Network analysis with the $W$-graph model
(via the Stochastic Block Model)

S. Robin

Joint work with P. Latouche and S. Ouadah

INRA / AgroParisTech

IMS, June 2015, Singapore
1. Modeling heterogeneity in interaction networks

2. Statistical inference of latent space models (focus on SBM)

3. From SBM to $W$-graph: Averaging models

4. Goodness-of-fit
Modeling heterogeneity in (biological) interaction networks
Heterogeneity in biological networks

Biological networks describe interactions between entities: genes, proteins, individuals, species...

Observed networks display heterogeneous topologies, that one would like to decipher and better understand.
Heterogeneity in biological networks

Biological networks describe interactions between entities: genes, proteins, individuals, species...

Observed networks display heterogeneous topologies, that one would like to decipher and better understand.

Dolphine social network.

[Newman and Girvan (2004)]

H. pylori PPI network.
Heterogeneous means ...

... not homogeneous, that is: different from an Erdös-Renyi (ER) graph.

Erdös-Renyi random graph $G(n, p)$: Consider $n$ nodes, node pairs $1 \leq i < j \leq n$ are independently connected with same probability $p$:

$$(Y_{ij}) \text{ iid, } Y_{ij} \sim \mathcal{B}(p).$$

- Very intensively studied.
- Fits very few real-life networks.
Latent space models

Latent variables allow to capture some underlying structure of a network (see review [Matias and R. (2014)]).
Latent space models

Latent variables allow to capture some underlying structure of a network (see review [Matias and R. (2014)]).

General setting for binary graphs. [Bollobás et al. (2007)]:

- A latent (unobserved) variable $Z_i$ is associated with each node:

  $$\{Z_i\} \text{ iid } \sim \pi$$

- Edges $Y_{ij} = \mathbb{I}\{i \sim j\}$ are independent conditionally to the $Z_i$'s:

  $$\{Y_{ij}\} \text{ independent } \mid \{Z_i\} : \Pr\{Y_{ij} = 1\} = \gamma(Z_i, Z_j)$$
Latent space models

Latent variables allow to capture some underlying structure of a network (see review [Matias and R. (2014)]).

General setting for binary graphs. [Bollobás et al. (2007)]:

- A latent (unobserved) variable $Z_i$ is associated with each node:
  \[
  \{Z_i\} \text{ iid } \sim \pi
  \]

- Edges $Y_{ij} = \mathbb{I}\{i \sim j\}$ are independent conditionally to the $Z_i$'s:
  \[
  \{Y_{ij}\} \text{ independent } \big| \{Z_i\} : \Pr\{Y_{ij} = 1\} = \gamma(Z_i, Z_j)
  \]

We focus here on model approaches, in contrast with, e.g.

- Spectral clustering [von Luxburg et al. (2008)].
Latent space models

State-space model: principle.
Latent space models

State-space model: principle.

- Consider $n$ nodes ($i = 1..n$);
Latent space models

State-space model: principle.

- Consider \( n \) nodes \((i = 1..n)\);
- \( Z_i \) = unobserved position of node \( i \),
  e.g.

\[
\{Z_i\} \sim \mathcal{N}(0, I)
\]
Latent space models

State-space model: principle.

- Consider \( n \) nodes \((i = 1..n)\);

- \( Z_i \) = unobserved position of node \( i \), e.g.
  \[
  \{Z_i\} \text{ iid } \sim \mathcal{N}(0, I)
  \]

- Edge \( \{Y_{ij}\} \) independent given \( \{Z_i\} \), e.g.
  \[
  \Pr\{Y_{ij} = 1\} = \gamma(Z_i, Z_j).
  \]
Latent space models

State-space model: principle.

- Consider $n$ nodes ($i = 1..n$);
- $Z_i =$ unobserved position of node $i$, e.g.
  \[ \{Z_i\} \text{iid } \sim \mathcal{N}(0, I) \]
- Edge $\{Y_{ij}\}$ independent given $\{Z_i\}$, e.g.
  \[ \Pr\{Y_{ij} = 1\} = \gamma(Z_i, Z_j). \]

\[ Y = \begin{pmatrix}
0 & 1 & 1 & 0 & 1 & \ldots \\
0 & 0 & 1 & 0 & 1 & \ldots \\
0 & 0 & 0 & 0 & 0 & \ldots \\
0 & 0 & 0 & 0 & 1 & \ldots \\
0 & 0 & 0 & 0 & 0 & \ldots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix} \]
A variety of state-space models

Latent position models.
- [Hoff et al. (2002)]:
  \[ Z_i \in \mathbb{R}^d, \quad \logit \gamma(z, z') = a - |z - z'| \]
- [Handcock et al. (2007)]:
  \[ Z_i \sim \sum_k p_k \mathcal{N}_d(\mu_k, \sigma_k^2 I) \]
- [Daudin et al. (2010)]:
  \[ Z_i \in S_K, \quad \gamma(z, z') = \sum_{k, \ell} z_k z'_\ell \gamma_{k\ell} \]
A variety of state-space models

Latent position models.

- [Hoff et al. (2002)]:
  \[ Z_i \in \mathbb{R}^d, \quad \text{logit } \gamma(z, z') = a - |z - z'| \]

- [Handcock et al. (2007)]:
  \[ Z_i \sim \sum_k p_k N_d(\mu_k, \sigma_k^2 I) \]

- [Daudin et al. (2010)]:
  \[ Z_i \in S_K, \quad \gamma(z, z') = \sum_{k, \ell} z_k z'_\ell \gamma_{k\ell} \]

In this talk, focus on
- the Stochastic Block Model (SBM) and
- the \( W \)-graph model (and its associated graphon).
Stochastic Block Model (SBM)

A mixture model for random graphs.
[Nowicki and Snijders (2001)]
Stochastic Block Model (SBM)

A mixture model for random graphs.
[Nowicki and Snijders (2001)]

- Consider \( n \) nodes \( (i = 1\ldots n) \);
Stochastic Block Model (SBM)

A mixture model for random graphs.
[Nowicki and Snijders (2001)]

- Consider $n$ nodes ($i = 1..n$);
- $Z_i =$ unobserved label of node $i$:

\[
\{Z_i\} \text{ iid } \sim \mathcal{M}(1; \pi)
\]

$\pi = (\pi_1, ... \pi_K)$;
Stochastic Block Model (SBM)

A mixture model for random graphs.
[Nowicki and Snijders (2001)]

- Consider $n$ nodes ($i = 1..n$);
- $Z_i$ = unobserved label of node $i$:
  \[ \{Z_i\} \text{ iid } \sim \mathcal{M}(1; \pi) \]
  \[ \pi = (\pi_1, ..., \pi_K) ; \]
- Edge $Y_{ij}$ depends on the labels:
  \[ \{Y_{ij}\} \text{ independent given } \{Z_i\}, \]
  \[ \Pr\{Y_{ij} = 1\} = \gamma(Z_i, Z_j) \]
Stochastic Block Model (SBM)

A mixture model for random graphs. [Nowicki and Snijders (2001)]

- Consider $n$ nodes ($i = 1..n$);

- $Z_i$ = unobserved label of node $i$:

  \[
  \{Z_i\} \text{ iid } \sim \mathcal{M}(1; \pi)
  \]

  \[\pi = (\pi_1, \ldots, \pi_K)\];

- Edge $Y_{ij}$ depends on the labels:
  \[
  \{Y_{ij}\} \text{ independent given } \{Z_i\},
  \]

  \[\Pr(Y_{ij} = 1) = \gamma(Z_i, Z_j)\]
**W-graph model**

Latent variables:

\[(Z_i) \text{ iid } \sim U_{[0,1]},\]

Graphon function \(\gamma\):

\[\gamma(z, z') : [0, 1]^2 \rightarrow [0, 1]\]

Edges:

\[\Pr\{Y_{ij} = 1\} = \gamma(Z_i, Z_j)\]
Interpreting the graphon function

The graphon function provides a global picture of the network’s topology.

'Scale free'

Community

Small world
Few words about the $W$-graph

**Probabilistic point of view.**

- $W$-graph have been mostly studied in the probability literature: [Lovász and Szegedy (2006)], [Diaconis and Janson (2008)]
- Motif (sub-graph) frequencies are invariant characteristics of a $W$-graph.
- Intrinsic un-identifiability of the graphon function $\gamma$ is often overcome by imposing that $u \mapsto \int \gamma(u, v) \, dv$ is monotonous increasing.
Few words about the $W$-graph

Probabilistic point of view.

- $W$-graph have been mostly studied in the probability literature: [Lovász and Szegedy (2006)], [Diaconis and Janson (2008)]
- Motif (sub-graph) frequencies are invariant characteristics of a $W$-graph.
- Intrinsic un-identifiability of the graphon function $\gamma$ is often overcome by imposing that $u \mapsto \int \gamma(u, v) \, dv$ is monotonous increasing.

Statistical point of view.

- Not much attention has been paid to its inference until recently: [Airoldi et al. (2013)], [Chatterjee et al. (2014)], [Olhede and Wolfe (2014)], ...
- SBM can be used as a proxy for $W$-graph.
Some generalizations of latent space graph models

Latent space models can be extended in various directions.
Some generalizations of latent space graph models

Latent space models can be extended in various directions.

**Weighted or directed networks.** Edges may have values: count, real, \( \{0, +, -, \pm\} \), ...

Latent space model can be adapted as

\[
Y_{ij} | Z_i, Z_j \sim F(\gamma(Z_i, Z_j))
\]

where \( F \) is can be any distribution: Poisson, normal, multinomial, etc.
Some generalizations of latent space graph models

Latent space models can be extended in various directions.

**Weighted or directed networks.** Edges may have values: count, real, $\{0, +, -, \pm\}$, ...
Latent space model can be adapted as

$$Y_{ij} | Z_i, Z_j \sim \mathcal{F}(\gamma(Z_i, Z_j))$$

where $\mathcal{F}$ is can be any distribution: Poisson, normal, multinomial, etc.

**Accounting for covariates.** Latent space model can also accommodate for covariates, via a regression term:

$$Y_{ij} | Z_i, Z_j \sim \mathcal{F}(\gamma(Z_i, Z_j) + x_{ij}' \beta)$$

where $x_{ij} = (x_{ij}^1, \ldots x_{ij}^d)'$. 

S. Robin  Joint work with P. Latouche and S. Ouadah (INRA / AgroParisTech)

Network analysis using W-graphs
Statistical inference of latent space models
**Incomplete data models**

**Aim.** Based on the observed network $Y = (Y_{ij})$, one wants typically to infer

- the parameters
  $$\theta = (\pi, \gamma)$$

- the hidden states
  $$Z = (Z_i)$$
Incomplete data models

**Aim.** Based on the observed network $Y = (Y_{ij})$, one want typically to infer

- the parameters
  $$\theta = (\pi, \gamma)$$
- the hidden states
  $$Z = (Z_i)$$

State space models belong to the class of incomplete data models as

- the edges $(Y_{ij})$ are observed,
- the latent positions (or status) $(Z_i)$ are not,
- and neither are the parameter.
Frequentist or Bayesian inference

Frequentist inference. $\theta$ is fixed and $Z$ is random. The aim is then to

- provide an estimate $\hat{\theta}$ of $\theta$,
- provide the conditional distribution $P_\theta(Z|Y)$ (for classification purposes and as a side product of the inference).
Frequentist or Bayesian inference

Frequentist inference. $\theta$ is fixed and $Z$ is random. The aim is then to
- provide an estimate $\hat{\theta}$ of $\theta$,
- provide the conditional distribution $P_\theta(Z|Y)$ (for classification purposes and as a side product of the inference).

Bayesian inference. Both $\theta$ and $Z$ are random. The aim is then to
- provide the joint conditional distribution $P(\theta, Z|Y)$. 
Frequentist or Bayesian inference

**Frequentist inference.** $\theta$ is fixed and $Z$ is random. The aim is then to
- provide an estimate $\hat{\theta}$ of $\theta$,
- provide the conditional distribution $P_\theta(Z|Y)$ (for classification purposes and as a side product of the inference).

**Bayesian inference.** Both $\theta$ and $Z$ are random. The aim is then to
- provide the joint conditional distribution $P(\theta, Z|Y)$.

Whatever the approach, we have to deal with conditional distributions:

$$P_\theta(Z|Y) \quad \text{or} \quad P(\theta, Z|Y).$$
Conditional distributions (1/2)

**Graphical models** describe the conditional independences between the random variables from a model [Lauritzen (1996)].
Conditional distributions (1/2)

Graphical models describe the conditional independences between the random variables from a model [Lauritzen (1996)].

Frequentist setting:
- iid $Z_i$'s,
Conditional distributions (1/2)

**Graphical models** describe the conditional independences between the random variables from a model [Lauritzen (1996)].

Frequentist setting:
- iid $Z_i$’s,
- $P(Y_{ij} | Z_i, Z_j)$,

![Diagram](image)
Conditional distributions (1/2)

Graphical models describe the conditional independences between the random variables from a model [Lauritzen (1996)].

Frequentist setting:
- iid $Z_i$'s,
- $P(Y_{ij}|Z_i, Z_j)$,
- $P(Z_i, Z_j|Y)$: graph moralization,
Conditional distributions (1/2)

Graphical models describe the conditional independences between the random variables from a model [Lauritzen (1996)].

Frequentist setting:
- iid $Z_i$’s,
- $P(Y_{ij}|Z_i, Z_j)$,
- $P(Z_i, Z_j|Y)$: graph moralization,
- this holds for each pair $(i,j)$,
Conditional distributions (1/2)

Graphical models describe the conditional independences between the random variables from a model [Lauritzen (1996)].

Frequentist setting:
- iid $Z_i$'s,
- $P(Y_{ij} | Z_i, Z_j)$,
- $P(Z_i, Z_j | Y)$: graph moralization,
- this holds for each pair $(i, j)$,

Conditional distribution. The dependency graph of $Z$ given $Y$ is a clique.
→ No factorization can be hoped (unlike for HMM).
→ $P_\theta(Z | Y)$ can not be computed (efficiently).
Bayesian perspective. Things get worst because $\theta = (\pi, \gamma)$ is also random.

Model:
Bayesian perspective. Things get worst because $\theta = (\pi, \gamma)$ is also random.

Model:
- $P(\theta)$. 
Conditional distributions (2/2)

Bayesian perspective. Things get worst because $\theta = (\pi, \gamma)$ is also random.

Model:
- $P(\theta)$
- $P(Z|\pi)$
Bayesian perspective. Things get worst because $\theta = (\pi, \gamma)$ is also random.

Model:
- $P(\theta)$
- $P(Z|\pi)$
- $P(Y|\gamma, Z)$
Bayesian perspective. Things get worst because $\theta = (\pi, \gamma)$ is also random.

Model:
- $P(\theta)$
- $P(Z|\pi)$
- $P(Y|\gamma, Z)$
- $P(\theta, Z|Y)$ is even more involved.
**Conditional distributions (2/2)**

Bayesian perspective. Things get worst because $\theta = (\pi, \gamma)$ is also random.

Model:

- $P(\theta)$
- $P(Z|\pi)$
- $P(Y|\gamma, Z)$
- $P(\theta, Z|Y)$ is even more involved.

Both frequentist and Bayesian inference require the calculation of conditional distributions that can not be computed.
→ Either sampling (MCMC) or approximation is required.
Variational (Bayes) inference

Variational approximations aim at replacing an intractable exact distribution $P$ with a tractable approximate distribution $\tilde{P}$. Typically:

$$P_\theta(Z|Y) \approx \prod_i \tilde{P}_{\theta,Y}(Z_i)$$
$$P(\theta, Z|Y) \approx \tilde{P}_Y(\theta) \times \tilde{P}_Y(Z)$$
$$P(\theta, Z|Y) \approx \tilde{P}_Y(\theta) \times \prod_i \tilde{P}_Y(Z_i)$$

Popular strategy: minimize the Kullback-Leibler divergence between $\tilde{P}$ and $P$:

$$\min KL[\tilde{P}(Z)||P_\theta(Z|Y)] \quad \text{or} \quad \min KL[\tilde{P}(\theta, Z)||P(\theta, Z|Y)]$$

→ Variational EM (VEM) algorithm [Wainwright and Jordan (2008)].
→ Variational Bayes EM (VBEM) algorithm [Beal and Ghahramani (2003)].
VBEM inference for SBM: *E. coli*’s operon network

[Picard et al. (2009)]:

Meta-graph representation.

Parameter estimates. $K = 5$
VBEM inference for SBM: *E. coli*’s operon network

Meta-graph representation.

[Picard et al. (2009)]
VBEM inference for SBM: *E. coli*’s operon network

Meta-graph representation.

Parameter estimates. $K = 5$

[Picard et al. (2009)]
Accuracy of VBEM estimates for SBM: Simulation study

Credibility intervals: $\pi_1$: $+$, $\gamma_{11}$: $\triangle$, $\gamma_{12}$: $\circ$, $\gamma_{22}$: $\bullet$

Network analysis using $W$-graphs
Accuracy of VBEM estimates for SBM: Simulation study

Credibility intervals: \( \pi_1: + \), \( \gamma_{11}: \triangle \), \( \gamma_{12}: \circ \), \( \gamma_{22}: \bullet \)

Width of the posterior credibility intervals. \( \pi_1, \gamma_{11}, \gamma_{12}, \gamma_{22} \)

[Gazal et al. (2012)]
Latent space graph models are useful to describe network heterogeneity. Their statistical inference raises some specific issues. Variational approximations help to circumvent these issues.
First half summary

- Latent space graph models are useful to describe network heterogeneity.
- Their statistical inference raises some specific issues.
- Variational approximations help to circumvent these issues.

And also

- Theoretical justifications of these approximations exist for SBM: [Celisse et al. (2012)], [Mariadassou and Matias (2014)]
- VEM and VBEM algorithms have been specifically developed for SBM: [Daudin et al. (2008)], [Latouche et al. (2012)]
- Model selection (choice of $K$ has also be addressed): [same refs as above].
SBM as a $W$-graph model

Latent variables:

$$(Z_i) \text{iid } \sim \mathcal{M}(1, \pi)$$

Blockwise constant graphon:

$$\gamma(z, z') = \gamma_{k\ell}$$

Edges:

$$\Pr\{Y_{ij} = 1\} = \gamma(Z_i, Z_j)$$

→ block widths $= \pi_k$, block heights $\gamma_{k\ell}$
Variational Bayes estimation of $\gamma(z, z')$

**VBEM inference** provides the approximate posteriors:

\[
(\pi | Y) \approx \text{Dir}(\pi^*) \\
(\gamma_{k\ell} | Y) \approx \text{Beta}(\gamma^0_{k\ell}, \gamma^1_{k\ell})
\]

**Estimate of $\gamma(u, v)$.** Due to the uncertainty of the $\pi_k$, the posterior mean of $\gamma_{K}^{SBM}$ is smooth

(Explicit integration using [Gouda and Szántai (2010)])
Bayesian model averaging (BMA). Consider a series of models $1, \ldots, K, \ldots$ in which a certain function of the parameter $f(\theta)$ can always be defined.
Bayesian model averaging (BMA). Consider a series of models $1, \ldots, K, \ldots$ in which a certain function of the parameter $f(\theta)$ can always be defined.

- Bayesian inference within each model $K$ provides the posterior

\[ P(\theta|K, Y) \quad \rightarrow \quad P(f(\theta)|K, Y). \]
Bayesian model averaging (BMA). Consider a series of models $1, \ldots, K, \ldots$ in which a certain function of the parameter $f(\theta)$ can always be defined.

- Bayesian inference within each model $K$ provides the posterior

$$P(\theta|K, Y) \rightarrow P(f(\theta)|K, Y).$$

- BMA [Hoeting et al. (1999)] relies on the marginal posterior of $f(\theta)$:

$$P(f(\theta)|Y) = \sum_K P(K|Y)P(f(\theta)|K, Y).$$
Variational Bayes model averaging

**Pushing it further:** Consider the model $K$ as an additional hidden variable:

$$P(Z, \theta, K|Y) \approx \tilde{P}(Z, \theta, K)$$

$$:= \tilde{P}(Z|K) \times \tilde{P}(\theta|K) \times \tilde{P}(K)$$

Note that no additional independence assumption is needed.

\[1\text{in terms of Küllback-Leibler divergence}\]
Variational Bayes model averaging

Pushing it further: Consider the model $K$ as an additional hidden variable:

$$P(Z, \theta, K|Y) \approx \tilde{P}(Z, \theta, K)$$

$$:= \tilde{P}(Z|K) \times \tilde{P}(\theta|K) \times \tilde{P}(K)$$

Note that no additional independence assumption is needed.

Variational Bayes model averaging (VBMA). The optimal\(^1\) approximation of $P(K|Y)$ satisfies [Volant et al. (2012)]:

$$\tilde{P}(K) \propto P(K)e^{\log P(Y|K) - KL(K)} = P(K|Y)e^{-KL(K)}$$

where $KL(K) = KL[\tilde{P}(Z, \theta|K); P(Z, \theta|Y, K)]$.

\(^1\)in terms of Küllback-Leibler divergence
Inferring the graphon function

Model averaging: There is no ‘true $K$’ in the $W$-graph model.

Apply VBMA recipe to $\gamma(z, z')$. For $K = 1..K_{\text{max}}$, fit an SBM model via VBEM and compute

$$\hat{\gamma}^{\text{SBM}}_{K}(z, z') = \tilde{E}[\gamma_{C(z),C(z')} | Y, K].$$
Inferring the graphon function

Model averaging: There is no 'true $K$' in the $W$-graph model.

Apply VBMA recipe to $\gamma(z, z')$. For $K = 1..K_{\text{max}}$, fit an SBM model via VBEM and compute

$$\hat{\gamma}_K^{\text{SBM}}(z, z') = \mathbb{E}[\gamma_{C(z), C(z')} | Y, K].$$

Then perform model averaging as

$$\hat{\gamma}(z, z') = \mathbb{E}[\gamma_{C(z), C(z')} | Y] = \sum_K \tilde{P}(K) \hat{\gamma}_K^{\text{SBM}}(z, z'),$$

[Latouche and R. (2013)].
PPI network

Like many PPI networks, *E. coli*’s network is highly concentrated around few nodes.
PPI network

Like many PPI networks, *E. coli*’s network is highly concentrated around few nodes.
Ecological network between fungal species

Link between 2 fungi if they are observed on one common host.
Ecological network between fungal species

Link between 2 fungi if they are observed on one common host.
Brain network

Links = connexions between areas of the macaque’s cortex
Brain network

Links = connexions between areas of the macaque’s cortex
Blog network (non-biological)

Links = connexions between French political blogs
Blog network (non-biological)

Links = connexions between French political blogs
Goodness-of-fit
Motifs frequency

- Network motifs have a biological (or sociological) interpretation in terms of building blocks of the global network
  → Triangles = 'friends of my friends are my friends'.

- Latent space graph models only describe binary interactions, conditional on the latent positions

→ Goodness of fit criterion based on motif frequencies?
Moments of motif counts

Moments under SBM: The first moments $\mathbb{E} N(m)$, $\mathbb{V} N(m)$ of the count are known for exchangeable graph models (incl. SBM) [Picard et al. (2008)]:

$$\mathbb{E}_{SBM} N(m) \propto \mu_{SBM}(m) =: f(\theta_{SBM})$$

where $\mu_{SBM}(m)$ is the motif occurrence probability under SBM.

Moments under $W$-graph: Motif probability under the $W$-graph can be estimated as

$$\hat{\mu}(m) = \sum_k \tilde{P}(K) \mathbb{E}(\mu_{SBM}(m)|X, K)$$

Estimates of $\mathbb{E}_W N(m)$ and $\mathbb{V}_W N(m)$ can be derived accordingly [Latouche and R. (2013)].
### Network frequencies in the blog network

<table>
<thead>
<tr>
<th>Motif</th>
<th>Count ($\times 10^3$)</th>
<th>Mean ($\times 10^3$)</th>
<th>Std. dev. ($\times 10^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Motif" /></td>
<td>29.7</td>
<td>39.7</td>
<td>8.3</td>
</tr>
<tr>
<td><img src="image2" alt="Motif" /></td>
<td>3.8</td>
<td>4.6</td>
<td>1.3</td>
</tr>
<tr>
<td><img src="image3" alt="Motif" /></td>
<td>608.7</td>
<td>968.3</td>
<td>336.8</td>
</tr>
<tr>
<td><img src="image4" alt="Motif" /></td>
<td>279.8</td>
<td>428.9</td>
<td>154.0</td>
</tr>
<tr>
<td><img src="image5" alt="Motif" /></td>
<td>47.4</td>
<td>74.5</td>
<td>35.1</td>
</tr>
<tr>
<td><img src="image6" alt="Motif" /></td>
<td>270.5</td>
<td>397.0</td>
<td>177.0</td>
</tr>
<tr>
<td><img src="image7" alt="Motif" /></td>
<td>62.1</td>
<td>87.8</td>
<td>47.4</td>
</tr>
<tr>
<td><img src="image8" alt="Motif" /></td>
<td>6.5</td>
<td>8.8</td>
<td>5.4</td>
</tr>
</tbody>
</table>

No specific structure seems to be exceptional wrt the model’s expectations.
Data: \( n = 51 \) tree species, 
\( Y_{ij} = \) number of shared parasites 
[Vacher et al. (2008)].
Covariates: Tree interaction (valued) network

Data:  $n = 51$ tree species,
$Y_{ij} =$ number of shared parasites
[Vacher et al. (2008)].

SBM: Given $Z_i = k, Z_j = \ell$,

$$Y_{ij} \sim \mathcal{P}(e^{\gamma_{k\ell}}),$$

$\gamma_{k\ell} =$ log-mean number of shared parasites.
Covariates: Tree interaction (valued) network

**Data:** $n = 51$ tree species, $Y_{ij} =$ number of shared parasites
[Vacher et al. (2008)].

**SBM:** Given $Z_i = k, Z_j = \ell$, $\gamma_{k\ell} = \log$-mean number of shared parasites.

**Results:** ICL selects $K = 7$ groups that are partly related with phylums.
Covariates: Tree interaction (valued) network

Data: \( n = 51 \) tree species, \( Y_{ij} = \) number of shared parasites [Vacher et al. (2008)].

SBM: Given \( Z_i = k, Z_j = \ell, \)

\[
Y_{ij} \sim \mathcal{P}(e^{\gamma_{k\ell}}),
\]

\( \gamma_{k\ell} \) = log-mean number of shared parasites.

Results: ICL selects \( K = 7 \) groups that are partly related with phylums.

<table>
<thead>
<tr>
<th></th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
<th>T7</th>
</tr>
</thead>
<tbody>
<tr>
<td>( e^{\gamma_{k\ell}} )</td>
<td>14.46</td>
<td>4.19</td>
<td>5.99</td>
<td>7.67</td>
<td>2.44</td>
<td>0.13</td>
<td>1.43</td>
</tr>
<tr>
<td>T2</td>
<td>14.13</td>
<td>0.68</td>
<td>2.79</td>
<td>4.84</td>
<td>0.53</td>
<td>1.54</td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td>3.19</td>
<td>4.10</td>
<td>0.66</td>
<td>0.02</td>
<td>0.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T4</td>
<td>7.42</td>
<td>2.57</td>
<td>0.04</td>
<td>1.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T5</td>
<td>3.64</td>
<td>0.23</td>
<td>0.83</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T6</td>
<td>0.04</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T7</td>
<td>0.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( \hat{\pi}_k \) | 7.8  | 7.8  | 13.7 | 13.7 | 15.7 | 19.6 | 21.6 |

S. Robin Joint work with P. Latouche and S. Ouada Network analysis using W-graphs IMS, Singapore
Covariates: Tree interaction (valued) network

Data: \( n = 51 \) tree species, \( Y_{ij} = \text{number of shared parasites} \) [Vacher et al. (2008)].

SBM: Given \( Z_i = k, Z_j = \ell, \)

\[ Y_{ij} \sim P(e^{\gamma_{k\ell}}), \]

\( \gamma_{k\ell} = \text{log-mean number of shared parasites}. \)

Results: ICL selects \( K = 7 \) groups that are partly related with phylums.
Accounting for the taxonomic distance

Model: \( x_{ij} = \text{distance}(i, j) \)

\[
Y_{ij} \sim \mathcal{P}(e^{\gamma k \ell + \beta x_{ij}}),
\]

[Mariadassou et al. (2010)].
Accounting for the taxonomic distance

Model: \( x_{ij} = \text{distance}(i, j) \)

\[
Y_{ij} \sim \mathcal{P}(e^{\gamma_{kl} + \beta x_{ij}}),
\]

[Mariadassou et al. (2010)].

Results: \( \hat{\beta} = -0.317 \).
\( \rightarrow \) for \( \bar{x} = 3.82 \),

\[
e^{\hat{\beta} \bar{x}} = .298
\]

\( \rightarrow \) The mean number of shared parasites decreases with taxonomic distance.
Accounting for the taxonomic distance

Model: \( x_{ij} = \text{distance}(i, j) \)

\[ Y_{ij} \sim P\left( e^{\gamma k \ell + \beta x_{ij}} \right), \]

[Mariadassou et al. (2010)].

Results: \( \hat{\beta} = -0.317. \)

\( \rightarrow \) for \( \bar{x} = 3.82, \)

\[ e^{\hat{\beta} \bar{x}} = 0.298 \]

\( \rightarrow \) The mean number of shared parasites decreases with taxonomic distance.
Accounting for the taxonomic distance

Model: \( x_{ij} = \text{distance}(i, j) \)

\[ Y_{ij} \sim \mathcal{P}(e^{\gamma k \ell + \beta x_{ij}}), \]

[Mariadassou et al. (2010)].

Results: \( \hat{\beta} = -0.317. \)

\( \rightarrow \) for \( \bar{x} = 3.82, \)

\[ e^{\hat{\beta} \bar{x}} = 0.298 \]

\( \rightarrow \) The mean number of shared parasites decreases with taxonomic distance.

<table>
<thead>
<tr>
<th>( e^{\hat{\lambda}_{k \ell}} )</th>
<th>T'1</th>
<th>T'2</th>
<th>T'3</th>
<th>T'4</th>
</tr>
</thead>
<tbody>
<tr>
<td>T'1</td>
<td>0.75</td>
<td>2.46</td>
<td>0.40</td>
<td>3.77</td>
</tr>
<tr>
<td>T'2</td>
<td>4.30</td>
<td>0.52</td>
<td>1.05</td>
<td>8.77</td>
</tr>
<tr>
<td>T'3</td>
<td>0.080</td>
<td>1.05</td>
<td>1.05</td>
<td>14.22</td>
</tr>
</tbody>
</table>

| \( \pi_k \) | 17.7 | 21.5 | 23.5 | 37.3 |
| \( \beta \) | -0.317 |
Accounting for the taxonomic distance

Model: \( x_{ij} = \text{distance}(i,j) \)

\[ Y_{ij} \sim \mathcal{P}(e^{\gamma_{k\ell} + \beta x_{ij}}), \]

[Mariadassou et al. (2010)].

Results: \( \hat{\beta} = -0.317 \).

\( \rightarrow \) for \( \bar{x} = 3.82 \),

\[ e^{\hat{\beta} \bar{x}} = 0.298 \]

\( \rightarrow \) The mean number of shared parasites decreases with taxonomic distance.

\( \rightarrow \) Groups are no longer associated with the phylogenetic structure.

\( \rightarrow \) SBM = residual heterogeneity of the regression.
'Residual' graphon

A simple graph model with covariates. When edge covariates $x_{ij}$ are available, simply fit a logistic regression [Pattison and Robins (2007)]:

\[
\begin{align*}
(Y_{ij}) \text{ independent} & \\
Y_{ij} \sim \mathcal{B}(p_{ij}) & \\
\text{logit } p_{ij} = x'_{ij} \beta.
\end{align*}
\]
'Residual’ graphon

A simple graph model with covariates. When edge covariates \( x_{ij} \) are available, simply fit a logistic regression [Pattison and Robins (2007)]:

\[
\begin{align*}
(Y_{ij}) & \text{ independent} \\
Y_{ij} & \sim \mathcal{B}(p_{ij}) \\
\logit p_{ij} & = x_{ij}' \beta.
\end{align*}
\]

Introducing a residual term. To assess the fit of the model, simply add a residual graphon-like term:

\[
\begin{align*}
(Z_i) & \text{ iid } \mathcal{U}[0, 1] \\
Y_{ij} | Z_i, Z_j & \sim \mathcal{B}(p_{ij}) \\
\logit p_{ij} & = x_{ij}' \beta + \gamma(Z_i, Z_j).
\end{align*}
\]

→ A VBEM algorithm can be designed to get \( \tilde{P}(\beta, \theta, Z) \approx P(\beta, \theta, Z | Y) \):

On-going work + [Jaakkola and Jordan (2000)].
Tree network

**Binary version:** Links between tree species if they host at least one common fungal parasite.
**Tree network**

**Binary version:** Links between tree species if they host at least one common fungal parasite.

**Regression:** covariates = genetic distance, taxonomic distance, geographic distance

→ The residual graphon is not flat: some heterogeneity remains.
Blog network: Already shown.
Blog network: Already shown.

Regression: covariates = same political party, pair includes a journalist

→ The residual graphon is still not flat.
Conclusion & future work

Some conclusions.

- The graphon provides a representation of the network topology.
- It can be estimated using variational Bayes inference.
  \[ \rightarrow \text{R packages 'mixer' and 'blockmodels'} \]
- It can be combined with covariates as a residual term.
Conclusion & future work

Some conclusions.
- The graphon provides a representation of the network topology.
- It can be estimated using variational Bayes inference → R packages ‘mixer’ and ‘blockmodels’
- It can be combined with covariates as a residual term.

Future work.
- Formal goodness-of-fit test
- Quality of variational Bayes estimates in SBM with covariates

Thank you for your attention.


