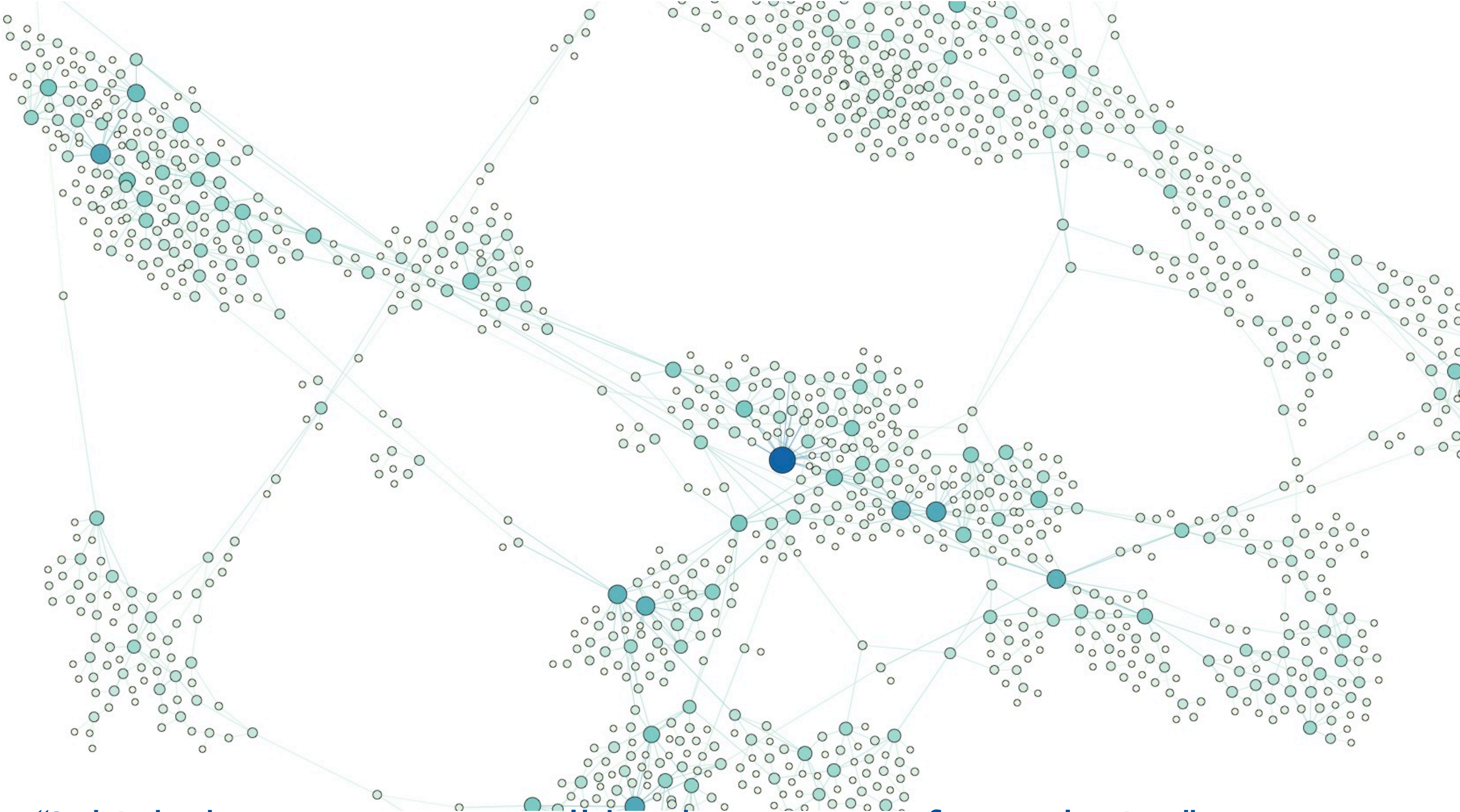
COMP6234: Network visualisation

Nora Alrajebah, Leslie Carr



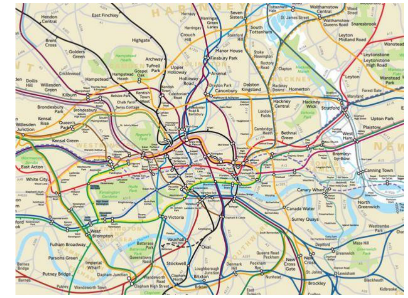
"I think the next century will be the century of complexity."

STEPHEN HAWKING

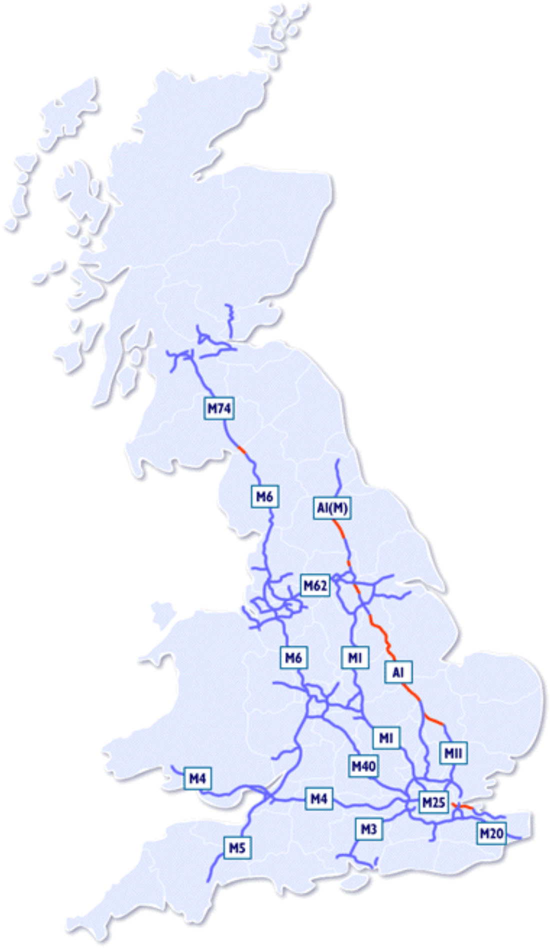
Connectedness

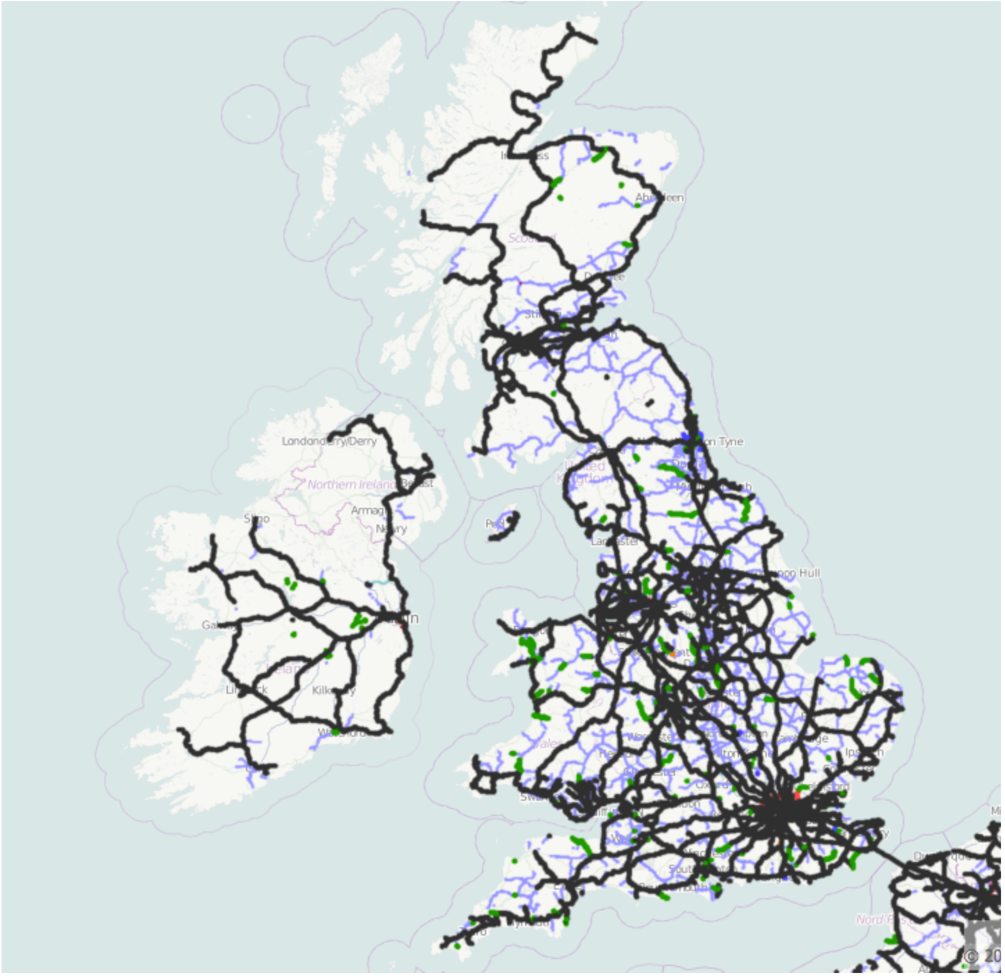
- There has been a great interest in the **complex connectedness** of the modern society
- The interconnected structures known as **networks** have been utilised in many contexts as a way of representing and visualising such complexity

London Underground

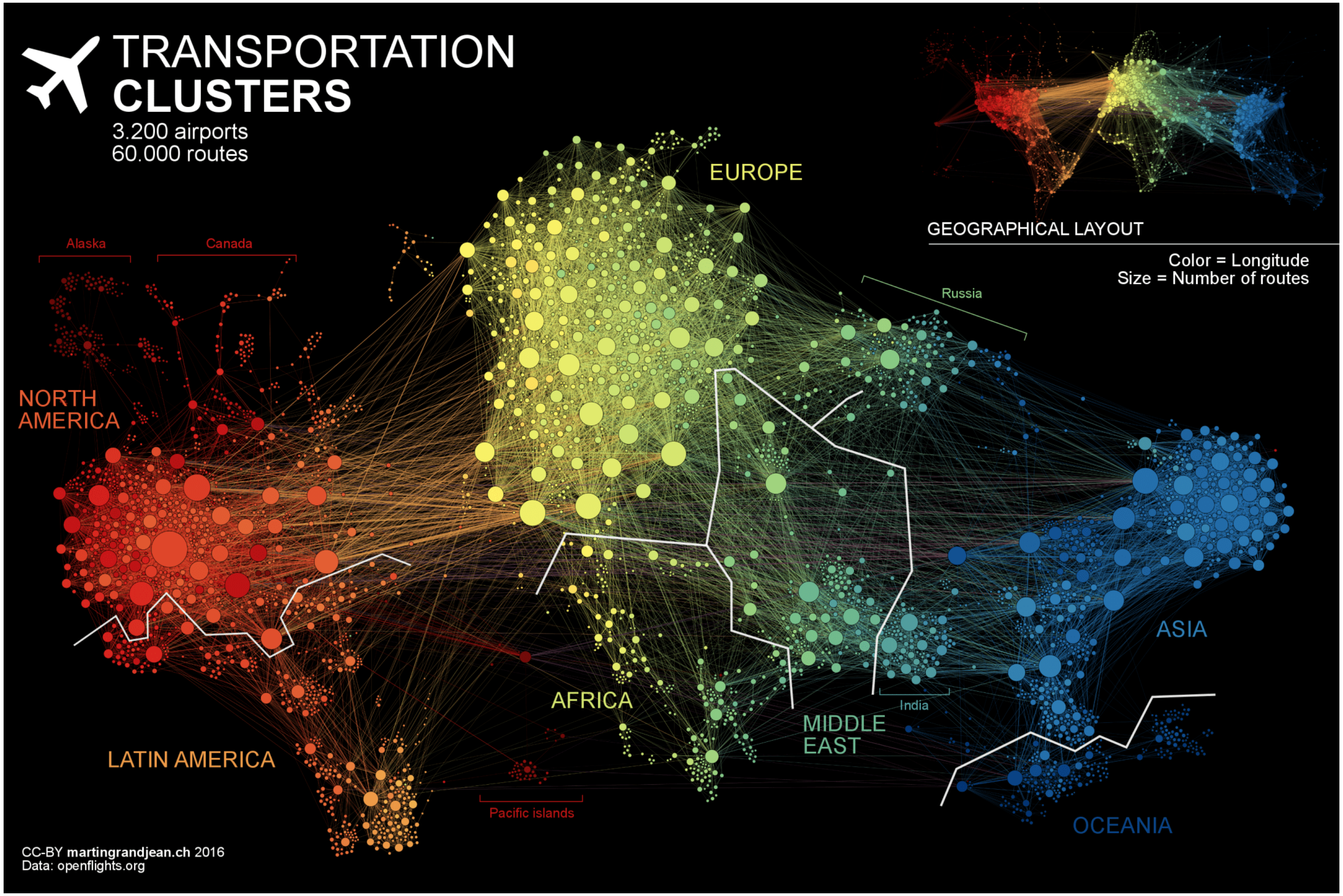


Motorway and Train Networks





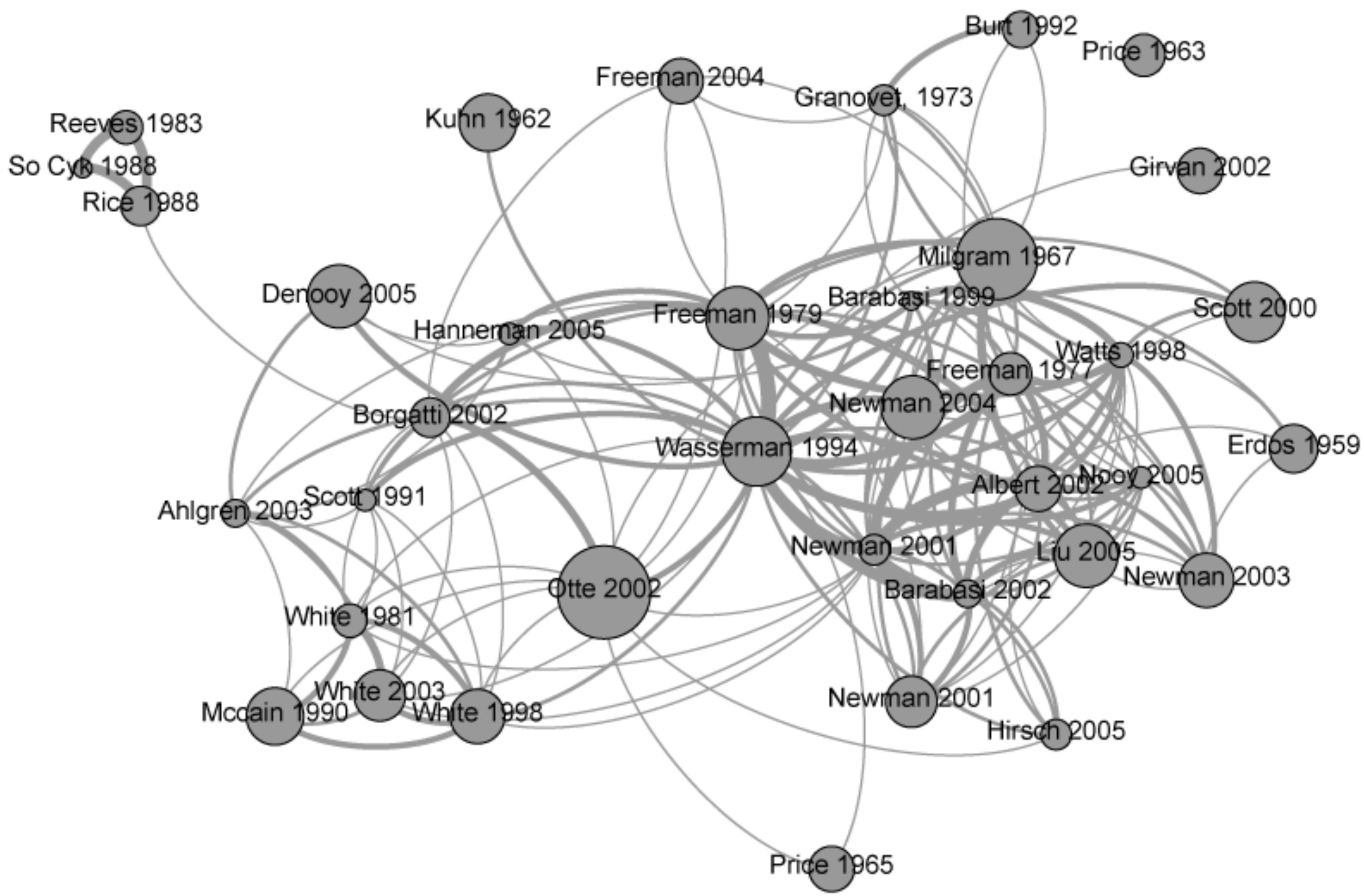
Air transport network



https://flowingdata.com/2016/05/31/air-transportation-network/

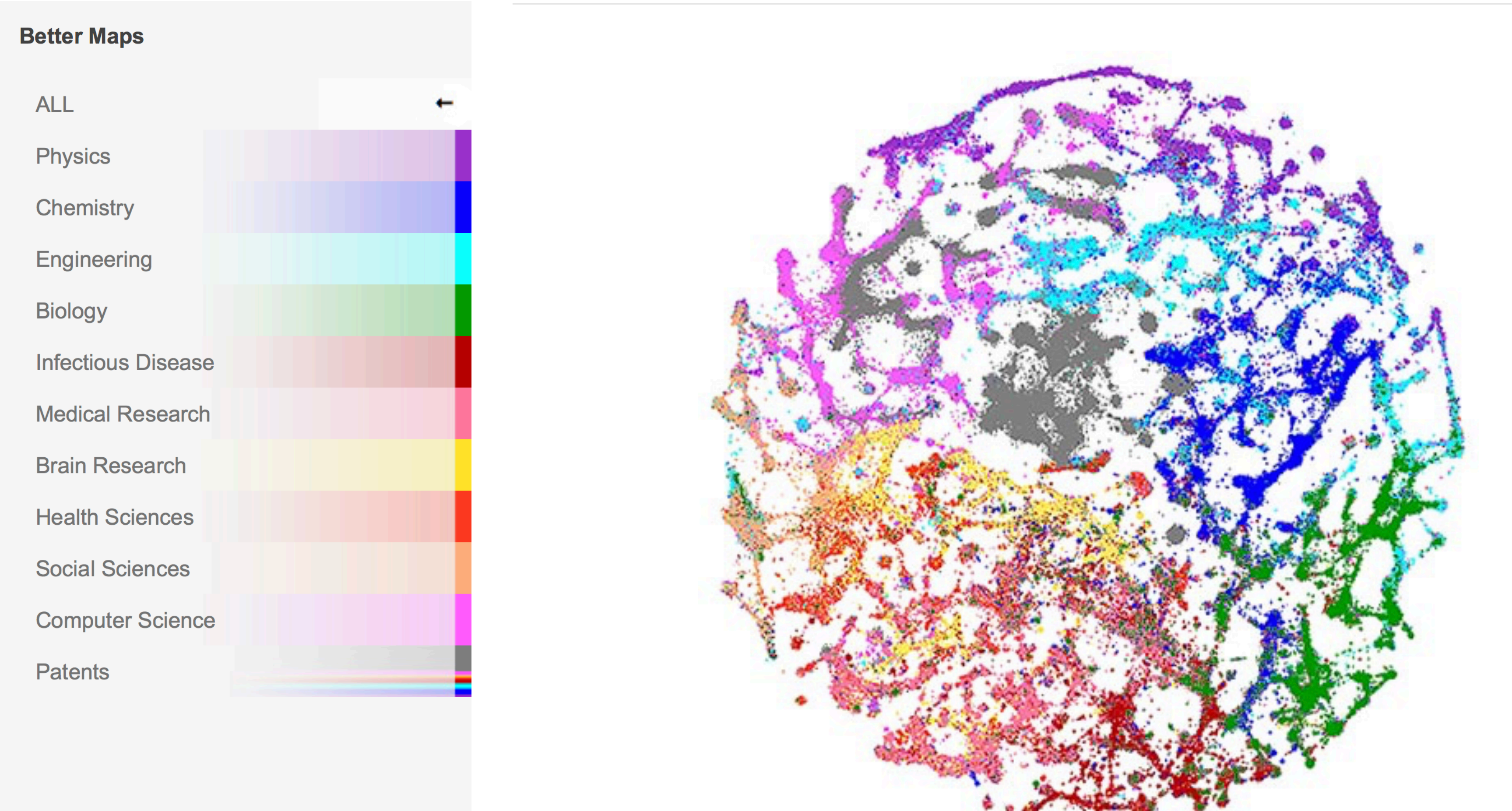
Citation network

- Co-citation analysis of the topic Social Network Analysis



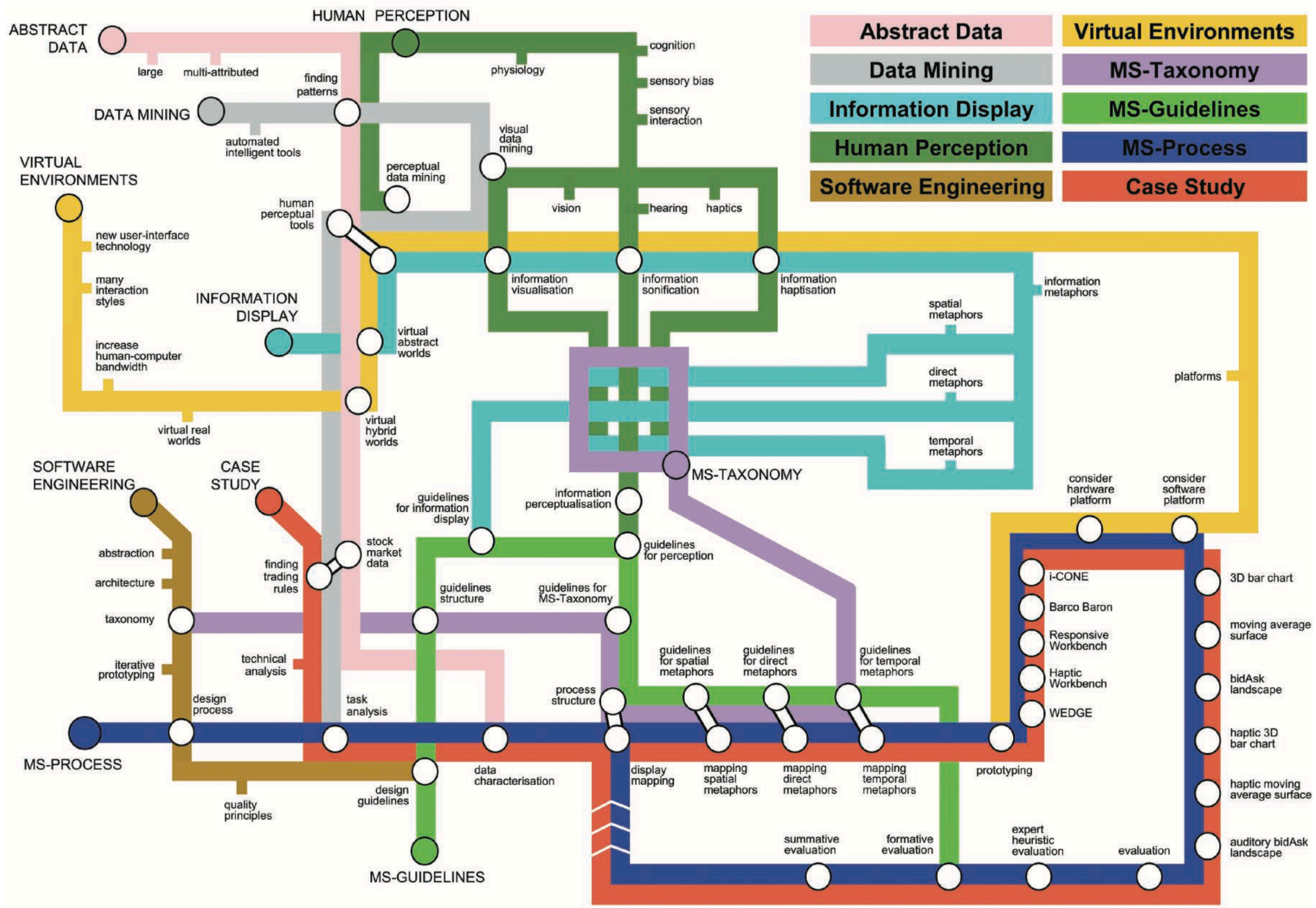
Ullmann, T. D. (2012) http://eduinf.eu/2012/03/15/co-citation-analysis-of-the-topic-social-network-analysis/

Network of Science



This map represents the relationships among and between 22 million documents – 20 million scientific and technical articles, and 2 million patents from 1996-2011. It was created by clustering the documents using a combination of citation analysis and text mining.
This map portrays the structure and dynamics of science and technology at a highly detailed level, with over 200,000 clusters of documents, each of which may last anywhere from two to sixteen years.

Network of Ideas in a PhD

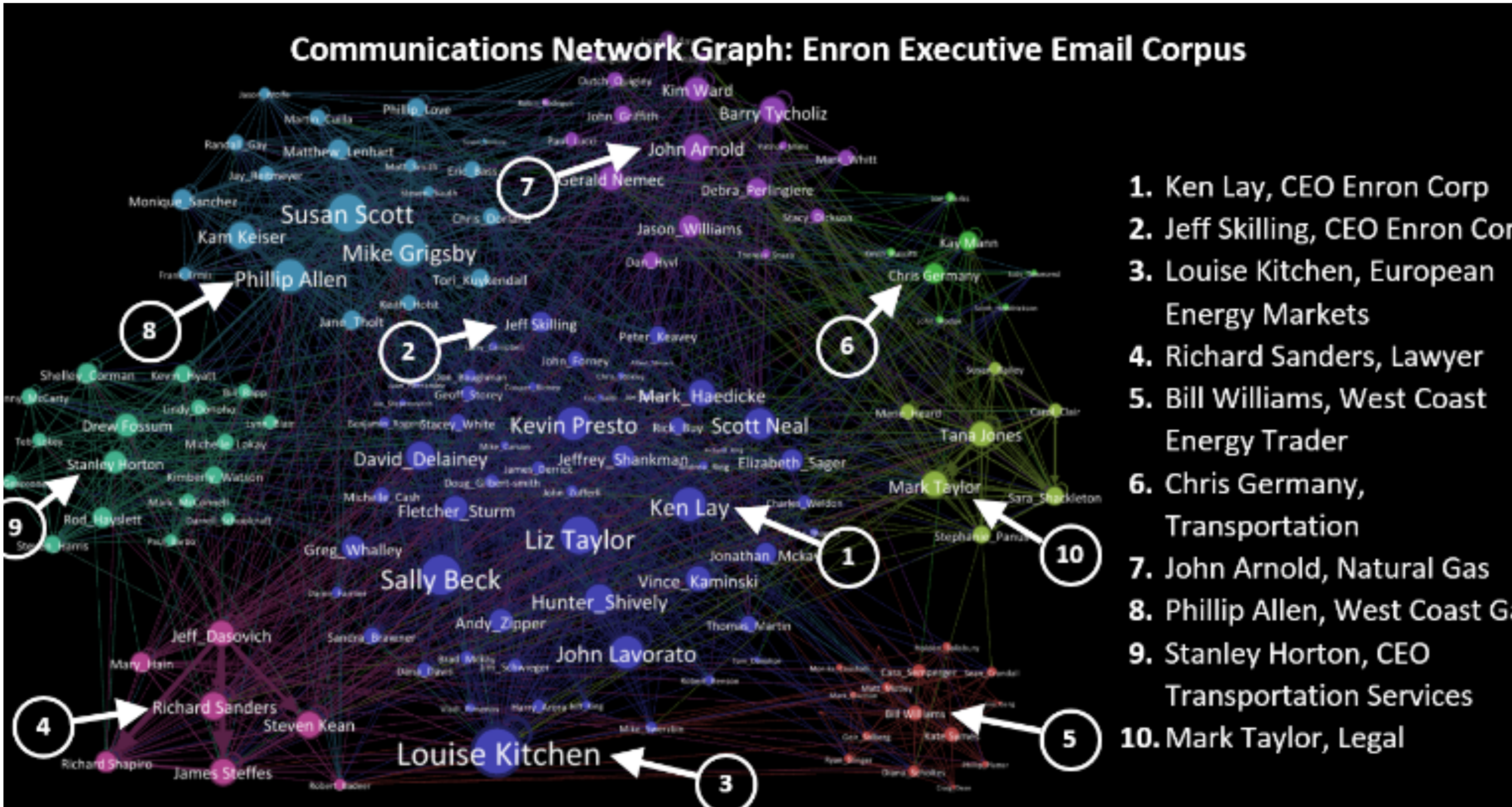


Social networks in films

- Forrest Gump 1994

Kaminski, J., Schober, M., Albaladejo, R., Zastupailo, O., Hidalgo, C. (2012). "Moviegalaxies - Social Networks in Movies", http://moviegalaxies.com

Communication networks



https://www.linkedin.com/pulse/visualizing-email-networks-damon-wade

Networks 101: Definition

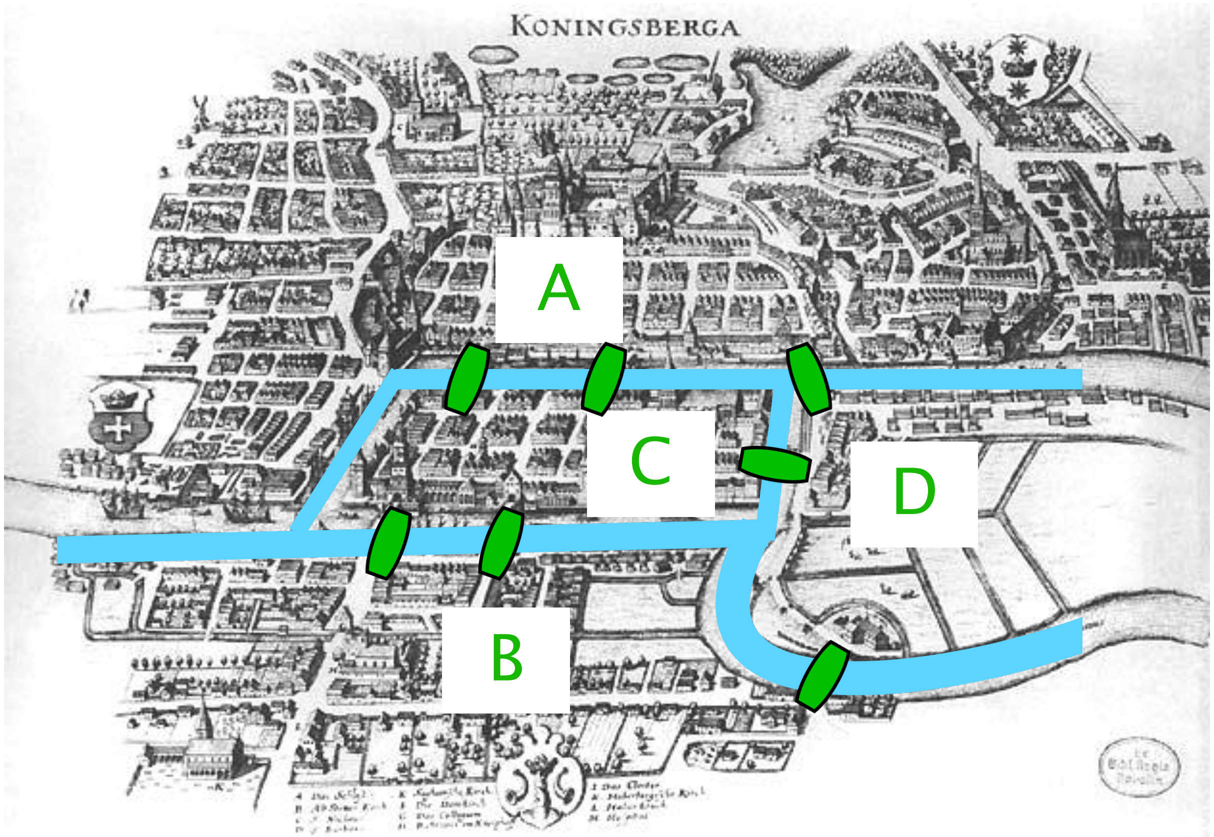
- A simple definition:
  - A network consist of a collection of **objects** where some pairs of them are connected by **links**
- The definition is **flexible**, so that it accommodates many contexts where objects can be anything: individuals, websites, usernames, companies, etc.
- Also, the links can be used to represent different types of **relationships** to associate objects with each other

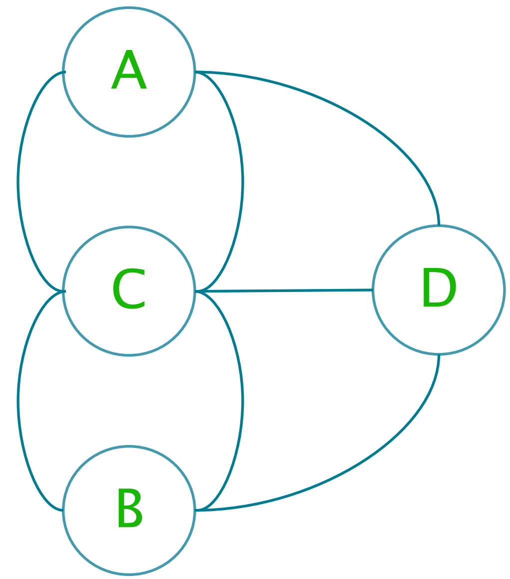
Networks 101: Examples

- There are different types of relationships between individuals that we can represent by links:
  1. Linking individuals who work with each other → **professional network**
  2. Linking individuals who co-author with each other → **collaboration network**
  3. Linking individuals who are friends with each other → **social network**
  4. Linking individuals who communicate with each other → **acquaintance network**

- Each network of the above allow us to explore different aspects of network analysis

Networks 101: The Bridges of Königsberg





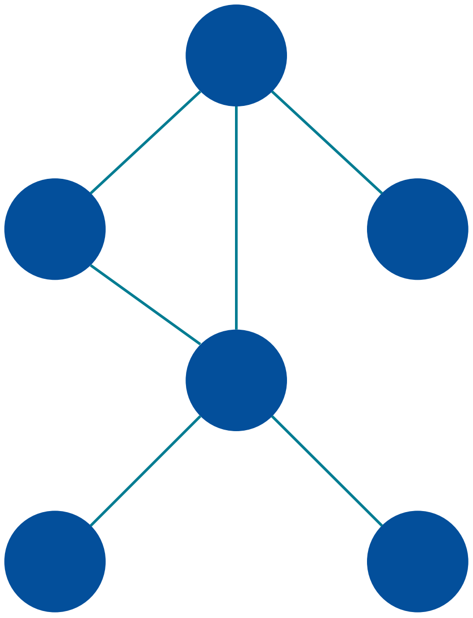
Can anyone walk across all seven bridges and never cross the same one twice? In 1735, Euler proved that such path does not exist!

Networks 101: Graph Theory (1)

- For his proof, Euler constructed a **graph** than links the pieces of land by their existing bridges
- In other words, Euler **represented (modelled)** the case using a graph
- Euler was the first to use graphs for a mathematical proof
- from Euler proof, a new branch of mathematics emerged known as **Graph Theory**

Networks 101: Graph Theory (2)

- The roots of Network Science comes from **Graph Theory**
- Graph theory form the mathematical background for Network Science
- In mathematical terms
  - networks = graphs
  - nodes = vertices
  - links = edges



Networks 101: Basic properties

- There are two basic properties for networks:
  - N = Number of nodes (vertices)
  - L = Number of links (edges)

- Links can be either **directed** or **undirected** and consequently the network will be either a directed network or an undirected one

- **For example:**
  - On Twitter, you can **follow** a user but that does not mean the other user will follow you
  - In contrast, Facebook's **friend** means that both you and your friend are following each other

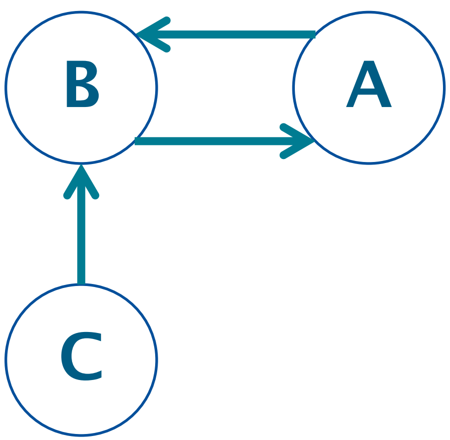
Networks 101: Links representation (1)

Twitter

A follow B

B follow A

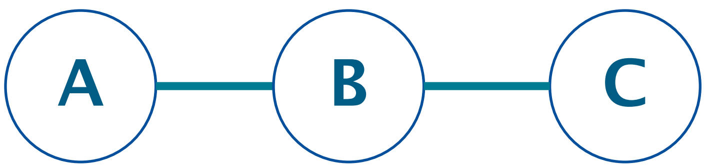
C follow B



Facebook

A is friend with B == B is friend with A

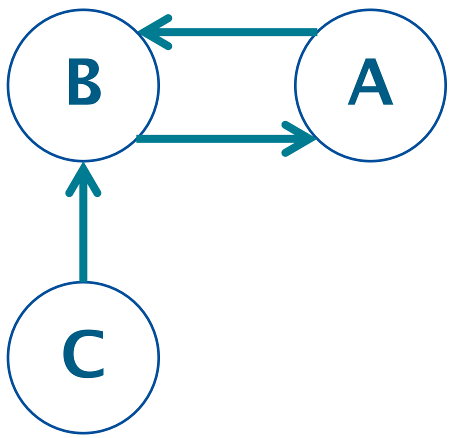
C is friend with B == B is friend with C



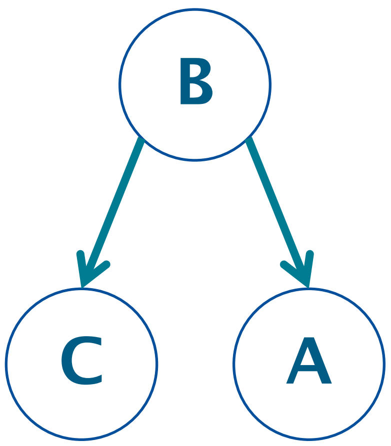
Both are considered as social networks!

Networks 101: Links representation (2)

- What if we want to represent the direction of information diffusion/flow between users i.e. retweets or shares?
- For example:
  - If user A retweet from user B and user C retweet from B

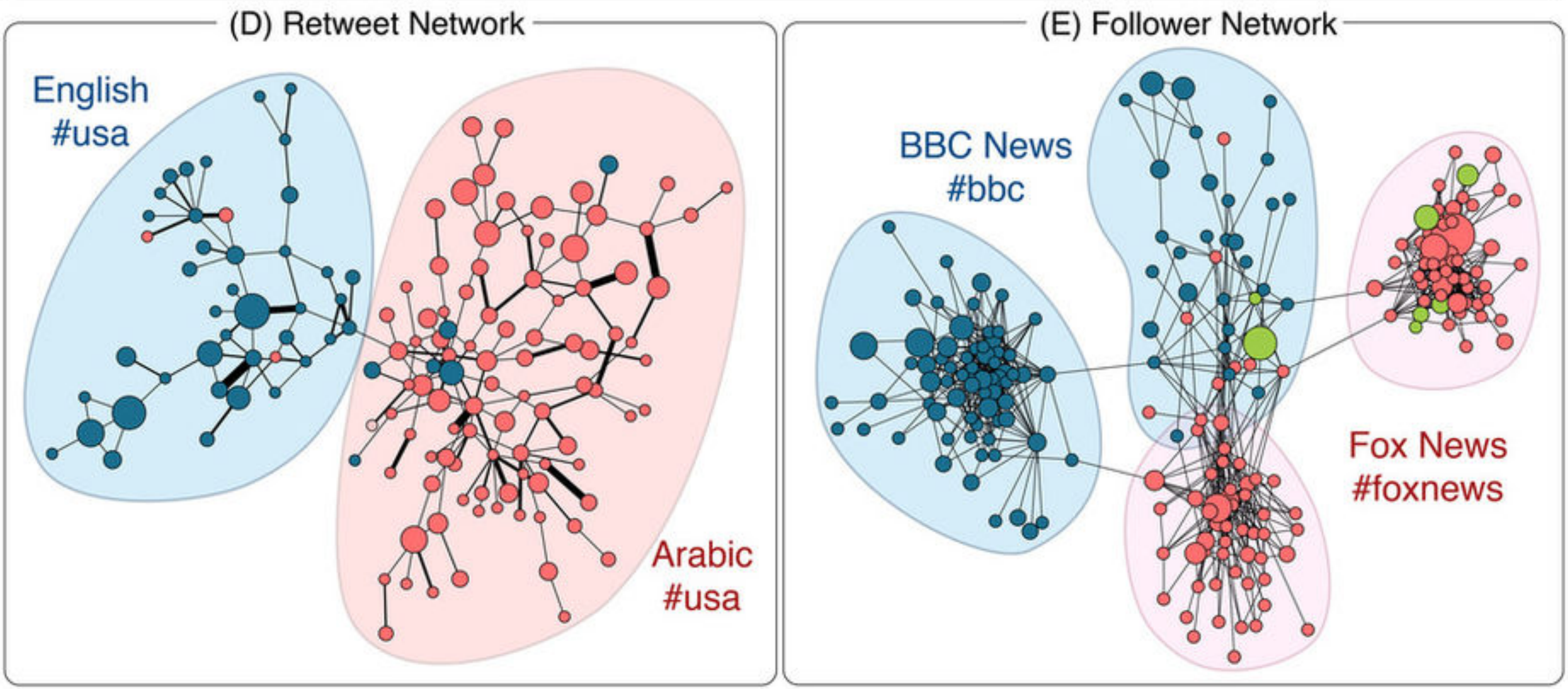


Social network



Retweets network
(information diffusion)

Networks 101: Different representations



[Weng et al. 2013]

Networks 101: Examples

Network	Nodes	Links	Directed/Undirected
Internet	Routers	Internet connections	Undirected
WWW	Webpages	Links (URLs)	Directed
Power Grid	Power plants	Cables	Undirected
Mobile phone calls	Subscribers	Calls	Directed
Email	Email addresses	Emails	Directed
Social network	Individuals/ users	Friend/Family/Follow etc.	Directed/Undirected
Information diffusion	Individuals/ users	Information flow	Directed
Science collaborations	Scientists	Co-authorships	Undirected
Citation networks	Papers	Citations	Directed
Actor network	Actors	Co-acting	Undirected
Character network	Characters	Relationship (friend, family, acquaintance, etc.)	Undirected
Protein interactions	Proteins	Binding interactions	Undirected
E. Coli metabolism	Metabolites	Chemical reactions	Directed

Table partially taken from: Barabási, A.L., 2016. *Network science*.

Networks properties: Degree (1)

- A **node's degree** is the number of links it has to other nodes
- However, it is different for directed networks and undirected ones
- For directed networks we differentiate between **in-degree** and **out-degree**
- We can also calculate the **average degree** for the whole network
  - For undirected network, divide the summation of all nodes' degree by the total number of nodes
  - Or = **2L/N**
  - For a directed network, divide the total number of links by the total number of nodes = **L/N**

Networks properties: Degree (2)

- For this **undirected** network the **degree** of each node is as follows:

  A = 3, B = 2, C = 1, D= 4,

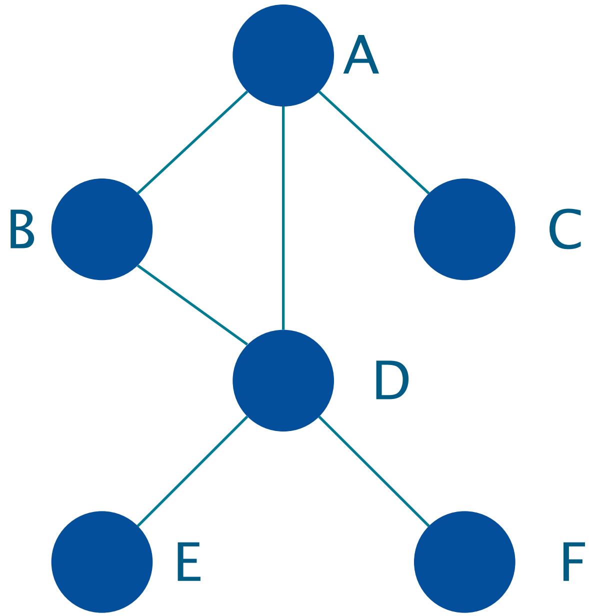
  E = 1, F = 1

- The **average degree** for this network:

  L = 3 + 2 + 1 + 4 + 1 + 1 = 12

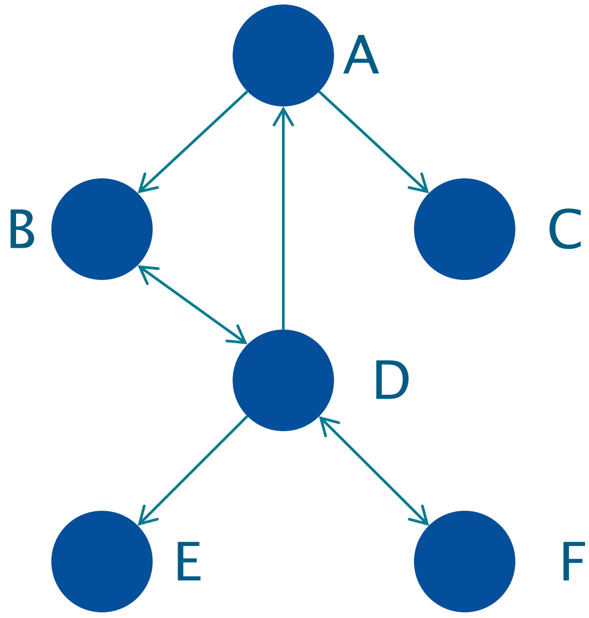
  Average degree = 12 / 6 = 2

  Or = (2 x 6) / 6 = 2



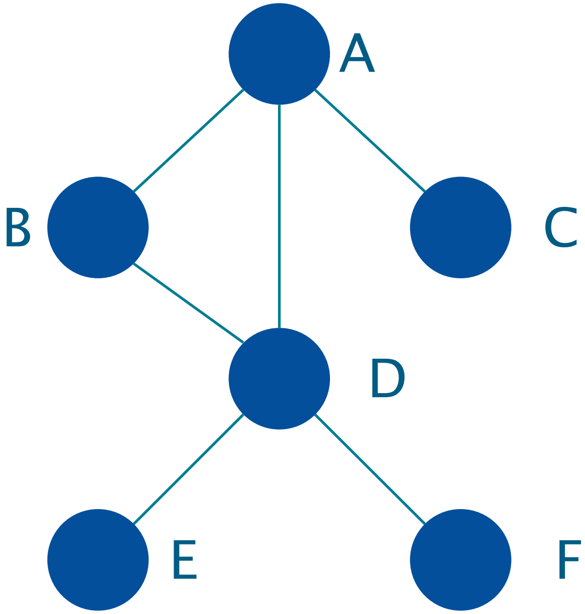
Networks properties: Degree (3)

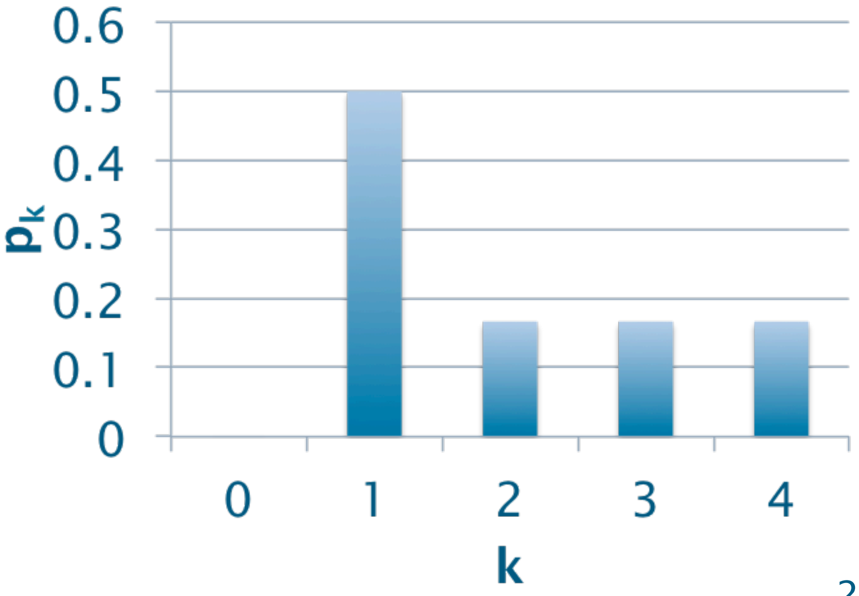
- For this **directed** network the **in-degree** of each node is as follows:
  A = 1, B = 2, C = 1, D= 2, E = 1, F = 1
- The **out-degree** of each node is as follows:
  A = 2, B = 1, C = 0, D= 4, E = 0, F = 1
- The **total-degree** of each node is as follows:
  A = 3, B = 3, C = 1, D= 6, E = 1, F = 2
- The **average degree** for this network:
  L = Σ in-degrees = Σ out-degrees
  Average degree = 8 / 6 = 1.33



Networks properties: Degree (4)

- **Degree distribution**
  - p_k is the probability that a randomly selected node in the network has degree k





Networks properties: Degree (5)

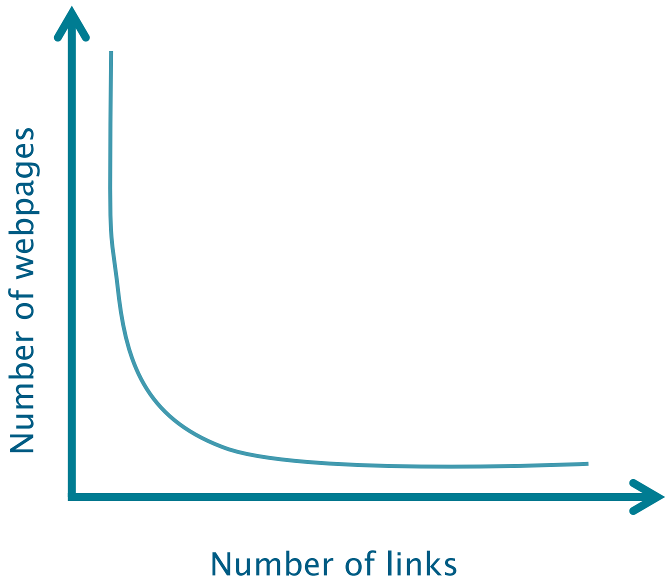
- The **degree distribution** is one of the fundamental concepts in graph theory because:
  - It is needed to calculate many network properties
  - It determines many network-related phenomena such as the spread of information and network robustness

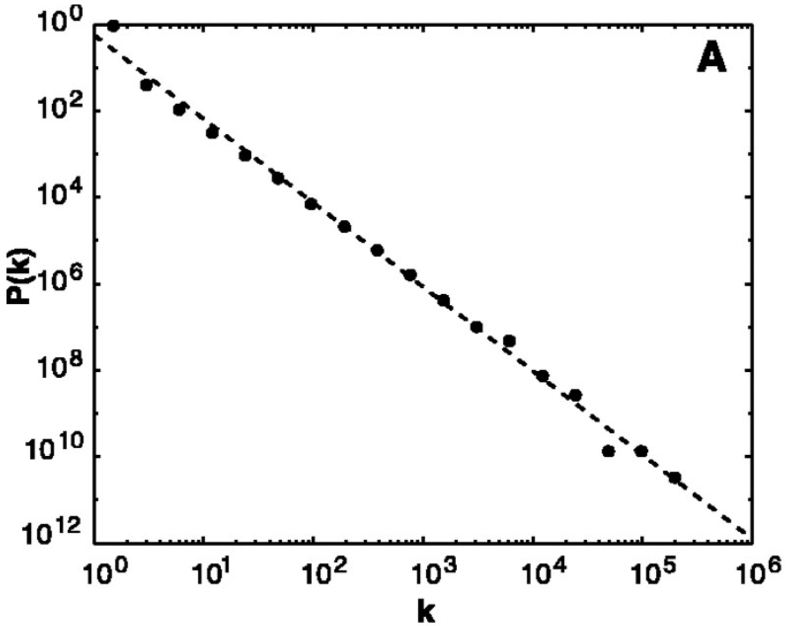
- In ***real networks*** such as the **Web** the degree distribution is sparse
  - The majority of webpages have a low degree while few webpages (hubs) have a very large degree

  → Power law distribution

Networks properties: Degree (6)

- **Power-Law Distribution of the World Wide Web**





[Adamic and Huberman, 2000]

Networks properties: Paths, Path lengths

- A **path** is a route from one node to another
- **Path length** is the number of hops from one node to another = the number of links in a path
- We can also identify the **shortest path** between any pair of nodes
- Paths and shortest path are fundamental to compute a number of network properties known as centrality measures

Networks properties: Centrality measures (1)

- By its name, **centrality** measures how central a node is
- I.e., they are ways to describe nodes' position in a network
- These measures help us identify the most central nodes i.e., **the most powerful/influential** nodes

Networks properties: Centrality measures (2)

- There are a number of centrality measures:
  1. **Degree centrality**
  2. **Betweenness centrality**
  3. **Closeness centrality**
  4. **Eigenvector centrality**

Networks properties: Centrality measures (3)

- **Degree centrality** → node's connectedness: how connected a node is?

- **in degree** – how many links to this node?
  – A web page is important if lots of things link to it
  – A twitter user is important if lots of other users follow it
  – A journal article is important if lots of other articles cite it

- **out degree** – how many links from this node
  – A review paper is useful if it links to lots of other papers
  – A search engine is useful if it links to other sites

Networks properties: Centrality measures (4)

- **Betweenness centrality:** How many pairs of nodes is a node between?
  - A node is between two others if it is on the shortest path from one to another
- A node that has high 'betweenness' is vital to the communication of the network
  - Many / most messages between two nodes must travel through it

Networks properties: Centrality measures (5)

- **Closeness centrality :** The farness of a node with respect to a network is the sum of its distances to every other nodes in the network
- The closeness of a node is the **inverse** of its **farness**
- Closeness determines how quickly a message can travel from one node to the whole network

Networks properties: Centrality measures (6)

- **Eigenvector centrality**: Instead of considering all nodes equivalent to each other, assume that some are more important
- There is some kind of *ranking* function
- Eigenvector centrality determines the connection of a node v to the *important* nodes in the network
- **PageRank** is an example

Network visualisation

- Network visualisation help us to visualise data in ways that allow us to get some insightful remarks

- There are two main aspects to **network visualisation**:
  1. Network layout: nodes placement in the network (position)
  2. Appearance: Nodes and edges styles: colour, size, shape

Network visualisation: Some useful tools

Desktop	Developer
NodeXL	Networkx
Gephi	graph-tool
Sci2 Tool	SNAP
Tableau	D3

DEMO

References

1. Adamic, L.A. and Huberman, B.A., 2000. Power-law distribution of the world wide web. science
2. Barabási, A.L., 2016. Network science. Cambridge university press.
3. Easley, D. & Kleinberg, J. (2010). Networks, Crowds, and Markets: Reasoning About a Highly Connected World
4. Newman, M. E. (2004). Analysis of weighted networks. Physical Review
5. Weng, L., Menczer, F. & Ahn, Y.-Y. (2013). Virality prediction and community structure in social networks. Scientific reports
6. Zhu, X., Gerstein, M. and Snyder, M., 2007. Getting connected: analysis and principles of biological networks. Genes & development