Introduction to biological molecular networks

Anthony Gitter
BMI/CS 576
www.biostat.wisc.edu/bmi576
gitter@biostat.wisc.edu
Fall 2016
Key concepts in networks

• **What** are molecular networks?
  – Graph-theoretic representation
  – Different types of networks

• **Why** are they useful for interpreting biological data?
  – Analysis of global network properties
  – Data integration
  – Hypothesis generation

• **How** can we learn networks from high-throughput data?
  – Classes of methods for expression-based network inference
  – Probabilistic graphical models
Understanding a cell as a system

• **Measure: identify the parts of a system**
  – Parts: different types of bio-molecules
    • Genes, proteins, metabolites
  – High-throughput assays to measure these molecules

• **Model: how these parts are put together**
  – Clustering
    – Network inference and analysis
Omics data provide comprehensive description of nearly all components of the cell.
Representing molecular networks

- Molecular networks typically represented by graphs
- Vertices/Nodes = a molecular part
- Edges = connections or interactions between parts
- Edges can have directionality, signs, and/or weight
Different types of molecular networks

• Depends on what
  – the vertices represent
  – the edges represent
  – whether edges directed or undirected
Transcriptional regulatory networks

Nodes: Regulatory protein like a TF, or target gene
Edges: TF A regulates gene C

Transcription factors (TFs)

E. coli: 153 TFs and 1319 target genes

S. cerevisiae: 157 TFs and 4410 target genes
Transcriptional regulatory networks

Nodes: Regulatory protein like a TF, or target gene
Edges: TF A regulates gene C

Transcription factors (TFs)

Gene C

Directed, weighted

Human: 26,070 high-confidence TF-gene edges
338 high-confidence interactions among 119 TFs (of ~2000)

Vargas and Santillan, 2008
Gerstein et al., 2012, Nature
Detecting protein-DNA interactions

- ChIP-chip and ChIP-seq binding profiles for transcription factors
- Determine the (approximate) locations in the genome where a protein binds

Peter Park, Nature Reviews Genetics, 2009
Protein-protein interaction networks

Vertices: Proteins
Edges: Protein U physically interacts with protein X

Yeast protein interaction network

Barabasi et al. 2004
Metabolic networks

Vertices: Enzymes
Edges: Enzyme M and N share a metabolite

Proteins (enzymes)
Metabolites

Undirected, weighted

Figure from KEGG database
Metabolic networks

Vertices: Enzymes
Edges: Enzyme M and N share a metabolite

Proteins (enzymes)

Metabolites

D-Galactonate

2-Dehydro-3-deoxy-D-galactonate

Enzyme: 4. Lyases
4.2 Carbon-oxygen lyases
4.2.1 Hydro-lyases
4.2.1.6 galactonate dehydratase

Undirected, weighted

Figure from KEGG database
Overview of the *E. coli* Metabolic Pathway Map

Image from the KEGG database
Signaling networks

Vertices: Enzymes and other proteins
Edges: Enzyme P modifies protein Q

Receptors

Directed

Vertices:
- CD3
- CD28
- LFA-1
- VAV
- LAT
- Lck
- Zap70
- SLP-76
- RAS
- Cytohesin
- JAB-1

Edges:
1. \(\alpha\)-CD3
2. \(\alpha\)-CD28
3. ICAM-2
4. PMA
5. \(\beta_2\)cAMP
6. G06976
7. AKT inh
8. Psitect
9. U0126
10. LY294002

Activators
1. \(\alpha\)-CD3
2. \(\alpha\)-CD28
3. ICAM-2
4. PMA
5. \(\beta_2\)cAMP

Inhibitors
6. G06976
7. AKT inh
8. Psitect
9. U0126
10. LY294002

Sachs et al., 2005, Science
Genetic interaction networks

Genetic interaction: If the phenotype of double mutant is significantly different than each mutant alone

Vertices: Genes
Edges: Genetic interaction between query (Q) and gene G

Undirected, signed

Step 1:
Generate double mutant

Saccharomyces cerevisiae

Mating X
Query strain
Deletion library

Step 2:
Score phenotype and identify interactions

1 2
3 4
Measure colony size

Colony 1: 98 pixels
Colony 2: 99 pixels
Colony 3: 95 pixels
Colony 4: 17 pixels

Dixon et al., 2009, Annu. Rev. Genet
Yeast genetic interaction network

23 million double mutants
550k negative and 350k positive genetic interactions

Costanzo et al., 2016, Science
Summary of different types of molecular networks

• Physical networks
  – *Transcriptional regulatory networks*: interactions between regulatory proteins (transcription factors) and genes
  – *Protein-protein*: interactions among proteins
  – *Signaling networks*: interactions between protein and small molecules, and among proteins that relay signals from outside the cell to the nucleus

• Functional networks
  – *Metabolic*: describe reactions through which enzymes convert substrates to products
  – *Genetic*: describe interactions among genes which when genetically perturbed together produce a more significant phenotype than expected
Computational problems in networks

• Analysis of network properties
  – Degree distributions
  – Network motifs
  – Network modules
  – Highly connected nodes and relationship to lethality

• Network applications
  – Interpretation of gene sets
  – Identification of condition-specific subnetworks or pathways
  – Using networks to infer functions of a gene

• Network reconstruction
  – Infer the structure and parameters of networks
  – We will examine this problem in the context of “expression-based network inference”
Subnetwork identification

• Given
  – Scores on nodes from a condition-specific experiment
  – A general background network

• Do
  – Select the relevant nodes and edges from the background network

• Only portions of a network are “active” in a specific condition
  – Stimulus response
  – Cell type
  – Disease

• Many different ways to formalize this optimization problem
  – Shortest paths
  – Network flow
  – Steiner tree
Subnetwork identification example

Scores on proteins, e.g. phosphorylation changes

Protein-protein interaction network

Overlay scores in the network

Predict connections among the high-scoring nodes

Tuncbag et al., 2016, PLOS Computational Biology
Network reconstruction

- Given
  - A set of attributes associated with network nodes
  - Typically attributes are mRNA levels
- Do
  - Infer which nodes interact
- Algorithms for network reconstruction can vary based on their meaning of interaction
  - Similarity
  - Mutual information
  - Predictive ability
Network reconstruction example

Y3K project
Stefely et al., 2016,
Nature Biotechnology
Computational methods to infer networks

• We will focus on transcriptional regulatory networks
• These networks are often inferred from gene expression data
• Many methods to do network inference
  – We will focus on probabilistic graphical models
Modeling a regulatory network

Who are the regulators?

How they determine expression levels?

Structure

Function

Hot1 regulates HSP12
HSP12 is a target of Hot1

HSP12

Ψ(X₁, X₂)

Y

ψ(X₁, X₂)

X₁

X₂

Boolean
Linear
DIFF. EQNS
PROBABILISTIC

….
Mathematical representations of networks

Models differ in the function that maps input system state to output state

Boolean Networks

Differential equations

\[ \frac{dX_3(t)}{dt} = \kappa g(X_1(t), X_2(t)) - rX_3(t) \]

Probabilistic graphical models

\[ P(X_3|X_1, X_2) = N(X_1a + X_2b, \sigma) \]