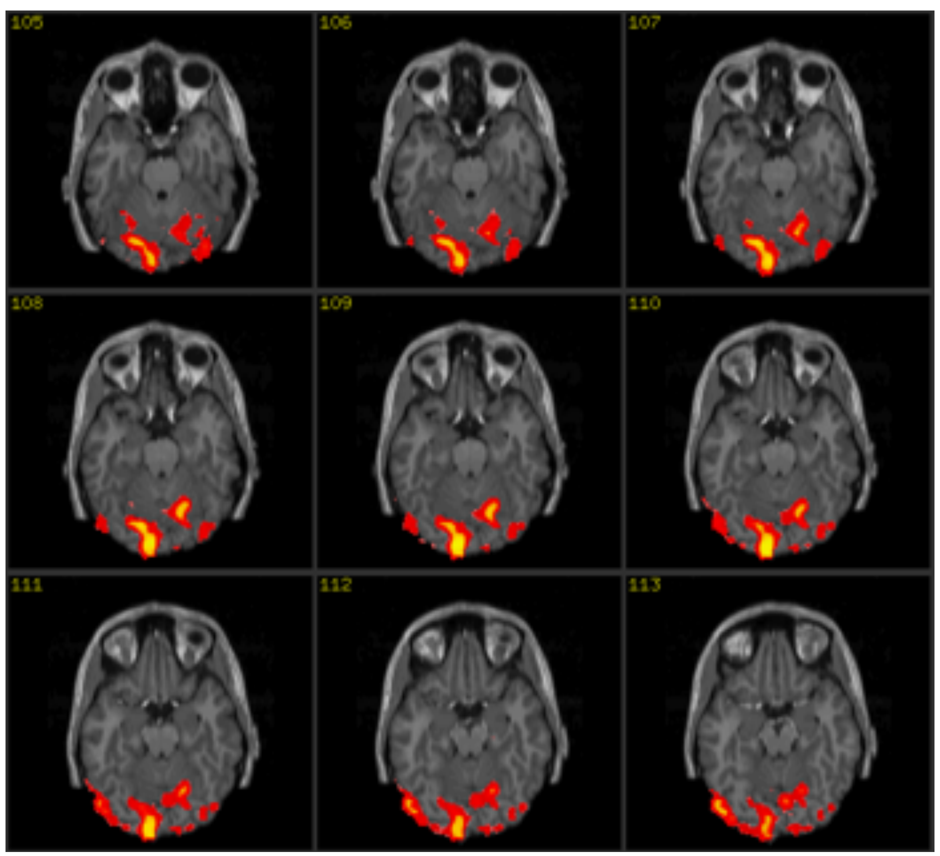
Learning Graphical Model Structure

Pradeep Ravikumar
UT Austin

School of ICASSP 2015

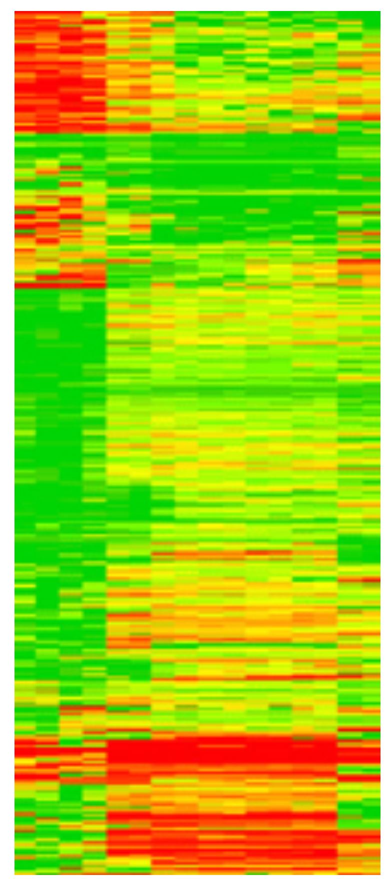
“Big-p” Data: large number of variables “p”

- Across modern applications {images, signals, networks}
  many^many variables



fMRI images

variables: image voxels



gene expression profiles

variables: genes

social networks

variables: users

"Big-p" Data

- A critical question given a large number of variables of interest:

“Big-p” Data

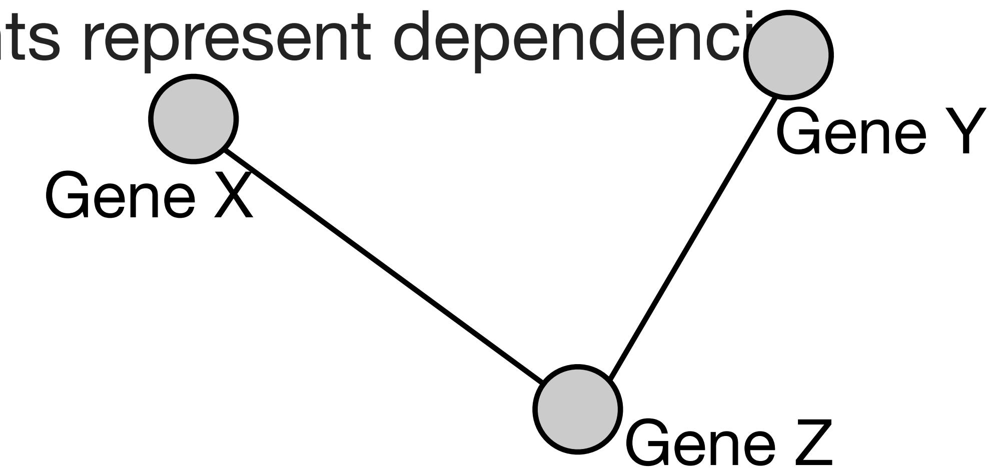
- A critical question given a large number of variables of interest:
  - What are the connections/dependencies among the variables?

“Big-p” Data

- A critical question given a large number of variables of interest:
  - What are the connections/dependencies among the variables?
- Consider a visual representation of this problem: where the variables are represented as nodes of a graph, and edge weights represent dependencies

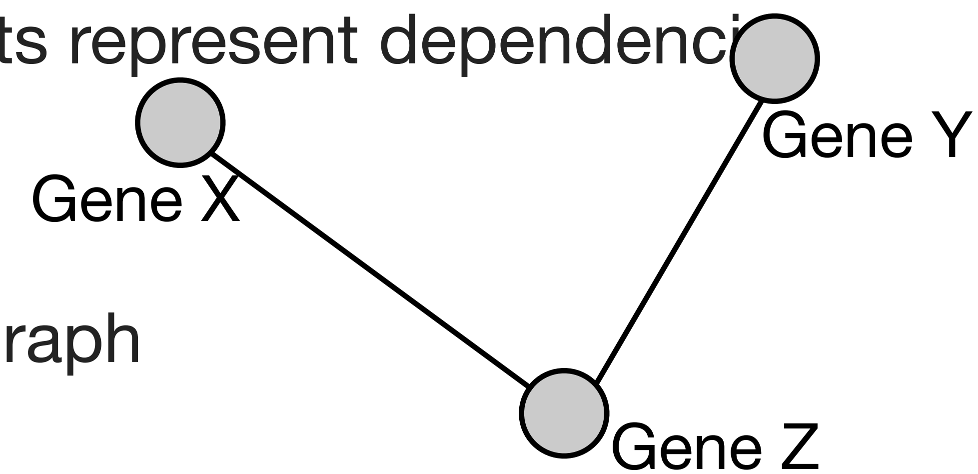
"Big-p" Data

- A critical question given a large number of variables of interest:
  - What are the connections/dependencies among the variables?
- Consider a visual representation of this problem: where the variables are represented as nodes of a graph, and edge weights represent dependenci

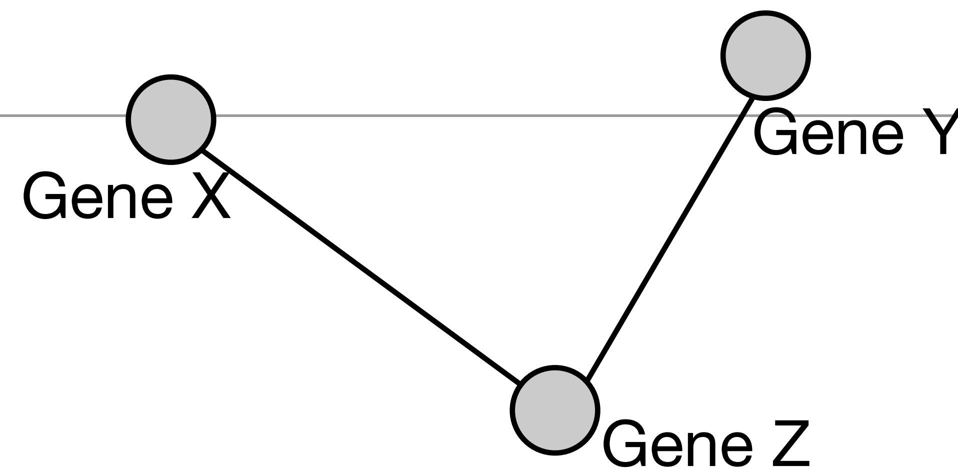


“Big-p” Data

- A critical question given a large number of variables of interest:
  - What are the connections/dependencies among the variables?
- Consider a visual representation of this problem: where the variables are represented as nodes of a graph, and edge weights represent dependenci
- Estimating the dependencies among the variables is then equivalent to estimating such a weighted graph

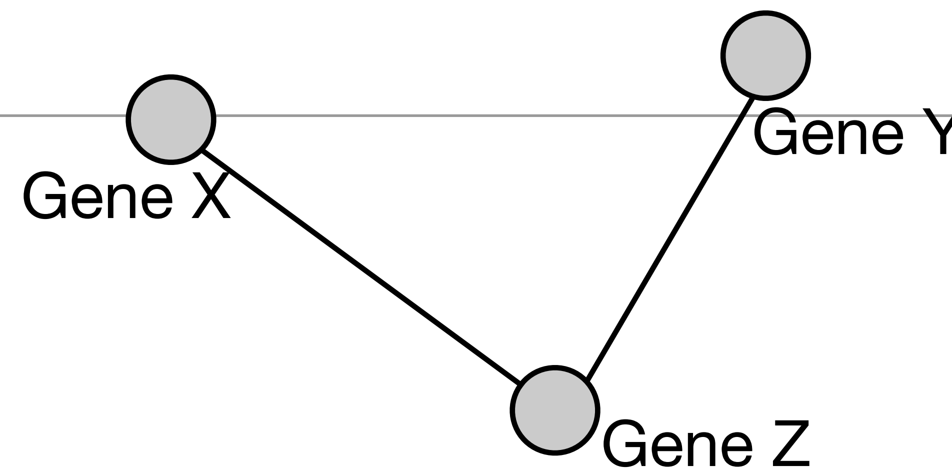


Graph Structure



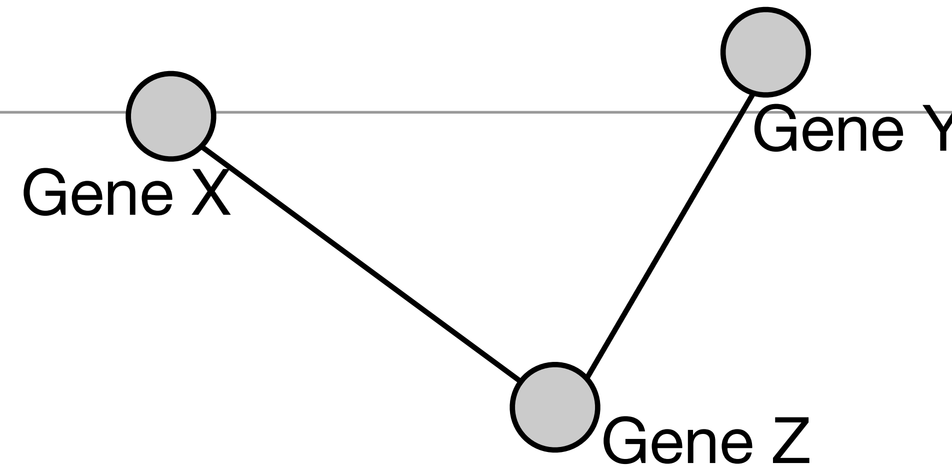
- What dependencies between variables could we be interested in?

Graph Structure



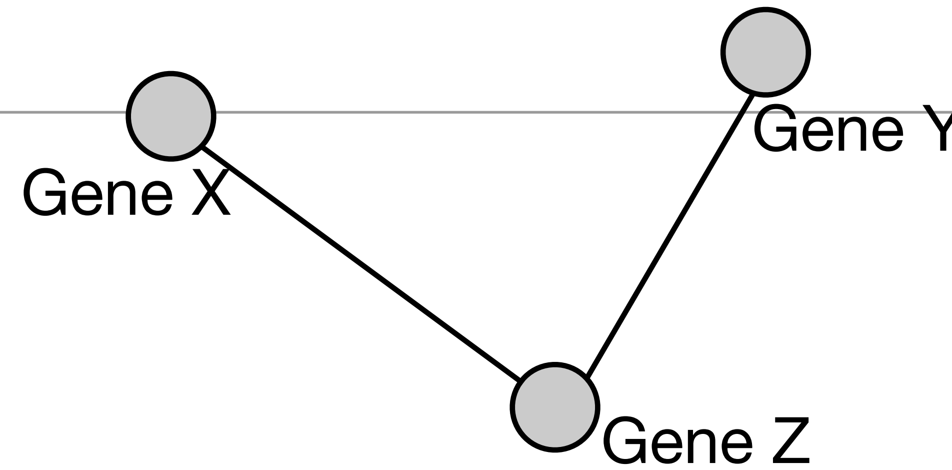
- What dependencies between variables could we be interested in?
    - Correlation? *Gene X activity is highly correlated with Gene Z activity*

Graph Structure



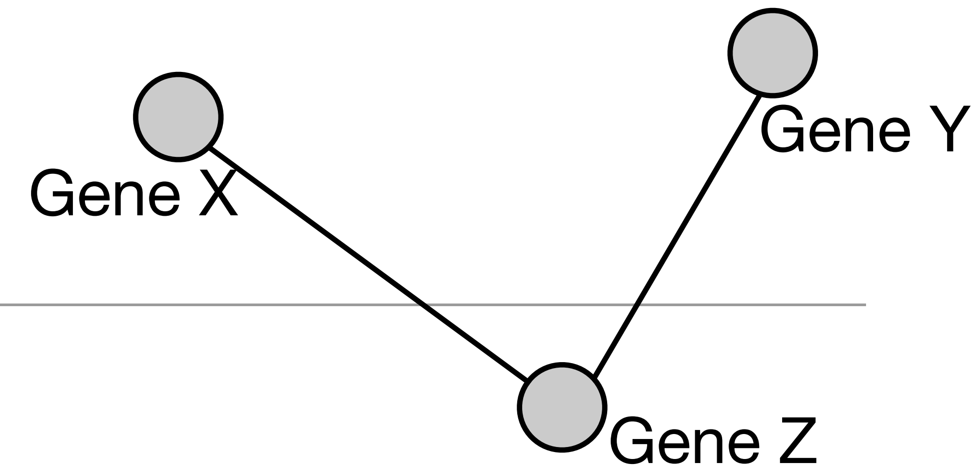
- What dependencies between variables could we be interested in?
  - Correlation? *Gene X activity is highly correlated with Gene Z activity*
  - Causation? *Gene X being active causes Gene Z to be active*

Graph Structure



- What dependencies between variables could we be interested in?
  - Correlation? *Gene X activity is highly correlated with Gene Z activity*
  - Causation? *Gene X being active causes Gene Z to be active*
  - **Conditional (In)dependence**: *Given all other genes, are Gene X and Gene Z (in)dependent?*

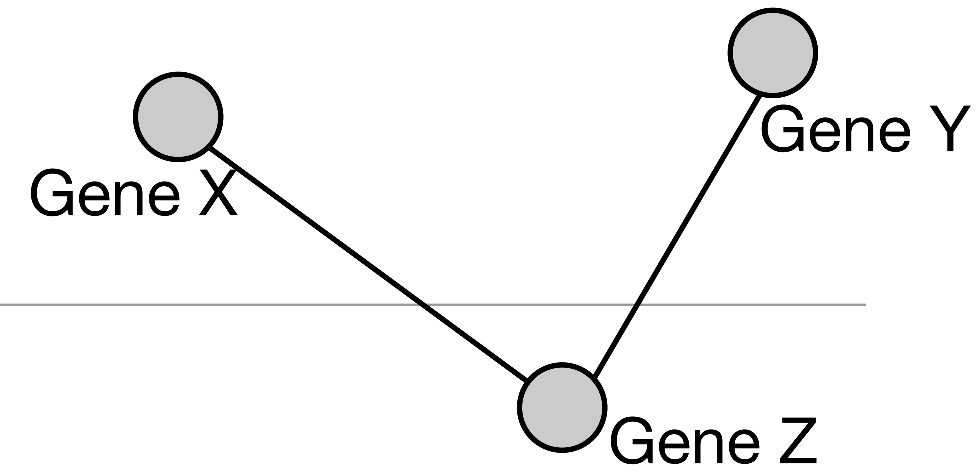
Graph Structure



- What dependencies between variables could we be interested in?

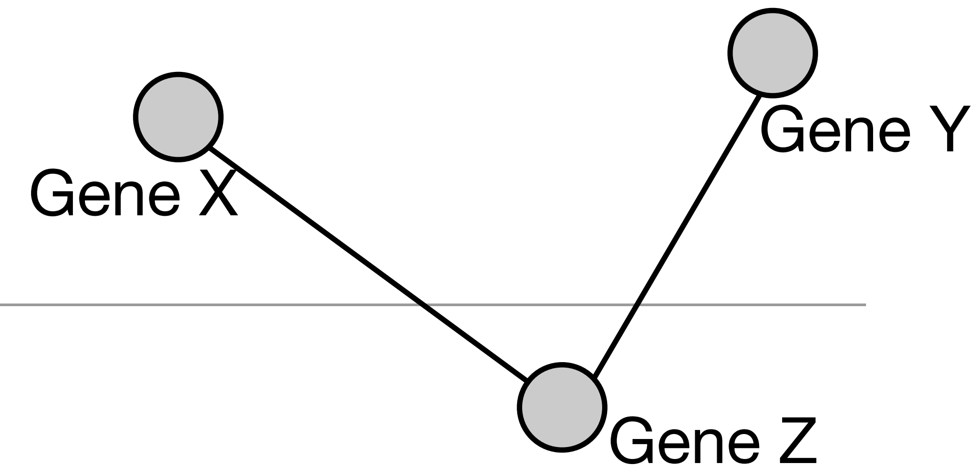
  - **Conditional (In)dependence**: *Given all other genes, are Gene X and Gene Y (in)dependent?*

Graph Structure



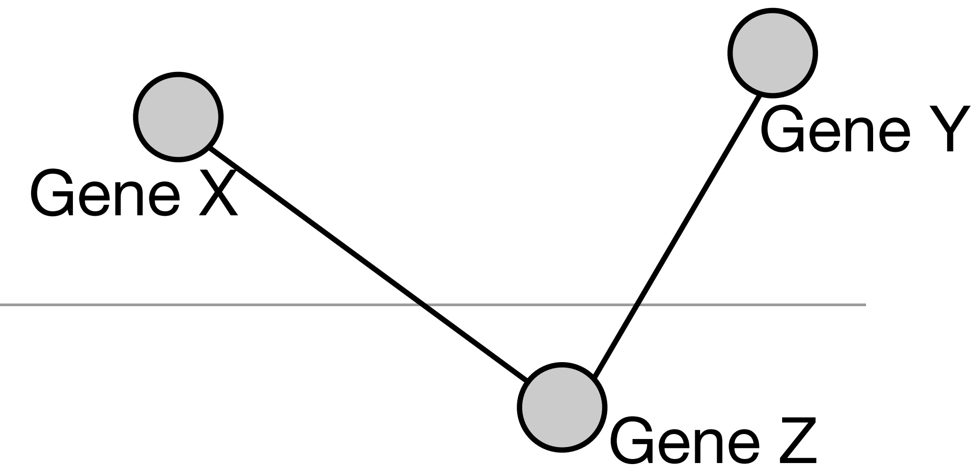
- What dependencies between variables could we be interested in?
  - **Conditional (In)dependence**: *Given all other genes, are Gene X and Gene Y (in)dependent?*
  - X = "shoe-size" and Y = "gray-hair" are "marginally" dependent (think of small children with small shoe-sizes and no gray-hair)

Graph Structure



- What dependencies between variables could we be interested in?
  - **Conditional (In)dependence**: *Given all other genes, are Gene X and Gene Y (in)dependent?*
  - X = “shoe-size” and Y = “gray-hair” are “marginally” dependent (think of small children with small shoe-sizes and no gray-hair)
    - But “shoe size” and “gray hair” are common-sensically not directly associated

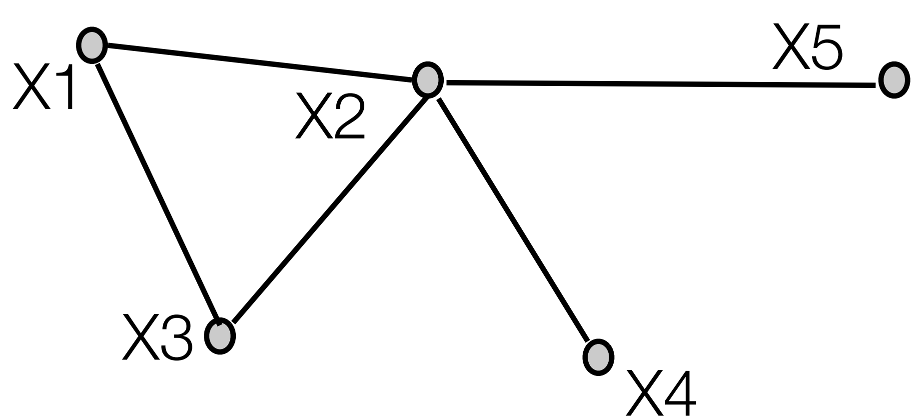
Graph Structure



- What dependencies between variables could we be interested in?
  - **Conditional (In)dependence**: *Given all other genes, are Gene X and Gene Y (in)dependent?*
  - X = “shoe-size” and Y = “gray-hair” are “marginally” dependent (think of small children with small shoe-sizes and no gray-hair)
    - But “shoe size” and “gray hair” are common-sensically not directly associated
    - Given Z = “age”, the dependence vanishes away: they are conditionally independent

Conditional Independence Graph Structure

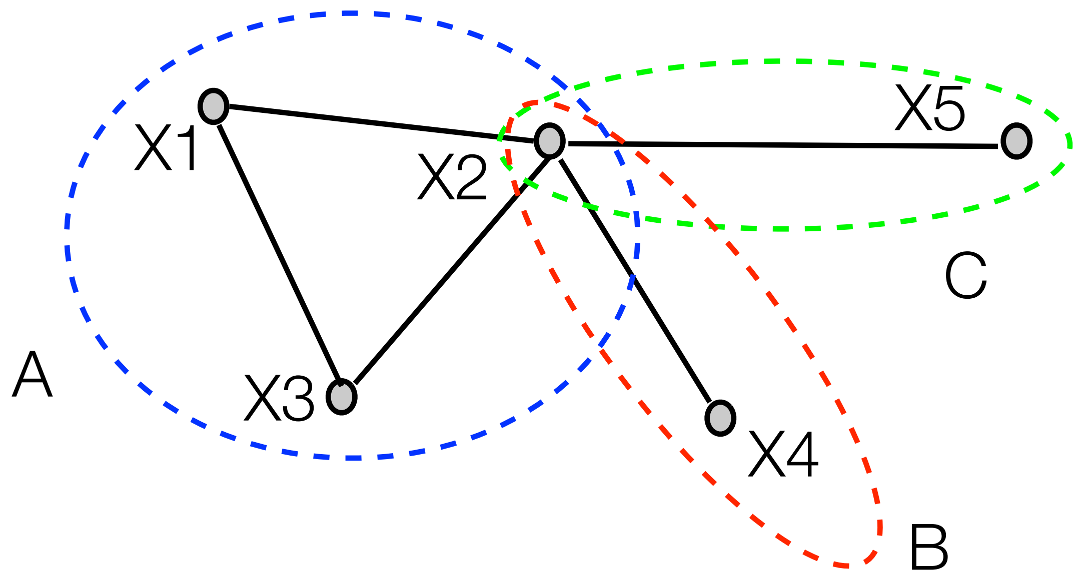
- Lack of an edge: lack of "direct dependence"
- no-edge(x,y) : x and y are independent given rest of nodes



$$X_3 \perp X_4 \,|\, \{X_1, X_2, X_5\}$$

Edges indicate Markov independence conditions

Graphical Model Structure

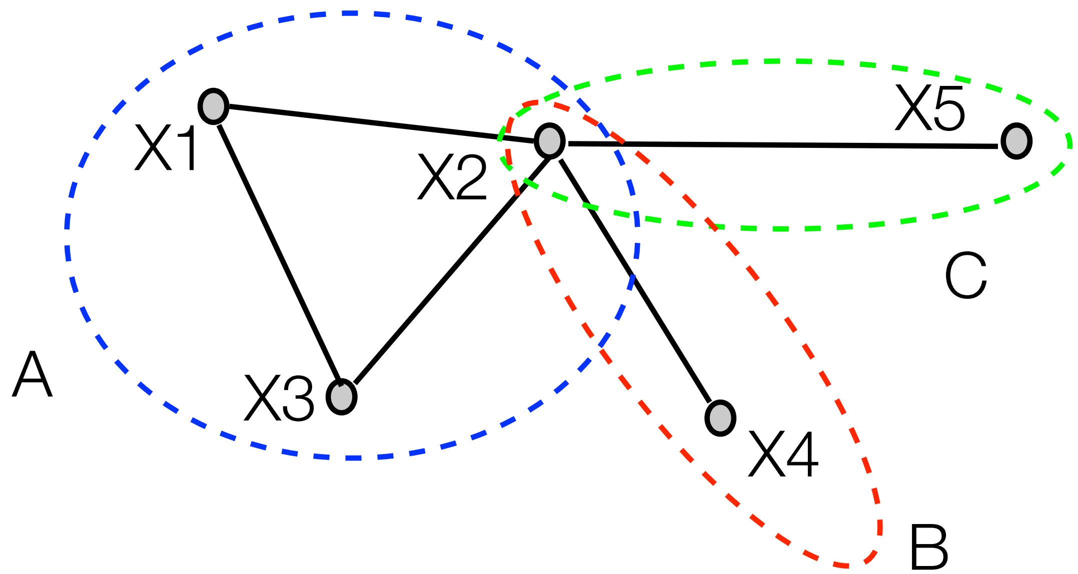


$$X_3 \perp X_4 \,|\, \{X_1, X_2, X_5\}$$

Edges indicate Markov independence conditions

- Given some graph G representing the conditional independence edge structure among the vector of random variables X

Graphical Model Structure

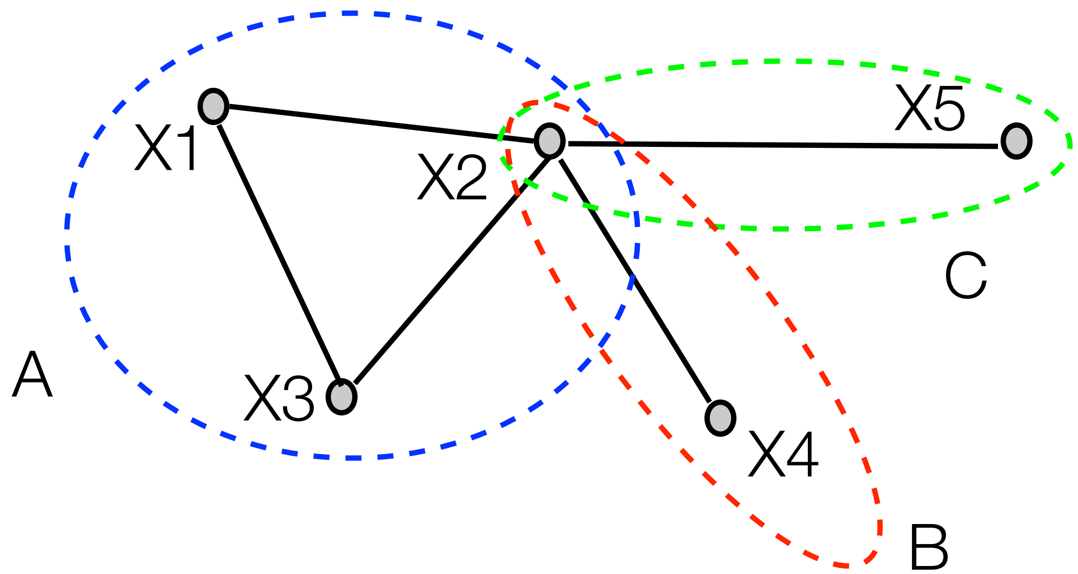


$$X_3 \perp X_4 \,|\, \{X_1, X_2, X_5\}$$

Edges indicate Markov independence conditions

- Given some graph G representing the conditional independence edge structure among the vector of random variables X
  - What is the set of distributions over X that respects this conditional independence structure (in other words, that satisfies all these conditional independences among the variables)

Graphical Model Structure

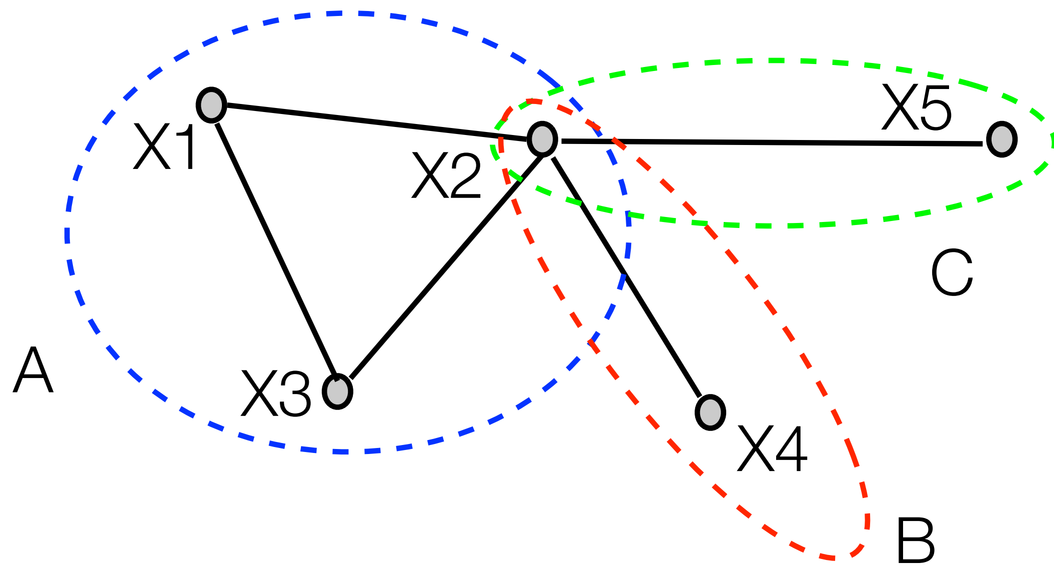


$$X_3 \perp X_4 \,|\, \{X_1, X_2, X_5\}$$

Edges indicate Markov independence conditions

- Given some graph G representing the conditional independence edge structure among the vector of random variables X
  - What is the set of distributions over X that respects this conditional independence structure (in other words, that satisfies all these conditional independences among the variables)
  - This set of distributions is called the graphical model represented by G

Graphical Model Structure

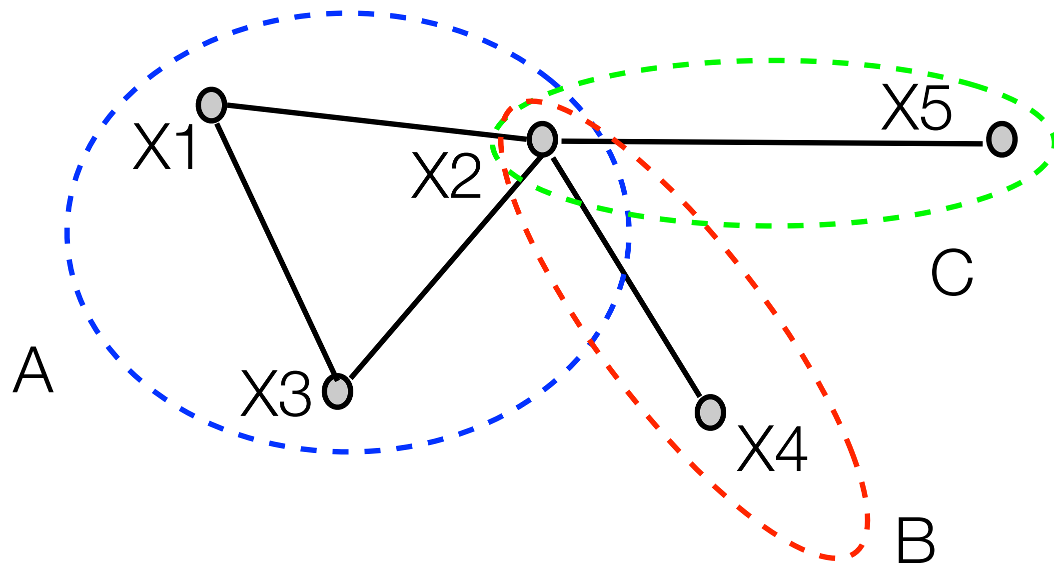


$$X_3 \perp X_4 \,|\, \{X_1, X_2, X_5\}$$

Edges indicate Markov independence conditions

- The graphical model represented by G is a family of distributions that respects the conditional independence structure specified by G

Graphical Model Structure

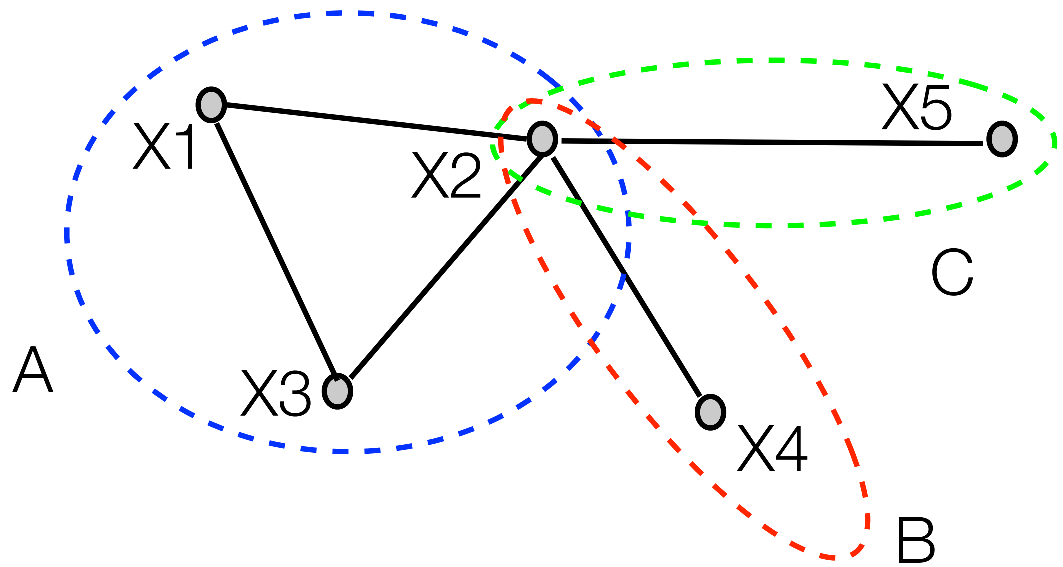


$$X_3 \perp X_4 \,|\, \{X_1, X_2, X_5\}$$

Edges indicate Markov independence conditions

- The graphical model represented by G is a family of distributions that respects the conditional independence structure specified by G

Graphical Model Structure



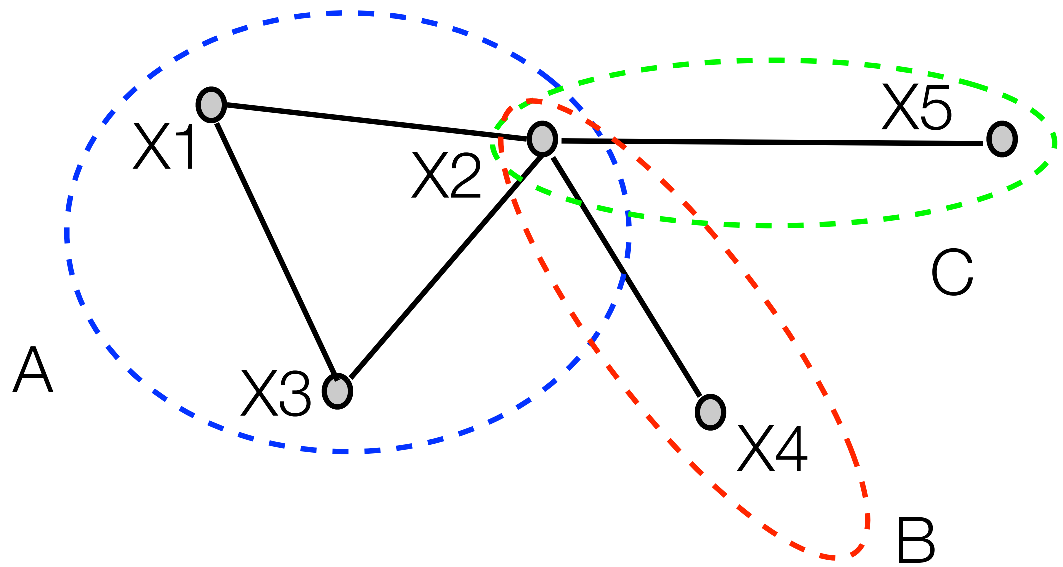
$$X_3 \perp X_4 \,|\, \{X_1, X_2, X_5\}$$

Edges indicate Markov independence conditions

- The graphical model represented by G is a family of distributions that respects the conditional independence structure specified by G

- Do these distributions have any particular algebraic form?

Graphical Model Structure



$$X_3 \perp X_4 \,|\, \{X_1, X_2, X_5\}$$

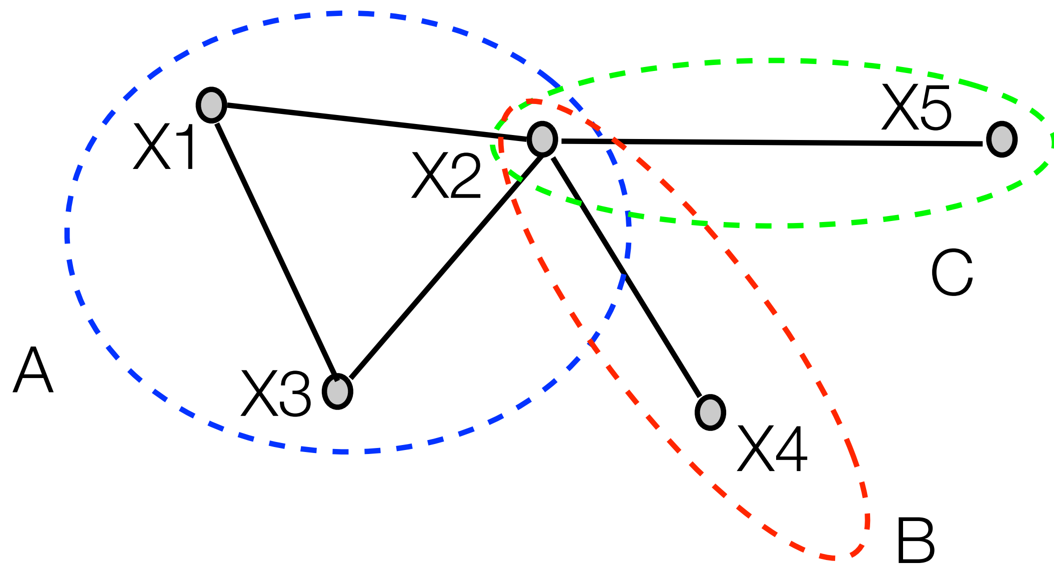
Edges indicate Markov independence conditions

- The graphical model represented by G is a family of distributions that respects the conditional independence structure specified by G

- Do these distributions have any particular algebraic form?

- **Hammersley Clifford:** they always take the form of a product of local factors, each of which depend only on a clique (fully connected subgraph)

Graphical Model Structure



$$X_3 \perp X_4 \,|\, \{X_1, X_2, X_5\}$$

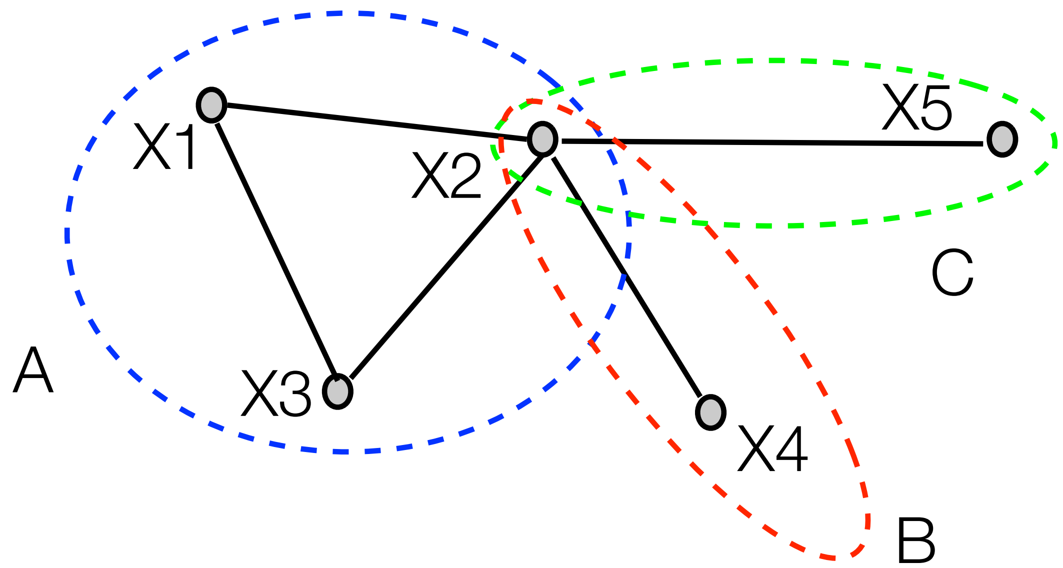
Edges indicate Markov independence conditions

- The graphical model represented by G is a family of distributions that respects the conditional independence structure specified by G

- Do these distributions have any particular algebraic form?

- **Hammersley Clifford:** they always take the form of a product of local factors, each of which depend only on a clique (fully connected subgraph)

Graphical Model Structure



$$X_3 \perp X_4 \,|\, \{X_1, X_2, X_5\}$$

Edges indicate Markov independence conditions

- The graphical model represented by G is a family of distributions that respects the conditional independence structure specified by G
- Do these distributions have any particular algebraic form?
- **Hammersley Clifford:** they take the form of a product of local factors, each of which depend only on a clique (fully connected subgraph)

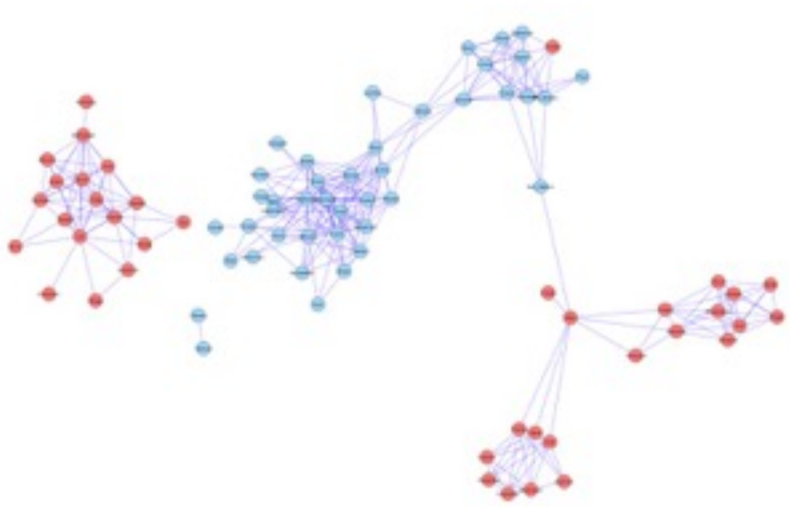
$$p(X) = \frac{1}{Z} \Psi_A(X_A) \Psi_B(X_B) \Psi_C(X_C)$$

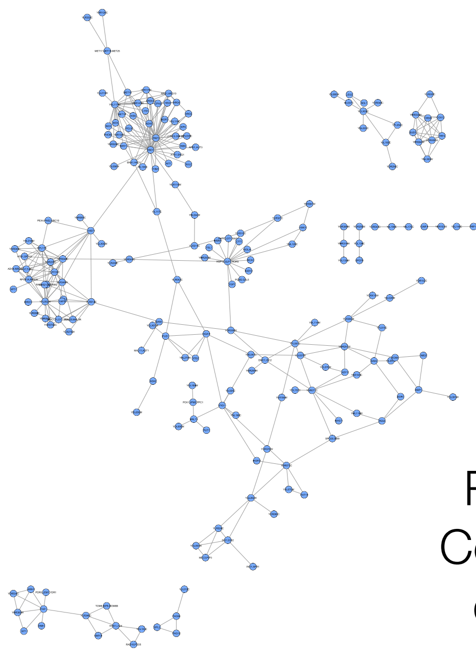
Graphical Model Structure

- The conditional independence graph structure, underlying a graphical model, is an object of interest in varied applications
  - network analysis, medical diagnosis, gene expression analyses, natural language processing, ...

US Senate 109th Congress

Banerjee et al, 2008



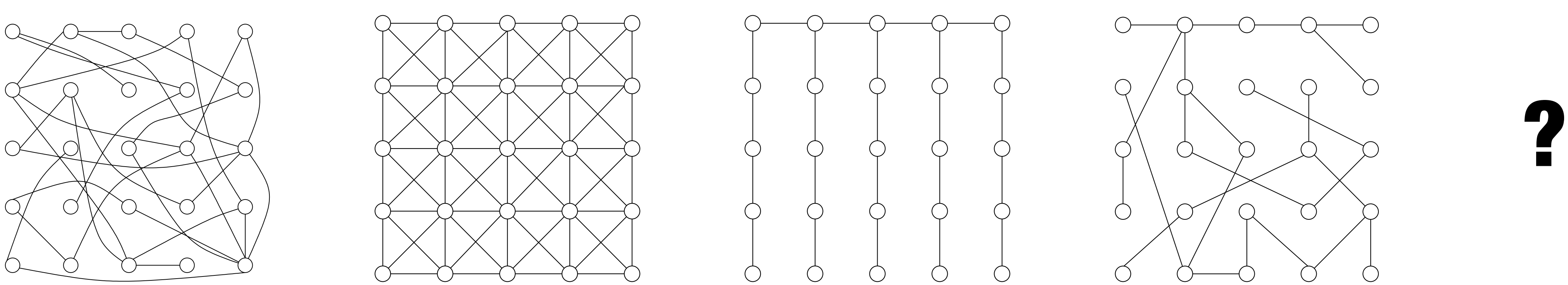


Rosetta Informatics Compendium of gene expression profiles

Graphical Model Structure Selection

GIVEN: n samples of $X = (X_1, \ldots, X_p)$

drawn from some unknown graphical model distribution P(X; G)
for some unknown graph G, recover the graph G.



Graphical Model Structure Selection

GIVEN: n samples of $X = (X_1, \ldots, X_p)$

drawn from some unknown graphical model distribution P(X; G)
for some unknown graph G, recover the graph G.

- It is common to further assume a parametric model form for P(X; G)
  - Ising Models, Multinomial (Discrete) Models,
    Gaussian Graphical Models, …

Examples: Parametric Graphical Models

$$p(X; \theta, G) = \frac{1}{Z(\theta)} \exp\Big(\sum_{(s,t) \in E(G)} \theta_{st}\, \phi_{st}(X_s, X_t) \Big)$$

$$\phi_{st}(x_s, x_t) \,:\, \text{arbitrary potential functions}$$

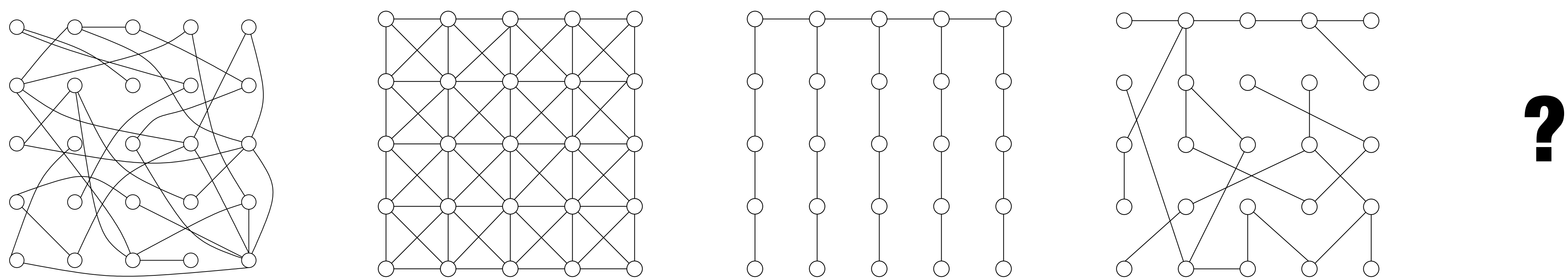
Ising	$x_s\, x_t$
Potts	$I(x_s = x_t)$
Indicator	$I(x_s, x_t = j, k)$

Parametric Graphical Model Selection

GIVEN: n samples of $X = (X_1, \ldots, X_p)$ with distribution $p(X;\theta^*;G)$, where

$$p(X;\theta^*) = \exp\left\{ \sum_{(s,t)\in E(G)} \theta_{st}\phi_{st}(x_s, x_t) - A(\theta^*) \right\}$$

PROBLEM: Estimate graph G given just the n samples.



Graphical Model Selection: Classical Approaches

- Score Based Approaches: **search** over space of graphs, with a score for any graph (based on learning the parametric graphical model given the graph)

Graphical Model Selection: Classical Approaches

- Score Based Approaches: **search** over space of graphs, with a score for any graph (based on learning the parametric graphical model given the graph)

- Constraint-based Approaches: estimate individual edges by **hypothesis tests** for conditional independences

Graphical Model Selection: Classical Approaches

- Score Based Approaches: **search** over space of graphs, with a score for any graph (based on learning the parametric graphical model given the graph)

- Constraint-based Approaches: estimate individual edges by **hypothesis tests** for conditional independences

- Caveats:

Graphical Model Selection: Classical Approaches

- Score Based Approaches: **search** over space of graphs, with a score for any graph (based on learning the parametric graphical model given the graph)

- Constraint-based Approaches: estimate individual edges by **hypothesis tests** for conditional independences

- Caveats:

  - difficult to provide guarantees for estimators

Graphical Model Selection: Classical Approaches

- Score Based Approaches: **search** over space of graphs, with a score for any graph (based on learning the parametric graphical model given the graph)

- Constraint-based Approaches: estimate individual edges by **hypothesis tests** for conditional independences

- Caveats:

  - difficult to provide guarantees for estimators

  - estimation problems they solve are NP-Hard

Graphical Model Selection

- Modern Approach: statistical estimation of the parametric graphical model subject to constraints on the underlying graph (e.g. edge bounds, degree bounds, etc.)
    - Caveats: such statistical estimation is not always computationally tractable; statistical guarantees plausible, but require advanced arguments

Graph-structure constrained MLE

$$\widehat{\theta} \in \arg \min_{\theta : \theta \in \Theta} \left\{ -\frac{1}{n} \sum_{i=1}^{n} \log p(x^{(i)}; \theta) \right\}$$

graph constraints **neg. log-likelihood**

- Statistical Estimation typically **intractable** because of
  - Graph Constraints: typically non-convex
  - Likelihood function: typically NP-Hard to **compute**

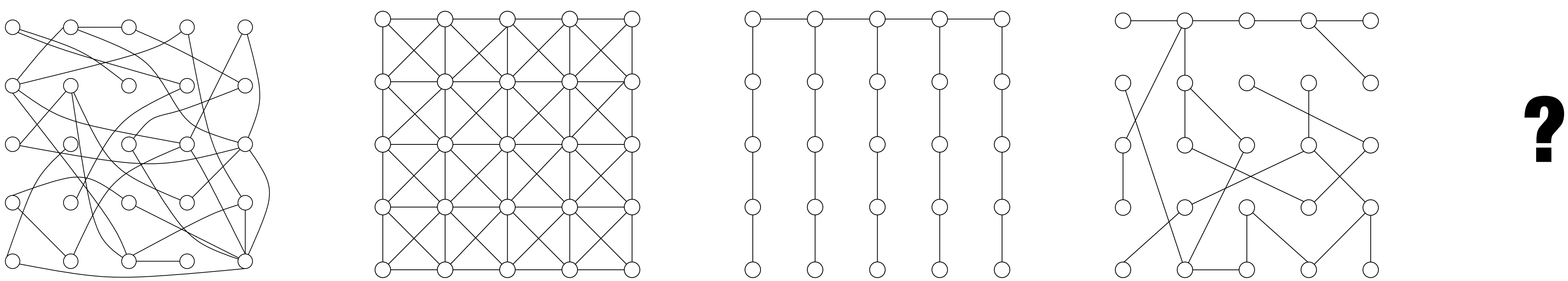
Outline: Graphical Model Selection

- Ising Models
- In brief: Gaussian Graphical Models, Multinomial Discrete Graphical Models
- In brief: a new class of parametric graphical models — exponential family graphical models

Ising Model Selection

GIVEN: n samples of $X = (X_1, \ldots, X_p)$ with distribution $p(X; \theta^*; G)$, where

$$p(X; \theta^*) = \exp\left\{ \sum_{(s,t) \in E(G)} \theta^*_{st} X_s X_t - A(\theta^*) \right\}$$

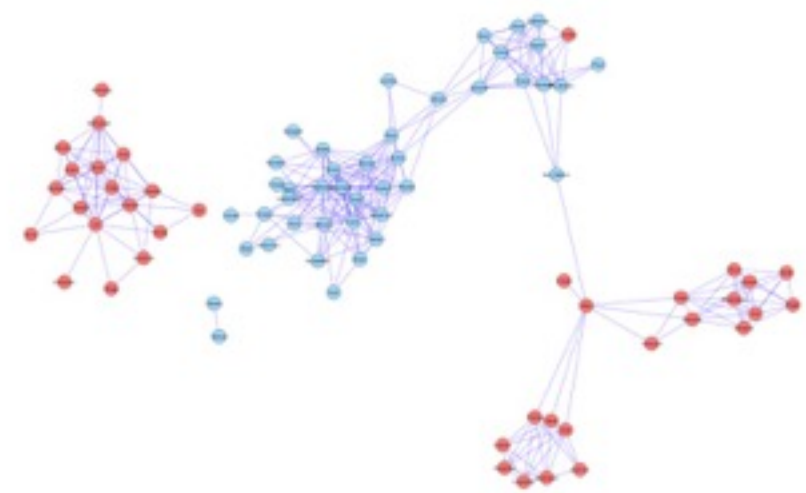


Ising Model Selection

GIVEN: n samples of $X = (X_1, \ldots, X_p)$ with distribution $p(X; \theta^*; G)$, where

$$p(X; \theta^*) = \exp\left\{ \sum_{(s,t) \in E(G)} \theta^*_{st} X_s X_t - A(\theta^*) \right\}$$

Applications: statistical physics, computer vision, social network analysis



US Senate 109th Congress

Banerjee et al, 2008

Ising Model Selection

- Just computing the likelihood of a **known** Ising model is NP Hard (since the normalization constant requires summing over exponentially many configurations)

$$Z(\theta) = \sum_{x \in \{-1,1\}^p} \exp\left(\sum_{st} \theta_{st}\, x_s\, x_t\right)$$

Ising Model Selection

- Just computing the likelihood of a **known** Ising model is NP Hard (since the normalization constant requires summing over exponentially many configurations)

$$Z(\theta) = \sum_{x \in \{-1,1\}^p} \exp\left(\sum_{st} \theta_{st}\, x_s\, x_t\right)$$

- Estimating the **unknown** Ising model parameters as well as graph structure might seem to be NP Hard as well

Ising Model Selection

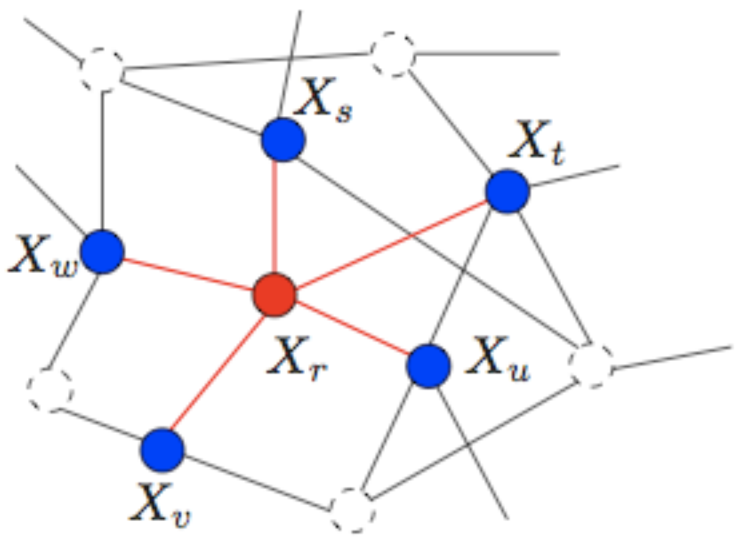
- Just computing the likelihood of a **known** Ising model is NP Hard (since the normalization constant requires summing over exponentially many configurations)

$$Z(\theta) = \sum_{x \in \{-1,1\}^p} \exp\left(\sum_{st} \theta_{st}\, x_s\, x_t \right)$$

- Estimating the **unknown** Ising model parameters as well as graph structure might seem to be NP Hard as well

- On the other hand, it is tractable to estimate the node-wise conditional distributions, of one variable conditioned on the rest of the variables

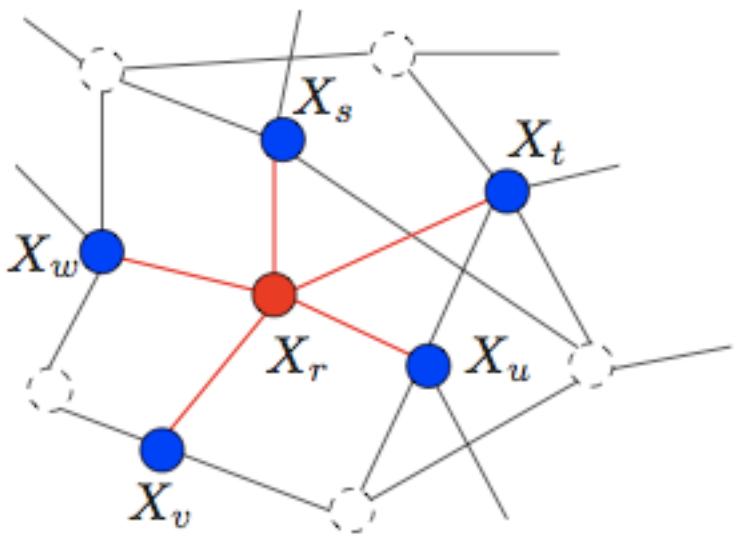
Neighborhood Estimation in Ising Models



For Ising models, node conditional distribution is just a logistic regression model:

$$p(X_r | X_{V\backslash r}; \theta, G) = \frac{\exp(\sum_{t\in N(r)} 2\,\theta_{rt} X_r X_t)}{\exp(\sum_{t\in N(r)} 2\,\theta_{rt} X_r X_t) + 1}$$

Neighborhood Estimation in Ising Models

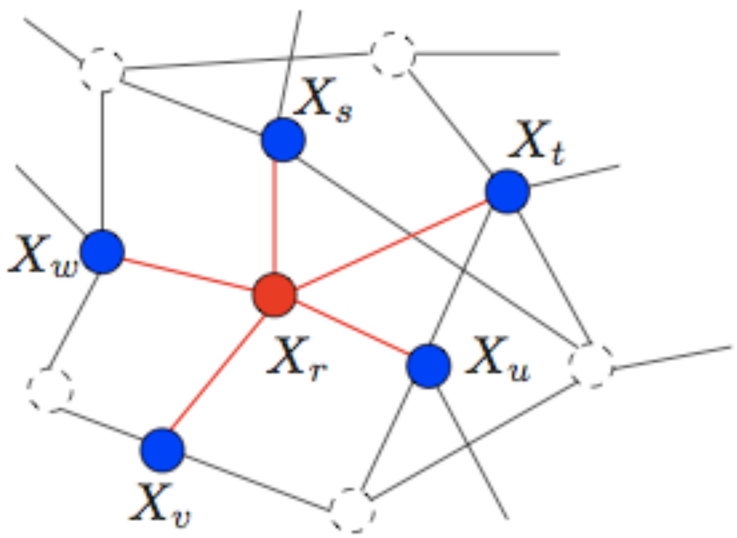


For Ising models, node conditional distribution is just a logistic regression model:

$$p(X_r | X_{V \backslash r}; \theta, G) = \frac{\exp(\sum_{t \in N(r)} 2\,\theta_{rt} X_r X_t)}{\exp(\sum_{t \in N(r)} 2\,\theta_{rt} X_r X_t) + 1}$$

- So instead of estimating graph structure constrained global Ising model, we could estimate structure constrained local node-conditional distributions — logistic regression models

Neighborhood Estimation in Ising Models



For Ising models, node conditional distribution is just a logistic regression model:

$$p(X_r | X_{V\backslash r}; \theta, G) = \frac{\exp(\sum_{t \in N(r)} 2\,\theta_{rt} X_r X_t)}{\exp(\sum_{t \in N(r)} 2\,\theta_{rt} X_r X_t) + 1}$$

- So instead of estimating graph structure constrained global Ising model, we could estimate structure constrained local node-conditional distributions — logistic regression models
- But would node-conditional distributions uniquely specify a consistent joint, or even be consistent with any joint at all?

Conditional and Joint Distributions

- Would node-conditional distributions uniquely specify a consistent joint, or even be consistent with any joint at all?

Conditional and Joint Distributions

- Would node-conditional distributions uniquely specify a consistent joint, or even be consistent with any joint at all?

- In general: no!

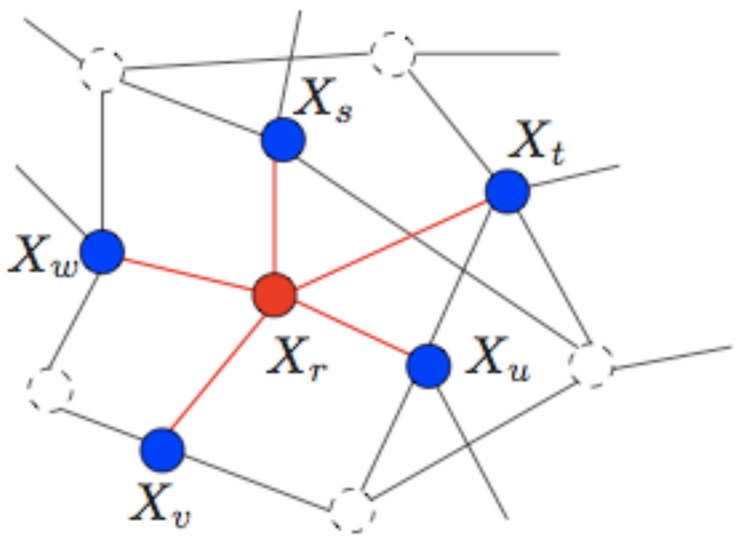
Conditional and Joint Distributions

- Would node-conditional distributions uniquely specify a consistent joint, or even be consistent with any joint at all?
- In general: no!
- But for the Ising model and node-wise logistic regression models: yes!

Conditional and Joint Distributions

- Would node-conditional distributions uniquely specify a consistent joint, or even be consistent with any joint at all?
- In general: no!
- But for the Ising model and node-wise logistic regression models: yes!
    - **Theorem (Besag 1974, R., Wainwright, Lafferty 2010):** An Ising model uniquely specifies and is uniquely specified by a set of node-wise logistic regression models.

Neighborhood Estimation in Ising Models



For Ising models, node conditional distribution is just a logistic regression model:

$$p(X_r|X_{V\backslash r};\theta,G) = \frac{\exp(\sum_{t\in N(r)} 2\,\theta_{rt}X_rX_t)}{\exp(\sum_{t\in N(r)} 2\,\theta_{rt}X_rX_t)+1}$$

- Global graph constraint of sparse, bounded degree graphs is equivalent to local constraint of bounded node-degrees (number of neighbors)
- Estimate node neighborhoods via constrained logistic regression models, and stich node-neighborhoods together to form global graph

Graph selection via neighborhood regression

Observation: Recovering graph G equivalent to recovering neighborhood set $N(s)$ for all $s \in V$.

Method: Given n i.i.d. samples $\{X^{(1)}, \ldots, X^{(n)}\}$, perform logistic regression of each node X_s on $X_{\setminus s} := \{X_s,\ t \neq s\}$ to estimate neighborhood structure $\widehat{N}(s)$.

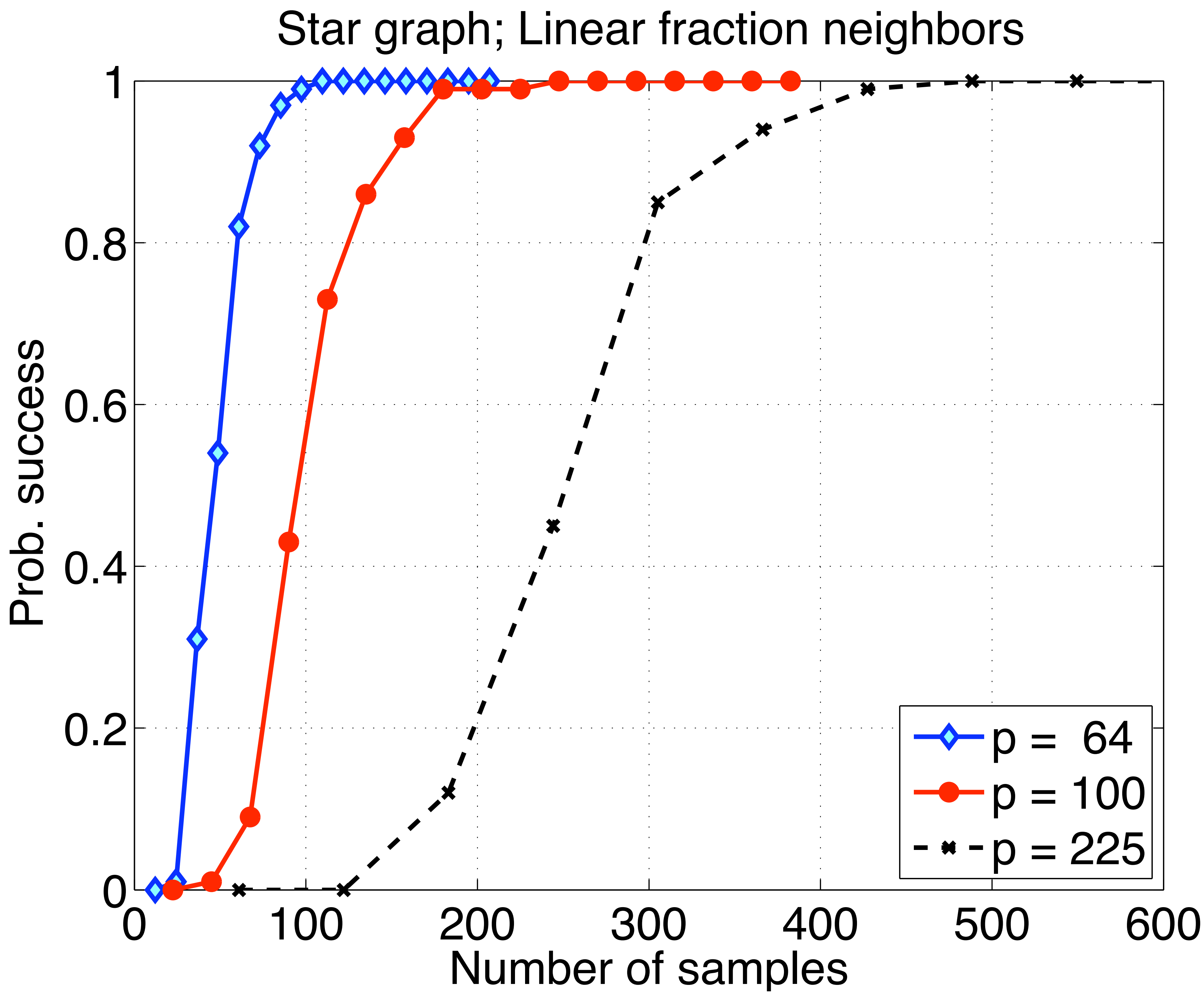
1. For each node $s \in V$, perform ℓ_1 regularized logistic regression of X_s on the remaining variables $X_{\setminus s}$:

$$\widehat{\theta}[s] \quad := \quad \arg \min_{\theta \in \mathbb{R}^{p-1}} \Big\{ \underbrace{\frac{1}{n} \sum_{i=1}^{n} f(\theta; X_{\setminus s}^{(i)})}_{\text{logistic likelihood}} \quad + \quad \rho_n \underbrace{\|\theta\|_1}_{\text{regularization}} \Big\}$$

2. Estimate the local neighborhood $\widehat{N}(s)$ as the support (non-negative entries) of the regression vector $\widehat{\theta}[s]$.

3. Combine the neighborhood estimates in a consistent manner (AND, or OR rule).

Empirical behavior: Unrescaled plots



Sufficient conditions for consistent model selection

- graph sequences $G_{p,d} = (V, E)$ with p vertices, and maximum degree d.
- edge weights $|\theta_{st}| \geq \theta_{\min}$ for all $(s,t) \in E$
- draw n i.i.d, samples, and analyze prob. success indexed by (n, p, d)

Theorem

(R., Wainwright, Lafferty, 2010)

Under incoherence conditions, for a rescaled sample size

$$\theta_{LR}(n,p,d) \quad := \quad \frac{n}{d^3 \log p} \; > \; \theta_{\text{crit}}$$

and regularization parameter $\rho_n \geq c_1 \tau \sqrt{\frac{\log p}{n}}$, *then with probability greater than* $1 - 2\exp\big(-c_2(\tau - 2)\log p\big) \to 1$*:*

(a) Uniqueness: *For each node $s \in V$, the ℓ_1-regularized logistic convex program has a unique solution. (Non-trivial since $p \gg n \implies$ not strictly convex).*

(b) Correct exclusion: *The estimated sign neighborhood $\widehat{N}(s)$ correctly excludes all edges not in the true neighborhood.*

(c) Correct inclusion: *For $\theta_{\min} \geq c_3 \tau \sqrt{d} \rho_n$, the method selects the correct signed neighborhood.*

Consequence: For $\theta_{\min} = \Omega(1/d)$, it suffices to have $n = \Omega(d^3 \log p)$.

Assumptions

Define Fisher information matrix of logistic regression:
$Q^* := \mathbb{E}_{\theta^*}\left[\nabla^2 f(\theta^*; X)\right]$.

A1. Dependency condition: Bounded eigenspectra:

$$C_{min} \;\leq\; \lambda_{min}(Q^*_{SS}), \qquad \text{and} \qquad \lambda_{max}(Q^*_{SS}) \;\leq\; C_{max}.$$
$$\lambda_{max}(\mathbb{E}_{\theta^*}[XX^T]) \quad \leq \quad D_{\max}.$$

A2. Incoherence There exists an $\nu \in (0, 1]$ such that

$$|||Q^*_{S^cS}(Q^*_{SS})^{-1}|||_{\infty,\infty} \;\;\leq\;\; 1 - \nu.$$

where $|||A|||_{\infty,\infty} := \max_i \sum_j |A_{ij}|$.

- bounds on eigenvalues are fairly standard
- incoherence condition:
  - partly necessary (prevention of degenerate models)
  - partly an artifact of ℓ_1-regularization
- incoherence condition is weaker than correlation decay

Multinomial, Gaussian Graphical Models

- Ising models are a specific parametric graphical model family, suited to the case where the variables are binary.

Multinomial, Gaussian Graphical Models

- Ising models are a specific parametric graphical model family, suited to the case where the variables are binary.

- When variables are categorical, taking values in a finite set:

  - Multinomial/Discrete Graphical Models (Jalali, **R.**, Vasuki, Sanghavi, 2011)

  - Applications: natural language processing, image analysis, bioinformatics

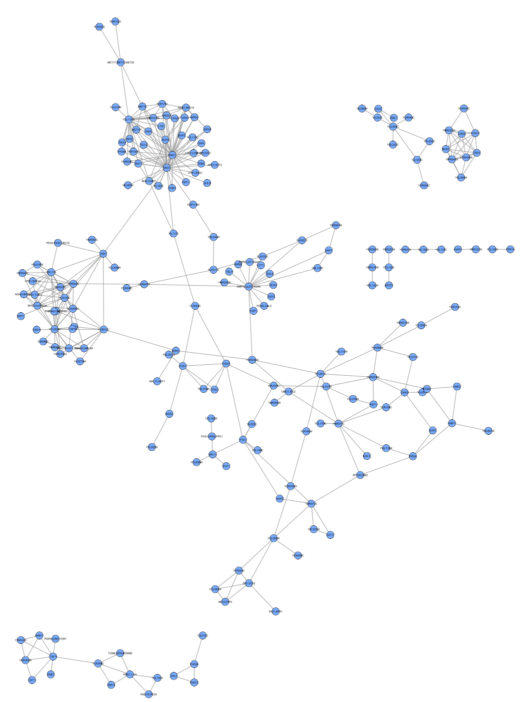
Multinomial, Gaussian Graphical Models

- Ising models are a specific parametric graphical model family, suited to the case where the variables are binary.

- When variables are thin-tailed continuous

  - Gaussian Graphical Models (**R.**, Raskutti, Wainwright, Yu, 2012)

  - Applications: widely used in bioinformatics e.g. genomic networks from micro-array data



Rosetta Informatics
Compendium of gene
expression profiles

Multinomial, Gaussian Graphical Models

- Ising models are a specific parametric graphical model family, suited to the case where the variables are binary.

- When variables are categorical, taking values in a finite set:

  - Multinomial/Discrete Graphical Models (Jalali, **R.**, Vasuki, Sanghavi, 2011)

- When variables are thin-tailed continuous

  - Gaussian Graphical Models (**R.**, Raskutti, Wainwright, Yu, 2012)

- **Similar results** as in the Ising model case: estimate constrained node-conditional distributions, and combine to estimate overall graph

Parametric Graphical Models

- Classical parametric graphical model families —Ising, Multinomial/discrete, Gaussian models

Parametric Graphical Models

- Classical parametric graphical model families —Ising, Multinomial/discrete, Gaussian models
  - suited for binary, categorical/discrete, and thin-tailed continuous data respectively

Parametric Graphical Models

- Classical parametric graphical model families —Ising, Multinomial/discrete, Gaussian models
    - suited for binary, categorical/discrete, and thin-tailed continuous data respectively
- What if we have data that does not fall into these categories: skewed continuous, or count-valued for instance

Parametric Graphical Models

- Classical parametric graphical model families —Ising, Multinomial/discrete, Gaussian models
  - suited for binary, categorical/discrete, and thin-tailed continuous data respectively
- What if we have data that does not fall into these categories: skewed continuous, or count-valued for instance
  - Are there more general parametric graphical model families?

Parametric Graphical Models

- Classical parametric graphical model families —Ising, Multinomial/discrete, Gaussian models
  - suited for binary, categorical/discrete, and thin-tailed continuous data respectively
- What if we have data that does not fall into these categories: skewed continuous, or count-valued for instance
  - Are there more general parametric graphical model families?
  - Exponential Family Graphical Models (Yang, **R.**, Allen, Liu 2012, 2014)

Recap: Classical Parametric Graphical Models

- Ising Models
  - node-conditional distribution: Bernoulli

Recap: Classical Parametric Graphical Models

- Ising Models
  - node-conditional distribution: Bernoulli
- Multinomial/Discrete Graphical Models
  - node-conditional distribution: Multinomial

Recap: Classical Parametric Graphical Models

- Ising Models
  - node-conditional distribution: Bernoulli
- Multinomial/Discrete Graphical Models
  - node-conditional distribution: Multinomial
- Gaussian Graphical model
  - node-conditional distribution: univariate Gaussian

Recap: Classical Parametric Graphical Models

- Ising Models
  - node-conditional distribution: Bernoulli
- Multinomial/Discrete Graphical Models
  - node-conditional distribution: Multinomial
- Gaussian Graphical model
  - node-conditional distribution: univariate Gaussian
- Perhaps there's a pattern here …

Background: Exponential Family Distributions

- Most common **univariate** distributions: Gaussian, Exponential, Bernoulli, Binomial, Poisson, Negative binomial, ...
- A broad class of distributions sharing a certain form:

$$P(X;\theta) = \exp\left\{\sum_{i\in\mathcal{I}} \theta_i B_i(X) + C(X) - A(\theta)\right\}$$

- Ingredients:

$\theta = \{\theta_i\}_{i\in\mathcal{I}}$	Parameters
$B(X) = \{B_i(X)\}_{i\in\mathcal{I}}$	Sufficient statistics
$C(X)$	Base measure
$A(\theta) = \log\left\{\sum_X \exp\langle\theta, B(X)\rangle + C(X)\right\}$	Log-partition function

Towards Exponential Family Graphical Models

- Suppose each node-conditional distribution is specified by *some* exponential family distribution:

$$
\begin{aligned}
P(X_s|X_{V\setminus s}) \quad &= \quad \exp\{E_s(X_{V\setminus s})\, B_s(X_s) + C_s(X_s) - \bar{A}_s(X_{V\setminus s})\} \\
E_s(X_{V\setminus s}) \quad & \quad \text{Parameters} \\
B_s(X) \quad & \quad \text{Sufficient statistics} \\
C_s(X) \quad & \quad \text{Base measure} \\
\bar{A}_s(\theta) \quad & \quad \text{Log-partition function}
\end{aligned}
$$

- **Key Question:** Does there exist a consistent joint distribution, and if so, is it unique?

Exponential Family Graphical Models

- **Theorem (Yang, R., Allen, Liu, 2012):** Suppose node-conditional distributions are specified by exponential family distributions as in previous slide. Then there exists a unique joint distribution consistent with these node-conditional distributions, and moreover it takes the following form:

$$P(X) = \exp\Bigg\{ \sum_{s} \theta_s B_s(X_s) + \sum_{s \in V} \sum_{t \in N(s)} \theta_{st} B_s(X_s) B_t(X_t) + \ldots + \sum_{s \in V} \sum_{t_2, \ldots, t_k \in N(s)} \theta_{s \ldots t_k} B_s(X_s) \prod_{j=2}^{k} B_{t_j}(X_{t_j}) + \sum_{s} C_s(X_s) - A(\theta) \Bigg\}$$

Exponential Family Graphical Models

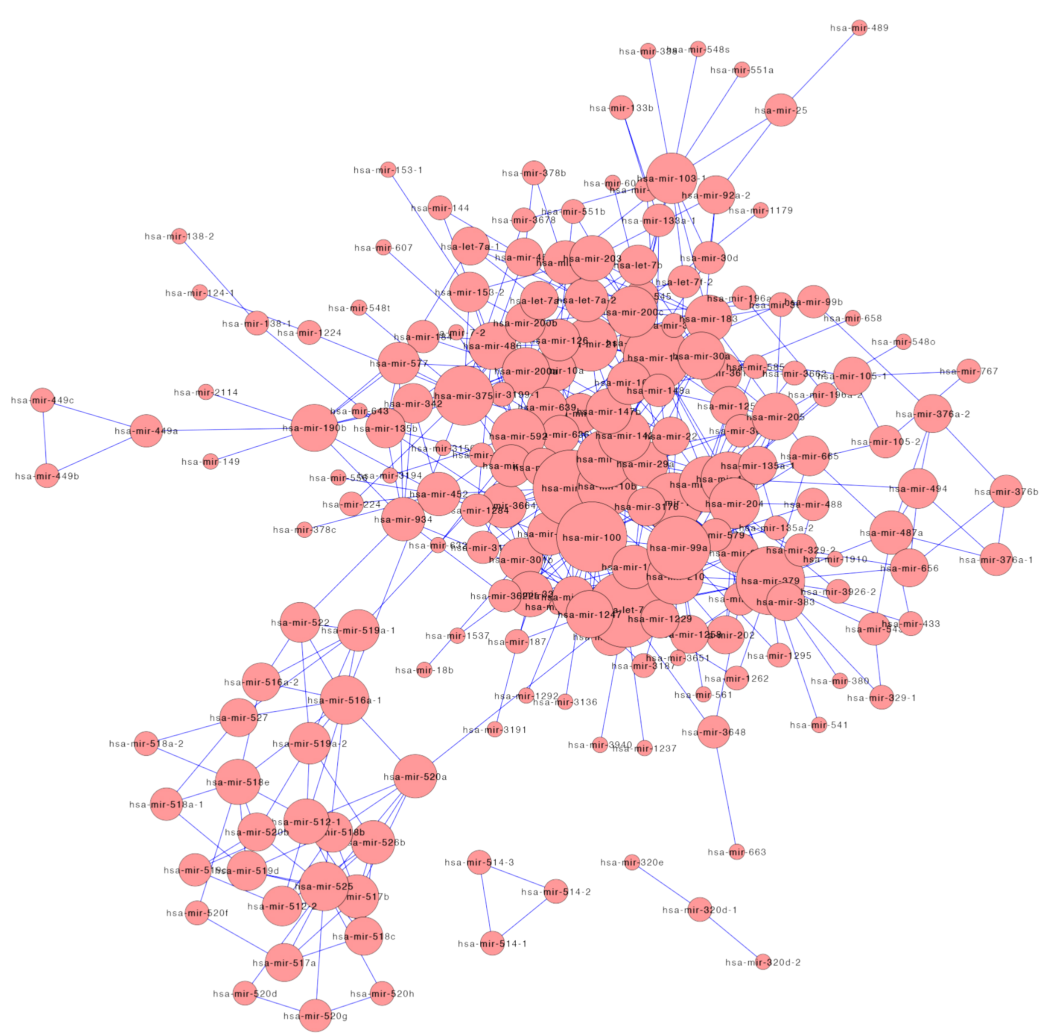
- **Theorem (Yang, R., Allen, Liu, 2012):** Suppose node-conditional distributions are specified by exponential family distributions as in previous slide. Then there exists a unique joint distribution consistent with these node-conditional distributions, and moreover it takes the following form:

$$P(X) = \exp\Bigg\{ \sum_{s} \theta_s B_s(X_s) + \sum_{s \in V} \sum_{t \in N(s)} \theta_{st}\, B_s(X_s) B_t(X_t) + \ldots + \sum_{s \in V} \sum_{t_2, \ldots, t_k \in N(s)} \theta_{s \ldots t_k}\, B_s(X_s) \prod_{j=2}^{k} B_{t_j}(X_{t_j}) + \sum_{s} C_s(X_s) - A(\theta) \Bigg\}$$

- The joint distribution moreover is a graphical model distribution with respect to a graph G specified by the local Markov independencies satisfied by the node-conditional distributions

Example: Poisson Graphical Models

$$P(X) = \exp\left\{ \sum_s \theta_s X_s + \sum_{(s,t)\in E} \theta_{st} X_s X_t + \sum_s \log(X_s!) - A(\theta) \right\}.$$



- MicroRNA network learnt from The Cancer Genome Atlas (TCGA) Breast Cancer Level II Data

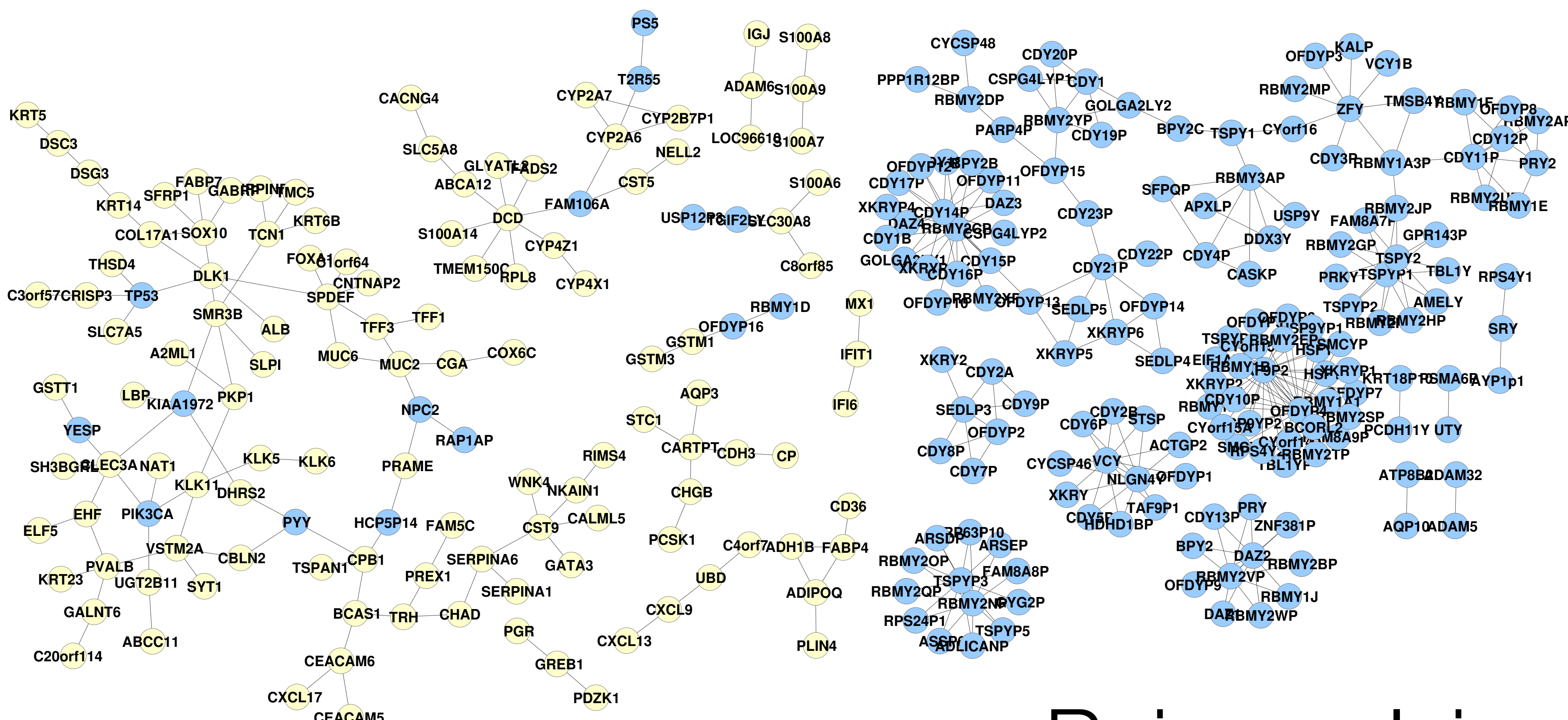
Example: Mixed Graphical Models

$$P(Y,Z) \propto \exp\Big\{ \sum_{s\in V_Y} \theta_s^Y Y_s + \sum_{s'\in V_Z} \theta_{s'}^z Z_{s'} + \sum_{(s,t)\in E_Y} \theta_{st}^{yy} Y_s Y_t$$
$$+ \sum_{(s',t')\in E_Z} \theta_{s't'}^{zz} Z_{s'} Z_{t'} + \sum_{(s,s')\in E_{YZ}} \theta_{ss'}^{yz} Y_s Z_{s'} - \sum_{s\in V_Y} \log(Y_s!) \Big\}.$$

Poisson-Ising Models

Example: Mixed Graphical Models

- Combine 'Level III RNA-sequencing' data and 'Level II non-silent somatic mutation and level III copy number variation data' for 697 breast cancer patients.



Poisson-Ising Models

- (Yellow) Gene expression via RNA-sequencing, count-valued
- (Blue) Genomic mutation, binary mutation status

Learning Exponential Family Graphical Models

- By construction, estimating exponential family graphical models is equivalent to estimate node-conditional univariate exponential family distributions

Learning Exponential Family Graphical Models

- By construction, estimating exponential family graphical models is equivalent to estimate node-conditional univariate exponential family distributions

- Graph Structure Learning Procedure:

  - Estimate graph-structure constrained node-conditional distributions, and estimate node-neighborhoods

  - Stitch node-neighborhoods together to form global graph estimate

Learning Exponential Family Graphical Models

- By construction, estimating exponential family graphical models is equivalent to estimate node-conditional univariate exponential family distributions

- Graph Structure Learning Procedure:

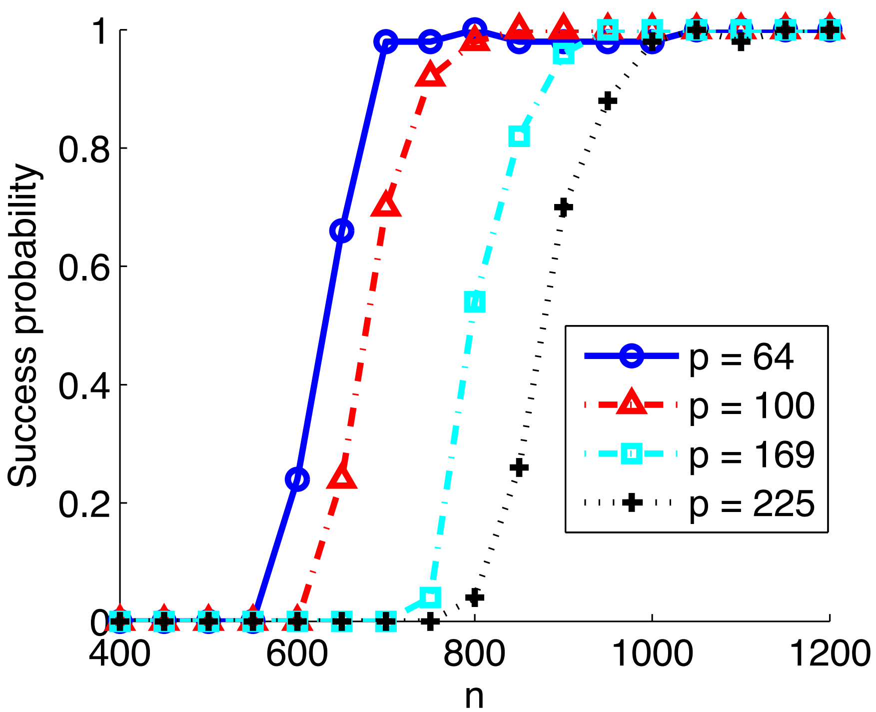
  - Estimate graph-structure constrained node-conditional distributions, and estimate node-neighborhoods

  - Stitch node-neighborhoods together to form global graph estimate

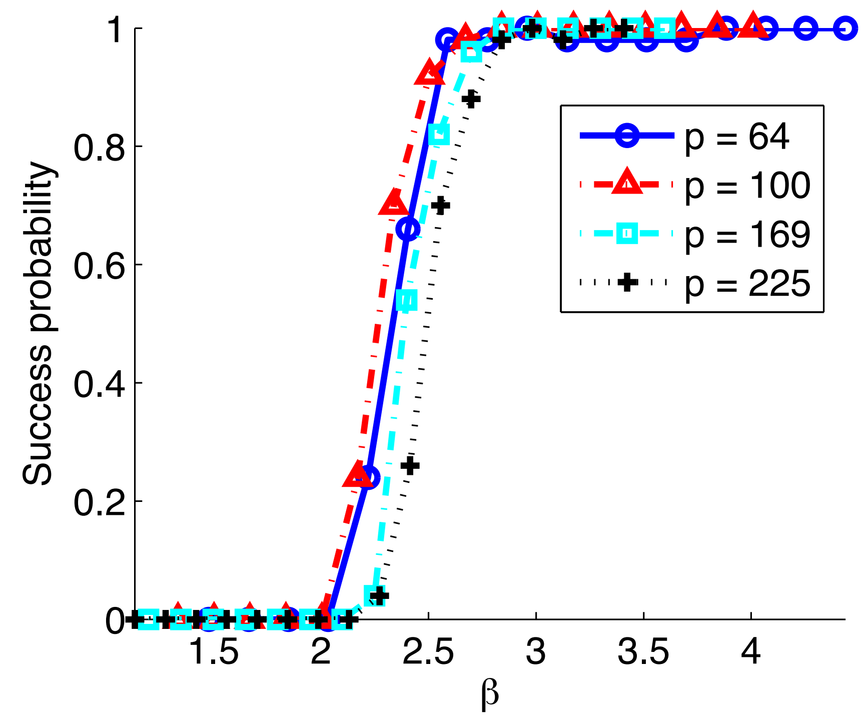
- Similar statistical guarantees for graphical model structure recovery as in Ising, Gaussian graphical model case can be showed even under this general setting (Yang, R., Allen, Liu 2014)

Experiments: Poisson Graphical Models

- Poisson Graphical Model: 4NN Grid structure



Prob. of successful graph recovery vs. number of samples n



Prob. of successful graph recovery vs. re-scaled sample size $\beta = n/(c \log p)$

Thank You!