

CineMAC: a Transnational Cinematic Citations Network Analysis

Maddalena Ghiotto

Digital Humanities and Digital Knowledge, ID number 0001030096

Chloe Papadopoulou

Digital Humanities and Digital Knowledge, ID number 0001058391

Contents

1 Abstract	2
2 Context	2
3 Problem and Motivation	3
4 Datasets	4
4.1 Citations' Network	4
4.2 Collaborations' Network	6
4.2.1 Network modeling	6
5 Validity and Reliability	7
5.1 Data Validity	7
5.2 Data Reliability	7
6 Measures	8
6.1 Citations' network	8
6.2 Collaborators' network	9
7 Results	10
7.1 Citations' Network	10
7.1.1 In-Degree Centrality	10
7.1.2 PageRank	12
7.1.3 Out-Degree Centrality	12
7.1.4 Betweenness Centrality	12
7.1.5 Weighted average of the measures	13
7.1.6 Decisions and criteria	14
7.1.7 Discussion	16
7.2 Collaborations' Network	18
7.2.1 Degree Centrality and Eigenvector Centrality	18
7.2.2 Closeness Centrality and Betweenness Centrality	18
7.2.3 Connected components	19
7.2.4 Clustering coefficient	19
7.2.5 Louvain community detection algorithm	19
8 Conclusion	24
9 Critique	24

1 Abstract

Project CineMAC is a network analysis project that aims to study the influence, trends and patterns of cinematic production through the creation and analysis of a network of film citations network and a corresponding collaborating network of various film agents. Unlike previous studies that failed to stress the domination of the USA cinematic industry in the available data, our project addresses the **power asymmetries** underlying the present westernised vision of cinematic history, undermining the role of East-Asian artistic influence. Our study did not manage to adequately account for the underrepresentation of East Asian cinema, against modern literature findings which call for its recognition as a hub for the transnational flow of ideas, due to lack of relevant information that prevented us from predicting accurate balancing factors. However, it allowed us to yield **alternative metrics for assessing** the world of cinematic production and highlighted **promising possibilities** for future studies that could contribute to this scope.

2 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. Specifically, we are interested in adopting a postcolonial perspective as a critical approach to the analysis of production, theoretical frameworks, contexts and creation of films. So far, most of postcolonial film studies focus on the qualitative analysis of the representation of the other as well as on how the impact of colonialism can be seen in film production practices and styles.¹ In the context of this paper we refer to the term *transnational* precisely under the lens of postcoloniality, hence not merely as a synonym of *world cinema* but adopting a critical, discursive stance on how film-making activities negotiate with nationality on all levels. As Will Higbee and Song Hwee Lim express:

“transnationalism in films and in the study of cinema cannot be taken as a given or for granted. The concept of ‘transnational cinema’ cannot be merely descriptive because all border-crossing activities are necessarily fraught with issues of power; neither can it be purely prescriptive as this often amounts to nothing more than wishful thinking.”²

To do so, our study will adopt the methods of citations analysis to uncover transnational influences through cinematic citations and address trends and patterns of cinema production. This is coherent with two main approaches of transnational film studies, respectively: the analysis of the cinematic representation of cultural identity, to challenge national cinema as stable and western-centric in its ideological norms as well as its narrative and aesthetic formations; the analysis of cultural formations that are rarely contained within national boundaries, such as questions of production, distribution and exhibition³

The boundaries of this study are those of transnational influences between North American film industry and East-Asian film industry. While it is less ambiguous what geographical area

¹for a comprehensive overview of the domain see Waller, S. P., Marguerite (Ed.). (2011). Postcolonial Cinema Studies. Routledge. <https://doi.org/10.4324/9780203181478>

²Hwee Lim, S. and Higbee, Will, Concepts of transnational cinema: Towards a critical transnationalism in film studies. <https://doi.org/10.1386/trac.1.1.7/1>

³Ibidem.

North America represents - the USA - defining East-Asia is not evident. As Leon Hunt and Leung Wing Fai also suggest: “East Asia is a case study in mutable regional identities. One can identify at least two ‘East Asias’. Arguably the most familiar maps onto the north-east Asia that emerged out of a post-war (Western) distinction from south-east Asia, and includes China, Hong Kong, Taiwan and Japan. However, another East Asia was imagined by Malaysian Prime Minister Dr Mahathir in his notional East Asian Economic Caucus, ‘a crescent of prosperous nations extending from north-east Asia to south-east Asia from Tokyo to Jakarta’.”⁴ For the sake of this study, and due to the data limitations that constrain it, when referring to East Asia we will be addressing a group of countries that contain China, Hong Kong, Japan, Mongolia, North Korea, South Korea, and Taiwan.

3 Problem and Motivation

Several studies have been carried out adopting the tools of network analysis on cinematic production, collaborations and citations. Even from a first literature review, we realised that the vast majority of these studies focus on the western industry. The reason can be easily inferred considering the data available, since most of these studies use imbd datasets as a starting material, which contains a large amount of structured metadata on movies and people that produced them.⁵ We consider in particular studies that used quantitative approaches to evaluate the influence of films in cinema-history by analysing connections between films. The problem we identified in almost all the studies known to us on this topic, is that they **failed to account for the partiality and westernised nature of the database**.

Both Canet and Valero’s study⁶ and Wasserman et al.⁷ refer to “film history.” In general, without addressing the unbalanced film representation of imbd. Bioglio and Pensa also omit this factor when identifying most influential movies “in the history of cinema”⁸, even though they acknowledge the western boudaries of their perspective in a later study⁹.

Our research aims at establishing a dialogue with these earlier studies on cinematic influence based on imbd database, by addressing the **power asymmetries** underlying the ever present vision of history of Cinema as America-centered, undermining the role of East-Asian artistic influence. Besides cinematic **citations**, we will research also **collaborations** and we will try to account for the misrepresentation of East-Asian cinema, counterbalancing with quantitative methods, where possible, the over-representation of American cinema.

Our specific contributions will be:

- The creation of a long-term impact rank that assesses USA and East Asian films according to their influence in Cinema History.

⁴Hunt, Leon and Wing Fai, Leung . East Asian Cinemas. Bloomsbury. 2008, p.5

⁵for an overview of the various studies carried out on the database see Canet, F., Valero, M. A., & Codina, L. (2016). Quantitative approaches for evaluating the influence of films using the IMDb database.p.153-154

⁶Ibidem.

⁷Spitz, A., & Horvát, E.Á. (2014). Measuring Long-Term Impact Based on Network Centrality: Unraveling Cinematic Citations. PLOS ONE, 9(10)

⁸Bioglio, L., & Pensa, R. (2018a). Is This Movie a Milestone? Identification of the Most Influential Movies in the History of Cinema.

⁹Bioglio, L., & Pensa, R. G. (2018b). Identification of key films and personalities in the history of cinema from a Western perspective. Applied Network Science, 3(1), Article 1.

- A general long-term impact film rank where the representation of East Asian cinema is more weighted and balanced.
- 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.

4 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.gz¹⁰, 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 implemented the use of Python libraries and the NetworkX package; the entire workflow and documentation can be found on our github repository folder.

4.1 Citations' Network

After processing and combining our data from the different sources, we modelled a directed graph in which every node corresponds to a film and every edge to a citation, whose direction is towards the film being cited¹¹. In total, the network consists of **26.588 nodes** with **86.944 edges** among them. Each film/node is identified by its respectful title-id containing its name and year of production, along with the following attributes retrieved from the imdb dataset:

- **year**: year of production,
- **country**: one or multiple countries of production,
- **genres**: a list of the film's genres,
- **tconst**: imdb identifier (for conducting a comparison with the collaborator's network).

¹⁰The Internet Movie Database (IMDb). Alternative interfaces. <http://imdb.com/interfaces/>

¹¹Please see our merging and network1 modelling notebooks for a detailed documentation.

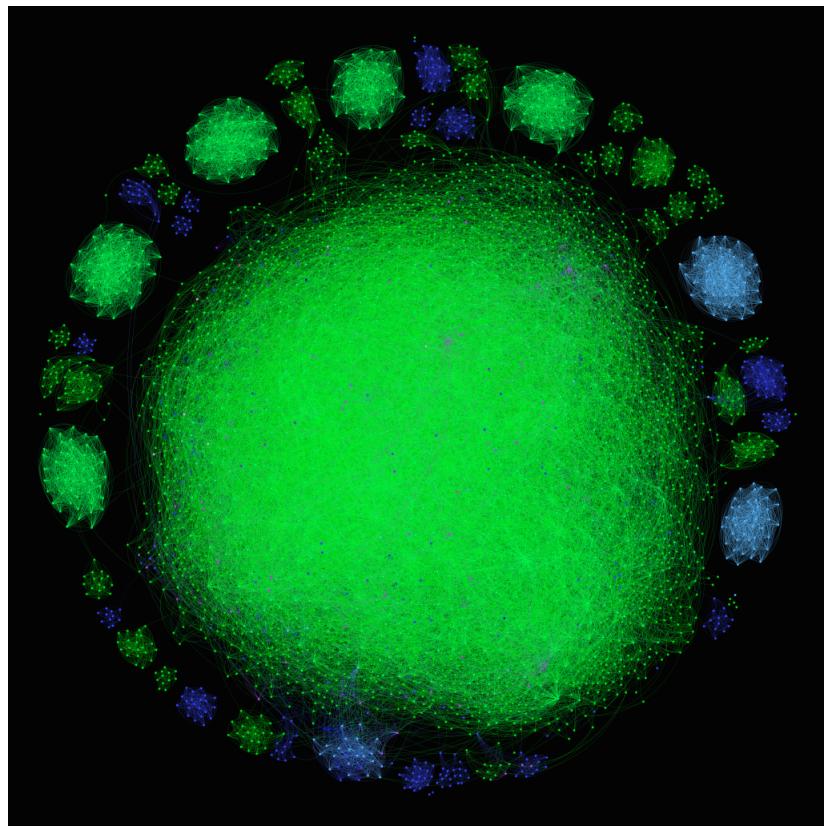


Figure 1: Citation's network.

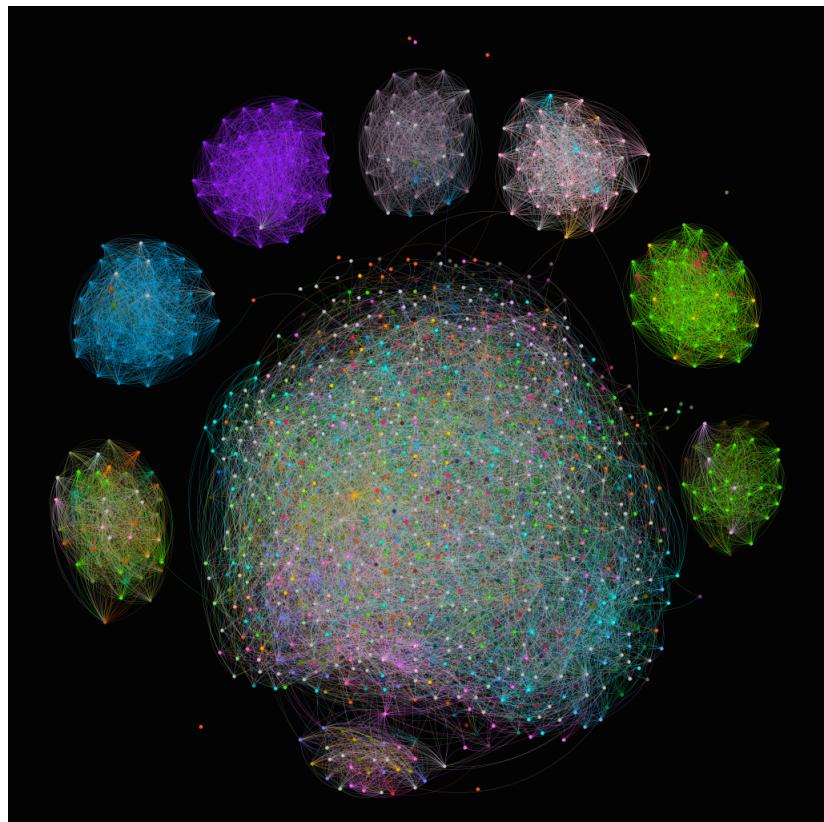


Figure 2: Citation's network rendered by genre partitioning.

4.2 Collaborations' Network

Before modeling the final Collaborations network, we needed to model a middle-step undirected bipartite network with edges from a set of agents to a set of movies, so that we could later project it on the people's set of nodes and obtain our final Collaborations' Network to analyze.

To model the bipartite network, some further steps of data cleaning and filtering were carried out to create a csv dataset fit for our necessities.

The complete pipeline is documented in a Jupyter Notebook `network2_data.ipynb`. The output of the notebook is the dataset `network2_data.csv`, that contains the following movie metadata: movie id, title, movie country, year, references, person id, person role, primary name, birth year and death year

4.2.1 Network modeling

The network modeling has been carried out using python library `networkx` and is documented in the script `network2modeling.py`. In the bipartite network, all movie nodes have the movie `tconst` identifier as label and “`title`”, “`country`”, “`year`”, “`cites`” as attributes; all people nodes have people `nconst` identifier as label and “`name`”, “`role`” as attributes.

Afterwards, this network was projected over the people's set of nodes to create our final Collaborations' network. Note that in this process, the `nconst` of movies that define collaborations have been assigned as edge labels.

Visualizations of our network with different colors for different roles:

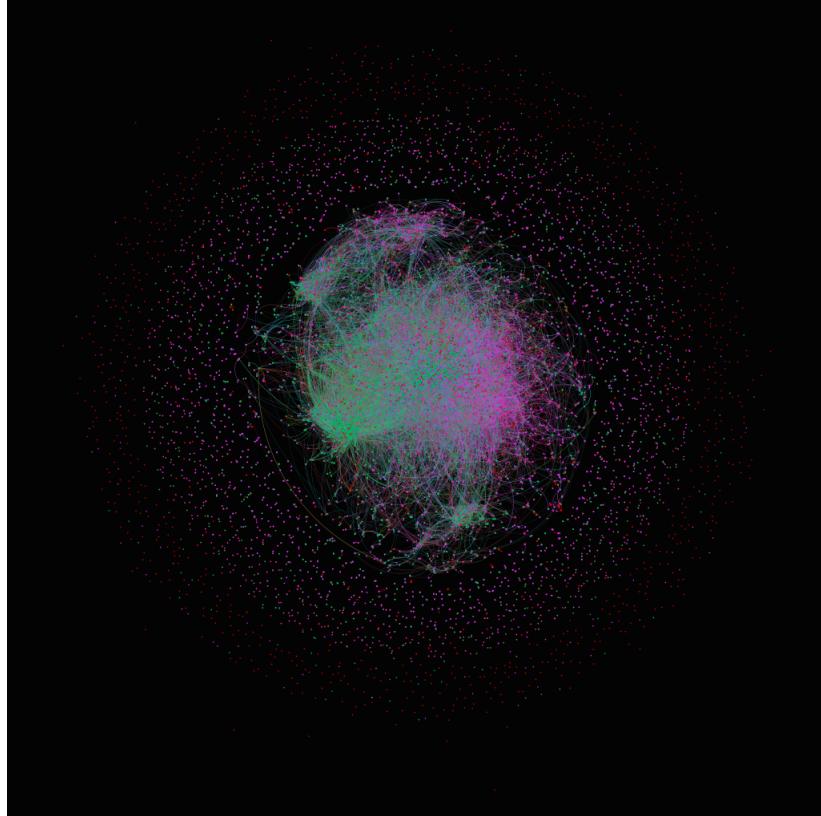


Figure 3: Collaborations Network

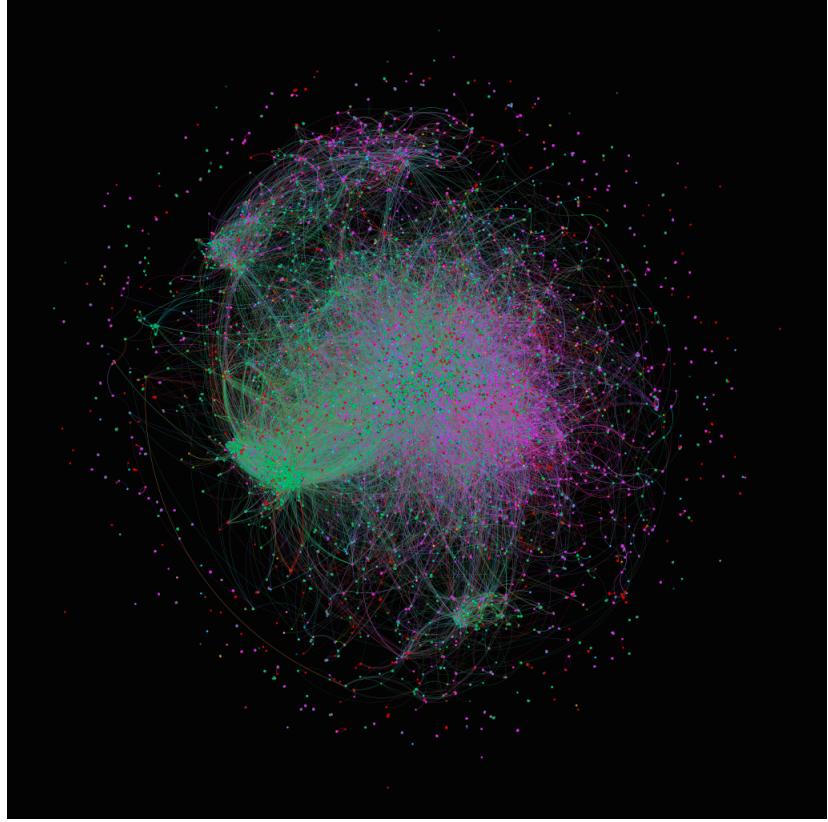


Figure 4: Collaborations’ network filtered Degree > 5

5 Validity and Reliability

5.1 Data Validity

In the context of our project, data validity concerns the extent to which it accurately reflects patterns of influence inside the film industry. Although the data reflects the state and impact of various cinematic industries across the world, we must point out two important issues that can result in a less than ideal representativeness.

Firstly, we observed the dominance of US movies within the dataset *movie-links.gz4*, which is due to the fact that US movies are indeed more widespread and therefore more likely to be noted by the datasets’ creator community. However, as we will discuss below, it hindered our efforts in

Furthermore, we encountered the challenge of inconsistent titles for certain Chinese and Japanese movies across our two dataset sources; although these were resolved to an extent, there were films that could not be matched in order to retrieve their imdb metadata; due to the limitations of this project, they were excluded from the network. We acknowledge that these issues may result in a lack of representativeness of the film industry as a whole, and would be addressed in the case of a further development of this project.

5.2 Data Reliability

The project’s data was obtained from IMDb, a widely recognized source for film-related information, accumulating a vast amount of data over many years. The accessibility of this data

from a trusted platform like IMDb enhances its reliability.

Moreover, the fact that the data has been utilized by existing projects¹², can be an indicator its reliability and suitability for research purposes; we plan to furtherly assess our analysis and results by comparing them to the independent project's outcomes.

6 Measures

6.1 Citations' network

We decided to use several techniques, each addressing a specific way of assessing a film's potential impact in cinematic production. **For a general long-term impact film rank we chose the following measures:**

- **In-Degree Centrality**, which in the context of our graph measures the number of incoming citations a film receives from other films in the network. Thus, we can measure the popularity and influence of a film on the basis of its incoming citations.
- **Eigenvector centrality**, which calculates the influence of a film based on both its own centrality and the centrality of its neighbors in the graph. Films that are connected to other highly influential films have higher eigenvector centrality scores, which reinforces the idea that the influence of a film is enhanced by its connections to other influential films.
- **PageRank**, an algorithm originally developed by Google to measure the importance of web pages in a network of hyperlinks. In the context of our network, PageRank assigns importance scores to films based on the structure of citations within the graph.

Further measures for assessing a film's contribution to the flow of information in cinema:

- **Out-degree centrality**, to see which movies and countries are most prone to citations, and can consequently contribute to the entire graph structure.
- **Betweenness centrality**, in order to investigate what movies occupy central positions in the network, having mediated the influence and impact of films throughout history, acting as "bridges" of information flow.

¹²Andreas Spitz, Emőke-Ágnes Horvát, Measuring Long-Term Impact Based on Network Centrality: Unraveling Cinematic Citations, published in Plos One, 8 October 2014

6.2 Collaborators' network

For the purpose of our project, we applied the following measures to our second network.

Degree Centrality and **Eigenvector Centrality**, in order to identify individuals with a high number of collaborations, meaning individuals who are well connected and likely to have worked at a larger number of movies involving different people. We can assume that these people bring together diverse talent and expertise in their sector which can make them more adaptable to work with different groups. Given the nature of the movie industry this is not necessarily an indicator of strong artistic influence, on the contrary, many artistically influential individuals in the field are known to have consistent collaboration patterns throughout their career. Individuals with a higher degree centrality certainly serve as hubs in the Collaborations network.

We decided to apply both Degree and Eigenvector centrality to be able to evaluate how the ranking of people would change including the quality of the connections - interpretable in term of *prestige* - that Eigenvector introduces, as it weights the score of a node on the relative score of its neighbour nodes.

Closeness Centrality, since it measures how quickly an individual can reach others in the network in terms of collaboration connections. Individuals with a higher closeness centrality have the potential to connect to a larger portion of the network more rapidly. Their influence lies in their ability to establish collaborations with diverse professionals, potentially leading to a broader range of projects and opportunities. They can quickly disseminate their ideas, creative concepts, or industry trends to a wide range of collaborators

Betweenness Centrality quantifies the extent to which an individual acts as a bridge or intermediary between others in the network. Individuals with high Betweenness centrality play a crucial role in **facilitating or gatekeeping** the transmission of information and ideas between groups of individuals who may not directly collaborate with each other. In the movie industry they can serve as key conduits for information flow, enabling exchange of knowledge and creative concepts. Being connected with diverse artistic communities, they are likely to be key individuals for the sharing of different artistic visions between groups that may not typically interact, creating opportunities for cross-pollination of ideas and the introduction of new artistic concepts.

Weighted average of the measures . Wanting to assess transnational collaboration and influence in the movie industry, we felt relevant to weigh each measure to giving a better representation of individuals who contributed more to **ideas-spreading and artistic cross-contamination**, all while trying to counterbalance the predominance of American personalities in our rank (see section 7 for the Results section).

To better understand the structure of the network and therefore giving context to the measures' result we computed:

Connected components, to assess the network's degree of fragmentation. A connected component is a subgraph in which each pair of nodes are connected with each other via a path. The components of a network usually represent distinct collaboration communities within the movie industry, it may identify groups with a strong history of working together with shared artistic interest or creative processes. Individuals within the same connected component may influence each other's work, creative decision and artistic approach.

Clustering coefficient on the bigger network component. The average clustering coefficient of a node is calculated by examining its neighbors and determining the proportion of connections among those neighbors. It measures how likely the neighbors of a node are connected to each other.

Louvain community detection algorithm, a heuristic method based on modularity optimization, which finds the community structure that maximizes modularity score.¹³ We applied this algorithm to look deeper into the structure of the connections of our network an to enable observation of non topological aspects on the different communities.

Finally, we decided to analyse the relation between the two networks, individuating the movie with the highest rank in the Citations' network of our twenty most influential people in the Collaborations' network. This is done to evaluate the correlation between the two ranks.

7 Results

7.1 Citations' Network

7.1.1 In-Degree Centrality

In the context of our graph, in-degree centrality measures the number of incoming citations a film receives from other films in the network. Thus, we can measure the direct impact of a film on the basis of its incoming citations, which is a very straight forward way to assess its influence and popularity in the history of cinema.

We observe that the scores indeed reflect the dominance of USA productions in our original dataset. Given that fact but also, the clustering trends of citations amongst films of the same countries, we wanted to explore the relationship of the two industries in a more meaningful way. Starting from the in-degree centrality of productions of East Asian and USA cinema, we thought to introduce a new evaluation filter according to their **incoming citations by foreign film industries**.

The application of this criterion proved insightful results, because the ranking order differs from the unfiltered data regarding films with the highest in-degree centrality(Table 2, Table 3. We therefore theorized that a further development could indeed weigh this factor into account, by developing a function that evaluates a film's citation by a foreign industry as an important parameter to its overall popularity and reach throughout the history of cinema.

¹³Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008

Rank	Film	In-Degree	Country	Genres
1	Star Wars (1977)	0.02899913	USA	Action,Adventure,Fantasy
2	The Wizard of Oz (1939)	0.02756986	USA	Adventure,Family,Fantasy
3	Psycho (1960)	0.01564674	USA	Horror,Mystery,Thriller
4	Jaws (1975)	0.01293865	USA	Adventure,Thriller
5	The Godfather (1972)	0.01293865	USA	Crime,Drama
6	Gone with the Wind (1939)	0.01102042	USA	Drama,Romance,War
7	Casablanca (1942)	0.01098281	USA	Drama,Romance,War
8	2001: A Space Odyssey (1968)	0.01060669	USA	Adventure,Sci-Fi
9	The Shining (1980)	0.01056908	USA	Drama,Horror
10	King Kong (1933)	0.01015534	USA	Adventure,Horror,Sci-Fi
11	Frankenstein (1931)	0.01008011	USA	Drama,Horror,Sci-Fi
12	Star Wars: Episode V (1980)	0.0100425	USA	Action,Adventure,Fantasy
13	The Terminator (1984)	0.00966638	USA	Action,Sci-Fi
14	Taxi Driver (1976)	0.00913981	USA	Drama
15	Night of the Living Dead (1968)	0.00895174	USA	Horror,Thriller
16	Citizen Kane (1941)	0.00887652	USA	Drama,Mystery
17	The Exorcist (1973)	0.00887652	USA	Horror
18	E.T. the Extra-Terrestrial (1982)	0.00861323	USA	Adventure,Family,Sci-Fi
19	Raiders of the Lost Ark (1981)	0.00857562	USA	Action,Adventure
20	Apocalypse Now (1979)	0.00812427	USA	Drama,Mystery,War

Table 1: Nodes with the highest In-Degree centrality

Rank	Film	Country	In-degree
1	Star Wars (1977)	USA	0.02899913
2	The Wizard of Oz (1939)	USA	0.02756986
3	Psycho (1960)	USA	0.01564674
4	The Godfather (1972)	USA	0.01293865
5	Jaws (1975)	USA	0.01293865
6	Gone with the Wind (1939)	USA	0.01102042
7	Casablanca (1942)	USA	0.01098281
8	2001: A Space Odyssey (1968)	USA	0.01060669
9	The Shining (1980)	USA	0.01056908
10	King Kong (1933)	USA	0.01015534
11	Frankenstein (1931)	USA	0.01008011
12	Star Wars: (1980)	USA	0.0100425
13	The Terminator (1984)	USA	0.00966638
14	Taxi Driver (1976)	USA	0.00913981
15	Night of the Living Dead (1968)	USA	0.00895174
16	Citizen Kane (1941)	USA	0.00887652
17	The Exorcist (1973)	USA	0.00887652
18	E.T. the Extra-Terrestrial (1982)	USA	0.00861323
19	Raiders of the Lost Ark (1981)	USA	0.00857562
20	Apocalypse Now (1979)	USA	0.00812427

Table 2: USA films with the highest in-citations from East Asian films.

Rank	Film	Country	In-Degree
1	Gojira (1954)	Japan	0.00692068
2	Blade Runner (1982)	USA, Hong Kong	0.004852
3	Enter the Dragon (1973)	Hong Kong, USA	0.00357317
4	Shichinin no samurai (1954)	Japan	0.00188062
5	Wo hu cang long (2000)	Taiwan, Hong Kong, USA, China	0.00176778
6	Mosura (1961)	Japan	0.00169256
7	SorSora no daikaijû Radon (1956)	Japan	0.00161733
8	Kingu Kongu tai Gojira (1962)	Japan, USA	0.00157972
9	Yôjinbô (1961)	Japan	0.00142927
10	Akira (1988)	Japan	0.00139166
11	Ringu (1998)	Japan	0.00135405
12	Tang shan da xiong (1971)	Hong Kong	0.00135405
13	Rashômon (1950)	Japan	0.00124121
14	Zatôichi monogatari (1962)	Japan	0.00124121
15	Dip huet seung hung (1989)	Hong Kong	0.0012036
16	Mosura tai Gojira (1964)	Japan	0.0012036
17	The Crying Game (1992)	Japan, USA	0.00109076
18	Game of Death (1978)	Hong Kong, USA	0.00109076
19	San daikaijû (1964)	Japan	0.00109076
20	Kaijû daisensô (1965)	Japan, USA	0.00105315

Table 3: East Asia films with the highest in-citations from USA films.

7.1.2 PageRank

We consider the PageRank algorithm, which was developed by Google to evaluate the importance of web pages in a network of hyperlinks, a highly appropriate measure in the context of our film citation network; in this way, we can assess nodes based on the **quality of their incoming citations** within the graph. A film's importance is determined not only by the number of citations it receives but also by the importance of the films that cite it, which reinforces the idea that a film's influence is enhanced by its connections to other influential films.

7.1.3 Out-Degree Centrality

In the case of out-degree centrality, the measure score is not indicative of a film's popularity but of the extent to which it references other films in the graph, which can suggest a significant contribution to the overall network of film citations. The ranking based on out-degree centrality score returned the films that are most prone to citing others in the network (Table 5; unsurprisingly, the recurring genres of the top results are Documentary and History, but interestingly, also Comedy and Crime).

7.1.4 Betweenness Centrality

Betweenness centrality quantifies the number of shortest paths amongst nodes, indicating its potential **control over the flow of information** or influence between other nodes. In our context, betweenness centrality identifies films that occupy central positions in the network, acting as important connectors or facilitators of information flow; we can theorize that films with a higher betweenness centrality play a crucial role in connecting different parts of the citation

Rank	Film	PageRank	Country	Genres
1	The Wizard of Oz (1939)	0.00758791	USA	Adventure,Family,Fantasy
2	Dracula (1931)	0.00491641	USA	Drama,Fantasy,Horror
3	Star Wars (1977)	0.00455868	USA	Action,Adventure,Fantasy
4	Psycho (1960)	0.0043758	USA	Horror,Mystery,Thriller
5	King Kong (1933)	0.00428598	USA	Adventure,Horror,Sci-Fi
6	The Birth of a Nation (1915)	0.00392208	USA	Drama,War
7	Snow White and the Seven Dwarfs (1937)	0.00322657	USA	Adventure,Animation,Family
8	Frankenstein (1931)	0.00290993	USA	Drama,Horror,Sci-Fi
9	Shadow of a Doubt (1943)	0.00271295	USA	Film-Noir,Thriller
10	2001: A Space Odyssey (1968)	0.00269954	USA	Adventure,Sci-Fi
11	The Merry Widow (1934)	0.00238175	USA	Comedy,Musical,Romance
12	Dr. Strangelove (1964)	0.00235889	USA	Comedy,War
13	Jaws (1975)	0.00233552	USA	Adventure,Thriller
14	Citizen Kane (1941)	0.0023169	USA	Drama,Mystery
15	Gone with the Wind (1939)	0.00231111	USA	Drama,Romance,War
16	Casablanca (1942)	0.00220961	USA	Drama,Romance,War
17	Touch of Evil (1958)	0.00214901	USA	Crime,Drama,Film-Noir
18	Chang: A Drama of the Wilderness (1927)	0.00213281	USA	Adventure,Documentary,Drama
19	The Searchers (1956)	0.00208214	USA	Adventure,Drama,Western
20	The Godfather (1972)	0.00201072	USA	Crime,Drama

Table 4: Nodes with the highest PageRank score

Rank	Film	Out-Degree	Country	Genres
1	Rewind This! (2013)	704	USA	Documentary
2	Adjust Your Tracking (2013)	562	USA	Documentary
3	Film Geek (2005)	502	USA	Comedy
4	Be Kind Rewind (2008)	455	USA	Comedy
5	American Grindhouse (2010)	293	USA	Documentary
6	Jersey Girl (2004)	261	USA	Comedy,Drama,Romance
7	Going to Pieces(2006)	257	USA	Documentary,Horror
8	That's Sexploitation! (2013)	236	USA	Documentary
9	Nightmares in Red, White and Blue(2009)	226	USA	Documentary,Horror
10	Cleanflix (2009)	219	USA	Crime,Documentary,History
11	Serial Mom (1994)	213	USA	Comedy,Crime,Thriller
12	A Decade Under the Influence (2003)	204	USA	Documentary
13	24x36: A Movie About Movie Posters (2016)	202	USA	Biography,Documentary,History
14	Los Angeles Plays Itself (2003)	192	USA	Documentary,History
15	There's Nothing Out There (1991)	188	USA	Comedy,Fantasy,Horror
16	Yes Man (2008)	185	USA	Comedy,Romance
17	The Benchwarmers (2006)	184	USA	Comedy,Sport
18	The Secret History of American Movies (2001)	175	USA	Documentary
19	The Holiday (2006)	166	USA	Comedy,Romance
20	Screaming in High Heels (2011)	150	USA	Documentary

Table 5: Nodes with the highest Out-Degree centrality

network and may have significant influence on the spread of ideas throughout the graph, as seen in Table 6.

7.1.5 Weighted average of the measures

Finally, we assigned weights to the In-Degree, PageRank, and Betweenness measures to reflect their relative importance in assessing node centrality. We believe that these measures

Rank	Film	Betweenness	Country	Genres
1	The Matrix (1999)	0.00118649	USA	Action,Sci-Fi
2	Pulp Fiction (1994)	0.00104666	USA	Crime,Drama
4	Scream (1996)	0.00085517	USA	Horror,Mystery
5	Serial Mom (1994)	0.00076564	USA	Comedy,Crime,Thriller
6	Natural Born Killers (1994)	0.00063377	USA	Action,Crime,Drama
7	Clerks (1994)	0.00055038	USA	Comedy
8	The Incredible Torture Show (1976)	0.0005167	USA	Comedy,Horror
9	Star Wars: Episode V (1980)	0.00045684	USA	Action,Adventure,Fantasy
10	Reservoir Dogs (1992)	0.00045271	USA	Crime,Thriller
11	Orgazmo (1997)	0.0004158	USA, Japan	Comedy
12	Shrek (2001)	0.00039918	USA	Adventure,Animation,Comedy
13	The Shining (1980)	0.00039899	USA	Drama,Horror
14	The Rocky Horror Picture Show (1975)	0.00039596	USA	Comedy,Horror,Musical
15	Raiders of the Lost Ark (1981)	0.00038883	USA	Action,Adventure
16	Desperado (1995)	0.00038395	USA	Action,Crime,Thriller
17	There's Nothing Out There (1991)	0.00035058	USA	Comedy,Fantasy,Horror
18	Kill Bill: Vol. 1 (2003)	0.00032161	USA	Action,Crime,Drama
19	Spider-Man (2002)	0.00030384	USA	Action,Adventure,Sci-Fi
20	Fight Club (1999)	0.00030297	USA	Drama
20	Star Wars (1977)	0.00029921	USA	Action,Adventure,Fantasy

Table 6: Nodes with the highest Betweenness centrality

capture different aspects of node importance and their contributions to the network dynamics. Our weight assignment is as follows:

- In-Degree: 0.6.
- PageRank: 0.8.
- Betweenness: 0.4.

7.1.6 Decisions and criteria

The Eigenvector centrality scores were excluded from the final score - and this study's results section - as we deemed that the **PageRank** assessment was a more appropriate algorithm to ascertain a film's influence because it takes into account the nature of *directed* links, ergo the in-citations to each node and their respective scores. Furthermore, although the Out-Degree centrality measures provided important insights to the entire study, we chose to exclude it for the simple reason that it is not related to a film's influence, but rather, in this context, to its contribution in the network.

Our weight assignment is based on a careful consideration of the unique characteristics and contributions of each centrality measure. We believe that **In-Degree** provides a good **starting point** for assessing node importance, while **PageRank** offers a more nuanced perspective by considering the **importance** of the nodes connecting to a given node. **Betweenness centrality**, although slightly less emphasized, still contributes valuable insights into the node's positioning and **influence** in the network's **information flow**.

With this decision, we aim to capture the multidimensional nature of node centrality in the citation's network analysis, to provide a more nuanced assessment of node importance in our network.

Rank	Film	Country	In-DC	PR	BC	CineMAC
1	The Wizard of Oz (1939)	USA	0.02756986	0.00758791	9.52E-05	1.40251242
2	Star Wars (1977)	USA	0.02899913	0.00455868	0.00029921	1.18077852
3	Psycho (1960)	USA	0.01564674	0.0043758	2.36E-05	0.7922602
4	Taxi Driver (1976)	USA	0.00913981	0.00164166	0.00097186	0.68841352
5	King Kong (1933)	USA	0.01015534	0.00428598	3.38E-05	0.67259395
6	The Matrix (1999)	USA	0.0075977	0.00080686	0.00118649	0.64065328
7	Dracula (1931)	USA	0.00582992	0.00491641	0	0.63832834
8	Pulp Fiction (1994)	USA	0.00801143	0.00083359	0.00104666	0.60489553
9	Jaws (1975)	USA	0.01293865	0.00233552	4.96E-05	0.52940021
10	The Godfather (1972)	USA	0.01293865	0.00201072	0.00014019	0.52563067
11	2001: A Space Odyssey (1968)	USA	0.01060669	0.00269954	5.32E-05	0.52085917
12	Frankenstein (1931)	USA	0.01008011	0.00290993	1.54E-06	0.51476394
13	Star Wars: Episode V (1980)	USA	0.0100425	0.00136834	0.00045684	0.50458284
14	The Shining (1980)	USA	0.01056908	0.00142452	0.00039899	0.50191039
15	Citizen Kane (1941)	USA	0.00887652	0.0023169	0.00020967	0.4973625
16	Gone with the Wind (1939)	USA	0.01102042	0.00231111	1.71E-06	0.47099782
17	Casablanca (1942)	USA	0.01098281	0.00220961	2.95E-05	0.46887217
18	The Terminator (1984)	USA	0.00966638	0.00171407	0.00025442	0.465089
19	The Birth of a Nation (1915)	USA	0.00203107	0.00392208	0	0.45465979
20	Snow White and the Seven Dwarfs (1937)	USA	0.00492722	0.00322657	0	0.44108887
20	Night of the Living Dead (1968)	USA	0.00895174	0.00192421	9.87E-05	0.41999777

Table 7: Total ranks for USA productions including the weighted scores.

Rank	Film	Country	In-DC	PR	BC	CineMAC
1	Gojira (1954)	Japan	0.006921	0.001502	0.000029	0.309822
2	Blade Runner (1982)	USA,Hong Kong	0.004852	0.000781	0.000105	0.216383
3	Shichinin no samurai (1954)	Japan	0.001881	0.001354	0.000009	0.183082
4	Enter the Dragon (1973)	Hong Kong,USA	0.003573	0.000814	0.000024	0.166181
5	Orgazmo (1997)	USA,Japan	0.000263	0.000032	0.000416	0.147241
6	Yôjinbô (1961)	Japan	0.001429	0.000581	0.000008	0.091998
7	Dip huet seung hung (1989)	Hong Kong	0.001204	0.000146	0.000156	0.091141
8	Otoko wa tsurai yo (1969)	Japan	0.001805	0.000509	0.000000	0.089310
9	Wo hu cang long (2000)	Taiwan, Hong Kong, USA, China	0.001768	0.000154	0.000078	0.077322
10	Hesokuri shachô (1956)	Japan	0.001504	0.000453	0.000000	0.077151
11	Sugata Sanshirô (1943)	Japan	0.000301	0.000623	0.000000	0.070278
12	Sora no daikaijû Radon (1956)	Japan	0.001617	0.000325	0.000004	0.067490
13	Zatôichi monogatari (1962)	Japan	0.001241	0.000375	0.000000	0.063469
14	Zoku otoko wa tsurai yo (1969)	Japan	0.001730	0.000268	0.000000	0.062315
15	Gojira no gyakushû (1955)	Japan	0.001166	0.000349	0.000009	0.062081
16	Akira (1988)	Japan	0.001392	0.000160	0.000050	0.060733
17	Mosura (1961)	Japan	0.001693	0.000234	0.000005	0.059429
18	Ringu (1998)	Japan	0.001354	0.000197	0.000035	0.058731
19	Miyamoto Musashi kanketsuhen (1956)	Japan	0.000113	0.000492	0.000006	0.054737
20	Otoko wa tsurai yo (1970)	Japan	0.001693	0.000187	0.000000	0.052937
20	Zoku hesokuri shacho (1956)	Japan	0.001429	0.000234	0.000000	0.052500

Table 8: Total ranks for East Asian productions including the weighted scores.

7.1.7 Discussion

The citations' network analysis results, as expected, reflect the structure of our dataset and the underlying impact of USA productions against East Asian films. Concerning the USA films, the centrality measures score rankings revealed a meaningful difference between a film's impact inside the cinematic network context, compared to other metrics such as box office revenues or viewer ratings.

The study also reinforced the idea that same region productions are more prone to influencing each other, which is reasonable given that the timespan studied is very wide, and in a large part of it features both the domination of Hollywood during a time where access and exposure to foreign films was more limited than today's.

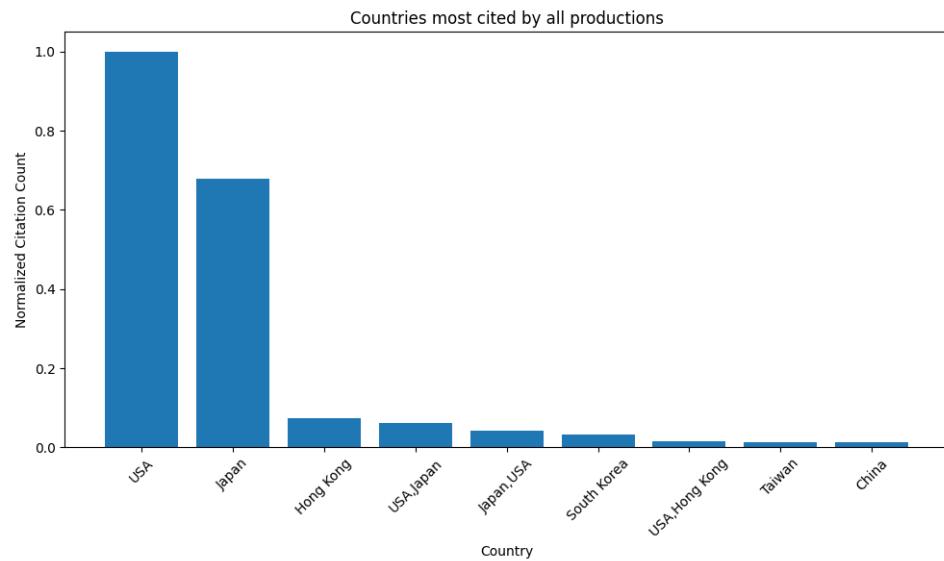


Figure 5: Countries most cited by all productions (USA and East Asian)

Films with a higher **in-degree** belong to an era described by as "New Hollywood Movies"¹⁴; occurring after the transition period of the 60s, during the decades of the 70s and 80s, these films gave popularity to now **world-wide known Hollywood filmmakers**, such as Steven Spielberg with Jaws, Raiders of the Lost Ark and E.T.: The Extra-Terrestrial, George Lucas with Star Wars, Francis Ford Coppola with The Godfather and Apocalypse Now, and Martin Scorsese with Taxi Driver. All of these films are listed in various ranks (such as the AFI Top100, the NFR, and the IMDb user rating Top250), have earned Academy Awards, yielded very good box office returns, and have the most incoming citations.

For what concerns **East Asian** only films, the ranking returns highly influential and famous movies, directed by renowned directors such as Akira Kurosawa, Ridley Scott, Bruce Lee, and Ang Lee. To investigate a few examples, Yôjinbô (1961), directed by Akira Kurosawa, is a masterpiece of samurai action that became a **benchmark** for the genre; Enter the Dragon (1973), not only brought enormous popularity to its starring actor, Bruce Lee, but also revolutionized

¹⁴Canet, F., Valero, M. A., & Codina, L. (2016). Quantitative approaches for evaluating the influence of films using the IMDb database.

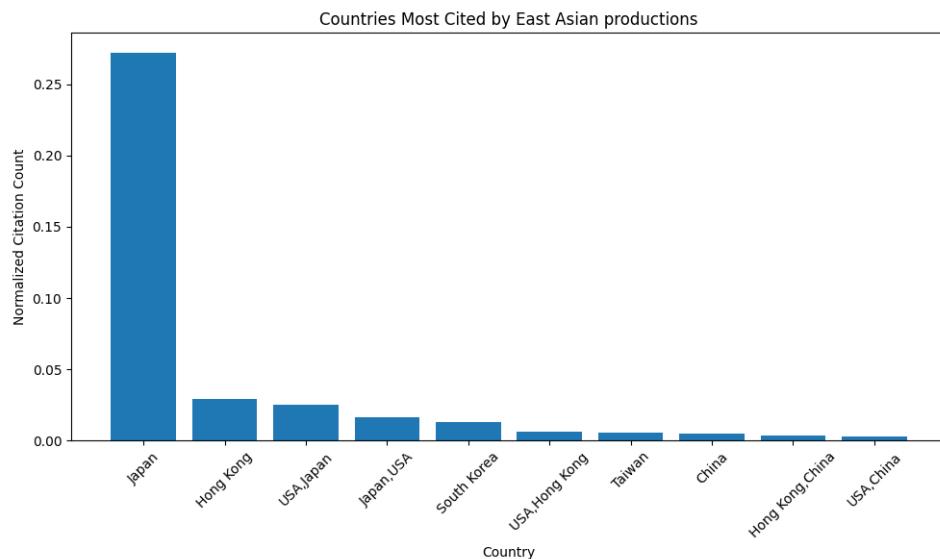


Figure 6: Countries most cited by East Asian productions

the martial arts genre and made it mainstream, influencing future generations of action films¹⁵. The rest of the films follow the same path, showcasing both artistry and intricate storytelling, and are responsible for bringing Asian cinema in the global market.

The PageRank based score returns a list of known "Classical Movies", with many Hollywood films from the 30s, 40s, and 50s, that feature **iconic scenes** (like "Wizard of Oz", "Casablanca", "Gone with the Wind", to name a few), along with unique, innovative Thriller and Sci-Fi films such as "Dracula", "King Kong", "Frankenstein", that paved the way for the **proliferation of such genres**. Overall, both categories were a stepping stone for the experimentation of other successful films in our graph, reflected by their in-degree centrality score.

The vast majority of the highest betweenness centrality scores are films of the 90s era, associated with proliferation of home viewing, the beginning of many movies that became a series of sequels, and the creation of common "**formulas**"¹⁶ belonging to Action, Crime, Sci-Fi, Horror and Comedy genres (as seen in Table 6), that influenced new generations of filmmakers. This rank features three films by Quentin Tarantino, who is well-known both for his movie references and his success as a director. As Peter Chumo points out, "*Tarantino's use of movie references goes beyond a simple postmodern recycling of old movie bits and generic plot lines to a thoughtful look at how such relics of the filmic past can come alive in the present*"¹⁷. Thus, we can infer that such directors do not only craft inspiration from existing films, but do so in a way that introduces a quality production that can enhance the **quality and diversity of experimentation** with cinematic references.

¹⁵Man-Fung Yip, Martial Arts Cinema and Hong Kong Modernity: Aesthetics, Representation, Circulation.2017, Hong Kong University Press, retrieved at <<https://hkupress.hku.hk/image/catalog/pdf-preview/9789888390717.pdf>>

¹⁶The History of Film The 1990s The Era of Mainstream Films and "Indie" Cinema, the Rise of Computer-Generated Imagery, the Decade of Re-makes, Re-releases, and More Sequels, retrieved in <<https://www.filmsite.org/90sintro.html>>

¹⁷Chumo, P.N. (1996). 'The Next Best Thing to a Time Machine': Quentin Tarantino's Pulp Fiction. Post Script: Essays in Film & the Humanities 15(3).

7.2 Collaborations' Network

7.2.1 Degree Centrality and Eigenvector Centrality

These ranks (see Table 9 and Table 10), highlight that all people in the top 20 rank are American, and specifically people that worked during the **classical Hollywood** cinema era, between the '20s and the '60s. This is a first hint into how **different historical production and collaboration patterns** are reflected in the network,¹⁸ as we will elaborate in the Critique section. Moreover, we notice how the Degree centrality rank returns more producers compared to the Eigenvector rank, as this is a role that may come with less prestige and fame, but certainly involves a **higher number of connections**, often with less influential people. On the contrary, directors and writers are more likely to be connected with less people, but more influential ones, as the Eigenvector rank reflects: we can infer that a producer will interact with several writers and directors, while a director may interact with only one or two producers throughout their entire career.

7.2.2 Closeness Centrality and Betweenness Centrality

Already from a first observation these measures appear more **diverse** nationality-wise, suggesting how in terms of influencing and cross-fertilization of cinema, East-Asian agents may emerge as more relevant and more well connected(Table 10, Table 11).

Weighted average of the measures (Table 12) This rank has been created assigning the following weights:

- **Degree Centrality:** 0.1
- **Eigenvector Centrality:** 0.2
- **Betweenness Centrality:** 0.4
- **Closeness Centrality:** 0.3

We notice how the rank provides a diverse outcome **nationality-wise**, even though none of the people in the top position are particularly renowned . We decided to investigate the outcoming rank excluding the producers, the role that, out of the three, is the **least linked to artistic** and creative aspects. As we suspected, the resulting rank displays people with a higher Eigenvector centrality, that are indeed more recognized(Table 14).

After retrieving the most influential film for each individual, we realized that, for a significant number of people, the movie was placed at the lowest places in the Citations' rank. To make sure that this was not due to very skewed results in weight assignment we tried comparing results with different weights: **Degree Centrality: 0.2 | 0.3** **Eigenvector Centrality: 0.3 | 0.3** **Betweenness Centrality: 0.1 | 0.2** **Closeness Centrality: 0.4 | 0.2**

Finally, we also computed the average of the measures without weights.

¹⁸Juhász, S., Tóth, G.,& Lengyel, B. (2020). Brokering the core and the periphery: Creative success and collaboration networks in the film industry. PLOS ONE, 15, e0229436.

The outcome did not significantly change for any of the tests, granted that it shows that top people in the collaboration rank are not necessarily connected to influential movies.

This result shows a strong pattern in how the two networks are related to each other, where **more influential people don't necessarily correspond to more influential films**. This could be explained by the fact that better connected people collaborated with more film agents because they were involved in the production of more films. Intuitively, one could hypothesize that quantity and quality are inversely proportional.

Beside two people - Michael Curtiz, director of *Casablanca* (CineMAC rank 8) and Ben Hecht, screenwriter of *Scarface*(1932, CineMAC rank 334) e “script doctor” di *Scarface*(1983, CineMAC rank 24) – every other person’s most influential film ranks lower than 4785 over 26849 films in the Citations’ network.

It's valuable to mention how the most influential films of the non-USA people present in our final rank are positioned:

Sang-hoon Ahn and Andy Yoon's *Beul-la-in-deu*(2011) positions 10755th (South Korea)

Gigi Qi's *Zhuo mi cang*(2016) positions 23718th (China), Peng Ren was also the screenplayer
Peng Ren's *Cui mian da shi* (2014) positions 24989th (China)

7.2.3 Connected components

In the Collaborations' network, 3477 connected components have been individuated, indicating that the network is highly fragmented and consists of multiple isolated groups or communities of individuals who have collaborated with each other but have limited or no connections with individuals outside their respective components.

Moreover, analyzing the size of the components we can see that there is one main large component of 28055 nodes, **suggesting a strong collaborative** network within a specific subset of the movie industry, while the second biggest component is already significantly smaller, comprehending only 21 nodes. The sum of the nodes in all the other components beside the main one is 9304, which represents 24% of the network, meaning that a not irrelevant subsection of the network is not well connected with the rest.

To better understand the interconnection of nodes in the network's bigger component we had to apply further measures such as the Clustering coefficient.

7.2.4 Clustering coefficient

The average clustering of the network's biggest component is 0.687, which suggests that the component has a tendency for nodes to be interconnected in local clusters, indicating the presence of cohesive subgroups or communities within the network.

7.2.5 Louvain community detection algorithm

The algorithm was able to identify 86 communities in the network, the biggest one having 4750 nodes, the smaller 3.

Looking into the top three community allows us to draw more solid interpretations on how the network is structured. It highlights a structural limitation of the network, namely the fact that it is composed by people who worked in **different time-spans** and under different production patterns

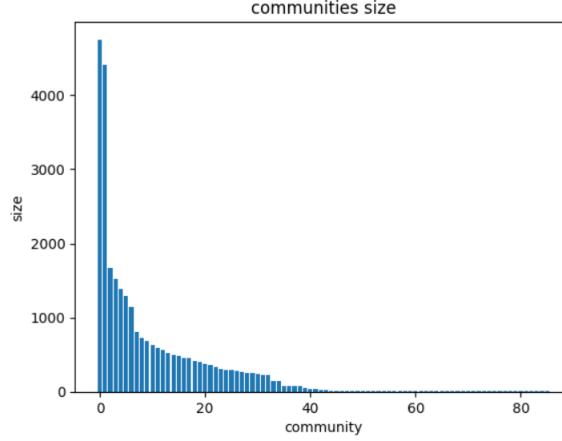


Figure 7: Size of communities in the collaborators’ network.

and paces. Unsurprisingly, the biggest community appears to be predominantly American and, from a first look, is composed of people that worked in earlier years. The second community is also mostly composed of American people but the members of this community worked in more recent years. The third and the fifth component, on the contrary, are composed predominantly of East-Asian people, and are also reflecting different time-spans. Finally, the fourth component appears to be composed of people with a diverse nationality range, for example including many European collaborators.

rank	name	roles	degree centrality
1	Ben Hecht	writer,director	0.006879383264628727
2	Michael Curtiz	director	0.005139461427271267
3	Raoul Walsh	director	0.004095508324856791
4	John Davis	producer	0.003988436211788639
5	Arnon Milchan	producer	0.0039081321269875255
6	John Huston	director,writer	0.0039081321269875255
7	Norman Taurog	director	0.00385459607045345
8	Darryl F. Zanuck	producer,writer,director	0.00385459607045345
9	Scott Rudin	producer	0.0038278280421864125
10	Mervyn LeRoy	director,producer	0.003720755929118261
11	Mенаhem Golan	producer,director,writer	0.0036136838160501093
12	William Beaudine	director,writer	0.0036136838160501093
13	John Ford	director,producer	0.0036136838160501093
14	William Shakespeare	writer	0.003586915787783072
15	Brian Grazer	producer,writer	0.003586915787783072
16	Dalton Trumbo	writer,director	0.003506611702981958
17	Yoram Globus	producer	0.0034798436747149205
18	Roy Del Ruth	director	0.0033995395899138066
19	Steven Spielberg	director,producer,writer	0.0032924674768456553
20	Jerry Wald	producer,writer	0.0032924674768456553

Table 9: Nodes with the highest Degree centrality

rank	name	roles	eigenvector centrality
1	Ben Hecht	writer,director	0.16769836434824734
2	Michael Curtiz	director	0.11702280101180426
3	Darryl F. Zanuck	producer,writer,director	0.09984826036129019
4	Norman Taurog	director	0.09353679952061277
5	Joseph L. Mankiewicz	writer,director,producer	0.09250452786362423
6	John Ford	director,producer	0.09137714078707712
7	John Lee Mahin	writer	0.09111510406994863
8	Mervyn LeRoy	director,producer	0.08580524228097451
9	Howard Hawks	director,writer	0.08564516199076809
10	George Cukor	director	0.08479871044656977
11	Richard Thorpe	director	0.0840986397534976
12	Jerry Wald	producer,writer	0.08117372688282819
13	George Marshall	director	0.07864400661304938
14	Roy Del Ruth	director	0.07840430433304797
15	Henry Hathaway	director	0.07797530705648269
16	Raoul Walsh	director	0.07727217077665534
17	W.S. Van Dyke	director,writer	0.07606355572952471
18	Sonya Levien	writer	0.07585057380098514
19	Jack Conway	director	0.07578508854820164
20	Herman J. Mankiewicz	writer	0.07331147136069123

Table 10: Nodes with the highest Eigenvector centrality

rank	name	roles	closeness centrality
1	William Shakespeare	writer	0.02606911777595779
2	Roger Corman	director,producer,writer	0.019059521608900093
3	John Davis	producer	0.017114773046759093
4	Richard Fleischer	director	0.016899771697342404
5	Takashige Ichise	producer,writer,director	0.015296300076494247
6	Menahem Golan	producer,director,writer	0.014969859729868602
7	Ben Hecht	writer,director	0.014218035735378269
8	Akira Kurosawa	writer,director	0.0121801170762063
9	Yoram Globus	producer	0.012128714292970768
10	Arnon Milchan	producer	0.01208197593191623
11	Steven Spielberg	director,producer,writer	0.01205003863363551
12	Robert Towne	writer,director	0.011743381076384625
13	Hark Tsui	director,producer,writer	0.011574292301976001
14	Stephen King	writer,director	0.011383899826026504
15	Brian Grazer	producer,writer	0.01137692908746441
16	Ishirō Honda	director,writer	0.011352399021130215
17	John Woo	director,writer,producer	0.011062367504596781
18	Scott Rudin	producer	0.01061693356733859
19	Sam Raimi	producer,director,writer	0.010544208933902528
20	Joel Silver	producer	0.010501650717345011

Table 11: Nodes with the highest Closeness centrality

rank	name	roles	betweenness centrality
1	Sang-hoon Ahn	writer,director	0.7083333333333334
2	Gigi Qi	producer	0.6538461538461539
3	Brian Gallagher	producer	0.6333333333333333
4	Darren Goldberg	producer	0.625
5	Tiffany Caprice	producer	0.59375
6	Andy Yoon	writer	0.5862068965517241
7	Anisa Qureshi	producer	0.5555555555555556
8	Xiaobai Gu	writer	0.5483870967741935
9	Shin Arum	producer	0.5483870967741935
10	Jae-Hoon Noh	producer	0.5483870967741935
11	B. Harrison Smith	director,writer	0.5277777777777778
12	Brant Sersen	director	0.5263157894736842
13	Kirk Roos	producer	0.5128205128205128
14	Eric England	director	0.5
15	Chris Marsh	producer	0.4878048780487805
16	Derek Lindeman	writer	0.48717948717948717
17	Jonathan Ilchert	producer	0.48717948717948717
18	John Guarneri	producer	0.48717948717948717
19	J.D. Lifshitz	producer	0.4857142857142857
20	Raphael Margules	producer	0.4857142857142857

Table 12: Nodes with the highest Betweenness centrality

Rank	previous	name	roles	DC	EC	CC	BC	CineMAC
1	12190	Sang-hoon Ahn	w,d	0.0	0.0	0.212	0.203	0.104
2	12187	Brian Gallagher	p	0.0	0.0	0.19	0.211	0.1
3	12186	Tiffany Caprice	p	0.0	0.0	0.178	0.22	0.1
4	12182	Darren Goldberg	p	0.0	0.0	0.188	0.208	0.099
5	12191	Gigi Qi	p	0.0	0.0	0.196	0.176	0.093
6	12194	Eric England	d	0.0	0.0	0.15	0.212	0.09
7	12183	Anisa Qureshi	p	0.0	0.0	0.167	0.158	0.081
8	12195	Craig Walendziak	w	0.0	0.0	0.131	0.176	0.077
9	12196	Eric B. Fleischman	p	0.0	0.0	0.134	0.156	0.073
10	12184	Kirk Roos	p	0.0	0.0	0.154	0.135	0.072
11	12188	B. Harrison Smith	d,w	0.0	0.0	0.158	0.129	0.072
12	12185	Brant Sersen	d	0.0	0.0	0.158	0.107	0.066
13	12197	Raphael Margules	p	0.0	0.0	0.146	0.093	0.06
14	12198	J.D. Lifshitz	p	0.0	0.0	0.146	0.093	0.06
15	12192	Peng Ren	w	0.0	0.0	0.142	0.088	0.057
16	12193	Andy Yoon	w	0.0	0.0	0.176	0.026	0.051
17	12189	Jonathan Ilchert	p	0.0	0.0	0.146	0.016	0.041
18	12199	Sean Tabibian	p	0.0	0.0	0.124	0.032	0.039
19	6	Ben Hecht	w,d	0.001	0.034	0.077	0.006	0.029
20	96	Michael Curtiz	d	0.001	0.023	0.072	0.002	0.025

Table 13: Nodes ranked after measures' weighted average

Rank	previous	name	roles	DC	EC	CC	BC	CineMAC
1	12190	Sang-hoon Ahn	w,d	0.0	0.0	0.212	0.203	0.104
2	12194	Eric England	d	0.0	0.0	0.15	0.212	0.09
3	12195	Craig Walendziak	w	0.0	0.0	0.131	0.176	0.077
4	12188	B. Harrison Smith	d,w	0.0	0.0	0.158	0.129	0.072
5	12185	Brant Sersen	d	0.0	0.0	0.158	0.107	0.066
6	12192	Peng Ren	w	0.0	0.0	0.142	0.088	0.057
7	12193	Andy Yoon	w	0.0	0.0	0.176	0.026	0.051
8	6	Ben Hecht	w,d	0.001	0.034	0.077	0.006	0.029
9	96	Michael Curtiz	d	0.001	0.023	0.072	0.002	0.025
10	59	Darryl F. Zanuck	p,w,d	0.0	0.02	0.074	0.003	0.024
11	65	Norman Taurog	d	0.0	0.019	0.073	0.002	0.024
12	141	Joseph L. Mankiewicz	w,d,p	0.0	0.019	0.074	0.002	0.023
13	0	William Shakespeare	w	0.0	0.005	0.077	0.01	0.023
14	142	George Cukor	d	0.0	0.017	0.073	0.002	0.023
15	196	Howard Hawks	d,w	0.0	0.017	0.072	0.001	0.023
16	27	John Huston	d,w	0.0	0.012	0.075	0.004	0.023
17	591	John Lee Mahin	w	0.0	0.018	0.071	0.001	0.023
18	293	Richard Thorpe	d	0.0	0.017	0.072	0.001	0.023
19	380	John Ford	d,p	0.0	0.018	0.07	0.001	0.022
20	130	Raoul Walsh	d	0.0	0.015	0.071	0.002	0.022

Table 14: Nodes ranked after measures' weighted average without producers

8 Conclusion

The citations' network analysis results, as expected, reflect the structure of our dataset and the underlying impact of USA productions against East Asian films. Concerning the USA films, the centrality measures score rankings revealed a meaningful difference between a film's impact inside the cinematic network context, compared to other metrics such as box office revenues or viewer ratings. The study also reinforced the idea that same-country productions are more prone to influencing each other, which is reasonable given that the time-span studied is very wide, and in a large part of it features both the domination of Hollywood during a time where access and exposure to foreign films was more limited than today's. Overall, the centrality measures accurately capture the artistic tendencies, cultural impact, and inter-textual references within the cinematic network, showcasing the diversity and evolution of film-making, for both USA and East Asian productions.

Our study provides a comprehensive demonstration of production practices in the movie industry throughout the years. It is notable that among the top 20 directors and writers in our measures scores, with the exception of Steven Spielberg, were active during the Classic Hollywood era. During this period, the focus was not on "author cinema," but rather on the mass production of movies, with directors producing a significant number of films, in contrast to renowned but less prolific artists like Stanley Kubrick (ranking at the 2074th position in the Degree centrality Rank and at the 3372nd in the Eigenvector) or Quentin Tarantino (ranking 999th in the Degree rank and 5796th in the Eigenvector).

9 Critique

We embarked in this project with ambitious research questions, wanting to unveil the dynamics of transnational collaborations and with the hope of being able to counterbalance the predominance of American Movie industry in most movie ranks. While our study gave insights to many **different interesting aspects** of cinema production and citations on a transnational level, it also highlighted the inequality of representation of East Asian movie industry with respect to the American industry. Unfortunately, given that this unbalanced representation springs from the data collected and the fact that we did not have enough information to predict any balancing factors, we were only partially able to counterbalance it, considering it **risky to arbitrarily impose a higher relevance** of East Asian films and personalities. Indeed, as mentioned in the measures discussion, a further development of this project could introduce weights to other factors such as a film's incoming citations from foreign industries.

To be able to create a comprehensive and balanced rank that gives justice to East Asian cinema production and properly highlights the long-term impact of East Asian films on a transnational level, a researcher would need to gather a dataset that reflects these kind of citations specifically. We envision that the best way to do it would probably be through crowdsourcing data gathering, using the same procedure through which the citations dataset used in the present study has been created but adding films present in East Asian movie databases. For example, *korean movie and drama database*, *Asian Wiki* and *Japanese movie database* are some sources that we could find but were less accessible and credited).

Another impediment that came from our data was the lack of nationality information for the people in the collaborations' network. Ideally it would have been useful to design a way of weighting relevant nodes that represent non-USA people, but the dataset was lacking these data. We did make an **attempt** to retrieve such information by means of API requests to wikidata, but this method was problematic for several reasons:

1. The large size of the network demanded a very high computational cost for such operation.
2. The lack of unique identifiers for people forced us to retrieve nationality on the basis of string matching, paving the way for a series of data errors such as homonyms or different nomenclatures that did not produce a match.
3. Reducing the graph to a subgraph including main components, although a viable option, was not ideal considering the networks' structures where the films and personalities were both parts of large but also smaller and isolated components.

Having this information would have enabled us to analyze connected components in a more meaningful way by retrieving same nationality components, for example, or to evaluate geographical representation nativeness in the main component compared to the one in the general network.

Lastly, It must be pointed out that due to the wide time-span covered by both the citations and collaborations network, they do not extensively account for different trends in **citational practices**, and **different cinema production practices over time**, respectively.

For the Citations' network, films cited by other films a long time after their release might be considered more influential than films from the same period citing each others. For the Collaborations network, the results reflect the intensity of American mass cinema production, especially during the Classic hollywood era. During this period, the focus was not on "author cinema," but rather on the mass production of movies, with directors producing a significant number of films, in contrast to renowned but less prolific artists like Stanley Kubrick (ranking at the 2074th position in the Degree centrality Rank and at the 3372nd in the Eigenvector) or Quentin Tarantino (ranking 999th in the Degree rank and 5796th in the Eigenvector). There are very influential people that produced a small number of films but tend to collaborate with the same group of people for the production of different films. For averaging the time span between movies that cite each other, a possible solution could be the one adopted by Wasserman et al who apply a correction factor: "they investigate the 'time lag' between edges in the connection network. For these authors, the time lag 'is the number of years between the releases of the edge's citing film and the release of the edge's cited film'" (2014b: 4).¹⁹

However, our study was able to make relevant East Asian films and personalities emerge from the networks we modeled, counterbalancing underrepresentation by filtering the data to exclude overrepresented nodes. To this extent, we presented in the Measures (section 6) the interpretation of relevant results with the underlying reasons for the higher position of USA personalities in the collaboration network, and by revealing the -often underrated- East-Asian movies that ranked high in our network. Overall, we believe our study could serve as an

¹⁹Wasserman, M., Zeng, X. & Nunes Amaral, L.A. (2014b). Cross-Evaluation of Metrics to Estimate the Significance of Creative Works. Proceedings of the National Academy of Sciences of the United States of America, 1–6.

interesting idea towards the further development of alternative metrics using network theory, for evaluating the world of cinematic production and its complexities.