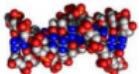


Why networks in biology?

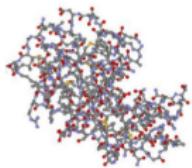
Molecules of life do not function in isolation



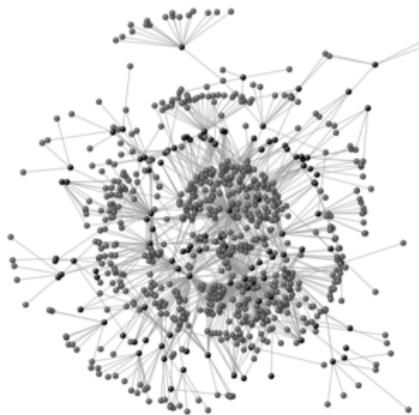
metabolites



genes



proteins



but form complex networks that define a cell

1 Introduction

- Learning goals
- Definitions

2 Graph Properties

- Definitions and explanations

3 Use Cases

- Use cases in biology

4 Algorithms

5 Online Resources

- Interaction Databases

6 Cytoscape

- Cytoscape

7 Summary

- Learning Goals

Learning goals

- ① What is a graph/network/pathway?

Learning goals

- ① What is a graph/network/pathway?
- ② What are the properties of a graph? What do they tell us?

Learning goals

- ① What is a graph/network/pathway?
- ② What are the properties of a graph? What do they tell us?
- ③ How can we use networks in biology?

Learning goals

- ① What is a graph/network/pathway?
- ② What are the properties of a graph? What do they tell us?
- ③ How can we use networks in biology?
- ④ How can you find the shortest path between two nodes in a network?

Learning goals

- ① What is a graph/network/pathway?
- ② What are the properties of a graph? What do they tell us?
- ③ How can we use networks in biology?
- ④ How can you find the shortest path between two nodes in a network?
- ⑤ Where can you find network/interaction data?

Learning goals

- ① What is a graph/network/pathway?
- ② What are the properties of a graph? What do they tell us?
- ③ How can we use networks in biology?
- ④ How can you find the shortest path between two nodes in a network?
- ⑤ Where can you find network/interaction data?
- ⑥ How to use Cytoscape, the bioinformatics tool for network analysis?

What is a graph? What is a network?

Definition

A graph/network is an abstract representation of a set of objects where some pairs of the objects are connected by links.

Graph is the more general and abstract term.



What is a graph? What is a network?

Definition

A graph/network is an abstract representation of a set of objects where some pairs of the objects are connected by links.

Graph is the more general and abstract term.



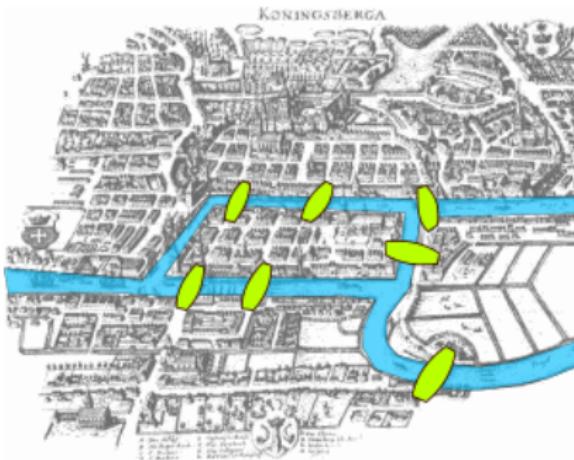
- object = **node or vertex**
- link = **line or edge**

Do you know any networks?

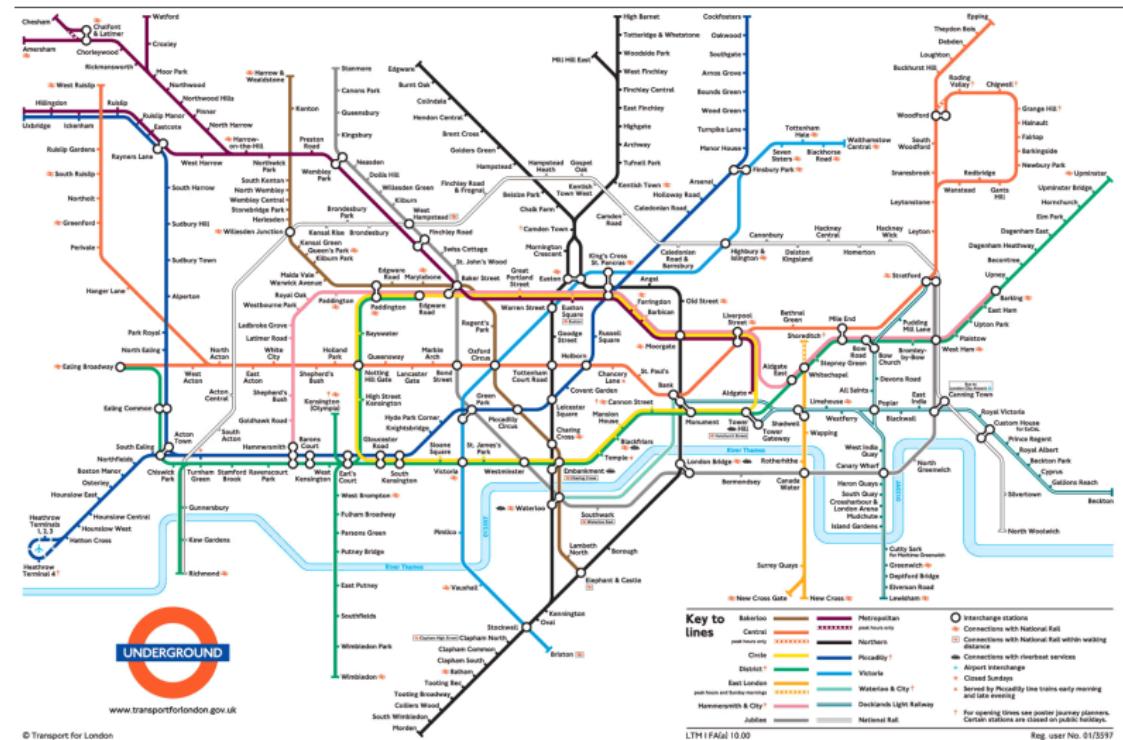


Graph theory - history

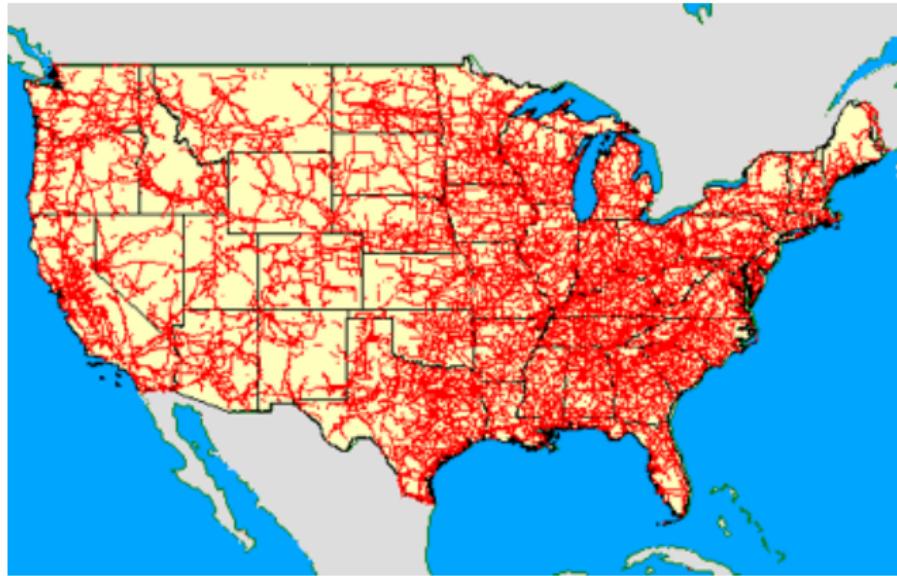
Seven Bridges of Königsberg, Leonhard Euler, 1736



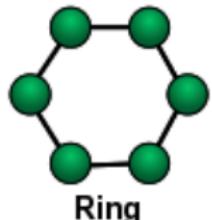
London Subway



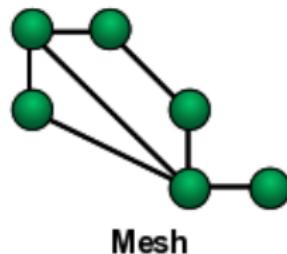
Power Grid



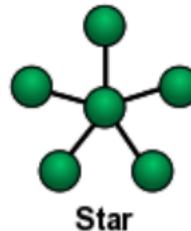
Network Topologies



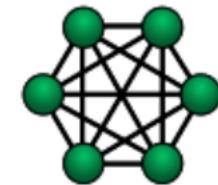
Ring



Mesh



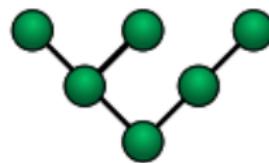
Star



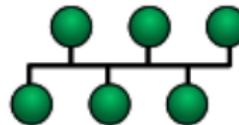
Fully Connected



Line



Tree



Bus

Social Networks

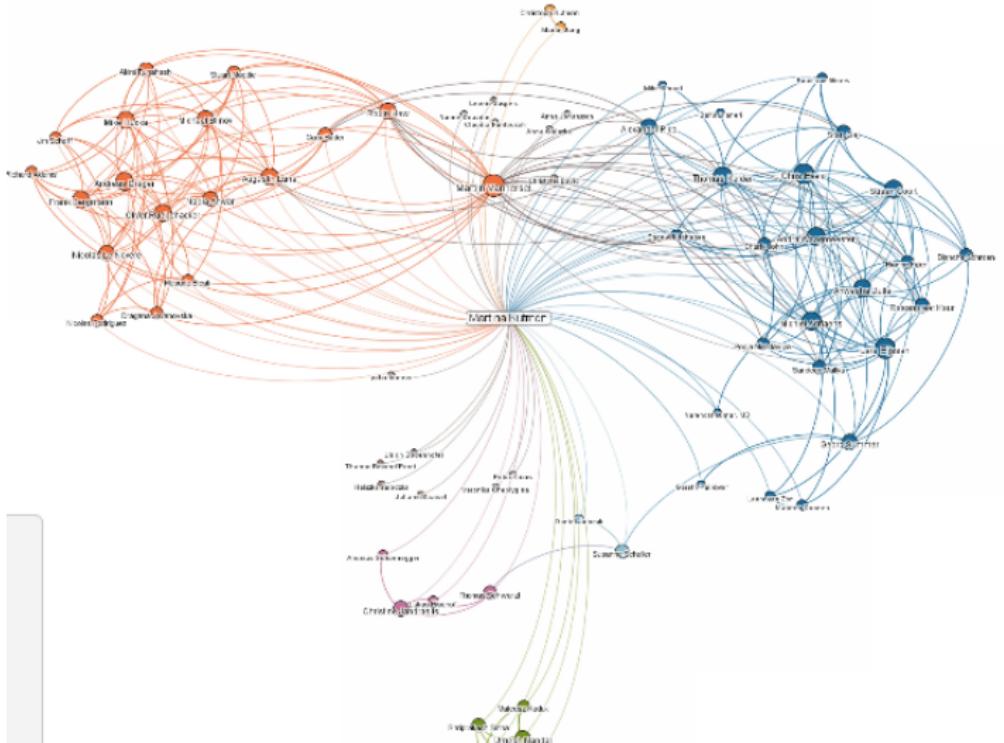
facebook.

g+

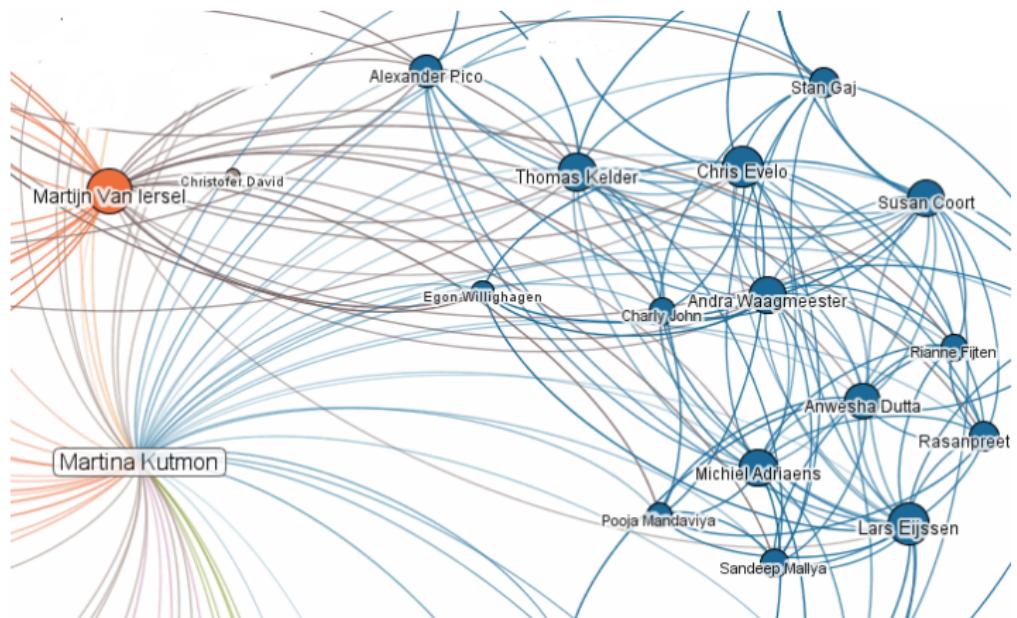
LinkedIn



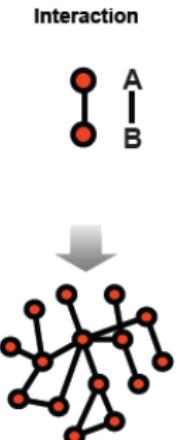
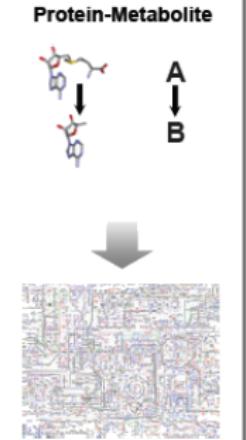
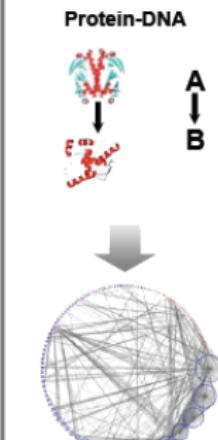
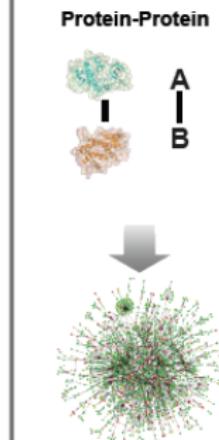
Social Networks



Social Networks



Biological Networks

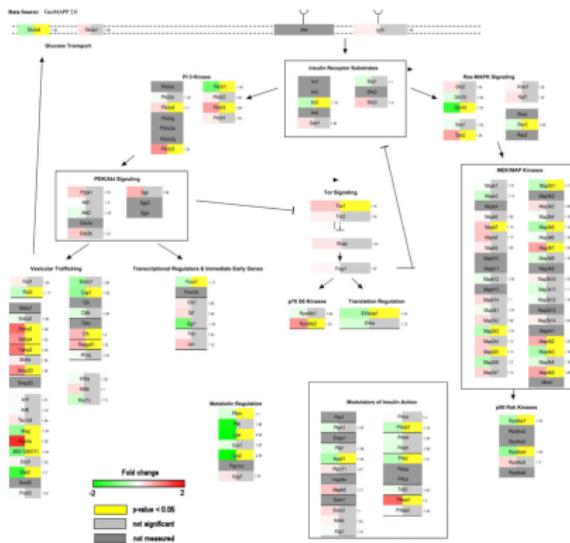
Network	Metabolic	Gene	Protein
Nodes	Metabolites	Transcription factor Target genes	Proteins
Links	Enzymatic conversion	Transcriptional Interaction	Physical Interaction
Interaction	Protein-Metabolite	Protein-DNA	Protein-Protein
			

What is a pathway?

Definition

A pathway is a diagram of a biological process.

It contains nodes (genes, proteins,...) and edges (interactions), but also descriptive elements, like labels or legends.



1 Introduction

- Learning goals
- Definitions

2 Graph Properties

- Definitions and explanations

3 Use Cases

- Use cases in biology

4 Algorithms

5 Online Resources

- Interaction Databases

6 Cytoscape

- Cytoscape

7 Summary

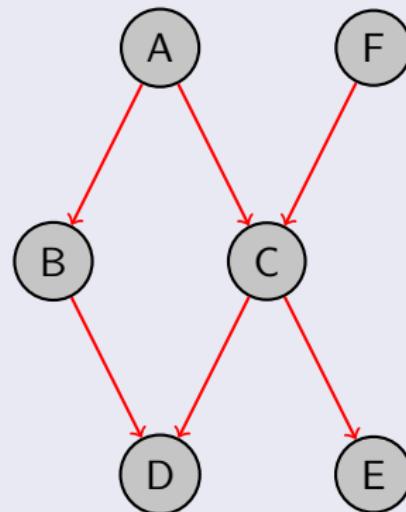
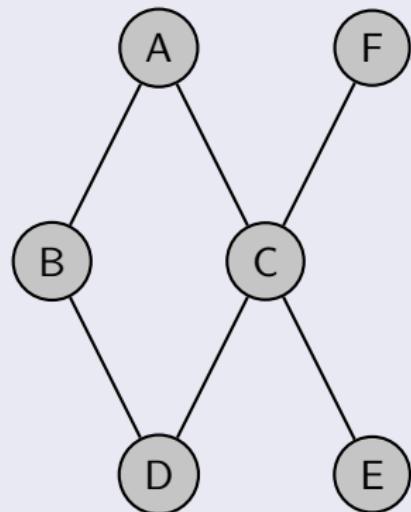
- Learning Goals

Direction

undirected \leftrightarrow directed

Direction

undirected \leftrightarrow directed

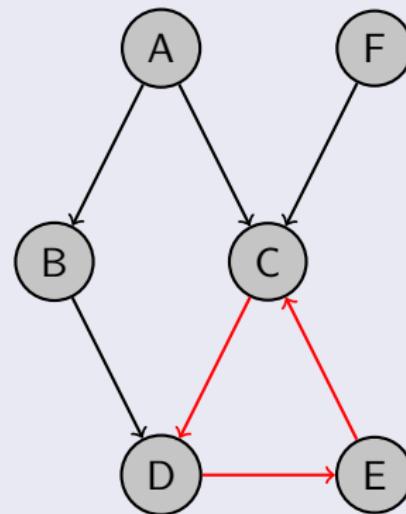
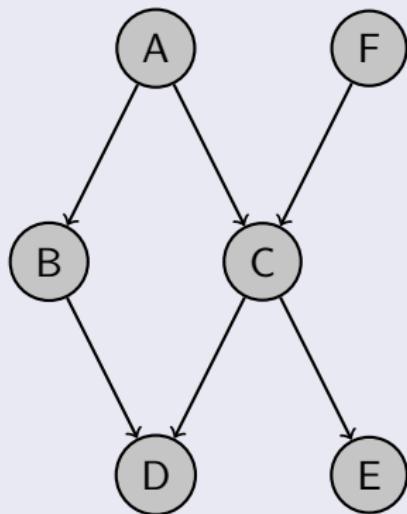


Cyclic

acyclic \leftrightarrow cyclic

Cyclic

acyclic \leftrightarrow cyclic

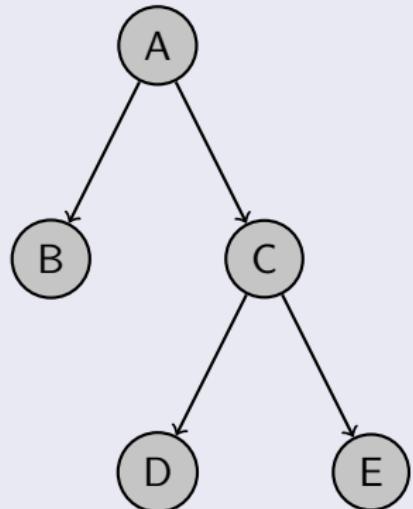


Hierarchical

ordered, directed, acyclic connected graph = tree

Hierarchical

ordered, directed, acyclic connected graph = tree



- **Root node**
topmost node without a parent
- **Inner node**
a node in the tree that has a parent and children
- **Leaf node**
the bottom node that does not have any child nodes.

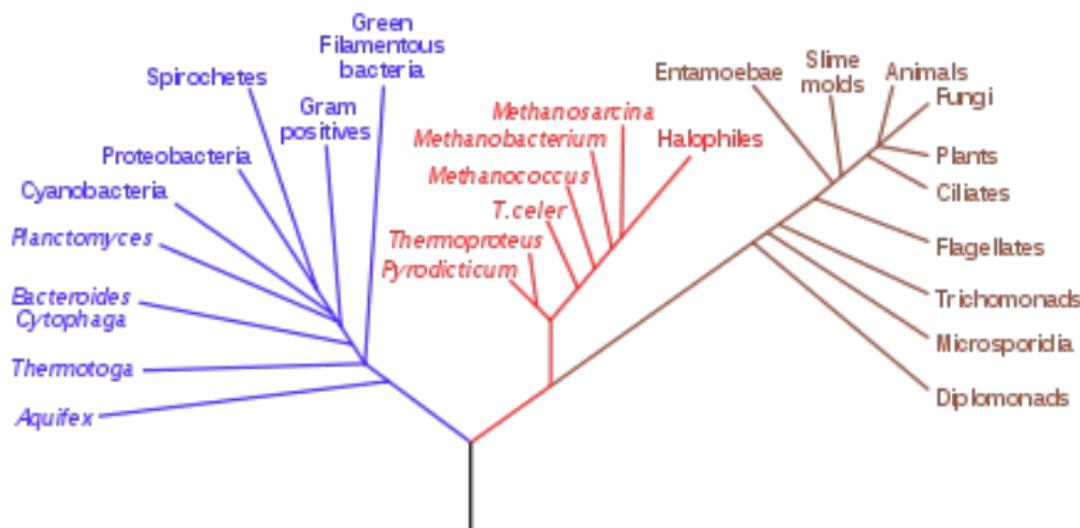
Phylogenetic tree

Phylogenetic Tree of Life

Bacteria

Archaea

Eucaryota



Centrality

Centrality measures the relative importance of a vertex within the graph

- how influential is a person within a social network?
- how important is a room within a building?
- how important is a transcription factor in a biological process?
- how much influence has a mutation of a protein?

Degree Centrality

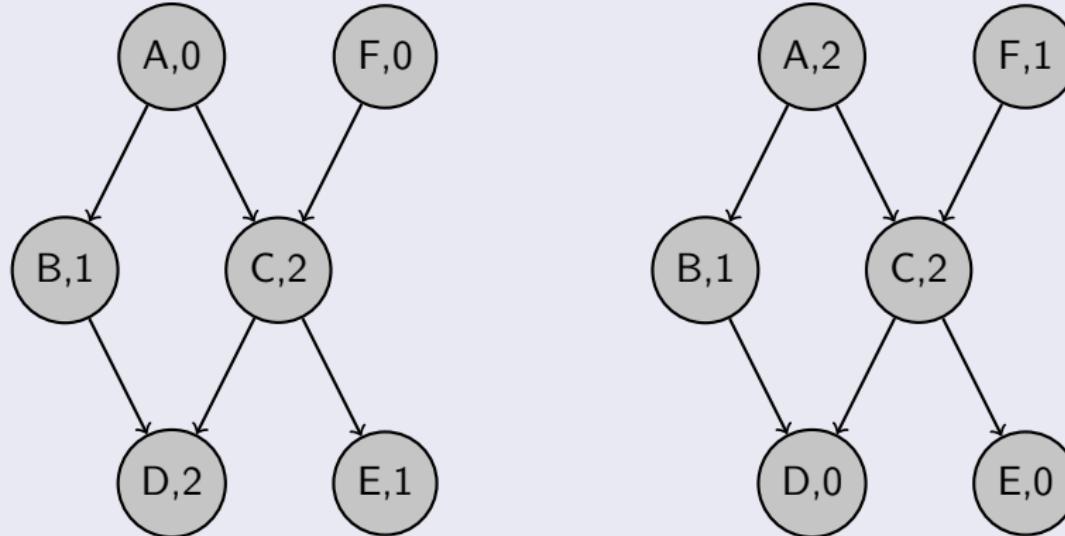
In-degree / Out-degree

Degree centrality is the total number of edges incident upon a node.

Degree Centrality

In-degree / Out-degree

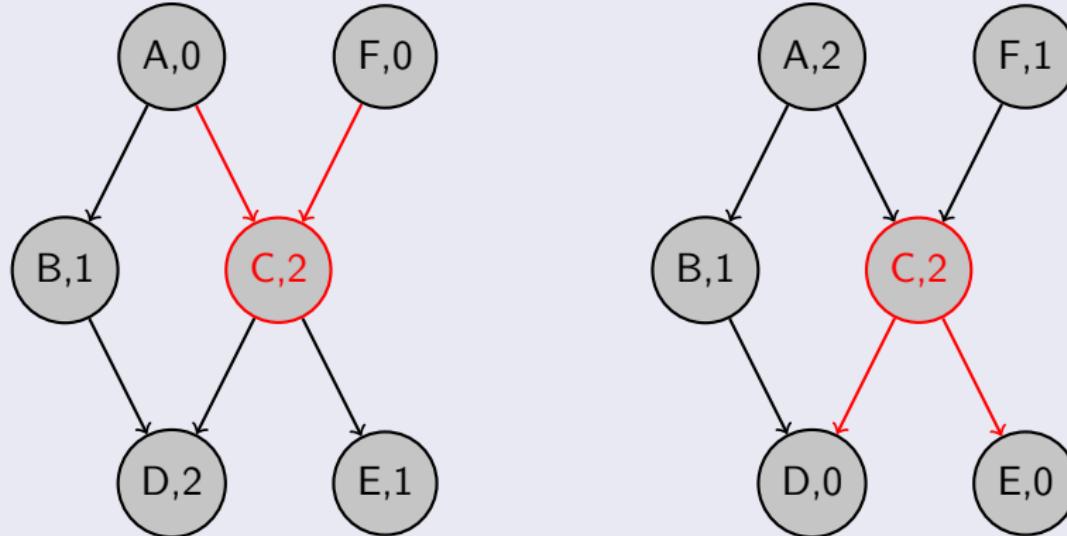
Degree centrality is the total number of edges incident upon a node.



Degree Centrality

In-degree / Out-degree

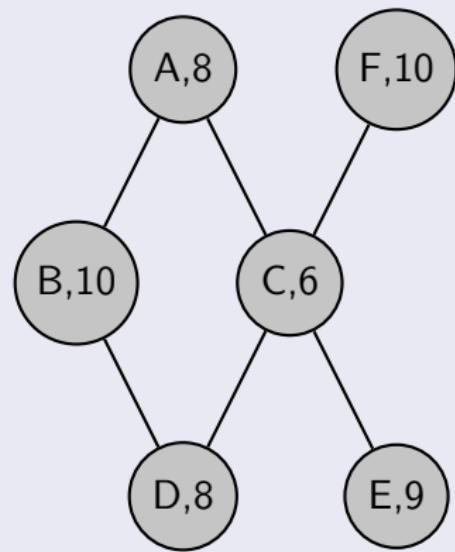
Degree centrality is the total number of edges incident upon a node.



Closeness Centrality

Closeness

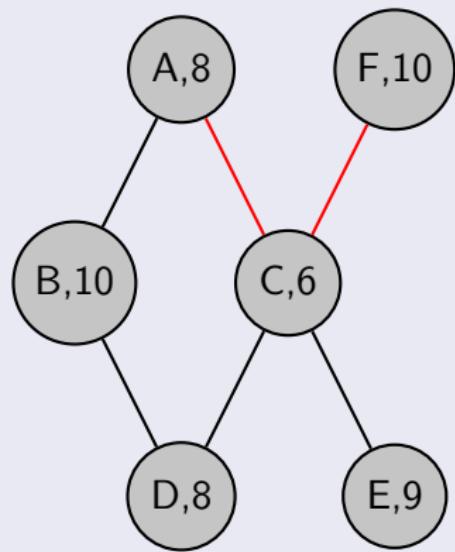
Closeness is the sum of the shortest-path distances to all other nodes in the network.



Closeness Centrality

Closeness

Closeness is the sum of the shortest-path distances to all other nodes in the network.

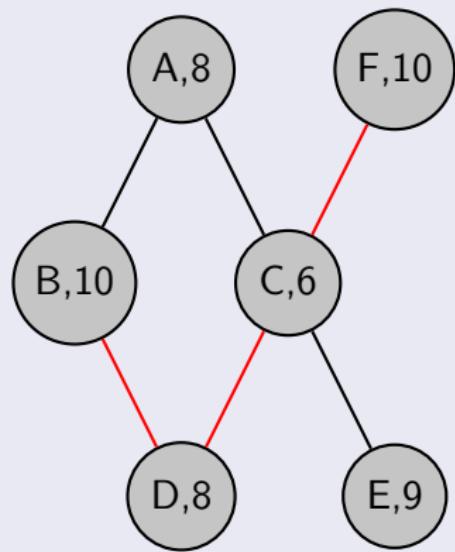


$f \rightarrow a = 2$ ($f-c-a$ or $f-c-d-b$)
 $f \rightarrow b = 3$ ($f-c-d-b$ or $f-c-a-b$)
 $f \rightarrow c = 1$
 $f \rightarrow d = 2$ ($f-c-d$ or $f-c-a-b-d$)
 $f \rightarrow e = 2$ ($f-c-e$)

Closeness Centrality

Closeness

Closeness is the sum of the shortest-path distances to all other nodes in the network.



$$f \rightarrow a = 2 \text{ (} f-c-a \text{ or } f-c-d-b \text{)}$$

$$f \rightarrow b = 3 \text{ (} f-c-d-b \text{ or } f-c-a-b \text{)}$$

$$f \rightarrow c = 1$$

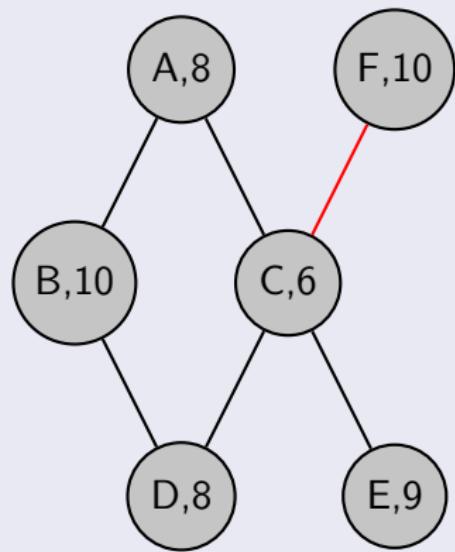
$$f \rightarrow d = 2 \text{ (} f-c-d \text{ or } f-c-a-b-d \text{)}$$

$$f \rightarrow e = 2 \text{ (} f-c-e \text{)}$$

Closeness Centrality

Closeness

Closeness is the sum of the shortest-path distances to all other nodes in the network.

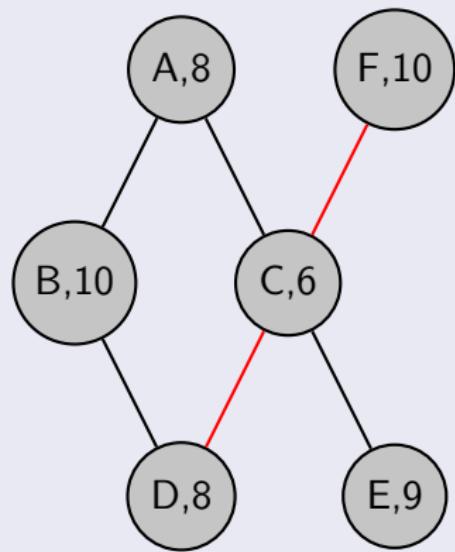


$f \rightarrow a = 2$ ($f-c-a$ or $f-c-d-b$)
 $f \rightarrow b = 3$ ($f-c-d-b$ or $f-c-a-b$)
 $f \rightarrow c = 1$
 $f \rightarrow d = 2$ ($f-c-d$ or $f-c-a-b-d$)
 $f \rightarrow e = 2$ ($f-c-e$)

Closeness Centrality

Closeness

Closeness is the sum of the shortest-path distances to all other nodes in the network.

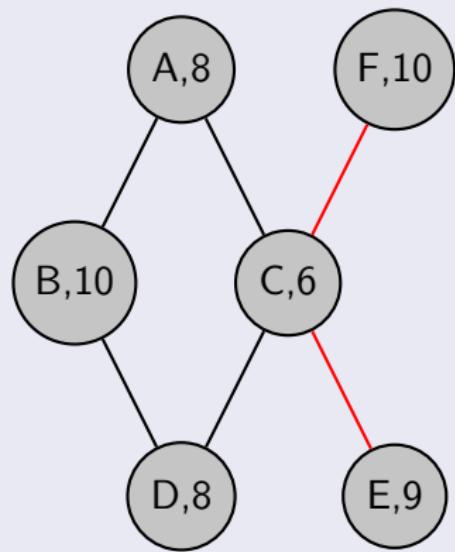


$f \rightarrow a = 2$ ($f-c-a$ or $f-c-d-b$)
 $f \rightarrow b = 3$ ($f-c-d-b$ or $f-c-a-b$)
 $f \rightarrow c = 1$
 $f \rightarrow d = 2$ ($f-c-d$ or $f-c-a-b-d$)
 $f \rightarrow e = 2$ ($f-c-e$)

Closeness Centrality

Closeness

Closeness is the sum of the shortest-path distances to all other nodes in the network.



$f \rightarrow a = 2$ ($f-c-a$ or $f-c-d-b$)
 $f \rightarrow b = 3$ ($f-c-d-b$ or $f-c-a-b$)
 $f \rightarrow c = 1$
 $f \rightarrow d = 2$ ($f-c-d$ or $f-c-a-b-d$)
 $f \rightarrow e = 2$ ($f-c-e$)

Betweenness Centrality

Betweenness

Betweenness of a node is the number of shortest paths from all vertices to all others that pass through that node.

- useful measure of the load placed on the given node in the network
- identification of key elements in a network (e.g. a transcription factor that needs to be present)

1 Introduction

- Learning goals
- Definitions

2 Graph Properties

- Definitions and explanations

3 Use Cases

- Use cases in biology

4 Algorithms

5 Online Resources

- Interaction Databases

6 Cytoscape

- Cytoscape

7 Summary

- Learning Goals

Use Cases

How are networks used in biology? Some examples:

- ① Identification of hubs
- ② Clustering
- ③ Network motifs and graphlets
- ④ Active subnetworks
- ⑤ Data integration

Use Cases

How are networks used in biology? Some examples:

- ① Identification of hubs
- ② Clustering
- ③ Network motifs and graphlets
- ④ Active subnetworks
- ⑤ Data integration

We will not discuss the topic of *network inference* in this lecture.

(1) Identification of hubs

Definition

A hub is a node in a network that is highly connected with the other nodes in the network, e.g. a node with a high betweenness centrality.

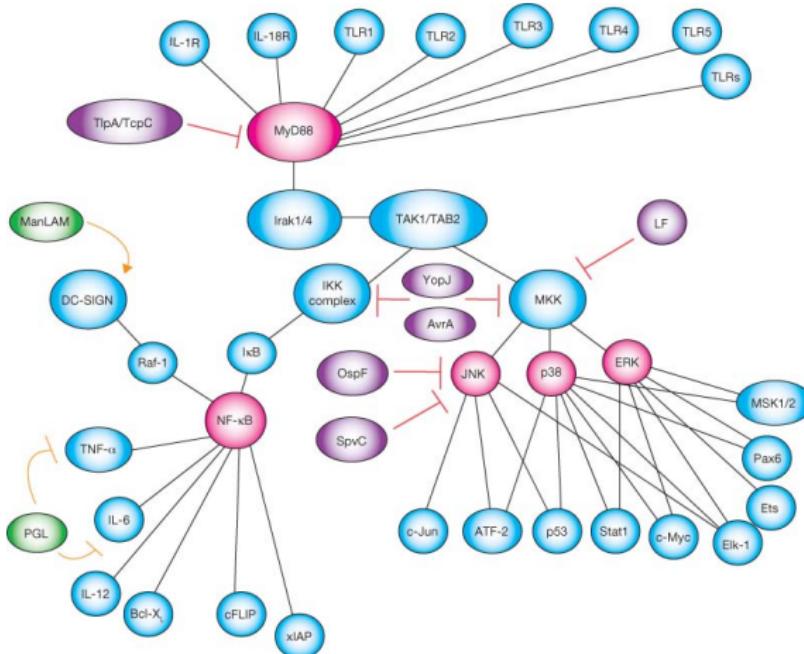
(1) Identification of hubs

Definition

A hub is a node in a network that is highly connected with the other nodes in the network, e.g. a node with a high betweenness centrality.

- Proteins that are hubs tend to be essential.
- Hubs are often involved in multiple processes.
- Deleting a hub reduces the overall connectivity of the network

(1) Identification of hubs



Targeting of immune signalling networks by bacterial pathogens, Igor E. Brodsky & Ruslan Medzhitov, Nature Cell Biology

(2) Clustering

Definition

Clustering is the finding of patterns in data, or the grouping of similar groups of nodes into clusters.

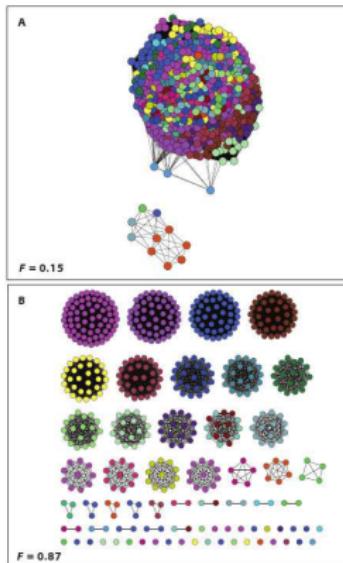
(2) Clustering

Definition

Clustering is the finding of patterns in data, or the grouping of similar groups of nodes into clusters.

- Many different algorithms (e.g. k-means, EM clustering, spectral clustering, etc.).
- Grouping of nodes that are close together.
- Usually the goal is to have many nodes *within* each cluster and few edges *between* clusters.

(2) Clustering



**Improving the Quality of Protein Similarity Network Clustering Algorithms
using the Network Edge Weight Distribution, Apeltsin et al, Bioinformatics**

(3) Network motifs and graphlets

Definition

Network motifs are significantly recurring small subnetworks. Each type of network seems to display its own set of characteristic motifs (ecological networks have different motifs than gene regulation networks, etc.).

(3) Network motifs and graphlets

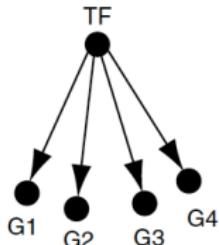
Definition

Network motifs are significantly recurring small subnetworks. Each type of network seems to display its own set of characteristic motifs (ecological networks have different motifs than gene regulation networks, etc.).

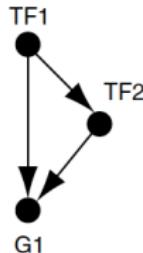
Definition

Graphlets are small connected non-isomorphic *induced* subgraphs of a large network. An induced subgraph must contain all edges between its nodes that are present in a large network.

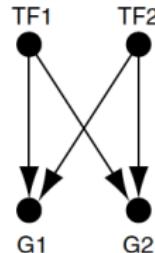
(3) Network motifs and graphlets



A



B

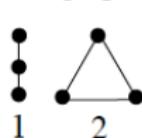


C

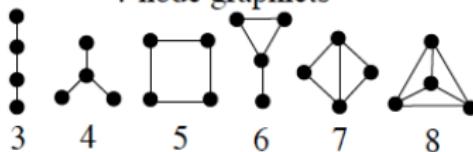
Network motifs are enriched with transcription factors whose transcripts have short half-lives, Edwin Wang, Enrico Purisima, *Trends in Genetics*

(3) Network motifs and graphlets

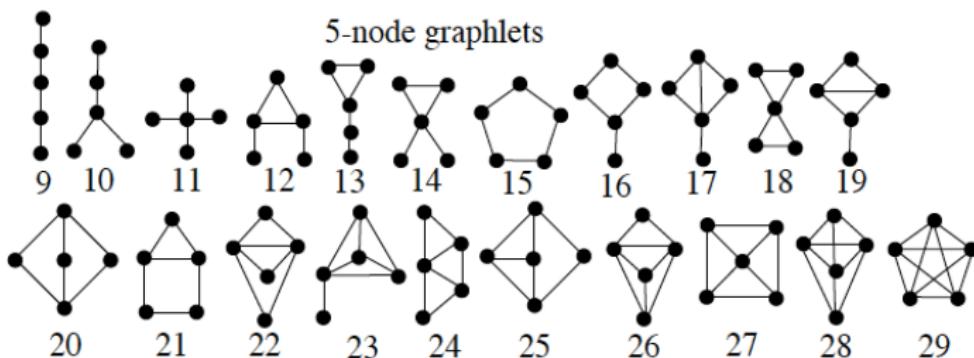
3-node graphlets



4-node graphlets



5-node graphlets



**Modeling interactome: scale-free or geometric?, N. Prulj, D. G. Corneil,
Bioinformatics**

(4) Active subnetworks

Definition

Active networks change over time and response to external signals.
Active networks are likely different between healthy and diseased cells.

(4) Active subnetworks

Definition

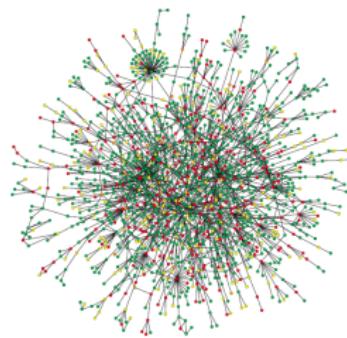
Active networks change over time and response to external signals.
Active networks are likely different between healthy and diseased cells.

- Gene expression data and interaction network necessary
- Identification of a connected subnetwork in which genes show a significant expression change between two samples
- Different tools and approaches available (jActiveModules in Cytoscape, heinz, DEGAS)

(5) Data Integration

Challenge

Construct and interpret biological networks of all interactions in a living cells, making use of all available data.



(5) Data Integration

Integration of

- genes
- transcripts
- SNPs
- proteins
- metabolites
- transcription factors
- miRNAs
- ...

(5) Data Integration

Integration of

- protein-protein interactions
- protein-DNA interactions
- regulatory interactions
- metabolic reactions
- metabolite-protein interactions
- ...

(5) Data Integration

Integration of

- official gene symbols/protein names
- GeneOntology terms
- protein families
- protein location
- chromosomal location
- ...

(5) Data Integration

Scale of complexity

-

Proteins as networks



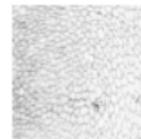
-

Molecular interactions as networks



-

Cell contacts as networks



1 Introduction

- Learning goals
- Definitions

2 Graph Properties

- Definitions and explanations

3 Use Cases

- Use cases in biology

4 Algorithms

5 Online Resources

- Interaction Databases

6 Cytoscape

- Cytoscape

7 Summary

- Learning Goals

Algorithms

Many different algorithms to solve problems in graphs

- find shortest path
- traveling salesman problem
- find reoccurring graphlets
- find hubs in networks (see *Network biology: Why do we need hubs?*, Nature Reviews Genetics)
- cluster nodes in a network
- ...

Dijkstra's algorithm to find the shortest path



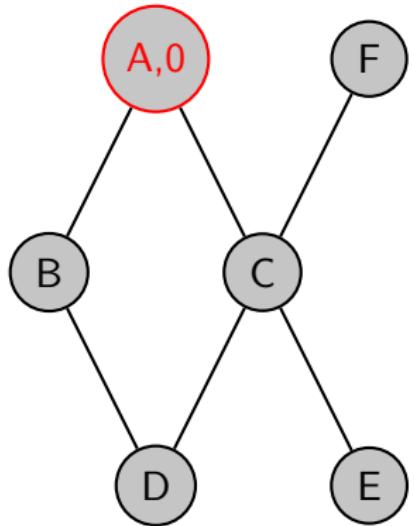
Edsger W. Dijkstra

- Dutch computer scientist
- published in 1959 a graph search algorithm that solves the single-source shortest path problem for a graph
- A note on two problems in connexion with graphs,
Numerische Mathematik 1: 269271.

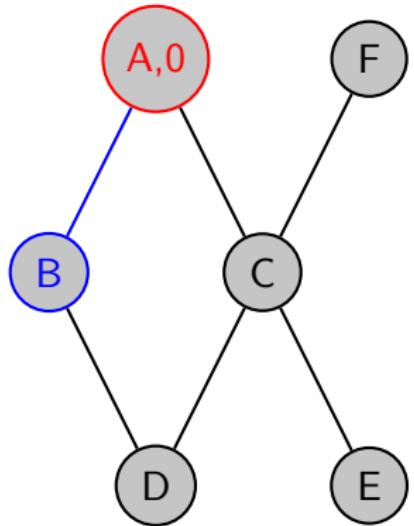
How can you find the shortest path between two nodes in a network?

- ① Initial node gets path length 0. All other nodes are unvisited.
Initial node is current node.
- ② Calculate the tentative distance for all neighbours of the current node. If the node has a tentative distance bigger than the current distance, replace it with the new distance. When all neighbours are examined, mark current node as visited.
- ③ Repeat this for all nodes. A visited node will never be checked again. Its distance recorded is final and minimal.

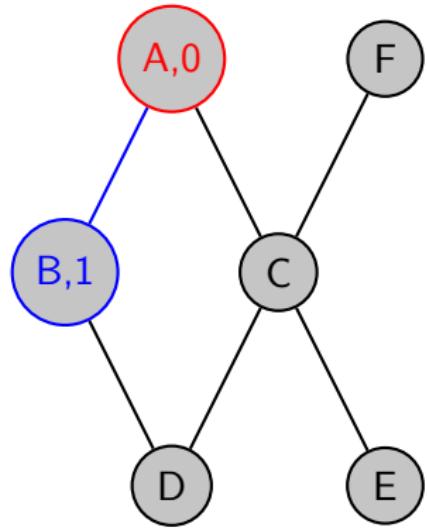
Dijkstra's algorithm to find the shortest path



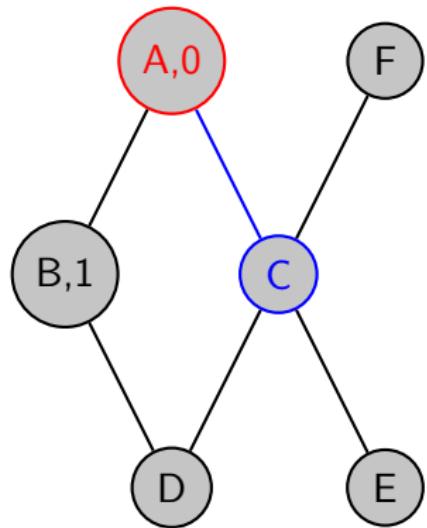
Dijkstra's algorithm to find the shortest path



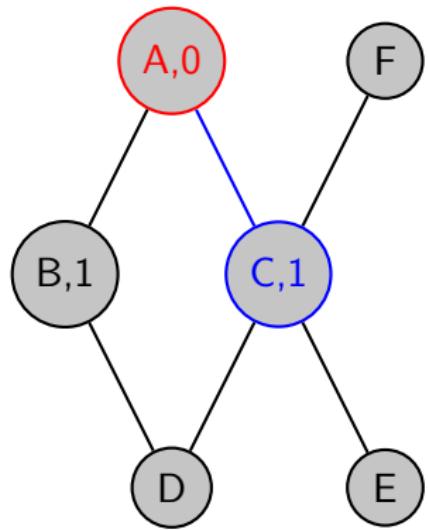
Dijkstra's algorithm to find the shortest path



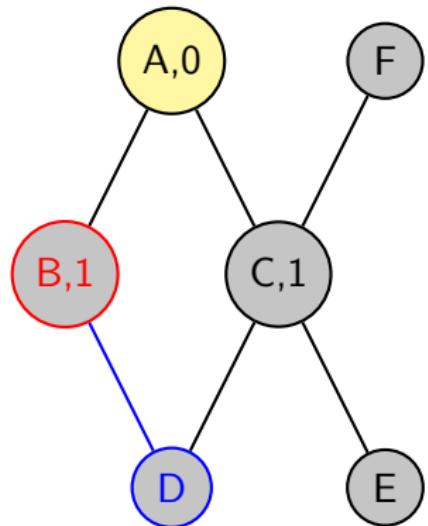
Dijkstra's algorithm to find the shortest path



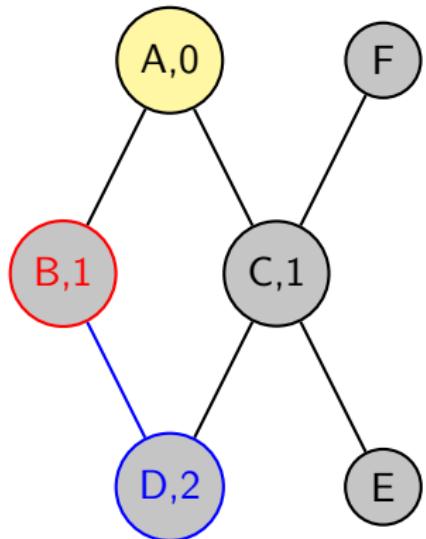
Dijkstra's algorithm to find the shortest path



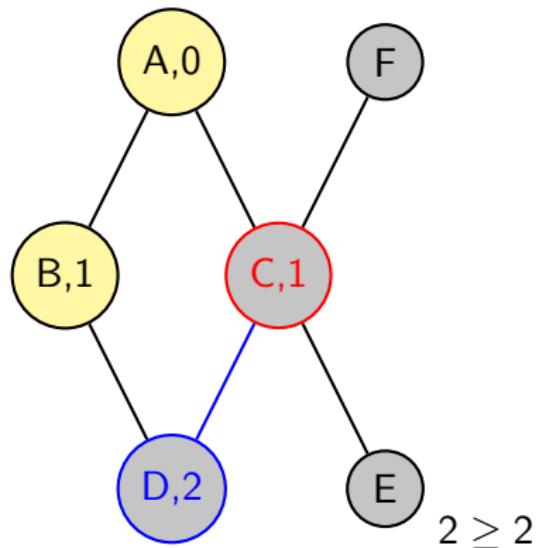
Dijkstra's algorithm to find the shortest path



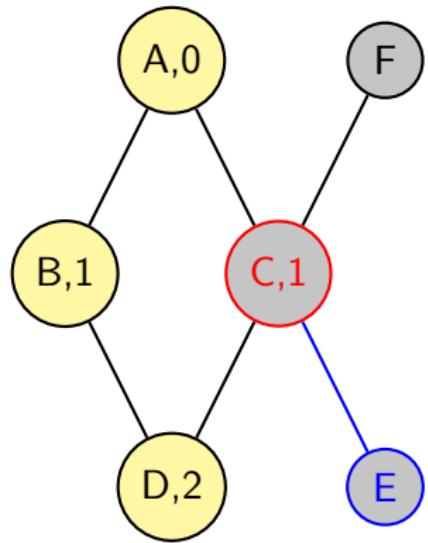
Dijkstra's algorithm to find the shortest path



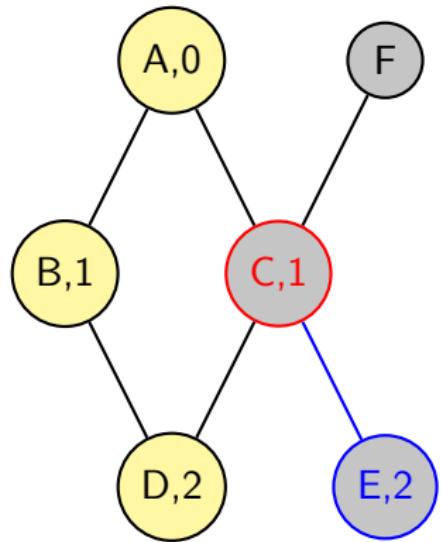
Dijkstra's algorithm to find the shortest path

 $2 \geq 2$

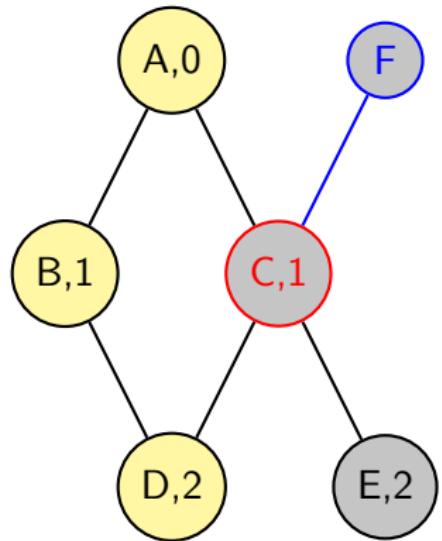
Dijkstra's algorithm to find the shortest path



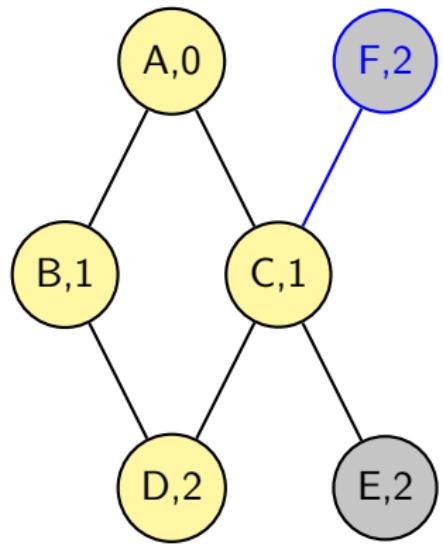
Dijkstra's algorithm to find the shortest path



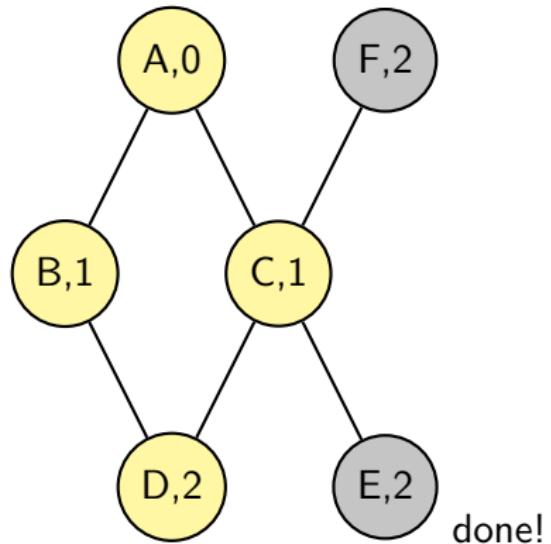
Dijkstra's algorithm to find the shortest path



Dijkstra's algorithm to find the shortest path



Dijkstra's algorithm to find the shortest path

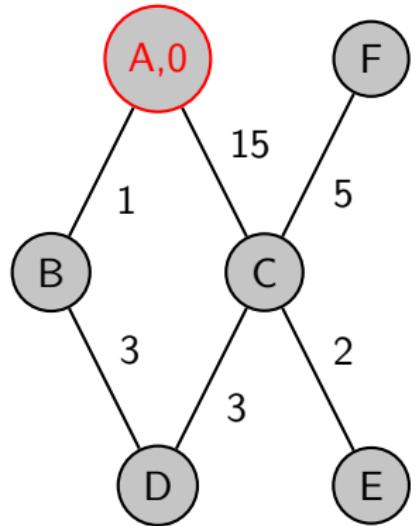


Weighted graph

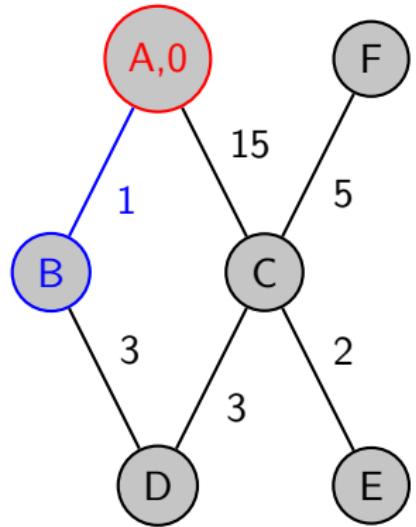
But what happens when the edges have a weight?



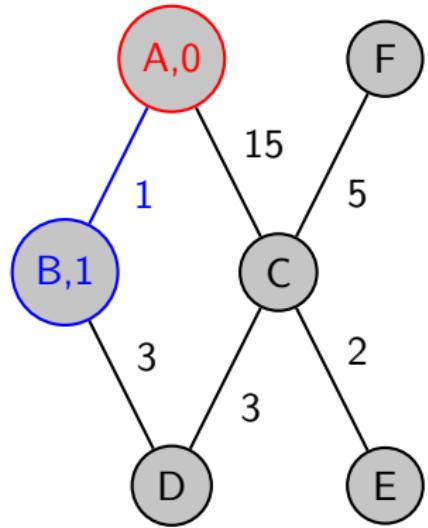
Dijkstra's algorithm to find the shortest path



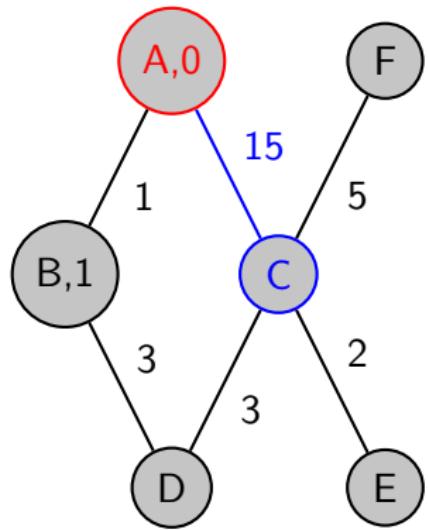
Dijkstra's algorithm to find the shortest path



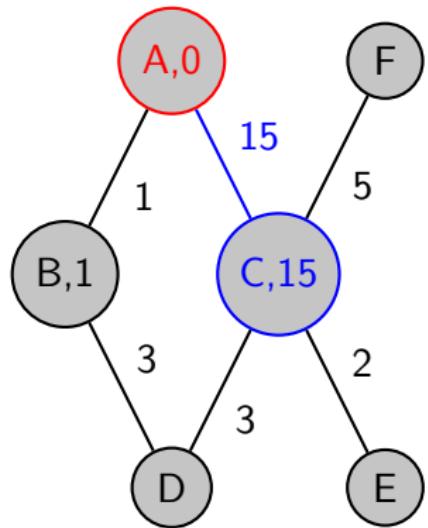
Dijkstra's algorithm to find the shortest path



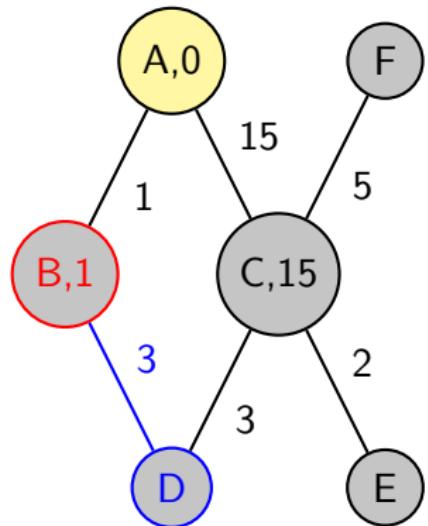
Dijkstra's algorithm to find the shortest path



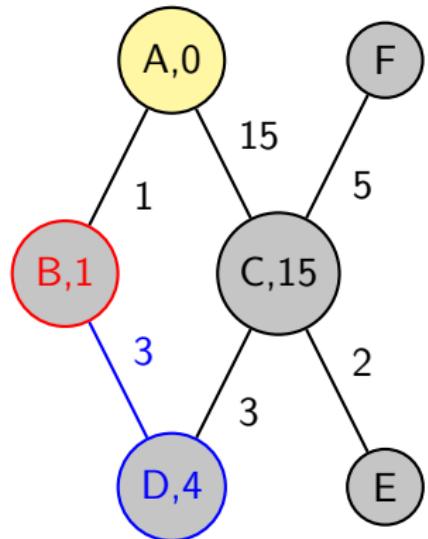
Dijkstra's algorithm to find the shortest path



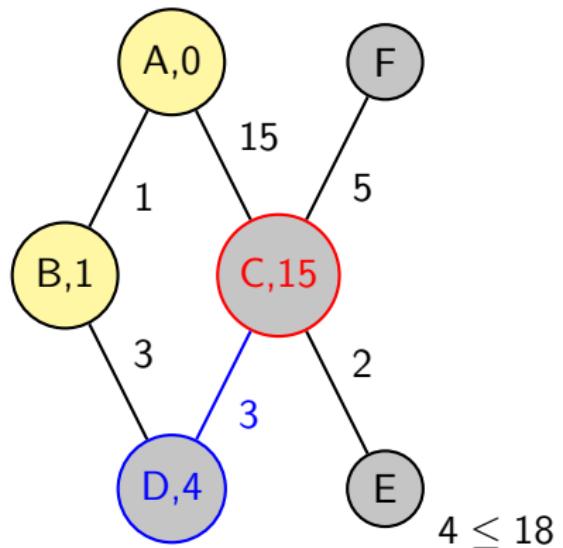
Dijkstra's algorithm to find the shortest path



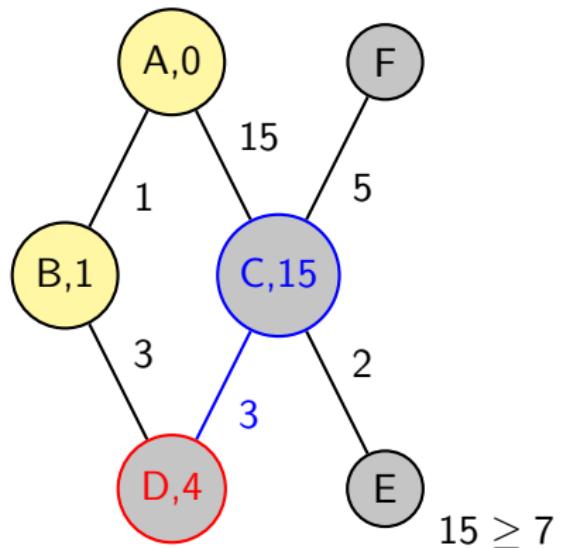
Dijkstra's algorithm to find the shortest path



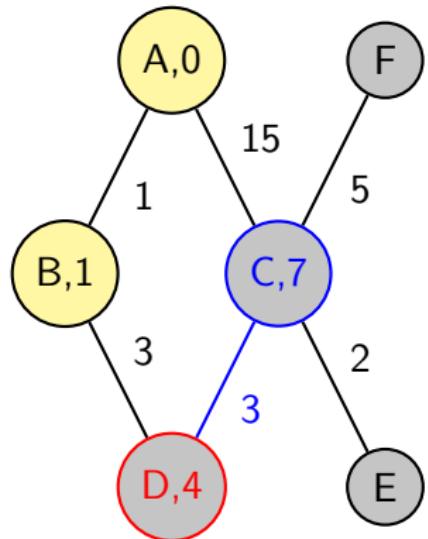
Dijkstra's algorithm to find the shortest path



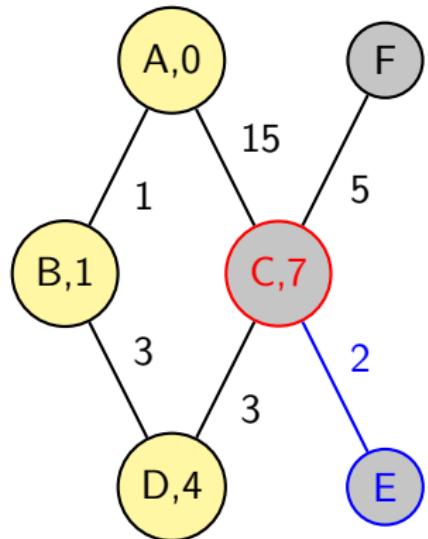
Dijkstra's algorithm to find the shortest path



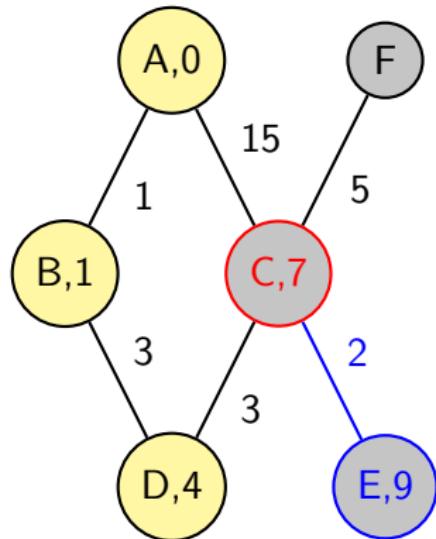
Dijkstra's algorithm to find the shortest path



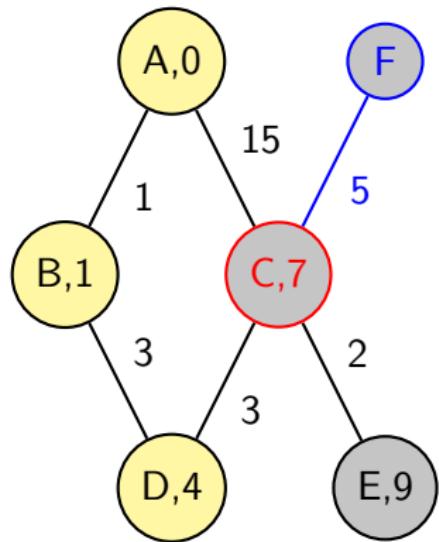
Dijkstra's algorithm to find the shortest path



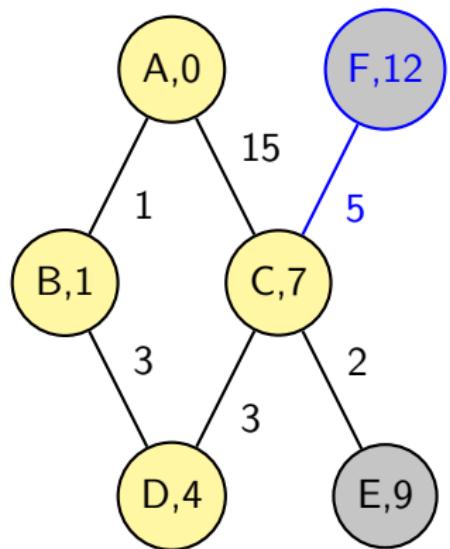
Dijkstra's algorithm to find the shortest path



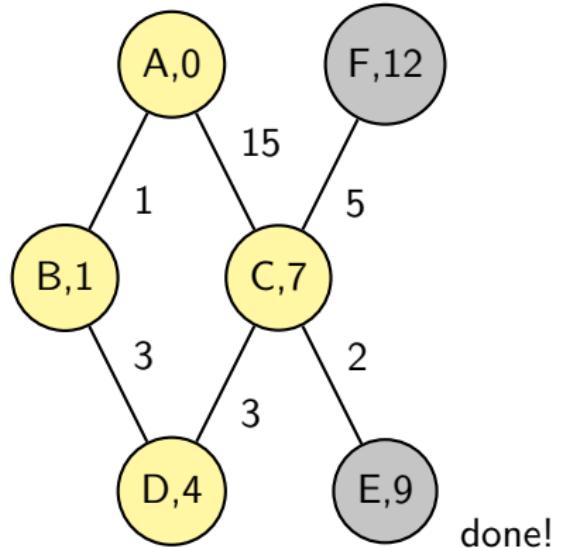
Dijkstra's algorithm to find the shortest path



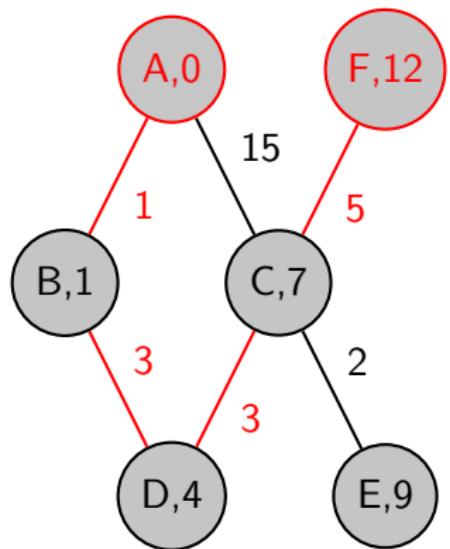
Dijkstra's algorithm to find the shortest path



Dijkstra's algorithm to find the shortest path



Dijkstra's algorithm to find the shortest path



Demo

Demo of DijkstraVis

<http://www.uweschmidt.org/projects/dijkstravis>

1 Introduction

- Learning goals
- Definitions

2 Graph Properties

- Definitions and explanations

3 Use Cases

- Use cases in biology

4 Algorithms

5 Online Resources

- Interaction Databases

6 Cytoscape

- Cytoscape

7 Summary

- Learning Goals

Online Resources

www.pathguide.org

Interaction type	Databases
Protein-Protein interactions	IntAct, STRING, MINT
Metabolic reactions	BRENDA, HMDB, Reactome, KEGG
Gene regulatory interactions	Pazar, JASPAR, RegulonDB, TFe, miRecords
Protein-Compound interactions	DrugBank, BioDrugScreen, Ligand Expo, Pathway Commons
Genetic interactions	BIND, GeneMANIA, Interactome

DrugBank

DRUGBANK

Open Data Drug & Drug Target Database

[Home](#)[Browse](#)[Search](#)[Downloads](#)[News & Updates](#)[About](#)[Help](#)[Contact Us](#)[Help / Advanced](#)

The DrugBank database is a unique bioinformatics and cheminformatics resource that combines detailed drug (i.e. chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e. sequence, structure, and pathway) information. The database contains 6712 drug entries including 1441 FDA-approved small molecule drugs, 135 FDA-approved biotech (protein/peptide) drugs, 84 nutraceuticals and 5084 experimental drugs. Additionally, 4227 non-redundant protein (i.e. drug target/enzyme/transporter/carrier) sequences are linked to these drug entries. Each DrugCard entry contains more than 150 data fields with half of the information being devoted to drug/chemical data and the other half devoted to drug target or protein data.



DrugBank is supported by [David Wishart](#), Departments of [Computing Science](#) & [Biological Sciences](#), [University of Alberta](#).

Pazar



A PUBLIC DATABASE OF TRANSCRIPTION FACTOR AND REGULATORY SEQUENCE ANNOTATION

[Sign in](#)

Search PAZAR Mall [View interactive mall map](#)
(requires Adobe Flash player)

[Transcription factors \(TFs\)](#) [Target genes](#) [Regulatory sequences](#) [Pre-computed TF profiles](#) [Browse projects](#)

Enter a TF to search

Select a TF from our list of reported TFs • [View advanced search options](#)

Welcome to PAZAR!

PAZAR is your one stop shopping experience for transcription factors and regulatory sequence annotations. It is a software framework for the construction and maintenance of regulatory sequence data annotations; a framework which allows multiple boutique databases to function independently within a larger system (or information mall). Our goal is to be the public repository for regulatory data. [View our publications](#)

[View our past webinars](#) (latest: January 25, 2012)

SEARCH
LOGIN
MY PROJECTS
SUBMIT
OVERVIEW

STRING database

[Home](#) · [Download](#) · [Help/Info](#) **STRING** 9.0**STRING - Known and Predicted Protein-Protein Interactions**

[search by name](#) [search by protein sequence](#) [multiple names](#) [multiple sequences](#)

protein name: (examples: #1 #2 #3)

(STRING understands a variety of protein names and accessions; you can also try a [random entry](#))

organism:

auto-detect

interactors wanted:

COGs Proteins

Reset

GO !

please enter your protein of interest...

What it does ...

STRING is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations; they are derived from four sources:

Genomic Context



High-throughput Experiments



(Conserved) Coexpression



Previous Knowledge



STRING quantitatively integrates interaction data from these sources for a large number of organisms, and transfers information between these organisms where applicable. The database currently covers 5'214'234 proteins from 1133 organisms.

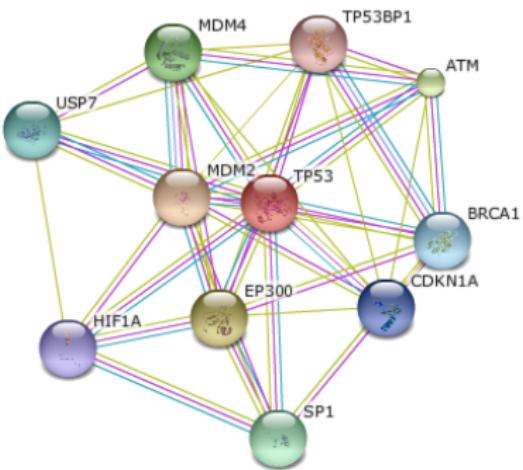
STRING database



What is the STRING database?

- www.string-db.org
- Known and predicted protein-protein interactions
- Direct (physical) and indirect (functional) associations
- Data is derived from four different sources
 - Genomic Context
 - High-throughput experiments
 - (conserved) Coexpression
 - Previous knowledge
- Version 9.0 contains 5,214,234 proteins from 1,133 organisms

STRING database

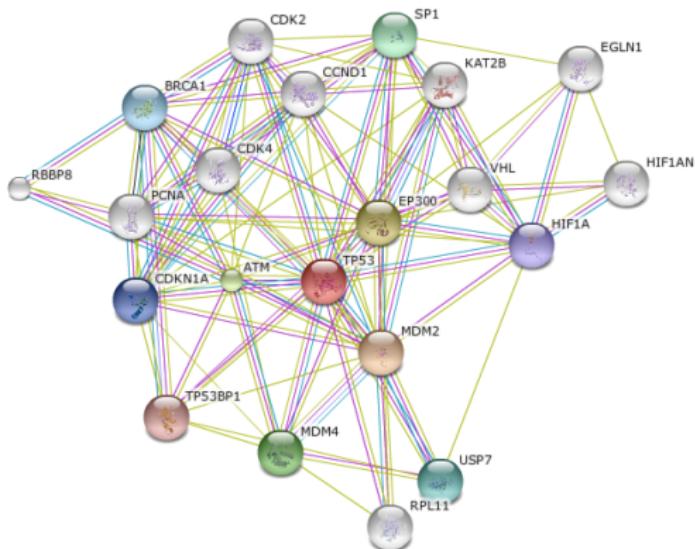


This is the **evidence view**. Different line colors represent the types of evidence for the association.



(requires Flash player 10 or better)

STRING database



This is the **evidence view**. Different line colors represent the types of evidence for the association.



(requires Flash player 10 or better)

STRING database

Predicted Functional Partners:

		Neighborhood	Gene Fusion	Cooccurrence	Experiments	Databases	Textmining	[Homology]	Score
● MDM2	Mdm2 p53 binding protein homolog (mouse); Inhibits TP53/p53- and TP73/p73-mediated cell cycle a [...] (497 aa)		●	●	●				0.999
● EP300	microRNA 1281; Functions as histone acetyltransferase and regulates transcription via chromatin [...] (2414 aa)		●	●	●				0.999
● ATM	ataxia telangiectasia mutated; Serine/threonine protein kinase which activates checkpoint signa [...] (3056 aa)		●	●	●				0.999
● MDM4	Mdm4 p53 binding protein homolog (mouse); Inhibits p53- and p73-mediated cell cycle arrest and [...] (490 aa)		●	●	●				0.999
● SP1	Sp1 transcription factor; Transcription factor that can activate or repress transcription in re [...] (785 aa)		●	●	●				0.999
● USP7	ubiquitin specific peptidase 7 (herpes virus-associated); Cleaves ubiquitin fusion protein subs [...] (1102 aa)		●	●	●				0.999
● BRCA1	breast cancer 1, early onset; The BRCA1-BARD1 heterodimer coordinates a diverse range of cell [...] (1863 aa)		●	●	●				0.999
● CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1); May be the important intermediate by which p5 [...] (164 aa)		●	●	●				0.999
● HIF1A	hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor); Functi [...] (826 aa)		●	●	●				0.999
● TP53BP1	tumor protein p53 binding protein 1; May have a role in checkpoint signaling during mitosis (By [...] (1977 aa)		●	●	●				0.999

1 Introduction

- Learning goals
- Definitions

2 Graph Properties

- Definitions and explanations

3 Use Cases

- Use cases in biology

4 Algorithms

5 Online Resources

- Interaction Databases

6 Cytoscape

- Cytoscape

7 Summary

- Learning Goals

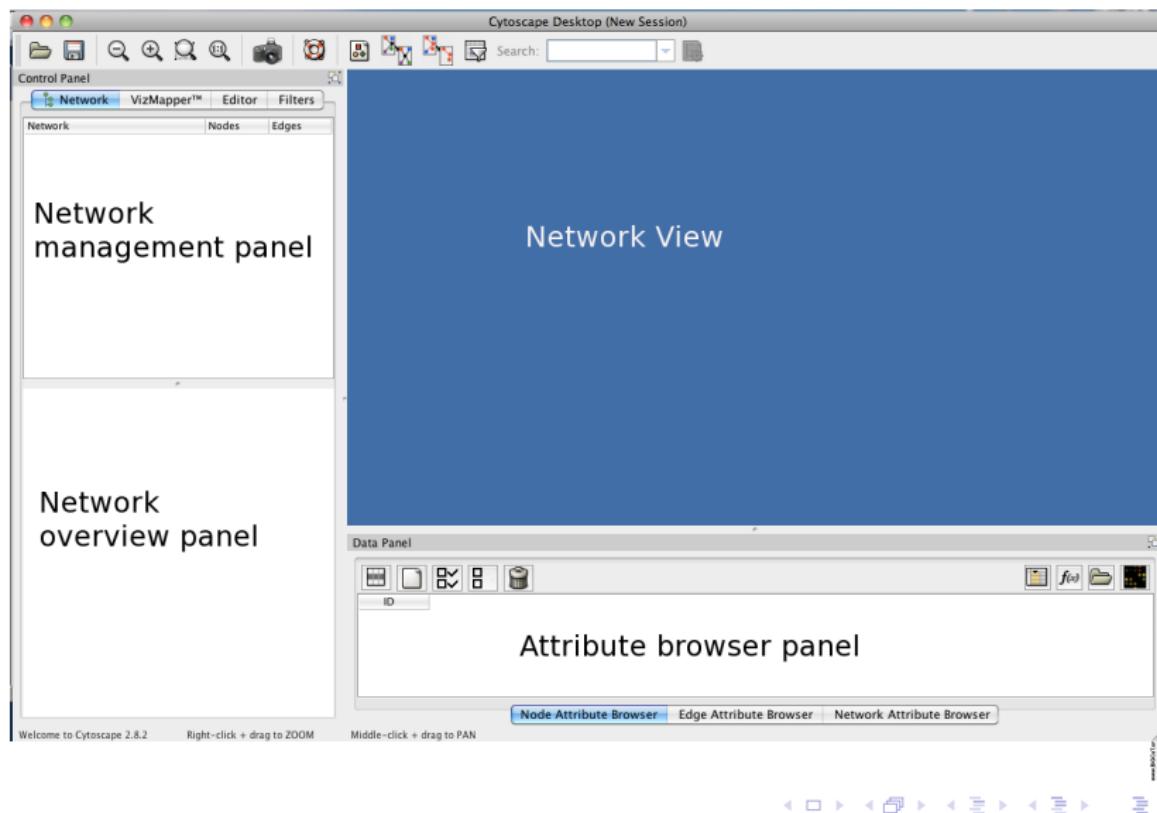
Cytoscape



What is Cytoscape?

- Cytoscape is a network visualization and analysis tool
- Functionality can be extended with plugins (currently 128)
 - Analysis
 - Network and Attribute I/O
 - Network Inference
 - Functional Enrichment
 - Communication / Scripting
 - Other
- 100% open source
- Current version is 2.8

Cytoscape



Practical Session



- Cytoscape online tutorial
- How to analyze a network (basic properties)?
- How to perform GO analysis in Cytoscape?
- How to extend a pathway with regulatory interactions?
- How to visualize expression data in a network?

1 Introduction

- Learning goals
- Definitions

2 Graph Properties

- Definitions and explanations

3 Use Cases

- Use cases in biology

4 Algorithms

5 Online Resources

- Interaction Databases

6 Cytoscape

- Cytoscape

7 Summary

- Learning Goals

What is a graph/network/pathway?

- A graph/network is an abstract representation of a set of objects where some pairs of the objects are connected by links.
- Objects are called nodes or vertices.
- Links are called edges or lines.
- Give some examples of networks.
- Pathway is mostly just a diagram, but can be used as network when all descriptive elements are removed.

What are the properties of a graph? What do they tell us?

- undirected ↔ directed
- acyclic ↔ cyclic
- hierarchical
- centrality: degree centrality, closeness centrality, betweenness centrality

How can we use networks in biology?

Just a few examples:

- Identification of hubs
- Clustering
- Network motifs and graphlets
- Active subnetworks
- Data integration

How can you find the shortest path between two nodes in a network?

- ① Initial node gets path length 0. All other nodes are unvisited.
Initial node is current node.
- ② Calculate the tentative distance for all neighbours of the current node. If the node has a tentative distance bigger than the current distance, replace it with the new distance. When all neighbours are examined, mark current node as visited.
- ③ Repeat this for all nodes. A visited node will never be checked again. Its distance recorded is final and minimal.

Where can you find network/interaction data?

- There are many different databases.
- In the practical session we will have a closer look at the STRING database.

How to use Cytoscape, the bioinformatics tool for network analysis?

- Cytoscape is a network visualization and analysis tool.
- We will use Cytoscape in the practical session to
 - analyze the properties of a network
 - perform a GO analysis
 - extend a pathway with regulatory interactions
 - visualize gene expression data on a network

Questions?

