Conclusions

In this thesis, we extended the framework of edge exchangeability, previously developed for unipartite networks, to the bipartite setting. In particular, we demonstrated how the Pitman–Yor process, a fundamental tool in Bayesian Nonparametrics, can be employed to generate bipartite edge exchangeable networks by modelling each edge as a pair of values drawn independently from distinct Pitman–Yor processes.

We further established that the sparsity property proven by Crane and Dempsey (2018) for unipartite unweighted networks also holds in the bipartite case, requiring only minor adjustments to account for the bipartite structure. Simulation studies supported this theoretical result, showing that network density tends to zero on average as the number of edges grows. Additional insights were gained by examining the special case of the Dirichlet process, which was found empirically not to induce sparsity in bipartite networks.

Furthermore, we illustrated a fully Bayesian approach to perform inference on the parameters using the software Stan, which allows for efficient MCMC sampling from the posterior distributions. We assessed convergence of the posteriors to true parameter values under additional simulation studies. We also discussed the matters of prior choice, posterior predictive checking for model evaluation and hypothesis testing, in particular through the practical application to real data. Inference on the discount parameters is particularly useful for identifying which group of the bipartite network drives the sparsity mechanism. This can be assessed by testing whether the group specific Pitman–Yor process reduces to the Dirichlet process, which we showed does not induce sparsity.

Clear limitation of the proposed model is the independence assumption between the two groups. Allowing for more complex dependence structures would certainly increase the applicability of the model, which as of now remains rather restricted.

Moreover, there remains to prove if the model creates sparse graph sequences even when retaining multiple edges, in accordance to the unipartite version of the model. Furthermore, there were hints suggesting the existence of a relationship between the discount parameters and the speed of convergence towards zero of the density statistics. Finding a closed form for such relationship would shed light on the precise effect of the discount parameters in inducing sparsity.

We would like to mention that experiments presented in this thesis were carried out using the Grid'5000 testbed, supported by a scientific interest group hosted by Inria and including CNRS, RENATER and several Universities as well as other organizations (see https://www.grid5000.fr).

Bibliography

Harry Crane and Walter Dempsey. Edge Exchangeable Models for Interaction Networks. *Journal of the American Statistical Association*, 113(523):1311–1326, 2018. doi: 10.1080/01621459.2017.1341413.