Course project
"Optimization approaches to community detection"

Marina Danilova, Alexander Podkopaev, Nikita Puchkin, Igor Silin

December 15, 2016

1 Introduction to community detection

2 Stochastic block model

3 Algorithms

3.1 Semidefinite relaxations

3.2 Conjugate gradients method

It is assumed, that latent group labels Z has a multinomial distribution with parameter π and each element of A has a Bernoulli distribution with parameter θ_{ij}:

$$
\begin{aligned}
&Z_i \sim \text{Poly}(\pi), \quad \pi^T = (\pi_1, \dots, \pi_k),\\
&a_{ij} \sim \text{Bernoulli}(P_{Z(i)Z(j)}), \quad i, j = \overline{1, n},
\end{aligned}
$$

A Bayesian approach is used to estimate the most probable configuration of Z given an adjacency matrix A

$$
Z^* = \arg\max_Z p(Z|A) = \arg\max_Z \iint p(Z, \pi, P|A) d\pi dP \tag{3.2.1}
$$

It treats parameters π and θ_{ij} as random variables with following prior distributions

$$
\begin{aligned}
&\pi \sim \text{Dirichlet}(\alpha)\\
&P_{ii} \sim \text{Beta}(\beta), \quad i = \overline{1, k},
\end{aligned}
$$

where α and β are predefined hyperparameters and P_{ij}, $i \neq j$ are set to equal to a small constant ε. Futhermore, to tackle the intractable integration in 3.2.1 a restriction on the family of factorized distributions is considered

$$
\mathcal{Q} = \{q : \ q(Z, \pi, P) = q(\pi) q(P) \prod_i q(Z_i)\},
$$

where

$$
\begin{aligned}
&q(Z_i) : Z_i \sim \text{Poly}(\tilde{\pi})\\
&q(\pi) : \pi \sim \text{Dirichlet}(\tilde{\alpha})\\
&q(P) : P_{ii} \sim \text{Beta}(\tilde{\beta}_i)
\end{aligned}
$$

The optimal distribution $q^* \in \mathcal{Q}$, that approximates the true posterior $p(Z, \pi, P|A)$ minimizes the Kullback-Leibler divergence

$$q^* = \arg\min_{q\in\mathcal{Q}} \mathrm{KL}\left(q \| p(Z, \pi, P|X)\right) \tag{3.2.2}$$

Define

$$\mathcal{L}(q) = \sum_{Z} \iint q(Z, \pi, P) \log \frac{p(A, Z, \pi, P)}{q(Z, \pi, P)} d\pi dP \tag{3.2.3}$$

Since

$$\mathcal{L}(q) + \mathrm{KL}\left(q \| p(Z, \pi, P|X)\right) = \log p(A)$$

the minimization problem 3.2.2 can be equivalently solved by maximizing $\mathcal{L}(q)$

$$q^* = \arg\max_{q\in\mathcal{Q}} \mathcal{L}(q) \tag{3.2.4}$$

Denote $\tilde{\Pi} = \|\tilde{\pi}_{ij}\|$, $i = \overline{1,n}$, $j = \overline{1,k}$. Given $\tilde{\Pi}$ values of parameters, that maximize $\mathcal{L}(q)$ can be found according to formulas

$$\begin{aligned} &\tilde{\alpha} = \alpha + \tilde{\Pi}^T 1 \\ &\tilde{\beta}_i = \beta + \frac{1}{2}\left(\tilde{\Pi}_{\cdot i}^T A \tilde{\Pi}_{\cdot i}, \tilde{\Pi}_{\cdot i}^T \overline{A} \tilde{\Pi}_{\cdot i}\right)^T, \quad i = \overline{1,k}, \end{aligned} \tag{3.2.5}$$

where 1 denotes an all-ones vector, $\overline{A} = 11^T - I - A$, and $\tilde{\Pi}_{\cdot i}$ stands for the i-th column of the matrix $\tilde{\Pi}$. A corresponding value of log-likelihood is equal to

$$\mathcal{L}(\tilde{\Pi}) = \sum_{i<j} \tilde{\Pi}_{\cdot i}^T A \tilde{\Pi}_{\cdot j} \log \varepsilon + \tilde{\Pi}_{\cdot i}^T \overline{A} \tilde{\Pi}_{\cdot j} \log(1-\varepsilon) - \sum_{i,j} \tilde{\pi}_{ij} \log \tilde{\pi}_{ij} + \log \frac{\mathcal{B}(\tilde{\alpha})}{\mathcal{B}(\alpha)} + \sum_{i} \log \frac{\mathcal{B}(\tilde{\beta})}{\mathcal{B}(\beta)}, \tag{3.2.6}$$

where $\tilde{\alpha} = \tilde{\alpha}(\tilde{\Pi})$ and $\tilde{\beta} = \tilde{\beta}(\tilde{\Pi})$ can be found from 3.2.5, $\mathcal{B}(\alpha) \triangleq \frac{\Gamma\left(\sum_i \alpha_i\right)}{\prod_i \Gamma(\alpha_i)}$ and Γ is gamma-function. Now the maximization problem 3.2.4 can be reformulated as follows

$$\begin{cases} \mathcal{L}(\tilde{\Pi}) \longrightarrow \max \\ \sum_j \tilde{\pi}_{ij} = 1, \quad i = \overline{1,n} \end{cases} \tag{3.2.7}$$

One can use reparametrization

$$\begin{aligned} &\tilde{\pi}_{ij} = e^{\theta_{ij} - \mathcal{A}_i}, \quad i = \overline{1,n}, j = \overline{1,k-1}, \\ &\tilde{\pi}_{ik} = e^{-\mathcal{A}_i}, \quad i = \overline{1,n}, \\ &\mathcal{A}_i = \log(1 + \sum_{j=1}^{k-1}) e^{\theta_{ij}}, \qquad i = \overline{1,n} \end{aligned} \tag{3.2.8}$$

and obtain a problem of unconstrained maximization

$$\mathcal{L}(\theta) \longrightarrow \max \tag{3.2.9}$$

The problem 3.2.9 can be solved via natural conjugate gradient method. Namely, given an initial value $\theta^{(0)}$, one iteratively finds optimal value of θ as follows

$$\theta^{(t+1)} = \theta^{(t)} + \lambda^{(t)} d^{(t)}, \quad t = 0, 1, 2, \ldots$$

Here $d^{(t)}$ is so called natural conjugate gradient. $d^{(t)}$ can be found according to formulas

$$d^{(t)} = \begin{cases} g^{(t)}, & t = 0 \\ g^{(t)} + \frac{\|g^{(t)}\|_\theta^2}{\|g^{(t-1)}\|_\theta^2} d^{(t-1)}, & t > 0, \end{cases} \tag{3.2.10}$$

where $g^{(t)}$ is a natural gradient of $\mathcal{L}(\theta)$. $\|\cdot\|$ stands for the norm with respect to Riemannian metrics

$$G(\theta) = \operatorname{diag}\left(\mathcal{I}(\theta_1), \ldots, \mathcal{I}(\theta_N)\right),$$

where $\theta_i = (\theta_{i1}, \ldots, \theta_{ik})^T$, $i = \overline{1, n}$ and (θ_i) is a Fischer information. This method is nothing else but a conjugate gradient method in a Riemannian space with metrics $G(\theta)$. Given θ, g and $\|g\|_\theta$ can be found as follows

$$\begin{aligned} g = \nabla_{\tilde{\Pi}} \mathcal{L}(\tilde{\Pi}) &= \begin{pmatrix} (I, -1)\nabla_{\tilde{\pi}_1} \mathcal{L}(\tilde{\Pi}) \\ \vdots \\ (I, -1)\nabla_{\tilde{\pi}_n} \mathcal{L}(\tilde{\Pi}) \end{pmatrix} \\ \|g\|_\theta^2 &= \sum_{i=1}^{n} \left(\nabla_{\tilde{\pi}_1} \mathcal{L}(\tilde{\Pi})\right)^T \left(\operatorname{diag}(\tilde{\pi}_i) - \tilde{\pi}_i^T \tilde{\pi}_i\right) \left(\nabla_{\tilde{\pi}_1} \mathcal{L}(\tilde{\Pi})\right), \end{aligned} \tag{3.2.11}$$

where $\tilde{\Pi} = \tilde{\Pi}(\theta)$ and

$$\frac{\partial \mathcal{L}(\tilde{\Pi})}{\partial \pi_{ij}} = \sum_{l \neq j} \left(A_{i\cdot} \tilde{\Pi}_{\cdot l} \log \varepsilon + \overline{A}_{i\cdot} \tilde{\Pi}_{\cdot j} \log(1 - \varepsilon) \right) + \left(A_{i\cdot} \tilde{\Pi}_{\cdot j}, \overline{A}_{i\cdot} \tilde{\Pi}_{\cdot j} \right) \nabla \log \mathcal{B}(\tilde{\beta}_j) + \psi(\tilde{\alpha}_j) - \log \tilde{\pi}_{ij} - 1, \tag{3.2.12}$$

where $\psi(\cdot)$ is digamma function.

The final algorithm is given in 1.

Algorithm 1: Natural Conjugate Gradient

Input: adjacency matrix A, maximum number of clusters k, tolerance η, maximum number of iterations $t_{\max}$
Output: array of predicted labels Z

Initialize θ;
$\mathcal{L}_{old} \leftarrow -\infty$;
$\lambda \leftarrow 1$;
for t *in range* ($t_{\max}$) **do**
  Calculate $\tilde{\Pi}$ using 3.2.8;
  Calculate $\tilde{\alpha}$, $\tilde{\beta}$ using 3.2.5;
  Calculate $\mathcal{L}$ using 3.2.6;
  if $0 \leq \frac{\mathcal{L}-\mathcal{L}_{old}}{|\mathcal{L}|} < \eta$ **then**
    break
  if $\eta \geq 0$ **then**
    Update d using 3.2.10, 3.2.11, 3.2.12;
    $\theta_{old} \leftarrow \theta$
    $\theta \leftarrow \theta_{old} + \lambda d$;
    $\mathcal{L}_{old} \leftarrow \mathcal{L}$;
  else
    $\lambda \leftarrow \frac{\lambda}{2}$
    $\theta \leftarrow \theta_{old} + \lambda|\eta|d$
  end
end
$Z = \left(\arg\max_{1\leq j\leq k} \tilde{\pi}_{1j}, \ldots, \arg\max_{1\leq j\leq k} \tilde{\pi}_{nj}\right)$
return Z

Initial value of θ was generated from a standart normal distribution $\mathcal{N}(0,1)$. Probability of occurence an inter-cluster edge ε was set to 10^{-10}. The maximal number of iterations and relative tolerance were taken equal to 100 and 10^{-6} respectively. Examples of performance of the natural conjugate gradient method can be found, for example, in [1].

3.3 Modularity-based methods

3.4 Spectral method

4 Data

5 Experimental results

6 Work split

7 References

References

[1] A fast inference algorithm for stochastic blockmodel, Zhiqiang Xu, Yiping Ke, Yi Wang, 2014.