Solutions for Week 9

Bruno Loureiro and Luca Saglietti

29.04.2021

1 Exercise 9.1

Show that the Bethe free entropy for the stochastic block model can be written using

$$Z^{ij} = \sum_{a<b} c_{ab}(\psi_a^{i\to j}\psi_b^{j\to i} + \psi_b^{i\to j}\psi_a^{j\to i}) + \sum_a c_{aa}\psi_a^{i\to j}\psi_a^{j\to i} \quad \text{for} \quad (i,j)\in E \tag{1.1}$$

$$Z^i = \sum_{t_i} n_{t_i} e^{-h_{t_i}} \prod_{j\in\partial i}\sum_{t_j} c_{t_j t_i}\psi_{t_j}^{k\to i} \tag{1.2}$$

as

$$\Phi_{\mathrm{BP}}(q,\{n_a\},\{c_{ab}\}) = \frac{1}{N}\sum_i \log Z^i - \frac{1}{N}\sum_{(i,j)\in E}\log Z^{ij} + \frac{c}{2} \tag{1.3}$$

where c is the average degree given by (9.2).

Solution: In order to obtain the sought expression for the BP free entropy we will work on the actual factor graph of the model, where each pair of nodes i and j will be connected to a pairwise interaction factor (ij). On the edges, we will denote as $\psi_a^{i\to(ij)}$ the message sent from i to (ij), representing the cavity marginal corresponding to "color" a, and as $\chi_a^{(ij)\to i}$ the message sent from (ij) to i, representing the probability of "color" a being the correct assignment for variable i given the constraint (ij). Remember that each variable also receives an external field of intensity $\log n_a$, representing the prior on the size of the communities. Finally, ψ_a^i will denote the marginal of variable i.

We are going to repeatedly use the BP equation:

$$\chi_a^{(ij)\to i} = \frac{1}{Z_{(ij)\to i}}\sum_b \psi_b^{j\to(ij)}\left[(1-A_{ij})\left(1-\frac{c_{ab}}{N}\right) + A_{ij}c_{ab}\right], \tag{1.4}$$

where $Z_{(ij)\to i}$ is simply obtained by summing over a the numerator.

The general expression for the Bethe free-entropy reads:

$$\Phi_{\mathrm{BP}} = \frac{1}{N}\sum_i \log Z_i + \frac{1}{N}\sum_{(ij)}\log Z_{(ij)} - \frac{1}{N}\sum_{i(ij)}\log Z_{i(ij)} \tag{1.5}$$

and thus we can proceed by evaluating all the terms separately and then putting all the pieces together (as usual we expect some cancellations). Note that the notation is a bit overloaded, but the difference is deliberate: $Z_i, Z_{(ij)}$ in the previous expression are actually different from Z^i, Z^{ij} in equation 1.1.

We start with the variable contribution:

$$\begin{aligned}
Z_i &= \sum_a n_a \prod_{(ij)} \chi_a^{(ij)\to i} && (1.6)\\
&\overset{(*)}{=} \sum_a n_a \prod_{(ij)} \frac{1}{Z_{(ij)\to i}} \sum_b \psi_b^{j\to(ij)} \left[(1-A_{ij})\left(1-\frac{c_{ab}}{N}\right)+A_{ij}c_{ab}\right] && (1.7)\\
&\overset{(**)}{=} \frac{1}{\prod_{(ij)} Z_{(ij)\to i}} \sum_a n_a \prod_{(ij)\in E}\left[\sum_b \psi_b^{j\to(ij)} c_{ab}\right] \prod_{(ij)\notin E}\left[\sum_b \psi_b^{j\to(ij)}\left(1-\frac{c_{ab}}{N}\right)\right], && (1.8)
\end{aligned}$$

using the BP equation for (*), and splitting over edge/non-edge contributions in (**).

Now, in the second term, we can use the normalization $\sum_b \psi_b^{j\to(ij)} = 1$ and write:

$$\prod_{(ij)\notin E}\left[1-\sum_b \psi_b^{j\to(ij)}\frac{c_{ab}}{N}\right] \overset{(*)}{\approx} e^{-\sum_{(ij)\notin E}\sum_b \psi_b^{j\to(ij)}\frac{c_{ab}}{N}} \overset{(**)}{\approx} e^{-h_a} \tag{1.9}$$

where, as in the lecture, we define $h_a = \sum_j \sum_b \psi_b^j \frac{c_{ab}}{N}$, and where in (*) we used the fact the we consider the high dimensional limit $N\to\infty$, and in (**) we approximated the sum (up to $\mathcal{O}(1/N)$) by replacing the sum over the non edges with the sum over all the links, and the cavity marginals with the marginals. Therefore we have:

$$Z_i = \frac{\sum_a n_a e^{-h_a}\prod_{(ij)\in E}\left[\sum_b \psi_b^{j\to(ij)} c_{ab}\right]}{\prod_{(ij)} Z_{(ij)\to i}} = \frac{Z^i}{\prod_{(ij)} Z_{(ij)\to i}} \tag{1.10}$$

Now we can look at the factor contribution:

$$\begin{aligned}
Z_{(ij)} &= \sum_{a,b} \psi_a^{i\to(ij)}\psi_b^{j\to(ij)}\left[(1-A_{ij})\left(1-\frac{c_{ab}}{N}\right)+A_{ij}c_{ab}\right] && (1.11)\\
&\overset{(*)}{=} \mathbb{I}((ij)\in E)\sum_{a,b}\psi_a^{i\to(ij)}\psi_b^{j\to(ij)}c_{ab} + \mathbb{I}((ij)\notin E)\sum_{a,b}\psi_a^{i\to(ij)}\psi_b^{j\to(ij)}\left(1-\frac{c_{ab}}{N}\right) && (1.12)\\
&\overset{(**)}{=} \mathbb{I}((ij)\in E)\sum_{a,b}\psi_a^{i\to(ij)}\psi_b^{j\to(ij)}c_{ab} + \mathbb{I}((ij)\notin E)\, e^{-\sum_{a,b}\psi_a^{i\to(ij)}\psi_b^{j\to(ij)}\frac{c_{ab}}{N}}, && (1.13)
\end{aligned}$$

where we have again split over edge/non-edge contributions in (*) and exploited the large N limit in (**).

Finally, the edge contribution yields:

$$\begin{aligned}
Z_{i(ij)} &= \sum_a \psi_a^{i\to(ij)}\chi_a^{(ij)\to i} && (1.14)\\
&\overset{(*)}{=} \sum_a \psi_a^{i\to(ij)}\frac{1}{Z_{(ij)\to i}}\sum_b \psi_b^{j\to(ij)}\left[(1-A_{ij})\left(1-\frac{c_{ab}}{N}\right)+A_{ij}c_{ab}\right] && (1.15)\\
&= \frac{Z_{(ij)}}{Z_{(ij)\to i}}, && (1.16)
\end{aligned}$$

Putting everything together we get:

$$\begin{aligned}
\Phi_{\rm BP} &= \frac{1}{N}\sum_i(\log Z^i - \sum_{(ij)}\log Z_{(ij)\to i}) + \frac{1}{N}\sum_{(ij)}\log Z_{(ij)} - \frac{1}{N}\sum_{i(ij)}(\log Z_{(ij)} - \log Z_{(ij)\to i}) && (1.17)\\
&= \frac{1}{N}\sum_i \log Z^i - \frac{1}{N}\sum_{(ij)}\log Z_{(ij)} && (1.18)\\
&= \frac{1}{N}\sum_i \log Z^i - \frac{1}{N}\sum_{(ij)\in E}\log\sum_{a,b}\psi_a^{i\to(ij)}\psi_b^{j\to(ij)}c_{ab} + \frac{1}{N}\sum_{(ij)\notin E}\sum_{a,b}\psi_a^{i\to(ij)}\psi_b^{j\to(ij)}\frac{c_{ab}}{N} && (1.19)\\
&\overset{(*)}{\approx} \frac{1}{N}\sum_i \log Z^i - \frac{1}{N}\sum_{(ij)\in E}\log Z^{ij} + \frac{1}{N}\sum_{(ij)}\sum_{a,b}\psi_a^i\psi_b^j\frac{c_{ab}}{N} && (1.20)
\end{aligned}$$

where in (*) we used the fact that the sum over colors can be split in $\sum_{a,b} f(a,b) = \sum_{a<b}(f(a,b) + f(b,a)) + \sum_a f(a,a)$ as in expression 1.1 in the assignment, and approximated the sum over non-edges with the sum over all links.

Finally, if we look at the last term in equation 1.20, we can recognize:

$$\sum_{(ij)}\sum_{a,b}\psi_a^i\psi_b^j\frac{c_{ab}}{N} = \langle|E|\rangle_{posterior} \overset{(*)}{=} N\frac{c}{2} \tag{1.21}$$

where in (*) the Nishimori condition guarantees the equivalence between the posterior average and the true generative model.

2 Exercise 9.2

Show that in the stochastic block model maximization of the Bethe free entropy with respect to the parameters n_a and c_{ab} at a BP fixed point leads to the following conditions for stationarity that can be then used for iterative learning of the parameters n_a and c_{ab}.

$$\begin{aligned}
n_a &= \frac{1}{N}\sum_i \psi_a^i && (2.1)\\
c_{ab} &= \frac{1}{N}\frac{1}{n_b n_a}\sum_{(i,j)\in E}\frac{c_{ab}(\psi_a^{i\to j}\psi_b^{j\to i} + \psi_b^{i\to j}\psi_a^{j\to i})}{Z^{ij}}\,. && (2.2)
\end{aligned}$$

Solution: Let's start by looking at the derivative w.r.t. n_a. Since we have to guarantee normalization of probabilities, we use the method of the multipliers and introduce the Lagrangian:

$$\mathcal{L} = \Phi_{\rm BP} + \lambda(1 - \sum_a n_a) \tag{2.3}$$

and set to zero the derivative:

$$\partial_{n_c}\mathcal{L} = \partial_{n_c}\Phi_{\rm BP} - \lambda = 0 \tag{2.4}$$

Remember that the Bethe free entropy is stationary with respect to the BP messages, so we only need to take explicit derivatives of Φ_{BP}. The only term where n_b appears is the first term of eq. 1.3:

$$\partial_{n_c}\Phi_{BP} = \frac{1}{N}\sum_i \partial_{n_c}\log Z_i = \frac{1}{N}\sum_i \frac{e^{-h_c}\prod_{j\in\partial i}\sum_b \psi_b^{j\to i}c_{cb}}{Z_i} = \frac{1}{N}\sum_i\frac{\psi_c^i}{n_c}. \tag{2.5}$$

Therefore we have:

$$\partial_{n_c}\mathcal{L} = 0 \quad \longrightarrow \quad \frac{1}{N}\sum_i \psi_c^i = \lambda n_c \tag{2.6}$$

Moreover we also require:

$$\partial_\lambda \mathcal{L} = 1 - \sum_a n_a = 0 \tag{2.7}$$

so, if we sum equation 2.6 over the colors we get:

$$\frac{1}{N}\sum_i\sum_c \psi_c^i \overset{(*)}{=} 1 = \lambda \sum_c n_c \overset{(**)}{=} \lambda, \tag{2.8}$$

where in (*) we used the normalization of the messages, and in (**) we used equation 2.7. So, substituting $\lambda = 1$ in eq. 2.6 we finally have:

$$n_a = \frac{1}{N}\sum_i \psi_a^i, \tag{2.9}$$

as one would have expected from the Nishimori condition.

Getting the equation for c_{ab} is a bit more involved. We will split the calculation in the three derivatives of the terms appearing in eq. 1.3. First we evaluate:

$$\begin{aligned}
\partial_{c_{cd}}\frac{1}{N}\sum_i \log Z^i &= \frac{\sum_i}{N}\frac{\partial_{c_{cd}}(\sum_a n_a e^{-h_a}\prod_{j\in\partial i}\sum_b c_{ab}\psi_b^{j\to i})}{Z^i} && (2.10)\\
&= \frac{\sum_i}{N}\frac{-(\partial_{c_{cd}}h_c)n_c e^{-h_c}\prod_{j\in\partial i}\sum_b c_{cb}\psi_b^{k\to i} + (c\leftrightarrow d)}{Z^i}\\
&+ \frac{\sum_i}{N}\frac{n_c e^{-h_c}\partial_{c_{cd}}(\prod_{j\in\partial i/k}\sum_b c_{cb}\psi_b^{j\to i})) + (c\leftrightarrow d)}{Z^i} && (2.11)\\
&= \frac{\sum_i}{N}\frac{-(\frac{1}{N}\sum_k \psi_d^k)\psi_c^i Z^i + (c\leftrightarrow d)}{Z^i}\\
&+ \frac{\sum_i}{N}\frac{n_c e^{-h_c}\sum_k \psi_d^{k\to i}(\prod_{j\in\partial i/k}\sum_b c_{cb}\psi_b^{j\to i})) + (c\leftrightarrow d)}{Z^i} && (2.12)\\
&\overset{(*)}{=} -2n_c n_d + \frac{1}{N}\sum_{i,k}\frac{\psi_d^{k\to i}\psi_c^{i\to k} + (c\leftrightarrow d)}{Z^{ik}} && (2.13)
\end{aligned}$$

where in (*) we used the fact that one can rewrite Z^i as:

$$\begin{aligned}
Z^i &= \sum_a n_a e^{-h_a}\prod_{j\in\partial i}\sum_b \psi_b^{j\to i}c_{ab} && (2.14)\\
&= \sum_{ab} c_{ab}\psi_b^{j\to i} n_a e^{-h_a}\prod_{k\in\partial i/j}\sum_c \psi_c^{j\to i}c_{ac} && (2.15)\\
&= \sum_{ab} c_{ab}\psi_b^{j\to i}\psi_a^{i\to j}Z^{i\to j} && (2.16)\\
&= Z^{ij}Z^{i\to j} && (2.17)
\end{aligned}$$

Then we have:

$$\begin{aligned}
\partial_{c_{cd}}\frac{1}{N}\sum_{(ij)\in E}\log Z^{ij} &= \frac{\sum_{(ij)\in E}}{N}\frac{\partial_{c_{cd}}(\sum_{ab}\psi_a^{i\to j}\psi_b^{j\to i}c_{ab})}{Z^{ij}} && (2.18)\\
&= \frac{1}{N}\sum_{(ij)\in E}\frac{\psi_c^{i\to j}\psi_d^{j\to i} + (c\leftrightarrow d)}{Z^{ij}} && (2.19)
\end{aligned}$$

and finally:

$$\partial_{c_{cd}} \left(\frac{c}{2} \right) = \frac{1}{2} \partial_{c_{cd}} \left(\sum_{ab} n_a n_b c_{ab} \right) = n_c n_d \tag{2.20}$$

Putting everything together (notice that the right term in 2.13 double counts 2.15):

$$\partial_{c_{cd}} \Phi_{\mathrm{BP}} = \frac{1}{N} \sum_{(ij) \in E} \frac{\psi_c^{i \to j} \psi_d^{j \to i} + (c \leftrightarrow d)}{Z^{ij}} - n_c n_d = 0. \tag{2.21}$$

Now, if we multiply both terms by c_{cd} and send $(cd) \to (ab)$ we get the sought result:

$$c_{ab} = \frac{1}{N n_a n_b} \sum_{(ij) \in E} \frac{c_{ab} \psi_a^{i \to j} \psi_b^{j \to i} + (a \leftrightarrow b)}{Z^{ij}}. \tag{2.22}$$