Estimating Latent Variable Densities for Exchangeable Network Models

Sharmodeep Bhattacharyya, Peter J. Bickel and Patrick J. Wolfe

September 26, 2014

Exchangeable network models provide a general non-parametric class of models for unlabeled random graphs. The model is derived from the representation of *exchangeable random infinite array* by Aldous and Hoover (1983). Given the adjacency matrix $\{A_{ij}\}_{i,j=1}^{n}$ as the data, we can define $\mathbb{P}(\{A_{ij}\}_{i,j=1}^{n})$ conditionally given latent variables $\{\xi_i\}_{i=1}^{n}$ associated with vertices $\{v_i\}_{i=1}^{n}$ respectively as (Bickel & Chen (2009), Bollobás et.al. (2007), Hoff et.al. (2002)).

$$\begin{aligned} \xi_1, \ldots, \xi_n &\overset{\text{iid}}{\sim} \mathcal{U}(0,1) \\ \Pr(A_{ij} = 1 | \xi_i = u, \xi_j = v) &= h_n(u,v) = \rho_n w(u,v), \end{aligned}$$

where $w(u,v)$ is the conditional *latent variable density* or *graphon* (Lovász and Szegedy (2006)) given $A_{ij} = 1$. h_n is not uniquely defined as $h_n\big(\varphi(u), \varphi(v)\big)$, with measure-preserving φ, gives the same model. Our main goal in this work is present an unified framework for estimation of the equivalence class of latent variable density $w(\cdot,\cdot)$ under measure-preserving transformations.

Recently there has been some focus in statistics on estimation of latent variable densities and its use as an exploratory tool for network data analysis (Wolfe and Olhede (2013), Latouche and Robin (2013), Chan and Airoldi (2014)). The main approach is that exchangeable network models are approximated by stochastic block models and latent variable densities are estimated using piece-wise constant or histogram-type estimators derived from the fitted block models. We show that if the latent variable density has some smoothness properties and the fitted block models, where the fitting can be done by any method including spectral method, profile likelihood, variational likelihood or modularity, satisfies certain consistency properties, then we can have consistent estimators for the equivalence class of latent variable densities under suitable metrics. We derive rates of convergence for the fitted block models estimating the latent variable density for appropriate metrics depending on the smoothness properties of the latent variable density as well as the error rates of the method of fitting the block models.

We also propose a cross-validation method using count statistics and their smooth functions to regularize the fitted block model and choose the size of the block model approximating the latent variable density. The cross-validation method provides us with a procedure to regularize the fitted block models independent of the algorithm used to fit the block model.

A simulation study and illustration on real networks to estimate the latent variable densities are also provided.