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

Bomin Kim

May 31, 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, $\lambda^{(c)}(t)$, 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 $N_{ij}^{(c)}(t)$, a counting process denoting the number of edges (emails) of interaction pattern c, from actor i to actor j up to time t. NOTE: We use the partition c since we expect that some interaction patterns have little variation among the pairs of actors (e.g. broadcasting), while some have large variation (e.g. meeting scheduling, personal affairs).

Combining the individual counting processes of all potential edges, $\mathbf{N}^{(c)}(t)$ is the multivariate counting process with $\mathbf{N}^{(c)}(t) = (N_{ij}^{(c)}(t) : i, j \in 1, ..., A, i \neq j)$. 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:

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

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:

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

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 $\mathcal{J}_{(i,t)}^{(c)}$ is the predictable receiver set of sender i at time t within all actors $A^{(c)}$. NOTE: We assume that all possible actor set $A^{(c)}$ varies depending on the interaction pattern (e.g. confidential communication do not move outside of certain actors).

2.2 Generative Process

The generative process of this model follows that of Blei et al. (2003) and 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 difference is that each documents is connected to one interaction pattern, and the topic distributions vary depending on the interaction pattern. Following are the generative process for each document in a corpus D and its plate notation (Figure 1):

1. $\phi^{(k)} \sim \text{Dir}(\delta, \mathbf{n})$
   - 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 [**See Algorithm 1**].

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 exist differences in the degree of influence from the network statistics $\boldsymbol{x}_t(i,j)$ that rely on the history of interactions.

   (b) $\boldsymbol{N}^{(c)}(t) \sim \text{CP}(\boldsymbol{\lambda}^{(c)}(t))$
       - The actual update of the counting process $N_{ij}^{(c)}(t)$ of the email d is
       $$N_{i^{(d)}j^{(d)}}^{(c^{(d)})}(t^{(d)}) = N_{i^{(d)}j^{(d)}}^{(c^{(d)})}(t^{(d)}-) + 1.$$

3. For each of the D documents [**See Algorithm 3**]:

   (a) $c^{(d)} \sim \text{Multinomial}(\gamma)$
       - Each document d is associated with one "interaction pattern" among C different types, with parameter γ.

   (b) $\boldsymbol{\theta}^{(d|c^{(d)})} \sim \text{Dir}(\alpha_{c^{(d)}}, \mathbf{m})$
       - Each email has a discrete distribution over topics $\boldsymbol{\theta}^{(d|c)}$, since the topic proportions for documents in the same cluster are drawn from the same distribution. Thus, A Dirichlet prior with cluster-specific concentration parameter $\alpha^{(c)}$ is placed.

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

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

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

Algorithm 1 Topic Word Distributions

---
for $k=1$ to K **do**
  | draw $\phi^{(k)} \sim \text{Dir}(\delta, \mathbf{n})$
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 $\lambda_{ij}^{(c)}(t) = \lambda_0 \cdot \exp\left\{\boldsymbol{\beta}^{(c)T} \boldsymbol{x}_t(i,j)\right\} \cdot 1\{j \in \mathcal{J}_{(i,t)}^{(c)}\}$
            end
            else
                set $\lambda_{ij}^{(c)}(t) = 0$
            end
        end
    end
    draw $\boldsymbol{N}^{(c)}(t) \sim \text{CP}(\boldsymbol{\lambda}^{(c)}(t))$
end

Algorithm 3 Document-Interaction Pattern Assginments

---

for _d=1 to D_ **do**
    draw $c^{(d)} \sim \text{Multinomial}(\gamma)$
    draw $\boldsymbol{\theta}^{(d|c^{(d)})} \sim \text{Dir}(\alpha^{c^{(d)}}, \mathbf{m})$

end

Algorithm 4 Tokens

---

for _d=1 to D_ **do**
    set $\bar{M}^{(d)} = \max(1, M^{(d)})$
    for _m=1 to_ $\bar{M}^{(d)}$ **do**
        draw $z_m^{(d)} \sim \text{Multinomial}(\boldsymbol{\theta}^{(d|c^{(d)})})$
        if $M^{(d)} \neq 0$ **then**
            draw $w_m^{(d)} \sim \text{Multinomial}(\phi^{(z_m^{(d)})})$
        end
    end
end

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.

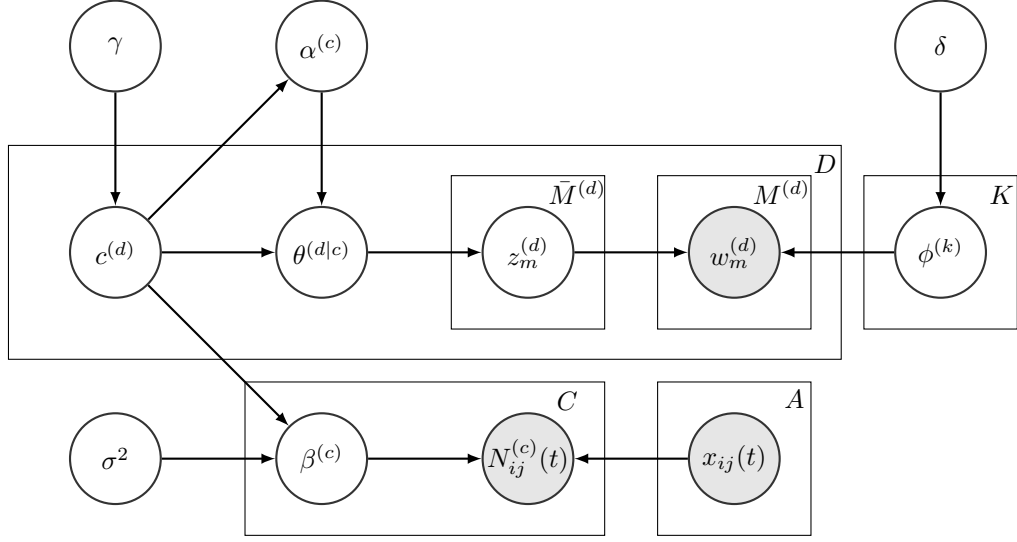Following Perry and Wolfe (2013), we use 6 effects as components of $\boldsymbol{x}_t(i,j)$.

Figure 1: Plate notation of IPTM

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. However, we take different approach by defining 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.

1. $\text{send}_t(i,j) = \sum\limits_{d:t_d<t} I\{i \to j\} \cdot g(t - t_d)$

2. $\text{receive}_t(i,j) = \sum\limits_{d:t_d<t} I\{j \to i\} \cdot g(t - t_d)$

3. $\text{2-send}_t(i,j) = \sum\limits_{h\neq i,j} \Big(\sum\limits_{d:t_d<t} I\{i \to h\} \cdot g(t - t_d) \Big)\Big(\sum\limits_{d:t_d<t} I\{h \to j\} \cdot g(t - t_d) \Big)$

4. $\text{2-receive}_t(i,j) = \sum\limits_{h\neq i,j} \Big(\sum\limits_{d:t_d<t} I\{h \to i\} \cdot g(t - t_d) \Big)\Big(\sum\limits_{d:t_d<t} I\{j \to h\} \cdot g(t - t_d) \Big)$

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

2.4 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 counting process $\mathcal{N} = \{\boldsymbol{N}(t_d)\}_{d=1}^D$. $\mathcal{X} = \{\boldsymbol{x}_{t_d}(i,j)\}_{d=1}^D$ is the metadata, and the latent variables are $\Phi = \{\phi^{(t)}\}_{t=1}^T, \Theta = \{\boldsymbol{\theta}^{(d|c)}\}_{d=1}^D, \mathcal{Z} = \{\boldsymbol{z}^{(d)}\}_{d=1}^D, \mathcal{C} = \{c^{(d)}\}_{d=1}^D$, and $\mathcal{B} = \{\boldsymbol{\beta}^{(c)}\}_{c=1}^C$.

Q1. What we observe is $\{\boldsymbol{N}(t_d)\}_{d=1}^D$, not $\{\boldsymbol{N}^{(c)}(t_d)\}_{d=1}^D$ which is of our interest. Which one should we use?
Q2. Do we use $\mathcal{A}$ as our metadata (such as CPME)? This requires the assumption that we know who the author is, but it is actually not in IPTM.
Q3. We now use $\alpha^{(c)}$ for Dirichlet prior. However, they are still hyper-parameter. So, is it fine to define $\boldsymbol{\alpha} = \{\alpha^{(c)}\}_{c=1}^C$ and use $\boldsymbol{\alpha}$ as before?

Below is the the big joint distribution

$$
\begin{aligned}
&P(\Phi, \Theta, \mathcal{Z}, \mathcal{W}, \mathcal{C}, \mathcal{B}, \mathcal{N} | \mathcal{X}, \delta, \boldsymbol{n}, \alpha, \boldsymbol{m}) \\
&= P(\mathcal{W}|\mathcal{Z}, \Phi) P(\mathcal{N}|\mathcal{X}, \mathcal{C}, \mathcal{B}) P(\mathcal{Z}|\Theta) P(\Phi|\delta, \boldsymbol{n}) P(\Theta, \alpha, \boldsymbol{m}) P(\mathcal{C}) P(\mathcal{B})
\end{aligned} \tag{3}
$$

Now we can integrate out Φ and Θ in latent Dirichlet allocation by applying Dirichlet-multinomial conjugacy as we did in CPME.

$$
\propto P(\mathcal{Z}|\delta, \boldsymbol{n}, \alpha, \boldsymbol{m}) P(\mathcal{W}|\mathcal{Z}) P(\mathcal{N}|\mathcal{X}, \mathcal{C}, \mathcal{B}) P(\mathcal{C}) P(\mathcal{B}) \tag{4}
$$

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

$$
P(\mathcal{Z}, \mathcal{C}, \mathcal{B} | \mathcal{W}, \mathcal{N}, \mathcal{X}, \delta, \boldsymbol{n}, \alpha, \boldsymbol{m}) \propto P(\mathcal{Z}, \mathcal{C}, \mathcal{B}, \mathcal{W}, \mathcal{N} | \mathcal{X}, \delta, \boldsymbol{n}, \alpha, \boldsymbol{m}) \tag{5}
$$

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

2.4.1 Resampling $\mathcal{Z}$

First, the new values of $z_m^{(d)}$ are sampled using the conditional posterior probability of being topic k:

$$
\begin{aligned}
&P(z_m^{(d)} = k | \mathcal{W}, \mathcal{Z}_{\backslash d,m}, \mathcal{C}, \mathcal{B}, \mathcal{N}, \mathcal{X}) \\
&\propto P(z_m^{(d)} = k, w_m^{(d)}, N(t_d) | \mathcal{W}_{\backslash d,m}, \mathcal{Z}_{\backslash d,m}, \mathcal{C}, \mathcal{B}, \mathcal{N}_{\backslash d}, \mathcal{X}) \\
&= P(z_m^{(d)} = k, w_m^{(d)} | \mathcal{W}_{\backslash d,m}, \mathcal{Z}_{\backslash d,m}, C, \delta, \boldsymbol{n}, \alpha, \boldsymbol{m}) \times P(N(t_d) | \mathcal{C}, \mathcal{B}, \mathcal{N}_{\backslash d}, \mathcal{X})
\end{aligned} \tag{6}
$$

where the subscript "$\backslash d, n$" denotes the exclsuion of position n in email d. In the last line of equation (6), the left part is the contribution of LDA, while the right part is the contribution of counting process. Similar to CPME, we can write the conditional probability in two ways based on the number of words in the email:

① The case where the email contains words $(M^{(d)} > 0)$

$$\propto (\bar{M}_{\backslash d,m}^{(k|d)} + \alpha^{(c^{(d)})} m_k) \cdot \frac{\bar{M}_{\backslash d,m}^{(w_m^{(d)}|k)} + \beta n_v}{\bar{M}_{\backslash d,m}^{(k)} + \beta}$$

$$\times \exp\Big\{-\Big(\sum_{i \in A^{(c^{(d)})}} \sum_{j \in \mathcal{J}_{(i,t)}^{(c^{(d)})}} \lambda_{ij}^{(c^{(d)})}\Big) t_d\Big\} \cdot \prod_{i \in A^{(c^{(d)})}} \prod_{j \in \mathcal{J}_{(i,t)}^{(c^{(d)})}} \frac{(\lambda_{ij}^{(c^{(d)})} t_d)^{n^{(ij)}}}{n^{(ij)}!}$$

$$(7)$$

with $\lambda_{ij}^{(c)} = \lambda_0 \cdot \exp(\boldsymbol{\beta}^{(c)T} \boldsymbol{x}_t(i,j)) \cdot 1\{j \in \mathcal{J}_{(i,t)}^{(c)}\}$. Note that the upper part of (7) comes from LDA, which is the same as CPME, while the lower part of (7) represents the multivariate counting process contribution based on Zocher (2006).

Q1. If we assume we know the author (Q2 in previous page), we can drop the sum/prod for i and only use j as CPME.
Q2. The term $1\{j \in \mathcal{J}_{(i,t)}^{(c)}\}$ is actually repetitive. But if I just use $A^{(c)}$ for j, the numerator in the last part of (7) becomes zero.

To avoid numerical problem coming form the large product, we can take the log of the equation (7) and work in log space using the log-conditional probability:

$$\log\Big\{(\bar{M}_{\backslash d,m}^{(k|d)} + \alpha^{(c^{(d)})} m_k) \cdot \frac{\bar{M}_{\backslash d,m}^{(w_m^{(d)}|k)} + \beta n_v}{\bar{M}_{\backslash d,m}^{(k)} + \beta}\Big\}$$

$$- t_d \Big(\sum_{i \in A^{(c^{(d)})}} \sum_{j \in \mathcal{J}_{(i,t)}^{(c^{(d)})}} \lambda_{ij}^{(c^{(d)})}\Big) + \sum_{i \in A^{(c^{(d)})}} \sum_{j \in \mathcal{J}_{(i,t)}^{(c^{(d)})}} \Big(n^{(ij)} \log(\lambda_{ij}^{(c^{(d)})} t_d) - \log(n^{(ij)}!)\Big)$$

$$(8)$$

Here, since we still contain product form in the later part of equaiton (8), we can apply Stirling's approximation for factorials, $\log(n!) \approx \log(n) - n + 1$, and obtain approximate log-conditional probability:

$$\approx \log\Big\{(\bar{M}_{\backslash d,m}^{(k|d)} + \alpha^{(c^{(d)})} m_k) \cdot \frac{\bar{M}_{\backslash d,m}^{(w_m^{(d)}|k)} + \beta n_v}{\bar{M}_{\backslash d,m}^{(k)} + \beta}\Big\}$$

$$- t_d \Big(\sum_{i \in A^{(c^{(d)})}} \sum_{j \in \mathcal{J}_{(i,t)}^{(c^{(d)})}} \lambda_{ij}^{(c^{(d)})}\Big) + \sum_{i \in A^{(c^{(d)})}} \sum_{j \in \mathcal{J}_{(i,t)}^{(c^{(d)})}} \Big(n^{(ij)} \log(\lambda_{ij}^{(c^{(d)})} t_d) - n^{(ij)} \log(n^{(ij)}) - n^{(ij)} + 1\Big)$$

$$= \log\Big\{(\bar{M}_{\backslash d,m}^{(k|d)} + \alpha^{(c^{(d)})} m_k) \cdot \frac{\bar{M}_{\backslash d,m}^{(w_m^{(d)}|k)} + \beta n_v}{\bar{M}_{\backslash d,m}^{(k)} + \beta}\Big\}$$

$$- t_d \Big(\sum_{i \in A^{(c^{(d)})}} \sum_{j \in \mathcal{J}_{(i,t)}^{(c^{(d)})}} \lambda_{ij}^{(c^{(d)})}\Big) + \sum_{i \in A^{(c^{(d)})}} \sum_{j \in \mathcal{J}_{(i,t)}^{(c^{(d)})}} \Big(n^{(ij)} \big(\log(\lambda_{ij}^{(c^{(d)})} t_d) - \log(n^{(ij)}) - 1\big) + 1\Big)$$

$$(9)$$

with $\lambda_{ij}^{(c)} = \lambda_0 \cdot \exp(\boldsymbol{\beta}^{(c)T} \boldsymbol{x}_t(i,j)) \cdot 1\{j \in \mathcal{J}_{(i,t)}^{(c)}\}$, same as before.

② The case where the email does not contain words ($M^{(d)} = 0$) In this case, we assign a dummy token that can take a topic assignment as a way to associate the document to an interaction pattern. Since the LDA contributions becomes constant, thus, we can igore it an only use the counting process part:

$$P(z_m^{(d)} = k | \mathcal{W}, \mathcal{Z}_{\backslash d,m}, \mathcal{C}, \mathcal{B}, \mathcal{N}, \mathcal{X})$$

$$\propto \exp\Big\{ -\big(\sum_{i \in A^{(c^{(d)})}} \sum_{j \in \mathcal{J}_{(i,t)}^{(c^{(d)})}} \lambda_{ij}^{(c^{(d)})} \big) t_d \Big\} \cdot \prod_{i \in A^{(c^{(d)})}} \prod_{j \in \mathcal{J}_{(i,t)}^{(c^{(d)})}} \frac{(\lambda_{ij}^{(c^{(d)})} t_d)^{n^{(ij)}}}{n^{(ij)}!}$$

$$(10)$$

Again, if we take th log, we end up with the approximate log-conditional probability:

$$\approx -t_d \big(\sum_{i \in A^{(c^{(d)})}} \sum_{j \in \mathcal{J}_{(i,t)}^{(c^{(d)})}} \lambda_{ij}^{(c^{(d)})} \big) + \sum_{i \in A^{(c^{(d)})}} \sum_{j \in \mathcal{J}_{(i,t)}^{(c^{(d)})}} \Big(n^{(ij)} \big(\log(\lambda_{ij}^{(c^{(d)})} t_d) - \log(n^{(ij)}) - 1 \big) + 1 \Big)$$

$$(11)$$

2.4.2 Resampling $\mathcal{C}$

The next variable we are going to resample is the document-interaction pattern assignments. The conditional posterior probability we want to calculate is:

$$P(c^{(d)} = c | \mathcal{Z}, \mathcal{W}, \mathcal{C}_{\backslash d}, \mathcal{B}, \mathcal{N}, \mathcal{X}, \delta, \boldsymbol{n}, \alpha, \boldsymbol{m})$$
$$\propto P(c^{(d)} = c, N(t_d) | \mathcal{Z}, \mathcal{W}, \mathcal{C}_{\backslash d}, \mathcal{B}, \mathcal{N}_{\backslash d} \mathcal{X}, \delta, \boldsymbol{n}, \alpha, \boldsymbol{m})$$
$$= P(c^{(d)} = c) \cdot P(N(t_d) | \mathcal{Z}, \mathcal{W}, \mathcal{C}_{\backslash d}, \mathcal{B}, \mathcal{N}_{\backslash d}, \mathcal{X}, \delta, \boldsymbol{n}, \alpha, \boldsymbol{m})$$
$$\propto P(N(t_d) | \mathcal{C}_{\backslash d}, \mathcal{B}, \mathcal{N}_{\backslash d}, \mathcal{X}) \qquad (12)$$
$$= \exp\Big\{ -\big(\sum_{i \in A^{(c^{(d)})}} \sum_{j \in \mathcal{J}_{(i,t)}^{(c^{(d)})}} \lambda_{ij}^{(c^{(d)})} \big) t_d \Big\} \cdot \prod_{i \in A^{(c^{(d)})}} \prod_{j \in \mathcal{J}_{(i,t)}^{(c^{(d)})}} \frac{(\lambda_{ij}^{(c^{(d)})} t_d)^{n^{(ij)}}}{n^{(ij)}!}$$

which is the same as the formula used for resampling of $\mathcal{Z}$. Of course, we can speed up the computation time by taking the log of equation (12), which is identical to (11).

2.4.3 Resampling $\mathcal{B}$

Finally, we wan to update the interaction pattern parameter $\boldsymbol{\beta}^{(c)}$. For this, we will use the Metropolis-Hastings algorithm with a proposal density Q being the multivariate Gaussian distribution, with variance σ^2, 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-ration and obtain the acceptance probability equal to:

$$\text{Acceptance Probability} = \begin{cases} \frac{P(\mathcal{B}' | \mathcal{Z}, \mathcal{W}, \mathcal{C}, \mathcal{N}, \mathcal{X})}{P(\mathcal{B} | \mathcal{Z}, \mathcal{W}, \mathcal{C}, \mathcal{N}, \mathcal{X})} & \text{if} < 1 \\ 1 & \text{else} \end{cases} \qquad (13)$$

After factorization, we get

$$\frac{P(\mathcal{B}'|\mathcal{Z},\mathcal{W},\mathcal{C},\mathcal{N},\mathcal{X})}{P(\mathcal{B}|\mathcal{Z},\mathcal{W},\mathcal{C},\mathcal{N},\mathcal{X})} = \frac{P(\mathcal{N}|\mathcal{B}',\mathcal{Z},\mathcal{W},\mathcal{C},\mathcal{X})P(\mathcal{B}')}{P(\mathcal{N}|\mathcal{B},\mathcal{Z},\mathcal{W},\mathcal{C},\mathcal{X})P(\mathcal{B})}$$
$$= \frac{P(\mathcal{N}|\mathcal{B}',\mathcal{C},\mathcal{X})P(\mathcal{B}')}{P(\mathcal{N}|\mathcal{B},\mathcal{C},\mathcal{X})P(\mathcal{B})} \qquad (14)$$

and apply the equation (12) for $P(\mathcal{N}|\mathcal{B},\mathcal{C},\mathcal{X})$, which is the probability of observing the edges under the interaction pattern parameters. For $P(\mathcal{B})$, we select a multivarate Gaussian priors as mentioned earlier. After taking the log, the equation we will use for the inference is:

$$\log(P(\mathcal{B}')) - \log(P(\mathcal{B}))$$
$$- t_d\Big(\sum_{i \in A^{(c^{(d)})}} \sum_{j \in \mathcal{J}_{(i,t)}^{(c^{(d)})}} \lambda_{ij}^{(c^{(d)})}\Big) + \sum_{i \in A^{(c^{(d)})}} \sum_{j \in \mathcal{J}_{(i,t)}^{(c^{(d)})}} \Big(n^{(ij)}\big(\log(\lambda_{ij}^{(c^{(d)})}t_d) - \log(n^{(ij)}) - 1\big) + 1\Big), \qquad (15)$$

and then the log of the acceptance ratio we have is:

$$\log(\text{Acceptance Probability}) = \min(\text{Log of (13)}, 0) \qquad (16)$$

To determine whether we accept the proposed update or not, we take the usual approacy, by comparing the log of acceptance ratio we have to the log of a sample from uniform(0,1).

3 Preliminary Analysis

Hurricane Sandy was the most destructive hurricane in 2012, which hit North Carolina on late October (October 28, Governor Bev Perdue declared a state of emergency in 24 western counties due to snow and strong winds). In our dataset, there are three counties which cover the date of Hurricane Sandy (October 22, 2012 – November 2, 2012), so we focus on the three counties, since the timestamp of email in this case is much more important than usual case without any disastrous event.

3.1 Dare County

Period	Before Sandy	During Sandy	After Sandy	Overall
# emails	1933	1563	1467	4963

Table 1: Summary of Dare county email data based on time period

Before Sandy ranges from 2012-09-01 to 2012-10-21 (7 weeks), During Sandy ranges from 2012-10-22 to 2012-11-02 (2 weeks), and After Sandy ranges from 2012-11-03 to 2012-11-30 (4 weeks).
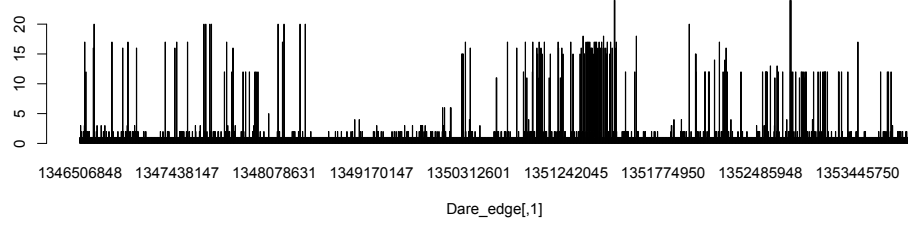
3.2 Lenoir County

Before Sandy ranges from 2012-10-01 to 2012-10-21 (3 weeks), During Sandy ranges from 2012-10-22 to 2012-11-02 (2 weeks), and After Sandy ranges from 2012-11-03 to 2012-12-31 (8 weeks).

Figure 2: Frequency of Dare county emails from 2012-09-01 to 2012-11-30

Time Interval	send	receive
$[-\infty, t)$	2.128, 2.659, 2.355, 2.919	0.292, 0.257, 0.047, 0.110
$[t - 30m, t)$	0.262, -0.064, 0.782, 0.317	2.087, 1.287 , 2.346, 1.870
$[t - 2h, t - 30m)$	0.383, 0.157 , 0.024, -0.045	0.553, 0.082, 0.794, 0.269
$[t - 8h, t - 2h)$	0.816, 0.054 , 0.077, 0.381	-0.221, 0.048, 0.298, -0.012
$[t - 32h, t - 8h)$	0.085, 0.014, 0.228, 0.070	0.101, 0.017, -0.033, 0.019
$[t - 5.33d, t - 32h)$	0.103, 0.025, 0.092, 0.008	-0.027, -0.016, -0.033, -0.009
$[t - 21.33d, t - 5.33d)$	0.052, 0.000, 0.059, 0.010	0.013, 0.030 , -0.016, 0.013
$[-\infty, t - 21.33d)$	0.052, 0.103, 0.027, 0.021	0.008, 0.000, 0.020, -0.005

Table 2: Estimated coefficients and approximate standard errors for dyadic effects of Dare county data (before Sandy, during Sandy, after Sandy, overall)

3.3   Vance County

Before Sandy ranges from 2012-09-04 to 2012-10-21 (7 weeks), During Sandy ranges from 2012-10-22 to 2012-11-02 (2 weeks), and After Sandy ranges from 2012-11-03 to 2012-11-30 (4 weeks).

References

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

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

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.

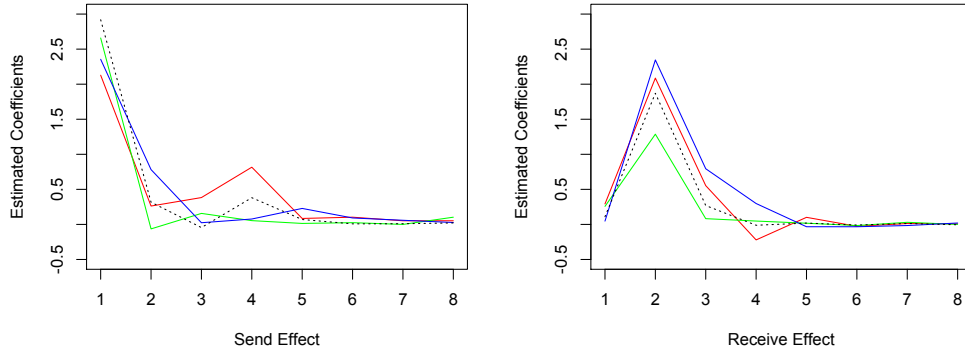Zocher, M. (2006). *Multivariate counting processes*. Techn. Univ., Inst. für Mathematische Stochastik.

Figure 3: Comparison of Send (left) and Receive (right) effect based on periods in Table 1. (Red=Before, Green=During, Blue=After, and dot=Overall)

Period	Before Sandy	During Sandy	After Sandy	Overall
# emails	216	83	302	601

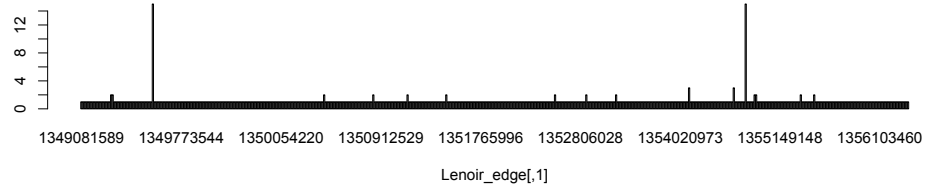Table 3: Summary of Lenoir county email data based on time period

Figure 4: Frequency of Lenoir county emails from 2012-10-01 to 2012-12-31

Period	Before Sandy	During Sandy	After Sandy	Overall
# emails	198	18	55	271

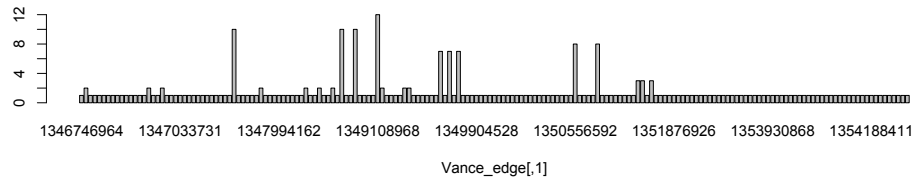Table 4: Summary of Vance county email data based on time period

Figure 5: Frequency of Vance county emails from 2012-09-04 to 2012-11-30