Bayesian Model-Based Clustering for Dynamic Count Networks

Kekempanos Angelos

Athens University of Economics and Business

October 25, 2021

Focus of this Talk

This talk aims at introducing Dynamic Count Networks (DCN) and as well as presenting a Bayesian Latent Space Model for community detection (clustering) in DCN.

The presentation consists predominantly of:

- Introduction to Dynamic Count Networks
- Definition of a Latent Space Model for DCN clustering
- Model parameter estimation using Bayesian inference
- Simulation of a DCN and model implementation

What is a Count Network?

Count Network Definition

Count Network is a mathematical structure, that uses count data (e.g. emails, calls) to describe pairwise relations between objects. It consists of vertices and the number of events between them.

Count Adjacency Matrix Definition

The count adjacency matrix Y for a count network of N nodes is an $N \times N$ matrix of which each cell is defined as:

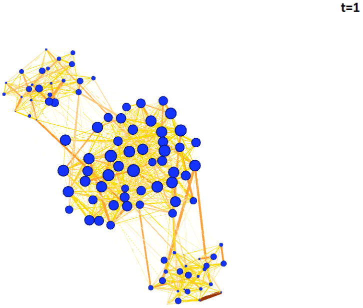
$$y_{ij} = \left\{ \text{number of events between the } i^{th} \text{ and } j^{th} \text{ actors of the network} \right\}.$$

What is a Dynamic Count Network?

Dynamic Count Network (DCN) Definition

Dynamic Count Network (DCN) is a count network, that changes over time and can be described by a count adjacency cube Y, where:

$$y_{ij}^{(t)} = \begin{Bmatrix} \text{number of events between the } i^{th} \text{ and } j^{th} \\ \text{actors of the network at time t .} \end{Bmatrix}$$



Outline

Model Definition

- We assume that the event rates between the i^{th} and j^{th} actors of the network at time t can be modelled as:

$$log(\lambda_{ij}^{(t)}) = \gamma^{(t)} \mathbb{1}_{\left\{y_{ij}^{(t-1)} - y_{ij}^{(t-2)} > 0\right\}} + \delta^{(t)} \mathbb{1}_{\left\{y_{ij}^{(t-1)} - y_{ij}^{(t-2)} \leq 0\right\}} - |W_i - W_j|$$

  where, $\gamma^{(t)}$ is an increasing count parameter and $\delta^{(t)}$ a decreasing-stable count parameter. W_i refers to the network actors latent positions in a $d-$dimensional Euclidean space and $|\cdot|$ is the Euclidean norm.

- We can express the likelihood of the above model as:

$$\mathcal{L}_y = \prod_{t=3}^{T} \prod_{i=1}^{N} \prod_{j:i<j}^{N} \left\{Poisson(\lambda_{ij}^{(t)})\right\}$$

Model Definition

- To represent clustering, we assume that W_i's are drawn from a finite mixture of G multivariate normal distributions (FMG). Each MVN has a different mean vector and a spherical covariance matrix with variances, that differ between groups:

$$W_i \sim \sum_{g=1}^{G} \pi_g MVN_d(\mu_g, \sigma_g^2 \mathrm{I}_d)$$

where, π_g is the probability that an actor belongs to the g^{th} group, so that $\pi_g \geq 0$ and $\sum_{g=1}^{G} \pi_g = 1$.

Outline

Prior Distributions

- For the FMG part of the model, we introduce a new variable k_i equal to g if the i^{th} actor belongs to the g^{th} group as is standard in Bayesian estimation of mixture models.

$$\pi \sim Dirichlet(\nu)$$
$$\sigma_g^2 \sim \sigma_0^2 InvX_\alpha^2$$
$$\mu_g \sim MVN_d(0, \omega^2 \mathrm{I}_d)$$

and

$$P(W, K|\pi_g, \mu_g, \sigma_g^2) = \prod_{g=1}^{G} \prod_{i=1}^{N} \left(\pi_g N_d(\mu_g, \sigma_g^2 \mathrm{I}_d)\right)^{\mathbb{1}_{\{k_i = g\}}}$$

Prior Distributions

- For the parameters γ and δ we assume random walk priors:
$$f(\gamma|\tau_\gamma,\tau_\gamma^0) = f\left(\gamma^{(1)};0,\frac{1}{\tau_\gamma^0}\right)\prod_{t=2}^{T} f\left(\gamma^{(t)};\gamma^{(t-1)},\frac{1}{\tau_\gamma}\right)$$
$$f(\delta|\tau_\delta,\tau_\delta^0) = f\left(\delta^{(1)};0,\frac{1}{\tau_\delta^0}\right)\prod_{t=2}^{T} f\left(\delta^{(t)};\delta^{(t-1)},\frac{1}{\tau_\delta}\right)$$
where, $f(x;m,v)$ is a Gaussian density estimated at x with mean m and variance v.
- For the prior information of the precision parameters τ_γ,τ_γ^0,τ_δ,τ_δ^0 we propose a $Gamma(A,B)$ density.

Posterior Distributions

The full conditional posterior distributions are:

- $W_i|K_i = g, others \propto \phi_d(w_i; \mu_g, \sigma_g^2 \mathrm{I}_d) \times P(Y|W, \beta, \gamma, \delta)$
- $\pi|others \propto Dirichlet(m + \nu)$
- $\mu_g|others \propto MVN_d\left(\frac{\mu_g \overline{w_g}}{\mu_g + \sigma_g^2/\omega^2}, \frac{\sigma_g^2}{\mu_g + \sigma_g^2/\omega^2}\mathrm{I}\right)$
- $\sigma_g^2|others \propto \left(\frac{\sigma_0^2 + ds_g^2}{\alpha + \mu_g d}\right) InvX^2_{\alpha + \mu_g d}$
- $P(K_i = g|others) = \frac{\pi_g \phi_d(w_i; \mu_g, \sigma_g^2 \mathrm{I}_d)}{\sum_{r=1}^{G} \phi_d(w_i; \mu_r, \sigma_r^2 \mathrm{I}_d)}$

where,

$$m_g = \sum_{i=1}^{n} \mathbb{1}_{\{K_i = g\}}$$

$$s_g^2 = \frac{1}{d} \sum_{i=1}^{n} (w_i - \mu_g)^T (w_i - \mu_g) \mathbb{1}_{\{K_i = g\}}$$

$$\overline{w_g} = \frac{1}{m_g} \sum_{i=1}^{n} w_i \mathbb{1}_{\{K_i = g\}}$$

Posterior Distributions

- $\gamma^{(t)}|others \propto P(Y|W,\beta,\gamma,\delta)\times f\left(\gamma^{(t)};0,\frac{1}{\tau_\gamma^0}\right)^{\mathbb{1}\{t=1\}} f\left(\gamma^{(t)};\gamma^{(t-1)},\frac{1}{\tau_\gamma}\right)^{\mathbb{1}\{t>1\}} f\left(\gamma^{(t+1)};\gamma^{(t)},\frac{1}{\tau_\gamma}\right)^{\mathbb{1}\{t<T\}}$
- $\delta^{(t)}|others \propto P(Y|W,\beta,\gamma,\delta)\times f\left(\delta^{(t)};0,\frac{1}{\tau_\delta^0}\right)^{\mathbb{1}\{t=1\}} f\left(\delta^{(t)};\delta^{(t-1)},\frac{1}{\tau_\delta}\right)^{\mathbb{1}\{t>1\}} f\left(\delta^{(t+1)};\delta^{(t)},\frac{1}{\tau_\delta}\right)^{\mathbb{1}\{t<T\}}$
- $\tau_\gamma \sim Gamma\left(A+\frac{T-1}{2}, B+\frac{\sum_{t=1}^{T}\left(\gamma^{(t)}-\gamma^{(t-1)}\right)^2}{2}\right)$
- $\tau_\gamma^0 \sim Gamma\left(A+\frac{1}{2}, B+\frac{\left(\gamma^{(1)}\right)^2}{2}\right)$
- $\tau_\delta \sim Gamma\left(A+\frac{T-1}{2}, B+\frac{\sum_{t=1}^{T}\left(\delta^{(t)}-\delta^{(t-1)}\right)^2}{2}\right)$
- $\tau_\delta^0 \sim Gamma\left(A+\frac{1}{2}, B+\frac{\left(\delta^{(1)}\right)^2}{2}\right)$

MCMC Algorithm

Step 1: use of Metropolis-Hastings algorithm to sample $W_i, \forall i = 1, ..., N$, setting as a proposal density a d-variate Gaussian distribution $N_d(w_{old}, \delta_w^2 \mathrm{I}_d)$.

Step 2: use of Gibbs sampling algorithm to generate samples for μ_g, σ_g^2 and π_g.

Step 3: use of random walk Metropolis algorithm to update the constant parameters γ and δ.

Step 4: use of Gibbs sampling algorithm to sample from the conjugate posterior densities of $\tau_\gamma, \tau_\gamma^0, \tau_\delta, \tau_\delta^0$.

Label Switching Problem

For Bayesian mixtures the invariance of the likelihood to permutations in the labelling is referred to as *label switching problem*.

ECR Algorithm (Papastamoulis and Iliopoulos, 2010)

It is based on the idea that equivalent allocation vectors are mutually exclusive from the label switching solution.

1. Define a pivot allocation vector $k^* = (k_1^*, \ldots, k_N^*)$. The pivot is selected by choosing a high-posterior density point.
2. For $iter = 1, \ldots, N_{iter}$ find a permutation $p^{(iter)} \in \mathcal{P}$ that maximizes $\sum_{i=1}^{N} \mathbb{1}_{\{pk^*_{iter} = k^*_i\}}$

Choosing the number of clusters

- BIC approximation for the dynamic Poisson regression part of the model:

$$BIC_{DPR} = 2log[P\{Y|\hat{W}, \hat{\gamma}, \hat{\delta}\}] - 2T\ log\left(\frac{TN(N-1)}{2}\right)$$

- BIC approximation for the FMG part of the model:

$$BIC_{FMG} = 2log[P\{\hat{W}|\hat{\theta}\}] - d_{FMG}log(N)$$

  where, d_{FMG} is the number of parameters in the clustering model and N the number of actors.

- BIC approximation for the entire model:

$$BIC = BIC_{DPR} + BIC_{FMG}$$

Outline

Simulation

Step 1: Assume that the network has the following fundamental properties $G = 3$, $d = 2$, $T = 20$ and $N = 100$.

Step 2: Generate samples for the actors latent positions W using as mixture proportions $\pi_g = (1/4, 1/2, 1/4)$, clusters centres $\mu_1 = (-3, -3)$, $\mu_2 = (0, 0)$, $\mu_3 = (3, 3)$ and variances $\sigma_g^2 = (0.4^2, 0.2^2, 0.3^2)$.

Step 3: Generate samples for the constant parameters $\gamma^{(t)}$ and $\delta^{(t)}$ using a random walk process.

$$\gamma^{(t)} \sim N(\gamma^{(t-1)}, 1/2)$$
$$\delta^{(t)} \sim N(\delta^{(t-1)}, 1/2)$$

Simulation

Step 4: For $t = 1, 2$ we have that:

$$Y_{ij}^{(t)} \sim Poisson\left(\frac{1}{|W_i - W_j|}\right)$$

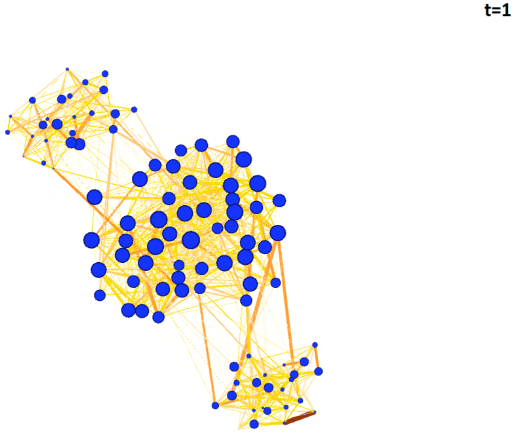
and for $t = 3, ..., T$

$$Y_{ij}^{(t)} \sim Poisson(\lambda_{ij}^{(t)})$$

where

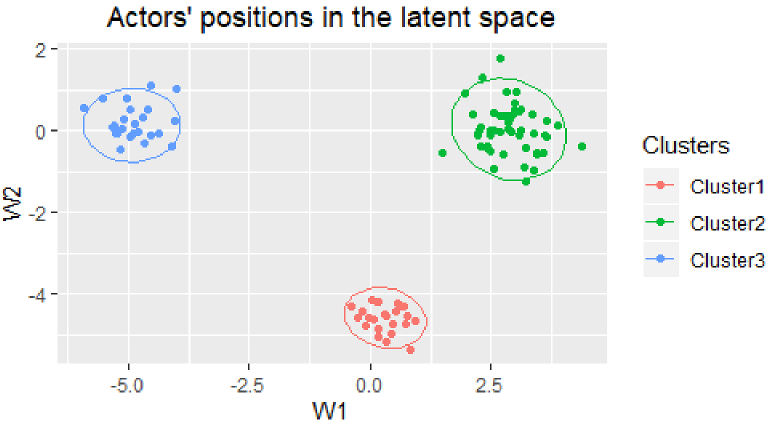
$$\lambda_{ij}^{(t)} = exp\left(\gamma^{(t)}\mathbb{1}_{\left\{y_{ij}^{(t-1)}-y_{ij}^{(t-2)}>0\right\}} + \delta^{(t)}\mathbb{1}_{\left\{y_{ij}^{(t-1)}-y_{ij}^{(t-2)}\leq 0\right\}} - |W_i - W_j|\right)$$

Simulation

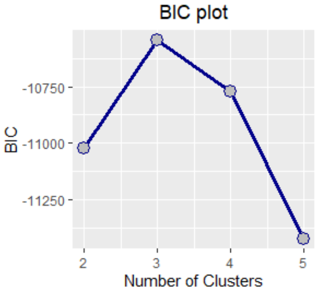


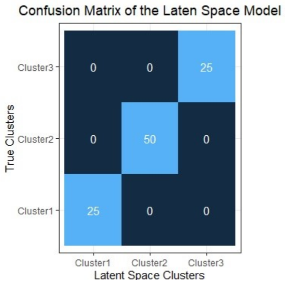
Model fitting

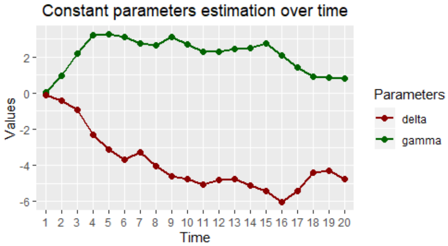
The MCMC algorithm ran for 20000 iterations, of which the first 8000 were used as burn in period. The values of the model hyper-parameters were defined as $\delta^2_W = 1$, $\nu = c(3,3,3)$, $\omega^2 = 2$, $\alpha = 2$ and $\sigma^2_0 = 0.1$.



Model fitting







Outline

Conclusions

- Accurate estimation of the dynamic count network communities.
- Time consuming runs of the MCMC algorithm.
- Difficulty of tuning the hyper-parameters.
- Fitting the DCN latent space model to real data.
- Use of a more time series approach for clustering DCN.

Outline

References

Mark S. Handcock, Adrian E. Raftery, Jeremy M. Tantrum (2007).
"Model-Based Clustering for Social Networks"
Royal Statistical Society, pages 302–308.
https://doi.org/10.1111/j.1467-985X.2007.00471.x

Nial Friel, Riccardo Rastelli, Jason Wyse and Adrian E. Raftery (2016).
"Interlocking directorates in Irish companies using a latent space model for bipartite networks"
Proceedings of the National Academy of Sciences, pages 302–308.
https://doi.org/10.1073/pnas.1606295113

Papastamoulis P. and Iliopoulos G. (2010).
An artificial allocations based solution to the label switching problem in Bayesian analysis of mixtures of distributions.
Journal of Computational and Graphical Statistics 19, 313-331