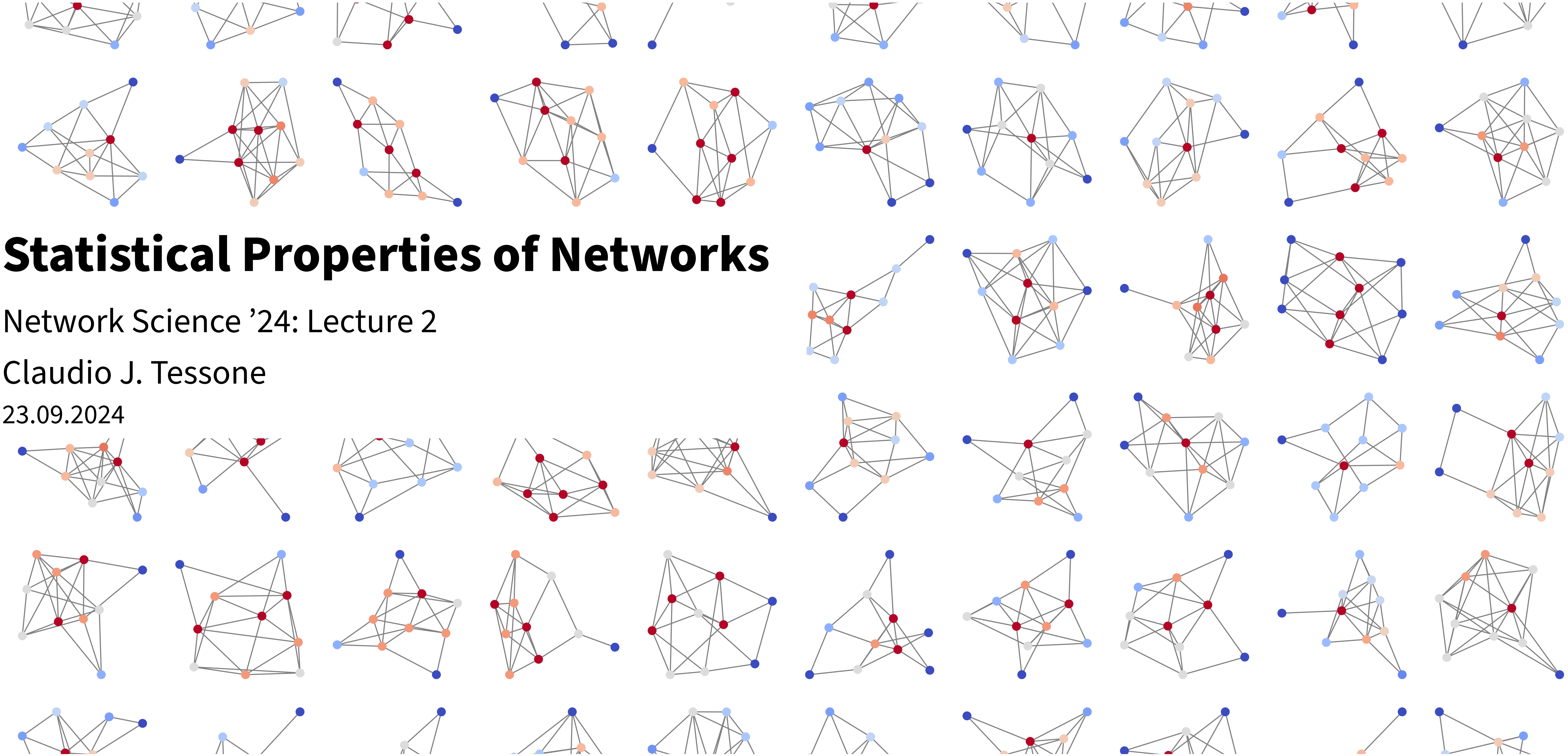
Universität Zürich UZH

UZH Blockchain Center

Blockchain & DLT Research Group



Statistical Properties of Networks

Network Science ’24: Lecture 2

Claudio J. Tessone

23.09.2024

1

Outlook

Lecture Objectives

- Which types of distributions are most commonly found in real-world networks?
- We will go beyond degree distributions to study *higher order network properties*: Assortativity and clustering
- We will see which networks display which kind of properties
- We will begin to understand the necessity of using benchmark models to compare with our data

Contents

1. Outlook

2. Recap Lecture 1

3. Simple networks

4. Paths

5. Higher-order properties
5.1 Reciprocity
5.2 Clustering
5.3 Assortativity

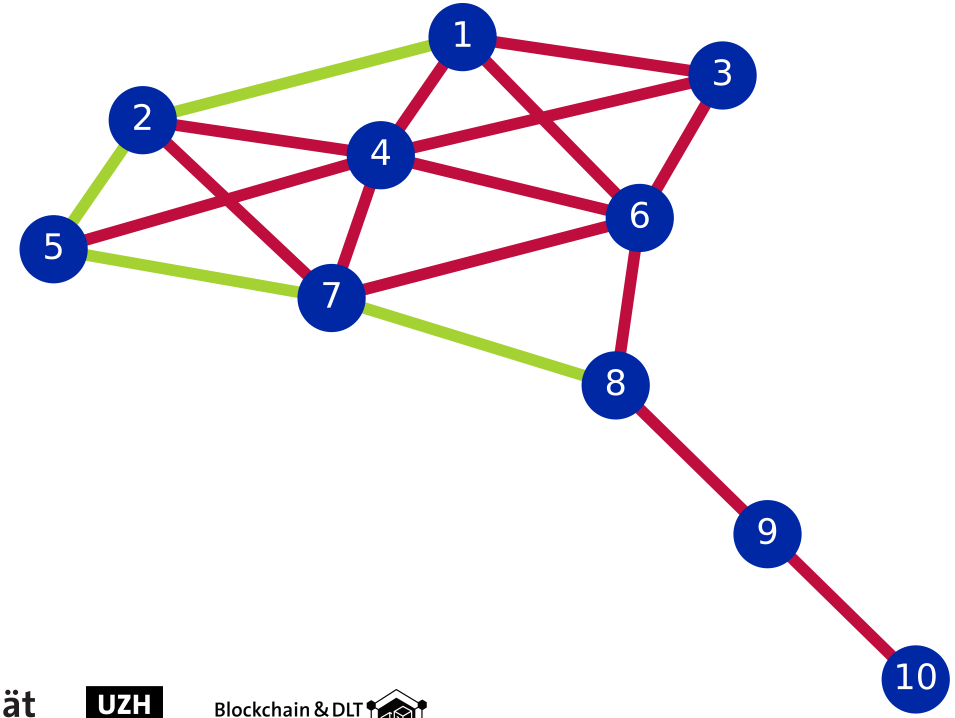
6. Network randomisation

Recap Lecture 1

2

Path Definition

A path is an alternating sequence of nodes and edges, beginning and ending at given node(s). *Nodes or vertices are visited only once*



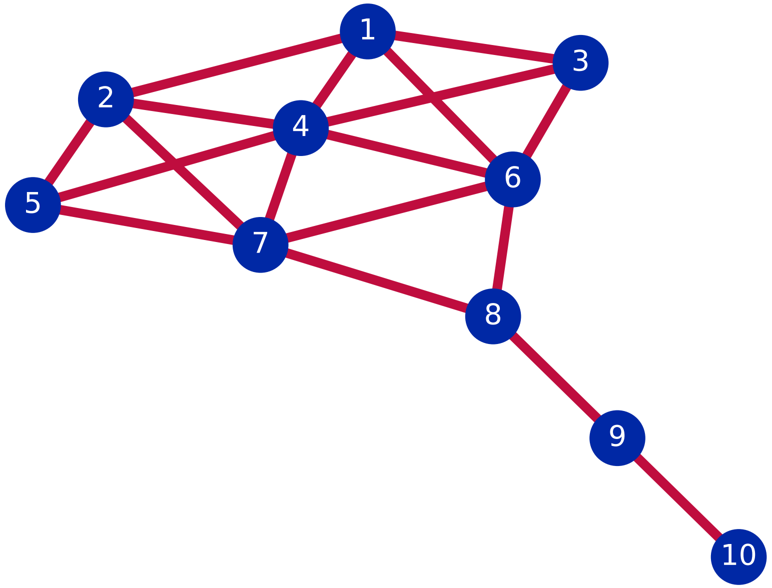
1 - 2 - 5 - 7 - 8
is a **path**

Network representation

Points to...

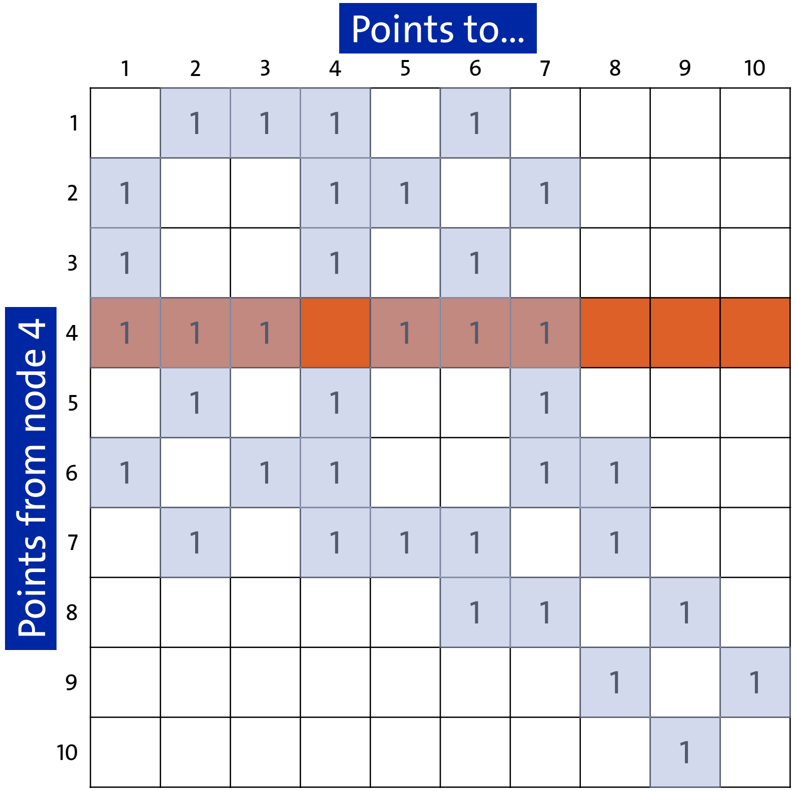
Points from ...

	1	2	3	4	5	6	7	8	9	10
1		1	1	1		1				
2	1			1	1		1			
3	1			1		1				
4	1	1	1		1	1	1			
5		1		1			1			
6	1		1	1			1	1		
7		1		1	1	1		1		
8						1	1		1	
9								1		1
10									1	



- The adjacency matrix $\mathbf{A} = \{a_{ij}\}_{i,j=1}^{N}$ is a representation of the network
- Symmetric for undirected networks

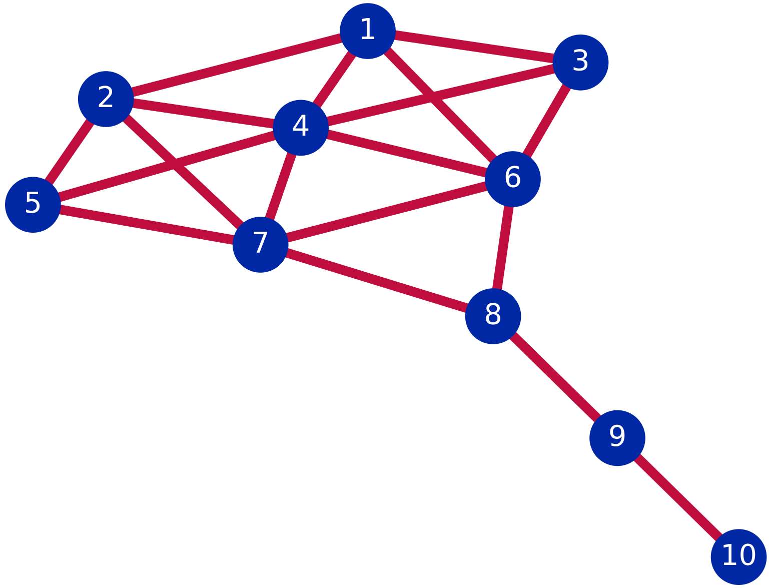
Node degree



Points to...

Points from node 4

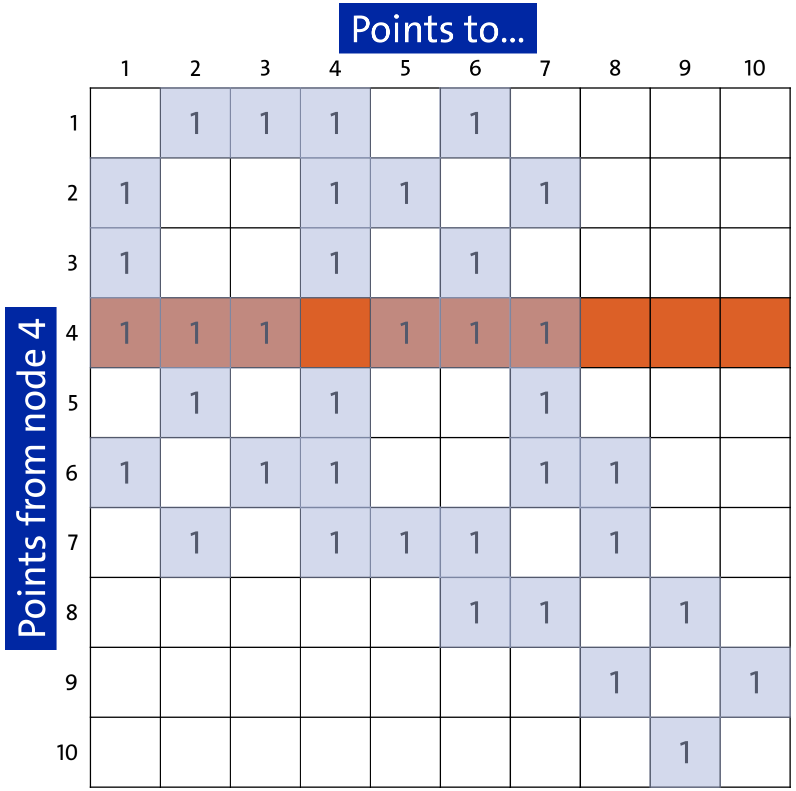
	1	2	3	4	5	6	7	8	9	10
1		1	1	1		1				
2	1			1	1		1			
3	1			1		1				
4	1	1	1		1	1	1			
5		1		1			1			
6	1		1	1			1	1		
7		1		1	1	1		1		
8						1	1		1	
9								1		1
10									1	



The degree of a node can be computed as a sum over the adjacency matrix

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

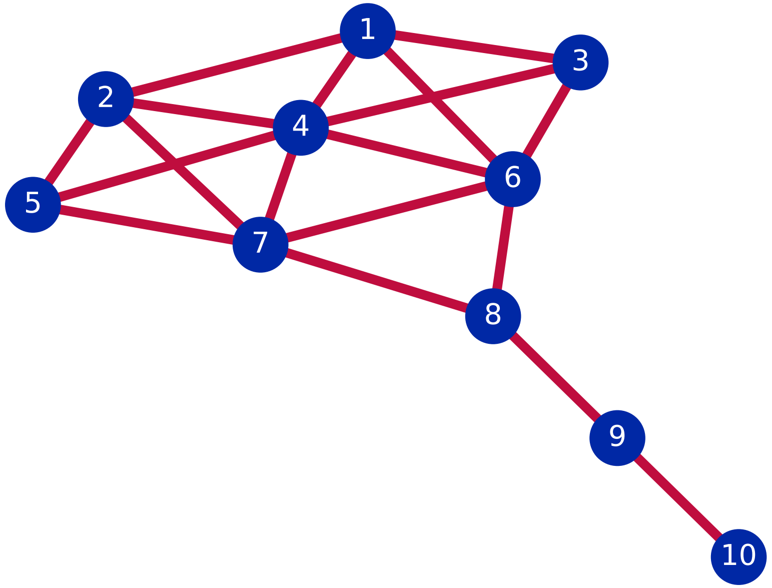
Node degree



Points to...

Points from node 4

	1	2	3	4	5	6	7	8	9	10
1		1	1	1		1				
2	1			1	1		1			
3	1			1		1				
4	1	1	1		1	1	1			
5		1		1			1			
6	1		1	1			1	1		
7		1		1	1	1		1		
8						1	1		1	
9								1		1
10									1	



The total number of links $L = |V|$ can be computed as a sum over the adjacency matrix (for unweighted/undirected networks)

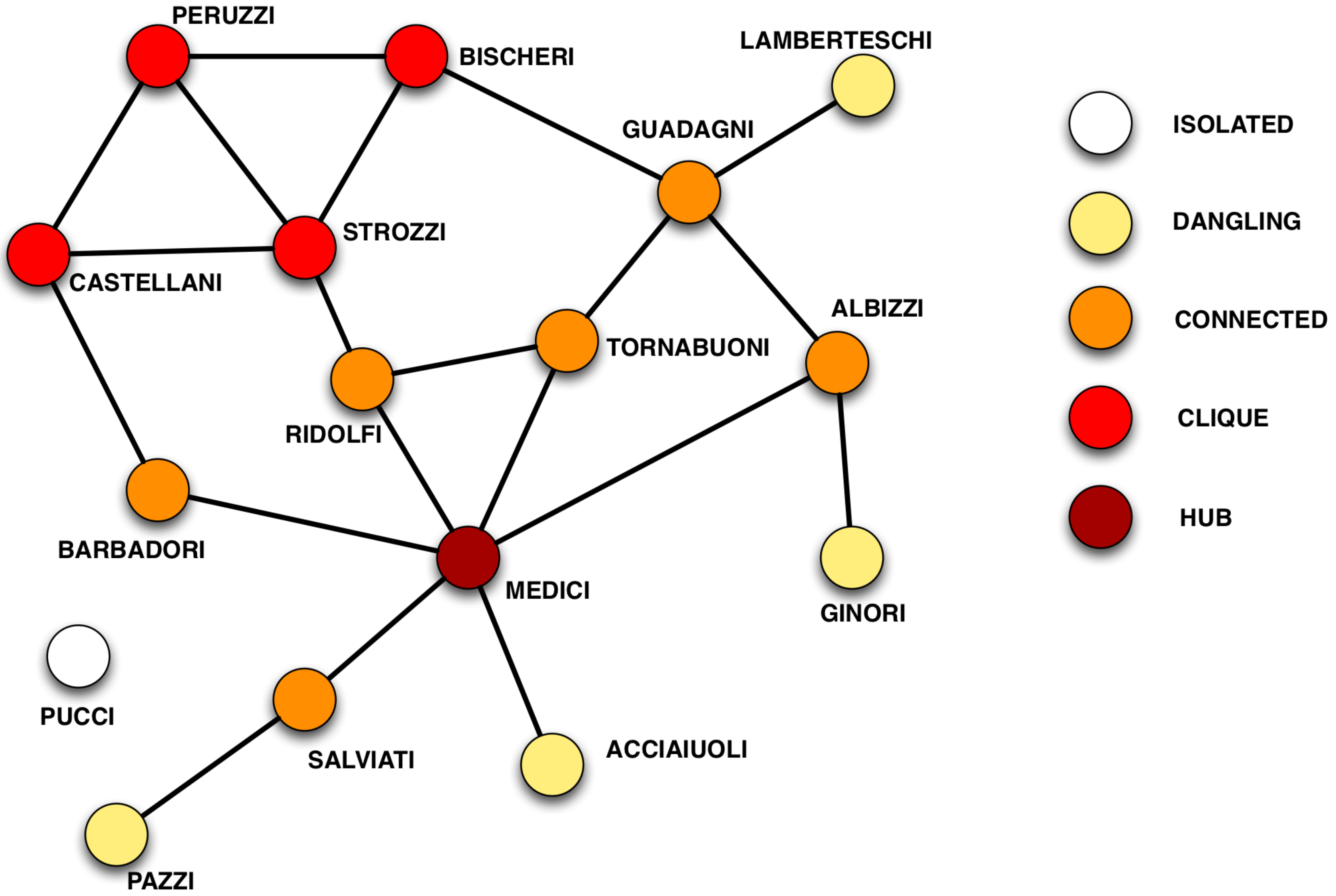
$$L = \frac{1}{2}\sum_{i,j=1}^{N} a_{ij}$$

Node degree: interpretation

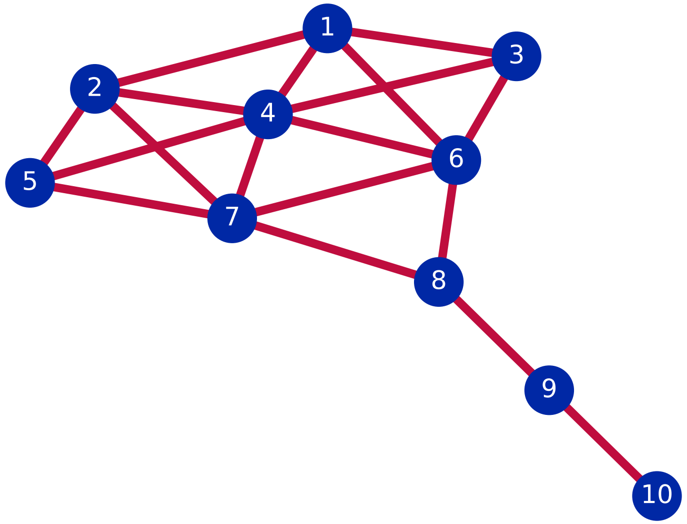
Node degree may clarify
the role of vertices.

Node degree: interpretation

Node degree may clarify the role of vertices.
Graph of the relationships of the Florence families in the Renaissance



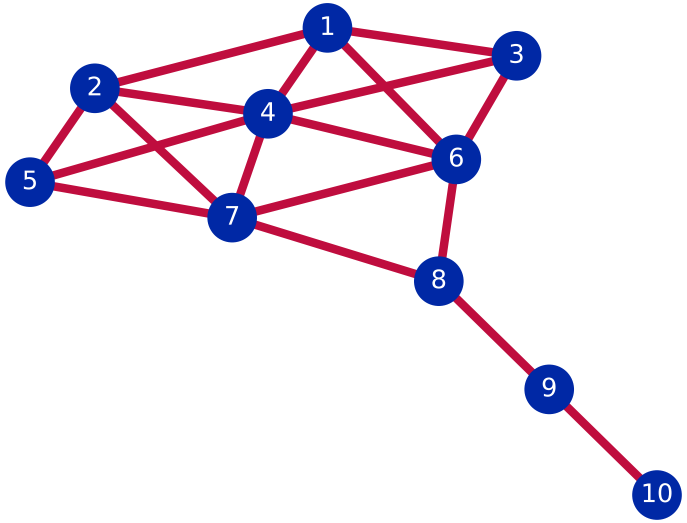
Degree distribution



Degree distribution is the fraction of nodes with a given degree $p(k)$

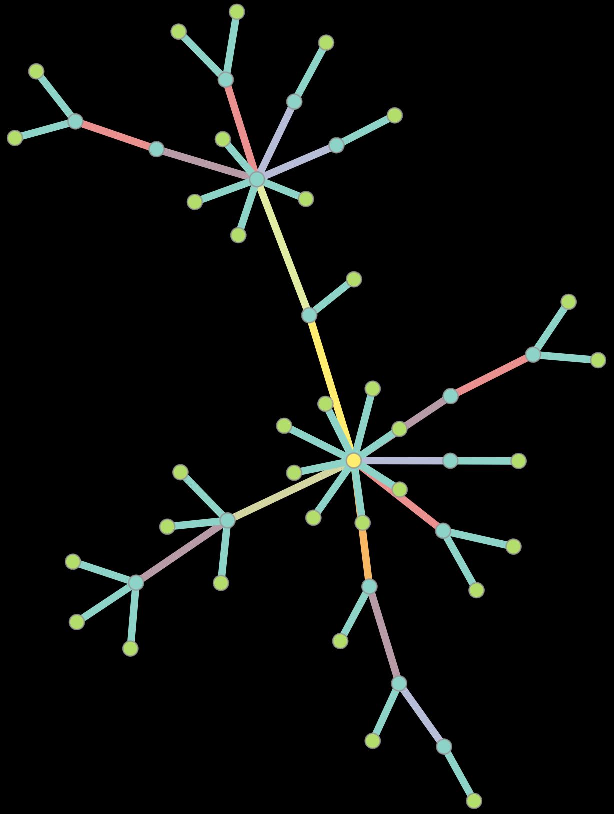
- $p(k = 1) = 1/10$ there is a single node with degree 1
- $p(k = 5) = 2/10$ there are two nodes with degree 5

Degree distribution

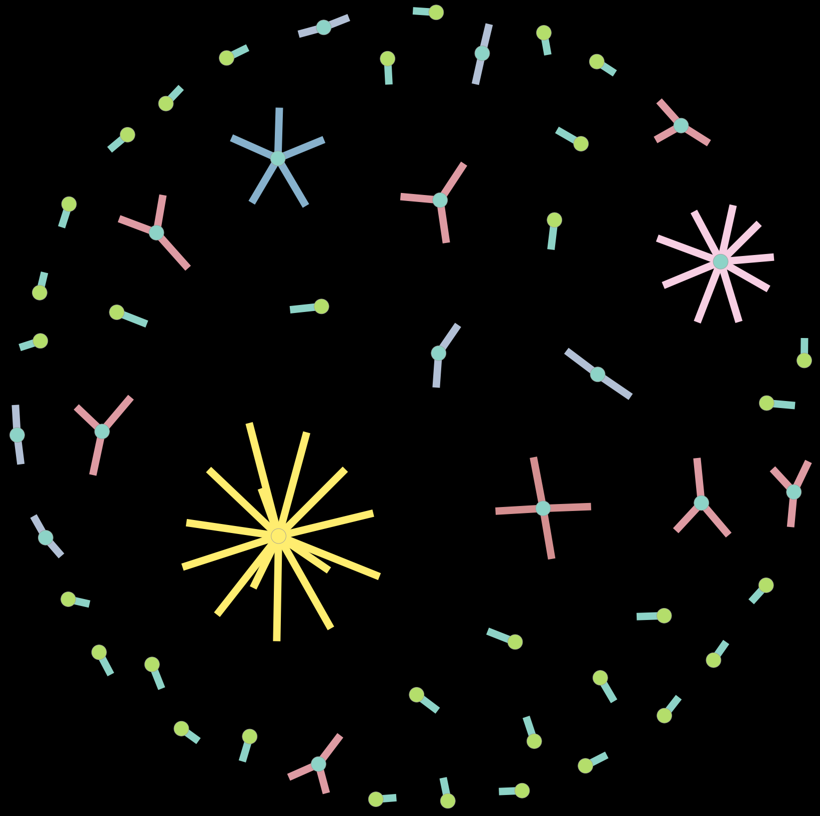


Degree	k	0	1	2	3	4	5	6
Degree distribution	$p(k)$	0	0.1	0.1	0.3	0.2	0.2	0.1

So far we have seen networks ...



… like this



Degree distribution

$$p(k) = \mathbb{P}(k_i = k)$$

In many contexts of interest, the degree distribution does not have a characteristic scale

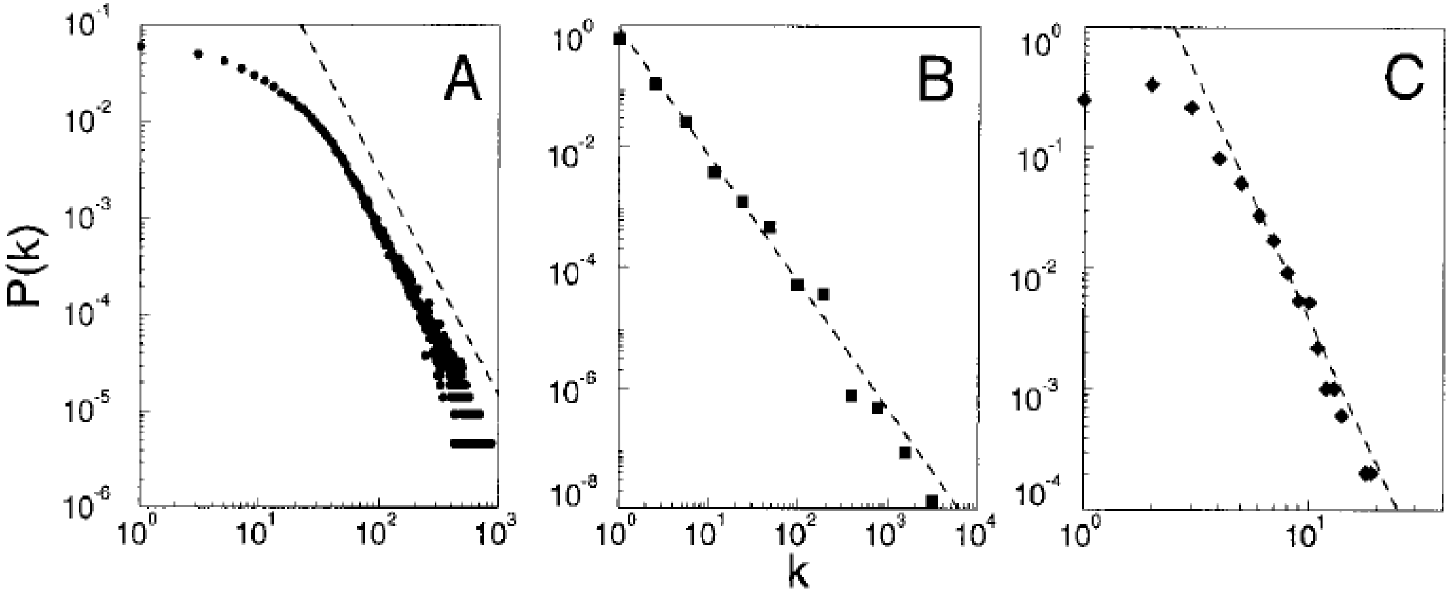


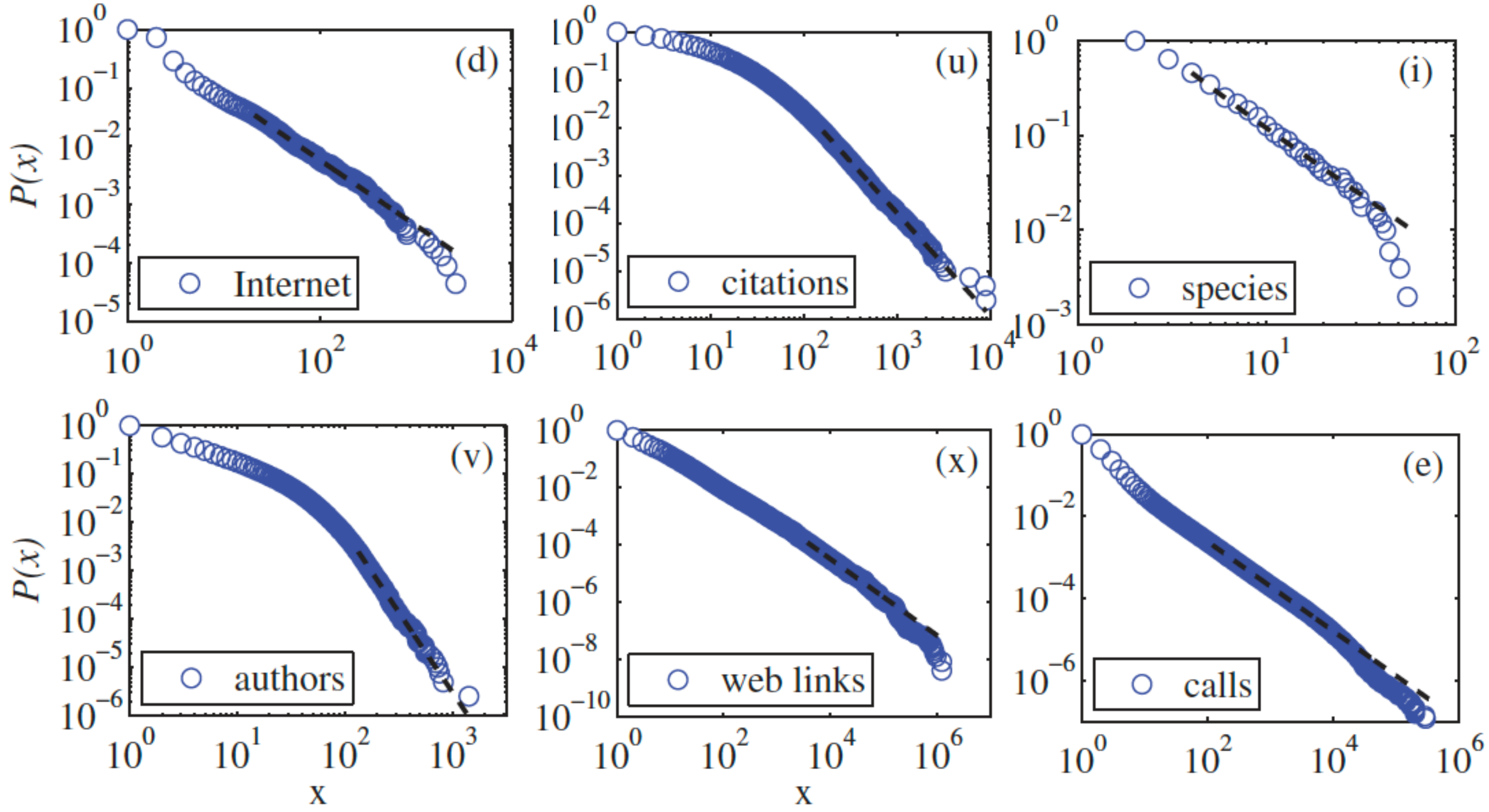
Fig. 1. The distribution function of connectivities for various large networks. (**A**) Actor collaboration graph with $N = 212{,}250$ vertices and average connectivity $\langle k \rangle = 28.78$. (**B**) WWW, $N = 325{,}729$, $\langle k \rangle = 5.46$ (*6*). (**C**) Power grid data, $N = 4941$, $\langle k \rangle = 2.67$. The dashed lines have slopes (A) $\gamma_{\text{actor}} = 2.3$, (B) $\gamma_{\text{www}} = 2.1$ and (C) $\gamma_{\text{power}} = 4$.

Barabási **Science**[1999]

Degree distribution

$$p(k) = \mathbb{P}(k_i = k)$$

In many contexts of interest, the degree distribution does not have a characteristic scale



Clauset **SIAM Review**[2009]

Important Notations (i)

$\langle \cdot \rangle$ represents the average of the argument

If the index is obvious, it is omitted

Examples

1. Average degree of a network can be denoted by

$$\langle \mathbf{k} \rangle = \langle k_i \rangle = \frac{1}{N} \sum_i k_i$$

Important Notations (ii)

δ_{ij} represents the Kronecker delta, where i, j are integer numbers

$$\delta_{ij} = \begin{cases} 1, & \text{if } \; i = j \\ 0, & \text{otherwise} \end{cases}$$

Examples

1. Number of nodes with degree k in a network is given by

$$\sum_i \delta_{k_i,k}$$

Some examples

- **Average network degree:**

$$\langle \mathbf{k} \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i$$

- **Average link strength (for weighted networks):**

$$\langle \mathbf{s} \rangle = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} s_{ij}$$

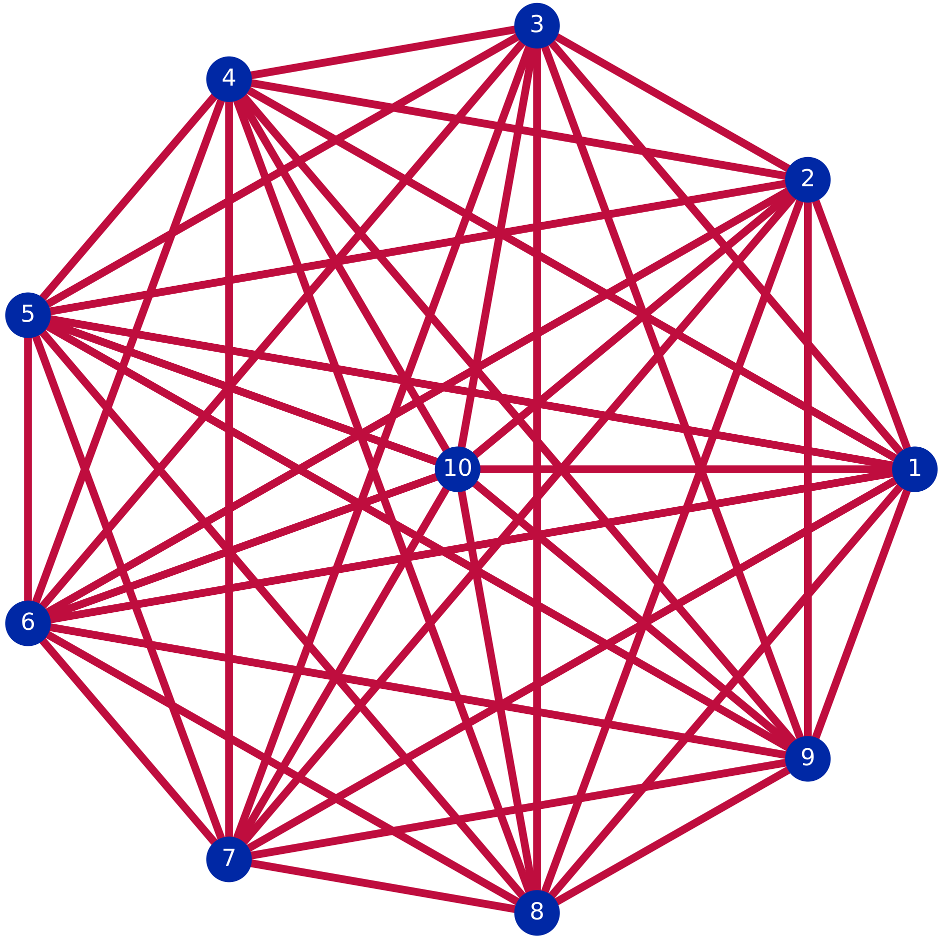
- Remember that N always represents the number of nodes

3

Simple networks

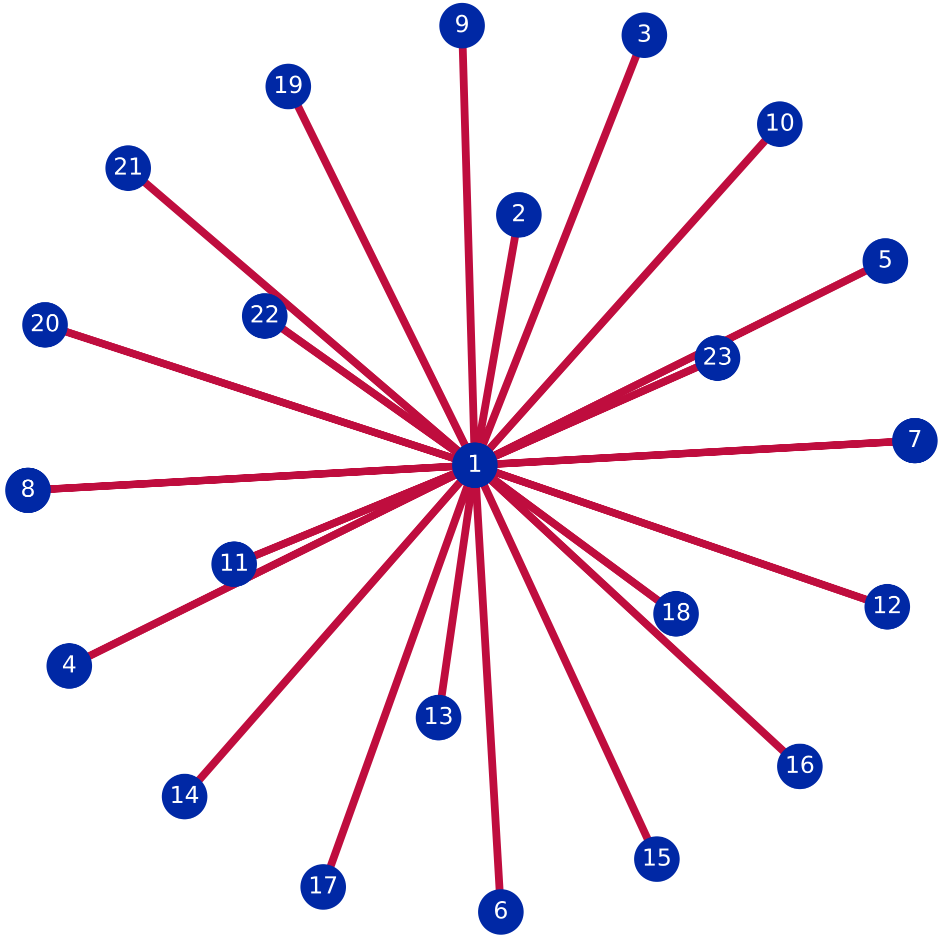
Basic networks: Fully connected

- Underlying network when all nodes are interconnected
- *In most situations: artificial, however, it is the base topology used in most analytical derivations*
- For N vertices: $k_i = N - 1$

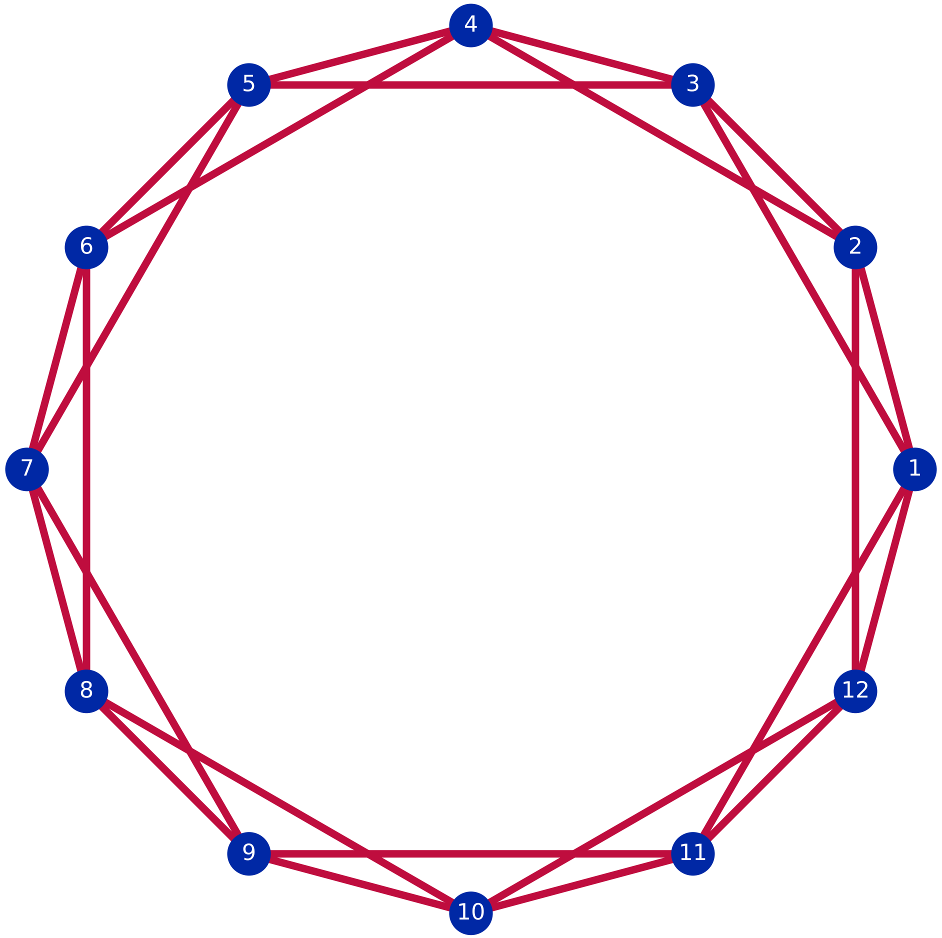


Basic networks: Star network

- Extreme centralisation / hierarchical organisation
- For N vertices: $k_1 = N - 1$; $k_i = 1$, for $i > 1$

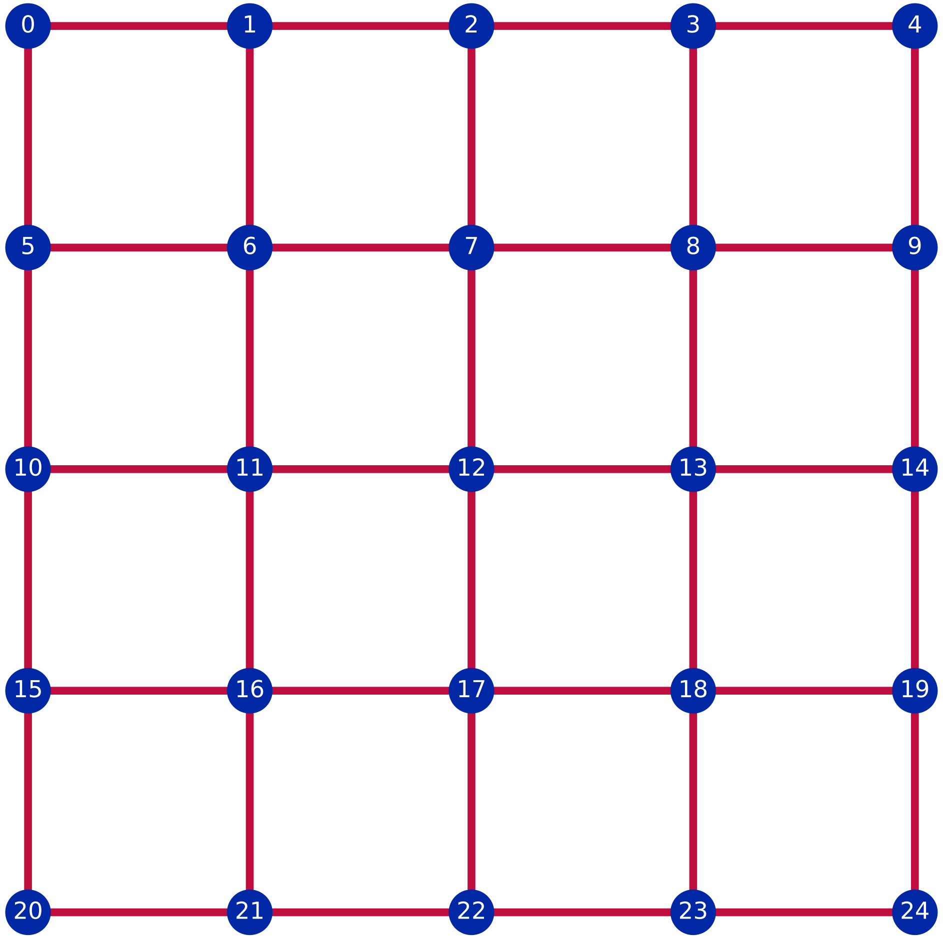


One-Dimensional Lattice



- Nodes are placed in a linear space. They are connected to κ nearest neighbours to the right and left
- κ is called the **coordination number**: $k_i = 2\kappa$
- **periodic boundary conditions**: the left-most node is a neighbour of the right-most node.

Two-Dimensional Lattice



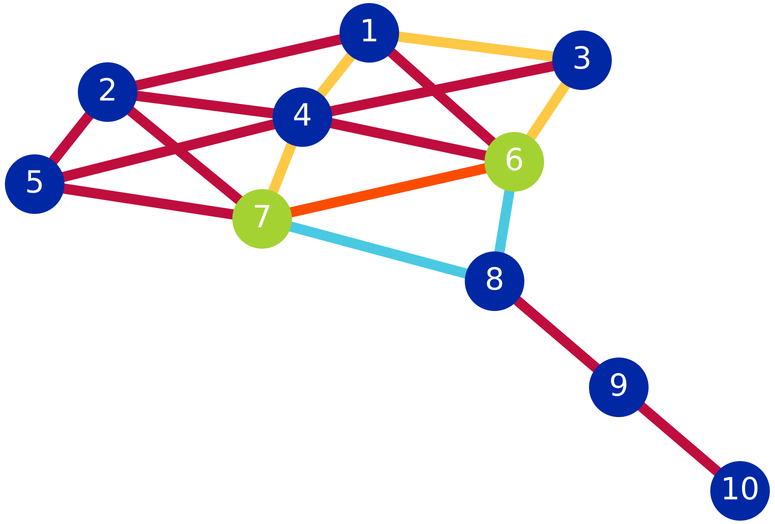
- nodes are placed in a bi-dimensional space. They are connected to κ nearest neighbours to east, west, north, south
- $k_i = 4\kappa$

4

Paths

Path Length

The number of links in a path is called the *Path Length*



Several paths and path length examples between node 6 and node 7:

- 6→3→1→4→7, path length: 4
- 6→8→7, path length: 2
- 6→7, path length: 1

There may be multiple paths between two nodes

The Shortest path between two nodes is the path with the shortest path length between them

Using Matrices to Count Walks

We can characterize the Shortest Path Length. Let's start with the number of walks:

Theorem

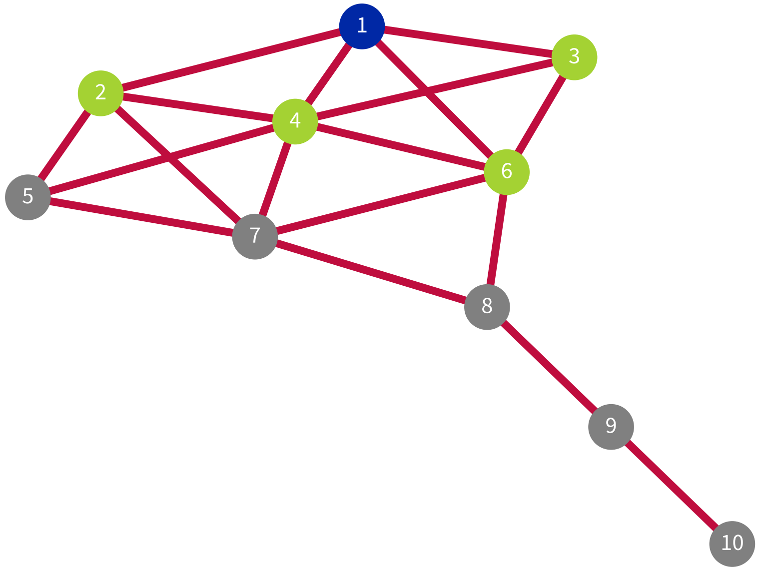
The number of walks of length r between two nodes i and j is equal to A^r_{ij}

Using Matrices to Count Walks

We can characterize the Shortest Path Length. Let’s start with the number of walks:

Theorem

The number of walks of length r between two nodes i and j is equal to A^r_{ij}



… to…

points from …

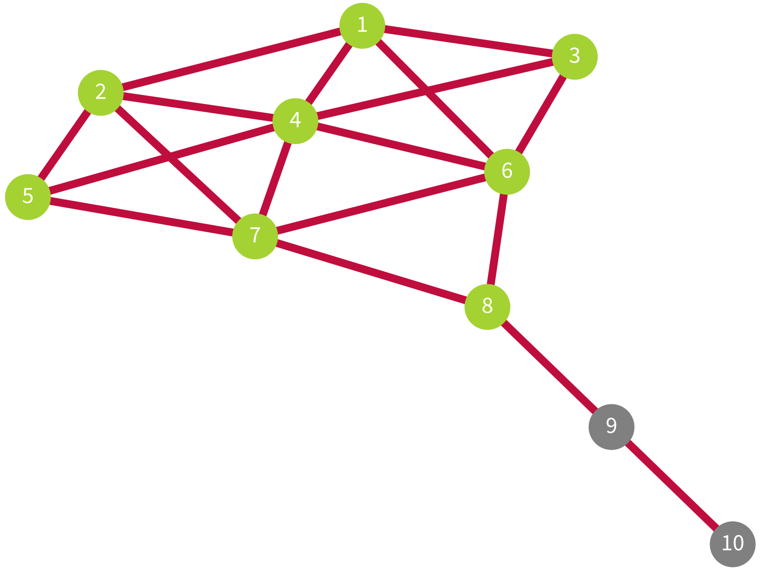
$$A = \begin{array}{c|cccccccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
\hline
1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
2 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\
3 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
4 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
5 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
6 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\
7 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\
8 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\
9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
10 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{array}$$

Using Matrices to Count Walks

We can characterize the Shortest Path Length. Let's start with the number of walks:

Theorem

The number of walks of length r between two nodes i and j is equal to A_{ij}^r



… to…

points from …

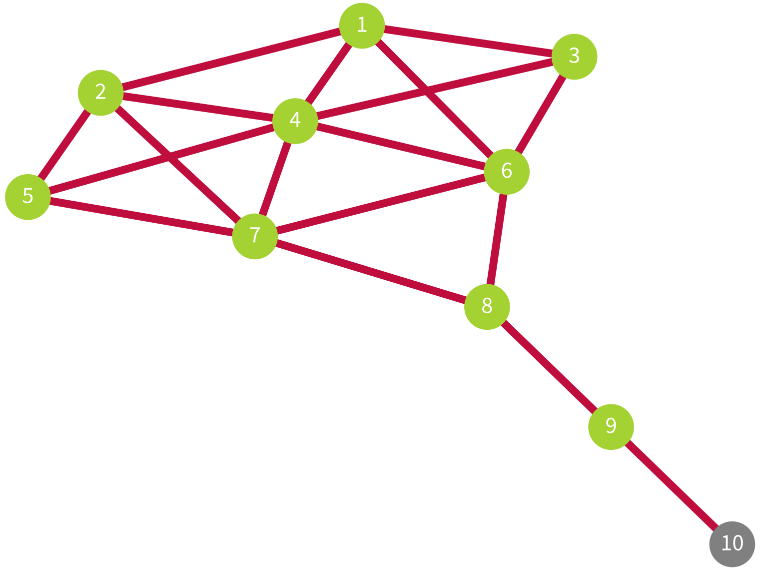
$$A^2 = \begin{array}{c|cccccccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
\hline
1 & 4 & 1 & 2 & 3 & 2 & 2 & 3 & 1 & 0 & 0 \\
2 & 1 & 4 & 2 & 3 & 2 & 3 & 2 & 1 & 0 & 0 \\
3 & 2 & 2 & 3 & 2 & 1 & 2 & 2 & 1 & 0 & 0 \\
4 & 3 & 3 & 2 & 6 & 2 & 3 & 3 & 2 & 0 & 0 \\
5 & 2 & 2 & 1 & 2 & 3 & 2 & 2 & 1 & 0 & 0 \\
6 & 2 & 3 & 2 & 3 & 2 & 5 & 2 & 1 & 1 & 0 \\
7 & 3 & 2 & 2 & 3 & 2 & 2 & 5 & 1 & 1 & 0 \\
8 & 1 & 1 & 1 & 2 & 1 & 1 & 1 & 3 & 0 & 1 \\
9 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 2 & 0 \\
10 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1
\end{array}$$

Using Matrices to Count Walks

We can characterize the Shortest Path Length. Let's start with the number of walks:

Theorem

The number of walks of length r between two nodes i and j is equal to A^r_{ij}



… *to*…

points from …

$$A^3 = \begin{array}{c|cccccccccc} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ \hline 1 & 8 & 12 & 9 & 14 & 7 & 13 & 9 & 5 & 1 & 0 \\ 2 & 12 & 8 & 7 & 14 & 9 & 9 & 13 & 5 & 1 & 0 \\ 3 & 9 & 7 & 6 & 12 & 6 & 10 & 8 & 4 & 1 & 0 \\ 4 & 14 & 14 & 12 & 16 & 12 & 16 & 16 & 6 & 2 & 0 \\ 5 & 7 & 9 & 6 & 12 & 6 & 8 & 10 & 4 & 1 & 0 \\ 6 & 13 & 9 & 10 & 16 & 8 & 10 & 14 & 8 & 1 & 1 \\ 7 & 9 & 13 & 8 & 16 & 10 & 14 & 10 & 8 & 1 & 1 \\ 8 & 5 & 5 & 4 & 6 & 4 & 8 & 8 & 2 & 4 & 0 \\ 9 & 1 & 1 & 1 & 2 & 1 & 1 & 1 & 4 & 0 & 2 \\ 10 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 2 & 0 \end{array}$$

Reachability with Matrices

As a consequence, we obtain this useful corollary:

Corollary

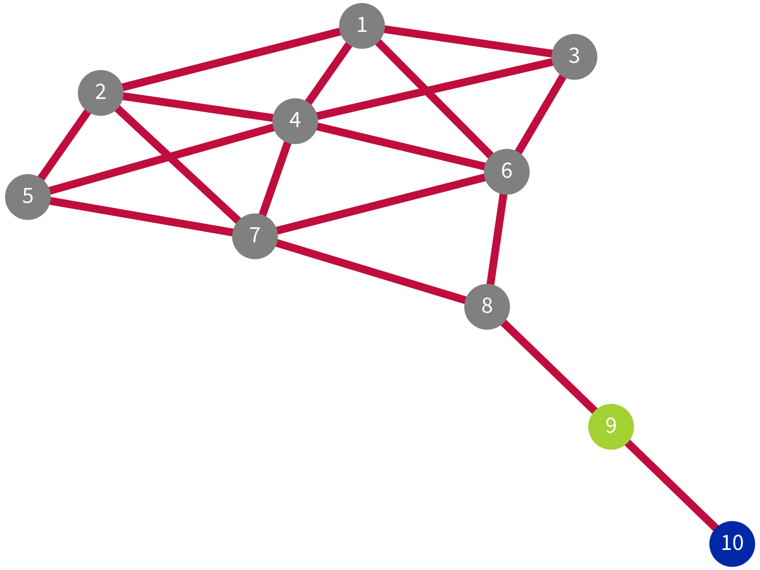
A node j is reachable from a node i if and only if it exists r such that $A^r_{ij} > 0$

Reachability with Matrices

As a consequence, we obtain this useful corollary:

Corollary

A node j is reachable from a node i if and only if it exists r such that $A^r_{ij} > 0$



… to…

points from …

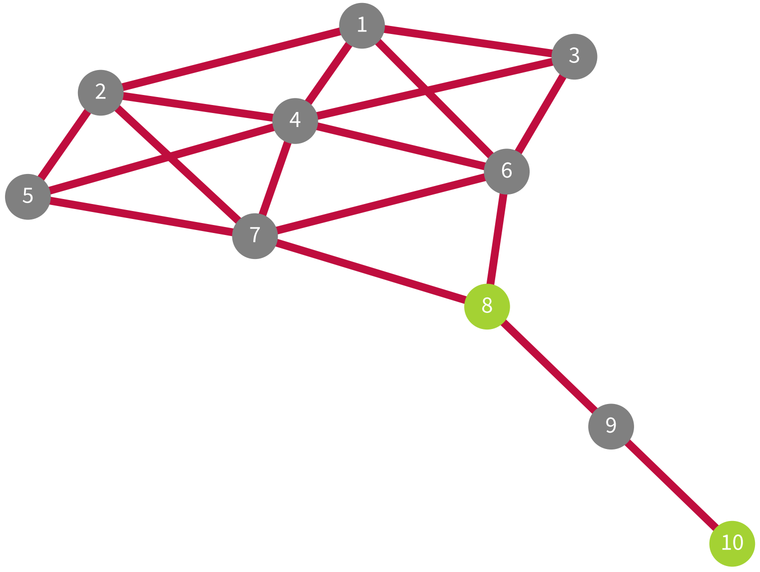
$$A = \begin{array}{c|cccccccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
\hline
1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
2 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\
3 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
4 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
5 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
6 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\
7 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\
8 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\
9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
10 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{array}$$

Reachability with Matrices

As a consequence, we obtain this useful corollary:

Corollary

A node j is reachable from a node i if and only if it exists r such that $A^r_{ij} > 0$



… to…

points from …

$$A^2 = \begin{array}{c|cccccccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
\hline
1 & 4 & 1 & 2 & 3 & 2 & 2 & 3 & 1 & 0 & 0 \\
2 & 1 & 4 & 2 & 3 & 2 & 3 & 2 & 1 & 0 & 0 \\
3 & 2 & 2 & 3 & 2 & 1 & 2 & 2 & 1 & 0 & 0 \\
4 & 3 & 3 & 2 & 6 & 2 & 3 & 3 & 2 & 0 & 0 \\
5 & 2 & 2 & 1 & 2 & 3 & 2 & 2 & 1 & 0 & 0 \\
6 & 2 & 3 & 2 & 3 & 2 & 5 & 2 & 1 & 1 & 0 \\
7 & 3 & 2 & 2 & 3 & 2 & 2 & 5 & 1 & 1 & 0 \\
8 & 1 & 1 & 1 & 2 & 1 & 1 & 1 & 3 & 0 & 1 \\
9 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 2 & 0 \\
10 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1
\end{array}$$

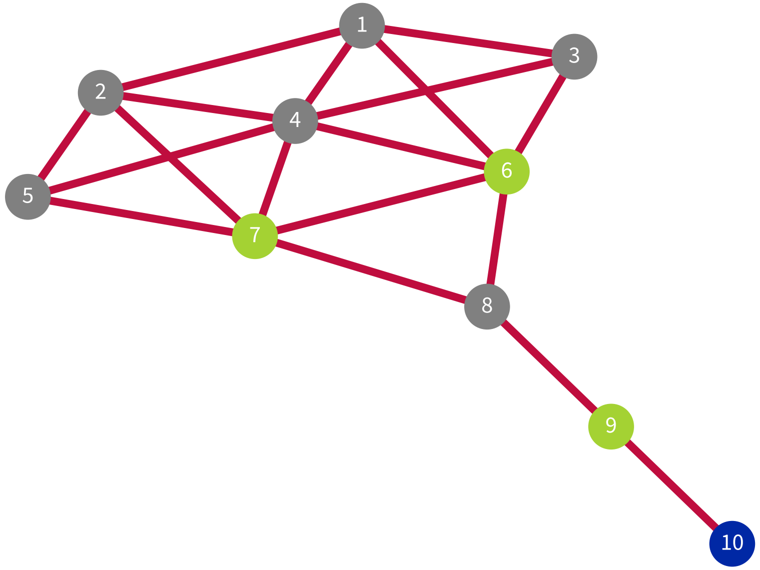
Universität Zürich$^{\text{UZH}}$ UZH Blockchain Center Blockchain & DLT Research Group

Reachability with Matrices

As a consequence, we obtain this useful corollary:

Corollary

A node j is reachable from a node i if and only if it exists r such that $A^r_{ij} > 0$



$\ldots$ *to*$\ldots$

points from $\ldots$

$$A^3 = \begin{array}{c|cccccccccc} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ \hline 1 & 8 & 12 & 9 & 14 & 7 & 13 & 9 & 5 & 1 & 0 \\ 2 & 12 & 8 & 7 & 14 & 9 & 9 & 13 & 5 & 1 & 0 \\ 3 & 9 & 7 & 6 & 12 & 6 & 10 & 8 & 4 & 1 & 0 \\ 4 & 14 & 14 & 12 & 16 & 12 & 16 & 16 & 6 & 2 & 0 \\ 5 & 7 & 9 & 6 & 12 & 6 & 8 & 10 & 4 & 1 & 0 \\ 6 & 13 & 9 & 10 & 16 & 8 & 10 & 14 & 8 & 1 & 1 \\ 7 & 9 & 13 & 8 & 16 & 10 & 14 & 10 & 8 & 1 & 1 \\ 8 & 5 & 5 & 4 & 6 & 4 & 8 & 8 & 2 & 4 & 0 \\ 9 & 1 & 1 & 1 & 2 & 1 & 1 & 1 & 4 & 0 & 2 \\ 10 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 2 & 0 \end{array}$$

Shortest Path Length

Corollary

The Shortest Path Length between two nodes is:

$$d_{ij} = argmin_r(A_{ij}^r \neq 0)$$

- By definition, the shortest path *between a node and itself* is zero $d_{ii} = 0$.
- If there is *no path* connecting two nodes, shortest path is defined to be *infinite* $d_{ij} = \infty$ if $A_{ij}^r = 0 \, \forall r$
- The Shortest Path between two nodes is also called **Distance**.

Related concepts

— The **Average Distance** is the Average Shortest Path between all pairs of nodes:

$$\langle \mathbf{d} \rangle = \langle d_{ij} \rangle = \frac{1}{N(N-1)} \sum_{i,j=1}^{N} d_{ij}$$

Related concepts

- The **Average Distance** is the Average Shortest Path between all pairs of nodes:

$$\langle \mathbf{d} \rangle = \langle d_{ij} \rangle = \frac{1}{N(N-1)} \sum_{i,j=1}^{N} d_{ij}$$

- The **Diameter** of a graph is the maximum distance between any pair of nodes in the graph:

$$d_{max} = \max_{i,j} d_{ij}$$

Related concepts

- The **Average Distance** is the Average Shortest Path between all pairs of nodes:

$$\langle \mathbf{d} \rangle = \langle d_{ij} \rangle = \frac{1}{N(N-1)} \sum_{i,j=1}^{N} d_{ij}$$

- The **Diameter** of a graph is the maximum distance between any pair of nodes in the graph:

$$d_{max} = \max_{i,j} d_{ij}$$

- The **Eccentricity** of a node v is the greatest distance between v and any other node:

$$e_i = \max_{j \in N} (d_{ij})$$

Related concepts

- The **Average Distance** is the Average Shortest Path between all pairs of nodes:

$$\langle \mathbf{d} \rangle = \langle d_{ij} \rangle = \frac{1}{N(N-1)} \sum_{i,j=1}^{N} d_{ij}$$

- The **Diameter** of a graph is the maximum distance between any pair of nodes in the graph:

$$d_{max} = \max_{i,j} d_{ij}$$

- The **Eccentricity** of a node v is the greatest distance between v and any other node:

$$e_i = \max_{j \in N} (d_{ij})$$

- A **peripheral node** is the one whose eccentricity is equal to the network diameter.

5

Higher-order properties

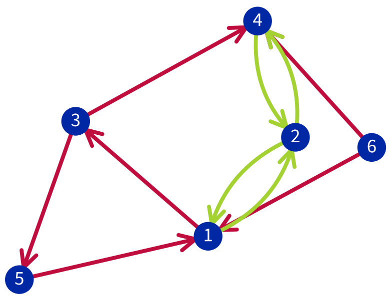
In a directed network: Are relationships **mutual**? If i claims to be friend of j, does j feel likewise about i?

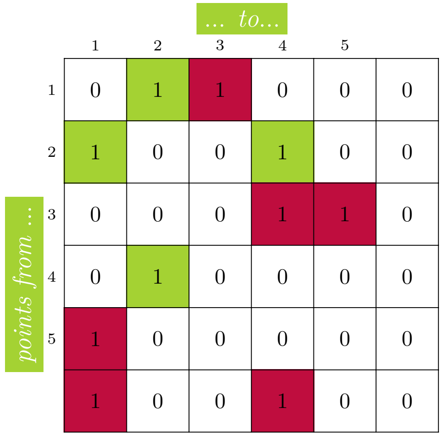
Reciprocity

For *directed* networks, reciprocity can be seen as a measure of how mutual the relationships are

Reciprocity

For *directed* networks, reciprocity can be seen as a measure of how mutual the relationships are





… to…

points from …

	1	2	3	4	5	
1	0	1	1	0	0	0
2	1	0	0	1	0	0
3	0	0	0	1	1	0
4	0	1	0	0	0	0
5	1	0	0	0	0	0
	1	0	0	1	0	0

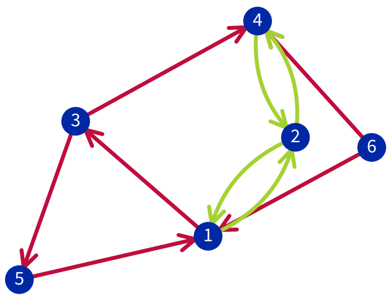
R = 0.4

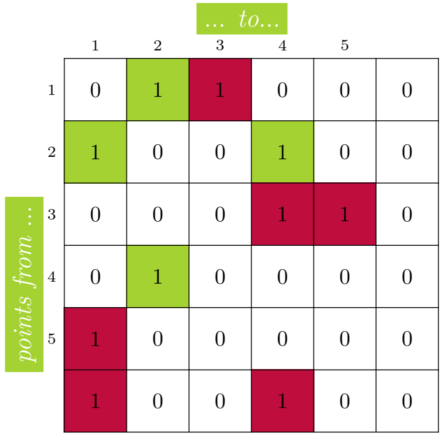
$$R = \frac{\sum_{u,v \in V} a_{u,v} a_{v,u}}{\sum_{u,v \in V} a_{u,v}}$$

- $L^{\leftrightarrow} = \sum_{u,v \in V} a_{u,v} a_{v,u}$ is the total number of reciprocated links.
- $a_{u,v} a_{v,u}$ is only equal to 1 if both edges exist.
- $L = \sum_{u,v \in V} a_{u,v}$ is the total number of links.

Reciprocity

For *directed* networks, reciprocity can be seen as a measure of how mutual the relationships are





… to…

points from …

	1	2	3	4	5	
1	0	1	1	0	0	0
2	1	0	0	1	0	0
3	0	0	0	1	1	0
4	0	1	0	0	0	0
5	1	0	0	0	0	0
	1	0	0	1	0	0

R = 0.4

$$R = \frac{\sum_{u,v \in V} a_{u,v} a_{v,u}}{\sum_{u,v \in V} a_{u,v}}$$

- $L^{\leftrightarrow} = \sum_{u,v \in V} a_{u,v} a_{v,u}$ is the total number of reciprocated links.
- $a_{u,v} a_{v,u}$ is only equal to 1 if both edges exist.
- $L = \sum_{u,v \in V} a_{u,v}$ is the total number of links.

Reciprocity is a *second order property* because it involves two nodes product in one calculation

Empirical Results

Dataset	L	R
Friends and Family	698	0.45
Reality Mining	82	0.34
Social Evolution	1596	0.35
Strongest Ties	208	0.49

Tabelle: Almaatouq *et al.*

Social, Cultural, and Behavioral Modeling[2016]

Empirical evidence shows that social networks display high reciprocity R

How likely is it that two friends of someone are also friends among themselves?

Clustering coefficient

The clustering coefficient is a node property. It measures how likely it is that if two nodes are neighbours of a focal one, there is an edge between them

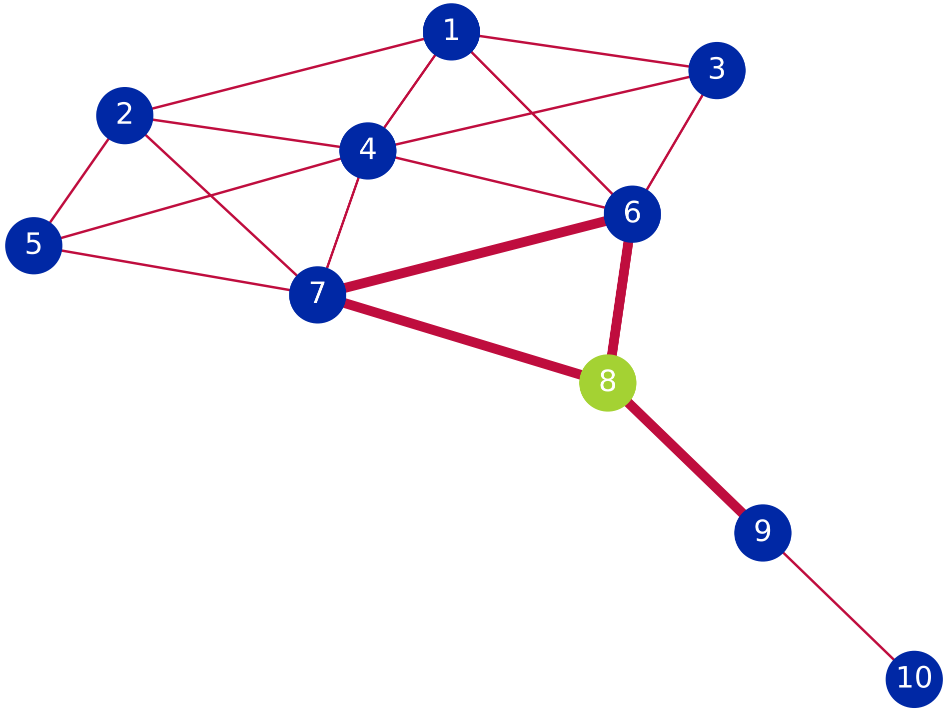
Clustering coefficient

The clustering coefficient is a node property. It measures how likely it is that if two nodes are neighbours of a focal one, there is an edge between them

$$C_i = \frac{\sum_{u,v \in V} a_{u,i} a_{v,i} a_{u,v}}{\sum_{u,v \in V} a_{u,i} a_{v,i}} = \frac{\sum_{u,v \in N_i} a_{u,i} a_{v,i} a_{u,v}}{\sum_{u,v \in N_i} a_{u,i} a_{v,i}} \tag{1}$$

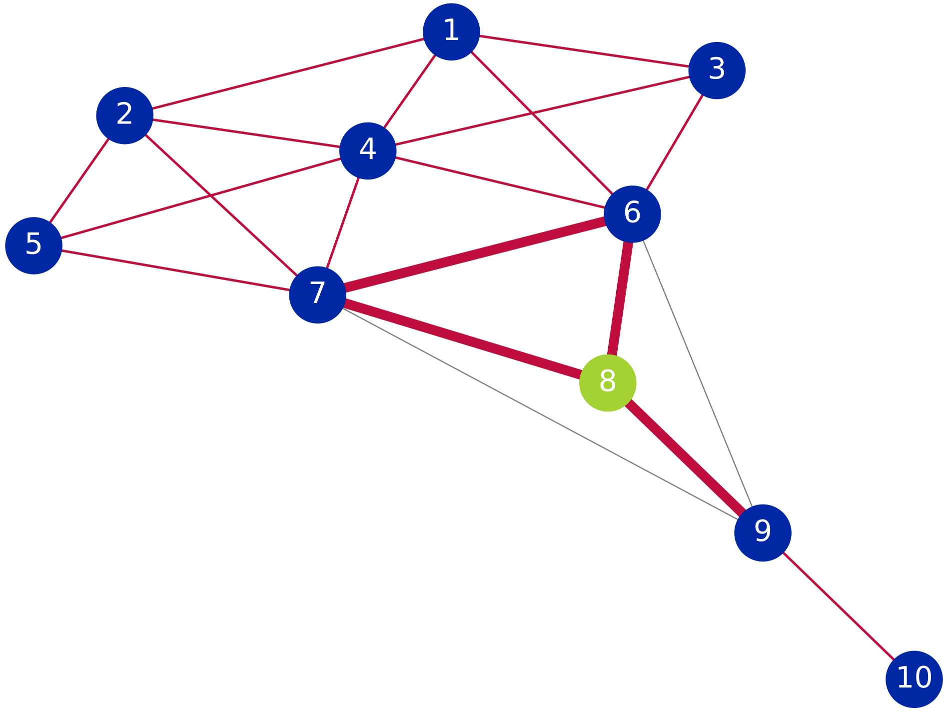
- N_i is the set of neighbours of node i

Clustering coefficient



$$C_8 = \frac{1}{3}$$

Clustering coefficient



Node	i	1	2	3	4	5	6	7	8	9	10
Clustering	C_i	2/3	2/3	1	16/30	1	1/2	1/2	1/3	0	0

Global clustering coefficient

$$\langle C \rangle = \frac{1}{N} \sum_i C_i$$

Clustering coefficient

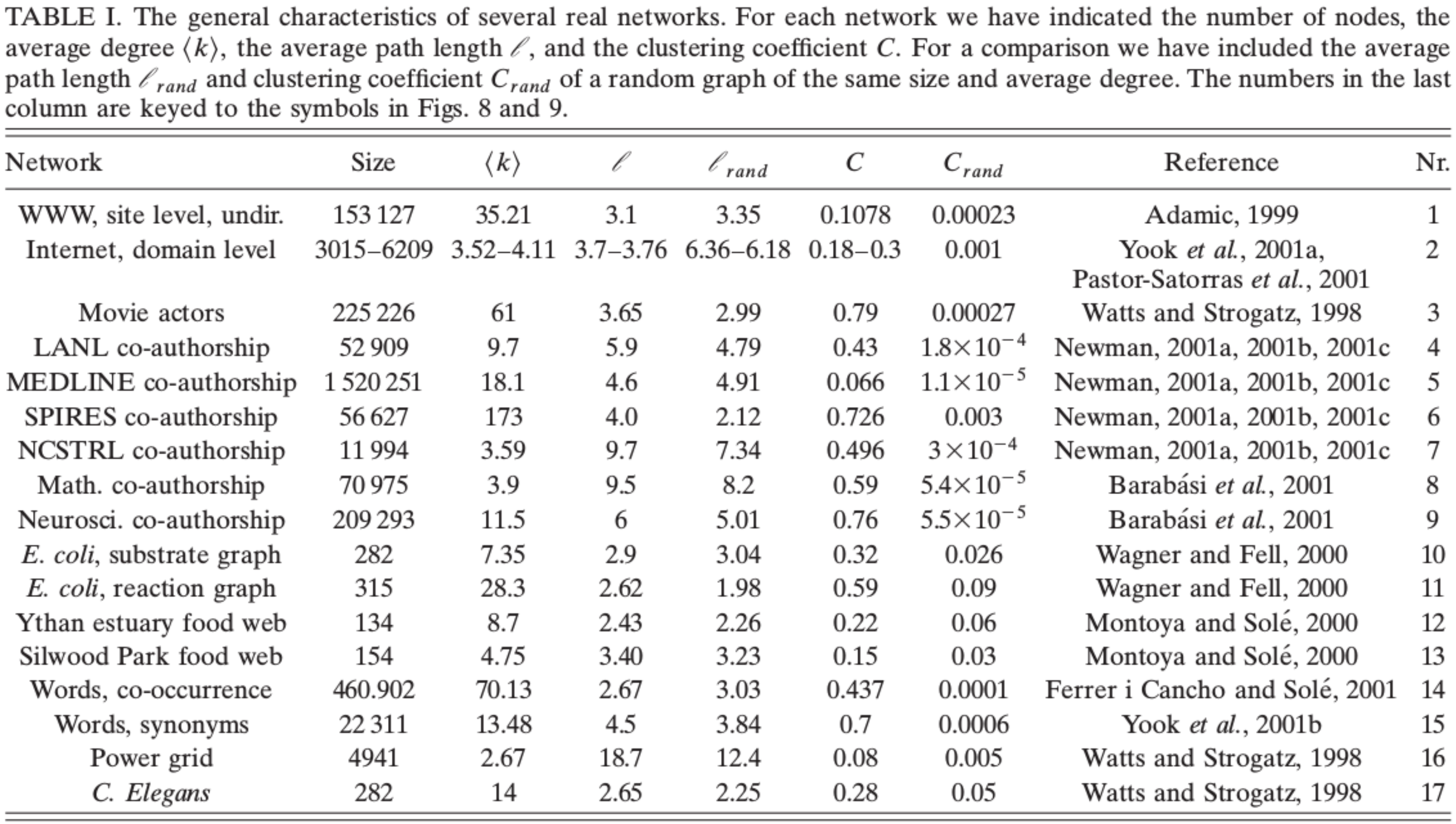


TABLE I. The general characteristics of several real networks. For each network we have indicated the number of nodes, the average degree $\langle k \rangle$, the average path length ℓ, and the clustering coefficient C. For a comparison we have included the average path length ℓ_{rand} and clustering coefficient C_{rand} of a random graph of the same size and average degree. The numbers in the last column are keyed to the symbols in Figs. 8 and 9.

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference	Nr.
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999	1
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook *et al.*, 2001a, Pastor-Satorras *et al.*, 2001	2
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998	3
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c	4
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c	5
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c	6
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c	7
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.*, 2001	8
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.*, 2001	9
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000	10
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000	12
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000	13
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001	14
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.*, 2001b	15
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998	16
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998	17

Most real world networks have **large** clustering coefficient

Is there homophily in the network? Are nodes connected to others similar to them or to others that are completely different?

Assortativity

1. Are low-degree nodes connected to other low-degree nodes? ... or are they mostly connected to high-degree nodes?
2. Are high-degree nodes connected to low-degree nodes or are they mostly connected to other high-degree nodes?
3. Or perhaps it does not matter at all?

A means of capturing the degree correlation is by examining the properties of neighbour connectivity

Nearest Neighbours's Degree

The **Average Nearest Neighbour Degree (ANND)** of node i is given by

$$k_{nn,i} = \frac{1}{k_i} \sum_{j \in N_i} k_j$$

Nearest Neighbours's Degree

The **Average Nearest Neighbour Degree (ANND)** of node i is given by

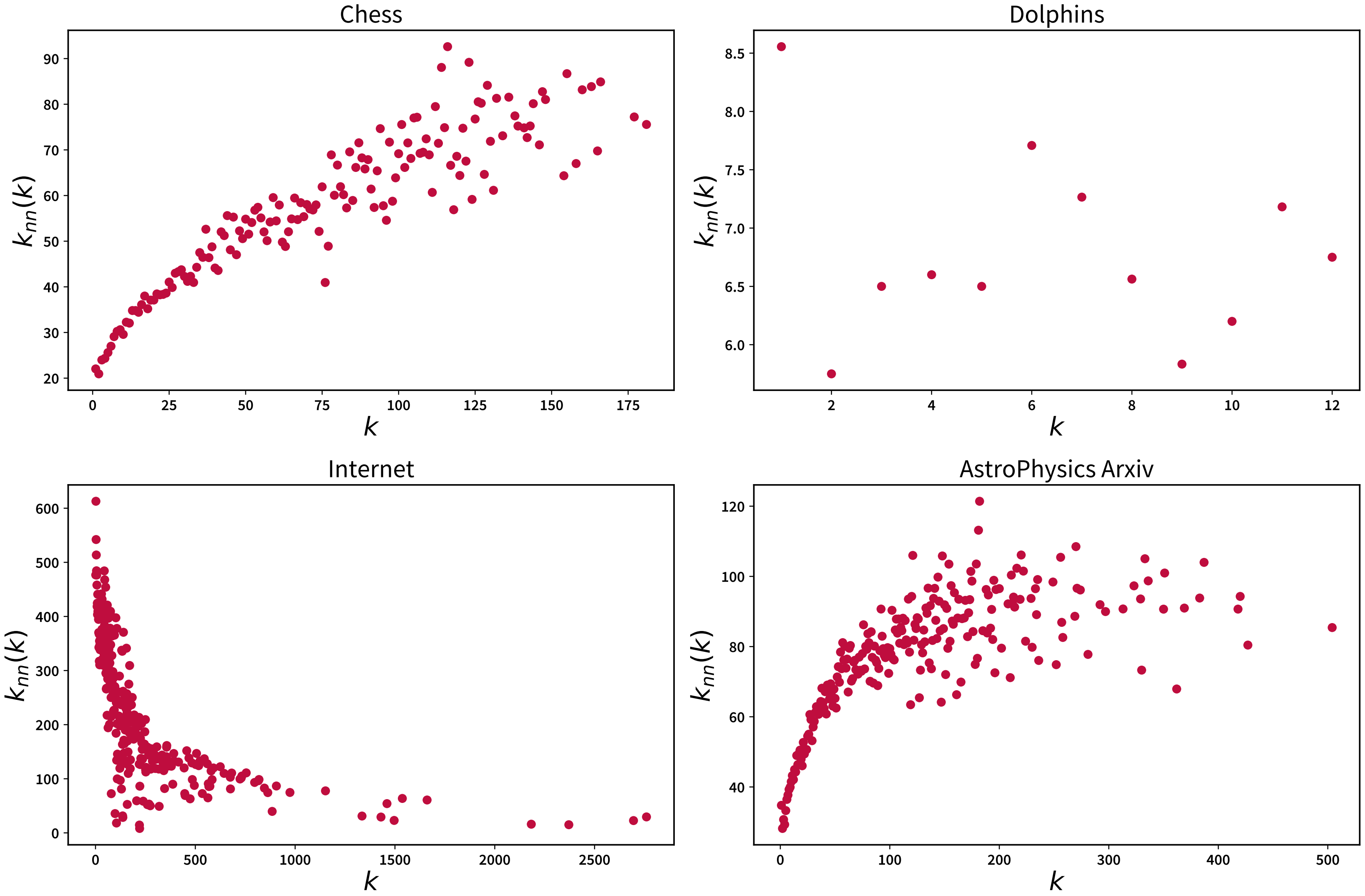
$$k_{nn,i} = \frac{1}{k_i} \sum_{j \in N_i} k_j$$

From the ANND of a node one can generalize to the Average Degree of the Nearest Neighbors as a function of the degree k:

$$k_{nn}(k) = \frac{\sum_i \delta_{k,k_i} k_{nn,i}}{\sum_i \delta_{k,k_i}}$$

- If $k_{nn}(k)$ is an increasing function on k, the network is **assortative**
- If $k_{nn}(k)$ is a decreasing function on k, the network is **disassortative**

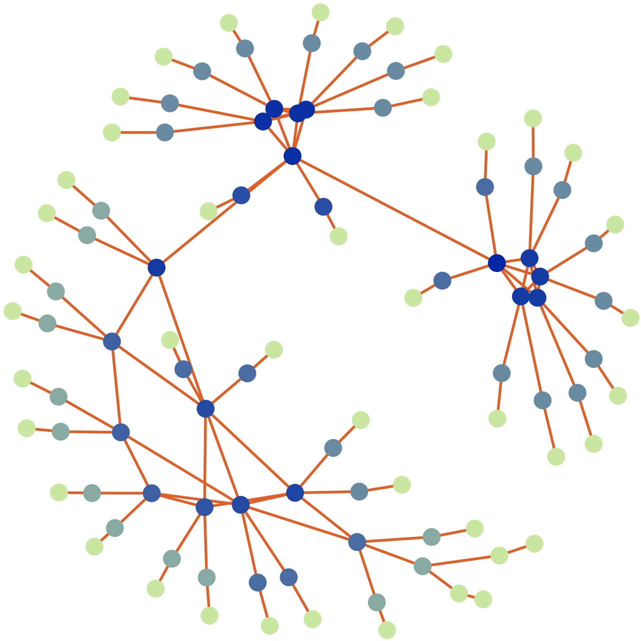
Nearest Neighbour Degree: Empirical Results



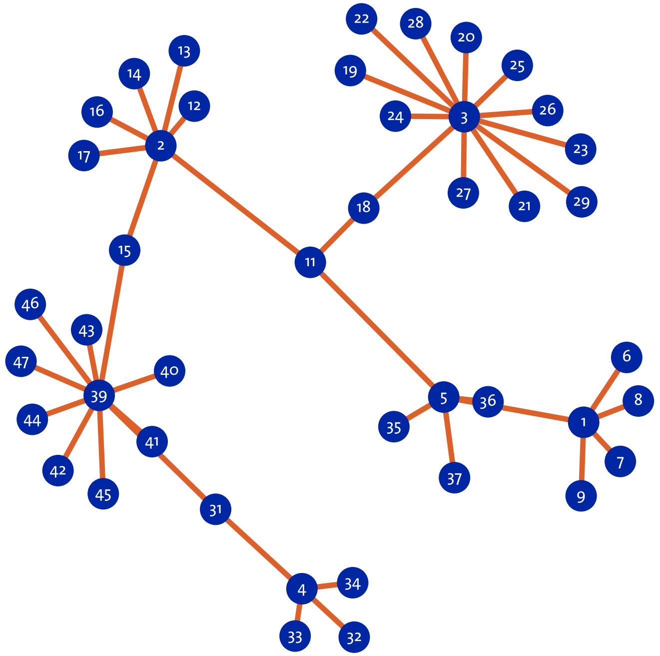
By plotting $k_{nn}(k)$ vs k we can visually recognise degree-degree correlations.

Assortativity

- Global property of the network describing the overall connectivity between nodes
- Assortativity is the Pearson correlation coefficient of degree between pairs of linked nodes



Assortativity



Dissortativity

Assortativity

We start from the definition of correlation function:

$$r = \text{Corr}(k_i, k_j) = \frac{Cov(k_i, k_j)}{Var(k_i)^2} = \frac{\langle k_i k_j \rangle - \langle k_i \rangle \langle k_j \rangle}{\sigma^2}$$

Assortativity

We start from the definition of correlation function:

$$r = \text{Corr}(k_i, k_j) = \frac{Cov(k_i, k_j)}{Var(k_i)^2} = \frac{\langle k_i k_j \rangle - \langle k_i \rangle \langle k_j \rangle}{\sigma^2}$$

where

$$\langle k_i k_j \rangle - \langle k_i \rangle \langle k_j \rangle = \sum_{(i,j) \in V} k_i k_j (P(k_i, k_j) - P(k_i)P(k_j))$$

and

$$\sigma^2 = \sum_k k^2 P(k) - \left(\sum_k k P(k) \right)^2$$

Assortativity: Empirical Results

Network	N	r
Physics coauthorship	52 909	0.363
Biology coauthorship	1 520 251	0.127
Mathematics coauthorship	253 339	0.120
Film actor collaborations	449 913	0.208
Company directors	7 673	0.276
Internet	10 697	−0.189
World-Wide Web	269 504	−0.065
Protein interactions	2 115	−0.156
Neural network	307	−0.163
Marine food web	134	−0.247
Freshwater food web	92	−0.276

Newman, M. E. J. (2002). Assortative Mixing in Networks. Physical Review Letters, 89(20), 208701.

Assortativity: Empirical Results

Network	N	r
Physics coauthorship	52 909	0.363
Biology coauthorship	1 520 251	0.127
Mathematics coauthorship	253 339	0.120
Film actor collaborations	449 913	0.208
Company directors	7 673	0.276
Internet	10 697	−0.189
World-Wide Web	269 504	−0.065
Protein interactions	2 115	−0.156
Neural network	307	−0.163
Marine food web	134	−0.247
Freshwater food web	92	−0.276

- Technological networks
  → *dissortative*
- Social networks
  → *assortative*

Newman, M. E. J. (2002). Assortative Mixing in Networks. Physical Review Letters, 89(20), 208701.

6

Network randomisation

The Typical Chain of Analysis

- Data collection
- Data analysis (e.g. measure clustering or assortativity)
- How significant are the results we find?

The Typical Chain of Analysis

- Data collection
- Data analysis (e.g. measure clustering or assortativity)
- How significant are the results we find?

If we quantify a network property, is the result statistically significant?

Later, we will also ask, what is the mechanism that originates the result?

To determine if the results are significant:

compare the results with those coming from a benchmark model

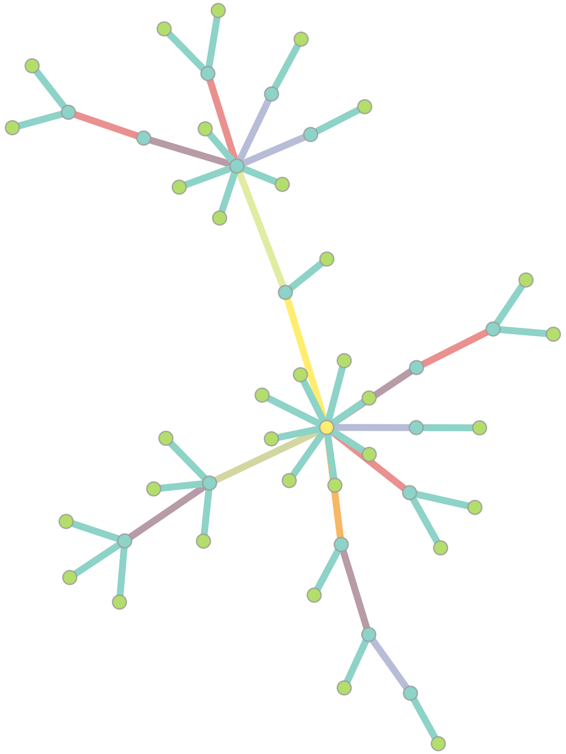
Are the results a simple effect of the degree distribution?

Degree-preserving network randomisation

If we change the network completely, but leave the nodes' degrees unaltered... do we obtain the same results for the property we are measuring?

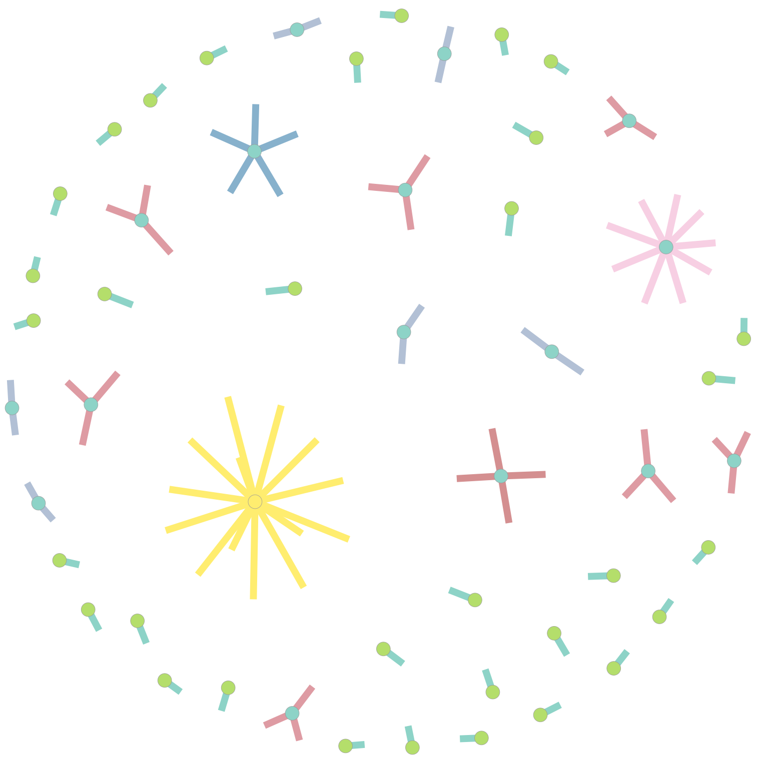
- *If the result of our observation does not change... then, the result stems from the degree distribution*
- If the result of our observation does change... then, the result comes from **higher order** network properties

Turning a network into pieces



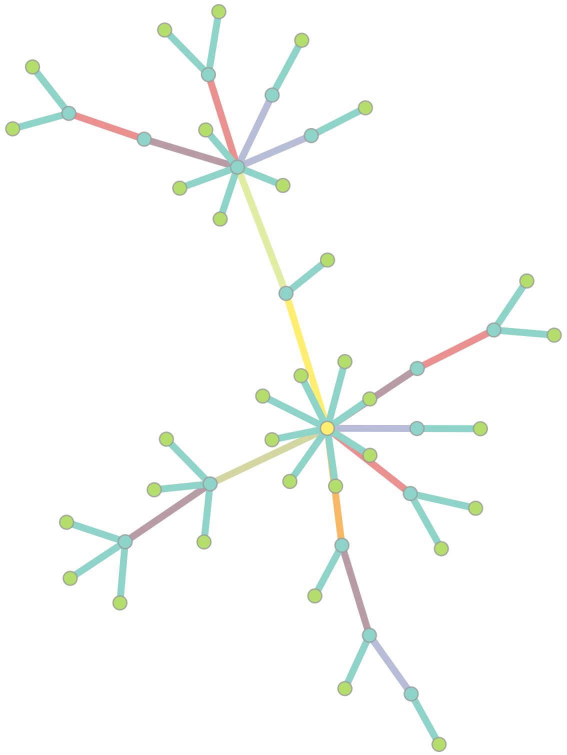
- Given an original network, we can cut all edges

Turning a network into pieces



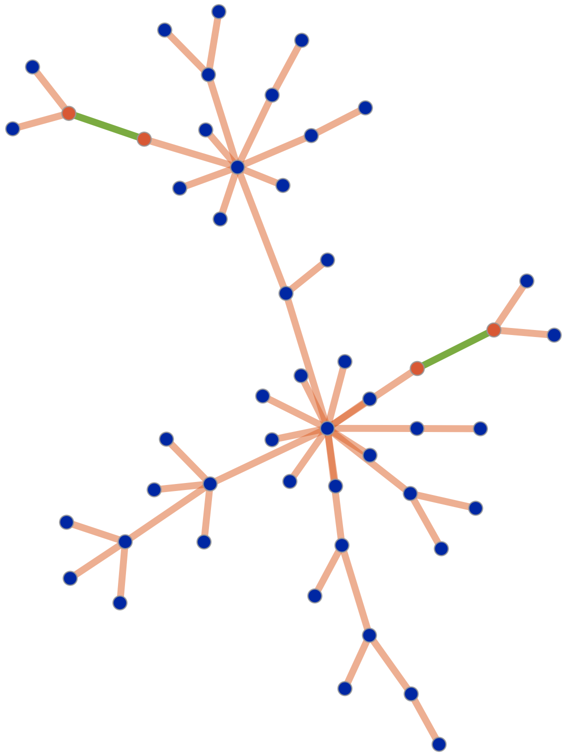
- Given an original network, we can cut all edges
- Each isolated node i has k_i *stubs*. Reconnecting them, allows us to build other networks with the same degree distribution.

A non-destructive approach



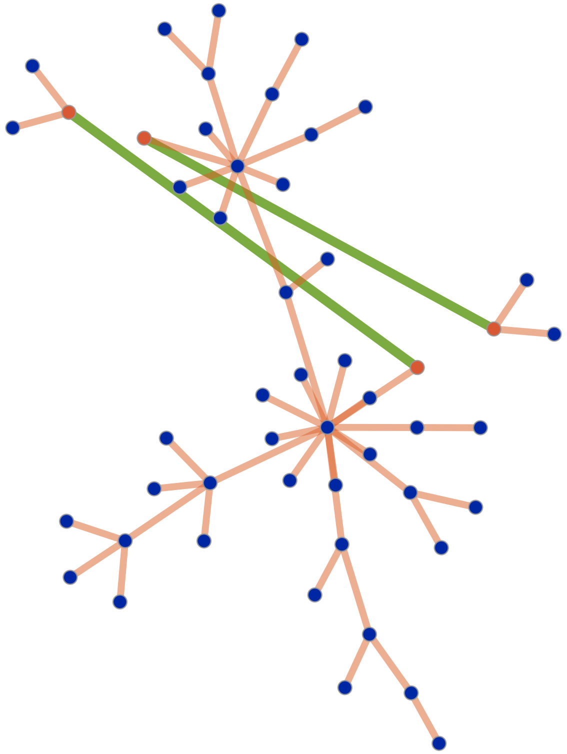
1. Start from the original network G

A non-destructive approach



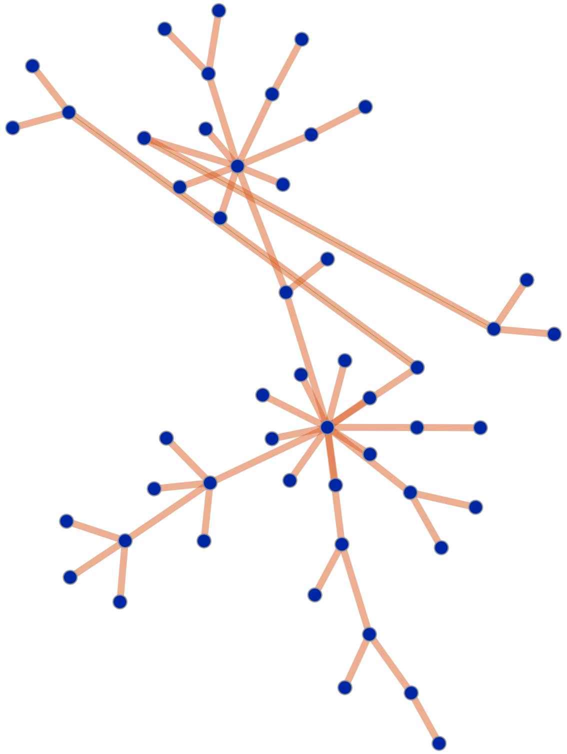
1. Start from the original network G
2. Select two edges at random

A non-destructive approach



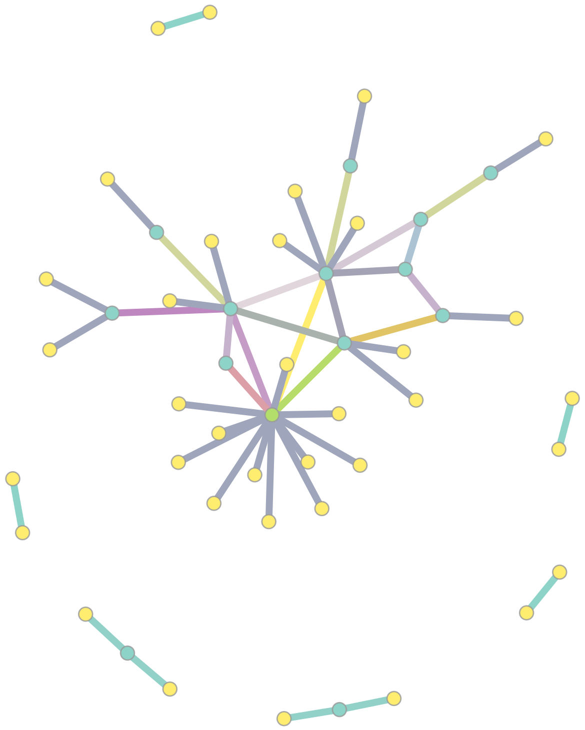
1. Start from the original network G
2. Select two edges at random
3. Exchange ends

A non-destructive approach



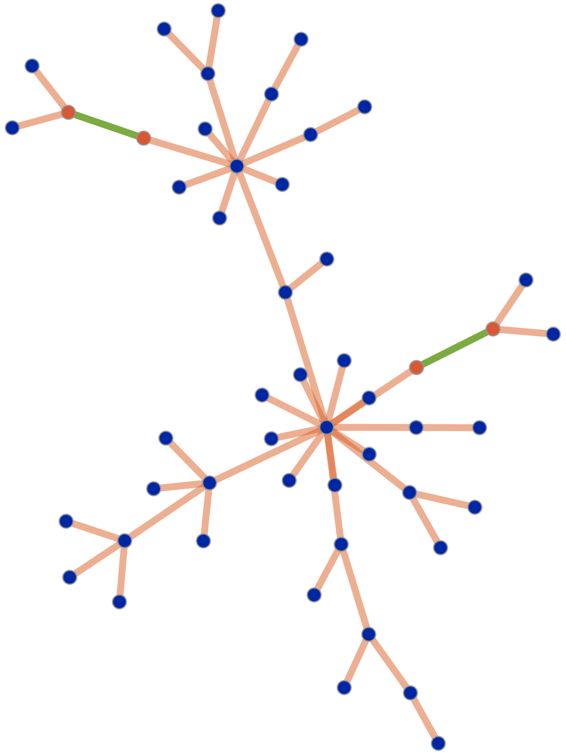
1. Start from the original network G
2. Select two edges at random
3. Exchange ends

A non-destructive approach



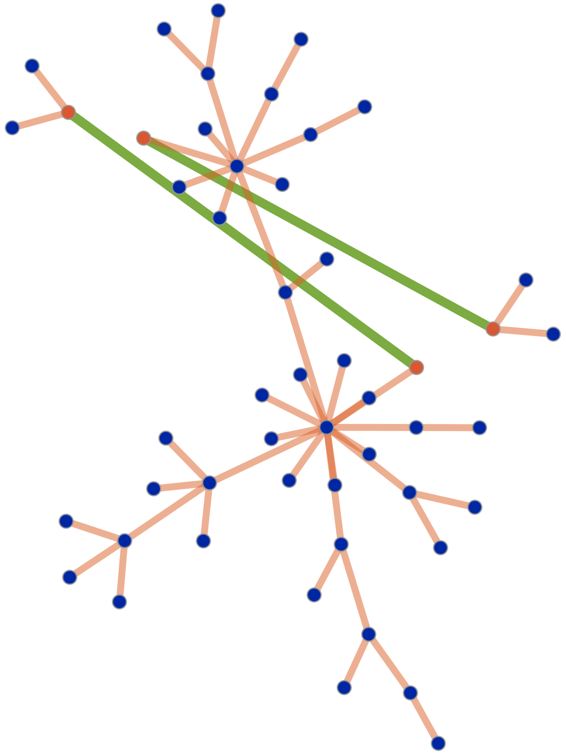
1. Start from the original network G
2. Select two edges at random
3. Exchange ends
4. Repeat steps 2-3 many times (at least, until each edge was selected - on average - once). Obtain G_{rnd}

No change in the degree distribution!



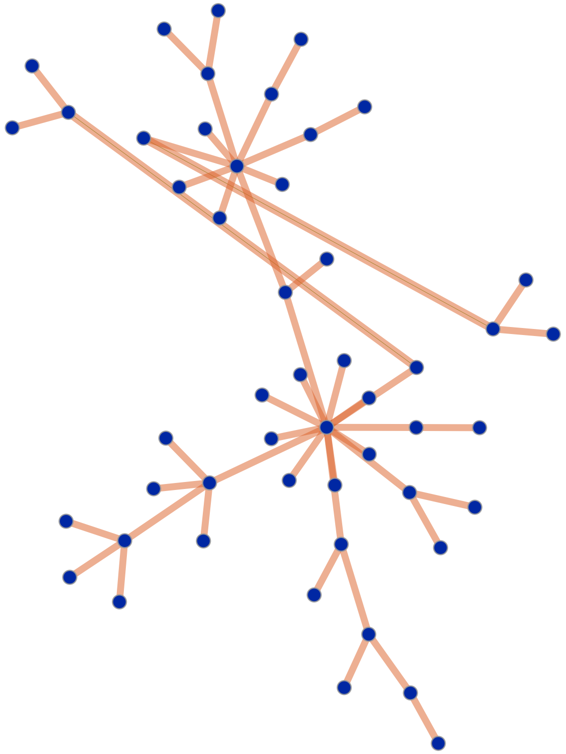
During an edge swap, the nodes affected do not change their degree!

No change in the degree distribution!



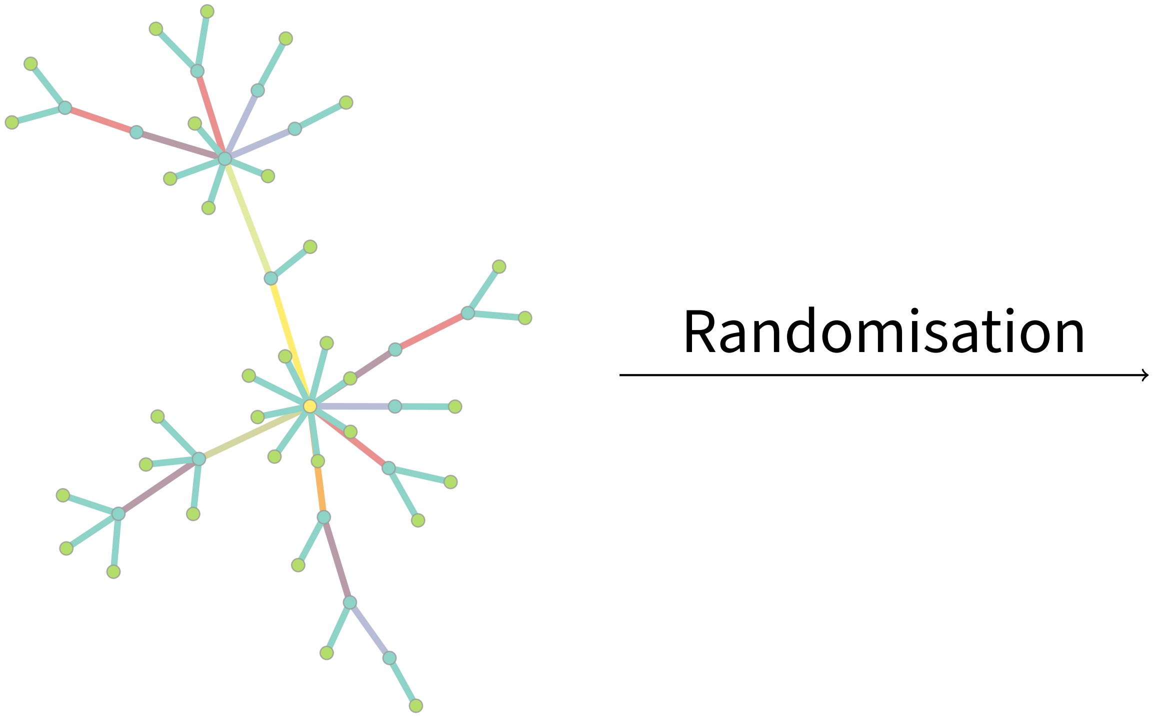
During an edge swap, the nodes affected do not change their degree!

No change in the degree distribution!



During an edge swap, the nodes affected do not change their degree!

Comparison of the results



Measure the properties in G, measures them in G_{rnd} and check if they are different

Why randomising? Because we want to understand if measurements we perform of the network are a network property or a trivial outcome of the degree distribution

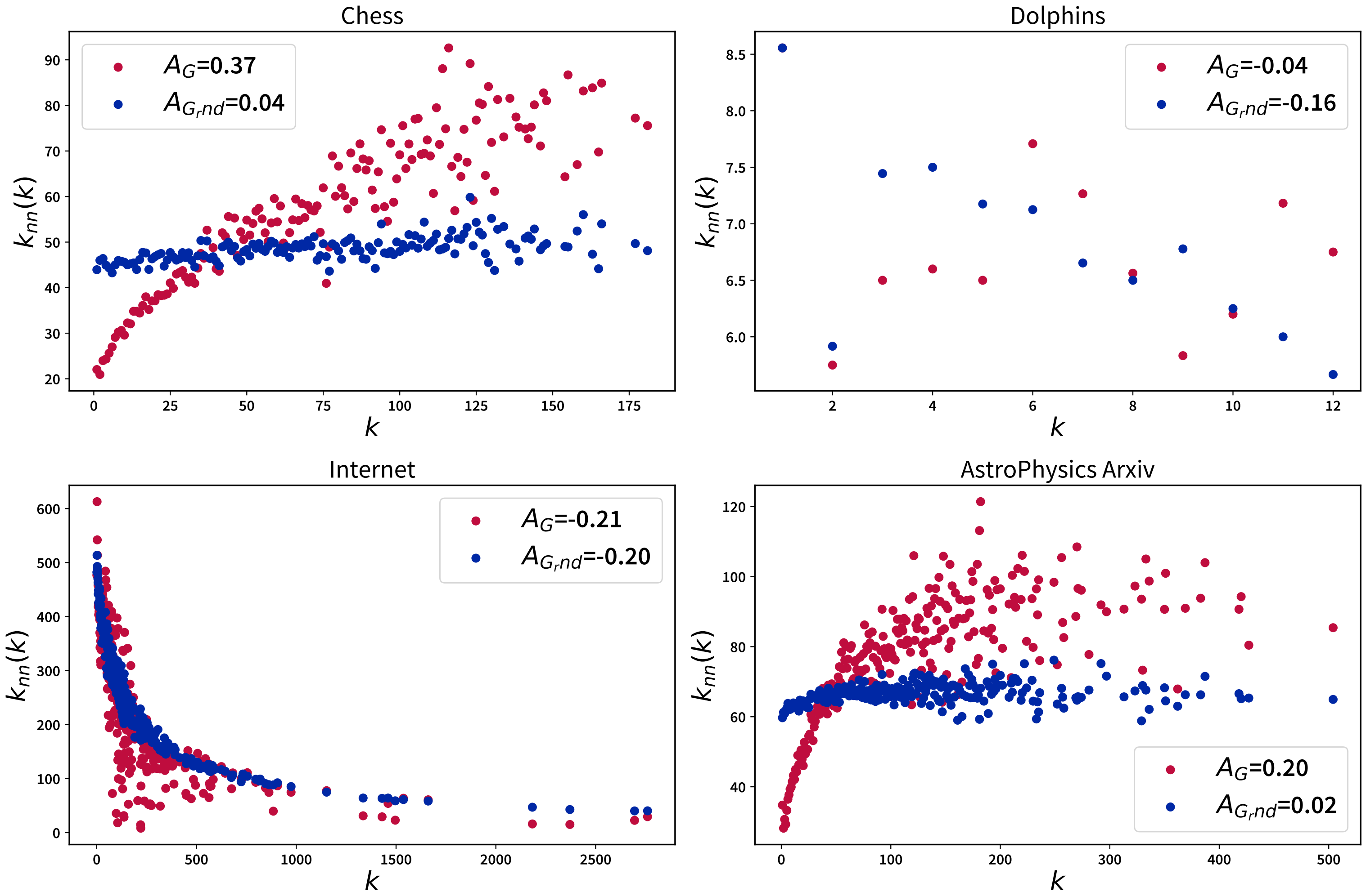
How to randomise the network?

In networkX there are multiple ways, (warning, it is in-place)

```
#!/usr/bin/python
nx.algorithms.smallworld.random_reference(G, niter=2, connectivity=False, seed=None)
```

To attain statistical significance, multiple realisations of G_{rnd} are necessary

Example: Assortativity



If, after randomization, the results remain unchanged, the degree correlation is a result of the degree distribution!

Assortativity and clustering coefficient are significant higher order characteristics of networks and can help to understand networks' structures well.

Take Home Messages

- We defined some simple graph typologies.
- Distance on a network can be derived from shortest paths.
- Homophily characterizes certain categories of real networks and can be measured with multiple methods.
- Network randomization is essential to determine the importance of observed properties.