Strategies for Online Inference of Network Mixtures

H. Zanghi, F. Picard , V. Miele and C. Ambroise

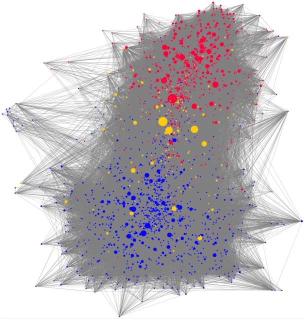
Lab. Biometry and Evolutionary Biology, Lyon
Lab. Statistics and Genome, Evry

Dublin, June 15th 2009

Outline

Motivations

- Study networks structure
- Find structural characteristics
- Find new structural information
- Reduce dimensionality
- Propose a synthetic representation



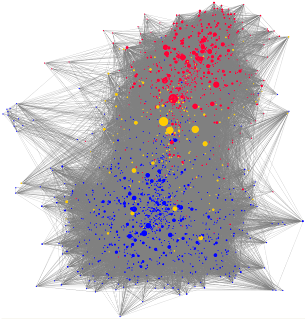
The US 2008 presidential Websphere

Avoiding the "module bottleneck"

- Single day snapshot of the websphere (Nov. 2007)
- Nodes : blogs, edges : citations
- 130,520 links and 1,870 sites

liberal	676
conserv.	1,026
indep.	168

- Political orientation defines unmistakable modules



The US 2008 presidential Websphere

Model-based strategies: a common objective

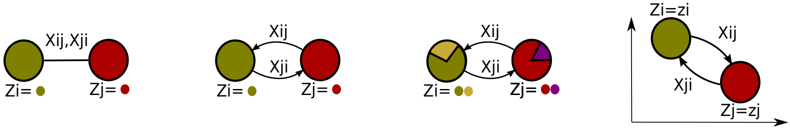
- V a set of vertices in $\{1, \ldots, n\}$,
- E a set of edges in $\{1, \ldots, n\}^2$,
- $\mathbf{X} = (X_{ij})$ the adjacency matrix such that $\{X_{ij} = 1\} = \mathbb{I}\{i \leftrightarrow j\}$.

Common hypothesis

- There exists a hidden label variable $\mathbf{Z}$ used to partition the network
- No *a priori* on the topologic features under investigation

Different modelling strategies for model-based clustering

Nowicki[4] Blocks	Daudin[2] MixNet	Airoldi[1] Mixed-Memb.	Handcock[3] LPCM



dyads	edges	mixed **Z**	continuous **Z**

Model-based clustering faces computational issues

- EM-like strategies are commonly used for ML estimation
- EM-like strategies are known to be poorly efficient from the speed of execution point of view
- Bayesian algorithms have been developed (BLOCKS, MMB)

Aim of the presentation

How to propose an efficient inference framework to perform model-based clustering on large networks ?

Outline

Presentation of MixNet

- $Z_{iq} = \mathbb{I}\{i \in q\}$ indep hidden variables:

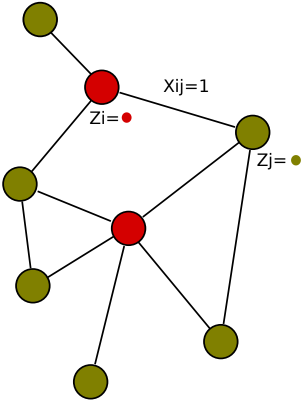
$$(\mathbf{Z}_i) \sim \mathcal{M}(1, \boldsymbol{\alpha})$$

- $\mathbf{X}|\mathbf{Z}$ are independent:

$$X_{ij}|\{Z_{iq}Z_{j\ell} = 1\} \sim \mathcal{B}(\pi_{q\ell})$$

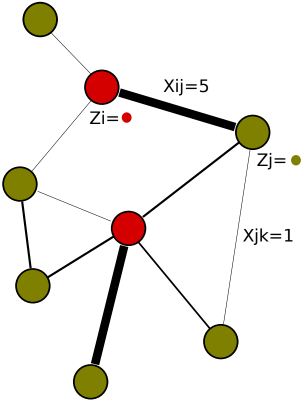
- Erdős-Rényi **Mix**ture for **Net**works

$$X_{ij} \sim \sum_{q\ell} \alpha_q \alpha_\ell \mathcal{B}_{q\ell}()$$



Generalization of MixNet to the exponential family

- Edges may contain information as well
- $\Pr\{\mathbf{X}|\mathbf{Z}\}$ may be more informative than Bernoulli
- $\mathbf{X}|\mathbf{Z}$ is supposed to belong to the exponential family, with natural parameter $\eta_{q\ell}$



Sufficient statistics of MixNet

- The model is defined by its conditional distribution $\Pr\{\mathbf{X}|\mathbf{Z}\}$

$$\log \Pr\{X_{ij}|Z_{iq}Z_{j\ell}=1;\eta_{q\ell}\} = \eta_{q\ell}^t h(X_{ij}) - a(\eta_{q\ell}) + b(X_{ij}),$$

- $(\mathbf{X},\mathbf{Z})$ belongs to the exponential family with sufficient statistics:

$$T(\mathbf{X},\mathbf{Z}) = \begin{cases} N_q & = \sum_i Z_{iq}, \\ H_{q\ell}(\mathbf{X},\mathbf{Z}) & = \sum_{ij} Z_{iq}Z_{j\ell}h(X_{ij}), \\ G_{q\ell}(\mathbf{Z}) & = N_q N_\ell, \end{cases}$$

- Reparametrization

$$\psi_{q\ell} = \frac{\partial a(\eta_{q\ell})}{\partial \eta_{q\ell}} = \frac{H_{ql}(\mathbf{X},\mathbf{Z})}{G_{ql}(\mathbf{Z})}$$

Estimation strategies

- EM-like algorithms to optimize the observed-data likelihood

$$\log \Pr(\mathbf{X};\beta) = \log\left(\sum_{\mathbf{Z}} \log \Pr(\mathbf{X},\mathbf{Z};\beta)\right).$$

- In the complete-data framework, the quantity of interest is:

$$\mathbb{E}\big(\log \Pr(\mathbf{X},\mathbf{Z};\beta)|\mathbf{X}\big).$$

- The difficult step : calculation of $\Pr(\mathbf{Z}|\mathbf{X};\beta)$ ($E-$step).

Handling missing data

- $\Pr(\mathbf{Z}|\mathbf{X};\beta)$ can not be factorized (*Daudin et al. 2008*)
- Prediction & Classification algorithms

$$\hat{\mathbf{Z}} = \arg\max \Pr(\mathbf{Z}|\mathbf{X})$$

Zanghi et al. 2008

- Mean field approximation & variational algorithm

$$\tilde{\mathbf{Z}} \simeq \mathbb{E}_{?}(\mathbf{Z}|\mathbf{X})$$

Daudin et al. 2008

- Gibbs sampling & Stochastic Approximation algorithms

$$\Pr\{\mathbf{Z}|\mathbf{X}\} \simeq \Pr\{Z_{iq} = 1|\mathbf{X}, \mathbf{Z}_{\setminus i}\}$$

Variational inference

- Approximate $\Pr(\mathbf{Z}|\mathbf{X};\boldsymbol{\beta})$ by $\mathcal{R}_{\mathbf{X}}[\mathbf{Z}]$ such that that minimizes

$$KL(\mathcal{R}_{\mathbf{X}}[\mathbf{Z}], \Pr(\mathbf{Z}|\mathbf{X};\beta))$$

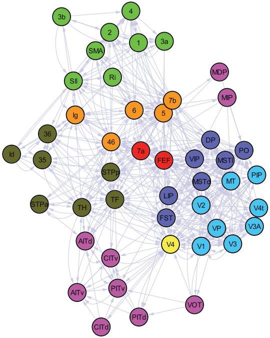
- $\mathcal{R}_{\mathbf{X}}[\mathbf{Z}]$ can be chosen among convenient distributions such that:

$$\log \mathcal{R}_{\mathbf{X}}[\mathbf{Z}] = \sum_{iq} Z_{iq} \log \tau_{iq}.$$

- $(\boldsymbol{\tau})$ are variational parameters updated during the variational-E-Step

Applications of MixNet to biological networks

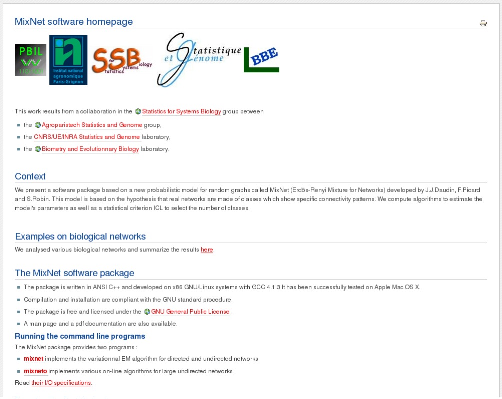
- Transcriptional Regulatory Network of E. Coli
- Macaque cortex network
- Metabolic network
- Foodweb



http://stat.genopole.cnrs.fr/software/mixnet/

Picard et al. 2009

The MixNet software



MixNet software homepage

This work results from a collaboration in the Statistics for Systems Biology group between

- the Agroparistech Statistics and Genome group,
- the CNRS/UE/INRA Statistics and Genome laboratory,
- the Biometry and Evolutionnary Biology laboratory.

Context

We present a software package based on a new probabilistic model for random graphs called MixNet (Erdős-Renyi Mixture for Networks) developed by J.J.Daudin, F.Picard and S.Robin. This model is based on the hypothesis that real networks are made of classes which show specific connectivity patterns. We compute algorithms to estimate the model's parameters as well as a statistical criterion ICL to select the number of classes.

Examples on biological networks

We analysed various biological networks and summarize the results here.

The MixNet software package

- The package is written in ANSI C++ and developed on x86 GNU/Linux systems with GCC 4.1.3 It has been successfully tested on Apple Mac OS X.
- Compilation and installation are compliant with the GNU standard procedure.
- The package is free and licensed under the GNU General Public License .
- A man page and a pdf documentation are also available.

Running the command line programs

The MixNet package provides two programs :

- **mixnet** implements the variationnal EM algorithm for directed and undirected networks
- **mixneto** implements various on-line algorithms for large undirected networks

Read their I/O specifications.

http://stat.genopole.cnrs.fr/software/mixnet/

Outline

Online Inference for large networks

- Every estimation method is concerned by the size of the dataset.
- 2 situations :
  $\longrightarrow$ the network is of fixed size and large,
  $\longrightarrow$ the network is growing over time.
- EM-like strategies are known to converge slowly and are very demanding from the memory point of view
- Online algorithms are incremental algorithms which recursively update parameters using current parameters and additional information provided by new observations

Principle of online algorithms - 1

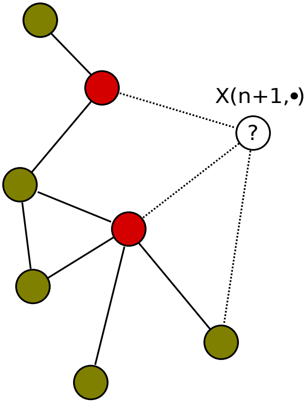
- The adjacency matrix change sequentially:

$$\mathbf{X}^{[n]} = \{X_{ij}\}_{i,j=1}^{n}, \ \mathbf{Z}^{[n]}$$

- The dimension of the added information is $\mathcal{O}(n)$

$$\mathbf{X}_{n+1,\bullet} = \{X_{n+1,j}, \ j \in \mathcal{V}\}$$

- $T(\mathbf{X}, \mathbf{Z})$ is additive when the network grows



$$T(\mathbf{X}^{[n+1]}, \mathbf{Z}^{[n+1]}) = T(\mathbf{X}^{[n]}, \mathbf{Z}^{[n]}) + T(\mathbf{X}_{n+1,\bullet}, \mathbf{Z}^{[n+1]}).$$

Principle of online algorithms - 2

- Recall that the aim is to maximize

$$\mathcal{Q}(\beta|\beta') \simeq \mathbb{E}_{\mathcal{R}_{\mathbf{X}}}\left(\log \Pr\{\mathbf{X},\mathbf{Z};\beta\}|\mathbf{X}\right)$$

- If $\mathcal{R}_{\mathbf{X}}$ is a factorizable distribution, the parameters can be updated using only the additional information provided by the new node

$$\begin{aligned}\mathbb{E}_{\mathcal{R}_{\mathbf{X}}}\left\{T(\mathbf{X}^{[n+1]},\mathbf{Z}^{[n+1]})|\mathbf{X}^{[n+1]}\right\} &= \mathbb{E}_{\mathcal{R}_{\mathbf{X}}}\left\{T(\mathbf{X}^{[n]},\mathbf{Z}^{[n]})|\mathbf{X}^{[n]}\right\} \\ &+ \mathbb{E}_{\mathcal{R}_{\mathbf{X}}}\left\{T(\mathbf{X}_{n+1,\bullet},\mathbf{Z}^{[n+1]})|\mathbf{X}_{n+1,\bullet}\right\}\end{aligned}$$

Online variational inference

- Since $\mathcal{R}_{\mathbf{X}}$ is chosen factorizable, the derivation of the online-variational EM is direct:

$$\tau_{n+1,q} \propto \alpha_q^{[n]} \exp\left\{ \sum_{\ell=1}^{Q} \sum_{j=1}^{n} \tau_{j\ell}^{[n]} \left(\eta_{q\ell}^{[n]} h(X_{n+1,j}) + a(\eta_{q\ell}^{[n]}) \right) \right\}.$$

- Parameters update is given by:

$$\begin{aligned} \alpha_q^{[n+1]} &= \frac{1}{n+1} \left(\sum_{i=1}^{n} \tau_{iq}^{[n]} + \tau_{n+1,q} \right), \\ \psi_{q\ell}^{[n+1]} &= \frac{\mathbb{E}_{\mathcal{R}^{[n]}} \left(H_{q\ell}(\mathbf{X}^{[n+1]}, \mathbf{Z}^{[n+1]}) \right)}{\mathbb{E}_{\mathcal{R}^{[n]}} \left(G_{q\ell}(\mathbf{Z}^{[n+1]}) \right)}. \end{aligned}$$

Online SAEM

- We use a Gibbs sampler generating a sequence of $\mathbf{Z}$ approaching

$$\Pr\{\mathbf{Z}|\mathbf{X}\} \simeq \Pr\{Z_{iq} = 1|\mathbf{X}, \mathbf{Z}_{\backslash i}\}$$

- In the online context we show that

$$\Pr\{Z_{n+1,q} = 1|\mathbf{X}^{[n+1]}, \mathbf{Z}^{[n]}\} \propto \exp\left\{\beta^t T(\mathbf{X}_{n+1,\bullet}, \mathbf{Z}^{[n]}, Z_{n+1,q})\right\}$$

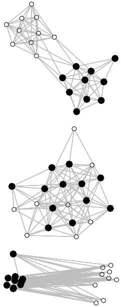
- Parameters are updated using $T\left(\mathbf{X}^{[n+1]}, \left\{\mathbf{Z}^{[n]}, \tilde{Z}_{n+1}\right\}\right)$

$$\begin{aligned}
\alpha_q^{[n+1]} &= \frac{1}{n+1}\left(\sum_{i=1}^{n} Z_{iq}^{[n]} + \tilde{Z}_{n+1,q}\right), \\
\psi_{q\ell}^{[n+1]} &= \frac{H_{q\ell}\left(\mathbf{X}^{[n+1]}, \left\{\mathbf{Z}^{[n]}, \tilde{Z}_{n+1}\right\}\right)}{G_{q\ell}\left(\left\{\mathbf{Z}^{[n]}, \tilde{Z}_{n+1}\right\}\right)}.
\end{aligned}$$

Outline

Simulation with affiliation networks

Model	ϵ	λ
1	0.1	0.9
2	0.5	0.5
3	0.6	0.4
4	0.7	0.3
5	0.8	0.2



Generated graphs with

- $Q \in \{2, 5, 20\}$ clusters in the same proportions
- $n \in \{100, 500, 1000\}$ nodes.

$$\boldsymbol{\pi} = \begin{bmatrix} \lambda & & \epsilon \\ & \ddots & \\ \epsilon & & \lambda \end{bmatrix}$$

Synthetic Data - Bias

	o-SAEM		o-Variational		o-CEM		MixNet	
(ϵ, λ)	$\hat{\epsilon}$	$\hat{\lambda}$	$\hat{\epsilon}$	$\hat{\lambda}$	$\hat{\epsilon}$	$\hat{\lambda}$	$\hat{\epsilon}$	$\hat{\lambda}$
(0.1,0.9)	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
(0.5,0.5)	-0.02	0.01	-0.00	0.00	-0.02	0.01	-0.00	0.00
(0.6,0.4)	-0.16	0.04	-0.15	0.04	-0.13	0.03	-0.00	-0.00
(0.7,0.3)	-0.01	0.00	-0.00	0.00	0.00	-0.00	0.00	-0.00
(0.8,0.2)	0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00

Bias is slighly more important in the online context compared with MixNet for intermediate situations

Synthetic Data - Variance

	o-SAEM		o-Variational		o-CEM		MixNet	
(ϵ, λ)	$\hat{\epsilon}$	$\hat{\lambda}$	$\hat{\epsilon}$	$\hat{\lambda}$	$\hat{\epsilon}$	$\hat{\lambda}$	$\hat{\epsilon}$	$\hat{\lambda}$
(0.1,0.9)	0.17	0.09	0.17	0.15	0.17	0.09	0.17	0.09
(0.5,0.5)	1.04	0.37	0.17	0.17	1.57	0.82	0.16	0.13
(0.6,0.4)	1.47	0.52	2.23	0.70	2.60	1.39	0.67	0.21
(0.7,0.3)	0.40	0.19	0.29	0.13	0.29	0.16	0.29	0.16
(0.8,0.2)	0.22	0.15	0.22	0.09	0.23	0.15	0.23	0.15

The online-variational strategy shows better results on variance
(not for intermediate situations)

Synthetic Data - Rand Index

	o-SAEM		o-Variational		o-CEM		MixNet	
(ϵ, λ)	mean	sd	mean	sd	mean	sd	mean	sd
(0.1,0.9)	1.00	0.01	1.00	0.01	1.00	0.01	1.00	0.01
(0.5,0.5)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(0.6,0.4)	0.14	0.13	0.10	0.15	0.25	0.16	0.96	0.04
(0.7,0.3)	0.98	0.01	0.98	0.01	0.98	0.01	0.98	0.01
(0.8,0.2)	0.98	0.02	0.98	0.02	0.98	0.02	0.98	0.02

Classification performance are better for CEM
(but biased estimators)

Synthetic Data - Gain in speed

	o-SAEM		o-Variational		o-CEM		MixNet	
N	*rand*	*time*	*rand*	*time*	*rand*	*time*	*rand*	*time*
100	0.14	0.07	0.14	0.11	0.14	0.07	0.26	0.21
250	0.47	0.76	0.48	0.77	0.48	0.74	0.99	2.08
500	0.64	0.97	0.67	1.02	0.66	0.95	1	25.00
750	0.82	2.20	0.83	2.36	0.83	2.14	1	125.30
1000	0.91	9.44	0.92	9.60	0.92	9.37	1	805.93
2000	0.98	147.67	0.99	148.31	0.99	147.33	1	13136.66

The gain in speed is of order 1000 !

First conclusions on simulations

- Every algorithm shows negligible bias and small variability for structured models
- Online Variational and online CEM give the best performance in terms of clustering accuracy
- The time of execution is drastically reduced

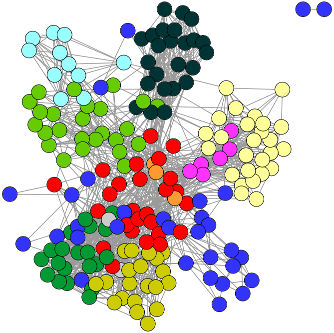
The online Variational is chosen with the best trade-off bias/accuracy/speed

Simulation of realistic structures

- Estimate MixNet parameters on a real network with complex structure
  $\hat{Q} = 11$

	1	2	3	4	5	6	7	8	9	10	11
1	8	.	.	13	.	.	15	.	.	.	.
2	.	100	.	.	.	19	89	6	.	66	.
3	.	.	39	83	12	6	10	.	.	.	.
4	13	.	83	100	72	38	67	.	.	.	.
5	.	.	12	72	83	17	20	.	.	.	.
6	.	19	6	38	17	15	60	.	.	.	.
7	15	89	10	67	20	60	100	21	.	19	19
8	.	6	.	.	.	.	21	35	.	.	.
9	.	.	.	.	.	.	.	.	93	.	.
10	.	66	.	.	.	.	19	.	.	22	.
11	.	.	.	.	.	.	19	.	.	.	55

Estimated between-group connectivity matrix

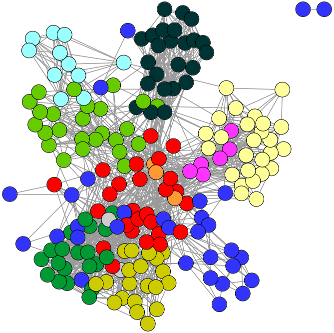


The French Political Websphere 2006

Performance on growing networks

- Generate network structures with increasing nodes
- Estimate the parameters using the online Variational algorithm

# nodes	# edges	rand	cpu (s)
0 + 200	3131.72	0.94	0.9
200 + 200	50316.32	0.998	0.4
400 + 400	12486.24	0.999	1.4
800 + 800	201009.5	1	5.7
1600 + 1600	803179.6	1	22.8
3200 + 3200	3202196	1	91.9
6400 +6400	12804008	1	371.1



The French Political Websphere 2006

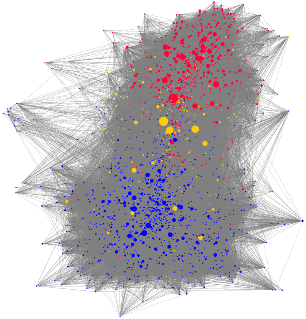
Outline

Uncovering structure in a WebSphere

- 130,520 links and 1,870 sites

liberal	676
conserv.	1,026
indep.	168

- Impossible to analyze with current algorithms due to the large size of the network



The US 2008 presidential Websphere

Estimated between-group connectivity matrix

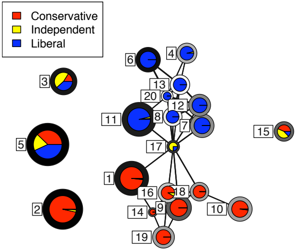
		Conservative								Liberal										
	17	1	2	10	9	14	16	18	19	11	4	6	7	8	12	13	20	3	5	15
17	.	54	.	.	64	.	52	66	.	62	.	.	67	67	62	43	100	.	.	.
1	54	.	.	.	.	54	.	.	.	.	.	.	.	.	.	.	.	.	.	.
2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
10	.	.	.	58	.	.	.	47	.	.	.	.	.	.	.	.	.	.	.	.
9	64	.	.	.	.	56	61	.	40	.	.	.	.	.	.	.	.	.	.	.
14	.	54	.	.	56	.	55	72	40	.	.	.	.	.	.	.	.	.	.	.
16	52	.	.	.	61	55	73	60	56	.	.	.	.	.	.	.	.	.	.	.
18	66	.	.	47	.	72	60	58	.	.	.	.	.	.	.	.	.	.	.	.
19	.	.	.	.	40	40	56	.	57	.	.	.	.	.	.	.	.	.	.	.
11	62	.	.	.	.	.	.	.	.	.	.	.	.	47	.	.	40	.	.	.
4	.	.	.	.	.	.	.	.	.	.	65	.	.	.	.	42	40	.	.	.
6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	88	99	.	.	.
7	67	.	.	.	.	.	.	.	.	.	.	.	47	76	49	.	76	.	.	.
8	67	.	.	.	.	.	.	.	.	47	.	.	76	90	74	81	98	.	.	.
12	62	.	.	.	.	.	.	.	.	.	.	.	49	74	45	92	98	.	.	.
13	43	.	.	.	.	.	.	.	.	.	42	88	.	81	92	95	99	.	.	.
20	100	.	.	.	.	.	.	.	.	40	42	99	76	98	98	99	100	.	.	.
3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
15	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N_q	4	214	407	66	56	1	24	19	36	26	58	207	51	20	37	23	3	81	514	23

Political communities only communicate through main US portals

(C17:`nytimes`,`wash.post`,`cnn`,`msn`)

MixNet Results

- "Cyber Balkanization" (very few cross-talks)
- Mass-media web-sites : central hubs that make oponent websites communicate

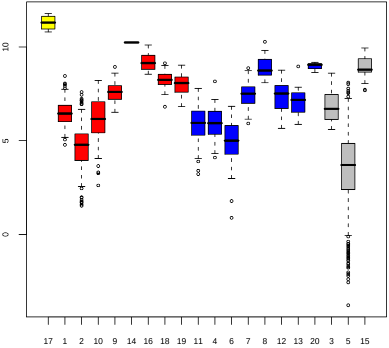


Communications between MixNet classes

Organization of substructures

- Each community is structured by a few set of influencial blogs

  `reachm`, `mahablog`, `juancole`, `foxnews`
- "Core" structure of blogs that are very efficient in terms of citation
- Professional way of using the web



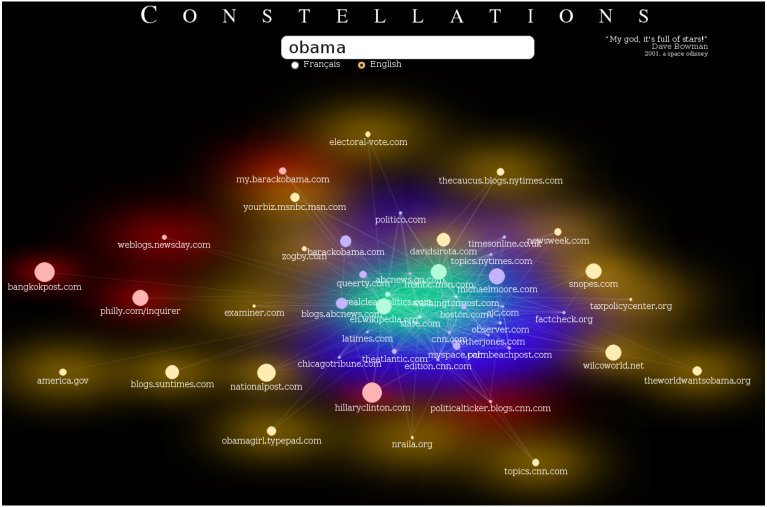
Centrality of MixNet classes

Outline

Conclusions

- We proposed online versions of different EM-based algorithms with application to the estimation of mixtures for Networks
- Online versions allows us to estimate the parameters on huge datasets or on time-growing networks
- The `MixNet` software already proposes this option
- Online methods could be applied to Bayesian algorithms (online-BLOCKS ?, online-MMB ?)

Constellations & MixNet classes on the web



http://constellations.labs.exalead.com

E.M. Airoldi, D.M. Blei, S.E. Fienberg, and E.P. Xing.
Mixed-membership stochastic blockmodels.
JMLR, 9:1981–2014, 2008.

J.J. Daudin, F. Picard, and S. Robin.
A mixture model for random graph.
Statistics and computing, 18(2):1–36, 2008.

MS. Handcock, A. Raftery, and J.M. Tantrum.
Model based clustering for social networks.
JRSS A, 2006.

K. Nowicki and T.A.B. Snijders.
Estimation and prediction for stochastic blockstructures.
Journal of the American Statistical Association, 96(455):1077–1090, 2001.

F. Picard, V. Miele, J.-J. Daudin, L. Cottret, and S. Robin.
Uncovering structure in biological networks using mixnet.
BMC Bioinformatics, to appear, 2009.

H. Zanghi, C. Ambroise, and V. Miele.
Fast online graph clustering via Erdős–Rényi mixture.
Pattern Recognition, 41(12):3592–3599, 2008.

Outline

Model selection

- When no information is available on the number of groups

$$\log \mathcal{L}(\mathbf{X}, \mathbf{Z}|m_Q) = \int_B \log \mathcal{L}(\mathbf{X}, \mathbf{Z}|\boldsymbol{\beta}, m_Q) g(\boldsymbol{\beta}|m_Q) \mathrm{d}\boldsymbol{\beta}.$$

$$\log \mathcal{L}(\mathbf{X}, \mathbf{Z}|m_Q) = \log \mathcal{L}(\mathbf{X}|\mathbf{Z}, m_Q) + \log \mathcal{L}(\mathbf{Z}|m_Q).$$

- These terms can be penalized separately : (Bernoulli ex.)

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

$$ICL(m_Q) = \max_{\boldsymbol{\beta}} \log \mathcal{L}(\mathbf{X}, \tilde{\mathbf{Z}}|\boldsymbol{\beta}, m_Q) - \mathrm{pen}_{\mathbf{X}|\mathbf{Z}} - \mathrm{pen}_{\mathbf{Z}}$$

Parameters update in the Bernoulli case for the online SAEM algorithm

$$
\begin{aligned}
\gamma_{q\ell}^{[n+1]} &= \frac{N_q^{[n]} N_\ell^{[n]}}{Z_{n+1,q} N_\ell^{[n]} + Z_{n+1,\ell} N_q^{[n]}}, \\
\xi_{q\ell}^{[n+1]} &= Z_{n+1,q} \sum_{j=1}^{n} Z_{j\ell}^{[n]} X_{n+1,j} + Z_{n+1,\ell} \sum_{i=1}^{n} Z_{iq}^{[n]} X_{i,n+1}, \\
\zeta_{q\ell}^{[n+1]} &= Z_{n+1,q} N_\ell^{[n]} + Z_{n+1,\ell} N_q^{[n]}.
\end{aligned}
$$

$$
\pi_{q\ell}^{[n+1]} = \gamma_{q\ell}^{[n+1]} \pi_{q\ell}^{[n]} + (1 - \gamma_{q\ell}^{[n+1]}) \frac{\xi_{q\ell}^{[n+1]}}{\zeta_{q\ell}^{[n+1]}}.
$$

Parameters update in the Bernoulli case for the online variational algorithm

$$
\begin{aligned}
\gamma_{q\ell}^{[n+1]} &= \frac{\mathbb{E}_{\mathcal{R}^{[n]}}\left(N_q^{[n]}\right)\mathbb{E}_{\mathcal{R}^{[n]}}\left(N_\ell^{[n]}\right)}{\tau_{n+1,q}\mathbb{E}_{\mathcal{R}^{[n]}}\left(N_\ell^{[n]}\right)+\tau_{n+1,\ell}\mathbb{E}_{\mathcal{R}^{[n]}}\left(N_q^{[n]}\right)}, \\
\mathbb{E}_{\mathcal{R}^{[n]}}\left(\xi_{q\ell}^{[n+1]}\right) &= \tau_{n+1,q}\sum_{j=1}^{n}\tau_{j\ell}^{[n]}X_{n+1,j}+\tau_{n+1,\ell}\sum_{i=1}^{n}\tau_{iq}^{[n]}X_{i,n+1}, \\
\mathbb{E}_{\mathcal{R}^{[n]}}\left(\zeta_{q\ell}^{[n+1]}\right) &= \tau_{n+1,q}\mathbb{E}_{\mathcal{R}^{[n]}}\left(N_\ell^{[n]}\right)+\tau_{n+1,\ell}\mathbb{E}_{\mathcal{R}^{[n]}}\left(N_q^{[n]}\right).
\end{aligned}
$$

$$
\pi_{q\ell}^{[n+1]} = \gamma_{q\ell}^{[n+1]}\pi_{q\ell}^{[n]} + (1-\gamma_{q\ell}^{[n+1]})\frac{\mathbb{E}_{\mathcal{R}^{[n]}}\left(\xi_{q\ell}^{[n+1]}\right)}{\mathbb{E}_{\mathcal{R}^{[n]}}\left(\zeta_{q\ell}^{[n+1]}\right)}.
$$