# Rebellion, crime and violence in Qing China, 1722–1911: A topic modeling approach

## Ian Matthew Miller

*Department of East Asian Languages and Civilizations, Harvard University, 9 Kirkland Street, Cambridge, MA 02118, USA*

## Abstract

Banditry and unrest in eighteenth and nineteenth century China have attracted substantial attention from several generations of researchers. Often, they apply particular ontologies a priori to the source base. Given their reliance on state documents, these studies are subject to the perspectives of record-keepers and their theories of violence. It is particularly difficult to apply fixed definitions to concepts like ''banditry'' and ''unrest''—a problem that applies as much to modern researchers as to our historical informants. To better view the nature of violence in the Qing Dynasty—as routine crime, and as rebellion and unrest—it is important to develop a model of how administrators understood it. Therefore, rather than assuming a fixed set of categories, this study models Qing administrators' typologies of violence based on the frequencies of term co-occurrence. Based on the term groupings in the model, five topics relate to violent unrest. Each topic accounts for a particular statistical pattern of word use corresponding to patterns of occurrence, observation and recording of related phenomena. These groupings give some insight into the ''crime rates'' of the eighteenth and nineteenth centuries, and more importantly, these groupings cast light on their understandings of crime, rebellion and unrest.
© 2013 Elsevier B.V. All rights reserved.

*Keywords:* China; Qing Dynasty; Rebellion; Unrest; Distant reading; Topic model

## 1. Introduction: the bandit problems

The administrators of the Qing Dynasty (1644–1911) had a bandit problem: gangs of robbers hid in the hills and marshes, European merchant adventurers brought opium to the coasts, and sectarian rebels spread through the countryside. On the centennial of its fall, the Qing Dynasty

still presents a bandit problem: these three very different phenomena were all recorded using much the same language: ''robber-bandits'' (*zeifei*), ''foreign bandits'' (*fanzei*) and ''rebellious bandits'' (*nizei*) all use the term *zei*—which roughly accords with multiple English words, including ''rebel'' and ''criminal,'' and which I have translated generally as ''bandit.'' How is an historian to differentiate between three such different phenomena? ''Bandits'' is just one of many historical problems that contemporaries found difficult to address, and modern scholars are equally at a loss to categorize and comprehend. What makes bandits so problematic? I will argue that they were difficult for their contemporaries to deal with, and remain challenging categories for historians, because they raise four separate-yet-related issues—bandits as a social phenomenon, bandits as a surveillance and reporting problem, bandits as an ontological question, and bandits as a textual processing challenge. To the contemporary historian, there is not a single ''bandit problem,'' but four separate-yet-interrelated bandit problems.

Bandits are a diffuse social problem because there are few full-time bandits. The boundaries between failed farmers and opportunistic robbers, and between criminal gangs and rebel armies, can shift unpredictably. Beyond the actual movement of individuals in and out of the ''economy of violence'' (Greenshields, 1994, pp. 1–2; Robinson, 2001, p. 2), the ontological category of ''bandit'' is itself largely a matter of perspective—it could cover a large set of activities with little in common aside from being ''illegitimate.'' Those that the emperor called a ''bandit gang'' might consider themselves to be a defensive militia, a righteous army, or even a group of remonstrating officials. ''Bandits'' are frequently, but not always, violent, they generally form gangs, but individual ''bandits'' also appear in the records, they may espouse seditious or heterodox causes or have no explicit political position. There is substantial overlap between terms used to describe economic, political and social roles with little in common. If illegal or heterodox groups are identified as ''bandits,'' there remains the question of whether and how they should be reported—if and when banditry attracts high-level notice, how do the authorities respond? For the historian, there is also the question of how and where government reports of violence were filed for posterity and future reference. I will begin by exploring the several difficulties that ''bandits'' and the related realms of crime, violence and rebellion pose to the historian before offering a means of addressing these ''bandit problems.'' Due to the diffuse and ontologically complex nature of banditry, any approach to these issues will take us into the *great unread*—the large body of historical texts too numerous, obscure or boring to attract much attention, but which yields new insight when read systematically.

## 1.1. Social problems

To Qing officials, and to the historians who study them, bandits are first and foremost a social problem. Why do peaceful citizens resort to violent behavior? How do small-time crooks cause large-scale unrest? When do money-hungry bandits become millenarian armies? Is rebellion the same as revolution? These questions loom large to scholars because they were practical concerns to the administrators of a violent era, but also because they inform our understanding of the following period—in particular, the nationalist revolution of 1911 and the communist revolution of 1949. Present day inquiry into questions of unrest and rebellion are therefore affected by the worldviews of eighteenth- and nineteenth-century Qing administrators, contemporary Western observers, and modern historians in both China and the West. Each of these traditions has made major contributions to our understandings of violence and unrest in early modern China, but each must be understood in light of its particular interests and biases.

Through the 1960s and 1970s, most Chinese- and Japanese-language scholarship on Ming (1368–1644) and Qing Dynasty uprisings was colored by the narratives of nationalism or class struggle; researchers compiled and read documents with an eye for ethnic- or class-based conflict. This produced an extensive body of scholarship on ''peasant uprisings,'' stressing domestic developments leading to the twentieth century revolutions (Harrison, 1969; Liu, 1981). On the other hand, Western scholarship from the same period was mostly concerned with cultural difference, and thus it was much more interested in the roots of rebel ritual. This included both obviously foreign cultural forms like the Heaven and Earth Society (the famous ''triads'' of Chinatowns worldwide; Murray and Baoqi, 1994, pp. 89–115; Ownby, 1996, pp. 7–11) and response to Western culture like Taiping Christianity (Cheng, 1963, p. vii). This discourse tended to reinforce notions of China's cultural backwardness prior to the arrival of the West (Cohen, 1984, pp. 1–3, 9–56). What these two approaches have in common is the emphasis on the ideological aspect of rebellion—regardless of whether it is proto-nationalism, class oppression or religion that makes people revolt.

For better or worse, both these threads of scholarship played a major role in the compilation of sources used by later historians. In the 1970s and 1980s, a new generation of Western historians incorporated Japanese and Chinese scholarship and a critique of older historiography, so as to produce substantial new research on rebellion. This included work on uprisings previously overlooked in the West, including the White Lotus (Kuhn, 1970), Eight Trigrams (Naquin, 1976), and Nian Rebellions (Perry, 1980)—as well as better-known ones like the Taipings (Kuhn, 1970) and Boxers (Esherick, 1987) and more work on rebellion generally (Chesneaux and Bianco, 1972; Yang, 1976). Despite this new perspective and the attempt to get away from previous works' shortcomings, this research tended to split into structuralist and culturalist camps. State documentation of rebellions and uprisings was still mostly categorized according to Marxist historiography (Robinson, 2001, p. 175; Teng, 1962, pp. 59–81), while the internal documents of rebellious groups like the Taipings or Heaven and Earth Societies was still largely based on translations by missionaries and colonial governments (Murray and Baoqi, 1994, pp. 116–150; Ownby, 1996, pp. 7–11). Even as they turned away from the explicit ideological positions of their predecessors, most of these scholars were affected by their underlying assumptions. Thus, work based on Chinese compilations of state sources tended to emphasize structural causes, while research based in the internal documents of rebellious groups tended to play up cultural causes of rebellions.

More recently, instead of searching for the roots of particular rebellions, researchers have turned to characterize the general climate of violence and unrest in the Ming and Qing. This approach was spurred by three major developments: the 1970s generation of scholarship detailed above, the new institutional scholarship—particularly Charles Tilly's work on collective violence—and the opening of major archives in Taiwan and mainland China in the 1980s. Following Tilly, Wong (1982, 2001), Tong (1991) and Hung (2011) have all produced work explicitly on ''collective violence,'' relying mainly on the same document bases as earlier work. Another group of legal scholars has focused on memorials to the Ministry of Justice (*xingke tiben*), now available in the First Historical Archives to document the role of violence in everyday social conflict (Buoye, 2000; Macauley, 1998; Theiss, 2004). Both of these approaches have avoided the pitfalls of searching for the roots of specific, outstanding events. Taken together, they tend to argue that violent conflict and upheaval were endemic to society—even outside of major uprisings—and the boundaries between phenomena we call ''crime,'' ''unrest,'' and ''rebellion'' are amorphous (see especially Robinson, 2001, pp. 20–26). While we can understand rebellions in terms of their individual histories, it is also worthwhile to understand them in the context of ''banditry'' more broadly defined.

## 1.2. Reporting problems

We should be aware that these last two generations of scholarship make use of sources organized according to another set of perspectives on society—those of the Ming and Qing bureaucracies. These include local archives, gazetteers (*difang zhi*) that served as occasional compilations of local documents, central archives of ordinary memorials (*tiben*, messages to the central bureaucracy) and the Veritable Records (*shilu*) used in this study—themselves largely compilations of a different type of ''secret memorials'' (*zouzhe*, messages directly to the emperor). This source base offers a huge advantage in approaching crime and rebellion through conceptual lenses contemporary to the events themselves. But this means that to address the phenomena of banditry through these documents, we must make inquiry into the nature of these documents, with an awareness that—like the historiography detailed above (Section 1.1)—each corpus represents the result of a distinct process of record-keeping. When were crimes and unrest noticed by low-level officials? When did officials report these events to their superiors and to the central bureaucracy? When and how were these reports compiled, preserved and annotated? The chain of reporting was different for each event, and each type of event. In some cases, we have good understanding of what was recorded and preserved, while in other cases we must speculate. Finally, we should be aware that each researcher uses a different document base, and that the practices differ according to corpus. Here, I will give a brief summary of the chain of reporting, the specifics of the corpus used in this study will be considered in Section 2.1.

To be reported, instances of crime or unrest must be observed; in the case of state records, that means that they must come to the attention of a representative of the state. While there were many people involved in local government in the Qing, the lowest level of official with authority to report directly to the central government was the magistrate or assistant magistrate at the county level. In an effort to prevent corruption, these officials were not permitted to serve in their home area and were rotated to new positions after three years of service; they generally did not know much of local customs, or even speak the dialect. This meant that magistrates were highly dependent on the sub-bureaucracy—including village heads, and semi-official clerks and runners—as well as the broader educated classes of local society. In turn, it meant that the events reported to the magistrate depended on the incentives of the people doing the reporting. It is probable that many minor crimes and conflicts never crossed the magistrate's desk, simply because the people involved preferred to handle affairs privately (Buoye, 2000, pp. 6–7). Nonetheless, most instances of illicit violence would almost certainly have at least one party interested enough to report it to the magistrate, and the magistrate was personally responsible to investigate cases of homicide (Buoye, 2000, pp. 7, 239).

Of the incidents reported to the magistrate's office, most crimes and lawsuits were handled locally, and would be unlikely to make it into higher-level documentation. Scholars have been able to capture a picture of crime and local society by using the few preserved archives at the county level—in Taiwan (Allee, 1994) and Sichuan (Sommer, 2000). A very few of these events—particularly those leading to larger problems of unrest or rebellion—might have made it into the local gazetteers (*difang zhi*) compiled every few decades. It is worth noting that these gazetteers used very concise and formulaic language, imposing categories that have been taken up by many scholars relying on this source base (Tong, 1991; Wong, 1982, 2001; Yang, 1976), a problem addressed in more detail in Section 1.3. Regardless, it is important to recognize that many types of crime would never make it into the documentation of the central bureaucracy, especially in cases where the state thought local society could manage its own business (Ownby, 1996, p. 147) —but instances of large-scale unrest or murderous violence almost certainly would.

More important to the present discussion, most cases involving violence or large-scale unrest—and any case involving homicide—would have been reported to the central bureaucracy (Buoye, 2000, pp. 239–244). Local officials might have had their own incentives to avoid reporting cases that would paint them in a negative light. However, magistrates were also subject to extensive oversight—including both lower- and higher-level officials within the local hierarchy, inspecting censors specifically tasked with surveillance, and even retired officials within their locality. Many local officials were chastised or cashiered for failing to report nascent problems, which suggests that they did hide problems but were also afraid of hiding them. It seems likely that these officials would be most likely to hide crimes least likely to spiral out of control—like property crimes or smuggling—and to report issues of more danger to both the state and their own reputation—like capital crimes and armed conflict.

Therefore, the central repositories of crime reports are a better source for studying violent crime than for studying crime generally. Several scholars have used these central archives to conduct broad studies of criminal violence, including property disputes gone wrong (Buoye, 2000) and retribution for violations of chastity (Theiss, 2004), using a large corpus of *ordinary memorials* (*tiben*) sent to the central bureaucracy. The disadvantage of this archive for the present study is that it is divided according to the branch of administration responsible. As recent scholarship has noted, popular unrest could and did become rebellion (Hung, 2011, pp. 102–134), and rebel armies—especially in the nineteenth century—were composed largely of criminal gangs and brotherhoods active prior to the breakout of large scale violence. As Kuhn (1970, p. 40) puts it, "there was no distinguishing the rebels from the human stream in which they swam." Thus, incidents that crossed the line between crime and rebellion—between the respective realms of the Ministry of Justice and the Ministry of War—are hard to survey on the basis of the Ministry of Justice archives alone.

More generally, the events considered of primary importance would all cross the desk of the emperor himself. These included rescripts from the central bureaucracy, as well as *secret memorials* (*zouzhe*—literally "folded memorials")—initially sent by an inner circle of trusted officials, but by the eighteenth century, largely replacing regular memorials (Wilkinson, 2013, pp. 283–284, 808). Ether way, these would be communications determined by lower levels of administration to require the personal attention of the emperor—including routine, high-level matters, as well as emergencies. Many, though not all, of these secret memorials made their way into the posthumous compilation of *Veritable Records*—the collection of documents important to an emperor's reign. This archive too has been widely used, especially to detail the history of specific rebellions and, increasingly, for more comprehensive studies (Hung, 2011). However, as I will explore in more detail in Section 2, much of this corpus remains part of the *great unread*.

If this were a study of theft or smuggling, there would be good reason to discount the information making it into even the lowest level of official reports, but as a study of violent crime, unrest and rebellion, we can be fairly secure that these incidents were reported to higher levels and recorded for later review. In fact, because "banditry" is the type of phenomenon of particular interest to the state, it provides a good test of the breadth of the archive and the ins and outs of official discourse (Robinson, 2001, p. 3). Nonetheless, we can see that summing all occasions of reported banditry gives us neither a "crime rate" nor a "reporting rate," but rather some interaction of the two. The contents of the archive is highly dependent on the ability of bandits to hide, the willingness of officials to report crime, and the degree to which reported instances of violence attracted the attention of the imperial court. Yet even if we were able to disentangle these functions to arrive at a true rate of "banditry" in society, we are left with the question of whether all these "bandits" are referring to the same thing, and if there are other related events that are not called "banditry."

## 1.3. Ontological problems

Aside from the issue of how and when events came to be reported to the imperial center, we are faced with the question of what language was used in reporting them. ''Banditry,'' like ''smuggling'' (Tagliacozzo, 2005, p. 5), forms a rather diffuse zone of meanings. In English, the root meaning of bandit relates to their legal status as outlaws, but it carries a strong connotation of ''armed band.'' In Chinese, *fei* is the closest term, denoting ''bandits, robbers, gangsters,'' but it has substantial overlap with *zei*, which covers a range from ''thief'' to ''traitor.'' These and related terms cover a range of behaviors unified by their illegal, illicit or rebellious nature, and they generally (but not always) involve large groups and the use of violence. Nonetheless, there are a range of ways in which groups could be called ''bandits'' and behavior called ''banditry'' (for a more complete list of Chinese terms describing men of violence, see Robinson, 2001, pp. 20–26).

First, people in illegitimate circumstances could be called outlaws. Soldiers away from their posts, escaped criminals, secret societies and the general floating population of ''bare sticks'' (*guanggun*, unmarried males) could all be considered ''bandits'' without committing acts that would be considered ''banditry.'' Social status and location factored in to this in ways that could change. For example, an unofficial militia might be condoned by one official and reported by his successor. Second, illegal actions could make people ''bandits.'' This was probably the most clear-cut definition—the Qing Code had fairly uniform standards of crime. Furthermore, Qing bureaucrats were careful social observers and did attempt to classify criminal behavior more closely over time, through the addition of sub-statutes (Ownby, 1996, p. 27, 147). Nonetheless, behavior that was violent and violated a statute or sub-statute could easily be called ''banditry,'' as well as more specific terms like ''feuding'' (*xiedou*) or ''robbery'' (*qiangdao*). Finally, ''banditry'' was often a matter of perspective—behavior considered legitimate to one group could be illegitimate in the eyes of others. Peasants resisting taxes that they considered excessive, merchants trading despite a newly imposed ban, or even officials accused of embezzlement might all be called ''bandits'' by the state administration, even as they judged themselves in the right. Emperors frequently signaled new standards of sedition by calling certain people ''bandits,'' but these varied a great deal based on the individual.

There have been numerous attempts in the literature to define categories with relation to violent uprisings and armed bands. Some researchers assigned these categories based on theoretical assessment of texts, coming up with terms like ''heterodox'' and ''orthodox'' groups, based on relations to the state (Kuhn, 1970, pp. 165–88). A few have used the language of the texts to classify tight categories like ''mountain bandits'' or ''troop mutiny'' (Tong, 1991, pp. 46–75; Yang, 1976). Others have done close studies of the Qing code to compare the treatment of one or more categories of violence and crime (Ownby, 1996, pp. 145–78; Theiss, 2004, pp. 98–117). As noted above (Sections 1.1 and 1.2), the language used in these texts was already the product of multiple historiographical processes—the compilation and editing of Qing officials and frequently of later scholars. Ultimately, these attempts at classification tend to proceed from the categories in the theory, historiography or legal literature. As a group of phenomena that crossed the boundaries of these classifications—both those of Qing observers and of later scholars—''banditry'' demands a more comprehensive reading. In place of imposing categories a priori—whether based on theory or on some subset of the corpus—I will use the association of words in the corpus as a whole to generate categories of understanding. The discourse on what constituted a ''bandit'' and when ''unrest'' became ''rebellion'' were not fixed and must be considered as part of our analysis.

## 1.4. Textual problems

Ultimately, researchers must start with the texts, but we must recognize that these are not straightforward representations of historical phenomena. They are highly biased perspectives resulting from several stages of imperfect observation, active deception and politically and socially informed epistemic processes. To complicate matters, there is no single, coherent vocabulary that aims to describe the phenomenon of violent unrest; any single term is either overly specific or overly general, and in some cases, both. Unlike the case of purely ''legal'' phenomena, we cannot simply fall back on code and precedent. In close reading, this problem can be addressed by attention to contextual, syntactic and symbolic cues to inform our interpretation of the semantic content. These cues can help us determine if *zei* is being used for the meaning ''robber'' or for the meaning ''rebel,'' or both. If we wish to venture into the *great unread*—to do a more comprehensive reading of the complete textual phenomenon—this approach is not tenable.

Generally speaking, distant reading approaches are limited to identifying semantic information. Many of these approaches have focused on keyword or tag-based methods. These work well when dealing with phenomena with well-defined ontologies—names, places, or specific terminology. However, to approach events and themes with more diffuse historical structure and with more loosely defined ontological structure, keyword approaches fail. I will argue that this is not because semantic information is insufficient to identify these phenomena, but because we cannot construct a list of terms that identifies all instances of a phenomenon without also identifying unrelated phenomena. In place of searching for vocabulary that is ontologically equivalent of the concept ''bandits,'' I take an approach that assumes probabilistic association between that concept and a broad set of vocabulary. Concretely, instead of assuming that *zei* or *fei* or *zeifei* always means ''bandit'' or ''robber,'' and instead of assuming that ''bandit'' or ''robber'' is always written using this well-defined vocabulary, I look for a frequency distribution of terms that often mean ''robber'' and terms that are otherwise associated with that idea. As we will see, this distribution includes words that can be translated as ''bandit''—*zei* and *fei*—with high frequency, and it also prominently features words associated with what bandits do—rob (*qiang*), form bands (*bang*), escape (*tao*) —and with how the state responds—by capturing them (*na*, *huo*, *bu*).

In this model, the realm of concepts captured by the term ''bandit'' falls approximately into six topics, roughly accounting for ideas of ''sedition,'' ''crime,'' ''unrest,'' ''border rebellion,'' ''rebellion,'' and ''major rebellion.'' Before examining the specific results, I will first explain the model and the corpus used to probe the bandit problems (Section 2). I will then explain the six topics individually, assessing how well they describe known historical phenomenon (Section 3). Finally, I will argue that this modeling approach allows for new conclusions about violence, unrest and rebellion—as social problems, ontological problems and textual problems (Section 4). Essentially, the topic model presents a novel epistemological approach: it parses ''banditry'' into a probabilistic ontology rather than a deterministic one. At the level of the individual document, this can be problematic, but at the level of the corpus, it allows for novel conclusions. It is a poor tool for close reading, but a promising plumb for the depths of the *great unread*.

## 2. The corpus and model

There is good reason to believe that the language of historical records provides sound footing for broad analysis of violence. The Qing Dynasty (1644–1911) was enormously concerned with

violence, especially mass violence, and with any other behavior that could lead to anti-state action. As noted above (Section 1.2), there is reason to believe that most instances of large-scale violence and unrest were reported to the central bureaucracy and recorded in its compiled records. I have chosen these posthumous compilations of secret memorials as the best single corpus of materials on events that led to rebellion or that the state believed might do so. Because of the broad nature of this corpus, it is difficult to define a set of vocabulary a priori. Instead, I use a statistical model to assign *topics* to the *documents* making up the *corpus*. In this model, the *topics* are determined by the association of multiple terms across documents. I believe this approach to this set of texts offers a useful way to look at the broader usage of concepts with diffuse ontologies.

## 2.1. The veritable records

The central bureaucracy of the Chinese state collected hundreds of thousands of records, including tens of thousands reporting violent crime and unrest. Any report important enough to cross the desk of the emperor was gathered in archives alongside records of the edicts by which the emperor responded. After the death of a sovereign, they were organized chronologically for future reference. These posthumous corpuses, collectively called the *Veritable Records* (*shilu*) were used by bureaucrats looking for precedence. They later formed a major source for official histories written by succeeding dynasties, including the Draft History of the Qing (*Qing shi gao*) written in the Republican period. Two *Veritable Records*, from the Ming and Qing Dynasties, are extant and digitized in their entirety (see Qing, 2000). It is important to note that these records are not comprehensive; they do not contain much of the ordinary paperwork of the central bureaucracy (including ordinary memorials and rescripts) or any record of local events that were not reported to the central administration. In addition, they are not always complete and their composition changed over time. In particular, the *Records* from the early Ming are quite sparse, and major changes in record keeping in the early Qing (including the innovation of secret memorials) and imperial interference with the compilation changed the nature of the early Qing *Records* (Wilkinson, 2013, pp. 611–612, 833–834). I have therefore started this project with the Yongzheng Reign (1722–1735), the first period when secret memorials were regularly in use. Despite its weaknesses, the *Veritable Records* is conceived as a complete corpus, and it represents a fairly comprehensive record of events of importance to the central court. Nonetheless, as noted above (Section 1.2), these sources offer a biased picture of violence, we cannot read them as an unproblematic reflection of society, but must rather consider the ways in which they reflect both social trends and trends in recognizing, reporting and recording these trends.

## 2.2. The Latent Dirichlet allocation

The statistical workings of this topic model (the Latent Dirichlet Allocation; Blei et al., 2003) and the software implementation (MALLET; McCallum, 2002) that I use will not be described in depth. The model can be summarized as follows:

- *Topics*[1] are distributions across all *word*s in the *corpus*. While each *topic* technically includes some proportion of every word in the *corpus*, it features some *words* with significantly higher

---

[1] Note that I have used *italics* to designate terms used for their specific meaning in the context of the topic model. Unitalicized words indicate the more general sense of the term.

frequency than others. These words are assumed to be topically related. Each *topic* is described by giving a list of its highest-frequency *words*.

- *Documents* are considered to be distributions across all *topics* in the corpus. While each *document* technically includes some proportion of every topic, it features some *topics* with much higher frequency than others. These *topics* should represent the main themes of the *document*. *Documents* are described by giving their *topic* proportions.
- In the *generative model*—the theoretical model of how texts are generated—a *document* is created by ''choosing'' *topics* according to the *topic* proportions in the *corpus* (i.e., high-proportion *topics* are more likely to be chosen), and then ''choosing'' *words* from those *topics* according to the *topic* proportions in the *document* and the *word* proportions in the *topic*. This means that the *words* most likely to be chosen for a given *document* are high-proportion *words* in high-proportion *topics*.
- The *topic model* is reverse engineered statistically based on the number of *topics*, and the observed occurrence of *words* within documents within the *corpus*. In effect, it is based on statistical observation of which *words* co-occur within *documents* more than they co-occur between *documents*. These words are then ''grouped'' into certain *topics*, and the *documents* where they are clumped are assigned high proportions of those *topics*.
- This research only specifies four inputs to the model:
- Number of *topics*—in this case fifty (chosen over forty and sixty for better precision);
- The corpus—the Qing Veritable Records from 1723 to 1911;
- *Document* delimitation—individual entries in the *Records* of more than twenty characters; and
- *Word* delimitation—individual Chinese characters, excluding grammatical characters and very high frequency characters.
- It should be noted that there are several assumptions made in using this model that may not be representative of the way *documents* were actually written and compiled in the *corpus*:
- *Word* order is not considered within *documents* (this is a ''bag of words'' model). Presumably *word* order does matter in the actual writing of *documents*, but this model assumes that it is not important in determining the *topics* used in a *document*; in other words, it assumes that grammar may contribute to meaning, but not to topicality.
- *Documents* are analyzed synchronically by the topic model (i.e., time is not integrated into the statistical model). These *documents* were actually written over a long period of time, so we should expect some change in the composition of *topics* over time. This is addressed after application of the model, rather than within the model.
- *Document* length is not considered except to ignore very short *documents* (under twenty characters). Even with the twenty-character cutoff, shorter *documents* are less likely to be assigned ''sensible'' *topic* proportions.
- Written Chinese in the late imperial period did not actually consist of single-character *words*, single and multiple-character words were both common. Single characters are analyzed because there were no good word-delimiting algorithms tested for classical Chinese at the time of research. Certain types of word formations, including many proper nouns, are actually reverse engineered by the topic analysis and placed into discrete *topics*.

Clearly there are some problematic assumptions in this document model. As noted above, the topic analysis proves capable of reverse engineering certain structures that are not part of the model, such as multi-character words. However, one especially problematic assumption is made in this analysis: it does not account for words, grammar or other intratextual structures. This analysis is very powerful at accounting for many corpus-level, intertextual phenomena—exactly

the type of analysis needed to probe the dark areas of the textual map. But it is not a substitute for other approaches to understanding the more proximate realms of meaning of individual texts within the corpus. I accounted for these shortcomings in two ways—by reimposing temporal meta-data on the records after running the topic model, and by close reading of as many documents as possible.

## 3. Crime and rebellion in the topic model

Of the fifty *topics* in this model, six are clearly related to areas where the dynastic apparatus encountered illegitimate violence. Not all of the *documents* covered by these *topics* deal with "banditry" explicitly, but these *topics* present a fundamental picture of how the state understood related concepts of violence, crime, rebellion, and illegitimacy. Closer examination will reveal a complex pattern of concern with different ways in which subjects violated the peace and sovereignty of the dynasty. The model provides two obvious approaches to understanding the meaning of these *topics*: it outputs the *words* of high proportion within each *topic*, as well as the proportion of *topics* in each *document*. To unpack these patterns, I use both sources of information. First, I give a general impression based on the proportion of *words* within each *topic*, and then I make a closer reading by surveying the temporal structure of each *topic*, along with a random selection of *documents* with high *topic* proportions. No labels are provided for the *topics* by the model. For convenience, I have given the *topics* titles based on my impression of their content. Also, the Appendix lists the top frequency *words* for each of the six *topics*.

Note that this method is quite different from other surveys of crime or collective violence. Tong (1991, pp. 27–31) has provided some discussion of the relative merits of using *incidents* or *events* (i.e., groups of associated *incidents*) as the unit of analysis. Buoye (2000) takes the legal *case*, roughly equivalent to the *event*, as his base unit for aggregation and analysis. Instead, I have summed *topic proportions* over *documents* to generate my index of *topic* occurrence. This is generally closer to the *incident* than to the *event*—that is to say that every *document* that references a *topic* adds to the count for that *topic*, even if the *event* detailed in the *document* has already been referenced elsewhere. However, there are also cases where many *incidents* or *events* are summarized in a single *document*—for example, the "autumn assizes" reviewed all the capital crimes of the year in a small number of *documents*. Therefore, the sum of *topic proportions* is a better indicator of interest in a *topic* than occurrence of a *topic*.

A survey of the common words in the six *topics* of interest suggests a neat split between ordinary incidents or crime that was reported to higher administrators as a matter of course, and extraordinary incidents or rebellion that involved the central bureaucracy because they required more resources and centralized planning. In the first category are two *topics* related to problems of law and order: Crime and Unrest.[2] As we might expect, documents with high topic proportions in Crime and Unrest feature *words* related to crime (*fan*—violate, *zui*—crime, *jian*—illicit), and law (*an*—case, *li*—sub-statute, *lü*—legal code). Unrest is more concerned with finding (*jiu*—investigate), and catching criminals (*huo*, *na*, *bu*), while Crime involves their interrogation and punishment (*xun*—interrogate, *shen*—trial, *gong*—confession, *xing*—punishment). Both of these *topics* attracted relatively stable interest from the dynasty over time, and it would appear than individual *incidents* tended to be recorded in a minimal number of *documents*. A third *topic*, Sedition is less about violence than about the intellectual and symbolic content of crime. In many

---

[2] For convenience, I have capitalized *topic* labels to separate them from the general usage of these terms.

ways, it bridges the law and order topics and the rebellion topics, containing relatively high frequency of words related to crime and punishment, as well as terms like rebel (*ni*) seen below.

Clearly in the second category are three *topics* concerned with rebellion and its suppression: Rebellion, Border Rebellion and Major Rebellion. These all feature high proportions of words describing rebellion (*ni*, *pan*), "bandits" (*zei*, *fei*), the military (*bing*, *jun*) and particularly to suppressing and wiping out rebels (*jiao*). Unsurprisingly, *incidents* captured by these *topics* tend to produce many *documents*. At face value, there are few major differences between these *topics*. Border Rebellion makes reference to border regions (*fan*—foreign, *chuan*—Sichuan, *zang*—Tibet, *mian*—Burma), minority peoples (*tu* and *si*—tribal administration, *qiu*—tribal chief, *la* and *ma*—Lama), and it contains many characters used to transliterate non-Chinese words. Clearly, it is referring to rebellion or warfare involving non-Han ethnic groups or border regions. The differences between Rebellion and Major Rebellion are less apparent. As we will see below, the differences between these *topics* are not at the level of *words*, but rather in the distribution of groups of *words*. Let us look at these *topics* one by one.

## 3.1. Crime

Crime is most clearly associated with judgment and punishment of criminals, as opposed to Unrest (Section 3.2), which frequently appears in documents concerned with the pursuit of criminals still at large. As noted above, the most closely associated *words* are mostly related to the processes of interrogation, legal precedence and judgment. A survey of *documents* featuring this *topic* reveals that it primarily captures the regular pattern of legal court proceedings. It covers a wide range of violations of the law of the state. Many of the crimes captured by this *topic* are nonviolent, albeit serious. These range from private digging of ginseng in violation of government monopoly (*si bao renshen*) to counterfeiting (*weizao*, *sizhu*). It even captures a case of taking bribes to take someone's place in sitting for an exam (*shou hui dingti*), a serious violation of the sanctity of the primary means of recruiting officials. This *topic* also includes all manner of violent crimes, ranging from individual manslaughter (*zhisi*) to inciting a mob (*yi zhong zishi*). One particular feature of the corpus is worth noting here: many of the entries featuring this *topic* are routine reports of executions as part of the regular "autumn assizes." The *documents* reporting the autumn assizes also contain long listing of names of criminals, which might be supposed to cause problems for the model. In fact, the *topic model* accounted for this pattern by way of three *topics*—the Crime *topic* and two Han Names *topics* whose high-frequency *words* consist largely of Han surnames. The *documents* reporting autumn assizes have high proportions of both the Crime *topic* and of these two Han Names *topics*.

The temporal pattern of Crime (see Fig. 1) further suggests that it shows the routine administration of judgments and punishments. Compared to other *topics*, its shows a moderate amount of variance. Its standard deviation of 1.72 is higher than some administrative *topics* like Personnel (0.83) and Infrastructure (0.99) that I have not discussed here. But this standard deviation is significantly less than the rebellion *topics*, which all have values above 2.5. Crime is not strongly correlated with any of the other five *topics* examined here. It is the only *topic* that has a strong seasonality: it has an $R$-squared value of around 0.20 against year and multi-year lags of itself, meaning that about one fifth of its variation is explained by the season. The local peak in this *topic* in the ninth and tenth lunar months largely accounts for this seasonal variation, and it is clearly a product of the autumn assizes noted above. While Crime shows a high degree of seasonality, it has the lowest degree of persistence: the $R$-squared values of its time-lag regressions against itself decay very quickly. There is an increase in Crime starting in the
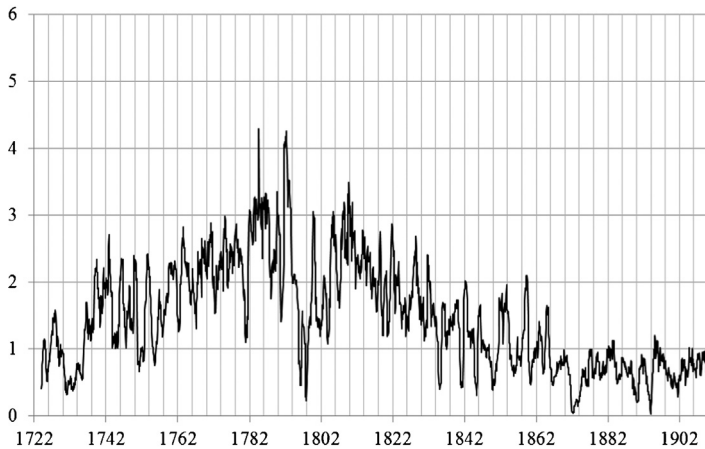
Fig. 1. Crime topic proportion over time (12 month moving average of monthly sum).

Qianlong Reign (mid to late eighteenth century) that is in keeping with a general increase in the number of entries in the *Veritable Records* in this period. After that, it shows a rather sharp drop-off around 1800, partial resurgence, and gradual secular decline over the course of the dynasty.

A review of Crime indicates that this *topic* captures the state's record-keeping of its law and order capacity, at least as pursued at the highest level of the bureaucracy. Most of the *words* in this *topic* are contained in the well-defined ontology of the legal code, and further analysis confirms the impression that this *topic* largely captures the everyday aspects of the emperor's direct jurisprudence. Given the contingencies of observing, reporting and recording crime and the relative clarity of the legal ontology, the specific variations in this *topic* probably tell us more about changes in state institutions and the capacity of the legal system than about the underlying crime rate.

## 3.2. Unrest

Unrest is dominated by *words* related to pursuing and capturing people disruptive to the social order, as well as terms detailing the organization (*huo* and *dang*—gang, *tu*—follower) and activities of criminals and malcontents. Like Crime, *documents* with high levels of Unrest encompass a wide variety of crimes and criminals. Some of these incidents are not violent, like illegal lumbering (*toukan*) or stealing transport grain, but even these are major crimes against state institutions. The vast majority of documents featuring high *topic* proportions in Unrest relate to group violence and large-scale disruption of law and order: robbery (*qiangduo*), kidnapping (*luxie*), resisting taxes (*jubu kanglian*) and armed feuds (*xiedou*). At face value, this *topic* is the counterpoint to Crime: Unrest entries are about catching criminals, and Crime is about judging and punishing them. However, a closer reading of the related documents suggests some clear differences. While both *topics* address a range of criminal behavior, Unrest has proportionally more large-scale violence. Unrest also has high frequencies of words that do not fit easily into legal ontologies, while some of these are present in the legal code, terms like *tou* and *jian* (illicit) and *qiang* (forceful), as well at the oft-referenced *zei* and *fei* (bandit) are highly contextual.

An appraisal of the temporal patterns substantiates this impression of difference between Crime (Fig. 1) and Unrest (Fig. 2). Unrest has a broadly similar pattern to Crime—moderate
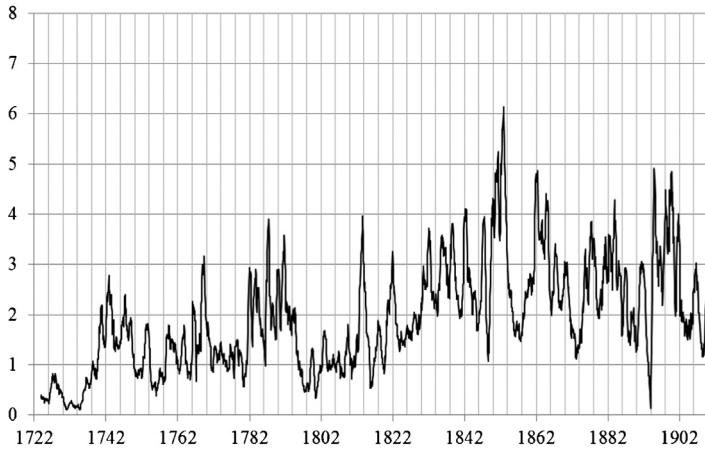
Fig. 2. Unrest topic proportion over time (12 month moving average of monthly sum).

variance around a fairly consistent mean throughout the period—with a standard deviation of 1.76. But Unrest is more persistent than Crime: it has $R$-squared values of above 0.15 against its one month and two month lags, and above 0.05 through its nine month lag. In rough terms, one sixth of the incidents recorded by *documents* dominated by this *topic* would last two months, and one in twenty would last nine months. And where Crime has a slight downward trend, Unrest generally has an upward trend. Unrest is also somewhat correlated with rebellion–including both Rebellion and Major Rebellion *topics*. Unrest is causally linked to Major Rebellion for at least two years ($R$-squared between 0.05 and 0.06 for all time lags between one month and two years) and to Rebellion for a year ($R$-squared of 0.05 for all time lags up to a year). Major Rebellion, but not Rebellion, is also somewhat causal of Unrest for up to six months—the regression of Unrest with lag of Major Rebellion up to six months has $R$-squared between 0.05 and 0.07. This suggests that major unrest could become rebellion, and rebellion could leave pockets of violence behind. It is interesting, but perhaps not statistically rigorous, to note that the violence from the really major rebellions of the nineteenth century left persistent violence for as much as six months, but that the administration seems to have been successful in quashing the aftershocks of smaller rebellions, which do not appear to lead to lasting violence.

These patterns further supports the impression that this *topic* is capturing state responses to unrest. It identifies records of a wide variety of events, at a variety of stages. It mostly addresses large-scale violent incidents and references semi-regular state policing, but records of non-violent unrest and other stages of state response also have high *topic* proportions in Unrest. I have theorized (Section 1.2) that the sort of large-scale violence reported in these *documents*— violence that could potentially lead to rebellion—was precisely the sort of event most likely to be reported by secret memorial. The secular rise in Unrest should therefore be interpreted more as an indication of social change than of changing patterns of state response—since this type of event would be reported regardless.

### 3.3. Sedition

From the list of common *words*—mostly terms referring to capture, interrogation and trial of criminals—this *topic* appears to be a generic indicator of law and order issues, much like the the two

*topics* above (Sections 3.1 and 3.2). A closer examination of documents with high Sedition topic proportions suggests that it is related to seditious use of language or symbols. Crimes like printing forbidden books (*nishu*) or posters (*jietie*), teaching heresy (*xiejiao*) or cutting off queues (*ge bian*) are captured by this *topic*. Sedition has major spikes in 1747, 1753, 1768, 1775 and 1787, all during the reign of the Qianlong Emperor. The second of these Sedition spikes is described in a full-length monograph; it relates to a rash of queue-cutting by supposed sorcerers (Kuhn, 1990). The hairstyle was mandated for all male subjects, and cutting queues incited Qianlong's paranoia about slights to his sovereignty. The flash of Sedition in 1775 probably shows his repeated crackdowns on language critical of the dynasty (Woodside, 2008, pp. 289–293), while the spike in 1787 is associated with the Lin Shuangwen uprising, which provoked a major crackdown on secret societies (Ownby, 1996, pp. 82–95). The *topic* generally declines after Qianlong's aggressive promotion of dynastic and religious order, but does not entirely disappear. The high frequency of relative terms like evil (*xie*) or rebellious (*ni*) suggests that the application of this *topic* depended closely on changing perceptions of what was illicit or heterodox. Unlike Unrest however, this shifting ontology probably had more to do with changing imperial visions of orthodoxy than with attempts to understand the general class of violent threats to social order.

It is curious to note that Sedition has essentially no relationship to any of the ''rebellion'' *topics*. Of the major rebellions in the mid- to late Qing (addressed below in Section 3.4), only the Lin Shuangwen Uprising is accompanied by an upswing in concern over sedition. This is quite in contrast to David Ownby's argument that the heterodox symbols of the Heaven and Earth Society only became specifically rebellious *after* the crackdown on the Lin Shuangwen Uprising (Ownby, 1996, pp. 26, 110–26), as well as Berand ter Haar's assertion that the characterization of groups as heterodox increased *after* the 1768 queue-cutting incident (ter Haar, 1992, pp. 248–250). In fact essentially all instances of rebellion, including the major events of the late nineteenth century, are accompanied by rising Unrest but not rising Sedition. This impression from the graph in Fig. 3 is borne out in the statistics: regressions of Sedition against Unrest, Rebellion, Major Rebellion and Border Rebellion and time lags of these *topics* all lack conclusions of statistical significance. The only correlation with Sedition is Crime, which has an $R$-squared value of 0.10. This suggests that this *topic* is capturing imperial *interest* in crimes of
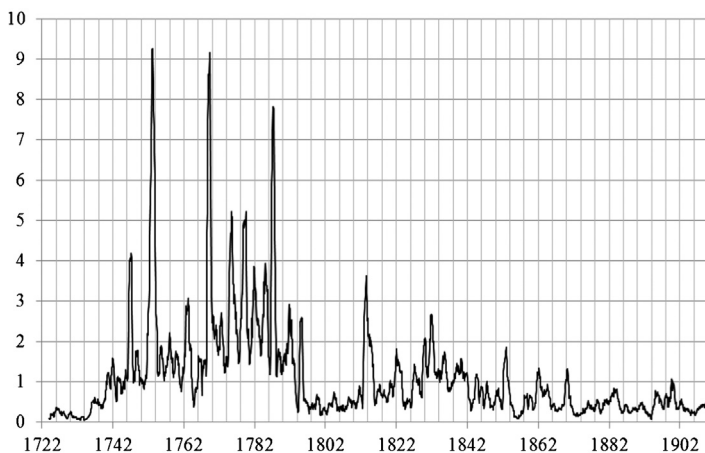


Fig. 3. Sedition topic proportion over time (12 month moving average of monthly sum).

sedition—language interpreted as strictly anti-dynasty—rather than a broader sense of heterodoxy in society at large. In fact, there is no *topic* in the model that accounts for religious heterodoxy, so we must assume that these issues either did not attract the same level of attention from the central court or were not discussed using a unified discourse over time. Sedition, in this sense, presents as a rather narrow *topic* that almost exclusively accounts for the Qianlong Emperor's obsession with perceived threats to his sovereignty. Qianlong stands out from all other emperors for his use of terms like ''bandit'' or ''rebellious'' in reference to crimes of discourse—whether or not they were violent in nature.

## 3.4. Border rebellion

From the list of most common *words*—primarily related to the military, rebels and non-Han people—it is clear that this *topic* is about warfare along the borders. On closer inspection, it is specific to a particular type of border wars. It has peaks in the middle and late eighteenth century when the Qing Empire expanded its reach into Central Asia, but these maxima highlight wars along the existing borders of the Qing state, not the campaigns that extended Qing rule into Central Asia. The Qing fought a series of border wars in the northwest in 1723, 1747–1749 and 1771–1776, and in the southwest in 1765–1769 (Burma), 1788–1789 (Vietnam), and in 1790–1792 (Tibet). All these wars were related to the Qing protecting existing interests in regions under its administration. Around the same time, there were a series of offensive campaigns in Xinjiang (1755–1759—Perdue, 2005; Woodside, 2008, pp. 50–52) that are not picked up by this *topic*. In Fig. 4, we can see two *topics* identifying different language used in the offensive wars in Xinjiang (Western Campaigns, spiking in the late 1750), and the conflicts seen to be suppressing rebellions in the near northwest and the southwest in the late 1740s, 1770s, 80s and 90s (Border Rebellion). This identifies an important distinction: unlike the wars in Xinjiang, the wars in the near northwest, and in particular in the southwest, were all seen as suppressing rebellions—reactions to eighteenth century efforts to stabilize and unify aboriginal areas and protect existing Qing interests (Woodside, 2008, pp. 253–281). At the same time, this *topic* does not capture the minority revolts in the late nineteenth century. The particular language used to record two other major rebellions in border areas—the Panthay Rebellion in the southwest (1856–1873) and the Dungan Revolt in the northwest (1862–1877)—are captured by the two *topics* below (Section 3.5),
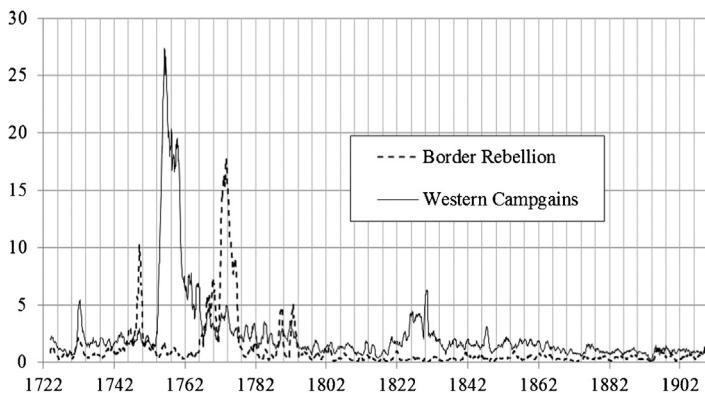


Fig. 4. Border Rebellion and Western Campaigns topic proportions over time (12 month moving average of monthly sum).

not by Border Rebellion, further indicating a major shift in the language and practice in the eighteenth and nineteenth centuries.

A look at the temporal patterns of Border Rebellion reinforces the sense that it accounts for a very different phenomenon than the two other rebellion *topics*. With a standard deviation of 2.54, Border Rebellion has a higher temporal variability than any of the crime-related *topics*, but less than the other forms of rebellion. It is equally persistent to the other rebellion *topics*, with *R*-squared values above 0.30 on its self-lag regressions to one year and above 0.10 to three years. But Border Rebellion is not correlated with either Rebellion or Major Rebellion. All of these campaigns—the eighteenth century border rebellions, the Xinjiang campaigns and the late nineteenth century border rebellions—ended up being major military operations against non-Han peoples. It would appear that the projects to incorporate and regulate the frontiers in the eighteenth century were the product of a very different environment than either the wars further west around the same time or the suppression of rebellions in similar areas in the late nineteenth century. This suggests that it is accurate to view the state-making of the Yongzheng and Qianlong emperors as very different processes than the attempts to preserve unity in the late Qing.

## 3.5. Rebellion and major rebellion

These two *topics* capture much the same phenomena, described in essentially the same language. In addition to sharing vocabulary, there is substantial overlap in their temporal aspect of their *topic* proportions, especially in the nineteenth century (see Fig. 5). In fact, a closer reading of *documents* containing these *topics* casts little light on the differences between them. It appears that instances of Major Rebellion contain more reference to the movements of large groups of rebels (*cuan*) and the extensive planning (*chou*) involved in fighting major campaigns against them. In addition, *documents* with high proportions of Major Rebellion tend to refer more to the use of militia (*yong*), a key difference between the eighteenth- and nineteenth-century responses to uprisings (Kuhn, 1970). Even this distinction is more a matter of degree than a clear-cut difference.

Examination of the temporal evidence (see Fig. 5) casts a bit more light on the situation: Major Rebellion is almost exclusively a nineteenth century *topic*, starting with the White Lotus Rebellion in the 1790s and accounting for the dynasty-shaking events of the 1850s and
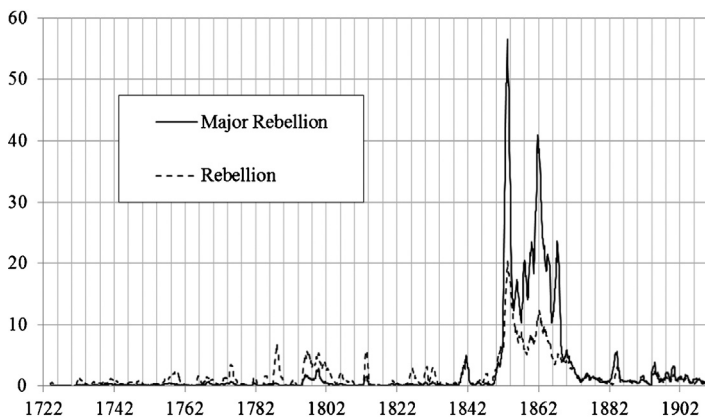


Fig. 5. Rebellion and Major Rebellion topic proportions over time (12 month moving average of monthly sum).

1860s—the Taiping and Nian Rebellions and the contemporary minority revolts, and the Boxer Rebellion in the 1890s. On the other hand, Rebellion picks up *all* of the known revolts throughout the period—including the major nineteenth-century rebellions, smaller-scale eighteenth-century rebellions, as well as several rebellions have not been topics of extensive research. These two *topics* are highly correlated with each other (with an *R*-squared value of 0.59), as well as with each other's time lags (above 0.27 on lags of up to two years). In the case of the late dynasty rebellions, it is clear that the same *incidents* are often described in *documents* with high Rebellion proportions as well as in *documents* with high proportions of Major Rebellion. In the 1860s, Rebellion simply accounts for those responses to these crises that were most similar to earlier instances, while Major Rebellion accounts for the records with the greatest differences. The higher instance of militia in Major Rebellion suggests that they were more important in later rebellions. Otherwise, the difference between the two *topics* is simply one of degree. Rebellion has a standard deviation of 3.12, higher than any of these six *topics* except for Major Rebellion, which boasts a staggering value of 7.61! Rebellion and Major Rebellion are both strongly persistent—much more so than the other *topics*. Major Rebellion is highly correlated with its self-lag for five years and beyond (*R*-squared of .42 on one-year lag, 0.30 on two-year lag, 0.20 on five-year lag); Rebellion is only slightly less so (0.39 at one year, 0.27 at two years, 0.09 at five years).

How can we interpret these ways in which these *topics* are intertwined? We should expect that any level of unrest that rose to a certain level came to be seen as full-on rebellion. The weak but significant causal relation between Unrest and the two rebellion *topics* suggests that one in twenty uprisings was significant enough to be considered "rebellion," but it should be stressed that there is no hard ontological boundary between the two. What about the difference between the two rebellion *topics*? It appears that Rebellion identifies a fairly consistent language related to state response to violent uprisings—capturing any unrest that was considered very serious by the state, while Major Rebellion tracks the institutional changes and new uses of language brought on by the still higher threat and changing social circumstances posed by the nineteenth century rebellions as compared to earlier uprisings. Did all rebellions after the White Lotus pose a larger existential threat to the dynasty, or did they merely introduce a new vocabulary of violence and illegitimacy that remained in use for the rest of the dynasty? Existing studies (e.g., Kuhn, 1970; Perry, 1980) suggest that this shift is more than just new language—that there was a rise in the level of violence in society that raised the threat level posed by any instance of unrest. This issue cannot be fully addressed here, but poses an interesting question for future research.

## 4. Conclusions

What does this evidence tell us about crime, banditry and unrest in Qing China? I would argue that these early modern "crime rates" tell us about both the underlying levels of criminal violence and about the patterns of crime reporting; they indicate something about the occurrence of revolts and about the preoccupation with sedition and rebellion. As a source base, the *Veritable Records* compound social and textual phenomena, and it is impossible to completely disentangle the evidence about Qing society from the evidence about Qing government, its language and reporting procedures. In fact, in almost every case, the changes in discourse cannot be separated from the changes in society and government that they describe. I will address several tentative conclusions about Qing state and society in more detail below. First, however, it is worth commenting on the success of the model.

Despite a rather simple approximation of the semantics of the Chinese language—including the highly problematic use of single characters as words—the LDA topic model used in this study shows great promise. The *topic proportions* examined above identify every major rebellion in the eighteenth and nineteenth centuries with spikes in the histograms. Furthermore, the periods of rebellion regarded as ''bigger'' or ''more threatening'' have correspondingly high peaks—a vindication of the use of *topic proportion* sums across *documents* as a unit of analysis. The *model* allows a clear distinction between *topics* describing endemic problems (Crime, Unrest), and those describing event-driven problems (the other four), while also identifying the relationships between the ordinary ''economy of violence'' and the extraordinary outbreaks of rebellion— including the correlation between Unrest and Rebellion. Not only does this correspond with the past several decades of research, it allows new analysis of the broader patterns of crime, unrest and rebellion. Now for the conclusions.

First, the level of state concern with unrest—and probably the overall level of violent upheaval—increased over the course of the eighteenth and nineteenth centuries, but not as much as previous research (e.g., Buoye, 2000; Kuhn, 1970; Perry, 1980) would indicate. The reporting of incidents involving mass unrest shows a small but significant secular increase during this period; at the same time, the more general law-and-order functions of the court show a small but significant secular decline. The increase in the everyday economy of violence may have been even greater than that indicated by Unrest precisely because the decline of state capacity to enforce policing functions in the provinces may have also reduced the rate at which these crimes were reported. On the other hand, the rise in Unrest may have been predicated on an increase in state concern with even ordinary banditry and small riots due to the threat posed by the nineteenth century rebellions.

Second, the Qing state's centralized surveillance process was more concerned with mass violence than other types of major criminal behavior. Local officials appear to have been relatively faithful in reporting the appearance of large, armed bands of criminals, and these reports generally attracted attention from higher-level administrators, who reported them by secret memorial. The court was less involved in the direct administration of other crimes, and was willing to leave most—but not all—prosecution to the lower levels of government. The trial records present in the *Veritable Records* tell us very little about the overall world of crime—as only serious crimes show up with any regularity. On the other hand, the *Records* give a more comprehensive picture of the scope of large-scale violence in early modern society, and the ways in which this violence could and did bleed into upheaval and rebellion.

Third, both the border wars of the seventeenth century and the cataclysmic upheavals of the mid-to-late nineteenth century were distinct phenomena that are only weakly described by the broader discourse on rebellion. The many border rebellions of the 1740s–1790s appear to have been the unique product of the Yongzheng and Qianlong Emperor's visions of empire, as seen both in their policies at the frontiers and the language used to record resistance. From the perspective of the imperial court, the threat of the nineteenth century rebellions—the White Lotus Rebellion of the late 1790s, and especially the Taiping and Nian Rebellions of the 1860s— was fundamentally different from earlier rebellions, and a new language developed to record these events and the state's response to them.

Fourth, as far as we can tell based on central court records, the only social phenomenon predictive of rebellion was mass violence, and this was only weakly predictive. The fluidity of language belies a great deal of confusion over how ''rebellion'' differed from mere ''rioting'' or ''banditry.'' Or perhaps the overlapping terminology for ''bandits'' and ''rebels'' reveals an understanding that they were frequently the same people, yet even from the state's perspective

there were many ''bandits'' who never rebelled. As far as we can tell based on this reading, state bureaucrats were right to closely report unrest, and the emperor was right to take notice—as this type of violent, mass action could and sometimes did turn into rebellion. But these administrators could never fully account for what made some forms of illicit violence more threatening than others.

Fifth, we can say very little about the role of ideology in shaping unrest. State surveillance reported little to no ''sedition'' prior to major uprisings. The volume of Sedition in the *Records* seems to have much more to do with imperial attempts to project orthodoxy than with any particular social phenomenon. The Qianlong Emperor generated a particularly impressive volume of verbiage on sedition, this probably reflects his attempts to project a particular vision of orthodoxy under which many phenomena were classed as perverse. Later emperors showed less interest in this project of cultural hegemony. This does not mean that ideology was never involved in rebellion, merely that the state apparatus was either unconcerned with ideology or poorly equipped to find and report it. When particular incidents surfaced, especially when the obsessive Qianlong Emperor demanded investigation into heterodoxy, officials could and did produce reports. At other times, these crimes were both easily hidden from local administrators and of little interest to these belabored bureaucrats.

This statistical approach to the corpus has reinforced and modified, rather than overturning existing narratives of crime and rebellion in the Qing. The general picture does not change: the mid- to late-Qing was defined by rising violence and declining state capacity, punctuated by the catastrophic events of the 1860s and a partial recovery thereafter. This is not an entirely new picture, but this analysis casts some light on how central administrators thought about crime, rebellion and unrest in the eighteenth and nineteenth centuries. Crime was largely a routine matter; it was discussed at court on particular schedules, was relatively easy to define, and merited less attention during the dynastic crisis. Acts of mass violence or unrest were more likely to attract attention from the central imperium. Rebellion was always a critical matter, and the nineteenth century seems to have generated a new level of concern, accompanying the larger scale of uprisings.

The more comprehensive reading does give a better sense of the ways official language described and reacted to the social setting. The language that officials used for *people*—terms like ''bandit'' (*zei* or *fei*) —is quite diffuse and applied in very different circumstances, with high frequency in all six *topics* examined here. The language they used to describe the *behavior* of these people is more specific, and parses more easily into individual *topics*. Strikingly, the ontological uncertainty of individual terms for these people seems to reflect the sociological uncertainty over whether these people would become full-fledged rebels or remain local thugs. The lack of clarity in language reflects uncertainty in the society it described.

Distant reading is no substitute for closer approaches to the individual events. This foray into the *great unread* has revealed as many new problems as it has solved. An expert reading a small number of texts is able to apply context and domain knowledge to parse linguistic phenomena— changing word uses—from social phenomena—changing patterns of events. This specialized knowledge is particularly helpful in understanding descriptions of well-known events. Context and expertise are much harder to apply to distant reading, especially when much of the corpus has yet to be closely examined by specialists. In this exploration, I have applied a combination of statistics, historiography and logic to try to understand the ways in which the ''bandit problems'' occlude our understanding. It is reassuring that my conclusions parallel those of other historians, and promising that they offer new perspectives, but these will ultimately need to be tested by future researchers. Hopefully, they will prompt further exploration of the *great unread*.

**Acknowledgements**

**Appendix. Top frequency words by topic**

Crime
*ni* 拟 – reference [precedent], draft
*zui* 罪 – crime
an 案 – case
*xing* 刑 – punishment, sentence
*bu* 部 – board, ministry
*shen* 審 – try, interrogate
*li* 例 – precedent, sub-statute
*zhao* 照 – reference [precedent]
*zhe* 斬 – behead
*si* 死 – die
*yi* 議 – discuss
*jiao* 絞 – hang by the neck
*shang* 傷 – wound, injure
*ding* 定 – decide
*ming* 命 – life, command
*gou* 勾 – checkmark (as in approve a judgment)
*jian* 監 – prison
*lü* 律 – legal code
*ou* 毆 – beat to death
*sha* 殺 – kill
*gai* 改 – correct
*li* 理 – principle, manage
*bi* 斃 – execute
*qiu* 秋 – autumn (as in autumn judgments)
*shi* 氏 – surname, maiden name

Unrest
*huo* 獲 – capture
*na* 拏 – capture
*fei* 匪 – bandit
*yan* 嚴 – strict
*fan* 犯 – violate, criminal
*di* 地 – location
*bu* 捕 – arrest

*ji* 緝 – arrest
*dao* 盜 – thief, robber
*chi* 飭 – order
*an* 案 – case
*tu* 徒 – follower
*xian* 縣 – county
*jie* 劫 – rob
*shou* 首 – head, chief
*guan* 官 – official
*cheng* 懲 – punish
*qiang* 搶 – rob
*tao* 逃 – escape
*min* 民 – the people
*jing* 經 – pass through, undergo
*zi* 滋 – spread, provoke
*du* 督 – supervise
*zei* 賊 – bandit, rebel
*bian* 弁 – low-ranking officer

Sedition
*huo* 獲 – capture
*na* 拏 – capture
*xun* 訊 – investigate
*gong* 供 – confession
*an* 案 – case
*jiao* 教 – teach, religion
*jie* 解 – send under guard, separate, explain
*yan* 嚴 – strict
*shen* 審 – try, interrogate
*jiu* 究 – investigate
*jing* 經 – pass through, undergo
*ni* 逆 – traitor, rebel, betray
*li* 李 – common surname, plum
*xian* 現 – present
*sheng* 省 – province
*xian* 縣 – county
*wang* 王 – common surname, king
*zhang* 張 – common surname, spread out
*liu* 劉 – common surname
*jia* 家 – family
*mi* 密 – secret
*shou* 首 – head, chief
*que* 確 – true
*wang* 往 – toward, proceed
*fang* 訪 – visit, inquire, investigate

Border Rebellion
*bing* 兵 – military

*fan* 番 – foreign, Tibet
*zei* 賊 – bandit, rebel
*lu* 路 – road, circuit
*chuan* 川 – Sichuan, river valley
*tu* 土 – ground, local, tribal administration
*jun* 軍 – military
*a* 阿 – transliterative character
*zang* 藏 – Tibet
*er* 爾 – transliterative character
*di* 地 – place
*gong* 攻 – attack, invade
*gui* 桂 – Guilin, laurel bay tree
*jin* 金 – gold, metal, transliterative character
*wang* 往 – toward, proceed
*jiao* 剿 – wipe out, destroy (specifically of bandit camps)
*guan* 官 – government
*ji* 機 – opportunity, machine
*li* 力 – power
*si* 司 – department, tribal administration
*ying* 營 – camp
*dai* 帶 – lead troops, bring
*xian* 現 – current
*zhong* 眾 – crowd, masses
*zhu* 駐 – halt, be stationed at

Rebellion
*zei* 賊 – bandit, rebel
*bing* 兵 – military
*fei* 匪 – bandit
*jiao* 剿 – wipe out, destroy (specifically of bandit camps)
*ni* 逆 – traitor, rebel, betray
*cheng* 城 – city, walls
*gong* 攻 – attack, invade
*jun* 軍 – military
*li* 力 – power
*fu* 復 – again, repeatedly
*cuan* 竄 – flee, exile, move (of bandits generally)
*yong* 勇 – brave, militia
*guan* 官 – government
*jing* 經 – pass through, undergo
*dai* 帶 – lead troops, bring
*ji* 擊 – attack, beat
*lu* 路 – road, circuit
*huo* 獲 – capture
*sheng* 勝 – victory, triumph, win
*ke* 克 – overcome, subdue, capture
*du* 督 – oversee

*bao* 保 – guard, protect
*shou* 首 – head, chief
*shu* 數 – number, many
*sha* 殺 – kill

Major Rebellion
*bing* 兵 – military
*jiao* 剿 – wipe out, destroy (specifically of bandit camps)
*jun* 軍 – military
*zei* 賊 – bandit, rebel
*fei* 匪 – bandit
*fang* 防 – guard, protect, prevent
cuan 竄 – flee, exile, move (of bandits generally)
dai 帶 – lead troops, bring
*li* 力 – power
*chi* 飭 – instruct
*xian* 現 – current
*lu* 路 – road, circuit
*bei* 北 – north
*zhou* 州 – sub-prefecture, part of many city names
*ni* 逆 – traitor, rebel, betray
*yong* 勇 – brave, militia
nan 南 – south
*sheng* 省 – province
*dong* 東 – east
*cheng* 城 – city, walls
*pai* 派 – group, clique, dispatch, appoint, send
*yan* 嚴 – strict
*sheng* 勝 – victory, triumph, win
*xun* 迅 – investigate
*du* 督 – oversee

# References

Allee, M.A., 1994. Law and Local Society in Late Imperial China: Northern Taiwan in the Nineteenth Century. Stanford University Press, Stanford, CA.

Blei, D.M., Ng, A.Y., Jordan, M.I., Lafferty, J., 2003. Latent Dirichlet allocation. Journal of Machine Learning Research 3, 993–1022.

Buoye, T.M., 2000. Manslaughter, Markets, and Moral Economy: Violent Disputes over Property Rights in Eighteenth Century China. Cambridge University Press, Cambridge, UK.

Cheng, J.C., 1963. Chinese Sources for the Taiping Rebellion, 1850–1864. Hong Kong University Press, Hong Kong.

Chesneaux, J., Bianco, L., 1972. Popular Movements and Secret Societies in China, 1840–1950. Stanford University Press, Stanford, CA.

Cohen, P.A., 1984. Discovering History in China: American Historical Writing on the Recent Chinese Past. Columbia University Press, New York.

Esherick, J., 1987. The Origins of the Boxer Uprising. University of California Press, Berkeley, CA.

Greenshields, M., 1994. An Economy of Violence in Early Modern France: Crime and Justice in the Haute Auvergne, 1587–1664. Pennsylvania State University Press, University Park, PA.

Harrison, J.P., 1969. The Communists and Chinese Peasant Rebellions: A Study in the Rewriting of Chinese history. Studies of the East Asian Institute. Atheneum, New York.

Hung, H., 2011. Protest with Chinese Characteristics: Demonstrations, Riots, and Petitions in the Mid-Qing Dynasty. Columbia University Press, New York.

Kuhn, P.A., 1970. Rebellion and its Enemies in Late Imperial China: Militarization and Social Structure, 1796–1864. Harvard University Press, Cambridge, MA.

Kuhn, P.A., 1990. Soulstealers: The Chinese Sorcery Scare of 1768. Harvard University Press, Cambridge, MA.

Liu, K.-C., 1981. World view and peasant rebellion: reflections on post-Mao historiography. Journal of Asian Studies 40, 295–326.

McCallum, A., 2002. MALLET: A Machine Learning for Language Toolkit. , http://mallet.cs.umass.edu.

Macauley, M.A., 1998. Social Power and Legal Culture: Litigation Masters in Late Imperial China. Stanford University Press, Stanford, CA.

Murray, D., Baoqi, Q., 1994. The Origins of the Tiandihui: The Chinese Triads in Legend and History. Stanford University Press, Stanford, CA.

Naquin, S., 1976. Millenarian Rebellion in China: The Eight Trigrams Uprising of 1813. Yale University Press, New Haven, CT.

Ownby, D., 1996. Brotherhoods and Secret Societies in Early and Mid-Qing China: The Formation of a Tradition. Stanford University Press, Stanford, CA.

Perry, E.J., 1980. Rebels and Revolutionaries in North China, 1845–1945. Stanford University Press, Stanford, CA.

Perdue, P.C., 2005. China Marches West: The Qing Conquest of Central Eurasia. Belknap Press of Harvard University Press, Cambridge, MA.

Qing, S., 2000. (Veritable Records of the Qing Dynasty), digital edition, 2000. Hanji dianzi wenxian ziliao ku (Electronic database of Chinese language textual resources) Academica Sinica, Taiwan Available at: www.hanchi.ihp.sinica.edu.tw.

Robinson, D.M., 2001. Bandits, Eunuchs, and the Son of Heaven: Rebellion and the Economy of Violence in Mid-Ming China. University of Hawaii Press, Honolulu.

Sommer, M.H., 2000. Sex, Law, and Society in Late Imperial China. Stanford University Press, Stanford, CA.

Tagliacozzo, E., 2005. Secret Trades, Porous Borders: Smuggling and States Along a Southeast Asian Frontier, 1865–1915. Yale University Press, New Haven, CT.

Teng, S., 1962. Historiography of the Taiping Rebellion. Harvard University Press, Cambridge, MA.

ter Haar, B.J., 1992. The White Lotus Teachings in Chinese Religious History. E.J. Brill, Leiden.

Theiss, J.M., 2004. Disgraceful Matters: The politics of Chastity in Eighteenth-Century China. University of California Press, Berkeley, CA.

Tong, J., 1991. Disorder under Heaven: Collective Violence in the Ming Dynasty. Stanford University Press, Stanford, CA.

Wilkinson, E., 2013. Chinese History: A New Manual. Harvard University Asia Center, Cambridge, MA.

Wong, R.B., 1982. Food riots in the Qing dynasty. Journal of Asian Studies 41, 767–788.

Wong, R.B., 2001. Tax resistance, economy and state transformation in China and Europe. Economics of Governance 2, 69–83.

Woodside, A., 2008. The Ch'ien-lung reign. In: Peterson, W.J. (Ed.), The Cambridge History of China. Vol. 9. Part 1, Ch'ing Empire to 1800. Cambridge University Press, Cambridge, UK, [electronic resource], pp. 230–309.

Yang, C.K., 1976. Some preliminary patterns of mass actions in nineteenth-century China. In: Wakeman, F., Grant, C. (Eds.), Conflict and Control in Late Imperial China. University of California Press, Berkeley, CA, pp. 174–210.

**Ian Matthew Miller** is a doctoral candidate in history and East Asian languages at Harvard University. His research focuses on the social and environmental history of Ming and Qing China and the use of digital humanities for classical Chinese texts.