I. Castillo and T. Rebafka *(Université Pierre et Marie Curie – Paris 6)*

The random graph model proposed by Caron and Fox has a number of remarkable properties, among others the possibility to combine sparsity and exchangeability and the scalability of computations within the associated Bayesian framework. In our view, their work rises a number of interesting questions on *statistical inference*. Adopting a frequentist approach, what can be said about posterior convergence assuming the data has been generated under a 'true' parameter?

Form a posterior distribution $\Pi[\cdot \mid Z]$ on $w \in (\mathbb{R}^+)^{\mathbb{N}^+} =: \mathcal{W}$ from

$$Z \mid w \sim \bigotimes_{i=1}^{n} \text{Bernoulli}(1 - e^{-2w_i w_j}) =: P_w, \qquad w \sim \Pi.$$

Consider the behaviour of $\Pi[\cdot \mid Z]$ under two frequentist settings: A) *well-specified*, where $Z \sim P_0 = P_{w_0}$ for an unknown fixed $w_0 \in \mathcal{W}$, B) possibly *misspecified*, where $Z \sim P_0 = Q$ for an arbitrary graph distribution Q.

In a simulation study, we considered estimation of two simple functionals, the edge density and the density of triangles

$$\psi_1 = \frac{1}{\binom{n}{2}} E_{P_0}\Big[\sum_{i<j} Z_{ij}\Big], \quad \psi_2 = \frac{1}{\binom{n}{3}} E_{P_0}\Big[\sum_{i<j<k} Z_{ij} Z_{jk} Z_{ki}\Big].$$

We used the default code under GGP with improper priors on hyperparameters.

One reason to consider setting B) is the specific exponential form of the link function considered in the paper, which may not hold for the data. Suppose for instance that Z has actually been generated from a stochastic blockmodel (SBM) with 2 groups, equiproportions and connectivity parameters $\alpha = \begin{pmatrix} 0.8 & 0.1 \\ 0.1 & 0.8 \end{pmatrix}$. Simulations suggest that the posterior is consistent for ψ_1, but inconsistent for ψ_2.

In setting A), we considered two cases: 1) an equiproportions SBM with connectivity matrix $\alpha = \begin{pmatrix} 0.8 & \approx \frac{1}{3} \\ \approx \frac{1}{3} & 0.1 \end{pmatrix}$ compatible with the exponential link function. Bayesian and frequentist behaviours of ψ_1 and ψ_2 are remarkably close and rapidly converging, see Table 1, suggesting a Bernstein–von Mises theorem holds. 2) a graph with 'correctly specified' link function and w a sample from the Cauchy distribution. The posterior still estimates the functionals ψ_1, ψ_2 well, but seems to underestimate large values of $w_{0,i}$, see Figure 1. To study sparsity, we repeated the two previous experiments but replacing w_0 by $\rho_n w_0$, where $\rho_n \to 0$. We noticed that the posterior on σ was concentrated on negative values, which suggests that σ may not universally quantify sparsity.

It would be interesting to determine which aspects of w (e.g. real valued functionals, or the complete vector w) can be estimated at minimax rate using priors as in the paper. Another question would be to extend the model and priors to cover arbitrary link functions, and possibly more general forms of the law of w.

Table 1: SBM with equiproportions and compatible link function, true $\psi_1 = 0.3938$ and $\psi_2 = 0.1026$. Mean lengths of 95% credible/confidence intervals and bias of posterior mean and frequentist estimates of ψ_1 and ψ_2 over $120, 120, 90, 60$ simulated graphs, resp.

number of nodes n		30	50	100	300
credible	int. length	0.0794	0.0470	0.0234	0.0078
$\hat{\psi}_1^{\text{bayes}}$	bias	-0.0025	-0.0011	0.0003	-0.0003
$\hat{\psi}_1^{\text{freq}}$	int. length	0.0816	0.0486	0.0234	0.0078
	bias	-0.0012	-0.0010	0.0001	-0.0004
credible	int. length	0.0502	0.0302	0.0153	0.0051
$\hat{\psi}_2^{\text{bayes}}$	bias	-0.0058	-0.0043	-0.0018	-0.0007
$\hat{\psi}_2^{\text{freq}}$	int. length	0.0485	0.0322	0.0149	0.0054
	bias	-0.0053	-0.0037	-0.0013	-0.0005

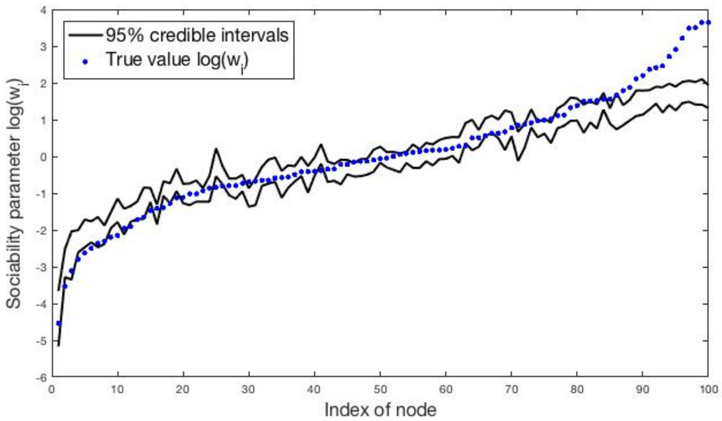

Figure 1: Credible intervals for $\log(w_i)$ and true values in a graph with 100 nodes (Nodes are ordered by increasing values of $\log(w_i)$).