Gender Representation in Romance Movies using Network Analysis

Dina Li Jiena (U1730712J), Wee Kim Wee School of Communication and Information

Ryan Tan Yi Wei (N2000457K), Faculty of Engineering, NUS

A final paper submitted in partial fulfilment of the requirements for

CS4264 Introduction to Network Analysis

Acknowledge

We would like to express our heartfelt appreciation and gratitude to our mentor, Professor Poong Oh, for his constant guidance and support throughout this project. This project would not have been possible without his advice.

Abstract

This study examines the importance of male and female characters in dialogue-based character networks of romance movies by measuring the centrality of characters. Betweenness and eigenvector centralities were used to determine a character's importance in the plot narrative and character world. We suggested a comparison of absolute values of centrality measures as well against typical centrality measures of random graphs generated. This helped us identify characters that were more central and possible gender-biased representation. We observed female protagonists having less importance in both the plot narrative and character world than male protagonists. The potential of this methodology lies in the ability to quantify the complex role of each character, allowing for more objective means to measure gender representation both within the movie and between movies.

Keywords: Gender representation; Romance movies; Protagonists; Characters; Character networks

Gender Representation in Romance Movies using Network Analysis

Gender representation in movies has been heavily analysed for its importance in influencing the mass culture and reflecting prevalent attitudes and beliefs in a given era (University of Minnesota Libraries Publishing, 2016), giving us insights into the role of women and men in society.

The issue of women underrepresentation in movies has been largely debated. In more recent times, we have witnessed the increase of representation of women on screen, with women comprising 34% of all speaking characters and accounting for 40% of protagonist roles in top-grossing films, achieving new historic highs (Lauzen, 2019). The rising popularity of strong female heroines in blockbuster movies such as Wonder Woman and Hunger Games have also given the impression of equal representation of women in movies. However, qualitative studies found women in movies limited to traditionally feminine roles (non-professional, homemakers, sexual gatekeepers, etc.) and are often sexualised and subordinated (Collins, 2011).

The Bechdel test is a simple litmus test by cartoonist Alison Bechdel measuring the representation of women in fiction. Based on only three conditions, the test asks whether a movie has at least two named women characters who speak to each other about something other than a man. Despite the seemingly low requirements, movies do not often pass the test. 57% of the hundred most widely distributed movies in the United States between 2000 to 2009 do not pass the Bechdel Test (Lindler & Schulting, 2017). The analysis of the 10 Oscar-winning movies for the original screenplay in 2018 revealed that only one of the movies could pass the Bechdel Test (Ryzik, 2018).

While increased representation is crucial to equal portrayal of women, it is necessary to look at how women are being portrayed and the agency of their roles in movies. Our study aims to quantify the importance of movie characters by using dialogue-based character interaction networks to determine their power in the narrative and character world. Such a measure allows for the exploration of vocal and relational disempowerment or empowerment based on the positioning of characters within the structure of movie narratives. Without delving into the specifics of context which is often subjective and time-consuming, we can conduct objective comparisons of gender representation across movies.

Method

Networks Studied

We studied a total of 11 character networks for each one of the chosen romance movies (Appendix 1):

1. Annie Hall (1977)
2. Princess Bride (1987)
3. Romeo and Juliet (1996)
4. Titanic (1997)
5. Runaway Bride (1999)
6. Eternal Sunshine of the Spotless Mind (2004)
7. 500 Days of Summer (2009)
8. Crazy Stupid Love (2011)
9. Silver Linings Playbook (2012)

10. Theory of Everything (2014)

11. La La Land (2016)

Nodes in the network represented individual characters and the edges represented utterance between two characters. Nodes are connected as long as there is a dialogue between the characters. The networks are undirected and unweighted with all self-loops removed as these have minimal influence on the centrality measures we will be using.

Data Processing

List of Variable Categories	
Node List	**Edge List**
nid (Node ID)	mid (Movie ID)
entity (Name of Character)	sid (Scene ID)
freq (Frequency of Utterance)	node_i (CharacterX)
scene_count	node_j (CharacterY)
gender	utterance
	uid (Dialogue)

Table 1: List of categories within node list/ edge list. Gender was labelled manually.

Labelling

Through cross-referencing with multiple online sources (Fandom, 2021; IMDb, 2021), the gender labels were added manually into the .csv node list (Table 1). Other character features such as age and profession were considered but were not included because of the lack of available data.

Cleaning the Network

We attempted to remove movie extras as we were unable to classify them. We determined extras as characters having 2 or fewer scene counts as we assumed their existence in the movie insignificant. However, this resulted in disconnected networks (Figure 1) due to their unexpectedly high betweenness. This led us to reconsider our criteria for determining "importance" since even minor characters with few utterances and scene counts had a role in connecting the character network.

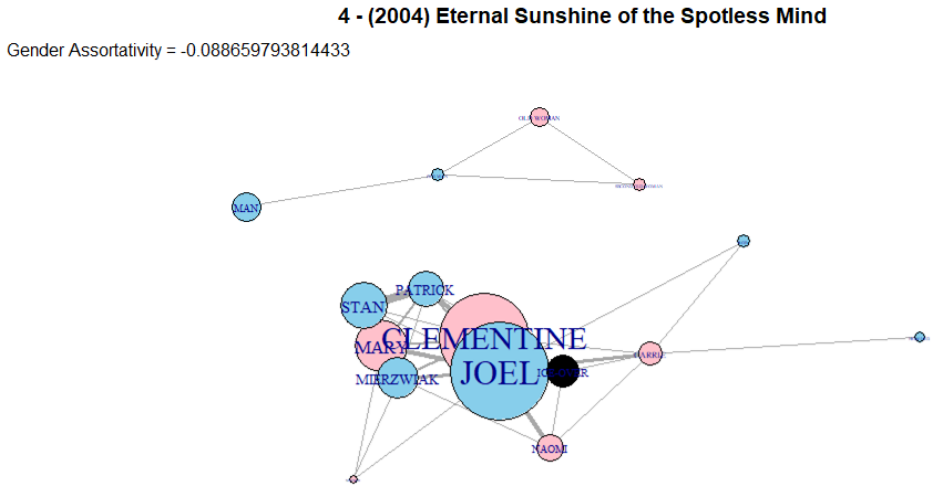


Figure 1: Example of a disconnected Network after characters with 2 or less scene count were filtered

Ultimately we chose to filter out characters if their gender data could not be found. Computationally, lack of information on gender is seen as a label by itself. However, these characters added no value to our analysis of gender representation.

Data Analysis

Techniques and measures

A challenge in our research comes from establishing a means to compare between and within character networks in a quantitative way. The following measures and tools were used:

1. Community Detection - used to identify clusters or cliques formed by characters in the network.
2. Assortative Coefficient - used to measure the tendency in which two characters will interact due to similarity in characteristics.
3. Centrality Measures - betweenness and eigenvector centrality were used to understand a character's importance in the movie.
4. Random Graphs - generated as a means to understand more objectively if the centrality measures of a character should be considered high or low (Erdős & Rényi, 1960). By comparing the character with a set number of random Graph with the same degree sequence, it can be determined if the character's value is higher or lower than typically expected.

Approach

The project analyses gender representation on 3 levels; Community (Macro), Characters (Meso) and Protagonists (Micro).

1. Community (Macro)

Community detection was used to understand how communities in all 11 character networks tend to form organically. The assortativity coefficient of gender was computed to understand if gender influences how communities were organised.

2. Characters (Meso)

   Three films were compared - 500 Days of Summer, Eternal Sunshine of the Spotless Mind, Annie Hall. They share a similar romance film narrative which is the presence of a Manic Pixie Dream Girl (Rabin, 2007); it can be expected that the character networks share very similar characteristics. As such, through comparing the distribution of relative centrality measures (random graphs) of all male and female characters in each of the movies, the differences might reveal certain key characteristics of the network which hints at a biased representation of gender amongst the characters.

3. Protagonists (Micro)

   Male and female protagonists were defined by characters with the highest utterance count. The absolute centrality measures (betweenness and eigenvector) of the male and female were compared. In romance films, the narrative is typically centered around the male and female protagonists. Since centrality measures broadly describe the *importance* and the *role* of the protagonists, it was hypothesised that gender representation within the film's narrative could be inferred through their centrality measures. This could help frame the unequal representation of a male and female protagonist's role in the romantic relationship.

Results

Community (Macro)

Distinct Communities

All 11 networks were composed of one giant component. We observed a range of modularity values among the networks. Movies with high modularities such as Theory of Everything and Annie Hall have distinct communities that indicate different parts of the narrative (Figure 2,3). These communities can be used to identify subplots involving the same characters based on spatial or temporal conditions.

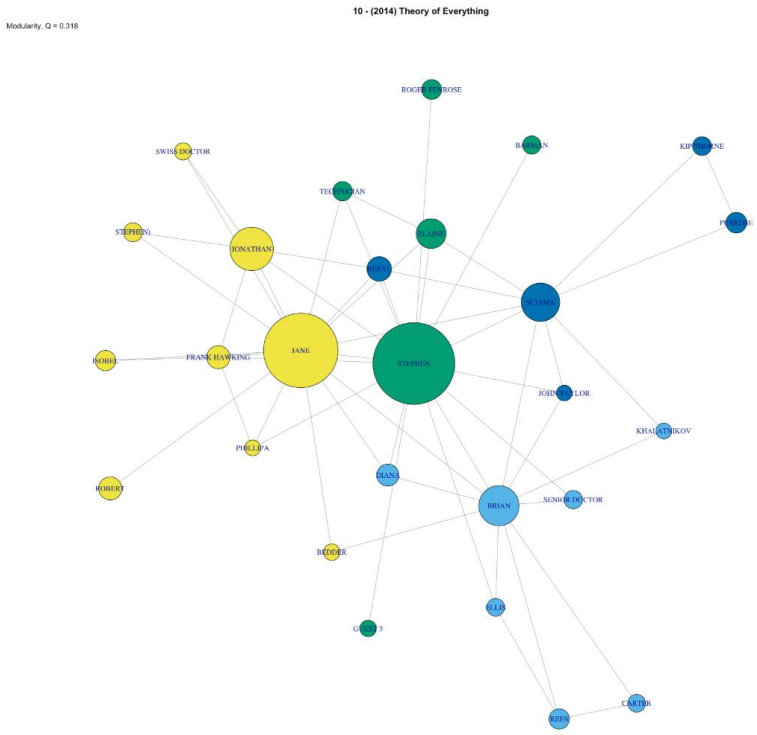


Figure 2: Character Network of *Theory of Everything* (2014)

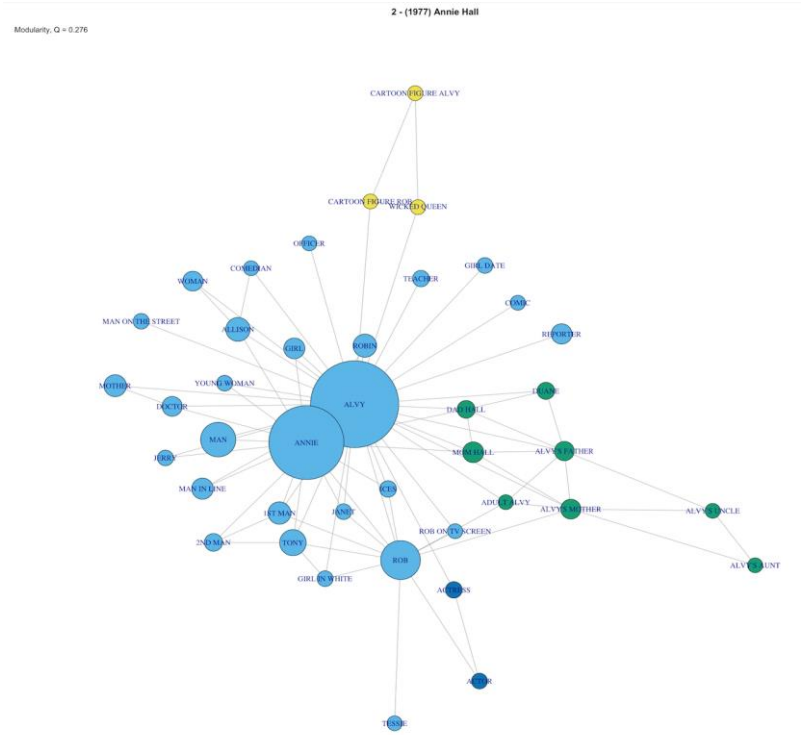


Figure 3: Character Network of *Annie Hall* (1977)

Increasing Gender Assortative/Disassortative

Through our analysis, the studied romance movies were becoming increasingly gender assortative and disassortative over time (Figure 4). Plotting the absolute values of assortativity, we can observe a general increasing trend (Figure 5). This could mean that gender has an increasing influence on determining how communities are formed within romance movies.

Figure 4: Gender Assortativity plotted against Years

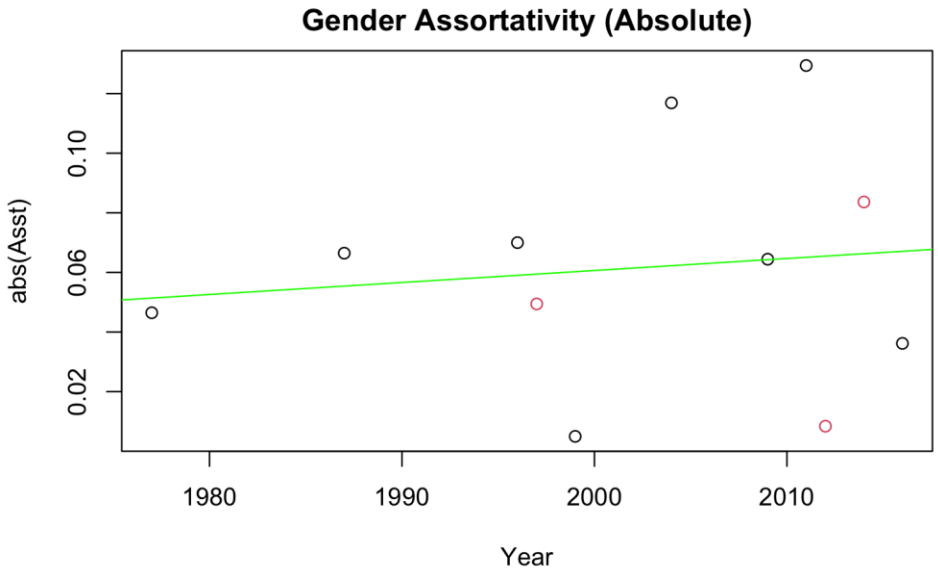


Figure 5: Gender Assortativity (absolute) plotted against Years

Centrality Measures of Character (Meso)

We analysed three movies that shared similar narratives featuring the Manic Pixie Dream Girl trope (Figure 6). By plotting the distribution of the relative centrality measures (against random graphs) of all male and female characters (Appendix 2) we were able to identify an anomaly in Annie Hall's betweenness centrality plot.

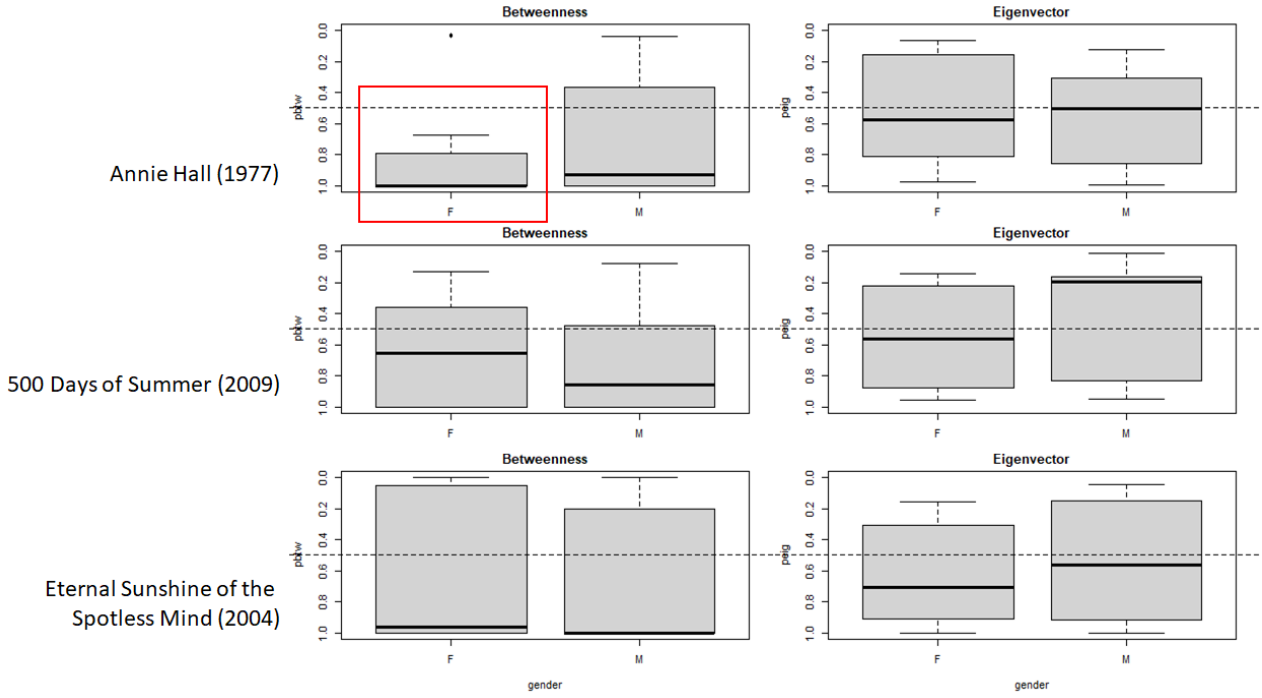


Figure 6: Distribution of Relative Centrality Measures for each Movie

Female characters, in particular, took on an unusually small range of p values between 0.8 to 1.0 which meant they tend to have lower betweenness centralities as compared to typical betweenness centralities. In fact, the movie's director, Woody Allen, had often been accused of his sexist portrayal of women as neurotics, shrewds and prostitutes (Itzkoff, 2013). This finding aligned with the criticisms of Woody Allen's movies and could be used to identify the biased representation of the genders.

Centrality Measures of Female and Male Protagonists (Micro)

We calculated betweenness centrality and eigenvector centrality measures of the female and male protagonists of each movie to determine their importance and to make claims about their relative positions within movies.

Betweenness centrality measures the amount of influence a node has over the flow of information in a network. In the context of a movie, betweenness centrality identifies characters that are important for their bridging role in connecting different parts of a movie's plot (Labatut & Bost, 2019). It is typically used to understand a character's importance in the plot narrative.

Eigenvector centrality measures the influence of a node based on the importance of its neighbours. In the context of a movie, eigenvector centrality identifies characters that are important for their connection to central characters in the movie world. This measure is often lower for characters who are connected to peripheral characters than for those who are connected to a few core characters (Algee-Hewitt, 2017). It is typically used to understand a character's importance in the character world (Jones et al., 2020).

Comparison of Absolute Values

The analysis of absolute values of betweenness and eigenvector centrality revealed that female protagonists tend to have lower eigenvector centrality than male protagonists (Appendix 3). Out of the 11 romance movies, nine movies reflected this result. As such, female protagonists are positioned to be less influential in romance movies as they are often connected to peripheral characters rather than core characters and are not as central as their male counterparts.

In terms of betweenness centrality, six out of the 11 romance movies had female protagonists with a lower score. Female protagonists had approximately an equal chance of

fairing higher or lower than their male counterparts and can be said to be equally important in acting as a narrative bridge.

We constructed a comparison matrix to further evaluate the relationship between betweenness and eigenvector centralities of female protagonists as compared to male protagonists (Figure 7).

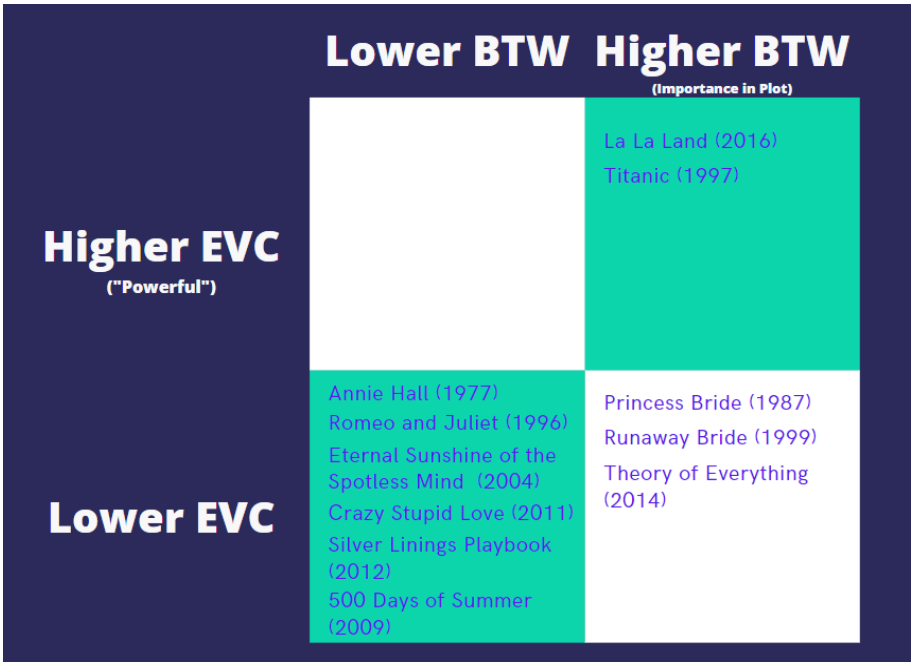


Figure 7: Matrix comparing centrality measures of the Female Protagonist against the Male Protagonist

We observed that a large portion of the movies had female protagonists with lower betweenness and eigenvector scores. These female protagonists are less important in the narrative as well as the movie world and often serve as supporting characters to the male protagonist. Movies with the Manic Pixie Dream Girl archetype (Schwyzer, 2013) can be observed in this quadrant; Annie Hall, Eternal Sunshine of the Spotless Mind, 500 Days of Summer. In the opposite quadrant, female protagonists with higher betweenness and eigenvector scores are less common.

No movies fell under the quadrant with lower betweenness and higher eigenvector scores, which would have meant a more important female protagonist with less narrative importance. Some movies could be found in the quadrant with higher betweenness and lower eigenvector scores where female protagonists were less important in the movie world but were narratively more important.

Discussion

The purpose of the study was to examine how gender is represented in romance movies. From our results, we observed a certain trend in the portrayal of male and female protagonists. Female protagonists tend to have lower betweenness and eigenvector centralities and were thus positioned to have less importance in the plot narrative and character world. Despite their seemingly important roles as protagonists, female protagonists were still less central and had less influence in the character network than male protagonists. Comparing distributions of centrality measures against typical values of random graphs was also useful in helping us identify biased gender representation that might exist in movies. Directors and producers could make use of such comparisons between female and male protagonists or all characters to determine if there is underlying skewed gender representation in their scripts.

Given the small sample size of movies analysed, we were unable to draw conclusions that can be generalised to all romance movies. Future research could look at analysing more movies and further categorising the romance movies into their respective subcategories for a more accurate comparison. Analysis of the text or utterance could also give us a better framing of the portrayal of female and male characters and their typical traits.

References

Algee-Hewitt, M. (2017). Distributed Character: Quantitative Models of the English Stage, 1550–1900. *New Literary History*, *48*(4), 751–782. https://doi.org/10.1353/nlh.2017.0038

Collins, R.L. (2011). Content Analysis of Gender Roles in Media: Where Are We Now and

Erdős, P., & Rényi, A. (1960). On the Evolution of Random Graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, 17(61).

Fandom. (2021). *Moviepedia: Your Film Guide*. Moviepedia. Retrieved from movies.fandom.com/wiki/

IMDb. (2021). *IMDb*. IMDb. https://www.imdb.com/

Itzkoff, D. (2013, July 17). *Woody Allen's Distinctive Female Characters*. The New York Times. Retrieved from https://www.nytimes.com/2013/07/21/movies/woody-allens-distinctive-female-characters.html

Jones, P., Quinn, E., & Koskinen, J. (2020). Measuring centrality in film narratives using dynamic character interaction networks. *Social Networks*, 63, 21–37. https://doi.org/10.1016/j.socnet.2020.03.003

Labatut, V., & Bost, X. (2019). Extraction and Analysis of Fictional Character Networks. *ACM Computing Surveys*, *52*(5), 1–40. https://doi.org/10.1145/3344548

Lauzen, M.M. (2020). *It's a Man's (Celluloid) World: Portrayals of Female Characters in the*

Lindner, A. M., & Schulting, Z. (2017). How Movies with a Female Presence Fare with Critics. *Socius: Sociological Research for a Dynamic World, 3*, 237802311772763. https://doi.org/10.1177/2378023117727636

Media and Culture. Pressbooks. Retrieved from https://open.lib.umn.edu/mediaandculture/chapter/8-3-movies-and-culture/

Rabin, N. (2007). *The Bataan Death March of Whimsy Case File #1: Elizabethtown*. The A.V. Club.

RDocumentation. (2019). *list.files: List the Files in a Directory/Folder*. Base (Version 3.6.2). https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/list.files

RDocumentation. (2019). nest: *Nest repeated values in a list-variable*. Tidyr. https://www.rdocumentation.org/packages/tidyr/versions/0.8.3/topics/nest

Ryzik, M. (2018, May 11). *Is Your Script Gender-Balanced? Try This Test*. The New York Times. Retrieved from https://www.nytimes.com/2018/05/11/movies/is-your-script-gender-balanced-try-this-test.html

Schwyzer, H. (2013, July 9). *The Real-World Consequences of the Manic Pixie Dream Girl Cliché*. The Atlantic. Retrieved from https://www.theatlantic.com/sexes/archive/2013/07/the-real-world-consequences-of-the-manic-pixie-dream-girl-clich-233/277645/

Top Grossing Films of 2019. San Diego State University. Retrieved from https://womenintvfilm.sdsu.edu/wp-content/uploads/2020/01/2019_Its_a_Mans_Celluloid_World_Report_REV.pdf

University of Minnesota Libraries Publishing. (2016). *8.3 Movies and Culture – Understanding*

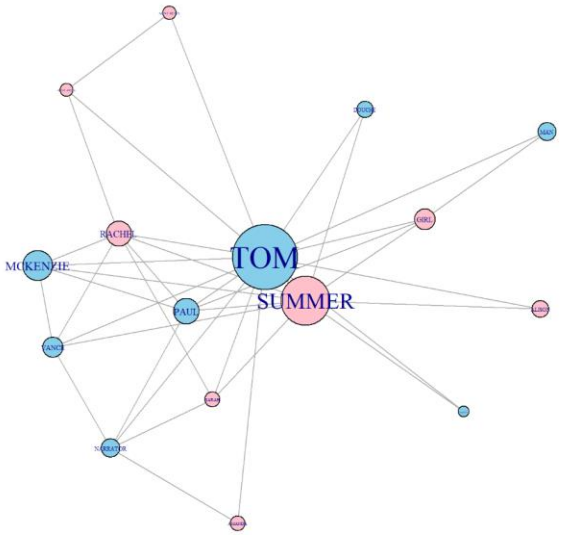
Where Should We Go?. *Sex Roles* 64, 290–298. https://doi.org/10.1007/s11199-010-9929-5

Appendix

Appendix 1 : Network Graph of All Plots

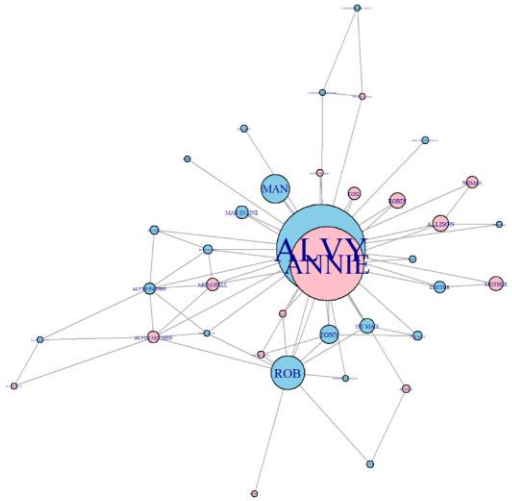
1 - (2009) 500 Days of Summer

Gender Assortativity = -0.0644257703081233



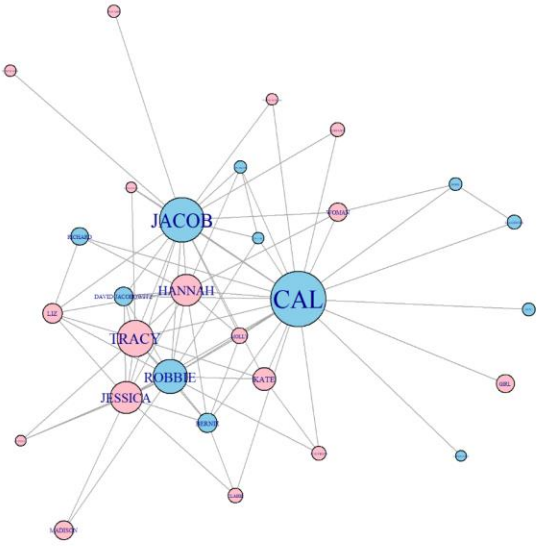
2 - (1977) Annie Hall

Gender Assortativity = -0.0464646464646464



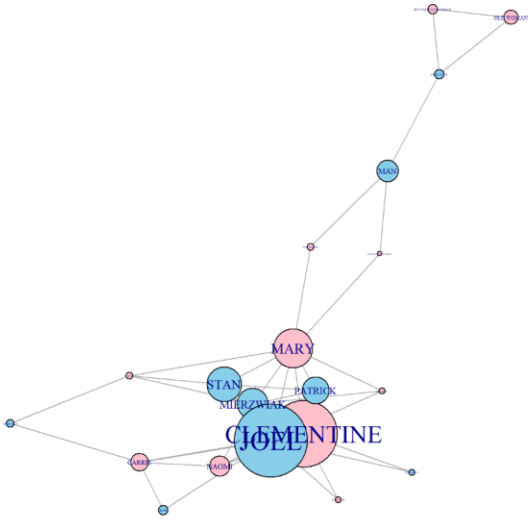
3 - (2011) Crazy, Stupid, Love

Gender Assortativity = -0.129411764705882



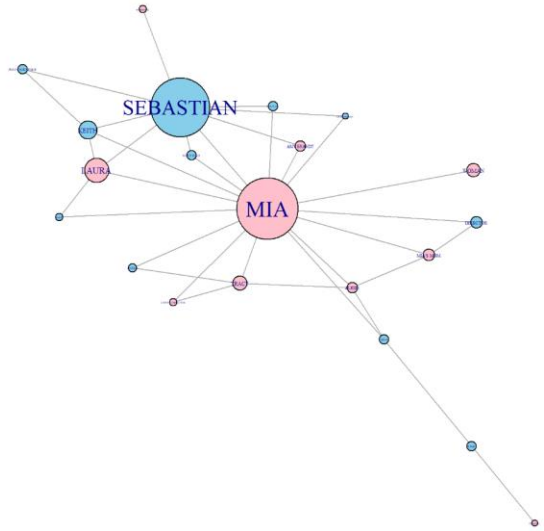
4 - (2004) Eternal Sunshine of the Spotless Mind

Gender Assortativity = -0.116883116883117

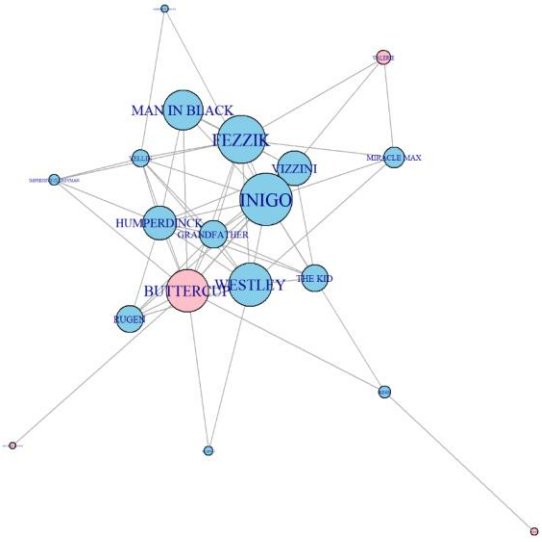


5 - (2016) La La Land

Gender Assortativity = -0.0361842105263157

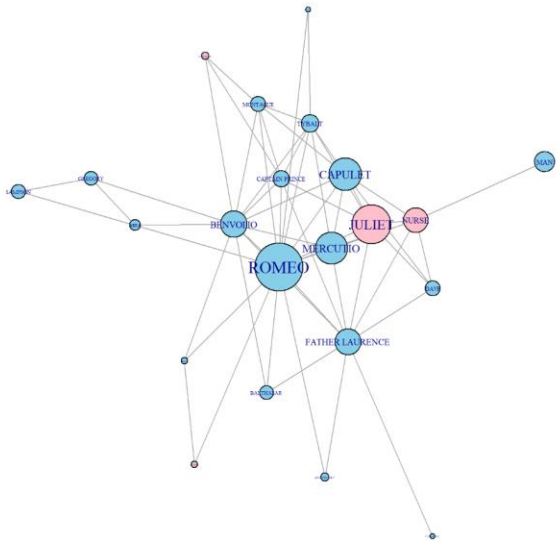


Gender Assortativity = -0.0664540138224347



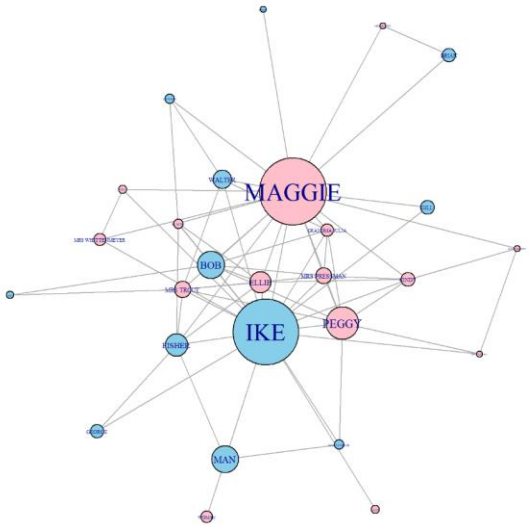
7 - (1996) Romeo & Juliet

Gender Assortativity = -0.0699945740640254



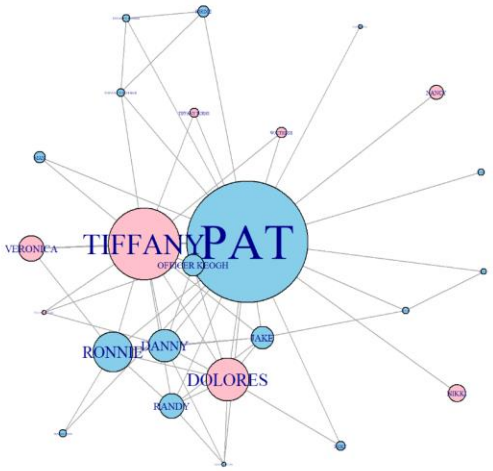
8 - (1999) Runaway Bride

Gender Assortativity = -0.00498504486540385



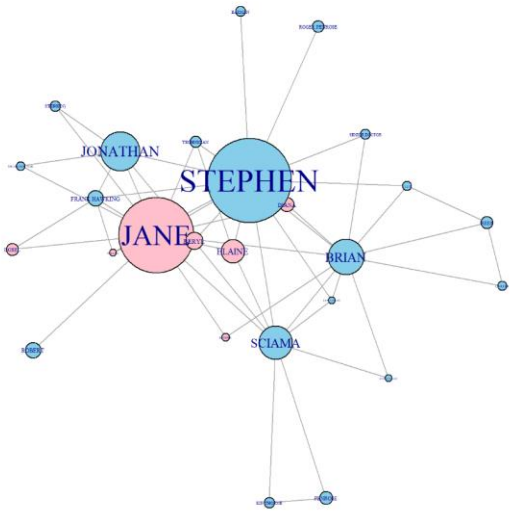
9 - (2012) Silver Linings Playbook

Gender Assortativity = 0.0084033613445379



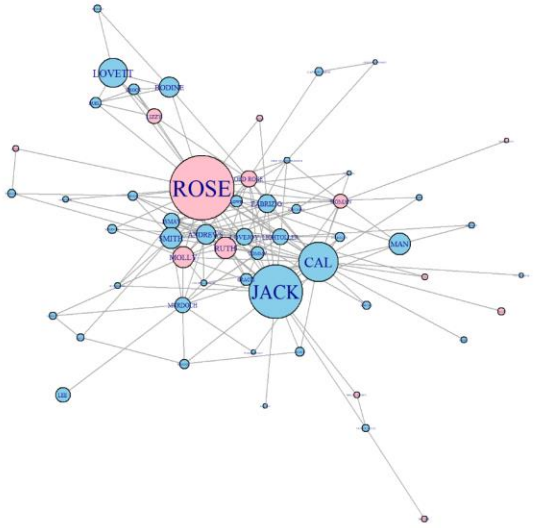
10 - (2014) Theory of Everything

Gender Assortativity = 0.0836363636363637

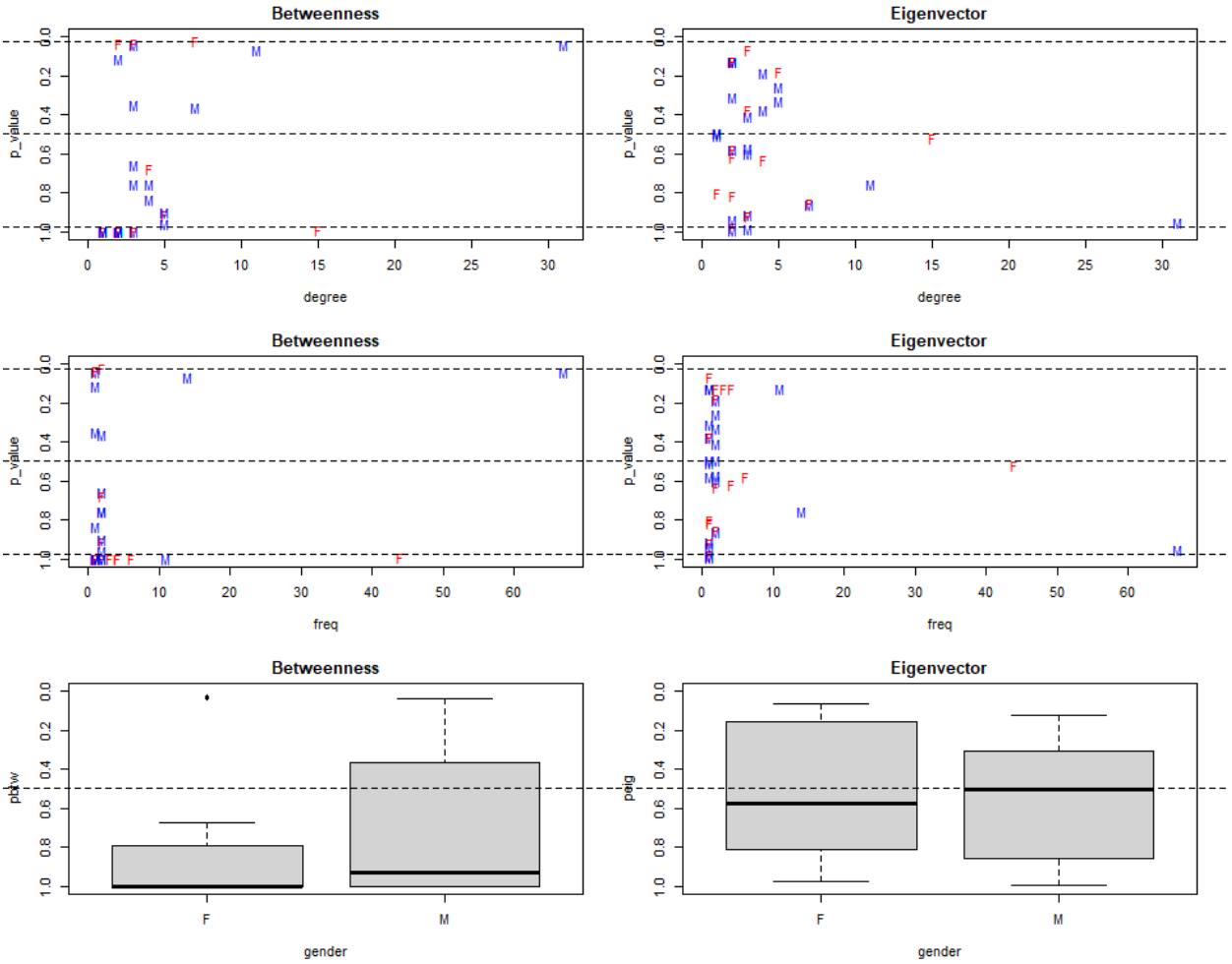


11 - (1997) Titanic

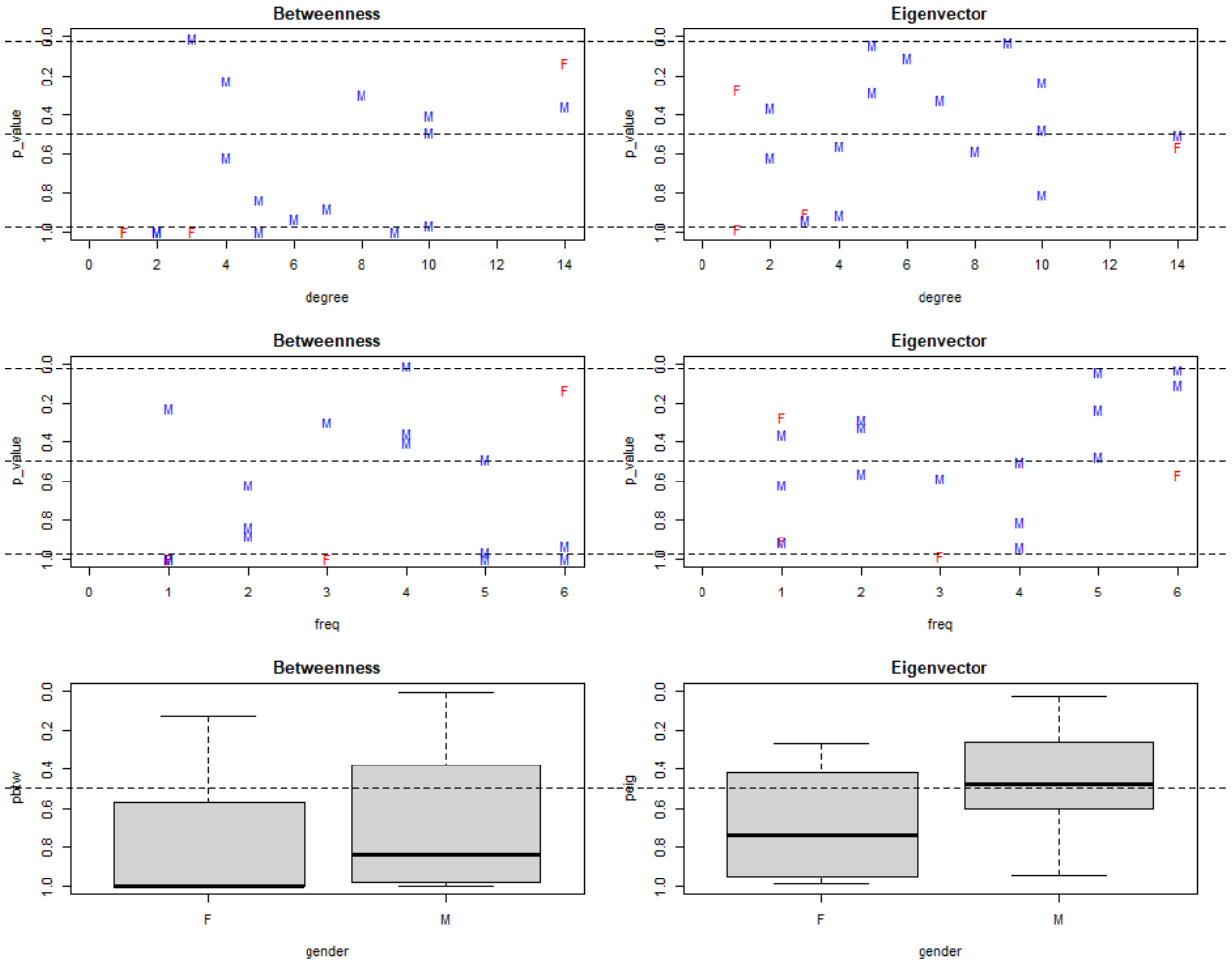
Gender Assortativity = 0.0494227994227991



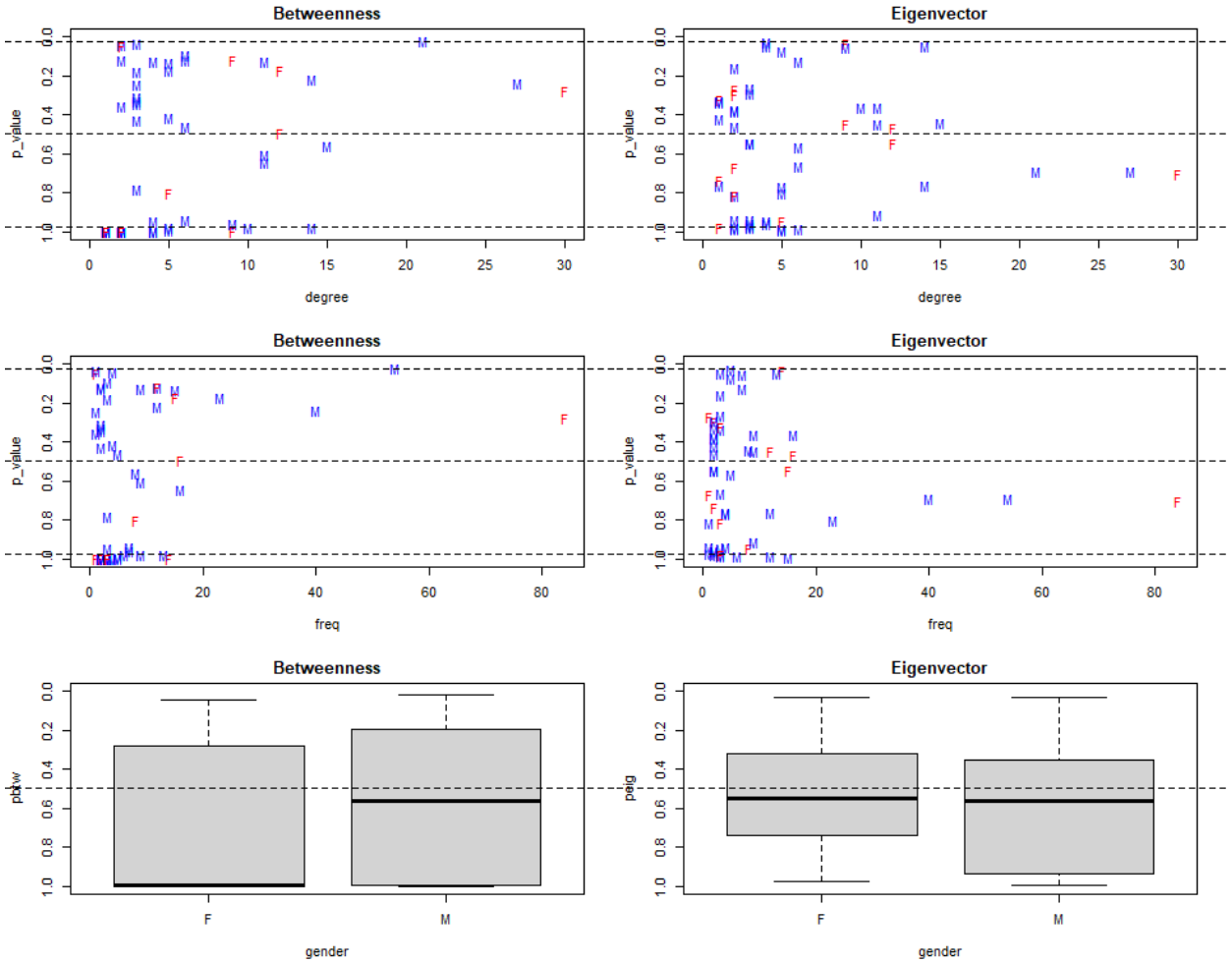
Appendix 2: Comparison of Side Characters



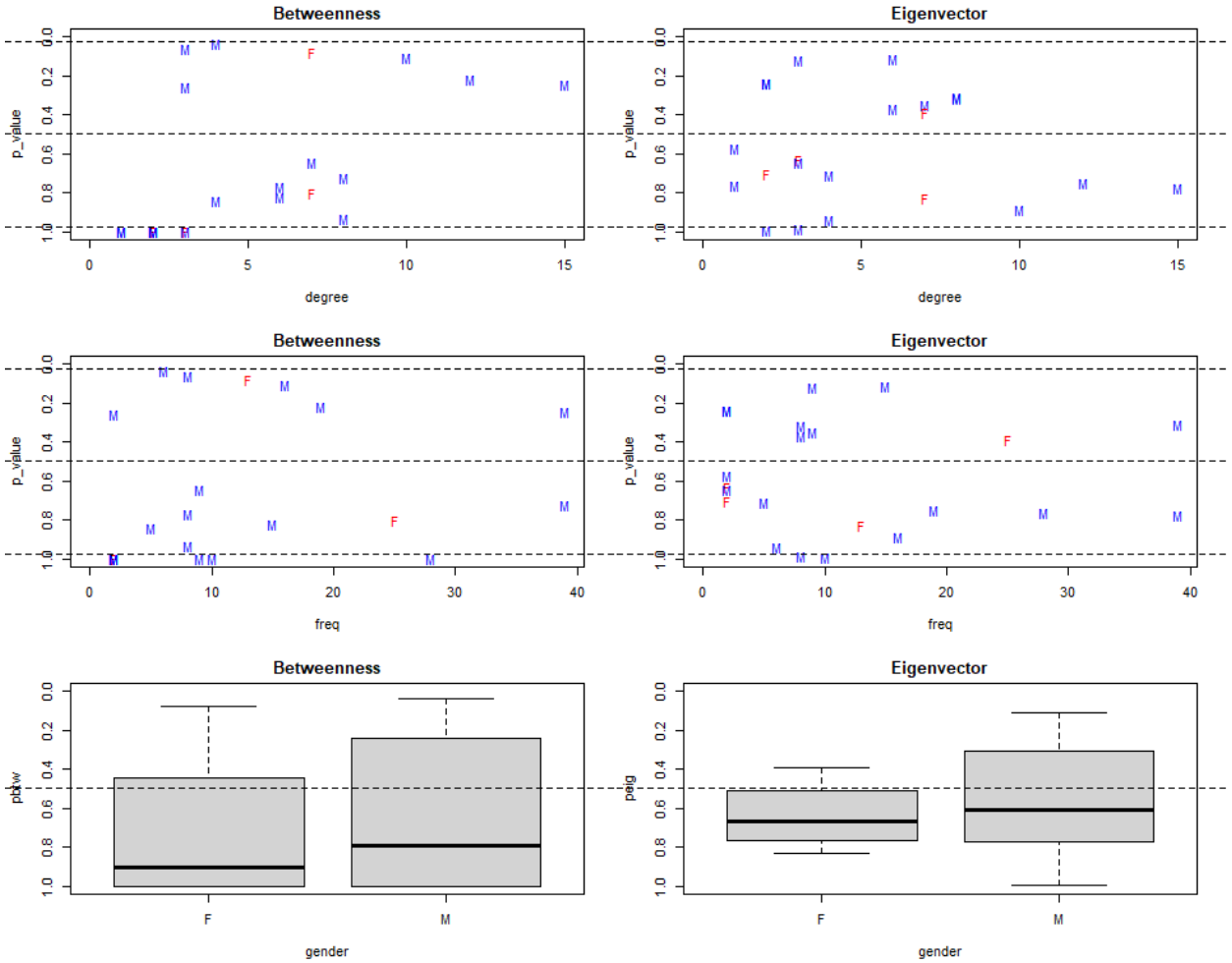
Annie Hall (1977)



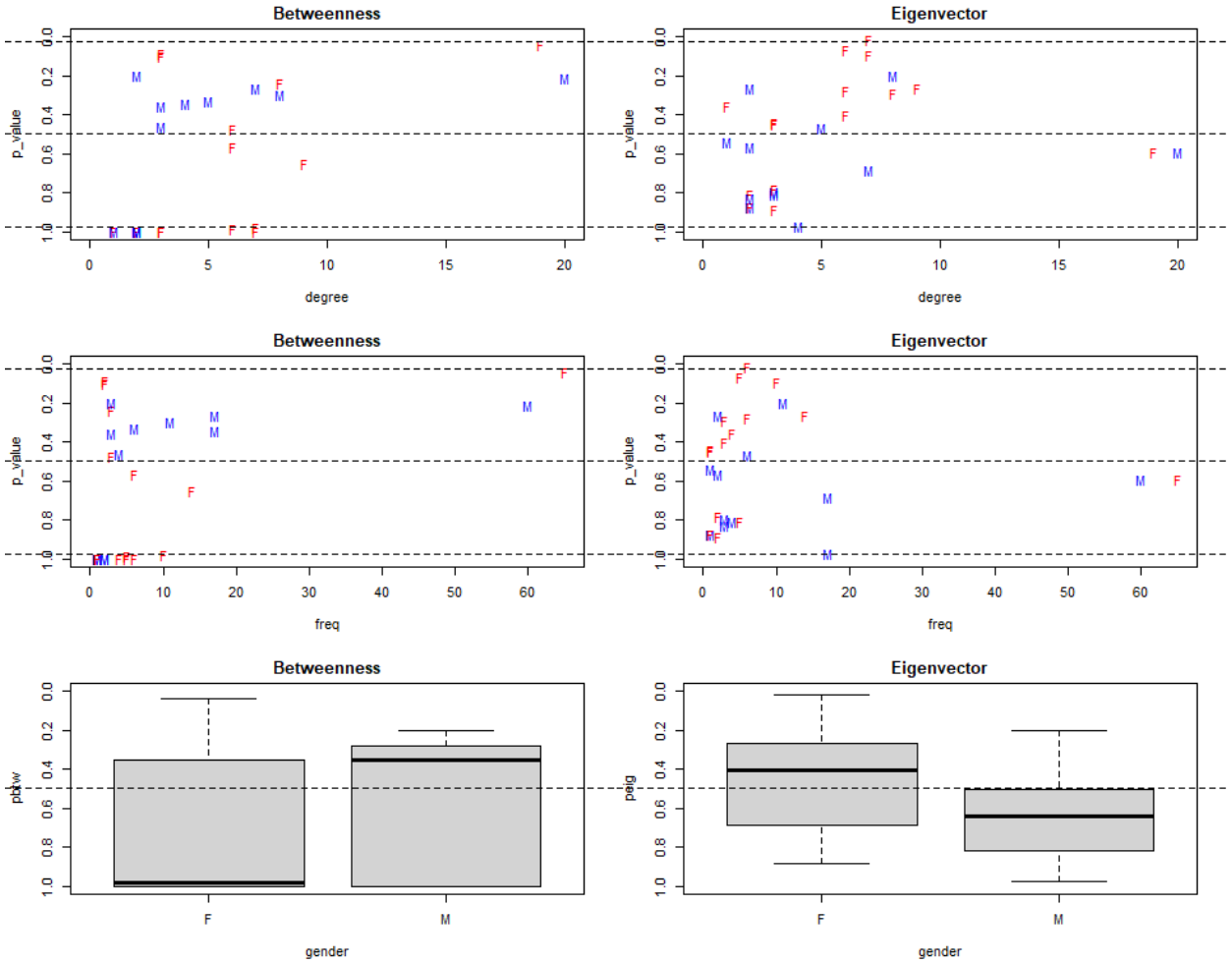
Princess Bride (1987)



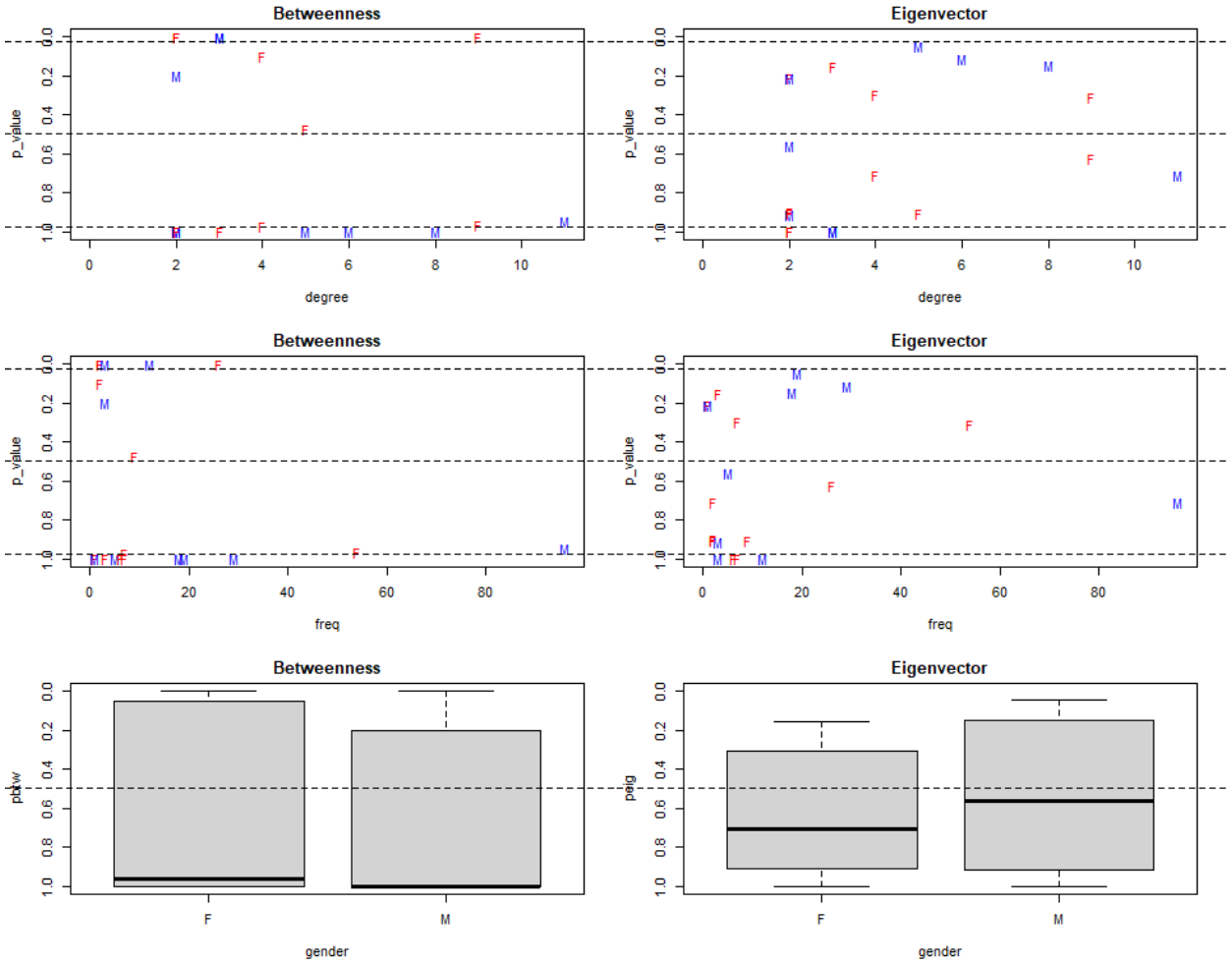
Romeo and Juliet (1996)



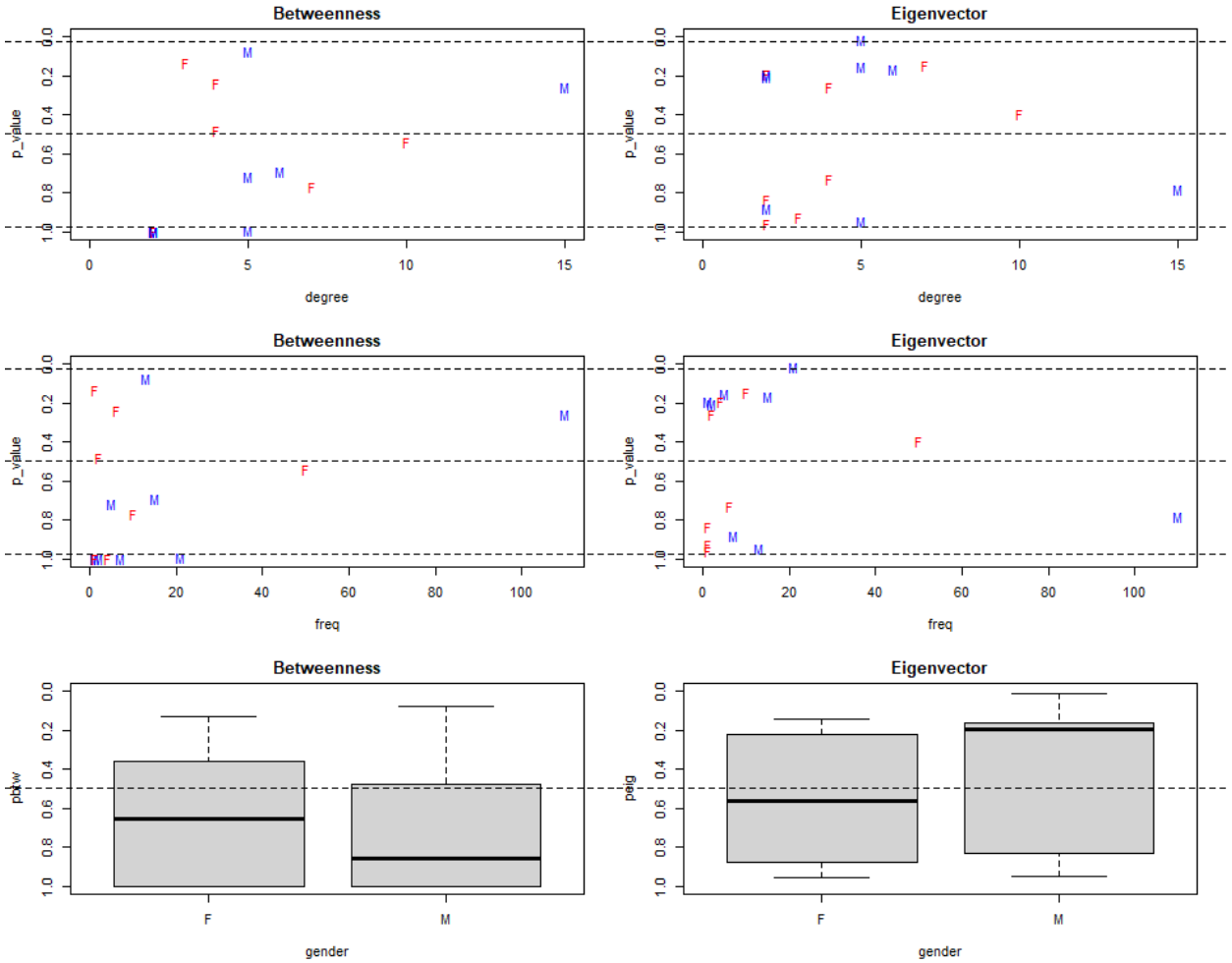
Titanic (1997)



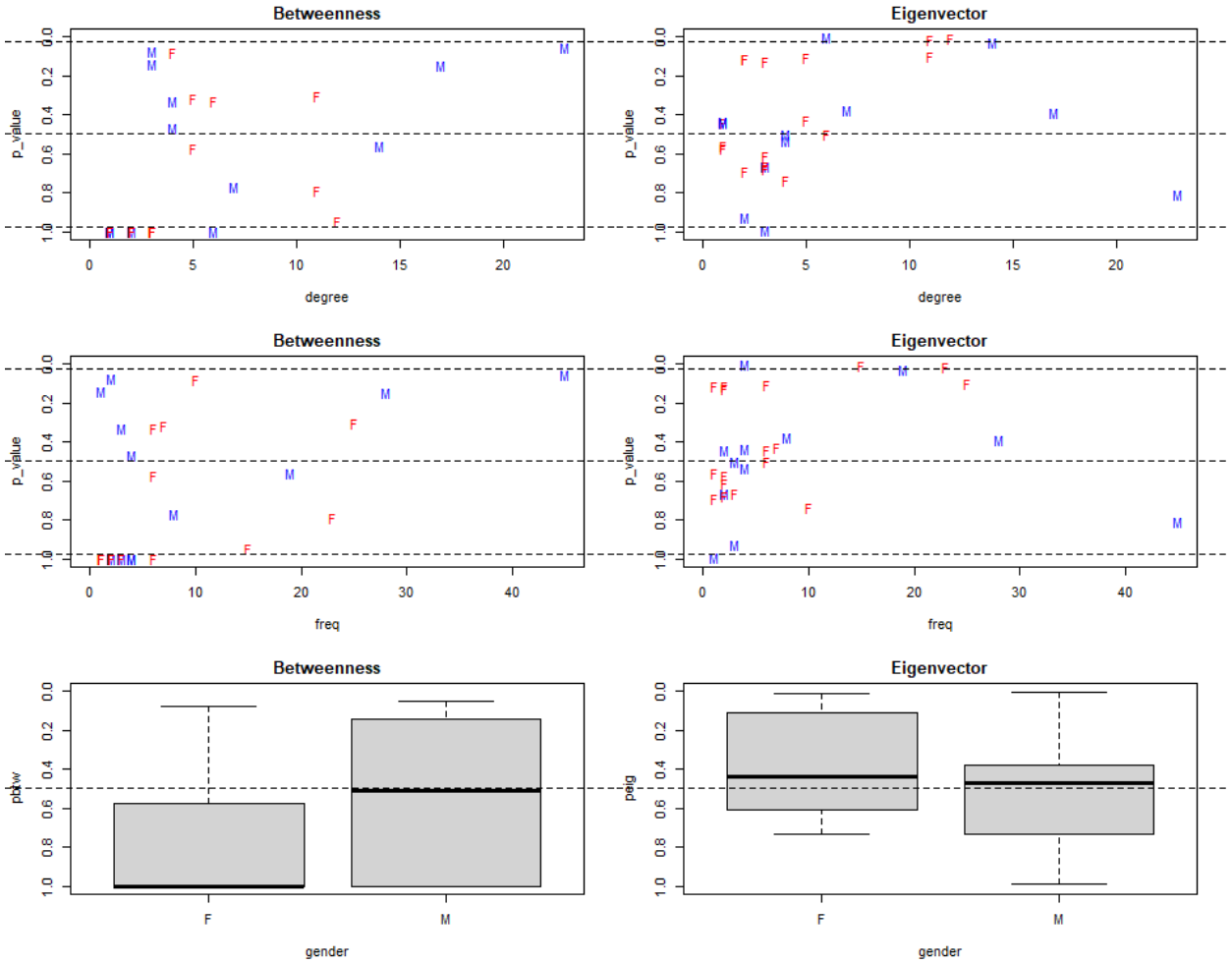
Runaway Bride (1999)



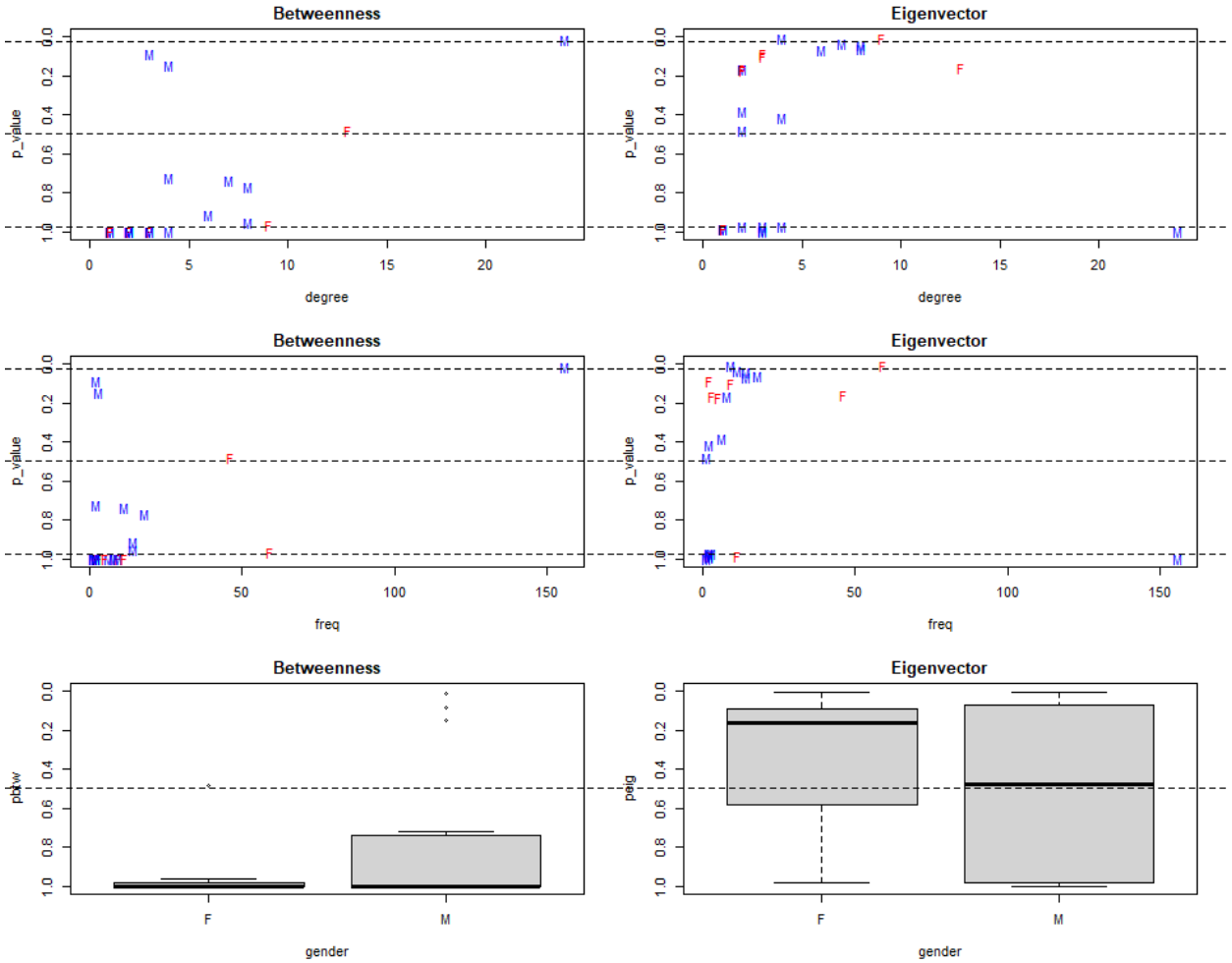
Eternal Sunshine of the Spotless Mind (2004)



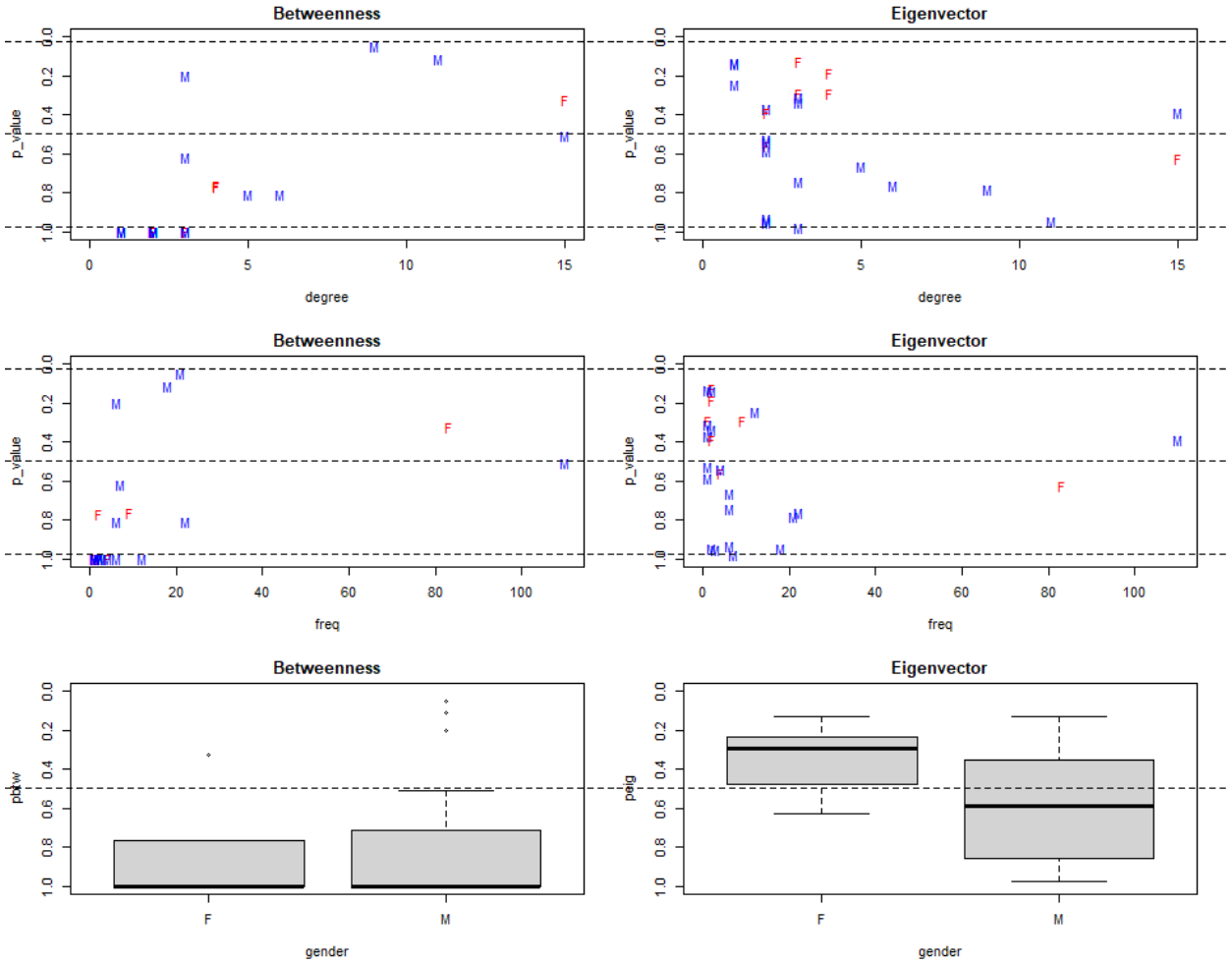
500 Days of Summer (2009)



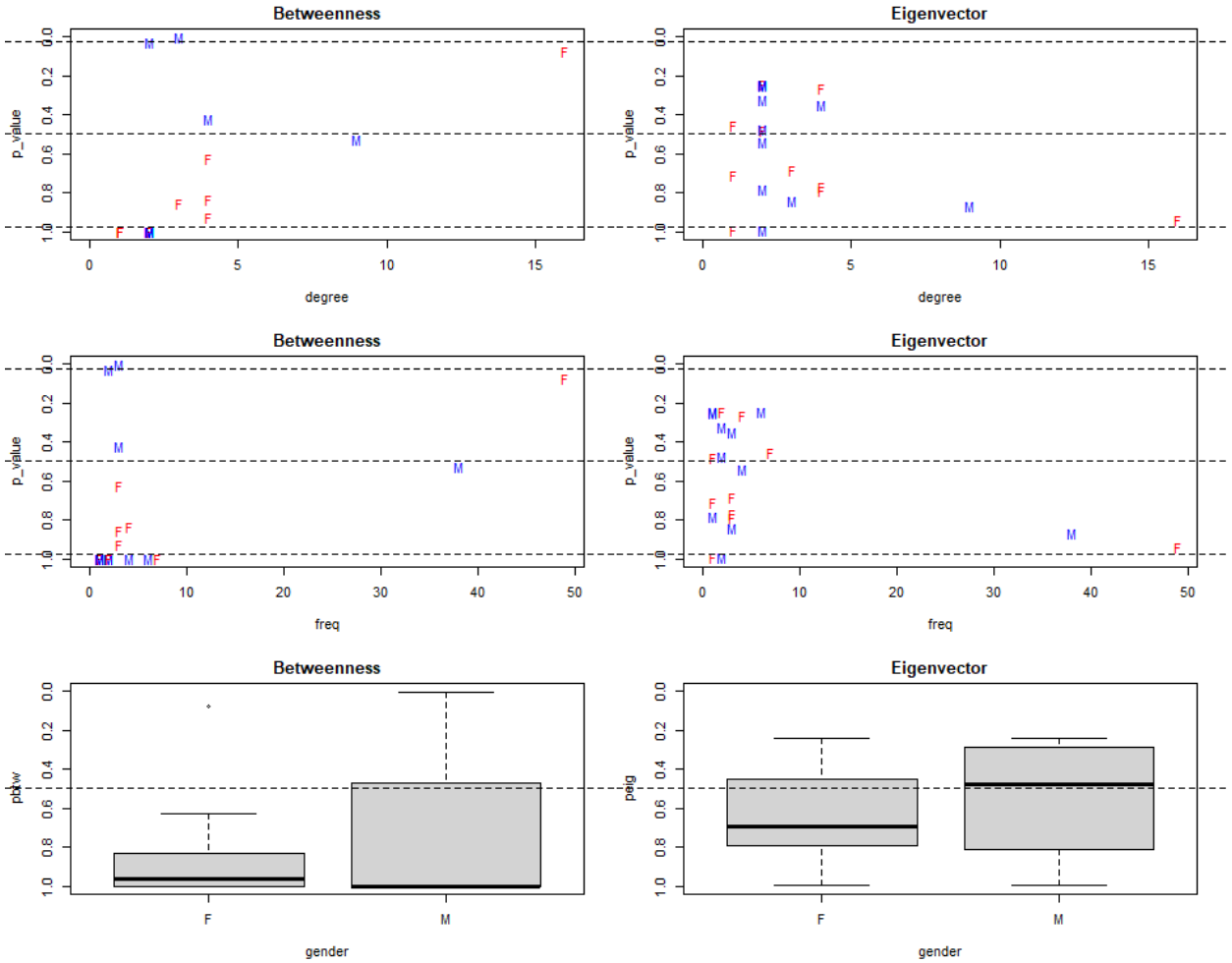
Crazy Stupid Love (2011)



Silver Linings Playbook (2012)

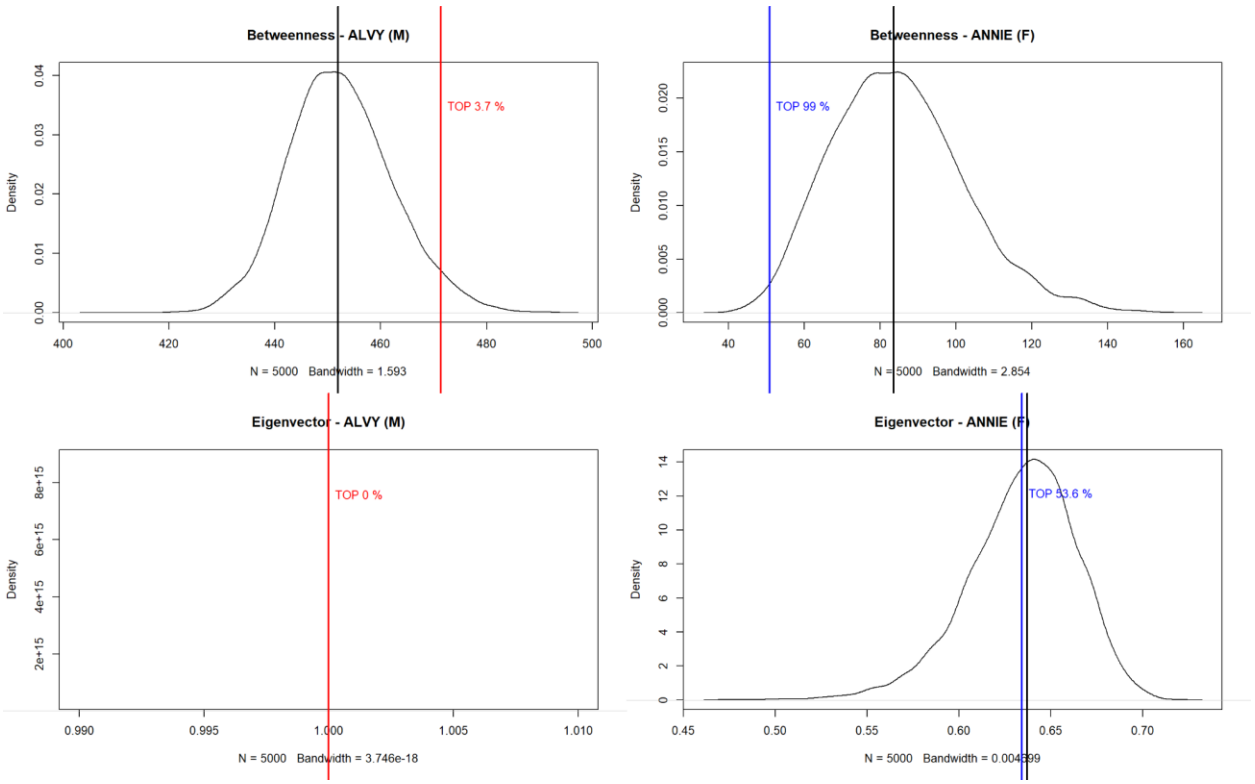


Theory of Everything (2014)

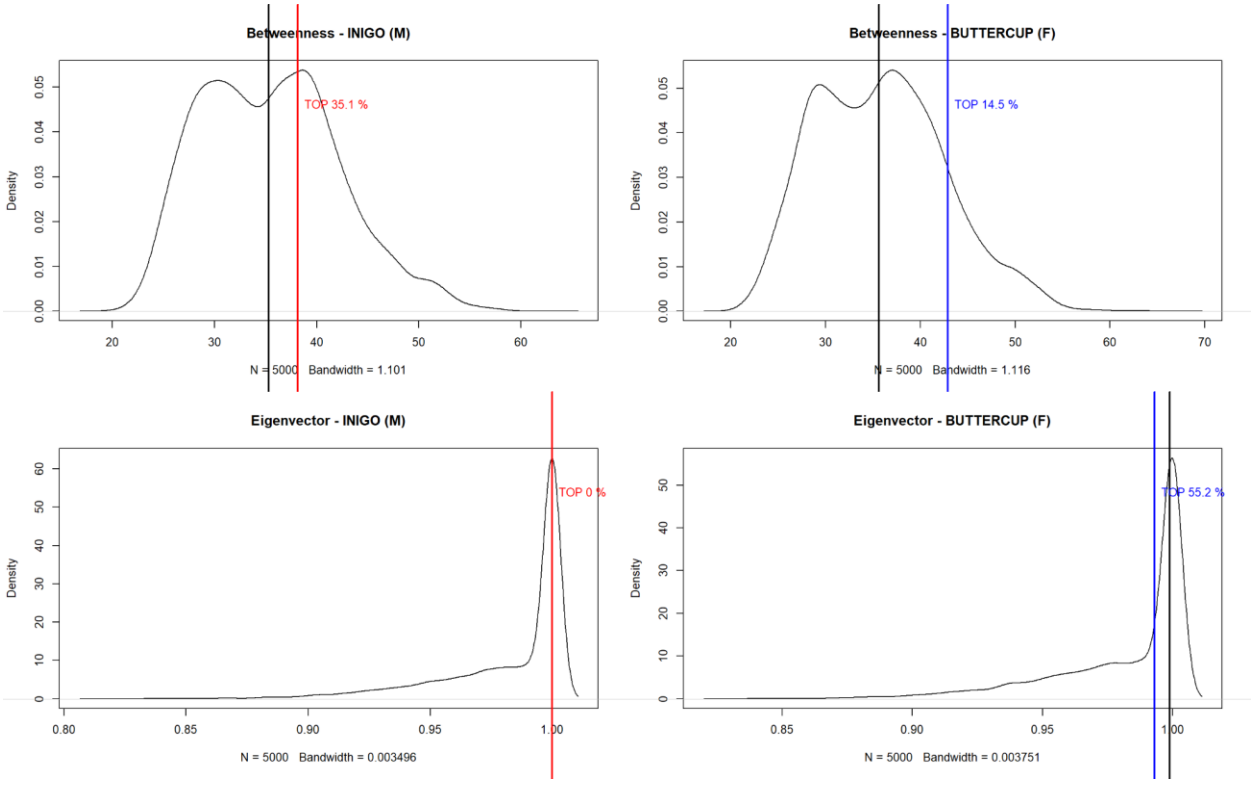


La La Land (2016)

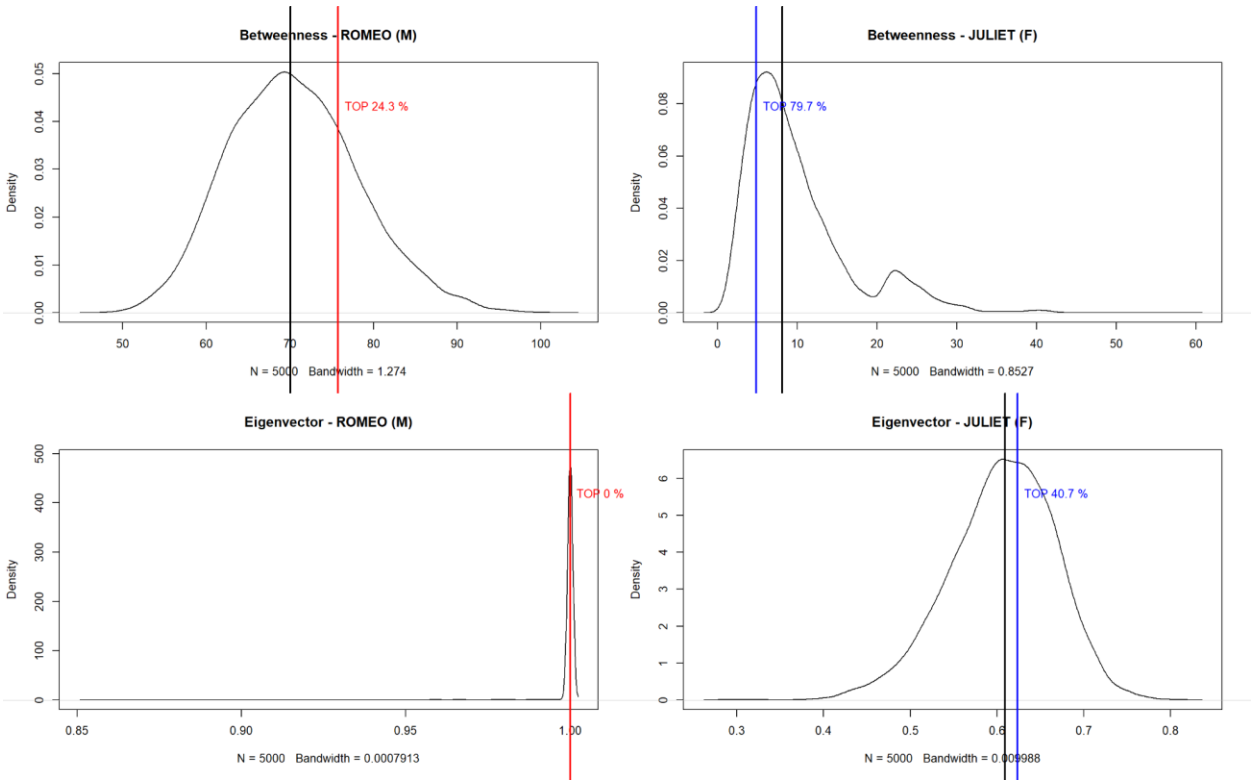
Appendix 3 : Comparison of Protagonists



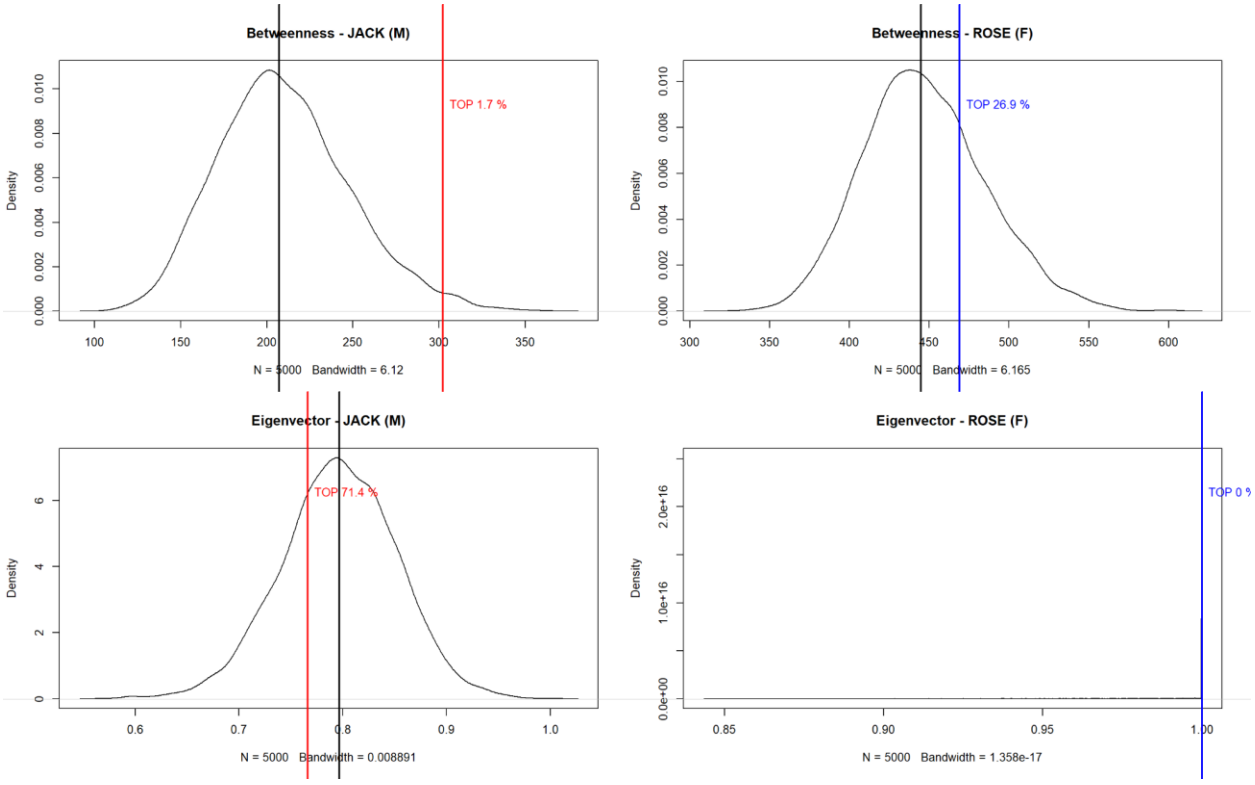
Annie Hall (1977)



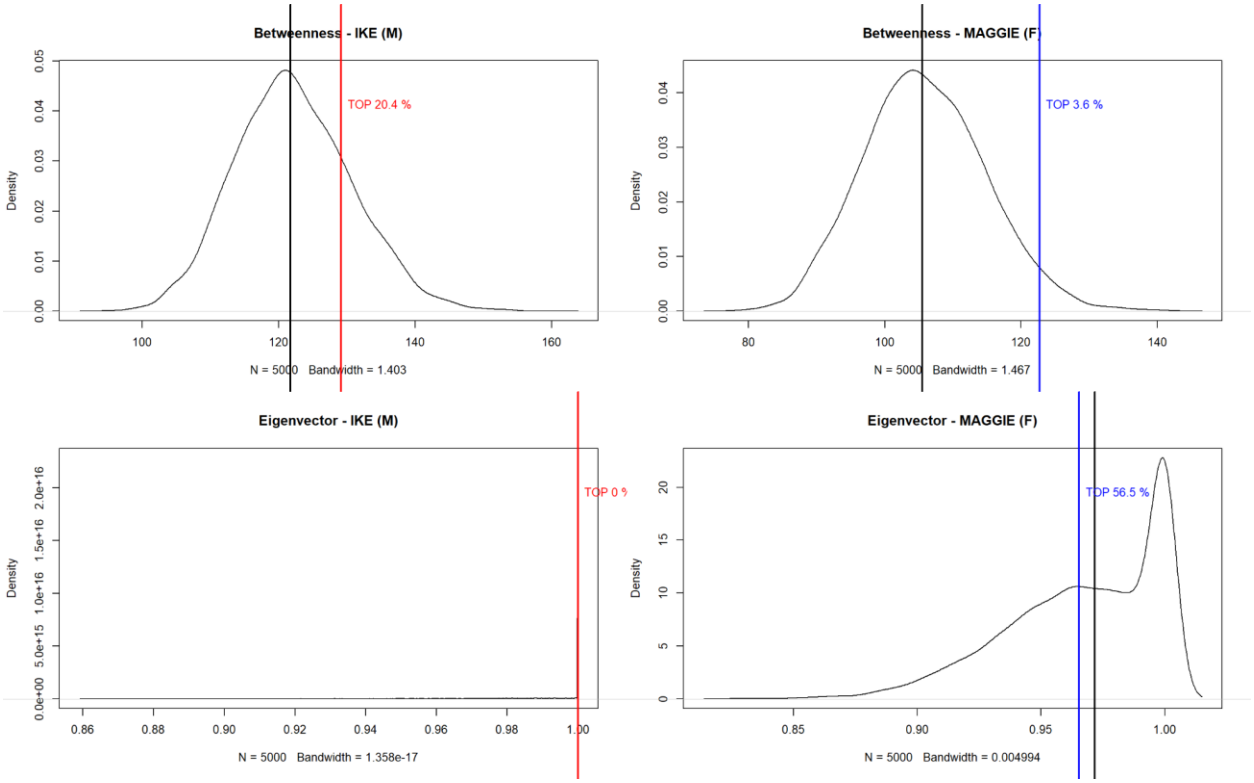
Princess Bride (1987)



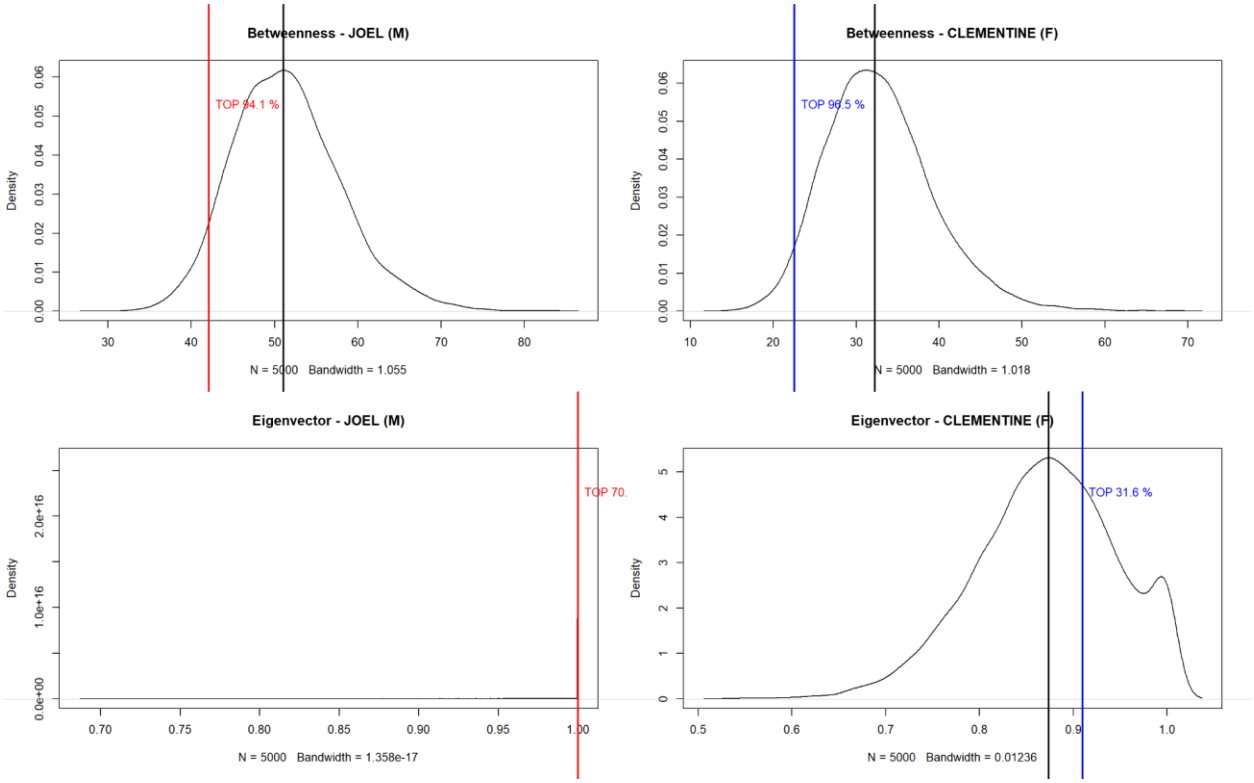
Romeo and Juliet (1996)



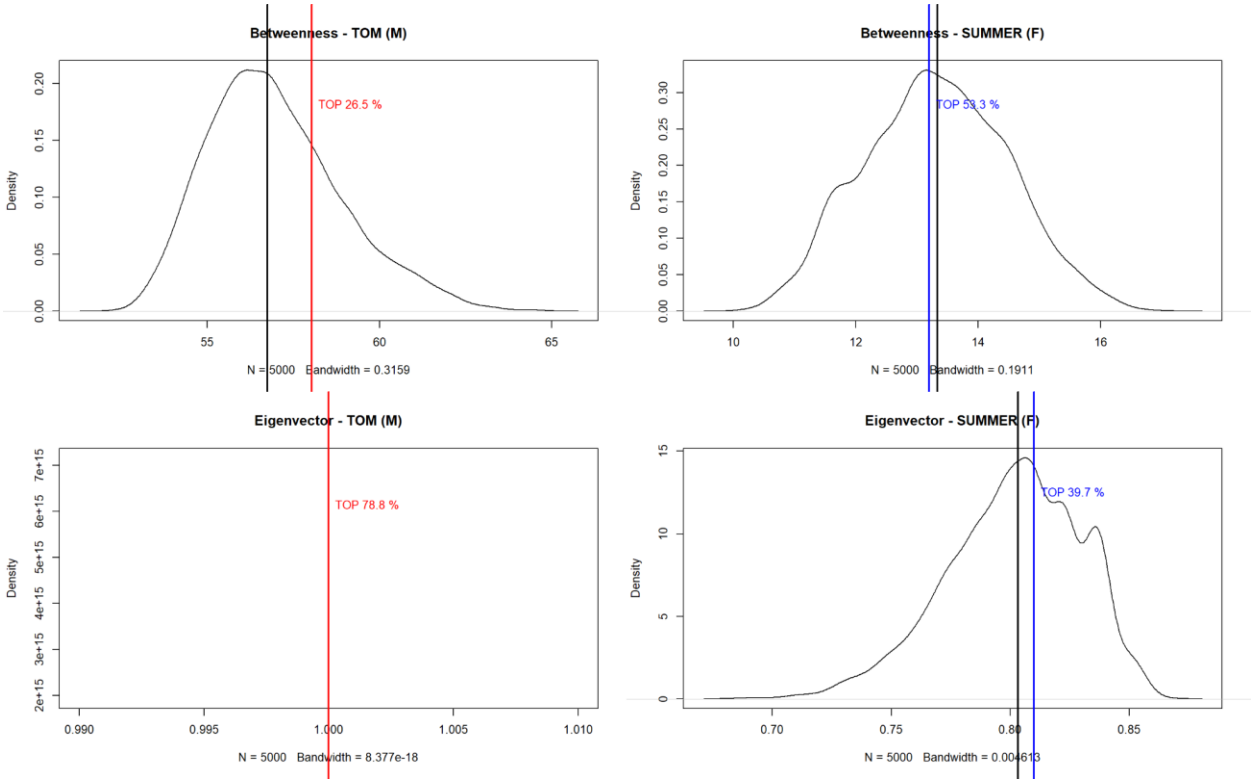
Titanic (1997)



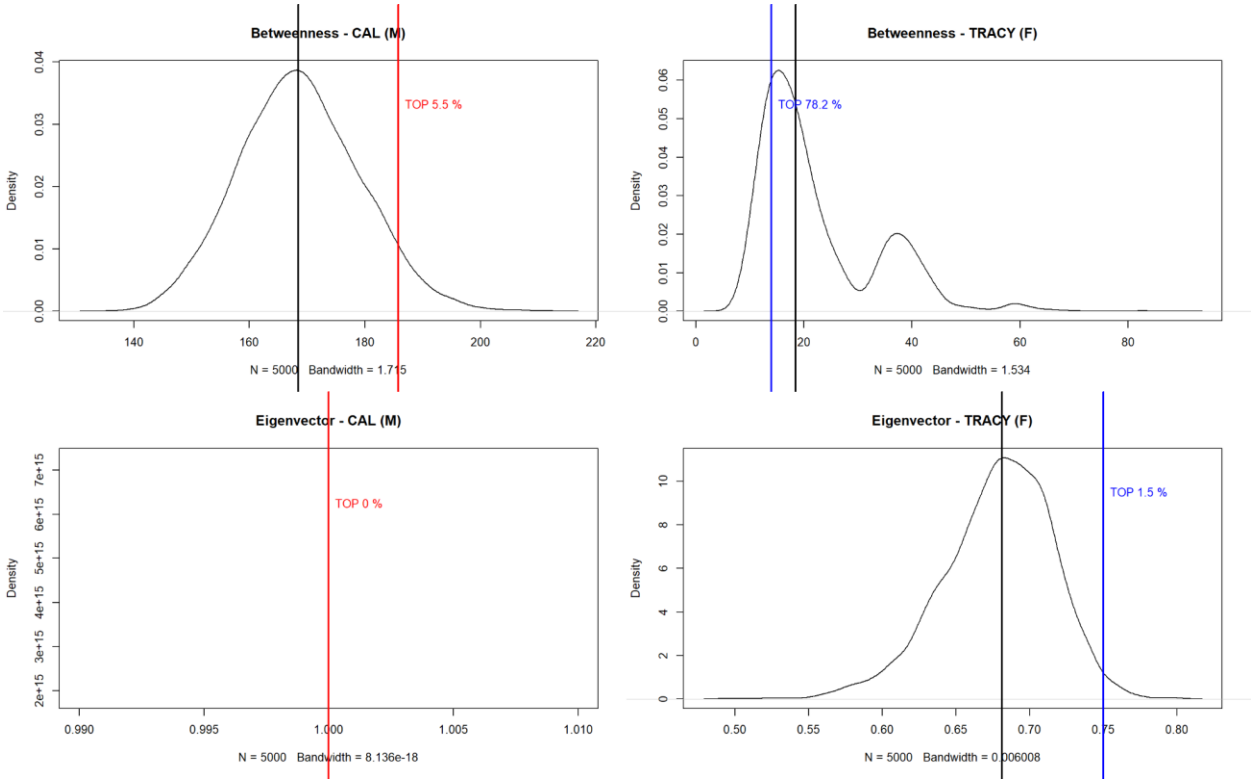
Runaway Bride (1999)



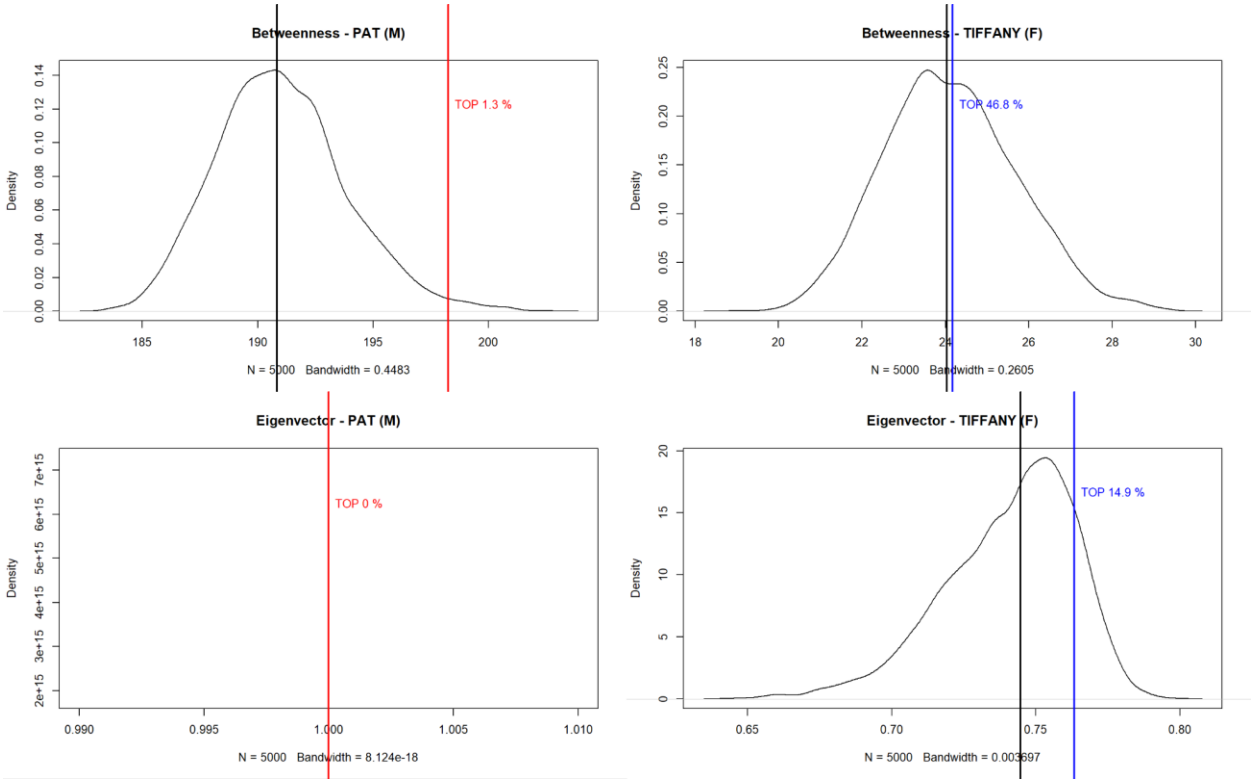
Eternal Sunshine of the Spotless Mind (2004)



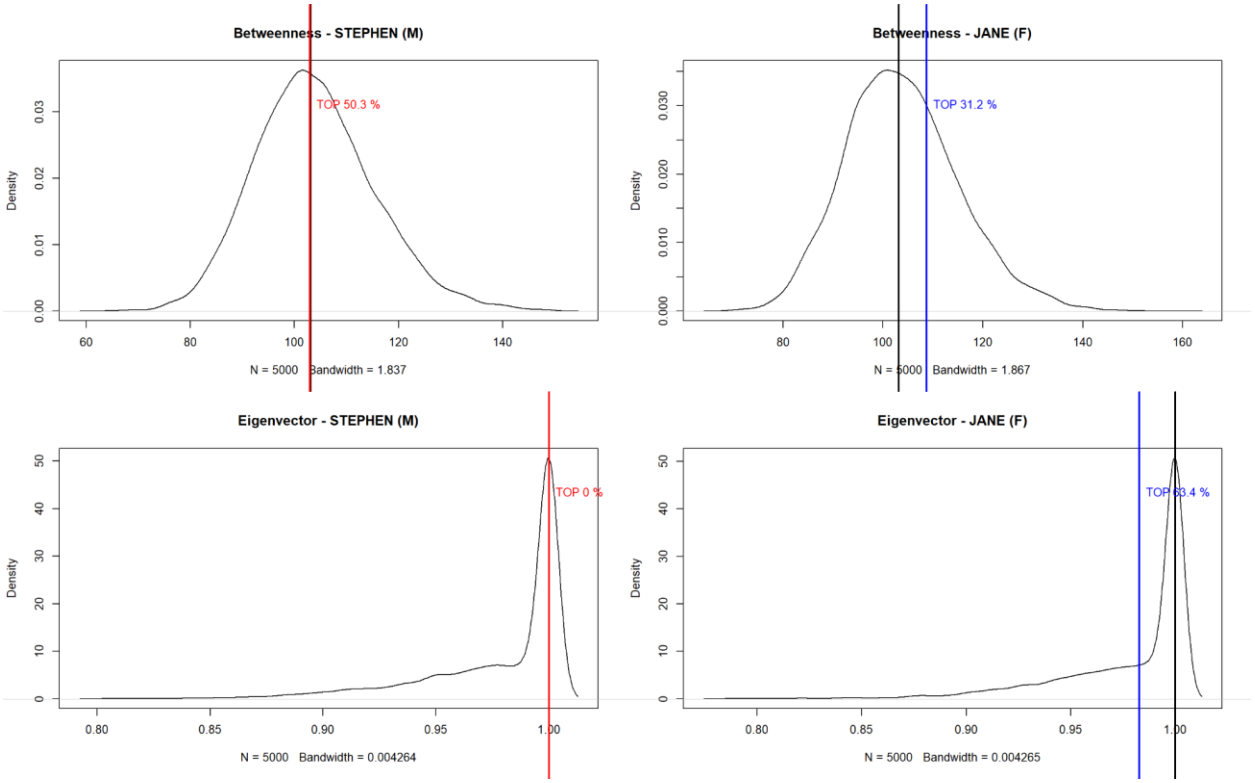
500 Days of Summer (2009)



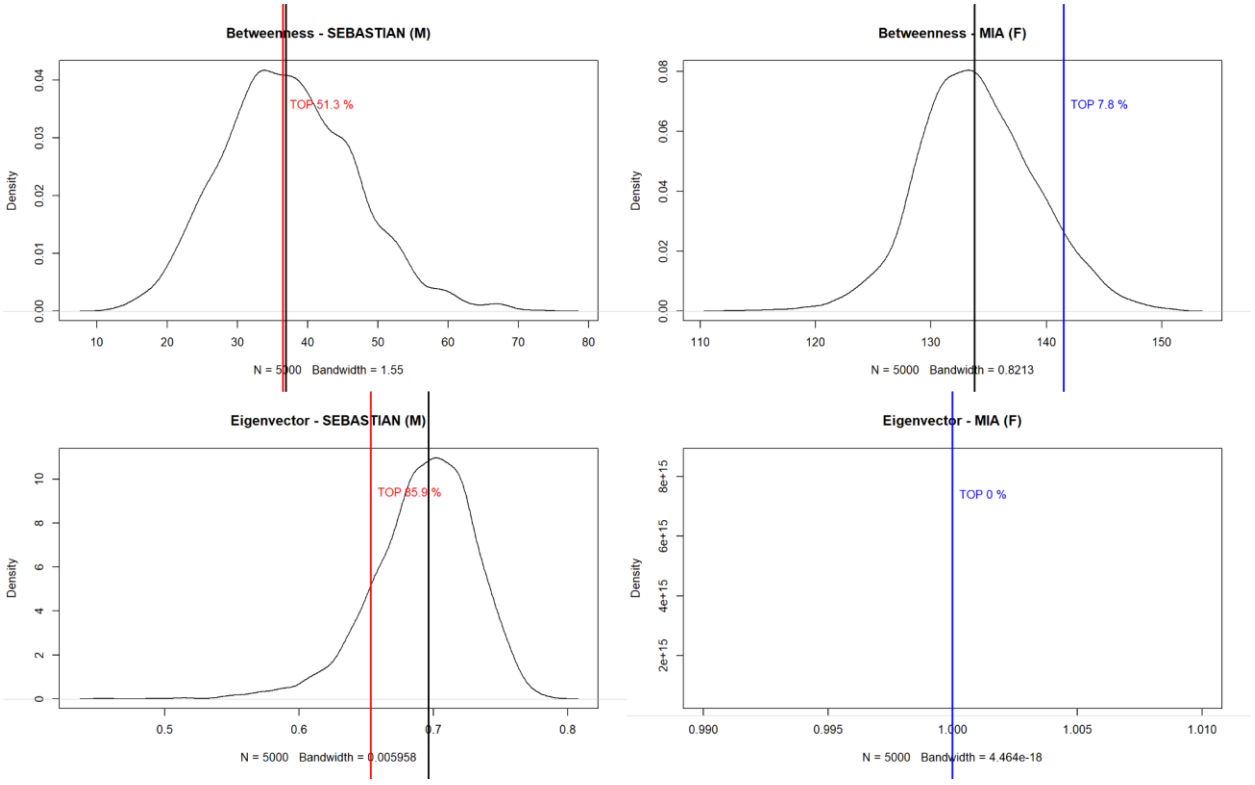
Crazy Stupid Love (2011)



Silver Linings Playbook (2012)



Theory of Everything (2014)



La La Land (2016)