Network analysis of gender-based stereotypical representations of movie character relationships

The project aims at analysing whether there are stereotypical representations of character relationships in movies of various genres released in the span of time between 1920s and 2010s. Specifically, the hypothesis is that the representation of male-female relationships between characters in movies is subjected to difficulties that concretize in the tendency to shoot scenes in which characters interact with only character of the same gender. Moreover, whenever a relationship between a male character and a female one exists, this is usually stronger than the average, mostly of the time related to a love story. Assumed that male and female characters have stereotypical behaviours in the development of relationships within a movie, the authors aim at demonstrate her hypothesis through the scientific method of network analysis.

The hypothesis will be proved against real data from movies collected in ‘[Moviegalaxies – Social Network in Movies](https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/T4HBA3)’ open source repository in Harvard Dataverse by Kaminski Jermain, Schober Michael, Albaladejo Raymond, Zastupailo Oleksandr and Hidalgo César. The authors created a large collection of datasets related to 773 films (1915–2012) gathered through the employment of a movie script parser of their own invention that determined same-scene appearance of characters as a proxy of their connectedness. Each co-appearance is measured as one degree unit per scene, which has then been assigned as the weight of edges within the network. Although not free of minor errors, as the authors themselves declare, the data provided are very rich and exhaustive. Provided in JSON format, each network is composed of a list of edges and a list of nodes. Each edge has its own weight given by the characters co-appearance measured in the movie; while each node is identified by an id and already provided with information about degree, pagerank, triangles, eccentricity, closeness centrality, betweenness centrality and eigenvector centrality measures. To these, the author of the project, in order to conduct her research, felt the necessity to manually add data related to character gender. By consulting [IMDb movie database](https://www.imdb.com/) page related to each selected movie, each character will be provided with gender information assigning them to the group of female or male.

The dataset used for conducting the research is thus composed of a representative sample of movies from the repository. The data will be processed in order to be read in Python and transformed in undirected weighted networks through networkx, a Python package for the creation, manipulation and study of complex networks.

On the data thus obtained, the following measures will be applied – through networkx built-in functions whenever possible or through further models realized by the author when necessary. Firstly, in order to analyse whether characters of the same gender tend to appear in the same scenes more than characters of different gender, modularity will be calculated respectively for the two groups. Secondly, it will be realized a model able to compute whether characters tend to develop stronger relationships with characters of the opposite gender. Finally, community detection algorithms will be applied in order to understand whether it would be possible to predict gender-based clusters of characters from the network structure, assumed that male and female characters tend to have different behaviour in the development of relationships in movies. The final purpose is to test whether these aspects describe stereotyped relationships between characters that make male-female connections and strength of their connections different from what would happen by chance. The computed measures from the sampled dataset will then be proved against the traditional random permutations.

Furthermore, an interesting aspect to be deepened is whether there are differences in movie character representations due to the movie historical period or movie genres.

Finally, a discussion will be raised about the validity of using scene co-appereance as proxy of character relationships in movies. Indeed, two characters may be related even if they do not often appear in the same scene, such as two lovers who are far apart due to the plot structure. Reflections will be conducted about possible alternatives to represent character relationships more accurately.