Statistical inference of generative network models

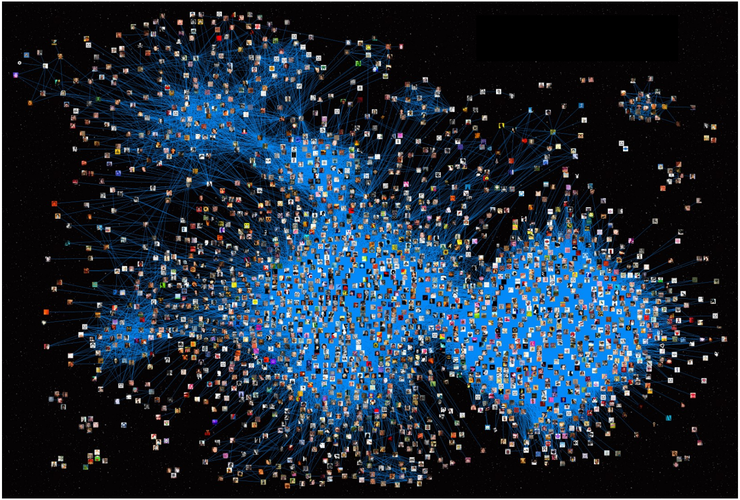
Tiago P. Peixoto

Universität Bremen
Germany

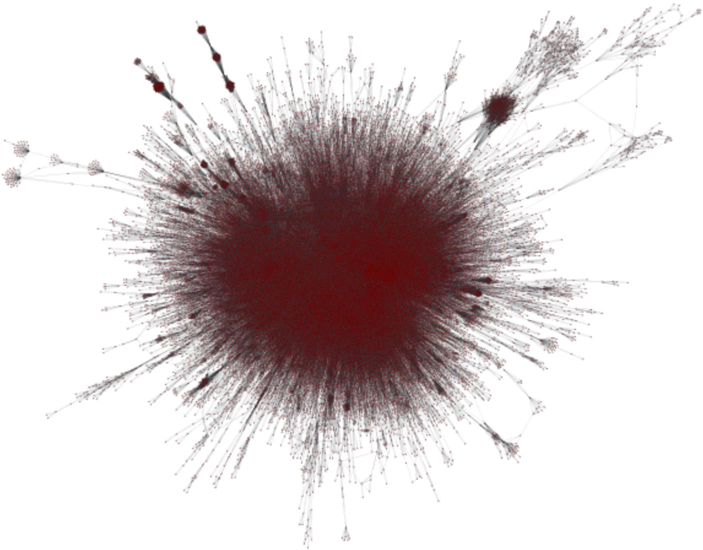
ISI Foundation
Turin, Italy

Como, May 2016

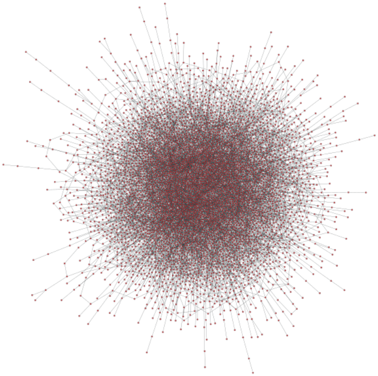
Large-scale network structure



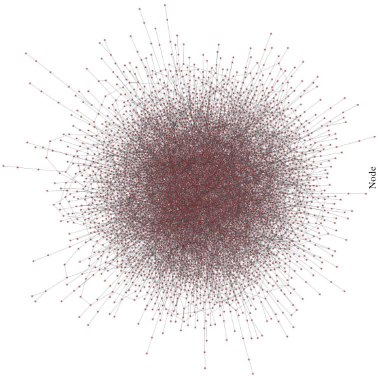
Problem: How to detect and characterize modular structure?

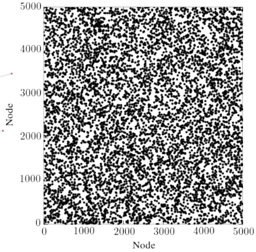


Structure vs. Noise

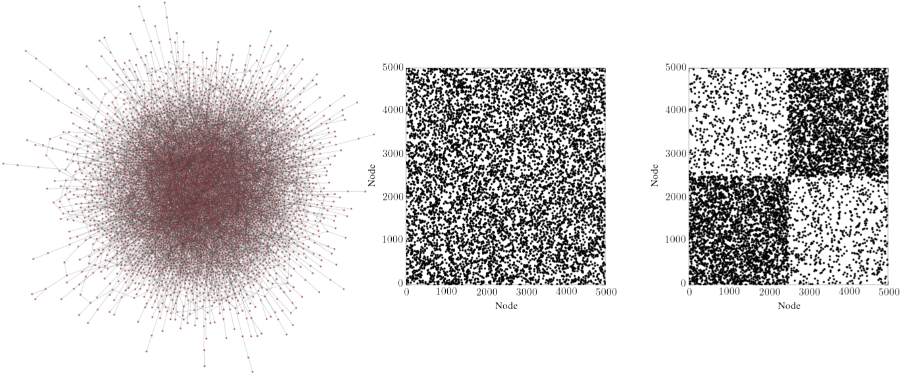


Structure vs. Noise

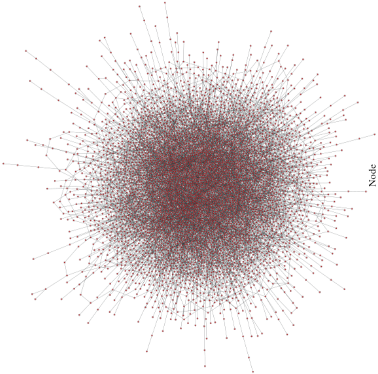


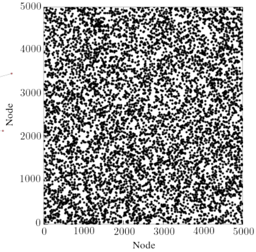


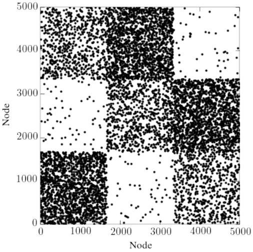
Structure vs. Noise



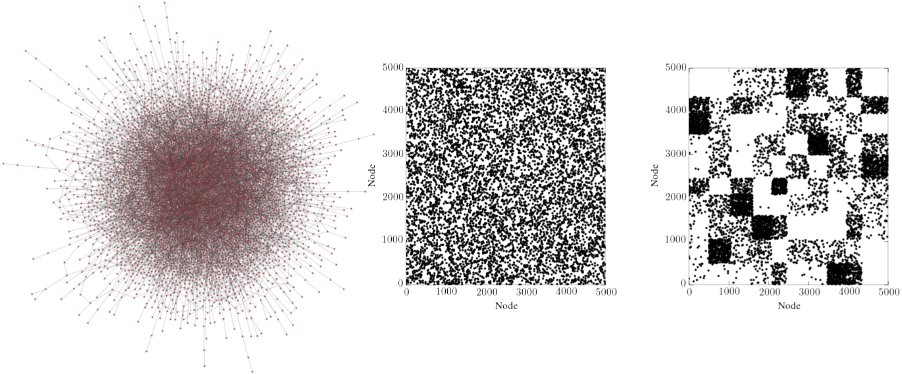
Structure vs. Noise



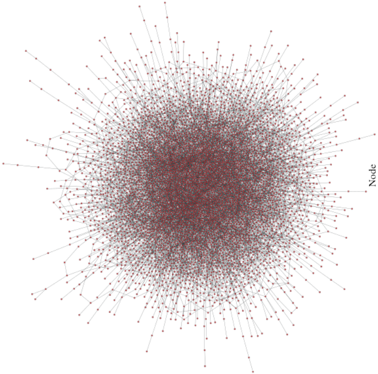


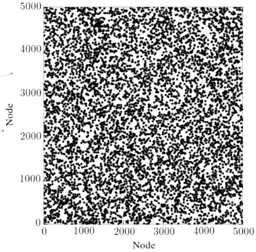


Structure vs. Noise

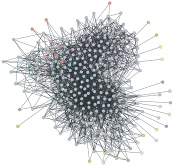


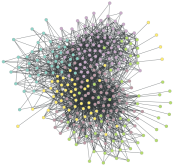
Structure vs. Noise

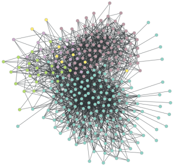


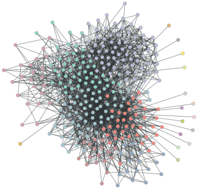


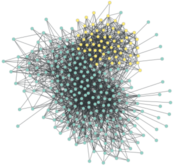
Different methods, different results...

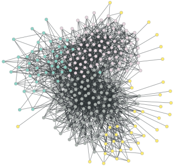

Betweenness

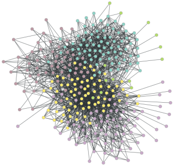

Modularity matrix

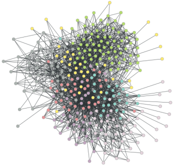

Infomap

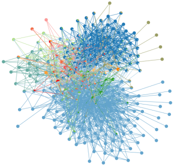

Walk Trap

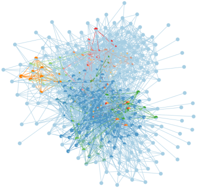

Label propagation

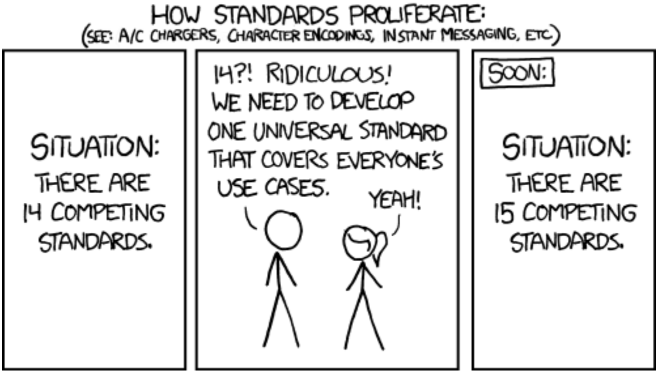

Modularity


Modularity (Blondel)


Spin glass


Infomap (overlapping)


Clique percolation



(XKCD, Randall Munroe)

Are we stuck on this?

A principled approach: Generative models

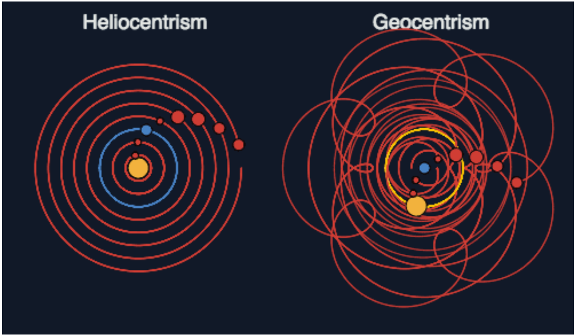
Before we devise algorithms, we must formulate *probabilistic models* for the formation of network structure.

- The parameters of the model describe the network structure.
- The actual values of the parameters are unknown beforehand, but can be inferred from data, using robust principles from statistics.

Ultimately, we seek to:

- Reduce the complexity of network data.
- Access the statistical significance of the results, and thus *differentiate structure from noise*.
- Compare different models as alternative hypotheses.
- Generalize from data and make predictions.

Nothing more than traditional science...



How do we model networks?

EXPONENTIAL RANDOM GRAPH MODEL (ERGM)

$G \rightarrow$ Network

$m(G) \rightarrow$ Observable (e.g. clustering, community structure, etc.)

We want a probabilistic model: $P(G)$

Maximum entropy principle

Ensemble entropy:

$$S = -\sum_G P(G) \ln P(G)$$

Maximize S conditioned on $\langle m \rangle = m^*$

$$\langle m \rangle = \frac{1}{N} \sum_G m(G) P(G)$$

Exponential random graph model (ERGM)

$G \to$ Network

$m(G) \to$ Observable (e.g. clustering, community structure, etc.)

We want a probabilistic model: $P(G)$

Maximum entropy principle

Ensemble entropy:

$$S = -\sum_G P(G) \ln P(G)$$

Maximize S conditioned on $\langle m \rangle = m^*$

$$\langle m \rangle = \frac{1}{N} \sum_G m(G) P(G)$$

Lagrange multipliers:

$$\Lambda = -\sum_G P(G) \ln P(G) - \theta \left(\sum_G m(G) P(G) - N m^* \right)$$

$$\frac{\partial \Lambda}{\partial P(G)} = 0, \quad \frac{\partial \Lambda}{\partial \theta} = 0$$

$$P(G|\theta) = \frac{e^{-\theta m(G)}}{Z(\theta)} \quad Z(\theta) = \sum_G e^{-\theta m(G)}$$

In Physics:

ERGM $\to$ Gibbs ensemble

$m(G) \to$ Hamiltonian

$\theta \to$ Inverse temperature

$Z(\theta) \to$ Partition function

ERGM WITH COMMUNITY STRUCTURE

N nodes divided into B groups.

Node partition, $b_i \in [1, B]$

Observables: Number of edges between groups r and s,
$e_{rs} = \sum_{ij} A_{ij}\delta_{b_i,r}\delta_{b_j,s}$

$$P(G|\theta) = \frac{\exp\left(-\sum_{r\leq s}\theta_{rs}e_{rs}\right)}{Z(\theta)} = \prod_{i<j}\frac{e^{-A_{ij}\theta_{b_ib_j}}}{e^{-\theta_{b_ib_j}}+1}$$

$$p_{rs} = \frac{1}{e^{\theta_{rs}}+1}$$

$$\boxed{P(G|b,\theta) = \prod_{i<j} p_{b_ib_j}^{A_{ij}}(1-p_{b_ib_j})^{1-A_{ij}}}$$

$p_{rs} \rightarrow$ Probability of an edge existing between two nodes of groups r and s.

$$\langle e_{rs}\rangle = n_r n_s p_{rs}$$

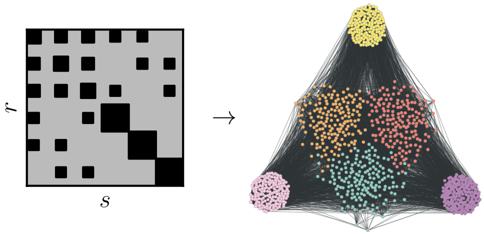
a.k.a. the stochastic block model (SBM)

THE STOCHASTIC BLOCK MODEL (SBM)

P. W. HOLLAND ET AL., SOC NETWORKS 5, 109 (1983)

Large-scale modules: N nodes divided into B groups.

Parameters: $b_i \to$ group membership of node i
$p_{rs} \to$ probability of an edge between nodes of groups r and s.



Properties:

- *General* model for large-scale structures. Traditional assortative communities are a special case, but it also admits arbitrary mixing patterns (e.g. bipartite, core-periphery, etc).
- Formulation for directed graphs is trivial.
- The meaning of "communities" or "groups" is well defined in the model.

BASIC SBM VARIATIONS

Bernoulli (simple graphs)

$$P(G|b,\theta) = \prod_{i<j} p_{b_ib_j}^{A_{ij}} (1-p_{b_ib_j})^{1-A_{ij}}$$

(Holland et al, 1983)

Poisson (multigraphs)

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

(Karrer and Newman, 2011)

Microcanonical

$$P(G|b,\{e_{rs}\}) = \frac{1}{\Omega(\{e_{rs}\},\{n_r\})}$$

Simple graphs

$$\Omega(\{e_{rs}\},\{n_r\}) = \prod_{r<s} \binom{n_rn_s}{e_{rs}} \prod_r \binom{\binom{n_r}{2}}{e_{rr}/2}$$

Multigraphs

$$\Omega(\{e_{rs}\},\{n_r\}) = \prod_{r<s} \left(\!\!\binom{n_rn_s}{e_{rs}}\!\!\right) \prod_r \left(\!\!\binom{\left(\!\binom{n_r}{2}\!\right)}{e_{rr}/2}\!\!\right)$$

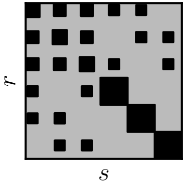
(Peixoto, 2012)

PARAMETRIC INFERENCE VIA MAXIMUM LIKELIHOOD

Data likelihood: $\mathcal{P}(G|\theta)$

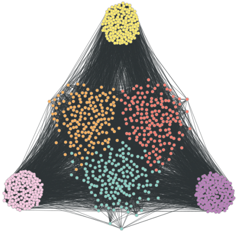
$G \rightarrow$ Observed network

$\theta \rightarrow$ Model parameters: $\{p_{rs}\}, \{b_i\}$



θ

$\xrightarrow[P(G|\theta)]{}$



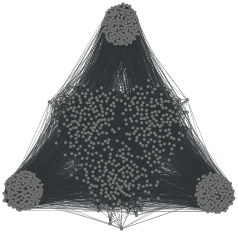
G

Parametric inference via maximum likelihood

Data likelihood: $\mathcal{P}(G|\theta)$

$G \rightarrow$ Observed network

$\theta \rightarrow$ Model parameters: $\{p_{rs}\}, \{b_i\}$



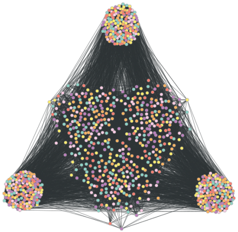
G

Parametric inference via maximum likelihood

Data likelihood: $\mathcal{P}(G|\theta)$

$G \rightarrow$ Observed network

$\theta \rightarrow$ Model parameters: $\{p_{rs}\}, \{b_i\}$



G

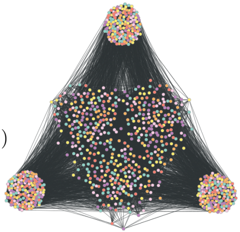
Parametric inference via maximum likelihood

Data likelihood: $\mathcal{P}(G|\theta)$

$G \rightarrow$ Observed network

$\theta \rightarrow$ Model parameters: $\{p_{rs}\}, \{b_i\}$

$\underset{\theta}{\operatorname{argmax}} P(G|\theta) \xleftarrow{}$



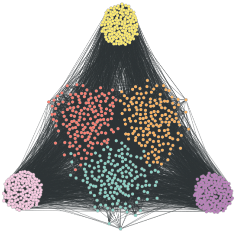
G

Parametric inference via maximum likelihood

Data likelihood: $\mathcal{P}(G|\theta)$

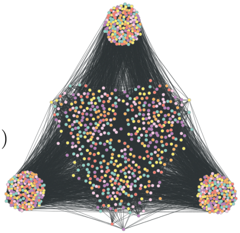
$G \rightarrow$ Observed network

$\theta \rightarrow$ Model parameters: $\{p_{rs}\}, \{b_i\}$



$\hat{\theta}$

$$\xleftarrow{} \underset{\theta}{\operatorname{argmax}} P(G|\theta)$$



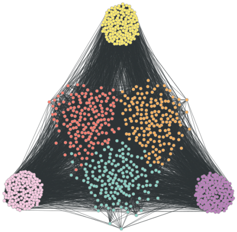
G

Parametric inference via maximum likelihood

Data likelihood: $\mathcal{P}(G|\theta)$

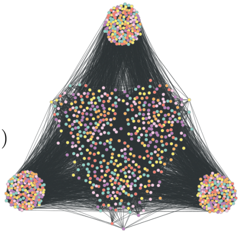
$G \rightarrow$ Observed network

$\theta \rightarrow$ Model parameters: $\{p_{rs}\}, \{b_i\}$



$\hat{\theta}$

$$\overset{\longleftarrow}{\underset{\theta}{\operatorname{argmax}} P(G|\theta)}$$



G

Efficient MCMC Inference algorithm

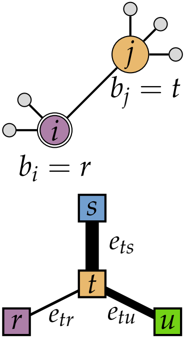
T. P. P., Phys. Rev. E 89, 012804 (2014)

Idea: Use the currently-inferred structure to guess the best next move.

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

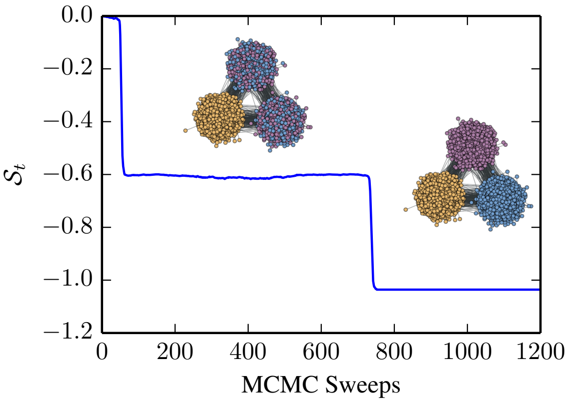
$$a = \min\left\{e^{-\beta\Delta\mathcal{S}}\frac{\sum_t p_t^i p(s \to r|t)}{\sum_t p_t^i p(r \to s|t)}, 1\right\}.$$

- Repeat.



EFFICIENT MCMC INFERENCE ALGORITHM

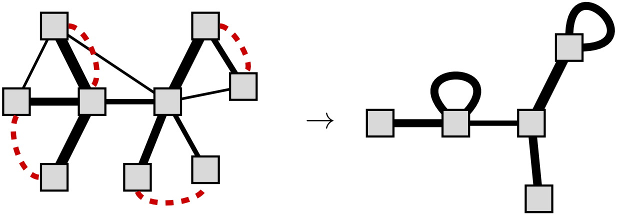
One remaining problem: Metastable states...



T.P.P., Phys. Rev. E 89, 012804 (2014)

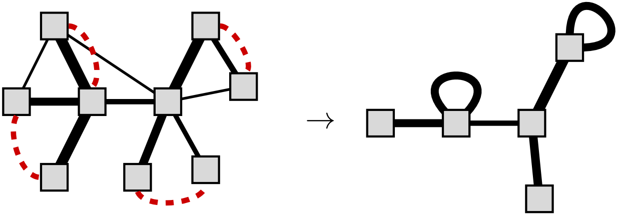
Efficient MCMC Inference algorithm

Solution: Agglomerative clustering.

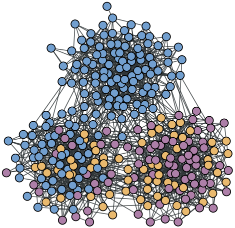


Efficient MCMC Inference algorithm

Solution: Agglomerative clustering.

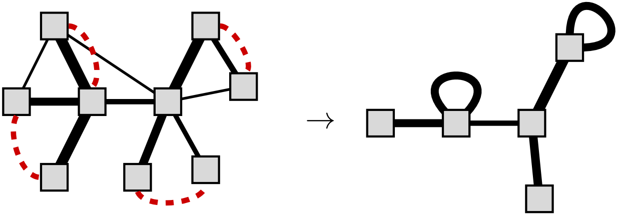


By making $\beta \to \infty$ we get a very reliable greedy heuristic!

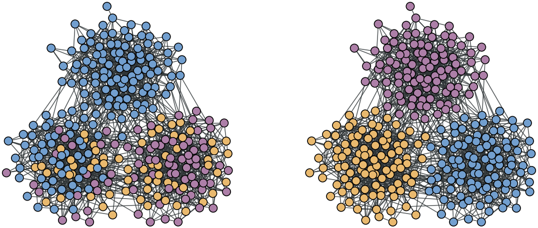


Efficient MCMC Inference algorithm

Solution: Agglomerative clustering.



By making $\beta \to \infty$ we get a very reliable greedy heuristic!



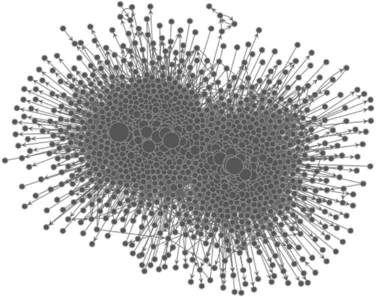
Algorithmic complexity: $O(N \ln^2 N)$ (independent of B)

Inference $\neq$ Bed of roses

There are some caveats...

Problem 1: Broad degree distributions

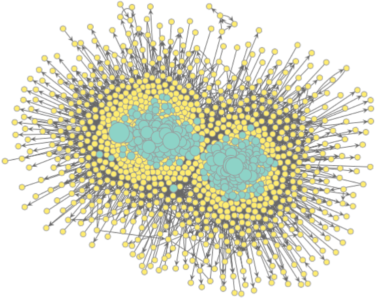
Brian Karrer and M. E. J. Newman, Phys. Rev. E 83, 016107, 2011



Political Blogs (Adamic and Glance, 2005)

Problem 1: Broad degree distributions

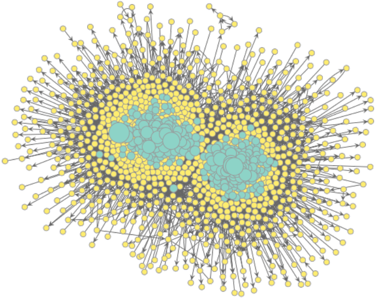
Brian Karrer and M. E. J. Newman, Phys. Rev. E 83, 016107, 2011



Political Blogs (Adamic and Glance, 2005)

Problem 1: Broad degree distributions

Brian Karrer and M. E. J. Newman, Phys. Rev. E 83, 016107, 2011



Political Blogs (Adamic and Glance, 2005)

Degree-corrected SBM

$$p_{ij} = \theta_i \theta_j \lambda_{b_i, b_j}$$

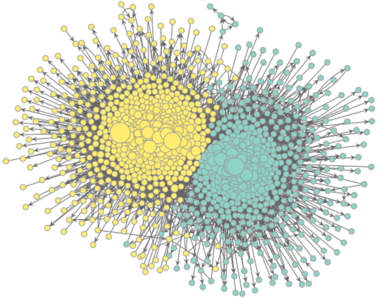
$\lambda_{rs} \to$ edges between groups
$\theta_i \to$ propensity of node i

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

$$\hat{\theta}_i = \frac{k_i}{e_{b_i}} \qquad \hat{\lambda}_{rs} = e_{rs}$$

Problem 1: Broad degree distributions

Brian Karrer and M. E. J. Newman, Phys. Rev. E 83, 016107, 2011



Political Blogs (Adamic and Glance, 2005)

Degree-corrected SBM

$$p_{ij} = \theta_i \theta_j \lambda_{b_i, b_j}$$

$\lambda_{rs} \to$ edges between groups
$\theta_i \to$ propensity of node i

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

$$\hat{\theta}_i = \frac{k_i}{e_{b_i}} \qquad \hat{\lambda}_{rs} = e_{rs}$$

(BIG) PROBLEM 2: OVERFITTING

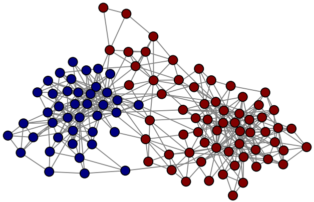
What if we don't know the number of groups, B?

$$\mathcal{S} = -\log_2 P(G|\{b_i\}, \{e_{rs}\}, \{k_i\})$$

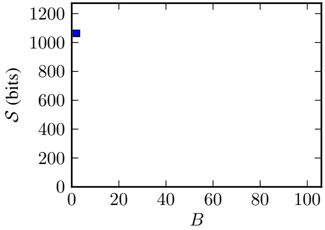
(BIG) PROBLEM 2: OVERFITTING

What if we don't know the number of groups, B?

$$\mathcal{S} = -\log_2 P(G|\{b_i\}, \{e_{rs}\}, \{k_i\})$$



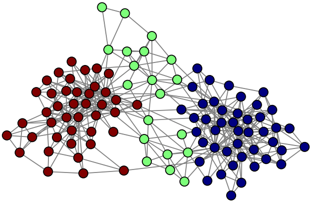
$B = 2$



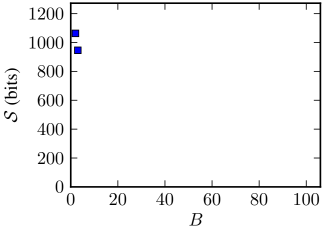
(BIG) PROBLEM 2: OVERFITTING

What if we don't know the number of groups, B?

$$\mathcal{S} = -\log_2 P(G|\{b_i\},\{e_{rs}\},\{k_i\})$$



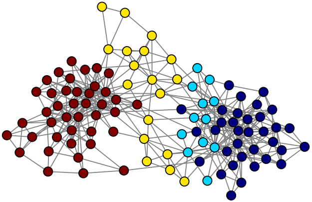
$B = 3$



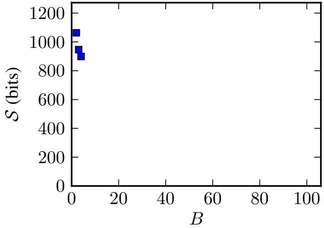
(BIG) PROBLEM 2: OVERFITTING

What if we don't know the number of groups, B?

$$\mathcal{S} = -\log_2 P(G|\{b_i\}, \{e_{rs}\}, \{k_i\})$$



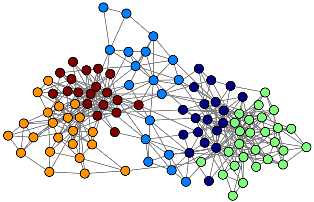
$B = 4$



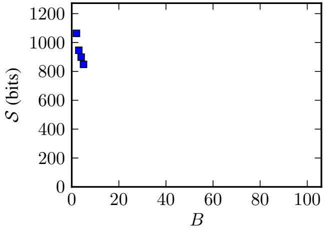
(BIG) PROBLEM 2: OVERFITTING

What if we don't know the number of groups, B?

$$\mathcal{S} = -\log_2 P(G|\{b_i\}, \{e_{rs}\}, \{k_i\})$$



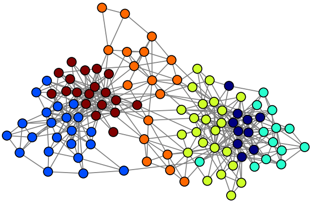
$B = 5$



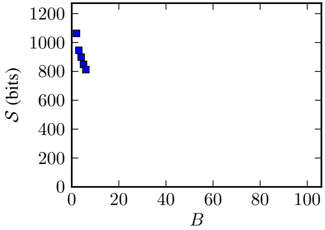
(BIG) PROBLEM 2: OVERFITTING

What if we don't know the number of groups, B?

$$\mathcal{S} = -\log_2 P(G|\{b_i\}, \{e_{rs}\}, \{k_i\})$$



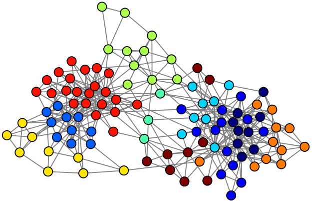
$B = 6$



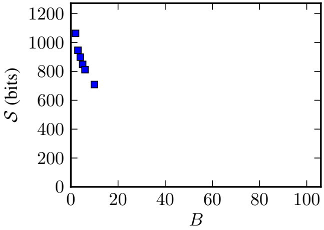
(BIG) PROBLEM 2: OVERFITTING

What if we don't know the number of groups, B?

$$\mathcal{S} = -\log_2 P(G|\{b_i\}, \{e_{rs}\}, \{k_i\})$$



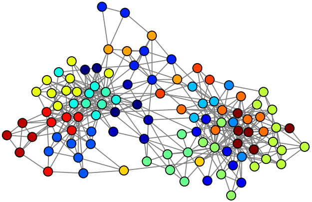
$B = 10$



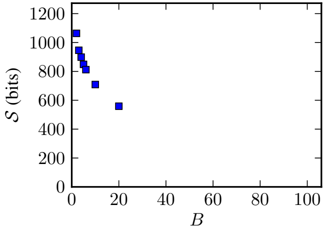
(BIG) PROBLEM 2: OVERFITTING

What if we don't know the number of groups, B?

$$\mathcal{S} = -\log_2 P(G|\{b_i\}, \{e_{rs}\}, \{k_i\})$$



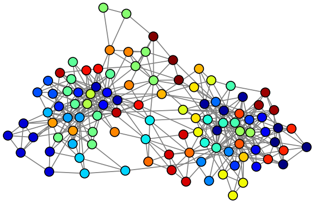
$B = 20$



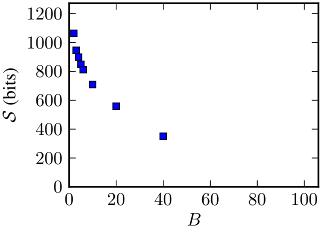
(BIG) PROBLEM 2: OVERFITTING

What if we don't know the number of groups, B?

$$\mathcal{S} = -\log_2 P(G|\{b_i\},\{e_{rs}\},\{k_i\})$$



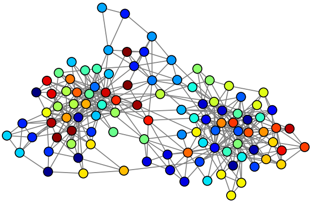
$B = 40$



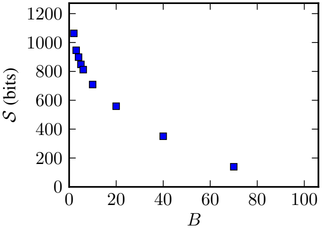
(BIG) PROBLEM 2: OVERFITTING

What if we don't know the number of groups, B?

$$\mathcal{S} = -\log_2 P(G|\{b_i\},\{e_{rs}\},\{k_i\})$$



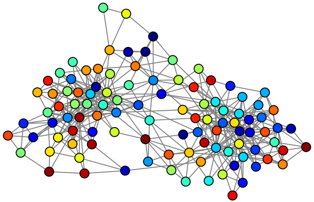
$B = 70$



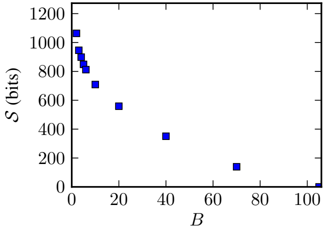
(BIG) PROBLEM 2: OVERFITTING

What if we don't know the number of groups, B?

$$\mathcal{S} = -\log_2 P(G|\{b_i\},\{e_{rs}\},\{k_i\})$$



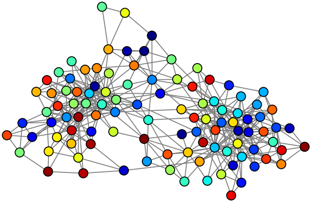
$B = N$



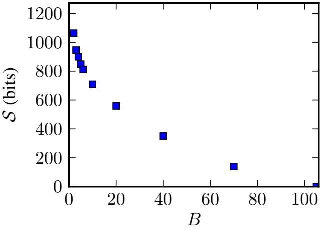
(BIG) PROBLEM 2: OVERFITTING

What if we don't know the number of groups, B?

$$\mathcal{S} = -\log_2 P(G|\{b_i\}, \{e_{rs}\}, \{k_i\})$$



$B = N$



... in other words: **Overfitting!**

Nonparametric Bayesian inference

Bayes' rule

$$\boxed{P(\theta|G) = \frac{P(G|\theta)P(\theta)}{P(G)}}$$

$$P(G|\theta) \rightarrow \text{Data likelihood}$$
$$P(\theta) \rightarrow \text{Prior}$$
$$P(\theta|G) \rightarrow \text{Posterior}$$

$$P(G) = \sum_\theta P(G|\theta)P(\theta) \rightarrow \text{Model evidence}$$

THE MINIMUM DESCRIPTION LENGTH PRINCIPLE (MDL)

Bayesian formulation

Posterior: $\mathcal{P}(\theta|G) = \dfrac{\mathcal{P}(G|\theta)\mathcal{P}(\theta)}{P(G)}$

$G \rightarrow$ Observed network

$\theta \rightarrow$ Model parameters: $\{e_{rs}\}, \{b_i\}$

$\mathcal{P}(\theta) \rightarrow$ Prior on the parameters

Inference vs. compression

$$-\ln \mathcal{P}(\theta|G) = \underbrace{-\ln P(G|\theta)}_{\mathcal{S}} \underbrace{-\ln P(\theta)}_{\mathcal{L}} + \ln \mathcal{P}(G)$$

$\mathcal{S} \rightarrow$ Information required to describe the network, *when the model is known*.

$\mathcal{L} \rightarrow$ Information required to describe *the model*.

$\Sigma = \mathcal{S} + \mathcal{L}$

Description length
Total information necessary to describe the data.

Rissanen, J. Automatica 14 (5): 465–658. (1978)

MDL for the SBM

T. P. Peixoto, Phys. Rev. Lett. 110, 148701 (2013)

Microcanonical model

$$\theta = \{b_i\}, \{e_{rs}\}$$

Data description length

$$S = -\ln P(G|\{b_i\}, \{e_{rs}\}) = \sum_{rs} \ln \binom{n_r n_s}{e_{rs}} + \sum_r \binom{\binom{n_r}{2}}{e_{rs}/2}$$

Partition description length

$$P(\{b_i\}) = P(\{b_i\}|\{n_r\})P(\{n_r\}) \quad \mathcal{L}_p = -\ln P(\{b_i\}) = \left(\!\!\binom{B}{N}\!\!\right) + \ln N! - \sum_r n_r!$$

Edge description length

$$P(\{e_{rs}\}) = \frac{1}{\Omega(B, E)} \quad \mathcal{L}_e = -\ln P(\{e_{rs}\}) = \ln \left(\!\!\binom{\left(\!\binom{B}{2}\!\right)}{E}\!\!\right)$$

$$\boxed{\Sigma = S + \mathcal{L}_p + \mathcal{L}_e}$$

Minimum Description Length

$\mathcal{S} \rightarrow$ Information required to describe the network, *when the model is known*.

$\mathcal{L} \rightarrow$ Information required to describe *the model*.

$$\Sigma = \mathcal{S} + \mathcal{L}$$

Description length
Total information necessary to describe the data.

T. P. Peixoto, Phys. Rev. Lett. 110, 148701 (2013)

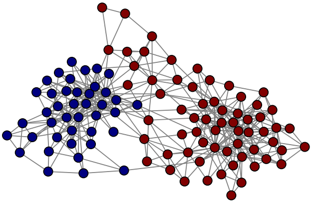
MINIMUM DESCRIPTION LENGTH

$\mathcal{S} \rightarrow$ Information required to describe the network, *when the model is known*.

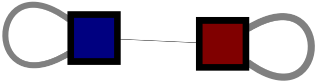
$\mathcal{L} \rightarrow$ Information required to describe *the model*.

$$\Sigma = \mathcal{S} + \mathcal{L}$$

Description length
Total information necessary to describe the data.



$B = 2, \mathcal{S} \simeq 1805.3$ bits



Model, $\mathcal{L} \simeq 122.6$ bits

$\Sigma \simeq 1927.9$ bits

T. P. Peixoto, Phys. Rev. Lett. 110, 148701 (2013)

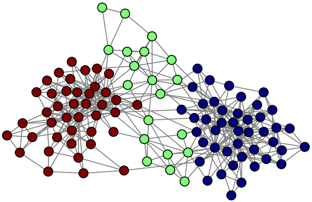
Minimum Description Length

$\Sigma = \mathcal{S} + \mathcal{L}$

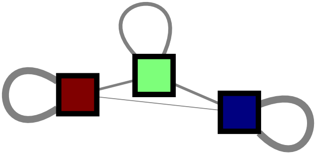
$\mathcal{S} \to$ Information required to describe the network, *when the model is known*.

$\mathcal{L} \to$ Information required to describe *the model*.

Description length
Total information necessary to describe the data.



$B = 3$, $\mathcal{S} \simeq 1688.1$ bits



Model, $\mathcal{L} \simeq 203.4$ bits

$\Sigma \simeq 1891.5$ bits

T. P. Peixoto, Phys. Rev. Lett. 110, 148701 (2013)

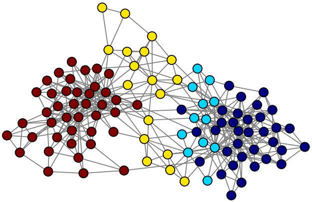
MINIMUM DESCRIPTION LENGTH

$\mathcal{S} \rightarrow$ Information required to describe the network, *when the model is known*.

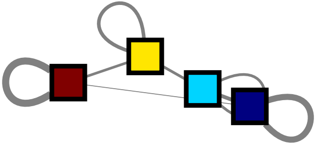
$\mathcal{L} \rightarrow$ Information required to describe *the model*.

$$\Sigma = \mathcal{S} + \mathcal{L}$$

Description length
Total information necessary to describe the data.



$B = 4$, $\mathcal{S} \simeq 1640.8$ bits



Model, $\mathcal{L} \simeq 270.7$ bits

$\Sigma \simeq 1911.5$ bits

T. P. Peixoto, Phys. Rev. Lett. 110, 148701 (2013)

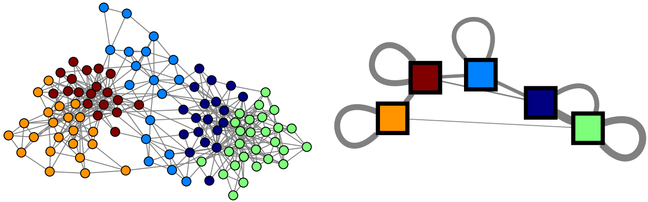
Minimum Description Length

$\Sigma = \mathcal{S} + \mathcal{L}$

$\mathcal{S} \rightarrow$ Information required to describe the network, *when the model is known*.

$\mathcal{L} \rightarrow$ Information required to describe *the model*.

Description length
Total information necessary to describe the data.



$B = 5$, $\mathcal{S} \simeq 1590.5$ bits

Model, $\mathcal{L} \simeq 330.8$ bits

$\Sigma \simeq 1921.3$ bits

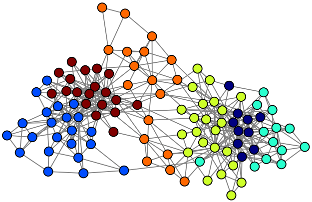
Minimum Description Length

$\Sigma = \mathcal{S} + \mathcal{L}$

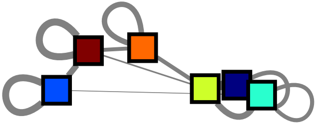
$\mathcal{S} \rightarrow$ Information required to describe the network, *when the model is known*.

$\mathcal{L} \rightarrow$ Information required to describe *the model*.

Description length
Total information necessary to describe the data.



$B = 6$, $\mathcal{S} \simeq 1554.2$ bits



Model, $\mathcal{L} \simeq 386.7$ bits

$\Sigma \simeq 1940.9$ bits

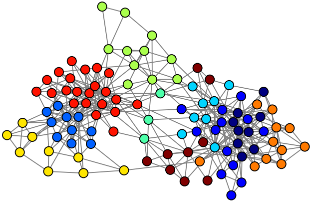
Minimum Description Length

$\Sigma = \mathcal{S} + \mathcal{L}$

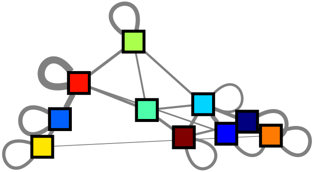
$\mathcal{S} \rightarrow$ Information required to describe the network, *when the model is known*.

$\mathcal{L} \rightarrow$ Information required to describe *the model*.

Description length
Total information necessary to describe the data.



$B = 10$, $\mathcal{S} \simeq 1451.0$ bits



Model, $\mathcal{L} \simeq 590.8$ bits

$\Sigma \simeq 2041.8$ bits

T. P. Peixoto, Phys. Rev. Lett. 110, 148701 (2013)

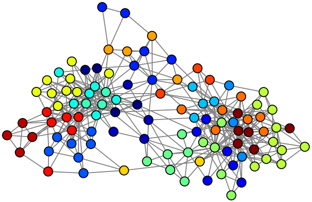
Minimum Description Length

$\Sigma = \mathcal{S} + \mathcal{L}$

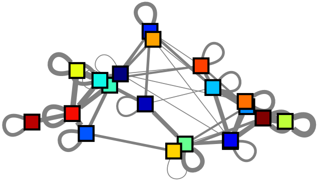
$\mathcal{S} \rightarrow$ Information required to describe the network, *when the model is known*.

$\mathcal{L} \rightarrow$ Information required to describe *the model*.

Description length
Total information necessary to describe the data.



$B = 20$, $\mathcal{S} \simeq 1300.7$ bits



Model, $\mathcal{L} \simeq 1037.8$ bits

$\Sigma \simeq 2338.6$ bits

T. P. Peixoto, Phys. Rev. Lett. 110, 148701 (2013)

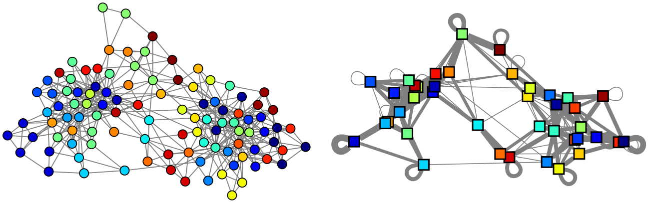
Minimum Description Length

$\mathcal{S} \rightarrow$ Information required to describe the network, *when the model is known*.

$\mathcal{L} \rightarrow$ Information required to describe *the model*.

$$\Sigma = \mathcal{S} + \mathcal{L}$$

Description length
Total information necessary to describe the data.



$B = 40$, $\mathcal{S} \simeq 1092.8$ bits   Model, $\mathcal{L} \simeq 1730.3$ bits

$\Sigma \simeq 2823.1$ bits

T. P. Peixoto, Phys. Rev. Lett. 110, 148701 (2013)

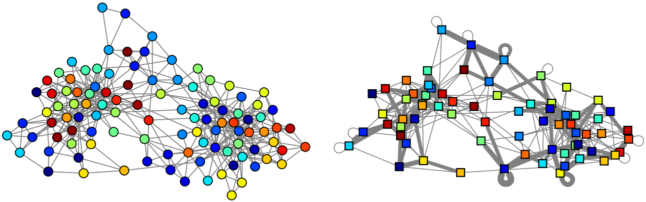
Minimum Description Length

$\mathcal{S} \rightarrow$ Information required to describe the network, *when the model is known*.

$\mathcal{L} \rightarrow$ Information required to describe *the model*.

$$\Sigma = \mathcal{S} + \mathcal{L}$$

Description length
Total information necessary to describe the data.



$B = 70$, $\mathcal{S} \simeq 881.3$ bits

Model, $\mathcal{L} \simeq 2427.3$ bits

$\Sigma \simeq 3308.6$ bits

T. P. Peixoto, Phys. Rev. Lett. 110, 148701 (2013)

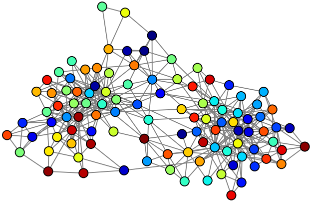
Minimum Description Length

$\mathcal{S} \rightarrow$ Information required to describe the network, *when the model is known*.

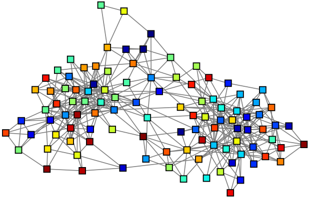
$\mathcal{L} \rightarrow$ Information required to describe *the model*.

$$\Sigma = \mathcal{S} + \mathcal{L}$$

Description length
Total information necessary to describe the data.



$B = N$, $\mathcal{S} = 0$ bits



Model, $\mathcal{L} \simeq 3714.9$ bits

$$\Sigma \simeq 3714.9 \text{ bits}$$

T. P. Peixoto, Phys. Rev. Lett. 110, 148701 (2013)

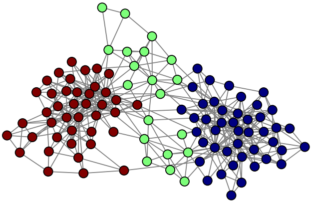
MINIMUM DESCRIPTION LENGTH

$\Sigma = \mathcal{S} + \mathcal{L}$

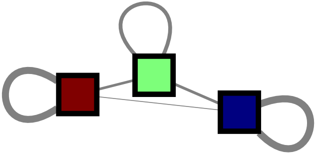
$\mathcal{S} \rightarrow$ Information required to describe the network, *when the model is known*.

$\mathcal{L} \rightarrow$ Information required to describe *the model*.

Description length
Total information necessary to describe the data.



$B = 3$, $\mathcal{S} \simeq 1688.1$ bits



Model, $\mathcal{L} \simeq 203.4$ bits

$\Sigma \simeq 1891.5$ bits

T. P. Peixoto, Phys. Rev. Lett. 110, 148701 (2013)

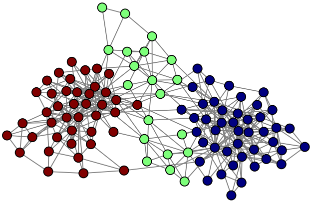
Minimum Description Length

$\Sigma = \mathcal{S} + \mathcal{L}$

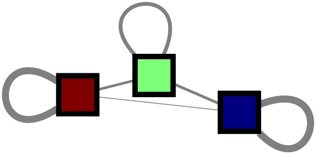
$\mathcal{S} \rightarrow$ Information required to describe the network, *when the model is known*.

$\mathcal{L} \rightarrow$ Information required to describe *the model*.

Description length
Total information necessary to describe the data.

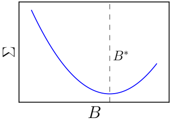


$B = 3$, $\mathcal{S} \simeq 1688.1$ bits



Model, $\mathcal{L} \simeq 203.4$ bits

$\Sigma \simeq 1891.5$ bits



$\rightarrow$

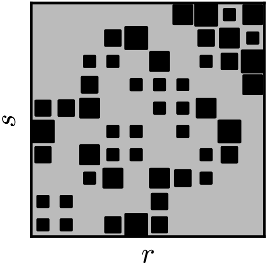
Occam's razor

The best model is the one which most compresses the data.

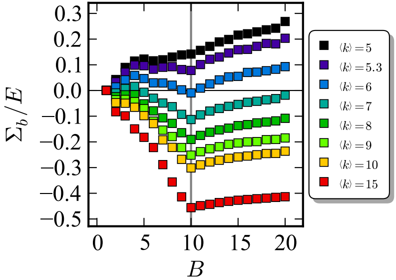
T. P. Peixoto, Phys. Rev. Lett. 110, 148701 (2013)

Works very well...

Generated ($B = 10$)

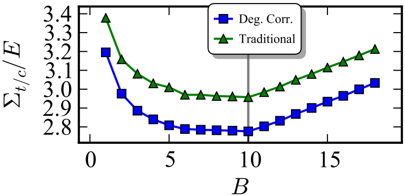


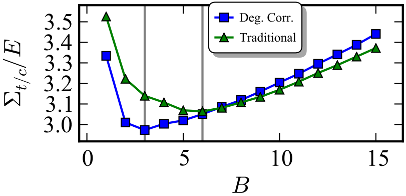
Planted

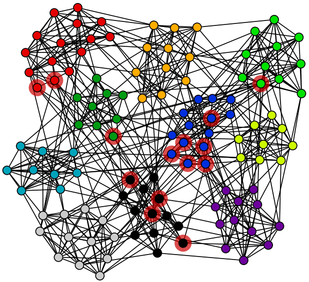


T. P. Peixoto, Phys. Rev. Lett. 110, 148701 (2013)

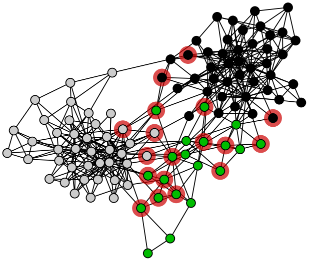
WORKS VERY WELL...







American football



Political Books

T. P. Peixoto, Phys. Rev. Lett. 110, 148701 (2013)

Example: The Internet Movie Database (IMDB)

Bipartite network of actors and films.
Fairly large: $N = 372,787$, $E = 1,812,657$

T. P. Peixoto, Phys. Rev. Lett. 110, 148701 (2013)

Example: The Internet Movie Database (IMDB)

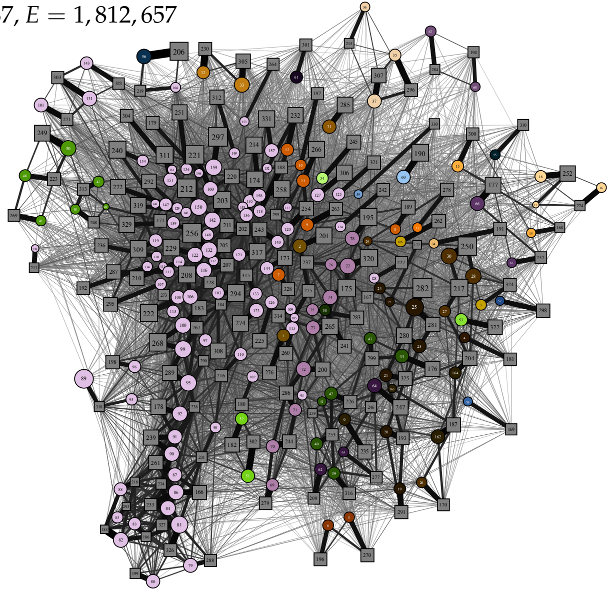
Bipartite network of actors and films.
Fairly large: $N = 372,787$, $E = 1,812,657$
MDL selects: $B = 332$

Example: The Internet Movie Database (IMDB)

Bipartite network of actors and films.

Fairly large: $N = 372,787$, $E = 1,812,657$

MDL selects: $B = 332$



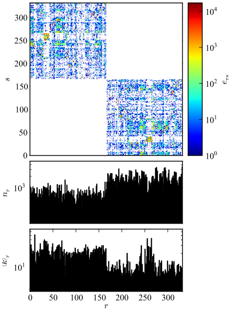
T. P. Peixoto, Phys. Rev. Lett. 110, 148701 (2013)

Example: The Internet Movie Database (IMDB)

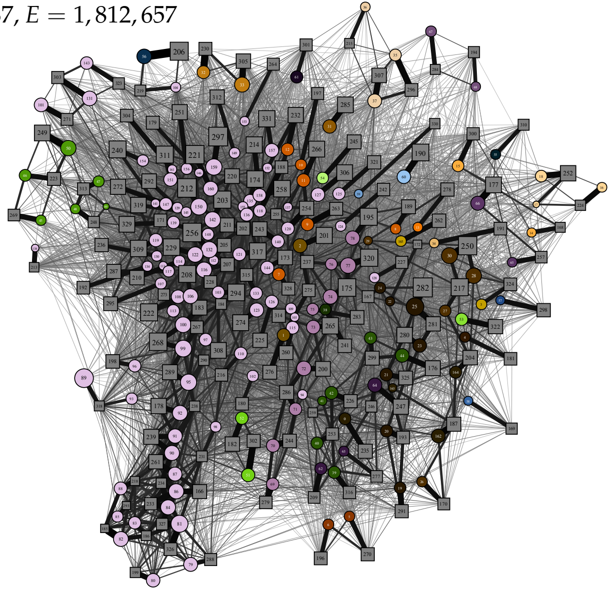
Bipartite network of actors and films.

Fairly large: $N = 372,787$, $E = 1,812,657$

MDL selects: $B = 332$



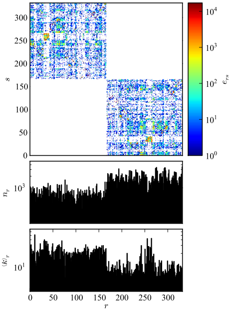
Bipartiteness is fully uncovered!



T. P. Peixoto, Phys. Rev. Lett. 110, 148701 (2013)

Example: The Internet Movie Database (IMDB)

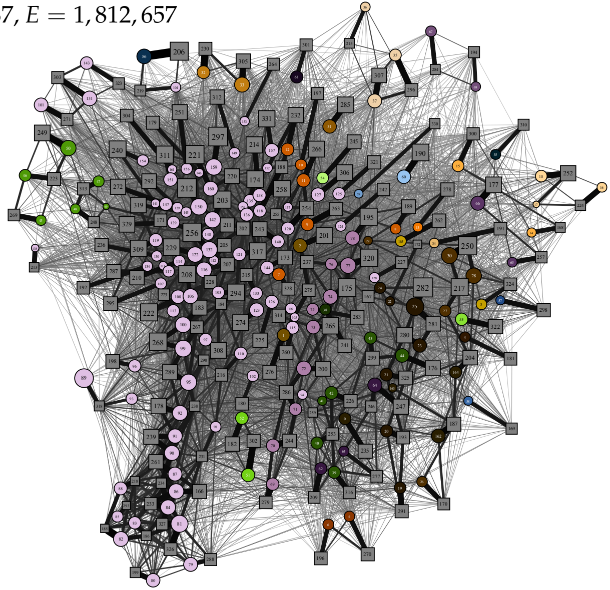
Bipartite network of actors and films.
Fairly large: $N = 372{,}787$, $E = 1{,}812{,}657$
MDL selects: $B = 332$



Bipartiteness is fully uncovered!

Detects meaningful features:

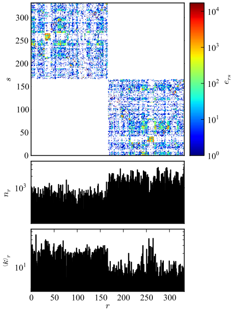
- Temporal
- Spatial (Country)
- Type/Genre



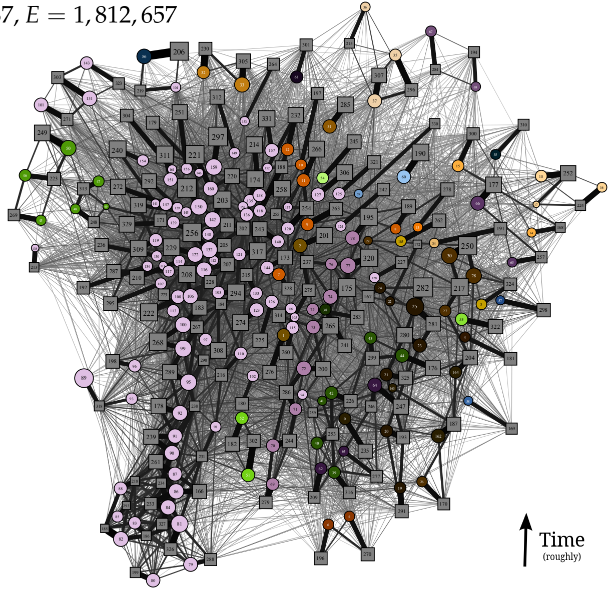
T. P. Peixoto, Phys. Rev. Lett. 110, 148701 (2013)

Example: The Internet Movie Database (IMDB)

Bipartite network of actors and films.
Fairly large: $N = 372,787$, $E = 1,812,657$
MDL selects: $B = 332$



Bipartiteness is fully uncovered!



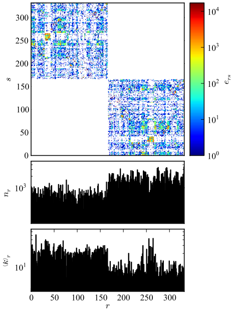
Detects meaningful features:

- Temporal
- Spatial (Country)
- Type/Genre

T. P. Peixoto, Phys. Rev. Lett. 110, 148701 (2013)

Example: The Internet Movie Database (IMDB)

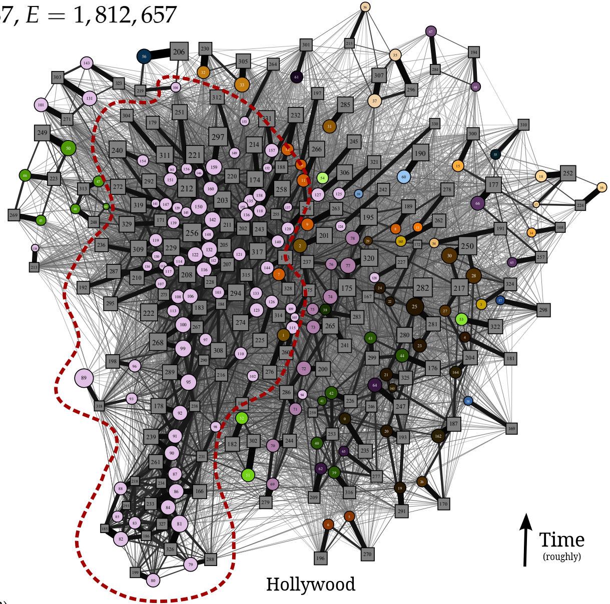
Bipartite network of actors and films.
Fairly large: $N = 372,787$, $E = 1,812,657$
MDL selects: $B = 332$



Bipartiteness is fully uncovered!

Detects meaningful features:

- Temporal
- Spatial (Country)
- Type/Genre



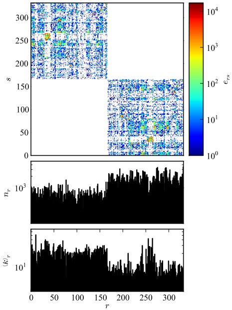
T. P. Peixoto, Phys. Rev. Lett. 110, 148701 (2013)

Example: The Internet Movie Database (IMDB)

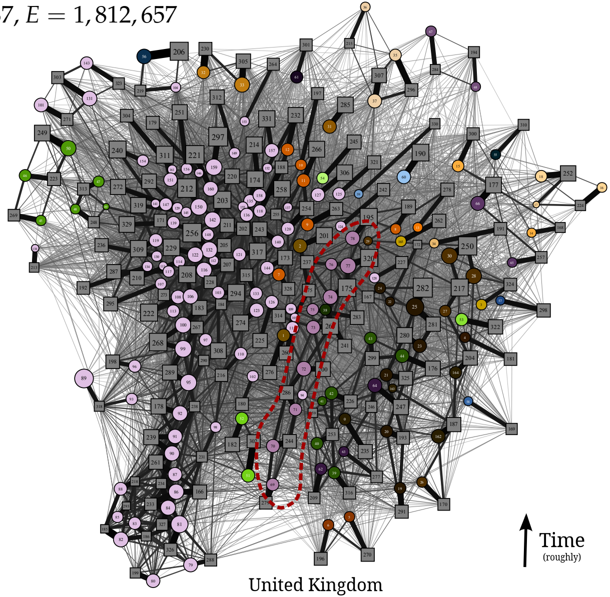
Bipartite network of actors and films.

Fairly large: $N = 372,787$, $E = 1,812,657$

MDL selects: $B = 332$



Bipartiteness is fully uncovered!



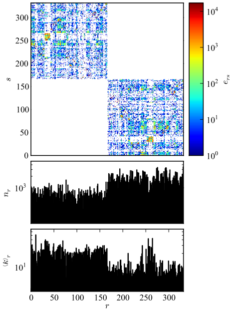
Detects meaningful features:

- Temporal
- Spatial (Country)
- Type/Genre

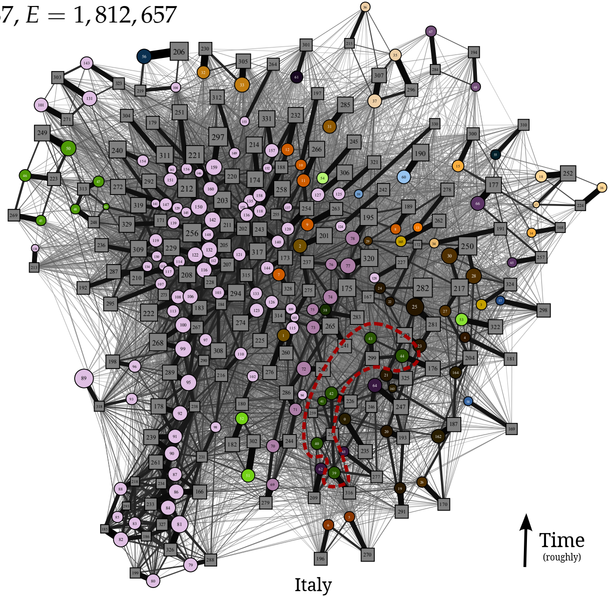
T. P. Peixoto, Phys. Rev. Lett. 110, 148701 (2013)

Example: The Internet Movie Database (IMDB)

Bipartite network of actors and films.
Fairly large: $N = 372,787$, $E = 1,812,657$
MDL selects: $B = 332$



Bipartiteness is fully uncovered!



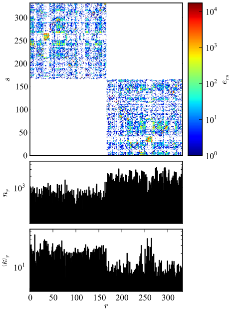
Detects meaningful features:

- Temporal
- Spatial (Country)
- Type/Genre

T. P. Peixoto, Phys. Rev. Lett. 110, 148701 (2013)

Example: The Internet Movie Database (IMDB)

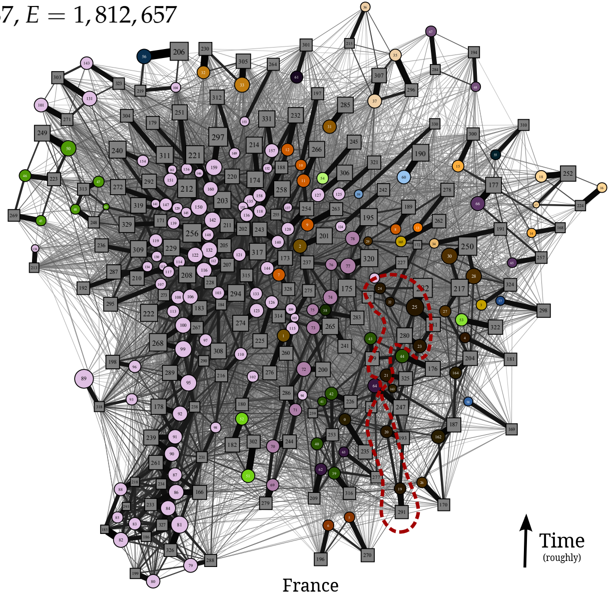
Bipartite network of actors and films.
Fairly large: $N = 372,787$, $E = 1,812,657$
MDL selects: $B = 332$



Bipartiteness is fully uncovered!

Detects meaningful features:

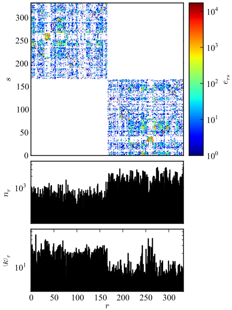
- Temporal
- Spatial (Country)
- Type/Genre



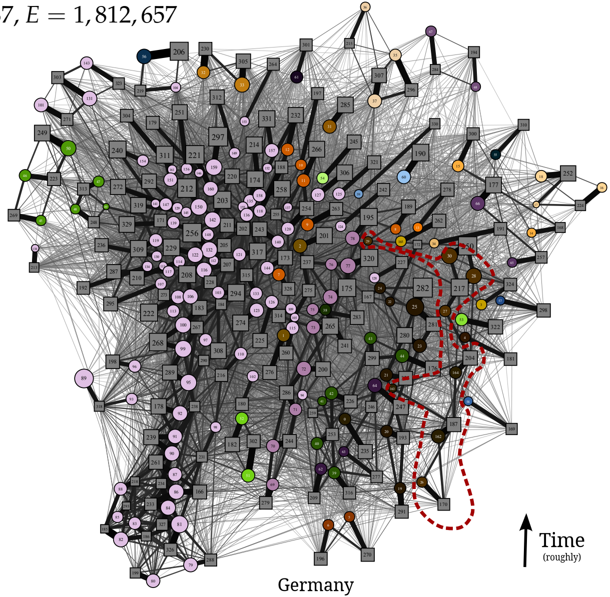
T. P. Peixoto, Phys. Rev. Lett. 110, 148701 (2013)

Example: The Internet Movie Database (IMDB)

Bipartite network of actors and films.
Fairly large: $N = 372,787$, $E = 1,812,657$
MDL selects: $B = 332$



Bipartiteness is fully uncovered!

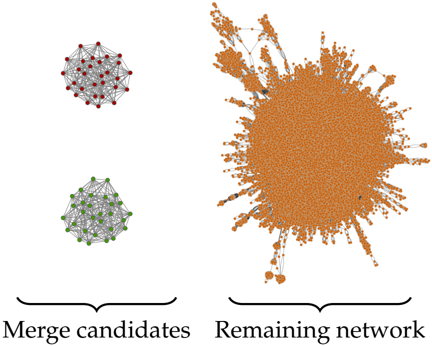


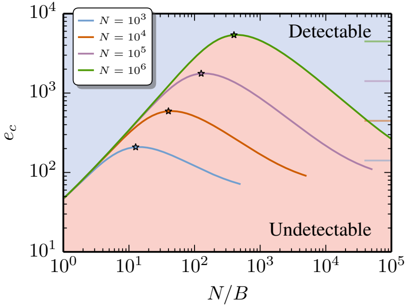
Detects meaningful features:

- Temporal
- Spatial (Country)
- Type/Genre

T. P. Peixoto, Phys. Rev. Lett. 110, 148701 (2013)

PROBLEM 3: RESOLUTION LIMIT !?





Minimum detectable block size $\sim \sqrt{N}$.

Why?

T. P. Peixoto, Phys. Rev. Lett. 110, 148701 (2013)

Resolution limit: Lack of prior information

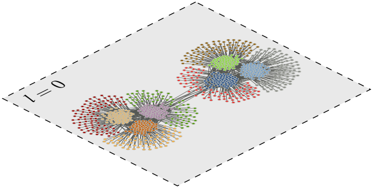
Assumption that all block structures (block graphs) occur with the same probability.

$$\Sigma \sim \underbrace{\mathcal{S}}_{\text{Data} \mid \text{Model}} + \underbrace{B^2 \ln E}_{\text{Edge counts}} + \underbrace{N \ln B}_{\text{Node partition}}$$

Noninformative priors $\rightarrow$ resolution limit

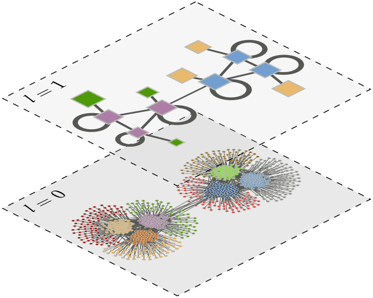
Solution: Model the model

T. P. Peixoto, Phys. Rev. X 4, 011047 (2014)



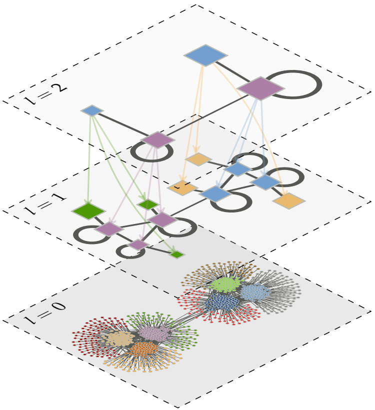
Solution: Model the model

T. P. Peixoto, Phys. Rev. X 4, 011047 (2014)



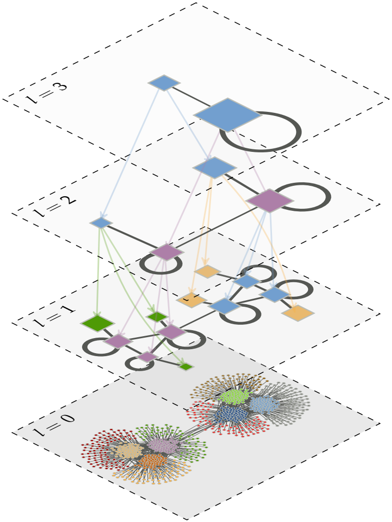
Solution: Model the model

T. P. Peixoto, Phys. Rev. X 4, 011047 (2014)



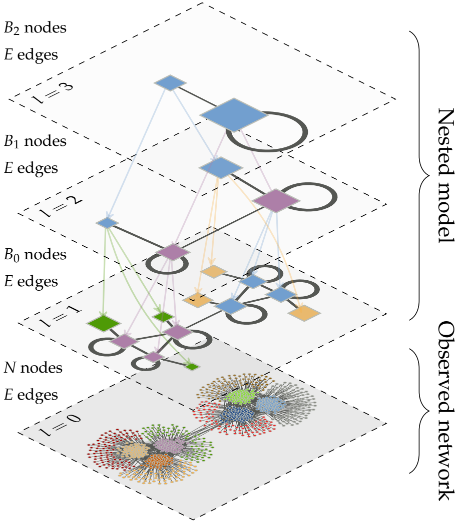
Solution: Model the model

T. P. Peixoto, Phys. Rev. X 4, 011047 (2014)



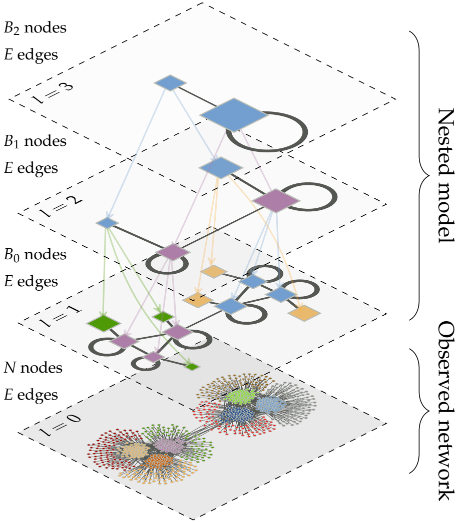
Solution: Model the model

T. P. Peixoto, Phys. Rev. X 4, 011047 (2014)



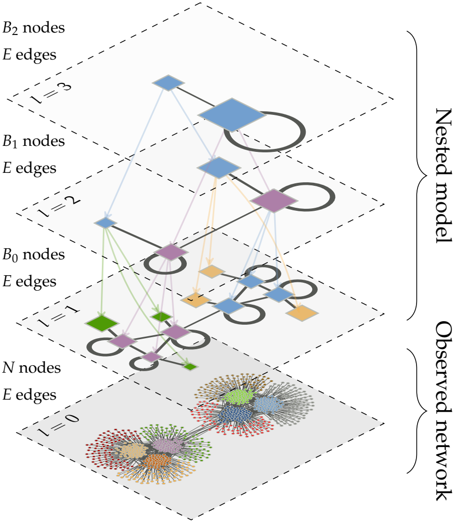
Solution: Model the model

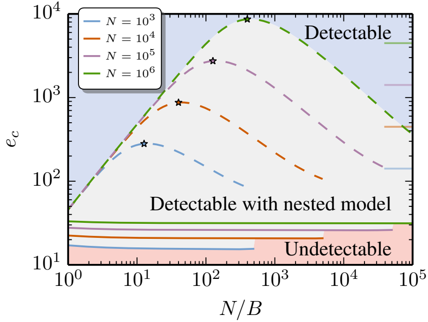
T. P. Peixoto, Phys. Rev. X 4, 011047 (2014)



Solution: Model the model

T. P. Peixoto, Phys. Rev. X 4, 011047 (2014)

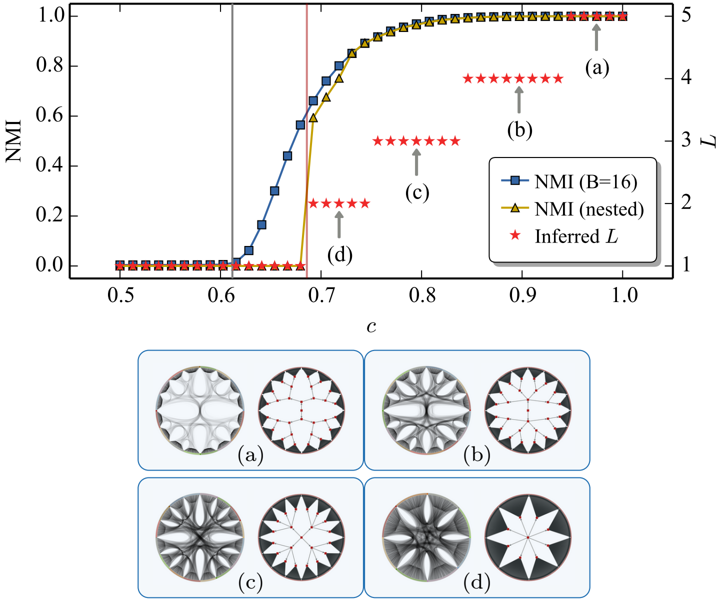




$$N/B_{\max} \sim \ln N$$

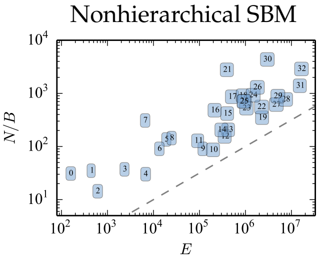
$$\ll \sqrt{N}$$

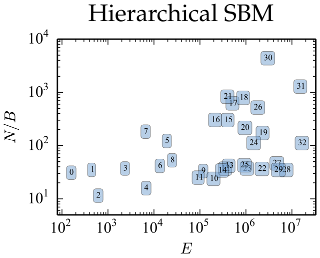
HIERARCHICAL MODEL: BUILT-IN VALIDATION



T. P. Peixoto, Phys. Rev. X 4, 011047 (2014)

EMPIRICAL NETWORKS





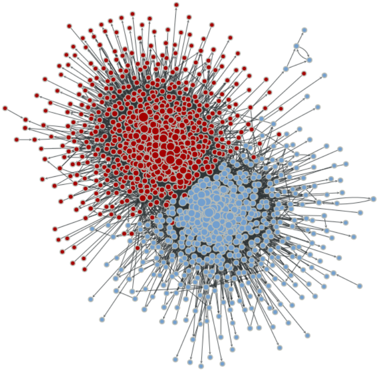
No.	N	E	Dir.	No.	N	E	Dir.	No.	N	E	Dir.
0	62	159	No	11	21,363	91,286	No	22	255,265	2,234,572	Yes
1	105	441	No	12	27,400	352,504	Yes	23	317,080	1,049,866	No
2	115	613	No	13	34,401	421,441	Yes	24	325,729	1,469,679	Yes
3	297	2,345	Yes	14	39,796	301,498	Yes	25	334,863	925,872	No
4	903	6,760	No	15	52,104	399,625	Yes	26	372,547	1,812,312	No
5	1,222	19,021	Yes	16	56,739	212,945	No	27	449,087	4,690,321	Yes
6	4,158	13,422	No	17	75,877	508,836	Yes	28	654,782	7,499,425	Yes
7	4,941	6,594	No	18	82,168	870,161	Yes	29	855,802	5,066,842	Yes
8	8,638	24,806	No	19	105,628	2,299,623	No	30	1,134,890	2,987,624	No
9	11,204	117,619	No	20	196,591	950,327	No	31	1,637,868	15,205,016	No
10	17,903	196,972	No	21	224,832	394,400	Yes	32	3,764,117	16,511,740	Yes

No.	Network	No.	Network	No.	Network
0	Dolphins	11	arXiv Co-Authors (cond-mat)	22	Web graph of `stanford.edu`.
1	Political Books	12	arXiv Citations (hep-th)	23	DBLP collaboration
2	American Football	13	arXiv Citations (hep-ph)	24	WWW
3	C. Elegans Neurons	14	PGP	25	Amazon product network
4	Disease Genes	15	Internet AS (Caida)	26	IMDB film-actor (bipartite)
5	Political Blogs	16	Brightkite social network	27	APS citations
6	arXiv Co-Authors (gr-qc)	17	Epinions.com trust network	28	Berkeley/Stanford web graph
7	Power Grid	18	Slashdot	29	Google web graph
8	arXiv Co-Authors (hep-th)	19	Flickr	30	Youtube social network
9	arXiv Co-Authors (hep-ph)	20	Gowalla social network	31	Yahoo groups (bipartite)
10	arXiv Co-Authors (astro-ph)	21	EU email	32	US patent citations

T.P.P., Phys. Rev. X 4, 011047 (2014)

Empirical networks

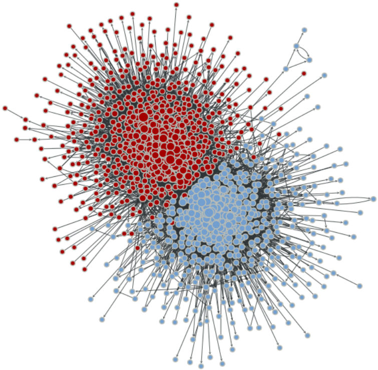
Political blogs ($N = 1,222, E = 19,027$)

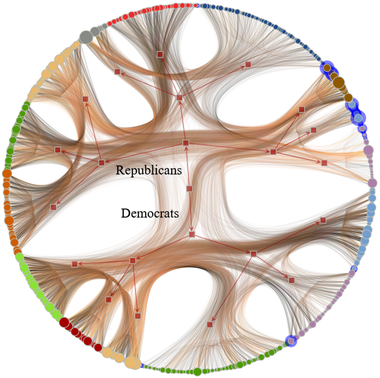


T. P. Peixoto, Phys. Rev. X 4, 011047 (2014)

Empirical networks

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

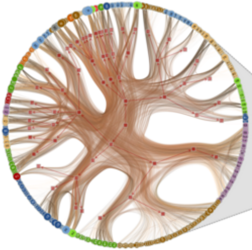


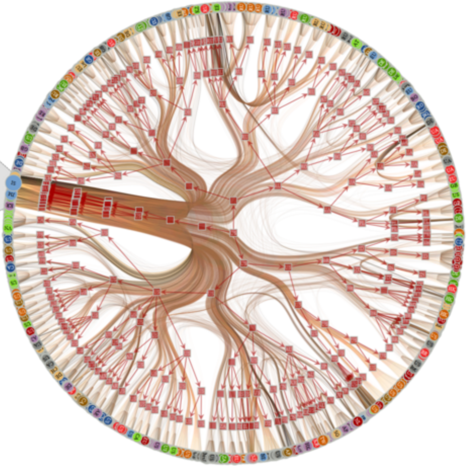


T. P. Peixoto, Phys. Rev. X 4, 011047 (2014)

EMPIRICAL NETWORKS

INTERNET (AUTONOMOUS SYSTEMS) ($N = 52,104, E = 399,625, B = 191$)

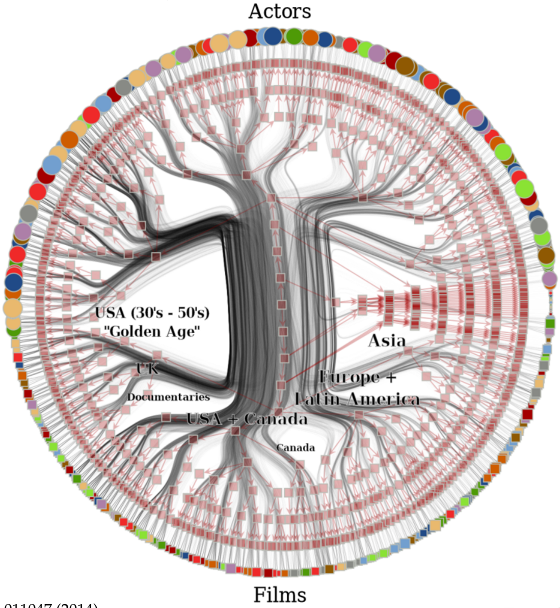




T. P. Peixoto, Phys. Rev. X 4, 011047 (2014)

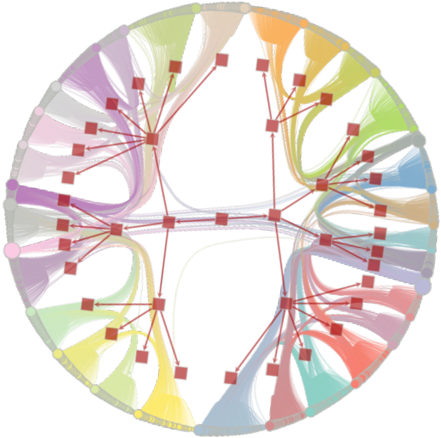
Empirical networks

IMDB Film-Actor Network ($N = 372,447, E = 1,812,312, B = 717$)



T. P. Peixoto, Phys. Rev. X 4, 011047 (2014)

Empirical networks



Human conectome

Part II

Limits on inference

How far can we recover what we put in?

Semi-parametric Bayesian inference

A. Decelle, F. Krzakala, C. Moore, L .Zdeborova, Phys. Rev. E 84, 066106 (2012)

$p_{rs}, \gamma_r \rightarrow$ Known parameters
$b_i \rightarrow$ Unknown parameters

$$P(\{b_i\}|G, \{p_{rs}\}, \{\gamma_r\}) = \frac{P(G|\{b_i\}, \{p_{rs}\})P(\{b_i\}|\{\gamma_r\})}{P(G|\{p_{rs}\}, \{\gamma_r\})}$$

$$P(G|\{b_i\}, \{p_{rs}\}) = \prod_{i<j} p_{b_i b_j}^{A_{ij}} (1 - p_{b_i b_j})^{1-A_{ij}}$$

$$P(\{b_i\}|\{\gamma_r\}) = \prod_i \gamma_{b_i}$$

$$P(G|\{p_{rs}\}, \{\gamma_r\}) = \sum_{\{b_i\}} P(G|\{b_i\}, \{p_{rs}\})P(\{b_i\}|\{\gamma_r\})$$

Semi-parametric Bayesian inference

A. Decelle, F. Krzakala, C. Moore, L .Zdeborova, Phys. Rev. E 84, 066106 (2012)

We have a Potts-like model...

$$P(\{b_i\}|G, \{p_{rs}\}, \{\gamma_r\}) = \frac{e^{-H(\{b_i\})}}{Z}$$

$$H(\{b_i\}) = -\sum_{i<j} A_{ij} \ln p_{b_i b_j} - (1 - A_{ij}) \ln(1 - p_{b_i b_j}) - \sum_i \ln \gamma_{b_i}$$

$$Z = \sum_{\{b_i\}} e^{-H(\{b_i\})}$$

Semi-parametric Bayesian inference

(Physicists doing complex systems.)

Cavity method: Solution on a tree

Bethe free energy

$$F = -\ln Z = -\sum_i \ln Z^i + \sum_{i<j} A_{ij} \ln Z^{ij} - E$$

$$Z^{ij} = N \sum_{r<s} p_{rs} (\psi_r^{i\to j} \psi_s^{j\to i} + \psi_s^{i\to j} \psi_r^{j\to i}) + N \sum_r p_{rr} \psi_r^{i\to j} \psi_r^{j\to i}$$

$$Z^i = \sum_r n_r e^{-h_r} \prod_{j\in\partial i} \sum_r N p_{rb_i} \psi_r^{j\to i}$$

Belief-propagation (BP): $$\psi_r^{i\to j} = \frac{1}{Z^{i\to j}} \gamma_r e^{-h_r} \prod_{k\in\partial i\setminus j} \left(\sum_s N p_{rs} \psi_s^{k\to i} \right)$$

Auxiliary fields: $h_r = \sum_i \sum_r p_{rb_i} \psi_r^i$

Node marginals:

$$P(b_i = r | p, \gamma) = \psi_r^i = \frac{1}{Z^i} \gamma_r \prod_{j\in\partial i} \left[\sum_s (N p_{rs})^{A_{ij}} (1 - p_{rs})^{1-A_{ij}} \psi_s^{j\to i} \right]$$

Exact on trees, and asymptotically exact on SBM networks with $N \gg B$.
Algorithmic complexity: $O(E \times B^2)$

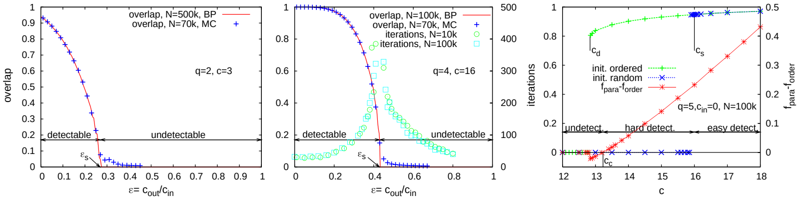
Decelle et al, 2012; M. Mézard, A. Montanari, “Information, Physics, and Computation”, 2009

PHASE TRANSITION IN SBM INFERENCE

A. DECELLE, F. KRZAKALA, C. MOORE, L .ZDEBOROVA, PHYS. REV. E 84, 066106 (2012)

Uniform SBM with B groups of size N/B, with an assortative structure,

$$Np_{rs} = c_{\text{in}}\delta_{rs} + c_{\text{out}}(1 - \delta_{rs}).$$



Detectable phase: $|c_{\text{in}} - c_{\text{out}}| > B\sqrt{\langle k \rangle}$

(Algorithmic independent!)

Rigorously proven: e.g. E. Mossel, J. Neeman, and A. Sly (2015).; E. Mossel, J. Neeman, and A. Sly (2014); C. Bordenave, M. Lelarge, and L. Massoulié, (2015)

Expectation Maximizaton + BP

Maximum likelihood for $P(G|\{p_{rs}\}, \gamma_r) = Z$

Expectation-maximization algorithm

Starting from an initial guess for $\{p_{rs}\}$, γ_r:

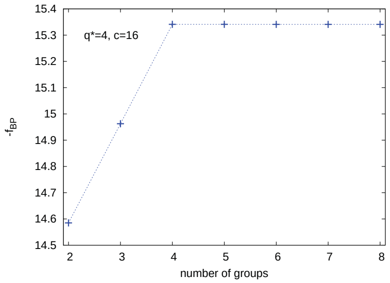
- Solve BP equations, and obtain node marginals. (Expectation step)
- Obtain new parameters estimates via $\{\hat{p}_{rs}\}, \{\hat{\gamma}_r\} = \underset{\{p_{rs}\},\{\gamma_r\}}{\operatorname{argmax}} P(G|\{p_{rs}\}, \gamma_r)$ (Maximization step)
- Repeat.

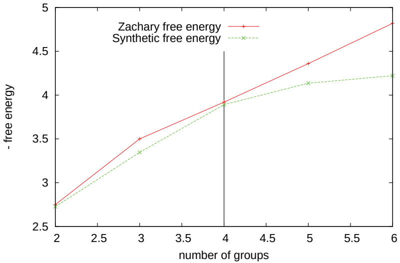
Algorithmic complexity: $O(EB^2)$

Problems:

- Converges to a local optimum.
- It is a parametric method, hence it cannot find the number of groups B.

EXPECTATION MAXIMIZATON + BP





Overfitting...

Part III

Model variations and model selection

Hypothesis testing

T.P.P, Phys. Rev. X 5, 011033 (2015)

Bayesian modelling allows us to select between different classes of models, according to statistical evidence.

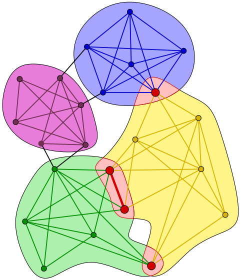
Posterior odds ratio:

$$\Lambda = \frac{P(\{\theta\}_a|G,\mathcal{H}_a)P(\mathcal{H}_a)}{P(\{\theta\}_b|G,\mathcal{H}_b)P(\mathcal{H}_b)}$$
$$= \exp(-\Delta\Sigma)$$

Subjective significance levels:

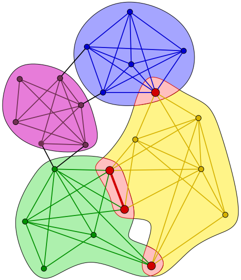
Λ	Evidence strength for $\mathcal{H}_b$
> 1	Favors $\mathcal{H}_a$
$1/3 < \Lambda < 1$	Very weak
$1/10 < \Lambda < 1/3$	Substantial
$1/30 < \Lambda < 1/10$	Strong
$1/100 < \Lambda < 1/30$	Very strong
$\Lambda < 1/100$	Decisive

Overlapping groups

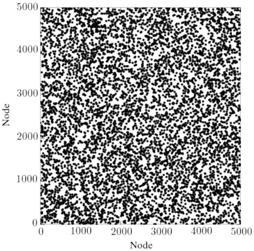


(Palla et al 2005)

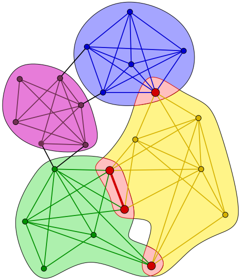
Overlapping groups



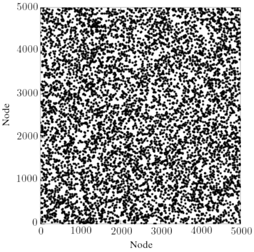
(Palla et al 2005)

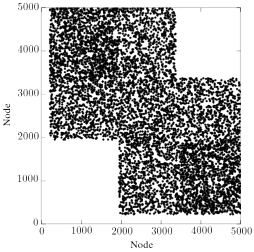


Overlapping groups



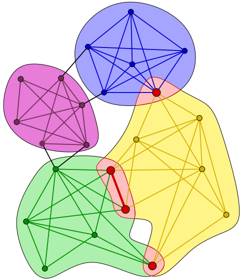
(Palla et al 2005)



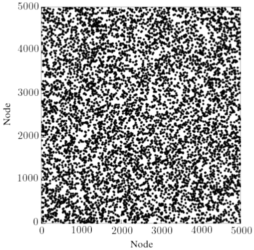


- Number of nonoverlapping partitions: B^N
- Number of overlapping partitions: 2^{BN}

OVERLAPPING GROUPS



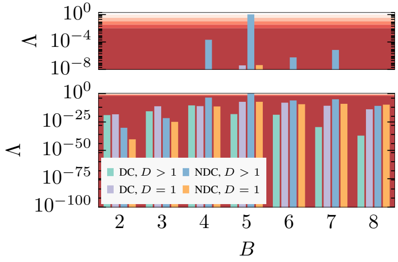
(Palla et al 2005)

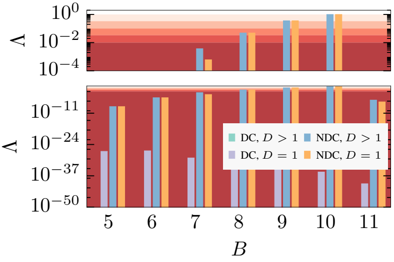


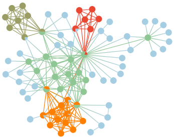
- Number of nonoverlapping partitions: B^N
- Number of overlapping partitions: 2^{BN}

Hypothesis testing

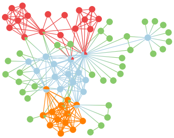
T.P.P, Phys. Rev. X 5, 011033 (2015)







$B = 5$,
non-degree-corrected,
overlapping, $\Lambda = 1$



$B = 4$,
non-degree-corrected,
overlapping,
$\Lambda \simeq 2 \times 10^{-4}$,

$B = 10$,
non-degree-corrected,
nonoverlapping, $\Lambda = 1$

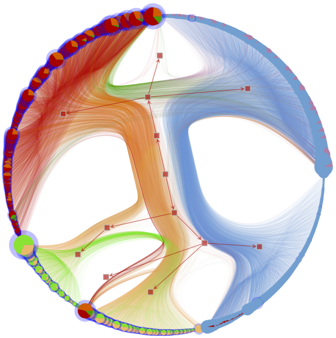
$B = 9$,
non-degree-corrected,
nonoverlapping,
$\Lambda \simeq 0.36$

(American College Football)

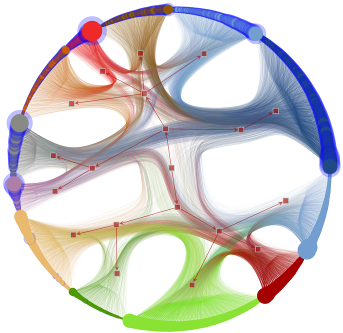
(Les Misérables)

HYPOTHESIS TESTING / MODEL COMPARISON

Political blogs



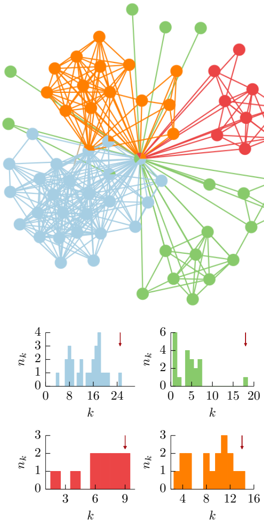
$B = 7$, overlapping, degree-corrected, $\Lambda = 1$



$B = 12$, nonoverlapping, degree-corrected, $\log_{10} \Lambda \simeq -747$

Hypothesis testing / Model comparison

Social "ego" network (from Facebook)

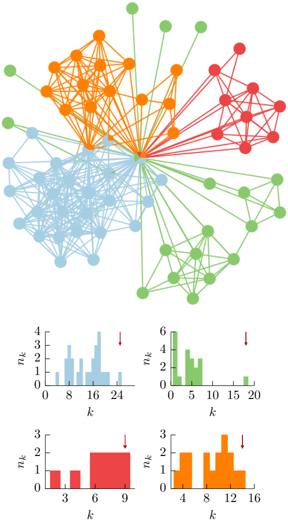


$B = 4, \Lambda \simeq 0.053$

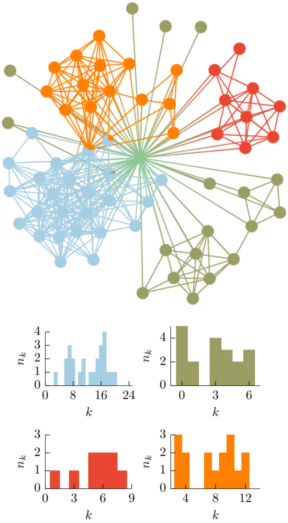
T.P.P, Phys. Rev. X 5, 011033 (2015)

Hypothesis testing / Model comparison

Social "ego" network (from Facebook)



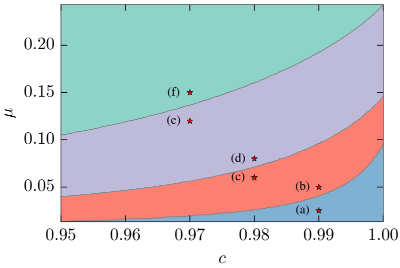
$B = 4, \Lambda \simeq 0.053$

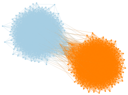


$B = 5, \Lambda = 1$

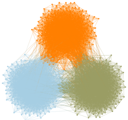
T.P.P, Phys. Rev. X 5, 011033 (2015)

OVERLAP VS. NONOVERLAP

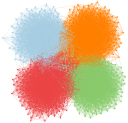




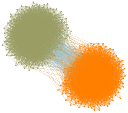
(a) $B = 2$, $c = 0.99$, $\mu = 0.025$



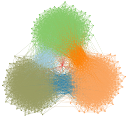
(c) $B = 3$, $c = 0.98$, $\mu = 0.06$



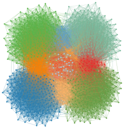
(e) $B = 4$, $c = 0.97$, $\mu = 0.12$



(b) $B = 3$, $c = 0.99$, $\mu = 0.05$



(d) $B = 7$, $c = 0.98$, $\mu = 0.08$



(f) $B = 15$, $c = 0.97$, $\mu = 0.15$

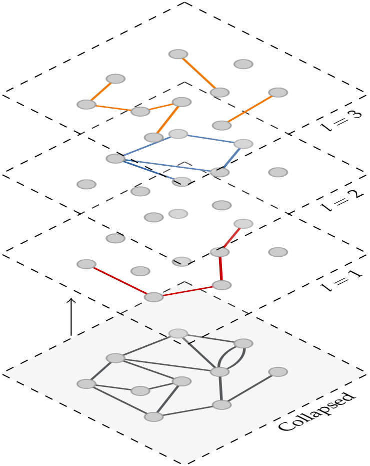
Hypothesis testing / Model comparison

No.	N	$\langle k \rangle$	$\log_{10} \Lambda_{DCO}$	$\log_{10} \Lambda_{DC}$	$\log_{10} \Lambda_{NDCO}$	$\log_{10} \Lambda_{NDC}$	B	$\langle d \rangle$	Σ/E
1	34	4.6	-2.1	-2.1	—	0	2	1	4
2	62	5.1	-4.6	-1.4	—	0	2	1	4.8
3	77	6.6	-17	-7.7	0	-7.3	5	1.1	4
4	105	8.4	-12	-2.8	-6.6	0	5	1	4.4
5	115	10.7	-79	-27	—	0	10	1	4.3
6	297	15.9	0	-61	-2.0×10^2	-2.1×10^2	5	1.8	5.1
7	379	4.8	-47	-6.6	0	-8.9	20	1.1	6.2
8	903	15.0	-3.8×10^2	-3.7×10^2	0	-3.7×10^2	60	1.2	3.1
9	1,278	2.8	-8.1	0	-1.5×10^2	-89	2	1	7.4
10	1,490	25.6	0	-5.2×10^2	-2.3×10^3	-2.3×10^3	7	1.8	4.4
11	1,536	3.8	-2.5×10^2	0	-65	-62	38	1	6.7
12	1,622	11.2	-4.3×10^2	0	-12	-82	48	1	3.3
13	1,756	4.5	-43	0	-4.0×10^2	-2.8×10^2	7	1	5.9
14	2,018	2.9	-9.2	0	-2.9×10^2	-2.1×10^2	2	1	8.5
15	4,039	43.7	-1.5×10^3	0	-8.1×10^2	-9.5×10^2	158	1	3.2
16	4,941	2.7	-2.2×10^2	0	-21	-25	25	1	11
17	7,663	17.8	0	-1.1×10^4	-5.3×10^3	-1.6×10^4	85	1	3.2
18	7,663	5.3	-1.8×10^3	0	-9.3×10^2	-7.3×10^2	63	1	5
19	8,298	25.0	-9.1×10^3	0	-1.4×10^4	-1.4×10^4	34	1	5.4
20	9,617	7.7	-4.2×10^3	0	-2.3×10^3	-2.5×10^3	34	1	9.3
21	26,197	2.2	-2.4×10^3	-1.2×10^3	0	-2.7×10^3	363	1.3	4.5
22	36,692	20.0	-4.1×10^4	0	-8.5×10^4	-2.8×10^4	1812	1	5.5
23	39,796	15.2	-6.1×10^4	0	-8.8×10^4	-4.5×10^4	1323	1	6.3
24	52,104	15.3	-1.5×10^5	0	-3.7×10^4	-4.0×10^4	172	1	6.4
25	58,228	14.7	0	-5.8×10^4	-1.8×10^5	-1.4×10^5	1995	3.2	7.3
26	65,888	305.2	-4.4×10^4	0	-4.6×10^5	-4.6×10^5	384	1	4.1
27	68,746	1.5	-4.8×10^3	-1.4×10^3	0	-7.0×10^3	719	1.4	6.4
28	75,888	13.4	-1.1×10^5	0	-8.2×10^4	-9.0×10^4	143	1	8.9
29	89,209	5.3	-1.0×10^4	0	-9.7×10^3	-1.1×10^4	848	1	3.2
30	108,300	3.5	-3.3×10^3	-5.2×10^3	0	-2.4×10^4	1660	1.8	5.7
31	133,280	5.9	0	-4.4×10^3	-7.4×10^4	-3.8×10^4	1944	5.3	4.4
32	196,591	19.3	0	-1.9×10^5	-7.1×10^5	-6.6×10^5	6856	3.7	7.8
33	265,214	3.2	-1.4×10^4	0	-9.2×10^4	-8.5×10^4	549	1	8.6
34	273,957	16.8	-5.4×10^5	0	-4.6×10^4	-7.2×10^4	727	1	5.8
35	281,904	16.4	-1.2×10^6	0	-2.8×10^5	-1.5×10^5	6655	1	4.3
36	317,080	6.6	-1.7×10^5	0	-3.9×10^5	-4.2×10^5	8766	1	11
37	325,729	9.2	-5.8×10^5	0	-1.1×10^6	-2.3×10^5	4293	1	5.8
38	325,729	9.2	-5.6×10^5	0	-1.2×10^6	-2.5×10^5	3995	1	5.8
39	334,863	5.5	-3.3×10^5	0	-3.6×10^5	-3.4×10^4	9118	1	11
40	372,787	9.7	-1.0×10^6	0	-1.3×10^5	-1.4×10^5	965	1	11
41	463,347	20.3	-6.4×10^5	0	-1.8×10^6	-1.5×10^6	9276	1	9.3
42	1,134,890	5.3	—	0	-4.5×10^5	-4.9×10^5	264	1	13

T.P.P, Phys. Rev. X 5, 011033 (2015)

SBM WITH LAYERS

T.P.P, PHYS. REV. E 92, 042807 (2015)



- Fairly straightforward. Easily combined with degree-correction, overlaps, etc.
- Edge probabilities are in general different in each layer.
- Node memberships can move or stay the same across layer.
- Works as a general model for discrete as well as *discretized* edge covariates.
- Works as a model for temporal networks.

SBM WITH LAYERS

T.P.P, PHYS. REV. E 92, 042807 (2015)

Edge covariates

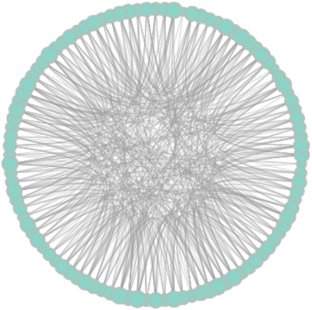
$$P(\{G_l\}|\{\theta\}) = P(G_c|\{\theta\}) \prod_{r \le s} \frac{\prod_l m_{rs}^l!}{m_{rs}!}$$

Independent layers

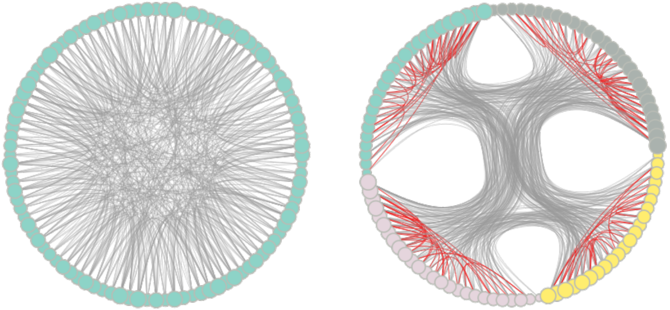
$$P(\{G_l\}|\{\{\theta\}_l\}, \{\phi\}, \{z_{il}\}) = \prod_l P(G_l|\{\theta\}_l, \{\phi\})$$

Embedded models can be of any type: Traditional, degree-corrected, overlapping.

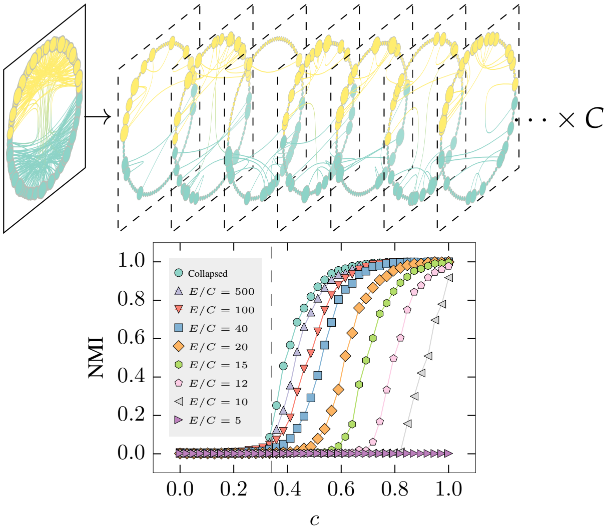
Layer information can reveal hidden structure



Layer information can reveal hidden structure



… BUT IT CAN ALSO HIDE STRUCTURE!



Model selection

Null model: Collapsed (aggregated) SBM + fully random layers

$$P(\{G_l\}|\{\theta\},\{E_l\}) = P(G_c|\{\theta\}) \times \frac{\prod_l E_l!}{E!}$$

(we can also aggregate layers into larger layers)

$N = 241$ Physicians

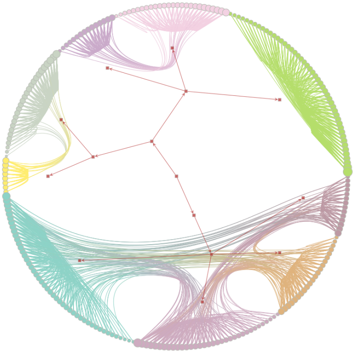
Survey questions:

- "When you need information or advice about questions of therapy where do you usually turn?"
- "And who are the three or four physicians with whom you most often find yourself discussing cases or therapy in the course of an ordinary week – last week for instance?"
- "Would you tell me the first names of your three friends whom you see most often socially?"

T.P.P, Phys. Rev. E 92, 042807 (2015)

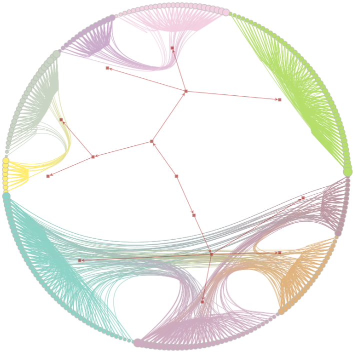
Model selection

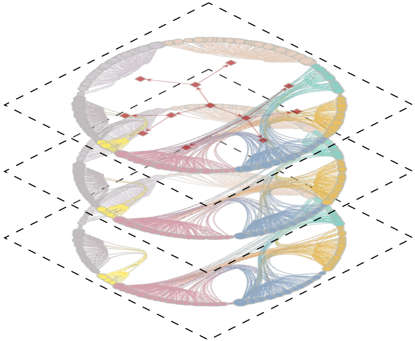
Example: Social network of physicians



Model selection

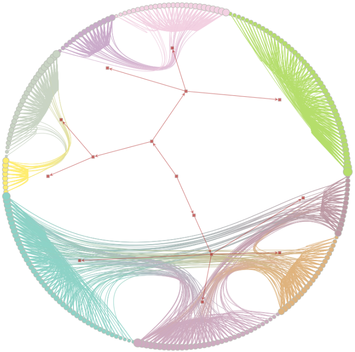
Example: Social network of physicians



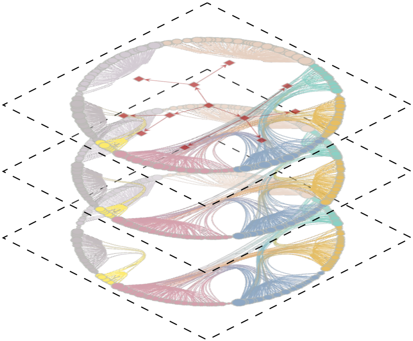


Model selection

Example: Social network of physicians



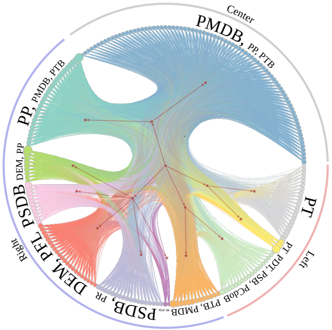
$\Lambda = 1$



$\log_{10} \Lambda \approx -50$

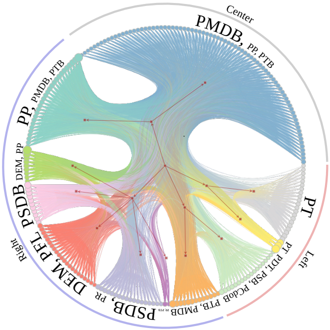
EXAMPLE: BRAZILIAN CHAMBER OF DEPUTIES

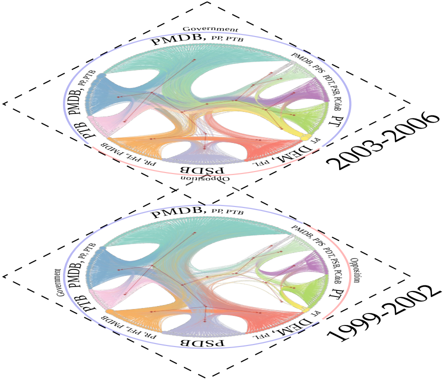
Voting network between members of congress (1999-2006)



Example: Brazilian chamber of deputies

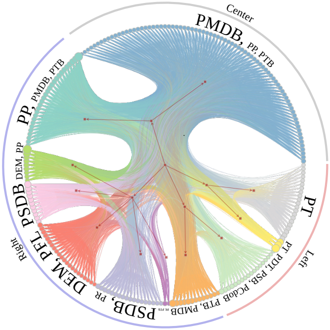
Voting network between members of congress (1999-2006)

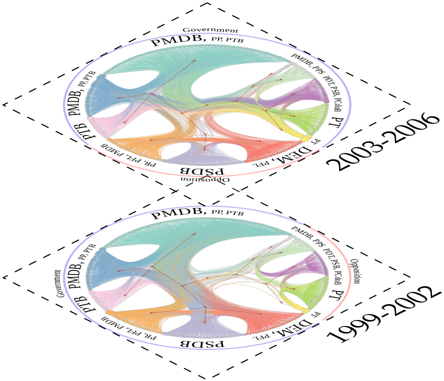




EXAMPLE: BRAZILIAN CHAMBER OF DEPUTIES

Voting network between members of congress (1999-2006)





$\log_{10} \Lambda \approx -111$

$\Lambda = 1$

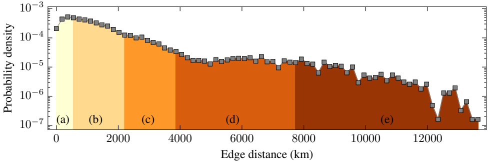
REAL-VALUED EDGES?

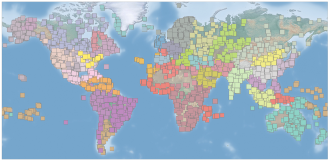
Idea: Layers $\{\ell\} \to$ bins of edge values!

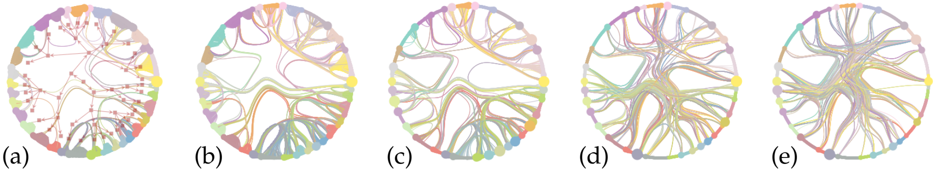
$$P(\{G_x\}|\{\theta\}_{\{\ell\}}, \{\ell\}) = P(\{G_l\}|\{\theta\}_{\{\ell\}}, \{\ell\}) \times \prod_l \rho(x_l)$$

Bayesian posterior $\to$ Number (and shape) of bins

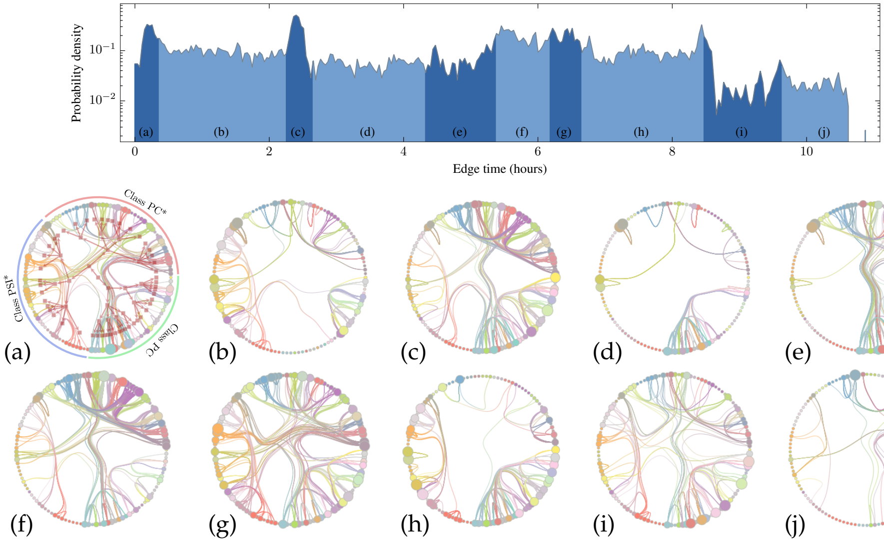
EXAMPLE: GLOBAL AIRPORT NETWORK





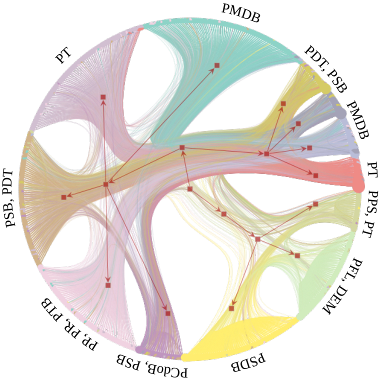


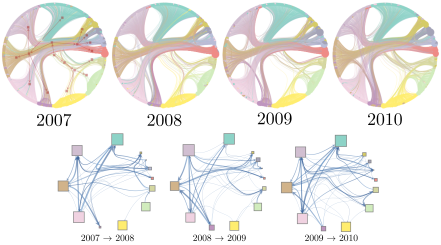
PROXIMITY BETWEEN HIGH-SCHOOL STUDENTS



J. Fournet, A. Barrat. PLoS ONE 9(9):e107878 (2014), `http://www.sociopatterns.org/`

Movement between groups...





TEMPORAL NETWORKS REDUX

AS AN ACTUAL DYNAMICAL SYSTEM THIS TIME

A more basic scenario: Sequences.

$$x_t \in [1, N] \rightarrow \text{ Token at time } t$$

$$x_t = \{x_0, x_1, x_2, x_3, \dots\}$$

(Tokens could be letters, base pairs, itinerary stops, etc.)

n-ORDER MARKOV CHAINS WITH COMMUNITIES

T. P. P. AND MARTIN ROSVALL, ARXIV: 1509.04740

Transitions conditioned on the last n tokens

$p(x_t|\vec{x}_{t-1}) \rightarrow$ Probability of transition from memory $\vec{x}_{t-1} = \{x_{t-n}, \ldots, x_{t-1}\}$ to token x_t

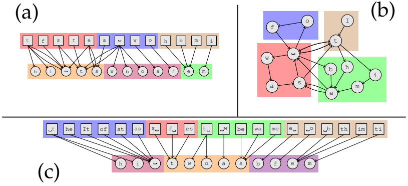
Instead of such a direct parametrization, we divide the tokens and memories into groups:

$$p(x|\vec{x}) = \theta_x \lambda_{b_x b_{\vec{x}}}$$

$\theta_x \rightarrow$ Overall frequency of token x

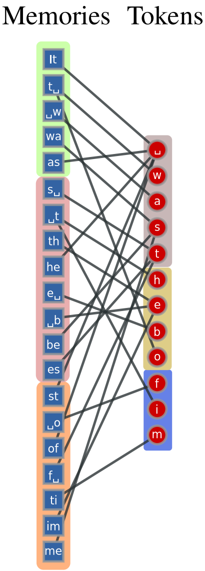
$\lambda_{rs} \rightarrow$ Transition probability from memory group s to token group r

$b_x, b_{\vec{x}} \rightarrow$ Group memberships of tokens and groups



$\{x_t\}$ = "It␣was␣the␣best␣of␣times"

n-ORDER MARKOV CHAINS WITH COMMUNITIES



$$\{x_t\} = \text{"It was the best of times"}$$

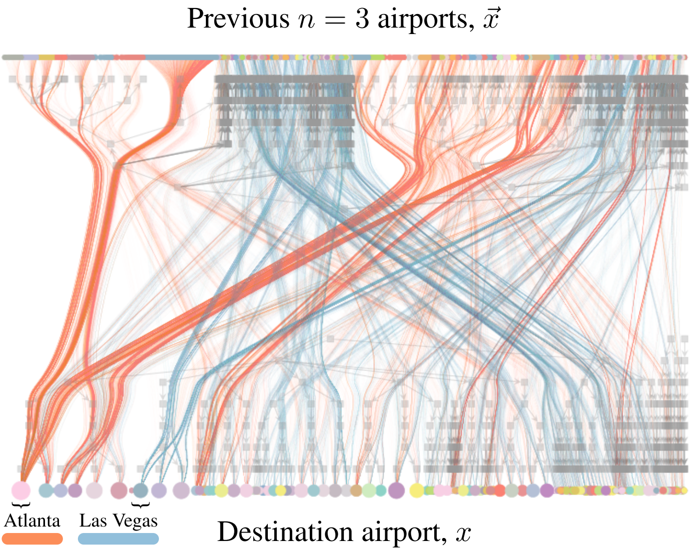
$$P(\{x_t\}|b) = \int d\lambda d\theta \, P(\{x_t\}|b, \lambda, \theta) P(\theta) P(\lambda)$$

The Markov chain likelihood is (almost) identical to the SBM likelihood that generates the bipartite graph.

Nonparametric → We can select the **number of groups** *and* the **Markov order** based on statistical evidence!

n-ORDER MARKOV CHAINS WITH COMMUNITIES

EXAMPLE: FLIGHT ITINERARIES



n-ORDER MARKOV CHAINS WITH COMMUNITIES

	US Air Flights				War and peace				Taxi movements				"Rock you" password list			
n	B_N	B_M	Σ	Σ'	B_N	B_M	Σ	Σ'	B_N	B_M	Σ	Σ'	B_N	B_M	Σ	Σ'
1	384	365	364, 385, 780	365, 211, 460	65	71	11, 422, 564	11, 438, 753	387	385	2, 635, 789	2, 975, 299	140	147	1, 060, 272, 230	1, 060, 385, 582
2	386	7605	319, 851, 871	326, 511, 545	62	435	9, 175, 833	9, 370, 379	397	1127	2, 554, 662	3, 258, 586	109	1597	984, 697, 401	987, 185, 890
3	183	2455	318, 380, 106	339, 898, 057	70	1366	7, 609, 366	8, 493, 211	393	1036	2, 590, 811	3, 258, 586	114	4703	910, 330, 062	930, 926, 370
4	292	1558	318, 842, 968	337, 988, 629	72	1150	7, 574, 332	9, 282, 611	397	1071	2, 628, 813	3, 258, 586	114	5856	889, 006, 060	940, 991, 463
5	297	1573	335, 874, 766	338, 442, 011	71	882	10, 181, 047	10, 992, 795	395	1095	2, 664, 990	3, 258, 586	99	6430	1, 000, 410, 410	1, 005, 057, 233
gzip	573, 452, 240				9, 594, 000				4, 289, 888				1, 315, 388, 208			
LZMA	402, 125, 144				7, 420, 464				2, 902, 904				1, 097, 012, 288			

(SBM can compress your files!)

T. P. P. and Martin Rosvall, arXiv: 1509.04740

DYNAMIC NETWORKS

Each token is an edge: $x_t \to (i, j)_t$

Dynamic network $\to$ Sequence of edges: $\{x_t\} = \{(i, j)_t\}$

Problem: Too many possible tokens! $O(N^2)$

Solution: Group the nodes into B groups.
Pair of node groups $(r, s) \to$ edge group.

Number of tokens: $O(B^2) \ll O(N^2)$

Two-step generative process:

$$\{x_t\} = \{(r, s)_t\}$$

(n-order Markov chain)

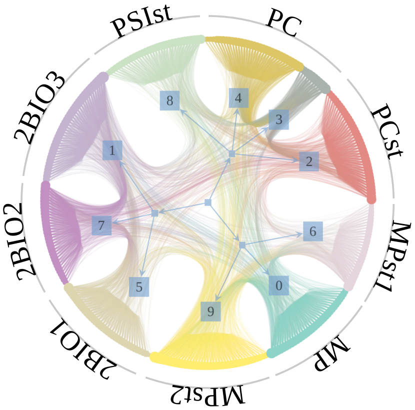
$$P((i, j)_t | (r, s)_t)$$

(static SBM)

Dynamic networks

Example: Student proximity

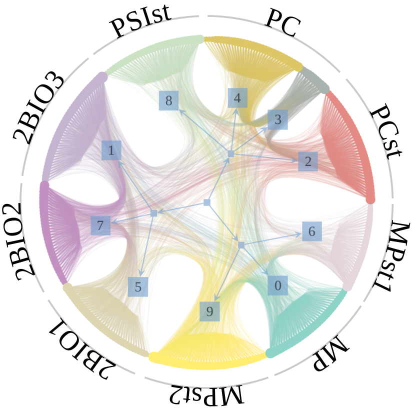
Static part



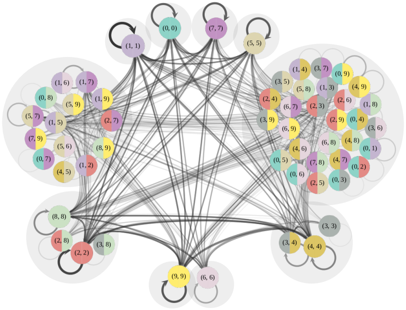
Dynamic networks

Example: Student proximity

Static part



Temporal part



Dynamic networks

Continuous time

$x_\tau \to$ token at continuous time τ

$$P(\{x_\tau\}) = \underbrace{P(\{x_t\})}_{\text{Discrete chain}} \times \underbrace{P(\{\Delta_t\}|\{x_t\})}_{\text{Waiting times}}$$

Exponential waiting time distribution

$$P(\{\Delta_t\}|\{x_t\},\lambda) = \prod_{\vec{x}} \lambda_{b_{\vec{x}}}^{k_{\vec{x}}} e^{-\lambda_{b_{\vec{x}}}\Delta_{\vec{x}}}$$

Bayesian integrated likelihood

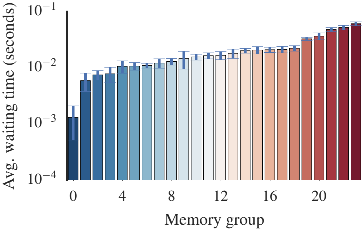
$$P(\{\Delta_t\}|\{x_t\}) = \prod_r \int_0^\infty \mathrm{d}\lambda\ \lambda^{e_r} e^{-\lambda\Delta_r} P(\lambda),$$
$$= \alpha^E \prod_r \frac{(e_r-1)!}{\Delta_r^{e_r}}.$$

Jeffreys prior: $P(\lambda) = \frac{\alpha}{\lambda}$

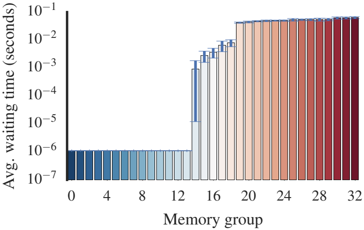
Dynamic networks

Continuous time

$\{x_\tau\} \rightarrow$ Sequence of notes in Beethoven's fifth symphony



Without waiting times
$(n = 1)$



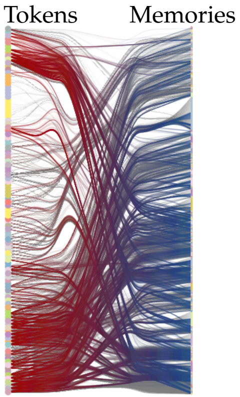
With waiting times
$(n = 2)$

Dynamic networks

Nonstationarity

$\{x_t\} \to$ Concatenation of "War and peace", by Leo Tolstoy, and "À la recherche du temps perdu", by Marcel Proust.

Unmodified chain



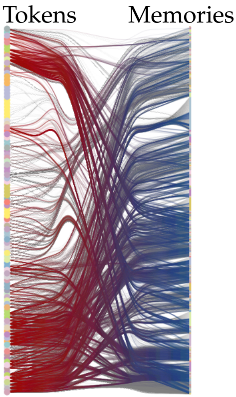
$$-\log_2 P(\{x_t\}, b) = 7,450,322$$

Dynamic networks

Nonstationarity

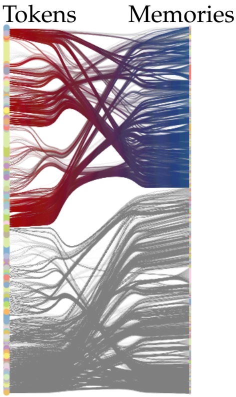
$\{x_t\} \to$ Concatenation of "War and peace", by Leo Tolstoy, and "À la recherche du temps perdu", by Marcel Proust.

Unmodified chain



$$-\log_2 P(\{x_t\}, b) = 7,450,322$$

Annotated chain $x'_t = (x_t, \text{novel})$



$$-\log_2 P(\{x_t\}, b) = 7,146,465$$

Model variations: Weighted graphs

C. Aicher, A. Z. Jacobs, and A. Clauset. Journal of Complex Networks 3(2), 221-248 (2015)

$$A_{ij} \in [0,1] \to \text{Adjacency matrix}$$
$$x_{ij} \in \mathbb{R} \to \text{Real-valued covariates}$$

Weights distributed according to the group memberships

$$P(\{x_{ij}\}|\{b_i\},\{\theta_{rs}\}) = \prod_{i<j} P(x_{ij}|\theta_{b_i b_j})$$

$$P(x|\theta) \to \text{Exponential family (Exponential, Gaussian, Pareto...)}$$

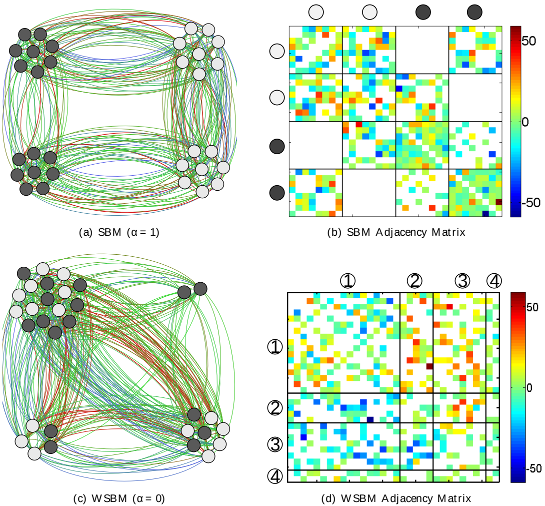
Caveats

- Shape of weight distribution must be defined beforehand.
- Parametric method (may overfit).
- Computationally expensive.

Model variations: Weighted graphs

C. Aicher, A. Z. Jacobs, and A. Clauset. Journal of Complex Networks 3(2), 221-248 (2015)

NFL games, weights: score difference



Model variations: Annotated networks

M.E.J. Newman and A. Clauset, arXiv:1507.04001

Main idea: Treat metadata as data, not "ground truth".

Annotations are partitions, $\{x_i\}$

Can be used as priors:

$$P(G, \{x_i\}|\theta, \gamma) = \sum_{\{b_i\}} P(G|\{b_i\}, \theta) P(\{b_i\}|\{x_i\}, \gamma)$$

$$P(\{b_i\}|\{x_i\}, \gamma) = \prod_i \gamma_{b_i x_u}$$

Drawbacks: Parametric (i.e. can overfit). Annotations are not always partitions.

MODEL VARIATIONS: ANNOTATED NETWORKS

M.E.J. NEWMAN AND A. CLAUSET, ARXIV:1507.04001

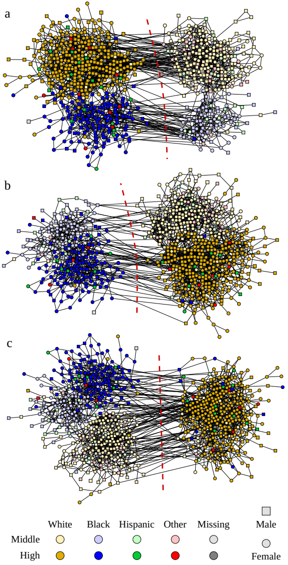


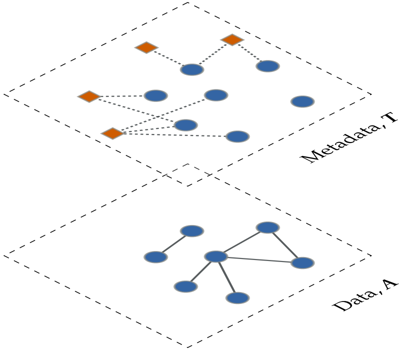
FIG. 2: Three divisions of a school friendship network, using as metadata (a) school grade, (b) ethnicity, and (c) gender.

Model variations: Annotated networks

Darko Hric, T. P. P., Santo Fortunato, arXiv:1604.00255

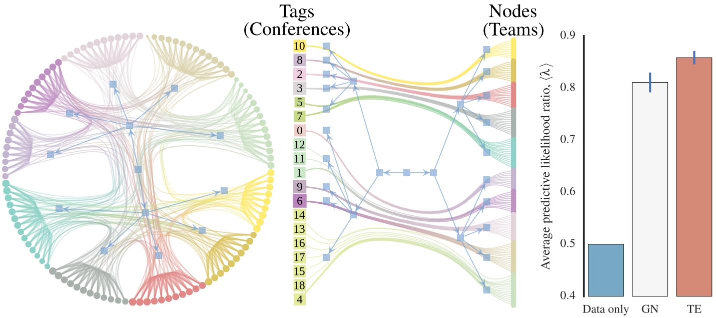
Generalized annotations

$$A_{ij} \to \text{Data layer}$$
$$T_{ij} \to \text{Annotation layer}$$



- Joint model for data and metadata.
- Arbitrary types of annotation.
- Both data and metadata are clustered into groups.
- Fully nonparametric.

Model variations: Annotated networks



(a) Data (b) Metadata (c) Node prediction

PREDICTION OF MISSING EDGES

$$G' = \underbrace{G}_{\text{Observed}} \cup \underbrace{\delta G}_{\text{Missing}}$$

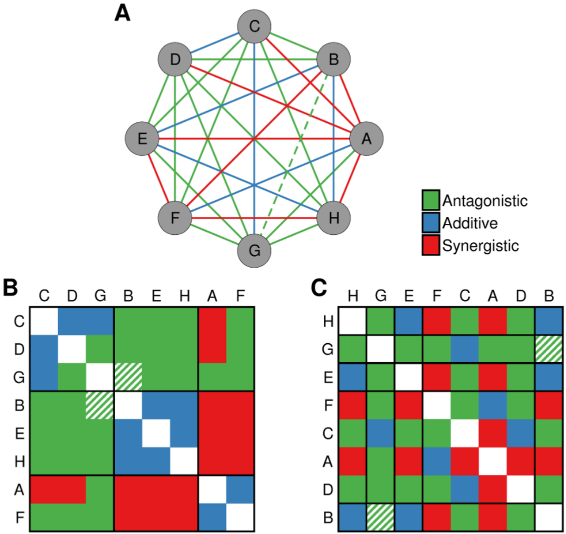
Bayesian probability of missing edges

$$P(\delta G|G,\{b_i\}) = \frac{\sum_\theta P(G \cup \delta G|\{b_i\},\theta)P(\theta)}{\sum_\theta P(G|\{b_i\},\theta)P(\theta)} = \exp(-\Delta\Sigma)$$

R. Guimerà, M Sales-Pardo, PNAS 2009; A. Clauset, C. Moore, MEJ Newman, Nature, 2008

Application: Drug-drug interactions

R Guimerà, M Sales-Pardo, PLoS Comput Biol, 2013



METADATA AND PREDICTION OF *missing nodes*

DARKO HRIC, T. P. P., SANTO FORTUNATO, ARXIV:1604.00255

Node probability, with known group membership:

$$P(\boldsymbol{a}_i|\boldsymbol{A}, b_i, \boldsymbol{b}) = \frac{\sum_\theta P(\boldsymbol{A}, \boldsymbol{a}_i|b_i, \boldsymbol{b}, \theta)P(\theta)}{\sum_\theta P(\boldsymbol{A}|\boldsymbol{b}, \theta)P(\theta)}$$

Node probability, with unknown group membership:

$$P(\boldsymbol{a}_i|\boldsymbol{A}, \boldsymbol{b}) = \sum_{b_i} P(\boldsymbol{a}_i|\boldsymbol{A}, b_i, \boldsymbol{b})P(b_i|\boldsymbol{b}),$$

Node probability, with unknown group membership, but known metadata:

$$P(\boldsymbol{a}_i|\boldsymbol{A}, \boldsymbol{T}, \boldsymbol{b}, \boldsymbol{c}) = \sum_{b_i} P(\boldsymbol{a}_i|\boldsymbol{A}, b_i, \boldsymbol{b})P(b_i|\boldsymbol{T}, \boldsymbol{b}, \boldsymbol{c}),$$

Group membership probability, given metadata:

$$P(b_i|\boldsymbol{T}, \boldsymbol{b}, \boldsymbol{c}) = \frac{P(b_i, \boldsymbol{b}|\boldsymbol{T}, \boldsymbol{c})}{P(\boldsymbol{b}|\boldsymbol{T}, \boldsymbol{c})} = \frac{\sum_\gamma P(\boldsymbol{T}|b_i, \boldsymbol{b}, \boldsymbol{c}, \gamma)P(b_i, \boldsymbol{b})P(\gamma)}{\sum_{b'_i}\sum_\gamma P(\boldsymbol{T}|b'_i, \boldsymbol{b}, \boldsymbol{c}, \gamma)P(b'_i, \boldsymbol{b})P(\gamma)}$$

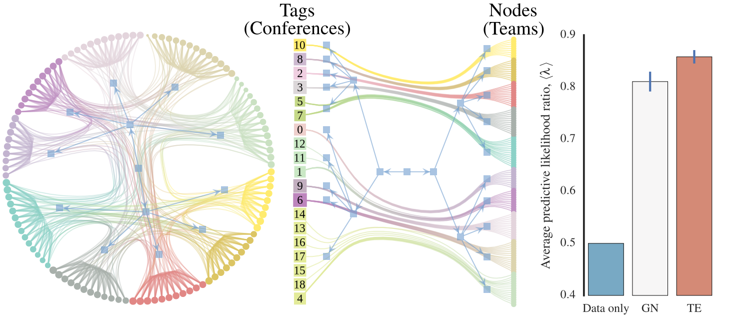
Predictive likelihood ratio:

$$\lambda_i = \frac{P(\boldsymbol{a}_i|\boldsymbol{A}, \boldsymbol{T}, \boldsymbol{b}, \boldsymbol{c})}{P(\boldsymbol{a}_i|\boldsymbol{A}, \boldsymbol{T}, \boldsymbol{b}, \boldsymbol{c}) + P(\boldsymbol{a}_i|\boldsymbol{A}, \boldsymbol{b})}$$

$\lambda_i > 1/2 \rightarrow$ the metadata improves the prediction task

Metadata and prediction of missing nodes

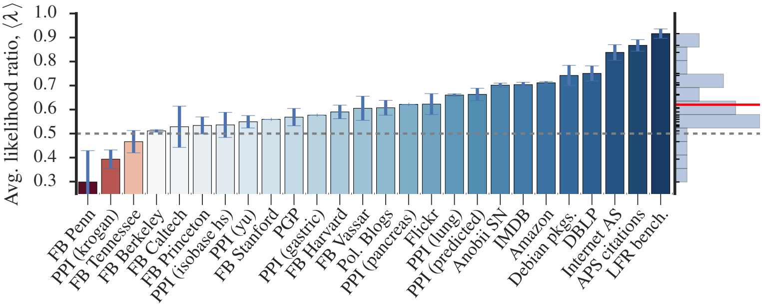
Darko Hric, T. P. P., Santo Fortunato, arXiv:1604.00255



(a) Data (b) Metadata (c) Node prediction

Metadata and prediction of missing nodes

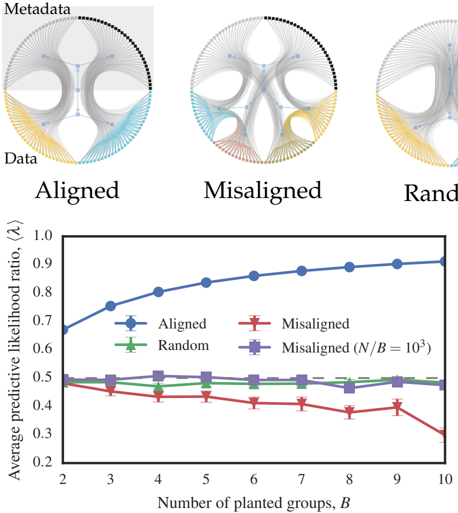
Darko Hric, T. P. P., Santo Fortunato, arXiv:1604.00255



$$\lambda_i = \frac{P(\boldsymbol{a}_i|\boldsymbol{A},\boldsymbol{T},\boldsymbol{b},\boldsymbol{c})}{P(\boldsymbol{a}_i|\boldsymbol{A},\boldsymbol{T},\boldsymbol{b},\boldsymbol{c}) + P(\boldsymbol{a}_i|\boldsymbol{A},\boldsymbol{b})}$$

Metadata and prediction of missing nodes

Darko Hric, T. P. P., Santo Fortunato, arXiv:1604.00255



METADATA AND PREDICTION OF MISSING NODES

DARKO HRIC, T. P. P., SANTO FORTUNATO, ARXIV:1604.00255

Neighbor probability:

$$P_e(i|j) = k_i \frac{e_{b_i,b_j}}{e_{b_i} e_{b_j}}$$

Neighbour probability, given metadata tag:

$$P_t(i) = \sum_j P(i|j) P_m(j|t)$$

Null neighbor probability (no metadata tag):

$$Q(i) = \sum_j P(i|j) \Pi(j)$$

Kullback-Leibler divergence:

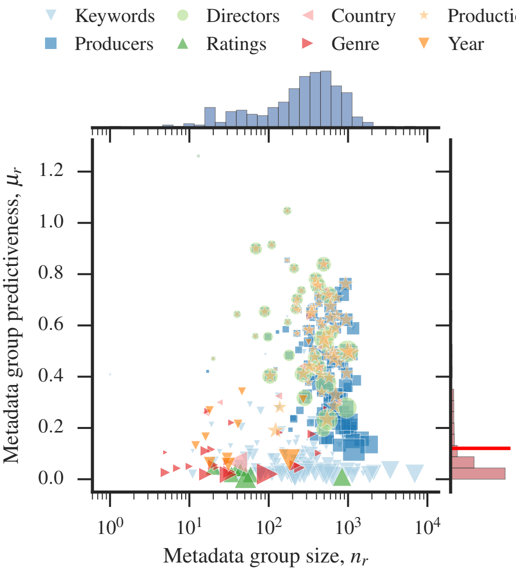
$$D_{\mathrm{KL}}(P_t||Q) = \sum_i P_t(i) \ln \frac{P_t(i)}{Q(i)}$$

Relative divergence:

$$\mu_r \equiv \frac{D_{\mathrm{KL}}(P_t||Q)}{H(q)} \to \text{Metadata group predictiveness}$$

Metadata and prediction of missing nodes

Darko Hric, T. P. P., Santo Fortunato, arXiv:1604.00255



Generalized community structure

M.E.J. Newman, T. P. Peixoto, Phys. Rev. Lett. 115, 088701 (2015)

Continuous generative models

- Nodes inhabit a latent-space: $x_u \in [0,1]$.

$$P(G|c,x) = \prod_{u<v} \omega(x_u,x_v)^{A_{uv}}(1-\omega(x_u,x_v))^{1-A_{uv}}$$

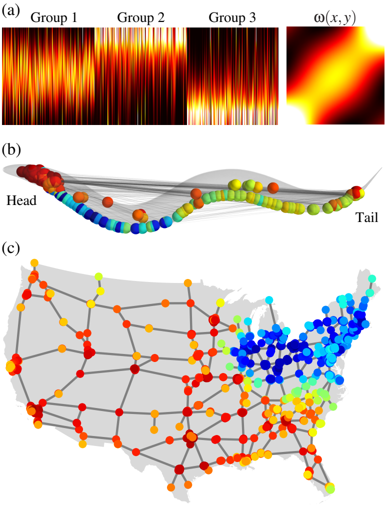
Polynomial expansion:

$$\omega(x,y) = \sum_{ij} c_{ij} B_i(x) B_j(y)$$

Bernstein basis:

$$B_i(x) = \binom{n}{i} x^i (1-x)^{n-i}$$

- Spatial information can be learned from network topology alone.
- No assortativity/homophily assumed!



Open problems

Knowns

- Better models...
- Mechanistic elements...

Known Unknowns

- Quality of fit?
- Network comparison?
- Cross-validation and bootstrapping?
- Combination of micro- and mesoscale structures?

Unknown Unknowns

- ?
- ?

The end

graph-tool

Very fast, freely available C++ code as part of the graph-tool Python library.

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