Time (and Space)-Varying Networks:

Reverse engineering rewiring social and genetic interactions

Eric Xing

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

Changing Social Networks

Corporativity, Antagonism, Cliques, ... over time?
Regulation of cell response to stimuli is paramount, but we can usually only measure (or compute) steady-state interactions.

Transcriptional interactions
- Chromatin IP
- Microarrays

Protein–protein interactions
- Protein coIP
- Yeast two-hybrid

Biochemical reactions
- Metabolic flux measurements

Biological regulations may be transient (in time and space) ...
Example II: Inflammatory Response in Endotoxinated Mice

- 100µg → 200µg → 300µg → 400µg → 600µg
  - Day 1 → Day 2 → Day 4 → Day 6 → Day 8

The Big-Picture Questions

- What pathway is active under certain extra-cellular stimuli or a certain point of a dynamic process?
- How does the network response to environmental perturbation and biomolecular/genetic therapy?
Current Practice …

Reverse engineer temporal/spatial-specific "rewiring" gene networks

Drosophila development
Modeling Time-Varying Graphs

- The temporal exponential graph models (Fan et al. ICML 2007)

**Transition Model:**

\[ P(A^t | A^{t-1}) = \frac{1}{Z(A^t, \theta)} \exp \left[ \sum_i \theta \Psi_i (A^t, A^{t-1}) \right] \]

**Emission Model:**

\[ P(X^t | A^t) = \exp \left\{ \theta \cdot \Psi(A^t, A^{t-1}) \right\} \]

"Dynamic" Potentials

\[ P(A^t | A^{t-1}) = \exp \{ \theta \cdot \Psi(A^t, A^{t-1}) - \ln Z(\theta, A^{t-1}) \} \]

- "Continuity":
  \[ \Psi_1 (A^t, A^{t-1}) = \sum_y (A_{iy}^t A_{iy}^{t-1} + (1 - A_{iy}^t)(1 - A_{iy}^{t-1})) \]

- "Reciprocity":
  \[ \Psi_2 (A^t, A^{t-1}) = \sum_y A_{iy}^t A_{iy}^{t-1} \]

- "Transitivity":
  \[ \Psi_3 (A^t, A^{t-1}) = \sum_{ykl} \frac{A_{iy}^t A_{ik}^{t-1} A_{lj}^{t-1}}{\sum_{ykl} A_{iy}^t A_{ik}^{t-1} A_{lj}^{t-1}} \]

- "Density":
  \[ \Psi_4 (A^t, A^{t-1}) = \sum_y A_{iy}^t \]

Network Workshop @ Santa Fe Inst. 12/4/2008
A tERGM is non-degenerate

- **Theorem 1**: when the transition distribution factors over the edges, a tERGM is non-degenerate:
  \[
  H(A') \geq \sum_{ij} H(A_{ij}'|A_{ij}^{t-1}) \geq p \ln \frac{1}{p} + (1 - p) \ln \frac{1}{1 - p}
  \]
  
- → Maximum likelihood estimator exists!
  (actually should not be taken for granted for arbitrary models, be careful!)

What’s it good for?

- Hypothesis Testing
- Data Exploration (e.g., node-classification)
- Foundation for Learning
  - **Current approaches**:
    - Time invariant structures
    - Global optimality?
    - Consistency guarantee?
Inferring Rewiring Biological Networks

- Networks rewire over discrete time-steps

Inferring Rewiring Biological Networks

- Networks rewire over epochs
Modeling Time-Varying Graphs

- The temporal exponential graph models (Fan et al. ICML 2007)

\[ P(A^{t+1} | A^{t}) = \frac{1}{Z(A^{t}, \theta)} \exp \left[ \sum_{s} \theta_{s} \psi(s, A^{t}, A^{t-1}) \right] \]

Transition Model:

\[ \lambda_{ij} = \sum_{t} \exp \left( \theta \Phi(A_{ij}, A_{ij}^{t}) \right) \]

Emission Model:

\[ P(X^{t} | A^{t}) = \sum_{A^{t}} \exp \left( \theta \Psi(X^{t}, A^{t}) \right) \]

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

Inference (1)

- Gibbs sampling:

  Need to evaluate the log-odds

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

  \[ = \log \frac{P(A_{ij} = 1 | A_{ij}^{t-1}, A_{ij}^{t+1}) + P(A_{ij} = 0 | A_{ij}^{t-1}, A_{ij}^{t+1})}{P(A_{ij} = 1 | A_{ij}^{t-1}, A_{ij}^{t+1}) + P(A_{ij} = 0 | A_{ij}^{t-1}, A_{ij}^{t+1})} \]

- Difficulty: Evaluate the ratio of Partition function \( Z(A^{t}) = \sum_{e} \exp(\theta \Phi(A, A')) \)

- So far scale to \( \sim 20 \) genes
Results on Simulated Data

- F1 scores on different parameter settings (varying $\theta_2, \eta$)

$(\theta_1 = -0.5, \theta_3 = 4, D = 5, 100k$ iterations of Gibbs sampling, $10$ repetitions)

$\eta = 0.5$  
$\eta = 1.0$  
$\eta = 1.5$

Graph Regression

Markov Random Fields

Graphical Gaussian Model

contains both the structure and parameters

\[ X' \sim \frac{1}{Z} \exp\left\{ \sum_i \theta_i x_i' + \sum_{i<j} \psi_{ij} x_i' x_j' \right\} \]

\[ X' \sim \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma'|} \exp\left\{ -\frac{1}{2} x'(\Sigma')^{-1} x' \right\} \]

Lasso:

\[ \hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{T} l(\theta) + \lambda \| \theta \| \]
Graph Regression

\[ \Theta^f \equiv (\Sigma^f)^{-1} \]
contains both the structure and parameters
Inference II

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

$$\hat{\theta}_i^1, \ldots, \hat{\theta}_i^T = \arg\min_{\theta_i^1, \ldots, \theta_i^T} \sum_{t=1}^T l_{\text{avg}}(\theta_i^t) + \lambda_1 \sum_{t=1}^T \| \theta_{i-1}^t \|_1$$

$$+ \lambda_2 \sum_{t=2}^T \| \theta_i^t - \theta_i^{t-1} \|_q,$$

where $l_{\text{avg}}(\theta_i^t) = \frac{1}{N_i} \sum_{d=1}^{N_i} \log P(x_i^d | x_{i-1}^d, \theta_i^t)$.

- Constrained convex optimization
  - Now scale to ~5000 nodes, how about 20K+?

Inference III

- Kernel Weighting (Kolar and Xing, 2008):

$$\hat{\theta}_i^1, \ldots, \hat{\theta}_i^T = \arg\min_{\theta_i^1, \ldots, \theta_i^T} \sum_{t=1}^T l_w(\theta_i^t) + \lambda_1 \sum_{t=1}^T \| \theta_i^{t-1} \|_1$$

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

- Constrained convex optimization
  - Could scale to ~$10^4$ genes, but under stronger smoothness assumptions
Consistency

- **Theorem 2**: for the kernel weighting method, under certain verifiable conditions (omitted here for simplicity):

\[
P \left[ \hat{G}(\lambda_n, t_0) \neq G_{t_0} \right] = O \left( \exp \left( -C' \frac{nh_n}{s_n^3} + C'' \log p \right) \right) \to 0
\]

- Consistency for TESLA is yet to be proven! (very hard!)
**F-Measure for the AND combination**

- **Dynamic**
- **Static**

**Number of samples per Epoch**

- **Over recovery:** The more samples in each epoch, the more all transient edges get recovered as if they appeared along all epochs.

- **Under recovery:** Few samples per epoch misses transient edges (not enough evidences for them when compared globally).

---

**NIPS academic social network**

- **1987**
- **1988**
- **1998**
- **1999**
Transient Interaction

Static Versus Dynamic
Evolution of Network Signatures

Evolution of Degree Distribution

Evolution of Edges

Evolution of Clustering

Transient Subgraph
Future Work

- Analyzing time-space data in biological processes
  - Drosophila life cycle
  - Breast cancer progression and reversal
  - Inflammatory response in endotoxinated mice
- Other dynamic behaviors of networks
  - Differentiation: tree of networks
  - Detection of sudden changes
  - Active learning – when to get more samples
- Open theoretical issues
  - Consistence (pattern, value, …)
  - Confidence
  - Stability
  - Sample complexity

Acknowledgement

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