Factionalism and the Red Guards under Mao's China:

Ideal Point Estimation Using Text Data[*]

Yi-Nung Tsai[†], Dechun Zhang[‡], and David, Yen-Chieh Liao [§]

Preliminary Working Draft / HEROEs-Essex PhD Seminar

Abstract

This article estimates the spatial positioning of the political elites and participants in the Cultural Revolution through analyzing expressed political views in propaganda publication in China. Prior theoretical research indicates that the political elites and the Red Guards lost control of the social movement, and it evolved into sheer verbal violence and physical skirmishes across all provinces in the end. We show that in this political chaos, the Red Guards and the elite were dynamically fractioned, as is reflected in self-printed propaganda, such as big-character posters (Dazibao 大字报) and tabloids (Xiaobao 小报). This paper also develops a novel approach that augments TextRank algorithm to extract keywords and Chinese phrases, on top of the Poisson scaling model (Wordfish) to estimate the differences of spatial positions using the extracted textual key terms. Results are shown to be consistent with the literature on the Cultural Revolution.

Keywords: Text as Data, Cultural Revolution, Mass movements, the Red Guards

1 Introduction

This article proposes an estimation strategy to measure politicians' ideological positions from textual data, together with an application to study the factionalism of the Red Guards during the Cultural Revolution. One of the most intriguing features of the mass movements that sparked the Cultural Revolution is the intensity of factionalism that has drawn a series of studies' attention to exploring the nature and the origin of factionalism. The literature has well explained how activists of the mass movement (the Red Guards) were divided into multiple factions that fought for controls for ruling power, political resources and other interests, such as educational institutes, workplaces, and local governments, etc. (MacFarquhar and Schoenhals 2006; Walder and Su 2003; Walder 2006a; Walder 2009; Dikötter 2016; Walder 2016; Walder 2015a). For example, Andreas (2009), Tang (2003), Wang (1995), and Xu (1999) explains structures of fragmented sub-factions; Walder documents internal conflicts and the coalition under the movement (e.g. Walder 2004; Walder 2006a; Walder 2006b; Walder 2009). The dynamic factionalism, that is, the underlying conflicts and disputes evolving rapidly over time between subgroups, was emerged when the Red Guards in similar positions made conflicting choices in ambiguous contexts in the status quo (Walder 2006a, pp. 712 – 713). To defend earlier choices, they mobilized in attempt to eliminate antagonistic factions and yet, dynamic shifts in the contexts for political choice across time could alter the defined antagonists, generating waves and waves of new conflicts within the Red Guards (Walder and Su 2003; Wang 2019; Su 2006; Walder and Lu 2017; Walder 2016).

In this sense, estimating the positions of the Red Guards and the political elite (officials) is key to understanding factionalism and political cleavage during the period across China. However, how different factions among the Red Guards selected positions of their own, or stances towards allies and against rivalries, or attitudes towards the political elite, during the mass movement remains relatively untouched. The scarcity of the corresponding literature (or literature on factionalism in general) is mainly attributed to the lack of appropriate numerical data sources and effective measures. One of the best alternatives to the numerical data is the textual data, when the actual numerical data becomes unavailable. Measuring spatial political

positions using textual data is a fundamental aspect to comprehend potential political phenomena, such as factual discourse. Consequentially, textual scaling methods, like Wordfish (Poisson scaling model) as in Slapin and Proksch (2008) and a great variant of Wordfish-based (or -like) methods (e.g., Imai, Lo, and Olmsted 2016; Lauderdale and Herzog 2016; Laver, Benoit, and Garry 2003; Lo, Proksch, and Slapin 2014a; Will et al. 2011; Vafa, Naidu, and Blei 2005), have emerged. These methods, through scaling the positions of parties, reliably facilitate in testing political theories, e.g., intra-party politics, polarization, and representation (Bäck and Debus 2016; Catalinac 2017; Kim et al. 2018; Ishima 2020; Proksch and Slapin 2010). While useful for many purposes, these approaches are viable only when topical variation constant is held and loosely considerate semantic structure and linguistic context are represented in the text.

To this end, this paper contributes to develop an estimation strategy that first constructs specific dictionaries in Simplified Chinese, containing keywords (key characters) and phrases, and then uses Wordfish as the scaling method for estimation. Particularly, in the first step, the latest Universal Dependencies 2.5 model tokenizes Simplified Chinese documents, based on linguistic usage patterns of Chinese language in natural language processing task, and TextRank, developed by Mihalcea and Tarau (2004), effectively extracts a set of representative terms from the input text, based on values of importance they carry, to compose domain-specific dictionaries.

Despite the lack of numerical records, textual resources during the Cultural Revolution is abundant. Representative textual data include tabloids (Xiaobao 小报), big-character posters (Dazibao 大字报) and party-related office documents. Existing literature typically look topics regarding the Cultural Revolution through the lens of interpreting participants' memoirs and historical bibliography, such as official publications, attempting to gather complete events. However, systematic analysis is empirically deficient in these topics, especially (dis)similarity of ideological positions between the political elite and the Red Guards and whether they strived for political initiatives or offended each other by making political statements, inflammatory speeches, and slogans. From this angle, this paper also contributes to fill in the slack: we apply the novel keyword and phrase extraction technique, by inputing those textual records, to construct a dictionary that reflects the elite's and the Red Guards's corresponding political

stances, respectively. We cross-validate it by showing that the trained dictionary composed from keyword extraction has higher accuracy in predicting faction groups than that from the tf-idf weighted word-counting approach. Political positions extrapolated from trails of terms in the dictionaries offer us a tool to analyze the heterogeneity across and within both political groups. On top of it, we empirically find various factions among the elite and the Red Guards, as well as collusions between certain factions.

The remain of paper proceeds as follows. Section 2 introduces the historical background of the Red Guards and the associated mass movement. Section 3 details the textual data sources. Section 4 explains the estimation strategy and based on it, constructs dictionaries using the key extraction algorithm and estimates political positions of the elite and the Red Guards. Section 5 concludes.

2   Historical Background

The mass movement initiated by commoners during the Cultural Revolution set off as a comprehensive attack on the system of bureaucracy, where the major political power was firmly grasped by the elite (Walder 2015a). Yet, in retrospective, the movement started by the politically powerless mass should have not evolved to such large scale without the assistance from external schemers: nor without the tight supervision and the scrutinization from the elite. After the establishment of the People's Republic, for a period, Mao Zedong and his central leadership were dis-supportive to plans that involved forming associations and groups independent from the Party and condemned any unauthorized action by the mass to remove politicians from their office. However, later on, Mao's attitude became so shifted that he openly endorsed and authorized the mass movement and even encouraged students into participation, without legal or political restriction imposed. Since then, Mao encouraged urban high school student groups and college student groups to form the Red Guards organization. During the early stage of the Cultural Revolution (1966-1968), the Red Guard organization as a political weapon fought against the "class enemies" and the Capitalist roader from the CCP. Several students organized mass movements and quickly launched the "red terror" in the city (MacFarquhar 1997; MacFar-

quhar and Schoenhals 2006; Lieberthal 2004).[1] Reasons behind his reneging are worth further inspection. Table 1 lists the organizations and the leaders of factions in the Red Guards for two faction types: the Conservative and the Rebel. For the complete description, please see Appendix A.

Table 1: Major Organizations and the Leader of Factions in the Red Guards

Faction Types	Faction Organizations / Leaders
Conservative	Tsinghua University High School Red Guards (清华附中红卫兵), Beijing University High School Red Flag Battle Group (北大附中红旗战斗小组), Chinese People's University High School Red Guards (北京人大附中红卫兵), Western District Picket Corps (西纠), The United Action Committee (联动), Eastern District Picket Corps (东纠), Tan Lifu (谭力夫), First Headquarters (一司), Second Headquarters (二司).
Rebel	Nie Yuanzi (聂元梓), Kuai Dafu (蒯大富), Tan Houlan (谭厚兰), Third Headquarters (三司), Tsinghua Jinggangshan Regiment (清华大学井冈山兵团), Aeronautics Institute Red Flag (北航红旗), Geology Institute East Is Red (地院东方红), New Beida Commune (新北大公社), Beishida Jinggangshan Combat Brigade(北师大井冈山战斗兵团), Capital Red Guards Congress (首都大专院校红卫兵代表大会), Capital Middle School Red Guards Congress (首都中学红代会), Shanghai Workers' Revolutionary Rebels General Headquarters (上海市工人革命造反总司令部).

First of all, aims and scopes of "the Red Guards", for Mao and the Central Cultural Revolution Group (the CCRG 中央文化革命小组), could be shaped accordingly to fulfill their needs along with the evolvement of the movement.[2] Consequentially, divergent interests within the Red Guards were originated from strategic decisions made by Mao and the CCRG to ally with different organizations and factions. Narrowly speaking, the Red Guards refer to active high

---

[1]Between April 23 and October 27, 1967, Minister of Public Security Xie Fuzhi alone approved the convening of more than 100 large municipal-level mass rallies all over Beijing at which deposed senior members of the central and Beijing municipal government and party organizations were struggled. Thousands of lesser rallies, organized by a city district, a factory, or perhaps jointly by a group of universities, were convened in Beijing alone(MacFarquhar and Schoenhals 2006, p. 123).

[2]News of the CCRG's existence and the names of its leading members were first revealed to an inner-party audience in the highly classified Zhongfa [1966] 281 on May 28, 1966. And its membership evolved rapidly over the next year. The CCRG's key players were familiar figures whose loyalty to Mao was beyond question, and whose positions depended entirely on Mao's patronage (MacFarquhar and Schoenhals 2006, pp. 99 – 102; Walder 2015b, pp. 202 – 3)

school student groups that were based in different provinces in China from May to late 1966. Particularly, the Red Guards from colleges and were active at the late stage of the movement is labelled as "the Rebels".[3] Widely speaking, the Red Guards include all active (in the movement) high school to college student groups in 1966, as well as a proportion of newly joined high school students after September 1966. The widely-defined Red Guards population matches participating student groups at various stages of the Cultural Revolution.

Seeking to maximize the participation of the student population, rationales behind Mao and the CCRG's decision for relaxing the restriction (that only high school student groups were allowed to join the red guards organization) are three-fold. First, the bloodline theory, which had been widely accepted and uniformly supported in the high school, proved difficult to circulate among college groups. The difference is rooted in the heterogeneity in family backgrounds: the majorities of high school group members grew up as a second generation from a "red" family, while the college students had received strict identity check upon being admitted to the college (Xu 1999). This partially explains the failure to form of the large-size college Red Guards, at least before August 1966. Second, the CCRG realized that in the early stage, college student population had already been divided into two subgroups: the majority, which was supported by the Cultural Revolution working team (pro the higher level of the party apparatus), and the minority, which was opposed to the working team. When Mao aimed at the higher level of the party apparatus as the criticizing target, he also manipulated the minority from colleges in order to escalate the revolution to higher level officials in the bureaucratic system (Walder 2015b). Third, after observing the shift in the target of Mao, the high school Red Guards opposed the criticism of Liu Shaoqi (刘少奇), Deng Xiaoping (邓小平) and other core party members. Ironically, this challenged the authority of the CCRG and forced Mao to withdraw his endorsement and link to the CCRG. As a result, an strong opponent to the high school Red Guards, the rebels from the minority in college opposing the high level officials, joined the fierce competition in the Cultural Revolution. Afterward, rebels became Mao's main force to attack the party and the bureaucracy.

---

[3]In the capital, Beijing where Mao resided, the Red Guards from colleges proudly referred themselves as the "the old Red Guards". They were strict about and prioritized the identity. For those the rebels who failed the identity check for their membership to join the old Red Guards, the Red Guards represented the "reactionary bloodline theory" and thus, the rebels considered them as "the Conservatives" (Xu 1999).

With the permission of the CCRG, the Rebels united and established the Third Headquarters in September 1966. During the gathering, the CCRG formally publicized their support towards the Third Headquarters. At the same time, Zhou Enlai also conveyed Mao's supportive attitude and consent to the Third Headquarters' attack on the bureaucracy. The United Action Committee (Inter-school organization of high school Red Guards) was accused as counterrevolutionaries by the Minister of Public Security Xie Fuzhi. Hence, the Third Headquarters assailed the base of the United Action Committee with the permission of the CCRG (MacFarquhar and Schoenhals 2006, 183 – 84). However, the Rebels' successive attacks on the high school Red Guards made political elites who oppose the radicalization of mass movement anxious. In February 1967, many PLA marshals and the vice-premier had serious disputes with the CCRG, accusing their lack of intention to restrain the rebels' aggressive behavior, in the meetings of Central Caucus chaired by Zhou Enlai. This conflict was later christened as the "February Countercurrent". The motivations behind the "February Countercurrent" are twofold: 1) the mass movement led by the rebels to criticize Liu and Deng triggered the political elite' anxiety; 2) many political elites' descendants as the members of the United Action Committee, were arrested by Xie Fuzhi. In response to the opposition of these elites (the CCRG and the rebels named them as the Reactionary later), Mao decided to keep the rebels' dominance in the movement and grant the green light to mass gatherings launched by the rebels to criticize the Reactionary.[4]

Mao and the CCRG were amazed by the ability of the rebels to suppress the high school Red Guards and criticize political elites within the party. In the following year, the Rebels fell into endless internal fighting. During the battle, the rebels were keen to use force to gain the leadership in university campuses, national ministries, and government agencies. Until the end of July 1968, Mao was extremely disappointed with his young disciples and then asked them to cede the campus's leadership. Since then, the main stage of the Red Guard movement ended.

Based on fore-mentioned background, it would be enlightening to systematically examine internal conflicts within the Red Guards and the similarity with the political elite in terms of political positions under the Cultural Revolution. In this paper, we combine TextRank algo-

---

[4]On 22 April 1967, Mao Zedong ordered a release of several members of The United Action Committee who are the children of political elites after arresting them for three months, along with requiring them "behave themselves" (MacFarquhar and Schoenhals 2006, p. 198).

rithm to extract key Chinese phrases, and Wordfish (Poisson scaling model) to estimate the differences of spatial positions using the extracted textual data covering the revolution. Prior theoretical research indicates that the political elites and the Red Guards lost control of the social movement, and it evolved into sheer verbal violence and physical skirmishes across all provinces in the end (MacFarquhar and Schoenhals 2006). The first question of our interest is the overall ideological dissimilarity between the political elite and the Red Guards: if overall, the political elite placed themselves differently vis-à-vis the Red Guards along the spectrum of ideological positions. Afterwards, we study, at the time interval of each major incident, the degree of disagreement among different factions in the Red Guards and investigate its spatial positioning in correspond to the political elite.

3   Data

This paper draws on the archive from Chinese Cultural Revolution Database (1966-1976) managed by the Chinese University of Hong Kong under the database named Databases for the History of Contemporary Chinese Political Movements. The main themed data sources utilized in analysis are the historical materials of the mass movement from 1966 to 1968, where relevant textual materials about revolutions of the Red Guards and related official speeches are classified by speaker and by year-month-date, with original posters or documents identified. On the one hand, the official speeches made by the elite (central officials) include meeting minutes, internal comments and decisions about issues related to Cultural Revolutions and circulated within the Chinese Communist Party (the CCP). In particular, these types of texts take forms of different formality: some are formal statements, such as then classified documents, and some are recorded on informal occasions, such as in mass gatherings. Depending on the scale of attendees, meetings at different occasions are of various purposes and carry different sentiments and yet, generally, large gatherings target the masses while meetings with fewer attendees aim to produce internal comments and decisions on leaders of the Red Guards, and discussion about the mass movement. Regardless of the size of meetings, we deem that these textual records reflect the position of the elite towards the social movement. On the other side, the mass (the Red Guards) typically uses columns, self-print tabloids (Xiaobao) and big-characters posters

(Dazibao) to express political stances and opinions of their own. This drastic difference in the means of expressing opinion between the two groups (the elite and the mass) allows us to analyze their speeches and textual records accordingly based on the sources of publication.[5]

Table 2 describes the document records used in this paper. The first column shows two types of documents for both political groups. the political elite and the Red Guards. Column 2 demonstrates the different faction labels in each political group. Three categories, the Radical, the Swing and the Reactionary are seen among the elite and two categories, the Rebel and the Conservative, belong to the Red Guards. Column 3, 4 and 5 shows the number of documents (corpus), the total number of tokens extracted and the average number of token extracted from per document, respectively in each category.

Table 2: Description of the Document Types in the Corpus

Types of Documents	Labels of Factions	The Number of Documents	The Total Number of Tokens	The Average Number of Tokens
The Political Elite	Radical	261	483,726	1853.356
	Swing	117	351,497	3004.248
	Reactionary	93	192,430	2069.140
The Red Guards	Rebel	247	1,075,177	4352.943
	Conservative	98	162,351	1656.643

4 Estimation Strategy

Textual scaling methods, such as in Wordfish (Slapin and Proksch 2008) and Wordscore (Laver, Benoit, and Garry 2003; Lowe 2008; Will et al. 2011; Lowe and Benoit 2013) are becoming main-

---

[5]The CCP's propaganda system proposed a measure of restrictive guiding concept behind official publications, initially during Yan'an Period (Gao 2019). The guiding concept placed a great emphasis on the principle of news confidentiality and access to classified reading. Newspapers edited by the CCP were required to reflect the Party's political principle, officially called "the principle of Party spirit" and these newspapers only became accessible to party members and yet, avoided the its exposure to the masses. Moreover, the access to information was also hierarchic to party members: the higher the ranking, the fewer the restrictions on reading materials, especially in terms of important and confidential information (Brady 2008; Gao 2019, 399-400). On the contrary, representative publications edited by the masses, tabloids and big-character posters, were guided by a different concept. Interestingly, these types of publications were developed in answering the need of the mass movement by Mao Zedong and his allies; yet they gradually lost their autonomy.

stream in text-as-data community. These applications have been applied to capture spatial ideological positions of political parties and comprehend political phenomena in political science. Typically, Wordfish (or Wordfish-like) methods (Slapin and Proksch 2008; Lo, Proksch, and Slapin 2014b; Lauderdale and Herzog 2016), through scaling the positions of parties in continuous ideological space, reliably facilitate in studying latent dimensions of political texts and test political theories, e.g., intra-party politics, polarization, and representation (e.g. Bäck and Debus 2016; Catalinac 2017; Kim et al. 2018; Ishima 2020; Proksch and Slapin 2010).

Subsequent research and discussion has extended to incorporate cutting edge techniques in natural language processing by considering semantic of aspect such as Glavas, Nanni, and Ponzetto (2017), Nanni and Glavaš (2019) and Handler et al. (2016). A similar approach to ours is also proposed by Nanni and Glavaš (2019), where they develop a semantically-aware phrase extraction technique (SemScale) on top of the Wordfish algorithm and it is designed for the European speech context. The novel algorithm proposed by Nanni and Glavaš (2019) shows validated evidence by considering semantic information performs better in measuring latent dimension from the European parliament speeches and suitable for cross-lingual political text scaling (Glavas, Nanni, and Ponzetto 2017, p. 692). Therefore, the estimates of original Wordfish may not provide much explicit information if a mixture of part of speech tagging are all included. Because bag of phrase based word are less ambiguous than unigrams, phrase based Wordfish model can unveil better semantic themes and generate more interpretable context, without considering a lengthy process to validate interpretation of the terms in each different incident in our paper. As a result, the Wordfish algorithm applied in our paper is generated from the corpus reduced only to keywords and phrases setting by applying extraction techniques on specific tagging (such as noun, verb and adjective), which can efficiently summarize the multitude of topics and improve semantic coherence. While original Wordfish applied in parliamentary speech or textual election manifest can provide useful information about the preferences of elected party members in most cases (Bäck and Debus 2016; Catalinac 2017; Kim et al. 2018; Ishima 2020; Proksch and Slapin 2010), such measures not only count solely word as unigram but also loosely capture semantic structure .

How to accurately measure the reflected positions and attitudes towards the Cultural Rev-

olution is of significance in this paper and in the text-as-data community. From this aspect, the advantages of using textual data over the traditional approach that uses roll-call votes to estimate average positions of a certain group are quite noticeable. First, generally the records of roll-call votes for China studies can be seldom found and the best alternative for position estimations is the text-as-data approach, which analyzes documents of subjective opinions and expression. Second, the traditional approach focuses on the average positions of a certain group and does not stress on the intra-group difference, such as within-party polarizations. Interpreting text as a special form of data permits us to track the differential of wordings and phrasings across heterogeneous groups or even within a seemingly homogeneous group and provides rich textual evidence for thorough ideological interpretation.

In this paper, we seek to find an algorithm that serves two main functions: i) extract key Chinese words and phrases (slogans); ii) estimate the differences of spatial positions using the extracted textual data. To this end, we develop a new strain of Wordfish algorithm that takes advantage of the facilitation of TextRank algorithm to extract Chinese characters and phrases.[6] It not only extracts sole key characters (as in Wordfish) but also considers semantic structures of Chinese characters to extract key phrases. The consideration of semantic structures of our proposed algorithm mitigates the biasedness and errors that are often found in the original Wordfish caused by counting sole words and neglecting the linguistic contexts. For instance, one of the key phrases in our study is "the Cultural Revolution" and Wordfish simply decomposes the phrase into two parts, "Cultural" and "Revolution", and counting them separately for further analysis.

4.1   TextRank: Automated Keyword Extraction

Measuring spatial political positions using textual data is a fundamental aspect to comprehend potential political phenomena. In this sense, estimating the positions of political actors is key to understanding factionalism under Moa's China. Specifically, in this paper, the first stage applies the TextRank algorithm that is commonly used in automated keywords extraction techniques of Natural Language Processing to create a subset of words and phrases that represents the most

---

[6]Wordfish was originally developed by Slapin and Proksch (2008) and implanted in the quanteda package (Benoit et al., 2018).

important or relevant information by calculating the weight for each token within the document (e.g., Barrios et al. 2016; Bałcerzak, Jaworski, and Wierzbicki 2014; Figueroa and Chen 2014; Wan and Xiao 2008). For the preparation of textual data for keyword extraction, we apply a pre-trained model from UDpipe, with parts of speech (POS) tagging and linguistic dependencies, to tokenize the collected textual data parsing as introduced in the previous section (Straka, Hajič, and Straková 2016; Straka and Straková 2017). The UDpipe is one of the popular NLP toolkits in open source community and produces annotations on not only the syntactic structure of a sentence based on CONLL-U format but also POS tagging — verb, noun, adjective, and so on—of each word in a sentence in Chinese. In our study, the latest pre-trained Universal Dependencies 2.5 model tokenizes Simplified Chinese documents and based on linguistic usage patterns of Chinese language, we reduce the tokens only to reserve specific tagging, i.e. noun, verb and adjective.[7] In particular, the tagging of linguistic features can be utilized to extract important phrases and pair n-grams as noun phrases via keyword extraction techniques, such as TextRank in this paper.

TextRank is an unsupervised graph-based algorithm for automated summarization of text containing keywords and initially introduced by Mihalcea and Tarau (2004) by converting Google's PageRank algorithm into feature selection to extract certain phrases. The ultimate goal of a keyword extraction is to effectively select a set of representative terms from the text that best describe and summarize the original document. Therefore, these representative terms are the near-minimum information set that can be utilized for text classification, document collection and information summarization, from which domain-specific dictionaries are constructed. TextRank reads original text as input and ranks a sequence of single or multiple lexical units extracted from the input, and these are stored as vertices in the text graph. Moreover, any relation between any pair of lexical units represents an informative relevance and can be potentially in-

---

[7]UDPipe is capable of performing excellent parts of speech tagging such as Universal POS tags and XPOS, morphological features, and syntactic dependencies, and syntactic analysis of CoNLL-U input based on Universal Dependencies (UD) framework, which is is cross-linguistically consistent treebank annotation of morphology and syntax for many languages, see https://ufal.mff.cuni.cz/udpipe/2. Trained models are provided for nearly all UD treebanks at this website, see https://universaldependencies.org. The pre-trained model for NLP task in this paper is UD Chinese-GSDSimp (trained and contributed by Qi and Yasuoka 2020) and can be found at the authors' Github repository, at https://github.com/yl17124/UD_Chinese-GSDSimp. For the natural language tasks, we use third-party R CRAN package implanted by Wijffels and BNOSAC (2020b) to process Simplified Chinese text. Its manual application and corresponding documents can be found at https://bnosac.github.io/udpipe/en/.

Figure 1: The Flowchart of the TextRank Algorithm

Step 1 :Merge all the document records by the time interval for each major political incidents during the Cultural Revolution.

Step 2 :The UDpipe/ UDpipe 2.5 language model annotates text based on CONLL-U format by sentence level and only keeps verb, noun, adjective.

Step 3 : Find vector representation (word embeddings) for each and every sentence/ words.

Articles/ Docs → (concate) → Text → (split) → Tokens with Sentence ID → Vector Representation → Similarity Matrix → Graph → Keyword Ranking → Keywords Phrases → Remove Repetitive Phrases → Phrase Frequency Matrix

Step 5 : Co-occurrence (similarities) between sentence vectors is then calculated and stored in a matrix.

Step 4 :a pair of lexical units are regarded as potentially linked if they co-occur within a window of 5 n-grams, and their representative vertices are linked in the text graph.

Step 6 : Compute similarity distance metrics between each phrase in each political incident and only keep one word/phrase if the similarity is higher than 0.85.

Step 7: Casting data into a word (phrases or keywords) frequency matrix

troduced to bridge the pair of vertices. Similar to Mihalcea and Tarau (2004), we particularly define the property of co-occurrence as such relation between two vertices: a pair of lexical units are regarded as potentially linked if they co-occur within a window of maximum n-grams (n is set to be 5 in our study), and their representative vertices are linked in the text graph. The importance of a vertex is determined by two factors, the number of (undirected) links it has and the weight of connected vertices. Mathematically, we can calculate the value of importance of a vertex i by

$$V(i) = (1 - d) + d \sum_{V(j) \in Link(i)} \frac{p_{ji}}{\sum_{V(l) \in Link(i)} p_{jl}} V(j), \tag{1}$$

where $V(i)$ is the value of importance of vertex i in the text, d represents the dampening factor, $Link(i)$ is the sample of vertices that share links with vertex i, p_{ji} measures the (absolute) weight

vertex j carries with respect to vertex i and therefore, $\frac{p_{ji}}{\sum_{V(l) \in Link(i)} p_{jl}}$ is the relative weight of vertex j with respect to i. From the above formula, the value of importance of vertex i in the text is the sum of weighted value of importances of all vertices that are linked to i. Moreover, the weight p_{ji} is positively related to the frequency at which the vertex j are linked to i.[8] Figure 1 illustrates a break-down of steps describing the flow of the TextRank algorithm utilized in our paper.

The structural identifiability and grammatical correctness of textual data from tabloids, big-character posters and official documents considered in our study enable us to apply TextRank to process these data. TextRank considers linguistic dependence among the tokened texts and excludes the texts with lower degree of co-occurrence and that are distance to the target lexical unit in a sentence. First, we merge all the document records by the time interval for each major political incidents as specified by (Walder 2009, see Appendix B). Figure 2 illustrates top 15 keywords/phases extracted from all documents at the point when each major incident happened. For computational manageability, then we only keep the top 25% quantile of relevant n-grams words from each incident, as the most representative terms. Stop words (the most common but the irrelevant) and some single and meaningless characters are automatically programmed to be excluded as they are too common to carry incident-specific information in Chinese.[9] At the same time, we compute similarity distance metrics between each word/ phrase in each political incident and keep one word/phrase if the similarity is higher than 0.85. This is due to that some of n-grams words or phrases are repeatedly constructed by TextRank but mean similarly. Finally, we use these incident-specific dictionary sets to build each phrase-frequency matrix for later analysis.[10]

---

[8]Under this specification of the value of important, we assume that the link between any two vertices are undirected. Mihalcea and Tarau (2004) using simulation to study the speed of convergency and accuracy during iteration, and conclude that as the connectivity of the graph increases, convergency is usually achieved after fewer iterations. Also, the directed and undirected graphs practically overlap and accuracy are reasonably similar.

[9]To find relevant keywords from each document records, we use Textrank package (version 0.3.1) implemented by Wijffels and BNOSAC (2020a) in R to summarize representative words from each text documents. The library allows us to find essential phrases and keywords by pairing them with part of speech tagging reduced to nouns, verbs, and adjectives only.

[10]We use stringdistmatrix() function in stringdist package (implanted by Loo et al. 2020) to computes the string distance matrix.

Figure 2: Top 15 Keywords Identified by TextRank at Each Major Incident
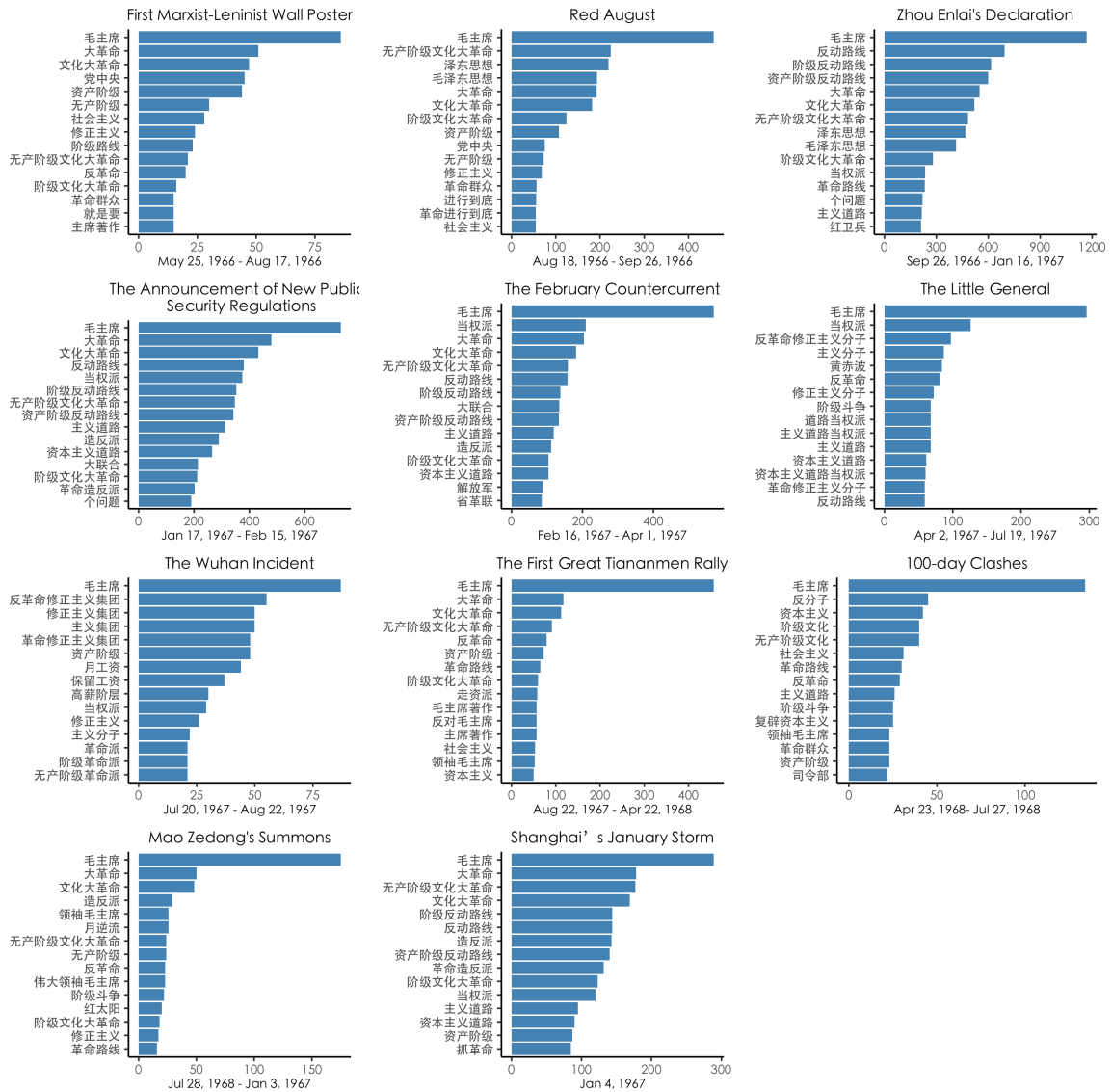
4.2   Estimating Positioning of Factions

In the second stage, the algorithm estimates positions of each faction based on the Poisson scaling model proposed by Slapin and Proksch (2008). The model produces the estimated positions by analyzing frequencies of extracted keywords and phrases from dictionaries constructed in the first stage. For simplicity, the distribution of the frequencies is assumed to be a Poisson process, as the Poisson process only has a single parameter, θ, for the model to estimate. θ represents both the mean and the variance of the distribution. The model takes the following functional form:

$$y_{jm} \sim Poisson(\theta_{jm})$$

$$\theta_{jm} = exp(a_m + b_j + \beta_j * w_i^m),$$

where y_{jm} is the frequency at which word j appears in document m and is stored at (j, m) element in the matrix word/phrase-document matrix Υ. θ_{jm} is the mean and the variance of the Poisson distribution. a_m is document m fixed effects, representing the length of the document. b_j is word/phrase fixed effects, representing the frequency at which word/phrase j appears across all documents. β_j represents word/phrase-specific weight on word/phrase j. ω_i^m is faction i's ideological position to which the document m relates to.[11] To complete the model, Bayesian formulation is introduced by the specification of prior multi-normal distribution for model parameters and w_i^m, for any document m:

$$\begin{pmatrix} p(a_m) \\ p(b_j) \\ p(\beta_j) \\ p(w_i^m) \end{pmatrix} \sim IIN \left(\begin{pmatrix} \mu_a \\ \mu_b \\ \mu_\beta \\ \mu_w \end{pmatrix}, \begin{pmatrix} \sigma_a^2 & 0 & 0 & 0 \\ 0 & \sigma_b^2 & 0 & 0 \\ 0 & 0 & \sigma_\beta^2 & 0 \\ 0 & 0 & 0 & \sigma_w^2 \end{pmatrix} \right). \tag{2}$$

---

[11]Unlike the original Wordfish model which assumes each document per actor, in the latest application of Imai, Lo, and Olmsted, the generalized Wordfish model enables multiple documents per actor for ideal point estimation. Given these notations, we can proceed to apply Imai, Lo, and Olmsted's application proposed to estimate an IRT model with phrase count on top of the Wordfish poisson scaling model developed by Slapin and Proksch (2008).

Further their joint posterior distribution is formulated as

$$
\begin{aligned}
p(\{a_m\}_{m=1}^M, \{b_j\}_{j=1}^J, & \{\beta_j\}_{j=1}^J, \{w_i^m\}_{i=1}^I | \Upsilon) \\
& \propto \left[\prod_{j=1}^J \left\{ \prod_{m=1}^M p(y_{jm}|a_m, \tilde{\beta}_j, w_i^m) p(a_m) \right\} p(\tilde{\beta}_j) \right] \\
& \times \prod_{i=1}^I p(w_i^m),
\end{aligned}
\tag{3}
$$

where $\tilde{\beta}_j = (b_j, \beta_j)$ and $p(\tilde{\beta}_j)$ denotes the joint probability density distribution of (b_j, β_j). [12]

5 Findings

5.1 Analysis with Pooled Documents

This section analyzes and summaries empirical findings obtained from our estimation strategy, as explained above. The first question of our interest is the overall ideological dissimilarity between the political elite and the Red Guards throughout the Cultural revolution: if overall, the political elite placed themselves differently vis-à-vis the Red Guards along the spectrum of ideological positions. Afterwards, we study, at the time interval of each major incident, the degree of disagreement among different factions in the Red Guards. We re-estimate the Red-Guards relevant documents by faction and separately for each incident, in order to compare the ideological position of factions among the Red Guards. Finally, by each incident and by faction, we juxtapose ideological positions of the political elite factions with that of the Red Guards factions to investigate the degree of disagreement among them.

Figure 3 plots smoothed density distributions of estimated outcomes with all documents pooled together for each of the Red Guards factions (left subfigures) and the political elite faction (right subfigures), respectively. Left subfigures illustrate density distributions of estimated ideological positions for three factions in the political elite, namely, the Radical (first left), the Reactionary (second left) and the Swing (third left). It is rather noticeable that overall the Radical and the Reactionary elite have similar distributions of ideological positions, and the

---

[12]See fuller information in Imai, Lo, and Olmsted (2006, pp. 649 – 650) and their supplementary appendix E (Imai, Lo, and Olmsted 2006, pp. 57 – 59).

Figure 3: Smoothed Density Distributions of Estimated Positions for the Political Elites (left subfigures) and the Red Guards (right subfigures)
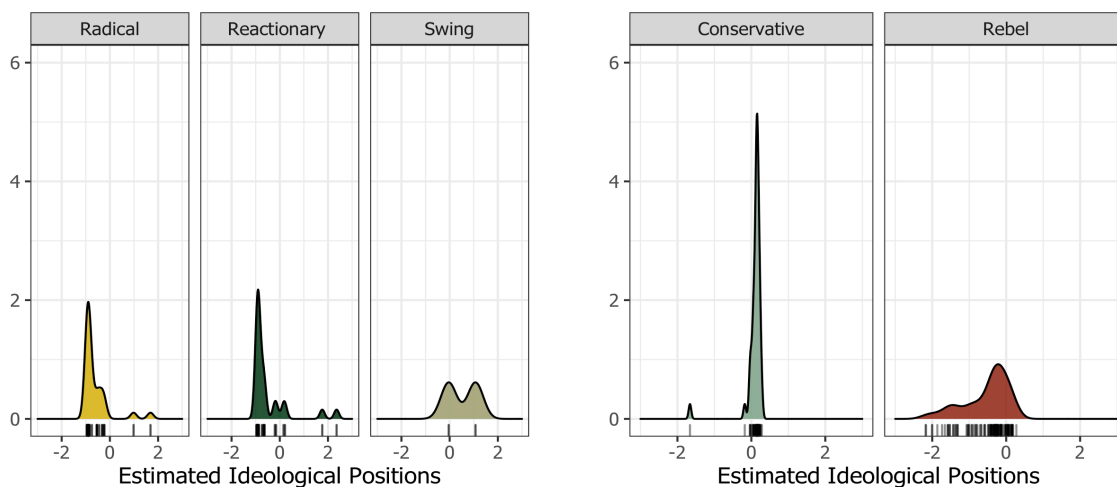
majority of the area is next to the position of the highest spike around value -1 on the spectrum (x-axis), reflecting the internal similarity in terms of ideological positions. Yet, we still observe other peaks that are rather distant from the highest spike. For example, for the Reactionary, two spikes are positioned around value 0 and other two around value 2. It means that despite of the ideological uniformity within the Radical or the Reactionary elite faction majority (as there exists a major center, the highest spike, for both distributions for both elite factions), inside each elite faction, there still exists subgroups holding different and rather contrasting ideologies, demonstrating by the long distance between the left and the right spikes. Further, if we stacked both distributions together, it is highly likely that we would record a distribution with a single peak centering around value -1. This shows the internal ideological similarity within the elite.[13] The Swing elite are a special faction consisting only one single politician, Zhou Enlai. As its name suggests, the ideological distribution of the Swing has two peaks. The double peakedness reflects the potential swing in terms of his ideological from time to time, and the height of both peaks is roughly the same, suggesting Zhou Enlai does not reside on a single position, instead his ideological positions swing from back and forth from around both peaks.

Right subfigures demonstrates density distribution of estimated ideological positions for two factions in the Red Guards, the Conservative (first right) and the Rebel (second right).

---

[13] Adding the Swing into the distribution wouldn't affect our results, as the Swing only consists of one single member.

Unlike the elite (Radical and the Reactionary), the estimated ideological positions of both factions in the Red Guards share less common shape, as the Conservative are rather united while the Rebel are rather dispersed along the spectrum. Particularly, the Conservative's density distribution is centered around its peak at value 0.1 and the majority of area is distributed around the center. The Rebel's distribution has one peak around value -0.2 and yet the area of the distribution is widely dispersed. The dispersion in terms of the ideological positions can be attributed to the high degree of internal disagreement in ideology within the Rebel. If we stacked the two distributions, it would be likely to produce the density distribution with two major peaks (one for the Conservative and one for the Rebel), reflecting the internally dispersed Red Guards ideology.

The most contrasting difference in the ideological distribution between both groups overall is that the Red Guards are more dispersed than the elite. In particular, the elite share common ideological distribution (in terms of center and shape), while in the Red Guards, the Conservative are ideologically united while the Rebel are rather dispersed (centering at different positions). The potential explanation might well reside with that overall the political elite factions have more unified ideology (at least reflected in textual statements), whereas the ideology of the Red Guards can be highly varied across individual factions.

To further illustrate the internal heterogeneity (disagreement) in the Red Guards in terms of ideological positions between the two different factions, Figure 4 plots the positions for faction organizations (or members) from the Rebel and the Conservative, where individual red dots represent organizations that are categorized in the Rebel and green dots represent organizations in the Conservative. The estimated positions for the Rebel (such as Nie Yuanzi and Kuai Dafu) are more widely spread, with the majority spanning from around -2.3 to slightly higher than 0, while the estimated positions for the Conservative (such as Western District Picket Corps 西纠, the United Action Committee 联动, and Beijing University High School Red Flag Battle Group 北大附中紅旗戰鬥小組, etc) are more concentrated from 0 to 0.5. This is in alignment with what we observed from right subfigures in Figure 3. From another perspective, the ideological dispersion can then be interpreted as the stability of allying bonds among different organiza-

Figure 4: Estimated Positions for Individual Members and Organizations in the Red Guards
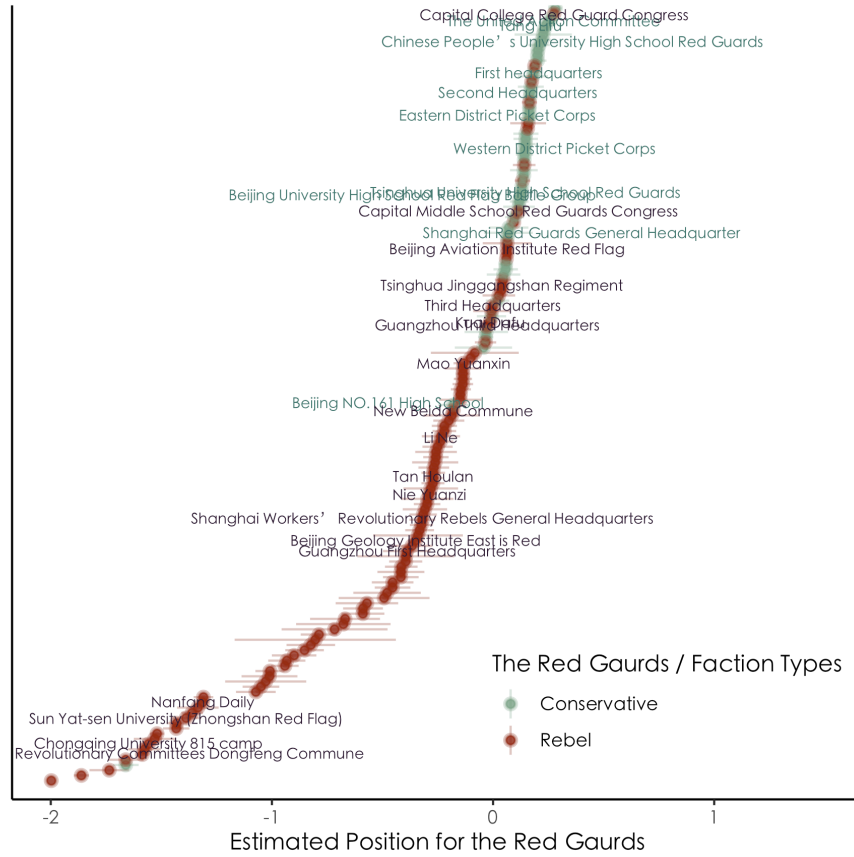

tions even in the same faction. Clearly, such bonds are rather loose in the Rebel and slightly firmer in the Conservative, resulting in a more unstable bonds in the Red Guards than the elite in general.

5.2   Incident-based Analysis

In this section, we select three major and representative political incidents occurring during the peak of the Red Guards movement. For each of the incident, we estimate ideological positions for members in factions from the elite and the Red Guards using documents recorded during the relevant event. This allows us to track dynamically the heterogeneity and the degree of disagreement in ideology seen from textual documents among factions over different events occurring at different time. In particular, these three incidents include: 1) Zhou Enlai's declaration (September 26, 1966 - January 16, 1967), when Zhou declared the support for the Rebel on behalf of Mao, 2) the announcement of new public security regulations (January 17, 1967 -

20

February 15, 1967), when the cause of the Conservative Red Guards was adjudicated as unauthorized, and 3) February Countercurrent (February 16, 1967 - April 1, 196), that is, criticism from the Reactionary elite against the CCRG after the meeting of the Central Caucus ignited Mao's anger that he directed the Rebel Red Guards to attack the Reactionary (see Appendix B). Outcomes are recorded and plotted in Figure 5, 6, and 7. At the first glimpse, there exists a significant categorization throughout incidents, despite the categorization varies across incidents. This agrees with the theory proposed by Walder (2006a, pp. 712 − 713) regarding dynamic factionalism, that is, in short, the Red Guards dynamically allied with different organizations to defend their social choice in the status quo, causing conflicts and disputes against different subgroups across time. Appendix C lists estimated ideological positions for each factions from the Red Guards and the political elite in each major historical incidents during the Cultural Revolution.

Figure 5: Estimated Positions for Zhou Enlai's Declaration
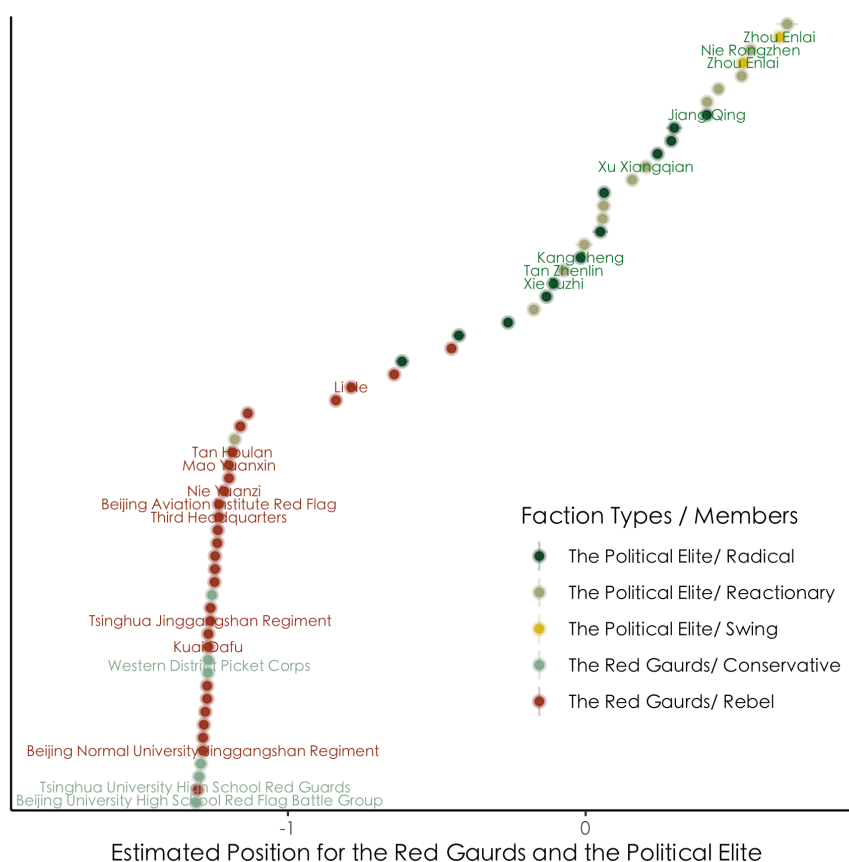
In the first event, Zhou Enlai attended the gathering organized by the Rebel named "Third

Headquarters (三司)" on behalf of Mao in Beijing, where he conveyed Mao's support and endorsement to the Rebel. This represents the official acknowledgment from Mao and the CCRG of the Rebel. Figure 5 depicts the estimated ideological positions for each members from each major faction. As is shown in the figure, during this event, there exists noticeable categorization of two groups: the Red Guards (the Rebel and the Conservative) occupy the spectrum to the left of value 0 on the x-axis, while the majority of the elite (the Radical, the Reactionary and the Swing) take up the spectrum to the right of 0. Although it is anticipated that the elite stay unified in terms of ideology (as is illustrated from the previous section), the overlap between the two factions in the Red Guards can be reasonably interpreted further. After Zhou conveyed Mao's favor to the Rebel, as their opponents, the Conservative lost Mao's (and the CCRG's) political support and altered their ideology to conform with the Rebel, in seek for retrieving Mao's approval. Further, positions of some organizations/members from the Radical elite overlap with the Red Guards (particularly, the Rebel), showing that the impact of Mao's support shifted not only the Conservative Red Guards, but also some from the elite towards the Rebel in terms of the ideology.

The second incident, that is, the minister of Public Security, Xie Fuzhi announced new public security regulations and designated the Conservatives as counterrevolutionaries. In the meanwhile, Third Headquarters (the Rebel) dominated the development of the movement and police started seeking and arresting members from the Conservative. Figure 6 illustrates estimated ideological position during this incident. As is clearly shown, the number of the Conservatives members shrinks due to the arrest. The elite and the Rebel form two contrasting groups, each taking up one side of the spectrum. Yet, the two remaining Conservative organizations/members have significant disagreement, one in alignment with the elite and other with the Rebel. More importantly, no the elite had overlap with the Rebel in terms of positions, implying the elite's concerns and disagreement over the current situation partial towards the Rebel Red Guards. Particularly, the position of the minister of Public Security, Xie Fuzhi, is the closest to the Rebel among all the elite, validating our estimation outcomes that Xie's speeches were indeed partial towards the Rebel.

February Countercurrent is the third incident we seek to explore and discuss. This incident

Figure 6: Estimated Positions for the Announcement of New Public Security Regulations
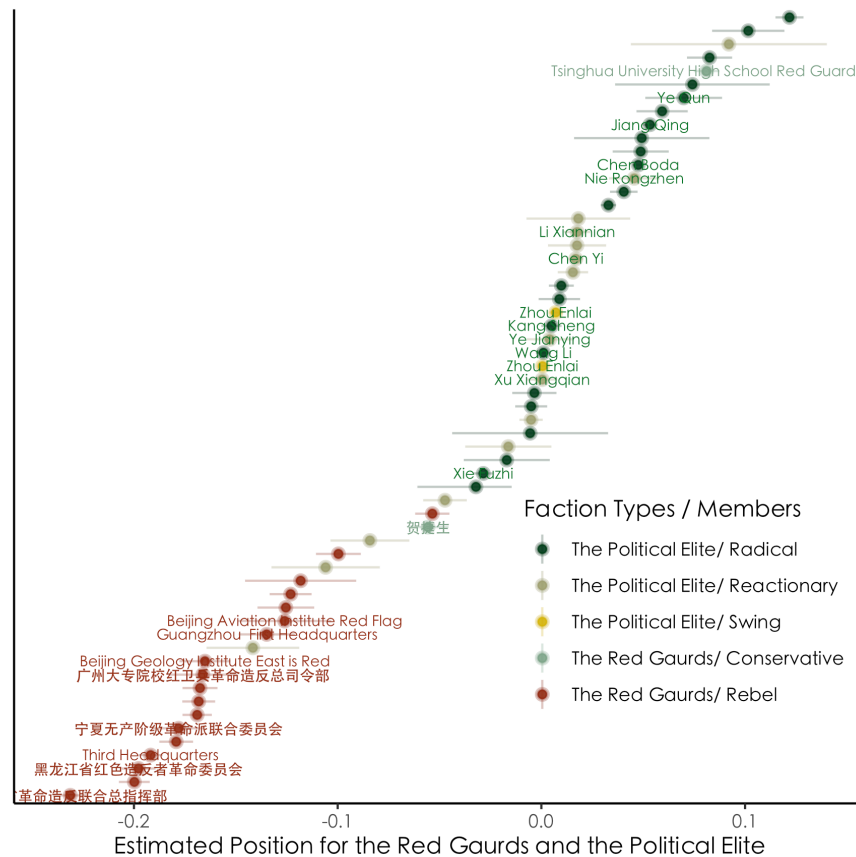
was triggered by questions and resentment made by leaders of the Reactionary elite in the meeting of the Central Caucus chaired by Zhou Enlai (周恩來). The Reactionary leaders, including PLA Marshals Chen Yi and Ye Jianying, etc, attributed the attack by the Rebel on the CCP central leadership to the CCRG's indulgent stance towards the Rebel. Note that, several leaders of the Conservative, Marshals Chen Yi's son included, were arrested in February during the previous incident relevant to Xie Fuzhi. Arrests of some second "red" family generation also intensified the Reactionary's resentment. As a result, after the meeting, Mao felt the need and determined to initiate attacks on a number of PLA Marshals. MacFarquhar and Schoenhals (2006) argues that at the end of this incident, Mao had no competent rivalry in the Party. From Figure 7, unlike the previous two incident we examined, the elite and the Red Guards move closer to each other, forming some overlaps. This implies decreasing disagreement in ideology extracted from textual documents between them, particularly for the Radical and the Rebel, after the incident.

23

Figure 7: Estimated Positions for February Countercurrent
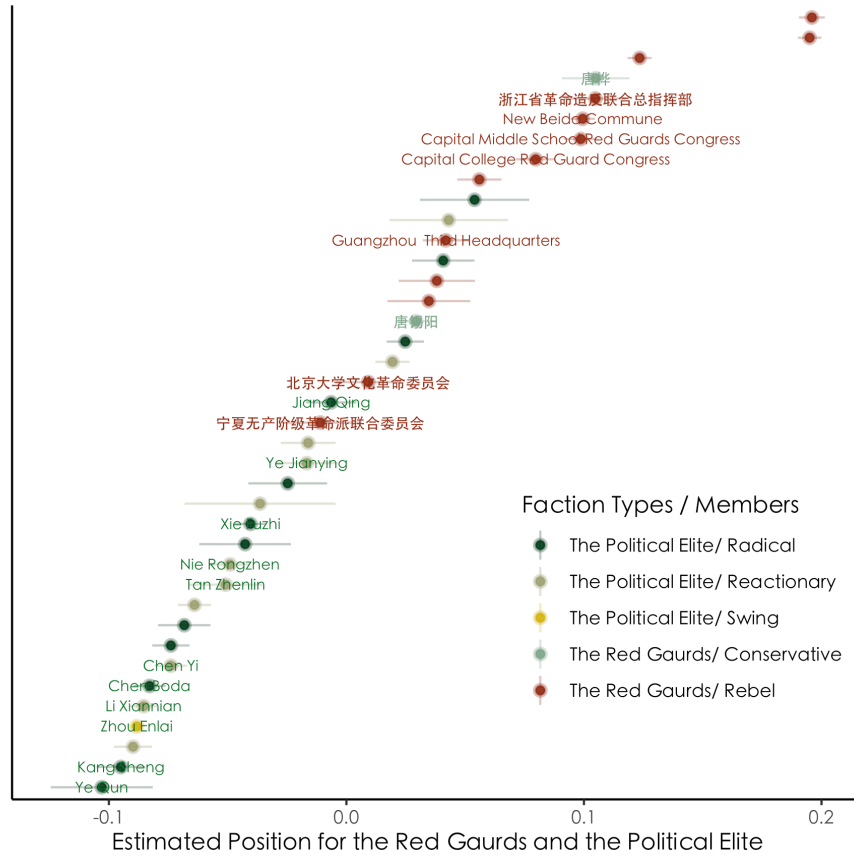
To sum up, these evidences illustrate the dynamic ideological positioning, as is reflected from the textual statements, of the elite and the Red Guards, was evolving across time and based on different events. Moreover, the political elite and the Red Guards with similar political views are inclined to echo and support each other when making statements. Naturally, the shift in political atmosphere is the dominating factor that causes leaders of the mass movement to adjust their ideological positions when making statements.

6   Conclusion

Accurate estimation of ideological positions is the primary and fundamental step towards the understanding of factionalism and cleavage in the Red Guards and the political elite. Yet, appropriate numerical data were seldom recorded and are rather scare at that point of time in China. This paper proposes a novel approach that analyzes textual data in Chinese, by utiliz-

ing the Red Guards' self-printed propaganda, such as big-character posters (Dazibao 大字报) and tabloids (Xiaobao 小报), and the elite's official statements and inflammatory speeches. The approach is implemented on top of some of widely applied Wordfish-based or -like scaling method and yet, TextRank algorithm is augmented in front of the scaling method to extract not only keywords but more reasonably key phrases. In addition, the latest version of Udpipe is introduced to facilitate the key extraction based on Chinese speech patterns and structures. Different from existing Wordfish algorithms, the scaling method (Poisson scaling model) in this paper only scales the extracted keywords and phrases. Two major advantages arise from our method. First, we augment TextRank so that we only care about keywords and phrase in each document and noisy terms have been filtered out before ideological position estimation. This is likely to mitigate estimation error due to less presence of the noise. Second, different from many languages, Chinese speeches are frequently phrase-based, rather than single word- (character-) based. TextRank is a graph-based algorithm that serves our purpose to interpret Chinese by extracting key phrases from texts.

Estimation outcomes of ideological positions demonstrate that the majority of the Red Guards factions and the elite factions are contrasting groups for most of time. This seems sensible and matches many evidences from previous literature (Walder and Su 2003; Wang 2019; Su 2006; Walder and Lu 2017; Walder 2016). However, the position of these individual factions dynamically varied according to the development of incidents and the shift of the political atmosphere. For instance, when the Rebel Red Guards received political acknowledgement from Mao, their rivalry, the Conservative Red Guards, conformed to the Rebel's ideological positions. This generally in alignment with Walder (2004), Walder (2006a), Walder (2006b), and Walder (2009). Moreover, when the Conservative Red Guards were deemed as counterrevolutionaries and under arrest, we indeed record less factions from the Conservative. These features indicate that application of our algorithm on Chinese textual data, such as statements from the Red Guards movement, is reasonable, and estimation outcomes are indeed reliable.

References

Andreas, Joel (2009). Rise of the Red Engineers: The Cultural Revolution and the Origins of China's New Class. Stanford: Stanford University Press.

Bäck, Hanna and Marc Debus (2016). Political Parties, Parliaments and Legislative Speechmaking. Palgrave Macmillan.

Bałcerzak, Bartomiej, Wojciech Jaworski, and Adam Wierzbicki (2014). "Application of Textrank Algorithm for Credibility Assessment." In: Proceedings - 2014 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Workshops, WI-IAT 2014.

Barrios, Federico et al. (2016). "Variations of the Similarity Function of TextRank for Automated Summarization." In: arXiv:1602.03606v1.

Benoit, Kenneth et al. (2018). "quanteda: An R package for the quantitative analysis of textual data." In: Journal of Open Source Software 3.30, p. 774.

Brady, Anne-Marie (2008). Marketing Dictatorship: Propaganda and Thought Work in Contemporary China. Lanham, Md: Rowman Littlefield.

Catalinac, Amy (2017). "Positioning under Alternative Electoral Systems: Evidence from Japanese Candidate Election Manifestos." In: American Political Science Review 112.1, pp. 31 – 48.

Dikötter, Frank (2016). The Cultural Revolution: A People's History, 1962 – 1976. London and New York: Bloomsbury Publishing.

Figueroa, Gerardo and Yi Shin Chen (2014). "Collaborative Ranking between Supervised and Unsupervised Approaches for Keyphrase Extraction." In: Proceedings of the 26th Conference on Computational Linguistics and Speech Processing, ROCLING 2014, pp. 110 – 124.

Gao, Hua (2019). How the Red Sun Rose: The Origins and Development of the Yan'an Rectification Movement, 1930 – 1945. Hong Kong: The Chinese University of Hong Kong Press.

Glavas, Goran, Federico Nanni, and Simone Paolo Ponzetto (2017). "Unsupervised Cross-Lingual Scaling of Political Texts." In: 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017 - Proceedings of Conference 2, pp. 688 – 693.

Handler, Abram et al. (2016). "Bag of What? Simple Noun Phrase Extraction for Text Analysis." In: e Proceedings of the Workshop on Natural Language Processing and Computational Social Science at the 2016 Conference on Empirical Methods in Natural Language Processing.

Imai, Kosuke, James Lo, and Jonathan Olmsted (2006). "Supplementary Appendix for " Fast Estimation of Ideal Points with Massive Data " Variational Inference for the Standard Ideal Point Model." In: 83, pp. 49 – 62.

— (2016). "Fast Estimation of Ideal Points with Massive Data." In: American Political Science Review 110.4, pp. 631 – 656.

Ishima, Hideo (2020). "How Electoral Reform Alters Legislative Speech: Evidence from the Parliament of Victoria, Australia 1992 – 2017." In: Electoral Studies 67.June, p. 102192.

Kim, In Song et al. (2018). "Estimating Spatial Preferences from Votes and Text." In: Political Analysis 26.2, pp. 210 – 229.

Lauderdale, Benjamin E. and Alexander Herzog (2016). "Measuring Political Positions from Legislative Speech." In: Political Analysis 24.3, pp. 374 – 394.

Laver, Michael, Kenneth Benoit, and John Garry (2003). "Extracting Policy Positions from Political Texts Using Words as Data." In: American Political Science Review 97.2, pp. 311 – 331.

Lieberthal, Kenneth G. (2004). Governing China: From Revolution Through Reform. New York: W.W. Norton.

Lo, James, Sven Oliver Proksch, and Jonathan B. Slapin (2014a). "Ideological Clarity in Multiparty Competition: A New Measure and Test Using Election Manifestos." In: British Journal of Political Science 46.3, pp. 591 – 610.

— (2014b). "Ideological Clarity in Multiparty Competition: A New Measure and Test Using Election Manifestos." In: British Journal of Political Science 46.3, pp. 591 – 610.

Loo, Mark van der et al. (2020). stringdist :Approximate String Matching, Fuzzy Text Search, and String Distance Functions. Tech. rep. The Comprehensive R Archive Network.

Lowe, Will (2008). "Understanding Wordscores." In: Political Analysis 16, pp. 356 – 371.

Lowe, Will and Kenneth Benoit (2013). "Validating Estimates of Latent Traits from Textual Data Using Human Judgment as a Benchmark." In: Political Analysis 21.3, pp. 298 – 313.

MacFarquhar, Roderick (1997). The Origins of the Cultural Revolution: Volume 3: The Coming of the Cataclysm 1961-1966. Columbia University Press.

MacFarquhar, Roderick and Michael Schoenhals (2006). Mao's Last Revolution. Cambridge: Harvard University Press.

Mihalcea, Rada and Paul Tarau (2004). "TextRank: Bringing Order into Texts." In: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics.

Nanni, Federico and Goran Glavaš (2019). "Political Text Scaling Meets Computational Semantics." In: arXiv:1904.06217.

Proksch, Sven Oliver and Jonathan B. Slapin (2010). "Position Taking in European Parliament Speeches." In: British Journal of Political Science 40.3, pp. 587 – 611.

Qi, Peng and Koichi Yasuoka (2020). Simplified Chinese Universal Dependencies dataset converted from the GSD (traditional) dataset with manual corrections. URL: https://github.com/yl17124/UD_Chinese-GSDSimp.

Slapin, Jonathan B. and Sven-Oliver Proksch (2008). "A Scaling Model for Estimating Time-Series Party Positions from Texts." In: American Journal of Political Science 52.3, pp. 705 – 722.

Straka, Milan, Jan Hajič, and Jana Straková (2016). "UDPipe: Trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, POS tagging and parsing." In: Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016, pp. 4290 – 4297.

Straka, Milan and Jana Straková (2017). "Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe." In: CoNLL 2017 - SIGNLL Conference on Computational Natural Language Learning, Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, pp. 88 – 99.

Su, Yang (2006). "Mass Killings in the Cultural Revolution: A Study of Three Provinces." In: China's Cultural Revolution as History. Ed. by Paul Pickowicz, Joseph Esherick, and Andrew G. Walder. Stanford, CA: Stanford University Press.

Tang, Shaojie (2003). Yi Ye Zhi Qiu: Qinghua Daxue 1968 Nian "Bai Ri Da Wudou" (The Falling of One Leaf Heralds the Autumn: The "Hundred Days of Great Violence" at Tsinghua University in 1968). Hong Kong: The Chinese University of Hong Kong Press.

Vafa, Keyon, Suresh Naidu, and David M. Blei (2005). "Text-Based Ideal Points." In: arXiv.

Walder, Andrew G. (2004). "Tan Lifu: A "Reactionary" Red Guard in Historical Perspective." In: The China Quarterly 180.June 2003, pp. 965 – 988.

— (2006a). "Ambiguity and Choice in Political Movements: The Origins of Beijing Red Guard Factionalism." In: American Journal of Sociology 112.3, pp. 710 – 750.

Walder, Andrew G. (2006b). "Factional Conflict at Beijing University, 1966-1968." In: China Quarterly 188, pp. 1023 – 1047.

— (2009). Fractured Rebellion: The Beijing Red Guard Movement. Cambridge, Mass: Harvard University Press.

— (2015a). China Under Mao: A Revolution Derailed. Cambridge, Mass.: Harvard University Press.

— (2015b). "Rebellion and Repression in China, 1966 – 1971." In: Social Science History 38.3-4, pp. 513 – 539.

— (2016). "Bending the Arc of Chinese History: The Cultural Revolution's Paradoxical Legacy." In: China Quarterly 227, pp. 613 – 631.

Walder, Andrew G. and Qinglian Lu (2017). "he Dynamics of Collapse in an Authoritarian Regime: China in 1967." In: American Journal of Sociology 122.4, pp. 1144 – 82.

Walder, Andrew G. and Yang Su (2003). "The Cultural Revolution in the Countryside: Scope, Timing and Human Impact." In: China Quarterly 15.173, pp. 74 – 99.

Wan, Xiaojun and Jianguo Xiao (2008). "CollabRank: Towards a collaborative approach to single-document keyphrase extraction." In: Coling 2008 - 22nd International Conference on Computational Linguistics, Proceedings of the Conference 1.August, pp. 969 – 976.

Wang, Shaoguang (1995). The Cultural Revolution in Wuhan: Failure of Charisma. Hong Kong: Oxford University Press.

Wang, Yuhua (2019). "The Political Legacy of Violence during China's Cultural Revolution." In: British Journal of Political Science, pp. 1 – 25.

Wijffels, Jan and BNOSAC (2020a). textrank: Summarize Text by Ranking Sentences and Finding Keywords. Tech. rep. The Comprehensive R Archive Network.

— (2020b). udpipe :Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the 'UDPipe' 'NLP' Toolkit. Tech. rep. The Comprehensive R Archive Network.

Will, Lowe et al. (2011). "Scaling Policy Preferences from Coded Political Texts." In: Legislative Studies Quarterly 36.1, pp. 123 – 155.

Xu, Youyu (1999). Xingxing Sese de Zaofan: Hongweibing Jingshen Suzhi de Xingcheng Ji Yanbian (Rebellion of All Hues: The Formation and Evolution of Red Guard Mentalities). Hong Kong: The Chinese University of Hong Kong Press.

Appendices

A   The Organizations and Leaders of Factions by Political Groups

Table 3: The Organizations and Leaders of Factions in the Political Elite

Faction Types	Faction Organizations / Leaders
Radical	Jiang Qing (江青), Chen Boda (陈伯达), Kang Sheng (康生), Zhang Chunqiao (张春桥), Yao Wenyuan (姚文元), Tan Zhenlin (谭震林), etc.
Swing	Zhou Enlai (周恩來)
Reactionary	Chen Yi (陈毅), Li Xiannian (李先念), Nie Rongzhen (聂荣臻), Ye Jianying (叶剑英), Tan Zhenlin (谭震林),, etc.

Note: the classification of these leaders and organizations are based on Walder (2006b), Walder (2009), Xu (1999), MacFarquhar and Schoenhals (2006), and Tang (2003).

Table 4: The Organizations and Leaders of Factions in the Red Guards

Faction Types	Faction Organizations / Leaders
Rebel	Tsinghua University High School Red Guards (清华附中红卫兵), Beijing University High School Red Flag Battle Group (北大附中红旗战斗小组), 北京人大附中红卫兵, Western District Picket Corps (西纠), Eastern District Picket Corps (东纠), Beijing NO.161 High School (北京一女中红卫兵), First Headquarters (一司), Second Headquarters (二司), The United Action Committee (联动), Tang Lifu (谭力夫), Shanghai Red Guards General Headquarter (上海市红卫兵总部), etc.
Conservative	Tsinghua Jinggangshan Regiment (清华大学井冈山兵团), Aeronautics Institute Red Flag (北航红旗), Geology Institute East Is Red (地院東方红), New Beida Commune (新北大公社), Third Headquarters(三司), Beishida Jinggangshan Combat Brigade(北师大井冈山战斗兵团), Nie Yuanzi (聂元梓), Capital Red Guards Congress (首都大专院校红卫兵代表大会), Capital College Red Guard Congress (首都中学红代会), Kai Dafu (蒯大富), Tan Houlan (谭厚兰), Shanghai Workers' Revolutionary Rebels General Headquarters (上海市工人革命造反总司令部), Nanfang Daily (南方日报), Wuhan Revolutionary Rebel Headquarters (武漢鋼二司), Li Ne (李訥), Guangdong Revolutionary Committees Dongfeng Commune (广东省委机关东风公社), Guangzhou First Headquarters (广州一司), Sun Yat-sen University (Zhongshan Red Flag) (中山大學「中大红旗」), Guangzhou Third Headquarters (广州三司), Chongqing University 815 camp (重庆大学八一五战斗团), Mao Yuanxin (毛远新), etc.

Note: the classification of these leaders and organizations are based on Walder (2006b), Walder (2009), Xu (1999), MacFarquhar and Schoenhals (2006), and Tang (2003).

B The Timeline of Major Historical Incidents in the Cultural Revolution

Table 5: The Timeline for Major Historical Incidents in the Cultural Revolution

Time	Incidents	Incidents Details
May 25, 1966 - Aug 17, 1966	First Marxist-Leninist wall poster	"First Marxist-Leninist wall poster" (第一张马列主义大字报) signed by Nie Yuanzi (聂元梓) and six others at Peking University denounces the party secretary.
Aug 18, 1966 - Sep 26, 1966	Red August	The first mass rally of Red Guards is held at Tiananmen Square. Mao Zedong expresses his support for the Red Guards. "Red August" (红八月) would start.
Sep 26, 1966 - Jan 16, 1967	Zhou Enlai's declaration	At a mass rally of the Third Headquarters (三司), Zhou Enlai (周恩来) declares the minority faction's cause correct and orders the rehabilitation of students punished by work teams and the destruction of materials in their case files.
Jan 17, 1967 - Feb 15, 1967	The announcement of new public security regulations	Xie Fuzhi (谢富治) releases new public security regulations that designate "The United Action Committee" (联动) students as counterrevolutionaries.
Feb 16, 1967 - Apr 1, 1967	February Countercurrent	A frontal attack upon the Central Cultural Revolution Group by the PLA Marshals Chen Yi (陈毅) and Ye Jianying (叶剑英) in meetings of the Central Caucus chaired by Zhou Enlai (周恩来). The historical event named the "February Countercurrent" (二月逆流).
Apr 2, 1967 - Jul 19, 1967	Adopt a Correct Attitude toward the Little General	The People's Daily publishes the editorial titled Adopt a Correct Attitude toward the Little Generals (正确对待革命小将). Later, the PLA was explicitly ordered not to fire on members of mass organizations; not to declare mass organizations reactionary; not to take revenge on rebels who had attacked the PLA in the past. The editorial has implied that the Red Guards' Armed Struggles (武斗) are allowed.
Jul 20, 1967 - Aug 22, 1967	Wuhan Incident	Wuhan incident (武汉事件) breaks out.
Aug 22, 1967 - Apr 22, 1968	The Burning of the British Mission	After days of the anniversary of the first great Tiananmen rally, the Red Guards invaded the British chargé d'affaires and burning the building.
Apr 23, 1968- Jul 27, 1968	100-day Clashes	At Tsinghua University, Jinggangshan Brigade (井冈山兵团总部) attacks a building held by the 414 faction (414 派), which repels the attack with rocks and spears. The historical event named "100-day clashes" (清华百日大武斗) is starting.
Jul 28, 1968 - Jan 3, 1967	Mao Zedong's summons	Leaders of the major Red Guards factions are summoned to an early morning meeting with Mao Zedong and other top officials; an angry Mao tells them that they have committed serious errors and they must return to campuses and welcome troops.
Jan 4, 1967	Shanghai's January Storm	Zhang Chunqiao (张春桥) and Yao Wenyuan (姚文元) arrived in Shanghai in their dual capacity as representatives of the CCRG and senior local cadres. Zhang later called upon a meeting of Red Guards leaders that the rebels to organize a "Down with the Shanghai Party Committee mass rally." Shanghai's January Storm (上海一月风暴) would start.

Note: the sequence of incidents in this table is taken from Walder (2009), at page 271-283.

C   The Estimated Positions of Each Major Historical Incident

Figure 8: The Estimated Positions of Each Major Historical Incident for Individuas of the Red Guards and the Political Elite