Reconstruction in the Sparse Labeled Stochastic Block Model

Jiaming Xu [1] Marc Lelarge [2] Laurent Massoulié [3]

[1] University of Illinois, Urbana-Champaign

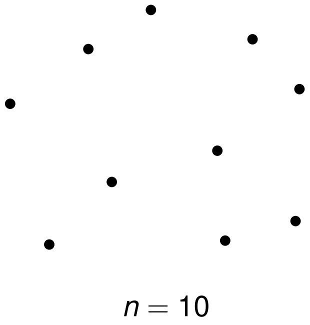
[2] INRIA-ENS

[3] INRIA-Microsoft Research Joint Centre

February 25, 2013

The sparse labeled stochastic block model

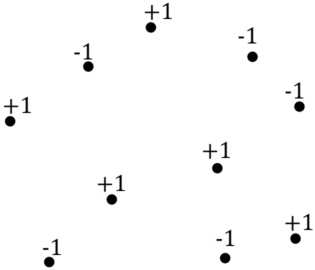
A random graph model on n nodes with two parameters, $a, b \geq 0$ and two discrete prob. distributions, μ, ν.



The sparse labeled stochastic block model

A random graph model on n nodes with two parameters, $a, b \geq 0$ and two discrete prob. distributions, μ, ν.

- Assign each vertex spin $+1$ or -1 uniformly at random.

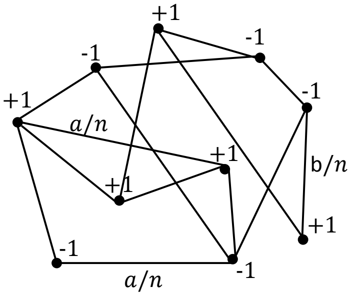


$n = 10$

The sparse labeled stochastic block model

A random graph model on n nodes with two parameters, $a, b \geq 0$ and two discrete prob. distributions, μ, ν.

- Independently for each pair (u, v):
  - if $\sigma_u = \sigma_v$, draw the edge w.p. a/n.
  - if $\sigma_u \neq \sigma_v$, draw the edge w.p. b/n.

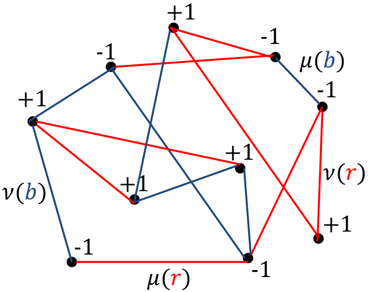


$a = 4, b = 2$

The sparse labeled stochastic block model

A random graph model on n nodes with two parameters, $a, b \geq 0$ and two discrete prob. distributions, μ, ν.

- Independently for each edge (u, v):
  - if $\sigma_u = \sigma_v$, label the edge w.d. μ.
  - if $\sigma_u \neq \sigma_v$, label the edge w.d. ν.



$\mu(r) = 0.6, \mu(b) = 0.4$
$\nu(r) = 0.4, \nu(b) = 0.6$

Reconstruction problem

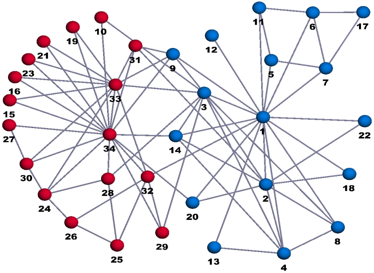
- Reconstruct the underlying spin configuration σ based on the observed labeled graph.
- Isolated nodes render exact reconstruction impossible. Focus on positively correlated reconstruction, i.e., $\hat{\sigma}$ agrees with σ in more than $1/2$ of all its entries.

Reconstruction problem

- Reconstruct the underlying spin configuration σ based on the observed labeled graph.
- Isolated nodes render exact reconstruction impossible. Focus on positively correlated reconstruction, i.e., $\hat{\sigma}$ agrees with σ in more than $1/2$ of all its entries.

Motivation

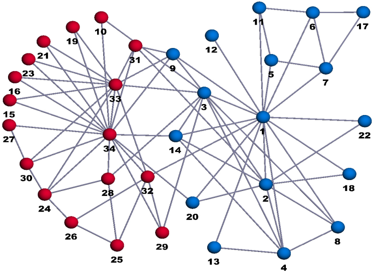
- Community detection in social or biological networks in the sparse regime with a bounded average degree.



- Labels to characterize various interaction types, e.g. strong and weak ties in friendship network. Edges across two different communities are more likely to be weak ties [OnnelaPNAS07].

Motivation

- Community detection in social or biological networks in the sparse regime with a bounded average degree.



- Labels to characterize various interaction types, e.g. strong and weak ties in friendship network. Edges across two different communities are more likely to be weak ties [OnnelaPNAS07].

Phase transition

Heimlicher-Lelarge-Massoulié '12:

Conjecture

If $\tau > 1$, then positively correlated reconstruction is possible.
If $\tau < 1$, then positively correlated reconstruction is impossible.

$$\tau = \frac{a+b}{2} \sum_{l \in \mathcal{L}} \frac{a\mu(l) + b\nu(l)}{a+b} \left(\frac{a\mu(l) - b\nu(l)}{a\mu(l) + b\nu(l)} \right)^2 .$$

- Generalize the conjectures in stochastic block model, [Decelle11],[Mossel12].
- τ comes from the local stability analysis of a fixed point of BP.

Phase transition

Heimlicher-Lelarge-Massoulié '12:

Conjecture

If $\tau > 1$, then positively correlated reconstruction is possible.
If $\tau < 1$, then positively correlated reconstruction is impossible.

$$\tau = \frac{a+b}{2} \sum_{l \in \mathcal{L}} \frac{a\mu(l) + b\nu(l)}{a+b} \left(\frac{a\mu(l) - b\nu(l)}{a\mu(l) + b\nu(l)} \right)^2 .$$

- Generalize the conjectures in stochastic block model, [Decelle11],[Mossel12].
- τ comes from the local stability analysis of a fixed point of BP.

Reconstruction method

Method	Regime	Complexity
BP	$\tau > 1$ [1]	depends on convergence time
MinBisection	$\tau > 64 \ln 2$	could be hard
Spectral	$\tau > C\sqrt{a+b}$	polynomial
Converse	$\tau < 1$	

[1] Unproven, validated via simulation [Heimlicher12]

Posterior distribution

- The posterior distribution resembles Ising model! Let $J_{uv} = \frac{1}{2}\log\frac{a\mu(L_{uv})}{b\nu(L_{uv})}$

$$\mathbb{P}(\sigma | A, L, \sum_u \sigma_u = 0) \approx \frac{1}{Z}\exp\left[\sum_{(u,v)\in E(G)} J_{uv}\sigma_u\sigma_v\right].$$

- The reconstruction threshold τ is equivalent to

$$\frac{a+b}{2}\mathbb{E}_J[\tanh^2(J)] = 1.$$

Posterior distribution

- The posterior distribution resembles Ising model! Let $J_{uv} = \frac{1}{2} \log \frac{a\mu(L_{uv})}{b\nu(L_{uv})}$

$$\mathbb{P}(\sigma | A, L, \sum_u \sigma_u = 0) \approx \frac{1}{Z} \exp \left[\sum_{(u,v) \in E(G)} J_{uv} \sigma_u \sigma_v \right].$$

- The reconstruction threshold τ is equivalent to

$$\frac{a+b}{2} \mathbb{E}_J [\tanh^2(J)] = 1.$$

Posterior distribution

- The posterior distribution resembles Ising model! Let $J_{uv} = \frac{1}{2} \log \frac{a\mu(L_{uv})}{b\nu(L_{uv})}$

$$\mathbb{P}(\sigma | A, L, \sum_u \sigma_u = 0) \approx \frac{1}{Z} \exp \left[\sum_{(u,v) \in E(G)} J_{uv} \sigma_u \sigma_v \right].$$

- The reconstruction threshold τ is equivalent to

$$\frac{a+b}{2} \mathbb{E}_J [\tanh^2(J)] = 1.$$

Belief Propagation

Let $h_{i\to j} = \frac{1}{2}\frac{m_{i\to j}(+1)}{m_{i\to j}(-1)}$, then BP update equations become

$$h_{i\to j}^{(t+1)} = -B^{(t)} + \sum_{k\in\partial i\setminus j} \text{arctanh}(\tanh(\beta J_{ik}) \tanh h_{k\to i}^{(t)}),$$

$$h_{i}^{(t+1)} = -B^{(t)} + \sum_{k\in\partial i} \text{arctanh}(\tanh(\beta J_{ik}) \tanh h_{k\to i}^{(t)}),$$

$$B^{(t+1)} = \frac{a-b}{2n}\sum_{k} \tanh h_{k}^{(t+1)}.$$

- $B^{(t)}$ is the mean field effect coming from all the disconnected pairs.
- The convergence of BP is unproven, but numerically validated.

Belief Propagation

Let $h_{i\to j} = \frac{1}{2}\frac{m_{i\to j}(+1)}{m_{i\to j}(-1)}$, then BP update equations become

$$h_{i\to j}^{(t+1)} = -B^{(t)} + \sum_{k\in\partial i\setminus j} \mathrm{arctanh}(\tanh(\beta J_{ik})\tanh h_{k\to i}^{(t)}),$$

$$h_{i}^{(t+1)} = -B^{(t)} + \sum_{k\in\partial i} \mathrm{arctanh}(\tanh(\beta J_{ik})\tanh h_{k\to i}^{(t)}),$$

$$B^{(t+1)} = \frac{a-b}{2n}\sum_{k}\tanh h_{k}^{(t+1)}.$$

- $B^{(t)}$ is the mean field effect coming from all the disconnected pairs.
- The convergence of BP is unproven, but numerically validated.

Minimum bisection

Minimum bisection on the weighted graph with weighting function $w(l) = \frac{a\mu(l) - b\nu(l)}{a\mu(l) + b\nu(l)}$.

Theorem

The minimum bisection finds a positively correlated partition if $\sum_l a\mu(l) w^2(l), \sum_l b\nu(l) w^2(l) > 8 \ln 2$ *and* $\tau > 64 \ln 2$.

- Proof uses Chernoff bound, and the weighting function $w(l)$ is chosen optimally.
- If instead $w(l) = \log \frac{a\mu(l)}{b\nu(l)}$, minimum bisection approximates MAP.
- The minimum bisection could be computationally expensive to solve.

Minimum bisection

Minimum bisection on the weighted graph with weighting function $w(l) = \frac{a\mu(l) - b\nu(l)}{a\mu(l) + b\nu(l)}$.

Theorem

The minimum bisection finds a positively correlated partition if $\sum_l a\mu(l) w^2(l), \sum_l b\nu(l) w^2(l) > 8 \ln 2$ *and* $\tau > 64 \ln 2$.

- Proof uses Chernoff bound, and the weighting function $w(l)$ is chosen optimally.
- If instead $w(l) = \log \frac{a\mu(l)}{b\nu(l)}$, minimum bisection approximates MAP.
- The minimum bisection could be computationally expensive to solve.

Minimum bisection

Minimum bisection on the weighted graph with weighting function $w(l) = \frac{a\mu(l) - b\nu(l)}{a\mu(l) + b\nu(l)}$.

Theorem

The minimum bisection finds a positively correlated partition if $\sum_l a\mu(l)w^2(l), \sum_l b\nu(l)w^2(l) > 8\ln 2$ *and* $\tau > 64\ln 2$.

- Proof uses Chernoff bound, and the weighting function $w(l)$ is chosen optimally.
- If instead $w(l) = \log \frac{a\mu(l)}{b\nu(l)}$, minimum bisection approximates MAP.
- The minimum bisection could be computationally expensive to solve.

Spectral method

- Spectral method as an integer relaxation of the minimum bisection: A is the weighted adjacency matrix

$$\max \sum_{(u,v)} (A_{uv} - \frac{a-b}{2n})\sigma_u \sigma_v$$
$$\text{s.t. } \|\sigma\|_2 = 1.$$

- Perturbed low-rank matrix A

$$\mathbb{E}[A|\sigma] = \frac{a-b}{2n}\mathbf{1}\mathbf{1}^\top + \frac{\tau}{n}\sigma\sigma^\top.$$

- Curse from vertices of high degrees $\Omega(\frac{\log n}{\log \log n})$.

Spectral method

- Spectral method as an integer relaxation of the minimum bisection: A is the weighted adjacency matrix

$$\max \sum_{(u,v)} (A_{uv} - \frac{a-b}{2n})\sigma_u \sigma_v$$
$$\text{s.t. } \|\sigma\|_2 = 1.$$

- Perturbed low-rank matrix A

$$\mathbb{E}[A|\sigma] = \frac{a-b}{2n}\mathbf{1}\mathbf{1}^\top + \frac{\tau}{n}\sigma\sigma^\top.$$

- Curse from vertices of high degrees $\Omega(\frac{\log n}{\log\log n})$.

Spectral method

- Spectral method as an integer relaxation of the minimum bisection: A is the weighted adjacency matrix

$$\max \sum_{(u,v)} (A_{uv} - \frac{a-b}{2n})\sigma_u \sigma_v$$
$$\text{s.t. } \|\sigma\|_2 = 1.$$

- Perturbed low-rank matrix A

$$\mathbb{E}[A|\sigma] = \frac{a-b}{2n}\mathbf{1}\mathbf{1}^\top + \frac{\tau}{n}\sigma\sigma^\top.$$

- Curse from vertices of high degrees $\Omega(\frac{\log n}{\log \log n})$.

Spectral method: regular graph

For regular case ($a = b$),

Theorem

If $\tau > C(\mu, \nu)$, then w.h.p., the largest eigenvector of A gives a positively correlated partition.

- Proof: spectrum of random weighted regular graph, ref. [Karger2011].

Spectral method: irregular graph

Remove nodes of degree greater than $(1+\alpha)\frac{a+b}{2}$.

Theorem

If $\tau^2 > C(\alpha)(a+b)$, then w.p. $1 - e^{-\Omega(\alpha^2(a+b))}$, the largest eigenvector of $A - \frac{a-b}{2n}\mathbf{1}\mathbf{1}^\top$ gives a positively correlated partition.

- Proof: spectrum of truncated ER random graph, ref. [Feige2005], [Keshavan2010].

Converse result

Theorem

If $\tau < 1$, then for any fixed vertices u and v, conditional on the spin of v, the spin of u is asymptotically uniformly distributed.

- It further implies that it is impossible to reconstruct a positively correlated partition.
- Proof: similar to [Mossel12], uses local tree argument, conditional independence property and the Ising spin model on labeled tree.

Converse result

Theorem

If $\tau < 1$, then for any fixed vertices u and v, conditional on the spin of v, the spin of u is asymptotically uniformly distributed.

- It further implies that it is impossible to reconstruct a positively correlated partition.
- Proof: similar to [Mossel12], uses local tree argument, conditional independence property and the Ising spin model on labeled tree.

Future work

- Apply sparse labeled stochastic block model into real data
- Convergence of belief propagation
- Polynomial-time reconstruction algorithm approaching the reconstruction threshold

Future work

- Apply sparse labeled stochastic block model into real data
- Convergence of belief propagation
- Polynomial-time reconstruction algorithm approaching the reconstruction threshold