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 18, 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, $i \rightarrow j$ 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, by fixing $\lambda_0 = 1$, 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 the deviation from 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

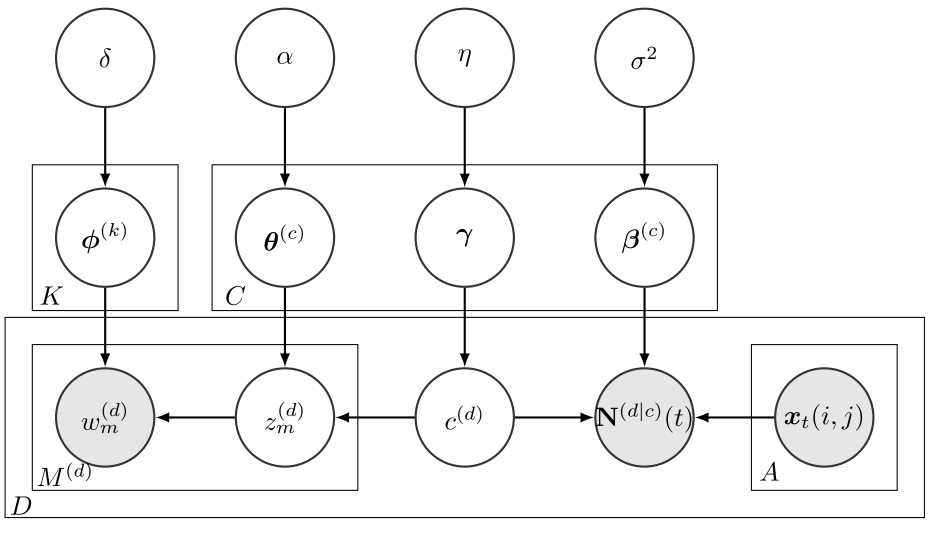

Figure 1: Plate notation of IPTM

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) as above, 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. However, one different thing from the existing specification 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. $\text{send}_t(i, j) = \sum\limits_{d:t^{(d)}<t} I\{i \rightarrow j\} \cdot g(t - t^{(d)})$
2. $\text{receive}_t(i, j) = \sum\limits_{d:t^{(d)}<t} I\{j \rightarrow i\} \cdot g(t - t^{(d)})$
3. $\text{2-send}_t(i, j) = \sum\limits_{h \neq i,j} \Big(\sum\limits_{d:t^{(d)}<t} I\{i \rightarrow h\} \cdot g(t - t^{(d)})\Big)\Big(\sum\limits_{d:t^{(d)}<t} I\{h \rightarrow j\} \cdot g(t - t^{(d)})\Big)$
4. $\text{2-receive}_t(i, j) = \sum\limits_{h \neq i,j} \Big(\sum\limits_{d:t^{(d)}<t} I\{h \rightarrow i\} \cdot g(t - t^{(d)})\Big)\Big(\sum\limits_{d:t^{(d)}<t} I\{j \rightarrow h\} \cdot g(t - t^{(d)})\Big)$

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

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

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 $\lambda = 1$ but it may vary based on the nature of document.

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

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

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:

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

3.2 Resampling $\mathcal{Z}$

Next, the new values of $z^{(d)}_m$ 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:

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

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:

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

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^2_B (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).

3.4 Pseudocode

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

Algorithm 5 MCMC($I, n_1, n_2, n_3, \delta_B$)

```
set initial values C^(0), Z^(0), and B^(0)
for i=1 to I do

    for n=1 to n_1 do
        fix Z = Z^(i-1) and B = B^(i-1)
        for d=1 to D do
            calculate p^C | z^(d), β^(c^(d)) = (p_1, ..., p_C), where p_c = exp(Eq. (11) corresponding to c)
            draw c^(d) ~ multinomial(p^C)
        end
    end

    for n=1 to n_2 do
        fix C = C^(i) and B = B^(i-1)
        for d=1 to D do
            for m=1 to M^(d) do
                calculate p^Z | c^(d), β^(c^(d)) = (p_1, ..., p_K), where p_k = exp(Eq. (13) corresponding to k)
                draw of z_m^(d) ~ multinomial(p^Z)
            end
        end
    end

    for n=1 to n_3 do
        fix C = C^(i), Z = Z^(i), and B^(0) = last value (n_3^th) of B^(i-1)
        for c=1 to C do
            draw β^(c) | C, Z, B^(n-1) using M-H algorithm in Section 3.3
        end
    end
end
summarize the results using:
the last value of C, the last value of Z, and the last n_3 length chain of B
```

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 $\delta = 5$. MCMC sampling was implemented based on the order and scheme illustrated in Section 3. We set the outer iteration number as $I = 100$, and inner iteration numbers as $n_1 = 10, n_2 = 10$, and $n_3 = 3500$, which took about 7.16 hours in total. In addition, after some experimentation, δ_B was set as 0.5, to ensure sufficient acceptance rate. In our case, the average acceptance rate for $\boldsymbol{\beta}$ was 0.526. 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, 500 were discarded as a burn-in, and every 3rd sample was taken for thinning. After these post-processing, MCMC diagnostic plots for IP1 and IP5 are attached in APPENDIX C as examples. There are some evidence of slightly bad mixing, which could be overcome if we sacrifice computation time and increase the size of thinning or iterations.

Below are the summary of IP-topic-word assignments. Each interaction pattern is paired with (a) posterior estimates 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 interaction pattern.

	IP1	**IP2**	**IP3**	**IP4**	**IP5**
intercept	-0.027 (0.031)	-0.278 (0.031)	-0.080 (0.032)	-0.024 (0.030)	0.165 (0.033)
send	0.626 (0.028)	0.255 (0.031)	0.394 (0.034)	-0.081 (0.031)	1.185 (0.027)
receive	-0.097 (0.029)	0.265 (0.027)	-0.021 (0.035)	0.070 (0.029)	0.777 (0.029)
2-send	0.070 (0.029)	0.110 (0.031)	-0.008 (0.032)	-0.037 (0.032)	0.021 (0.029)
2-receive	0.022 (0.030)	0.025 (0.032)	-0.043 (0.032)	-0.072 (0.034)	-0.017 (0.030)
sibling	-0.172 (0.033)	0.056 (0.033)	-0.204 (0.029)	-0.119 (0.030)	-0.076 (0.031)
cosibling	0.041 (0.028)	0.057 (0.030)	-0.071 (0.031)	-0.009 (0.031)	0.195 (0.033)

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

First, Table 2 summarizes the posterior means and standard errors for $\boldsymbol{\beta}^{(c)}$ corresponding to each interaction patterns. Below are the several examples of the interpretation of estimates, in the context of point process framework. Refer to Fig.3 of Perry and Wolfe (2013) attached below for better understanding of the interpretation.

- (**Intercept**) Assuming no history at all between the sender and receiver, the document is $\frac{e^{(-0.027)}}{e^{(-0.278)}} \approx 1.285$ times more likely to be IP1 relative to IP2.
- (**Send**) If i sends an email to j at time t, the likelihoods of i sends email of IP3 to j at time $t+1$ and $t+2$ are multiplied by $e^{(0.394 \times e^{(-1)})} \approx 1.156$ and $e^{(0.394 \times e^{(-2)})} \approx 1.055$, respectively.

(**Receive**) If j sends an email to i at time t, the likelihoods of i sends email of IP4 to j at time $t+1$ and $t+2$ are multiplied by $e^{(0.070\times e^{(-1)})} \approx 1.026$ and $e^{(0.070\times e^{(-2)})} \approx 1.010$, respectively.

- (**2-send**) If i sends an email to k at time t, and k sends an email to j at time $t+1$, then i sends email to j at time $t+2$ at a lower rate if IP4 (likelihood multiplied by $e^{(-0.037\times e^{(-1)}\times e^{(-1)})} \approx 0.995$), and at a higher rate if IP5 (likelihood multiplied by $e^{(0.021\times e^{(-1)}\times e^{(-1)})} \approx 1.003$).
  (**2-receive**) If j sends an email to k at time t, and k sends an email to i at time $t+1$, then i sends email to j at time $t+2$ at a lower rate if IP3 (likelihood multiplied by $e^{(-0.043\times e^{(-1)}\times e^{(-1)})} \approx 0.994$), and at a higher rate if IP2 (likelihood multiplied by $e^{(0.025\times e^{(-1)}\times e^{(-1)})} \approx 1.003$).
  (**sibling**) If k sends i and j an email at time t and $t+1$, respectively, then i sends an email to j at time $t+2$ at a lower rate if IP4 (likelihood multiplied by $e^{(-0.119\times e^{(-1)}\times e^{(-1)})} \approx 0.984$), and at a higher rate if IP2 (likelihood multiplied by $e^{(0.025\times e^{(-1)}\times e^{(-1)})} \approx 1.008$).
  (**cosibling**) If k receives an email from i and j at time t and $t+1$, respectively, then i sends an email to j at time $t+2$ at a lower rate if IP3 (likelihood multiplied by $e^{(-0.071\times e^{(-1)}\times e^{(-1)})} \approx 0.990$), and at a higher rate if IP5 (likelihood multiplied by $e^{(0.195\times e^{(-1)}\times e^{(-1)})} \approx 1.027$).

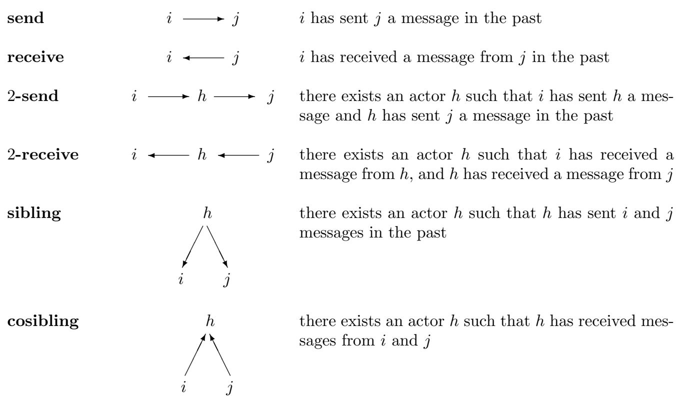

Fig. 3. Dynamic covariates to measure network effects

By examining the estimates in Table 2 and their corresponding interpretaiton, it seems that there exist significant differences in the effect of some dynamic network covariates, especially dyadic effects (send and receive). In order to see these differences more clearly, we compared the posterior distribution using the boxplots in Figure 2. Now, it is more apparent that the intercept and dyadic effects are different across the interaction patterns. Specifically, IP5 seems to be highly dependent on the history of dyadic interactions; that is, whether the sender had sent to or received from the receiver strongly affects the rate of their interactions.

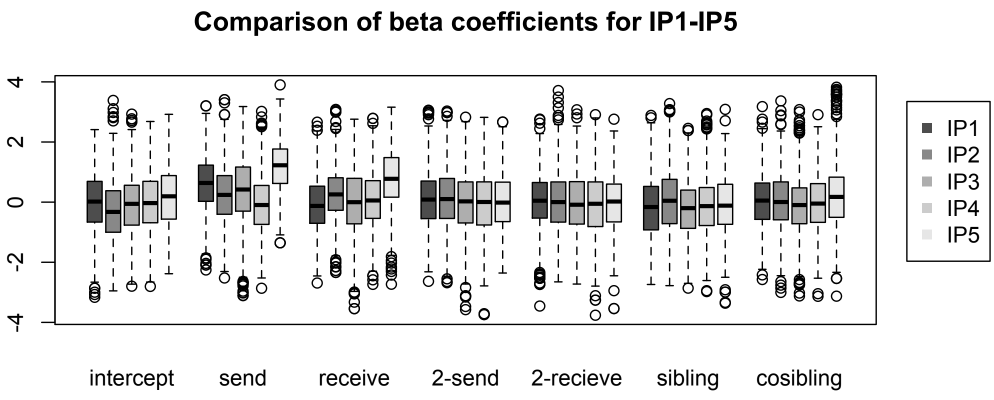

Figure 2: Posterior distribution of $\boldsymbol{\beta}^{(c)}$ for Vance county emails

Next, we scrutinize the topic-word distributions corresponding to each interaction patterns in Table 3. Unlike $\boldsymbol{\beta}$, there is no distinctive difference in the topic distributions $\mathcal{Z}$, given the assignment of interaction patterns to the documents $\mathcal{C}$. The top three topics given the interaction patterns are quite overlapping, possibly because the unconditional distribution of topics reveal that Topic 2, 1, 5, 6, and 7 dominate the whole topic-word assignments, with the probabilities of these top five topics sum up to 0.533. Moreover, the actual words are not significantly different, having several words like 'director', 'phones', 'church', 'street', or 'henderson' (county seat of Vance county) appeared repetitively across the topics as well as interaction patterns.

IP1 (61 emails)		**IP2** (38 emails)		**IP3** (57 emails)		**IP4** (56 emails)		**IP5** (57 emails)	
Topic 1	0.141	Topic 2	0.143	Topic 1	0.151	Topic 6	0.123	Topic 2	0.186
will	0.0313	will	0.0718	will	0.0410	will	0.0676	will	0.0842
october	0.0313	director	0.0276	director	0.0273	morning	0.0338	director	0.0330
description	0.0223	system	0.0276	description	0.0239	electronic	0.0338	opeartions	0.0293
director	0.0179	center	0.0221	latest	0.0239	church	0.0270	phone	0.0219
planning	0.0179	directory	0.0221	phones	0.0205	suite	0.0270	church	0.0183
public	0.0179	henderson	0.0166	henderson	0.0171	henderson	0.0203	october	0.0183
week	0.0179	operations	0.0166	attached	0.0171	enp	0.0203	street	0.0147
henderson	0.0134	suite	0.0166	suite	0.0137	hereto	0.0203	development	0.0147
phone	0.0134	meeting	0.0166	office	0.0137	description	0.0135	center	0.0147
development	0.0134	phone	0.0166	cem	0.0137	street	0.0135	system	0.0147
Topic 2	0.134	Topic 17	0.115	Topic 5	0.130	Topic 2	0.105	Topic 7	0.137
will	0.0657	will	0.0342	will	0.0873	operations	0.0476	will	0.0448
director	0.0329	week	0.0342	street	0.0278	director	0.0397	jail	0.0299
phones	0.0282	street	0.0274	operations	0.0278	suite	0.0317	description	0.0199
department	0.0235	message	0.0274	church	0.0238	message	0.0317	emergency	0.0199
heads	0.0235	church	0.0205	description	0.0198	street	0.0238	center	0.0199
review	0.0188	time	0.0205	phone	0.0198	planning	0.0238	director	0.0149
october	0.0188	lines	0.0205	system	0.0198	will	0.0238	henderson	0.0149
system	0.0188	operations	0.0137	phones	0.0198	emergency	0.0238	suite	0.0149
extension	0.0188	meeting	0.0137	extension	0.0198	office	0.0238	phone	0.0149
description	0.0141	system	0.0137	center	0.0159	today	0.0238	office	0.0149
Topic 6	0.102	Topic 5	0.107	Topic 6	0.116	Topic 16	0.100	Topic 5	0.107
will	0.0432	will	0.0889	will	0.0982	message	0.0417	street	0.0446
director	0.0370	street	0.0222	fax	0.0223	will	0.0333	operations	0.0255
street	0.0247	system	0.0222	henderson	0.0179	church	0.0250	meeting	0.0255
electronic	0.0247	questions	0.0222	system	0.0179	operations	0.0250	folks	0.0255
heads	0.0247	department	0.0222	copies	0.0179	electronic	0.0250	coming	0.0255
fax	0.0185	chapter	0.0222	description	0.0134	coming	0.0250	henderson	0.0191
church	0.0185	phones	0.0222	church	0.0134	latest	0.0250	planning	0.0191
instructions	0.0185	cutting	0.0222	meeting	0.0134	director	0.0167	will	0.0191
time	0.0185	director	0.0148	heads	0.0134	henderson	0.0167	emergency	0.0191
message	0.0185	henderson	0.0148	sure	0.0134	street	0.0167	system	0.0191

Table 3: Summary of MCMC sampling results for Vance county emails. Each interaction pattern is shown with the top 3 topics and 10 words with the highest probability conditioned on that topic.

Although Vance county email data did not display distinctive idiosyncrasy across the interaction patterns and their corresponding topic assignments, it is not surprising because Vance county is a small county (land area: 253.52 sq. mi and population: 44,998), and our exploratory data analysis did not find any significant change in the email exchanges of department managers during the period of hurricand Sandy. Yet, it is definitely worthwhile to further look at this in terms of showing the applicability of interaction-partitioned topic model (IPTM), in case of email data. In the next section, we apply the same method to another corpus, Dare county email data, in hope of finding more interesting results and also comparing the outcomes between the two counties.

4.2 Dare county email data

Application to Dare county email data was conducted in the exactly same manner as previous section. However, the size of Dare county data is much larger than that of Vance county data; that is, Dare county data contains 4,845 emails between 27 actors, including 2,907 vocabulary in total, after post-processing as before (i.e. multicast and word count > 0). Considering the huge expected computation time, we specified smaller number of topics, $K = 10$, and smaller number of interaction patterns, $C = 3$. Except those, all other parameters such as Dirichlet priors or MCMC parameters were identically specified as Section 4.1.

	IP1	**IP2**	**IP3**	**IP4**	**IP5**
intercept	0.900	-0.451	-0.795	0.259	-0.856
send	0.409	-0.536	2.755	0.459	1.972
receive	0.077	-0.616	1.125	-0.185	0.721
2-send	0.194	-1.201	0.062	-0.264	-0.997
2-receive	-1.823	-0.411	-1.026	-1.203	0.873
sibling	-0.678	0.981	0.638	0.302	-0.403
cosibling	-0.344	-1.402	-1.753	-1.333	-1.141

Table 4: Summary of posterior $\boldsymbol{\beta}^{(c)}$ estimates for Vance county emails

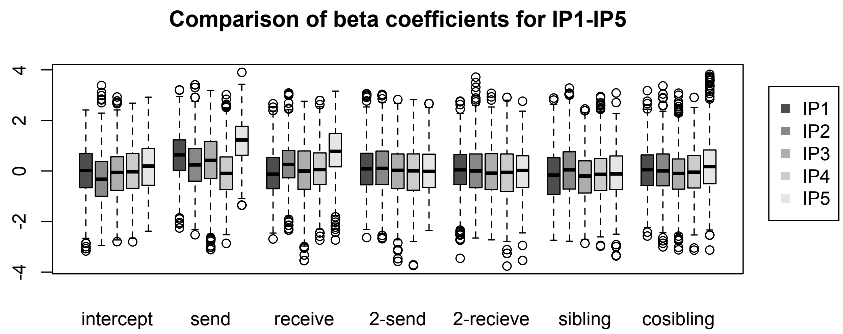

Figure 3: Posterior distribution of $\boldsymbol{\beta}^{(c)}$ for Vance county emails

IP1 (56 emails)		**IP2** (47 emails)		**IP3** (69 emails)		**IP4** (32 emails)		**IP5** (65 emails)	
Topic 15	0.212	Topic 18	0.181	Topic 9	0.218	Topic 14	0.302	Topic 1	0.215
operations	0.0563	phone	0.0395	electronic	0.0789	will	0.1250	will	0.0523
center	0.0387	development	0.0395	heads	0.0366	phones	0.0724	directory	0.0494
office	0.0352	henderson	0.0316	ncgs	0.0338	week	0.0395	jail	0.0465
communications	0.0317	planning	0.0316	attachments	0.0310	system	0.0373	extension	0.0436
enp	0.0317	description	0.0277	review	0.0282	cutting	0.0307	will	0.0262
suite	0.0282	fax	0.0277	chapter	0.0282	rest	0.0307	attached	0.0262
emergency	0.0282	suite	0.0237	pursuant	0.0254	october	0.0285	folks	0.0262
henderson	0.0563	e-mail	0.0237	department	0.0225	provided	0.0285	technology	0.0262
street	0.0246	attached	0.0237	tomorrow	0.0225	department	0.0241	excel	0.0262
director	0.0211	director	0.0198	time	0.0197	phone	0.0219	director	0.0233
Topic 4	0.208	Topic 19	0.171	Topic 7	0.208	Topic 12	0.298	Topic 16	0.214
emergency	0.0466	description	0.0546	message	0.0531	will	0.1064	will	0.1023
suite	0.0430	henderson	0.0336	request	0.0501	phones	0.0643	extension	0.0409
director	0.0358	director	0.0252	electronic	0.0442	october	0.0377	folks	0.0322
fax	0.0323	street	0.0252	time	0.0295	training	0.0377	directory	0.0292
operations	0.0323	church	0.0210	review	0.0295	department	0.0310	call	0.0263
office	0.0287	phone	0.0210	department	0.0295	provided	0.0310	latest	0.0263
cem	0.0287	goldvance...	0.0210	response	0.0265	system	0.0288	cutover	0.0205
henderson	0.0215	fax	0.0168	manager	0.0236	week	0.0288	number	0.0205
will	0.0179	suite	0.0168	director	0.0236	cutting	0.0266	henderson	0.0175
phone	0.0143	project	0.0168	public	0.0206	day	0.0222	advised	0.0175
Topic 8	0.199	Topic 3	0.161	Topic 11	0.204	Topic 17	0.271	Topic 2	0.202
operations	0.0489	description	0.0446	department	0.0631	will	0.0854	will	0.1022
emergency	0.0489	e-mail	0.0357	message	0.0511	phone	0.0390	latest	0.0310
director	0.0338	developement	0.0313	records	0.0420	october	0.0390	extension	0.0279
henderson	0.0338	director	0.0268	heads	0.0360	week	0.0341	jail	0.0279
fax	0.0338	goldvance...	0.0268	will	0.0330	phones	0.0268	updated	0.0248
street	0.0301	church	0.0223	electronic	0.0330	folks	0.0268	director	0.0217
center	0.0301	henderson	0.0179	pursuant	0.0330	rest	0.0244	attached	0.0217
office	0.0263	street	0.0179	chapter	0.0300	senior	0.0244	coming	0.0171
church	0.0188	phone	0.0179	manager	0.0270	system	0.0220	henderson	0.0163
enp	0.0188	semprius	0.0179	response	0.0240	directory	0.0220	cutover	0.0154

Table 5: Summary of MCMC sampling results for Vance county emails. Each interaction pattern is shown with the top 3 topics and words that have the highest probability conditioned on that topic.

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 $P(z_m^{(d)} = k, w_m^{(d)} = w, c^{(d)} = c | \mathcal{W}_{\backslash d}, \mathcal{Z}_{\backslash d}, \mathcal{C}_{\backslash d}, \delta, \boldsymbol{n}, \alpha, \boldsymbol{m})$, From Bayes' theorem and Gamma identity $\Gamma(k+1) = k\Gamma(k)$,

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

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

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

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

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

APPENDIX C: MCMC Diagnostics for Vance county emails

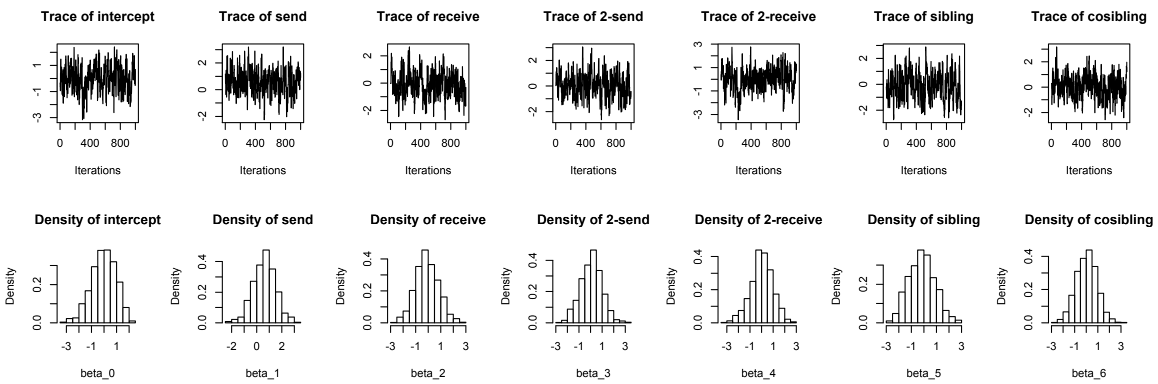

Figure 4: Traceplots and density plots of $\boldsymbol{\beta}^{(1)}$

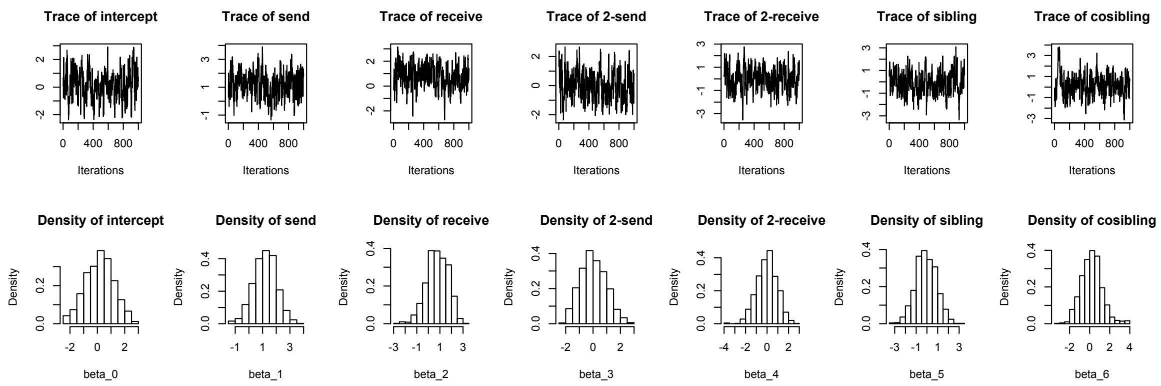

Figure 5: Traceplots and density plots of $\boldsymbol{\beta}^{(5)}$

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.

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.

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.