Bayesian Model-Based Clustering for Dynamic Count Networks

Kekempanos Angelos†

†Athens University of Economics and Business

Objectives

Objectives for today:

- Introduction to DCN.
- Definition of a Latent Space Model for DCN clustering.
- Bayesian Inference.
- Simulation and Model Implementation.

What is a DCN?

Count Network

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

Dynamic Count Network

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

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

Latent Space Model

Model Definition

We assume that:

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

$\forall i, j = 1, ..., N$ and $t = 1, ..., T$

The functional form of the model is:

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

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.

Clustering definition

To represent clustering, we assume that W_i's are drawn from a finite mixture of G multivariate normal distributions (FMG).

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

where, π_g is the probability that an actor belongs to the g^{th} group.

Model Selection

We choose the model with the maximum value of a BIC approximation

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

Label Switching

We solve the label switching problem using the Equivalence Classes Representatives (ECR) algorithm.

Bayesian Inference

MCMC Algorithm

Step 1

Use of Metropolis-Hastings algorithm to sample $W_i, \forall i = 1, ..., N.$

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

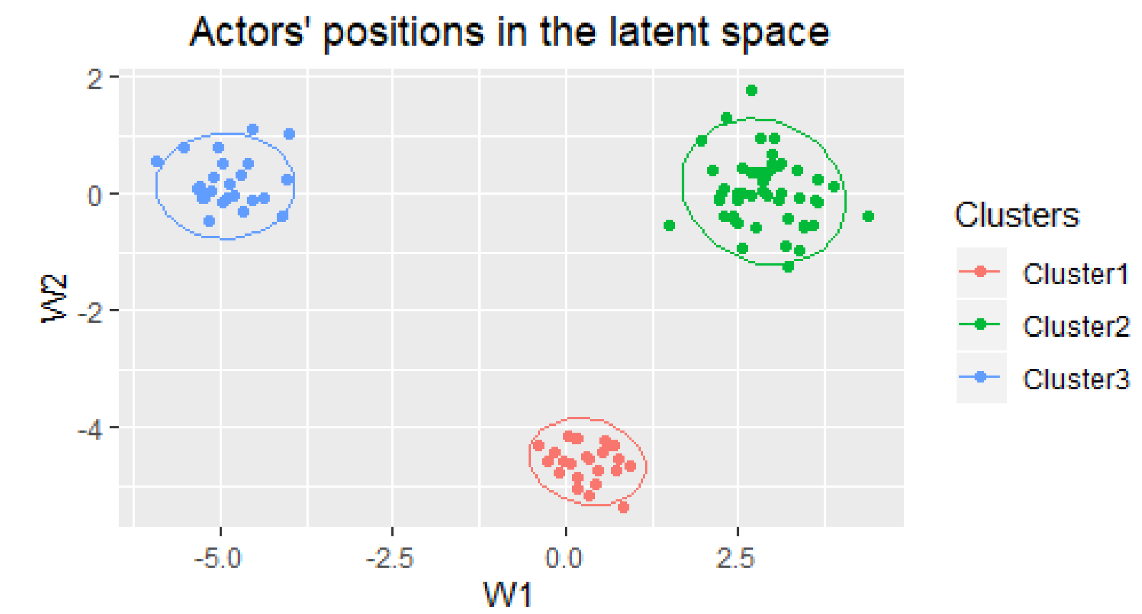
Use of Gibbs sampling algorithm to sample from the conjugate posterior densities of the precision parameters.

Implementation

Simulation

We simulate a DCN with $G = 3$ clusters, $d = 2$ euclidean space dimensions, $T = 20$ time periods and $N = 100$ nodes, without overlapping between the clusters.

Model Fitting



- Adjusted Rand Index = 1
- BIC = -10545.09

References

- Mark S. Handcock, Adrian E. Raftery, Jeremy M. Tantrum (2007). "Model-Based Clustering for Social Networks" *Royal Statistical Society*, pages 302–308.
- 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
- Papastamoulis, P. (2014). Handling the label switching problem in latent class models via the ECR algorithm. *Communications in Statistics-Simulation and Computation*, 43(4):913–927.

.