

Efficient Bayesian inference of multi-scale network structures

Tiago P. Peixoto

*Universität Bremen
Germany*

(Soon: University of Bath, UK)

*ISI Foundation
Turin, Italy*

Cambridge, July 2016

IDENTIFICATION OF MODULAR STRUCTURE

Generative models

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

\mathbf{A} \rightarrow network

\mathbf{b} \rightarrow partition of the nodes into groups

θ \rightarrow more model parameters

Bayesian inference

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

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



THE STOCHASTIC BLOCK MODEL (SBM)

COMMON FORMULATIONS

Bernoulli

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

p_{rs} → edge probability between two nodes of groups r and s

(simple graphs)

Poisson

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

λ_{rs} → average number of edges between two nodes of groups r and s

(multigraphs)

THE STOCHASTIC BLOCK MODEL (SBM)

COMMON FORMULATIONS

Bernoulli

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

p_{rs} → edge probability between two nodes of groups r and s

(simple graphs)

Poisson

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

λ_{rs} → average number of edges between two nodes of groups r and s

(multigraphs)

Microcanonical

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

e_{rs} → 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:

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

MARGINAL LIKELIHOOD

“Canonical” marginal likelihood:

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

Microcanonical marginal likelihood:

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

Marginal \leftrightarrow Joint

Deeper Bayesian hierarchies:

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

MINIMUM DESCRIPTION LENGTH (MDL)

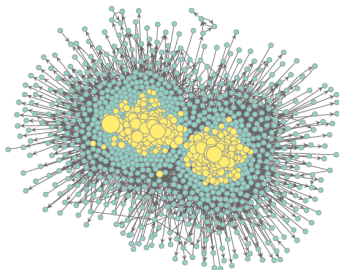
$$P(\mathbf{b}|\mathbf{A}) = \frac{P(\mathbf{A}, \mathbf{b})}{P(\mathbf{A})} = \frac{P(\mathbf{A}, \theta, \mathbf{b})}{P(\mathbf{A})} = \frac{\exp(-\Sigma)}{P(\mathbf{A})}$$

Description length:

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

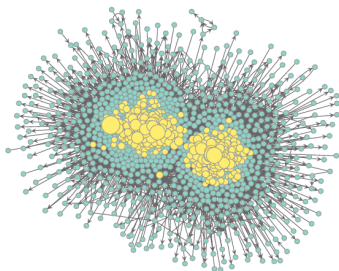
DEGREE-CORRECTION

Traditional SBM

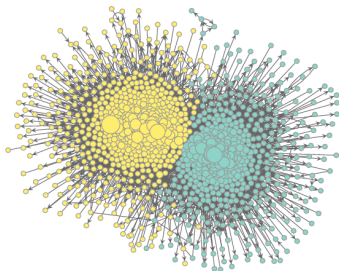


DEGREE-CORRECTION

Traditional SBM



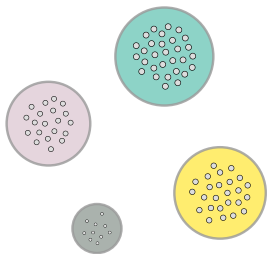
Degree-corrected SBM



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

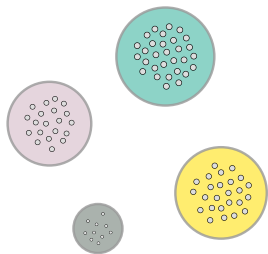
$\theta_i \rightarrow$ average degree of node i
(Karrer and Newman 2011)

DEGREE-CORRECTED MICROCANONICAL SBM

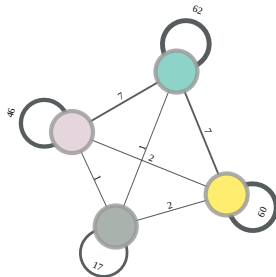


Node partition, \mathbf{b} .

DEGREE-CORRECTED MICROCANONICAL SBM

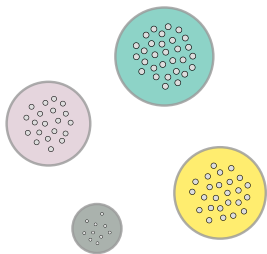


Node partition, \mathbf{b} .

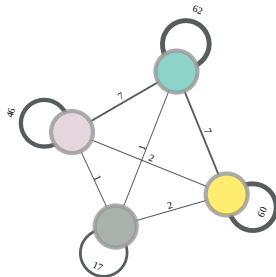


Edge counts between groups, \mathbf{e} .

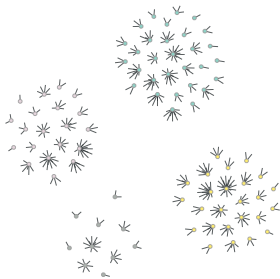
DEGREE-CORRECTED MICROCANONICAL SBM



Node partition, \mathbf{b} .

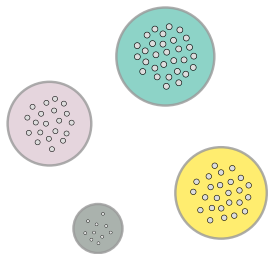


Edge counts between groups, \mathbf{e} .

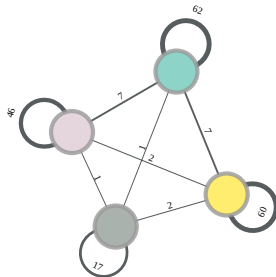


Degree sequence, \mathbf{k} .

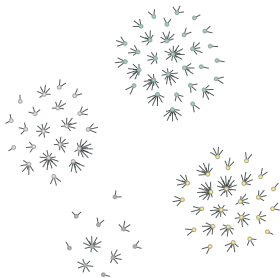
DEGREE-CORRECTED MICROCANONICAL SBM



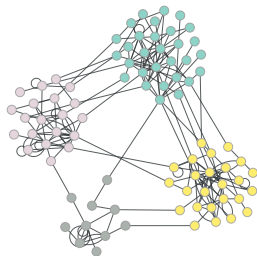
Node partition, \mathbf{b} .



Edge counts between groups, \mathbf{e} .



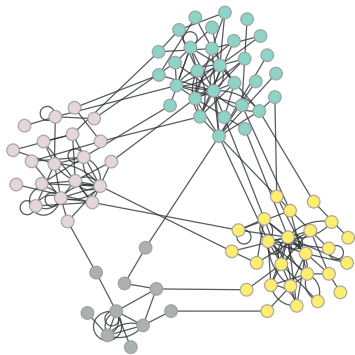
Degree sequence, \mathbf{k} .



Adjacency matrix, \mathbf{A} .

MODEL LIKELIHOOD: HALF-EDGE PAIRINGS

CONFIGURATION MODEL WITH GROUP STRUCTURE

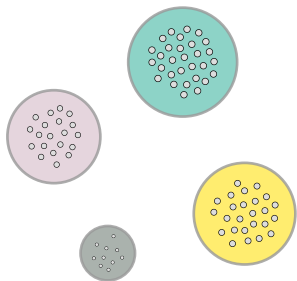


$$P(\mathbf{A}|e, \mathbf{k}, \mathbf{b}) = \frac{\Xi(\mathbf{A})}{\Omega(e)}$$

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

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

PRIOR FOR THE PARTITION

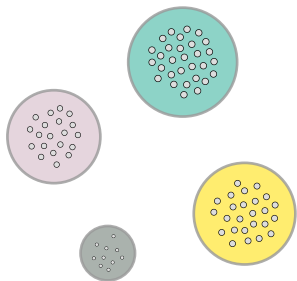


PRIOR FOR THE PARTITION

Option 1: Noninformative

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

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



PRIOR FOR THE PARTITION

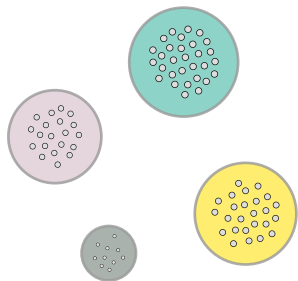
Option 1: Noninformative

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

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

Option 2: Conditioned on group sizes

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



PRIOR FOR THE PARTITION

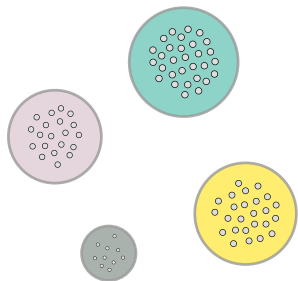
Option 1: Noninformative

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

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

Option 2: Conditioned on group sizes

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



For $N \gg B$:

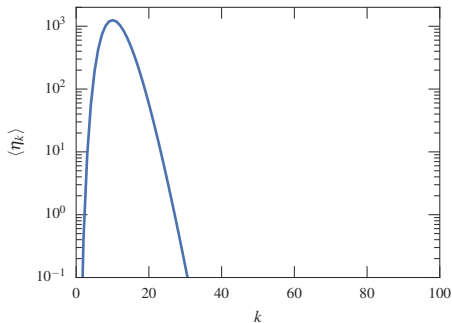
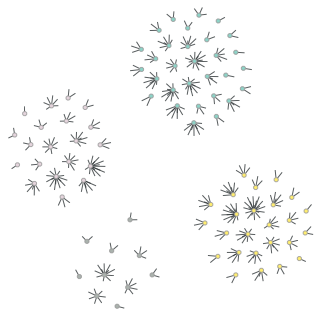
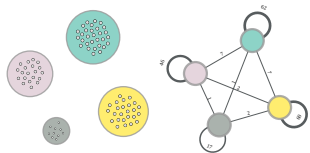
$$-\ln P(\mathbf{b}|B) \approx NH(\mathbf{n}) + O(B \ln N)$$

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

PRIOR FOR THE DEGREES

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

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



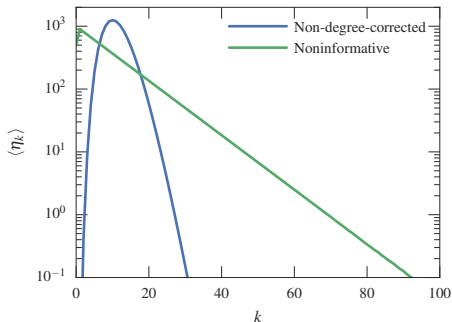
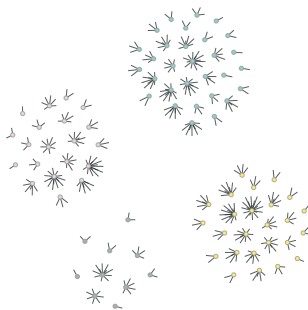
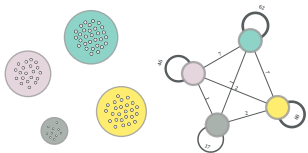
PRIOR FOR THE DEGREES

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

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

Option 1: Noninformative

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



PRIOR FOR THE DEGREES

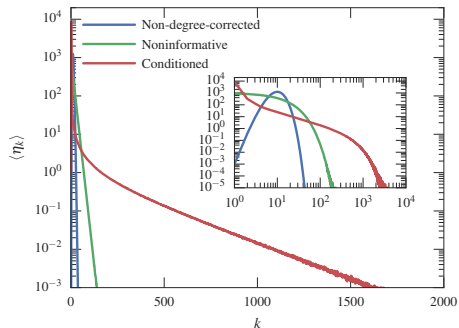
Option 2: Conditioned on degree distribution

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

PRIOR FOR THE DEGREES

Option 2: Conditioned on degree distribution

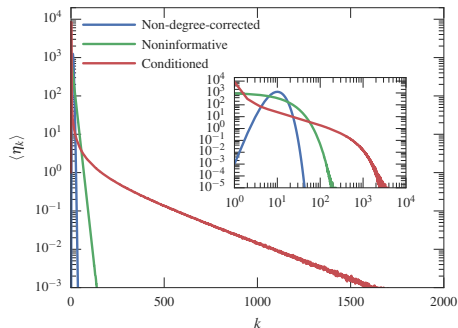
$$\begin{aligned} P(\mathbf{k}|\mathbf{b}, \mathbf{e}) &= P(\mathbf{k}|\boldsymbol{\eta})P(\boldsymbol{\eta}) \\ &= \prod_r \frac{n_r!}{\prod_r \eta_k^r!} \times \prod_r q(e_r, n_r)^{-1} \end{aligned}$$



PRIOR FOR THE DEGREES

Option 2: Conditioned on degree distribution

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



Bose-Einstein statistics

N “bosons” (nodes)

$k_i \rightarrow$ “energy level” (degree)

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

$N \gg 1, E \sim N$

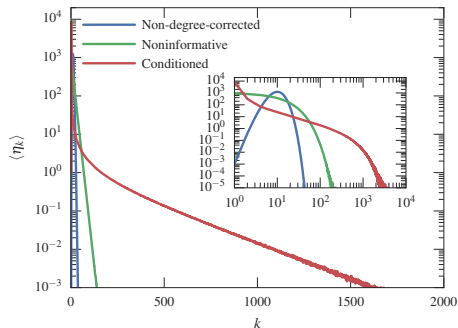
$\mu \rightarrow 1/\sqrt{E}, \lambda \rightarrow 0$

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

PRIOR FOR THE DEGREES

Option 2: Conditioned on degree distribution

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



Bose-Einstein statistics

N “bosons” (nodes)

$k_i \rightarrow$ “energy level” (degree)

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

$N \gg 1, E \sim N$

$\mu \rightarrow 1/\sqrt{E}, \lambda \rightarrow 0$

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

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

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

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

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

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

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

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

(Szekeres, 1951)

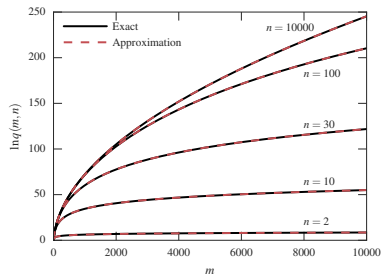
PRIOR FOR THE DEGREES: INTEGER PARTITIONS

$q(m, n) \rightarrow$ 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 $n \leq m \leq M$ in time
 $O(M^2)$.



Approximation

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

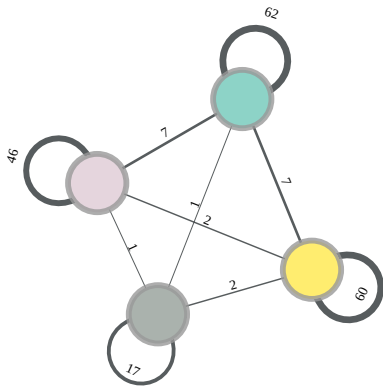
$$f(u) = \frac{v(u)}{2^{3/2}\pi u} \left[1 - (1 + u^2/2)e^{-v(u)} \right]^{-1/2}$$

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

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

(Szekeres, 1951)

PRIOR FOR THE EDGE COUNTS

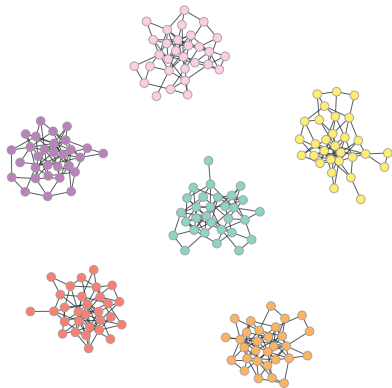


Option 1: Noninformative

$$P(\mathbf{e}) = \left(\left(\begin{pmatrix} B \\ 2 \\ E \end{pmatrix} \right) \right)^{-1}$$

$$\binom{n}{m} = \binom{n+m-1}{m}$$

“RESOLUTION” PROBLEM



Planted partition

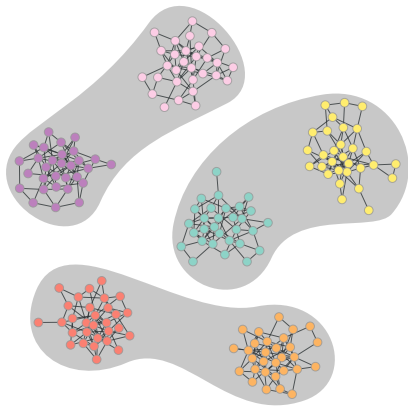
B groups of uniform size N/B ,
perfectly assortative:

$$e_{rr} = 2E/B, e_{rs} = 0, r \neq s.$$

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

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

“RESOLUTION” PROBLEM



Planted partition

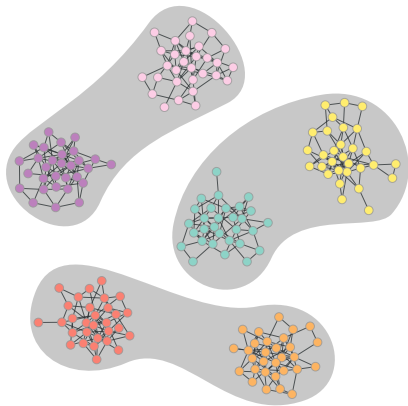
B groups of uniform size N/B ,
perfectly assortative:

$$e_{rr} = 2E/B, e_{rs} = 0, r \neq s.$$

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

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

“RESOLUTION” PROBLEM



Planted partition

B groups of uniform size N/B ,
perfectly assortative:

$$e_{rr} = 2E/B, e_{rs} = 0, r \neq s.$$

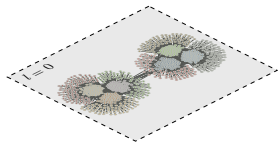
For $N \gg 1, E \sim N, B \gg 1$

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

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

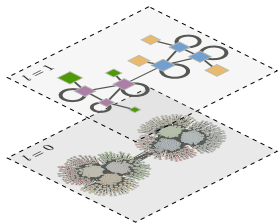
PRIOR FOR THE EDGE COUNTS: GROUP HIERARCHIES

Option 2: Conditioned on... another SBM!



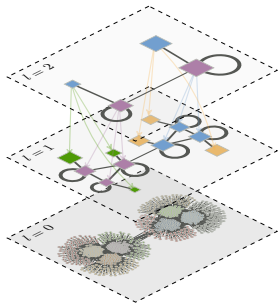
PRIOR FOR THE EDGE COUNTS: GROUP HIERARCHIES

Option 2: Conditioned on... another SBM!



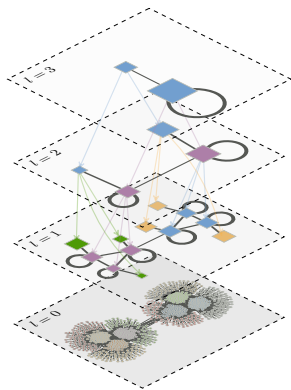
PRIOR FOR THE EDGE COUNTS: GROUP HIERARCHIES

Option 2: Conditioned on... another SBM!



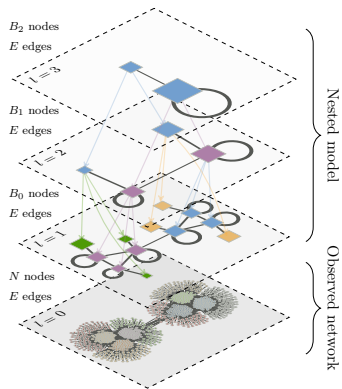
PRIOR FOR THE EDGE COUNTS: GROUP HIERARCHIES

Option 2: Conditioned on... another SBM!

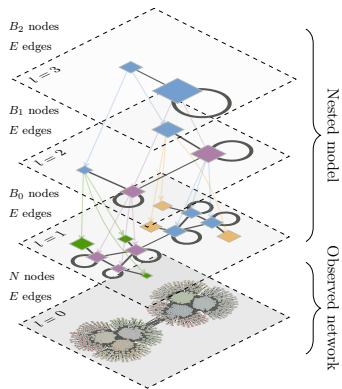


PRIOR FOR THE EDGE COUNTS: GROUP HIERARCHIES

Option 2: Conditioned on... another SBM!



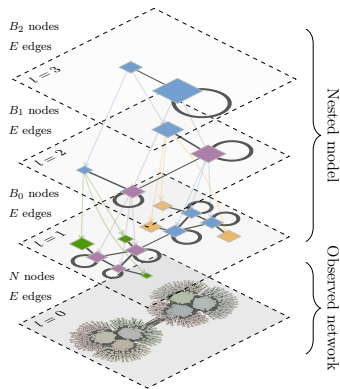
PRIOR FOR THE EDGE COUNTS: GROUP HIERARCHIES



Option 2: Conditioned on... another SBM!

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

PRIOR FOR THE EDGE COUNTS: GROUP HIERARCHIES



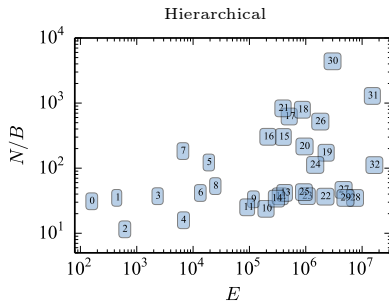
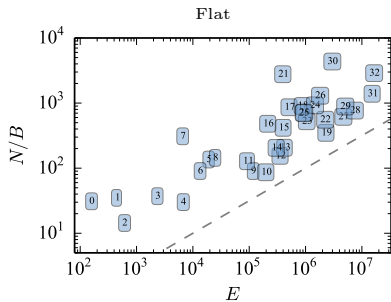
Option 2: Conditioned on... another SBM!

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

For planted partition:

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

RESOLUTION PROBLEM IN THE WILD



THE WHOLE MODEL

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

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

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

INFERENCE ALGORITHM: METROPOLIS-HASTINGS

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

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

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 .

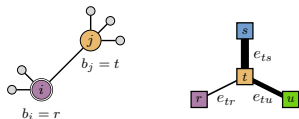
In contrast to:

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

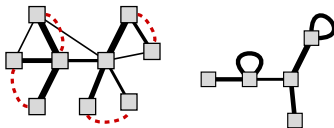
- ▶ Choose a random vertex i (happens to belong to group r).
- ▶ Move it to a random group $s \in [1, B]$, chosen with a probability $p(r \rightarrow s|t)$ proportional to $e_{ts} + \epsilon$, where t is the group membership of a randomly chosen neighbour of i .



Fast mixing times.

Agglomerative initialization

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



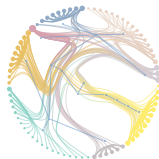
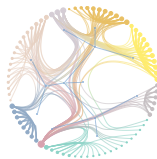
Avoids metastable states.

SAMPLING HIERARCHIES

$$\begin{aligned}P(\{\mathbf{b}_l\}|\mathbf{A}) &= \frac{P(\mathbf{A}, \mathbf{k}, \mathbf{b}_0|e_1) \prod_l P(e_{l-1}, \mathbf{b}_l|e_l)}{P(\mathbf{A})} \\ &= P(\mathbf{b}_0|\mathbf{A}, e_1, \mathbf{k}) \prod_l P(\mathbf{b}_l|e_{l-1}, e_l)\end{aligned}$$

with per-level posteriors

$$\begin{aligned}P(\mathbf{b}_0|\mathbf{A}, \mathbf{k}, e_1) &= \frac{P(\mathbf{A}|e_1, \mathbf{k}, \mathbf{b}_0)P(\mathbf{k})P(\mathbf{b}_0)}{P(\mathbf{A}|e_1)} \\ P(\mathbf{b}_l|e_{l-1}, e_l) &= \frac{P(e_l|e_{l+1}, \mathbf{b}_l)P(\mathbf{b}_l)}{P(e_l|e_{l+1})}.\end{aligned}$$



SAMPLING VS. MAXIMIZATION

Maximization (MDL)

$$\{\hat{\mathbf{b}}_l\} = \operatorname{argmax}_{\{\mathbf{b}_l\}} P(\{\mathbf{b}_l\}|\mathbf{A})$$

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

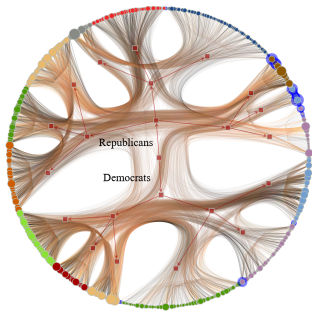
Sampling

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

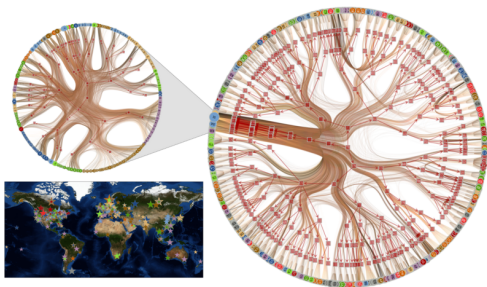
Finds the best posterior ensemble; marginal posterior $P(B_l|\mathbf{A})$.

SOME RESULTS FOR MAXIMIZATION

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

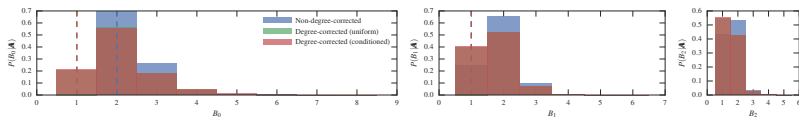


Internet (AS)
($N = 52,104$, $E = 399,625$)

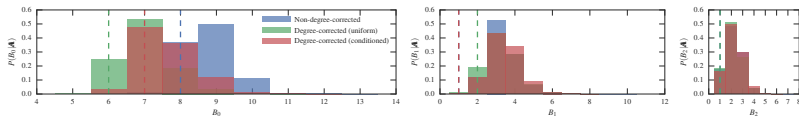


SAMPLING VS. MAXIMIZATION

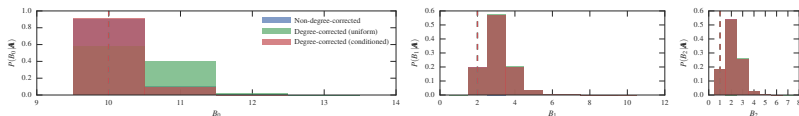
Karate Club ($N = 34, E = 78$)



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

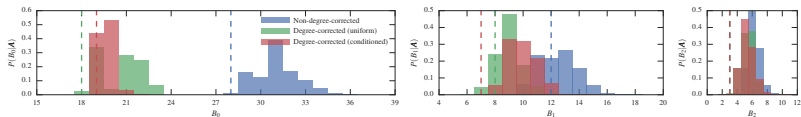


American Football ($N = 115, E = 613$)

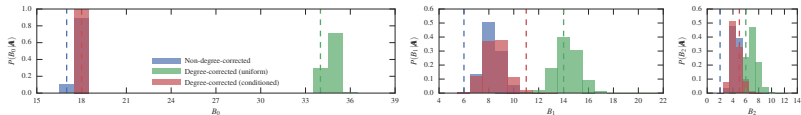


SAMPLING VS. MAXIMIZATION

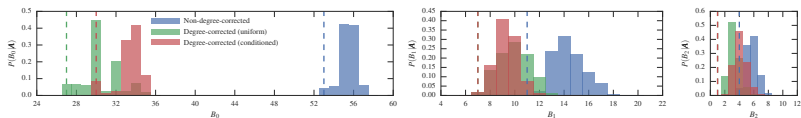
C. elegans neural network ($N = 297, E = 2,359$)



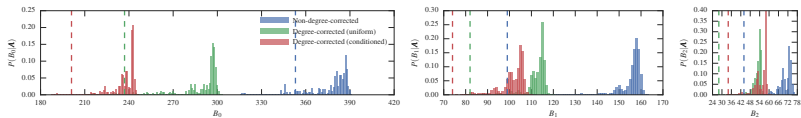
Power grid ($N = 4,941, E = 6,594$)



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

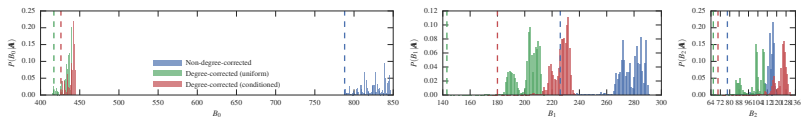


Airports ($N = 3,286, E = 68,344$)

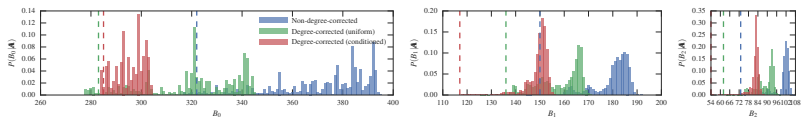


SAMPLING VS. MAXIMIZATION

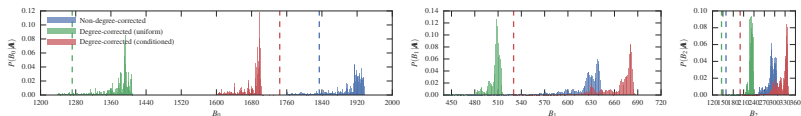
Linux source code ($N = 30,837, E = 213,954$)



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

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

MODEL SELECTION

1. Best model fit

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

2. Average over models

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

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

MODEL SELECTION

1. Best model fit

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

2. Average over models

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

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

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

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

MODEL SELECTION

1. Best model fit

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

2. Average over models

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

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

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

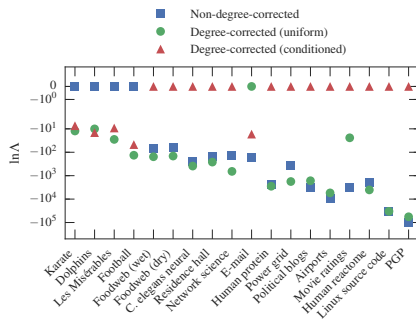
$$q(\{\mathbf{b}_l\}) = P(\{\mathbf{b}_l\} | \mathbf{A}) = \frac{P(\{\mathbf{b}_l\}, \mathbf{A})}{P(\mathbf{A})}$$

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

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

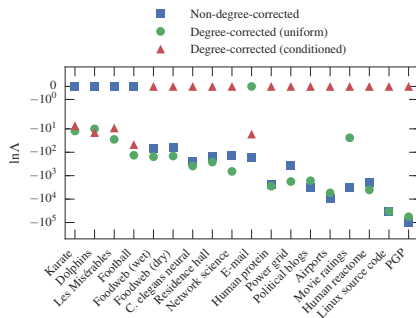
MODEL SELECTION

1. Best model fit (MDL)

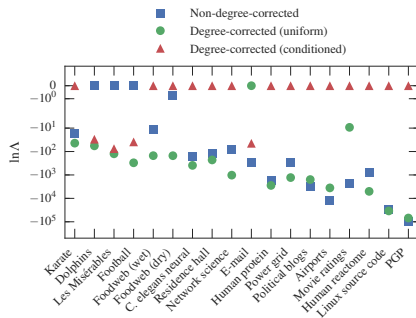


MODEL SELECTION

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)