

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

Bomin Kim¹, Aaron Schein³, Bruce Desmarais¹, and Hanna Wallach^{2,3}

¹Pennsylvania State University

²Microsoft Research NYC

³University of Massachusetts Amherst

May 29, 2017

1 Tie Generating Process

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

1. Choose the number of recipients

$$R_i^{(d)} \sim \text{zero-truncated Binomial}(A - 1, \delta_i), \quad (1)$$

where $A - 1$ comes from excluding the sender himself as a possible receiver (self-loop) and δ_i is the sender-specific probability of success. For example, we can use R function

```
library(actuar)
R_i = rztbinom(n = 1, size = 3, prob = 0.1)
```

2. (Data augmentation) For each sender $i \in \{1, \dots, A\}$, create a list of receivers J_i by applying multivariate Wallenius' noncentral hypergeometric distribution (MWNCHypergeo) to every $j \in \mathcal{A}_{\setminus i}$

$$J_i^{(d)} \sim \text{MWNCHypergeo}\left(\mathbf{m} = \mathbf{1}_{A-1}, N = R_i^{(d)}, \boldsymbol{\omega} = \{\lambda_{ij}^{(d)}\}_{j \in \mathcal{A}_{\setminus i}}\right), \quad (2)$$

where \mathbf{m} is the vector of availability (we have maximum 1 available for each actor except the sender), N is the total number of receivers to be sampled, and $\boldsymbol{\omega}$ is the weight for each actor to be sampled. Same as before, $\lambda_{ij}^{(d)}$ is evaluated at time $t_+^{(d-1)}$. Note that $+$ denotes including the timepoint itself, meaning that λ_{ij} is obtained using the history of interactions until and including the timestamp $t^{(d-1)}$. For example, we can use R function

```
library(BiasedUrn)
J_i = rMFNCHypergeo(nran = 1, m = c(1,1,1,1), n = 2, odds = c(0.1, 0.2, 0.3, 0.4))
```

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

$$\Delta T_{iJ_i}^{(d)} \sim \text{Exp}(\lambda_{iJ_i}^{(d)}), \quad (3)$$

where $\lambda_{iJ_i}^{(d)}(t) = \sum_{c=1}^C p_c^{(d)} \cdot \exp\left\{\lambda_0^{(c)} + \frac{1}{|J_i|} \sum_{j \in J_i} \mathbf{b}^{(c)T} \mathbf{x}_t^{(c)}(i, j)\right\} \cdot \prod_{j \in J_i} 1\{j \in \mathcal{A}_{\setminus i}\}.$

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

$$\begin{aligned} t^{(d)} &= t^{(d-1)} + \min(\Delta T_{iJ_i}), \\ i^{(d)} &= i_{\min(\Delta T_{iJ_i})}, \\ J^{(d)} &= J_{i^{(d)}}. \end{aligned} \quad (4)$$

2 Inference

$$\begin{aligned}
& P(\mathcal{R}^{(d)}, \mathcal{J}_a^{(d)}, \mathcal{T}_a^{(d)}, i_o^{(d)}, J_o^{(d)}, t_o^{(d)} | \mathcal{I}_o^{(<d)}, \mathcal{J}_o^{(<d)}, \mathcal{T}_o^{(<d)}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \delta) \\
&= P(\text{Number of recipients}) \times P(\text{Edge generation}) \times P(\text{Time generation}) \times P(\text{choose the observed}) \\
&= \prod_{i \in \mathcal{A}} \left(R_i^{(d)} \sim \text{ztbinom}(A-1, \delta_i) \right) \times \prod_{i \in \mathcal{A}} \left(J_i^{(d)} \sim \text{MWNCHypergeo}(\mathbf{1}_{A-1}, R_i^{(d)}, \{\lambda_{ij}^{(d)}\}_{j \in \mathcal{A}_{\setminus i}}) \right) \\
&\quad \times \prod_{i \in \mathcal{A}} \left(\Delta T_{iJ_i}^{(d)} \sim \text{Exp}(\lambda_{iJ_i}^{(d)}) \right) \times \prod_{i \in \mathcal{A}_{\setminus i_o}^{(d)}} P\left(\Delta T_{iJ_i}^{(d)} > \Delta T_{i_o J_o}^{(d)}\right) \\
&= \left(\prod_{i \in \mathcal{A}} \binom{A-1}{R_i^{(d)}} \frac{\delta_i^{R_i^{(d)}} (1-\delta_i)^{A-1-R_i^{(d)}}}{1 - (1-\delta_i)^{A-1}} \right) \times \left(\prod_{i \in \mathcal{A}} \text{dMWNCHypergeo}(J_i^{(d)}; \mathbf{1}_{A-1}, R_i^{(d)}, \{\lambda_{ij}^{(d)}\}_{j \in \mathcal{A}_{\setminus i}}) \right) \\
&\quad \times \left(\prod_{i \in \mathcal{A}} \lambda_{iJ_i}^{(d)} e^{-\Delta T_{iJ_i}^{(d)} \lambda_{iJ_i}^{(d)}} \right) \times \left(\prod_{i \in \mathcal{A}_{\setminus i_o}^{(d)}} e^{-\Delta T_{i_o J_o}^{(d)} \lambda_{i_o J_o}^{(d)} \lambda_{iJ_i}^{(d)}} \right), \tag{5}
\end{aligned}$$

with the probability mass function of dMWNCHypergeo with $\mathbf{m} = \mathbf{1}_{A-1}$ given as

$$\text{dMWNCHypergeo}(\mathbf{x}; \mathbf{m} = \mathbf{1}_{A-1}, n = R_i^{(d)}, \boldsymbol{\omega} = \{\lambda_{ij}^{(d)}\}_{j \in \mathcal{A}_{\setminus i}}) = \int_0^1 \prod_{i=1}^{A-1} (1 - t^{\omega_i/d})^{x_i} dt,$$

where $d = \sum_{i=1}^{A-1} \omega_i(1-x_i)$, \mathbf{x} is the $(A-1)$ length vector indicating the receivers, $\boldsymbol{\omega} = (\omega_1, \dots, \omega_{A-1})$ is the weight or odds of each receiver to be chosen, and $n = \sum_{i=1}^{A-1} x_i$ is the total number of receivers chosen.

We can simplify this further by integrating out the latent time $\mathcal{T}_a^{(d)} = \{\Delta T_{iJ_i}^{(d)}\}_{i \in \mathcal{A}_{\setminus i_o}^{(d)}}$ in the last two terms:

$$\begin{aligned}
& \int_0^\infty \cdots \int_0^\infty \left(\prod_{i \in \mathcal{A}_{\setminus i_o}^{(d)}} \lambda_{iJ_i}^{(d)} e^{-(\Delta T_{iJ_i}^{(d)} + \Delta T_{i_o J_o}^{(d)} \lambda_{i_o J_o}^{(d)} \lambda_{iJ_i}^{(d)})} \right) d\Delta T_{1J_1}^{(d)} \cdots d\Delta T_{AJ_A}^{(d)} \\
&= \prod_{i \in \mathcal{A}_{\setminus i_o}^{(d)}} e^{-\Delta T_{i_o J_o}^{(d)} \lambda_{i_o J_o}^{(d)} \lambda_{iJ_i}^{(d)}} \left(\int_0^\infty \lambda_{iJ_i}^{(d)} e^{-\Delta T_{iJ_i}^{(d)} \lambda_{iJ_i}^{(d)}} d\Delta T_{iJ_i}^{(d)} \right) \\
&= \prod_{i \in \mathcal{A}_{\setminus i_o}^{(d)}} e^{-\Delta T_{i_o J_o}^{(d)} \lambda_{i_o J_o}^{(d)} \lambda_{iJ_i}^{(d)}} \left(\left[-e^{-\Delta T_{iJ_i}^{(d)} \lambda_{iJ_i}^{(d)}} \right]_{\Delta T_{iJ_i}^{(d)}=0}^\infty \right) \\
&= e^{-\Delta T_{i_o J_o}^{(d)} \lambda_{i_o J_o}^{(d)} \sum_{i \in \mathcal{A}_{\setminus i_o}^{(d)}} \lambda_{iJ_i}^{(d)}}, \tag{6}
\end{aligned}$$

where $\Delta T_{i_o J_o}^{(d)}$ is the observed time difference between d^{th} and $(d-1)^{\text{th}}$ document (i.e. $t^{(d)} - t^{(d-1)}$). Therefore, we can simplify Equation (5) as below:

$$\begin{aligned}
& P(\mathcal{R}^{(d)}, \mathcal{J}_a^{(d)}, i_o^{(d)}, J_o^{(d)}, t_o^{(d)} | \mathcal{I}_o^{(<d)}, \mathcal{J}_o^{(<d)}, \mathcal{T}_o^{(<d)}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \delta) \\
&= \left(\prod_{i \in \mathcal{A}} \binom{A-1}{R_i^{(d)}} \frac{\delta_i^{R_i^{(d)}} (1-\delta_i)^{A-1-R_i^{(d)}}}{1 - (1-\delta_i)^{A-1}} \right) \times \left(\prod_{i \in \mathcal{A}} \int_0^1 \prod_{j=1}^{A-1} (1 - t^{\lambda_{ij}^{(d)} / \sum_{j=1}^{A-1} \lambda_{ij}^{(d)} (1-J_{ij}^{(d)})})^{J_{ij}^{(d)}} dt \right) \\
&\quad \times \left(\lambda_{i_o J_o}^{(d)} \right) \times \left(e^{-\Delta T_{i_o J_o}^{(d)} \lambda_{i_o J_o}^{(d)} \sum_{i \in \mathcal{A}_{\setminus i_o}^{(d)}} \lambda_{iJ_i}^{(d)}} \right), \tag{7}
\end{aligned}$$

where this joint distribution can be interpreted as 'probability of choosing the number of recipients from zero-truncated Binomial distribution \times probability of choosing the latent receivers from Wallenius' noncentral hypergeometric distribution \times probability of the observed time comes from Exponential distribution \times probability of all latent time greater than the observed time, given that the latent time also come from Exponential distribution.'

2.1 Inference on δ

We can assign Beta prior on each δ_i . Does Beta-Binomial conjugacy still hold for zero-truncated Binomial case? I don't think so... (due to the additional denominator) should we use M-H instead?

2.2 Inference on the augmented data \mathcal{J}_a

Given the observed sender of the document $i_o^{(d)}$, we sample the latent receivers for each sender $i \in \mathcal{A}_{\setminus i_o^{(d)}}$. 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 $i \in \mathcal{A}_{\setminus i_o^{(d)}}$, we derive the conditional probability:

$$\begin{aligned} & P(\mathcal{J}_i^{(d)} = J_i^{(d)} | \mathcal{R}^{(d)}, i_o^{(d)}, J_o^{(d)}, t_o^{(d)}, \mathcal{I}_o^{(<d)}, \mathcal{J}_o^{(<d)}, \mathcal{T}_o^{(<d)}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \delta) \\ & \propto P(\mathcal{J}_i^{(d)} = J_i^{(d)}, i_o^{(d)}, J_o^{(d)}, t_o^{(d)} | \mathcal{R}^{(d)}, \mathcal{I}_o^{(<d)}, \mathcal{J}_o^{(<d)}, \mathcal{T}_o^{(<d)}, \mathcal{Z}, \mathcal{C}, \mathcal{B}, \delta) \\ & \propto \left(\int_0^1 \prod_{j=1}^{A-1} (1 - t^{\lambda_{ij}^{(d)} / \sum_{j=1}^{A-1} \lambda_{ij}^{(d)} (1 - J_{ij}^{(d)})})^{J_{ij}^{(d)}} dt \right) \times \left(e^{-\Delta T_{i_o^{(d)} J_o^{(d)}}^{(d)} \sum_{i \in \mathcal{A}_{\setminus i_o^{(d)}}} \lambda_{i J_i^{(d)}}^{(d)}} \right), \end{aligned} \quad (8)$$

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).

2.3 Inference on \mathcal{Z}

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

2.4 Inference on \mathcal{C}

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

2.5 Inference on \mathcal{B}

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