Network Science
SICSS Summer School

Instructors

Vincent Buskens
Sociology
Utrecht University

Jiamin Ou
Sociology
Utrecht University
(on leave this year)

Leto Peel
Data Analytics & Digitalisation
Maastricht University

Javier Garcia-Bernardo
Methodology & Statistics
Utrecht University

Mahdi Shafiee Kamalabad
Methodology & Statistics
Utrecht University

Elena Candellone
Methodology & Statistics
Utrecht University

Participants

Why are you taking this summer school?

Program

Monday:

Introduction to networks

Network centrality

Tuesday:

Statistical models (Mahdi)

Link prediction (Javier & Leto)

Wednesday:

Network inference (Mahdi)

Thursday:

Network models (Leto)

Community detection (Leto)

Friday:

Dynamics in networks (Vincent)

Day program

09:00–10:00:

Introductions

10:00–12:00:

Introduction to network science

Practical + discussion

12:00-13:00

Lunch

13:00–16:30:

Network representation

Centrality

Introduction to networks

Network game

Introduce yourself, and find one thing you have in common:

- Countries (apart from the NLD) that you have lived in
- Favorite cuisine
- Sports you practice
- Programming languages you use
- …

Draw a line in the whiteboard, write the names in this spreadsheet: https://tinyurl.com/network-game

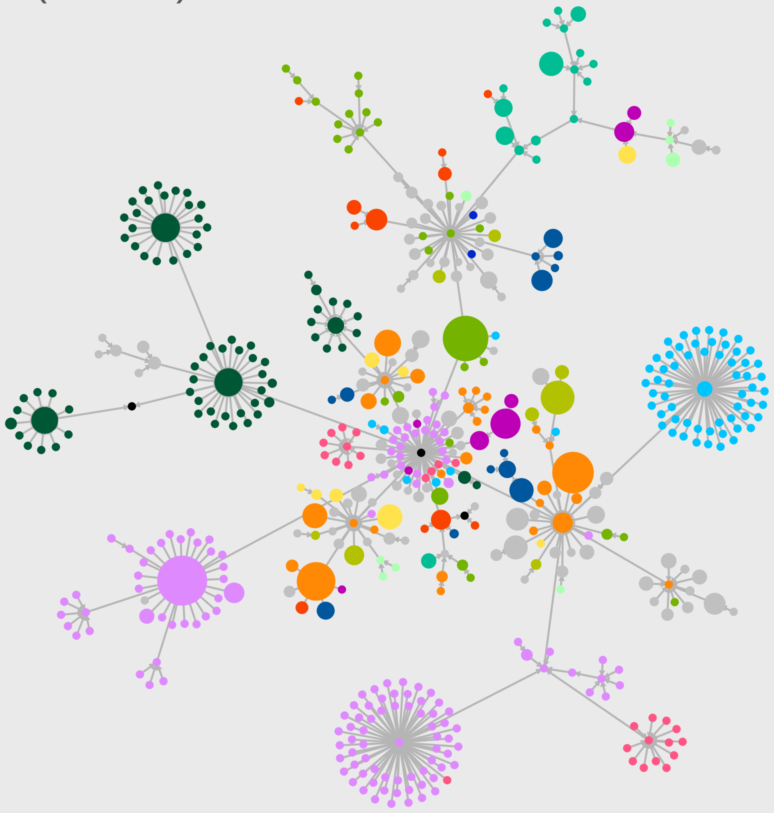
It's a public file, please don't use your full name

What is a network?

Mathematical representation of the relationships (edges) between entities (nodes)

The most important question to ask yourself:

What are the nodes and what are the edges?



Types of networks

	Network	Nodes	Edges
Social/ Behavioral	Friendship	People	Friendships
	Follower	Online accounts	Followers/likes
	Psychological	Symptoms	Co-ocurrence
Biology	Gene regularory	Genes	Activations/inhibations
	Food web	Animals	Predating
Economic	Trade	Countries/companies	Money flows
	Ownership	Companies	Ownership stakes
Intrastructure	Internet	Computers (IPs)	Data transmission
	Power grid	Power stations	Power lines
	Airplane network	Airports	Flights

https://aaronclauset.github.io/courses/5352/csci5352_F21_L1.pdf

Type of networks and characteristics

Type 1: Interaction and flow → “Real networks”.

- Offline interactions
- Online interactions

Type 2: Affiliation → Node 1 is part of/related to node 2

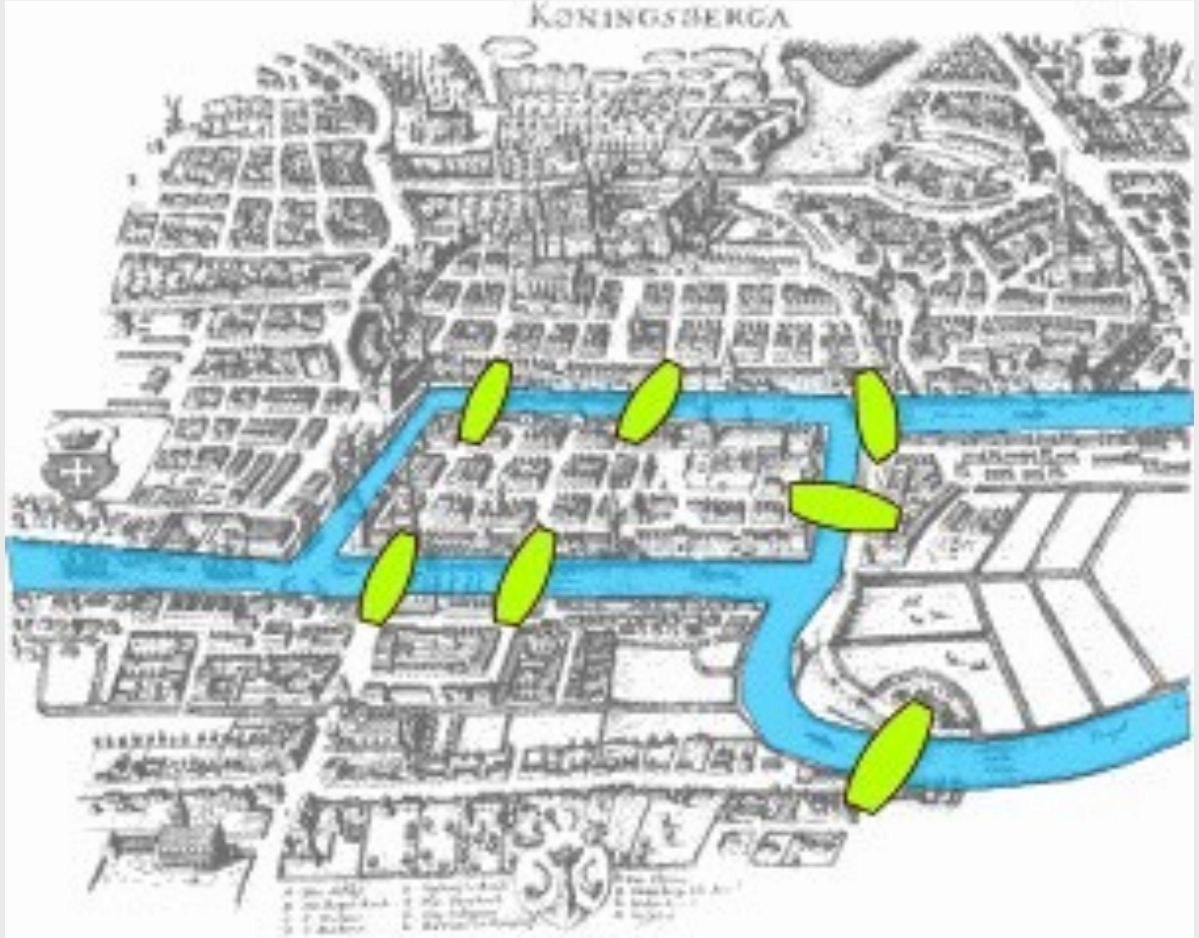
- Most administrative data: e.g. students in classrooms
- Bipartite networks

Type 3: Co-occurrence → Node 1 is correlated with node 2

- Stock market networks
- Brain networks

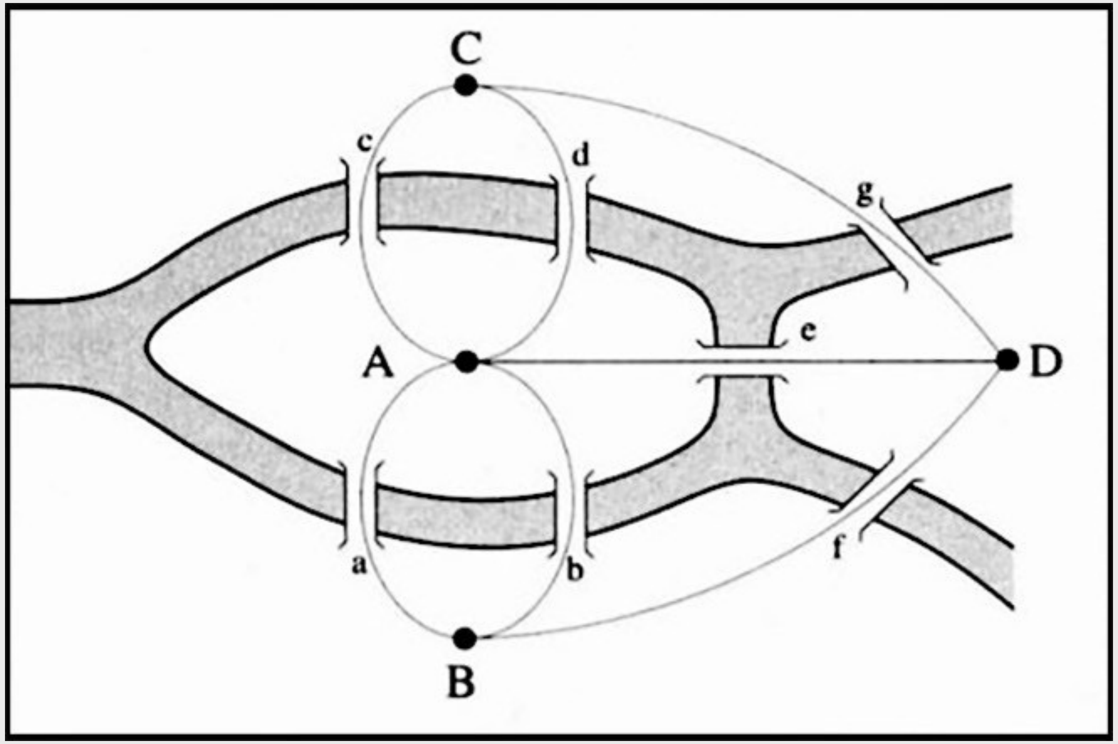
What about family networks?

Bridges of Königsberg

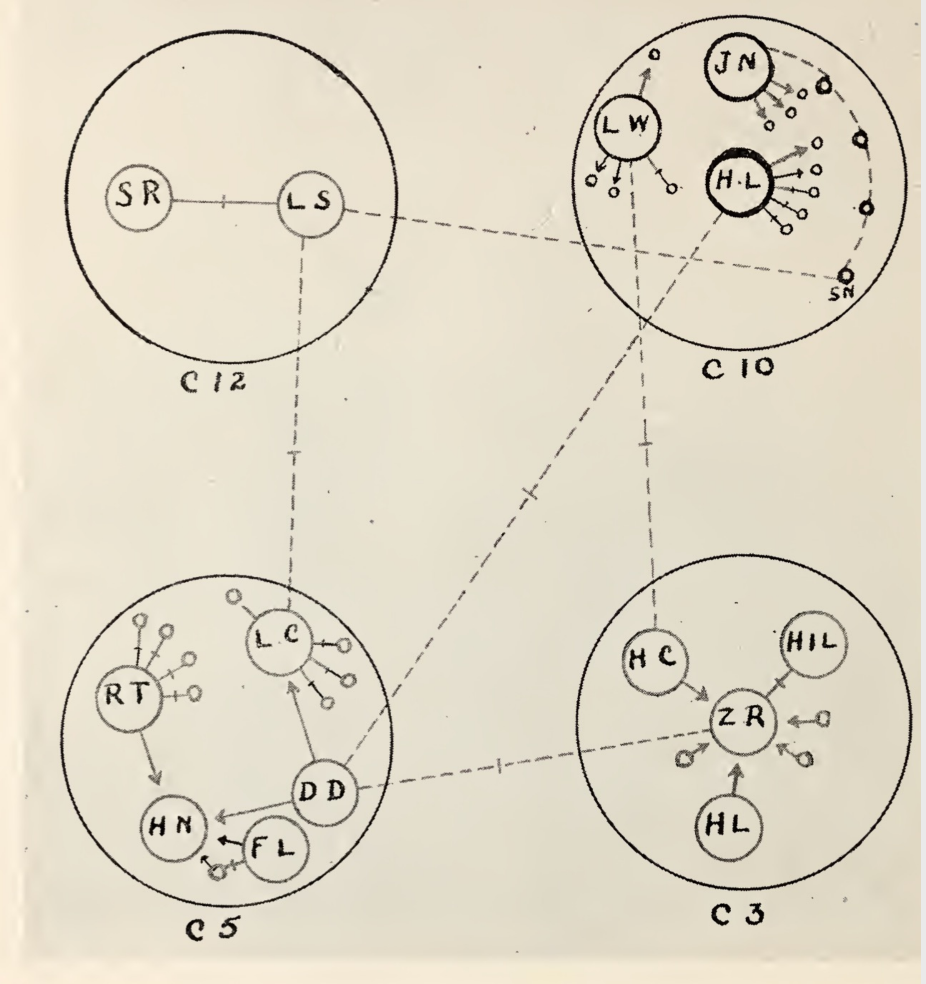


Is there a way you crosses each bridge exactly once and returns to the starting point

Euler (1736)



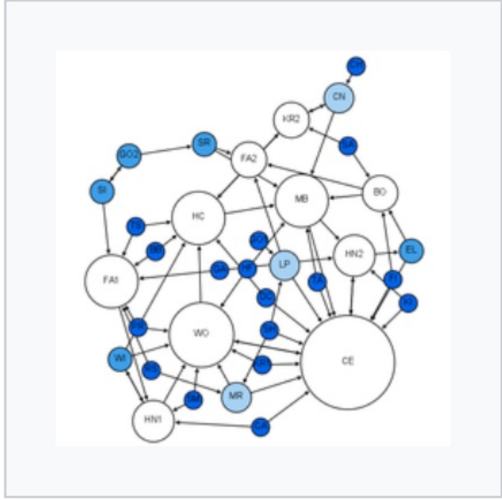
Network Science has roots in sociology



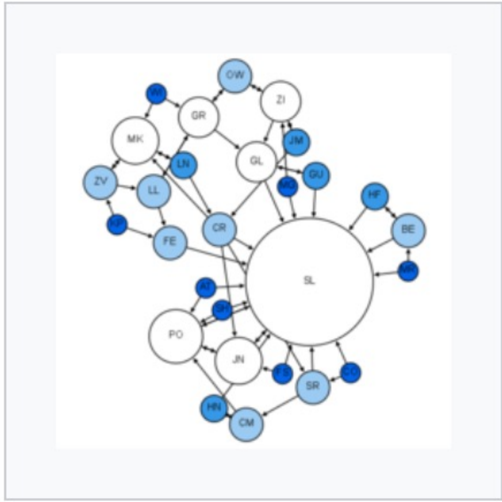
Moreno. Who shall survive?

Network science: Social and behavioral scientists in the XX century (e.g. Jacob Moreno & Hellen Hall Jennings, Harrison White, Mark Granovetter)

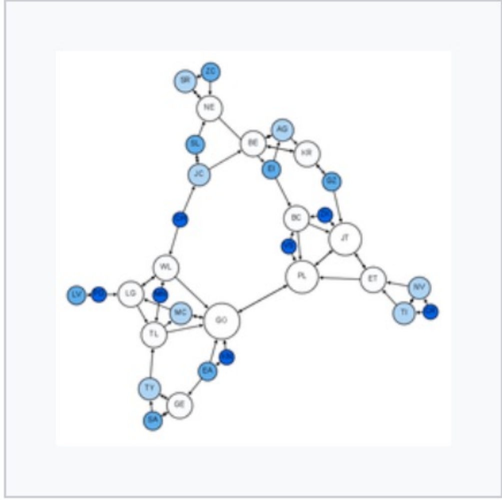
- Hellen Hall Jennings and Jacob Moreno (1930s): Hudson School for girls: Sociometry. Networks can represent the systems and how information spreads
- Manfred Kochen & Ithiel de Sola Pola (1950s), Jeffrey Travers and Stanley Milgram's (1969): Small-world studies
- Nancy Howell (1969): *The Search for an Abortionist*, women acquired scarce information through short chains of weak ties.
- Mark Granovetter (1973) *The Strength of Weak Ties*. Diffusion of information takes place primarily through bridges (weak ties). Strong links are redundant.
- François Lorrain & Harrison White (1976): Blockmodels for networks
- Duncan Watts, Steven Strogatz (1998): Next wave of network science



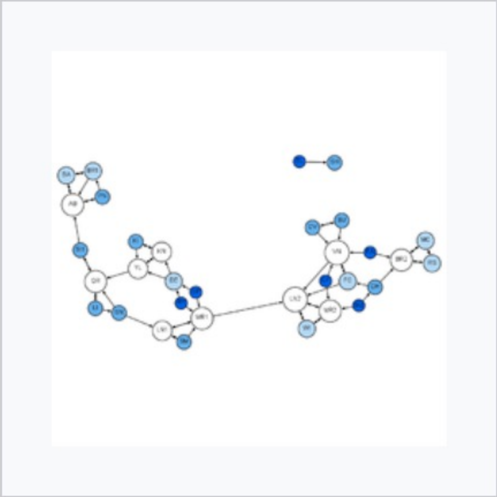
1st Grade



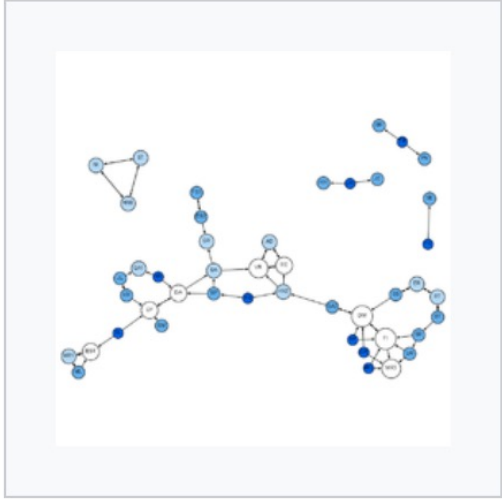
2nd Grade



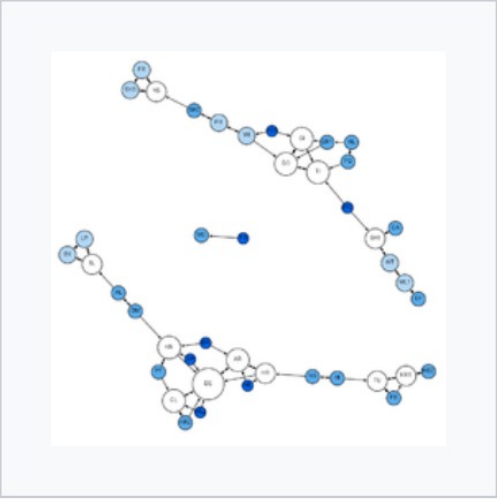
3rd Grade



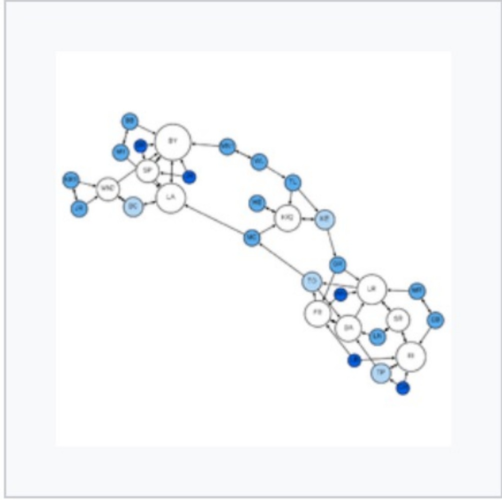
4th Grade



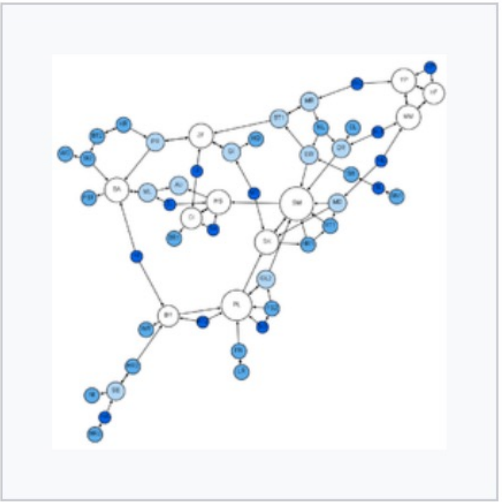
5th Grade



6th Grade



7th Grade



8th Grade

Moreno, source: wikipedia. Edges: who wants to be sitting next to whom?

Why do we care about networks?

Main goal: Understanding Complex Systems: Find insights that we would miss if we would study the nodes independently (one person != society)

Social science: Networks reflect preferences (selection) and influence us (spread of information, diseases, opportunities)

- Networks allow us to study sociological theories (dangerous generalizations below!):
  - Social capital: The position of an individual in their social network (embeddeness) determines opportunities and outcomes. e.g. structural holes and network closure (Burt, 2001)
  - Structural holes: social capital is created by a network in which people can broker connections between otherwise disconnected segments ~ betweenness centrality
  - Network closure: social capital is created by a network of strongly interconnected element ~ clustering coefficient
  - Understanding dynamics (e.g when does information spread?)

Epidemiology: How to stop disease transmission?

Criminology: How to detect crime?

Biotechnology: Understanding disease in gene regulatory network, finding targets to treat diseases.

Ecology: Constrains and dynamics of ecosystems

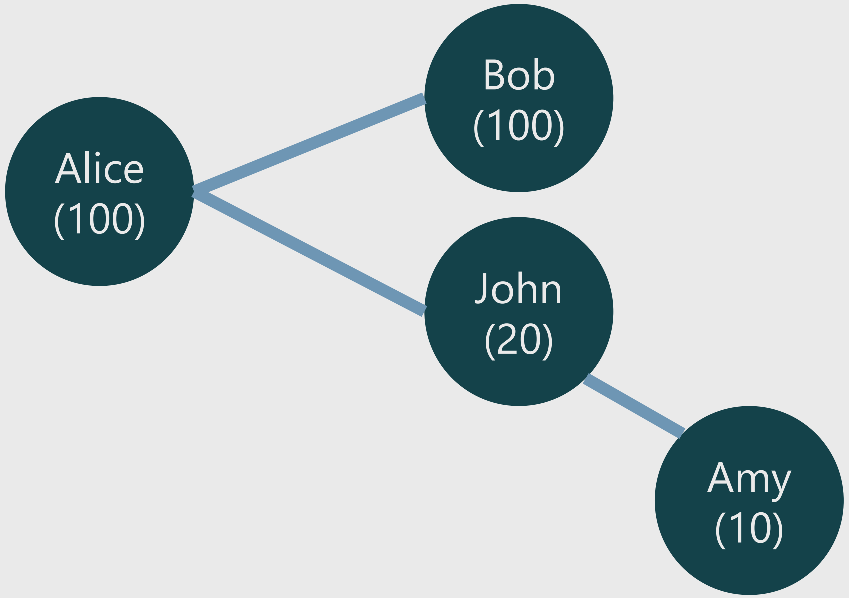
Psychology: How mental processes depend on our interactions (esp. important in experimental psychology)?

Engineering: Improve network performance and reliability (e.g., power grids, transportation)

Physics view: Dependence on topology (reliability, dynamics, emergent behavior and phase transitions)

Basic definitions

Networks (graphs)



Nodes (vertices, actors) connected by **edges** (links, connections, relationships)

N: **Nodes** = {Alice, Bob, John, Amy}

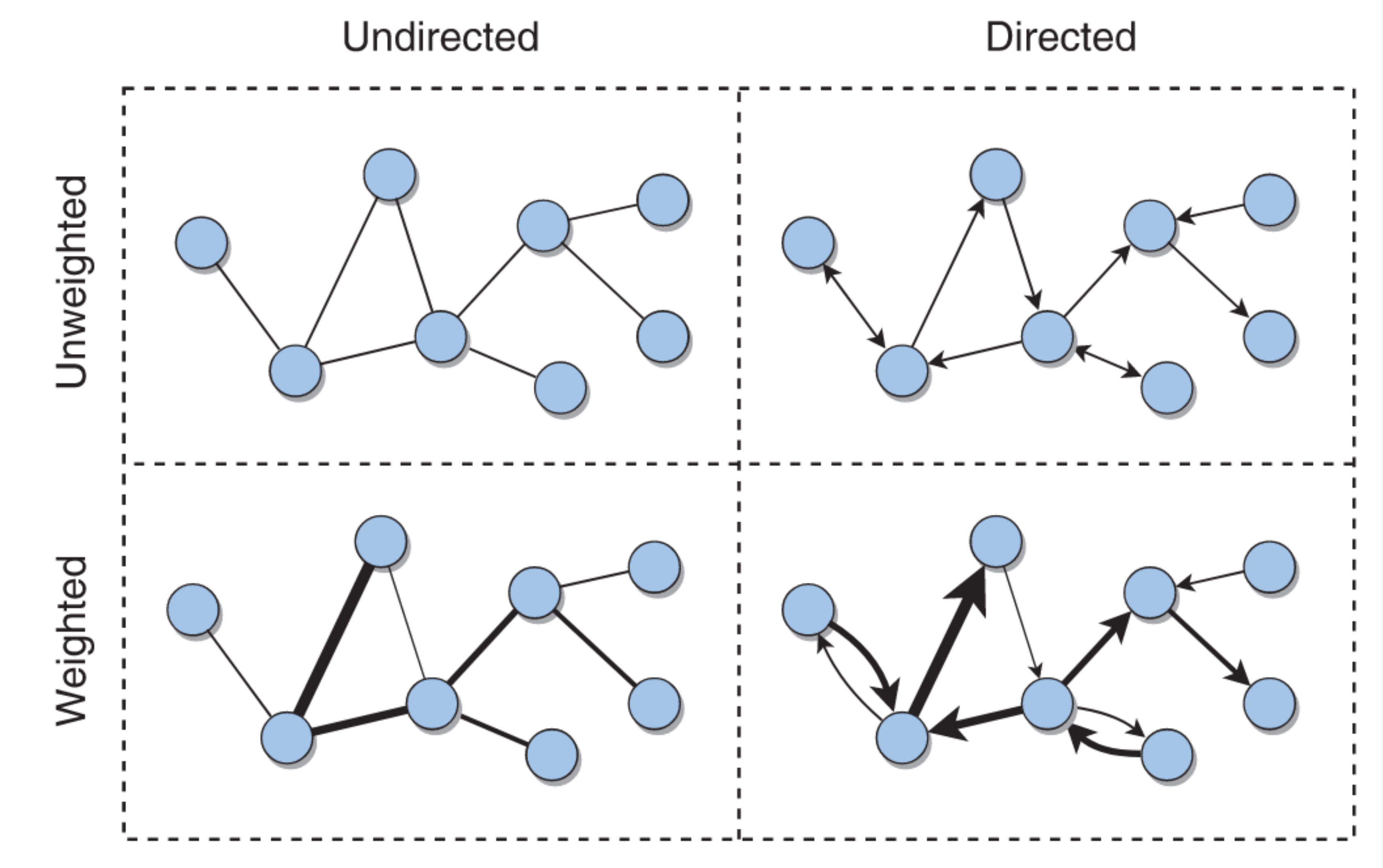
E: **Edges** = {(Alice, Bob), (Alice, John), (John, Amy)}

The edge (i,j) connects node i to node j

Nodes can have **attributes** (e.g. gender, income, etc)

Edges can have **attributes** (e.g. type, strength, etc)

Directed vs undirected; weighted vs unweighted



Source: A first course in network science (2020)

Undirected: The link (i,j) connects node i to node j in both directions

Directed: The link (i,j) connects node i (source) to node j (target)

Weighted: There is a weight associated to each edge

Degree in undirected networks

Definition: Number of neighbors in the network

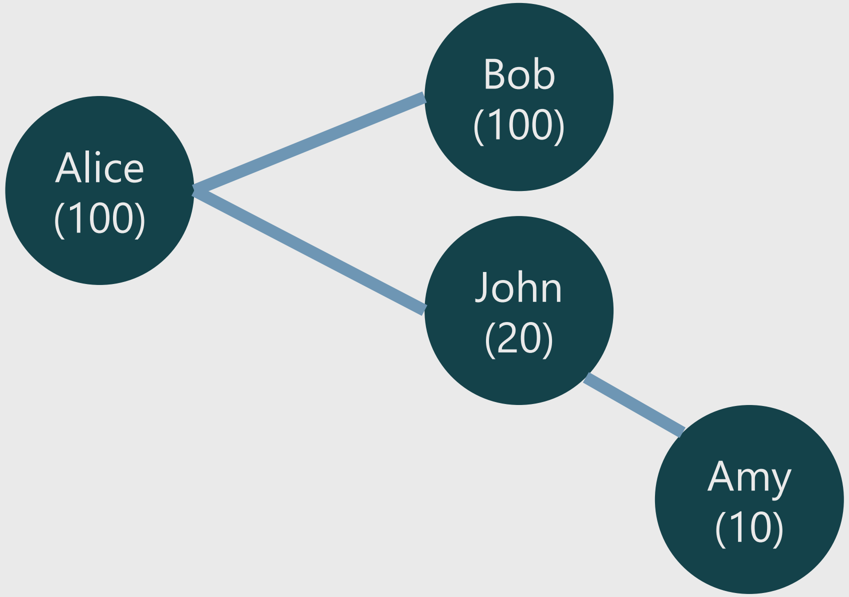
Node: degree

Alice: 2

Bob: 1

John: 2

Amy: 1



Degree in directed networks

Out-degree: Number of outgoing edges

In-degree: Number of incoming edges

Total degree: Sum of out and in degree

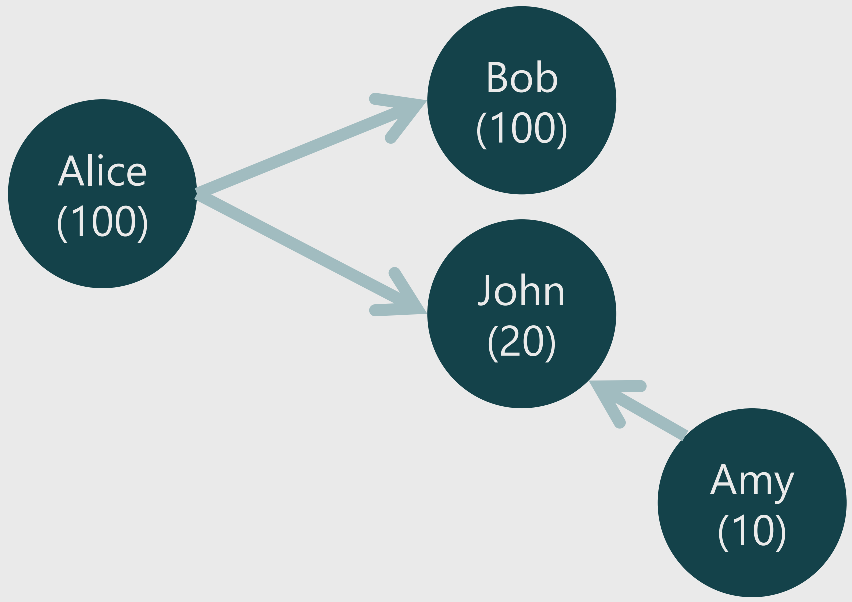
Node: (out, in, total)

Alice: (2, 0, 2)

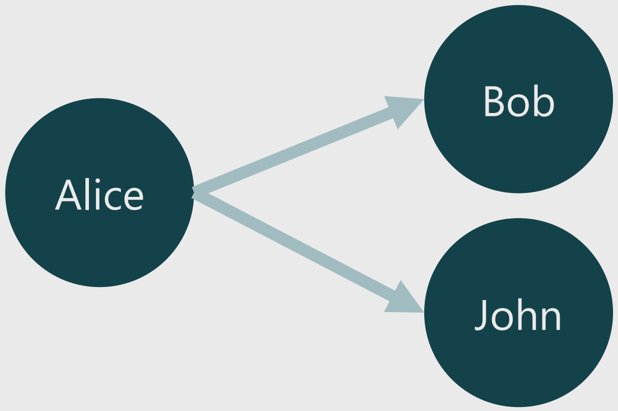
Bob: (0,1, 1)

John: (0,2, 2)

Amy: (1,0, 1)



Network representation



Adjacency list (edgelist):

- Adv: It is dense: Only keeping edges
- Disadvantage: Hard to work with

Origin	Target	Weigth
Alice	Bob	1
Alice	John	1

Adjacency matrix:

- Adv: Linear algebra is easy
- Disadvantage: It is sparse (mostly zeros).1E6 nodes → 1 trillion options

Target → ↓ Source	Alice	Bob	John
Alice	0	1	1
Bob	0	0	0
John	0	0	0

In computer → Sparse matrices: Best of both worlds

Practical 1, exercise 1

1. Download materials:
- https://github.com/jgarciab/NetworkScience
- (click on code -> Download Zip)

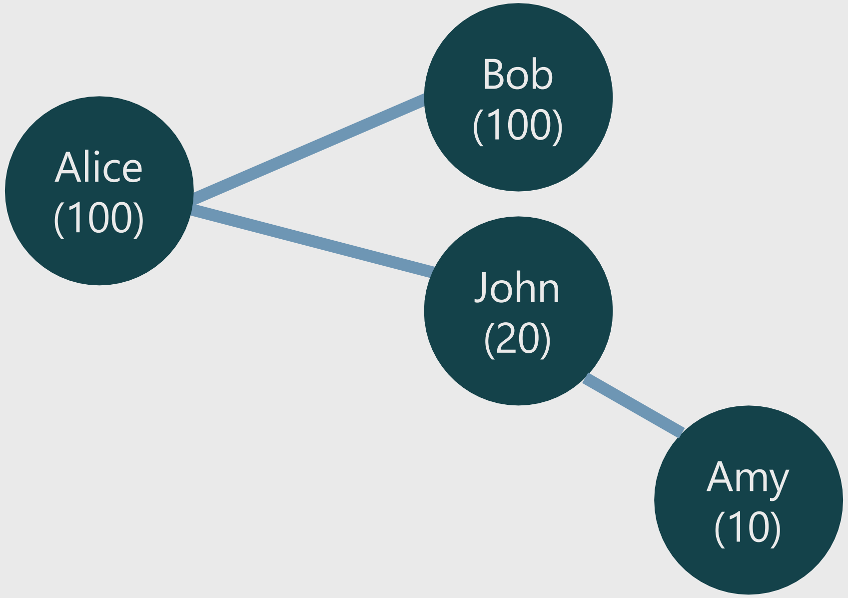
2. Extract ZIP

3. Set up Python. On Windows & Mac using the graphical interface:
- Open Anaconda
- Go to “Environments” (left menu)
- Click on “Import” and specify the file “environment.yml” (it’s one of the files that you downloaded)
- Activate environment by clicking in the “play” button next to the environment.

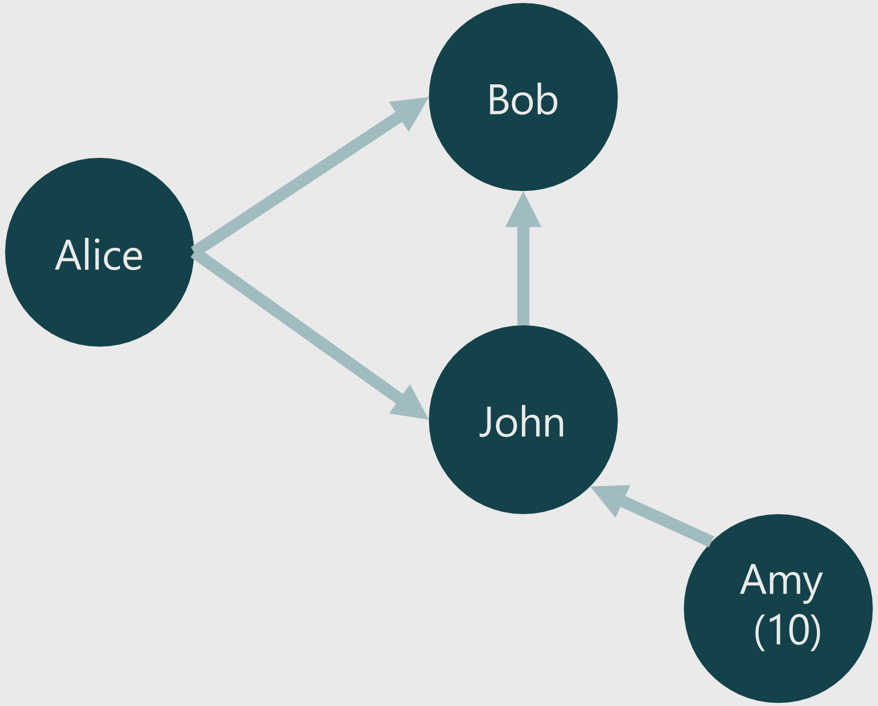
On Linux (also works for Windows/Mac):
- Open a conda terminal (open a terminal)
- Navigate to the directory with the code using dir (ls) to list the files and cd XXX (cd XXX) to enter directory XXX.
- Create a new environment: *conda env create -f environment.yml*
- Activate environment: *conda activate networks*
- Launch jupyter notebook: *jupyter notebook*

4. Do exercise 1 in: day*1a_intro_networks.ipynb*

Types of networks

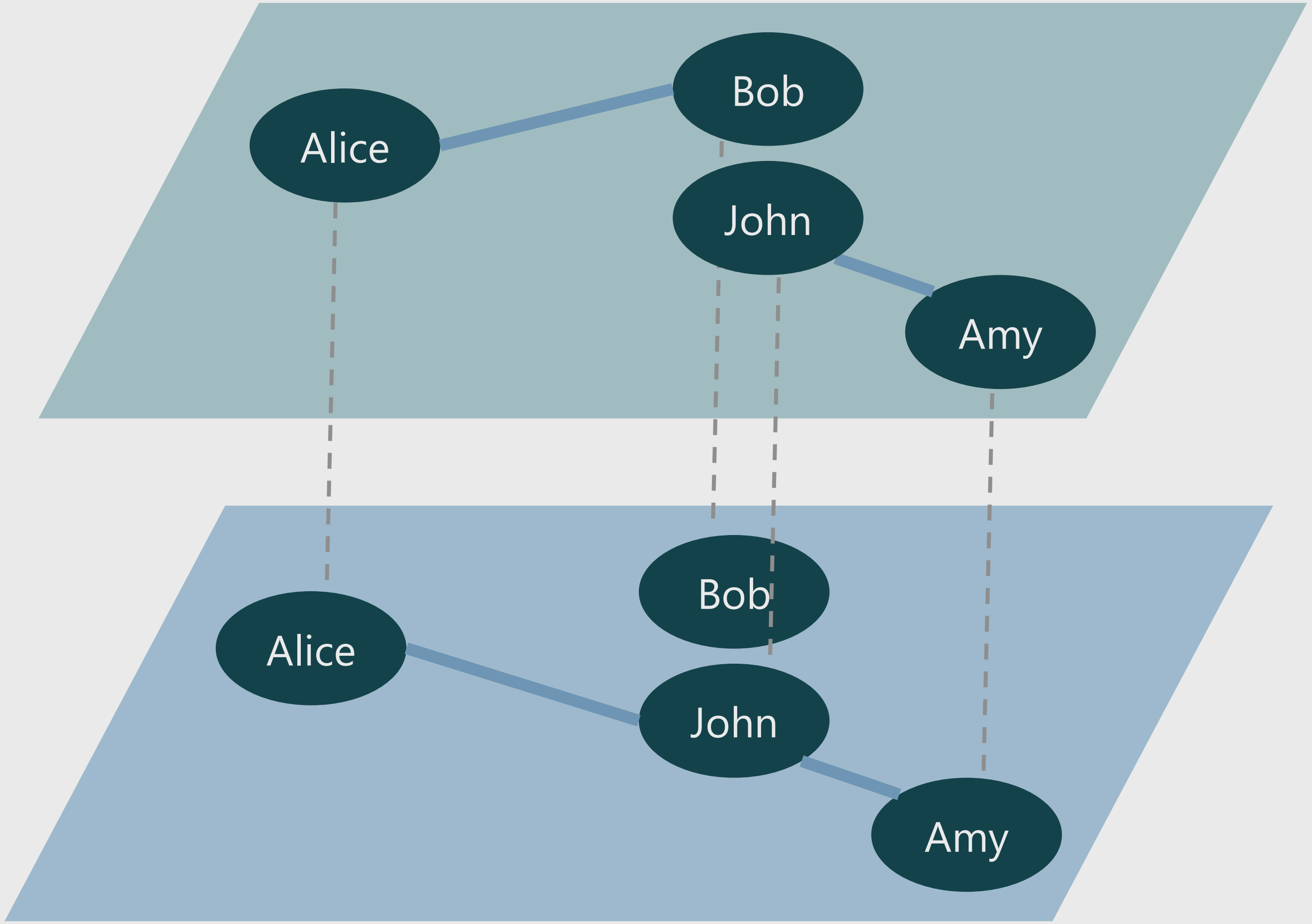


Undirected Acyclic Graphs (Trees)



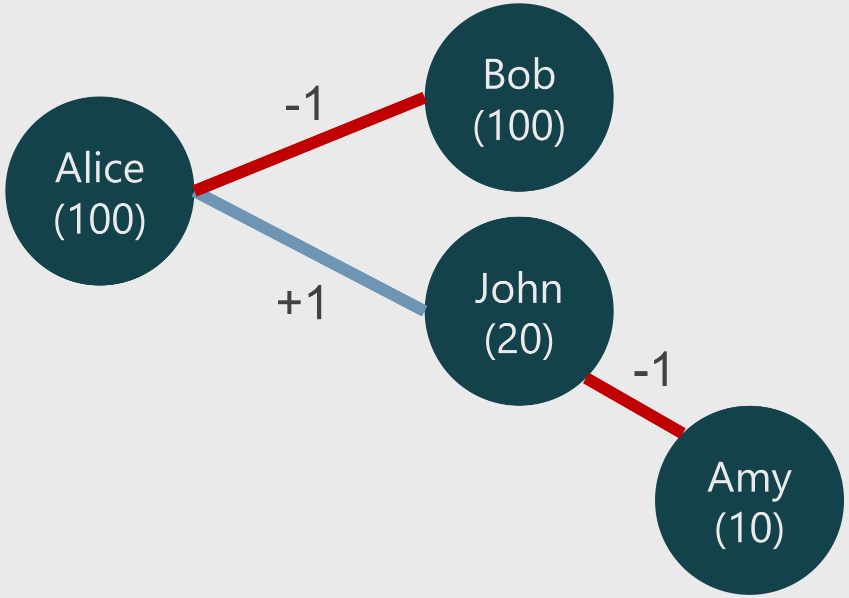
Directed Acyclic Graphs (DAGs)

Other types of networks: Multiplex



Other types of networks: Signed

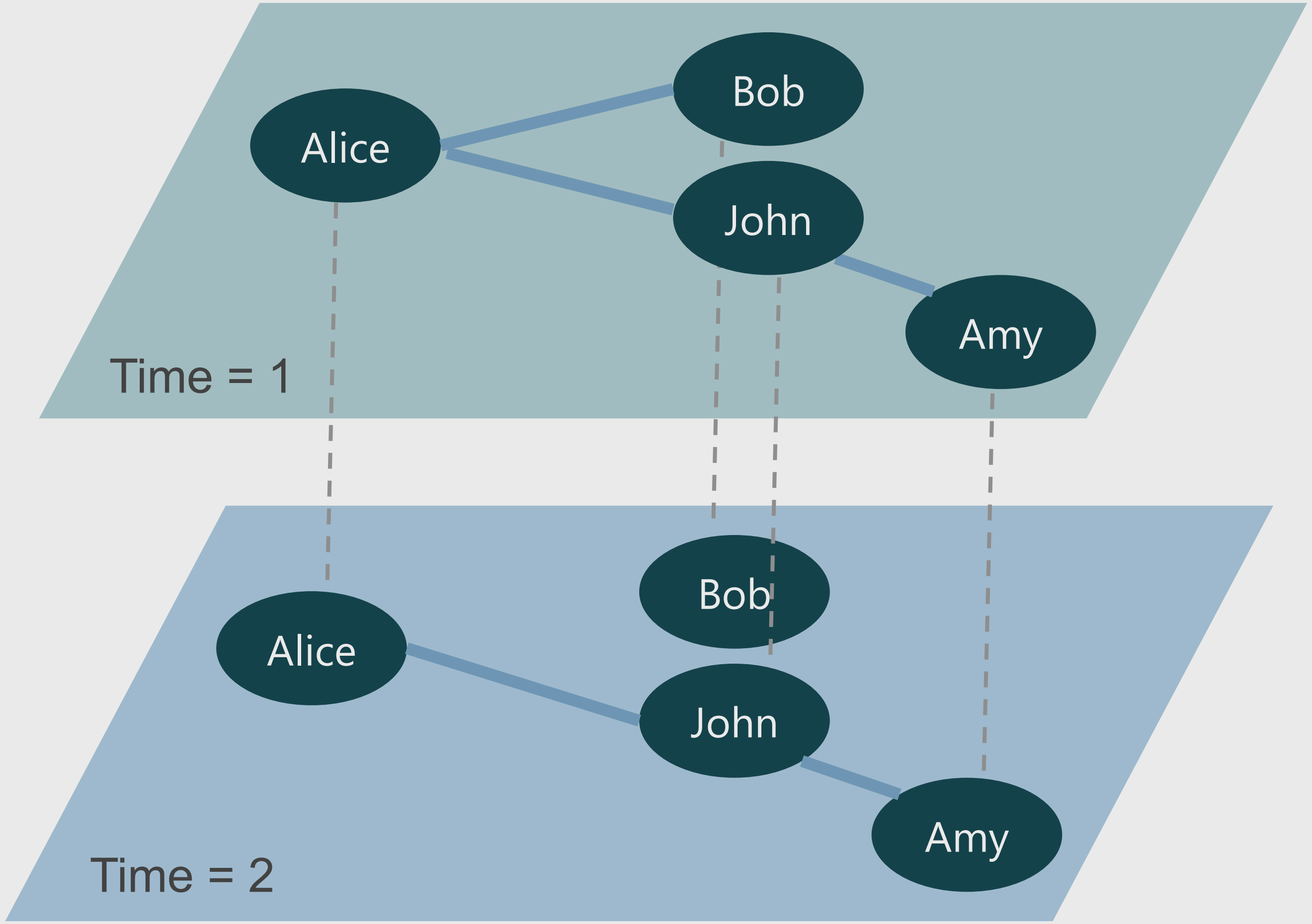
Structural balance



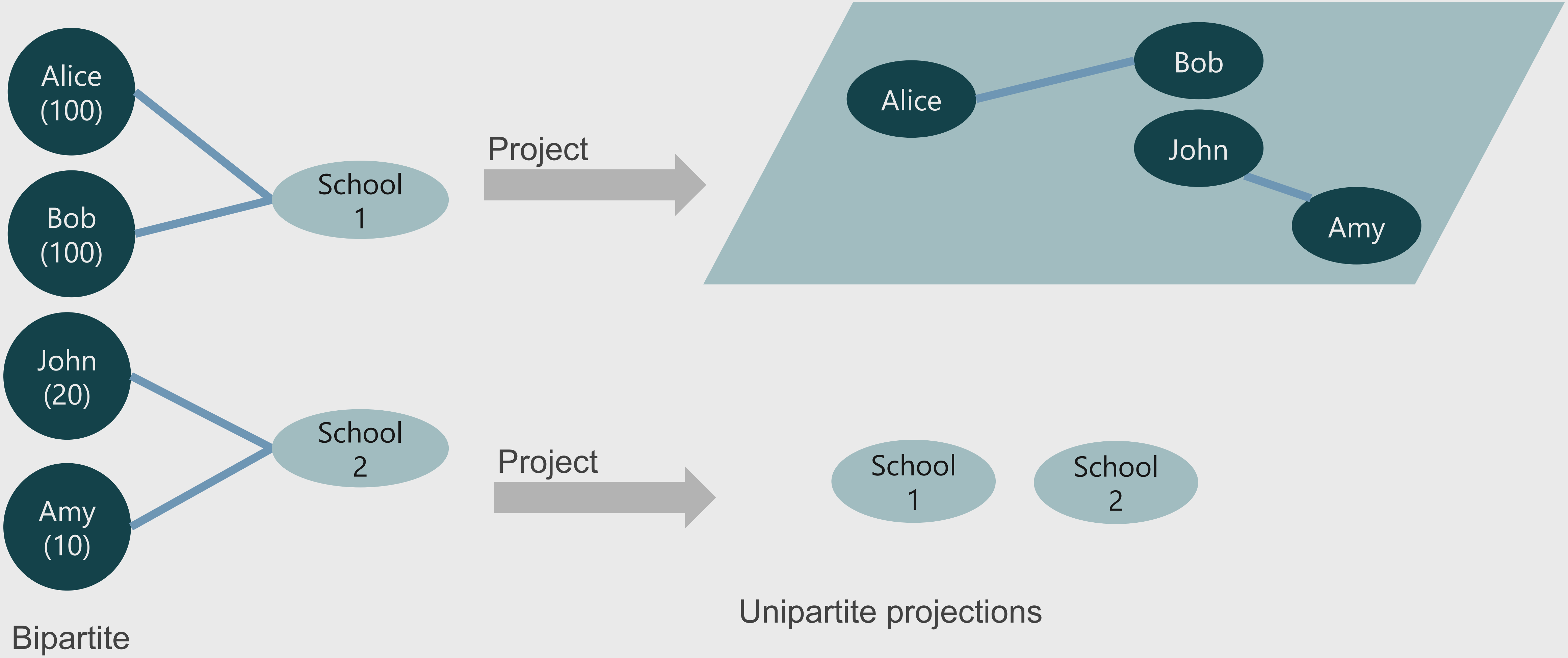
Other types of networks: Temporal

Either:

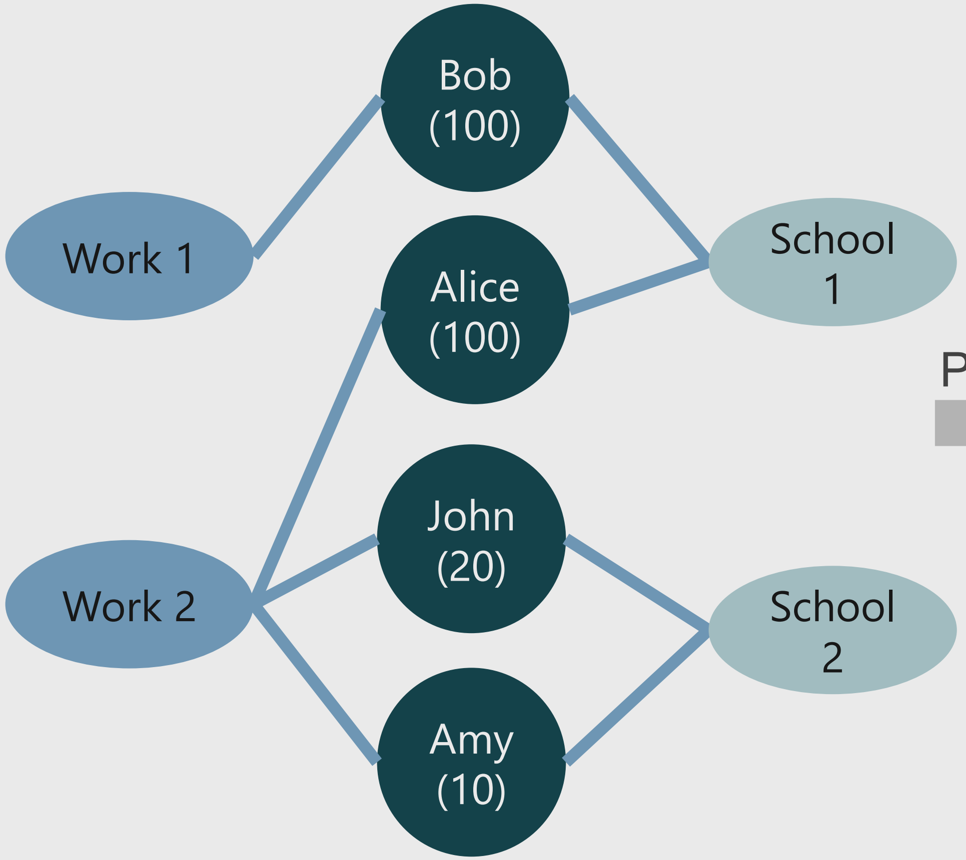
- Snapshots
- Time of events



Other types of networks: Bipartite

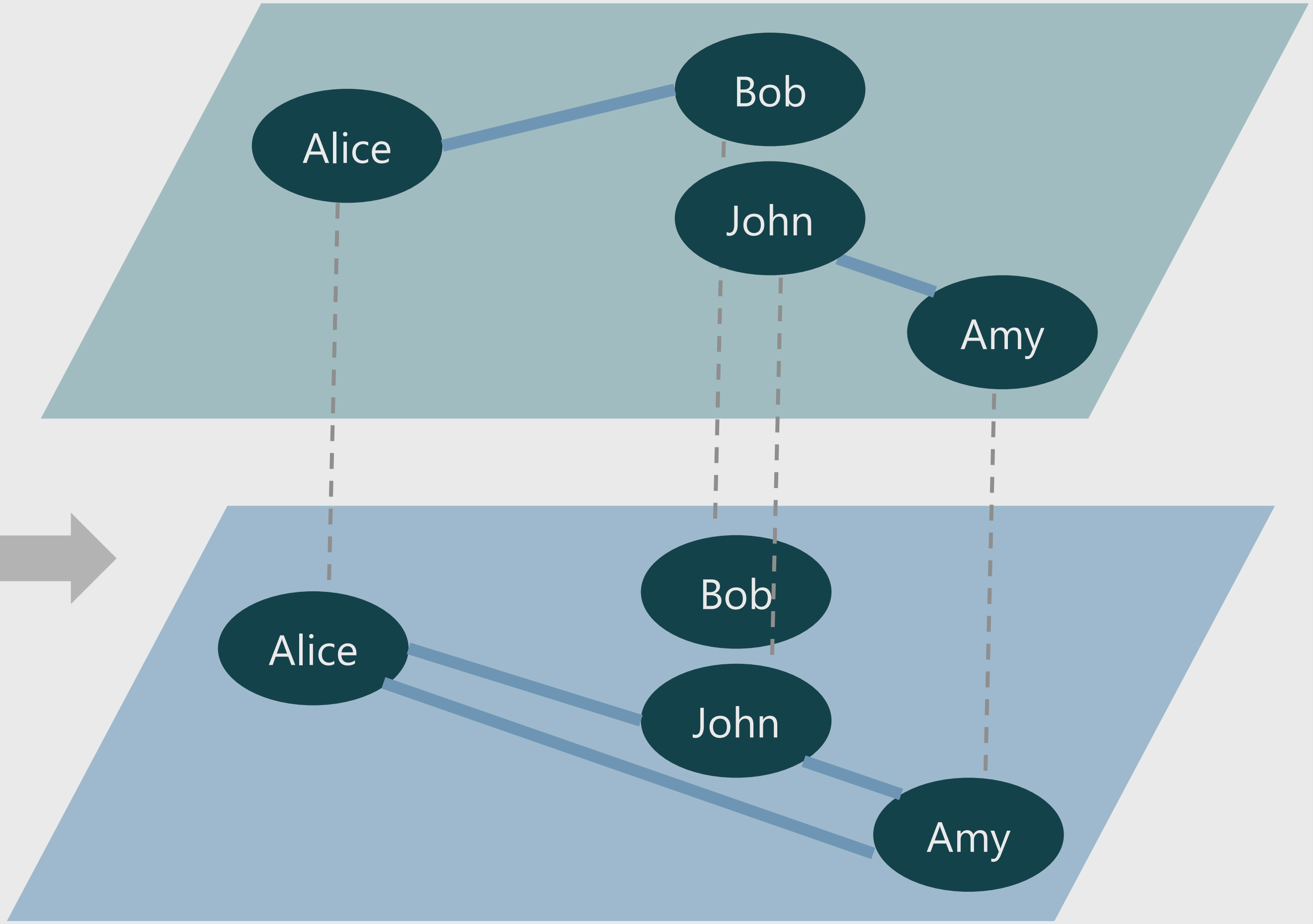


Other types of networks: Multipartite



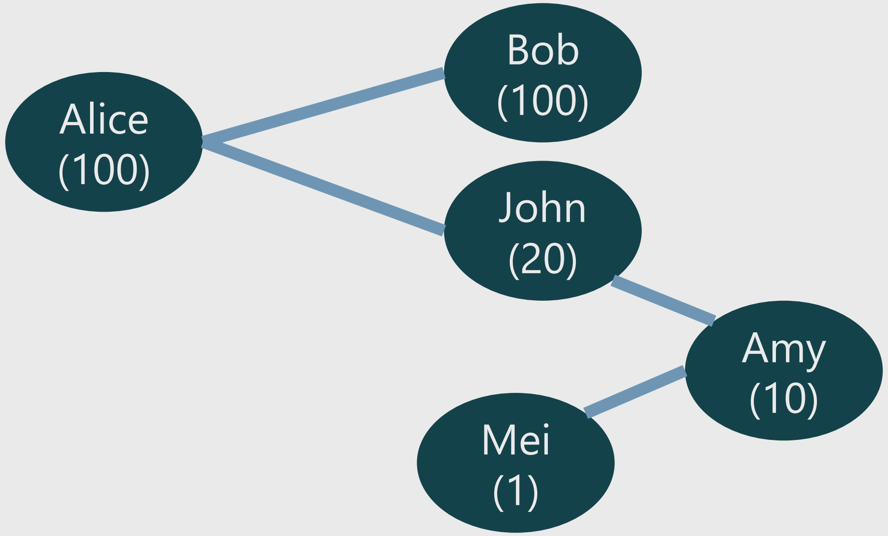
Multipartite network

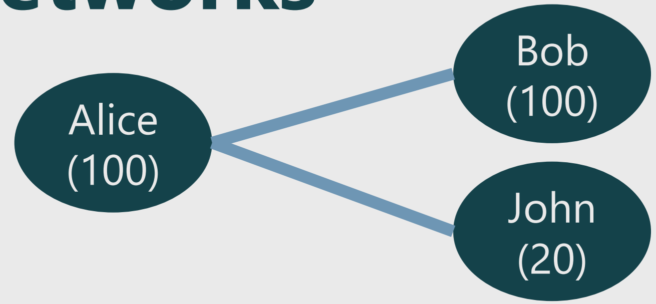
Project



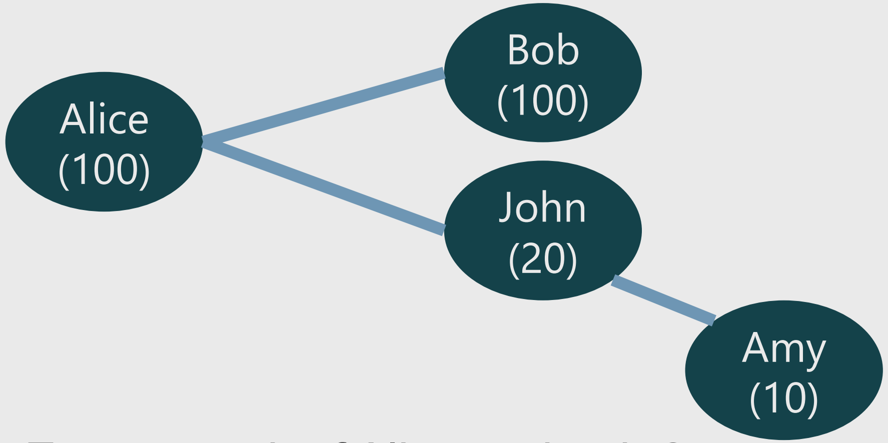
Multiplex projection

Other types of networks: Ego-networks



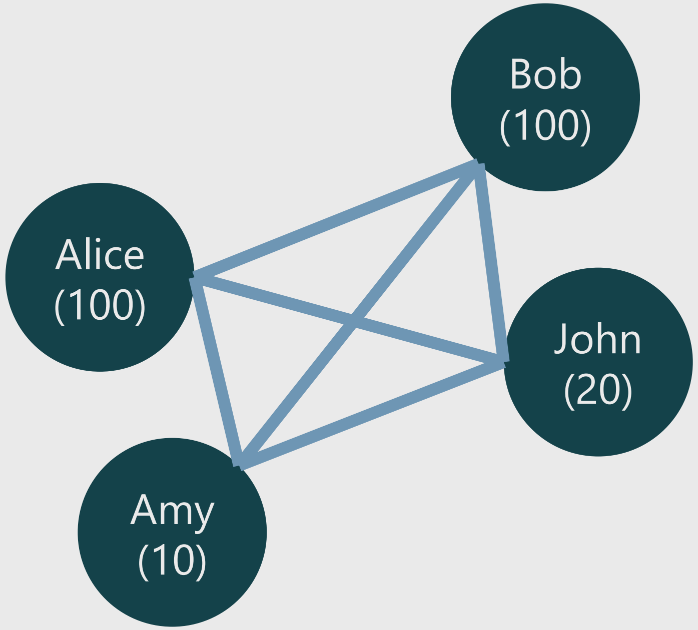


Ego network of Alice at depth 1



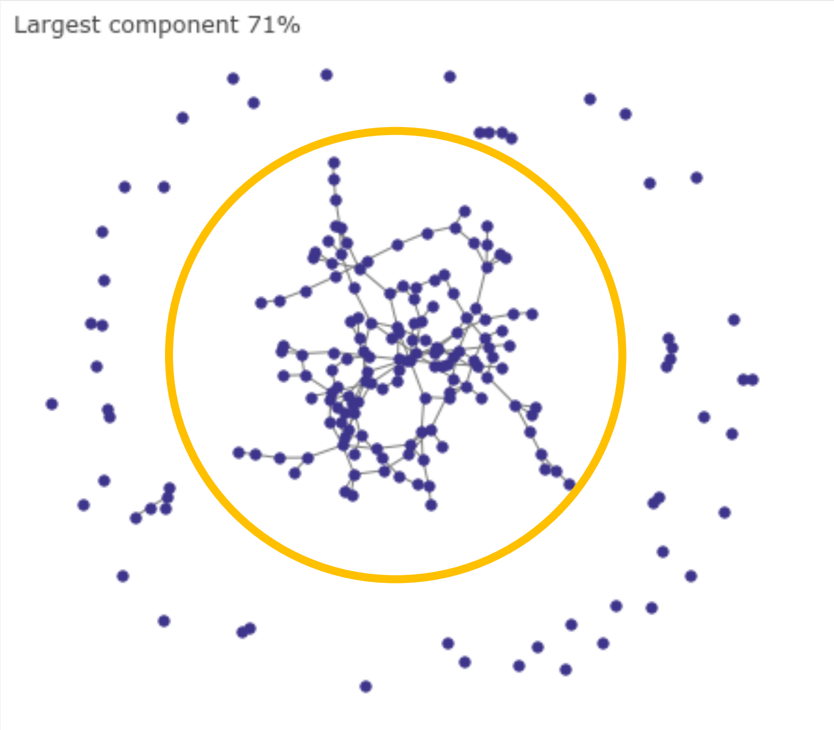
Ego network of Alice at depth 2

Other types of networks: Clique



Network characteristics

Connectedness



Real networks are typically connected, forming a **“giant component”**

If the average degree < 1 → many small components

If the average degree > 1 → suddenly the system becomes connected

Let’s try this!

Small world: six degrees of separation

Milgram's experiment: six degrees of separation

Strogatz, Watts: small number of random links are enough to create small world networks

In real networks: Hubs are responsible of the "smallness"

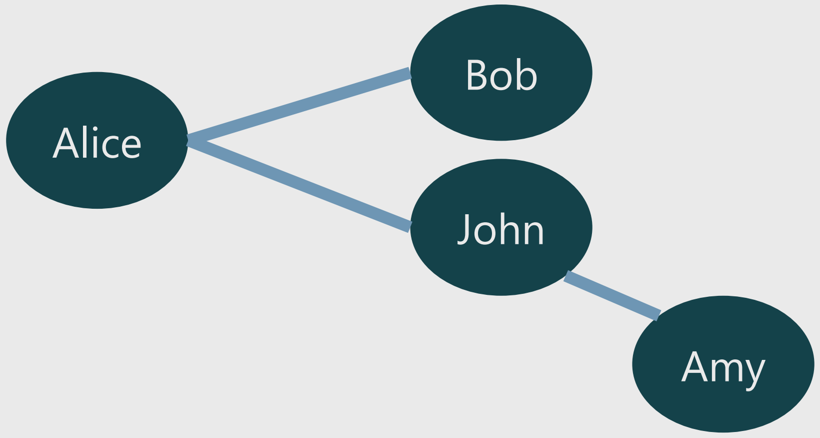
Shortest path between node 1 and node 2:

- Minimum number of steps requires to go from node 1 to node 2
- Between Alice, Amy → 2

Diameter:

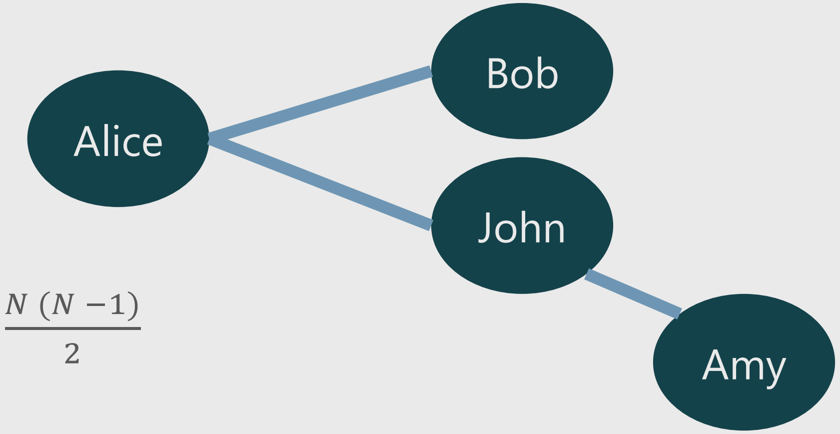
- Longest "shortest path" between two nodes
- In our network: 2 (Alice -> John -> Amy)

Real networks have **small diameters**



Density

Definition: Number of edges present / potential number of edges



- Number of edges = 3
- Potential number of edges in directed network = (4*3)
- Potential number of edges in undirected network = (4*3)/2 $= \binom{N}{2} = \frac{N\,(N-1)}{2}$

Density = 3/6 = 50%

Real networks are typically **sparse**

As size increases density decreases (average degree is usually fixed)

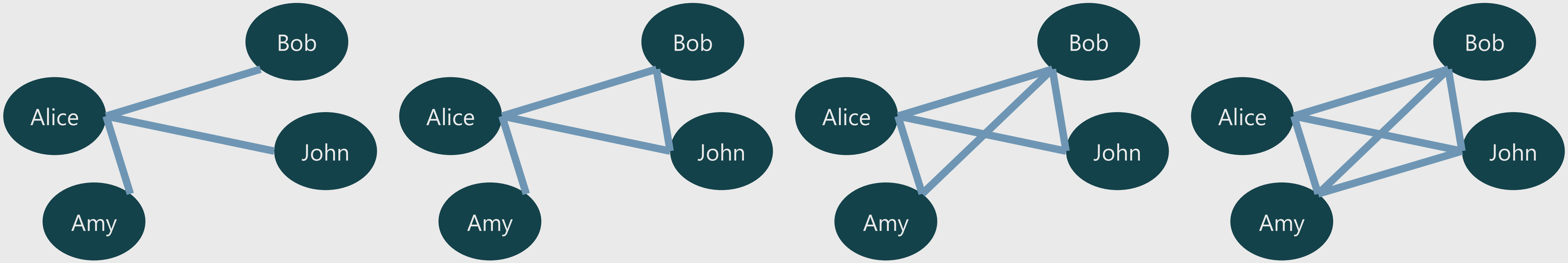
Local clustering (~transitivity)

Strogatz, Watts (1998): How many of your neighbors are connected to each other

Average clustering of a network: Average clustering of the nodes

Real networks have **high clustering**

Clustering of Alice:

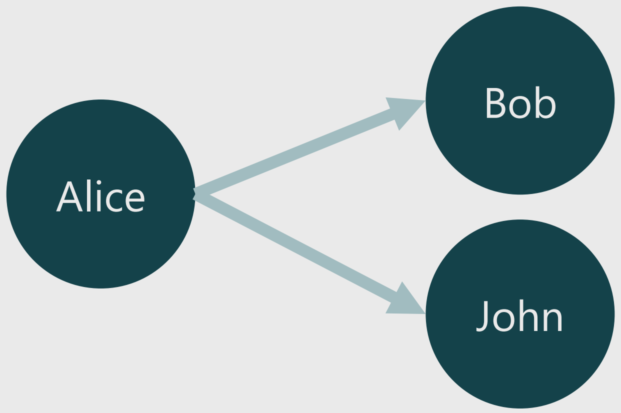


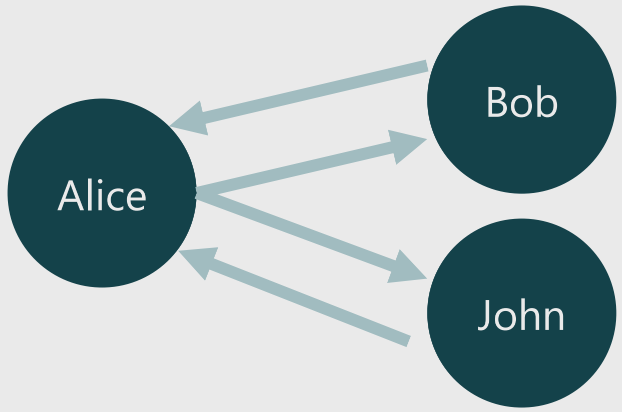
0/3 1/3 2/3 3/3

Reciprocity

Directed networks

Ratio of the number of edges pointing in both directions to the total number of edges in the graph.

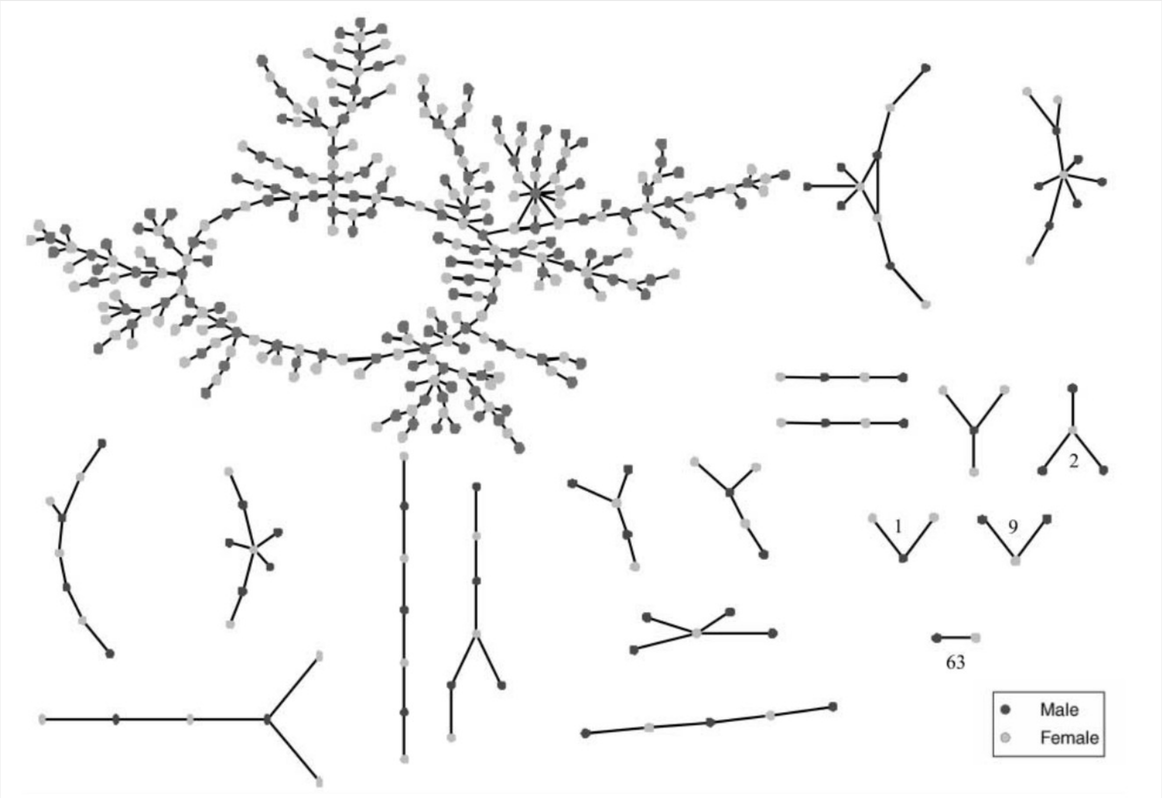




Assortativity (homophily)

Preference for nodes to attach to others that are similar in some way

Paraisópolis favela and Morumbi, in São Paulo
Photography by Tuca Vieira (the guardian)



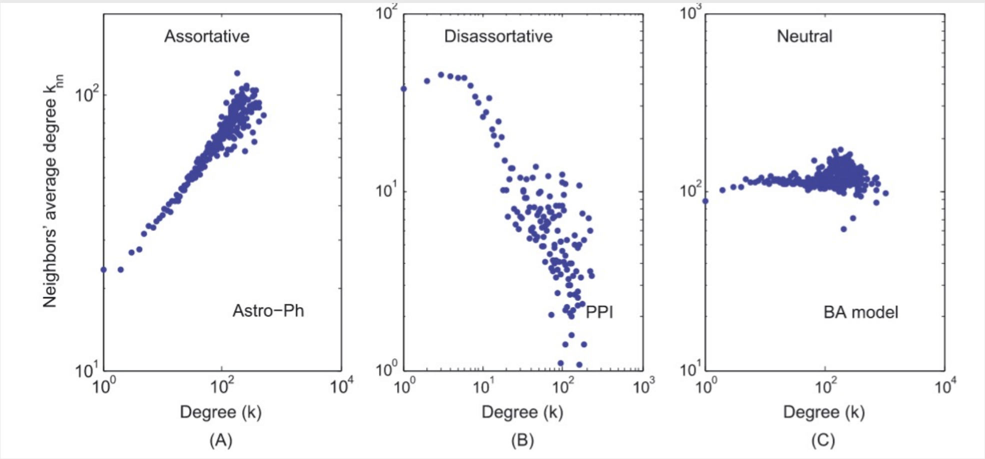
Romatic links between teenagers
Bearman, Moody, Stovel (1991)v

Assortativity (homophily)

At the network level:

- Categorial unordered variable = Modularity
  - (Actual links between edges between nodes of same type – expected number of links between nodes of same type)/number of links
- Continuous variable: Pearson's correlation across edges.

Mixing patterns in networks, Newman, Physical Review E, 67 026126, 2003



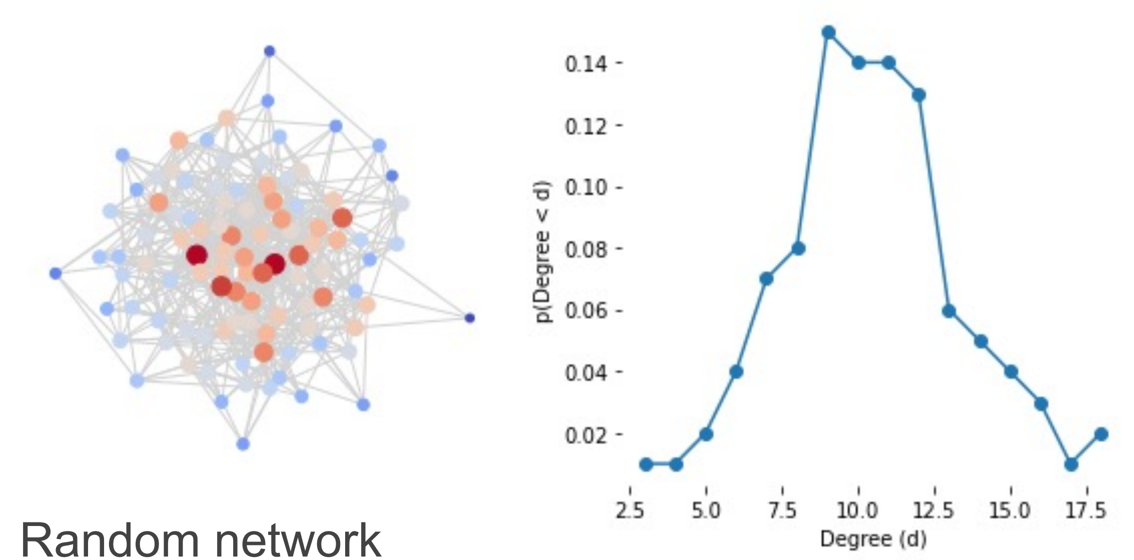
Jiang et al (2016)

At the local level:

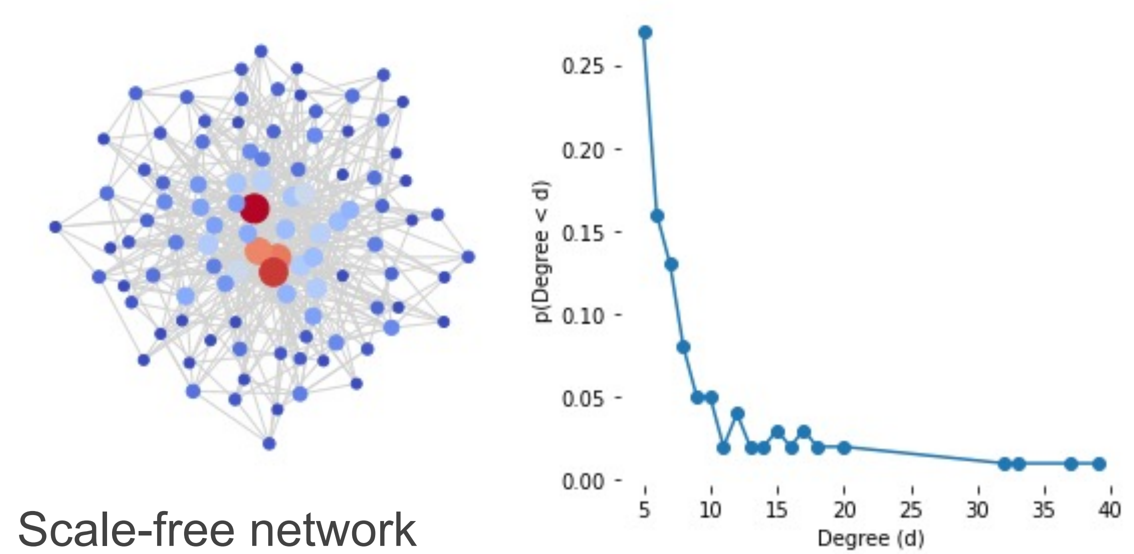
- Real networks can be locally assortative or disassortative
- Exercise: Draw a degree-assortative network

Multiscale mixing patterns in networks, Peel, Delvenne and Lambiotte (2018)

Heavy tails / scale-free



Random network



Scale-free network

Networks are not random, they have heavy degree distributions

PDF (probability density function)

→ Degree vs probability of degree

→ Represented by histogram

Many possible mechanisms:

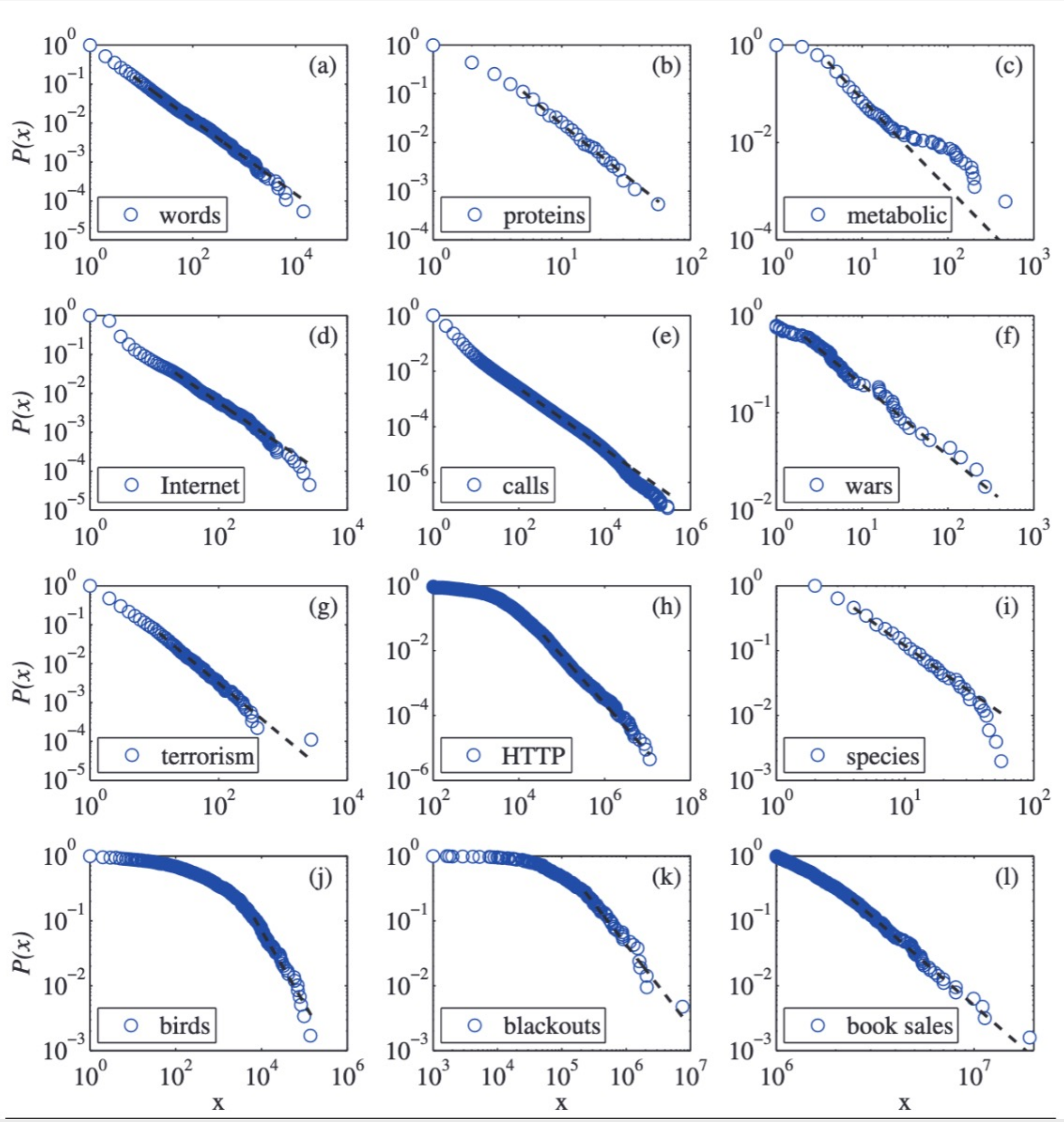
- Multiplicative growth
- Preferential attachment (Rich get richer, Mathew effect)
- Copying models

Growing networks: https://www.stat.cmu.edu/~cshalizi/networks/16-1/lectures/08/li.pdf

Heavy tails

Most complex systems have **heavy tail distributions**

Most real networks have heavy tail degree distributions



Clauset, Shalizi & Newman (2009)

Random networks don’t have heavy tails

PDF (probability density function)

→ Degree vs probability of degree

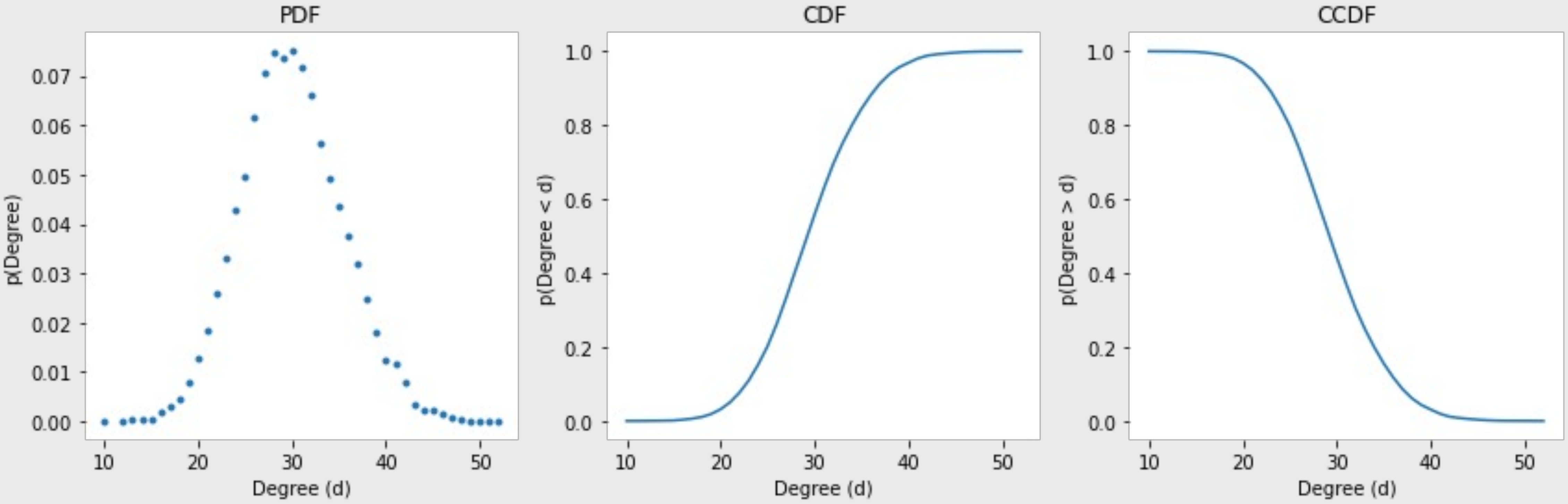
→ Represented by histogram

CDF (cumulative density function)

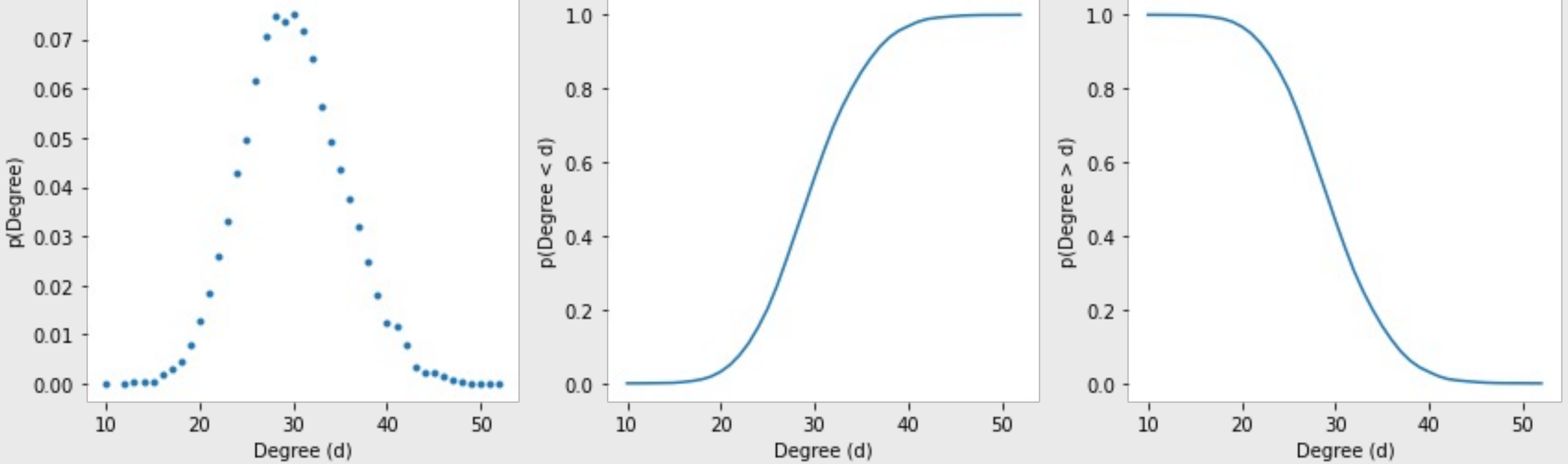
→ Degree s vs probability degree < s

CCDF: Complementary CDF

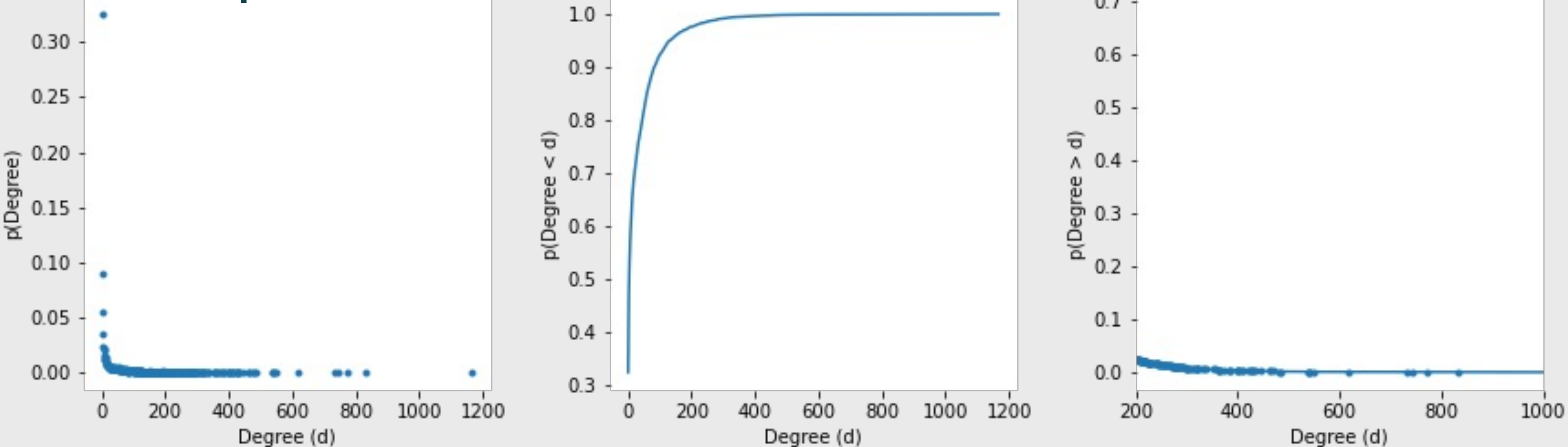
→ Degree s vs probability degree > s



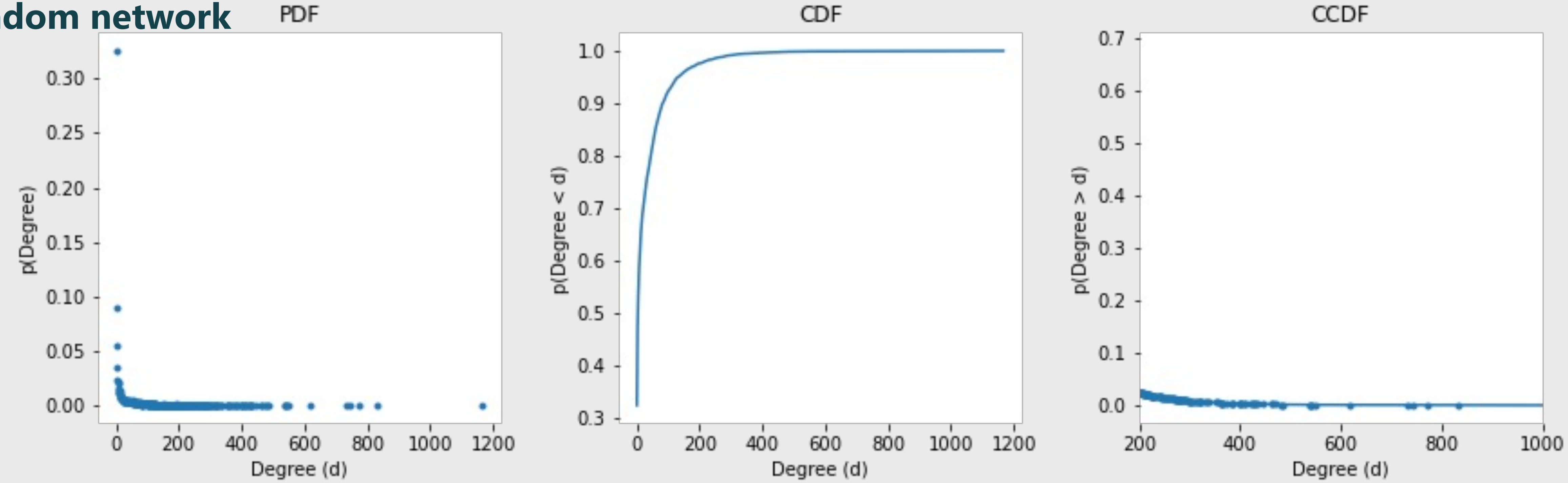
Random network



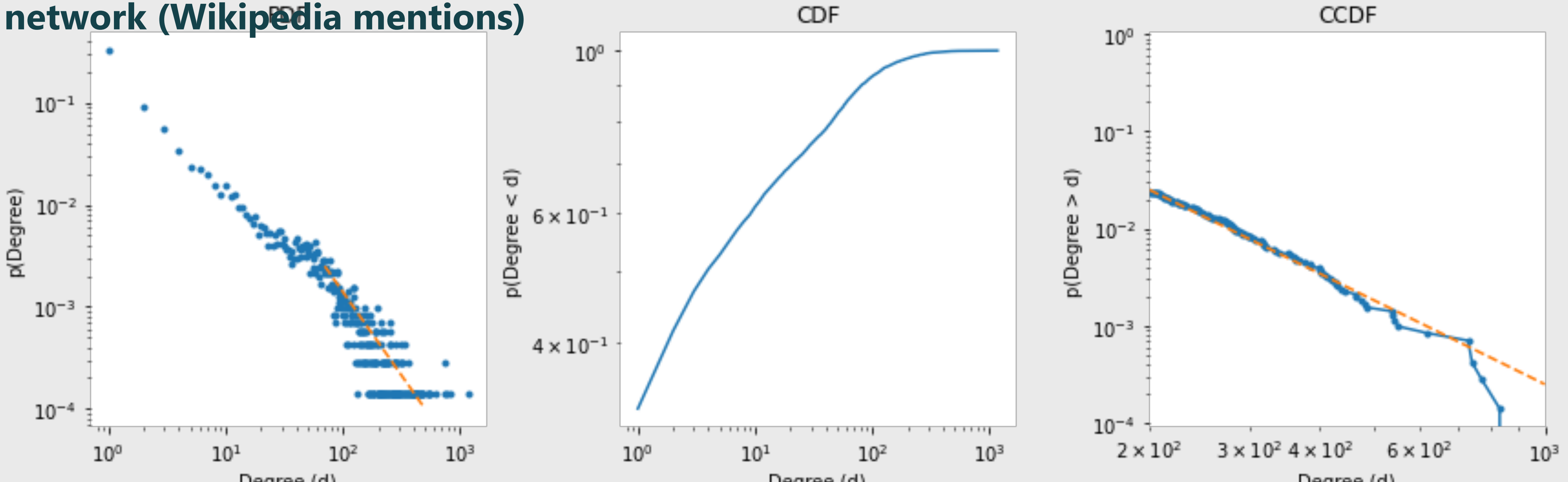
Real network (Wikipedia mentions)



Random network



Real network (Wikipedia mentions)



Is it a power-law? $P(d) \sim d^{-\alpha}$

Critical Truths About Power Laws

Most reported power laws lack statistical support and mechanistic backing.

MICHAEL P. H. STUMPF AND MASON A. PORTER

SCIENCE • 10 Feb 2012 • Vol 335, Issue 6069 • pp. 665-666 • DOI: 10.1126/science.1216142

Scale-Free Networks

Scientists have recently discovered that various complex systems have an underlying architecture governed by shared organizing principles. This insight has important implications for a host of applications, from drug development to Internet security

BY ALBERT-LÁSZLÓ BARABÁSI AND ERIC BONABEAU

Article | Open Access | Published: 04 March 2019

Scale-free networks are rare

Anna D. Broido & Aaron Clauset

Nature Communications **10**, Article number: 1017 (2019)

Love is All You Need

Clauset's fruitless search for scale-free networks

by Albert-László Barabási, March 6, 2018

Comment | Open Access | Published: 04 March 2019

Rare and everywhere: Perspectives on scale-free networks

Petter Holme

Nature Communications **10**, Article number: 1016 (2019) | Cite this article

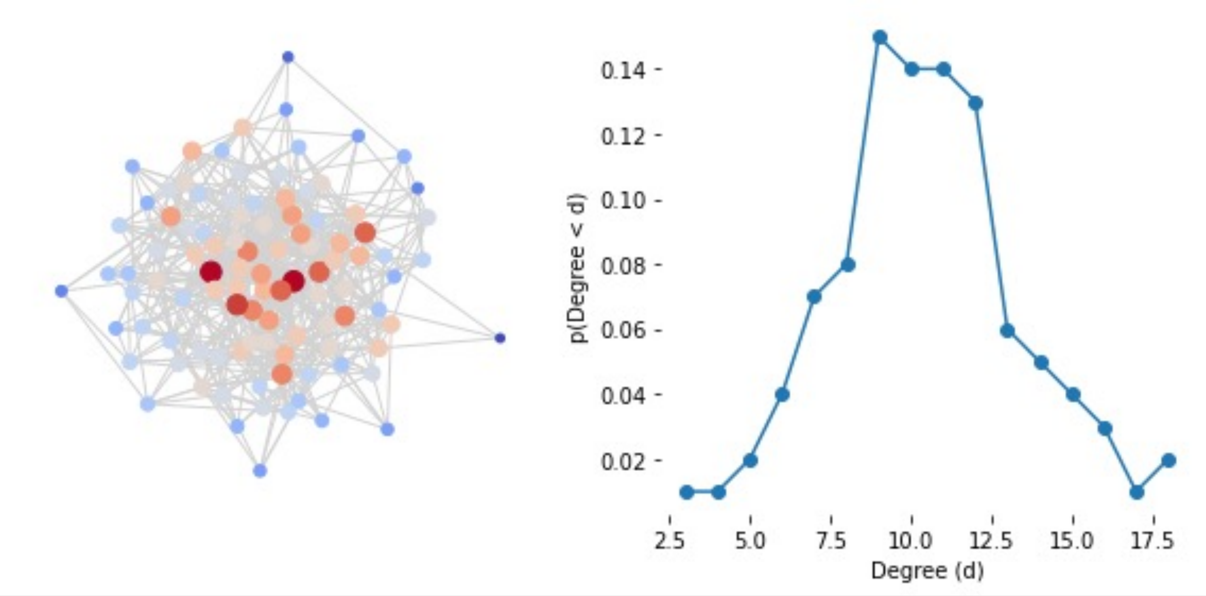
True scale-free networks hidden by finite size effects

Matteo Serafino, Giulio Cimini, Amos Maritan, +3 , and Guido Caldarelli Authors Info & Affiliations

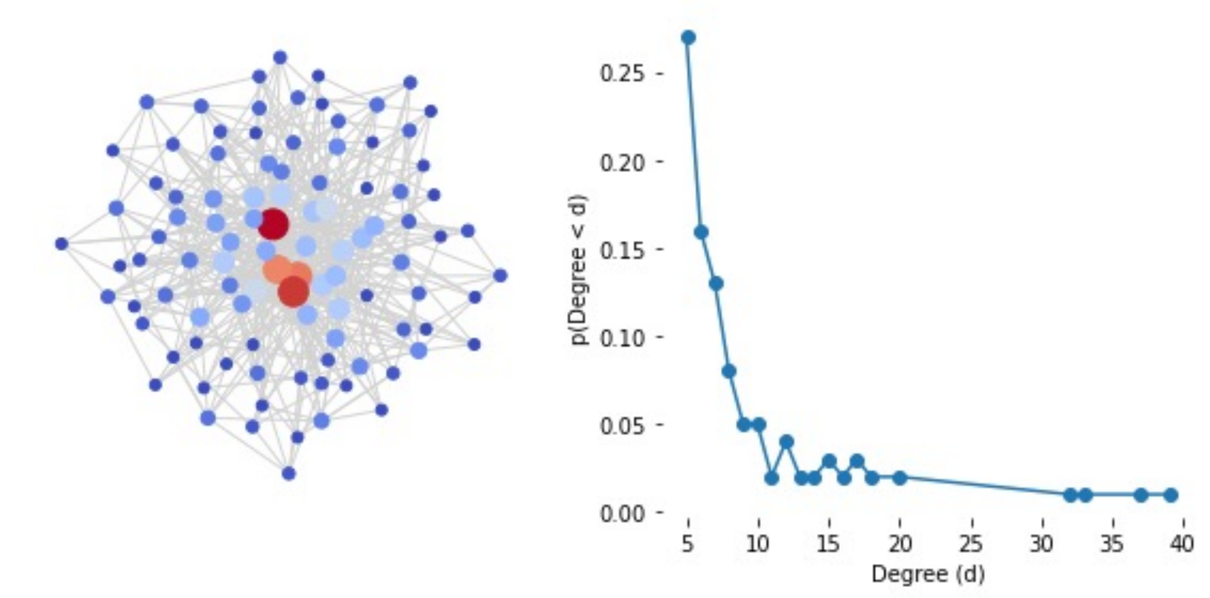
Edited by Lai-Sang Young, New York University, New York, NY, and approved November 2, 2020 (received for review July 3, 2020)

December 30, 2020 | 118 (2) e2013825118 | https://doi.org/10.1073/pnas.2013825118

Robustness to failures
Fragility to targeted attacks



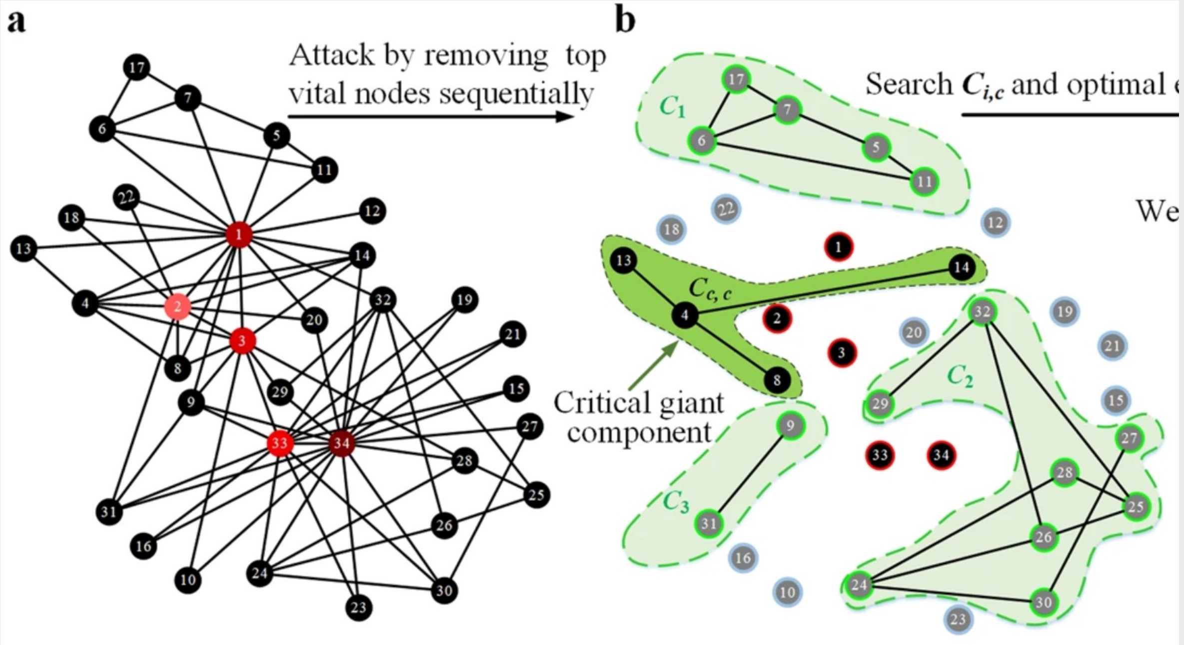
Random network



Power-law network

Robustness to failures
Fragility to targeted attacks

Albert, Jeong, Barabasi (2000) Attack and error tolerance of complex networks



Li et al (2011)

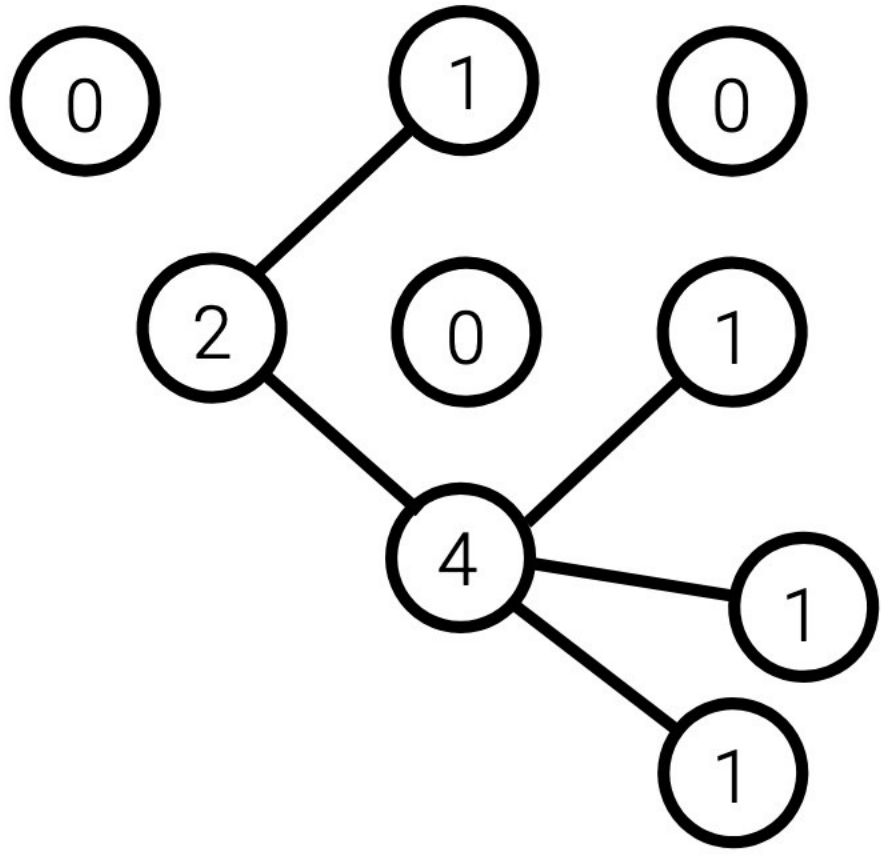
Friendship paradox

Your friends are more popular than you are

Some people have no friends.

But because they appear in nobody's friendship circles, they're not making anyone else feel popular.

The same applies to other people: the more friends you have, the likelier you are to be represented in people's friendship circles.



Feld, S. L. (1991). Why your friends have more friends than you do. American Journal of Sociology, 96(6), 1464-1477.

Friendship paradox

Your friends are more popular than **you** are

average friend
(count node proportional
to their degree)

average person
(count each node once)

Feld, S. L. (1991). Why your friends have more friends than you do. American Journal of Sociology, 96(6), 1464-1477.

Practical 1, exercise 2

Compare the PPI network with the Twitter network (ic2s2_netsci_3.tsv).

What characteristics apply to both? Which don’t? Why?

Practical 1, exercise 3

In 3.7 compare the Wikipedia network (default example), the PPI network and the Twitter network (IC2S2).

Types of analysis

They should fit your research question

Types of analysis: Descriptive statistics

Describe the network characteristics (density, diameter, average degree, clustering, etc)

Types of analysis: Centralities

What are the most important nodes in the network?

- The one with more connections ➔ **Degree centrality**
- The one linked to more important neighbors ➔ **Pagerank / Eigenvector / Katz centrality**
- The one closest to all other nodes ➔ **Closeness centrality**
- The ones that act as brokerage? ➔ **Betweeness centrality**

More on this this afternoon

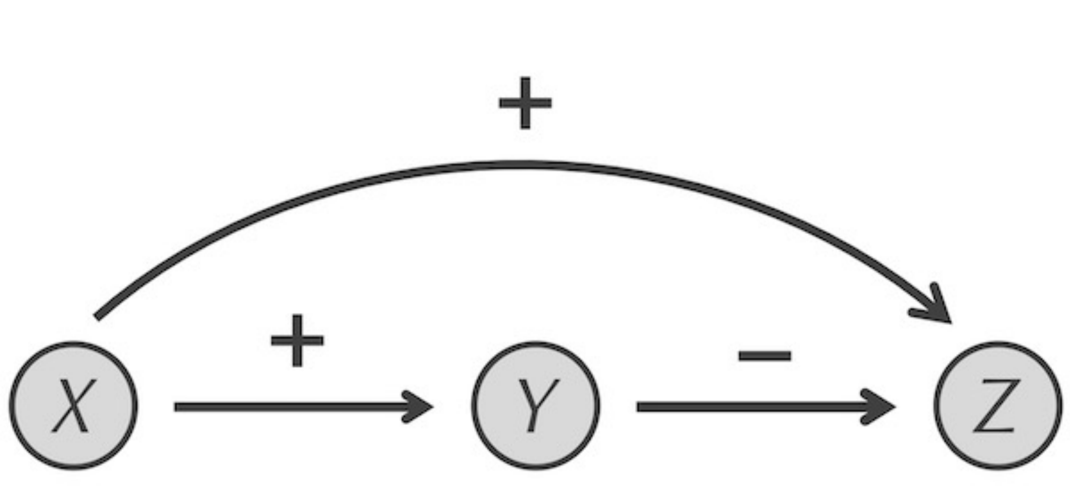
Types of analysis: Node-level regression

Calculate node-level features:

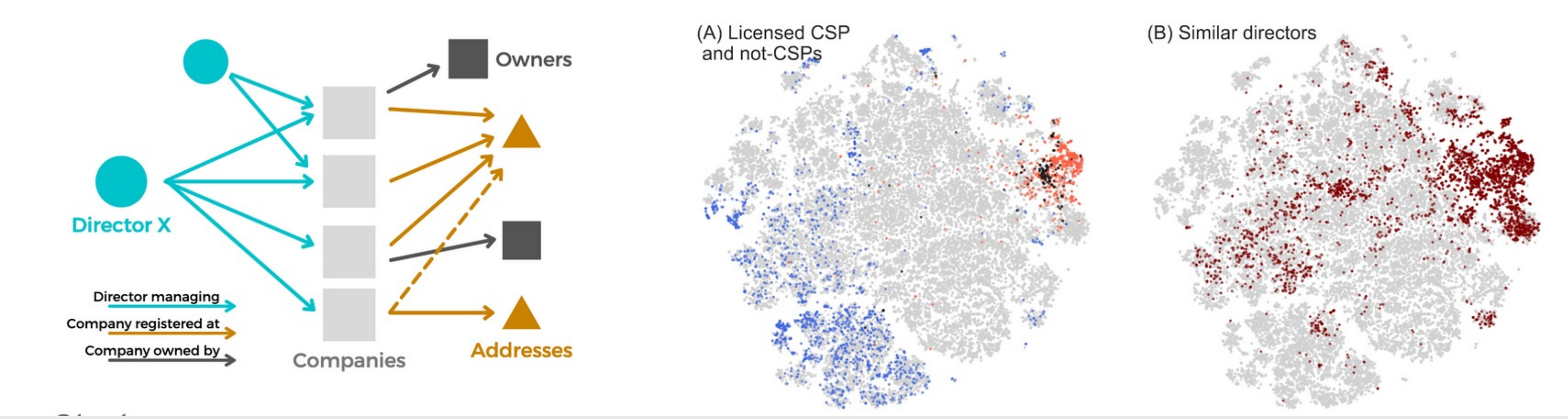
- Centrality
- Local clustering (transitivity / embeddedness)
- Local reciprocity
- Local assortativity (homophily)
- …
- Include in your prediction model (e.g. a regression)

Types of analysis: Motif detection

Find overrepresented patterns



Feed-forward loop (https://biologicalmodeling.org/motifs/feedforward)



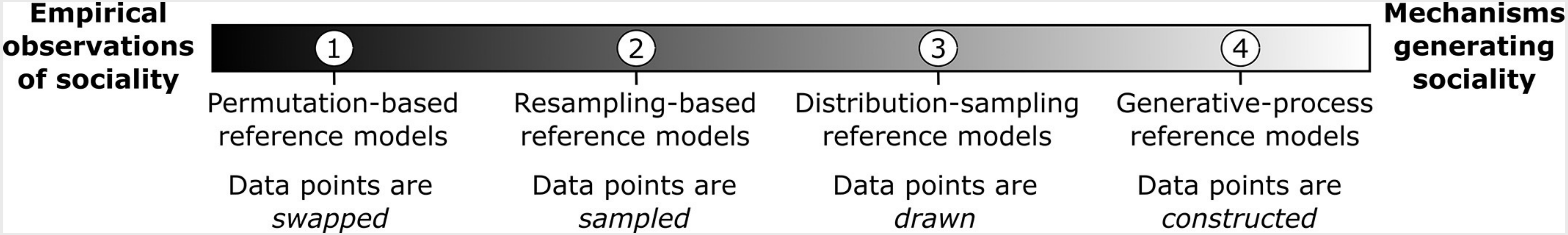
(Garcia-Bernardo et al, 2022)

Types of analysis: Testing hypothesis

We observe some behavior in the network (e.g. the clustering is 0.5). Is this relevant?

Approach: Create a reference model (see *Hobson 2021* for a great guide) to compare with it

- Configuration model (permuting edges)
- Generative models (e.g. rich get richer model + homophily + triadic closure)
- ERGM (model how the features of dyads affect the presence or strength of edges.)
- ABM



Hobson 2021

More on this on Thursday

Permutation of attributes

Calculate significance by resampling.
Avoids running a regression (which is difficult with all the interdependences)

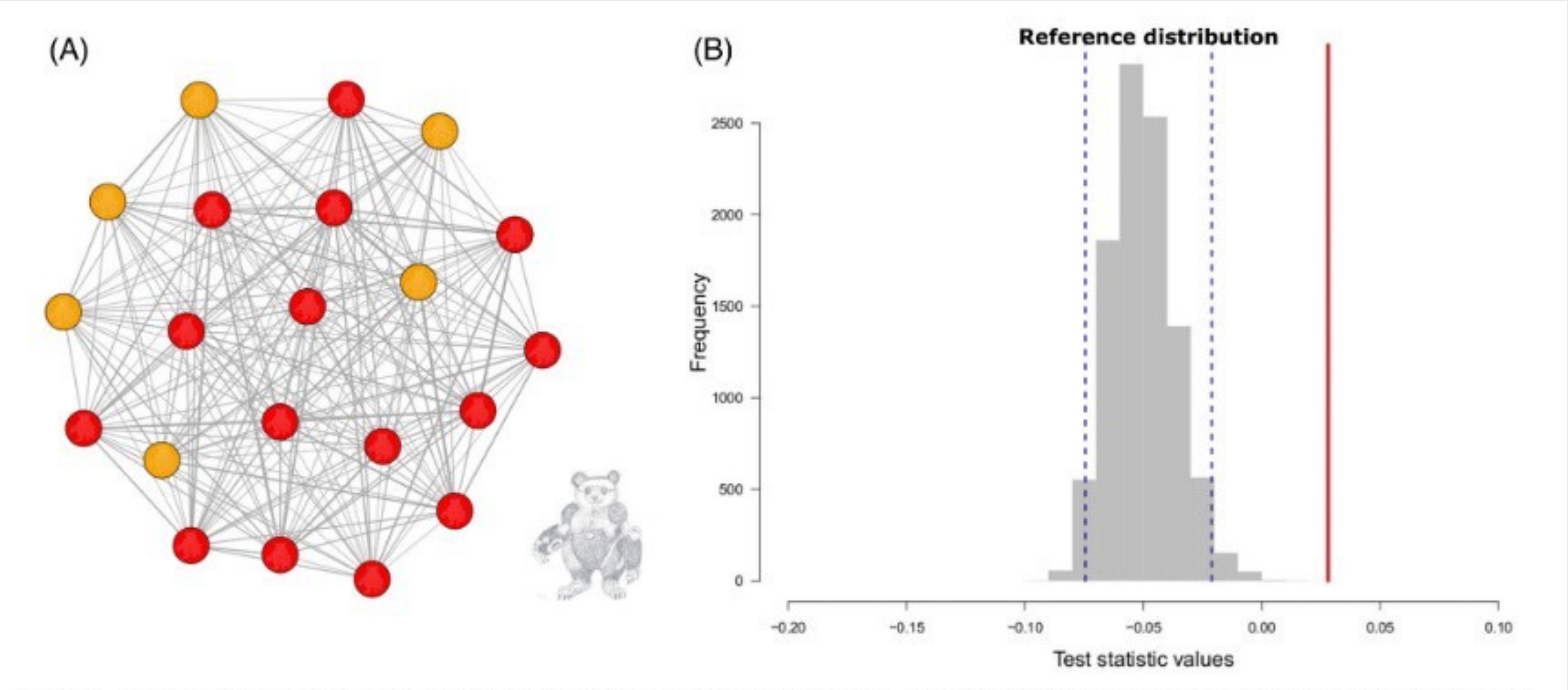
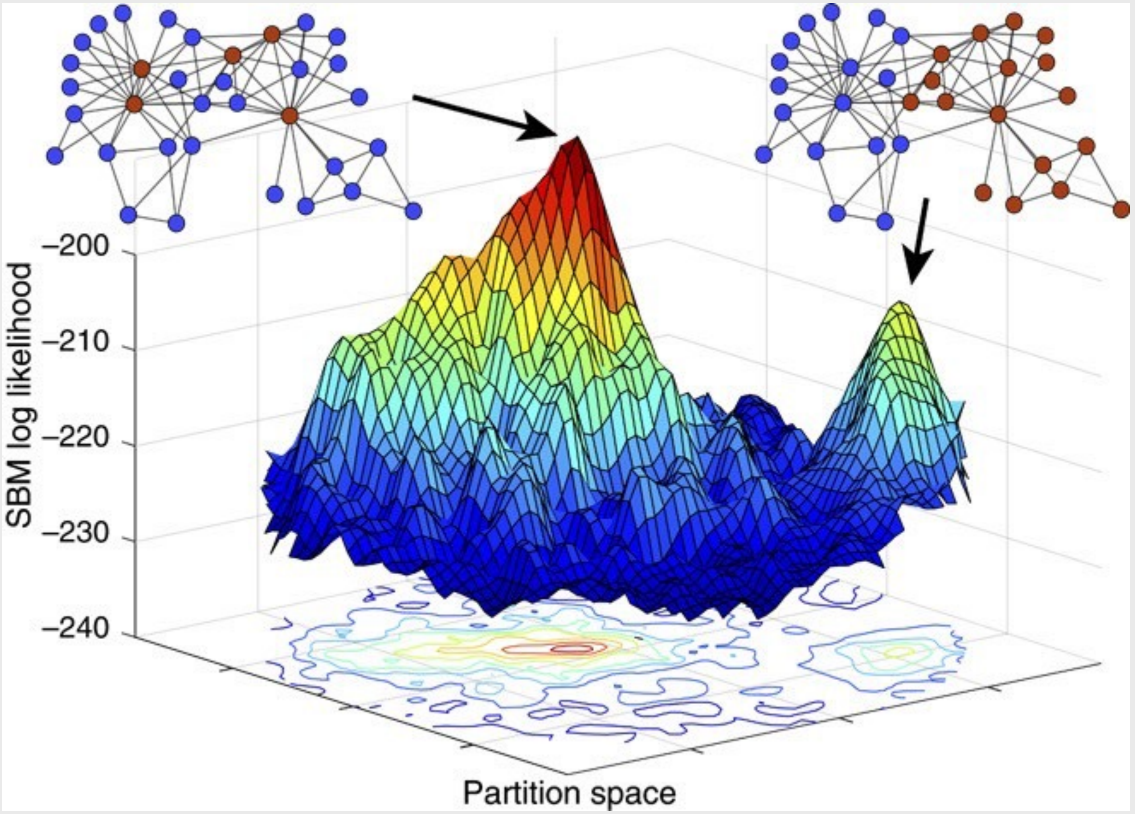


Fig 2. An example of study approaches: do burbils socially assort by nose colour? (A) Association network of burbils, with nodes colour-coded by nose colour and (B) distribution of values based on the permutation procedure of Team 1; observed value of the test statistic shown as a red solid line and the 2.5 and 97.5% quantiles of the reference distribution as blue dashed lines.

Hobson 2021

Types of analysis: Community detection



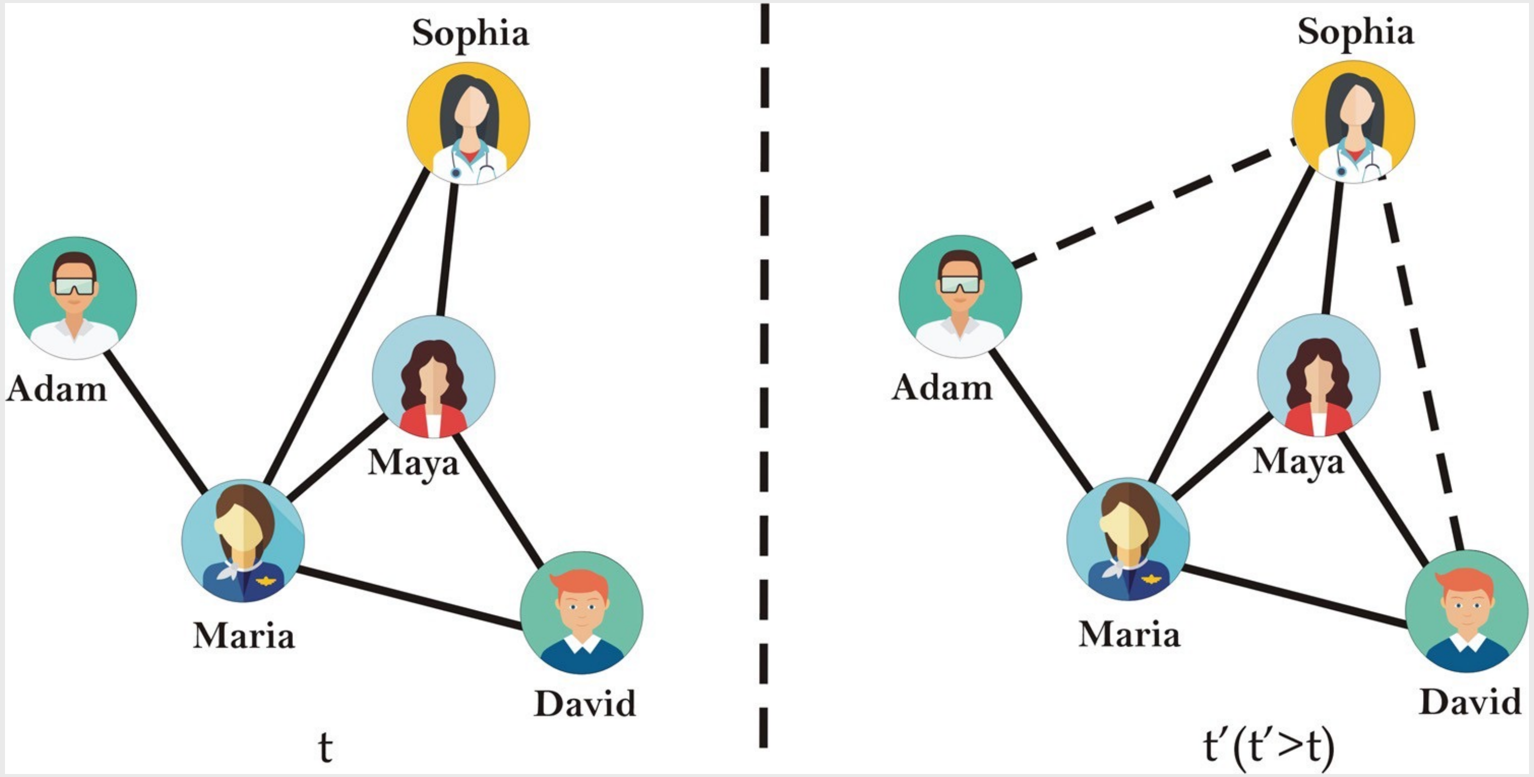
Peel et al

- **What clusters of nodes can we find in the network?**

- **Stochastic Blockmodels** (Harrison White, structural equivalence, core-periphery)
- **Modularity minimization**

More on this on Thursday

Types of analysis: Link/metadata prediction



Ahmad et al 2020

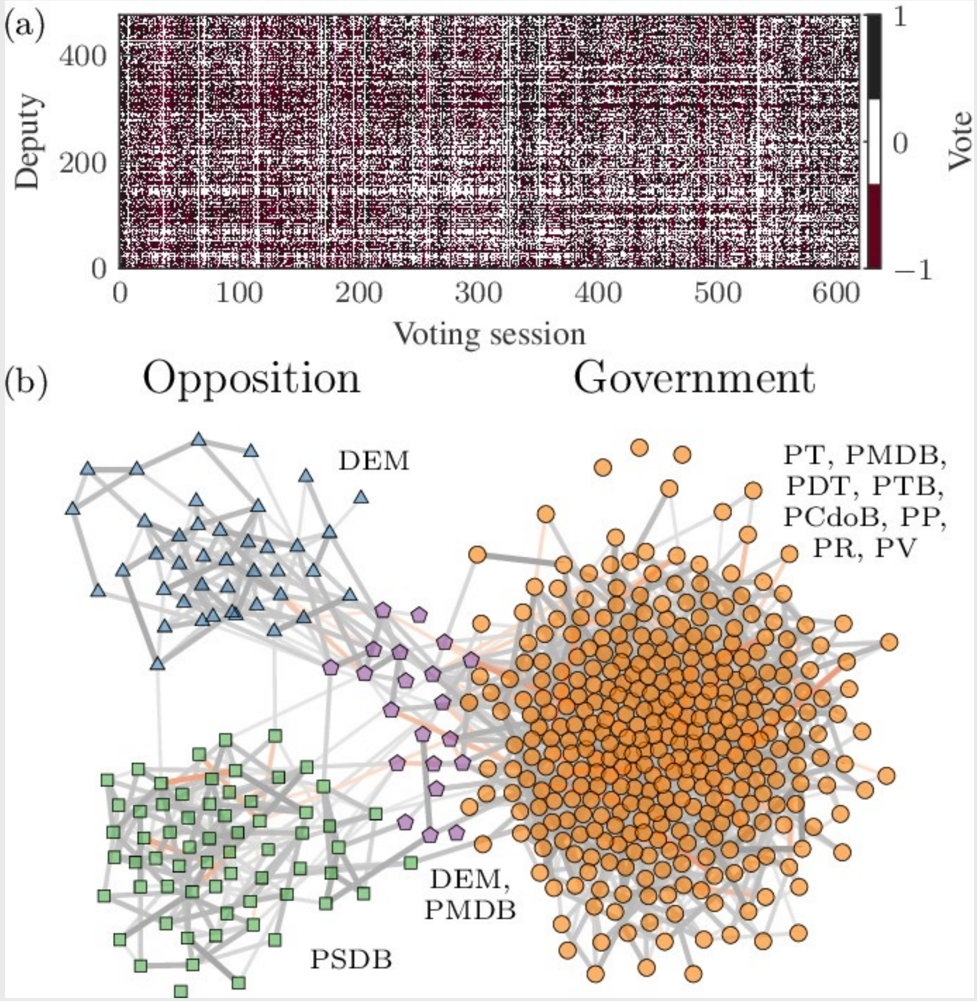
Networks are rarely complete

Approaches such as triangle closure, SBM or node embeddings

More on this on Tuesday

Types of analysis: Network inference

Network from co-occurrences



Network Reconstruction and Community Detection from Dynamics, Peixoto 2019

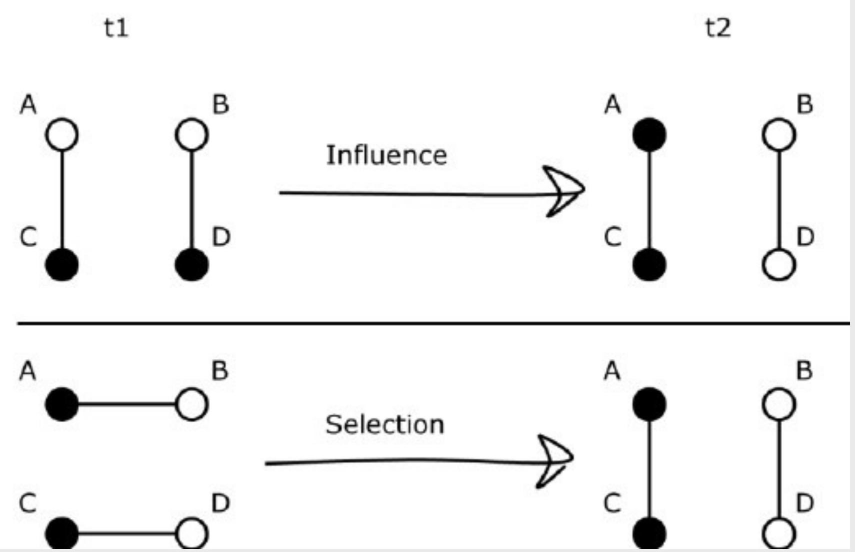
More on this on Wednesday

Types of analysis: Dynamics

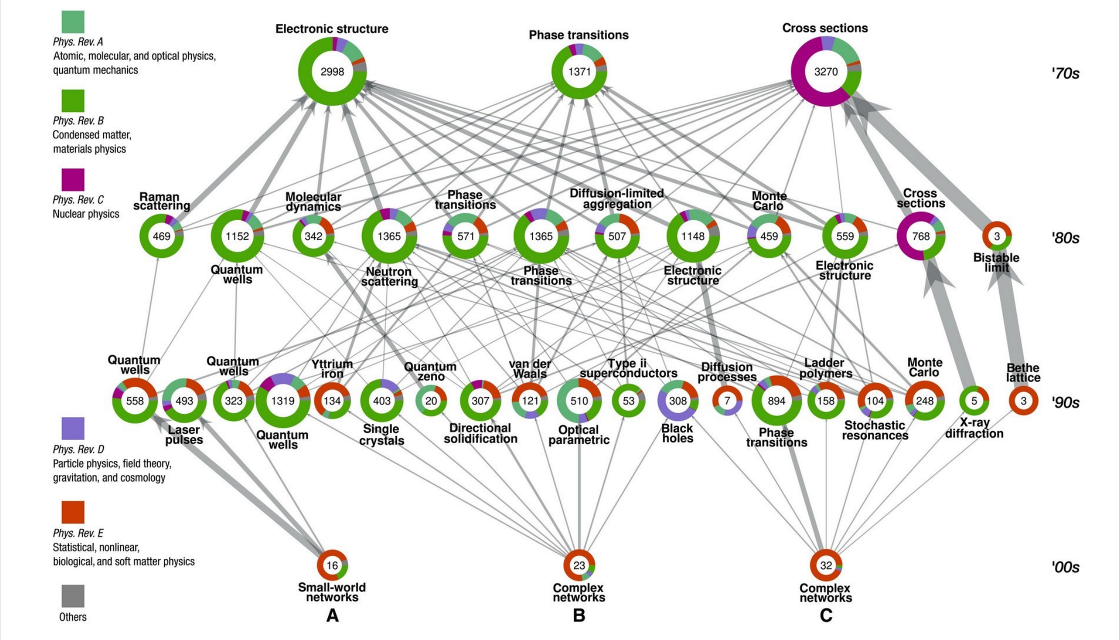
How does behavior/diseases/information spread?

Model matters: Simple contagion vs complex contagion

Longitudinal data: Allow to test selection vs influence, evolution of communities over time, co-evolution of network topology and ideas



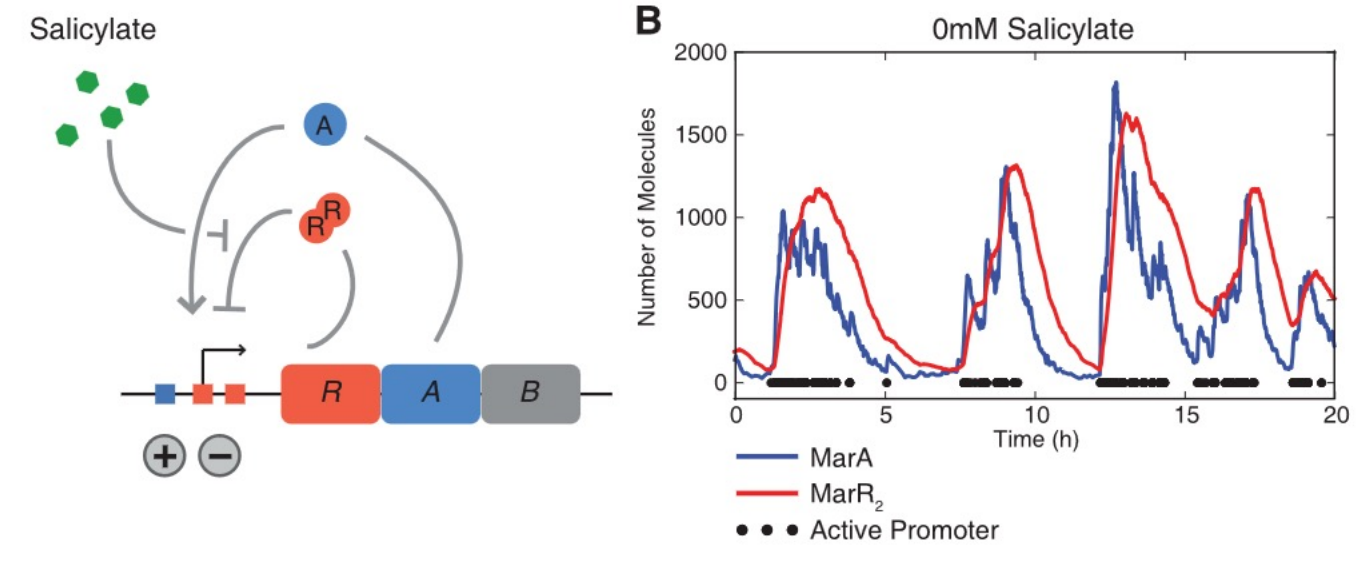
Friemel, 2015



Bovet et al, 2022

More on this on Friday

Types of analysis: Dynamics



$$P_{00} + A \underset{k_{-a}}{\overset{k_a}{\rightleftharpoons}} P_{10}$$

$$P_{10} \xrightarrow{\alpha_{10}} P_{10} + M + R_{uf} + A_{uf}$$

$$P_{11} \xrightarrow{\alpha_{11}} P_{11} + M + R_{uf} + A_{uf}$$

$$P_{01} + A \underset{k_{-a}}{\overset{k_a/\beta}{\rightleftharpoons}} P_{11}$$

$$P_{12} \xrightarrow{\alpha_{12}} P_{12} + M + R_{uf} + A_{uf}$$

$$R \xrightarrow{\lambda_r} 0$$

$$P_{02} + A \underset{k_{-a}}{\overset{k_a/\beta'}{\rightleftharpoons}} P_{12}$$

$$P_{02} \xrightarrow{\alpha_{02}} P_{02} + M + R_{uf} + A_{uf}$$

$$R_2 \xrightarrow{\lambda_r} 0$$

$$P_{00} + R_2 \underset{k_{-r}}{\overset{2 \cdot k_r}{\rightleftharpoons}} P_{01}$$

$$M \xrightarrow{\beta_r} M + R_{uf}$$

$$M \xrightarrow{\lambda_M} 0$$

$$M \xrightarrow{\beta_a} M + A_{uf}$$

$$P_{01} + R_2 \underset{2 \cdot k_{-r}}{\overset{k_r}{\rightleftharpoons}} P_{02}$$

$$P_{00} \xrightarrow{\alpha_{00}} P_{00} + M + R_{uf} + A_{uf}$$

$$A_{uf} \xrightarrow{\lambda_{auf}} 0$$

$$A_{uf} \xrightarrow{k_{fa}} A$$

$$R_2 \underset{k_{-sal}}{\overset{[Sal] \cdot k_{Sal}}{\rightleftharpoons}} R_2 \cdot Sal$$

$$P_{10} + R_2 \underset{k_{-r}}{\overset{2 \cdot k_r/\alpha}{\rightleftharpoons}} P_{11}$$

$$A \xrightarrow{\lambda_a} 0$$

$$R_{uf} \xrightarrow{k_{fr}} R$$

$$P_{01} \xrightarrow{\alpha_{01}} P_{01} + M + R_{uf} + A_{uf}$$

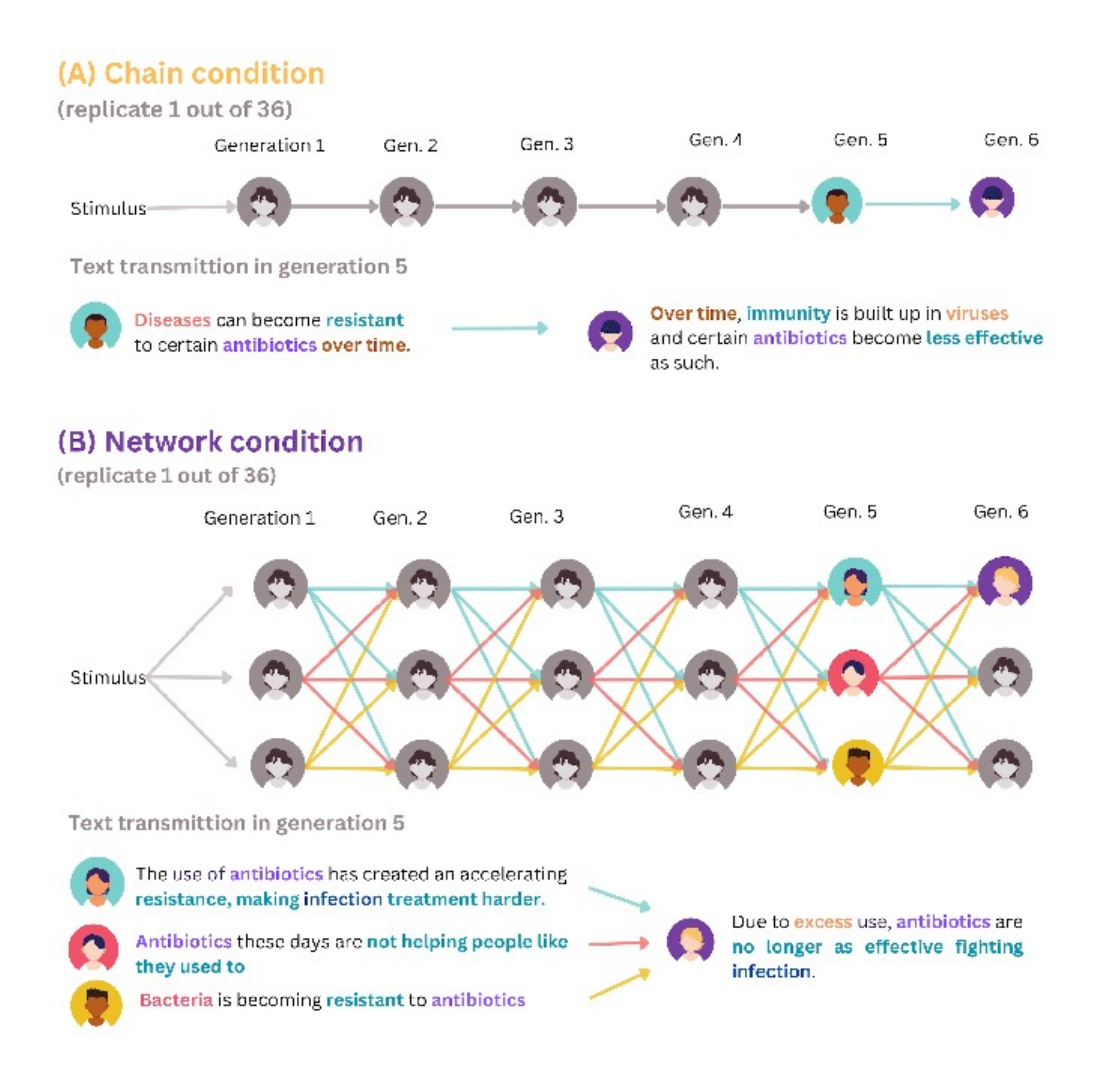
$$R_{uf} \xrightarrow{\lambda_{ruf}} 0$$

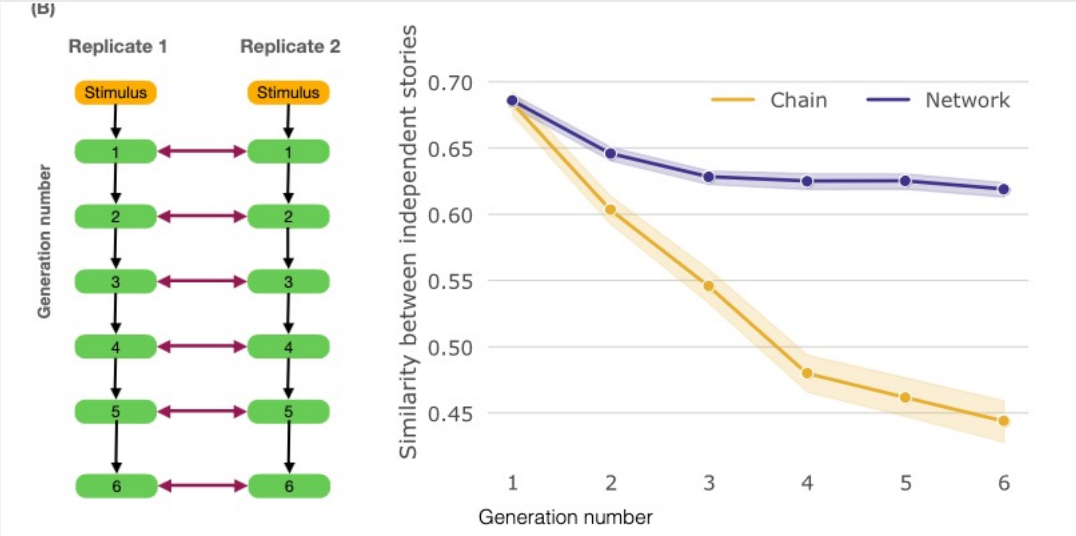
$$P_{11} + R_2 \underset{2 \cdot k_{-r}}{\overset{k_r/\alpha'}{\rightleftharpoons}} P_{12}$$

$$2 \times R \underset{k_{-dr}}{\overset{k_{dr}}{\rightleftharpoons}} R_2$$

Garcia-Bernardo and Dunlop (2015,2016, 2017)

Types of analysis: Dynamics





Resources

Tools

- Libraries:
- igraph (C, Python & R wrappers)
- Networkx (Pyhon)
- graph-tool (Python (UNIX))
- statnet (R)

- Gephi: open-source network analysis and visualization software package

- Interactive network visualization:
- visNetwork (R) – see e.g. here
- pyVis (Python)
- Bokeh + networkx (Python)

Data

Stanford Large Network Dataset Collection: https://snap.stanford.edu/data/

Network repository: https://networkrepository.com/networks.php

Index of Complex Networks:

https://icon.colorado.edu

Practical 1, exercise 4