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

Bomin Kim

May 28, 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:

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

where λ_0 is the common baseline hazards for the overall interaction, $\boldsymbol{\beta}^{(c)}$ is an unknown vector of coefficients in $\boldsymbol{R}^p$, $\boldsymbol{x}(i, j, t)$ 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 $\boldsymbol{x}(i, j, t)$ 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 $\lambda_{ij}^{(c)}(t) = \lambda_0 \cdot \exp(\boldsymbol{\beta}^{(c)T}\boldsymbol{x}(i,j,t)) \cdot 1\{j \in \mathcal{J}_{(i,t)}^{(c)}\}$
            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 $\boldsymbol{x}(i,j,t)$ 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 $\boldsymbol{x}(i,j,t)$.

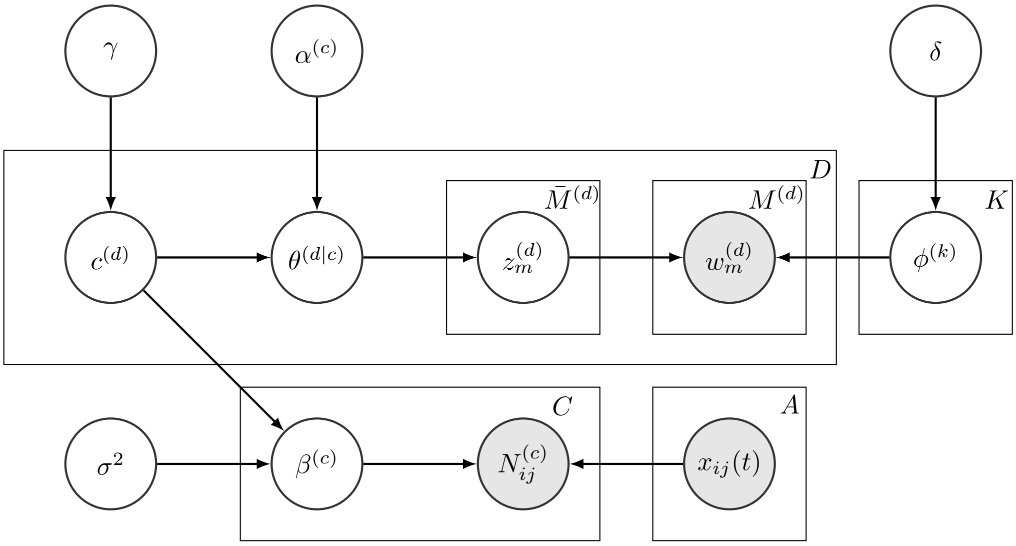

Figure 1: Plate notation of IPME

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. send$(i, j, t) = \sum\limits_{d:t_d<t} I\{i \to j\} \cdot g(t - t_d)$
2. receive$(i, j, t) = \sum\limits_{d:t_d<t} I\{j \to i\} \cdot g(t - t_d)$
3. 2-send$(i, j, t) = \sum\limits_{h \neq i,j} (\sum\limits_{d:t_d<t} I\{i \to h\} \cdot g(t - t_d))(\sum\limits_{d:t_d<t} I\{h \to j\} \cdot g(t - t_d))$
4. 2-receive$(i, j, t) = \sum\limits_{h \neq i,j} (\sum\limits_{d:t_d<t} I\{h \to i\} \cdot g(t - t_d))(\sum\limits_{d:t_d<t} I\{j \to h\} \cdot g(t - t_d))$
5. sibling$(i, j, t) = \sum\limits_{h \neq i,j} (\sum\limits_{d:t_d<t} I\{h \to i\} \cdot g(t - t_d))(\sum\limits_{d:t_d<t} I\{h \to j\} \cdot g(t - t_d))$
6. cosibling$(i, j, t) = \sum\limits_{h \neq i,j} (\sum\limits_{d:t_d<t} I\{i \to h\} \cdot g(t - t_d))(\sum\limits_{d:t_d<t} I\{j \to h\} \cdot g(t - t_d))$

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

First of all, we figure out how to obtain the 6 covariates in $x^{(c)}(a,r,t)$ defined in the previous section. Notice that one special thing about this approach is that for each term we put weights in front of typical network statistics, that is,

$$\frac{\bar{N}^{(c|d)}}{\bar{N}^{(d)}}.$$

Similar to the generative process of CPME, define

$$\bar{N}^{(d)} = \max(1, N^{(d)}) \quad \text{and} \quad \bar{N}^{(c|d)} = \sum_{n=1}^{\bar{N}^{(d)}} \delta(l_{z_n^{(d)}} = c).$$

In order to obtain $\bar{N}^{(c|d)}$, we need the probability that a token is assgined to a particular topic related to specific interaction pattern conditional on the topic assignments of all other tokens, the word-types of all other tokens, and hyper-parameters. Since we use uniform distribution for topic-interaction pattern assignments (i.e. $l_t \sim \text{Unif}(1, C)$),

$$P(l_{z_n^{(d)}} = c|\mathcal{W}, \mathcal{Z}_{\backslash d,n}, \beta, \boldsymbol{n}, \alpha, \boldsymbol{m}) = \sum_{t:l_t=c} P(z_n^{(d)} = t|\mathcal{W}, \mathcal{Z}_{\backslash d,n}, \beta, \boldsymbol{n}, \alpha, \boldsymbol{m}).$$

Now following the existing derivation from the work of Matthew Denny, by applying Bayes rule, we get

$$\begin{aligned}\sum_{t:l_t=c} P(z_n^{(d)} = t|\mathcal{W}, \mathcal{Z}_{\backslash d,n}, \beta, \boldsymbol{n}, \alpha, \boldsymbol{m}) &= \sum_{t:l_t=c} P(z_n^{(d)} = t, w_n^{(d)}|\mathcal{W}_{\backslash d,n}, \mathcal{Z}_{\backslash d,n}, \beta, \boldsymbol{n}, \alpha, \boldsymbol{m}) \\ &= \sum_{t:l_t=c} \frac{N_{\backslash d,n}^{(t|d)} + \alpha m_t}{N^d - 1 + \alpha} \cdot \frac{N_{\backslash d,n}^{(w_n^{(d)}|t)} + \beta n_v}{N_{\backslash d,n}^{(t)} + \beta}\end{aligned}$$

After we fully construct the matrix of covariates $x^{(c)}(a,r,t)$, we move to estimation steps. Assuming each interaction has a single receiver (no multicast), we model the stochastic intensity of the multivariate counting process N as

$$\lambda_t(i,j) = \bar{\lambda}_t(i) \cdot \exp\{\beta^T x_t(i,j)\} \cdot 1\{j \in \mathcal{J}_t(i)\}, \tag{3}$$

where β is an unknown vector of coefficients in $\boldsymbol{R}^p$. After treating the baseline rate $\bar{\lambda}_t(i)$ as a nuisance parameter, estimation for β proceeds by maximizing the so-called partial likelihood of Cox (1992):

$$PL_t(\beta) = \prod_{t_m \leq t} \frac{\exp(\beta^T x_{t_m}(i_m, j_m))}{\sum_{j \in \mathcal{J}_{t_m}(i_m)} \exp(\beta^T x_{t_m}(i_m, j))}, \tag{4}$$

where t_m, i_m, and j_m are the time, sender, and receiver of the mth event. Further, we can estimate the coefficient vector β using the log-partial likelihood at time t:

$$\log PL_t(\beta) = \sum_{t_m \leq t} \Big\{\beta^T x_{t_m}(i_m, j_m) - \log\big[\sum_{j \in \mathcal{J}_{t_m}(i_m)} \exp\{\beta^T x_{t_m}(i_m, j)\}\big]\Big\}. \tag{5}$$

The function $\log PL_t(\cdot)$ is concave, so it can be maximized using the first two derivatives, the gradient $U_t(\beta) = \bigtriangledown[\log PL_t(\beta)]$ and negative Hessian $I_t(\beta) = -\bigtriangledown^2 [\log PL_t(\beta)]$ via the Newton-Raphson algorithm. See Perry and Wolfe (2013) for the details.

The model can be extended for interactions that possibly involve a single sender and multiple receivers such as emails, which we call a multicast interaction. Let the tuple (t, i, J) indicate that at time t, sender i interacted with receiver set J. In addition, let $|J|$ denote the cardinality of the receiver set J. Then, we can write the modified version of (3) and (5) as below:

$$\lambda_t(i, J) = \bar{\lambda}_t(i; |J|) \cdot \exp\{\sum_{j \in J} \beta_0^T x_t(i, j)\} \cdot \prod_{j \in J} 1\{j \in \mathcal{J}_t(i)\}, \tag{6}$$

$$\log PL_t(\beta) = \sum_{t_m \leq t} \Big\{ \sum_{j \in J_m} \beta^T x_{t_m}(i_m, j) - \log\big[\sum_{\substack{j \subseteq \mathcal{J}_{t_m}(i_m) \\ |J| = |J_m|}} \exp\{\sum_{j \in J} \beta^T x_{t_m}(i_m, j)\}\big] \Big\}. \tag{7}$$

Fitting the model above is quite complicated due to the double sums. Thus, instead of using the multicast model, we can use the alternative, the approximate partial loglikelihood, which is obtained by treating the multicast interaction as multiple pairwise interactions. For instance, an email with the tuple $(t, i, J = (j_1, j_2))$ is treated as two separate emails (t, i, j_1) and (t, i, j_2). Under this transformation, the obtained approximate partial loglikelihood is:

$$\log \widetilde{PL}_t(\beta) = \sum_{t_m \leq t} \Big\{ \sum_{j \in J_m} \beta^T x_{t_m}(i_m, j) - |J_m| \log\big[\sum_{j \subseteq \mathcal{J}_{t_m}(i_m)} \exp\{\beta^T x_{t_m}(i_m, j)\}\big] \Big\} \tag{8}$$

and $\log PL_t(\beta) \approx \log \widetilde{PL}_t(\beta)$. To fit the model in a simpler way, we use (8) when estimating the coefficitients β. For implementation and simulation processes in detail, see Perry and Wolfe (2013). This framework can be viewed as one version of relational event model Butts (2008), accounting for repeated directed actions and multicast inteactions. Therefore, the coefficients measuring the common effects such as homophily or transitivity can be interpreted in the same context as in Butts (2008).

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

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).

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

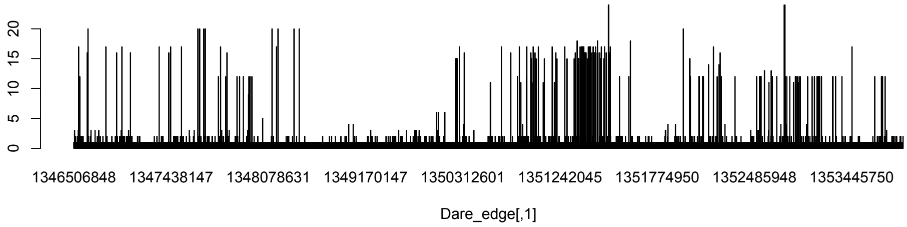

Figure 2: Frequency of Dare county emails from 2012-09-01 to 2012-11-30

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).

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)

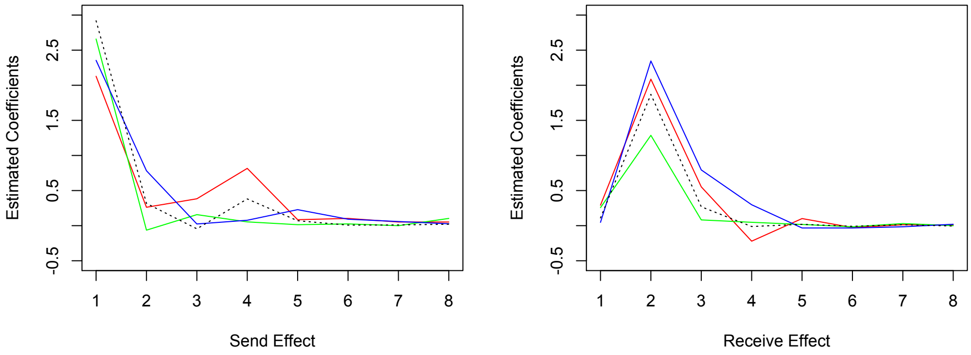

Figure 3: Comparison of Send (left) and Receive (right) effect based on periods in Table 1. (Red=Before, Green=During, Blue=After, and dot=Overall)

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.

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

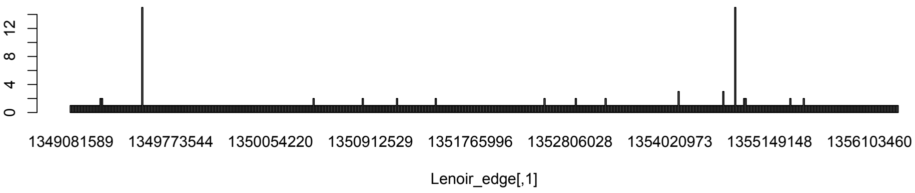

Figure 4: Frequency of Lenoir county emails from 2012-10-01 to 2012-12-31

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.

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

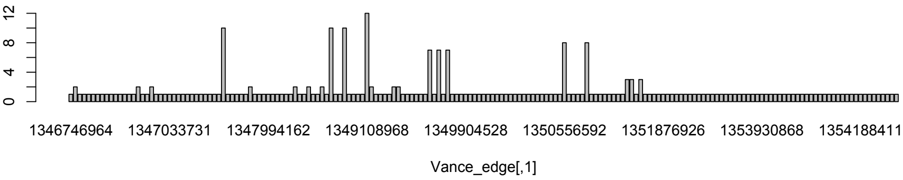

Figure 5: Frequency of Vance county emails from 2012-09-04 to 2012-11-30