Modeling Movie Success from Collaboration Networks Using Graph Neural Networks

David Corcoran, Gentry Lamb, Adam Stein
Georgetown University

Abstract

In the film industry, success often hinges not only on individual talent but also on the strength and structure of collaboration among cast and crew. This project explores the predictive power of movie collaboration networks – graphs composed of actors, directors, and producers connected via shared film projects – in determining box office success and critical acclaim. We construct two types of networks from a dataset of feature films: a crew collaboration network linking actors and directors, and a movie similarity network based on shared collaborators. Then we calculate centrality measures for individuals in the collaboration network and integrate them with traditional metadata to build regression and classification models predicting both financial and critical success. While regression models demonstrate modest predictive power for normalized revenue, classification models more effectively distinguish financial success. Applying Graph Neural Networks (GNNs) to the movie similarity network further demonstrated the utility of attention-based approaches for capturing structural dynamics, improving classification accuracy for both financial and critical outcomes. Our findings suggest that network centrality features capture valuable structural information that improves predictive performance, particularly for financial outcomes.

Keywords: graph neural networks, social network analysis, collaboration modeling, success prediction, node centrality, predictive modeling, network science

Introduction

The film industry has long been a powerful force in shaping culture and society, while also providing significant economic contributions. The collaborative efforts of a carefully assembled team of actors, directors, producers, and other creatives greatly shape a film's outcome. In recent years, researchers have increasingly recognized the film industry as a complex system that can be modeled as a network, where nodes are defined as individuals and edges between them represent shared projects. This networked perspective opens the door to studying how collaboration patterns correlate with both commercial success and critical acclaim. This project investigates whether the structure of a film's collaboration network can serve as a reliable predictor of its success. Using graph neural networks (GNNs) and social network analysis tools, we aim to model and quantify how a film's cast and crew configuration impacts its performance.

Understanding the relationship between team structure and movie outcomes is not only of theoretical interest but also has practical implications. For casting directors, producers, and studios, predictive insights derived from collaboration graphs could inform more strategic decisions during the planning and pre-production phases. More broadly, this research contributes to the study of team dynamics, creative industries, and graph-based learning by applying advanced machine learning methods to real-world social systems.

Recent work highlights the power of network analysis and machine learning in predicting movie success. Borsboom et al.[1] establish the value of using centrality and network topology to uncover latent dynamics in interdependent systems, a framework directly applicable to modeling collaboration among film professionals. Building on this, Vitelli[6] demonstrates that structural features such as degree and betweenness centrality significantly improve box office prediction when combined with traditional metadata, validating actor-movie networks as effective predictive tools. Complementing these structural approaches, Zhang et al.[9] show that integrating sentiment analysis and emotional mining into neural network models boosts prediction performance, suggesting that audience perception and engagement are important contextual signals.

Choudhary et al.[2] construct collaboration graphs from IMDb and Netflix data to analyze centrality, clustering, and community structures within the film industry. Their work reveals patterns in collaboration networks that reflect both historical trends and emerging content platforms, such as the rise of over-the-top (OTT) media. These findings reinforce the value of structural attributes, such as team connectivity and role prominence, in explaining film performance, offering a methodological foundation for the use of social network analysis (SNA) in our models. Parallel research in recommendation systems also illustrates how graph-based architectures can capture complex relationships within media data. Nesmaoui et al.[5] and Yang et al.[8] apply LightGCN and knowledge graph convolutional networks (KGCN), respectively, to model interactions between users and films. Their results show that incorporating relational structures alongside metadata improves predictive accuracy over traditional baselines. Although these models focus on consumer behavior, their success in leveraging graph-based signals supports our approach of using creative team relationships for prediction.

More broadly, foundational work by Watts and Strogatz[7] on small-world networks highlights how high clustering and short path lengths, both common in real-world social systems, can influence the flow of information and collaboration. These properties likely characterize film industry networks as well, making them especially suitable for graph-based analysis. Similarly, Gao et. al.[3] and Mondal et al.[4] emphasize the flexibility of GNNs in handling sparse and multi-relational data, with applications ranging from recommender systems to multimodal movie analysis. While our project focuses specifically on collaboration networks, their work underscores the general effectiveness of GNNs in modeling complex, real-world structures. Together, these studies provide a strong foundation for our approach, supporting the hypothesis that a film's creative team structure, when modeled as a network, can offer predictive insights into its financial and critical success.

Our project builds on these foundations by constructing a heterogeneous graph consisting of actors, directors, and producers linked via shared film collaborations. We then train GNN models to predict two primary outcomes: (1) box office success, categorized into revenue tiers, and (2) critical acclaim, based on ratings or award recognition. We hypothesize that the structure of a movie's collaboration network influences its likelihood of success, and that graph-based learning can capture these structural patterns more effectively than traditional flat-feature models. Ultimately, this project seeks to build and analyze large-scale collaboration networks drawn from thousands of films, using them as the foundation for predictive modeling. We aim to train models that rely on network-derived features, either alone or in combination with traditional metadata, to predict commercial and critical outcomes. In doing so, we also strive to identify interpretable structural patterns that characterize successful creative teams. By applying graph-based learning to the film industry, our goal is to contribute both practical insights for decision-makers and methodological advances in the application of GNNs to real-world social systems.

Data Collection

The dataset for this study was constructed using two publicly available sources: The Movie Database (TMDB) and the Open Movie Database (OMDB). TMDB served as the primary source, offering extensive coverage of film-level metadata, while OMDB was incorporated to provide additional critical reception metrics and technical attributes. Together, these sources enabled the creation of a comprehensive dataset capturing financial, critical, and production-related aspects of contemporary cinema.

Data from TMDB were obtained through the official Python client and direct API calls. To ensure a diverse and representative sample, films were collected across all available genres, excluding "TV Movie" to focus on theatrical releases. Within each genre, up to 35 pages of results were retrieved, ranked by revenue to prioritize commercially impactful titles. For each film, the collection included identifiers (title, release date, IMDb ID) as well as detailed attributes such as budget, revenue, genres, thematic keywords, and the names of key creative personnel (directors, producers, writers, composers, cinematographers, and top-billed actors). Franchise membership was also recorded. Deduplication was achieved by maintaining a set of visited TMDB IDs, ensuring no film was queried more than once.

Each film was then cross-referenced with OMDB via its IMDb ID to enrich the dataset with attributes not consistently available in TMDB, including IMDb and Metascore ratings, Rotten Tomatoes scores, runtime, awards, and country and language of origin. The resulting dataset contained 7,333 unique theatrical films, exported as a structured CSV for analysis. This raw dataset served as the foundation for all subsequent network construction, feature engineering, and predictive modeling in this study.

Data Cleaning

Following the initial data collection, extensive preprocessing was conducted to ensure the dataset was consistent, interpretable, and suitable for statistical modeling. Duplicate entries were first addressed by distinguishing films with identical titles but different release years (e.g., *Hercules* (1997) vs. *Hercules* (2014)). True duplicates, films with the same title and year, were removed, retaining only the first instance.

Next, data types were standardized. Runtime values were converted to integers, and ratings from IMDb, TMDB, Rotten Tomatoes, and Metacritic were normalized to a 10-point scale. From these, a composite Average Rating was calculated. Release dates were parsed into datetime format, and a *Release_Month* variable was derived to capture potential seasonal effects. Award information was processed to extract the number of Oscar nominations and wins, with an additional binary indicator for whether a film had won at least one Oscar. To support comparability, budget and revenue were z-score standardized, and ROI was calculated as the revenue-to-budget ratio. Invalid or infinite values were replaced with NaN values, and films lacking essential financial data were removed.

To support predictive modeling, list-type variables (e.g., genres, directors, actors, producers, writers, composers, cinematographers, and production companies) were parsed into Python lists, and count-based features (e.g., *num_directors*, *num_writers*, *num_production_companies*) were derived. These features provided concise measures of collaboration and production complexity. For classification tasks, two binary outcomes were created. *Success_Financial* was defined as ROI greater than 1.0, indicating profitability. *Success_Critical* was defined as having an *Average Rating* in the top 25th percentile. To prevent data leakage, the continuous ROI and *Average Rating* variables were excluded once their corresponding binary targets were created.

After cleaning, the final dataset contained 6,947 films, encompassing a diverse set of financial, critical, and production-related attributes, ensuring analytical utility. Separate versions of the dataset were prepared to support regression and classification tasks, each tailored with the features most relevant to its respective modeling framework. The variables included in the analysis are summarized in Table 1 below, with several explained in greater detail in later sections.

Table 1. Final dataset attributes and descriptions

Variable Name	Data Type	Description
IMDB_ID	str	Unique identifier assigned to the movie by IMDb
Title	str	Title of the movie
Year	int	Year the movie was released
Release_Month	int	Month of release in numeric format
Age_Rating	str	Official age classification
Runtime	int	Duration of the movie in minutes
Country	str	Country or countries where the movie was produced
Average_Rating	float	Average of all available ratings
Budget_Normalized	float	Budget normalized by a reference
Budget	int	Production budget of the movie in USD
Revenue_Normalized	float	Revenue normalized using z-score standardization
num_genre	int	Count of genres associated with the movie
num_directors	int	Number of directors listed
num_producers	int	Number of producers listed
num_writers	int	Number of writers listed
num_composers	int	Number of composers listed
num_cinematographers	int	Number of cinematographers listed
num_production_companies	int	Number of production companies involved
DegreeCentrality_mean	float	Mean degree centrality across cast and crew
DegreeCentrality_max	float	Maximum degree centrality across cast and crew
ClosenessCentrality_mean	float	Mean closeness centrality across cast and crew
ClosenessCentrality_max	float	Maximum closeness centrality across cast and crew
BetweennessCentrality_mean	float	Mean betweenness centrality across cast and crew

BetweennessCentrality_max	float	Maximum betweenness centrality across cast and crew
Success_Financial	int	1 if financially successful, 0 otherwise
Success_Critical	int	1 if critically successful, 0 otherwise

Methods

Network Creation

Two networks were created from the cleaned tabular movie dataset to explore patterns of collaboration and structural relationships within the film industry: a crew collaboration network and a movie similarity network. These networks encode different relational structures, one focused on individuals and the other on films, but are derived from the same underlying data.

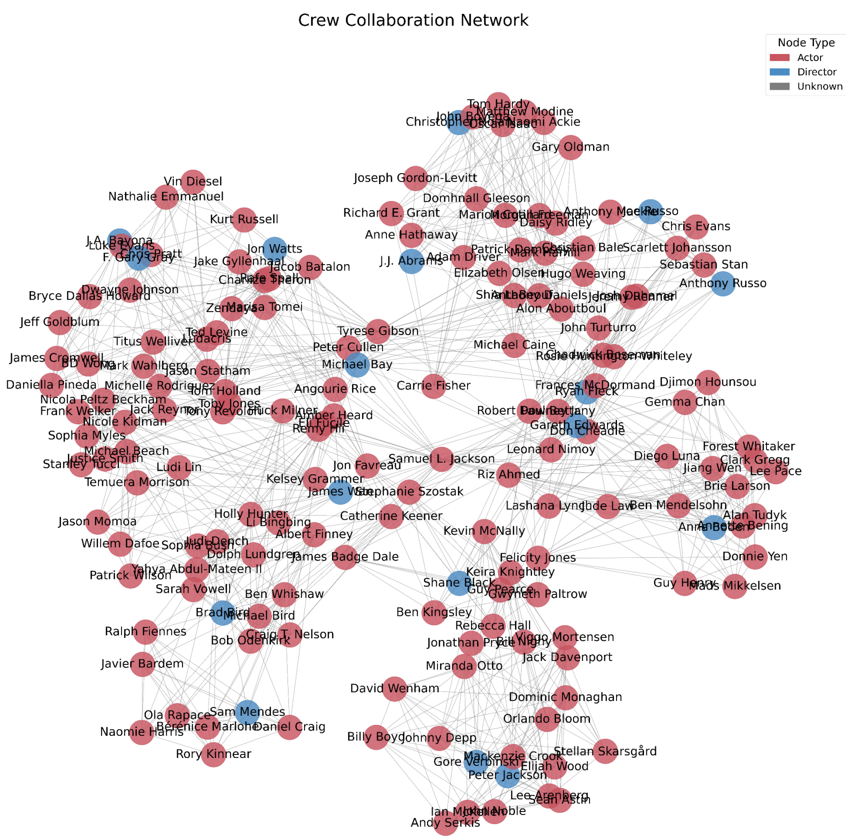

Figure 1. Crew Collaboration Network (Sample of 15 Movies)

The crew collaboration network model's connections between individuals, specifically actors and directors, are based on shared participation in the same films. Each person is represented as a node, enriched with features such as the average rating of their films, normalized budget and revenue, and return on investment (ROI), computed across all associated projects. Edges are added between collaborators (actor–actor, actor–director, or director–director) who have worked on at least one film together. Each edge stores the list of shared films, a weight denoting collaboration frequency, genres, and release years, and aggregated statistics like revenue, ratings, ROI, and Oscar performance. The resulting graph is serialized in GraphML format with structured attributes stored as delimited strings or JSON. For visualization, a spring layout is used to highlight structural clusters, with actors rendered in red and directors in blue. Figure 1 presents a version of this network based on 15 sampled films to enhance interpretability.

In contrast, the movie similarity network models relationships between films based on shared creative personnel. Each movie is a node, annotated with metadata such as ratings (IMDb, Rotten Tomatoes, TMDB), financial metrics (budget, revenue, ROI), runtime, popularity, genre, Oscar performance, and release year. Edges are created between films that share one or more actors or directors, with each edge storing the number of shared collaborators (used as edge weight), a list of common names, and overlapping genres. Like the crew network, this graph is saved in GraphML format. Visualization again uses a spring layout, with node color reflecting normalized degree centrality to highlight highly connected films. Figure 2 displays a sampled version based on 50 movies, shown for illustrative purposes.

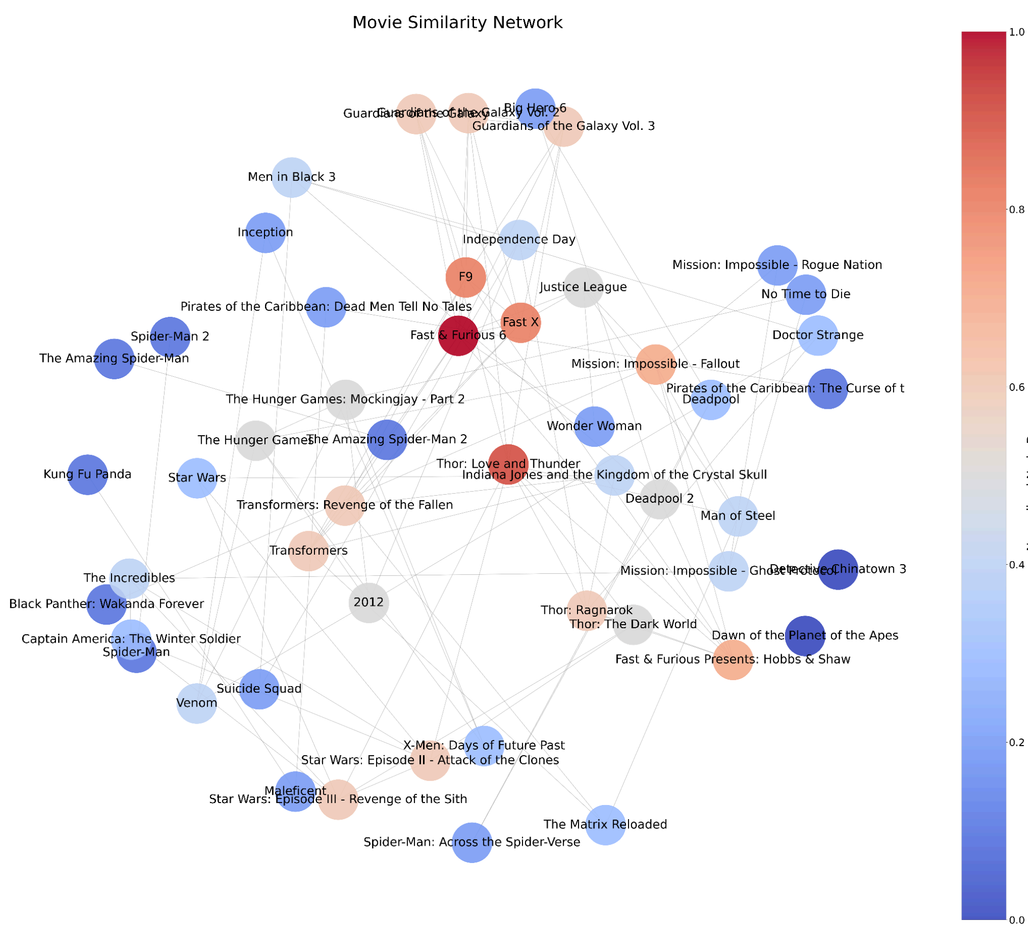

Figure 2. Movie Similarity Network (Sample of 50 Movies)

Together, these two networks offer complementary perspectives: the crew collaboration network reveals professional and social dynamics among individuals in the film industry, while the movie similarity network provides insight into structural and creative proximity among films. Both graphs serve as valuable analytical assets for downstream applications such as community detection, influence modeling, recommendation systems, or graph neural network (GNN) training.

Centrality Measures

To incorporate insights from network analysis into the predictive modeling, centrality measures were derived from the crew collaboration network constructed from the cleaned dataset. In this network, nodes represented individuals involved in film production who collaborated on the same film. This structure captures the extent and nature of professional connections within the film industry and provides a framework for quantifying the prominence of contributors based on their structural positions in the network.

Three centrality metrics were calculated for each individual: degree centrality, closeness centrality, and betweenness centrality. Degree centrality reflects the number of direct collaborations an individual has, providing a measure of professional breadth and immediate access to others. Closeness centrality captures the average distance from a given individual to all others in the network, indicating how quickly that person can reach the broader industry through their connections. Betweenness centrality measures the extent to which an individual lies on the shortest paths between others, thereby identifying contributors who serve as bridges between otherwise disconnected groups. Together, these measures characterize different dimensions of professional influence, ranging from direct connectivity to brokerage roles within the industry.

To link these network-based attributes to individual films, the centrality measures of all credited contributors were aggregated at the film level. For each movie, both the mean and maximum values of each centrality measure were computed across its associated personnel. The mean values indicate the overall connectedness and influence of the film's creative team, while the maximum values highlight the potential impact of particularly prominent individuals, such as a highly connected lead actor or director. During preliminary analyses, the inclusion of both mean and maximum measures led to multicollinearity, as the two variants of each centrality metric were strongly correlated. To mitigate this issue, only the maximum centrality values were retained in the final regression and classification models, as these were more strongly associated with the outcome variables of financial and critical success.

Regression Models

To investigate the determinants of both financial and critical success of films, multiple linear regression models were employed. Two continuous outcome variables were specified: *Revenue_Normalized*, which represents standardized box office revenue, and *Average_Rating*, the mean of available critical and audience ratings. These variables were chosen to capture the two distinct dimensions of success we were interested in.

Before modeling, several steps were taken to ensure the dataset was suitable for regression analysis. First, categorical variables were encoded using one-hot encoding with the first category dropped to avoid multicollinearity. Non-pertinent textual or list-based fields such as Title, IMDB_ID, Actors, and Producers were excluded, as they did not offer standardized numeric information. To account for differing scales across predictors, numerical variables were standardized using a z-score transformation via the StandardScaler function from scikit-learn. This ensured comparability of regression coefficients and improved numerical stability during model estimation. Each regression model was trained using an 80/20 train-test split. The training set was used to fit an ordinary least squares (OLS) regression model using scikit-learn's

LinearRegression, and the testing set provided out-of-sample evaluation. Model performance was assessed using several standard metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). Scatterplots comparing predicted and true values were generated for each target to visualize model fit.

Following the baseline analysis, an extended dataset was constructed that incorporated network centrality measures derived from the cast and crew collaboration networks as mentioned in the above section (Degree Centrality, Closeness Centrality, and Betweenness Centrality). Both the mean and maximum values of these measures across a movie's credited personnel were initially included. To assess potential multicollinearity introduced by these additional variables, a Variance Inflation Factor (VIF) analysis was performed. The results revealed substantial collinearity, particularly between the mean and maximum variants of centrality measures. Given this, the regression models were re-estimated using only the maximum centrality measures, which exhibited stronger correlations with the outcomes in the preliminary analysis. This stratified approach allowed for a rigorous evaluation of whether incorporating network-based attributes meaningfully improved predictive performance beyond traditional budgetary and production features.

Classification Models

Following the regression analysis, classification models were developed to evaluate whether binary definitions of film success could be more effectively predicted using the available features. As mentioned in the Data Cleaning section, two outcome variables for classification were defined: *Success_Financial* and *Success_Critical*. These outcomes allowed for the investigation of both financial and critical success as distinct but related dimensions. Pre-modeling preparation steps followed those for regression. Categorical variables were transformed into dummy variables using one-hot encoding, while continuous predictors were standardized with StandardScaler to ensure comparability across features of differing scales. Dummy variables were retained in their binary form.

Feature selection included both baseline attributes and, in a second set of models, network centrality measures (degree, closeness, and betweenness centrality). Consistent with the regression analysis, additional specifications using only the maximum values of centrality measures were later considered to address issues of multicollinearity observed in preliminary analyses. Logistic regression was chosen as the modeling approach given its interpretability and suitability for binary classification tasks. While other approaches, such as decision trees, were used, we will not discuss them in detail here. Models were trained using the training subset and evaluated on the testing subset.

Model performance was assessed using a range of standard evaluation metrics: accuracy, precision, recall, and F1-score. These metrics provide a balanced understanding of model performance across different types of classification errors, particularly given the imbalanced nature of the success categories. In addition, receiver operating characteristic (ROC) curves and the area under the ROC curve (ROC-AUC) were calculated to assess the trade-off between sensitivity and specificity across varying classification thresholds. Confusion matrices were also generated to visualize classification outcomes. This approach allowed for a direct comparison of predictive power across financial and critical success targets, as well as across model specifications with and without the inclusion of centrality-based predictors.

GNN Models

To explore how graph-based learning could enhance movie success prediction, we implemented and evaluated two types of graph neural networks: Graph Convolutional Networks (GCN) and Graph Attention Networks (GAT). GCNs aggregate information from a node's neighbors using fixed weighting based on graph structure, making them effective at capturing patterns across a network. GATs use a self-attention mechanism to learn dynamic weights for each neighbor during aggregation. This makes them particularly suitable for learning from heterogeneous graph structures, where different connections may carry different levels of importance. By using both models, we aimed to understand whether attention mechanisms provide a meaningful performance advantage over traditional graph convolution in the context of movie data.

We applied both GCN and GAT models to a movie similarity graph, where each node represents a movie and edges indicate collaborative ties between movies based on shared cast or crew members. The task was to classify movies into three buckets for average movie rating and four buckets for box office revenue. For both targets, we divided the target values into approximately equal quantiles to create class labels, allowing us to cast this as a multi-class node classification problem. Each node was described by a set of input features combining both node-level attributes and graph-based structural descriptors. The node-level attributes included numeric features such as budget, runtime, and release year, along with one-hot or multi-hot encoded categorical features for genre and age rating. We also incorporated several centrality measures (degree, closeness, betweenness, and eigenvector centrality) as structural features to help the model capture each movie's importance or role in the network topology.

To ensure a thorough evaluation, we randomly split the nodes into training, validation, and test sets using an 80% / 10% / 10% partition. Both models were trained using masked node classification, where only the training nodes were used to compute loss and gradients. Model performance was monitored on the validation set using cross-entropy loss, and final evaluation was conducted on the test set. To optimize the models, we implemented a hyperparameter tuning function. For both GCN and GAT, we looped through dictionaries of hyperparameter value combinations, varying learning rate, dropout rate, hidden layer sizes, and weight decay. Each configuration was trained and validated independently, and the model with the lowest validation loss was selected as the final configuration. This best model was then retrained and evaluated on the test set to obtain the final metrics.

To avoid overfitting, we applied dropout within the network layers and used early stopping during training. We monitored validation loss after each epoch and terminated training if no improvement was observed for 10 consecutive epochs. This combination of structured input features, graph-aware modeling, and robust training protocols allowed us to assess the effectiveness of GCNs and GATs for predicting both revenue and rating outcomes in a graph-structured movie dataset.

Results

Regression Analysis

Table 2 presents the results of the regression models for predicting both *Revenue_Normalized* and *Average_Rating*. The baseline model, which included budgetary, temporal, and production-related variables, served as the initial benchmark. The inclusion of network centrality measures was then evaluated in two stages: first, using both the mean and maximum values of each centrality metric, and subsequently restricting the model to maximum values only to reduce multicollinearity.

Table 2. Regression model performance for predicting film revenue and ratings.

Target	Model Specification	MSE	RMSE	MAE	R²
Revenue_Normalized	Baseline	0.5421	0.7363	0.4761	0.4513
Revenue_Normalized	With Centrality (All)	0.4853	0.6967	0.4639	0.5088
Revenue_Normalized	With Centrality (Max Only)	0.4811	0.6936	0.4573	0.5130
Average_Rating	Baseline	1.8077	1.3445	1.0871	0.0476
Average_Rating	With Centrality (All)	1.7052	1.3058	1.0599	0.1016
Average_Rating	With Centrality (Max Only)	1.6857	1.2983	1.0522	0.1119

The baseline model explained approximately 45% of the variance in normalized revenue, with moderate predictive accuracy as indicated by an RMSE of 0.74. The inclusion of network centrality measures improved performance, raising R^2 to 0.51 when restricted to maximum centrality values. The corresponding reduction in error metrics (MSE, RMSE, and MAE) indicates that centrality measures contribute modestly to the prediction of financial outcomes.

In contrast, the models predicting *Average_Rating* performed poorly across all specifications. The baseline model accounted for less than 5% of the variance in ratings, and even with centrality features, the R^2 increased only marginally to 0.11. This suggests that critical success, as measured through aggregated ratings, is influenced by factors not captured in the dataset, such as narrative quality, cultural relevance, or promotional campaigns.

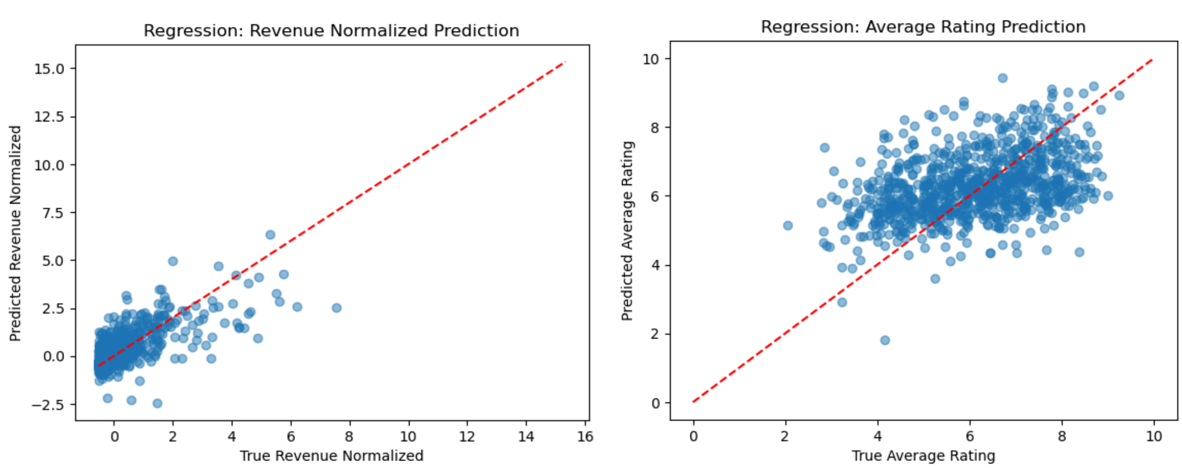

Figure 3. Scatterplot of True vs Predicted Values for Regression Models

Scatterplots of predicted versus observed values reinforced these findings. While predictions for revenue showed some alignment along the line of equality, particularly in the centrality enhanced models, the rating predictions exhibited wide dispersion, indicating limited predictive capacity. Overall, the regression results demonstrate that network attributes add some explanatory power for financial performance, but their effect on predicting critical reception is minimal. These findings motivated the subsequent shift toward classification modeling to better capture patterns of success.

Classification Analysis

Logistic regression models were estimated for the two binary outcomes: *Success_Critical* and *Success_Financial*. Model performance was evaluated using accuracy, precision, recall, F1-score, and ROC-AUC. For *Success_Critical*, the model achieved an accuracy of 0.76 with relatively high specificity but limited sensitivity, as evidenced by a recall of 0.21. This indicates that while the model correctly identified a majority of unsuccessful films, it struggled to identify critically successful films. For *Success_Financial*, accuracy was also 0.76, with much stronger recall (0.95) but at the expense of lower precision (0.77). This suggests the model was more effective in detecting financially successful films, though with a higher rate of false positives.

Table 3. Performance of baseline vs. enhanced classification models.

Target	Model Specification	Accuracy	Precision	Recall	F1 Score	ROC AUC
Success_Financial	Baseline	0.76	0.77	0.95	0.85	0.71
Success_Financial	With Centrality	0.75	0.78	0.90	0.83	0.71
Success_Critical	Baseline	0.76	0.55	0.21	0.30	0.71
Success_Critical	With Centrality	0.76	0.56	0.21	0.31	0.74

Adding centrality measures yielded modest improvements, particularly for the prediction of *Success_Critical*. As shown in Table 3, the inclusion of centrality features raised the ROC-AUC for *Success_Critical* from 0.71 to 0.74, with a small improvement in F1-score (from 0.30 to 0.31). Gains were less evident for *Success_Financial*, where accuracy slightly decreased from 0.76 to 0.75 and the F1-score declined marginally. ROC-AUC remained unchanged.

Examination of the ROC curves confirmed these findings. For *Success_Critical*, the addition of centrality attributes resulted in a visibly higher curve, consistent with the ROC-AUC increase. Confusion matrices highlighted the persistent challenge of detecting critically successful films: while the majority of unsuccessful films were classified correctly, a substantial portion of successful films continued to be misclassified. In contrast, the financial success models performed considerably better at identifying successful films, albeit with some over-prediction.

The confusion matrices in Figure 4 provide further insight into the classifiers' behavior. For critical success, the model showed strong performance in identifying non-successful films but struggled to correctly identify critically successful ones, suggesting a bias toward the majority class. In contrast, the financial success model demonstrated more balanced performance, effectively distinguishing between successful and unsuccessful films, though it tended to overpredict financial success in some cases.

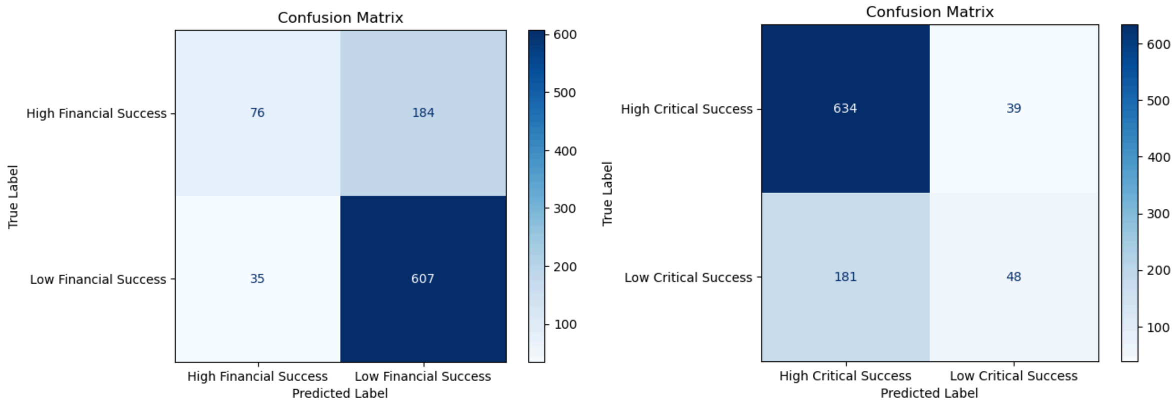

Figure 4. Confusion Matrices for Classification Models

Overall, the inclusion of centrality-based predictors modestly improved the models' discriminatory power, particularly for critical success. However, the results also underscore the difficulty of predicting critical acclaim compared to financial outcomes, even with enriched feature sets.

GNN Analysis - Ratings

To evaluate the performance of graph neural networks in classifying movies into rating buckets, we implemented both a Graph Convolutional Network (GCN) and a Graph Attention Network (GAT) using the movie similarity graph. The goal was to predict which of three rating classes each movie falls into, based on both its attributes (budget, revenue, runtime, genre, awards, etc.) and its structural position within the network (centrality measures). Average ratings were divided into three quantile-based categories. The "Low Rating" class contained movies with a rating less than 5.73, the "Medium Rating" class contained movies rated from 5.73 to 7.08, and the "High Rating" class contained movies with a rating above 7.08.

Table 4. Final model configuration for the GCN classifying average ratings.

Hidden Layer Size	Learning Rate	Weight Decay	Dropout Rate	Optimizer	Validation Loss	Test Accuracy
[128]	0.01	0.001	0.2	Adam	0.9928	46.75%

As shown in Table 4, the final GCN rating classification model was trained using a single hidden layer of 128 units, a learning rate of 0.01, weight decay of 0.001, and a dropout rate of 0.2, optimized with the Adam optimizer. The model was trained for 64 epochs and reached a validation loss of 0.9928. The GCN mode achieved a final test accuracy of 46.75%, demonstrating a moderate ability to predict movie ratings using both feature and structural information.

As illustrated in Figure 5, the model was most effective in identifying Low-rated and High-rated movies, achieving 142 and 63 correct predictions, respectively. These classes also had the highest AUC scores, at 0.67 and 0.70, indicating moderate separability from the other categories. Performance was notably weaker on Medium-rated movies, with only 31 correct predictions and an AUC of 0.56, reflecting the difficulty in capturing this middle category due to its semantic and statistical overlap with the adjacent classes. The confusion matrix highlights that most misclassifications occurred between neighboring classes, particularly between Medium and both Low and High, suggesting the model's sensitivity to boundary cases. The ROC curves further confirm this, showing the flattest curve for the Medium class.

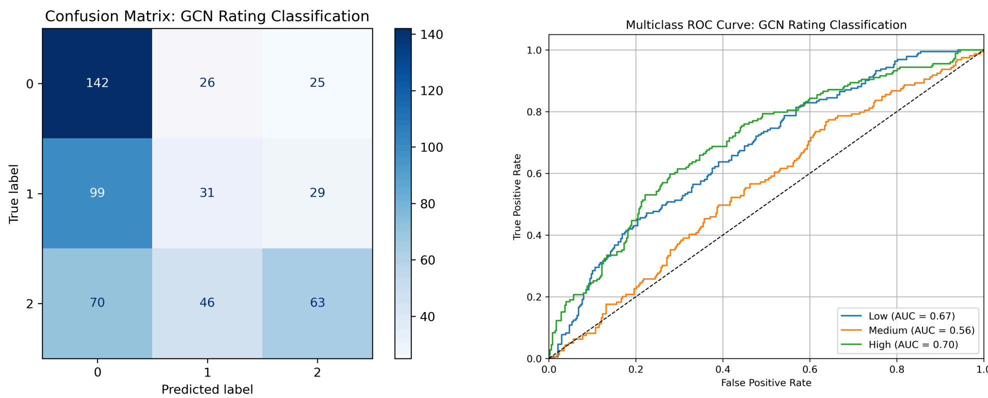

Figure 5. Confusion Matrix and ROC Curve for the GCN rating classification model

Table 5. Final model configuration for the GAT classifying average ratings.

Hidden Layer Size	Learning Rate	Weight Decay	Dropout Rate	Optimizer	Validation Loss	Test Accuracy
[128]	0.01	0.001	0.2	Adam	0.9445	53.25%

As shown in Table 4, the final GAT rating classification model was trained using a single hidden layer of 128 units, a learning rate of 0.01, weight decay of 0.001, and a dropout rate of 0.2, optimized with the Adam optimizer. The model was trained for 64 epochs and reached a validation loss of 0.9445. The GAT model achieved a final test accuracy of 53.25%, demonstrating a moderate ability to predict movie ratings using both feature and structural information.

As illustrated in Figure 6, the model was most effective in identifying Low-rated and High-rated movies, achieving 144 and 104 correct predictions, respectively. These classes also had the highest AUC scores, at 0.74 and 0.76, indicating moderate separability from the other categories. Performance was notably weaker on Medium-rated movies, with only 45 correct predictions and an AUC of 0.56, reflecting the difficulty in capturing this middle category due to its semantic and statistical overlap with the adjacent classes. The confusion matrix highlights that most misclassifications occurred between neighboring classes, particularly between Medium and both Low and High, suggesting the model's sensitivity to boundary cases. The ROC curves further confirm this, showing the flattest curve for the Medium class.

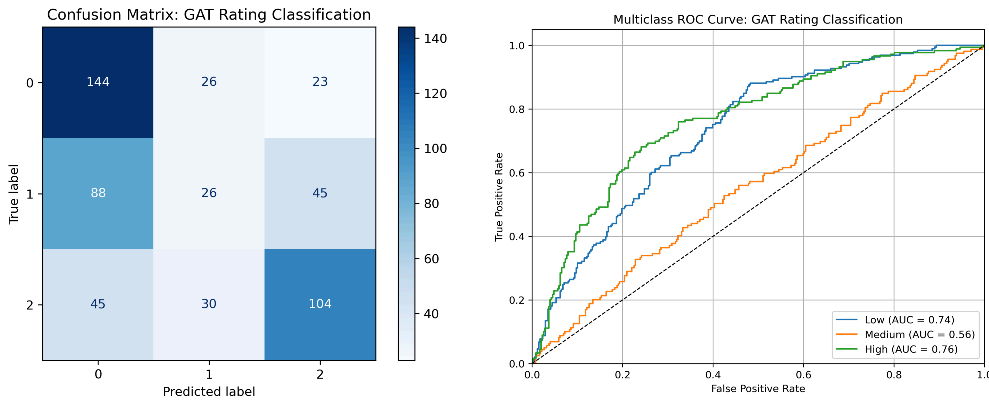

Figure 6. Confusion Matrix and ROC Curve for the GAT rating classification model

Overall, the GAT model demonstrated stronger generalization in the rating classification task compared to the GCN. By leveraging attention mechanisms to dynamically weigh the influence of neighboring nodes, the GAT achieved a higher test accuracy (53.25% vs. 46.75%) and improved AUC scores for both the Low (0.74 vs. 0.67) and High (0.76 vs. 0.70) rating classes. While both models struggled with the Medium category, the GAT still outperformed the GCN in correctly classifying Medium-rated movies. These findings underscore the advantages of attention-based graph learning in scenarios where node connectivity and attribute relevance vary across the network. The results reinforce the value of graph-based neural architectures for movie rating prediction, particularly when combining traditional metadata with rich relational and structural information.

GNN Analysis - Revenue

To evaluate the performance of graph neural networks in classifying movies into revenue buckets, we implemented and tuned both a Graph Convolutional Network (GCN) and a Graph Attention Network (GAT) using the movie similarity graph. The goal was to predict which of four revenue classes each movie falls into, based on both its attributes (budget, runtime, genre, rating, etc.) and its structural position within the network (centrality measures). Revenue values were quantile-bucketized into four approximately equally sized classes. The "Low Revenue" class contained movies with revenue below $8.2M, the "Lower-Mid Revenue" class contained movies with revenue between $8.2M and $38.1M, the "Upper-Mid Revenue" class contained movies with revenue between $38.1M and $111.9M, and the "High Revenue" class contained movies with revenue revenue above $111.9M.

Table 6. Final model configuration for GCN classifying box office revenue.

Hidden Layer Size	Learning Rate	Weight Decay	Dropout Rate	Optimizer	Validation Loss	Test Accuracy
[128]	0.01	0.001	0.1	Adam	0.8937	63.61%

As depicted in Table 6, the final model was trained using a single hidden layer of 128 units, aAs shown in Table 6, the final GCN revenue classification model was trained using a single hidden layer of 128 units, a learning rate of 0.01, weight decay of 0.001, and a dropout rate of 0.1, optimized with the Adam optimizer. The model was trained for 151 epochs and achieved a final validation loss of 0.8937. It demonstrated strong performance, accurately classifying movies into their revenue categories over 63% of the time.

The model excelled particularly in distinguishing movies in the Low and High revenue categories, achieving high true positive counts (123 and 137, respectively) and strong AUC scores of 0.90 and 0.93. Performance was moderate for the Lower-Mid and Upper-Mid revenue classes, with 84 and 86 correct predictions and corresponding AUCs of 0.79 and 0.77. Misclassifications primarily occurred between adjacent categories, indicating the model's sensitivity to borderline cases. The ROC curves further support this trend, highlighting the model's strong discriminative capability at the revenue extremes, while also suggesting room for improvement in capturing the subtler differences within mid-range revenue categories.

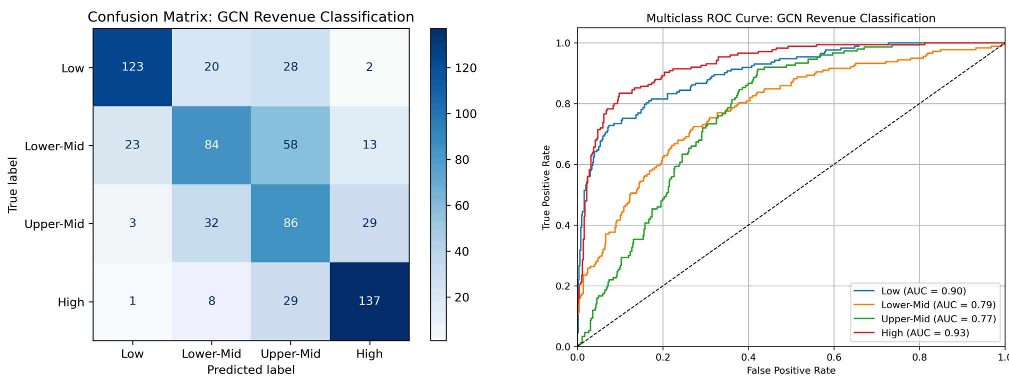

Figure 7. Confusion Matrix and ROC Curve for the GCN revenue classification model

Table 7. Final model configuration for the GAT classifying box office revenue.

Hidden Layer Size	Learning Rate	Weight Decay	Dropout Rate	Optimizer	Validation Loss	Test Accuracy
[64]	0.01	0.001	0.1	Adam	0.7249	70.41%

As shown in Table 7, the final GAT revenue classification model was trained using a single hidden layer of 64 units, a learning rate of 0.01, weight decay of 0.001, and a dropout rate of 0.1, optimized with the Adam optimizer. The model was trained for 200 epochs and achieved a validation loss of 0.7249, with a notable increase in overall test accuracy to 70.41%, a significant improvement over the GCN model. The GAT model showed excellent performance in classifying both Low and High revenue movies, with 137 and 129 correct predictions, respectively. These classes also generated the highest AUC scores, each reaching 0.96. The Lower-Mid and Upper-Mid categories were also better captured compared to the GCN results, with 109 and 101 correct predictions and improved AUCs of 0.86 and 0.85, respectively. The

ROC curves in Figure 8 reinforce this finding, showing steeper curves and higher AUCs across all categories.

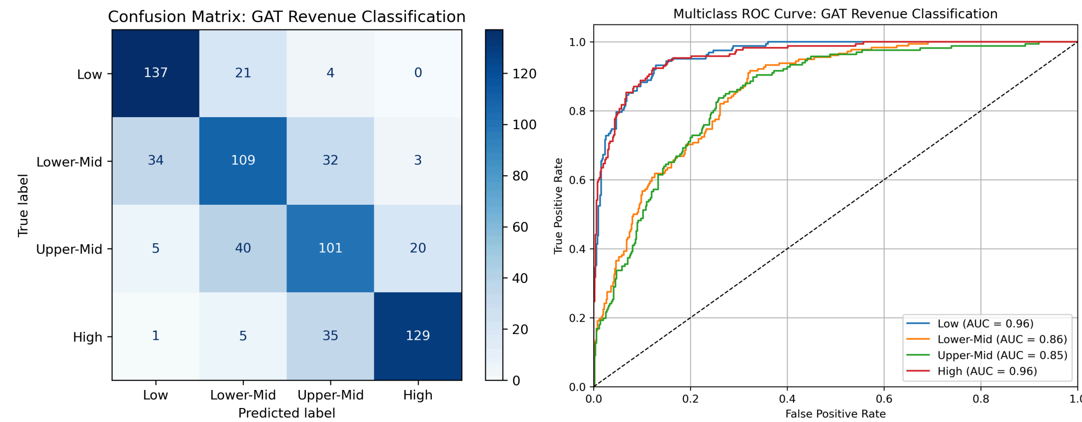

Figure 8. Confusion Matrix and ROC Curve for the GAT revenue classification model

Overall, the GAT model demonstrated superior performance in the movie revenue classification task compared to the GCN. While both models utilized the movie similarity graph and incorporated node features such as budget, runtime, genre, and centrality measures, the GAT's ability to assign dynamic attention weights to neighboring nodes resulted in more effective learning and generalization. The GAT achieved a higher test accuracy (70.41% vs. 63.61%) and outperformed the GCN across all four revenue categories, particularly in the Low and High revenue classes. Additionally, the GAT improved classification in the Lower-Mid and Upper-Mid tiers, where the GCN showed only moderate success. The ROC curves further illustrate the GAT's stronger discriminative power across revenue levels. These results highlight the advantage of attention-based graph neural networks in leveraging both feature attributes and structural context, reinforcing their potential for complex prediction tasks.

Discussion

This study sought to examine the predictors of both financial and critical success in the film industry using a combination of traditional production-related attributes and novel network centrality measures derived from cast and crew collaborations. The regression analyses demonstrated that while financial outcomes could be modeled with moderate success, predicting average critical ratings remained a challenge, even after accounting for network centrality. The relatively low R^2 values for the rating models underscore the complexity of critical reception, which may be influenced by less tangible factors such as cultural context, thematic resonance, or distribution strategies that were not captured in the present dataset.

The classification analyses offered additional insight. Financial success was modeled with higher recall, suggesting the models could effectively identify financially successful films, though sometimes at the cost of increased false positives. In contrast, critical success was more difficult to predict, with models struggling to balance sensitivity and specificity. The inclusion of network centrality measures provided modest improvements, particularly for critical success, as reflected in increased ROC-AUC values. This suggests that social and professional networks within the film industry may exert an influence, though the effect size was smaller than anticipated.

Several limitations should be acknowledged. First, the dataset was restricted to films available through the TMDB and OMDB APIs, which may exclude smaller productions or films not widely distributed. Second, budget and revenue figures may be inconsistently reported across sources, introducing measurement error. Third, the centrality measures derived from crew and cast networks did not account for qualitative aspects of collaboration, such as the historical success of partnerships or the reputation of the individuals involved. These limitations may explain why the gains from incorporating centrality were relatively modest.

Future research could address these limitations by integrating richer data sources, such as box office trends over time, marketing expenditures, or sentiment analysis from social media and reviews. Network analyses could also be refined by weighting ties based on the prior success of collaborations or by incorporating multiplex relationships (e.g., individuals working together across different roles or media formats). Moreover, experimenting with more complex machine learning models such as random forests, gradient boosting, or neural networks may reveal non-linear relationships that logistic regression and linear models are not well-suited to capture.

Conclusion

The findings of this study highlight both the promise and the challenges of predicting film success using a data-driven approach. While financial performance was modeled with reasonable accuracy, critical success proved more elusive, reflecting the multifaceted nature of artistic evaluation. Incorporating network centrality features offered some incremental improvements, particularly in distinguishing critically successful films, but the effect sizes remained limited.

Ultimately, this project underscores the value of combining traditional film attributes with innovative measures of professional networks while also illustrating both the promise of network-based prediction and the ongoing need for methodological refinement to better capture the multifaceted drivers of film success. As richer data become available and more sophisticated analytical techniques are applied, predictive modeling of film success may yield stronger and more actionable insights for both researchers and industry stakeholders.

References

1. Borsboom, D., Deserno, M. K., Rhemtulla, M., et al. (2021). Network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers, 1*, 58.
2. Chaudhary, S., Giri, S., & Gautam, B. (2025). Analyzing social networks of actors in movies and TV shows. In *International Conference on Applied Algorithms. Cham: Springer Nature Switzerland.*
3. Gao, C., Zheng, Y., Li, N., Li, Y., Qin, Y., Piao, J., Quan, Y., Chang, J., Jin, D., He, X., & Li, Y. (2023). A survey of graph neural networks for recommender systems: Challenges, methods, and directions. *ACM Transactions on Recommender Systems, 1(1)*, Article 1.
4. Mondal, P., Das, A., Mishra, S., & Debnath, N. C. (2023). Graph convolutional neural network for multimodal movie recommendation. In *Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing (SAC '23)*. ACM.

5. Nesmaoui, R., Louhichi, M., & Lazaar, M. (2023). A collaborative filtering movies recommendation system based on graph neural network. *Procedia Computer Science, 220*, 456–461.
6. Vitelli, M. (2015). Predicting box office revenue for movies. *Stanford Network Analysis Project.*
7. Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature, 393*(6684), 440–442.
8. Yang, H., Li, J., & Jahng, S. (2024). The application of knowledge graph convolutional network-based film and television interaction under artificial intelligence. *IEEE Access.*
9. Zhang, Z., Meng, Y., & Xiao, D. (2024). Prediction techniques of movie box office using neural networks and emotional mining. *Scientific Reports, 14,* 21209.
10. The Movie Database (TMDB), www.themoviedb.org/?language=en-US. Accessed 4 Aug. 2025.
11. “OMDb API.” OMDb API - The Open Movie Database, www.omdbapi.com/. Accessed 4 Aug. 2025.

Appendix

Table 1a. Auxiliary variables removed from final data.

Variable Name	Data Type	Description
Genre	list	List of genres associated with the movie
Directors	list	List of directors credited for the movie
Actors	list	List of main actors in the movie
Producers	list	List of producers involved in making the movie
Writers	list	List of individuals credited for writing the movie
Composers	list	List of people who composed the movie’s music score
Cinematographers	list	List of cinematographers responsible for the movie’s visual direction
Production_Companies	list	List of companies involved in producing the film
Language	list	Languages spoken in the movie
Metascore_Rating	float	Metascore rating from Metacritic, normalized to a 0–10 scale
IMDB_Rating	float	IMDb user rating on a 0–10 scale
Rotten_Tomatoes_Rating	float	Rotten Tomatoes critic score, normalized to a 0–10 scale
TMDB_Rating	float	TMDB user rating on a 0–10 scale
Won_Oscar	bool	Boolean indicating if the film won at least one Oscar
Oscar_Wins	int	Number of Oscars the film won
Oscar_Nominations	int	Number of Oscar nominations received
num_actors	int	Number of actors listed
Revenue	int	Total gross revenue generated by the film in USD

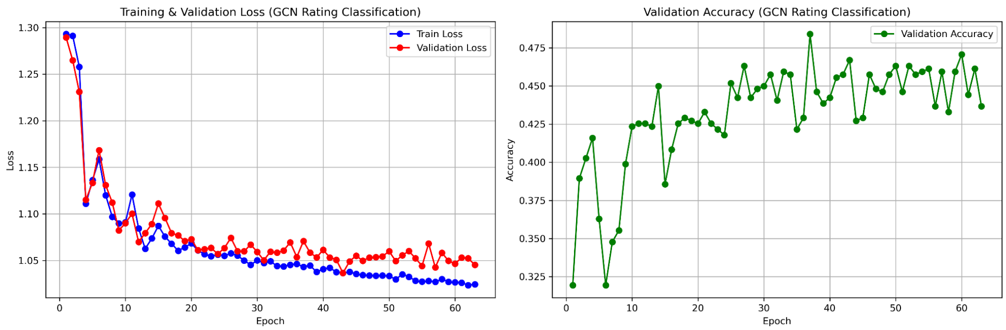

Figure 1a. Training, Validation, and Accuracy curves for the GCN rating classification model

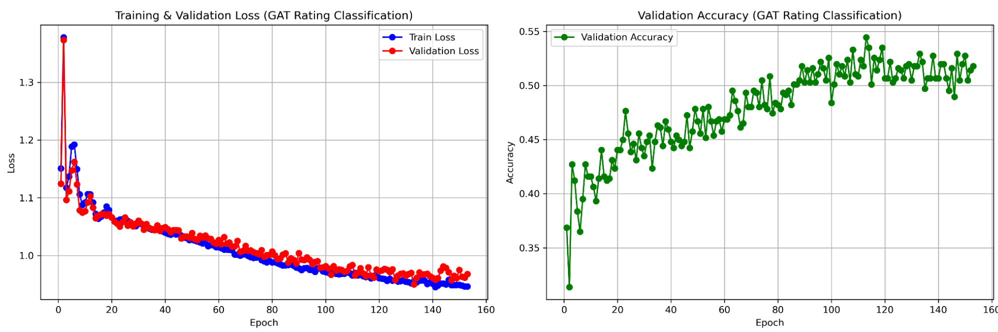

Figure 2a. Training, Validation, and Accuracy curves for the GAT rating classification model

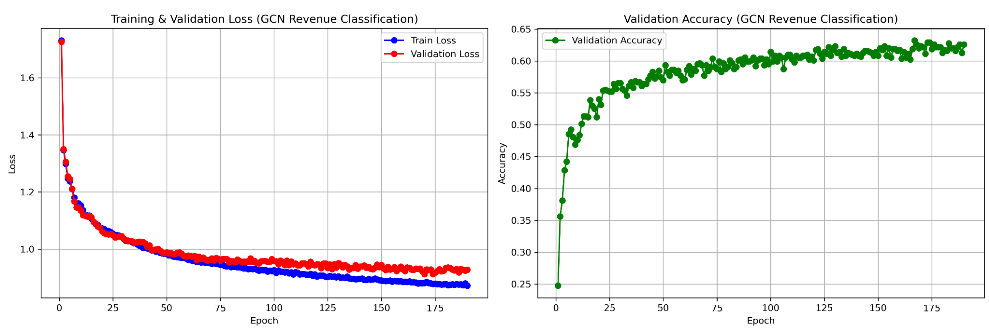

Figure 3a. Training, Validation, and Accuracy curves for the GCN revenue classification model

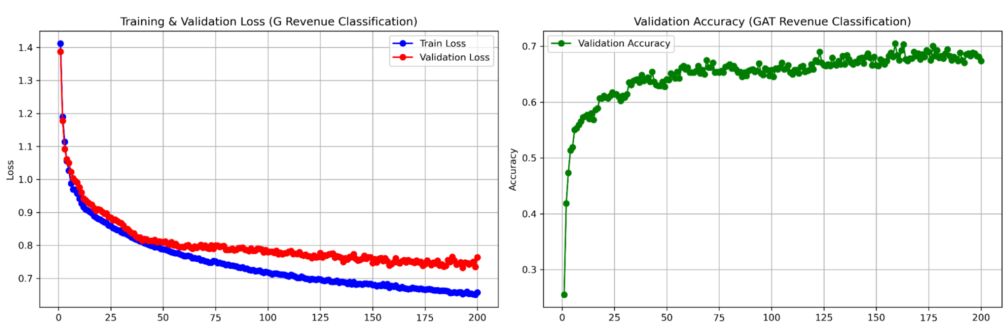

Figure 4a. Training, Validation, and Accuracy curves for the GAT revenue classification model