

Analyzing the Impact of Nepotism on the Success of Bollywood Actors

Background

The Bollywood film industry, which is the primary film industry in India featuring Hindi-language films, is notorious for its history of nepotism. Certain influential families, such as the Kapoors, Bachchans, and Khans, have almost all of their family members working in the industry, as if it is their birthright. In recent years, actors and actresses such as Ananya Panday and Sara Ali Khan have been facing backlash from the media regarding the role of nepotism in their blossoming careers. Therefore, our group wanted to investigate the role of nepotism and the extent to which it affects the careers of Bollywood actors and actresses.

Experimental Design and Procedure

We hypothesize that nepotism, although a guarantee of opportunities, is not a guarantee of success in Bollywood.

In order to test this hypothesis, we first created a web scraping software which was able to gather data from various Wikipedia web pages regarding the familial relationships of various Bollywood actors (i.e. parents, children, spouses, etc.). Using this data, we created a graph with the adjacency list structure such that each node represented a unique actor, and the adjacency lists for each node listed relatives of the actor. Then, we implemented the BFS algorithm to traverse the graph and determine the total number of connected components in our graph, as well as the size of the connected component that each actor was a part of. Using these metrics, we were able to calculate the “nepotism ratio” in the Bollywood industry, which we defined as the ratio between the number of connected components in our graph and the total number of nodes.

After collecting data from our BFS traversal and the Wikipedia pages, we moved on to scrape the Rotten Tomatoes website to get the average box office revenue, average movie rating (based on audience scores), and number of movies for each actor. Using this data, we created graphs which compared the size of each actor’s connected component (which represents the size of their

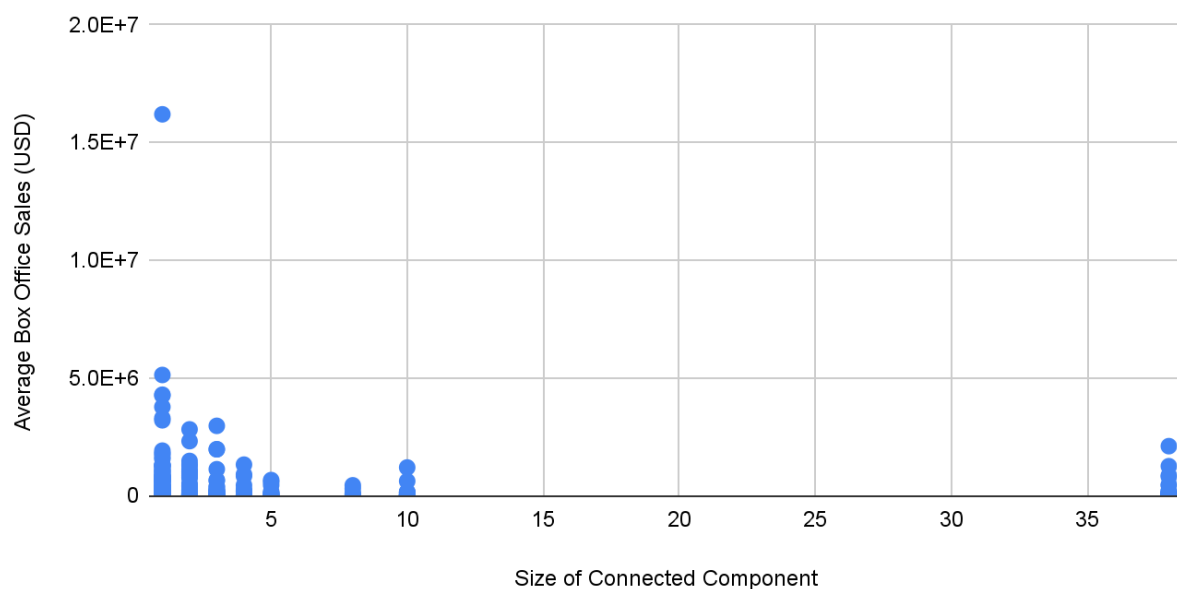
“network” of actors, giving a sense of the amount of nepotism they are subject to) to both their average box office revenue and average movie rating for each actor.

Data/Results

Nepotism Ratio ≈ 0.51062

This is defined as the number of actors in a connected component with more than 1 node divided by the total number of actors. Even a number this high is indicative of high nepotism. Almost 1 in every 2 actors is a product of nepotism! This can be seen by looking at the graph, wherever size of CCs is greater than 1.

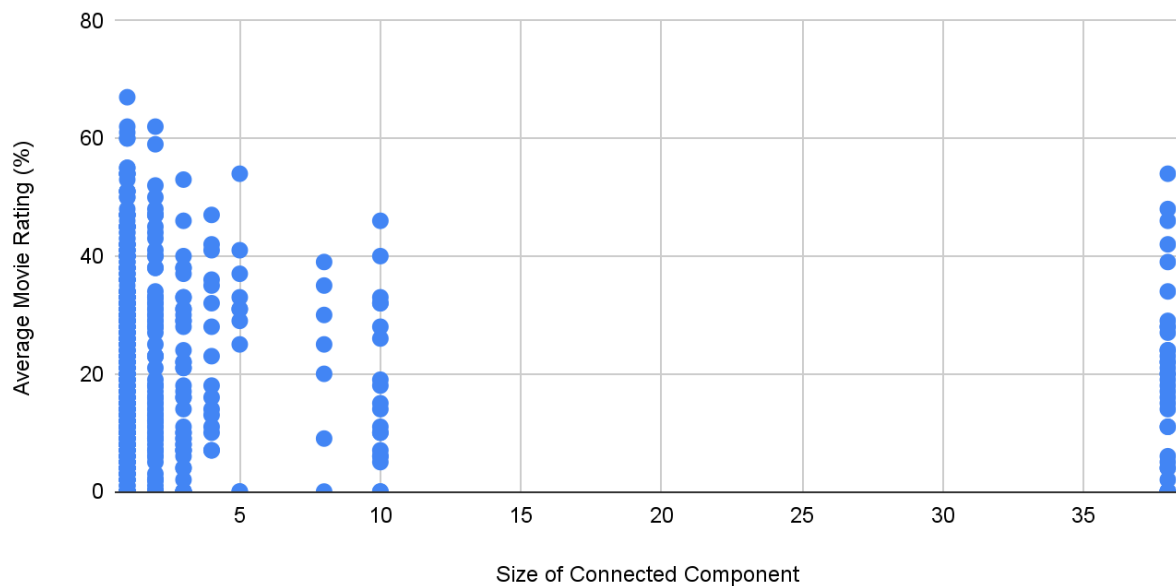
Figure A: Size of an Actor's Connected Component vs. Average Box Office Sales



In Figure A, we plotted the relationship between the size of an actor's connected component, which is the size of their network of relatives within the Bollywood industry, and their average box office sales. We can see that there is not a clear correlation between the size of an actor's connected component and the average box office sales for that actor. In fact, if we are looking at the first part of the graph with connected components ranging from size 0 to 10, there is actually a slight downwards trend, possibly suggesting that actors with a smaller sized connected component have higher box office sales. However, this might simply be due to the fact that there

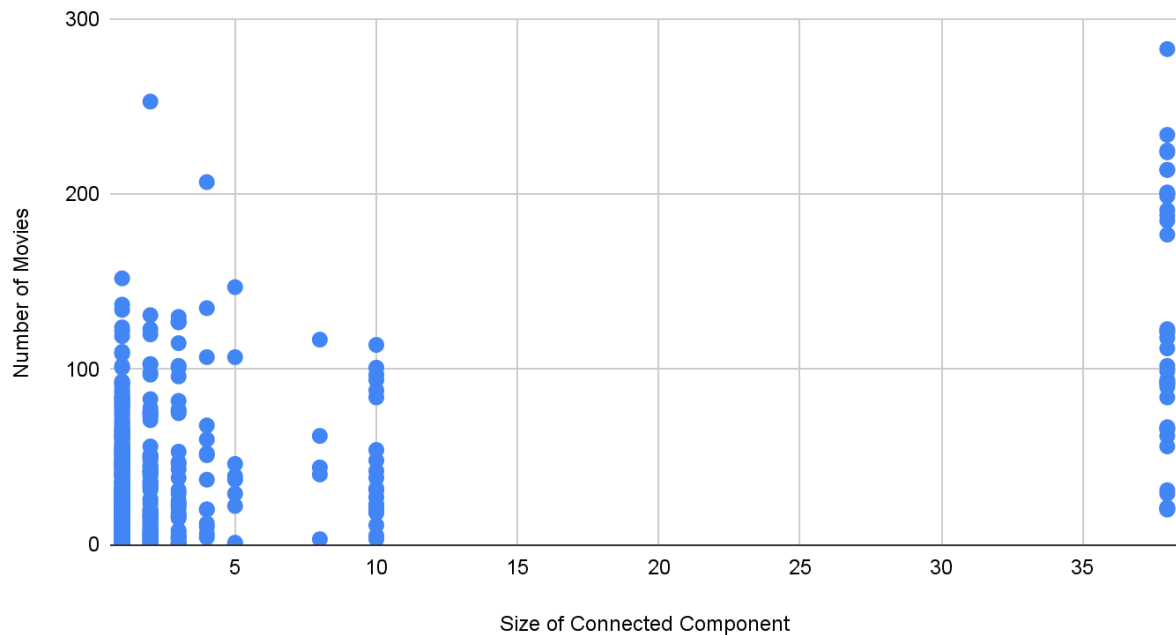
are a greater number of data points which have smaller connected components, so there is therefore a larger range of box office sales in that area.

Figure B: Size of Actor's Connected Component vs. Average Movie Rating



In Figure B, we plotted the relationship between the size of an actor's connected component and their average movie rating on Rotten Tomatoes. As you can see, there is no real trend that suggests that the size of an actor's connected component affects their average movie rating. There is however a slight indication of movie ratings dropping as nepotism increases, which lines in with our hypothesis.

Figure C: Size of an Actor's Connected Component vs. Number of Movies



In Figure C, we plotted the relationship between the size of an actor's connected component and the number of movies that they have been in. While it is not very strong, you can still see that there is a trend between the size of an actor's connected component and the number of movies that they have starred in. Even though we have more data in the lower half of our range, we can still see that actor's with a larger connected component tend to star in a greater number of movies.

Extending the analysis

We decided to extend the analysis to make a nepotism predictor in Bollywood. Based on the box-office performance of a movie, how connected is an actor? We tried running a Betweenness analysis, but this took an extensive amount of computational time and was beyond the scope of the class. The code we wrote is available on our git. Instead, we predicted the size of the connected components based on box-office performance. This was our effort to make a predictor for new debuts in bollywood.

What are the chances that a new debutant can perform well? Please look at the attached code to play with this analysis. While this wasn't linked to proving our hypothesis, it is an extension of our analysis.

The app runs locally on a shiny server, and was programmed using R. Though the relationship between debut performances is skewed by actors performing in movies together (non- connected actors debuting with actors that are otherwise connected), this was an interesting excursion in exploratory data analysis and programming the interactive application. Please upload the file with the actor_name, and box_office_score data, and a predictions column will be created accordingly. The application is viewable below.

The screenshot shows a web browser window with the address bar displaying 'http://127.0.0.1:3876'. The page title is 'How Connected Are You? Based on Debut Box Office Data'. The main content area has a light gray background and contains two primary interactive elements. The first is a file upload section titled 'upload csv file here' in bold. It includes a 'Browse...' button and a status indicator that says 'No file selected'. The second element is a button with a download icon and the text 'Download the Predicted Data from Here'. In the top right corner of the browser window, there is a 'Publish' button with a dropdown arrow.

Conclusion

Based on the data we collected, we found that there is not a statistically significant relationship between the size of an actor's connected component and their average movie rating nor their average box office sales. One important aspect to make note of is that in our analysis, we were not able to include any familiar relationships that actors had with important, non-actor figures in

the Bollywood industry (i.e. directors, producers, etc.), which could have impacted the size of the actors' connected components. We also were unable to include friendships, which could also help actors gain success and earn opportunities in the film industry. We also did not consider the length of an actor's career, which would likely affect their movie count as well as their average metrics.