



ARTICLE



<https://doi.org/10.1057/s41599-022-01143-2>

OPEN

# A Chinese tale of three regions: a century of China in thousands of films

Zhuo Chen<sup>1,4</sup>, Wen Ma<sup>1,4</sup>, Wei Guo<sup>2,3</sup>  & Yunsong Chen<sup>1</sup>  

Over the past 70 years, because of their different historical and social contexts, the film industries of China's three regions, namely mainland China, Taiwan, and the Hong Kong Special Administrative Region (SAR), have exhibited distinct characteristics that grow out of a single cultural root. Based on data from the Internet Movie Database (IMDb), this study investigates how the image of China has spread overseas through cultural products by applying big data analysis and machine learning methods to compare the content, topics, and attitudes toward the image of China disseminated by films originating in these three regions. The results show that the three areas have sought to express their subjectivity during periods of flux and striven to connect with the world in films. The macroscopic analysis of large-scale content allows the exploration of hidden cultural phenomena and compensates for the lack of objectivity of traditional research methods, while understanding embedded historical and cultural contexts of the three regions helps to clarify the regions' ethnic, cultural, and emotional connections. Therefore, this study has contributed to cultural sociology both in topic and methodology.

<sup>1</sup>Department of Sociology, School of Social and Behavioral Sciences, Nanjing University, Nanjing, P. R. China. <sup>2</sup>Department of Social Work and Social Policy, School of Social and Behavioral Sciences, Nanjing University, Nanjing, P. R. China. <sup>3</sup>The Centre for Asia-Pacific Development Studies, Nanjing University, Nanjing, P. R. China. <sup>4</sup>These authors contributed equally: Zhuo Chen, Wen Ma. ✉email: [weiguo@nju.edu.cn](mailto:weiguo@nju.edu.cn); [yunsong.chen@nju.edu.cn](mailto:yunsong.chen@nju.edu.cn)

## Introduction

Films represent both film industry and pieces of visual art. The film industry has thrived in mainland China, Taiwan, and the Hong Kong Special Administrative Region (SAR) since it was first introduced to China at the end of the 19th century (Lu, 1997). The films from the above three regions have developed in different directions and, while they are mainly independent, each region exerts some influence on the others (Tan, 2020). The differences and changes in the sociopolitical environments of the three regions have created complex contexts for the films they have produced. The history of film in China began with the Shanghai film industry, which boomed before 1949. After the founding of the People's Republic of China (PRC) in 1949, the Communist Party of China (CCP) ruled mainland China and the Nationalist Party of China (CNP) occupied Taiwan, while until 1997 Hong Kong was a British colony, which explained the reason that the three regions and their films took very different paths. However, with the reforms and the opening up of mainland China, collaboration between the three regions has increased since the 1990s. A golden age for the Chinese film industry is emergent.

The different histories and cultures of the three regions have generated distinct film styles. However, their films share the same cultural roots and show robust national features. Some scholars have argued that the Chinese have used the Western concept of the nation-state to participate in the international order and to resist the West (Berry and Farquhar, 2006, p. 2). Therefore, an analysis of China's image as evinced in the films from the three regions may elucidate the process of China's cultural connection with the Western world.

A national image is shaped by both objective and subjective factors. According to Kenneth Boulding (1959), a country's national image represents the amalgamation of its understanding of itself as well as outsiders' view of the nation. Thus, the subjective impression is pivotal to the interpretation of a national image. The cognitive subject belongs to a specific society, and individual impressions must bear the brand of the local culture, as is particularly evident in Taiwan and Hong Kong. As former colonies, Taiwan and Hong Kong were influenced by the states that dominated them. A separation between "self" and "the other" was thus bound to appear in their standpoints vis-à-vis China.

Films can integrate national images into specific plots and pictures and thus denote one of the most vivid ways to express such cognitive representations. The audience effects model of agenda-setting theory states that media such as films influence audience perceptions; in turn, audience feedback modifies the focus of these media (Erbring et al., 1980; McCombs, 2005). In reality, in the era of the Internet, editors of leading Western websites place selected films about/from China on their online platforms and this selection governs Western audiences' initial view of China. The cinematic portrayals of China subsequently enable a second construction, the outsiders' view of China. The two processes reflect the influence of the media on the consciousness of audiences. Understanding the choice of film topics and the expression of ideas facilitates the study of the process through which the Western mind forms its impression of China.

This study conducted a big data analysis of China's image in the Western world. In so doing, it covered a century of films produced in the three regions mentioned, thus signifying a novel deconstruction of China's film history; it also represents a pioneering use of context analysis methodology. Traditional film studies have focused on case analysis (e.g., Lu, 1997; Li, 2016; Yeh, 2018), but the conclusions deduced from specific instances cannot guarantee objectivity, generalizability, and comprehensiveness. In this study, the above issues were resolved using the Word2vec toolkit, the latent Dirichlet allocation (LDA) topic model, and

Google's sentiment analysis model to study the synopses of pre-2019 film plots on the Internet movie database (IMDb). A spatiotemporal analysis framework was constructed to explore changes in the conception of China in the content, topics, and sentiments of films from mainland China, Taiwan, and Hong Kong.

## Background

When describing Chinese films in the past, scholars usually constructed a linear historical narrative to explore a unified and coherent national boundary (Yoshimoto, 1991). However, the grand conceptual construct of China actually comprises varied content types such as multiple spoken languages, former colonies, and religious affiliations (Berry and Farquhar, 2006). The emergent notion of the Chinese-language film thus transcends the singular nation-state narrative, encompassing all types of local or global cinematic renditions related to the Chinese language (Sun, 2016). However, some scholars have argued that viewing Chinese films through the lens of a Western cultural system could cause their "de-China-lization" (Li, 2016), which means abandoning Chinese cultural connotation to cater to western understanding. Chinese films face the challenge of balancing globalization and nationality as they attempt to establish a new subjectivity.

The confrontation between globalization and nationality is particularly evident in mainland China. Since the founding of the PRC, filmmakers have made great efforts to resist foreign cultural invasion. Films have thus become a crucial political component of nationalism (Lu, 1997) and a powerful instrument in the promotion of socialist ideology. After the reform and opening-up launched in the late 1970s, a new generation of directors including Zhang Yimou described a charming and barbaric East in films such as *Raise the Red Lantern* (1991). Although narratives that showed the eeriness of the East achieved massive success in the West, postcolonialists were highly critical of these renditions (Li and Wang, 2009). The situation changed again as China entered the 21st century: the economy and the film industry in mainland China soared together. Some films, such as *The Wandering Earth* (2019), combined a Hollywood narrative style and national confidence with mature film technology, expressing the ambition to dominate the international order and successfully constructing a China-centric subjectivity.

The portrayal of China is more complicated in Taiwanese films than in films originating from mainland China. As a former Japanese colony, Taiwan belongs to the mainland China-Taiwan-Japan cultural triangle, and its colonial history and postcolonial context have caused an identity crisis (Li and Wang, 2009). Homi Bhabha's (1994) postcolonial theory posits that the colonist's culture creates a continued pain in the colony, forming a hybrid culture. Taiwanese filmmakers have validated their identity through just such a cross-cultural construct. Directors such as Hsiao-Hsien Hou and Edward Yang contemplated Taiwan's turbulent history in the 1980s, exhibiting an intense nativeness and discovering a narrative path between Chinese and Western cultures. However, in establishing its nativeness, Taiwan emotionally and culturally distanced itself from the mainland. Films like *Cape No.7* (2008) manifested this de-China-lization and even evoked nostalgia for Japanese culture, revealing the lingering presence of the ghosts of the Japanese colonial period.

A former British colony, Hong Kong's early films followed British practices of live entertainment and included vaudevilles and cabarets (Yeh, 2018, p. 6). The Hong Kong film industry was the first of the three regions to commercialize, producing many

comedies and kung fu films in the 1970s, when films starring Bruce Lee became immensely popular worldwide. For instance, movies such as *Way of the Dragon* (1972), were set in a foreign country and represented “ethnic Chinese.” They showcased a powerful national character by portraying the development of Chinese kung fu, and attained global fame as they defended Chinese identity. As an urban center representing ideological confrontations, Hong Kong has always avoided politics and prioritized entertainment, which helped the tide of consumerism sweep through its film industry. Nevertheless, consumerism could not dispel the identity-related dilemma of the colonized, and numerous films continued to discuss this issue after Hong Kong’s return to the PRC. For example, *Port of Call* (2015) described the integration of mainland China residents into Hong Kong society, showing that this region still sought its identity through the story of “the other.”

Mainland China, Taiwan, and Hong Kong share their origins but have taken different paths. Their cultural expressions of China have thus diverged under the joint influence of various factors, including the economy and politics. To a great extent, Taiwan leans toward the Japanese; Hong Kong has an affinity with Great Britain and the United States; and mainland China has embarked on the socialist road guided by Marxism. Faced with the West, however, all three regions have been forced to show their nativeness as “the other.” According to Bhabha (1990, p. 294; 1994, p. 34), the problematic boundaries of modernity are sketched out in these ambivalent temporalities of the nation-space. This ambivalence has evoked an irreversible inclination to construct cultural differences that can revise the history of critical theory. Therefore, the trend toward globalization drives the need to grasp the cultural distinctions within a nation. After a long cultural collision with the suzerain state, the former colony generates new power through mimicry and finally achieves cultural hybridity.

As a vital aspect of postcolonialism, numerous ideological and social forces have informed the impression of China expressed in the films of the past 70 years. It is difficult for theoretical studies to accurately compare the discrete features of the films of the three regions and to grasp their changes over time. Therefore, using big data analysis and machine learning methods, we explored the transformations observed in the image of China in the films produced in mainland China, Taiwan, and Hong Kong SAR over the last century from a postcolonial theoretical perspective.

## Data and methods

**Data sources.** Traditional methods using questionnaire surveys cannot accurately measure a country’s image, which is diffused through an entire society and entails both subjectivity and universality. However, based on massive content, the current capabilities of unsupervised machine learning open the door to new methodologies. Natural language processing technologies permit the analysis of an extensive range of content and deliver a better measure of cultural concepts for social science disciplines.

As the data source, we selected the plot synopses of China-related films from mainland China, Taiwan, and Hong Kong collected via IMDb. Founded in 1990, IMDb is a high-quality international online database of films, providing information including film plots, casts, regions of production and release, ratings, and reviews. As of June 2020, the IMDb collection consisted of 552,366 films. The database has been widely used in scholarly research. For example, researchers have applied statistical models to classify the sentiments expressed in reviews of films (Tripathy et al., 2016) and to predict the popularity of future films (Abidi et al., 2020). Film reviews have also been used

to study the attitudes of audiences toward elite values (Ridderström, 2018) or to explore the politicized role of humor (Ridanpää, 2014). New quantitative and qualitative research breakthroughs have been reported using large datasets such as IMDb.

This study used the terms “China” and “Chinese” as keywords to screen pre-2019 film data from the three regions to determine the image of China. All of the selected films articulated an idea of China for international audiences. Some of the films involved multi-regional collaborations; thus, to avoid duplication, the first region where the film is made was considered the standard of judgment. After removing error values, 1047 relevant films were found: 682 from mainland China, 104 from Taiwan, and 261 from Hong Kong.<sup>1</sup> For historical and political reasons, we mainly compared the films of the three regions between 1949 and 2018. We used web crawler technology to grab titles and plot synopses from IMDb, reflecting the plot and setting of each film. Furthermore, information about each film’s region of production and release year was compiled to explore the process of transforming China’s image, as well as the factors related to the films’ social contexts.

**Measurement methods.** The 1047 films selected from IMDb showed a multi-dimensional image of China. The quantitative analysis performed in this study endeavored to address three aspects of the main topics of the films from the three regions. We looked at their mentions of China or the Chinese people; their specific propensities to treat particular topics; and their positive or negative attitudes in describing China. We used text mining in the analysis because film titles and plot synopses represent unstructured textual data.

**Contextual measurement.** We used word embedding technology to predict the contexts of “China.” Google’s open-source toolkit, the Word2vec algorithm, can turn unstructured textual data into high-dimensional vector data. The similarity between words can be measured, and the context can be predicted by calculating the distance between the high-dimensional vectors.

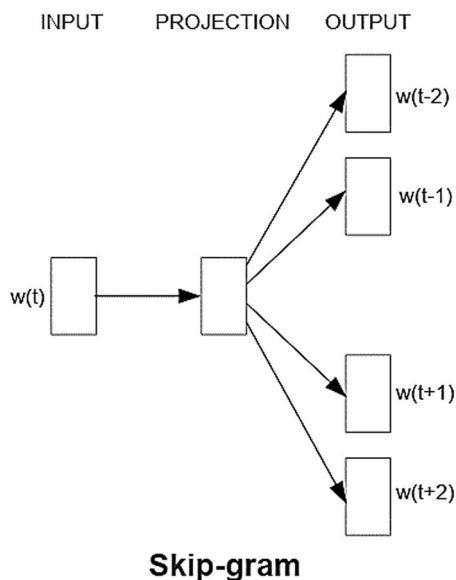
The Skip-gram algorithm of Word2vec, based on a shallow neural network, was selected for the current study. The cosine similarity between the input vector of the input word (the current word) and the output vector of the target word (the context of the current word) was calculated and normalized after words were encoded. Specifically, the Python Gensim library was used for training the Word2vec word vector model. After carrying out word segmentation, stemming, and removal of stop words, each word was transformed into a 128-dimensional vector through Skip-gram. “China” was selected as the focal word, and the cosine distance between the focal and other words was calculated to measure the contexts. Figure 1 displays the Skip-gram algorithm (Mikolov et al., 2013).

**Topic analysis.** Films have changed according to the historical development of the different regions. The special relationship between mainland China, Taiwan, and Hong Kong in the past has left the regions with closely related, yet very different, images of China. The films made in the three regions often showcase their preferred topics or events in their descriptions of China. LDA topic mining can help tease these out.

LDA is a three-level hierarchical Bayesian model for document topic generation using unsupervised learning techniques (Blei et al., 2003). The application of LDA is based on three nested concepts: a corpus, or the text set to be modeled; a document, or an item in the corpus; and a term, or a word in the document. Documents are nested in the corpus, and terms are nested in the document.

The LDA model assumes the existence of some topics within the corpus, and each topic is defined as the probability distribution of fixed words. Each document selects a topic with a certain probability, and a specific word is selected from this chosen topic. A corpus is described as the probability distribution of potential topics, and each topic is a probability distribution of terms. Data analysis was performed by using the joint distribution to calculate the conditional (posterior) distribution of the hidden variable (subject structure) within a given observed variable (word in the document). The formula was as follows:

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D}) = \prod_{i=1}^K p(\beta_i) \prod_{d=1}^D p(\theta_d) \left( \prod_{n=1}^N p(z_{d,n} | \theta_d) p(w_{d,n} | \beta_{1:K}, z_{d,n}) \right)$$



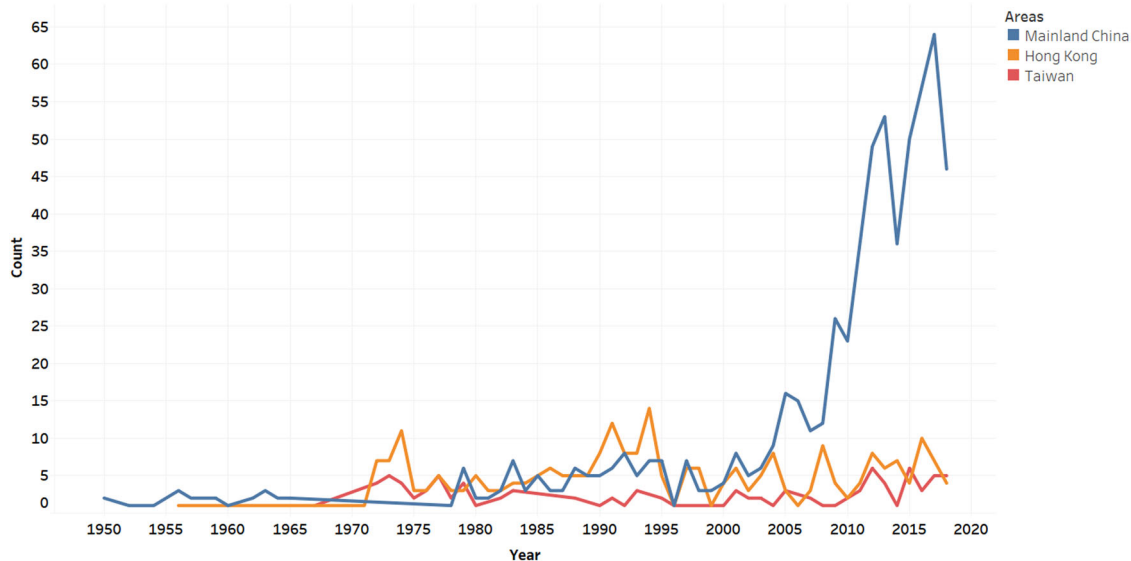
**Fig. 1 The skip-gram algorithm of Word2vec.** The skip-gram algorithm uses each current word ( $w(t)$ ) as an input to a log-linear classifier with continuous projection layer and predicts words ( $w(t-2)$ - $w(t+2)$ ) within a certain range before and after the current word.

In this formula,  $\beta_{(1:K)}$  denotes all topics, while  $\beta_K$  represents the distribution of the words for the  $K$ th topic.  $\theta_{(1:D)}$  indicates all documents, while  $\theta_d$  is the proportion of the topic in the  $d$ th document, and  $\theta_{d,k}$  reflects the proportion of the  $K$ th topic in the  $d$ th document. All topics of the  $d$ th document are represented by  $z_{d,b}$  while  $z_{d,n}$  denotes the topic of the  $n$ th word in the  $d$ th document. All words in the  $d$ th document constitute  $w_{d,b}$  and  $w_{d,n}$  represents the  $n$ th word in the  $d$ th document. Compared with a simple co-occurrence analysis, the LDA topic model reveals the underlying semantic relationships between words, even if the words never appear together in a document. Unlike other topic cluster models, the LDA topic model uses a distinct mixed-membership approach that allows a document to span various topics, ensuring the accuracy of the analysis.

**Sentiment analysis.** We also applied sentiment analysis to calculate the overall sentiment score of the relevant films from a more abstract perspective. Sentiment analysis entails the use of opinions or emotions (affective states) in the examination, cognition, and inference of subjective texts. The text hints at the direction of the feelings of different people. Researchers can use the trends that are discerned to analyze the public's views or evaluations on a specific issue using two methods of analysis: the sentiment dictionary can be used to assess the rule matching method; and the sentiment classification task can be set up based on the sentiment dictionary and text features.

The sentiment analysis for this study was performed through Google Cloud Platform, which allows users to run such an evaluation via its application programming interface (API). After the text information is entered, the API can assess the overall attitude of the text, using a score ranging between  $-1.0$  (negative) and  $+1.0$  (positive) to denote the tendency of the sentiment expressed. Negative scores represent adverse emotions and opinions; positive scores convey constructive sentiments. The greater the absolute value, the stronger the sentiment, while the closer the number is to 0, the more neutral the sentiment.

**Contexts in China-related films.** Figure 2 illustrates the number of China-related films originating annually from mainland China, Hong Kong, and Taiwan. The number of films produced in mainland China was relatively low for the three decades following the founding of the PRC, possibly due to the immaturity of film



**Fig. 2 Number of China-related films from mainland China, Taiwan, and Hong Kong, by year.** The line chart reflects the changes in the number of China-related films in mainland China (blue), Hong Kong (orange), and Taiwan (red) from 1950 to 2020.

**Table 1 Distance of words to “China”.**

Mainland China		Taiwan		Hong Kong	
single	0.999557	life	0.999683	revolve	0.999557
cinema	0.999541	Chinese	0.999647	story	0.999504
overcome	0.999537	Taiwan	0.999584	hit	0.999504
resident	0.999527	one	0.999573	conflict	0.999492
province	0.999517	family	0.999562	canton	0.999489
development	0.999505	world	0.999548	nothing	0.999489
sound	0.999492	time	0.999538	lead	0.999488
poverty	0.999492	work	0.999534	camp	0.999475
comrade	0.999490	film	0.999533	lonely	0.999468
truth	0.999489	Taiwanese	0.999521	emerge	0.999467
stand	0.999489	Japanese	0.999520	journey	0.999465
enemy	0.999489	young	0.999491	series	0.999465
market	0.999488	run	0.999485	little	0.999461
minority	0.999485	sword	0.999485	worker	0.999460
freedom	0.999483	girl	0.999479	follow	0.999458
course	0.999482	old	0.999471	search	0.999458
secret	0.999481	village	0.999443	lady	0.999458
today	0.999481	war	0.999437	sun	0.999457
common	0.999479	love	0.999437	fortune	0.999456
economic	0.999475	martial	0.999434	hero	0.999454

technology and the Cultural Revolution. After China’s reform and opening-up, the number of films increased explosively, given the context of the booming economy and the need to build national confidence.

China-related films peaked in Taiwan and Hong Kong in the 1970s and 1990s, respectively, perhaps because of the cognitive reconstruction of the concept “China” by their citizens. These periods coincided with turning points in relations between the two regions and the mainland. In 1971, the PRC resumed its lawful seat in the United Nations, prompting people in Taiwan and Hong Kong to pay attention to China’s position and build on China’s legacy in their films. The Sino–British Joint Declaration that returned Hong Kong to the PRC in 1997 was announced in 1984. Hong Kong, which had been ruled by Great Britain since 1841, re-examined its relationship with the mainland. This reassessment was directly reflected in films made in Hong Kong, and the number of relevant films increased in the following decade (1985–1995).

To further explore the image of China constructed by films, the Skip-gram algorithm of Word2vec was used to train a 128-dimensional word vector model based on the film plot corpus using “China” as the central word to predict the principal context. Table 1 lists the words with the shortest cosine distance to the central word “China,” which are also the words with the greatest similarity to the focal word.

The results of Word2vec reveal that “single” denotes the term closest in films made in mainland China to the focal word. In terms of specific context, “single” is mostly combined with “child” and “mother,” revealing the films’ preference for family-related settings and their attention to the one-child policy and single mothers. In addition, the contextual terms around the word “China” primarily encompass three categories: first, *politics*, a classification involving mentions of the country’s military and foreign policy or words such as “comrade”; second, the category of the *enemy*, referring to opponents of the nation and socialism; and third, the grouping of *minority*, signifying issues pertaining to ethnic populations and their governance. The second category focuses on the provincial economy and livelihoods and typically includes words such as “resident,” “province,” “poverty,” “market,” and “economic.” The third group discusses China’s past and future, describing its historical memory and prospective

development, and incorporates words like “overcome,” “development,” and “today.”

The China-related context of Taiwanese films seems more oriented toward everyday life than the deliberate positive discourse on China in mainland films, using words such as “life,” “family,” “work,” “girl,” “love,” and “village.” Most of the relevant films focus on cross-strait exchanges, describing individual life experiences in Taiwan and mainland China. The listed words “Chinese” and “Taiwanese” exhibit the distinction drawn between the identities of the people of Taiwan and “the other” character of mainland Chinese citizens. Meanwhile, the term “Japanese” reflects the complicated Japanese context of Taiwanese films, and the words “war” and “martial” hint at the significance of military power to Taiwan.

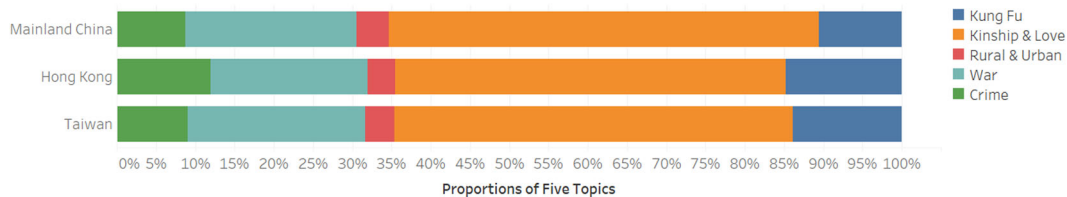
The scenes of Hong Kong films are more relevant to the administrative status of the metropolitan area and include words like “revolve” and “canton,” depicting its past status as a British colony and its current standing as a SAR of the PRC. Furthermore, terms such as “hit,” “conflict,” and “hero” reveal Hong Kong’s tradition of making films featuring kung fu, police actions, and bandits. Additionally, the words “worker” and “fortune” encompass the exchange of labor and capital between Hong Kong and the mainland.

Notably, the image of China constructed by these films spans the different perceptions held by the three regions and incorporates the attitudes of IMDb editors in summarizing the content of the films. The synopses of films produced in mainland China include words such as “Comrade,” “freedom,” and “poverty”; summaries of films originating in Taiwan use terms such as “world” and “war”; and the outlines of Hong Kong movies contain terms such as “canton” and “fortune.” To some extent, these words reveal the subjective understanding of the international audience represented by IMDb editors. Influenced by their ideologies, the editors reveal certain sociopolitical tendencies, for instance by emphasizing liberal values and regional sovereignty.

**Topics of China-related films.** Although the films of the three regions influence each other to some degree, we postulated that the topics covered by the films would be different because of their respective social backgrounds and political environments.

**Table 2 LDA topic analysis of China-related films.**

Topic 1 Kung Fu	Topic 2 Kinship & Love	Topic 3 Rural & Urban	Topic 4 War	Topic 5 Crime					
art	0.058	life	0.029	village	0.105	Japanese	0.039	Hong	0.076
martial	0.047	love	0.024	modern	0.098	war	0.027	school	0.066
master	0.047	young	0.021	night	0.063	shanghai	0.026	Kong	0.066
fu	0.038	family	0.020	foreign	0.061	force	0.024	student	0.057
team	0.032	live	0.016	road	0.051	army	0.019	police	0.040
gang	0.031	woman	0.015	mission	0.048	help	0.019	order	0.037
great	0.026	city	0.013	say	0.047	work	0.019	kill	0.035
kung	0.025	old	0.012	shot	0.043	town	0.018	brother	0.032
fight	0.022	girl	0.011	hand	0.039	fight	0.018	government	0.023
money	0.022	man	0.009	question	0.039	group	0.018	short	0.022
result	0.019	home	0.008	deadly	0.037	back	0.018	head	0.022
yuan	0.018	series	0.008	body	0.037	leave	0.017	hero	0.021
rural	0.018	friend	0.008	outside	0.030	American	0.017	truth	0.020
power	0.017	Beijing	0.008	catch	0.029	learn	0.016	doctor	0.020
dragon	0.017	traditional	0.008	train	0.027	couple	0.016	leader	0.019
international	0.017	parent	0.008	local	0.019	return	0.015	murder	0.018
artist	0.017	movie	0.008	back	0.016	end	0.015	teacher	0.016
win	0.016	son	0.008	fail	0.003	red	0.014	officer	0.016



**Fig. 3 Proportions of films on the five topics in the three regions.** The bar chart compares the proportions of kung fu (blue), kinship and love (orange), rural and urban (red), war (sky blue), and crime (green) topics in Mainland China, Hong Kong, and Taiwan.

Bringing out these distinctions would facilitate understanding of the three regions’ national identities. The LDA topic model was deployed to perform unsupervised cluster analysis based on the corpus of film synopses. After taking into account the overall length of the text and the model’s actual effect,<sup>2</sup> the model with five topics was selected to carry out the analysis.

Table 2 presents the top 20 words with the highest frequency corresponding to each topic after excluding words with no practical significance. Each word in each plot synopsis was assigned to a topic to calculate the distribution of synopses on each topic. Words pertaining to each topic were extracted, and the five topic names were summarized as “kung fu,” “kinship and love,” “rural and urban,” “war,” and “crime.”

Most of the words in Topic 1 are related to kung fu, with “martial art,” and “kung fu,” as examples. Other terms include fighting verbs (e.g., “fight” and “win”), fighting objects (e.g., “master,” “team,” “gang,” and “man”), and evaluations (e.g., “great” and “international”). Words such as “evil” and “dragon” are also commonly used in this topic, revealing the indomitable spirit of Chinese people in the face of evil as “Chinese dragon’s (i.e., long) offspring.” Films belonging to this topic are primarily action-oriented, with thrilling stunts at their core to showcase traditional Chinese kung fu. They often involve scenes of fighting, destruction, and rescue. They call for social justice and strike the audience’s senses. Representative movies include *Along Comes the Tiger* (1977) from Taiwan and *Kung Fu Quest* (2013) from Hong Kong.<sup>3</sup>

Topic 2, kinship and love, is generally exemplified by words of affection (e.g., “love,” “young,” “woman,” “girl,” and “man”) and family associations (e.g., “family,” “home,” “friend,” “parent,” and “son”). The contexts of this topic relate to everyday life and often

focus on the lives of ordinary people, resonating with audiences through emotional expression. Typical works include *Rong and Her Mother in Law* (2012), *The Naughty boy Ma Xiaotiao* (2009), and *The Romance beside the Well* (2013).<sup>4</sup>

The representative words for Topic 3, rural and urban, include “village,” “modern,” “night,” “foreign,” “road,” “outside,” “local,” “shot,” and “catch.” These words may reflect the collisions between the rural and the urban, the domestic and the international, and the local and the foreign, capturing their mutual integration and adaptation in the context of modernization. Terms such as “shot,” “catch,” and “deadly” reveal crime and violence as related to public security or guerrilla warfare. This topic’s two highest-rated films are *Bruce Lee’s Sister* (1974) and *Vampire VS Vampire* (1989). However, only 28.6% and 22.6% of the words in their plot synopses describe the topic, implying that the topic often appears as the background to the plot rather than as its primary content.

Topic 4 predominantly discusses war through words like “war,” “force,” “army,” and “fight.” Topic 4 is often anchored in significant historical military actions, describing scenes of battle, establishing the image of a hero, and reflecting on the disasters and trauma triggered by war. Words alluding to foreigners are also observed, for example, “Japanese” and “American,” reflecting the fact that war in mainland China, Taiwan, and Hong Kong often involved Japan and the United States. Typical works are *Eternal Wave* (2017) and *Black Sun 731* (1988), in the context of the Anti-Japanese War (hereafter referred to as “AJW”).<sup>5</sup>

Most words used for Topic 5 relate to police activities and crime, for instance, “police,” “order,” “kill,” “hero,” “murder,” “officer,” and “cop.” This topic focuses on the antagonism between the police force and criminal elements, justice and evil,

**Table 3 ANOVA and LSD tests on the proportion of film topics in the three regions.**

	MC	HK	TW	ANOVA-F	MC vs. HK	MC vs. TW	HK vs. TW
Kung Fu	0.106	0.148	0.138	32.467***	-0.042***	-0.032***	0.010
Kinship & Love	0.547	0.497	0.507	21.658***	0.050***	0.040***	-0.010
Rural & Urban	0.041	0.035	0.038	2.756*	0.006**	0.003	-0.003
War	0.219	0.200	0.226	5.570***	0.018***	-0.007	-0.026***
Crime	0.087	0.119	0.090	26.686***	-0.032***	-0.003	0.029***

MC represents the average proportion of relevant topics in mainland China; HK represents the average proportion of relevant topics in Hong Kong; TW represents the average proportion of relevant topics in Taiwan; ANOVA-F represents the F value in the ANOVA test; MC vs. HK represents the LSD test between MC and HK; MC vs. TW represents the LSD test between MC and TW; and HK vs. TW represents the LSD test between HK and TW.  
 \*\*\*p < 0.01, \*\*p < 0.05, \*p < 0.1.

or law and wrongdoing, involving the detection of illegal acts and the fight against gang-related offenses. Meanwhile, “Hong Kong” is the topic’s core word, demonstrating that the topic is usually set in Hong Kong society. Films with the highest proportion of plot synopses classified in this topic are all produced in Hong Kong: *Long Arm of the Law: Part 2* (1987), *Fist of Fury II* (1992), and *Police Story III* (1992).<sup>6</sup>

The average proportion of each topic was calculated by region to further compare the differences between each region’s films and the distribution of the five topics between the regions. Figure 3 shows that the three regions share elements of the distribution of topics. The topic of kinship and love is the most common context, accounting for more than 50% of the films and denoting the highest proportion of films. The topic of war describes nearly 20% of the films, and films on kung fu and crime represent around 10% of the total. The topic of rural and urban has the lowest share in all three regions at <5%.

Meanwhile, the distribution of topics in the three regions also differs. To ensure the robustness of the results, we performed an analysis of variance (ANOVA) and least significant difference (LSD) tests on the proportion of film topics in the different regions. The detailed results are shown in Table 3.

First, the proportion of films about kinship and love in mainland China is significantly higher than that in Taiwan and Hong Kong. Specifically, films from mainland China on this topic account for 55% of the total, 4% and 5% higher than Taiwan and Hong Kong, respectively. Therefore, mainland Chinese films are more inclined to focus on the everyday context and to depict the feelings, social relations, and destiny of characters at the micro level. There are also significantly more films about rural and urban issues in mainland China and in Taiwan than in Hong Kong, which is more modern and internationalized. In addition, the long-standing household registration (*Hukou*) system that divides the mainland Chinese population into urban and rural residents with strikingly different living conditions and social welfare support schemes may have led to more films depicting both the conflicts and the connections between rural and urban areas.

Second, most Hong Kong films focus on kung fu and crime. Table 3 shows that the number of Hong Kong films about kung fu (15%) is significantly higher than that of mainland Chinese films (11%), while the proportion of crime films in Hong Kong (12%) is greater than that in the mainland (9%) and in Taiwan (9%). The film industry in Hong Kong showed increased commercialization and marketization characteristics in the early stages of film development, meaning that high-drama action films flourished there. Thus, kung fu and crime films in Hong Kong were of high quality and well received both at home and abroad. In addition, as a former British colony and an international port city, the Hong Kong film industry learned from the crime films that were popular in Europe and America.

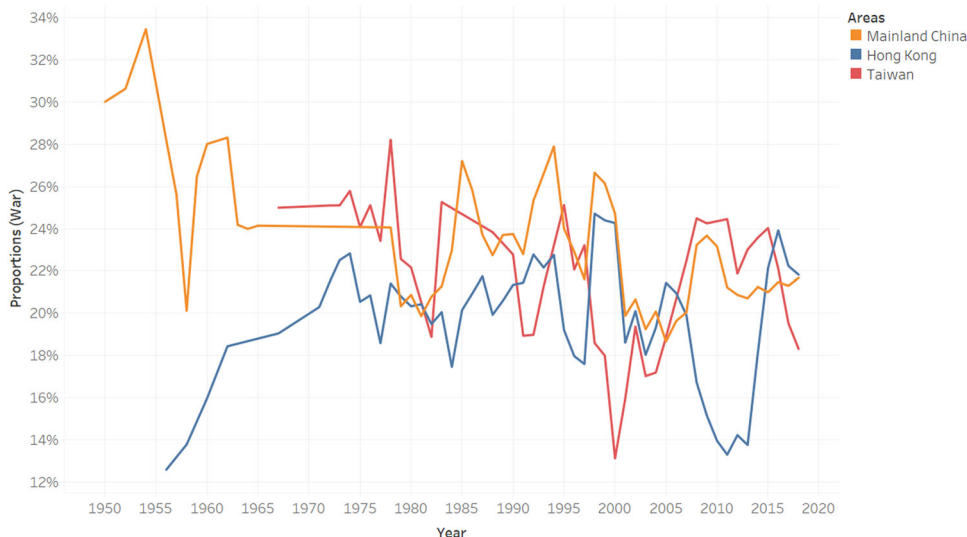
Third, Taiwan’s proportion of war films is around 23%, slightly higher than the percentage of mainland China (22%) and significantly higher than that of Hong Kong (20%). Taiwan has not been the main battlefield of any modern war, but it shows a preference for war stories in its films, especially the AJW. The Taiwanese share China’s historical memory of wars as an aspect of their ethnicity. Thus, a group of patriotic filmmakers in Taiwan uphold the ethnic, kinship, and geographical relationships between Taiwan and the mainland through military films set in various modern wars (Chen, 2008). However, Japan colonized Taiwan from 1895 to 1945, and many Japanese elements therefore persist in Taiwanese films, partly validating the Homi Bhabha (1994)’s postcolonialism.

The above analysis focused primarily on the different proportions of major China-related topics articulated in the cinema of the three regions and did so at a static level. Films favor different topics at different times, depending on their social contexts, meaning that the topics that are prevalent at any given point may change; thus, the topic of war is used as an example to introduce time to complete a dynamic comparative analysis.

Some war films are filled with nationalist emotions, stimulating anger toward aggressors, nostalgia for the homeland, and sympathy for the tragic destinies of victims. Some films about war raise crisis awareness, strengthen historical identity, and enhance social unity. Other war-related films construct fantastic images of heroes, showcasing their wisdom and courage. The three regions commonly give wars significance through historical consciousness and efforts, to transform China into a modern nation-state. According to the Western ethos, only countries with such consciousness may claim their rights in the international arena (Duara, 1995, p. 4). Hence, war is a topic suitable for discussion in all Chinese regions. War films are often based on historical facts and influenced by political, diplomatic, and other social characteristics. The distribution of this topic, as shown by the proportion of films dealing with it, was extracted by year. To smooth out extreme values, the moving average was calculated using the proportion values from the preceding and following year, and the values and curves were plotted in Fig. 4.

Figure 4 highlights the following phenomena relating to the topic of war in the three regions.

First, victories in war often lead to peaks in the number of war films in certain periods, such as in the 1950s in mainland China. China was able to end its long history of war and establish a new nation after its victory in the AJW in 1945 and the Chinese Civil War (CCW) in 1949. Filled with joy, the people praised their war heroes and their new life, and this pride and happiness were reflected in the films of that era. The proportion of films on this topic rose significantly and peaked at 33% in 1954, right after the Korean Armistice Agreement was signed of the Korean War (KW), which the PRC considered a victory. Similarly, the proportion of war films peaked in mainland China around

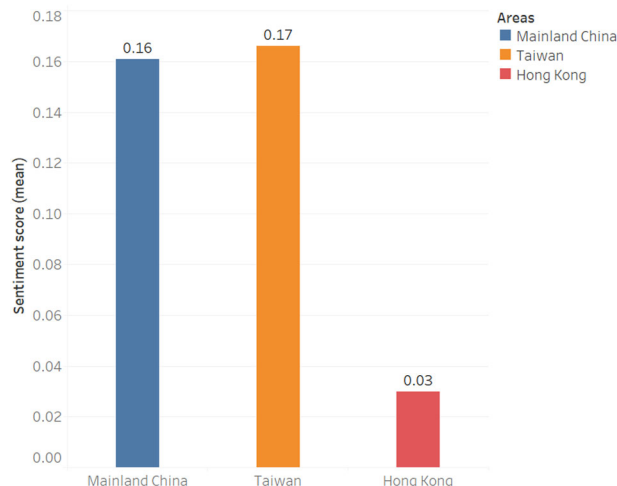


**Fig. 4 Proportions of films on the topic of war by year in the three regions.** The line chart reflects the changes in proportions of China-related films on the topic of war in mainland China (orange), Hong Kong (blue), and Taiwan (red) from 1950 to 2020.

1985 and 1995, the two years coinciding with the 40th and 50th anniversaries of the victory in the AJW. With the exception of a peak at nearly 24% in 2009, which met the 60th anniversary of the founding of the PRC, since the beginning of the 21st century, the proportion of war films in mainland China has mostly remained below 22%. During this period, the mainland’s economy was growing rapidly, and its films’ focus on peaceful development may have reduced the number of war films.

Second, international circumstances significantly influenced the extent to which this topic was highlighted in the films made in the three regions. War films often fulfill the political function of building nationalist feelings, which may explain the peak achieved by war films in the 1970s in Taiwan, when it confronted diplomatic turmoil. The United Nations ended Taiwan’s membership in 1971. In 1972, Japan severed diplomatic relations with Taiwan, triggering a series of severances involving 27 countries. Japan then became the main object of resentment for Taiwan (Li, 1997). War films reshaped historical memory at that time. The Taiwanese authorities hoped to relieve people’s disappointment and anxiety and rebuild national confidence through such films. The proportion of Taiwanese war films reached its highest level, about 28%, in 1978. In contrast, war films declined to a low level (about 20%) in mainland China in the late 1970s and early 1980s, a fact that could be associated with the establishment of diplomatic relations between the PRC, Japan, and the United States. At that time, mainland China needed to show an attitude of avoidance and restraint toward war.

Third, the war film curve of Taiwan reflects the change in its identity. Taiwan’s attitude toward Japan and the AJW reflects these temporal shifts. Lee Teng-hui, who was strongly pro-Japanese and pro-American, was elected Chairman of the Kuomintang (KMT, or Chinese Nationalist Party) in 1990. He visited the United States in 1995 as the leader of Taiwan and broadcast his support for Taiwan’s independence. From 1995 to 2000, the proportion of war films continued to decline, reaching a low of 13% in 2000. Taiwan’s sense of identity with China weakened during his tenure, and the trauma of being colonized subsided. Instead, postcolonial nostalgia was evoked, and a new culture with a pro-Japanese color gained prominence. War films fell out of favor in Taiwan as a result. However, the proportion of war films was relatively high after the turn of the 21st century, especially under KMT rule between 2008 and 2016 when it rose to more than 22%. In contrast, the number of war films fell when the



**Fig. 5 Average film sentiment scores of the three regions.** The bar chart compares the average film sentiment scores of China-related films in mainland China (blue), Hong Kong (red), and Taiwan (orange).

Democratic Progressive Party (DPP) came to power and advocated Taiwan’s independence (e.g., 2000–2008, 2016–present), perhaps because of the attitude toward the PRC of the administration concerned. At present, the KMT aspires to normalize relations with the PRC, while the DPP desires the opposite, calling for a distinctive national identity for Taiwan.

The people of Hong Kong demonstrate relatively little interest in war films compared with mainland China and Taiwan, perhaps because Hong Kong films have long been marketized and are less subject to political influences. Few war films have recently been produced in Hong Kong. In 2011, for example, the proportion was only 13%. The reason is presumably that the region’s war history is gradually receding and therefore young audiences have no memory of war.

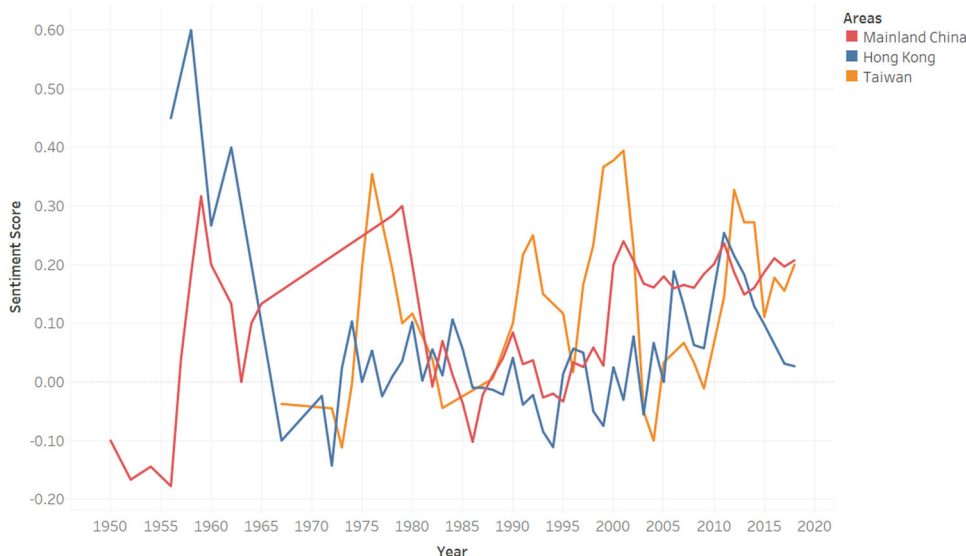
**Sentiments of China-related films.** The sentiment analysis of a text directly reflects the general positive and negative tones of films from the three regions. Figure 5 demonstrates that the general average sentiment scores of films made in the three



**Table 4 ANOVA and LSD tests on Sentiment scores in the three regions.**

	MC	HK	TW	ANOVA-F	MC vs. HK	MC vs. TW	HK vs. TW
Sentiment score	0.161	0.030	0.166	14.474***	0.131***	-0.005	-0.136***

MC represents the average sentiment score in mainland China; HK represents the average sentiment score in Hong Kong; TW represents the average sentiment score in Taiwan; ANOVA-F represents the F value in the ANOVA test; MC vs. HK represents the LSD test between MC and HK; MC vs. TW represents the LSD test between MC and TW; and HK vs. TW represents the LSD test between HK and TW.  
 \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ .



**Fig. 6 Film sentiment scores of the three regions by year.** The line chart reflects the changes in average film sentiment scores of China-related films in mainland China (red), Hong Kong (blue), and Taiwan (orange) from 1950 to 2020.

regions are all >0, implying that China is positively viewed in all three regions. Table 4 shows the results of the ANOVA and LSD tests on the sentiment scores in the three regions. No significant difference is found between the scores of mainland China and Taiwan at 0.16 and 0.17, respectively. Hong Kong has the lowest score (0.03), reflecting less positive sentiments. We calculated the moving averages of the scores of the three regions by year and drew curves to accurately reflect the changes in the sentiments of the films produced in the three regions (Fig. 6).

The sentiment scores calculated for the films from the mainland are intricately related to China’s political and economic circumstances. In the early days of the PRC, film sentiment in mainland China was very low, ranging from -0.2 to -0.1. Subsequently, the political movements from 1958 to 1978 prompted particularly high positive sentiments in mainland films. For example, the films of 1959 reached the highest sentiment score of 0.32, a level almost 1.8 standard deviations above the average value (0.10).<sup>7</sup> However, as political bubbles such as the Great Leap Forward and the Cultural Revolution burst by the 1980s, the moods depicted in the films also turned sour: the score was -0.01 in 1982 and -0.1 in 1986. Often dubbed “Scar culture,” many of the films of this period focused on tragic destinies and traumas, reflecting the distortion of the national spirit and the deterioration of interpersonal relationships. Since the late 1990s, with the return of Hong Kong and Macao and the economic boom, the sentiment score of films produced in mainland China has risen significantly, reaching 0.24, about 1 standard deviation above the mean, in 2001. This increase reveals the exploration of subjectivity by mainland films, which explains their optimistic and nationalist sentiments.

A complex and changeable cross-strait relationship has been forged by Taiwan and mainland China because of their historical

and political contexts. The early Minnan people who migrated to Taiwan, the Japanese colonial government, and the later Republic of China administration, regarded Taiwan merely as a small island on the edge of their territories, deeming it far less critical than the mainland (Berry and Lu, 2005, p. 4). Given the influence of different policies, such contrasts present diverse social and cultural states, as well as varying opinions and sentiments.

Taiwanese films produced during the period of martial law (1949–1987) were anti-Communist but not anti-China, exhibiting a “low–high–low” trend in their sentiment scores. In 1949, the KMT retreated to Taiwan and implemented a strict martial law policy. Films were used as ideological tools for political propaganda and were increasingly controlled by the KMT government. A substantial number of films made in Taiwan in the late 1960s portrayed hatred toward the mainland, and their corresponding sentiment scores were also very negative. For example, the scores from 1967 to 1974 were all lower than 0. In the mid-1970s, the focus of Taiwan-made films shifted from an anti-Communist to an anti-Japanese stance because of the diplomatic turmoil at the time. Resentment over the severance of diplomatic relations with Japan was expressed through the memory of the victory in the AJW, which created the growing sentiment toward China seen during this period, reaching 0.35 in 1976, about 1.5 standard deviations higher than Taiwan’s annual average score. The sentiment scores declined in the late 1970s, and their low values continued into the 1980s, falling to -0.04 in 1983. Taiwan’s film industry began to produce relatively negative films to construct a cultural orthodoxy in the context of the mainland’s Taiwan policy such as *Message to Compatriots in Taiwan* (1979).

The KMT government announced the end of martial law in 1987. The same year, the isolation of the two sides of the

strait ended. The sentiment scores rose rapidly immediately after the period of martial law (1987–1990), reaching a local peak of 0.25 in 1993, almost 1 standard deviation above the mean, reflecting both sides' desire to communicate. The Taiwanese could visit relatives on the mainland and could see the thus-far imagined China: the higher sentiment scores of Taiwanese films in the late 1980s to the early 1990s may partly be attributable to this ability. After martial law ended, cinema no longer had the power to construct its version of China's reality; rather, it instinctively revealed Taiwanese feelings.

Since the 1990s, the emotional changes noted in Taiwanese films have been intimately associated with the power politics of the region. Cross-strait relations fell to a freezing point in 1995 following Lee Teng-hui's remarks on Taiwan's independence. The 1995–1996 Taiwan Strait Crisis made the situation even tenser, which explains that the sentiment scores of Taiwanese films related to China touched a nadir of only 0.02 in 1996. However, after the crisis, the sentiment score of Taiwanese films rose sharply, holding at about 0.38 from 1999 to 2001. In 2000, the DPP came to power for the first time, and Chen Shui-bian was elected Taiwan's leader. He reiterated Taiwan's independence, as well as carrying out several de-China-lization policies such as constitutional reform, corresponding to the downward trend of the sentiment scores in that period. The result can be seen in the downward trend of film sentiment scores from 2001 to 2004, which reached a low of  $-0.1$  in 2004. Then, in 2008 the KMT candidate, Ma Ying-jeou, was elected Taiwan's leader. At the same time as he promoted a peaceful exchange between Taiwan and the mainland based on the 1992 Consensus, from 2009 the sentiment score rose, peaking at 0.33 in 2012. Tsai Ing-Wen (member of the DPP) has been in power since 2016 and has stoked hostility between the two sides of the strait. Consequently, the moods of the films show a downward trend again, even as the overall sentiment remains positive.

Hong Kong and the mainland have made different journeys and developed discrete social environments, creating separate emotional cognitions and cultural expressions. This separation is reflected in Hong Kong films, in which sentiment sometimes moves in the opposite direction to that of the mainland. Hong Kong's economy was booming in the 1960s and 1970s, while the opposite was true of the political and fiscal environment on the mainland. Some Hong Kong leftists launched an armed uprising against the British colonial government in 1967, but this drive gave rise to Hong Kong's aversion to nationalist ideology. It might cause the negative sentiments on China from 1967 to 1974, with sentiment scores below 0. The PRC and Great Britain issued a joint declaration in 1984, agreeing to the return of Hong Kong's sovereignty to the PRC in 1997. The significant changes they were facing caused anxiety among Hong Kong people. In the words of author Leung Ping-kwan (2012, p. 243), "what most people worried about in the 1980s was losing their existing way of life and the obliteration of the existing culture." This assertion may explain the emotional low points registered in Hong Kong films from 1987 to 1995, most of whose values were negative, with a low of  $-0.11$  in 1994. The increasingly frequent exchanges and growing cooperation with the mainland after the turn of the century led to progressively more positive feelings toward China in Hong Kong films. During this time the sentiment score rose, reaching 0.25 in 2011. However, the mood turned negative again, probably because of political protests like Occupy Central with Love and Peace, and in 2018 the sentiment score was just 0.03.

The above analysis to some extent showed that there are many factors that influence the China-related emotions articulated in the films of the three regions, although our aim was not to establish clear and robust causality. Among these factors, the relationship between the three regions is the most significant,

followed by the economic and social environments. In this sense, the image of China is not merely a cultural concept; it is also greatly influenced by politics. Economic and social factors may also determine the sentiments expressed in films. The amplification of national strengths in periods of economic optimism and the resulting relaxed social environment represent ideal conditions for the proliferation of the arts, which shift from their reliance on the West to focus on indigenous expressions with a positive mindset.

## Discussion

This study presented a panoramic image of China (in the West) in a century's films from mainland China, Taiwan, and Hong Kong SAR, from the perspective of postcolonialism. It did so by adopting machine learning methods (word embedding technology and the LDA topic model) to explore the transformations observed in the image of China in films produced in the three regions over the last 100 years. The discrete historical contexts of the three regions mean that they have constructed distinct societies even though all are rooted in Chinese culture. These social existences are articulated as divergent cultural characteristics in the films of the three regions, which convey different images of China to the international community. Their manifestations of China and Chinese culture have also transformed in tandem with the major historical events that the three regions have experienced over the last century.

The image of China is positive and realistic in films produced in mainland China, and their number is correlated with China's economic and political circumstances. For example, explosive growth may be observed in the number of films produced in China in periods of social stability and economic prosperity such as the 21st century. In addition, most contexts in which the word "China" appears in mainland films are related to the country's historical memory and future development. The sentiment scores of mainland films also correspond to the times. Greater optimism may be noted in films made during the Cultural Revolution, echoing political demands, and after the turn of the 21st century, as a natural expression of improved living standards and rising morale. Conversely, low moods were depicted in films made during the 1980s, a period of recovery from the traumas of the Cultural Revolution.

In Taiwan-made films, China's image is ambiguous and changeable, and the perspective is more individual than the standpoint of the mainland. The number of films produced in Taiwan fluctuates, but only a little, and much less than the number of mainland and Hong Kong films. Taiwanese films focus largely on personal topics, stories that reveal the changing times. In particular, Japanese and war elements proliferate in Taiwan-produced films, reflecting the persistent influence of Japanese colonization on Taiwanese society. In terms of sentiment, Taiwanese films are influenced by politics. Their sentiment scores related to China touched nadirs at important historical moments, such as the island losing its membership of the United Nations in 1972 and Lee Teng-hui's visit to the United States in 1995.

Hong Kong films showcase China through its traditions and international elements. Hong Kong was the first of the three regions to commercialize its films and achieve worldwide success. Consequently, the number of films produced in Hong Kong peaked between the 1970s and the 1990s. Subsequently, Hong Kong film production began to decline because of the boom in the mainland Chinese film industry and the increase in co-productions. Hong Kong-made films typically transcend reality, focusing on kung fu and crime, combining elements of traditional Chinese culture with Hollywood business models, and expressing the unyielding character of the Chinese people. Nevertheless, the

sentiment scores of Hong Kong films are affected by political events, and reveal public confusion about the erstwhile colony's identity and future after its return to China. Sentiment scores increased after the beginning of the current century but have dropped again because of the turmoil of recent years.

The results of our analysis of the content, topics, and sentiments of the films of the three regions show that political, economic, cultural, and ideological factors combine to influence China's image in the West. Cinema is a cultural product, and culture is both a practice and a system of symbols and meanings (Sewell, 2005, pp. 160–161). The reality is distorted by the consciousness of creators when films reflect society. In addition, truth and falsehood cannot be defined via simple binary distinctions: they are complex, multi-layered, and contextual systems influenced by national strengths and diplomacy. The vast increase in China's national strengths has made filmmakers in the three regions increasingly eager to hone their image of China and spread local awareness of change in the West through their films.

The results of the quantitative analyses also validate some of the theories outlined in the section "Background" of this paper, particularly postcolonialism. In the 1980s, mainland films experienced a period of emotional depression because of the historical introspection of their directors, but they also offered the West an opportunity to understand mainland China. Numerous Japanese elements persist in Taiwanese films, revealing the region's close relationship with its colonial history. Likewise, Hong Kong films demonstrate a cultural attachment to the suzerain state similar to that of Taiwanese films. To echo Bhabha (1994)'s concepts, Taiwan and Hong Kong represent regions where multiple cultures congregate, allowing intercultural communications to occur in a third space and construct a new intersubjectivity.

**Limitations.** Some limitations merit attention. First, it is necessary to briefly acknowledge the status quo of the three regions' film industries. Films produced in Taiwan and Hong Kong have recently distanced themselves from those made on the mainland. For example, no Chinese-language films, whether local or mainland productions, achieved a top 10 box office ranking in the two regions in 2019. Conversely, 8 of the top 10 mainland films were locally made. Some mainland films have succeeded locally but have failed to replicate their success in Taiwan, Hong Kong, and overseas markets, perhaps because of political factors or a lack of openness and inclusiveness. Second, classification-related difficulties meant that we were unable to evaluate the features of co-productions.

Nevertheless, in spite of the drawbacks, this study makes several contributions to the literature. First, its methodology contributes to cultural sociology. It directs cultural studies away from the analysis of micro and individual cases to large-scale cultural data. The joint use of three quantitative methods means that the results are more comprehensive, and the features of the three regions' cultures are more objective and accurate than conclusions drawn from single cultural phenomena. Thus, this study better reflects social reality at the macro level. Second, the findings of this study increase understanding of the relationship between the three regions, a relationship that changed dramatically in the second half of the past century. Knowledge of the similarities and idiosyncrasies in the respective cultures of the three regions is not only useful for a deeper understanding of the distinctive development process of Chinese culture, but also helps to clarify the regions' ethnic, cultural, and emotional connections, which go beyond politics. Finally, from the perspective of cultural sociology, the results of this study show that the image of China in the West as it is portrayed in movies is a combination of its

self-portrayal and selection by IMDb editors and users worldwide who all have their own tastes and ideologies.

### Data availability

The materials from this research are available on request from the corresponding authors.

Received: 15 November 2021; Accepted: 23 March 2022;

Published online: 12 April 2022

### Notes

- 1 We also compared the results obtained from IMDb with those from Douban (a cloned version of IMDb and the most popular online movie review website in China) and found that the number of Douban-related films after screening was 1219 from mainland China, 187 from Taiwan, and 358 from Hong Kong. This finding indirectly proves that there exists potential selection bias regarding films produced in Greater China (including Hong Kong SAR and Taiwan) by IMDb editors and users worldwide, who have their own tastes and ideologies.
- 2 The number of topics is a vital parameter for training LDA topic models. In this study, models with 5, 8, 10, and 15 topics were tested. After a comprehensive evaluation of each topic's connotation, topic differentiation, and the research purposes, five topics were finally chosen to train the model.
- 3 According to the LDA topic model, around 51.1% of the plot synopses of *Along Comes the Tiger* (1977) classify it to Topic 1, ranking first among 1047 films. Around 51.1% of the plot synopses of *Kung Fu Quest* (2010) categorize the film into this topic, suggesting that both movies are representative works of Topic 1.
- 4 According to the LDA topic model, the proportions of the three films belonging to this topic are 85.1%, 79.6%, and 78.6%, respectively, and all denote representative works.
- 5 According to the LDA topic model, the proportions of the two films belonging to this topic are 51% and 50%, respectively, meaning that both are representative works.
- 6 According to the LDA topic model, the proportions of the three films belonging to this topic are 44.2%, 42.4%, and 37.9%, respectively, and they are all representative works.
- 7 The means and standard deviations here are calculated based on the smoothed time series, which differs from the means calculated based on films.

### References

- Abidi SMR et al. (2020) Popularity prediction of movies: from statistical modeling to machine learning techniques. *Multimed Tools Appl* 79(47):35583–35617
- Berry C, Farquhar MA (2006) *China on screen: cinema and nation*. Columbia University Press.
- Berry C, Lu F (2005) *Island on the edge: Taiwan new cinema and after*. Hong Kong University Press.
- Bhabha HK (1990) *Nation and narration*. Routledge.
- Bhabha HK (1994) *The location of culture*. Routledge.
- Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. *J Mach Learn Res* 3:993–1022
- Boulding KE (1959) National images and international systems. *J Conflict Resolut* 3(2):120–131
- Chen F (2008) *History of Taiwanese film*. China Film Press, Beijing (in Chinese).
- Duara P (1995) *Rescuing history from the nation: questioning narratives of modern China*. University of Chicago Press.
- Erbring L, Goldenberg EN, Miller AH (1980) Front-page news and real-world cues: a new look at agenda-setting by the media. *Am J Political Sci* 24(1):16–49
- Leung P (2012) *Ten Essays on Hong Kong culture*. Zhejiang University Press, Hangzhou (in Chinese).
- Li D (2016) Subjectivity, integrity and concretization in the study of Chinese film history. *Lit Art Stud* 8:82–92. (in Chinese)
- Li X, Wang X (2009) Criticism of Chinese postcolonial film criticism. *Soc Sci Yunnan* 2:142–145. (in Chinese)
- Li T (1997) *Taiwanese films, society, and history*. Asiatic Books, Taipei (in Chinese).
- Lu SH (1997) *Transnational Chinese cinemas: identity, nationhood, gender*. University of Hawaii Press, Honolulu.
- McCombs M (2005) A look at agenda-setting: past, present and future. *Journal Stud* 6(4):543–557
- Mikolov T et al. (2013) Efficient estimation of word representations in vector space. ICLR (Workshop Poster), Arizona, USA
- 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
- Ridderstrom H (2018) Class concerns in a heritage setting: viewers' responses to downtown abbey on IMDb. *Series* 4(2):91–101
- Sewell Jr WH (2005) *Logics of history: social theory and social transformation*. University of Chicago Press.

- Sun S (2016) Chinese-language film or Chinese cinema? review of an ongoing debate in the Chinese mainland. *J Chin Cine* 10(1):61–66
- Tan Q (2020) ‘Bifurcated history’ and rewriting Chinese film history. *Film Art* 2:149–154. (in Chinese)
- Tripathy A, Agrawal A, Rath SK (2016) Classification of sentiment reviews using n-gram machine learning approach. *Expert Syst Appl* 57:117–126
- Yeh EYY (2018) Early film cultures in Taiwan, Hong Kong, and Republican China: kaleidoscopic histories. University of Michigan Press.
- Yoshimoto M (1991) The difficulty of being radical: the discipline of film studies and the postcolonial world order. *Boundary 2* 18(3):242–257

### Acknowledgements

This research has received funding from the Major Project of the National Social Science Fund of China (grant No. 19ZDA149) and the National Natural Science Foundation of China (grant No. 71921003).

### Author contributions

Conception/design of the work (YC and WG); analysis of data and interpretation of findings (ZC and WM); drafted and revised the work (all authors).

### Competing interests

The authors declare no competing interests.

### Ethical approval

Ethical assessment is not required prior to conducting the research reported in this paper, as the present study does not have experiments on human subjects and animals, and does not contain any sensitive and private information.

### Informed consent

This article does not contain any studies with human participants performed by any of the authors.

### Additional information

**Correspondence** and requests for materials should be addressed to Wei Guo or Yunsong Chen.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022