Sleeping Lion or Sick Man? Machine Learning Approaches to Deciphering Heterogeneous

Images of Chinese in North America

Dr. Qiang Fu, Associate Professor

Department of Sociology, The University of British Columbia

Vancouver, BC, V6T 1Z1, Canada, Email: qiang.fu@ubc.ca

Mr. Yufan Zhuang, PhD student

Department of Computer Science and Engineering, University of California, San Diego

La Jolla, CA, 92093, USA, Email: y5zhuang@ucsd.edu

Dr. Yushu Zhu, Assistant Professor

Urban Studies Program and School of Public Policy, Simon Fraser University

Vancouver, BC, V6B 5K3, Canada, Email: yushu_zhu@sfu.ca

Dr. Xin Guo, Senior Lecturer

School of Mathematics and Physics, The University of Queensland,

Brisbane, QLD, 4072, Australia,

and Assistant Professor

Department of Applied Mathematics, The Hong Kong Polytechnic University,

Hong Kong, China, Email: xin.guo@uq.edu.au

Acknowledgement

Introduction

For researchers who are keen to explore how people perceive others in a transnational setting, the many Western images of Chinese offer one of the most compelling and puzzling cases. After China launched its Open Door policy in the late 1970s, the mosaic of images of Chinese people, culture, and society in North America were particularly influenced by a wide variety of sociocultural factors, including the influx of Chinese immigrants, the development of multiculturalism, the history of Chinese railroad workers in North America, global security concerns, the growth of international trade, and the rise and fall of Confucian ethics. Research inquiries into the ever-changing and diverse Chinese images in North America greatly enhance our understanding of the dynamic interactions between people and place in a global context.

Yet while such research inquiries appear to be rich in theories, they are poor in methods. Despite significant efforts to theorize place-based sense, meanings, and identities, research has tended to focus on an egocentric or individualistic approach to characterizing one's own assessment and experience (Wang 1991, Knight 1982, Secor 2004, Tuan 1990), where social sentiments are often treated as a homogeneous construct either abstracted from topics of discussion or tied to a specific topic. If social sentiments toward others constitute a symbolic system from which one's present sense of place is derived (Bourdieu 1989: 5), scholars need to capitalize on the rich amount of media data and carefully examine the trends, patterns, and variability of the everyday expressions of others in public discourse.

Traditional regression or survey-based methods may be adequate for studying social sentiments in a given time period, in a specific location, or on a certain topic, but few methods are capable of investigating sentiments across time, on a broader geographical scale, or across

discussion topics. Moreover, these methods cannot provide a fine-grained temporal analysis, and their performance is subject to different choices of survey design and instruments. In recent years pioneering studies in geography and other social sciences have suggested a possible shift toward computational methods and big data analytics (e.g., Martin and Schuurman 2017, Graham, Hale and Gaffney 2014, Hu, Deng and Zhou 2019, DeLyser and Sui 2013, Kwan 2016). Based on over 280,000 newspaper articles published in North America from 1978 to 2019, this study proposes in detail an innovative synthesis of state-of-the-art tools in machine learning to explore heterogeneous topic-specific sentiments over time. In particular, we introduce the Bidirectional Encoder Representations from Transformers (BERT) for sentiment analysis. The BERT, which makes the processing of long sequences (e.g., newspaper articles) possible, has clear methodological advantages over lexicon-based or traditional neural-network methods for sentiment analysis. By challenging a simplified view of pro- or anti-Chinese sentiments in North America, this study stimulates a multifaceted understanding of social sentiments in a transnational setting, provides a unified computational framework for scholars to decipher complex temporal interactions in place-making, and calls for a paradigm shift from ontological or causal reductionism to understanding the intertwined social and discursive formation of spatiality.

Sentiment, Place, and Territory

A basic human instinct is to ascribe meaning to a place, even if that meaning is sometimes negative (Tuan 1990). In line with the premise that space does not in itself make place (Martin 2003, Harvey 1996), place, as envisioned by classic thinkers, is primarily defined by one's

position in society rather than the more abstract notion of spatial location (Tuan 1977, Sorokin 1943). The people, artifacts, and architecture of a place become meaningful when they are relevant to one's social life. If the place is primarily imagined, meanings derived from symbolic interactions with a space embody one's imagination of place (Gregory 1994). Among the different meanings that regulate our relations with place, sentiments refer to the more affective or emotional aspects, and can provide a powerful discursive force for place-making. For example, topophilia (love of place) characterizes the intense feeling that sports fans have for their home stadium, which influences the spatial distribution of these stadiums (Bale 1996, Tuan 1990). On the eve of the Romans' total destruction of Carthage during the third Punic War, Tuan described the feeling of a citizen toward Roman vengeance as "we can respond to this sentiment and yet feel disturbed when the fate of others, rather than of ourselves, is in question" (Tuan 1980: 464). Appealing to ideals (e.g., imagining what a livable and attractive neighborhood should be) is also found to foster place-based activism (Martin 2003).

However, the dyadic interplay between discursive and social processes is not unidirectional; rather, the two mutually inform each other. To understand this interplay, it is useful to bring social relations back in. Geographers understand territory as a bounded space consisting of a constellation of social relations among actors and between actors and space, in which the operation of power is manifested and communicated via meanings (Crevoisier 2004, Delaney 2008). Since individuals clearly sense the existence of a boundary when feeling a threat posed by others, what particularly makes sentiments territorial is their involvement in the discursive formation of ourselves and others (Delaney 2008, Tuan 1977). Depending on one's social relations with the dominant class, the approval and disapproval conveyed by sentiments suggest some are treated as insiders and the rest remain on the margins of society (Stock 1983,

Delaney 2008). In this regard, social sentiments exert control over space and represent a popular form of territoriality (Sack 1986). Meanwhile, when meanings are subject to multiple interpretations, the negotiation of social relations is reflected by a complex array of sentiments that assert or deny one's relevance to a bounded space (Lyman and Scott 1967). In other words, the spatialization of power is expressed in ways that are both more explicit (rules and laws) and less explicit (attitudes, beliefs, and values). Sentiments, especially those expressed by the media, make the latter instantaneously visible in everyday life. Compared with institutional arrangements, social sentiments provide a more subtle and timely way to reflect social relations and, in response to changing social relations, function as a potential harbinger of social changes.

Others in Public Discourse

The existing literature on (negative) attitudes toward immigrants reveals two primary mechanisms—material and ethnic/cultural—that shape the perception of others in a transnational context, where nation provides an overarching narrative in public discourse (Mughan and Paxton 2006, Hiers, Soehl and Wimmer 2017, Ceobanu and Escandell 2008, Semyonov, Raijman and Gorodzeisky 2006). Viewing themselves as the guardians of economic/material welfare or ethnic/cultural integrity, members of a dominant class employ material or ethnic rhetoric to frame the images of immigrants or foreigners. Realistic group conflict theory posits that increasing competition with locals for job opportunities, welfare support, education resources, and other material benefits leads to the rise of xenophobia in host countries (Meuleman, Davidov and Billiet 2009, Mughan and Paxton 2006). Negative sentiments toward immigrants respond to changing temporal contexts and tend to be more serious during economic downturns (Olzak

4

1994). For example, postwar Canadians appeared to show less support for immigration when the country's unemployment rates were high (Tienhaara 1974), although a later study found little evidence that Canadian attitudes toward immigrants were affected by the country's unemployment rates (Schissel, Wanner and Frideres 1989). A cross-country analysis suggests that in European countries anti-foreigner sentiment decreases with improved economic conditions but increases with more foreigners (Semyonov et al. 2006).

Collective imagination of what *our* own nation should be—East or West, traditional or modern, ethnic or civic—can also drive social sentiments toward immigrants/foreigners. Kohn's pioneering work distinguishes between civic and ethnic forms of nationalism (Kohn 1967). The Western view of nationalism emphasizes civil liberties, political inclusion, and voluntary participation in the exercise of state power. The boundary of a nation-state is permeable and independent of identification with a specific cultural or ethnic group (Ceobanu and Escandell 2008). In contrast, the Eastern view of nationalism pays meticulous attention to the ethnic nature of a nation-state. This view has an emotional resonance: when citizenship is defined by one's affiliation with a dominant ethnic group, negative sentiments toward immigrants are used to display natives' cultural supremacy over others (Hiers et al. 2017). Emotional defensiveness fueled by radical nationalism suggests that the very presence of outsiders in a nation-state is illegitimate simply because of their alienness. Yet outright antagonism toward immigrants, foreigners, or newcomers based on the Eastern view of nationalism is clearly at odds with modern societies that advocate for multiculturalism and inclusion (Mughan and Paxton 2006, Li 1998). Therefore, right-wing parties in democracies often express their ethnicity-based concern about immigrants as being grounded in cultural considerations, rather than making a direct claim that "you are not one of us." Similarly, right-wing politicians repeatedly blame immigrants for

their lack of effort to conform to local values or failure in social assimilation (Semyonov et al. 2006, Mughan and Paxton 2006).

Images of Chinese in North America: Realities, Myths, or Politics?

"Chinese" is used as an umbrella term to describe diverse groups of individuals living in areas central or peripheral to a unified but heterogeneous culture (Kuhn 1990, Hughes 2013, Li 1998, Tu 1991). According to Hughes (2013), "Chinese" can refer to either some or all of the following groups: 1) Han people living in the People's Republic of China (PRC); 2) ethnic minorities living within the PRC's jurisdiction; 3) Chinese compatriots in greater China, such as Taiwan, Hong Kong, and Macau; and 4) the Chinese diaspora and migrants living in other countries. The breadth of this definition implies that the sense of Chineseness is centered on one's subjective identification and cultural orientation rather than imposed by ethnicity (Liang 1949, Hughes 2013, Kuhn 1990). As indicated by Pye's observation that "China is a civilization pretending to be a state" (Pye 1990: 58), this complex cultural system has long developed distinctive approaches to the arts, philosophy, social relations, citizenship, bureaucracy, family life, and food (Kuhn 1990, Lin 2001, Fei 1946, Liang 1949, Bian 2019, Zhou 2021, Xie 2016). Chinese culture provides a key psychosocial anchor that confers security, a sense of belonging, and identity (Hughes 2013, Xiang, Yeoh and Toyota 2013, Liang 1949). Unlike Kohn's (1967) notions of civic and ethnic nationalism, it is argued that Confucianism, as the combination of "a

worldview, a social ethic, a political ideology, a scholarly tradition, and a way of life" (Tu 1998: 3), constitutes a shared basis of cultural nationalism (Guo 2004).

Given the crucial nexus between filial piety (e.g., the unconditional respect and care for parents) and unflinching loyalty to the emperor, migration, especially international migration, may appear to be at odds with traditional Chinese values (Hwang 1999, Liang 1949). When Chinese relocated as officials, soldiers, workers, or businessmen, they did so with the hope that they would return home rich, die on home soil, be buried in a family grave, and be worshipped by their descendants (Wang 1991, Li 1998). This transient sojourner mentality, which stressed a connection with rather than separation from home, was widely observed among the early generations of Chinese migrants to North America and possibly discouraged their integration in their host societies (Li 1998, Wang 1991, Kuhn 2009). After migrating to the United States, many continued to wear the Manchu tonsure (the long braided queue worn by Chinese males). However the queue made them a visible target and tended to invite harsh discrimination, it symbolized their strong desire to return to China and loyalty to the Qing empire (Metzger 2004). These early migrants did not share in the "American Dream" or "Canadian Dream"; rather, they had a "Chinese dream" even though imperial rulers were very skeptical of their loyalty towards the empire and often punished them upon their return (Kuhn 2009). Due to the barriers to social mobility and the lack of job opportunities in China, they had to seek these elsewhere in the hope of improving their living standards at home and achieving upward mobility upon their return. In traditional Chinese society, people were not simply attached to their families; rather, they were embedded in an extended web of familial obligations (Lin 1995, Bian 2019). Physical distance did not absolve one from fulfilling familial, kin, and ethnic obligations while abroad, but cooking in separate kitchens while living under the same roof might (see Chapter 1, Kuhn 2009).

Historical images of the Chinese, as perceived by Westerners, used to be exceedingly

negative, possibly reflecting a combination of colonialism, Orientalism, and white supremacy

(Ang 2001, Wang 1991, Li 1998, Stanley 2011). For example, colonial authorities in Manila and

Batavia had to rely on Chinese immigrants as the middleman minority to extract local wealth,

sustain their urban lifestyles, and trade via the corridors between China and Southeast Asia,

while often, albeit not always (Raffles 1830, Purcell 1965), perceiving the Chinese as culturally

alienated and ethnically threatening (Blussé 1986, Wickberg 1965). This dilemma, together with

pre-emptive strikes undertaken by Chinese immigrants against European rulers, led to massacres

of Chinese (Kuhn 2009). Nevertheless, perceptions of the Chinese in early modern Southeast

Asia also appeared to be diverse: In Siam and Malacca, social acceptance of Chinese immigrants

and settlers was largely driven by their active integration into the host societies rather than

cultural or ethnic issues (Purcell 1965, Kuhn 2009, Keyes 2019).

Without revenue-generating networks in the host society or active connections with their

home communities, Chinese immigrants to North America were largely perceived as racially

inferior, economically disadvantaged, and culturally undesirable; even if Chinese workers were

found to be useful to the economy, the "unmeltable" ethnic was perceived as needing to be

spatially and temporally contained (Anderson 1991, Li 1998, Kuhn 2009). The employment of

Chinese immigrants facing harsh working conditions was mostly confined to either labor-

intensive industries (e.g., railway construction, farming, and logging) or low-paying jobs in

Chinatowns (e.g., hostess, waitress, cashier, laundry worker). While the former made Chinese

immigrants the targets of overt hostility and fierce competition from European immigrant

labourers or the white working class (Kuhn 2009), "Chinatown" was depicted in public

narratives as congested, crime-ridden, and exotic (Anderson 1991). Chinese in the New World

were particularly haunted by one paradoxical form of racism: Chinese immigrants who endured the most exhausting, unpleasant, and dangerous work with inadequate subsistence and low wages were intentionally excluded from the process of building of a more civilized society due to their alleged lack of British civic values (Kuhn 2009, Hsu 2000, Lee 2003). In an effort to mitigate the influx of Chinese immigrants, institutional discrimination was striking. In Canada, under the Chinese Immigration Act, an exorbitant head tax was levied on Chinese immigrants from 1885 to 1923, but not on immigrants from anywhere else (Mawani 2004, Li 1998).

With the civil rights movement in the United States, the emergence of multiculturalism in Canada, and the rejection of racialized stereotypes and exclusion in North America, there has been a shift toward positive images of Chinese, who are described as hardworking, artistic, intelligent, and honest (Isaacs 1980). Since the focus of immigration policies in North America has shifted gradually from privileging white European immigrants to favoring skilled workers, attracting talents, and enabling family reunions, both the United States and Canada have become popular destinations for Chinese immigrants, and both countries have witnessed the growth of vibrant Chinese communities, especially in metropolitan areas such as New York, Toronto, Los Angeles, Vancouver, and San Francisco (Kuhn 2009, Li 1998, Fong and Shibuya 2005). Yet social sentiments toward the Chinese may still be informed by entrenched historical inertia. Chinese immigrants often become scapegoats for various social problems in host societies and are blamed for inflating housing prices, competing for school spaces, and blatant displays of wealth (Ley 2011, Li 1998).

Since China's first diplomatic contact with the West, the ways in which Western observers perceive the Chinese and China have increasingly become a controversial and politically charged issue. When the Macartney Mission (the first British mission to China)

arrived in China in 1793, Chinese officials from different levels and bureaucracies in the Qing Empire debated vigorously with members of the delegation on seemingly bizarre yet fundamental issues: whether the delegation had brought gifts as tribute to the emperor; whether Lord Macartney should address himself as the ambassador or a conveyor of tribute; whether members of the delegation should kowtow to the Qianlong Emperor; and, most critically, whether there was only one emperor under heaven, namely, the hitherto invincible, glorious, and benevolent Qianlong Emperor, who had reigned for more than half a century (Peyrefitte 2013). To deter any subtle or far-fetched ethnic challenge to the Manchu regime from the 1750s to the 1770s, Qianlong was determined to imprison, exile, or execute literati who suggested even the "faintest hint of disrespect" (Kuhn 2002: 9). For both the monarch and his circumspect officials, the Western images of the regime appeared to be too sensitive to ignore and too delicate to maneuver, especially if the "Middle Kingdom" (Rossabi 1983) traditionally viewed its expeditions to distant countries (such as the naval expeditions in the Ming dynasty) as a way to awe foreign powers, showcase cultural achievements, and expand its tributary system (Arrighi, Hui and Hung 2004, Kuhn 2009).

During China's tumultuous transition from the Qing dynasty to the Republican period in the early 1900s, two metaphors about Western images of the Chinese/China became popular. One was the "sick man of Asia," a phrase likely borrowed from the European metaphor of the Ottoman Empire as the "sick man of Europe," which suggested a pushover who was easily manipulated by foreign forces. The other was the "sleeping lion," suggesting a formidable power that no one wanted to offend (Huang 2011). Ironically, the introduction of both of these contradictory metaphors into Chinese literature can be traced back to one of the most influential thinkers of his time, Liang Qichao (Liang Ch'i-Ch'ao, 1873–1929), who maintained that the

10

consideration of national unity called for broader political participation and the arousal of nationalism among Chinese (Yang 2005, Kuhn 2002, Lee 2003).

A number of scholars, especially political scientists, view the selective portrayal of negative Western images of Chinese, along with China's media censorship, as part of a territorial strategy to foster nationalism and consolidate political control in China (Huang 2015, Kim 2018, King, Pan and Roberts 2017, Han 2015, Deibert et al. 2011). According to the relative deprivation theory (Brockmann et al. 2009), one's subjective feeling of happiness derives from a comparative appraisal of others' experience. In this vein, a survey experiment conducted in China suggests that a person's positive assessment of socioeconomic conditions in foreign countries leads to a negative domestic evaluation (Huang 2015). Huang (2015) further argues that social conditions in other countries, especially advanced democracies, become a critical yardstick for Chinese citizens to assess the performance of their own government. Thus, the selective portrayal of foreign countries' mistreatment and harassment of Chinese is a deliberate strategy to justify the status quo of the party–state and attenuate public criticism (Huang 2015, King et al. 2017, Han 2015).

When alternative sources of information and competing views become scarce or even unavailable due to media censorship, there is likely to be a feedback loop that perpetuates and amplifies negative coverage over time. State-run media (Walker and Orttung 2014), undercover pro-government Web commenters (King et al. 2017), and Web users who voluntarily defend the regime (Han 2015) in China are selective in their coverage of extreme negative stereotypes, biases, and behaviors against Chinese in Western democracies, thereby helping to enhance public awareness of an external foe regardless of whether the collective threat is realistic, hypothetical, or imaginary. When more readers begin to hold an antagonistic view and favor negative

coverage of foreign countries, there is in turn a stronger demand for negatively biased media reports (Kim 2018).

Our discussion above clearly suggests that researchers from different schools of thought tend to focus on negative Western images of Chinese. Next, we use big data analytics to examine whether this focus can be substantiated.

Data

Since this research covers a time span of more than forty years, newspaper coverage of Chinese probably presents the most viable data source for text analysis. As illustrated by DiMaggio, Nag and Blei (2013), newspapers are preferred over other media to study social sentiments because newspaper articles shape the knowledge base of the reading public, represent a major source of information over time, provide focal points for casual conversation, reflect the views of opinion leaders, and disseminate information through other media channels. While the long period of study means that newspapers provide the only source of digitized texts, we further illustrate why social media (e.g., Twitter and Facebook) cannot be an adequate source to explore heterogeneous images of Chinese in North America, or even place-based discourse, despite the fact that it is methodologically easier and computationally cheaper to analyze short social media texts.

In their review of advances in computational social sciences, Lazer and Radford (2017) note that "Twitter has become to social media scholars what the fruit fly is to biologists" (page 29), but this practice of measuring what is easily measurable rather than what we intend to measure is subject to serious methodological pitfalls (Tufekci 2014, Lazer and Radford 2017).

One fundamental (ideal-user) assumption underlying the use of social media data is that they are created by true human beings who faithfully express their opinions and attitudes (Lazer and Radford 2017). Unfortunately, this assumption is highly questionable when we explore images of Chinese in North America. First, it is well known that many social media posts are created by bots instead of real people (Ferrara et al. 2016), and one person may possess several accounts to conceal his/her true identity or opinions (Lazer and Radford 2017). As illustrated by Graham et al. (2014), human geographers should be particularly cautious in the use of Twitter data because a user's profile location is often different from the physical location from which one tweets. Second, social media can be easily manipulated by organizations to assert their interests and influence public opinion (Lazer and Radford 2017, King et al. 2017). In recent years political scientists and big data researchers have vigorously investigated Beijing's deliberate foreign propaganda strategy (the Global Propaganda Campaign, or *Da Waixuan*) and the regime's massive army of pro-government Web commenters (also known as the 50 Cent Army), which annually fabricates millions of social media posts as if they were genuine opinions expressed by ordinary people in cyberspace (Bolsover and Howard 2019, King et al. 2017, Brady 2015, Deibert et al. 2011).

The text corpus used in this study was obtained as follows. We first established the following criteria to guide our selection of mainstream newspapers in North America for text analysis: 1) the newspapers should have national coverage in the United States or Canada; 2) the newspapers should have the widest circulation in their respective countries; 3) they should be English-language newspapers; 4) they should be general in their coverage rather than focused on specific topics (e.g., business and finance); and 5) their circulation should be comparable to each other in size. For the selected newspapers, every article published from 1 January 1977 to 31

13

May 2019 was retrieved as long as it contained the word "Chinese." However, we noticed that there was limited newspaper coverage of Chinese prior to the beginning of China's Open Door policy in 1978. As a result, the reference period of study was from 1 January 1978 to 31 May 2019. Because the public in North America may not distinguish between different types of Chinese (e.g., Chinese in Canada and Chinese-Canadians) (Li 1998), retrieving all articles containing the word "Chinese" allowed us to explore the full breadth of social sentiments in different domains. The text corpus contained 169,084 American newspaper articles (104,007, 54,549, and 10,528 articles published in *the New York Times*, *the Washington Post*, and *USA Today*, respectively) and 119,480 Canadian newspaper articles (52,317, 43,529, and 23,634 articles published in *The Globe and Mail*, *Toronto Star* and *National Post*, respectively). This procedure generated a corpus of 288,564 documents in total. From 2017 to 2019 members of our research team manually retrieved these newspaper articles from electronic databases.

To achieve a compact representation of the corpus, we removed common stop words such as "and," "but," and "the." The RAKE (Rapid Automatic Keyword Extraction) algorithm was applied to extract meaningful phrases and combine words into phrases (Rose et al. 2010). For example, adjacent words such as "united" and "states" were combined into "united states." Based on the results from preliminary topic-modeling analysis using Latent Semantic Analysis (Landauer, Foltz and Laham 1998) and Latent Dirichlet Allocation (Blei, Ng and Jordan 2003), we performed multiple rounds of data cleaning to remove additional stop words and meaningless words (e.g., reporters' names, physical addresses). The text corpus is represented using a document-word matrix in which each column of the matrix corresponds to a document and each row of the matrix corresponds to a word.

Methods

Topic Modeling via the Hierarchical Dirichlet Process

We used topic modeling to process the whole text corpus and extract all discussion topics related to Chinese. According to our discussion above, "Chinese," just like "nation" and "ethnicity," is a "thick" concept (Coppedge 1999), which is complex in nature and subject to individual interpretation. Moreover, human geographers find that a thick concept is often socially constructed and reflects power relations at work, which corresponds to the identity politics of naming places (Berg and Kearns 1996). Cyr (2016) suggests focus groups as an effective tool to reveal how the public perceives or understands different notions of a thick concept. However, conclusions based on a small-scale qualitative method are highly context-specific and tend to vary with the size of a focus group, the composition of participants, and the specific time–space location in which a focus group is conducted. In contrast, by processing large volumes of texts from different sources and locations and over a long timespan, discussion topics automatically generated by topic modeling provide a more comprehensive characterization of a thick concept and are not vulnerable to subjective information offered by a few interviewees. Moreover, topic modeling is able to identify key topics among all discussion topics.

We compared results from different methods of topic modeling and chose a nonparametric Bayesian method, the hierarchical Dirichlet process (HDP) (Teh et al. 2006). The HDP method is an extension of the classical latent Dirichlet allocation (LDA), whose statistical generative process is described as follows (Blei et al. 2003). With a number of optimal topics K specified by researchers, the distribution of K topics in the jth document follows a multinomial

15

distribution $M(\theta_j)$ where the parameter θ_j is drawn from a Dirichlet distribution $D(\alpha)$. Here the

hyper-parameter α is a K-dimensional vector whose elements denote the weights/importance of

all the K topics in the corpus and are summed to 1. Meanwhile, the distribution of all V words in

the kth topic follows another multinomial distribution $M(\varphi_k)$, where the parameter φ_k is drawn

from a second Dirichlet distribution $D(\beta)$. The hyper-parameter β is a V-dimensional vector

whose elements denote the weights of V words in any topic. A word $x_{i,j}$ in the ith location of

the jth document is generated by a mixture process: the word's corresponding topic $t_{i,j}$ is first

generated from a multinomial distribution $M(\theta_j)$ and $x_{i,j}$ is then generated from a multinomial

distribution $M(\varphi_{t_{i,j}})$. If we use L_i and N to denote the word length of document j and the total

number of documents, respectively, the total probability of the LDA model is

$$P(X, t, \theta, \varphi; \alpha, \beta) = \prod_{k=1}^{K} P(\varphi_k; \alpha) \prod_{j=1}^{N} P(\theta_j; \alpha) \prod_{i=1}^{L_j} P(t_{i,j} | \theta_j) P(x_{i,j} | \varphi_{t_{i,j}}) \ . \tag{1}$$

In equation (1), only $x_{i,j}$ is observed and all other variables are latent variables.

The HDP as a nonparametric method uses a third Dirichlet process to take into account

uncertainties in the number of optimal topics K (Teh et al. 2006). To use the HDP, researchers

only need to specify an upper limit on K instead of K itself so that all integers lower than the

upper limit will be considered in the search for the optimal number of topics. We chose 150 as

K's upper limit; this upper limit has been used in many previous studies (e.g., Wang, Paisley and

Blei 2011). We compared the results from the HDP with those from six alternative methods (i.e.,

latent semantic analysis, principal component analysis, factor analysis, nonnegative matrix

factorization, the neural autoregressive model, and the neural variational document model), and

used held-out likelihood and several coherence measures to assess the performance of different models (Fu et al. 2021). The auxiliary analysis suggested that the HDP produced the most meaningful and coherent topics over time.

For each article, the HDP calculates its likelihood (posterior probabilities summed to 1) of belonging to these 150 discussion topics. We assigned this article to one discussion topic with the largest posterior probability. Figure 1 shows the distribution of these articles' topic-assignment probabilities (i.e., the largest probabilities we used to determine the discussion topics of newspaper articles). Since most of the articles had a topic assignment with a probability greater than 0.5, which is much higher than the random assignment probability (1/150), the majority of these articles (94.3 percent) were assigned to their main discussion topics. Our subsequent analysis was restricted to those articles assigned to their main discussion topics (i.e., topic-assignment probabilities > 0.5).

[Figure 1 about here]

Sentiment Analysis using Bidirectional Encoder Representations from Transformers

The shift from modeling short social media texts to long newspaper texts poses a methodological challenge that calls for a groundbreaking method for sentiment analysis. Previous tools such as recurrent neural networks (RNNs) cannot adequately handle long sequences (texts). RNNs tend to "forget" information processed earlier, even for sentences appearing in the same paragraph. In such cases, the relevant information may be one or a few sentences away, and may thus provide little information that helps in understanding the focal sentence (Liu et al. 2017). If only short sequences such as Twitter texts can be adequately analyzed, the sentiment of long texts such as

17

newspaper articles can only be modelled as a simple average of sentiment scores from separate sentences. As a result, an article's overall sentiment score might not be informed by the context of a sentence, the order of the sentences, or the organization of paragraphs.

By being able to integrate multiple novel deep-learning techniques including word embedding (Levy and Goldberg 2014), bidirectional language models (Peters et al. 2018), and the Transformers (Vaswani et al. 2017), the BERT (Bidirectional Encoder Representations from Transformers) is widely viewed as a milestone in natural language processing and also marks a new era for sentiment analysis (Clark et al. 2019, Devlin et al. 2018). This deep-learning method sets a record of achieving 94.9 percent accuracy on the Stanford Sentiment Treebank dataset (Devlin et al. 2018). It performs well with long sequences such as newspaper articles and has clear advantages over the lexicon-based methods adopted by several pioneering studies (Flores 2017, Hu et al. 2019, Zou et al. 2018).

The BERT is capable of handling long sequences. By connecting hidden layers of opposite directions to capture the underlying meaning of sentences (Peters et al. 2018), bidirectional language models draw on information of past and future contexts to better understand a text. Moreover, due to the presence of a large number of Transformers, the BERT as an effective tool for handling long sequences is able to process an entire article at once. By combining Feed-forward Neural Networks and attention models, the Transformers not only learn the long-term dependencies of words in sentences but also prioritize different semantic features of words based on the specific queries/questions at stake (Vaswani et al. 2017).

In contrast, existing lexicon-based methods are highly manual and make several strong assumptions about text analysis. First, they assume that all nouns are entities. However, as linguists have noted, the fact that all conceptual entities are always represented by nouns does

not necessarily mean that all nouns are entities (Black and Chiat 2003). Second, the sentiment expressed by a word is constant across documents. Third, the effect of one word (e.g., an adjective) on another (e.g., a noun) decreases substantially with their distance within a sentence. Fourth, sentiments expressed by different words and sentences add up linearly. To overcome these methodological issues, the BERT does not focus on definite meanings assigned to specific words. Instead, it provides a contextualized understanding of words, phrases, and sentences by capturing subtle semantic changes and complex linguistic features. For the example sentence "Chinese paid a head tax to enter and work in Canada" (Li 1998, Mawani 2004), a lexicon-based method may completely fail to detect its negative meaning and even indicate a positive sentiment due to the presence of "paid," "enter," and "work." However, the BERT readily captures the sentence's highly negative meaning.

We fine-tuned the BERT on a binary sentiment classification dataset from the Internet Movie Database (IMDb) with an additional binary classification layer (Maas et al. 2011). This dataset consists of 50,000 reviews dichotomized as either positive or negative. Since the original IMDb rating scale ranges from one to ten, Maas et al. (2011) labeled any review with a rating less than four as negative, and any review with a rating higher than seven as positive. Given the binary nature of these labels, they only suggest whether a review is positive or negative, not how positive or negative it is. Scholars find that the BERT and related sentiment-analysis models trained on the IMDb generalize well to other datasets and contexts (Xie et al. 2019, Kayal, Singh and Goyal 2020). When Kayal et al. (2020) generalized the BERT model fine-tuned on the IMDb to sentiment analysis based on various Amazon, Yelp, weather, and scientific reviews, their corresponding F_1 scores, defined as the harmonic mean of the precision and recall (sensitivity), varied from 80.5 percent to 92.5 percent (the average is 86.5 percent).

According to the binary nature of labels and our model specification (as described below), the value of a sentiment score should not be interpreted as how negative/positive a document is; rather, it represents how confident we are that a document reflects a positive or negative sentiment. A higher absolute value means higher confidence in the prediction of positive or negative sentiments. More specifically, the BERT was trained to minimize the binary cross-entropy/log loss and generate a sentiment score of each article:

$$L(y, A) = -\frac{1}{N}\sum_{i=1}^{N}(y_i \cdot \log(A_i) + (1 - y_i) \cdot \log(1 - A_i)) \ , \tag{2}$$

where y_i corresponds to positive ($y_i = 1$) or negative labels ($y_i = 0$) obtained from the IMDb dataset, and A_i is a predicted probability from the BERT. Based on the definition of the loss function, the binary cross-entropy loss is minimized when the distribution of the predicted probabilities is closest to the distribution of labels: if y_i is 1 and A_i is also close to 1, we are more confident that a document reflects a positive sentiment; if y_i is 0 and A_i is also close to 0, we are more confident that a document reflects a negative sentiment. Given the definition of the binary cross-entropy/log loss, any wrong prediction (e.g., y_i is 1 but A_i is close to 0 or 0.5) is severely penalized. In other words, the relationship between y_i and A_i is similar to that between a binary outcome y_i^L and a continuous variable p_i^L, indicating the probability of y_i^L to be 1 in a logistic regression model (Long and Freese 2006). The observed binary outcome y_i^L has a value of either success (e.g., positive label) or failure (e.g., a negative label), while the unobserved latent probability p_i^L is being fitted in a logistic regression model. Depending on whether a positive or

negative sentiment is at stake, we rescaled A_i to a sentiment score ranging from -1 (a negative sentiment) to 1 (a positive sentiment).

The distribution of sentiment scores has an average positive value of 0.485, but most of the sentiment scores are very close to 1 or -1, which suggests that the BERT performs well on the text corpus. In particular, articles with sentiment scores larger than 0.95 or smaller than -0.95 accounted for 94.7 percent of the corpus. This bimodal distribution suggests a very high confidence that the majority (94.7 percent) of these articles reflect either positive or negative sentiments. A sample of labelled documents is provided in the supplemental materials. We used Gensim, a Python library for topic modeling, to implement the HDP and TensorFlow to implement the BERT (Devlin et al. 2018).

Results

To specify distinct discussion topics in a country, we performed the HDP separately on the American and Canadian text corpuses. The main discussion topics specified by the HDP are presented in Table 1. While results from the HDP suggested that the optimal number of discussion topics was two for both text corpora, we present the results pertaining to the first eight discussion topics, ranked by the Dirichlet hyper-parameter α, or each topic's relative weight in the overall topics. According to the meaning of the top keywords associated with a specific topic, we determined that the top eight discussion topics in American newspapers correspond to the following themes (each topic's corresponding value of α is noted in parentheses): culture (17.2 percent), economy and trade (16.6 percent), international security (10.9 percent), Chinese politics

21

(10.9 percent), life (7.6 percent), sports (5.9 percent), American politics (5.8 percent), and food (4.5 percent). In terms of their relative importance, these top eight topics are a major component (79.4 percent) of all 150 topics specified by the HDP. For the Canadian corpus the top eight discussion topics focus on the following themes: economy and trade (16.9 percent), life (10.8 percent), Canadian society and politics (8.8 percent), international politics (8.5 percent), food (8.2 percent), culture (4.8 percent), civil affairs (4.3 percent), and international security (3.9 percent). In terms of their relative importance, these eight topics account for 66.4 percent of all 150 discussion topics of the Canadian text corpus.

[Table 1 about here]

Our subsequent analysis focused on five out of the eight discussion topics: culture, economy and trade, international security, Chinese (or international) politics, and life. We chose to focus on these five topics for two reasons. First, they are present in both American and Canadian corpora, which permits a direct comparison of images of Chinese in these two countries. Second, as suggested by Table 1, culture, economy and trade, international security, Chinese politics, and life are the top five topics in the American corpus (as ranked by their relative importance α) and are central to social, economic, cultural, and political aspects of Chinese images in Western societies.

The annual distribution of the number of American newspaper articles assigned to these five topics (with topic-assignment probabilities > 0.5) is shown in Figure 2.1. News coverage on Chinese increased since the late 1970s. Several spikes in the distribution deserve special attention. First, around the June Fourth (Tiananmen Square) Incident in 1989, Chinese politics

22

immediately became the most pervasive topic; there were more articles on this subject than on any other topic. A closer investigation of articles assigned to the topic of Chinese politics shows that the average word frequency of "Tiananmen" (per document) in 1989 was 14.5 times of that in other years over the period of study. Second, international security was the primary concern when the Hainan Island incident occurred in 2001 and one of the recent North Korea crises occurred in 2017. The average word frequencies of "North Korea" and "nuclear" in articles assigned to the topic of international security were 32.3 and 15.3 times higher, respectively, than in other years. In contrast, the number of articles on culture, economy and trade, and life steadily increased over the last four decades and showed less variation.

The temporal pattern of Canadian newspaper coverage of Chinese is different from that in the United States (see Figure 2.2). Since the late 1970s there have been more articles on all topics except for international politics. Despite two spikes in the number of articles on international politics when the June Fourth Incident in 1989 and the handover of Hong Kong in 1997 happened, more recent reports containing the word "Chinese" tend not to be situated in the context of international politics. There has been a dramatic increase in articles on the economy and trade since China's entry into the World Trade Organization (WTO) in the early 2000s. This topic attracted attention particularly when a formal "strategic partnership" between Canada and China was announced during Chinese President Hu Jintao's visit to Canada in 2005, and again around the time of the 2015 Canadian federal election.

[Figure 2 about here]

As indicated by overall positive sentiment scores produced by the BERT (see Figure 3 with three-point moving average smoothing), newspaper coverage related to Chinese was generally favorable. The temporal trends in both countries appear to be stable and similar to each other over the period of study (the Pearson's correlation coefficient is 0.436). However, the aggregate analysis concealed substantial topic-specific heterogeneity in social sentiments. Based on partial analysis of variance (see Table 2), the partial sum of squares explained by discussion topics is 2968.7, which accounts for more than 75 percent of the total variance explained by the model. In other words, the discussion topics specified in Table 1 explain much more variance than any other factors (e.g., year of publication, country of publisher, length of the article, and newspaper publisher). After adjusting for degrees of freedom, the country of the publisher is associated with the highest mean of squares explained (443.7).

[Figure 3 and Table 2 about here]

Figures 4.1 and 4.2 illustrate topic- and country-specific variations in sentiments with three-point moving average smoothing. For the U.S. corpus (see Figure 4.1), sentiment scores on culture are more positive than on other topics. Sentiments on Chinese politics and life are generally more positive than are the sentiments on international security and the economy and trade. With Deng Xiaoping's active engagement with the world system immediately after the Cultural Revolution (1966–76), Americans were exposed to a more positive media portrayal of all five topics relevant to Chinese. The continuous improvement of media sentiments reached its peak around the year 1984, when the Sino-British Joint Declaration was signed in Beijing. Except for sentiments on culture, newspaper sentiments on other topics began to worsen until the

24

early 1990s, a period when there was more uncertainty about China's reforms. The worsening of sentiments is particularly striking for topics on international security and on the economy and trade. On the topic of Chinese politics, one noteworthy observation is that a massive crackdown on dissidents in China (e.g., the June Fourth Incident) did not necessarily lead to a sharp decline in sentiment scores. Our closer investigation of opinions expressed by these observers in North America reveals that their harsh criticism of authoritarian rule was also accompanied by support for citizen participation, admiration for courageous resistance, and hope for progressive social changes in China. A gradual improvement (with fluctuations) in sentiments is observed in more recent years. In and after the 1990s, more drastic changes in sentiments on international security coincided with the handover of Hong Kong in 1997, the Beijing Olympics (and Tibetan unrest) in 2008, and the 2017 North Korea nuclear crisis. For sentiments on the economy and trade, the opposite changes are observed during China's entry into the WTO in the early 2000s and the U.S.-China trade war.

While the trend and levels of Canadian sentiments across different topics share some similarities with the U.S. case, there are several differences. First, sentiments on culture are not as positive in Canada as they are in the United States. Second, Canadian sentiments in general exhibit more temporal variations than their American counterparts. Third, sentiments in the two countries may react differently in a specific time period. For example, Canadian sentiments on the economy and trade, international security, and international politics exhibited a sharp decline starting in the mid-1980s and then a recovery after 1989, when the June Fourth Incident took place. The decrease in sentiments on related topics was more sustained in the United States over the same period.

Conclusion

Conventional studies on place-based discourse face several methodological and epistemological questions. First, is there a standard tool that consistently measures discursive processes across people and place? Second, how can we effectively process a huge amount of digitized data and track discourse over time? Third, should discourse be understood as a homogeneous construct or a heterogeneous mixture? To address these issues, this study proposes a unified computational framework for the analysis of media sentiments in a transnational setting. We find that discussion topics explain much more variation in sentiments than does any other factor. There are distinct trajectories of sentiments across different topics. Sentiments on culture appear to be more positive and stable over the period of study, whereas sentiments on international security, the economy and trade, and Chinese (international) politics show more temporal variations. In a specific time period, sentiments on the same topic may exhibit different patterns in Canada and the United States. By combining state-of-the-art tools for topic modeling and sentiment analysis, we find that the Western images of Chinese can hardly be reduced to a "sleeping lion" or a "sick man." We also used an alternative sentiment-neuron-based model for sentiment analysis (Radford, Jozefowicz and Sutskever 2017). Although sentiment scores generated by this alternative method are slightly lower than those produced by the BERT, we still do not observe an overall anti-Chinese sentiment, and the trajectories of sentiment scores remain highly topic-specific. These findings signify that media construction cannot exhaust other aspects of discursive and social formation of realities. The obsession with pro- or anti-Chinese

26

sentiments, especially the latter, may reflect one's inability or unwillingness to distinguish

between reality and observation, or an epistemic fallacy that our observations are always

overdetermined by theories and underdetermined by facts (Bhaskar [1975] 2013).

Place is both imagined through symbolic interactions and constructed via social relations

(Gregory 1994). Based on the premise that spatiality is the medium, product, presupposition, and

embodiment of social actions (Soja 1985), the duality of place suggests that our attention should

shift away from the question of "which causes which" to a more fine-grained analysis of the

intertwined interactions between social and discursive processes over time. If place is believed to

provide a legitimizing discourse for actions, we must articulate how the social reality of

everyday life is jointly constructed by one's social and imagined relations with a place.

This paradigm shift toward a synthesis of social and discursive formation of realities is

informed by three lines of reasoning. First, (shared) meanings are constrained by social

structures. According to phenomenologists such as Schutz (1970), shared meanings constructed

through social interactions shape how people interpret social realities in an intersubjective world.

The intersubjectivity argument further suggests that the actual content of a reality is of less

interest to social scientists. Rather, because the creation of social realities is "constrained by the

pre-existing social and cultural structures" (Ritzer and Stepnisky 2017: 76), we should attempt to

explain how a reality comes into being through a better understanding of social structures.

Second, social structures are embedded in and reproduced by actions (Bourdieu 1977, Collins

1981). While Collins (1981) does not deny the symbolic basis of social structures, he contends

that "patterns of repetitive micro-interaction" (Collins 1981: 985) constitute the empirical

realities of social structures, where the continuity of social structures relies on everlasting

behaviors repeated by individuals and structural changes are attributable to changes in

microbehaviors. Finally, actions are in turn informed by discursive processes. Bauman's inquiry into *modernity and ambivalence* suggests that the dichotomy between good and bad (evil) people cannot occur in any realm of social life except for the discursive realm (Bauman 1993). These dichotomous codes are schematic and act as a cultural anchoring for the reproduction of taken-for-granted narratives and the repetition of everyday microinteractions (Alexander 2013). The discursive formation of realities can be territorial: assigning dichotomous codes to various actors also allows for identification, categorization, and differentiation in a politicized space (Alexander 2013).

The interactions between social and discursive processes further suggest that our understanding of meaning, like the concepts of space and power, is best characterized as relational (Delaney 2008). Meaning is not independent of the relational realm of everyday lived experience in which it is always embedded (Alexander 2002). Rather, ambiguous or vague meanings attached to different subjects are settled through social interactions. The apparent gap between actual media portrayals and our common understanding of Western images of Chinese, as well as the topic-specific patterns of sentiments and the coevolution of history and sentiments, all provide important clues about how meanings are contested, suppressed, and advocated by interconnected actors for the politicization of space.
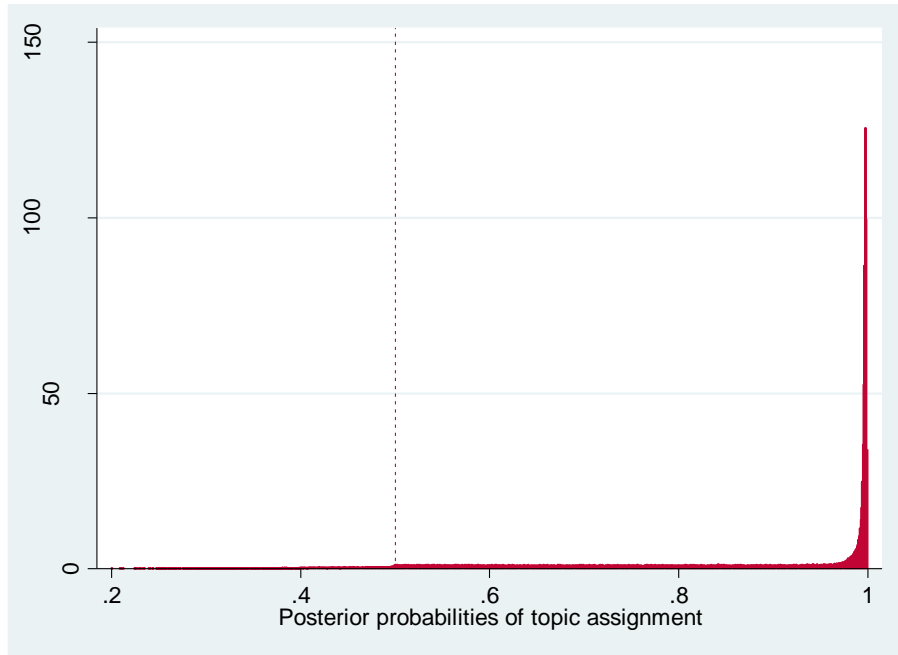
Figures and Tables

Figure 1 The distribution of articles' topic‑assignment probabilities based on the Hierarchical Dirichlet

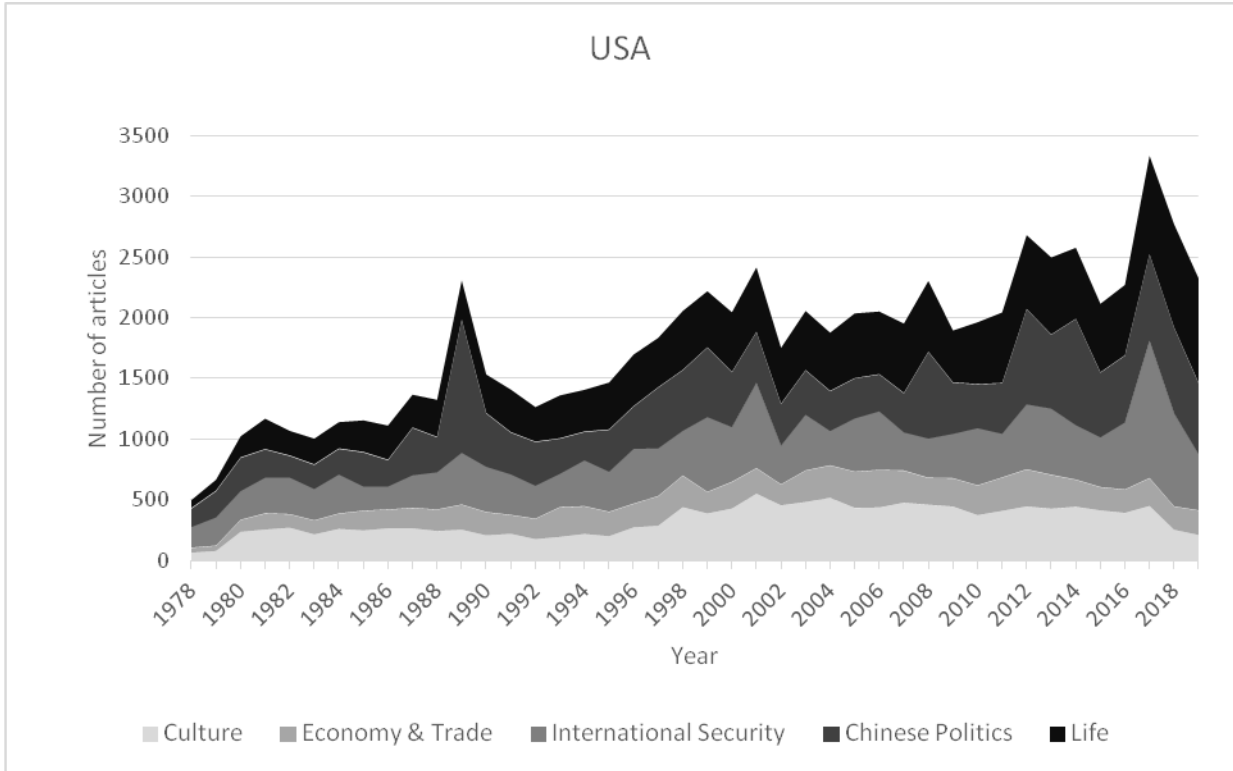**Process (the dashed line is drawn at 0.5)**

Figure 2.1 Number of American newspaper articles assigned to five key topics by the HDP, 1978–2019

Figure 2.2 Number of Canadian newspaper articles assigned to five key topics by the HDP, 1978–2019
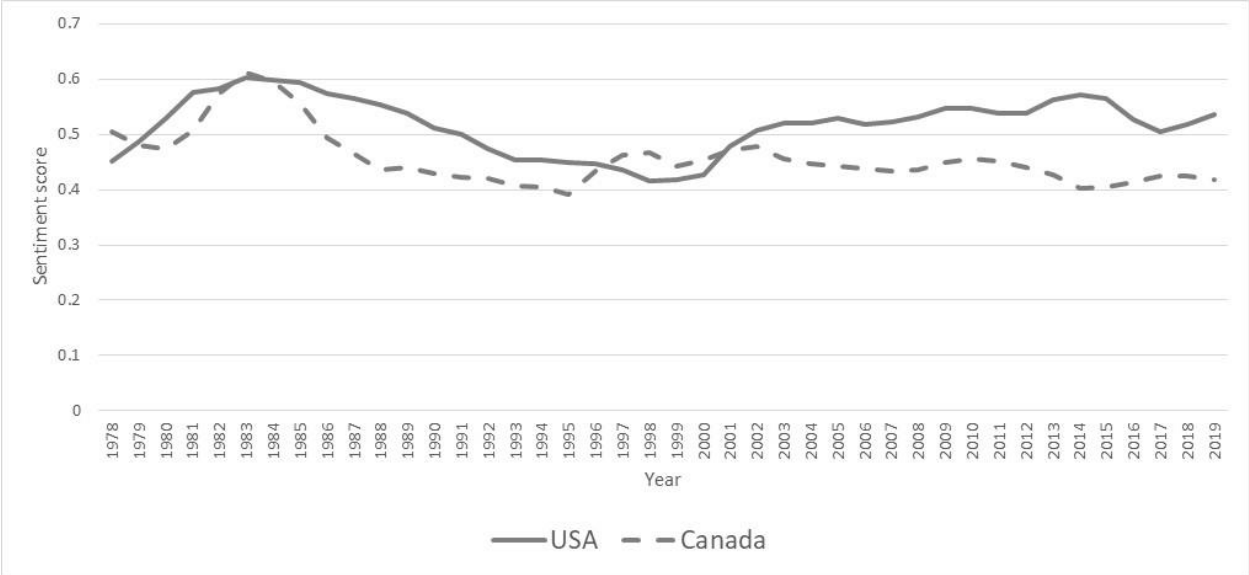
Figure 3 Overall trajectories of social sentiments related to Chinese produced by the BERT, 1978–2019
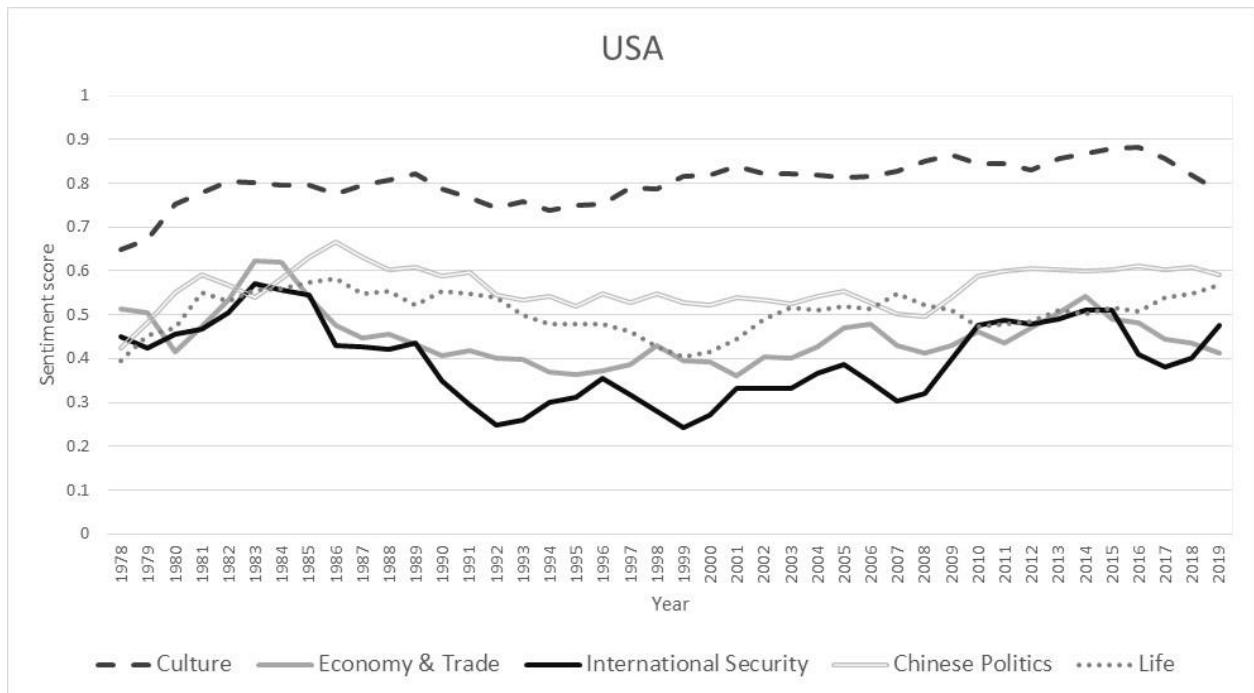
Figure 4.1 Topic-specific trajectories of American sentiments produced by the BERT, 1978–2019
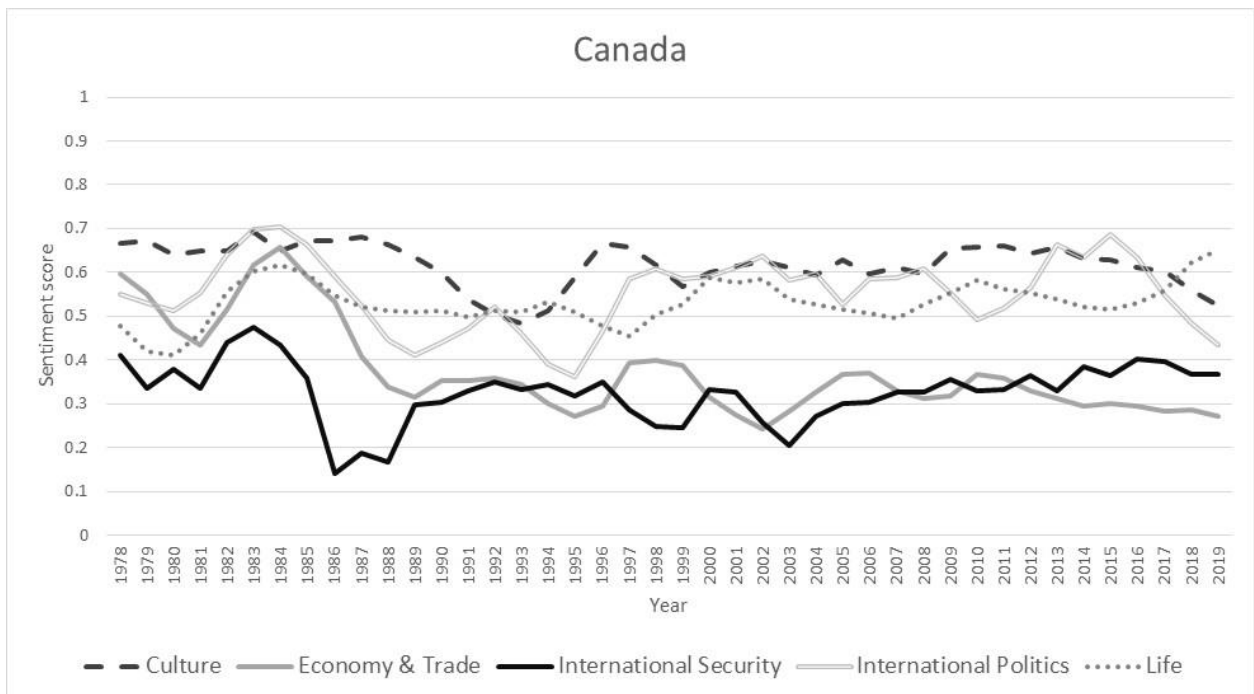

Figure 4.2 Topic-specific trajectories of Canadian sentiments produced by the BERT, 1978–2019

Table 1 Main discussion topics and their top keywords produced by the HDP

α	Topic	Top keywords ranked by loadings
	American Newspapers	
17.2%	Culture	art, museum, free, center, work, American, New York, city, park, house, gallery, open, artist, theater, music, child, painting, world, life, book, exhibition, collection, film, …
16.6%	Economy & Trade	company, American, government, market, country, million, business, school, city, world, United States, official, bank, state, billion, president, economy, price, economic, …
10.9%	International Security	United States, official, president, American, nuclear, country, military, administration, Taiwan, North Korea, government, Beijing, foreign, soviet, war, missile, weapon, …
10.9%	Chinese Politics	government, party, official, Beijing, leader, country, political, communist, police, group, student, Hong Kong, right, city, president, state, military, power, force, protest, …
7.6%	Life	film, book, life, world, story, woman, movie, American, work, family, child, New York, know, school, man, back, home, young, novel, love, good, director, character, city, …
5.9%	Sports	team, game, world, player, woman, American, Olympic, sport, medal, back, play, coach, second, home, point, men, know, United States, Olympics, city, national, four, win, …
5.8%	American Politics	official, government, president, American, Clinton, state, United States, right, security, company, case, law, department, national, republican, country, campaign, Trump, …
4.5%	Food	restaurant, food, dish, sauce, chicken, menu, chef, good, wine, dinner, room, New York, place, served, cooking, fish, shrimp, rice, bar, soup, table, pork, salad, beef, kitchen, …
	Canadian Newspapers	
16.9%	Economy & Trade	company, market, Canada, price, Canadian, million, bank, world, stock, oil, country, government, billion, economy, business, investment, growth, investor, share, trade, …
10.8%	Life	film, world, woman, Canadian, game, Canada, life, movie, team, story, play, back, know, good, work, book, second, best, man, American, love, family, director, music, player, …
8.8%	Canadian Society & Politics	Canada, Canadian, community, school, government, city, student, group, child, family, woman, immigrant, liberal, work, party, university, right, home, Ontario, minister, …
8.5%	International Politics	government, country, party, world, official, Canada, Hong Kong, leader, political, soviet, communist, war, Canadian, right, foreign, Beijing, minister, Taiwan, state, power, …
8.2%	Food	food, restaurant, sauce, cup, place, good, city, chicken, water, room, cooking, dish, hotel,

4.8%	Culture	add, home, oil, green, back, small, chef, fresh, rice, cook, white, wine, taste, hot, …
		art, work, city, gallery, museum, world, artist, home, house, book, life, Canadian, place,
		Canada, building, design, room, look, painting, collection, centre, good, history, wall, …
4.3%	Civil Affairs	police, Canada, family, home, man, Canadian, woman, official, case, child, city, back,
		government, court, officer, life, Hong Kong, country, men, know, death, immigration, …
3.9%	International Security	country, nuclear, war, president, United States, world, military, Beijing, government,
		official, North Korea, trade, Canada, minister, security, weapon, missile, Taiwan, …

Table 2 Factors affecting newspaper sentiments related to Chinese: Partial analysis of variance

Factor	Partial sum of squares	Degrees of freedom	Mean of squares	F-value
Year of publication	258.6	41	6.3	8.8***
Country of publisher	443.7	1	443.7	621.8***
Length of article [a]	74.9	51	1.5	2.1***
Newspaper publisher	138.0	4	34.5	48.4***
Discussion topic	2968.7	14	212.0	297.2***
Model	3953.4	111	35.6	49.9
Residual	120585.2	168981	0.7	
Total	124538.6	169092	0.7	

*** Significant at the 0.001 level.

[a] The length of the articles has been rounded to the nearest 100 words.

References

Alexander, J. C. (2002) On the social construction of moral universals: theHolocaust'from war crime to trauma drama. *European journal of social theory,* 5**,** 5-85.

---. 2013. *The dark side of modernity.* John Wiley & Sons.

Anderson, K. J. 1991. *Vancouver's Chinatown: racial discourse in Canada, 1875-1980.* Montreal, Canada: McGill-Queen's University Press.

Ang, I. 2001. *On not speaking Chinese: Living between Asia and the West.* London, UK: Routledge.

Arrighi, G., P.-k. Hui & H.-F. Hung. 2004. Historical capitalism, east and west. In *The resurgence of East Asia: 500, 150 and 50 year perspectives,* eds. G. Arrighi, T. Hamashita & M. Selden. London, UK: Routledge.

Bale, J. (1996) Space, place and body culture: Yi-Fu Tuan and a geography of sport. *Geografiska Annaler: Series B, Human Geography,* 78**,** 163-171.

Bauman, Z. 1993. *Modernity and Ambivalence.* Oxford, UK: Polity.

Berg, L. D. & R. A. Kearns (1996) Naming as norming:'race', gender, and the identity politics of naming places in Aotearoa/New Zealand. *Environment and Planning D: Society and Space,* 14**,** 99-122.

Bhaskar, R. [1975] 2013. *A realist theory of science.* New York, NY, USA: Routledge.

Bian, Y. 2019. *Guanxi, how China works.* Cambridge, UK: Polity Press.

Black, M. & S. Chiat (2003) Noun–verb dissociations: A multi-faceted phenomenon. *Journal of Neurolinguistics,* 16**,** 231-250.

Blei, D. M., A. Y. Ng & M. I. Jordan (2003) Latent dirichlet allocation. *Journal of machine learning research,* 3**,** 993-1022.

Blussé, L. 1986. *Strange company: Chinese settlers, mestizo women, and the Dutch in VOC Batavia.* Dordrecht, The Netherlands: Foris Publications.

Bolsover, G. & P. Howard (2019) Chinese computational propaganda: automation, algorithms and the manipulation of information about Chinese politics on Twitter and Weibo. *Information, communication & society,* 22**,** 2063-2080.

Bourdieu, P. 1977. *Outline of a Theory of Practice.* Cambridge, UK: Cambridge university press.

--- (1989) Social space and symbolic power. *Sociological theory,* 7**,** 14-25.

Brady, A.-M. (2015) Authoritarianism Goes Global (II): China's Foreign Propaganda Machine. *Journal of Democracy,* 26**,** 51-59.

Brockmann, H., J. Delhey, C. Welzel & H. Yuan (2009) The China puzzle: Falling happiness in a rising economy. *Journal of Happiness Studies,* 10**,** 387-405.

Ceobanu, A. M. & X. Escandell (2008) East is West? National feelings and anti-immigrant sentiment in Europe. *Social science research,* 37**,** 1147-1170.

Clark, K., U. Khandelwal, O. Levy & C. D. Manning (2019) What Does BERT Look At? An Analysis of BERT's Attention. *arXiv preprint arXiv:.04341.*

Collins, R. (1981) On the microfoundations of macrosociology. *American journal of sociology,* 86**,** 984-1014.

Coppedge, M. (1999) Thickening thin concepts and theories: combining large N and small in comparative politics. *Comparative Politics***,** 465-476.

Crevoisier, O. (2004) The innovative milieus approach: toward a territorialized understanding of the economy? *Economic geography,* 80**,** 367-379.

Cyr, J. (2016) The pitfalls and promise of focus groups as a data collection method. *Sociological methods & research,* 45**,** 231-259.

Deibert, R., J. Palfrey, R. Rohozinski & J. Zittrain. 2011. *Access contested: security, identity, and resistance in Asian cyberspace*. Cambridge, MA: MIT Press.

Delaney, D. 2008. *Territory: a short introduction*. Malden, MA: John Wiley & Sons.

DeLyser, D. & D. Sui (2013) Crossing the qualitative-quantitative divide II: Inventive approaches to big data, mobile methods, and rhythmanalysis. *Progress in human geography,* 37**,** 293-305.

Devlin, J., M.-W. Chang, K. Lee & K. Toutanova (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:.04805*.

DiMaggio, P., M. Nag & D. Blei (2013) Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding. *Poetics,* 41**,** 570-606.

Fei, H.-T. (1946) Peasantry and gentry: An interpretation of Chinese social structure and its changes. *American Journal of Sociology***,** 1-17.

Ferrara, E., O. Varol, C. Davis, F. Menczer & A. Flammini (2016) The rise of social bots. *Communications of the ACM,* 59**,** 96-104.

Flores, R. D. (2017) Do anti-immigrant laws shape public sentiment? A study of Arizona's SB 1070 using Twitter data. *American Journal of Sociology,* 123**,** 333-384.

Fong, E. & K. Shibuya (2005) Multiethnic Cities in North America. *Annu. Rev. Sociol.,* 31**,** 285-304.

Fu, Q., Y. Zhuang, J. Gu, Y. Zhu & X. Guo (2021) Agreeing to disagree: choosing among eight topic-modeling methods. *Big Data Research,* 23**,** 1-9.

Graham, M., S. A. Hale & D. Gaffney (2014) Where in the world are you? Geolocation and language identification in Twitter. *The Professional Geographer,* 66**,** 568-578.

Gregory, D. 1994. *Geographical imaginations*. Oxford, UK: Blackwell.

Guo, Y. 2004. *Cultural nationalism in contemporary China*. London, UK: Routledge.

Han, R. (2015) Defending the authoritarian regime online: China's "voluntary fifty-cent army". *The China Quarterly,* 224**,** 1006-1025.

Harvey, D. 1996. *Justice, nature, and the geography of difference*. Cambridge, MA: Blackwell.

Hiers, W., T. Soehl & A. Wimmer (2017) National trauma and the fear of foreigners: How past geopolitical threat heightens anti-immigration sentiment today. *Social Forces,* 96**,** 361-388.

Hsu, M. Y. 2000. *Dreaming of gold, dreaming of home: Transnationalism and migration between the United States and South China, 1882-1943*. Stanford, CA: Stanford University Press.

Hu, Y., C. Deng & Z. Zhou (2019) A semantic and sentiment analysis on online neighborhood reviews for understanding the perceptions of people toward their living environments. *Annals of the American Association of Geographers,* 109**,** 1052-1073.

Huang, H. (2015) International knowledge and domestic evaluations in a changing society: the case of China. *American Political Science Review,* 109**,** 613-634.

Huang, Y. (2011) The Sick Man of Asia-China's Health Crisis. *Foreign Aff.,* 90**,** 119.

Hughes, C. 2013. *Taiwan and Chinese nationalism: National identity and status in international society*. Oxford, UK: Routledge.

Hwang, K. K. (1999) Filial piety and loyalty: Two types of social identification in Confucianism. *Asian Journal of Social Psychology,* 2**,** 163-183.

Isaacs, H. R. 1980. *Scratches on our minds: American views of China and India*. Armonk, NY, USA: ME Sharpe.

Kayal, P., M. Singh & P. Goyal. 2020. Weakly-supervised deep learning for domain invariant sentiment classification. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, 239-243. Hyderabad, India.

Keyes, C. F. 2019. *Thailand: Buddhist kingdom as modern nation state*. New York, NY: Routledge.

Kim, S. E. (2018) Media bias against foreign firms as a veiled trade barrier: Evidence from Chinese newspapers. *American Political Science Review,* 112**,** 954-970.

King, G., J. Pan & M. E. Roberts (2017) How the Chinese government fabricates social media posts for strategic distraction, not engaged argument. *American political science review,* 111**,** 484-501.

Knight, D. B. (1982) Identity and territory: geographical perspectives on nationalism and regionalism. *Annals of the Association of American Geographers,* 72**,** 514-531.

Kohn, H. 1967. *The idea of nationalism: A study in its origins and background*. London, UK: Transaction Publishers.

Kuhn, P. A. 1990. *Soulstealers*. Cambridge, MA, USA: Harvard University Press.

---. 2002. *Origins of the modern Chinese state*. Stanford University Press.

---. 2009. *Chinese among others: Emigration in modern times*. Lanham, MD: Rowman & Littlefield.

Kwan, M.-P. (2016) Algorithmic geographies: Big data, algorithmic uncertainty, and the production of geographic knowledge. *Annals of the American Association of Geographers,* 106**,** 274-282.

Landauer, T. K., P. W. Foltz & D. Laham (1998) An introduction to latent semantic analysis. *Discourse processes,* 25**,** 259-284.

Lazer, D. & J. Radford (2017) Data ex machina: introducing big data. *Annual Review of Sociology,* 43**,** 19-39.

Lee, E. 2003. *At America's gates: Chinese immigration during the exclusion era, 1882-1943*. Chapel Hill, NC: University of North Carolina Press.

Levy, O. & Y. Goldberg. 2014. Neural word embedding as implicit matrix factorization. In *Advances in neural information processing systems*, 2177-2185. Montreal, Canada.

Ley, D. 2011. *Millionaire migrants: Trans-Pacific life lines*. Oxford, UK: John Wiley & Sons.

Li, P. S. 1998. *The Chinese in Canada*. Don Mills, ON: Oxford University Press.

Liang, S. 1949. *The essential meanings of Chinese culture*. Hong Kong: Zheng Zhong Press.

Lin, N. (1995) Local market socialism: local corporatism in action in rural China. *Theory and Society,* 24**,** 301-354.

--- (2001) Guanxi: A conceptual analysis. *Contributions in Sociology,* 133**,** 153-166.

Liu, C., Z. Jin, J. Gu & C. Qiu. 2017. Short-term load forecasting using a long short-term memory network. In *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, 1-6. Torino, Italy.

Long, J. S. & J. Freese. 2006. *Regression Models for Categorical Dependent Variables Using Stata*. College Station: Stata Press.

Lyman, S. M. & M. B. Scott (1967) Territoriality: A neglected sociological dimension. *Social problems,* 15**,** 236-249.

Maas, A. L., R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng & C. Potts. 2011. Learning word vectors for sentiment analysis. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies*, 142-150. Portland, OR, USA: Association for Computational Linguistics.

Martin, D. G. (2003) "Place-framing" as place-making: Constituting a neighborhood for organizing and activism. *Annals of the Association of American Geographers,* 93**,** 730-750.

Martin, M. E. & N. Schuurman (2017) Area-based topic modeling and visualization of social media for qualitative GIS. *Annals of the American Association of Geographers,* 107**,** 1028-1039.

Mawani, R. (2004) "Cleansing the Conscience of the People": Reading Head Tax Redress in Multicultural Canada. *Canadian Journal of Law & Society/La Revue Canadienne Droit et Société,* 19**,** 127-151.

Metzger, S. (2004) Charles Parsloe's Chinese Fetish: An Example of Yellowface Performance in Nineteenth-Century American Melodrama. *Theatre Journal,* 56**,** 627-651.

Meuleman, B., E. Davidov & J. Billiet (2009) Changing attitudes toward immigration in Europe, 2002–2007: A dynamic group conflict theory approach. *Social science research,* 38**,** 352-365.

Mughan, A. & P. Paxton (2006) Anti-immigrant sentiment, policy preferences and populist party voting in Australia. *British Journal of Political Science,* 36**,** 341-358.

Olzak, S. 1994. *The dynamics of ethnic competition and conflict*. Palo Alto, CA, USA: Stanford University Press.

Peters, M. E., M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee & L. Zettlemoyer (2018) Deep contextualized word representations. *arXiv preprint arXiv:.05365*.

Peyrefitte, A. 2013. *The immobile empire*. New York, NY, USA: Vintage.

Purcell, V. 1965. *The Chinese in Southeast Asia*. London, UK: Oxford University Press.

Pye, L. W. (1990) China: erratic state, frustrated society. *Foreign Affairs,* 69**,** 56-74.

Radford, A., R. Jozefowicz & I. Sutskever (2017) Learning to generate reviews and discovering sentiment. *arXiv preprint arXiv:1704.01444*.

Raffles, T. S. 1830. *The history of Java*. London, UK: John Murray.

Ritzer, G. & J. Stepnisky. 2017. *Modern sociological theory*. Thousand Oaks, CA, USA: Sage publications.

Rose, S., D. Engel, N. Cramer & W. Cowley (2010) Automatic keyword extraction from individual documents. *Text mining: applications and theory,* 1**,** 1-20.

Rossabi, M. 1983. *China among Equals: The Middle Kingdom and Its Neighbors, 10th-14th Centuries*. Univ of California Press.

Sack, R. D. 1986. *Human territoriality: its theory and history*. Cambridge, UK: Cambridge University Press.

Schissel, B., R. Wanner & J. S. Frideres (1989) Social and economic context and attitudes toward immigrants in Canadian cities. *International Migration Review,* 23**,** 289-308.

Schutz, A. 1970. *Alfred Schutz on phenomenology and social relations*. Chicago, IL, USA: University of Chicago Press.

Secor, A. (2004) "There is an Istanbul that belongs to me": citizenship, space, and identity in the city. *Annals of the association of American Geographers,* 94**,** 352-368.

Semyonov, M., R. Raijman & A. Gorodzeisky (2006) The rise of anti-foreigner sentiment in European societies, 1988-2000. *American sociological review,* 71**,** 426-449.

Soja, E. W. 1985. The spatiality of social life: towards a transformative retheorisation. In *Social relations and spatial structures*, 90-127. New York, NY, USA: Springer.

Sorokin, P. A. 1943. *Sociocultural Causality, Space, Time*. Durham, NC: Duke University Press.

Stanley, T. J. 2011. *Contesting white supremacy: School segregation, anti-racism, and the making of Chinese Canadians*. Vancouver, Canada: UBC Press.

Stock, B. 1983. *The implications of literacy: Written language and models of interpretation in the eleventh and twelfth centuries*. Princeton, NJ, USA: Princeton University Press.

Teh, Y. W., M. I. Jordan, M. J. Beal & D. M. Blei (2006) Hierarchical Dirichlet Processes. *Journal of the American Statistical Association,* 101**,** 1566-1581.

Tienhaara, N. 1974. *Canadian Views on Immigration and Population: An Analysis of Post-War Gallup Polls*. Ottawa, Canada: Information Canada.

Tu, W.-M. (1991) Cultural China: The periphery as the center. *Daedalus,* 120**,** 1-32.

---. 1998. Confucius and confucianism. In *Confucianism and the family,* eds. W. H. Slote & G. A. D. Vos, 3-36. Albany, NY, USA: State University of New York Press.

Tuan, Y.-F. 1977. *Space and Place: The Perspective of Experience*. Minneapolis, USA: University of Minnesota Press.

--- (1980) The significance of the artifact. *Geographical review,* 70**,** 462-472.

---. 1990. *Topophilia: A study of environmental perceptions, attitudes, and values*. New York, USA: Columbia University Press.

Tufekci, Z. (2014) Big questions for social media big data: Representativeness, validity and other methodological pitfalls. *arXiv preprint arXiv:1403.7400*.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser & I. Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*, 5998-6008. Long Beach, USA.

Walker, C. & R. W. Orttung (2014) Breaking the news: The role of state-run media. *Journal of Democracy,* 25**,** 71-85.

Wang, C., J. Paisley & D. M. Blei. 2011. Online variational inference for the hierarchical Dirichlet process. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. Fort Lauderdale, FL, USA.

Wang, L. L.-c. (1991) Roots and changing identity of the Chinese in the United States. *Daedalus,* 120**,** 181-206.

Wickberg, E. 1965. *The Chinese in Philippine life, 1850-1898*. New Haven, CT: Yale University Press.

Xiang, B., B. S. Yeoh & M. Toyota. 2013. *Return: nationalizing transnational mobility in Asia*. Durham, NC: Duke University Press.

Xie, Q., Z. Dai, E. Hovy, M.-T. Luong & Q. V. Le (2019) Unsupervised data augmentation for consistency training. *arXiv preprint arXiv:1904.12848*.

Xie, Y. (2016) Understanding inequality in China. *Chinese Journal of Sociology,* 2**,** 327-347.

Yang, J.-s. (2005) Imaging National Humiliation:'Sick Man of East Asia'in Modern Chinese Intellectual and Cultural History. *Journal of History of National Chengchi University,* 23**,** 1-44.

Zhou, X. (2021) Chinese bureaucracy through three lenses: Weberian, Confucian, and Marchian. *Management and Organization Review***,** 1-28.

Zou, L., N. S. Lam, H. Cai & Y. Qiang (2018) Mining Twitter data for improved understanding of disaster resilience. *Annals of the American Association of Geographers,* 108**,** 1422-1441.