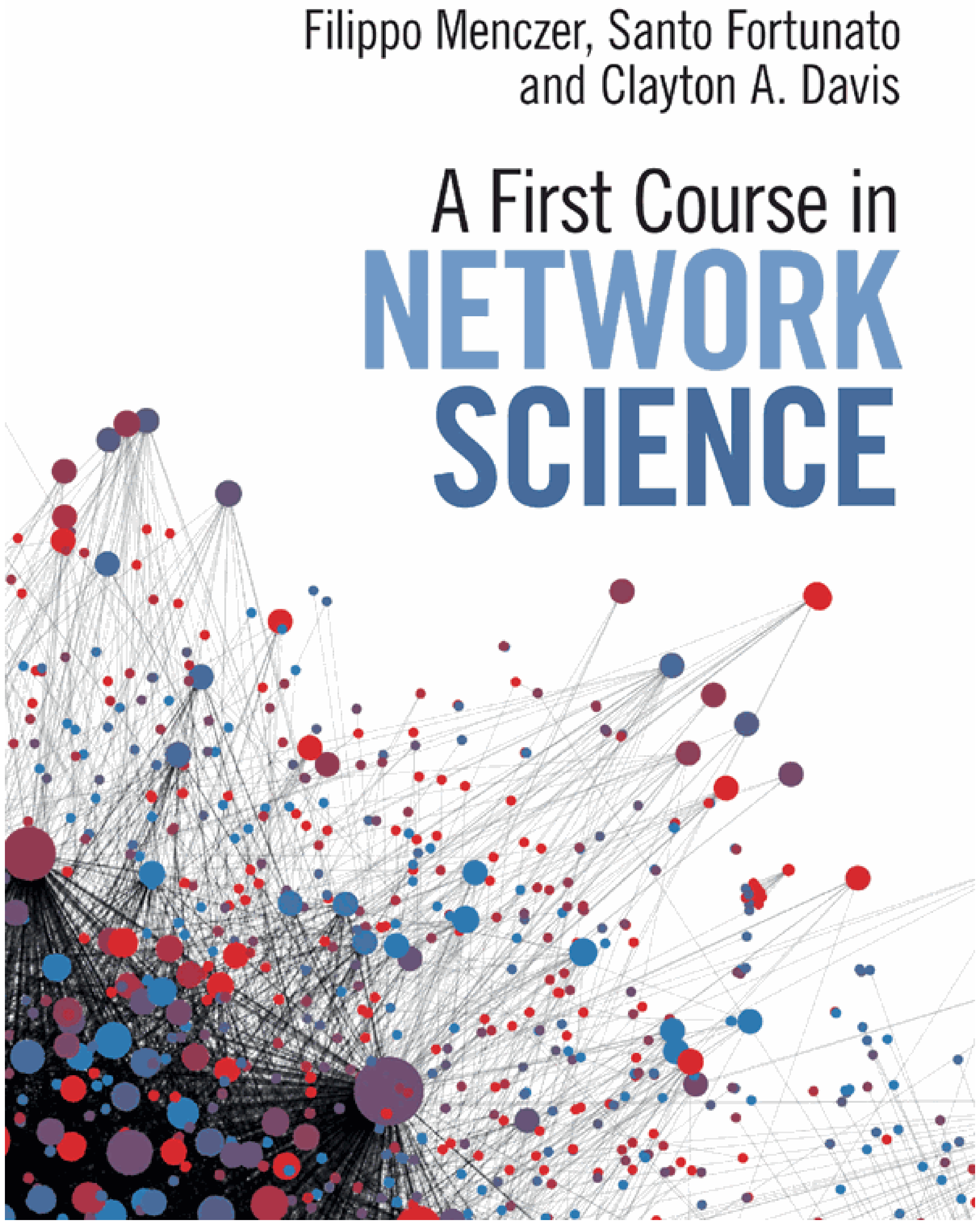
netherlands eScience center

First course in Network Science

Sections:
2.7 Six Degrees of Separation
2.8 Friend of a Friend

Sarah Alidoost
Analytics SIG, 23/11/2020



- similarity between neighbors,
- short paths connecting nodes,
- triangles.

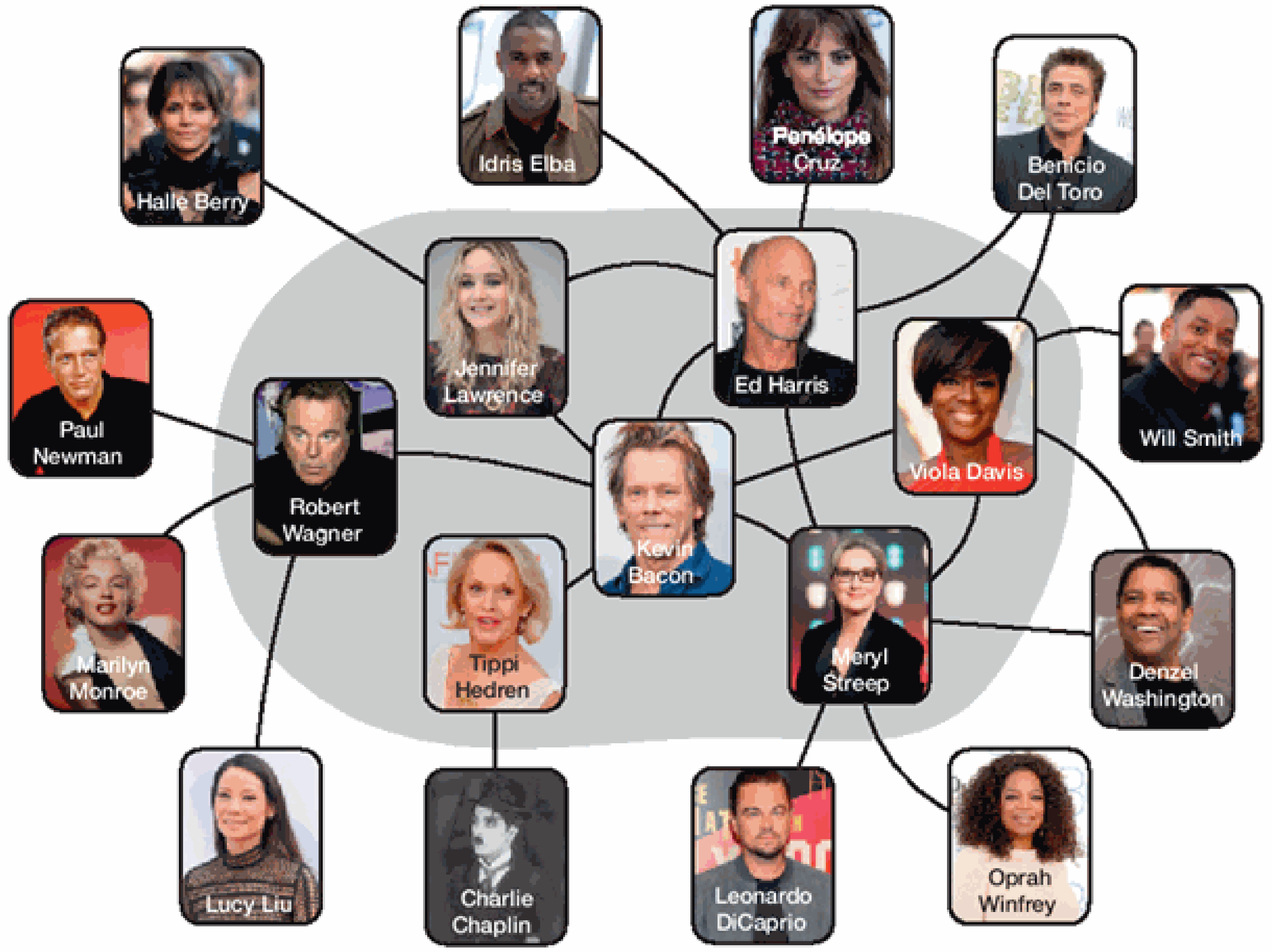


Illustration of the *Six Degrees of Kevin Bacon* game. A few of the nodes connected to Kevin Bacon in the co-star network are shown in the shaded area, along with links among them. A small sample of the nodes at distance $\ell = 2$ is also included. Photo credit: Getty Images.

2.7 Six Degrees of Separation

- Any two people in the world are connected by a short chain of acquaintances.
- Social networks have a diameter and an even shorter average path length.

Guglielmo Marconi

1900

Frigyes Karinthy

1920

Stanley Milgram

1960

Source: wikipedia

2.7 Six Degrees of Separation

We watch: https://www.youtube.com/watch?v=a99ry70CnRs

When can we call a path "short"? Compared to what?

The average path length $< \ell >$ is short when it grows very slowly with the size of the network (N).

$$< \ell > \sim \log N$$

Source: https://www.pexels.com/

What does it mean?

- The network could have tens millions of nodes, and yet its average path length would be in the single digit.

- Also, the network could multiply many times in size while the average path length would only get a few steps longer.

$$\langle \ell \rangle \sim \log N$$

Searching for path is something we do routinely in all kinds of networks.

Some games on finding network paths:

- Wikiracing,
- The Wiki Game

This tells us that Wikipedia has short paths.

Table 2.1 The average path length of various networks.

Network	Nodes (N)	Links (L)	Average path length ($\langle \ell \rangle$)	Clustering coefficient (C)
Facebook Northwestern Univ.	10,567	488,337	2.7	0.24
IMDB movies and stars	563,443	921,160	12.1	0
IMDB co-stars	252,999	1,015,187	6.8	0.67
Twitter US politics	18,470	48,365	5.6	0.03
Enron email	87,273	321,918	3.6	0.12
Wikipedia math	15,220	194,103	3.9	0.31
Internet routers	190,914	607,610	7.0	0.16
US air transportation	546	2,781	3.2	0.49
World air transportation	3,179	18,617	4.0	0.49
Yeast protein interactions	1,870	2,277	6.8	0.07
C. elegans brain	297	2,345	4.0	0.29
Everglades ecological food web	69	916	2.2	0.55

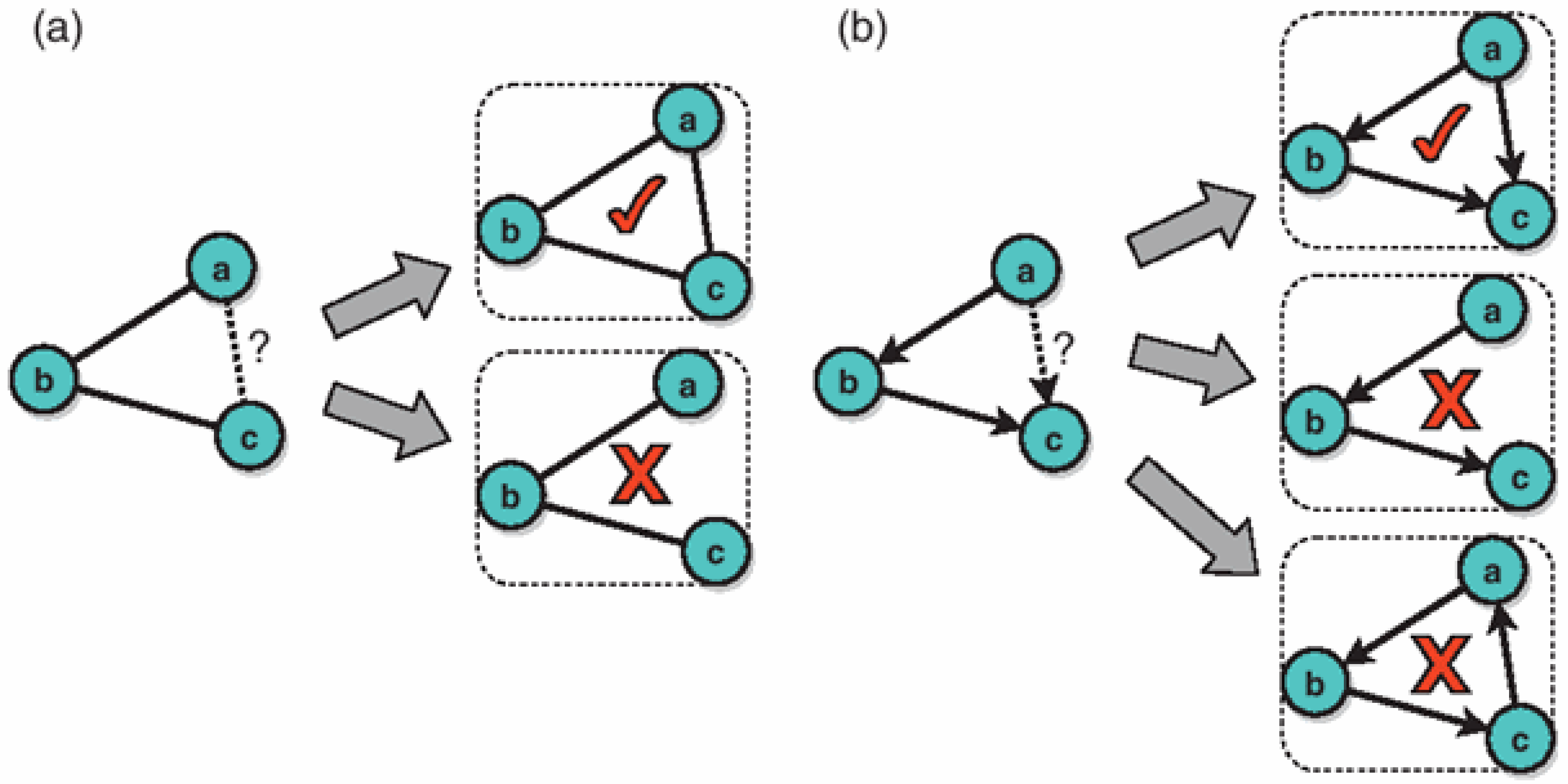
In all of these examples, the average path length is only a few steps.

Source: https://unsplash.com/

2.8 Friend of a Friend

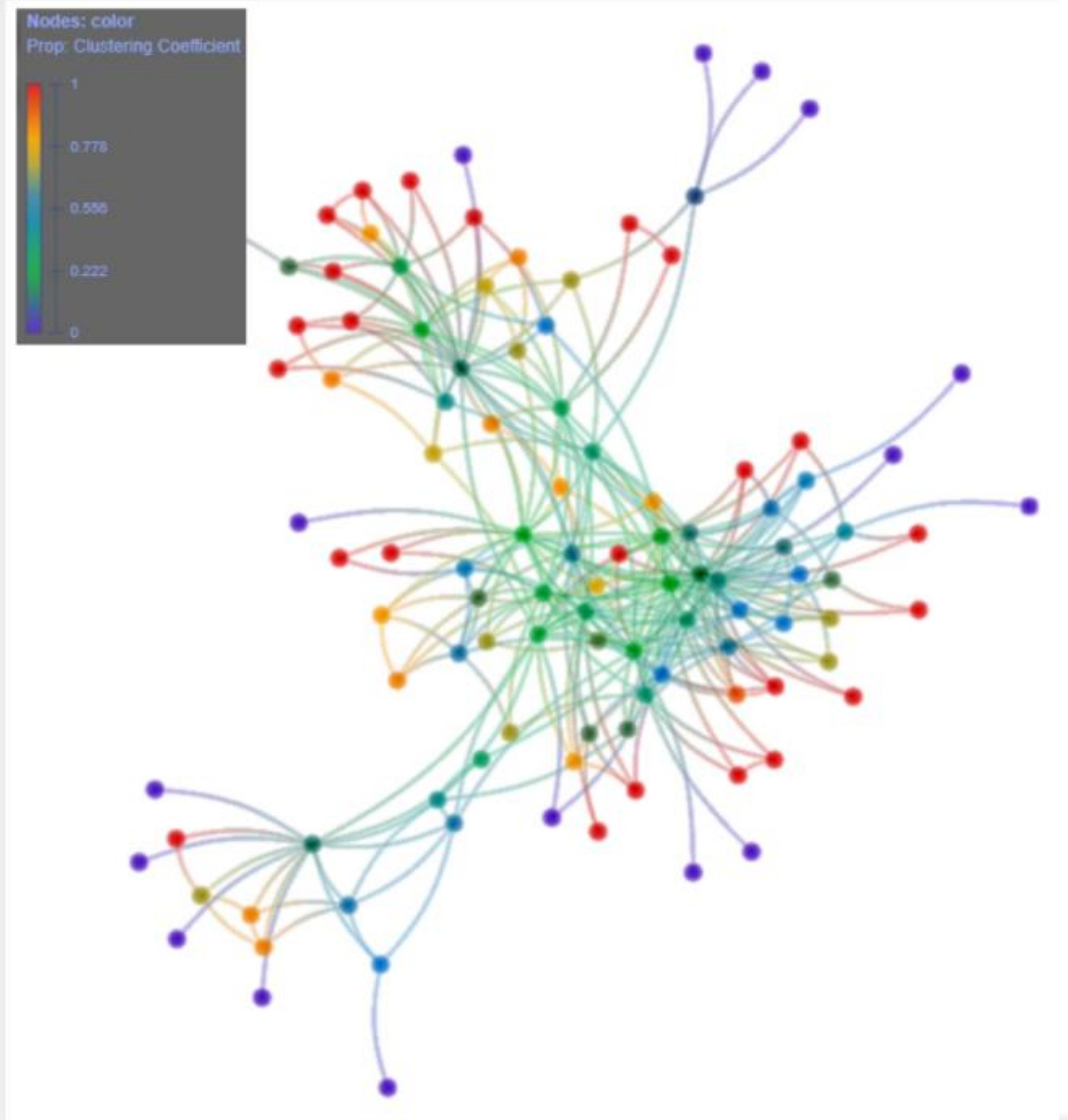
- There is a good chance that a friend of my friend is also my friend.
- This translates into the presence of many triangles in the network.

A triangle is a triad (set of three nodes) where each pair of nodes is connected.



Triads and triangles. (a) In an undirected network, node **b** has neighbors **a** and **c**. They may or may not form a triangle, depending on whether or not **a** and **c** are connected to each other. (b) In a directed network, node **a** links to **b** and node **b** links to **c**. A shortcut link from **a** to **c** would form a directed triangle.

The connectivity among the neighbors of the nodes is an important feature of the local structure of the network. It captures how tightly clustered the nodes are.



Source: M.E.J. Newman, Networks, an introduction, Oxford University Press, 2010, ISBN: 978-0-19-920665-0, sec. 7.9

Source: https://www.pexels.com/

Clustering coefficient

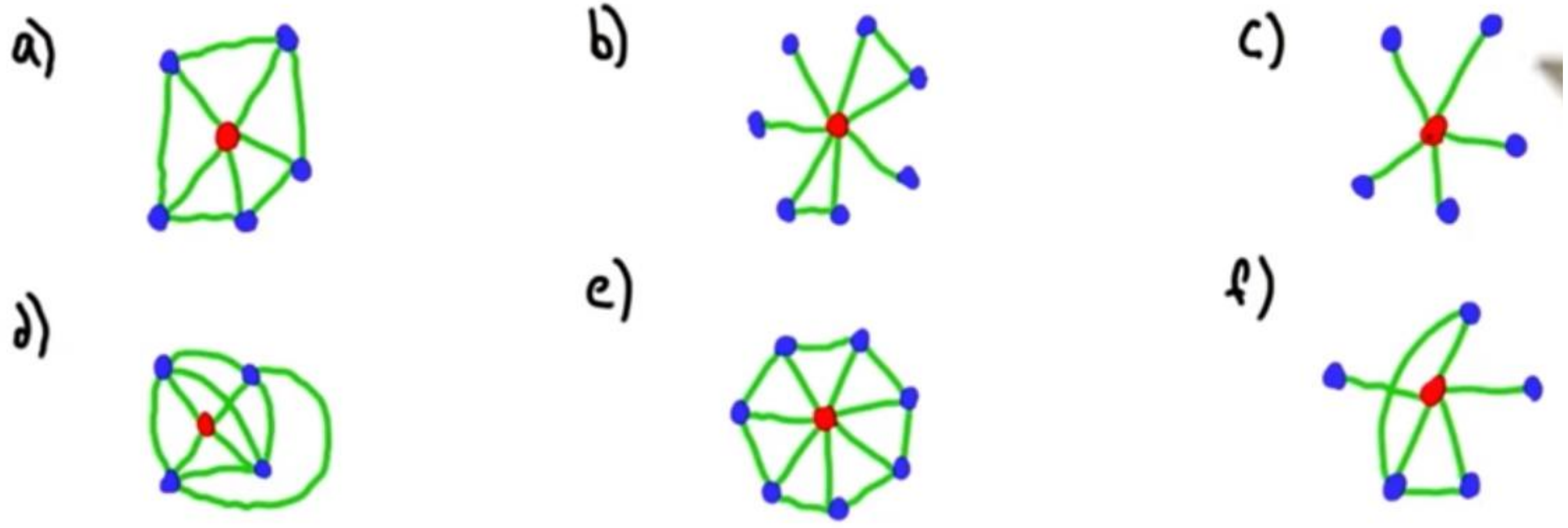
- For a node: the fraction of pairs of neighbors that are connected.

- For the entire network: the average of the clustering coefficients of its nodes.

the number of triangles that include the node i

$$C(i) = \frac{\tau(i)}{\tau_{max}(i)} = \frac{\tau(i)}{\binom{k_i}{2}} = \frac{2\tau(i)}{k_i(k_i-1)},$$

the max. number of triangles for i = the number of pairs formed by its ki neighbours.

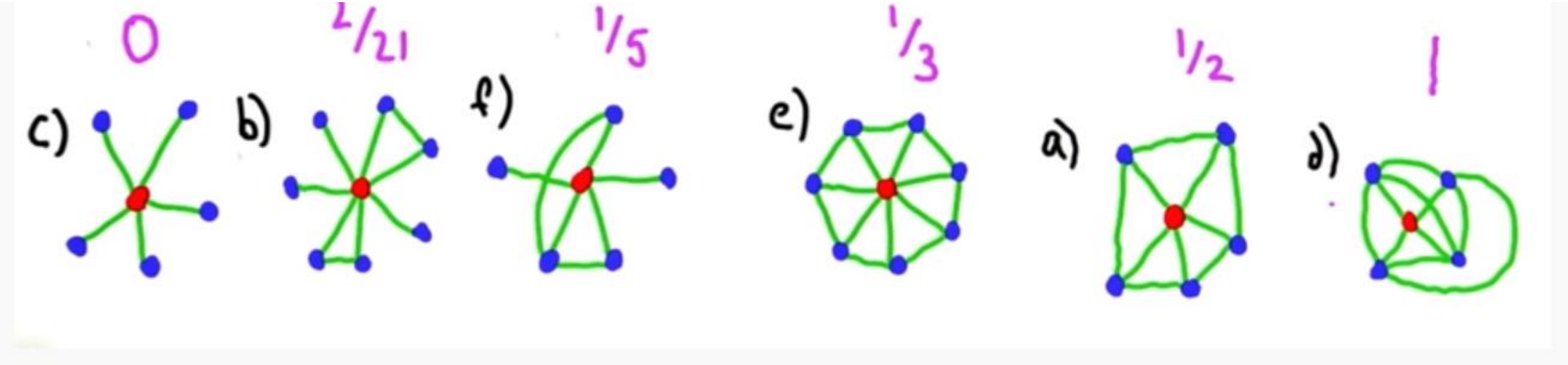


the number of triangles that include the node i

$$C(i) = \frac{\tau(i)}{\tau_{max}(i)} = \frac{\tau(i)}{\binom{k_i}{2}} = \frac{2\tau(i)}{k_i(k_i-1)},$$

the max. number of triangles for i = the number of pairs formed by its ki neighbours.

Exercise: Clustering coefficient?



Answer: Clustering coefficient

source: https://www.youtube.com/watch?v=bcxXlP5pe0U


Source: https://www.pexels.com/

In this book we only deal with the clustering coefficient in undirected networks.

For directed networks we can simply ignore the direction of links and treat them as if they were undirected when calculating the clustering coefficient.

- In social networks, we meet people through shared contacts, thus closing triangles. This mechanism called triadic closure.

- Online social networks make suggestions based on triadic closure. Facebook recommends "people you may know"; these recommendations result in high clustering.

Source: https://www.pexels.com/

Table 2.1 The clustering coefficient of various networks.

Network	Nodes (N)	Links (L)	Average path length ($\langle \ell \rangle$)	Clustering coefficient (C)
Facebook Northwestern Univ.	10,567	488,337	2.7	0.24
IMDB movies and stars	563,443	921,160	12.1	0
IMDB co-stars	252,999	1,015,187	6.8	0.67
Twitter US politics	18,470	48,365	5.6	0.03
Enron email	87,273	321,918	3.6	0.12
Wikipedia math	15,220	194,103	3.9	0.31
Internet routers	190,914	607,610	7.0	0.16
US air transportation	546	2,781	3.2	0.49
World air transportation	3,179	18,617	4.0	0.49
Yeast protein interactions	1,870	2,277	6.8	0.07
C. elegans brain	297	2,345	4.0	0.29
Everglades ecological food web	69	916	2.2	0.55

Exercise: Why C of the network "IMDB movies and starts" is 0?

Table 2.1 The clustering coefficient of various networks.

Network	Nodes (N)	Links (L)	Average path length ($\langle \ell \rangle$)	Clustering coefficient (C)
Facebook Northwestern Univ.	10,567	488,337	2.7	0.24
IMDB movies and stars	563,443	921,160	12.1	0
IMDB co-stars	252,999	1,015,187	6.8	0.67
Twitter US politics	18,470	48,365	5.6	0.03
Enron email	87,273	321,918	3.6	0.12
Wikipedia math	15,220	194,103	3.9	0.31
Internet routers	190,914	607,610	7.0	0.16
US air transportation	546	2,781	3.2	0.49
World air transportation	3,179	18,617	4.0	0.49
Yeast protein interactions	1,870	2,277	6.8	0.07
C. elegans brain	297	2,345	4.0	0.29
Everglades ecological food web	69	916	2.2	0.55

Answer: The network is bipartie; no triangles. Links between pairs of movie or stars are not present in this network.

```
nx.triangles(G)              # dict node -> no. triangles
nx.clustering(G, node)       # clustering coefficient of node
nx.clustering(G)             # dict node -> clustering coeff.
nx.average_clustering(G)     # network's clustering coeff.
```

NetworkX functions to count triangles and calculate coefficients.

Summary:

- Most of networks have a very short paths on average. This is known as the small world property.

- Social networks have high clustreing due to friend-of-a-friend triangles.

Let’s stay in touch