Movie rating prediction using heterogeneous graph neural networks

Sebouh Kafalian
University of Leeds
od20skk@leeds.ac.uk,
kafalian@adobe.com,
sebouhkafalian@gmail.com

ABSTRACT

In recent years, graph neural networks (GNNs) have gained prominence in various domains due to their ability to model complex relationships within graph-structured data. In this paper, we propose a novel approach for predicting movie ratings by leveraging heterogeneous graph. Our method combines movie attributes, and their relations to people and countries to construct a heterogeneous graph representation. We then employ few GNN architectures to forecast movie ratings. Experimental results on IMDb dataset for Netflix movies demonstrate the effectiveness of our approach. Our work aims to predict movie ratings prior to release, potentially saving investors millions of dollars in production and marketing costs.

1 INTRODUCTION

The movie industry is a multi-billion-dollar industry, generating approximately $10 billion of revenue annually. It is estimated that 80% of the industry's profits over the last decade is generated from just 6% of the films released; 78% of movies have lost money of the same time period [1]. Therefore, predicting movie ratings before their release is essential to help reduce financial risk by guiding marketing strategies and investment decisions. Additionally, by focusing on films with higher predicted ratings, studios can improve the quality of their releases and increase the percentage of profitable movies. This not only optimizes cost but also increases the proportion of successful films, contributing to more sustainable profitability in the industry.

Several studies have explored movie success prediction using various machine learning techniques and pre-release features, such as genre, budget, and director, to estimate profitability and ratings, including the work by Im and Nguyen (2011) on box office profitability prediction [2]. Similarly, Latif and Afzal (2016) employed movie attributes such as crew, release date, and number of screens to predict movie popularity using classifiers like logistic regression, achieving high accuracy [3].

Previous research on movie rating prediction often relied on traditional methods, such as collaborative filtering or simple machine learning models, which lacked the ability to capture the intricate relationships between movies, actors, genres, and other important features. These methods were limited by their inability to model the underlying graph structure of data, resulting in a loss of crucial relational information [4]. In contrast, Graph Neural Networks (GNNs) are explicitly designed to handle such graph-structured data [5].

Graph Neural Networks has been proven to work great for recommendation systems [6, 7, 8] and has been applied on movies dataset for recommendation systems [9, 10] but not for pre-release movie rating forecasting. In this paper we propose a solution by representing IMDb movies dataset as heterogeneous graph structure and training a model based on heterogeneous graph neural networks (HGNNs) [11]. HGNNs allow us to capture complex relationships and interactions among different entities in the dataset, leading to improved performance in movie rating prediction tasks.

2 DATA

IMDb, established in 1990 and acquired by Amazon in 1998, is a leading platform for movie, TV series, actor, and director information. It allows users to rate content on a scale from 0 to 10 and leave comments, providing valuable insights for others. Many users rely on these ratings and reviews to discover new films and shows, making IMDb an essential tool for making informed viewing choices.

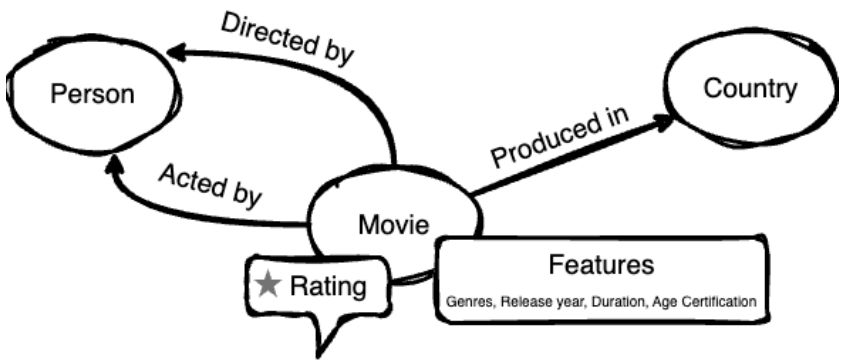

*Figure 1. A graph representation of Netflix's movie database. The graph has relationships between movies and people via **Produced in** and **Acted by** edges, as well as **Directed by** edges between movies and countries. Movies include attributes such as **genre**, **release year**, and **duration**, with their **rating** used as labels.*

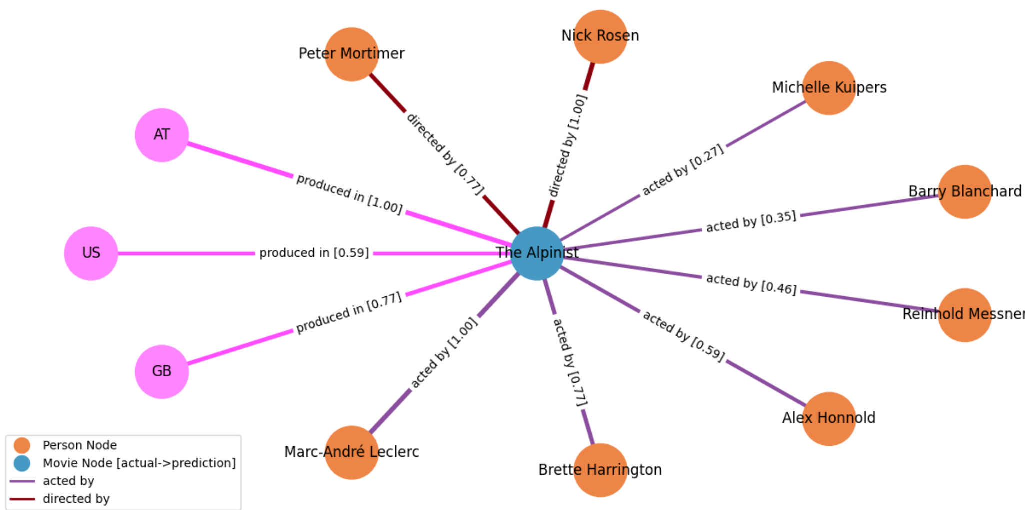

Figure 2. Graph representation of an example movie: 'The Alpinist'. The graph has few acted_by relations to actors and a single directed_by relation. The movie also has 3 produced_by relations to different countries.

In this research we use IMDb dataset for only Netflix movies. Netflix is a global streaming platform that provides a vast selection of TV shows and movies to its subscribed users, offering instant access to content with just a click. The "Netflix TV shows and Movies" dataset available on Kaggle consists of movies with its attributes like the title, genres, release date, runtime and of course the IMDb rating.
The dataset contains approximately 39K nodes, including over 2100 popular movies.

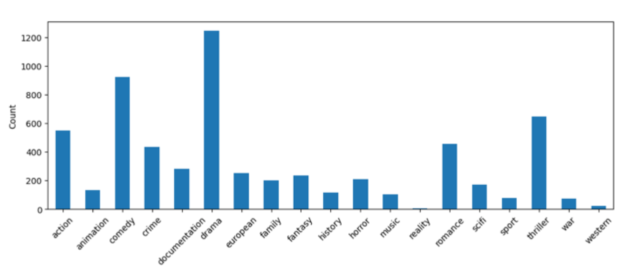

Figure 3. Count of Movies by Genre

Those movies are connected to their associated actors, directors, and production countries. In total, there are around 50 K edge between these nodes.

As shown in Figure 1, our dataset contains three types of nodes and three types of bi-directional edges. Each movie has attributes such as genre, release date, age certification, and runtime, while country and person nodes lack any features. Movies are connected to people via two relations—actor and director—with edge weights determined by their order in the cast. As seen in Figure 3, the decay coefficient ensures that main actors, directors or production countries (appearing earlier) receive lower weights, preventing weight saturation and allowing the model to learn effectively from both leading and supporting roles or countries.

3 RELATED WORKS

Marović et al. (2011) used movie and user data from IMDb to compare various methods for movie rating prediction. They examined content-based, collaborative, and hybrid approaches, implementing models such as regression trees, neural networks, k-Nearest Neighbor (k-NN), personality diagnosis, SVD-kNN. Their study aimed to identify the optimal approach for predicting user ratings on a given movie [12]. Other studies have similarly utilized user data to predict movie ratings, focusing on delivering a personalized user experience by employing collaborative filtering techniques to achieve this. [13, 14]. In another study, R. Parimi and D. Caragea (2013) [15] applied machine learning algorithms to predict the box-office success of movies before their release. They introduced a graph-based network to model the dependency relationships between movies, considering factors like shared actors, directors, and genres. By utilizing these connections to extract features for their classification model, they were able to enhance prediction accuracy. This approach outperforms traditional methods that treat each movie as an independent entity.

On the other hand, in the field of representation learning for heterogeneous networks, Dong et al. (2017) introduced the metapath2vec model [16], which significantly advances scalable representation learning. The model utilizes meta-path-based random walks to capture the structural and semantic relationships in heterogeneous networks. By applying a skip-gram model to these walks, metapath2vec effectively learns low-dimensional embeddings for nodes. The experimental results demonstrate that metapath2vec outperforms existing methods in various tasks, such as node classification and clustering, highlighting its efficacy in capturing rich information from heterogeneous graphs.

the Heterogeneous Graph Attention Network (HAN) introduced by Wang et al. (2019) [11] presents a significant advancement. HAN leverages hierarchical attention mechanisms to effectively manage the complexity of heterogeneous graphs, which consist of diverse node and link types. The model employs node-level attention to assess the importance of a node's neighbors based on meta-paths, and semantic-level attention to evaluate the significance of different meta-paths. This dual attention mechanism enables the generation of node embeddings that encapsulate rich semantic information.

Hamilton et al. (2017) introduced GraphSAGE [17], a novel inductive framework for generating node embeddings in large graphs. Unlike traditional transductive methods, GraphSAGE leverages node feature information to generate embeddings for previously unseen nodes by sampling and aggregating features from a node's local neighborhood. This approach allows for efficient and scalable representation learning, which is particularly useful for evolving graphs and applications requiring generalization to new data.

4 APPLYING GRAPH NEURAL NETWORK

Due to the relatively small size of our graph, we followed the splitting approach outlined in the paper "FastGCN: Fast Learning with Graph Convolutional Networks via Importance Sampling" by J. Chen et al. (2018) [18]. The dataset was divided into training, validation, and testing sets, with 20% allocated for testing and 10% for validation.

To test different graph architectures, we developed a two-layer model framework that is implemented by all models. The first layer is dynamically initialized, followed by a hidden layer with four dimensions and an output layer of one dimension. We apply a Leaky ReLU activation function between these layers. Utilizing this framework, we created five models: GCN, SAGE, GAT, FiLM, and GraphTransformer [19, 17, 20, 21, 22].

For benchmarking, we created a dummy model that simply returns the mean rating as its prediction, which in our case is rating of 6.4.

Model	MSE	MAE	R^2 score
Benchmark	0.01987	0.11352	0
GraphGCN	0.01608	0.09444	0.19038
GraphSAGE	0.01628	0.09556	0.17994
GraphGAT	**0.01560**	0.09262	**0.21430**
GraphFiLM	0.01732	0.09518	0.12757
GraphTrsfmr	0.01607	**0.09178**	0.19089

Table 1. Comparison of the Mean Square Error, Mean Absolute Error and R^2 score *between our models and the benchmark model.*

While all our models outperformed the benchmark (see Table 1), the improvements were smaller than we had anticipated, indicating that there is still work to be done. As Table 1 shows, the best-performing model is based on the GAT architecture; therefore, we will proceed to evaluate that model next.

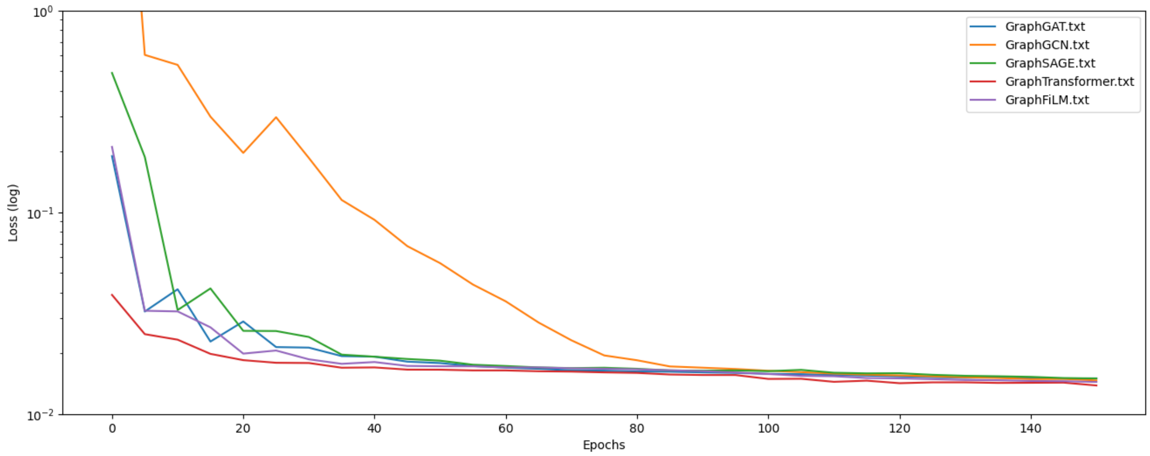

Figure 4. Training loss over epochs (log scale).

5 EVALUATION

As depicted in Figure 5 the actual ratings (green) exhibit a wider spread, with greater variance, whereas our predictions (blue) are concentrated around the center. This indicates that while the model predicts the central tendency reasonably well, it has difficulty capturing the full range of the ratings, especially for movies with very low scores.

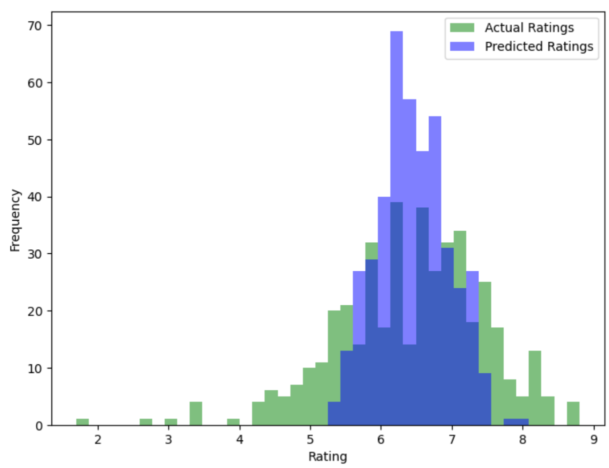

Figure 5: The distributions of actual (green) and predicted (blue) movie ratings

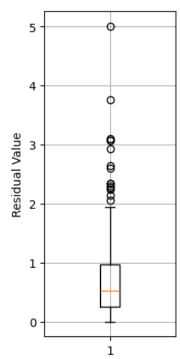

Figure 6. Error box for residuals.

In figure 6, we observe that most residuals are closely clustered around 0.5, suggesting that the model's predictions are generally accurate for most movies. However, the presence of several outliers reaching up to 5 indicates instances where the model significantly underperforms. This reinforces the idea that while the model performs well for the central distribution of ratings, it struggles with extreme cases.

6 CASE STUDIES

First let's look at the most inaccurate predictions made by our model (see tabel 3).

Movie Title	IMDB Rating	Predicted rating	residual
Himmatwala	1.7	6.707263	5.007263
Indoo Ki Jawani	3.0	6.765801	3.765801
365 Days	3.3	6.407984	3.107984
Main Aurr Mrs Khanna	3.4	6.475968	3.075968
Cuties	3.4	6.469944	3.069944

Table 3. Inaccurate predicted movies.

As anticipated, from table 3 we observe that top 5 inaccurately predicted movies have ratings below 5, confirming that our model struggles to perform well on lower-rated movies.
The movie with the worst predicted rating was "Himmatwala".

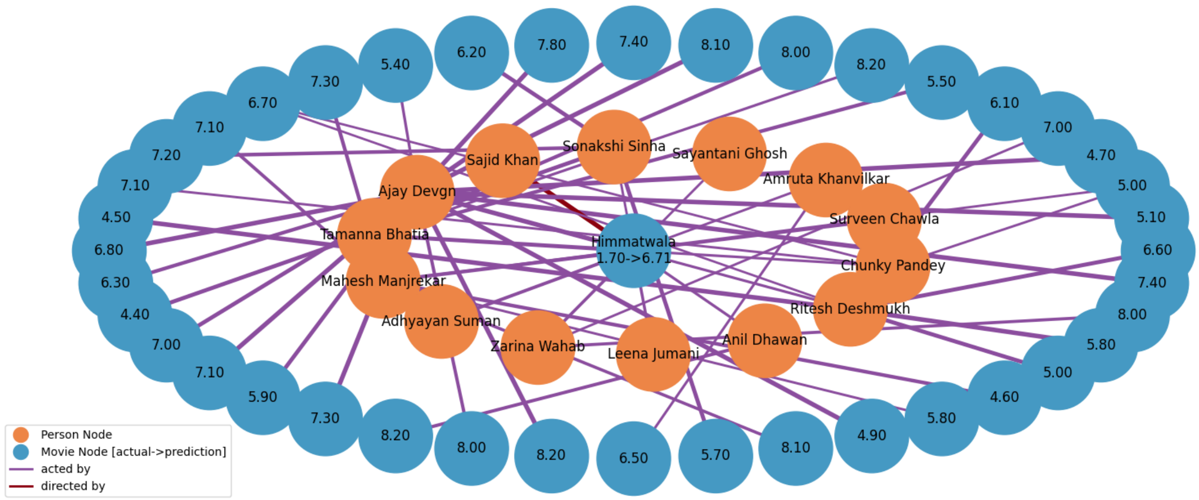

Figure 8 The inaccurate prediction: The relations of movie "Himmatwala" without country node and radius of 2.

As we can see in Figure 8, almost all actors involved in that film had previously acted in movies rated higher than 4.5. Therefore, the model was correct in predicting that this movie would be a successful movie as well, since most of casted actors are somewhat successful. This indicates that we are missing additional information about movies that influences their success, such as the planned budget, screenplay quality, or other external factors.

Now let's look at the most accurate predicitons of our model.

Movie Title	IMDB rating	Predicted rating	Residual
David Cross: Making America Great Again	6.5	6.500573	0.000573
Burning Sands	6.0	5.997471	0.002529
Phir Hera Pheri	7.1	7.105605	0.005605
The Replacements	6.6	6.591538	0.008462
All Together Now	6.5	6.489522	0.010478

Table 2. Most Accurate predicted movies.

Again we can see the same pattern here, all correctly predicted movies are around 6.4 mean rating. Let's see what relations The movie "David Cross: Making America Great Again" have.

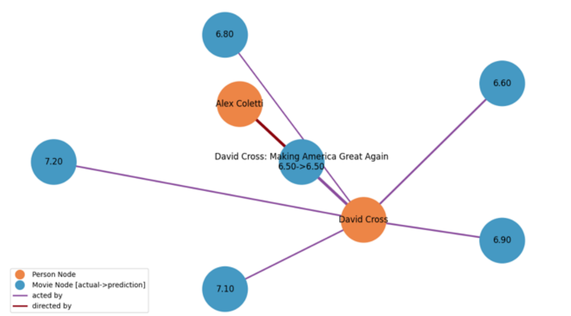

Figure 7. Most accurate prediction: The movie "David Cross: Making America Great Again".

In Figure 7, we can observe all the actors and the director connected to the movie, with the country node excluded to avoid cluttering the graph due to its large number of connections. The graph shows that the movie had one director, who did not participate in any other films, and one actor involved in five different movies.
Next lets see all of the movies that Tom Cruise and Tom Hanks participated in.

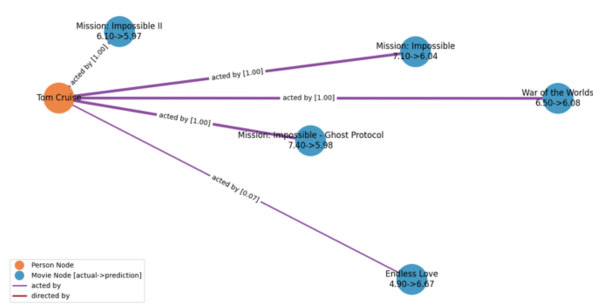

Figure 9. Tom Cruise Movies

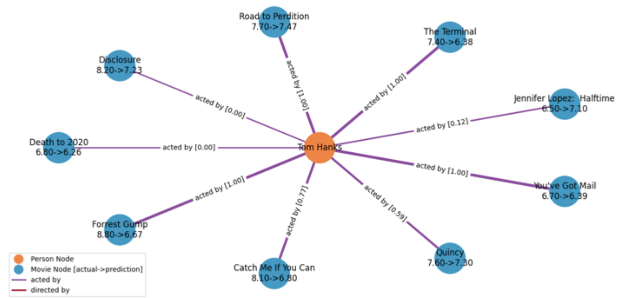

Figure 10. Tom Hanks Movies

Immediately, we can see that the model roughly predicted most of Tom Hanks' movies but performed less well for Tom Cruise's films. Notably, "Mission Impossible: Ghost Protocol" received a much lower predicted rating of 5.98 compared to its actual rating of 7.4. We will leave further investigation of this discrepancy for future work.

7 FUTURE WORK

Through our movie-watching experiences, we've all noticed that some actors excel in certain genres but may perform poorly in others. For example, an actor who shines in action roles might not deliver the same level of performance in dramatic or comedic settings. These genre-specific strengths and weaknesses are often reflected in audience ratings and critical reviews. Therefore, we suggest incorporating movie genres as edge attributes to enhance the model's ability to capture actors' performance across different genres.

Another potential improvement is to expand the graph by adding edges between movies for their sequels and prequels. This will allow the model to capture relationships between related films, which may influence a movie's rating based on the success or failure of its predecessors or successors.

We can further improve the graph by incorporating connections between movies, actors, and their awards or nominations. Adding this data would provide the model with valuable insights into how critical acclaim and industry recognition affect a movie's success.

Finally, we could enrich the movie features by adding embeddings of the screenplay and financial information. Incorporating screenplay embeddings would allow the model to capture the narrative structure, genre conventions, and thematic elements, potentially revealing patterns related to a movie's success or critical reception. Adding financial data, such as production budgets, marketing expenses, and box office earnings, could help the model understand the economic factors influencing a film's performance, providing a more comprehensive picture of the variables impacting movie ratings.

8 ACKNOWLEDGMENT

We would like to thank Professor Abdulrahman his for guidance and support.

9 REFERENCES

[1] Box Office Pro, 2019. MPA 2019 global box office and home entertainment surpasses $100 billion. Available at: https://www.boxofficepro.com/mpa-2019-global-box-office-and-home-entertainment-surpasses-100-billion/

[2] Im Darin, Minh Thao Nguyen, 2011. Predicting Box-Office Success of Movies in the U.S. Market. Stanford University. Available at: https://cs229.stanford.edu/proj2011/ImNguyen-PredictingBoxOfficeSuccess.pdf.

[3] Yew Jin Lim, Yee Whye Teh, 2007. Variational Bayesian Approach to Movie Rating Prediction. Available at: https://www.stats.ox.ac.uk/~teh/research/bayesml/kddcup2007.pdf.

[4] Keyulu Xu, Weihua Hu, Jure Leskovec, Stefanie Jegelka, 2019. How Powerful Are Graph Neural Networks?. In International Conference on Learning Representations (ICLR). Available at: https://arxiv.org/abs/1810.00826.

[5] Scarselli Franco, Gori Marco, Tsoi Ah Chung, Hagenbuchner Markus, Monfardini Gabriele, 2009. The Graph Neural Network Model. *IEEE* Transactions on Neural Networks. Available at: https://ieeexplore.ieee.org/document/4700287.

[6] Kaige Yang, Laura Toni, 2018. Graph-Based Recommendation System. IEEE Global Conference on Signal and Information Processing. Available at: https://ieeexplore.ieee.org/document/8646359.

[7] Tinglin Huang, Yuxiao Dong, Ming Ding, Zhen Yang, Wenzheng Feng, Xinyu Wang, Jie Tang, 2021. MixGCF: An Improved Training Method for Graph Neural Network-based Recommender Systems. Available at: https://doi.org/10.1145/3447548.3467408.

[8] Ruiping Yin, Kan Li, Guangquan Zhang, Jie Lu, 2019. A Deeper Graph Neural Network for Recommender Systems. Knowledge-Based Systems, 185, p.105020. Available at: https://doi.org/10.1016/j.knosys.2019.105020.

[9] Akhter Moaz, Vandana Bhagat, Ashaq Hussain Ganie, 2024. Framework for Movie Recommendation System using GNN and Textual Data. 2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies (TQCEBT). Available at: https://ieeexplore.ieee.org/document/10545127.

[10] CheonSol Lee, DongHee Han, Keejun Han, Mun Yi, 2022. Improving Graph-Based Movie Recommender System Using Cinematic Experience. Applied Sciences, 12(1493). Available at: https://doi.org/10.3390/app12031493.

[11] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, Philip S. Yu, 2019. Heterogeneous Graph Attention Network. *arXiv preprint arXiv:1903.07293*. Available at: https://arxiv.org/pdf/1903.07293.

[12] Mladen Marovic, Marko Mihokovic, Mladen Miksa, Sinisa Pribil, Alan Tus, 2011. Automatic movie ratings prediction using machine learning. Proceedings of the 34th International Convention MIPRO, Opatija, 2011, pp. 1640-1645. Available at: https://ieeexplore.ieee.org/document/5967324.

[13] O. Bora Fikir, İlker O. Yaz, Tansel Özyer, 2010. A Movie Rating Prediction Algorithm with Collaborative Filtering. 2010 International Conference on Advances in Social Networks Analysis and Mining. Available at: https://ieeexplore.ieee.org/document/5562751.

[14] Yew Jin Lim, Yee Whye Teh, 2007. Variational Bayesian approach to movie rating prediction. Proceedings of KDD cup and workshop, pp. 15-21, 2007. Available at: https://www.stats.ox.ac.uk/~teh/research/bayesml/kddcup2007.pdf.

[15] Rohit Parimi, Doina Caragea, 2013. Pre-release Box-Office Success Prediction for Motion Pictures. Proceedings of the 2013 Machine Learning and Data Mining Conference (MLDM), Lecture Notes in Artificial Intelligence (LNAI) 7988, Springer-Verlag Berlin Heidelberg. Available at: https://link.springer.com/chapter/10.1007/978-3-642-39712-7_44

[16] Yuxiao Dong, Nitesh V. Chawla, Ananthram Swami. 2017. Metapath2vec: Scalable Representation Learning for Heterogeneous Networks. Available at: https://doi.org/10.1145/3097983.3098036

[17] William L. Hamilton, Rex Ying, Jure Leskovec. 2017. Inductive Representation Learning on Large Graphs. In *31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA*. Available at: https://proceedings.neurips.cc/paper_files/paper/2017/file/5dd9db5e033da9c6fb5ba83c7a7ebea9-Paper.pdf

[18] Jie Chen, Tengfei Ma, Cao Xiao, 2018. FastGCN: Fast Learning with Graph Convolutional Networks via Importance Sampling. International Conference on Learning Representations (ICLR). Available at: https://arxiv.org/abs/1801.10247

[19] Christopher Morris, Martin Ritzert, Matthias Fey, William L. Hamilton, Jan Eric Lenssen, Gaurav Rattan, Martin Grohe, 2019. Weisfeiler and Leman Go Neural: Higher-order Graph Neural Networks. In Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19). Available at: https://arxiv.org/pdf/1810.02244.

[20] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, Yoshua Bengio. 2018. Graph Attention Networks. Published as a conference paper at ICLR 2018. Available at: https://arxiv.org/abs/1710.10903.

[21] Marc Brockschmidt, 2020. GNN-FiLM: Graph Neural Networks with Feature-wise Linear Modulation. *Proceedings of the 37th International Conference on Machine Learning*, Vienna, Austria. Available at: https://arxiv.org/abs/1906.12192.

[22] Yunsheng Shi, Zhengjie Huang, Shikun Feng, Hui Zhong, Wenjing Wang, Yu Sun. 2021. "Masked Label Prediction: Unified Message Passing Model for Semi-Supervised Classification." Baidu Inc., China. Available at: https://arxiv.org/abs/2009.03509.