Programming today is a race between software engineers striving to build bigger and better idiot-proof programs, and the Universe trying to produce bigger and better idiots. So far, the Universe is winning. - Unknown
Biological Networks  - Introduction

- The Concept of System Biology
  - Systems-level understanding of biological systems
  - Analyse not only individual components, but their interactions as well and emergent behaviour
  - System biology focus on the complex interactions in biological Systems
Biological Networks - Introduction

- The Concept of System Biology
  - System Biology Vs traditional cell Biology and Molecular Biology
  - Carl Wuench’s work
    - It is increasingly recognized that complex biological systems cannot be described in a reductionist view.
    - Understanding the behavior of such systems starts with understanding the topology of the corresponding network.
    - Topological information is fundamental in constructing realistic models for the function of the network.
The higher-order properties and functions that arise from the interaction of the parts of a system are called **emergent properties**.

- **human brain** can thought by the interaction of brain cells,
- **a single brain cell** is incapable of the property of thought.
What is a **network** (or **graph**)?

- A set of nodes (vertices) and edges (links)
- Edges describe a relationship between the nodes
Biological Networks - Introduction

Networks model many real-world phenomena
Biological Networks - Introduction

- E.g., Facebook
Biological Networks - Introduction

• E.g., WWW
Biological Networks - Introduction

- E.g., Internet
Biological Networks - Introduction

- E.g., Airline routes
Biological networks - Introduction

- What are Biological networks?
  - An often hypothetical formulation of interactions between biological entities inside organisms
  - Some do have physical forms
  - Few such networks are known in anything approaching their complete structure
  - Still less is known on the parameters governing the behavior of such networks over time
Why are Biological networks important?

- **For Biology**
  - If the topology and dynamics of all biological networks is known, the state of an organism could be completely predicted.
  - As such Biological networks are important to understanding organisms:
    - Treating and preventing ailments
    - Genetic Engineering
- **For Network Science**
  - Biological networks evolved over millions of years - only the fittest survive:
    - Robustness
    - Attack tolerance
    - Efficiency of information transfer
  - Design features can be emulated for better network design.
Why Study Networks?

- It is increasingly recognized that complex systems cannot be described in a reductionist view.
- Understanding the behavior of such systems starts with understanding the topology of the corresponding network.
- Topological information is fundamental in constructing realistic models for the function of the network.
Biological networks - Introduction

- **Biological nets**
  E.g., Protein structure networks
Biological networks - Introduction

- **Biological nets**
  E.g., Protein-protein interaction (PPI) networks
Biological networks - Introduction

- **Biological nets**
  E.g., Metabolic networks

Metabolic network of *A. thaliana*
Biological networks - Introduction

- **Biological nets**
  Other network types

$X \rightarrow Y$ represents

transcription network

neuron synaptic connection network

ecological food web

gene $x$  
gene $y$  

$X$  

Eat  

$Y$  

fish
Biological Network Model

- Network
  - A linked list of interconnected nodes.

- Node
  - Protein, peptide, or non-protein biomolecules.

- Edges
  - Biological relationships, etc., interactions, regulations, reactions, transformations, activation, inhibitions.
Biological Network Model

- It is usually represented by a 2-D diagram with characteristic symbols linking the protein and non-protein entities.

  - A circle indicates a protein or a non-protein biomolecule.
  - An symbol in between indicates the nature of molecule-molecule process (activation, inhibition, association, disassociation, etc.)
Introduction: biological networks

- Types of biological networks:
  - Intra-cellular networks
    - Metabolic networks
    - Transcriptional regulation networks / Gene Regulatory networks (GRN)
  - Cell signaling networks
    - Transcription networks and various representations
  - Protein-protein interaction (PPI) networks
  - Protein structure networks
Introduction: biological networks

- Types of biological networks:
  - Other biological networks
    - Neuronal synaptic connection and cortex networks
    - Brain functional networks
    - Phylogenetic networks
    - Correlation networks (e.g., gene expression)
    - Disease – “disease gene” association networks
    - Drug – “drug target” networks
  - On the fringe - Not inside an organism
    - Ecological and food webs
    - Word web of human language?
How do we analyze biological networks?

- Topology and large scale organization
  - Degree and remaining degree distributions
  - Network diameter
  - Assortativeness
  - Mutual information
  - Clustering
  - Modularity
  - Growth, evolution and modelling large scale structure

- Small scale organization and motifs
- Design patterns
Introduction: biological networks

- All of the intra-cellular networks describe cellular functioning at different levels and often “overlap”
  - Cell relies on numerous highly interconnected interactions and chemical reactions between various types of molecules, e.g., proteins, DNA, RNA, metabolites, etc.
  - Various activities of cells are controlled by the action of molecules upon molecules
  - Proteins – central players
Intra-cellular networks
Introduction: biological networks

Types of biological networks:

- Intra-cellular networks
  - **Metabolic networks**
  - Transcriptional regulation networks / Gene Regulatory networks (GRN)
  - Cell signaling networks
    - Transcription networks and various representations
  - Protein-protein interaction (PPI) networks
  - Protein structure networks
Metabolic networks

- Used for studying and modeling metabolism
  - Biochemical reactions in cells that allow an organism to:
    - Respond to the environment
    - Grow
    - Maintain its structure
    - ...
  - i.e., the main biochemical reactions needed to keep an organism in homeostasis
    - An internal regulation that maintains a stable, constant condition of a living system
Metabolic Networks

- There are two types
  - **Substrate networks:**
    - Nodes are substrates
    - Links are interactions (reactions)
  - **Reaction networks:** Vice versa
  - How do we represent enzymes?

- **A cellular Pathway:** A Pathway can be defined as a modular unit of interacting molecules to fulfill a cellular function.

- **A metabolic Pathway:** A series of chemical reactions occurring within a cell, catalyzed by enzymes, resulting in either the formation of a metabolic product to be used or stored by the cell, or the initiation of another metabolic pathway

- A metabolic network consists of metabolic pathways

- The global structure of metabolism is highly conserved through evolution (wagner)
Metabolic networks

- **Metabolites**
  - Small molecules such as glucose and amino acids
  - Also, macromolecules such as polysaccharides and glycans (carbohydrates)

- **Metabolic pathways**
  - Series of successive biochemical reactions for a specific metabolic function,
    - e.g., glycolysis or penicillin synthesis, that convert one metabolite into another
  - **Enzymes**: proteins that catalyze (accelerate) chem. reactions

- Thus, in a metabolic pathway:
  - Nodes correspond to metabolites and enzymes
    - In an alternate order ➔ bipartite graphs
  - Directed edges correspond to metabolic reactions
  - Simpler approaches: nodes are metabolites, directed edges are reactions that convert one metabolite into the other
Metabolic networks of *H. pylori* (left) and *Homo sapiens* (right) (intermediate substrates omitted)
A Pathway Example: Citrate cycle
Major metabolic pathways in humans

- Cellular Respiration - breaking down molecules to ATP (Video)
  - Glycolysis
  - Anaerobic respiration
  - Citric acid cycle / Krebs cycle
  - Oxidative phosphorylation
- Fatty acid metabolism
- Urea cycle
- Other organisms - Photosynthesis etc.
- Important to understand:
  - We are talking about cell metabolism – Not about human metabolic system
Metabolic pathways in humans
Constructing a metabolic Network - Glycolysis example

- Two major phases - preparatory and payoff
  - Glucose → Pyruvate, giving ATP molecules
  - Initial ATP necessary as investment
  - Presence of loops and short cuts
Glycolysis Example

- Let us abbreviate
  - A = Glucose
  - B = Glucose 6 phosphate
  - C = Fructose 6 phosphate
  - D = Fructose – 1,6 – Bi phosphate
  - E = Di hydro acetone phosphate
  - F = Glycerandehyde 3 phosphate
  - G = 1,3 – Bi phospho glycerate
  - H = 3-phospho glycerate
  - I = 2 – Phopho glycerate
  - J = Phopho enol pyruvate
  - K = pyruvate

- Preparatory Phase:
  - A +ATP \rightarrow B + ADP
  - B \leftrightarrow C
  - C + ATP \rightarrow D + ADP
  - D \leftrightarrow E + F
  - E + F \leftrightarrow F

- Pay off phase
  - F + NAD^+ + P \leftrightarrow G + NADH + H^+
  - G + ADP \leftrightarrow H + ATP
  - H \leftrightarrow I
  - I + H2O \leftrightarrow J
  - J + ADP \leftrightarrow K + ATP
Glycolysis example
The Large scale organization of metabolic networks

- Substrate networks are scale free with exponents typically between 2.0 and 3.0
  - Reaction networks are not exactly scale free
- Network diameter is the same for all metabolic networks, irrespective of their size
  - This is normally not the case with all scale free networks
  - The substrates are increasingly more connected as the network grows
- Network diameter is robust to random attacks
  - Not so for targeted attacks against hubs
- Assortativeness - typically slightly disassortative when all substrates are considered
  - Becomes assortative when intermediate substrates are omitted
Metabolic networks

- All metabolic pathways of a cell form a **metabolic network**
  - Complete view of cellular metabolism and material/mass flow through the cell
  - Cell relies on this network to digest substrates from the environment, generate energy, and synthesize components needed for its growth and survival
- Used to, for example:
  - Cure human metabolic diseases through better understanding of the metabolic mechanisms
  - Control infections of pathogens by understanding the metabolic differences between human and pathogens
Introduction: biological networks

- Types of biological networks:
  - Intra-cellular networks
    - Metabolic networks
  - Transcriptional regulation networks / Gene Regulatory networks (GRN)
  - Cell signaling networks
    - Transcription networks and various representations
  - Protein-protein interaction (PPI) networks
  - Protein structure networks
Gene Regulatory Networks

- Model regulation of gene expression
  - Recall: gene $\rightarrow$ mRNA $\rightarrow$ protein

- Gene regulation
  - Gives a cell control over its structure and function, e.g.:
    - **Cellular differentiation** - a process by which a cell turns into a more specialized cell type
    - **Morphogenesis** (a process by which an organism develops its shape)
    - ...

July 2013
N. H. N. D. de Silva
Gene Regulatory Networks

- **Nodes: Genes**
  - DNA sequences which are transcribed into mRNAs that translate into proteins

- **Links: Control of expression**
  - Directed edges correspond to interactions through which the products of one gene affect those of another
    - Protein-protein, protein-DNA and protein-mRNA interactions
Gene Regulatory Networks

- Signal \(\rightarrow\) Receptor Proteins \(\rightarrow\) Activate transcription factor \(\rightarrow\) attach to cis-regulatory region of a gene \(\rightarrow\) output mRNA \(\rightarrow\) output Protein

- Some of these proteins are receptor proteins for other genes
  - A gene controls another gene’s expression levels – resulting in GRNs

- Transcription factor X (protein product of gene X) binds regulatory DNA regions of gene Y to regulate the production rate (i.e., stimulate or repress transcription) of protein Y
  - Note: proteins are products of gene expression that play a key role in regulation of gene expression
Transcriptional regulation networks

- Problem
  - Stimulation and repression of gene transcription are both represented the same way in the network

- Available for model organisms
  - Non-human species manipulated and studied to get insights into workings of other organisms, e.g.:
    - Baker's yeast, S. cerevisiae (Milo et al., 2002)
    - E. coli (Shen-Orr et al., 2002)
    - Sea urchin (Davidson et al., 2002)
    - Fruitfly, D. melanogaster
  - Available from: EcoCyc, GeneNet, KEGG, RegulonDB, Reactom, TRANSPATH, TRANSFAC
Representation of the *E. coli* transcriptional regulatory network. a) Representation of the transcription-factor gene regulatory network of *E. coli*. Green circles represent transcription factors, brown circles denote regulated genes, and those with both functions are coloured in red. Projections of the network onto b) transcription factor and onto c) regulated gene nodes are also shown.

Modeling GRNs: The challenge

- Logical Models: State variables are represented discretely
- Boolean networks / Random Boolean Networks (Ensemble approach)
  - Attractors and attractor cycles – The eventual states of the network from which it cannot escape
  - Garden of Eden States - there is no path to reach that state
  - Transient Trees
  - Attractors as cell types - Cancer cells etc
  - Designing drugs: breaking away from a ‘bad’ limit cycle?
Modelling GRNs as RBNs

- The Wiring diagram
- The updating rules / Truth table
- State transition table - Synchronous updates
- State space Diagram
  - Two point attractors
  - Limit cycle (period - 2)
  - Garden of Eden States
RBN models - strengths and weaknesses

- **Strength**
  - Ease of simulation and manageable state complexity

- **Weaknesses**
  - The Boolean assumption - Various expression levels
  - Issue of synchrony

- **Other Logical models:**
  - Generalized Logic: Variables can assume multiple logical values
  - Asynchronous update
Continuous Models

- Differential Equation models:
  - Since one Gene’s expression concentration depends often on the rate of change of another gene’s expression, DDE models are a natural choice.
  - Eg: Two Node feedback Loops (an often occurring motif in GRNs): Can be modeled as:

\[
\frac{dr_i}{dt} = f_i(p) \\
\frac{dp_i}{dt} = g_i(r)
\]
Artificial Neural Network Models

- Mathematically, it is possible to create a mapping between a neural network and a system of Ordinary Differential Equations (ODEs)
- Nodes of the ANN should not be confused with nodes in the GRN.
  - A corresponding Gene may not exist for a Node in the hidden layer of the ANN, for example
Pros and Cons of Continuous Models

- **Advantage:** No assumption about synchronous update, continuous expression levels can be modeled

- **Disadvantages:**
  - Systems of ODEs are not always easily solvable
  - Less intuitive
Stochastic Models for GRNs

- Justification:
  - evidence suggests they are actually produced stochastically in short “bursts”
  - the time taken for a concentration to reach its critical threshold will vary stochastically.
  - Also, there is noise in cell signals: Introduces randomness

- Stochasticity can be introduced to the above mentioned models:
  - Eg: The ODE representation can be modified to:

\[ \frac{dx_i}{dt} = f_i(x_i) + \nu_i(t) \]

- Where \( \nu(t) \) is a stochastic function
Introduction: biological networks

- Types of biological networks:
  - Intra-cellular networks
    - Metabolic networks
    - Transcriptional regulation networks / Gene Regulatory networks (GRN)
  - **Cell signaling networks**
    - Transcription networks and various representations
  - Protein-protein interaction (PPI) networks
  - Protein structure networks
Cell Signaling Networks

- **Cell signaling**
  - Complex communication system that governs basic cellular activities, e.g., development, repair, immunity

- **Errors in signaling cause diseases**
  - E.g., cancer, autoimmune diseases, diabetes…

E.g.: Transforming growth factor beta (TGF-β) is a protein that controls proliferation, cellular differenciation, and other functions in most cells.
Cell Signalling Networks

- Cells are Nodes – signals link them (inter cellular network)
- Cell signaling is part of a complex system of communication that governs basic cellular activities
- Many types of cell signals exist
  - Some cells can form gap junctions that connect their cytoplasm to the cytoplasm of adjacent cells
  - juxtacrine signalling (also known as contact dependent signaling) - Cells must make contact for signalling
  - Many cell signals are carried by molecules that are released by one cell and move to make contact with another cell
Cell signaling networks

- **Signaling pathways**
  - Ordered sequences of signal transduction reactions in a cell, as shown in the previous figure
  - Cascade of reversible chemical modifications of proteins
    - E.g., phosphorylation catalyzed by protein kinases
  - Signaling pathways in the cell form the **cell signaling network**
    - Nodes are proteins and edges are directed
Cell signaling networks

Famous examples (lots of literature on them):

- **Mitogen-activated protein kinase (MAPK) pathway**
  - Originally called “ERK” pathway
  - MAPK protein: an enzyme, a **protein kinase**, which can attach phosphate groups to a target protein, causing its spatial reorganization and affecting its function
    - Other enzymes can restore protein’s initial function
  - E.g.:
    - MYC
      - An **oncogene** transcription factor expressed in a wide range of human cancers (oncogene – when mutated or over-expressed, the gene helps turn a normal into a tumor cell)
      - MAPK can **phosphorylate** (attach phosphate group to) MYC and alter gene transcription and cell cycle progression
    - EGFR = “epidermal growth factor receptor”
      - Activates MAPK pathway
      - Mutations affecting its expression/activity can result in cancer
Cell signaling networks

Famous examples (lots of literature on them) cont’d:

- **Hedgehog signaling pathway**
  - One of the key regulators of animal development
  - Conserved from fly to human
  - Establishes basis of fly body plan
  - Important during embryogenesis (the process by which the embryo develops) and metamorphosis (from larva to pupa to adult)

- **TGF-beta signaling pathway**
  - The “transforming growth factor” (TGF) signaling pathway
  - Involved in:
    - Cell growth
    - Cell differentiation
    - Apoptosis (programmed cell death)
Cell signaling networks

- Compared to metabolic networks:
  - Limited mass flow
  - Instead, sig. nets provide information transmission along a sequence of reactions – one enzyme modulates the activity of another, which then modulates the activity of the third enzyme, etc., but enzymes are not consumed in the reactions they catalyze

- Compared to transcriptional reg. networks:
  - They overlap, but gene expression, i.e., transcription factors, can be seen as the “final targets” of signaling pathways

- Compared to PPI networks:
  - Signal transduction is indeed mediated between proteins, but PPIs are undirected without a defined input and output (as we will discuss soon)
  - Not all PPIs are involved in chemical reactions or part of signal transduction
  - Also, many components of signaling are not proteins
  - These networks have much in common
  - At the same time, they reflect different aspects of cellular activity
Introduction: biological networks

- Types of biological networks:
  - Intra-cellular networks
    - Metabolic networks
    - Transcriptional regulation networks / Gene Regulatory networks (GRN)
  - Cell signaling networks
    - Transcription networks and various representations
  - **Protein-protein interaction (PPI) networks**
  - Protein structure networks
Proteins in a cell

- There are thousands of different active proteins in a cell acting as:
  - enzymes, catalysors to chemical reactions of the metabolism
  - components of cellular machinery (e.g. ribosomes)
  - regulators of gene expression
  - Certain proteins play specific roles in special cellular compartments.
  - Others move from one compartment to another as “signals”.
Protein Interactions

- Proteins perform a function as a complex rather as a single protein.

- Knowing whether two proteins interact can help us discover unknown proteins’ functions:
  - If the function of one protein is known, the function of its binding partners are likely to be related—“guilt by association”.
  - Thus, having a good method for detecting interactions can allow us to use a small number of proteins with known function to characterize new proteins.
Protein Interactions

Finding Proteins That Interact

One technique, called the yeast two-hybrid system, relies on bringing into close proximity two halves (a and b) of a protein that activates a gene that causes a yeast cell to turn blue. It is used to determine which of a pool of unknown “prey” proteins binds to a known “bait” protein.

1 Insert DNA encoding a known “bait” protein linked to DNA for half (a) of the activator protein.

2 Insert DNA for the other half (b) of the activator protein linked to DNA encoding random “prey” proteins.

3 Look for color change, which indicates “prey” protein binding to “bait”.

Protein-protein interaction (PPI) networks
Protein-protein interaction (PPI) networks

- A protein-protein interaction (PPI) usually refers to a physical interaction, i.e., binding between proteins
- Can be other associations of proteins such as functional interactions – e.g., synthetic lethality
Protein-protein interaction (PPI) networks

- PPIs are very important for structure and function of a cell:
  - Participate in signal transduction
    - Play a role in many diseases (e.g., cancer)
  - Can be stable interactions forming a protein complex
    (a form of a quaternary protein structure, set of proteins which bind to do a particular function, e.g., ribosome, hemoglobin – illustrated below)
Yeast Protein Interaction Network

Nodes: proteins
Links: physical interactions (binding)
Protein-protein interaction (PPI) networks

- PPIs are very important for structure and function of a cell:
  - Can be transient interactions
    - Brief interactions that modify a protein that can further change PPIs e.g., protein kinases (add a phosphate group to a target protein)
    - A protein can carry another protein, e.g., nuclear pore importins (proteins that carry other proteins from cytoplasm to nucleus and vice versa)
    - Transient interaction form the dynamic part of PPI networks
  - Some estimates state that about 70% of interactions are stable and 30% are dynamic
- PPI are essential to almost every process in a cell
- Thus, understanding PPIs is crucial for understanding life, disease, development of new drugs (most drugs affect PPIs)
Protein-protein interaction (PPI) networks

Methods to detect PPIs

- Biological and computational approaches
- None are perfect
  - High rates of false positives
    - Interactions present in the data sets that are not present in reality
  - High rates of false negatives
    - Missing true interactions
Protein-protein interaction (PPI) networks

**Methods to detect PPIs**

- PPIs initially studied individually by small-scale biochemical techniques (SS)
- However, large-scale (high-throughput) interaction detection methods (HT) are needed for high discovery rates of new protein interactions
- SS of better “quality,” i.e., less noisy than HT
- However, HT are more standardized, while SS are performed differently each time
- SS are biased – the focus is on the subsets of proteins interesting to particular researchers
- HT – view of the entire proteome
Protein-protein interaction (PPI) networks

Methods to detect PPIs

- Physical binding
  - Yeast 2-hybrid (Y2H) screening
  - Mass spectrometry of purified complexes
- Functional associations
  - Correlated mRNA expression profiles
- Genetic interactions
  - In silico (computational) methods

- In many cases, functional associations do take the form of physical binding
Protein-protein interaction (PPI) networks

Yeast two-hybrid assay

- Binary PPIs
- Pairs of proteins to be tested for interaction are expressed as artificial (genetically engineered) fusion proteins in yeast:
  - One protein is fused to a reporter gene (a gene attached to another gene of interest)
  - The other is fused to a transcription factor
  - Any interaction between them is detected by the transcriptional activation of the reporter gene
Protein-protein interaction (PPI) networks

Yeast two-hybrid assay
- One protein (in PPI) is “bait”, the other is “prey”
- Potential problem:
  - Interest in a particular pathway of, say 15 proteins
  - These 15 proteins are all “baits”
  - There is an order of magnitude more “preys”
  - This imposes a particular structure on the PPI network by experimental design without reflecting the underlying network topology
- To avoid this, a matrix of $n \times n$ needs to be probed, where each bait is also a prey (Mark Vidal’s lab, Harvard)
Protein-protein interaction (PPI) networks

Yeast two-hybrid assay
- This method is scalable to entire proteome
- Directly tests a protein pair for an interaction
- But high noise rate (50%, even up to 70%)
  - Because Y2H investigates interactions between:
    - artificial, fusion proteins
    - in the yeast
    - in the yeast’s nucleus
- Each of these steps is noisy
- Proteins need to be in their native environment, not in nucleus
  - E.g., although proteins can physically bind, they never do so inside a cell, because of different localization, or because they are never simultaneously expressed
Protein-protein interaction (PPI) networks

Mass spectrometry of purified complexes

- Individual proteins are tagged and used as hooks to biochemically purify whole protein complexes

- Complexes separated and components identified by mass spectrometry (MS)
  - MS measures mass-to-charge ratio of ions

- TAP (Tandem Affinity Purification)

- HMS-PCI (High-Throughput MS Protein Complex Identification)

- Not binary but co-complex data
Protein-protein interaction (PPI) networks

Mass spectrometry of purified complexes

- We know what proteins are in the complexes, but not how they are connected
  - Spoke model
  - Matrix model
Protein-protein interaction (PPI) networks

Mass spectrometry of purified complexes

- **Pros:**
  - Detects real complexes in their physiological settings
  - Consistency check is possible by tagging several members of a complex
  - Good for screening permanent/stable interactions

- **Cons:**
  - Might miss some complexes that are not present under given cellular conditions
  - Tagging may disturb complex formation
  - Loosely associated components can be washed off during purification
Protein-protein interaction (PPI) networks

Functional associations
  ▶ Correlated mRNA expression profiles
    ▶ Results in a gene expression correlation network
Protein-protein interaction (PPI) networks

Functional associations

- Genetic interactions
  - Two non-essential genes that cause lethality when mutated at the same time form a synthetic lethal interaction
  - Such genes are often functionally associated and their encoded proteins may also interact physically
  - Charles Boone’s group from University of Toronto published genetic interaction networks
Protein-protein interaction (PPI) networks

Functional associations

Genetic interactions

Figure 1. Gene Interactions in Cancer
(A) Extreme forms of genetic interaction are defined by synthetic lethality (in which a combination or synthesis of gene mutations causes cell death) and the reverse scenario, synthetic viability (in which a combination of gene effects rescues the lethal effects of a single gene change).
(B) Different modes of genetic interaction defined by quantitative effects on a phenotype, such as cell fitness. Here the value 1 represents the maximal fitness of cells, and the individual effects of changes in genes A (0.7) or B (0.5) are shown. When no interaction between genes A and B exists, the simple combination of effects (shown here as $0.7 \times 0.5 = 0.35$) is expected; any deviation from this value suggests an interaction between genes A and B.
Protein-protein interaction (PPI) networks

Functional associations

- In silico (computational) methods
  - Gene fusion (if two genes are present in one species and fused in another)
  - ...

July 2013
N. H. N. D. de Silva
Protein-protein interaction (PPI) networks

Biases within PPI networks

- The following is lost:
  - Spatial information
  - Temporal information
  - Information about experimental conditions
  - Strength of interactions
  - Number of experiments confirming interactions

- PPI network: proteome + interactome
  - Proteome: a set of all unique proteins in an organism;
  - How does protein concentration affect the topology:
    - More instances of a protein in the cell \(\rightarrow\) more interacting partners in the network?
Protein-protein interaction (PPI) networks

Quality and completeness of PPI data

- Data sets produced by different methods are often complementary.
- Even data sets obtained by the same technique complement each other to some (large) extent.
- Completeness of data sets:
  - Yeast: ~50% (~6K proteins, ~30K-60K interactions)
  - Human: ~10% (~25K proteins, ~260K interactions; ~300 million pairs to test)
  - Fly
  - Worm
Introduction: biological networks

- Types of biological networks:
  - Intra-cellular networks
    - Metabolic networks
    - Transcriptional regulation networks / Gene Regulatory networks (GRN)
    - Cell signaling networks
      - Transcription networks and various representations
  - Protein-protein interaction (PPI) networks
  - Protein structure networks
Protein structure networks

- PDB (Protein Data Bank): http://www.pdb.org/
Protein structure networks

- “Residue interaction graphs” (RIGs) model protein structures
  - Nodes are amino acid residues
  - Undirected, unweighted edges exist between amino acids that are in close proximity in the protein’s 3-dimensional structure
    - E.g., within 5 Angstroms (1 Å = 10^{-10} meters)

Introduction: biological networks

- Types of biological networks:
  - **Intra-cellular networks**
    - Metabolic networks
    - Transcriptional regulation networks / Gene Regulatory networks (GRN)
  - Cell signaling networks
    - Transcription networks and various representations
  - Protein-protein interaction (PPI) networks
  - Protein structure networks
Intra-cellular networks: summary

Proteins
Metabolites
Metabolism
Gene regulation
Cell signaling
PPIs
Other biological networks
Introduction: biological networks

- Types of biological networks:
  - Other biological networks
    - Neuronal synaptic connection and cortex networks
    - Brain functional networks
    - Ecological food webs
    - Phylogenetic networks
    - Correlation networks (e.g., gene expression)
    - Disease – “disease gene” association networks
    - Drug – “drug target” networks
  - On the fringe - Not inside an organism
    - Ecological and food webs
    - Word web of human language?
Neuronal networks

- Neural networks are the largest Biological networks (or any other networks) we study – The number of neurons in human brain is $O(100 \text{ billion})$

- The neural network of C.Elegans, however, has only about 300 (2000) neurons - Widely studied!

- Neural networks show small world effect – How about scale free exponents?
Introduction: biological networks

- Types of biological networks:
  - Other biological networks
    - Neuronal synaptic connection and cortex networks
    - **Brain functional networks**
    - Phylogenetic networks
    - Correlation networks (e.g., gene expression)
    - Disease – “disease gene” association networks
    - Drug – “drug target” networks
  - On the fringe - Not inside an organism
    - Ecological and food webs
    - Word web of human language?
Brain functional networks

- Simultaneous (correlated) activities of brain regions during a task
Introduction: biological networks

- Types of biological networks:
  - Other biological networks
    - Neuronal synaptic connection and cortex networks
    - Brain functional networks
  - Phylogenetic networks
    - Correlation networks (e.g., gene expression)
    - Disease – “disease gene” association networks
    - Drug – “drug target” networks
  - On the fringe - Not inside an organism
    - Ecological and food webs
    - Word web of human language?
Phylogenetic networks (trees)

- Evolutionary relationships between species
Introduction: biological networks

- Types of biological networks:
  - Other biological networks
    - Neuronal synaptic connection and cortex networks
    - Brain functional networks
    - Phylogenetic networks
  - Correlation networks (e.g., gene expression)
  - Disease – “disease gene” association networks
  - Drug – “drug target” networks
- On the fringe - Not inside an organism
  - Ecological and food webs
  - Word web of human language?
Correlation networks (e.g., gene expression)

- Different from transcriptional regulation networks
- Not a direct result of experiments
- Determined by:
  - Collecting large amounts of high-throughput data
  - Calculating the correlations between all elements
- Biolayout Express 3-D: a tool for generating correlation networks
Introduction: biological networks

- Types of biological networks:
  - Other biological networks
    - Neuronal synaptic connection and cortex networks
    - Brain functional networks
    - Phylogenetic networks
    - Correlation networks (e.g., gene expression)
  - Disease - “disease gene” association networks
  - Drug – “drug target” networks
- On the fringe - Not inside an organism
  - Ecological and food webs
  - Word web of human language?
Disease – “disease gene” association networks

- Link diseases that are caused by the same gene
- Link genes if they cause the same disease

Introduction: biological networks

- Types of biological networks:
  - Other biological networks
    - Neuronal synaptic connection and cortex networks
    - Brain functional networks
    - Phylogenetic networks
    - Correlation networks (e.g., gene expression)
    - Disease – “disease gene” association networks
  - **Drug - “drug target” networks**
  - On the fringe - Not inside an organism
    - Ecological and food webs
    - Word web of human language?
Drug – “drug target” association networks

- Link drugs if they target the same gene (protein)
- Link genes (proteins) if they are targeted by the same drug
On the fringe - Not inside an organism
Introduction: biological networks

- Types of biological networks:
  - Other biological networks
    - Neuronal synaptic connection and cortex networks
    - Brain functional networks
    - Phylogenetic networks
    - Correlation networks (e.g., gene expression)
    - Disease – “disease gene” association networks
    - Drug – “drug target” networks
  - On the fringe - Not inside an organism
    - Ecological and food webs
    - Word web of human language?
Food webs

- Vertex are species; Links are predatory connections
- Rather small (a few hundreds) – therefore statistical analysis is comparatively difficult
- Seem to suggest small-worldness and scale free distributions - even though the later conclusion is contested
- Cannibalism and mutual eating are widespread
Introduction: biological networks

- Types of biological networks:
  - Other biological networks
    - Neuronal synaptic connection and cortex networks
    - Brain functional networks
    - Phylogenetic networks
    - Correlation networks (e.g., gene expression)
    - Disease – “disease gene” association networks
    - Drug – “drug target” networks
  - On the fringe - Not inside an organism
    - Ecological and food webs
    - Word web of human language
Other real-world networks

- Technological networks:
  - WWW
  - Internet
  - Electric circuits
  - Software call graphs

- Transportation networks:
  - Roads, airlines, railways

- Social networks:
  - Friendships/relationships (Facebook, MySpace)
  - Collaborations between scientists/movie stars
  - Spread of infections and diseases
  - Economic networks
  - Relationships between organizations (companies, NGOs, etc.)
  - City/country trading relationships
  - Migrations
  - Disaster response networks
Inference techniques
Inference techniques

- How can we infer network properties when we do not know the complete topology?
  - Eg: scale free exponent, assortativeness, degree distribution etc
  - Inference techniques
  - Micro array data - Time Vs signal axis
  - We can see the propagation of signals
  - Develop rule sets to determine the topology from this. (Perturbation Analysis)
Biological Networks
Properties
Biological Networks
Properties

- Power law degree distribution: Rich get richer
- Small World: A small average path length
  - Mean shortest node-to-node path
- Robustness: Resilient and have strong resistance to failure on random attacks and vulnerable to targeted attacks
- Hierarchical Modularity: A large clustering coefficient
  - How many of a node’s neighbors are connected to each other
PREFERENTIAL ATTACHMENT on Growth: the probability that a new vertex will be connected to vertex i depends on the connectivity of that vertex:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$
The Barabási-Albert [BA] model

(a) Random Networks
(b) Power law Networks

Power Law Network (Scale Free)

- The probability of finding a highly connected node decreases exponentially with k:
  \[ P(K) \sim K^{-\gamma} \]
Small World Property

- A small average path length
- Any node can be reached within a small number of edges, 4~5 hops.
Power Law Network

- Power-law degree distribution & Small world phenomena also observed in:
  - communication networks
  - web graphs
  - research citation networks
  - social networks
- Classical -Erdos-Renyi type random graphs do not exhibit these properties:
  - Links between pairs of fixed set of nodes picked uniformly:
  - Maximum degree logarithmic with network size
  - No hubs to make short connections between nodes
Complex systems maintain their basic functions even under errors and failures
(cell $\rightarrow$ mutations; Internet $\rightarrow$ router breakdowns)
Attack Tolerance

- Robust. For $\gamma < 3$, removing nodes does not break network into islands.
- Very resistant to random attacks, but attacks targeting key nodes are more dangerous.
References

- Slides prepared by Dr. Mahendra Piraveenan for this class in previous years.
- “341: Introduction to Bioinformatics” - Dr. Nataša Pržulj, Department of Computing, Imperial College London (http://www.doc.ic.ac.uk/~natasha/course2012/docs/7_introduction-biolog_nets.ppt)
- “Biological Networks” - Woochang Hwang (http://www.cse.buffalo.edu/DBGROUP/Seminar/Huwang-2-7.ppt)
- “Microarrays” - Mustafa Kahraman (http://www.atilim.edu.tr/~mkahraman/SystemBiology.ppt)