Bayesian Inference for Mixtures of Stochastic Block Models

Cassiano O. Becker¹, Francisco J. Barreras¹, and Victor M. Preciado¹

1. University of Pennsylvania, Philadelphia, PA, USA

Analyzing the community structure in networks is an essential tool for describing the organization of complex systems. In that regard, a common statistical framework for describing communities in networks is the Stochastic Block Model (SBM) [1]. The SBM has been comprehensively studied, both in terms of its theoretical guarantees and in seeking extensions to consider richer models including weighted edges, skewed degree distributions, mixed communities, etc. Besides analyzing community structure on a single network, an extension that has become increasingly relevant consists of analyzing multiple networks simultaneously. While there have been attempts to integrate multiple networks to this analysis [2], there is still a need for a probabilistic generative model that allows a description of multiple networks according to the SBM framework. In particular, with such model we could inquire whether a set of networks can be clustered by grouping networks with similar community structure. Towards that goal, we propose a probabilistic model consisting of a mixture of SBMs, in which every network is associated with a probability of belonging to each of multiple models. Because exact inference for mixture models of this type is intractable, we develop a variational inference framework [3] to simultaneously cluster the networks and learn the SBM model associated with each cluster. Further, to overcome multiple local maxima associated with the objective function (ELBO), we perform gradient ascent on the natural parameter space and adopt a deterministic annealing strategy. The learned model can be used for predictive tasks including classification of networks, link prediction and data reduction. We illustrate the effectiveness of our algorithm through experiments on synthetic data as well as real networks derived from human brains.


Figure 1: A unified probabilistic generative model for mixtures of SBMs enables simultaneously learning model prototypes and clustering networks according to their latent community structure. In (a), we see a graphical model where \( x_{k,i,j} \) are the observed edges for networks \( k = 1, \ldots, K \); \( z_{k,m,i} \) and \( y_i \) are latent variables representing, respectively, community and cluster (model) membership; and \( \gamma_m \) and \( \pi_m \), \( m = 1, \ldots, M \) represent different SBMs. In (b), we show a schematic of the problem to be solved on a set of 25 networks, represented by their binary adjacency matrices. We wish to simultaneously identify the clusters of networks and the block model associated with each cluster.