COMMUNICATION & SOCIETY

Quantitative approaches for evaluating the influence of films using the IMDb database

Fernando Canet
fercacen@upv.es
Associate Professor in Film Studies. Fine Arts College. Politechnic University of Valencia. Spain.

Miguel Ángel Valero
miguel.valero@upv.es
Researcher. Politechnic University of Valencia. Spain.

Lluís Codina
lluis.codina@upf.edu
Associate Professor. Pompeu Fabra University. Barcelona. Spain.

Submitted
February 2, 2016
Approved
March 15, 2016

ISSN 0214-0039
E ISSN 2386-7876
doi: 10.15581/003.29.2.151-172
www.communication-society.com

2016 – Vol. 29(2),
pp. 151-172

How to cite this article:
Canet, F., Valero, M.A. & Codina, L. (2016). Quantitative approaches for evaluating the influence of films using the IMDb database. *Communication & Society 29*(2), 151-172.

This work is part of the Project "Active Audiences and Journalism. Interactivity, Web Integration and Findability of Journalistic Information". CSO2012-39518-C04-02. National Plan for R+D+i, Spanish Ministry of Economy and Competitiveness

Abstract

Why do films certain remain influential throughout film history? The purpose of this paper is to attempt to answer this question. To do so, we adopt some quantitative approaches that facilitate an objective interpretation of the data. The data source we have chosen for this study is the Internet Online Movie Database (IMDb), and in particular, one of its sections called "Connections", which lists references made to a film in subsequent movies and references made in the film itself to previous ones. The extraction and analysis of these networks of citations allows us to draw some conclusions about the most influential movies in film history, identifying their distinguishing features, and considering how their popularity has evolved over time.

Keywords

Influential films, film history, IMDb, connections, database, quantitative approaches

1. Introduction

What are the most influential films in film history? What features define these films? Why films do certain remain popular over time? In other words, why are some films kept alive in the collective memory while others are consigned to oblivion? The main purpose of this paper is to attempt to answer these questions. To do so, we adopt some quantitative approaches that facilitate an objective interpretation of the data. The data source we have chosen for this study is the Internet Online Movie Database (IMDb), for two reasons: first, it is the biggest and the most well-known movie database on the Internet, and has an extensive range of descriptive categories that allow for a multilayered analysis; second, a significant number of previous studies have used this dataset, which supports its value as a resource for scientific research.

One of these previous studies is by Max Wasserman, Xiao Han T. Zeng and Luís A. Nunes Amaral, who, in their paper titled "Cross-evaluation of metrics to estimate the significance of creative works", define the concept of significance as "the lasting importance of a creative work". According to these authors, a film's significance can be measured by taking three factors into account: quality, impact and

influence. While the concept of quality is self-explanatory, the other two terms require some clarification. Impact is defined by the authors as "the overall effect of a creative work on an individual, industry, or society at large", which "can be measured as sales, downloads, media mentions, or other possible means." Influence, on the other hand, is defined "as the extent to which a creative work is a source of inspiration for later works" (2014b: 1). In this article, we will be focusing mainly on this last factor, as we consider "influence" to be a central concept for explaining the significance of a movie over the course of film history.

Filmmakers are thus the ones who determine the influence of films over time, when they look back on their own film heritage and choose one movie or another as a source of inspiration for their own works. According to Marijke de Valck and Malte Hagener (2005: 15), this tendency to use film history "as a limitless warehouse that can be plundered for tropes, objects, expressions, styles, and images from former works" is a well-known phenomenon. In any academic milieu, the most influential scholars are the ones whose works are quoted most by their colleagues in subsequent articles. Hence the current popularity of the "h index", which attempts to measure a researcher's impact through the citations that his or her works receive. It would seem logical that if we can define "influential articles" in this way, we can similarly speak of "influential movies" as those which are cited most often in subsequent films.

However, while scholars explicitly state their sources in their texts, filmmakers do not, which poses the problem of how to identify citations made by filmmakers and who is responsible for identifying them. The obvious answer is that this responsibility falls upon the audience. Thus, in cinema, a citation is only effective if someone recognizes it, as filmmakers invite spectators to play a game of identifying what is familiar. Sigmund Freud describes this discovery of the familiar as a source of human pleasure,[1] while Umberto Eco (2005: 108) suggests that it activates the spectator's "encyclopedic competence". In this respect, it is important to note that the recognition of a cinematic citation depends on the spectator's personal competence, a fact also pointed out by film scholars such as Vera Dika (2003: 103) and Pam Cook (2005: 168).

In recent years, the potential for encyclopedic competence has been greatly enhanced by the Internet and the birth of social networks that invite users to participate. IMDb, for example, is not only a database but also a social network whose users share their opinions and knowledge, contributing content to the largest digital collection of data not only on films but also on television programs and video games. Citations recognized by spectators are collected on IMDb under its category of "Connections". Thus, as Wasserman et al. point out, "by analyzing this citation network obtained from user-edited data, we can investigate the suitability of metrics to estimate film significance based on the spread of influence in the world of motion pictures" (2014b: 2). It is precisely this that is our main objective in this paper: through the extraction and analysis of these data, to draw some conclusions about the most influential films in cinema history, identifying their distinguishing features, and considering how their popularity has evolved over time. These quantitative results will be interpreted from a film studies perspective, which will facilitate a better understanding within the context of film history.

2. Intertextuality and allusion to film history

The idea of viewing film history as a network of citations naturally leads us to the concept of "intertextuality", a term coined by Julia Kristeva in 1966, drawing on Mikhail Bakhtin´s concept of dialogism, referring to the idea that "any text is constructed as a mosaic of quotations" (Kristeva, 1986: 37). For Roland Barthes "the text is a tissue of quotations," (1977:

[1] See Freud's 1905 essay "Jokes and the Unconscious" (SE: Vol. VIII: 128).

ISSN 2386-7876 – © 2016 Communication & Society 29(2), 151-172

46-47) while Umberto Eco notes that "a given text echoes previous texts" (2005: 197). Gerard Genette identifies intertextuality as one of the five categories that fall under broader concept of transtextuality. In his book *Palimpsest*, Genette defines transtextuality as "all that sets the text in a relationship, whether obvious or concealed, with other texts." For our purposes here, Genette´s most suggestive category is hypertextuality, defined as "any relationship uniting a text B (which I shall call the hypertext) to an earlier text A (I shall, of course, call it the hypotext)" (1997: 1-6). Based on this idea, for this study films can be classified in two categories: the films cited (text A), and the films citing them (text B), all of which form the connected nodes that shape this network of citations. These connections between movies, as Graham Allen suggests, move out "from the independent text into a network of textual relations" so that "[t]he text becomes the intertext" (2006: 1).

The nodes of this network are the movies, while the links between them are the connections made by filmmakers, who in so doing are employing the expressive device referred to by Noël Carrol as "allusion to film history", which he defines as "an umbrella term covering a mixed lot of practices including quotations, the memorialization of past genres, the reworking of past genres, homages, and the recreation of 'classic' scenes, shots, plot motifs, lines of dialogue, themes, gestures [...]" (1982: 52). Again, as noted above, it is the spectators who are tasked with identifying these allusions. In this way, intertextuality calls upon the participation of both parties in the filmic experience: the filmmakers, who allude to their film heritage, and the spectators, who recognize the allusion. In both cases, an awareness of film history is required.

3. The study of reception using IMDb: challenges and possibilities

Clearly, as Vivian Sobchack suggests, "IMDb is an important resource for improving our understanding of spectator response" (2013: 38). A great deal of research, conducted mainly in the last three years, has made use of this dataset for which the spectator's participation is crucial, pursuing different objectives and applying different methodologies. Below is a review of this research, classified according to the different attributes analyzed in each study.

One of the more widely studied IMDb attributes is that of film user reviews, which collects opinions offered by spectators about movies (and is therefore is not numeric data). Klaus Dodds has explored how spectators engage with *The James Bond* film series with the objective of better understanding how they interpret the geopolitics of these films (2006). In the same vein, Juha Ridanpää also analyzes geopolitical issues, but in this case focusing on the political nature of humor by dissecting the IMDb film reviews of Sacha Baron Cohen's comedy, *The Dictator* (2012). Ridanpää also investigates how IMDb can become a platform for political participation (2014). On the other hand, Jahna Otterbacher, focusing on gender issues, uses logistical classification to compare reviews written by men and women with respect to writing style, content and metadata features. As predicted by sociolinguistic theory, she finds differences between genders in stylistic features and content of the reviews analyzed (2013).

Another methodology used in these studies is "opinion mining", which uses information technologies to analyze the user opinions. This methodology has mainly been used to analyze attitudes towards products and services as expressed in user reviews. For instance, Jonas Krauss *et al.* use this approach to apply a model for predicting a movie's success (success being defined according to box office returns and Academy Award nominations). They measure user opinions according to two variables: "intensity" (meaning the frequency with which a film is discussed); and "positivity" (i.e. the level of positivity in the attitudes expressed towards a movie) (2008). However, as Yang Liu *et al.* point out in their paper, this research does not consider the fact that attitudes change over time. They try to fill this gap

ISSN 2386-7876 – © 2016 Communication & Society 29(2), 151-172

by proposing two adaptive methods to capture the evolution of attitudes expressed by users in their reviews. In addition, they propose to apply these methods to sales predictions. To this end, they compare selected user reviews of films with their box office earnings (2013). In another study, Jason J. Jung also applies mining methods, in this case for identifying "short-head" users. The opinions of these expert users are very important for improving the effectiveness of recommendation systems (2012).

Another attribute analyzed in this kind of research is user rating. On the IMDb database, user opinions on movies are quantified on a scale from 1 to 10. Like Jung in his research, Sung Moon Bae *et al.* focus on recommendation systems; however, in this case, the IMDb user rating is the category to be applied to the improvement of predictions (2014). The development of a user rating prediction model is also the objective of Ping-Yu Hsu *et al.*, for which they also adopt the technique of data mining. In both studies, the researchers seek to demonstrate that user rating is a good indicator for predicting box office success (2014). In 2004, Saraee *et al.* had already begun using data mining techniques to analyze user rating, which they also argued was a good indicator of a film's popularity. With this approach they sought, on the one hand, to determine whether big-budget films are more popular than low-budget ones, and, on the other, to identify whether there was a "golden age" of film (2004). Francisco Fraile and Juan Carlos Guerri also connected user rating to popularity, showing how parametric popularity models can be linked to user rating distribution. In their case, they found that the identification of a film's popularity is very useful for determining the best system for delivering it to viewers (2014).

Miscellaneous attributes have been another popular category for analysis in recent research. This category includes research that uses more than one attribute. Due to the large number of attributes associated with each movie, rather than examining only one attribute, Khalid Ibnal Asad *et al.* propose a suitable approach for identifying which attributes are useful for evaluating the pre-release popularity of a movie, as well as its post-release prospects. Correlation coefficients are established between datasets in order to predict the popularity and commercial success of movies (2012). Similarly, S. Kabinsingha *et al.* study a range of attributes in order to classify movies. In their case, the objective of classification is to propose movie ratings as a guide for parents. The most useful attributes for this purpose are genres and words used in movie reviews (2012).

Different attributes can be analyzed for a diverse range of purposes. For instance, Jungsik Park *et al.* examine synopses and movie keyword tags in order to identify the major recurrent themes in Hollywood action movies released from 1930 to 2009 (2014). Sameet Sreenivasan uses crowdsourced keywords in order to analyze the evolution of novelty patterns in movies over time (2013). Meanwhile, Luca Canini *et al.* use the IMDb genre category to test their proposed movie classification system based on different emotional patterns established by mapping filming and editing techniques, such as light source color, motion dynamics and audio track energy (2009). On the other hand, Randy A. Nelson and Robert Glotfelty examine the usefulness of movie star status for predicting the success of movies at the box office. They propose a measurement of the influence of the film star based on the number of visits to his or her web page on IMDb (2012).

Three other recent studies all explore the classification of co-actor networks, which connect actors based on the movies in which they have been cast together. Bruce W. Herr *et al.* propose different ways of visualizing the complex co-actor network that take into account movie classifications (2007). Similarly, Adel Ahmed *et al.* propose a network analysis method that integrates visualization with the aim of evaluating this kind of relationship system (2007). On the other hand, Lazaros K. Gallos *et al.* consider the co-actor network as a small-world prototype and propose their own fractal method for evaluating it as such (2013).

Finally, of special relevance to our research are two studies conducted largely by the same group of researchers. Both published in 2014, these two studies are to our knowledge

ISSN 2386-7876 – © 2016 Communication & Society 29(2), 151-172

the only ones that use the "connections" attribute as their dataset. Wasserman *et al.* begin their analysis of IMDb by exploring the possibilities of diverse metadata, such as the number of user votes, production budget or box office returns, as well as proposing correlations between these attributes. As a novelty, they also introduce "connections" and the type of relations between movies listed in this section (2014a). However, it was not until their next paper on this topic that they took full advantage of the potential of the connections classification. As noted in the introduction, Wasserman *et al.* use the metadata taken from the connections category as quantifiable variables for estimating a film's significance. In order to confirm the accuracy of this approach, they consider other sources that may also reflect the significance of a movie, such as the movies selected for preservation in the National Film Register[2]. For this purpose they apply a cross-evaluation of metrics based on three different approaches: expert opinions, "wisdom of the crowd" and automated measures. The procedure followed by Wasserman *et al.* is the main reference for our research.

4. The "Connections" classification

The connections classification on IMDb can be separated into two categories: on the one hand, the references made to a film in subsequent movies; and on the other, the references made in this film to previous ones. While the former measures the influence that a film exerts on the future, the latter measures the influence that the cinematic past has exerted on it. For Mikhail Iampolski, the "semantic fullness" of a film can be best understood by analyzing its possible connections in both directions (1998: 100). As Wasserman *et al.* note, "[i]n network theory, these values are known, respectively, as the in-degree and out-degree of a node," (2014a: 4) the former referring to "incoming connections" and the latter to "outgoing connections". Incoming connections (the references made to a film in subsequent movies) are measured on IMDb through seven variables: "Followed by", "Remade as", "Edited into", "Spin off", "Referenced in", "Featured in" and "Spoofed in"; while their respective counterparts measure outgoing connections (references made in this film to previous ones): "Follows", "Remake of", "Edited from", "Spin off from", "References", "Features" and "Spoofs".

To clarify how these seven variables are used, we will take as an example the unquestionable classic *Psycho* (Hitchcock, 1960). According to the "connections" category, *Psycho* was followed by three movies: *Psycho II* (1983), *Psycho III* (1986), and *Psycho IV: The Beginning* (1990), the first two being sequels, and the latter being a prequel. In 1998, Gus van Sant released his controversial remake of *Psycho*. Two spin-offs have sprung up from Psycho, the film *Bates Motel* (1987) and the recent TV series of the same name (2013-). Hitchcock´s movie has been referenced in 920 subsequent films, the famous shower scene being the main source of this popularity. Actual footage from *Psycho* has been edited into six other movies (*Psycho II*, for example, opens with the original shower scene), and extracts from *Psycho* have been featured in another 109. In this case, unlike the "edited into" variable, it is clear that the extract is from another movie. The extract may be viewed by the film's characters on TV or in a movie theatre, or featured in a review program[3]. *Psycho* has been spoofed in 170 movies[4]. In short, *Psycho* is an extremely influential movie, since it has 1,212 incoming connections but only 3 outgoing connections.

---

[2] Since 1988, the US National Film Preservation Board, as part of the Library of Congress, is the institution in charge of the selection of American films for preservation.

[3] For instance, Mr. Gold in the pilot episode of Dawson´s Creek (1998) is watching *Psycho*, and an extract of this movie is shown in From the Journals of Jean Seberg (1995) as a way of illustrating the madness of men on film.

[4] For example, in the *The Simpsons* episode "Itchy and Scratchy and Marge" (1990), the shower scene is accurately recreated shot-by-shot when Maggie knocks Homer unconscious with a mallet.

ISSN 2386-7876 – © 2016 Communication & Society 29(2), 151-172

Unlike Wasserman *et al*, our analysis is not limited only to references, spoofs and features, but takes into consideration all seven variables to measure film influence. Although, as illustrated by the above example of *Psycho*, these three variables have greater weight than the other four, we have taken the decision to include all of them in our study since each variable in one way or another reflects connections to subsequent movies and is therefore representative of a movie's influence over time.

5. Data

The first task of this research involved the collection of appropriate shared metadata provided by IMDb through its file sharing system[5]. We retrieved data from IMDb on 3,135,600 nodes, including films, TV series, video games and other types of media. Of these, barely 3% have connections, so that our initial dataset was 89,551 nodes with 811,609 connections. This data was collected in January 2015 and was processed during February and March 2015. From these nodes, we selected the 100 movies with the most incoming connections, identifying them as the 100 most influential movies (Top100_I). These 100 nodes have 58,274 incoming connections and only 1,344 outgoing connections, which is why they are considered influential movies. Scripting programming techniques provided on-line querying capacities to obtain only important data related to the chosen sample. Finally, we developed a small local MS Access database, powerful enough to handle SQL (Structured Query Language) queries such as grouping, total summing and comparison operations, for further analysis.

6. Results and discussion

Table 1. Top100_I

Year	Movie	Incoming
1977	Star Wars	3592
1939	The Wizard of Oz	2570
1972	The Godfather	1339
1980	Star Wars: The Empire Strikes Back	1289
1960	Psycho	1212
1975	Jaws	1141
1942	Casablanca	1118
1981	Raiders of the Lost Ark	1044
1982	E.T. The Extra-Terrestrial	964
1984	The Terminator	950
1939	Gone with the Wind	917
1968	2001: A Space Odyssey	906
1933	King Kong	842
1983	Star Wars: Return of the Jedi	825
1984	Ghost Busters	822
1931	Frankenstein	803
1997	Titanic	794
1999	The Matrix	783
1979	Apocalypse Now	770

[5] Available at www.imdb.com/interfaces

ISSN 2386-7876 – © 2016 Communication & Society 29(2), 151-172

1980	The Shining	757
1994	Pulp Fiction	739
1973	The Exorcist	728
1985	Back to the Future	722
1979	Alien	705
1976	Rocky	698
1993	Jurassic Park	667
1937	Snow White and the Seven Dwarfs	666
1976	Taxi Driver	632
1941	Citizen Kane	621
1991	The Silence of the Lambs	616
1983	Scarface	613
1991	Terminator 2: The Judgment Day	605
1968	Night of the Living Dead	599
1984	A Nightmare on Elm Street	588
1946	It´s a Wonderful Life	585
1965	The Sound of Music	572
1954	Gojira	565
2001	The Lord of the Rings: The Fellowhip of the Ring	565
1999	Star Wars: The Phantom Menace	555
1971	A Clockwork Orange	539
1994	The Lion King	519
1988	Die Hard	492
1964	Goldfinger	483
1978	Superman	479
1974	The Texas Chain Saw Massacre	476
1978	Halloween	473
1989	Batman	473
1982	First Blood	463
1964	Mary Poppins	462
1986	Aliens	456
1971	Dirty Harry	452
1968	Planet of the Apes	440
1986	Top Gun	440
1931	Dracula	431
1967	The Graduate	431
1966	The Good, the Bad and the Ugly	422
1942	Bambi	417
1994	Forrest Gump	416
1984	The Karate Kid	414
1977	Saturday Night Fever	393
1980	Friday the 13th	389
1940	Pinocchio	382
1958	Vertigo	382
1982	Blade Runner	377

ISSN 2386-7876 – © 2016 Communication & Society 29(2), 151-172

1987	RoboCop	374
1964	Dr. Strangelove or: How I Learned to	371
1972	Deliverance	365
2008	Twilight	363
1978	Grease	361
1990	Goodfellas	357
1992	Reservoir Dogs	355
1952	Singin´ in the Rain	352
1977	Close Encounters of the Third Kind	350
1989	The Little Mermaid	349
1922	Nosferatu	345
1996	Jerry Maguire	345
1984	Indiana Jones and the Temple of Doom	344
1961	West Side Story	343
1975	Monty Python and the Holy Grail	343
1995	Toy Story	331
2009	Avatar	330
1999	Fright Club	324
1976	Carrie	319
1981	The Evil Dead	319
1987	Predator	318
1987	Full Metal Jacket	316
1968	Rosemary's Baby	312
1988	Rain Man	309
1975	One Flew over the Cuckoo´s Nest	308
1999	The Blair Witch Project	308
2001	Harry Potter and the Sorcerer´s Stone	302
1978	Dawn of the Dead	296
1935	Bride of Frankenstein	293
1950	Sunset Blvd	288
1927	Metropolis	287
1940	Fantasia	286
1962	Dr. No	284
1992	Aladdin	281
1995	Braveheart	281
1953	Peter Pan	279

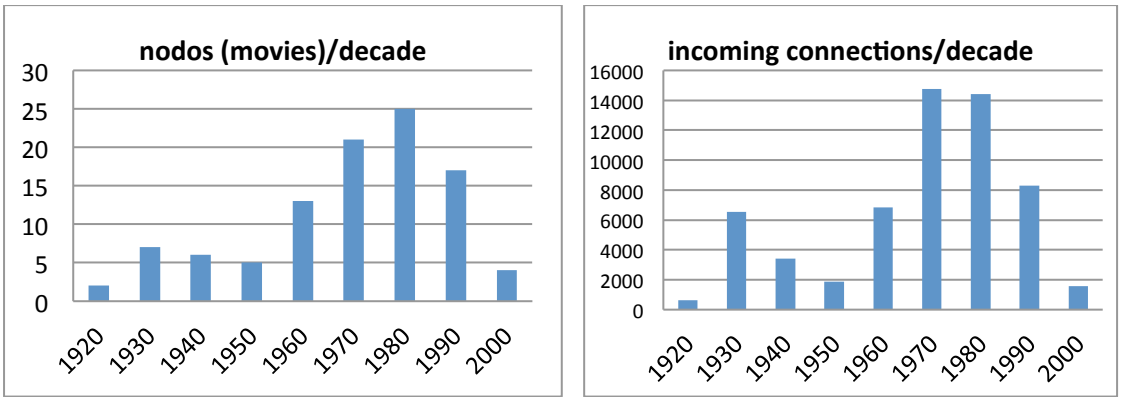
6.1. *The allusionist cinematic practice*

As shown in the table (Top100_I), the most influential movie in the history of film is *Star Wars*, with nearly 1,000 incoming connections more than the movie that ranks second. We can even go a little further by saying that the *Star Wars* series is the most influential, since three of its movies are among the Top100_I. In second place is *The Wizard of Oz* and in the third is *The Godfather*. The preponderance of US cinema is overwhelming: 90 movies are from the United States, 3 are UK/US co-productions (all of which could also be considered US films since their director is Stanley Kubrick), and only 7 are foreign films (3 from the UK, 2 from Germany, 1 from Japan and 1 from Italy). With respect to the most popular decades,

ISSN 2386-7876 – © 2016 Communication & Society 29(2), 151-172

using the number of movies per decade or the summation of incoming connections, the 70s and 80s are the most influential (see Figure 1). However, the order of these two decades changes depending on the value applied; thus, while the 70s has more total connections, the 80s produced more movies (nodes) in the Top 100. The reason for this is that the 70s produced two of the three most influential films: *Star Wars* and *The Godfather*. Likewise, the weight of the 30s is more significant in the connections graphics because of the presence of *The Wizard of Oz*, which skews the data on the importance of this decade.

Figure 1



The Hollywood revival of the 70s, known as "New Hollywood" or "Post-classical Hollywood", prevails in the list over the golden age of classical Hollywood cinema. Movies made in the 40s and 50s have clearly given way to their successors. The best-known of the New Hollywood filmmakers, who began their careers in the 70s, have directed the most influential movies listed in Top100_I. Steven Spielberg, with six movies, and James Cameron and Stanley Kubrick, both with five, stand out among them.

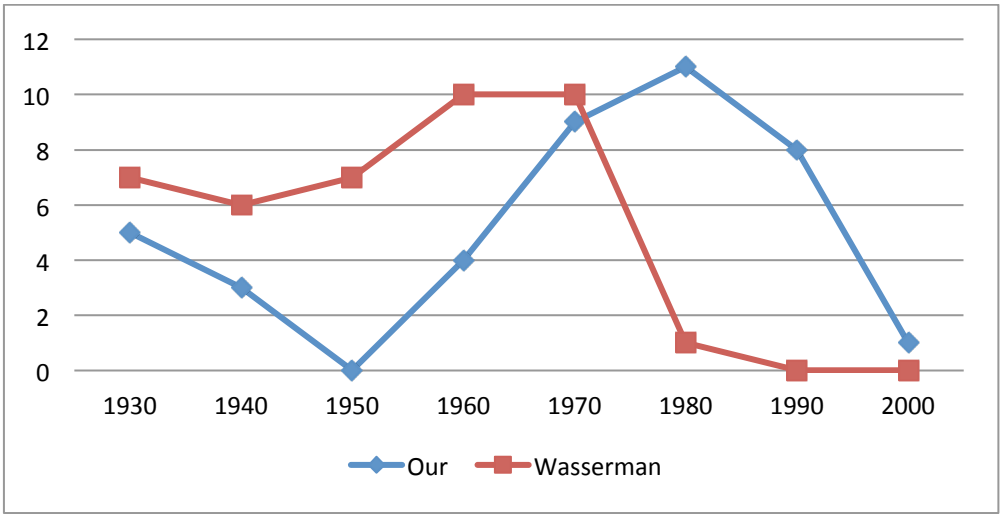
Wasserman *et al.* also detect this tendency in their analysis, but only up to the 70s. The reason for this difference is that they apply a correction factor. In order to research how connections operate over time, they investigate the "time lag" between edges in the connection network. For these authors, 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). Their results demonstrate that more films receive "shorter-gap citations" (i.e. less than 20 years between cited film and citing film), while the frequency of citations decreases as time goes on. However, and very importantly, there is one kind of film that does not follow this trend, since for this type of film it seems as if "timeliness does not matter" (2014b: 5). These are films which, despite the passage of time, remain influential on future generations, and it is therefore these films that the research aims to identify. Wasserman´s results show that for time lags of 25 years or more the likelihood of receiving connections increases again. For this reason, they propose to use only the citations that a movie receives in other movies released 25 or more years later to determine its significance. However, this decision is problematic, as in one fell swoop it eliminates movies from the 80s and 90s. By examining the data more closely, our aim is to determine whether the behavioral patterns of certain movies made in the 80s and 90s are comparable with older ones, which would consequently mean that they should be considered equally influential.

Wasserman *et al.* only list 41 movies and so for the purpose of making comparisons we had to reduce our list to this same number. As shown in Figure 2, there is a total coincidence of movies from the 70s, making this decade the period that modulates the two different distributions. While Wasserman´s distribution gives more weight to earlier decades, our

ISSN 2386-7876 – © 2016 Communication & Society 29(2), 151-172

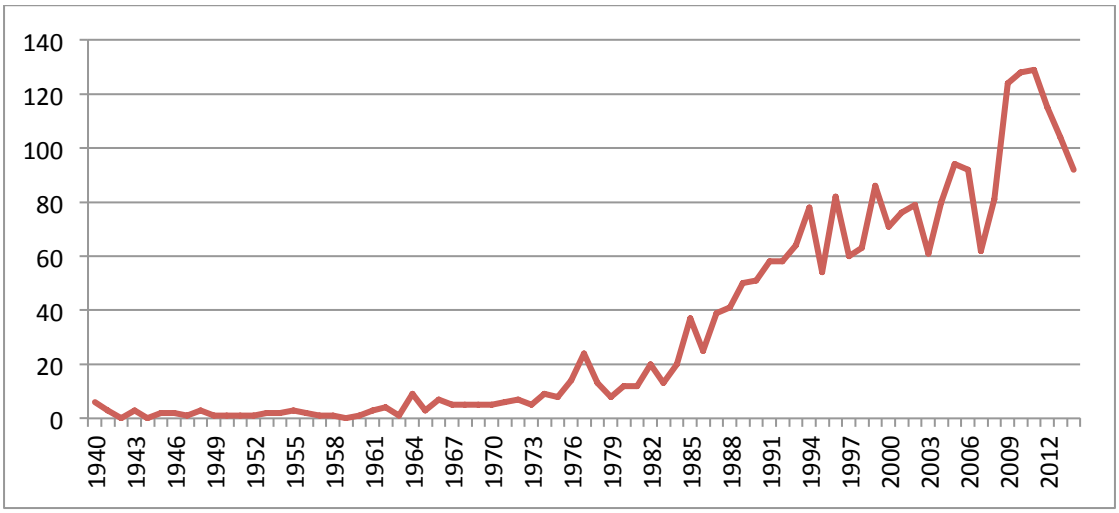
distribution gives more relevance to subsequent decades. Wasserman's list includes hardly any movies from the 80s or 90s (only one from the 80s), whereas our list includes none from the 50s. The two lists only coincide in 19 movies. However, our Top100_I list includes all of the films on Wasserman´s list except six, which are mainly from the 50s.

Figure 2. Nodes (movies)/decades.
Comparison between Wassserman's list and our proposal



Movies from the 30s and 40s don't become influential exactly 25 years after their release; for example, while *The Wizard of Oz*´s 25-year time lag starts in 1964 (see Figure 3), this film doesn't begin an upward tendency in citations until 1973.

Figure 3. The Wizard of Oz (1939). Time lag 25y (1964)



For films from the 50s and 60s this gap becomes smaller and even in some cases coincides exactly with the 25-year mark, as in *Vertigo* (see Figure 4) or *2001: A Space Odyssey*. Indeed, as early as 1960, the onset of this tendency may be earlier than the 25-year point; for instance, *Psycho* sees an upward trend in connections in 1975 whereas its 25 year time lag is

ISSN 2386-7876 – © 2016 Communication & Society 29(2), 151-172

in 1985 (see Figure 5). Thus, citation patterns in these movies are dependent not so much a certain time lag as due to the emergence of the allusionist cinematic practice in the 70s and 80s.

Figure 4. Vertigo (1958). Time lag 25y (1983)

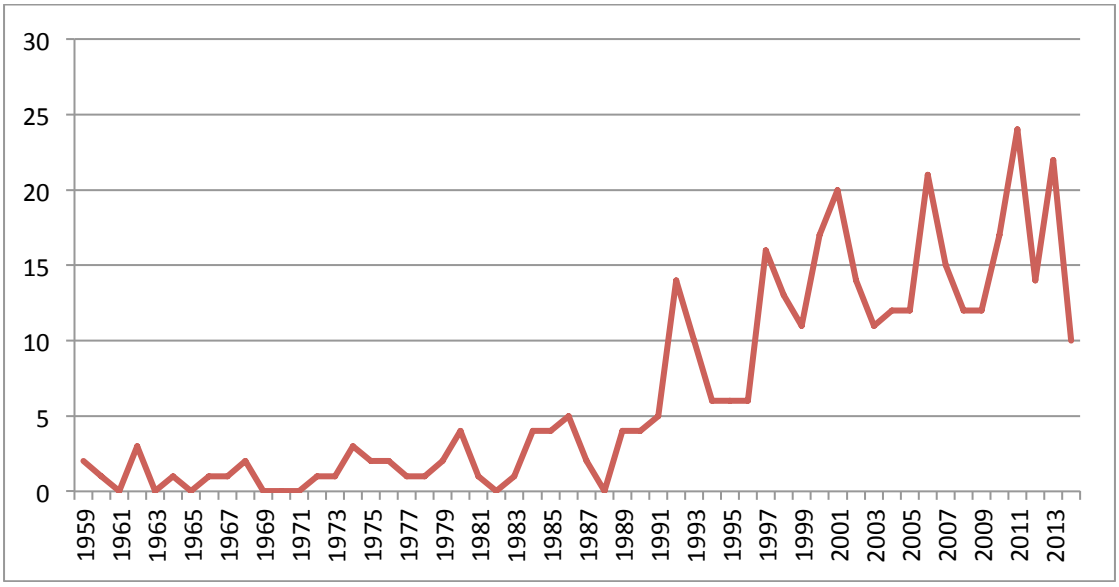
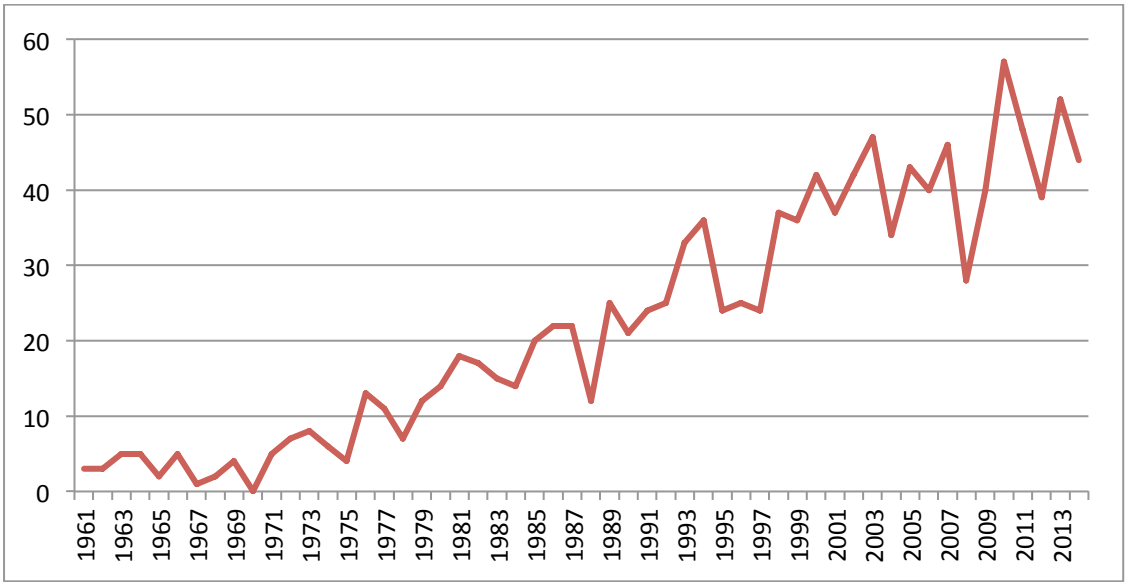


Figure 5. Psycho (1960). Time lag 25y (1985)



As Brian Ott and Cameron Walter point out, "in the early 1980s, media critics observed that films and television shows had increasingly begun quoting and referencing other popular cultural artifacts" (2000: 429). According to Carroll, this practice begins earlier, in the 60s and early 70s, as a trend that he defines as "the boom of allusionism" (1982: 54). However, this “boom” was actually the modest beginning of a bigger, more sustained trend that developed further in the following decades. If we take a look at the distribution of the outgoing connections of the Top100_I, we can see how this tendency begins in the 70s and

ISSN 2386-7876 – © 2016 Communication & Society 29(2), 151-172

80s but culminates in the 90s (see Figure 6). On the other hand, if we consider incoming connections, the boom moves very significantly to the first decade but especially the early part of the second decade of this century, reaching its peak in 2011 (see Figure 7). It is worth noting that a significant number of quotes are from TV series, which have increased exponentially in popularity over the last two decades. We argue that the movies from the 30s through to the 60s are not so much influential due to the long-time lag identified by Wasserman *et al.* as they take full advantage of this specific period in film history, in which quotation became an increasingly popular practice.

Figure 6. Outgoing connections/Decade

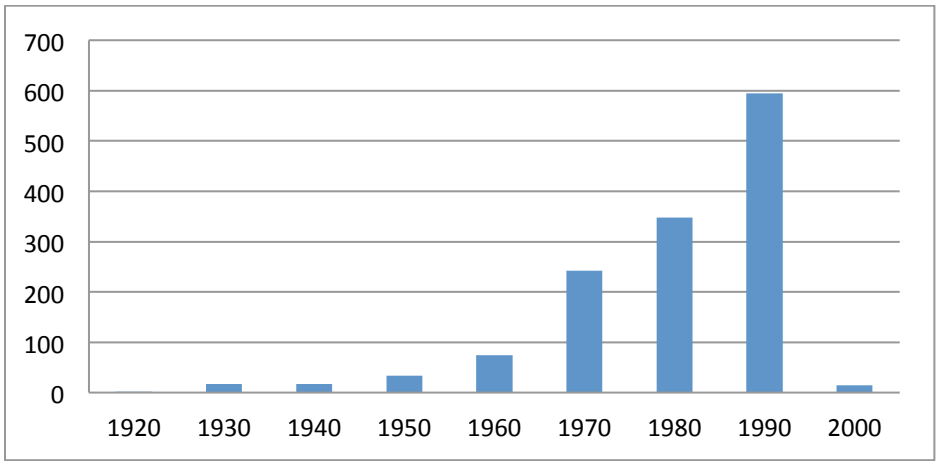
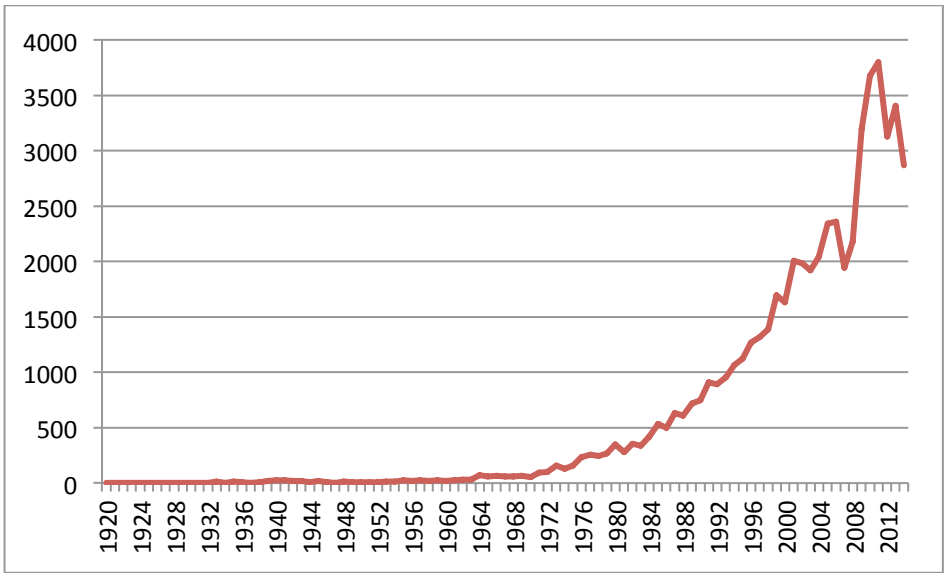


Figure 7. Incoming connections/year

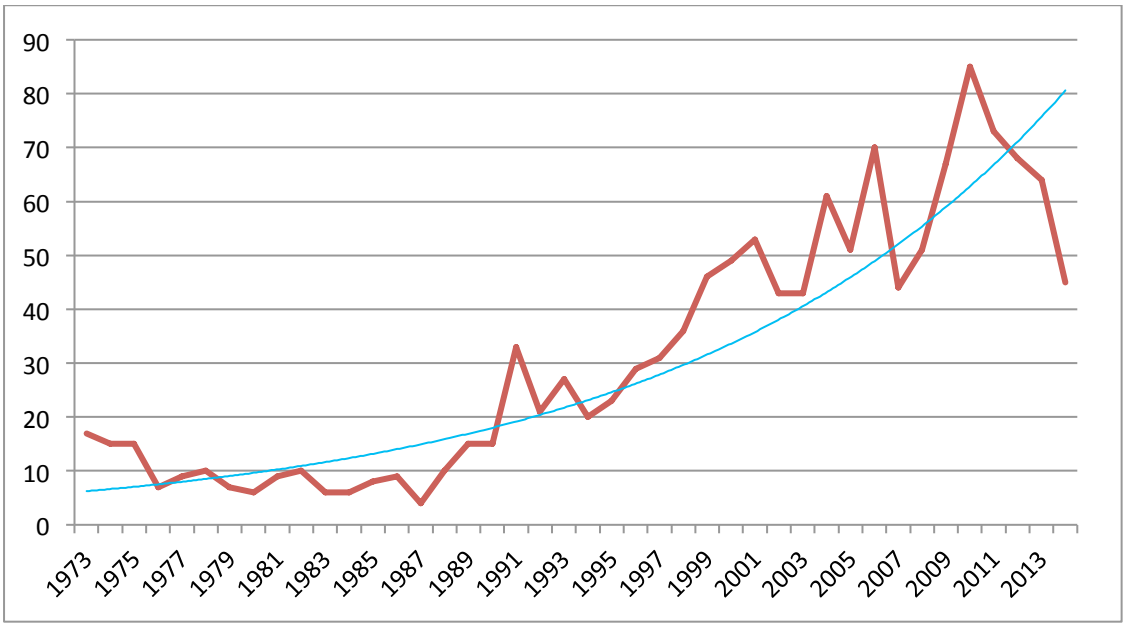


The movies of the 70s still benefit from this explosion, since their 25-year time lag starts at the end of the last century or even earlier. The distribution pattern of citations of these movies follows the exponential curve of growth more closely, as illustrated by the graph for *The Godfather* (see Figure 8). This movie is a perfect example of how influential movies evolve over time. Its release in 1972 and the release of Part Two in 1974 facilitated an initial period of popularity along the exponential curve. This early enthusiasm gives way to a period under the exponential curve until the release of Part Three in 1990, with the result

ISSN 2386-7876 – © 2016 Communication & Society 29(2), 151-172

that in the following year quotations significantly overtake the exponential curve, breaking the barrier of 30 connections. Later, the distribution follows more or less the course of the curve until coinciding with its 25 year time lag, 1997, and rebounds upwards to its maximum value in 2010. The last period is downward as is usual in the rest of the movies being studied. The explanation for this general descent, from 2011 to 2014 as we can see again in Figure 7, has less to do with filmmakers giving up citation practices than the fact that spectators need time to recognize the allusions made in the newest movies. This tendency confirms the data obtained by Wasserman *et al.*, who note "that there is a latency period for the reporting of film connections" (2014a: 7).

Figure 8. The Godfather (1972). Time lag 25y (1997)

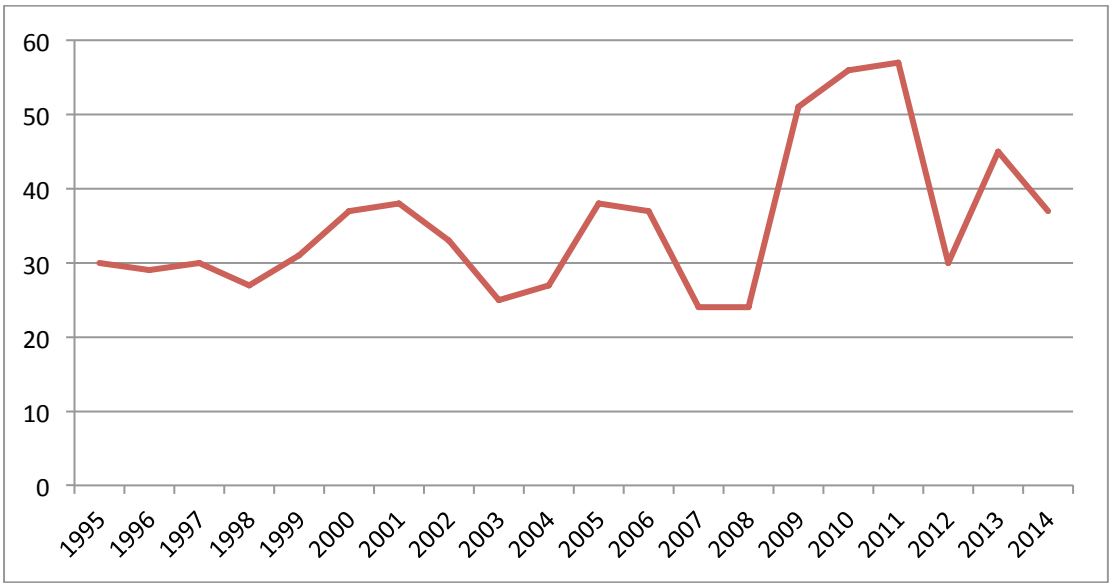


Obviously, the movies of the 80s and later cannot benefit from the quotation boom phenomenon because their 25-year time-lag is at the end of this period. *E.T.: The Extra-Terrestrial*, for example, follows a pattern similar to *The Godfather*'s; but unlike this movie, the rise in citations of *E.T.* is only partially identifiable, since this starts in 2000 and its 25-year time-lag begins in 2007. *Pulp Fiction* is another interesting example in this case from the 90s (see Figure 9). Its pattern perfectly matches the patterns of earlier films patterns considered influential. Wasserman *et al.* take into consideration the selections made by the National Film Registry (NFR) as films of significance for American culture[6]. This institution already lists *Pulp Fiction*, along with other films from the 1990s such as *The Matrix* and *The Silence of the Lambs*. The time it has taken for these movies to be selected or the NFR list has been 19, 13 and 20 years respectively; relatively short periods if we take into consideration that the average time it takes for a movie to be included in this list is more than 50 years. As Wasserman *et al.* demonstrate, their long-gap citation model is a strong predictor for inclusion in the NFR. The reason is that both indicators are very conservative, which favors older movies at the expense of newer ones. Examples of this bias are *Pulp Fiction*, *The Matrix* and *The Silence of the Lambs*, which for their own merit should be incorporated within the group of influential films even though they are relatively new.

[6] The NFR currently (25/03/2015) comprises 649 films that are culturally, historically and aesthetically significant for American culture.

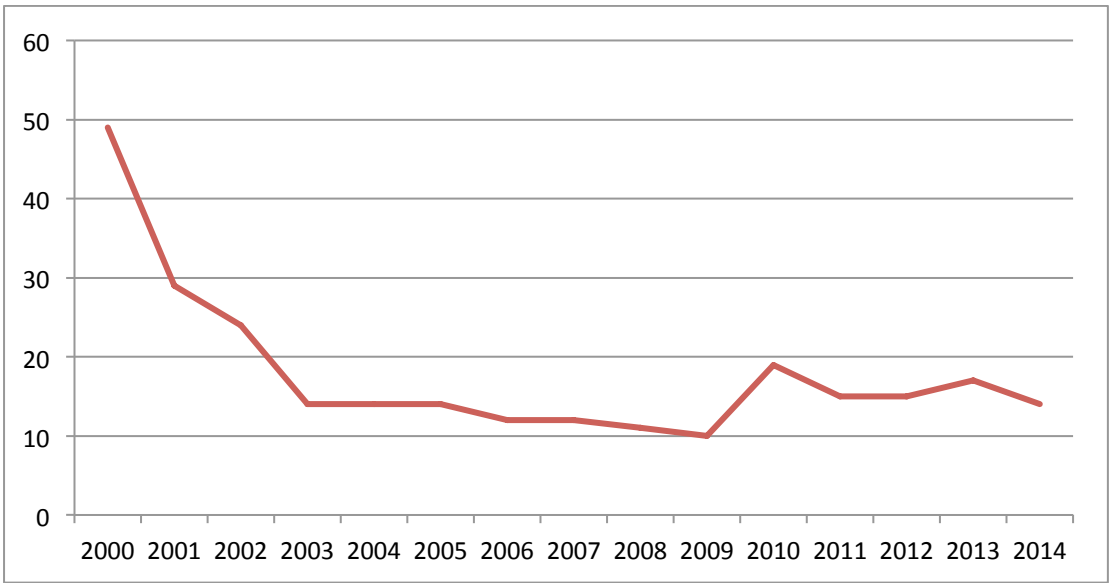
ISSN 2386-7876 – © 2016 Communication & Society 29(2), 151-172

Figure 9. Pulp Fiction (1994). Time lag 25y (2019)



Quite different is the case of *The Blair Witch Project*, another movie from the 90s (see Figure 10). This case reflects the tendency detected by Wasserman, as noted above, whereby films receive numerous shorter-gap citations and then the frequency decreases over time. The year after its release, in 2000, this movie received a very significant number of connections (almost 50), reflecting a strong initial impact, but a year later its influence had already decreased significantly, falling to 30 connections, and so on until it stabilized to just above 10 connections. If we compare the patterns of *Pulp Fiction* and *The Blair Witch Project*, we can predict, taking into account this one source, that the former has more probability of continuing to be influential than the latter, whose significance over time seems to be fading.

Figure 10. The Blair With Project (1999). Time lag 25y (2024)



6.2. *Proposal for classification of influential movies*

In this section, our goal is to propose a classification that would make it possible to define different types of influential movies. To do this, we consider the other two variables that

ISSN 2386-7876 – © 2016 Communication & Society 29(2), 151-172

define a film's significance according to Wasserman *et al.*: quality and impact. For the evaluation of quality, we propose to use the opinion both of experts (critics, academics) and of spectators (i.e. the general public). Both sources have advantages and disadvantages: the expert´s opinion may be more reasoned than that of the general audience, but, on the other hand, the latter can offer a less biased view than the experts, because they constitute a larger number of opinions.

According to different research, tapping into the "wisdom of crowds" can be a useful method for evaluating products and services (Surowiecki, 2004; Manski, 2006; and Krauss *et al.*, 2008). We propose to use IMDB´s user rating as a quantitative indicator of this wisdom. By using this score, IMDb elaborates its own 250 best movies (Top250). In relation to expert opinions we propose to take into consideration five indicators: Metascore (Metacritic´s ratings by critics); "The 50 Greatest Films of All Time" listed by the British Film Institute and Sight and Sound (Top50); the American Film Institute's Top 100[7]; inclusion in the National Film Registry (NFR); and, finally, awards and nominations received. In addition, in order to evaluate the economic impact we propose to use the box office gross indicator. For the purpose of avoiding bias we use the inflation-adjusted ticket price provided by The Numbers, whose web site presents a chart listing the "all time highest grossing movies in the domestic market since 1977"; 980 movies that made over $100 million in the US box office[8].

We use the term "Foundational Movies" to refer to our proposed first group of influential movies. This group covers movies from the 20s and early 30s, which, due to their particular qualities, and in spite of their age, have remained influential over time. As foundational movies in film history, they are referenced in subsequent movies and make very little reference to previous ones. In this group, we include *Nosferatu*, *Metropolis*, *Frankenstein*, *Dracula* and *King Kong*. The first two of these are German Expressionist works while the latter three are Hollywood productions (all three are listed in the NFR).

The construction of fantastic universes is common to all of these foundational movies. The depiction of monstrous creatures (with the exception of *Metropolis*) is another common feature. These monsters, who are taken out of their usual environment to threaten humankind, are not only the plot focus of their respective movies, but also the main reason for the films' continued popularity over time. They are the main source of citations and their images have been updated over the course of film history. For example, *King Kong* has been updated, among other occasions, in 1976 and more recently in 2005. These two moments are significant in the distribution of its connections, as at these times there is a notable increase in the number of the incoming connections received by the original *King Kong*[9].

All these movies are also seminal works of the horror genre, making them especially influential for the establishment of the codes and conventions thereof. *Nosferatu* and *Dracula* are unquestionably the foundational movies of the vampire genre. However, due to their completely different production styles (the former based on German Expressionist horror-fantasy, the latter on the Universal Studio style) each one has produced a lineage of its own. It is the latter that has proven more influential thanks to heirs like the Hammer Vampires of the 50s, 60s and 70s, the updated versions directed by John Badham in 1979 and by Coppola in 1992, and the recent prequel to the Dracula story, titled *Dracula Untold* (2014).

The second group we propose is "Classical Movies". In this group, we include eight films from 30s, 40, and 50s, all produced by Hollywood studios. In this group, the reasons for their

[7] http://www.afi.com/100Years/movies10.aspx
[8] 26/03/2015 (http://www.the-numbers.com/movie/records/All-Time-Inflation-Adjusted).
[9] It is common for milestones in the afterlife of a movie to see significant spikes in its connections; for example, 2001: A Space Odyssey has a major spike in the year 2001.

ISSN 2386-7876 – © 2016 Communication & Society 29(2), 151-172

popularity are not the characters but certain memorable scenes. These films are notable for the recognition of their quality by experts: all are listed in the NFR and in the AFI Top100, 37.5% of them are in the BFI Top50, and 62.5% received Academy Awards. Significant among these films are: *Citizen Kane*, listed in first place on the AFI list, second place on the BFI list and the winner of one Academy Award; *Vertigo*, which tops the BFI list and finishes ninth on the AFI chart; *Singin´ in the Rain*, fifth on the AFI list and twentieth on the BFI; *Casablanca*, third on the AFI chart and the winner of three Academy Awards including best picture; and *Gone with the Wind*, sixth on the AFI list and the winner eight Academy Awards including best picture. The spectators also endorse the quality of this group of movies, all of which are found in the Top250. Like the former group, these films are characterized mostly by incoming connections; however, there are two movies in this group that tentatively begin the practice of revisiting the cinematic past: *Sunset Blvd* and *Singin´ in the Rain*, both from the 50s. The latter, for example, draws on earlier films in the musical genre, such as *The Jazz Singer* (1927) or *The Broadway Melody* (1929).

The third group is the "Disney Factory": nine movies produced by Walt Disney that are found among the 100 most influential films. While five of them belong to Hollywood's classical period, four are from the early 80s or the 90s. It is interesting to note that the Academy Awards that these films have won or been nominated for are in almost all cases in the categories of Best Music, Best Original Song and Best Original Score. These films are not particularly visible on the lists of best films, with the exception of the oldest, *Snow White and the Seven Dwarfs* (1937) and the newest, *Toy Story* (1995). The latter in particular has better ratings in the quality indicators.

Toy Story was actually the product of a joint venture between Disney and Pixar Animation Studios, the first result of this productive alliance between these two animation studios. The success of this movie led to the production of two sequels, and the popular expression that "the sequels are never any good" is not applicable in this case. All three films grossed more than $100 million, and each sequel outdid the box office success of its predecessor. The last in the series not only fared better in terms of economic impact but also in quality indicators: *Toy Story 3* won two Academy Awards (Best Animated Feature Film and Best Original Song), giving it significantly more awards and nominations than the other two films in the series, and it holds a higher position in the user Top250.

Indeed, in all cases for which we have data, the Disney Factory group is very profitable because all of its films exceeded $100 million in gross revenues. Of special note in this respect is the economic success of *The Lion King*, which is the seventh highest grossing film on The Numbers list. This movie also presents above acceptable quality indicators. In short, it is fair to say that the Disney Factory has produced some significant movies, not only influential but also with good quality ratios and considerable box office success, becoming a historical benchmark in the world of commercial animation while maintaining its supremacy today thanks to its astute association with Pixar.

We call the fourth group "Successful Sagas", as the success of a representative number of influential movies originates from the respective series of films to which they belong. *Toy Story* is an example of this group; however, the pattern followed by this trilogy is not typical, as it is more common that the original movie is the one that not only remains the most influential but also the most significant over time. *Star Wars*, *Raiders of the Lost Ark*, *Alien*, *Back to the Future* and *Jurassic Park* are all examples of this tendency. Without a doubt, *Star Wars* is the outstanding case: the first film is not only the most influential, as noted above, but also the biggest box office success. Moreover, this movie launched the most successful film saga ever produced; all the movies of this franchise (six to date) are among the fifty films with the highest box office revenues and four are among the top ten. However, although its successors have also been successful, it is indisputably the original *Star Wars* film that remains the most influential.

ISSN 2386-7876 – © 2016 Communication & Society 29(2), 151-172

Not all sagas follow this tendency, as we can also find examples, albeit less frequent, where the successor surpasses the original. Christopher Nolan´s reinterpretation of *Batman* in 2008, for example, receives better indicators in all areas than Tim Burton´s 1989 version. Moreover, even though Nolan's film is relatively new, it has already received 278 incoming connections. Of course, there is a previous version of Batman, but it is Tim Burton's that was the most successful and therefore it is this and not previous versions that has become the point of reference for future generations. Another example is James Cameron's *The Terminator* series, as *Terminator 2: Judgment Day* achieves better indicators than its predecessor. With more spectacular special effects, Cameron improved on his previous work, thereby achieving much better results at the box office, among spectators and in terms of awards, winning four Oscars related mainly to visual and sound effects.

The fifth group is what we refer to as "New Hollywood Movies". Practically all the original movies mentioned in the last paragraph are examples of this group. As noted above, the decline of the old filmic models gave rise to a re-interpretation of those models, occurring mostly after the transition period of the 60s, during the decades of the 70s and 80s. Above all, this period includes the films that made the names of the new generation of Hollywood filmmakers, who, after their first experimental steps, updated old formulas to carve out their own place in film history, 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 the AFI Top100, the NFR, and the IMDb user rating Top250; all have good scores on Metacritic, have earned Academy Awards, have yielded very good box office returns, and have accumulated the most incoming connections. We therefore consider these movies to be the best examples of significant movies of all the films in the sample. *The Godfather* deserves special mention, since this movie is the best example of those cases in which the audience and critics are in perfect harmony; spectators rate it the second best film of all time, as do the experts at the AFI; moreover, it is included in the BFI and NFR lists and Metacritic gives it its highest score (100).

But we also propose the inclusion in this group of movies that are not as significant for film history as a whole as for particular genres. These movies, like the ones mentioned above, establish new models that serve as benchmarks for future generations. Most striking is the number of films of this type belonging to the horror genre. 1968 can be recognized as a watershed year for modern horror with the release of films like George A. Romero's *Night of the Living Dead* and Roman Polanski´s *Rosemary´s Baby*, followed later by *The Exorcist* (1973), *The Texas Chain Saw Massacre* (1974), *Halloween* (1978), *Friday the 13th* (1980) and *The Shining* (1980).

The sixth group we propose is "Link Movies", referring to films that form links in the chain of film history intertextuality. While in the 70s and 80s directors started looking back admiringly on their filmic heritage, by the 90s they were doing so obsessively. In this decade, the practice of recovering forgotten and outdated movies became common while at the same time these films became references for a new generation. The eight movies with the most outgoing connections are from this decade, and in Figure 6 we can see the upward trend, reaching its highest point in the 90s, as noted above. Thus, this decade, the hinge between the old century and the new one, offers the best examples of link movies, since they accumulate both incoming and outgoing connections.

Perhaps the best example of this practice is *Pulp Fiction*, the film with by far the most outgoing connections, directed by Quentin Tarantino, who is well-known both for his predilection for referencing movies and his ability to produce highly successful films in their own right. As M. Keith Booker, rightly notes, *Pulp Fiction* "was overtly (and extremely successfully) marketed as a film about film, and as a film made by a film buff for film buffs." This is why Amanda Lipman defines this film as a "ragbag of film references" while Booker

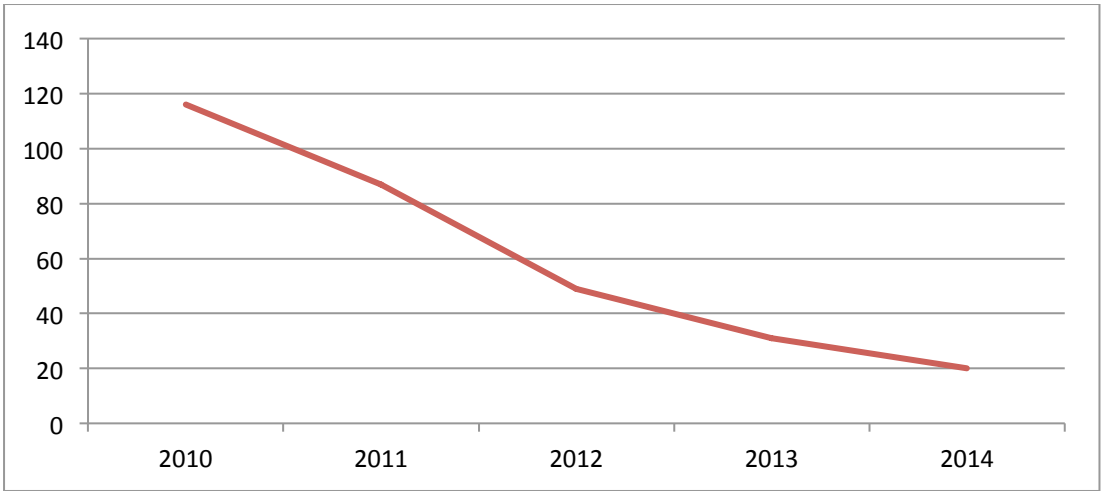
ISSN 2386-7876 – © 2016 Communication & Society 29(2), 151-172

describes it as "a collection of references to earlier movies, and as such [...] a perfect illustration of the phenomenon of pastiche in postmodern film" (2007: 89). However, 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" (1996: 17). It is precisely this element of updating rather than mere repetition of old formulas that allows film history to move forward.

In addition, this updated "present" subsequently becomes a point of reference for later proposals, such as the *Breaking Bad* TV series, which features numerous references to Tarantino´s filmography. For instance, the appearance of Jesse Pinkman´s girlfriend is reminiscent of that of Mia Wallace, which in turn cites that of Nana in Godard´s *Vivre sa Vie* (*My Life to Live*, 1962), which itself refers back to what could be the original, the Louise Brooks look. Although she may be the original source, Brooks is probably not the reference for the *Breaking Bad* character, which is more likely to be Tarantino´s Mia given that she has become more famous than her predecessors, being the one who has left an indelible mark on the collective imaginary at the expense of the original source.

Finally, we propose a seventh group, which we have labeled "Instantly Popular Movies". This group includes recent movies which, in very little time, have managed to achieve notable significance and popularity. For example, the year after its release in 2010 *Avatar* received no fewer than 116 incoming connections, although this avalanche immediately waned to 87 in the following year to just 20 connections in 2014 (see Figure 11). However, this significant descent may be partly explained by the general tendency that characterizes this period, from 2011 to 2014, as discussed above.

Figure 11. Avatar (2009). Time lag 25y (2034)



It is impossible to predict whether this downward trend will continue, or if it will recover in subsequent years as happened to *Titanic*, another James Cameron movie. As we can see in Figure 12, *Titanic*, like *Avatar*, sank to its lowest level of connections in the fifth year after its release and then began a rising trend, albeit without reaching its initial notoriety. In both cases, Cameron achieved a huge economic impact, as *Titanic* is the second biggest grossing movie at the box office while *Avatar* is the fourth. Moreover, both movies have received numerous awards and nominations; for instance, *Titanic* is the movie with the most Academy Awards (eleven in total) of all the films in Top100_I.

ISSN 2386-7876 – © 2016 Communication & Society 29(2), 151-172

Figure 12. Titanic (1997). Time lag 25y (2022)

On the other hand, *Twilight*, although it follows the same pattern as *Avatar*, has poorer scores than Cameron´s movie in all of the indicators. For example, *Twilight* has the lowest spectator rating of the Top100_I movies and a very low Metascore. With these indicators it is unlikely that this film will see the kind of turnaround enjoyed by *Titanic*. Although it is not possible to predict the future behavior of movies, as this depends on filmmakers and spectators, based on these criteria it is reasonable to posit that *Twilight* has less chance than *Avatar* of transcending a fleeting impact on the present to have a lasting influence on future generations.

7. Conclusion

As part of the equation, quality and impact are important parameters for determining a movie's potential influence on future generations. As we have seen, the most influential films have good scores in these two variables and therefore they can be considered good indicators of a film's future significance. Moreover, both variables have been useful for the purposes of classifying influential films. Seven categories have been proposed here: Foundational Movies, Classical Movies, the Disney Factory, Successful Sagas, New Hollywood Movies, Link Movies, and Instantly Popular Movies. The aim of this classification is to better understand the features that define the most influential movies.

However, before establishing this classification, it was necessary to identify the 100 most influential films. This was done by viewing film history as a network of citations, the filmmakers being the ones responsible for quoting and the spectators being the ones responsible for recognizing the quotes. Many of these citations are listed in the various categories of the IMDb section "connections". In recent years, the participation of spectators in this social network has become an object of analysis for the purpose of making predictions and drawing conclusions about cinematic behavior. Different attributes are used, such as user reviews, user ratings, box office gross, awards and nominations, movie genres, movie stars, co-actor networks, and so on. Different methodologies are applied, such as opinion mining, mapping techniques, network analysis, visualization methods and cross-evaluation metrics, and the different studies have diverse focuses, such as geopolitical interpretations, gender differences, forecasting movie success, sales prediction, improving the effectiveness of recommendation systems, predicting the popularity of movies, classifying movies, or determining the evolution of new patterns.

ISSN 2386-7876 – © 2016 Communication & Society 29(2), 151-172

Our focus in this case has been to identify, define and classify the 100 most influential movies in film history. Our research has found that although some movies from the early and classical periods of cinema still retain some capacity for influence, films from the post-classical era of the 70s and 80s clearly predominate, and over time this predominance will shift to the 90s. The references of the filmmakers of these three decades came mainly from classicism, while for new generations post-classical cinema is the main point of reference. At the same time, the filmmakers of the 70s and 80s, with an already rich film history behind them, began the practice of revisiting the cinematic past, which continued to grow in the 90s and reached its peak by the early part of the second decade of this century. In many cases, the new generations of filmmakers turn to the recent past to reference motifs that they believe were created by their sources, when in reality they themselves were updates of long forgotten motifs from the past that those more recent filmmakers re-popularized, turning them into the sources for the younger filmmakers, for whom, it would seem, film history began some time in the 1970s.

References

Ahmed, A. *et al.* (2007). Visualisation and Analysis of the Internet Movie Database. *Asia-Pacific Symposium on Visualisation* 2007, APVIS 2007, Proceedings, 17–24.

Allen, G. (2006). *Intertextuality*. London: Routledge.

Asad, K.I., Tanvir A. & Md Saiedur Rahman (2012). Movie Popularity Classification Based on Inherent Movie Attributes Using C4.5, PART and Correlation Coefficient. *2012 International Conference on Informatics, Electronics and Vision, ICIEV* 2012, 747–752.

Barthes, R. (1977). The Death of the Author. *Image—Music—Text*. Trans. Stephen Heath. London: Fontana.

Booker, M.K. (2007). *Postmodern Hollywood: What's New in Film and Why It Makes Us Feel so Strange*. Wesport: Praeger.

Canini, L., Benini, S., Migliorati, P. & Leonardi, R. (2009). Emotional Identity of Movies. *Proceedings - International Workshop on Content-Based Multimedia Indexing*, 1821–1824.

Carroll, N. (1982). The Future of Allusion: Hollywood in the Seventies (and Beyond). *October 20*, 51-81.

Chumo, 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).

Cook, P. (2005). *Screening the Past: Memory and Nostalgia in Cinema*. London and New York: Routledge.

De Valck, M. & Hagener, M. (2005). *Cinephilia: Movies, Love and Memory*. Amsterdam: Amsterdam University Press.

Dika, V. (2003). *Recycled Culture in Contemporary Art and Film The Uses of Nostalgia*. Cambridge: Cambridge University Press.

Dodds, K. (2006). Popular Geopolitics and Audience Dispositions: James Bond and the Internet Movie Database (IMDb). *Transactions of the Institute of British Geographers 31*, 116–130.

ISSN 2386-7876 – © 2016 Communication & Society 29(2), 151-172

Eco, U. (2005). Innovation & Repetition: Between Modern & Postmodern Aesthetics. *Daedalus 134*(4), 191–207.

Fraile, F. & Guerri, J.C. (2014). Simple Models of the Content Duration and the Popularity of Television Content. *Journal of Network and Computer Applications 40*,12–20.

Gallos, L.K., Potiguar, F.K., Andrade, J.S. Jr. & Makse, H.A. (2013). IMDB Network Revisited: Unveiling Fractal and Modular Properties from a Typical Small-World Network. *PLoS ONE 8*(6), 26–28.

Genette, G. (1997). *Palimpsests: Literature in the Second Degree*, trans. Channa Newman and Claude Doubinsky. University of Nebraska Press.

Herr, B.W., Ke, W., Hardy, E. & Börner, K. (2007). Movies and Actors: Mapping the Internet Movie Database. *Proceedings of the International Conference on Information Visualisation*, IEE Computer Society Conference Publishing Services, 465–469.

Iampolski, M. (1998). *The memory of Tiresias. Intertextuality and Film*, trans. Harsha Ram. Berkeley: University of California Press.

Jung, J.J. (2012). Attribute selection-based recommendation framework for short-head user group: An empirical study by MovieLens and IMDB. *Expert Systems with Applications 39* (4), 4049-4054.

Kabinsingha, S., Chindasorn, S. & Chantrapornchai, C. (2012). A Movie Rating Approach and Application Based on Data Mining. *International Journal of Engineering and Innovative Technology* (IJEIT) *2*(1), 77–83.

Krauss, J., Nann, S., Simon, D., Kai, F. & Gloor, P. (2008). Predicting Movie Success and Academy Awards through Sentiment and Social Network Analysis. *Proceedings of 16th European Conference on Information Systems (ECIS)*, Galway, Ireland.

Kristeva, J. (1986). Word, Dialogue and Novel. In T. Moi (Ed.). *The Kristeva Reader*. New York: Columbia University Press.

Liu, Y., Yu, X., An. A. & Huang. X. (2013). Riding the tide of sentiment change: sentiment analysis with evolving online reviews. *World Wide Web 16*(4), 477-496.

Manski, C.F. (2006). Interpreting the Predictions of Prediction Markets. *Economic Letters 91*, 425-429.

Moon Bae, S., Chun Lee, S. & Park, J.H. (2014). Utilization of Demographic Analysis with IMDB User Ratings on the Recommendation of Movies. *Journal of Society for e-Business Studies 19*(3), 125-141

Ott, B., & Walter, C. (2000). Intertextuality: Interpretive Practice and Textual Strategy. *Critical Studies in Media Communication 17*(4), 429-446.

Otterbacher, J. (2013). Gender, writing and ranking in review forums: a case study of the IMDb. *Knowledge and Information Systems 35*(3), 645-664.

Park, J., Kim, M. & Jun, Y. (2014). Interdisciplinary Research on Hollywood Action Movies from 1930 to 2009. *English 21*, 27, 2, 369-386.

Hsu, P., Shen, Y. & Xie, X. (2014). Predicting Movies User Ratings with Imdb Attributes. *Rough Sets and Knowledge Technology 8818*, 444-453

Nelson, R. & Glotfelty, R. (2012). Movie stars and box office revenues: an empirical analysis. *Journal of Cultural Impact 36*(2), 141-166.

Ridanpää, J. (2014). 'Humour is Serious' as a Geopolitical Speech Act: IMDb Film Reviews of Sacha Baron Cohen's The Dictator. *Geopolitics 19*(1), 140-160.

Saraee, M., White, S. & Eccleston, J. (2004). A Data Mining Approach to Analysis and Prediction of Movie Ratings. *Data Mining V*, 343-352.

Sobchack, V. (2013). WHY I (LOVE) IMDb. *Film Comment 38*.

Sreenivasan, S. (2013). Quantitative analysis of the evolution of novelty in cinema through crowdsourced keywords. *Scientific Reports* 3, 1-11.

Surowiecki, J. (2004). *The Wisdom of Crowds*. New York: Doubleday.

Wasserman, M., Mukkerjee, S., Scott, K., Zeng, X.H.T., Radicchi, F & Amaral. La.A.N. (2014a). Correlations between User Voting Data, Budget, and Box Office for Films in the Internet Movie Database. *Journal of the Association for Information Science and Technology*, 1–14.

Wasserman, 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.

ISSN 2386-7876 – © 2016 Communication & Society 29(2), 151-172