Pathway & Network Analysis of Omics Data: Networks in Biology

Ali Shojaie
Department of Biostatistics
University of Washington
`faculty.washington.edu/ashojaie`

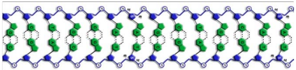
Summer Institute for Statistical Genetics – Australia, 2017

Why Study Networks?

- Components of biological systems, e.g. genes, proteins, metabolites, interact with each other to carry out different functions in the cell.
- Examples of such interactions include signaling, regulation and interactions between proteins.
- We cannot understand the function and behavior of biological systems by studying individual components ($2 + 2 \neq 4!$).
- Networks provide an efficient representation of complex reaction in the cells, as well as basis for mathematical/statistical models for the study of these systems.

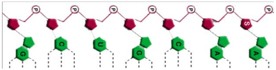
Central Dogma of Molecular Biology (Extended)

DNA



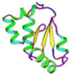
Genomics – 25,000 Genes

RNA



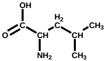
Transcriptomics – 100,000 Transcripts

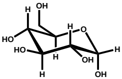
Protein

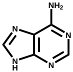


Proteomics – 1,000,000 Proteins

Biochemicals (Metabolites)

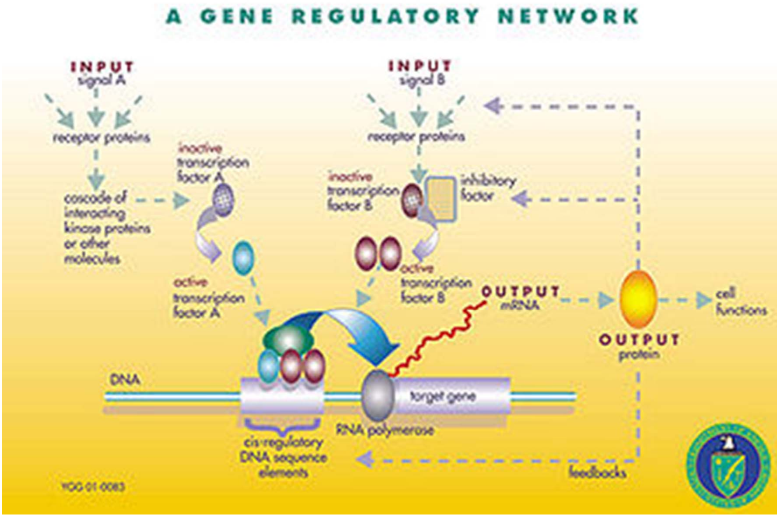




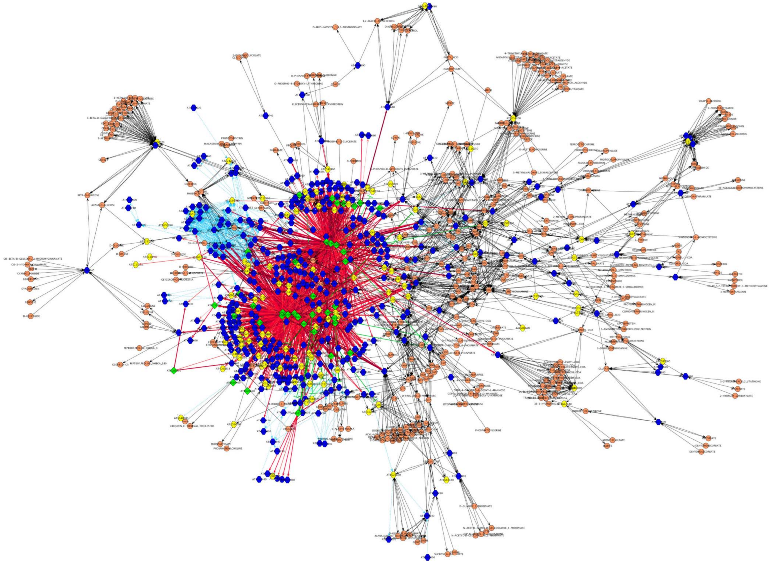


Metabolomics ~2000 Compounds

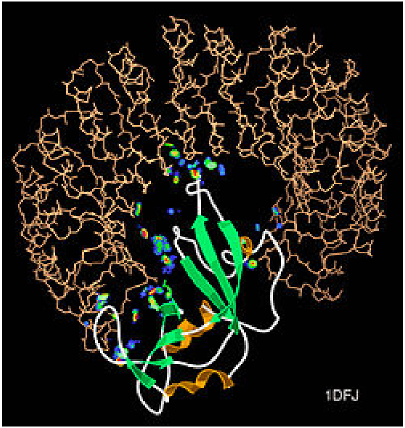
Networks in Biology: Gene Regulatory Interactions



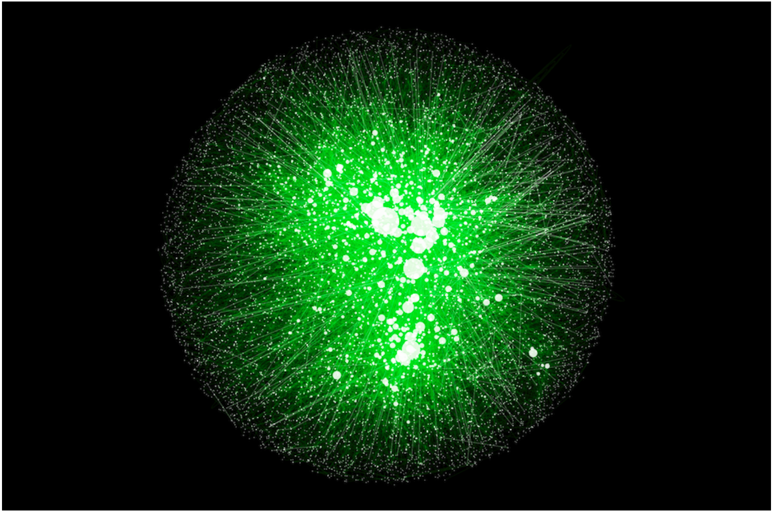
Networks in Biology: Gene Regulatory Networks



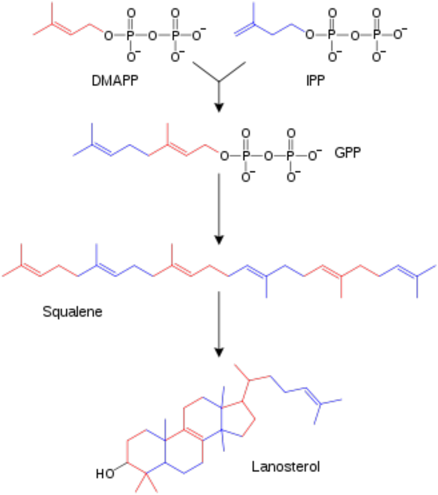
Networks in Biology: Protein-Protein Interactions



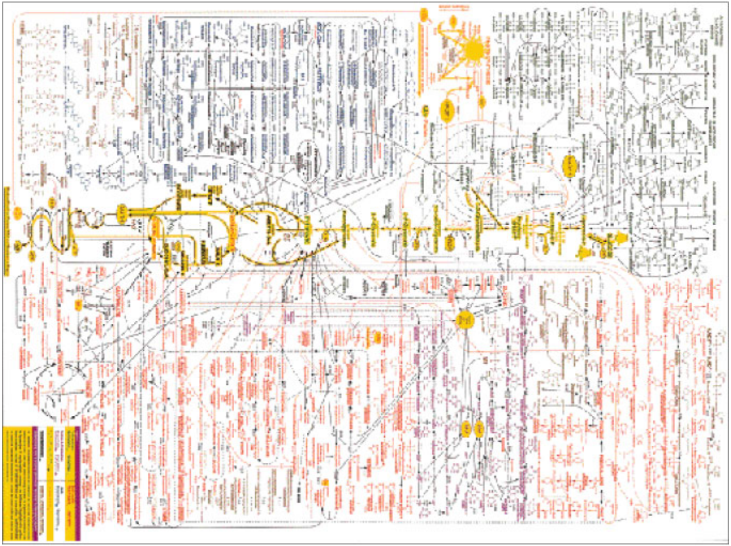
Networks in Biology: Protein-Protein Interaction (PPI) Networks



Networks in Biology: Metabolic Reactions



Networks in Biology: Metabolic Pathways



But Do Networks Matter?

- They Do!
- Recent studies have linked changes in gene/protein networks with many human diseases.

Systems Biology and Emerging Technologies

Gene Networks and microRNAs Implicated in Aggressive Prostate Cancer

Liang Wang,[1] Hui Tang,[2] Venugopal Thayanithy,[3] Subbaya Subramanian,[3] Ann L. Oberg,[2] Julie M. Cunningham,[1] James R. Cerhan,[2] Clifford J. Steer,[4] and Stephen N. Thibodeau[1]

[1]Departments of Laboratory Medicine and Pathology and [2]Health Sciences Research, Mayo Clinic, Rochester, Minnesota; and Departments of [3]Laboratory Medicine and Pathology, [4]Medicine, and Genetics, Cell Biology, and Development, University of Minnesota, Minneapolis, Minnesota

But Do Networks Matter?

0888-8809/07/$15.00/0
Printed in U.S.A.

Molecular Endocrinology 21(9):2112–2123

doi: 10.1210/me.2006-0474

Estrogen-Regulated Gene Networks in Human Breast Cancer Cells: Involvement of E2F1 in the Regulation of Cell Proliferation

Joshua D. Stender, Jonna Frasor, Barry Komm, Ken C. N. Chang, W. Lee Kraus, and Benita S. Katzenellenbogen

Departments of Biochemistry (J.D.S.) and Molecular and Integrative Physiology (J.F., B.S.K.), University of Illinois at Urbana-Champaign, Urbana, Illinois 61801-3704; Women's Health and Musculoskeletal Biology (B.K., K.C.N.C.), Wyeth Research, Collegeville, Pennsylvania 19426; and Department of Molecular Biology and Genetics (W.L.K.), Cornell University, Ithaca, New York 14853-4203

But Do Networks Matter?

Cancer Cell
Article

A Transcriptional Signature and Common Gene Networks Link Cancer with Lipid Metabolism and Diverse Human Diseases

Heather A. Hirsch,[1,7] Dimitrios Iliopoulos,[1,7] Amita Joshi,[1,7] Yong Zhang,[2] Savina A. Jaeger,[3] Martha Bulyk,[3,4,5] Philip N. Tsichlis,[6] X. Shirley Liu,[2] and Kevin Struhl[1,*]
[1]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA
[2]Department of Biostatistics and Computational Biology, Dana Farber Cancer Institute, Harvard School of Public Health, Boston, MA 02115, USA
[3]Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA
[4]Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA
[5]Harvard/MIT Division of Health Sciences and Technology (HST), Harvard Medical School, Boston, MA 02115, USA
[6]Molecular Oncology Research Institute, Tufts Medical Center, Boston, MA 02111, USA
[7]These authors contributed equally to this work
*Correspondence: kevin@hms.harvard.edu
DOI 10.1016/j.ccr.2010.01.022

But Do Networks Matter?

And, incorporating the knowledge of networks improves our ability to find causes of complex diseases.

Molecular Systems Biology 3; Article number 140; doi:10.1038/msb4100180
Citation: ***Molecular Systems Biology*** **3:140**

www.molecularsystemsbiology.com

REPORT

Network-based classification of breast cancer metastasis

Han-Yu Chuang[1,5], Eunjung Lee[2,3,5], Yu-Tsueng Liu[4], Doheon Lee[3] and Trey Ideker[1,2,4,*]

[1] Bioinformatics Program, University of California San Diego, La Jolla, CA, USA, [2] Department of Bioengineering, University of California San Diego, La Jolla, CA, USA, [3] Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Korea and [4] Cancer Genetics Program, Moores Cancer Center, University of California San Diego, La Jolla, CA, USA
[5] These authors contributed equally to this work
* Corresponding author. Department of Bioengineering, University of California San Diego, La Jolla, CA 92093, USA. Tel.: +1 858 822 4558; Fax: +1 858 534 5722; E-mail: trey@bioeng.ucsd.edu

Why Do We Need Network Inference?

- Despite progress, our knowledge of interactions in the genome is limited.
- The entire genome is a vast landscape, and experiments for discovering networks are very expensive
- From a statistical point of view, network estimation is related to estimation of covariance matrices, which has many independent applications in statistical inference and prediction (*more about this later*)
- Finally, and perhaps most importantly, gene and protein networks are dynamic and changes in these networks have been attributed to complex diseases.

Networks: A Short Premier

Networks: A Short Premier

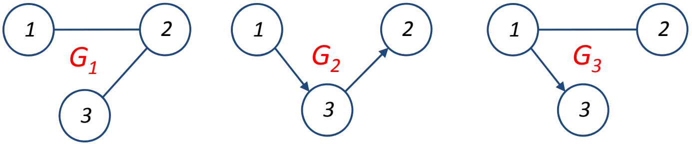
- A network is a collection of nodes V and edges E.

Networks: A Short Premier

- A network is a collection of nodes V and edges E.
- We assume there are p nodes in the network, and that the nodes correspond to random variables $X_1, \ldots X_p$.

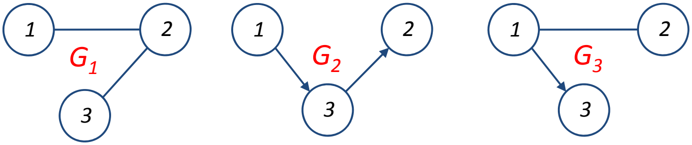
Networks: A Short Premier

- A network is a collection of nodes V and edges E.
- We assume there are p nodes in the network, and that the nodes correspond to random variables $X_1, \ldots X_p$.
- Edges in the network can be directed $X \to Y$ or undirected $X - Y$.



Networks: A Short Premier

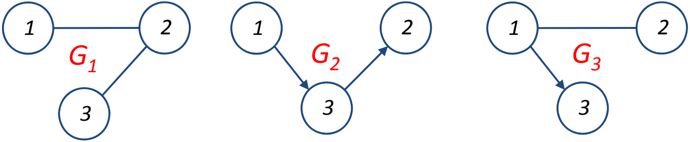
- A network is a collection of nodes V and edges E.
- We assume there are p nodes in the network, and that the nodes correspond to random variables $X_1, \ldots X_p$.
- Edges in the network can be directed $X \to Y$ or undirected $X - Y$.



- In all these example, the nodes are $V = \{1, 2, 3\}$.

Networks: A Short Premier

- A network is a collection of nodes V and edges E.
- We assume there are p nodes in the network, and that the nodes correspond to random variables $X_1, \ldots X_p$.
- Edges in the network can be directed $X \rightarrow Y$ or undirected $X - Y$.



- In all these example, the nodes are $V = \{1, 2, 3\}$.
- The edges are:

$$
\begin{aligned}
E_1 &= \{1 - 2, 2 - 3\} \\
E_2 &= \{1 \rightarrow 3, 3 \rightarrow 2\} \\
E_3 &= \{1 - 2, 1 \rightarrow 3\}
\end{aligned}
$$

Networks: A Short Premier

Networks: A Short Premier

- A convenient way to represent the edges of the network is to use an adjacency **matrix** A

Networks: A Short Premier

- A convenient way to represent the edges of the network is to use an adjacency **matrix** A
- A matrix is a rectangular array of data (similar to a table)

Networks: A Short Premier

- A convenient way to represent the edges of the network is to use an adjacency **matrix** A
- A matrix is a rectangular array of data (similar to a table)
- Values in each entry are shown by indeces of row and column

$$A = \left[\begin{array}{ccc} . & \mathbf{x} & . \\ . & . & . \\ . & . & . \end{array}\right]$$ Here, $\mathbf{x}$ is in row 1 and column 2

Networks: A Short Premier

- A convenient way to represent the edges of the network is to use an adjacency **matrix** A
- A matrix is a rectangular array of data (similar to a table)
- Values in each entry are shown by indeces of row and column

$$A = \begin{bmatrix} . & \mathbf{x} & . \\ . & . & . \\ . & . & . \end{bmatrix}$$ Here, **x** is in row 1 and column 2

- Adjacency matrix is a square matrix, which has a **1 if there is an edge** from a node in one row to a node in another column, and **0** otherwise

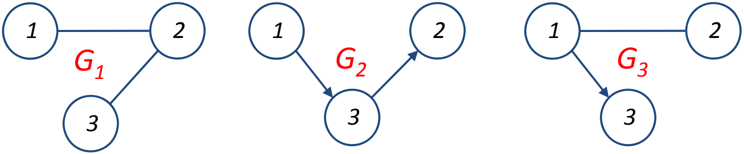
Networks: A Short Premier

- A convenient way to represent the edges of the network is to use an adjacency **matrix** A
- A matrix is a rectangular array of data (similar to a table)
- Values in each entry are shown by indeces of row and column

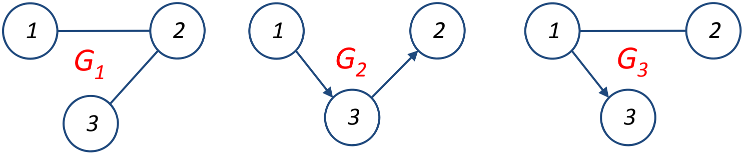
$$A = \begin{bmatrix} . & \mathbf{x} & . \\ . & . & . \\ . & . & . \end{bmatrix}$$ Here, **x** is in row 1 and column 2

- Adjacency matrix is a square matrix, which has a **1 if there is an edge** from a node in one row to a node in another column, and **0** otherwise
- For undirected edges, we add a **1** in both directions

Networks: A Short Premier



Networks: A Short Premier



$$A = \begin{bmatrix} 0 & \mathbf{1} & 0 \\ \mathbf{1} & 0 & \mathbf{1} \\ 0 & \mathbf{1} & 0 \end{bmatrix} \quad A = \begin{bmatrix} 0 & 0 & \mathbf{1} \\ 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 \end{bmatrix} \quad A = \begin{bmatrix} 0 & \mathbf{1} & \mathbf{1} \\ \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

What Do Edges in Biological Networks Mean?

- In gene regulatory networks, an edge from gene i to gene j often means that i affects the expression of j; i.e. as i's expression changes, we expect that expression of j to increase/decrease.

What Do Edges in Biological Networks Mean?

- In gene regulatory networks, an edge from gene i to gene j often means that i affects the expression of j; i.e. as i's expression changes, we expect that expression of j to increase/decrease.
- In protein-protein interaction networks, an edge between proteins i and j often means that *the two proteins bind together and form a protein complex*. Therefore, we expect that these proteins are generated at similar rates.

What Do Edges in Biological Networks Mean?

- In gene regulatory networks, an edge from gene i to gene j often means that i affects the expression of j; i.e. as i's expression changes, we expect that expression of j to increase/decrease.
- In protein-protein interaction networks, an edge between proteins i and j often means that *the two proteins bind together and form a protein complex*. Therefore, we expect that these proteins are generated at similar rates.
- In metabolic networks, an edge between compound i and j often means that *the two compounds are involved in the same reaction*, meaning that they are generated at relative rates.

What Do Edges in Biological Networks Mean?

- In gene regulatory networks, an edge from gene i to gene j often means that i affects the expression of j; i.e. as i's expression changes, we expect that expression of j to increase/decrease.
- In protein-protein interaction networks, an edge between proteins i and j often means that *the two proteins bind together and form a protein complex*. Therefore, we expect that these proteins are generated at similar rates.
- In metabolic networks, an edge between compound i and j often means that *the two compounds are involved in the same reaction*, meaning that they are generated at relative rates.
- Thus, edges represent some type of association among genes, proteins or metabolites, defined generally to include *linear or nonlinear* associations; more later....

Statistical Models for Biological Networks

Statistical Models for Biological Networks

- We use the framework of graphical models

Statistical Models for Biological Networks

- We use the framework of graphical models
- In this setting, nodes correspond to “random variables”

Statistical Models for Biological Networks

- We use the framework of graphical models
- In this setting, nodes correspond to "random variables"
- In other words, each node of the network represents one of the variables in the study
  - In gene regulatory networks, nodes $\equiv$ genes
  - In PPI networks, nodes $\equiv$ proteins
  - In metabolic networks, nodes $\equiv$ metabolites

Statistical Models for Biological Networks

- We use the framework of graphical models
- In this setting, nodes correspond to "random variables"
- In other words, each node of the network represents one of the variables in the study
  - In gene regulatory networks, nodes $\equiv$ genes
  - In PPI networks, nodes $\equiv$ proteins
  - In metabolic networks, nodes $\equiv$ metabolites
- In practice, we observe n measurements of each of the variables (genes/proteins/ metabolites) for say different individuals, and want to determine which variables are connected, or use their connection for statistical analysis

An Overview of Methods for Network Inference

Two general classes of network inference methods :

- Methods based on marginal measures of association:
  - Co-expression Networks (uses linear measures of association)
  - Methods based on mutual information (can accommodate non-linear associations)
- Methods based on conditional measures of association:
  - Methods assuming (multivariate) normality (`glasso`, etc)
  - Generalizations to allow for nonlinear dependencies

Our Plan

In the remainder of this module, we will cover the following topics

- Methods for reconstructing undirected networks
  - Marginal association (co-expression) nets (WGCNA, ARACNE)
  - Conditional independence graphs (CIGs)
- Network analysis (and more on WGCNA)
- Methods for reconstructing directed networks
  - Bayesian Networks (basic concepts, reconstruction algorithm)
  - Reconstructing directed networks from time-course data and perturbation screens (time permitting)
- Network-based pathway enrichment analysis
- Network-based omics data integration