CASE STUDIES IN NETWORKS

Dr. David O'Sullivan

University of Limerick

Department of Maths and Stats

Email: David.OSullivan@ul.ie

Twitter: @DavidJPOS

Welcome

- Who am I!?
  - Dr. David O'Sullivan
  - University of Limerick
  - Department of Maths and Stats

- What do I do? Network and industrial data science
  - How information spreads
    - Data -> Irish marriage referendum
    - Modelling -> information spread

  - Industrial projects
    - Churn prediction
    - Electricity forecasting

Integrating sentiment and social structure to determine preference alignments: the Irish Marriage Referendum

David J. P. O'Sullivan, Guillermo Garduño-Hernández, James P. Gleeson and Mariano Beguerisse-Díaz

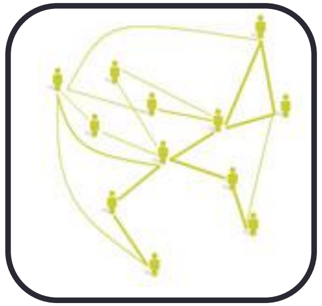
Published: 12 July 2017 https://doi.org/10.1098/rsos.170154

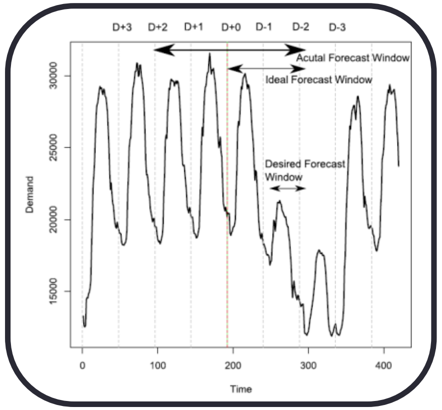
A multi-type branching process method for modelling complex contagion on clustered networks

Leah A. Keating, James P. Gleeson, David J.P. O'Sullivan

Mathematical modeling of complex contagion on clustered networks

David J. P. O'Sullivan*, Gary J. O'Keeffe, Peter G. Fennell and James P. Gleeson
Mathematics Applications Consortium for Science and Industry, Department of Mathematics and Statistics, University of Limerick, Limerick, Ireland



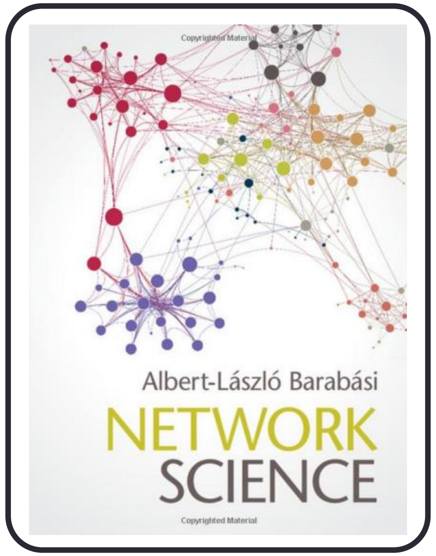


Welcome

- Short course in networks
  - Give a, brief, overview of network science
  - Focus on application of methods and theory
  - Format: introduce theory, followed by coding example (In R)

  - Early examples are taken from ->

- What will we cover?
  - Little graph theory
  - Properties of networks
  - Models of network
  - Simple spreading process on networks (if we have time)



Welcome: rough time table – subject to chance

- 9:30 – 11:30; lecture
  - Intro to R
  - Basics of networks
  - Creating network from data
- 11:30 – 12:30; coding exercise
- 12:30 – 13:30; lunch

- 13:30 – 15:30; lecture
  - Community detection
  - Spreading process on networks
- 15:30 – 16:30; coding exercise and round up
- (If were a ahead, ill add a little on an advanced topic.)

Why R?

- Build by statisticians for statisticians
- Has all the latest and greats statistical methods
- Plays well with Python, C++ and stan
- Access to the ‘tidyverse’ suite of packages

What were going to talk about

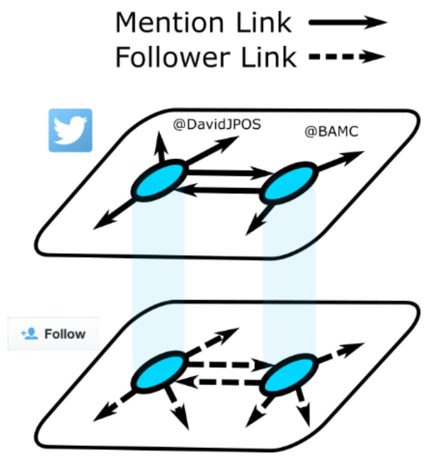
- Case studies in network analysis
  - Going to go through some of my work

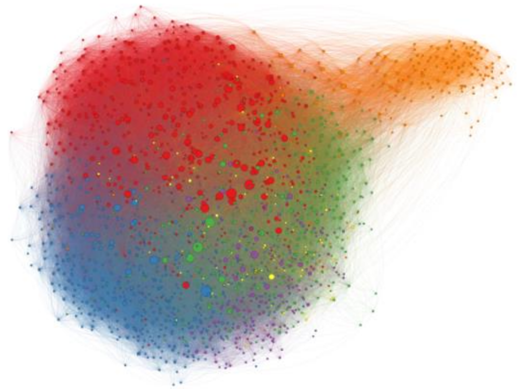
Integrating sentiment and social structure to determine preference alignments: the Irish Marriage Referendum

David J. P. O'Sullivan, Guillermo Garduño-Hernández, James P. Gleeson and Mariano Beguerisse-Díaz

Published: 12 July 2017 | https://doi.org/10.1098/rsos.170154

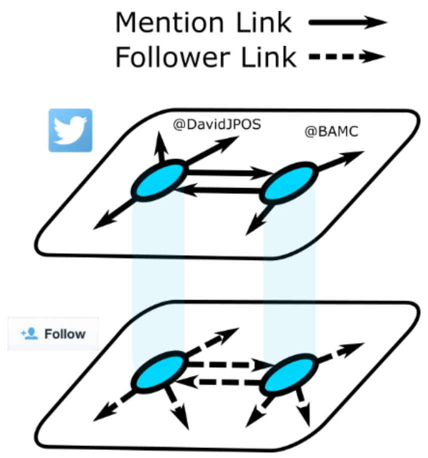
  - Use this as an excuse to introduce some ideas

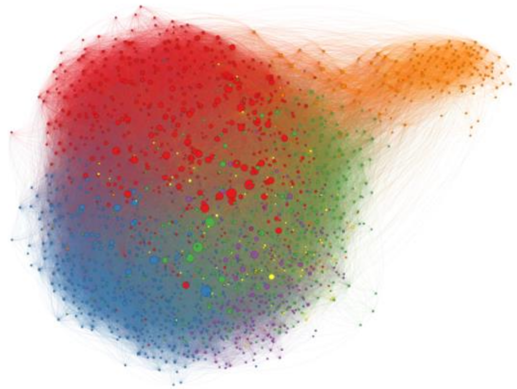




What were going to talk about

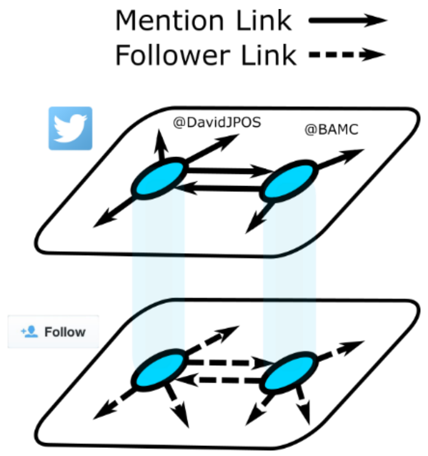
- GitHub
  - Code to recreate some of the analysis here
  - Data to have a play with
    - More on that later
  - Papers that I talk about
  - Links to books
  - Links tutorials in R
  - https://github.com/DavidJPOS/Case-studies-in-networks

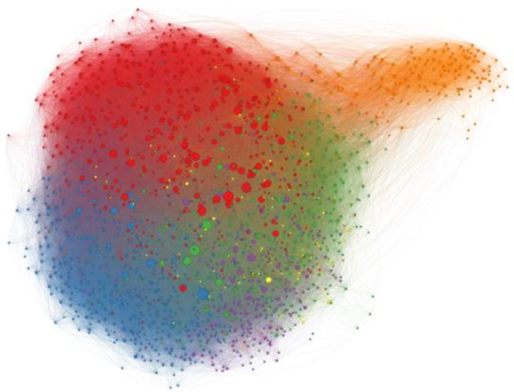




What were going to talk about

- The kind of things I'm interested in
  - Using network science and statistics to answer interesting questions and building tools for social science
  - How people self sort into groups (communities)
  - What this means for the spread of information on networks
  - Collective behaviour



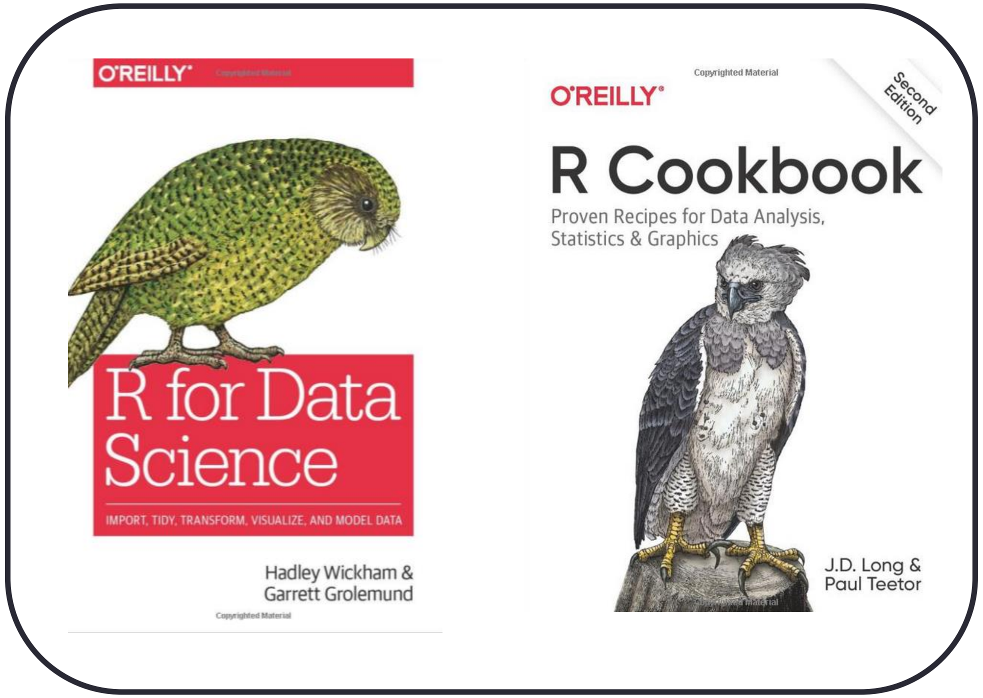


THE BASICS

First R

A little break to play with R

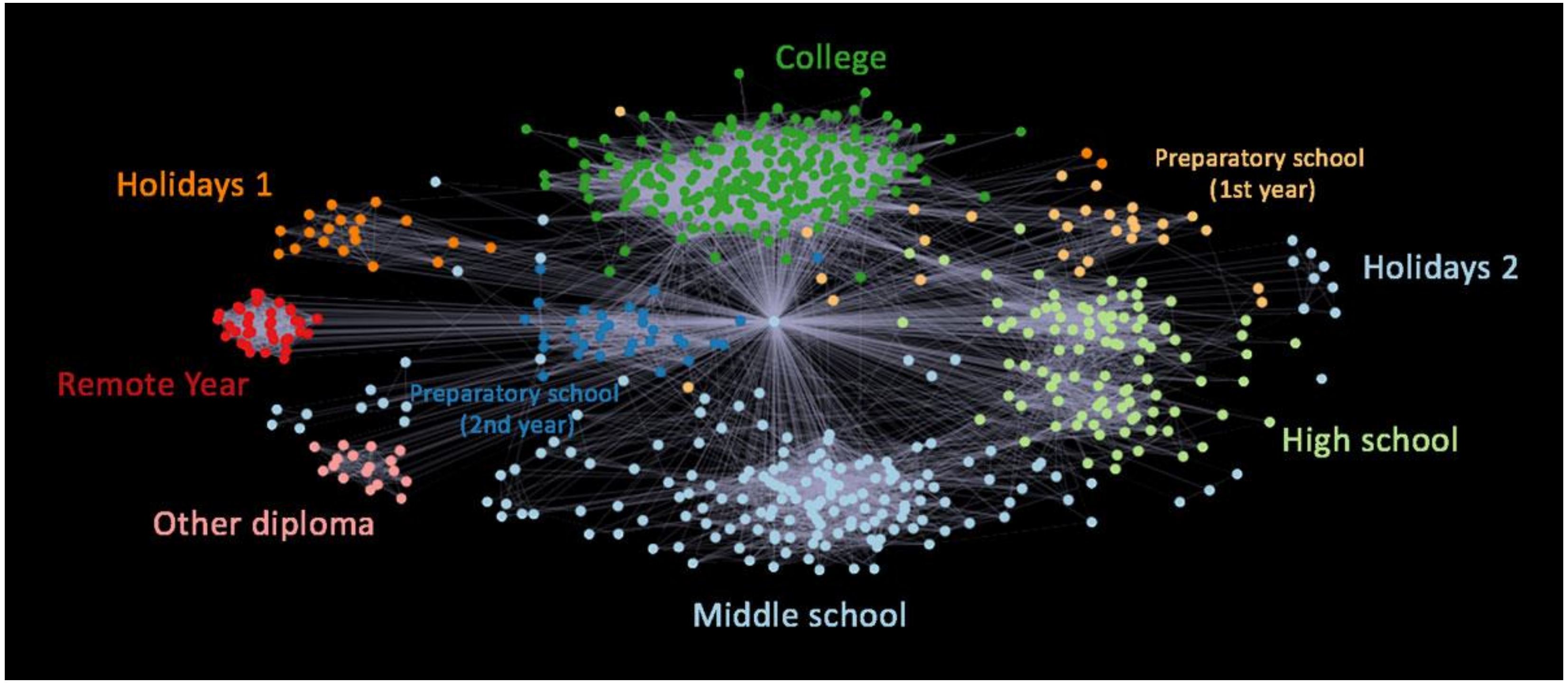
- First, we going to introduce R and the ‘tidyverse’
- Compressed version of ‘R for data science’
- Excellent books (and there free!)
- Longer tutorial for those who are interested after this
- 0_introduction_to_R_and_tidyverse.R



THE BASICS

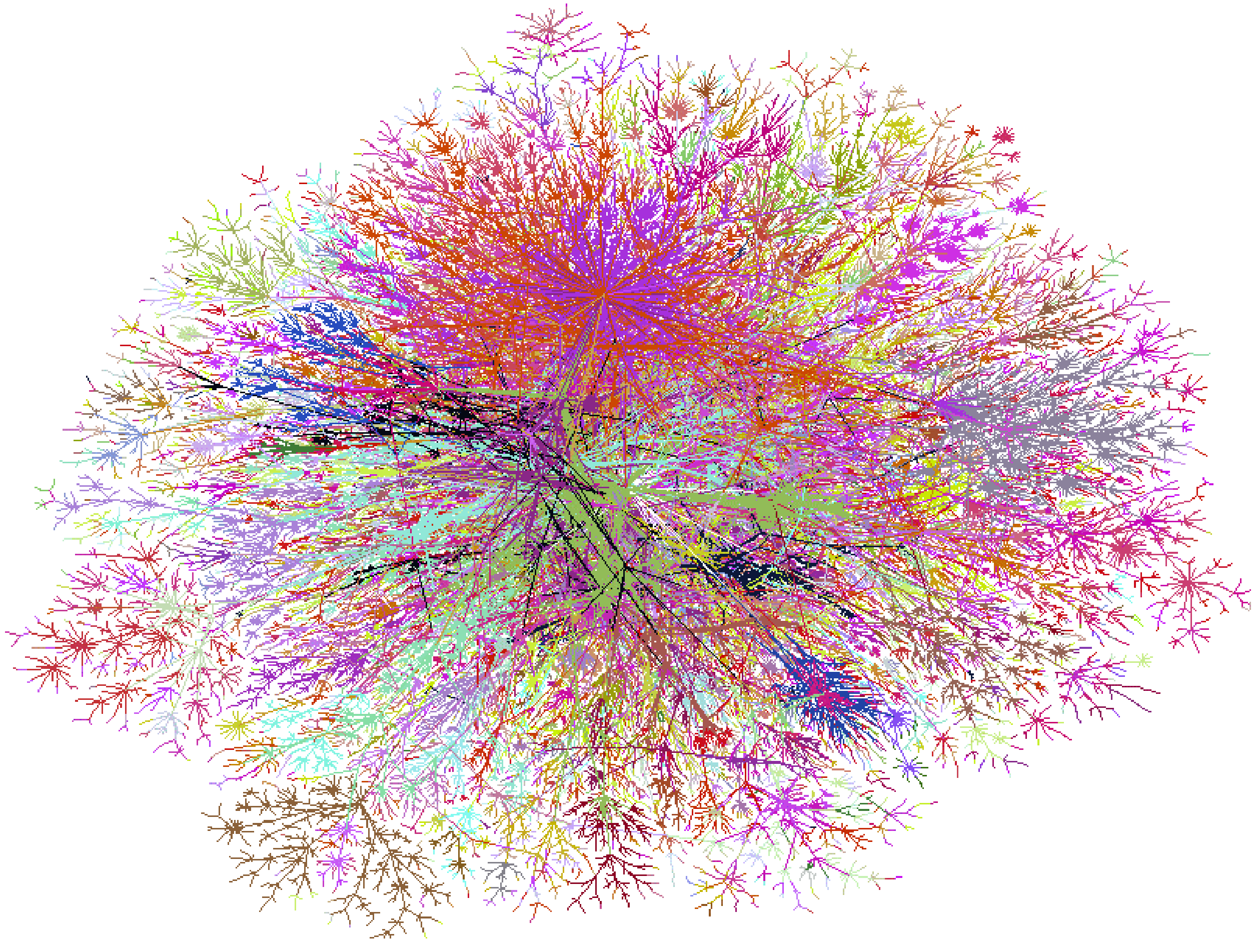
Now networks

Network theory: first some examples



From: Eliot Andres' blog "Plotting your Facebook friend network using NetworkX and python-Louvain"
https://ndres.me/post/friend-graph-tutorial/

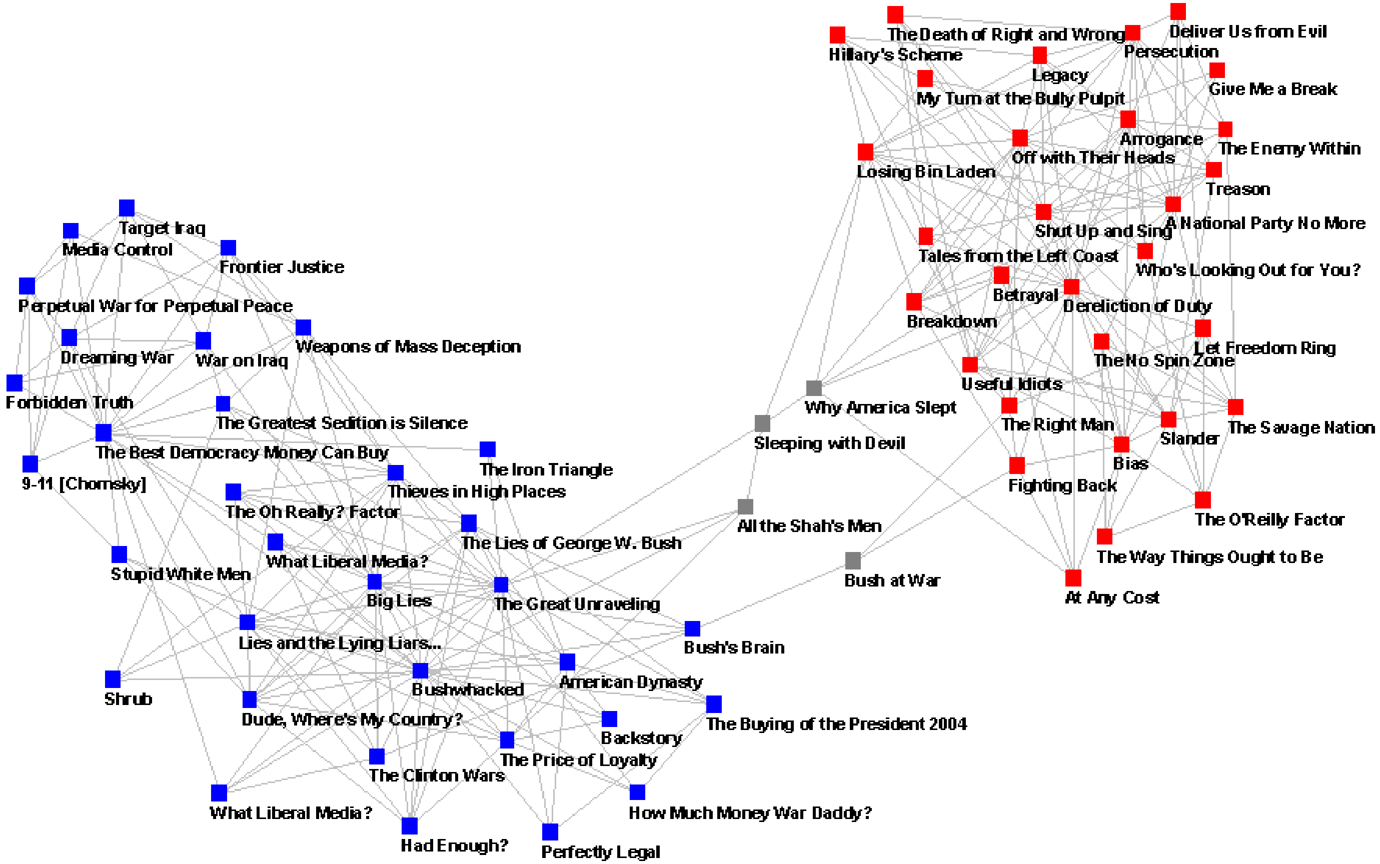
Network theory: first some examples



From: Mark Newman's gallery of network images,
http://www-personal.umich.edu/~mejn/networks/

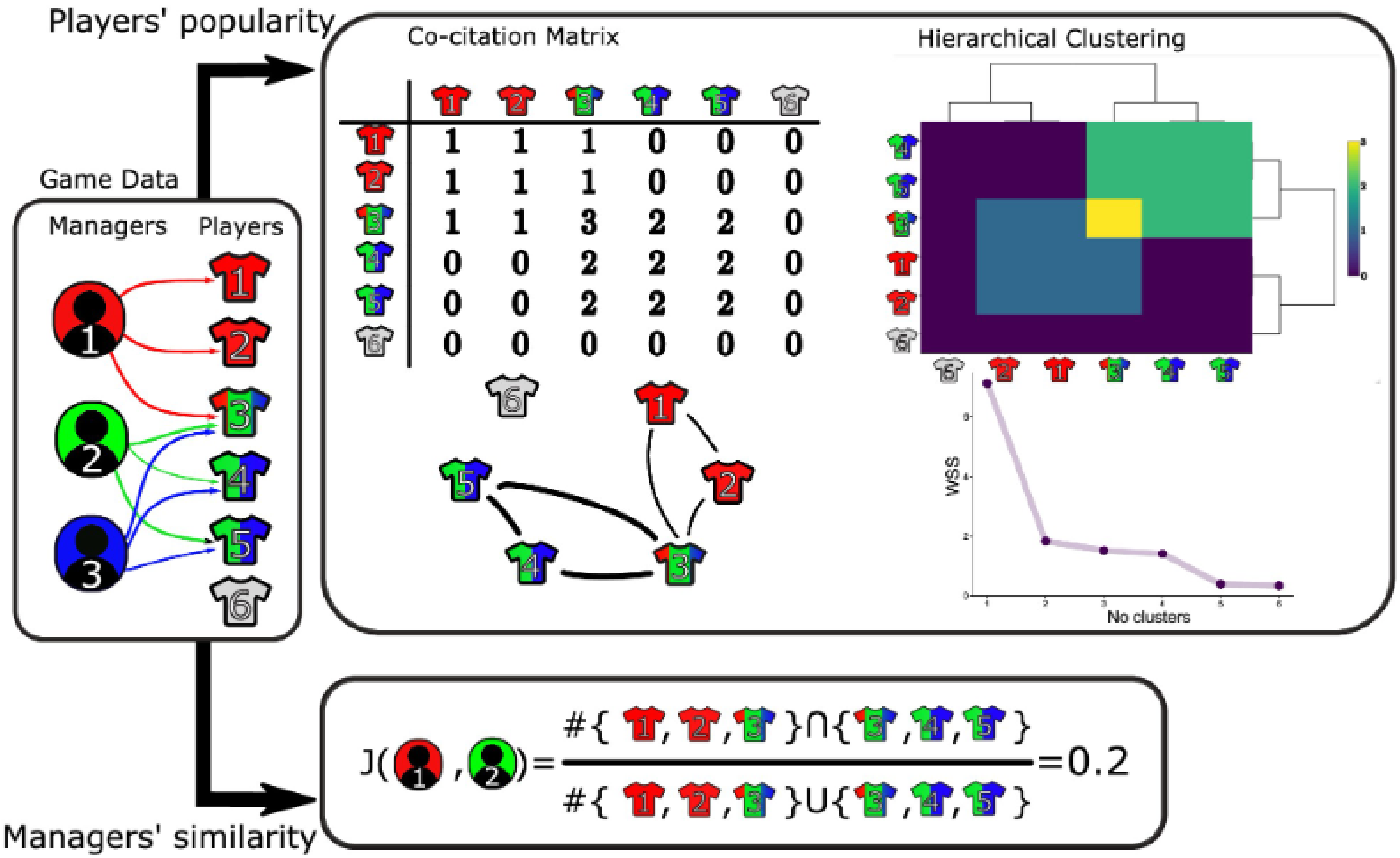
Network theory: first some examples

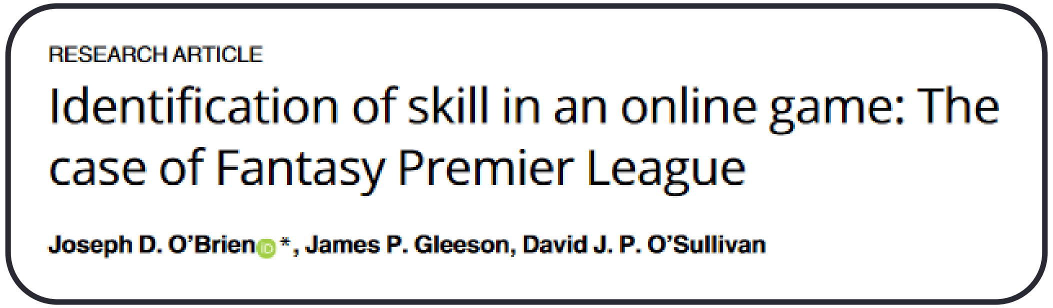
Books
online



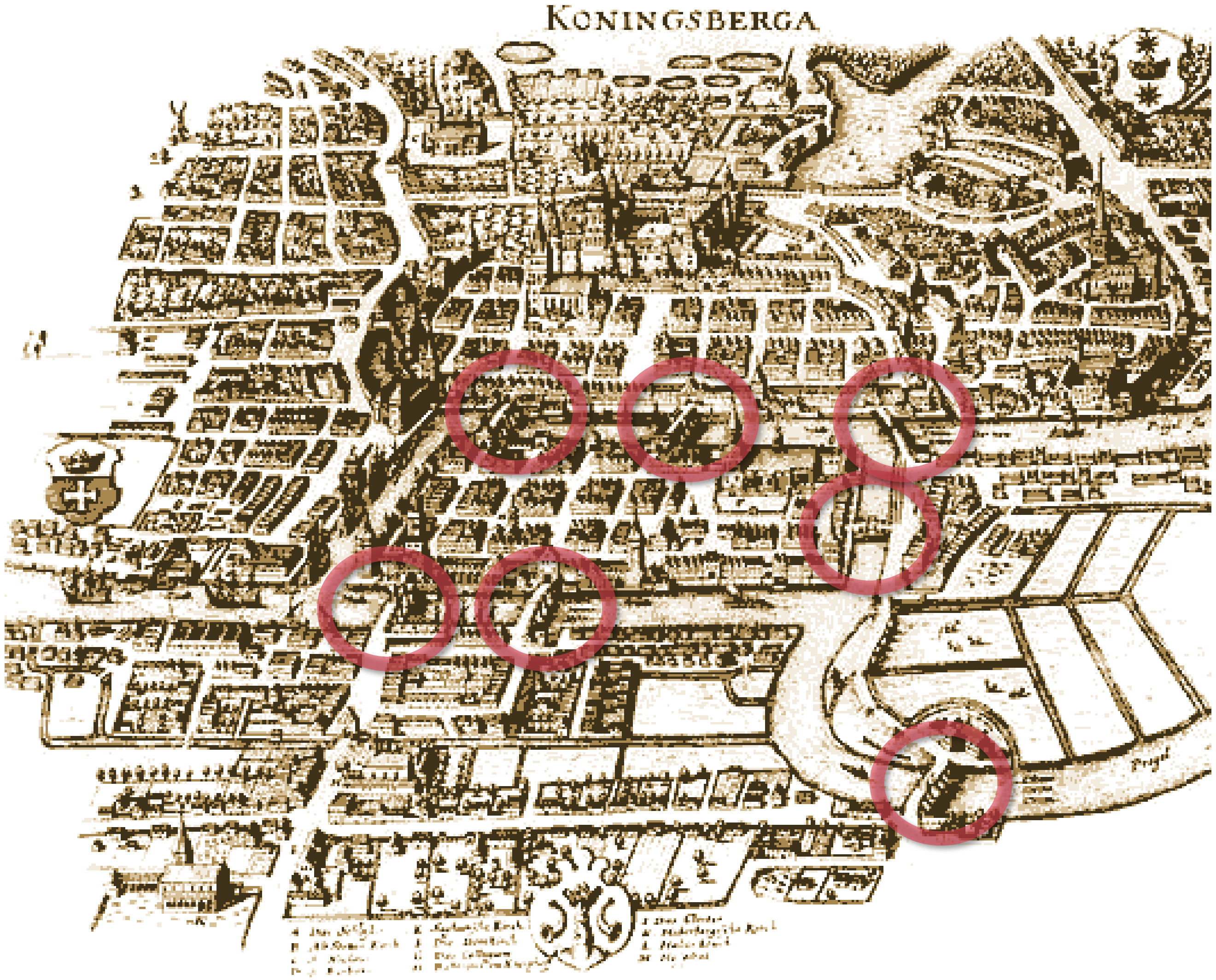
From: Mark Newman's gallery of network images,
http://www-personal.umich.edu/~mejn/networks/

Network theory: first some examples



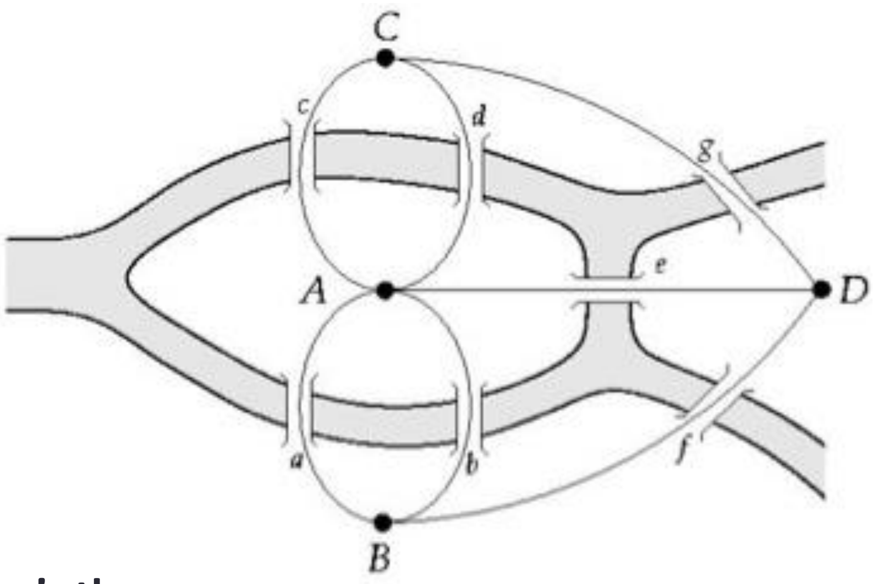


One of the very first examples: The bridges of Konigsberg



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

One of the very first examples: The bridges of Konigsberg



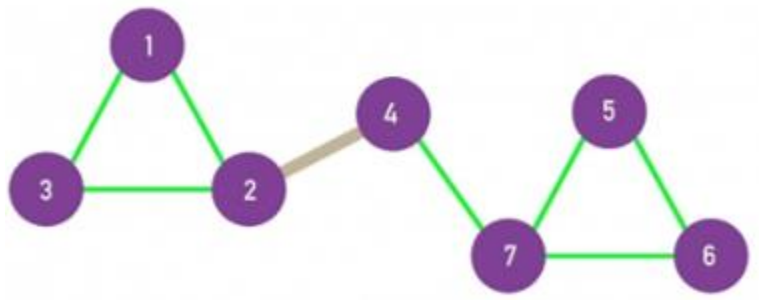
Can one walk across the seven bridges and never cross the same bridge twice?

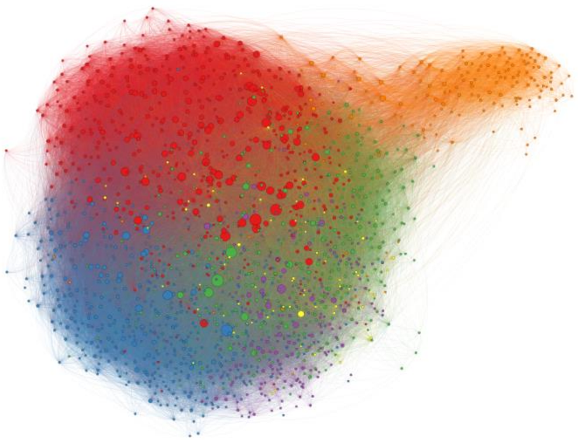
1735: Euler's theorem:

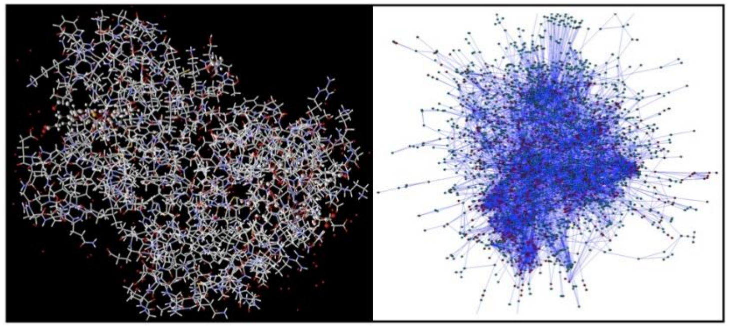
(a) If a graph has more than two nodes of odd degree, there is no path.

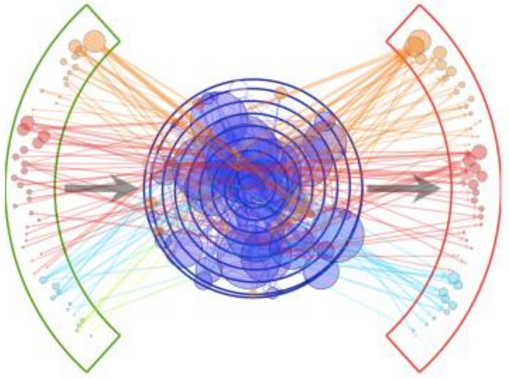
(b) If a graph is connected and has no odd degree nodes, it has at least one path.

Basics of network theory





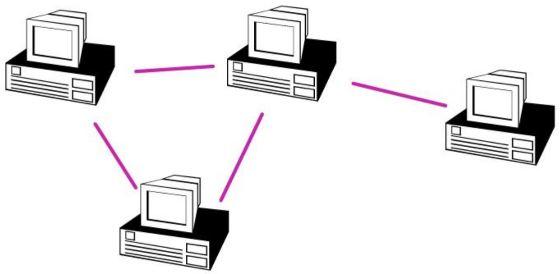


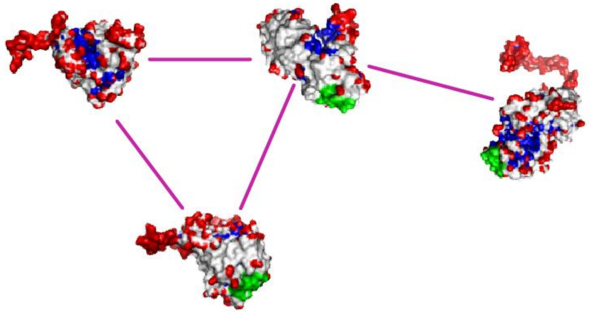


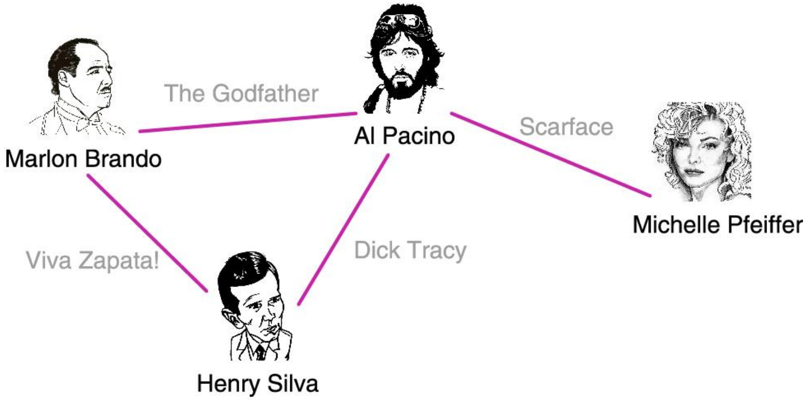
- components: nodes, vertices $\{V\}$
- interactions: links, edges $\{E\}$
- system: network, graph $G = \{V, E\}$

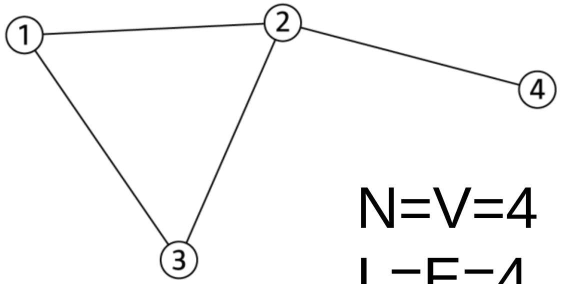
Basics of network theory

- A common language, a common tool kit









N=V=4

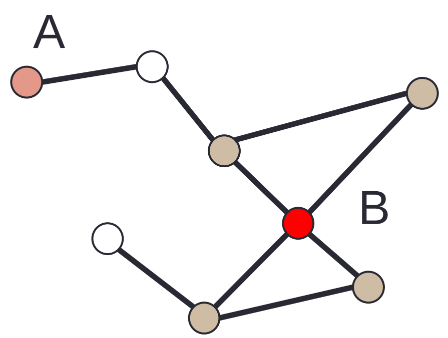
L=E=4

Basics of network theory

- Node degree: number of links that a node has

$$k_B = 4$$
$$k_A = 1$$

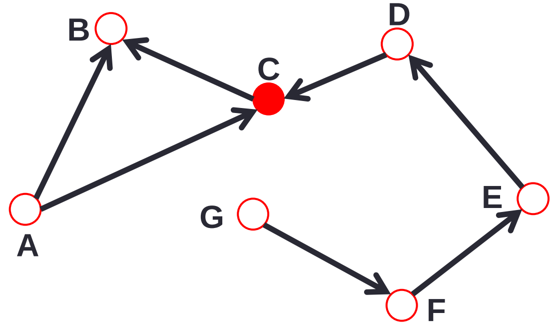
Undirected



- *Directed networks*
  - in-degree and out-degree
  - The (total) degree is the sum of in- and out-degree

$$k_C^{in} = 2$$
$$k_C^{out} = 1$$
$$k_C = 3$$

Directed



Some simple summary stats

Average (mean):

- $E[X] =< X >= \frac{x_1+\cdots+x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i$

The n^{th} moment:

- $\mathrm{E}[X^n] \ =< X^n >= \frac{x_1^n+\cdots+x_N^n}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i^n$

Standard deviation:

- $\sigma_X = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\,x_i - E[X])^2}$

Distribution of X:

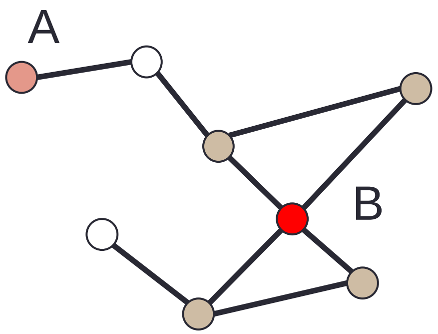
- $p_x = \frac{1}{N}\sum_i \delta_{x,x_i}$

Basics of network theory

- Average degree

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

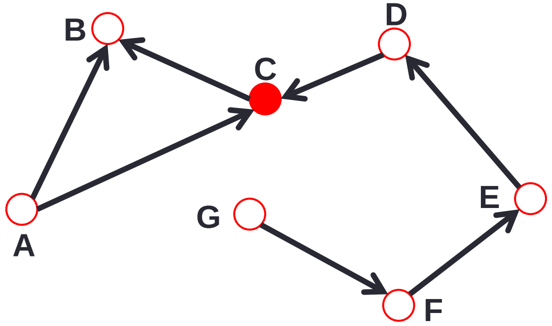
N – the number of nodes in the graph



Undirected

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

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

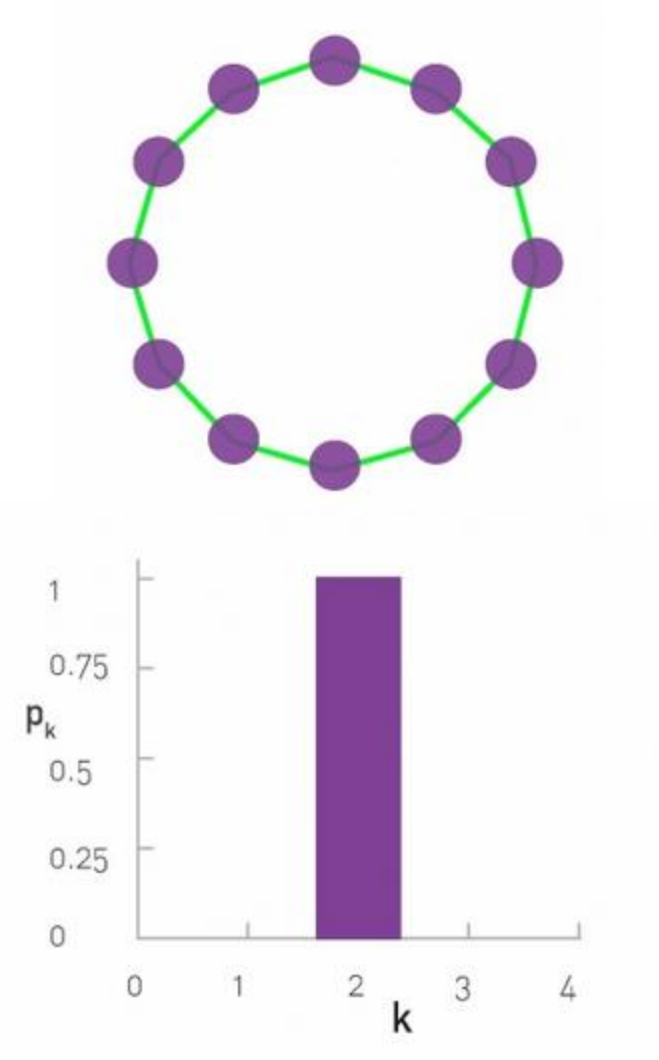


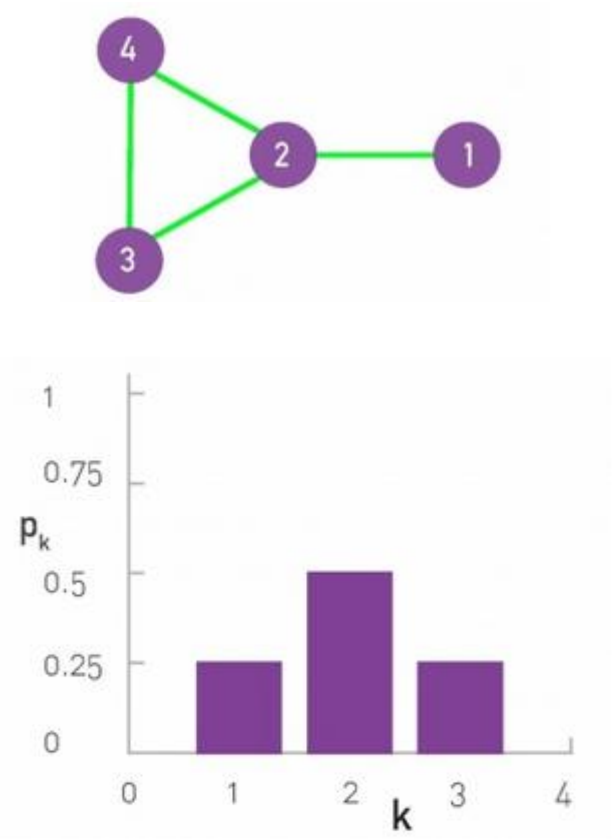
Directed

Basics of network theory

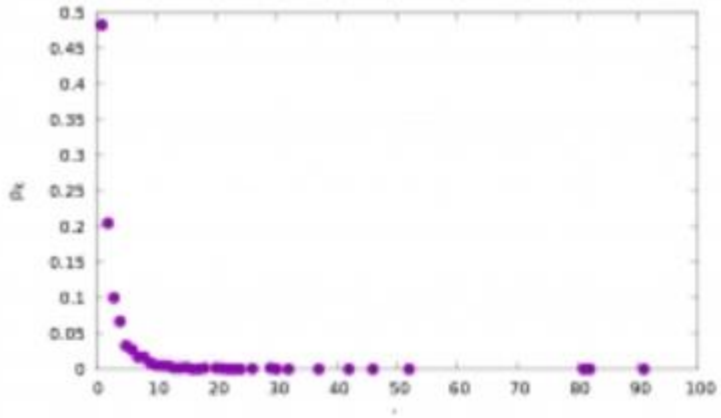
NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L	⟨k⟩
Internet	Routers	Internet connections	Undirected	192,244	609,066	6.33
WWW	Webpages	Links	Directed	325,729	1,497,134	4.60
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594	2.67
Mobile Phone Calls	Subscribers	Calls	Directed	36,595	91,826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

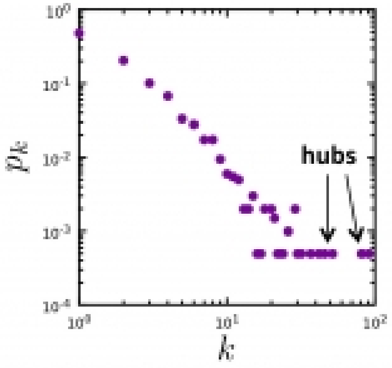
Basics of network theory





Degree distribution

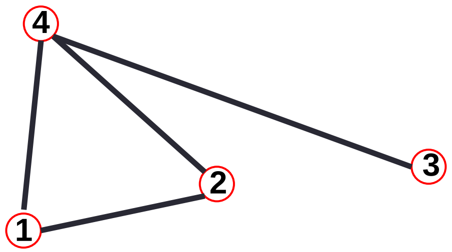


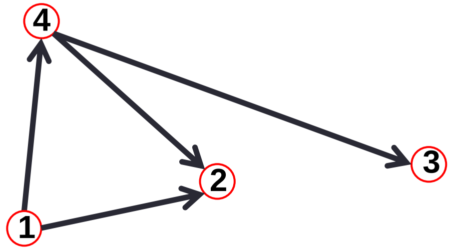


Adjacency matrix

- But to do computation would we need a good repentations

- Adjacency matrix ($\mathbf{A}$):
  - an $N \times N$ matrix whose elements are given by
  - $A_{ij} = \begin{cases} 1 & \{v_i, v_j\} \in E, \\ 0 & \{v_i, v_j\} \notin E. \end{cases}$

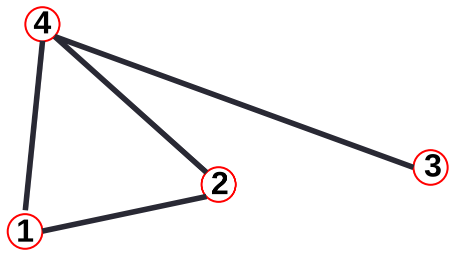




$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \qquad A_{ij} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

Adjacency matrix and node degrees

Undirected



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

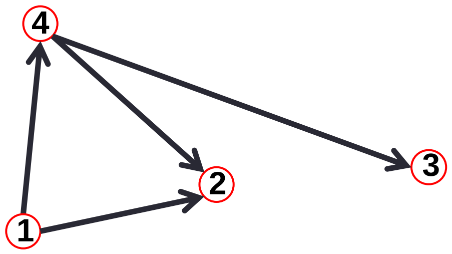
$$A_{ij} = A_{ji}$$
$$A_{ii} = 0$$

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

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

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

Directed



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

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

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

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

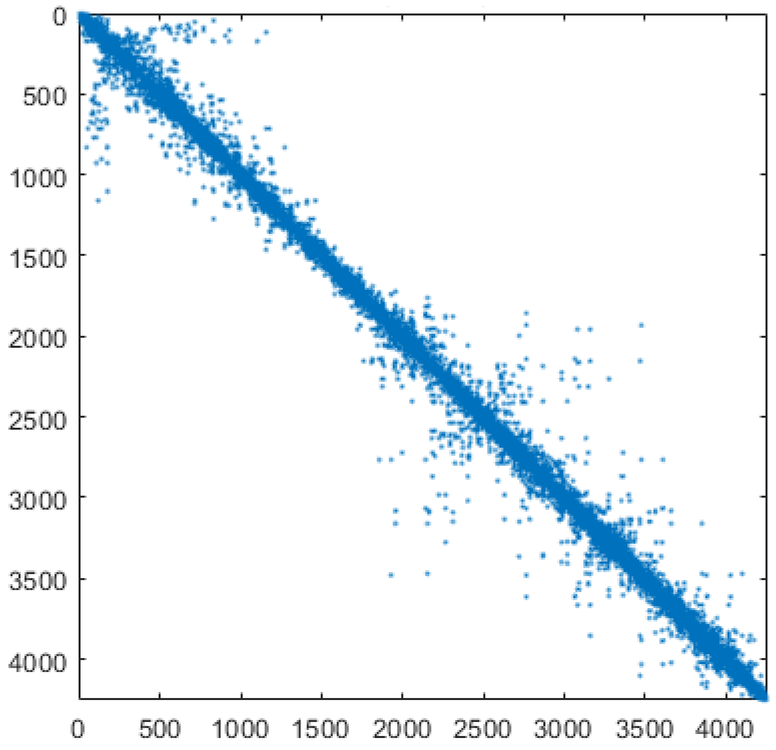
$$L = \sum_{i=1}^{N} k_i^{in} = \sum_{j=1}^{N} k_j^{out} = \sum_{i,j}^{N} A_{ij}$$

Adjacency matrix are sparse

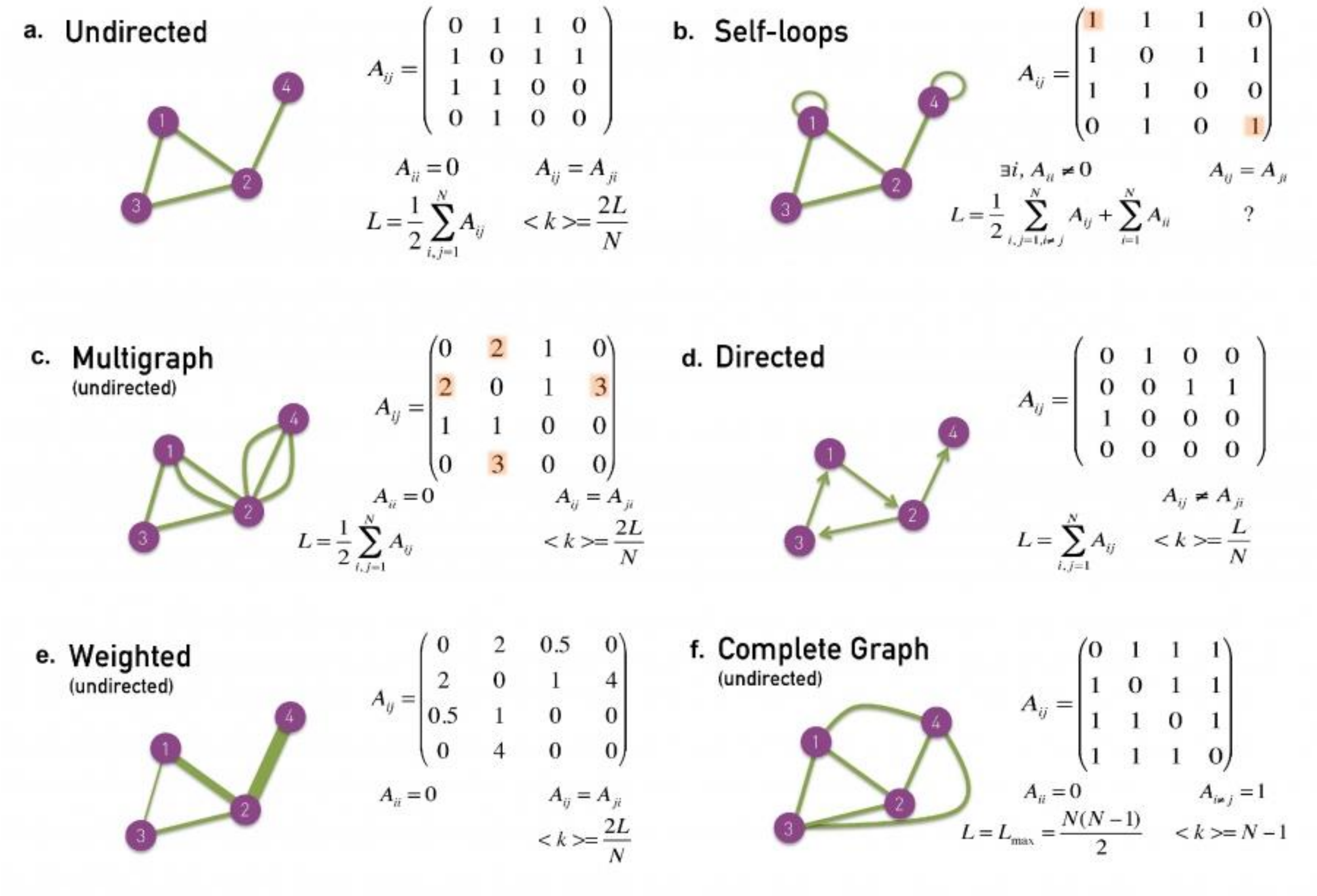
- Real empirical network are sparse, and so is its adjacency matrix

- Most possible links do not exist

- So most of the entries in the matrix are zero

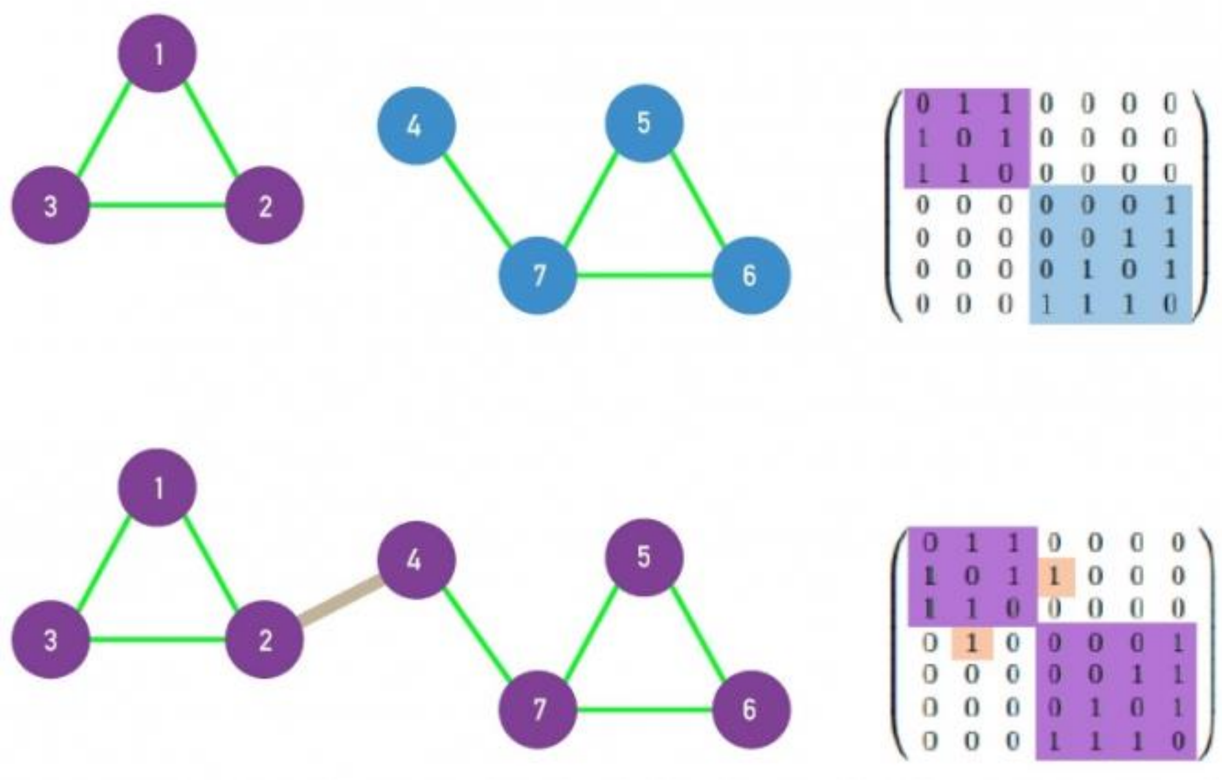


Basics of network theory



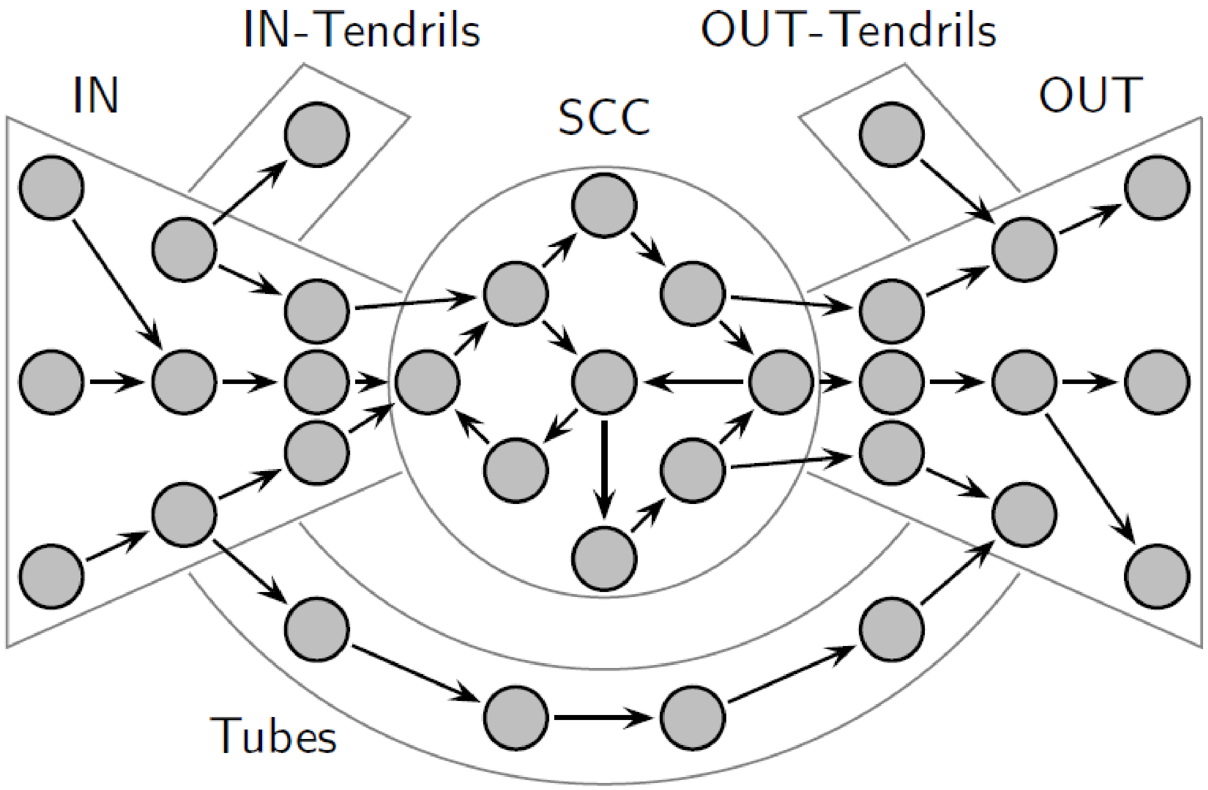
Basics of network theory

- Connectedness
  - Is the graph fractured into smaller parts?
  - One big group?
  - Fully connected



Basics of network theory

- The bow tie



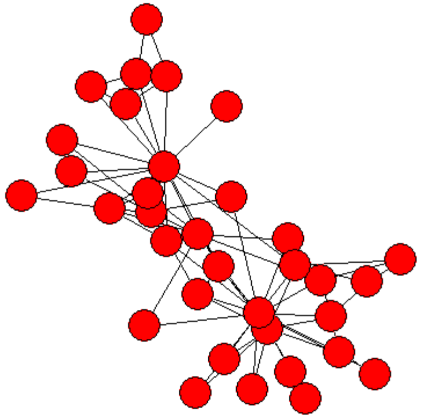
Basics of network theory

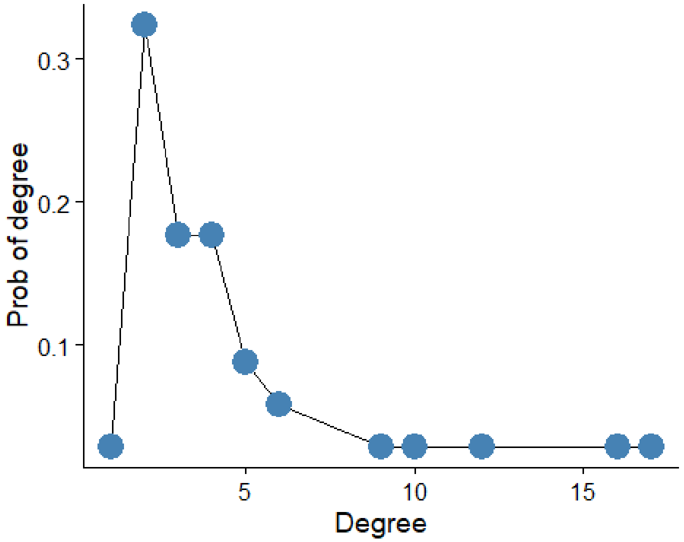
- Degree distribution

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

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

- Regular graph?
- Presentations of hubs?
  - Impacts how process spreads on the networks



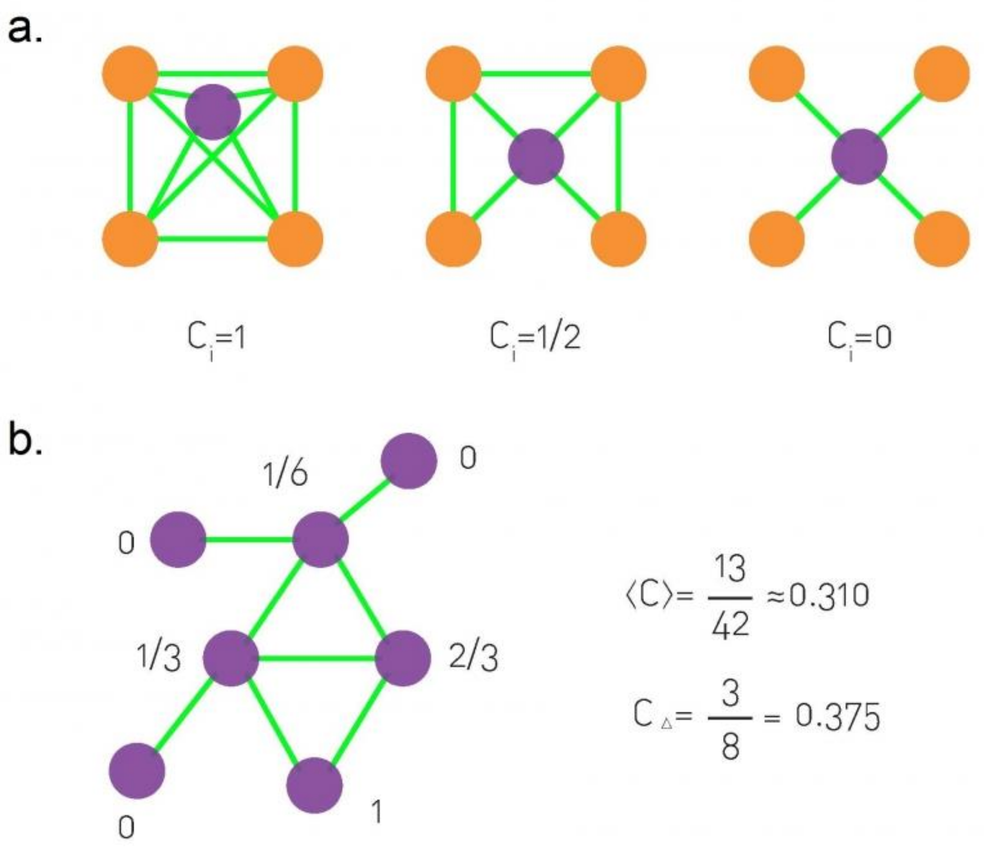


Basics of network theory

$$C_{\Delta} = \frac{3 \times N_{\Delta}}{N_3}$$

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

- Really useful for social networks



Basics of network theory

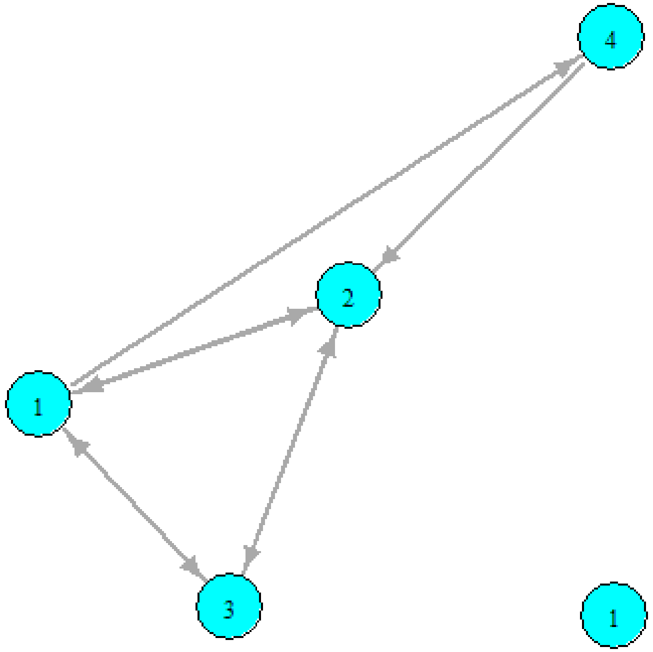
- Mutual links (also referred to as reciprocal links)

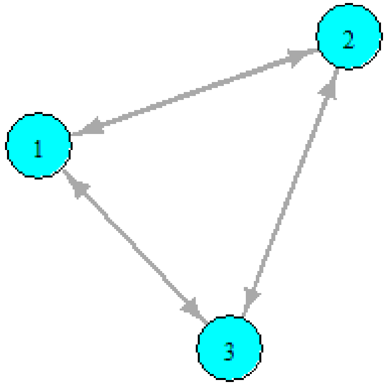
- Important social

$$r = \frac{1}{|E|}\sum_{ij} A_{ij}\, A_{jl} = \frac{1}{|E|} Tr A^2$$

- Filter networks by mutual links can reduce the computational cost of calculations drastically

- And no to forget shortest paths!





A little break to play with R

- Creating and plotting simple graphs in R

- 1_networks_basics_and_Igraph.R

igraph – The network analysis package

- Handy package for doing lots of network analysis
  - Easy to code in and fast
  - Alternatives -> networkx and graphtool

CREATING NETWORK FROM DATA