Network Modeling of Dynamic Textual Communication with Application to Government Email Corpora

Bomin Kim[1], Zachary Jones[1], Bruce Desmarais[1], and Hanna Wallach[2,3]

[1]Pennsylvania State University
[2]Microsoft Research NYC
[3]University of Massachusetts Amherst

March 11, 2017

Abstract

In this paper, we introduce the interaction-partitioned topic model (IPTM)—a probabilistic model of who communicates with whom about what, and when. Broadly speaking, the IPTM partitions the content in time-stamped textual communications, such as emails, according to both the network dynamics that they reflect and their content. To define the IPTM, we integrate a dynamic version of the exponential random graph model—a generative model for ties that tend toward structural features such as triangles—and latent Dirichlet allocation—a generative model for topic-based content. The IPTM assigns each topic to an "interaction pattern," characterized by a set of dynamic network generative processes. We use the IPTM to analyze emails sent between department managers in two county governments in North Carolina; one of these email corpora covers the Outer Banks during the time period surrounding Hurricane Sandy. Via this application, we demonstrate that the IPTM is effective at predicting and explaining continuous-time textual communications.

1 Introduction

In recent decades, real-time digitized textual communication has developed into a ubiquitous form of social and professional interaction (see, e.g., 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 networks consisting of text-valued edges that arise in continuous time (e.g., e-mail messages). A number of models that build upon topic modeling through Latent Dirichlet Allocation (Blei et al., 2003) to incorporate link data as well as textual content have been developed in response to this need [**ASK HANNA FOR CITES**]. The exponential random graph model (ERGM) (Robins et al., 2007; Chatterjee et al., 2013; Hunter et al., 2008), which is the canonical model for network structure as it is flexible enough to account for nearly any pattern of tie formation (Desmarais and Cranmer, 2017), has yet to be adapted to the setting of time-stamped text-valued edges. We build upon recent extensions of ERGM that model time-stamped ties (Perry and Wolfe, 2013; Butts, 2008), and develop the interaction-partitioned topic model (IPTM) to simultaneously model the network structural patterns that govern tie formation, and the content in the communications.

2 Generative Process

Assume we have a collection of documents, consisting of D number of unique documents. A single email, indexed by $d \in \{1, ..., D\}$, is represented by the four components $(i^{(d)}, J^{(d)}, t^{(d)}, \boldsymbol{w}^{(d)})$. The first two are the sender and receiver of the email: an integer $i^{(d)} \in \{1, ..., A\}$ indicates the identity of the sender out of A number of actors (or nodes) and an integer vector $J^{(d)} = \{j_r^{(d)}\}_{r=1}^{|J^{(d)}|}$ indicates the identity of the receiver (or receivers) out of $A-1$ number of actors (by excluding self-loop), where $|J^{(d)}| \in \{1, ..., A-1\}$ denotes the total number of the receivers. Next, $t^{(d)}$ is the (unix time-based) timestamp of the email d, and $\boldsymbol{w}^{(d)} = \{w_m^{(d)}\}_{n=1}^{N^{(d)}}$ is a set of tokens that comprise the text of the email. In this section, we illustrate how the words $\boldsymbol{w}^{(d)}$ are generated according to the variant of latent Dirichlet allocation (Blei et al., 2003), and then how the rest three components, $(i^{(d)}, J^{(d)}, t^{(d)})$, are simulatneously generated from the stochastic intensity and topic assignments of the document.

2.1 Content Generating Process

The content generating process is a simple addition of the interaction pattern assignment to the existing generative process of Latent Dirichlet Allocation Blei et al. (2003). This concept is also very similar to the content-partitioned multinetwork embeddings (CPME) model [**ASK BRUCE FOR CITES**], in a way that each topic k is also associated with a cluster assignment C_k, where C_k can take one of $c = \{1, ..., C\}$ values.

First we generate the global (corpus-wide) variables. There are two main sets of global variables: those that describe the topics people talk about and those that describe how people interact (interaction patterns). These variables are linked by a third set of variables that associate each topic with the pattern that best describes how people interact when talking about that topic.

There are K topics. Each topic k is a discrete distribution over V word types.

1. $\boldsymbol{\phi}^{(k)} \sim \text{Dir}(\beta, \mathbf{u})$ [**See Algorithm 1**]
   - A "topic" k is characterized by a discrete distribution over V word types with probability vector $\phi^{(k)}$. A symmetric Dirichlet prior $\mathbf{u}$ with the concentration parameter β is placed.

There are C interaction patterns. Each interaction pattern consists of a vector of coefficients $\boldsymbol{\gamma}^{(c)}$ in $\boldsymbol{R}^P$ and a vector of P-dimensional dynamic network statistics for directed edge (i, j) at time t $\boldsymbol{x}_t^{(c)}(i, j)$; however, we assume that our generative process is conditioned on these covariates.

2. $\boldsymbol{\beta}^{(c)} \sim \text{Normal}(\mathbf{0}, \sigma^2 I_P)$ [**See Algorithm 2**]:
   - The vector of coefficients depends on the interaction pattern c. This means that there is variation in the degree of influence from the dynamic network statistics.

The topics and interaction patterns are tied together via a set of K categorical variables.

3. $C_k \sim \text{Unif}(1, C)$ [**See Algorithm 3**]:
   - Each topic is associated with a single interaction pattern.

Then, we generate the local variables. We assume the following generative process for each document d in a corpus D [**See Algorithm 4**]:

4-1. Choose the number of words $N^{(d)} \sim \text{Poisson}(\zeta)$

4-2. Choose document-topic distribution $\boldsymbol{\theta}^{(d)} \sim \text{Dir}(\alpha, \boldsymbol{m})$

4-3. For each of the $N^{(d)}$ words $w_n^{(d)}$:

   (a) Choose a topic $z_n^{(d)} \sim \text{Multinomial}(\boldsymbol{\theta}^{(d)})$

   (b) Choose a word $w_n^{(d)} \sim \text{Multinomial}(\phi^{(z_n^{(d)})})$

2.2 Stochastic Intensity

Here we illustrate how a set of dynamic network features and topic-interaction assignments jointly identify the stochastic intensity of a document, which plays a key role in tie generating process in 2.3. Assume that each document $d \in \{1, ..., D\}$ has an $A \times A$ stochastic intensity (or hazard) matrix of $\boldsymbol{\lambda}^{(d)}(t) = \{\{\lambda_{ij}^{(d)}(t)\}_{i=1}^{A}\}_{j=1}^{A}$, where $\lambda_{ij}^{(d)}(t)$=P{for document d, $i \rightarrow j$ occurs in time interval $[t, t+dt)$, given that it has not ben sent until time t}.

To calculate the distribution of interaction patterns within a document, we estimate the (empirical) proportion of the words in document d which are assigned the topics corresponding to the interaction pattern c from 2.1:

$$p_c^{(d)} = \frac{\sum_{n=1}^{N^{(d)}} I(C_{z_n^{(d)}} = c)}{N^{(d)}}, \tag{1}$$

where I is the indicator function (1/0) and $\sum\limits_{c=1}^{C} p_c^{(d)} = 1$.

Now, we define the $(i, j)^{th}$ element of the stochastic intensity matrix $\boldsymbol{\lambda}^{(d)}(t)$ forms:

$$\lambda_{ij}^{(d)}(t) = \lambda_0 \sum_{c=1}^{C} p_c^{(d)} \cdot \exp\Big\{\boldsymbol{\beta}^{(c)T} \boldsymbol{x}_t^{(c)}(i, j)\Big\} \cdot 1\{j \in \mathcal{A}_{\backslash i}\}, \tag{2}$$

where λ_0 is the common baseline hazards for the overall interaction (assume that λ_0 does not depend on t), $p_c^{(d)}$ is as defined above, $\boldsymbol{\beta}^{(c)}$ is an unknown vector of coefficients in $\boldsymbol{R}^p$ corresponding to the interaction pattern c, $\boldsymbol{x}_t^{(c)}(i, j)$ is a vector of the p-dimensional dynamic network statistics for directed edge (i, j) at time t corresponding to the interaction pattern c, and $\mathcal{A}_{\backslash i}$ is the predictable receiver set of sender i within the set of all possible actors $\mathcal{A}$ (no self-loop). After we include intercept into $\boldsymbol{x}_t^{(c)}(i, j)$, we rewrite (2) as:

$$\lambda_{ij}^{(d)}(t) = \sum_{c=1}^{C} p_c^{(d)} \cdot \exp\Big\{\boldsymbol{\beta}^{(c)T} \boldsymbol{x}_t^{*(c)}(i, j)\Big\} \cdot 1\{j \in \mathcal{A}_{\backslash i}\}, \tag{3}$$

where the first element of $\boldsymbol{\beta}^{(c)}$ corresponds to the baseline intensity of interaction pattern c by including the intercept term and setting $\boldsymbol{x}_t^{*(c)}(i, j) = (\mathbf{1}, \boldsymbol{x}_t^{(c)}(i, j))$. This can be seen as the weighted average of stochastic intensities across the interaction patterns. Next, since multicast interactions—those involving a single sender but multiple receivers—are allowed for this model, we expand the rate of interaction between sender i and the receivers in a set J as:

$$\lambda_{iJ}^{(d)}(t) = \sum_{c=1}^{C} p_c^{(d)} \cdot \exp\Big\{\sum_{j \in J} \boldsymbol{\beta}^{(c)T} \boldsymbol{x}_t^{*(c)}(i, j)\Big\} \cdot \prod_{j \in J} 1\{j \in \mathcal{A}_{\backslash i}\}. \tag{4}$$

In case of single receivers ($|J| = 1$), Equation (4) is reduced to Equation (3), thus in the following sections we use the Equation (4) of multicast cases as a general form of the stochastic intensity between the sender and receivers.

2.3 Tie Generating Process

We assume the following generative process for each document d in a corpus D:

1. (Data augmentation) For each sender $i \in \{1, ..., A\}$, create a list of receivers J_i by applying the Bernoulli probabilities to every $j \in \mathcal{A}_{\backslash i}$

$$I(i \rightarrow j) \sim \text{Ber}\Big(1 - \exp(-\delta \lambda_{ij}^{(d)}(t_+^{(d-1)}))\Big), \tag{5}$$

where the probability is called a Bernoulli-Poisson (BerPo) link function (Zhou, 2015) and δ is a tuning parameter to control the number of recipients and the urgency of the document to be sent. Note that $_+$ denotes including the timepoint itself, meaning that λ_{ij} is obtained using the history of interactions until and including the timestamp $t^{(d-1)}$. (i.e. $\lambda_{ij}^{(d)}(t_+^{(d-1)}) = \sum_{c=1}^{C} p_c^{(d)} \cdot \exp\Big\{\boldsymbol{\beta}^{(c)T} \boldsymbol{x}^{*(c)}_{t_+^{(d-1)}}(i,j)\Big\} \cdot 1\{j \in \mathcal{A}_{\backslash i}\}$)

2. For each sender $i = 1, ..., A$, generate the time increments $\Delta T_{iJ_i} \sim \text{Exp}(\lambda_{iJ_i}^{(d)}(t_+^{(d-1)}))$, where $\lambda_{iJ_i}^{(d)}(t_+^{(d-1)}) = \sum_{c=1}^{C} p_c^{(d)} \cdot \exp\Big\{\sum_{j \in J_i} \boldsymbol{\beta}^{(c)T} \boldsymbol{x}^{*(c)}_{t_+^{(d-1)}}(i,j)\Big\} \cdot \prod_{j \in J_i} 1\{j \in \mathcal{A}_{\backslash i}\}$
3. Set the timestamp $t^{(d)} = t^{(d-1)} + \min(\Delta T_{iJ_i})$, $i^{(d)} = i_{\min(\Delta T_{iJ_i})}$, and $J^{(d)} = J_{i^{(d)}}$.

To simplify the notation, we use $(i, J, t)^{(d)} \sim \text{GP}(\boldsymbol{\lambda}^{(d)}(t_+^{(d-1)}))$ to refer to the joint generating process of $(i^{(d)}, J^{(d)}, t^{(d)})$ above.

2.4 Joint Generative Process of Document

Below are the joint generative process for each document in a corpus D and the corresponding plate notation (Figure 1).

Algorithm 1 Topic Word Distributions

```
for k=1 to K do
    draw φ^(k) ~ Dir(β, u)
end
```

Algorithm 2 Interaction Pattern Parameters

```
for c=1 to C do
    draw β^(c) ~ Normal(0, σ^2 I_P)
end
```

Algorithm 3 Topic Interaction Pattern Assginments

```
for k=1 to K do
    draw C_k ~ Unif(1, C)
end
```

Algorithm 4 Document Generating Process

```
for d=1 to D do
    draw N^(d) ~ Poisson(ζ)
    draw θ^(d) ~ Dir(α, m)
    for n=1 to N^(d) do
        draw z_n^(d) ~ Multinom(θ^(d))
        draw w_n^(d) ~ Multinom(φ^(z_n^(d)))
    end
    for c=1 to C do
        set p_c^(d) = (Σ_{n=1}^{N^(d)} I(C_{z_n^(d)} = c)) / N^(d)
    end
    draw (i, J, t)^(d) ~ GP(λ^(d)(t_+^(d-1)))
end
```

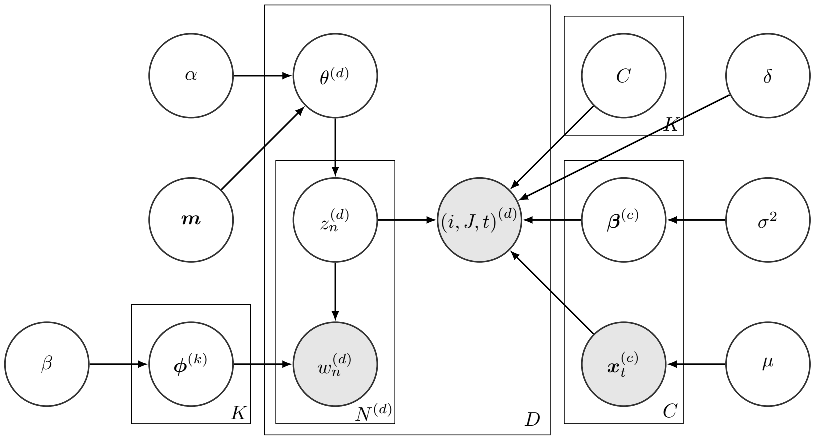

Figure 1: Plate notation of IPTM

3 Dynamic covariates to measure network effects

The network statistics $\boldsymbol{x}_t^{*(c)}(i, j)$ of Equation (3), corresponding to the ordered pair (i, j), can be time-invariant (such as gender) or time-dependent (such as the number of two-paths from i to j just before time t). There could be various static and dynamic covariates $\boldsymbol{x}_t^{(c)}(i, j)$ that affects the stochastic intensity, however, we decide to use the covariates that depend on the history of the process, considering the strong recency and reciprocity effects of textual communications, especially emails.

send	$i \longrightarrow j$	i has sent j a message in the past
receive	$i \longleftarrow j$	i has received a message from j in the past
2-send	$i \longrightarrow h \longrightarrow j$	there exists an actor h such that i has sent h a message and h has sent j a message in the past
2-receive	$i \longleftarrow h \longleftarrow j$	there exists an actor h such that i has received a message from h, and h has received a message from j
sibling	$i \longleftarrow h \longrightarrow j$	there exists an actor h such that h has sent i and j messages in the past
cosibling	$i \longrightarrow h \longleftarrow j$	there exists an actor h such that h has received messages from i and j

Fig. 3. Dynamic covariates to measure network effects

Following Perry and Wolfe (2013) (refer to Figure above), we use 4 effects as components of $\boldsymbol{x}_t^{*(c)}(i, j)$, including the intercept to estimate the baseline intensities. The two behaviors (send and receive) are dyadic, involving exactly two actors, while the one is triadic (sum of 2-send, 2-receive, sibling, and cosibling), involving three actors. In addition to those, we further include the indegree for receiver and outdegree for sender effects to measure the popularity and centrality. However, one novel point in this paper is that we define the effects not to be based on finite sub-interval, which require large number of dimension. Instead, we create a time-weighting function so that we can use a single statistic for each effect by incorporating the recency of event into the statistic itself. The time-weighting function estimated at time t is specified as:

$$\mathrm{TW}_t(\{t_n^{(a\to b)}\}_{n=1}^N, \mu) = (1 + (\prod_{n=1}^N \Delta t_n)^{\frac{1}{N}})^{\mu}, \tag{6}$$

where the first input $\{t_n^{(a\to b)}\}_{n=1}^N$ denotes the timestamps of the documents where the sender equals to a and the receiver set contains b before time t (of total length $N = \sum_{d:t^{(d)}<t} I\{a \to b\}$), and the second input μ is the parameter of decay, measuring how fast the past interactions lose their influence on the future interactions. Then, the output is a function of the geometric mean of the time differences $\Delta t_n = t - t_n^{(a\to b)}$, for $n = 1, .., N$. Note that we add 1 before taking μ^{th} power, in order to ensure that the base is always greater than 1.

As a result, all of the statistics below can be seen as time-weighted dynamic network statistics.

1. $\text{intercept}_t^{(c)}(i,j) = 1$

2. $\text{send}_t^{(c)}(i,j) = \frac{\sum\limits_{d:t^{(d)}<t} p_c^{(d)} \cdot I\{i \rightarrow j\}}{\text{TW}_t(\{t_n^{(i\rightarrow j)}\}_{n=1}^N, \mu)}$, where $N = \sum\limits_{d:t^{(d)}<t} I\{i \rightarrow j\}$

3. $\text{receive}_t^{(c)}(i,j) = \frac{\sum\limits_{d:t^{(d)}<t} p_c^{(d)} \cdot I\{i \rightarrow j\}}{\text{TW}_t(\{t_n^{(j\rightarrow i)}\}_{n=1}^N, \mu)}$, where $N = \sum\limits_{d:t^{(d)}<t} I\{j \rightarrow i\}$

4. $\text{triangle}_t^{(c)}(i,j) = \frac{\sum\limits_{d:t^{(d)}<t} \sum\limits_{h\neq i,j} \Big(p_c^{(d)} \cdot I\{i\rightarrow h \text{ or } h\rightarrow i\}\Big)\Big(p_c^{(d)} \cdot I\{j\rightarrow h \text{ or } h\rightarrow j\}\Big)}{\text{TW}_t(\{t_n^{(i\rightarrow h \text{ or } h\rightarrow i \text{ or } j\rightarrow h \text{ or } j\rightarrow h)}\}_{n=1}^N, \mu)}$,
   where $N = \sum\limits_{d:t^{(d)}<t} I\{i \rightarrow h \text{ or } h \rightarrow i \text{ or } j \rightarrow h \text{ or } j \rightarrow h\}$

5. $\text{outdegree}_t^{(c)}(i) = \frac{\sum\limits_{d:t^{(d)}<t} p_c^{(d)} \cdot I\{i \rightarrow \forall j\}}{\text{TW}_t(\{t_n^{(i\rightarrow \forall j)}\}_{n=1}^N, \mu)}$, where $N = \sum\limits_{d:t^{(d)}<t} I\{i \rightarrow \forall j\}$

6. $\text{indegree}_t^{(c)}(j) = \frac{\sum\limits_{d:t^{(d)}<t} p_c^{(d)} \cdot I\{\forall i \rightarrow j\}}{\text{TW}_t(\{t_n^{(\forall i\rightarrow j)}\}_{n=1}^N, \mu)}$, where $N = \sum\limits_{d:t^{(d)}<t} I\{\forall i \rightarrow j\}$

4 Inference

In this case, what we actually observe are the tokens $\mathcal{W} = \{\boldsymbol{w}^{(d)}\}_{d=1}^D$ and the sender, recipient, and timestamps $(i = i^{(d)}, j = J^{(d)}, t = t^{(d)})$ of the document d, which we jointly denote as $\mathcal{N} = \{\boldsymbol{N}^{(d)}\}_{d=1}^D = \{(i,J,t)^{(d)}\}_{d=1}^D$ in the inference, meaning the network features. Next, $\mathcal{X} = \{\{\boldsymbol{x}_{t^{(d)}}^{(c)}(i,j)\}_{c=1}^C\}_{d=1}^D$ is the metadata, and the latent variables are $\Phi = \{\boldsymbol{\phi}^{(k)}\}_{k=1}^K, \Theta = \{\boldsymbol{\theta}^{(d)}\}_{d=1}^D, \mathcal{Z} = \{\boldsymbol{z}^{(d)}\}_{d=1}^D, \mathcal{C} = \{C_k\}_{k=1}^K$, and $\mathcal{B} = \{\boldsymbol{\beta}^{(c)}\}_{c=1}^C$.

Below is the the big joint distribution

$$\begin{aligned}
&P(\Phi, \Theta, \mathcal{W}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \mathcal{N} | \mathcal{X}, \beta, \boldsymbol{u}, \alpha, \boldsymbol{m}, \sigma^2, \mu, \delta) \\
&= P(\mathcal{W}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \mathcal{N} | \Phi, \Theta, \mathcal{X}, \sigma^2, \mu, \delta) P(\Phi, \Theta | \beta, \boldsymbol{u}, \alpha, \boldsymbol{m}) \\
&= P(\mathcal{W}|\mathcal{Z}, \Phi) P(\mathcal{Z}|\Theta) P(\mathcal{N}|\mathcal{Z}, \mathcal{C}, \mathcal{X}, \mathcal{B}, \mu, \delta) P(\mathcal{B}|\mathcal{C}, \sigma^2) P(\Phi|\beta, \boldsymbol{u}) P(\Theta|\alpha, \boldsymbol{m}) P(\mathcal{C})
\end{aligned} \tag{7}$$

Now we can integrate out Φ and Θ in latent Dirichlet allocation by applying Dirichlet-multinomial conjugacy. See APPENDIX B for the detailed steps. After integration, our inference goal is to draw samples from the posterior distribution:

$$\propto P(\mathcal{W}|\mathcal{Z}) P(\mathcal{Z}|\beta, \boldsymbol{u}, \alpha, \boldsymbol{m}) P(\mathcal{N}|\mathcal{Z}, \mathcal{C}, \mathcal{X}, \mathcal{B}, \mu, \delta) P(\mathcal{B}|\mathcal{C}, \sigma^2) P(\mathcal{C}|\boldsymbol{\gamma}) \tag{8}$$

In practice, we can achieve this inference goal by sequentially resampling the value of each latent variable $\mathcal{Z}, \mathcal{C}$, and $\mathcal{B}$ from its conditional posterior distribution, using the equation below:

$$P(\mathcal{Z}, \mathcal{C}, \mathcal{B} | \mathcal{W}, \mathcal{N}, \mathcal{X}, \beta, \boldsymbol{u}, \alpha, \boldsymbol{m}, \sigma^2, \mu, \delta) \propto P(\mathcal{W}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \mathcal{N} | \mathcal{X}, \beta, \boldsymbol{u}, \alpha, \boldsymbol{m}, \sigma^2, \mu, \delta) \tag{9}$$

Either Gibbs sampling or Metropolis-Hastings algorithm is applied by sequentially resampling each latent variables from their respective conditional posterior.

4.1 Data augmentation

As mentioned earlier in Section 2.3, we use data augmentation in the tie generating process. We should consider those auxilary variables (or generated latent data) when it comes to make inferences on the related latent variables. Here are the derivation steps for the conditional probability of the sender, recipient, and timestamps $\mathcal{N} = \{\boldsymbol{N}^{(d)}\}_{d=1}^D = \{(i,J,t)^{(d)}\}_{d=1}^D$.

In latent data generating step, the sets (i, J_i) should account for the probability $P(\text{latent receiver}|\text{latent time} < \text{observed time})$. Let $\Delta T_{i^{(d)}J^{(d)}}$ be the observed time increment, ΔT_{iJ_i} be the latent time increment associated with the latent email sent by i, and r_{ij} be an indicator of whether sender i added j to the latent email receivers (i.e. $j \in J_i$), and $J_i^{(-j)}$ denote the vector of recipient indicators other than j.

$$
\begin{aligned}
&P\Big(r_{ij}=1|(\Delta T_{iJ_i}>\Delta T_{i^{(d)}J^{(d)}})\cap(J_i^{(-j)})\Big)\\
&=\frac{P\Big((r_{ij}=1)\cap(\Delta T_{iJ_i}>\Delta T_{i^{(d)}J^{(d)}})\cap(J_i^{(-j)})\Big)}{P\Big((\Delta T_{iJ_i}>\Delta T_{i^{(d)}J^{(d)}})\cap(J_i^{(-j)})\Big)}\\
&=\frac{P\Big(\Delta T_{iJ_i}>\Delta T_{i^{(d)}J^{(d)}}|(r_{ij}=1)\cap(J_i^{(-j)})\Big)P\Big((r_{ij}=1)\cap(J_i^{(-j)})\Big)}{P\Big((\Delta T_{iJ_i}>\Delta T_{i^{(d)}J^{(d)}})\cap(J_i^{(-j)})|r_{ij}=1\Big)P\Big(r_{ij}=1\Big)+P\Big((\Delta T_{iJ_i}>\Delta T_{i^{(d)}J^{(d)}})\cap(J_i^{(-j)})|r_{ij}=0\Big)P\Big(r_{ij}=0\Big)},
\end{aligned}
\tag{10}
$$

where $P\Big(\Delta T_{iJ_i}>\Delta T_{i^{(d)}J^{(d)}}|(r_{ij}=1)\cap(J_i^{(-j)})\Big)$, $P\Big((\Delta T_{iJ_i}>\Delta T_{i^{(d)}J^{(d)}})\cap(J_i^{(-j)})|r_{ij}=1\Big)$, and $P\Big((\Delta T_{iJ_i}>\Delta T_{i^{(d)}J^{(d)}})\cap(J_i^{(-j)})|r_{ij}=0\Big)$ are given by the cumulative exponential distributions associated with the respective $\lambda_{iJ_i}(t_+^{(d-1)})$ and $P\Big((r_{ij}=1)\cap(J_i^{(-j)})\Big)$, $P\Big(r_{ij}=1\Big)$, and $P\Big(r_{ij}=0\Big)$ are given by the Bernoulli equation in Section 2.3. We plug in the equations and update the latent ties r_{ij} with the probability:

$$
\begin{aligned}
&=\frac{e^{-(t^{(d)}-t^{(d-1)})\sum\limits_{j\in J_i^{(+j)}}\lambda_{ij}^{(d)}(t_+^{(d-1)})}\times(1-e^{-\delta\lambda_{ij}^{(d)}(t_+^{(d-1)})})\times\Big(\prod_{a\in\mathcal{A}_{\backslash i,j}}(e^{\delta\lambda_{ia}^{(d)}(t_+^{(d-1)})}-1)^{I(a\in J_i)}e^{-\delta\lambda_{ia}^{(d)}(t_+^{(d-1)})}\Big)}{e^{-(t^{(d)}-t^{(d-1)})\sum\limits_{j\in J_i^{(+j)}}\lambda_{ij}^{(d)}(t_+^{(d-1)})}\times(1-e^{-\delta\lambda_{ij}^{(d)}(t_+^{(d-1)})})+e^{-(t^{(d)}-t^{(d-1)})\sum\limits_{j\in J_i^{(-j)}}\lambda_{ij}^{(d)}(t_+^{(d-1)})}e^{-\delta\lambda_{ij}^{(d)}(t_+^{(d-1)})}}\\
&=\frac{e^{-(t^{(d)}-t^{(d-1)})\lambda_{ij}^{(d)}(t_+^{(d-1)})}\times(1-e^{-\delta\lambda_{ij}^{(d)}(t_+^{(d-1)})})\times\Big(\prod_{a\in\mathcal{A}_{\backslash i,j}}(e^{\delta\lambda_{ia}^{(d)}(t_+^{(d-1)})}-1)^{I(a\in J_i)}e^{-\delta\lambda_{ia}^{(d)}(t_+^{(d-1)})}\Big)}{e^{-(t^{(d)}-t^{(d-1)})\lambda_{ij}^{(d)}(t_+^{(d-1)})}\times(1-e^{-\delta\lambda_{ij}^{(d)}(t_+^{(d-1)})})+e^{-\delta\lambda_{ij}^{(d)}(t_+^{(d-1)})}}.
\end{aligned}
\tag{11}
$$

Next, when it comes to the inference, conditioned upon the existence of the d^{th} document at some particular time $t^{(d)}$, the probability that the document is sent from $i^{(d)}$ to $J^{(d)}(=J_{i^{(d)}})$ is

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{N}^{(d)})&=P\Big(\Delta T_{i^{(d)}J^{(d)}}\sim\text{Exp}(\lambda_{i^{(d)}J^{(d)}}^{(d)}(t_+^{(d-1)}))\Big)\times P\Big(\min(\Delta T_{iJ_i})=\Delta T_{i^{(d)}J^{(d)}}\Big)\times P\Big(\text{Edge creations}\Big)\\
&=\Big(\lambda_{i^{(d)}J^{(d)}}^{(d)}(t_+^{(d-1)})e^{-(t^{(d)}-t^{(d-1)})\lambda_{i^{(d)}J^{(d)}}^{(d)}(t_+^{(d-1)})}\Big)\times\Big(e^{-(t^{(d)}-t^{(d-1)})\sum\limits_{i\neq i^{(d)}}\lambda_{iJ_i}^{(d)}(t_+^{(d-1)})}\Big)\\
&\quad\times\Big(\prod_{i\in\mathcal{A}}\prod_{j\in\mathcal{A}_{\backslash i}}(1-e^{-\delta\lambda_{ij}^{(d)}(t_+^{(d-1)})})^{I(j\in J_i)}(e^{-\delta\lambda_{ij}^{(d)}(t_+^{(d-1)})})^{1-I(j\in J_i)}\Big)\\
&=\Big(\lambda_{i^{(d)}J^{(d)}}^{(d)}(t_+^{(d-1)})e^{-(t^{(d)}-t^{(d-1)})\sum\limits_{i\in\mathcal{A}}\lambda_{iJ_i}^{(d)}(t_+^{(d-1)})}\Big)\times\Big(\prod_{i\in\mathcal{A}}\prod_{j\in\mathcal{A}_{\backslash i}}(e^{\delta\lambda_{ij}^{(d)}(t_+^{(d-1)})}-1)^{I(j\in J_i)}e^{-\delta\lambda_{ij}^{(d)}(t_+^{(d-1)})}\Big),
\end{aligned}
\tag{12}
$$

4.2 Resampling $\mathcal{Z}$

The first variable we are going to resample is the topic assignments, one words in a document at a time. The new values of $z_n^{(d)}$ are sampled using the conditional posterior probability of being topic k as we derived in APPENDIX C:

$$
\begin{aligned}
&P(z_n^{(d)}=k|\mathcal{W},\mathcal{Z}_{\backslash d,n},\mathcal{C},\mathcal{X},\mathcal{B},\mathcal{N},\beta,\boldsymbol{u},\alpha,\boldsymbol{m},\sigma^2,\mu,\delta)\\
&\propto P(z_n^{(d)}=k,w_n^{(d)}|\mathcal{W}_{\backslash d,n},\mathcal{Z}_{\backslash d,n},\beta,\boldsymbol{u},\alpha,\boldsymbol{m})
\end{aligned}
\tag{13}
$$

where the subscript "$\backslash d,n$" denotes the exclsuion of position n in d^{th} email. We write the conditional probability as:

$$
\propto(N_{k|d,\backslash d,n}+\alpha\boldsymbol{m}_k)\times\frac{N^{WK}_{w_n^{(d)}k,\backslash d,n}+\frac{\beta}{W}}{\sum_{w=1}^{W}N^{WK}_{wk,\backslash d,n}+\beta}.
\tag{14}
$$

which is the well-known form of collapsed Gibbs sampling equation for LDA. After taking the log, the sampling equation we use for the update of topic assignment is:

$$
\log(N_{k|d,\backslash d,n}+\alpha\boldsymbol{m}_k)+\log(N^{WK}_{w_n^{(d)}k,\backslash d,n}+\frac{\beta}{W})-\log(\sum_{w=1}^{W}N^{WK}_{wk,\backslash d,n}+\beta).
\tag{15}
$$

4.3 Resampling $\mathcal{C}$

The next variable we are going to resample is the topic-interaction pattern assignments, one topic at a time. To obtain the Gibbs sampling equation, which is the posterior conditional probability for the interaction pattern $\mathcal{C}$ for k^{th} topic, i.e. $P(C_k = c|\mathcal{W}, \mathcal{Z}, \mathcal{C}_{\backslash k}, \mathcal{X}, \mathcal{B}, \mathcal{N}, \beta, \boldsymbol{u}, \alpha, \boldsymbol{m}, \sigma^2, \mu, \delta)$. We can derive the equation as below:

$$
\begin{aligned}
&P(C_k = c|\mathcal{W}, \mathcal{Z}, \mathcal{C}_{\backslash k}, \mathcal{X}, \mathcal{B}, \mathcal{N}, \beta, \boldsymbol{u}, \alpha, \boldsymbol{m}, \sigma^2, \mu, \delta)\\
&\propto P(C_k, \mathcal{N}|\mathcal{W}, \mathcal{Z}, \mathcal{C}_{\backslash k}, \mathcal{X}, \mathcal{B}, \mathcal{N}, \beta, \boldsymbol{u}, \alpha, \boldsymbol{m}, \sigma^2, \mu, \delta)\\
&\propto P(C_k|\mathcal{C}_{\backslash k})P(\mathcal{N}|\mathcal{Z}, c^{(k)} = c, \mathcal{C}_{\backslash k}, \mathcal{X}, \mathcal{B}, \mu, \delta),
\end{aligned}
\tag{16}
$$

where $P(C_k = c|\mathcal{C}_{\backslash k}) = \frac{1}{C}$ so this term disappears. We also know that $P(\mathcal{N}|\mathcal{Z}, c^{(k)} = c, \mathcal{C}_{\backslash k}, \mathcal{X}, \mathcal{B}, \mu, \delta)$ is the probability of observing a document with the sender, receiver, and time equal to $(i = i^{(d)}, j = J^{(d)}, t = t^{(d)})$, respectively, given a set of parameter values. As shown in 4.1, we plug in:

$$
\begin{aligned}
&P(\mathcal{N}|\mathcal{Z}, C_k = c, \mathcal{C}_{\backslash k}, \mathcal{X}, \mathcal{B}, \mu, \delta)\\
&= \prod_{d=1}^{D} \Big(\lambda^{(d)}_{i^{(d)}J^{(d)}}(t_+^{(d-1)}) e^{-(t^{(d)}-t^{(d-1)})\sum_{i\in\mathcal{A}} \lambda^{(d)}_{iJ_i}(t_+^{(d-1)})}\Big) \times \Big(\prod_{i\in\mathcal{A}}\prod_{j\in\mathcal{A}_{\backslash i}} (e^{\delta\lambda^{(d)}_{ij}(t_+^{(d-1)})} - 1)^{I(j\in J_i)} e^{-\delta\lambda^{(d)}_{ij}(t_+^{(d-1)})}\Big),
\end{aligned}
\tag{17}
$$

where $\lambda^{(d)}_{ij}(t_+^{(d-1)})$ equals to $\sum_{c=1}^{C} p_c^{(d)} \cdot \exp\Big\{\boldsymbol{\beta}^{(c)T}\boldsymbol{x}^{*(c)}_{t_+^{(d-1)}}(i,j)\Big\} \cdot 1\{j \in \mathcal{A}_{\backslash i}\}$. After replacing λ to the function of $\boldsymbol{\beta}$ and $\boldsymbol{x}$ and then taking the log of Equation (18) to avoid numerical issue from exponentiation and increase the speed of computation, the sampling equation used for the update of interaction pattern assignment is:

$$
\begin{aligned}
&\sum_{d=1}^{D}\Big\{\sum_{c=1}^{C}\Big(\log(p_c^{(d)}) + \sum_{j\in J_{i^{(d)}}} \boldsymbol{\beta}^{(c)T}\boldsymbol{x}^{*(c)}_{t_+^{(d-1)}}(i^{(d)}, j)\Big) - (t^{(d)} - t^{(d-1)})\sum_{i\in\mathcal{A}}\sum_{c=1}^{C}\Big(p_c^{(d)} e^{\sum_{j\in J_i}\boldsymbol{\beta}^{(c)T}\boldsymbol{x}^{*(c)}_{t_+^{(d-1)}}(i,j)}\Big)\\
&+\sum_{i\in\mathcal{A}}\sum_{j\in\mathcal{A}_{\backslash i}}\Big(I(j\in J_i)\cdot\log(e^{\delta\sum_{c=1}^{C}\Big(p_c^{(d)}e^{\boldsymbol{\beta}^{(c)T}\boldsymbol{x}^{*(c)}_{t_+^{(d-1)}}(i,j)}\Big)} - 1) - \delta\sum_{c=1}^{C}\Big(p_c^{(d)}e^{\boldsymbol{\beta}^{(c)T}\boldsymbol{x}^{*(c)}_{t_+^{(d-1)}}(i,j)}\Big)\Big)\Big\},
\end{aligned}
\tag{18}
$$

with $C_k = c$ throughout, making only $\{p_c^{(d)}\}_{d=1}^{D}$ vary by C_k. Note that in the product over d, we need only consider those emails that actually use topic k; the others will have no terms involving C_k.

4.4 Resampling $\mathcal{B}$

Finally, we update the interaction pattern parameter $\boldsymbol{\beta}^{(c)}$, one interaction pattern at a time. For this, we use the Metropolis-Hastings algorithm with a proposal density Q being the multivariate Gaussian distribution, with variance $\boldsymbol{\beta}_B^2$ (proposal distirbution variance parameters set by the user), centered on the current values of $\boldsymbol{\beta}^{(c)}$. Then we draw a proposal $\boldsymbol{\beta}'^{(c)}$ at each iteration. Under symmetric proposal distribution (such as multivariate Gaussian), we cancel out Q-ratio and obtain the acceptance probability equal to:

$$
\text{Acceptance Probability} = \begin{cases} \frac{P(\boldsymbol{\beta}'^{(c)}|\mathcal{W},\mathcal{Z},\mathcal{C},\mathcal{X},\mathcal{B}_{\backslash c},\mathcal{N},\beta,\boldsymbol{u},\alpha,\boldsymbol{m},\sigma^2,\mu,\delta)}{P(\boldsymbol{\beta}^{(c)}|\mathcal{W},\mathcal{Z},\mathcal{C},\mathcal{X},\mathcal{B}_{\backslash c},\mathcal{N},\beta,\boldsymbol{u},\alpha,\boldsymbol{m},\sigma^2,\mu,\delta)} & \text{if} < 1 \\ 1 \quad \text{else} \end{cases}
\tag{19}
$$

After factorization, we get

$$
\begin{aligned}
&\frac{P(\boldsymbol{\beta}'^{(c)}|\mathcal{W},\mathcal{Z},\mathcal{C},\mathcal{X},\mathcal{B}_{\backslash c},\mathcal{N},\beta,\boldsymbol{u},\alpha,\boldsymbol{m},\sigma^2,\mu,\delta)}{P(\boldsymbol{\beta}^{(c)}|\mathcal{W},\mathcal{Z},\mathcal{C},\mathcal{X},\mathcal{B}_{\backslash c},\mathcal{N},\beta,\boldsymbol{u},\alpha,\boldsymbol{m},\sigma^2,\mu,\delta)}\\
&= \frac{P(\boldsymbol{\beta}'^{(c)}|\mathcal{W},\mathcal{Z},\mathcal{C},\mathcal{X},\mathcal{B}_{\backslash c},\mathcal{N},\beta,\boldsymbol{u},\alpha,\boldsymbol{m},\sigma^2,\mu,\delta)}{P(\mathcal{W},\mathcal{Z},\mathcal{C},\mathcal{B},\mathcal{N}|\mathcal{X},\beta,\boldsymbol{u},\alpha,\boldsymbol{m},\sigma^2,\mu,\delta)}\\
&= \frac{P(\boldsymbol{\beta}'^{(c)})}{P(\boldsymbol{\beta}^{(c)})} \times \frac{P(\mathcal{N}|\boldsymbol{\beta}'^{(c)},\mathcal{Z},\mathcal{C},\mathcal{X},\mathcal{B}_{\backslash c},\mu,\delta)}{P(\mathcal{N}|\mathcal{Z},\mathcal{C},\mathcal{X},\mathcal{B},\mu,\delta)},
\end{aligned}
\tag{20}
$$

where $P(\mathcal{N}|\mathcal{Z},\mathcal{C},\mathcal{X},\mathcal{B},\mu,\delta)$ is Equation (18). As in Section 4.3, we obtain the log of acceptance ratio:

$$\begin{aligned}
&\log\Big(\phi_d(\boldsymbol{\beta}'^{(c)};\mathbf{0},\sigma^2 I_P)\Big)-\log\Big(\phi_d(\boldsymbol{\beta}^{(c)};\mathbf{0},\sigma^2 I_P)\Big)\\
&+\sum_{d=1}^{D}\Big\{\sum_{c=1}^{C}\Big(\log(p_c^{(d)})+\sum_{j\in J_{i^{(d)}}}\boldsymbol{\beta}'^{(c)T}\boldsymbol{x}^{*(c)}_{t_+^{(d-1)}}(i^{(d)},j)\Big)-(t^{(d)}-t^{(d-1)})\sum_{i\in\mathcal{A}}\sum_{c=1}^{C}\Big(p_c^{(d)}e^{\sum_{j\in J_i}\boldsymbol{\beta}'^{(c)T}\boldsymbol{x}^{*(c)}_{t_+^{(d-1)}}(i,j)}\Big)\\
&\quad+\sum_{i\in\mathcal{A}}\sum_{j\in\mathcal{A}_{\setminus i}}\Big(I(j\in J_i)\cdot\log(e^{\delta\sum_{c=1}^{C}\big(p_c^{(d)}e^{\boldsymbol{\beta}'^{(c)T}\boldsymbol{x}^{*(c)}_{t_+^{(d-1)}}(i,j)}\big)}-1)-\delta\sum_{c=1}^{C}\Big(p_c^{(d)}e^{\boldsymbol{\beta}'^{(c)T}\boldsymbol{x}^{*(c)}_{t_+^{(d-1)}}(i,j)}\Big)\Big)\Big\}\\
&-\sum_{d=1}^{D}\Big\{\sum_{c=1}^{C}\Big(\log(p_c^{(d)})+\sum_{j\in J_{i^{(d)}}}\boldsymbol{\beta}^{(c)T}\boldsymbol{x}^{*(c)}_{t_+^{(d-1)}}(i^{(d)},j)\Big)-(t^{(d)}-t^{(d-1)})\sum_{i\in\mathcal{A}}\sum_{c=1}^{C}\Big(p_c^{(d)}e^{\sum_{j\in J_i}\boldsymbol{\beta}^{(c)T}\boldsymbol{x}^{*(c)}_{t_+^{(d-1)}}(i,j)}\Big)\\
&\quad-\sum_{i\in\mathcal{A}}\sum_{j\in\mathcal{A}_{\setminus i}}\Big(I(j\in J_i)\cdot\log(e^{\delta\sum_{c=1}^{C}\big(p_c^{(d)}e^{\boldsymbol{\beta}^{(c)T}\boldsymbol{x}^{*(c)}_{t_+^{(d-1)}}(i,j)}\big)}-1)-\delta\sum_{c=1}^{C}\Big(p_c^{(d)}e^{\boldsymbol{\beta}^{(c)T}\boldsymbol{x}^{*(c)}_{t_+^{(d-1)}}(i,j)}\Big)\Big)\Big\},
\end{aligned}\tag{21}$$

where $\phi_d(\cdot;\mu,\Sigma)$ is the d-dimensional multivariate normal density. After some simplification, we finally use:

$$\begin{aligned}
&\log\Big(\phi_d(\boldsymbol{\beta}'^{(c)};\mathbf{0},\sigma^2 I_P)\Big)-\log\Big(\phi_d(\boldsymbol{\beta}^{(c)};\mathbf{0},\sigma^2 I_P)\Big)\\
&+\sum_{d=1}^{D}\Big\{\sum_{c=1}^{C}\Big(\sum_{j\in J_{i^{(d)}}}(\boldsymbol{\beta}'^{(c)}-\boldsymbol{\beta}^{(c)})^T\boldsymbol{x}^{*(c)}_{t_+^{(d-1)}}(i^{(d)},j)\Big)\\
&-(t^{(d)}-t^{(d-1)})\sum_{i\in\mathcal{A}}\sum_{c=1}^{C}p_c^{(d)}\Big(e^{\sum_{j\in J_i}\boldsymbol{\beta}'^{(c)T}\boldsymbol{x}^{*(c)}_{t_+^{(d-1)}}(i,j)}-e^{\sum_{j\in J_i}\boldsymbol{\beta}^{(c)T}\boldsymbol{x}^{*(c)}_{t_+^{(d-1)}}(i,j)}\Big)\\
&+\sum_{i\in\mathcal{A}}\sum_{j\in\mathcal{A}_{\setminus i}}\Big(I(j\in J_i)\cdot\log(e^{\delta\sum_{c=1}^{C}p_c^{(d)}\big(e^{\boldsymbol{\beta}'^{(c)T}\boldsymbol{x}^{*(c)}_{t_+^{(d-1)}}(i,j)}-e^{\boldsymbol{\beta}^{(c)T}\boldsymbol{x}^{*(c)}_{t_+^{(d-1)}}(i,j)}\big)}-1)\\
&\qquad-\delta\sum_{c=1}^{C}p_c^{(d)}\Big(e^{\boldsymbol{\beta}'^{(c)T}\boldsymbol{x}^{*(c)}_{t_+^{(d-1)}}(i,j)}-e^{\boldsymbol{\beta}^{(c)T}\boldsymbol{x}^{*(c)}_{t_+^{(d-1)}}(i,j)}\Big)\Big)\Big\}.
\end{aligned}\tag{22}$$

Then the log of acceptance ratio we have is:

$$\log(\text{Acceptance Probability})=\min((22),\,0)\tag{23}$$

To determine whether we accept the proposed update or not, we take the usual approach, by comparing the log of acceptance ratio we have to the log of a sample from uniform(0,1). Also, note that in the sum over d, we need only consider those emails that actually use topics corresponding to interaction pattern c.

4.5 Sampling of tuning parameters (μ,δ)

μ and δ are the two tuning parameters; μ is a parameter of time-decay, reflecting the duration of influence of past interactions on future interactions, and δ is a parameter in the tie generating process that controls the number of recipients and the urgency of the document. Since these two parameters both are only involved in $P(\mathcal{N}|\mathcal{Z},\mathcal{C},\mathcal{X},\mathcal{B},\mu,\delta)$, sampling of (μ,δ) could be incorporated into sampling of $\mathcal{B}$.

Should we transform (μ,δ) to $(\mu^{(c)},\delta^{(c)})$? Current setup does not allow simultaneous sampling since we update $\boldsymbol{\beta}^{(c)}$, one interaction pattern at a time. Alternatively, we can sample $\mathcal{B}$ all at once, instead of one interaction at a time. CPME tried both, but latter (all at once) may yield low acceptance rate since we have $C\times P+2$ elements updated at once, while the former $((\mu^{(c)},\delta^{(c)}))$ does C different sampling of $P+2$ elements.

4.6 Pseudocode

To implement the inference procedure outlined above, we provide a pseudocode for Markov Chain Monte Carlo (MCMC) sampling. Note that we use two loops, outer iteration and inner iteration, in order to avoid the label switching problem (Jasra et al., 2005), which is an issue caused by the nonidentifiability of the components under symmetric priors in Bayesian mixture modeling. When summarizing model results, we will only use the values from the last I^{th} outer loop because there is no label switching problem within the inner iteration.

Algorithm 5 MCMC

```
set initial values C^(0), Z^(0), and B^(0)
for i=1 to I do
    optimize α and m using hyperparameter optimization in Wallach (2008)
    for d=1 to D do
        sample the augmented data (i, J_i) for i ∈ A_{\i^(d)} (See Section 4.1)
    end
    for n=1 to n_1 do
        fix C = C^(i) and B = B^(i−1)
        for d=1 to D do
            for n=1 to N^(d) do
                calculate p^Z|others = (p_1, ..., p_K) using Equation (15)
                draw of z_n^(d) ∼ multinomial(p^Z)
            end
        end
    end
    for n=1 to n_2 do
        fix Z = Z^(i−1) and B = B^(i−1)
        for k=1 to K do
            calculate p^C|others = (p_1, ..., p_C) using Equation (18)
            draw C_k ∼ multinomial(p^C)
        end
    end
    for n=1 to n_3 do
        fix C = C^(i), Z = Z^(i), and B^(0) = last value (n_3^th) of B^(i−1)
        calculate X = {x*^(c)_{t^(d)}(i, j)}_{d=1}^D according to Section 3 conditioned on others
        for c=1 to C do
            sample β^(c)|others using Equation (22) and (23)
            draw δ, a tuning parameter that controls the number of multicasts
            draw μ, the time decay parameter
        end
    end
end
summarize the results using:
the last value of C, the last value of Z, and the last n_3 length chain of B
```

5 Appliction to North Carolina email data

To see the applicability of the model, we used the North Carolina email data using two counties, Vance county and Dare county, which are the two counties whose email corpus cover the date of Hurricane Sandy (October 22, 2012 – November 2, 2012). Especially, Dare county geographically covers the Outer Banks, so we would like to see how the communication pattern changes during the time period surrounding Hurricane Sandy. Here we apply IPTM to both data and demonstrate the effectiveness of the model at predicting and explaining continuous-time textual communications.

5.1 Vance county email data

Vance county data contains $D = 185$ emails sent between $A = 18$ actors, including $W = 620$ vocabulary in total. We used $K = 10$ topics and $C = 2$ interaction patterns. MCMC sampling was implemented based on the order and scheme illustrated in Section 4. We set the outer iteration number as $I = 1000$, the inner iteration numbers as $n_1 = 3, n_2 = 3$, and $n_3 = 3300$. First 100 outer iterations and first 300 iterations of third inner iteration were used as a burn-in, and every 10^{th} sample was taken as a thinning process of third inner iteration. In addition, after some experimentation, δ_B was set as 0.5, to ensure sufficient acceptance rate. MCMC diagnostic plots are attached in APPENDIX D, as well as the geweke test statistics.

Below are the summary of IP-topic-word assignments. Each interaction pattern is paired with (a) posterior estimates of dynamic network effects $\boldsymbol{\beta}^{(c)}$ corresponding to the interaction pattern, (b) the top 3 topics most likely to be generated conditioned on the interaction pattern, and (c) the top 10 most likely words to have generated conditioned on the topic and interaction pattern. By examining the estimates in Table 2 and

	IP1 (54 emails)	**IP2** (107 emails)	**IP3** (108 emails)
intercept	0.515 [-0.523, 1.546]	-0.364 [-2.108, 1.934]	-1.230 [-1.948, 0.194]
send	1.916* [1.130, 2.937]	2.843* [1.969, 3.885]	2.531* [1.595, 3.568]
receive	0.158 [-1.126, 1.098]	3.068* [2.427, 4.509]	1.067* [0.488, 1.781]
triangles	1.483 [-0.507, 2.558]	-1.478* [-2.038, -0.918]	-1.787* [-3.062, -0.958]
outdegree	0.514 [-0.570, 1.377]	0.492 [-0.804, 1.665]	0.771 [-1.152, 2.544]
indegree	2.166* [1.534, 2.895]	1.397* [0.720, 2.187]	2.464* [1.840, 3.327]

Table 1: Summary of posterior estimates of $\boldsymbol{\beta}^{(c)}$ for Vance county emails

Figure 2: Posterior distribution of $\boldsymbol{\beta}^{(c)}$ for Vance county emails

their corresponding interpretaiton, it seems that there exist strong effects of dynamic network covariates. That is, whether the sender and receiver previously had dyadic or triangle interaction strongly increase the rate of their interactions. Moreover, to see the differences across the interaction patterns more clearly, we compared the posterior distribution using the boxplots in Figure 2 and it seems that there exists notable differences in dynmic network covariates across the interaction patterns. For example, IP2 has the highest send and receive effect and IP3 has the highest outdegree and indegree effect, while its baseline intensity (i.e. intercept) or triangle effect are not as high as other interaction patterns. Later, multiple hypothesis testing could be applied in order to test the significance of the differences in $\boldsymbol{\beta}^{(c)}$ across the C number of interaction patterns.

Next, we scrutinize the topic distributions corresponding to each interaction patterns in Figure 3. There is some distinctive differences in the topic distributions $\mathcal{Z}$, given the assignment of interaction patterns to the documents $\mathcal{C}$. Specifically, each interaction pattern has different topics as the topic with highest probability.

Figure 3: Posterior distribution of $\mathcal{Z}$ for Vance county emails, with (upper) and without (lower) considering IP

Furthermore, we look at the distribution of words given the topics, which corresponds to Algorithm 4 in the generative process. Since the topic-word distribution ϕ does not depend on the interaction patterns as previous cases, Table 3 lists top 10 topics with top 10 words that have the highest probability conditioned on the topic. In addition, this time we try to check the interaction pattern-word distribution by listing top 10 words that have the highest probability conditioned on the interaction pattern. It seems that the words are not significantly different, having several words like 'director', 'phones', 'department', 'description', or 'henderson' (county seat of Vance county) appeared repetitively across the most of the topics or interaction patterns. The word 'will' was ranked the top in most of the lists, probably because it was not deleted during the text mining process while other similar type of words like ‘am’, ‘is’, ‘are’, or ‘can’ are all removed.

IP1 (54 emails)	**IP2** (107 emails)	**IP3** (108 emails)
K=2 (0.40), K=4 (0.17), K=9 (0.15)	K=8 (0.38), K=5 (0.24)	K=1 (0.31), K=3 (0.17), K=6 (0.15)
message, electronic, records	phones, phone	henderson, street, description
time, response, ncgs	week, department	latest, fax, church
department, hereto, attachments	system, rest	planning, suite, emergency
heads, finance, director	october, advised	attached, director, center
request, financial, operations	training, jail	extension, goldvancesealimprovedjordan, phone
manager, system, work	cutting, send	development, phase, morning
pursuant, additional, office	cutover, center	e-mail, board, email
chapter, class, helped	folks, tuesday	good, rural, excel
public, local, internal	instructions, monday	young, funds, lease
subject, communications, reporting	dss, thursday	list, turn, form

Table 2: Summary of top 5 topics with top 10 words that have the highest probability conditioned on the topic (Symmetric)

5.2 Dare county email data

Dare county data contains $D = 2247$ emails between $A = 27$ actors, including $W = 2907$ vocabulary in total. Again, we used $K = 10$ topics and $C = 3$ interaction patterns. MCMC sampling was implemented based on the order and scheme illustrated earlier. We set the outer iteration number as $I = 1000$, and inner iteration numbers as $n_1 = 3, n_2 = 3$, and $n_3 = 3300$. In addition, after some experimentation, δ_B was set as 0.02, to ensure sufficient acceptance rate. In our case, the average acceptance rate for $\boldsymbol{\beta}$ was 0.277. As demonstrated in Algorithm 5, the last value of $\mathcal{C}$, the last value of $\mathcal{Z}$, and the last n_3 length chain of $\mathcal{B}$ were taken as the final posterior samples. Among the $\mathcal{B}$ samples, 300 were discarded as a burn-in and every 10^{th} samples were taken. After these post-processing, MCMC diagnostic plots are attached in APPENDIX D, as well as geweke test statistics.

APPENDIX

APPENDIX A: Notations in IPTM

Sender of the d^{th} document	$i^{(d)}$	Scalar
Receivers of the d^{th} document	$J^{(d)}$	$\lvert J^{(d)}\rvert$-dimensional vector
Individual receiver of the d^{th} document	$j^{(d)}$	Scalar
Time of the d^{th} document	$t^{(d)}$	Scalar
Authors of the corpus	$\mathcal{A}$	Set
Number of authors	A	Scalar
Number of documents	D	Scalar
Number of words in the d^{th} document	$N^{(d)}$	Scalar
Number of topics	K	Scalar
Vocabulary size	W	Scalar
Number of interaction patterns	C	Scalar
Number of words assigned to word and topic	N^{WK}	Scalar
Interaction pattern of the k^{th} topic	C_k	Scalar
Tuning parameter in tie generative process	δ	Scalar
Time decay parameter	μ	Scalar
Poisson parameter for number of words $N^{(d)}$	ζ	Scalar
Words in the d^{th} document	$\boldsymbol{w}^{(d)}$	$N^{(d)}$-dimensional vector
n^{th} word in the d^{th} document	$w_n^{(d)}$	n^{th} component of $\boldsymbol{w}^{(d)}$
Topic assignments in the d^{th} document	$\boldsymbol{z}^{(d)}$	$N^{(d)}$-dimensional vector
Topic assignments for n^{th} word in the d^{th} document	$z_n^{(d)}$	n^{th} component of $\boldsymbol{z}^{(d)}$
Dirichlet concentration prior for document topic distribution	α	Scalar
Dirichlet base prior for document topic distribution	$\boldsymbol{m}$	K-dimensional vector
Dirichlet concentration prior for topic word distribution	β	Scalar
Dirichlet base prior for topic word distribution	$\boldsymbol{u}$	W-dimensional vector
Variance of Normal prior	σ^2	Scalar
Probabilities of the words given topic k	$\boldsymbol{\phi}^{(k)}$	W-dimensional vector
Probabilities of the topics given the d^{th} document	$\boldsymbol{\theta}^{(d)}$	K-dimensional vector
Coefficient of the intensity process given interaction pattern c	$\boldsymbol{\beta}^{(c)}$	p-dimensional vector
Network statistics for (i, j) at time t given interaction pattern c	$\boldsymbol{x}_t^{(c)}(i, j)$	p-dimensional vector
Stochastic intensity of document d at time t	$\boldsymbol{\lambda}^{(d)}(t)$	$A \times A$ matrix
Stochastic intensity of (i, j) in document d at time t	$\lambda_{ij}^{(d)}(t)$	Scalar
Proportion of topics in document d corresponding to interaction pattern c	$p_c^{(d)}$	Scalar
Time increments associated to (i, J) pair	ΔT_{iJ}	Scalar

Table 3: Symbols associated with IPTM, as used in this paper

APPENDIX B: Integrating out Φ and Θ in latent Dirichlet allocation

$$\begin{aligned}
&P(\Phi, \Theta, \mathcal{W}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \mathcal{N} \mid \mathcal{X}, \beta, \boldsymbol{u}, \alpha, \boldsymbol{m}, \sigma^2, \mu, \delta) \\
&= P(\mathcal{W}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \mathcal{N} \mid \Phi, \Theta, \mathcal{X}, \sigma^2, \mu, \delta) P(\Phi, \Theta \mid \beta, \boldsymbol{u}, \alpha, \boldsymbol{m}) \\
&= P(\mathcal{W} \mid \mathcal{Z}, \Phi) P(\mathcal{Z} \mid \Theta) P(\mathcal{N} \mid \mathcal{C}, \mathcal{X}, \mathcal{B}, \mu, \delta) P(\mathcal{B} \mid \mathcal{C}, \sigma^2) P(\Phi \mid \beta, \boldsymbol{u}) P(\Theta \mid \alpha, \boldsymbol{m}) P(\mathcal{C}) \\
&= \Big[\prod_{d=1}^{D} \prod_{n=1}^{N^{(d)}} P(w_n^{(d)} \mid \phi_{z_n^{(d)}})\Big] \times \Big[\prod_{d=1}^{D} \prod_{n=1}^{N^{(d)}} P(z_n^{(d)} \mid \boldsymbol{\theta}^{(d)})\Big] \times \Big[\prod_{d=1}^{D} P(\boldsymbol{N}^{(d)}(t^{(d)}) \mid \mathcal{C}, \mathcal{X}, \mathcal{B}, \mu, \delta)\Big] \\
&\quad \times \Big[\prod_{c=1}^{C} P(\boldsymbol{\beta}^{(c)} \mid \sigma^2)\Big] \times \Big[\prod_{k=1}^{K} P(\boldsymbol{\phi}^{(k)} \mid \beta, \boldsymbol{u})\Big] \times \Big[\prod_{d=1}^{D} P(\boldsymbol{\theta}^{(d)} \mid \alpha, \boldsymbol{m})\Big] \times \Big[\prod_{k=1}^{K} P(C_k)\Big]
\end{aligned} \tag{24}$$

Dropping the terms independent of tokens out, we further rewrite the equation (24) as below:

$$
\begin{aligned}
&\propto \Big[\prod_{d=1}^{D}\prod_{n=1}^{N^{(d)}} P(w_n^{(d)}|\phi_{z_n^{(d)}})\Big] \times \Big[\prod_{d=1}^{D}\prod_{n=1}^{N^{(d)}} P(z_n^{(d)}|\boldsymbol{\theta}^{(d)})\Big] \times \Big[\prod_{k=1}^{K} P(\boldsymbol{\phi}^{(k)}|\beta,\boldsymbol{u})\Big] \times \Big[\prod_{d=1}^{D} P(\boldsymbol{\theta}^{(d)}|\alpha,\boldsymbol{m})\Big] \\
&\propto \Big[\prod_{d=1}^{D}\prod_{n=1}^{N^{(d)}} \phi_{w_n^{(d)} z_n^{(d)}}\Big] \times \Big[\prod_{d=1}^{D}\prod_{n=1}^{N^{(d)}} \boldsymbol{\theta}^{(d)}_{z_n^{(d)}}\Big] \\
&\quad \times \Big[\prod_{k=1}^{K}\Big(\frac{\Gamma(\sum_{w=1}^{W}\beta u_w)}{\prod_{w=1}^{W}\Gamma(\beta u_w)}\prod_{w=1}^{W}\phi_{wk}^{\beta u_w -1}\Big)\Big] \times \Big[\prod_{d=1}^{D}\Big(\frac{\Gamma(\sum_{k=1}^{K}\alpha m_k)}{\prod_{k=1}^{K}\Gamma(\alpha m_k)}\prod_{k=1}^{K}(\boldsymbol{\theta}_k^{(d)})^{\alpha m_k -1}\Big)\Big] \\
&\propto \Big[\frac{\Gamma(\sum_{w=1}^{W}\beta u_w)}{\prod_{w=1}^{W}\Gamma(\beta u_w)}\Big]^{K} \times \prod_{d=1}^{D}\Big[\frac{\Gamma(\sum_{k=1}^{K}\alpha m_k)}{\prod_{k=1}^{K}\Gamma(\alpha m_k)}\Big] \times \Big[\prod_{k=1}^{K}\prod_{w=1}^{W}\phi_{wk}^{N_{wk}^{WK}+\beta u_w -1}\Big] \times \Big[\prod_{d=1}^{D}\prod_{k=1}^{K}(\boldsymbol{\theta}_k^{(d)})^{N_{k|d}+\alpha m_k -1}\Big]
\end{aligned}
\tag{25}
$$

where N_{wk}^{WK} is the number of times the w^{th} word in the vocabulary is assigned to topic k, and $N_{k|d}$ is the number of times topic k shows up in the document d. By looking at the forms of the terms involving Θ and Φ in Equation (21), we integrate out the random variables Θ and Φ, making use of the fact that the Dirichlet distribution is a conjugate prior of multinomial distribution. Applying the well-known formula $\int \prod_{n=1}^{M}[x_m^{k_m-1}dx_m] = \frac{\prod_{n=1}^{M}\Gamma(k_m)}{\Gamma(\sum_{n=1}^{M}k_m)}$ to (22), we have:

$$
\begin{aligned}
&P(\mathcal{W},\mathcal{Z},\mathcal{C},\mathcal{B},\mathcal{N}|\mathcal{X},\beta,\boldsymbol{u},\alpha,\boldsymbol{m},\boldsymbol{\gamma},\boldsymbol{\eta},\sigma^2) \\
&= \text{Const.}\int_{\Theta}\int_{\Phi}\Big[\prod_{k=1}^{K}\prod_{w=1}^{W}\phi_{wk}^{N_{wk}^{WK}+\beta u_w -1}\Big]\Big[\prod_{d=1}^{D}\prod_{k=1}^{K}(\boldsymbol{\theta}_k^{(d)})^{N_{k|d}+\alpha m_k -1}\Big]d\Phi d\Theta \\
&= \text{Const.}\Big[\prod_{k=1}^{K}\int_{\phi_{:k}}\prod_{w=1}^{W}\phi_{wk}^{N_{wk}^{WK}+\beta u_w -1}d\phi_{:k}\Big] \times \Big[\prod_{d=1}^{D}\int_{\theta_{:d}}\prod_{k=1}^{K}(\boldsymbol{\theta}_k^{(d)})^{N_{k|d}+\alpha m_k -1}d\theta_{:d}\Big] \\
&= \text{Const.}\Big[\prod_{k=1}^{K}\frac{\prod_{w=1}^{W}\Gamma(N_{wk}^{WK}+\frac{\beta}{W})}{\Gamma(\sum_{w=1}^{W}N_{wk}^{WK}+\beta)}\Big] \times \Big[\prod_{d=1}^{D}\frac{\prod_{k=1}^{K}\Gamma(N_{k|d}+\alpha m_k)}{\Gamma(N_{\cdot|d}+\alpha)}\Big].
\end{aligned}
\tag{26}
$$

APPENDIX C: Conditional probability of $\mathcal{Z}$

$$
\begin{aligned}
&P(\boldsymbol{w}^{(d)},\boldsymbol{z}^{(d)}|\mathcal{W}_{\backslash d},\mathcal{Z}_{\backslash d},\beta,\boldsymbol{u},\alpha,\boldsymbol{m}) \\
&\propto \prod_{n=1}^{N^{(d)}} P(z_n^{(d)}=k, w_n^{(d)}=w|\mathcal{W}_{\backslash d,n},\mathcal{Z}_{\backslash d,n},\beta,\boldsymbol{u},\alpha,\boldsymbol{m})
\end{aligned}
\tag{27}
$$

To obtain the Gibbs sampling equation, we need to obtain an expression for $P(z_n^{(d)}=k, w_n^{(d)}=w|\mathcal{W}_{\backslash d,n},\mathcal{Z}_{\backslash d,n},\beta,\boldsymbol{u},\alpha,\boldsymbol{m})$, From Bayes' theorem and Gamma identity $\Gamma(k+1)=k\Gamma(k)$,

$$
\begin{aligned}
&P(z_n^{(d)}=k, w_n^{(d)}=w|\mathcal{W}_{\backslash d,n},\mathcal{Z}_{\backslash d,n},\beta,\boldsymbol{u},\alpha,\boldsymbol{m}) \\
&\propto \frac{P(\mathcal{W},\mathcal{Z}|\beta,\boldsymbol{u},\alpha,\boldsymbol{m})}{P(\mathcal{W}_{\backslash d,n},\mathcal{Z}_{\backslash d,n}|\beta,\boldsymbol{u},\alpha,\boldsymbol{m})} \\
&\propto \frac{\prod_{k=1}^{K}\frac{\prod_{w=1}^{W}\Gamma(N_{wk}^{WK}+\beta u_w)}{\Gamma(\sum_{w=1}^{W}N_{wk}^{WK}+\beta)} \times \prod_{k=1}^{K}\frac{\Gamma(N_{k|d}+\alpha m_k)}{\Gamma(N_{\cdot|d}+\alpha)}}{\prod_{k=1}^{K}\frac{\prod_{w=1}^{W}\Gamma(N_{wk,\backslash d,n}^{WK}+\beta u_w)}{\Gamma(\sum_{w=1}^{W}N_{wk,\backslash d,n}^{WK}+\beta)} \times \prod_{k=1}^{K}\frac{\Gamma(N_{k|d,\backslash d,n}+\alpha m_k)}{\Gamma(N_{\cdot|d,\backslash d,n}+\alpha)}} \\
&\propto \frac{N_{wk,\backslash d,n}^{WK}+\frac{\beta}{W}}{\sum_{w=1}^{W}N_{wk,\backslash d,n}^{WK}+\beta} \times \frac{N_{k|d,\backslash d,n}+\alpha m_k}{N^{(d)}-1+\alpha}
\end{aligned}
\tag{28}
$$

Then, same as for LDA, we also know that the topic assignment $z_n^{(d)}=k$ is obtained by:

$$
P(z_n^{(d)}=k|w_n^{(d)}=w,\mathcal{W}_{\backslash d,n},\mathcal{Z}_{\backslash d,n},\beta,\boldsymbol{u},\alpha,\boldsymbol{m}) \propto \frac{N_{k|d,\backslash d,n}+\alpha m_k}{N^{(d)}-1+\alpha}
\tag{29}
$$

In addition, the conditional probability that a new word generated in the document would be $w_n^{(d)} = w$, given that it is generated from topic $z_n^{(d)} = k$ is obtained by:

$$P(w_m^{(d)} = w | z_m^{(d)} = k, \mathcal{W}_{\backslash d,n}, \mathcal{Z}_{\backslash d,nm}, \beta, \boldsymbol{u}, \alpha, \boldsymbol{m}) \propto \frac{N_{wk,\backslash d,n}^{WK} + \frac{\beta}{W}}{\sum_{w=1}^{W} N_{wk,\backslash d,n}^{WK} + \beta} \tag{30}$$

NOTE: Using Equation (28), the unnormalized constant we use to check the model convergence and the corresponding log-constant is,

$$\begin{aligned} &\prod_{d=1}^{D} \prod_{n=1}^{N^{(d)}} P(z_n^{(d)} = k, w_n^{(d)} = w | \mathcal{W}_{\backslash d,n}, \mathcal{Z}_{\backslash d,m}, \beta, \boldsymbol{u}, \alpha, \boldsymbol{m}) \\ &\propto \prod_{d=1}^{D} \prod_{n=1}^{N^{(d)}} \frac{N_{w_n^{(d)} z_n^{(d)},\backslash d,n}^{WK} + \frac{\beta}{W}}{\sum_{w=1}^{W} N_{w z_n^{(d)},\backslash d,n}^{WK} + \beta} \times \frac{N_{k|d,\backslash d,n} + \alpha m_{z_n^{(d)}}}{N^{(d)} - 1 + \alpha}, \end{aligned} \tag{31}$$

$$\begin{aligned} &\sum_{d=1}^{D} \sum_{n=1}^{N^{(d)}} \log\Big(P(z_n^{(d)} = k, w_n^{(d)} = w | \mathcal{W}_{\backslash d,n}, \mathcal{Z}_{\backslash d,n}, \beta, \boldsymbol{u}, \alpha, \boldsymbol{m})\Big) \\ &\propto \sum_{d=1}^{D} \sum_{n=1}^{N^{(d)}} \log\Big(N_{w_m^{(d)} z_m^{(d)},\backslash d,n}^{WK} + \frac{\beta}{W}\Big) - \log\Big(\sum_{w=1}^{W} N_{w z_m^{(d)},\backslash d,n}^{WK} + \beta\Big) \\ &+ \log\Big(N_{k|d,\backslash d,n} + \alpha m_{z_m^{(d)}}\Big) - \log\Big(N^{(d)} - 1 + \alpha\Big) \end{aligned} \tag{32}$$

APPENDIX D: MCMC Diagnostics

References

Blei, D. M. and Lafferty, J. D. (2006). Dynamic topic models. In *Proceedings of the 23rd international conference on Machine learning*, pages 113–120. ACM.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022.

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, page 205. IGI Global.

Butts, C. T. (2008). A relational event framework for social action. *Sociological Methodology*, 38(1):155–200.

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

Desmarais, B. A. and Cranmer, S. J. (2017). Statistical inference in political networks research. In Victor, J. N., Montgomery, A. H., and Lubell, M., editors, *The Oxford Handbook of Political Networks*. Oxford University Press.

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.

Jasra, A., Holmes, C., and Stephens, D. (2005). Markov chain monte carlo methods and the label switching problem in bayesian mixture modeling. *Statistical Science*, pages 50–67.

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

McCallum, A., Corrada-Emmanuel, A., and Wang, X. (2005). Topic and role discovery in social networks.

Minka, T. (2000). Estimating a dirichlet distribution.

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.

Pew, R. C. (2016). Social media fact sheet. *Accessed on 03/07/17.*

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.

Rosen-Zvi, M., Griffiths, T., Steyvers, M., and Smyth, P. (2004). The author-topic model for authors and documents. In *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, pages 487–494. AUAI Press.

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

Snijders, T. A. (2017). Stochastic actor-oriented models for network dynamics. *Annual Review of Statistics and Its Application*, (0).

Szóstek, A. M. (2011). ?dealing with my emails?: Latent user needs in email management. *Computers in Human Behavior*, 27(2):723–729.

Wallach, H. M. (2008). *Structured topic models for language*. PhD thesis, University of Cambridge.

Wallach, H. M., Mimno, D. M., and McCallum, A. (2009). Rethinking lda: Why priors matter. In *Advances in neural information processing systems*, pages 1973–1981.

Wang, X. and McCallum, A. (2006). Topics over time: a non-markov continuous-time model of topical trends. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 424–433. ACM.

Zhou, M. (2015). Infinite edge partition models for overlapping community detection and link prediction. In *AISTATS*.