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

May 30, 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 J_i by applying the Bernoulli probabilities to every $j \in \mathcal{A}_{\backslash i}$ with random order:

$$J_{ij}|J_{i\backslash j} \sim \text{Ber}\Big(\frac{\frac{\delta}{|J_{i\backslash j}^{(d)}|}\lambda_{ij}^{(d)}}{\frac{\delta}{|J_{i\backslash j}^{(d)}|}\lambda_{ij}^{(d)}+1}\Big), \tag{1}$$

   where $|J_{i\backslash j}^{(d)}|$ is the number of 1's currently in the receiver set J_i, excluding the j^{th} element. We divide δ by this quantity to ensure at least one receiver is chosen for each sender. For example, if $|J_{i\backslash j}^{(d)}| = 0$, J_{ij} becomes 1 with probability 1. We no longer have independent asumption across the edges, instead, every j^{th} element is sampled conditioned on the rest of components in J_{ij} vector, $J_{i\backslash j}$.

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}_{\mathrm{a}}^{(d)}, \mathcal{T}_{\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) \\
&= 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}} \prod_{j\in\mathcal{A}_{\backslash i}} \Big(J_{ij} \sim \mathrm{Ber}\Big(\frac{\frac{\delta}{|J_{i\backslash j}^{(d)}|}\lambda_{ij}^{(d)}}{\frac{\delta}{|J_{i\backslash j}^{(d)}|}\lambda_{ij}^{(d)} + 1}\Big)\Big) \times \prod_{i\in\mathcal{A}} \Big(\Delta T_{iJ_i}^{(d)} \sim \mathrm{Exp}(\lambda_{iJ_i}^{(d)})\Big) \times \prod_{i\in\mathcal{A}_{\backslash 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}} \prod_{j\in\mathcal{A}_{\backslash i}} \big(\frac{\frac{\delta}{|J_{i\backslash j}^{(d)}|}\lambda_{ij}^{(d)}}{\frac{\delta}{|J_{i\backslash j}^{(d)}|}\lambda_{ij}^{(d)} + 1}\big)^{I(j\in J_i^{(d)})} \big(\frac{1}{\frac{\delta}{|J_{i\backslash j}^{(d)}|}\lambda_{ij}^{(d)} + 1}\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}_{\backslash i_o^{(d)}}} e^{-\Delta T_{i_o^{(d)} J_o^{(d)}}^{(d)} \lambda_{iJ_i^{(d)}}^{(d)}}\Big) \\
&= \Big(\prod_{i\in\mathcal{A}} \prod_{j\in\mathcal{A}_{\backslash i}} \frac{(\frac{\delta}{|J_{i\backslash j}^{(d)}|}\lambda_{ij}^{(d)})^{I(j\in J_i)}}{\frac{\delta}{|J_{i\backslash j}^{(d)}|}\lambda_{ij}^{(d)} + 1}\Big) \times \Big(\lambda_{i_o^{(d)} J_o^{(d)}}^{(d)} e^{-\Delta T_{i_o^{(d)} J_o^{(d)}}^{(d)} \lambda_{i_o^{(d)} J_o^{(d)}}^{(d)}}\Big) \times \Big(\prod_{i\in\mathcal{A}_{\backslash i_o^{(d)}}} \lambda_{iJ_i}^{(d)} e^{-(\Delta T_{iJ_i}^{(d)} + \Delta T_{i_o^{(d)} J_o^{(d)}}^{(d)}) \lambda_{iJ_i}^{(d)}}\Big),
\end{aligned} \tag{4}
$$

We can simplify this further by integreting out the latent time $\mathcal{T}_{\mathrm{a}}^{(d)} = \{\Delta T_{iJ_i}^{(d)}\}_{i\in\mathcal{A}_{\backslash i_o^{(d)}}}$ in the last term as before, then we can simplify Equation (4) as below:

$$
\begin{aligned}
&P(\mathcal{J}_{\mathrm{a}}^{(d)}, i_{\mathrm{o}}^{(d)}, J_{\mathrm{o}}^{(d)}, t_{\mathrm{o}}^{(d)} | \mathcal{I}_{\mathrm{o}}^{(<d)}, \mathcal{J}_{\mathrm{o}}^{(<d)}, \mathcal{T}_{\mathrm{o}}^{(<d)}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \delta) \\
&= \Big(\prod_{i\in\mathcal{A}} \prod_{j\in\mathcal{A}_{\backslash i}} \frac{(\frac{\delta}{|J_{i\backslash j}^{(d)}|}\lambda_{ij}^{(d)})^{I(j\in J_i)}}{\frac{\delta}{|J_{i\backslash j}^{(d)}|}\lambda_{ij}^{(d)} + 1}\Big) \times \Big(\lambda_{i_o^{(d)} J_o^{(d)}}^{(d)}\Big) \times \Big(e^{-\Delta T_{i_o^{(d)} J_o^{(d)}}^{(d)} \sum_{i\in\mathcal{A}_{\backslash i_o^{(d)}}} \lambda_{iJ_i^{(d)}}^{(d)}}\Big),
\end{aligned} \tag{5}
$$

where this joint distribution can be interpreted as 'probability of latent and observed edges from Bernoulli distribution × 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 $\mathcal{J}_{\mathrm{a}}$

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 $J_i^{(d)}$ be the $(A-1)$ length vector of indicators (0/1) representing the latent receivers corresponding to the sender i in the document d. For each sender i, we are going to resample the receiver vector $J_i^{(d)}$, one at a time. For a latent sender $i \in \mathcal{A}_{\backslash i_o^{(d)}}$, we derive the conditional probability:

$$
\begin{aligned}
&P(\mathcal{J}_{ij}^{(d)} = J_{ij}^{(d)} | \mathcal{J}_{i\backslash j}^{(d)}, i_{\mathrm{o}}^{(d)}, J_{\mathrm{o}}^{(d)}, t_{\mathrm{o}}^{(d)}, \mathcal{I}_{\mathrm{o}}^{(<d)}, \mathcal{J}_{\mathrm{o}}^{(<d)}, \mathcal{T}_{\mathrm{o}}^{(<d)}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \boldsymbol{\delta}) \\
&\propto P(\mathcal{J}_{i}^{(d)} = J_{i}^{(d)}, \mathcal{J}_{i\backslash j}^{(d)}, i_{\mathrm{o}}^{(d)}, J_{\mathrm{o}}^{(d)}, t_{\mathrm{o}}^{(d)} | \mathcal{I}_{\mathrm{o}}^{(<d)}, \mathcal{J}_{\mathrm{o}}^{(<d)}, \mathcal{T}_{\mathrm{o}}^{(<d)}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \boldsymbol{\delta}) \\
&\propto \Big(\frac{(\frac{\delta}{|J_{i\backslash j}^{(d)}|}\lambda_{ij}^{(d)})^{I(j\in J_i)}}{\frac{\delta}{|J_{i\backslash j}^{(d)}|}\lambda_{ij}^{(d)} + 1}\Big) \times \Big(e^{-\Delta T_{i_o^{(d)} J_o^{(d)}}^{(d)} \sum_{i\in\mathcal{A}_{\backslash i_o^{(d)}}} \lambda_{iJ_i^{(d)}}^{(d)}}\Big),
\end{aligned} \tag{6}
$$

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.

No idea how to choose the proposal distribution for the indicator vector $J_i^{(d)}$. Possibly sample each element $J_{ij}^{(d)}$ as we did before, using M-H sampling with the choice of univariate Wallenius' distribution as the proposal density which relies on approximation (Fog, 2008).

2.2 Inference on $\mathcal{Z}$

same as before but edge probability part changed to $\left(\prod_{i\in\mathcal{A}} \text{dMWNCHypergeo}\left(J_i^{(d)}; \mathbf{1}_{A-1}, R_i^{(d)}, \{\lambda_{ij}^{(d)}\}_{j\in\mathcal{A}_{\setminus i}}\right)\right)$

2.3 Inference on $\mathcal{C}$

same as before but edge probability part changed to $\left(\prod_{i\in\mathcal{A}} \text{dMWNCHypergeo}\left(J_i^{(d)}; \mathbf{1}_{A-1}, R_i^{(d)}, \{\lambda_{ij}^{(d)}\}_{j\in\mathcal{A}_{\setminus i}}\right)\right)$

2.4 Inference on $\mathcal{B}$

same as before but edge probability part changed to $\left(\prod_{i\in\mathcal{A}} \text{dMWNCHypergeo}\left(J_i^{(d)}; \mathbf{1}_{A-1}, R_i^{(d)}, \{\lambda_{ij}^{(d)}\}_{j\in\mathcal{A}_{\setminus i}}\right)\right)$