Generalized Communities in Networks

Tiago P. Peixoto

Universität Bremen
Germany

ISI Foundation
Turin, Italy

M. E. J. Newman

University of Michigan & Santa Fe Institute
USA

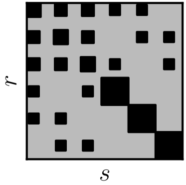
Seoul, June 2016

The stochastic block model (SBM)

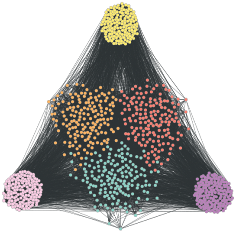
P. W. Holland et al., Soc Networks 5, 109 (1983)

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.



$\rightarrow$



Properties:

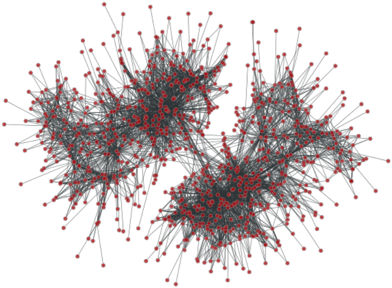
- *General* model for large-scale structures. Traditional assortative communities are a special case, but it also admits arbitrary mixing patterns (e.g. bipartite, core-periphery, etc).
- Formulation for directed graphs is trivial.

Latent space models

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

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

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



(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

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

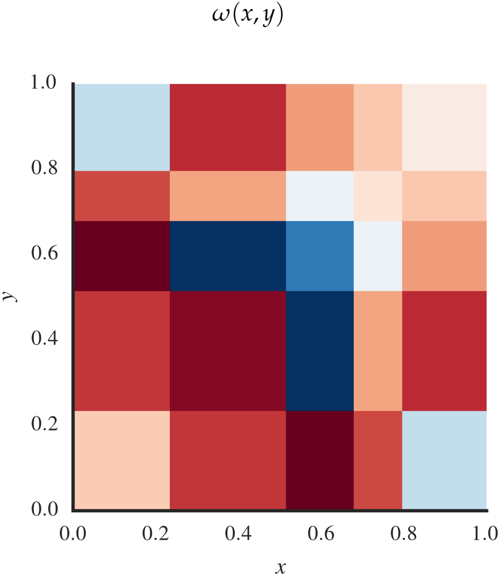
$$p_{ij} = \omega(x_i, x_j)$$

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

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



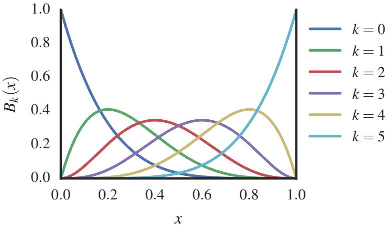
A "soft" graphon parametrization

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

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

Bernstein polynomials:

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



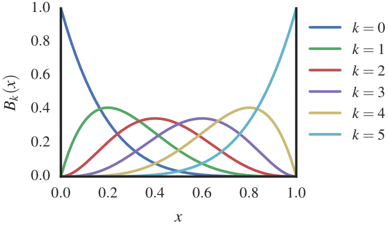
A "SOFT" GRAPHON PARAMETRIZATION

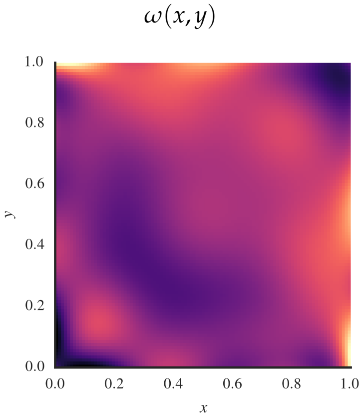
$$p_{uv} = \frac{d_u d_v}{2m}\omega(x_u, x_v)$$

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

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

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

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

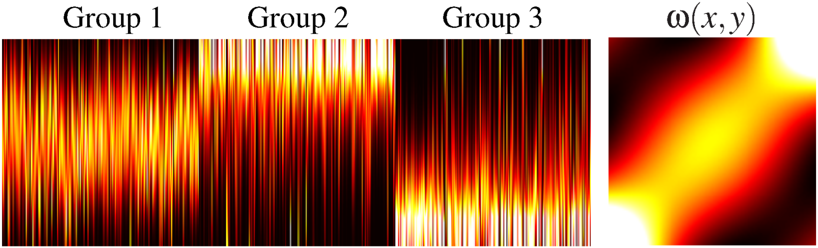
Belief-Propagation

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

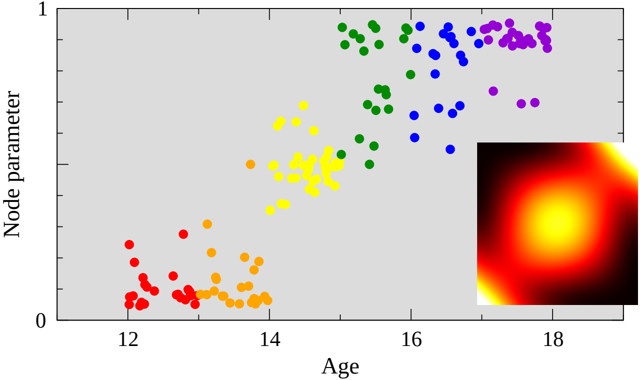
$$q_{uv}(x,y) = \frac{\eta_{u\to v}(x)\eta_{v\to u}(y)\omega(x,y)}{\iint_0^1 \eta_{u\to v}(x)\eta_{v\to u}(y)\omega(x,y)\mathrm{d}x\mathrm{d}y}.$$

Algorithmic complexity: $O(mN^2)$

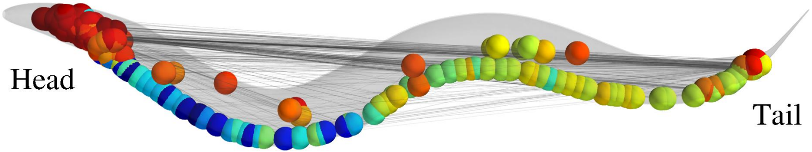
Example: SBM sample



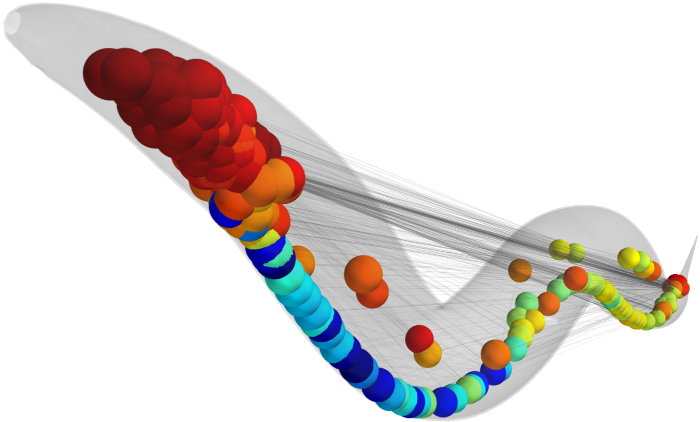
Example: School friendships



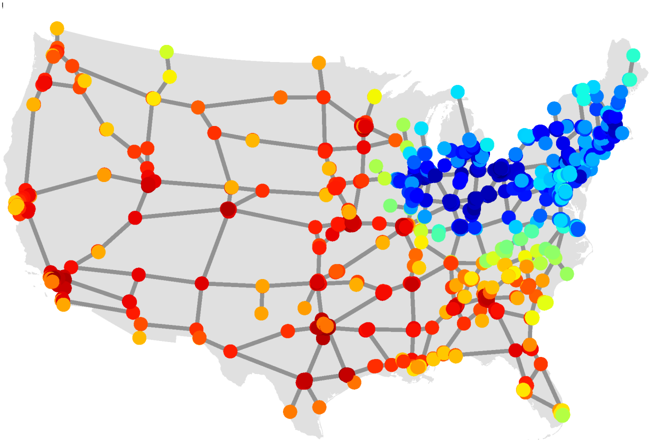
Example: C. elegans worm



Example: C. elegans worm



Example: Interstate highway



The End

Main message:

- "Soft" generative model.
- Continuous latent space.
- Arbitrary mixing patterns.

Todo:

- Fully Bayesian inference & model selection.
- More spatial dimensions.

M.E.J. Newman, T. P. Peixoto, PRL 115, 088701 (2015)