

Introduction to Systems Biology

Lecture 8 Part A-1

Iyengar

Network Building & Analysis and Data Organization

Graph Theory

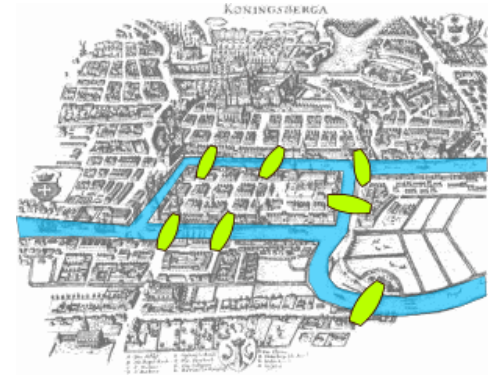
Graphs mathematical structure of entities (*vertices or nodes*) and connections (edges)

Started in 1700s by Euler who tried to solve the problem of how to walk through the city such that one crosses each bridge but only once

Actually, this is not possible – however the reasoning of why it is impossible to led to some of the original concepts of how relationships between nodes and edges lead restriction of how one can walk through a graph

In cell biology /biochemistry nodes are cellular components and edges are interactions, reactions and relationships between nodes

Casting cellular systems as networks is very useful in understanding how the systems are organized (topology)



Seven Bridges of Königsberg
Adapted figure from Wikipedia
Author : Bogdan Giușcă

Introduction to Systems Biology

Lecture 8 Part A-2

Iyengar

Bayesian Networks

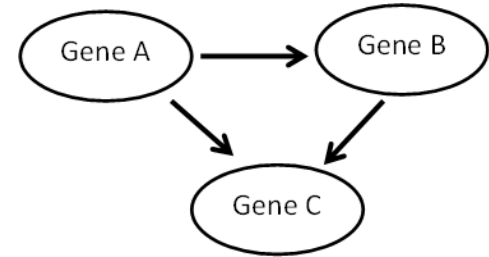
Sometimes relationships (edges) between nodes are conditional

Specifying the conditional relationship allows us to compute inferences regarding systems behavior

e.g. – in the small gene regulatory network shown here, both Gene A and Gene B can induce the expression of Gene C with different probabilities and Gene A can also induce Gene B with a third distinct probability

This system can be modeled as a Bayesian network

From the levels of C one can infer the probabilities of the active levels of Gene A and B



Introduction to Systems Biology

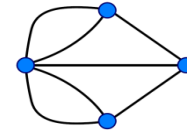
Lecture 8 Part A-3

Iyengar

Different types of networks

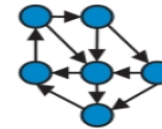
Undirected graphs

The Euler seven bridges problem shown as a graph. Note there are 4 nodes and 7 edges and the direction of the edges are not specified. Such undirected graphs are commonly constructed from high throughput data and often represents partially our knowledge of cell biological systems



Directed graphs

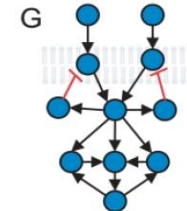
Nodes are connected by directed edges restricting capability to traverse the graph or direct signal flow. But relationship of the effect between nodes is not specified. Directed edges can represent reversible reactions where reaction in one direction is dominant



Sign-Specified Directed Graphs

Nodes are connected by black arrows that represent stimulatory relationship and red plungers that represent inhibitory relationship

Such graphs are the most useful... can be used to develop dynamical models of signal flow



Introduction to Systems Biology

Lecture 8 Part A-4

Iyengar

A 2005 network of a mammalian brain cell (hippocampal neuron)

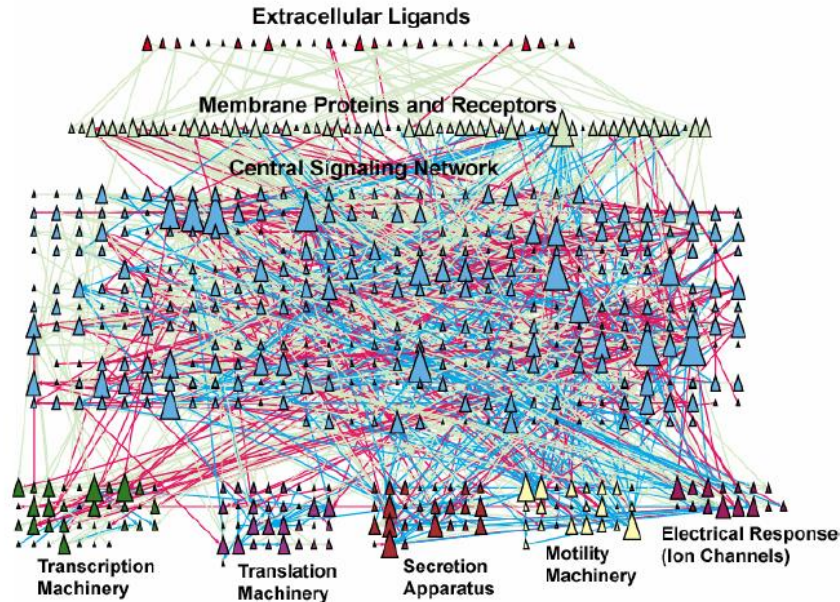
Although the network diagram is not visually tractable it is fully computable graph is drawn by a computer program

Each triangle is a cellular protein . Size of the triangle represents the number of edges (connections)

Green lines - activating edges

Red lines - inhibitory edges

Blue – neutral - interaction but no activation or inhibitionsuch as interactions with scaffold proteins.



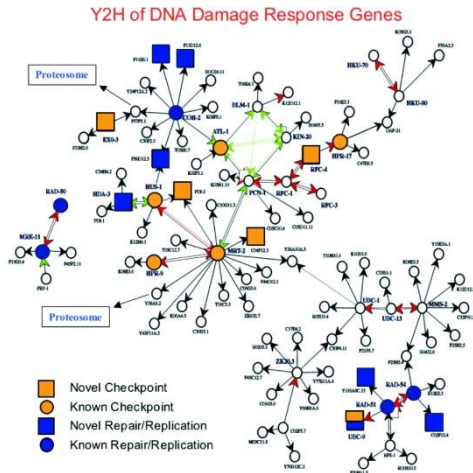
Ma'ayan A et. al. (2005) Science 309: 1078

Introduction to Systems Biology

Lecture 8 Part A-5

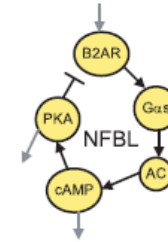
Iyengar

Types of Networks relevant to cellular systems biology



Cusick M E et al.
Hum. Mol. Genet. 2005;14:R171-R181

1. Cell Signaling Networks



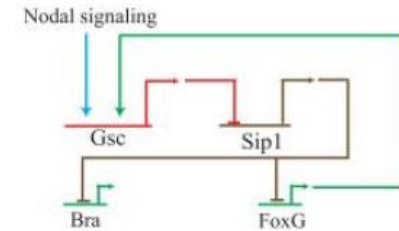
Ma'ayan A (2009) JBC 284:5451

2. Protein-Protein interaction Networks

Y2H - yeast -2-hybrid – a method for detecting direct protein-protein interactions
a network of direct interactors involved in DNA damage repair in yeast

3. Gene Regulatory Networks

a small network of four genes involved in oral ectoderm development of sea urchin



Davidson EH Curr Opin Genet Dev (2009)19: 535

Introduction to Systems Biology

Lecture 8 Part A-6

Iyengar

Bioinformatics

A field focused on developing methods for computer-based curating (storing, organizing, & annotating) and retrieving data...most often in a computable format. The structure of a database is called schema.

Development of software tools to process data to extract knowledge is a major focus of bioinformatics research

Data input and curation can be automated or manual

Major classes of biological databases

Genes- Genomics

DNA Sequences and Sequence Analysis - GenBank

Proteins

Database of Protein Structures - PDB

Protein characteristics - UniProtKB

National Center for Biotechnology Information at the National Library of Medicine www.ncbi.nlm.nih.gov

Introduction to Systems Biology

Lecture 8 Part A-7

Iyengar

Database of Cell Signaling Cell Signaling Pathways
Science Signaling--- Signal Transduction Knowledge Environment

KEGG Kyoto Encyclopedia of Genes and Genomes: a database of biological functions and systems including pathways

Pathway Commons Biological pathways from multiple organisms

GEO Gene Expression Omnibus genomics data base supported by NCBI - microarray and sequence based data

OMIM Online Mendelian Inheritance in Man - catalog of human genes and genetic disorders and traits

ENCODE Encyclopedia of DNA Elements – all functional elements in the human DNA sequence

Introduction to Systems Biology

Lecture 8 Part A-8

Iyengar

Uses of Databases

1. Computationally building large systems (networks) representing cellular functions
2. Use as prior knowledge to discover new pathways and subnetworks associated with functions of interest
3. Integration of high throughput data with multiple databases allows for development of multiscale understanding of physiological functions and identification of disease functions