Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Dynamic Network Analysis

model, algorithm, theory and applications

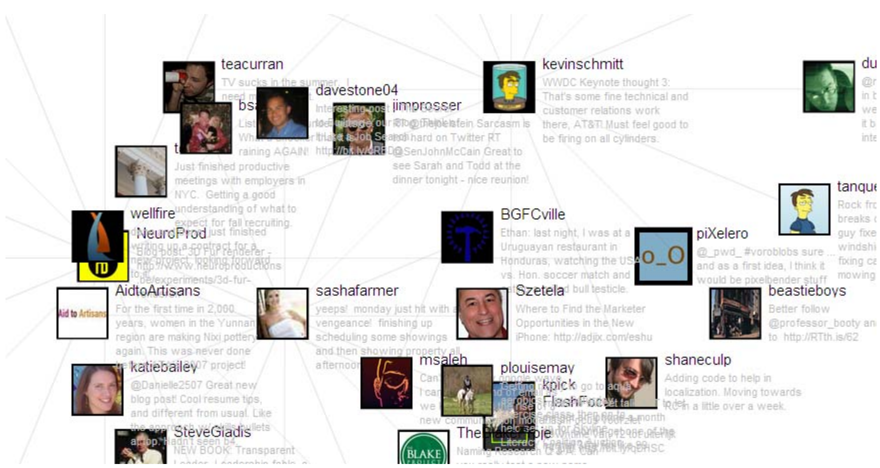
Eric Xing

epxing@cs.cmu.edu

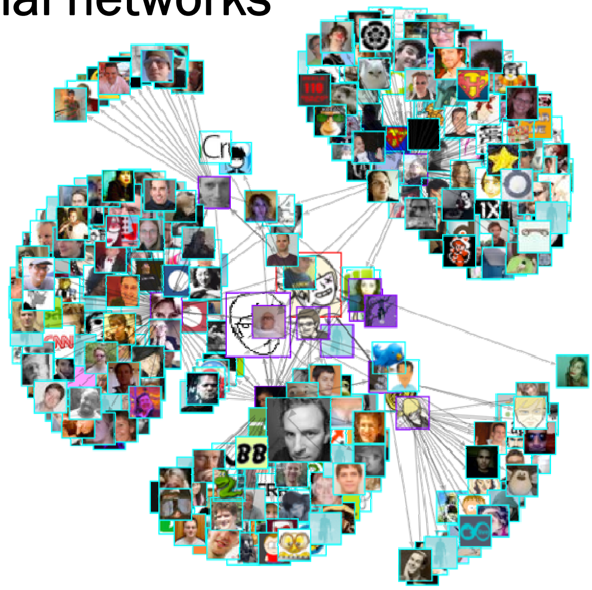
Machine Learning Dept./Language Technology Inst./Computer Science Dept.
Carnegie Mellon University

Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

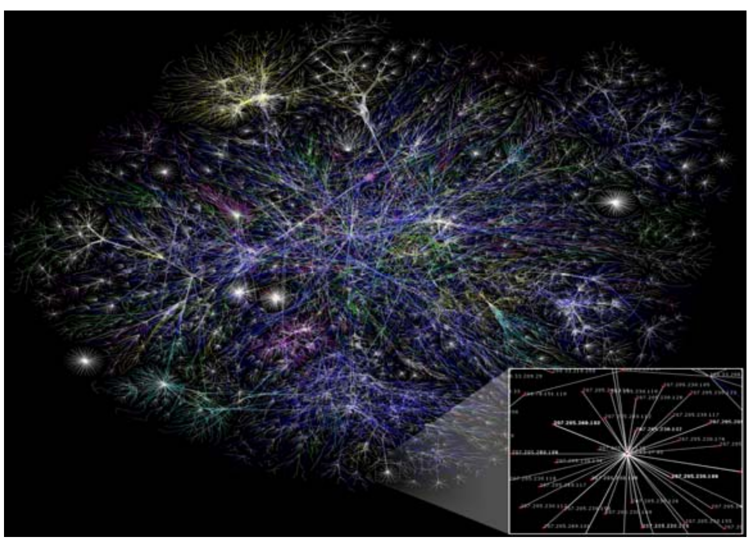
Our experience with networks



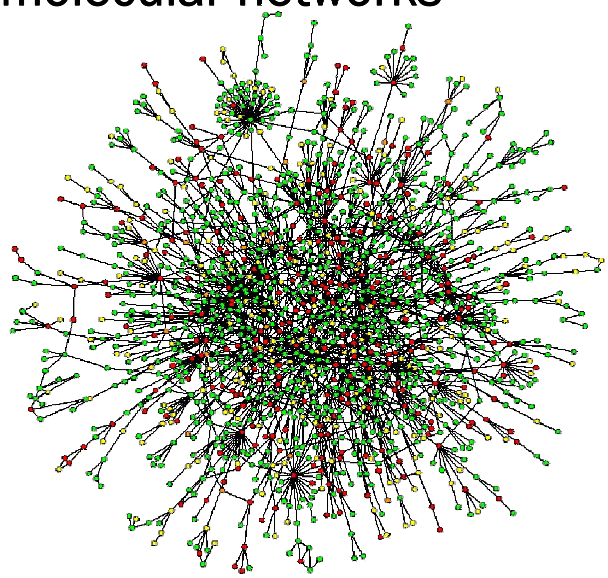
Social networks



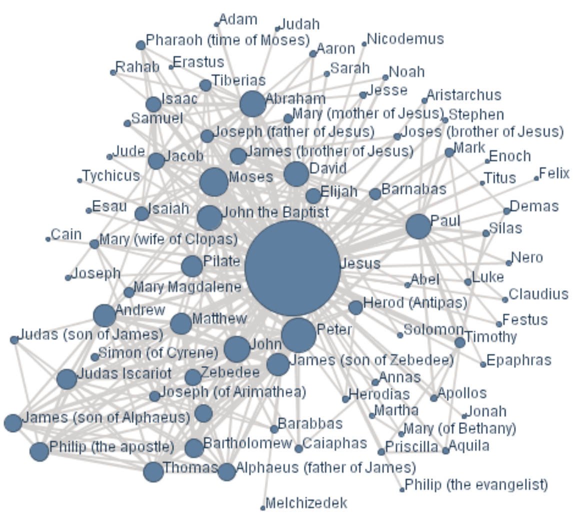
The Internet



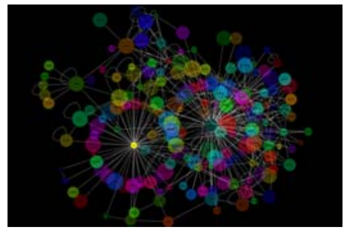
Bio-molecular networks

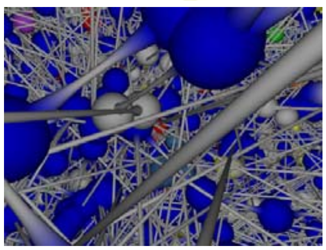


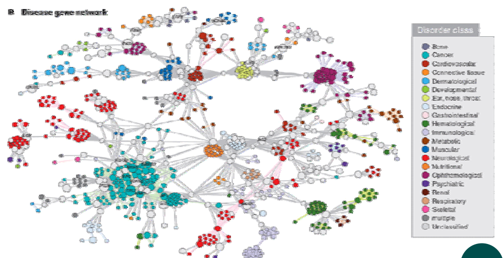
Jesus network



Network analysis -- visualization

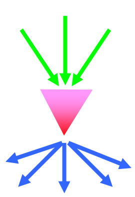




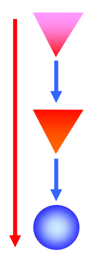


Global topological measures

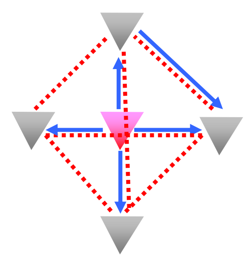
- Indicate the gross topological structure of the network



Connectivity
(Degree)



Path length

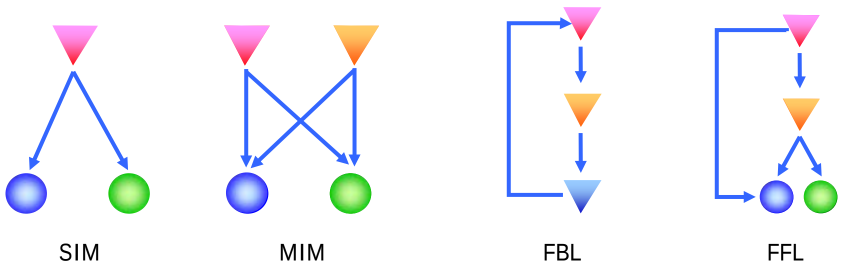


Clustering coefficient

[Barabasi]

Local network motifs profiles

- Regulatory modules within the network

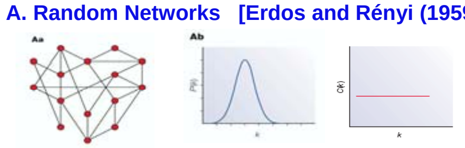


6/29/2010 9

[Alon]

Models for Global Network Analysis

A. Random Networks [Erdos and Rényi (1959, 1960)]



$$P(k) = \frac{e^{-\bar{k}}\,\bar{k}^{k}}{k!}$$

Mean path length ~ ln(k)

Phase transition:

Connected if: $p \geq \ln(k)/k$

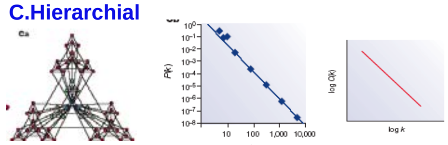
B. Scale Free [Price,1965 & Barabasi,1999]

$P(k) \sim k^{-\gamma},\ k >> 1,\ 2 < \gamma$

Mean path length ~ lnln(k)

Preferential attachment. Add proportionally to connectedness

C.Hierarchial



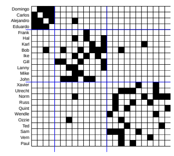
Copy smaller graphs and let them keep their connections.

6/29/2010 10

Models for Meta Analysis of Networks

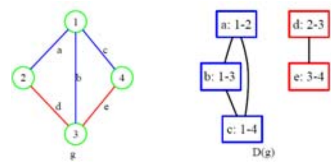
A. Stochastic Block Models [Holland, Laskey, and Leinhardt 1983]

$$\Pr(\mathbf{X}=\mathbf{x}) = \frac{1}{\kappa}\exp\left(\sum \lambda_{ij} x_{ij}\right)$$



B. Exponential Radom Graph Models [Bahadur 1961, Besag 1974, Frank & Strauss 1986]

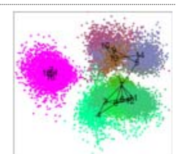
$$\Pr(\mathbf{X}=\mathbf{x}) = \frac{\exp\{\theta' \mathbf{z}(\mathbf{x})\}}{c}$$



C. Latent Space Models [Hoff, Raftery, and Handcock, 2002]

$$z_i \sim \sum_{g=1}^{G} \lambda_g \, \mathrm{MVN}_d(\mu_g, \sigma_g^2 I_d)$$

$$\text{log-odds}(y_{i,j}=1 \mid z_i, z_j, x_{i,j}, \beta) = \beta_0^{\mathrm{T}} x_{i,j} - \beta_1 |z_i - z_j|$$



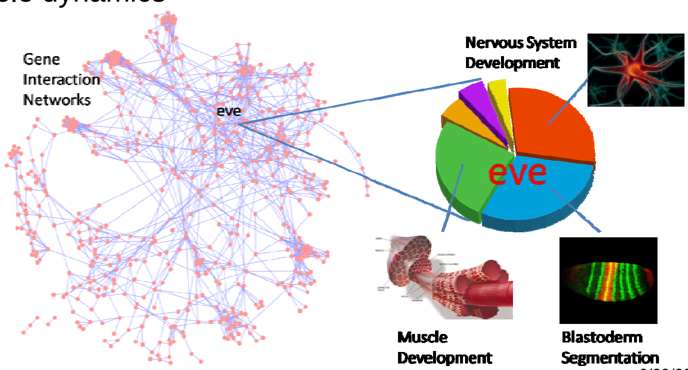
Summary

- Impressive graphs!
- Seemingly impressive statements about nwk properties

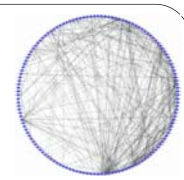
  --- scale free, small world, …
- Can even fit data with some models
- But … What about details?
  - Can we **infer** …
    - **e.g., the role of every node, the meaning of every edge?**
    - **the network topology itself?**
  - Can we **predict** …
  - Can we **simulate** …
- Most current analyses tell **BIG** stories, or obvious stories, but not so useful to serious detail-hunters

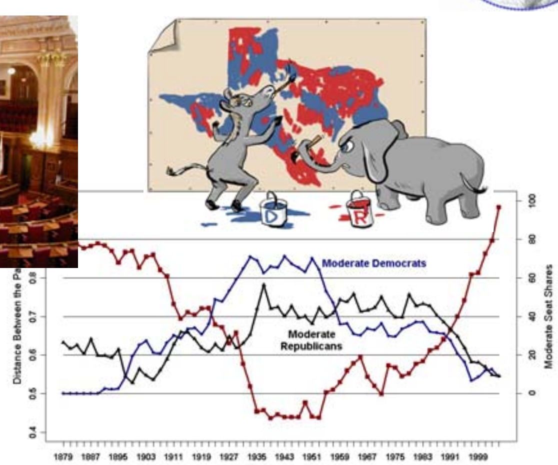
I: Network tomography

- Micro-inference vs. Meso- or Macro-inference
- Multi-role of every node
- Context dependent role-instantiation
- Role dynamics



Evolving Social Networks





Corporativity,

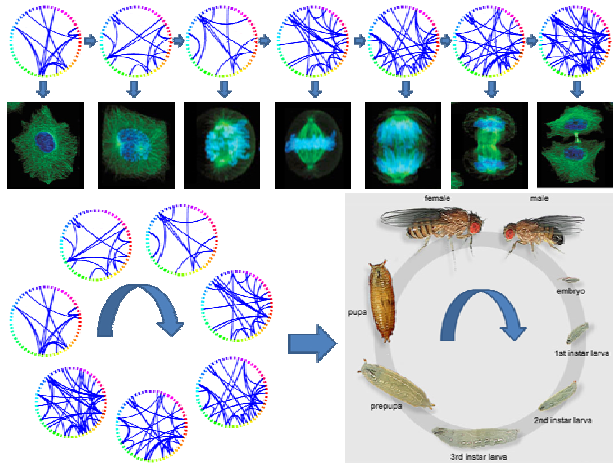
Antagonism,

Cliques,

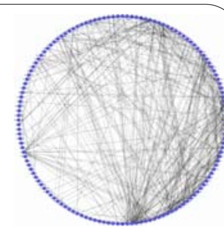
…

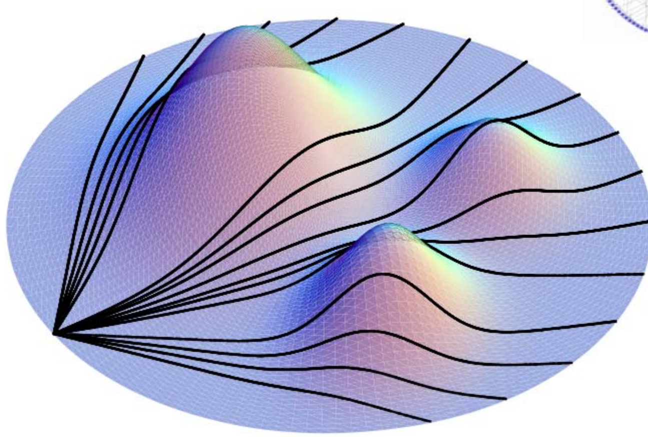
over time?

Evolving Gene Networks



II: Travel time tomography



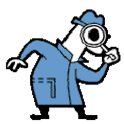


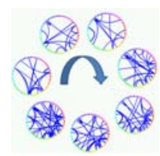
III: Where do networks come from?

- Existing work:
  - Assuming ***iid*** sample, and **a single** ***static*** network

$$\mathbf{x}^t \sim P_\theta(\mathbf{X}; G) \qquad t = 1, \ldots, T$$

  - Assuming networks or network time series are observable and given



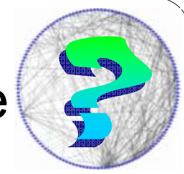


  - Then model/analyze the generative and/or dynamic mechanisms

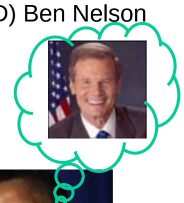
This Talk:

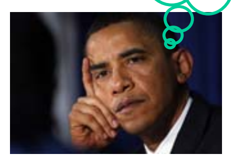
- Dynamic Network Model
  - Temporal exponential random graph model (tERGM)
    [Hanneke and Xing, **ICML 06**; Fan, Hanneke, Fu and Xing, **ICML 07**; Hanneke, Fu, and Xing, **EJS 10**]
- Reverse Engineer Latent Time-Evolving Networks
  - The TESLA and KELLER algorithms, and beyond
    [Fan, Hanneke, Fu, and Xing, **ICML 07**; Ahmed and Xing, **PNAS 09**; Song, Mladen and Xing, **ISMB 09**; Mladen, Song, Ahmed and Xing, **AOAS 09**; Mladen, Song, and Xing, **NIPS 09**; Song, Mladen and Xing, **NIPS 09**]
- Network tomography: modeling latent "multi-role" of vertices
  - Mixed Membership Stochastic Blockmodel (MMSB)
    [Airoldi, Blei, Xing and Fienberg, **LinkKDD 05**; Airoldi, Blei, Fienberg and Xing, **JMLR 08**]
- Dynamic tomography underlying evolving network
  - dynamic MMSBs
    [Xing, Fu, and Song, **AOAS 09**; Ho, Le, and Xing, **submitted 10**]

Political network in the Senate

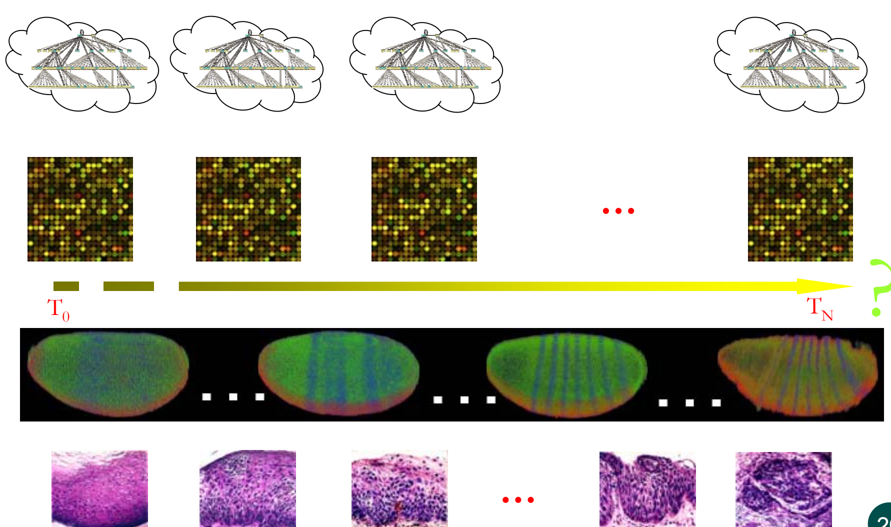


Sen. (D) Ben Nelson



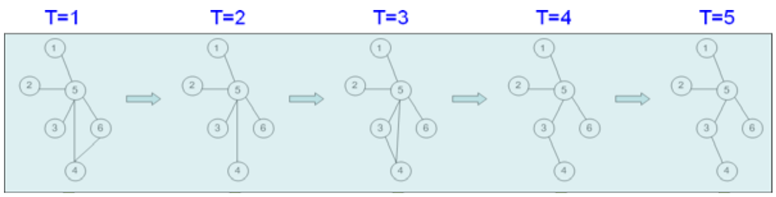


"Rewiring" Pathways in Biology



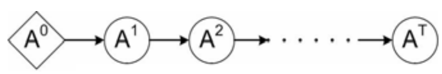
Modeling Evolving Networks

- We observe the network at discrete, evenly spaced time points $t = 1,2,\ldots,T$.



- The observed network at time t: A^t.

- We want to design a class of statistical models for the evolution of networks over a **fixed** set of nodes.



Markov Assumption

- To simplify things, assume the network observed at timestep t $(1 \leqslant t \leqslant T)$ is independent of the rest of history given the knowledge of the network at timestep t-1, then

$$P\left(A^t, A^{t-1}, \ldots, A^2, A^1\right) = P\left(A^t \middle| A^{t-1}\right) \cdots P\left(A^2 \middle| A^1\right) P\left(A^1\right)$$

- What should the conditional look like?

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Exponential Random Graphs

[Holland and Leinhardt 1981, Wasserman and Pattison 1996]

- Very general families for modeling a single static network observation.

$$P(A) = \exp\{\theta \cdot \phi(A) - \ln Z(\theta)\}$$

- $\phi(A)$ are known as a *potential*, and θ its *weight*
- Can estimate the θ parameters by MCMC MLE

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

ERGM Example

- A Classic example: (Frank & Strauss 1986)
  - $\phi_1(A)$ = # edges in A
  - $\phi_2(A)$ = # 2-stars in A
  - $\phi_3(A)$ = # triangles in A

$$P(A) \propto \exp\{\theta_1\phi_1(A) + \theta_2\phi_2(A) + \theta_3\phi_3(A)\}$$

Temporal Extension of ERGMs

- Can we build all the work on ERGMs when designing a temporal model?

$$P\left(A^t \middle| A^{t-1}\right) = \exp\left\{\theta \cdot \Psi\left(A^t, A^{t-1}\right) - \ln Z\left(\theta, A^{t-1}\right)\right\}$$

  - $\Psi(A^t, A^{t-1})$ is a *temporal potential*
  - Say the network has a single relation, and its value is either 0 or 1 (e.g., “friends” or “not friends”).
  - Let A_{ij} denote the value of the relation between ith actor and *j*th actor.
  - Then we can use $\Psi(A^t, A^{t-1})$ to capture the dynamic properties of all A_{ij}'s

An idea for specifying a model

- A network might be decomposable into different types of “motifs” (e.g., “hub & spokes”, “k-clique”, “triangle”,…).
- Write the potential functions to encode your understanding about how each motif evolves.
- It’s nice because we can “plug in” our intuition about the data.

An Example

$$P\left(A^t \middle| A^{t-1}\right) = \exp\left\{\theta \cdot \Psi\left(A^t, A^{t-1}\right) - \ln Z\left(\theta, A^{t-1}\right)\right\}$$

- "Continuity": $\Psi_1\left(A^t, A^{t-1}\right) = \sum_{ij}\left(A_{ij}^t A_{ij}^{t-1} + \left(1 - A_{ij}^t\right)\left(1 - A_{ij}^{t-1}\right)\right)$
- "Reciprocity": $\Psi_2\left(A^t, A^{t-1}\right) = \sum_{ij} A_{ij}^t A_{ji}^{t-1}$
- "Transitivity": $\Psi_3\left(A^t, A^{t-1}\right) = \frac{\sum_{ijk} A_{ij}^t A_{ik}^{t-1} A_{kj}^{t-1}}{\sum_{ijk} A_{ik}^{t-1} A_{kj}^{t-1}}$
- "Density": $\Psi_4\left(A^t, A^{t-1}\right) = \sum_{ij} A_{ij}^t$

An Example (cont.)

$$\begin{aligned} P\left(A^t \middle| A^{t-1}\right) \propto \exp\Bigg\{ & \theta_1 \sum_{ij}\left(A_{ij}^t A_{ij}^{t-1} + \left(1 - A_{ij}^t\right)\left(1 - A_{ij}^{t-1}\right)\right) \\ & + \theta_2 \sum_{ij} A_{ij}^t A_{ji}^{t-1} \\ & + \theta_3 \frac{\sum_{ijk} A_{ij}^t A_{ik}^{t-1} A_{kj}^{t-1}}{\sum_{ijk} A_{ik}^{t-1} A_{kj}^{t-1}} \\ & + \theta_4 \sum_{ij} A_{ij}^t \Bigg\} \end{aligned}$$

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Maximum Likelihood Estimation

- Approximate MLE by MCMC (*Z* intractable)

$$\ln P\left(A^T, A^{T-1}, \ldots, A^2 \middle| A^1\right) = \theta \cdot \sum_{t=2}^{T} \Psi\left(A^t, A^{t-1}\right) - \sum_{t=2}^{T} \ln Z\left(\theta, A^{t-1}\right)$$

$$\nabla \ln P\left(A^T, A^{T-1}, \ldots, A^2 \middle| A^1\right) = \sum_{t=2}^{T} \Psi\left(A^t, A^{t-1}\right) - \sum_{t=2}^{T} E_\theta\left[\Psi\left(A, A^{t-1}\right) \middle| A^{t-1}\right]$$

- Use gradient ascent, using MCMC to estimate the expectation on each iteration (as in ERGM).

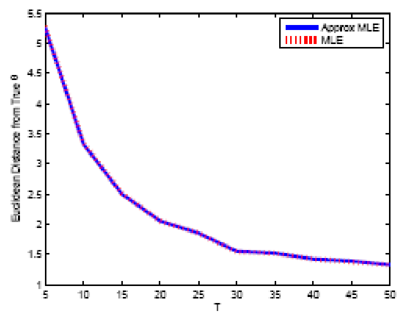
SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Estimation Toy Example

- Generate a series of 10 networks from the example.
- True model has θ_1=0, θ_2=5, θ_3=0, θ_4=-20 (i.e., reciprocity and density only)
- Estimated parameters:
  θ_1=-0.5, θ_2=4.2, θ_3=-0.08, θ_4=-20.2

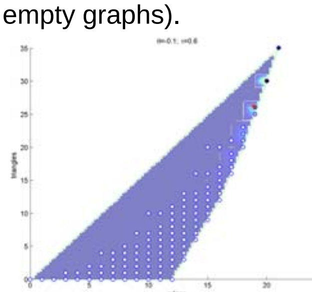
Simulation

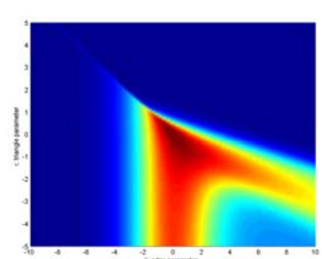
- Uses the example model
- True parameters random in [0,10)
- 100 actors



Degeneracy (Handcock et al. 2008)

- For many ERGMs, most of the parameter space is populated by distributions that place almost all of the probability mass on a subset of the sample space containing networks that bear no resemblance to the observed networks (typically the complete or empty graphs).

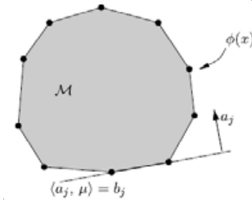




- For such models, an MLE does not exist, resulting in poor fit
  - e.g., When the observed statistics do not lie inside of the convex hull of the set of all realizable $u(A)$.

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegreative Genomics

Marginal Polytope



- ERGM

$$P(X) = \exp\{\theta \cdot \phi(X) - \ln Z(\theta)\}$$

- Realizable mean parameter set

$$\mathcal{M} := \left\{\mu \in \mathbb{R}^d \mid \exists p \text{ s.t. } \mathbb{E}_p[\phi_\alpha(X)] = \mu_\alpha, \ \forall \alpha \in \mathcal{I}\right\}$$

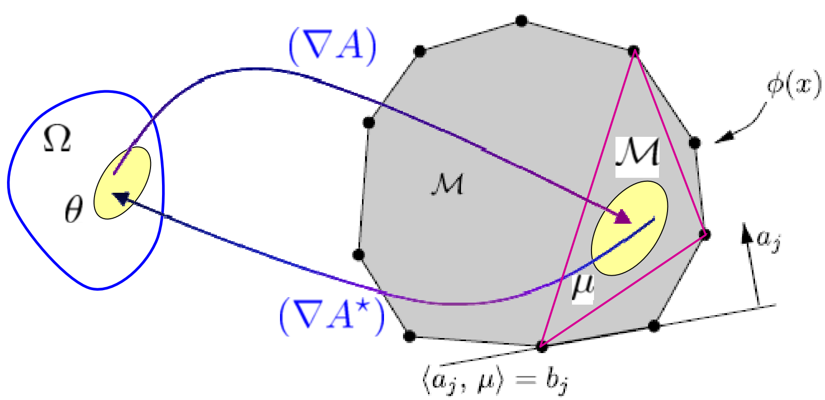
  - A convex subset of $\mathbb{R}^d$
  - Convex hull for discrete case

$$\mathcal{M} = \left\{\mu \in \mathbb{R}^d \mid \sum_{x \in \mathcal{X}^m} \phi(x)p(x) = \mu, \text{ for some } p(x) \geq 0, \sum_{x \in \mathcal{X}} p(x) = 1\right\}$$
$$\triangleq \operatorname{conv}\left\{\phi(x), x \in \mathcal{X}\right\}$$

  - Convex polytope when $|\mathcal{X}|$ is finite

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegreative Genomics

Bijection for minimal exponential family

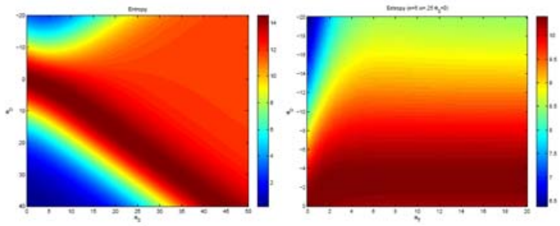


A tERGM is non-degenerate

[*Hanneke, Fu and Xing, EJS 2010*]

- **Theorem 1**: when the transition distribution factors over the edges, a tERGM is non-degenerate:

$$H(A^t) \geq \sum_{i,j} H(A^t_{i,j}|A^{t-1}) \geq p \ln \frac{1}{p} + (1-p) \ln \frac{1}{(1-p)}$$

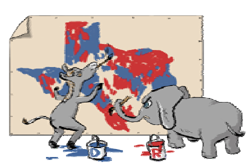


- → Maximum likelihood estimator exists!

  (should not be taken for granted for arbitrary models, be careful!)

Assessing statistic importance and quality of fit: A case study

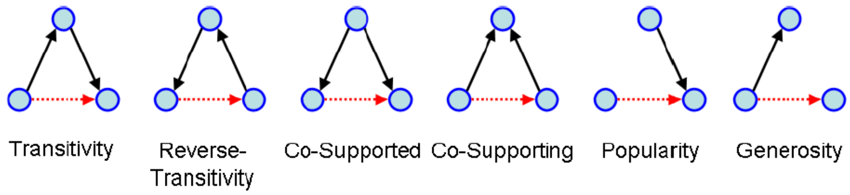
- Senate network – 109th congress



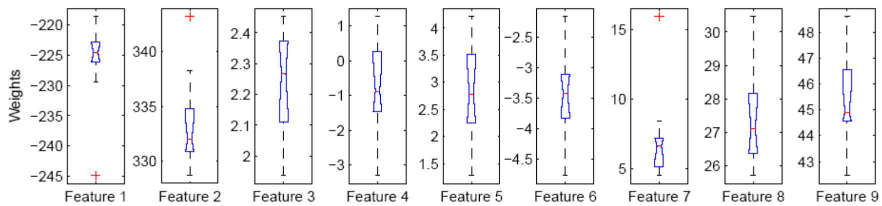
- Voting records from 109th congress (2005 - 2006)
- There are 100 senators whose votes were recorded on the 542 bills, each vote is a binary outcome

Statistic importance

- Temporal potentials:

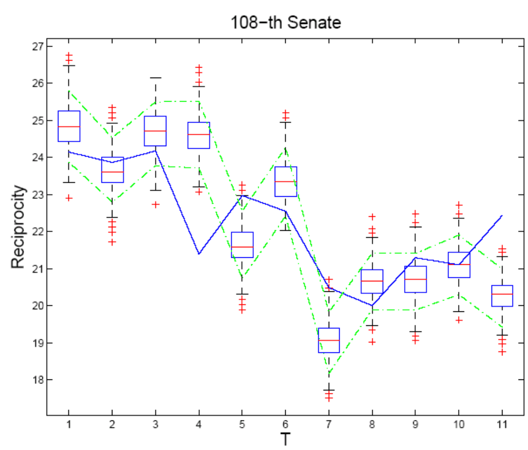


- "importance" of t-potentials:

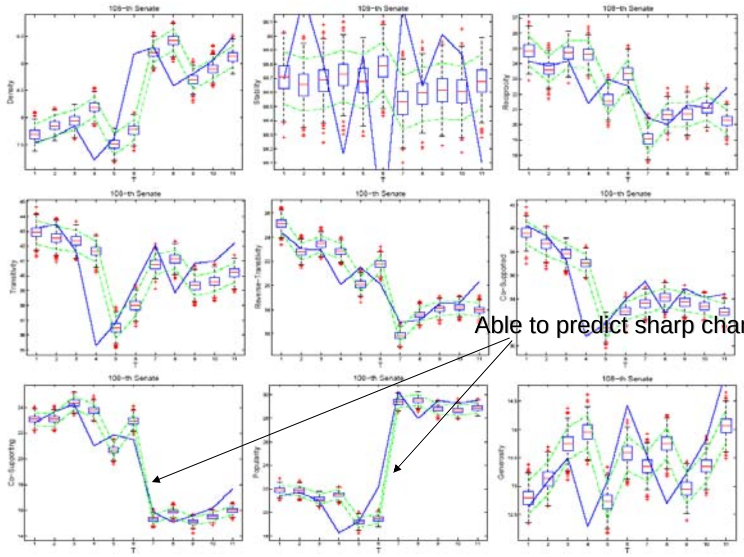


Quality of fit

- Statistic of $\Psi(A^t, A^{t-1})$ from sampled A^t from $P(A^t|A^{t-1})$ versus that from the true A^t



Quality of fit



Able to predict sharp changes

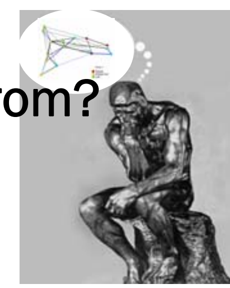
What's it good for?

- Hypothesis Testing
- Data Exploration (e.g., node-classification)
- Foundation for Learning

This Talk:

- Dynamic Network Model
  - Temporal exponential random graph model (tERGM)
    [Hanneke and Xing, **ICML 06**; Fan, Hanneke, Fu and Xing, **ICML 07**; Hanneke, Fu, and Xing, **EJS 10**]
- Reverse Engineer Latent Time-Evolving Networks
  - The TESLA and KELLER algorithms, and beyond
    [Fan, Hanneke, Fu, and Xing, **ICML 07**; Ahmed and Xing, **PNAS 09**; Song, Mladen and Xing, **ISMB 09**; Mladen, Song, Ahmed and Xing, **AOAS 09**; Mladen, Song, and Xing, **NIPS 09**; Song, Mladen and Xing, **NIPS 09**]
- Network tomography: modeling latent “multi-role" of vertices
  - Mixed Membership Stochastic Blockmodel (MMSB)
    [Airoldi, Blei, Xing and Fienberg, **LinkKDD 05**; Airoldi, Blei, Fienberg and Xing, **JMLR 08**]
- Dynamic tomography underlying evolving network
  - dynamic MMSBs
    [Xing, Fu, and Song, **AOAS 09**; Ho, Le, and Xing, **submitted 10**]

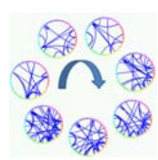
Where do networks come from?



- Existing work:
  - Assuming ***iid*** sample, and **a single *static*** network

$$\mathbf{x}^t \sim P_\theta(\mathbf{X}; G) \qquad t = 1, \ldots, T$$

  - Assuming networks or network time series are observable and given

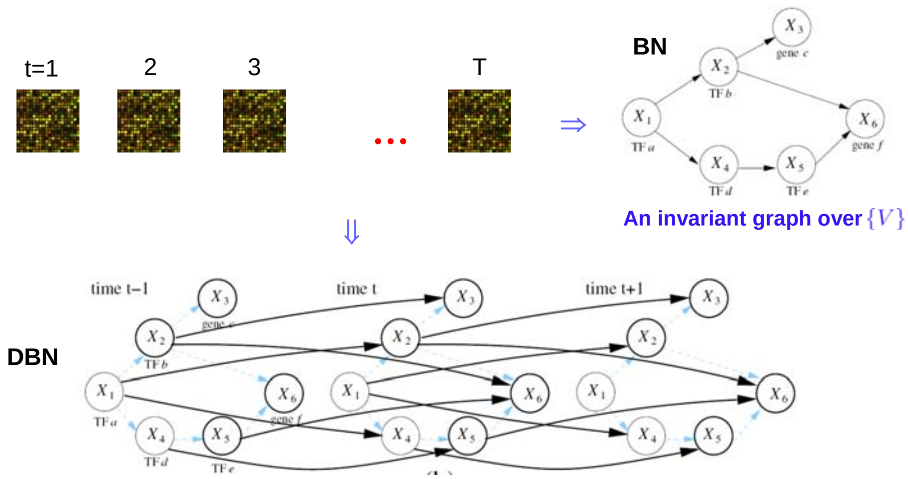


  - Then model/analyze the generative and/or dynamic mechanisms
- **We assume:**

$$\mathbf{x}^t \sim P_{\theta^t}(\mathbf{X}; G^t) \qquad t = 1, \ldots, T$$

Non-iid | Varying Coefficient | Varying Structure

Background: network inference



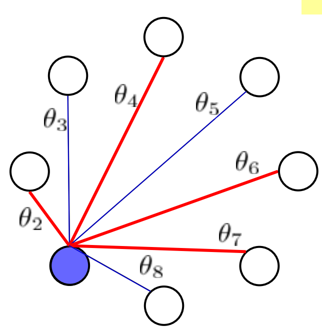
Graph Regression

$$\mathbf{X} \sim \frac{1}{Z} \exp\{\sum_i \theta_i x_i + \sum_{i<j} \theta_{i,j} x_i x_j\}$$

Markov Random Fields

$$\mathbf{X} \sim \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \exp\{-\frac{1}{2}\mathbf{x}'(\Sigma)^{-1}\mathbf{x}\}$$

Graphical Gaussian Model



$$\Theta^t \equiv (\Sigma^t)^{-1}$$

contains both the **structure** and **parameters**

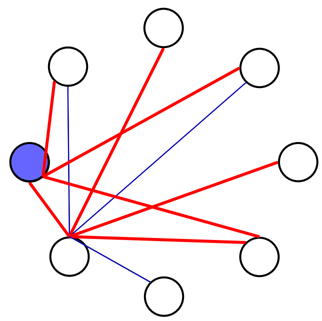
$$p(X_i|\mathbf{X}_{-i}) = GLIM(\theta^T \mathbf{X}_{-i})$$

Neighborhood selection

Lasso:

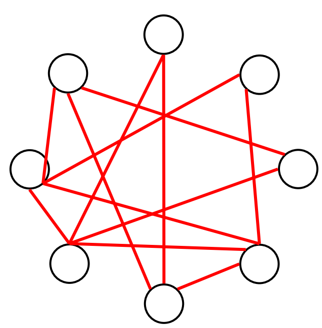
$$\hat{\theta} = \arg\min_{\theta} \sum_{n=1}^{N} \gamma(\mathbf{x}^{(n)}; \theta) + \lambda_1 \| \theta \|_1$$

Graph Regression



Neighborhood selection

Graph Regression

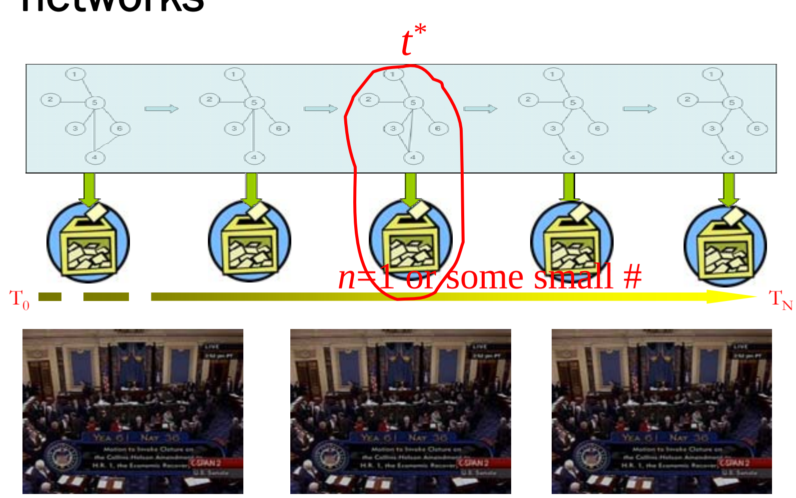


It can be shown that:
given ***iid*** samples, and under several technical conditions (e.g., "irrepresentable"), the recovered structured is "**sparsistent**" even when p >> n

Political network in the Senate

Sen. (D) Ben Nelson

Reverse engineer "rewiring" social networks



SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegreative Genomics

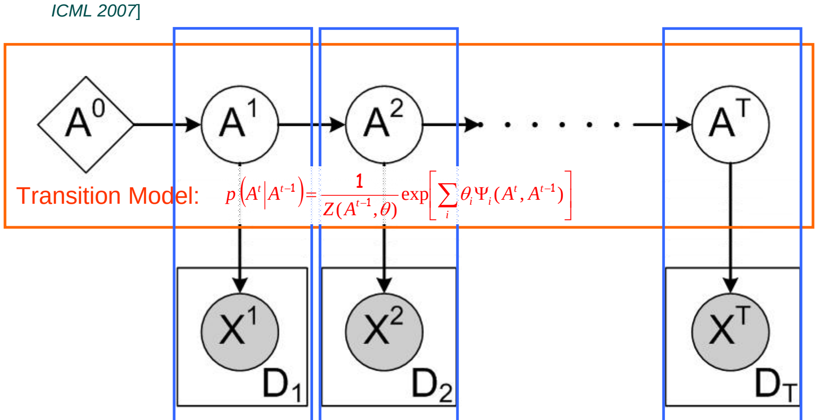
Challenges

- Very small sample size
  - observations are scarce and costly
- Noisy data
- Large dimensionality of the data
  - usually $p \gg n$
  - complexity regularization is required to avoid curse of dimensionality, e.g. sparsity
- And now the data are non-iid since underlying probability distribution is changing !

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegreative Genomics

Modeling Time-Varying Graphs

- The *hidden* temporal exponential graph models [*Fan et al. ICML 2007*]

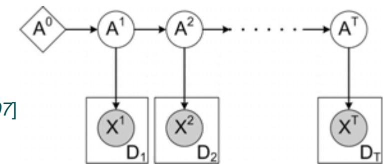


Transition Model: $p\left(A^t \middle| A^{t-1}\right) = \frac{1}{Z(A^{t-1},\theta)} \exp\left[\sum_i \theta_i \Psi_i(A^t, A^{t-1})\right]$

Emission Model: $p\left(x^t \middle| A^t, \Lambda\right) = \frac{1}{Z(A^t,\Lambda,\eta)} \exp\left[\eta \sum_{ij} \Phi_{ij}(x_i^t, x_j^t, A_{ij}^t, \Lambda_{ij})\right]$

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Inference (0) [Fan et al. ICML 2007]



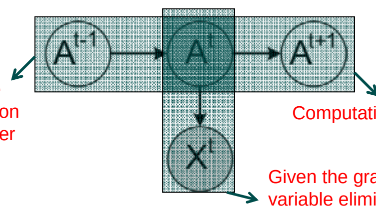
- Gibbs sampling:

P(Network|Data) ?

- Need to evaluate the log-odds

$$\mu_{ij}^t = \log \frac{P(A_{ij}^t = 1 | A_{t-1}, A_{-ij}^t, A^{t+1}, x^t)}{P(A_{ij}^t = 0 | A_{t-1}, A_{-ij}^t, A^{t+1}, x^t)}$$
$$= \log \frac{P(A_{-ij}^t, A_{ij}^t = 1 | A^{t-1})}{P(A_{-ij}^t, A_{ij}^t = 0 | A^{t-1})} + \log \frac{P(A^{t+1} | A_{-ij}^t, A_{ij}^t = 1)}{P(A^{t+1} | A_{-ij}^t, A_{ij}^t = 0)} + \log \frac{P(x^t | A_{-ij}^t, A_{ij}^t = 1)}{P(x^t | A_{-ij}^t, A_{ij}^t = 0)}$$

- Difficulty: Evaluate the ratio of Partition function $Z(A') = \Sigma_A \exp(\theta \Phi(A, A'))$
- **So far scale to ~20 nodes !!!**



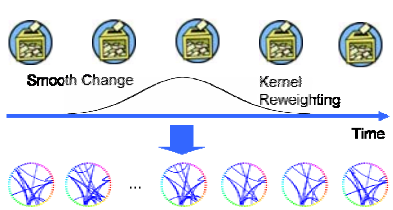
Straightforward -- tractable transition model; the partition function is the product of per edge terms

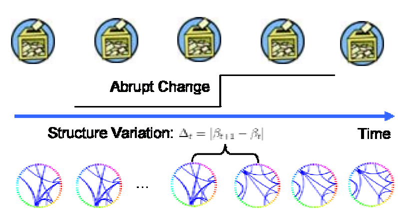
Computation is non-trivial

Given the graphical structure, run variable elimination algorithms, works well only for small graphs

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Two Scenarios





Smoothly evolving graphs

Abruptly evolving graphs

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Lasso:

$$\hat{\theta} = \arg\min_{\theta} \sum_{n=1}^{N} \gamma(\mathbf{x}^{(n)};\theta) + \lambda_1 \| \theta \|_1$$

Inference I

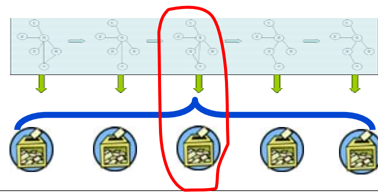
[Song, Kolar and Xing, Bioinformatics 09]

- **KELLER**: Kernel Weighted L_1-regularized Logistic Regression

$$\hat{\theta}_i^t = \arg\min_{\theta_i^t} l_w(\theta_i^t) + \lambda_1 \| \theta_i^t \|_1 \quad \forall t$$

$$\text{where} \quad l_w(\theta_i^t) = \sum_{t'=1}^{T} w(\mathbf{x}^{t'};\mathbf{x}^t) \log P(x_i^{t'} | \mathbf{x}_{-i}^{t'}, \theta_i^t).$$

- Constrained convex optimization
  - **Estimate time-specific nets one by one, based on "virtual iid" samples**
  - Could scale to ~10^4 genes, but under stronger smoothness assumptions



SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Problem formulation

- Formulate as structure learning problem of a time-evolving Markov Random Fields

$$\mathcal{D}^n = \{X^t \sim \mathbb{P}_{\theta^t} \mid t = 1/n, 2/n, \ldots, 1\}$$

$$\mathbb{P}_{\theta^t}(X) = \frac{1}{Z(\theta^t)} \exp\left(\sum_{(u,v)\in E^t} \theta_{uv}^t x_u x_v \right)$$

- Idea: maximize the likelihood to obtain the structure

$$\hat{\theta}^{t^*} = \arg \min_{\|\theta\|_1 \le C(\lambda_n)} \left\{ -\sum_{t\in\mathcal{T}^n} w_t(t^*) \log \mathbb{P}_{\theta^t}(x^t) \right\}$$

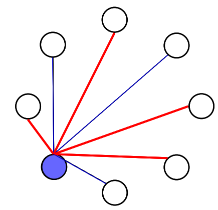
- Calculation of likelihood: intractable (because of the Z)

Algorithm - neighborhood selection

- Conditional likelihood

$$\mathbb{P}_{\theta^t}(x_i^t | x_{\setminus i}^t) = \text{logistic}\left(2x_i^t \left\langle \theta_{\setminus i}^t, x_{\setminus i}^t \right\rangle\right)$$

  - Neighborhood: $S(x_i) = \{j \mid \theta_{i,j}^t \neq 0\}$

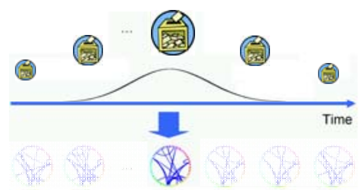


- Time-specific graph regression:
  - Estimate at $t^* \in [0,1]$

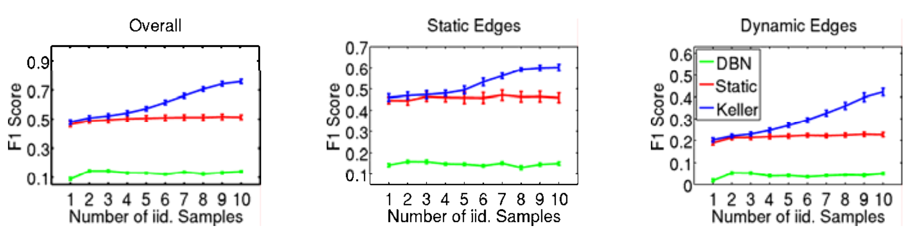
$$\min_{\theta \in \mathbb{R}^{p_n - 1}} \left\{ - \sum_{t \in \mathcal{T}^n} w_t(t^*) \gamma(\theta_i; x^t) + \lambda_1 \|\theta_i\|_1 \right\}$$

Where $\gamma(\theta_i^t; x^t) = \log \mathbb{P}_{\theta_i^t}(x_i^t | x_{\setminus i}^t)$

and $w_t(t^*) = \dfrac{K_{h_n}(t - t^*)}{\sum_{t' \in \mathcal{T}^n} K_{h_n}(t' - t^*)}$



Synthetic data



Structural consistency of KELLER

Assumptions

- Define: $Q_u^t := \mathbb{E}\left[\nabla^2 \log \mathbb{P}_{\theta^t}[X_u | X_{\backslash u}]\right], \quad \forall u \in V$
  $\Sigma_u^t := \mathbb{E}\left[X_{\backslash u}^t {X_{\backslash u}^t}^T\right], \quad \forall u \in V$
  $s = \max_u \max_t |S_u^t|, \quad \theta_{\min} = \min_{e \in E} \max |\theta_e^t|$
- A1: Dependency Condition
  $$\Lambda_{\min}(Q_{SS}^{t^*}) \geq C_{\min}, \quad \forall t \in [0,1]$$
  $$\Lambda_{\max}\left(\Sigma^{t^*}\right) \leq D_{\max}, \quad \forall t \in [0,1]$$
- A2: Incoherence Condition $\exists \alpha \in (0,1]$ such that
  $$\left\| Q_{S^c S}^{t^*} (Q_{SS}^{t^*})^{-1} \right\|_\infty \leq 1 - \alpha, \quad \forall t^* \in [0,1]$$
- A3: Smoothness Condition
  $$\max_{u,v} \sup_{t^*} |\sigma'_{uv}(t^*)| \leq A_0, \quad \max_{u,v} \sup_{t^*} |\sigma''_{uv}(t^*)| \leq A$$
  $$\max_{u,v} \sup_{t^*} |\theta'_{uv}(t^*)| \leq B_0, \quad \max_{u,v} \sup_{t^*} |\theta''_{uv}(t^*)| \leq B$$

Theorem

[Kolar and Xing, 09]

Assume that A1, A2, A3 hold. Furthermore, assume that the following conditions hold:

1. $h_n = \mathcal{O}(n^{-\frac{1}{3}})$
2. $s_n h_n = o(1)$,
3. $\frac{s_n^3 \log p_n}{n h_n} = o(1)$
4. $\lambda_1 = \mathcal{O}(\sqrt{\frac{\log p}{n h_n}})$
5. $\theta^*_{\min} = \Omega(\sqrt{\frac{s_n \log p_n}{n h_n}})$

then

$$\mathbb{P}\left[\hat{G}(\lambda_1, h_n, t^*) \neq G^{t^*}\right] = \mathcal{O}\left(\exp\left(-C \frac{n h_n}{s_n^3} + C' \log p\right)\right) \to 0$$

Inference II [Ahmed and Xing, PNAS 09]

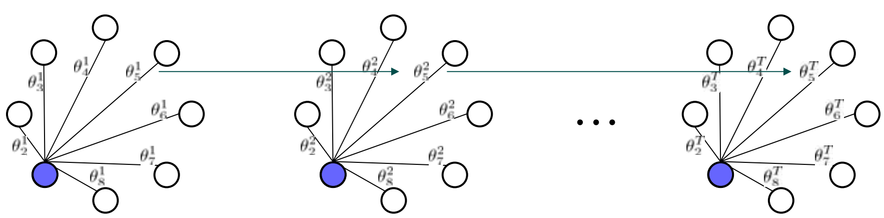
- **TESLA**: Temporally Smoothed L_1-regularized logistic regression

$$\hat{\theta}_i^1, \ldots, \hat{\theta}_i^T = \arg\min_{\theta_i^1,\ldots,\theta_i^T} \sum_{t=1}^{T} l_{avg}(\theta_i^t) + \lambda_1 \sum_{t=1}^{T} \| \theta_i^t \|_1 + \lambda_2 \sum_{t=2}^{T} \| \theta_i^t - \theta_i^{t-1} \|_q^q,$$

where $l_{avg}(\theta_{\mathbf{i}}^{\mathbf{t}}) = \frac{1}{N^t} \sum_{d=1}^{N^t} \log P(x_{d,i}^t | \mathbf{x}_{\mathbf{d},-\mathbf{i}}^{\mathbf{t}}, \theta_{\mathbf{i}}^{\mathbf{t}})$.

- Constrained convex optimization
  - **Estimate time-specific nets jointly, based on original "non-iid" samples**
  - Scale to ~5000 nodes, does not need smoothness assumption, can accommodate abrupt changes.

Temporally Smoothed Graph Regression



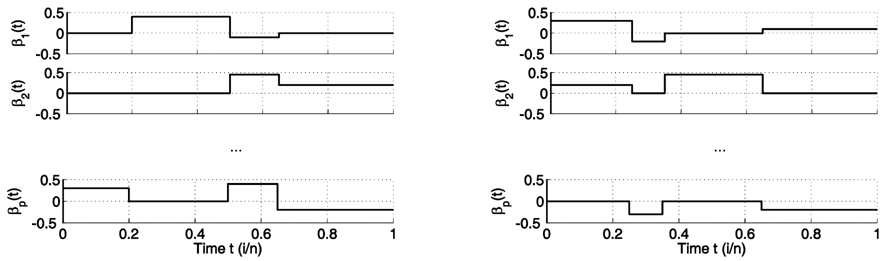
TESLA:

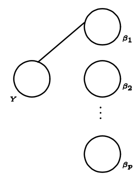
$$\min_{\substack{\theta_i^1,\ldots,\theta_i^T \\ \mathbf{u}_i^1,\ldots,\mathbf{u}_i^T;\mathbf{v}_i^2,\ldots,\mathbf{v}_i^T}} \sum_{t=1}^{T} \ell(\mathbf{x}^t; \theta_i^t) + \lambda_1 \sum_{t=1}^{T} \mathbf{1}'\mathbf{u}_i^t + \lambda_2 \sum_{t=2}^{T} \mathbf{1}'\mathbf{v}_i^t$$

$$\text{s.t.} \quad -u_{i,j}^t \leq \theta_{i,j}^t \leq u_{i,j}^t, \; t = 1, \ldots, T, \; \forall j \in V \setminus i,$$

$$\text{s.t.} \quad -v_{i,j}^t \leq \theta_{i,j}^t - \theta_{i,j}^{t-1} \leq v_{i,j}^t, \; t = 2, \ldots, T, \; \forall j \in V \setminus i,$$

Coefficients as functions





Modified estimation procedure

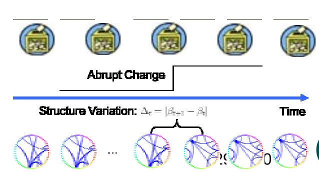
[Kolar, Le and Xing, NIPS 09]

- estimate block partition on which the coefficient functions are constant

$$\min_{\beta} \sum_{i=1}^{n} (Y_i - \mathbf{X}_i \beta(t_i))^2 + 2\lambda_2 \sum_{k=1}^{p} ||\beta_k||_{\mathrm{TV}} \qquad (*)$$

- estimate the coefficient functions on each block of the partition

$$\min_{\gamma \in \mathbb{R}^p} \sum_{t_i \in{}_j} (Y_i - \mathbf{X}_i \gamma)^2 + 2\lambda_1 ||\gamma||_1 \qquad (**)$$



SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Structural Consistency of TESLA

[Kolar, Le and Xing, NIPS 09]

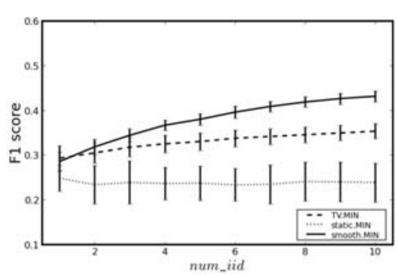
I. It can be shown that, by applying the results for model selection of the Lasso on a *temporal difference transformation* of (*), **the block are estimated consistently**

II. Then it can be further shown that, by applying Lasso on (**), **the neighborhood of each node on each of the estimated blocks consistently**

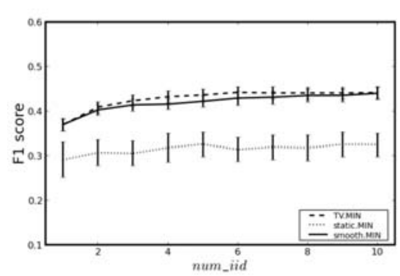
- Further advantages of the two step procedure
  - choosing parameters easier
  - faster optimization procedure

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Comparison of KELLER and TESLA



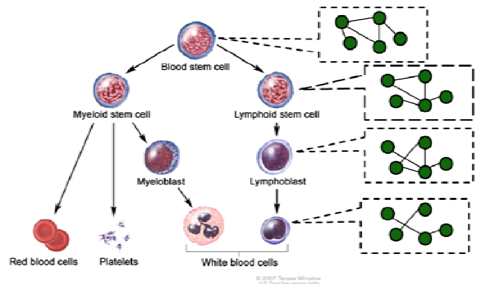
Smoothly varying



Abruptly varying

Inference III [Parikh et. al, 2010]

- **TREEGL**: Tree-Smoothed Graph Lasso



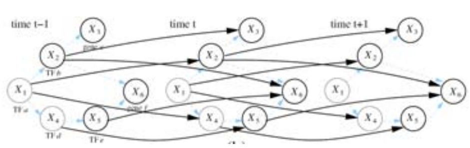
$$\hat{\theta}^{(1)}_{\setminus u},\dots,\hat{\theta}^{(n)}_{\setminus u} = argmin_{\theta^{(1)},\dots,\theta^{(n)}}\Big(-\sum_{n=1}^{N}\sum_{s=1}^{S_n}\gamma(\theta^{(n)},\mathbf{x}^{(n,s)}) + \lambda_1\sum_{n=1}^{N}\|\theta^{(n)}\|_1 + \lambda_2\sum_{n=2}^{N}\|\theta^{(n)}-\theta^{(\pi(n))}\|_1\Big)$$

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Time-Varying Dynamic Bayesian Networks [Song, Kolar and Xing, NIPS 09]

- Autoregressive model with time-varying coeff/struct

$$X^t = A^t \cdot X^{t-1} + \epsilon,\ \epsilon \sim \mathcal{N}(0, \sigma^2 I)$$

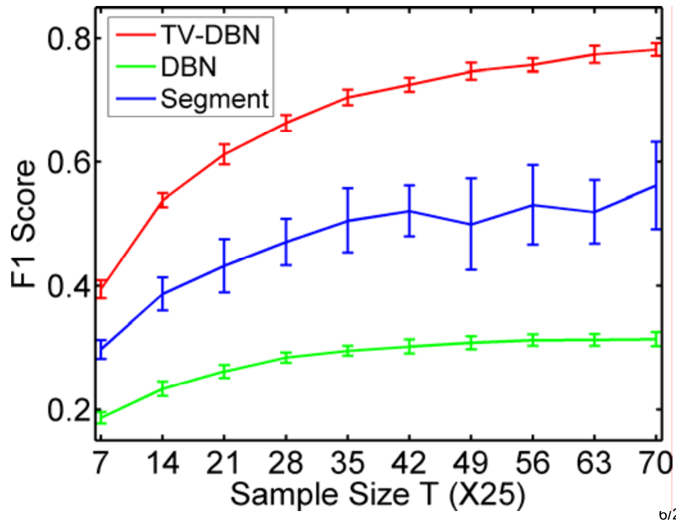


"Granger causality"

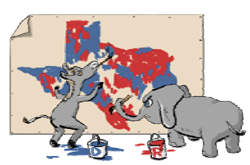
- Sparsity pattern in *A* corresponds to a *directed* TV-network

$$\mathcal{E}^t = \{(i,j) \in \mathcal{V}\times\mathcal{V} \mid i \neq j, \boldsymbol{A}^t_{ij} \neq 0\}$$

Performance on Synthetic TV-DBN



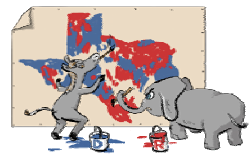
Senate network – 109^{th} congress

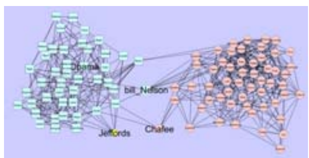


- Voting records from 109th congress (2005 - 2006)
- There are 100 senators whose votes were recorded on the 542 bills, each vote is a binary outcome

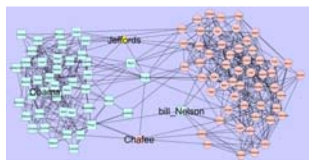
- Estimating parameters:
  - KELLER: bandwidth parameter to be h_n = 0.174, and the penalty parameter λ_1 = 0.195
  - TESLA: λ_1 = 0.24 and λ_2 = 0.28

Senate network – 109th congress

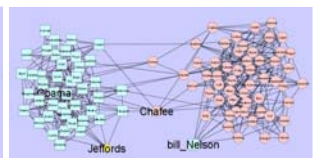




March 2005

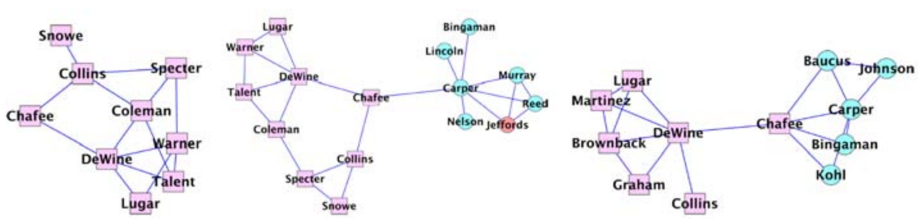


January 2006



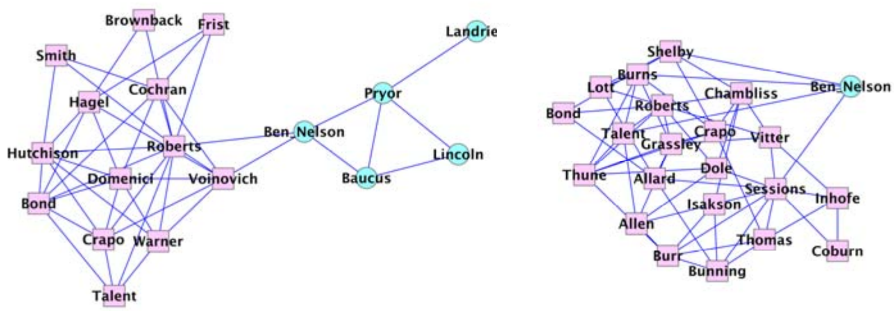
August 2006

Senator Chafee



(a) t = 0.1 (b) t = 0.4 (c) t = 0.8

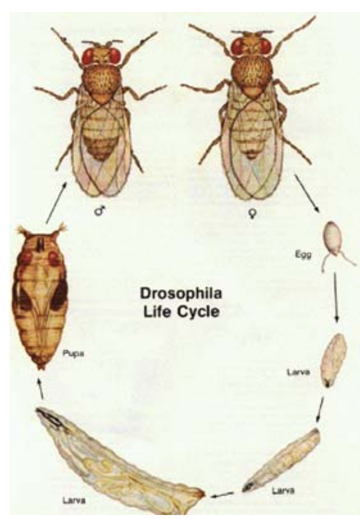
Senator Ben Nelson



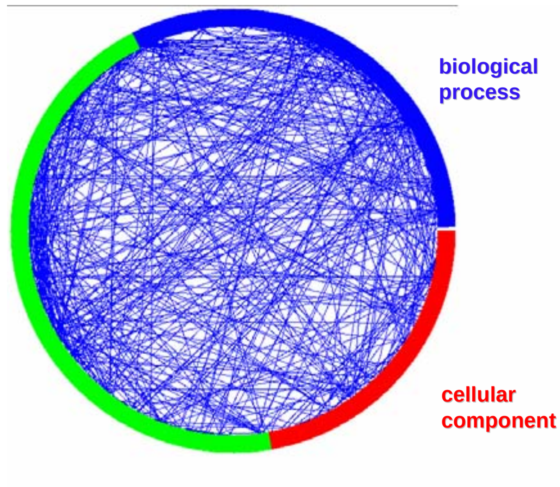
T=0.2 **T=0.8**

Drosophila life cycle

- From Arbeitman et al. (2002)
- Four stages:
  - embryo, larva, pupa, adult
- 66 microarray measured across full life cycle
- Focus on 588 development related genes

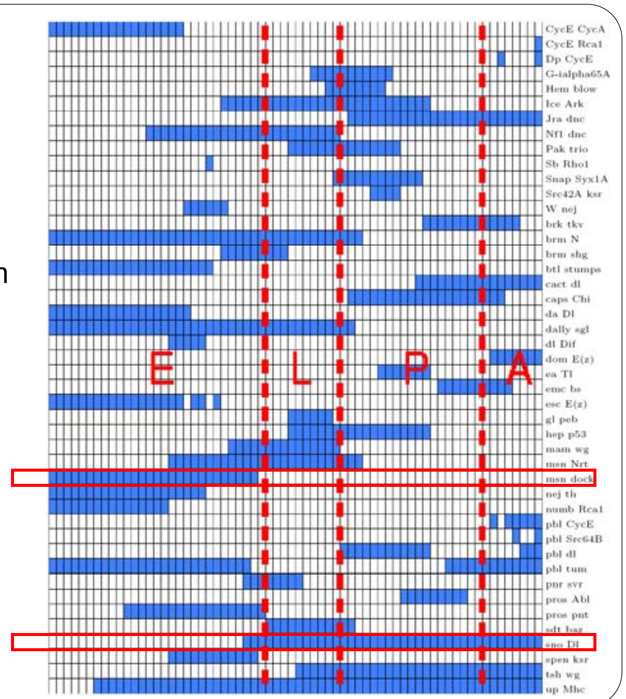


Dynamic Gene Interactions Networks of *Drosophila Melanogaster*



Known Gene Interactions

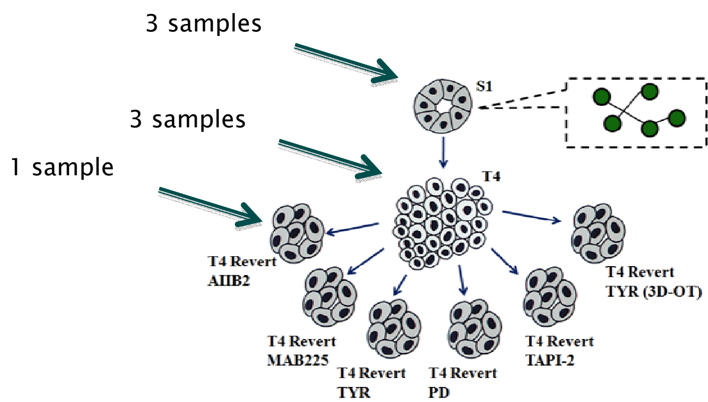
- Visualizing Time-span of known gene interactions



SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Breast cancer progression/reversal

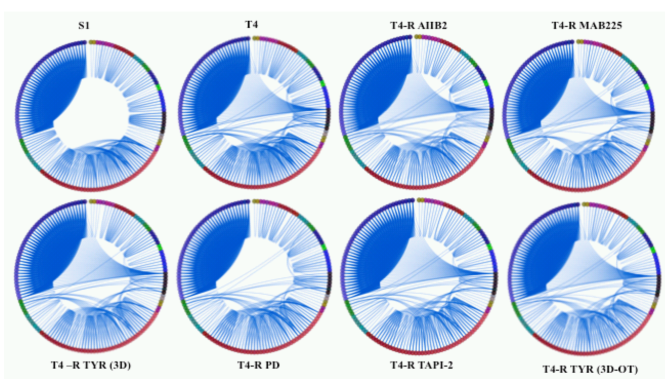
[Ankur et al, Submitted 2010]



Based on 3-dimensional Organotypic cell culture

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Tree-evolving networks

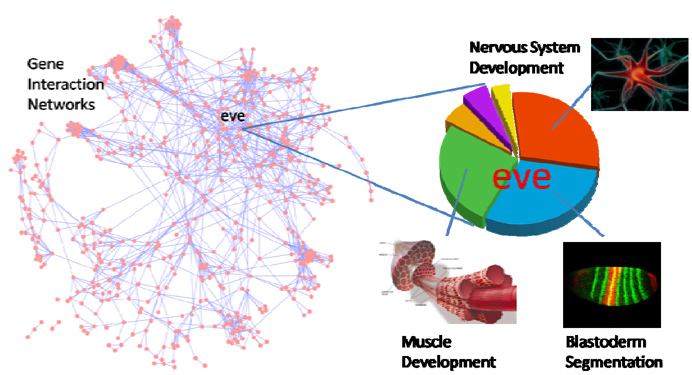


This Talk:

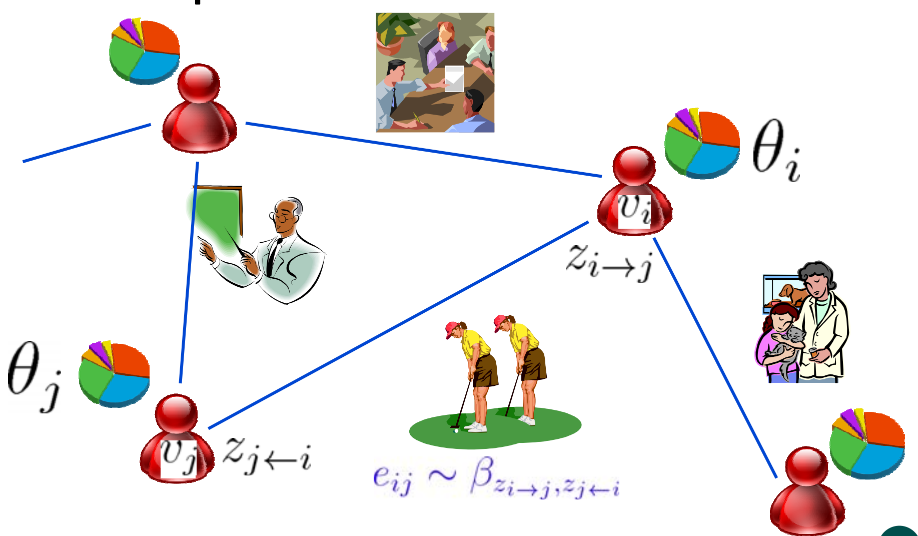
- Dynamic Network Model
  - Temporal exponential random graph model (tERGM)
    [Hanneke and Xing, **ICML 06**; Fan, Hanneke, Fu and Xing, **ICML 07**; Hanneke, Fu, and Xing, **EJS 10**]
- Reverse Engineer Latent Time-Evolving Networks
  - The TESLA and KELLER algorithms, and beyond
    [Fan, Hanneke, Fu, and Xing, **ICML 07**; Ahmed and Xing, **PNAS 09**; Song, Mladen and Xing, **ISMB 09**; Mladen, Song, Ahmed and Xing, **AOAS 09**; Mladen, Song, and Xing, **NIPS 09**; Song, Mladen and Xing, **NIPS 09**]
- Network tomography: modeling latent “multi-role" of vertices
  - Mixed Membership Stochastic Blockmodel (MMSB)
    [Airoldi, Blei, Xing and Fienberg, **LinkKDD 05**; Airoldi, Blei, Fienberg and Xing, **JMLR 08**]
- Dynamic tomography underlying evolving network
  - dynamic MMSBs
    [Xing, Fu, and Song, **AOAS 09**; Ho, Le, and Xing, **submitted 10**]

Network tomography

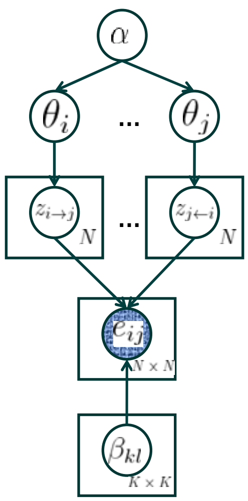
- Multi-role of every node
- Context dependent role-instantiation
- Role dynamics



Example:



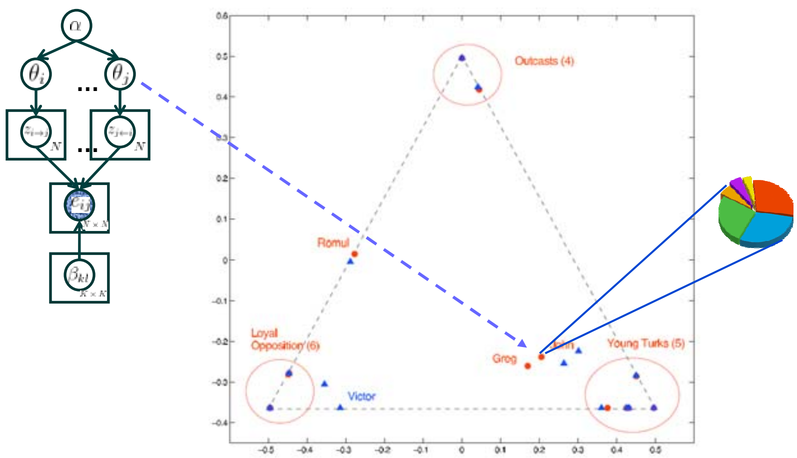
Mixed Membership Stochastic Blockmodel [Airoldi, Blei, Fienberg and Xing, 2008]



1. $\{\theta_i\}_{i=1}^N \sim p(\theta|\alpha) \equiv \text{Dirichlet}(\theta;\alpha)$ sample mixed membership vectors.
2. For each actor v_j that actor v_i possibly interacts with:
   - $z_{i\to j} \sim \text{Multinomial}(z|\theta_i)$ sample an indicator for v_i;
   - $z_{i\leftarrow j} \sim \text{Multinomial}(z|\theta_j)$ sample an indicator for v_j;
   - $e_{ij} \sim \text{Bernoulli}(e|z_{i\to j}^\top B z_{i\leftarrow j})$ sample a link.

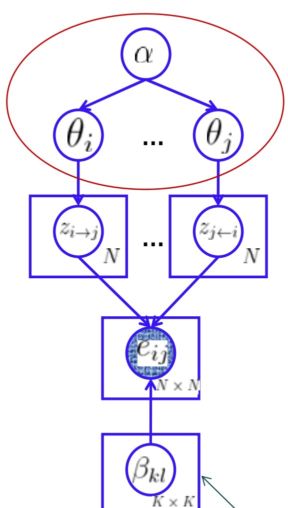
SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

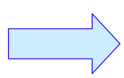
In the mixed-membership simplex



SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Logistic Normal Prior



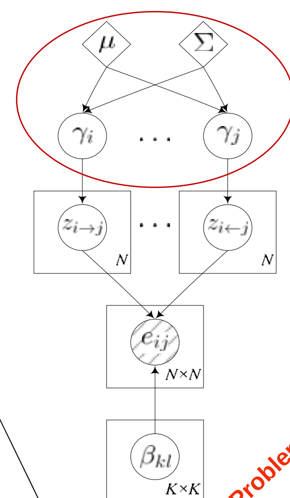


$\theta \sim LN_K(\mu,\Sigma)$

$\Rightarrow$

$\gamma_{1:K-1} \sim N_{K-1}(\mu,\Sigma), \quad \gamma_K = 0$

$\theta_i = \exp\left\{\gamma_i - \log\left(1 + \sum_{i=1}^{K-1} e^{\gamma_i}\right)\right\}$

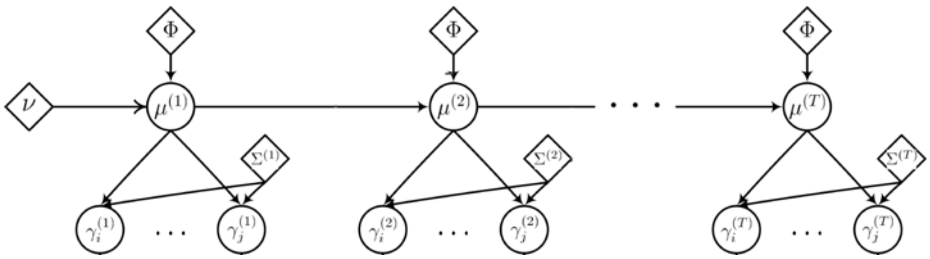


Role Compatibility Matrix

Problem

- Non-conjugate
- Laplace Approximation

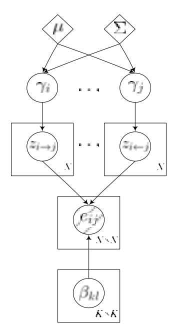
Can be combined with SSM



$$\mu^{(1)} = \nu + w^{(1)}, \quad w^{(1)} \sim \mathcal{N}(0, \Phi)$$

$$\mu^{(t)} = A\mu^{(t-1)} + w^{(t)}, \quad w^{(t)} \sim \mathcal{N}(0, \Phi)$$

Inference and Learning

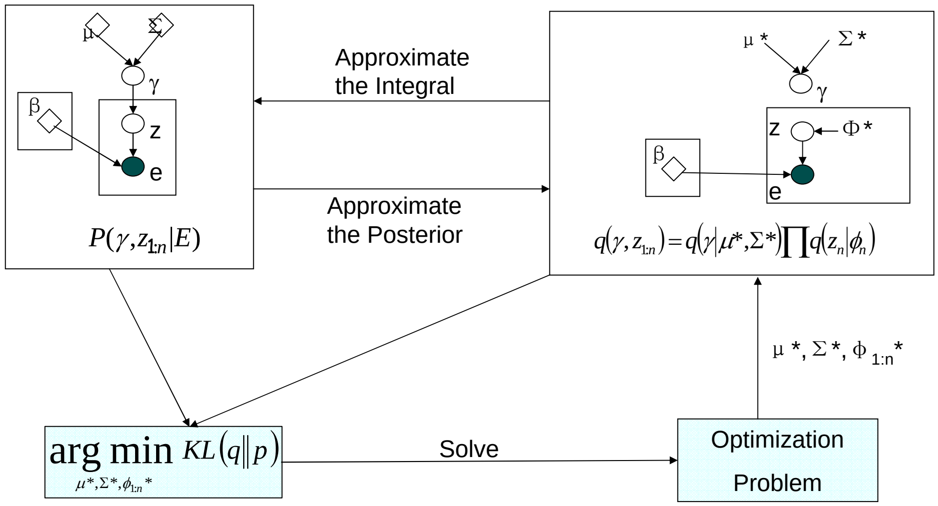


- Inference
  - Given a Network $E = \{e_{ij}\}$
    - Evaluation:
      $P(E \mid \mu, \Sigma, \beta)$
    - Posterior:
      $P(\gamma, \mathbf{z} \mid \mu, \Sigma, \beta, E)$
- Learning
  - Given a Network $E = \{e_{ij}\}$
    - Parameter estimation

$$\{\mu^*, \Sigma^*, \beta^*\} = \arg\max_{(\mu,\Sigma,\beta)} \sum \log\left(P\left(\{e_{ij}\} \mid \mu, \Sigma, \beta\right)\right)$$

Intractable!
e.g. $P(E \mid \mu, \Sigma, \beta)$

Variational Inference



Laplace Variations Inference

- Fully Factored Distribution

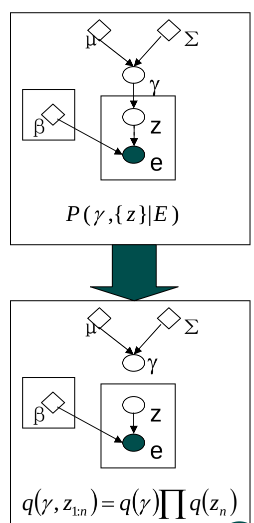
$$q(\gamma, z_{1:n}) = q(\gamma) \prod q(z_n)$$

- Fixed Point Equations

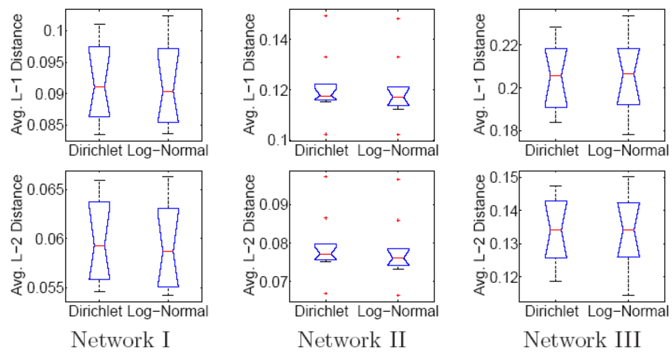
$$q_\gamma^*(\gamma) = P\left(\gamma \middle| \langle S_z \rangle_{q_z}, \mu, \Sigma\right) \approx N(\mu_\gamma, \Sigma_\gamma)$$

$$q_z^*(z) = P\left(z \middle| \langle S_\gamma \rangle_{q\gamma}, \beta_{1:k}\right) \approx \text{Multi}(\theta_z)$$

Laplace approximation



Dirichlet vs. LN MMSB



Prior	Avg. ℓ_2 distance	Log-likelihood
Dirichlet	0.091	-5755.8
Logistic Normal	0.092	-5691.7

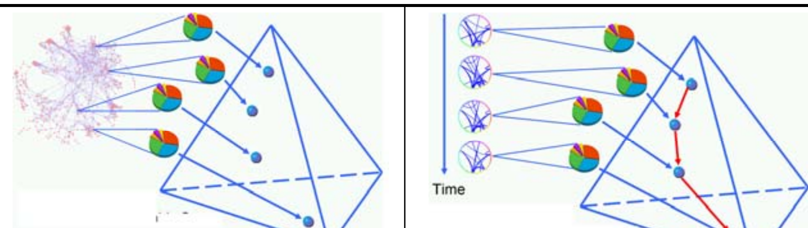
This Talk:

- Dynamic Network Model
  - Temporal exponential random graph model (tERGM)
    [Hanneke and Xing, **ICML 06**; Fan, Hanneke, Fu and Xing, **ICML 07**; Hanneke, Fu, and Xing, **EJS 10**]
- Reverse Engineer Latent Time-Evolving Networks
  - The TESLA and KELLER algorithms, and beyond
    [Fan, Hanneke, Fu, and Xing, **ICML 07**; Ahmed and Xing, **PNAS 09**; Song, Mladen and Xing, **ISMB 09**; Mladen, Song, Ahmed and Xing, **AOAS 09**; Mladen, Song, and Xing, **NIPS 09**; Song, Mladen and Xing, **NIPS 09**]
- Network tomography: modeling latent “multi-role" of vertices
  - Mixed Membership Stochastic Blockmodel (MMSB)
    [Airoldi, Blei, Xing and Fienberg, **LinkKDD 05**; Airoldi, Blei, Fienberg and Xing, **JMLR 08**]
- Dynamic tomography underlying evolving network
  - dynamic MMSBs
    [Xing, Fu, and Song, **AOAS 09**; Ho, Le, and Xing, **submitted 10**]

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Dynamic tomography

- How to model dynamics in a simplex?

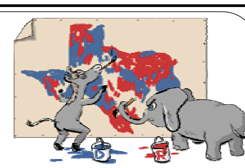


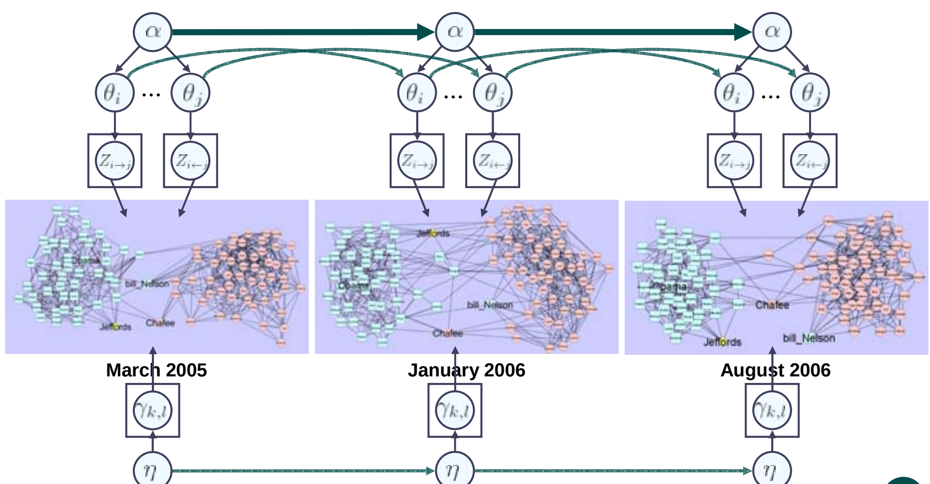
Project an individual/stock in network into a "tomographic" space

Trajectory of an individual/stock in the "tomographic" space

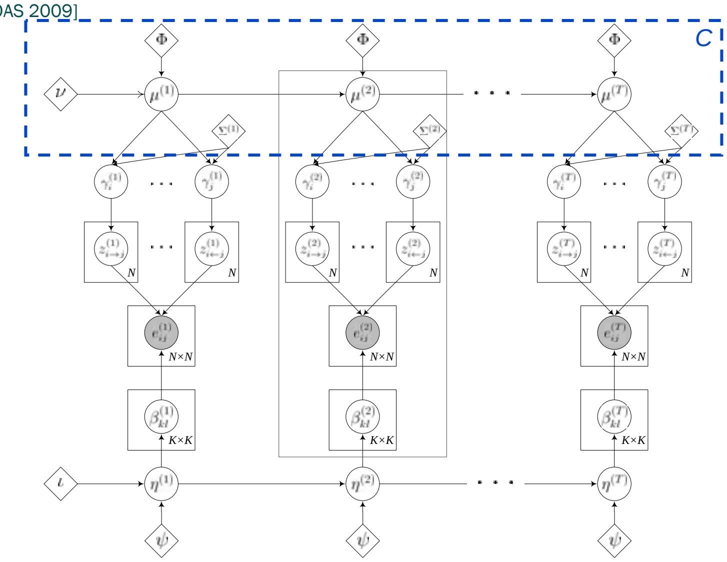
SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Evolving networks



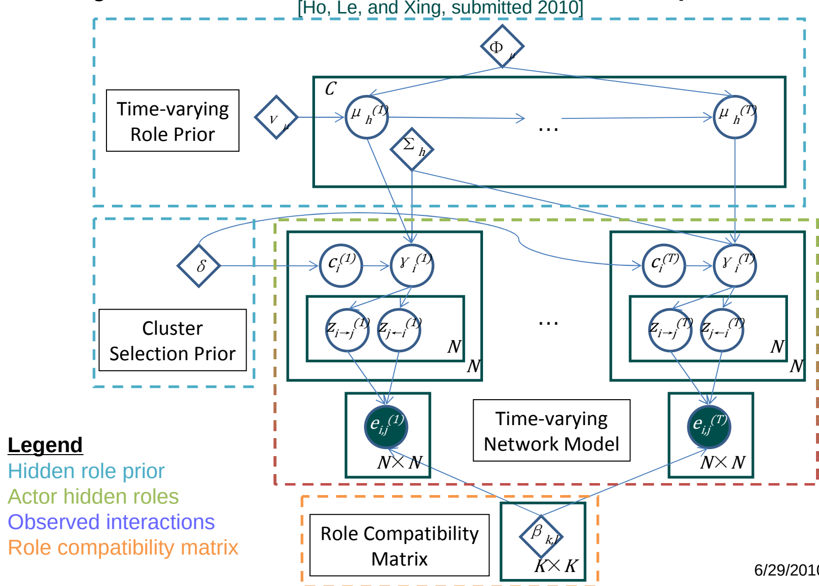


Dynamic MMSB (dMMSB) [Xing, Fu, and Song, AOAS 2009]



Dynamic Mixture of MMSB (dM³SB)

[Ho, Le, and Xing, submitted 2010]



SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

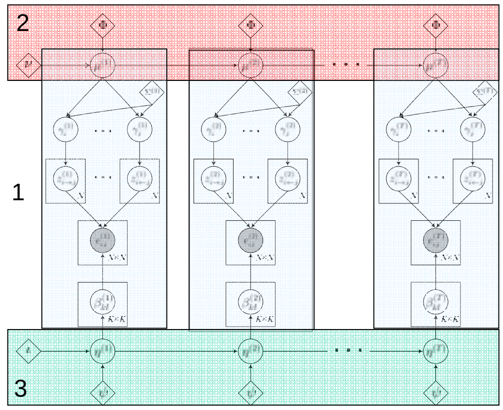
Algorithm: Generalized Mean Field

(xing et al. 2004)

Approximate the joint posterior $p\Big(\{\vec{\mathbf{z}}^{(t)}, \vec{\pi}^{(t)}, \vec{\mu}^{(t)}, B^{(t)}\}_{t=1}^{T} | \Theta, \{G^{(t)}\}_{t=1}^{T}\Big)$ where Θ denotes the model parameters, by a factored approximate distribution:

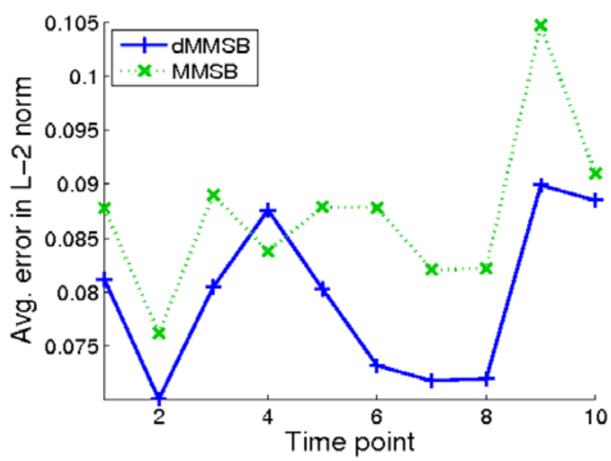
$$q\Big(\{\vec{\mathbf{z}}^{(t)}, \vec{\pi}^{(t)}, \vec{\mu}^{(t)}, B^{(t)}\}_{t=1}^{T}\Big) = q_1\big(\{\vec{\mathbf{z}}^{(t)}, \vec{\pi}^{(t)}\}_{t=1}^{T}\big) \times q_2\big(\{\vec{\mu}^{(t)}\}_{t=1}^{T}\big) \times q_3\big(\{B^{(t)}\}_{t=1}^{T}\big),$$

- Inference via variational EM
  - Generalized mean field
  - Laplace approximation
  - Kalman filter & RTS smoother



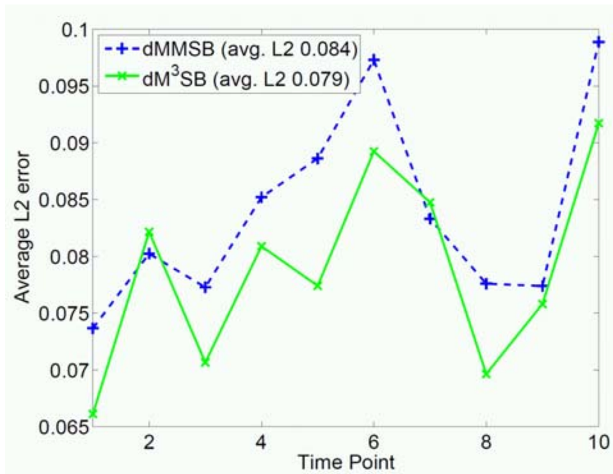
SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

dMMSB vs. MMSB



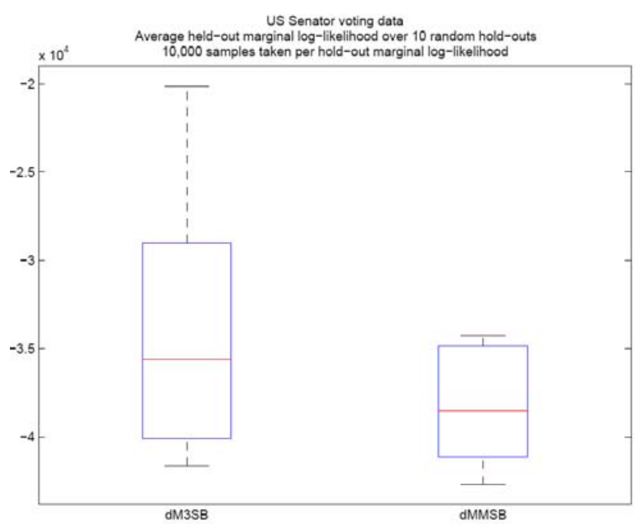
SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

dM^{3}SB vs. dMMSB

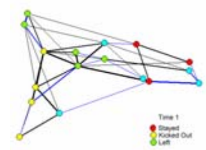


SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Goodness of fit

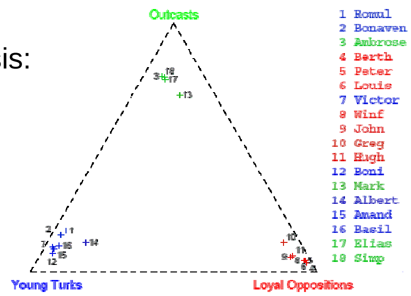


Case Study 0: Sampson's Monk Network



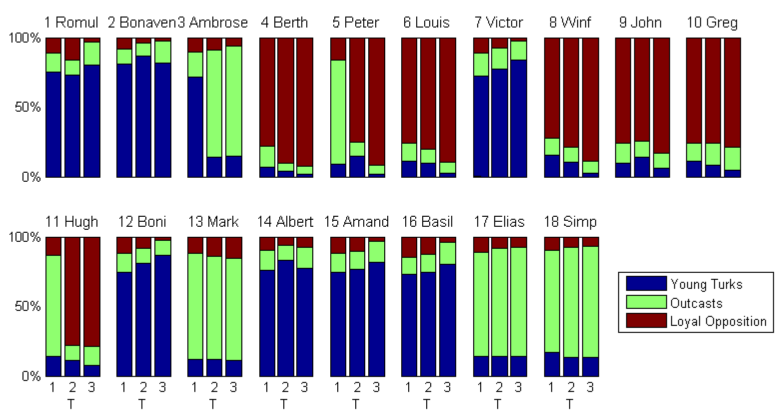
- Dataset Description
  - 18 monks (junior members in a monastery)
  - Liking relations recorded
  - 3 time-points in one year period
  - Timing: before a major conflict outbreak

- Recall static analysis:



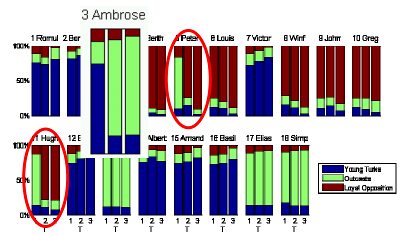
Sampson's Monk Network: role trajectories

- The trajectories of the varying role-vectors over time



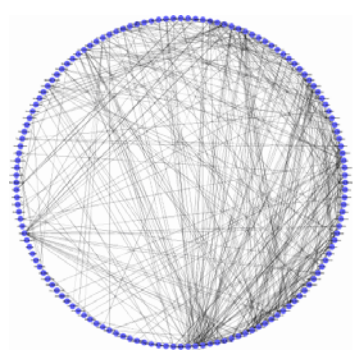
Sampson's Monk Network: Dynamic Analysis

- Observations
  - Big changes in time 1 to time 2
  - From time 2 to time 3, role-vectors purifying
    - More isolated
    - Led to the separation



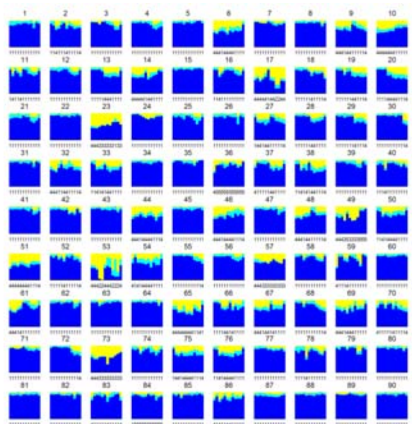
Case Study 1: The Enron Network

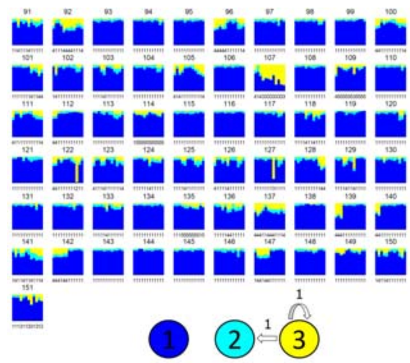
- Dataset Description
  - 151 persons considered
  - used emails from 2001, and built an email network for each month,
  - so the dynamic network has 12 time points.
  - we learned a dMMSB of 5 latent roles.



SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Enron Network: role trajectories





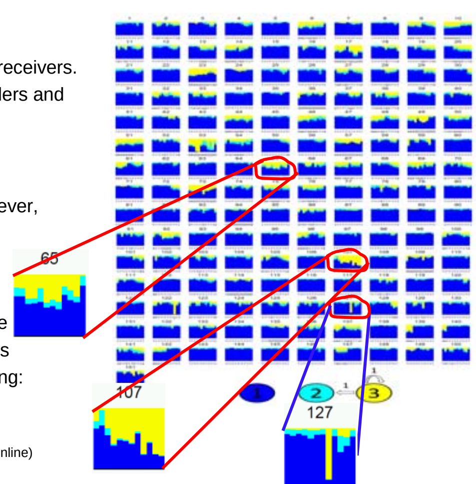
Role Compatibility Matrix *B*
Role 1 = Passive, 2 = Receiver, 3 = Sender/Receiver

Colored bars: Estimated latent space vector (for each time point)
Numbers under bars: Estimated cluster (for each time point)

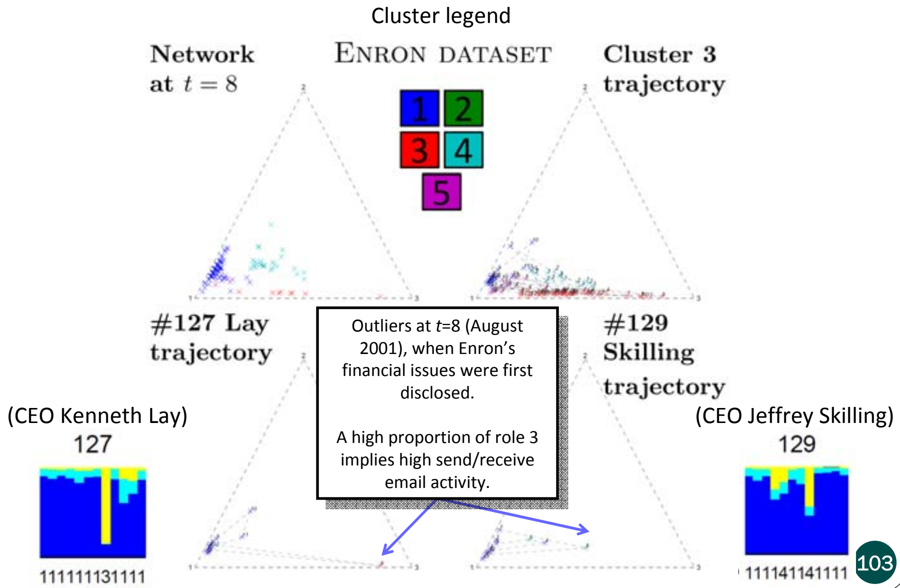
SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Enron Network: dynamic analysis

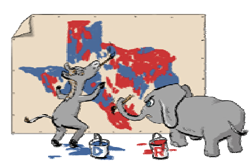
- Roles
  - The first role (blue) stands for inactivity.
  - The second role (cyan) corresponds to receivers.
  - Role 3 (yellow) correspond to both senders and receivers.
- Dynamic changes
  - Most actors are smooth over time. However,
- Individual activity
  - All people are dominated by role 1
  - Apart from role, most people have a little and no role 3 --- they only receive emails
  - A few send actively and receive, including:
    - *Mark Haedicke* (#65) (Managing Director of the Legal Department)
    - *Louise Kitchen* (#107) (President of Enron Online)
  - Yet some are more interesting …
    - *Kenneth Lay* (#127) (Chairman and CEO of Enron)

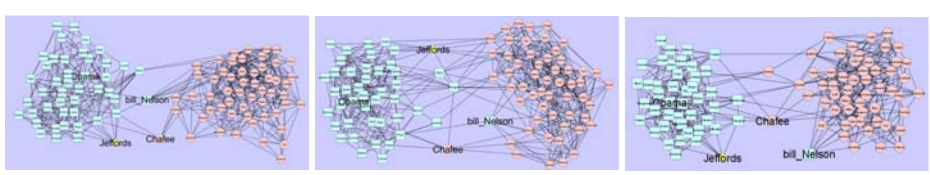


Enron Network: dynamic analysis



Case Study 2: The 109th congress





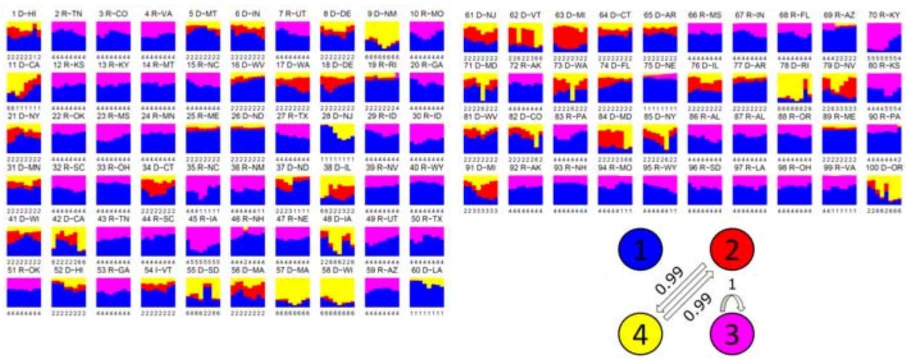
March 2005 | January 2006 | August 2006

US senator voting records

100 senators, 109th Congress (Jan 2005 – Dec 2006) in 8 epochs

Senate Network: role trajectories

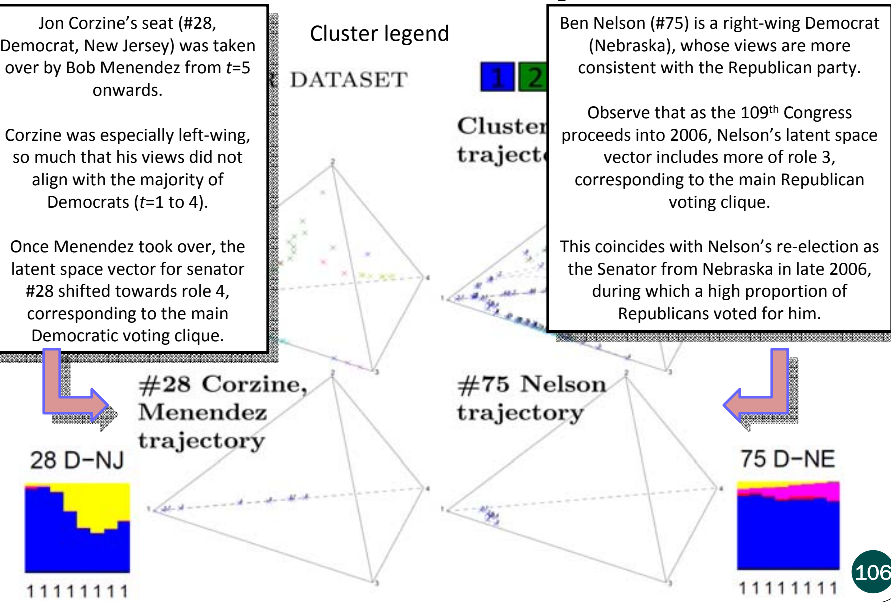
Voting data preprocessed into a network graph using (Kolar *et al*., 2008)



Colored bars: Estimated latent space vector
Numbers under bars: Estimated cluster
Letters beside actor index: Political party and State

Role Compatibility Matrix *B*
Role 1 = Passive, 2/4 = Democratic clique, 3 = Republican clique

Senate Network: role trajectories



Jon Corzine's seat (#28, Democrat, New Jersey) was taken over by Bob Menendez from t=5 onwards.

Corzine was especially left-wing, so much that his views did not align with the majority of Democrats (t=1 to 4).

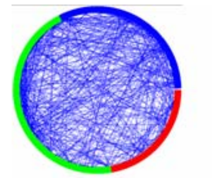
Once Menendez took over, the latent space vector for senator #28 shifted towards role 4, corresponding to the main Democratic voting clique.

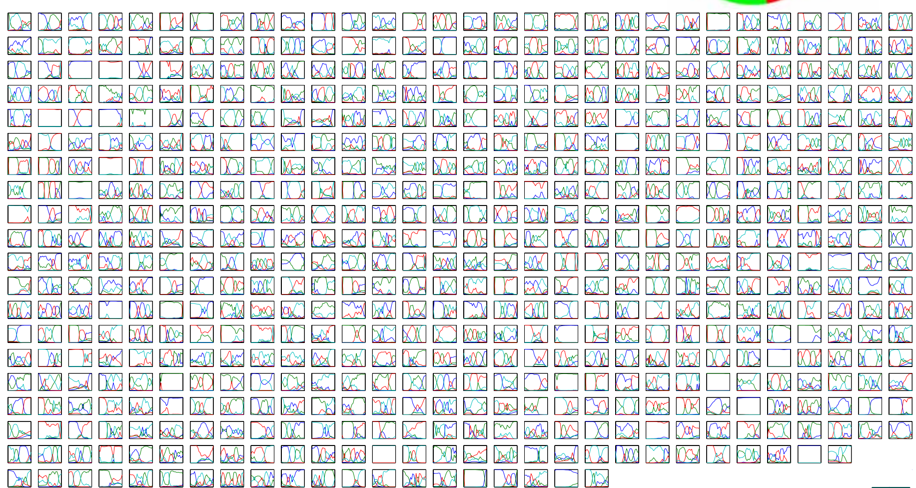
Ben Nelson (#75) is a right-wing Democrat (Nebraska), whose views are more consistent with the Republican party.

Observe that as the 109th Congress proceeds into 2006, Nelson's latent space vector includes more of role 3, corresponding to the main Republican voting clique.

This coincides with Nelson's re-election as the Senator from Nebraska in late 2006, during which a high proportion of Republicans voted for him.

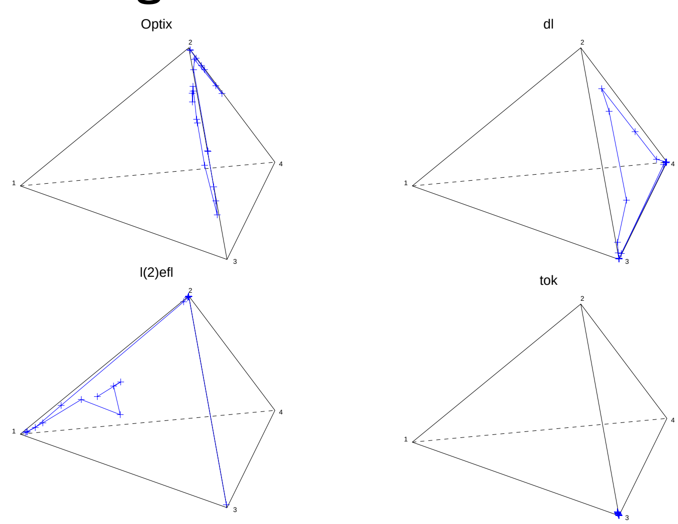
Drosophila Network: dynamic role vectors





6/29/2010

Drosophila Network: trajectories of selected genes



6/29/2010

Summary

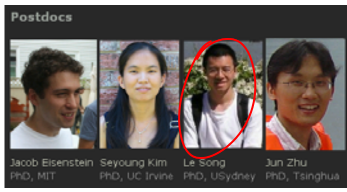
- Non-degenerate model for network evolution
- Efficient and Sparsistent algorithms for recovering latent time-evolving networks from nodal attributes
- Multi-role estimation from network topology
  - Roles undertaken by actors are not independent of each other; they can have internal dependency structures
  - An actor in the network can be fractionally assigned to multiple roles
  - Mixed memberships of actors vary temporally
- Visualization and analysis tool for dynamic networks
- Discovered interesting patterns from Email/Vote/Gene Nwks

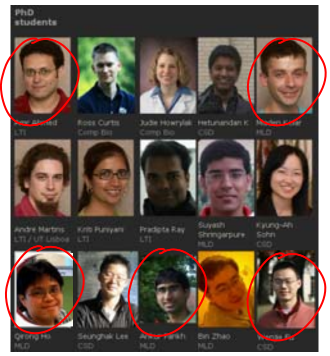
Discussion: future directions

- How about estimating time varying "causal" graphs, or general time-varying graphical models?
  - The *time-varying DBN* based on an *nonstationary auto-regressive model (NIPS 2009) entails 1st-order "Granger causality"*
  - Consistency proof is difficult (sample no longer conditionally independent), but still possible under assumption of local stationarity
  - Est. of arbitrary TV-GM remains an open problem in both algorithm and theory
- Web-scale inference/modeling
  - Scalability to million-node network problem remains hard
  - Are nodal/edge level inference still interesting in mega networks?
  - Predictive models for large graph evolution, community organization, information diffusion
- Socio-media modeling and prediction
  - Facebook, Twitter, Flicker: many interesting problems
  - Data integration: Graph + Text + image + …

Acknowledgement

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics





http://www.sailing.cs.cmu.edu/

Funding: