Models of networks and mixed membership stochastic blockmodels

Edo Airoldi

Department of Statistics, Harvard
Broad Institute of Harvard and MIT

Guest lecture for EE380L at UT Austin (Prof. J Ghosh), November 10, 2011

Agenda

- Overview
- Models of networks
- Mixed membership blockmodels
  1. Inference
  2. Results
- Concluding remarks

Overview

- Structured data vs. latent dependence structure
  - Leveraging observed (noisy) structure for estimation
  - As opposed to dim redux, graphical models, sparsity, …
- Technical challenges
  - Abandon convenient representations of dependence
  - Deal with structured measurements and interfering units
- This talk
  - Statistical problems when structure is expressed by a graph

What is a complex network?

- Define as a collection of measurements on pairs of sampling units and of unit-specific attributes
- Traditionally, can only choose 2 out of 3
  1. Large scale, e.g. millions of nodes
  2. Realistic
  3. Completely mapped, or to a large extent
- Today, a number of systems fall under this data setting that satisfy all three characteristics

A few examples

- Internet, WWW and Wikipedia
- Signaling pathways and metabolic networks
- JStor and scientific literature
- Cell-phone data, e.g. Rwanda, UK, ATT
- Yahoo and other instant messaging systems
- Linked-In and Facebook
- Blogs and Twitter

Rich, interdisciplinary literature

- Historical notes

  Moreno formalizes the sociogram (’34), Sociometry (‘37)

  50s: Sociology (Coleman et al. ‘57), Mathematics (Erdos & Reniy ‘59, Gilbert ‘59), Psychology (Milgram ‘67, ’69)

  70s: Statistics (Holland, Leinhardt, Fienberg, Wasserman)

  90s: Computer Science (Faloutsos[3] ’99), Physics (Huberman & Adamic ‘99, Albert & Barabasi ‘99)

Statistical issues in network analysis

- Representation and compressed sensing

  How to smoothly represent the space of all graph structures?

  Motifs, metrics, spectral, …, semi-parametric

- Population models

  Sample size? Notions of variability? *(See survey paper)*

- Diffusion of information on a network

  How to infer who talks to whom from aggregate traffic?

Statistical issues in network analysis

- Confidence sets, tests, GoF, model selection
  - How to establish confidence sets for network structure?
  - The Newman-Girvan modularity score is inconsistent
- Inference from a sample
  - CDC sponsored more than 90 studies to date using RDS
  - Are network sampling designs ignorable? No.
- Causal inference with interference
  - How to separate peer-influence effects from homophily?

Some details to think about

- Easy to measure things. Hard to pose questions.
  May not really know what any node or link means.
- What does $Y_{ij}=0$ mean?
- Valued measurements and censoring.
- Notion of variability. (sample size, populations)
- Global properties must be non-trivial outcomes of the composition of local properties and structures

Agenda

- Overview
- Models of networks
- Mixed membership blockmodels
  1. Inference
  2. Results
- Concluding remarks

Network modeling 101

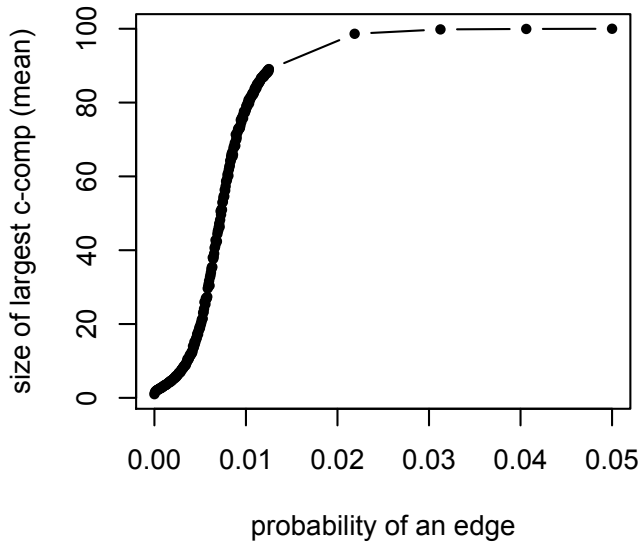
- Graphs or networks?
- Usually a graph is defined as, $G = (V,E)$
- For the purpose of this seminar, $G = (1{:}N, Y_{N\times N})$
- Complex networks, $G = (1{:}N, Y_{N\times N}, X_{N\times P})$
- Random graphs via $P(G|\Theta)$ or $P(Y|\Theta)$
- Frequentist or Bayes?

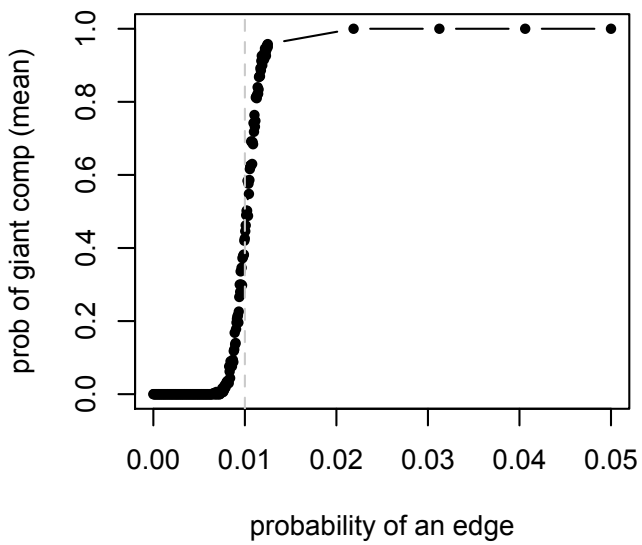
Erdös-Renyi-Gilbert

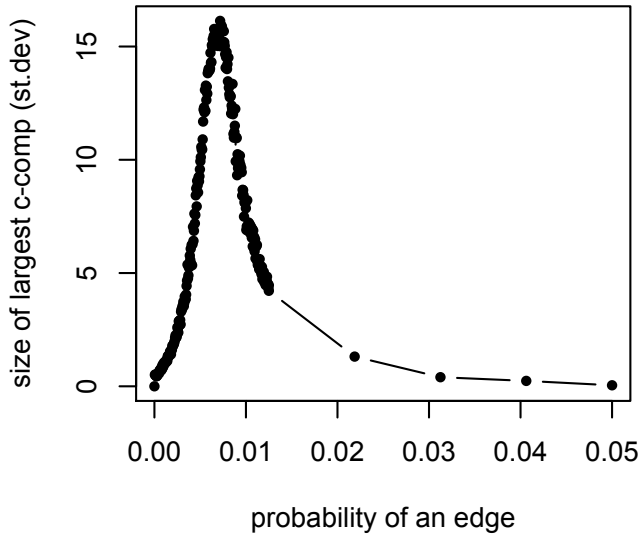
- The most widely known random graph model
- Binary edges are sampled independently
  G(N,θ): sample Y_{ij} from Bernoulli(θ) for i,j=1..N
  G(N,M): sample Y from SRS(θ,M)
- Likelihood for G(N,θ)
  $P(Y|\Theta) = \Pi_{ij}\ \theta^{Yij}\ (1-\theta)^{(1-Yij)}$

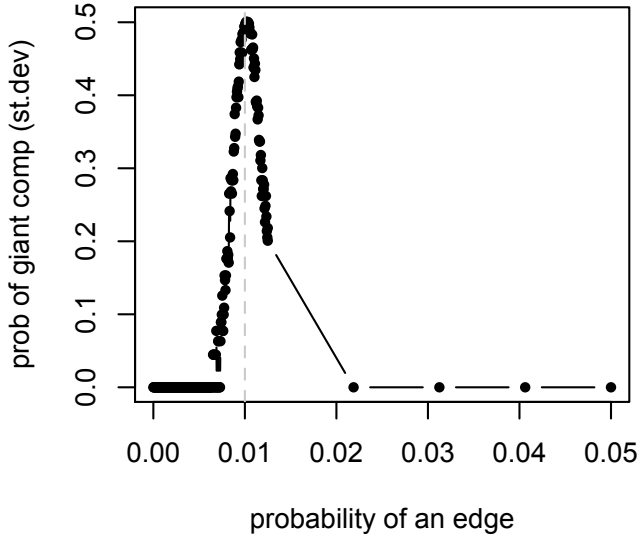
Emergence of the giant component

- ER studied G(N,M) as $\theta=M/\binom{N}{2}$ increases in [0,1]
- For a graph with N nodes, $\theta=1/N$ is a critical value
  1. If $\theta<1/N$, no connected components of size larger than O(log N) will exist in the graph, as $N\uparrow\infty$
  2. If $\theta=1/N$, largest connected component of size $O(N^{2/3})$ will exist in the graph, as $N\uparrow\infty$
  3. If $\theta>1/N$, unique connected component of size O(N) will exist in the graph, as $N\uparrow\infty$. No other components with more than O(log N) will exist, as $N\uparrow\infty$


size of largest c-comp (mean)
0
20
40
60
80
100
0.00
0.01
0.02
0.03
0.04
0.05
probability of an edge


prob of giant comp (mean)
0.0
0.2
0.4
0.6
0.8
1.0
0.00
0.01
0.02
0.03
0.04
0.05
probability of an edge


size of largest c-comp (st.dev)
0
5
10
15
0.00
0.01
0.02
0.03
0.04
0.05
probability of an edge


prob of giant comp (st.dev)
0.0
0.1
0.2
0.3
0.4
0.5
0.00
0.01
0.02
0.03
0.04
0.05
probability of an edge

p* or ERG models

$$\Pr(Y=y|\Theta=\theta) = \exp\{ \Sigma_k \theta_k S_k(y) + A(\theta) \}$$

where $S_k(y)$ counts specific structure k, such as

- edges $S_1(y) = \Sigma_{1\le i\le j\le n} y_{ij}$
- triangles $S_3(y) = \Sigma_{1\le i\le j\le h\le n} y_{ij}\, y_{ih}\, y_{jh}$.

Frank & Strauss (JASA, 1986), Snijders et al. (Soc. Met., 2004), Hanneke & Xing (LNCS, 2007)

Towards exchangeable graphs

- Symmetry suggests the nodes should be treated as exchangeable in the following sense

$$\Pr(\{y_{i,j} : 1 \leq i < j \leq n\} \in A) = \Pr(\{y_{\pi i,\pi j} : 1 \leq i < j \leq n\} \in A)$$

- A result by Hoover and Aldous: any model that satisfies this condition for any N is of the form

$$y_{i,j} = h(\mu, u_i, u_j, \epsilon_{i,j})$$

for u_i,u_j i.i.d. and ε_{ij} i.i.d node/pair-specific effects

Exchangeable graph models

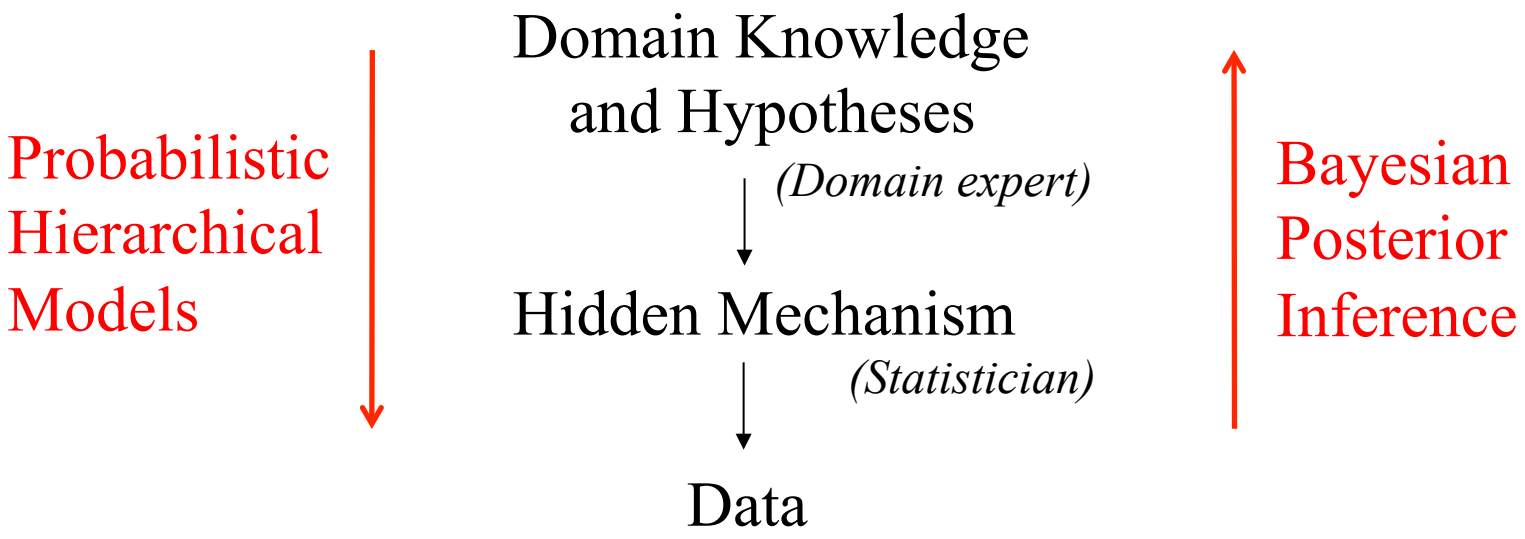
- Alternative specifications of $h(\mu,u_i,u_j,\varepsilon_{ij})$ lead to different models. With some generality

$$P(Y_{ij}=1|\mu,u_i,u_j,\varepsilon_{ij}) = h'(\mu + \alpha(u_i,u_j) + \varepsilon_{ij}) = \theta_{ij}$$

- Likelihood

$$P(Y|c) = \int_{\Theta} P(\Theta|c) \cdot \Pi_{ij} \theta_{ij}^{Yij} (1-\theta_{ij})^{(1-Yij)} d\theta_{ij}$$

Approach



- Issues: scalability, global vs. local perspectives

Three basic models

- Latent space model

  $\alpha(u_i,u_j) = -|u_i-u_j|$; u_i real vectors, for i=1…N

- Latent eigenmodel

  $\alpha(u_i,u_j) = u_i'\Lambda u_j$; u_i real vectors, for i=1…N; Λ diag. K×K

- Latent class model

  $\alpha(u_i,u_j) = B_{ui,uj}$; u_i =1…K, for i=1…N; B symm. K×K

Latent space models

log-odds $(Y_{ij}=1|u_i,u_j,\mu) = \mu - |u_i - u_j| = \eta_{ij}$
where u_i is a point in $\mathbb{R}^k$, for all nodes i in N.

Idea: close points in $\mathbb{R}^k$ are likely to be connected.

Here u_is are constants; $\theta_{ij} = [1+\exp\{-\eta_{ij}\}]^{-1}$ and
likelihood is $P(Y|U,\mu) = \Sigma_{ij} [\eta_{ij} Y_{ij} - \log(1+\exp\{\eta_{ij}\})]$

Hoff et at. (JASA, 2002), Handcock et al. (JRSS/A, 2007), Krivitsky et al. (Soc. Net., 2009)

Shortcomings so far

- ERG models *(Wasserman et al., Handcock et al.)*

  Summarize graphs using exp model on motif-counts

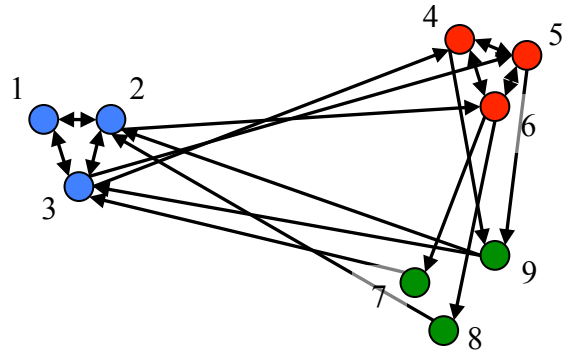
  Issues: cannot offer node-specific predictions, ..

- Latent space models *(Hoff et al. 02; Hoff 03)*

  Project adjacency matrix onto a latent $\mathbb{R}^K$ via logistic regression; closer points increase chance of connectivity

  Issues: MCMC does not scale, hard identifiability problem, no clustering effect

Model specifications



$\pi_i \sim$ Dirichlet (α), for all nodes i=1..N

$y_{ij}|\pi_i,\pi_j \sim$ Bernoulli (π_i`B π_j), for all pairs (i,j)

where π_i is a point in the K-simplex, and B is K×K.

Nodes in the same block share similar connectivity.

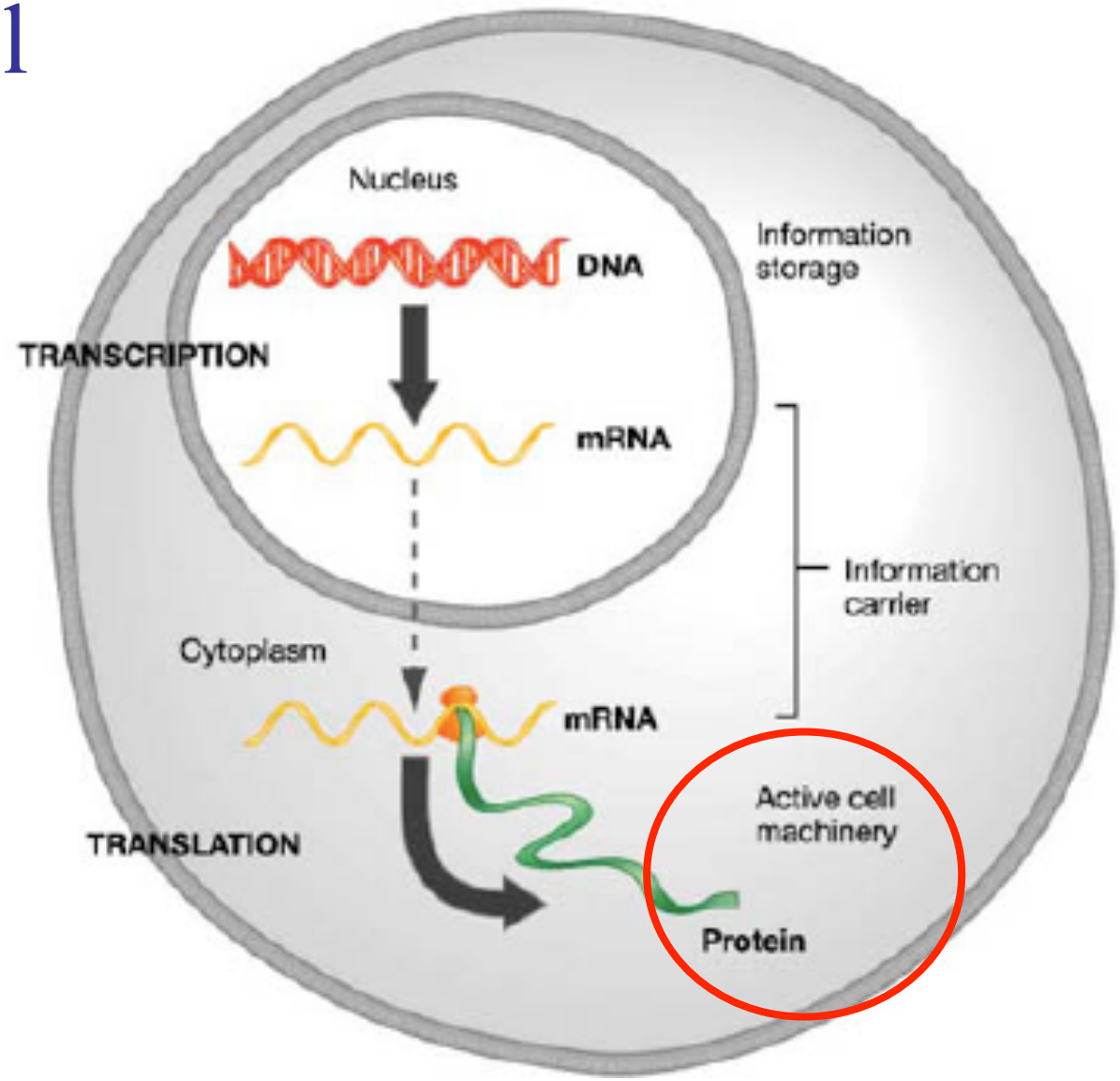
Loraine & White (JMS, 1971), Fienberg et al. (JASA, 1985), Nowicki & Snijders (JASA, 2001), Airoldi et al. (JMLR, 2008)

Agenda

- Overview
- Models of networks
- Mixed membership blockmodels
  1. Inference
  2. Results
  3. Remarks
- Concluding remarks

(Source: fig.cox.miami.edu)

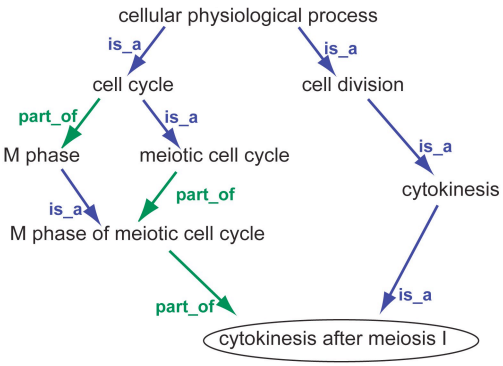
The cell

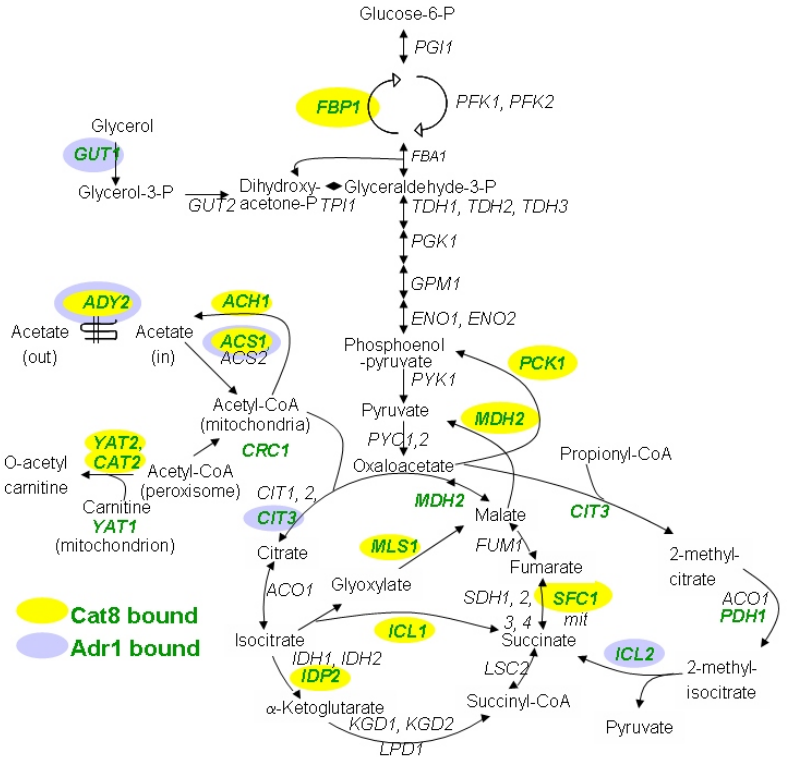


(Source: SGD and own work)

Functions & mechanisms

- Cytoplasm is a busy place
  Proteins, small molecules
- Taxonomy of functions
  Gene Ontology annotations
  (e.g., cell division)
- Mechanisms
  Pathways as complex graphs
  (e.g., carbon metabolism)

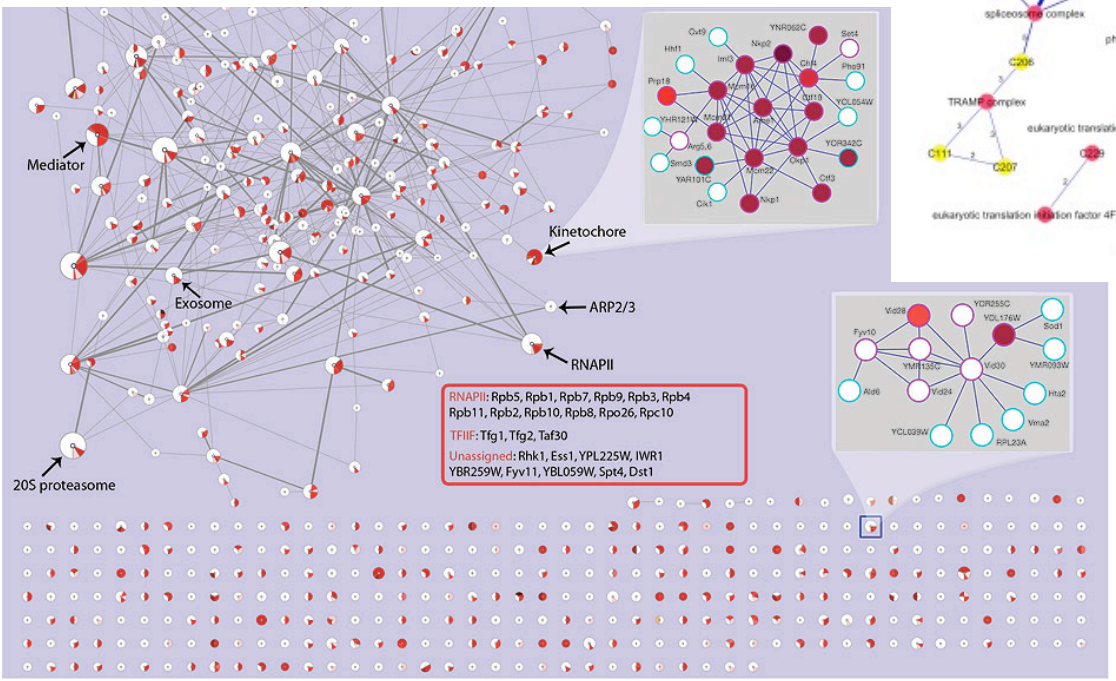


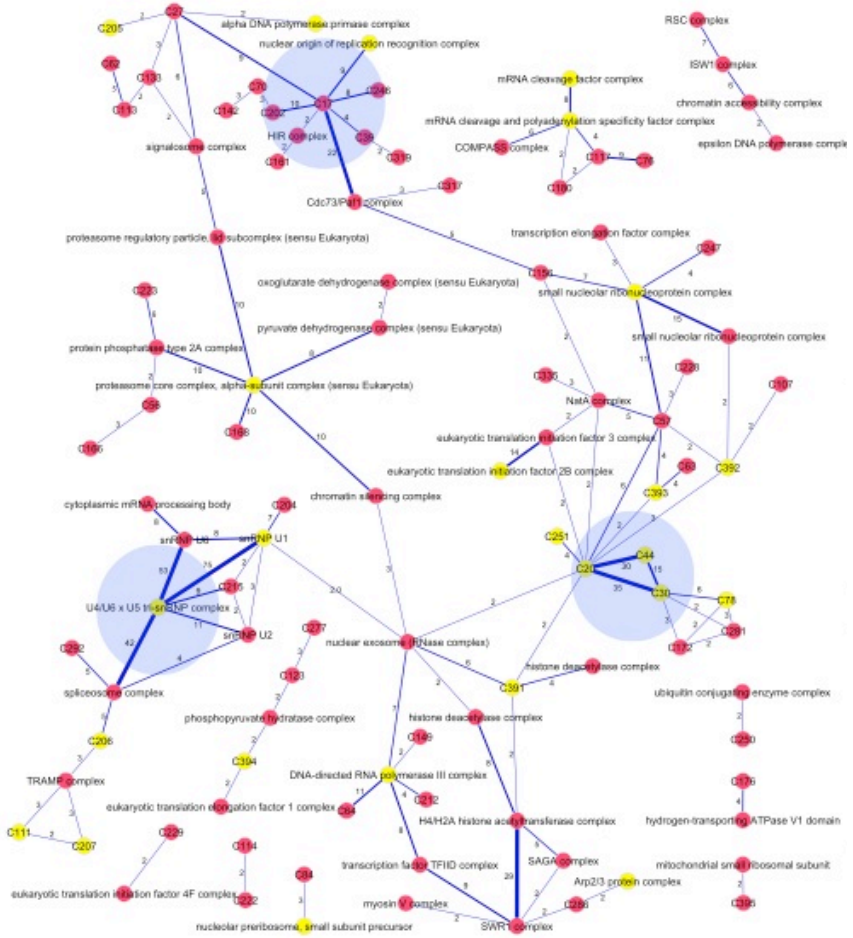


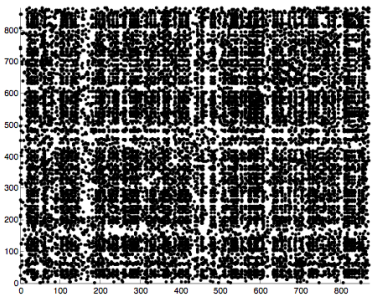
(Source: Nature, and BMC Bioinformatics)

Domain knowledge

Proteins form stable protein complexes to carry out functions in the cell



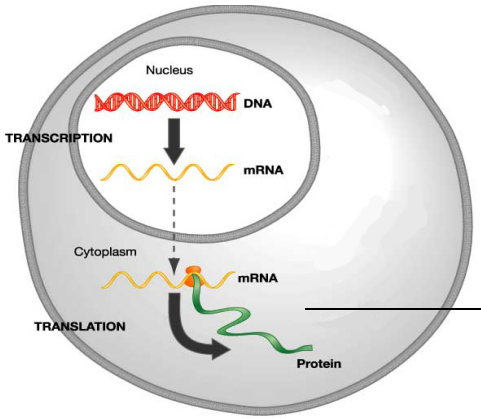


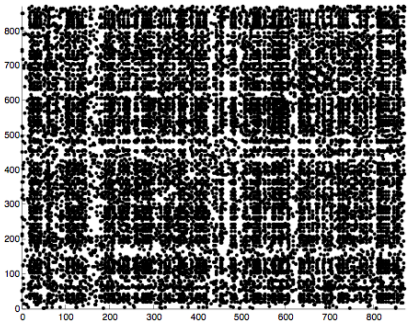


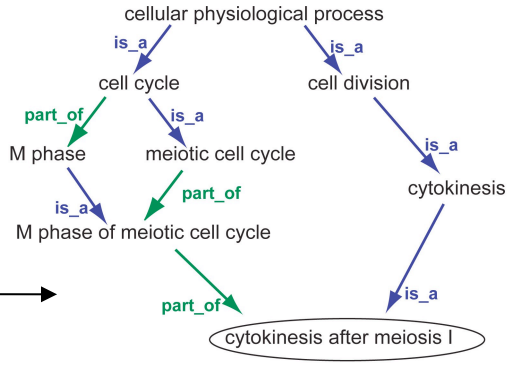
Protein interaction data

(Source: fig.cox.miami.edu, SGD, and own work)

Scientific questions







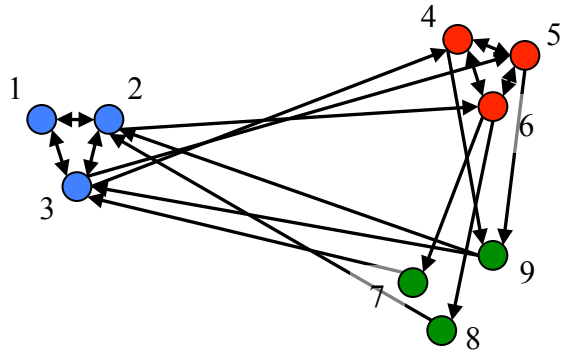
Yeast cell

Protein interaction data

Functions (GO Slim)

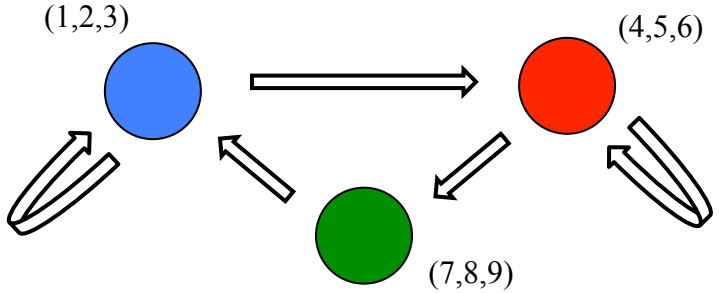
- Can interaction motifs:
  - indicate proteins' multifaceted functional role?
  - reveal protein complexes and relations among them?

Two modeling ideas



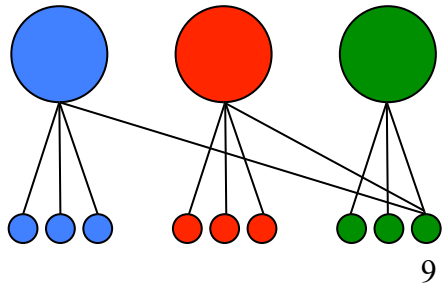
- Structural equivalence *(Lorrain & White, 1971)*
  - Nodes with similar connectivity collapsed into a block
- Instantiated by
  - Blockmodel *(B)*

  (≈ Nowiki & Snijders, 01, Airoldi et al. 05, 07, 08)

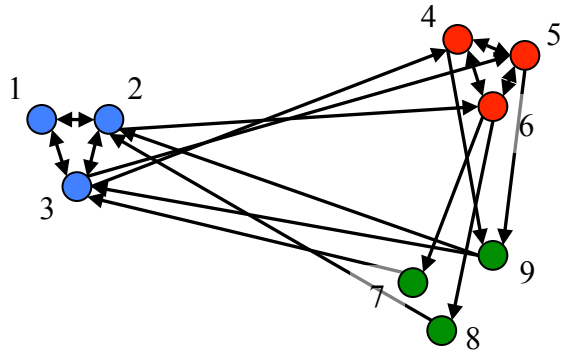


- Combined with
  - Mixed membership *(Π)*

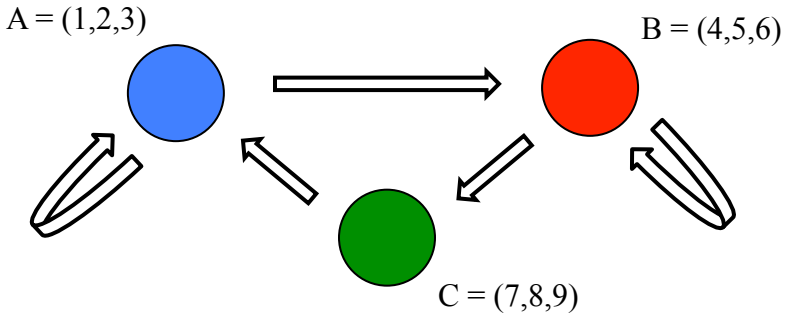
  (Airoldi et al. 05, 07, 08)



Blockmodel, *B*

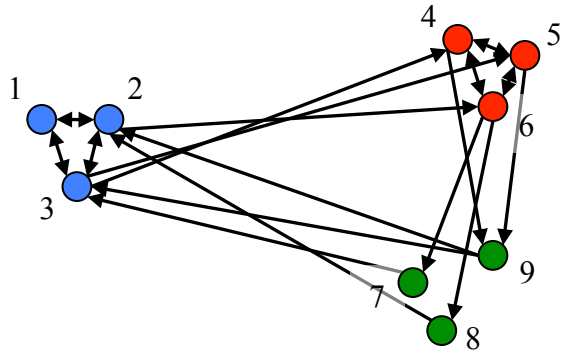


- Captures salient structure at the block level
- Connectivity among nodes within the same block (across blocks) is only specified on average

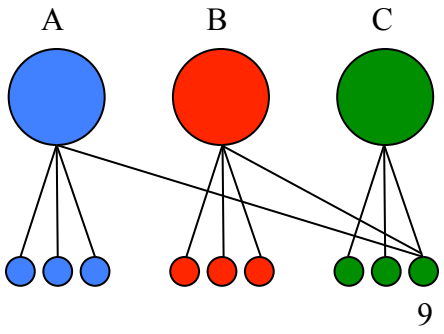


	From			
A	B	C		
1.0	0	0.3	A	To
0.3	1.0	0	B	
0	0.3	0	C	

Mixed membership, Π



- Nodes can be mapped to multiple blocks
- Extends the idea of a mixture (i.e., local weights)
- Node-specific weights useful for prediction

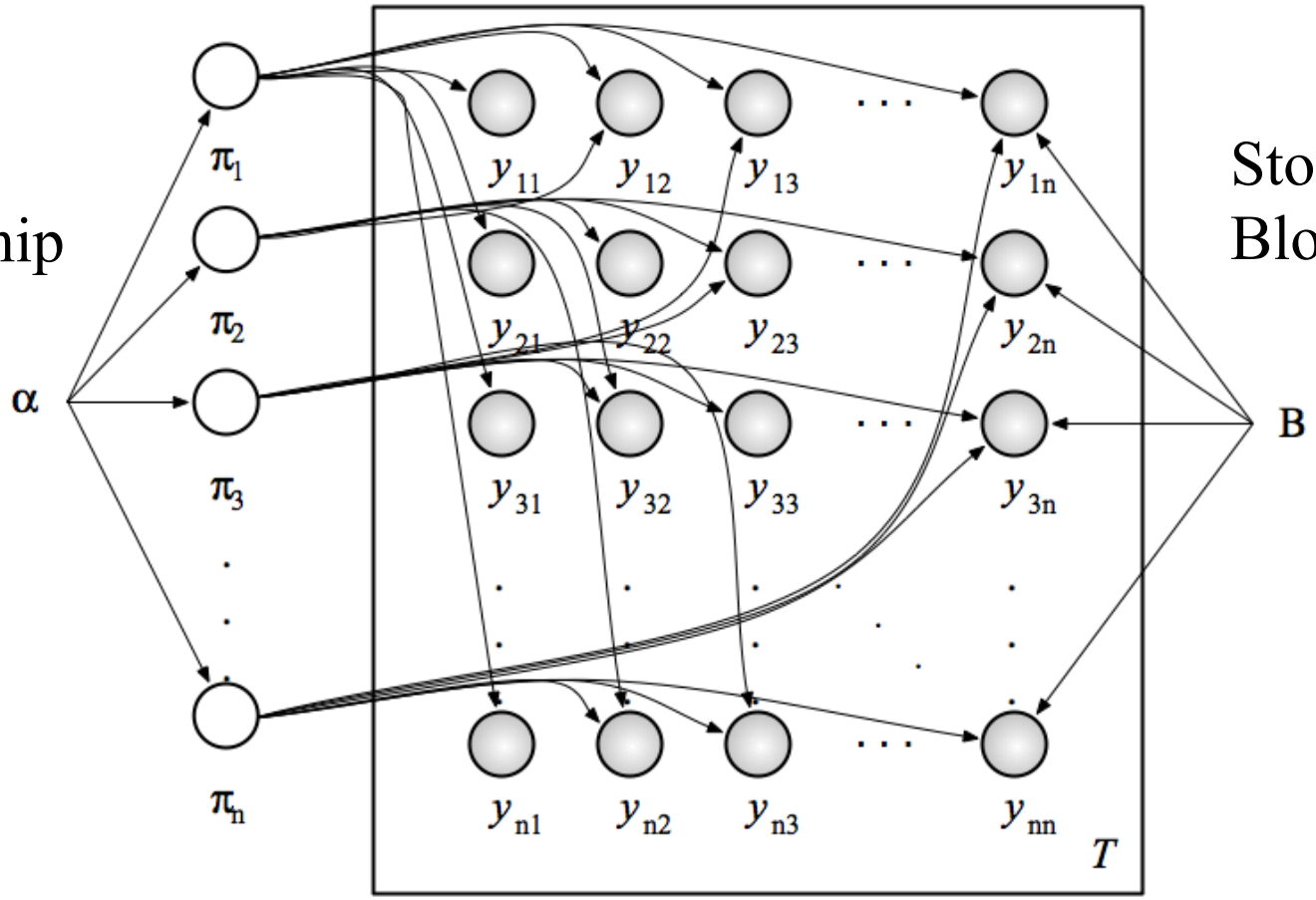


A	B	C	node
1.0	0	0	1
1.0	0	0	2
	. .		. .
0.1	0.1	0.8	9

Model: projecting *Y* onto *B* via *Π*

Mixed Membership

Stochastic Blockmodel



Model: variant for prediction

Blockmodel + node-specific memberships

$$\vec{\pi}_n \sim Dirichlet\ (\alpha), \quad n \in [1, N]$$

$$Y(n,m) \sim Bernoulli\ (\vec{\pi}'_n B\ \vec{\pi}_m), \quad (n,m) \in [1,N]^2$$

Likelihood

$$\ell(Y|\alpha, B) = \int_\Pi \prod_n p(\vec{\pi}_n|\alpha) \prod_{nm} p(Y(n,m)|\vec{\pi}_n, \vec{\pi}_m, B)\ d\Pi$$

Note: the matrix B has size K×K

Model: variant for de-noising

Blockmodel + relation-specific memberships

$$\vec{\pi}_n \sim Dirichlet\ (\alpha), \quad n \in [1, N]$$

$$\vec{z}_{nm\rightarrow} \sim multinomial\ (\vec{\pi}_n, 1), \quad (n, m) \in [1, N]^2$$

$$\vec{z}_{nm\downarrow} \sim multinomial\ (\vec{\pi}_m, 1), \quad (n, m) \in [1, N]^2$$

$$Y(n, m) \sim Bernoulli\ (\vec{z}'_{nm\rightarrow} B\ \vec{z}_{nm\downarrow}), \quad (n, m) \in [1, N]^2$$

Note: the matrix B has size K×K

Agenda

- Overview
- Models of networks
- Mixed membership blockmodels
  1. Inference
  2. Results
- Concluding remarks

Revisiting EM

- Data Y, latent variables $X = (\Pi, Z)$, and constants $\Theta = (\alpha, B)$

$$
\begin{aligned}
\log p(Y|\Theta) &= \log \int_{\mathcal{X}} p(Y, X|\Theta)\, dX \\
&= \log \int_{\mathcal{X}} q(X)\, \frac{p(Y, X|\Theta)}{q(X)}\, dX \qquad \text{(for any } q\text{)} \\
&\geq \int_{\mathcal{X}} q(X)\, \log \frac{p(Y, X|\Theta)}{q(X)}\, dX \qquad \text{(Jensen's)} \\
&= \mathbb{E}_q \left[\log p(Y, X|\Theta) - \log q(X) \right] \;=:\; \mathcal{L}(q, \Theta)
\end{aligned}
$$

Variational EM

- EM maximizes the lower bound over *(q,Θ)*
- In EM we set

$$q = p(X \mid Y, \Theta)$$

- If not feasible, we can posit approximation for *q* using free parameters *Δ* — this is vEM

$$q \approx q_{\Delta}(X) \;\rightarrow\; p(X \mid Y) \text{ at } \Delta^* = \Delta^*(Y)$$

Variational EM (cont.)

- Leads to approximate lower bound

$$\mathbb{E}_{q_\Delta}\left[\log p(Y, X \mid \Theta) - \log q_\Delta(X)\right] \quad =: \quad \mathcal{L}(q_\Delta, \Theta)$$

- Iterate

Variational E-step: $\Delta^* = \arg\max_\Delta \ \mathcal{L}(q_\Delta, \Theta)$

M-step: $\Theta^* = \arg\max_\Theta \ \mathcal{L}(q_{\Delta^*}, \Theta)$

Nested variational EM

- Mean field: $q_{\Delta}(\Pi, Z) = \prod_n q_{\vec{\gamma}_n}(\vec{\pi}_n) \cdot \prod_{nm} q_{\vec{\phi}_{nm}}(\vec{z}_{nm})$

Vanilla vEM *(Jordan et al. 99)*

E-step:

initialize $\boldsymbol{\gamma}_{1:N}$, $\boldsymbol{\phi}_{1:N,1:N}$

1. update $\boldsymbol{\phi}_{1:N,1:N}$
2. update $\boldsymbol{\gamma}_{1:N}$

M-step:

update α, B

Nested vEM *(Airoldi et al. 05, 08)*

E-step:

initialize $\boldsymbol{\gamma}_{1:N}$

loop pairs *(n,m)*

1. init & optimize $\boldsymbol{\phi}_{n,m}$
2. partially update $\boldsymbol{\gamma}_n, \boldsymbol{\gamma}_m$

M-step:

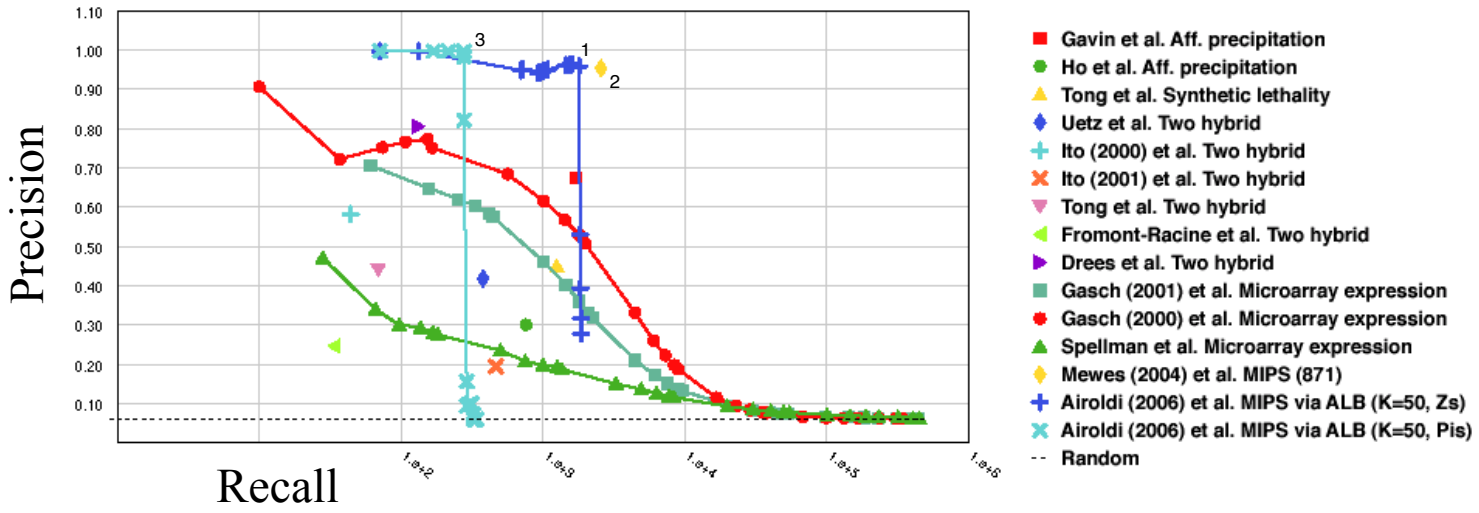
update α, B

Agenda

- Overview
- Models of networks
- Mixed membership blockmodels
  1. Inference
  2. Results
- Concluding remarks

Evaluation: recovering function

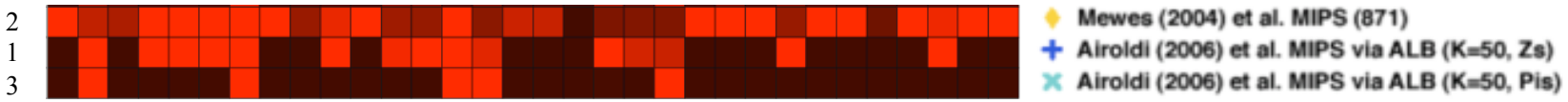
- Functional content in $\mathbb{P}(\mathbf{Y} \mid \hat{\Theta})$



- Model reveals information about functional modules
  (cross-validation: K=50; gold standard in Myers et al. 06)*

Evaluation: identifying blocks

- Two model variants capture a different number of functional processes, with equally high accuracy

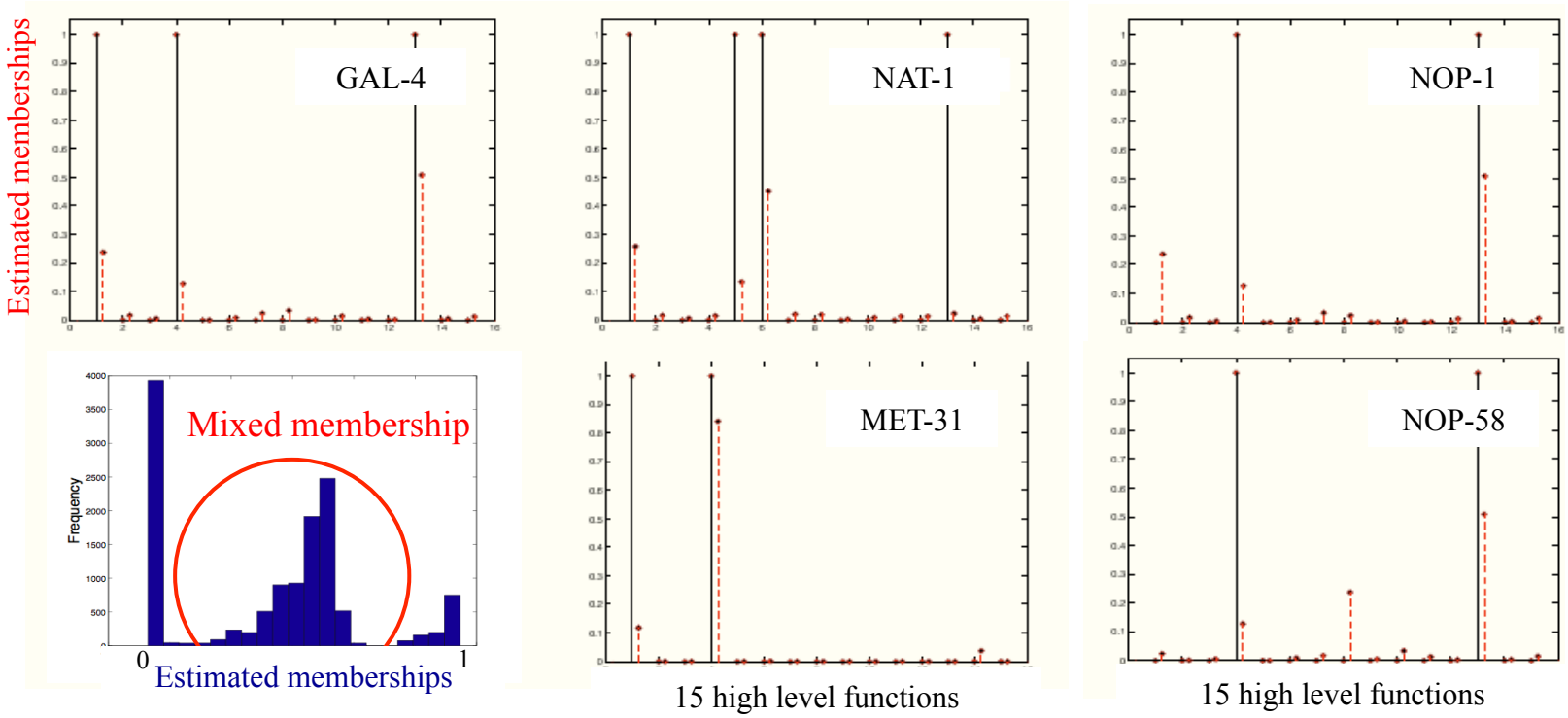


GO functional processes *(Area under the curve, red = high)*

Description	Pred.
Biopolymer catabolism	561
Transcription from RNA polymerase II promoter	341
Protein biosynthesis	281
DNA replication	196
Protein complex assembly	191
Chromatin modification	172
Protein amino acid acetylation	91
Transcription from RNA polymerase I promoter	78
Homeostasis	78

Evaluation: mixed membership

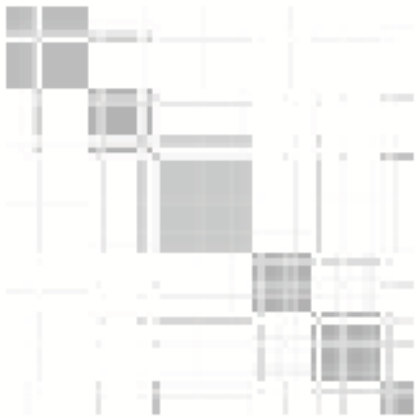
- Amount of mixed membership is substantial
- Membership reveals multifaceted functional roles



National study on adolescents

- A friendship network among 69 students in grades 7-12

Original data



node-specific
(prediction)

relation-specific
(de-noising)

	MMSB Clusters						MSB Clusters						LSCM Clusters					
Grade	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
7	13	1	0	0	0	0	13	1	0	0	0	0	13	1	0	0	0	0
8	0	9	2	0	0	1	0	10	2	0	0	0	0	11	1	0	0	0
9	0	0	16	0	0	0	0	0	10	0	0	6	0	0	7	6	3	0
10	0	0	0	10	0	0	0	0	0	10	0	0	0	0	0	0	3	7
11	0	0	1	0	11	1	0	0	1	0	11	1	0	0	0	0	3	10
12	0	0	0	0	0	4	0	0	0	0	0	4	0	0	0	0	0	4

Table 1: Grade levels versus (highest) expected posterior membership for the 69 students, according to three alternative models. MMSB is the proposed mixed membership stochastic blockmodel, MSB is a simpler stochastic block mixture model (Doreian et al., 2007), and LSCM is the latent space cluster model (Handcock et al., 2007).

$$\hat{B} = \begin{bmatrix} 0.3235 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.3614 & 0.0002 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.2607 & 0.0 & 0.0 & 0.0002 \\ 0.0 & 0.0 & 0.0 & 0.3751 & 0.0009 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0002 & 0.3795 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.3719 \end{bmatrix}$$

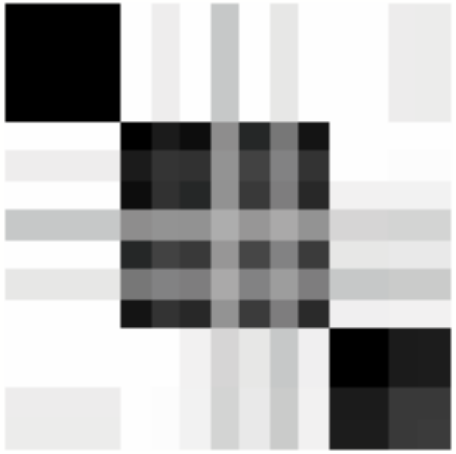
Sampson's monastery data

- Multivariate sociometric relations among novices in a NE monastery, over two years.
- Anthropological observations as ground truth
- Two factions, plus social outcasts and waverers
- After two years John and Greg get expelled, most young turks leave and the order dissolves

Expressing connectivity

- Two variants provide increasing levels of definition

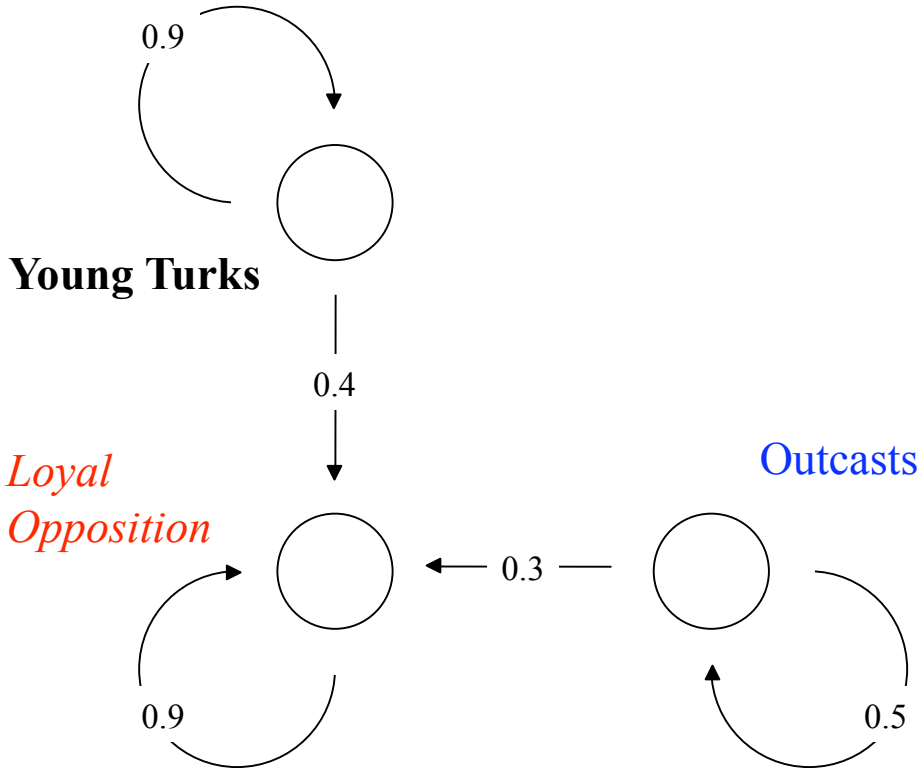
Original data



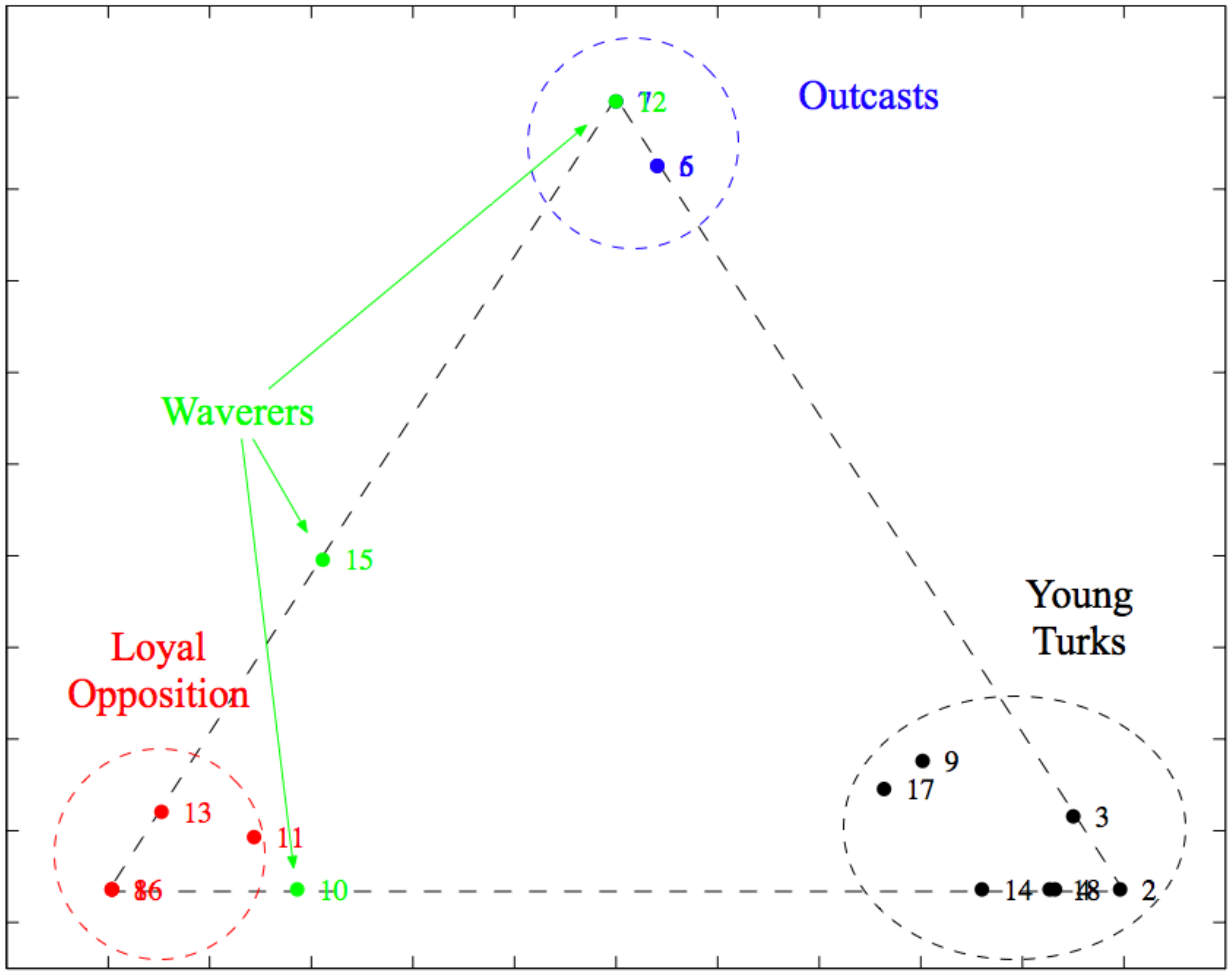
node-specific
(prediction)

relation-specific
(de-noising)

Social structure: blockmodel

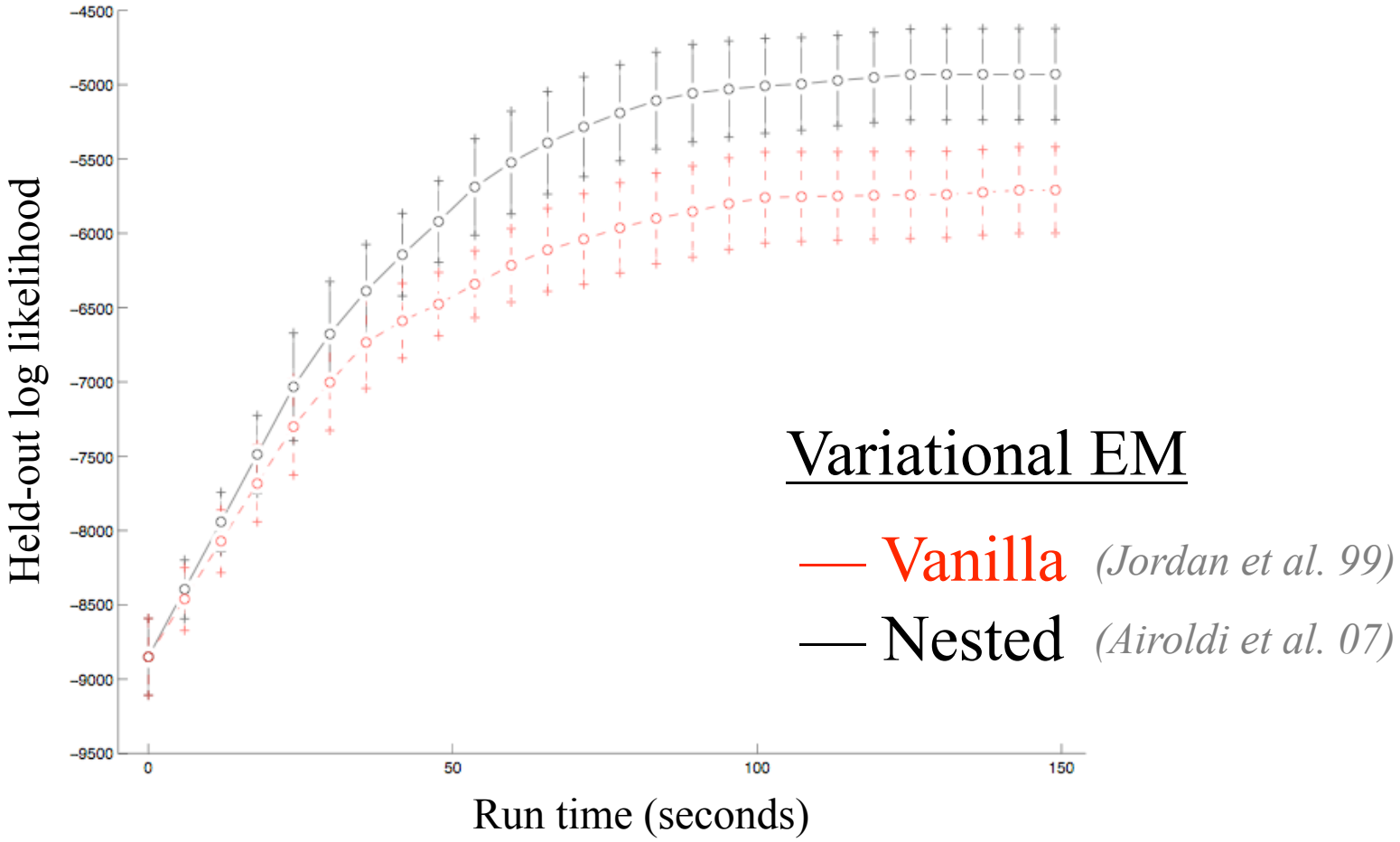


Social structure: membership



1	Ambrose
2	Boniface
3	Mark
4	Winfrid
5	Elias
6	Basil
7	Simplicius
8	Berthold
9	John Bosco
10	Victor
11	Bonaventure
12	Amand
13	Louis
14	Albert
15	Ramuald
16	Peter
17	Gregory
18	Hugh

Evaluation: nested variational EM



(Simulated data; 300 nodes, 10 blocks)

Model extensions

- Sparsity, general formulation, informative priors and full Bayes *(Airoldi, Blei, Fienberg & Xing, 05, 06, 08)*
- Node attributes *(Airoldi, Markowetz, Blei & Troyanskaya)*
- Dynamic *(Airoldi, Fienberg & Krackhardt, 08)*
- Extensions by others *(Hofman & Wiggins 07; Eliassi-Rad, Griffiths & Jordan; Nallapati, Cohen & Lafferty; Frey et al., 06, Chang & Blei)*

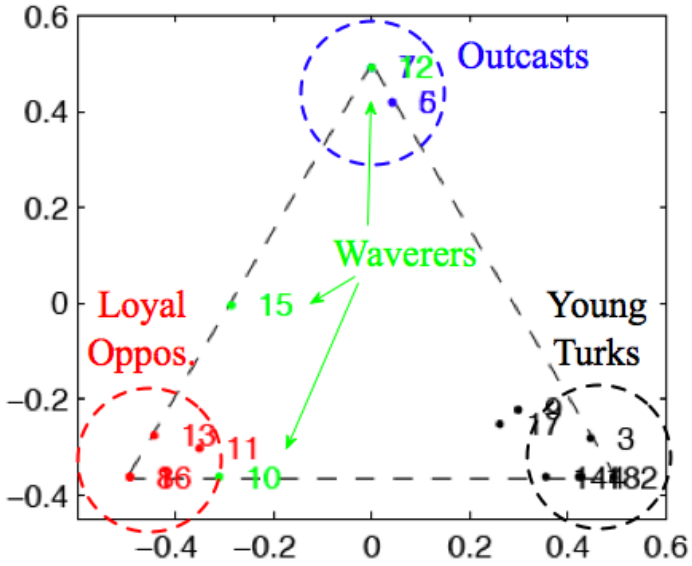
Dynamics of social failure

- Analysis suggests a theory of social failure in isolated communities. Try longitudinal model

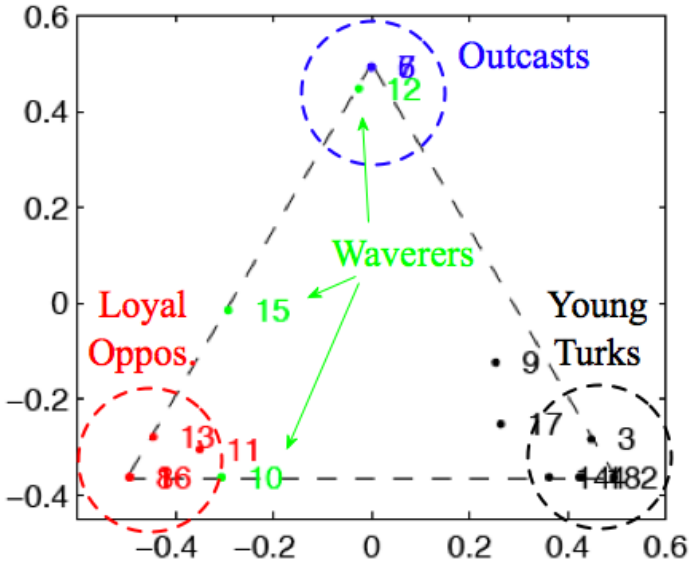
$$
\begin{aligned}
P\left(\vec{\pi}_0(n) \mid \Theta\right) &\sim \mathbf{f} \circ \mathit{Gaussian}\left(\vec{0}, A\right), \\
P\left(\vec{\pi}_t(n) \mid \vec{\pi}_{t-1}(n), \Theta\right) &\sim \mathbf{f} \circ \left[\, \mathit{Gaussian}\left(\vec{0}, A\right) + \mathbf{f}^{-1} \circ \vec{\pi}_{t-1}(n) \,\right], \\
P\left(\, Y_t(n,m) \mid \Pi_t, \Theta\right) &\sim \text{Bernoulli}\left(\vec{\pi}_t(n)' B\, \vec{\pi}_t(m)\right),
\end{aligned}
$$

- Data: $Y_t(n,m)$ s.t. $n, m = 1, \ldots, N = 18$ and $t = 1, 2, 3.$

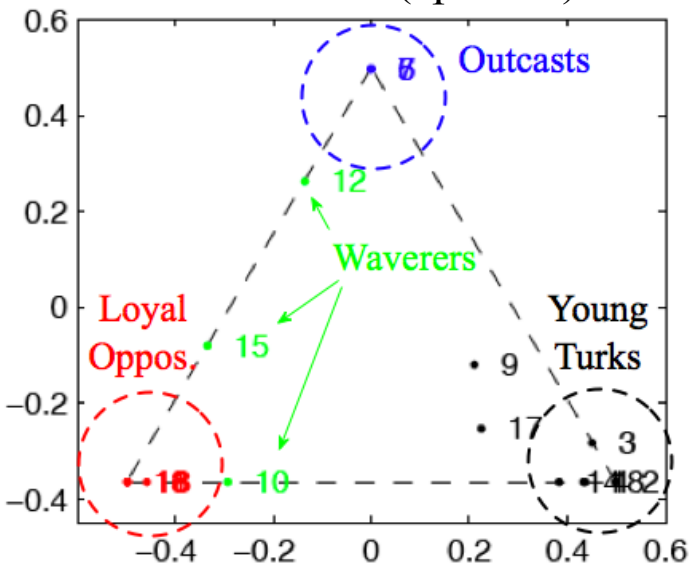
Whom do like (epoch 1)



Whom do like (epoch 2)



Whom do like (epoch 3)



1	Ambrose	(9)
2	Boniface	(15)
3	Mark	(7)
4	Winfrid	(12)
5	Elias	(17)
6	Basil	(3)
7	Simplicius	(18)
8	Berthold	(6)
9	John Bosco	(1)
10	Victor	(8)
11	Bonaventure	(5)
12	Amand	(13)
13	Louis	(11)
14	Albert	(16)
15	Ramuald	(10)
16	Peter	(4)
17	Gregory	(2)
18	Hugh	(14)

Agenda

- Overview
- Models of networks
- Mixed membership blockmodels
- Concluding remarks

Take home points

- Complex networks are an exciting research area that is generating new statistical problems
- The familiar notions of sampling variability and sampling designs are challenged
- Potential for impact in the sciences, from biology to communications, and from computational social science to healthcare survey design and analysis

Acknowledgements and pointers

CDC, Facebook, Bell Labs. S Fienberg, E Xing, D Blei, B Singer, A Gelman, Z Ghahramani, J Leskovec, J Kleinberg, D Rubin.

1. Getting started in probabilistic graphical models. Airoldi, *PLoS Computational Biology*, 2007.
2. Mixed membership stochastic blockmodels. Airoldi, Blei, Fienberg & Xing, *Journal of Machine Learning Research*, 2008. (in R: iGraph, LDA)
3. A survey of statistical network models. Goldenberg, Zheng, Fienberg & Airoldi. *Foundations & Trends in Machine Learning*, 2009.
4. Deconvolution of mixing time series on a graph. Blocker & Airoldi. *Uncertainty in Artificial Intelligence (UAI)*, 2011.