Supplementary materials for "Model-based clustering of time-evolving networks through temporal exponential-family random graph models"

Kevin H. Lee[a], Lingzhou Xue[b,*], David R. Hunter[b]

[a]*Department of Statistics, Western Michigan University, Kalamazoo, MI 49008, USA*
[b]*Department of Statistics, The Pennsylvania State University, University Park, PA 16802, USA*

In this supplementary note, we provide additional details of the proposed variational EM algorithm and also a simulation study with unbalanced clusters, i.e., clusters whose marginal probabilities are unequal.

1. More details of variational EM algorithm

1.1. Obtaining a lower bound of the log-likelihood (9)

Using Jensen's inequality, the log-likelihood function can be bounded as

$$\begin{aligned}\ln \Pr(\mathbf{y}_1,\ldots,\mathbf{y}_T \mid \mathbf{y}_0) &= \ln\left\{\sum_{z\in\mathcal{Z}} \frac{\Pr(\mathbf{y}_1,\ldots,\mathbf{y}_T, z \mid \mathbf{y}_0)}{A(z)} A(z)\right\} \\ &\geq \sum_{z\in\mathcal{Z}} \left\{\ln \frac{\Pr(\mathbf{y}_1,\ldots,\mathbf{y}_T, z \mid \mathbf{y}_0)}{A(z)}\right\} A(z). \\ &= \mathrm{E}_A\left\{\ln \Pr(\mathbf{y}_1,\ldots,\mathbf{y}_T, z \mid \mathbf{y}_0)\right\} - \mathrm{E}_A\left\{\ln A(z)\right\}.\end{aligned} \tag{1}$$

Some choices of $A(z)$ give rise to better lower bounds than others. The difference between the log-likelihood and the lower bound is equal to the Kullback-Leibler divergence from $A(z)$ to $\Pr(\mathbf{Z} = z \mid \mathbf{y}_T,\ldots,\mathbf{y}_0)$, since

$$\ln \Pr(\mathbf{y}_1,\ldots,\mathbf{y}_T \mid \mathbf{y}_0) - \sum_{z\in\mathcal{Z}} \left\{\ln \frac{\Pr(\mathbf{y}_1,\ldots,\mathbf{y}_T, z \mid \mathbf{y}_0)}{A(z)}\right\} A(z) = \sum_{z\in\mathcal{Z}} \left\{\ln \frac{A(z)}{\Pr(\mathbf{Z} = z \mid \mathbf{y}_T,\ldots,\mathbf{y}_0)}\right\} A(z).$$

Hence, we would obtain the best lower bound, yielding equality in (1), when $A(z) = \Pr(\mathbf{Z} = z \mid \mathbf{y}_T,\ldots,\mathbf{y}_0)$, which reduces the Kullback-Leibler divergence to zero. However, this $A(z)$ is computationally intractable since it cannot be further factored over nodes.

To achieve tractability, we constrain $A(z)$ to the mean-field variational family where the $\mathbf{Z}_i$ are mutually independent, i.e.,

$$A(z) = \Pr(\mathbf{Z} = z) = \prod_{i=1}^{n} \Pr(\mathbf{Z}_i = z_i).$$

To allow for achieving the optimal such $A(z)$ in terms of Kullback-Leibler divergence, we maximize the flexibility of our model by specifying that each $\mathbf{Z}_i$ has its own multinomial parameter vector. That is, we assume that $\Pr(\mathbf{Z}_i = z_i)$ is Multinomial$(1; \gamma_{i1},\ldots,\gamma_{iK})$ for $i = 1,\ldots,n$. Therefore, $\mathbf{\Gamma} = (\boldsymbol{\gamma}_1,\ldots,\boldsymbol{\gamma}_n)$ is the variational parameter.

*Corresponding author. Email address: `lzxue@psu.edu`

With $A(z) = \Pr(\mathbf{Z} = z)$, we obtain the following explicit lower bound of the log-likelihood:

$$\begin{aligned}
\mathrm{LB}(\boldsymbol{\pi}, \boldsymbol{\theta}; \boldsymbol{\Gamma}) =& \mathrm{E}_{\boldsymbol{\Gamma}} \{\ln \Pr(\mathbf{y}_1, \ldots, \mathbf{y}_T, z \mid \mathbf{y}_0)\} - \mathrm{E}_{\boldsymbol{\Gamma}} \{\ln \Pr(\mathbf{Z} = z)\} \\
=& \mathrm{E}_{\boldsymbol{\Gamma}} \{\ln \Pr(\mathbf{y}_1, \ldots, \mathbf{y}_T \mid \mathbf{y}_0, z) + \ln \Pr(\mathbf{Z} = z \mid \mathbf{y}_0)\} - \mathrm{E}_{\boldsymbol{\Gamma}} \{\ln \Pr(\mathbf{Z} = z)\} \\
=& \mathrm{E}_{\boldsymbol{\Gamma}} \left\{ \ln \prod_{t=1}^{T} \Pr(\mathbf{Y}_t = \mathbf{y}_t \mid \mathbf{y}_0, z) \right\} + \mathrm{E}_{\boldsymbol{\Gamma}} \{\ln \Pr(\mathbf{Z} = z \mid \mathbf{y}_0) - \ln \Pr(\mathbf{Z} = z)\} \\
=& \sum_{t=1}^{T} \sum_{i<j}^{n} \sum_{k=1}^{K} \sum_{l=1}^{K} \gamma_{ik} \gamma_{jl} \ln \Pr(Y_{t,ij} = y_{t,ij} \mid y_{t-1,ij}, Z_{ik} = Z_{jl} = 1) + \sum_{i=1}^{n} \sum_{k=1}^{K} \gamma_{ik} (\ln \pi_k - \ln \gamma_{ik}).
\end{aligned}$$

1.2. Constructing a surrogate function

To construct the surrogate function $Q(\boldsymbol{\pi}^{(\tau)}, \boldsymbol{\theta}^{(\tau)}, \boldsymbol{\Gamma}^{(\tau)}; \boldsymbol{\Gamma})$ of equation (10), which satisfies equations (11) and (12), we separate the lower bound of the log-likelihood, $\mathrm{LB}(\boldsymbol{\pi}^{(\tau)}, \boldsymbol{\theta}^{(\tau)}; \boldsymbol{\Gamma})$, into two parts,

$$\sum_{t=1}^{T} \sum_{i<j}^{n} \sum_{k=1}^{K} \sum_{l=1}^{K} \gamma_{ik} \gamma_{jl} \ln \Pr(Y_{t,ij} = y_{t,ij} \mid y_{t-1,ij}, Z_{ik} = Z_{jl} = 1)$$

and

$$\sum_{i=1}^{n} \sum_{k=1}^{K} \gamma_{ik} (\ln \pi_k^{(\tau)} - \ln \gamma_{ik}).$$

For the first part, since $\ln \Pr(Y_{t,ij} = y_{t,ij} \mid y_{t-1,ij}, Z_{ik} = Z_{jl} = 1) < 0$, the arithmetic-geometric mean inequality $A^2 + B^2 \geq 2AB$ implies that

$$\left(\gamma_{ik}^2 \frac{\gamma_{jl}^{(\tau)}}{2\gamma_{ik}^{(\tau)}} + \gamma_{jl}^2 \frac{\gamma_{ik}^{(\tau)}}{2\gamma_{jl}^{(\tau)}} \right) \ln \Pr(Y_{t,ij} = y_{t,ij} \mid y_{t-1,ij}, Z_{ik} = Z_{jl} = 1) \leq \gamma_{ik} \gamma_{jl} \ln \Pr(Y_{t,ij} = y_{t,ij} \mid y_{t-1,ij}, Z_{ik} = Z_{jl} = 1).$$

Hence, the first part of our surrogate function is written as

$$\sum_{t=1}^{T} \sum_{i<j}^{n} \sum_{k=1}^{K} \sum_{l=1}^{K} \left(\gamma_{ik}^2 \frac{\gamma_{jl}^{(\tau)}}{2\gamma_{ik}^{(\tau)}} + \gamma_{jl}^2 \frac{\gamma_{ik}^{(\tau)}}{2\gamma_{jl}^{(\tau)}} \right) \ln \Pr(Y_{t,ij} = y_{t,ij} \mid y_{t-1,ij}, Z_{ik} = Z_{jl} = 1).$$

For the second part, the concavity of the logarithm function means that its graph is bounded above by any tangent line, giving the inequality

$$-\log \gamma_{ik} \geq -\log \gamma_{ik}^{(\tau)} - \frac{\gamma_{ik}}{\gamma_{ik}^{(\tau)}} + 1.$$

Therefore, the second part of our surrogate function is written as

$$\sum_{i=1}^{n} \sum_{k=1}^{K} \gamma_{ik} \left(\log \pi_k^{(\tau)} - \log \gamma_{ik}^{(\tau)} - \frac{\gamma_{ik}}{\gamma_{ik}^{(\tau)}} + 1 \right).$$

Combining these two parts, we have the surrogate function $Q(\boldsymbol{\pi}^{(\tau)}, \boldsymbol{\theta}^{(\tau)}, \boldsymbol{\Gamma}^{(\tau)}; \boldsymbol{\Gamma})$, which equals

$$\sum_{t=1}^{T} \sum_{i<j}^{n} \sum_{k=1}^{K} \sum_{l=1}^{K} \left(\gamma_{ik}^2 \frac{\gamma_{jl}^{(\tau)}}{2\gamma_{ik}^{(\tau)}} + \gamma_{jl}^2 \frac{\gamma_{ik}^{(\tau)}}{2\gamma_{jl}^{(\tau)}} \right) \ln \Pr(Y_{t,ij} = y_{t,ij} \mid y_{t-1,ij}, Z_{ik} = Z_{jl} = 1) + \sum_{i=1}^{n} \sum_{k=1}^{K} \gamma_{ik} \left(\log \pi_k^{(\tau)} - \log \gamma_{ik}^{(\tau)} - \frac{\gamma_{ik}}{\gamma_{ik}^{(\tau)}} + 1 \right).$$

2. A simulation study with unbalanced clusters

We provide an additional simulation study to check the performance of our proposed algorithm in an unbalanced clusters setting, called Model 7, in which the mixing parameters are unequal but all other parameters are the same as in Model 1.

Table 1
Parameter values for a model with unequal mixing proportions.

	Model 7	
	G_1	G_2
Mixing proportion π_k	0.2	0.8
Stability parameter θ_k^s	-0.5	0.5
Initial network density parameter θ_k^d	-0.5	0.5

We check the performance of our criterion functions in choosing the correct number of groups. As shown in Table 2, both CL-BIC and modified ICL perform well. Next, we assess the clustering performance. The average value of RI and NMI results are reported in Table 3.

Table 2
Frequencies of min CL-BIC and max ICL over 100 repetitions. K_0 represents the true number of groups.

	Model 7 ($K_0 = 2$)			
	$K = 1$	$K = 2$	$K = 3$	$K = 4$
min CL-BIC	0	98	0	2
max ICL	0	100	0	0

Table 3
Mean values of Rand Index (RI) and Normalized Mutual Information (NMI) for 100 repetitions for various models and values of K, with sample standard deviations in parentheses, where K_0 is the true number of groups.

	Model 7 ($K_0 = 2$)		
	$K = 2$	$K = 3$	$K = 4$
RI	1.000 (0.000)	0.686 (0.033)	0.586 (0.053)
NMI	1.000 (0.000)	0.479 (0.040)	0.373 (0.046)

Finally, Table 4 summarizes estimation performance of our algorithm. The results of Tables 2 through 4 together tell us that our algorithm performs convincingly on this test case.

Table 4
Average values of ℓ^2 norm loss for estimated mixing proportions and network parameters over 100 repetitions with standard deviations shown in parentheses. K_0 represents the true number of groups.

Model 7 ($K_0 = 2$)	
$\text{RSE}_{\boldsymbol{\pi}}$	$\text{RSE}_{\boldsymbol{\theta}^s}$
0.045 (0.032)	0.020 (0.013)