1. Supporting Information: Additional information for this article is available

Appendix S1

Posterior computation steps

Conditional on the clusters $S_1, \ldots, S_H$, the full posterior satisfies

$$L \propto \prod_{i=1}^{n} N(\mathbf{x}_i; \mathbf{z}_i(\boldsymbol{\eta}_{s_1}, \ldots, \boldsymbol{\eta}_{s_p}) + \boldsymbol{\alpha}_i, \delta I_p) N(\boldsymbol{\alpha}_i; 0, \Sigma_G),$$
$$\times \prod_{h=1}^{H} \{N(\boldsymbol{\eta}_h; 0, diag(\tau_{h1}, \ldots, \tau_{hq})) \prod_{qq=1}^{q} Exp(\tau_{hq}; \lambda_h^2/2)\} \pi(\Sigma_G \mid G) \pi(G|S_1, \ldots, S_H) \pi(S_1, \ldots, S_H),$$

where $s_1, \ldots, s_p$, denote the cluster memberships. For updating the cluster memberships, we utilize the slice sampler of Walker (2007), which speeds up the computations by restricting the possible cluster allocations for each column. The slice sampler introduces uniform latent variables which allows the following form for the density

$$f(\mathrm{x}_i \mid \mathrm{z}_i, \boldsymbol{\alpha}_i, \delta) = \prod_{j=1}^{p} \sum_{l=1}^{\infty} w_l N(x_{ij}; \mathrm{z}_i \boldsymbol{\eta}_l + \alpha_{ij}; \delta) \pi(\boldsymbol{\eta}_l) = \prod_{j=1}^{p} \sum_{l \in N_w(u)} w_l N(x_{ij}; \mathrm{z}_i \boldsymbol{\eta}_l + \alpha_{ij}; \delta) \pi(\boldsymbol{\eta}_l), \; i = 1, \ldots, n, \quad (1)$$

where $N_w(u) = \{l : w_l > u\}$ and $w_l = v_l \prod_{k<l}(1 - v_k)$, $v_k \sim Be(1, M)$ as in equation (1). Let $\boldsymbol{\tau}_h = \{\tau_{hj} : j \in S_h\}$, $h = 1, \ldots, H$, and $D_{\boldsymbol{\tau}_h} = \mathrm{diag}(\tau_{h1}, \ldots, \tau_{hq})$. Note that $\pi(\delta)$ is chosen such that the posterior samples of $\delta = O(10^{-3})$, so that all posterior distributions involving a δ^{-1} term can be computed without difficulty. The MCMC steps are detailed below.

Step 1.1: Update the ν's after marginalizing out the augmented uniform variable using $\pi(\nu_h|-) = Be(1 + p_h, \sum_{j>h} p_j + M)$, where p_h is the cardinality of S_h, $h = 1, \ldots, H$.
Step 1.2: Update the augmented uniform variables from its full conditional as described in Walker (2007).
Step 2: Update the allocation of atoms to different subjects using $f(s_j = h|X, u_j) \propto N(\mathrm{x}^{(j)}; \boldsymbol{\eta}_h + \alpha_j 1_n, \delta I_n) I(h \in N_w(u_j))$, $j = 1, \ldots, p$, $h = 1, \ldots, H$, with $N_w(u_j)$ defined as in (**??**).
Step 3: Update the precision parameter using $\pi(M|-) = Ga(a_m + H, b_m - \sum_{l=1}^{H} \log(1 - \nu_l))$, where H is the number of clusters in the particular iteration and where $M \sim Ga(a_m, b_m)$.
Step 4: For the h-th cluster update the atom $\boldsymbol{\eta}_h$ using $\pi(\boldsymbol{\eta}_h|-) = N\Big(\boldsymbol{\eta}_h; (D_{\boldsymbol{\tau}_h}^{-1} + \delta^{-1} p_h I_n)^{-1} (\delta^{-1} Z^T \sum_{j \in S_h} (\mathrm{x}^{(j)} - \boldsymbol{\alpha}^{(j)})), (D_{\boldsymbol{\tau}_h}^{-1} + \delta^{-1} p_h I_n)^{-1}\Big)$, $h = 1, \ldots, H$.
Step 5: Update δ using $\pi(\delta^{-1}|-) = \mathrm{Ga}\Big(a_\delta + np/2, b_\delta + \frac{1}{2}(Y - XB - A)^T (Y - XB - A)\Big)$, where $A = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n)^T$.
Step 6: Update $\boldsymbol{\alpha}_i$ using $\pi(\boldsymbol{\alpha}_i|-) = N\Big(\boldsymbol{\alpha}_i; (\delta^{-1} I_p + \Omega)^{-1} \delta^{-1} (\mathrm{x}_i - \mathrm{z}_i B), (\delta^{-1} I_p + \Omega)^{-1}\Big)$.
Step 7: Given the graph $G_{(S_1, \ldots, S_H)}$, update Ω using $\pi(\Omega \mid -) = \mathrm{HIW}_G\Big(b + n, D + \sum_{i=1}^{n} \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^T\Big)$. .
Step 8: Propose a new graph G^* by either (a) adding an edge to G with probability α_G, or (b) deleting an edge from G with probability $1 - \alpha_G$. Accept G^* with probability $\min\Big(1, \frac{L(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n | Z, B, G) p(G^*) q(G|G^*)}{L(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n | Z, B, G) p(G) q(G^*|G)}\Big)$, where $q(G|G^*)/q(G^*|G) = a_G/(1 - a_G)$ for (a) and $(1 - a_G)/a_G$ for (b), and $p(G) = \Gamma(a_{\omega,c} + t_G)\Gamma(b_{\omega,c} + p(p-1)/2 - t_G)/\Gamma(a_{\omega,c} + b_{\omega,c} + p(p-1)/2)$. Here t_G is the total number of unique edges in G, $\Gamma(t) = (t-1)!$ denotes the Gamma function, and $a_{\omega,c} = a_{\omega 0}$ or $a_{\omega 1}$ depending on whether the nodes corresponding to the edge in question belongs to the same cluster or not, and similarly for $b_{\omega,c}$.
Step 9: Update λ_h using $\pi(\lambda_h|-) \sim Ga(q + a_\lambda, b_\lambda + 0.5 \sum_{l=1}^{q} |\beta_{hl}|)$, $h = 1, \ldots, H$.
Step 10: Update hyperparameters $\boldsymbol{\tau}_j$, $j = 1, \ldots, H$, similarly as in Park and Casella (2008).

References

Park, T, & Casella, G, (2008), The Bayesian Lasso, *Journal of the American Statistical Association*, **103**, 681–686.

Walker, SG, (2007), Sampling the Dirichlet mixture model with slices, . *Communications in Statistics: Simulation and Computation*, **36**, 45–54.