Mixture models for random graphs

Jean-Jacques Daudin
UMR AgroParisTech-INRA518

GDR Statistique et Sante,
13 November 2008,
Why do people represent data by a network (1)?

- Real networks do exist: electric, transport or www networks...They have been represented for a long time by virtual networks.
- Virtual network is a nice way for representing or even "modelling" many scientific phenomenons: social relations, metabolic pathways, chemical reactions...

FIG. 5: The karate club network of Zachary (figure taken from Girvan and Newman [18]).
Why do people represent data by a network (2)?

- an overall representation of the interactions between many nodes
- the plot reveals the topology of the networks
- nodes may be colored, adding more information

Figure 2 | Yeast protein interaction network. A map of protein–protein interactions in *Saccharomyces cerevisiae*, which is based on early yeast two-hybrid measurements, illustrates that a few highly connected nodes (which are also known as hubs) hold the network together. The largest cluster, which contains ~78% of all proteins, is shown. The colour of a node indicates the phenotypic effect of removing the corresponding protein (red = lethal, green = non-lethal, orange = slow growth, yellow = unknown). Reproduced with permission from REF 18 © Macmillan Magazines Ltd.
An unusual data set structure

<table>
<thead>
<tr>
<th>item</th>
<th>X_1</th>
<th>...</th>
<th>...</th>
<th>X_p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>x_{11}</td>
<td>...</td>
<td>...</td>
<td>x_{1p}</td>
</tr>
<tr>
<td>2</td>
<td>x_{21}</td>
<td>...</td>
<td>...</td>
<td>x_{2p}</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>n</td>
<td>x_{n1}</td>
<td>...</td>
<td>...</td>
<td>x_{np}</td>
</tr>
</tbody>
</table>

Structure for relational data

<table>
<thead>
<tr>
<th>item1</th>
<th>item2</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>r_{12}</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>r_{13}</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>n−1</td>
<td>n</td>
<td>r_{n−1,n}</td>
</tr>
</tbody>
</table>

- In the relational data set, the core information is the relation between two items.
- Lines are not independent.
- The data structure is similar to distance, similarity, covariance or correlation matrices.
What do we want?

A simple representation of a complex graph, using meta-vertices and meta-edges.

FIG. 5: The karate club network of Zachary (figure taken from Girvan and Newman [18]).
Mixtnet: a mixture model for random graphs

- \( i = 1, n \) nodes
- \( q = 1, Q \) classes
- \( X_{ij} = 1 \) if there is an edge from node \( i \) to node \( j \).
- \( Z = Z_{iq} \) discrete latent variable, \( Z_{iq} = 1 \) if node \( i \) pertains to class \( q \)
- \((Z_{i1}, Z_{i2}...Z_{iQ}) \sim \mathcal{M}(1, \alpha_1, \alpha_2, ...\alpha_Q)\)
- Conditionally to \( Z \), \( X_{ij} \) are independent Bernoulli RV with

\[
P(X_{ij} = 1|Z_i = q, Z_j = l) = \pi_{ql}
\]
Mixnet: a flexible model

<table>
<thead>
<tr>
<th>Description</th>
<th>Graph</th>
<th>Q</th>
<th>Π</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erdos, no cluster</td>
<td><img src="image1" alt="Erdos Graph" /></td>
<td>1</td>
<td>$p$</td>
</tr>
</tbody>
</table>
| Hubs                         | ![Hubs Graph](image2) | 4  | \[
\begin{pmatrix}
0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 \\
0 & 0 & 1 & 0 \\
\end{pmatrix}
\] |
| cluster in "usual sense"     | ![Cluster Graph](image3) | 2  | \[
\begin{pmatrix}
1 & \varepsilon \\
\varepsilon & 1 \\
\end{pmatrix}
\] |
| Hierarchical                 | ![Hierarchical Graph](image4) | 5  | \[
\begin{pmatrix}
0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 \\
\end{pmatrix}
\] |
First Idea: Use maximum likelihood estimators

- **Complete data likelihood**

\[
\mathcal{L}(X, Z) = \sum_i \sum_q Z_{iq} \ln \alpha_q + \sum_{i<j} \sum_{q,l} Z_{iq} Z_{jl} \ln b(\pi_{ql}, X_{ij})
\]

where \( b(\pi_{ql}, X_{ij}) = \pi^{X_{ij}} q l (1 - \pi_{ql})^{(1-X_{ij})} \)

- **Observed data likelihood**

\[
\mathcal{L}(X) = \ln \sum_Z \exp \mathcal{L}(X, Z)
\]

- observed data likelihood requires a sum over \( Q^n \) terms: untractable
- EM-like strategies require \( \Pr(Z|X) \): untractable (no conditional independence).
Log-Likelihood of the model

**First Idea:** Use maximum likelihood estimators

- **Complete data likelihood**
  
  \[ \mathcal{L}(X, Z) = \sum_i \sum_q Z_{iq} \ln \alpha_q + \sum_{i < j} \sum_{q,l} Z_{iq} Z_{jl} \ln b(\pi_{ql}, X_{ij}) \]

  where \( b(\pi_{ql}, X_{ij}) = \pi_{ql}^{X_{ij}} (1 - \pi_{ql})^{(1-X_{ij})} \)

- **Observed data likelihood**
  
  \[ \mathcal{L}(X) = \ln \sum_{Z} \exp \mathcal{L}(X, Z) \]

  - observed data likelihood requires a sum over \( Q^n \) terms: untractable
  - EM-like strategies require \( \Pr(Z|X) \): untractable (no conditional independence).
Log-Likelihood of the model

**First Idea:** Use maximum likelihood estimators

- **Complete data likelihood**

  \[ L(X, Z) = \sum_i \sum_q Z_{iq} \ln \alpha_q + \sum_{i<j} \sum_{q,l} Z_{iq} Z_{jl} \ln b(\pi_{ql}, X_{ij}) \]

  where \( b(\pi_{ql}, X_{ij}) = \pi_{ql}^{X_{ij}} (1 - \pi_{ql})^{(1-X_{ij})} \)

- **Observed data likelihood**

  \[ L(X) = \ln \sum_Z \exp L(X, Z) \]

- observed data likelihood requires a sum over \( Q^n \) terms: **untractable**

- EM-like strategies require \( \Pr(Z|X) \): **untractable** (no conditional independence).
Variational Inference: Pseudo Likelihood

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

- Optimize in $\mathcal{R}_X$ the function $\mathcal{I}(\mathcal{R}_X)$ given by:
  
  $$\mathcal{I}(\mathcal{R}_X[Z]) = \mathcal{L}(X) - \text{KL}(\mathcal{R}_X[Z], \Pr(Z|X))$$
  
  $$= \mathcal{H}(\mathcal{R}_X[Z]) - \sum_Z \mathcal{R}_X[Z] \mathcal{L}(X, Z)$$

- For simple $\mathcal{R}_X$, $\mathcal{I}(\mathcal{R}_X[Z])$ is tractable.

- At best, $\mathcal{R}_X = \Pr(Z|X)$ and $\mathcal{I}(\mathcal{R}_X[Z]) = \mathcal{L}(X)$. 
Variational Inference: Pseudo Likelihood

**Main Idea:** Replace complicated $\Pr(Z|X)$ by a simple $R_X[Z]$ such that $\text{KL}(R_X[Z], \Pr(Z|X))$ is minimal.

- Optimize in $R_X$ the function $J(R_X)$ given by:

$$J(R_X[Z]) = L(X) - \text{KL}(R_X[Z], \Pr(Z|X)) = H(R_X[Z]) - \sum_Z R_X[Z]L(X, Z)$$

- For simple $R_X$, $J(R_X[Z])$ is tractable,

- At best, $R_X = \Pr(Z|X)$ and $J(R_X[Z]) = L(X)$. 
**Main Idea:** Replace complicated $Pr(Z|X)$ by a simple $R_X[Z]$ such that $KL(R_X[Z], Pr(Z|X))$ is minimal.

- Optimize in $R_X$ the function $J(R_X)$ given by:

  $$J(R_X[Z]) = L(X) - KL(R_X[Z], Pr(Z|X))$$

  $$= H(R_X[Z]) - \sum_Z R_X[Z]L(X, Z)$$

- For simple $R_X$, $J(R_X[Z])$ is tractable,

- At best, $R_X = Pr(Z|X)$ and $J(R_X[Z]) = L(X)$. 
2 Steps Iterative Algorithm

- **Step 1** Optimize $J(R_X[Z])$ w.r.t. $R_X[Z]$:
  - Restriction to a "comfortable" class of functions,
  - $R_X[Z] = \prod_i h(Z_i; \tau_i)$, with $h(., \tau_i)$ the multinomial distribution,
  - $\tau_{iq}$ is a variational parameter to be optimized using a fixed point algorithm:

\[
\tilde{\tau}_{iq} \propto \alpha_q \prod_{j \neq i} \prod_{l=1}^{Q} b(\pi_{ql}, X_{ij}) \tilde{\tau}_{jl}
\]

- **Step 2** Optimize $J(R_X[Z])$ w.r.t. $(\alpha, \pi)$:
  - Constraint: $\sum_q \alpha_q = 1$

\[
\tilde{\alpha}_q = \frac{\sum_i \tilde{\tau}_{iq}}{n}
\]

\[
\tilde{\pi}_{ql} = \frac{\sum_{ij} \tilde{\tau}_{iq} \tilde{\tau}_{jl} X_{ij}}{\sum_{ij} \tilde{\tau}_{iq} \tilde{\tau}_{jl}}
\]
2 Steps Iterative Algorithm

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

  \[
  \tilde{\tau}_{iq} \propto \alpha_q \prod_i \prod_{j \neq i} b(\pi_{ql}, X_{ij}) \tilde{\tau}_{jl}
  \]

- **Step 2** Optimize $\mathcal{J}(\mathcal{R}_X[Z])$ w.r.t. $(\alpha, \pi)$:  
  → Constraint: $\sum_q \alpha_q = 1$

  \[
  \tilde{\alpha}_q = \sum_i \tilde{\tau}_{iq} / n \\
  \tilde{\pi}_{ql} = \sum_{ij} \tilde{\tau}_{iq} \tilde{\tau}_{jl} X_{ij} / \sum_{ij} \tilde{\tau}_{iq} \tilde{\tau}_{jl}
  \]
Model Selection Criterion

- BIC-like criterion to select the number of classes:

- The likelihood can be split: $\mathcal{L}(X,Z|Q) = \mathcal{L}(X|Z, Q) + \mathcal{L}(Z|Q)$.

- These terms can be penalized separately:

  \[
  \mathcal{L}(X|Z, Q) \rightarrow \text{pen}_{X|Z} = \frac{Q(Q+1)}{2} \log \frac{n(n-1)}{2}
  \]

  \[
  \mathcal{L}(Z|Q) \rightarrow \text{pen}_{Z} = (Q-1) \log(n)
  \]

  \[
  \text{ICL}(Q) = \max_{\theta} \mathcal{L}(X, \tilde{Z}|\theta, m_Q) - \frac{1}{2} \left( \frac{Q(Q+1)}{2} \log \frac{n(n-1)}{2} - (Q-1) \log(n) \right)
  \]
Model Selection Criterion

- BIC-like criterion to select the number of classes:

- The likelihood can be split: \( \mathcal{L}(X, Z|Q) = \mathcal{L}(X|Z, Q) + \mathcal{L}(Z|Q) \).

- These terms can be penalized separately:

\[
\mathcal{L}(X|Z, Q) \rightarrow \text{pen}_{X|Z} = \frac{Q(Q + 1)}{2} \log \frac{n(n - 1)}{2}
\]

\[
\mathcal{L}(Z|Q) \rightarrow \text{pen}_{Z} = (Q - 1) \log(n)
\]

\[
\text{ICL}(Q) = \max_{\theta} \mathcal{L}(X, \tilde{Z}|\theta, m_Q) - \frac{1}{2} \left( \frac{Q(Q+1)}{2} \log \frac{n(n-1)}{2} - (Q - 1) \log(n) \right)
\]
Karate Club

- nodes are members of the club
- edges between 2 members if they have social relation outside the club
- known properties: the club has split away in two parts (cercles and squares).

Mixnet results for Karate Club

The split is recovered and the role of the leaders is underlined.
Transcriptional regulatory network of E. Coli

- nodes are operons
- edges between 2 operons if one regulates the other
- known properties: sparseness, no feed-back circuits, hierarchical organization.

Data from Shen-Or et al. Nature genetics, 2002
Mixnet results for RTN of E. Coli

Meta Hierarchical structure, Meta Single Input Modules and Feed Forward Loops.
Macaque Cortex Network

- nodes are cortical regions
- edges between 2 regions if one is connected to the other
- known properties: highly connected network, central and "provincial hubs".

Data from Sporns et al. PLoS one, 2007
Mixnet results for Cortex network

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>75.0</td>
<td>58.9</td>
<td>100.0</td>
<td>43.7</td>
<td>2.8</td>
<td>3.6</td>
<td>10.0</td>
<td>.</td>
</tr>
<tr>
<td>2</td>
<td>44.7</td>
<td>76.1</td>
<td>71.4</td>
<td>85.7</td>
<td>3.2</td>
<td>12.2</td>
<td>25.7</td>
<td>.</td>
</tr>
<tr>
<td>3</td>
<td>100.0</td>
<td>42.9</td>
<td>45.7</td>
<td>50.0</td>
<td>55.5</td>
<td>28.6</td>
<td>20.0</td>
<td>.</td>
</tr>
<tr>
<td>4</td>
<td>6.2</td>
<td>92.8</td>
<td>50.0</td>
<td>100.0</td>
<td>11.1</td>
<td>42.9</td>
<td>100.0</td>
<td>.</td>
</tr>
<tr>
<td>5</td>
<td>4.2</td>
<td>6.4</td>
<td>66.6</td>
<td>27.8</td>
<td>23.6</td>
<td>4.8</td>
<td>4.4</td>
<td>.</td>
</tr>
<tr>
<td>6</td>
<td>8.9</td>
<td>12.2</td>
<td>28.6</td>
<td>42.9</td>
<td>12.7</td>
<td>76.2</td>
<td>31.4</td>
<td>1.8</td>
</tr>
<tr>
<td>7</td>
<td>15.0</td>
<td>45.7</td>
<td>.</td>
<td>80.0</td>
<td>6.7</td>
<td>42.9</td>
<td>100.0</td>
<td>45.0</td>
</tr>
<tr>
<td>8</td>
<td>.</td>
<td>.</td>
<td>18.7</td>
<td>.</td>
<td>7.1</td>
<td>62.5</td>
<td>57.1</td>
<td>.</td>
</tr>
<tr>
<td>alpha</td>
<td>17.0</td>
<td>14.9</td>
<td>2.1</td>
<td>4.3</td>
<td>19.2</td>
<td>14.9</td>
<td>10.6</td>
<td>17.0</td>
</tr>
</tbody>
</table>

Central and provincial hubs well identified.
Food-web network

- The food web shows 5 levels of organization: plants (circle), herbivores (box), parasitoids (parallelogram), hyperparasitoids (triangle) and hyper-hyperparasitoids (diamond).
- A trophic link is considered between two insects when one insect is observed within one host.
- Known properties: hierarchic organization.

Data from Dawah et al. Journal of animal ecology, 1995, and Martinez...
Mixnet results for Food-Web network

The 5 levels are well identified plus a specific community. Local hierarchies are detected.
Conclusions

• Mixnet is a flexible model which allows to replace a complicated network by a simple meta-network.
• Freely available package, URL: http://pbil.univ-lyon1.fr/software/MixNet.
• Extension to Poisson and Gaussian X is already made.
• properties of variational estimates is in progress

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Franck Picard</td>
<td>UMR CNRS 5558, UCB Lyon 1</td>
<td>Method + Biological examples + Mixnet package</td>
</tr>
<tr>
<td>Vincent Miele</td>
<td>UMR CNRS 5558, UCB Lyon 1</td>
<td>Mixnet package</td>
</tr>
<tr>
<td>Stéphane Robin</td>
<td>UMR 518 AgroParisTech/INRA</td>
<td>Method</td>
</tr>
<tr>
<td>Mahendra Mariadassou</td>
<td>UMR 518 AgroParisTech/INRA</td>
<td>Method</td>
</tr>
<tr>
<td>Ludovic Cottret</td>
<td>UMR CNRS 5558, UCB Lyon 1</td>
<td>Mixnet package</td>
</tr>
</tbody>
</table>
Why do people represent data by a network (3) ?

- nodes may be of different sizes, adding more information
- edges may be colored, adding more information
- a movie may report the evolution of the network in time.

2200 persons (nodes) from the Framingham Heart study. Circles with red borders = women, circles with blue borders = men.

The size of each circle is proportional to BMI. The color of the circles indicates the person’s obesity status: yellow = obese person and green = nonobese person. The colors of the ties between the nodes indicate the relationship between them: purple denotes a friendship or marital tie and orange denotes a familial tie. N.A. Christakis et al. N Engl J Med, 2007, July
### Research Fronts rankings in Mathematics, sorted by Citations

<table>
<thead>
<tr>
<th>Rank</th>
<th>Research Front</th>
<th>Core Papers</th>
<th>Citations</th>
<th>Citations / paper</th>
<th>Mean year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>Complex networks</strong>, Statistical mechanics, Structure, Function, Evolution</td>
<td>3</td>
<td>3415</td>
<td>1138</td>
<td>2002.3</td>
</tr>
<tr>
<td>2</td>
<td>Current-driven magnetic domain wall motion</td>
<td>27</td>
<td>1385</td>
<td>51.3</td>
<td>2004.6</td>
</tr>
<tr>
<td>3</td>
<td>False discovery rate procedure</td>
<td>9</td>
<td>936</td>
<td>104</td>
<td>2003.3</td>
</tr>
<tr>
<td>4</td>
<td>Evolutionary game dynamics, <strong>Social networks</strong></td>
<td>12</td>
<td>382</td>
<td>31.83</td>
<td>2004.8</td>
</tr>
<tr>
<td>5</td>
<td>Porous plate using homotopy analysis method</td>
<td>12</td>
<td>359</td>
<td>29.92</td>
<td>2003.7</td>
</tr>
</tbody>
</table>
Biological systems show emergent properties that are not readily explainable by the study of their constituent parts...

New mathematical formalisms, based on graph theory, are emerging in order to represent and study the underlying interaction networks present in the cell. The understanding of the evolution and organization of these networks is changing the way scientists look at biology. More specifically:

- gene regulation
- metabolic pathways analysis
- protein-protein interactions
Representations of a correlation matrix by a graph

\[
\begin{pmatrix}
1 & 0.8 & 0.5 & 0.2 \\
1 & 0.1 & 0.1 \\
1 & 0.1 \\
1 \\
\end{pmatrix}
\]

Correlation matrix

Not readable if more than 10 items

Valued graph

Not readable if more than 10 items

Binary graph using 0.15 threshold

Clusters and hubs appear, but the result depends on the threshold used and the representation software.

Other graphical representations are possible: colored intensity matrix, PCA...
The focus is on topological questions

Are there clusters of similar nodes?

• existence of functional modules in biological networks, i.e. sub-networks which can produce outputs independently.

• Each cluster may be studied independently and thus we are allowed to work on lower sized networks.

• If a node A, with unknown properties, is in the same cluster that a well known node B, then we are tempted to attribute B’s properties to A.

• How many hubs? they are the most important nodes in the network.

• Resilience of the network’s connectivity to edge suppression.
Clustering methods

Hierarchical clustering  a well-known set of algorithms for a similarity matrix, many criteria proposed


Markov Clustering Algorithm  MCL simulates a flow on the graph by calculating successive powers of $X$. At each iteration an inflation step is applied to enhance the contrast between regions of strong or weak flow in the graph. The process converges towards a partition of the graph, with a set of high flow regions separated by boundaries with no flow, package from Von Dongen

Edge-betweenness clustering  Divisive method using the edge betweenness EB, (for an edge) : the number of shortest paths that pass through the edge. The edge with the highest EB is removed, Girvan & Newman
Clustering of nodes

Santa Fe Institute collaboration network: the nodes of the network represent scientists from the Santa Fe Institute and an edge is drawn between two nodes if the corresponding scientists have coauthored at least one publication during the calendar year 1999 or 2000.
Statistical models for networks
Statistical models for networks

More probabilistic models for networks are necessary to:

1. summarize the information contained in one network and compare networks
2. have a reference model to say if a particular motif is exceptional
3. predict the label of a node
4. predict the value of an edge
5. create artificial networks by simulation