Proposal for spatial HDP

To incorporate the spatial information in the hdp model we set a prior on the topics that assures correlation between near points. In Gelfand et al. (2005) they have a Dirichlet process of multivariate normal distributions, where the dimension of the normal distribution reflects the number of locations. As we need our topics to be non-negative it would make sense to use the log-normal instead. Below, we are following the notation of the HDP model in Wang et al. (2011), where they have made a fast implementation of HDP. The base distribution is given by

$$\beta'_k \sim \text{Beta}(1, \gamma)$$
$$\beta_k = \beta'_k \prod_{l=1}^{k-1}(1 - \beta'_k)$$
$$\phi_k = e^{f_k} \text{ where } f_k \sim N_N(\mu_k(s), \sigma^2 H_n(\xi))$$
$$G_0 = \sum_{k=1}^{\infty} \beta_k \delta_{\phi_k}$$

where

$$(H_n(\xi))_{ij} = exp(-\xi||s_i - s_j||).$$

The construction for each observation is given by

$$\pi'_{jk} \sim \text{Beta}\left(\alpha_0 \beta_k, \alpha_0\left(1 - \sum_{l=1}^{k} \beta_l\right)\right)$$
$$\pi_{jk} = \pi'_{jk} \prod_{l=1}^{k-1}(1 - \pi'_{jk})$$
$$G_j = \sum_{k=1}^{\infty} \pi_{jk} \delta_{\phi_k}$$

What is highlighted in red was a symmetric Dirichlet in the original HDP model. Inspired by Townes and Engelhardt (2023) the topics/factors are vectors over the cells, which means the j in this case represents a gene. This way we can use the gaussian process approximation from that paper(using the new code from Luis) for the topics to incorporate the spatial information. The rest of the updates for the weights would still be updated as before using MFVB.

References

Gelfand, A. E., Kottas, A., and MacEachern, S. N. (2005). Bayesian nonparametric spatial modeling with dirichlet process mixing. *Journal of the American Statistical Association*, 100(471):1021–1035.

Townes, F. W. and Engelhardt, B. E. (2023). Nonnegative spatial factorization applied to spatial genomics. *Nature Methods*, 20(2):229–238.

Wang, C., Paisley, J., and Blei, D. M. (2011). Online variational inference for the hierarchical dirichlet process. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 752–760. JMLR Workshop and Conference Proceedings.