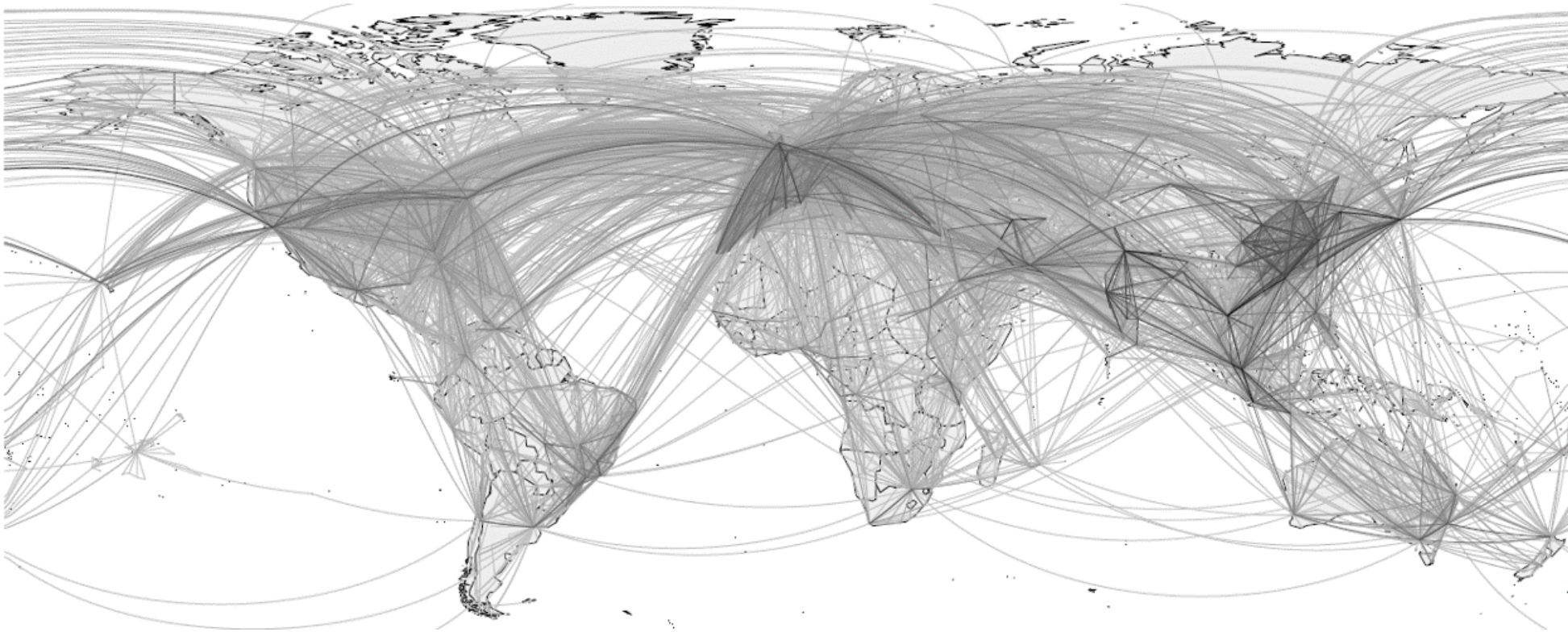
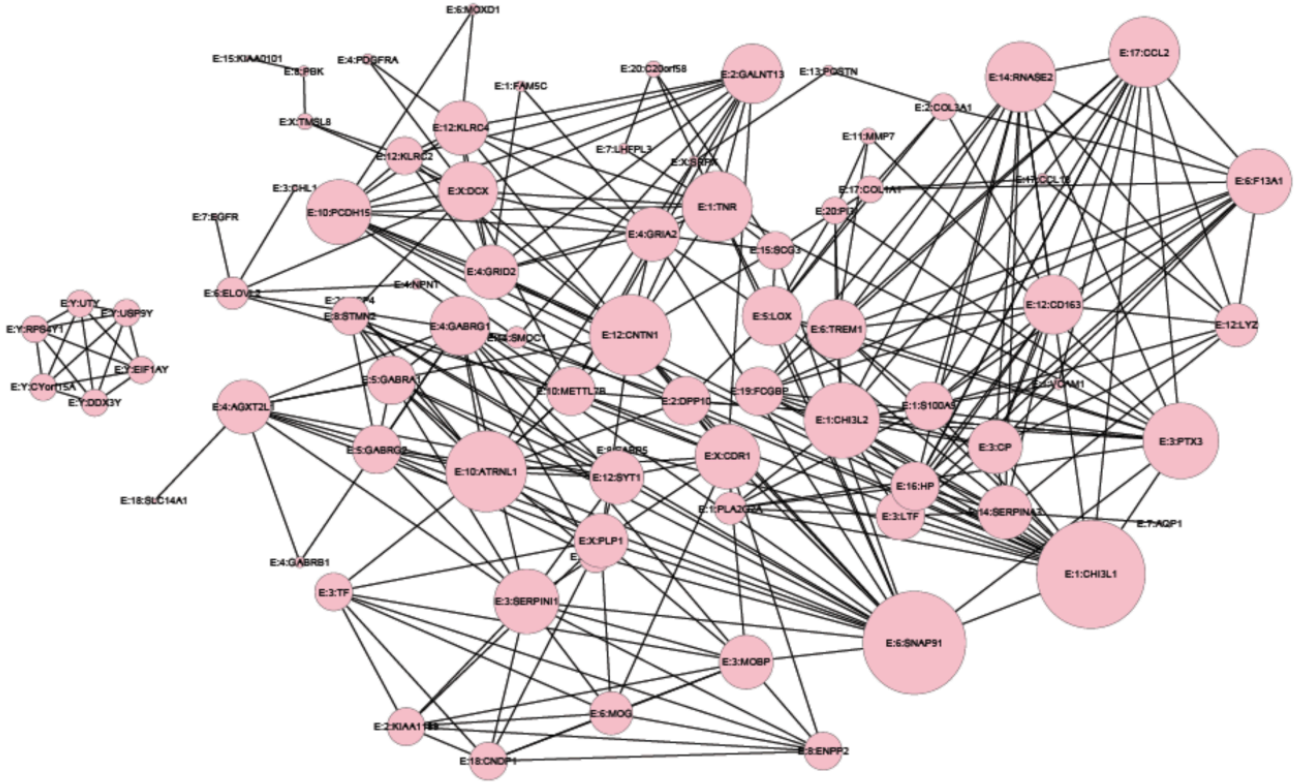
Unsupervised Analysis: Graphical Models

Networks



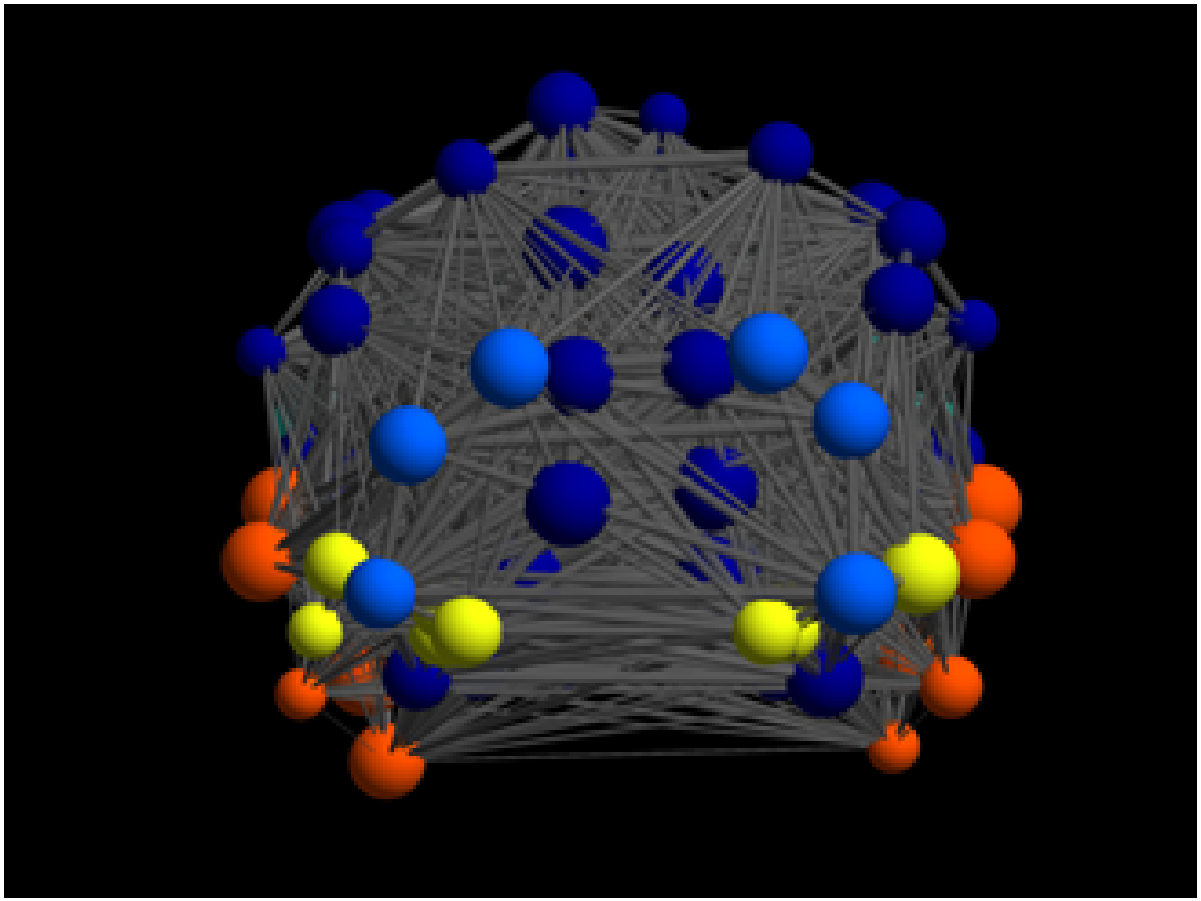
Depicts relationships (edges) between features (nodes).

Networks



Genomics: Relationships between genes.

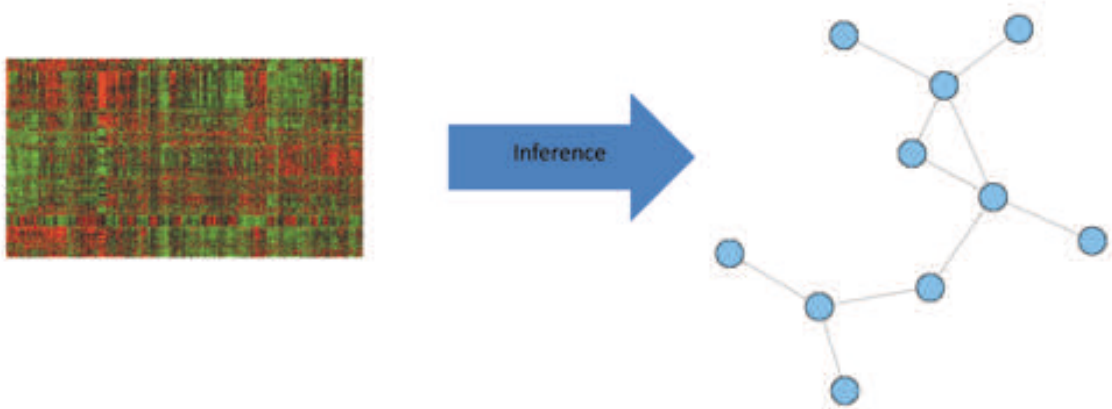
Networks



Neuroimaging: Relationships between brain regions.

Networks in Unsupervised Learning

1. Network Data.
   - Examples: Social networks, twitter, citations, surveillance, web links, etc.
2. **Our focus**: Learn network from data (structural network learning).
   - Data matrix: $\boldsymbol{X}_{n \times p}$.
   - Features form p nodes.
   - Goal: Learn edges between nodes (i.e. learn the relationships between features).



Graphical Models

Probabilistic Graphical Models

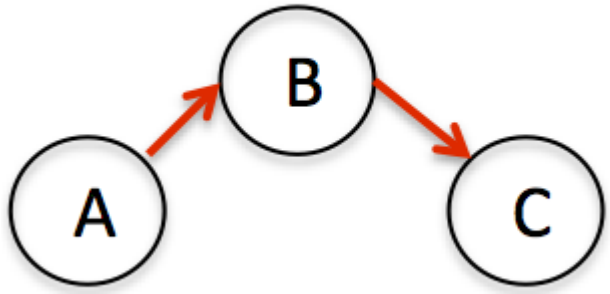
Joint multivariate probability distribution where dependencies can be represented as a network.

Advantages:

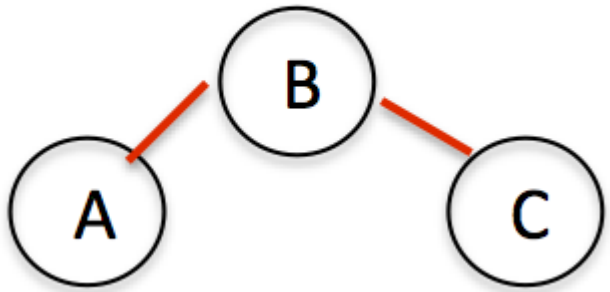
- Graphical models offer efficient factorized forms for joint distributions with easily interpretable dependencies.
  - **Conditional dependencies** denoted via an edge in network.
- Convenient visual representation.

Two Major Types of Graphical Models

1. Undirected - Markov Networks.
   - **Our Focus!**

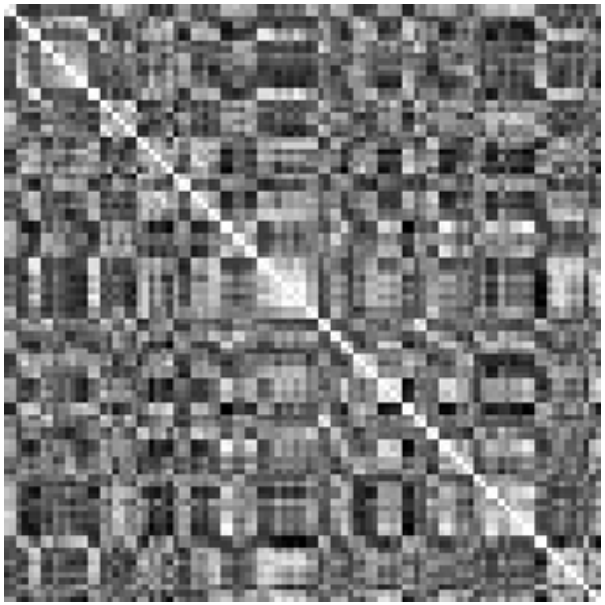


2. Directed - Bayesian Networks.
   - Directed Acyclic Graphs.

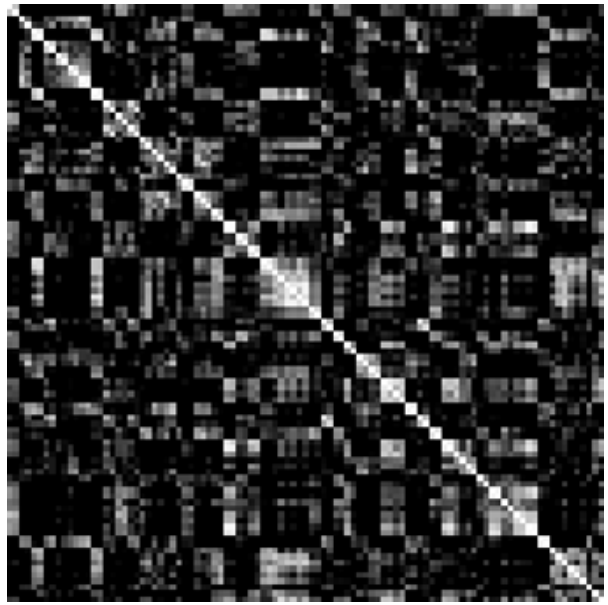


Other Network Estimation Strategies (Undirected)

- Correlation Networks (association networks).



Correlation matrix.



Thresholded correlation matrix.

- Mutual information networks.

NOT Graphical Models!

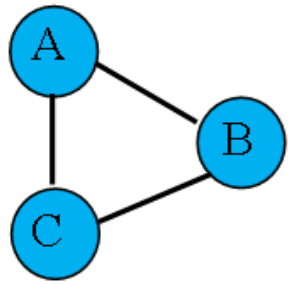
- Even though models are sometimes depicted as a network, they are not always graphical models (e.g. decision trees, neural networks, HMM, etc.).

Markov Networks

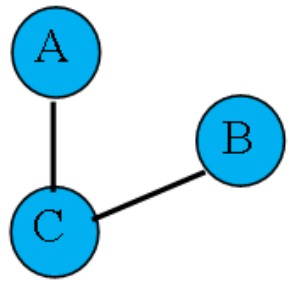
Markov Network

An *undirected graphical model* that characterizes conditional dependence (direct) relationships.

- *Edge*: Two nodes are **conditionally dependent**.
- *No edge*: Two nodes are **conditionally independent**.
- Conditions on all other nodes.



$$A \perp B \mid C$$



Markov Networks - Conditional Dependence

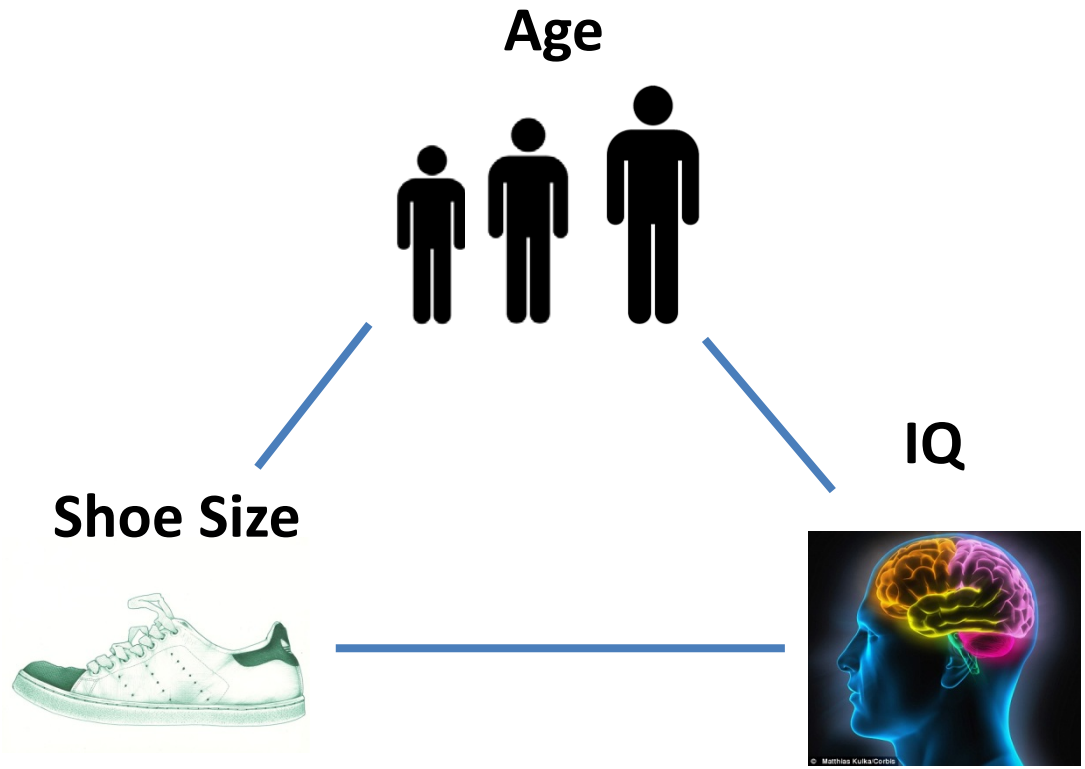
Regression Interpretation:

- Imagine trying to predict the observations in Node A (response) by the observations of all other nodes (predictors).

- Node B predictive of Node A (with all other nodes in model).
  - A is conditionally dependent on B.
  - Edge.

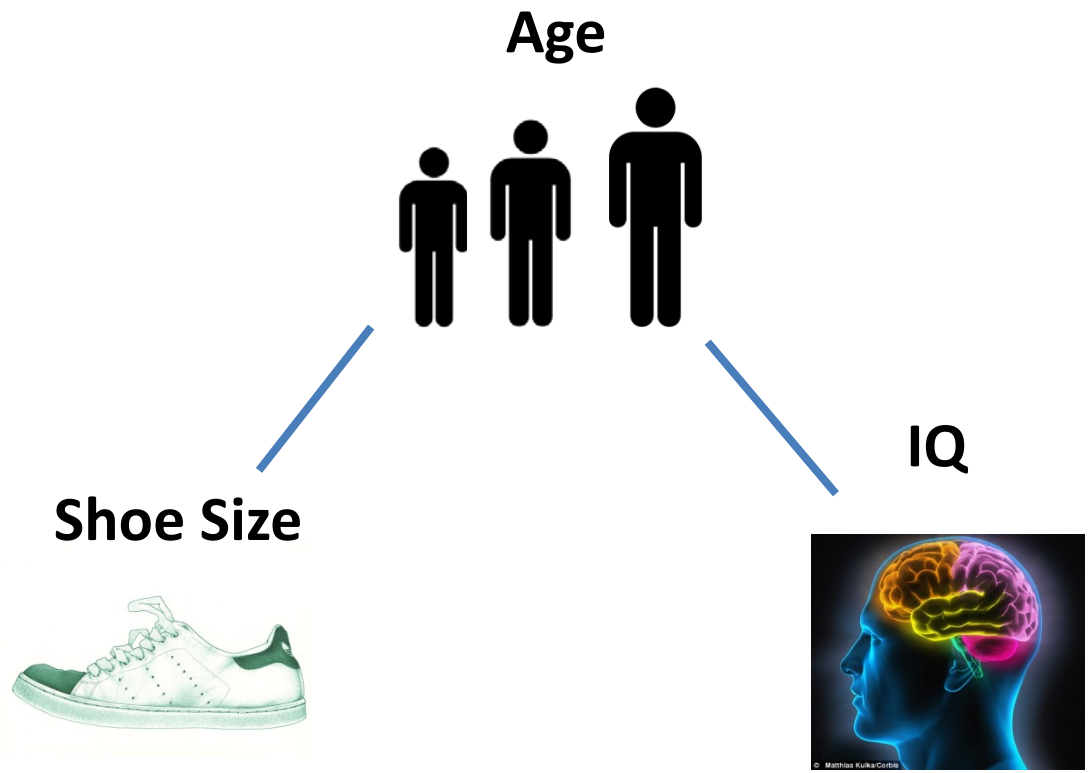
- Because of other nodes in model, Node B does not add any predictive value for Node A.
  - A is conditionally independent of B.
  - No Edge.

Markov Networks - Conditional Dependence



Correlation.

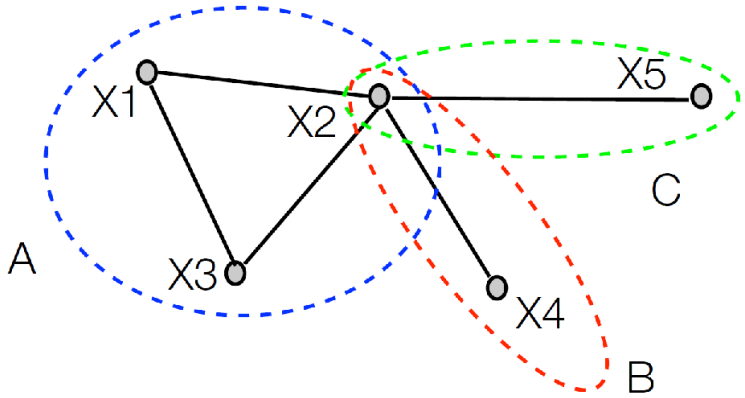
Markov Networks - Conditional Dependence



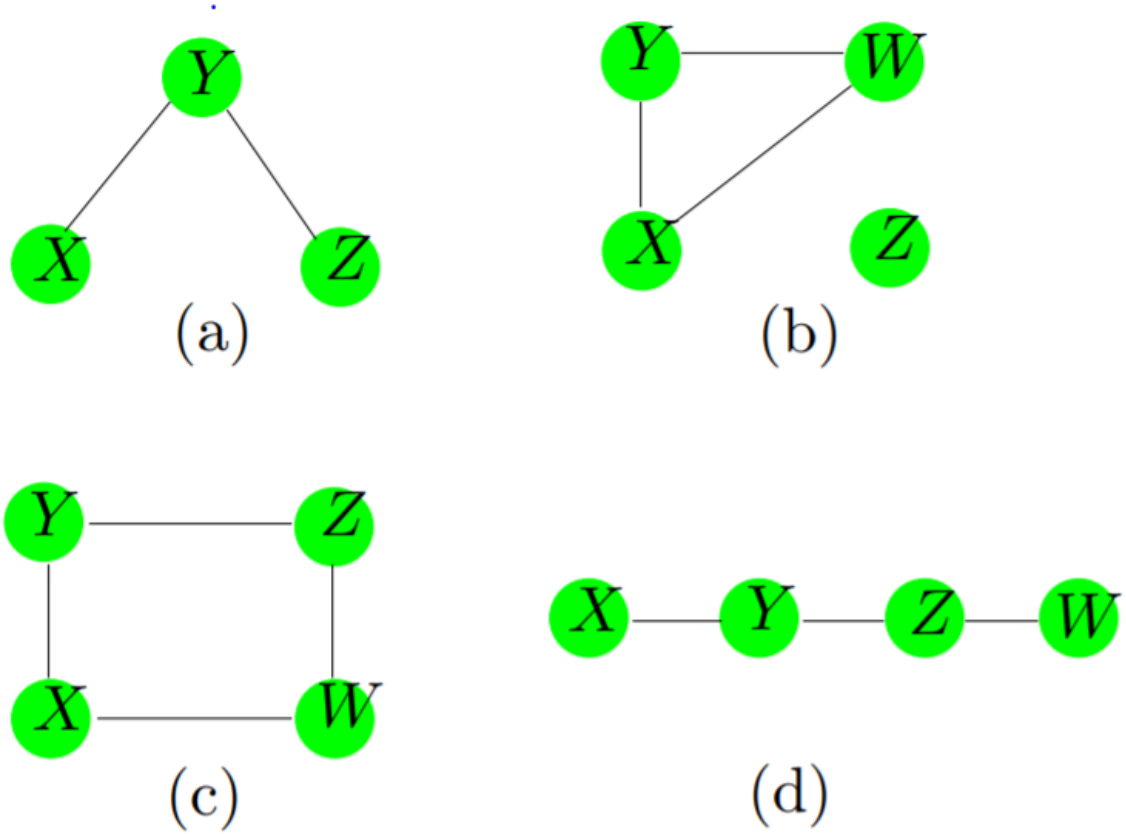
Conditional Dependence (proportional to Partial Correlation).

Markov Networks

- Local Markov Property: Conditional dependencies defined by node-neighborhoods, or the set of nodes connected to a given node via an edge.
- Global Markov Property: Pairwise conditional dependencies and neighborhoods jointly define the global dependence structure (formally defined by separators).
- Hammersley-Clifford Theorem: Density on graph factorizes according to sufficient statistics on cliques. (Probabilistic model!)



Markov Networks



- Practice conditional dependence / independence relationships.

Gaussian Graphical Models

Gaussian Graphical Models

- Multivariate normal: $\boldsymbol{X} \sim N(0, \boldsymbol{\Theta}^{-1})$
- $\boldsymbol{\Theta}$ the inverse covariance matrix.
  - Precision or concentration matrix.
  - If standardized, $\boldsymbol{\Theta}$ the *partial correlation* matrix.
    - ★ Partial Correlation: Regress X_1 on $X_3, \ldots X_p$ and get the residual, r_1; do the same for X_2 yielding residual r_2. The partial correlation between X_1 and X_2 is $\text{Cor}(r_1, r_2)$.
- Zeros in $\boldsymbol{\Theta} \Longrightarrow$ conditional independence!
  - Edges correspond to non-zeros in $\boldsymbol{\Theta}$.

Gaussian Graphical Models

Inference Goals:

- Graph Selection (Structural graph learning).
  - Estimate the zeros in $\mathbf{\Theta}$ (e.g. find all of the edges in the graph).
- Parameter estimation.

Two major algorithms:

1. Graphical Lasso
2. Neighborhood Selection.

Graphical Lasso

Estimate sparse $\mathbf{\Theta}$ via Penalized Maximum Likelihood Estimation (MLE).

Graphical Lasso (Glasso)

$$\underset{\mathbf{\Theta}}{\text{maximize}} \quad \log|\mathbf{\Theta}| - \text{tr}(\boldsymbol{X}^T\boldsymbol{X}\mathbf{\Theta}) - \lambda||\mathbf{\Theta}||_1$$

- Blue: Log-likelihood.
- Red: Penalty that encourages zeros in off-diagonal elements of $\mathbf{\Theta}$.

`R`: `glasso` or `huge` package.

Neighborhood Selection

- Estimate sparse $\mathbf{\Theta}$ via penalized conditional MLE - by estimating zeros in one row / column of $\mathbf{\Theta}$ at a time.
- For each node X_j, find it's node-neighbors (L_1-penalized regression or Lasso):

$$\underset{\beta^j}{\text{minimize}} \quad ||X_j - \boldsymbol{X}_{\neq j}\beta^j||_2^2 + \lambda||\beta^j||_1$$

- Symmetry - β_i^j not always same as β_j^i.
  - Min or max rule.

`R`: `huge` package.

Graph Sparsity

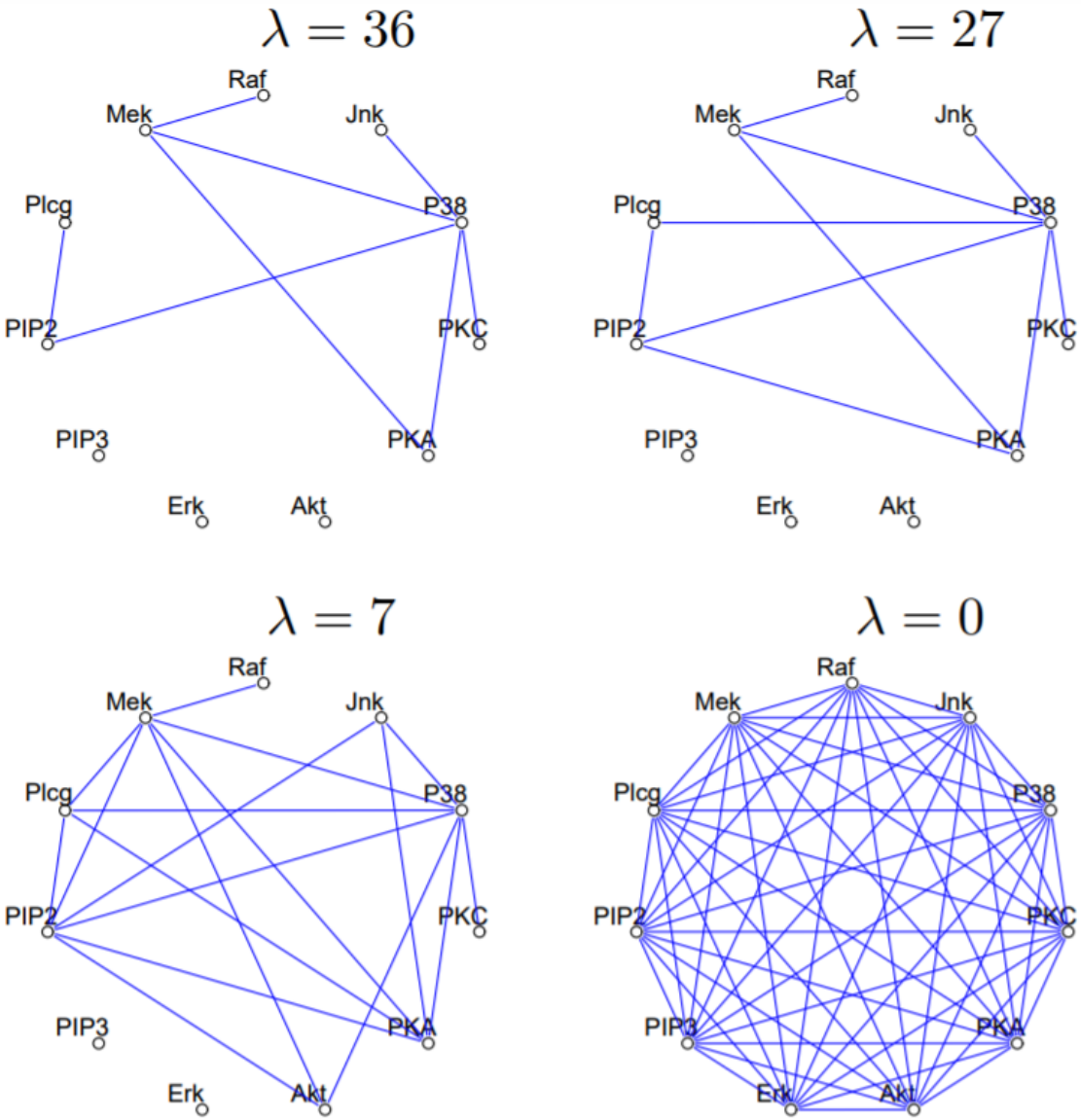
λ Controls Sparsity.

$$\underset{\boldsymbol{\Theta}}{\text{maximize}} \quad \log|\boldsymbol{\Theta}| - \text{tr}(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\Theta}) - \lambda||\boldsymbol{\Theta}||_1$$

- $\lambda = 0$ gives a dense network (no sparsity).
- As λ increases, network becomes more sparse.

- Modulates trade-off between model fit and network sparsity.

Graph Sparsity



How to Choose λ?

- Cross-Validation - tends to yield overly dense networks.
- Extended BIC - adjusted BIC for high-dimensional settings.

Stability Selection

- Idea: Choose λ that gives the most **stable network**.

Procedure:

1. Repeatedly re-sample (bootstrapping or sub-sampling) observations.
2. Choose λ that results in the smallest network variability across re-samples.
3. *Stability Score*: For each edge, the proportion of re-samples in which edge was selected.

`R`: `huge` package.

Other Types of Graphical Models

Non-parametric GGMs

Rank-based Graphical Models:

- $\tilde{\mathbf{\Sigma}}$ a non-parameteric estimator of $\mathbf{\Sigma}$ or the correlation matrix.
  - Spearman's ρ.
  - Kendall's τ.
- Plug $\tilde{\mathbf{\Sigma}}$ into the glasso algorithm in place of $\hat{\mathbf{\Sigma}}$.

Copula Graphical Models:

- Assume $f(X_1), \ldots, f(X_p)$ follows a GGM.
- Transform data via $f()$ and then fit a GGM.
- Typically use a Gaussian copula model for $f()$.

Other extensions of GGMs

- Multiple Graphical Models.
  - For groups of observations, estimate graphical models with both shared structure across groups and individual structure within groups.
- Time Varying Graphical Models.
  - Smoothly varying graph over time estimated via local kernel smoothers.
  - Change points in graph structure over time estimated via fusion penalties.
- Latent Variable Graphical Models
  - Assume observed features are dependent on latent variables which exhibit a low-rank effect. Estimate a sparse (graph structure) plus low-rank inverse covariance matrix.

Graphical Models via Exponential Families

Key Idea: What if we can leverage univariate distributions appropriate for different data types?

Examples:

- Gaussian, Bernoulli, Poisson, Binomial, Negative Binomial, Exponential, ...

$$P(Z) = \exp\left(\theta\, B(Z) + C(Z) - D(\theta)\right)$$

- θ is the canonical parameter.
- $B(Z)$ is the sufficient statistic.
- $C(Z)$ is the base measure.
- $D(\theta)$ is the log-partition function.

Graphical Models via Exponential Families

Assumption:

- Node-conditional distributions are univariate exponential family densities.

Joint distribution has the following form:

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

Graph Selection and Estimation:

- Neighborhood Selection via ℓ_1-penalized Generalized Linear Models.

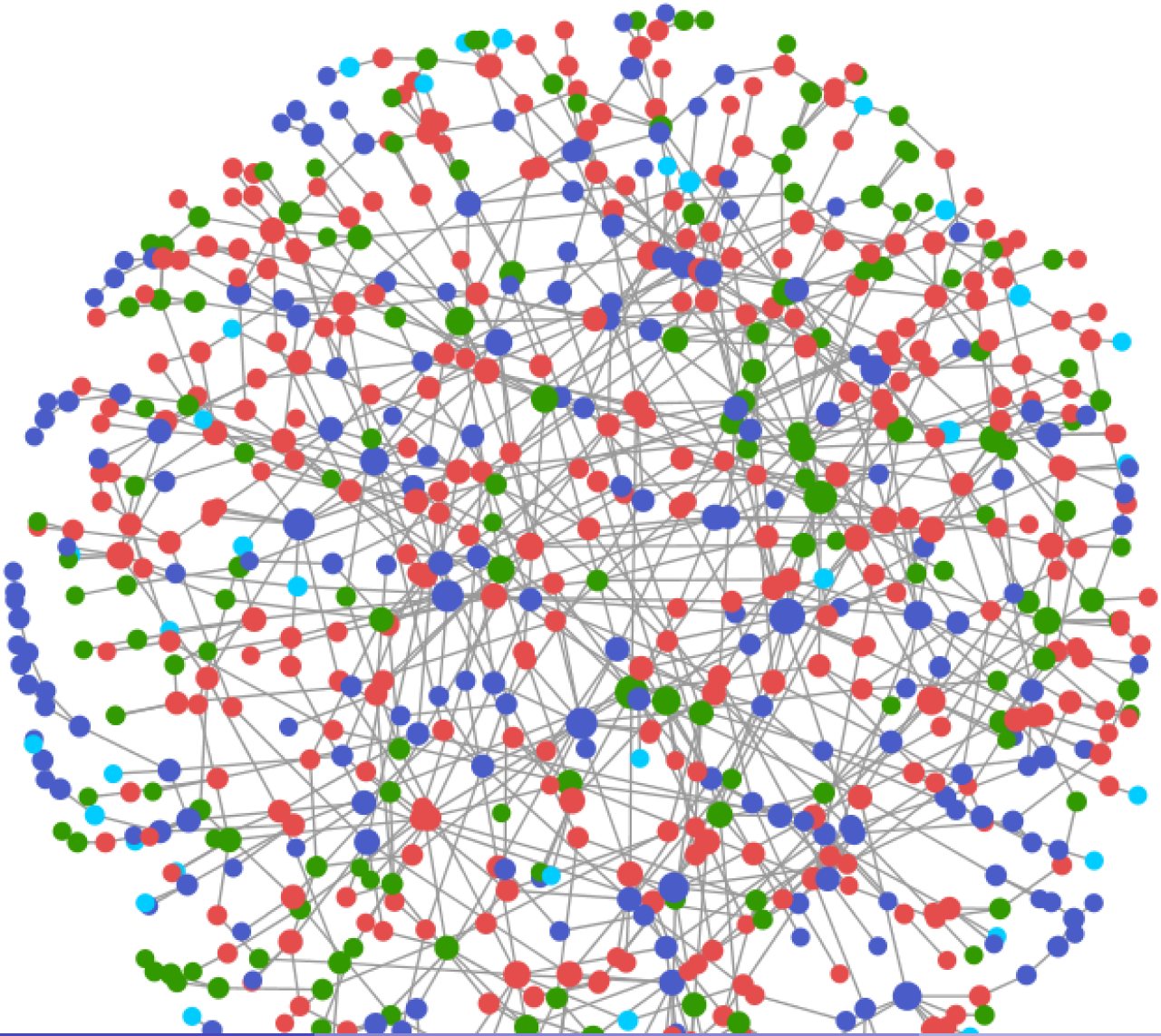
Graphical Models via Exponential Families

Special Cases:

- Gaussian Graphical Model.
- Ising Model.
  - Assume conditional distributions are Bernoulli.
  - Graphical Models for binary data.
- Poisson Graphical Model.
  - Assume conditional distributions are Poisson.
  - Graphical Models for count data.
  - Some restrictions on types of dependencies; unrestricted dependencies require modified Poisson-like distributions.

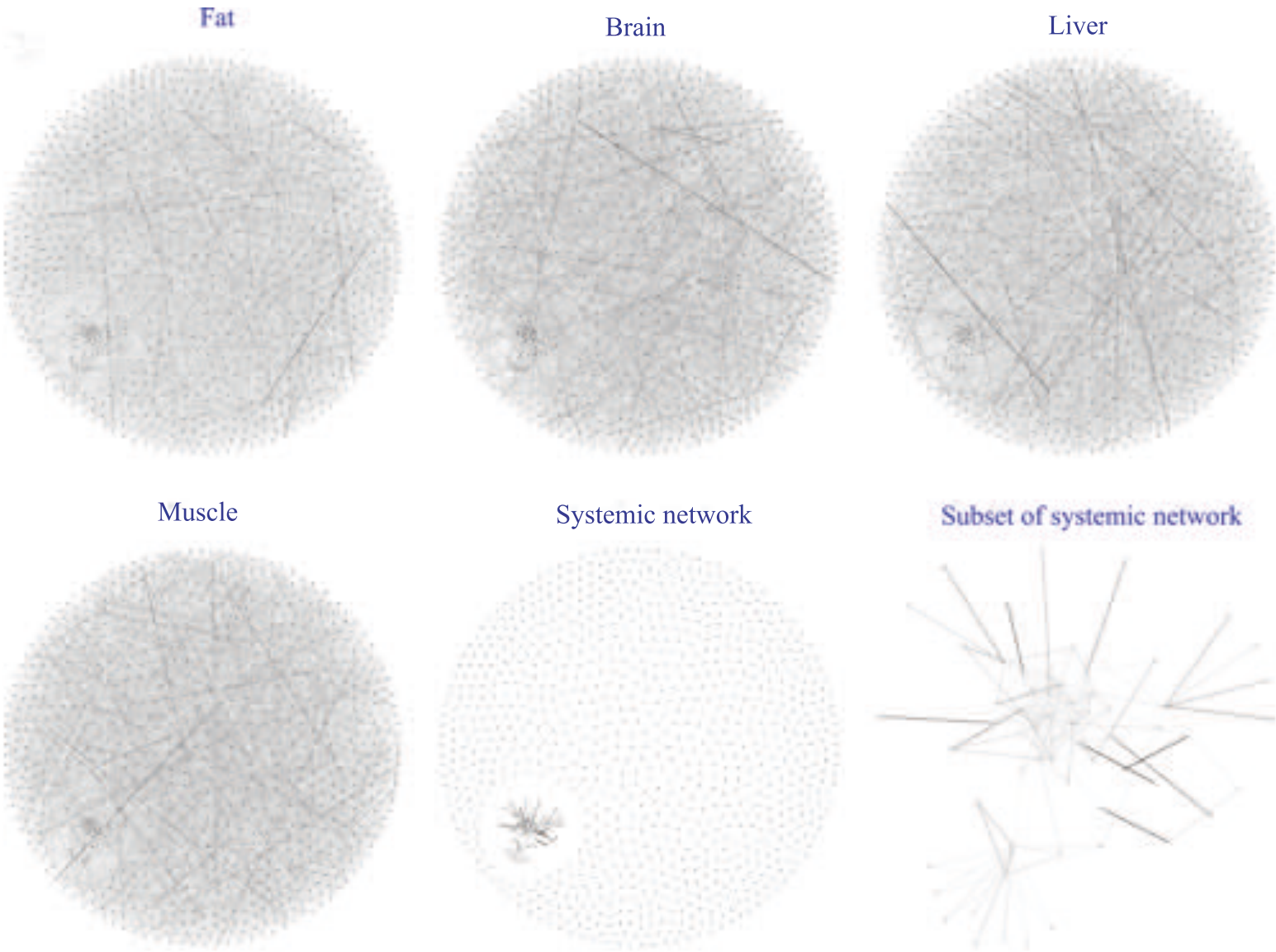
Research Highlight

Mixed Graphical Models



Research Highlight

Multiple Gaussian Graphical Models



References

Textbooks:

- Elements of Statistical Learning by Hastie, Tibshirani & Friedman. `http://statweb.stanford.edu/~tibs/ElemStatLearn/`

Some of the figures in this presentation are taken from this textbook with permission from the authors.