Network inference

Part 3

Tiago P. Peixoto

University of Bath

Tehran, August 2018

Weighted graphs

C. Aicher et al. Journal of Complex Networks 3(2), 221-248 (2015); T.P.P
arXiv: 1708.01432

Adjacency: $A_{ij} \in \{0,1\}$ or $\mathbb{N}$

Weights: $x_{ij} \in \mathbb{N}$ or $\mathbb{R}$

SBMs with edge covariates:

$$P(\boldsymbol{A},\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{b}) = P(\boldsymbol{x}|\boldsymbol{A},\boldsymbol{\gamma},\boldsymbol{b})P(\boldsymbol{A}|\boldsymbol{\theta},\boldsymbol{b})$$

Adjacency:

$$P(\boldsymbol{A}|\boldsymbol{\theta}=\{\boldsymbol{\lambda},\boldsymbol{\kappa}\},\boldsymbol{b}) = \prod_{i<j}\frac{\mathrm{e}^{-\lambda_{b_i,bj}\kappa_i\kappa_j}\left(\lambda_{b_i,bj}\kappa_i\kappa_j\right)^{A_{ij}}}{A_{ij}!},$$

Edge covariates:

$$P(\boldsymbol{x}|\boldsymbol{A},\boldsymbol{\gamma},\boldsymbol{b}) = \prod_{r\le s} P(\boldsymbol{x}_{rs}|\boldsymbol{\gamma}_{rs})$$

$P(\boldsymbol{x}|\boldsymbol{\gamma}) \to$ Exponential, Normal, Geometric, Binomial, Poisson, ...

Weighted graphs

T.P.P arXiv: 1708.01432

Nonparametric Bayesian approach

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

Marginal likelihood:

$$\begin{aligned}
P(\boldsymbol{A},\boldsymbol{x}|\boldsymbol{b}) &= \int P(\boldsymbol{A},\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{b})P(\boldsymbol{\theta})P(\boldsymbol{\gamma})\,\mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\gamma}\\
&= P(\boldsymbol{A}|\boldsymbol{b})P(\boldsymbol{x}|\boldsymbol{A},\boldsymbol{b}),
\end{aligned}$$

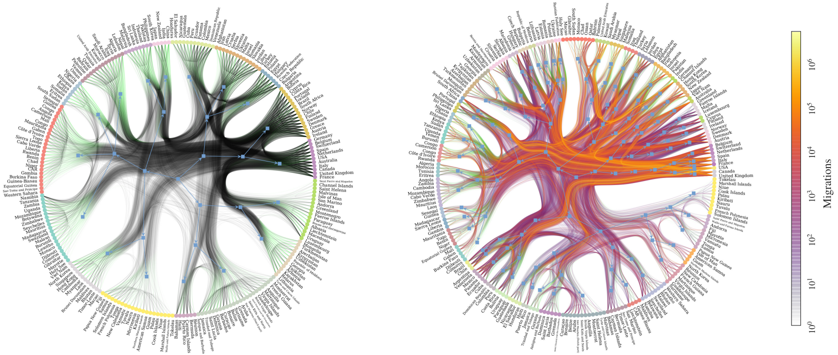
Adjacency part (unweighted):

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

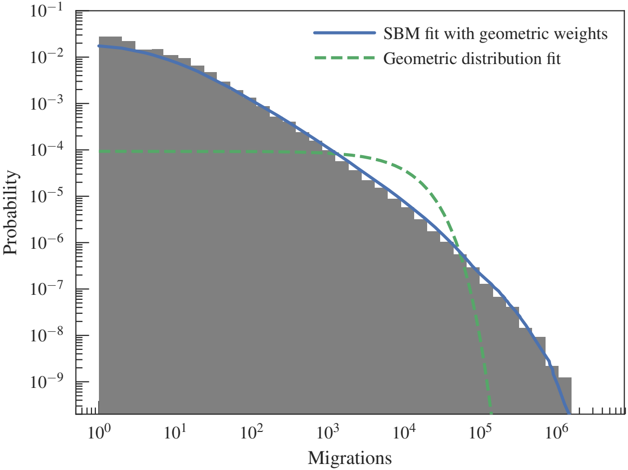
Weights part:

$$\begin{aligned}
P(\boldsymbol{x}|\boldsymbol{A},\boldsymbol{b}) &= \int P(\boldsymbol{x}|\boldsymbol{A},\boldsymbol{\gamma},\boldsymbol{b})P(\boldsymbol{\gamma})\,\mathrm{d}\boldsymbol{\gamma}\\
&= \prod_{r\le s}\int P(\boldsymbol{x}_{rs}|\boldsymbol{\gamma}_{rs})P(\boldsymbol{\gamma}_{rs})\,\mathrm{d}\boldsymbol{\gamma}_{rs}
\end{aligned}$$

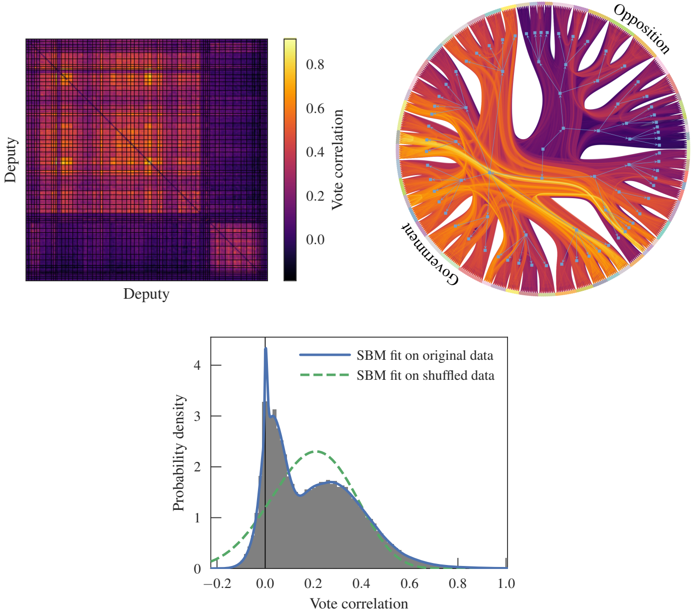
UN Migrations



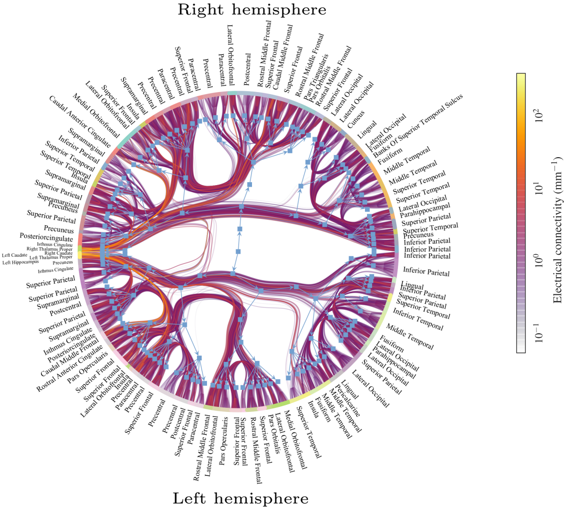
UN Migrations

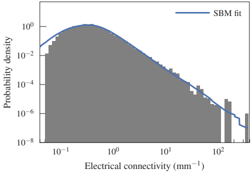


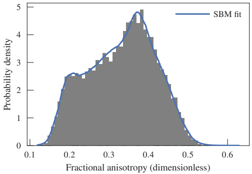
Votes in congress



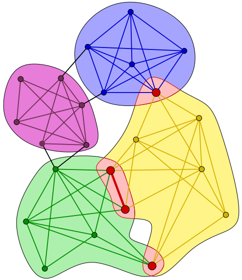
Human connectome





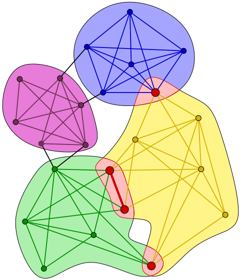


Overlapping groups

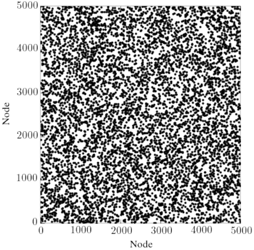


(Palla et al 2005)

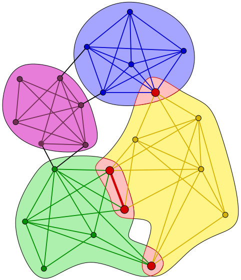
Overlapping groups



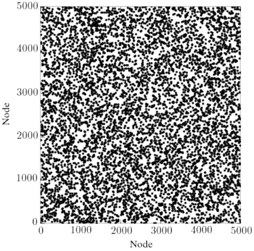
(Palla et al 2005)

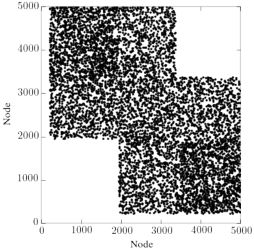


Overlapping groups



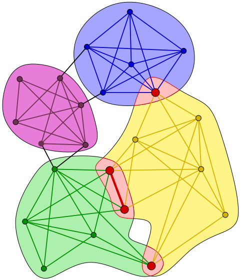
(Palla et al 2005)



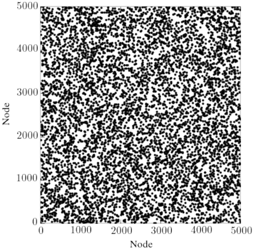


- Number of nonoverlapping partitions: B^N
- Number of overlapping partitions: 2^{BN}

Overlapping groups



(Palla et al 2005)



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

Group overlap

$$P(\boldsymbol{A}|\boldsymbol{\kappa},\boldsymbol{\lambda}) = \prod_{i<j}\frac{e^{-\lambda_{ij}}\lambda_{ij}^{A_{ij}}}{A_{ij}!} \times \prod_{i}\frac{e^{-\lambda_{ii}/2}(\lambda_{ii}/2)^{A_{ii}/2}}{A_{ii}/2!}, \quad \lambda_{ij} = \sum_{rs}\kappa_{ir}\lambda_{rs}\kappa_{js}$$

Labelled half-edges: $A_{ij} = \sum_{rs} G_{ij}^{rs}, \quad P(\boldsymbol{A}|\boldsymbol{\kappa},\boldsymbol{\lambda}) = \sum_{\boldsymbol{G}} P(\boldsymbol{G}|\boldsymbol{\kappa},\boldsymbol{\lambda})$

Group overlap

$$P(\boldsymbol{A}|\boldsymbol{\kappa},\boldsymbol{\lambda}) = \prod_{i<j}\frac{e^{-\lambda_{ij}}\lambda_{ij}^{A_{ij}}}{A_{ij}!}\times\prod_{i}\frac{e^{-\lambda_{ii}/2}(\lambda_{ii}/2)^{A_{ii}/2}}{A_{ii}/2!},\quad \lambda_{ij}=\sum_{rs}\kappa_{ir}\lambda_{rs}\kappa_{js}$$

$$\text{Labelled half-edges: } A_{ij}=\sum_{rs}G_{ij}^{rs},\quad P(\boldsymbol{A}|\boldsymbol{\kappa},\boldsymbol{\lambda})=\sum_{\boldsymbol{G}}P(\boldsymbol{G}|\boldsymbol{\kappa},\boldsymbol{\lambda})$$

$$\begin{aligned}P(\boldsymbol{G}) &= \int P(\boldsymbol{G}|\boldsymbol{\kappa},\boldsymbol{\lambda})P(\boldsymbol{\kappa})P(\boldsymbol{\lambda}|\bar{\lambda})\;\mathrm{d}\boldsymbol{\kappa}\mathrm{d}\boldsymbol{\lambda},\\ &= \frac{\bar{\lambda}^{E}}{(\bar{\lambda}+1)^{E+B(B+1)/2}}\frac{\prod_{r<s}e_{rs}!\prod_{r}e_{rr}!!}{\prod_{rs}\prod_{i<j}G_{ij}^{rs}!\prod_{i}G_{ii}^{rs}!!}\times\prod_{r}\frac{(N-1)!}{(e_r+N-1)!}\times\prod_{ir}k_i^r!,\end{aligned}$$

Group overlap

$$P(\boldsymbol{A}|\boldsymbol{\kappa},\boldsymbol{\lambda}) = \prod_{i<j}\frac{e^{-\lambda_{ij}}\lambda_{ij}^{A_{ij}}}{A_{ij}!}\times\prod_i\frac{e^{-\lambda_{ii}/2}(\lambda_{ii}/2)^{A_{ii}/2}}{A_{ii}/2!},\quad \lambda_{ij}=\sum_{rs}\kappa_{ir}\lambda_{rs}\kappa_{js}$$

Labelled half-edges: $A_{ij}=\sum_{rs}G_{ij}^{rs}$, $\quad P(\boldsymbol{A}|\boldsymbol{\kappa},\boldsymbol{\lambda})=\sum_{\boldsymbol{G}}P(\boldsymbol{G}|\boldsymbol{\kappa},\boldsymbol{\lambda})$

$$\begin{aligned}P(\boldsymbol{G}) &= \int P(\boldsymbol{G}|\boldsymbol{\kappa},\boldsymbol{\lambda})P(\boldsymbol{\kappa})P(\boldsymbol{\lambda}|\bar{\lambda})\,\mathrm{d}\boldsymbol{\kappa}\mathrm{d}\boldsymbol{\lambda},\\ &= \frac{\bar{\lambda}^E}{(\bar{\lambda}+1)^{E+B(B+1)/2}}\frac{\prod_{r<s}e_{rs}!\prod_r e_{rr}!!}{\prod_{rs}\prod_{i<j}G_{ij}^{rs}!\prod_i G_{ii}^{rs}!!}\times\prod_r\frac{(N-1)!}{(e_r+N-1)!}\times\prod_{ir}k_i^r!,\end{aligned}$$

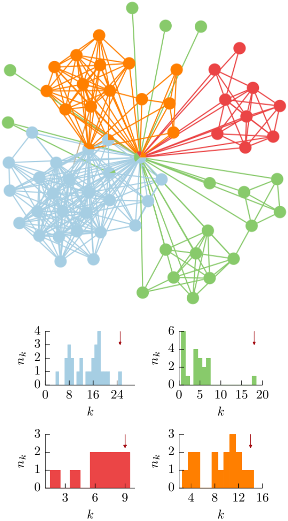
Microcanonical equivalence:

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

$$\begin{aligned}P(\boldsymbol{G}|\boldsymbol{k},\boldsymbol{e}) &= \frac{\prod_{r<s}e_{rs}!\prod_r e_{rr}!!\prod_{ir}k_i^r!}{\prod_{rs}\prod_{i<j}G_{ij}^{rs}!\prod_i G_{ii}^{rs}!!\prod_r e_r!},\\ P(\boldsymbol{k}|\boldsymbol{e}) &= \prod_r \left(\!\!\binom{e_r}{N}\!\!\right)^{-1}\end{aligned}$$

Overlap vs. non-overlap

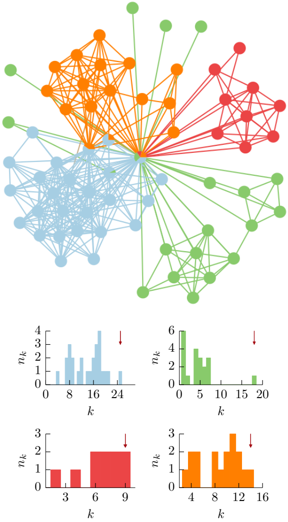
Social "ego" network (from Facebook)



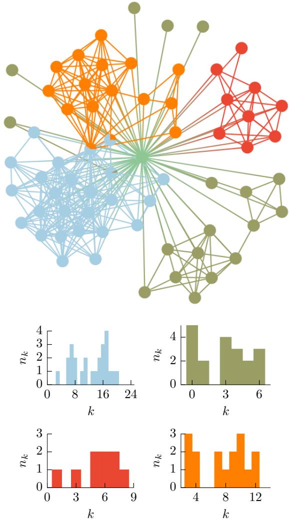
$B = 4,\ \Lambda \simeq 0.053$

Overlap vs. non-overlap

Social "ego" network (from Facebook)

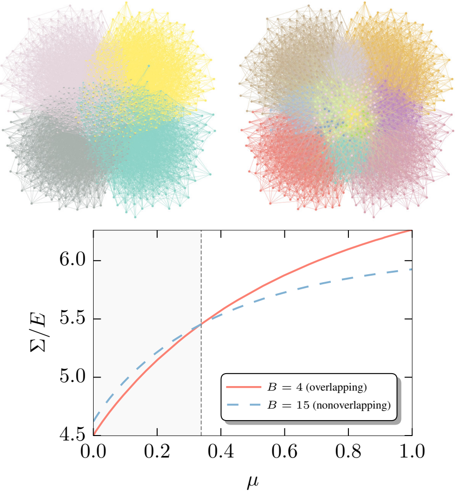


$B = 4,\ \Lambda \simeq 0.053$



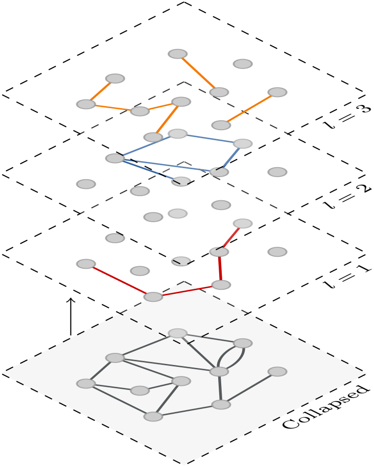
$B = 5,\ \Lambda = 1$

Overlap vs. non-overlap



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

Edge covariates

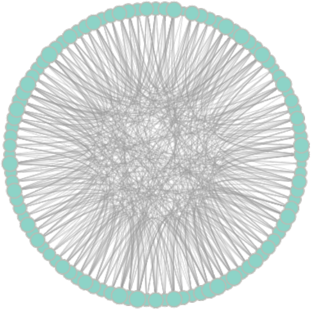
$$P(\{\boldsymbol{A}_l\}|\{\boldsymbol{\theta}\}) = P(\boldsymbol{A}_c|\{\boldsymbol{\theta}\}) \prod_{r\leq s} \frac{\prod_l m_{rs}^l!}{m_{rs}!}$$

Independent layers

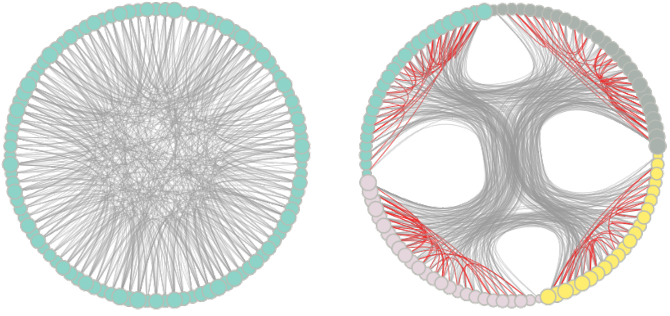
$$P(\{\boldsymbol{A}_l\}|\{\{\boldsymbol{\theta}\}_l\},\{\boldsymbol{\phi}\},\{\boldsymbol{z_{il}}\}\}) = \prod_l P(\boldsymbol{A}_l|\{\boldsymbol{\theta}\}_l,\{\boldsymbol{\phi}\})$$

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

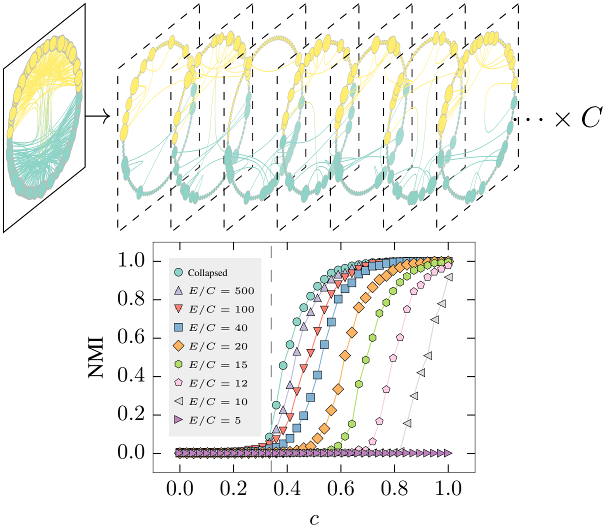
Layer information can reveal hidden structure



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)

Model selection

Example: Social network of physicians

$N = 241$ Physicians

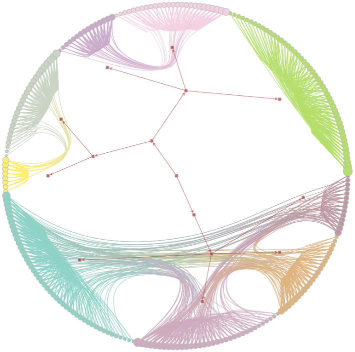
Survey questions:

- "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)

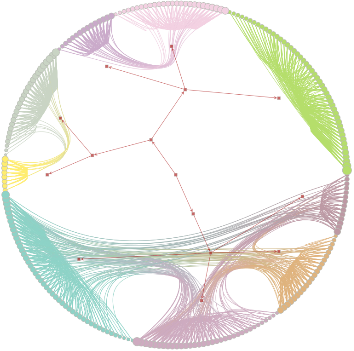
Model selection

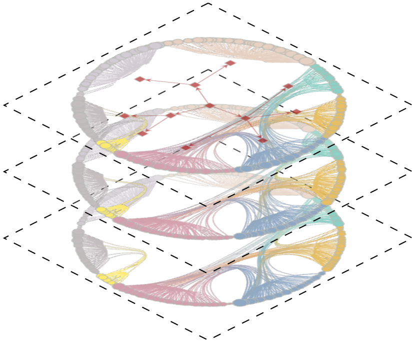
Example: Social network of physicians



Model selection

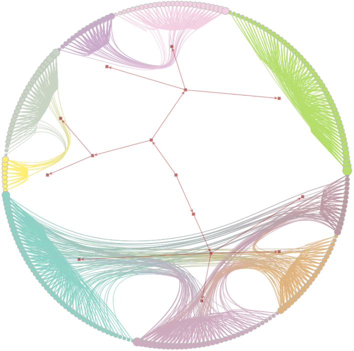
Example: Social network of physicians



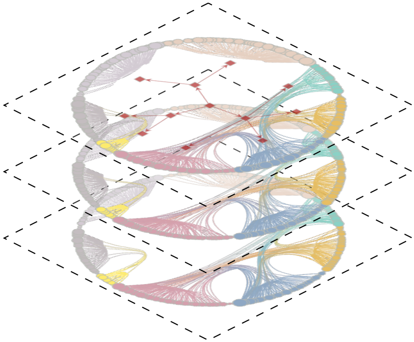


Model selection

Example: Social network of physicians



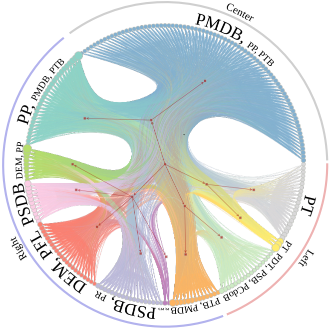
$\Lambda = 1$



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

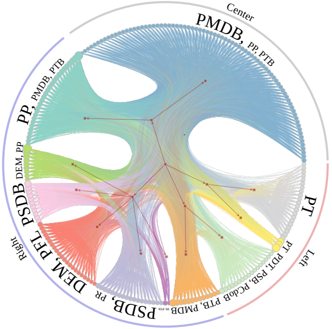
Example: Brazilian chamber of deputies

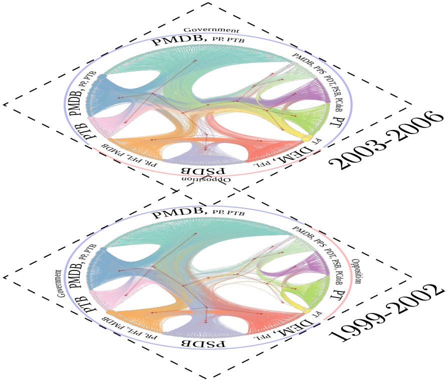
Voting network between members of congress (1999-2006)



Example: Brazilian chamber of deputies

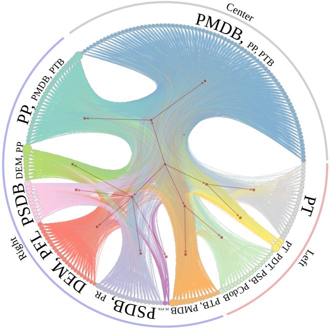
Voting network between members of congress (1999-2006)



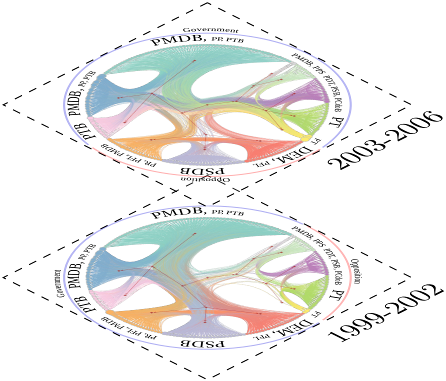


Example: Brazilian chamber of deputies

Voting network between members of congress (1999-2006)



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



$\Lambda = 1$

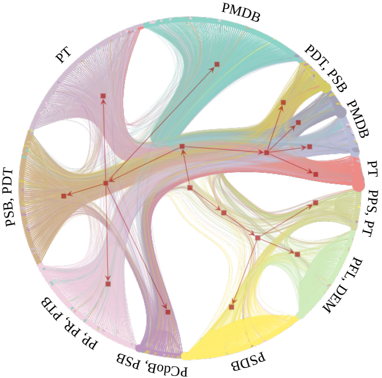
Real-valued edges?

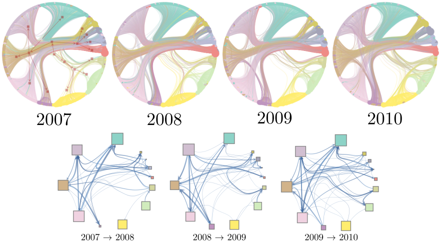
Idea: Layers $\{\ell\}$ $\to$ bins of edge values!

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

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

Movement between groups...





Networks with metadata

Many network datasets contain *metadata*: Annotations that go beyond the mere adjacency between nodes.

Often assumed as indicators of topological structure, and used to *validate* community detection methods. A.k.a. “ground-truth”.

Example: American college football

Metadata (Conferences)

Example: American college football

SBM fit

Example: American college football

Discrepancy

Example: American college football

Discrepancy

Why the discrepancy?

Some hypotheses:

Example: American college football

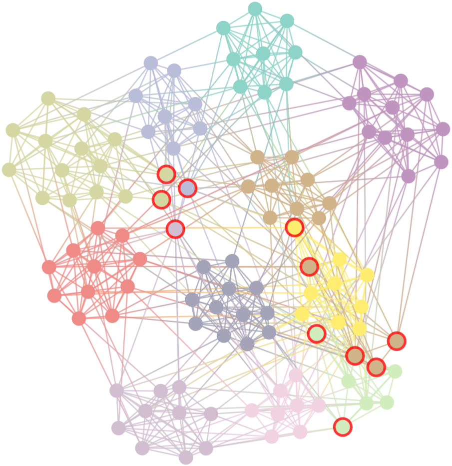
Discrepancy

Why the discrepancy?

Some hypotheses:

- The model is not sufficiently descriptive.

Example: American college football



Discrepancy

Why the discrepancy?

Some hypotheses:

- The model is not sufficiently descriptive.
- The metadata is not sufficiently descriptive or is inaccurate.

Example: American college football

Discrepancy

Why the discrepancy?

Some hypotheses:

- The model is not sufficiently descriptive.
- The metadata is not sufficiently descriptive or is inaccurate.
- Both.

Example: American college football

Discrepancy

Why the discrepancy?

Some hypotheses:

- The model is not sufficiently descriptive.
- The metadata is not sufficiently descriptive or is inaccurate.
- Both.
- Neither.

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.

Metadata is often very heterogeneous

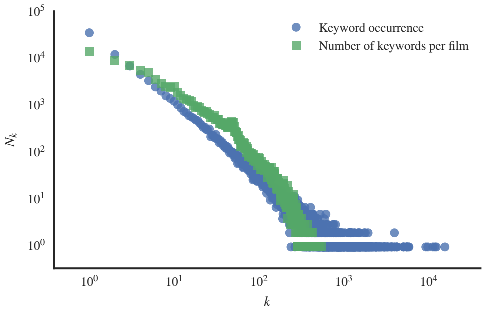
Example: IMDB Film-Actor network

Data: $96,982$ Films, $275,805$ Actors, $1,812,657$ Film-Actor Edges

Film metadata: Title, year, genre, production company, country, user-contributed keywords, etc.

Actor metadata: Name, Age, Gender, Nationality, etc.

User-contributed keywords $(93,448)$



Metadata is often very heterogeneous

Example: IMDB Film-Actor network

Keyword	Occurrences
'independent-film'	15513
'based-on-novel'	12303
'character-name-in-title'	11801
'murder'	11184
'sex'	9759
'female-nudity'	9239
'nudity'	5846
'death'	5791
'husband-wife-relationship'	5568
'love'	5560
'violence'	5480
'police'	5463
'father-son-relationship'	5063

Metadata is often very heterogeneous

Example: IMDB Film-Actor network

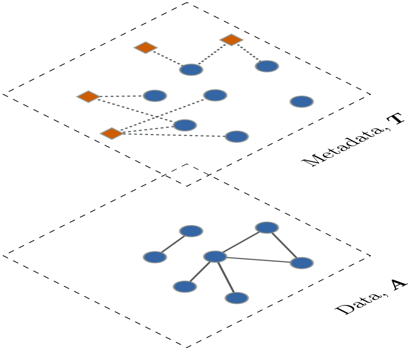
Keyword	Occurrences	Keyword	Occurrences
'independent-film'	15513	'discriminaton-against-anteaters'	1
'based-on-novel'	12303	'partisan-violence'	1
'character-name-in-title'	11801	'deliberately-leaving-something-behind'	1
'murder'	11184	'princess-from-outer-space'	1
'sex'	9759	'reference-to-aleksei-vorobyov'	1
'female-nudity'	9239	'dead-body-on-the-beach'	1
'nudity'	5846	'liver-failure'	1
'death'	5791	'hit-with-a-skateboard'	1
'husband-wife-relationship'	5568	'helping-blind-man-cross-street'	1
'love'	5560	'abandoned-pet'	1
'violence'	5480	'retired-clown'	1
'police'	5463	'resentment-toward-stepson'	1
'father-son-relationship'	5063	'mutilating-a-plant'	1

Better approach: Metadata as data

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

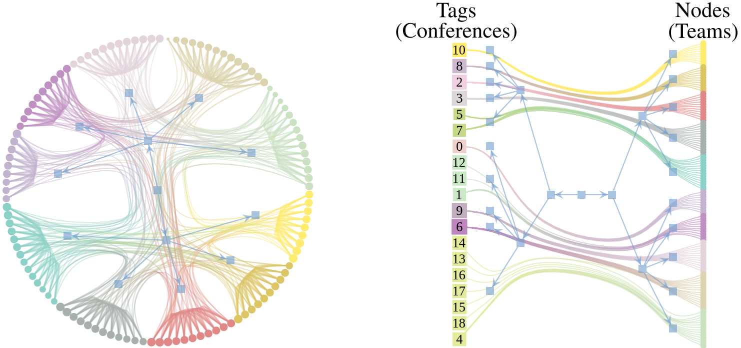
Generalized annotations

$A_{ij} \rightarrow$ Data layer
$T_{ij} \rightarrow$ Annotation layer



- Joint model for data and metadata (the layered SBM [1]).
- Arbitrary types of annotation.
- Both data and metadata are clustered into groups.
- Fully nonparametric.

Example: American college football



(a) Data (b) Metadata

Prediction of missing edges

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

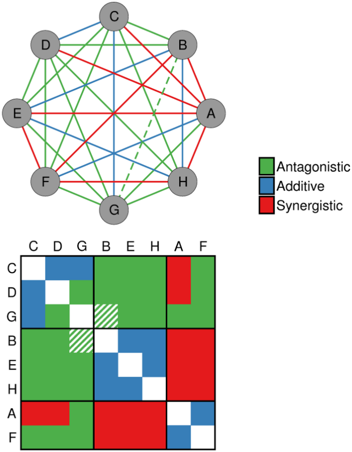
Posterior probability of missing edges

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

A. Clauset, C. Moore, MEJ Newman, Nature, 2008

R. Guimerà, M Sales-Pardo, PNAS 2009

Drug-drug interactions



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

METADATA AND PREDICTION OF *missing nodes*

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:

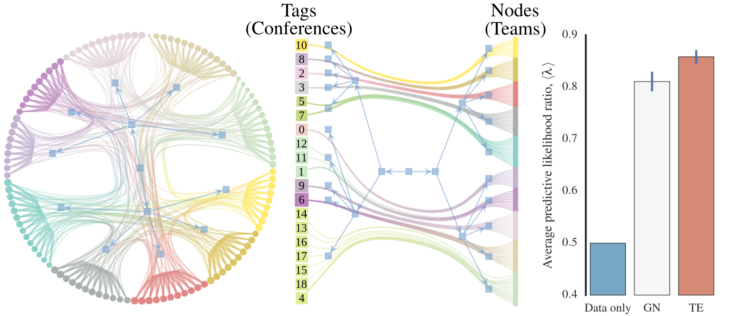
$$P(b_i|\boldsymbol{T}, \boldsymbol{b}, \boldsymbol{c}) = \frac{P(b_i, \boldsymbol{b}|\boldsymbol{T}, \boldsymbol{c})}{P(\boldsymbol{b}|\boldsymbol{T}, \boldsymbol{c})} = \frac{\sum_\gamma P(\boldsymbol{T}|b_i, \boldsymbol{b}, \boldsymbol{c}, \gamma)P(b_i, \boldsymbol{b})P(\gamma)}{\sum_{b_i'}\sum_\gamma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

Predictive likelihood ratio:

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

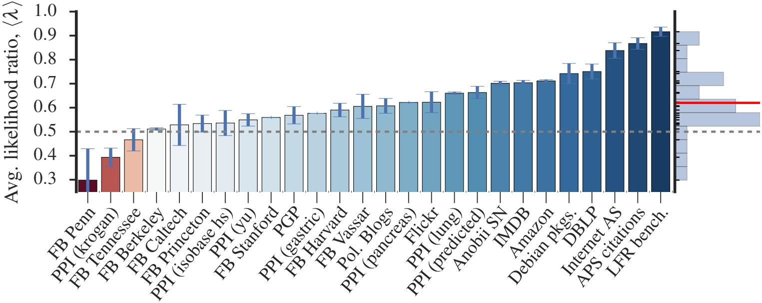
$\lambda_i > 1/2 \rightarrow$ the metadata improves the prediction task

Metadata and prediction of missing nodes



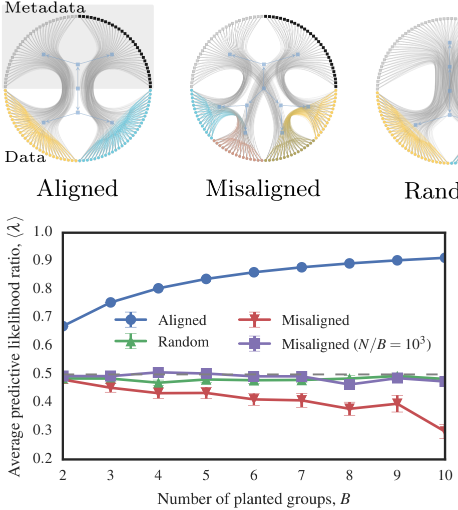
(a) Data (b) Metadata (c) Node prediction

Metadata and prediction of missing nodes



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

Metadata and prediction of missing nodes



Metadata predictiveness

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):

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

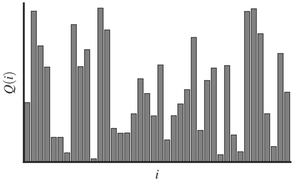
Kullback-Leibler divergence:

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

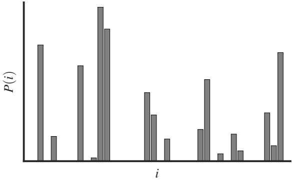
Relative divergence:

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

Neighbour prob. without metadata

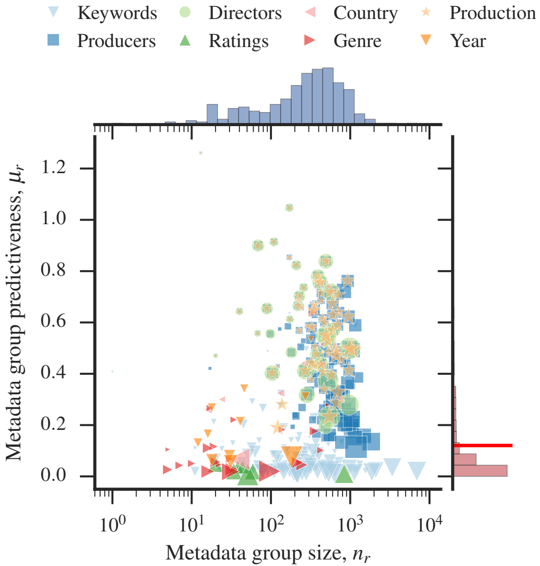


Neighbour prob. with metadata



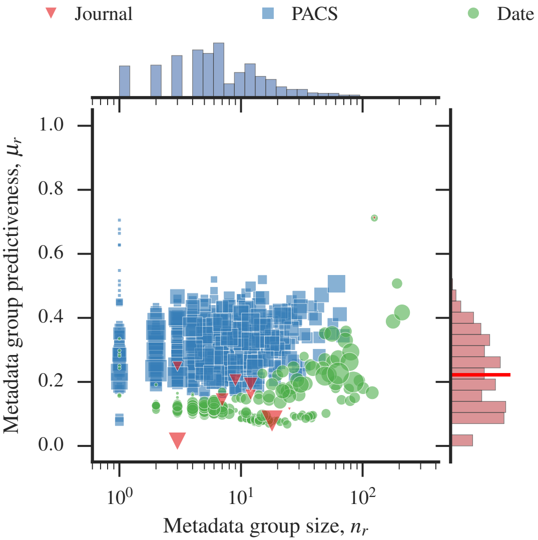
Metadata predictiveness

IMDB film-actor network



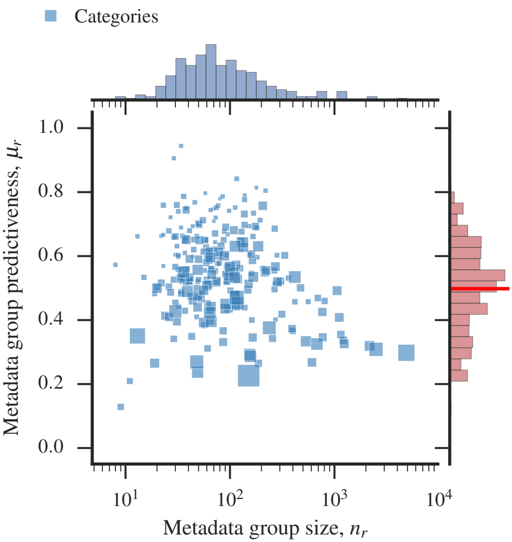
Metadata predictiveness

APS citation network



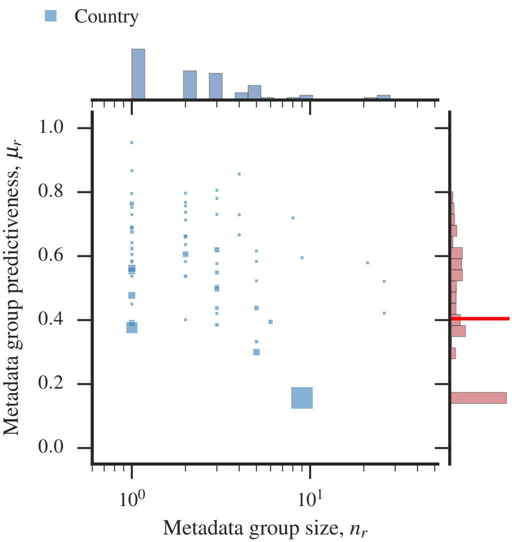
Metadata predictiveness

Amazon co-purchases



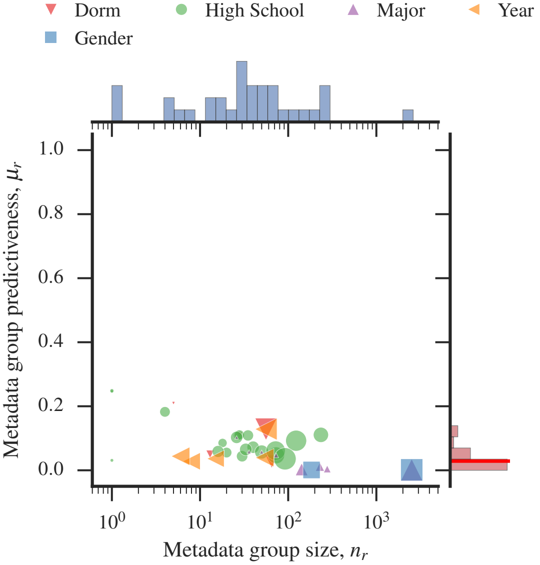
Metadata predictiveness

Internet AS



Metadata predictiveness

Facebook Penn State



n-order Markov chains with communities

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

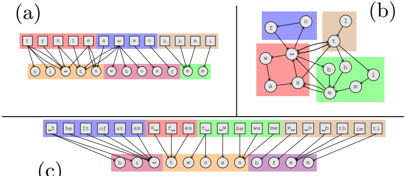
Transitions conditioned on the last n tokens

$p(x_t|\vec{x}_{t-1}) \rightarrow$ 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:

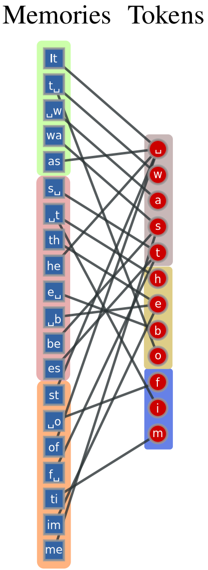
$$p(x|\vec{x}) = \theta_x \lambda_{b_x b_{\vec{x}}}$$

$\theta_x \rightarrow$ 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



$\{x_t\} =$ `"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 transition graph.

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

T. P. P. and Martin Rosvall, Nature Communications (in press)

Bayesian formulation

$$P(\{x_t\}|b) = \int \mathrm{d}\theta\,\mathrm{d}\lambda\, P(\{x_t\}|b,\lambda,\theta)\prod_r \mathcal{D}_r(\{\theta_x\})\prod_s \mathcal{D}_s(\{\lambda_{rs}\})$$

Noninformative priors $\rightarrow$ Microcanonical model

$$P(\{x_t\}|b) = P(\{x_t\}|b,\{e_{rs}\},\{k_x\}) \times P(\{k_x\}|\{e_{rs}\},b) \times P(\{e_{rs}\}),$$

where

$$\begin{aligned} P(\{x_t\}|b,\{e_{rs}\},\{k_x\}) &\rightarrow \text{Sequence likelihood},\\ P(\{k_x\}|\{e_{rs}\},b) &\rightarrow \text{Token frequency likelihood},\\ P(\{e_{rs}\}) &\rightarrow \text{Transition count likelihood}, \end{aligned}$$

$-\ln P(\{x_t\},b) \rightarrow$ ***Description length*** of the sequence

Inference $\leftrightarrow$ Compression

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!)

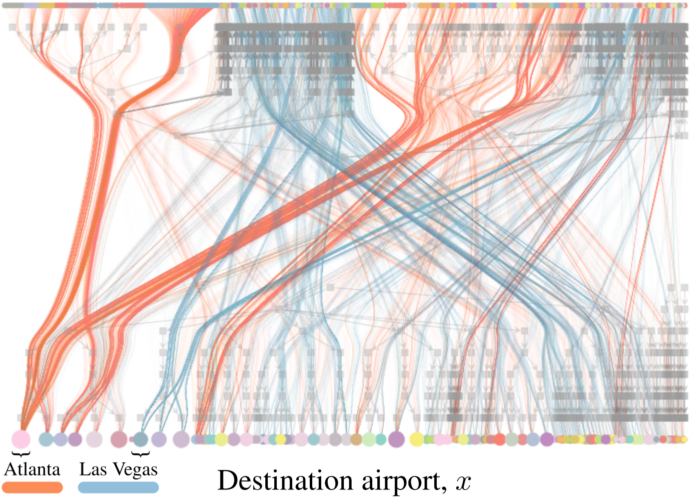
T. P. P. and Martin Rosvall, arXiv: 1509.04740

n-order Markov chains with communities

Example: Flight itineraries

$$\vec{x}_t = \{x_{t-3}, \text{Altanta}|\text{Las Vegas}, x_{t-1}\}$$

Previous $n = 3$ airports, $\vec{x}$



Destination airport, x

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 of pairs of group labels)

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

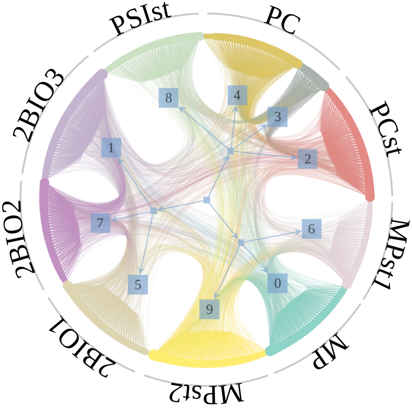
(static SBM generating edges from group labels)

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

Dynamic networks

Example: Student proximity

Static part

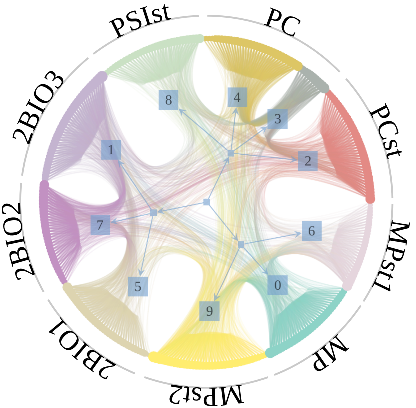


T. P. P. and Martin Rosvall, arXiv: 1509.04740

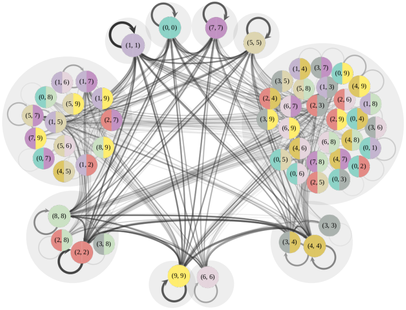
Dynamic networks

Example: Student proximity

Static part



Temporal part



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

Dynamic networks in 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

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

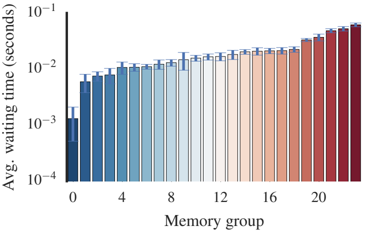
Hyperparameters: α, β. Noninformative limit $\alpha \to 0$, $\beta \to 0$ leads to Jeffreys prior: $P(\lambda) \propto \frac{1}{\lambda}$

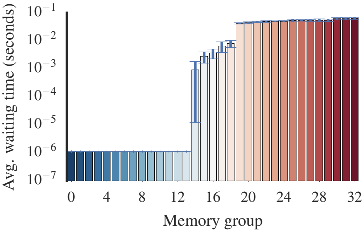
T. P. P. and Martin Rosvall, arXiv: 1509.04740

Dynamic networks

Continuous time

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





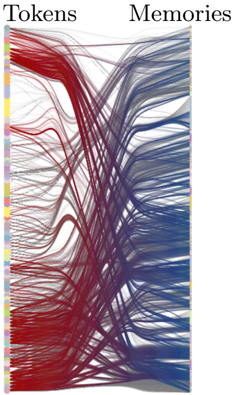
Without waiting times ($n = 1$)

With waiting times ($n = 2$)

Nonstationarity Dynamic networks

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

Unmodified chain

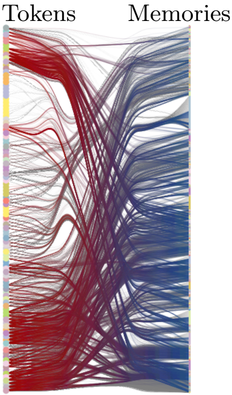


$$-\log_2 P(\{x_t\}, b) = 7,450,322$$

Nonstationarity Dynamic networks

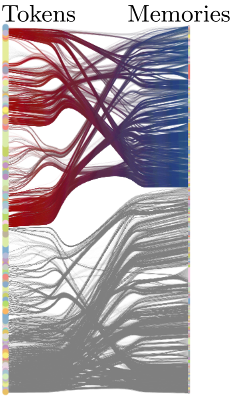
$\{x_t\} \to$ Concatenation of "War and peace," by Leo Tolstoy, and "À la recherche du temps perdu," by Marcel Proust.

Unmodified chain



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

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



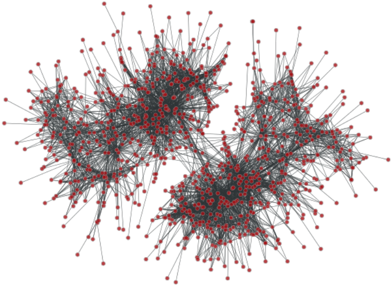
$$-\log_2 P(\{x_t\}, b) = 7,146,465$$

Latent space models

P. D. Hoff, A. E. Raferty, and M. S. Handcock, J. Amer. Stat. Assoc. 97, 1090–1098 (2002)

$$P(G|\{\vec{x}_i\}) = \prod_{i>j} p_{ij}^{A_{ij}}(1-p_{ij})^{1-A_{ij}}$$

$$p_{ij} = \exp\left(-(\vec{x}_i - \vec{x}_j)^2\right).$$



(Human connectome)

Many other more elaborate embeddings (e.g. hyperbolic spaces).
Properties:

- *Softer approach:* Nodes are not placed into discrete categories.
- Exclusively *assortative* structures.
- Formulation for directed graphs less trivial.

Discrete vs. continuous

Can we formulate a unified parametrization?

The Graphon

$$P(G|\{x_i\}) = \prod_{i>j} p_{ij}^{A_{ij}} (1 - p_{ij})^{1-A_{ij}}$$

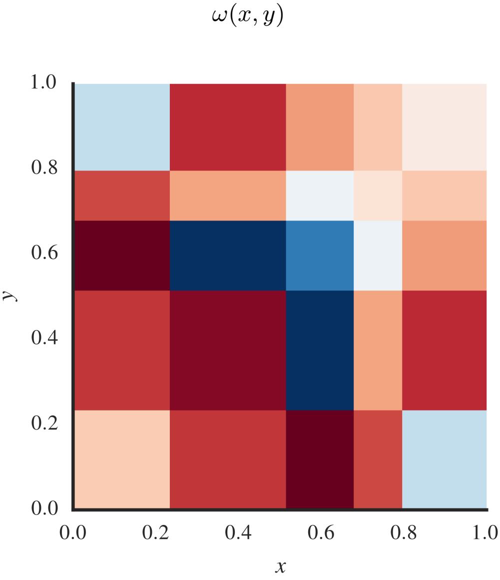
$$p_{ij} = \omega(x_i, x_j)$$

$$x_i \in [0, 1]$$

Properties:

- Mostly a theoretical tool.
- Cannot be directly inferred (without massively overfitting).
- Needs to be parametrized to be practical.

The SBM → a piecewise-constant Graphon



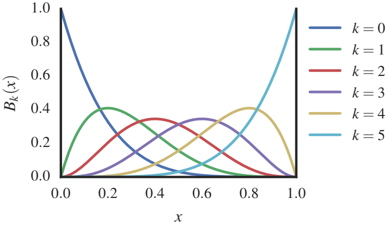
A "soft" graphon parametrization

$$p_{uv} = \frac{d_u d_v}{2m}\omega(x_u, x_v)$$

$$\omega(x,y) = \sum_{j,k=0}^{N} c_{jk} B_j(x) B_k(y)$$

Bernstein polynomials:

$$B_k(x) = \binom{N}{k} x^k (1-x)^{N-k}, \qquad k = 0 \dots N$$



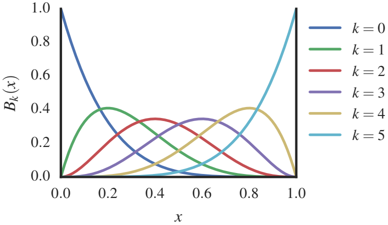
A "soft" graphon parametrization

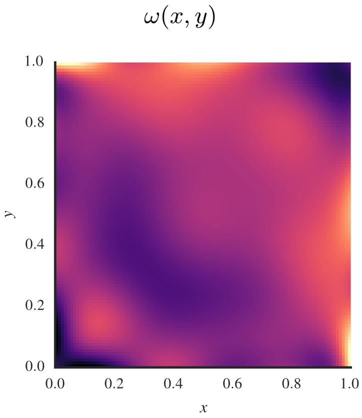
$$p_{uv} = \frac{d_u d_v}{2m} \omega(x_u, x_v)$$

$$\omega(x, y) = \sum_{j,k=0}^{N} c_{jk} B_j(x) B_k(y)$$

Bernstein polynomials:

$$B_k(x) = \binom{N}{k} x^k (1-x)^{N-k}, \qquad k = 0 \dots N$$





Inferring the model

Semi-parametric Bayesian approach

Expectation-Maximization algorithm

1. Expectation step

$$q(\mathbf{x}) = \frac{P(\mathbf{A},\mathbf{x}|\mathbf{c})}{\int P(\mathbf{A},\mathbf{x}|\mathbf{c})\mathrm{d}^n\mathbf{x}}$$

2. Maximization step

$$P(\mathbf{A}|\mathbf{c}) = \int P(\mathbf{A},\mathbf{x}|\mathbf{c})\mathrm{d}^n\mathbf{x}$$

$$\hat{c}_{jk} = \underset{c_{jk}}{\mathrm{argmax}}\, P(\mathbf{A}|\mathbf{c})$$

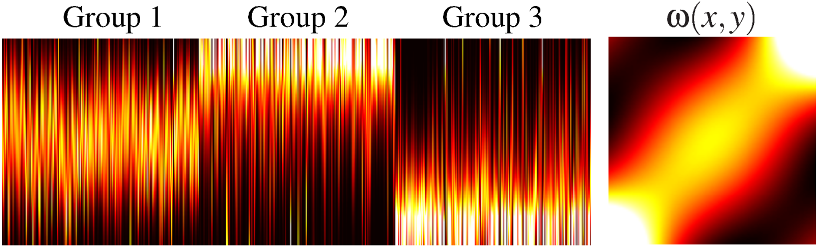
Belief-Propagation

$$\eta_{u\to v}(x) = \frac{1}{Z_{u\to v}} \exp\biggl(-\sum_w d_u d_w \int_0^1 q_w(y)\omega(x,y)\mathrm{d}y\biggr) \times \prod_{\substack{w(\neq v)\\ a_{uw}=1}} \int_0^1 \eta_{w\to u}(y)\omega(x,y)\mathrm{d}y,$$

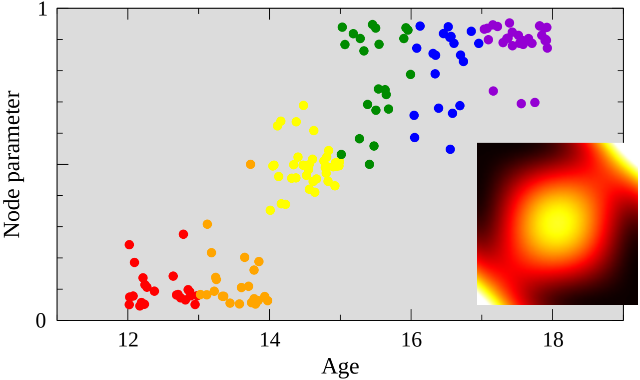
$$q_{uv}(x,y) = \frac{\eta_{u\to v}(x)\eta_{v\to u}(y)\omega(x,y)}{\iint_0^1 \eta_{u\to v}(x)\eta_{v\to u}(y)\omega(x,y)\mathrm{d}x\mathrm{d}y}.$$

Algorithmic complexity: $O(mN^2)$

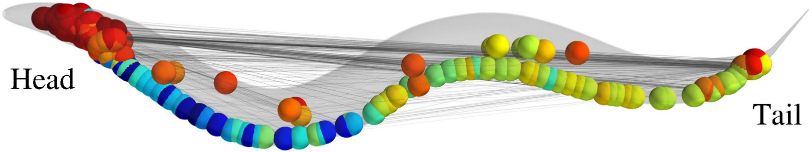
Example: SBM sample



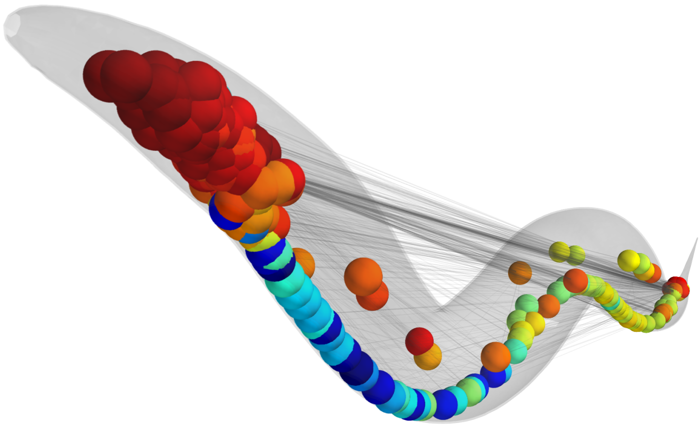
Example: School friendships



Example: C. elegans worm



Example: C. elegans worm



Example: Interstate highway