Ensemble symmetries and the detectability limit of finite size stochastic block models

Jean-Gabriel Young,[1] Laurent Hébert-Dufresne,[2] Edward Laurence,[1] Patrick Desrosiers,[1] and Louis J. Dubé[1].

1. Département de Physique, de Génie Physique, et d'Optique, Université Laval, Québec, Qc, Canada
2. Santa Fe Institute, New Mexico, USA

The stochastic block model (SBM) is a useful model of structured random networks [1]: nodes are partitioned in blocks and edges are created independently between pairs of nodes with a probability that only depends on the block memberships of each node. Consequently, within a block, the network is indistinguishable from an Erdős-Rényi (ER) graph, but, between blocks, the connection probability can be chosen arbitrarily. The choice of different connection probabilities gives the model its versatility; for instance, one can generate networks with a community structure, a fuzzy multipartite structure or a core-periphery structure.

In the last few years, it has been shown that the model exhibits a *detectability limit* (DL) in the *infinite* size ensemble, for both the sparse and the dense regimes [2,3]. In essence, it has been shown that the block memberships cannot be recovered if the connection probabilities are too close to each other. However, since networks are *finite* objects, our understanding of the DL is not complete.

With an information theoretic approach, we derive first, for the infinite case, a set of constraints on the functional form of the DL. Our method is based on the natural symmetries of the SBM, and uses transformations of the generated networks to show that the DL must have certain properties. We then turn our attention to the finite size case, and argue that the Kullback-Leibler divergence (KL) of the SBM ensemble and its *random equivalent* (an ER graph of equivalent density) is a natural proxy for the difficulty of the detection problem. We finally show that a critical curve of constant KL correctly predicts the position of the DL and furthermore that it possesses all the desirable properties derived in the infinite limit.

[1] Holland, P. W., Laskey, K. B. and Leinhardt, S., *Soc. Networks*, **5** (1983), p. 109–137.
[2] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborov, *Phys. Rev. Lett.*, **107** (2011), p. 065701.
[3] Peixoto, T. P., *Phys. Rev. Lett.*, **111** (2013), p 098701

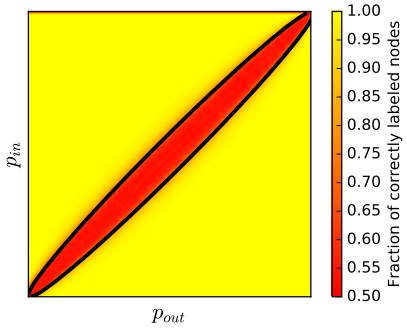

Figure 1: Spectral detection on networks generated by the SBM with 2 blocks of size $n = 180$, an internal connection probability p_{in} for both blocks and an external connection probability p_{out}. In the undetectable region (red), the inferred block memberships are as good as a random assignment of the nodes to blocks (accuracy of 50%, on average). The black curve is a line of constant KL.