Uncovering Latent Structure in Valued Graphs

M. Mariadassou, S. Robin

UMR AgroParisTech/INRA MIA 518, Paris

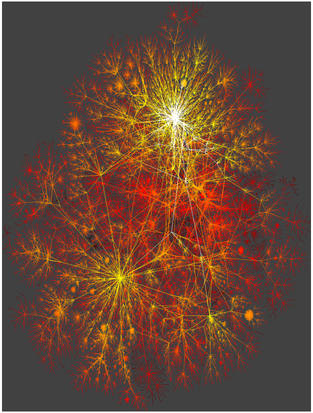
ECCS07, October 2007

Outline

Motivations for the study of networks

Networks...

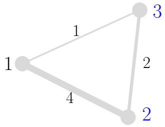
- Arise in many fields:
  - → Biology, Chemistry
  - → Physics, Internet.
- Represent an interaction pattern:
  - → $O(n^2)$ interactions
  - → between n elements.
- Have a topology which:
  - → reflects the structure/function relationship



From Barabási website

Some Notations

- **Notations:**
  - → V a set of vertices in $\{1, \dots, n\}$;
  - → E a set of edges in $\{1, \dots, n\}^2$;
  - → $\mathbf{X} = (X_{ij})$ the adjacency matrix, with X_{ij} the value of the edge between i and j.
- **Random graph definition:**
  - → To describe the network, we need the joint distribution of the X_{ij}.
- **Example:**

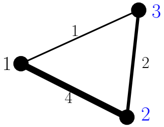


$$V = \{1, 2, 3\}$$
$$E = \{\{1, 2\}, \{2, 3\}, \{3, 1\}\}$$

$$\begin{pmatrix} . & 4 & 1 \\ . & . & 2 \\ . & . & . \end{pmatrix}$$

Some Notations

- **Notations:**
  - → V a set of vertices in $\{1, \dots, n\}$;
  - → E a set of edges in $\{1, \dots, n\}^2$;
  - → $\mathbf{X} = (X_{ij})$ the adjacency matrix, with X_{ij} the value of the edge between i and j.
- **Random graph definition:**
  - → To describe the network, we need the joint distribution of the X_{ij}.
- **Example:**



$$V = \{1, 2, 3\}$$
$$E = \{\{1, 2\}, \{2, 3\}, \{3, 1\}\}$$

$$\begin{pmatrix} . & 4 & 1 \\ . & . & 2 \\ . & . & . \end{pmatrix}$$

Explicit Random Graph Model (vertices)

- **Vertices heterogeneity**
  - → Hypothesis: the vertices are distributed among Q classes with different connectivity;
  - → $\mathbf{Z} = (\mathbf{Z}_i)_i$; $Z_{iq} = \mathbb{1}\{i \in q\}$ are indep. hidden variables;
  - → $\alpha = \{\alpha_q\}$, the *prior* proportions of groups;
  - → $(\mathbf{Z}_i) \sim \mathcal{M}(1, \alpha)$.
- **Example:**
  - → Example for 8 nodes and 3 classes with $\alpha = (0.25, 0.25, 0.5)$

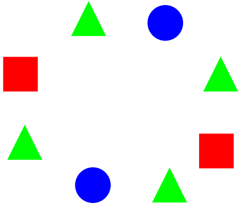
Explicit Random Graph Model (vertices)

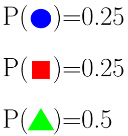
- **Vertices heterogeneity**
  - → Hypothesis: the vertices are distributed among Q classes with different connectivity;
  - → $\mathbf{Z} = (\mathbf{Z}_i)_i$; $Z_{iq} = \mathbb{1}\{i \in q\}$ are indep. hidden variables;
  - → $\boldsymbol{\alpha} = \{\alpha_q\}$, the *prior* proportions of groups;
  - → $(\mathbf{Z}_i) \sim \mathcal{M}(1, \alpha)$.
- **Example:**
  - → Example for 8 nodes and 3 classes with $\alpha = (0.25, 0.25, 0.5)$

Explicit Random Graph Model (vertices)

- **Vertices heterogeneity**
  - → Hypothesis: the vertices are distributed among Q classes with different connectivity;
  - → $\mathbf{Z} = (\mathbf{Z}_i)_i$; $Z_{iq} = \mathbb{1}\{i \in q\}$ are indep. hidden variables;
  - → $\boldsymbol{\alpha} = \{\alpha_q\}$, the *prior* proportions of groups;
  - → $(\mathbf{Z}_i) \sim \mathcal{M}(1, \boldsymbol{\alpha})$.

- **Example:**
  - → Example for 8 nodes and 3 classes with $\boldsymbol{\alpha} = (0.25, 0.25, 0.5)$



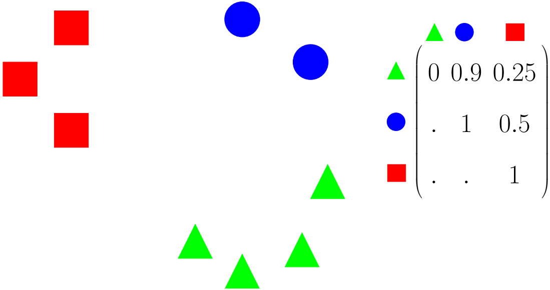


Explicit Random Graph Model (edges)

- **$\mathbf{X}$ distribution**
  - → conditional distribution : $X_{ij}|\{i \in q, j \in \ell\} \sim f(., \theta_{q\ell})$;
  - → $\boldsymbol{\theta} = (\theta_{q\ell})$ is the connectivity paramater matrix;
  - → ERMG : "Erdös-Rényi Mixture for Graphs".
- **Example:**
  - → Example for 3 classes with Bernoulli-valued edges;

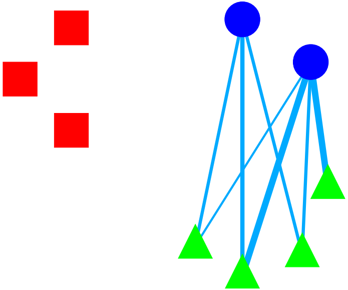
Explicit Random Graph Model (edges)

- **$\mathbf{X}$ distribution**
  - → conditional distribution : $X_{ij}|\{i \in q, j \in \ell\} \sim f(., \theta_{q\ell})$;
  - → $\boldsymbol{\theta} = (\theta_{q\ell})$ is the connectivity paramater matrix;
  - → ERMG : "Erdös-Rényi Mixture for Graphs".
- **Example:**
  - → Example for 3 classes with Bernoulli-valued edges;



Explicit Random Graph Model (edges)

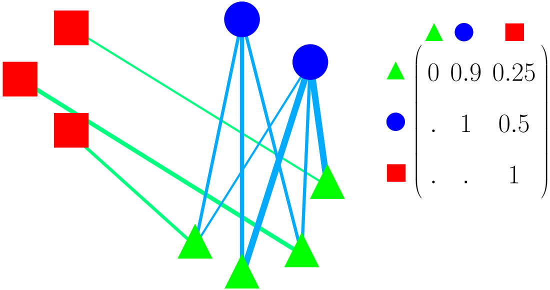
- **$\mathbf{X}$ distribution**
  - → conditional distribution : $X_{ij}|\{i \in q, j \in \ell\} \sim f(., \theta_{q\ell})$;
  - → $\boldsymbol{\theta} = (\theta_{q\ell})$ is the connectivity paramater matrix;
  - → ERMG : "Erdös-Rényi Mixture for Graphs".
- **Example:**
  - → Example for 3 classes with Bernoulli-valued edges;



	▲	●	■
▲	0	0.9	0.25
●	.	1	0.5
■	.	.	1

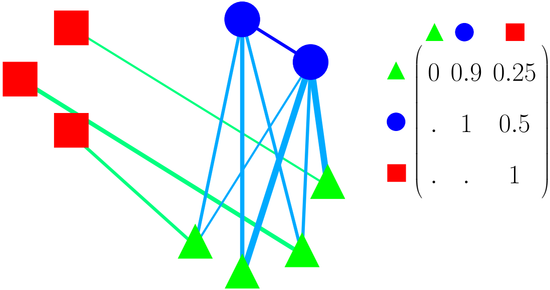
Explicit Random Graph Model (edges)

- **X distribution**
  - → conditional distribution : $X_{ij}|\{i \in q, j \in \ell\} \sim f(., \theta_{q\ell})$;
  - → $\boldsymbol{\theta} = (\theta_{q\ell})$ is the connectivity paramater matrix;
  - → ERMG : "Erdös-Rényi Mixture for Graphs".
- **Example:**
  - → Example for 3 classes with Bernoulli-valued edges;



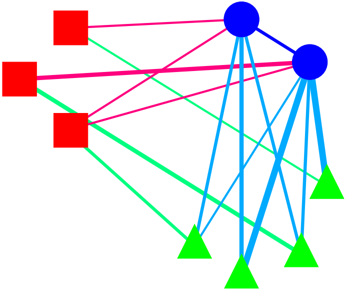
Explicit Random Graph Model (edges)

- **$\mathbf{X}$ distribution**
  - → conditional distribution : $X_{ij}|\{i \in q, j \in \ell\} \sim f(., \theta_{q\ell})$;
  - → $\boldsymbol{\theta} = (\theta_{q\ell})$ is the connectivity paramater matrix;
  - → ERMG : "Erdös-Rényi Mixture for Graphs".
- **Example:**
  - → Example for 3 classes with Bernoulli-valued edges;



Explicit Random Graph Model (edges)

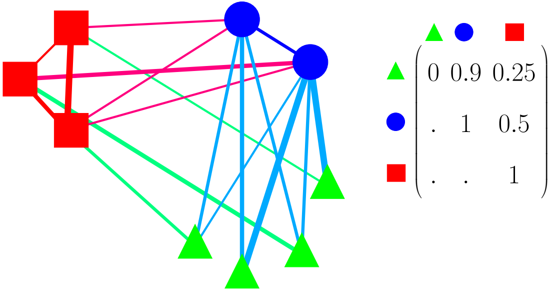
- **$\mathbf{X}$ distribution**
  - → conditional distribution : $X_{ij}|\{i \in q, j \in \ell\} \sim f(., \theta_{q\ell})$;
  - → $\boldsymbol{\theta} = (\theta_{q\ell})$ is the connectivity paramater matrix;
  - → ERMG : "Erdös-Rényi Mixture for Graphs".
- **Example:**
  - → Example for 3 classes with Bernoulli-valued edges;



$$
\begin{array}{c} \\ \blacktriangle \\ \bullet \\ \blacksquare \end{array}
\begin{array}{c}
\begin{array}{ccc} \blacktriangle & \bullet & \blacksquare \end{array} \\
\begin{pmatrix} 0 & 0.9 & 0.25 \\ . & 1 & 0.5 \\ . & . & 1 \end{pmatrix}
\end{array}
$$

Explicit Random Graph Model (edges)

- **$\mathbf{X}$ distribution**
  - → conditional distribution : $X_{ij}|\{i \in q, j \in \ell\} \sim f(., \theta_{q\ell})$;
  - → $\boldsymbol{\theta} = (\theta_{q\ell})$ is the connectivity paramater matrix;
  - → ERMG : "Erdös-Rényi Mixture for Graphs".
- **Example:**
  - → Example for 3 classes with Bernoulli-valued edges;



Random Edge Values

- **Classical Distributions:**

→ $f(., \theta_{q\ell})$ can be any probability distribution;

→ Bernoulli: presence/absence of an edge;

→ Multinomial: nature of the connection (friend, lover, colleague);

→ Poisson: in coauthorship networks, number of copublished papers;

→ Gaussian: intensity of the connection (airport network);

→ Bivariate Gaussian: directed networks where forward and backward edges are correlated;

→ Etc.

Mixture Model to easily generate graphs

Random Edge Values

- **Classical Distributions:**

→ $f(., \theta_{q\ell})$ can be any probability distribution;

→ Bernoulli: presence/absence of an edge;

→ Multinomial: nature of the connection (friend, lover, colleague);

→ Poisson: in coauthorship networks, number of copublished papers;

→ Gaussian: intensity of the connection (airport network);

→ Bivariate Gaussian: directed networks where forward and backward edges are correlated;

→ Etc.

Mixture Model to easily generate graphs

Random Edge Values

- **Classical Distributions:**

→ $f(., \theta_{q\ell})$ can be any probability distribution;

→ Bernoulli: presence/absence of an edge;

→ Multinomial: nature of the connection (friend, lover, colleague);

→ Poisson: in coauthorship networks, number of copublished papers;

→ Gaussian: intensity of the connection (airport network);

→ Bivariate Gaussian: directed networks where forward and backward edges are correlated;

→ Etc.

Mixture Model to easily generate graphs

Random Edge Values

- **Classical Distributions:**

→ $f(., \theta_{q\ell})$ can be any probability distribution;

→ Bernoulli: presence/absence of an edge;

→ Multinomial: nature of the connection (friend, lover, colleague);

→ Poisson: in coauthorship networks, number of copublished papers;

→ Gaussian: intensity of the connection (airport network);

→ Bivariate Gaussian: directed networks where forward and backward edges are correlated;

→ Etc.

Mixture Model to easily generate graphs

Random Edge Values

- **Classical Distributions:**

→ $f(., \theta_{q\ell})$ can be any probability distribution;

→ Bernoulli: presence/absence of an edge;

→ Multinomial: nature of the connection (friend, lover, colleague);

→ Poisson: in coauthorship networks, number of copublished papers;

→ Gaussian: intensity of the connection (airport network);

→ Bivariate Gaussian: directed networks where forward and backward edges are correlated;

→ Etc.

Mixture Model to easily generate graphs

Random Edge Values

- **Classical Distributions:**

→ $f(., \theta_{q\ell})$ can be any probability distribution;

→ Bernoulli: presence/absence of an edge;

→ Multinomial: nature of the connection (friend, lover, colleague);

→ Poisson: in coauthorship networks, number of copublished papers;

→ Gaussian: intensity of the connection (airport network);

→ Bivariate Gaussian: directed networks where forward and backward edges are correlated;

→ Etc.

Mixture Model to easily generate graphs

Random Edge Values

- **Classical Distributions:**

→ $f(., \theta_{q\ell})$ can be any probability distribution;

→ Bernoulli: presence/absence of an edge;

→ Multinomial: nature of the connection (friend, lover, colleague);

→ Poisson: in coauthorship networks, number of copublished papers;

→ Gaussian: intensity of the connection (airport network);

→ Bivariate Gaussian: directed networks where forward and backward edges are correlated;

→ Etc.

Mixture Model to easily generate graphs

Random Edge Values

- **Classical Distributions:**

→ $f(., \theta_{q\ell})$ can be any probability distribution;

→ Bernoulli: presence/absence of an edge;

→ Multinomial: nature of the connection (friend, lover, colleague);

→ Poisson: in coauthorship networks, number of copublished papers;

→ Gaussian: intensity of the connection (airport network);

→ Bivariate Gaussian: directed networks where forward and backward edges are correlated;

→ Etc.

Mixture Model to easily generate graphs

Log-Likelihood of the model

First Idea: Use maximum likelihood estimators

- **Complete data likelihood**

$$\mathcal{L}(\mathbf{X},\mathbf{Z}) = \sum_i \sum_q Z_{iq} \ln \alpha_q + \sum_{i<j} \sum_{q,\ell} Z_{iq} Z_{j\ell} \ln f_{\theta_{q\ell}}(X_{ij})$$

  with $f_{\theta_{q\ell}}(X_{ij})$ likelihood of edge value X_{ij} under $i \sim q$ and $j \sim \ell$.

- **Observed data likelihood**

$$\mathcal{L}(\mathbf{X}) = \ln \sum_{\mathbf{Z}} \exp \mathcal{L}(\mathbf{X},\mathbf{Z})$$

- The observed data likelihood requires a sum over Q^n terms, and is thus untractable;
- EM-like strategies require the knowledge of $\Pr(\mathbf{Z}|\mathbf{X})$, also untractable (no conditional independence) and thus also fail.

Log-Likelihood of the model

First Idea: Use maximum likelihood estimators

- **Complete data likelihood**

$$\mathcal{L}(\mathbf{X}, \mathbf{Z}) = \sum_{i} \sum_{q} Z_{iq} \ln \alpha_q + \sum_{i<j} \sum_{q,\ell} Z_{iq} Z_{j\ell} \ln f_{\theta_{q\ell}}(X_{ij})$$

  with $f_{\theta_{q\ell}}(X_{ij})$ likelihood of edge value X_{ij} under $i \sim q$ and $j \sim \ell$.

- **Observed data likelihood**

$$\mathcal{L}(\mathbf{X}) = \ln \sum_{\mathbf{Z}} \exp \mathcal{L}(\mathbf{X}, \mathbf{Z})$$

- The observed data likelihood requires a sum over Q^n terms, and is thus untractable;

- EM-like strategies require the knowledge of $\Pr(\mathbf{Z}|\mathbf{X})$, also untractable (no conditional independence) and thus also fail.

Log-Likelihood of the model

First Idea: Use maximum likelihood estimators

- **Complete data likelihood**

$$\mathcal{L}(\mathbf{X}, \mathbf{Z}) = \sum_i \sum_q Z_{iq} \ln \alpha_q + \sum_{i<j} \sum_{q,\ell} Z_{iq} Z_{j\ell} \ln f_{\theta_{q\ell}}(X_{ij})$$

  with $f_{\theta_{q\ell}}(X_{ij})$ likelihood of edge value X_{ij} under $i \sim q$ and $j \sim \ell$.

- **Observed data likelihood**

$$\mathcal{L}(\mathbf{X}) = \ln \sum_{\mathbf{Z}} \exp \mathcal{L}(\mathbf{X}, \mathbf{Z})$$

- The observed data likelihood requires a sum over Q^n terms, and is thus untractable;

- EM-like strategies require the knowledge of $\Pr(\mathbf{Z}|\mathbf{X})$, also untractable (no conditional independence) and thus also fail.

Variational Inference: Pseudo Likelihood

Main Idea: Replace complicated $\Pr(\mathbf{Z}|\mathbf{X})$ by a simple $\mathcal{R}_{\mathbf{X}}[\mathbf{Z}]$ such that $KL(\mathcal{R}_{\mathbf{X}}[\mathbf{Z}], \Pr(\mathbf{Z}|\mathbf{X}))$ is minimal.

- Optimize in $\mathcal{R}_{\mathbf{X}}$ the function $\mathcal{J}(\mathcal{R}_{\mathbf{X}})$ given by :

$$\begin{aligned}\mathcal{J}(\mathcal{R}_{\mathbf{X}}[\mathbf{Z}]) &= \mathcal{L}(\mathbf{X}) - KL(\mathcal{R}_{\mathbf{X}}[\mathbf{Z}], \Pr(\mathbf{Z}|\mathbf{X})) \\ &= \mathcal{H}(\mathcal{R}_{\mathbf{X}}[\mathbf{Z}]) - \sum_{\mathbf{Z}} \mathcal{R}_{\mathbf{X}}[\mathbf{Z}]\mathcal{L}(\mathbf{X}, \mathbf{Z})\end{aligned}$$

- At best, $\mathcal{R}_{\mathbf{X}} = \Pr(\mathbf{Z}|\mathbf{X})$ and $\mathcal{J}(\mathcal{R}_{\mathbf{X}}[\mathbf{Z}]) = \mathcal{L}(\mathbf{X})$;
- For simple $\mathcal{R}_{\mathbf{X}}$, $\mathcal{J}(\mathcal{R}_{\mathbf{X}}[\mathbf{Z}])$ is tractable.

Variational Inference: Pseudo Likelihood

Main Idea: Replace complicated $\Pr(\mathbf{Z}|\mathbf{X})$ by a simple $\mathcal{R}_{\mathbf{X}}[\mathbf{Z}]$ such that $KL(\mathcal{R}_{\mathbf{X}}[\mathbf{Z}], \Pr(\mathbf{Z}|\mathbf{X}))$ is minimal.

- Optimize in $\mathcal{R}_{\mathbf{X}}$ the function $\mathcal{J}(\mathcal{R}_{\mathbf{X}})$ given by :

$$\begin{aligned} \mathcal{J}(\mathcal{R}_{\mathbf{X}}[\mathbf{Z}]) &= \mathcal{L}(\mathbf{X}) - KL(\mathcal{R}_{\mathbf{X}}[\mathbf{Z}], \Pr(\mathbf{Z}|\mathbf{X})) \\ &= \mathcal{H}(\mathcal{R}_{\mathbf{X}}[\mathbf{Z}]) - \sum_{\mathbf{Z}} \mathcal{R}_{\mathbf{X}}[\mathbf{Z}]\mathcal{L}(\mathbf{X}, \mathbf{Z}) \end{aligned}$$

- At best, $\mathcal{R}_{\mathbf{X}} = \Pr(\mathbf{Z}|\mathbf{X})$ and $\mathcal{J}(\mathcal{R}_{\mathbf{X}}[\mathbf{Z}]) = \mathcal{L}(\mathbf{X})$;

- For simple $\mathcal{R}_{\mathbf{X}}$, $\mathcal{J}(\mathcal{R}_{\mathbf{X}}[\mathbf{Z}])$ is tractable.

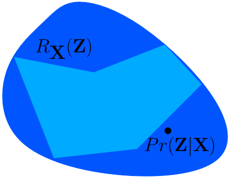
Variational Inference: Pseudo Likelihood

Main Idea: Replace complicated $\Pr(\mathbf{Z}|\mathbf{X})$ by a simple $\mathcal{R}_{\mathbf{X}}[\mathbf{Z}]$ such that $KL(\mathcal{R}_{\mathbf{X}}[\mathbf{Z}], \Pr(\mathbf{Z}|\mathbf{X}))$ is minimal.

- Optimize in $\mathcal{R}_{\mathbf{X}}$ the function $\mathcal{J}(\mathcal{R}_{\mathbf{X}})$ given by :

$$\begin{aligned} \mathcal{J}(\mathcal{R}_{\mathbf{X}}[\mathbf{Z}]) &= \mathcal{L}(\mathbf{X}) - KL(\mathcal{R}_{\mathbf{X}}[\mathbf{Z}], \Pr(\mathbf{Z}|\mathbf{X})) \\ &= \mathcal{H}(\mathcal{R}_{\mathbf{X}}[\mathbf{Z}]) - \sum_{\mathbf{Z}} \mathcal{R}_{\mathbf{X}}[\mathbf{Z}]\mathcal{L}(\mathbf{X}, \mathbf{Z}) \end{aligned}$$

- At best, $\mathcal{R}_{\mathbf{X}} = \Pr(\mathbf{Z}|\mathbf{X})$ and $\mathcal{J}(\mathcal{R}_{\mathbf{X}}[\mathbf{Z}]) = \mathcal{L}(\mathbf{X})$;
- For simple $\mathcal{R}_{\mathbf{X}}$, $\mathcal{J}(\mathcal{R}_{\mathbf{X}}[\mathbf{Z}])$ is tractable.



2 Step Algorithm

- **Step 1 Optimize $\mathcal{J}(\mathcal{R}_{\mathbf{X}}[\mathbf{Z}])$ w.r.t. $\mathcal{R}_{\mathbf{X}}[\mathbf{Z}]$:**
  - → Restriction to a "comfortable" class of functions;
  - → $\mathcal{R}_{\mathbf{X}}[\mathbf{Z}] = \prod_i h(\mathbf{Z}_i; \boldsymbol{\tau}_{i,\mathbf{X}})$, with $h(.; \boldsymbol{\tau}_{i,\mathbf{X}})$ the multinomial distribution;
  - → $\tau_{iq,\mathbf{X}}$ is a variational parameter to be optimized using a fixed point algorithm:

$$\tilde{\tau}_{iq,\mathbf{X}} \propto \alpha_q \prod_{j \neq i} \prod_{\ell=1}^{Q} f_{\theta_{q\ell}}(X_{ij})^{\tilde{\tau}_{j\ell,\mathbf{X}}}$$

- **Step 2 Optimize $\mathcal{J}(\mathcal{R}_{\mathbf{X}}[\mathbf{Z}])$ w.r.t. (α, θ):**
  - → Constraint: $\sum_q \alpha_q = 1$

$$\begin{aligned} \tilde{\alpha}_q &= \sum_i \tilde{\tau}_{iq,\mathbf{X}} / n \\ \tilde{\theta}_{q\ell} &= \arg\max_\theta \sum_{ij} \tilde{\tau}_{iq,\mathbf{X}} \tilde{\tau}_{j\ell,\mathbf{X}} \log f_\theta(X_{ij}) \end{aligned}$$

  - → Closed expression of $\tilde{\theta}_{q\ell}$ for classical distributions.

2 Step Algorithm

- Step 1 **Optimize** $\mathcal{J}(\mathcal{R}_{\mathbf{X}}[\mathbf{Z}])$ **w.r.t.** $\mathcal{R}_{\mathbf{X}}[\mathbf{Z}]$:
  - → Restriction to a "comfortable" class of functions;
  - → $\mathcal{R}_{\mathbf{X}}[\mathbf{Z}] = \prod_i h(\mathbf{Z}_i; \boldsymbol{\tau}_{i,\mathbf{X}})$, with $h(.; \boldsymbol{\tau}_{i,\mathbf{X}})$ the multinomial distribution;
  - → $\tau_{iq,\mathbf{X}}$ is a variational parameter to be optimized using a fixed point algorithm:

$$\tilde{\tau}_{iq,\mathbf{X}} \propto \alpha_q \prod_{j \neq i} \prod_{\ell=1}^{Q} f_{\theta_{q\ell}}(X_{ij})^{\tilde{\tau}_{j\ell,\mathbf{X}}}$$

- Step 2 **Optimize** $\mathcal{J}(\mathcal{R}_{\mathbf{X}}[\mathbf{Z}])$ **w.r.t.** (α, θ):
  - → Constraint: $\sum_q \alpha_q = 1$

$$\begin{aligned} \tilde{\alpha}_q &= \sum_i \tilde{\tau}_{iq,\mathbf{X}}/n \\ \tilde{\theta}_{q\ell} &= \arg\max_{\theta} \sum_{ij} \tilde{\tau}_{iq,\mathbf{X}} \tilde{\tau}_{j\ell,\mathbf{X}} \log f_{\theta}(X_{ij}) \end{aligned}$$

  - → Closed expression of $\tilde{\theta}_{q\ell}$ for classical distributions.

Model Selection Criterion

- We derive a statistical BIC-like criterion to select the number of classes:
- The likelihood can be split: $\mathcal{L}(\mathbf{X}, \mathbf{Z}|Q) = \mathcal{L}(\mathbf{X}|\mathbf{Z}, Q) + \mathcal{L}(\mathbf{Z}|Q)$.
- These terms can be penalized separately:

$$\begin{aligned} \mathcal{L}(\mathbf{X}|\mathbf{Z}, Q) &\rightarrow \text{pen}_{\mathbf{X}|\mathbf{Z}} = \frac{Q(Q+1)}{2} \log \frac{n(n-1)}{2} \\ \mathcal{L}(\mathbf{Z}|Q) &\rightarrow \text{pen}_{\mathbf{Z}} = (Q-1)\log(n) \end{aligned}$$

$$ICL(Q) = \max_{\theta} \mathcal{L}(\mathbf{X}, \tilde{\mathbf{Z}}|\theta, m_Q) - \frac{1}{2}\left(\frac{Q(Q+1)}{2} \log \frac{n(n-1)}{2} - (Q-1)\log(n)\right)$$

Model Selection Criterion

- We derive a statistical BIC-like criterion to select the number of classes:
- The likelihood can be split: $\mathcal{L}(\mathbf{X}, \mathbf{Z}|Q) = \mathcal{L}(\mathbf{X}|\mathbf{Z}, Q) + \mathcal{L}(\mathbf{Z}|Q)$.
- These terms can be penalized separately:

$$\begin{aligned} \mathcal{L}(\mathbf{X}|\mathbf{Z}, Q) &\rightarrow \text{pen}_{\mathbf{X}|\mathbf{Z}} = \frac{Q(Q+1)}{2} \log \frac{n(n-1)}{2} \\ \mathcal{L}(\mathbf{Z}|Q) &\rightarrow \text{pen}_{\mathbf{Z}} = (Q-1)\log(n) \end{aligned}$$

$$\boxed{ICL(Q) = \max_{\boldsymbol{\theta}} \mathcal{L}(\mathbf{X}, \tilde{\mathbf{Z}}|\boldsymbol{\theta}, m_Q) - \frac{1}{2}\left(\frac{Q(Q+1)}{2}\log\frac{n(n-1)}{2} - (Q-1)\log(n)\right)}$$

Simulation Setup

→ Undirected graph with $Q = 3$ classes;

→ Poisson-valued edges;

→ $n = 100, \ 500$ vertices;

→ $\alpha_q \propto a^q$ for $a = 1, \ 0.5, \ 0.2$;

- $a = 1$: balanced classes;
- $a = 0.2$: unbalanced classes $(80.6\%, \ 16.1\%, \ 3.3\%)$

→ Connectivity matrix of the form $\begin{pmatrix} \lambda & \gamma\lambda & \gamma\lambda \\ \gamma\lambda & \lambda & \gamma\lambda \\ \gamma\lambda & \gamma\lambda & \lambda \end{pmatrix}$ for $\gamma = 0.1, \ 0.5, \ 0.9, \ 1.5$ and $\lambda = 2, \ 5$.

- $\gamma = 1$: all classes equivalent (same connectivity pattern);
- $\gamma <> 1$: classes are different;
- λ: mean value of an edge;

→ 100 repeats for each setup.

Simulation Setup

→ Undirected graph with $Q = 3$ classes;

→ Poisson-valued edges;

→ $n = 100, \ 500$ vertices;

→ $\alpha_q \propto a^q$ for $a = 1, \ 0.5, \ 0.2$;

- $a = 1$: balanced classes;
- $a = 0.2$: unbalanced classes $(80.6\%, \ 16.1\%, \ 3.3\%)$

→ Connectivity matrix of the form $\begin{pmatrix} \lambda & \gamma\lambda & \gamma\lambda \\ \gamma\lambda & \lambda & \gamma\lambda \\ \gamma\lambda & \gamma\lambda & \lambda \end{pmatrix}$ for $\gamma = 0.1, \ 0.5, \ 0.9, \ 1.5$ and $\lambda = 2, \ 5$.

- $\gamma = 1$: all classes equivalent (same connectivity pattern);
- $\gamma <> 1$: classes are different;
- λ: mean value of an edge;

→ 100 repeats for each setup.

Simulation Setup

→ Undirected graph with $Q = 3$ classes;

→ Poisson-valued edges;

→ $n = 100, 500$ vertices;

→ $\alpha_q \propto a^q$ for $a = 1, 0.5, 0.2$;
- $a = 1$: balanced classes;
- $a = 0.2$: unbalanced classes $(80.6\%, 16.1\%, 3.3\%)$

→ Connectivity matrix of the form $\begin{pmatrix} \lambda & \gamma\lambda & \gamma\lambda \\ \gamma\lambda & \lambda & \gamma\lambda \\ \gamma\lambda & \gamma\lambda & \lambda \end{pmatrix}$ for $\gamma = 0.1, 0.5, 0.9, 1.5$ and $\lambda = 2, 5$.
- $\gamma = 1$: all classes equivalent (same connectivity pattern);
- $\gamma <> 1$: classes are different;
- λ: mean value of an edge;

→ 100 repeats for each setup.

Simulation Setup

→ Undirected graph with $Q = 3$ classes;

→ Poisson-valued edges;

→ $n = 100,\ 500$ vertices;

→ $\alpha_q \propto a^q$ for $a = 1,\ 0.5,\ 0.2$;

- $a = 1$: balanced classes;
- $a = 0.2$: unbalanced classes $(80.6\%,\ 16.1\%,\ 3.3\%)$

→ Connectivity matrix of the form $\begin{pmatrix} \lambda & \gamma\lambda & \gamma\lambda \\ \gamma\lambda & \lambda & \gamma\lambda \\ \gamma\lambda & \gamma\lambda & \lambda \end{pmatrix}$ for

$\gamma = 0.1,\ 0.5,\ 0.9,\ 1.5$ and $\lambda = 2,\ 5$.

- $\gamma = 1$: all classes equivalent (same connectivity pattern);
- $\gamma <> 1$: classes are different;
- λ: mean value of an edge;

→ 100 repeats for each setup.

Simulation Setup

- → Undirected graph with $Q = 3$ classes;
- → Poisson-valued edges;
- → $n = 100, \ 500$ vertices;
- → $\alpha_q \propto a^q$ for $a = 1, \ 0.5, \ 0.2$;
  - $a = 1$: balanced classes;
  - $a = 0.2$: unbalanced classes $(80.6\%, \ 16.1\%, \ 3.3\%)$
- → Connectivity matrix of the form $\begin{pmatrix} \lambda & \gamma\lambda & \gamma\lambda \\ \gamma\lambda & \lambda & \gamma\lambda \\ \gamma\lambda & \gamma\lambda & \lambda \end{pmatrix}$ for $\gamma = 0.1, \ 0.5, \ 0.9, \ 1.5$ and $\lambda = 2, \ 5$.
  - $\gamma = 1$: all classes equivalent (same connectivity pattern);
  - $\gamma <> 1$: classes are different;
  - λ: mean value of an edge;
- → 100 repeats for each setup.

Results

- Root Mean Square Error (RMSE) = $\sqrt{Bias^2 + Variance}$

Results

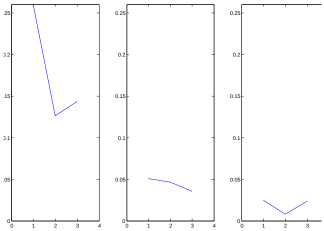
- Root Mean Square Error (RMSE) = $\sqrt{Bias^2 + Variance}$

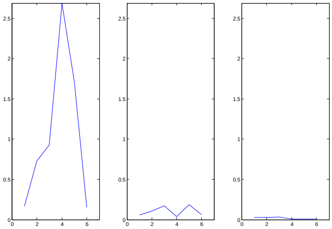
RMSE for the α_q

x-axis: $\alpha_1, \alpha_2, \alpha_3$

RMSE for the λ_{ql}

x-axis: $\lambda_{11}, \lambda_{22}, \lambda_{33}, \lambda_{12}, \lambda_{13}, \lambda_{23}$





(n,λ,γ,a) from left (hard) to right (easy):

$(100, 2, 0.9, 0.2)$, $(100, 2, 0.5, 0.5)$, $(500, 5, 0.1, 1)$

Simulation Setup and Results

- → Undirected graph with $Q^\star = 3$ classes;
- → Poisson-valued edges;
- → $n = 50,\ 100,\ 500,\ 1000$ vertices;
- → $\alpha_q = (57.1\%, 28,6\%, 14,3\%)$ (or $a = 0.5$);
- → $\lambda = 2, \gamma = 0.5$;
- → Retrieve Q that maximizes ICL;
- → 100 repeats for each value of n;

	Q		
n	2	3	4
50	82	**17**	1
100	7	**90**	3
500	0	**100**	0
1000	0	**100**	0

Frequency (in %) at which Q is selected for various n.

Simulation Setup and Results

- → Undirected graph with $Q^\star = 3$ classes;
- → Poisson-valued edges;
- → $n = 50,\ 100,\ 500,\ 1000$ vertices;
- → $\alpha_q = (57.1\%, 28, 6\%, 14, 3\%)$ (or $a = 0.5$);
- → $\lambda = 2, \gamma = 0.5$;
- → Retrieve Q that maximizes ICL;
- → 100 repeats for each value of n;

	Q		
n	2	3	4
50	82	**17**	1
100	7	**90**	3
500	0	**100**	0
1000	0	**100**	0

Frequency (in %) at which Q is selected for various n.

Summary

Flexibility of ERMG

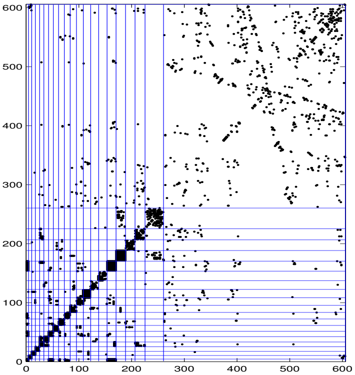
- A simple way to simulate networks;
- Many distributions to model different networks;
- Probabilistic model which captures features of real-networks (data not-shown).

Estimation and Model selection

- Variational approaches to compute approximate MLE when dependencies are complex,
- A statistical criterion to choose the number of classes (ICL).

E. Coli reaction network `http://www.biocyc.org/`

- Dot-plot representation (605 nodes and $1,782$ vertices)
  - $\rightarrow$ adjacency matrix (sorted)
- Biological interpretation:
  - $\rightarrow$ Groups 1 to 20 gather reactions involving all the same compound either as a substrate or as a product,
  - $\rightarrow$ A compound (chorismate, pyruvate, ATP,*etc*) can be associated to each group.
- The structure of the metabolic network is governed by the compounds.



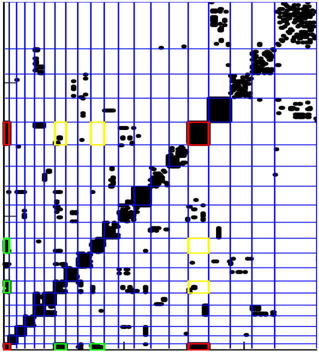
E. Coli reaction network `http://www.biocyc.org/`

→ Classes 1 and 16 constitute s single clique corresponding to a single compound (pyruvate),

→ They are split into two classes because they interact differently with classes 7 (CO_2) and 10 (AcetylCoA)

→ Connectivity matrix (sample):

q, l	1	7	10	16
1	1.0			
7	.11	.65		
10	.43		.67	
16	1.0	.01	ϵ	1.0



Adjacency matrix (sample)