Network Analysis and Propagation Dynamics of the Most Successful Films

Jay Sharma
jksharma@my.yorku.ca
York University
Toronto, Ontario, Canada

Harsimran Saini
saini19@my.yorku.ca
York University
Toronto, Ontario, Canada

Saiz Prasla
saiz2001@my.yorku.ca
York University
Toronto, Ontario, Canada

Zuhair Siddiqi
zuhairs@my.yorku.ca
York University
Toronto, Ontario, Canada

Keywords

films, box office, actors, network analysis, graph models, tmdb, louvain

1 ABSTRACT

The dynamics of film networks, particularly through shared actors, hold potential insights for predicting blockbuster success and optimizing film investments. The COVID-19 pandemic significantly impacted global box office revenues, underscoring the need for strategic decision-making by studios. In this study, we utilize data from TMDB, encompassing the top 9,400 highest-rated films, to construct a network where films are nodes connected by weighted undirected edges based on the number of shared actors. Employing network analysis techniques such as PageRank and community detection through the Louvain algorithm, we aim to identify promising film concepts. Our comprehensive examination considers factors like ratings, box office performance, shared actors, and production studios, uncovering communities and centralities within the network that correlate with successful films. This approach offers valuable insights for filmmakers, producers, and industry stakeholders to maximize the potential success of future film projects.

2 INTRODUCTION

In this project, our focus lies in exploring the dynamics of film networks through shared actors, aiming to uncover insights into potential blockbuster success for film studios to achieve higher returns on investment in their productions. Due to the effects of the COVID-19 lockdown, which include filming delays and theatre closures, average global annual box office revenue has declined [1]. This situation necessitates a strategic approach by studios to optimize their film investments. Using data from TMDB comprising the top 9,400 highest-rated films of all time, we will construct a network where films are nodes connected by weighted undirected edges based on the number of shared actors. Our objective is to identify the most promising film concepts by employing network analysis techniques such as PageRank and exploring film attributes across different areas using community detection.

By considering various factors including ratings, box office performance, shared actors, and production studios, we will conduct a comprehensive examination of the network of films. Through this analysis, we aim to uncover communities or centralities within the network that correlate with the most successful films. This approach will offer valuable insights for filmmakers, producers, and industry stakeholders seeking to optimize their creative decisions and maximize the potential success of future film projects.

3 Motivation / Problem

The motivation of this project is to find similarities and connections in a large set of the top films to see what factors into the most successful films of all time, in an era where theatrical releases are on a decline and Hollywood is struggling, we want to find ways to make people go back to theatres.

- Given a dataset of films, what similarities and structures can we identify to connect these films?
- What factors can we use to assess each film's performance?
- What graphical analyses and evaluations can we employ to measure success and identify films worth replicating?

4 RELATED WORK

4.1 Centrality of Networks

One relevant study to our analysis of film networks [2] is based on centrality and communities of the most successful films. This study uses network graphs of actors based on their common appearances in popular films. By retrieving data from IMDb, the project constructs a network where the nodes represent actors and the edges indicate shared film roles. The study also explores additional scenarios, such as the popularity of actors in less well-received films within the Belgian film industry.

4.2 Clustering Model

This study [3] on networks of productions and co-productions is also relevant to our analysis of the film network as it demonstrates the funding structures of different Hollywood studios. The study reveals that approximately 20% of productions generate 80

This study can help determine if shared personnel, such as actors, directors, and producers, contribute to success, with large clusters of films featuring similar actors potentially indicating success. Smaller clusters of critically acclaimed films may suggest that specific directors or actors contribute significantly to a film's artistic intent and review score.

4.3 Network Model Patterns

Another study [4] is also relevant to the analysis of the film network analysis as this study shows the investigating patterns in the film industry using the data retrieved from IMDb. The study is focused upon the network centrality and finds that rather the film ratings, high centrality is associated with active statistics. It also focuses on the sentiment analysis of film scripts, relationships between film costs and revenue and community structures in actor collaborations. Some of the important discoveries include associations between larger budgets and higher ratings, the recognition of popular genres and seasonal patterns in the box office performance of films.

5 METHODOLOGY

In order to answer the probelm and address all queries we had, we took the steps as listed:

- 1. Find and cleanse dataset, as discussed in **Section 5.1**
- 2. With the data, create the network, including weighting the edges and adding attributes to the nodes as discussed in **Section 5.2 and Section 5.3**
- 3. Perform several network analysis algorithms as discussed in **Section 5.4** and report our analysis on the data we find.

5.1 Dataset

Our first step was to find a suitable dataset and cleanse the data of any inconsistencies. Using the TMDB (The film Database) API [5] we were able to get the dataset of the top 9400 top rated films of all time, with this information, we decided to take into account various different attributes. We removed films with no recorded budget or revenue, if either was missing, we decided not to include it in our graph, since these attributes were crucial to our calculations of node weight.

For the actual dataset, due to the method in which TMDB's API functions, we got the top rated films using their "top_rated" endpoint, and we went through each page until we hit the end of the list, this gave us a dataset of around 9400 films, in this list we got each of the "film_ids" that were available so we could search for details of each film. Initially, we had manually scraped another website, "BoxOfficeMojo", but they only had lists available for the top 1000 films, and we wanted a more robust dataset. After getting the ids, we fetched each of the film's details including:

- Budget - How much the film costed to produce
- Genres - What genres are attributed to the film
- Actors - Who acted in the film
- Production Companies - What studios produced the film
- Viewer Rating - The rating of the film
- Revenue - How much the film made at the Box Office
- Series - Whether or not the film is a part of a series
- Release Year - The year the film released

This was the data that we felt was relevant to our analysis and was needed to show commonalities in our graph. Initially, we wanted to use IMDB, but they had extremely expensive API costs that weren't feasible for this project, and manually scraping would have caused issues including rate limiting and potentially having our IP addresses blocked. One thing we did with the TMDB API was used multithreading to have it so that the API calls were run concurrently after having the list of "film_ids" split into chunks, this improved the speed of the API calls and allowed for a lot more testing when running the calls.

5.2 The Mathematical Formulation of Node Weights

For each film node, we implemented a scoring strategy to measure "success," which we refer to as the "success score." Initially, the scoring was based on a ratio of revenue to budget. However, we encountered a significant issue with this method: it favored films with very small budgets. For instance, the film "The Blair Witch Project" had a budget of $60,000 and earned $250,000,000 in theaters, resulting in an abnormally high success score of over 5000, which was 10 times higher than any other score.

To address this issue, we revised our approach. We now calculate the profit, normalize its value, and incorporate a normalized user rating. Finally, we scale the combined score by multiplying it by 10 to obtain a score out of 10. Normalization is a method of scaling down a set of numbers so that they fall between a range of values, in our case the numbers fell between 0 and 1. [6]. In our code, we used the MinMaxScaler() function from the scikit-learn library, this automatically normalized our values for us using the standard formula. The steps we took to find our normalized profit and ratings were:

Algorithm 1 Calculate Success Score

Require: $revenue_i, budget_i, rating_i$
Ensure: $success$

1: $profit_i = revenue_i - budget_i$
2: $normalized_{profit} = \frac{profit_i - profit_{\min}}{profit_{\max} - profit_{\min}}$
3: $normalized_{rating} = \frac{rating_i - rating_{\min}}{rating_{\max} - rating_{\min}}$
4: $success = \left(\frac{normalized_{profit}}{2}\right) + \left(\frac{normalized_{rating}}{2}\right) \times 10$

For the score, we decided to evenly weigh both profit and rating, as we felt that both factors are important indicators of a film's success. For example, films like "Suicide Squad" and "Venom" were received rather poorly but performed well at the box office, while others like "Blade Runner 2049" received critical acclaim but performed poorly financially. With this score as a node attribute, we introduce a new factor into algorithms such as PageRank, allowing for a personalized factor to be considered.

5.3 The Mathematical Formulation of Edge Weights

For our edge weights, we experimented with several formulas until settling on one that we found most logical and appropriate. Initially, we considered only common actors, and then we tested a method involving actors, genres, and production companies, which led to a graph with too many nodes. We determined that common genres alone were insufficient to justify an edge, so we ultimately settled on the following formula:

Algorithm 2 CreateWeightedEdge(node1, node2)

Require: $node1, node2$
Ensure: weighted edge or "none"
$commonActors \leftarrow$ intersect($node1.actors, node2.actors$)
$commonStudios \leftarrow$ intersect($node1.studios, node2.studios$)
if $|commonActors| > 2$ **then**
    $weightedEdge \leftarrow (|commonActors| \times 2) + |commonStudios|$
else
    $weightedEdge \leftarrow$ "none"
end if
Return $weightedEdge$

We placed emphasis on common actors by assigning them an edge weight of 2, considering each film listed only the top 20 highest billed actors in our dataset. When viewing a film poster, audiences often look for recognizable actors as a major deciding factor. Therefore, we prioritized actors, acknowledging their significant influence in attracting viewers in the contemporary film industry. While film studios can also wield considerable influence, actors are typically perceived as a more prominent selling point compared to studios.

5.4 The Algorithms

With the network we've created, there are several network analysis algorithms we have run to analyze the network and extrapolate some information from it. The algorithms we've run so far are Personalized PageRank, Node Degree Centrality, and we plan to use the Louvain algorithm for further analysis.

Personalized PageRank For Personalized PageRank, we utilized NetworkX's `pagerank(G, personalization='success')` function, where G represents our film graph and 'success' is the attribute from each node representing its success score. PageRank, originally developed to rank web pages by Google, ranks nodes based on incoming link structure, assuming that important nodes are linked by other important nodes. Personalized PageRank extends this by incorporating a personalization vector that biases scores towards nodes with higher success scores [7].

Node Degree Centrality Node Degree Centrality was assessed using NetworkX's built-in `degree_centrality()` function. This metric measures the importance of nodes based solely on the number of connections each node has within the film graph. Unlike PageRank, which emphasizes interconnectedness, degree centrality provides an unbiased evaluation of node significance solely based on link count [8].

Louvain Algorithm Next, we employed the Louvain algorithm using NetworkX's `louvain_communities(G, weight)` function. This algorithm detects communities within our graph, taking into account edge weights to emphasize stronger connections, such as frequent collaborations between actors. Louvain operates efficiently with a time complexity of $O(n \log n)$, making it suitable for large datasets compared to alternatives like Girvan-Newman, which has a higher time complexity of $O(n^3)$ [9]. Initially assigning each node to its own community, Louvain iteratively optimizes community assignments to maximize modularity, resulting in cohesive and meaningful clusters. This approach enhances clustering efficiency and yields more balanced community sizes compared to other algorithms.

—

Initially, we experimented with Girvan-Newman, which produced a single large community followed by numerous smaller ones. In contrast, Louvain provided a more even distribution of communities with comparable sizes, aligning better with our analytical goals.

5.5 Robustness and Challenges

Some challenges that we have faced during this project so far include the need to transition to a much larger graph. Initially, we planned to use only 100 nodes, representing the top 100 highest grossing films of all time. However, this smaller dataset lacked robustness and provided insufficient information. To address this issue, we opted instead to utilize data from the top 9,400 highest rated films of all time.

Another significant challenge was the inability to access IMDb's API. IMDb is typically the primary source for film information, but their API is prohibitively expensive, costing over $150,000 [10], primarily designed for commercial usage.

6 EXPERIMENTS / EVALUATION

6.1 The Network

The nodes and edges of our final network were as follows:

Table 1: Nodes and Edges of film Graph

Nodes	Edges
5,755	13,998

We felt this network was a sizeable subset of our initial dataset which had all relevant data that we needed for further evaluation and analysis.

6.2 Evaluations

We evalulated the graph using various algorithms including PageRank, Node Degree centrality, and the Louvain Community Detection algorithm. Below are the top 10 nodes for each of the respective algorithms, as well as summaries for the 5 communities with the highest average success scores with different attributes.

Table 2: Top 10 films by PageRank

films	PageRank
Harry Potter and the Order of the Phoenix	0.0013234276868831663
Avengers: Infinity War	0.0012905504009495543
Avengers: Endgame	0.0012624523143546574
Alice in Wonderland	0.0011773257260922498
Harry Potter and the Deathly Hallows: Part 1	0.0011429170249350508
Avengers: Age of Ultron	0.001110811313672943
Harry Potter and the Deathly Hallows: Part 2	0.0011096608904956203
Harry Potter and the Half-Blood Prince	0.001031892073510828
Harry Potter and the Prisoner of Azkaban	0.0010214331826948306
The Lego film	0.0010023023070782578

Table 3: Top 10 Nodes by Degree Centrality

films	Degree Centrality
Harry Potter and the Order of the Phoenix	0.007994438651372959
The Lego film	0.007820646506777894
Sausage Party	0.0076468543621828295
Avengers: Endgame	0.0074730622175877654
Avengers: Infinity War	0.0074730622175877654
Alice in Wonderland	0.0074730622175877654
Harry Potter and the Deathly Hallows: Part 1	0.006951685783802572
This Is the End	0.006951685783802572
Avengers: Age of Ultron	0.006256517205422315
Knocked Up	0.006256517205422315

Table 4: "Lord of the Rings" Community Summary

Most Common Actors By Appearance in Community	
David Wenham	12
Hugo Weaving	11
Brad Dourif	10
Most Common Genres by Films in Community	
Drama	50
Action	38
Thriller	35
Most Common Production Companies	
Universal Pictures	14
Warner Bros. Pictures	10
Columbia Pictures	9
Additional Metrics	
Community Size	436 films
Average Success Score	3.560
Series Percentage	33%
Average PageRank	0.000185
Average Degree Centrality	0.000791
Average Release Year	2006
Top 5 Highest Scoring Films by Success Score	
The Lord of the Rings: The Return of the King	6.933
The Lord of the Rings: The Two Towers	6.528
The Lord of the Rings: The Fellowship of the Ring	6.426
The Hobbit: The Desolation of Smaug	5.533
The Hobbit: An Unexpected Journey	5.438

Table 5: "Marvel Cinematic Universe" Community Summary

Most Common Actors By Appearance in Community	
Actor	Count
Samuel L. Jackson	16
Anthony Mackie	14
Idris Elba	13
Most Common Genres by Films in Community	
Action	70
Drama	66
Adventure	61
Most Common Production Companies	
Marvel Studios	27
Columbia Pictures	13
Warner Bros. Pictures	13
Additional Metrics	
Community Size	171 films
Average Success Score	3.762
Series Percentage	34%
Average PageRank	0.000239
Average Degree Centrality	0.001351
Average Release Year	2009
Top 5 Highest Scoring Films by Success Score	
Avengers: Endgame	9.165
Avengers: Infinity War	7.917
Spider-Man: No Way Home	7.680
Barbie	6.123
Spider-Man: Far From Home	5.855

Table 6: "Harry Potter" Community Summary

Most Common Actors by Appearance in Community	
Johnny Depp	25
Colin Firth	23
Jim Broadbent	22
Most Common Genres by Films in Community	
Drama	246
Action	122
Thriller	114
Most Common Production Companies	
Working Title Films	42
Warner Bros. Pictures	39
Universal Pictures	31
Additional Metrics	
Community Size	436 films
Average Success Score	3.751
Series Percentage	20%
Average PageRank	0.000222
Average Degree Centrality	0.00122
Average Release Year	2004
Top 5 Highest Scoring Films by Success Score	
Harry Potter and the Deathly Hallows: Part 2	6.895
Bohemian Rhapsody	6.186
Harry Potter and the Philosopher's Stone	6.092
Harry Potter and the Prisoner of Azkaban	5.856
Harry Potter and the Goblet of Fire	5.819

Table 7: Nolan/Modern Classics Community Summary

Most Common Actors by Apperance in Community	
Morgan Freeman	14
Penelepe Cruz	12
Michael Caine	11
Most Common Genres by Films in Community	
Drama	96
Thriller	82
Action	53
Most Common Production Companies	
Warner Bros. Pictures	18
Universal Pictures	13
TV 2	8
Additional Metrics	
Community Size	173 films
Average Success Score	3.734
Series Percentage	20%
Average PageRank	0.0001820
Average Degree Centrality	0.0007846
Average Release Year	2011
Top 5 Highest Scoring Films by Success Score	
The Dark Knight	6.5916
Oppenheimer	6.2587
Inception	6.1886
The Dark Knight Rises	5.9345
The Batman	5.4153

Table 8: Classic Sci-Fi/Inspirational Community Summary

Most Common Actors by Apperance in Community	
Clint Eastwood	21
Mel Gibson	12
Arnold Schwarzenegger	10
Most Common Genres by Films in Community	
Drama	55
Action	55
Thriller	49
Most Common Production Companies	
Warner Bros. Pictures	33
20th Century Fox	18
Malpaso Productions	17
Additional Metrics	
Community Size	149 films
Average Success Score	3.7102
Series Percentage	34%
Average PageRank	0.000208
Average Degree Centrality	0.000844
Average Release Year	1980
Top 5 Highest Scoring Films by Success Score	
Back to the Future	5.6161
Terminator 2: Judgment Day	5.5133
12 Angry Men	5.2035
The Exorcist	5.1946
The Good, the Bad and the Ugly	5.1893

6.3 Analysis and Visualization

The top 10 films identified by PageRank mostly consist of some of the most successful films of all time. Despite some variations, notable similarities between PageRank and node degree centrality are apparent, particularly with three different Avengers films appearing in both metrics. These films serve as central nodes within the graph, exerting influence across the entire network and within their immediate neighborhoods.

Distinct characteristics and success factors can be observed across different communities within our analysis. For instance, communities like "Lord of the Rings" and "Harry Potter" are characterized by their large size and interconnectedness, encompassing genres such as drama and action. The "Marvel Cinematic Universe" stands out for its unity and influence in the action-adventure category, supported by recurring performers and thematic consistency.

In contrast, smaller communities like "Nolan/Modern Classics" and "Classic Sci-Fi/Inspirational" also exhibit popular films in genres like drama, thriller, and action. These communities demonstrate a blend of economic and critical success, underpinned by strong writing and star power. Notably, the "Nolan" community is distinctive for prominently featuring films associated with a single director, reflecting the significant influence of Christopher Nolan's directorial style and storytelling prowess.

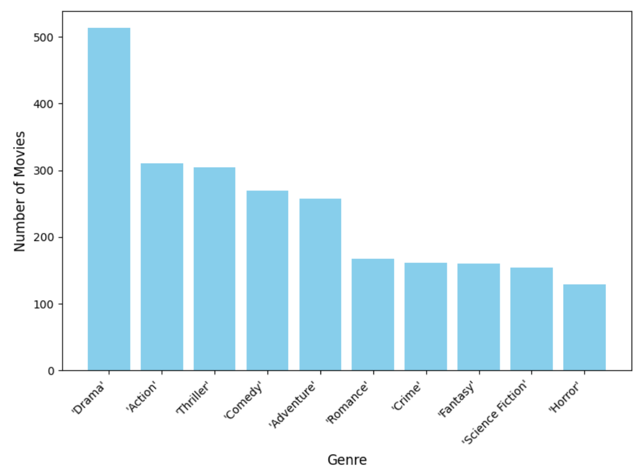

Figure 1: Top 10 genres in the top 5 communities by number of films

We used the Louvain algorithm to conduct community detection analysis to find common genres in communities, checking if the most successful films are in larger series of films or not, as well as other factors that are statistically significant in the network, such as the most common actors. As **Figures 1** and **2** depict data from the community summaries highlighting that the top five have similar attributes in the form of recurring actors and same genres. Specifically on the same genre point, it is important to note how the analyses using the Louvain algorithm enabled us to see the disparity of Drama as an outlier relative to all other genres, due to most films being classified as 'Drama' due to it becoming an all encompassing genre for most films.

Looking at **Figure 2**, we observe a notable presence of prominent actors within the top 10. Iconic figures such as Johnny Depp, Samuel L. Jackson, and Clint Eastwood feature prominently, underscoring their substantial influence and widespread appeal across

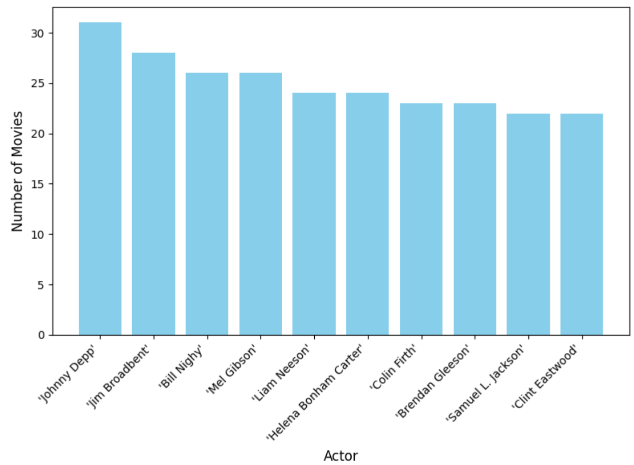

Figure 2: Top 10 actors in the top 5 communities by number of films

multiple successful films. These actors are not only recognized for their versatile performances but also for their ability to attract audiences and contribute significantly to the success of the films they star in.

Johnny Depp, known for his distinctive roles in franchises like "Pirates of the Caribbean," exemplifies how star power can elevate a film's box office performance and cultural impact. Similarly, Samuel L. Jackson's prolific career spans across various genres, cementing his status as one of the most bankable actors in Hollywood history. Clint Eastwood's dual roles as both actor and director have shaped numerous critically acclaimed and commercially successful films, showcasing his enduring influence in the industry.

Beyond these individual actors, Figure 2 highlights the strategic use of star casting as a contributing factor to the success of top-ranking films. The presence of these actors in multiple successful films underscores their ability to draw audiences and enhance the film's marketability and appeal.

Furthermore, by analyzing metrics of films in a given cluster, with evaluations on a film's success score, series percentage, and average release year we were able to statistically determine which film was the most optimal in return on investment for the producers, given its respective budget size. Communities which are franchise-centric such as the Harry Potter or Marvel Cinematic Universe community and communities which are genre-specific such as Modern Classics or Science-Fiction (Sci-Fi) were approximately similar in actor count. This similarity in actor counts across different community types facilitated a comparative analysis of the networks, revealing the impact of casting choices on a film's success.

Based on the evaluation metrics, we noticed that the average PageRank for the Marvel Cinematic Universe was the highest, followed by Harry Potter, then Classic Sci-Fi, then Nolan/Modern Classics, and finally in last place was Lord of the Rings. We also observed in our analysis that degree centrality was highest in Nolan/Modern Classics, followed by Marvel Cinematic Universe, then by Classic Sci-Fi, then Lord of the Rings, and finally Harry Potter had the lowest average degree centrality. These findings from our analysis are later translated into insights in our conclusion section.

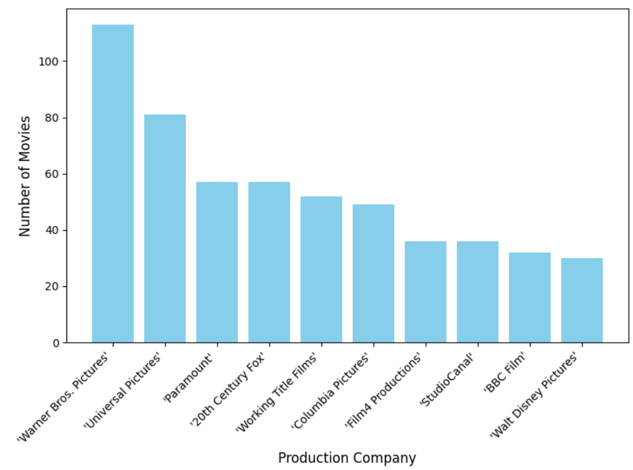

Figure 3: Top 10 production companies in the top 5 communities by number of films

Looking at the production companies listed, Warner Bros., Universal, and Paramount seem to be some of the most prolific and common, Walt Disney Pictures has a much lower number most likely due to the communities that featured the Disney Animated films not being in the top 5 by success score, animation in the modern age is not as well received as live action blockbusters, both by user ratings and by box office. Warner Bros. owning Harry Potter, Lord of the Rings and the film rights to Batman also helped within these communities since these were some of the prominent films in the list, some notable films that were missing from these communities include the Star Wars and Avatar films, most likely due to the fact that these film's actors aren't as commonly seen in films as actors from the films in the communities we had found.

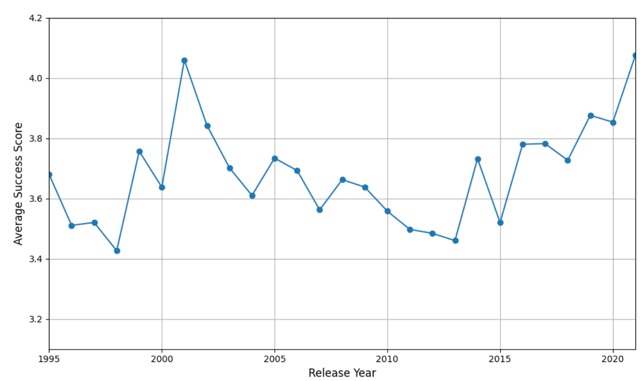

Figure 4: Average success score for films in the top 5 communities over time (1995-2022)

Lastly, by evaluating the average release year within clusters a temporal bias was detected as newer films dominated the community summary charts, specifically the average success score. However, to possibly account for this bias, a potential future direction would be to account for inflation of currency and how that could have contribute dot this temporal bias. Overall, the comprehensive analysis provided insights into optimal investment strategies and aiding in data-driven production decisions, enhancing our understanding of the cinematic landscape.

7 CONCLUSIONS

To conclude, we've constructed our network of films, with each film being a film and each edge being weighted with common actors. We developed a scoring system for the success of each film that uses a normalized profit and rating to attribute each node with a success score, we used different algorithms such as PageRank, Node degree centrality, and the Louvain community detection algorithm on our graph to see some commonalities and correlations with our various scoring mechanisms and common films. The small budget that Lord of the Rings, which was produced on a relatively small budget yet achieved a large box office performance, highlighting the potential for high profitability even with limited initial funding if there is a large commonality and continuity of actors amongst films.

Possible future works could be based on the fact there is the aforementioned temporal bias, with newer films scoring better on average in most metrics relative to older films, with the notable exception of the Lord of the Rings series. This bias indicates a trend where contemporary films generally outperform their older counterparts, which the visualization in **Figure 4** clearly depicts average success scores increasing over time. This discrepancy could be attributed to changes in currency inflation or advancements in film technology, but to fundamentally address this bias, one potential future step is to incorporate currency inflation calculations. Adjusting financial metrics for inflation would provide a more accurate comparison across different time periods. This refinement of the algorithm could enhance the algorithm's ability to fairly evaluate cinematographic success in terms of the budget-to-box office revenue ratio. Sorting this inconvenience out would help see a pattern as to what continues to be successful films across decades regardless of the vast differences in film-making techniques, technological advancements, and the growth of film culture.

References

[1] P. Frater, 2024. CinemaCon: Worldwide Box Office Headed for Marginal Decline in 2024. *Variety Magazine*. Retrieved from https://variety.com/2024/film/news/cinema-box-office-decline-2024/amp/.

[2] F. Zeshan. IMDb Project. Retrieved from https://zesh0012.github.io/IMDB-project/.

[3] K. Chen, 2021. Network graphs of actors based on popular films in common. *Medium*. Retrieved from https://medium.com/web-mining-is688-spring-2021/network-graphs-of-actors-based-on-popular-films-in-common-69d30e7b5e07.

[4] J. Miller, 2008. Structure and evolution of online social networks. *International Journal of Communication*, 2, 831-891. Retrieved from https://ijoc.org/index.php/ijoc/article/viewFile/896/580.

[5] The film Database (TMDB). TMDB API. Accessed June 1, 2024. https://www.thefilmdb.org/documentation/api.

[6] WallStreetMojo. Normalization Formula - What Is It, How To Calculate. WallStreetMojo. Accessed June 1, 2024. https://www.wallstreetmojo.com/normalization-formula/.

[7] L. Page, S. Brin, R. Motwani, and T. Winograd, 1999. The PageRank Citation Ranking: Bringing Order to the Web. Technical Report, Stanford InfoLab.

[8] NetworkX Developers. 2024. networkx.algorithms.centrality.degree_centrality. NetworkX Documentation. Retrieved from https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.centrality.degree_centrality.html.

[9] NetworkX Developers. 2024. networkx.algorithms.community.louvain.louvain_communities. NetworkX Documentation. Retrieved from https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.community.louvain.louvain_communities.html.

[10] IMDb and Box Office Mojo for films/TV/OTT (API). Amazon Web Services. Retrieved from https://aws.amazon.com/marketplace/pp/prodview-yeuyizioqmfsy?sr=0-2\&ref_=beagle&applicationId=AWSMPContessa#offers