Mixture Models for Graph Clustering

Sofiane Ezzehi [1] Bastien Le Chenadec [1] Theïlo Terrisse [1]

École des Ponts ParisTech

Context

The Stochastic Block Model

The **Stochastic Block Model** is a *mixture* model for graphs. Each node is assigned to a *class*, and conditionally on these classes the probability of an edge between two nodes i and j only depends on their class memberships. This model draws inspiration from mixture models for distribution of degrees, and the Erdös-Rényi model which deals with the probability for two given nodes of being connected.

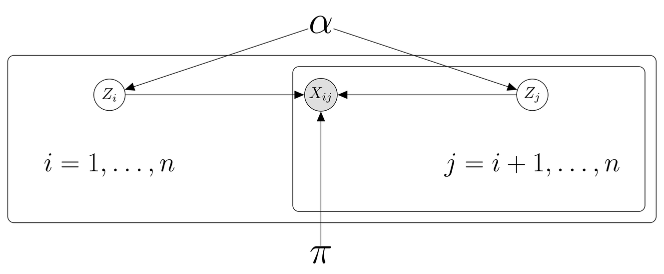

Figure 1. Graphical model of the SBM

Formally, we consider an undirected graph with n nodes and no self-loops. We denote X the adjacency matrix of this graph, thus $X_{ij} \in \{0, 1\}$ denotes the existence of an edge between nodes i and j.

The model is parametrized by Q the number of classes, $\alpha \in [0, 1]^Q$ the prior distribution on the classes, and $\pi \in [0, 1]^{Q \times Q}$ the probability of an edge between two nodes of different classes. We also introduce the random variables $Z_i \in \{0, 1\}^Q$ for $i \in [\![1, n]\!]$, that represent the membership of node i to each class. The prior distributions on Z and X are given by :

$$\begin{cases} \forall i \in [\![1, n]\!], \quad \sum_{q=1}^{Q} Z_{iq} = 1 \quad \text{(unique class)} \\ \forall q \in [\![1, Q]\!], \quad \mathbb{P}(Z_{iq} = 1) = \alpha_q \quad \text{(class distribution)} \\ \forall q, l \in [\![1, Q]\!], \quad \forall i \neq j \in [\![1, n]\!], \quad \mathbb{P}(X_{ij} = 1 \mid Z_{iq} = 1, Z_{jl} = 1) = \pi_{ql} \quad \text{(edge probability)} \end{cases} \tag{1}$$

The variational Expectation-Maximization algorithm

Variants

Experiments on an SBM dataset

Second set of experiments

References

[1] J.-J. Daudin, F. Picard, and S. Robin.
A mixture model for random graphs.
Statistics and Computing, 18(2):173–183, June 2008.