A Network Model for Continuous Time Textual Communications 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 1, 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 time-stamped textual communications, such as emails, according to both the network dynamics that they reflect and their content. To do this, it draws on the Cox multiplicative intensity model—a generative model for ties that tend toward structural features such as reciprocated dyads and triangles—and latent Dirichlet allocation—a generative model for topic-based content. The IPTM assigns each communication to an "interaction pattern," characterized by a set of dynamic network features and a distribution over a shared set of topics. 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

Since Latent Dirichlet Allocation (**?**), which models of words alone, a variety of useful topic models have been developed by considering additional components other than words. One direction incorporated the network aspect by using the author and receiver information such as Author-Topic Models (**?**), Author-Recipient-Topic Models and Role-Author-Recipient-Topic Models (**?**), however, all of the models treat the author and recipients as observed variable thus generative processes only involve topics and words in a document. On the other side, there have been a lot of topic models that focus on the timestamps of documents. For example, Dynamic Topic Models (**?**) introduced logistic normal topic models that rely on Markov assumptions with discretization of time, and Topics Over Time (**?**) suggested a countinuous-time model by generating a timestamp from the Beta distribution with its parameter depending on each topic assignment. However none of the above mentioned topic models are jointly dealing with the author-recipient and timestamps of the documents, although it is common in the field of dynamic network analysis to study who connects to whom and when (i, j, t).

In this paper, we develop the interaction-partitioned topic model (IPTM), a dynamic topic model which reflects the network dynamics (sender or author, recipients, and timestamp) and the contents of the document (topics and words). By assigning each communication to an "interaction pattern," IPTM connects the two separate processes, tie generating process from the Cox multiplicative intensity model and content generating process from latent Dirichlet allocation, and model the process of who communicates with whom about what, and when.

2 IPTM Model

We first introduce the multiplicative Cox intensity model in the context of tie formation process in a continuous-time textual communication network. We then illustrate the generative process of the model which incorporates the generative process of stochastic actor-oriented models and latent Dirichlet allocation. Lastly, specification of the dynamic network statistics used is demonstrated. For concreteness, we frame our discussion of this model in terms of email data, although it is generally applicable to any similarly-structured communication data.

2.1 Cox Multiplicative Intensity

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 $(i^{(d)}, J^{(d)}, t^{(d)}, W^{(d)})$. 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 $W^{(d)} = \{w_m^{(d)}\}_{n=1}^{N^{(d)}}$ is a set of tokens that comprise the text of the email. In this section, we only consider the first three , $(i^{(d)}, J^{(d)}, t^{(d)})$, and explain how we apply multiplicative intensity model to the generating process of a document (or a tie).

The interaction-partitioned topic model (IPTM) assigns each communication to an "interaction pattern," characterized by a set of dynamic network features and a distribution over a shared set of topics. Here we illustrate how a set of dynamic network features contribute uniquely identifies each interaction pattern. Assume that each interaction pattern $c \in \{1, ..., C\}$ has an $A \times A$ stochastic intensity (or hazard) matrix of $\boldsymbol{\lambda}^{(c)}(t) = \{\{\lambda_{ij}^{(c)}(t)\}_{i=1}^{A}\}_{j=1}^{A}$, where $\lambda_{ij}^{(c)}(t)$=P{for interaction pattern c, $i \to j$ occurs in time interval $[t, t+dt)$, given that it has not ben sent until 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. The detailed specifications of the dynamic network covariates are illustrated in Section **??**.

Following the multiplicative Cox model, the $(i,j)^{th}$ element of the intensity matrix $\boldsymbol{\lambda}^{(c)}(t)$ forms:

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

where λ_0 is the common baseline hazards for the overall interaction (assume that λ_0 does not depend on t), $\boldsymbol{\beta}^{(c)}$ is an unknown vector of coefficients in $\boldsymbol{R}^p$, $\boldsymbol{x}_t^{(c)}(i,j)$ is a vector of the p-dimensional dynamic network statistics for directed edge (i,j) at time t, and $\mathcal{A}_{\setminus i}$ is the predictable receiver set of sender i within the set of all possible actors $\mathcal{A}$ (no self-loop). Equivalently, by fixing $\lambda_0 = 1$, we can avoid unknown baseline hazard rate and rewrite (1) as:

$$\lambda_{ij}^{(c)}(t) = \exp\Big\{\boldsymbol{\beta}^{(c)T}\boldsymbol{x}_t^{*(c)}(i,j)\Big\} \cdot 1\{j \in \mathcal{A}_{\setminus i}\}, \tag{2}$$

where the first element of $\boldsymbol{\beta}^{(c)}$ corresponds to the deviation from λ_0, by including the intercept term and setting $\boldsymbol{x}_t^{*(c)}(i,j) = (\mathbf{1}, \boldsymbol{x}_t^{(c)}(i,j))$. 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:

$$\lambda_{iJ}^{(c)}(t) = \exp\Big\{\sum_{j \in J}\boldsymbol{\beta}^{(c)T}\boldsymbol{x}_t^{*(c)}(i,j)\Big\} \cdot \prod_{j \in J} 1\{j \in \mathcal{A}_{\setminus i}\}. \tag{3}$$

In case of single receivers ($|J| = 1$), Equation (3) is reduced to Equation (2), thus in the following sections we use the Equation (3) of multicast cases as a general form of the stochastic intensity between the sender and receivers.

2.2 Generative Process

The interaction-partitioned topic model (IPTM) is a probabilistic model of who communicates with whom about what, and when. The generative process of IPTM consists of two parts: 1) generation of the ties (i.e. 'who', 'whom', and 'when') and 2) generation of content (i.e. 'what'). The tie generating process resembles that of stochastic actor-oriented models (SAOMs) of **?**, and the content generating process directly follows latent Dirichlet allocation (LDA) of **?**. In this section, we illustrate the two generative process separately, and show how the two processes can jointly generate a document.

2.2.1 Tie Generating Process

Motivated from stochastic actor-oriented model (SAOM) of **?**, a model which assumes that the network evolves as a stochastic process 'driven by the actors', we assume the following generative process for each document d in a corpus D:

1. Choose the interaction pattern $c^{(d)} \sim \text{Multinomial}(\gamma)$
2. (Data augmentation) For each sender $i \in \{1, ..., A\}$, create a list of receivers J_i by applying the Bernoulli probabilities to every $j \in \mathcal{A}_{\setminus i}$

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

   where the probability is called a Bernoulli-Poisson (BerPo) link function (**?**) and δ is a tuning parameter to control the number of multicasts. Note that $_+$ denotes including the timepoint itself, meaning that λ_{ij} is obtained using the history of interactions until and including $t^{(d-1)}$. (i.e. $\lambda_{ij}^{(c^{(d)})}(t_+^{(d-1)}) = \exp\Big\{\boldsymbol{\beta}^{(c^{(d)})T} \boldsymbol{x}^{*(c^{(d)})}_{t_+^{(d-1)}}(i,j)\Big\} \cdot 1\{j \in \mathcal{A}_{\setminus i}\}$)
3. For each sender $i = 1, ..., A$, generate the time increments $\Delta T_{iJ_i} \sim \text{Exp}(\lambda_{iJ_i}(t_+^{(d-1)}))$, where $\lambda_{iJ_i}(t_+^{(d-1)}) = \exp\Big\{\sum_{j \in J_i} \boldsymbol{\beta}^{(c^{(d)})T} \boldsymbol{x}^{*(c^{(d)})}_{t_+^{(d-1)}}(i,j)\Big\} \cdot \prod_{j \in J_i} 1\{j \in \mathcal{A}_{\setminus i}\}$
4. Set the timestamp $t^{(d)} = t^{(d-1)} + \min(\Delta T_{iJ_i})$, $i^{(d)} = i_{\min(\Delta T_{iJ_i})}$, and $J^{(d)} = J_{i^{(d)}}$.

2.2.2 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 **?**. This concept is also very similar to the Cluster-Based Topic Modelling (**?**), in a way that every document is assigned its group $c^{(d)}$ and then each document-specific topic distribution $\boldsymbol{\theta}^{(d)}$ can instead be drawn from a group-specific $\text{Dir}(\alpha^{(c^{(d)})}, \boldsymbol{m}^{(c^{(d)})})$, thereby respecting the document groupings. We assume the following generative process for each document d in a corpus D:

1. Choose the interaction pattern $c^{(d)} \sim \text{Multinomial}(\gamma)$
2. Choose the number of words $N^{(d)} \sim \text{Poisson}(\zeta)$
3. Choose document-topic distribution $\boldsymbol{\theta}^{(d)} \sim \text{Dir}(\alpha^{(c^{(d)})}, \boldsymbol{m}^{(c^{(d)})})$
4. 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 $w_n^{(d)} \sim \text{Multinomial}(\phi^{(z_n^{(d)})})$

2.2.3 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).

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.

2. For the interaction pattern $c = 1, ..., C$, **[See Algorithm 2]**:

   (a) $\boldsymbol{\beta}^{(c)} \sim \text{Normal}(\mathbf{0}, \sigma^2 I_P)$
   - 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.

   (b) Set $\alpha^{(c)}$ and $\boldsymbol{m}^{(c)}$
   - The topic proportions for documents in the same cluster share the same parameters in the Dirichlet distribution. How we choose these parameters will be explained in Section **??**.

3. For the document $d = 1, ..., D$ **[See Algorithm 3]**:

   (a) $c^{(d)} \sim \text{Multinomial}(\boldsymbol{\gamma})$
   - Each document d is associated with one "interaction pattern" among C different types, with the parameter $\boldsymbol{\gamma}$. Here, we assign the prior for the multinomial parameter $\boldsymbol{\gamma} \sim \text{Dir}(\eta, \boldsymbol{l})$, where $\boldsymbol{l}$ is a symmetric base prior by default.

   (b) $N^{(d)} \sim \text{Poisson}(\zeta)$
   - The number of words in the document is chosen from Poisson distribution, and ζ can be either fixed or estimated.

   (c) Calculate $\boldsymbol{x}^{*(c^{(d)})}_{t_+^{(d-1)}}(i, j)$ and the corresponding $\boldsymbol{\lambda}^{(c^{(d)})}(t)$
   - The dynamic network statistics are calculated based on the documents of the same interaction pattern, using the history of interactions until (and including) the last document.

   (d) Choose $t^{(d)}$, $i^{(d)}$, and $J^{(d)}$ following Section 1.2.1. (i.e. $\mathbf{N}^{(d|c^{(d)})}(t^{(d)}) \sim \text{CP}(\boldsymbol{\lambda}^{(c^{(d)})}(t_+^{(d-1)}))$)
   - $\mathbf{N}^{(d|c^{(d)})}(t^{(d)})$ is a $A \times A$ matrix where $(i^{(d)}, j)^{th}$ $(j \in J^{(d)})$ elements are 1 and the rest are 0.

   (e) $\boldsymbol{\theta}^{(d)} \sim \text{Dir}(\alpha^{(c^{(d)})}, \boldsymbol{m}^{(c^{(d)})})$
   - Each email has a discrete distribution over topics $\boldsymbol{\theta}^{(d)}$, and the topic proportions for documents in the same cluster share the same parameters in the Dirichlet distribution.

   (f) For each of the word $n = 1, ..., N^{(d)}$:

   (f1) $z_n^{(d)} \sim \text{Multinomial}(\boldsymbol{\theta}^{(d)})$

   (f2) $w_n^{(d)} \sim \text{Multinomial}(\phi^{(z_n^{(d)})})$

Algorithm 1 Topic Word Distributions

for *k=1 to K* **do**
  draw $\boldsymbol{\phi}^{(k)} \sim \text{Dir}(\beta, \mathbf{u})$
end

Algorithm 2 Interaction Pattern-unique Parameters

for *c=1 to C* **do**
  draw $\boldsymbol{\beta}^{(c)} \sim \text{Normal}(\mathbf{0}, \sigma^2 I_P)$
  set $\alpha^{(c)}$ and $\boldsymbol{m}^{(c)}$
end

Algorithm 3 Document Generating Process

for *d=1 to D* **do**
  draw $c^{(d)} \sim \text{Multinomial}(\boldsymbol{\gamma})$
  draw $N^{(d)} \sim \text{Poisson}(\zeta)$
  draw $(t^{(d)}, i^{(d)}, J^{(d)})$ using $\mathbf{N}^{(d|c^{(d)})}(t^{(d)}) \sim \text{CP}(\boldsymbol{\lambda}^{(c^{(d)})}(t_+^{(d-1)}))$
  draw $\boldsymbol{\theta}^{(d)} \sim \text{Dir}(\alpha^{(c^{(d)})}, \boldsymbol{m}^{(c^{(d)})})$
  for *n=1 to* $N^{(d)}$ **do**
    draw $z_n^{(d)} \sim \text{Multinomial}(\boldsymbol{\theta}^{(d)})$
    draw $w_n^{(d)} \sim \text{Multinomial}(\boldsymbol{\phi}^{(z_n^{(d)})})$
  end
end

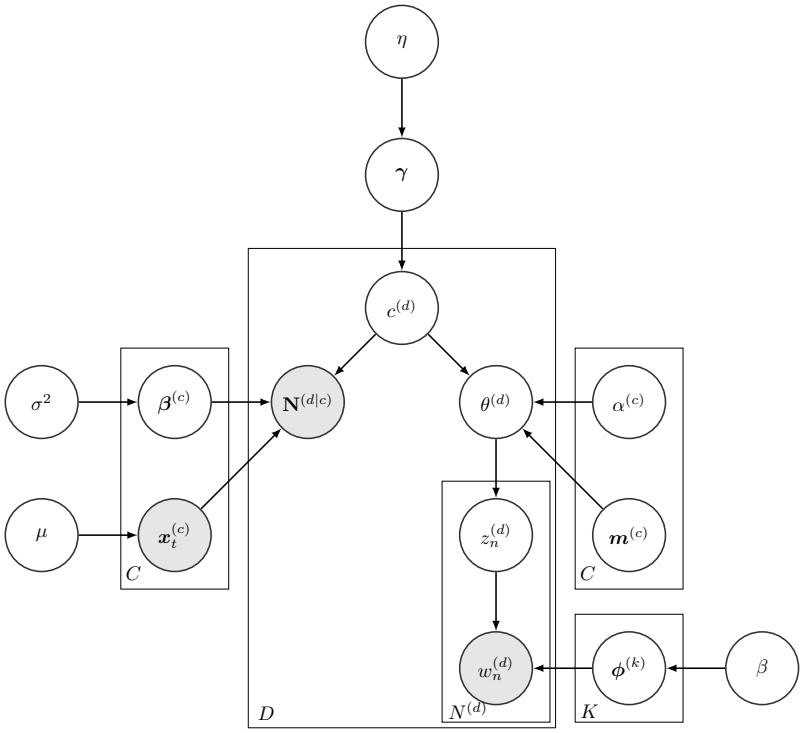

Figure 1: Plate notation of IPTM

2.3 Dynamic covariates to measure network effects

The network statistics $\boldsymbol{x}_t^{*(c)}(i,j)$ of Equation (2), 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). Since time-invariant covariates can be easily specified in various manners (e.g. homophily or group-level effects), here we only consider specification of dynamic covariates.

Following **?** (refer to Fig.3 of **?** attached 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 exactly three actors. In addition to the ones from **?**, we also include the indegree for receiver and outdegree for sender effects to measure the popularity and centrality. However, one different point from the existing specification is that we define the effects not to be based on finite sub-interval, which require large number of dimension. Instead, we create a single statistic for each effect by incorporating the recency of event into the statistic itself. As a result, all of the statistics can be seen as time-weighted dynamic network statistics.

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

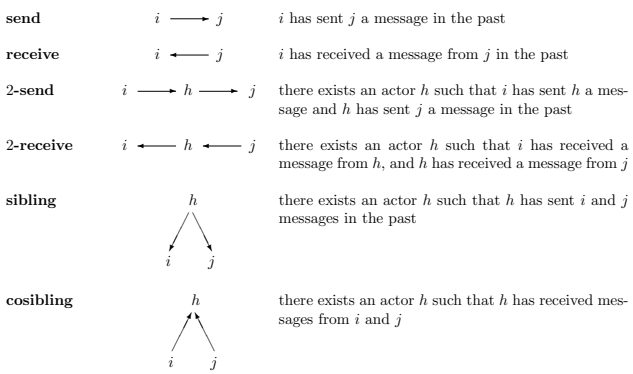

Fig. 3. Dynamic covariates to measure network effects

2. $\text{send}_t^{(c)}(i,j) = \sum_{d:c^{(d)}=c} \sum_{d:t^{(d)}<t} I\{i \to j\} \cdot g(t - t^{(d)})$
3. $\text{receive}_t^{(c)}(i,j) = \sum_{d:c^{(d)}=c} \sum_{d:t^{(d)}<t} I\{j \to i\} \cdot g(t - t^{(d)})$
4. $\text{triangle}_t^{(c)}(i,j) = \sum_{d:c^{(d)}=c} \sum_{h \neq i,j} \Big(\sum_{d:t^{(d)}<t} I\{i \to h \text{ or } h \to j\} \cdot g(t - t^{(d)}) \Big) \Big(\sum_{d:t^{(d)}<t} I\{j \to h \text{ or } h \to j\} \cdot g(t - t^{(d)}) \Big)$
5. $\text{outdegree}_t^{(c)}(i) = \sum_{d:c^{(d)}=c} \sum_{j \neq i} \sum_{d:t^{(d)}<t} I\{i \to j\} \cdot g(t - t^{(d)})$
6. $\text{indegree}_t^{(c)}(j) = \sum_{d:c^{(d)}=c} \sum_{i \neq j} \sum_{d:t^{(d)}<t} I\{j \to i\} \cdot g(t - t^{(d)})$

Here, time decaying function $g(t - t^{(d)})$ reflects the difference between current time t and the timestamp of previous email $t^{(d)}$, thus measuring the recency. Inspired by the self-exciting Hawkes process, which is often used to model the temporal effect of email data, we can take the exponential kernel $g(t - t^{(d)}) = e^{-\mu(t-t^{(d)})}$ where μ is the parameter of speed at which sender replies to emails, with larger values indicating faster response times. Indeed, μ^{-1} is the expected number of hours it takes to reply to a typical email. For simplicity, in our simulation we fixed $\mu = 0.05$ (i.e. $g(t - t^{(d)}) = e^{-0.05(t-t^{(d)})}$), but μ can also be esitimated.

3 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 email in the form of the counting process $\mathcal{N} = \{\boldsymbol{N}^{(d)}(t^{(d)})\}_{d=1}^{D}$. Next, $\mathcal{X} = \{\boldsymbol{x}_{t^{(d)}}^{(c)}(i,j)\}_{d=1}^{D}$ is the metadata, and the latent variables are $\Phi = \{\boldsymbol{\phi}^{(k)}\}_{k=1}^{K}, \Theta = \{\boldsymbol{\theta}^{(d)}\}_{d=1}^{D}, \mathcal{Z} = \{\boldsymbol{z}^{(d)}\}_{d=1}^{D}, \mathcal{C} = \{c^{(d)}\}_{d=1}^{D}$, and $\mathcal{B} = \{\boldsymbol{\beta}^{(c)}\}_{c=1}^{C}$.

Below is the the big joint distribution

$$\begin{aligned}
&P(\Phi, \Theta, \mathcal{W}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \mathcal{N} | \mathcal{X}, \beta, \boldsymbol{u}, \alpha, \boldsymbol{m}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \sigma^2) \\
&= P(\mathcal{W}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \mathcal{N} | \Phi, \Theta, \mathcal{X}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \sigma^2) P(\Phi, \Theta | \beta, \boldsymbol{u}, \alpha, \boldsymbol{m}) \\
&= P(\mathcal{W} | \mathcal{Z}, \Phi) P(\mathcal{Z} | \Theta) P(\mathcal{N} | \mathcal{C}, \mathcal{X}, \mathcal{B}) P(\mathcal{B} | \mathcal{C}, \sigma^2) P(\Phi | \beta, \boldsymbol{u}) P(\Theta | \mathcal{C}, \alpha, \boldsymbol{m}) P(\mathcal{C} | \boldsymbol{\gamma}) P(\boldsymbol{\gamma} | \boldsymbol{\eta})
\end{aligned} \tag{4}$$

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, we obtain below:

$$\propto P(\mathcal{W} | \mathcal{Z}) P(\mathcal{Z} | \mathcal{C}, \beta, \boldsymbol{u}, \alpha, \boldsymbol{m}) P(\mathcal{N} | \mathcal{C}, \mathcal{B}, \mathcal{X}) P(\mathcal{B} | \mathcal{C}, \sigma^2) P(\mathcal{C} | \boldsymbol{\gamma}) \tag{5}$$

Then, we only have to perform inference over the remaining unobserved latent variables $\mathcal{Z}, \mathcal{C}$, and $\mathcal{B}$, using the equation below:

$$P(\mathcal{Z}, \mathcal{C}, \mathcal{B} | \mathcal{W}, \mathcal{N}, \mathcal{X}, \beta, \boldsymbol{u}, \alpha, \boldsymbol{m}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \sigma^2) \propto P(\mathcal{W}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \mathcal{N} | \mathcal{X}, \beta, \boldsymbol{u}, \alpha, \boldsymbol{m}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \sigma^2) \tag{6}$$

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

3.1 Data augmentation

As mentioned earlier in Section **??**, we use data augmentation to make inference on $\mathcal{B} = \{\boldsymbol{\beta}^{(c)}\}_{c=1}^{C}$. That is, for every document d, we generate the latent sender-receivers (i, J_i) for all $i \in \mathcal{A}_{\setminus i}$ and plug the generated sets into the likelihood function of $\boldsymbol{\beta}^{(c)}$. Note that in latent data generating step, the sets (i, J_i) should account for the probability $P(\text{latent receiver}|\text{latent time} < \text{observed time})$. This probability can be derived fairly easily using Bayes' theorem:

Let $t^{(d)}$ be the observed time, t_i be the latent time associated with the latent email sent by i, and $i \to j$ be an indicator of whether sender i added j to the latent email receivers (i.e. $j \in J_i$).

$$\begin{aligned} P(i \to j | \Delta T_{iJ_i} > \Delta T_{i^{(d)}J^{(d)}}) &= \frac{P(\Delta T_{iJ_i} > \Delta T_{i^{(d)}J^{(d)}} | i \to j) P(i \to j)}{P(\Delta T_{iJ_i} > t^{(d)})} \\ &= \frac{P(\Delta T_{iJ_i} > \Delta T_{i^{(d)}J^{(d)}} | i \to j) P(i \to j)}{P(\Delta T_{iJ_i} > \Delta T_{i^{(d)}J^{(d)}} | i \to j) P(i \to j) + P(\Delta T_{iJ_i} > \Delta T_{i^{(d)}J^{(d)}} | i \not\to j) P(i \not\to j)} \end{aligned} \tag{7}$$

where $P(\Delta T_{iJ_i} > \Delta T_{i^{(d)}J^{(d)}} | i \to j)$ and $P(\Delta T_{iJ_i} > \Delta T_{i^{(d)}J^{(d)}} | i \not\to j)$ are given by the cumulative exponential distributions associated with the respective $\lambda_{iJ_i}(t_+^{(d-1)})$ and $P(i \to j)$ and $P(i \not\to j)$ are given by the Bernoulli transformation in Section **??**. Therefore, we determine the latent ties $I(i \to j)$ from Bernoulli distribution of probability:

$$\frac{e^{-(t^{(d)} - t^{(d-1)})\lambda^{j}_{iJ_i}(t_+^{(d-1)})}(1 - e^{-\delta\lambda_{ij}^{(c^{(d)})}(t_+^{(d-1)})})}{e^{-(t^{(d)} - t^{(d-1)})\lambda^{j}_{iJ_i}(t_+^{(d-1)})}(1 - e^{-\delta\lambda_{ij}^{(c^{(d)})}(t_+^{(d-1)})}) + e^{-(t^{(d)} - t^{(d-1)})\lambda^{-j}_{iJ_i}(t_+^{(d-1)})}(e^{-\delta\lambda_{ij}^{(c^{(d)})}(t_+^{(d-1)})})}, \tag{8}$$

QUESTION: We need baseline λ_{iJ_i} before we add j into the recipient set.....
where $\lambda^{j}_{iJ_i}$ is calculated from the receiver set of i that includes j, while $\lambda^{-j}_{iJ_i}$ is calculated from the receiver set of i that excludes j.

Perhaps Bayes' Theorem is not the correct approach given the rest of the recipient set for sender j. Let $R_i^{(-j)}$ denote the vector of recipient indicators other than r_{ij}. Then what we are looking for is

$$P(r_{ij} = 1 | \Delta T_{iJ_i} > \Delta T_{i^{(d)}J^{(d)}} \cap R_i^{(-j)}) = \frac{P(r_{ij} = 1 \cap \Delta T_{iJ_i} > \Delta T_{i^{(d)}J^{(d)}} \cap R_i^{(-j)})}{P(\Delta T_{iJ_i} > \Delta T_{i^{(d)}J^{(d)}} \cap R_i^{(-j)})}.$$

Both the numerator and denominator are tractable if we rewrite as

$$\frac{P(\Delta T_{iJ_i} > \Delta T_{i^{(d)}J^{(d)}} | r_{ij} = 1 \cap R_i^{(-j)}) P(r_{ij} = 1 \cap R_i^{(-j)})}{P(\Delta T_{iJ_i} > \Delta T_{i^{(d)}J^{(d)}} \cap R_i^{(-j)} | r_{ij} = 1) P(r_{ij} = 1) + P(\Delta T_{iJ_i} > \Delta T_{i^{(d)}J^{(d)}} \cap R_i^{(-j)} | r_{ij} = 0) P(r_{ij} = 0)}.$$

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

$$\begin{aligned} \mathcal{L}(\boldsymbol{\beta}^{(c^{(d)})}) &= P\Big(\Delta T_{i^{(d)}J^{(d)}} \sim \text{Exp}(\lambda_{i^{(d)}J^{(d)}}(t_+^{(d-1)}))\Big) \times P\Big(\min(\Delta T_{iJ_i}) = \Delta T_{i^{(d)}J^{(d)}}\Big) \times P\Big(\text{Edge creations}\Big) \\ &= \Big(\lambda_{i^{(d)}J^{(d)}}(t_+^{(d-1)}) e^{-(t^{(d)} - t^{(d-1)})\lambda_{i^{(d)}J^{(d)}}(t_+^{(d-1)})}\Big) \times \Big(e^{-(t^{(d)} - t^{(d-1)}) \sum\limits_{i \neq i^{(d)}} \lambda_{iJ_i}(t_+^{(d-1)})}\Big) \\ &\quad \times \Big(\prod_{i \in \mathcal{A}} \prod_{j \in \mathcal{A}_{\setminus i}} (1 - e^{-\delta\lambda_{ij}(t_+^{(d-1)})})^{I(j \in J_i)} (e^{-\delta\lambda_{ij}(t_+^{(d-1)})})^{1 - I(j \in J_i)}\Big) \\ &= \Big(\lambda_{i^{(d)}J^{(d)}}(t_+^{(d-1)}) e^{-(t^{(d)} - t^{(d-1)}) \sum\limits_{i \in \mathcal{A}} \lambda_{iJ_i}(t_+^{(d-1)})}\Big) \times \Big(\prod_{i \in \mathcal{A}} \prod_{j \in \mathcal{A}_{\setminus i}} (e^{\delta\lambda_{ij}(t_+^{(d-1)})} - 1)^{I(j \in J_i)} e^{-\delta\lambda_{ij}(t_+^{(d-1)})}\Big), \end{aligned} \tag{9}$$

Applying this to the every document in the corpus and taking the log, we obtain the full likelihood function

used for the inference of $\{\boldsymbol{\beta}^{(c)}\}_{c=1}^{C}$ as:

$$\begin{aligned}\ell(\boldsymbol{\beta}^{(c^{(d)})}) = \sum_{d=1}^{D} \Big\{ \sum_{j \in J_{i^{(d)}}} \boldsymbol{\beta}^{(c^{(d)})T} \boldsymbol{x}^{*(c^{(d)})}_{t_+^{(d-1)}}(i^{(d)}, j) - (t^{(d)} - t^{(d-1)}) \sum_{i \in \mathcal{A}} e^{\sum_{j \in J_i} \boldsymbol{\beta}^{(c^{(d)})T} \boldsymbol{x}^{*(c^{(d)})}_{t_+^{(d-1)}}(i,j)} \\ + \sum_{i \in \mathcal{A}} \sum_{j \in \mathcal{A}_{\setminus i}} \Big(I(j \in J_i) \cdot \log(e^{\delta e^{\boldsymbol{\beta}^{(c^{(d)})T} \boldsymbol{x}^{*(c^{(d)})}_{t_+^{(d-1)}}(i,j)}} - 1) - \delta e^{\boldsymbol{\beta}^{(c^{(d)})T} \boldsymbol{x}^{*(c^{(d)})}_{t_+^{(d-1)}}(i,j)} \Big) \Big\},\end{aligned} \tag{10}$$

and since Equation (10) above is the joint likelihood function of the sender, receivers and timestamp $(i^{(d)}, J^{(d)}, t^{(d)})$, it will be used for $P(\mathbf{N}^{(d)}(t^{(d)})|c^{(d)} = c, \mathcal{C}_{\setminus d}, \mathcal{B}, \mathcal{N}_{\setminus d}, \mathcal{X})$ in the next following steps.

3.2 Resampling $\mathcal{C}$

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

$$\begin{aligned}
&P(c^{(d)} = c|\mathcal{W}, \mathcal{Z}, \mathcal{C}_{\setminus d}, \mathcal{B}, \mathcal{N}, \mathcal{X}, \beta, \boldsymbol{u}, \alpha, \boldsymbol{m}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \sigma^2) \\
&\propto P(c^{(d)} = c, \boldsymbol{w}^{(d)}, \boldsymbol{z}^{(d)}, \mathbf{N}^{(d)}(t^{(d)})|\mathcal{W}_{\setminus d}, \mathcal{Z}_{\setminus d}, \mathcal{C}_{\setminus d}, \mathcal{B}, \mathcal{N}_{\setminus d}, \mathcal{X}, \beta, \boldsymbol{u}, \alpha, \boldsymbol{m}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \sigma^2) \\
&\propto P(c^{(d)} = c|\mathcal{C}_{\setminus d}, \boldsymbol{\gamma}) P(\mathbf{N}^{(d)}(t^{(d)})|c^{(d)} = c, \mathcal{C}_{\setminus d}, \mathcal{B}, \mathcal{N}_{\setminus d}, \mathcal{X}) P(\boldsymbol{w}^{(d)}, \boldsymbol{z}^{(d)}|c^{(d)} = c, \mathcal{W}_{\setminus d}, \mathcal{Z}_{\setminus d}, \mathcal{C}_{\setminus d}, \beta, \boldsymbol{u}, \alpha, \boldsymbol{m}),
\end{aligned} \tag{11}$$

where $P(c^{(d)} = c|\mathcal{C}_{\setminus d}, \boldsymbol{\gamma})$ comes from the multinomial prior γ and $P(\mathbf{N}^{(d)}(t^{(d)})|c^{(d)} = c, \mathcal{C}_{\setminus d}, \mathcal{B}, \mathcal{N}_{\setminus d}, \mathcal{X})$ 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. We will replace this by the partial likelihood in Equation (4) (without the product term since resampling of c is document-specific). For the last term $P(\boldsymbol{w}^{(d)}, \boldsymbol{z}^{(d)}|c^{(d)} = c, \mathcal{W}_{\setminus d}, \mathcal{Z}_{\setminus d}, \mathcal{C}_{\setminus d}, \beta, \boldsymbol{u}, \alpha, \boldsymbol{m})$, we will follow typical LDA approach.

Using Bayes' theorem (See APPENDIX C for conditional probabilty of the last term), we have

$$\begin{aligned}
&= \Big[\gamma_c\Big] \times \Big[\Big(\lambda_{i^{(d)}J^{(d)}}(t_+^{(d-1)}) e^{-(t^{(d)} - t^{(d-1)}) \sum_{i \in \mathcal{A}} \lambda_{iJ_i}(t_+^{(d-1)})}\Big) \times \Big(\prod_{i \in \mathcal{A}} \prod_{j \in \mathcal{A}_{\setminus i}} (e^{\delta \lambda_{ij}(t_+^{(d-1)})} - 1)^{I(j \in J_i)} e^{-\delta \lambda_{ij}(t_+^{(d-1)})}\Big)\Big] \\
&\qquad \times \Big[\prod_{n=1}^{N^{(d)}} \frac{N_{z_n^{(d)}|d,\setminus d,n} + \alpha^{(c)} \boldsymbol{m}^{(c)}_{z_n^{(d)}}}{N_{\cdot|d,\setminus d,n} + \alpha^{(c)}}\Big],
\end{aligned} \tag{12}$$

where $N_{k|d}$ is the number of times topic k shows up in the document d. Furthermore, we can take the log of Equation (10) to avoid numerical issue from exponentiation and increase the speed of computation, which becomes:

$$\begin{aligned}
&\Big[\log(\gamma_c)\Big] + \Big[\sum_{j \in J_{i^{(d)}}} \boldsymbol{\beta}^{(c^{(d)})T} \boldsymbol{x}^{*(c^{(d)})}_{t_+^{(d-1)}}(i^{(d)}, j) - (t^{(d)} - t^{(d-1)}) \sum_{i \in \mathcal{A}} e^{\sum_{j \in J_i} \boldsymbol{\beta}^{(c^{(d)})T} \boldsymbol{x}^{*(c^{(d)})}_{t_+^{(d-1)}}(i,j)} \\
&\quad + \sum_{i \in \mathcal{A}} \sum_{j \in \mathcal{A}_{\setminus i}} \Big(I(j \in J_i) \cdot \log(e^{\delta e^{\boldsymbol{\beta}^{(c^{(d)})T} \boldsymbol{x}^{*(c^{(d)})}_{t_+^{(d-1)}}(i,j)}} - 1) - \delta e^{\boldsymbol{\beta}^{(c^{(d)})T} \boldsymbol{x}^{*(c^{(d)})}_{t_+^{(d-1)}}(i,j)}\Big)\Big] \\
&\quad + \Big[\sum_{n=1}^{N^{(d)}} \log\Big(N_{z_n^{(d)}|d,\setminus d,n} + \alpha^{(c)} \boldsymbol{m}^{(c)}_{z_n^{(d)}}\Big) - \log\Big(N_{\cdot|d,\setminus d,n} + \alpha^{(c)}\Big)\Big].
\end{aligned} \tag{13}$$

3.3 Resampling $\mathcal{Z}$

Next, the new values of $z_m^{(d)}$ are sampled for all of the token topic assignments (one token at a time), using the conditional posterior probability of being topic k as we derived in APPENDIX C:

$$\begin{aligned}
&P(z_n^{(d)} = k|\mathcal{W}, \mathcal{Z}_{\setminus d,n}, \mathcal{C}, \mathcal{B}, \mathcal{N}, \mathcal{X}, \beta, \boldsymbol{u}, \alpha, \boldsymbol{m}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \sigma^2) \\
&\propto P(z_n^{(d)} = k, w_n^{(d)}|\mathcal{W}_{\setminus d,n}, \mathcal{Z}_{\setminus d,m}, C, \beta, \boldsymbol{u}, \alpha, \boldsymbol{m})
\end{aligned} \tag{14}$$

where the subscript "$\setminus d, n$" denotes the exclsuion of position n in d^{th} email. In the last line of equation (13), it is the contribution of LDA, so we can write the conditional probability:

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

which is the well-known form of collapsed Gibbs sampling equation for LDA. After taking the log, we finally obtain:

$$\log(N_{k|d,\setminus d,n} + \alpha^{(c^{(d)})} \boldsymbol{m}_k^{(c^{(d)})}) + \log(N^{WK}_{w_n^{(d)}k,\setminus d,n} + \beta u_w) - \log(\sum_{w=1}^{W} N^{WK}_{wk,\setminus d,n} + \beta). \tag{16}$$

3.4 Resampling $\mathcal{B}$

Finally, we wan to update the interaction pattern parameter $\boldsymbol{\beta}^{(c)}$, one interaction pattern at a time. For this, we will 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:

$$\text{Acceptance Probability} = \begin{cases} \frac{P(\mathcal{B}'|\mathcal{W},\mathcal{Z},\mathcal{C},\mathcal{N},\mathcal{X})}{P(\mathcal{B}|\mathcal{W},\mathcal{Z},\mathcal{C},\mathcal{N},\mathcal{X})} & \text{if} < 1 \\ 1 \quad \text{else} \end{cases} \tag{17}$$

After factorization, we get

$$\begin{aligned} \frac{P(\mathcal{B}'|\mathcal{W},\mathcal{Z},\mathcal{C},\mathcal{N},\mathcal{X})}{P(\mathcal{B}|\mathcal{W},\mathcal{Z},\mathcal{C},\mathcal{N},\mathcal{X})} &= \frac{P(\mathcal{N}|\mathcal{B}',\mathcal{W},\mathcal{Z},\mathcal{C},\mathcal{X})P(\mathcal{B}')}{P(\mathcal{N}|\mathcal{B},\mathcal{W},\mathcal{Z},\mathcal{C},\mathcal{X})P(\mathcal{B})} \\ &= \frac{P(\mathcal{N}|\mathcal{C},\mathcal{X},\mathcal{B}')P(\mathcal{B}')}{P(\mathcal{N}|\mathcal{C},\mathcal{X},\mathcal{B})P(\mathcal{B})}, \end{aligned} \tag{18}$$

where $P(\mathcal{N}|\mathcal{C},\mathcal{X},\mathcal{B})$ is the partial likelihood in Equation (4).

For $P(\mathcal{B})$, we select a multivarate Gaussian priors as mentioned earlier. Similar to what we did in Section **??**, we can take the log and obtain the log of acceptance ratio as following:

$$\begin{aligned} &\log\Big(\phi_d(\boldsymbol{\beta}'^{(c)};\mathbf{0},\sigma^2 I_P)\Big) - \log\Big(\phi_d(\boldsymbol{\beta}^{(c)};\mathbf{0},\sigma^2 I_P)\Big) \\ &+ \sum_{d:c^{(d)}=c} \Big[\sum_{j\in J_{i^{(d)}}} \boldsymbol{\beta}'^{(c^{(d)})T} \boldsymbol{x}^{*(c^{(d)})}_{t_+^{(d-1)}}(i^{(d)},j) - (t^{(d)}-t^{(d-1)}) \sum_{i\in\mathcal{A}} e^{\sum_{j\in J_i} \boldsymbol{\beta}'^{(c^{(d)})T} \boldsymbol{x}^{*(c^{(d)})}_{t_+^{(d-1)}}(i,j)} \\ &+ \sum_{i\in\mathcal{A}} \sum_{j\in\mathcal{A}_{\setminus i}} \Big(I(j\in J_i)\cdot \log(e^{\delta e^{\boldsymbol{\beta}'^{(c^{(d)})T} \boldsymbol{x}^{*(c^{(d)})}_{t_+^{(d-1)}}(i,j)}} - 1) - \delta e^{\boldsymbol{\beta}'^{(c^{(d)})T} \boldsymbol{x}^{*(c^{(d)})}_{t_+^{(d-1)}}(i,j)} \Big) \Big] \\ &- \sum_{d:c^{(d)}=c} \Big[\sum_{j\in J_{i^{(d)}}} \boldsymbol{\beta}^{(c^{(d)})T} \boldsymbol{x}^{*(c^{(d)})}_{t_+^{(d-1)}}(i^{(d)},j) - (t^{(d)}-t^{(d-1)}) \sum_{i\in\mathcal{A}} e^{\sum_{j\in J_i} \boldsymbol{\beta}^{(c^{(d)})T} \boldsymbol{x}^{*(c^{(d)})}_{t_+^{(d-1)}}(i,j)} \\ &+ \sum_{i\in\mathcal{A}} \sum_{j\in\mathcal{A}_{\setminus i}} \Big(I(j\in J_i)\cdot \log(e^{\delta e^{\boldsymbol{\beta}^{(c^{(d)})T} \boldsymbol{x}^{*(c^{(d)})}_{t_+^{(d-1)}}(i,j)}} - 1) - \delta e^{\boldsymbol{\beta}^{(c^{(d)})T} \boldsymbol{x}^{*(c^{(d)})}_{t_+^{(d-1)}}(i,j)} \Big) \Big], \end{aligned} \tag{19}$$

where $\phi_d(\cdot;\mu,\Sigma)$ is the d-dimensional multivariate normal density. Then the log of acceptance ratio we have is:

$$\log(\text{Acceptance Probability}) = \min((18),\, 0) \tag{20}$$

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).

3.5 Asymmetric Dirichlet prior over Θ

? demonstrated that the typical implementations of topic models using symmetric Dirichlet priors with fixed concentration parameters often result in less practical results, and the model fitting can be improved by applying an asymmetric Dirichlet prior over the document–topic distributions (i.e. Θ). Therefore, we assign an asymmetric Dirichlet prior over the interaction pattern-topic distributions, $\Theta = \{\boldsymbol{\theta}^{(d)}\}_{d=1}^{D}$, where $\boldsymbol{\theta}^{(d)}$ is drawn from $\text{Dir}(\alpha^{(c^{(d)})}, \boldsymbol{m}^{(c^{(d)})})$. While **?** illustrates two different methods, adding a hierarchy to Θ and optimizing the hyperparameters (α and $\boldsymbol{m}$), we choose to use hyperparameter optimization steps since it is computationally efficient and also sufficient to achieve the desired performance gains. Now, we assume $\boldsymbol{m}^{(c)}$ to be non-uniform base measures (while $\alpha^{(c)}$ is still a fixed concentration parameter), and implement the hyperparameter optimization technique called "new fixed-point iterations using the Digamma recurrence relation" in **?** based on Minka's fixed-point iteration (**?**).

Here we summarize Chapter 2 of **?** and its extension to our IPTM, to illustrate the basic steps and equations

used for our optimization. Basically, we want to find the optimal hyperparameter $[\alpha \boldsymbol{m}]^*$ given the data $\mathcal{D}$ such that the probability of the data given the hyperparameters $P(\mathcal{D}|\alpha \boldsymbol{m})$ is maximized at $[\alpha \boldsymbol{m}]^*$. After incorporating the interaction pattern component, the evidence is now given by

$$P(\mathcal{D}^{(c)}|\alpha^{(c)}\boldsymbol{m}^{(c)}) = \prod_{d:c^{(d)}=c} \frac{\Gamma(\alpha^{(c)})}{\Gamma(N_{\cdot|d} + \alpha^{(c)})} \prod_{k=1}^{K} \frac{\Gamma(N_{k|d} + \alpha^{(c)} m_k^{(c)})}{\Gamma(\alpha^{(c)} m_k^{(c)})} \tag{21}$$

and is concave in $\alpha^{(c)}\boldsymbol{m}^{(c)}$, thus we will estimate $[\alpha^{(c)}\boldsymbol{m}^{(c)}]^*$ within each outer runs of MCMC.

First, the starting point is derived by Minka's fixed-point iteration which takes the derivative of the lower bound $B([\alpha^{(c)}\boldsymbol{m}^{(c)}]^*)$ of $\log P(\mathcal{D}^{(c)}|[\alpha^{(c)}\boldsymbol{m}^{(c)}]^*)$ with respect to $[\alpha^{(c)} m_k^{(c)}]^*$:

$$[\alpha^{(c)} m_k^{(c)}]^* = \alpha^{(c)} m_k^{(c)} \frac{\sum_{d:c^{(d)}=c} \Psi(N_{k|d} + \alpha^{(c)} m_k^{(c)}) - \Psi(\alpha^{(c)} m_k^{(c)})}{\sum_{d:c^{(d)}=c} \Psi(N_{\cdot|d} + \alpha^{(c)}) - \Psi(\alpha^{(c)})}, \tag{22}$$

where $\Psi(\cdot)$ is the first derivative of the log gamma function, known as the digamma function, and the quantity $N_{k|d}$ is the number of times that outcome k was observed in the document d. Moreover, the quantity $N_{\cdot|d} = \sum_{k=1}^{K} N_{k|d}$ is the total number of words in the document d. The value $\alpha^{(c)} m_k^{(c)}$ acts as an initial "pseudocount" for outcome k across the documents of interaction pattern c.

Next, Wallach's new method rewrites the equation above using the notation $C_k(n) = \sum_{d:c^{(d)}=c} \beta(N_{k|d} - n)$ and $C_{\cdot}(n) = \sum_{d:c^{(d)}=c} \beta(N_{\cdot|d} - n)$:

$$[\alpha^{(c)} m_k^{(c)}]^* = \alpha^{(c)} m_k^{(c)} \frac{\sum_{n=1}^{\max_d N_{k|d}} C_k(n)[\Psi(n + \alpha^{(c)} m_k^{(c)}) - \Psi(\alpha^{(c)} m_k^{(c)})]}{\sum_{n=1}^{\max_d N_{\cdot|d}} C_{\cdot}(n)[\Psi(n + \alpha^{(c)}) - \Psi(\alpha^{(c)})]}. \tag{23}$$

Finally, applying the digamma recurrence relation (for any positive integer n)

$$\Psi(n + z) - \Psi(z) = \sum_{f=1}^{n} \frac{1}{f - 1 + z},$$

we subtitute Equation (20) for below:

$$[\alpha^{(c)} m_k^{(c)}]^* = \alpha^{(c)} m_k^{(c)} \frac{\sum_{n=1}^{\max_d N_{k|d}} C_k(n) \sum_{f=1}^{n} \frac{1}{f-1+\alpha^{(c)} m_k^{(c)}}}{\sum_{n=1}^{\max_d N_{\cdot|d}} C_{\cdot}(n) \sum_{f=1}^{n} \frac{1}{f-1+\alpha^{(c)}}}. \tag{24}$$

This method is as accurate as Mika's fixed-point iteration method, but it acheives computational efficiency since the digamma recurrence relation reduces the number of new calculations required for each successive n to one. Pseudocode for the complete fixed-point iteration is given in algorithm 2.2 of **?**.

3.6 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 (**?**), 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 4 MCMC($I, n_1, n_2, n_3, \beta_B$)

set initial values $\mathcal{C}^{(0)}, \mathcal{Z}^{(0)}$, and $\mathcal{B}^{(0)}$
for *i=1 to I* **do**
  optimize $\alpha^{(c)}$ and $\boldsymbol{m}^{(c)}$ using hyperparameter optimization in Section **??**, for $c = 1, ..., C$
  for *d=1 to D* **do**
    sample the augmented data (i, J_i) for $i \in \mathcal{A}_{\backslash i^{(d)}}$ (See Section **??**)
  end
  for *n=1 to* n_1 **do**
    fix $\mathcal{Z} = \mathcal{Z}^{(i-1)}$ and $\mathcal{B} = \mathcal{B}^{(i-1)}$
    for *d=1 to D* **do**
      calculate $\boldsymbol{x}^{*(c)}_{t^{(d)}}(i^{(d)}, j)$ according to Section **??**, for every $c = 1, ..., C$
      calculate $p^{\mathcal{C}}|\boldsymbol{z}^{(d)}, \boldsymbol{\beta}^{(c^{(d)})} = (p_1, ..., p_C)$, where $p_c = \exp$(Eq. (11) corresponding to c)
      draw $c^{(d)} \sim$ multinomial$(p^{\mathcal{C}})$
    end
  end
  for *n=1 to* n_2 **do**
    fix $\mathcal{C} = \mathcal{C}^{(i)}$ and $\mathcal{B} = \mathcal{B}^{(i-1)}$
    for *d=1 to D* **do**
      for *n=1 to* $N^{(d)}$ **do**
        calculate $p^{\mathcal{Z}}|\boldsymbol{c}^{(d)}, \alpha^{(c^{(d)})}, \boldsymbol{m}^{(c^{(d)})}, \boldsymbol{\beta}^{(c^{(d)})} = (p_1, ..., p_K)$, where $p_k =$ (Eq. (13) corresponding to k)
        draw of $z_n^{(d)} \sim$ multinomial$(p^{\mathcal{Z}})$
      end
    end
  end
  for *n=1 to* n_3 **do**
    fix $\mathcal{C} = \mathcal{C}^{(i)}$, $\mathcal{Z} = \mathcal{Z}^{(i)}$, and $\mathcal{B}^{(0)} =$ last value (n_3^{th}) of $\mathcal{B}^{(i-1)}$
    calculate $\mathcal{X} = \{\boldsymbol{x}^{*(c)}_{t^{(d)}}(i, j)\}_{d=1}^{D}$ according to Section **??** given fixed $\mathcal{C}$
    for *c=1 to C* **do**
      draw $\boldsymbol{\beta}^{(c)}|\mathcal{C}, \mathcal{Z}, \mathcal{B}^{(n-1)}$ using M-H algorithm in Section **??**
      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 $\mathcal{C}$, the last value of $\mathcal{Z}$, and the last n_3 length chain of $\mathcal{B}$

4 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.

4.1 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 **??**. 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 pro-

cess 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)

4.2 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

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 interaction pattern and topic	N^{CK}	Scalar
Number of words assigned to word and topic	N^{WK}	Scalar
Interaction pattern of the d^{th} document	$c^{(d)}$	Scalar
Time of the d^{th} document	$t^{(d)}$	Scalar
Tuning parameter in tie generative process	δ	Scalar
Time decay parameter	μ	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 given interaction pattern c	$\alpha^{(c)}$	Scalar
Dirichlet base prior given interaction pattern c	$\boldsymbol{m}^{(c)}$	K-dimensional vector
Dirichlet concentration prior	β	Scalar
Dirichlet base prior	$\boldsymbol{u}$	W-dimensional vector
Dirichlet concentration prior	η	Scalar
Dirichlet base prior	$\boldsymbol{l}$	C-dimensional vector
Multinomial prior	γ	C-dimensional vector
Variance of Normal prior	σ^2	Scalar
Probabilities of the words given topics	Φ	$W \times K$ matrix
Probabilities of the words given topic k	$\boldsymbol{\phi}^{(k)}$	W-dimensional vector
Probabilities of the topics	Θ	$K \times D$ matrix
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 directed edge (i,j) given interaction pattern c	$\boldsymbol{x}_t^{(c)}(i,j)$	p-dimensional vector
Counting process in the d^{th} document given interaction pattern	$\mathbf{N}^{(d\|c)}(t)$	$A \times A$ matrix

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

APPENDIX B: Deriving the sampling equations for IPTM

$$
\begin{aligned}
&P(\Phi,\Theta,\mathcal{W},\mathcal{Z},\mathcal{C},\mathcal{B},\mathcal{N}|\mathcal{X},\beta,\boldsymbol{u},\alpha,\boldsymbol{m},\boldsymbol{\gamma},\boldsymbol{\eta},\sigma^2)\\
&= P(\mathcal{W},\mathcal{Z},\mathcal{C},\mathcal{B},\mathcal{N}|\Phi,\Theta,\mathcal{X},\boldsymbol{\gamma},\boldsymbol{\eta},\sigma^2)P(\Phi,\Theta|\beta,\boldsymbol{u},\alpha,\boldsymbol{m})\\
&= P(\mathcal{W}|\mathcal{Z},\Phi)P(\mathcal{Z}|\Theta)P(\mathcal{N}|\mathcal{C},\mathcal{B},\mathcal{X})P(\mathcal{B}|\mathcal{C},\sigma^2)P(\Phi|\beta,\boldsymbol{u})P(\Theta|\mathcal{C},\alpha,\boldsymbol{m})P(\mathcal{C}|\boldsymbol{\gamma})P(\boldsymbol{\gamma}|\boldsymbol{\eta})\\
&= \Big[\prod_{d=1}^{D}\prod_{n=1}^{N^{(d)}} P(w_n^{(d)}|\phi_{z_n^{(d)}})\Big] \times \Big[\prod_{d=1}^{D}\prod_{n=1}^{N^{(d)}} P(z_n^{(d)}|\boldsymbol{\theta}^{(d)})\Big] \times \Big[\prod_{d=1}^{D} P(\mathbf{N}^{(d)}(t^{(d)})|c^{(d)},\boldsymbol{x}^{(c^{(d)})}(t^{(d)}),\boldsymbol{\beta}^{(c)})\Big]\\
&\quad \times \Big[\prod_{c=1}^{C} P(\boldsymbol{\beta}^{(c)}|\sigma^2)\Big] \times \Big[\prod_{k=1}^{K} P(\boldsymbol{\phi}^{(k)}|\beta,\boldsymbol{u})\Big] \times \Big[\prod_{d=1}^{D} P(\boldsymbol{\theta}^{(d)}|\alpha^{(c^{(d)})},\boldsymbol{m}^{(c^{(d)})})\Big] \times \Big[\prod_{d=1}^{D} P(c^{(d)}|\boldsymbol{\gamma})\Big]\\
&\quad \times P(\boldsymbol{\gamma}|\boldsymbol{\eta})
\end{aligned}
\tag{25}
$$

Since $P(\boldsymbol{\beta}^{(c)}|\sigma^2)$ is Normal$(\mathbf{0}, \sigma^2)$ and $P(\boldsymbol{\gamma}|\boldsymbol{\eta})$ is Dirichlet$(\boldsymbol{\eta})$, we can drop the two terms out and 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_{d=1}^{D} P(\mathbf{N}^{(d)}(t^{(d)})|c^{(d)}, \boldsymbol{x}^{(c^{(d)})}(t^{(d)}), \boldsymbol{\beta}^{(c)})\Big] \\
&\quad \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^{(c^{(d)})}, \boldsymbol{m}^{(c^{(d)})})\Big] \times \Big[\prod_{d=1}^{D} P(c^{(d)}|\boldsymbol{\gamma})\Big] \\
&= \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_{d=1}^{D}\Big(\lambda_{i^{(d)}J^{(d)}}(t_+^{(d-1)}) e^{-(t^{(d)}-t^{(d-1)})\sum\limits_{i\in\mathcal{A}}\lambda_{iJ_i}(t_+^{(d-1)})}\Big) \times \Big(\prod_{i\in\mathcal{A}}\prod_{j\in\mathcal{A}_{\setminus i}} (e^{\delta\lambda_{ij}(t_+^{(d-1)})}-1)^{I(j\in J_i)} e^{-\delta\lambda_{ij}(t_+^{(d-1)})}\Big)\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^{(c^{(d)})}m_k^{(c^{(d)})})}{\prod_{k=1}^{K}\Gamma(\alpha^{(c^{(d)})}m_k^{(c^{(d)})})}\prod_{k=1}^{K}(\boldsymbol{\theta}_k^{(d)})^{\alpha^{(c^{(d)})}m_k^{(c^{(d)})}-1}\Big)\Big] \\
&\quad \times \Big[\prod_{d=1}^{D}\gamma_c^{I(c^{(d)}=c)}\Big] \\
&= \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^{(c^{(d)})}m_k^{(c^{(d)})})}{\prod_{k=1}^{K}\Gamma(\alpha^{(c^{(d)})}m_k^{(c^{(d)})})}\Big] \\
&\quad \times \Big[\prod_{d=1}^{D}\Big(\lambda_{i^{(d)}J^{(d)}}(t_+^{(d-1)}) e^{-(t^{(d)}-t^{(d-1)})\sum\limits_{i\in\mathcal{A}}\lambda_{iJ_i}(t_+^{(d-1)})}\Big) \times \Big(\prod_{i\in\mathcal{A}}\prod_{j\in\mathcal{A}_{\setminus i}} (e^{\delta\lambda_{ij}(t_+^{(d-1)})}-1)^{I(j\in J_i)} e^{-\delta\lambda_{ij}(t_+^{(d-1)})}\Big)\Big] \\
&\quad \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^{(c^{(d)})}m_k^{(c^{(d)})}-1}\Big] \times \Big[\prod_{d=1}^{D}\gamma_{c^{(d)}}\Big]
\end{aligned}
\tag{26}
$$

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:

$$
\begin{aligned}
&P(\mathcal{W}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \mathcal{N}|\mathcal{X}, \beta, \boldsymbol{u}, \alpha, \boldsymbol{m}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \sigma^2) \\
&= \text{Const.}\int_{\Theta}\int_{\Phi}\Big[\prod_{k=1}^{K}\prod_{w=1}^{W}\phi_{wk}^{N_{wk}^{WK}+\beta u_w-1}\Big]\Big[\prod_{d=1}^{D}\prod_{k=1}^{K}(\boldsymbol{\theta}_k^{(d)})^{N_{k|d}+\alpha^{(c^{(d)})}m_k^{(c^{(d)})}-1}\Big]d\Phi d\Theta \\
&= \text{Const.}\Big[\prod_{k=1}^{K}\int_{\phi_{:k}}\prod_{w=1}^{W}\phi_{wk}^{N_{wk}^{WK}+\beta u_w-1}d\phi_{:k}\Big] \times \Big[\prod_{d=1}^{D}\int_{\theta_{:d}}\prod_{k=1}^{K}(\boldsymbol{\theta}_k^{(d)})^{N_{k|d}+\alpha^{(c^{(d)})}m_k^{(c^{(d)})}-1}d\theta_{:d}\Big] \\
&= \text{Const.}\Big[\prod_{k=1}^{K}\frac{\prod_{w=1}^{W}\Gamma(N_{wk}^{WK}+\beta u_w)}{\Gamma(\sum_{w=1}^{W}N_{wk}^{WK}+\beta)}\Big] \times \Big[\prod_{d=1}^{D}\frac{\prod_{k=1}^{K}\Gamma(N_{k|d}+\alpha^{(c^{(d)})}m_k^{(c^{(d)})})}{\Gamma(N_{\cdot|d}+\alpha^{(c^{(d)})})}\Big].
\end{aligned}
\tag{27}
$$

APPENDIX C: Computing conditional probability

$$
\begin{aligned}
&P(\boldsymbol{w}^{(d)}, \boldsymbol{z}^{(d)}|c^{(d)} = c, \mathcal{W}_{\setminus d}, \mathcal{Z}_{\setminus d}, \mathcal{C}_{\setminus d}, \beta, \boldsymbol{u}, \alpha^{(c)}, \boldsymbol{m}^{(c)}) \\
&\propto \prod_{n=1}^{N^{(d)}} P(z_m^{(d)} = k, w_m^{(d)} = w|c^{(d)} = c, \mathcal{W}_{\setminus d,n}, \mathcal{Z}_{\setminus d,m}, \mathcal{C}_{\setminus d}, \beta, \boldsymbol{u}, \alpha^{(c)}, \boldsymbol{m}^{(c)})
\end{aligned}
\tag{28}
$$

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

$k\Gamma(k)$,

$$
\begin{aligned}
&P(z_m^{(d)}=k, w_m^{(d)}=w, c^{(d)}=c|\mathcal{W}_{\setminus d,n}, \mathcal{Z}_{\setminus d,m}, \mathcal{C}_{\setminus d}, \beta, \boldsymbol{u}, \alpha^{(c)}, \boldsymbol{m}^{(c)}) \\
&\propto \frac{P(\mathcal{W}, \mathcal{Z}, \mathcal{C}|\beta, \boldsymbol{u}, \alpha, \boldsymbol{m})}{P(\mathcal{W}_{\setminus d,n}, \mathcal{Z}_{\setminus d,n}, \mathcal{C}|\beta, \boldsymbol{u}, \alpha, \boldsymbol{m})} \\
&\propto \frac{\prod_{k=1}^{K} \frac{\prod_{w=1}^{W}\Gamma(N_{wk}^{WK}+\beta u_w)}{\Gamma(\sum_{w=1}^{W} N_{wk}^{WK}+\beta)} \times \prod_{k=1}^{K} \frac{\Gamma(N_{k|d}+\alpha^{(c)}m_k^{(c)})}{\Gamma(N_{\cdot|d}+\alpha^{(c)})}}{\prod_{k=1}^{K} \frac{\prod_{w=1}^{W}\Gamma(N_{wk,\setminus d,n}^{WK}+\beta u_w)}{\Gamma(\sum_{w=1}^{W} N_{wk,\setminus d,n}^{WK}+\beta)} \times \prod_{k=1}^{K} \frac{\Gamma(N_{k|d,\setminus d,n}+\alpha^{(c)}m_k^{(c)})}{\Gamma(N_{\cdot|d,\setminus d,n}+\alpha^{(c)})}} \\
&\propto \frac{N_{wk,\setminus d,n}^{WK}+\beta u_w}{\sum_{w=1}^{W} N_{wk,\setminus d,n}^{WK}+\beta} \times \frac{N_{k|d,\setminus d,n}+\alpha^{(c)}m_k^{(c)}}{N_{\cdot|d,\setminus d,n}+\alpha^{(c)}}
\end{aligned}
\tag{29}
$$

Then, the conditional probability that a novel word generated in the document of interaction pattern $c^{(d)}=c$ would be assigned to topic $z_n^{(d)}=k$ is obtained by:

$$
\begin{aligned}
&P(z_m^{(d)}=k|w_m^{(d)}=w, c^{(d)}=c, \mathcal{W}_{\setminus d,n}, \mathcal{Z}_{\setminus d,m}, \mathcal{C}_{\setminus d}, \beta, \boldsymbol{u}, \alpha^{(c)}, \boldsymbol{m}^{(c)}) \\
&\propto \frac{N_{k|d,\setminus d,n}+\alpha^{(c)}m_k^{(c)}}{N_{\cdot|d,\setminus d,n}+\alpha^{(c)}}
\end{aligned}
\tag{30}
$$

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:

$$
\begin{aligned}
&P(w_m^{(d)}=w|z_m^{(d)}=k, c^{(d)}=c, \mathcal{W}_{\setminus d,n}, \mathcal{Z}_{\setminus d,m}, \mathcal{C}_{\setminus d}, \beta, \boldsymbol{u}, \alpha^{(c)}, \boldsymbol{m}^{(c)}) \\
&\propto \frac{N_{wk,\setminus d,n}^{WK}+\beta u_w}{\sum_{w=1}^{W} N_{wk,\setminus d,n}^{WK}+\beta}
\end{aligned}
\tag{31}
$$

NOTE: Using Equation (26), the unnormalized constant we use to check the model convergence and the corresponding log-constant are,

$$
\begin{aligned}
&\prod_{d=1}^{D}\prod_{n=1}^{N^{(d)}} P(z_m^{(d)}=k, w_m^{(d)}=w|\mathcal{W}_{\setminus d,n}, \mathcal{Z}_{\setminus d,m}, \mathcal{C}, \beta, \boldsymbol{u}, \alpha^{(c)}, \boldsymbol{m}^{(c)}) \\
&\propto \prod_{d=1}^{D}\prod_{n=1}^{N^{(d)}} \frac{N_{w_m^{(d)}z_m^{(d)},\setminus d,n}^{WK}+\beta u_{w_m^{(d)}}}{\sum_{w=1}^{W} N_{wz_m^{(d)},\setminus d,n}^{WK}+\beta} \times \frac{N_{k|d,\setminus d,n}+\alpha^{(c^{(d)})}m_{z_m^{(d)}}^{(c^{(d)})}}{N_{\cdot|d,\setminus d,n}+\alpha^{(c^{(d)})}},
\end{aligned}
\tag{32}
$$

$$
\begin{aligned}
&\sum_{d=1}^{D}\sum_{n=1}^{N^{(d)}} \log\Big(P(z_m^{(d)}=k, w_m^{(d)}=w|\mathcal{W}_{\setminus d,n}, \mathcal{Z}_{\setminus d,m}, \mathcal{C}, \beta, \boldsymbol{u}, \alpha^{(c)}, \boldsymbol{m}^{(c)})\Big) \\
&\propto \sum_{d=1}^{D}\sum_{n=1}^{N^{(d)}} \log\Big(N_{w_m^{(d)}z_m^{(d)},\setminus d,n}^{WK}+\beta u_{w_m^{(d)}}\Big) - \log\Big(\sum_{w=1}^{W} N_{wz_m^{(d)},\setminus d,n}^{WK}+\beta\Big) \\
&+ \log\Big(N_{k|d,\setminus d,n}+\alpha^{(c^{(d)})}m_{z_m^{(d)}}^{(c^{(d)})}\Big) - \log\Big(N_{\cdot|d,\setminus d,n}+\alpha^{(c^{(d)})}\Big)
\end{aligned}
\tag{33}
$$

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.

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.

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.

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).

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*.