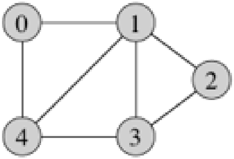
Community Identification in Weighted Networks

Daniel Kessler

April 22nd, 2019

Outline

Networks and Adjacency Matrices



	0	1	2	3	4
0	0	1	0	0	1
1	1	0	1	1	1
2	0	1	0	1	0
3	0	1	1	0	1
4	1	1	0	1	0

Figure: Undirected Binary Graph and Adjacency Matrix from Stack Exchange

Notation

- Observe a graph $G = (V, E)$, $n = |V|$
- Adjacency matrix $A \in \{0, 1\}^{n \times n}$
- $A_{i,j} = 1 \iff (i, j) \in E$

Stochatic Block Model

Classical SBM

- Proposed in (Holland, Laskey, and Leinhardt 1983)
- Suppose each $v \in V$ can be assigned to one of K communities
- Let $\boldsymbol{z} \in \{1, 2, \ldots, K\}^n$ give community assignments
- $z_i \sim \text{Categorical}(p_1, p_2, \ldots, p_K)$
- $A_{i,j} \mid \boldsymbol{z} \overset{ind}{\sim} \text{Bernoulli}(P_{z(i),z(j)})$
- $P \in [0,1]^{K \times K}$ parameterizes the blocks

Weighted SBM

- All edges present: $E = \{(i,j) : i \neq j, \in [n]\})$
- Entries of A take values in $\mathcal{S} \subseteq \mathbb{R}$
- $A_{i,j} | \boldsymbol{z} \overset{ind}{\sim} F(\theta_{z(i),z(j)})$
- F is some law parameterized by θ

Bayesian Model

- Proposed in (Aicher, Jacobs, and Clauset 2013; Aicher, Jacobs, and Clauset 2015)
- Suppose that all edge distributions come from common exponential family
- Use conjugate priors: $\pi_{\tau_r}(\theta_r) = \frac{1}{Z(\tau_r)} \exp(\tau_r \cdot \eta(\theta_r))$
- Prior (flat) for $\boldsymbol{z}$: $\pi_i(z_i) = \text{Categorical}\left(\frac{1}{K}, \ldots, \frac{1}{K}\right)$
- Suppose $\boldsymbol{z} \perp \boldsymbol{\theta}$ (based on DAG from paper)
- $\pi(z, \theta \mid \boldsymbol{\tau}) = \prod_i^n \frac{1}{K} \prod_r^R \frac{1}{Z(\tau_r)} \exp(\tau_r \cdot \eta(\theta_r))$
- Let $\pi^\star(\boldsymbol{z}, \theta)$ be the posterior

$$\pi^\star \propto P(A \mid \boldsymbol{z}, \theta)\pi(\boldsymbol{z}, \theta)$$

Formal Model

- $\pi(\theta_r \mid \tau_r) = \frac{1}{Z(\tau_r)} \exp(\tau_r \cdot \eta(\theta_r))$
- $\pi(z_i \mid \mu_i) = \mu_i(z_i) : \sum_{k=1}^{K} \mu_i(k) = 1$
- In practice, $\mu_i = \frac{1}{K} \implies \pi(z_i \mid \mu_i) = \frac{1}{K}$
- $\pi(\boldsymbol{z}, \theta) = \pi(\boldsymbol{z})\pi(\theta) = \prod_{i<j} \frac{1}{K} \prod_r \frac{1}{Z(\tau_r)} \exp(\tau_r \cdot \eta(\theta_r))$
- $P(A \mid \boldsymbol{z}, \theta) = \left[\prod_{i<j} h(A_{i,j})\right] \exp(\sum_{i<j} T(A_{i,j})\eta(\theta_{z_i,z_j}))$
- Posterior $\pi^*(\boldsymbol{z}, \theta) = \frac{P(A|\boldsymbol{z},\theta)\pi(z)\pi(\theta)}{\int_\Theta \sum_{\boldsymbol{z}\in\mathcal{Z}} P(A|\boldsymbol{z},\theta)\pi(\boldsymbol{z})\pi(\theta)d\theta}$
- Approximation q to π^*
  $q(\boldsymbol{z}, \theta \mid \boldsymbol{\mu}^*, \boldsymbol{\tau}^*) = \prod_i \mu_i^*(z_i) \times \prod_r \frac{1}{Z(\tau_r^*)} \exp(\tau_r^* \cdot \eta(\theta_r))$

Variational Estimation

This approach proposed in (Aicher, Jacobs, and Clauset 2013; Aicher, Jacobs, and Clauset 2015) (and subsequent derivations adapted therefrom)

- $\pi^\star(\boldsymbol{z}, \theta) \propto \Pr(A \mid \boldsymbol{z}, \theta)\pi(\boldsymbol{z}, \theta)$
- Approximate $\pi^\star(\boldsymbol{z}, \theta)$ by factorizable $q(\boldsymbol{z}, \theta) = q_{\boldsymbol{z}}(\boldsymbol{z})q_\theta(\theta)$
- Choose q that minimizes KL-divergence with posterior

$$D_{KL}(q \parallel \pi^\star) = -\int q \log \frac{\pi^\star}{q}$$

- Doing this directly is hard, but there's another way...

Variational Inference Trick

Recall that the marginal log likelihood of the data is a fixed quantity (since A) has been observed

$$
\begin{aligned}
C &= \log P(A) \\
&= \int_\Theta \sum_{\mathbf{z}\in\mathcal{Z}} q(\mathbf{z},\theta)d\theta \log P(A) && \text{(Multiply by 1)} \\
&= \int_\Theta \sum_{\mathbf{z}\in\mathcal{Z}} q(\mathbf{z},\theta) \log \frac{P(A,\mathbf{z},\theta)}{P(\mathbf{z},\theta \mid A)} d\theta && \text{(Conditioning tricks)} \\
&= \underbrace{\int_\Theta \sum_{\mathbf{z}\in\mathcal{Z}} q(\mathbf{z},\theta) \log \frac{P(A,\mathbf{z},\theta)}{q(\mathbf{z},\theta)} d\theta}_{\mathcal{G}(q)} - \underbrace{\int_\Theta \sum_{\mathbf{z}\in\mathcal{Z}} q(\mathbf{z},\theta) \frac{P(\mathbf{z},\theta \mid A)}{q(\mathbf{z},\theta)} d\theta}_{D_{\mathrm{KL}}(q\|\pi^*)}
\end{aligned}
$$

$D_{\mathrm{KL}}(q \parallel \pi^*) = \underbrace{\log P(A)}_{\text{Constant}} - \mathcal{G}(q)$, so maximizing $\mathcal{G}$ minimizes KL

Maximizing $\mathcal{G}(q)$

$$
\begin{aligned}
\mathcal{G}(q) &= \int_\Theta \sum_{\boldsymbol{z}\in\mathcal{Z}} q(\boldsymbol{z},\theta) \log \frac{P(A,\boldsymbol{z},\theta)}{q(\boldsymbol{z},\theta)} d\theta \\
&= \int_\Theta \sum_{\boldsymbol{z}\in\mathcal{Z}} q(\boldsymbol{z},\theta) \log \frac{P(A \mid \boldsymbol{z},\theta)\pi(\boldsymbol{z},\theta)}{q(\boldsymbol{z},\theta)} d\theta \\
&= \int_\Theta \sum_{\boldsymbol{z}\in\mathcal{Z}} q(\boldsymbol{z},\theta) \log P(A \mid \boldsymbol{z},\theta) d\theta + \int_\Theta \sum_{\boldsymbol{z}\in\mathcal{Z}} q(\boldsymbol{z},\theta) \log \frac{\pi(\boldsymbol{z},\theta)}{q(\boldsymbol{z},\theta)} d\theta \\
&= \mathbb{E}_q \left[\log P(A \mid \boldsymbol{z},\theta)\right] + \underbrace{\mathbb{E}_q \left[\log \frac{\pi(\boldsymbol{z},\theta)}{q(\boldsymbol{z},\theta)}\right]}_{-D_{\mathrm{KL}}(q\|\pi)}
\end{aligned}
$$

We want to choose a q that maximizes our log likelihood, but without straying too far from the prior π

Algorithm and Updates

- We have that $G \propto \sum_r^R E_q(T_r + \tau_r - \tau_r^*) E_q(\eta(\theta_r)) + \sum_r^R \log \frac{Z(\tau_r)}{Z(\tau_r^*)} + \sum_i + \sum_{z_i \in [K]} \mu_i^* \log \frac{\mu_i(z_i)}{\mu_i^*(z_i)}$
- $E_q T_r = \sum_{i<j} \sum_{(z_i,z_j)=r} \mu_i^*(z_i) \mu_j^*(z_j) T(A_{i,j})$
- $E_q \eta(\theta_r) = \frac{\partial}{\partial \tau} \log Z(\tau) \mid_{\tau=\tau_r^*}$
- Taking gradients yields update rules
- $\tau_r^* = \tau_r + E_q T_r$: makes sense given conjugacy
- $\mu_i^*(z) \propto \exp\left(\sum_r \frac{\partial E_q T_r}{\partial \mu_i(z)} \cdot E_q \eta(\theta_r)\right)$
- Normalize the μ_i^* to sum to 1

Results and Next Steps

- (Aicher, Jacobs, and Clauset 2015) Provides a MATLAB-based implementation
- Experiments with small graphs show good recovery of true labels

Work in Progress (See final report)

- (Naïve) Sampling-Based Approach to find MAP
- Compare sampling-approach with recent github package by Jean-Gabriel Young (Postdoc at UMich)
- More comprehensive experiments
- Comparison of Variational Method with Sampling-Based Methods (in terms of runtime/efficiency as well as accuracy)

References I

- Christopher Aicher, Abigail Z. Jacobs, and Aaron Clauset. "Adapting the Stochastic Block Model to Edge-Weighted Networks". In: *arXiv:1305.5782 [physics, stat]* (May 2013). arXiv: 1305.5782.
- Christopher Aicher, Abigail Z. Jacobs, and Aaron Clauset. "Learning Latent Block Structure in Weighted Networks". In: *Journal of Complex Networks* 3.2 (June 2015). arXiv: 1404.0431, pp. 221–248. ISSN: 2051-1310, 2051-1329. DOI: 10.1093/comnet/cnu026.
- Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. "Stochastic blockmodels: First steps". In: *Social Networks* 5.2 (June 1, 1983), pp. 109–137. ISSN: 0378-8733. DOI: 10.1016/0378-8733(83)90021-7.