

Bayesian Dynamic Network Modeling for Streaming Social Media Political Talk

Haohan Chen^{*}

December 15, 2020

Abstract

Streaming social media network data has been used in recent studies on political behavior and institutions. Modeling their time dynamics helps political scientists produce robust results and efficiently manage their data collection process. However, existing political science methods are yet to provide researchers with the tools to analyze and monitor streaming social media network data. In this paper, I introduce Bayesian dynamic network modeling for political science research. An extension of the recent development of dynamic modeling techniques, the method enables political scientists to track trends and detect anomalies in streaming social media network data. I illustrate the method with an application to an original dataset of political discourse from a Chinese social networking site. The model detects citizens' behavioral responses to political and non-political events. It also suggests the Chinese government censors and fabricates online discourse during politically sensitive periods.

Data from political discourse on social media streams contain patterns that await discovery and explanation. However, political scientists lack the tools to track and analyze the data's time dynamics.

In recent years, social media data have been used to answer questions about political behavior and institutions that conventional forms of data have limited capacity to address. Social media data generate high-quality measures and serve as a platform for large-scale behavioral experiments. They also reveal strategies to manipulate the flow of information.

^{*}Postdoctoral Fellow, Center for Social Media and Politics, New York University, 60 5th Avenue, New York, NY 10011 (Email: haohan.chen@nyu.edu)

Most observational studies and online field experiments with social media data are built on data collection efforts that last for an extended period. As a result, observations are timestamped, and they arrive at researchers' databases as streaming data. This property has implications for both the data analysis and management of research projects. For robust data analysis, the timestamps of data generation should be considered for valid measures and unconfounded estimates. For efficient project management, researchers benefit from tracking patterns in real-time as data streams instead of waiting until the completion of data collection to perform analysis.

Methods for modeling the time dynamics of streaming social media data have not received enough attention in the political science community. For data analysis, social media data have often been aggregated into cross-sectional data or repeatedly measured data of only a few time points. For project management, researchers mostly rely on ad hoc visualization to examine their intermediate results during the data collection process.

In this chapter, I introduce Bayesian dynamic network modeling (BDNM), a scalable and interpretable method that helps political scientists track and analyze the time dynamics of social media network data. BDNM has the capacity to perform real-time inference, prediction, and anomaly detection as data streams in. The model applies the latest development of Bayesian forecasting modeling methods with an original extension to account for the eruptive and sporadic feature of social media data. I illustrate the method with its application to an original dataset of social media discourse, studying how political attention shifts in authoritarian China.

The rest of this paper is organized as follows. Section 1 reviews political science literature on social media and discusses the motivation for the new method. Section 2 introduces the Bayesian dynamic network modeling method. Section 3 presents an application of the method. Section 4 concludes.

1 Modeling Social Media Network Data

In recent years, empirical studies based on social media data have garnered significant interest in the political science community. This increasingly accessible type of data has been mined to answer a broad range of questions about political behavior and institutions. Social media data

have shed light on political behavior by providing new measures of public opinion and new platforms for field studies examining how individuals respond to political information. The data can also be used to study political institutions, revealing the motives and consequences of elites' and regimes' systematic intervention in mass political communication. This rich data source challenges political scientists to develop new methods to analyze it. State-of-the-art methods for political talk do not pay enough attention to an important dimension of the data: the time dynamics. The missing dimension limits political scientists' capacity to explore big social media data and to make inferences and predictions using these data. The rest of this section provides an overview of a selection of the important political science literature on social media political talk and describes the gap I attempt to fill with this paper.

1.1 Social Media Data for Political Science Research

Political scientists use social media data to study political behavior and political institutions. Studies of political behavior use social media data to measure mass opinion: Bond and Messing (2015) uses friendship, following, and demographic data of 6 million Facebook users to estimate their ideological position and validate their results by their self-reported political views in surveys; Barberá et al. (2015) creates measures of the ideological positions of 3.8 million Twitter users with their following network.

Political behavior research also uses data of social media activities to study how individuals process and communicate political information. Studies recently conducted in a democratic context focus on polarization and the electoral consequences of social media political talk. These studies show that the flow of information on the social media can be polarizing. Individuals have a low chance of being exposed to news representing views different from their own because their online friends tend to be in same ideological camp as them (Bakshy et al., 2015). Even when individuals are exposed to differing views, they tend to consolidate their original views (Bail et al., 2018). A substantial amount of incivility and racism is observed on social media (Munger, 2017a,b). Beyond its polarizing effect on political opinion, social media has electoral consequences. A large-scale experimental study shows that virtual recognition of voting can boost turnout (Bond et al., 2012).

Outside of democratic countries, studies of political behavior using social media data focus

on how citizens respond to government control of information. Censorship by authoritarian governments evidently adds friction to spread of information among citizens, which deters citizens from sharing anti-regime information (Roberts, 2014). However, when techniques for accessing blocked websites are available, sudden censorship can increase citizens' access to blocked websites (Hobbs and Roberts, 2018).

Finally, social media data are used to understand political institutions, especially in authoritarian regimes. For example, observational and experimental studies based on data from a large number of social media posts find that the Chinese government strategically censors social media content: posts with collective action potential are more likely to be removed while those simply criticizing the government are more likely to be tolerated (King et al., 2013, 2014). Leaked backend data from one of China's largest social networking sites help identify how companies censor social media content at the government's request (Miller, 2018). Leaked communication data from a propaganda department in China show that the government hires large numbers of Internet commentators to fabricate social media posts, in order to distract citizens from criticism of the regime (King et al., 2017). And Twitter data from Venezuela show elites in an authoritarian regime strategically use social media posts to discredit their political opponents and opposition protests (Munger et al., 2018).

This selection of literature demonstrates the importance of social media data in recent empirical studies in political science. Social media data can help political scientists understand politics at both the micro (behavioral) and the macro (institutional) level across contexts. To use this new type of data to answer substantive questions, political scientists apply and develop new methods for observational and experimental studies online.

1.2 Methods for Social Media Network Data

Political scientists use social media data for both observational and experimental research. Observational studies focus on exploring the data "as-is" and create measures to test theories of interest with statistical inference. For example, among studies reviewed in Section 1.1, observational social media data are employed to measure ideological position (Bond and Messing, 2015; Barberá et al., 2015), reveal citizens' responses to censorship (Roberts, 2014; Hobbs and Roberts, 2018), and regimes' censorship and propaganda strategies (King et al., 2013; Miller,

2018; King et al., 2017). In these studies, researchers do not intervene in the data generating process. In comparison, recent scholarship has seen the rise of field experiments using social media data, in which researchers design interventions and examine their treatment effect. For example, political scientists alter participants’ political news feeds (Bakshy et al., 2015; Bail et al., 2018) and online interactions (Munger, 2017b,a). Online field experiments have gained in popularity because they allow more control over the data than observational studies do, while providing a more realistic context for participants than lab experiments do (see Parigi et al., 2017; Baldassarri and Abascal, 2017; Muise and Pan, 2019; de Rooij et al., 2009, for reviews).

Social network modeling is an essential tool for both observational studies and field experiments with social media data. Many studies draw inferences from analysis of network data, for example, by creating ideological measurement with online friendship network, or by analyzing patterns of communication within an online messaging network. In addition, even studies that are not focused on the network commonly analyze network structure for exploratory analysis or, in field experiments, track connections among participants in the network to check for interference that may bias causal identification.

The most popular method for social network data in political science is latent space modeling. Latent space models estimate positions of observations in a low-dimensional latent space based on observed data (?). Such models can fit networks whose edges are continuous, ordinal data, or censored data (Hoff, 2009; Hoff et al., 2013; Hoff, 2015). To measure opinion with social media data, researchers extend the model to efficiently estimate latent ideological space of millions of Twitter users (Barberá, 2015; Barberá et al., 2015). In addition, variants of latent space models are used to measure politicians’ ideological positions based on congressional voting (Clinton et al., 2004) and campaign donation (?). In international relation studies, they are used to model trade and conflicts among states (Dorff and Ward, 2013).

This paper focuses on a different dimension of social media network data: time dynamics. Most existing political science studies analyze social media network data as cross-sectional data or at most as repeated measures data with only a few time points. Observational studies scrape data generated within a window of time and analyze it without regard to the exact timestamps at which datapoints were generated. Experimental studies, similarly, take snapshots of data at different phases of the study or group data into bins based on the time it

was generated. The lack of attention to the time dynamics of social media network data causes significant loss of information for researchers, especially considering how data are collected in current online observational studies and field experiments.

1.3 Tracking the Time Dynamics of Streaming Social Media Network Data

Methods to track the time dynamics of social media network data are an important yet underexplored area for political science studies on social media network data. I argue that modeling time can serve two purposes: generating more robust and richer empirical findings; and enabling effective monitoring of research projects. To serve these purposes, a model to track the time dynamics of social media networks should be scalable and amenable to streaming data.

Incorporating time in analyses helps generate more robust and richer findings from data. First, researchers can control for time as a confounder for robust results. For example, in studies on social media political communication, the quality of communication may be confounded by the time of day and day of week the talk takes place. If time of communication is not randomly assigned, not controlling for time can bias the estimated effect of the variable of interest. Second, the variation in social media activities explained by time may be theoretically interesting to political scientists. For example, in studies on individual responses to information, the timing of actions, such as following a politician or sharing news, may indicate user attitudes: those who receive information as treatment from the researcher but are slower to take actions on it may be more skeptical and more reluctant to do so.

Incorporating time can also enable researchers to more effectively monitor research projects using social media network data. Collection of original data from the social network can take months or even years. The long time required for data collection is sometimes due to limitations in computational power. More often, it is by design: in observational studies, researchers observe online communities for an extended time period to obtain larger sample sizes; in field experiments, they may want to study the long-term effects of their treatments. Regardless of the reason for choosing a long data collection process, researchers usually benefit from real-time intermediate output. High-quality intermediate output enables researchers to access preliminary results before data collection is completed. It also helps researchers detect

anomalies early and respond with supplemental research designs.

To serve these two purposes, political scientists need a new type of model that is scalable and amenable to streaming data. The computation of the model should be scalable as data gathered from large social media networks measured over a long time period can be computationally overwhelming. In addition, to monitor streaming data in the data collection process, the model should have online learning capacity. In the following section, I introduce Bayesian dynamic network modeling, a method that has both features.

2 Bayesian Dynamic Network Modeling

I introduce Bayesian dynamic network modeling for political science studies of social media network data. Scalable and capable of online learning with streaming data, the method has three integral components: a data processing strategy, decoupling/recoupling, dividing repeatedly measured network data into univariate sequences; a Bayesian model, the dynamic generalized linear model, for decoupled binary and count network flows; and the dynamic gravity model for reconstruction of network structure with the learned parameters. The first part of this section provides an overview of the method with reference to the literature. The second part introduces the decoupling/recoupling concept. The third part introduces the dynamic generalized linear model. The fourth part introduces the dynamic gravity model.

2.1 An Overview of Bayesian Dynamic Modeling

I start the discussion with an overview of Bayesian dynamic modeling to provide background information to readers from the political science community.¹ Bayesian dynamic network modeling considers repeatedly measured network flow data as a multivariate time series. The method is a recent extension of traditional (univariate) dynamic models, a family of statistical model for time series data. Since its introduction decades ago, dynamic models have had wide application in medical research and finance. However, they have been underutilized in political science.

The basic setup of a univariate dynamic models can be demonstrated by an example of its simplest form, a normal dynamic linear model. Importantly, it models time series data with a

¹The overview refers to two textbooks of Bayesian dynamic modeling: ?, Section 1, 2, ?, Chapter 4.

system of two equations. Consider a time series $Y = [y_1, y_2, \dots, y_T]^T$. A normal dynamic linear model is formalized as follows:

$$\text{(observation model)} \quad y_t = \mathbf{F}'_t \boldsymbol{\theta}_t + \nu_t \quad \nu_t \sim N(0, V_t)$$

$$\text{(evolution model)} \quad \boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t \quad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{W}_t)$$

$$\text{(initial information)} \quad \boldsymbol{\theta}_0 \sim N(\mathbf{0}, \mathbf{W}_0)$$

To elaborate: at time t , the distribution of observed data y_t is dependent on an unobserved state vector $\boldsymbol{\theta}_t$ of the current period; the distribution of unobserved state vector $\boldsymbol{\theta}_t$ evolves over time depending on previous state vectors. The model has four hyperparameters, predetermined by the researcher: \mathbf{F}_t determines the relation between observed data and the current state vector, containing known constants and regressors at time t . \mathbf{G}_t , known as the state evolution matrix, determines in what way a current state vector is dependent on previous state vectors. The specification of \mathbf{G}_t is flexible. Traditional time domain models AR, MA, and ARMA can all be modeled with special cases of \mathbf{G}_t (? , Chapter 2). It can also model seasonality (? , Chapter 3). ν_t is the observation noise, controlling how much the model attributes variance of the data at certain time point to measurement error. \mathbf{w}_t is the state evolution noise, controlling how much fluctuation of data over time is considered signals of time-dependent evolution as oppose to noise.

Fitting a Bayesian dynamic model, two processes are performed to learn the posterior state vector: sequential updating and retrospective updating. Sequential updating learns the posterior distribution of a state vector $\boldsymbol{\theta}_t$ based on data points observed at all previous and current periods. Note that at the first time period $t = 1$, prior information $\boldsymbol{\theta}_0$ is provided. After sequential updating, researchers can perform retrospective analysis to smooth the learned state vectors. Retrospective updating learns posterior distribution of a state vector $\boldsymbol{\theta}_{t|T}$ based on all data points of past, current, and future time periods. Researchers can make inferences based on the two sets of learned posterior state vectors. For use of the model to forecast, interests lie at using sequentially updated state vectors $\boldsymbol{\theta}_t$ to forecast data of future k time periods y_{t+k} .

Bayesian dynamic modeling has been widely applied to analysis of medical and financial data. However, it is underutilized in political science, especially research on big social media

network data. The lack of attention is evident in its absence from recent major methodological contributions to time series data analysis in political science (Keele and Kelly, 2006; Keele and DeBoef, 2008; Beck and Katz, 2011; Box-Steffensmeier et al., 2014). To my knowledge, its only applications are in models of the ideological positions of the justices of the U.S. Supreme Court (?) and in text-as-data models for legislative agendas (Quinn et al., 2010). In both applications, dynamic linear models are applied to capture the time dynamics.

The dynamic linear model cannot be directly applied to social media network data because the data are multivariate and usually not continuous or normally distributed. I build on recent developments of the decoupling/recoupling data processing strategy and techniques handling over-dispersion in the dynamic modeling literature to develop new dynamic network models for political science research using social media data.

2.2 Decoupling/Recoupling

Using a decoupling/recoupling strategy, I operationalize social media networks into individual univariate time series for scalable computation. The strategy is straightforward: the decoupling step divides network data into individual time series, each fits into an individual dynamic model; the recoupling step models the fitted parameters of dynamic models of the decoupling stage.

The straightforward strategy can be illustrated with a hypothetical example. Assume a researcher collects data on uncivil political exchanges among a group of users on a social media platform. Let y_{ijt} be the number of uncivil exchanges between users i and j observed at time t . Assume there are I users observed in T time periods. In the decoupling step, an individual dynamic model is fitted to a sequence of data $Y_{ij} = [y_{ij,1}, y_{ij,2}, \dots, y_{ij,T}]$ for each pair of users $i, j \in \{1, 2, \dots, I\}$. The maximum total number of individual models fit is the total number of possible edges $I(I - 1)$. Results of the decoupling step show how the level of uncivil communication between each pair of users changes over time. In the recoupling step, parameters learned in all these models are used to attribute the sources of edge formation (i.e., engagement in uncivil online talk in this example). This is a stepwise process in which the decoupling step models the time dependency, while the recoupling step models the network dependency.

My application of the decoupling/recoupling strategy is built on recent developments in dynamic modeling methods. For example, dynamic dependence network models by ? use the strategy for modeling and forecasting high-dimensional financial time series data. ? use the strategy to model multi-scale data of retail inventory, making use of information about categories of products. Closer to my effort, ? and ? model click-through data on news websites to analyze online traffic over time.²

With network data decoupled into individual time series, the next steps constitute the development of appropriate dynamic models for decoupled data and an appropriate recoupling method.

2.3 Decoupling: Dynamic Generalized Linear Model

The edges of social media networks are often binary or weighted by some count data. Examples of binary (unweighted) edges include whether a person follows another and whether a person sends a message to one another during a certain time period. Examples of edges weighted by count data include the number of common friends or common page visits between two people and the number of conversation between them. To model these types of data, dynamic models for binary and count outcome are required. An additional challenge, these data can be volatile and sporadic: social media activities can be eruptive (e.g., discussion around a piece of breaking news); the number of non-zero data points between a pair of nodes throughout time can be rare (e.g., many people remain silent on the social media most of the time). Therefore, these feature require additional modeling design. I introduce dynamic generalized linear models for binary and count data and extend them to better fit volatile and sporadic social media data.

2.3.1 DGLM for Binary and Count Data in a Network

I model the time series of binary and count data with dynamic generalized linear models (?, Section 14). With DGLM, observations are assumed to be drawn from a distribution in the

²For more examples, see West (2020).

exponential family:

$$p(Y_t|\eta_t, V_t) = \exp\{V_t^{-1}[y_t(Y_t)\eta_t - a(\eta_t)]\}b(Y_t, V_t)$$

where η_t is the natural parameter of the distribution and V_t is the scale parameter.

A dynamic generalized linear model is specified as below:

$$\begin{aligned} \text{(observation)} \quad & p(Y_t|\eta_t) \\ \text{(link)} \quad & g(\eta_t) = \mathbf{F}'_t \boldsymbol{\theta}_t \\ \text{(evolution)} \quad & \boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t \quad \mathbf{w}_t \sim [\mathbf{0}, \mathbf{W}_t] \end{aligned}$$

Compared to the dynamic linear models introduced in Section 2.1, the generalized linear model adds a link function between the observation and evolution models. In the link function, the natural parameter of the observation model η_t links the observation and evolution models. In the evolution model, the normality assumption of the evolution noise is removed. It is instead defined in terms of its first and second moments.

In my application, I focus on two cases of dynamic generalized linear models: the Bernoulli logistic model and the Poisson loglinear model. The former models binary data, while the latter models count data. A Bernoulli logistic model is specified as below:

$$\begin{aligned} \text{(observation)} \quad & y_t \sim \text{Bernoulli}(p_t) \quad p_t \sim \text{Beta}(\alpha_t^b, \beta_t^b) \\ \text{(link)} \quad & \log \frac{p_t}{1-p_t} = \mathbf{F}'_t \boldsymbol{\theta}_t \\ \text{(evolution)} \quad & \boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t \quad \mathbf{w}_t \sim [\mathbf{0}, \mathbf{W}_t] \end{aligned}$$

A Poisson loglinear model is specified as below:

$$\begin{aligned} \text{(observation)} \quad & y_t \sim \text{Poisson}(\lambda_t) \quad \lambda_t \sim \text{Gamma}(\alpha_t^g, \beta_t^g) \\ \text{(link)} \quad & \log \lambda_t = \mathbf{F}'_t \boldsymbol{\theta}_t \\ \text{(evolution)} \quad & \boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t \quad \mathbf{w}_t \sim [\mathbf{0}, \mathbf{W}_t] \end{aligned}$$

2.3.2 Handling Over-dispersion: Dynamic Count Mixture Model

Social media network count data over time can be eruptive and sporadic. Modeling such data with the Poisson loglinear model can cause over-dispersion, underestimate variance, and fail to predict unusually high or low data points. I tackle this challenge by using a dynamic count mixture model with random effects. The dynamic count mixture model was originally developed by ? to model sales data where the sales volume of different products was volatile and of different scales. I apply it to social media network data, which arguably has similar properties.

The dynamic count mixture model considers generation of the observed data as a two-step process: First, a binary series indicating whether observation y_t is non-negative $z_t = \mathbb{1}(y_t > 0)$ is drawn from a Bernoulli distribution. Then, if $z_t = 0$, $y_t = 0$. If $z_t = 1$, y_t is drawn from a Poisson distribution shifted by 1. The observation function is specified below:

$$\begin{aligned} \text{(observation)} \quad z_t \sim \text{Bern}(p_t) \quad y_t | z_t &= \begin{cases} 0 & \text{if } z_t = 0 \\ 1 + x_t \quad x_t \sim \text{Pois}(\lambda_t) & \text{if } z_t = 1 \end{cases} \\ p_t \sim \text{Beta}(\alpha_t^b, \beta_t^b) \quad \lambda_t \sim \text{Gamma}(\alpha_t^g, \beta_t^g) \end{aligned}$$

The two stages of the observational model have separate dynamic structures. Their link functions and evolution equations follow the model structures of the Bernoulli logistic and Poisson loglinear models respectively, as specified below:

$$\begin{aligned} \text{(link-bern)} \quad \log \frac{p_t}{1 - p_t} &= \mathbf{F}_t^{0'} \boldsymbol{\xi}_t \\ \text{(link-pois)} \quad \log \lambda_t &= \mathbf{F}_t^{+'} \boldsymbol{\theta}_t \\ \text{(evolution-bern)} \quad \boldsymbol{\xi}_t &= \mathbf{G}_t^{0'} \boldsymbol{\xi}_{t-1} + \mathbf{w}_t^0 \quad \mathbf{w}_t^0 \sim [\mathbf{0}, \mathbf{W}_t^0] \\ \text{(evolution-pois)} \quad \boldsymbol{\theta}_t &= \mathbf{G}_t^{+'} \boldsymbol{\theta}_{t-1} + \mathbf{w}_t^+ \quad \mathbf{w}_t^+ \sim [\mathbf{0}, \mathbf{W}_t^+] \end{aligned}$$

The above specification may still be insufficient for volatile and sporadic social media network data. I further add a time-specific random effect to the conditional Poisson model. Specifically, a time-specific element that can capture a part of the variation at certain time point that cannot be captured by the original state vector. The link function and evolution model for

the conditional Poisson loglinear model with random effect is specified below (the part of the Bernoulli logistic model remains unchanged):

$$\begin{aligned}
& \text{(link-pois-re)} \quad \log(\lambda_t) = \tilde{\mathbf{F}}_t^{+'} \tilde{\boldsymbol{\theta}}_t \quad \text{where } \tilde{\mathbf{F}}_t^+ = [1, \mathbf{F}_t^{+'}]' \text{ and } \tilde{\boldsymbol{\theta}}_t = [\zeta_t, \boldsymbol{\theta}'_t]' \\
& \text{(evolution-pois-re)} \quad \boldsymbol{\theta}_t = \tilde{\mathbf{G}}_t^{+'} \boldsymbol{\theta}_{t-1} + \tilde{\mathbf{w}}_t^+ \quad \tilde{\mathbf{w}}_t^+ \sim [0, \tilde{\mathbf{W}}_t^+] \\
& \quad \text{where } \tilde{\mathbf{G}}_t^+ = \text{blockdiag}(1, \mathbf{G}_t^+) \quad \tilde{\mathbf{W}}_t^+ = \text{blockdiag}(w_t^{RE}, \mathbf{W}_t^+)
\end{aligned}$$

2.3.3 Fitting Dynamic Generalized Linear Models

Dynamic generalized linear models can learn the posterior parameters for inferences and predictions as data streams in, a key feature making them suitable for observational studies and online field experiments with social media. Fitting a DGLM, there are two major processes: sequential updating and retrospective updating. The former learns parameters based on all previous and current data point, while the latter update parameters using all data. This section outlines the updating processes of the dynamic Bernoulli logistic and Poisson loglinear models, ending with a brief discussion of how they extend to mixture models.

Sequential updating fits models as data streams in. The process loops through data points from the beginning to the end to learn parameters. To learn posterior parameters at each time point, it takes six steps, as is detailed in the following block of equations :³

$$\begin{aligned}
& \text{(initialize)} \quad \text{Parameters of the initial stage } \mathbf{m}_0 \text{ and } \mathbf{C}_0 \text{ are pre-determined.} \\
& \text{(prior)} \quad \boldsymbol{\theta}_t | \mathcal{D}_{t-1} \sim [\mathbf{a}_t, \mathbf{R}_t] \quad \text{where } \mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1}, \mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t \\
& \text{(predict)} \quad \eta_t | \mathcal{D}_{t-1} \sim [f_t, q_t] \quad \text{where } f_t = \mathbf{F}_t' \mathbf{a}_t, q_t = \mathbf{F}_t' \mathbf{R}_t \mathbf{F}_t \\
& \text{(link)} \quad \text{Get prior observation model param. } \psi_t | \mathcal{D}_{t-1} \text{ by moment approx.} \\
& \quad \text{Bern: solve } \begin{cases} f_t = \gamma(\alpha_t^b) - \gamma(\beta_t^b) \\ q_t = \dot{\gamma}(\alpha_t^b) + \dot{\gamma}(\beta_t^b) \end{cases} \quad \text{for } \alpha_t^b, \beta_t^b \Rightarrow p_t | \mathcal{D}_{t-1} \sim \text{Bern}(\alpha_t^b, \beta_t^b) \\
& \quad \text{Pois: solve } \begin{cases} f_t = \gamma(\alpha_t^g) - \log \beta_t^g \\ q_t = \dot{\gamma}(\alpha_t^g) \end{cases} \quad \text{for } \alpha_t^g, \beta_t^g \Rightarrow \lambda_t | \mathcal{D}_{t-1} \sim \text{Gamma}(\alpha_t^g, \beta_t^g)
\end{aligned}$$

³The Bernoulli logistic model and the Poisson loglinear model have the same operation in all steps but the “link” step. The equation starting with “Bern” is for the former, while the one starting with “Pois” is for the latter.

where $\gamma(\cdot)$ and $\dot{\gamma}(\cdot)$ are digamma and trigamma functions.

- (update) Get posterior observation model param. $\psi_t|\mathcal{D}_t$ with data
 Bern: $p_t|\mathcal{D}_t \sim \text{Bern}(\alpha_t^{\text{b}*}, \beta_t^{\text{b}*})$ where $\alpha_t^{\text{b}*} = \alpha_t^{\text{b}} + z_t, \beta_t^{\text{b}*} = \beta_t^{\text{b}} + 1 - z_t$
 Pois: $\lambda_t|\mathcal{D}_t \sim \text{Gamma}(\alpha_t^{\text{g}*}, \beta_t^{\text{g}*})$ where $\alpha_t^{\text{g}*} = \alpha_t^{\text{g}} + x_t, \beta_t^{\text{g}*} = \beta_t^{\text{g}} + 1$
- (link-back) Get posterior link $\eta_t|\mathcal{D}_t \sim [f_t^*, q_t^*]$
 Bern: $f_t^* = \gamma(\alpha_t^{\text{b}*}) - \gamma(\beta_t^{\text{b}*}) \quad q_t^* = \dot{\gamma}(\alpha_t^{\text{b}*}) + \dot{\gamma}(\beta_t^{\text{b}*})$
 Pois: $f_t^* = \gamma(\alpha_t^{\text{g}*}) - \log(\beta_t^{\text{g}*}) \quad q_t^* = \dot{\gamma}(\alpha_t^{\text{g}*})$
- (posterior) $\theta_t|\mathcal{D}_t \sim [\mathbf{m}_t, \mathbf{C}_t]$ where
 $\mathbf{m}_t = \mathbf{a}_t + \mathbf{R}_t \mathbf{F}_t (f_t^* - f_t) / q_t \quad \mathbf{C}_t = \mathbf{R}_t - \mathbf{R}_t \mathbf{F}_t \mathbf{F}_t' \mathbf{R}_t' (1 - q_t^* / q_t) / q_t$

Retrospective updating, the second block of processes in fitting dynamic models, can be performed at the end of a phase of data collection. The learning process updates parameters learned in sequential updating by considering all data points collected. For sequences with considerable volatility, it can significantly change the estimation of some parameters. Researchers can use parameters learned with retrospective updating as the “finalized” results for interpretation.

The process starts from the end of a sequence and loops backward to the start. Updating at each time point requires three steps, as detailed below:⁴.

- (retro) $\theta_t|\mathcal{D}_T \sim [\mathbf{a}_{t|T}, \mathbf{R}_{t|T}]$ where, with $\mathbf{B}_t = \mathbf{C}_t \mathbf{G}'_{t+1} \mathbf{R}_{t+1}^{-1}$
 $\mathbf{a}_{t|T} = \mathbf{m}_t - \mathbf{B}_t (\mathbf{a}_{t+1} - \mathbf{a}_{t+1|T}) \quad \mathbf{R}_{t|T} = \mathbf{C}_t - \mathbf{B}_t (\mathbf{R}_{t+1} - \mathbf{R}_{t+1|T}) \mathbf{B}_t'$
- (predict-s) $\eta_t|\mathcal{D}_T \sim [f_{t|T}, q_{t|T}]$ where $f_{t|T} = \mathbf{F}'_t \mathbf{a}_{t|T}, q_{t|T} = \mathbf{F}'_t \mathbf{R}_{t|T} \mathbf{F}_t$
- (link-s) Get smoothed observation model param. $\psi_t|\mathcal{D}_T$ by moment approx.
 Bern: solve $\begin{cases} f_{t|T} = \gamma(\alpha_{t|T}^{\text{b}}) - \gamma(\beta_{t|T}^{\text{b}}) \\ q_{t|T} = \dot{\gamma}(\alpha_{t|T}^{\text{b}}) + \dot{\gamma}(\beta_{t|T}^{\text{b}}) \end{cases}$ for $\alpha_{t|T}^{\text{b}}, \beta_{t|T}^{\text{b}}$
 $\Rightarrow p_t|\mathcal{D}_T \sim \text{Bern}(\alpha_{t|T}^{\text{b}}, \beta_{t|T}^{\text{b}})$

⁴The same as Footnote 3.

$$\text{Pois: solve } \begin{cases} f_{t|T} = \gamma(\alpha_{t|T}^g) - \log \beta_{t|T}^g \\ q_{t|T} = \dot{\gamma}(\alpha_{t|T}^g) \end{cases} \quad \text{for } \alpha_{t|T}^g, \beta_{t|T}^g$$

$$\Rightarrow \lambda_t | \mathcal{D}_T \sim \text{Gamma}(\alpha_{t|T}^g, \beta_{t|T}^g)$$

where $\gamma(\cdot)$ and $\dot{\gamma}(\cdot)$ are digamma and trigamma functions.

The above discussion outlines the processes of sequential updating and retrospective updating for both the dynamic Bernoulli logistic model and the dynamic Poisson loglinear model.

Fitting the dynamic count mixture model only requires a simple extension from the above two models. As discussed in Section 2.3.2, the two stages of the observation model have separate evolution processes. The binary indicators for non-zero counts z_t are simply fitted to a dynamic Bernoulli logistic model. The shifted counts x_t are fitted to a separate dynamic Poisson loglinear model. At a time period where zero is observed, x_t is treated as missing, for which parameters are not updated in either sequential or retrospective updating.

2.3.4 Choosing Hyperparameters

Fitting a dynamic model, the choice of hyperparameters determines the model's assumptions (which reflects the researcher's understanding of the data) about the structure of time dependency (\mathbf{G}_t), the structure of local covariates (\mathbf{F}_t), and the level of noise as data evolve over time (\mathbf{W}_t). Dynamic models are flexible and provide researchers with a variety of options for these hyperparameters. Some specifications make a dynamic model equivalent to traditional time domain models such as ARMA (see ?, Chapter 4, 5). I focus on a simple specification that suffices for my purpose.

Working with streaming social media data, researchers are interested in finding meaningful trends of edge formation in the network. However, with the messiness of social media data, manually examining the data points can confuse signals with noise. For this reason, I introduce two simple specifications of hyperparameters \mathbf{F}_t and \mathbf{G}_t that provide interpretable results on how fast the data of interest evolve over time.

The two specifications, Local Linear Growth Model (LLGM) and Local Quadratic Growth Model (LQGM), are among the most popular dynamic models (see ?, Section 7; see also ? for a recent application). The two models both take constant $\mathbf{F}_t \equiv \mathbf{F}, \mathbf{G}_t \equiv \mathbf{G}$. They are

specified as follows:

$$\begin{aligned}
 (\text{LLGM}) \quad \mathbf{F} &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} & \mathbf{G} &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} & \boldsymbol{\theta}_t &= \begin{bmatrix} \theta_{0,t} \\ \theta_{\Delta,t} \end{bmatrix} \\
 (\text{LQGM}) \quad \mathbf{F} &= \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} & \mathbf{G} &= \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} & \boldsymbol{\theta}_t &= \begin{bmatrix} \theta_{0,t} \\ \theta_{\Delta,t} \\ \theta_{\Delta^2,t} \end{bmatrix}
 \end{aligned}$$

The models are of interest particularly because elements of their state vectors $\boldsymbol{\theta}_t$ are interpretable. In LLGM, the first element of state vector $\theta_{0,t}$ shows the absolute current level, while the second element $\theta_{\Delta,t}$ captures the local rate of change. In LQGM, besides the first two elements that have the same interpretation with LLGM, the third element $\theta_{\Delta^2,t}$ captures the change in rate of change.

The definition of the parameter for state evolution noise \mathbf{w}_t is more straightforward. Following the conventional practice of dynamic modeling, I model it as a proportion of the prior variance of the state, introducing a discount factor $\delta \in (0, 1]$ (Klein, Chapter 4).

$$\mathbf{W}_t = \frac{1 - \delta}{\delta} \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}'_t$$

With this specification, a larger discount factor δ is associated with smaller evolution noise, leading to a more stable fitted trend. When the discount rate is 1, it means the researcher assume no noise in evolution. In this case, all noise is attributed to the data generating process in the observation model.

For dynamic count mixture models, the discount rate for the Bernoulli logistic model and the Poisson loglinear model are modeled separately. In addition, the random effect in the Poisson model has its own discount rate, controlling the degree to which the researcher can attribute variation to time-specific effect. The setup results in three discount rates as hyperparameters: δ^{bern} for the Bernoulli model, δ^{pois} for the Poisson model, and ρ^{pois} for the random effect of the Poisson model. Formally, it changes the (prior) step of the sequential

updating algorithm to the following:

$$\begin{aligned}
(\text{prior-bern}) \quad \mathbf{R}_t &= \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}'_t / \delta^{\text{bern}} \\
(\text{prior-pois}) \quad \mathbf{R}_t &= \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}'_t / \delta^{\text{pois}} \\
(\text{prior-pois-re}) \quad \mathbf{R}_t &= \tilde{\mathbf{G}}_t \mathbf{C}_{t-1} \tilde{\mathbf{G}}'_t / \text{blockdiag}([\rho^{\text{pois}}, \delta^{\text{pois}} \mathbf{1}_{(|\tilde{\mathbf{G}}_{t-1}|^2)}])
\end{aligned}$$

In more complex modeling, researchers may use different discount rates for different elements of the state vector. I do not explore this complexity in this paper.

2.3.5 One-step-ahead Prediction for Model Evaluation and Anomaly Detection

Like other Bayesian methods, predictions of dynamic generalized linear models are distributions. In existing applications of dynamic modeling, one-step-ahead prediction is often performed for forecasting, model evaluation and anomaly detection. One-step-ahead prediction obtains the distribution of the outcome at a current time point based on data in all previous time points. The distributions of one-step-ahead predictions for Bernoulli logistic and Poisson loglinear models are specified as follows:

$$\begin{aligned}
(\text{Bern-predict}) \quad \hat{z}_t | \mathcal{D}_{t-1} &\sim \text{Beta-Bernoulli}(1, \alpha_t^b, \beta_t^b) \\
(\text{Pois-predict}) \quad \hat{x}_t | \mathcal{D}_{t-1} &\sim \text{Negative-Binomial}(\alpha^{\text{gt}}, \frac{\beta_t^g}{1 + \beta_t^g})
\end{aligned}$$

Applications of dynamic models to financial data use one-step or k -step ahead prediction for forecasting. However, since forecasting is not the objective of my application for political science research, I do not discuss it in detail.

My application uses one-step prediction to evaluate models and detect anomalies. With the one-step-ahead predictive distribution, the marginal predictive likelihood of each data point can be obtained. The cumulative marginal predictive likelihood is used to evaluate a model up to a certain time point. The marginal predictive likelihood at a time point is used to detect anomalies: when there is a low marginal predictive likelihood, the flow is likely abruptly changing to an abnormally low or high value.⁵

Aside from this, the distribution of fitted data at each time point after retrospective

⁵Other methods for model monitoring are also available. See ?, Section 11.

updating can be obtained with the same operationalization of one-step-ahead prediction.

$$\begin{aligned}
 \text{(Bern-fitted-retro)} \quad \hat{z}_{t|T}|\mathcal{D}_T &\sim \text{Beta-Bernoulli}(1, \alpha_{t|T}^b, \beta_{t|T}^b) \\
 \text{(Pois-fitted-retro)} \quad \hat{x}_{t|T}|\mathcal{D}_T &\sim \text{Negative-Binomial}(\alpha_{t|T}^g, \frac{\beta_{t|T}^g}{1 + \beta_{t|T}^g})
 \end{aligned}$$

2.3.6 Summary

In this section, I introduced dynamic generalized linear modeling for decoupled social media network data. Three types of models were discussed: the Bernoulli logistic model, the Poisson loglinear model, and the count mixture model. The first model fits binary network flows while the second and third fit count network data. I highlighted the count mixture model which captures the eruptive and sporadic nature of social media data.

The decoupling step fits models describing edge formation between each pair of nodes in the network as data streams in. As discussed, it can help researchers explore trends and detect anomalies during the long process of data collection. However, the step’s result is not yet informative of the network structure. In the next section, I introduce a method for unveiling network dependency: the dynamic gravity model.

2.4 Recoupling: Dynamic Gravity Model

Researchers working with streaming social media network data are often interested in attributing edge formation between nodes. For example, during data collection for a study on the incivility of social media political communication, researchers may observe a spike of uncivil exchange between two users at a time point. A question of interest is whether the observation is attributable to a general rise of incivility in the community, a systematic behavioral change in either user, or a disruption of the two user’s relationship. Answering questions of this kind requires the second step of Bayesian dynamic network modeling: recoupling.

I use a dynamic gravity model to recouple. The model assumes that edge formation between a pair of nodes is attributable to the multiplicative effect of four factors: a time-specific effect (μ), an original effect (α), a destination effect (β), and an affinity effect (γ).

For nodes i, j at time t , the edge formation model is specified below:

$$y_{ijt} = \mu_t \alpha_{it} \beta_{jt} \gamma_{ijt}$$

The gravity model is familiar to political scientists studying international political economy. For example, studies use the gravity model to fit trade data among countries to study the effect of WTO (Rose, 2004; Tomz et al., 2007). A fully Bayesian dynamic latent space model based on the gravity model is developed to model international trade networks (Ward et al., 2013). My application of the model directly adapts the recent development of dynamic modeling for online traffic data (??). My application slightly deviates by defining the model in terms of the observed outcome instead of the parameters of the observation model, to intuitively link with the mixture model and fully capture the variance.

The dynamic gravity model approximates the distribution of the four types of effect by empirically decomposing samples of predictive distribution produced in the decoupling step. For each pair of nodes i, j at each time point t , a sample of predicted or fitted outcome \hat{y}_{ijt} is drawn from its distribution. Putting all samples of individual sequences together results in a sample of fitted networks. For a sample network s , the four types of effect are calculated below:

$$\begin{aligned} f_t^{(s)} &= \sum_{i \in I, j \in I} \log(\hat{y}_{ijt}^{(s)}) / I^2 \\ a_{it}^{(s)} &= \sum_{j=1}^I \log(\hat{y}_{ijt}^{(s)}) / I - f_t^{(s)} \\ b_{jt}^{(s)} &= \sum_{i=1}^I \log(\hat{y}_{ijt}^{(s)}) / I - f_t^{(s)} \\ g_{ijt}^{(s)} &= \log(\hat{y}_{ijt}^{(s)}) - f_t^{(s)} - a_{it}^{(s)} - b_{jt}^{(s)} \end{aligned}$$

The model outputs samples approximating the distribution of the four decomposed network effects. Researchers can use their summary statistics (e.g., mean and 95% credible interval) for inference and anomaly detection in a way that is similar to that discussed in Section 2.3.

3 Application: Attention on Chinese Social Media

In this section, I apply Bayesian dynamic network modeling to a dataset of online political discourse from one of China’s largest social networking sites. My empirical results show that individual attention shifts in response to political and non-political events over time. In addition, my results suggest the existence of strategic censorship and fabrication of online discourse by the Chinese government. This section is organized as follows: Section 3.1 introduces how the substantive question is linked to the method. Section 3.2 and 3.3 introduce the data and the models. Section 3.4 elaborates my findings. Section 3.5 discusses.

3.1 Attention

Attention is a scarce resource in the social media era. Politicians fight for it with various motives. How public attention on politics develops and is manipulated in different contexts receives close scrutiny in the political science literature.

In a democratic context, researchers are interested in how public attention and elites’ political and policy agendas shape one another over time. For example, Jennings and John (2009) analyze time-series survey data and coded the Queen’s speech in the U.K. to find that the British government has short-run responsiveness to public attention on certain issues. Other works explore the dynamics of attention shifts of politicians and political parties (Grimmer, 2010; Quinn et al., 2010; Ramirez, 2009).

I focus on political attention in an authoritarian context. Recent research shows that authoritarian governments both respond to public opinion and actively manipulate it, for a common aim: avoiding anti-regime collective action. For example, experiments show that Chinese local officials are more likely to respond to an inquiry that comes with a threat of collective action (Chen et al., 2016; Meng et al., 2017) and the Chinese government censors social media posts with collective action potential (King et al., 2013, 2014). Of particular interest to my inquiry, a recent study shows that the Chinese government has hired an “army” of commentators to distract political discussion on social media away from criticism of the government (King et al., 2017).

New social media data provide new opportunities for the study of public political attention, especially for researchers of authoritarian politics. Data collection for research projects on

attention shift often takes months if not years to complete. Efficiently processing streaming network data can help researchers understand the dynamics of attention shift during the course of data collection. It can also help researchers better monitor their data collection. For example, in the event of abnormally high attention surrounding certain issues, researchers may be alerted and implement supplemental programs (e.g., a follow-up survey or experiment).

In this case study, I demonstrate the application of Bayesian dynamic network modeling to analyze the dynamics of political attention on a Chinese social networking site.

3.2 Data

I study attention on Chinese social media using a dataset of political discussion on *Zhihu*, one of the largest social networking sites in China. *Zhihu* is a question-and-answer site. Users can post questions, answer questions, and vote or comments on answers. Figure 1 shows a screenshot of a *Zhihu* question page: a question is displayed with one or more tags in the header and answers to the question are displayed below. Key to my network operationalization, all questions on the site are associated with tags that indicate their topics of interest. Tags are assigned by users as well as the site’s administrators. Here, my interest in the data is in how users’ attention flows from one tag to another at the aggregate level. In this section, I introduce the *Zhihu* website, the data collection process, and the operationalization of network data of interest.

3.2.1 *Zhihu*

Zhihu is one of the most popular social networking sites in China. It ranks 24th among all websites in China and 105th in the world, according to *Alexa*, an online service of Amazon that tracks real-time traffic of websites globally. A closer look at the 50 top visited sites in China shows that the only two social network sites with larger traffic than *Zhihu* are *Weibo* (7th) and *Tianya* (22st). The site’s management says that it has over 100 million registered users and 26 million active users per day, who spend an average of one hour on the site per day.

Political discussion on the website has influenced several political events in China. For example, in 2016, users of the website raised public awareness of the Lei Yang Incident, in

Figure 1: A Screenshot of of *Zhihu*



which a young man died suspiciously while in custody in Beijing. Friends of Lei posted statements and questions about Lei's death on *Zhihu* and suggested that the policemen handling the case engaged in misconduct. Fellow users of the site discussed inconsistency in the police statements, which were later quoted in reports by traditional media outlets and discussions on other Chinese social media platforms such as *Weibo* and *WeChat*. The role of *Zhihu* as a platform for expressing political opinions was later reported in the *Economist*. More recently, during the March 2018 National People's Congress, users on the website subtly discussed and criticized the lifting of presidential term limits by asking and answering questions about the harm of "driver fatigue." Following this discussion, the mobile app for the site was temporarily suspended. Beyond these high-profile cases, users of the site are actively engaged in everyday discussions about governance, policies, and international relations.

3.2.2 Data Collection

I developed a web scraper to collect political discourse posted on *Zhihu* from December 1, 2011 (when the website launched), to March 30, 2016. My scraper visited 101,532 pages with questions tagged “politics,” collecting a total of 511,137 answers to these questions. In total, these answers received over 9.7 million upvotes and over 2 million comments by over 1.8 million unique users. I also scraped the profiles of users who: asked a question tagged “politics;” posted an answer to any of these questions; or upvoted or commented on any of these answers. Information was collected from user profiles, including a set of self-reported demographic information, namely, gender, location (province and city), education (name of schools attended), and occupation.

3.2.3 Operationalization of Network Data

I examine how users of *Zhihu* shift their attention across political topics by looking at traffic flows among tags. On *Zhihu*, tags on questions summarize political topics. A user posting answers under certain tags signals attention to the topic summarized by the tag. A user who previously posted an answer under one tag then posts under another tag is considered to have shifted attention from the previous tag to the current one. The count of the number of users shifting attention is operationalized as the network flow between the two tags. In addition to the tags in the political section, I create an “outside” tag, indicating traffic flow from outside the community. If a user has never posted an answer before, his contribution to any tag is considered a flow from “outside” to that tag.

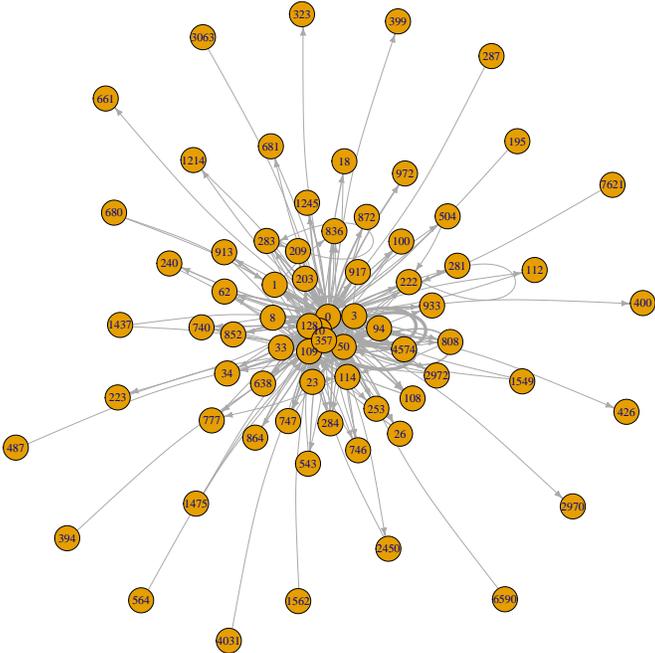
With this coding rule, I obtain a dataset of over 7 million “shifts of attention” between tags. Note that many answers have multiple tags. I handle the multi-tagging problem by including all pairs of combinations between the origin and the destination answers. For example, consider a user who posts an answer at time t_0 under tags A, B, and C and then posts an answer at t_1 under tag D, E ($t_1 > t_0$). Then attention shifts of the user at t_1 are: $A \rightarrow D$, $A \rightarrow E$, $B \rightarrow D$, $B \rightarrow E$, $C \rightarrow D$, $C \rightarrow E$.

To reduce sparsity, I coarsen the unit of time measurement into days, resulting in 365 time periods. I also take a subset of traffic between tags. First, I choose 100 tags with the activity throughout the year and subset edges flowing to one of them. Second, I count the number of

each pair of tags per day, for all 365 days in 2015. The two steps result in 1 million edges. Finally, I subset edges that have traffic in at least 20% of the days throughout the year of 2015, resulting in 1811 edges, a final subset of data I fit into the model.

With some simple exploratory analysis, snapshots of the daily network shows a high level of centrality. Most traffic comes from outside the political community. This is intuitive as this is a large social networking site where the political section is just one of its many communities. Also, within tags in the community, the level of centrality is high. Tags about major political entities and topics are much more popular than tags about more niche issues. Figure 2 shows a network of tags on January 10, 2015.

Figure 2: Attention Shifts among Tags on January 10, 2015



Note: A snapshot of flows of political discussion among tags with large traffic (top 100) in *Zhihu* on January 10, 2015. The graph shows 256 edges among 71 nodes. The width of an edge ($i \rightarrow j$) is proportional to $\log(\text{count})$.

3.3 Models

I apply a Bayesian dynamic network model to the *Zhihu* network dataset. In the decoupling step, I use the dynamic count mixture model with random effects, given the high level of centrality discovered in exploratory analyses and the eruptive features of social media political

discourse. For each pair of tags i and j , their edge formation at time t is modeled as follows:

$$\begin{aligned}
& \text{(observation)} \quad z_{ijt} \sim \text{Bern}(p_{ijt}) \\
& y_{ijt}|z_{ijt} = \begin{cases} 0 & \text{if } z_{ijt} = 0 \\ 1 + x_{ijt} & x_{ijt} \sim \text{Pois}(\lambda_{ijt}) \text{ if } z_{ijt} = 1 \end{cases} \\
& p_{ijt} \sim \text{Beta}(\alpha_{ijt}^b, \beta_{ijt}^b) \quad \lambda_{ijt} \sim \text{Gamma}(\alpha_{ijt}^g, \beta_{ijt}^g) \\
& \text{(link)} \quad \log \frac{p_{ijt}}{1 - p_{ijt}} = \mathbf{F}_{ijt}^0{}' \boldsymbol{\xi}_{ijt} \\
& \log(\lambda_{ijt}) = \tilde{\mathbf{F}}_{ijt}^+{}' \tilde{\boldsymbol{\theta}}_{ijt} \quad \text{where } \tilde{\mathbf{F}}_{ijt}^+ = [1, \mathbf{F}_t^+{}']' \text{ and } \tilde{\boldsymbol{\theta}}_{ijt} = [\zeta_{ijt}, \boldsymbol{\theta}_t']' \\
& \text{(evolution)} \quad \boldsymbol{\xi}_{t-1} + \mathbf{w}^0 \quad \mathbf{w}_t^0 \sim [\mathbf{0}, \mathbf{W}_t^0] \\
& \boldsymbol{\theta}_{ijt} = \tilde{\mathbf{G}}_{ijt}^+{}' \tilde{\boldsymbol{\theta}}_{ij,t-1} + \tilde{\mathbf{w}}_{ijt}^+ \quad \tilde{\mathbf{w}}_{ijt}^+ \sim [\mathbf{0}, \tilde{\mathbf{W}}_{ijt}^+]
\end{aligned}$$

For F and G , I use the Local Linear Growth Model. For discount factors, I let $\delta^{bern} = 0.95$, $\delta^{pois} = 0.90$, $\rho^{pois} = 0.90$. Other discount factors are experimented with, for evaluation purposes.

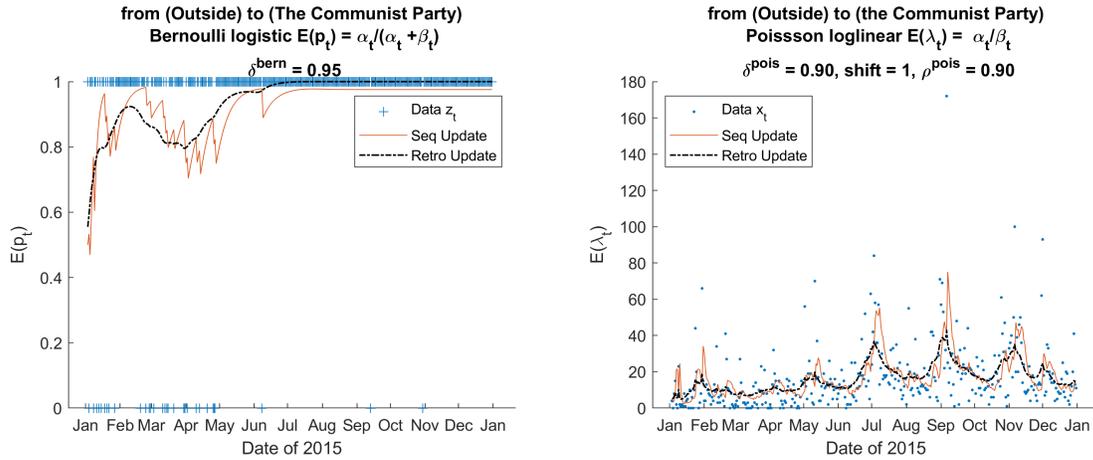
3.4 Results

The fitted model detects interesting patterns of shifts of attention among a diverse set of topics on *Zhuhu*. In this section, I present the results of an example pair of tags: traffic flow from outside the community to the tag “the communist party.” The traffic between the pair of tags suggests behavior patterns and regime strategies that are of interest to observers of Chinese politics.

3.4.1 Examining the Posteriors

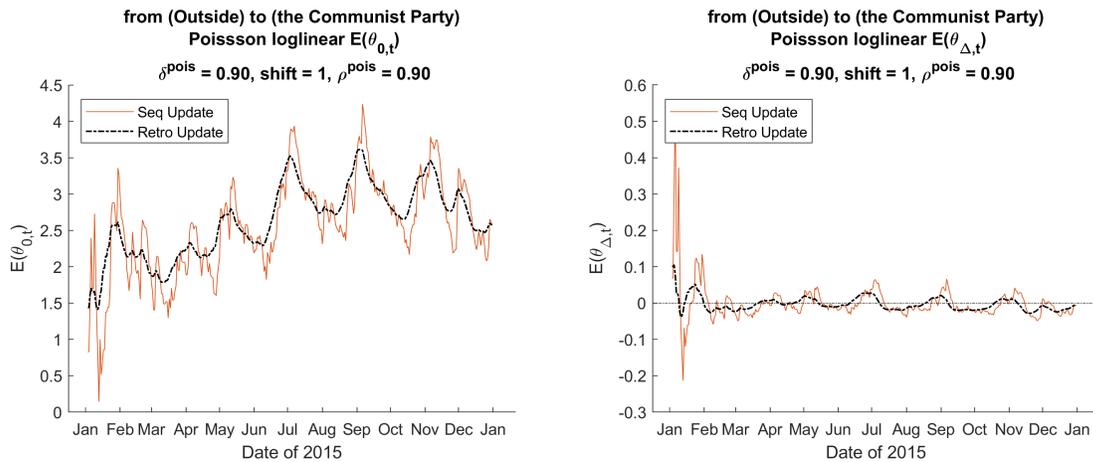
The fitted parameters of the observation model show two patterns: traffic started out sparse but quickly become regularly non-zero; several spikes of traffic are observed throughout the years. The patterns are shown in Figure 3. The left panel shows fitted parameters of the beta distribution of the Bernoulli distribution. The expected probability of non-negative flow fluctuates until August, after which it converges to 1. The right panel shows the fitted parameters of the conditional shifted Poisson distribution. Spikes are spotted in July, August, and November.

Figure 3: Sequential and Retrospective Updating of Parameters



The fitted state vector demonstrates the absolute value and the rate of change of the network flow. As discussed in Section 2.3.4, the Local Linear Growth Model outputs interpretable state vectors. As shown in the left panel of Figure 4, the expected value of $\theta_{0,t}$ shows a pattern similar to the Poisson mean. The right panel of Figure 4 shows that the rate of change remains relatively stable throughout the year. Except for the model’s “burn-in” period at the start, February observes a spike in the rate of change $\theta_{\Delta,t}$. This is the time when the pair of nodes start to have traffic. Other local maxima of $\theta_{\Delta,t}$ are located in July, September and November, in accordance to the spike in absolute values.

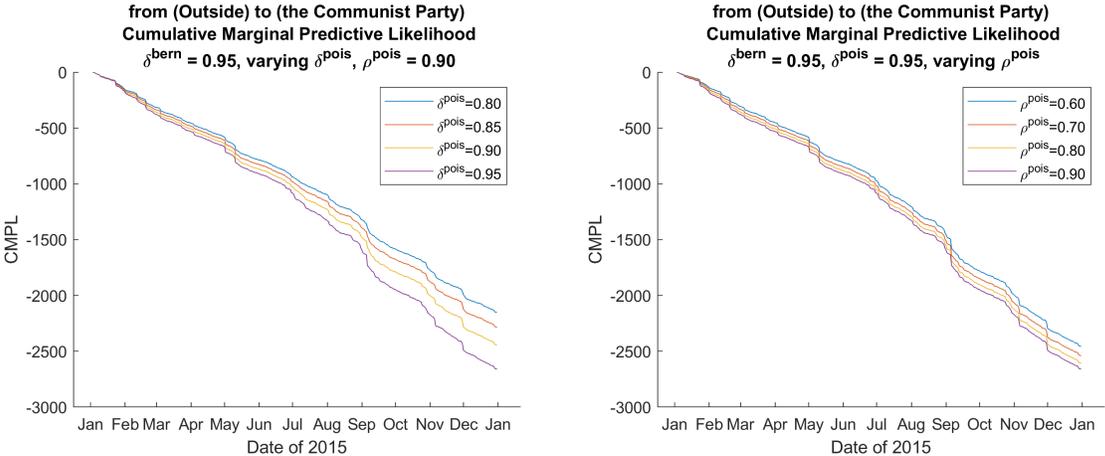
Figure 4: The State Vector of the Local Linear Growth Model



The models are evaluated with cumulative marginal predictive likelihood (CMPL). As

shown in Figure 5, CMPL decreases when the discount rates increase, as expected. Two observations are of interest. First, the impact of the discount factor for random effect on the result is small compared to that of the main discount factor. The lack of impact of the random effect on the result is also evident in my other experiments. Second, CMPL drops abruptly at a few points in time, suggesting sudden shifts that cannot be captured by the model.

Figure 5: Model Evaluation and Selection



3.4.2 Behavioral Patterns Detected

Detecting sudden shifts of attention as data streams in would be of interest to researchers studying behavior of political communication. I use marginal predictive likelihood of one-step-ahead prediction to detect abnormally abrupt shifts of attention between the example pair of tags. Figure 6 shows one-step-ahead prediction, marginal predictive likelihood (MPL), and cumulative marginal predictive likelihood (CMPL) (from top to bottom in the figure). Points whose MPLs are lower than the 5th percentile of all are marked with blue 'x'. Upon further review, it becomes clear that most of the highlighted points map to important events in China.

The first group of low-MPL points appears around February. The high traffic may be trivially attributable to the holiday effect: the Chinese new year is on February 19. Around the holiday, people may have the time and inclination to participate in online discussion. The high traffic is unlikely politically motivated. The second group of low-MPL points appears

in May, right after another major public holiday in China: the labor day. The cause may be a mixture of holiday and politics. The labor day is a week-long holiday, giving people the time to talk online. In addition, this is a holiday deeply associated with communism, which likely prompts people's interest in joining discussions about the communist party. The third low-MPL point appears at the start of July, when the Chinese Communist Party celebrates its founding day, another politically relevant anomaly.

The next two anomalies in MPL are likely associated with special political events in 2015. A spike in August appears around the time of a tragic explosion in Tianjin, China, that killed 173 people, for which the local government was severely criticized. The anomaly in September, the largest spike throughout the whole series, likely corresponds to a grand military parade on September 3 to commemorate the 70th anniversary of the end of World War II. Interestingly, the traffic spikes and then abruptly drops.⁶

The above interpretation suggests two types of sources for anomalies: regularly-scheduled events such as holidays and political events that increase interest in the communist party. The dynamic gravity model in the recoupling step can help to distinguish the two types of effect. Figure 7 shows the samples of one-step-ahead prediction decomposed into four types of effect. Panel (A) shows the overall traffic in this community of political discussion. Mid-February shows two data points much higher than the upper bound of the credible interval, confirming the hypothesis about the Chinese new year holiday effect. Such anomalies are also evident in the July, August, and September spikes, suggesting that discussion triggered by the political events can take place in other topics in the community. Panel (C) shows the destination effect of the tag "the communist party." Data points significantly beyond the 95% credible interval suggest support for my hypotheses about the politically-motivated spikes in early July, mid August, and September. However, it finds other spikes that have not been detected by the decoupled data. Panel (D) shows the affinity effect. Variances are large and signals of patterns are weak in this sequence. However, the spikes in July, August, and September are still in evidence, meaning that the anomalies may be driven by new-comers to the community rather than existing contributors.

⁶To find explanation for the anomalies, I refer to a news report on China's top news search in 2015 by the New York Times: <https://cn.nytimes.com/china/20151225/c25searches/dual/>

3.4.3 Censorship and Fabrication Suggested

Moving beyond individual outliers to a more intriguing finding, results of the dynamic graphic model suggest the existence of strategic censorship and fabrication of online discourse on the site at extremely politically sensitive time periods. First, an anomaly of traffic in June suggests the existence of censorship. As shown in Panel (A) of Figure 7, the traffic around June slumps: it starts to decrease in mid-May, reaches a minimum at the beginning of June, and remains low until late-June. The sudden decrease is abnormal, considering the steady growth in the preceding months and the stably high level in the months afterwards. Students of Chinese politics can intuitively attribute it to the regime’s annual mass censorship campaign around June 4th. Ever since the Tiananmen incident on June 4th, 1989, discussion of the event has been a political taboo. In the social media era, the authorities put a lot of effort into removing any discussion about the event, especially in the days around its anniversary. The time-specific effect is evidence of this effort.

Furthermore, the origin, destination, and affinity effects show that the censorship is strategic and suggest evidence of fabrication of social media discourse by the regime. As evident in Panel (B) of Figure 7, the origin effect significantly increases in June, amid the slumped time-specific effect. This signals a large inflow of outsiders (i.e., first-time contributors) to this community of political discussion. A possible explanation, consistent with findings of King et al. (2017), is that commentators working for the government flood into the community in this politically sensitive period to distract other members away from political discussions.⁷ Panel (C) shows that the destination effect of the tag “the communist party” also experiences an increase in the extremely politically sensitive period. This suggests that censorship is strategic and discussion about the ruling party is tolerated and even encouraged during this period. It would take close examination of posts to tell what content is encouraged. Finally, Panel (D) shows that the affinity effect is steady in June. Its lack of pattern indicates that the tag “the communist party” is probably not the main target of the outsiders’ inflow in June.

⁷Admittedly, an alternative explanation is that people’s interest in politics is high in general in this special period, leading to a general inflow. Ruling out this possibility requires close examination of account information of the new contributors.

3.5 Discussion

I apply Bayesian dynamic network modeling to study how attention shifts in political discussion on a Chinese social networking site, *Zhihu*. I use a count mixture model and a dynamic gravity model to model traffic among 100 most popular tags in the community for political discussion. I demonstrate the model by describing the flow of new contributors into the tag “the communist party.” My results detect both behavioral patterns of individuals and suggest possible intervention from the regime. Spikes of traffic during holidays and around political events suggest participation in political discussion is driven by both political and non-political motives. A slump of time-specific traffic accompanied by a high origin effect and normal destination effect in June suggest the existence of strategic censorship and the fabrication of social media discourse during politically the sensitive period.

It is worth re-iterating that, thanks to its online learning capacity, the dynamic network model can alert researchers to these anomalies as data stream in. This is especially relevant to studies on authoritarian politics, as many unexpected political events can create risks that researchers want to control and opportunities which researchers want to seize during their data collection.

4 Conclusion

In recent years, social media has become a new data source for empirical political science research. Researchers use social media to study political behavior and institutions. Among recent substantive and methodological contributions to political science research with social media network data, few works have explored its time dynamics. Most existing works analyze this type of data as cross-sectional data or repeated measures with only a few time points. Without attention to the time dynamics, political scientists may miss opportunities to understand an important source of variation.

In this chapter, I introduce the Bayesian dynamic network modeling method, a tool helping political scientists analyze trends of network flows in streaming social media data. The method uses a decoupling/recoupling strategy: it first decouples network data into univariate time series sequences to learn the time dynamics, and then it recouples the fitted models to learn the network structure. In the decoupling step, I fit data to a dynamic count mixture model

to account for variances in eruptive and sporadic social media data. In the recoupling step, I fit samples of posterior outcomes to a dynamic gravity model to attribute network flows to time-specific, origin, destination, and affinity effects.

I illustrate the method with its application to an original dataset of political discourse on a Chinese social networking site, *Zhihu*. The application studies how political attention shifts in authoritarian China and how the government controls it. With the model, I find evidence of individual behavior and governmental intervention: spikes of discussion about the communist party are associated with major political events; an anomaly of traffic in June suggests strategic censorship and fabrication by the government.

The method can be extended in both the decoupling and recoupling steps to accommodate more complex research designs. In the decoupling step, I have so far focused on local growth models that assume short time dependency and incorporate no covariates. In future extensions, the model can consider a more complex structure of time dependency (e.g., seasonality) and explanatory variables. In the recoupling step, the gravity model can be extended to consider node-level and edge-level covariates in explaining edge formation.

Finally, with this chapter, I attempt to contribute to an ongoing discussion on big data in political science research. New sources of big data deviate from the conventional forms of data in this field as they are generated every minute and exceed researchers' storage capacity. This challenges political scientists to develop scalable computational methods that can inform them about data's real-time patterns and alert them to anomalies. This is important for both the generation of research ideas and hypotheses as well as the monitoring of ongoing data collection efforts. The Bayesian dynamic network modeling method introduced in this chapter attempts to address this challenge. The potential of this family of models for big-data political science studies using social media data deserves more attention in future research.

Figure 6: Prediction and Anomaly Detection

from (outside) to (the Communist Party)
Dynamic Count Mixture Model

$$\delta^{\text{bern}} = 0.95, \delta^{\text{pois}} = 0.90, \text{shift} = 1, \rho^{\text{pois}} = 0.90$$

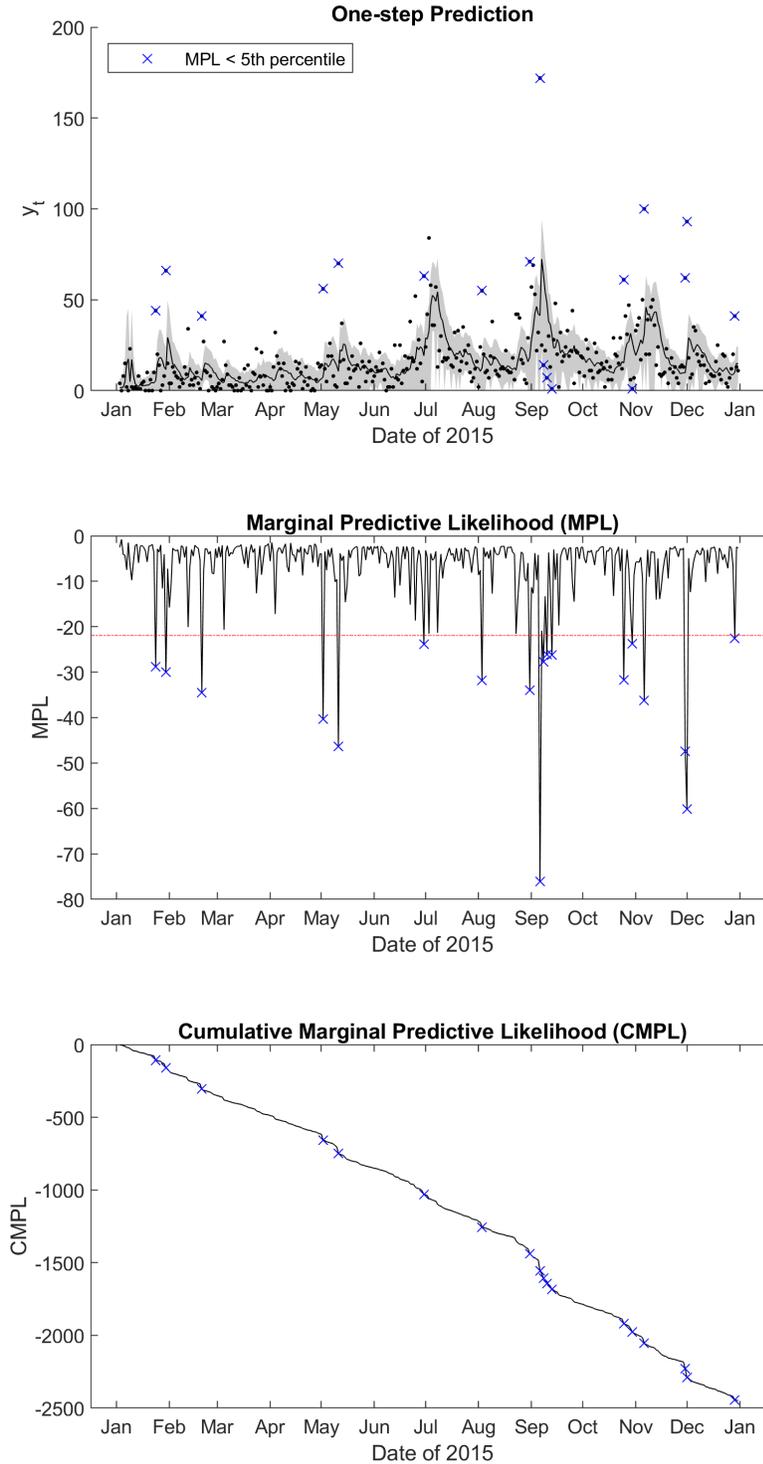
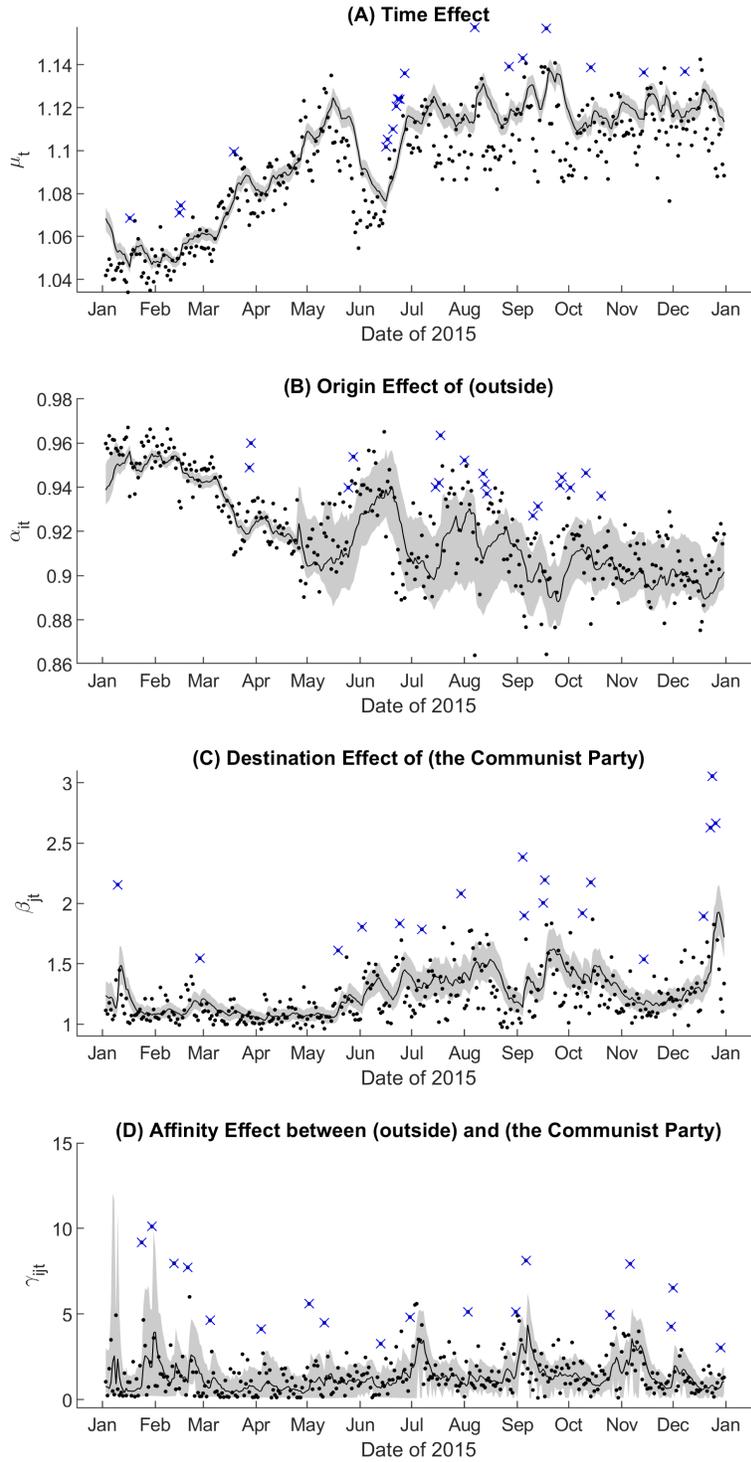


Figure 7: Dynamic Gravity Model with Sequentially Updated DCMM

from (outside) to (the Communist Party)
 Dynamic Gravity Model
 with DCMM seq. update $\delta^{\text{bern}}=.95$, $\delta^{\text{pois}}=.90$, $\rho^{\text{pois}}=.90$



References

- Bail, C. A., L. P. Argyle, T. W. Brown, J. P. Bumpus, H. Chen, M. B. F. Hunzaker, J. Lee, M. Mann, F. Merhout, and A. Volfovsky (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences* 115(37), 9216–9221.
- Bakshy, E., S. Messing, and L. A. Adamic (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science* 348(6239), 1130–1132.
- Baldassarri, D. and M. Abascal (2017). Field experiments across the social sciences. *Annual Review of Sociology* 43(1), 41–73.
- Barberá, P. (2015). Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Political Analysis* 23(1), 76–91.
- Barberá, P., J. T. Jost, J. Nagler, J. A. Tucker, and R. Bonneau (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological science* 26(10), 1531–42.
- Beck, N. and J. N. Katz (2011). Modeling dynamics in time-series - cross-section political economy data. *Annual Review of Political Science* 14(1), 331–352.
- Bond, R. and S. Messing (2015). Quantifying social media’s political space: estimating ideology from publicly revealed preferences on Facebook. *American Political Science Review* 109(1), 62–78.
- Bond, R. M., C. J. Fariss, J. J. Jones, A. D. Kramer, C. Marlow, J. E. Settle, and J. H. Fowler (2012). A 61-million-person experiment in social influence and political mobilization. *Nature* 489(7415), 295–298.
- Box-Steffensmeier, J. M., J. R. Freeman, M. P. Hitt, and J. C. W. Pevehouse (2014). *Time Series Analysis for the Social Sciences*. Cambridge University Press.
- Chen, J., J. Pan, and Y. Xu (2016). Sources of authoritarian responsiveness: A field experiment in China. *American Journal of Political Science* 60(2), 383–400.
- Clinton, J., S. Jackman, and D. Rivers (2004). The statistical analysis of roll call data. *American Political Science Review* 98(2), 355–370.
- de Rooij, E. A., D. P. Green, and A. S. Gerber (2009). Field experiments on political behavior and collective action. *Annual Review of Political Science* 12(1), 389–395.
- Dorff, C. and M. D. Ward (2013). Networks, dyads, and the social relations model. *Political Science Research and Methods* 1(2), 159–178.
- Grimmer, J. (2010). A Bayesian hierarchical topic model for political texts: Measuring expressed agendas in senate press releases. *Political Analysis* 18(1), 1–35.
- Hobbs, W. and M. E. Roberts (2018). How sudden censorship can increase access to information. *American Political Science Review* 112(3), 621–636.

- Hoff, P., B. Fosdick, A. Volfovsky, and K. Stovel (2013). Likelihoods for fixed rank nomination networks. *Network Science* 1(3), 253–277.
- Hoff, P. D. (2009). Multiplicative latent factor models for description and prediction of social networks. *Computational and Mathematical Organization Theory* 15(4), 261–272.
- Hoff, P. D. (2015). Dyadic data analysis with amen.
- Jennings, W. and P. John (2009). The dynamics of political attention: public opinion and the Queen’s speech in the United Kingdom. *American Journal of Political Science* 53(4), 838–854.
- Keele, L. and S. DeBoef (2008). Taking time seriously. *American Journal of Political Science* 52(1), 184–200.
- Keele, L. and N. J. Kelly (2006). Dynamic models for dynamic theories: The ins and outs of lagged dependent variables. *Political Analysis* 14(2), 186–205.
- King, G., J. Pan, and M. Roberts (2013). How censorship in China allows government criticism but silences collective expression. *American Political Science Review* 107(917), 326–343.
- King, G., J. Pan, and M. E. Roberts (2014). Reverse-engineering censorship in China : Randomized experimentation and participant observation. *Science* 345(6199), 981.
- King, G., J. Pan, and M. E. Roberts (2017). How the Chinese government fabricates social media posts for strategic distraction, not engaged argument. *American Political Science Review* 111(3), 484–501.
- Meng, T., J. Pan, and P. Yang (2017). Conditional receptivity to citizen participation: Evidence from a survey experiment in China. *Comparative Political Studies* 50(4), 399–433.
- Miller, B. (2018). The limits of commercialized censorship in China. *Working paper*.
- Muise, D. and J. Pan (2019). Online field experiments. *Asian Journal of Communication* 29(3), 217–234.
- Munger, K. (2017a). Experimentally reducing partisan incivility on Twitter. *Working Paper*.
- Munger, K. (2017b). Tweetment effects on the tweeted: Experimentally reducing racist harassment. *Political Behavior* 39(3), 629–649.
- Munger, K., R. Bonneau, J. Nagler, and J. A. Tucker (2018). Elites tweet to get feet off the streets: Measuring regime social media strategies during protest. *Political Science Research and Methods*, 1–20.
- Parigi, P., J. J. Santana, and K. S. Cook (2017). Online field experiments: Studying social interactions in context. *Social Psychology Quarterly* 80(1), 1–19.
- 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.

- Ramirez, M. D. (2009). The dynamics of partisan conflict on congressional approval. *American Journal of Political Science* 53(3), 681–694.
- Roberts, M. E. (2014). Fear or friction? How censorship slows the spread of information in the digital age.
- Rose, A. K. (2004). Do we really know that the WTO increases trade? *American Economic Review* 94(1), 98–114.
- Tomz, M., J. L. Goldstein, and D. Rivers (2007). Do we really know that the WTO increases trade? Comment. *American Economic Review*, 97(5), 2005–2018.
- Ward, M. D., J. S. Ahlquist, and A. Rozenas (2013). Gravity’s rainbow: A dynamic latent space model for the world trade network. *Network Science* 1(1), 95–118.
- West, M. (2020). Bayesian forecasting of multivariate time series: Scalability, structure uncertainty and decisions (with discussion). *Annals of the Institute of Statistical Mathematics* (forthcoming).