Network Theory and Dynamic Systems

03. Small Worlds

SOSE 2025

Dr. -Ing. Stefania Zourlidou

Institute for Web Science and Technologies
Universität Koblenz

Recap from Previous Lecture

- The Language of Networks
  - The Components: Nodes, Links
  - Types of Networks and Representations
  - Features of Nodes and Links
- Properties that Characterize Structure & Behavior of Networks
- Roles of Networks in Affecting Processes Occurring on Network Structures

Objectives of this Lecture

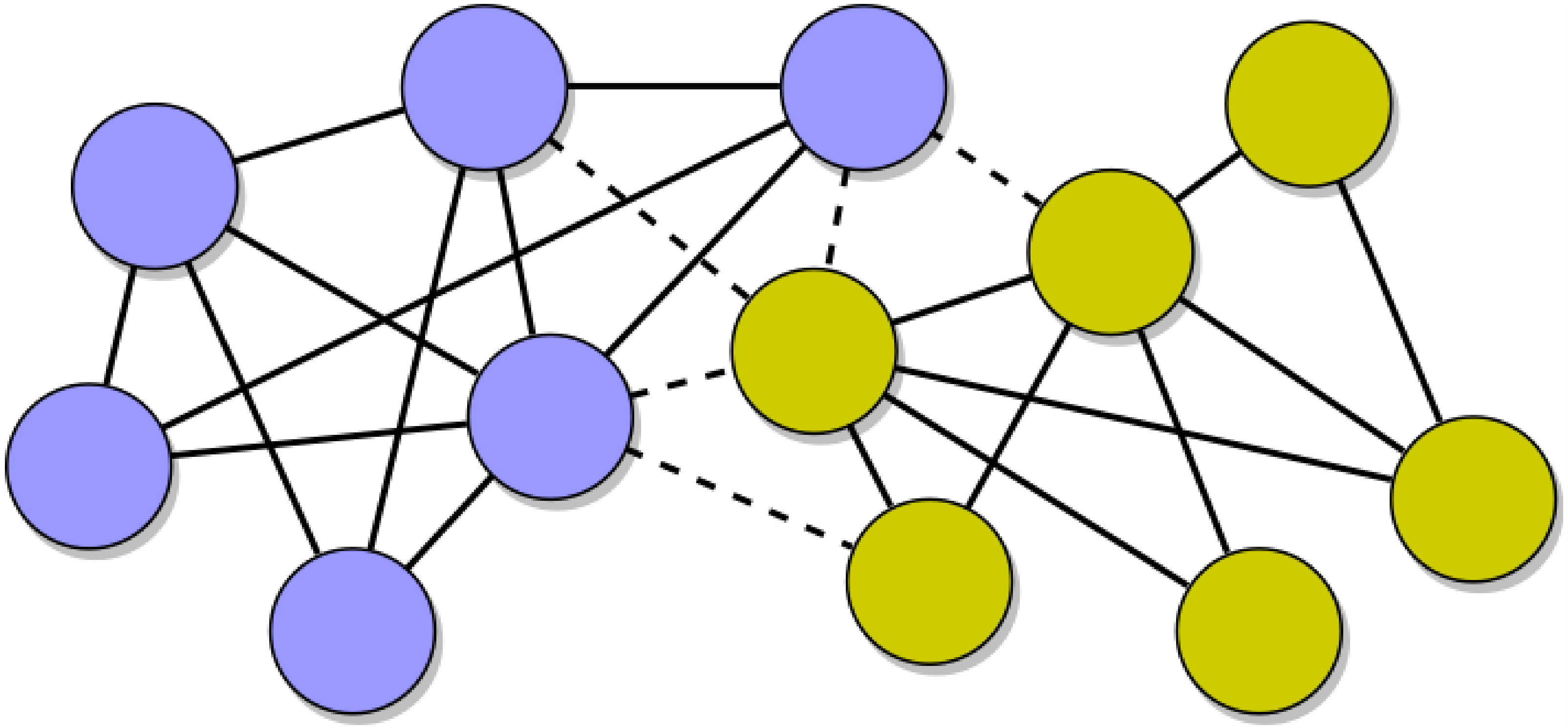
- Birds of a Feather
- Paths and Distances
- Connectedness and Components
- Trees
- Finding Shortest Paths
- Social Distance
- Six Degrees of Separation
- Small Worlds
- Friend of a Friend

universität koblenz
weiter:denken

1. Birds of a Feather

Birds of a Feather

Understanding the Tendency of Similar Individuals to Form Connections



- This network graph illustrates how nodes—each representing an individual with a specific trait (such as political views or musical preferences)—tend to connect more frequently with others who share similar attributes
- The color of each node denotes a particular characteristic
- As shown, nodes of the same color are more likely to be linked, highlighting the principle of homophily: the tendency for individuals to associate with those who are similar to themselves

Introduction to Assortativity

- Assortativity in social networks describes the tendency of nodes (such as individuals or entities) to form connections with others who share similar attributes
  - These attributes may include age, geographic location, interests, or social background
- It quantifies the extent to which similar nodes are more likely to be connected compared to dissimilar ones
- Understanding assortativity provides insight into how cohesive subgroups emerge within broader social structures, and reveals the underlying dynamics that drive these patterns

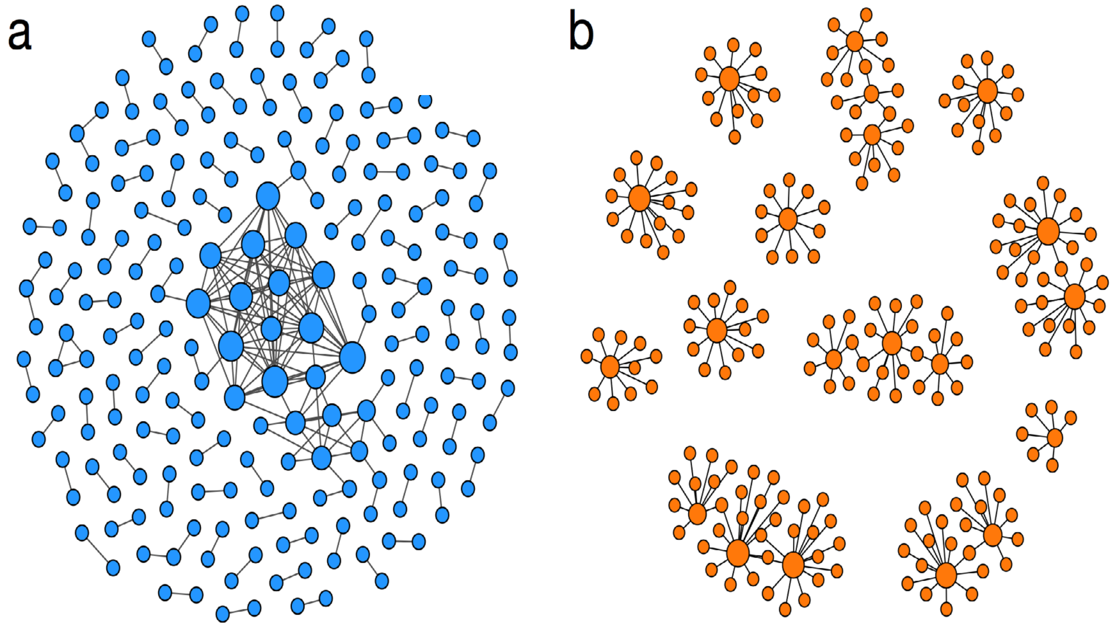
Assortativity

- Two possible mechanisms by which assortativity emerges naturally:
  1. **Selection** or **homophily**: similar nodes become connected
  2. (Social) **influence**: connected nodes become more similar
- It can also be a bad thing
  - For example "**echo chambers**" and "**groupthink**" are situations where your friends are like you, diversity is killed, and you are only exposed to opinions that reinforce your pre-existing beliefs…

We're going to explain it in more detail later

Degree Assortativity

- A.k.a. **degree correlation**
- Assortative networks have a **core-periphery** structure with hubs in the core
  - Example: social networks
- Disassortative networks have **hub-and-spoke** (or **star**) structure
  - Example: Web, Internet, food webs, bio networks



Assortativity in NetworkX (1/2)

universität koblenz
weiter:denken

```
# based on an a categorical attribute, such as gender
assort_a = nx.attribute_assortativity_coefficient(G, category)

# based on a numerical attribute, such as age
assort_n = nx.numeric_assortativity_coefficient(G, quantity)

# based on degree (Pearson correlation
# between degree of adjacent nodes)
r = nx.degree_assortativity_coefficient(G)
```

Assortativity in NetworkX (2/2)

Another way to compute the degree assortativity is by measuring the correlation between the degree and the average degree of the neighbors of nodes with that degree:

$$k_{nn}(i) = \frac{1}{k_i} \sum_j a_{ij} k_j$$

$$\langle k_{nn}(k) \rangle = \langle k_{nn}(i) \rangle_{i:k(i)=k}$$

```
import  scipy.stats
knn_dict = nx.k_nearest_neighbors(G)
k, knn = list(knn_dict.keys()), list(knn_dict.values ())
r, p_value = scipy.stats.pearsonr(k, knn)
```

The Dark Side of Homophily (1/2)

- **Homophily**—the tendency of individuals to form ties with others who are similar to them in key dimensions (e.g., interests, beliefs, or demographics)— can also have negative consequences:

?

The Dark Side of Homophily (2/2)

- **Homophily**—the tendency of individuals to form ties with others who are similar to them in key dimensions (e.g., interests, beliefs, or demographics)— can also have negative consequences:

the formation of homogeneous groups or clusters within larger networks, which can significantly impact the flow and integrity of information

⬇

Polarization, Misinformation, and Social Manipulation

Polarization

The process by which opinions within groups become increasingly extreme, driving groups further apart in values and beliefs.

- Homophily reduces exposure to differing perspectives and reinforces *confirmation bias*, where individuals predominantly encounter information that supports their existing views
- This dynamic contributes to a fragmented society in which shared understanding diminishes, compromise becomes more difficult, and the risk of social conflict and instability increases

Spread of Misinformation

universität koblenz
weiter:denken

- **Role of Echo Chambers**
  - **Echo Chamber Effect:** Social networks with high homophily foster environments where the same ideas, beliefs, or misinformation are continuously reinforced, reducing exposure to alternative or corrective viewpoints
  - **Vulnerability to Fake News:** These echo chambers accelerate the spread of misinformation, as individuals are more likely to accept content that aligns with their pre-existing beliefs without questioning its validity.
- **Exploitation by Malicious Actors**
  - Manipulative agents (e.g., state-sponsored groups, malicious organizations) can exploit homophilous networks by strategically injecting misinformation that spreads rapidly within like-minded communities, deepening social divides
  - **Example**: Coordinated misinformation campaigns during election periods that target specific ideological or demographic groups to influence public opinion or suppress voter turnout

Social Bots and Manipulation Tactics

- **Social Bots**
  - Automated accounts designed to imitate human behavior in online spaces
  - They are used to manipulate discourse, amplify particular viewpoints, and disseminate misinformation at scale
- **Sybil Attacks:**
  - The creation of numerous fake identities to fabricate consensus or propagate specific narratives, distorting public perception and potentially influencing real-world outcomes
- The orchestrated use of such tactics can undermine democratic processes by fostering political polarization and misleading the public on critical issues

Concluding Thoughts on Network Assortativity

- Implications for Social Dynamics
  - Social Cohesion and Segregation
    - **Homophily** can enhance social cohesion by nurturing close-knit communities formed around shared beliefs or attributes
    - At the same time, it can contribute to **social fragmentation**, reinforcing echo chambers and deepening societal divides by limiting interaction across differing viewpoints
  - Information Dissemination
    - Assortative networks tend to facilitate efficient information flow within homogeneous clusters, yet **may hinder communication** across heterogeneous groups
    - This pattern is especially relevant for understanding the spread of information and **misinformation** in online social platforms, where echo chambers and algorithmic curation play significant roles

Concluding Thoughts on Network Assortativity

- **Practical Applications**
  - **Marketing and Advertising**
    - Businesses can leverage assortativity measurements to refine targeted marketing strategies, ensuring that promotional materials reach the most receptive audiences
  - **Public Health**
    - Understanding social network structures can enhance the effectiveness of public health campaigns, particularly in promoting healthy behaviors or in vaccination drives
  - **Political Campaigning**
    - Politicians and activists can use knowledge of network assortativity to craft messages that resonate within and across ideological clusters

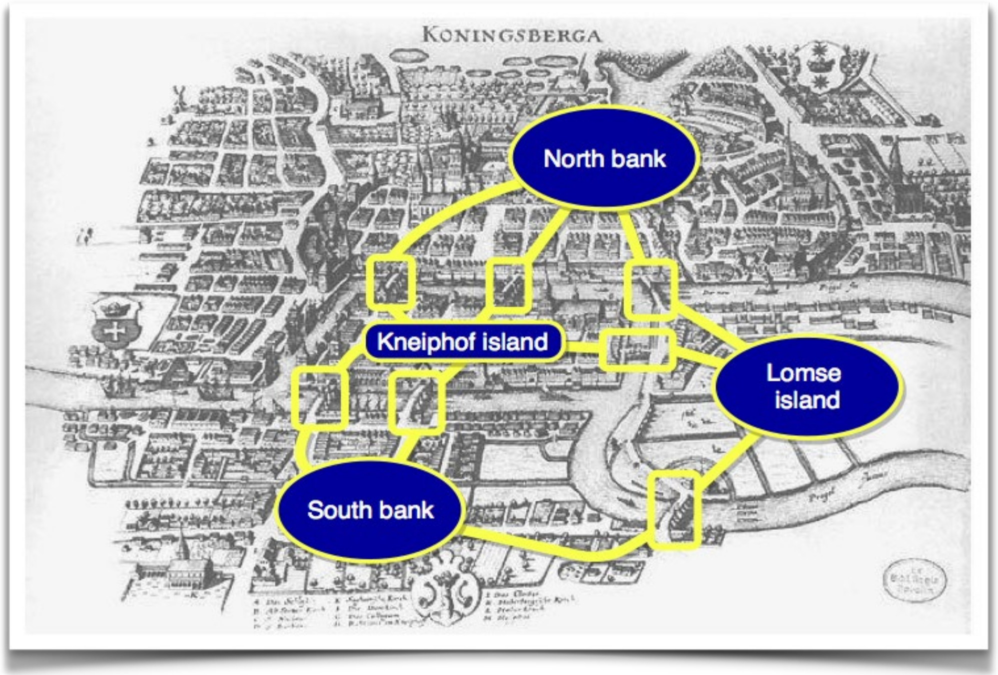
2. Paths and Distances

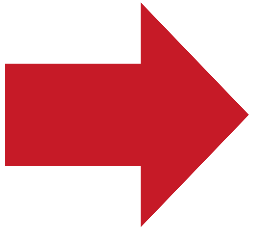
Paths: Definitions

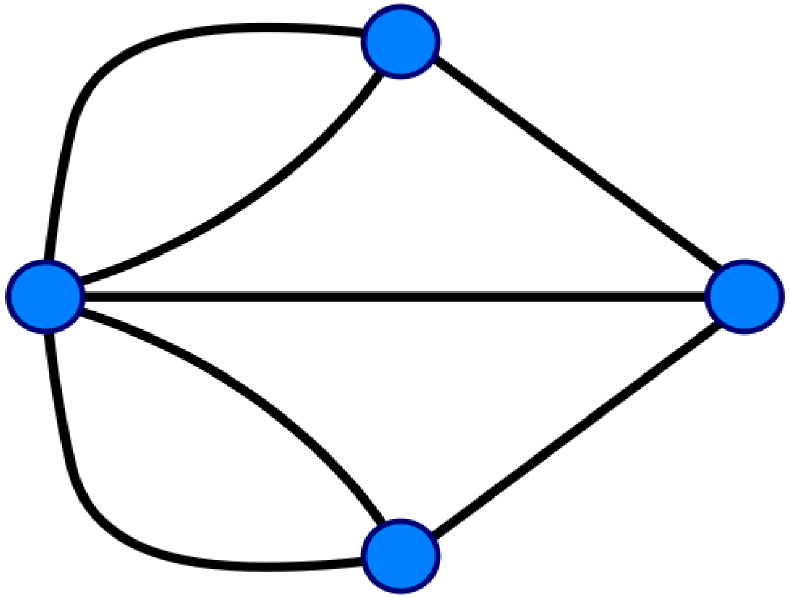
universität koblenz
weiter:denken

- **Path**: sequence of links traversed to go from a **source** to a **target** node
  - In a directed network, links must be traversed according to their direction
  - Note: A path may not always exist between two nodes
- **Cycle**: path where source and target node are the same
- **Simple path**: A path that does **not** revisit any link
  - We will only deal with simple paths
- **Path length**: number of links in path
- Finding paths was the earliest problem studied in network science

Euler circa 1736: Koningsberg Bridges





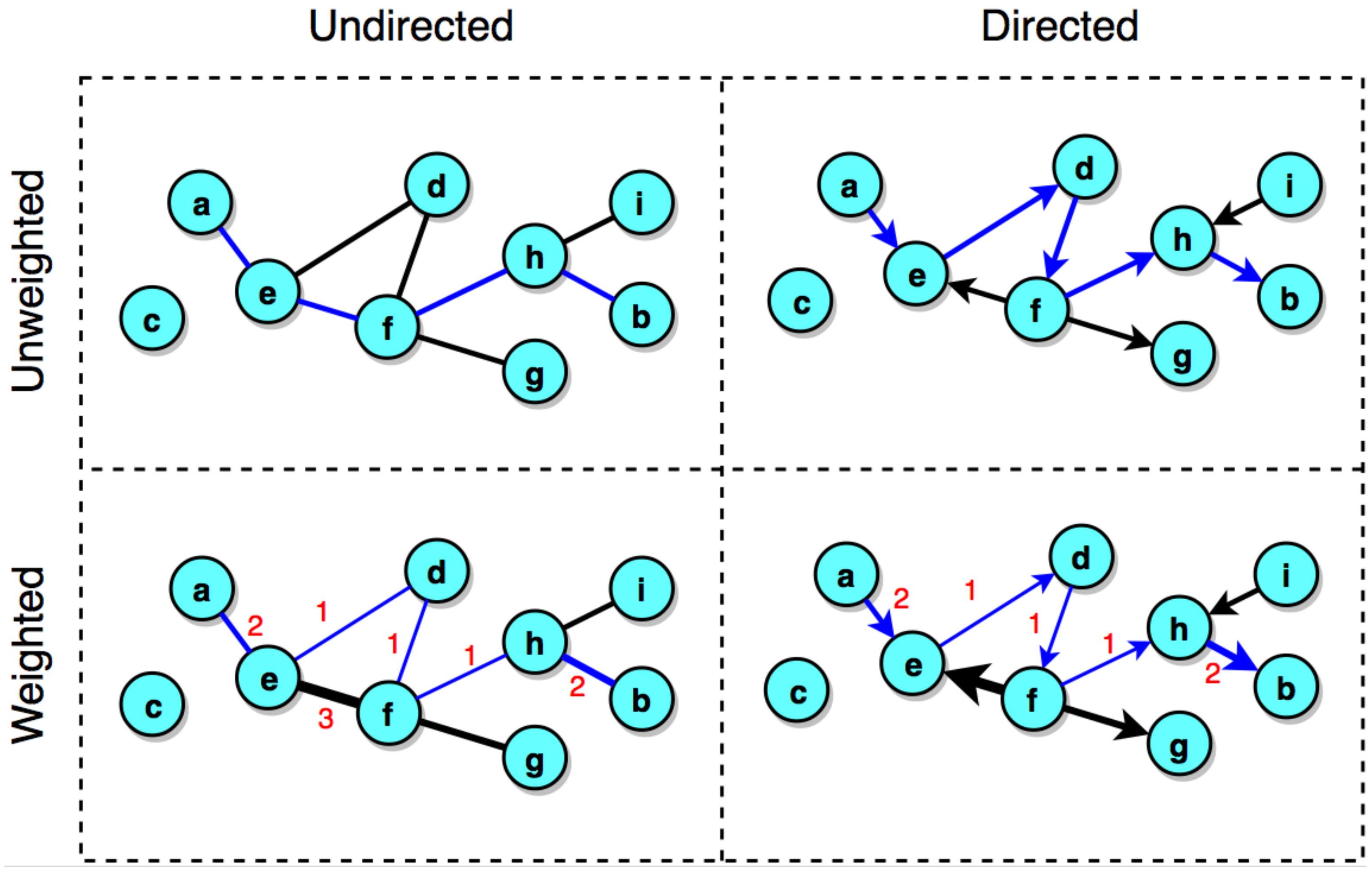


Q: Can you cross all 7 bridges just once each?

A: No. At most two nodes (start, end) may have odd degree

Shortest Paths

- **Shortest path** between two nodes: minimal length (there may be more than one)
  - In weighted networks, weights may represent distances
- **Shortest path length** or **distance**: length of shortest path
  - Undefined (∞) if there is no path



APL and Diameter (1/2)

- We can use the shortest paths to characterize a network:
  - The **diameter** is the longest shortest-path length, or the maximum of the shortest path lengths across all pairs of nodes:

$$\ell_{max} = \max_{i,j} \ell_{ij}$$

  - The **average path length** (APL) is the average of the shortest path lengths across all pairs of nodes
    - Undirected network: $\langle \ell \rangle = \frac{\sum_{i,j} \ell_{ij}}{\binom{N}{2}} = \frac{2\sum_{i,j} \ell_{ij}}{N(N-1)}$
    - Directed network:

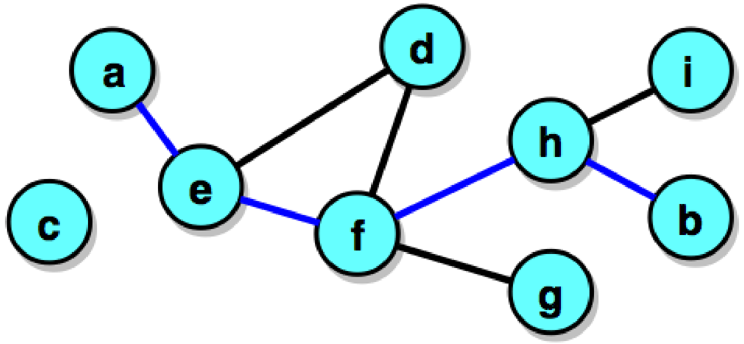
$$\langle \ell \rangle = \frac{\sum_{i,j} \ell_{ij}}{N(N-1)}$$

APL and Diameter (2/2)

- What if there is not a path between one or more pairs of nodes?
  - We can say APL and diameter are undefined (as NetworkX does)
  - We can measure APL and diameter within the largest connected component (defined later)
  - We can use a mathematical trick:

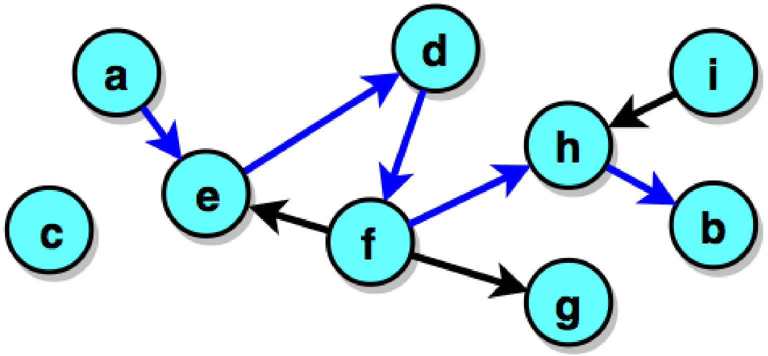
$$\langle \ell \rangle = \left(\frac{\sum_{i,j} \frac{1}{\ell_{ij}}}{\binom{N}{2}} \right)^{-1}$$

Paths and APL (1/2)

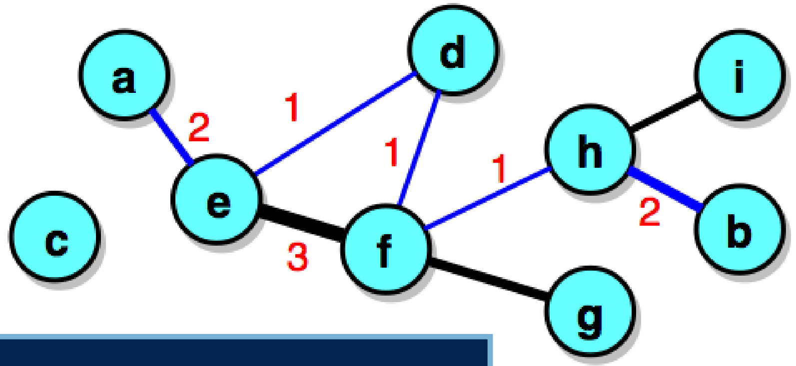



```
nx.has_path(G, 'a', 'c')            # False
nx.has_path(G, 'a', 'b')            # True
nx.shortest_path(G, 'a', 'b')       # ['a','e','f',h','b']
nx.shortest_path_length(G,'a','b')  # 4
nx.shortest_path(G, 'a')            # dictionary
nx.shortest_path_length(G, 'a')     # dictionary
nx.shortest_path(G)                 # all pairs
nx.shortest_path_length(G)          # all pairs
nx.average_shortest_path_length(G)  # error
G.remove_node('c')                  # make G connected
nx.average_shortest_path_length(G)  # now okay
```

Paths and APL (2/2)




```
nx.has_path(D, 'b', 'a')       # False
nx.has_path(D, 'a', 'b')       # True
nx.shortest_path(D, 'a', 'b') # ['a','e','d','f',h','b']
```




```
nx.shortest_path_length(W, 'a', 'b')           # 4
nx.shortest_path_length(W, 'a', 'b', 'weight') # 7
```

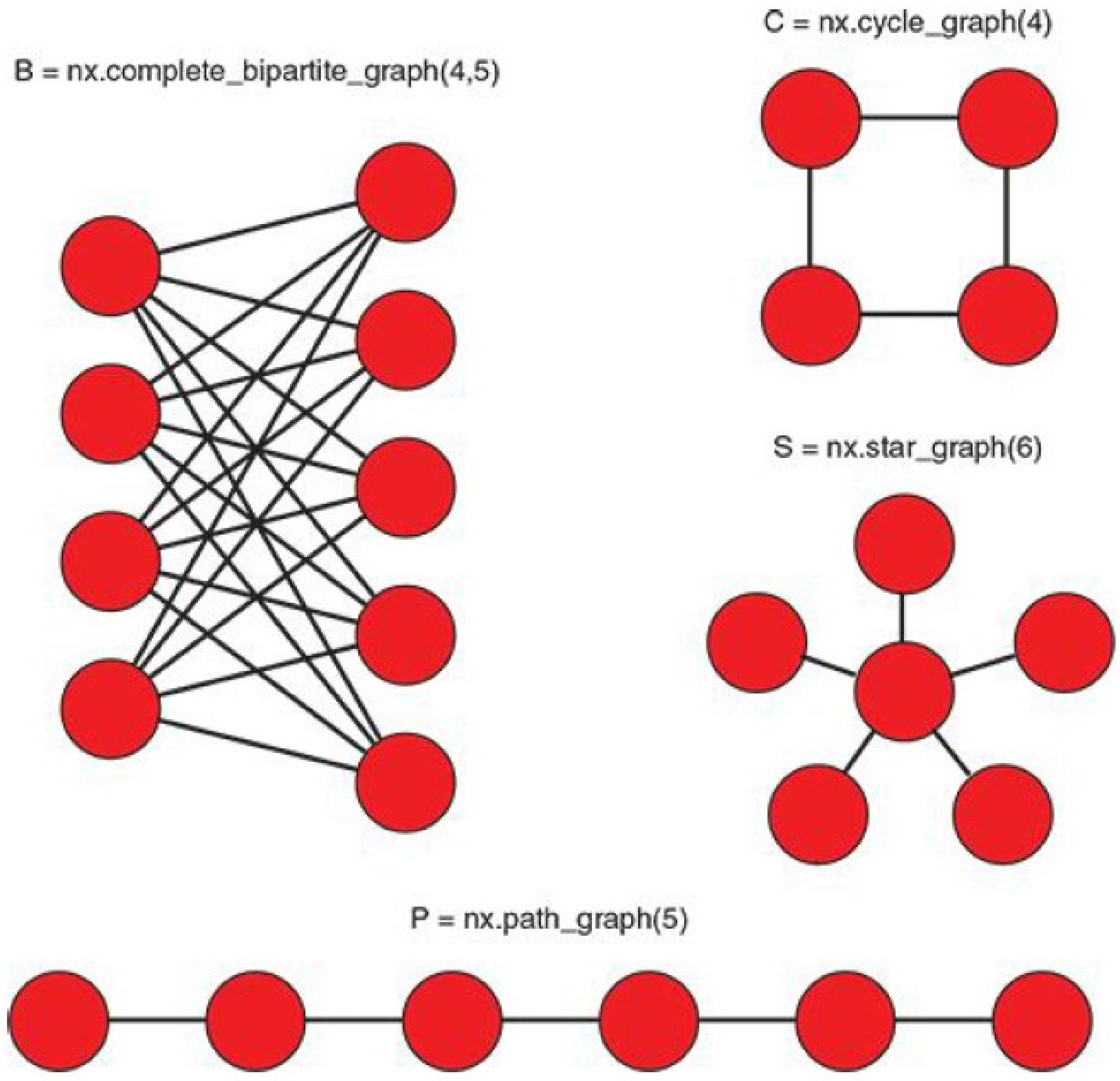
3. Connectedness and Components

Concept of Connectedness

- **Connectedness** in a network refers to the ability to reach any node from any other node via paths of intermediate nodes and links
- It is a critical measure that defines the integrity and usability of the network, influencing how information or processes flow through the network
- **Density and Connectivity**: Higher density generally implies better connectedness, meaning that the network can more likely sustain connectivity even if some links are removed or fail

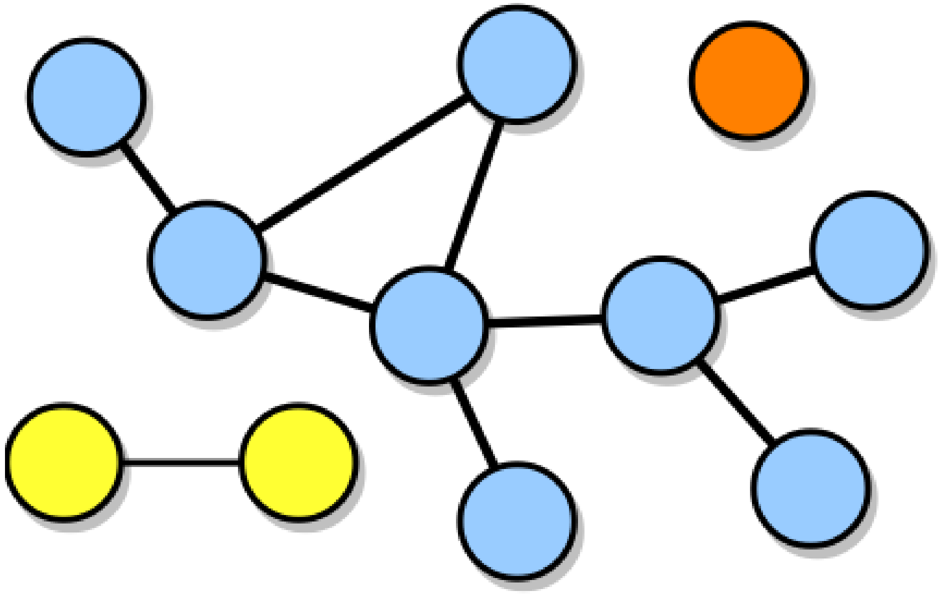
Connectedness

```
K4 = nx.complete_graph(4)
nx.is_connected(K4) # True
C = nx.cycle_graph(4)
nx.is_connected(C) # True
P = nx.path_graph(5)
nx.is_connected(P) # True
S = nx.star_graph(6)
nx.is_connected(S) # True
```



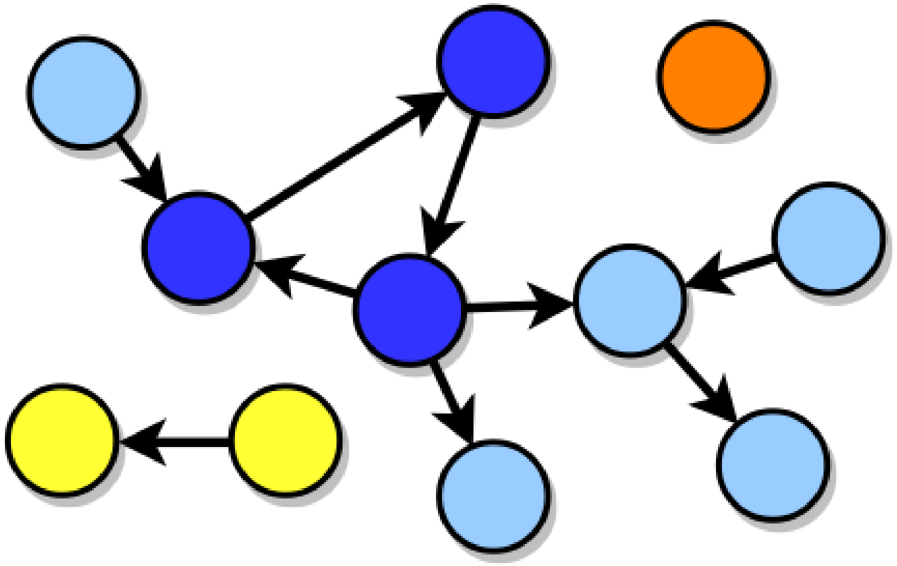
Connectedness and Components (1/3)

- A network is **connected** if there is a path between any two nodes
- If a network is not connected, it is **disconnected** and has multiple connected components
- A **connected component** is a connected subnetwork
  - The largest one is called **giant component**; it often includes a substantial portion of the network
  - A **singleton** is the smallest-possible connected component

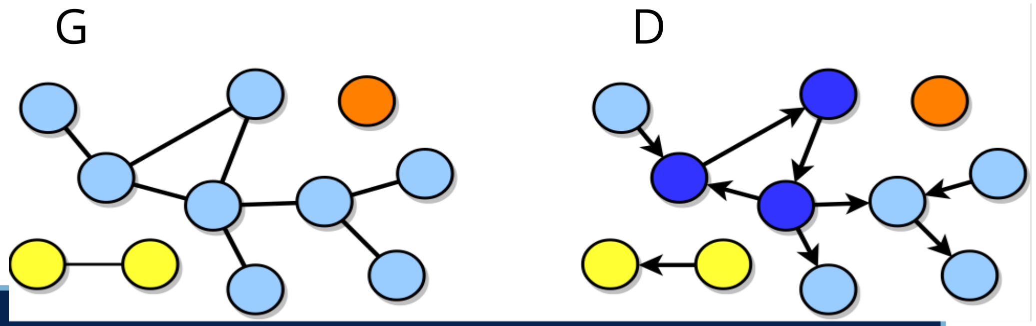


Connectedness and Components (2/3)

- **A directed network** can be **strongly connected** or **weakly connected** if there is a path between any two nodes, respecting or disregarding the link directions, respectively
- Similarly for **strongly connected** or **weakly connected components**
- The **in-component** of a strongly connected component S is the set of nodes from which one can reach S, but that cannot be reached from S
- The **out-component** of a strongly connected component S is the set of nodes that can be reached from S, but from which one cannot reach S



Connectedness and Components (3/3)




```
nx.is_connected(G)                            # False
comps = sorted(nx.connected_components(G),
               key=len, reverse=True)
nodes_in_giant_comp = comps[0]
GC = nx.subgraph(G, nodes_in_giant_comp)
nx.is_connected(GC)                           # True
nx.is_strongly_connected(D)                   # False
nx.is_weakly_connected(D)                     # False
list(nx.weakly_connected_components(D))
list(nx.strongly_connected_components(D))     # lots of
                                              # singletons
```

4. Trees

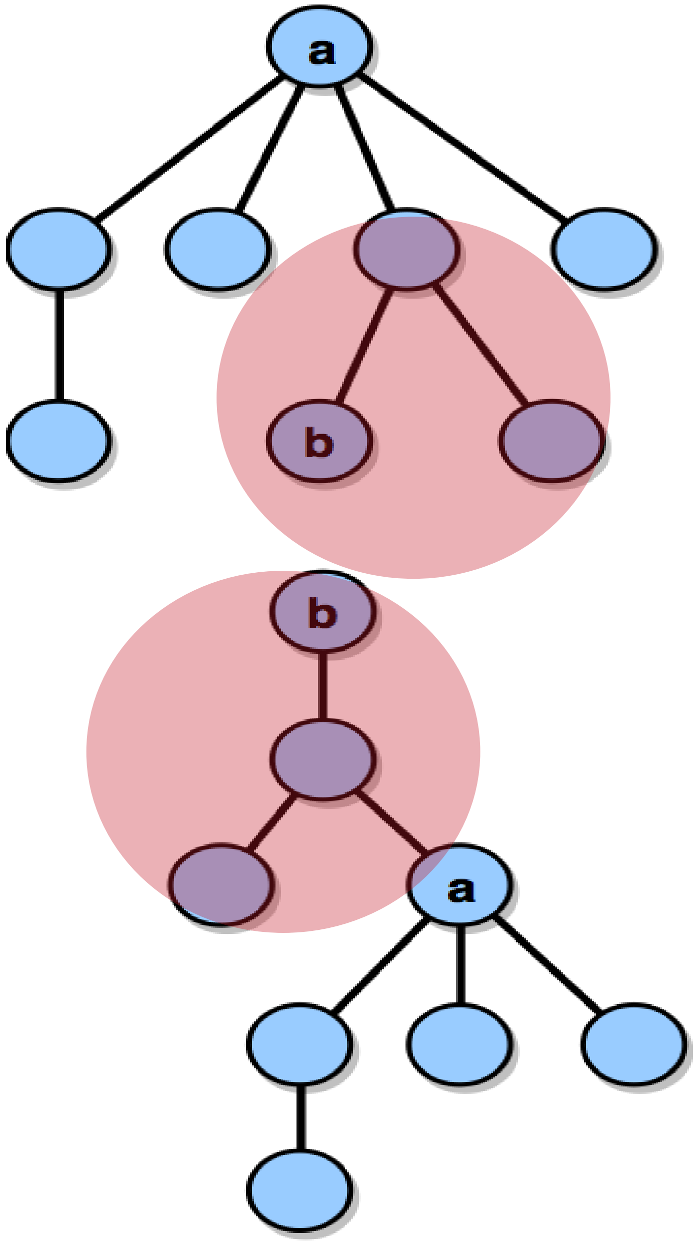
Tree (in Graph Theory)

universität koblenz
weiter:denken

- A **tree** is a special class of undirected, connected networks characterized by the property that the removal of any single link (edge) results in the network splitting into exactly two distinct **components**
- This structural feature implies that trees contain no cycles (closed loops), making them a minimal connected structure; they have the minimum number of links required to maintain connectivity among the nodes, which equals the number of nodes minus one (N-1 links for N nodes)
- Trees are fundamental in various fields such as data structures, algorithm design, network theory, and more, providing frameworks for processes and analyses that require efficient, hierarchical connectivity without redundancies

Trees (1/2)

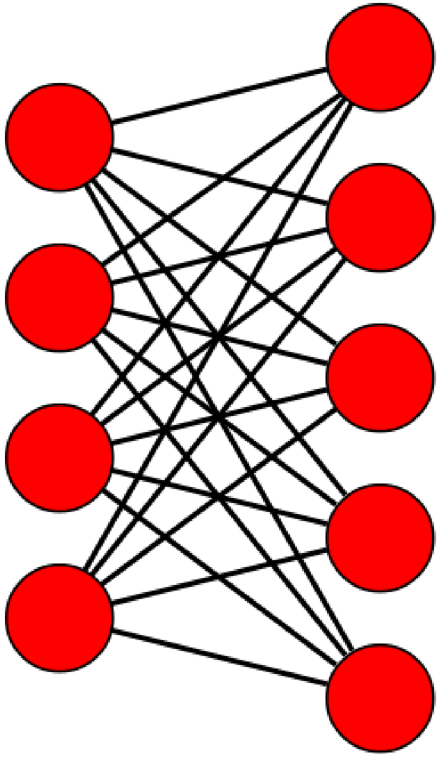
- A **tree** is a **connected** network **without cycles**
- A **tree** is a **connected** network **with *N*-1 links**
  - Exercise: prove that these two definitions are equivalent
- In a tree there is a **single path** between any two nodes
- Trees are **hierarchical**: you can pick a node as the **root**. Each node is connected to a **parent** node (toward the root) and to one or more **children** nodes (away from the root). Exceptions:
  - The root has no parent
  - The **leaves** have no children



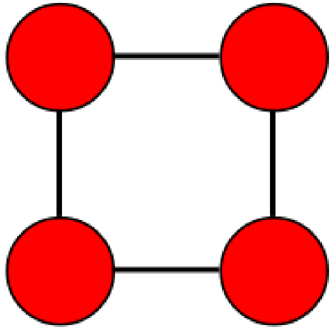
Trees (2/2)

```
K4 = nx.complete_graph(4)
nx.is_tree(K4)                # False

nx.is_tree(B)                 # False
nx.is_tree(C)                 # False

nx.is_tree(S)                 # True
nx.is_tree(P)                 # True
```

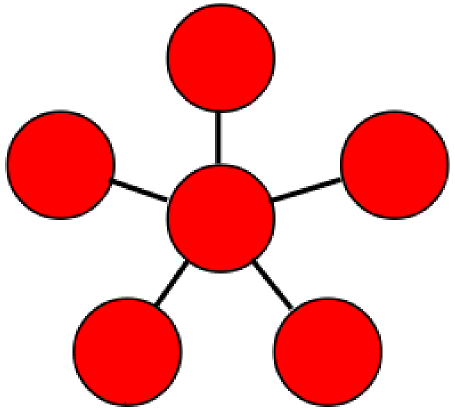
B = nx.complete_bipartite_graph(4,5)



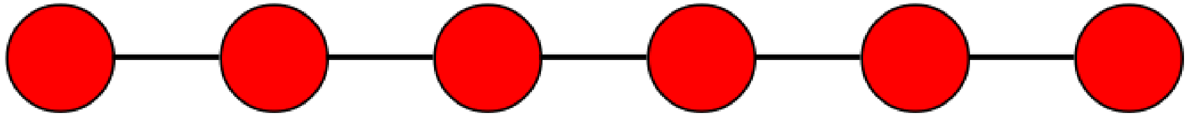
C = nx.cycle_graph(4)



S = nx.star_graph(6)



P = nx.path_graph(5)



5. Finding Shortest Paths

Introduction to Finding Shortest Paths

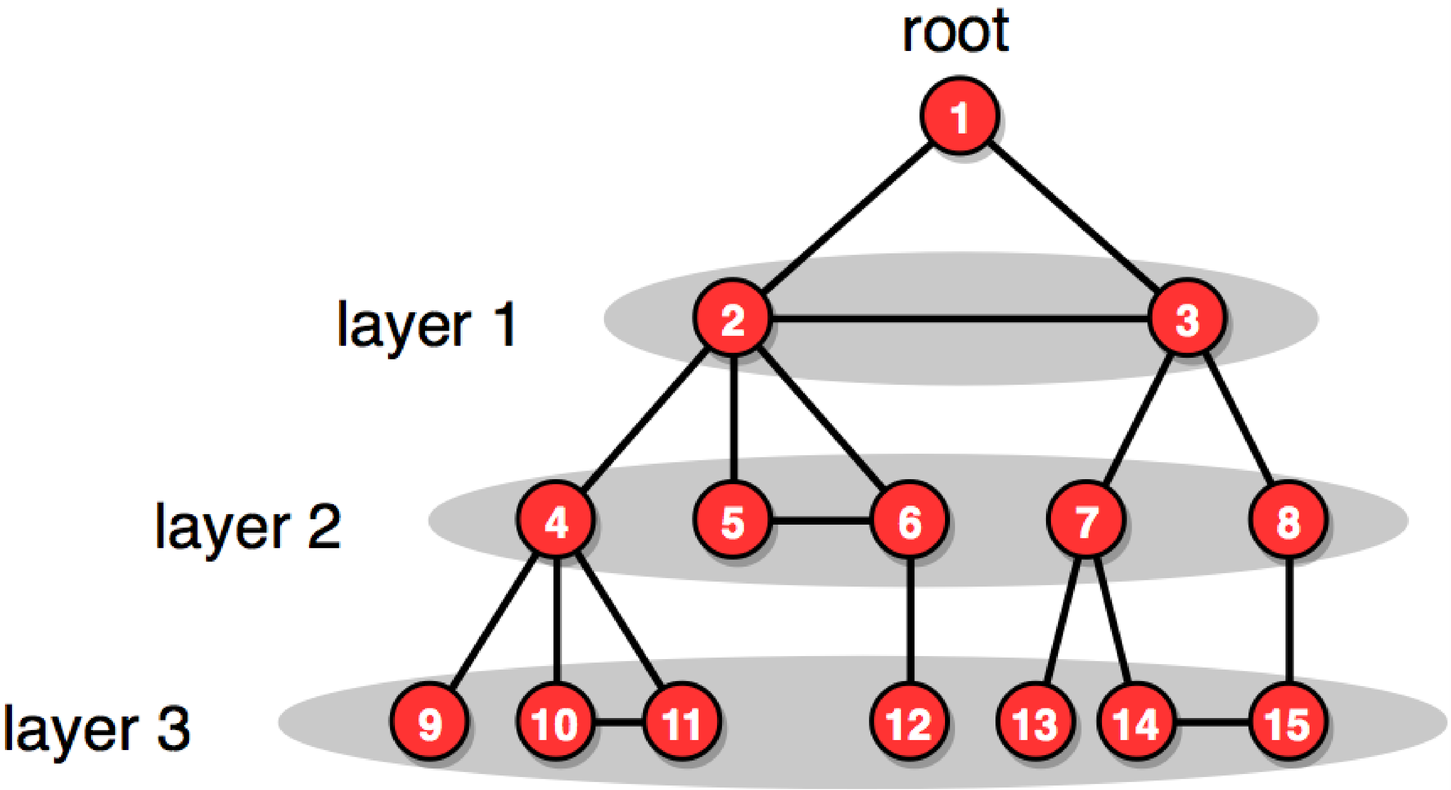
- To determine the shortest path between two nodes, network analysis tools like NetworkX systematically explore the network from a starting point (source node) to every other node
- This is essential in various applications, including routing and navigation systems, data organization, and optimizing network communications
- **Web Crawlers**
  - Search engines use automated programs known as web crawlers that navigate the web to find and index new pages
  - The process involves mapping the web as a network of pages (nodes) linked by hyperlinks (edges) and using shortest path algorithms to efficiently traverse this network

Breadth-First Search

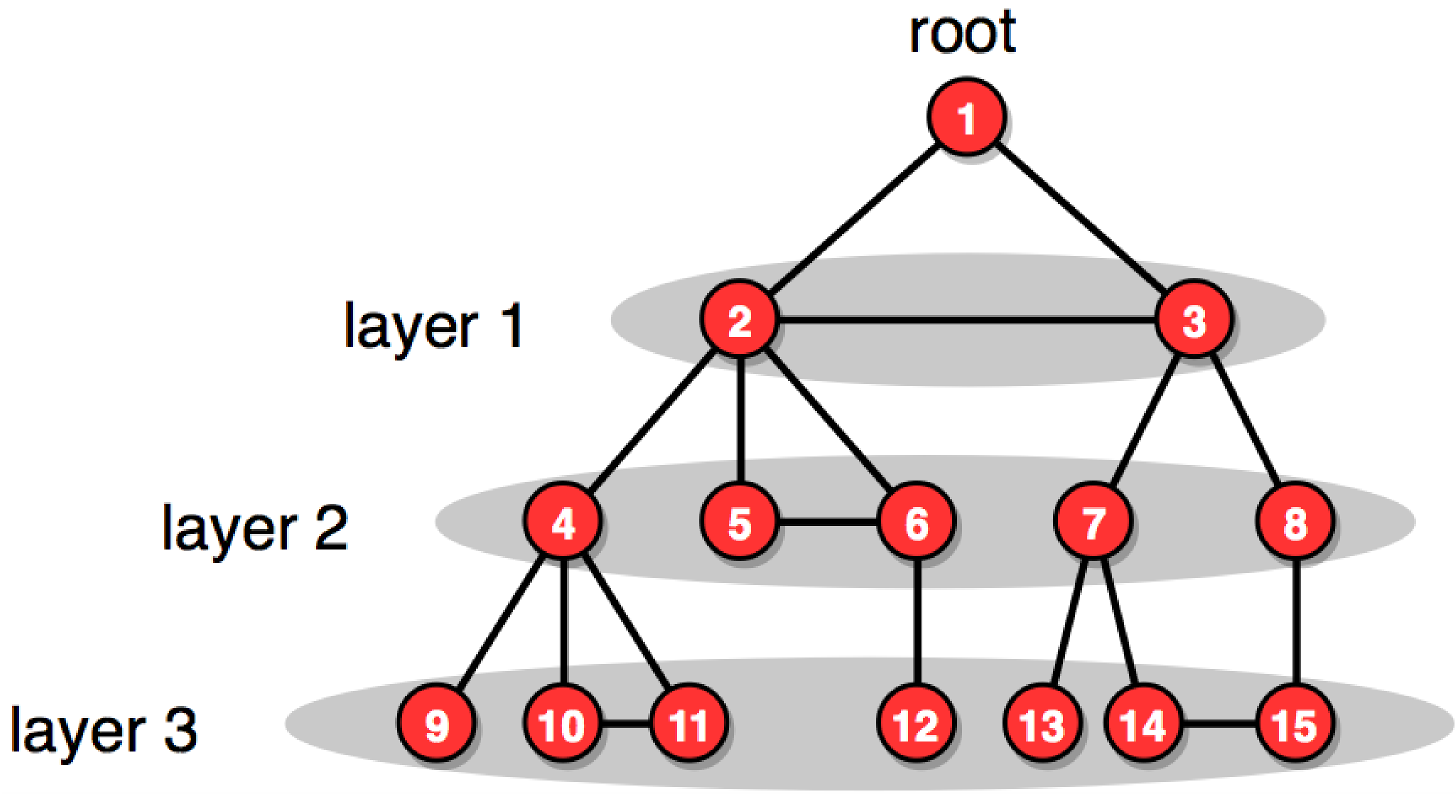
The algorithm, or procedure for navigating through a network starting from a *source* node and finding the shortest path between the source and every other node in the network is called ***breadth-first search***

The Breadth-First Search Algorithm (1/3)

- Start from a source node (root)
- Visit the entire breadth of the network, within some distance from the source, before we move to a greater depth, farther away from the source
- Start from each node to find all-pairs-shortest-paths (slow: $O(N^2)$)



The Breadth-First Search Algorithm (2/3)

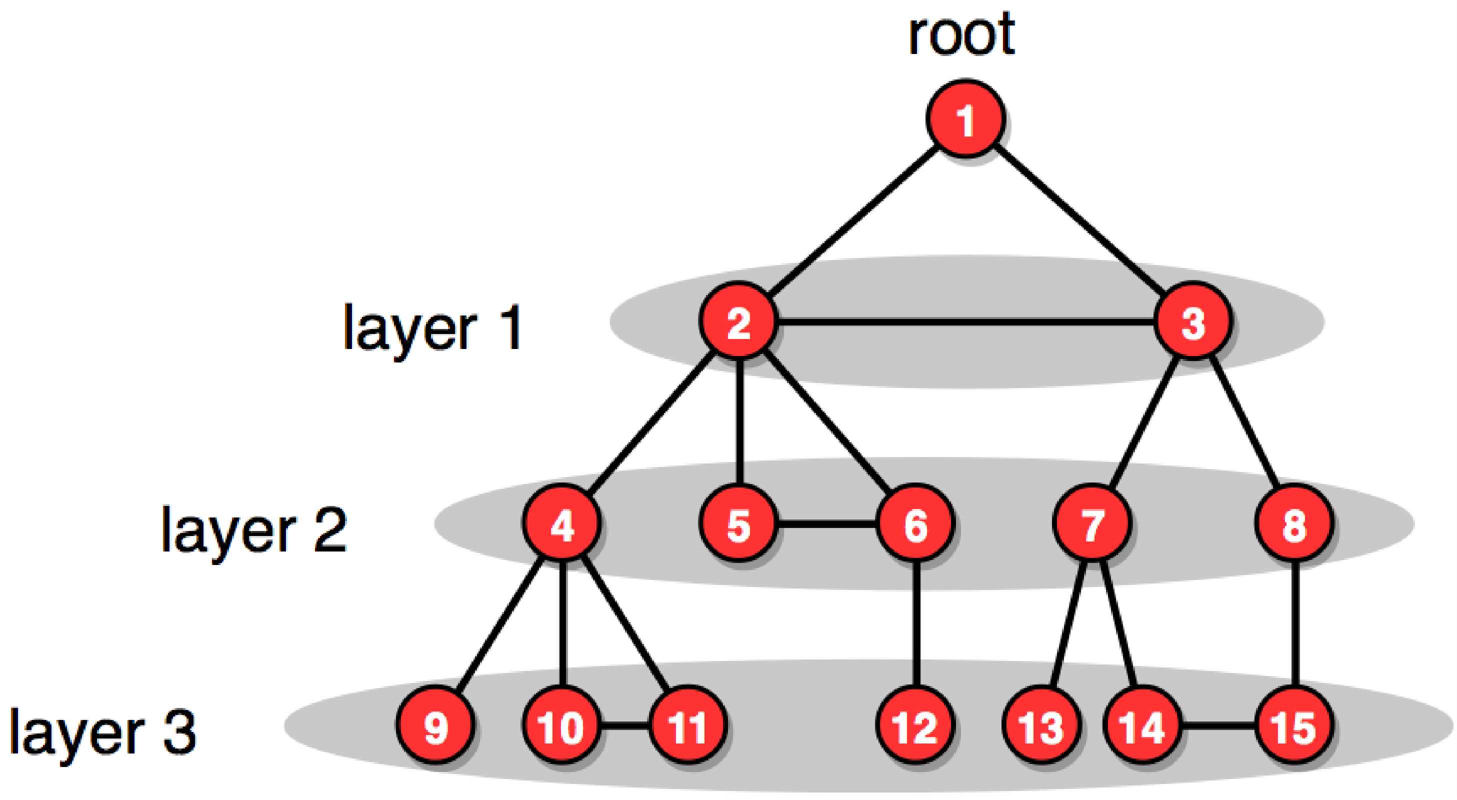


Node **1** is selected as the source. First, we visit the neighbors of **1**, which are **2** and **3**. This is layer 1, including all nodes one step away from the source. Then we move to their neighbors **4**, **5**, **6**, **7**, **8**, which are two steps away from the source (layer 2). Finally, we reach nodes **9**, **10**, **11**, **12**, **13**, **14**, **15**, at distance three from the source (layer 3).

The Breadth-First Search Algorithm (3/3)

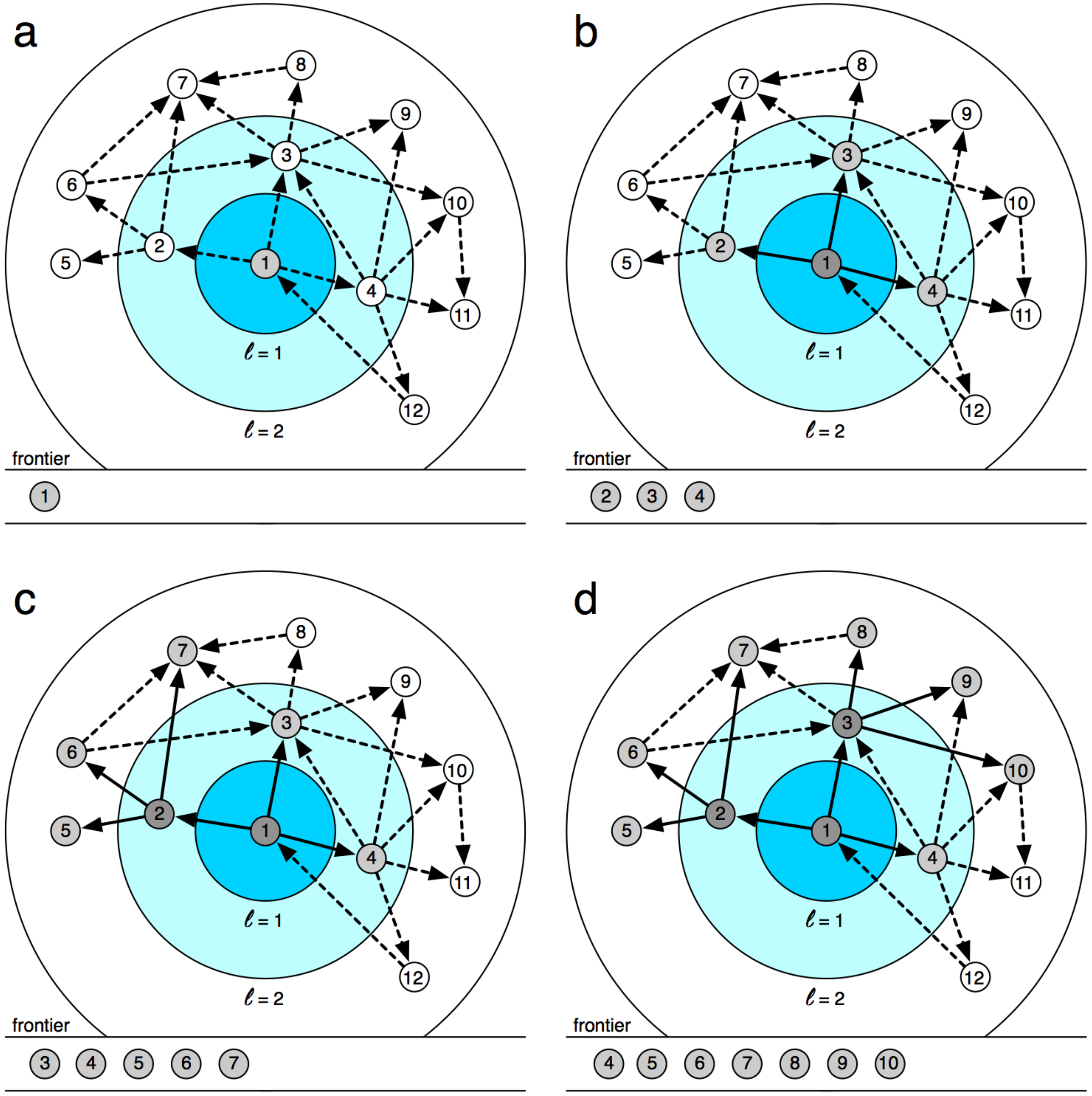
- To find the shortest paths from the source to the other nodes, the breadth-first search algorithm builds a directed *shortest-path tree*, containing the same nodes as the original network but only a subset of the links

- The tree maps the shortest paths between its root (the source node) and all other nodes



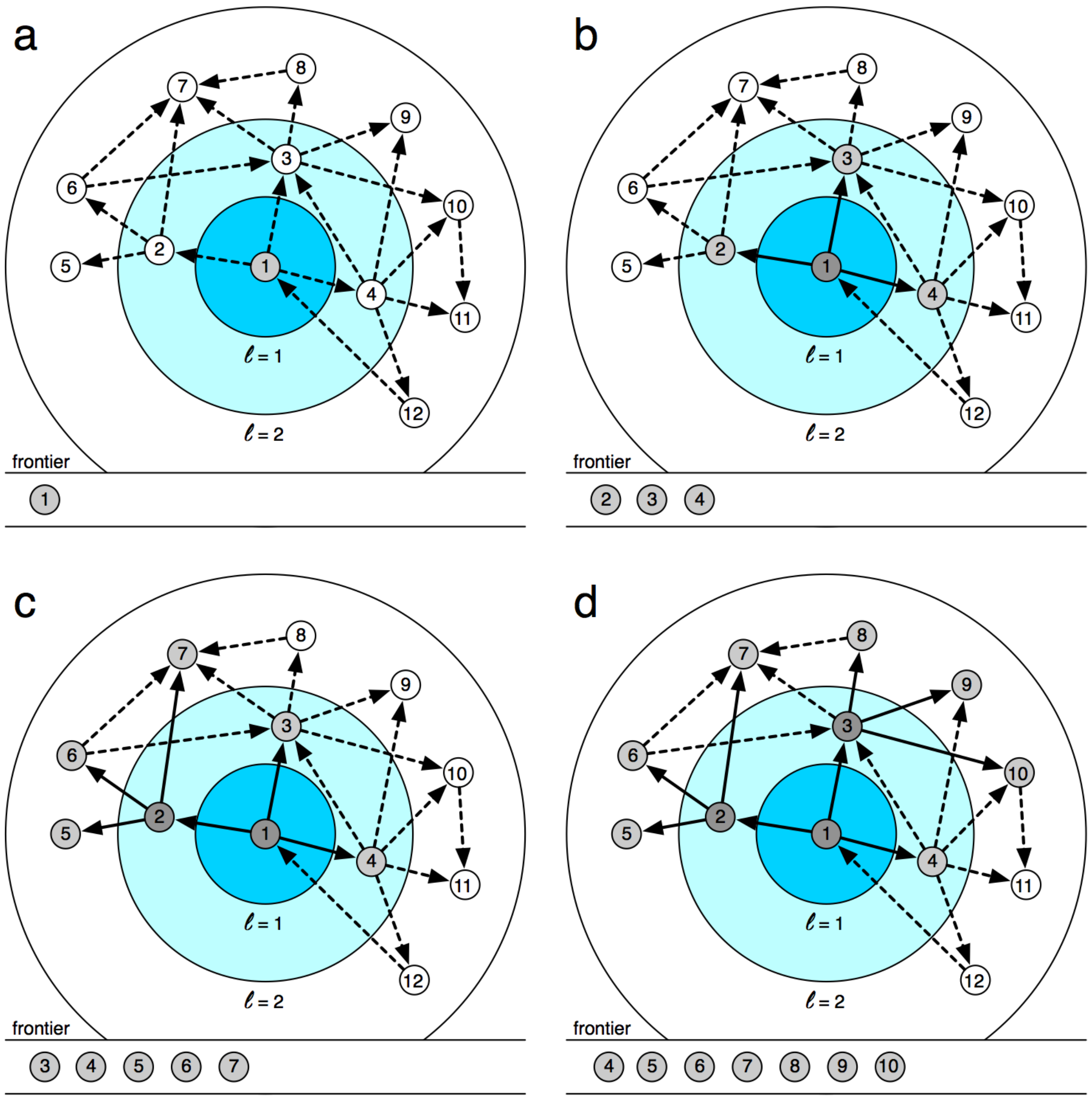
Breadth-First Search (BFS)

- Each node has an attribute storing its **distance** ℓ from the source, initially ℓ = –1 except ℓ(source) = 0
- A queue (FIFO) holds the **frontier**, initially contains the source
- A directed **shortest path tree**, initially all the nodes and no links
- Iterate until the frontier is empty:
  1. Remove next node *i* in frontier
  2. For each neighbor/successor *j* of *i* with $\ell(j) = -1$:
     1. Queue *j* into frontier
     2. $\ell(j) = \ell(i) + 1$
     3. Add link $(i \rightarrow j)$ to shortest-path tree



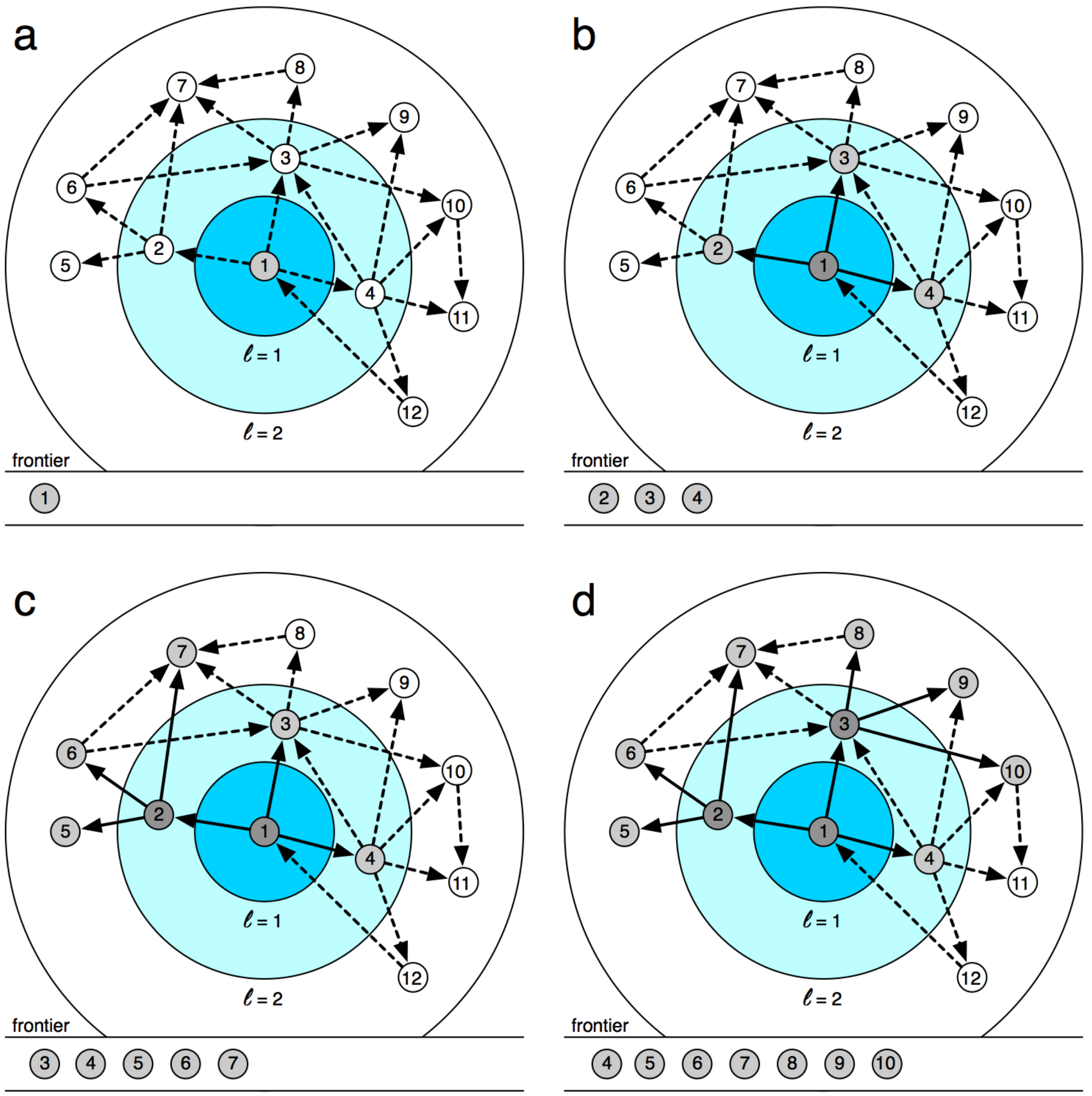
Breadth-First Search (BFS)

- To find the shortest path from the source to any target node, we have to follow the links in the shortest-path tree backward from the target node through predecessors in the upper layers, until we arrive at the source
- Recall that in a tree there is a single path to the root; each node has a single predecessor. Then we have to reverse the path to obtain the shortest path from the source to the target
  - In an undirected network this is the same as the path from the target to the source, but in a directed network they may be different



BFS Example

- What is the shortest path from node **1** to node **7**?
- We go from **7** to its predecessor in the shortest-path tree, which is node **2**, and then to its predecessor, which is the root node **1**. Reversing this path we obtain the shortest path **1 2 7**
- Note that this is not the only short path — the path **1 3 7** has the same length, but the algorithm only identifies *one* shortest path from the source
- Note also that in this directed network, the shortest path from node **7** to node **1** is not the same; in fact there is no such path



BFS, Large Networks and Weighted Networks

- BFS is optimal for finding the shortest paths in unweighted networks from a single source node to all other nodes
  - It explores all neighbors at the current depth prior to moving on to nodes at the next depth level
- While BFS is straightforward, it becomes computationally burdensome when applied to large networks or when paths are weighted
  - Utilizing the shortest_path(G) or shortest_path_length(G) functions in NetworkX on large networks demonstrates significant computational delays, illustrating the practical challenges in real-world applications
  - Even if short paths exist, they are not necessarily easy to find
- Slightly more complicated algorithms also exist for shortest paths in weighted networks

6. Social Distance, Six Degrees of Seperation

Social Distance

- How close or distant are two nodes in a network?
  - This question is fundamentally about the average path length between nodes
  - Average path length provides a measure of node closeness—offering insights into how efficiently information or influence can spread through a network
- The question has been explored extensively in social networks
  - Unlike grid-like structures (e.g., road or power networks), which often have long paths, social networks tend to show **surprisingly short average path lengths**, due to high **connectivity** and **clustering**
  - Let us start by considering coauthorship networks, in which nodes are scholars and links represent two people having coauthored one or more publications
  - Coauthorship data is both **accessible** and **reliable**, making these networks a valuable case study for examining **social collaboration dynamics**

Paul Erdős

Image by Kmhkmh - CC BY 3.0
https://commons.wikimedia.org/w/index.php?curid=38087162

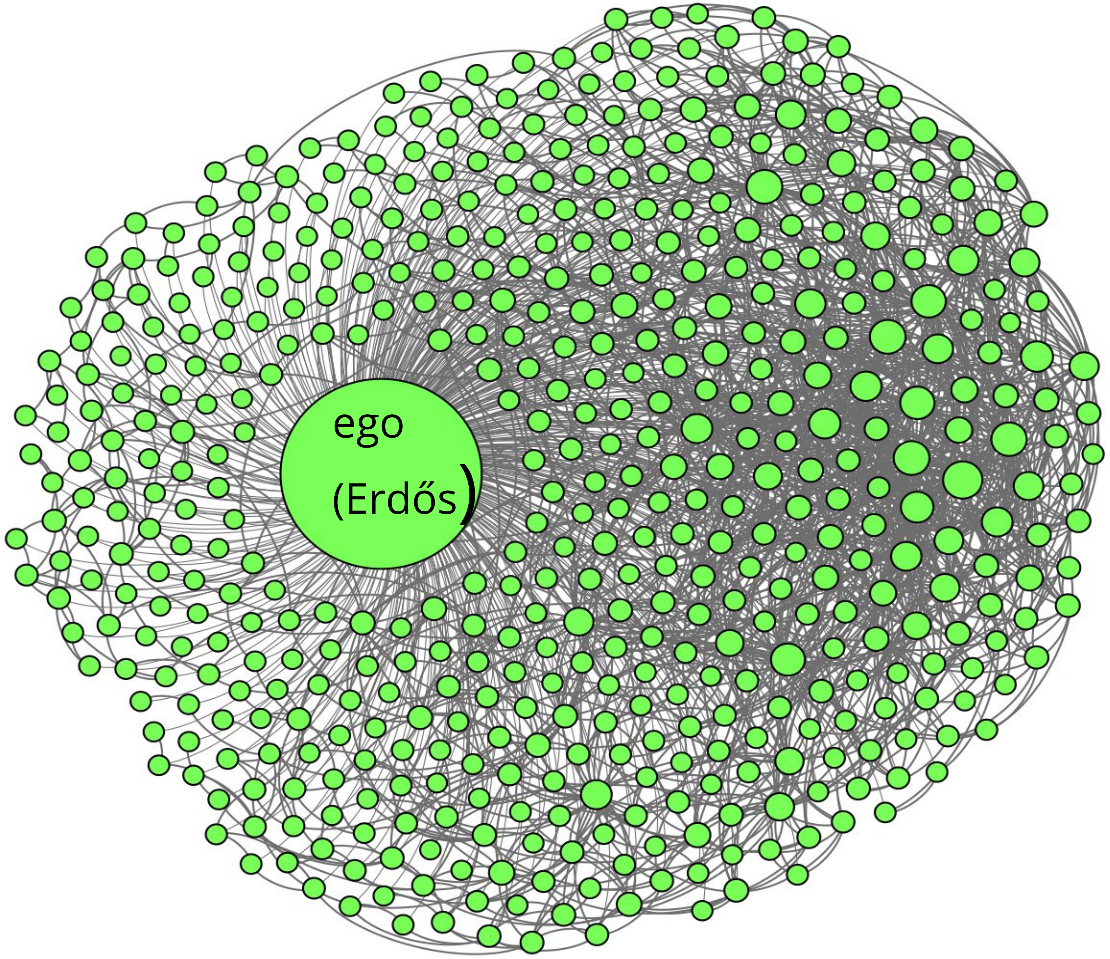
- Paul Erdős was a prolific Hungarian mathematician known for his extensive contributions across various areas of mathematics
- Considered the father of graph theory together with Alfréd Rényi
- He collaborated with over 500 coauthors: a hub in the coauthorship network!

Erdős Number

- The Erdős number represents the collaborative distance in authorship between Paul Erdős and another researcher, as measured by the shortest path through co-author links in the coauthorship network
- For instance, if a scholar wrote a paper with Erdős, their Erdős number is 1. If they wrote with someone who wrote with Erdős, their number is 2, and so forth
- Many mathematicians are proud to have a small Erdős number
- Tool to compute one's Erdős number:

mathscinet.ams.org/mathscinet/collaborationDistance.html

Erdős ego coauthorhip network



Erdős numbers

Davis

BY EMILIO FERRARA, ONUR VAROL, CLAYTON DAVIS, FILIPPO MENCZER, AND ALESSANDRO FLAMMINI

The Rise of Social Bots

4

Menczer

Fortunato

Paul Erdős
(Image by Kmhkmh - CC BY 3.0 commons.wikimedia.org/w/index.php?curid=38087162)

0

Journal of Graph Theory

Article | Full Access

Highly irregular graphs†

Yousef Alavi, Gary Chartrand, F. R. K. Chung, Paul Erdös, R. L. Graham, Ortrud R. Oellermann

1

Topical interests and the mitigation of search engine bias

S. Fortunato, A. Flammini, F. Menczer, and A. Vespignani

PNAS August 22, 2006 103 (34) 12684-12689; https://doi.org/10.1073/pnas.0605525103

Communicated by Elinor Ostrom, Indiana University, Bloomington, IN, July 1, 2006 (received for review March 2, 2006)

3

Fan Chung (Image by Che Graham - CC BY 3.0 commons.wikimedia.org/w/index.php?curid=19475582)

The Workshop on Internet Topology (WIT) Report

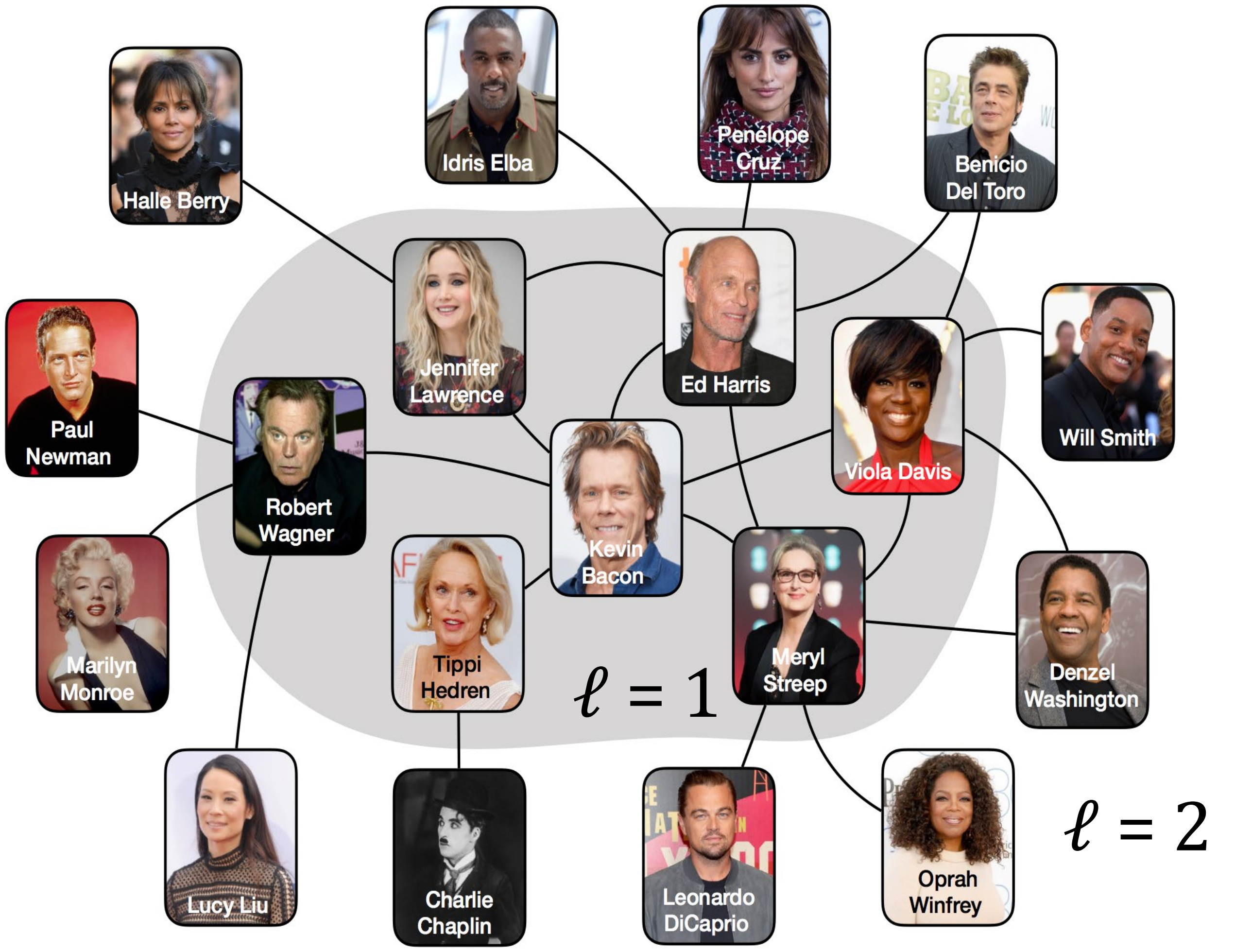
Dmitri Krioukov CAIDA dima@caida.org	Fan Chung UCSD fan@math.ucsd.edu	kc claffy CAIDA kc@caida.org
Marina Fomenkov CAIDA marina@caida.org	Alessandro Vespignani Indiana University alexv@indiana.edu	Walter Willinger AT&T Research walter@research.att.com

2

Alessandro Vespignani - Image CC BY-SA 2.0 commons.wikimedia.org/w/index.php?curid=60299947

Six Degrees of Kevin Bacon

- Short paths are found among all authors, not just Erdős…
  - .. and in **all** social networks, not just coauthorship
- Consider the movie co-star network as a second example
- Let's play the Oracle of Bacon game: oracleofbacon.org
  - Not just Kevin Bacon…
  - Can you find two stars separated by more that four links?



Six Degrees of Seperation

- What have we observed? Social networks tend to exhibit very short paths between nodes
- The concept of six degrees of separation suggests that any two people in a social network are connected through at most six intermediate steps
- The idea first appeared in the 1929 short story “Chains” by Hungarian author Frigyes Karinthy
- In 1967, psychologist Stanley Milgram provided empirical support through an influential experiment measuring social distance between randomly selected individuals in the United States
- The phrase “six degrees of separation” was later popularized by John Guare in his 1991 play (and subsequent film adaptation)

7. Small Worlds

Social Networks

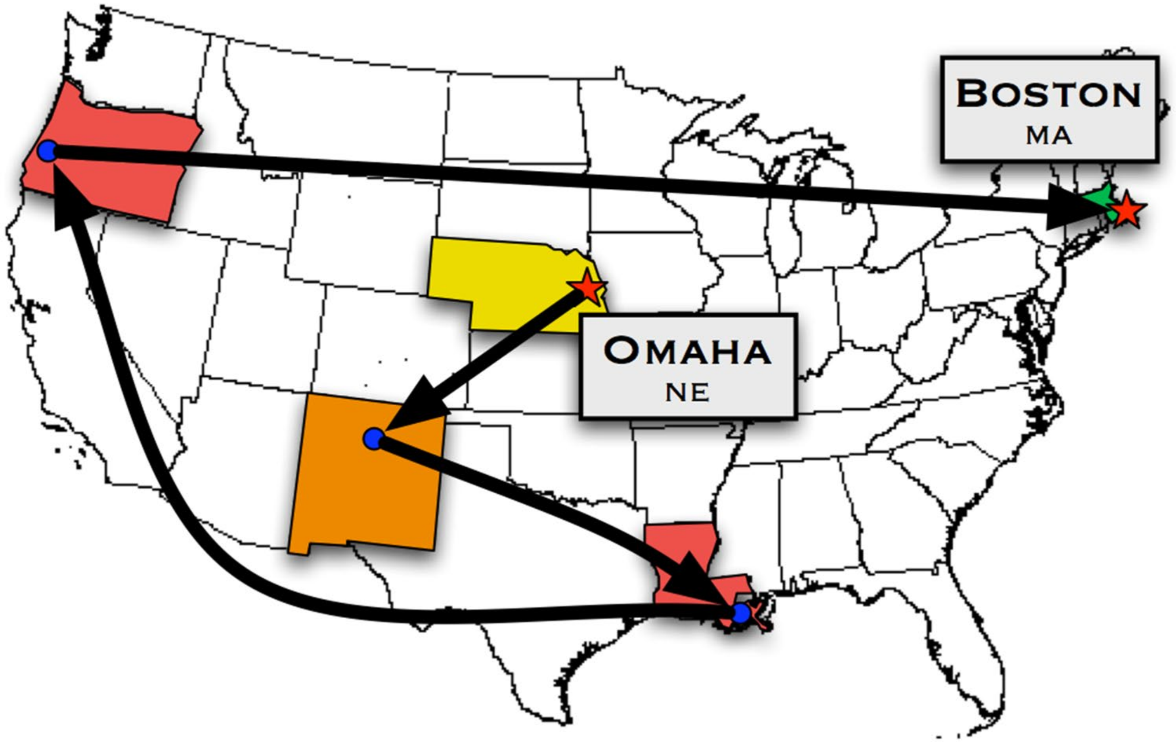
- Erdős Number and Six Degrees of Kevin Bacon
  - These examples illustrate that real-world networks—whether among mathematicians or actors—are often surprisingly small in terms of connection steps
- Social Networks and their Implications for Social Dynamics
  - The chance of encountering a friend of a friend is higher than we typically assume, highlighting the dense interconnectivity of social circles
  - This interconnectedness enables fast information dissemination and shapes social, professional, and cultural dynamics, making social networks powerful channels for communication and influence

Concept of Small Worlds

- **Common Experience:** It’s a frequent occurrence to meet someone new only to discover a mutual acquaintance, illustrating the unexpectedly short social distances within large populations
- The "small world" phenomenon suggests that social networks are characterized by short path lengths between individuals, typically quantified by the concept of "degrees of separation"

Milgram's Experiment

- Instructions: send to personal acquaintance who is more likely to know target
- 160 letters to people in Omaha, NE and Wichita, KS
- 2 targets in Mass: the wife of a student in Sharon and a stockbroker in Boston
- 42 letters reached the target(only 26%)
- Average: 6.5 steps (range: 3-12 steps)
- Much lower than most people expected!
- “Small world” effect is still surprising



More Small World Experiments

- Milgram's experiment was replicated in 2003 by Yahoo Research using email
  - 18 targets in 13 countries
  - 384 completed chains out of more than 24 thousand started
  - APL = 4 but when accounting for broken chains, estimated median PL of 5–7 steps
- Replicated by researchers at Facebook and University of Milan in 2011
  - 721 million active Facebook users
  - 69 billion friendships
  - APL = 4.74 steps: even shorter!


YAHOO! RESEARCH
SMALL WORLD EXPERIMENT

About the Experiment

The Small World Experiment is designed to test the hypothesis that anyone in the world can get a message to anyone else in just "six degrees of separation" by passing it from friend to friend. Sociologists have tried to prove (or disprove) this claim for decades, but it is still unresolved.

Now, using Facebook we finally have the technology to put the hypothesis to a proper scientific test. By participating in this experiment, you'll not only get to see how you're connected to people you might never otherwise encounter, you will also be helping to advance the science of social networks.

Become a Sender

We have already recruited a number of Target Persons from around the world.

Now we want you to try to reach them by becoming a Sender

Click on the Participate Button below, and you'll be shown your assigned target. Then you'll get to choose a friend to pass the message to. That person will then get the same instructions, and so on....

If everyone passes the messages along, your message will reach the target. How many steps will it take? There's only one way to find out.

Continue


The New York Times

Business Day
Technology

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPI

SAMSUNG The Next Big Thing Is Here GALAXY S4

Separating You and Me? 4.74 Degrees

By JOHN MARKOFF and SOMINI SENGUPTA
Published: November 21, 2011

The world is even smaller than you thought.

Enlarge This Image

Adding a new chapter to the research that cemented the phrase "six degrees of separation" into the language, scientists at Facebook and the University of Milan reported on Monday that the average number of acquaintances separating any two people in the world was not six but 4.74.

The original "six degrees" finding, published in 1967 by the psychologist Stanley Milgram,

RECOMMEND
TWITTER
LINKEDIN
SIGN IN TO E-MAIL
PRINT
REPRINTS
SHARE

Enough Sa
Now Playi

Short Paths

- What do we mean by "short paths"? When can we call a path "short"?
- It depends on the size of the network!
- Observe the relationship between APL and network size when considering networks (or subnetworks) of different sizes
- We say that the average path length is **short** when it **grows very slowly** with the size of the network, say, logarithmically:

$$\langle \ell \rangle \sim \log N$$

Small Worlds

- Many other types of networks are small worlds, too
- Air transportation networks, the Internet, the Web, and Wikipedia, all have short paths
  - Play Wikiracing games to convince yourself
  - Example: The Wiki Game (thewikigame.com)
- Most real-world networks are small worlds
- Exceptions: grid-like networks

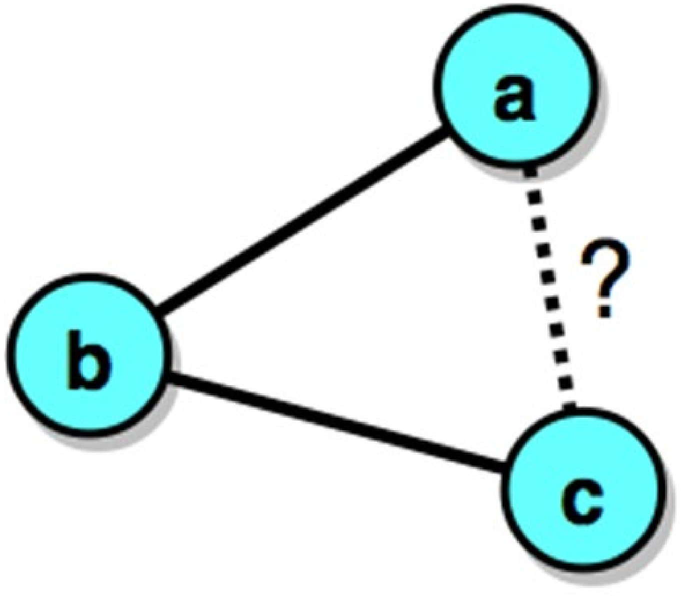
Table 2.1 Average path length and clustering coefficient of various network examples. The networks are the same as in Table 1.1, their numbers of nodes and links are listed as well. Link weights are ignored. The average path length is measured only on the giant component; for directed networks we consider directed paths in the giant strongly connected component. To measure the clustering coefficient in directed networks, we ignore link directions.

Network	Nodes (N)	Links (L)	Average path length ($\langle \ell \rangle$)	Clustering coefficient (C)
Facebook Northwestern Univ.	10,567	488,337	2.7	0.24
IMDB movies and stars	563,443	921,160	12.1	0
IMDB co-stars	252,999	1,015,187	6.8	0.67
Twitter US politics	18,470	48,365	5.6	0.03
Enron Email	87,273	321,918	3.6	0.12
Wikipedia math	15,220	194,103	3.9	0.31
Internet routers	190,914	607,610	7.0	0.16
US air transportation	546	2,781	3.2	0.49
World air transportation	3,179	18,617	4.0	0.49
Yeast protein interactions	1,870	2,277	6.8	0.07
C. elegans brain	297	2,345	4.0	0.29
Everglades ecological food web	69	916	2.2	0.55

8. Friend of a Friend

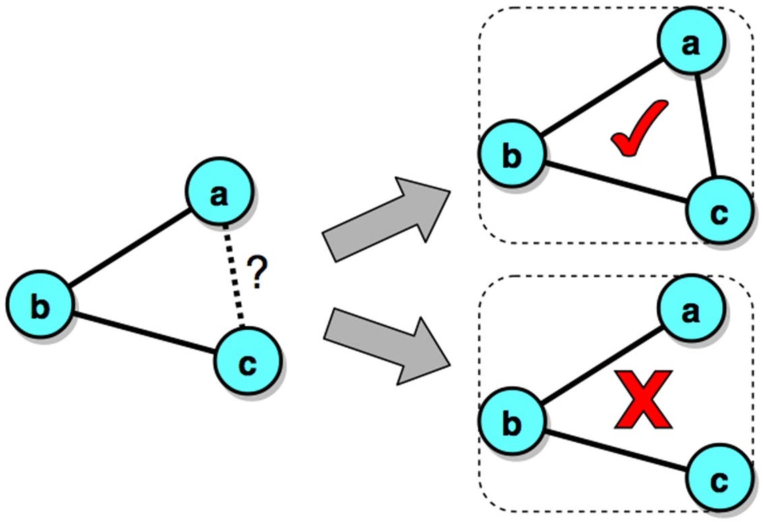
Friend of a Friend (1/3)

- Another feature of social (and some other) networks is the presence of **triangles**: if Alice (A) and Bob (C) are both friends with Charlie (B), they are also likely friends of each other
- In other words, many friends of my friends are also my friends



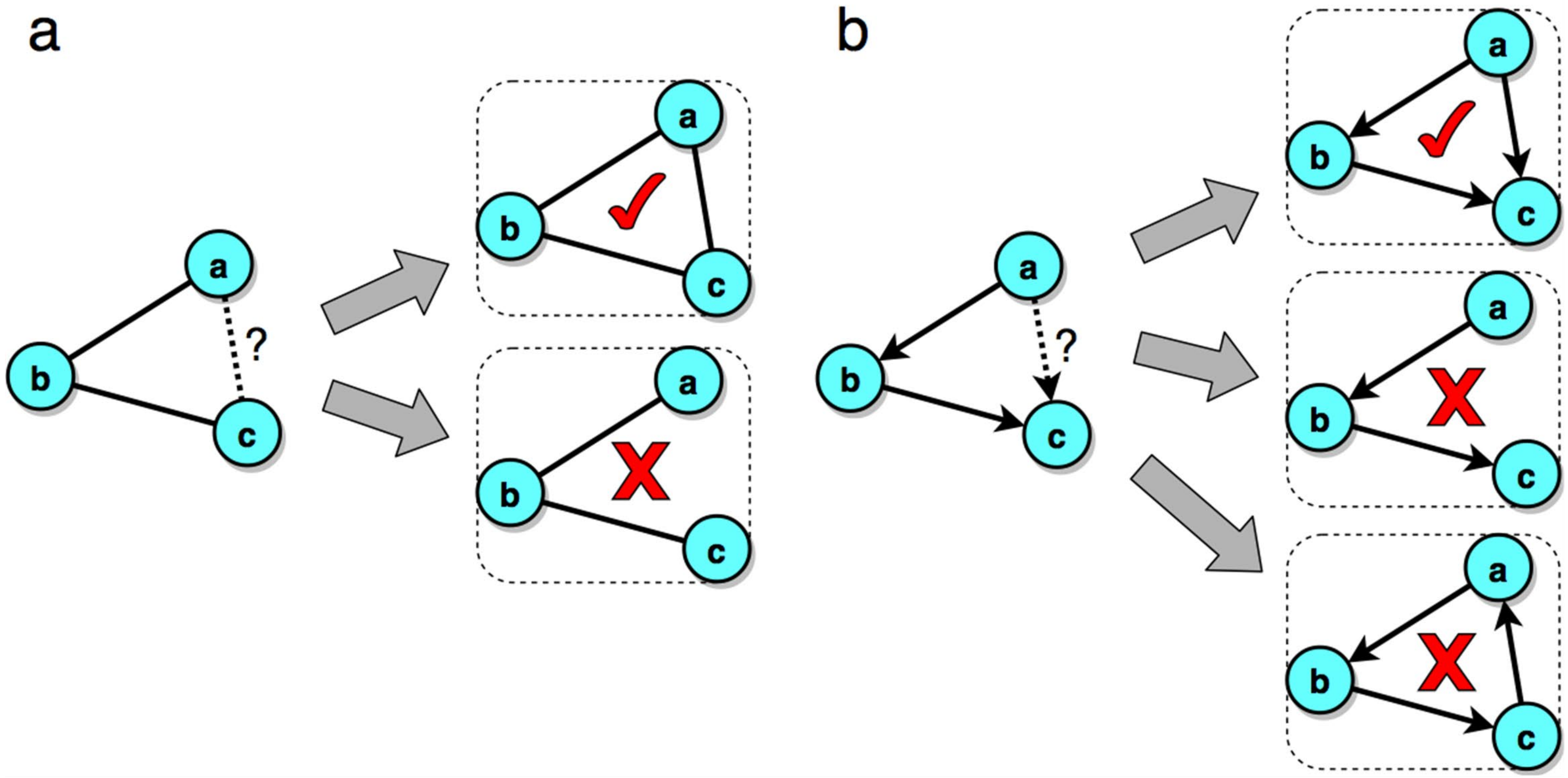
Friend of a Friend (2/3)

- A triangle is a triad (set of three nodes) where each pair of nodes is connected
- The connectivity among the neighbors of the nodes is crucial for understanding the local structure of a network
- This connectivity highlights how tightly knit or *clustered* the nodes are within the network



Friend of a Friend (3/3)

- In directed networks, we can consider only certain types of directed triangles, like shortcuts in Twitter



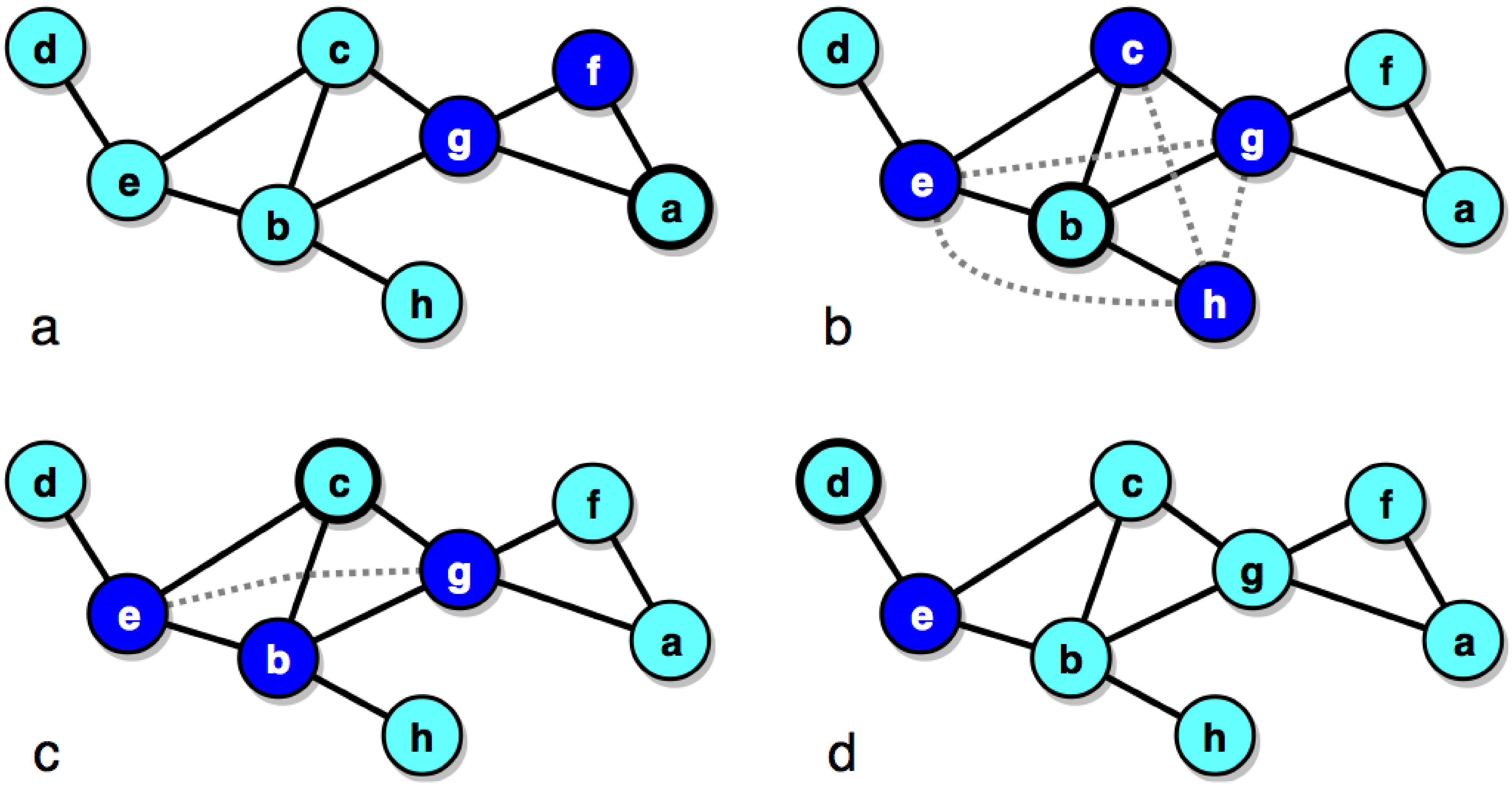
Clustering Coefficient

- We can measure the number of triangles that a node actually has relative to how many it could have
- The **clustering coefficient** of a node is the **fraction of pairs of the node's neighbors that are connected to each other:**

$$C(i) = \frac{\tau(i)}{\tau_{max}(i)} = \frac{\tau(i)}{\binom{k_i}{2}} = \frac{2\tau(i)}{k_i(k_i - 1)}$$

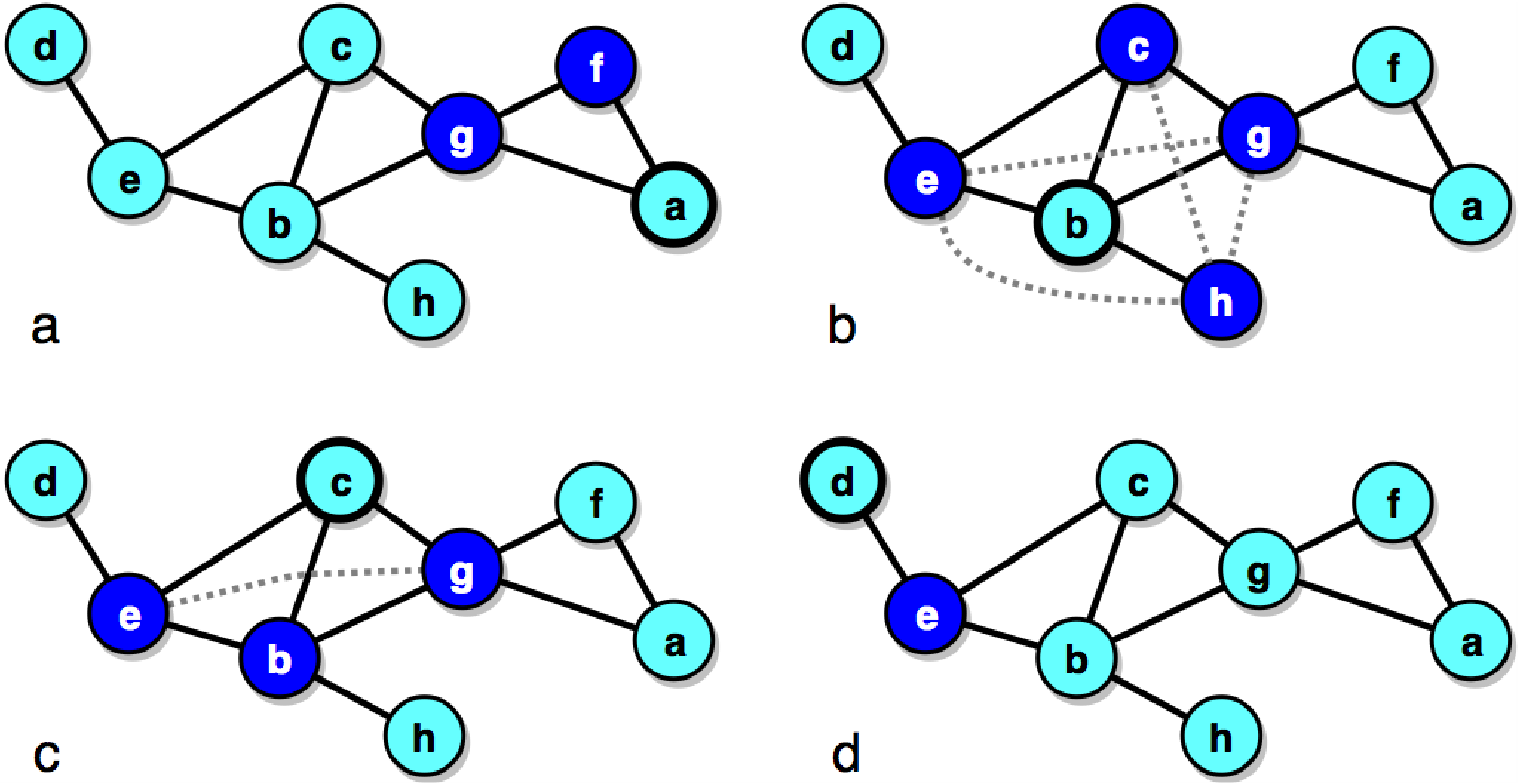
where $\boldsymbol{\tau}(i)$ tau is the number of triangles involving *i*. Note that in this definition, the clustering coefficient is undefined if $k_i < 2$: a node must have at least degree 2 to have any triangles. However NetworkX assumes C=0 if k=0 or k=1.

Clustering Coefficient Exercises (1/2)



Clustering Coefficient Exercises (2/2)

$C(a) = \frac{1}{1}$



$C(b) = \frac{2}{6}$

$C(c) = \frac{2}{3}$

C(d) undefined

The Clustering Coefficient in Networks

- Measures how close nodes are to forming a complete triangle (triadic closure) within their neighborhood
- Averaging this measure across all nodes gives an overall clustering coefficient for the network
- Low Coefficient (near zero): Few triangles exist, indicating sparse connectivity
- High Coefficient (near one): Many triangles exist, indicating dense connectivity

Clustering Coefficient in Social Networks (1/2)

- Typically high, indicating a strong presence of triangles
  - Example: Coauthorship networks where scholars frequently collaborate in group have clustering coefficient often above 0.5

Why?

Clustering Coefficient in Social Networks (2/2)

universität koblenz
weiter:denken

- Typically high, indicating a strong presence of triangles
  - Example: Coauthorship networks where scholars frequently collaborate in group have clustering coefficient often above 0.5

Why?

- Mechanism Behind High Clustering: Triadic Closure
  - A social mechanism where individuals tend to form ties with friends-of-friends, thus naturally closing triangles

Practical Application of Triadic Closure

- Platforms like Facebook and Twitter use algorithms based on triadic closure to recommend new connections

- This leads to an increase in the clustering coefficient as users connect with friends of friends

Network Clustering Coefficient (1/2)

- The clustering coefficient of the network is the average of the clustering coefficients of the nodes:

$$C = \frac{\sum_{i:k_i>1} C(i)}{N_{k>1}}$$

  - Again, we should exclude singletons and nodes with k=1, but NetworkX assumes those have C=0

```
nx.triangles(G)             # dict node -> no. triangles
nx.clustering(G, node)      # clustering coefficient of node
nx.clustering(G)            # dict node -> clustering coefficient
nx.average_clustering(G) # network's clustering coefficient
```

Network Clustering Coefficient (2/2)

- Some networks, e.g., social networks, tend to have high clustering coefficients because of **triadic closure**: we meet through common friends
- Other networks, e.g., bipartite and tree-like networks, have low clustering coefficient

Table 2.1 Average path length and clustering coefficient of various network examples. The networks are the same as in Table 1.1, their numbers of nodes and links are listed as well. Link weights are ignored. The average path length is measured only on the giant component; for directed networks we consider directed paths in the giant strongly connected component. To measure the clustering coefficient in directed networks, we ignore link directions.

Network	Nodes (N)	Links (L)	Average path length ($\langle \ell \rangle$)	Clustering coefficient (C)
Facebook Northwestern Univ.	10,567	488,337	2.7	0.24
IMDB movies and stars	563,443	921,160	12.1	0
IMDB co-stars	252,999	1,015,187	6.8	0.67
Twitter US politics	18,470	48,365	5.6	0.03
Enron Email	87,273	321,918	3.6	0.12
Wikipedia math	15,220	194,103	3.9	0.31
Internet routers	190,914	607,610	7.0	0.16
US air transportation	546	2,781	3.2	0.49
World air transportation	3,179	18,617	4.0	0.49
Yeast protein interactions	1,870	2,277	6.8	0.07
C. elegans brain	297	2,345	4.0	0.29
Everglades ecological food web	69	916	2.2	0.55

universität koblenz
weiter:denken

9. Summary

Summary

universität koblenz
weiter:denken

- **Assortativity**: Describes the tendency of nodes to connect with similar others, driven by homophily (like attracts like) and social influence (connections induce similarity)
- **Paths & Distances**
  - Path: Sequence of links connecting nodes
  - Path Length: Number of links in the shortest path (natural distance)
- **Tree Structures**: Acyclic, minimally connected graphs representing the simplest connected network form
  - **Connected Components**: Subnetworks where all nodes are reachable from one another. In directed networks, can be strongly or weakly connected
- **Average Path Length**
  - Indicates how "close" nodes are on average
  - Small-world networks have surprisingly short average paths (e.g., six degrees of separation)
- **Clustering Coefficient**
  - For a node: ratio of existing to possible triangles
  - For a network: average across all nodes
  - Social networks often have high clustering due to "friend-of-a-friend" connections

References

[1] Menczer, F., Fortunato, S., & Davis, C. A. (2020). A First Course in Network Science Cambridge: Cambridge University Press.

- Chapter 2 Small Worlds

[2] OLAT course page: https://olat.vcrp.de/url/RepositoryEntry/4669112833

Further Readings

- Samoylenko, I., Aleja, D., Primo, E., Alfaro-Bittner, K., Vasilyeva, E., Kovalenko, K., Musatov, D., Raigorodskii, A.M., Criado, R., Romance, M. and Papo, D., 2023. Why are there six degrees of separation in a social network?. Physical Review X, 13(2), p.021032.
- Zhang, Z., Zhou, F. and Ning, H., 2022. When Six Degrees of Separation Meets Online Social Networks: How Low Can the Degree Be?. arXiv preprint arXiv:2209.06577.