Friendly Graph Theory: Degree correlations

Introduction to Network Science

Dr. Michele Starnini — https://github.com/chatox/networks-science-course

Universitat Pompeu Fabra Barcelona

Contents:

- Assortativity: Degree correlations
- Friendship Paradox

all related to friendship in social networks!

Degree correlations

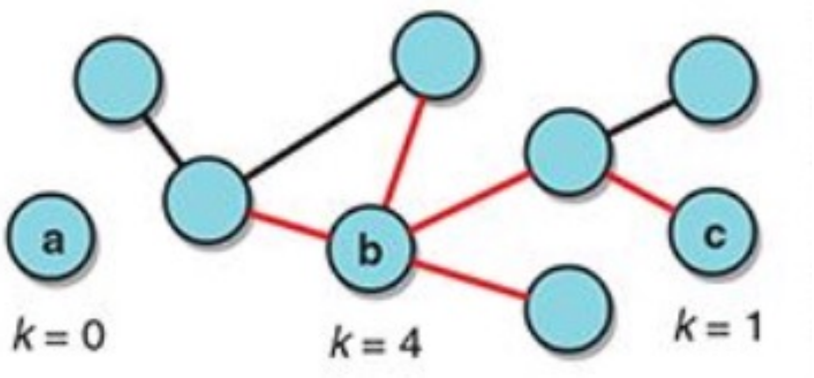
Who is a friend? [Assortativity]

- Degree is the main feature of nodes
- In social networks, degree correlations can determine connections: **assortativity**
- Example: very famous people (with millions of followers) follow each other

Degrees

Node i has degree k_i:
number of links incident on this node

High-degree nodes are called **hubs, k >>> 1**



Prob. that a randomly chosen node has degree k (**k-node**)

$$p(k) \quad \propto \quad N_k$$

\# of k-nodes

Normalization

$$\sum_k N_k = N$$

Prob. of sampling a k-node

$$p(k) = \frac{N_k}{N}$$

Average degree

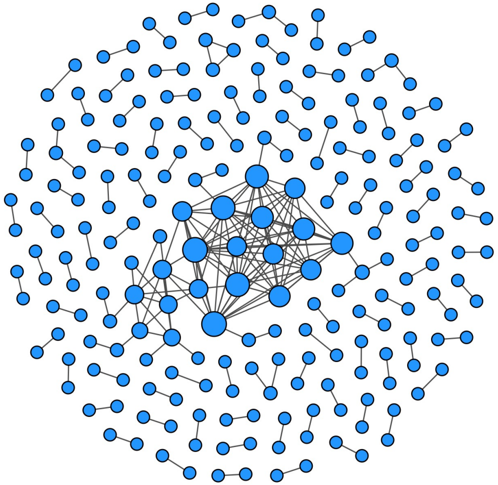
$$\langle k \rangle = \frac{\sum_k k p(k)}{N}$$

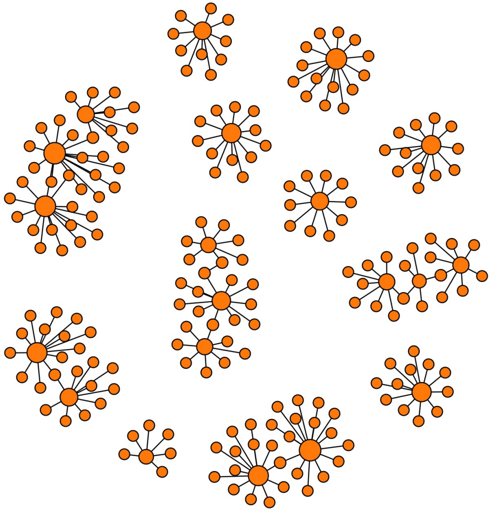
Degree assortativity

A.k.a. **degree correlation:**

- Assortative networks have a **core-periphery** structure with hubs in the core
  (Ex: social networks)

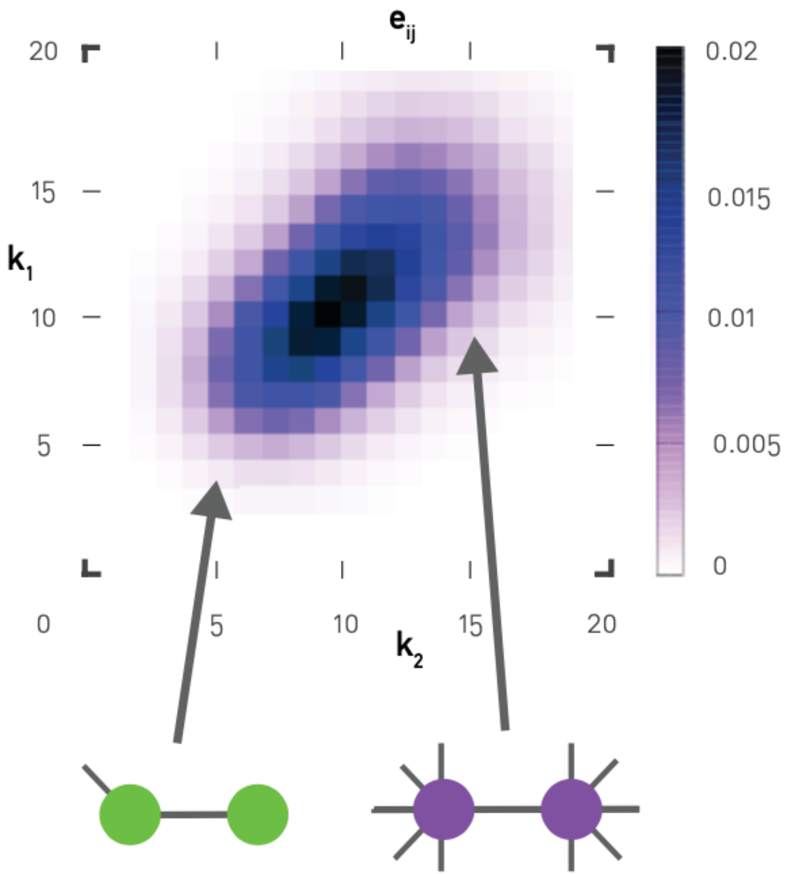
- Disassortative networks have **hub-and-spoke** (or **star**) structure
  (Ex: Web, Internet, food webs, bio networks)





Assortative networks

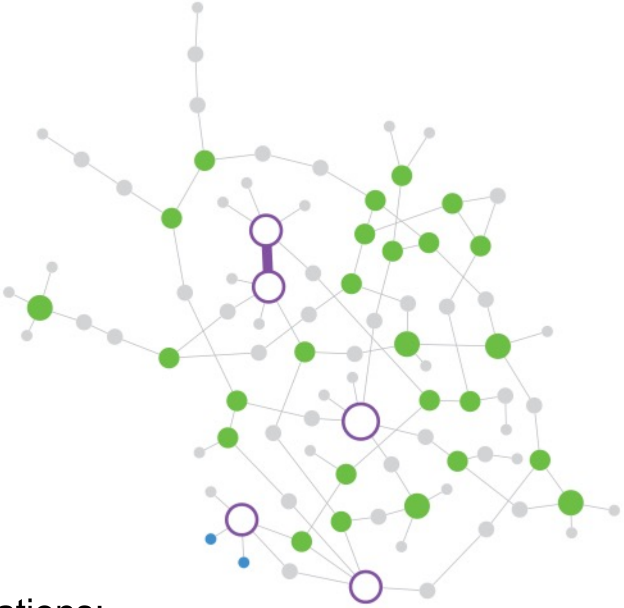
$E_{k,k'}$ = #links between k-nodes & k'-nodes

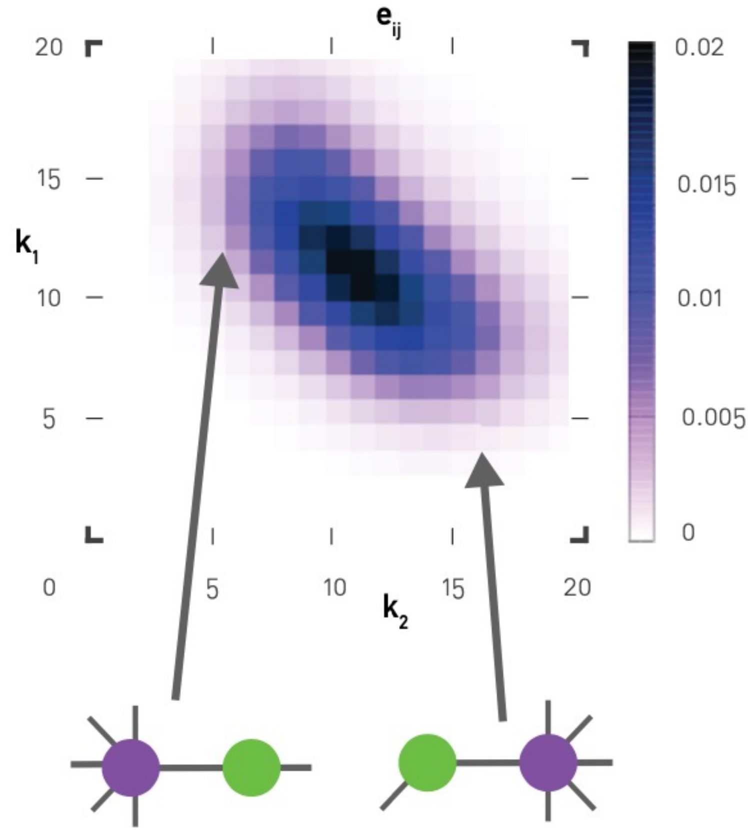


Positive degree correlations:
small-degree nodes to small-degree nodes,
hubs to hubs

Dis-assortative networks

$E_{k,k'}$ = #links between k-nodes & k'-nodes





Negative degree correlations:
small-degree nodes to hubs

Degree correlations

$E_{k,k'}$ = #links between k-nodes & k'-nodes

$$\sum_{k,k'} E_{k,k'} = N\langle k \rangle = 2E$$

Sum over all nodes twice

Joint prob. that a random link is connected to a k-node & a k'-node

$$p(k',k) \propto E_{k,k'}$$

Normalization

$$\sum_{k,k'} p(k',k) = 1$$

Joint prob. that a random link connects a k-node & a k'-node

$$p(k,k') = \frac{E_{k,k'}}{\sum_{k,k'} E_{k,k'}} = \frac{E_{k,k'}}{N\langle k \rangle}$$

Degree correlations

$E_{k,k'}$ = #links between k-nodes & k'-nodes

$$\sum_{k,k'} E_{k,k'} = N\langle k \rangle = 2E$$

Sum over all nodes twice

Prob. that a random link is connected to a k-node

Total # links from k-nodes

$$q_k \propto \sum_{k'} E_{k,k'} = kN_k$$

Normalization

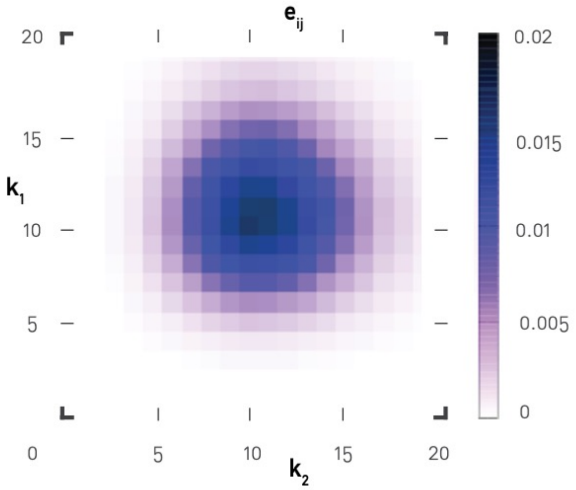
$$\sum_k q_k = 1$$

Prob. that a random link is connected to a k-node

$$q_k = \sum_{k'} p(k', k) = \frac{\sum_{k'} E_{k',k}}{N\langle k \rangle} = \frac{kp(k)}{\langle k \rangle}$$

No degree correlations

No correlations:
degree of a node is **independent** from the degree of others



Joint prob. that a random link is connected to a k-node & a k'-node,
If no degree corralations

$$p_0(k', k) \;\propto\; E^0_{k,k'} = kN_k \times k'N_{k'}$$

Total # links from k-nodes

Total # links from k'-nodes

Joint prob. that a random link connects a k-node & a k'-node in uncorrelated networks

$$p_0(k', k) = q_k \times q'_k = \frac{kp(k)k'p(k')}{\langle k \rangle^2}$$

$$q_k = \frac{kp(k)}{\langle k \rangle}$$

Prob. that a random link is connected to a k-node

Measuring degree correlations

Conditional prob. that a random link, **already connected to a k-node,** is also connected to a k'-node $=$ $\dfrac{\text{Joint prob. that a random link is connected to a k-node \& a k'-node}}{\text{Prob. that a random link is connected to a k-node}}$

$$p(k'|k) = \frac{p(k',k)}{\sum_{k'} p(k',k)}$$

Remember:

$$p(k',k) = \frac{E_{k,k'}}{N\langle k\rangle}$$

$$\sum_{k'} p(k',k) = q_k = \frac{kp(k)}{N\langle k\rangle}$$

$\Rightarrow$

Conditional prob. that a random link is connected to a k'-node, given that it is connected to a k-node

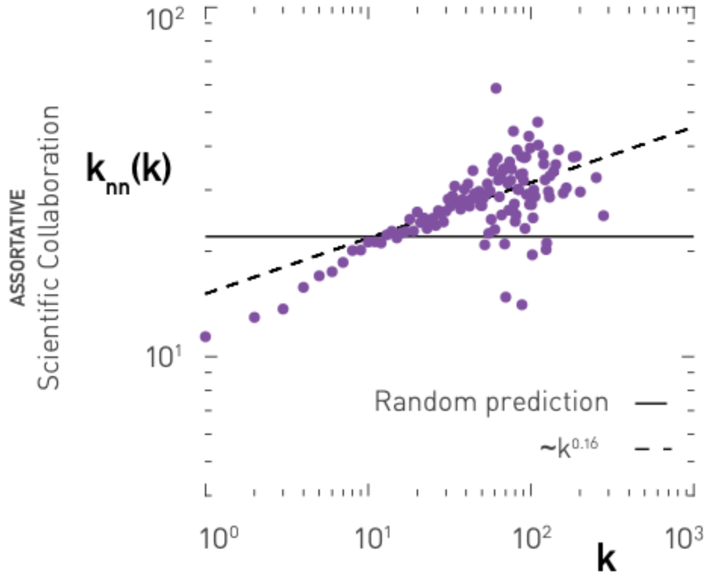
$$p(k'|k) = \frac{E_{k,k'}}{kp(k)}$$

Measuring degree correlations

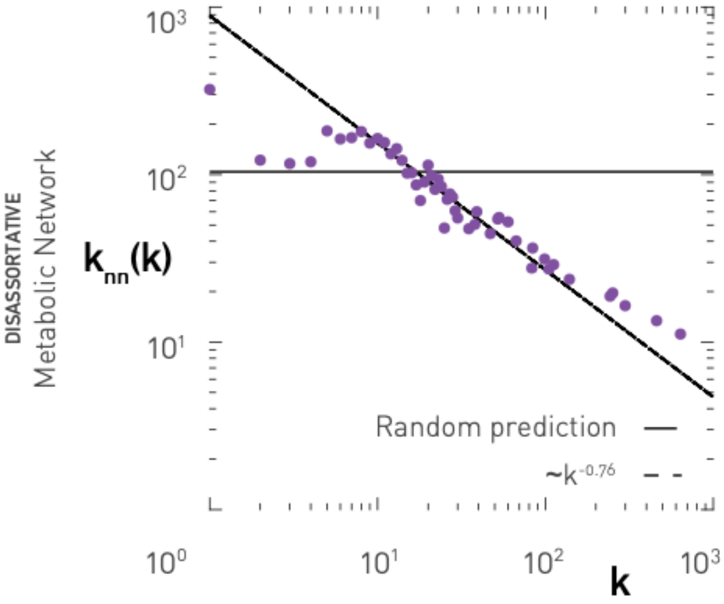
Average degree of the nearest neighbors (NN) of nodes with degree k:

$$k_{NN}(k) \equiv \sum_{k'} k' p(k'|k)$$

Assortative networks: Increasing $k_{NN}(k)$



Disassortative networks: Decraesing $k_{NN}(k)$



Exercise

Find $k_{NN}(k)$ if there are no degree correlations!

$$k_{NN}(k) \equiv \sum_{k'} k' p(k'|k)$$

$$p(k'|k) = \frac{p(k',k)}{\sum_{k'} p(k',k)}$$

$$p_0(k',k) = q_k \times q'_k = \frac{kp(k)k'p(k')}{\langle k \rangle^2}$$

Pin board: https://upfbarcelona.padlet.org/chato/v0apheshv2l4hbot

No degree correlations

$$k_{NN}(k) \equiv \sum_{k'} k' p(k'|k)$$

$$p(k'|k) = \frac{p(k',k)}{\sum_{k'} p(k',k)}$$

$$p_0(k',k) = q_k \times q_k' = \frac{kp(k)k'p(k')}{\langle k \rangle^2}$$

$$p_0(k'|k) = \frac{q_k q_{k'}}{q_k} = q_{k'}$$

Definition of no-correlations:
$p_0(k'|k)$ is **independent** of k!!

$$k_{NN}^0(k) = \sum_{k'} k' p_0(k'|k) = \sum_{k'} k' q_{k'} = \frac{\sum_{k'} k'k'p(k')}{N\langle k \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle}$$

Friendship paradox

Average degree (number of friends)
of a randomly chosen node (you): **<k>**

<

Average degree (number of friends)
of a k-node (your friend with k friends): **k_{NN}(k)**

If no degree correlations:
k_{NN}(k) = <k^2> / <k>

$$\langle k \rangle = \sum_k k p(k)$$

average of
degree distribution

$$\langle k^2 \rangle = \sum_k k^2 p(k)$$

$$\sigma^2 = \langle k^2 \rangle > \langle k \rangle^2 > 0$$

Variance of
degree distribution

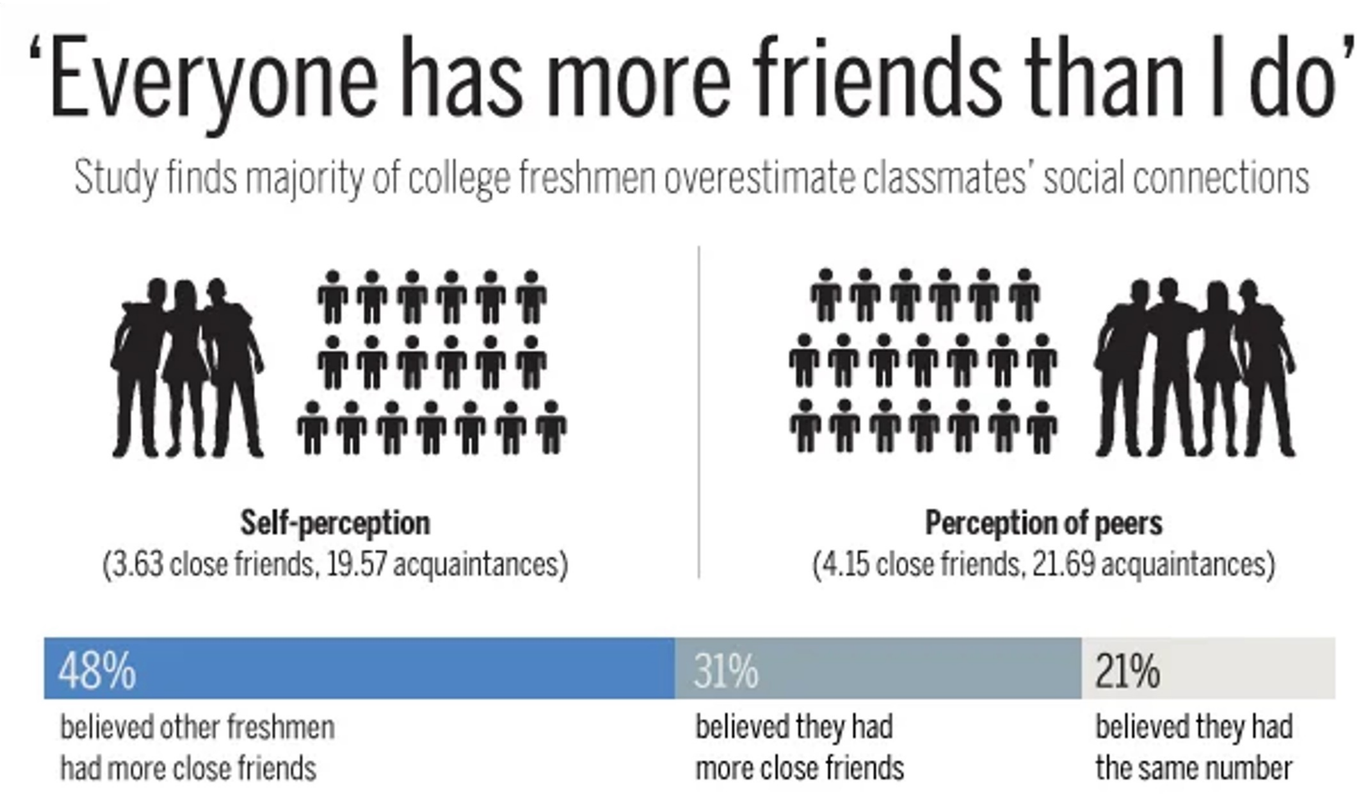
$$\langle k^2 \rangle > \langle k \rangle^2$$

Your friends' average
number of friends

Your average
number of friends

$$\frac{\langle k^2 \rangle}{\langle k \rangle} > \langle k \rangle$$

Your friends have more friends than you have!



Source: [Whillans et al. 2017](). Image credit: [The Harvard Gazette]().

Another way to look at the friendship paradox:

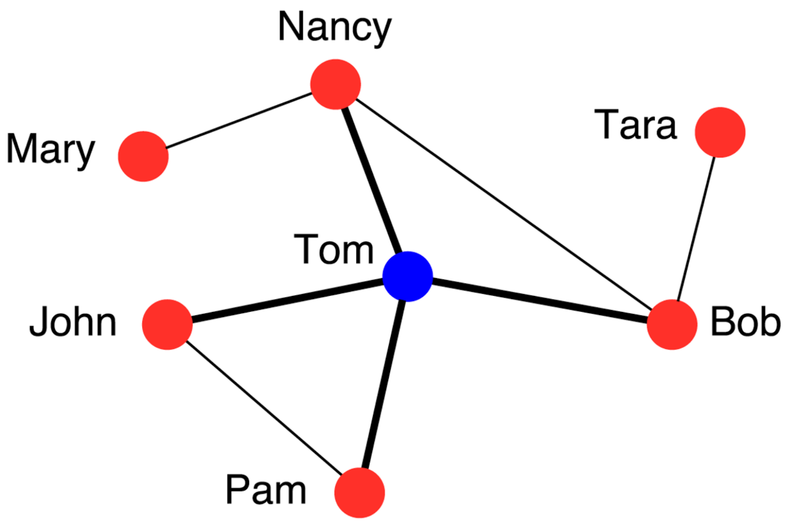
The consequences of
different sampling methods

Exercise

Consequences of sampling methods

What is the probability of selecting Tom **if we select a random node**?

What is the probability of selecting Tom **if we select a random edge and then randomly one of the two nodes attached to it?**



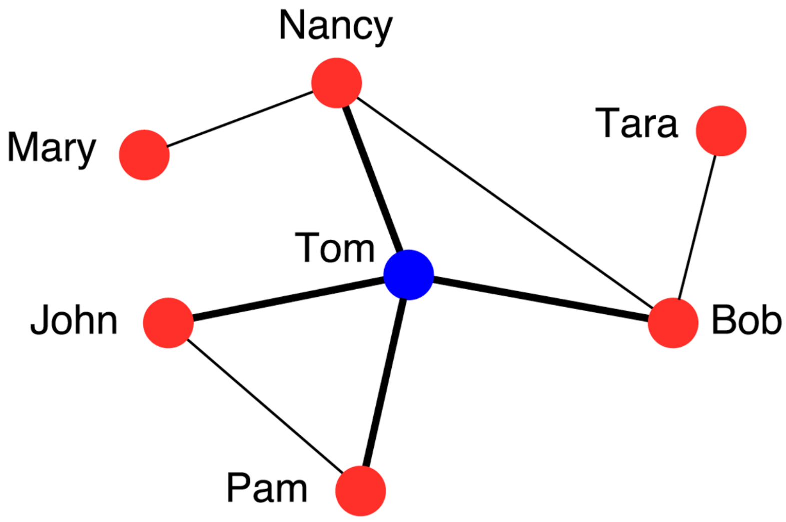
Pin board: https://upfbarcelona.padlet.org/chato/ocpl5n14i8hrkr4i

Sampling a random node

vs

Sampling a random neighbor
of a random node

Average degree of friends



Average degree

$(1 + 3 + 3 + 1 + 4 + 2 + 2)/7 = 16/7 \simeq$ **2.29**

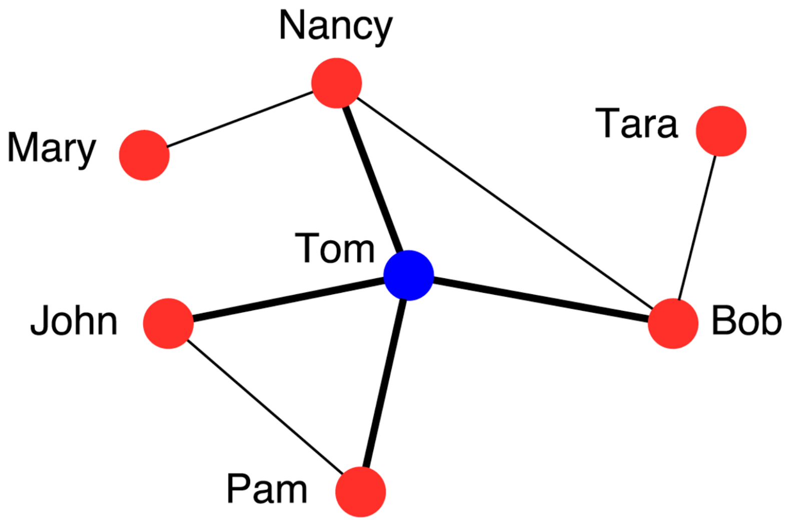
Average degree of friends of ...

… Mary: 3

… Nancy: (1+4+3)/3 = 8/3

…

Average degree of friends



Average degree

$(1 + 3 + 3 + 1 + 4 + 2 + 2)/7 = 16/7 \simeq$ **2.29**

Average degree of friends of ...

… Mary: 3

… Nancy: (1+4+3)/3 = 8/3

… Tara: 3

… Bob: (1+3+4)/3 = 8/3

… Tom: (3+3+2+2)/4 = 10/4

… John: (4+2)/2 = 3

… Pam: (4+2)/2 = 3

Average degree of friends $\simeq$ **2.83** (> 2.29)

The friendship paradox

Take a random person x; what is the expected degree of this person?

It is $<k>$

Take a random person x, now pick one of x’s neighbors, let’s say y; what is the expected degree of y?

It is not $<k>$, it is $<k^2>/<k>$

The friendship paradox can be useful

Examples:

- **As a marketing strategy:** if *u* invites a friend *v* to buy/use a product, it is likely that *v* has many friends, and hence it is relevant for marketing that v buys/use the product
- **As a vaccination strategy:** instead of offering a vaccine to random people, ask them to name a friend, offer the vaccine to those people, who will have larger degree

Imagine you’re at a random airport on earth

Is it more likely to be …
a large airport or a small airport?

If you take a random flight out of it …
will it go to a large airport or a small airport?

An example of friendship paradox

Pick a random airport on Earth

Most likely it will be a small airport

However, no matter how small it is, it **will** have flights to big airports

On average **those airports will have much larger degree**

Summary

Things to remember

- How to quantify degree correlations:
  - Positive: assortative networks
  - Negative: disassortative networks
  - Neutral networks
- Friendship paradox

Sources

- A. L. Barabási (2016). Network Science – Chapter 07
- F. Menczer, S. Fortunato, C. A. Davis (2020). A First Course in Network Science – Chapter 02
- URLs cited in the footer of specific slides