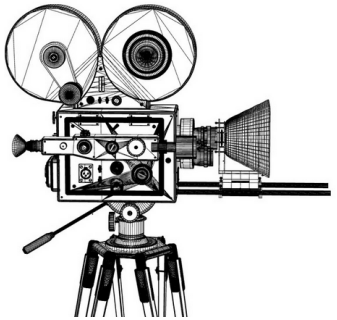
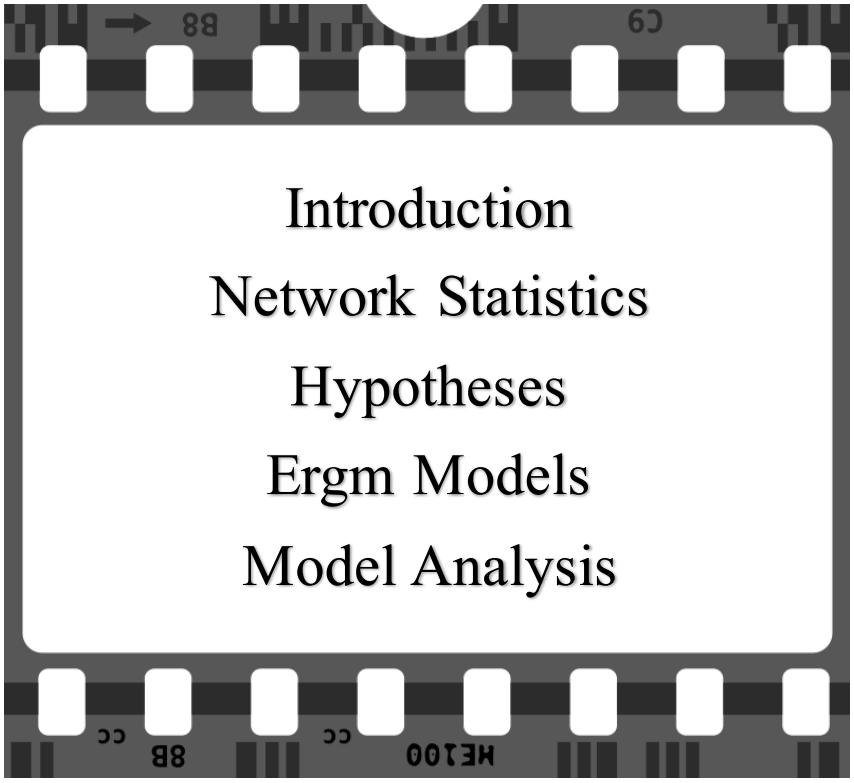
U.S. Movies Citation Network

Ameer Khan Shawn Li Ye (Iris) Tu



Overview


Introduction
Network Statistics
Hypotheses
Ergm Models
Model Analysis

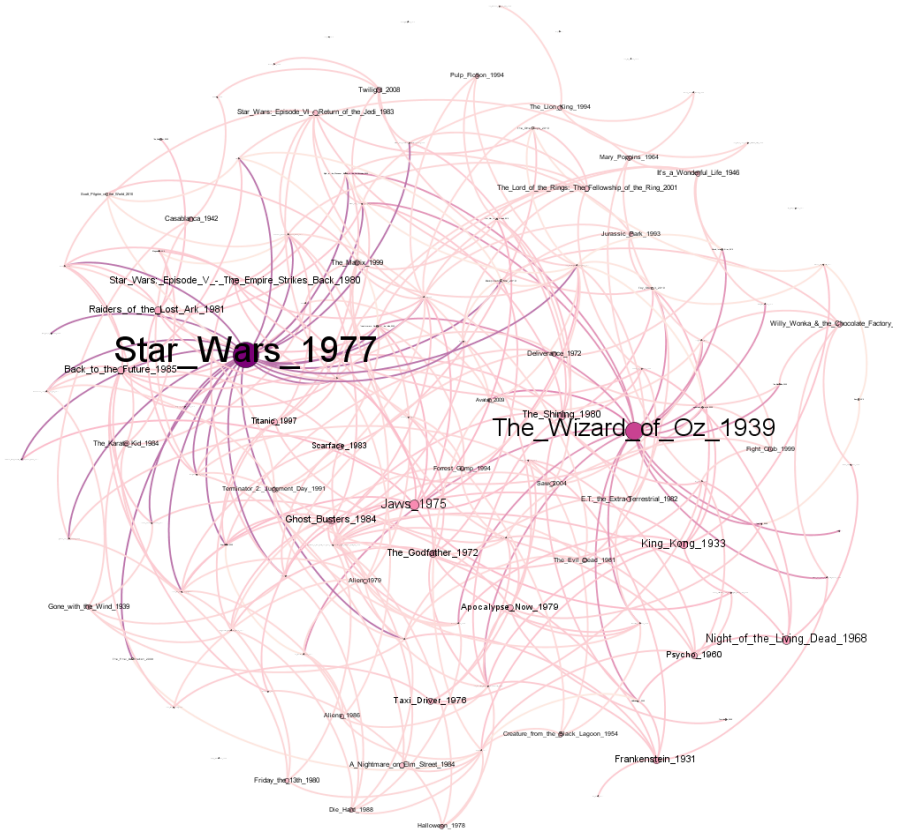
Introduction

Movie Citation Network

- 3185 Movies
- 4634 References
- Time Span: 2009 – 2011
- Additional data from OMDB API, including imdb ratings, tomato meter, production house, genre and etc.

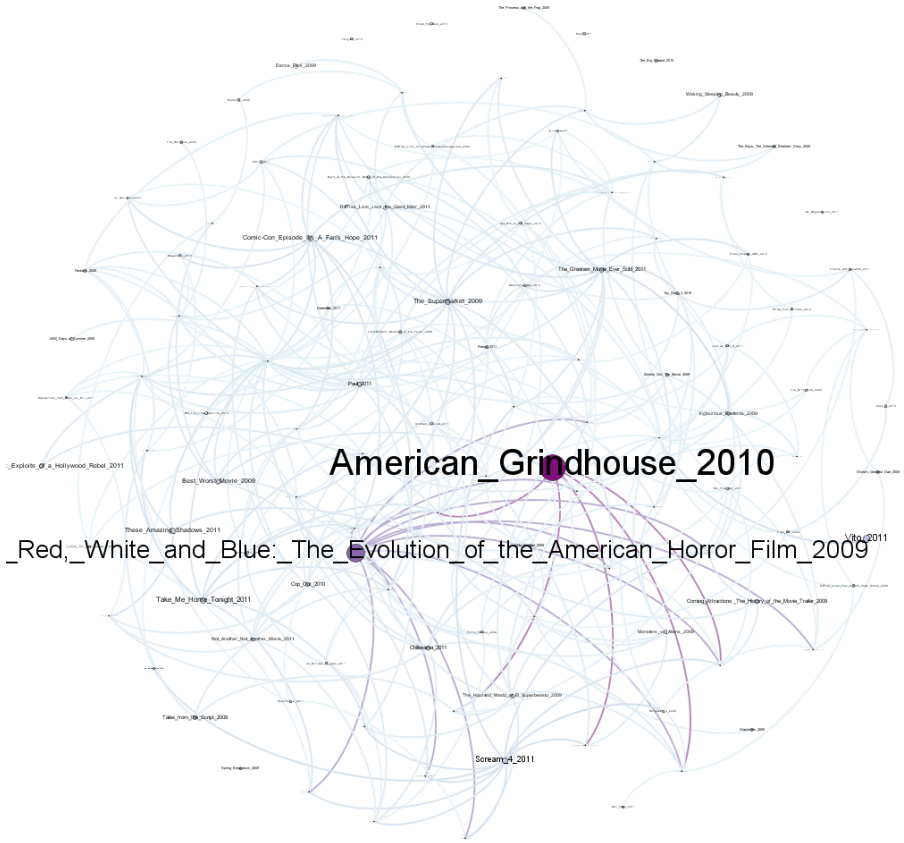
Network Statistics

In-Degree Centrality



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

Out-Degree Centrality



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

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


A long time ago in a galaxy far, far away...
STAR WARS
TWENTIETH CENTURY-FOX Presents
A LUCASFILM LTD. PRODUCTION
STAR WARS
Starring MARK HAMILL HARRISON FORD CARRIE FISHER
PETER CUSHING
and
ALEC GUINNESS
Written and Directed by GEORGE LUCAS Produced by GARY KURTZ Music by JOHN WILLIAMS
PANAVISION® PRINTS BY DE LUXE® TECHNICOLOR®
PG
DOLBY SYSTEM


The WIZARD of OZ

Hypotheses

Hypotheses

- Are highly-rated movies more likely to be cited?
- Do popular movies get more citations?
- Do critically acclaimed movies get cited more?
- Are movies from top production houses referenced more frequently?
- Are movies tend to cite others from the same production house?
- Are movies likely to cite others from the same genre?

ERGM Model

ERGM Model: Effects of Ratings and Popularity

Effects	Estimate	Significance	Interpretation
edges	-8.525	***	Movie references are not likely to be random
transitive	1.337	***	Movies tend to cite movies referened by their citations
IMDB Rating of Cited Movie	0.0524	***	Highly rated movies tend to be cited more
IMDB Votes of Cited Movie	0.000002	***	Popular movies tend to be cited more
Tomatometer of Cited Movie	0.0111	***	Critical acclaim increases the likelihood for a movie to be cited

ERGM Model: Effects of Production House

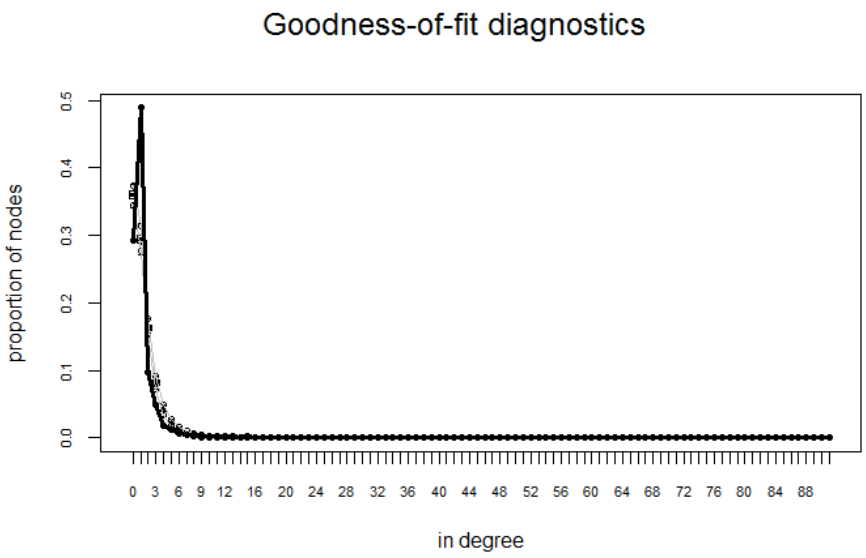
Effects	Estimate	Significance	Interpretation
References within same Production House	0.6289	***	Movies are more likely to reference others from the same production house
Citations for 20th Century Fox Movies	0.2949	***	20th Centry Fox movies are more likely to be cited
Citations for Paramount Pictures Movies	0.3051	***	Paramount Pictures movies are more likely to be cited
Citations for Universal Pictures Movies	0.2951	***	Universal Pictures movies are more likely to be cited
Citations for Warner Bros. Pictures Movies	0.2577	***	Warner Bros. Pictures movies are more likely to be cited

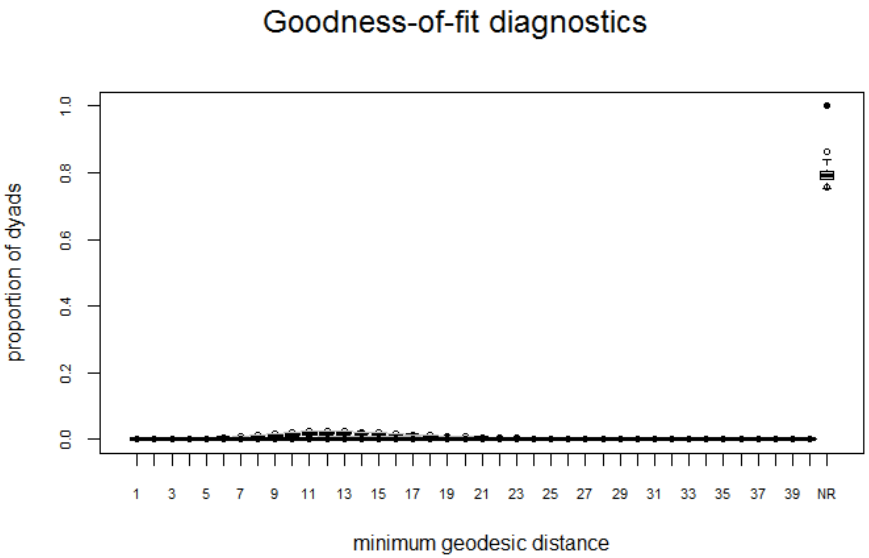
ERGM Model: Effects of Movie Genres

Effects	Estimate	Significance	Interpretation
References within Comedy Genre	-0.2309	***	Comedy movies are less likely to cite other comedies
References within Crime Genre	0.5167	***	Crime movies tend to reference movies from the same genre
References within Romance Genre	0.2455	***	Romance movies tend to reference movies from the same genre
References within Sci-Fi Genre	0.1299	**	Sci-fi movies tend to reference movies from the same genre
References within Documentaries	-1.2210	***	Documentaries are not likely to reference other documentaries

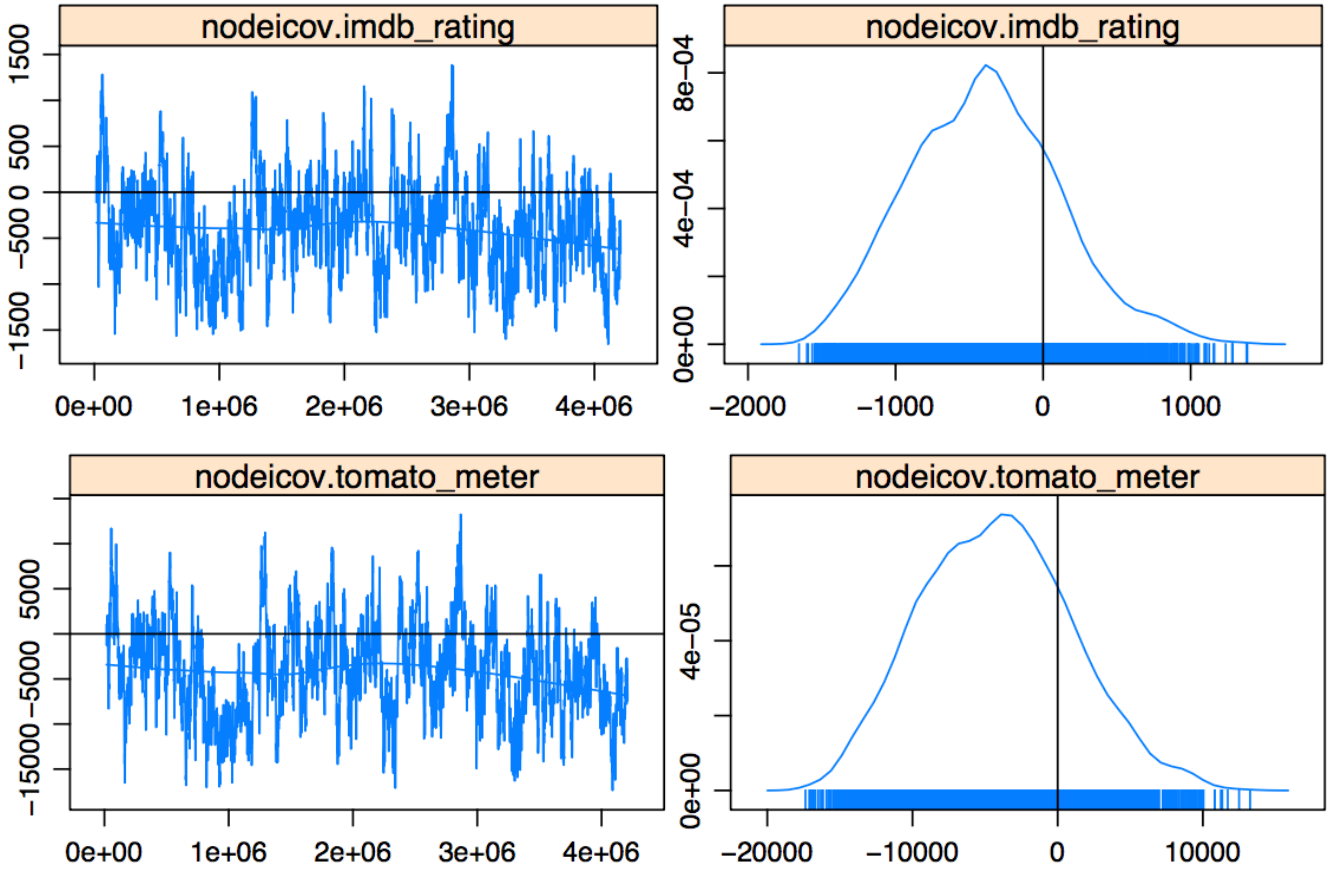
Model Diagnostics

Goodness of Fit





MCMC Diagnostics



Thank you!

Appendix

```
==========================
Summary of model fit
==========================

Formula:   citationsNetwork ~ edges + transitive + nodeicov("imdb_rating") +
    nodeicov("imdb_votes") + nodeicov("tomato_meter") + nodematch("production",
    keep = c(1:5, 7:11)) + nodeifactor("production", base = c(2:6,
    8, 11)) + nodematch("comedy", keep = 1) + nodematch("crime",
    keep = 1) + nodematch("romance", keep = 1) + nodematch("scifi",
    keep = 1) + nodematch("docum", keep = 1)

Iterations:  6 out of 20

Monte Carlo MLE Results:
                                          Estimate Std. Error MCMC %  p-value
edges                                   -8.525e+00  1.308e-01      0  < 1e-04 ***
transitive                               1.337e+00  2.241e-01      0  < 1e-04 ***
nodeicov.imdb_rating                     5.237e-02  1.371e-02      0 0.000133 ***
nodeicov.imdb_votes                      2.080e-06  1.003e-07      0  < 1e-04 ***
nodeicov.tomato_meter                    1.107e-02  6.231e-04      0  < 1e-04 ***
nodematch.production                     6.289e-01  1.060e-01      1  < 1e-04 ***
nodeifactor.production.20th Century Fox  2.949e-01  7.700e-02      0 0.000128 ***
nodeifactor.production.Paramount Pictures 3.051e-01  6.424e-02      0  < 1e-04 ***
nodeifactor.production.Universal Pictures 2.951e-01  7.959e-02      0 0.000209 ***
nodeifactor.production.Warner Bros. Pictures 2.577e-01  6.462e-02      0  < 1e-04 ***
nodematch.comedy                        -2.309e-01  3.092e-02      0  < 1e-04 ***
nodematch.crime                          5.167e-01  4.999e-02      0  < 1e-04 ***
nodematch.romance                        2.455e-01  4.759e-02      0  < 1e-04 ***
nodematch.scifi                          1.299e-01  4.605e-02      0 0.004805 **
nodematch.docum                         -1.221e+00  4.557e-02      0  < 1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance: 14058467  on 10141040  degrees of freedom
 Residual Deviance:    76232  on 10141025  degrees of freedom

AIC: 76262    BIC: 76474    (Smaller is better.)
```