Interaction-Partitioned Topic Models (IPTM) using a Point Process Approach

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

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

July 12, 2016

1 Ideas

Current CPME model does not involve any of temporal component, which plays a key role in email interactions. Intuitively, past interaction behaviors significantly influence future ones; for example, if an actor i sent an email to actor j, then j is highly likely to send an email back to i as a response (i.e. reciprocity). Moreover, the recency and frequency of past interactions can also be considered to effectively predict future interactions. Thus, as an exploratory data analysis, point process model for directional interaction is applied to the North Carolina email data. Starting from the existing framework focused on the analysis of content-partitioned subnetworks, I would suggest an extended approach to analyze the data using the timestamps in the email, aiming to develop a joint dynamic or longitudinal model of text-valued ties.

CPME model is a Bayesian framework using two well-known methods: Latent Dirichlet Allocation (LDA) and Latent Space Model (LSM). Basically, existence of edge depends on topic assignment k (LDA) and its corresponding interaction pattern c. Each topic $k = 1, \ldots, K$ has one interaction pattern c=1,. . . ,C, and each interaction pattern posits unique latent space (LSM), thus generating $A \times A$ matrix of probabilities $P^{(c)}$ that a message author a will include recipient r on the message, given that it is about a topic in cluster c. Incorporating point process approach, now assume that under each interaction pattern, we have $A \times A$ matrix of stochastic intensities at time t, $\boldsymbol{\lambda}^{(c)}(t)$, which depend on the history of interaction between the sender and receiver. We will refer this as interaction-partitioned topic models (IPTM).

2 IPTM Model

In this section, we introduce multiplicative Cox regression model for the edge formation process in a longitudinal communication network. 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 Point Process Framework

A single email, indexed by d, is represented by a set of tokens $w^{(d)} = \{w_m^{(d)}\}_{m=1}^{M^{(d)}}$ that comprise the text of that email, an integer $i^{(d)} \in \{1, ..., A\}$ indicating the identity of that email's sender, an integer $j^{(d)} \in \{1, ..., A\}$ indicating the identity of that email's receiver, and an integer $t^{(d)} \in [0, T]$ indicating the (unix time-based) timestamp of that email. To capture the relationship between the interaction patterns expressed in an email and that email's recipients, documents that share the interaction pattern c are associated with an $A \times A$ matrix of $\boldsymbol{\lambda}^{(c)}(t) = \{\{\lambda_{ij}^{(c)}(t)\}_{i=1}^{A}\}_{j=1}^{A}$, the stochastic intensity where $\lambda_{ij}^{(c)}(t)dt$=P{for interaction pattern c, $i \to j$ occurs in time interval $[t, t+dt)$}. We will model the counting process $\mathbf{N}^{(d|c)}(t)$ through $\boldsymbol{\lambda}^{(c)}(t)$ using a version of the Cox proportional intensity model, where $N_{ij}^{(d|c)}(t)$ denotes the number of edges (emails) for document d from actor i to actor j up to time t (from the starting point 0) given that the document corresponds to interaction pattern c. Since this counting proess $\mathbf{N}$ is document-based, each element is either 0 or 1, and only one element of the matrix is 1 while all the rests are 0 (assuming no multicast).

Combining the individual counting processes of all potential edges, $\mathbf{N}^{(d|c)}(t)$ is the multivariate counting process with $\mathbf{N}^{(d|c)}(t) = (N_{ij}^{(d|c)}(t) : i, j \in 1, ..., A, i \neq j)$. Here we make no assumption about the independence of individual edge counting process. As in Vu et al. (2011), we model the multivariate counting process via Doob-Meyer decomposition:

$$\mathbf{N}^{(d|c)}(t) = \int_0^t \boldsymbol{\lambda}^{(c)}(s)ds + \mathbf{M}(t) \tag{1}$$

where essentially $\boldsymbol{\lambda}^{(c)}(t)$ and $\mathbf{M}(t)$ may be viewed as the (deterministic) signal and (martingale) noise, respectively.

Following the multiplicative Cox model of the intensity process $\boldsymbol{\lambda}^{(c)}(t)$ given $\boldsymbol{H}_{t-}$, the entire past of the network up to but not including time t, we consider for each potential directed edge (i, j) the intensity forms:

$$\lambda_{ij}^{(c)}(t|\boldsymbol{H}_{t-}) = \lambda_0 \cdot \exp\Big\{\boldsymbol{\beta}^{(c)T}\boldsymbol{x}_t(i,j)\Big\} \cdot 1\{j \in \mathcal{A}^{(c)}\} \tag{2}$$

where λ_0 is the common baseline hazards for the overall interaction, $\boldsymbol{\beta}^{(c)}$ is an unknown vector of coefficients in $\boldsymbol{R}^p$, $\boldsymbol{x}_t(i,j)$ is a vector of p statistics for directed edge (i, j) constructed based on $\boldsymbol{H}_{t-}$, and $\mathcal{A}^{(c)}$ is the predictable receiver set of sender i corresponding to the interaction pattern c within the set of all possible actors $\mathcal{A}$. Equivalently, we can rewrite (2):

$$\lambda_{ij}^{(c)}(t|\boldsymbol{H}_{t-}) = \exp\Big\{\boldsymbol{\beta}^{(c)T}\boldsymbol{x}_t^*(i,j)\Big\} \cdot 1\{j \in \mathcal{A}^{(c)}\} \tag{3}$$

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

Based on the framework illustrated so far, the likelihood we will use for inference procedure is that of Perry and Wolfe (2013). For each type of interaction pattern $c = 1, ..., C$, estimation for $\boldsymbol{\beta}^{(c)}$ proceeds by maximizing the so-called partial likelihood of Cox (1992):

$$PL_t(\boldsymbol{\beta}^{(c)}) = \prod_{d:c^{(d)}=c} \frac{\exp\{\boldsymbol{\beta}^{(c)T}x_{t^{(d)}}(i^{(d)}, j^{(d)})\}}{\sum_{j\in\mathcal{A}^{(c)}}\exp\{\boldsymbol{\beta}^{(c)T}x_{t^{(d)}}(i^{(d)}, j)\}}, \tag{4}$$

where $t^{(d)}$, $i^{(d)}$, and $j^{(d)}$ are the time, sender, and receiver of the dth document. For computational efficiency, we will use the log-partial likelihood:

$$\log PL_t(\boldsymbol{\beta}^{(c)}) = \sum_{d:c^{(d)}=c} \Big\{\boldsymbol{\beta}^{(c)T}x_{t^{(d)}}(i^{(d)}, j^{(d)}) - \log\big[\sum_{j\in\mathcal{A}^{(c)}} \exp\{\boldsymbol{\beta}^{(c)T}x_{t^{(d)}}(i^{(d)}, j)\}\big]\Big\}. \tag{5}$$

2.2 Generative Process

The generative process of this model follows the topic model (LDA) of Blei et al. (2003) and the author-topic model of Rosen-Zvi et al. (2004). Same as LDA, documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words. However, one crucial difference is that each document is connected to one type of interaction pattern, and the topic distributions vary depending on the assigned interaction pattern.

Conditioned on the interaction pattern and their distributions over topics, the process by which a document is generated can be summarized as follows: first, an interaction pattern is chosen by multinomial for each document; next, a topic is sampled for each word from the distribution over topics associated with the interaction pattern of the document; finally, words themselves are sampled from the distribution over words associated with each topic. At the same time, the unique sender-recipient pair of the document is determined by the rate of intensities associated with the interaction pattern and history of interactions until the time the document is written. Below are the detailed generative process for each document in a corpus D and its plate notation (Figure 1), and Table 1 summarizes the notations used in this paper:

1. $\boldsymbol{\phi}^{(k)} \sim \text{Dir}(\delta, \mathbf{n})$ **[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 with concentration parameter δ is placed .

2. For each of the C interaction patterns [**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 network statistics $\boldsymbol{x}_t(i,j)$ that rely on the history of interactions.

   (b) Using $\boldsymbol{\beta}^{(c)}$ in (a), update $\boldsymbol{\lambda}^{(c)}(t)$
   - We use the equation $\lambda_{ij}^{(c)}(t) = \exp\Big\{\boldsymbol{\beta}^{(c)T}\boldsymbol{x}_t^*(i,j)\Big\} \cdot 1\{j \in \mathcal{A}^{(c)}\}$ for all $i \in \mathcal{A}, j \in \mathcal{A}, i \neq j$.

   (c) $\boldsymbol{\theta}^{(c)} \sim \text{Dir}(\alpha, \mathbf{m})$
   - Each email has a discrete distribution over topics $\boldsymbol{\theta}^{(c)}$, since the topic proportions for documents in the same cluster are drawn from the same distribution. The Dirichlet parameters α and $\mathbf{m}$ may or may not vary by interaction patterns.

3. For each of the D documents [**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 parameter $\boldsymbol{\gamma}$. Here, we assign the prior for the multinomial parameter $\boldsymbol{\gamma} \sim \text{Dir}(\eta, \boldsymbol{l})$

   (b) $\mathbf{N}^{(d|c^{(d)})}(t^{(d)}) \sim \text{CP}(\boldsymbol{\lambda}^{(c^{(d)})}(t^{(d)}))$
   - The actual update of the counting process $\mathbf{N}^{(d|c^{(d)})}(t)$ of the email d is $N_{i^{(d)}j^{(d)}}^{(d|c^{(d)})}(t^{(d)}) = 1$ and the rest $N_{(i,j)\neq(i^{(d)},j^{(d)})}^{(d|c^{(d)})}(t^{(d)}) = 0$.

4. For each of the M words [**See Algorithm 4**]:

   (a) $z_m^{(d)} \sim \text{Multinomial}(\boldsymbol{\theta}^{(c^{(d)})})$

   (b) $w_m^{(d)} \sim \text{Multinomial}(\phi^{(z_m^{(d)})})$

Algorithm 1 Topic Word Distributions

---

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

Algorithm 2 Interaction Patterns

for *c=1 to C* **do**
  draw $\boldsymbol{\beta}^{(c)} \sim \text{Normal}(\mathbf{0}, \sigma^2 I_P)$
  for *i=1 to A* **do**
    for *j=1 to A* **do**
      if $i \neq j$ **then**
        set $\lambda_{ij}^{(c)}(t) = \exp\Big\{\boldsymbol{\beta}^{(c)T}\boldsymbol{x}_t^*(i,j)\Big\} \cdot 1\{j \in \mathcal{A}^{(c)}\}$
      end
      else
        set $\lambda_{ij}^{(c)}(t) = 0$
      end
    end
  end
  draw $\boldsymbol{\theta}^{(c)} \sim \text{Dir}(\alpha, \mathbf{m})$
end

Algorithm 3 Document-Interaction Pattern Assignments

for *d=1 to D* **do**
  draw $c^{(d)} \sim \text{Multinomial}(\boldsymbol{\gamma})$
  draw $\mathbf{N}^{(d|c^{(d)})}(t^{(d)}) \sim \text{CP}(\boldsymbol{\lambda}^{(c^{(d)})}(t^{(d)}))$
end

Algorithm 4 Tokens

for *d=1 to D* **do**
  set $M^{(d)}$ = the number of words in document d
  for *m=1 to* $M^{(d)}$ **do**
    draw $z_m^{(d)} \sim \text{Multinomial}(\boldsymbol{\theta}^{(c^{(d)})})$
    draw $w_m^{(d)} \sim \text{Multinomial}(\boldsymbol{\phi}^{(z_m^{(d)})})$
  end
end

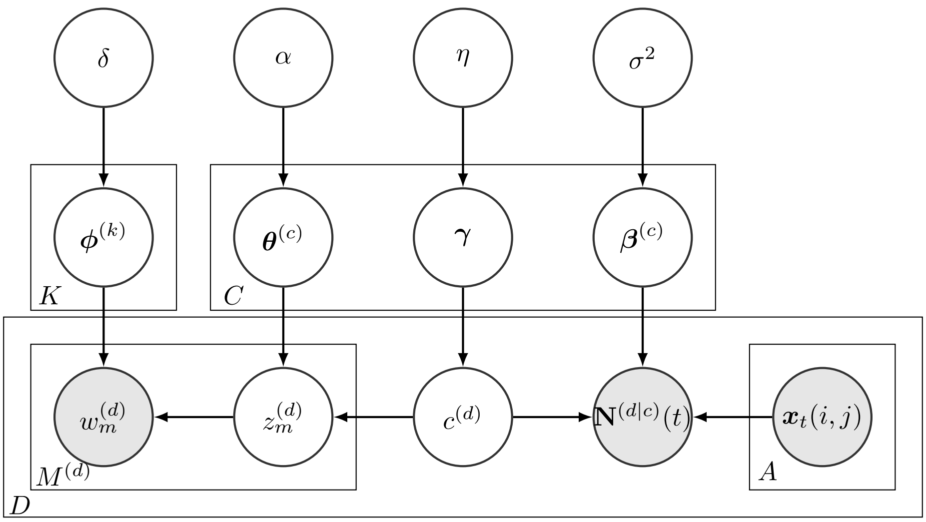

Figure 1: Plate notation of IPTM

Authors of the corpus	$\mathcal{A}$	Set
Authors of the corpus given interaction pattern c	$\mathcal{A}^{(c)}$	Set
Number of authors	A	Scalar
Number of documents	D	Scalar
Number of words in the d^{th} document	$M^{(d)}$	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	M^{CK}	Scalar
Number of words assigned to word and topic	M^{WK}	Scalar
Interaction pattern of the d^{th} document	$c^{(d)}$	Scalar
Time of the d^{th} document	$t^{(d)}$	Scalar
Words in the d^{th} document	$\boldsymbol{w}^{(d)}$	$M^{(d)}$-dimensional vector
m^{th} word in the d^{th} document	$w_m^{(d)}$	m^{th} component of $\boldsymbol{w}^{(d)}$
Topic assignments in the d^{th} document	$\boldsymbol{z}^{(d)}$	$M^{(d)}$-dimensional vector
Topic assignments for m^{th} word in the d^{th} document	$z_m^{(d)}$	m^{th} component of $\boldsymbol{z}^{(d)}$
Dirichlet concentration prior	α	Scalar
Dirichlet base prior	$\boldsymbol{m}$	K-dimensional vector
Dirichlet concentration prior	δ	Scalar
Dirichlet base prior	$\boldsymbol{n}$	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 given interaction patterns	Θ	$K \times C$ matrix
Probabilities of the topics given interaction pattern c	$\boldsymbol{\theta}^{(c)}$	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)	$\boldsymbol{x}_t(i, j)$	p-dimensional vector
Counting process in the d^{th} document given interaction pattern	$\mathbf{N}^{(d\|c)}(t)$	$A \times A$ matrix

Table 1: Symbols associated with IPTM, as used in this work

2.3 Dynamic covariates to measure network effects

The network statistics $\boldsymbol{x}_t(i, j)$ of equations (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 Perry and Wolfe (2013), we use 6 effects as components of $\boldsymbol{x}_t(i, j)$. The first two behaviors (send and receive) are dyadic, involving exactly two actors, while the last four (2-send, 2-receive, sibling, and cosibling) are triadic, involving exactly three actors. In addition, we include intercept term and use $\boldsymbol{x}_t^*(i, j)$ so that we can estimate the baseline intensities at the same time. One different thing is that we define the effects not to be based on finite sub-interval, which require large number of dimention. Instead, we create a single statistic for each effect by incorporating the recency of event into the statistic itself.

0. $\text{intercept}_t(i, j) = 1$
1. $\text{send}_t(i, j) = \sum_{d: t^{(d)} < t} I\{i \rightarrow j\} \cdot g(t - t^{(d)})$
2. $\text{receive}_t(i, j) = \sum_{d: t^{(d)} < t} I\{j \rightarrow i\} \cdot g(t - t^{(d)})$
3. $\text{2-send}_t(i, j) = \sum_{h \neq i, j} \Big(\sum_{d: t^{(d)} < t} I\{i \rightarrow h\} \cdot g(t - t^{(d)})\Big)\Big(\sum_{d: t^{(d)} < t} I\{h \rightarrow j\} \cdot g(t - t^{(d)})\Big)$
4. $\text{2-receive}_t(i, j) = \sum_{h \neq i, j} \Big(\sum_{d: t^{(d)} < t} I\{h \rightarrow i\} \cdot g(t - t^{(d)})\Big)\Big(\sum_{d: t^{(d)} < t} I\{j \rightarrow h\} \cdot g(t - t^{(d)})\Big)$

5. $\text{sibling}_t(i,j) = \sum_{h\neq i,j} \Big(\sum_{d:t^{(d)}<t} I\{h \to i\} \cdot g(t - t^{(d)}) \Big) \Big(\sum_{d:t^{(d)}<t} I\{h \to j\} \cdot g(t - t^{(d)}) \Big)$

6. $\text{cosibling}_t(i,j) = \sum_{h\neq i,j} \Big(\sum_{d:t^{(d)}<t} I\{i \to h\} \cdot g(t - t^{(d)}) \Big) \Big(\sum_{d:t^{(d)}<t} I\{j \to h\} \cdot g(t - t^{(d)}) \Big)$

Here, $g(t - t^{(d)})$ 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 $g(t-t^{(d)}) = \lambda e^{-\lambda(t-t^{(d)})}$ where λ is the parameter of speed at which sender replies to emails, with larger values indicating faster response times. Indeed, λ^{-1} is the expected number of hours it takes to reply to a typical email. For simplicity, we can fix $\lambda = 1$.

3 Inference

The inference for IPTM is similar to that of CPME. In this case, what we actually observe are the tokens $\mathcal{W} = \{\boldsymbol{w}^{(d)}\}_{d=1}^{D}$ and the sender, recipient, and timestamps of the email in the form of the counting process $\mathcal{N} = \{\boldsymbol{N}^{(d)}(t^{(d)})\}_{d=1}^{D}$. Next, $\mathcal{X} = \{\boldsymbol{x}_{t^{(d)}}(i,j)\}_{d=1}^{D}$ is the metadata, and the latent variables are $\Phi = \{\boldsymbol{\phi}^{(k)}\}_{k=1}^{K}$, $\Theta = \{\boldsymbol{\theta}^{(c)}\}_{c=1}^{C}$, $\mathcal{Z} = \{\boldsymbol{z}^{(d)}\}_{d=1}^{D}$, $\mathcal{C} = \{c^{(d)}\}_{d=1}^{D}$, and $\mathcal{B} = \{\boldsymbol{\beta}^{(c)}\}_{c=1}^{C}$.

Below is the the big joint distribution

$$\begin{aligned}
&P(\Phi, \Theta, \mathcal{W}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \mathcal{N} | \mathcal{X}, \delta, \boldsymbol{n}, \alpha, \boldsymbol{m}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \sigma^2) \\
&= P(\mathcal{W}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \mathcal{N} | \Phi, \Theta, \mathcal{X}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \sigma^2) P(\Phi, \Theta | \delta, \boldsymbol{n}, \alpha, \boldsymbol{m}) \\
&= P(\mathcal{W}|\mathcal{Z}, \Phi) P(\mathcal{Z}|\Theta) P(\mathcal{N}|\mathcal{C}, \mathcal{X}, \mathcal{B}) P(\mathcal{B}|\mathcal{C}, \sigma^2) P(\Phi|\delta, \boldsymbol{n}) P(\Theta|\mathcal{C}, \alpha, \boldsymbol{m}) P(\mathcal{C}|\boldsymbol{\gamma}) P(\boldsymbol{\gamma}|\boldsymbol{\eta})
\end{aligned} \tag{6}$$

Now we can integrate out Φ and Θ in latent Dirichlet allocation by applying Dirichlet-multinomial conjugacy as we did in CPME. See APPENDIX A for the detailed steps. After integration, we obtain below:

$$\propto P(\mathcal{W}|\mathcal{Z}) P(\mathcal{Z}|\mathcal{C}, \delta, \boldsymbol{n}, \alpha, \boldsymbol{m}) P(\mathcal{N}|\mathcal{C}, \mathcal{B}, \mathcal{X}) P(\mathcal{B}|\mathcal{C}, \sigma^2) P(\mathcal{C}|\gamma) \tag{7}$$

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

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

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

3.1 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. $P(c^{(d)} = c | \mathcal{W}, \mathcal{Z}, \mathcal{C}_{\backslash d}, \mathcal{B}, \mathcal{N}, \mathcal{X}, \delta, \boldsymbol{n}, \alpha, \boldsymbol{m}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \sigma^2)$. We can derive the equation as below:

$$\begin{aligned}
&P(c^{(d)} = c | \mathcal{W}, \mathcal{Z}, \mathcal{C}_{\backslash d}, \mathcal{B}, \mathcal{N}, \mathcal{X}, \delta, \boldsymbol{n}, \alpha, \boldsymbol{m}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \sigma^2) \\
&\propto P(c^{(d)} = c, \boldsymbol{w}^{(d)}, \boldsymbol{z}^{(d)}, \mathbf{N}^{(d)}(t^{(d)}) | \mathcal{W}_{\backslash d}, \mathcal{Z}_{\backslash d}, \mathcal{C}_{\backslash d}, \mathcal{B}, \mathcal{N}_{\backslash d}, \mathcal{X}, \delta, \boldsymbol{n}, \alpha, \boldsymbol{m}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \sigma^2) \\
&\propto P(c^{(d)} = c | \mathcal{C}_{\backslash d}, \boldsymbol{\gamma}) P(\mathbf{N}^{(d)}(t^{(d)}) | c^{(d)} = c, \mathcal{C}_{\backslash d}, \mathcal{B}, \mathcal{N}_{\backslash d}, \mathcal{X}) P(\boldsymbol{w}^{(d)}, \boldsymbol{z}^{(d)} | c^{(d)} = c, \mathcal{W}_{\backslash d}, \mathcal{Z}_{\backslash d}, \mathcal{C}_{\backslash d}, \delta, \boldsymbol{n}, \alpha, \boldsymbol{m}),
\end{aligned} \tag{9}$$

where $P(c^{(d)} = c | \mathcal{C}_{\backslash d}, \boldsymbol{\gamma})$ comes from the multinomial prior γ and $P(\mathbf{N}^{(d)}(t^{(d)}) | c^{(d)} = c, \mathcal{C}_{\backslash d}, \mathcal{B}, \mathcal{N}_{\backslash d}, \mathcal{X})$ 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 product term since resampling of c is document-specific). For the last term $P(\boldsymbol{w}^{(d)}, \boldsymbol{z}^{(d)} | c^{(d)} = c, \mathcal{W}_{\backslash d}, \mathcal{Z}_{\backslash d}, \mathcal{C}_{\backslash d}, \delta, \boldsymbol{n}, \alpha, \boldsymbol{m})$, we will follow typical LDA approach.

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

$$= \Big[\gamma_c\Big] \times \Big[\frac{\exp\{\boldsymbol{\beta}^{(c)T} x_{t^{(d)}}(i^{(d)}, j^{(d)})\}}{\sum_{j \in \mathcal{A}^{(c)}} \exp\{\boldsymbol{\beta}^{(c)T} x_{t^{(d)}}(i^{(d)}, j)\}}\Big] \times \Big[\prod_{m=1}^{M^{(d)}} \frac{M^{CK}_{cz_m^{(d)},\backslash d,m} + \alpha m_k}{\sum_{k=1}^{K} M^{CK}_{ck,\backslash d,m} + \alpha}\Big], \tag{10}$$

where M^{CK}_{ck} is the number of times topic k shows up given the interaction pattern c over the entire corpus. Furthermore, we can take the log of Equation (10) to avoid numerical issue from exponentiation and increase the speed of computation, which becomes:

$$\log(\gamma_c) + \Big(\boldsymbol{\beta}^{(c)T} x_{t^{(d)}}(i^{(d)}, j^{(d)}) - \log\big[\sum_{j \in \mathcal{A}^{(c)}} \exp\{\boldsymbol{\beta}^{(c)T} x_{t^{(d)}}(i^{(d)}, j)\}\big]\Big) + \sum_{m=1}^{M^{(d)}} \log(\frac{M^{CK}_{cz_m^{(d)},\backslash d,m} + \alpha m_k}{\sum_{k=1}^{K} M^{CK}_{ck,\backslash d,m} + \alpha}). \tag{11}$$

3.2 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 B:

$$\begin{aligned} &P(z_m^{(d)} = k | \mathcal{W}, \mathcal{Z}_{\backslash d,m}, \mathcal{C}, \mathcal{B}, \mathcal{N}, \mathcal{X}, \delta, \boldsymbol{n}, \alpha, \boldsymbol{m}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \sigma^2) \\ &\propto P(z_m^{(d)} = k, w_m^{(d)} | \mathcal{W}_{\backslash d,m}, \mathcal{Z}_{\backslash d,m}, C, \delta, \boldsymbol{n}, \alpha, \boldsymbol{m}) \end{aligned} \tag{12}$$

where the subscript "$\backslash d, m$" denotes the exclsuion of position m in email d. In the last line of equation (10), it is the contribution of LDA, so similar to CPME we can write the conditional probability:

$$\propto (M^{CK}_{c^{(d)}k,\backslash d,m} + \alpha m_k) \cdot \frac{M^{WK}_{w_m^{(d)}k,\backslash d,m} + \delta n_w}{\sum_{w=1}^{W} M^{WK}_{wk,\backslash d,m} + \delta} \tag{13}$$

which is the well-known form of collapsed Gibbs sampling equation for LDA.

3.3 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 δ_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:

$$\text{Acceptance Probability} = \begin{cases} \frac{P(\mathcal{B}'|\mathcal{W},\mathcal{Z},\mathcal{C},\mathcal{N},\mathcal{X})}{P(\mathcal{B}|\mathcal{W},\mathcal{Z},\mathcal{C},\mathcal{N},\mathcal{X})} & \text{if} < 1 \\ 1 \quad \text{else} \end{cases} \tag{14}$$

After factorization, we get

$$\begin{aligned} \frac{P(\mathcal{B}'|\mathcal{W},\mathcal{Z},\mathcal{C},\mathcal{N},\mathcal{X})}{P(\mathcal{B}|\mathcal{W},\mathcal{Z},\mathcal{C},\mathcal{N},\mathcal{X})} &= \frac{P(\mathcal{N}|\mathcal{B}',\mathcal{W},\mathcal{Z},\mathcal{C},\mathcal{X})P(\mathcal{B}')}{P(\mathcal{N}|\mathcal{B},\mathcal{W},\mathcal{Z},\mathcal{C},\mathcal{X})P(\mathcal{B})} \\ &= \frac{P(\mathcal{N}|\mathcal{C},\mathcal{X},\mathcal{B}')P(\mathcal{B}')}{P(\mathcal{N}|\mathcal{C},\mathcal{X},\mathcal{B})P(\mathcal{B})}, \end{aligned} \tag{15}$$

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 3.1, we can take the log and obtain the log of acceptance ratio as following:

$$\begin{aligned}
&\log\Big(\phi_d(\boldsymbol{\beta}'^{(c)};\mathbf{0},\sigma^2 I_P)\Big) - \log\Big(\phi_d(\boldsymbol{\beta}'^{(c)};\mathbf{0},\sigma^2 I_P)\Big) \\
&+ \sum_{d:c^{(d)}=c} \Big\{\boldsymbol{\beta}'^{(c)T} x_{t^{(d)}}(i^{(d)}, j^{(d)}) - \log\big[\sum_{j\in\mathcal{A}^{(c)}} \exp\{\boldsymbol{\beta}'^{(c)T} x_{t^{(d)}}(i^{(d)}, j)\}\big]\Big\} \\
&- \sum_{d:c^{(d)}=c} \Big\{\boldsymbol{\beta}^{(c)T} x_{t^{(d)}}(i^{(d)}, j^{(d)}) - \log\big[\sum_{j\in\mathcal{A}^{(c)}} \exp\{\boldsymbol{\beta}^{(c)T} x_{t^{(d)}}(i^{(d)}, j)\}\big]\Big\},
\end{aligned} \tag{16}$$

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

$$\log(\text{Acceptance Probability}) = \min((16),\, 0) \tag{17}$$

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).

4 Application: 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). Exploratory analysis revealed that Dare county experienced significant change in the pattern of email exchanges; specifically, during the emergency period, email interactions significanty less rely on previous history of interactions, compared to the normal period. On the other hand, Vance county did not experience any distinctive change, and the possible reason for the difference is the locations of two counties. Here we apply IPTM to both data to see the differences in detail, in terms of the interaction patterns and topics of the corpus.

4.1 Vance county email data

After treating multicast emails (those involving a single sender but multiple receivers) as multiple distinct emails, Vance county data contains 269 emails (only count the email with the number of words greater than 0) between 18 actors, including 620 vocabulary in total. We used $K = 20$ topics assuming symmetric Dirichlet prior with the concentration parameter $\alpha = 5$, and $C = 5$ interaction patterns assuming multinomial prior with parameter γ (coming from symmetric Dirichlet prior with the concentration parameter $\eta = 5$). For topic-word distributions, we assumed that ϕ follows symmectic Dirichlet distribution with the concentration parameter $\delta = 5$. Markov chain Monte Carlo (MCMC) sampling was implemented based on the order and scheme illustrated in Section 3. We ran total 20,000 iterations, which took about 10.36 hours. Among the entire samples, 5,000 were discarded as a burn-in, and every 15th sample was taken for thinning. Below are the summary of interaction pattern-topic-word assignments. Each interaction pattern is illustrated with (a) posterior estimates (means) of dynamic network effects 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 the interaction pattern.

	IP1	**IP2**	**IP3**	**IP4**	**IP5**
intercept	-0.036	-0.113	0.054	-0.024	-0.010
send	0.555	0.306	0.869	0.065	0.532
receive	0.175	0.090	0.091	0.119	0.019
2-send	-0.055	-0.021	0.072	0.044	-0.034
2-receive	-0.003	-0.010	0.051	-0.003	0.052
sibling	0.019	0.019	0.054	-0.071	-0.008
cosibling	-0.028	0.073	0.019	0.042	0.007

Table 2: Summary of posterior $\boldsymbol{\beta}^{(c)}$ estimates for Vance county

IP1 (51 emails)		**IP2** (46 emails)		**IP3** (79 emails)		**IP4** (29 emails)		**IP5** (64 emails)	
Topic 10	0.148	Topic 19	0.145	Topic 9	0.194	Topic 17	0.286	Topic 2	0.194
emergency	0.0448	description	0.0431	message	0.0452	will	0.1045	will	0.0807
operations	0.0439	director	0.0259	electronic	0.0440	phones	0.0587	extension	0.0333
office	0.0325	phone	0.0241	department	0.0320	week	0.0357	latest	0.0324
henderson	0.0283	planning	0.0229	heads	0.0285	october	0.0346	directory	0.0271
director	0.0277	henderson	0.0196	request	0.0235	system	0.0285	jail	0.0267
fax	0.0277	development	0.0194	time	0.0231	phone	0.0252	folks	0.0253
street	0.0259	street	0.0191	review	0.0214	department	0.0238	attached	0.0251
church	0.0245	suite	0.0186	records	0.0213	rest	0.0229	director	0.0170
center	0.0245	church	0.0183	public	0.0212	cutting	0.0227	cutover	0.0156
suite	0.0240	fax	0.0181	ncgs	0.0211	provided	0.0227	henderson	0.0156
Topic 13	0.123	Topic 18	0.134	Topic 5	0.193	Topic 12	0.286	Topic 1	0.193
emergency	0.0429	description	0.0437	message	0.0455	will	0.1064	will	0.0791
operations	0.0420	director	0.0257	electronic	0.0432	phones	0.0592	extension	0.0336
office	0.0323	phone	0.0240	department	0.0317	week	0.0352	latest	0.0323
director	0.0288	planning	0.0227	heads	0.0284	october	0.0345	jail	0.0268
henderson	0.0279	street	0.0202	request	0.0231	system	0.0282	directory	0.0264
fax	0.0267	development	0.0195	time	0.0230	phone	0.0252	folks	0.0253
street	0.0255	henderson	0.0190	response	0.0217	cutting	0.0232	attached	0.0250
suite	0.0252	fax	0.0184	records	0.0212	training	0.0231	director	0.0172
church	0.0249	church	0.0184	review	0.0210	rest	0.0231	henderson	0.0161
center	0.0240	suite	0.0181	hereto	0.0209	provided	0.0230	cutover	0.0153
Topic 4	0.123	Topic 3	0.106	Topic 7	0.193	Topic 14	0.285	Topic 20	0.192
emergency	0.0432	description	0.0442	message	0.0450	will	0.1036	will	0.0798
operations	0.0432	director	0.0275	electronic	0.0428	phones	0.0600	extension	0.0341
office	0.0331	phone	0.0240	department	0.0320	week	0.0352	latest	0.0321
fax	0.0297	planning	0.0234	heads	0.0283	october	0.0347	jail	0.0265
director	0.0295	street	0.0210	request	0.0241	system	0.0275	attached	0.0254
henderson	0.0291	henderson	0.0209	time	0.0232	phone	0.0251	directory	0.0254
street	0.0265	development	0.0205	review	0.0213	department	0.0242	folks	0.0253
suite	0.0261	fax	0.0199	hereto	0.0212	provided	0.0227	director	0.0171
church	0.0260	church	0.0199	records	0.0210	training	0.0227	henderson	0.0163
center	0.0245	suite	0.0198	ncgs	0.0209	rest	0.0226	cutover	0.0154

Table 3: Summary of MCMC sampling results for Vance county. Each interaction pattern is shown with the top 3 topics and words that have the highest probability conditioned on that topic.

4.2 Dare county email data

APPENDIX

APPENDIX A: Deriving the sampling equations for IPTM

$$\begin{aligned}
&P(\Phi, \Theta, \mathcal{W}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \mathcal{N} | \mathcal{X}, \delta, \boldsymbol{n}, \alpha, \boldsymbol{m}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \sigma^2) \\
&= P(\mathcal{W}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \mathcal{N} | \Phi, \Theta, \mathcal{X}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \sigma^2) P(\Phi, \Theta | \delta, \boldsymbol{n}, \alpha, \boldsymbol{m}) \\
&= P(\mathcal{W} | \mathcal{Z}, \Phi) P(\mathcal{Z} | \Theta) P(\mathcal{N} | \mathcal{C}, \mathcal{B}, \mathcal{X}) P(\mathcal{B} | \mathcal{C}, \sigma^2) P(\Phi | \delta, \boldsymbol{n}) P(\Theta | \mathcal{C}, \alpha, \boldsymbol{m}) P(\mathcal{C} | \boldsymbol{\gamma}) P(\boldsymbol{\gamma} | \boldsymbol{\eta}) \\
&= \Big[\prod_{d=1}^{D} \prod_{m=1}^{M^{(d)}} P(w_m^{(d)} | \phi_{z_m^{(d)}})\Big] \times \Big[\prod_{d=1}^{D} \prod_{m=1}^{M^{(d)}} P(z_m^{(d)} | \boldsymbol{\theta}^{(c)})\Big] \times \Big[\prod_{d=1}^{D} P(\mathbf{N}^{(d)}(t^{(d)}) | c^{(d)}, \boldsymbol{x}(t^{(d)}), \boldsymbol{\beta}^{(c)})\Big] \\
&\quad \times \Big[\prod_{c=1}^{C} P(\boldsymbol{\beta}^{(c)} | \sigma^2)\Big] \times \Big[\prod_{k=1}^{K} P(\boldsymbol{\phi}^{(k)} | \delta, \boldsymbol{n})\Big] \times \Big[\prod_{c=1}^{C} P(\boldsymbol{\theta}^{(c)} | \alpha, \boldsymbol{m})\Big] \times \Big[\prod_{d=1}^{D} P(c^{(d)} | \boldsymbol{\gamma})\Big] \times P(\boldsymbol{\gamma} | \boldsymbol{\eta})
\end{aligned} \tag{18}$$

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 (20) as below:

$$
\begin{aligned}
&\propto \Big[\prod_{d=1}^{D}\prod_{m=1}^{M^{(d)}} P(w_m^{(d)}|\phi_{z_m^{(d)}})\Big] \times \Big[\prod_{d=1}^{D}\prod_{m=1}^{M^{(d)}} P(z_m^{(d)}|\boldsymbol{\theta}^{(c)})\Big] \times \Big[\prod_{d=1}^{D} P(\mathbf{N}^{(d)}(t^{(d)})|c^{(d)}, \boldsymbol{x}(t^{(d)}), \boldsymbol{\beta}^{(c)})\Big] \\
&\quad \times \Big[\prod_{k=1}^{K} P(\boldsymbol{\phi}^{(k)}|\delta, \boldsymbol{n})\Big] \times \Big[\prod_{c=1}^{C} P(\boldsymbol{\theta}^{(c)}|\alpha, \boldsymbol{m})\Big] \times \Big[\prod_{d=1}^{D} P(c^{(d)}|\boldsymbol{\gamma})\Big] \\
&= \Big[\prod_{d=1}^{D}\prod_{m=1}^{M^{(d)}} \phi_{w_m^{(d)} z_m^{(d)}}\Big] \times \Big[\prod_{d=1}^{D}\prod_{m=1}^{M^{(d)}} \boldsymbol{\theta}^{(c)}_{z_m^{(d)}}\Big] \times \Big[\prod_{d=1}^{D} \frac{\exp\{\boldsymbol{\beta}^{(c)T} x_{t^{(d)}}(i^{(d)}, j^{(d)})\}}{\sum_{j\in\mathcal{A}^{(c)}} \exp\{\boldsymbol{\beta}^{(c)T} x_{t^{(d)}}(i^{(d)}, j)\}}\Big] \\
&\quad \times \Big[\prod_{k=1}^{K}\Big(\frac{\Gamma(\sum_{w=1}^{W}\delta n_w)}{\prod_{w=1}^{W}\Gamma(\delta n_w)}\prod_{w=1}^{W}\phi_{wk}^{\delta n_w - 1}\Big)\Big] \times \Big[\prod_{c=1}^{C}\Big(\frac{\Gamma(\sum_{k=1}^{K}\alpha m_k)}{\prod_{k=1}^{K}\Gamma(\alpha m_k)}\prod_{k=1}^{K}(\boldsymbol{\theta}_k^{(c)})^{\alpha m_k - 1}\Big)\Big] \times \Big[\prod_{d=1}^{D}\gamma_c^{I(c^{(d)}=c)}\Big] \\
&= \Big[\frac{\Gamma(\sum_{w=1}^{W}\delta n_w)}{\prod_{w=1}^{W}\Gamma(\delta n_w)}\Big]^K \times \Big[\frac{\Gamma(\sum_{w=1}^{W}\delta n_w)}{\prod_{w=1}^{W}\Gamma(\delta n_w)}\Big]^C \times \Big[\prod_{d=1}^{D} \frac{\exp\{\boldsymbol{\beta}^{(c)T} x_{t^{(d)}}(i^{(d)}, j^{(d)})\}}{\sum_{j\in\mathcal{A}^{(c)}} \exp\{\boldsymbol{\beta}^{(c)T} x_{t^{(d)}}(i^{(d)}, j)\}}\Big] \\
&\quad \times \Big[\prod_{d=1}^{D}\gamma_{c^{(d)}}\Big] \times \Big[\prod_{k=1}^{K}\prod_{w=1}^{W}\phi_{wk}^{M_{wk}^{WK}+\delta n_w - 1}\Big] \times \Big[\prod_{c=1}^{C}\prod_{k=1}^{K}(\boldsymbol{\theta}_k^{(c)})^{M_{ck}^{CK}+\alpha m_k - 1}\Big]
\end{aligned} \tag{19}
$$

where M_{wk}^{WK} is the number of times the w^{th} word in the vocabulary is assigned to topic k, and M_{ck}^{CK} is the number of times topic k shows up given the interaction pattern c. 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 $\int \prod_{m=1}^{M}[x_m^{k_m - 1}dx_m] = \frac{\prod_{m=1}^{M}\Gamma(k_m)}{\Gamma(\sum_{m=1}^{M}k_m)}$ to (22), we have:

$$
\begin{aligned}
&P(\mathcal{W}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \mathcal{N}|\mathcal{X}, \delta, \boldsymbol{n}, \alpha, \boldsymbol{m}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \sigma^2) \\
&= \text{Const.} \int_\Theta\int_\Phi \Big[\prod_{k=1}^{K}\prod_{w=1}^{W}\phi_{wk}^{M_{wk}^{WK}+\delta n_w - 1}\Big]\Big[\prod_{c=1}^{C}\prod_{k=1}^{K}(\boldsymbol{\theta}_k^{(c)})^{M_{ck}^{CK}+\alpha m_k - 1}\Big] d\Phi d\Theta \\
&= \text{Const.}\Big[\prod_{k=1}^{K}\int_{\phi_{:k}}\prod_{w=1}^{W}\phi_{wk}^{M_{wk}^{WK}+\delta n_w - 1} d\phi_{:k}\Big] \times \Big[\prod_{c=1}^{C}\int_{\theta_{:c}}\prod_{k=1}^{K}(\boldsymbol{\theta}_k^{(c)})^{M_{ck}^{CK}+\alpha m_k - 1} d\theta_{:c}\Big] \\
&= \text{Const.}\Big[\prod_{k=1}^{K}\frac{\prod_{w=1}^{W}\Gamma(M_{wk}^{WK}+\delta n_w)}{\Gamma(\sum_{w=1}^{W}M_{wk}^{WK}+\delta)}\Big] \times \Big[\prod_{c=1}^{C}\frac{\prod_{k=1}^{K}\Gamma(M_{ck}^{CK}+\alpha m_k)}{\Gamma(\sum_{k=1}^{K}M_{ck}^{CK}+\alpha)}\Big].
\end{aligned} \tag{20}
$$

APPENDIX B: Computing conditional probability

$$
\begin{aligned}
&P(\boldsymbol{w}^{(d)}, \boldsymbol{z}^{(d)}|c^{(d)} = c, \mathcal{W}_{\backslash d}, \mathcal{Z}_{\backslash d}, \mathcal{C}_{\backslash d}, \delta, \boldsymbol{n}, \alpha, \boldsymbol{m}) \\
&\propto \prod_{m=1}^{M^{(d)}} P(z_m^{(d)} = k, w_m^{(d)} = w|c^{(d)} = c, \mathcal{W}_{\backslash d,m}, \mathcal{Z}_{\backslash d,m}, \mathcal{C}_{\backslash d}, \delta, \boldsymbol{n}, \alpha, \boldsymbol{m})
\end{aligned} \tag{21}
$$

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

$$\begin{aligned}
&P(z^{(d)}_m = k, w^{(d)}_m = w, c^{(d)} = c|\mathcal{W}_{\backslash d,m}, \mathcal{Z}_{\backslash d,m}, \mathcal{C}_{\backslash d}, \delta, \boldsymbol{n}, \alpha, \boldsymbol{m}) \\
&\propto \frac{P(\mathcal{W}, \mathcal{Z}, \mathcal{C}|\delta, \boldsymbol{n}, \alpha, \boldsymbol{m})}{P(\mathcal{W}_{\backslash d,m}, \mathcal{Z}_{\backslash d,m}, \mathcal{C}|\delta, \boldsymbol{n}, \alpha, \boldsymbol{m})} \\
&\propto \frac{\prod_{k=1}^{K} \frac{\prod_{w=1}^{W} \Gamma(M^{WK}_{wk} + \delta n_w)}{\Gamma(\sum_{w=1}^{W} M^{WK}_{wk} + \delta)} \times \prod_{c=1}^{C} \frac{\prod_{k=1}^{K} \Gamma(M^{CK}_{ck} + \alpha m_k)}{\Gamma(\sum_{k=1}^{K} M^{CK}_{ck} + \alpha)}}{\prod_{k=1}^{K} \frac{\prod_{w=1}^{W} \Gamma(M^{WK}_{wk,\backslash d,m} + \delta n_w)}{\Gamma(\sum_{w=1}^{W} M^{WK}_{wk,\backslash d,m} + \delta)} \times \prod_{c=1}^{C} \frac{\prod_{k=1}^{K} \Gamma(M^{CK}_{ck,\backslash d,m} + \alpha m_k)}{\Gamma(\sum_{k=1}^{K} M^{CK}_{ck,\backslash d,m} + \alpha)}} \\
&\propto \frac{M^{WK}_{wk,\backslash d,m} + \delta n_w}{\sum_{w=1}^{W} M^{WK}_{wk,\backslash d,m} + \delta} \times \frac{M^{CK}_{ck,\backslash d,m} + \alpha m_k}{\sum_{k=1}^{K} M^{CK}_{ck,\backslash d,m} + \alpha}
\end{aligned} \tag{22}$$

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

$$\begin{aligned}
&P(z^{(d)}_m = k|w^{(d)}_m = w, c^{(d)} = c, \mathcal{W}_{\backslash d,m}, \mathcal{Z}_{\backslash d,m}, \mathcal{C}_{\backslash d}, \delta, \boldsymbol{n}, \alpha, \boldsymbol{m}) \\
&\propto \frac{M^{CK}_{ck,\backslash d,m} + \alpha m_k}{\sum_{k=1}^{K} M^{CK}_{ck,\backslash d,m} + \alpha}
\end{aligned} \tag{23}$$

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

$$\begin{aligned}
&P(w^{(d)}_m = w|z^{(d)}_m = k, c^{(d)} = c, \mathcal{W}_{\backslash d,m}, \mathcal{Z}_{\backslash d,m}, \mathcal{C}_{\backslash d}, \delta, \boldsymbol{n}, \alpha, \boldsymbol{m}) \\
&\propto \frac{M^{WK}_{wk,\backslash d,m} + \delta n_w}{\sum_{w=1}^{W} M^{WK}_{wk,\backslash d,m} + \delta}
\end{aligned} \tag{24}$$

References

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

Cox, D. R. (1992). Regression models and life-tables. In *Breakthroughs in statistics*, pages 527–541. Springer.

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.

Vu, D. Q., Hunter, D., Smyth, P., and Asuncion, A. U. (2011). Continuous-time regression models for longitudinal networks. In Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., and Weinberger, K., editors, *Advances in Neural Information Processing Systems 24*, pages 2492–2500. Curran Associates, Inc.