Localization and Coordination: How Propaganda and Censorship Converge in Chinese Newspapers*

Margaret E. Roberts† and Brandon M. Stewart‡

October 7, 2014

Abstract

Increased commercialization of Chinese newspapers has surfaced two questions for scholars of Chinese media: to what extent does censorship still exist within newspapers in China, and what is the logic of propaganda in an increasingly marketized information environment? Despite the importance of these topics, the estimation of the prevalence and content of censorship and propaganda in Chinese media has been forestalled by the Sisyphean task of manually reading and annotating thousands of news articles that are published daily in China. In this paper, we offer the first unsupervised large-scale study of propaganda and censorship in the content of Chinese newspapers. Our study draws on a new full-text corpus of four million news reports spanning 31 provincial and city-level papers over the last seven years. We also develop a new estimation strategy for the Structural Topic Model to enable topical discovery in the texts. We uncover two strategies that the Chinese government uses to address politically sensitive events within China: localization and coordination. In the first strategy (localization), local papers report on sensitive events happening within their immediate area, but these reports are not be picked up by provincial level papers- effectively containing the spread of information about the event. In the second strategy (coordination), all papers report identical stories on sensitive events which nevertheless have high public awareness. This coordination reduces citizens' access to different perspectives on the event.. Thus we find that the Chinese government uses synchronization of news stories not only for traditional propaganda, but also as a form of censorship that reduces the amount of disagreement on sensitive but widely-known issues.

*Our thanks to Ken Benoit, Dan Hopkins, Kyle Jaros, Gary King, Brendan O'Connor, Jennifer Pan, and Dustin Tingley for helpful comments. Our thanks to Antonio Coppola and Amy Jiang for excellent research assistance. The Chinese Newspaper corpus described here was collected by a joint effort with Kyle Jaros and Jennifer Pan.

†Assistant Professor, Department of Political Science, University of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093 http://margaretroberts.net

‡Ph.D. Candidate, Department of Government, 1737 Cambridge Street, Harvard University, Cambridge MA 02134; http://scholar.havard.edu/bstewart

1 Introduction

To what extent does censorship and propaganda still exist in Chinese news media? What form does it take and what does it target? In a marketized media environment, newspapers no longer completely reflect Chinese government strategy as newspapers pander to their audiences. As the influence of government on news media occurs behind closed doors, differentiating between market-driven and state-driven news is a continual challenge of Chinese media scholars, particularly those who seek to understand variation in how papers follow central government strategy or the logic behind propaganda (Qin, Wu and Strömberg, 2012; Huang and Li, 2013; Stockmann, 2012; Brady, 2009). While evidence exists that the Propaganda Department in China still enforces coordination of propaganda within newspapers and that newspapers still omit some major events, the frequency and nature of these strategies are a matter of much debate.

Up until now, adjudicating between arguments about how much the state still has a hand in Chinese media has been limited by the difficulty of estimating the extent and targets of censorship and propaganda within newspapers. Estimating censorship and propaganda strategies in China requires sifting through the entirety of newspapers to determine whether reporting has been coordinated across all papers (evidence of propaganda), or if event omission is consistent across all papers (indications of censorship). Because this is a virtual impossibility, scholars have instead relied on sampled reading of newspapers, counts of keywords, and interviews with journalists and editors of newspapers to back-out strategies of media control. While each of these strategies points to important trends in Chinese media, they rely on a-priori assumptions of what is controlled and where the locus of that control originates.

We fill this gap by conducting the first unsupervised large-scale study of propaganda and censorship in Chinese newspapers. We collect provincial and city-level Chinese newspapers from 2007-2013 and compare coverage across a corpus of over four million news report spanning 31 newspapers in China. Leveraging recent advances in stochastic variational inference (Hoffman et al., 2013), we develop a scalable estimation procedure for the Structural Topic Model (Roberts et al., 2013) which enables us to estimate hundreds of topics for our 500 million word collection on a standard desktop. To estimate censorship, we identify changes in media reporting surrounding large-scale, sensitive events on the ground in China. To estimate propaganda, we identify moments of coordination where all papers reprint the same news story.

In doing so, we uncover two main strategies the Chinese government uses to report on politically sensitive events within China. For locally sensitive events the news relies on *localization* where only papers in the immediate affected area will address the event, likely to retain credibility. Despite their potential appeal to audiences outside of the local area, these events are not picked up by higher-level papers. Second, the government relies on *coordination* for politically sensitive events that have already gained wide-spread public attention. Such coordination reduces the number of perspectives on the event, advocating for the perspective devised by the government.

Interestingly, coordination of news articles that pertain to nationally sensitive events suggests that censorship takes the form not only of omission, but also of synchronization. Put another way, propaganda is not just for promoting positive stories about the government, but also for enforcing one particular view of negative news and filling newspapers during sensitive time periods. Thus propaganda and censorship lie on one connected continuum. By measuring coordination, our method reveals not only heart-warming stories about government deeds, but also the most important concerns and underlying strategies of the Chinese government. It suggests an alternative motivation for the continued use of propaganda: reducing the scope of potential viewpoints and room for dissent within Chinese public sphere.

The paper is structured as follows. First, we discuss the current puzzles within the Chinese media literature about censorship and propaganda within Chinese newspapers and outline how the estimation of the Chinese government's media strategy could speak to this literature. Then we describe our data and data collection process. Third, we introduce the scalable Structural Topic Model (STM). We then summarize our findings and the logic we hypothesize underlies the Chinese governments' media strategy. We conclude with a discussion of future research.

2 Newspaper Commercialization and Media Control in China

Over the past ten years, newspapers in China have undergone rapid commercialization (Stockmann, 2012). The state, unable to continue financing all newspapers and responding to public demand for more information, encouraged the partial privatization of newspapers, reducing regulation of papers as well as encouraging newspapers to find their own sources of funding. Reporting in China has therefore become increasingly market-driven.

Commercialization of the media market in China has segmented the market into different types of newspapers. Stockmann (2012) identifies a spectrum of newspapers within China, ranging from official to commercialized. Official newspapers are artifacts of the Leninist political system, associated with different levels of state government, and are still run by the state. These newspaper exist at the national, province, city, county, township, and village levels. They are distributed only within the area of their political entities' jurisdiction and their content is managed by the political entity to which they belong. However, they still retain funds from sales of newspapers and from advertising and therefore have incentives to increase their readership.

Commercialized newspapers are not associated with a particular bureaucracy within government, and tend to be new entrants into the market. These newspapers are more privately owned (they may be up to 49% privately owned) and tend to be more driven by the market, generating revenue entirely through the market. However, like their official counterparts, commercialized papers are still subject to directives by state propaganda units.

For all newspapers, propaganda directives issued to newspapers can contain both instructions to omit certain stories and instructions to print particular articles. For news that casts a negative light on the national government, directives will instruct provincial newspapers not to print or emphasize these stories. On the other hand, for news that the Party wants everyone to read, provincial newspapers are often instructed to reprint articles from the national newswire *Xinhua* (Brady, 2009). This synchronization is thought by scholars to be the main mechanism through which propaganda is distributed in Chinese newspapers.

The incentive-structure for newspapers presents a quandary where newspapers have to please the public and the government, both who have very different preferences. While consumers of news notoriously have a preference for negative stories (Trussler and Soroka, 2014), the Chinese Propaganda department is known to prefer that newspapers report on positive stories that reflect well on the government (Stockmann, 2012; Brady, 2009). Such conflicting preferences make it difficult to identify which articles are propaganda and which events are omitted by design.

Although scholars of newspapers in democracies have argued that newspaper ownership and competition are the primary determinants of media coverage (Besley and Prat, 2006; Djankov et al., 2001), the result of commercialization on media bias within China has been decidedly mixed. While commercialization has encouraged newspapers to pander to their audiences and to report on local government scandals (Tong and Sparks, 2009; Tong, 2011; Qin, Wu and Strömberg, 2012), it has also buttressed the perceived independence of the Chinese media, despite continuing central government control over reporting on sensitive issues (Stockmann, 2012). Whether the majority of media reporting is dictated by consumer preferences or top-down control and how newspapers set the agenda is still a subject of much debate.

Increased commercialization of newspapers in China has also created a puzzle of why propaganda still exists in an information age. Evidence suggests that propaganda is sometimes sufficiently transparent that citizens do not believe in it (Chen and Shi, 2001). Stockmann (2012) finds that consumers of media do not turn to official sources on certain topics specifically because they doubt the credibility of official sources. Huang and Li (2013) argues that there could be alternative motivations for propaganda, including signaling the strength of the Chinese government, and finds that while ideological articles do not change Chinese citizens' perceptions of the government, they do change Chinese citizens' perceptions of how powerful the government is.

2.1 Our Contribution

A missing empirical approach in these preceding studies is a large-scale estimation of topical content in Chinese newspapers. Up until now, measuring censorship and propaganda in newspapers has been constrained by the physical impossiblity of manually sifting through newspapers to determine both omission of events and coordination across newspapers. Using a corpus of 4 million newspaper articles, we discover continued strategies of censorship in Chinese newspapers. We also find that propaganda is as likely to be used during sensitive events and on sensitive topics than it is on topics that reflect well on the Chinese government, implying continued rationale for propaganda within newspapers in China. We turn to our data and analysis in the next section.

3 Newspaper Dataset

We collect articles from official newspapers within China at the provincial and cities level. Provincial and city newspapers are run by their respective governments and thus provide the most direct window into the sought after media strategy of the government. As with all newspapers in China, although much financing comes from the market, each newspaper is responsible for writing in accordance with national and local propaganda directives. Local and national propaganda departments also have the authority to screen newspaper articles yet to be published, and also frequently review articles after publication to ensure that the papers are following directions (Brady, 2009).

The key to our analysis is that the extent of coordination or glaring omissions across newspapers will reveal aspects of the government's strategy. In the absence of national or local direction, newspapers should reflect local popular opinion, which will likely vary substantially by province. We would therefore expect that the majority of glaring omissions in the topics across all newspapers and coordinated printing across all newspapers would be the result of central government directives. We use this insight to estimate censorship and propaganda within the newspapers.

Our corpus of newspapers includes all printed news articles from 31 different provincial and city papers scraped from the web. We do not scrape from typical newspaper websites, which contain ads and may contain articles not included in the print form. Instead, we collect the articles from the "print" website version for each of these newspapers, which contain the newspaper articles actually printed in the newspaper in database-like format. Not only does this make the newspaper content easier to collect, it also reflects the articles actually printed within the paper and does not contain online advertisements.

In addition to the text of the post, we collected the date of the newspaper article and the page of the article, the author of the article, and keywords, and pictures associated with the article when available. Not all papers contain all years, so we collect as many years as available. Our entire corpus contains 3.9 million articles. A table of the papers and their date ranges is provided in Table 1.

Statistical text analysis in Chinese careful segmentation of the text into distinct tokens. We use the suite of tools available in the Stanford Natural Language Processing toolkit (Chang, Galley and Manning, 2008). For an exposition of the issues involved in processing non-English texts for analysis we defer to Lucas et al. (2013). Before analysis we also remove extremely rare words, a standard set of Chinese stop words, and a number of stray markup language tags. After segmentation and stopword removal, we process the text into a term-document matrix that indicates words within each of the newspaper documents.

Newspaper	Date Range
Beijing	2011-09-02 to 2013-10-30
Changchun	2009-09-15 to 2014-05-30
Chengdu	2009-08-01 to 2014-05-30
Chongqing	2007-12-27 to 2013-10-30
Fujian	2011-01-01 to 2013-10-30
Fuzhou	2013-08-01 to 2014-06-20
Gansu	2009-11-10 to 2013-10-30
Guangdong	2007-11-20 to 2013-10-30
Guangxi	2008-11-08 to 2013-10-30
Guangzhou	2007-08-01 to 2014-05-30
Guilin	2008-09-02 to 2014-05-30
Hainan	2008-03-01 to 2013-10-30
Heilongjiang	2007-10-08 to 2013-10-30
Henan	2007-10-14 to 2013-10-29
Hengyang	2009-09-27 to 2013-11-04
Hunan	2009-01-13 to 2013-10-29
Jiangxi	2007-03-19 to 2013-10-30
Jilin	2011-07-01 to 2013-10-30
Jincheng	2008-01-01 to 2014-06-24
Kunming	2008-10-01 to 2014-05-30
Lanzhou	2009-08-03 to 2014-05-30
Liaoning	2008-03-11 to 2013-10-30
Macao	2013-05-01 to 2013-10-30
Nanchang	2011-08-01 to 2014-05-30
Nanjing	2012-03-14 to 2014-06-24
Inner Mongolia	2008-01-01 to 2013-10-28
Qinghai	2013-05-22 to 2013-10-30
Shandong	2011-05-15 to 2013-08-28
Tibet	2008-06-15 to 2013-10-30
Yunnan	2008-11-19 to 2013-10-30
Zhejiang	2006-01-01 to 2013-10-30

Table 1: Newspapers and corresponding dates ranges

4 Methods for the Scalable Structural Topic Model

In order to examine strategies of propaganda and censorship, we need a measure of the topical content of newspaper articles. Due to the size of the corpus, we turn to statistical methods for automated content analysis which provide low-dimensional summaries of the text (Grimmer and Stewart, 2013).

We use an unsupervised mixed-membership approach called the Structural Topic Model

(Roberts et al., 2013). While effective inference is possible for small to moderate size corpora the scale of our problem requires us to develop a new estimation strategy for the model.

In Section 4.1 we provide a brief overview of the model and provide the rationale for its use in this application. We next outline the particular model specification choices (4.2). We then describe the modifications necessary to allow inference at the scale of millions of documents (4.3). Readers uninterested in the statistical and technical innovations of the paper can skip to Section 5 where we discuss results.

4.1 Structural Topic Model

The Structural Topic Model (STM) is a mixed membership model in the style of latent Dirichlet Allocation (LDA). STM differs from LDA in being specifically designed to facilitate the analysis of how document metadata (such as newspaper and date of publication) drive the prevalence of latent topics and the content of those topics. The method is perhaps best understood by analogy to the simpler LDA model. We provide a very general description here and refer the reader to (Roberts, Stewart and Airoldi, 2014) for technical details.[1]

LDA is a generative model of text documents which uses the bag-of-words representation, where each document is represented as a vector containing the counts of each distinct vocabulary word contained in the document (Blei, 2012). These counts are modeled as arising from a fixed number topics which are distributions over words. Each document is itself a mixture over topics where the mixing weights indicate the proportion of words within the document expected to come from each topic. The LDA framework is different from the single membership models previously used in political science where each document is assumed to come from exactly one topic (Grimmer, 2010; Quinn et al., 2010).

[1] Roberts, Stewart and Airoldi (2014) is the primary technical manuscript with details on the estimation procedure directed at statisticians. Roberts et al. (Forthcoming) contains an explanation oriented to political scientists with extensive simulation evidence and applications in open-ended survey response. Roberts et al. (2013) provides a compact technical introduction to the method and applications. All manuscripts as well as an R package implementing the model are available at `structuraltopicmodel.com`.

These models focus on modeling the hierarchical structure of the corpus by nesting documents within author (Grimmer, 2010) and time unit (Quinn et al., 2010).

STM provides a framework for specifying hierarchical structures for the data using document-level covariates. Covariates enter the model in two ways: *topic prevalence* and *topical content*. Topic prevalence covariates model the expected frequency with which a topic is discussed, in our case allowing us to capture variation in topic use attributable to time trends within each newspaper. Topical content covariates allow systematic differences in word use within a topic, in our case capturing city/province specific rates of word use. Both types of covariates induce a partial pooling of the parameters that allow us to model corpus structure.

More formally the model can be described by the following generative process

$$
\begin{aligned}
\sigma_k^2 &\sim \text{InvGamma}(1,1) \\
\gamma_{k,p} &\sim \text{Normal}(0,\sigma_k^2) \\
\theta_d &\sim \text{LogisticNormal}(\mathbf{X}_d\gamma_k,\Sigma) \\
z_{d,n} &\sim \text{Multinomial}(\theta_d) \\
w_{d,n} &\sim \text{Multinomial}(\beta_{d,k=z_{d,n}}) \\
\beta_{d,k,v} &= \frac{\exp(m_v + \kappa_v^{\cdot,k} + \kappa_v^{Y_d,\cdot} + \kappa_v^{Y_d,k})}{\sum_v \exp(m_v + \kappa_v^{\cdot,k} + \kappa_v^{Y_d,\cdot} + \kappa_v^{Y_d,k})} \\
\kappa_i &\sim \text{Laplace}(0,\tau_i)
\end{aligned}
$$

where we denote the number of topics K, observed words $w_{d,n}$, topic prevalence covariates $\mathbf{X}$ topical content covariates $\mathbf{Y}$, and m_v the empirical log probability of word v. Figure 1 summarizes the data generating process.

4.2 Model Specification

The STM model requires specification of the number of the number of topics, a design matrix for topical prevalence and a partition over documents for topical content. We set

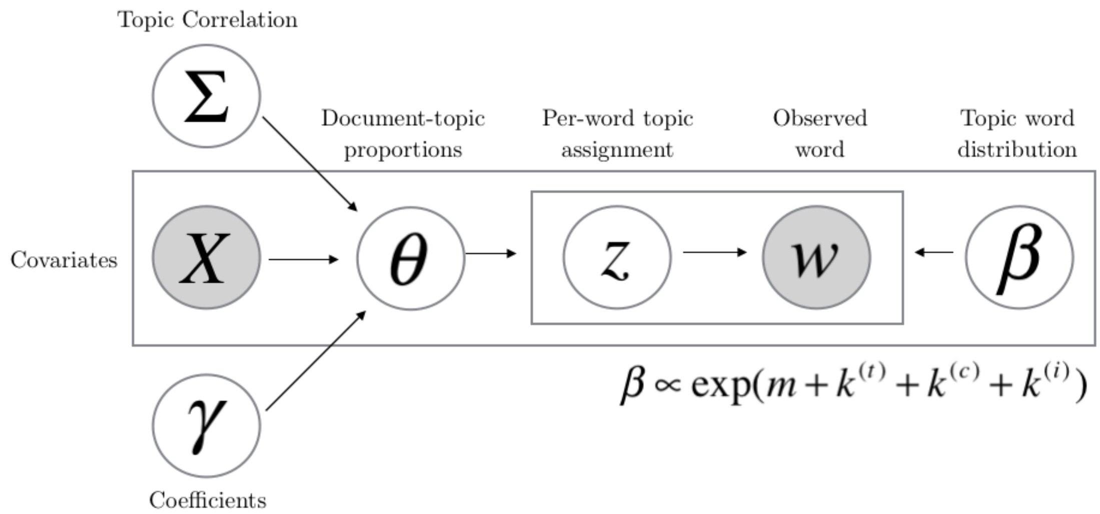

Figure 1: The Structural Topic Model

the number of topics to $K = 300$. This choice was made heuristically by examining model fits with increasing numbers of topics until a suitable granularity of the data achieved. More systematic approaches are a subject of future research.

For the topic prevalence portion of the model we want to allow for paper specific time trends. To capture these trends we use piece-wise cubic B-splines at 50 evenly spaced knots over the 2861 days in our coverage interval. These spline bases are interacted with newspaper indicators allowing for each paper to exhibit a unique time trend pooled towards the global mean. The coefficients are drawn from a zero mean Normal distribution with a topic-specific precision parameter drawn from a Gamma(1,1) prior. This applies a mild regularization which draws the effects to zero.

The topical content model allows for word rates to differ across city and provincial levels. Thus for a given topic k the distribution is formed by combining the baseline word rate (m), topic specific effect for the word ($\kappa^{\cdot,k}$), newspaper effect κ^{Y_d} and the topic-specific newspaper effect $\kappa^{Y_d,l}$:

$$\beta_{k,v} = \frac{\exp(m_v + \kappa_v^{\cdot,k} + \kappa_v^{Y_d,\cdot} + \kappa_v^{Y_d,k}}{\sum_v \left(\exp(m_v + \kappa_v^{\cdot,k} + \kappa_v^{Y_d,\cdot} + \kappa_v^{Y_d,k}\right)}$$

The κ parameters are estimated with the sparsity promoting L_1 penalty (equivalently MAP estimation under a Laplace prior).

4.3 Scalable Inference

The scale of the document corpus presents substantial challenges for inference. The sparse word count representation of the data is sufficiently large that it cannot even by held in active memory on a standard desktop. Our `R` package for the STM uses a fast partially-collapsed Variational EM algorithm in which local latent variables for each document are updated in the E-step and global topic parameters are updated in the M-step. Because each document in the collection must be visited during the E-step the inference procedure could take days to complete a single iteration. Furthermore this batch procedure is wasteful; we don't need to iterate over every document in order to effectively update the global parameters.

To allow for inference at this scale we use stochastic variational inference recently developed in Hoffman et al. (2013) for conjugate exponential family models. Due to nonconjugacy and the expensive normalization in calculating β the standard SVI is not directly applicable. We describe our adapted approach below which is summarized in its entirety in Figure 2.

Stochastic Variational Inference Batch, or standard, variational inference finds the parameters of a tractable approximate posterior which minimizes the Kullback-Liebler divergence to the true posterior of interest (Grimmer, 2011; Wainwright and Jordan, 2008). To do so we maximize the Evidence Lower Bound (ELBO) which is a lower bound on the log marginal probability of the data formed by Jensen's inequality. Thus inference becomes an optimization problem. For conjugate exponential families and factorized approximating distributions, a simple iterative procedure of closed-form updates can be shown to provably converge to a local optimum.

In stochastic variational inference (SVI) we optimize the ELBO using stochastic gra-

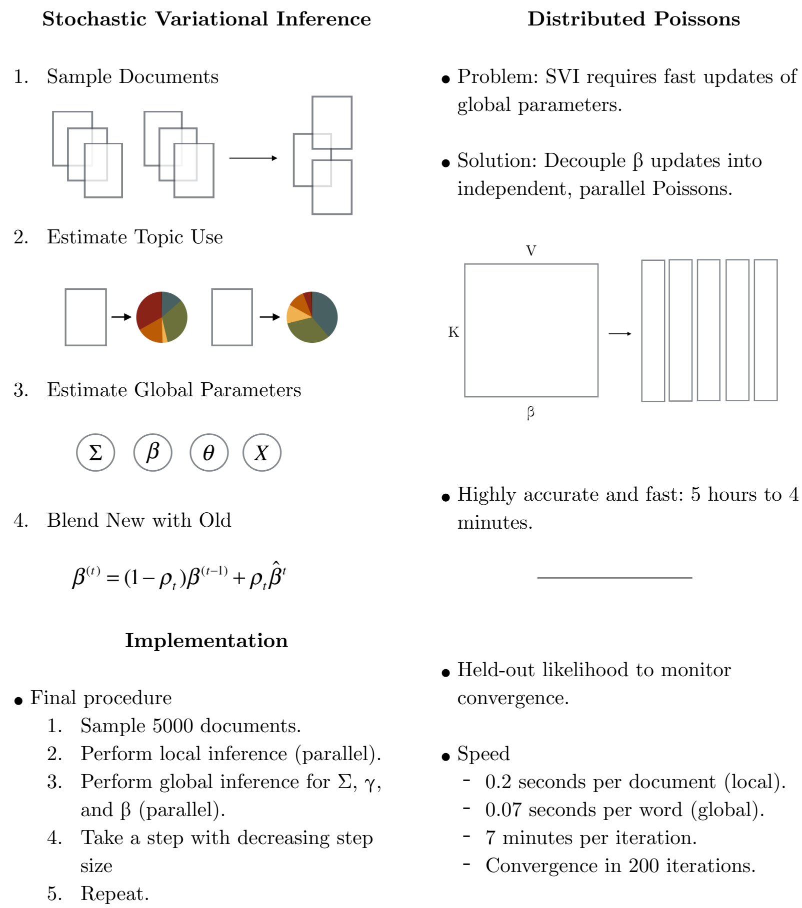

Figure 2: A summary of our stochastic variational inference scheme

dient descent (Robbins and Monro, 1951; Hoffman et al., 2013). Using a subsample of the documents, or minibatch, we form an unbiased estimate of the natural gradient of the global parameter space. We then take a step in this direction and repeat. As long as the sequence of step sizes ρ_t meet the necessary conditions ($\sum \rho_t = \infty, \sum \rho_t^2 < \infty$) then the procedure will converge to a local optimum (Robbins and Monro, 1951). Convergence guarantees also hold if we multiply the gradient by a preconditioning matrix (so long as that matrix is positive definite with bounded eigenvalues).

Hoffman et al. (2013) show using the properties of exponential families that preconditioning by the Fisher information leads to updates with a simple and intuitive structure. Denote our global parameters at iteration $t-1$ as $\beta^{(t-1)}$. We take a subsample of the data and form an estimate of the global parameters reweighting the subsample to equal the number of data points in the full dataset. Denoting this estimate $\hat{\beta}^t$, the update step of size ρ_t which follows the (noisy) natural gradient can be given as:

$$\beta^{(t)} = (1 - \rho_t)\beta^{(t-1)} + \rho_t\hat{\beta}^t$$

which is simply the convex combination of the original estimate with our approximation based on the sub-sample. With each iteration we take steps of decreasing size. Here we use the step schedule given in Hoffman et al. (2013)

$$\rho_t = (t + \tau)^{-\kappa}$$

where the forgetting rate is controlled by κ (set here to .9) and the delay is controlled by τ is set to 1.

The resulting procedure scales well to large document corpora because we can update the global parameters without first needing to iterate through all the documents. When the global updates are fast, this can result in remarkable speed gains. Furthermore we need not be able to hold all the documents in memory. We only need the global parameters and the documents within the current subsample.

Limitations of Prior Work The stochastic variational inference strategy outlined above is designed only for conjugate exponential family models. It also implicitly assumes that the global parameter updates are fast to compute relative to update the local latent variables. Neither of these situations holds in our model. The logistic normal is not conjugate to the multinomial likelihood. Furthermore computing the update for the topical content parameters (κ) involves an expensive normalization involving a log-sum-exp term over the vocabulary.[2]

The non-conjugacy arises both in updates for the local document-level latent variables as well as the global parameters. As in the batch algorithm for STM we use the

[2]Concurrently with the development of this work numerous extensions have been proposed for particular cases of nonconjugacy. However, few approaches have dealt with the issue of the expensive M-step calculation. A notable exception isRabinovich and Blei (2014) which takes a different stylistic approach to the one used here.

Laplace approximation for the local nonconjugacy (Wang and Blei, 2013). For the global parameters, we could simply follow the gradient of the ELBO without preconditioning. However, this runs the risk of inefficient use of the sample information. Instead of attempting to identify the optimal preconditioning matrix we simply solve for the optimal global parameters of the reweighted sub-sample data and update the parameter using the convex combination formulation as though it were conjugate. This implicitly solves for and uses the optimal preconditioning matrix.

In the STM model the global parameters are the parameters controlling topic prevalence (coefficients γ which parameterize the mean of topic prevalence and topic covariance matrix Σ) and the parameters controlling topic content (κ). Updates for γ and Σ involve standard conjugate updates for a multivariate linear regression model. A few computational tricks are employed to allow for efficient reweighting of the data but computations are essentially standard. We address the updates of κ next.

Topical Content Updates In our batch variational algorithm for STM with content covariates, we compute MAP estimates for the topic parameters κ under sparsity promoting prior. The update is equivalent to a multinomial logistic regression with an L1 penalty which we solve using the cyclical coordinate ascent algorithm in the `glmnet` package (Friedman, Hastie and Tibshirani, 2010). Estimation is slow due to the high dimensional parameter space and the expensive normalizing constant in the softmax function. With a larger vocabulary such as the one used here, the κ update using `glmnet` can take hours to compute.

This bottleneck is problematic for the SVI algorithm which implicitly assumes that global updates will be extremely fast to compute relative to the local updates. Indeed the algorithm does nothing to decrease the scale of the global inference problem- all the parameters must still be updated. Furthermore while the document level updates can be trivially parallelized, the global updates are coupled, requiring serial calculation.

To combat this problem we use the Distributed Multinomial Regression framework of Taddy (2013). The central idea is to use the Multinomial/Poisson dual formulation to approximate the complete multinomial regression using a set of V separate poisson regressions (one for each item in the vocabulary). The formulation uses a plugin estimator for the observation-level fixed effects that induce the coupling in the multinomial likelihood. Taddy (2013) shows that this simple approximation is effective even for extremely adversarial cases.[3] Not only does the Poisson formulation allow for faster computations, the decoupling of the parameters allows for each of the V regressions to be computed in parallel.

We compute each of these modified Poisson regressions using the cyclical coordinate ascent algorithm in `glmnet`. The algorithm computes an entire regularization path of penalty parameters. We choose the penalty parameter using an Akaike information criterion (Taddy, 2013).[4]

Implementation and Estimation We base our implementation on the existing `stm` R package for batch variational inference making use of the `parallel` package for multicore processing (R Core Team, 2014). In each iteration we take a random sample of 5000 documents from the collection and estimate the local variational distribution for each in parallel. These local updates take around 0.2 seconds per document, with the 5000 document collection taking anywhere from 125 to 300 seconds for the collection on a 20-core machine. Global updates are dominated by the cost of updating κ with each of the 4761 poisson regressions taking approximately 0.07 seconds. These updates take anywhere from 325 to 400 seconds on a 20-core machine. The average cost of a single iteration

[3] *Ex ante*, we would expect the approximation to break down when the multinomial observation has few total counts distributed across the categories. Examples in Taddy (2013) show accurate computations even when the total observation counts are 1 per observation. In our case each multinomial observation is the sum over all counts for a particular newspaper topic combination, meaning the counts are generally at least in the thousands.

[4] While note a purely Bayesian approach, the information criterion provides a rough approximation to the marginal likelihood. An area of future research is to more directly connect this approach to MAP estimation under a proper prior for the Laplace parameter. We opt for the penalized likelihood approach to take advantage of the incredibly fast estimation algorithms in the `glmnet` package. In practice the regularization path is giving us a discrete approximation to the local posterior which we could use to perform numerical integration over the penalty parameter in the style of Rue, Martino and Chopin (2009).

is 481 seconds (8 minutes). The entire model fit takes approximately 24 hours which involves approximately 200 iterations.

We initialize the algorithm using a deterministic spectral algorithm for LDA (Arora et al., 2012). See Roberts, Stewart and Tingley (N.d.) for a justification of spectral algorithm in this setting and simulation studies indicating strong performance. We emphasize that while the initialization is stochastic unlike the batch setting the resulting scalable algorithm is stochastic due to the subsampling.

To assess convergence we follow (Hoffman et al., 2013) in monitoring expected log-likelihood on a fixed held-out set. At the beginning of the procedure we set aside 5000 documents which are never used in updating parameters. These documents are themselves split into half. At each iteration the first half of the document is used to estimate the document's mixture over topics (θ) and then the second half is used to assess the predictive log-likelihood of the estimated topics. Because we only need a subset of the local parameters and they are computed on documents which are half as short on average, the heldout metric takes on 60-90 seconds to computer per iteration (2-3 times as fast as the E-step on a comparably sized step).

After completing the inference procedure we take a final pass through all documents to obtain estimates of the topic proportions for every document in the corpus. These form the basis of our analysis.

4.4 Summary of Methodological Contributions

In order to facilitate inference of our multi-million article newspaper corpus we developed a new estimation strategy for the structural topic model. Our approach builds on prior work in stochastic variational inference, extending and adapting these methods to the difficult setting of our model.

Due to the heavy use of parallelism, it is difficult to include the methods as described here in our `R` package `stm` and guarantee that they will work on all machines. We have however included numerous features which are inspired by the design here including the distributed poisson estimation of content covariate models as well as memoized inference procedure which provides some of the benefits of stochastic estimation in the batch setting.[5]

5 Results

We estimate 300 topics from the newspapers newspapers. For each document, we estimate the proportion of the document in each of these 300 topics. Using these estimates, we look for glaring omissions and coordination of articles as indications of propaganda and censorship within Chinese newspapers.

5.1 Localization of Negative Content in Chinese Newspapers

At first glance, we unsurprisingly notice the existence of topics that would positively reflect on the Chinese government. For example, there is a topic about Communist Party heroes, including the words “communist”, “hero”, “noble”, “ancestors”, “role models”, and including such people as Yangshan Zhou, the Baoshan party secretary who made the 2011 list of models for good politicians. Other topics describe about having good values and working hard as Party members, with words such as “values”, “dedication”, “diligence”, “responsibility”, “party”, and “red”.

On the other hand, papers do frequently address long-standing problems within Chinese society. Our model uncovers three pollution topics, one specifically focused on air pollution with words such as “particulate matter”, “pollution”, “coal”, “heavy metals”, and “excessive”; one focused on water pollution with words such as “water”, “water supply”, “drinking”, “samples”, “pass rate”, and “quarantine bureau”; and a general pollution topic that takes on “garbage”, “sewage”, “emissions”, “water pollution”, and “carbon dioxide”.

[5]The core idea in memoized inference (Hughes and Sudderth, 2013) is to divide the data into a set of groups and update global parameters after completing each set. This is more memory intensive but leads to faster convergence for datasets with large numbers of documents.

Similarly we find four topics about corruption, including one about accepting bribes ("bribe", "accept", "corruption", "bribery"), one about fighting corruption ("fighting corruption", "honest politics", "inspection committee"), one about using public money to fund private events, and one about ethics in politics. A close reading of articles within both the environment and corruption topics reveals that newspapers do not just talk about these topics broadly, but also mention specifics of these events.

In contrast to pollution and corruption, we find very little discussion of specific protest events in China. We don't find a topic directly addressing protests, although we do find a few topics related to stability and conflict, which are euphemisms for conflict. A closer reading of these articles showed that they are typically addressing stability in general, without referring to particulars of protest.

Which papers tend to report on negative topics and which seem to ignore them? We label topics that are associated with "sensitive" events, like pollution or corruption in the corpus. We then estimate the differences in reporting between papers on these topics. As shown in Figure 3, we find that city-level papers contain far more negative topics than provincial level papers. In particular, provincial papers focus more on pollution and corruption topics, which are largely ignored in higher-level papers despite their appeal to larger audiences.

This is consistent with a strategy of localization, where negative events are addressed in the specific locality, but not picked up by higher level papers. Whereas in a country with newspapers driven by the market, negative topics would be selected for higher level papers, provincial papers in China seem to report more on topics that reflect positively on the government.

We then study the content of newspapers during sensitive events in China. King, Pan and Roberts (2013) find that online censorship in China is focused mainly on protest events, to stop the spread of information about collective action. But how do newspapers deal with large-scale protest events happening in the very city where the newspaper is set on reporting?

Figure 3: Differences in Topic Proportions on Negative Events Between Provincial and City Papers

To study this, we collect information about protest events that occurred within the cities in our dataset. We find 16 protest events in locales that correspond to the papers we are studying. Each of these protest events was large enough to merit online and international news attention, which is how we identified them. We find that only two out of sixteen of these protests were covered by the newspapers. In each case, the protest event was only covered by the city-level newspaper and not picked up by the provincial-level newspaper. In one case, the city-level paper reported on the protest 10 days *after* the event occurred. This points to large-scale omission of protest events within China. Only when protest events are large enough (and perhaps resolved enough) do papers report on them, and at that, only at the very localized, not national level.

5.2 Coordination of Negative Content in Chinese Newspapers

Next, we study the prevalence of propaganda within Chinese newspapers. As we know that propaganda in China takes the form of coordinated articles across newspapers, we estimate coordination of papers on a daily level between the newspapers of focus. To do this, we need enough newspapers within the time period to make sure that coordination is not simply a coincidence, but a larger propaganda strategy of the Chinese government. Since the greatest overlap of our papers is between the end of 2011 and the beginning of 2013, we use this time period as the focus of our study of propaganda.

5.2.1 Measuring Coordination

Newspaper coordination in China does not mean perfect plagiarism, in that when the Propaganda Department sends out an article for reprint, the newspapers do not have to copy the reprint directly. We found that reprints often come with a header or new title that the newspaper creates itself, and sometimes a new concluding paragraph. In addition, some papers will choose to remove various paragraphs within the article that was issued by the Propaganda Department. Two example reprints are included in Figure 4, one entitled "Carry forward the popularity of Marxism with Chinese characteristics" and the other entitled "New circumstances continue to advance the Party's advanced nature". Even though the articles contain two different titles and different numbers of paragraphs, the majority of the text between the two articles is identical. As a result, our objective for a measure of coordination is not a perfect match between newspaper articles, but rather a similarity measure that uncovers newspaper articles where large portions of the text are identical.

For each day within this time period, we subset the newspaper dataset to articles published on that day. For each of these newspapers, we estimate the level of similarity between the topical vectors that resulted from the STM output for each article. We use two measures of similarity between these topical vectors. First, we use cosine similarity between the topical vectors of each article to determine similarity. Cosine similarity measure the angle between two vectors a and b and is defined as:

$$\cos(a,b) = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n}(a_i)^2}\sqrt{\sum_{i=1}^{n}(b_i)^2}}$$

南方日報 高度决定影响力 返回南方报网

2011年6月29日 星期三 返回首页 版面导航 标题导航

3 上一篇 放大 缩小 默认

胡锦涛强调充分运用党建历史经验不断推进新形势下党的先进性建设

推进马克思主义中国化时代化大众化

据新华社北京6月28日电 中共中央政治局6月28日下午就中国共产党保持和发展党的先进性研究进行第三十次集体学习。中共中央总书记胡锦涛在主持学习时强调，加强党的先进性建设，是加强和改进党的建设的长期任务和永恒课题。我们要充分运用党的建设历史经验，及时总结党的建设新鲜经验，深化对加强和改进新形势下党的建设的规律性认识，继续以改革创新精神全面加强和改进党的建设，不断推进党的先进性建设。

中央党史研究室曲青山教授、中国人民大学马克思主义学院杨凤城教授就这个问题进行讲解，并谈了他们的意见和建议。

中共中央政治局各位同志认真听取了他们的讲解，并就有关问题进行了讨论。

胡锦涛在主持学习时发表了讲话。他指出，新形势下推进党的先进性建设，必须深刻认识和充分运用坚持马克思主义基本原理同中国具体实际相结合、坚持科学理论指导、坚定不移走自己的路的历史经验，始终把马克思主义作为自己的指导思想，坚持解放思想、实事求是、与时俱进，坚持用马克思主义立场观点方法观察当今世界、观察当代中国，准确把握时代脉搏，不断推进马克思主义中国化时代化大众化，使党的理论和实践始终体现时代性、把握规律性、富于创造性，使党的路线方针政策不断与时俱进，保证党始终站在时代发展的前列和中国社会发展进步的潮头。必须深刻认识和充分运用紧紧依靠人民群众、诚心诚意为人民谋利益、从人民群众中汲取前进力量的历史经验，坚持立党为公、执政为民，坚持以人为本，把最广大人民根本利益作为党全部工作的出发点和落脚点，深入做好群众工作，办好顺民意、解民忧、惠民生的实事，不断实现好、维护好、发展好最广大人民根本利益，使党始终得到人民群众支持和拥护。必须深刻认识和充分运用自觉加强和改进党的建设、不断增强党的创造力凝聚力战斗力、永葆党的生机活力的历史经验，坚持党要管党、从严治党，紧紧围绕党的中心任务，以改革创新精神全面推进党的思想建设、组织建设、作风建设、制度建设和反腐倡廉建设，着力加强党的执政能力建设和先进性建设，不断解决好提高党的领导水平和执政水平、提高拒腐防变和抵御风险能力这两大历史

上一版 下一版

胡锦涛在中共中央政治局第三十次集体学习时强调

新形势下不断推进党的先进性建设

要充分运用党的建设历史经验，及时总结党的建设新鲜经验，深化对加强和改进新形势下党的建设的规律性认识，继续以改革创新精神全面加强和改进党的建设，不断推进党的先进性建设

中共中央政治局6月28日下午就中国共产党保持和发展党的先进性研究进行第三十次集体学习。中共中央总书记胡锦涛在主持学习时强调，90年来，我们党在加强和改进自身建设实践中积累的历史经验，对推进党的先进性建设具有长期指导作用，必须十分珍惜并发扬光大。加强党的先进性建设，是加强和改进党的建设的长期任务和永恒课题。我们要充分运用党的建设历史经验，及时总结党的建设新鲜经验，深化对加强和改进新形势下党的建设的规律性认识，继续以改革创新精神全面加强和改进党的建设，不断推进党的先进性建设。

胡锦涛在主持学习时发表了讲话。他指出，面对深刻变化的国际国内形势，面对改革发展稳定的繁重任务，面对各族人民过上更好生活的新期待，面对党员队伍和党的建设遇到的新情况新问题，我们党肩负着新的历史责任，面临着新的重大考验。执政时间越久，改革开放和社会主义市场经济发展越深入，国内外环境变化越深刻，越要高度重视保持和发展党的先进性，越要把党的先进性建设作为一项重大战略任务切实推进，使党始终成为思想上政治上组织上完全巩固、站在时代前列带领人民团结奋斗、开拓进取的坚强领导核心。

胡锦涛指出，新形势下推进党的先进性建设，必须深刻认识和充分运用坚持马克思主义基本原理同中国具体实际相结合、坚持科学理论指导、坚定不移走自己的路的历史经验，始终把马克思主义作为自己的指导思想，坚持解放思想、实事求是、与时俱进，坚持用马克思主义立场观点方法观察当今世界、观察当代中国，准确把握时代脉搏，不断推进马克思主义中国化时代化大众化，使党的理论和实践始终体现时代性、把握规律性、富于创造性，使党的路线方针政策不断与时俱进，保证党始终站在时代发展的前列和中国社会发展进步的潮头。据新华社

下一篇4

Figure 4: Example of Two Reprinted Articles

Cosine similarity measures the similarity between topical proportions, weighting topics with higher topic proportions more than those with lower topic proportions. This measure is important to us because we want to make sure that the main topic of the text has close to identical weight in order to consider the article a reprint. However, for shorter documents, we found that almost perfect measures of cosine similarity would count two newspaper articles as reprints when they were actually not because shorter documents have fewer highly weighted topics. Therefore, in addition to cosine similarity, we used Hellinger distance, which puts weight on the entire distribution of the document (Blei and Lafferty, 2007). Hellinger distance is a distance between distributions and is defined as:

$$\text{Hellinger}(a, b) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^{n} (sqrt(a_i) = \sqrt{(b_i)})^2}$$

For two articles to count as reprints of one another, we required a cosine similarity measure of at least .99 and a Hellinger distance of less than .2. We found the combination of cosine similarity and Hellinger distance between the topical vectors of articles most closely uncovered reprints of newspaper articles. To verify, we randomly sampled pairs of newspaper articles from our dataset. Our measure recovered coordination in 100% of the sample pairs.

5.2.2 Topics of Coordination

How prevalent is propaganda in Chinese newspapers? Which topics are most likely to be coordinated across Chinese newspapers? We estimate coordinated clusters that contained at least 15 newspaper reprints on the same day, or articles that were reprinted in at least half of the newspapers. Newspaper articles are coordinated on this level 232 times within the time period, or about once every four days.

Figure 5 plot the topics associated with these coordinated clusters. Unsurprisingly, we find that many of the clusters are related to topics we would typically think were propaganda. For example, many of the newspaper articles are coordinated on CCP History, or on slogans associated with party propaganda.

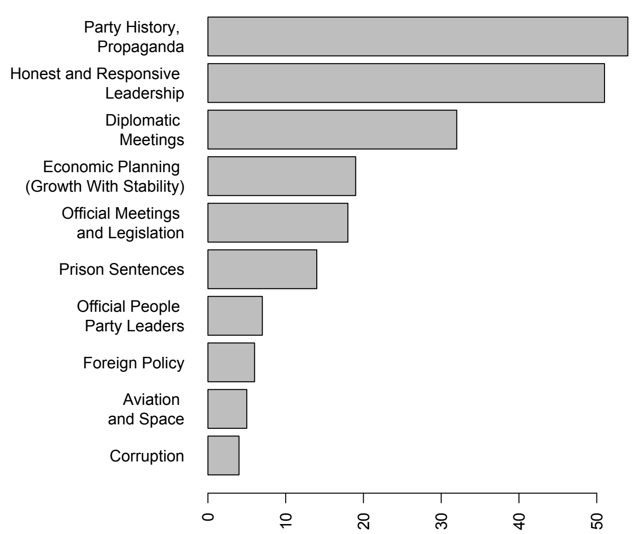

Figure 5: Coordinated Topics

However, interestingly we also find topics that are related to extremely negative and sensitive events within Chinese politics. In particular, a large portion of coordinated articles in this sample focus on the jailing of high level CCP officials, evidence of internal division and corruption within the upper-echelons of the CCP. If we take a closer look at these events (Table 2) related to the prison topics, we find that the majority of these coordinated events are related to the downfall of former Chongqing mayor Bo Xilai. Others include the downfall of extremely high level CCP officials including Lai Changxin and Liu Zhijun. This suggests that even though city-level papers take on the most reporting about corruption in general, on the most high-level cases, the central government uses coordination instead of localization to control information about the event.

Date	Event
2012-02-14	Lai Changxin – smuggling and bribery
2012-07-27	Gu Kailai – murder of Neil Heywood
2012-08-10	Gu Kailai trial murder of Neil Heywood
2012-08-21	Gu Kailai trial again, in more detail
2012-09-06	Wang Lijun trial (indicted)
2012-09-19	Wang Lijun trial (details)
2012-09-25	Wang Lijun Sentence
2013-07-09	Liu Zhijun – Minister of Railways bribery
2013-08-23	Bo Xilai hearing – first day
2013-08-24	Bo Xilai hearing – second day
2013-08-25	Bo Xilai hearing – third day
2013-08-26	Bo Xilai hearing – conclusion
2013-09-23	Bo Xilai hearing – sentence
2013-10-26	Bo Xilai hearing – maintaining final sentence

Table 2: Description and date of coordinated events in the prison topic.

A closer look at the coordinated events related to foreign policy reveals the same pattern (Table 3). The most sensitive international events include international news that could stoke nationalist protest, including news about disputed islands in the South China Sea, weapons sales to enemies, or visits with the exiled Dalai Lama. We find that five out of six of the coordinated events about foreign policy are related to these types of international events. Such coordination prevents the emergence of alternative viewpoints of the correct course of action for the Chinese government and continues to place the blame on foreign governments.

Date	Event
2011-11-06	Greek government's difficulty coming to agreement, Euro crisis
2012-05-16	Firmly opposing British leader meeting with the Dalai Lama
2012-05-22	Russia states that countries should not meddle in South China Sea conflict
2012-11-02	Japan shouldn't pull other countries into the Diaoyu conflict
2012-12-24	Opposes U.S.-Japan security agreement because of Diaoyu Islands

Table 3: Description and date of coordinated events in the foreign policy topic.

5.2.3 Timeline of Sensitive Coordination

Certainly, not all coordinated events focus on negative or sensitive events, as evidenced by the existence of coordination about topics such as propaganda. However, even coordinated articles that contain what we would traditionally think of as propaganda seem to be crowding out discussion of other topics. In particular, the highest levels of coordination occur at time periods that are most sensitive for the Chinese government.

Figure 6 plots the number of coordinated news articles by day across the time period we study. The day with the most coordination happens to be on the day of leadership transition between Hu Jintao and Xi Jinping, considered by scholars to be the most sensitive day within our study. The second highest cluster of coordinated articles occurs in March of 2012 and 2013, during the meetings of the National People's Congress.

Interestingly, high levels of coordination between newspapers is correlated with what we know about background censorship during these time periods. During these three meetings, the Chinese government throttled websites, jailed activists, and increased censorship.[6]. Even though the coordinated articles during these events are almost completely what we would think was propaganda, the coordination mechanisms here is serving a similar role as censorship, throttling the publication of alternative stories by requiring the front page of newspapers to print stories related to party ideology and slogans. Propaganda in this case seems to be serving an identical role as censorship.

[6]See Wall Street Journal Nov 10, 2012 (online.wsj.com) and Quartz Nov 12, 2012 (qz.com)

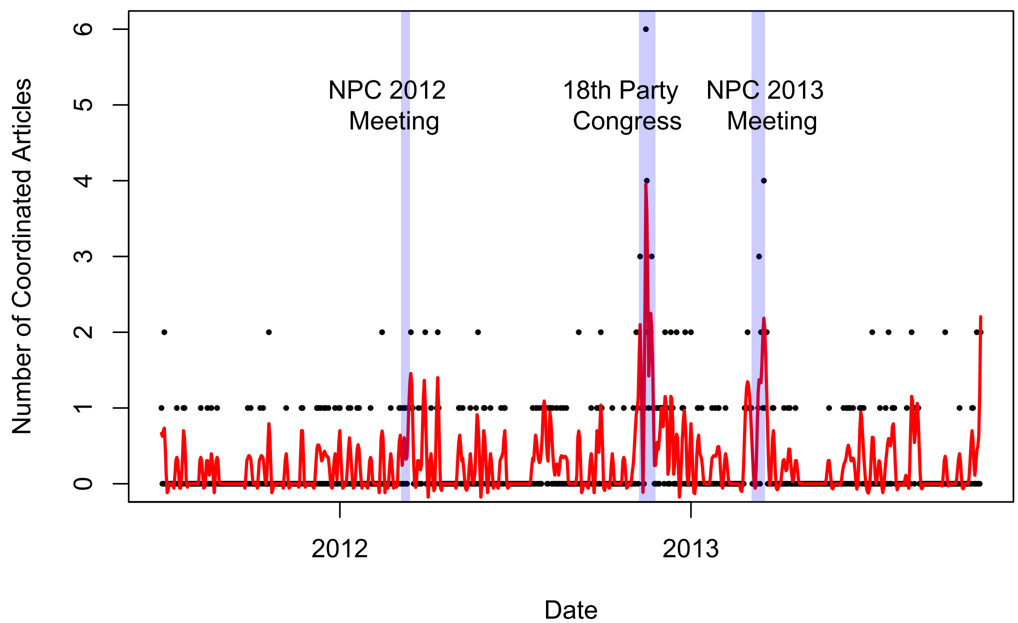

Figure 6: Timeline of Coordination

5.3 Discussion: The Logic of Localization and Coordination

Why would the Chinese media use localization on some sensitive events and coordination on others? The logic behind this lies in the conflict between bottom-up and top-down incentives. Newspapers are working to tradeoff between credibility to audiences that they receive revenue from and the necessity of following government directives. While negative news and large-scale events are attractive to newspaper audiences, they cast a negative light on government performance.

For newspapers, the credibility problem is aggravated when the event is more observable to the public. Pollution, large-scale protest events, or corruption scandals that become well-known to the public might seem blatantly omitted when newspapers fail to include them. While newspapers may not go searching for negative news to report based on government directives, missing sensitive events that their audience finds well-known could be potentially damaging.[7] For the government, the image problem is aggravated

[7] Stockmann (2012) finds that consumers of media already select sources based on their perceived cred-

the larger the paper's audience. The higher-level or more followed the paper is, the more impact the negative news will have on government image.

Based on these incentives, the logic of localization is quite clear. For local papers, the credibility problem for locally sensitive events is aggravated, since their local audience already has knowledge of the event. For the government, the image problem is lessened for local papers, since the audience is less widespread. Overall, it makes sense that there would be more negative reporting at the local level than at higher levels since consumers of news are more aware of their immediate surroundings. Unlike the Western press, where protests and negative events at local levels are quickly picked up and reported to a wider audience, the government does not want information about these events spreading to other areas.

In the cases where sensitive events happen at the national level, though, the credibility problem shifts to the national government. At this point the national government has two options: to ignore the issue, omitting it from the press, or to allow papers to address it. Interestingly, as we discuss below, we find that they do something in between: address it in all papers, with a unified and coordinated message. This both solves the credibility problem and lessens the space for alternative perspectives. The government's usual form of propaganda, synchronization, is in this case used as a form of censorship.

High levels of coordination on sensitive news articles and during sensitive time periods also suggests a new logic for propaganda. Perhaps the purpose of propaganda lies not in what the articles themselves contain, but what articles they prevent from appearing in the news media; what articles they crowd out. In that case propaganda is not ideological promotion, as many have suggested in the past, but rather a form of censorship of alternative perspectives.

---

ibility.

6 Conclusion

In this article we have uncovered and described two main strategies adopted by the Chinese government to exert control of the Chinese media environment: *localization*, where negative events are either omitted or only reported in local papers, and *coordination*, where sensitive events that are nationally recognized are reported in order to crowd out other perspectives. This suggests a new theoretical perspective on the use of propaganda that extends beyond simple ideological promotion, to controlling and shaping the media environment by crowding out opposing views.

To identify these strategies we conducted the first unsupervised large-scale study of propaganda and censorship content of Chinese newspapers. This enabled us to uncover the strategies of localization and coordination which are only observable by making systematic comparisons across a large number of texts. This represents to us the promise of "big data" to provide insights into the hidden strategies of otherwise opaque actors such as the Chinese government.

In order to facilitate this large scale analysis we joined with our collaborators Kyle Jaros and Jennifer Pan to collect a novel corpus of four million newspaper articles in 31 provincial and city-level papers over the last seven years. We also developed a new estimation strategy for the Structural Topic Model suitable to corpora of this scale. Many of the components of this estimation strategy have been incorporated into our open-source software `stm`.

There are numerous exciting avenues for future research. We have outlined a theoretical conception of propaganda that extends beyond the traditional idea of ideological promotion. In future work we hope to explore variations on these strategies further. For example, not all reprinted articles are the same. Are there systematic difference in choices of article title, deleted or added paragraphs, or placement within the newspaper? Further, are there clusters of newspapers that consistently speak together or in a more unified way?

There are also several interesting methodological directions for future research. Several of the model parameters are set by approximate manual search, notably the learning rate and the number of topics. We could instead consider automated data-driven approaches to estimating these parameters (Snoek, Larochelle and Adams, 2012), or in the case of the learning rate setting it adaptively (Ranganath et al., 2013). These types of automated procedures will help us to further extend our open-source software package `stm` to include even more of the methodological tools we describe here.

Beyond our theoretical interests in propaganda the newspaper data provide a remarkable opportunity to explore new questions in Chinese politics. We have demonstrated that when coupled with the right methodological tools, this data can yield new insights into the Chinese propaganda apparatus. We believe that this is just the beginning of the types of questions we can answer using this type of data.

References

Arora, Sanjeev, Rong Ge, Yoni Halpern, David Mimno, Ankur Moitra, David Sontag, Yichen Wu and Michael Zhu. 2012. "A practical algorithm for topic modeling with provable guarantees." *arXiv preprint arXiv:1212.4777* .

Besley, Timothy and Andrea Prat. 2006. "Handcuffs for the grabbing hand? The role of the media in political accountability." *American Economic Review* 96(3):720–736.

Blei, D. M. 2012. "Probabilistic topic models." *Communications of the ACM* 55(4):77–84.

Blei, David M. and John D. Lafferty. 2007. "A Correlated Topic Model of Science." *The Annals of Applied Statistics* 1(1):17–35.

Brady, Anne-Marie. 2009. *Marketing dictatorship: Propaganda and thought work in contemporary China*. Rowman & Littlefield.

Chang, Pi-Chuan, Michel Galley and Christopher D Manning. 2008. Optimizing Chinese word segmentation for machine translation performance. In *Proceedings of the Third Workshop on Statistical Machine Translation*. Association for Computational Linguistics pp. 224–232.

Chen, Xueyi and Tianjian Shi. 2001. "Media effects on political confidence and trust in the People's Republic of China in the post-Tiananmen period." *East Asia* 19(3):84–118.

Djankov, Simeon, Caralee McLiesh, Tatiana Nenova and Andrei Shleifer. 2001. Who owns the media? Technical report National Bureau of Economic Research.

Friedman, Jerome, Trevor Hastie and Rob Tibshirani. 2010. "Regularization paths for generalized linear models via coordinate descent." *Journal of statistical software* 33(1):1.

Grimmer, J. 2010. "A Bayesian hierarchical topic model for political texts: Measuring expressed agendas in Senate press releases." *Political Analysis* 18(1):1.

Grimmer, J. and B. M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21(3):267–297.

Grimmer, Justin. 2011. "An introduction to Bayesian inference via variational approximations." *Political Analysis* 19(1):32–47.

Hoffman, Matthew D, David M Blei, Chong Wang and John Paisley. 2013. "Stochastic variational inference." *The Journal of Machine Learning Research* 14(1):1303–1347.

Huang, Haifeng and Zhi Li. 2013. "Propaganda and Signaling." *Available at SSRN 2325101* .

Hughes, Michael C and Erik Sudderth. 2013. Memoized Online Variational Inference for Dirichlet Process Mixture Models. In *Advances in Neural Information Processing Systems*. pp. 1133–1141.

King, G., J. Pan and M. E. Roberts. 2013. "How Censorship in China Allows Government Criticism but Silences Collective Expression." *American Political Science Review* 107:1–18.

Lucas, Christopher, Richard Nielsen, Margaret E. Roberts, Brandon M Stewart, Alex Storer and Dustin Tingley. 2013. "Computer assisted text analysis for comparative politics.".

Qin, Bei, Yanhui Wu and David Strömberg. 2012. The Determinants of Media Bias in China. Technical report working paper.

Quinn, K.M., B.L. Monroe, M. Colaresi, M.H. Crespin and D.R. Radev. 2010. "How to analyze political attention with minimal assumptions and costs." *American Journal of Political Science* 54(1):209–228.

R Core Team. 2014. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
URL: *http://www.R-project.org/*

Rabinovich, Maxim and David Blei. 2014. The Inverse Regression Topic Model. In *Proceedings of The 31st International Conference on Machine Learning*. pp. 199–207.

Ranganath, Rajesh, Chong Wang, Blei David and Eric Xing. 2013. An adaptive learning rate for stochastic variational inference. In *Proceedings of The 30th International Conference on Machine Learning*. pp. 298–306.

Robbins, Herbert and Sutton Monro. 1951. "A stochastic approximation method." *The annals of mathematical statistics* pp. 400–407.

Roberts, M. E., B. M. Stewart, D. Tingley, C. Lucas, J. Leder-Luis, S. Gadarian, B. Albertson and D. Rand. Forthcoming. "Structural topic models for open-ended survey responses." *Am. Journal of Political Science* .

Roberts, Margaret, Brandon Stewart, Dustin Tingley and Edoardo Airoldi. 2013. "The structural topic model and applied social science." *Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation* .

Roberts, Margaret E, Brandon M Stewart and Dustin Tingley. N.d. Navigating the Local Modes of Big Data: The Case of Topic Models. Technical report Working Paper. Export BibTex Tagged XML.

Roberts, Margaret E., Brandon M. Stewart and Edoardo M. Airoldi. 2014. "Structural Topic Models.".

Rue, Håvard, Sara Martino and Nicolas Chopin. 2009. "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations." *Journal of the royal statistical society: Series b (statistical methodology)* 71(2):319–392.

Snoek, Jasper, Hugo Larochelle and Ryan P Adams. 2012. Practical Bayesian Optimization of Machine Learning Algorithms. In *NIPS*. pp. 2960–2968.

Stockmann, Daniela. 2012. *Media commercialization and authoritarian rule in China*. Cambridge University Press.

Taddy, Matt. 2013. "Distributed multinomial regression." *arXiv preprint arXiv:1311.6139* .

Tong, Jingrong. 2011. *Investigative journalism in China: journalism, power, and society*. Continuum.

Tong, Jingrong and Colin Sparks. 2009. "Investigative journalism in China today." *Journalism Studies* 10(3):337–352.

Trussler, Marc and Stuart Soroka. 2014. "Consumer Demand for Cynical and Negative News Frames." *The International Journal of Press/Politics* p. 1940161214524832.

Wainwright, Martin J and Michael I Jordan. 2008. "Graphical models, exponential families, and variational inference." *Foundations and Trends® in Machine Learning* 1(1-2):1–305.

Wang, C. and D. M. Blei. 2013. "Variational Inference in Nonconjugate Models." *JMLR* 14:1005–1031.