Statistical inference with correlated disorder in community detection

ndrea Vincenzo Dell' bate and Louise Budzynski

DI ENS, École Normale Supérieure, PSL, CNRS, INRI

Most theoretical work on Bayesian Inference model inputs as i.i.d. component-wise draws from some probability distribution. Despite providing valuable insights, these approaches are blind to the structure of real-world data. Determining the effects of the structure in the signal on the inference process is a priori an open question and will depend on the particular problem to be considered. While it might make inference easier, as the knowledge of some structure in the signal can help to restrict the search to a smaller space, it might as well make the problem harder from an algorithmic viewpoint, as observed in [1, 2] when an ergodicity breaking phenomenon prevents numerical methods (such as message-passing approaches) to provide a reasonable approximation of the posterior probability distribution of the signal given the data. This is particularly surprising, as it was conjectured that Bayes optimality implied the absence of glass phase at equilibrium [3].

In order to gain theoretical insights, we focus on the Censored Block Model model, also known as planted spin glass, with a structured signal. In particular, we consider the signal, or planted configuration, to be sampled from the Boltzmann distribution of an homogeneous Ising model with inverse temperature κ. This model is minimal, analytically tractable and its theoretical and computational limits are well understood in the case of unstructured ($\kappa = 0$) signal [3]. We study the information-theoretic and algorithmic properties, focusing on the Bayes optimal setting.

We provide a quantitative analysis of the feasibility of the inference task in the form of a phase diagram depending on κ quantifying the structure of the signal and the signal-to-noise ratio . The analysis is comprised of both a theoretical and a numerical part: in the former we studied the properties of the posterior distribution in typical realizations of the inference problem on random regular graphs, focusing on the thermodynamic limit in order to use the cavity method, while in the latter we performed numerical resolutions of the cavity equations and compared with the results obtained by running Belief Propagation on finite size instances of the problem.

The planted configuration undergoes a ferromagnetic phase transition at $\kappa \; = \mathrm{atanh}\ 1/c \quad 1)$ and a replica symmetry breaking phase transition at $\kappa_{\mathrm{KS}} = \mathrm{atanh}\ 1/\sqrt{c \quad 1})$. The phase diagram can therefore be divided into 3 regions, one for each phase of the prior. When the prior is replica symmetric the phase diagram is divided into a region in which inference is "easy" (information-theoretically possible and with optimal performance achieved with Belief Propagation) and a region in which it is "impossible" (information-theoretically impossible).

In the region where the prior is paramagnetic the critical value dividing the easy and impossible phases is lower than the one obtained with an unstructured prior $\kappa = 0)$: the inference process is made easier as the introduction of the structure lowers the critical value .

In the region in which the prior is ferromagnetic the critical value is $\kappa) = \kappa$, therefore the introduction of structure hinders the inference process for large enough values of κ.

Finally, in the region where the prior is in a replica symmetry breaking phase the posterior undergoes a RSB phase transition for small enough values of .

[1] . Braunstein, L. Budzynski, M. Mariani, Physical Review E, **108**(6), (2023).
[2] . Braunstein, L. Budzynski, M. Mariani, F. Ricci-Tersenghi. Physical Review E, **111**(6), (2025).
[3] L. Zdeborová and F. Krzakala. dvances in Physics, **65**(5), 453–552 (2016)