`

# MSIA 490 Social Network Analysis

*Project Report: US Film Citation Network*

**Ameer Khan**

**Xiang (Shawn) Li**

**Ye (Iris) Tu**

`

# Introduction

The focus of this project is to analyze the characteristics of network formed by film citations. A film is said to cite another film if a sequence, sentence, character, or other part of the referenced film has been adopted, used, or imitated in the referencing film [1]. Analyzing the film citation network gives insights into what characteristics of a film makes it an influential piece of art, and determines the significance of those works. The dataset for this project has been sourced from the Internet film Database (IMDb) and obtained from the Amaral Lab [2] at Northwestern University. It contains citation information as an edge list for over 15,000 US films, and 42,000 citation links.

# Data Preprocessing

Due to size limitations, only a subset of the network consisting of citing films released during the period 2009 to 2011, corresponding to the most recent citing films in the dataset, was considered for this project. This subset consists of 4,634 film citations by 1,001 citing films. There are 2,253 unique cited films from 1894 (Corbett and Courtney Before the Kinetograph) to 2010. In order to perform network analysis on the data, additional film attributes were obtained for each film in the dataset from the Open Movie Database, including IMDb rating, IMDb review counts, Rotten Tomato Tomatometer, production house, and film genre, using the OMDb API [3]. The genre of a film is an attribute with several tags in no particular order; so dummy boolean variables were created for representing major genres such as action, comedy, crime, documentary, drama, horror, romance and sci-fi.

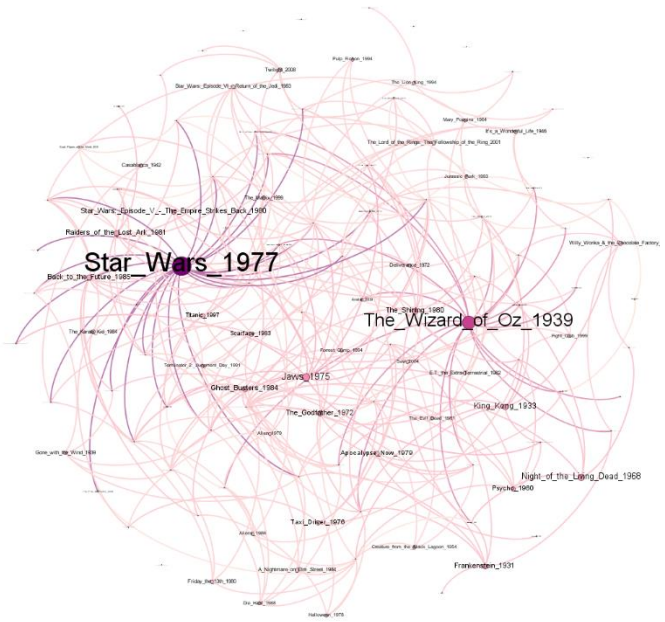
# Network Statistics

The film reference network is a directed graph with 3,185 films (nodes) and 4,634 citations (edges). The average degree of this network is 1.456, which indicates the average reference links (to or from the node) of a film. The network diameter is 3, which is the largest geodesic distance between any two films through references. The average length of geodesic distance between any two films in the network is 1.16.

Indegree centrality measures how many times a film has been referenced by other films. Since the subset network considered for the project is also fairly large, nodes with degree less than 12 have been filtered out in order to visualize indegree of films. The size of the nodes represents their indegree. In the figure below, Star Wars 1977 and The Wizard of Oz 1939 pops out as the nodes with the highest indegree. The top 10 films in terms of indegree centrality are listed below:

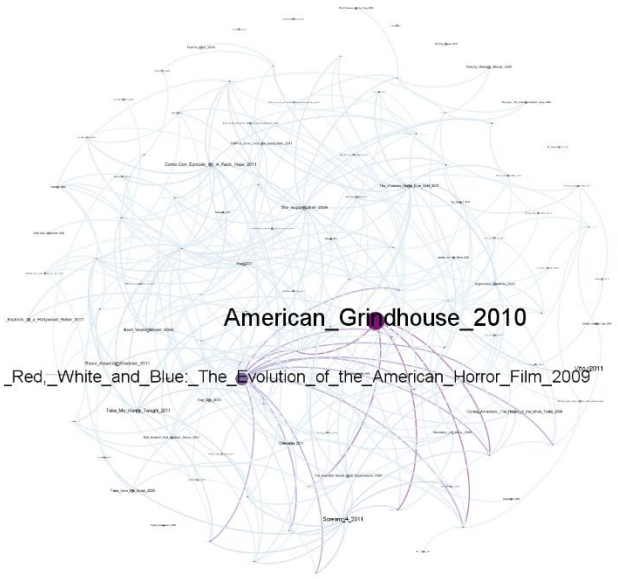
`

| Movies | In-Degree |
|---|---|
| Star Wars 1977 | 88 |
| The Wizard of Oz 1939 | 60 |
| Jaws 1975 | 32 |
| Night of the Living Dead 1968 | 26 |
| King Kong 1933 | 25 |
| The Shining 1980 | 23 |
| Raiders of the Lost Ark 1981 | 23 |
| Frankenstein 1931 | 22 |
| The Godfather 1972 | 22 |
| Star Wars: Episode V - The Empire Strikes Back 1980 | 21 |




Outdegree centrality indicates how many films a film refers to. As before, nodes with degree less than 12 were filtered out to visualize the films with high outdegree. The size of the nodes represents their outdegree. In the figure below, American Grindhouse 2010 and Nightmares in Red, White and Blue (both of them documentaries about American film history) stand out as the two central nodes with the highest outdegree.

| Movies | Out-Degree |
|---|---|
| American Grindhouse 2010 | 305 |
| Nightmares in Red, White and Blue 2009 | 207 |
| Vito 2011 | 75 |
| Scream 4 2011 | 62 |
| Take Me Home Tonight 2011 | 55 |
| These Amazing Shadows 2011 | 51 |
| Comic-Con Episode IV: A Fan's Hope 2011 | 48 |
| The Supermarket 2009 | 47 |
| Corman's World: Exploits of a Hollywood Rebel 2011 | 41 |
| Best Worst Movie 2009 | 41 |




Eigenvector centrality is another crucial network statistic that measures a node's centrality as a function of its neighbors' centrality. Films with high eigenvector centralities are the ones that are referenced by other significant and highly cited works. The top nodes in terms of indegree - Star Wars 1977, The Wizard of Oz 1939, Jaws 1975 - are the top three for eigenvector centrality as well. The top ten films with highest eigenvector centrality are listed below:

`

| Movies | Eigenvector Centrality |
|---|---|
| Star Wars 1977 | 1.000000000 |
| The Wizard of Oz 1939 | 0.761937155 |
| Jaws 1975 | 0.345318645 |
| The Shining 1980 | 0.300250218 |
| King Kong 1933 | 0.287003499 |
| Raiders of the Lost Ark 1981 | 0.260112046 |
| The Godfather 1972 | 0.251133561 |
| Star Wars: Episode V | 0.25108956 |
| Night of the Living Dead 1968 | 0.22375082 |
| Ghost Busters 1984 | 0.215351626 |

# Hypotheses

Based on the initial exploration of the data and network statistics, the following characteristics about the film citation network were hypothesized:

- Are films with high IMDb ratings more likely to be cited?
- Do popular films (with a large number of reviews) get more citations?
- Do critically acclaimed films (in terms of Tomatometer) get cited more?
- Are films from top production houses referenced more frequently?
- Do films tend to cite others from the same production house?
- Are films likely to cite others from the same genre?

# Citation Network Model

Exponential random graph models (ERGM) were built to test the above proposed hypotheses. Effects for validating network structure properties such as edges and transitive triads, along with node attributes were added to the ERGM model. The results from the final model are presented below:

***ERGM Model: Effects for Network Structure***

| **Effect** | **Property** | **Estimate** | **Probability** | **Interpretation** |
|---|---|---|---|---|
| Edges | edges | -8.551*** | 0.000193 | Film references are not likely to be random |
| Transitivity | transitive | 1.25*** | 3.490 | Cited films tend to have iconic/memorable aspects |

`

The edges effect is to condition the baseline for the model, and its parameter measures the density of the network. The parameter for edges is strongly negative, indicating that citations between films are not random as the network density is very low.

The transitivity effect measures local transitive triads in the network. To interpret this effect for the citation network, it is essential to understand how the citation links are created between films. For instance, if an action film A features an innovative action sequence, and another action film B adapts that scene, B is said to cite A. If a comedy film C spoofs the same original scene from film A, it will refer to both films B and A, forming a transitive triad. Thus, transitivity indicates if the same scenes/concepts from a film are repeatedly adapted by different films. In other words, it indicates if cited movies tend to have iconic or memorable aspects about them.

Only transitivity is tested for in the network because it is a directed graph and contains citation links with a temporal characteristic (they only go backwards in time, from a more recent film to an older film), and hence, only transitive triangles are possible in the network.

The triangles parameter can be interpreted such that if a dyad is added to the network that forms no transitive triangles, the probability of this happening is 0.00019. If adding the dyad results in addition of one or two transitive triangles in the network, the probabilities are 0.00067 and 0.00235 respectively [4].

***ERGM Model: Effects for Ratings and Popularity***

| **Effect** | **Property** | **Estimate** | **Odds** | **Interpretation** |
|---|---|---|---|---|
| IMDb Rating of Cited Film | nodeicov | 0.05219*** | 1.05358 | Highly rated films are likely to be cited more |
| IMDb Votes of Cited Film | nodeicov | 0.000002*** | 1.00000 | Film popularity in terms of number of ratings tend to be cited more |
| Tomatometer of Cited Film | nodeicov | 0.010980*** | 1.01104 | Critical acclaim increases likelihood of citation |

The IMDb rating of a film is the overall user rating ranging from 1 through 10, while the IMDb vote count is the number of users who rated the film, which is a proxy for popularity of a film. Both rating and vote counts were included in the model since popular films may not always be rated highly. The positive parameter for IMDb rating implies that highly rated films tend to be cited more. Specifically, a film having one point higher than another has 5% more chance to be cited, controlling for all other factors. The parameter for IMDb votes is very small, and the odds are nearly 1. However, this parameter is indicating the effect of one vote for a film on IMDb, where popular films can garner votes in the order of millions. The odds can be interpreted as for every other factor being equal; a film with 10,000 more votes has a 2% more chance to be cited.

Tomatometer is a film rating metric by Rotten Tomatoes that measures how many film critics have given positive reviews for a film. If a film gets 40 positive reviews from critics and 10 negative ones, its tomatometer would be 80%. The model suggests that for every one percent increase in tomatometer increases the chances of citation for that film by 1%, all else being equal.

`

***ERGM Model: Effects for Production House***

| Effect | Property | Estimate | Odds | Interpretation |
|---|---|---|---|---|
| Production House Homophily | nodematch | 0.568*** | 1.7642 | Films from a production house are likely to cite older films from same production, for major production houses |
| 20th Century Fox | nodeifactor | 0.324*** | 1.3832 | Films from 20th Century Fox, MGM, Paramount Pictures, Universal Pictures, Warner Bros. and Warner Home studios are more likely to be referenced by other films. Films from Universal and Paramount have the highest likelihood of being cited |
| Columbia Pictures | nodeifactor | 0.0508 | 1.0521 | |
| MCA Universal Home | nodeifactor | 0.1727 | 1.1885 | |
| MGM | nodeifactor | 0.3076* | 1.3602 | |
| MGM Home | nodeifactor | 0.2107 | 1.2345 | |
| Paramount Pictures | nodeifactor | 0.336*** | 1.3999 | |
| Sony Home | nodeifactor | 0.0985 | 1.1035 | |
| Universal Pictures | nodeifactor | 0.357*** | 1.4293 | |
| Warner Bros. Pictures | nodeifactor | 0.2898*** | 1.3362 | |
| Warner Home Video | nodeifactor | 0.316*** | 1.3718 | |

Production house homophily measures how likely are films from major production houses to cite films produced by the same studios. The major studios were selected as the top 10 studios in the citation network subset based on frequency and are listed below:

- 20th Century Fox
- Columbia Pictures
- MCA Universal Home Video
- MGM
- MGM Home Entertainment
- Paramount Pictures
- Sony Home Entertainment
- Universal Pictures
- Warner Bros. Pictures
- Warner Home Video

`

The parameter for production house homophily indicates that movies from these studios are 76% more likely to cite other movies produced by the same production house. The likelihood that a movie is cited given that it is produced by a specific studio from the top ten list was also measured by the model. Movies produced by 20th Century Fox, MGM, Universal Pictures, Paramount Pictures and Warner Bros. studios tend to be cited more, while there isn't sufficient evidence to draw conclusions about the other production houses. Controlling for other factors, movies from Universal (43%) and Paramount (40%) have the highest tendency to be cited.

***ERGM Model: Effects for Genre Tag***

| **Effect** | **Property** | **Estimate** | **Odds** | **Interpretation** |
|---|---|---|---|---|
| Action Homophily | nodematch | -0.0152 | 0.98488 | Crime, romance and sci-fi films tend to cite others that share the same genre tag. Comedies and documentaries are much less likely to cite films that have the same respective tags |
| Comedy Homophily | nodematch | -0.236*** | 0.79002 | |
| Crime Homophily | nodematch | 0.548*** | 1.72996 | |
| Documentary Homophily | nodematch | -1.235*** | 0.29083 | |
| Horror Homophily | nodematch | -0.0045 | 0.99551 | |
| Romance Homophily | nodematch | 0.2498*** | 1.28377 | |
| Sci-Fi Homophily | nodematch | 0.126* | 1.13383 | |

Films have several genre tags, and those having the most common tags were analyzed for genre homophily. These genre categories included action, comedy, crime, documentary, horror, romance and sci-fi. Effects were added to the model to measure homophily for each category under consideration, that is, do films with a specific genre tag get cited by other movies which also have the same tag. Drama was not included as a genre in the model because it raised multicollinearity issues due to high cross-correlation. This may be due to drama being a very broad genre tag, overlapping heavily with other genre tags.

The model indicates genre homophily for crime, romance and sci-fi movies, that is, citation links are more likely to be between two romance films than between a romance and an non-romance movie, for instance. However, in case of comedies and documentaries, citation links tend to be heterophilic, that is, documentaries are likely to cite non-documentaries than another documentary (81% likelihood), and comedies are less likely to cite other comedies (21%). Intuitively, these results make sense as documentaries might refer to movie scenes and dialogs, especially if the focus is film history, filmmaking, etc. Comedies also refer to a lot of non-

`

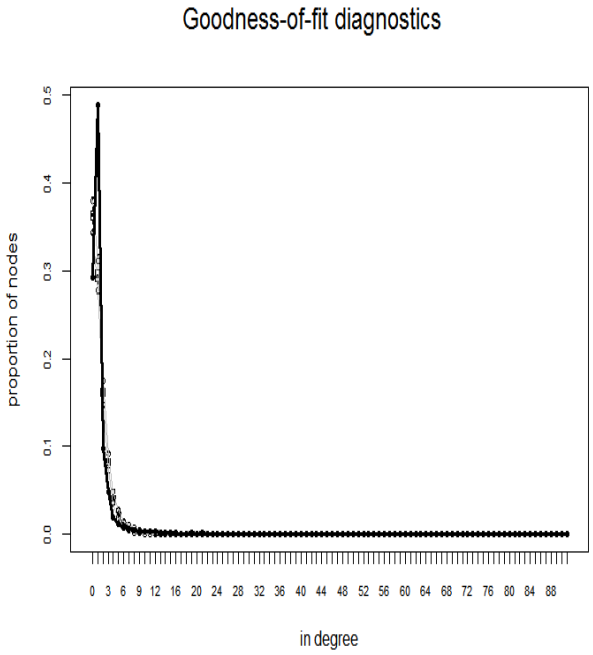
comedy movies as they spoof or comically recreate scenes from famous films, but rarely refer to other comedies, unless they are sequels. There wasn't enough evidence to draw conclusions about action and horror homophily.

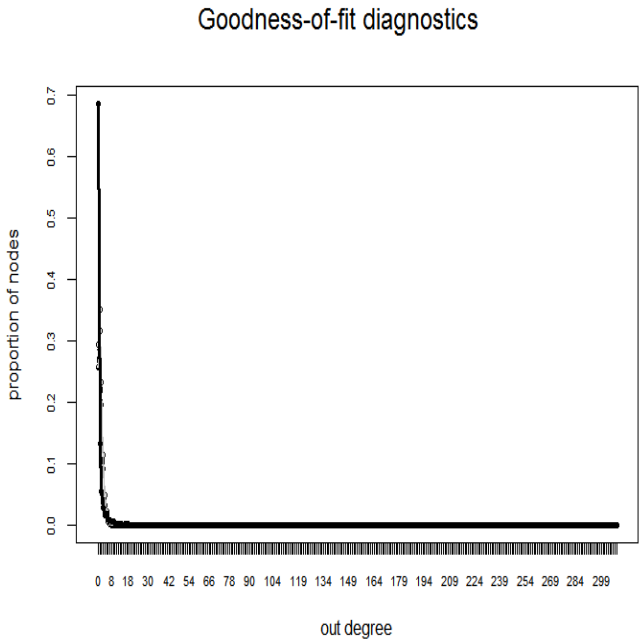
# Model Diagnostics

### *Goodness-of-Fit*

The goodness-of-fit measures tests the models for how well they fit the observed data, by comparing network characteristics that were not used for fitting with randomized networks generated using the ERGM model. A model is a good fit if the observed network has similar statistics as randomly generated networks, i.e., the p-values are high. The model is tested for goodness of fit using indegree, outdegree and minimum geodesic distance as measures.

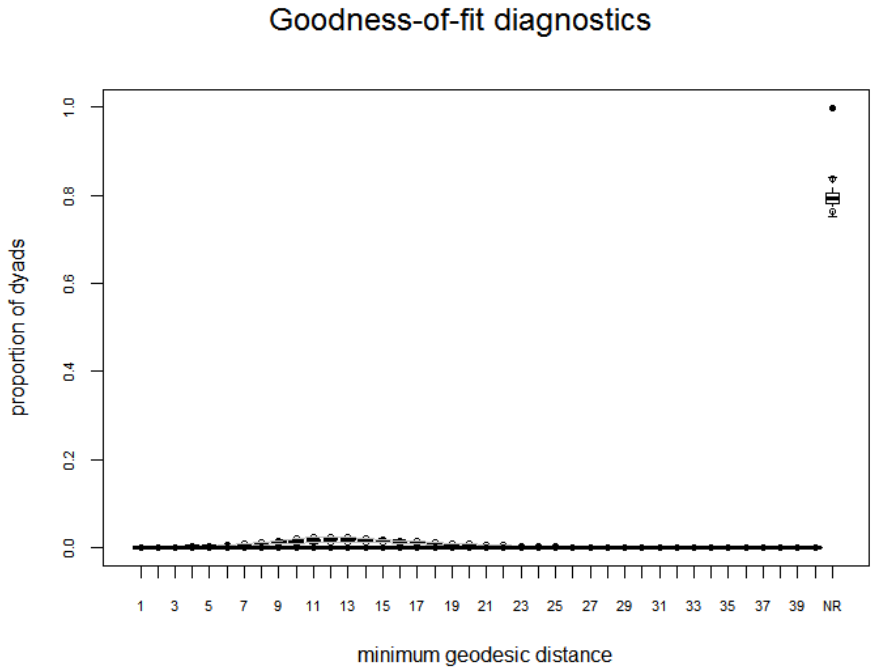
It is noticed that for indegree, the observed network does not lie within the 95% confidence band of the probability distribution of randomized networks, especially when the indegree is less than 8. Similar results are observed for outdegree goodness-of-fit statistics as well. This is because the randomized networks are following a power law distribution for indegree and outdegree, while the citation network has some structural idiosyncrasies, which may be due to the temporal characteristic of links (edges only point backwards in time) or the way the data was subset (latest citing movies).



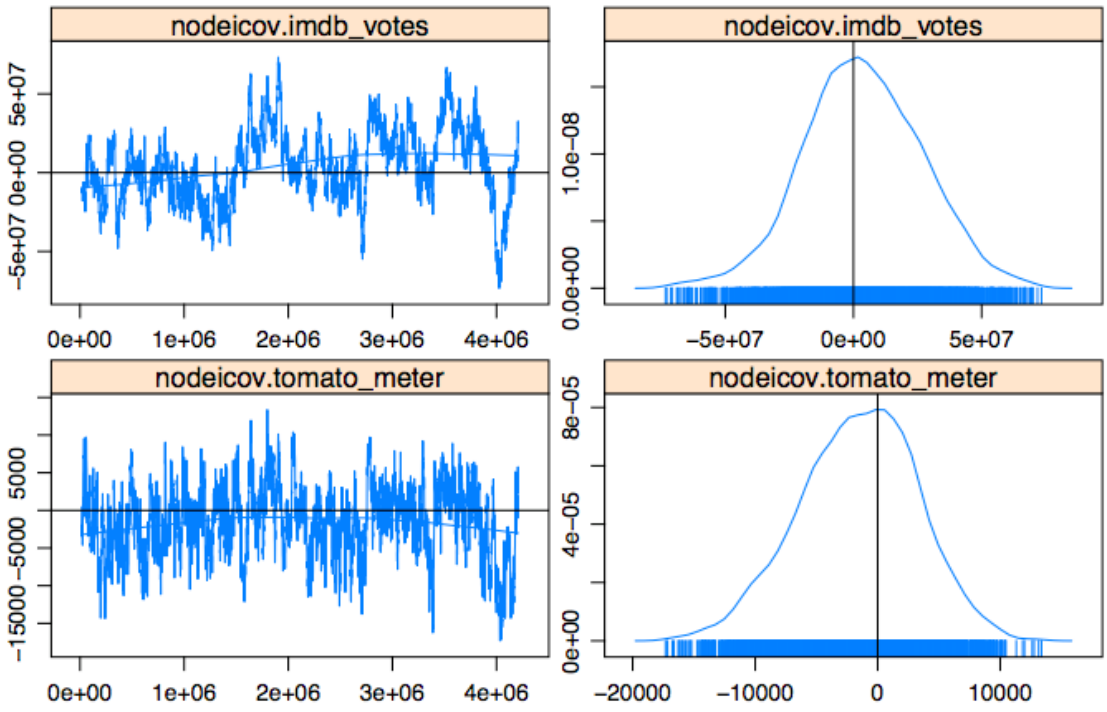

The geodesic distance was also used to measure goodness-of-fit, and the results showed that the observed network had a much different distribution of dyads based on geodesic distance than randomized networks generated using the model. Again, peculiar properties of the citation network might be responsible for the poor fit.

`




***MCMC Diagnostics***

Statistics for MCMC diagnostics are shown below for a few of the model parameters. The plots indicate that the MCMC estimates for all the effects included in the model have a smooth, quasi-normal distribution, indicating that their corresponding coefficients have converged to stable values. The complete diagnostics have been attached as a PDF document.



`

# References


1. US film citation network, amaral-lab.org/resources/data-sets/us-film-citation-network
2. Wasserman M, Zeng XHT, Amaral LAN. "Cross-evaluation of metrics to estimate the significance of creative works", Proceedings of the National Academy of Sciences of the U.S.A.
3. Python wrapper for OMDb API, https://pypi.python.org/pypi/omdb/0.2.0
4. Triad formation in ERGM models, https://statnet.org/workshops/SUNBELT/EUSN/ergm/ergm_tutorial.html#triad-formation

`

# Appendix

## *Model Summary*

```
Formula:   citationsNetwork ~ edges + transitive + nodeicov("imdb_rating") +
    nodeicov("imdb_votes") + nodeicov("tomato_meter") + nodematch("production",
    keep = c(1:5, 7:11)) + nodeifactor("production", base = 6) +
    nodematch("action", keep = 1) + nodematch("comedy", keep = 1) +
    nodematch("crime", keep = 1) + nodematch("doc", keep = 1) +
    nodematch("horror", keep = 1) + nodematch("romance", keep = 1) +
    nodematch("scifi", keep = 1)

Iterations:  56 out of 100

Monte Carlo MLE Results:
                                                    Estimate Std. Error MCMC %  p-value
edges                                             -8.551e+00  1.517e-01      0  < 1e-04 ***
transitive                                         1.250e+00  1.754e-01      0  < 1e-04 ***
nodeicov.imdb_rating                               5.219e-02  1.544e-02      0 0.000727 ***
nodeicov.imdb_votes                                2.107e-06  7.351e-08      1  < 1e-04 ***
nodeicov.tomato_meter                              1.098e-02  5.973e-04      0  < 1e-04 ***
nodematch.production                               5.677e-01  1.310e-01      0  < 1e-04 ***
nodeifactor.production.20th Century Fox            3.244e-01  8.041e-02      0  < 1e-04 ***
nodeifactor.production.Columbia Pictures           5.082e-02  1.127e-01      0 0.651920
nodeifactor.production.MCA Universal Home Video    1.727e-01  1.561e-01      0 0.268483
nodeifactor.production.MGM                         3.076e-01  1.213e-01      0 0.011242 *
nodeifactor.production.MGM Home Entertainment      2.107e-01  1.479e-01      0 0.154319
nodeifactor.production.Paramount Pictures          3.364e-01  6.100e-02      0  < 1e-04 ***
nodeifactor.production.Sony Pictures Home Entertainment  9.846e-02  1.268e-01      0 0.437582
nodeifactor.production.Universal Pictures          3.572e-01  7.603e-02      0  < 1e-04 ***
nodeifactor.production.Warner Bros. Pictures       2.898e-01  7.723e-02      0 0.000175 ***
nodeifactor.production.Warner Home Video           3.161e-01  1.353e-01      0 0.019441 *
nodematch.action                                  -1.524e-02  4.561e-02      0 0.738212
nodematch.comedy                                  -2.357e-01  3.712e-02      0  < 1e-04 ***
nodematch.crime                                    5.481e-01  4.677e-02      0  < 1e-04 ***
nodematch.doc                                     -1.235e+00  4.548e-02      0  < 1e-04 ***
nodematch.horror                                  -4.496e-03  4.458e-02      0 0.919667
nodematch.romance                                  2.498e-01  4.416e-02      0  < 1e-04 ***
nodematch.scifi                                    1.256e-01  5.325e-02      0 0.018355 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance: 14058467  on 10141040  degrees of freedom
 Residual Deviance:    76215  on 10141017  degrees of freedom

AIC: 76261    BIC: 76586    (Smaller is better.)
```

## *Goodness-of-fit for in-degree*

| | obs | min | mean | max | MC p-value |
|---|---|---|---|---|---|
| 0 | 932 | 1091 | 1155.99 | 1225 | 0 |
| 1 | 1558 | 869 | 938.71 | 1034 | 0 |
| 2 | 310 | 456 | 513.77 | 584 | 0 |
| 3 | 154 | 224 | 262.81 | 296 | 0 |
| 4 | 59 | 108 | 131.26 | 163 | 0 |
| 5 | 39 | 48 | 70.88 | 92 | 0 |
| 6 | 23 | 27 | 38.59 | 55 | 0 |
| 7 | 17 | 14 | 24.24 | 39 | 0.1 |

`

| | | | | | |
|---|---|---|---|---|---|
| 8 | 15 | 6 | 14.61 | 25 | 1 |
| 9 | 12 | 3 | 9.33 | 16 | 0.46 |
| 10 | 10 | 1 | 5.85 | 12 | 0.14 |
| 11 | 12 | 0 | 3.95 | 9 | 0 |
| 12 | 8 | 0 | 2.9 | 7 | 0 |
| 13 | 6 | 0 | 2.3 | 7 | 0.04 |
| 14 | 4 | 0 | 1.67 | 5 | 0.2 |
| 15 | 6 | 0 | 1.06 | 4 | 0 |
| 16 | 2 | 0 | 1.12 | 5 | 0.6 |
| 17 | 1 | 0 | 0.84 | 3 | 1 |
| 18 | 2 | 0 | 0.65 | 4 | 0.32 |
| 19 | 3 | 0 | 0.62 | 2 | 0 |
| 20 | 0 | 0 | 0.48 | 3 | 1 |
| 21 | 3 | 0 | 0.47 | 2 | 0 |
| 22 | 2 | 0 | 0.46 | 2 | 0.12 |
| 23 | 2 | 0 | 0.32 | 3 | 0.16 |
| 24 | 0 | 0 | 0.28 | 2 | 1 |
| 25 | 1 | 0 | 0.14 | 1 | 0.28 |
| 26 | 1 | 0 | 0.21 | 1 | 0.42 |
| 27 | 0 | 0 | 0.16 | 1 | 1 |
| 28 | 0 | 0 | 0.23 | 2 | 1 |
| 29 | 0 | 0 | 0.22 | 3 | 1 |
| 30 | 0 | 0 | 0.08 | 1 | 1 |
| 31 | 0 | 0 | 0.07 | 1 | 1 |
| 32 | 1 | 0 | 0.08 | 1 | 0.16 |
| 33 | 0 | 0 | 0.19 | 1 | 1 |
| 34 | 0 | 0 | 0.06 | 1 | 1 |
| 35 | 0 | 0 | 0.07 | 1 | 1 |
| 36 | 0 | 0 | 0.07 | 1 | 1 |
| 38 | 0 | 0 | 0.09 | 1 | 1 |
| 39 | 0 | 0 | 0.02 | 1 | 1 |
| 40 | 0 | 0 | 0.01 | 1 | 1 |
| 42 | 0 | 0 | 0.02 | 1 | 1 |
| 43 | 0 | 0 | 0.01 | 1 | 1 |
| 45 | 0 | 0 | 0.01 | 1 | 1 |
| 48 | 0 | 0 | 0.01 | 1 | 1 |
| 52 | 0 | 0 | 0.01 | 1 | 1 |
| 54 | 0 | 0 | 0.01 | 1 | 1 |
| 57 | 0 | 0 | 0.01 | 1 | 1 |
| 60 | 1 | 0 | 0 | 0 | 0 |

`

| | | | | | |
|---|---|---|---|---|---|
| 66 | 0 | 0 | 0.01 | 1 | 1 |
| 70 | 0 | 0 | 0.01 | 1 | 1 |
| 71 | 0 | 0 | 0.01 | 1 | 1 |
| 75 | 0 | 0 | 0.01 | 1 | 1 |
| 80 | 0 | 0 | 0.01 | 1 | 1 |
| 85 | 0 | 0 | 0.01 | 1 | 1 |
| 88 | 1 | 0 | 0 | 0 | 0 |

## *Goodness-of-fit for out-degree*

| | obs | min | mean | max | MC p-value |
|---|---|---|---|---|---|
| 0 | 2184 | 797 | 863.6 | 1238 | 0 |
| 1 | 420 | 1002 | 1056.71 | 1127 | 0 |
| 2 | 177 | 467 | 694.22 | 752 | 0 |
| 3 | 94 | 186 | 329.08 | 383 | 0 |
| 4 | 55 | 89 | 133.06 | 170 | 0 |
| 5 | 62 | 36 | 57.86 | 86 | 0.6 |
| 6 | 32 | 15 | 26.03 | 40 | 0.28 |
| 7 | 30 | 4 | 12.45 | 23 | 0 |
| 8 | 20 | 1 | 6.21 | 11 | 0 |
| 9 | 19 | 0 | 2.35 | 9 | 0 |
| 10 | 17 | 0 | 1.29 | 7 | 0 |
| 11 | 8 | 0 | 0.75 | 8 | 0.02 |
| 12 | 7 | 0 | 0.22 | 2 | 0 |
| 13 | 5 | 0 | 0.22 | 4 | 0 |
| 14 | 6 | 0 | 0.13 | 2 | 0 |
| 15 | 2 | 0 | 0.11 | 4 | 0.06 |
| 16 | 8 | 0 | 0.06 | 2 | 0 |
| 17 | 3 | 0 | 0.09 | 3 | 0.02 |
| 18 | 5 | 0 | 0.05 | 2 | 0 |
| 19 | 1 | 0 | 0.05 | 1 | 0.1 |
| 20 | 1 | 0 | 0.03 | 1 | 0.06 |
| 21 | 1 | 0 | 0.05 | 2 | 0.08 |
| 22 | 1 | 0 | 0.04 | 3 | 0.04 |
| 23 | 3 | 0 | 0.01 | 1 | 0 |
| 24 | 0 | 0 | 0.03 | 1 | 1 |
| 25 | 1 | 0 | 0.03 | 1 | 0.06 |
| 26 | 1 | 0 | 0.03 | 1 | 0.06 |
| 27 | 1 | 0 | 0.02 | 1 | 0.04 |
| 28 | 1 | 0 | 0.03 | 1 | 0.06 |

`

| | | | | | |
|---|---|---|---|---|---|
| 29 | 1 | 0 | 0.02 | 1 | 0.04 |
| 30 | 1 | 0 | 0.02 | 1 | 0.04 |
| 31 | 1 | 0 | 0 | 0 | 0 |
| 32 | 1 | 0 | 0.02 | 2 | 0.02 |
| 33 | 3 | 0 | 0 | 0 | 0 |
| 34 | 0 | 0 | 0.03 | 1 | 1 |
| 37 | 1 | 0 | 0 | 0 | 0 |
| 39 | 0 | 0 | 0.01 | 1 | 1 |
| 40 | 2 | 0 | 0.01 | 1 | 0 |
| 41 | 2 | 0 | 0 | 0 | 0 |
| 46 | 0 | 0 | 0.01 | 1 | 1 |
| 47 | 1 | 0 | 0 | 0 | 0 |
| 48 | 1 | 0 | 0.01 | 1 | 0.02 |
| 49 | 1 | 0 | 0 | 0 | 0 |
| 50 | 0 | 0 | 0.01 | 1 | 1 |
| 55 | 1 | 0 | 0 | 0 | 0 |
| 62 | 1 | 0 | 0 | 0 | 0 |
| 65 | 0 | 0 | 0.01 | 1 | 1 |
| 75 | 1 | 0 | 0 | 0 | 0 |
| 81 | 0 | 0 | 0.01 | 1 | 1 |
| 87 | 0 | 0 | 0.01 | 1 | 1 |
| 120 | 0 | 0 | 0.01 | 1 | 1 |
| 145 | 0 | 0 | 0.01 | 1 | 1 |
| 207 | 1 | 0 | 0 | 0 | 0 |
| 305 | 1 | 0 | 0 | 0 | 0 |

## *Goodness-of-fit for minimum geodesic distance*

| | obs | min | mean | max | MC p-value |
|---|---|---|---|---|---|
| 1 | 4634 | 4429 | 4610.61 | 4738 | 0.84 |
| 2 | 750 | 5230 | 7546.42 | 8336 | 0 |
| 3 | 60 | 6207 | 12029.36 | 14083 | 0 |
| 4 | 0 | 7045 | 19032.03 | 23832 | 0 |
| 5 | 0 | 7316 | 29609.71 | 39605 | 0 |
| 6 | 0 | 8212 | 44935.04 | 63581 | 0 |
| 7 | 0 | 9125 | 65849.44 | 98381 | 0 |
| 8 | 0 | 10500 | 92265.07 | 142482 | 0 |
| 9 | 0 | 11220 | 122194.25 | 190640 | 0 |

`

| | | | | | |
|---|---|---|---|---|---|
| 10 | 0 | 10875 | 151807.25 | 233941 | 0 |
| 11 | 0 | 10112 | 176138.82 | 263395 | 0 |
| 12 | 0 | 9235 | 190613.03 | 271630 | 0 |
| 13 | 0 | 8513 | 192843.55 | 255714 | 0 |
| 14 | 0 | 7572 | 183523.77 | 228111 | 0 |
| 15 | 0 | 6311 | 165237.46 | 204530 | 0 |
| 16 | 0 | 5061 | 141754.2 | 180728 | 0 |
| 17 | 0 | 4254 | 116713.05 | 150975 | 0 |
| 18 | 0 | 3632 | 92695.96 | 132546 | 0 |
| 19 | 0 | 3073 | 71445.56 | 112989 | 0 |
| 20 | 0 | 2352 | 53703.55 | 94095 | 0 |
| 21 | 0 | 1845 | 39511.37 | 77199 | 0 |
| 22 | 0 | 1321 | 28549.28 | 62385 | 0 |
| 23 | 0 | 954 | 20292.64 | 49350 | 0 |
| 24 | 0 | 689 | 14242.19 | 37961 | 0 |
| 25 | 0 | 493 | 9891.27 | 28482 | 0 |
| 26 | 0 | 353 | 6832.76 | 21691 | 0 |
| 27 | 0 | 113 | 4705 | 17133 | 0 |
| 28 | 0 | 32 | 3224.78 | 13915 | 0 |
| 29 | 0 | 8 | 2203.39 | 10890 | 0 |
| 30 | 0 | 1 | 1503.05 | 8479 | 0 |
| 31 | 0 | 0 | 1024.59 | 6792 | 0.04 |
| 32 | 0 | 0 | 701.93 | 6640 | 0.1 |
| 33 | 0 | 0 | 484.78 | 6123 | 0.14 |
| 34 | 0 | 0 | 337.13 | 5646 | 0.3 |
| 35 | 0 | 0 | 235.28 | 5093 | 0.42 |
| 36 | 0 | 0 | 160.72 | 4247 | 0.72 |
| 37 | 0 | 0 | 111.55 | 3348 | 0.92 |
| 38 | 0 | 0 | 76.77 | 2653 | 1 |
| 39 | 0 | 0 | 51.85 | 1932 | 1 |
| 40 | 0 | 0 | 33.4 | 1304 | 1 |
| 41 | 0 | 0 | 21.4 | 866 | 1 |
| 42 | 0 | 0 | 13.28 | 521 | 1 |
| 43 | 0 | 0 | 8.7 | 335 | 1 |
| 44 | 0 | 0 | 6.21 | 323 | 1 |
| 45 | 0 | 0 | 4.86 | 310 | 1 |
| 46 | 0 | 0 | 3.67 | 273 | 1 |
| 47 | 0 | 0 | 2.73 | 234 | 1 |
| 48 | 0 | 0 | 2.02 | 179 | 1 |
| 49 | 0 | 0 | 1.32 | 122 | 1 |

`

| | | | | | |
|---|---|---|---|---|---|
| 50 | 0 | 0 | 0.69 | 65 | 1 |
| 51 | 0 | 0 | 0.27 | 27 | 1 |
| 52 | 0 | 0 | 0.06 | 6 | 1 |
| 53 | 0 | 0 | 0.02 | 2 | 1 |
| Inf | 10135596 | 7628439 | 8072252.91 | 9994221 | 0 |