Vincent Bianchi, Brett Lobsinger, Ben Lortie, Dian Wang
Professor Traeger
28 April 2024
Networks: Theory and Analysis

Using Network Theory to Cast Quentin Tarantino's Next Film

Introduction

Quentin Tarantino is one of the most celebrated directors of the past 30 years. His unique directorial style - which consists of rich dialogues, non-linear stories, and intense violence - makes him stand out when compared to other directors. He has demonstrated numerous times through his nine-feature filmography that he can gather an elite cast and deliver a thrilling spectacle to his audience. Amidst speculation about his next film is the idea of "Kill Bill: Volume 3", a continuation of the two-part feature starring Uma Thurman in 2003 and 2004. While many will argue that the story ended after the events of "Kill Bill: Volume 2", it has not stopped fans from guessing who could star in the next (hypothetical) installment. Tarantino tends to use a core group of actors across his films, with about 13% of his cast appearing in more than one of his films. Due to the interconnected nature of his casts, we theorize that network theory and analysis can provide substantial insight into who could appear in "Kill Bill: Volume 3". By examining the actor-to-actor relationships within the network structure, we are confident that we can suggest actors that not only fit Tarantino's casting preferences but also have the potential to deliver an epic performance to close out the Kill Bill franchise.

Network Creation

With so many actors being utilized across so many movies, hard-coding the individual data frames would have been needlessly difficult. So, we web-scraped Wikipedia to get a list of each actor used in a Tarantino film, as seen in lines 13 - 51 of the "scrape" R script. After cleaning each data frame, we combined the separate data frames to get one master list of all the actors. Though we were removing duplicate entries in our final data frame, we made sure to note how frequently an actor appeared in a Tarantino film to help us later on with our recommendations. After that, we used a combination function to find each actor-to-actor pairing across all of his films, as seen in lines 18 - 27 in the “actors df” R script. Each time an actor X appeared with another actor Y, the X-Y pairing received a point. Some pairings appeared multiple times, indicating that X and Y appeared together multiple times.

Our network nodes were the unique 187 actors, and the edges were the actor-to-actor combinations. Figure 1 in the appendix shows that our network was visually divided into distinct groups. This can be explained by how most actors have only appeared once in Tarantino’s films, so these actors are grouped together based on said film. To distinguish certain actors from each other, we used color to designate the number of times an actor appeared in Tarantino’s filmography. Completing the network creation allowed us to move on to the next step where we analyzed the network structure to make informed decisions about the next cast.

Methodology & Analysis

With so many network analysis methods at our disposal, there were plenty of different routes we could have gone down. To start, we analyzed the centralization of our network. This not only provided us with information about our network’s structure, but we were also able to find which actors had the highest centrality scores (or, lowest in the case of closeness centrality). By analyzing the degree, closeness, betweenness, and eigenvector centralities, we could find the most influential and connected actors in our network.

First, we found the overall centralization scores to be 0.69, 0.35, 0.17, and 0.67, respectively, as noted by the code in lines 11 - 21 of the “analysis” R script. With the degree score being so high, we acknowledge that the majority of the actors have lots of edges and are well-connected throughout the network. For closeness, the score is moderate-low, meaning a good handful of the actors are in the shortest path between nodes. While some actors may be further removed, most actors in the network are closely connected. As for betweenness, a low score indicates that there are only a few actors that act as a “bridge”. Some actors are like major highways, connecting lots of other actors, while others are more like smaller roads, connecting fewer actors. So, the network isn't evenly balanced in terms of who connects to whom. Finally, the high eigenvector score means that there's a pretty strong concentration of influence or importance among a few key actors. We imagine this score refers to how these films have a few A-list actors surrounded by actors with lower importance/influence.

After analyzing the relevant centrality measures, we decided to run an Exponential Random Graph Model (ERGM) to understand how different variables affect the ability to create a tie, as seen in lines 31 - 39 of the “analysis” R script. In addition to analyzing the density of the graph, we looked at the effect of numerical variables like IMDB rating and number of previous appearances. Furthermore, we created a binary variable to be analyzed in the ERGM as well, where actors that appeared more than once received a value of 1. Figure 2 in the appendix shows our ERGM model with coefficients. Ignoring significance, we see that as rating and frequency increase, so do the log odds of a tie forming. Additionally, the log odds increase if one of the actors has appeared in multiple films, as noted by the binary “if_mult” variable. We found that, when accounting for edges, the probability of forming a tie is about 5.3%. After accounting for the different variables in the ERGM, this value increases to about 8.9%. While this change may seem small at first glance, it is worth noting that each coefficient (for frequency, rating, and if_mult) positively affected the ability to form a tie. So, our group rationalized that Tarantino will likely pick a handful of actors that have a high number of previous appearances, as well as a high average IMDB rating.

Recommendations

Based on the outcomes of the network analysis, we were able to narrow down our cast suggestions for Kill Bill: Volume 3. The one selection that is non-negotiable is Uma Thurman, who would reprise her role as the Bride. Following her, the selections for the main cast are purely based on the ERGM analysis results. The next two actors are Kurt Russell and Zoë Bell, and their inclusion is based on the combination function in R - that pairing had the most appearances in Tarantino's filmography. Following them is Harvey Keitel who, among actors who appeared multiple times, has the highest average IMDB ratings from Tarantino films. Rounding out the main cast is Tim Roth, who had the highest number of appearances in the data frame (excluding Samuel L. Jackson and Michael Madsen, whose characters died in the events of "Kill Bill: Volume 2") Additionally, Tim Roth had the lowest closeness centrality, indicating that he lies in the shortest paths to get to other actors.

For the supporting cast, we used the ID numbers tied to the highest value for each centrality measure (except for closeness, in which the lowest value was used and found to be Tim Roth). Christian Brückner, who had a small role in "Inglourious Basterds", had the highest value for degree centrality. Based on betweenness, Austin Butler will make an appearance, following his minor role in "Once Upon a Time…in Hollywood". Finally, Amanda Plummer had the highest eigenvector score, marking her first Tarantino appearance since "Pulp Fiction".

Conclusion

Using network analysis, our team created a list of recommended actors to hypothetically feature in the ensemble cast of "Kill Bill: Volume 3". Analyzing centralization metrics allowed us to get a better idea of the structure of the actor network which, in turn, allowed us to find the most connected and influential actors. We followed that with an ERGM analysis, getting a better understanding of the factors that play into a tie being formed within the network. Given Quentin Tarantino's tendency to use a group of core actors in all of his films, we believe that our recommendations are plausible. While further analysis would be needed to boost the validity of these recommendations, we believe that the network analysis and insights provide a solid foundation for his next cast. Quentin Tarantino has a lot of actors at his disposal, and given the success of his previous films, we are confident that our findings can help him create a film that achieves further financial success and box office acclaim.

Appendix

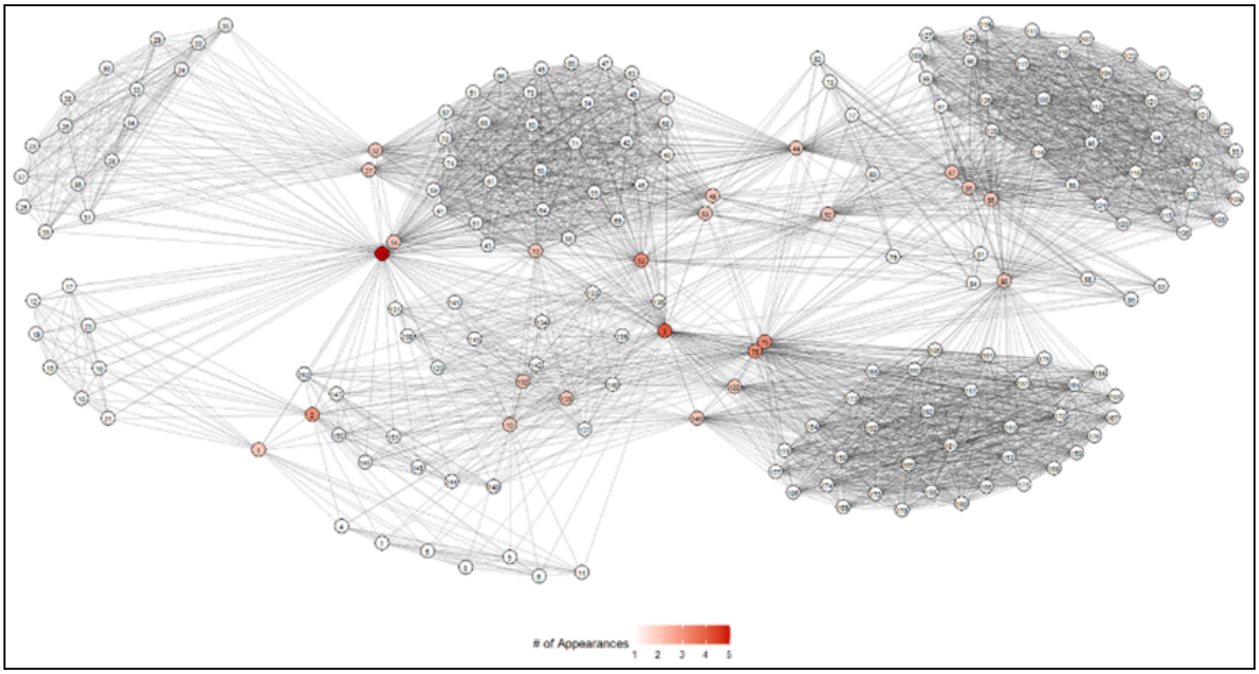

Figure 1: Network structure of the actors in Quentin Tarantino's films.

```
Call:
ergm(formula = net_sna ~ edges + nodecov("Freq") + nodecov("rating") +
    nodematch("if_mult") + nodefactor("if_mult"))

Maximum Likelihood Results:

                    Estimate Std. Error MCMC % z value Pr(>|z|)
edges               -2.87771    0.36341      0  -7.919  < 1e-04 ***
nodecov.Freq         0.35962    0.02949      0  12.193  < 1e-04 ***
nodecov.rating       0.01575    0.02208      0   0.713 0.475658
nodematch.if_mult    0.18589    0.05110      0   3.638 0.000275 ***
nodefactor.if_mult.1 0.42910    0.05929      0   7.237  < 1e-04 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 48218  on 34782  degrees of freedom
 Residual Deviance: 32308  on 34777  degrees of freedom

AIC: 32318  BIC: 32360  (Smaller is better. MC Std. Err. = 0)
```

Figure 2: ERGM model summary.