Bayesian Analysis (0000) **00**, Number 0, pp. 1

A Network Model for Dynamic Textual Communications with Application to Government Email Corpora

Bomin Kim*, Aaron Schein†, Bruce A. Desmarais‡, and Hanna Wallach§

Abstract. We introduce the interaction-partitioned topic model (IPTM)—a probabilistic model for who communicates with whom about what, and when. Broadly speaking, the IPTM partitions timestamped textual communications, according to both the network dynamics that they reflect and their content. To define the IPTM, we integrate the hyperedge event model (HEM)—a generative model for events that can be represented as directed edges with one sender and one or more receivers or one receiver and one or more senders—and latent Dirichlet allocation (LDA)—a generative model for topic-based content. The IPTM assigns each document to an "interaction pattern"—a generative process for contents and ties that is governed by a topic distribution and a set of dynamic network features. We use the IPTM to analyze emails sent between department managers in Dare county government in North Carolina, and demonstrate that the model is effective at predicting and explaining continuous-time textual communications.

MSC 2010 subject classifications: Primary 60K35, 60K35; secondary 60K35.

Keywords: dynamic network model, topic modeling, text analysis, email data analysis.

1 Introduction

In recent decades, real-time digitized textual communication has developed into a ubiquitous form of social and professional interaction (Kanungo and Jain, 2008; Szóstek, 2011; Burgess et al., 2004; Pew, 2016). From the perspective of the computational social scientist, this has lead to a growing need for methods of modeling interactions that manifest as text exchanged in continuous time. A number of models that build upon topic modeling through Latent Dirichlet Allocation (LDA) (Blei et al., 2003) to incorporate link data as well as textual content have been developed recently (McCallum et al., 2005; Lim et al., 2013; Krafft et al., 2012). These models are innovative in their extensions that incorporate network tie information. However, none of the models that are currently available in the literature integrate the rich random-graph structure offered by state of the art models for network structure—such as the exponential random graph model (ERGM) (Robins et al., 2007; Chatterjee et al., 2013; Hunter et al., 2008). The ERGM is the canonical model for modeling the structure of a static network. It is

*Department of Statistics, Pennsylvania State University bzk147@psu.edu
†College of Information and Computer Sciences, UMass Amherst aschein@cs.umass.edu
‡Department of Political Science, Pennsylvania State University bdesmarais@psu.edu
§ Microsoft Research NYC hanna@dirichlet.net

 DOI: 0000

flexible enough to specify a generative model that accounts for nearly any pattern of tie formation (e.g., reciprocity, clustering, popularity effects) (Desmarais and Cranmer, 2017). Several models have been developed that handle time-stamped ties in which tie formation is governed by structural dynamics similar to those used in ERGMs (Perry and Wolfe, 2013; Butts, 2008; Snijders, 1996). We develop the interaction-partitioned topic model (IPTM) which simultaneously models the network structural patterns that govern time-stamped tie formation, and the content in the communications.

The models on which we build, including the relational event model (Butts, 2008), the point process model (Perry and Wolfe, 2013), and most closely the hyperedge event model (HEM), provide frameworks for explaining or predicting ties between nodes using the network sub-structures in which the two nodes are embedded (e.g., predict a tie is highly likely to form between two nodes if those two nodes have many shared partners). Models based on network structure have been used for many applications in which the ties between nodes are annotated with text. The text, despite providing rich information regarding the strength, scope, and character of the ties, has been largely excluded from these analyses, due to the inability of these network models to incorporate textual attributes of ties. These application domains include, among other applicaitons, the study of legislative networks in which networks reflect legislators' co-support of bills, but exclude bill text (Bratton and Rouse, 2011; Alemán and Calvo, 2013); the study of alliance networks in which networks reflect countries' co-signing of treaties, but exclude treaty text (Camber Warren, 2010; Cranmer et al., 2012b,a; Kinne, 2016); the study of scientific co-authorship networks that exclude the text of the co-authored papers (Kronegger et al., 2011; Liang, 2015; Fahmy and Young, 2016); and the study of text-based interaction on social media (e.g., users tied via 'mentions' on twitter) (Yoon and Park, 2014; Peng et al., 2016; Lai et al., 2017).

In defining and testing the IPTM we embed core conceptual property—interaction pattern—to link the content component of the model, and network component of the model such that knowing who is communicating with whom at what time (i.e., the network component) provides information about the content of communication, and vice versa (Section 2). The IPTM leads to an efficient inference using the Markov Chain Monte Carlo (MCMC) algorithm (Section 3) and acheives good predictive peformance (Section **??**). Finally, the IPTM discovers interesting and interpretable latent structure through application to email corpora of internal communications by government officials in Dare County, NC (Section **??**).

2 Interaction-partitioned Topic Model

In this section we define the IPTM by describing a process according to which documents are generated in continuous time. Data generated under the IPTM consists of D unique documents. A single document indexed by $d \in [D]$—where $[D]$ denotes a categorical set with D levels $[D] = \{1, \dots, D\}$—is represented by the four components: the sender $s_d \in [A]$, an indicator vector of receivers $\boldsymbol{r}_d$, where $r_{dj} = 1$ if $j \in [A]$ is a receiver of document d and 0 otherwise, the timestamp $t_d \in (0, \infty)$, and a set of word counts $\boldsymbol{w}_d = \{w_{dv}\}_{v=1}^{V}$—where each w_{dv} is an integer greater or equal to zero and V

is the size of vocabularies—that comprise the text of the document. For simplicity, we assume that documents are ordered by time such that $t_d < t_{d+1}$.

2.1 Interaction Patterns

They key idea that combines the IPTM component modeling "what" with the component modeling "who," "whom," and "when" is that different documents comes from the introduction of "interaction patterns." Each interaction pattern $c \in [C]$ is characterized by a set of features that affects networks and timestamps—such as the number of messages sent from i to j in some time interval—and corresponding coefficients. We associate each document with the interaction pattern that best describes how people interact, and that is reflected to what people talk about via topic assignments. To be specific, each document $d \in [D]$ draws an interaction pattern c_d

$$c_d \sim \text{Categorical}(\frac{\psi_1}{\sum_{c=1}^{C} \psi_c}, \ldots, \frac{\psi_C}{\sum_{c=1}^{C} \psi_c}), \tag{2.1}$$

and we assume the interaction pattern-specific factors are gamma-distributed

$$\psi_c \sim \Gamma(\epsilon_0, \epsilon_0), \tag{2.2}$$

where we place an uninformative gamma prior over these shape and rate parameters.

2.2 Generative Process for Network Dynamics

The generative process for network portion of the documents—the senders, receivers, and timestamps—exactly follows that of the hyperedge event model (HEM), except that now we consider the interaction pattern assignments of the documents in order to understand the differences in network dynamics across the interaction patterns. While the HEM can be applied for two types of hyperedges—edges with (1) one sender and one or more receivers, and (2) one or more senders and one receiver—here we only present the generative process for those involving one sender and one or more receivers, considering our email applications.

Candidate receivers

Given the interaction pattern of document d—i.e., c_d, we define the "receiver intensity" for every possible sender–receiver pair (i, j) where $i \neq j$ as a linear combination of statistics relevant to the receiver selection process:

$$\lambda_{idj} = \boldsymbol{b}_{c_d}{}^{\top} \boldsymbol{x}_{idjc_d}, \tag{2.3}$$

where $\boldsymbol{b}_c$ is a P–dimensional vector of coefficients and $\boldsymbol{x}_{idjc}$ is a set of receiver selection features which vary depending on the hypotheses regarding canonical processes relevant to network theory such as popularity, reciprocity, and transitivity, as well as the effects of attributes of the sender and receivers, or sender–receiver pairs. We place a Normal prior $\boldsymbol{b}_c \sim N(\boldsymbol{\mu}_b, \Sigma_b)$.

We then draw a binary receiver vector $\boldsymbol{u}_{id} = (u_{id1}, \dots, u_{idA})$ from a probability measure "MB_G", which helps us to 1) allow a sender to select multiple receivers for a single document, 2) force the sender to select at least one receiver, and 3) ensure a tractable normalizing constant for the receiver selection distribution. To be specific, we assume

$$\boldsymbol{u}_{id} \sim \mathrm{MB}_G(\boldsymbol{\lambda}_{id}), \tag{2.4}$$

where $\boldsymbol{\lambda}_{id} = (\lambda_{id1}, \dots, \lambda_{idA})$. In particular, we define $\mathrm{MB}_G(\boldsymbol{\lambda}_{id})$ as

$$\Pr(\boldsymbol{u}_{id}|\boldsymbol{b}, \boldsymbol{x}_{id}, c_d) = \frac{1}{Z(\boldsymbol{\lambda}_{id})} \exp\Big(\log\big(\mathrm{I}(\|\boldsymbol{u}_{id}\|_1 > 0)\big) + \sum_{j \neq i} \lambda_{idj} u_{idj}\Big), \tag{2.5}$$

where $Z(\boldsymbol{\lambda}_{id}) = \prod_{r \neq a} \big(\exp(\lambda_{idj}) + 1\big) - 1$ is the normalizing constant and is the ℓ_1-norm, and the log-indicator term $\log\big(\mathrm{I}(\|\boldsymbol{u}_{id}\|_1 > 0)\big)$ ensures that empty receiver sets are excluded from the distribution's support. This is approximately equivalent to assuming that each u_{idj} is drawn with the probability of 1 being $\mathrm{logit}(\lambda_{idj})$, excluding the case when $u_{idj} = 0$ for all $j \in [A]$.

Candidate timestamps

Similarly, for each sender and document combination, our model draws a candidate timestamp at which the document would be sent given the candidate sender and receiver combinations. Given the interaction pattern assignment c_d, the timing rate for sender i and document d is

$$\mu_{id} = g^{-1}(\boldsymbol{\eta}_{c_d}^{\top} \boldsymbol{y}_{idc_d}), \tag{2.6}$$

where $\boldsymbol{\eta}_c$ is a Q–dimensional vector of coefficients with a Normal prior $\boldsymbol{\eta}_c \sim N(\boldsymbol{\mu}_\eta, \Sigma_\eta)$, $\boldsymbol{y}_{adc}$ is a set of event timing features—covariates that could affect timestamps of the document, and $g(\cdot)$ is the appropriate link function such as identity, log, or inverse.

Next, following the generalized linear model (GLM) framework (Nelder and Baker, 1972), we assume the mean and variance of the τ_{id}—i.e., each sender's "time increment" from document $d-1$ to document d or $t_d - t_{d-1}$—satistify

$$\begin{aligned} E(\tau_{id}) &= \mu_{id}, \\ V(\tau_{id}) &= V(\mu_{id}), \end{aligned} \tag{2.7}$$

where τ_{id} here is a positive real number drawn from a specific distribution among the possible choices including exponential, Weibull, gamma, and log-normal. We may need other latent variables to draw the time increment based on the choice of distribution, to account for the variance of time increments, beyond the coefficients for the features used to model the rate. $V(\mu)$—e.g., the shape parameter k for the Weibull, the shape parameter θ for the gamma, and the variance parameter σ_τ^2 for the log-normal. We use $f_\tau(\cdot; \mu, V(\mu))$ and $F_\tau(\cdot; \mu, V(\mu))$ to denote the probability density function (p.d.f) and cumulative density function (c.d.f), respectively, with mean μ and variance $V(\mu)$.

Senders, receivers, and timestamps

Finally, under the IPTM the observed sender, receivers, and timestamp of document d is generated by selecting the sender–receiver-set pair with the smallest time increment (Snijders, 1996), along with the words generated according to Section 2.3:

$$\begin{aligned} s_d &= \text{argmin}_i(\tau_{id}), \\ \boldsymbol{r}_d &= \boldsymbol{u}_{s_d d}, \\ t_d &= t_{d-1} + \tau_{s_d d}, \end{aligned} \tag{2.8}$$

Therefore, it is a sender-driven process in that the receivers and timestamps of a document are jointly determined by the sender's urgency to send the document to the selected receivers. Note that our generative process accounts for tied events such that in case of tied documents—i.e., multiple senders draw exactly the same candidate timestamps—we assume that all documents are generated and occur simultaneously.

2.3 Generative Process for Contents

Given the sender-receiver-timestamps, the words $\boldsymbol{w}_d$ are generated by extending the well-known Bayesian topic model, latent Dirichlet allocation (LDA) (Blei et al., 2003), to follow the form of Bayesian Poisson Tucker decomposition (BPTD) (Schein et al., 2016). As in LDA, we generate the corpus-wide global variables that describe the content via topics. First, the topic-word factors for each topic $k \in [K]$ and word type $v \in [V]$ and the interaction pattern-topic factors for each interaction pattern $c \in [C]$ are both gamma-distributed, where we again assume that these factors are drawn directly from an uninformative gamma prior

$$\begin{aligned} \phi_{kv} &\sim \Gamma(\epsilon_0, \epsilon_0), \\ \theta_{ck} &\sim \Gamma(\epsilon_0, \epsilon_0). \end{aligned} \tag{2.9}$$

Next, we generate the sender-specific factors for each $i \in [A]$ from uninformative gamma prior

$$\pi_i \sim \Gamma(\epsilon_0, \epsilon_0), \tag{2.10}$$

which serves as the weight of sender i such that a person with higher weight tends to write a document using more words. Finally, given the specifications above, the number of words of type $v \in [V]$ in document d is drawn from

$$w_{dv} \sim \text{Poisson}(\pi_{s_d} \sum_{c=1}^{C} \sum_{k=1}^{K} I_{dc}\theta_{ck}\phi_{kv}), \tag{2.11}$$

where I_{dc} is an indicator for $I(c_d = c)$. Note that Equation (2.11) above is identical to $w_{dv} \sim \text{Poisson}(\pi_{s_d} \sum_{k=1}^{K} \theta_{c_d k}\phi_{kv})$. Also note that $\boldsymbol{w}_d = (w_{d1}, \ldots, w_{dV})$ is a very sparse vector with $\sum(\boldsymbol{w}_d) = N_d$, where N_d denotes the total number of words in a document. Algorithm 1 provides a summary on Section 2, which presents the entire generative process for documents given the gamma-distributed factors.

Algorithm 1 Generative Process: one sender and one or more receivers

Input: number of documents and nodes (D, A), gamma weights $(\boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\pi})$, covariates $(\boldsymbol{x}, \boldsymbol{y})$, and coefficients $(\boldsymbol{b}, \boldsymbol{\eta})$

for $d = 1$ to D **do**
  draw $c_d \sim \text{Categorical}(\frac{\psi_1}{\sum_c \psi_c}, \ldots, \frac{\psi_C}{\sum_c \psi_c})$
  for $i = 1$ to A **do**
    for $j = 1$ to A $(j \neq i)$ **do**
      set $\lambda_{idj} = \boldsymbol{b}_{c_d}^\top \boldsymbol{x}_{idj}$
    end for
    draw $\boldsymbol{u}_{id} \sim \text{MB}_G(\boldsymbol{\lambda}_{id})$
    set $\mu_{id} = g^{-1}(\boldsymbol{\eta}_{c_d}^\top \boldsymbol{y}_{id})$
    draw $\tau_{id} \sim f_\tau(\mu_{id}, V(\mu))$
    for $v = 1$ to V **do**
      $w_{dv} \sim \text{Poisson}(\pi_{\text{argmin}_i(\tau_{id})} \sum_{c=1}^{C} \sum_{k=1}^{K} I_{dc}\theta_{ck}\phi_{kv})$
    end for
  end for
  if $n \geq 2$ tied events **then**
    set $s_d, \ldots, s_{d+n-1} = \text{argmin}_i(\tau_{id})$,
    set $\boldsymbol{r}_d = \boldsymbol{u}_{s_d d}, \ldots, \boldsymbol{r}_{d+n-1} = \boldsymbol{u}_{s_{d+n-1} d}$
    set $t_d, \ldots, t_{d+n-1} = t_{d-1} + \min_i \tau_{id}$
    set $\boldsymbol{w}_d = (w_{d1}, \ldots, w_{dV})$
    jump to $e = d + n$
  else
    set $s_d = \text{argmin}_i(\tau_{id})$
    set $\boldsymbol{r}_d = \boldsymbol{u}_{s_d d}$
    set $t_d = t_{d-1} + \min_i \tau_{id}$
    set $\boldsymbol{w}_d = (w_{d1}, \ldots, w_{dV})$
  end if
end for

3 Posterior inference

In this section we describe how we invert the generative process to obtain the posterior distribution over the latent variables—interaction pattern factors $\{\psi_c\}_{c=1}^C$ and assignments $\{c_d\}_{d=1}^D$, topic-word factors $\{\{\phi_{kv}\}_{v=1}^V\}_{k=1}^K$, interaction pattern-topic factors $\{\{\theta_{ck}\}_{k=1}^K\}_{c=1}^C$, document-specific factors $\{\pi_i\}_{i=1}^A$, candidate receivers $\{\boldsymbol{u}_d\}_{d=1}^D$, coefficients for receiver selection features $\{\boldsymbol{b}_c\}_{c=1}^C$, and coefficients for event timing features $\{\boldsymbol{\eta}_c\}_{c=1}^C$—conditioned on the observed data $\{(s_d, \boldsymbol{r}_d, t_d, \boldsymbol{w}_d)\}_{d=1}^D$, covariates $\{(\boldsymbol{x}_d, \boldsymbol{y}_d)\}_{d=1}^D$, and hyperparamters $(\boldsymbol{\mu}_b, \Sigma_b, \boldsymbol{\mu}_\eta, \Sigma_\eta, \gamma_0, \xi, \epsilon_0)$. We draw the samples using Markov chain Monte Carlo (MCMC) methods, repeatedly resampling the value of each latent variable from its conditional posterior via a Metropolis-within-Gibbs sampling algorithm. In the following, we provide each latent variable's conditional posterior along with pseudocode of MCMC in Algorithm 2.

Interaction pattern factors

$$
\begin{aligned}
\Pr(\psi_c|\boldsymbol{\psi}_{\setminus c}, \boldsymbol{c}) &\propto \Pr(\boldsymbol{c}|\psi_c, \boldsymbol{\psi}_{\setminus c}) \times \Pr(\psi_c) \\
&\propto (\prod_{d=1}^{D} \frac{\psi_{c_d}}{\sum_{c=1}^{C} \psi_c}) \times \psi_c^{\epsilon_0 - 1} \times \exp(-\epsilon_0 \psi_c) \\
&\propto (\psi_c)^{\sum_{d=1}^{D} I_{dc} + \epsilon_0 - 1} \times \exp(-\epsilon_0 \psi_c) \\
&\sim \Gamma(\sum_{d=1}^{D} I_{dc} + \epsilon_0, \epsilon_0)
\end{aligned} \tag{3.1}
$$

Topic-word factors

$$
\begin{aligned}
\Pr(\phi_{kv}|\, \boldsymbol{\phi}_{\setminus k,v}, \boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{\pi}, \boldsymbol{c}) &\propto \Pr(\boldsymbol{w}_{kv}|\, \phi_{kv}, \boldsymbol{\phi}_{\setminus k,v}, \boldsymbol{\theta}, \boldsymbol{\pi}, \boldsymbol{c}) \times \Pr(\phi_{kv}) \\
&\propto (\prod_{d=1}^{D} \frac{(\pi_{s_d}\theta_{c_d k}\phi_{kv})^{w_{dkv}} \times \exp(-\pi_{s_d}\theta_{c_d k}\phi_{kv})}{w_{dkv}!}) \times \phi_{kv}^{\epsilon_0 - 1} \times \exp(-\epsilon_0 \phi_{kv}) \\
&\propto (\phi_{kv})^{\sum_{d=1}^{D} w_{dkv} + \epsilon_0 - 1} \times \exp(-(\epsilon_0 + \sum_{d=1}^{D} \pi_{s_d}\theta_{c_d k})\phi_{kv}) \\
&\sim \Gamma(\sum_{d=1}^{D} w_{dkv} + \epsilon_0, \sum_{d=1}^{D} \pi_{s_d}\theta_{c_d k} + \epsilon_0),
\end{aligned} \tag{3.2}
$$

where $(w_{d1v}, \ldots, w_{dKv}) \sim \text{Multinomial}(w_{dv}, \{\theta_{c_d k}\phi_{kv}\}_{k=1}^{K})$.

Interaction pattern-topic factors

$$
\begin{aligned}
\Pr(\theta_{ck}|\, \boldsymbol{\theta}_{\setminus c,k}, \boldsymbol{\phi}, \boldsymbol{w}, \boldsymbol{\pi}, \boldsymbol{c}) &\propto \Pr(\boldsymbol{w}_{ck}|\, \theta_{ck}, \boldsymbol{\theta}_{\setminus c,k}, \boldsymbol{\phi}, \boldsymbol{\pi}, \boldsymbol{c}) \times \Pr(\theta_{ck}) \\
&\propto (\prod_{d:c_d=c} \prod_{v=1}^{V} \frac{(\pi_{s_d}\theta_{ck}\phi_{kv})^{w_{dkv}} \times \exp(-\pi_{s_d}\theta_{ck}\phi_{kv})}{w_{dkv}!}) \times \theta_{ck}^{\epsilon_0 - 1} \times \exp(-\epsilon_0 \theta_{ck}) \\
&\propto (\theta_{ck})^{\sum_{d:c_d=c} \sum_{v=1}^{V} w_{dkv} + \epsilon_0 - 1} \times \exp(-(\epsilon_0 + \sum_{d:c_d=c} \sum_{v=1}^{V} \pi_{s_d}\phi_{kv})\theta_{ck}) \\
&\sim \Gamma(\sum_{d:c_d=c} \sum_{v=1}^{V} w_{dkv} + \epsilon_0, \sum_{d:c_d=c} \sum_{v=1}^{V} \pi_{s_d}\phi_{kv} + \epsilon_0),
\end{aligned} \tag{3.3}
$$

where $(w_{d1v}, \ldots, w_{dKv}) \sim \text{Multinomial}(w_{dv}, \{\theta_{c_d k}\phi_{kv}\}_{k=1}^{K})$.

Document-specific factors

$$\begin{aligned}
\Pr(\pi_i|\,\boldsymbol{\pi}_{\setminus i},\boldsymbol{\phi},\boldsymbol{\theta},\boldsymbol{w},\boldsymbol{c}) &\propto \prod_{d:s_d=i}\Pr(\boldsymbol{w}_d|\,\pi_i,\boldsymbol{\pi}_{\setminus i},\boldsymbol{\phi},\boldsymbol{\theta},\boldsymbol{c})\times\Pr(\pi_i)\\
&\propto \Big(\prod_{d:s_d=i}\prod_{k=1}^{K}\prod_{v=1}^{V}\frac{(\pi_i\theta_{c_dk}\phi_{kv})^{w_{dkv}}\times\exp(-\pi_i\theta_{c_dk}\phi_{kv})}{w_{dkv}!}\Big)\times\pi_i^{\epsilon_0-1}\times\exp(-\epsilon_0\pi_i)\\
&\propto (\pi_i)^{\sum_{d:s_d=i}\sum_{k=1}^{K}\sum_{v=1}^{V}w_{dkv}+\epsilon_0-1}\times\exp(-(\epsilon_0+\sum_{d:s_d=i}\sum_{k=1}^{K}\sum_{v=1}^{V}\theta_{c_dk}\phi_{kv})\pi_i)\\
&\sim\Gamma(N_i+\epsilon_0,\sum_{d:s_d=i}\sum_{k=1}^{K}\sum_{v=1}^{V}\theta_{c_dk}\phi_{kv}+\epsilon_0),
\end{aligned}\tag{3.4}$$

where N_i is the number of words written by the sender i over the corpus—i.e., $N_i = \sum_{d:s_d=i}\sum_{k=1}^{K}\sum_{v=1}^{V}w_{dkv}$.

Candidate receivers

In the IPTM, direct computation of the posterior densities for the latent variables $\boldsymbol{b}$ and $\boldsymbol{\eta}$—i.e., $\Pr(\boldsymbol{b}|\boldsymbol{x},\boldsymbol{s},\boldsymbol{r},\boldsymbol{t},\boldsymbol{c})$ and $\Pr(\boldsymbol{\eta}|\boldsymbol{y},\boldsymbol{s},\boldsymbol{r},\boldsymbol{t},\boldsymbol{c})$—is not possible. However, it is possible to augment the data by candidate receivers $\boldsymbol{u}$ such that we can obtain their conditional posterior by collapsing the known distributions—$\Pr(\boldsymbol{b},\boldsymbol{u}|\boldsymbol{x},\boldsymbol{s},\boldsymbol{r},\boldsymbol{t},\boldsymbol{c})$ and $\Pr(\boldsymbol{\eta},\boldsymbol{u}|\boldsymbol{y},\boldsymbol{s},\boldsymbol{r},\boldsymbol{t},\boldsymbol{c})$—through integrating out $\boldsymbol{u}$. We adapt this common tool in Bayesian statistics called "data augmentation" (Tanner and Wong, 1987; Neal and Kypraios, 2015). Since u_{idj} is a binary random variable, it may be sampled from a Bernoulli distribution with probability $p_{idj}=\frac{\exp(\lambda_{idj})}{\exp(\lambda_{idj})+\mathrm{I}(\|\boldsymbol{u}_{id\setminus j}\|_1>0)}$, since

$$\begin{aligned}
\Pr(u_{idj}=1|\,\boldsymbol{u}_{id\setminus j},\boldsymbol{b},\boldsymbol{x},\boldsymbol{s},\boldsymbol{r},\boldsymbol{t},\boldsymbol{c}) &\propto \exp(\lambda_{idj})\\
\Pr(u_{idj}=0|\,\boldsymbol{u}_{id\setminus j},\boldsymbol{b},\boldsymbol{x},\boldsymbol{s},\boldsymbol{r},\boldsymbol{t},\boldsymbol{c}) &\propto \mathrm{I}(\|\boldsymbol{u}_{id\setminus j}\|_1>0),
\end{aligned}\tag{3.5}$$

where the subscript "$\setminus j$" denotes a quantity excluding data from position j and $\mathrm{I}(\cdot)$ is the indicator function that prevents empty receiver sets.

Coefficients for receiver selection features

Unlike the candidate receivers above, the conditional posterior for $\boldsymbol{b}=\{\boldsymbol{b}_c\}_{c=1}^{C}$ does not have a closed form; however each vector $\boldsymbol{b}_c$ may instead be re-sampled using the Metropolis–Hastings (MH) algorithm. Assuming an uninformative prior (i.e., $N(0,\infty)$), the conditional posterior for $\boldsymbol{b}_c$ is proportional to

$$\Pr(\boldsymbol{b}_c|\,\boldsymbol{u},\boldsymbol{x},\boldsymbol{s},\boldsymbol{r},\boldsymbol{t},\boldsymbol{c})\propto\prod_{d:c_d=c}\prod_{i=1}^{A}\frac{1}{Z(\boldsymbol{\lambda}_{id})}\exp\Big(\log\big(\mathrm{I}(\|\boldsymbol{u}_{id}\|_1>0)\big)+\sum_{j\neq i}\lambda_{idj}u_{idj}\Big).\tag{3.6}$$

Coefficients for event timing features

Likewise, we use the MH algorithm to update the latent variable $\boldsymbol{\eta} = \{\boldsymbol{\eta}_c\}_{c=1}^C$. Assuming an uninformative prior for each $\boldsymbol{\eta}_c$ (i.e., $N(0, \infty)$), the conditional posterior for an untied event case is proportional to

$$\Pr(\boldsymbol{\eta}_c | \boldsymbol{u}, \boldsymbol{y}, \boldsymbol{s}, \boldsymbol{r}, \boldsymbol{t}, \boldsymbol{c}) \propto \prod_{d:c_d=c} \Big(f_\tau(\tau_d; \mu_{s_d d}, V(\mu)) \times \prod_{i \neq s_d} \big(1 - F_\tau(\tau_d; \mu_{id}, V(\mu))\big)\Big), \quad (3.7)$$

where $f_\tau(\tau_d; \mu_{s_d d}, V(\mu))$ is the probability that the d^{th} observed time increment comes from the specified distribution $f_\tau(\cdot)$ with the observed sender's mean $\mu_{s_d d}$, and $\prod_{i \neq s_d} \big(1 - F_\tau(\tau_d; \mu_{id}, V(\mu))\big)$ is the probability that the rest of (unobserved) senders for document d all draw time increments greater than τ_d. Moreover, under the existence of tied events, the conditional posterior of $\boldsymbol{\eta}_c$ is written as proportional to

$$\begin{aligned}\Pr(\boldsymbol{\eta}_c | \boldsymbol{u}, \boldsymbol{y}, \boldsymbol{s}, \boldsymbol{r}, \boldsymbol{t}, \boldsymbol{c}) \propto \prod_{m=1}^{M} \Big(&\prod_{d:t_d=t_m^*} f_\tau(t_m^* - t_{m-1}^*; \mu_{s_d d}, V(\mu)) \\ &\times \prod_{i \notin \{s_d\}_{d:t_d=t_m^*}} \big(1 - F_\tau(t_m^* - t_{m-1}^*; \mu_{id}, V(\mu))\big)\Big),\end{aligned} \quad (3.8)$$

where $t_1^*, \ldots, t_M^*$ are the unique timepoints across D documents ($M \leq E$). If $M = D$ (i.e., no tied events), equation (3.8) reduces to equation (3.7). Note that when we have the latent variable to quantify the variance in time increments $V(\mu)$ (based on the choice of timestamp distribution in Section 2.2), we also use equation (3.7) (or equation (3.8) in case there exist tied events) for the additional MH update—e.g., $\Pr(k | \boldsymbol{\eta}, \boldsymbol{u}, \boldsymbol{y}, \boldsymbol{s}, \boldsymbol{r}, \boldsymbol{t}, \boldsymbol{c})$ for Weibull, $\Pr(\theta | \boldsymbol{\eta}, \boldsymbol{u}, \boldsymbol{y}, \boldsymbol{s}, \boldsymbol{r}, \boldsymbol{t}, \boldsymbol{c})$ for gamma, and $\Pr(\sigma_\tau^2 | \boldsymbol{\eta}, \boldsymbol{u}, \boldsymbol{y}, \boldsymbol{s}, \boldsymbol{r}, \boldsymbol{t}, \boldsymbol{c})$ for log-normal.

Interaction pattern assignments

Conditional posterior of the interaction pattern assignments $\boldsymbol{c} = \{c_d\}_{d=1}^D$ is quite complicated since it is the only component that connects the network dynamics and contents. Despite its complexity, we can still perform efficient inference by sampling from categorical distribution.

$$\begin{aligned}&\Pr(c_d = c | \boldsymbol{c}_{\backslash d}, \boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\pi}, \boldsymbol{w}, \boldsymbol{u}, \boldsymbol{b}, \boldsymbol{x}, \boldsymbol{\eta}, \boldsymbol{y}, \boldsymbol{s}, \boldsymbol{r}, \boldsymbol{t}) \\ &\propto \Pr(\psi_c | \boldsymbol{\psi}_{\backslash c}, \boldsymbol{c}) \times \prod_{k=1}^{K} \prod_{v=1}^{V} \Pr(\phi_{kv} | \boldsymbol{\phi}_{\backslash k,v}, \boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{\pi}, \boldsymbol{c}) \times \prod_{c=1}^{C} \prod_{k=1}^{K} \Pr(\theta_{ck} | \boldsymbol{\theta}_{\backslash c,k}, \boldsymbol{\phi}, \boldsymbol{w}, \boldsymbol{\pi}, \boldsymbol{c}) \\ &\times \prod_{i=1}^{A} \Pr(\pi_i | \boldsymbol{\pi}_{\backslash i}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{c}) \times \prod_{i=1}^{A} \Pr(\boldsymbol{u}_{id} | \boldsymbol{b}, \boldsymbol{x}, \boldsymbol{c}) \times \Pr(\boldsymbol{b}_c | \boldsymbol{u}, \boldsymbol{x}, \boldsymbol{s}, \boldsymbol{r}, \boldsymbol{t}, \boldsymbol{c}) \times \Pr(\boldsymbol{\eta}_c | \boldsymbol{u}, \boldsymbol{y}, \boldsymbol{s}, \boldsymbol{r}, \boldsymbol{t}, \boldsymbol{c}),\end{aligned} \quad (3.9)$$

Algorithm 2 MCMC algorithm

Input: number of outer and inner iterations (O, I_1, I_2) and initial values of $(\boldsymbol{\psi}, \boldsymbol{u}, \boldsymbol{b}, \boldsymbol{\eta})$

```
for o = 1 to O do
  for d = 1 to D do
    for i = 1 to A do
      for j = 1 to A (j ≠ i) do
        update u_idj using Gibbs update —equation (3.5)
      end for
    end for
  end for
  for n = 1 to I_1 do
    update b using MH algorithm—equation (3.6)
  end for
  for n = 1 to I_2 do
    update η using MH algorithm—equation (3.7) or (3.8)
  end for
  if extra parameter for V(μ) then
    update the variance parameter using MH algorithm—equation (3.7) or (3.8)
  end if
end for
summarize the results with the last chain of b and η
```

which is the product of the entire equations we used in Section 3. We could further simplify this as below:

$$\begin{aligned}
&\Pr(c_d = c \,|\, \boldsymbol{c}_{\backslash d}, \boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\pi}, \boldsymbol{w}, \boldsymbol{u}, \boldsymbol{b}, \boldsymbol{x}, \boldsymbol{\eta}, \boldsymbol{y}, \boldsymbol{s}, \boldsymbol{r}, \boldsymbol{t}) \\
&\propto \psi_c \times \prod_{k=1}^{K}\prod_{v=1}^{V} \frac{(\pi_{s_d}\theta_{ck}\phi_{kv})^{w_{dkv}} \times \exp(-\pi_{s_d}\theta_{ck}\phi_{kv})}{w_{dkv}!} \\
&\times \prod_{i=1}^{A} \frac{1}{Z(\boldsymbol{\lambda}_{id})} \exp\Big(\log\big(\mathrm{I}(\|\boldsymbol{u}_{id}\|_1 > 0)\big) + \sum_{j\neq i} \lambda_{idj} u_{idj}\Big) \text{ using } \boldsymbol{b}_c \text{ and } \boldsymbol{x}_c \\
&\times \Big(f_\tau(\tau_d; \mu_{s_d d}, V(\mu)) \times \prod_{i\neq s_d} \big(1 - F_\tau(\tau_d; \mu_{id}, V(\mu))\big)\Big) \text{ using } \boldsymbol{\eta}_c \text{ and } \boldsymbol{y}_c \\
&\propto \psi_c \times \prod_{k=1}^{K}\prod_{v=1}^{V} (\theta_{ck})^{w_{dkv}} \times \exp(-\pi_d \sum_{k=1}^{K}\sum_{v=1}^{V} \theta_{ck}\phi_{kv}) \\
&\times \prod_{i=1}^{A} \frac{1}{Z(\boldsymbol{\lambda}_{id})} \exp\Big(\log\big(\mathrm{I}(\|\boldsymbol{u}_{id}\|_1 > 0)\big) + \sum_{j\neq i} \lambda_{idj} u_{idj}\Big) \text{ using } \boldsymbol{b}_c \text{ and } \boldsymbol{x}_c \\
&\times \Big(f_\tau(\tau_d; \mu_{s_d d}, V(\mu)) \times \prod_{i\neq s_d} \big(1 - F_\tau(\tau_d; \mu_{id}, V(\mu))\big)\Big) \text{ using } \boldsymbol{\eta}_c \text{ and } \boldsymbol{y}_c.
\end{aligned} \tag{3.10}$$

References

Alemán, E. and Calvo, E. (2013). "Explaining policy ties in presidential congresses: A network analysis of bill initiation data." *Political Studies*, 61(2): 356–377. 2

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). "Latent Dirichlet Allocation." *J. Mach. Learn. Res.*, 3: 993–1022. 1, 5

Bratton, K. A. and Rouse, S. M. (2011). "Networks in the legislative arena: How group dynamics affect cosponsorship." *Legislative Studies Quarterly*, 36(3): 423–460. 2

Burgess, A., Jackson, T., and Edwards, J. (2004). "Email overload: Tolerance levels of employees within the workplace." In *Innovations Through Information Technology: 2004 Information Resources Management Association International Conference, New Orleans, Louisiana, USA, May 23-26, 2004*, volume 1, 205. IGI Global. 1

Butts, C. T. (2008). "A RELATIONAL EVENT FRAMEWORK FOR SOCIAL ACTION." *Sociological Methodology*, 38(1): 155–200. 2

Camber Warren, T. (2010). "The geometry of security: Modeling interstate alliances as evolving networks." *Journal of Peace Research*, 47(6): 697–709. 2

Chatterjee, S., Diaconis, P., et al. (2013). "Estimating and understanding exponential random graph models." *The Annals of Statistics*, 41(5): 2428–2461. 1

Cranmer, S. J., Desmarais, B. A., and Kirkland, J. H. (2012a). "Toward a network theory of alliance formation." *International Interactions*, 38(3): 295–324. 2

Cranmer, S. J., Desmarais, B. A., and Menninga, E. J. (2012b). "Complex dependencies in the alliance network." *Conflict Management and Peace Science*, 29(3): 279–313. 2

Desmarais, B. A. and Cranmer, S. J. (2017). "Statistical Inference in Political Networks Research." In Victor, J. N., Montgomery, A. H., and Lubell, M. (eds.), *The Oxford Handbook of Political Networks*. Oxford University Press. 2

Fahmy, C. and Young, J. T. (2016). "Gender Inequality and Knowledge Production in Criminology and Criminal Justice." *Journal of Criminal Justice Education*, 1–21. 2

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., and Morris, M. (2008). "ergm: A package to fit, simulate and diagnose exponential-family models for networks." *Journal of statistical software*, 24(3): nihpa54860. 1

Kanungo, S. and Jain, V. (2008). "Modeling email use: a case of email system transition." *System Dynamics Review*, 24(3): 299–319. 1

Kinne, B. J. (2016). "Agreeing to arm: Bilateral weapons agreements and the global arms trade." *Journal of Peace Research*, 53(3): 359–377. 2

Krafft, P., Moore, J., Desmarais, B., and Wallach, H. M. (2012). "Topic-Partitioned Multinetwork Embeddings." In Pereira, F., Burges, C., Bottou, L., and Weinberger, K. (eds.), *Advances in Neural Information Processing Systems 25*, 2807–2815. Curran Associates, Inc. 1

Kronegger, L., Mali, F., Ferligoj, A., and Doreian, P. (2011). "Collaboration structures in Slovenian scientific communities." *Scientometrics*, 90(2): 631–647. 2

Lai, C.-H., She, B., and Tao, C.-C. (2017). "Connecting the dots: A longitudinal observation of relief organizations' representational networks on social media." *Computers in Human Behavior*, 74: 224–234. 2

Liang, X. (2015). "The Changing Impact of Geographic Distance: A Preliminary Analysis on the Co-author Networks in Scientometrics (1983-2013)." In *System Sciences (HICSS), 2015 48th Hawaii International Conference on*, 722–731. IEEE. 2

Lim, K. W., Chen, C., and Buntine, W. (2013). "Twitter-Network Topic Model: A Full Bayesian Treatment for Social Network and Text Modeling." In *NIPS2013 Topic Model workshop*, 1–5. 1

McCallum, A., Corrada-Emmanuel, A., and Wang, X. (2005). "The Author-Recipient-Topic Model for Topic and Role Discovery in Social Networks, with Application to Enron and Academic Email." In *Workshop on Link Analysis, Counterterrorism and Security*, 33. 1

Neal, P. and Kypraios, T. (2015). "Exact Bayesian inference via data augmentation." *Statistics and Computing*, 25(2): 333–347. 8

Nelder, J. A. and Baker, R. J. (1972). *Generalized linear models*. Wiley Online Library. 4

Peng, T.-Q., Liu, M., Wu, Y., and Liu, S. (2016). "Follower-followee network, communication networks, and vote agreement of the US members of congress." *Communication Research*, 43(7): 996–1024. 2

Perry, P. O. and Wolfe, P. J. (2013). "Point process modelling for directed interaction networks." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(5): 821–849. 2

Pew, R. C. (2016). "Social Media Fact Sheet." *Accessed on 03/07/17*. 1

Robins, G., Pattison, P., Kalish, Y., and Lusher, D. (2007). "An introduction to exponential random graph (p*) models for social networks." *Social networks*, 29(2): 173–191. 1

Schein, A., Zhou, M., Blei, D. M., and Wallach, H. (2016). "Bayesian poisson tucker decomposition for learning the structure of international relations." *arXiv preprint arXiv:1606.01855*. 5

Snijders, T. A. (1996). "Stochastic actor-oriented models for network change." *Journal of mathematical sociology*, 21(1-2): 149–172. 2, 5

Szóstek, A. M. (2011). "?Dealing with My Emails?: Latent user needs in email management." *Computers in Human Behavior*, 27(2): 723–729. 1

Tanner, M. A. and Wong, W. H. (1987). "The calculation of posterior distributions by data augmentation." *Journal of the American statistical Association*, 82(398):

528–540. 8

Yoon, H. Y. and Park, H. W. (2014). "Strategies affecting Twitter-based networking pattern of South Korean politicians: social network analysis and exponential random graph model." *Quality & Quantity*, 1–15. 2