Efficient Bayesian inference of multi-scale network structures

Tiago P. Peixoto

University of Bath

Bath, September 2016

Identification of modular structure

Generative models

Model likelihood: $P(\boldsymbol{A}|\theta, \boldsymbol{b})$

$\boldsymbol{A} \to$ network
$\boldsymbol{b} \to$ partition of the nodes into groups
$\theta \to$ more model parameters

Bayesian inference

$$P(\boldsymbol{b}|\boldsymbol{A}) = \frac{P(\boldsymbol{A}|\boldsymbol{b})P(\boldsymbol{b})}{P(\boldsymbol{A})}$$

$$P(\boldsymbol{A}|\boldsymbol{b}) = \int P(\boldsymbol{A}|\theta, \boldsymbol{b})P(\theta)\mathrm{d}\theta$$

The stochastic block model (SBM)

Common formulations

Bernoulli

$$P(\boldsymbol{A}|\boldsymbol{p},\boldsymbol{b}) = \prod_{i<j} p_{b_i b_j}^{A_{ij}} (1 - p_{b_i b_j})^{1-A_{ij}}$$

$p_{rs} \to$ edge probability between two nodes of groups r and s

(simple graphs)

Poisson

$$P(\boldsymbol{A}|\boldsymbol{\lambda},\boldsymbol{b}) = \prod_{i<j} \frac{\lambda_{b_i b_j}^{A_{ij}} e^{-\lambda_{b_i b_j}}}{A_{ij}!}$$

$\lambda_{rs} \to$ average number of edges between two nodes of groups r and s

(multigraphs)

The stochastic block model (SBM)

Common formulations

Bernoulli

$$P(\boldsymbol{A}|\boldsymbol{p},\boldsymbol{b}) = \prod_{i<j} p_{b_i b_j}^{A_{ij}}(1-p_{b_i b_j})^{1-A_{ij}}$$

$p_{rs} \to$ edge probability between two nodes of groups r and s

(simple graphs)

Poisson

$$P(\boldsymbol{A}|\boldsymbol{\lambda},\boldsymbol{b}) = \prod_{i<j} \frac{\lambda_{b_i b_j}^{A_{ij}} e^{-\lambda_{b_i b_j}}}{A_{ij}!}$$

$\lambda_{rs} \to$ average number of edges between two nodes of groups r and s

(multigraphs)

Microcanonical

$$P(\boldsymbol{A}|\boldsymbol{e},\boldsymbol{b})$$

$e_{rs} \to$ total number of edges between groups r and s

(i.e. hard constraints)

Why Microcanonical?

Two good advantages:

- Efficiency: Inference becomes possible not only for networks with a very large number of nodes and edges, but also with an **unlimited number of groups.**

- Deeper Bayesian hierarchies: Noninformative priors can be replaced by conditioned priors and hyperpriors.

Marginal likelihood

"Canonical" marginal likelihood:

$$P(\boldsymbol{A}|\boldsymbol{b}) = \int P(\boldsymbol{A}|\theta, \boldsymbol{b})P(\theta)\mathrm{d}\theta$$

Marginal likelihood

"Canonical" marginal likelihood:

$$P(\boldsymbol{A}|\boldsymbol{b}) = \int P(\boldsymbol{A}|\theta, \boldsymbol{b})P(\theta)\mathrm{d}\theta$$

Microcanonical marginal likelihood:

$$\begin{aligned} P(\boldsymbol{A}|\boldsymbol{b}) &= \sum_{\theta} P(\boldsymbol{A}|\theta, \boldsymbol{b})P(\theta) \\ &= P(\boldsymbol{A}|\hat{\theta}, \boldsymbol{b})P(\hat{\theta}) = P(\boldsymbol{A}, \hat{\theta}|\boldsymbol{b}) \end{aligned}$$

$$\boxed{\text{Marginal} \leftrightarrow \text{Joint}}$$

Deeper Bayesian hierarchies:

$$\begin{aligned} P(\boldsymbol{A}|\boldsymbol{b}) &= \sum_{\theta}\sum_{\alpha} P(\boldsymbol{A}|\theta, \boldsymbol{b})P(\theta|\alpha)P(\alpha) \\ &= P(\boldsymbol{A}|\hat{\theta}, \boldsymbol{b})P(\hat{\theta}|\hat{\alpha})P(\hat{\alpha}) = P(\boldsymbol{A}, \hat{\theta}, \hat{\alpha}|\boldsymbol{b}) \end{aligned}$$

Minimum description length (MDL)

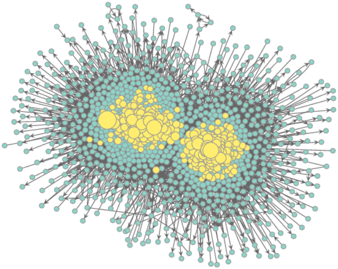
$$P(\boldsymbol{b}|\boldsymbol{A}) = \frac{P(\boldsymbol{A}, \boldsymbol{b})}{P(\boldsymbol{A})} = \frac{P(\boldsymbol{A}, \theta, \boldsymbol{b})}{P(\boldsymbol{A})} = \frac{\exp(-\Sigma)}{P(\boldsymbol{A})}$$

Description length:

$$\Sigma = -\underbrace{\ln P(\boldsymbol{A}|\theta, \boldsymbol{b})}_{\text{Data}} - \underbrace{\ln P(\theta, \boldsymbol{b})}_{\text{Model}}$$

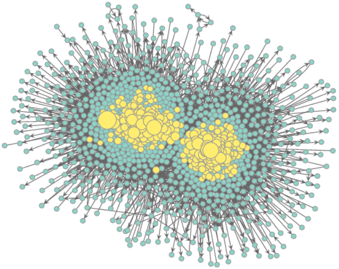
Degree-correction

Traditional SBM

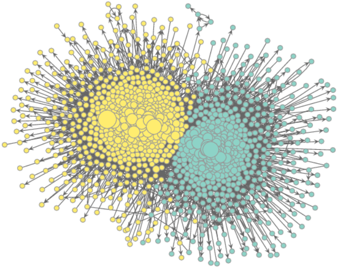


Degree-correction

Traditional SBM



Degree-corrected SBM

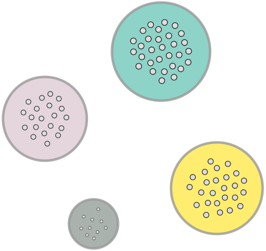


$$P(\boldsymbol{A}|\boldsymbol{\lambda},\boldsymbol{\theta},\boldsymbol{b}) = \prod_{i<j} \frac{(\theta_i\theta_j\lambda_{b_ib_j})^{A_{ij}} e^{-\theta_i\theta_j\lambda_{b_ib_j}}}{A_{ij}!}$$

$\theta_i \to$ average degree of node i

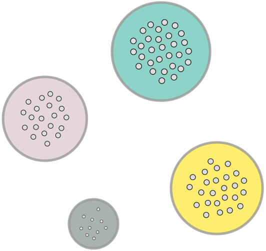
(Karrer and Newman 2011)

Degree-corrected microcanonical SBM

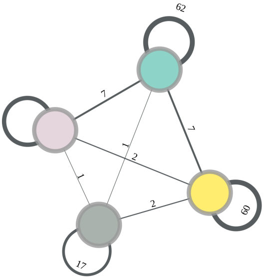


Node partition, $\boldsymbol{b}$.

Degree-corrected microcanonical SBM

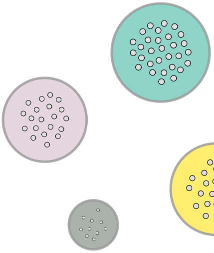


Node partition, $\boldsymbol{b}$.

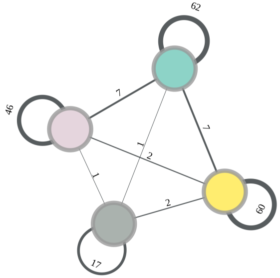


Edge counts between groups, $\boldsymbol{e}$.

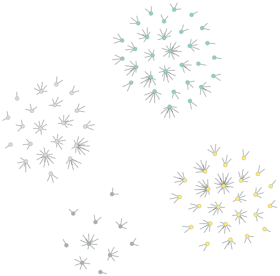
Degree-corrected microcanonical SBM



Node partition, $\boldsymbol{b}$.

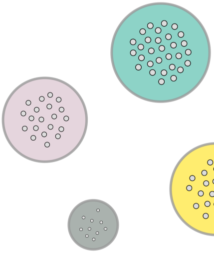


Edge counts between groups, $\boldsymbol{e}$.

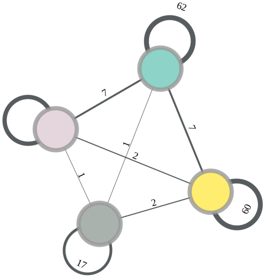


Degree sequence, $\boldsymbol{k}$.

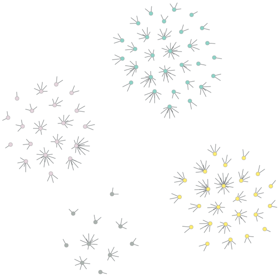
Degree-corrected microcanonical SBM



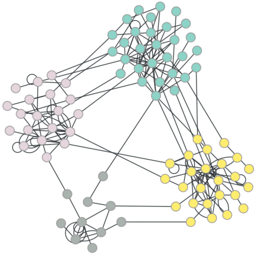
Node partition, $\boldsymbol{b}$.



Edge counts between groups, $\boldsymbol{e}$.



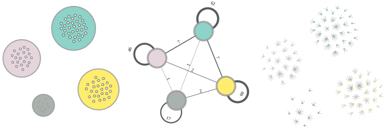
Degree sequence, $\boldsymbol{k}$.

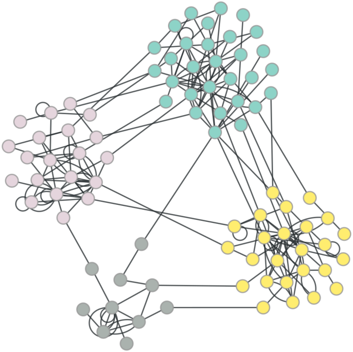


Adjacency matrix, $\boldsymbol{A}$.

Model likelihood: Half-edge pairings

Configuration model with group structure



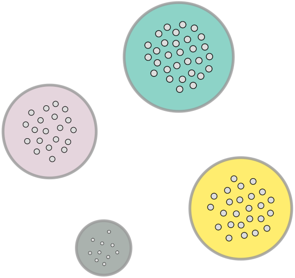


$$P(\boldsymbol{A}|\boldsymbol{e},\boldsymbol{k},\boldsymbol{b}) = \frac{\Xi(\boldsymbol{A})}{\Omega(\boldsymbol{e})}$$

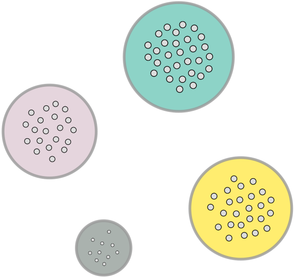
$$\Omega(\boldsymbol{e}) = \frac{\prod_{r<s} e_{rs}! \prod_r e_{rr}!!}{\prod_r e_r!}$$

$$\Xi(\boldsymbol{A}) = \frac{\prod_{i<j} A_{ij}! \prod_i A_{ii}!!}{\prod_i k_i!}$$

Prior for the partition



Prior for the partition

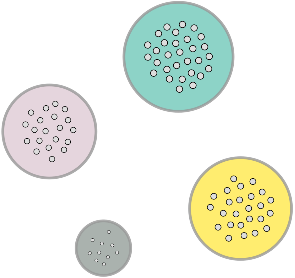


Option 1: Noninformative

$$P(\boldsymbol{b}|B) = \frac{1}{B^N(1 - B(1-1/B)^N)}$$

$B \to$ Number of **nonempty** groups

Prior for the partition



Option 1: Noninformative

$$P(\boldsymbol{b}|B) = \frac{1}{B^N(1 - B(1-1/B)^N)}$$

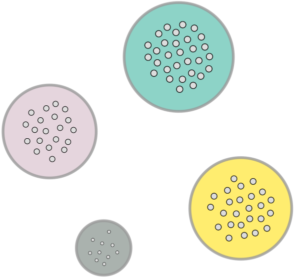
$B \to$ Number of **nonempty** groups

Option 2: Conditioned on group sizes

$$P(\boldsymbol{b}|B) = P(\boldsymbol{b}|\boldsymbol{n})P(\boldsymbol{n})$$

$$= \frac{N!}{\prod_r n_r!} \times \binom{N-1}{B-1}^{-1}$$

Prior for the partition



Option 1: Noninformative

$$P(\boldsymbol{b}|B) = \frac{1}{B^N(1 - B(1-1/B)^N)}$$

$B \rightarrow$ Number of **nonempty** groups

Option 2: Conditioned on group sizes

$$\begin{aligned} P(\boldsymbol{b}|B) &= P(\boldsymbol{b}|\boldsymbol{n})P(\boldsymbol{n}) \\ &= \frac{N!}{\prod_r n_r!} \times \binom{N-1}{B-1}^{-1} \end{aligned}$$

For $N \gg B$:

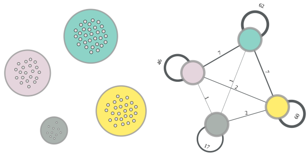
$$-\ln P(\boldsymbol{b}|B) \approx NH(\boldsymbol{n}) + O(B \ln N)$$

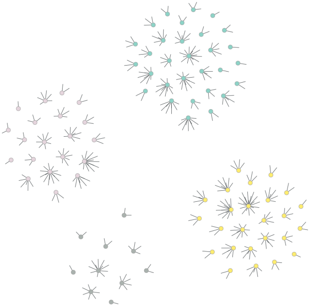
$$H(\boldsymbol{n}) = -\sum_r \frac{n_r}{N} \ln \frac{n_r}{N}$$

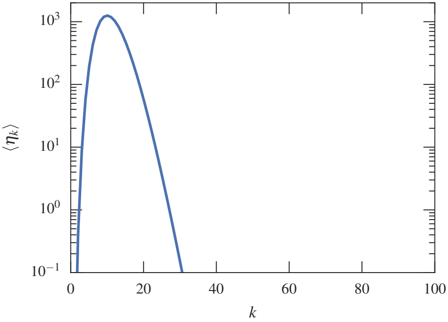
Prior for the degrees

Option 0: Random half-edges
(Non-degree-corrected)

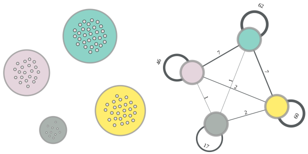
$$P(\boldsymbol{k}|\boldsymbol{e},\boldsymbol{b}) = \prod_r \frac{e_r!}{n_r^{e_r}\prod_{i\in r} k_i!}$$

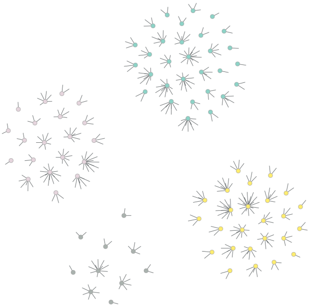






Prior for the degrees



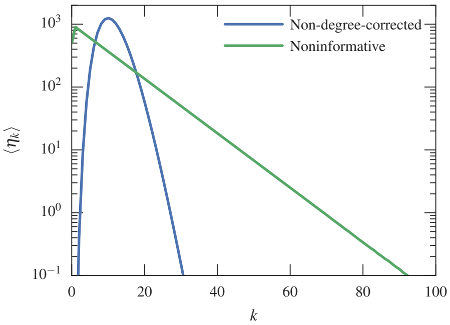


Option 0: Random half-edges
(Non-degree-corrected)

$$P(\boldsymbol{k}|\boldsymbol{e},\boldsymbol{b}) = \prod_r \frac{e_r!}{n_r^{e_r}\prod_{i\in r} k_i!}$$

Option 1: Noninformative

$$P(\boldsymbol{k}|\boldsymbol{e},\boldsymbol{b}) = \prod_r \left(\!\!\binom{n_r}{e_r}\!\!\right)^{-1}$$



Prior for the degrees

Option 2: Conditioned on degree distribution

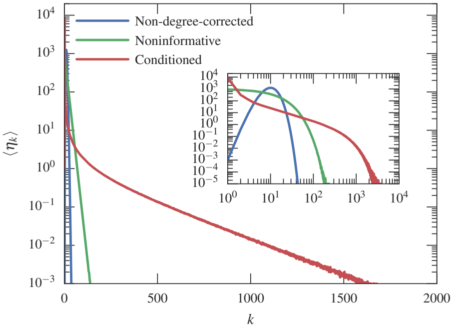
$$
\begin{aligned}
P(\boldsymbol{k}|\boldsymbol{b},\boldsymbol{e}) &= P(\boldsymbol{k}|\boldsymbol{\eta})P(\boldsymbol{\eta}) \\
&= \prod_r \frac{n_r!}{\prod_r \eta_k^r!} \times \prod_r q(e_r, n_r)^{-1}
\end{aligned}
$$

Prior for the degrees

Option 2: Conditioned on degree distribution

$$P(\boldsymbol{k}|\boldsymbol{b},\boldsymbol{e}) = P(\boldsymbol{k}|\boldsymbol{\eta})P(\boldsymbol{\eta})$$

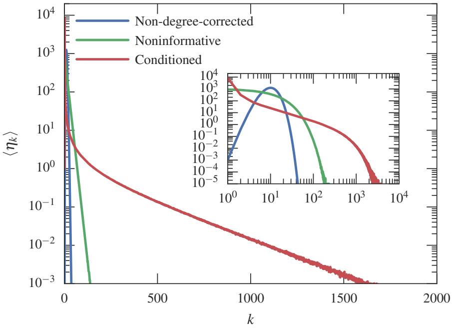
$$= \prod_r \frac{n_r!}{\prod_r \eta_k^r!} \times \prod_r q(e_r, n_r)^{-1}$$



Prior for the degrees

Option 2: Conditioned on degree distribution

$$P(\boldsymbol{k}|\boldsymbol{b},\boldsymbol{e}) = P(\boldsymbol{k}|\boldsymbol{\eta})P(\boldsymbol{\eta})$$
$$= \prod_r \frac{n_r!}{\prod_r \eta_k^r!} \times \prod_r q(e_r, n_r)^{-1}$$



<u>Bose-Einstein statistics</u>

N "bosons" (nodes)

$k_i \to$ "energy level" (degree)

$$\langle \eta_k \rangle = \frac{1}{\exp(\mu k + \lambda) - 1}$$

$$N \gg 1,\ E \sim N$$
$$\mu \to 1/\sqrt{E},\ \lambda \to 0$$

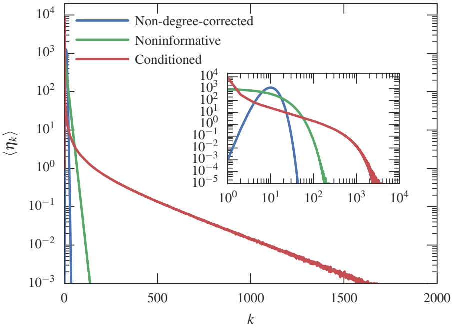
$$\langle \eta_k \rangle \approx \frac{1}{\lambda + \mu k} \approx \frac{1}{k} \qquad k \ll E$$

Prior for the degrees

Option 2: Conditioned on degree distribution

$$P(\boldsymbol{k}|\boldsymbol{b},\boldsymbol{e}) = P(\boldsymbol{k}|\boldsymbol{\eta})P(\boldsymbol{\eta})$$

$$= \prod_r \frac{n_r!}{\prod_r \eta_k^r!} \times \prod_r q(e_r, n_r)^{-1}$$



<u>Bose-Einstein statistics</u>

N "bosons" (nodes)

$k_i \to$ "energy level" (degree)

$$\langle \eta_k \rangle = \frac{1}{\exp(\mu k + \lambda) - 1}$$

$$N \gg 1,\ E \sim N$$
$$\mu \to 1/\sqrt{E},\ \lambda \to 0$$

$$\langle \eta_k \rangle \approx \frac{1}{\lambda + \mu k} \approx \frac{1}{k} \qquad k \ll E$$

$$-\ln P(\boldsymbol{k}|\boldsymbol{e},\boldsymbol{b}) \approx \sum_r n_r H(\boldsymbol{\eta}_r) + O(\sqrt{n_r})$$

$$H(\boldsymbol{\eta}_r) = -\sum_k (\eta_k^r/n_r)\ln(\eta_k^r/n_r)$$

Prior for the degrees: Integer partitions

$q(m, n) \rightarrow$ number of partitions of integer m into at most n parts

Exact computation

$$q(m, n) = q(m, n-1) + q(m-n, n)$$

Full table for $n \leq m \leq M$ in time $O(M^2)$.

Prior for the degrees: Integer partitions

$q(m,n) \rightarrow$ number of partitions of integer m into at most n parts

Exact computation

$$q(m,n) = q(m,n-1) + q(m-n,n)$$

Full table for $n \leq m \leq M$ in time $O(M^2)$.

Approximation

$$q(m,n) \approx \frac{f(u)}{m} \exp(\sqrt{m}g(u)),$$

$$f(u) = \frac{v(u)}{2^{3/2}\pi u}\left[1-(1+u^2/2)e^{-v(u)}\right]^{-1/2}$$

$$g(u) = \frac{2v(u)}{u} - u\ln(1-e^{-v(u)})$$

$$v = u\sqrt{-v^2/2 - \mathrm{Li}_2(e^v)}$$

(Szekeres, 1951)

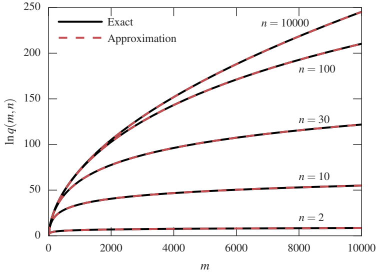
Prior for the degrees: Integer partitions

$q(m, n) \rightarrow$ number of partitions of integer m into at most n parts

Exact computation

$$q(m, n) = q(m, n-1) + q(m-n, n)$$

Full table for $n \le m \le M$ in time $O(M^2)$.



Approximation

$$q(m, n) \approx \frac{f(u)}{m} \exp(\sqrt{m} g(u)),$$

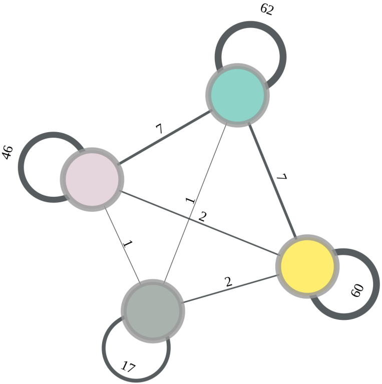
$$f(u) = \frac{v(u)}{2^{3/2}\pi u} \left[1 - (1 + u^2/2)e^{-v(u)}\right]^{-1/2}$$

$$g(u) = \frac{2v(u)}{u} - u \ln(1 - e^{-v(u)})$$

$$v = u\sqrt{-v^2/2 - \mathrm{Li}_2(e^v)}$$

(Szekeres, 1951)

Prior for the edge counts

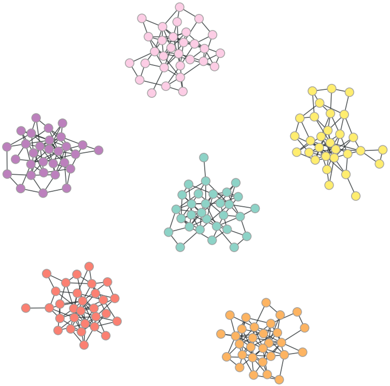


Option 1: Noninformative

$$P(\boldsymbol{e}) = \left(\!\!\left(\begin{matrix}\left(\!\!\binom{B}{2}\!\!\right)\\ E\end{matrix}\right)\!\!\right)^{-1}$$

$$\left(\!\!\binom{n}{m}\!\!\right) = \binom{n+m-1}{m}$$

"Resolution" problem



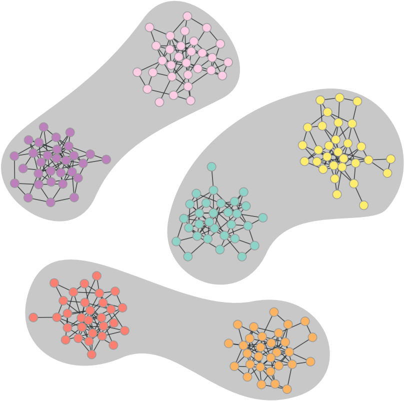
Planted partition

B groups of uniform size N/B,
perfectly assortative:
$e_{rr} = 2E/B$, $e_{rs} = 0, r \neq s$.

For $N \gg 1, E \sim N, B \gg 1$

$$-\ln P(\boldsymbol{A}, \boldsymbol{e}, \boldsymbol{b}) \approx \underbrace{2E \ln N/B}_{\text{Network}} + \underbrace{N \ln B}_{\text{Partition}} + \underbrace{B^2 \ln 2E}_{\text{Edge counts}}$$

"Resolution" problem



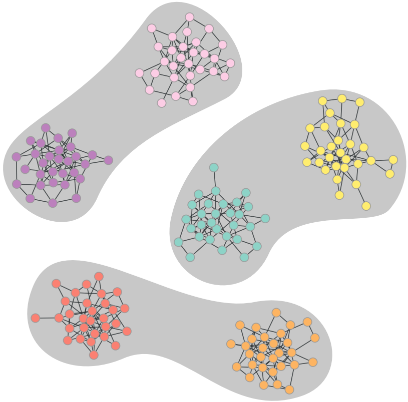
Planted partition

B groups of uniform size N/B,
perfectly assortative:
$e_{rr} = 2E/B$, $e_{rs} = 0, r \neq s$.

For $N \gg 1, E \sim N, B \gg 1$

$$-\ln P(\boldsymbol{A}, \boldsymbol{e}, \boldsymbol{b}) \approx \underbrace{2E \ln N/B}_{\text{Network}} + \underbrace{N \ln B}_{\text{Partition}} + \underbrace{B^2 \ln 2E}_{\text{Edge counts}}$$

"Resolution" problem



Planted partition

B groups of uniform size N/B,
perfectly assortative:
$e_{rr} = 2E/B$, $e_{rs} = 0, r \neq s$.

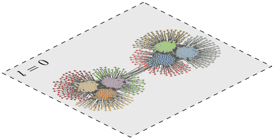
For $N \gg 1, E \sim N, B \gg 1$

$$-\ln P(\boldsymbol{A}, \boldsymbol{e}, \boldsymbol{b}) \approx \underbrace{2E \ln N/B}_{\text{Network}} + \underbrace{N \ln B}_{\text{Partition}} + \underbrace{B^2 \ln 2E}_{\text{Edge counts}}$$

$$B_{\max} = O(\sqrt{N})$$

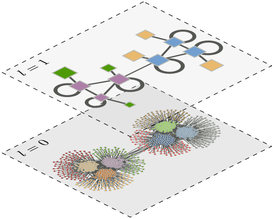
Prior for the edge counts: Group hierarchies

Option 2: Conditioned on... another SBM!



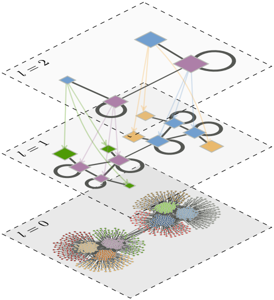
Prior for the edge counts: Group hierarchies

Option 2: Conditioned on... another SBM!

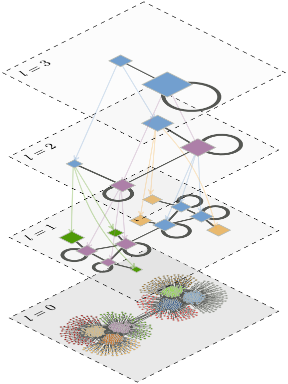


Prior for the edge counts: Group hierarchies

Option 2: Conditioned on... another SBM!

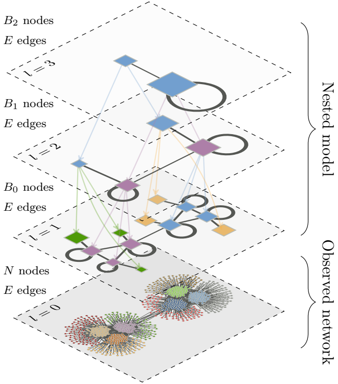


Prior for the edge counts: Group hierarchies



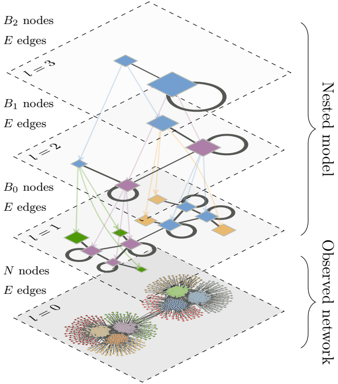
Option 2: Conditioned on... another SBM!

Prior for the edge counts: Group hierarchies



Option 2: Conditioned on... another SBM!

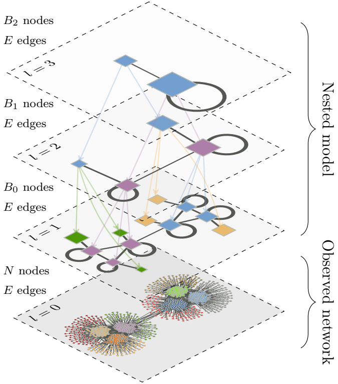
Prior for the edge counts: Group hierarchies



Option 2: Conditioned on... another SBM!

$$P(\boldsymbol{e}_l|\boldsymbol{b}_l,\boldsymbol{e}_{l+1}) = \prod_{r<s}\left(\!\!\binom{n_r^l n_s^l}{e_{rs}^{l+1}}\!\!\right)^{-1} \times \prod_r \left(\!\!\binom{n_r^l(n_r^l+1)/2}{e_{rr}^{l+1}/2}\!\!\right)^{-1}$$

Prior for the edge counts: Group hierarchies



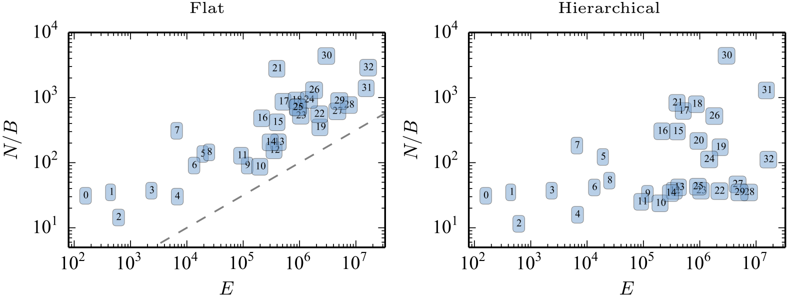
Option 2: Conditioned on... another SBM!

$$P(\boldsymbol{e}_l|\boldsymbol{b}_l,\boldsymbol{e}_{l+1}) = \prod_{r<s}\left(\!\!\binom{n_r^l n_s^l}{e_{rs}^{l+1}}\!\!\right)^{-1}\times \prod_r \left(\!\!\binom{n_r^l(n_r^l+1)/2}{e_{rr}^{l+1}/2}\!\!\right)^{-1}$$

For planted partition:

$$B_{\max} = O\left(\frac{N}{\log N}\right)$$

Resolution problem in the wild



The whole model

$$P(\boldsymbol{A}, \boldsymbol{k}, \{\boldsymbol{b}_l\}, \{\boldsymbol{e}_l\} | B, E, L) = P(\boldsymbol{A} | \boldsymbol{k}, \boldsymbol{b}_0, \boldsymbol{e}_1) P(\boldsymbol{k} | \boldsymbol{b}_0, \boldsymbol{e}_1) \prod_{l=1}^{L} P(\boldsymbol{e}_l | \boldsymbol{b}_l, \boldsymbol{e}_{l+1}) P(\boldsymbol{b}_l | B_l)$$

Boundary: $B_L = 1$, $\boldsymbol{e}_{L+1} = \{2E\}$

$$P(B) = \frac{1}{N} \qquad P(E) = \binom{N}{2}^{-1} \qquad P(L) = 1/N$$

Inference algorithm: Metropolis-Hastings

- Move proposal, $b_i = r \to b_i = s$
- Accept with probability

$$a = \min\left(1, \frac{P(\boldsymbol{b}'|\boldsymbol{A})P(\boldsymbol{b} \to \boldsymbol{b}')}{P(\boldsymbol{b}|\boldsymbol{A})P(\boldsymbol{b}' \to \boldsymbol{b})}\right).$$

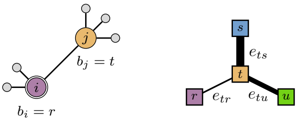
Move proposal of node i requires $O(k_i)$ operations. A whole MCMC sweep can be performed in $O(E)$ time, **independent of the number of groups,** B.

In contrast to:
1. EM + BP with Bernoulli SBM: $O(EB^2)$ (Semiparametric) [Decelle et al, 2011]
2. Variational Bayes with (overlapping) Bernoulli SBM: $O(EB)$ (Semiparametric) [Gopalan and Blei, 2011]
3. Bernoulli SBM with noninformative priors: $O(EB^2)$ (Greedy) [Côme and Latouche, 2015]
4. Poisson SBM with noninformative priors: $O(EB^2)$ (Gibbs) [Newman and Reinert, 2016]

Efficient inference: other improvements

Smart move proposals

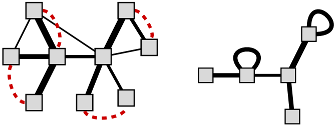
- Choose a random vertex i (happens to belong to group r).
- Move it to a random group $s \in [1, B]$, chosen with a probability $p(r \to s|t)$ proportional to $e_{ts} + \epsilon$, where t is the group membership of a randomly chosen neighbour of i.



Fast mixing times.

Agglomerative initialization

- Start with $B = N$.
- Progressively merge groups.



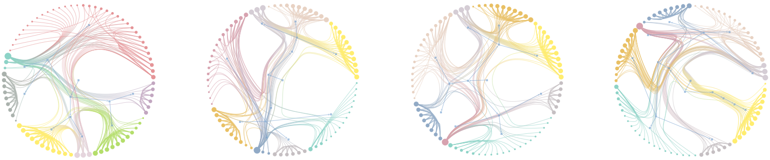
Avoids metastable states.

Sampling hierarchies

$$
\begin{aligned}
P(\{\boldsymbol{b}_l\}|\boldsymbol{A}) &= \frac{P(\boldsymbol{A}, \boldsymbol{k}, \boldsymbol{b}_0|\boldsymbol{e}_1)\prod_l P(\boldsymbol{e}_{l-1}, \boldsymbol{b}_l|\boldsymbol{e}_l)}{P(\boldsymbol{A})} \\
&= P(\boldsymbol{b}_0|\boldsymbol{A}, \boldsymbol{e}_1, \boldsymbol{k})\prod_l P(\boldsymbol{b}_l|\boldsymbol{e}_{l-1}, \boldsymbol{e}_l)
\end{aligned}
$$

with per-level posteriors

$$
\begin{aligned}
P(\boldsymbol{b}_0|\boldsymbol{A}, \boldsymbol{k}, \boldsymbol{e}_1) &= \frac{P(\boldsymbol{A}|\boldsymbol{e}_1, \boldsymbol{k}, \boldsymbol{b}_0)P(\boldsymbol{k})P(\boldsymbol{b}_0)}{P(\boldsymbol{A}|\boldsymbol{e}_1)} \\
P(\boldsymbol{b}_l|\boldsymbol{e}_{l-1}, \boldsymbol{e}_l) &= \frac{P(\boldsymbol{e}_l|\boldsymbol{e}_{l+1}, \boldsymbol{b}_l)P(\boldsymbol{b}_l)}{P(\boldsymbol{e}_l|\boldsymbol{e}_{l+1})}.
\end{aligned}
$$



Sampling vs. Maximization

Maximization (MDL)

$$\{\hat{\boldsymbol{b}}_l\} = \underset{\{\boldsymbol{b}_l\}}{\operatorname{argmax}}\, P(\{\boldsymbol{b}_l\}|\boldsymbol{A})$$

Finds the best partition, and number of groups $\{B_l\}$.

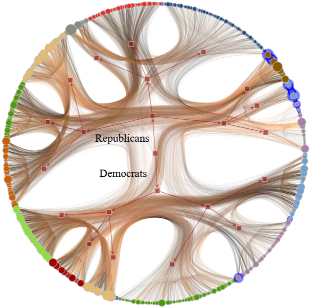
Sampling

$$\{\boldsymbol{b}_l\} \sim P(\{\boldsymbol{b}_l\}|\boldsymbol{A})$$

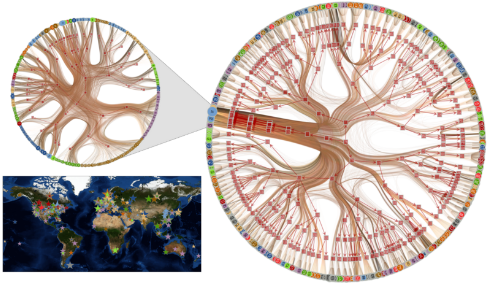
Finds the best posterior ensemble; marginal posterior $P(B_l|\boldsymbol{A})$.

Some results for maximization

Political blogs
$(N = 1,222, E = 19,027)$

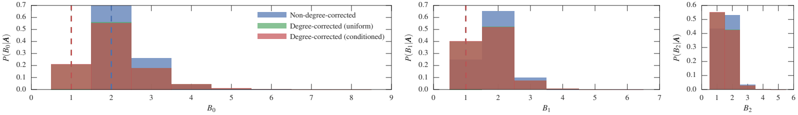


Internet (AS)
$(N = 52,104, E = 399,625)$

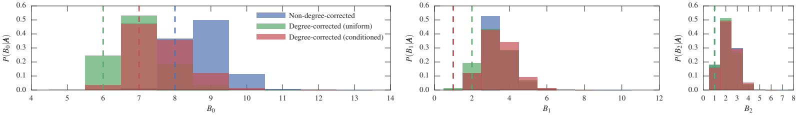


Sampling vs. maximization

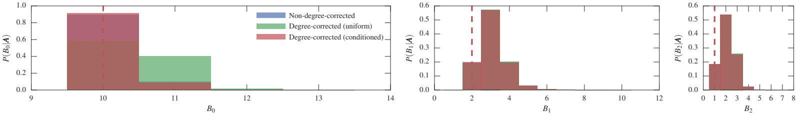
Karate Club $(N = 34, E = 78)$



Les Misérables character co-appearances $(N = 77, E = 254)$

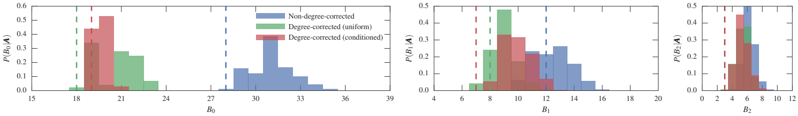


American Football $(N = 115, E = 613)$

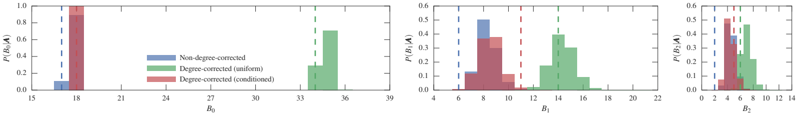


Sampling vs. maximization

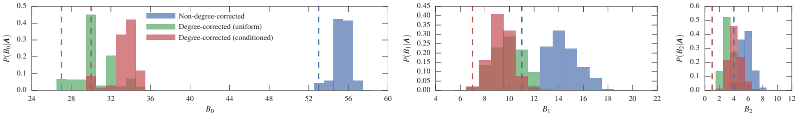
C. elegans neural network ($N = 297, E = 2,359$)



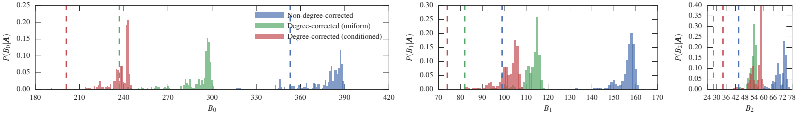
Power grid ($N = 4,941, E = 6,594$)



E-mail ($N = 1,133, E = 5,451$)

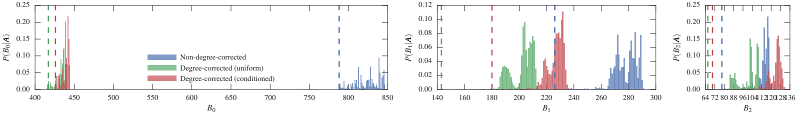


Airports ($N = 3,286, E = 68,344$)

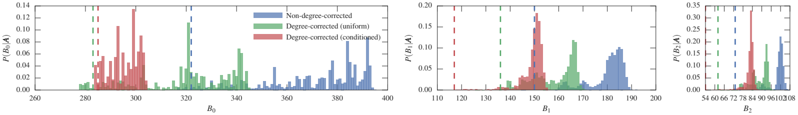


Sampling vs. maximization

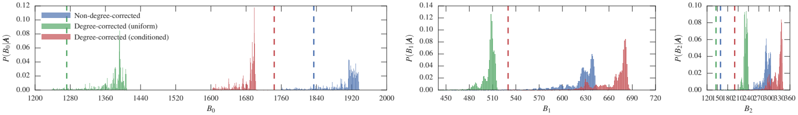
Linux source code ($N = 30,837, E = 213,954$)



Human protein interactions ($N = 6,327, E = 147,547$)



PGP web of trust ($N = 39,796, E = 301,498$)



Model selection

1. Best model fit

$$\Lambda = \frac{P(\boldsymbol{A}, \{\boldsymbol{b}_l\}|\mathcal{H}_1)}{P(\boldsymbol{A}, \{\boldsymbol{b}'_l\}|\mathcal{H}_2)} = \exp(-\Delta\Sigma)$$

Model selection

1. Best model fit

$$\Lambda = \frac{P(\boldsymbol{A}, \{\boldsymbol{b}_l\} | \mathcal{H}_1)}{P(\boldsymbol{A}, \{\boldsymbol{b}'_l\} | \mathcal{H}_2)} = \exp(-\Delta\Sigma)$$

2. Average over models

$$\Lambda = \frac{P(\boldsymbol{A} | \mathcal{H}_1)}{P(\boldsymbol{A} | \mathcal{H}_2)}$$

$$P(\boldsymbol{A} | \mathcal{H}) = \sum_{\{\boldsymbol{b}_l\}} P(\boldsymbol{A}, \{\boldsymbol{b}_l\} | \mathcal{H})$$

Model selection

1. Best model fit

$$\Lambda = \frac{P(\boldsymbol{A}, \{\boldsymbol{b}_l\}|\mathcal{H}_1)}{P(\boldsymbol{A}, \{\boldsymbol{b}'_l\}|\mathcal{H}_2)} = \exp(-\Delta\Sigma)$$

2. Average over models

$$\Lambda = \frac{P(\boldsymbol{A}|\mathcal{H}_1)}{P(\boldsymbol{A}|\mathcal{H}_2)}$$

$$P(\boldsymbol{A}|\mathcal{H}) = \sum_{\{\boldsymbol{b}_l\}} P(\boldsymbol{A}, \{\boldsymbol{b}_l\}|\mathcal{H})$$

$$\ln P(\boldsymbol{A}) = \underbrace{\sum_{\{\boldsymbol{b}_l\}} q(\{\boldsymbol{b}_l\}) \ln P(\boldsymbol{A}, \{\boldsymbol{b}_l\})}_{-\langle\Sigma\rangle} \underbrace{- \sum_{\{\boldsymbol{b}_l\}} q(\{\boldsymbol{b}_l\}) \ln q(\{\boldsymbol{b}_l\})}_{-\mathcal{S}}$$

$$q(\{\boldsymbol{b}_l\}) = P(\{\boldsymbol{b}_l\}|\boldsymbol{A}) = \frac{P(\{\boldsymbol{b}_l\}, \boldsymbol{A})}{P(\boldsymbol{A})}$$

Model selection

1. Best model fit

$$\Lambda = \frac{P(\boldsymbol{A},\{\boldsymbol{b}_l\}|\mathcal{H}_1)}{P(\boldsymbol{A},\{\boldsymbol{b}'_l\}|\mathcal{H}_2)} = \exp(-\Delta\Sigma)$$

2. Average over models

$$\Lambda = \frac{P(\boldsymbol{A}|\mathcal{H}_1)}{P(\boldsymbol{A}|\mathcal{H}_2)}$$

$$P(\boldsymbol{A}|\mathcal{H}) = \sum_{\{\boldsymbol{b}_l\}} P(\boldsymbol{A},\{\boldsymbol{b}_l\}|\mathcal{H})$$

$$\ln P(\boldsymbol{A}) = \underbrace{\sum_{\{\boldsymbol{b}_l\}} q(\{\boldsymbol{b}_l\}) \ln P(\boldsymbol{A},\{\boldsymbol{b}_l\})}_{-\langle\Sigma\rangle} - \underbrace{\sum_{\{\boldsymbol{b}_l\}} q(\{\boldsymbol{b}_l\}) \ln q(\{\boldsymbol{b}_l\})}_{-\mathcal{S}}$$

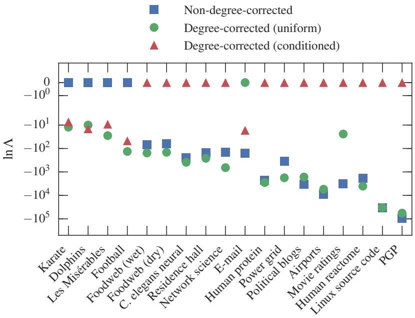
$$q(\{\boldsymbol{b}_l\}) = P(\{\boldsymbol{b}_l\}|\boldsymbol{A}) = \frac{P(\{\boldsymbol{b}_l\},\boldsymbol{A})}{P(\boldsymbol{A})}$$

$$q(\{\boldsymbol{b}_l\}) \approx \prod_l q_l(\boldsymbol{b}_l)$$

$$q_0(\boldsymbol{b}_0) \approx \prod_{i<j} q_{ij}(b_i^0, b_j^0)^{A_{ij}} \prod_i q_i(b_i^0)^{1-k_i}, \qquad q_l(\boldsymbol{b}_l) \approx \prod_i q_i^l(b_i^l), \quad l > 0$$

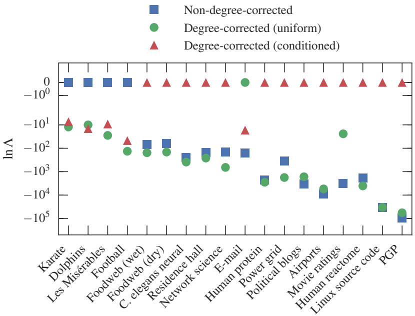
Model selection

1. Best model fit (MDL)

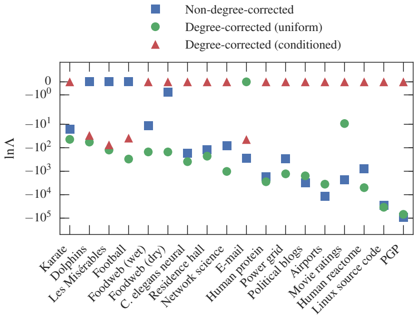


Model selection

1. Best model fit (MDL)



2. Average over models



For code, see:

`http://graph-tool.skewed.de`

Papers:

`http://skewed.de/tiago`

T.P.P., Phys. Rev. E 92, 042807 (2015)
T.P.P., Phys. Rev. X 5, 011033 (2015)
T.P.P., Phys. Rev. X 4, 011047 (2014)
T.P.P., Phys. Rev. E 89, 012804 (2014)
T.P.P., Phys. Rev. Lett. 110 148701 (2013)