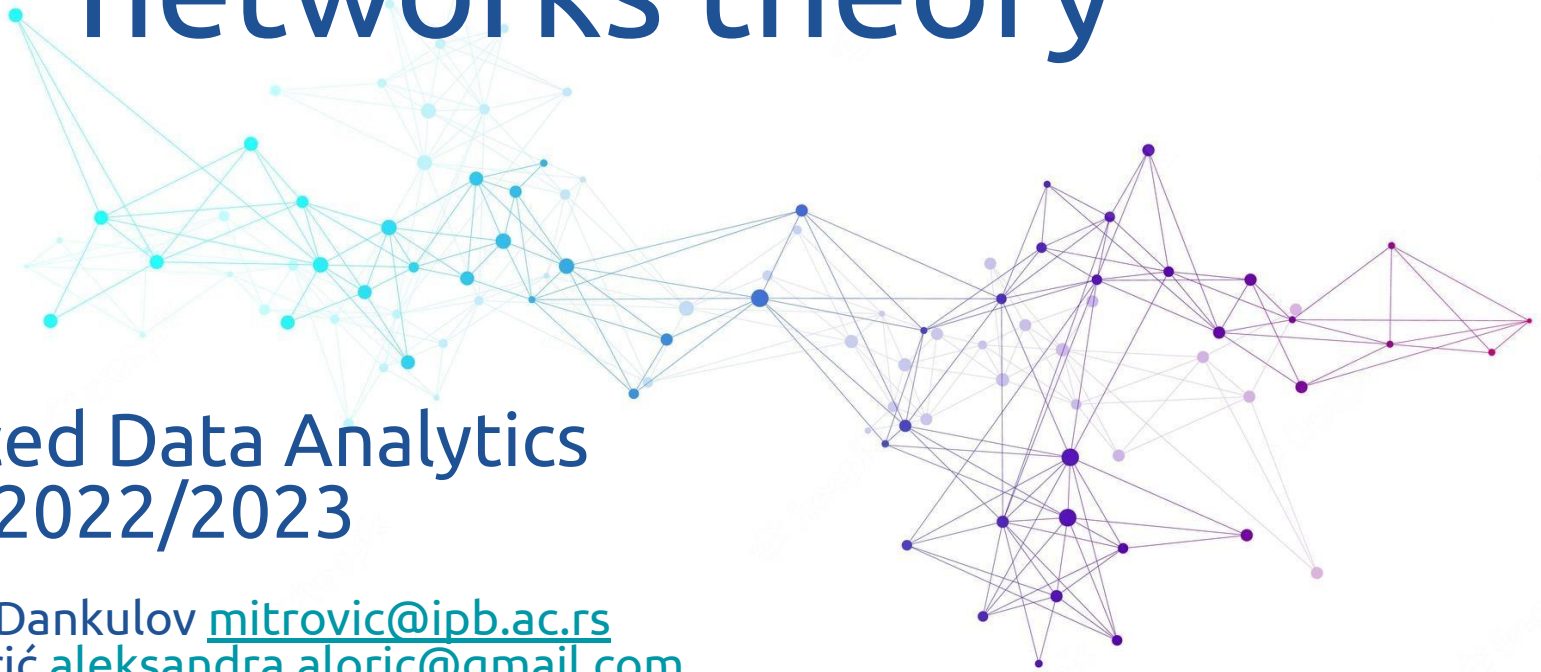
Introduction to complex networks theory



Advanced Data Analytics 2022/2023

Marija Mitrović Dankulov mitrovic@ipb.ac.rs
Aleksandra Alorić aleksandra.aloric@gmail.com

Who are we?

Aleksandra Alorić

Marija Mitrović Dankulov

- Two desperados
- King's College London
- Institute of Physics Belgrade
- Faculty of Physics, UB
- Jozef Stefan institute
- Aalto University

Aim of this course

- Develop an intuition for reasoning about problems using networks.
- Understand network representations, basic terminology and concepts.
- Understand and use methods for describing and clustering network data.
- Analyze and model real-world network data, using math and computation.

Course materials

- Github repo link
  - Presentations and other reading materials
  - Notebooks and data for practical part of the course
- Tools:
  - Python (notebooks)
  - Networkx
- Book: http://networksciencebook.com/
  - During every lecture we'll point out appropriate chapters
  - Provide reading materials whenever we cover topics outside of the book

Course structure

- 06/12/2022: Introduction - what are networks and why are they useful
- 08/12/2022: Intro to python and selected libraries for data manipulation & visualization
- 10/12/2022: Node features, degrees, the degree distribution; Math refresher
- 12/12/2022: Types of links - (un)directed, weighted
- 14/12/2022: Paths in networks
- 16/12/2022: Network connectedness; Clustering & motifs
- 21/12/2022: Temporal & multiplex networks
- 23/12/2022: Network project demo - from research question over data to results
- 26/12/2022: Null models for complex networks
- 28/12/2022: Network models cont.
- 11/01/2023: Network robustness
- 13/01/2023: Communities in networks
- 17/01/2023: Communities cont.
- 19/01/2023: Spreading phenomena
- 24/01/2023: Final thoughts & outlook

Grade structure

- 10% mini homeworks - after (every) lesson there will be a short homework, something that can be easily done/answered with the knowledge from the lecture
- 40% homework reports
  - 2 homeworks: winter break homework, end of course
  - mini versions of exam
- 60% take home exam
  - We will prepare dataset & jupyter notebook with questions, your goal will be to go through the notebook, write codes and texts answering the questions and send us back the jupyter + pdf

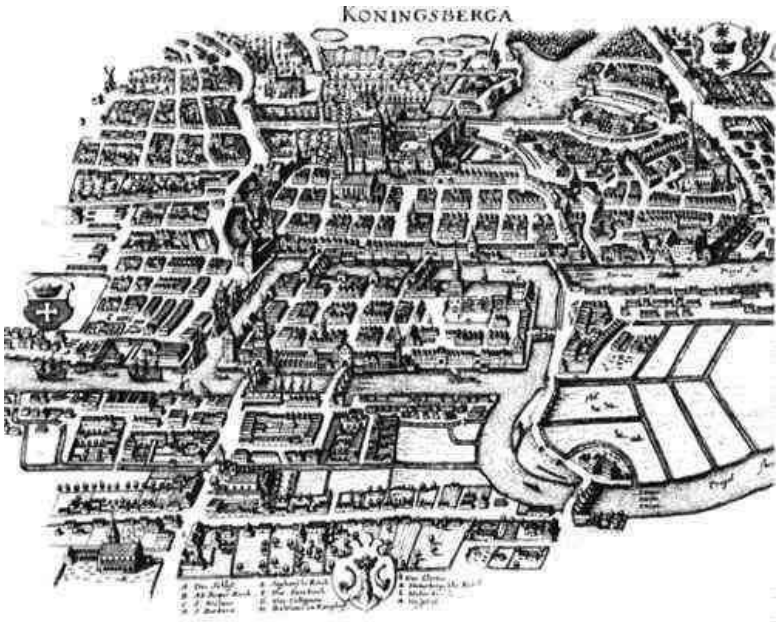
Logistics

- Can we work 2x90mins with 15mins break?
- Can we start at 6pm?
- Laptops?
- Other questions

First Network problem

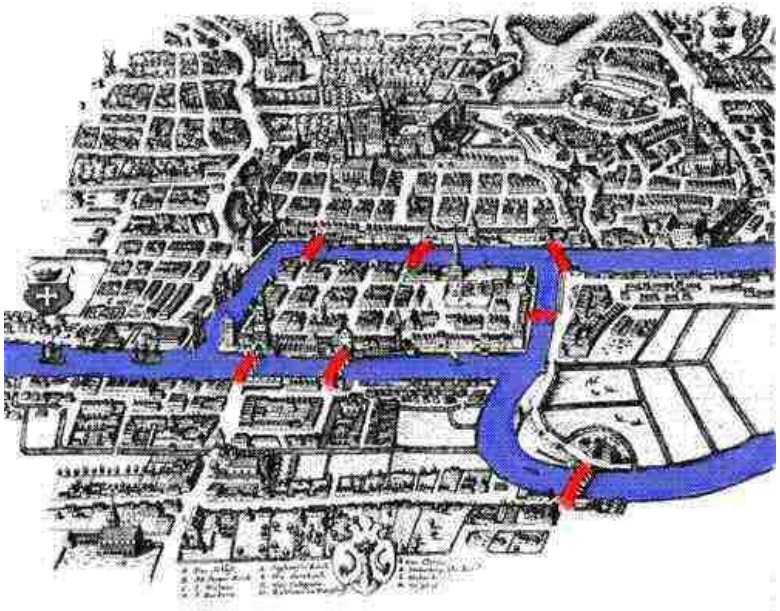
Königsberg bridges

- The city of Königsberg in Prussia (now Kaliningrad, Russia) was set on both sides of the Pregel River, and included two large islands—Kneiphof and Lomse—which were connected to each other, and to the two mainland portions of the city, by seven bridges. **The problem was to devise a walk through the city that would cross each of those bridges once and only once.**

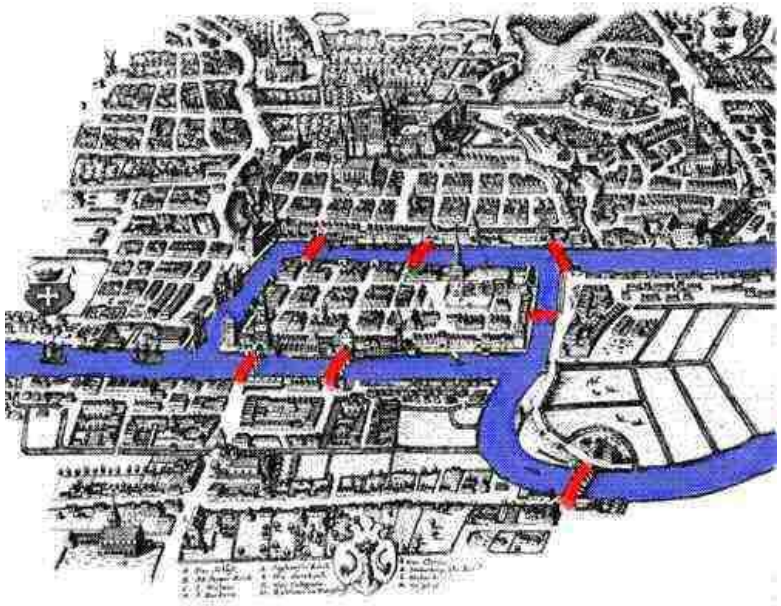


Königsberg bridges

- The city of Königsberg in Prussia (now Kaliningrad, Russia) was set on both sides of the Pregel River, and included two large islands—Kneiphof and Lomse—which were connected to each other, and to the two mainland portions of the city, by seven bridges. **The problem was to devise a walk through the city that would cross each of those bridges once and only once.**



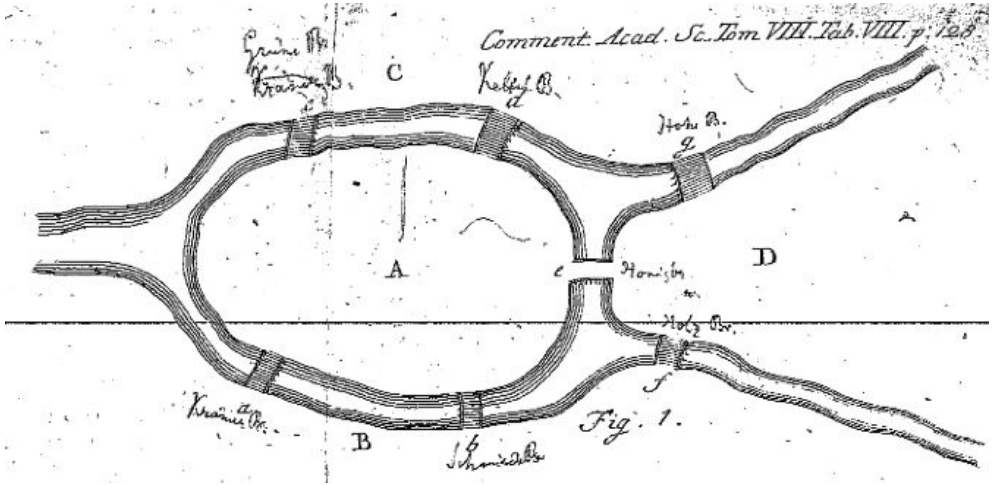
Euler’s approach - problem abstraction



- Characteristics of the city are irrelevant
  - the exact shape, size of the land pieces separated by the river are not of value when searching for the path - **we care about the brides and what they connect**

- *‘This question is so banal, but seemed to me worthy of attention in that [neither] geometry, nor algebra, nor even the art of counting was sufficient to solve it.’*

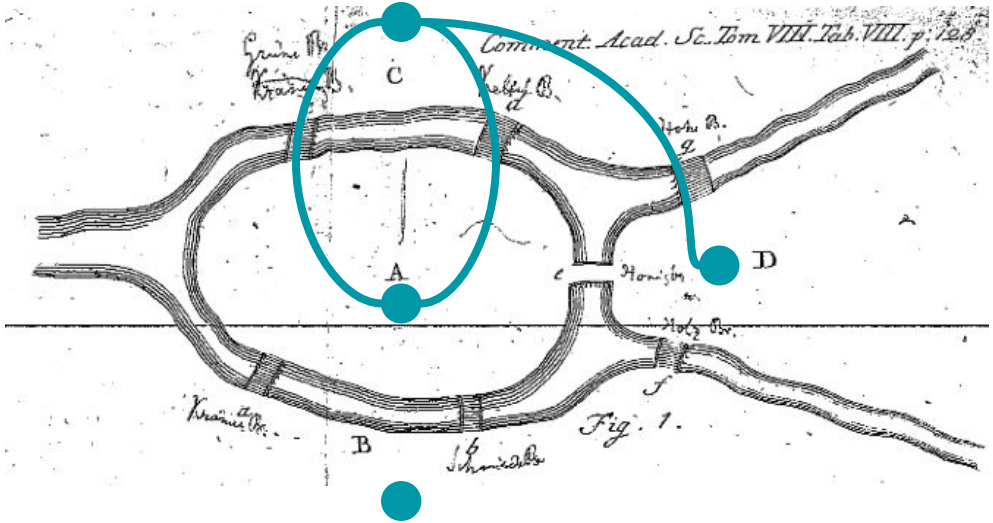
Euler’s approach - problem abstraction



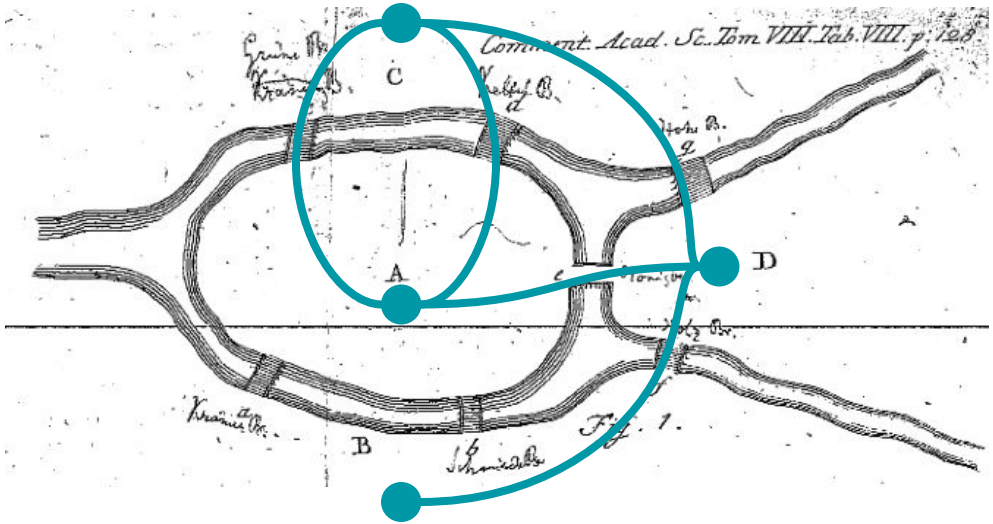
- Characteristics of the city are irrelevant
  - the exact shape, size of the land pieces separated by the river are not of value when searching for the path - **we care about the brides and what they connect**

Euler's Figure 1 from ‘Solutio problematis ad geometriam situs pertinentis,’ Enestrom 53 [source: **MAA Euler Archive**]

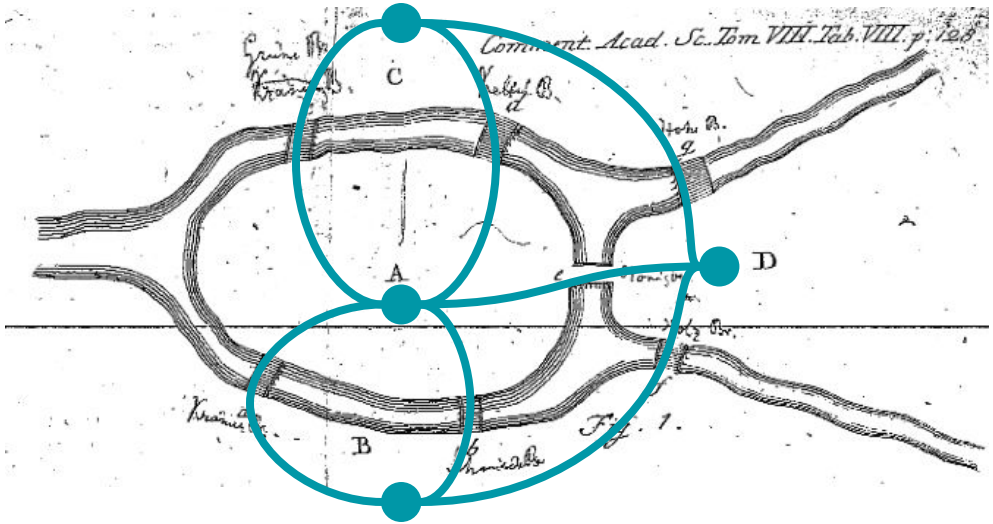
Euler’s approach - problem abstraction



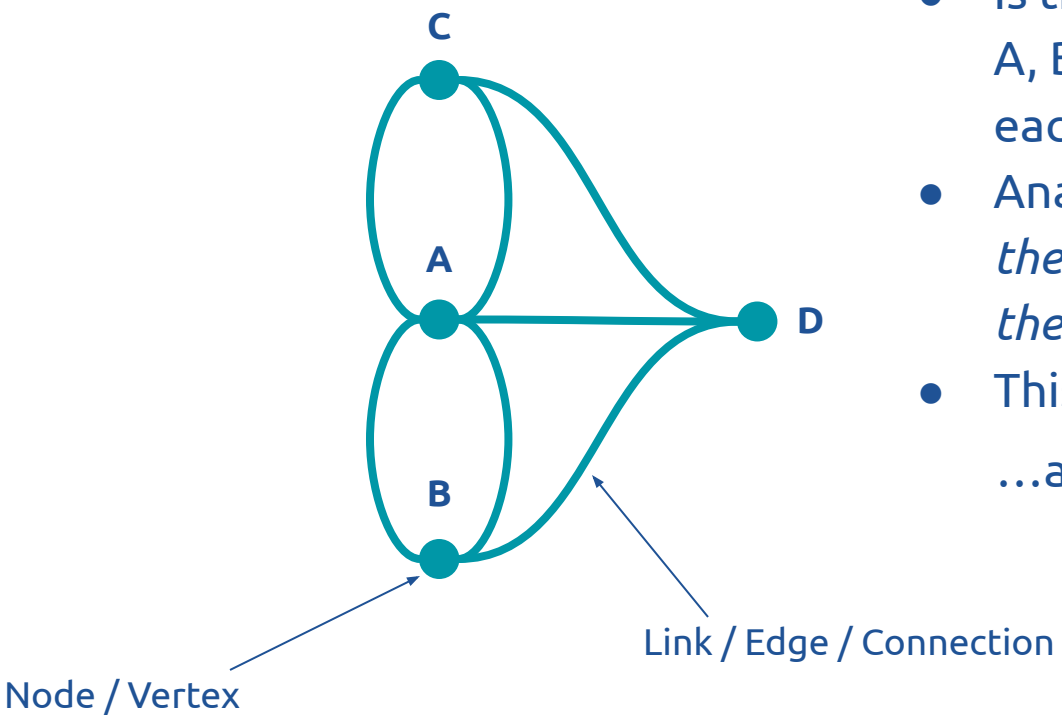
Euler’s approach - problem abstraction



Euler’s approach - problem abstraction



Euler’s approach - problem abstraction

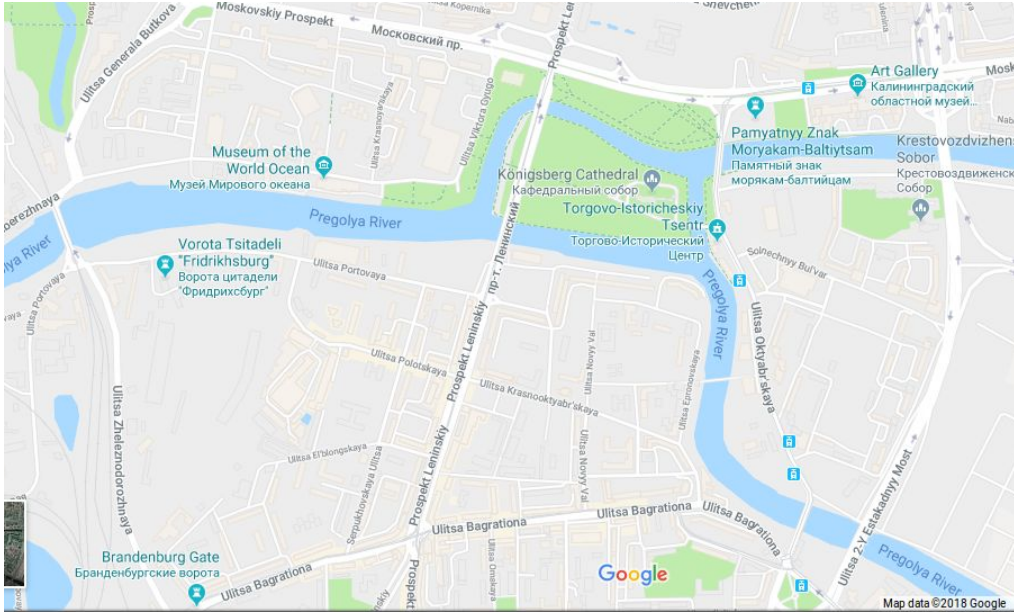


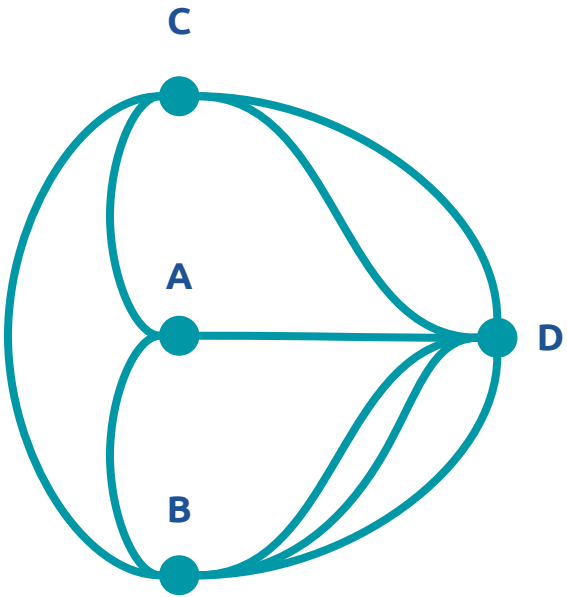
- Is there a path such that every node A, B, C, D is visited while crossing each link only once?
- Analogous question: *Can you draw the figure on the left without lifting the pen from the paper?*
- This is called Eulerian path

…and it does not exist for this network

Kaliningrad today

- Bridge situation changed - is the walk possible now?

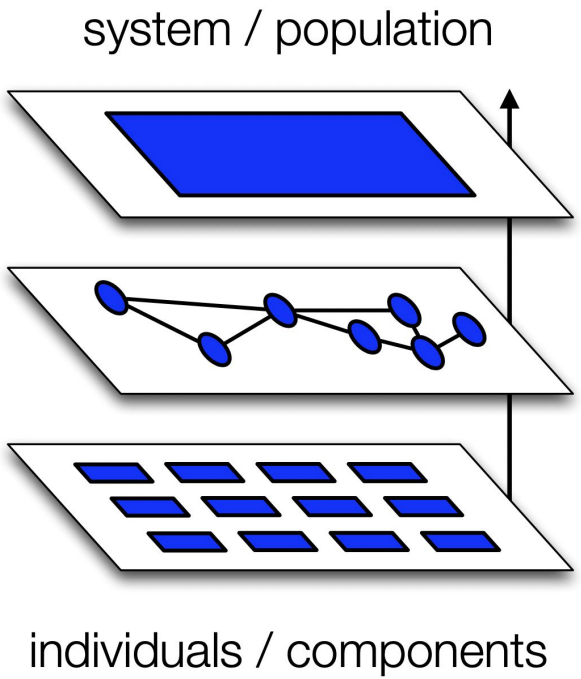




Königsberg bridges - takeaway messages

- Network abstractions can help us solve some problems
  - Distill only the most relevant bits for the problem at hand
  - Realise that the solvability of Konigsberg path is related to structure of the bridge network and not the cleverness of a city walker trying to find a solution

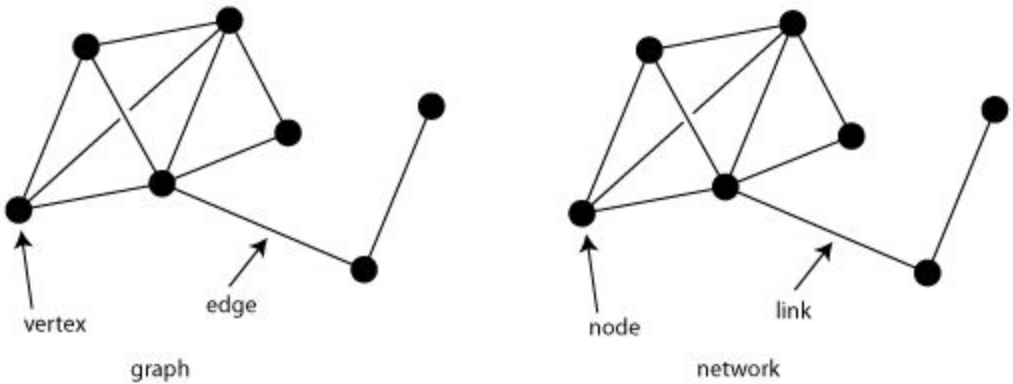
- Think of a network as:
  - an approach
  - a structure that exists above individuals / components or: structure that exists below system / population



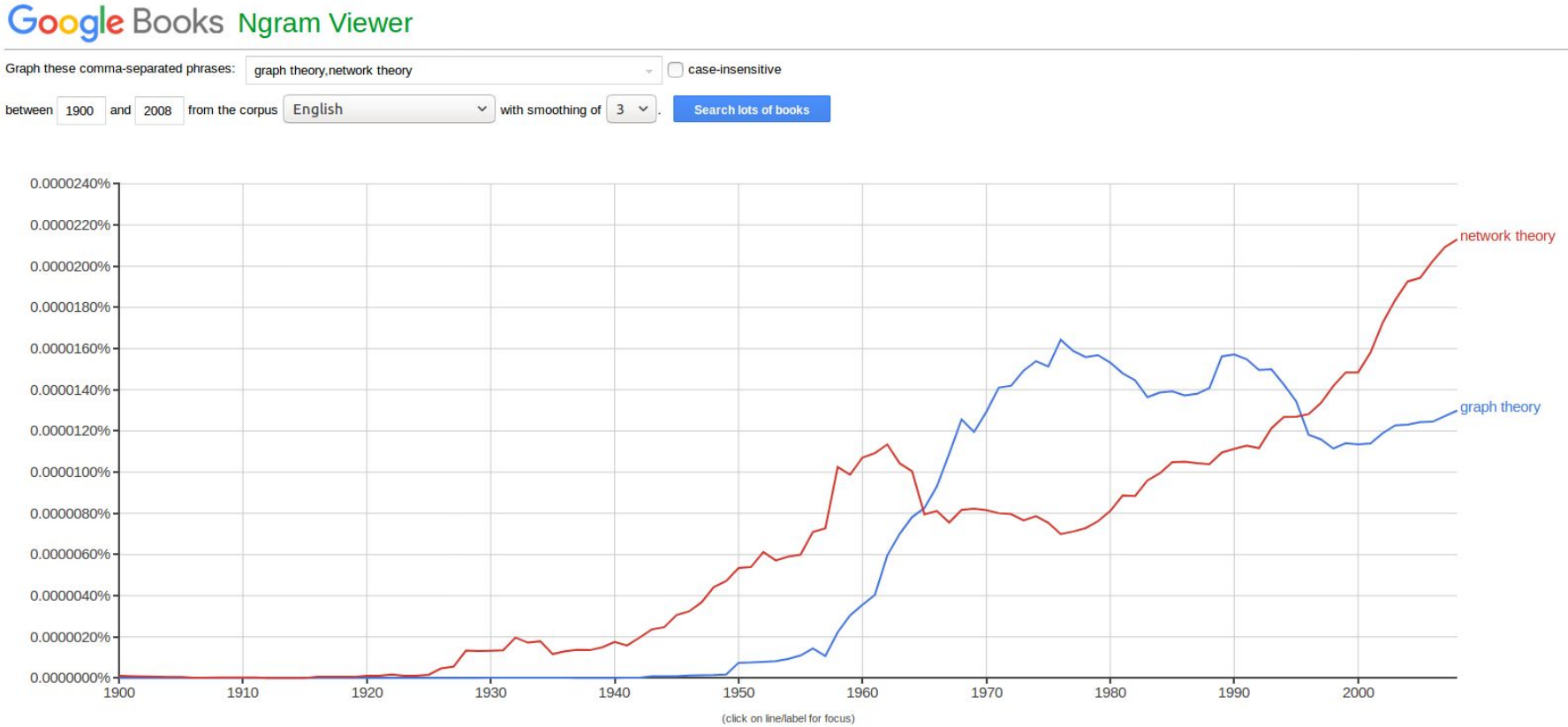
Further reading about Konigsberg bridges:
https://www.maa.org/press/periodicals/convergence/leonard-eulers-solution-to-the-konigsberg-bridge-problem

Terminology: Graph vs Network

- Most of the time dependent on who do you speak to
- Graph more widely used among mathematicians (**graph theory**)
- **Social network analysis** started as a field mostly by 20th century sociologists
- Abundance of data more recently prompted an interdisciplinary field of **network science**
- Graph more widely used to denote abstract structure, while networks are used in data-driven contexts



Terminology



Break

Intro questionnaire (~15mins)

LINK: https://forms.gle/D5WZNFt6mGucyvHg6

Problems we can’t solve without networks

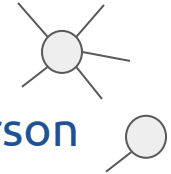
…or we would solve them wrongly

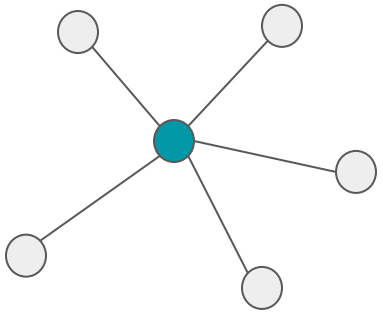
Who should share information

Imagine there’s an important information that you’re aiming to share with a group of people who know each other. You can talk with only one person and you rely on them sharing the information further. Who would you choose to communicate the information?

You ask them how many people they know:

- 1 person knows 5 other people
- 5 people know only 1 other person





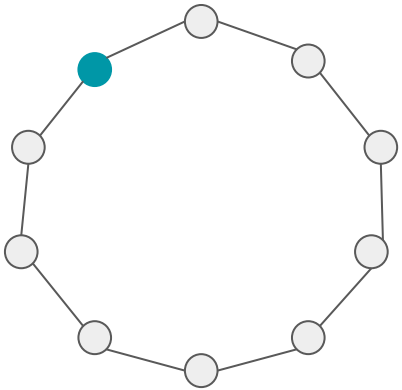
Who should share information

Imagine there’s an important information that you’re aiming to share with a group of people who know each other. You can talk with only one person and you rely on them sharing the information further. Who would you choose to communicate the information?

You ask them how many people they know:

- 10 people, all know exactly 2 other people

Scenario 1:



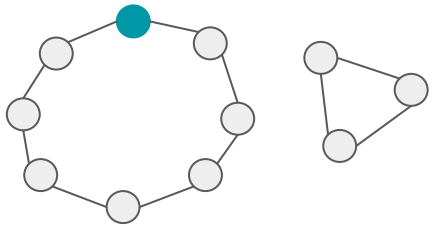
Who should share information

Imagine there’s an important information that you’re aiming to share with a group of people who know each other. You can talk with only one person and you rely on them sharing the information further. Who would you choose to communicate the information?

You ask them how many people they know:

- 10 people, all know exactly 2 other people

Scenario 2:

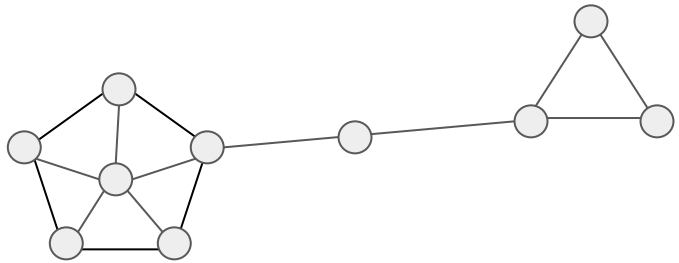


Who to isolate to stop rumor reaching everyone?

Imagine there’s a false information spreading in a group of people who know each other. You can talk with only one person aiming to prevent that most other people hear this misinformation. Who would you choose to educate?

You ask them how many people they know:

- 1 person knows 5 other people
- 1 person knows 4 other people
- 5 people know 3 others
- 3 people know 2 others



Can you meet Taylor Swift by asking a friend of a friend to connect you?

Even if you knew:

- number of friends Taylor has
- and number of friends of all their friends
- and number of their friends

…knowing friendship numbers only would not help you find the friends in common and the path over mutual friends that lead you to them.

But knowing who connect to whom, e.g. having the acquaintance network would!

Small world experiment (1969)

- Use acquaintance chain to send letter from randomly selected individuals to a target person
- Origin of ‘six degrees of separation’

An Experimental Study of the Small World Problem*

JEFFREY TRAVERS

Harvard University

AND

STANLEY MILGRAM

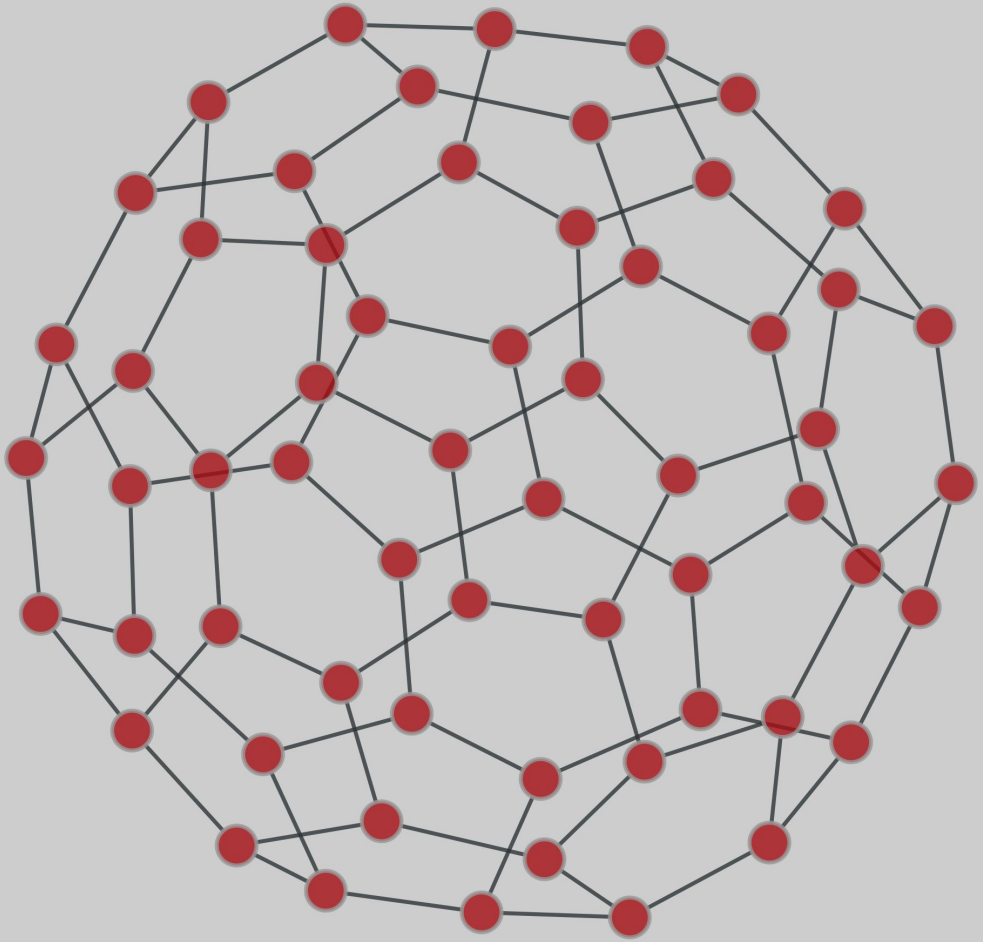
The City University of New York

Arbitrarily selected individuals (N=296) in Nebraska and Boston are asked to generate acquaintance chains to a target person in Massachusetts, employing "the small world method" (Milgram, 1967). Sixty-four chains reach the target person. Within this group the mean number of intermediaries between starters and targets is 5.2. Boston starting chains reach the target person with fewer intermediaries than those starting in Nebraska; subpopulations in the Nebraska group do not differ among themselves. The funneling of chains through sociometric "stars" is noted, with 48 per cent of the chains passing through three persons before reaching the target. Applications of the method to studies of large scale social structure are discussed.

Other studied networks

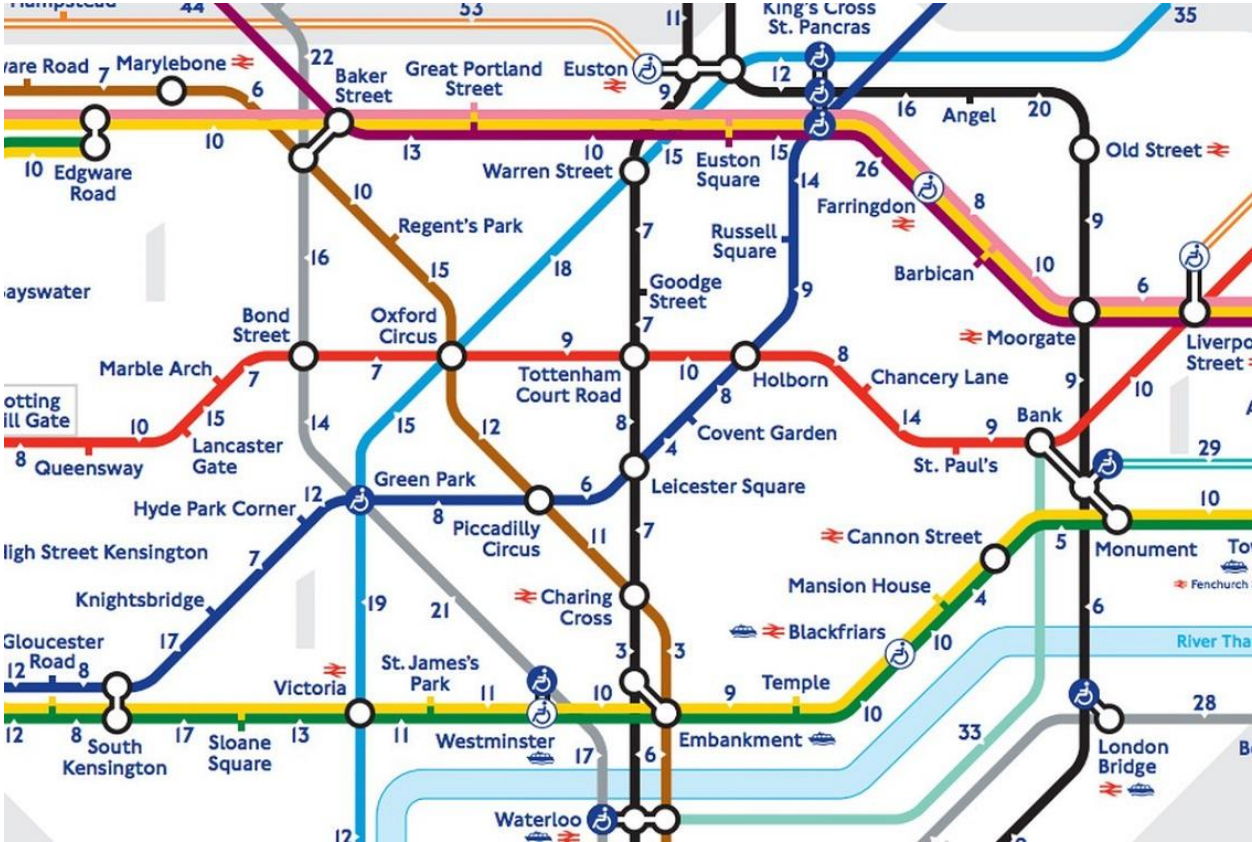
Molecular structure

- Mostly regular networks
- Fullerene C60 is shown on the right, but other carbon structures are interesting (graphene, carbon nanotubes…), as well as other other non-carbon molecules
- Material properties are related to the network structure
- Nodes - carbon atoms
- Links - chemical bonds



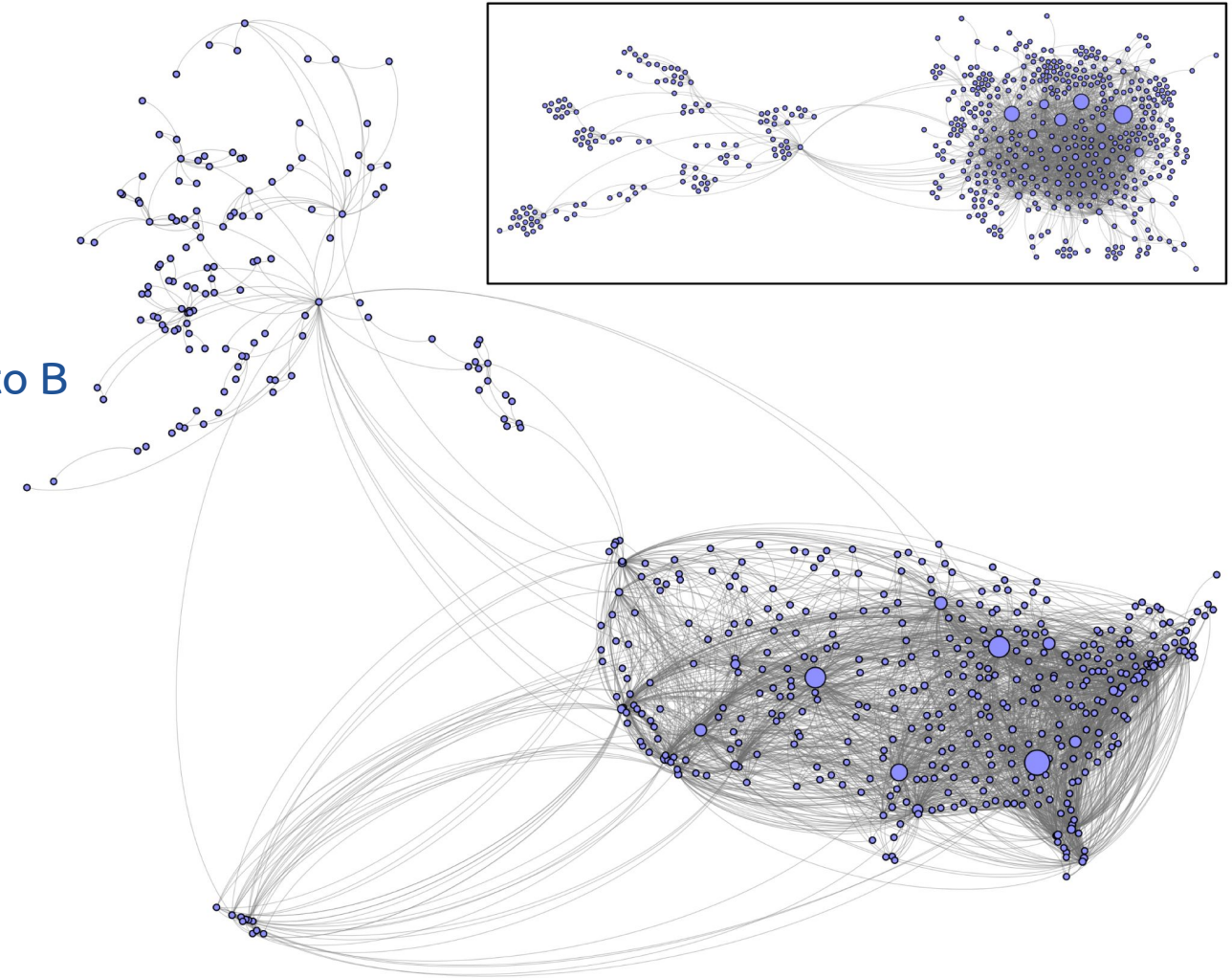
Metro network

- London tube network
- Useful for finding your way around the city
- Nodes - stations
- Links - metro lines operating between stations



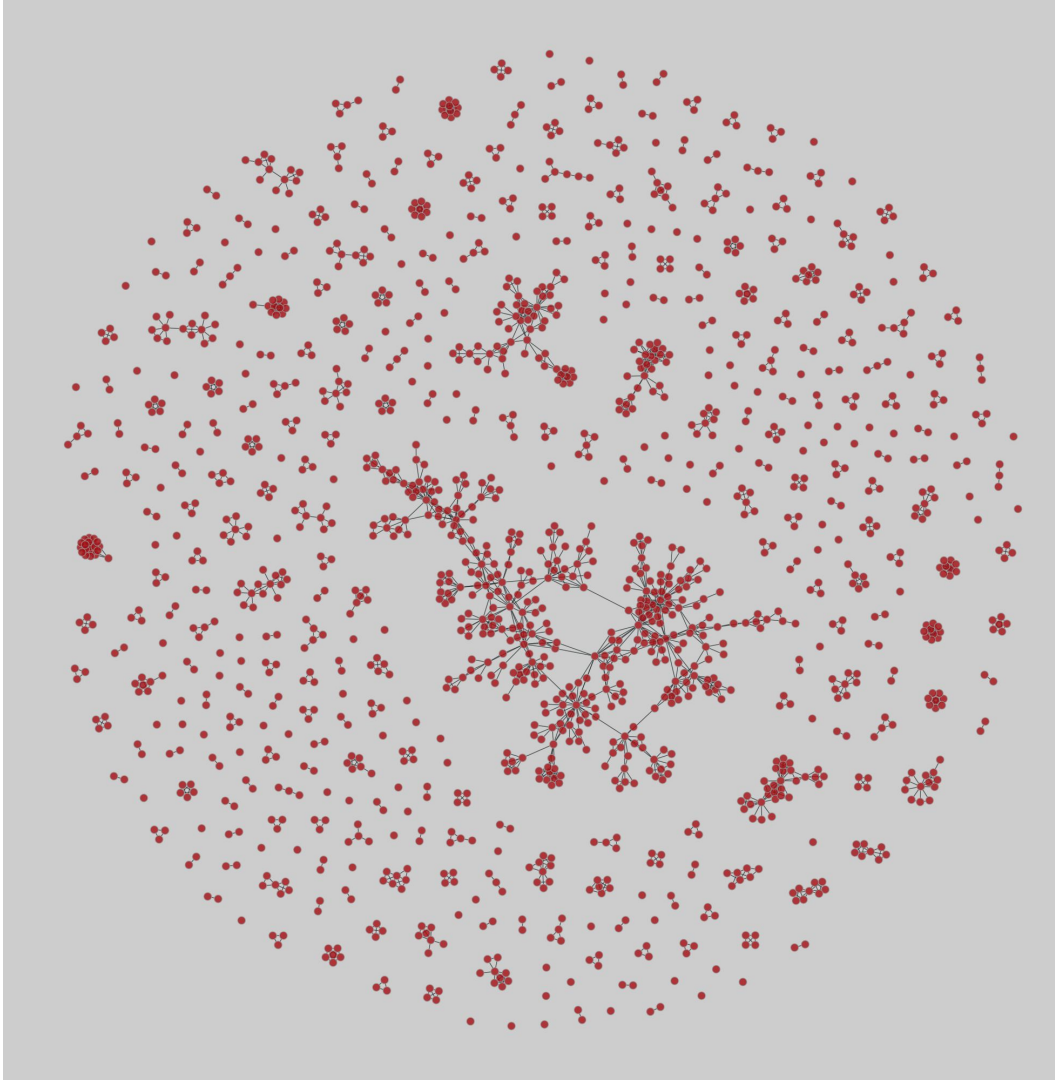
Flight network

- Network of direct flights within USA
- How to get from A to B with least airport transfers?
- Nodes - airports
- Links - direct flights



Scientific collaboration

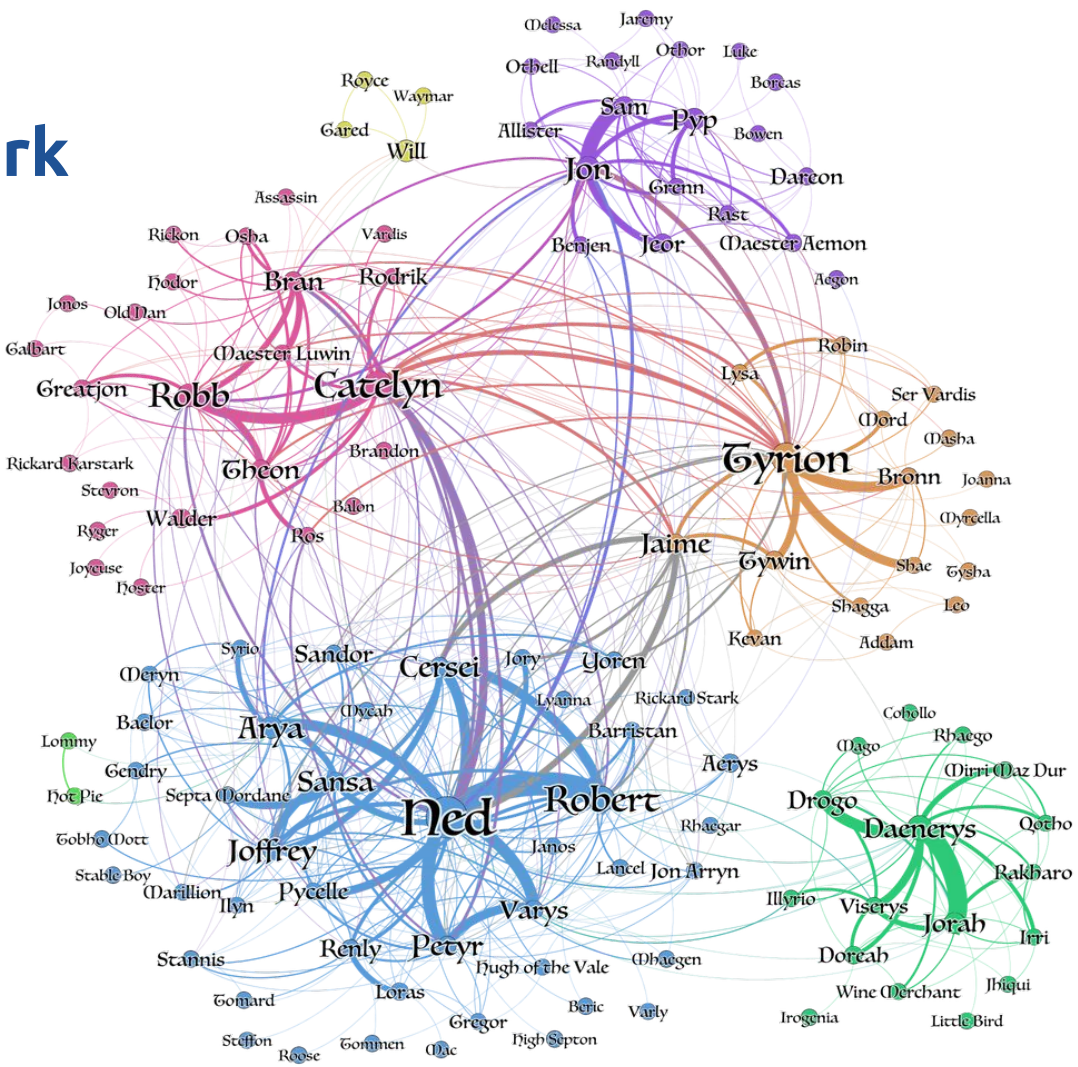
- Collaboration of network scientists (2006) - two persons are connected if they published a paper together
- Nodes - scientists
- Links - joint publications



Fiction character network

- Game of thrones (season 1)
- Useful for automatic detection of ‘important’ characters
- Nodes - characters
- Links - interaction (dialog, joint appearance in a scene, mention…)

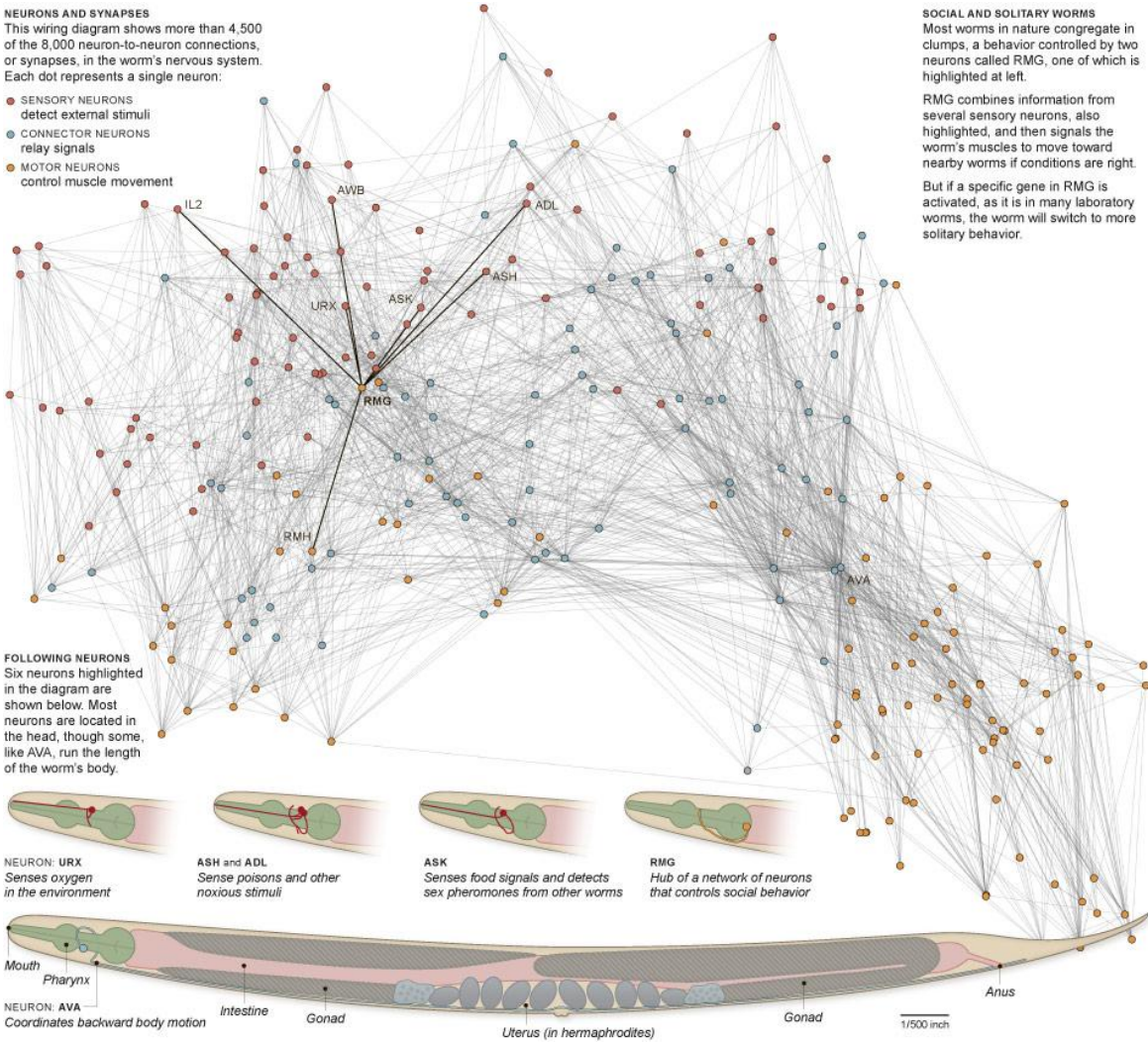
- All books and seasons here: https://networkofthrones.wordpress.com/



Networks of neurons

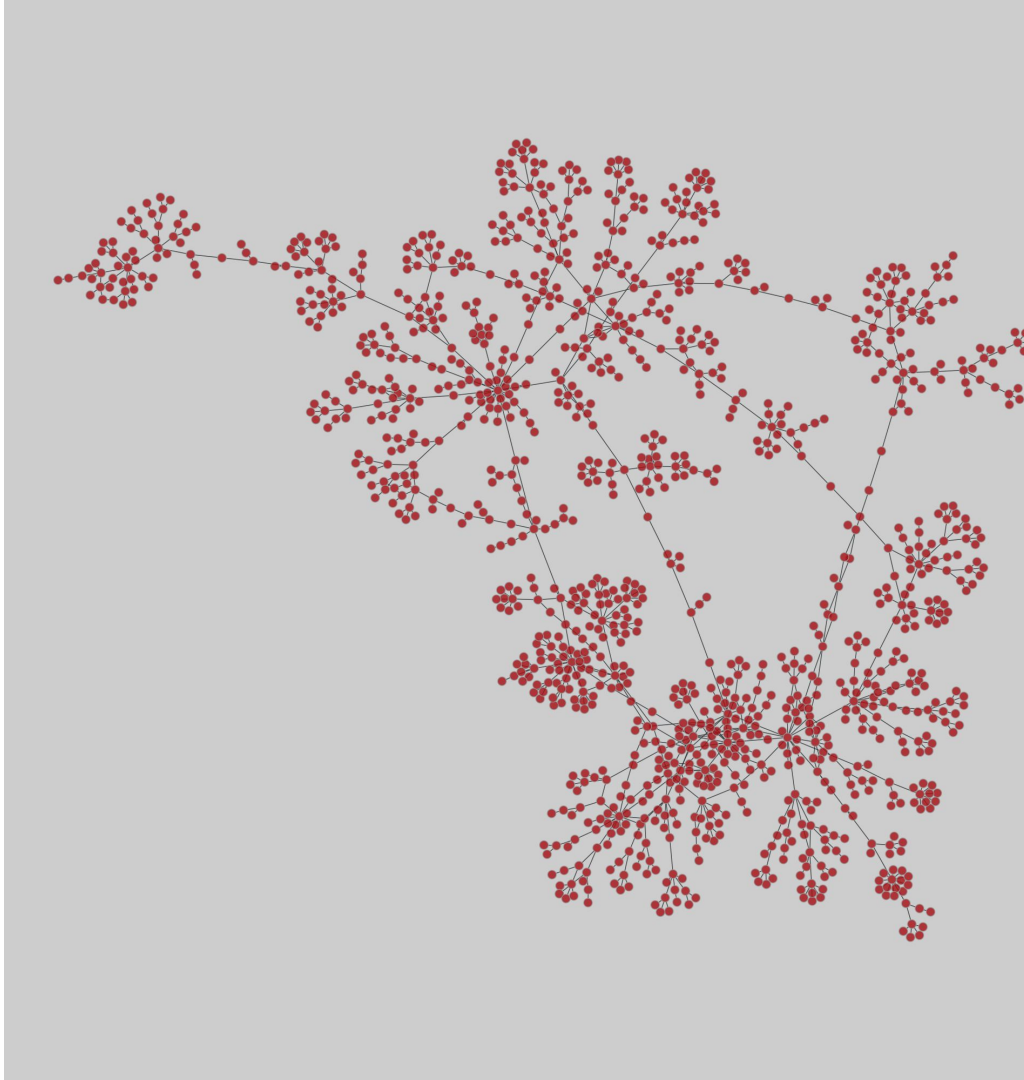
- C. elegans nervous system
- Nodes - neurons
- Links - synapses

- Only 302 neurons, but more than 8000 links
  - For reference human brain has 86 billion neurons



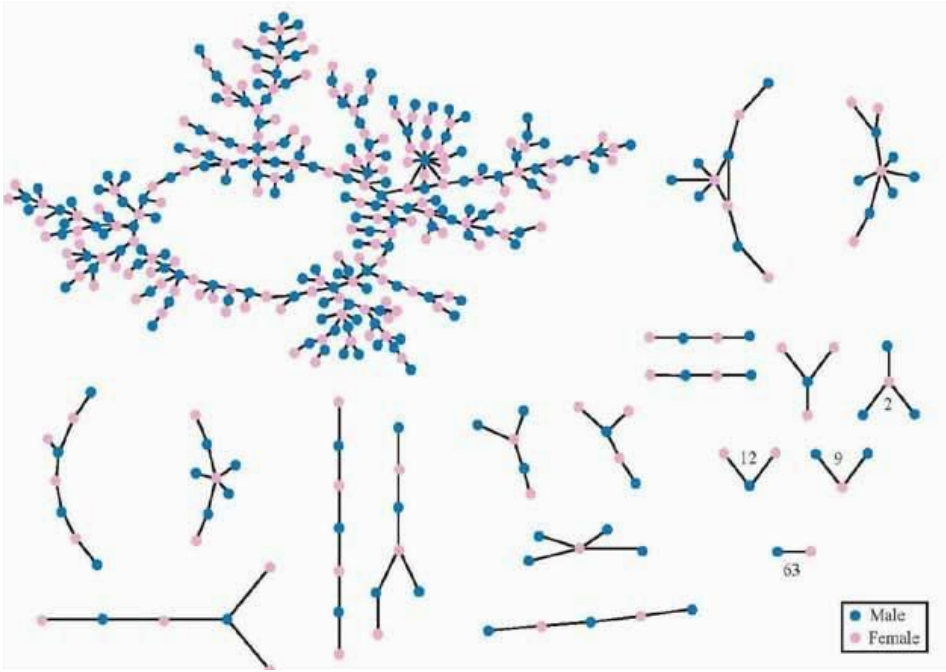
Crime network

- Rosenfeld crime network (1991)
- People involved or affected by crimes in St. Louis in the 1990s
- Nodes - people - suspects, witnesses and victims of crimes
- Links - exist between people who were involved/affected by the same crime



Teenage relationships

- Romantic & Sexual relationships at Jefferson Highshool within 6 months period
- Useful for assessment of STD transmission rates
- Nodes - students
- Links - relationship

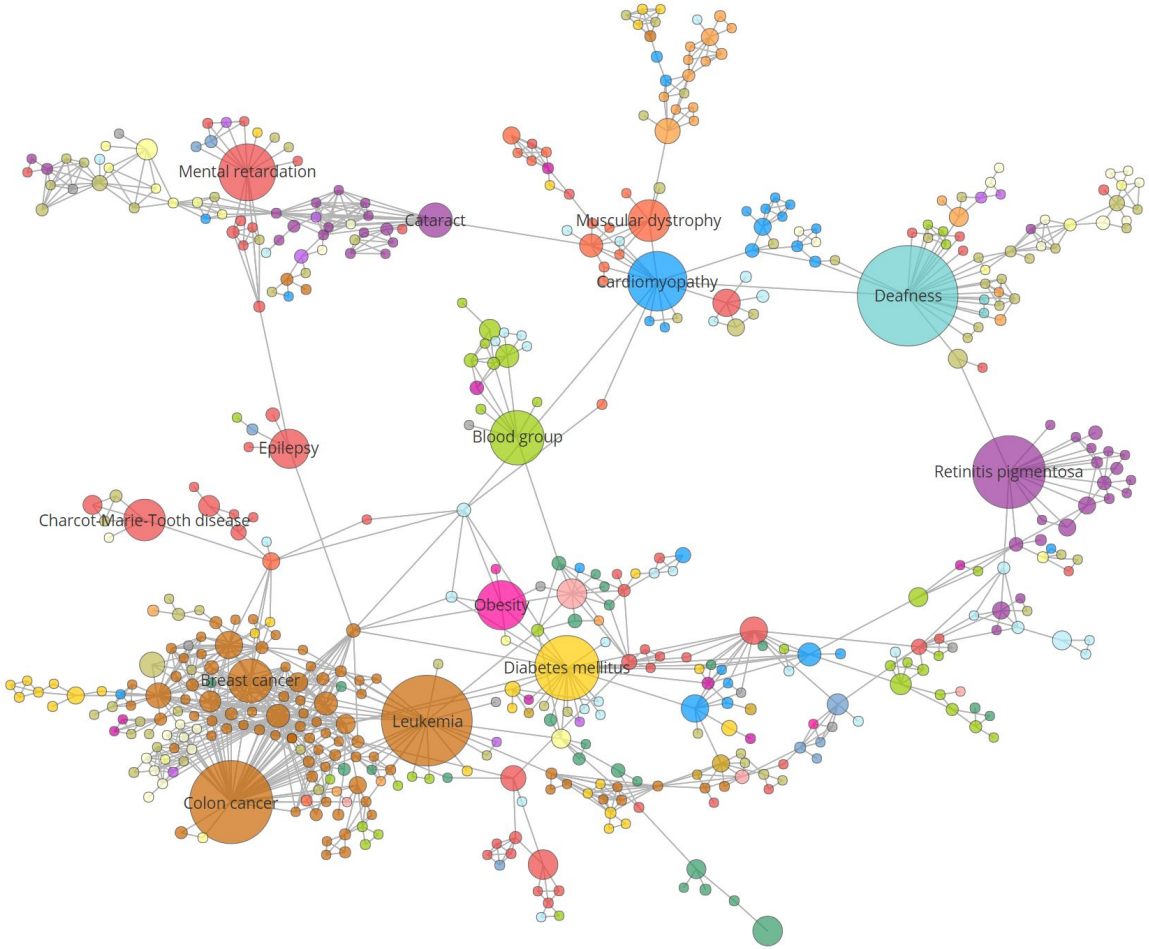


Company network

- Board members are shared among top companies
- Important to assess capital flows and companies interdependencies
- Nodes - companies
- Links - shared board member

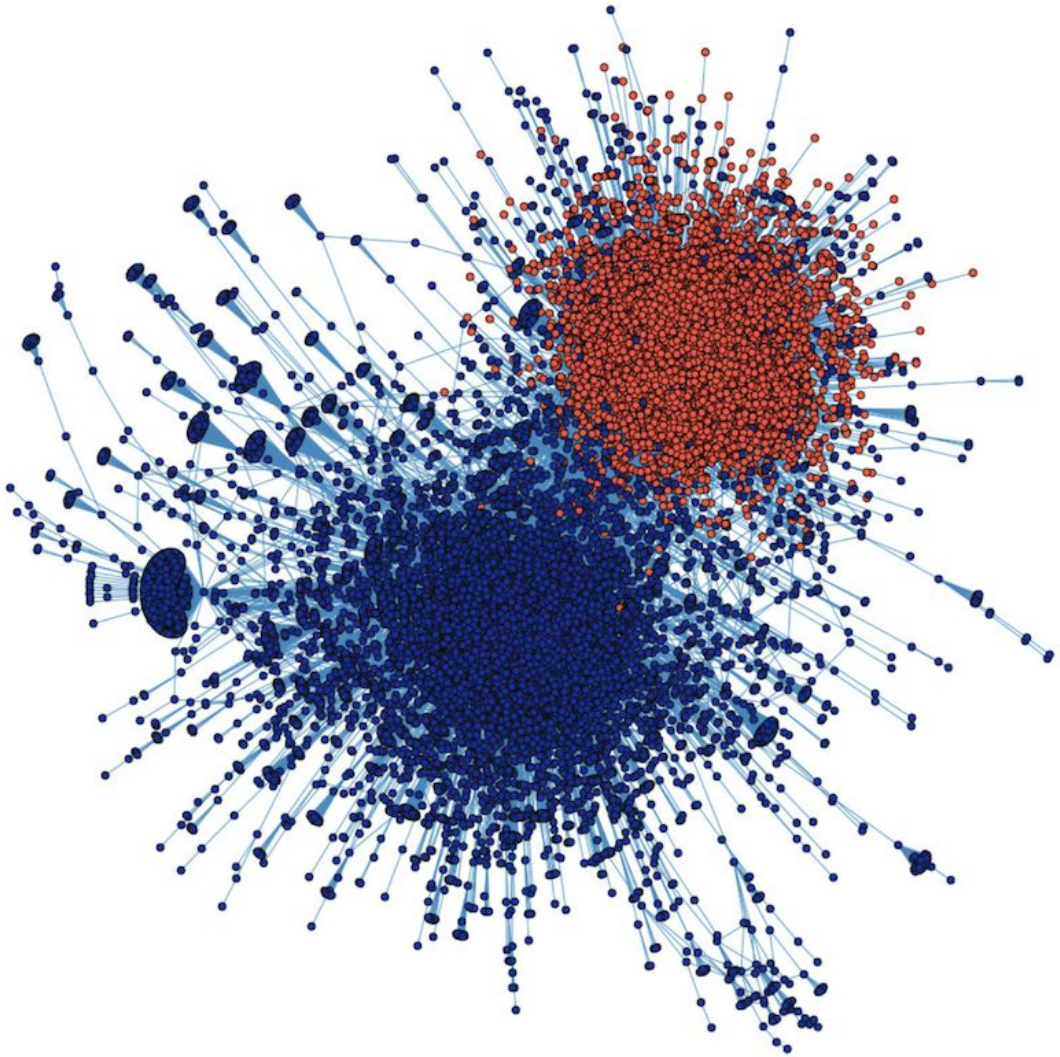
Human disease network

- Network of disease based on their common genetic origin
- Nodes - disease
- Links - gene in common



Retweet network

- Political retweet network during 2010 US midterm elections
- Useful for assessing political polarisation in public communication
- Nodes - twitter users
- Links - retweet



What do we do once we have a network?

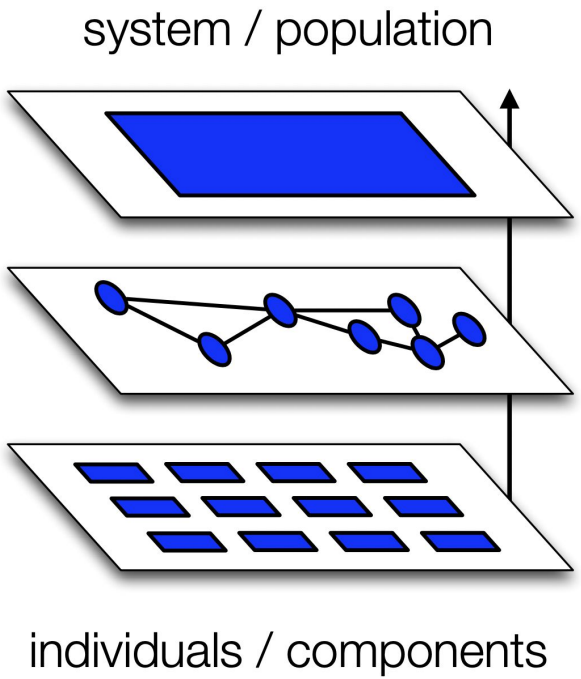
- …beyond nice visualisations :)
- Structure measurements
- Network robustness
- Prediction of missing data
- Find communities
- Comparison with null models
- Dynamics on networks

- … and we’ll cover all of these during the following weeks

There’s no a single network in a system

Among you (so the same nodes) we can construct multiple types of networks:

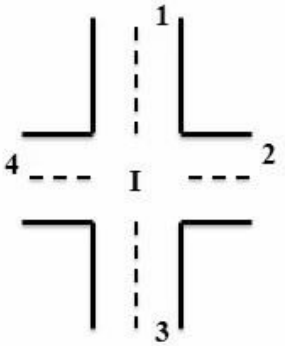
- Link means you had a conversation
- Link means you follow each other on instagram
- Link means you shook hands within the last week
- Link means you worked on a homework together



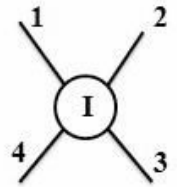
Network representation of a system is not unique!

Think about the road network, depending on the problem you are trying to solve, the same data about intersections and roads can be transformed into network in two ways:

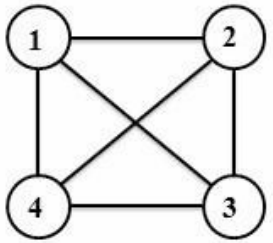
- Finding the most vulnerable intersection
- Finding road that lead to the most other roads



Road Network with an intersection



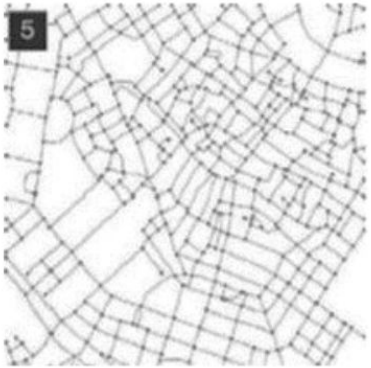
Primal Graph



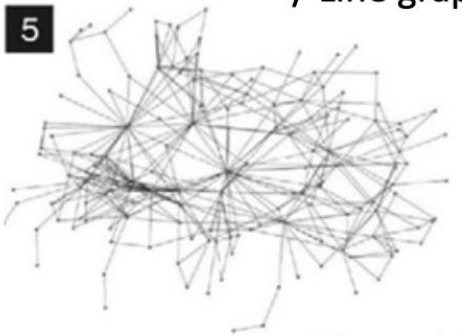
Dual Graph

Network representation of a system is not unique!

Primal graph / primary representation



Dual representation / Line graph



S. Lammer et al (2006)

Further reading ideas

- Chapter 1 of Network science book: http://networksciencebook.com/chapter/1
- Further reading about Konigsberg bridges: https://www.maa.org/press/periodicals/convergence/leonard-eulers-solution-to-the-konigsberg-bridge-problem
- Vax game https://vax.herokuapp.com/tour learn about networks by trying to stop virus spread over network.
- A network of science: 150 years of Nature papers https://www.youtube.com/watch?v=GW4s58u8PZo

Homework

Try to think of three different networks examples that we did not mention in this lecture. State the nodes and links for each of them.

Submission LINK