Chapter 3

Methods and Data
Chapter Three

Methods and Data

3.1 Introduction

In the introduction and chapter two, the central research focuses and distinctive features of Chinese populism have been outlined. This has served as the starting point of a research line to unravel the dynamics of the triangular relationship of “people-elite-government” in an ongoing and intensifying digitalized, platformized, and datafied society. In such a society, most of what we say and what we do has the capacity of being observed, recorded, stored, and analyzed (Cheney-Lippold, 2017). This has transformed processes of human actions into traceable and analyzable data, termed datafication. The datafication of society manifests itself mainly in three aspects: dematerialisation, liquification, and density (Normann, 2001).

Dematerialization emphasizes the capacity to dissociate an asset’s or resource’s informational component and its contextual use from the physical world. Liquification highlights the idea that data can be easily governed, filtered, manipulated, and moved around through digital platforms. Density is the result of the value creation process and is the best (re)combination of resources that are mobilized for a specific context at a given time and location (Lycett, 2013). This nature of datafication in our digital age has had the effect of prompting researchers to rethink the core elements of the research enterprise and the logic of value creation within research.

Datafication provides revolutionary opportunities as well as challenges for researchers examining populism. On the one hand, the transformation of social actions into quantified data creates new possibilities to record, track, and analyze populist discourse. Digital platforms (e.g., Sina Weibo, Facebook, Twitter, and Tiktok) provide populists with a largely free space and cost-effective channel through which they can both generate and disseminate their populist message. In this way, digital platforms have the potential to function as databases (Smit, Heinrich & Broersma, 2017), where populist discourse can be better recorded and tracked and, later, accessed for analysis. On the other hand, while datafication brings opportunities, it also brings challenges. Prominent among these is when the consistent accumulation
of data reaches an unprecedented scale (i.e. big data), beyond the human processing capability. A further challenge concerns the variety of data collected and created. Variety proposes that data, coming from multiple sources in different forms, is “often unstructured, error ridden, and inconsistent in nature” (Lycett, 2013, p. 381). As growing piles of datasets are being produced, monitored, and collected on digital platforms, there is a pressing need for methodological innovation to deal with the various genres and the large scale of data sets in ways that provide order. In response to this challenge, there has been efforts to develop analytical approaches to big datasets that utilize the processing power of computers, but there are also pleas for methods that combine computational methodology with the expert knowledge and hermeneutic skills of human beings (Nelson, 2020).

This chapter explores how the research in this thesis addresses these challenges on two fronts. Regarding the challenges of various genres and large scale of data sets, it first explains the necessity and advantages of the applied mixed methodological approaches, which combine discourse analysis with multimodal discourse analysis (Kress & Van Leeuwen, 2001) and grounded theory (Glaser & Strauss, 1999, 2005) with computational grounded theory (Nelson, 2020). Then, it demonstrates the methodological challenges of identifying populist cases in the context of China, which is followed by a section on detailed data collection and sampling. In doing so, it goes on not only to explain why these mixed methods can help to address the challenges outlined above, but also how they can be applied to analyze the various genres of data, such as textual and visual data, and how they can be employed to investigate a relatively big data set and help to mine the potentialities of the data set by using a mixed methodological approach. Finally, this chapter ends by explaining the potential limitations and ethical concerns of this research and how they have been addressed.

3.2 Mixed methods and triangulation

In the face of these challenges, mixed methodological approaches have been designed for this thesis, with individual methodological approaches adapted to respond to specific research questions and the nature of the data in each of the empirical studies. These include the qualitative methodologies of discourse analysis (chapter four), multimodal discourse analysis (chapter six), and grounded theory
(chapter five), and the mixed method of computational grounded theory (chapter five). This section explains the benefits of combining these several research techniques for the current inquiry.

Using a mixed-method approach, both within and across studies, can be beneficial for a variety of reasons. For the purpose of this thesis, mixing diverse methods and tools for data analysis is advantageous for two primary reasons. First, the mixed methodological approach allows for effective mining of data and makes full use of the affordances of the data being explored so that findings can be triangulated. This increases the validity and reliability of the research and its contributions to the wider field of populism research. To give one example, in chapter five, Latent Dirichlet Allocation (LDA) topic modeling is the first step of computational grounded theory (this will be illustrated in Section 3.5.2). LDA is then combined with grounded theory to analyze data in the form of posts and hashtags gathered from Sina Weibo. LDA topic modeling is a computer-assisted content analysis technique that is widely used in research to explore the thematic structure of a large data set in media and communication research. However, when using LDA topic modeling alone, it faces at least four challenging questions (Maier et al., 2018): (a) how to appropriately pre-process unstructured text data, e.g., by deciding whether to delete hashtags or not when analyzing Twitter or Weibo data; (b) how to appropriately choose algorithmic parameters, e.g., by identifying the number of topics; (c) how to evaluate the validity and improve the reliability of the topic model; and, (d) how to appropriately interpret the resulting topics.

The responses taken to each of these challenges can have a particular effect on the reliability and validity of the research. And, in many cases, these challenges cannot be solved by using LDA topic modeling alone. When analyzing data (hashtags and posts) from Twitter or Sina Weibo, the results are often deemed invalid if hashtags are not deleted or filtered out during the LDA topic modeling process, since hashtags are the most frequently appearing in a data set. However, if the hashtags are deleted, then the results are not considered as reliable because they will face the contradiction of investigating the thematic structure of a data set while deleting important references to those themes, as hashtags themselves indicate topics. To avoid these outcomes, and based on the nature of the Twitter and Sina Weibo data,
a combination is made of grounded theory analysis of hashtags (which is within the analyzing capacity of human researchers) and LDA topic modeling of the posts or tweets (which is often a larger dataset, beyond the reasonable analyzing capacity of human beings in the constraints of specific projects). This mixed method approach not only fully exploits the potential of the data set but also makes it possible to triangulate the results by comparing human and machine coding, which helps to choose the number of topics to be generated. In this thesis, this is achieved by adapting Nelson’s (2020) methodology for computational grounded theory, outlined further below (section 3.5.2) and in chapter five.

Second, a mixed methodological approach has the potential to “offset the respective strengths and weaknesses of individual methods” (Swart, 2018, p.43). In particular, the need for using mixed methods is increasingly important when dealing with various genres and hybrid data. For instance, when exploring online populist visual communication in China in chapter six, an essential set of data is weaponized internet memes, which often combine texts and images. Multimodal discourse analysis (Kress & Van Leeuwen, 2001) can help disentangle the visual semiotics of these internet memes. However, it is insufficient in explaining how distinctive semiotic resources are interwoven with one another and how they are, then, recontextualized to form a new arrangement of meaning. The methodological approach to analyzing internet memes proposed by Wiggins (2019) responds in part to this challenge by providing a forceful framework to reveal how ideology is generated through intertextualizing semiotic resources. At the same time, Wiggins (2019) does not offer a method for how to identify semiotics from internet memes. Therefore, a combination of multimodal discourse analysis from Kress and Van Leeuwen and the internet memes analysis framework from Wiggins offsets the respective strengths and weaknesses when using them alone.

3.3 Methodological challenges

One specific methodological challenge that has been encountered in this research has been the difficulty of identifying populist cases for further empirical analysis. In the previous chapter, analysis demonstrated that Chinese populism is distinct from other forms of populism by the absence of charismatic leaders (He, Eldridge & Broersma, 2021). This feature of Chinese populism poses challenges for researchers.
Methods and Data

Put simply, whereas in liberal democratic contexts charismatic leaders as well as their associated populist parties serve as mediators between the people and the elite, portraying themselves as the true representatives of the people (Moffitt, 2016; Couperus, Tortola & Rensmann, 2022), this cannot be used as a criterion for identifying cases in China. To further detail the challenges this poses, a great deal of research on populism revolves around the speech (Wodak, 2021), performance (Moffitt, 2016), and media use of these charismatic leaders (Ernst, et al., 2019), such as Donald Trump in the United States, Geert Wilders in the Netherlands, Marine Le Pen in France, or Jair Bolsonaro in Brazil. However, the absence of charismatic leaders has not only left research on populism in China without a clear person to center on in defining the object of analysis, it also makes it difficult to identify whether a case is a populist case or not because the contexts in which populist performances might emerge are, comparatively, different outside democracies. In order to ensure the cases selected for this study are populist cases, rather than some other form of online collective actions, this thesis adopts an indirect identification strategy by focusing on cases that have been accepted as populist by academia. The identification process can be broadly separated into two stages: (1) identifying populist research articles in CNKI (zhiwang, Chinese National Knowledge Infrastructure, which provides the largest electronic collection of full-text research articles across disciplines) and from within that research, (2) identifying populist cases.

In the stage of identifying populist research articles, four criteria were used to select articles that would be relevant to the study of Chinese populism. First, we limit our analysis to full-length articles that include “民粹” (mincui, populism) in their titles. This term was selected because it is the root word for a wide range of possible terms being used in populism research. For instance, “民粹主义” (mincui zhuyi, populism) “民粹主义政党” (mincui zhuyi zhengdang, populist party), and “民粹主义领袖” (mincui zhuyi lingxiu, populist leader) all include “民粹” as a root word. This resulted in 914 articles. Second, the sample is further narrowed down, to only include those articles which are indexed by the Chinese Social Science Citation Index (CSSCI) from 1998 to 2020 (CSSCI index was built in 1998). When an article is indexed by CSSCI this is a reasonable measure of its academic reputation since the CSSCI is the most reliable and comprehensive citation index in China. This narrowed
the sample to 357 articles. Then inclusion and exclusion criteria were applied. Since this study focuses on populist phenomena and related discussion in mainland China, research articles concentrating on populism outside China were excluded. This resulted in 160 articles in our sample. The fourth criteria, as a final check, cross-compared the sample to the list of most-cited articles on populism. This ensures our sample reflects the state of the art of Chinese populism research. The results demonstrated that our 160 sampled articles were among the most referenced and discussed in the field. Considering its influence, we also included Yu’s (1997) article “Populism in the Process of Modernization,” which was published in 1997, before the formation of the CSSCI. This resulted in 161 populism research articles to be analyzed.

In the next stage of identifying populist cases, the sample was further narrowed to academic articles that explicitly addressed contemporary populism cases in mainland China (128 articles), from which we identified cases that were studied by at least two scholars (a case mentioned by the same author in different papers counts as one case). A total of 61 cases from between 2003 and 2020 were identified and included. For the convenience of future research, we then identified a keyword to label each case. We choose the label that is most frequently used by scholars as the keyword for each case. These selecting procedures have, to some extent, ensured that the cases chosen for this study are populism cases that are academically accepted by scholars, and that they have been separated by scholars from other forms of online collective actions. The selection criteria and procedures are demonstrated in Table 3.1, and the labels of all the cases are demonstrated in Table 3.2. For a complete discussion of these labels, I refer to chapter two.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Criteria</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identify articles</td>
<td>Keyword search: 民粹 (mincui)</td>
<td>914 articles</td>
</tr>
<tr>
<td></td>
<td>Citation index check: Narrow to articles indexed in CSSCI</td>
<td>357 articles</td>
</tr>
<tr>
<td></td>
<td>Inclusion criteria applied: Focused on populism in mainland China, or discussed theories on populism in general and related to the nature of Chinese populism 160 articles Exclusion criteria applied: Excluded all articles that addressed populism outside mainland China</td>
<td>160 articles</td>
</tr>
</tbody>
</table>
Final check: Do these articles reflected the state of the art of Chinese populism research based on citation figures? 161 articles

Contemporary populist cases that have been explored by at least two scholars from 1990 to 2020 61 cases

Table 3.2 The Labels of Populist Cases from 2003 to 2020 in China

<table>
<thead>
<tr>
<th>Categories</th>
<th>Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anti-Japan March</td>
<td>Anti-Japan March (2012); Boycott Carrefour (2008); Chou Tzu-yu case (2016); Diba Expedition (2016); Southern China Sea case (2016); Boycott Letian supermarket (2016)</td>
</tr>
<tr>
<td>Exclusion</td>
<td>Sun Zhigang case (2003); Dingzihu case (2007); Yang Jia case (2008); Hangzhou 70km/h Drag Racing case (2009); Lei Yang case (2016); Nie Shubin case (2016); Ha’erbin BMW case (2003); Deng Yujiao case (2009); Shanghai “fishing” entrapment case (2009); Li Changkui case (2009); Yao Jiaxin case (2010); Qian Yunhui case (2010); Yihuang Self-immolation incident (2010); Jia Jinglong case (2015); Yang Gailan case (2016);</td>
</tr>
<tr>
<td>Anti-establishment</td>
<td>Tang Hui case (2006); Shangxi Brick Kiln case (2007); Ha’erbin Police Bit Student case (2008); Luo Caixia case (2009); “My father is Li Gang” case (2010); Ha’erbin Kill Doctor case (2012); Gang Rape case (2013); Dongguan Prostitutes case (2014); Xu Chunhe case (2015); Wei Zexi case (2016); Liu Yong case (2003); Huanan Tiger case (2007); The Dispute on Retirement Age case (2008); Luxury Cigarette case (2008); 723 High Speed Train case (2011); Guo Meimei case (2011); Wu Ying case (2012); Yang Dacai case (2012); Cai Bin case (2012); Lei Zhengfu (2012); Tanggu Explosion case (2015); Vaccine Scandal case (2018)</td>
</tr>
<tr>
<td>Anti-intellectual</td>
<td>Chongqing Wanzhou case (2004); Anhui Chizhou Case (2005); Xiamen PX case (2007); Sichuan Dazhu Case (2007); Weng’an Case (2008); Shishou case (2009); Fangshen case (2012); Peng Yu (2007); Xiao Yueyue case (2011); Yu Huan case (2016)</td>
</tr>
<tr>
<td></td>
<td>Furong sister (2004); A Murder triggered by a Steamed Roll (2005); Super Girl (2006); Bombard Yushi Mao (2007); Jiang Ping (2007); CCSTV New Year Gala (2009); Thesis Plagiarism (2009); 叫兽(roaring beast, Jiaoshou, another name of professor, 教授 jiaoshou), 砖家(speciousist, zhuanjia, another name of specialist 专家, zhuanjia); Xili Brother (2010);</td>
</tr>
</tbody>
</table>

3.4 Data collection

Based on the specific research questions and an increasingly narrow research focus within each study in this thesis, various types of data were gathered. Much of the
empirical data that was used in this study was gathered from digital media platforms. These include: Baidu Tieba, which literally means “post bar,” an online forum affiliated with the Chinese search engine company Baidu; Sina Weibo, the Chinese equivalent of Twitter, with 249 million daily active users in December 2021, and Zhihu, a Quora-like question-and-answer platform in China, with 68 million registered users and 18.5 million daily active users. These platforms were selected for several reasons. First, these online spaces encourage and empower netizens to comment on, discuss, and even satirize political power (Yang, 2009). That makes them ideal channels for netizens to address their voices, appeals, and discontent. Furthermore, the posts and comments areas of these platforms function as “discursive arenas where members of subordinated social groups invent and circulate counter-discourses, which in turn permit them to formulate oppositional interpretations of their identities, interests, and needs” (Fraser, 1992, p.67). The contentious nature of discourse in this online public space aligns with the contentious features of populist politics. Consequently, they become platforms that allow populist cases to emerge and to be discussed.

Second, these digital platforms are ideal spaces for investigating the dynamic discourse polarization of multiple actors online. On the one hand, Weibo, Tieba, and Zhihu afford individuals opportunities to increase their autonomy and independence through expression (Shi & Yang, 2016). On the other hand, they also afford authorities the means to constrain this expression by monitoring, filtering, and deleting “unfriendly” discourse that may potentially threaten, or be seen as threatening, the stability of the state. Thus, populist anti-establishment appeals are often censored, filtered, and deleted (King et al., 2013). Furthermore, with the affordances of digital technology and the changing digital ecology in China, “the government has updated their online strategies, moving beyond censorship by taking a proactive role” to shape, redirect public opinion (Sun, 2018, p. 112). As such, online discussions on these digital platforms offer us the opportunity to explore how various social actors, such as netizens, governments, and digital platforms themselves, interact with each other to shape the formation of public opinion. In doing so, it is worth bearing in mind that working with data from these platforms entails working with data that has necessarily been affected by the censorship regime, and as such some potential data might have been deleted or blocked.
Methods and Data

post-publication (King, Pan & Roberts, 2013). These conditions of censorship, however, also allow us an opportunity and a context in which scholars can investigate any apparent tensions between resistance and control on Chinese social media, as discussed in chapter five.

3.4.1 Case and data for online populists’ articulation of the people

When turning to the purpose of disentangling the meanings attached to the people, the first step was to search for popular discourse using the keywords that were identified for each recognized populism case (see Table 3.2) on Baidu Tieba, Sina Weibo, and Zhihu. Particular attention was paid to the process of articulation to understand how the people are being articulated in netizens’ posts and comments. For earlier cases that predated these platforms or their available data, we rely on scholarly publications to investigate how scholars discussed popular discourse, but also references made to these previous cases that surfaced in our analysis of contemporary online discourse. Based on the antagonistic people (netizens)-elite (other) relationship, the 61 cases were classified into three supra-categories: exclusion (6 cases), anti-intellectualism (9 cases), and anti-establishment (46 cases). Exclusion is reflected in those cases where people were seen as revolting against foreign others (or “ultranationalism”) (Schroeder, 2021) and where elites were seen as betraying their Chinese identity. Anti-intellectualism cases include not only those which were associated with the affective aversion to scientists and experts (Hofstadter, 1963), but also where anti-intellectualism emerged through mass support for a satirical sub-culture that challenges the hegemony of elite-dominated cultural production and cultural institutions (Sun, 2006). Anti-establishment cases include those where appeals and expressions of opposition to those wielding power were made. This grouping allowed the analysis to further identify three types of populism in China: national populism, cultural populism, and online populism.

For each category, one specific case was then selected to further examine different meanings of the people based on specific political and cultural settings. These cases were selected first because they triggered wide public discussion online. Various actors are stirred and get involved in certain actions, like reposting, commenting, and appealing, further raising public attentions and concerns. Second, cases were also selected due to their rebellious nature. Netizens not only show their
sympathy to the disadvantaged groups, but are also rebellious towards the corrupt elite, established-institutions, states, and “others” through online connective actions. Thus, they allow this research to investigate the dynamics of the “people-elites-government” relationship.

For the category of national populism, the Diba Expedition to Facebook in 2016 was selected (Diba is a Baidu Tieba-based online community with more than 30 million followers). At the beginning of 2016, Diba netizens were mobilized to revolt against Chou Tzu-Yu, a 16-year-old Taiwanese singer, for her behavior of waving the flag of the Republic of China (the formal name for Taiwan) on a South Korean TV show, and Tsai Ing-wen, Apple Daily, and SET News who support Chou.

CCSTV Spring Festival Gala (China Countryside Television, Shanzhai Chunwan) was selected as an illustrative case study for cultural populism due to its impact on the rise of Shanzhai culture, which has a rebellious nature (de Kloet & Chow, 2017), and has the intention of delegitimizing the authority of elite-dominant cultural institutions through comedic imitation (Chen, 2014). The CCSTV Spring Festival Gala was initiated by Mengqi Shi. After attending the official CCTV Spring Festival Gala, he criticized it for being “not designed for the ordinary audience... the audience was nothing but high officials and rich people.” So, he organized a program produced by and for the ordinary people: the CCSTV Gala.

Dongguan Anti-Pornography Movement in 2014 was selected for the category of online populism. In 2014, CCTV journalists uncovered pornographic KTV (karaoke) using hidden cameras in Dongguan. After this was reported, more than six thousand police agents wiped out local KTVs (Tai, 2015), and thousands of prostitutes were arrested. Although prostitution is illegal in China, Dongguan prostitutes gained sympathy from netizens who regarded them as members of the collective “us” of socially vulnerable groups (Xia, 2014).

3.4.2 Case and data for discursive logic of online populists’ articulation

To study the formation of online public opinion and the dynamic discourse polarization of multiple actors online, an exemplar case of #DrivingIntoThePalaceMuseum was selected. This is one of several recent,
prominent cases in China where individual misbehavior has sparked public outrage. In this and similar cases, a surge in populist rhetoric emerges online, expressing outrage at the misbehavior, but this online outrage is quickly censored. Such cases allow us to explore how various powers have emerged and how they are interwoven in a dynamic polarization process.

On January 17, 2020, a Sina Weibo user posted a number of images of herself and a friend posing with a luxury car on the grounds of the Palace Museum. This online posting served as the catalyst for this incident. She added in her post: “On Monday, the Palace Museum was closed, so I hurried over, hid from the crowds, and went to play in the Palace Museum.” The museum, located in the Forbidden City, is one of the most important historical imperial palaces and world heritage sites in China. Since 2013, vehicles have been prohibited from entering its grounds. Normally, the museum is closed on Monday for maintenance. The woman appeared to have entered illegally. Her photos quickly went viral and set off a public outcry. Weibo users not only condemned her for possibly causing damage to the grounds, but also questioned how this could happen. Hyper-active netizens responded with a steady outpouring of indignation, directed at the privileges and freebies that were given to rich elites, and the exceptions to the rules they were granted. Eight hours after the woman made her initial posts, the Palace Museum apologized in a Weibo post, saying it would implement stricter management to avoid similar incidents in the future. However, rather than putting an end to the netizens’ indignation, this evasive statement fueled even more public outrage and led to further online discussion.

Weibo posts and hashtags revolving around this case were collected. The hashtag collection began with a People’s Daily Weibo post that included the hashtag #TheCuratorofThePalaceMuseumWangXuDongAppolizedtothePublic. A snowball sampling approach, characterized as a method for exploring dynamics of networking and referral (Parker, Scott & Geddes, 2019), was used to collect as many hashtags as possible. For this case, 35 hashtags were collected. Hashtags are best known as topic-makers, which are used to bring ad hoc publics together around a particular topic by encapsulating public opinion and sentiment (Weston, Chopra & Adams, 2014). Since hashtags capture and reflect consensus, they offer a lens through which to examine complaints from netizens and public opinion surrounding
an incident.

The second type of data in the sample is posts. By searching through the hashtags that were collected from Weibo, the 20 most commented-upon posts under each hashtag were manually collected. In the process of gathering posts, two criteria were adhered to. First, unique posts were gathered, and duplicates were removed. Second, a post is classed under the first mentioned hashtag if it contains several hashtags. In total, 657 posts were gathered. These data gathering selections are made based on the posts that received a large amount of public attention. For instance, one post from CCTV news attracted more than 33,000 comments and 600,000 “likes.”

3.4.3 Cases and data for populist visual communication

Finally, to study netizens’ affinity with the people and dislike of the elite or “others,” three Diba Expeditions (帝吧出征, diba chuzheng) in 2016, 2018, and 2019 were selected as case studies. On January 20, 2016, the online community ‘Diba’ initiated an expedition and mobilized its followers to bypass China’s firewall to access the Facebook pages of Taiwan’s then presidential candidate, Ing-wen Tsai, Apple Daily, and SET News, and to bombard these pages with memes, reacting to their pro-independence stance. On September 24, 2018, Diba summoned an expedition again, this time flooding the Facebook pages of Sweden’s Ministry of Foreign Affairs, Swedish national broadcaster SVT, and SVT host Jesper Rönndahl. This expedition was a protest against SVT for airing a television show that the Chinese embassy in Sweden condemned for provoking “racial hatred and confrontation targeting China and some other ethnic groups” (Embassy of the People’s Republic of China in the Kingdom of Sweden, 2018). Diba netizens organized a third expedition to support the Hong Kong police on August 17, 2019. In collaboration with those belonging to the community Fangirls (similar to fandom, referring to young female fans of celebrities in popular culture. 饭圈女孩, “fanquan nvhai” in Chinese), they targeted the Facebook, Twitter, and Instagram accounts of “violent” Hong Kong protesters.

These studies illustrate unique dimensions of online populist dynamics in several ways. First, the Diba Expeditions show how online populism interacts with digital nationalism (Schneider, 2021) or what Liu (2019) describes as fandom nationalism.
Methods and Data

This provides a chance to examine how populist “people-elite” and nationalist “we-other” antagonisms emerge through socially and culturally ingrained visual semiotics. Second, as part of the expedition, a large number of internet memes were made, turned into digital weapons, and spread. These memes were used to attack the social media accounts of people who were being targeted, not just during the expedition but also afterward, as these memes continue to spread online.

A corpus of internet memes was established to examine the production, dissemination, and evolution of expedition weapons. The memes in this corpus were collected in various ways. Sampling was first and foremost achieved by searching for the Diba community on Baidu. Because Diba not only serves as the base camp for online expedition movements, but also a base where meme packages are distributed and circulated that followers can use. Sampling was based on two criteria: (1) the meme packages contained at least one of the keywords “Diba Expedition” (帝吧出征), “Cyber Expedition” (网络出征) or “Fangirls Expedition” (饭圈出征) in their titles or posts before and during the period of the expedition; (2) meme packages gathered from Diba should contain the signifier seals. This search resulted in 456 internet memes gathered. After deleting duplicated memes, a sample of 193 internet memes was used for the analysis.

Table 3.3 Cases, Sources, and Data for the Studies of This Thesis

<table>
<thead>
<tr>
<th>Study</th>
<th>Cases</th>
<th>Source &amp; Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Who are the people?</td>
<td>Diba Expedition to Facebook in 2016</td>
<td>Source: Baidu Tieba (Diba) and Zhihu</td>
</tr>
<tr>
<td>(Chapter Four)</td>
<td></td>
<td>Data: (for instance) <a href="https://zhuanlan.zhihu.com/p/20522264">https://zhuanlan.zhihu.com/p/20522264</a></td>
</tr>
<tr>
<td></td>
<td>CCSTV Spring Festival Gala</td>
<td>Source: Baidu Tieba</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data: (for instance) <a href="https://tieba.baidu.com/p/5099530262?pid=106784691983">https://tieba.baidu.com/p/5099530262?pid=106784691983</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td>&amp;cid=0#106784691983</td>
</tr>
<tr>
<td></td>
<td>Dongguan Anti-pornography Movement in 2014</td>
<td>Source: Zhihu and Sina Weibo</td>
</tr>
<tr>
<td></td>
<td>#DrivingIntoThePalaceMuseum</td>
<td>Data: (for instance) <a href="https://www.zhihu.com/question/22708604">https://www.zhihu.com/question/22708604</a></td>
</tr>
<tr>
<td>Discursive logics</td>
<td></td>
<td>Source: Sina Weibo</td>
</tr>
<tr>
<td>(Chapter Five)</td>
<td></td>
<td>Data: 35 hashtags and 657 posts</td>
</tr>
<tr>
<td>Visual</td>
<td>Diba Expedition in</td>
<td>Source: Tieba and Zhihu</td>
</tr>
</tbody>
</table>
communication (Chapter Six) | 2016, 2018, and 2019 | Data: (for instance) 193 internet memes
|  |  | https://zhuanlan.zhihu.com/p/20522264
|  |  | https://www.zhihu.com/question/39679370/answer/825753

## 3.5 Data analysis

### 3.5.1 Discourse analysis and multimodal discourse analysis

Discourse analysis is widely used in qualitative social science research. Many scholars utilize the term discourse analysis to describe both what they do and how they do it. To understand what discourse analysis is, it is first necessary to know what discourse is. There are two ways of defining discourse considered in this thesis. The first approach stems from De Saussure (2011), who distinguishes between parole (linguistics of speaking) from langue (linguistics of language), and Mikhail Bakhtin (2010), who, in The Dialogic Imagination, argued that meaning is generated through the “primacy of context over text” and through intertextual relation (Maranhão, 1990). This approach defines discourse as both speech as well as its context, an approach based on analyzing linguistic data. The second is Foucault’s discourse/power approach (2019), defining discourse as a power. He argued that discourse is always influenced by historically established discursive practices. Thus, discourse analysis is employed to investigate how power emerges from and within discursive practices.

Following the “text-discourse-power” understanding of discourse leads us to the body of scholarship on critical discourse analysis (CDA), a type of discourse analysis that “primarily studies the way social power abuse, dominance, and inequality are enacted, reproduced, and resisted by text and talk in the social and political context” (Van Dijk, 2001, p. 352). According to Van Dijk (2001), CDA has the capacity to bridge the gap between the microlevel of social order (such as language use, communication, verbal interaction, etc.) and a macrolevel of analysis (such as power, structure, dominance, and inequality between social groups). Furthermore, CDA defines social power in terms of control, which means that “groups have (more or less) power if they are able to (more or less) control the acts and minds of (members of) other groups” (Van Dijk, 2001, p. 355). “Power as control” enables us to analyze how charismatic populist leaders and populist parties seize power through
controlling public discourse, to mobilize their followers. Similar to “power as control,” another key concept in CDA is *manipulation*, which can be tied to Laclau’s idea of *articulation* (Laclau & Mouffe, [1985] 2001) as a process of bringing discursive elements together to construct a new arrangement of meaning.

Adopting a constructionist understanding of discourse enables us to investigate how populists’ discursive power emerges through manipulated articulation. First, regarding populist discourse as an *articulation* process means that populist discourse is constructed through the selection and recontextualization of discourse elements to form a new meaning. For instance, understanding populist discourse as manipulated allows scholarship to investigate how the discursive power of *the people* emerges from netizens’ daily online discursive practices. Second, addressing populist discourse through the lens of manipulated articulation sees populist discourse polarization as a discourse that has been constructed with certain aims, goals, and intentions. These approaches to populist discourse also help reveal populist discourse polarization process by exploring why and how misinformation and disinformation are manipulated and spread online. Based on the textual and visual discourse that were gathered from digital platforms, the analysis first disentangles the specific meanings that are attached to *the people*, or who the people refer to in a specific context, and then reveals how the discursive power of *the people* emerges by manipulated recontextualizing within a broad historical and socio-cultural context (chapter four). In chapter six, discourse analysis is combined with multimodal discourse analysis (Kress & Van Leeuwen, 2001) to explore how meanings of visual semiotics are generated through the chain of populist visual communication.

**Mutimodal discourse analysis**

Moving beyond textual analysis, multimodal discourse analysis (Kress & Van Leeuwen, 2001) provides a methodological framework to investigate how several semiotic modes are interwoven and used to communicate messages in the highly mediated world. Semiotic modes refer to the “semiotic resources which allow the simultaneous realization of discourse” (Kress & Van Leeuwen, 2001, p. 21). The basis of multimodal discourse theory lies in two different but coherently related
ideas: *content stratum* and *expression stratum*. The content stratum explores how semiotic resources are designed, combined, and situated to form a new arrangement of meaning or discourse, which is “socially constructed knowledge of reality” (Kress & Van Leeuwen, 2001, p. 4). The expression stratum highlights the production and distribution of semiotic modes. Specifically, it investigates how semiotic modes are materially organized and produced and how they are technically preserved and distributed.

In this thesis, multimodal discourse analysis is applied first to identify the socially and culturally embedded visual semiotics from the widespread and weaponized internet memes. Then, Wiggins’ analytical framework of *ideology*, *semiotics*, and *intertextuality* is combined with multimodal discourse analysis to investigate how visual and textual semiotics are combined, designed, and contextualized to form a new arrangement of meaning. *Semiotics* primarily concerns how meaning is constructed through visual signs, produced for the intention of signification, meaning-making, and communication. *Intertextuality* refers to how a text is recontextualized to generate an additional layer of meaning. By tracing the meaning that is generated from the design and organization of visual and textual semiotics, it reveals how internet memes are weaponized to express netizens’ affinity to the people and their aversion to the elite and others.

<table>
<thead>
<tr>
<th>Content Stratum</th>
<th>Language</th>
<th>Visual semiotics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A:</strong> Political slogans: “We must liberate Taiwan.”</td>
<td><strong>U:</strong> Cartoon images;</td>
<td><strong>Mode one:</strong> Playful jiong style.</td>
</tr>
<tr>
<td><strong>B:</strong> Shameless language: “Dumbass.”</td>
<td><strong>V:</strong> Political iconography;</td>
<td><strong>Mode two:</strong> Political iconography</td>
</tr>
<tr>
<td><strong>C:</strong> Playful words: “Don’t push me, or you will loss your dad.”</td>
<td><strong>W:</strong> Exaggerated celebrities performances;</td>
<td><strong>Mode three:</strong> Colour</td>
</tr>
<tr>
<td><strong>D:</strong> Seals (language): “Diba Anti-Taiwan Independence Dog Exclusive seal”</td>
<td><strong>X:</strong> Political figures;</td>
<td></td>
</tr>
<tr>
<td><strong>E:</strong> Poets: “I live at the head of the Yangtze River, you live at the end of the Yangtze River. Thinking of you but not seeing you, building a new”</td>
<td><strong>Y:</strong> Scenery and images of traditional mainland Chinese food;</td>
<td></td>
</tr>
<tr>
<td><strong>Z:</strong> Seals (color);</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.5.2 Grounded theory and computational grounded theory

Grounded theory

This thesis aims to investigate the discursive logics that are followed during online discourse polarization and depolarization processes. Based on the data (hashtags and posts) that were gathered from Sina Weibo (see Table 3.3), it studies what new topics are emerging at different stages of an online bottom-up populist case and how these topics are discussed. These questions can be explored by the inductive approach of grounded theory. Grounded theory is a paradigm which provides systematic, yet flexible guidelines for gathering and analyzing data, allowing themes and categories to inductively emerge from data (Glaser & Strauss, 1999), and ultimately lead to a data-driven but abstract theoretical understanding of society (Charmaz, 2014). Grounded theory methods have long been applied by researchers to conduct rigorous theory-producing research for its advantages of demystifying “the conduct of qualitative inquiry” and promoting “gaining both analytic control and momentum” (Charmaz, 2014, p. 4). According to Glaser and Strauss (1999), rigorous grounded theory starts with sampling data with an open mind and letting interesting themes emerge from the data organically, rather than being restricted by research questions and extensive existing literature (Glaser & Holton, 2004). The practice of grounded theory includes several components, such as “simultaneous involvement in data collection and analysis,” “constructing analytic codes and categories from
data” constant comparisons during every stage of the analysis, and elaborate categories through memo-writing (Charmaz, 2014, p. 7). However, no matter how inductive a research set-up is, it is impossible for a researcher to start with a blank slate, as his or her mind is always occupied with knowledge of the field, and “research is generally guided by at least a basic empirical interest” (Miles & Huberman, 1994, p. 16-17). For this reason, instead of entering the field without any points of focus, “which may make it difficult to decide on a line of inquiry during data collection when phrasing questions or to interpret the data” (Swart, 2018, p. 54), grounded theory is mainly employed as a data analysis method to explore the semantic structure of data (hashtags) gathered from Sina Weibo.

Despite the advantages that grounded theory has, it also faces several challenges. The primary challenge of grounded theory is “subjectivities” (Saldana, 2015, p. 8), as: “The nature of grounded theory necessitates a series of ‘judgement calls’” (Nelson, 2020, p. 3). Researchers also bring “subjectivities,” “predispositions,” and “personalities” into the analytical process, in terms of sampling, coding, and interpreting data. As a result, validating and reproducing the research has become unusually difficult (Biernacki, 2012). In addition, grounded theory has been criticized for its inability to process large amounts of unstructured data (Bail, 2014). Due to the ongoing and intensifying datafication of our society, this criticism has only grown louder as big and bigger data sets are being compiled, growing beyond the analytical capacity of individual researchers working manually. This calls for a new approach to deal with big sets of unstructured data.

**Computational grounded theory**

In response to this challenge, Nelson (2020) has proposed the novel methodological approach of computational grounded theory to mitigate the drawbacks of purely qualitative research. The new approach “combines expert human knowledge and hermeneutic skills with the processing power and pattern recognition of computers, producing a more methodologically rigorous but interpretive approach to content analysis” (Nelson, 2020, p. 3). Compared with grounded theory, computational grounded theory is advanced not only for its ability to incorporate large amounts of data but also for scholars’ ability to reproduce these studies, which brings the
Methods and Data

analysis closer towards measures of validity and reliability.

According to Nelson (2020), computational grounded theory entails a three-step methodological framework, which includes pattern detection (computational), pattern refinement (manual), and pattern confirmation (computational). The step of pattern detection involves deconstructing, simplifying, and clustering complicated, unstructured, and often massive, amounts of data (often “text”) into simpler, frequently appearing, and more interpretable lists of keywords. With the processing power of a computer, computational methods can reveal patterns and semantic structures that would not be immediately available to human readers. Furthermore, they can process the data in a quicker and more reliable manner, and “they are completely reproducible” (Nelson, 2020, p. 23). The pattern refinement step refers to reengaging with data hermeneutically through computationally guided deep reading. Replicating the traditional grounded theory is necessary as it (1) assures the validity and reliability of the preliminary patterns that are identified from the previous step; (2) parses and interprets the preliminary patterns; and (3) modifies the preliminary patterns if necessary to better fit the whole nature of the data (Nelson, 2020). The third, pattern confirmation step, tests whether the patterns detected from previous steps are reliable and generalizable by utilizing a different computational method.

Latent Dirichlet Allocation (LDA) is widely utilized in the first pattern detection step (Nelson, 2020). LDA is a three-level hierarchical Bayesian model, which has been widely utilized in document modeling, text classification, and collaborative filtering (Blei, Ng & Jordan, 2003). As a computational content-analysis technique, LDA topic modeling is often used to detect the underlying themes of an unstructured collection of texts. This computational and data-driven nature of LDA makes it “attractive for communication research” (Maier et al., 2018, p. 1), because it allows for topic identification and semantic mining from large unstructured data in natural language processing. It decontextualises and simplifies the texts into keywords and calculates the co-occurrence of the keywords to yield better disambiguation of words and, thus, a more precise assignment of documents to topics. However, there are limits. As Nelson (2020) argues, “unsupervised topic modeling is an excellent research tool for some purposes but not for others and so should always be used with caution” (p. 19), as it inevitably faces the challenges of preprocessing data, subjectivities in choosing
the number of topics, evaluating the topic model, and interpreting the results (Maier et al., 2018).

This calls for a researcher to be cognizant of the nature of the data set one is exploring, and considering the merits and limits of approaches, including those within computational grounded theory. Given the nature of Sina Weibo data (including hashtags and posts), and the pros and cons of grounded theory and computational grounded theory, this thesis combines grounded theory with computational grounded theory to develop a novel approach to investigate the semantic structure of Weibo data. As noted in the outset of this chapter, applying LDA topic modeling alone would call for investigating the thematic structure of a data set while, simultaneously, deleting important hashtags in the data that would indicate important topics. By combining grounded theory analysis at the level of hashtags (which is within the analyzing capacity of human beings) and LDA topic modeling at the level of the posts (which is often beyond the analyzing capacity of human beings), this is mitigated. This novel combination not only helps to fully exploit the potential of the data set but also enables triangulation of the results of human coding and machine coding, thus contributing to selecting the number of topics to be generated as well as interpreting the resulting topics.
3.6 Limitations

The methodological designs described above allow for a solid investigation of the characteristics, discursive logics, and “pressure valve” functions of Chinese populism. However, like all research set-ups, they have limitations. First, as mentioned in the section of methodological challenges (Section 3.3), this thesis is limited in its case identification criteria. A case was selected only if it was addressed by at least two scholars and had been published in CSSCI indexed journals with “mincui” in the article titles. This risks excluding cases that are interesting illustrations of populism, but which are addressed by only one scholar or have not been published in CSSCI-indexed journals. Furthermore, the selection criteria focus on cases in contemporary China which excludes those cases that occurred before 1990. This restricts our understanding of communist populism and its impacts on contemporary online bottom-up populism in the digital world. Based on this study,
future research can broaden select criteria and deepen understanding of Chinese populism by taking cases before 1990 into consideration.

Second, this thesis aims to unravel the discourse polarization and depolarization processes on social media, as well as the dynamics between “people (netizens) - elite (others) - government” relationships. Thus, a novel methodological framework has been designed based on the aims of this study and the nature of Sina Weibo data, as presented in chapter five. This new approach is valid for its ability to triangulate and bring interpretation to the results, and advances our ability to explore big data. Nevertheless, these findings are still limited to a relatively small data set and a singular case. In the future, the results could be tested against a larger data set that includes a wider range of cases.

A final potential limitation pertains to the insufficient analysis of populist visual communication in a global environment. In chapter six, analysis identified visual semiotics that were utilized in weaponizing internet memes. Findings indicate how netizens demonstrate their affinity for the people and their aversion towards the elite. However, these findings are highly contextual. Since these internet memes were posted on widely accessible social media platforms, such as Facebook and Twitter, it remains unknown whether netizens outside of China would adopt a similar understanding of the visual semiotics that are utilized in weaponizing internet memes. In the future, more research can be done to further explore this, which will contribute to our knowledge of heterogeneous people on a global scale.

3.7 Research ethics

Due to dynamics of digitalization, platformization, and datafication, an increasing number of studies are conducted based on data gathered from digital platforms, such as Sina Weibo, Facebook, Twitter, and Instagram. The online and digital features of data generate new ethical concerns for social science research. Compared with traditional social science research, which has a longer history with explicit ethical regulations that scholars can follow, research on media and communication in a digitalized and platformized society have fewer clear rules to follow. For instance, some scholars (Marcus et al., 2012) argue that researchers should contact bloggers or internet users to get their consent before quoting their
content in research. Other scholars (Eysenbach & Till, 2001) argue that it is not necessary to get consent as the data produced by users is publicly available (Association of Internet Researchers (AoIR) ethics working committee, 2002). Facing such challenges, scholars suggest that traditional ethical regulations could be applied to the field of internet research by reflecting upon potential ethical tensions at different stages of sampling, collecting, storing, and analyzing.

Ethical concern has also been one of the central challenges in this study. First, if a consent form is required before the data collection, then researchers may have to contact hundreds to millions of internet users, which is impossible. For practical reasons, informed consent is not acquired from the internet users, not only due to the large scale of the data but also because of the semi-anonymous feature of the internet, which itself constrains the possibility for the researchers to identify specific individuals. Furthermore, in the big data projects, data are not generated on the level of individuals. Researchers cannot trace them based on the outcomes. Second, the other concern of studying the internet regards the potential risks to individuals, particularly in terms of cases which involve “human flesh search” (Pan, 2010) and weaponized internet memes in social-political contexts of China. “Human flesh search” and protests with weaponized internet memes themselves are malicious activities by netizens, which may lead to potential harm to the targeted individuals. However, if research on these malicious activities is against ethic rules, then it will result in no research being conducted. Thus, what we should do is to anonymise as much as possible to avoid the identification of specific individuals.

This research also takes ethical concerns into consideration at different stages of research practices. In terms of research design, sampling, and data collection, we only select platforms that are open to the public. Audiences can access the same content analyzed here, even without registration, and no approval is needed to get access to the data. To mitigate any potential harm or inconvenience that could be caused to relevant individuals who create these posts, all personal and confidential information that may lead to the identification of the specific individuals is either obscured in this thesis, or anonymized. In some cases, this is not possible. In the study of the # DrivingIntoThePalaceMuseum and of weaponized internet memes, content involving political figures and celebrities could be identified either by context
or signature (the pattern of posting). However, this involves material that features public commentary about public officials engaging in public life, individuals whose expectations of anonymity are lesser in these contexts.

Furthermore, the cases utilized in this study have attracted wide public attention and have been reported by official news agents in and outside of China. Finally, the data is archived on a private drive and can only be accessed by the researchers involved in the study, who follow the Research Data Policy (RDP) of the University of Groningen (2021 version) and the Research Data Management Protocol and Guidelines (RDMPG) of the Faculty of Arts. In particular, all data, including raw preliminary data, secondary processed data, is archived in a secure and fixed format based on the RDP and RDMPG. Finally, a metadata file is constructed to ensure the data is findable, accessible, interoperable, and reusable. This metadata file is first archived on my university workplace X drive, then backed up on a shared network Y drive\(^1\). The data is only used for academic purposes. When the results were written up and sent to be published, the personal profiles and information were changed so that no one could be identified.

### 3.8 Conclusion

This chapter has outlined the methodological design that the thesis utilized to investigate the discursive power and the meanings attached to the people, discursive logics and online populist visual communication. Through a mixed methodological framework which combines discourse analysis with multimodal discourse analysis and grounded theory with computational grounded theory, it explicitly takes the genre and size of data into consideration as challenges to researching populism which this thesis aims to overcome. On at least two levels, the data-driven methodological framework benefits the research. First, it addresses the pressing need to combine expert human knowledge and hermeneutic skills of human beings with the processing power of computers to deal with the ongoing and intensifying datafication of society. In particular, it provides a novel methodological

framework to explore the data from Sina Weibo, which includes hashtags and posts (and could also be used to explore Twitter data). This new method mitigates the shortcomings of LDA topic modeling for its inability to identify topics over time and the difficulty of choosing the number of topics. Second, the research design addresses the idea that, based on the genre and size of the data, various methodological approaches, such as qualitative, quantitative, and computational methods, can be combined to make research processes more transparent, adding transparency and interpretation to the research. This chapter also points out the challenges and potential shortcomings of the methodological design. Future research should move the understanding of Chinese populism forward by overcoming these drawbacks.