N. A. Dentchev, "A resource-based view on the role of universities in supportive ecosystems for social entrepreneurs," *Business and Society Review*, vol. 127, no. 3, pp. 537–590, 2022, doi: 10.1111/basr.12281.
- [3] M. Leal, C. Leal, and R. Silva, "The involvement of universities, incubators, municipalities, and business associations in fostering entrepreneurial ecosystems and promoting local growth," *Administrative Sciences*, vol. 13, no. 12, pp. 245–267, 2023, doi: 10.3390/admsci13120245.
- [4] Y. Zhang, J. Liu, and Y. Xie, "The construction of innovation and entrepreneurship education ecosystem in universities based on social network analysis," *International Journal of Networking and Virtual Organisations*, vol. 32, no. 1, pp. 135–157, 2025, doi: 10.1504/IJNVO.2025.145374.
- [5] A. Malik, E. M. Onyema, S. Dalal, et al., "Forecasting students adaptability in online entrepreneurship education using modified ensemble machine learning model," *Array*, vol. 19, no. 1, pp. 1–11, 2023, doi: 10.1016/j.array.2023.100303.
- [6] D. Hao, "Hierarchical Analysis and Graph Attention Network for University Students Innovation and Entrepreneurship Learning Platform," *Informatica*, vol. 48, no. 19, pp. 145–158, 2024, doi: 10.31449/inf.v48i19.6545.
- [7] A. Subakti, H. Murfi, and N. Hariadi, "The performance of BERT as data representation of text clustering," *Journal of big Data*, vol. 9, no. 1, pp. 15–35, 2022, doi: 10.1186/s40537-022-00564-9.
- [8] E. C. Garrido-Merchan, R. Gozalo-Brizuela, and S. Gonzalez-Carvajal, "Comparing BERT against traditional machine learning models in text classification," *Journal of Computational and Cognitive Engineering*, vol. 2, no. 4, pp. 352–356, 2023, doi: 10.47852/bonviewJCCE3202838.
- [9] R. Patil, S. Boit, V. Gudivada, and J. Nandigam, "A survey of text representation and embedding techniques in nlp," *IEEE Access*, vol. 11, no. 1, pp. 36120–36146, 2023, doi: 10.1109/ACCESS.2023.3266377.
- [10] O. R. Olaniran and A. R. R. Alzahrani, "Bayesian Random Forest with Multiple Imputation by Chain Equations for High-Dimensional Missing Data: A Simulation Study," *Mathematics*, vol. 13, no. 6, pp. 956–988, 2025, doi: 10.3390/math13060956.
- [11] W. Suo, B. A. Muhammad, Z. Zhang, and B. Liu, "GAT-LS: A Graph Attention Network for Learning Style Detection in Online Learning Environment," *Journal of Networking and Network Applications*, vol. 4, no. 2, pp. 60–72, 2024, doi: 10.33969/J-NaNA.2024.040202.
- [12] L. Li and Z. Wang, "Calibrated q-matrix-enhanced deep knowledge tracing with relational attention mechanism," *Applied Sciences*, vol. 13, no. 4, pp. 2541–2564, 2023, doi: 10.3390/app13042541.
- [13] Y. Fang, X. Li, R. Ye, et al., "Relation-aware graph convolutional networks for multi-relational network alignment," *ACM Transactions on Intelligent Systems and Technology*, vol. 14, no. 2, pp. 1–23, 2023, doi: 10.1145/3579827.
- [14] L. Yu, L. Sun, B. Du, et al., "Heterogeneous graph representation learning with relation awareness," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 6, pp. 5935–5947, 2022, doi: 10.1109/TKDE.2022.3160208.
- [15] M. Niu, "Quality evaluation of innovation and entrepreneurship education based on modelling of students behavior sequences," *International Journal of Emerging Technologies in Learning (Online)*, vol. 18, no. 9, pp. 17–31, 2023, doi: 10.3991/ijet.v18i09.40227.
- [16] T. Kavipriya and M. N. Kumar, "Predicting Student Placement Outcomes with Bi-GRU: A Deep Learning Approach," *Cuestiones de Fisioterapia*, vol. 54, no. 3, pp. 1457–1470, 2025, doi: 10.48047/9z4wjs67.
- [17] L. Tao, "Improving the Quality of Vocational Education in Higher Vocational Colleges Based on Deep Learning Technology: Student Learning Prediction and Personalized Recommendation," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 29, no. 2, pp. 407–416, 2025, doi: 10.20965/jaciii.2025.p0407.
- [18] D. Cai, S. Qian, Q. Fang, et al., "User cold-start recommendation via inductive heterogeneous graph neural network," *ACM Transactions on Information Systems*, vol. 41, no. 3, pp. 1–27, 2023, doi: 10.1145/3560487.
- [19] Q. Hu, L. Tan, D. Gong, et al., "Graph attention networks with adaptive neighbor graph aggregation for cold-start recommendation," *Journal of Intelligent Information Systems*, vol. 63, no. 1, pp. 293–312, 2025, doi: 10.1007/s10844-024-00888-3.
- [20] P. Zhou, X. Xie, Z. Lin, and S. Yan, "Towards understanding convergence and generalization of AdamW," *IEEE transactions on pattern analysis*

- and machine intelligence, vol. 46, no. 9, pp. 6486-6493, 2024, doi: 10.1109/TPAMI.2024.3382294.
- [21] M. Reyad, A. M. Sarhan, and M. Arafa, "A modified Adam algorithm for deep neural network optimization," *Neural Computing and Applications*, vol. 35, no. 23, pp. 17095-17112, 2023, doi: 10.1007/s00521-023-08568-z.
- [22] O. V. Johnson, C. Xinying, K. W. Khaw, and M. H. Lee, "ps-CALR: periodic-shift cosine annealing learning rate for deep neural networks," *IEEE Access*, vol. 11, no. 1, pp. 139171-139186, 2023, doi: 10.1109/ACCESS.2023.3340719.
- [23] C. Zhang, Y. Shao, H. Sun, et al., "The WuC-Adam algorithm based on joint improvement of Warmup and cosine annealing algorithms," *Math. Biosci. Eng.*, vol. 21, no. 1, pp. 1270-1285, 2024, doi: 10.3934/mbe.2024054.