Network Analysis:

The Hidden Structures behind the Webs We Weave

17-338 / 17-668

Homophily and Degree Correlation

Thursday, September 19, 2024

Patrick Park & Bogdan Vasilescu

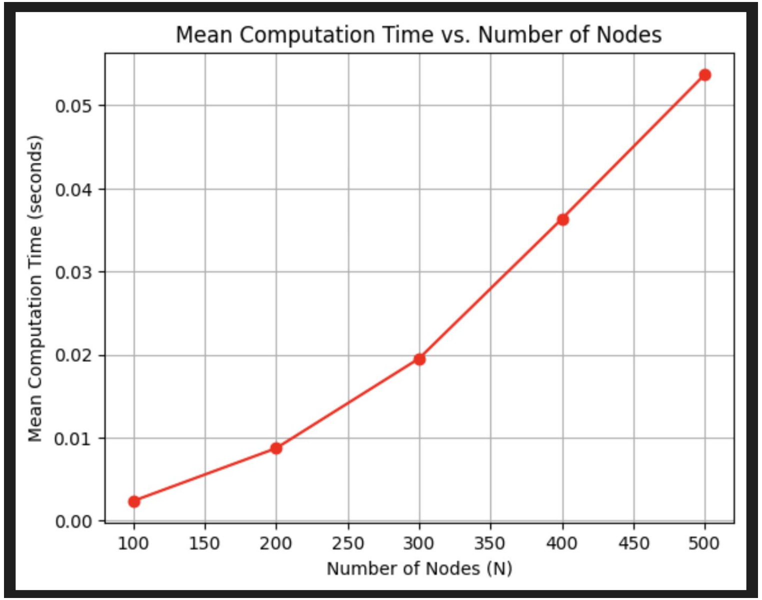
Carnegie Mellon University
School of Computer Science

S3D
Software and Societal
Systems Department

2-min Quiz, on Canvas

A few notes on HW2

Range of answers… 2 minutes, 3.9 minutes, 17 minutes, 164132 minutes, 34606 minutes



Is this a linear or a quadratic relationship?

Korean Twitter network has 1660554 nodes.

- Linear, $r^2 \approx 0.92$ → predicts 3.69 minutes
- Quadratic, $r^2 \approx 0.99$ → predicts 14287 minutes

Quick Recap – Last Thursday's Lecture

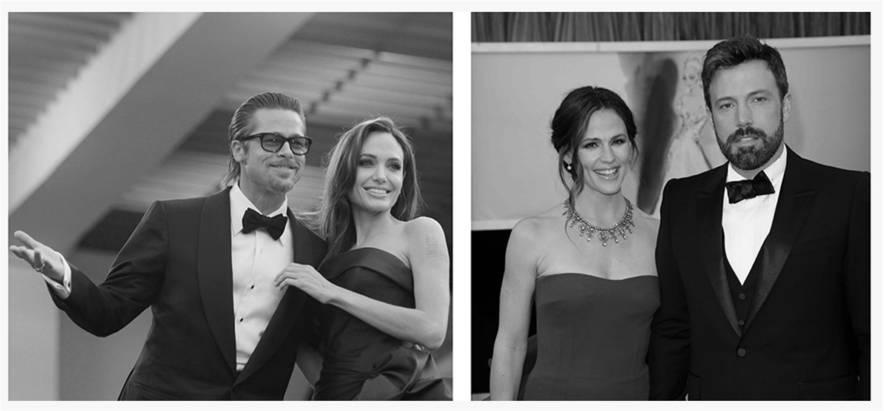
Homophily and how to measure

The natural sciences perspective

Homophily: Status & Power

Degree homophily: “degree assortativity” or “degree correlation” – high-degree nodes tend to be connected to other high-degree nodes and vice versa.

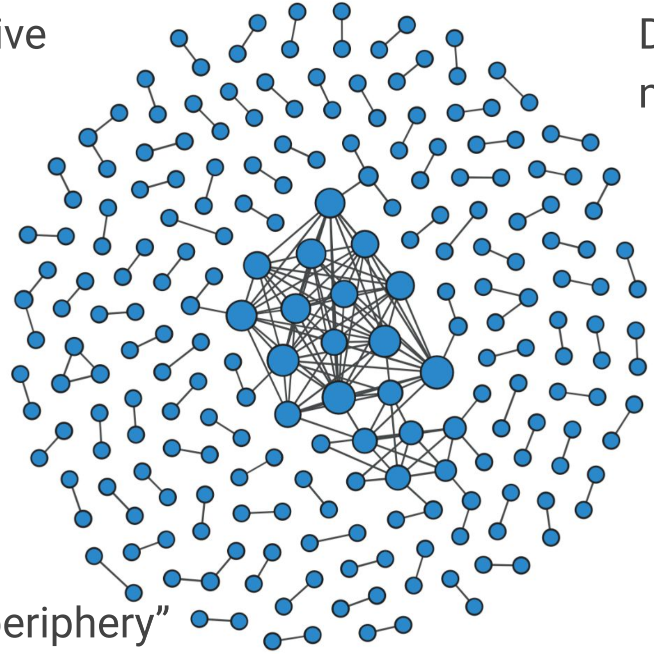
Extensively studied from a graph-theoretic perspective.



Degree Assortativity / Disassortativity

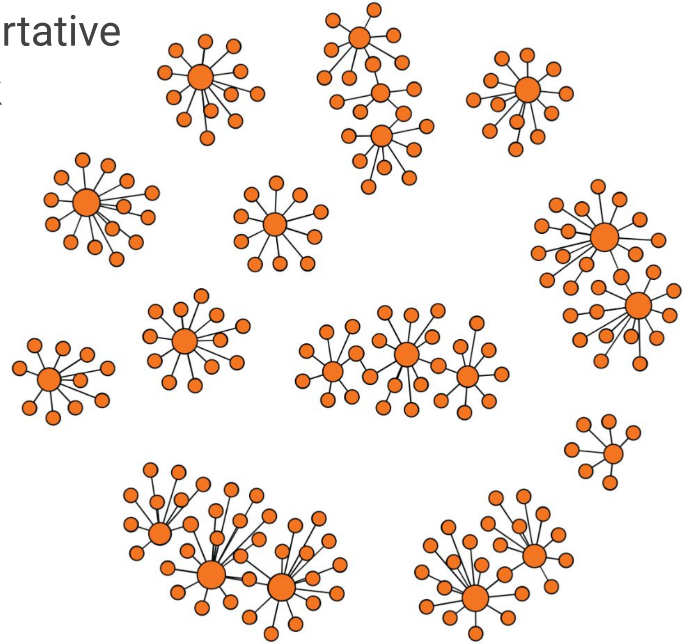
Example:

Assortative network



“Core-periphery”

Disassortative network

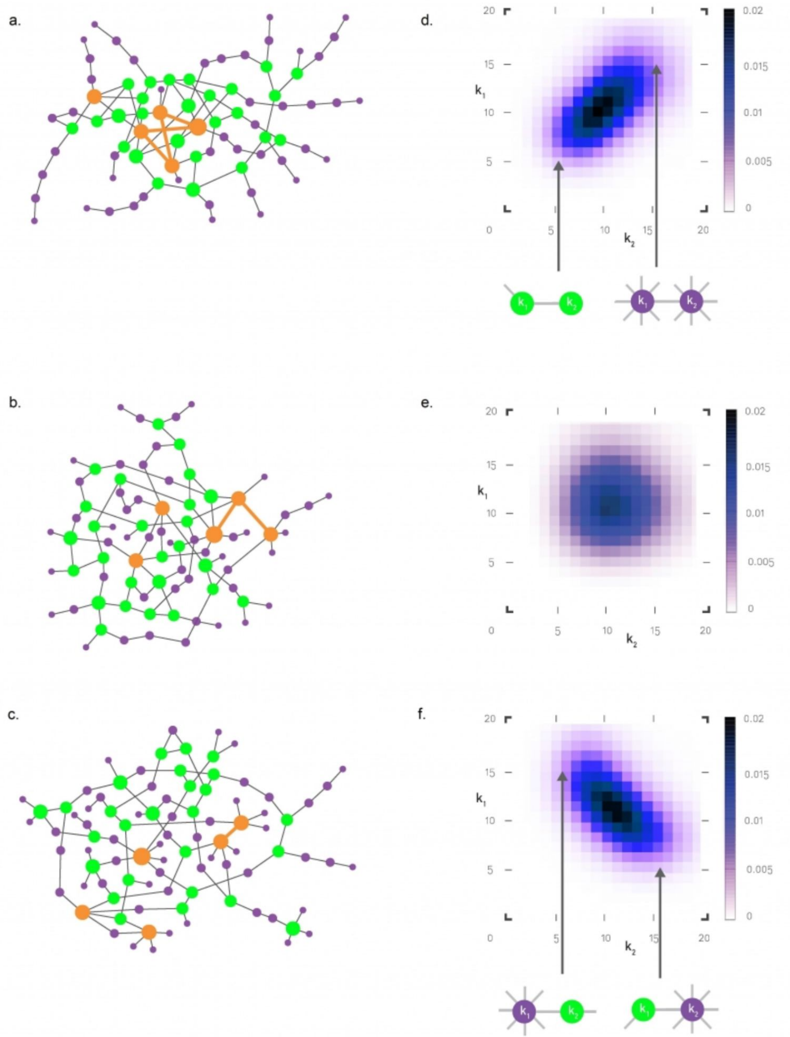


Degree Assortativity / Disassortativity

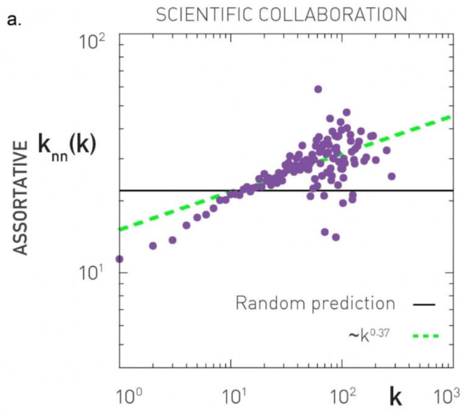
(a) **Positive** degree correlation: Connected nodes have similar degree

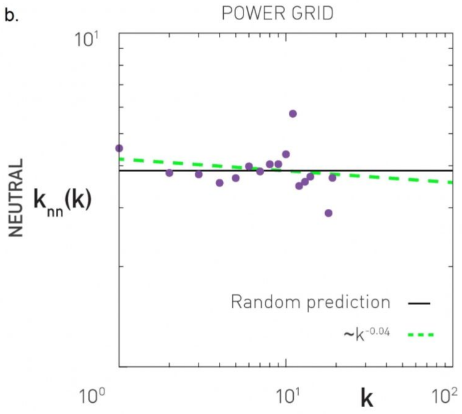
(b) **Neutral**: The degree of connected nodes have no correlation

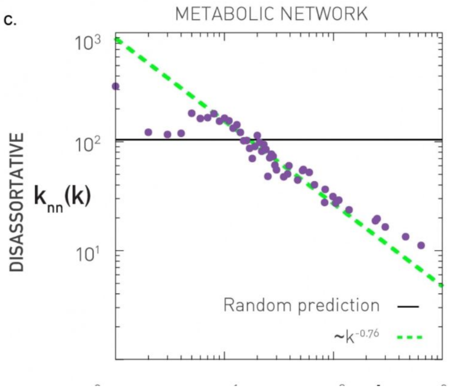
(c) **Negative** degree correlation: Connected nodes have dissimilar degree



Measuring degree correlation:
Average degree of the neighbors of a node of degree k







Average degree of neighbors increases as k increases → assortative network

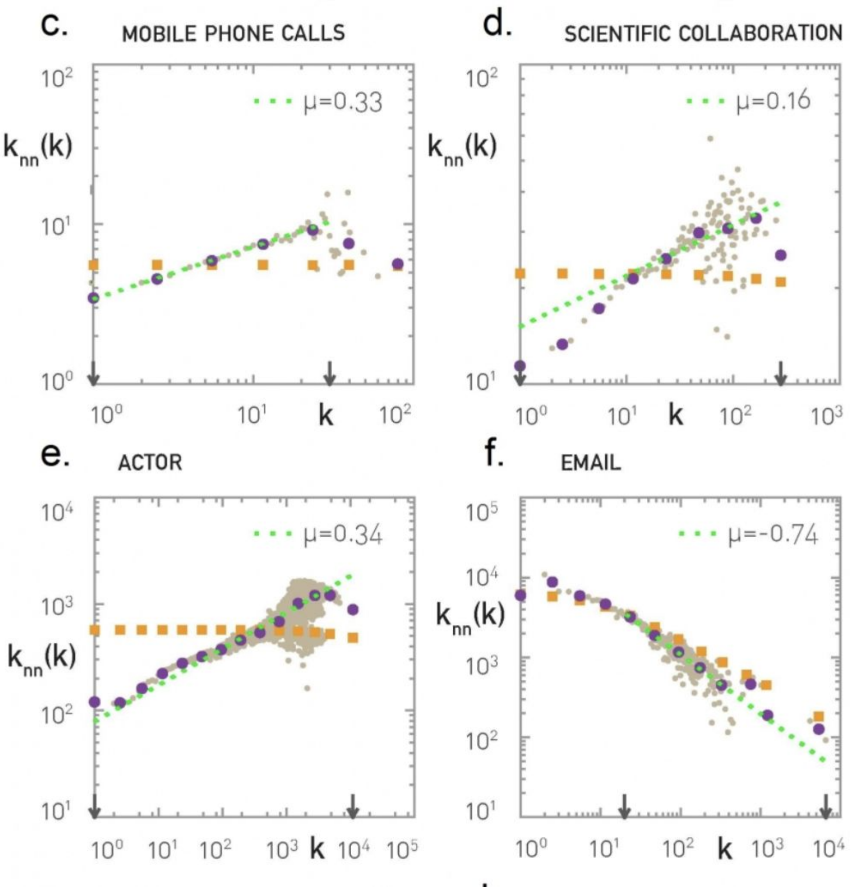
Average degree of neighbors neither increases nor decreases as k increases → degree neutral network

Average degree of neighbors decreases as k increases → disassortative network

Human social networks tend to exhibit positive degree correlations

Why positive?

Why is the email network negative?



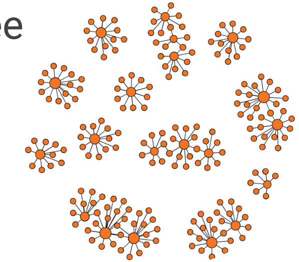
Human social networks tend to exhibit positive degree correlations

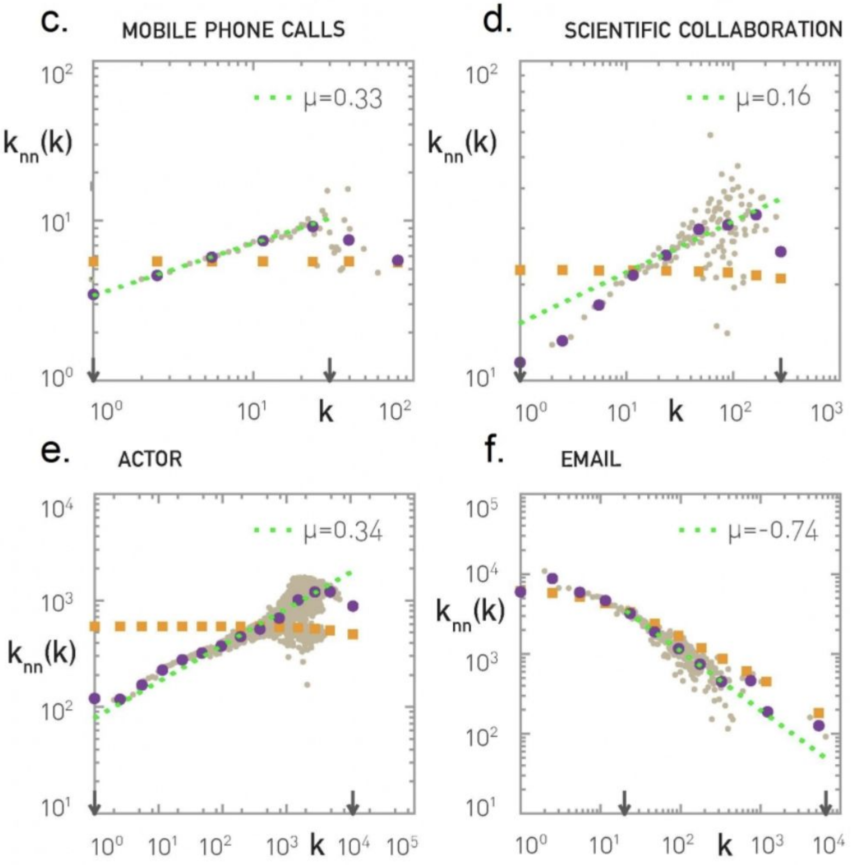
Why positive?
→ Open question. Several studies argue that it is related to the fact that humans form groups

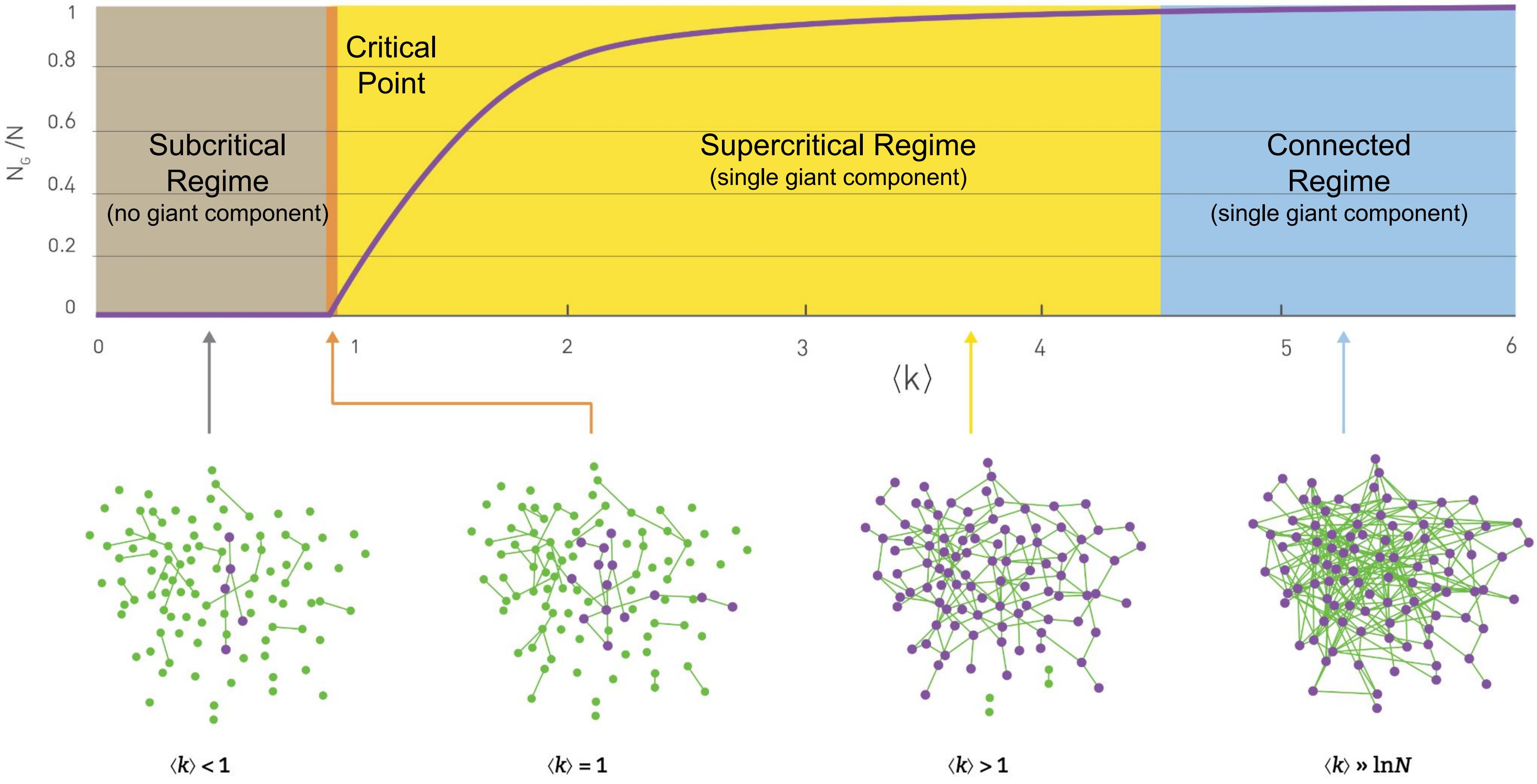
→ People in large groups tend to have high degree (more group members to connect with) and those in small groups are constrained in forming ties - hence low degree

Why is the email network negative?

→ Networks with skewed degree distributions tend to exhibit negative degree correlations





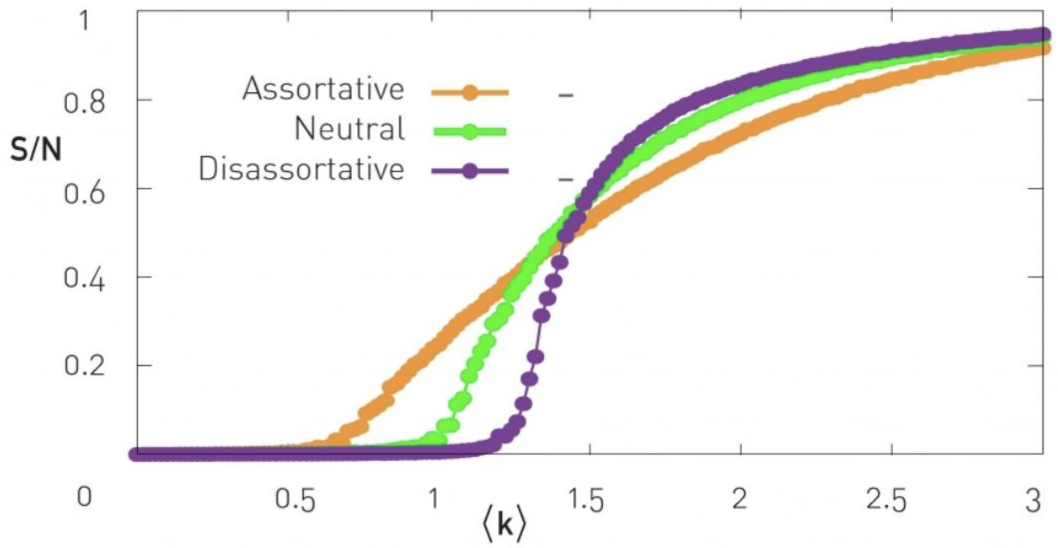


(Barabasi Ch. 3.6; Erdős & Rényi, 1959)

Impact of Assortativity: Higher connectivity

Giant component can emerge at lower mean degree <k>

Size of largest component / Size of entire network

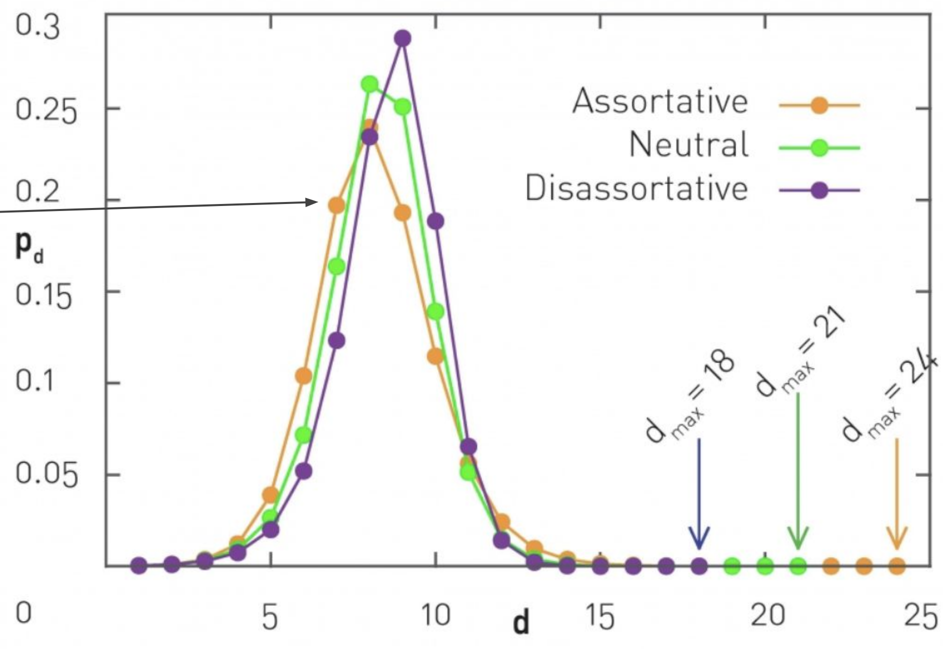


This means connectivity increases even if people do not have many connections

Impact of Assortativity: Higher connectivity

Giant component can emerge at lower mean degree <k>

Assortative networks have shorter average path length



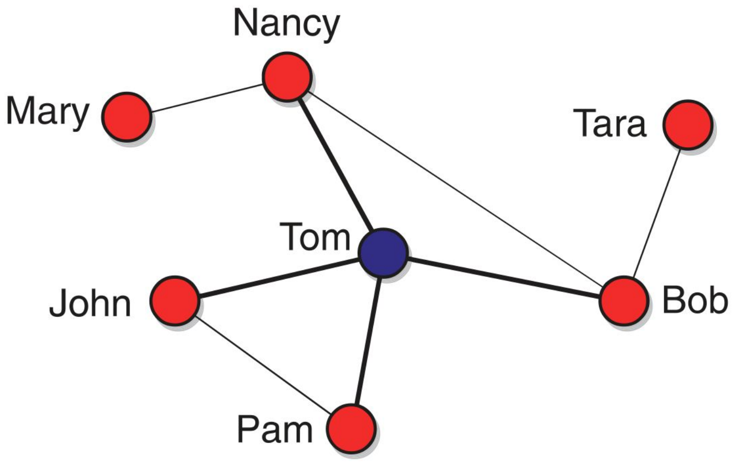
Case Study: The Friendship Paradox

Suppose you are looking for the person with the most friends

You only have a directory of phone numbers

Option 1: Call a person randomly

The chance that you pick Tom is … ?

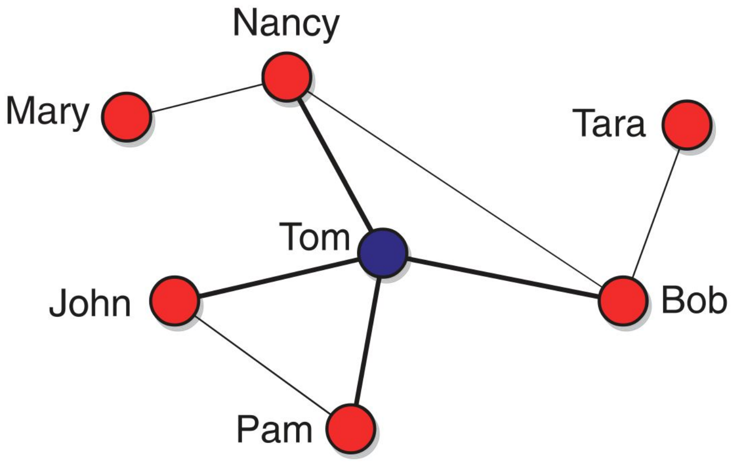


Suppose you are looking for the person with the most friends

You only have a directory of phone numbers

Option 1: Call a person randomly

The chance that you pick Tom is 1/7 ~ 14%

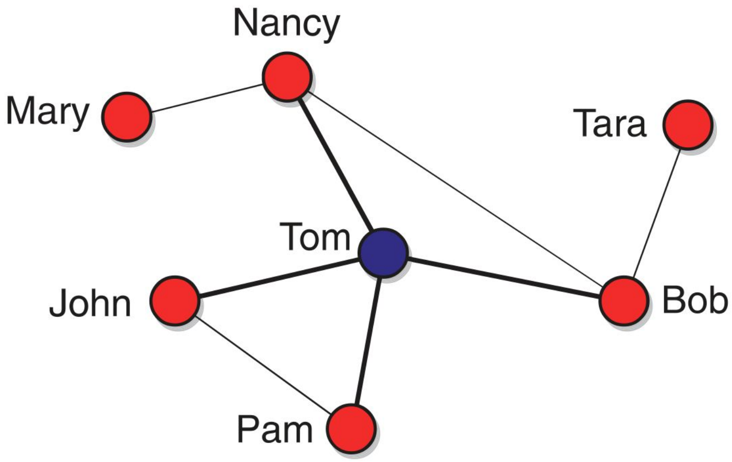


Suppose you are looking for the person with the most friends

You only have a directory of phone numbers

Option 2: Call a person randomly, and ask them about a random friend

The chance that you pick Tom is …?

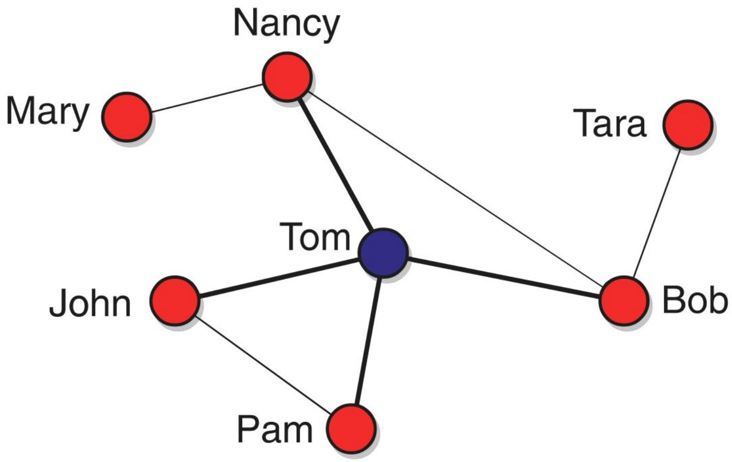


Suppose you are looking for the person with the most friends

You only have a directory of phone numbers

Option 2: Call a person randomly, and ask them about a random friend

The chance that you pick Tom is 5/21 ~ 24%



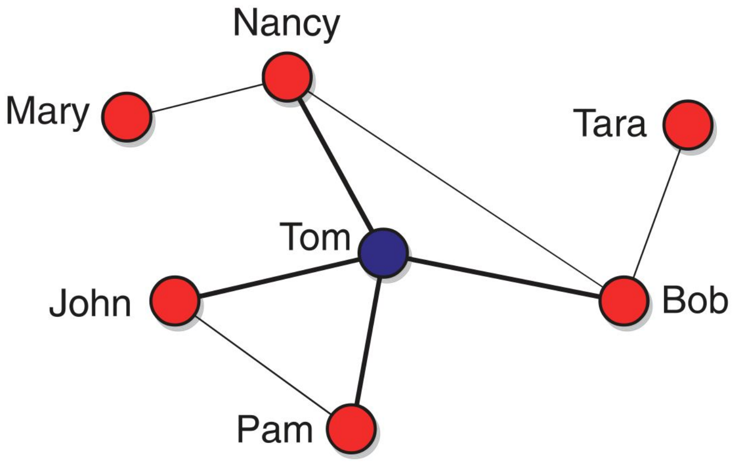
Mary: 0/1, Nancy: ⅓, John: ½, Pam: ½, Bob: ⅓, Tara: 0/1, Tom: 0/4

Probability of being called: 1/7

Therefore: (0/1+⅓+½+½+⅓+0/1+0/4)*1/7 = 5/21

Now, the paradox:

Average degree: ?

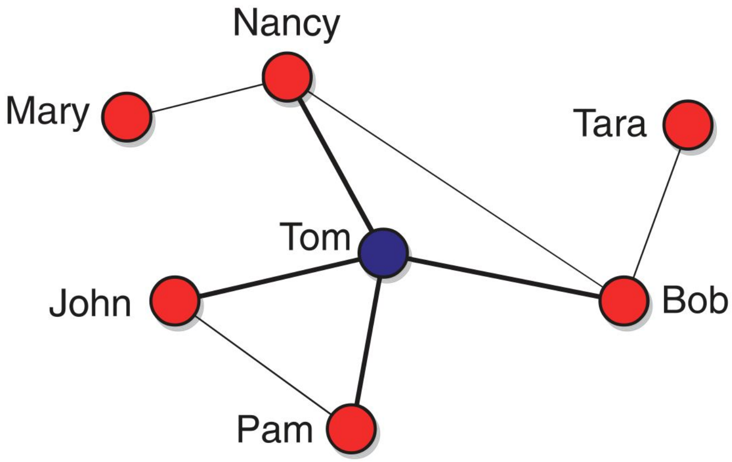


Now, the paradox:

Average degree: (1+3+4+2+2+3+1)/7

= 16 / 7 = 2.29

Average degree of neighbors: ?



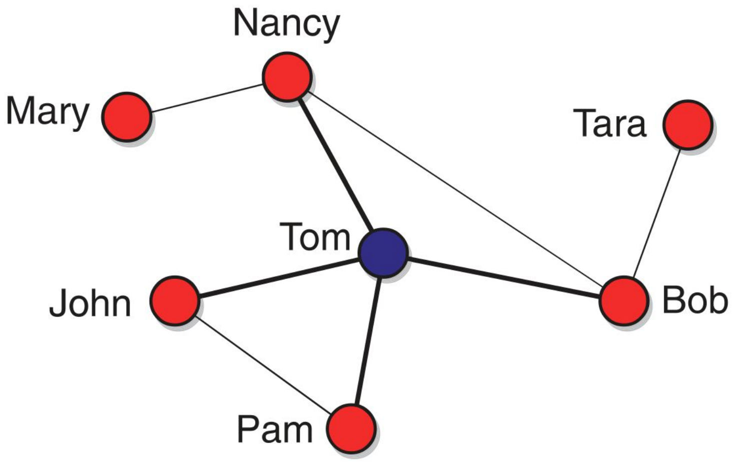
Now, the paradox:

Average degree: (1+3+4+2+2+3+1)/7

= 16 / 7 = 2.29

Average degree of neighbors:

(3+8/3+10/4+3+3+8/3+3) / 7 = 2.83

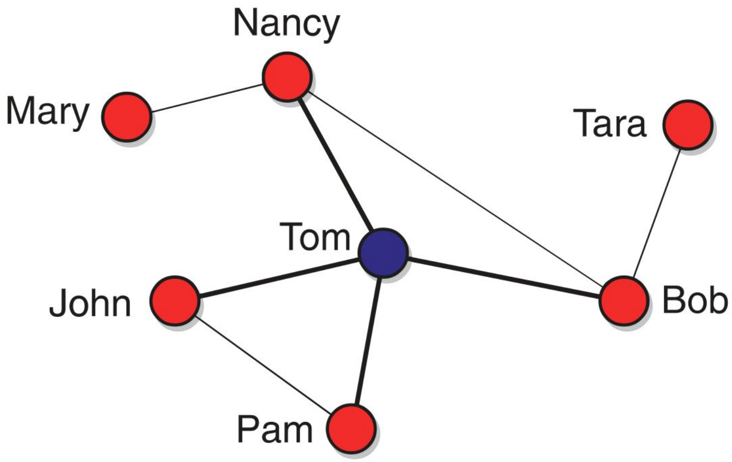


Now, the paradox:

Average degree: 2.29

Average degree of neighbors: 2.83

Your friends have more friends than you, on average!

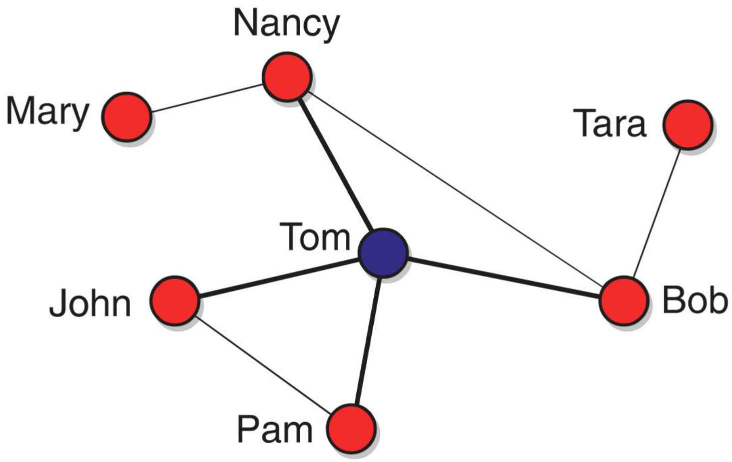


But it doesn't hold for everyone:

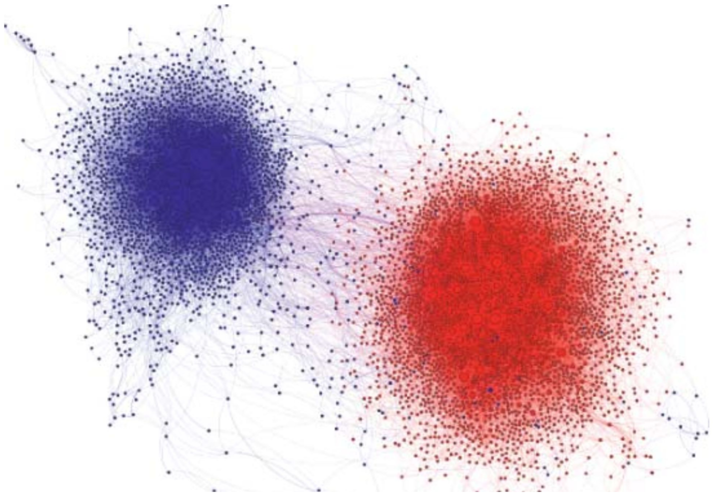
Nancy has 3 friends: Mary, Tom, Bob

They have in total 1 + 4 + 3 = 8 friends

→ Nancy's friends have on average 8/3 friends (i.e., less than Mary)



Aside: The dark side of homophily



Exceedingly easy to connect with people who share our worldviews and unfriend / unfollow people with different opinions.

Information can be shared and consumed in such a selective and efficient way as to influence our opinions very effectively.

Result: segregation and polarization of our online communities.

High risk of manipulation by misinformation and social bots.

Aside: Networks can also exhibit inverse homophily

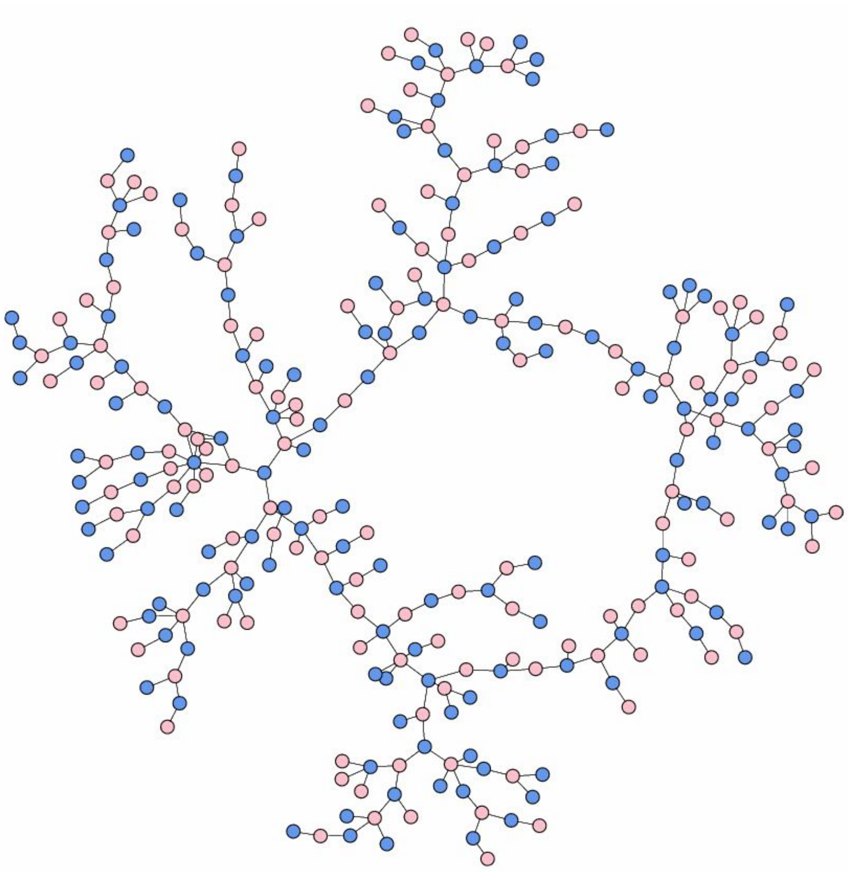
If the fraction of cross-gender edges is significantly more than 2pq.

Do you remember any example?

Aside: Networks can also exhibit heterophily

If the fraction of cross-gender edges is significantly <u>more</u> than 2pq.

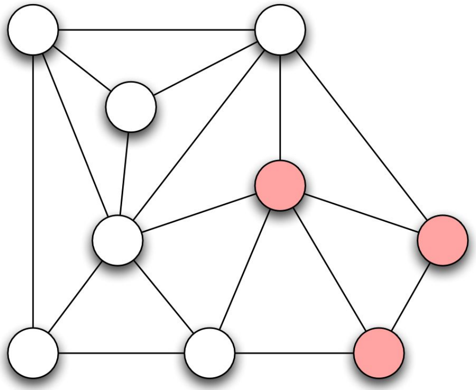
Yes! The high school dating network



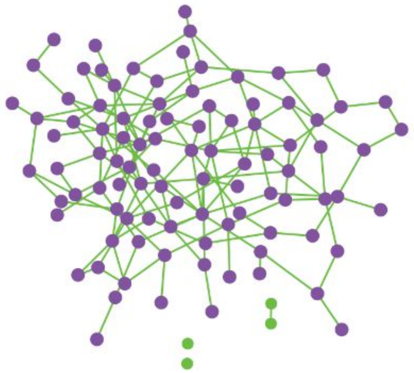
Comparing Homophily between Groups

Problem:

If groups X and Y have different levels of homophily, how can we measure them separately and compare them to each other?



Approach 1: Compare the observed probability of a red-red tie to a random baseline, do the same for the white-white tie, and see which observed probability deviates farther from random.



Measuring homophily

What is the observed probability of a tie between two nodes from group x?

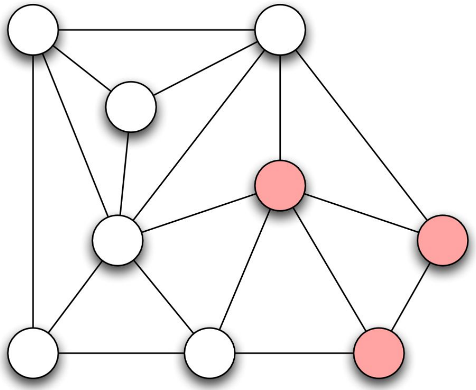
→ $$\widehat{Pr_{xx}} = \frac{L_{xx}}{L}$$

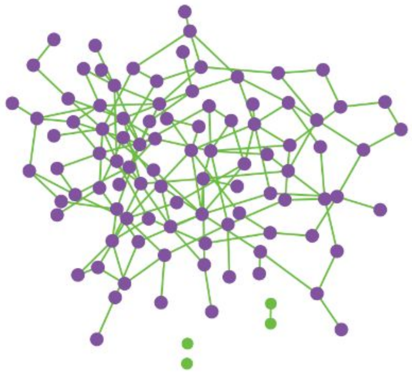
What is the random baseline probability?

→ $$Pr_{xx} = \left(\frac{N_x}{N}\right)^2$$

Hc oes the observed deviate from the random baseline?

→ $$\theta_{xx} = \frac{\widehat{Pr_{xx}}}{Pr_{xx}}$$

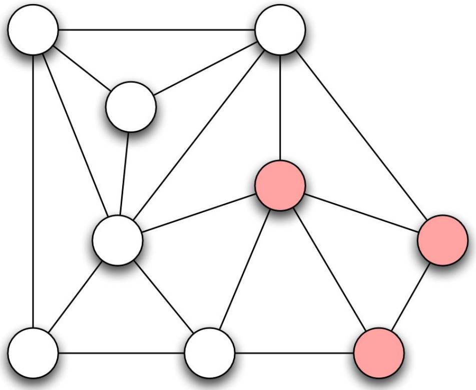




Measuring homophily

Q: What is a hidden assumption in this homophily test?

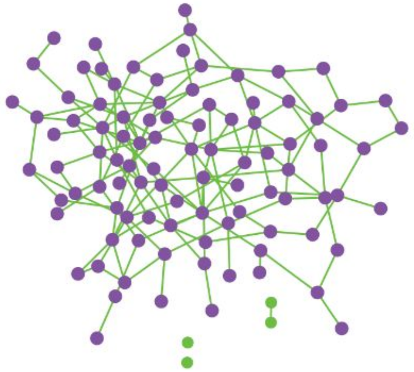
Hint: Recall how Erdos-Renyi random graphs are constructed.



Every dyad has equal probability, *p*, of getting connected

So, **both groups will have the same average degree**

$$D_x = D_y$$



Measuring homophily

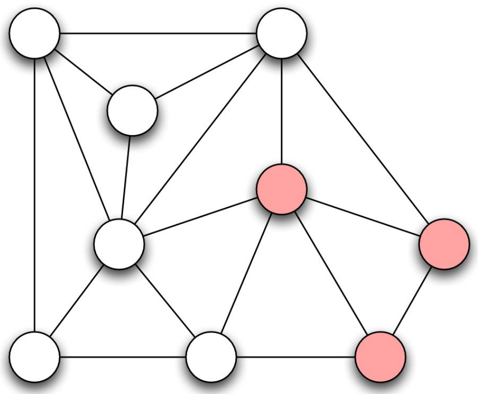
Actual average degrees of x and y

x: red, y: white

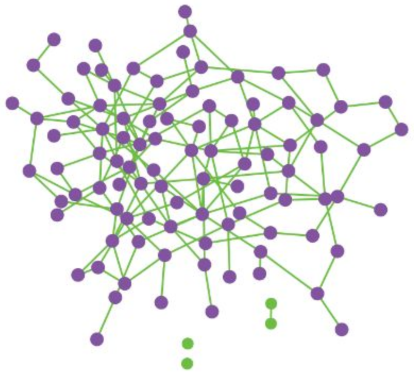
Dx = 10/3 = 20/6

Dy = 25/6

Dx < Dy



So, even if group x and y have the same homophilous tendency, group y will have more friends, so they may appear more homophilous



Summary

We’ve seen another fundamental property of networks: similarity between neighbors

(Recall short paths connecting nodes and triangles formed by common neighbors)

Two <u>extremely</u> powerful analysis techniques: comparison to a random (shuffled) network and longitudinal analysis!