A Network Model for Dynamic Textual Communications with Application to Government Email Corpora

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

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

June 1, 2017

1 Tie Generating Process

We assume the following generative process for each document d in a corpus D:

1. (Data augmentation) For each sender $i \in \{1, ..., A\}$, create a list of receivers denoted by an indicator vector J_i according to the probability mass function of Gibbs measure:

$$J_i^{(d)} = \frac{1}{Z(J_i^{(d)})} \exp\Big(\sum_{j \in \mathcal{A}_{\setminus i}} J_{ij}^{(d)} \lambda_{ij}^{(d)} - \text{dlognorm}(\log(\sum_{j \in \mathcal{A}_{\setminus i}} J_{ij}^{(d)}); 0, \delta)\Big), \qquad J_i^{(d)} \in \mathcal{J} \tag{1}$$

where $Z(J_i^{(d)}) = \sum\limits_{\forall J_i^{(d)} \in \mathcal{J}} \exp\Big(\sum_{j \in \mathcal{A}_{\setminus i}} J_{ij}^{(d)} \lambda_{ij}^{(d)} - \text{dlognorm}(\log(\sum\limits_{j \in \mathcal{A}_{\setminus i}} J_{ij}^{(d)}); 0, \delta)\Big)$ is the partition function assuring that the probabilities sum to unity, and $\mathcal{J}$ be the sample space in which $J_i^{(d)}$ may exist (in this case, the sample space consists of all binary $(A-1)$ length vectors). δ is a positive real-valued parameter that is used as a density penalty.

2. For every sender $i \in \mathcal{A}$, generate the time increments

$$\Delta T_{iJ_i} \sim \text{Exp}(\lambda_{iJ_i}^{(d)}). \tag{2}$$

3. Set timestamp, sender, and receivers simultaneously (NOTE: $t^{(0)} = 0$):

$$\begin{aligned} t^{(d)} &= t^{(d-1)} + \min(\Delta T_{iJ_i}), \\ i^{(d)} &= i_{\min(\Delta T_{iJ_i})}, \\ J^{(d)} &= J_{i^{(d)}}. \end{aligned} \tag{3}$$

2 Inference

$$\begin{aligned}
&P(\mathcal{J}_\text{a}^{(d)}, \mathcal{T}_\text{a}^{(d)}, i_\text{o}^{(d)}, J_\text{o}^{(d)}, t_\text{o}^{(d)} | \mathcal{I}_\text{o}^{(<d)}, \mathcal{J}_\text{o}^{(<d)}, \mathcal{T}_\text{o}^{(<d)}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \delta) \\
&= P\Big(\text{latent receivers generation}\Big) \times P\Big(\text{latent time generation}\Big) \times P\Big(\text{choose the observed}\Big) \\
&= \prod_{i \in \mathcal{A}} \Big(J_i^{(d)} \sim \text{Gibbs measure}(\{\lambda_{ij}^{(d)}\}_{j=1}^{A}, \delta)\Big) \times \prod_{i \in \mathcal{A}} \Big(\Delta T_{iJ_i}^{(d)} \sim \text{Exp}(\lambda_{iJ_i}^{(d)})\Big) \times \prod_{i \in \mathcal{A}_{\setminus i_o^{(d)}}} P\Big(\Delta T_{iJ_i}^{(d)} > \Delta T_{i_o^{(d)} J_o^{(d)}}\Big) \\
&= \Big(\prod_{i \in \mathcal{A}} e^{\sum_{j \in \mathcal{A}_{\setminus i}} J_{ij}^{(d)} \lambda_{ij}^{(d)} - \text{dlognorm}(\log(\sum\limits_{j \in \mathcal{A}_{\setminus i}} J_{ij}^{(d)}); 0, \delta)}\Big) \times \Big(\prod_{i \in \mathcal{A}} \lambda_{iJ_i}^{(d)} e^{-\Delta T_{iJ_i}^{(d)} \lambda_{iJ_i}^{(d)}}\Big) \times \Big(\prod_{i \in \mathcal{A}_{\setminus i_o^{(d)}}} e^{-\Delta T_{i_o^{(d)} J_o^{(d)}}^{(d)} \lambda_{iJ_i^{(d)}}^{(d)}}\Big),
\end{aligned} \tag{4}$$

which we can simplify further by integreting out the latent time $\mathcal{T}_{\mathrm{a}}^{(d)} = \{\Delta T_{iJ_i}^{(d)}\}_{i\in\mathcal{A}_{\backslash i_o^{(d)}}}$. Then we can simplify Equation (4) as below:

$$
\begin{aligned}
&P(\mathcal{J}_{\mathrm{a}}^{(d)}, i_{\mathrm{o}}^{(d)}, J_{\mathrm{o}}^{(d)}, t_{\mathrm{o}}^{(d)} | \mathcal{I}_{\mathrm{o}}^{(<d)}, \mathcal{J}_{\mathrm{o}}^{(<d)}, \mathcal{T}_{\mathrm{o}}^{(<d)}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \delta) \\
&= \Big(\prod_{i\in\mathcal{A}} e^{\sum_{j\in\mathcal{A}_{\backslash i}} J_{ij}^{(d)}\lambda_{ij}^{(d)} - \mathrm{dlognorm}(\log(\sum_{j\in\mathcal{A}_{\backslash i}} J_{ij}^{(d)});0,\delta)}\Big) \times \Big(\lambda_{i_o^{(d)} J_o^{(d)}}^{(d)}\Big) \times \Big(e^{-\Delta T_{i_o^{(d)} J_o^{(d)}}^{(d)} \sum_{i\in\mathcal{A}} \lambda_{iJ_i^{(d)}}^{(d)}}\Big),
\end{aligned} \tag{5}
$$

where this joint distribution can be interpreted as 'probability of latent and observed edges from Gibbs measure × probability of the observed time comes from Exponential distribution × probability of all latent time greater than the observed time, given that the latent time also come from Exponential distribution.'

2.1 Inference on the augmented data $\mathcal{J}_{\mathbf{a}}$

Given the observed sender of the document $i_o^{(d)}$, we sample the latent receivers for each sender $i \in \mathcal{A}_{\backslash i_o^{(d)}}$. Here we illustrate how each sender-receiver pair in the document d is updated.

Define $\mathcal{J}_i^{(d)}$ be the $(A-1)$ length random vector of indicators (0/1) with its realization being $J_i^{(d)}$, representing the latent receivers corresponding to the sender i in the document d. For each sender i, we are going to resample $J_{ij}^{(d)}$, which is the j^{th} element of the receiver vector $J_i^{(d)}$, one at a time with random order.

$$
\begin{aligned}
&P(\mathcal{J}_{ij}^{(d)} = J_{ij}^{(d)} | \mathcal{J}_{i\backslash j}^{(d)}, i_{\mathrm{o}}^{(d)}, J_{\mathrm{o}}^{(d)}, t_{\mathrm{o}}^{(d)}, \mathcal{I}_{\mathrm{o}}^{(<d)}, \mathcal{J}_{\mathrm{o}}^{(<d)}, \mathcal{T}_{\mathrm{o}}^{(<d)}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \boldsymbol{\delta}) \\
&\propto P(\mathcal{J}_{i}^{(d)} = J_{i}^{(d)}, \mathcal{J}_{i\backslash j}^{(d)}, i_{\mathrm{o}}^{(d)}, J_{\mathrm{o}}^{(d)}, t_{\mathrm{o}}^{(d)} | \mathcal{I}_{\mathrm{o}}^{(<d)}, \mathcal{J}_{\mathrm{o}}^{(<d)}, \mathcal{T}_{\mathrm{o}}^{(<d)}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \boldsymbol{\delta}) \\
&\propto \exp\Big(\sum_{j\in\mathcal{A}_{\backslash i}} J_{ij}^{(d)}\lambda_{ij}^{(d)} - \mathrm{dlognorm}(\log(\sum_{j\in\mathcal{A}_{\backslash i}} J_{ij}^{(d)}); 0, \delta)\Big) \times \Big(\lambda_{i_o^{(d)} J_o^{(d)}}^{(d)}\Big) \times \Big(e^{-\Delta T_{i_o^{(d)} J_o^{(d)}}^{(d)} \lambda_{iJ_i^{(d)}}^{(d)}}\Big) \\
&\propto \exp\Big(\sum_{j\in\mathcal{A}_{\backslash i}} J_{ij}^{(d)}\lambda_{ij}^{(d)} - \mathrm{dlognorm}(\log(\sum_{j\in\mathcal{A}_{\backslash i}} J_{ij}^{(d)}); 0, \delta)\Big) \times \Big(e^{-\Delta T_{i_o^{(d)} J_o^{(d)}}^{(d)} \lambda_{iJ_i^{(d)}}^{(d)}}\Big),
\end{aligned} \tag{6}
$$

where we replace typical use of $(-d)$ to $(< d)$ on the right hand side of the conditional probability, due to the fact that $d^{(th)}$ document only depends on the past documents, not on the future ones.

To be more specific, since $J_{ij}^{(d)}$ could be either 1 or 0, we divide into two cases as below:

$$
\begin{aligned}
&P(\mathcal{J}_{ij}^{(d)} = 1 | \mathcal{J}_{\backslash ij}^{(d)}, i_{\mathrm{o}}^{(d)}, J_{\mathrm{o}}^{(d)}, t_{\mathrm{o}}^{(d)}, \mathcal{I}_{\mathrm{o}}^{(<d)}, \mathcal{J}_{\mathrm{o}}^{(<d)}, \mathcal{T}_{\mathrm{o}}^{(<d)}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \delta) \\
&\propto \exp\Big(\sum_{j\in\mathcal{A}_{\backslash i}} J_{i[+j]}^{(d)}\lambda_{ij}^{(d)} - \mathrm{dlognorm}(\log(\sum_{j\in\mathcal{A}_{\backslash i}} J_{i[+j]}^{(d)}); 0, \delta) - \Delta T_{i_o^{(d)} J_o^{(d)}}^{(d)} \lambda_{iJ_{i[+j]}^{(d)}}^{(d)}\Big),
\end{aligned} \tag{7}
$$

where $J_{i[+j]}^{(d)}$ meaning that the j^{th} element of $J_i^{(d)}$ is fixed as 1. On the other hand,

$$
\begin{aligned}
&P(\mathcal{J}_{ij}^{(d)} = 0 | \mathcal{J}_{\backslash ij}^{(d)}, i_{\mathrm{o}}^{(d)}, J_{\mathrm{o}}^{(d)}, t_{\mathrm{o}}^{(d)}, \mathcal{I}_{\mathrm{o}}^{(<d)}, \mathcal{J}_{\mathrm{o}}^{(<d)}, \mathcal{T}_{\mathrm{o}}^{(<d)}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \delta) \\
&\propto \exp\Big(\sum_{j\in\mathcal{A}_{\backslash i}} J_{i[-j]}^{(d)}\lambda_{ij}^{(d)} - \mathrm{dlognorm}(\log(\sum_{j\in\mathcal{A}_{\backslash i}} J_{i[-j]}^{(d)}); 0, \delta) - \Delta T_{i_o^{(d)} J_o^{(d)}}^{(d)} \lambda_{iJ_{i[-j]}^{(d)}}^{(d)}\Big),
\end{aligned} \tag{8}
$$

where $J_{i[-j]}^{(d)}$ meaning similarly that the j^{th} element of $J_i^{(d)}$ is fixed as 0.

Now we can use multinomial sampling using the two probabilities, Equation (7) and Equation (8). Again, we would calculate the probabilities in the log space to prevent from numerical underflow.

3 Non-empty Gibbs measure

The probability that vertex i selects the binary receiver vector of length $n-1$, $J_i^{(d)}$ is given by

$$\mathrm{P}(J_i^{(d)}) = \frac{1}{Z(\delta, \lambda_i^{(d)})} \exp\left[\ln\left(\mathrm{I}\left(\left(\sum_{j\neq i} J_{ij} \right) > 0 \right) \right) + \sum_{j\neq i} (\delta + \lambda_{ij}^{(d)}) J_{ij}^{(d)} \right],$$

where δ is real-valued intercept that controls the expected value of $J_{ij}^{(d)}$, and $\lambda_{ij}^{(d)}$ is a positive dyad-specific function of the network histories of i and j, and any additional attributes included in the model (note, it may fit the data better to include lambda as $\ln\left(\lambda_{ij}^{(d)}\right)$, and $\lambda_i^{(d)}$ is the vector of dyadic weights in which i is the sender.

To use this distribution efficiently, we need to derive a closed-form expression for $\frac{1}{Z(\delta,\lambda_i^{(d)})}$ that does not require brute-force summation over the support of $J_i^{(d)}$. We begin by recognizing that if $J_i^{(d)}$ were drawn via independent Bernoulli distributions in which $\mathrm{P}(J_{ij}^{(d)}{=}1)$ was given by $\mathrm{logit}\left(\delta + \lambda_{ij}^{(d)}\right)$, then

$$\mathrm{P}(J_i^{(d)}) \propto \exp\left[\sum_{j\neq i} (\delta + \lambda_{ij}^{(d)}) J_{ij}^{(d)} \right].$$

This is straightforward to verify by looking at

$$\mathrm{P}(J_{ij}^{(d)} = 1 | J_{i,-j}) = \frac{\exp(\delta + \lambda_{ij}) \exp\left[\sum_{h\neq i,j} (\delta + \lambda_{ih}^{(d)}) J_{ih}^{(d)}\right]}{\exp(\delta + \lambda_{ij}) \exp\left[\sum_{h\neq i,j} (\delta + \lambda_{ih}^{(d)}) J_{ih}^{(d)}\right] + \exp(0) \exp\left[\sum_{h\neq i,j} (\delta + \lambda_{ih}^{(d)}) J_{ih}^{(d)}\right]},$$

$$\mathrm{P}(J_{ij}^{(d)} = 1 | J_{i,-j}) = \frac{\exp(\delta + \lambda_{ij})}{\exp(\delta + \lambda_{ij}) + 1}.$$

We denote the logistic-Bernoulli normalizing constant as

$$Z^l(\delta, \lambda_i^{(d)}) = \sum_{J_i \in [0,1]^{(n-1)}} \exp\left[\sum_{j\neq i} (\delta + \lambda_{ij}^{(d)}) J_{ij}^{(d)} \right].$$

Now, since

$$\exp\left[\ln\left(\mathrm{I}\left(\left(\sum_{j\neq i} J_{ij} \right) > 0 \right) \right) + \sum_{j\neq i} (\delta + \lambda_{ij}^{(d)}) J_{ij}^{(d)} \right] = \exp\left[\sum_{j\neq i} (\delta + \lambda_{ij}^{(d)}) J_{ij}^{(d)} \right],$$

except when $\sum_{j\neq i} J_{ij} = 0$, in which case

$$\exp\left[\ln\left(\mathrm{I}\left(\left(\sum_{j\neq i} J_{ij} \right) > 0 \right) \right) + \sum_{j\neq i} (\delta + \lambda_{ij}^{(d)}) J_{ij}^{(d)} \right] = 0.$$

As such, we note that

$$Z(\delta, \lambda_i^{(d)}) = Z^l(\delta, \lambda_i^{(d)}) - \exp\left[\sum_{j\neq i, J_{ij}=0} (\delta + \lambda_{ij}^{(d)}) J_{ij}^{(d)} \right],$$

$$Z(\delta, \lambda_i^{(d)}) = Z^l(\delta, \lambda_i^{(d)}) - 1.$$

We can therefore derive a closed form expression for $Z(\delta, \lambda_i^{(d)})$ via a closed form expression for $Z^l(\delta, \lambda_i^{(d)})$. This can be done by looking at the probability of the zero vector under the logistic-Bernoulli model.

$$\frac{\exp\left[\sum_{j\neq i, J_{ij}=0} (\delta + \lambda_{ij}^{(d)}) J_{ij}^{(d)}\right]}{Z^l(\delta, \lambda_i^{(d)})} = \prod_{i\neq j} \frac{\exp(-(\delta + \lambda_{ij}))}{\exp(-(\delta + \lambda_{ij})) + 1},$$

$$\frac{1}{Z^l(\delta, \lambda_i^{(d)})} = \prod_{i \neq j} \frac{\exp\left(-(\delta + \lambda_{ij})\right)}{\exp\left(-(\delta + \lambda_{ij})\right) + 1},$$

$$Z^l(\delta, \lambda_i^{(d)}) = \frac{1}{\prod_{i \neq j} \frac{\exp\left(-(\delta + \lambda_{ij})\right)}{\exp\left(-(\delta + \lambda_{ij})\right) + 1}}.$$

The closed form expression for the normalizing constant under the non-empty Gibbs measure is

$$Z(\delta, \lambda_i^{(d)}) = \frac{1}{\prod_{i \neq j} \frac{\exp\left(-(\delta + \lambda_{ij})\right)}{\exp\left(-(\delta + \lambda_{ij})\right) + 1}} - 1.$$