CineMAC
a Transnational Cinematic Citations Network Analysis

Andrea D'Arpa, Computer Science Master Degree (Curriculum C), ID number 0000983830
Chloe Papadopoulou, Digital Humanities and Digital Knowledge, ID number 0001058391
Maddalena Ghiotto, Digital Humanities and Digital Knowledge, ID number (matricola)

Context

The general field in which our research can be situated is Film Studies, a sector concerned with exploring the narrative, artistic, cultural, economic, and political implications of the cinema.
In particular we are interested in adopting a postcolonial perspective as a critical approach to the analysis of production, theoretical frameworks, contexts and creation of films.
To do so, our study will consider the history of films and the reciprocal influence between USA and East Asian cinema production, hence adopting the tools and methods of citation analysis studies.

Problem and Motivation

The problem we aspire to address within this study is the power asymmetries underlying the ever present vision of history of Cinema as Hollywood-centered, as well as how eastern cinema has been incorporated in western cinema, often without fully acknowledging the sources, hence undermining the role of the foreign artistic influence; or how it affected the representation of the Other, shaping the perception and history of foreign identities.
As a compass to better navigate this field of studies we draw knowledge from two main texts:

- *Postcolonial Cinema Studies*, edited by Sandra Ponzanesi, and Marguerite Waller, Taylor & Francis Group, 2011. *ProQuest Ebook Central*
- *East Asian Cinemas, Exploring Transnational Connections on Film*, edited by Leon Hunt & Leung Wing-Fai, I.B Tauris & Co Ltd, New York, 2008

Moreover, we are taking inspiration from Spitz and Horvàt's Network Analysis study *Unraveling Cinematic Citation*[1] where light is shed «on the potential of a systematic quantitative investigation based on cinematic influences in identifying the most inspiring creations in world cinema».[2] Their main contribution was to create for the first time a movie rank based on the historical importance of movies, beyond traditional measures of success, such as box-office revenue and critical acclaim, which are external to the craft of the movie itself.

---

[1] Andreas Spitz, Emőke-Ágnes Horvát, *Measuring Long-Term Impact Based on Network Centrality: Unraveling Cinematic Citations*, published in Plos One, 8 October 2014 (https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0108857#references).
[2] *Ibidem.*

Nonetheless, their result is not critical to the underrepresentation of East Asian Cinema, and their results, as one could expect, reflect a western narration of cinema history.
Our specific contribution will be:

- A long-term impact analysis that specifically ranks East Asian films according to their influence on American cinema.
- A general long-term impact film rank where the representation of East Asian cinema is weighted and balanced against the overrepresentation of American Cinema.
- The individuation of central figures in transnational film production collaboration in order to assess how production with transnational teams relate to the success of the film, and to what extent it agrees with the most famous ranks of movie's success.

Datasets

The project's data is acquired by the Internet Movie Database, which provides datasets in tab-separated values format, freely accessible on their dedicated website for personal and non commercial use.

In particular, we used the subset title.akas.tsv.gz[3], which contains a variety of information regarding film titles and their country of origin.

Regarding information on movie citations, it appears that the dataset movie-links.list used in our reference paper had been discontinued by IMDb, as we could not find an updated version on their platform. Luckily, we were able to find it in another repository in a text file format.
This dataset, created and expanded by millions of users, records six types of citations between over 40,000 international feature films starting from the beginning of cinema; the types of citations range from very subtle references to explicitly featured sequences of previous films.

The combination of these datasets provides an image of a film's basic information - including its country of origin - and its relation to other films by means of inward or outward citations.

For handling our project's data and computing our analysis measures, we implement the use of Python libraries and the NetworkX package.

Measures

We will consider two graphs:

[3] 1. The Internet Movie Database (IMDb). Alternative interfaces. http://imdb.com/interfaces/

1. **Network 1:** a directed multiplex graph in which every node corresponds to a movie and every edge to a citation. The edge is directed from the citing movie to the cited one. Citations, hence edges, are of six distinct qualitative types of citations (as already differentiated in IMDB data): references, features, remake of, spoof, edited from, follows.
   We also plan to assign weights to the edges, as to give more relevance to more impactful types of citations.
2. **Network 2:** it's a projection of a bipartite network with directed edges from a set of agents to a set of movies. Is an undirected graph, in which all nodes represent people that worked on a movie (the director, the writer, the producer) and every edge connecting them represent two nodes co-worked in the creation of a movie.

We decided to use several techniques, each addressing a specific topic of the project.
To Network 1 we will apply:

1. A long-term **impact analysis** of East Asian films on American cinema:
   - **Bibliographic Coupling Analysis**: This technique measures the similarity of the reference lists of two films and the frequency of citation between them. This method can be used to analyze the degree of influence that East Asian films have had on American films based on the extent of citations in the film's reference lists.
   - **Co-occurrence Analysis**: This technique measures the frequency of co-occurrence of films within a defined set of data sources. This method can be used to identify the key films that are most frequently referenced and cited in relation to East Asian and American films.
2. A general long-term impact film rank:
   - **PageRank Algorithm**: This well-known algorithm used in network analysis ranks nodes in a network based on the number and importance of incoming links. This method can be used to rank films based on their overall impact, considering both the representation of East Asian films and the overrepresentation of American films.
   - **Out-degree** to see which one are the movies most prone to citations
3. We will apply **betweenness centrality** in order to investigate what movies in particular have mediated the influence and impact of east asian movies throughout history, functioning as a "bridge" to other movies that maybe didn't have primary access to the source (if there is any way to analyze film similarity this would be great, because then we can really see when a movie citing another american movie with a strong east asiatic influence still unknowingly acquired the same features that came from the asian source)

To Network 2, to inquire about the influence that may have been transmitted by people's collaborations and not by images themselves, we will apply:

- **Centrality Measures**: This is a set of measures that quantify the centrality of a node in a network. Different centrality measures can be used to identify the most influential and central figures in a network of transnational film production collaboration. For the sake of the project we decided to go with **Degree Centrality**

(which assesses the importance of a node based on the number of other nodes connected to it**)** and **Eigenvector Centrality (**that measure the importance of a node based on both number and quality[4] of the other nodes connected to it**)** measures. Finally, **closeness centrality** to assess which are the most well connected agents.

- We will then evaluate what are the movies in which our most important (influential and well connected) agents of Network 2 participated and how that is related to the influence ranking of the movie.
- We will calculate **components, clustering coefficient**, **local clustering coefficient** and **redundancy** to see the connectedness of the network as well as individuate structural holes, hence those nodes, whose missing links give it more control over diffusion of knowledge and information between neighbors. This will be done in order to better understanding how the degree of connectedness of the network is related to the nationality of the agents.

[4] Quality can be for example the number of other nodes connected to the node wich i'm connecting with. This concept is used, for example, in most blockchains.