SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Estimating Time-Varying Networks

Eric Xing

epxing@cs.cmu.edu

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

with Mladen Kolar, Le Song and Amr Ahmed

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Outline

- Background
- Motivation and challenge
- Algorithms
  - Keller
  - Tesla
  - Formal analysis: asymptotic consistency
- Empirical analysis
  - Senate network
  - Drosophila network
- Discussions

SAILING LAB

Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Background

- Classical asymptotic theory in statistical inference:
  - number of observations $n \longrightarrow +\infty$
  - model dimension *p* is fixed
- Problems in real world, e.g., computational biology:
  - models are large, and observations are scarce and costly
  - usually $p = \Theta(n) \text{ or } p >> n$
- Complexity regularization is required to avoid curse of dimensionality, e.g. *sparsity*

SAILING LAB

Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Background, cont'd

- Recently, lots of methods
  - Lasso
  - Elastic net
  - Dantzig selector
  - Graphical Lasso
  - Nonnegative Garrote Estimator
  - …
- **Assumption:**
  - data is independent and identically-distributed

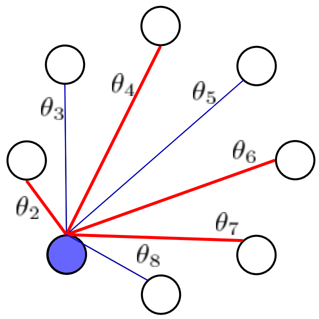
Graph Regression

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

Markov Random Fields

$$\mathbf{X}^t \sim \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma^t|^{\frac{1}{2}}}\exp\{-\frac{1}{2}\mathbf{x}'^t(\Sigma^t)^{-1}\mathbf{x}^t\}$$

Graphical Gaussian Model



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

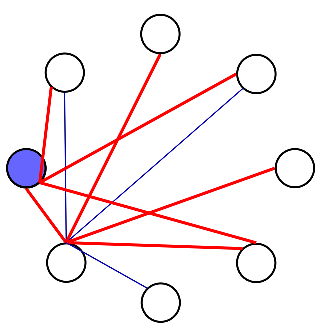
contains both the structure and parameters

Neighborhood selection

Lasso:

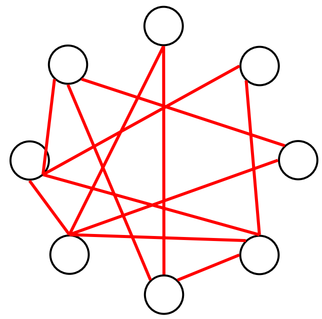
$$\hat{\theta} = \arg\min_{\theta} \sum_{t=1}^{T} l(\theta) + \lambda_1 \| \theta \|_1$$

Graph Regression



SAILING LAB

Graph Regression

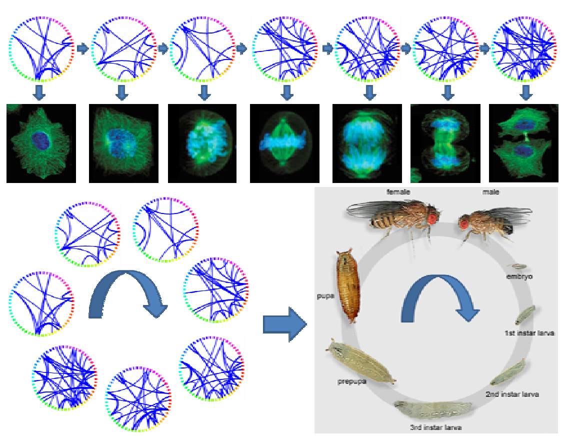


It can be shown that:
given iid samples, and under several technical conditions (e.g., "irrepresentable"), the recovered structured is "**sparsistent**" even when p >> n

SAILING LAB

Our problem

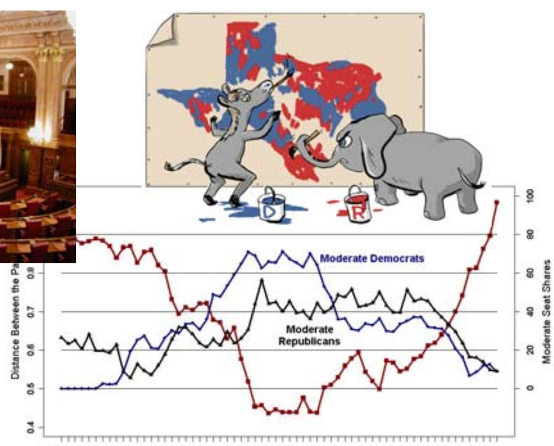
- Inferring Time-Varying Networks



Outline

- Background
- Motivation and challenge
- Algorithms
  - Keller
  - Tesla
  - Formal analysis: asymptotic consistency
- Empirical analysis
  - Senate network
  - Drosophila network
- Discussions

Changing Social Networks



Corporativity,

Antagonism,

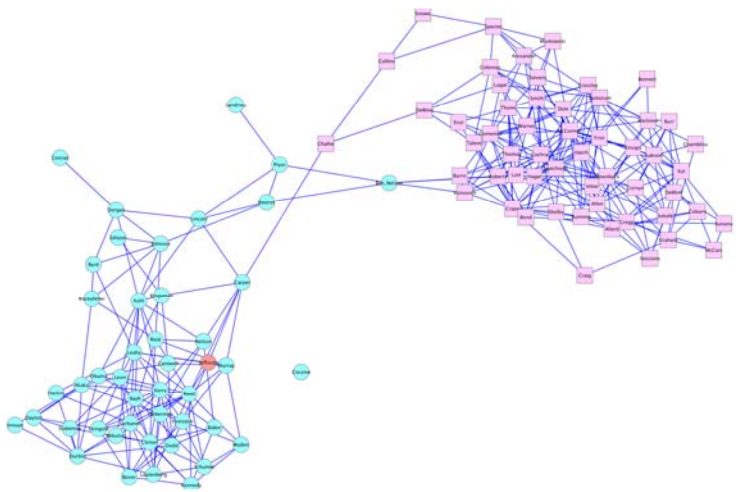
Cliques,

…

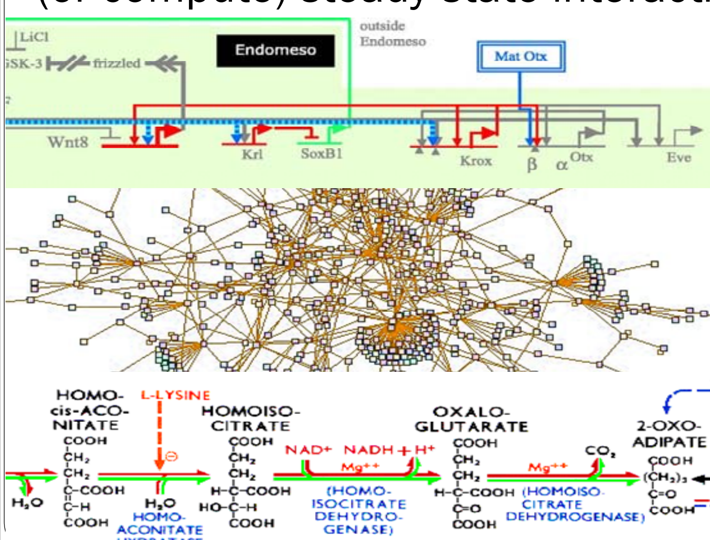
over time?

SAILING LAB

Senate Voting Records



Regulation of cell response to stimuli is paramount, but we can usually only measure (or compute) steady-state interactions

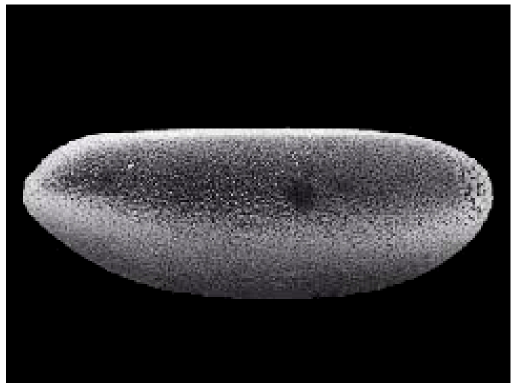


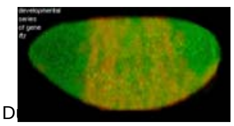
Transcriptional interactions
- Chromatin IP
- Microarrays

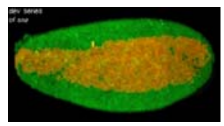
Protein—protein interactions
- Protein coIP
- Yeast two-hybrid

Biochemical reactions
- Metabolic flux measurements

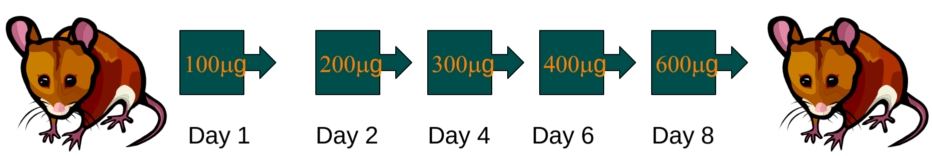
Biological regulations may be transient (in time and space) …

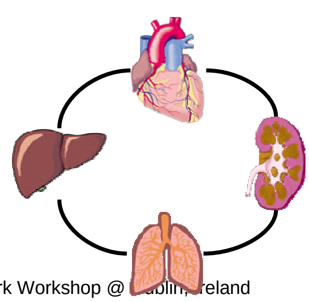






Example II: Inflammatory Response in Endotoxinated Mice

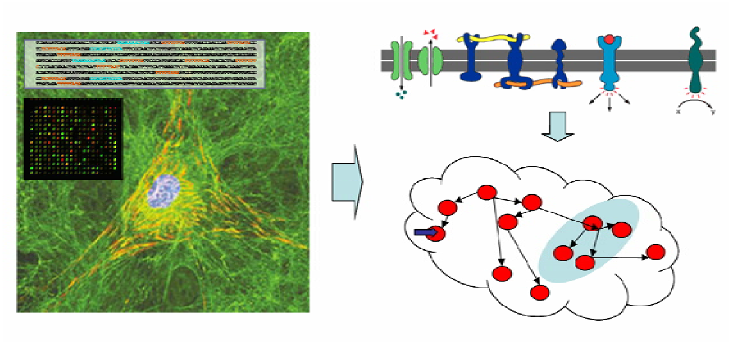




SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

The Big-Picture Questions

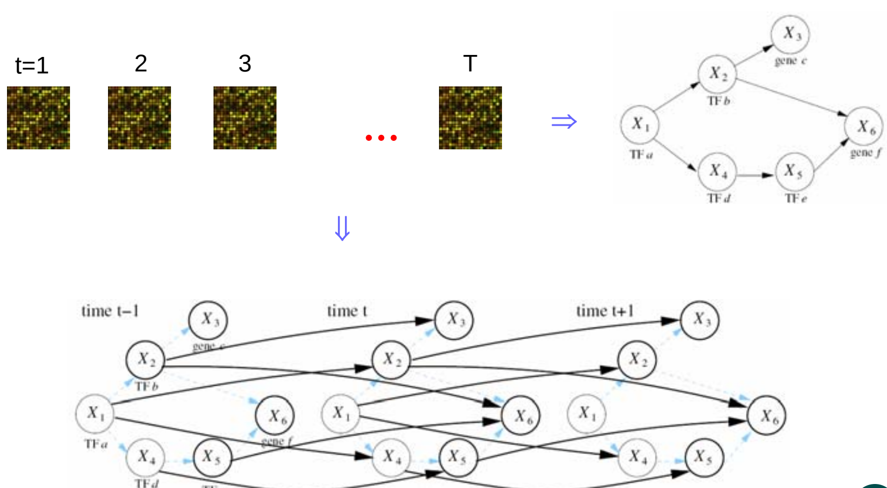
- What pathway is active under certain extra-cellular stimuli or a certain point of a dynamic process?



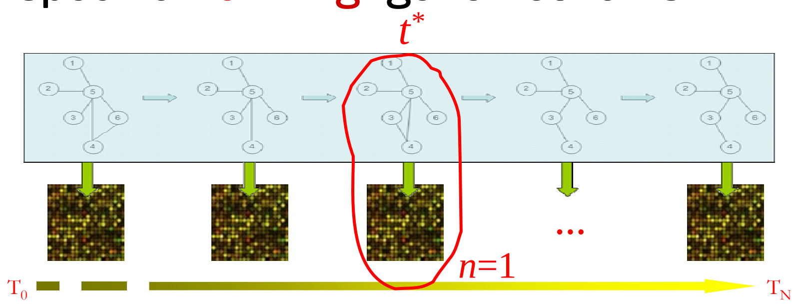
- How does the network response to environmental perturbation and biomolecular/genetic therapy?

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Current Practice …



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



Drosophila development

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 & Integrative Genomics

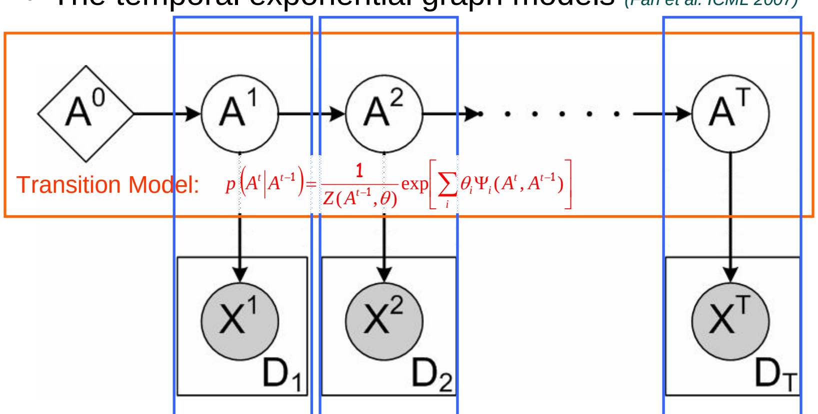
Outline

- Background
- Motivation and challenge
- Algorithms
  - Keller
  - Tesla
  - Formal analysis: asymptotic consistency
- Empirical analysis
  - Senate network
  - Drosophila network
- Discussions

SAILING LAB
Laboratory for Statistical Artificial Intelligence & Integrative Genomics

Modeling Time-Varying Graphs

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



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

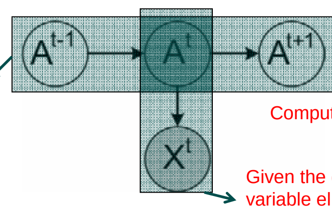
Inference (0)

$P(\text{Network}|\text{Data})$?

- Gibbs sampling:
  - 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 genes



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

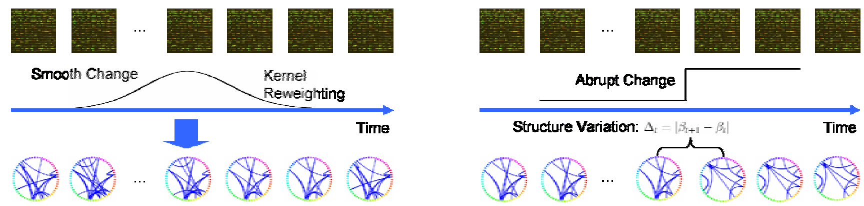
Computation is non-trivial

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

Problem

- Computational cost!
- Global optimality?
- Consistency guarantee?

Two Scenarios



Smoothly evolving graphs

Abruptly evolving graphs

Inference I

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

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

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

- Constrained convex optimization
  - Estimate time-specific one by one
  - Could scale to ~10^4 genes, but under stronger smoothness assumptions

Problem formulation

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

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

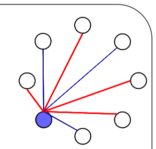
$$\mathbb{P}_{\theta^t}(X) = \frac{1}{Z(\theta^t)} \exp\left(\sum_{(u,v)\in E^t} \theta^t_{uv} x_u x_v \right)$$

- Idea: maximize the likelihood to obtain the structure

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

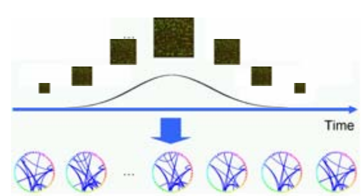
- Calculation of likelihood: intractable (because of the Z)

Algorithm - neighborhood selection



- Conditional likelihood

$$\mathbb{P}_{\theta^t}(x^t_u | x^t_{\backslash u}) = \frac{\exp\left(2x^t_u \left\langle \theta^t_{\backslash u}, x^t_{\backslash u}\right\rangle\right)}{\exp\left(2x^t_u \left\langle \theta^t_{\backslash u}, x^t_{\backslash u}\right\rangle\right) + 1},$$

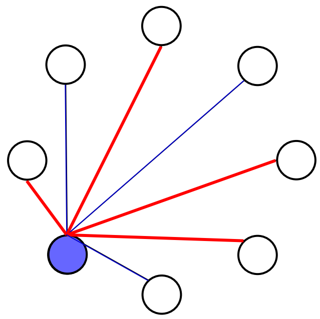


- Neighborhood: $S(x_u) = \{j \mid \theta^t_{u,j} \neq 0\}$
- Estimate at $t^* \in [0, 1]$

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

Where $\gamma(\theta^t; x^t) = \log \mathbb{P}_{\theta^t}(x^t_u | x^t_{\backslash u})$ and $w_t(t^*) = \frac{K_{h_n}(t - t^*)}{\sum_{t'\in\mathcal{T}^n} K_{h_n}(t' - t^*)}$

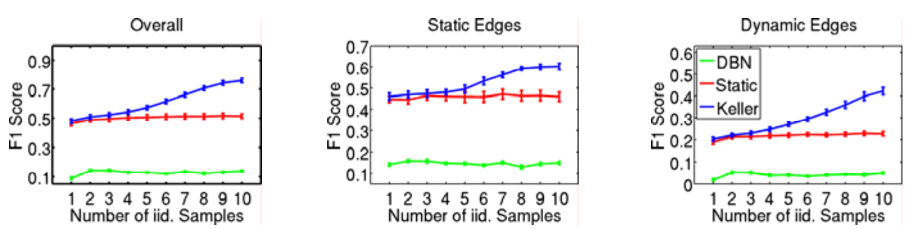
Graph Regression



Lasso

$$\min_{\theta \in \mathbb{R}^{p-1}} \left\{ -\sum_t w_t \gamma(\theta; x^t) + \lambda_n \|\theta\|_1 \right\}$$

Synthetic data



SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Structural consistency of KELLER

- When does the method succeed in recovering the unknown structure?
- Under which conditions on $(\mathrm{n},\ \mathrm{p},\ \mathrm{s},\ \theta_{\min})$ can we estimate the structure consistently?

$$s = \max_u \max_t |S_u^t|, \quad \theta_{\min} = \min_{e \in E} \max |\theta_e^t|$$

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Assumptions

- Define:

$$Q_u^t := \mathbb{E}\left[\nabla^2 \log \mathbb{P}_{\theta^t}[X_u | X_{\setminus u}]\right], \qquad \forall u \in V$$

$$\Sigma_u^t := \mathbb{E}\left[X_{\setminus u}^t {X_{\setminus u}^t}^T\right], \qquad \forall u \in V$$

$$s = \max_u \max_t |S_u^t|, \quad \theta_{\min} = \min_{e \in E} \max |\theta_e^t|$$

- 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

$$\left\| Q_{S^c S}^{t^*} (Q_{SS}^{t^*})^{-1} \right\|_\infty \leq 1 - \alpha, \quad \forall t^* \in [0,1]$$

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Assumptions

- A3: Smoothness Condition

$$\Sigma^{t^*} = [\sigma_{uv}(t^*)]$$

$$\max_{u,v} \sup_{t^*} |\sigma'_{uv}(t^*)| \leq A_0, \quad \max_{u,v} \sup_{t^*} |\sigma''_{uv}(t^*)| \leq A$$
$$\max_{u,v} \sup_{t^*} |\theta'_{uv}(t^*)| \leq B_0, \quad \max_{u,v} \sup_{t^*} |\theta''_{uv}(t^*)| \leq B$$

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Theorem

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

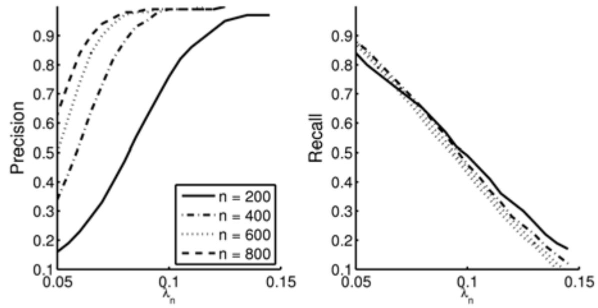
1. $h_n = \mathcal{O}(n^{-\frac{1}{3}})$
2. $s_n h_n = o(1)$,
3. $\frac{s_n^3 \log p_n}{n h_n} = o(1)$
4. $\lambda_1 = \mathcal{O}(\sqrt{\frac{\log p}{n h_n}})$
5. $\theta^*_{\min} = \Omega(\sqrt{\frac{s_n \log p_n}{n h_n}})$

then

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

Experiments

- Experiments on synthetic data p = 50, 100 edges. Every 100 discrete time steps we add and remove 20 edges.



Precision and recall as a function of the regularization parameter

Inference II

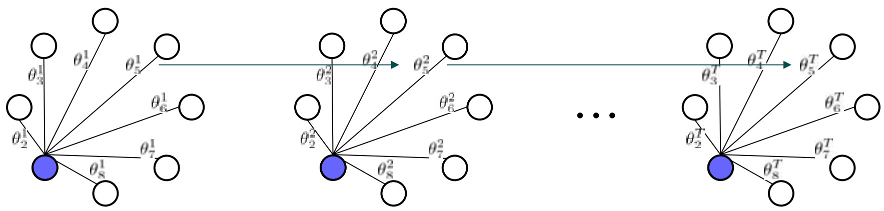
- **TESLA**: Temporally Smoothed L_1-regularized logistic regression

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

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

- Constrained convex optimization
  - Scale to ~5000 nodes, does not need smoothness assumption, can accommodate abrupt changes.

Temporally Smoothed Graph Regression



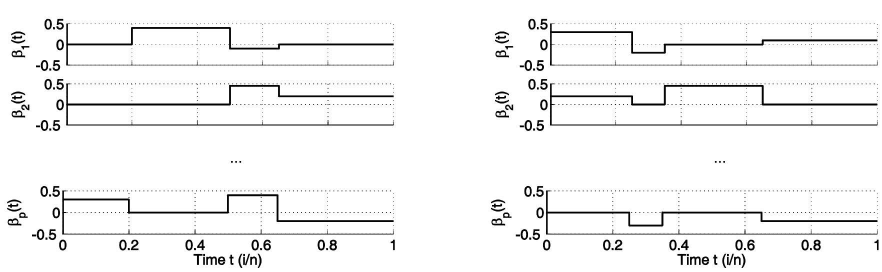
TESLA:

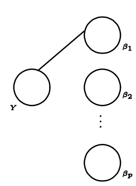
$$\min_{\substack{\theta_i^1,\dots,\theta_i^T \\ \mathbf{u}_i^1,\dots,\mathbf{u}_i^T;\mathbf{v}_i^2,\dots,\mathbf{v}_i^T}} \sum_{t=1}^{T} \ell(\mathbf{x}^t;\theta_i^t) + \lambda_1 \sum_{t=1}^{T} \mathbf{1}'\mathbf{u}_i^t + \lambda_2 \sum_{t=2}^{T} \mathbf{1}'\mathbf{v}_i^t$$

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

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

Coefficients as functions





Modified estimation procedure

- 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

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

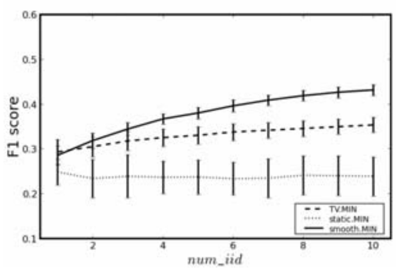
Structural Consistency of TESLA

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**

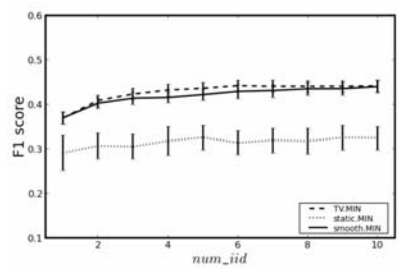
II. Then it can be further shown that, by applying Lasso on (**), **the neighborhood of each node on each of the estimated blocks consistently**

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

Comparison of KELLER and TESLA



Smoothly varying

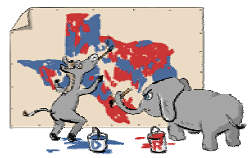


Abruptly varying

Outline

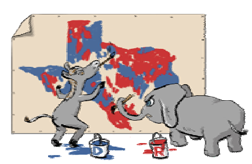
- Background
- Motivation and challenge
- Algorithms
  - Keller
  - Tesla
  - Formal analysis: asymptotic consistency
- Empirical analysis
  - Senate network
  - Drosophila network
- Discussions

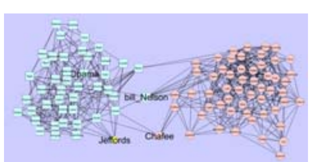
Senate network – 109th congress



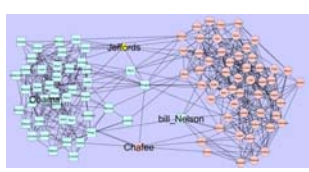
- Voting records from 109th congress (2005 - 2006)
- There are 100 senators whose votes were recorded on the 542 bills, each vote is a binary outcome
- Estimating parameters:
  - KELLER: bandwidth parameter to be h_n = 0.174, and the penalty parameter λ_1 = 0.195
  - TESLA: λ_1 = 0.24 and λ_2 = 0.28

Senate network – 109th congress

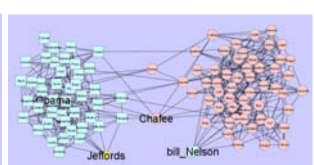




March 2005

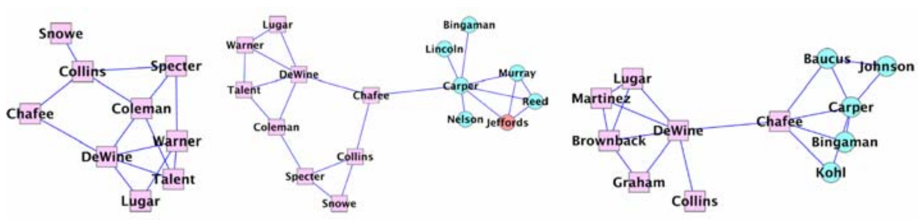


January 2006



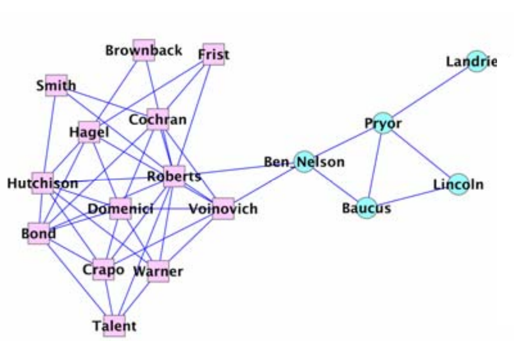
August 2006

Senator Chafee

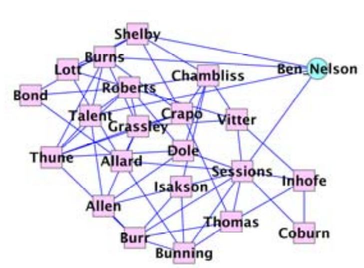


(a) t = 0.1 (b) t = 0.4 (c) t = 0.8

Senator Ben Nelson



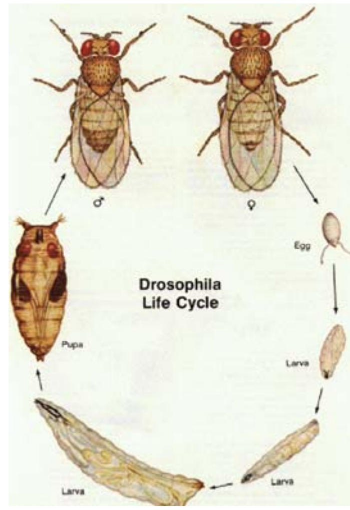
T=0.2

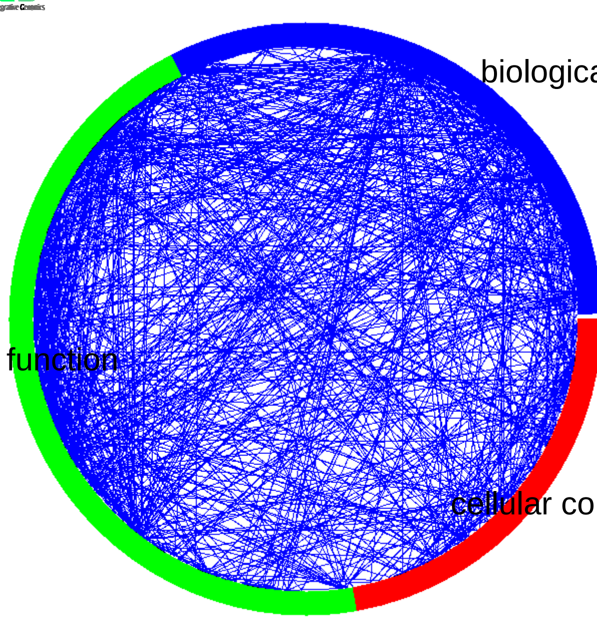


T=0.8

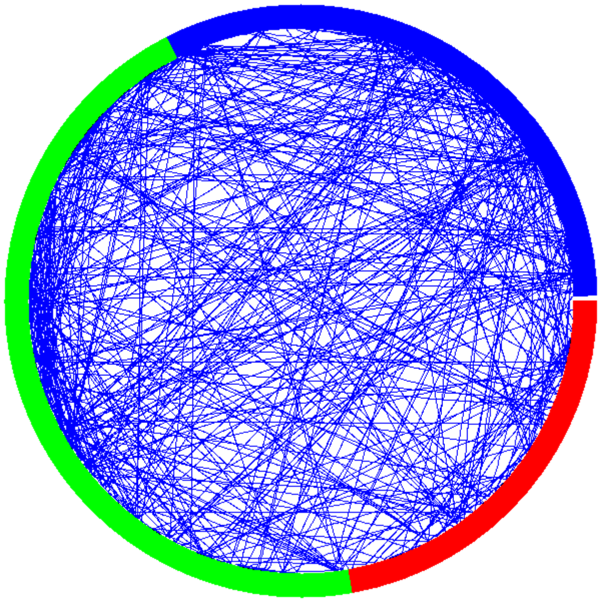
Drosophila life cycle

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



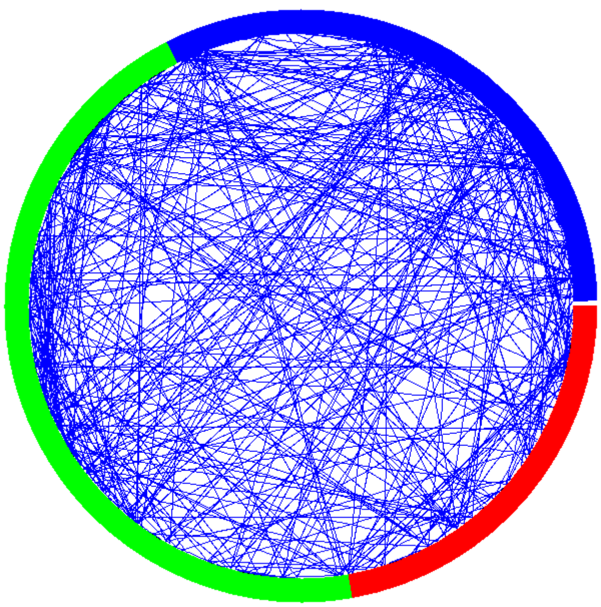


SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics



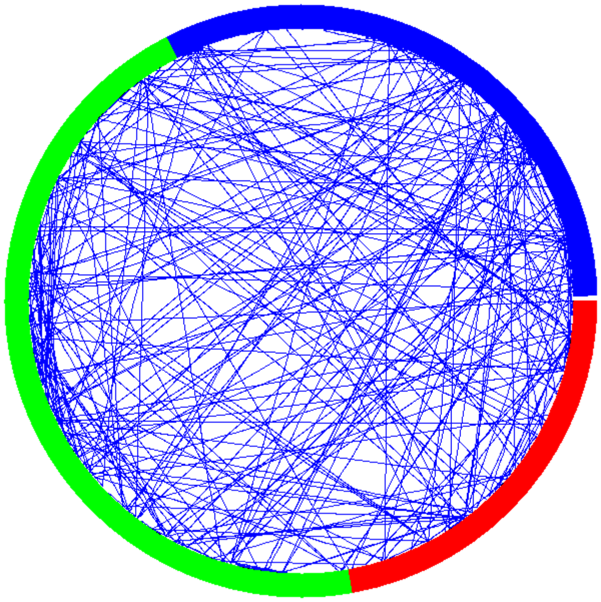
T=2

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics



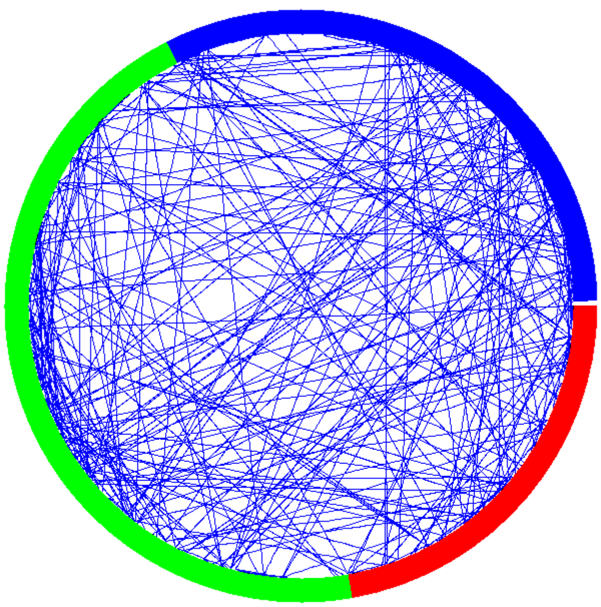
T=3

SAILING LAB
Laboratory for Statistical Artificial Intelligence & Integrative Genomics

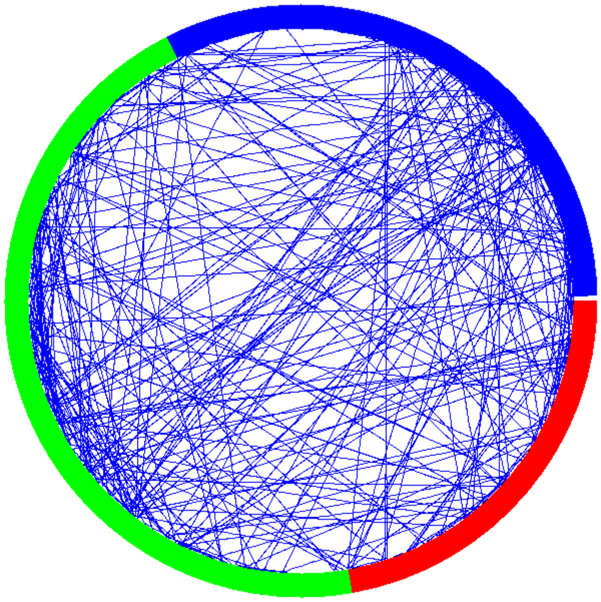


T=4

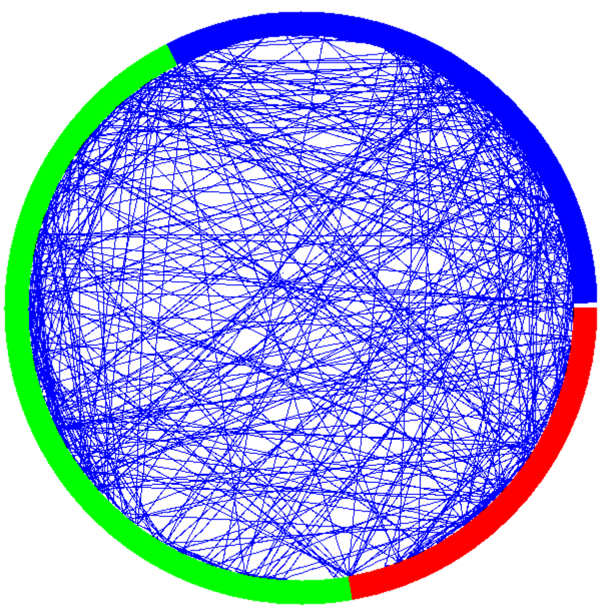
SAILING LAB
Laboratory for Statistical Artificial Intelligence & Integrative Genomics



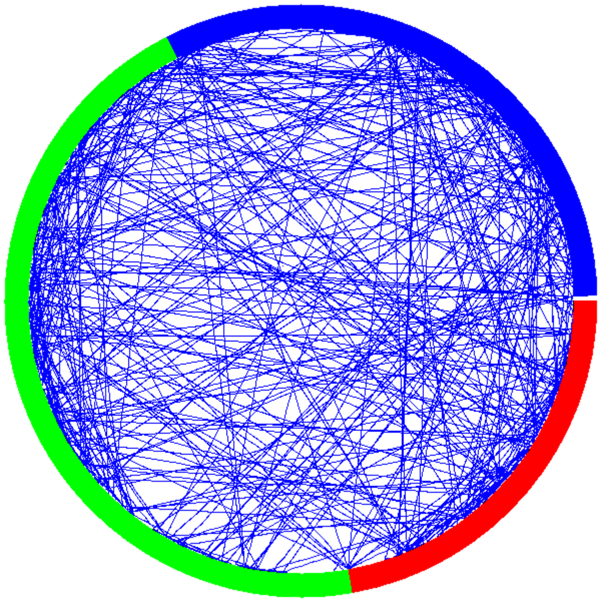
T=5

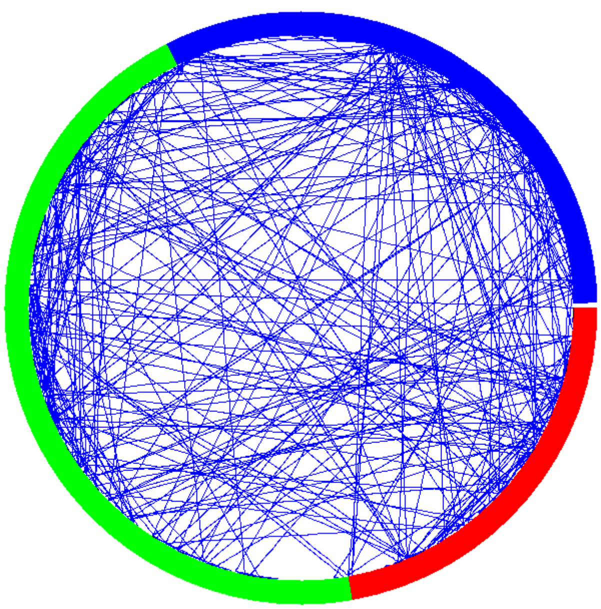


T=6

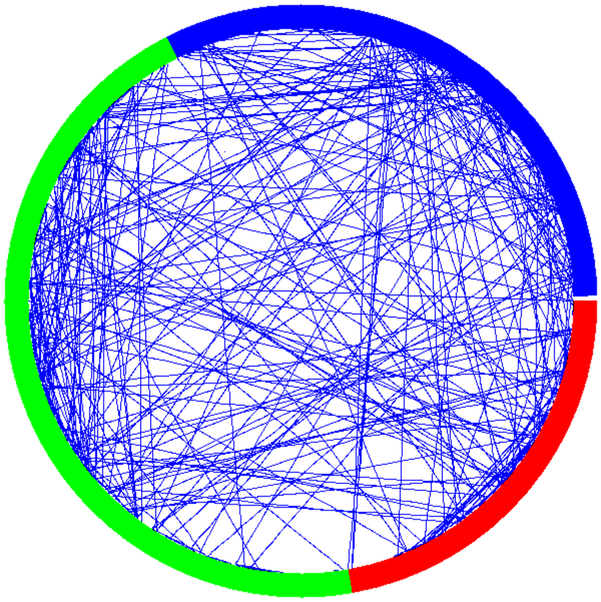


T=7


T=8

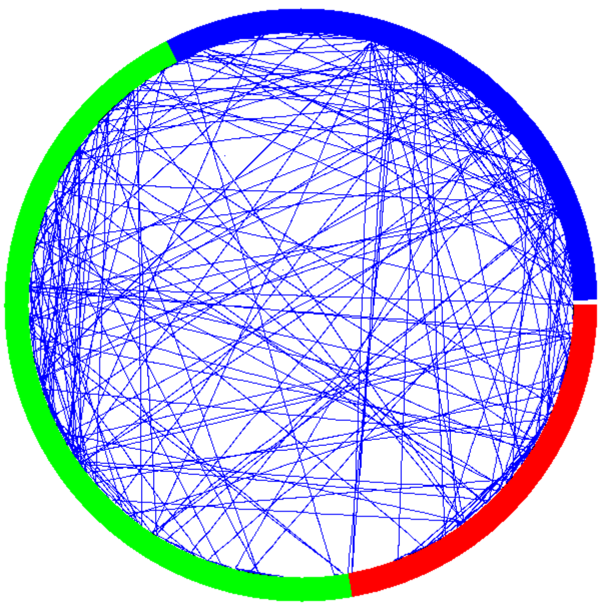

T=9

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics



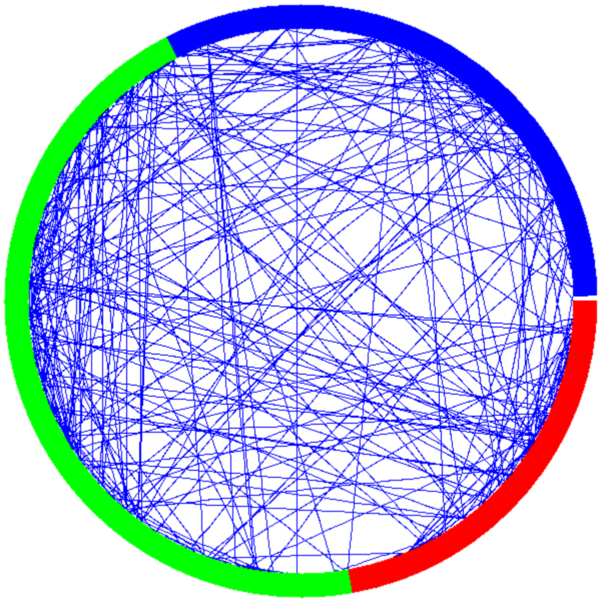
T=10

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics



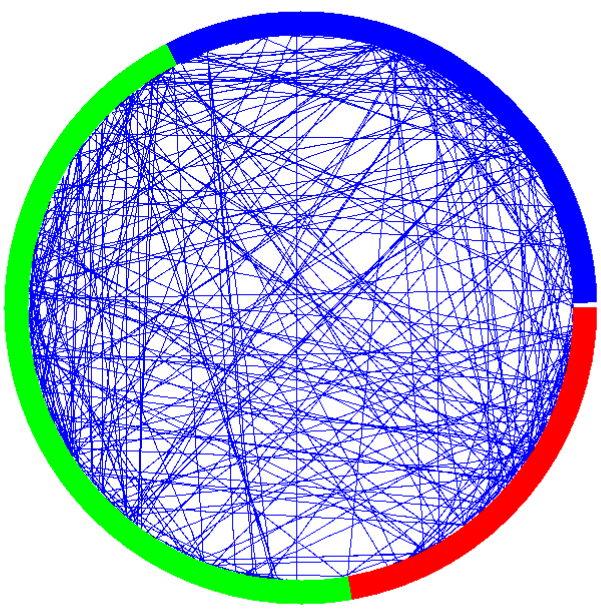
T=11

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics



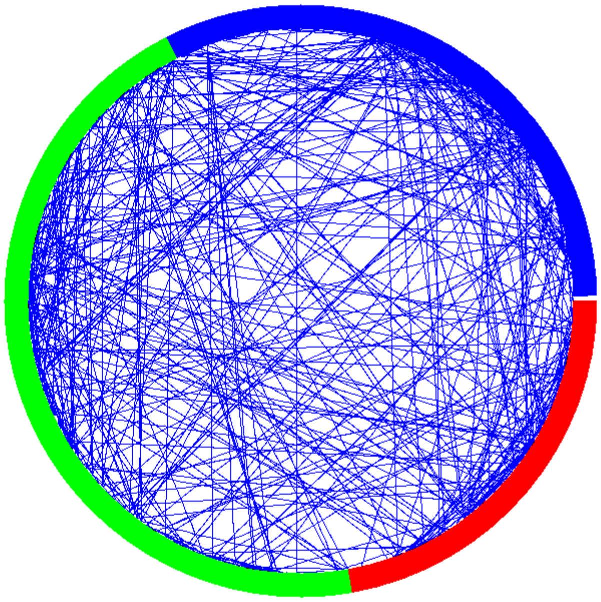
T=12

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics



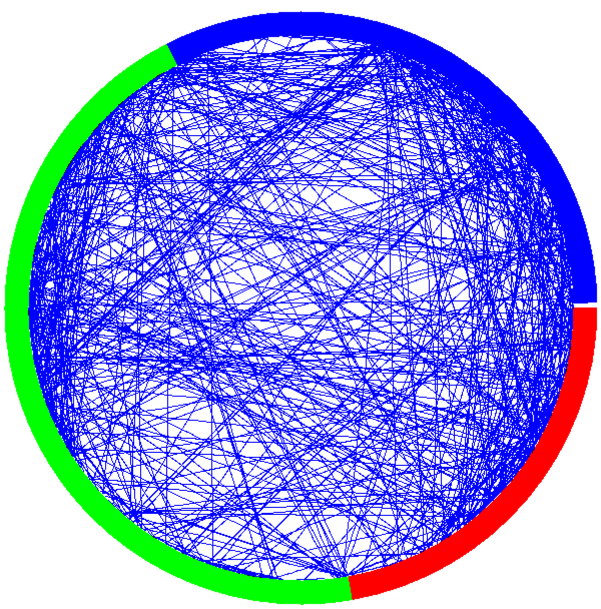
T=13

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics



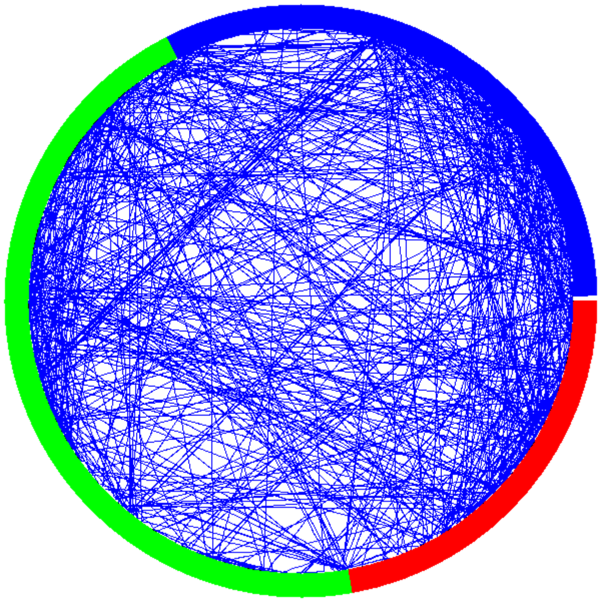
T=14

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics



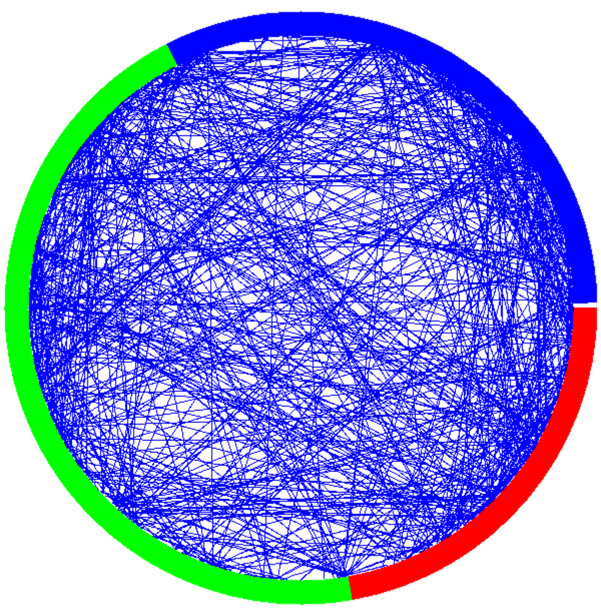
T=15

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

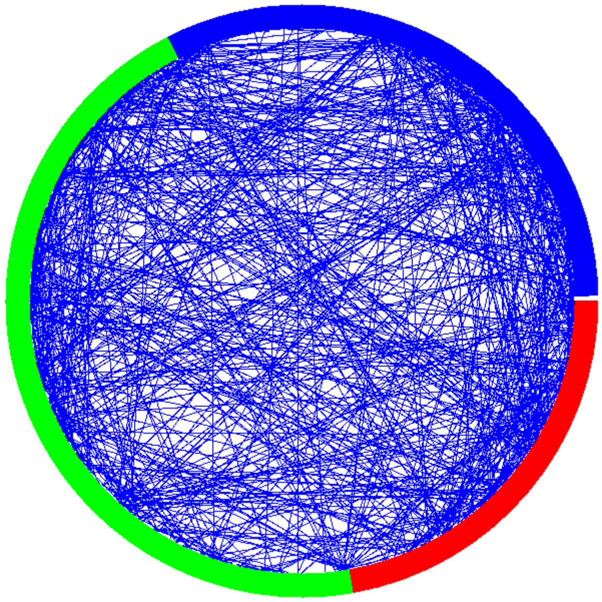


T=16

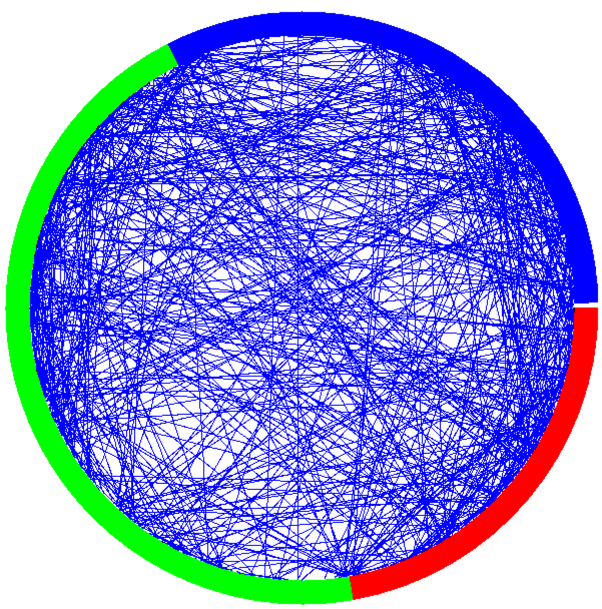
SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics



T=17

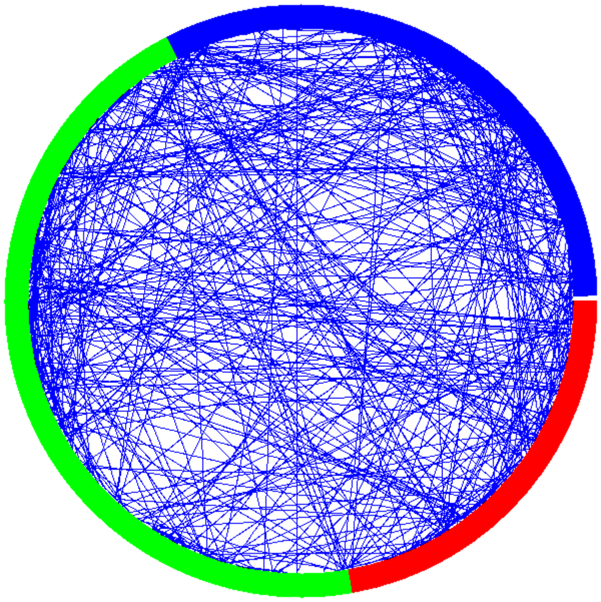


T=18



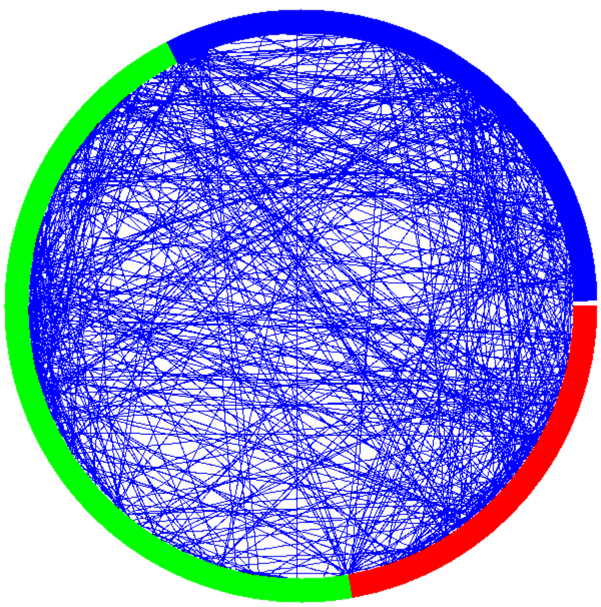
T=19

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics



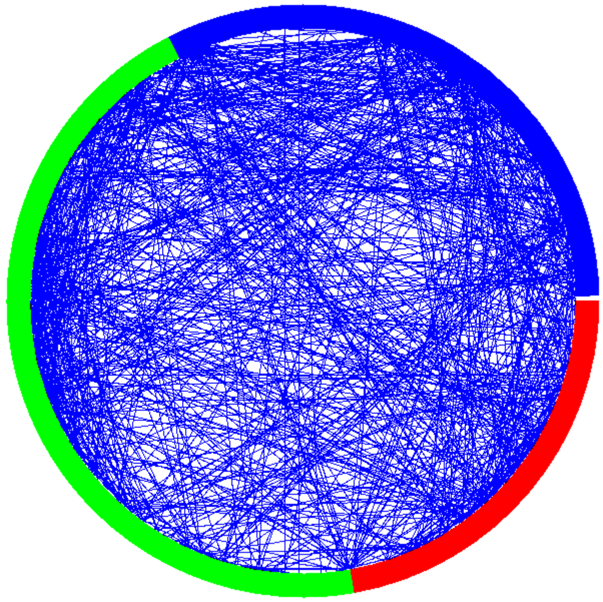
T=20

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics



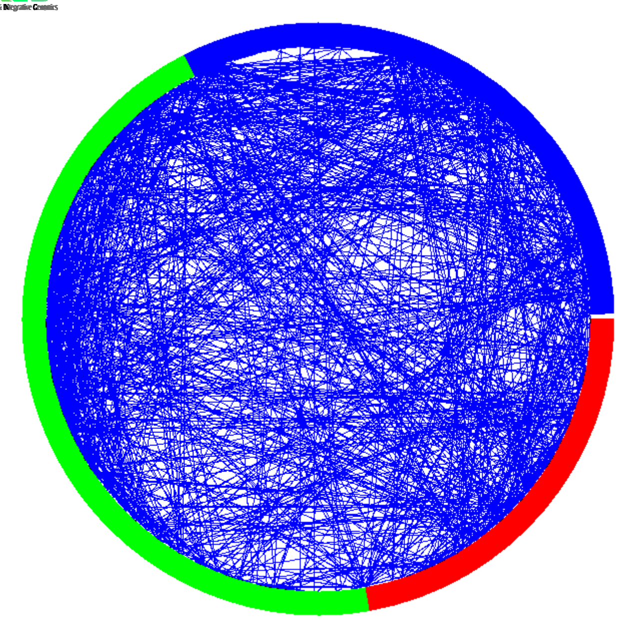
T=21

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics



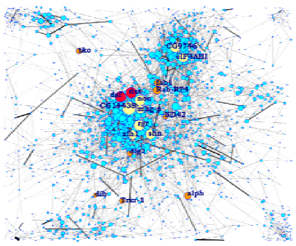
T=22

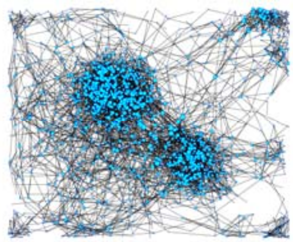
SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

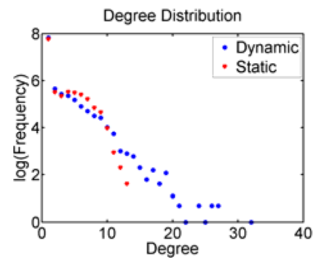


T=23

Static Versus Dynamic

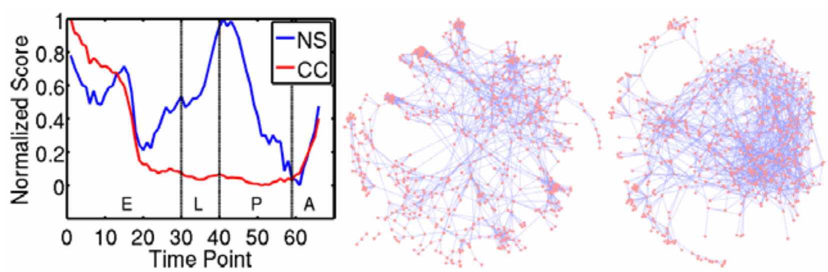






Network Statistics

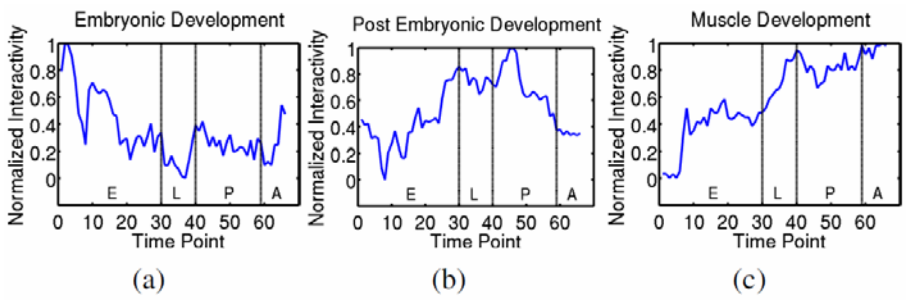
- Network size (NS) and local clustering coefficient (CC) follow different trends



(a) Network statistics (b) Embryonic stage (c) Pupal stage

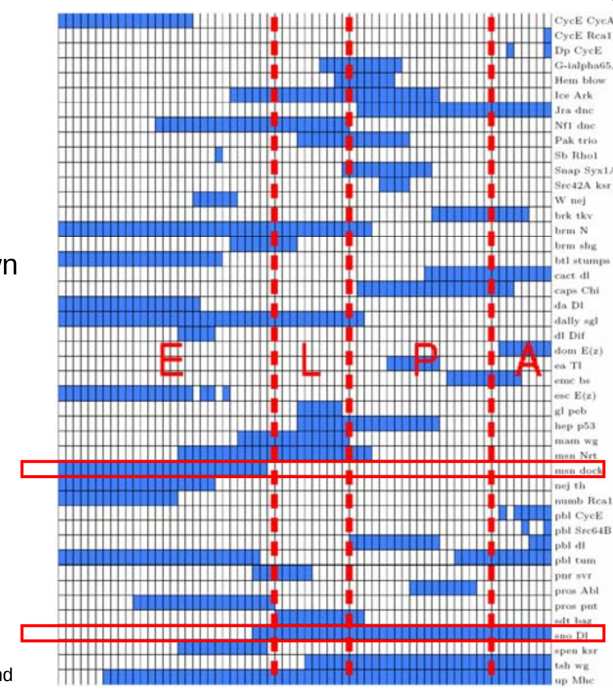
Stage-specific Gene Sets

- Stage-specific genes show stage-specific interactivity



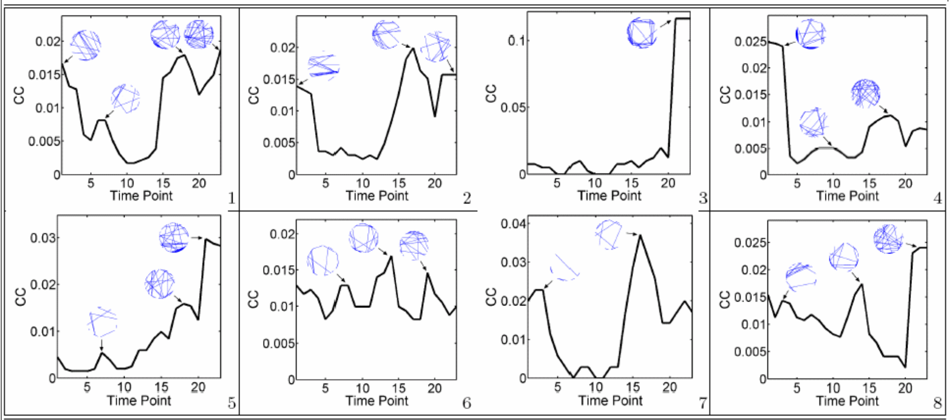
Known Gene Interactions

- Visualizing Time-span of known gene interactions



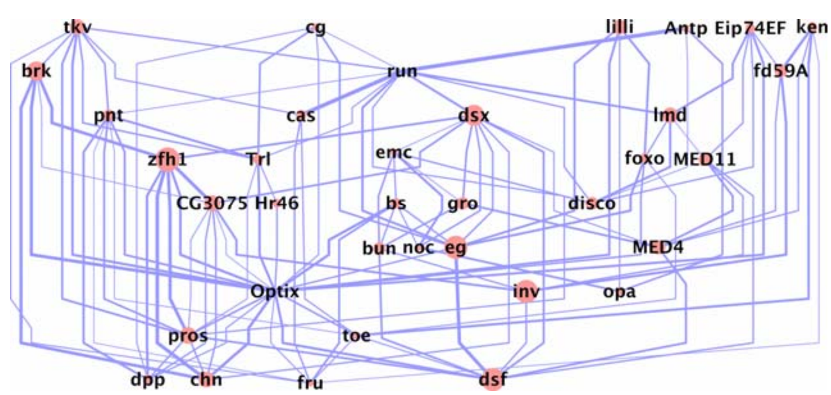
SAILING LAB
Laboratory for Statistical Artificial Intelligence & Integrative Genomics

Transient Subgraph



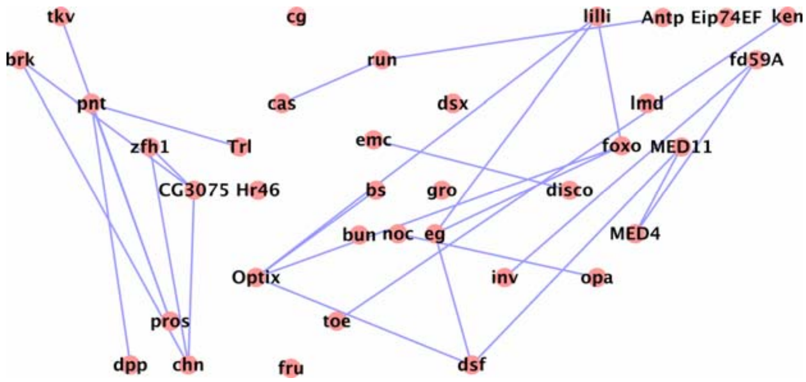
SAILING LAB
Laboratory for Statistical Artificial Intelligence & Integrative Genomics

Transcriptional Factor Cascade



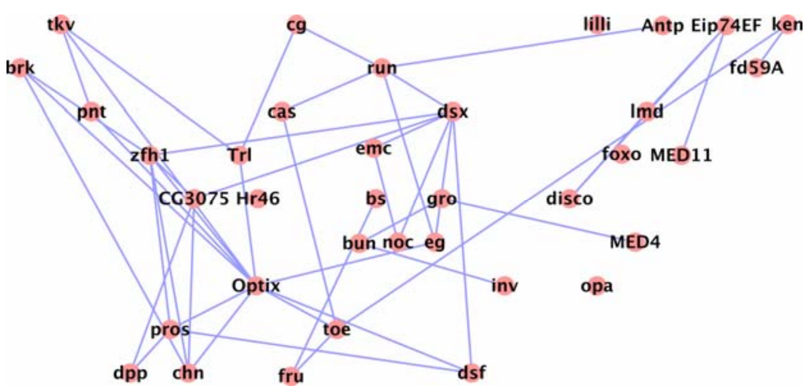
SAILING LAB

TF Cascade – mid-embryonic stage

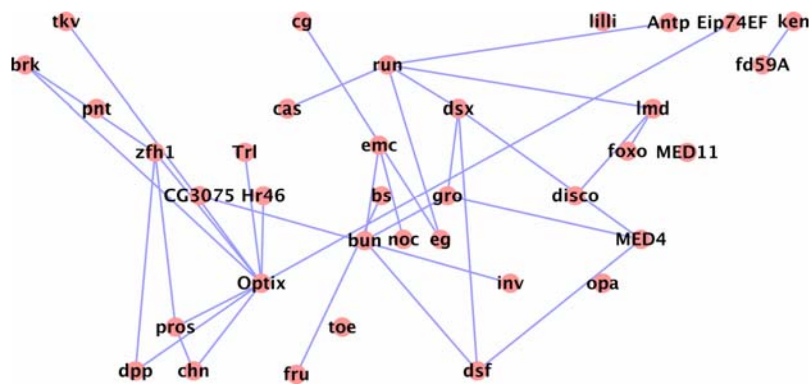


SAILING LAB

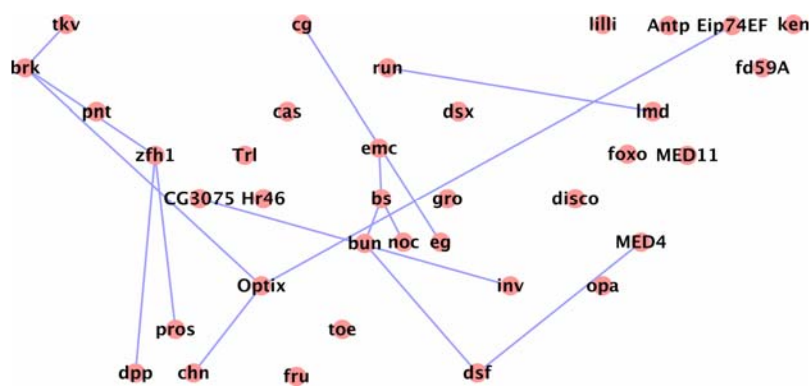
TF Cascade – mid-larva stage



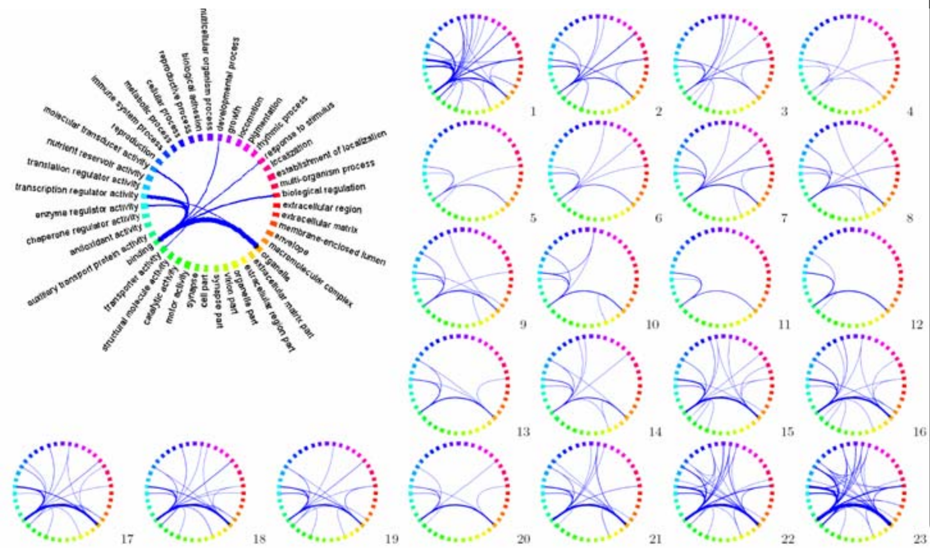
TF Cascade – mid-pupal stage



TF Cascade – mid-adult stage



Transient Group Interactions



Discussion:

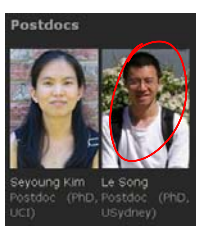
- How about estimating continuous-valued graphs?
  - Both KELLER and TEALA can be applied to such case --- change the logistic regression to linear regression
- How about estimating directed graphs?
  - We have been able to extend KELLER to estimating *time-varying DBN* based as an *nonstationary auto-regressive model*
  - Consistency proof is difficult (sample no longer conditionally independent), but still possible under assumption of local stationarity
- How about general time-varying graphical models?
  - Open problem in both algorithm and theory

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

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics



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

Funding: