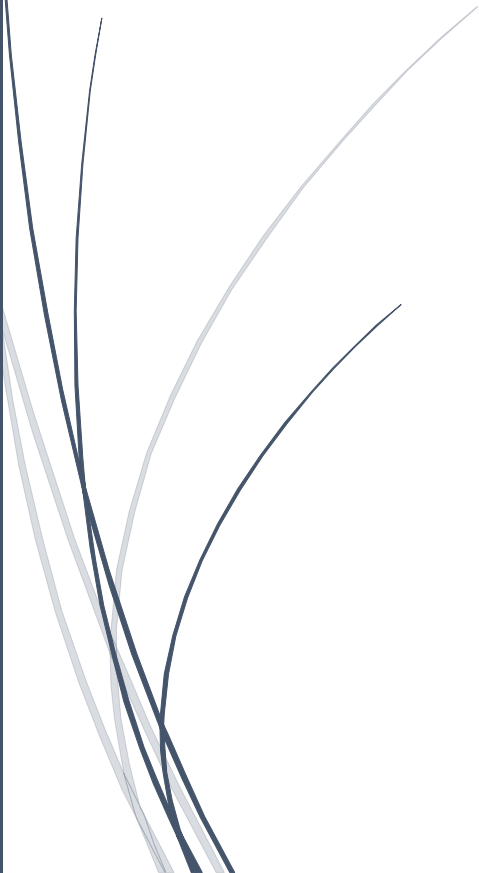


The logo consists of a dark blue vertical bar on the left and a blue arrow pointing right, containing the text "RADemics" in white.

RADemics

Graph Neural Networks for Predicting Course Performance Trends and Curriculum Optimization

Abstract line art consisting of several thin, curved lines in dark blue and light grey, originating from the bottom left and extending upwards and to the right.

A. Thangam, Rajesh

Pondicherry University Community College,
Bharat Institute of Technology

Graph Neural Networks for Predicting Course Performance Trends and Curriculum Optimization

¹A. Thangam, Assistant Professor, Pondicherry University Community College, Lawspet, Pondicherry, India. thangamgri@yahoo.com

²Rajesh, Assistant Professor, Department of Computer science and engineering, Bharat Institute of Technology, Meerut UP, India, rajesh.vision01@gmail.com

Abstract

Graph Neural Networks (GNNs) have emerged as a transformative approach for modeling complex relationships in academic data, enabling predictive analytics for student performance trends and curriculum optimization. Real-world educational datasets often suffer from sparsity, missing records, and class imbalance, which significantly impact model accuracy and generalization. This book chapter explores advanced GNN-based methodologies to address these challenges, focusing on data augmentation, semi-supervised learning, and graph structure refinement techniques. Feature engineering strategies, including imputation-based enhancements, embedding transformations, and multi-source data integration, are presented to improve predictive performance in sparse academic graphs. Methods for handling skewed distributions in multi-class performance prediction are discussed, ensuring fair and unbiased student outcome analysis. The proposed approaches contribute to a more robust, scalable, and interpretable framework for academic performance forecasting, facilitating data-driven decision-making in higher education. Future research directions emphasize adaptive learning mechanisms, self-supervised graph modeling, and explainable AI techniques for enhancing the reliability of educational analytics.

Keywords: Graph Neural Networks, student performance prediction, curriculum optimization, data sparsity, semi-supervised learning, feature engineering.

Introduction

GNNs have emerged as a powerful tool for modeling complex, structured data, making them particularly well-suited for analyzing academic performance trends and optimizing curricula [1]. In traditional educational analytics, student records, course enrollments, and assessment scores are often treated as isolated data points, failing to capture the inherent relationships between different academic elements [2]. Real-world educational systems operate as interconnected networks, where students are linked to courses, instructors, and peer groups in a dynamic manner [3]. GNNs leverage these relationships by propagating information across the graph structure, enabling more comprehensive and context-aware predictions [4]. This capability makes them highly effective in forecasting student success, identifying at-risk learners, and optimizing academic pathways based on performance trends [5]. Despite their advantages, deploying GNNs in educational settings presents several challenges, primarily due to data sparsity, class imbalance, and incomplete student

records. Addressing these issues was crucial for enhancing the accuracy and reliability of academic performance predictions [6].

One of the most significant challenges in applying GNNs to education was data sparsity, which arises due to missing student records, incomplete course enrollment histories, and gaps in assessment data [7]. Sparse data weakens node connectivity in academic graphs, limiting the ability of GNN models to learn meaningful representations [8]. Unlike traditional machine learning models, which often rely on fully structured datasets, GNNs depend on node relationships to propagate information effectively. When key student-course interactions are missing, the performance of predictive models was compromised [9]. For example, students who have taken unconventional academic paths have fewer direct connections to their peers, making it difficult for the model to infer their performance accurately [10,11]. To mitigate this, various graph-based data augmentation techniques, such as edge completion, synthetic node generation, and embedding interpolation, can be employed to enhance connectivity. These methods help create a more representative academic network, improving the robustness of GNNs in handling sparse educational data [12].

Another major concern in student performance prediction was class imbalance, where certain academic outcomes, such as high or failing grades, are significantly underrepresented in the dataset [13]. Most students tend to achieve average grades, leading to a skewed distribution that biases predictive models toward dominant performance categories. This imbalance affects the ability of GNNs to correctly classify students who fall into minority-grade categories, limiting their usefulness for early intervention strategies [14-17]. Addressing this requires specialized techniques, including cost-sensitive learning, resampling methods, and synthetic data generation to balance class distributions. In graph-based learning, reweighting node importance, oversampling underrepresented student groups, and leveraging adversarial training approaches can significantly improve predictive accuracy [18]. Ensuring that GNN models fairly represent all performance levels was essential for providing equitable learning opportunities and enhancing personalized education strategies.