

Graph theory: Connectivity

Introduction to Network Science

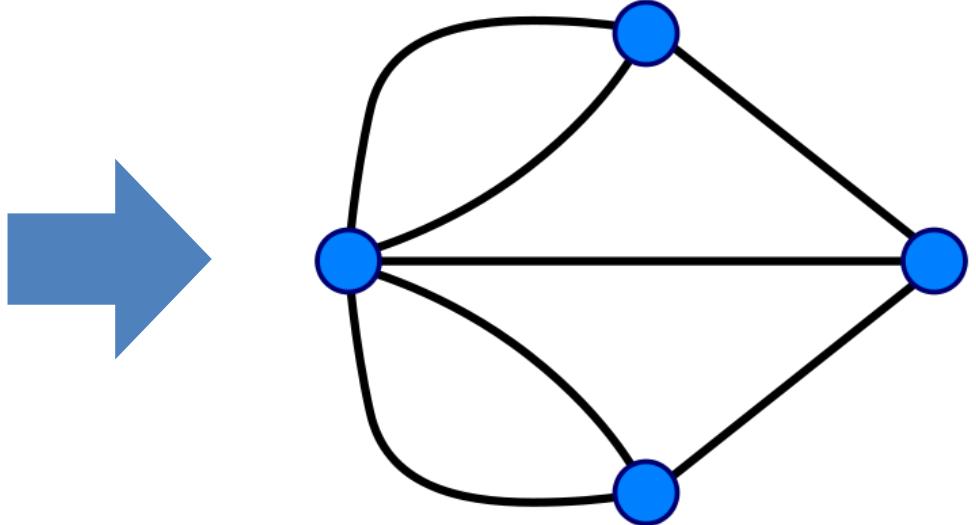
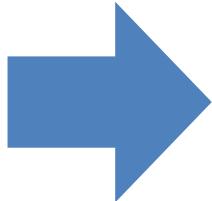
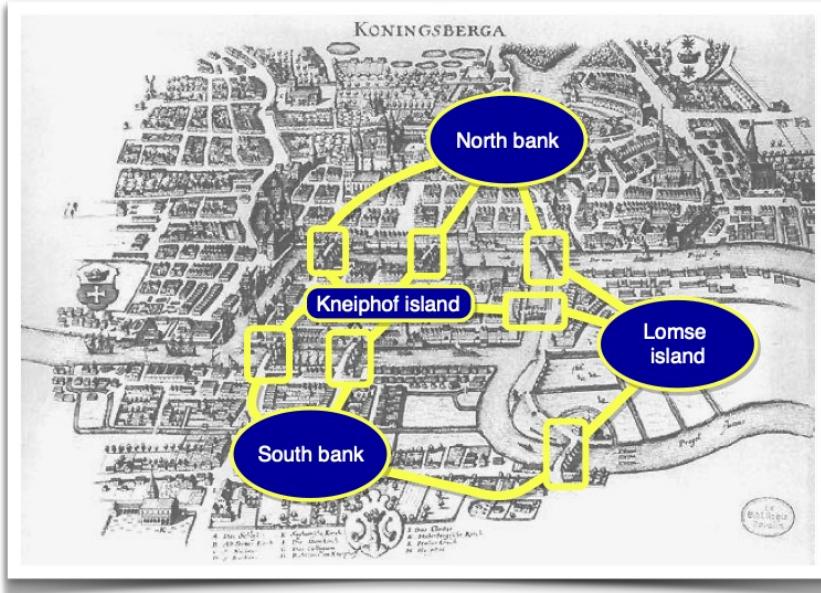
Instructor: Michele Starnini — <https://github.com/chatox/networks-science-course>

Contents

- Paths and distances
- Connected components
- Small worlds

Paths and distances

Euler circa 1736: Koningsberg bridges

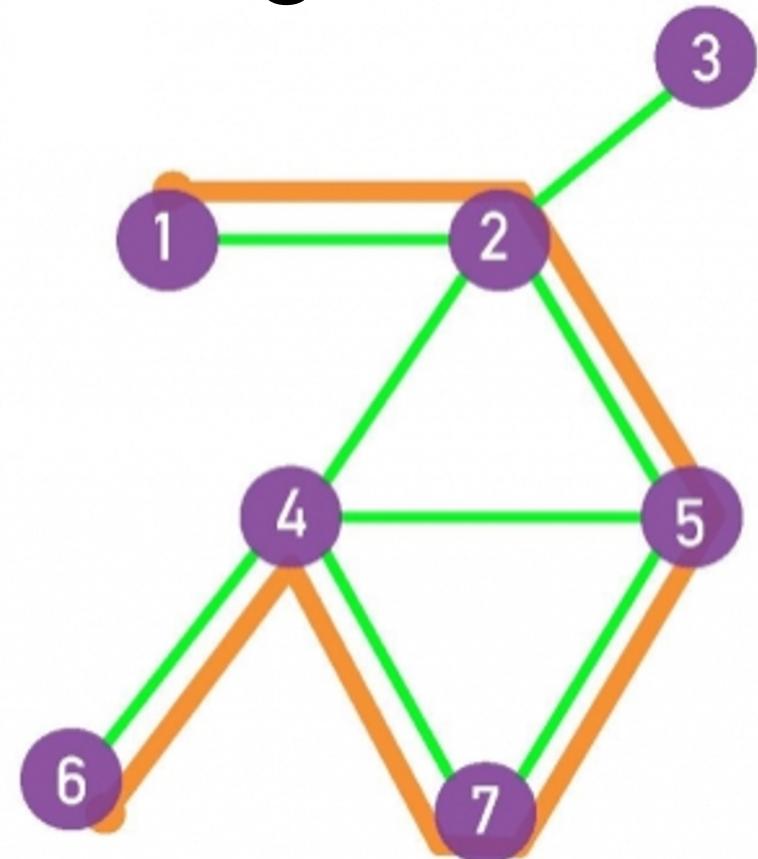


Q: Can you cross all 7 bridges just once each?

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

Paths: sequences of edges

- The destination of each edge is the origin of the next edge
 - In directed graphs, paths follow the direction of the edges
- The length of the path is the number of edges on it
 - Example: path in orange has length 5



Paths and diameter

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

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

Undirected network:

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

Paths and diameter

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)

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

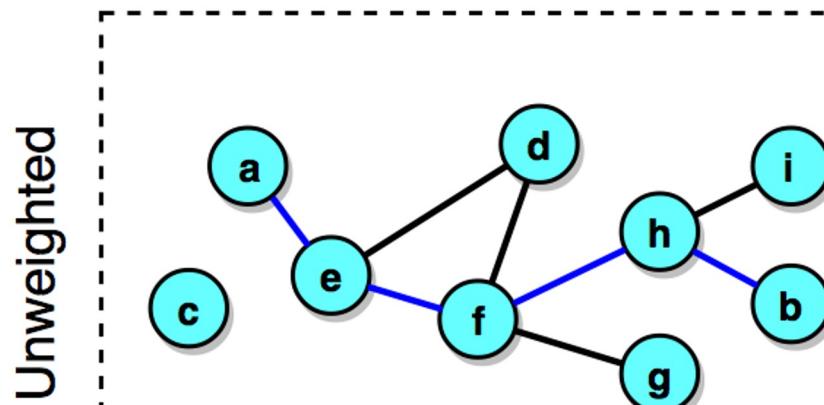
We can use a mathematical trick:

Distance

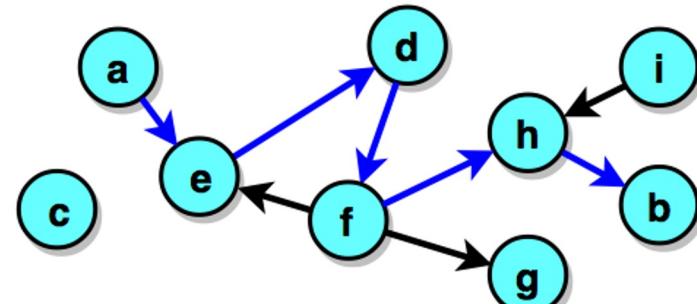
- If two nodes i, j are in the same connected component:
 - the **distance** between i and j , denoted by d_{ij} is the **length of the shortest path** between them
- If they are not in the same connected component, the distance is by definition infinite (∞)

Blue =
shortest path
between
nodes a and b

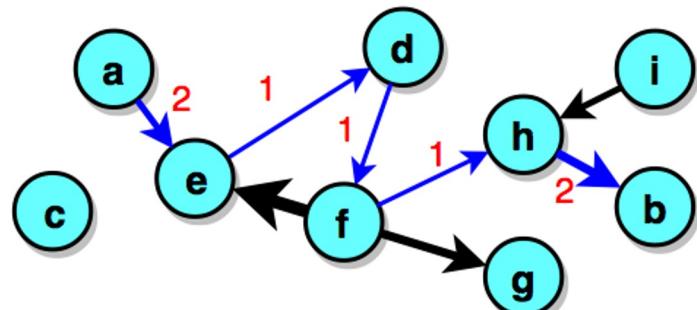
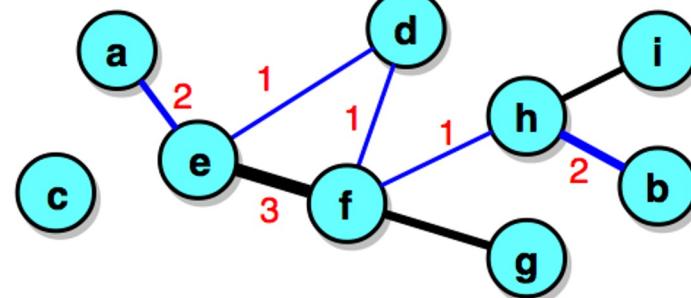
Undirected



Directed



Weighted



Diameter

- The **diameter** of a network is the maximum distance between two nodes on it, d_{\max}
- The **effective diameter** (or effective-90% diameter) is a number d such that 90% of the pairs of nodes (i,j) are at a distance smaller than d
- The **average distance** is $\langle d \rangle$, and is measured only for nodes that are in the same connected component

Connected components

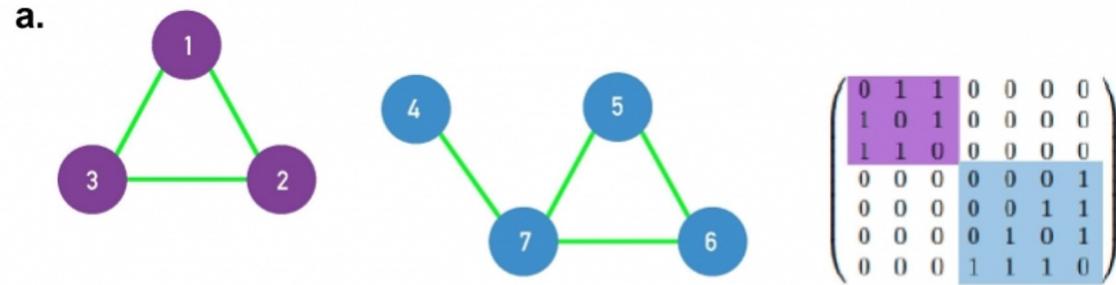
Connectedness

- If a path exists between two nodes i, j : those nodes are part of the same **connected component**
- A **connected graph** has only one connected component
- A **singleton** is a connected component with only one node

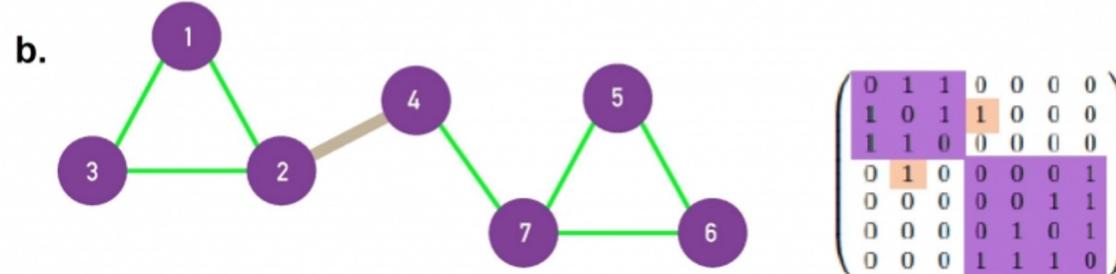
Connected graphs

A **disconnected graph** has an adjacency matrix that can be arranged in block diagonal form

a. disconnected



b. connected

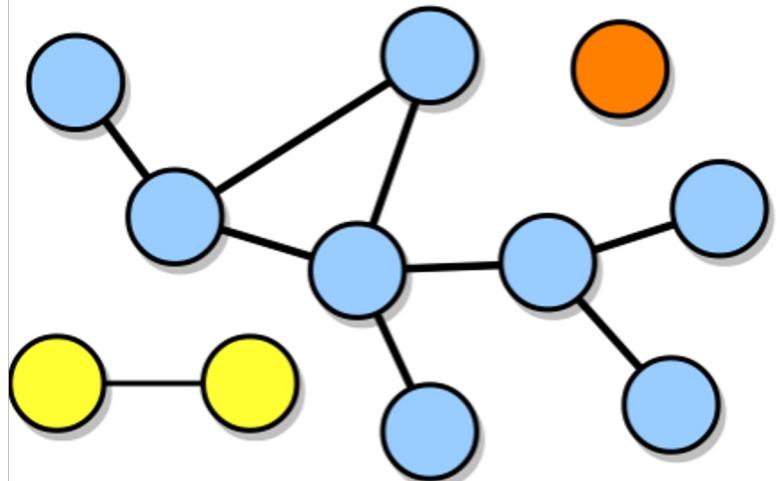


Connectedness in directed graphs

- A directed graph is **strongly connected** if it has only one connected component
- A directed graph is **weakly connected** if, when seen as undirected, has only one connected component

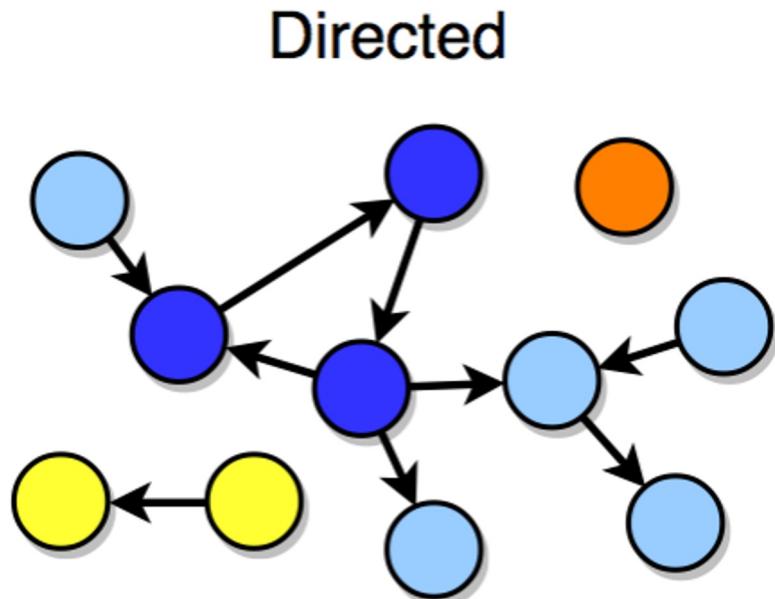
Connectedness example (undirected)

Undirected



- Is not connected
- Has 3 connected components
- One of the connected components is a singleton

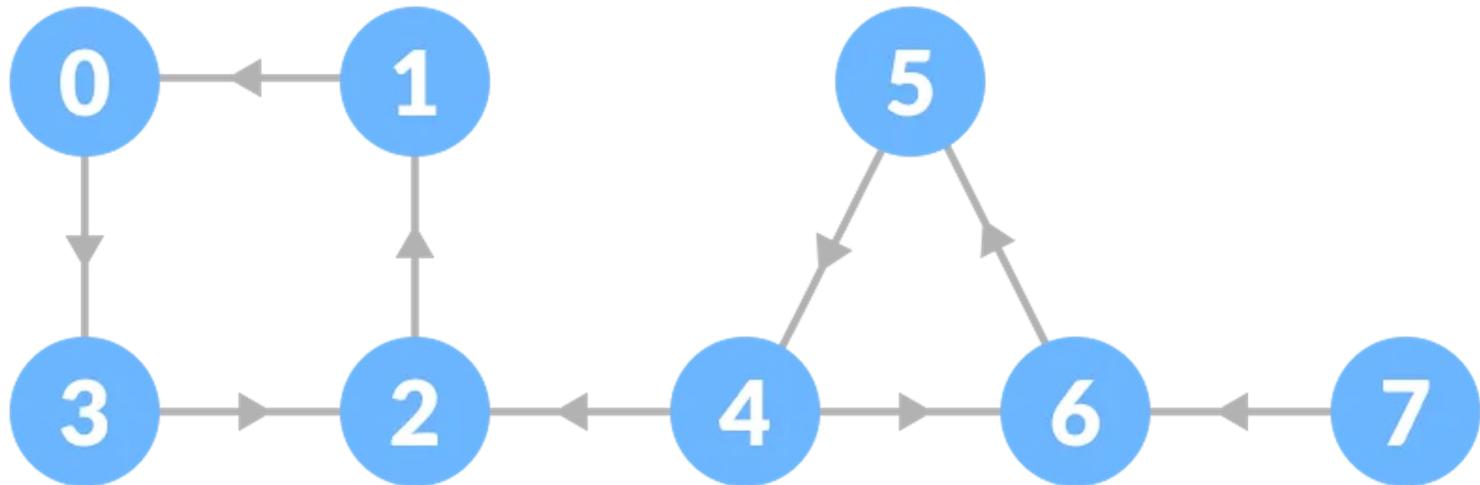
Connectedness example (directed)



- Is not strongly connected
- Is not weakly connected
- Has 3 connected components

Quick exercise

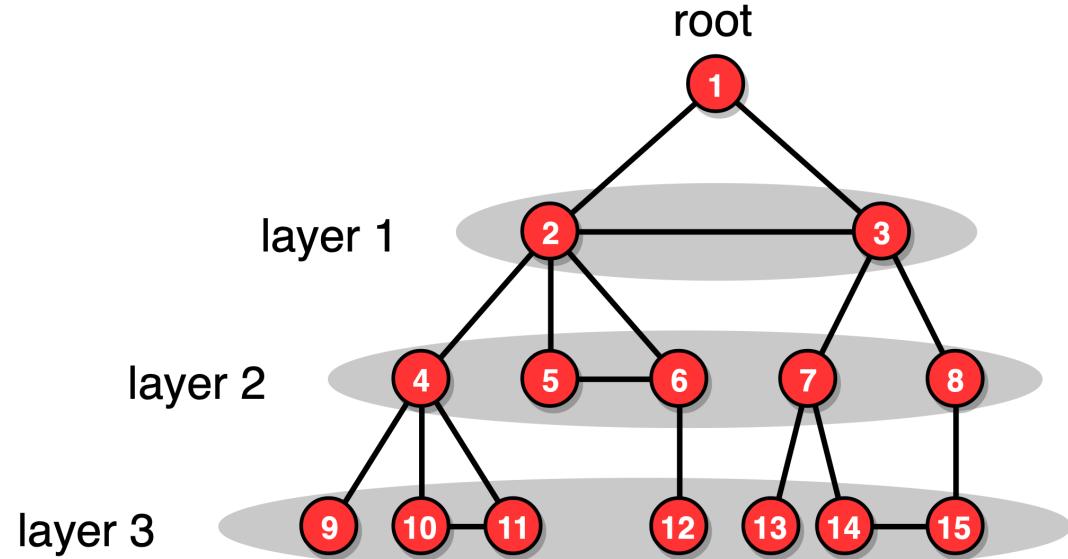
Find the strongly connected components



Finding shortest paths

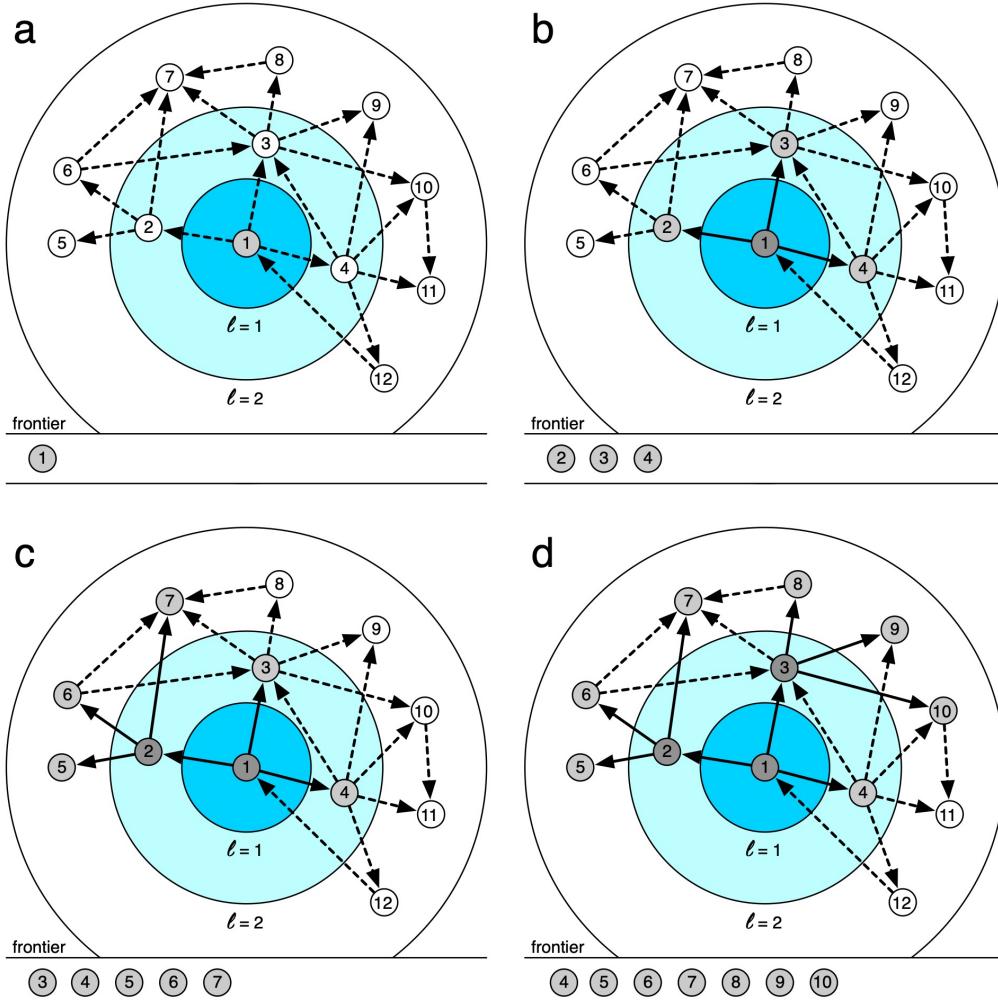
One algorithm used to find shortest paths is called **breadth-first search**

- 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: $O(N^2)$)



Breadth-first search (BFS)

- Each node has an attribute storing its **distance** ℓ from the source, initially $\ell = -1$ except $\ell(\text{source}) = 0$
- 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:
 - Remove next node i in frontier
 - For each successor j of i with $\ell(j) = -1$:
 - Queue j into frontier
 - $\ell(j) = \ell(i) + 1$
 - Add link $(i \rightarrow j)$ to shortest-path tree



Distance in real networks



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

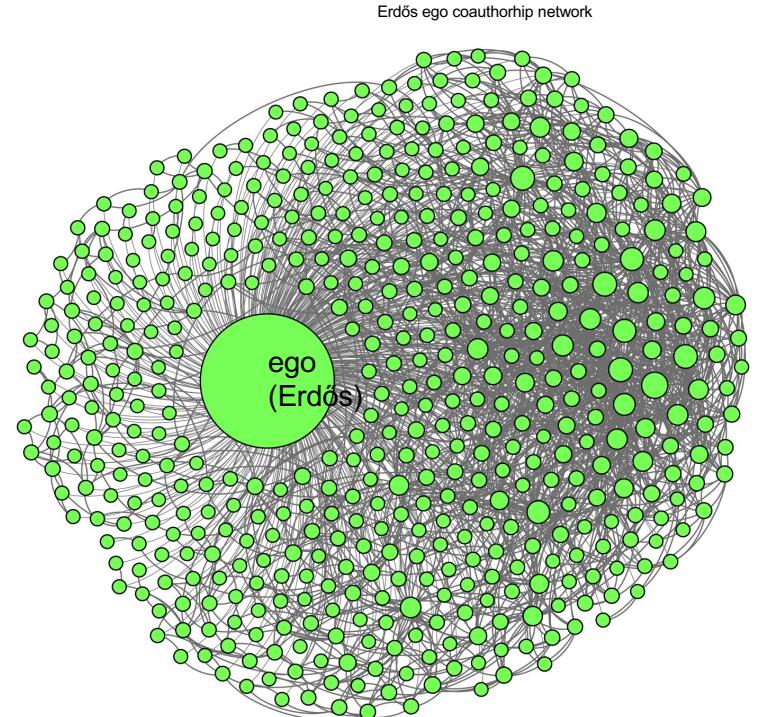
- How close or far are two nodes in a social network?
- Let us start by considering coauthorship networks, in which nodes are scholars and links represent two people having coauthored one or more publications
- Paul Erdos: Considered the father of graph theory together with Alfréd Rényi
- He collaborated with over 500 coauthors: a hub in the coauthorship network!

Erdős number

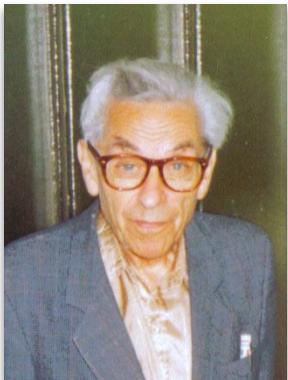
- An author's Erdős number is the length of the shortest path between them and Erdős in the coauthorship network
- Many mathematicians are proud to have a small Erdős number

Tool to compute one's Erdős number:

mathscinet.ams.org/mathscinet/collaborationDistance.html



Erdős numbers



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

0



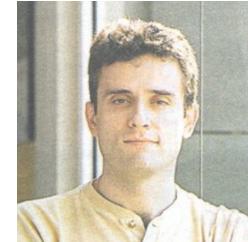
Topological properties of a time-integrated activity driven network

Michele Starnini and Romualdo Pastor-Satorras

Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord B4, 08034 Barcelona, Spain

(Dated: September 27, 2018)

4



3

Journal of Graph Theory

Article | Full Access

Highly irregular graphs[†]

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

1



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
Marina Fomenkov
CAIDA
marina@caida.org

Fan Chung
UCSD
fan@math.ucsd.edu
Alessandro Vespignani
Indiana University
alexv@indiana.edu

kc claffy
CAIDA
kc@caida.org
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=6029947

VOLUME 86, NUMBER 14

PHYSICAL REVIEW LETTERS

2 APRIL 2001

Epidemic Spreading in Scale-Free Networks

Romualdo Pastor-Satorras¹ and Alessandro Vespignani²

¹Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord, Mòdul B4, 08034 Barcelona, Spain

²The Abdus Salam International Centre for Theoretical Physics (ICTP), P.O. Box 586, 34100 Trieste, Italy

(Received 20 October 2000)

Six Degrees of Kevin Bacon

Short paths are found among all authors, not just Erdős...

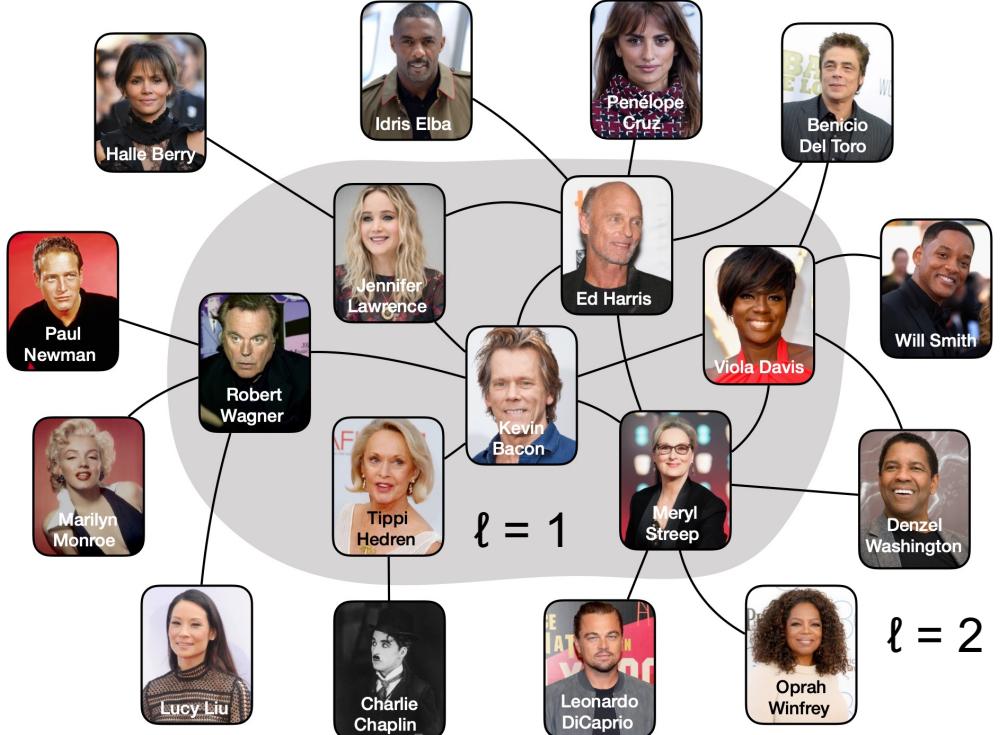
...And in all social networks, not just coauthorship

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 than four links? Play the game and try!

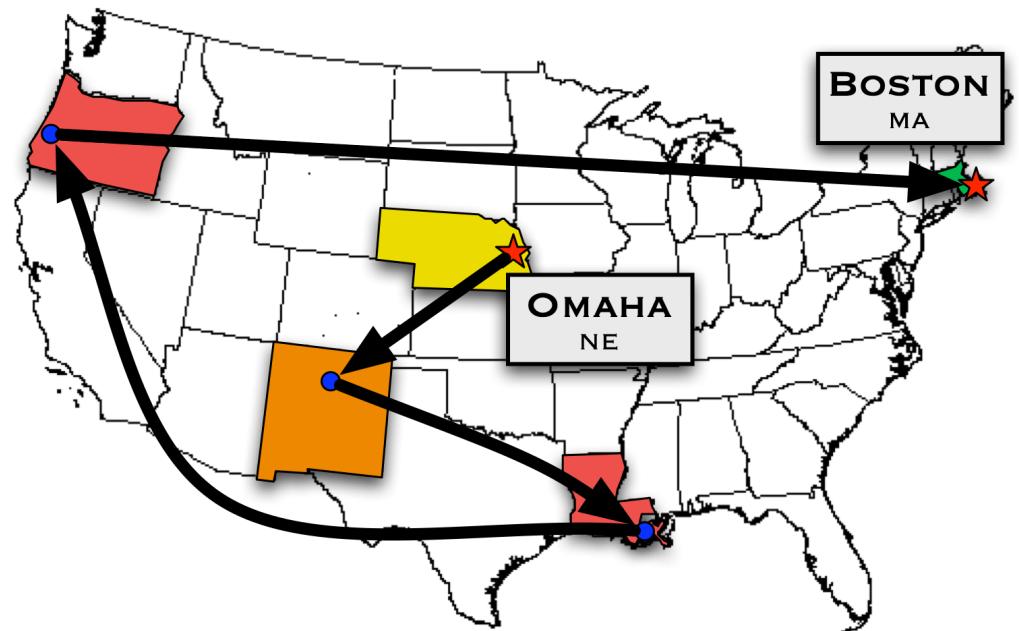


Small worlds

- What have we learned? Social networks tend to have very **short paths**
- **Six degrees of separation:** the idea that any two people are at most six steps away from each other in the social network
- First idea was in the short story “Chains” by Hungarian writer Frigyes Karinthy in 1929
- Psychologist **Stanley Milgram** provided first evidence in 1967 through a famous experiment to measure the **social distance** between any two people in the US
- John Guare coined the term “six degrees of separation” in a 1991 play (movie, too)

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 made it back (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!

The screenshot shows the Yahoo! Research Small World Experiment homepage. At the top, the Yahoo! logo is displayed above the text "RESEARCH" and "SMALL WORLD EXPERIMENT". Below this is a map of the world where several orange dots represent participants. To the right, a section titled "About the Experiment" provides information about the hypothesis of six degrees of separation and the goal of the experiment. It mentions that the claim has been unresolved for decades. A paragraph explains the use of Facebook to conduct the test, stating that participants will try to connect to a target person through their social network. A "Continue" button is visible at the bottom right of this section. On the far right, a sidebar titled "Become a Sender" encourages users to recruit friends to participate.

The screenshot shows a news article from The New York Times' Business Day Technology section. The headline reads "Separating You and Me? 4.74 Degrees". The article is by John Markoff and Somini Sengupta, published on November 21, 2011. The text discusses the replication of the small world experiment by Facebook and the University of Milan, finding an average path length of 4.74 steps. A photo of a smiling man with glasses is included. To the right of the article is a sidebar with social sharing options (Facebook, Twitter, LinkedIn, etc.) and a link to "Enough Said Now Playing".

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

Small worlds Networks examples:

Air transportation, the Internet, the Web, and Wikipedia.

Exceptions: grid-like networks

Most real, complex networks are small worlds!

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

Summary

Things to remember

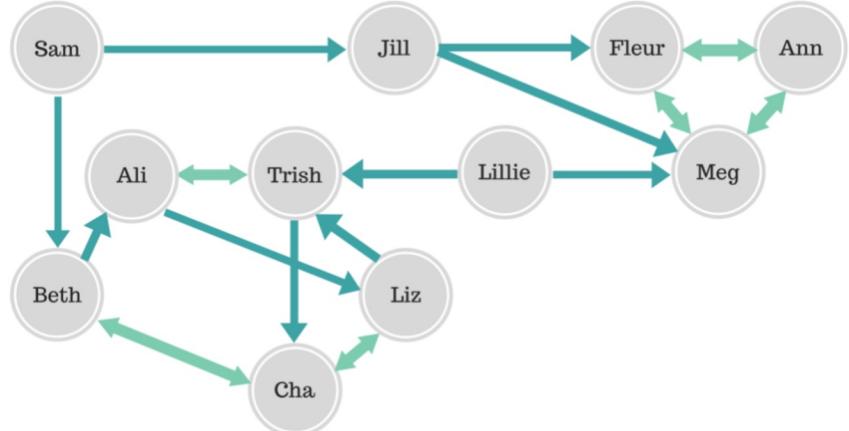
- Distance, diameter, effective diameter
 - In directed and undirected graphs
- Connected components
 - In directed and undirected graphs
 - d graphs
- Small words networks
 - Milgram's experiment
 - Examples of MW networks

Sources

- A. L. Barabási (2016). Network Science -
[Chapter 01](#)
- F. Menczer, S. Fortunato, C. A. Davis (2020). A First Course in Network Science – Chapter 02
- URLs cited in the footer of specific slides

Practice on your own

- Measure the sparsity of this graph L/L_{\max}
- (ignore direction of links)



Practice on your own (cont.)

- Compute the distance between two nodes
- Compute the diameter of a graph
- Identify connected components