Data Structures and Algorithms for Engineers

Module 9: Complex Networks

Lecture 1: The importance of complex networks and network science, review of graph theory.

David Vernon
Carnegie Mellon University Africa

vernon@cmu.edu
www.vernon.eu

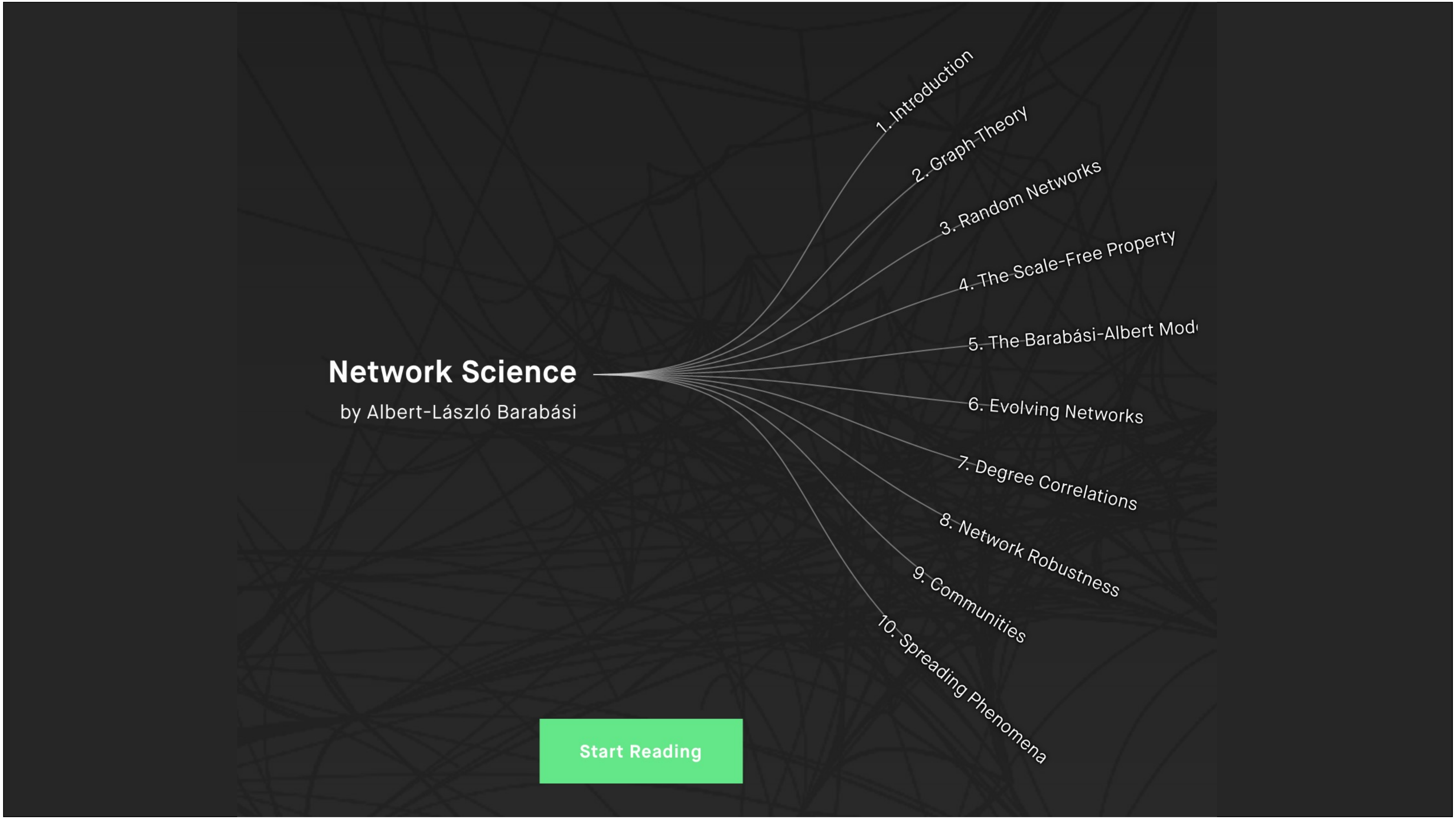
Lecture DSA09-01

Complex Networks

- The importance of complex networks and network science

- Review of graph theory
  - Euler's theorem: the Bridges of Königsberg
  - Networks vs. graphs
  - Degree, average degree, and degree distribution
  - Bipartite networks
  - Path length, BFS, Connectivity, Components
  - Clustering coefficient

This lecture is based on Chapters 1 and 2 of *Network Science* by A.-L. Barabási
(see https://networksciencebook.com/)


Network Science
by Albert-László Barabási
1. Introduction
2. Graph Theory
3. Random Networks
4. The Scale-Free Property
5. The Barabási-Albert Mod
6. Evolving Networks
7. Degree Correlations
8. Network Robustness
9. Communities
10. Spreading Phenomena
Start Reading

Complex Networks

Network Science

Economic Impact: From Web Search to Social Networking

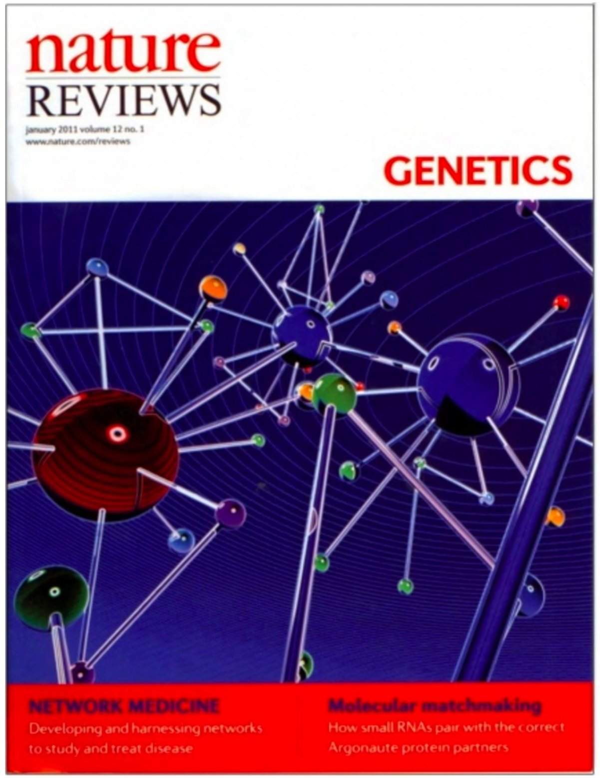
"The most successful companies of the 21st century, from Google to Facebook, Twitter, LinkedIn, Cisco, Apple and Akamai, base their technology and business model on networks"

A.-L. Barabási

Complex Networks

Network Science

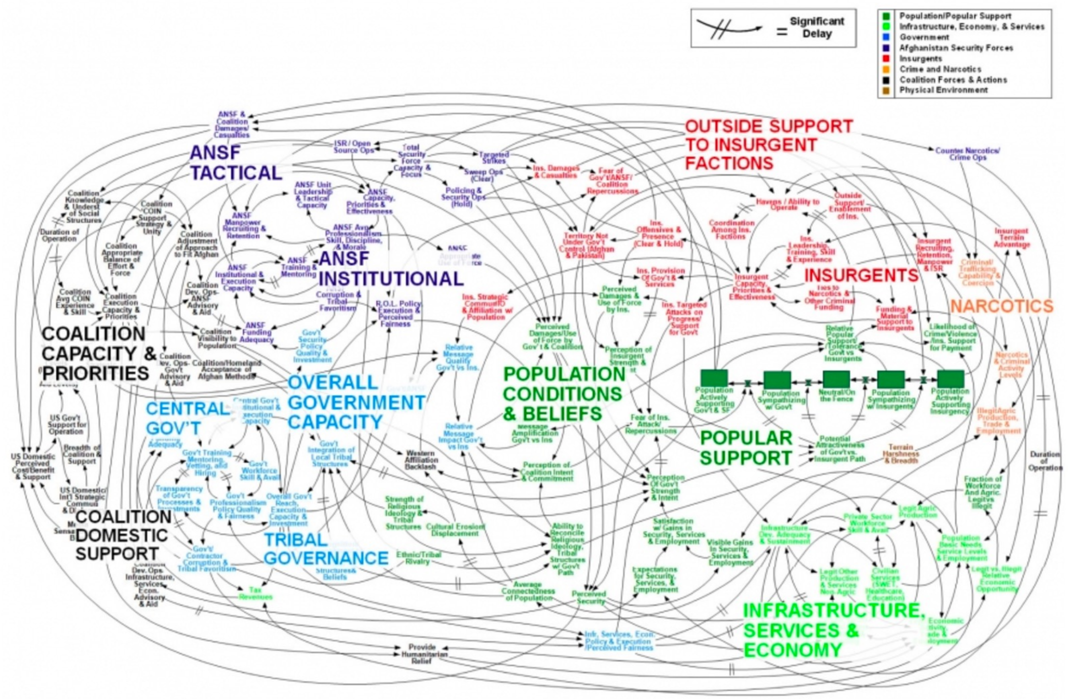
Health: From Drug Design to Metabolic Engineering



Complex Networks

Network Science

Security: Fighting Terrorism

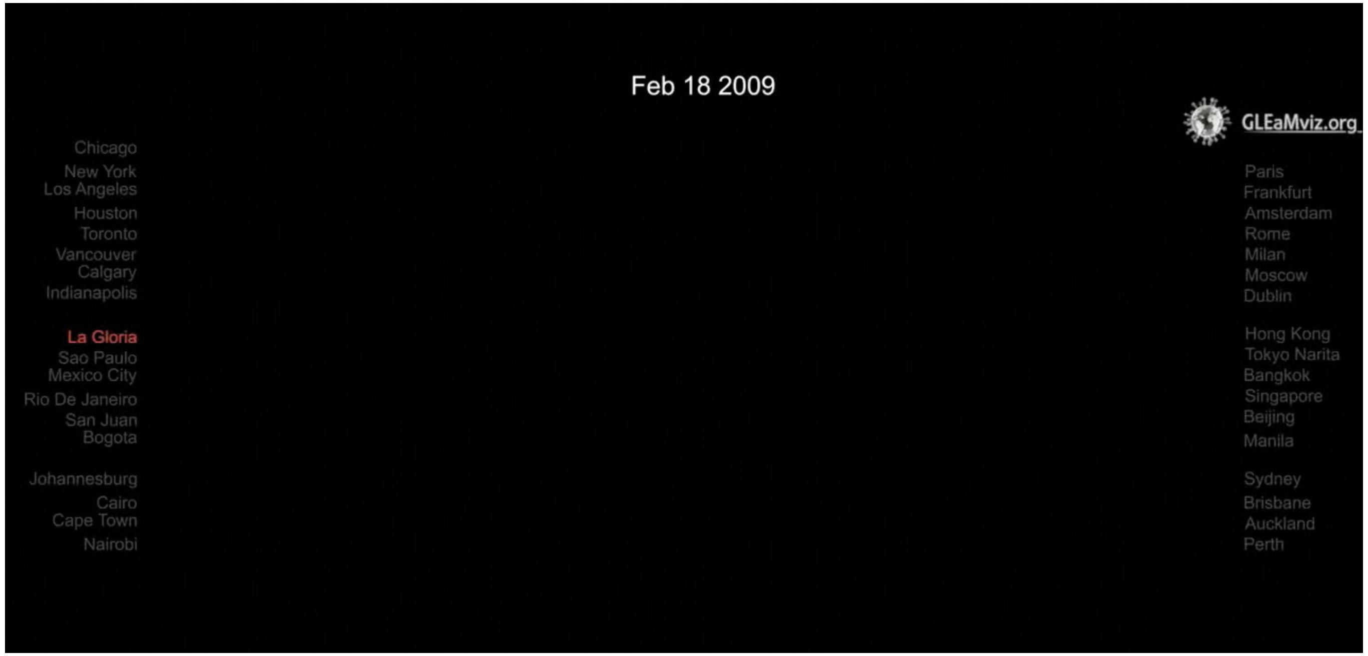


This diagram was designed during the Afghan war in 2012 to portray the American operational plans in Afghanistan

Complex Networks

Network Science

Epidemics: from Forecasting to Halting Deadly Viruses

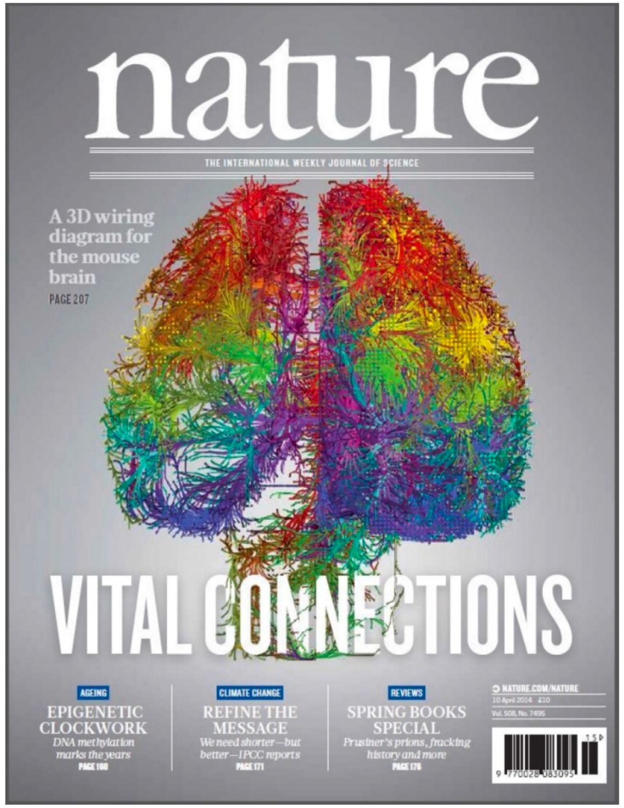


The predicted spread of the H1N1 epidemics during 2009, representing the first successful real-time prediction of a pandemic

Complex Networks

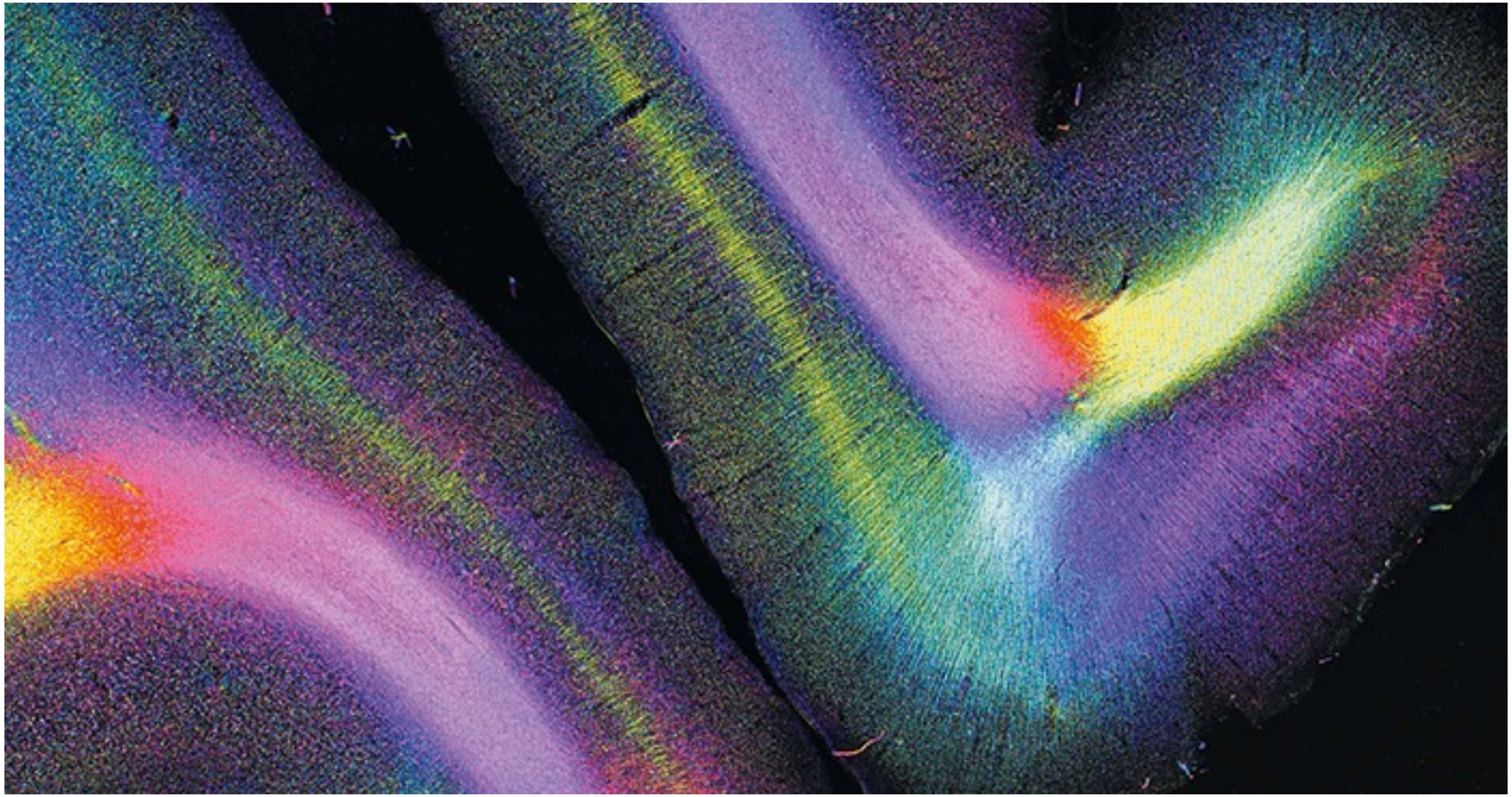
Network Science

Neuroscience: Mapping the Brain



Complex Networks

Network Science

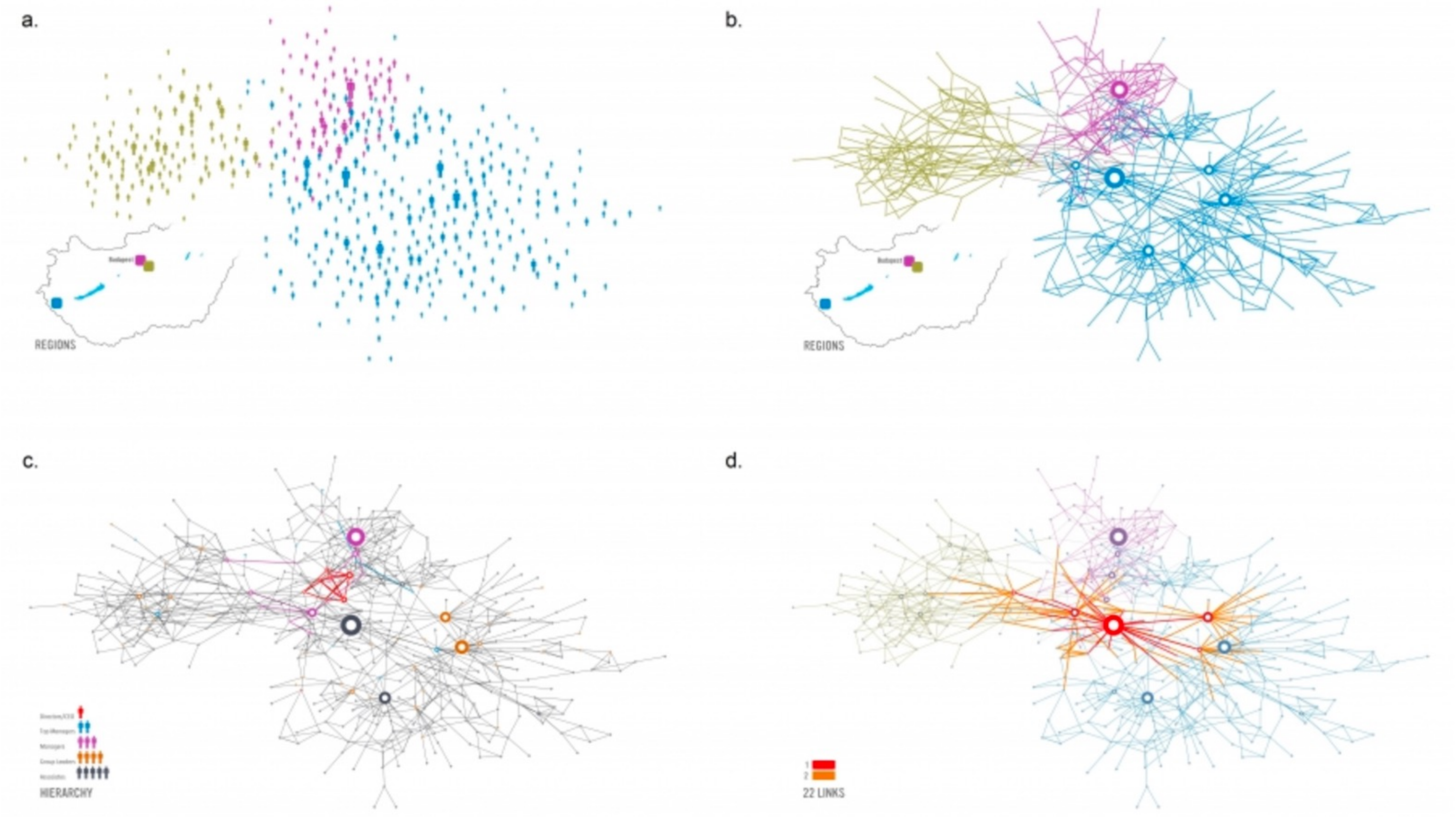


Human Brain Project

Complex Networks

Network Science

Management: Uncovering the Internal Structure of an Organization



Complex Networks

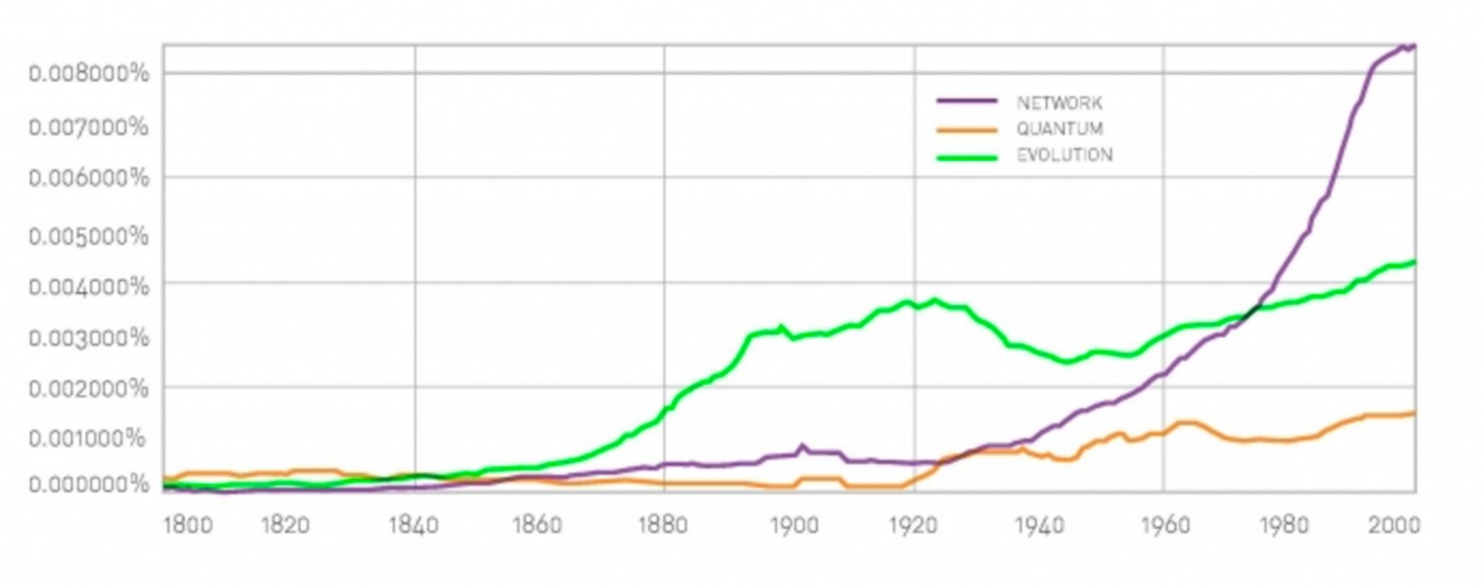
Network Science

Complex Networks

Network Science

The Rise of Networks:

The frequency of use of the words evolution, quantum, and network in books since 1880



Complex Networks

Network Science

“Network science is an enabling platform, offering novel tools and perspectives for a wide range of scientific problems, from social networking to drug design.”

A.-L. Barabási

Complex Networks

Network Science

"A key discovery of network science is that the architecture of networks emerging in various domains of science, nature, and technology are similar to each other,

a consequence of being governed by the same organizing principles.

Consequently, we can use a common set of mathematical tools to explore these systems."

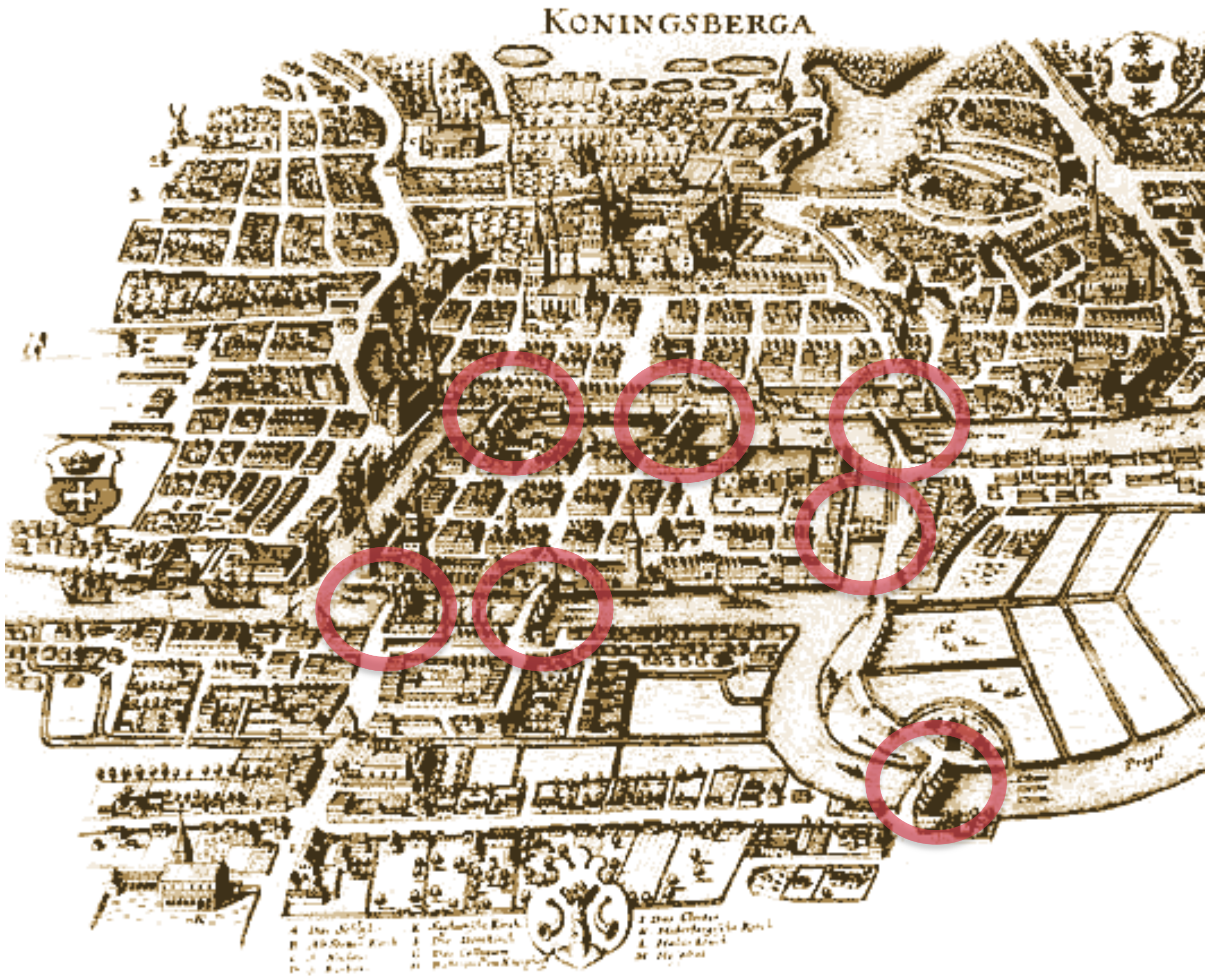
A.-L. Barabási

Complex Networks

The origin of graph theory: the Bridges of Königsberg

Complex Networks

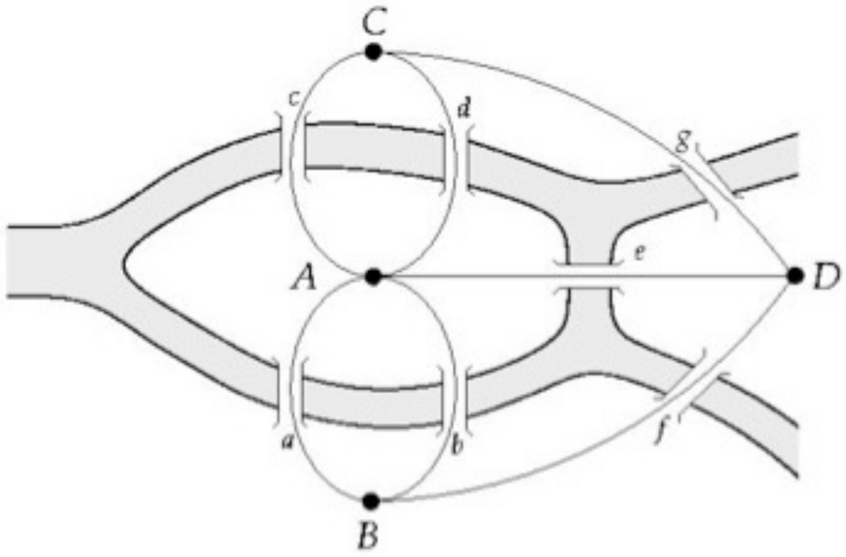
The origin of graph theory: the Bridges of Konigsberg



Can one walk across the seven bridges and never cross the same bridge twice?

Complex Networks

The origin of graph theory: the Bridges of Königsberg



Can one walk across the seven bridges and never cross the same bridge twice?

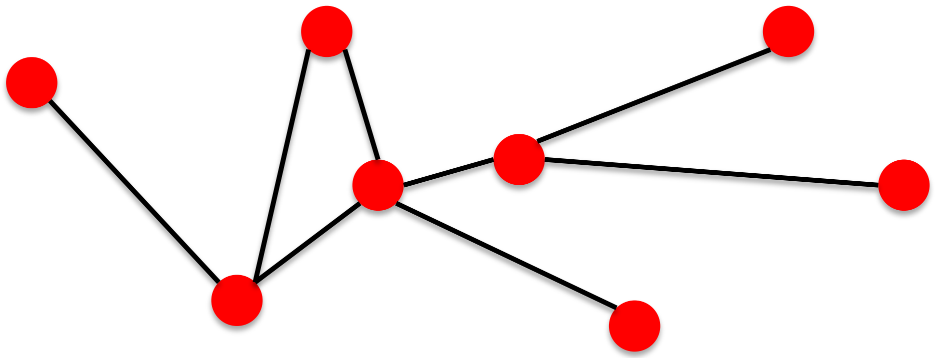
1735: **Euler's theorem:**

(a) If a graph has more than two nodes of odd degree, there is no path.

(b) If a graph is connected and has no odd degree nodes, it has at least one path.

Complex Networks

Networks and Graphs



- **components**: nodes, vertices N
- **interactions**: links, edges L
- **system**: network, graph (N,L)

Network Science: Graph Theory

Complex Networks

Networks and Graphs

Networks or Graphs?

In the scientific literature the terms *network* and *graph* are used interchangeably:

Network Science	Graph Teory
Network	Graph
Node	Vertex
Link	Edge

network often refers to real systems

graph: mathematical representation of a network

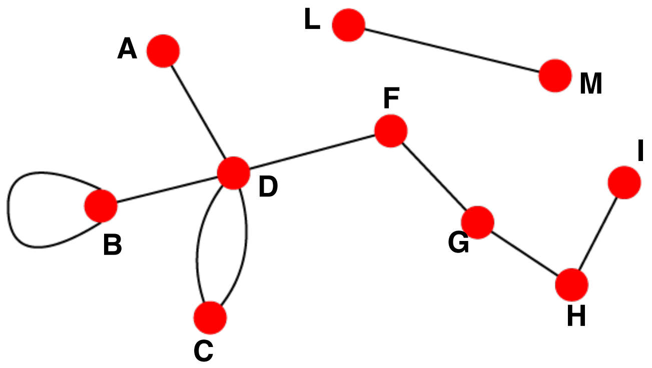
Complex Networks

Networks and Graphs

Undirected

Links: undirected (*symmetrical*)

Graph:

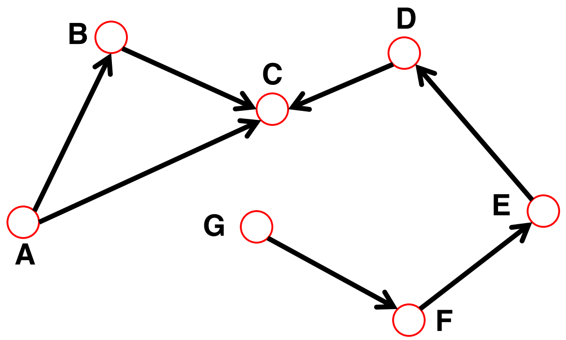


Undirected links :
coauthorship links
Actor network
protein interactions

Directed

Links: directed (*arcs*).

Digraph = directed graph:



An undirected link is the superposition of two opposite directed links.

Directed links :
URLs on the www
phone calls
metabolic reactions

Network Science: Graph Theory

Complex Networks

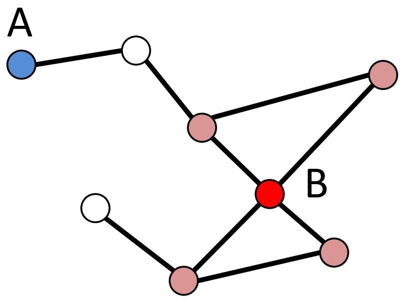
Networks and Graphs

NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L
Internet	Routers	Internet connections	Undirected	192,244	609,066
WWW	Webpages	Links	Directed	325,729	1,497,134
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594
Mobile Phone Calls	Subscribers	Calls	Directed	36,595	91,826
Email	Email addresses	Emails	Directed	57,194	103,731
Science Collaboration	Scientists	Co-authorship	Undirected	23,133	93,439
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908
Citation Network	Paper	Citations	Directed	449,673	4,689,479
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930

Complex Networks

Degree, Average Degree, and Degree Distribution

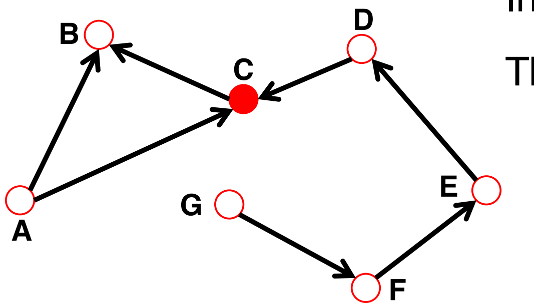
Undirected



Node degree: the number of links connected to the node.

$$k_A = 1 \qquad k_B = 4$$

Directed



In *directed networks* we can define an in-degree and out-degree.

The (total) degree is the sum of in- and out-degree.

$$k_C^{in} \square 2 \qquad k_C^{out} \square 1 \qquad k_C \square 3$$

Source: a node with $k^{in} = 0$; Sink: a node with $k^{out} = 0$.

Complex Networks

Degree, Average Degree, and Degree Distribution

BRIEF STATISTICS REVIEW

Four key quantities characterize a sample of N values $x_1, \ldots, x_N$:

Average (mean):

$$\langle x \rangle = \frac{x_1 + x_2 + \ldots + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

The n^th moment:

$$\langle x^n \rangle = \frac{x_1^n + x_2^n + \ldots + x_N^n}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i^n$$

Standard deviation:

$$\sigma_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \langle x \rangle\right)^2}$$

Distribution of x:

$$p_x = \frac{1}{N}\sum_i \delta_{x,x_i}$$

where p_x follows

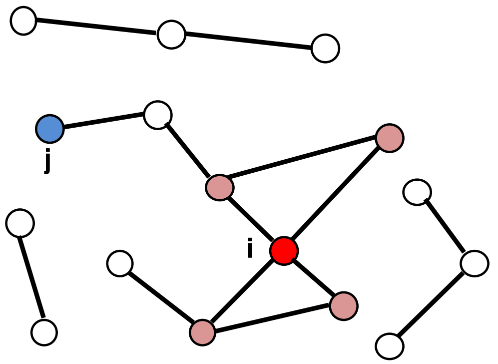
$$\sum_i p_x = 1 \; \left(\int p_x \, dx = 1 \right)$$

Network Science: Graph Theory

Complex Networks

Degree, Average Degree, and Degree Distribution

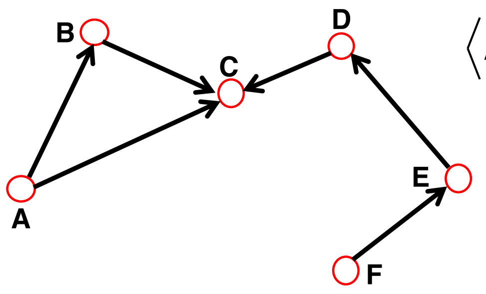
Undirected



$$\langle k \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} k_i \qquad \langle k \rangle \equiv \frac{2L}{N}$$

N – the number of nodes in the graph

Directed



$$\langle k^{in} \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} k_i^{in}, \quad \langle k^{out} \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} k_i^{out}, \quad \langle k^{in} \rangle = \langle k^{out} \rangle$$

$$\langle k \rangle \equiv \frac{L}{N}$$

Network Science: Graph Theory

Complex Networks

Degree, Average Degree, and Degree Distribution

NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L	⟨k⟩
Internet	Routers	Internet connections	Undirected	192,244	609,066	6.33
WWW	Webpages	Links	Directed	325,729	1,497,134	4.60
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594	2.67
Mobile Phone Calls	Subscribers	Calls	Directed	36,595	91,826	2.51
Email	Email addresses	Emails	Directed	57,194	103,731	1.81
Science Collaboration	Scientists	Co-authorship	Undirected	23,133	93,439	8.08
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908	83.71
Citation Network	Paper	Citations	Directed	449,673	4,689,479	10.43
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802	5.58
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930	2.90

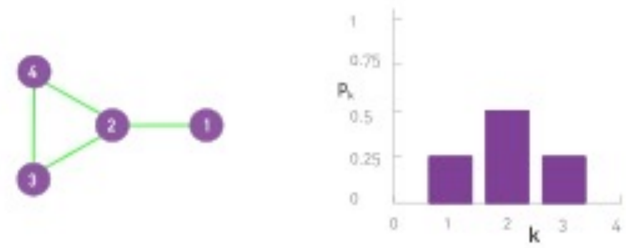
Network Science: Graph Theory

Complex Networks

Degree, Average Degree, and Degree Distribution

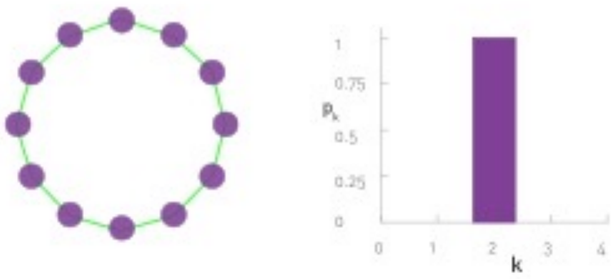
Degree distribution

P(k): probability that a randomly chosen node has degree *k*



N_k = # nodes with degree *k*

$P(k) = N_k / N$ ➔ plot



Complex Networks

Degree, Average Degree, and Degree Distribution

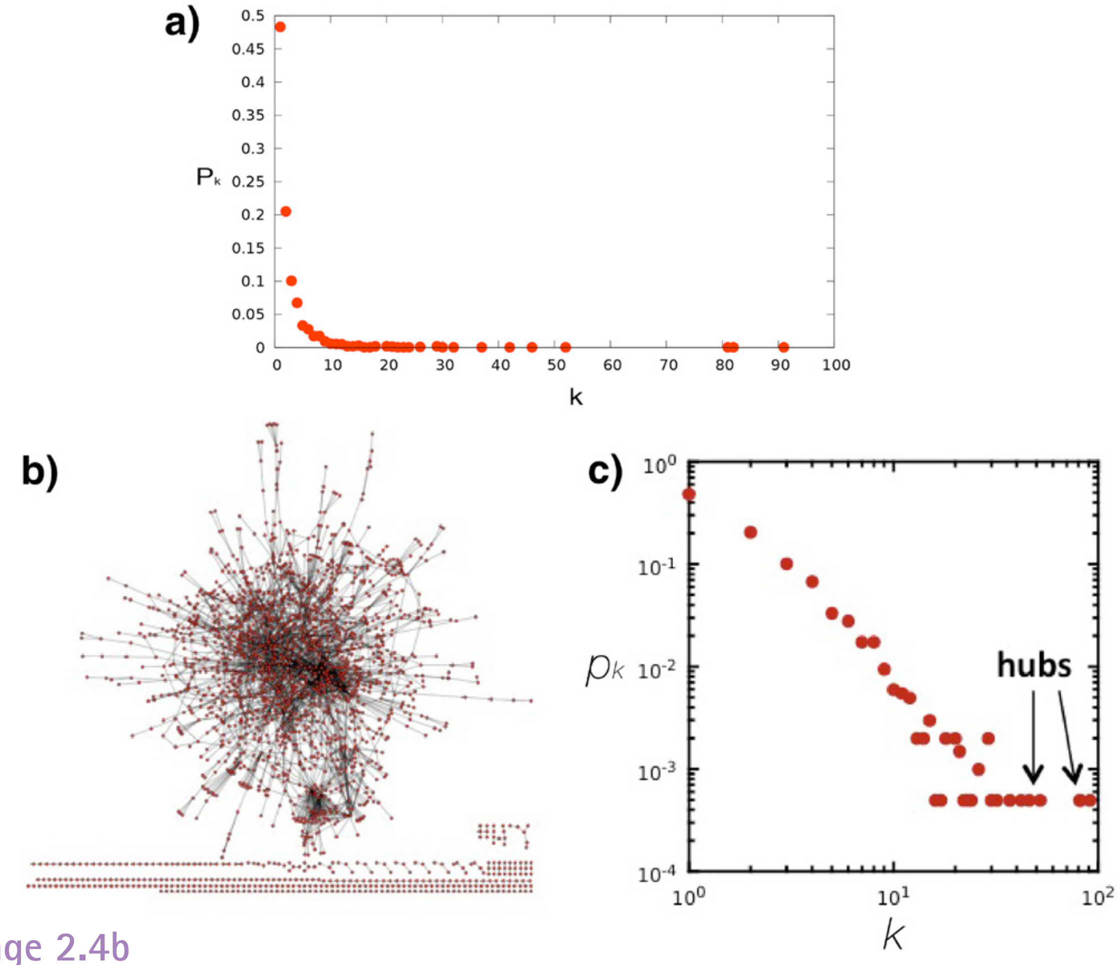


Image 2.4b

Complex Networks

Degree, Average Degree, and Degree Distribution

Discrete Representation: p_k is the probability that a node has degree k.

Continuum Description: $p(k)$ is the pdf of the degrees, where

$$\int_{k_1}^{k_2} p(k)dk$$

represents the probability that a node's degree is between k_1 and k_2.

Normalization condition:

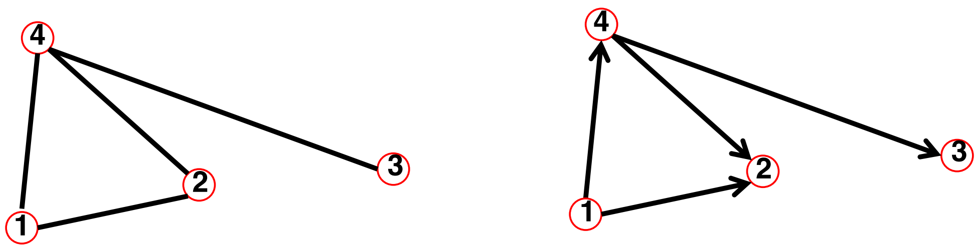
$$\sum_{0}^{\infty} p_k = 1 \qquad \int_{K_{\min}}^{\infty} p(k)dk = 1$$

where K_{min} is the minimal degree in the network.

Network Science: Graph Theory

Complex Networks

Adjacency Matrix Representation



$A_{ij}=1$ if there is a link between node i and j

$A_{ij}=0$ if nodes i and j are not connected to each other.

$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \qquad A_{ij} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

Note that for a directed graph (right) the matrix is not symmetric.

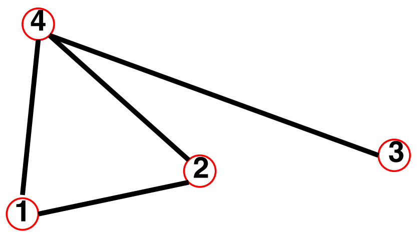
$A_{ij} = 1$ if there is a link pointing from node j and i

$A_{ij} = 0$ if there is no link pointing from j to i

Complex Networks

Adjacency Matrix Representation

Undirected



$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

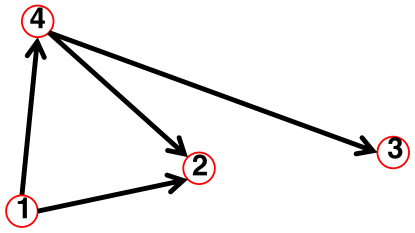
$$A_{ij} = A_{ji}$$
$$A_{ii} = 0$$

$$k_i = \sum_{j=1}^{N} A_{ij}$$

$$k_j = \sum_{i=1}^{N} A_{ij}$$

$$L = \frac{1}{2}\sum_{i=1}^{N} k_i = \frac{1}{2}\sum_{ij}^{N} A_{ij}$$

Directed



$$A_{ij} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

$$A_{ij} \neq A_{ji}$$
$$A_{ii} = 0$$

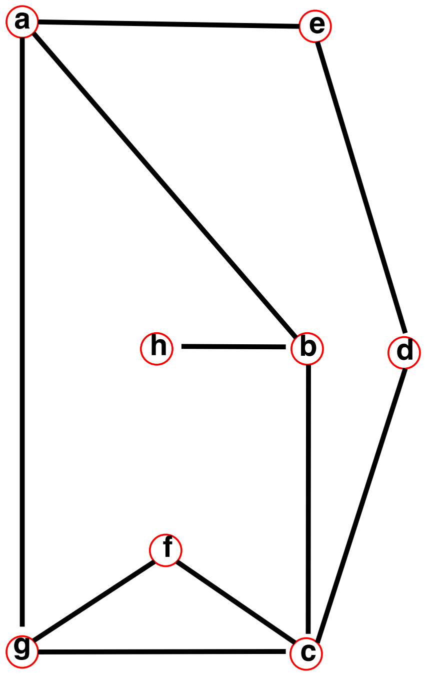
$$k_i^{in} = \sum_{j=1}^{N} A_{ij}$$

$$k_j^{out} = \sum_{i=1}^{N} A_{ij}$$

$$L = \sum_{i=1}^{N} k_i^{in} = \sum_{j=1}^{N} k_j^{out} = \sum_{i,j}^{N} A_{ij}$$

Complex Networks

Adjacency Matrix Representation

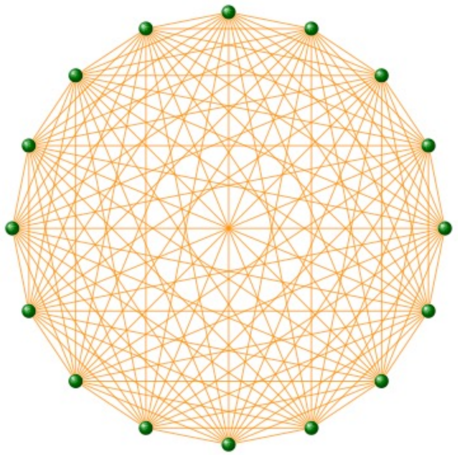


	a	b	c	d	e	f	g	h
a	0	1	0	0	1	0	1	0
b	1	0	1	0	0	0	0	1
c	0	1	0	1	0	1	1	0
d	0	0	1	0	1	0	0	0
e	1	0	0	1	0	0	0	0
f	0	0	1	0	0	0	1	0
g	1	0	1	0	0	0	0	0
h	0	1	0	0	0	0	0	0

Complex Networks

Real Networks are Sparse

The maximum number of links a network of N nodes can have is: $L_{\max} = \binom{N}{2} = \frac{N(N-1)}{2}$



A graph with degree $L = L_{max}$ is called a complete graph, and its average degree is $\langle k \rangle = N\text{-}1$

Complex Networks

Real Networks are Sparse

Most networks observed in real systems are sparse:

$$L \ll L_{max}$$

or

$$\langle k \rangle \ll N-1$$

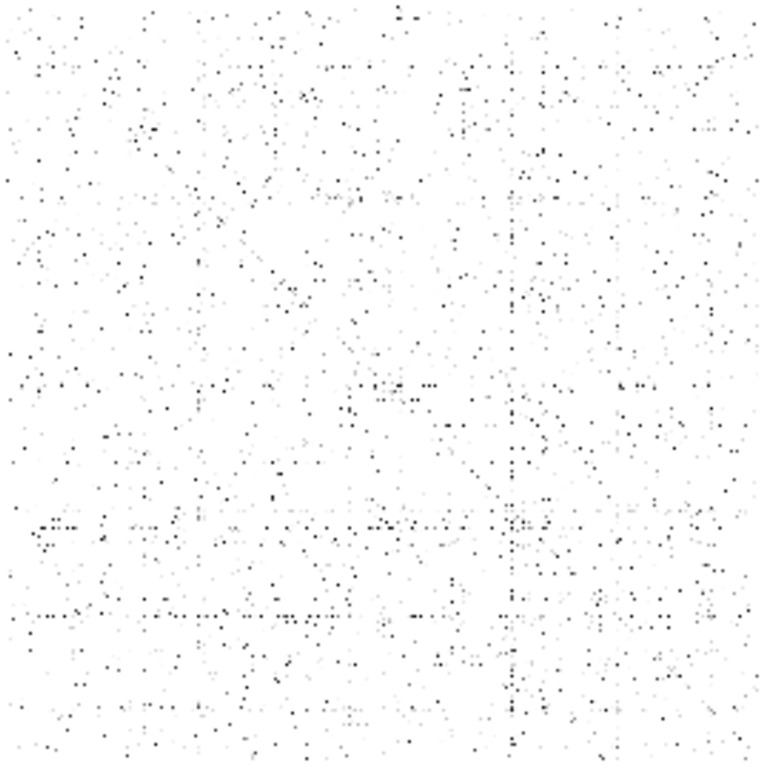
WWW (ND Sample):	N=325,729;	L=1.4 10^6	$L_{max}=10^{12}$	<k>=4.51
Protein (*S. Cerevisiae*):	N= 1,870;	L=4,470	$L_{max}=10^7$	<k>=2.39
Coauthorship (Math):	N= 70,975;	L=2 10^5	$L_{max}=3\ 10^{10}$	<k>=3.9
Movie Actors:	N=212,250;	L=6 10^6	$L_{max}=1.8\ 10^{13}$	<k>=28.78

(Source: Albert, Barabasi, RMP2002)

Network Science: Graph Theory

Complex Networks

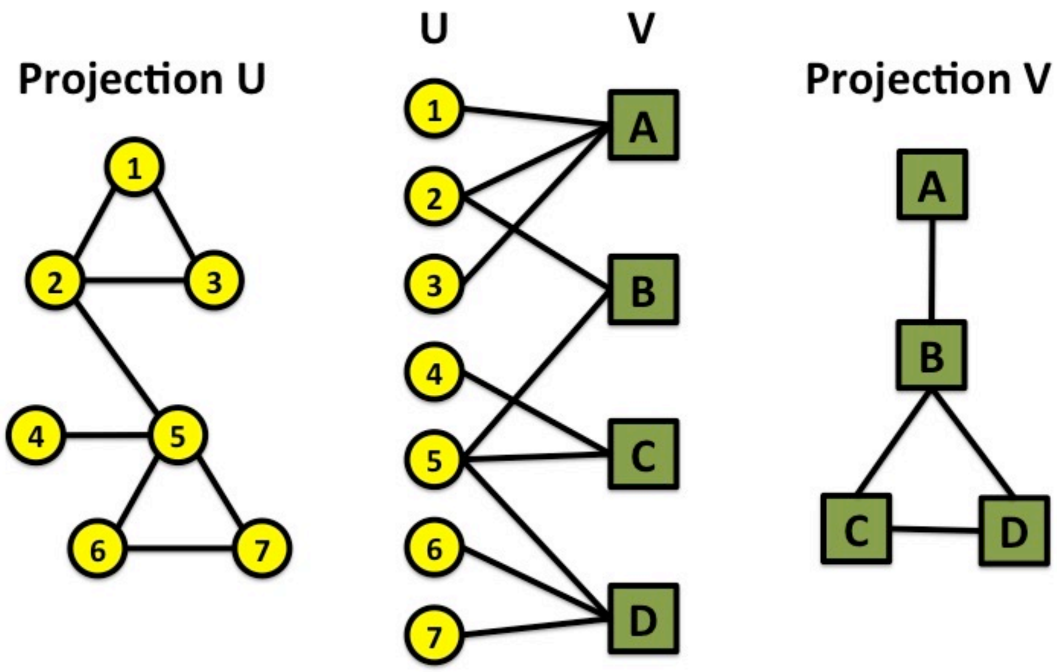
Real Networks are Sparse



Complex Networks

Bipartite Networks

A bipartite graph (or bigraph) is a graph whose nodes can be divided into two disjoint sets U and V such that every link connects a node in U to one in V; that is, U and V are independent sets.

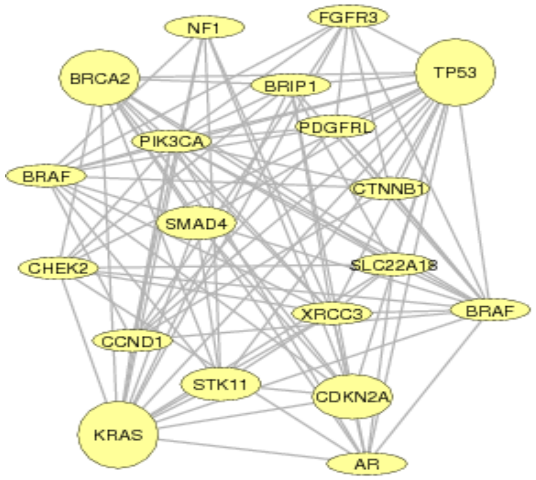


Examples:

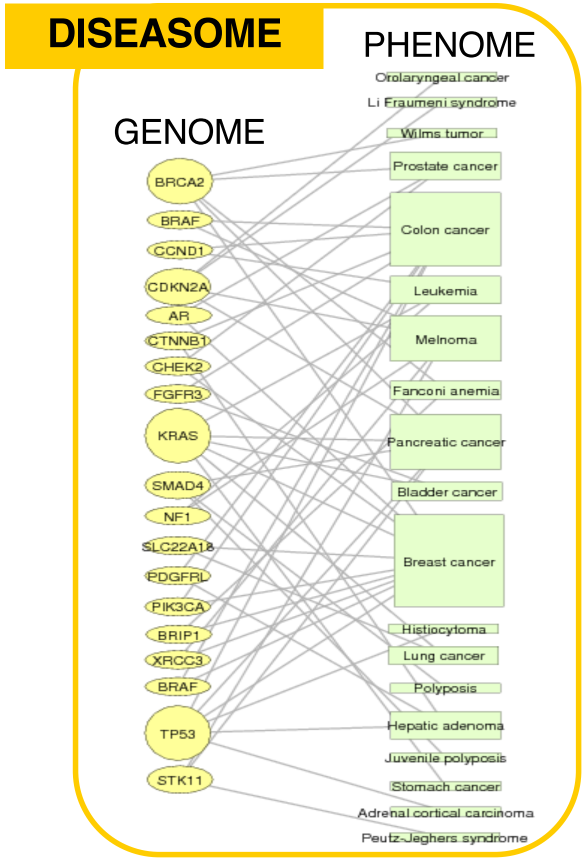
Hollywood actor network
Collaboration networks
Disease network (diseasome)

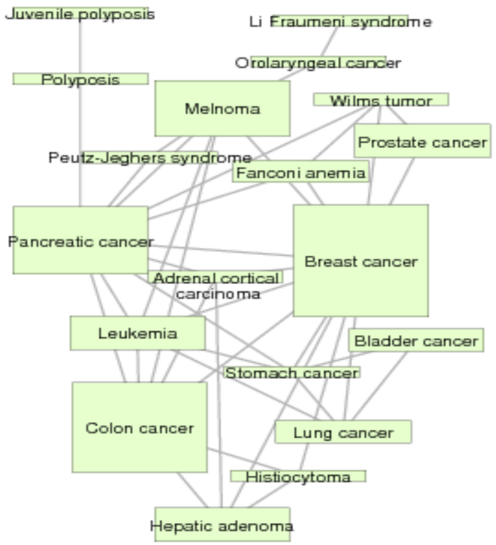
Complex Networks

Bipartite Networks



Gene network





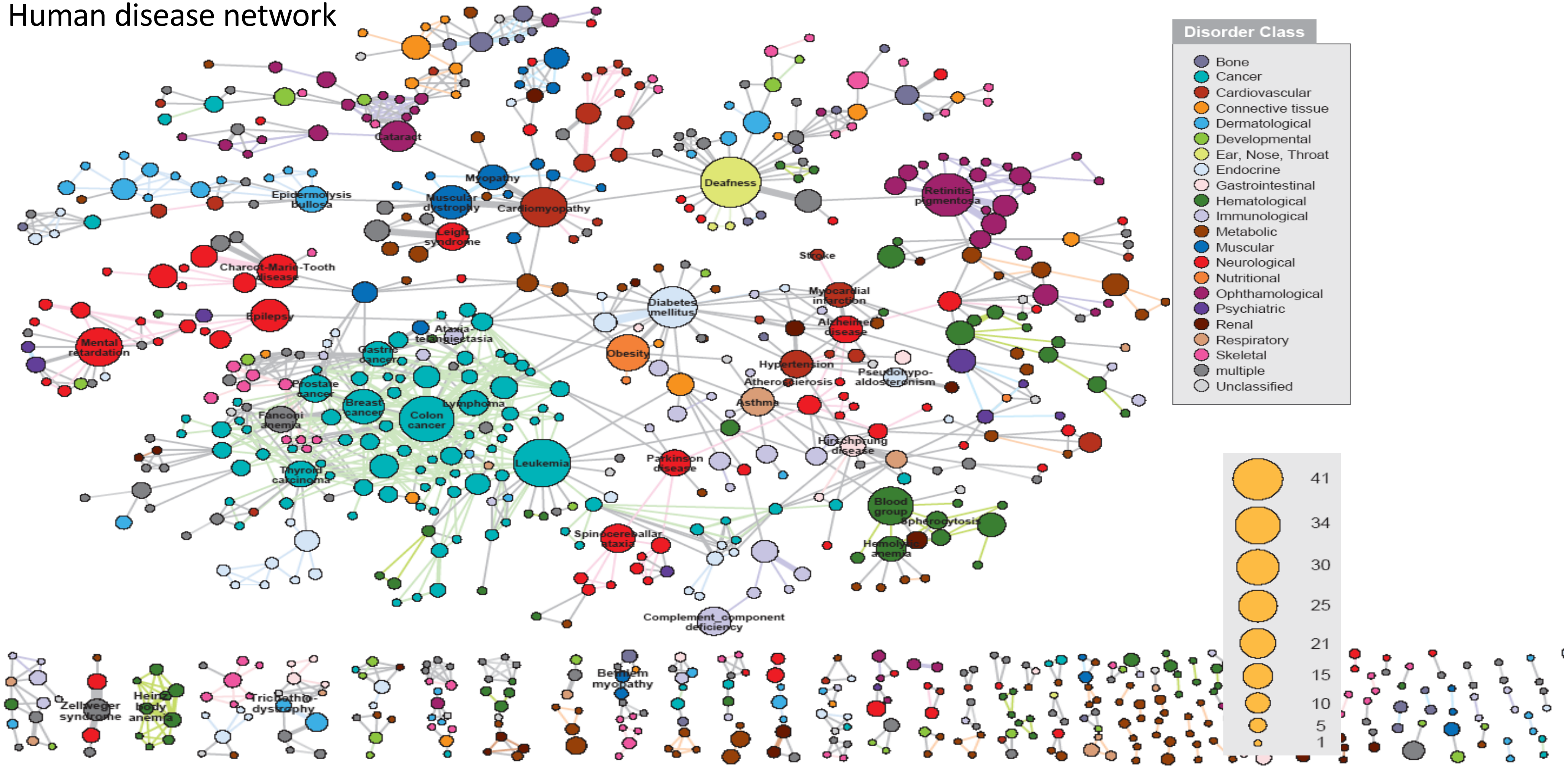
Disease network

Goh, Cusick, Valle, Childs, Vidal & Barabási, PNAS (2007)

Complex Networks

Bipartite Networks

Human disease network



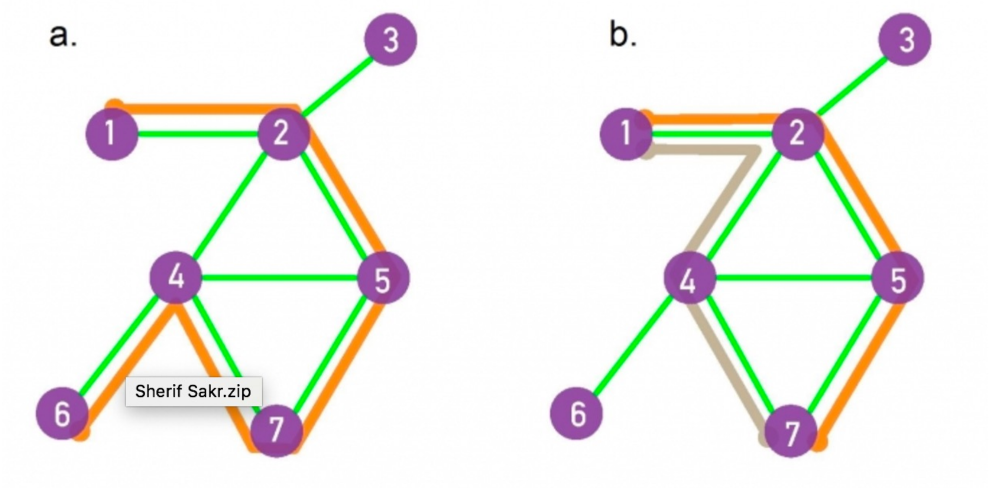
Complex Networks

Paths

A *path is* a sequence of nodes in which each node is adjacent to the next one

$P_{i0,in}$ of length n between nodes i_0 and i_n is an ordered collection of $n+1$ nodes and n links

$$P_n = \{i_0, i_1, i_2, \ldots, i_n\} \qquad P_n = \{(i_0, i_1), (i_1, i_2), (i_2, i_3), \ldots, (i_{n-1}, i_n)\}$$



The path shown in orange in (a) follows the route 1→2→5→7→4→6, hence its length is $n = 5$.

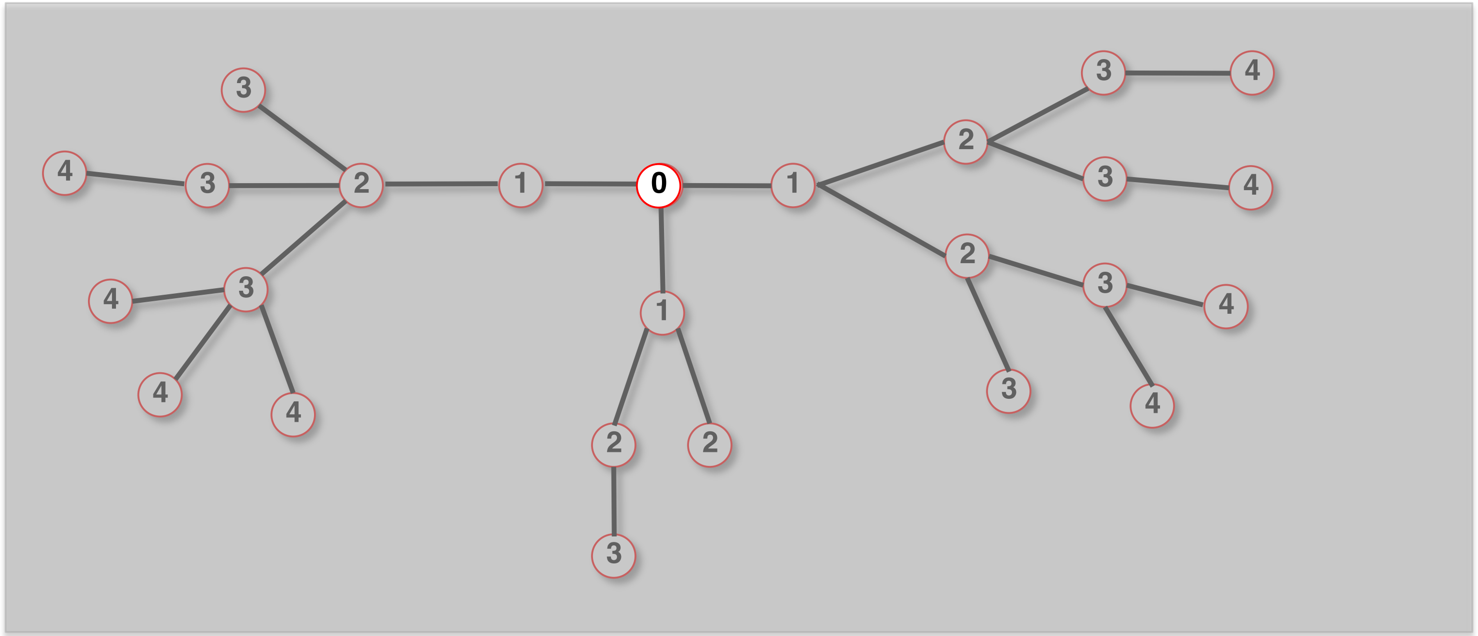
The network diameter is the largest distance in the network, being $d_{max} = 3$ here.

Complex Networks

Paths – Breadth-First Search

Distance between node 0 and node 4:

1.Start at 0.



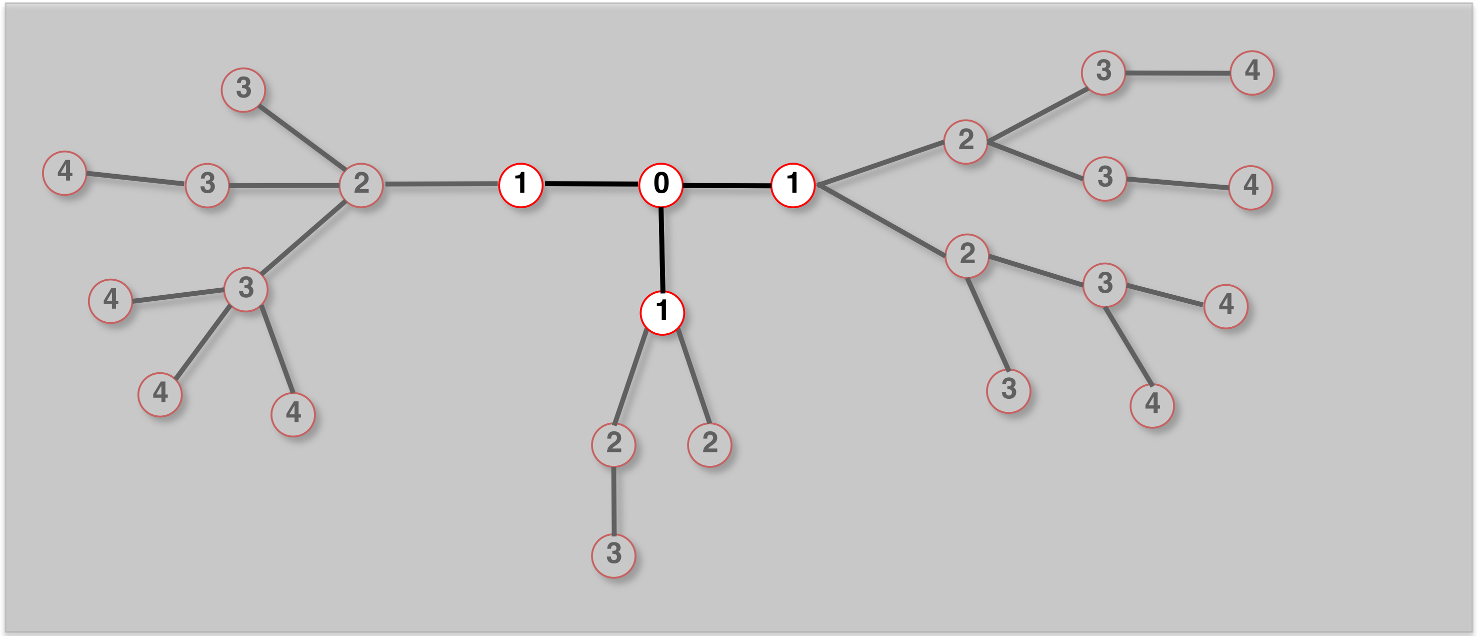
Network Science: Graph Theory

Complex Networks

Paths – Breadth-First Search

Distance between node 0 and node 4:

1. Start at 0.
2. Find the nodes adjacent to 1. Mark them as at distance 1. Put them in a queue.



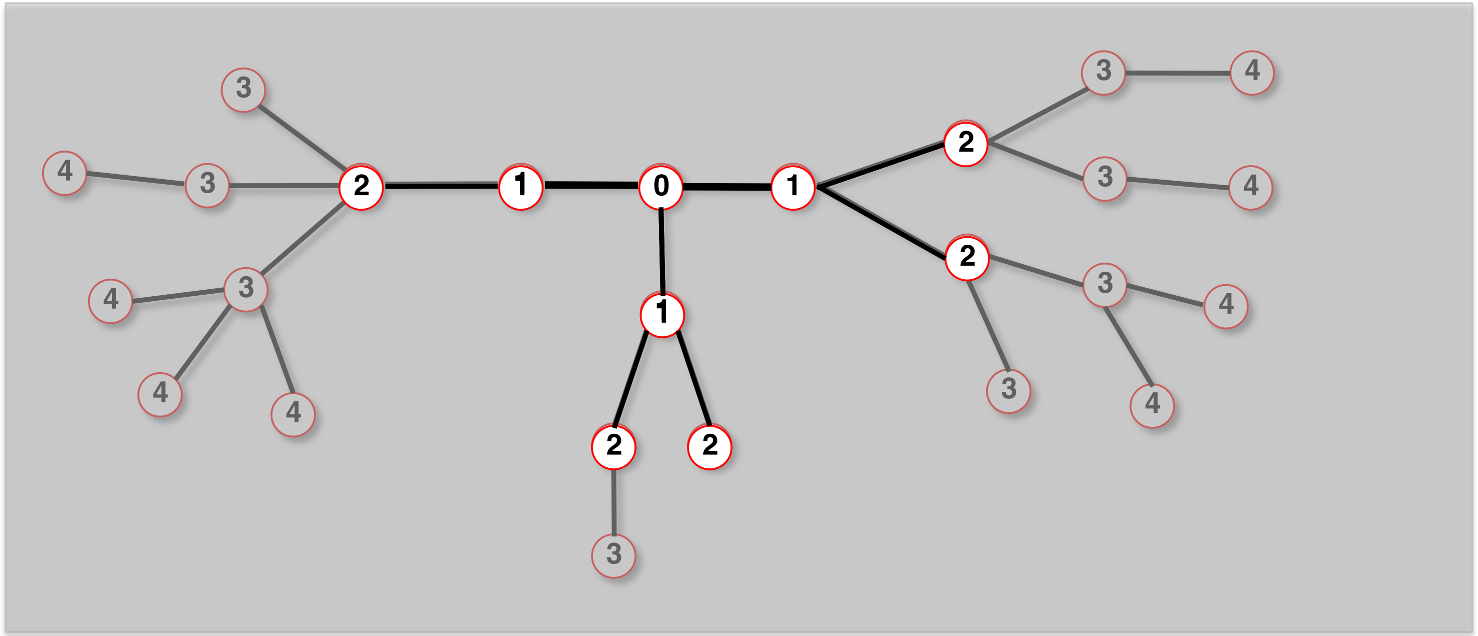
Network Science: Graph Theory

Complex Networks

Paths – Breadth-First Search

Distance between node 0 and node 4:

1. Start at 0.
2. Find the nodes adjacent to 0. Mark them as at distance 1. Put them in a queue.
3. Take the first node out of the queue. Find the unmarked nodes adjacent to it in the graph. Mark them with the label of 2. Put them in the queue.



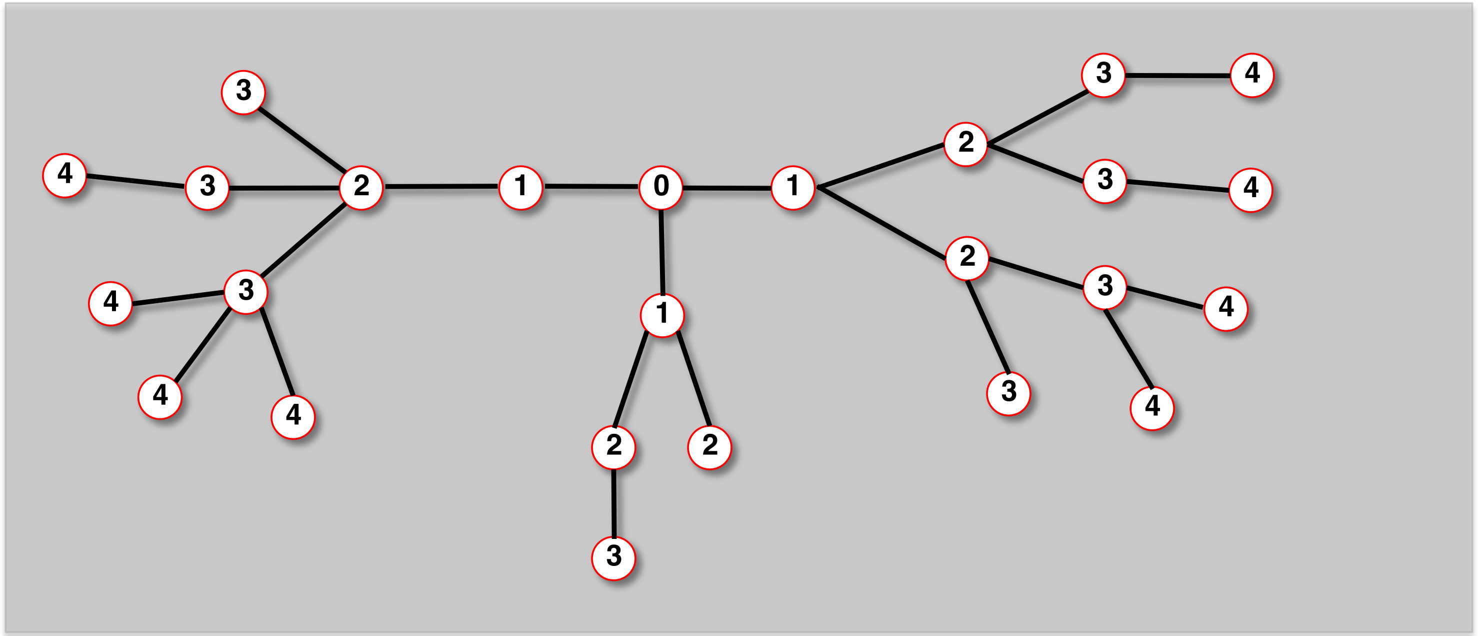
Network Science: Graph Theory

Complex Networks

Paths – Breadth-First Search

Distance between node 0 and node 4:

1. Repeat until you find node 4 or there are no more nodes in the queue.
2. The distance between 0 and 4 is the label of 4 or, if 4 does not have a label, infinity.



Network Science: Graph Theory

Complex Networks

Paths

Diameter: d_{max} the maximum (shortest) distance between any pair of nodes in the graph

Average path length/distance, $\langle d \rangle$ *,* for a directed graph:

$$\langle d \rangle \equiv \frac{1}{2L_{\max}} \sum_{i,j \neq i} d_{ij}$$

where d_{ij} is the distance from node i to node j

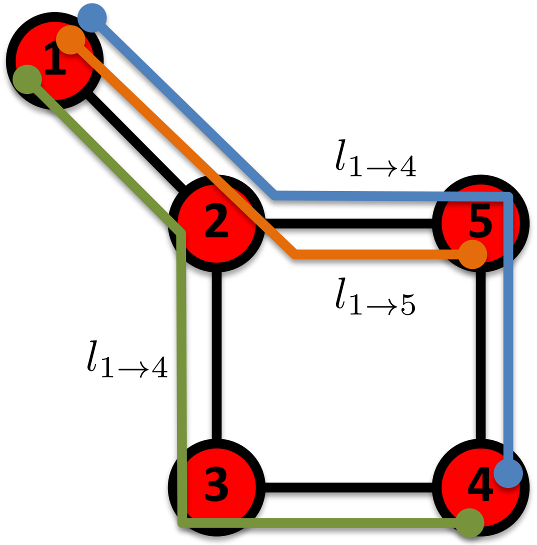
In an *undirected graph* $d_{ij} = d_{ji}$*, so* we only need to count them once:

$$\langle d \rangle \equiv \frac{1}{L_{\max}} \sum_{i,j>i} d_{ij}$$

Complex Networks

Paths

Shortest Path



$l_{1\to4} = 3$

$l_{1\to5} = 2$

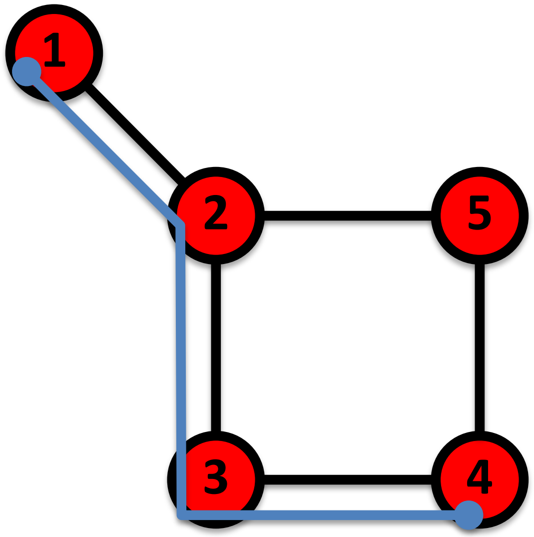
The path with the shortest length between two nodes (distance)

Network Science: Graph Theory

Complex Networks

Paths

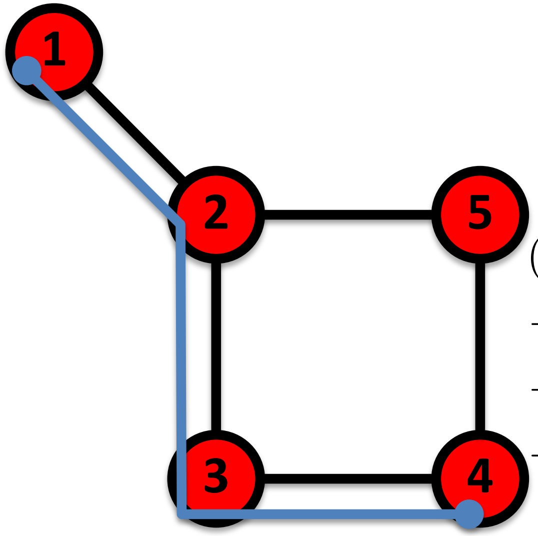
Diameter



$l_{1\to 4} = 3$

The longest shortest path in a graph

Average Path Length



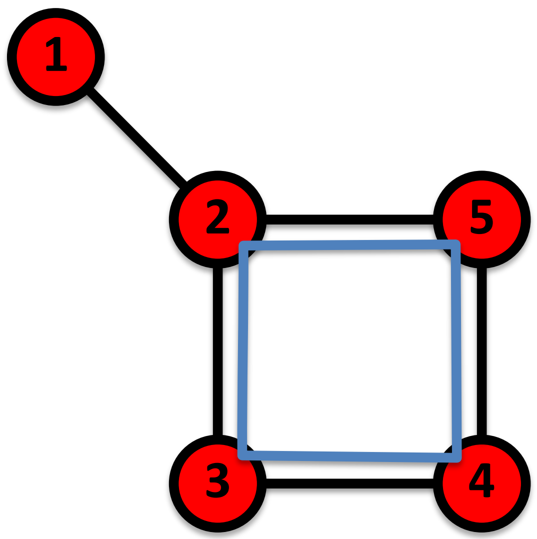
$$(l_{1\to 2} + l_{1\to 3} + l_{1\to 4} + \\ + l_{1\to 5} + l_{2\to 3} + l_{2\to 4} + \\ + l_{2\to 5} + l_{3\to 4} + l_{3\to 5} + \\ + l_{4\to 5})/10 = 1.6$$

The average of the shortest paths for all pairs of nodes.

Complex Networks

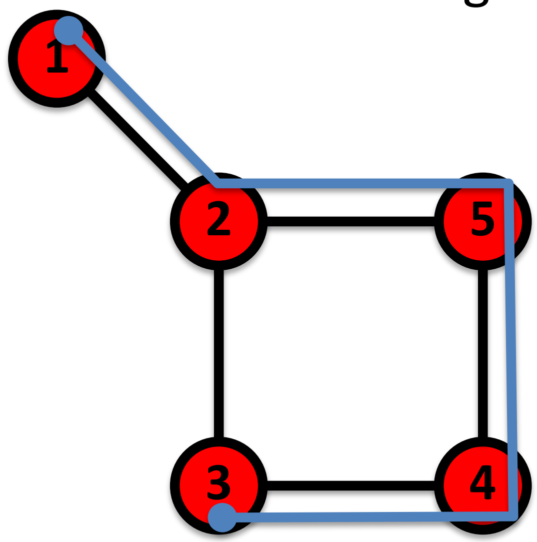
Paths

Cycle



A path with the same start and end node.

Self-avoiding Path

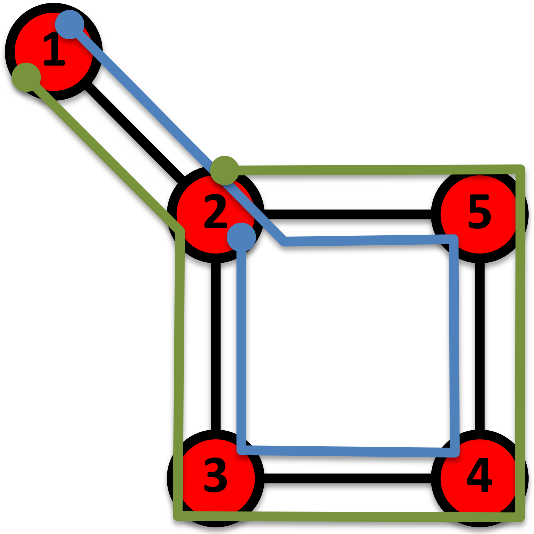


A path that does not intersect itself.

Complex Networks

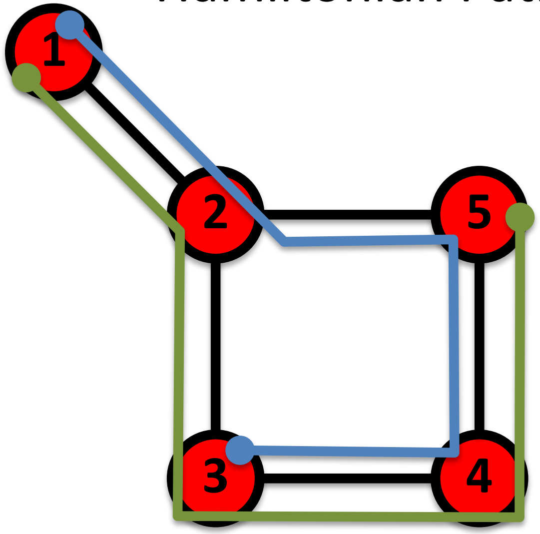
Paths

Eulerian Path



A path that traverses each
link exactly once

Hamiltonian Path

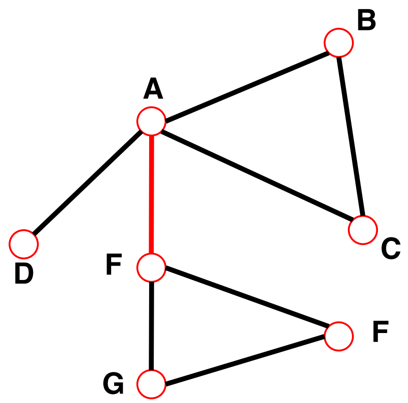


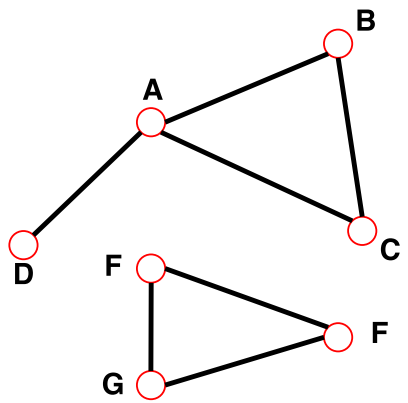
A path that visits each
node exactly once

Complex Networks

Connectivity & Components: Undirected Graphs

Connected (undirected) graph: any two vertices can be joined by a path.
A disconnected graph is made up by two or more connected components.





Largest Component:
Giant Component

The rest: **Isolates**

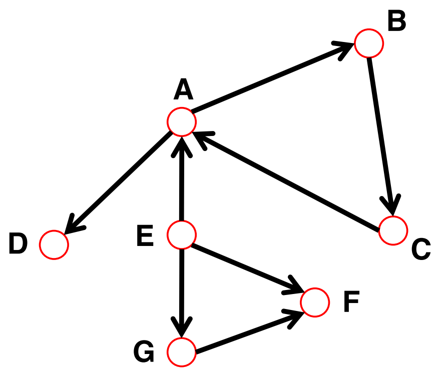
Bridge: if we erase it, the graph becomes disconnected.

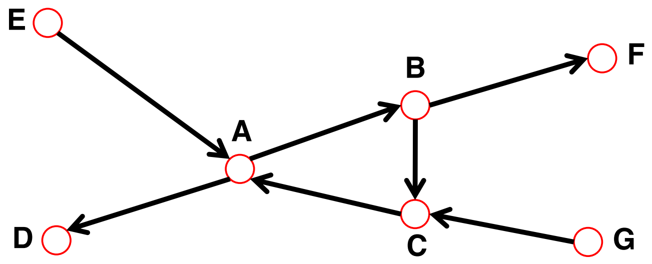
Complex Networks

Connectivity & Components: Directed Graphs

Strongly connected directed graph: has a path from each node to every other node and vice versa (e.g. AB path and BA path).
Weakly connected directed graph: it is connected if we disregard the edge directions.

Strongly connected components can be identified, but not every node is part of a nontrivial strongly connected component.





In-component: nodes that can reach the scc,
Out-component: nodes that can be reached from the scc.

Complex Networks

Connectivity & Components: Directed Graphs

Finding the Connected Components of a Network

- Start from a randomly chosen node i and perform a BFS

Label all nodes reached this way with $n = 1$.

- If the total number of labeled nodes equals N, then the network is

connected. If the number of labeled nodes is smaller than N, the network consists of several components. To identify them, proceed to step 3.

- Increase the label $n \to n + 1$. Choose an unmarked node j, label it

with n. Use BFS to find all nodes reachable from j, label them all with n. Return to step 2.

Complex Networks

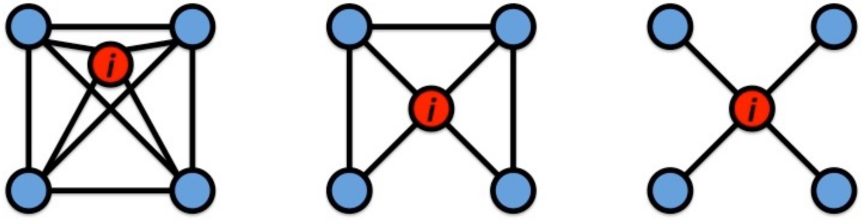
Clustering Coefficient

Local clustering coefficient: what fraction of your neighbors are connected?

$$C_i = \frac{2L_i}{k_i(k_i-1)}$$

L_i represents the number of links between the k_i neighbors of node i

C_i measures the network's local link density: the more densely interconnected the neighborhood of node i, the higher is its local clustering coefficient. C_i in [0,1]



$C_i = 1$ $C_i = 1/2$ $C_i = 0$

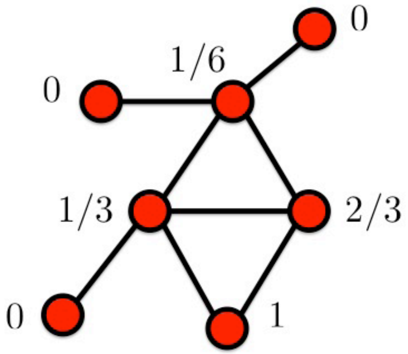
Complex Networks

Clustering Coefficient

The degree of clustering of a whole network is captured by the
average clustering coefficient:

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^{N} C_i$$

$\langle C \rangle$ is the probability that two neighbors of a randomly selected node link to each other.



$$\langle C \rangle = \frac{13}{42} \approx 0.310$$

Complex Networks

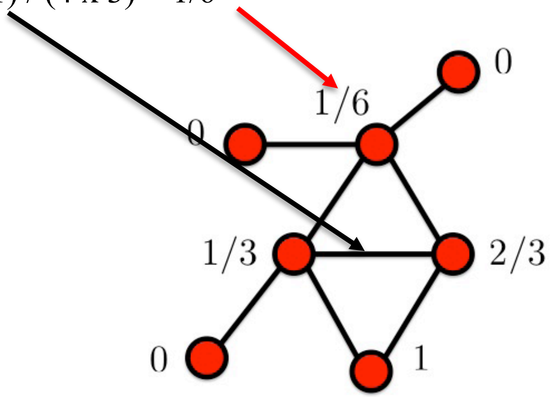
Clustering Coefficient

The degree of clustering of a whole network is captured by the
average clustering coefficient:

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^{N} C_i$$

$\langle C \rangle$ is the probability that two neighbors of a randomly selected node link to each other.

$$C_i = \frac{2L_i}{k_i(k_i-1)}$$ ⇒ (2 x 1) / (4 x 3) = 1/6



$$\langle C \rangle = \frac{13}{42} \approx 0.310$$

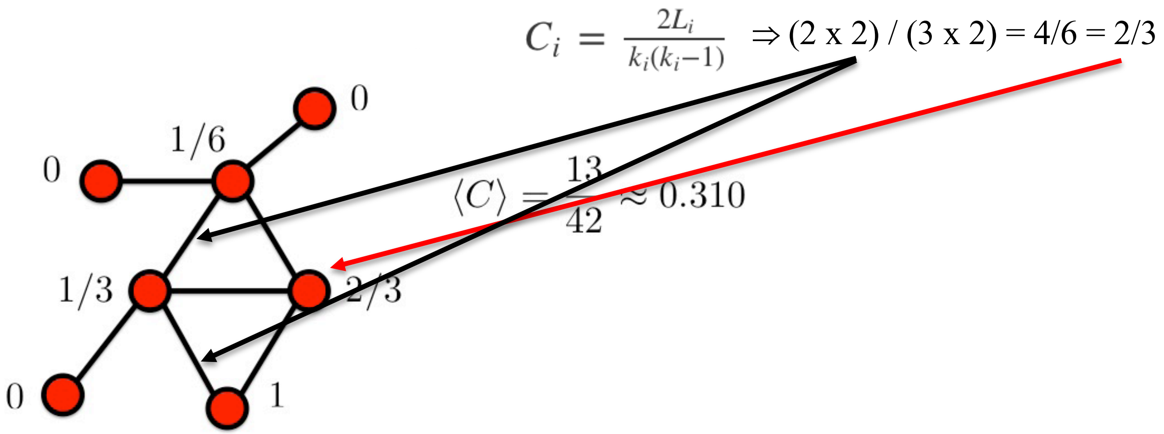
Complex Networks

Clustering Coefficient

The degree of clustering of a whole network is captured by the average clustering coefficient:

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^{N} C_i$$

$\langle C \rangle$ is the probability that two neighbors of a randomly selected node link to each other.



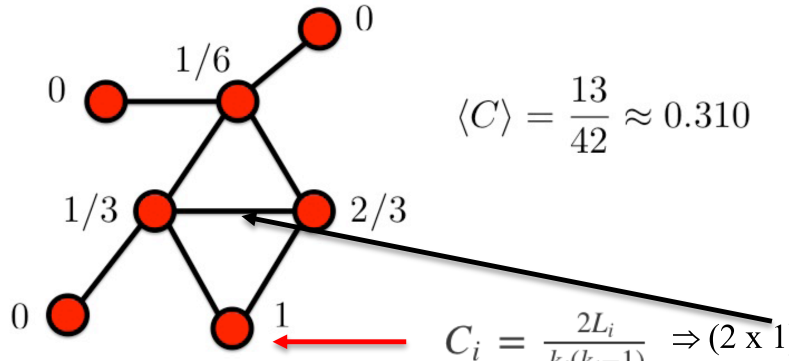
Complex Networks

Clustering Coefficient

The degree of clustering of a whole network is captured by the
average clustering coefficient:

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^{N} C_i$$

$\langle C \rangle$ is the probability that two neighbors of a randomly selected node link to each other.



$$\langle C \rangle = \frac{13}{42} \approx 0.310$$

$$C_i = \frac{2L_i}{k_i(k_i - 1)} \Rightarrow (2 \text{ x } 1) / (2 \text{ x } 1) = 1$$

Complex Networks

Clustering Coefficient

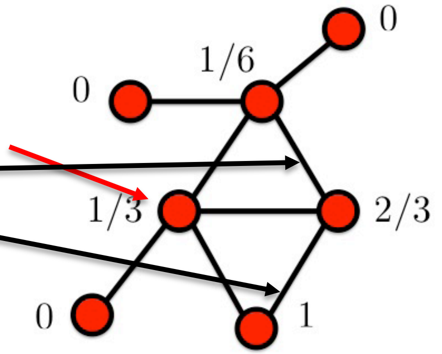
The degree of clustering of a whole network is captured by the average clustering coefficient:

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^{N} C_i$$

$\langle C \rangle$ is the probability that two neighbors of a randomly selected node link to each other.

$$C_i = \frac{2L_i}{k_i(k_i-1)}$$

⇒ (2 x 2) / (4 x 3) = 4/12 = 1/3



$$\langle C \rangle = \frac{13}{42} \approx 0.310$$

Complex Networks

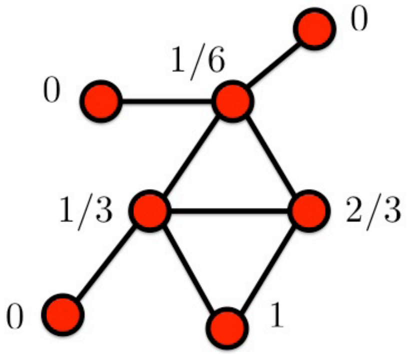
Clustering Coefficient

The degree of clustering of a whole network is captured by the average clustering coefficient:

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^{N} C_i$$

⇒ (1/7) x ((1/6) + (1/3) + (2/3) + (1/1))
= (1/7) x ((1/6) + (2/6) + (4/6) + (6/6))
= (13 / 42)

$\langle C \rangle$ is the probability that two neighbors of a randomly selected node link to each other.



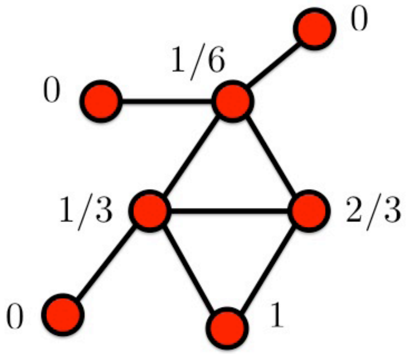
$$\langle C \rangle = \frac{13}{42} \approx 0.310$$

Complex Networks

Clustering Coefficient

The degree of global clustering of a whole network is captured by the global clustering coefficient:

$$C_\Delta = \frac{3 \times NumberOfTriangles}{NumberOfConnectedTriples}$$



$$\langle C \rangle = \frac{13}{42} \approx 0.310$$

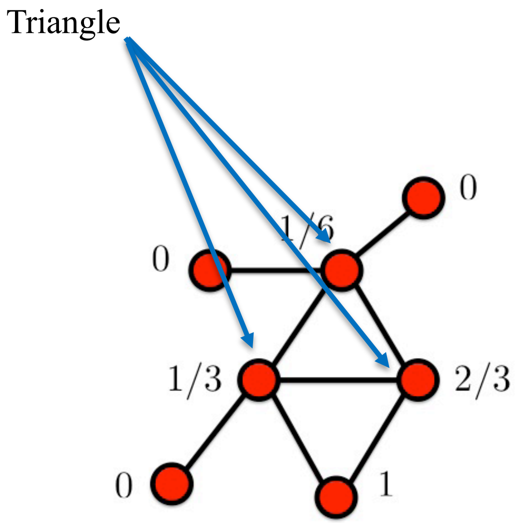
$$C_\Delta = \frac{3}{8} = 0.375$$

Complex Networks

Clustering Coefficient

The degree of global clustering of a whole network is captured by the global clustering coefficient:

$$C_\Delta = \frac{3 \times NumberOfTriangles}{NumberOfConnectedTriples}$$



$$\langle C \rangle = \frac{13}{42} \approx 0.310$$

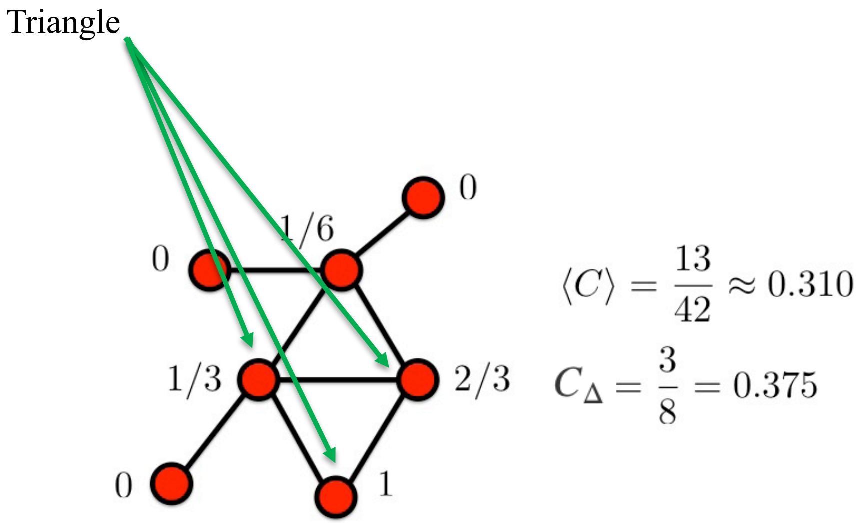
$$C_\Delta = \frac{3}{8} = 0.375$$

Complex Networks

Clustering Coefficient

The degree of global clustering of a whole network is captured by the global clustering coefficient:

$$C_\Delta = \frac{3 \times NumberOfTriangles}{NumberOfConnectedTriples}$$

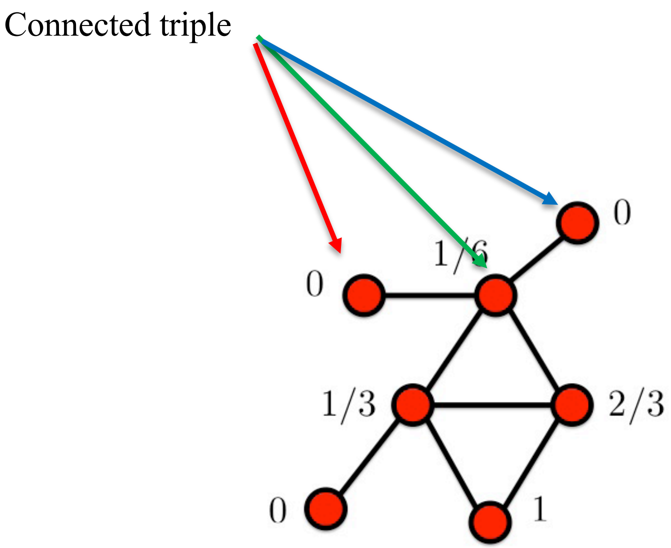


Complex Networks

Clustering Coefficient

The degree of global clustering of a whole network is captured by the global clustering coefficient:

$$C_\Delta = \frac{3 \times NumberOfTriangles}{NumberOfConnectedTriples}$$



$$\langle C \rangle = \frac{13}{42} \approx 0.310$$

$$C_\Delta = \frac{3}{8} = 0.375$$

Complex Networks

Clustering Coefficient

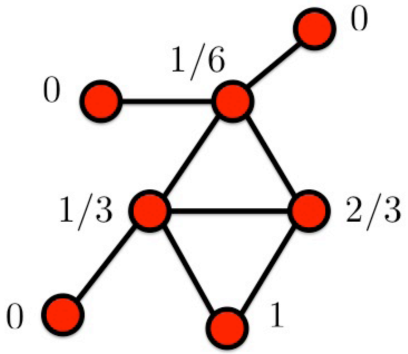
The degree of global clustering of a whole network is captured by the global clustering coefficient:

$$C_\Delta = \frac{3 \times NumberOfTriangles}{NumberOfConnectedTriples}$$

(3 x 2) / (10 + 6) = (6/16) = 3 / 8

open triples

closed triples



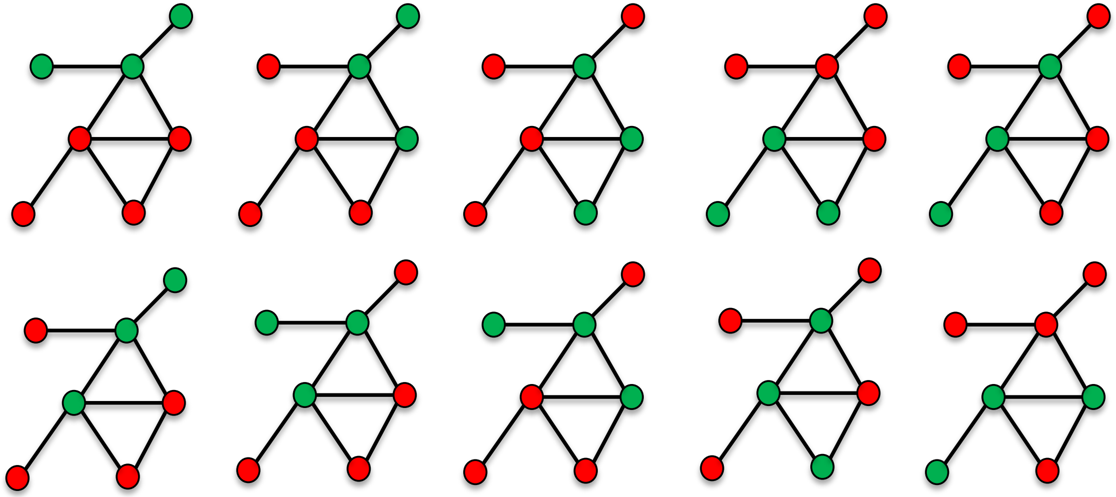
$$\langle C \rangle = \frac{13}{42} \approx 0.310$$

$$C_\Delta = \frac{3}{8} = 0.375$$

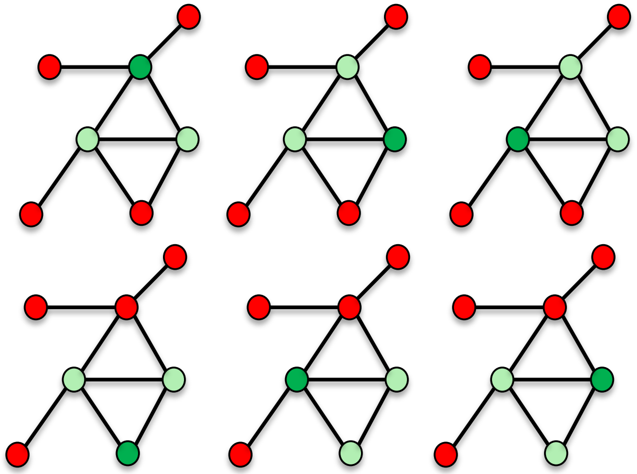
Complex Networks

Clustering Coefficient

10 open triples

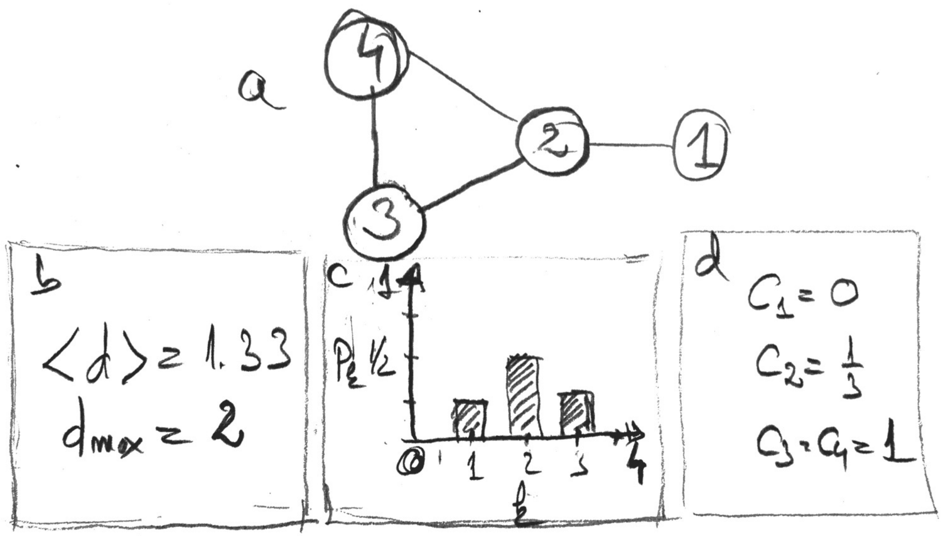


6 closed triples



Complex Networks

Three Central Quantities in Network Science



Average path length: $\langle d \rangle$

Degree distribution: $p(k)$ p_k

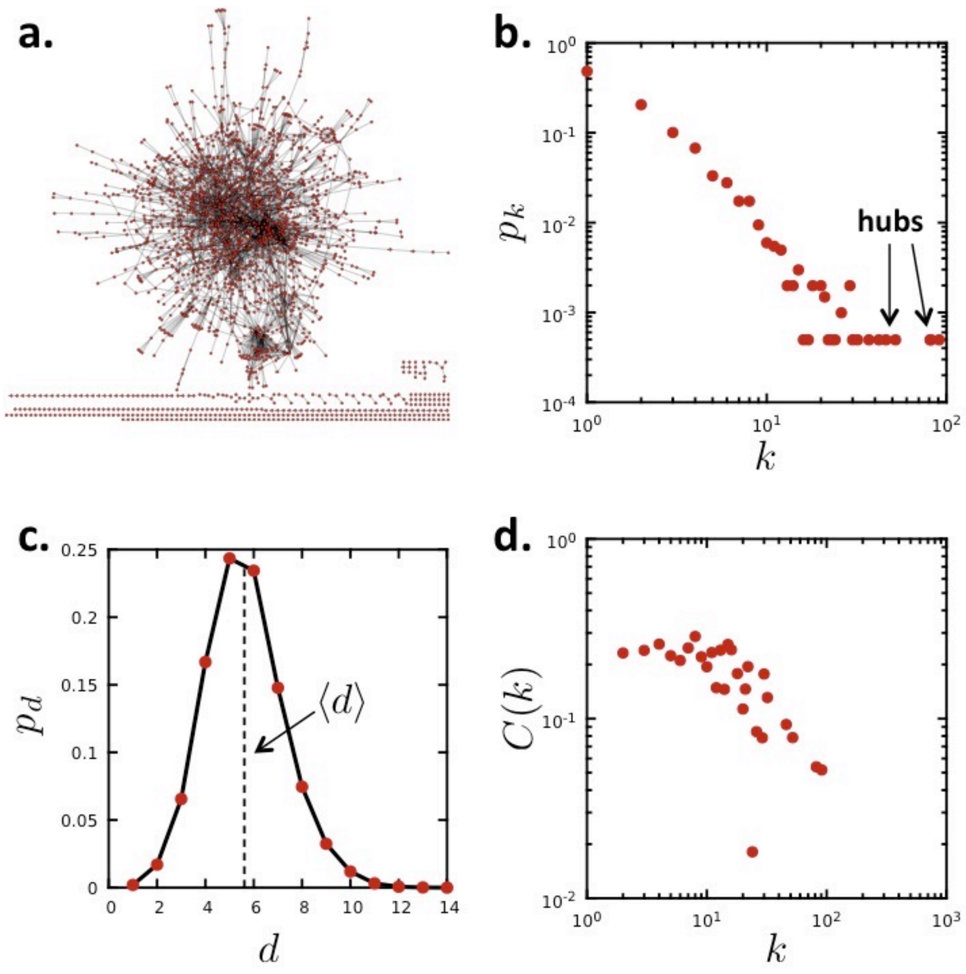
Clustering coefficient: $C_i = \frac{2L_i}{k_i(k_i-1)}$

Complex Networks

Case Study: Protein-Protein Interaction Network

Complex Networks

Case Study: Protein-Protein Interaction Network

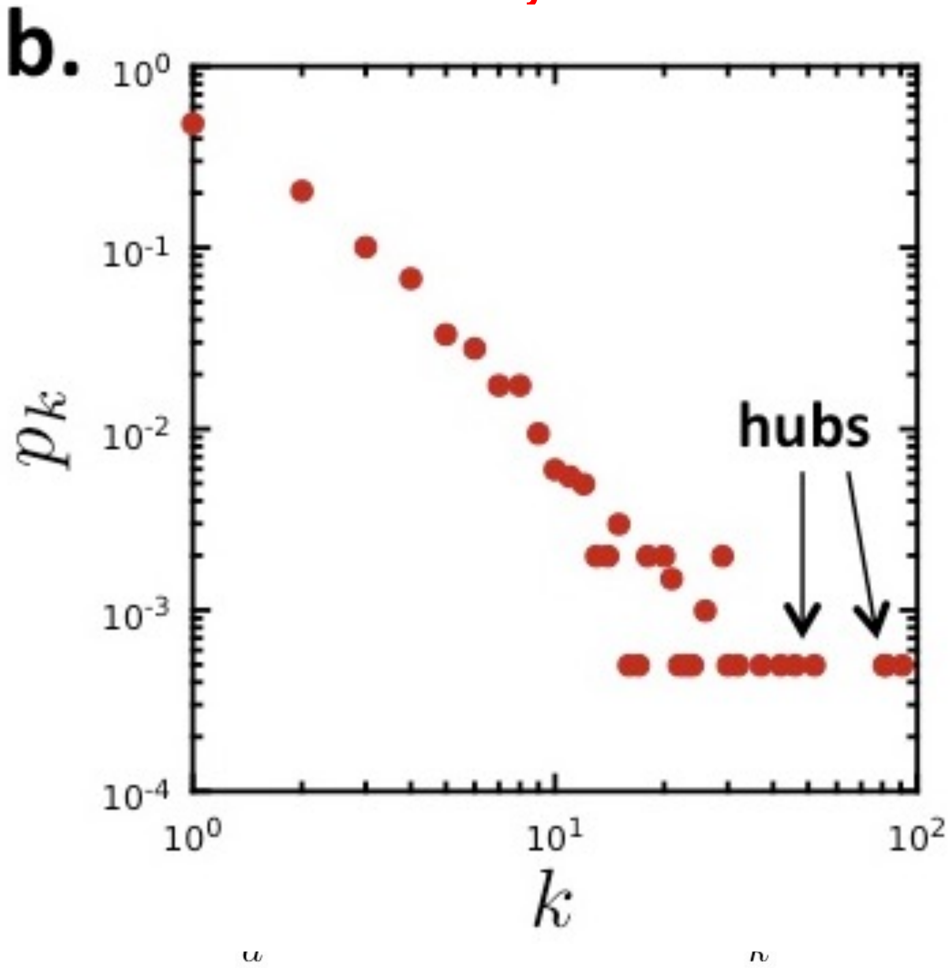


Undirected network
N=2,018 proteins as nodes
L=2,930 binding interactions as links.
Average degree <k>=2.90.

Not connected: 185 components
the largest (giant component) 1,647 nodes

Complex Networks

Case Study: Protein-Protein Interaction Network



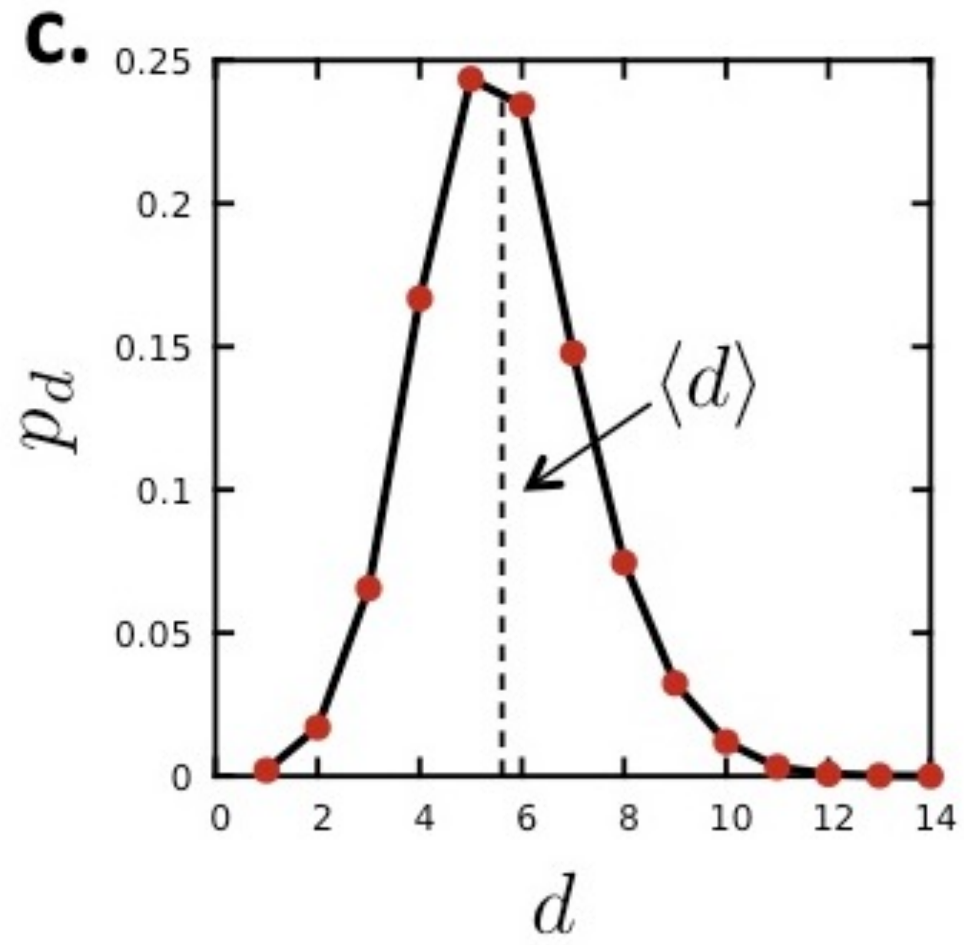
p_k is the probability that a node has degree k

N_k = # nodes with degree k

$p_k = N_k / N$

Complex Networks

Case Study: Protein-Protein Interaction Network

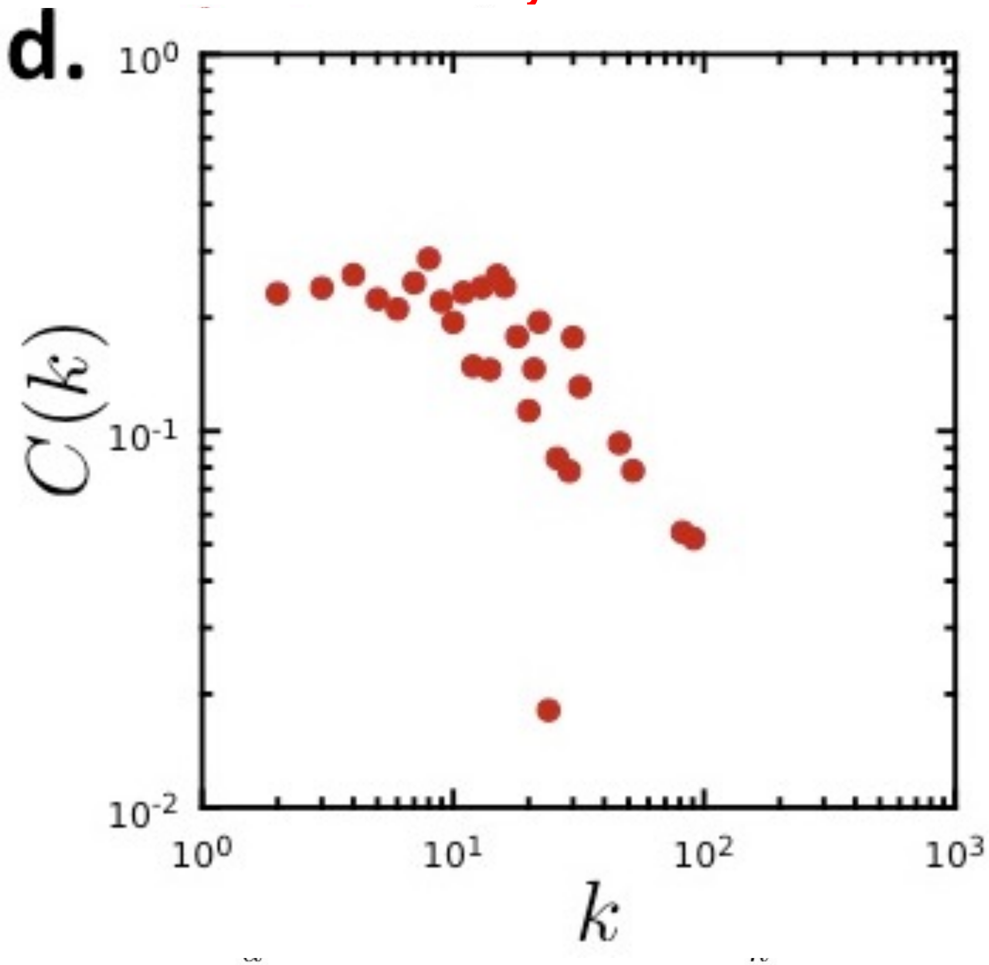


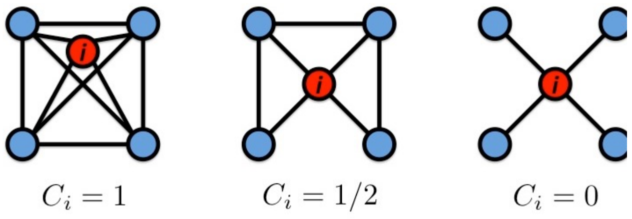
d_{max}=14

$\langle d \rangle$ = 5.61

Complex Networks

Case Study: Protein-Protein Interaction Network





$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$

$$\langle C \rangle = 0.12$$