Interaction Network Analysis

Young-Rae Cho
Assistant Professor
Department of Computer Science
Baylor University

Biological Networks

- **Definition**
  - Maps of biochemical reactions, interactions, regulations between genes or proteins

- **Importance**
  - Show the mechanisms of molecular function in a cell
  - Key resource for functional characterization of molecules in a systematic view

- **Examples**
  - Metabolic networks
  - Protein-protein (or genetic) interaction networks
  - Gene regulatory networks
  - Signal transduction networks
Overview

- Backgrounds
- Interaction Network Modeling
- Functional Module Detection
- Signaling Pathway Prediction
- Essential Gene Selection
- Conserved System Identification

Protein-Protein Interactions (1)

- **Protein-Protein Interactions**
  - Physical interactions to form protein complexes
  - Functional relationships to perform the same molecular functions

- **Interactome**
  - The entire set of protein-protein interactions in the genome scale
  - Problem: large scale
  - Requires systematic, computational analysis
Protein-Protein Interactions (2)

- **Protein Structures**
  - Primary Structure
  - Secondary Structure: α-helix, β-strand
  - Tertiary Structure
  - Quaternary Structure
  - Protein-protein interactions

- **Protein Interaction Networks**
  - Interaction Networks
    - Undirected, unweighted graph representation with a set of nodes $V$ as proteins and a set of edges $E$ as interactions between them
    - Problem: complex connectivity
  - Example
Interaction Determination (1)

➢ Types of Interactions
  ▪ Permanent (stable) interactions
  ▪ Transient interactions

➢ Experimental Methods
  ▪ Mass spectrometry
    ▪ Identification of components in protein complexes
  ▪ Two-hybrid system
    ▪ Determination of binary protein-protein interactions
    ▪ Problem: a large number of false positives
  ▪ Protein microarray

Interaction Determination (2)

➢ Computational Methods
  ▪ Sequence-based approaches
    ▪ Phylogenetic profile analysis
    ▪ Gene fusion analysis
    ▪ Gene neighborhood analysis
  ▪ Structure-based approaches
    ▪ Homolog search
    ▪ Interface similarity search
  ▪ Expression-based approaches
    ▪ Expression profile correlation
  ▪ Interaction-based approaches
    ▪ Interaction prediction from known interactions
Phylogenetic Profile Analysis (1)

- **Main idea**
  - if two proteins in one organism have orthologs in another organism, they are likely to interact with each other and be functionally linked

![Genome and Phylogenetic Profile Diagram]

**Profile Clusters**
- P4: 1 0 0
- P1: 1 0 1
- P5: 1 1 1
- P3: 0 1 1
- P6: 0 1 1
- P2: 1 1 0
- P7: 1 1 0

**Conclusion**
- P2 and P7 are functionally linked
- P3 and P6 are functionally linked

Phylogenetic Profile Analysis (2)

- **Process**
  - Input: a pair of proteins
  - Find orthologs across species
  - Build scoring matrices
  - Calculate correlation between two proteins

![Protein Sequence Alignment and Scoring Matrix Diagram]
Gene Fusion Analysis (1)

Main Idea
- If two proteins in different organisms have orthologs in one organism in consecutive way, they are likely to interact with each other and be functionally linked.

Gene Fusion Analysis (2)

Process
- Input: a query genome (Q) and a reference genome (R)
- Detect all similarity within Q
- Build a matrix T
- Detect similarity between Q & R
- Build a matrix Y
- Detect Fusion using an Algorithm:
  - Find A, B, C such that A is similar to C, B is similar to C, and A is not similar to B
Interaction-based Approaches

- **Common Neighbor Analysis**
  - If two proteins have many common interacting partners, they are likely to interact with each other.
  - Jaccard coefficient:
    \[ S(x, y) = \frac{|N(x) \cap N(y)|}{|N(x) \cup N(y)|} \]
  - Geometric coefficient:
    \[ S(x, y) = \frac{|N(x) \cap N(y)|^2}{|N(x)| \cdot |N(y)|} \]
  - Dice coefficient:
    \[ S(x, y) = \frac{2|N(x) \cap N(y)|}{|N(x)| + |N(y)|} \]
  - Simpson coefficient:
    \[ S(x, y) = \frac{|N(x) \cap N(y)|}{\min(|N(x)|, |N(y)|)} \]
  - Hyper-geometric coefficient (P-value):
    \[ P = \frac{\binom{T}{Z} \left( \binom{T-Z}{T-|N(x)|} \binom{T-|N(y)|}{Z} \right)}{\binom{T}{Z} \binom{T-|N(x)|}{T} \binom{T-|N(y)|}{T}} \]

- **Majority-based Method**
  - Assigning majority of functions of interacting partners to the unknown gene.
  - Called "guilt-by-association"

![Diagram of local analysis for function prediction]

**Local Analysis for Function Prediction**

- **Majority-based Method**
  - Assigning majority of functions of interacting partners to the unknown gene.
  - Called "guilt-by-association"
Global Analysis for Function Prediction

- **Extension of Majority-based Method**
  - Assigning functions of all unknown genes in an interaction network
  - Requires optimization

![Interaction Network Diagram]

Overview

- **Backgrounds**
- *Interaction Network Modeling*
- Functional Module Detection
- Signaling Pathway Prediction
- Essential Gene Selection
- Conserved System Identification
Random Graph Model (1)

- Erdős and Rényi, 1960 (E-R Model)
  - Random graph as N nodes connected by m edges that are randomly chosen from N(N-1)/2 possible edges
  - m = \( p[N(N-1)/2] \) where \( p \) is the probability of each pair of nodes being connected
  - Degree distribution \( P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} \)
    - Degree of \( v \): the number of links from \( v \) to other nodes
    - Degree distribution \( P(k) \): probability that a node has \( k \) links
  - Expected number of nodes with degree \( k \)
    \[ E(X_k) = N \cdot p(k) = N \binom{N-1}{k} p^k (1-p)^{N-1-k} = \lambda_k \rightarrow P(X_k = r) = e^{-\lambda_k} \lambda_k^r / r! \]
    (Poisson distribution)

Random Graph Model (2)

- Example of E-R Model
  - Poisson distribution with \( N = 1000 \) and \( p = 0.0015 \)
Barabasi and Albert, 1999 (B-A Model)

- Focused on network dynamics based on these two steps:
  - Growth: networks are continuously expanded by the addition of new nodes with a link to the nodes already present
  - Preferential attachment: the nodes are likely to be linked to high-degree nodes
- Power law degree distribution: \( P(k) \sim k^{-\gamma} \) where \( 2 < \gamma < 3 \)
- Features
  - A very few high-degree nodes and many low-degree nodes → hub-oriented network structure
  - Very low average shortest path length → small-world network

Scale-Free Network Model (2)

- Example of B-A Model
  - Power-law degree distribution with the best fit of \( \gamma = 2.9 \) on the dashed line

![Graph showing power-law degree distribution with a best fit line at \( \gamma = 2.9 \).]
Modular Network Model

- **Modular Networks**
  - Verified by high average clustering coefficient
    - Density of a graph $G(V,E)$: the number of actual edges over the number of all possible edges, $D(G) = 2|E| / |V|(|V|-1)$
    - Clustering coefficient of a node $v$: the density of a subgraph with neighbors of $v$
  - Dense intra-connections among the nodes in the same modules
  - Sparse inter-connections between two nodes in different modules

Hierarchical Network Model

- **Hierarchical Networks**
  - Integrated of scale-free topology with modular structure
  - Hierarchy of modules is controlled by hubs
  - Clustering coefficient distribution $C$
    - Scale-free network & Modular network: $C$ is independent of degree $k$
    - Hierarchical network: $C \sim k^{-1}$
Schematic View

A | B | C

Scale-free Networks | Modular Networks | Hierarchical Networks

Overview

- Backgrounds
- Interaction Network Modeling
- Functional Module Detection
- Signaling Pathway Prediction
- Essential Gene Selection
- Conserved System Identification
Clustering Interaction Network

- **Protein Complex**
  - A group of proteins having physical interactions at the same place, same time

- **Functional Module**
  - A group of proteins having the same function
  - A group of proteins having interactions even at different place, different time

- **Protein Complex / Functional Module Detection**
  - Clustering interaction networks by graph clustering algorithms
  - Detecting densely connected subgraphs
  - Grouping densely connected proteins and adding peripheral proteins

Overview

- Backgrounds
- Interaction Network Modeling
- Functional Module Detection

- **Signaling Pathway Prediction**
- Essential Gene Selection
- Conserved System Identification
Predicting Pathways on Interaction Network

- **Signaling Pathway**
  - A series of proteins having signaling and response relationship

- **Signaling Network**
  - A combined form of linear signaling pathways
  - A directed acyclic graph

- **Signaling Pathway / Signaling Network Prediction**
  - Given starting and ending nodes, searching the strongest paths
  - Searching the most frequent paths
  - Mapping frequent connection patterns and selecting common connected paths

Overview

- Backgrounds
- Interaction Network Modeling
- Functional Module Detection
- Signaling Pathway Prediction
- **Essential Gene Selection**
- Conserved System Identification
Measuring Essentiality of Proteins

- **Essential Genes / Proteins**
  - Functional core genes or proteins in a functional module
  - Hubs in an interaction network
  - Significance in biomedical applications: drug target detection

- **Measurement Essentiality of Genes / Proteins**
  - Using centrality measures
  - Selecting bridging points of functional modules
  - Selecting merging points of signaling pathways

Overview

- Backgrounds
- Interaction Network Modeling
- Functional Module Detection
- Signaling Pathway Prediction
- Essential Gene Selection
- Conserved System Identification
Network Alignment

- **Main Idea & Goal**
  - Aligning two or more evolutionary distal interaction networks to identify evolutionary conserved connection patterns
  - Measure sequential similarity between molecules (orthologs), AND topological similarity between interaction networks

Comparison between Sequence & Network Alignment

- **Sequence Alignment**
  - Aligning two or more sequences
  - Searches matches (identical letters), mismatches (non-identical letters), and gaps
  - Returns alignment in a two-row representation including gaps

- **Network Alignment**
  - Aligning two or more networks
  - Searches matches (orthologs), mismatches (non-orthologs), and gaps
  - Returns an alignment network having ortholog pairs as nodes and conserved interactions as edges
Issues in Network Alignment

- **Technical Issues**
  - How to map two or more networks to detect a common sub-network
  - How to optimize the alignment network for multiple orthologs
  - How to improve efficiency of network alignment

- **Network Alignment Types**
  - Global network alignment
    - Aligning two or more entire networks
  - Local network alignment
    - Detecting maximally (strongly) conserved sub-networks

Questions?

- Lecture Slides are found on the Course Website,
  web.ecs.baylor.edu/faculty/cho/5330