Expert Systems With Applications 250 (2024) 123750

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Modeling the opinion dynamics of superstars in the film industry

Jesús Giráldez-Cru [a,c,*], Ana Suárez-Vázquez [d], Carmen Zarco [b,c], Oscar Cordón [a,c]

[a] Department of Computer Science and Artificial Intelligence (DECSAI), University of Granada (UGR), Spain
[b] Department of Marketing and Market Research, University of Granada (UGR), Spain
[c] Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), University of Granada (UGR), Spain
[d] Department of Business Management, University of Oviedo (UNIOVI), Spain

ARTICLE INFO

Keywords:
Film industry
Movie superstars ranking
Opinion dynamics
Bounded confidence
Repulsion mechanisms

ABSTRACT

One of the most challenging questions in the film industry is to rank superstars, which ultimately affects some performance indicators like movie success. In this work, we address this question by means of opinion dynamics models, where the evolution of opinions in a population is analyzed. We apply a model of this kind to study the evolution of opinions about a set of well-known movie superstars in a real-world population. Also, we use real-world data from a specialized cinema website to model mass communication processes (representing film releases and their related news and marketing campaigns), and to measure the performance of our model. Our results show that the proposed model is able to accurately represent this complex system, where the opinion dynamics of superstars are mostly driven by emotional mechanisms, and reveal that film releases and their corresponding marketing campaigns only have a short term effect on those opinions. To the best of our knowledge, this is the first work that applies opinion dynamics models to the study of opinions about superstars in the film industry.

1. Introduction

Being able to predict a product's success prior to launch has always been a key concern for marketers (Hauser et al., 2006). Estimating the performance of a new product is especially challenging in creative industries, such as the motion picture industry (Karniouchina, 2011), as traditional techniques for assessing audience response are unable to capture the full experiential consumption aspects of movies (Eliashberg & Sawhney, 1994).

Previous research suggests that movie success is a complex phenomenon resulting from the interplay of movie characteristics (e.g., star cast), post filming studio actions in terms of communication and distribution activities, and external factors such as critics' reviews or word of mouth recommendations (Hennig-Thurau et al., 2006). Digitalization has increased competition in the audiovisual content market and determining the value of each new film has become even more essential now (Kübler et al., 2021).

Among the myriad factors that influence the success of a film, one that has attracted particular attention in both the industry and academic literature is the value of superstars (De Vany & Walls, 1999). To this end, and in contrast to dichotomous approaches that attempt to determine the value of actor and actresses using dummy variables (star/non-star), superstar rankings have proven to be more effective (Nelson & Glotfelty, 2012). Thus, the motivation of this work is to address one of the most challenging questions in the film industry: how to establish a ranking of superstars. Previous methodologies employed to address this question have led to mixed conclusions about the effect of superstars on films' success and the results of different rankings are difficult to compare (Ghiassi et al., 2015). Moreover, these rankings are based primarily on the box office results of stars, ignoring the fact that movies are experiential products whose consumption is driven by emotions (Hennig-Thurau et al., 2007). In the current contribution, we analyze the role of emotions about movie superstars from the lens of opinion dynamics (OD). OD models use agent-based modeling (Epstein, 2006; Farmer & Foley, 2009) to analyze the evolution of opinions in a population (Dong et al., 2018; Wang et al., 2020; Xia et al., 2011). They mostly rely on an opinion fusion rule which defines how each agent's opinion is updated after an interaction (representing, for instance, a word-of-mouth process between agents, or a mass communication broadcast) (Noorazar et al., 2018). OD models have been successfully used to analyze many problems, including the study of biased opinions (Anagnostopoulos et al., 2022), multi-attribute group decision-making (Li et al., 2021), the identification of opinion leaders in social networks (Chen et al., 2021), the gap between people's voting result and their collective opinion (Jiao & Li, 2021), and

* Corresponding author at: Department of Computer Science and Artificial Intelligence (DECSAI), University of Granada (UGR), Spain.
E-mail addresses: jgiraldez@ugr.es (J. Giráldez-Cru), anasv@uniovi.es (A. Suárez-Vázquez), carmen.zarco@ugr.es (C. Zarco), ocordon@decsai.ugr.es (O. Cordón).

https://doi.org/10.1016/j.eswa.2024.123750
Received 21 April 2023; Received in revised form 12 February 2024; Accepted 18 March 2024
Available online 28 March 2024

the consensus reaching with trust evolution in social network group decision-making (Zhang et al., 2022), among others.

One of the most studied OD models is bounded confidence (BC) (Deffuant et al., 2000; Hegselmann & Krause, 2002), where agents' opinions evolve as a consequence of interactions between agents with similar opinions (i.e., agents in their confidence area). BC has been used to explain the OD reaching both consensus and fragmentation of opinions (Castellano et al., 2009). Nevertheless, this model is unable to capture some kinds of emotions that drive opinion evolution in certain systems, e.g., extremization (Isenberg, 1986). In contrast, the Agent-independent Time-based Bounded Confidence and Repulsion (ATBCR) model (Giráldez-Cru et al., 2022) has been recently proposed to overcome this inconvenience by means of an extension in the classical BC model based on a repulsion mechanism. In fact, this model has been already applied to model opinions in marketing campaigns (Giráldez-Cru et al., 2022).

Therefore, OD models arise as an alternative modeling strategy to represent the ranking of superstars. However, and due to the experiential nature of movies, we conjecture that emotional mechanisms (as the repulsion rule of the ATBCR model) are required to capture the underlying OD about superstars. Emotional mechanisms are the cultural and adaptive process by which individuals react to environmental contingencies in a flexible and dynamic way (Scherer, 2009). For example, the process by which an individual change his/her opinion of a superstar due to the divergence of this opinion from that of others is not rational and emotions play a role. Consequently, we adapt the ATBCR model of OD to study the evolution of opinions about a set of well-known superstars in the real-world population from Suárez-Vázquez (2015) and benchmark the obtained model with others resulting from the use of alternative OD approaches. To better capture the real dynamics of the phenomenon, we use real-world data both from a cross-sectional survey, to properly model the spectators' opinions, and from the specialized cinema website IMDb,[1] to enrich the model incorporating mass communication processes. These processes represent an information exchange to a large portion of the population, including news and marketing campaigns, for instance. In our case study, these mass communication processes represent film releases and all the related communication generated for these events. In addition, we also use real-world data from this website as a ground-truth to measure the accuracy of our model, which allows us to validate its outputs with an external, independent source of information. To the best of our knowledge, this is the first work that applies an OD model using real-world data to the study of opinions about superstars in the film industry.

In summary, these are the main contributions of our work:

- We present an innovative study of film performance indicators using opinions about superstars.
- Our analysis is based on an interaction-based OD model to rank a set of well-known superstars in the film industry, fed with data from a survey of spectators.
- We introduce a mechanism of mass communication, which is used to represent film releases and related news. These mass communication processes affect opinions about those superstars considering real-world data from a specialized cinema website.
- We present an extensive experimental analysis comparing the accuracy of different OD models using the said real-world data.
- Our experimental results show that the OD of this system are mostly driven by emotional mechanisms.

The rest of this manuscript is organized as follows. Section 2 motivates the present work and describes other related references. In Section 3 some preliminaries on the ATBCR model are defined, along with a brief description of some classical OD models which will be considered alternative approaches to modeling the spectators' opinions on the considered problem. Section 4 is devoted to the description of the adaptation of our model to cope with the real-world phenomenon tackled, including how real-world data is extracted and used. The experimental analysis is developed in Section 5. Finally, we conclude in Section 6.

2. Related works

The celebrate quote of Rita Hayworth "*men go to bed with Gilda but wake up with me*" represents the strength of the movie stars-spectators emotional relationship. Indeed, previous research has described stars as emotional competent objects (Luo et al., 2010) and identified the key emotional drivers of their influence (Moraes et al., 2019). This emotional dimension explains the difficulties associated with determining the origins of stardom (Adler, 1985; Rosen, 1981), the so-called power of stars (Elberse, 2007). It is related to the experience good property of movies (Chang & Ki, 2005) which underlines the importance of the psychological approach (Eliashberg et al., 2006), that is, focusing on the individual moviegoer's decisions, when analyzing the success factors of movies (Hadida, 2009).

2.1. Emotional dimension of superstars: secondary data

Film is the industry of making-believe by interacting with the human emotional system (Tan, 2013). It provides stimuli that influence spectators' emotions (Zillmann, 2015). One of these forces are the superstars that have been positioned as personalities with box-office power (King, 1987). The source of such power has raised a great interest in previous literature from the original explanations of Rosen (Rosen, 1981), Adler (Adler, 1985), Frank and Cook (Frank & Cook, 1995), and Borghans and Groot (Borghans & Groot, 1998) to more recent proposals of applied nature (Harashima, 2016).

Superstars are one of the few tangible features of film quality (Eliashberg et al., 2000; Ravid, 1999) and their mere presence benefits the promotional opportunities of the film (Suárez-Vázquez, 2011). Thus, the actress Sarah Bernhardt is frequently entitled as the world's first superstar precisely for her ability to anticipate the importance of image and buzz (Isaac-Goizé, 2023). Stardom is not only a cultural reality but a commercial one founded on the marketability of human identities (McDonald, 2012). Behind this phenomenon is the distinction between the artistic and commercial value of the star, that is, stars' popular appeal vs. expert judgments about the stars (Holbrook, 1999). Capturing popular appeal is not an easy endeavor. At the film level, online user-generated information is a valuable source of spectators' awareness and feelings. During the film pre-release period, star power affects the volume and valence of the online conversations about the film (Liu, 2006).

In fact, the cast of the film is one of the marketing variables with potential impact on spectators' decisions (Eliashberg et al., 2000). To approximate the value of the cast, academic researchers have relied on secondary sources based on (Elberse, 2007): industry magazines (Sawhney & Eliashberg, 1996); ratings from members of the industry (Ainslie et al., 2005), and previous awards and box-office success (Ravid, 1999). Actually, the accessibility of data is behind the huge increase in published research related to the film industry in the last decade (Behrens et al., 2021; McKenzie, 2023). Although the secondary nature of this data makes easier to conduct large-scale studies, it does not provide information about moviegoers' emotions towards the stars. Other kind of data, such the one that result from primary cross-sectional methods of research, can measure the emotional dimension of superstars from an individual perspective.

[1] https://www.imdb.com/

2.2. Emotional dimension of superstars: survey data

Star power is one of the key film-specific attributes examined in previous literature on the demand side of the movie market (McKenzie, 2023). Existing research has pointed out that focusing on audience information processing is crucial to fully understand superstars power (Hofmann et al., 2017). In this sense, Suárez (Suárez-Vázquez, 2015) conducted an empirical study collecting data from a cross-sectional survey for a set of 17 superstars. Due to the relevance of the young and highly educated segment in the composition of film audience (Terry et al., 2010), the population under study in that survey were people between 24 and 34 years old, with university studies in an European country. The sample was gender balance (51.2% female), mean age was 22.5 (s.d. 2.6) and 63.7% of the respondents went to a movie at the cinema occasionally (less than one a month). This dataset provides a homogeneous population in terms of age and educational level, which minimizes noise, i.e., the influence of extraneous variables (Peterson, 2001). It also allows reducing possible response bias in terms of, for example, acquiescence (Rammstedt et al., 2013). Besides, this data was coherent with the theatrical attendance demographics provided by the Motion Picture Association (MPA) (2021). Thus, from a typological point of view, the considered dataset represents the most significant segment of the cinema market (Broekhuizen et al., 2011; Cuadrado & Frasquet, 1999; Díaz et al., 2018).

The final sample included 5,440 responses from 320 spectators. This signifies, for a level of confidence —data accuracy— of 95%, a margin of error —data precision— of 5%, which is considered acceptable in social research (Taherdoost, 2017). Furthermore, this margin of error decreases to 1% if the full survey dataset is considered. The following information was collected for each spectator and each superstar in the data sample: (1) emotions elicited by each of the 17 superstars; (2) intention of watching a film starring each of the 17 superstars measured on a scale from 1 to 5 with anchors 1 "*The presence of this star in the cast of the film will not encourage me to watch the film at all*" and 5 "*The presence of this star in the cast of the film will be a very important stimulus for me to watch the film*". Following (Laros & Steenkamp, 2005), the emotions experienced for each of the 17 superstars were measured across eight basic emotions: anger, fear, sadness, shame, contentment, happiness, love, and pride. Spectators were asked the degree to which they felt each of these emotions for each of the superstars considered on a scale from 1 to 5 where 1 meant "*I do not feel this emotion at all*" and 5 stood for "*I feel this emotion very strongly*". Among the insights drawn from this study, the following seem particularly relevant in the context of the current paper: (1) when discussing the success of movie stars, the impact of emotions of different valence does not necessarily differ (i.e. the influence on the intention to watch a movie of both happiness and sadness evoked by a star are positive, despite their different valence); (2) changes in positive valence emotions have a greater impact on moviegoers intentions than changes in negative valence emotions; and (3) the degree of substitutability among superstars is affected by their emotional profile, superstars with a similar emotional profile are more substitutable in terms of predicted effect on spectators' intentions.

3. Preliminaries on opinion dynamics models

In this section, we provide a general summary of OD models, followed by a brief overview of both the ATBCR model (Giráldez-Cru et al., 2022) and of some classical OD models, which are used along the experimental analysis. In every case, we use standard notation in OD models (Dong et al., 2018).

3.1. Models of opinion dynamics

The goal of OD models is to study the dynamics from a set of initial opinions to a set of final opinions during a lapse of time, and to understand how this set of final opinions is achieved.

In OD models, a population of N agents interact in a social network.[2] This social network is represented by a graph $G(V, A)$, where V is the set of nodes (with $|V| = N$), and A is the $N \times N$ adjacency matrix (i.e., $A_{ij} = 1$ iff. there is an edge between nodes i and j in G, $A_{ij} = 0$ otherwise). Each agent is represented by a distinct node i in the graph, and interacts with other agents within its neighborhood, i.e., with some agent j in the set $\{j \in V \mid A_{ij} = 1\}$.

Every agent has an opinion on a certain subject, which evolves as a consequence of these interactions. The model is executed during a finite number of time steps T. Let $x_i(t) \in [0, 1]$ be the opinion of agent i at time step t, with $1 \leq i \leq N$ and $0 \leq t \leq T$. Moreover, let $X(t)$ be the opinion profile at time step t, i.e., $X(t) = \{x_i(t)\}_{1 \leq i \leq N}$. Therefore, OD models are proposed to study the underlying dynamics to achieve $X(T)$ from $X(0)$.

3.2. The ATBCR opinion dynamics model

The ATBCR model is an extension of the classical Deffuant–Weisbuch (DW) model of BC (Deffuant et al., 2000). The DW model is able to explain both consensus and fragmentation of opinions within a population. ATBCR extends it with a repulsion mechanism in order to also explain the extremization of opinions.

In the ATBCR model, each time step simulates an interaction between a randomly selected pair of agents i and j, connected in the social network (i.e., $A_{ij} = 1$). An agent i participating in an interaction with agent j in time step $t+1$ updates its opinion according to the following opinion fusion rule:

$$x_i(t+1) = \begin{cases} x_i(t) + \mu(x_j(t) - x_i(t)) \text{ iff } |x_i(t) - x_j(t)| < \varepsilon_i(t) \\ x_i(t) \text{ iff } \varepsilon_i(t) \leq |x_i(t) - x_j(t)| \leq \vartheta_i(t) \\ min(1, max(0, rep(i, j))) \text{ iff } |x_i(t) - x_j(t)| > \vartheta_i(t) \end{cases} \tag{1}$$

where the repulsion rule defined in the third case of Eq. (1) is:

$$rep(i, j) = x_i(t) - \mu(x_j(t) - x_i(t)) \tag{2}$$

with $\mu \in [0, 0.5]$ being the convergence speed of the model, and $\varepsilon_i(t)$ and $\vartheta_i(t)$ being the confidence and the repulsion thresholds of agent i at time step t, respectively. Notice that Eq. (1) always return a scalar value in the interval $[0, 1]$, i.e., the interval where agents' opinions take values. Notice also that both thresholds are agent-dependent and time-based, i.e., they both depend on the agent i and the time step t. For more details, we address the reader to the original definition of the ATBCR model (Giráldez-Cru et al., 2022).

The ATBCR model has been shown to preserve the behavior of the classical BC models (consensus and fragmentation of opinions) (Deffuant et al., 2000) while it also introduces a new mechanism to explain the extremization of opinions in a population with heterogeneous agents' behaviors (Giráldez-Cru et al., 2022). Nevertheless, and without loss of generality, in this work we only consider homogeneous and time-invariant thresholds of confidence and repulsion $\varepsilon = \varepsilon_i(t)$ and $\vartheta = \vartheta_i(t)$, $\forall i \in [1, N], t \in [0, T]$. This is due to the complexity to obtain fine-grained data to model these thresholds in our case study. We emphasize that we only use real-world data in our experimental analysis.

The rationality of the system can be derived from the values of ε and ϑ. In particular, systems with high confidence thresholds ε are *rational* (i.e., only rational interactions affect the OD of the system), whereas systems with high repulsion thresholds ϑ are *emotional* (i.e., opinions evolve as a consequence of emotional decisions).

[2] Notice that fully-mixed OD models can be simulated using a fully connected social network.

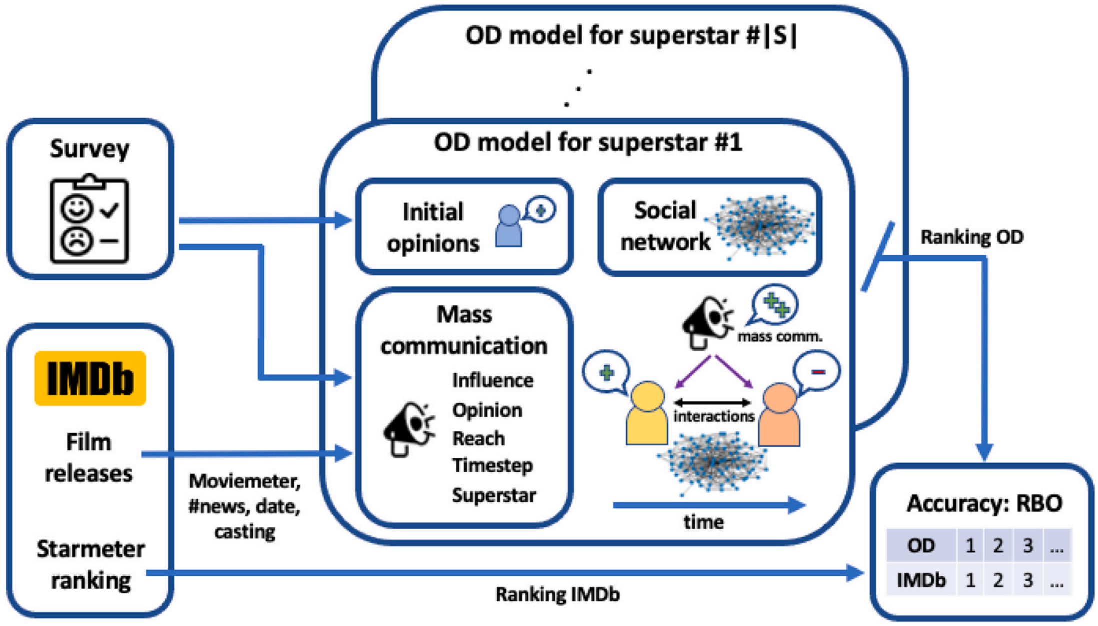

Fig. 1. Overview of the proposed model. The real-world data from the survey is used to initialize spectator agents' opinions and mass communication influences. The real-world data about film releases available at IMDb is used to model the remaining parameters of mass communication processes. A realistic social network is used to model agents' interactions. The OD model is executed, considering both mass communication processes and agents' interactions, which alter agents' opinions, and a ranking of superstars is computed based on those opinions. Finally, such a ranking is compared to the Starmeter ranking from IMDb.

3.3. Classical opinion dynamics models

The ATBCR model presented in the previous subsection was proposed as an extension of the DW model. Hence, the DW model (Deffuant et al., 2000) becomes a particular case of the former when $\vartheta = 1$. This model thus represents a purely rational system, where OD are only driven by a BC mechanism. Mathematically, it can be expressed just using the first case of Eq. (1).[3]

The DeGroot model (Degroot, 1974) is another highly rational system. But, in contrast to the DW model, agents are fully susceptible to any other opinion (i.e., the confidence of any other opinion is complete). Mathematically, the opinion $x_i(t+1)$ is updated as

$$x_i(t+1) = \sum_{j=1}^{N} w_{ij} x_j(t) \tag{3}$$

where w_{ij} is the weight agent i gives to agent j. This model has been extensively used to study the sufficient and necessary conditions to reach a consensus in a population (Berger, 1981).

The Friedkin–Johnsen (FJ) model (Friedkin & Johnsen, 1990) is a generalization of the DeGroot model in the sense it also considers that agents may be attached to a certain extent to their original opinion. The opinion fusion rule in the FJ model is:

$$x_i(t+1) = \xi_i \sum_{j=1}^{N} \left(w_{ij} x_j(t) \right) + (1-\xi_i) x_i(0) \tag{4}$$

where ξ_i is the susceptibility of agent i (i.e., $1-\xi_i$ is the stubbornness of agent i), and the remaining parameters are the same than in the DeGroot model.

4. Model description

This section describes all the decisions taken to design our model. First, it describes a light extension of the ATBCR model to handle multidimensional opinions in Section 4.1. Next, we describe how the real-world data from the survey by Suárez-Vázquez (2015) about superstars is used to initialize the opinion profile of the population in Section 4.2. The social network used to model spectator agents' interaction is described in Section 4.3. Section 4.4 describes how mass communication processes are modeled in our system. Finally, we use an external source of information from the specialized cinema website IMDb[4] to measure the accuracy of our model, as described in Section 4.5. An overview of the model is depicted in Fig. 1. A summary of the notation used in our model is reported in Table 1.

4.1. A multidimensional extension of the ATBCR model

In our analysis, we consider a multidimensional extension of the ATBCR model. In particular, let $S = \{s_1, \dots, s_{|S|}\}$ be a set of $|S|$ subjects, and $x_i^{s_k}(t)$ be the opinion of spectator agent i at time step t on subject $s_k \in S$. For the sake of simplicity, we consider that an interaction between spectator agents i and j only affects a particular subject (i.e., s_k). Therefore, we model the complex dynamics of this multidimensional system by executing in parallel $|S|$ independent instances of the original ATBCR model, one for each subject s_k, during $T/|S|$ time steps each. This way, the same interaction between two agents can happen in different instances of the model (each affecting a particular subject), and hence, this multidimensional extension is able to model interactions affecting multiple subjects. Moreover, notice that, besides the global performance of this multidimensional system, this extension also allows us to study the OD on each subject independently.

[3] Since the DW model is a particular case of the ATBCR model, it is guaranteed that the second and the third cases of Eq. (1) are never used in the DW model (i.e., when $\vartheta = 1$ in the ATBCR model).

[4] Star power is very difficult to measure objectively. In a particular film, star status is usually shown in the way the name of the star is deployed on the screen credits and on the promotional material of the film (McDonald, 2012). On an aggregate level, different measures have been proposed along the time, such as the Ulmer Scale or the Top Money-Making starts by Quigley Publishing Company. However, with the advent of accessible data from public websites, online industry resources, in particular, IMDb is behind most of the recent published research.

Table 1
Summary of the notation used in our model.

ATBCR model (Section 3.2)	
N	Number of agents
T	Number of time steps
$x_i(t)$	Opinion of agent i at time step t (with $1 \leq i \leq N$ and $0 \leq t \leq T$)
μ	Convergence speed
$\varepsilon_i(t)$	Confidence threshold of agent i at time step t
$\vartheta_i(t)$	Repulsion threshold of agent i at time step t
$X(t)$	Opinion profile at time step t
V	Set of nodes (with $\|V\| = N$)
A	Adjacency matrix of size $N \times N$
$G(V, A)$	Graph representing the social network
Multidimensional extension of the ATBCR model (Section 4.1)	
$S = \{s_1, \dots, s_{\|S\|}\}$	Set of subjects on which agents have opinions
$x_i^{s_k}(t)$	Opinion of spectator agent i at time step t on subject $s_k \in S$
Initial opinions from a real-world population (Section 4.2)	
$u_r^{s_k}$	Utility of survey respondent r about superstar s_k
α^{s_k}	Coefficient of superstar s_k
β_e	Coefficient of emotion e
$a_{r,e}^{s_k}$	Answer of respondent r about emotion e and superstar s_k
Mass communication (Section 4.4)	
Ψ	Opinion communicated in the mass communication process
υ	(Effective) Reach of a mass communication process
$\kappa(i)$	(Effective) Influence that a mass communication process has on spectator agent i
$m = \langle \Psi, \upsilon, \kappa(i), s_k, t \rangle$	Mass communication process with Ψ, υ, and $\kappa(i)$, on subject s_k occurring at time step t
μ_c	Convergence speed of mass communication processes

4.2. Modeling opinions from a real-world population

As seen in Section 2.2, the work by Suárez-Vázquez (2015) surveys the emotions of a population about a set of superstars. The sixteen superstars analyzed (i.e., S) are: Christian Bale, Gerard Butler, Nicholas Cage, George Clooney, Russell Crowe, Johnny Depp, Leonardo DiCaprio, Robert Downey Jr., Will Ferrell, Megan Fox, Tom Hanks, Robert Pattinson, Brad Pitt, Zoe Saldana, Will Smith, Kristen Stewart, and Reese Witherspoon; all of them are well-known superstars in the film industry.[5] The eight considered emotions are: anger, fear, sadness, shame, contentment, happiness, love, and pride; and were proposed as the main emotions to be analyzed in this context (Laros & Steenkamp, 2005). Based on them, the logit model is used to compute the *utility* u of the survey respondent r about superstar $s_k \in S$ as:

$$u_r^{s_k} = \alpha^{s_k} + \sum_{e \in Emotions} \beta_e \cdot a_{r,e}^{s_k} \tag{5}$$

where α^{s_k} (with $k \in \{1, \dots, 16\}$) is the coefficient of each superstar, β_e is the coefficient of the emotions $e \in \{1, \dots, 8\}$, and $a_{r,e}^{s_k}$ is the answer of respondent r about emotion e and superstar s_k in the survey (Suárez-Vázquez, 2015).[6] Notice that these utilities can be seen as the respondents' opinions about the superstars.

Since the survey by Suárez-Vázquez (2015) was conducted in a sample of 320 respondents, we amplify this population 10 times, i.e., $N = 3,200$ spectator agents in our model. This amplification process is a common procedure extensively used in agent-based modeling, where the number of agents is usually in the order of thousands (Chica & Rand, 2017). We emphasize that this process increases the heterogeneity of the population without altering its original average values, since the resulting population is composed of many agents *similar* to the original ones, but slightly different from each other. In particular, for each respondent r and superstar s_k in the survey, we generate a Gaussian distribution $\mathcal{N}_r(\bar{X}, \sigma)$ of initial opinions with size $|\mathcal{N}| = 10$, $\bar{X} = u_r^{s_k}$, and a small variability with $\sigma = 0.1$. This way, the initial opinion profile $X(0)$ of the population is the union of these distributions, i.e., $X(0) = \cup_r \mathcal{N}_r$. Notice that this process produces 10 spectator agents *similar* to each survey respondent (which is an actual person), allowing us to use a real-world distribution of initial opinions, amplified to simulate a larger, heterogeneous population.

Moreover, we use this survey (Suárez-Vázquez, 2015) to model the influence of each superstar on each respondent (we recall that the survey contains an specific question on the intention of watching a film starred by each superstar), following the same amplification process. This influence is used in the mass communications processes described in Section 4.4.

In Fig. 2, we represent the distribution of initial opinions, in both the survey and our model, as well as the influence of each superstar in the population. As it can be seen, both distributions are very similar, being the one used in our model slightly smoother.

4.3. Modeling spectator agents' interactions

In our model, we represent spectator agents' interactions in a social network using a synthetic graph following the Preferential Attachment (PA) model (Barabási & Albert, 1999). This model has been proposed to explain the growth of complex networks, and produces graphs with scale-free structure, i.e., graphs where node degree follows a power-law distribution $P(i) \sim i^{-\gamma}$, characterized by the exponent γ. As a consequence, most of the nodes have a low degree (i.e., they are only connected to a small subset of nodes), whereas a few nodes are connected to a big majority of the graph (i.e., they are *hubs*). In practice, many real-world networks exhibit this scale-free structure, with γ usually ranging in the interval $(2, 3]$. Hence, this social network topology is appropriate to represent the real fan interactions among spectators in our case study. The resulting PA graphs used in our analysis has 3,200 nodes (i.e., N), 31,900 edges, an average node degree $\langle k \rangle = 19.94$, and an average clustering coefficient $\langle C \rangle = 0.03$. We recall that interactions between spectator agents i and j only occur when these agents are connected in the social network, i.e., $A_{ij} = 1$ (there is an edge between nodes i and j in G).

[5] The survey by Suárez-Vázquez (2015) also considered Brad Pitt, who was used to calibrate the logit model. For this reason, we do not consider him in our analysis.

[6] The values of these coefficients are directly extracted from (Suárez-Vázquez, 2015).

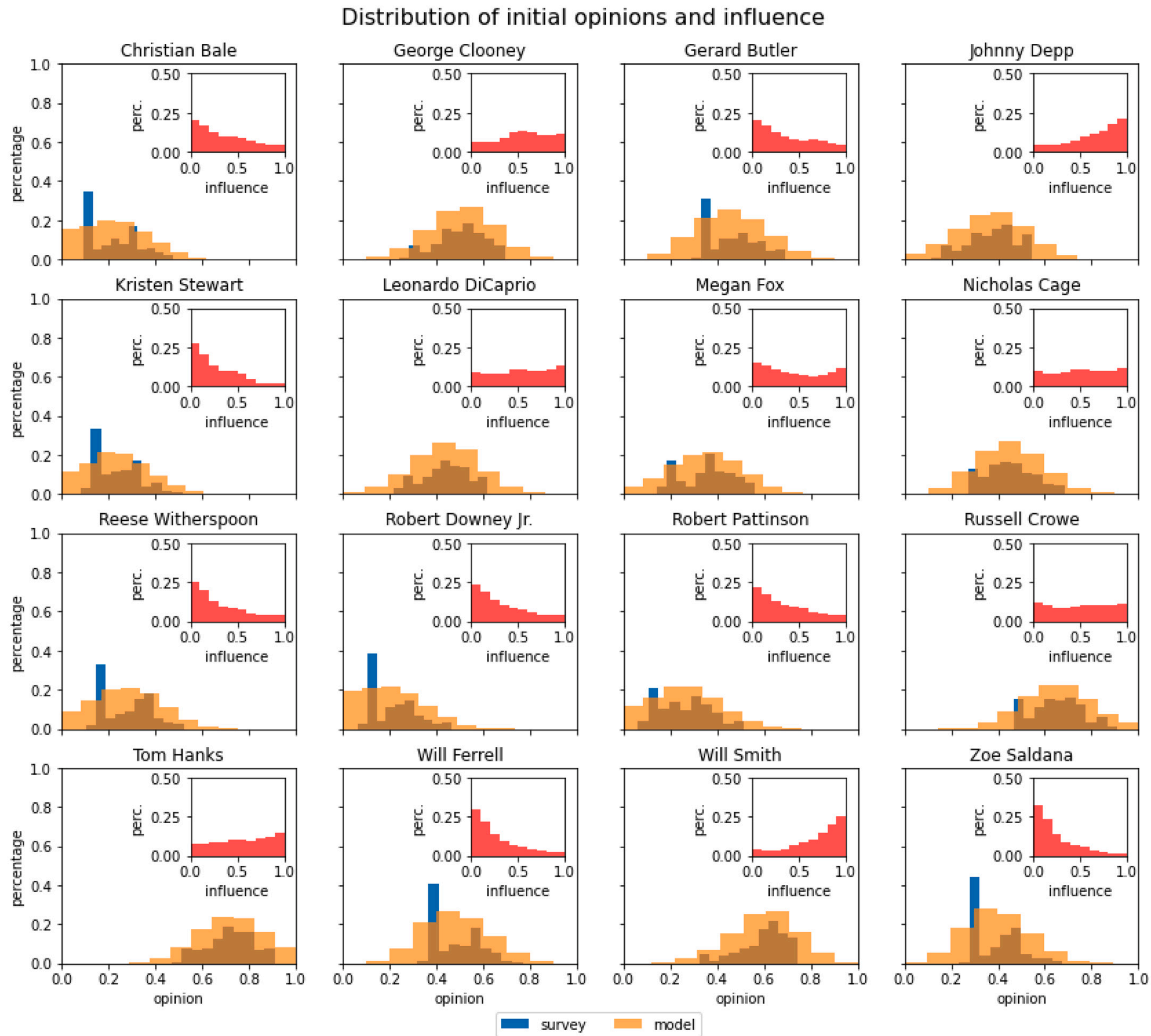

Fig. 2. Distribution of survey answers (blue) and their corresponding initial opinions in the model (orange) about the considered superstars. Inner plots represent the influence of each superstar in the population of spectator agents.

4.4. Modeling mass communication in the ATBCR model

The model previously described allows us to study the OD about superstars. These dynamics are the consequence of interactions between spectator agents (i.e. the audience). However, it lacks the effects of mass communication, i.e., the processes of exchanging information to a large portion of the population. In our context, film releases (and all the news and marketing campaigns related to them) represent these mass communication processes, which also have a major impact on the audience's opinion. This section presents an extension of our model to also capture the effects of mass communications.

The mass communication mechanism we define is inspired by Carletti et al. (2006). Each process m of mass communication is defined by a tuple $m = \langle \Psi, \upsilon, \kappa(i), s_k, t \rangle$, where the components of m are the following:

- Ψ is the opinion communicated in this process, which must be in the same representation and scale than spectator agents' opinions (in our model, a real number in the interval $[0,1]$).
- $\upsilon \in [0,1]$ is the (effective) reach that this process has in the population.
- $\kappa(i) \in [0,1]$ is the (effective) influence that this process has on spectator agent i.
- s_k is the subject m targets.
- t is the time step when this process m occurs.

Note that both υ and κ are the *effective* reach and influence, i.e., they do not model the potential impact that a marketing campaign may have in the population, but the actual values that they have.

The multidimensional extension of the ATBCR model proceeds as described in Section 4.1 with an additional mechanism to consider these mass communication processes. In particular, at every time step t that a mass communication process $m = \langle \Psi, \upsilon, \kappa(i), s_k, t \rangle$ takes place,

Table 2
Mass communication processes considered in the model. The influence $\kappa(i)$ of these processes is represented in Fig. 2.

Period	Superstar (s_k)	Ψ	υ	t
P1	J. Depp	0.7800	0.40	938
P1	T. Hanks	0.8000	0.30	2812
P2	Z. Saldana	0.7800	0.24	5312
P2	G. Butler	0.7900	0.50	6562
P2	G. Clooney	0.6700	0.36	7188
P2	N. Cage	0.8200	0.14	7500
P2	J. Depp	0.7250	0.44	8125
P2	L. DiCaprio	0.7100	0.40	8750
P2	G. Clooney	0.5850	0.40	9062
P2	K. Stewart	0.7800	0.32	9064
P2	R. Pattinson	0.7800	0.32	9065
P3	R. Downey Jr.	0.7650	0.36	10 312
P3	T. Hanks	0.7750	0.16	10 625
P3	C. Bale	0.7750	0.22	11 875
P3	G. Butler	0.6100	0.20	11 876
P3	N. Cage	0.8350	0.40	13 125
P3	R. Witherspoon	0.8500	0.42	13 126
P3	W. Ferrell	0.8050	0.08	13 750
P3	M. Fox	0.7300	0.16	14 062
P3	W. Ferrell	0.7450	0.26	14 375
P4	R. Downey Jr.	0.6600	0.32	16 562
P4	J. Depp	0.7300	0.32	16 875
P4	M. Fox	0.7150	0.22	17 188
P4	W. Smith	0.7150	0.42	17 500
P4	K. Stewart	0.7200	0.22	17 812
P4	R. Pattinson	0.7950	0.40	18 125
P4	C. Bale	0.6150	0.90	19 732
P5	W. Ferrell	0.7550	0.38	20 938
P5	R. Pattinson	0.7150	0.32	21 250
P5	Z. Saldana	0.8200	0.30	22 188
P5	N. Cage	0.7900	0.52	22 500

$\upsilon \cdot N$ spectator agents are randomly selected – the ones reached by this process –, and they all update their opinions as:

$$x_i^{s_k}(t+1) = x_i^{s_k}(t) + \mu_c \cdot \kappa(i) \cdot (\Psi - x_i^{s_k}(t)) \tag{6}$$

where $\mu_c \in [0, 0.5]$ is the convergence speed of the mass communication processes, and i is a spectator agent reached by this mass communication process m.

Combining agents' interactions (first and third cases of Eq. (1)) and mass communication processes (Eq. (6)), the resulting model is able to capture the complex dynamics of an evolving multidimensional scenario like opinions about superstars.

In order to model these mass communication processes, we use some information available at IMDb. In particular, we use the Starmeter rankings of IMDb, which ranks the popularity of movie superstars along time. For a mass communication process $m = \langle \Psi, \upsilon, \kappa(i), s_k, t \rangle$ representing a film release, the opinion Ψ is computed as the position of this film in the ranking of the best 200 films. For instance, top films transmit a very good opinion (close to 1). The reach υ is directly proportional to the number of news of this release (normalized over 500k). This information is also extracted from IMDb. Moreover, the time step t of this release is directly computed from its release date, and the subject s_k is the superstar starring such a film. Finally, the influence $\kappa(i)$ is directly extracted from the survey (Suárez-Vázquez, 2015) and modeled, for each agent i, as described in Section 4.2. This influence $\kappa(i)$ is depicted in Fig. 2. In Table 2 we report all the film releases considered in our analysis.

4.5. Benchmarking the model accuracy: comparison to specialized cinema real-world data

In our analysis, we compare the output of our OD model designed from real-world data from the survey about superstars used in Suárez-Vázquez (2015) with respect to the information available in the specialized cinema website IMDb.

In our simulation, we consider the six Starmeter rankings of IMDb published between May 1st, 2011 (the closest date to the survey by Suárez-Vázquez (2015)) and November 11th, 2012. In particular, IMDb publishes the Starmeter ranking every 16 weeks. Therefore, we use the rankings published on 05/01/2011, 08/21/2011, 12/11/2011, 04/01/2012, 07/22/2012, and 11/11/2012. The period between these six Starmeter rankings comprise approximately one year and a half. We also consider all the films released in this interval. In this way, our simulation covers a significant period where the initial opinions reflected in the survey can evolve as a consequence of the opinion dynamics and mass communication processes. Our model is independently executed for each superstar during $T = 25{,}000$ time steps, generating intermediate rankings of opinions every 5,000 time steps (one for each period). Notice that each period of 5,000 time steps corresponds to 16 weeks (i.e., the time between the publication of two Starmeter rankings), hence each day comprises around 45 interactions in our model.

In order to measure the accuracy of the model, the superstars are ranked according to the spectator agents' opinions about them. In particular, the final opinion profile of the population about each superstar is averaged, sorting these average opinions to produce the OD ranking. This ranking is equal to the one obtained with a positional voting system (with Borda count) (Saari, 1995). Finally, the OD ranking is compared to the actual ranking of IMDb (Starmeter) containing only the 16 superstars at study. In order to compare these two conjoint rankings, we use the Rank Biased Overlap (RBO) (Webber et al., 2010). This metric returns a value in [0,1], with 0 indicating a total discrepancy, and 1 indicating a complete correlation, and it is computed as:

$$RBO(R_1, R_2, p) = (1-p) \sum_{d=1}^{|R_1|} p^{d-1} \cdot AGG_d \tag{7}$$

where R_1 and R_2 are the rankings to be compared, p is the steepness of discrepancy weights, and AGG_d is the agreement of rankings R_1 and R_2 at depth d, defined as:

$$AGG_d = \frac{|R_{1[:d]} \cap R_{2[:d]}|}{d} \tag{8}$$

with $R_{[:d]}$ denoting the first d elements of ranking R.

As discussed by Webber et al. (2010), RBO solves several disadvantages of other ranking comparison metrics, including the Kendall's τ coefficient. For instance, τ gives the same weight to every discrepancy,[7] regardless the position of the ranking where they occur (e.g., a discrepancy at the top of the ranking has the same weight than one at the bottom), whereas RBO overcomes this drawback. RBO is also able to handle non-conjoint pairs of rankings, although in our case this requirement is not necessary since both rankings contain exactly the same set of superstars.

The global accuracy of our model is the average RBO of the six ranking comparisons, one for each period considered.

5. Experimental analysis

In this section we present the experimental analysis on the OD about the 16 very well-known superstars considered.

Since spectator agents only interact with other agents in their neighborhood (when they are connected in the social network, i.e., $A_{ij} = 1$), the distribution of initial opinions in the graph can have a major impact on the OD. To solve this, we perform a number of independent Monte Carlo (MC) executions of the model, differing in the seeding of the opinions within the nodes of the graph. Each MC execution returns differences in the OD about superstars. However, these differences only

[7] A discrepancy is a pair of two elements in different order in each ranking.

produced negligible differences in the output OD rankings. Therefore, and for the sake of reducing the computational complexity of the study, in the rest of our analysis the reported results represent the average accuracy of 3 MC executions.

In the following subsections, we first present a sensitivity analysis of the proposed model. Then, we present a fine-grained analysis of the most accurate configuration of our model. Finally, we present a comparison between our model and other state-of-the-art OD methods.

5.1. Sensitivity analysis of the proposed model

In a first experiment, we perform a sensitivity analysis of our model. In particular, we analyze the confidence and repulsion thresholds, and how they affect the accuracy. Notice that these thresholds model the rationality of the system, i.e., whether the evolution of spectator agent's opinions is driven by emotional (or rational) mechanisms. As commonly analyzed in the literature, the rest of the parameters of our model are fixed to $\mu = 0.2$ and $\mu_c = 0.5$.

In Fig. 3 we represent the results of this sensitivity analysis. The most accurate scenario, with an average $RBO = 0.473833$, is found with a confidence threshold $\varepsilon = 0.3$ and a repulsion threshold $\vartheta = 0.5$. This represents a system with high confidence but also high repulsion. This scenario is analyzed in more details in the following subsection.

The sensitivity analysis also shows that this system is highly ruled by the repulsion mechanism, i.e., executions with a higher repulsion threshold return, in general, more accurate results (see the right area of Fig. 3). These results match the expected behavior of the OD about superstars, which are, in general, based on emotional sentiments, with a relatively low degree of rationality.

This result is in line with the hedonic perspective of movie goers' behavior, under what the emotional component of behavior dominates the cognitive component (Eliashberg & Sawhney, 1994). It also gives an explanation to the classic statement "*nobody knows anything*" by screenwriter William Goldman. Our OD study shows that, at least as far as individual moviegoers' decisions are concerned, superstars power is not a matter of knowledge, but of emotions. Thus, it could be said that in the film industry "*nobody knows anything*" but "*everybody feels something*" and those feelings strongly influence the power of superstars. When measuring the value of superstars, in addition to characteristics such as experience or awards (Wei, 2006), their emotional value must be taken into account. On an aggregate level, this result may offer an explanation for the weak relationship between expert judgments and popular appeal (Holbrook, 2005), as the latter is expected to be more strongly affected by the emotional value of stars.

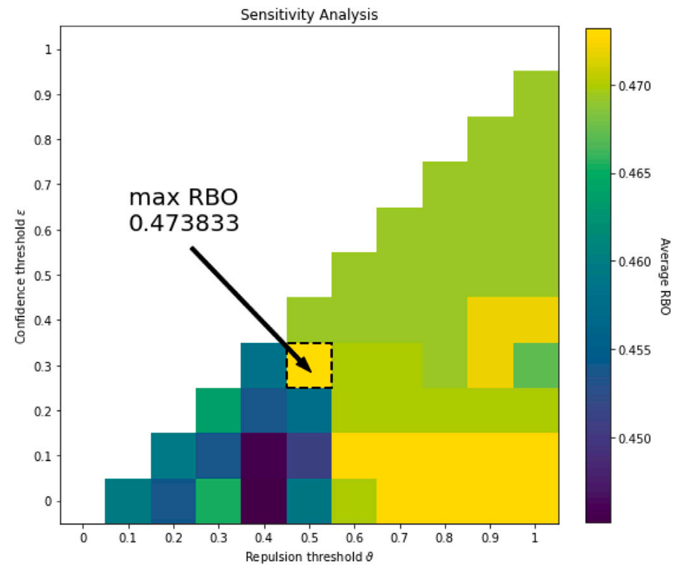

Fig. 3. Sensitivity analysis of OD about superstars in terms of the confidence and repulsion thresholds of the proposed model.

5.2. Fine-grained analysis of the proposed model

Next, we analyze the detailed results of our model executed with $\varepsilon = 0.3$ and $\vartheta = 0.5$ (the most accurate scenario found previously). In Fig. 4 we report the OD about each superstar, where blue points represent the opinion evolution (each point reflects the opinion of a single spectator agent at a specific time in a $[0, 1]$ scale) as a consequence of spectator agents' interactions along time, represented by the X axis; red points represent the final opinions at each period (including the initial opinions); and green points represent the impact of mass communication processes. We also include subplots with the distribution of final opinions at the end of the last period.

An interesting observation from these detailed results is the distinct effect on the opinions of spectator agents' interactions and mass communication processes. On the one hand, agents' interaction tend to form a major consensus in the long term. See, for instance, the distribution of final opinions (right subplots of each superstar), which exhibits small differences. On the other hand, mass communication only has a small impact on the dynamics of the opinions. In particular, the opinions resulting from mass communication processes diverge from the average opinion, producing both bad and good opinions in the spectator agents (see the green points in Fig. 4). However, this phenomenon only has a short term effect.

These results allow us to explain the different behaviors of the audience in both the short and the long term. In the short term, just after a release, the audience may be highly influenced by (aggressive) marketing campaigns. However, these campaigns may be counterbalanced by the spectator's experience, which is shared with other spectators influencing each other, and this explains the success of the films (and the superstars) in the long term. This phenomenon has been observed in several films and can be interpreted as a consequence of how success film drivers change between short- and long-term box office (Hennig-Thurau et al., 2006). A paradigmatic example of this phenomenon was the first movie "My bit fat Greek wedding", released in 2002. This romantic comedy became "*one of the most profitable films in history*" that "*went viral when there was not such a thing as going viral*" (Goldenberg, 2016).

Also, in Table 3 we report the OD and the IMDb rankings for the six periods, as well as the RBO value for their comparison. The evolution of these RBO values is represented in Fig. 5. These results show that the RBO value is improved in all the periods, except the second one (although this degradation is small). We conjecture that this is due to the number of films released in this period, most of them having a very large influence. Nevertheless, the RBO of the other four periods shows improvements with respect to the baseline RBO of initial opinions. This suggests that our model is able to adequately represent the OD in this system. We must emphasize that, although these RBO values do not show a total consensus, the ranking of opinions is based on a particular population whereas the IMDb ranking is based on a specialized cinema website. Therefore, certain divergences are expected. Moreover, there exist some differences in the nature of the data. On the one hand, the survey (initial opinions in the model) provides a transversal snapshot of the opinions of this population in a particular moment. On the other hand, the ranking of IMDb is the result of behavioral analysis of users in this website (thus not strictly based on opinions) in a lapse of time. In fact, this reveals one of the main open problems in this field: how to establish a ranking of superstars, which motivates the present work. Even so, the performance of the model is pretty satisfactory according to the RBO values and their evolution.

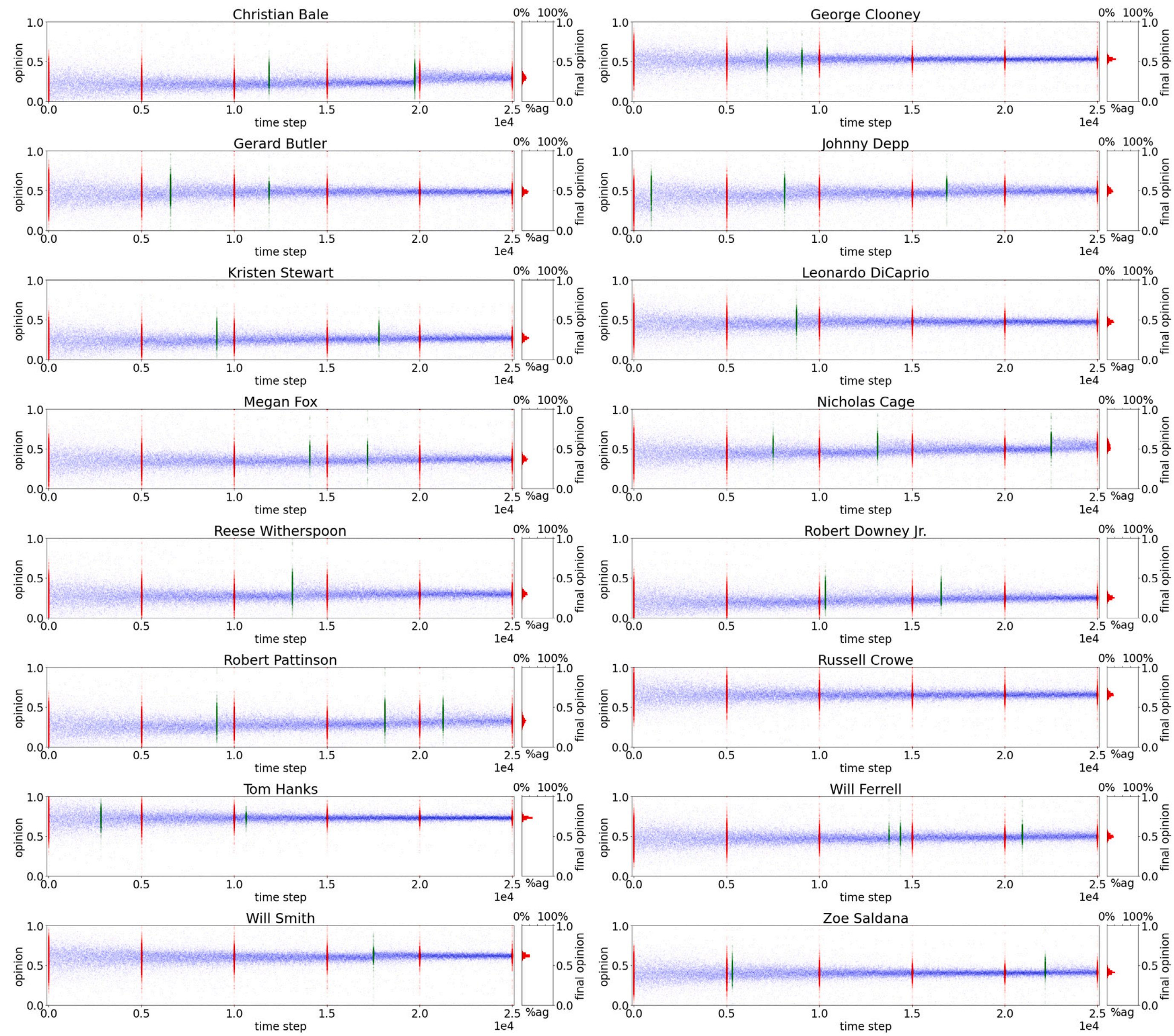

Fig. 4. OD about the 16 analyzed superstars in the model executed with $\varepsilon = 0.3$ and $\vartheta = 0.5$.

5.3. Comparison with other methods

In this subsection, we finally present a comparison between the proposed method based on the ATBCR model and other state-of-the-art OD models in the literature. In particular, we implement adaptations of our methodology based on the DW model (Deffuant et al., 2000), the DeGroot model (Degroot, 1974), and the FJ model (Friedkin & Johnsen, 1990).

In order to adapt our methodology to these OD models, we follow the same steps discussed in Section 4, i.e., we implement a multidimensional extension of the specific OD model, we initialize the initial opinions using the real-world data from the existing survey (Suárez-Vázquez, 2015), we model agents' interactions using a PA network, we model mass communication processes representing film releases, and we measure the accuracy of the output comparing the obtained ranking of superstars versus the real-world ranking of the specialized website IMDb, using the RBO value. The only difference between these adaptations is the underlying OD model that captures how opinions are updated along time.

Table 4 reports a comparison on the RBO value obtained for these adaptations, based on the ATBCR, the DW, the DeGroot, and the (two variants of the) FJ models. In most of the six periods analyzed, as well as in the aggregate global average RBO value of the six periods, the proposed methodology based on the ATBCR model is the most accurate one, or it shows a performance very close to the optimal model.

In a sensitivity analysis of the DW model adaptation, we found that the best configuration is achieved with $\varepsilon = 0.1$. This surprising value represents a very low confidence system, composed of very *stubborn* spectator agents, i.e., in general these agents only change their opinions after an interaction with another agent having a very similar opinion, thus their resulting opinions are almost unchanged. We emphasize that, in contrast to this DW-based adaptation, the adaptation based on the ATBCR model returned, as the most accurate configuration, a highly emotional system with a relatively high degree of confidence as well (see Section 5.1). Qualitatively, the DW-based model is unable to capture the OD about superstars, which are expected to be mostly driven by emotional mechanisms (as the ATBCR model is able to capture, as showed above).

Table 3
Rankings of superstars using the proposed OD model and the IMDb Starmeter in the analyzed periods.

	Period P0		Period P1		Period P2	
Rank.	OD	IMDb	OD	IMDb	OD	IMDb
1	T. Hanks	C. Bale	T. Hanks	G. Clooney	T. Hanks	G. Clooney
2	R. Crowe	J. Depp	R. Crowe	J. Depp	R. Crowe	K. Stewart
3	W. Smith	K. Stewart	W. Smith	M. Fox	W. Smith	J. Depp
4	G. Clooney	R. Pattinson	G. Clooney	K. Stewart	G. Clooney	R. Pattinson
5	W. Ferrell	L. DiCaprio	W. Ferrell	T. Hanks	G. Butler	L. DiCaprio
6	G. Butler	N. Cage	G. Butler	L. DiCaprio	W. Ferrell	Z. Saldana
7	N. Cage	R. Witherspoon	N. Cage	C. Bale	L. DiCaprio	R. Downey Jr.
8	L. DiCaprio	R. Downey Jr.	L. DiCaprio	G. Butler	J. Depp	W. Smith
9	Z. Saldana	T. Hanks	J. Depp	Z. Saldana	N. Cage	C. Bale
10	J. Depp	G. Butler	Z. Saldana	N. Cage	Z. Saldana	G. Butler
11	M. Fox	W. Smith	M. Fox	R. Witherspoon	M. Fox	N. Cage
12	R. Witherspoon	R. Crowe	R. Witherspoon	R. Pattinson	R. Pattinson	M. Fox
13	R. Pattinson	W. Ferrell	R. Pattinson	R. Downey Jr.	R. Witherspoon	T. Hanks
14	K. Stewart	M. Fox	K. Stewart	W. Smith	K. Stewart	R. Witherspoon
15	C. Bale	Z. Saldana	C. Bale	W. Ferrell	C. Bale	W. Ferrell
16	R. Downey Jr.	G. Clooney	R. Downey Jr.	R. Crowe	R. Downey Jr.	R. Crowe
RBO	0.433		0.449		0.393	

	Period P3		Period P4		Period P5	
Rank.	OD	IMDb	OD	IMDb	OD	IMDb
1	T. Hanks	G. Clooney	T. Hanks	C. Bale	T. Hanks	C. Bale
2	R. Crowe	J. Depp	R. Crowe	J. Depp	R. Crowe	J. Depp
3	W. Smith	C. Bale	W. Smith	K. Stewart	W. Smith	K. Stewart
4	G. Clooney	K. Stewart	G. Clooney	G. Clooney	N. Cage	R. Downey Jr.
5	N. Cage	L. DiCaprio	N. Cage	R. Downey Jr.	G. Clooney	G. Clooney
6	G. Butler	R. Downey Jr.	J. Depp	L. DiCaprio	W. Ferrell	T. Hanks
7	W. Ferrell	N. Cage	G. Butler	W. Smith	J. Depp	L. DiCaprio
8	L. DiCaprio	Z. Saldana	W. Ferrell	W. Ferrell	G. Butler	N. Cage
9	J. Depp	W. Smith	L. DiCaprio	T. Hanks	L. DiCaprio	G. Butler
10	Z. Saldana	W. Ferrell	Z. Saldana	M. Fox	Z. Saldana	W. Ferrell
11	M. Fox	T. Hanks	M. Fox	R. Pattinson	M. Fox	W. Smith
12	R. Witherspoon	R. Witherspoon	R. Pattinson	G. Butler	R. Pattinson	R. Pattinson
13	R. Pattinson	M. Fox	R. Witherspoon	N. Cage	R. Witherspoon	R. Crowe
14	K. Stewart	R. Pattinson	C. Bale	Z. Saldana	C. Bale	Z. Saldana
15	C. Bale	G. Butler	K. Stewart	R. Witherspoon	K. Stewart	M. Fox
16	R. Downey Jr.	R. Crowe	R. Downey Jr.	R. Crowe	R. Downey Jr.	R. Witherspoon
RBO	0.492		0.507		0.569	

Table 4
Comparison between adaptations of the proposed methodology based on several OD models, computed as the RBO value for each period and the global average $\langle RBO \rangle$.

Model	P0	P1	P2	P3	P4	P5	$\langle RBO \rangle$
ATBCR-based		0.449	0.393	**0.492**	0.507	**0.569**	**0.473833**
DW-based		0.449	**0.428**	**0.492**	0.467	0.448	0.452833
DeGroot-based	0.433	0.449	0.410	0.481	**0.513**	0.531	0.469619
FJ-based (low stubbornness)		0.449	0.401	**0.492**	0.473	**0.569**	0.469603
FJ-based (high stubbornness)		**0.475**	0.401	0.481	0.484	0.479	0.459020

In the uniform adaptation of the DeGroot model to our system, we recognized that it is less accurate that the one based on the ATBCR model. This suggests that modeling spectator agents with a mostly rational behavior, even when they are fully susceptible to other opinions, is inaccurate to model the OD about superstars. Again, this result reinforces the need to use emotional mechanisms (as those in the ATBCR model) to represent this complex system.

Finally, we analyzed a uniform adaptation of the FJ model to our methodology. In particular, we analyzed two scenarios differing in the degree of stubbornness of the model: low (stubbornness, we use ($\xi = 0.6$) and high ($\xi = 0.3$) stubbornness. Notice that $\xi = 1$ represents a case without stubbornness, which is exactly the DeGroot model. Empirically, we found that both adaptations exhibit an overall worse performance than the ATBCR model. This suggests that neither a fully rational behavior nor a highly stubborn one are able to capture the underlying OD about superstars, where emotions play a key role in these dynamics (that the ATBCR model is able to capture).

In summary, our global study suggests that, as expected, the OD about superstars is driven by emotional mechanisms and hence the underlying OD model to represent it needs a repulsion rule (as the one proposed in the ATBCR model; see the third case of Eq. (1)) to capture it. In contrast, any model ignoring such a repulsion nature of OD in the real-world (e.g., the DW model, the DeGroot model, and the FJ model) would be unable to reproduce the emotional mechanisms of this system and, hence, it would be unable to accurately capture the OD about superstars.

Our proposal reflects the fact that cinema business is "*a business of extremes*" due in part to "*the way moviegoers dynamically influence one another*" (De Vany & Walls, 2004). Indeed, the ATBCR model can capture the dynamic process of the cinema market overcoming proposals, such as Suárez-Vázquez (2015), that offer a fixed picture of the market in a given point in time.

Existing survey-based studies – which fall into what Eliashberg et al. (2006) categorized as the psychological approach to movie-going behavior literature – seek to explain the individual decision-making processes of moviegoers. These survey-based studies focus on the individual by analyzing how various perceptual variables, such as attitudes or emotions, affect key cinema behaviors. Our model can integrate this knowledge about individual opinions into a dynamic process in which an individual's opinions evolve as they are affected by the opinions of

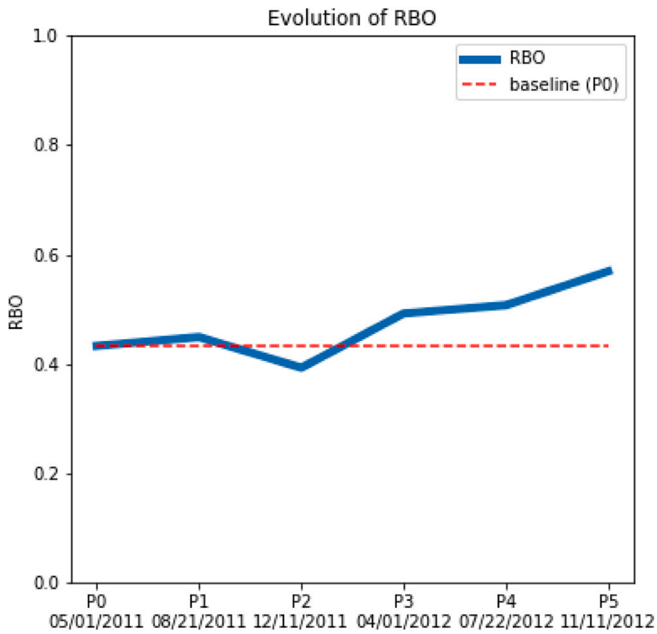

Fig. 5. Evolution of the RBO value in the analyzed periods.

others. To this end, and to model the mass communication processes, information about film releases, news and marketing campaigns is taken into account, which enriches the proposal with aggregated data on movie-going behavior. The use of survey-based data and industry data provides a deeper understanding of the complex and dynamic nature of cinema demand. The integrated approach deepens the results of pure survey-based data, making it possible to explain not only how specific emotions elicited by superstars affect individual moviegoers' decisions but also how interaction with others affects the relative value of superstars, i.e. their ranking position.

6. Conclusions and future works

In this work, we presented a study of OD about superstars in the film industry in order to establish a ranking of them based on a set of eliciting emotions of the audience, represented via spectator agents. In particular, we use a survey on a real-world population (Suárez-Vázquez, 2015) to initialize opinions about a set of well-known superstars, and extend the ATBCR model of OD (Giráldez-Cru et al., 2022) with a mass communication process to represent the evolution of these opinions along time. To model these mass communication processes representing film releases and marketing campaigns related to them, we use real-world information available in the specialized cinema website IMDb. Additionally, a social network is used to model spectator agent's interactions in a realistic manner. Using this OD model, a ranking of superstars is computed, and compared to the external, independent Starmeter ranking of IMDb. We emphasize that both the initial data and the information used to validate the model are real-world, in contrast to other OD model where only synthetic data is used. To the best of our knowledge, this is the first work that proposes a ranking of superstars based on an OD model using real-world data for both the initialization of opinions and the validation of the results. Moreover, we consider that all the design decisions of our model may be useful to implement other real-world applications based on OD models, including, e.g., models for politics, marketing, and sociology, among others.

Our results reveal that this model is able to accurately capture the OD of this complex system. In fact, the opinion evolution about superstars mostly evolves by repulsion mechanisms, i.e., OD about superstars are, in general, based on emotional sentiments, with a relatively low degree of rationality. Moreover, our analysis shows the distinct impact of spectator agents' interactions and mass communication processes. In particular, we found out that film releases only have a major impact on the audience's opinions in the short term, i.e., they may be able to alter spectator agents' opinions substantially just after the mass communication process occurs, but their effects do not endure in the long term, where spectator agents' interactions are the main influence factors that rule the dynamics of this complex system.

Future testings of our model could involve some modifications of it to improve or increase its advantages. When focusing on actors and actresses we are using a narrow definition of superstar. Directors and producers could also be considered (Wei, 2006). This study analyzes the impact on spectators of only one leading actor per movie. Possible model replications could consider the synergistic effect between different stars (Elberse, 2007). It would also be convenient to replicate the study using a different dataset – focusing on another type of population – to initialize spectators' opinions. Another important research direction would be to integrate a more fine-grained indicator for publicity by considering aspects such as valence and quality of the marketing campaigns (Hofmann & Opitz, 2019). Finally, it would be interesting to analyze other graph topologies to represent real-world interactions, including the small-world (Watts & Strogatz, 1998) and the Lancichinetti–Fortunato–Radicchi (Lancichinetti et al., 2008) models.

Besides, the proposed methodology based on the design of OD models from user survey data including mass communication processes also designed from real-world data can be applied to other marketing and consumer behavior analysis areas. On the one hand, it can be used to model how consumers react to marketing campaigns and viral word-of-mouth processes (Chica et al., 2023; Chica & Rand, 2017; Delre et al., 2010; Suárez-Vázque & Chica, 2021). The perceptions of the consumers about the different brands in a market can also be obtained from (tracking data) survey data, as the opinions about superstars in the current study. The spreading of those perceptions within the social network of consumers and the effects caused by both this local diffusion process and the global action of the marketing campaigns can be modeled using our OD approach. On the other hand, it can also be applied to model and analyze the user intention to adopt digital payment services, making use of survey data as that considered in Bhatia et al. (2023), among others consumer behavior applications.

CRediT authorship contribution statement

Jesús Giráldez-Cru: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing. **Ana Suárez-Vázquez:** Conceptualization, Validation, Investigation, Resources, Writing – original draft, Writing – review & editing. **Carmen Zarco:** Conceptualization, Validation, Resources, Writing – review & editing. **Oscar Cordón:** Conceptualization, Validation, Resources, Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research has been developed within the R&D project CONFIA (PID2021-122916NB-I00), funded by MICIU/AEI/10.13039/5011000 11033/ and FEDER, EU. J. Giráldez-Cru is also supported through the Juan de la Cierva program (IJC2019-040489-I). Funding for open access charge: Universidad de Granada / CBUA.

References

Adler, M. (1985). Stardom and talent. *The American Economic Review*, *75*(1), 208–212.

Ainslie, A., Dreze, X., & Zufryden, F. (2005). Modeling movie lifecycles and market share. *Marketing Science*, *24*, 508–517.

Anagnostopoulos, A., Becchetti, L., Cruciani, E., Pasquale, F., & Rizzo, S. (2022). Biased opinion dynamics: when the devil is in the details. *Information Sciences*, *593*, 49–63.

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*, 509–512.

Behrens, R., Foutz, N. Z., Franklin, M., Funk, J., Gutierrez-Navratil, F., Hofmann, J., & Leibfried, U. (2021). Leveraging analytics to produce compelling and profitable film content. *Journal of Cultural Economics*, *45*, 171–211.

Berger, R. L. (1981). A necessary and sufficient condition for reaching a consensus using DeGroot's method. *Journal of the American Statistical Association*, *76*(374), 415–418.

Bhatia, S., Singh, N., & Liébana-Cabanillas, F. (2023). Intermittent continued adoption of digital payment services during the covid-19 induced pandemic. *International Journal of Human–Computer Interaction*, *39*(14), 2905–2919.

Borghans, L., & Groot, L. (1998). Superstardom and monopolistic power: Why media stars earn more than their marginal contribution to welfare. *Journal of Institutional and Theoretical Economics*, *154*(3), 546–571.

Broekhuizen, T. L., Delre, S. A., & Torres, A. (2011). Simulating the cinema market: How cross-cultural differences in social influence explain box office distributions. *Journal of Product Innovation Management*, *28*(2), 204–217.

Carletti, T., Fanelli, D., Grolli, S., & Guarino, A. (2006). How to make an efficient propaganda. *Europhysics Letters*, *74*(2), 222–228.

Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of Modern Physics*, *81*, 591–646.

Chang, B. H., & Ki, E. J. (2005). Devising a practical model for predicting theatrical movie success: Focusing on the experience good property. *Journal of Media Economics*, *18*(4), 247–269.

Chen, J., Kou, G., Wang, H., & Zhao, Y. (2021). Influence identification of opinion leaders in social networks: an agent-based simulation on competing advertisements. *Information Fusion*, *76*, 227–242.

Chica, M., Cordón, O., Robles, J. F., Garrido, A., Mingot, J., & Damas, S. (2023). Zio: an artificial intelligence digital twin to build virtual markets. In *2023 IEEE conference on artificial intelligence* (pp. 173–174).

Chica, M., & Rand, W. (2017). Building agent-based decision support systems for word-of-mouth programs: A freemium application. *Journal of Marketing Research*, *54*(5), 752–767.

Cuadrado, M., & Frasquet, M. (1999). Segmentation of cinema audiences: An exploratory study applied to young consumers. *Journal of Cultural Economics*, *23*, 257–267.

De Vany, A., & Walls, W. D. (1999). Uncertainty in the movie industry: Does star power reduce the terror of the box office? *Journal of Cultural Economics*, *23*, 285–318.

De Vany, A. S., & Walls, W. D. (2004). Motion picture profit, the stable paretian hypothesis, and the curse of the superstar. *Journal of Economic Dynamics & Control*, *28*(6), 1035–1057.

Deffuant, G., Neau, D., Amblard, F., & Weisbuch, G. (2000). Mixing beliefs among interacting agents. *Advances in Complex Systems*, *3*(01n04), 87–98.

Degroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, *69*(345), 118–121.

Delre, S. A., Jager, W., Bijmolt, T. H., & Janssen, M. A. (2010). Will it spread or not? The effects of social influences and network topology on innovation diffusion. *Journal of Product Innovation Management*, *27*(2), 267–282.

Díaz, A., Gómez, M., Molina, A., & Santos, J. (2018). A segmentation study of cinema consumers based on values and lifestyle. *Journal of Retailing and Consumer Services*, *41*, 79–89.

Dong, Y., Zhan, M., Kou, G., Ding, Z., & Liang, H. (2018). A survey on the fusion process in opinion dynamics. *Information Fusion*, *43*, 57–65.

Elberse, A. (2007). The power of stars: Do star actors drive the success of movies? *Journal of Marketing*, *71*(4), 102–120.

Eliashberg, J., Elberse, A., & Leenders, M. A. (2006). The motion picture industry: Critical issues in practice, current research, and new research directions. *Marketing Science*, *25*(6), 638–661.

Eliashberg, J., Jonker, J., Sawhney, M. S., & Wierenga, B. (2000). MOVIEMOD: An implementable decision-support system for prerelease market evaluation of motion pictures. *Marketing Science*, *19*(3), 226–243.

Eliashberg, J., & Sawhney, M. S. (1994). Modeling goes to hollywood: Predicting individual differences in movie enjoyment. *Management Science*, *40*(9), 1151–1173.

Epstein, J. M. (2006). *Generative social science: Studies in agent-based computational modeling*. Princeton University Press.

Farmer, J. D., & Foley, D. (2009). The economy needs agent-based modelling. *Nature*, *460*, 685–686.

Frank, R., & Cook, P. (1995). *The Winner-Take-All Society*. New York: Free Press.

Friedkin, N. E., & Johnsen, E. C. (1990). Social influence and opinions. *Journal of Mathematical Sociology*, *15*(3–4), 193–206.

Ghiassi, M., Lio, D., & Moon, B. (2015). Pre-production forecasting of movie revenues with a dynamic artificial neural network. *Expert Systems with Applications*, *42*(6), 3176–3193.

Giráldez-Cru, J., Zarco, C., & Cordón, O. (2022). Analyzing the extremization of opinions in a general framework of bounded confidence and repulsion. *Information Sciences*, *609*, 1256–1270.

Goldenberg, D. (2016). The first "My big fat Greek wedding" was the blockbuster nobody saw coming. FiveTrhirtyEight. https://fivethirtyeight.com/features/the-first-my-big-fat-greek-wedding-was-the-blockbuster-nobody-saw-coming/.

Hadida, A. L. (2009). Motion picture performance: A review and research agenda. *International Journal of Management Reviews*, *11*(3), 297–335.

Harashima, T. (2016). Ranking value and preference: A model of superstardom. *The Munich Personal RePEC Archive*, 74626.

Hauser, J., Tellis, G., & Griffin, A. (2006). Research on innovation: A review and agenda for marketing science. *Marketing Science*, *25*(6), 687–717.

Hegselmann, R., & Krause, U. (2002). Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation*, *5*(3).

Hennig-Thurau, T., Houston, M. B., & Walsh, G. (2006). The differing roles of success drivers across sequential channels: An application to the motion picture industry. *Journal of the Academy of Marketing Science*, *34*(4), 559–575.

Hennig-Thurau, T., Houston, M. B., & Walsh, G. (2007). Determinants of motion picture box office and profitability: an interrelationship approach. *Review of Managerial Science*, *1*, 65–92.

Hofmann, J., Clement, M., Völckner, F., & Hennig-Thurau, T. (2017). Empirical generalizations on the impact of stars on the economic success of movies. *International Journal of Research in Marketing*, *34*(2), 442–461.

Hofmann, K. H., & Opitz, C. (2019). Talent and publicity as determinants of superstar incomes: empirical evidence from the motion picture industry. *Applied Economics*, *51*(13), 1383–1395.

Holbrook, M. B. (1999). Popular appeal versus expert judgments of motion pictures. *Journal of Consumer Research*, *26*(2), 144–155.

Holbrook, M. B. (2005). The role of ordinary evaluations in the market for popular culture: Do consumers have "good taste"? *Marketing Letters*, *16*, 75–86.

Isaac-Goizé, T. (2023). The original influencer was French – A show dedicated to Sarah Bernhardt celebrates the world's first superstar. https://www.vogue.com/article/sarah-bernhardt-paris-petit-palais-exhibition.

Isenberg, D. J. (1986). Group polarization: A critical review and meta-analysis. *Journal of Personality and Social Psychology*, *50*(6), 1141–1151.

Jiao, Y., & Li, Y. (2021). An active opinion dynamics model: the gap between the voting result and group opinion. *Information Fusion*, *65*, 128–146.

Karniouchina, E. V. (2011). Are virtual markets efficient predictors of new product success? The case of the Hollywood stock exchange. *Journal of Product Innovation Management*, *28*(4), 470–484.

King, B. (1987). The star and the commodity: Notes towards a performance theory of stardom. *Cultural Studies*, *1*(2), 145–161.

Kübler, R., Seifert, R., & Kandziora, M. (2021). Content valuation strategies for digital subscription platforms. *Journal of Cultural Economics*, *45*(2), 295–326.

Lancichinetti, A., Fortunato, S., & Radicchi, F. (2008). Benchmark graphs for testing community detection algorithms. *Physical Review E*, *78*, Article 046110.

Laros, F. J., & Steenkamp, J.-B. E. (2005). Emotions in consumer behavior: a hierarchical approach. *Journal of Business Research*, *58*(10), 1437–1445.

Li, Y., Liu, M., Cao, J., Wang, X., & Zhang, N. (2021). Multi-attribute group decision-making considering opinion dynamics. *Expert Systems with Applications*, *184*, Article 115479.

Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing*, *70*(3), 74–89.

Luo, L., Chen, X., Han, J., & Park, C. W. (2010). Dilution and enhancement of celebrity brands through sequential movie releases. *Journal of Marketing Research*, *47*(6), 1114–1128.

McDonald, P. (2012). *Hollywood stardom*. John Wiley and Sons.

McKenzie, J. (2023). The economics of movies (revisited): A survey of recent literature. *Journal of Economic Surveys*, *37*(2), 480–525.

Moraes, M., Gountas, J., Gountas, S., & Sharma, P. (2019). Celebrity influences on consumer decision making: New insights and research directions. *Journal of Marketing Management*, *35*(13–14), 1159–1192.

Motion Picture Association (MPA) (2021). 2021 THEME report. https://www.motionpictures.org/research-docs/2021-theme-report/.

Nelson, R. A., & Glotfelty, R. (2012). Movie stars and box office revenues: An empirical analysis. *Journal of Cultural Economics*, *36*, 141–166.

Noorazar, H., Sottile, M. J., & Vixie, K. R. (2018). An energy-based interaction model for population opinion dynamics with topic coupling. *International Journal of Modern Physics C*, *29*(11), Article 1850115.

Peterson, R. A. (2001). On the use of college students in social science research: Insights from a second-order meta-analysis. *Journal of Consumer Research*, *28*(3), 450–461.

Rammstedt, B., Kemper, C. J., & Borg, I. (2013). Correcting Big Five personality measurements for acquiescence: An 18–country cross–cultural study. *European Journal of Personality*, *27*(1), 71–81.

Ravid, S. A. (1999). Information, blockbusters, and stars: A study of the film industry. *Journal of Business*, *72*(4), 463–492.

Rosen, S. (1981). The economics of superstars. *The American Economic Review*, *71*(5), 845–858.

Saari, D. G. (1995). *Basic geometry of voting*. Springer.

Sawhney, M. S., & Eliashberg, J. (1996). A parsimonious model for forecasting gross box-office revenues of motion pictures. *Marketing Science*, *15*, 113–131.
Scherer, K. R. (2009). Emotions are emergent processes: they require a dynamic computational architecture. *Philosophical Transactions of the Royal Society, Series B (Biological Sciences)*, *364*(1535), 3459–3474.
Suárez-Vázque, A., & Chica, M. (2021). Others' fortune in online vs offline settings: How envy affects people's intention to share information. *Internet Research*, *31*(5), 1641–1655.
Suárez-Vázquez, A. (2011). Critic power or star power? The influence of hallmarks of quality of motion pictures: An experimental approach. *Journal of Cultural Economics*, *35*, 119–135.
Suárez-Vázquez, A. (2015). Superstars as emotion-eliciting objects. An examination of the effect of the emotion mix of movie stars. *Journal of Media Economics*, *28*(2), 84–99.
Taherdoost, H. (2017). Determining sample size; how to calculate survey sample size. *International Journal of Economics and Management Systems*, *2*, 237–239.
Tan, E. S. (2013). *Emotion and the structure of narrative film: Film as an emotion machine*. Routledge.
Terry, N., Cooley, J. W., & Zachary, M. (2010). The determinants of foreign box office revenue for English language movies. *Journal of International Business and Cultural Studies*, *2*, 1.
Wang, W.-T., He, Y.-L., Huang, J. Z., & Ma, L.-H. (2020). A new approach to solve opinion dynamics on complex networks. *Expert Systems with Applications*, *145*, Article 113132.
Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*, 440–442.
Webber, W., Moffat, A., & Zobel, J. (2010). A similarity measure for indefinite rankings. *ACM Transactions on Information Systems*, *28*(4), 20:1–20:38.
Wei, L. (2006). Invited commentary — Making sense of these million-dollar babies—Rationale behind superstar profit participation contracts. *Marketing Science*, *25*(6), 678–680.
Xia, H., Wang, H., & Xuan, Z. (2011). Opinion dynamics: A multidisciplinary review and perspective on future research. *International Journal of Knowledge and Systems Science*, *2*(4), 72–91.
Zhang, Y., Chen, X., Gao, L., Dong, Y., & Pedryczc, W. (2022). Consensus reaching with trust evolution in social network group decision making. *Expert Systems with Applications*, *188*, Article 116022.
Zillmann, D. (2015). Mood management: Using entertainment to full advantage. *Communication, Social Cognition, and Affect*, *31*, 147–171.