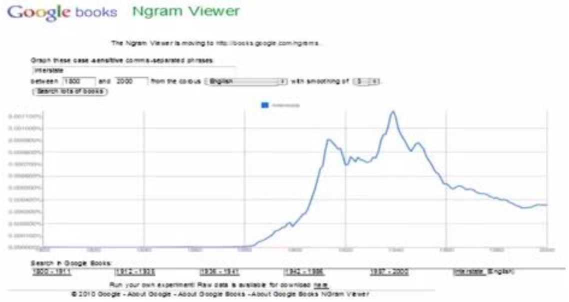
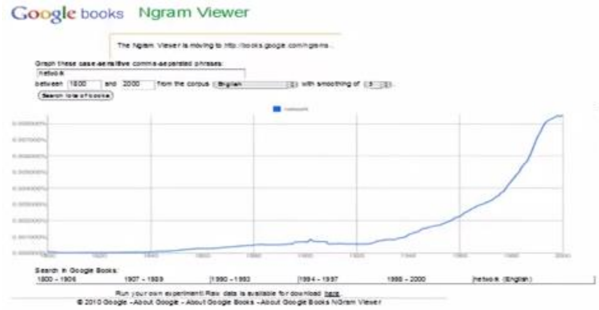
Session 14 - Networks

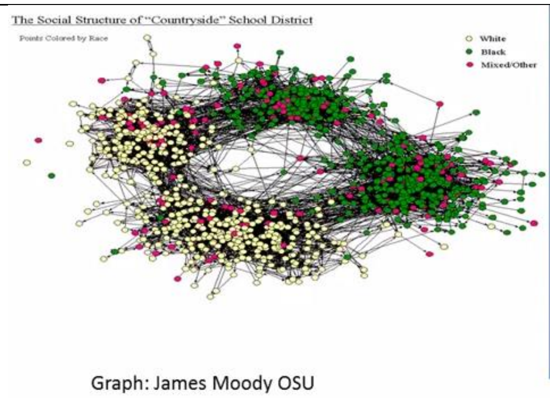
14.1 Networks



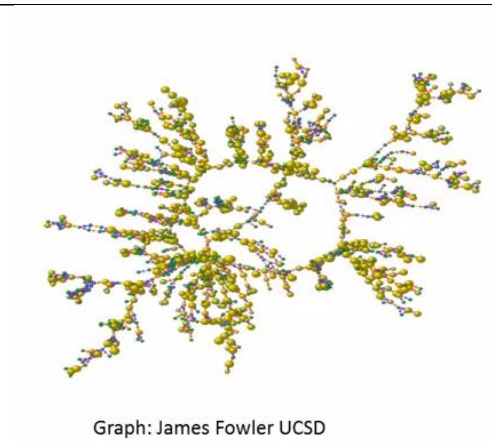
Search for 'hiway'



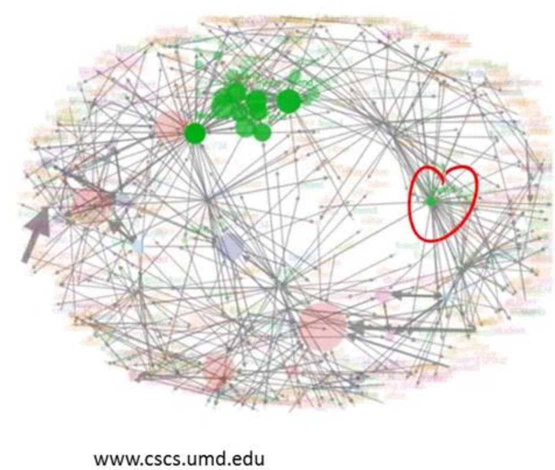
Search for 'network'



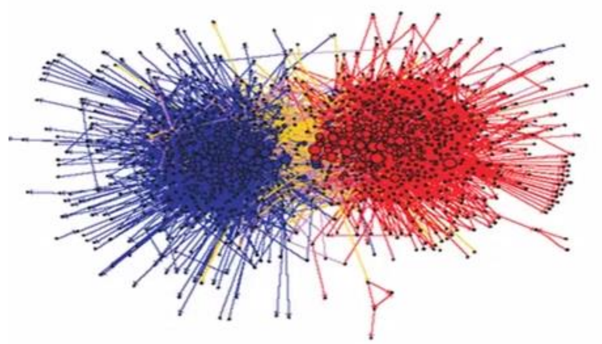
Middle school:
yellow⇔ black, green ⇔ caucasian, red ⇔ girls



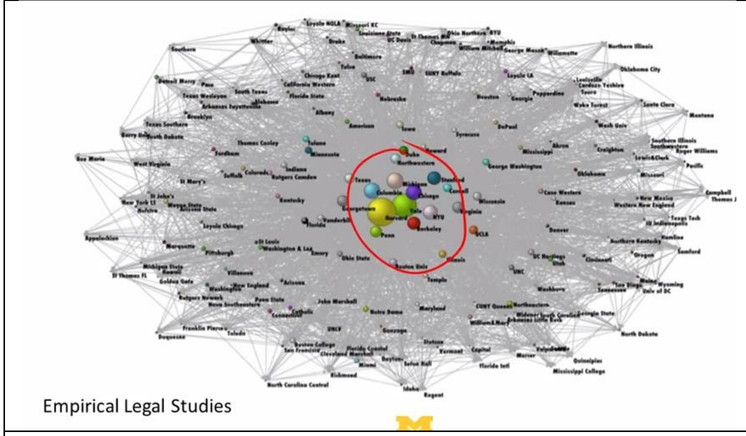
Adult friends network

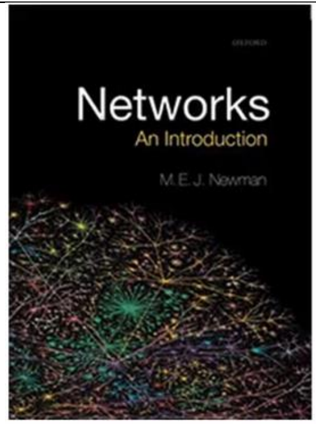


Corporate network



659 Conservative(red) / 676 liberal(blue) bloggers with respective links plus yellow (liberal to conservative) and purple (conservative to liberal). Lada Adamic ➔ polarized discourse

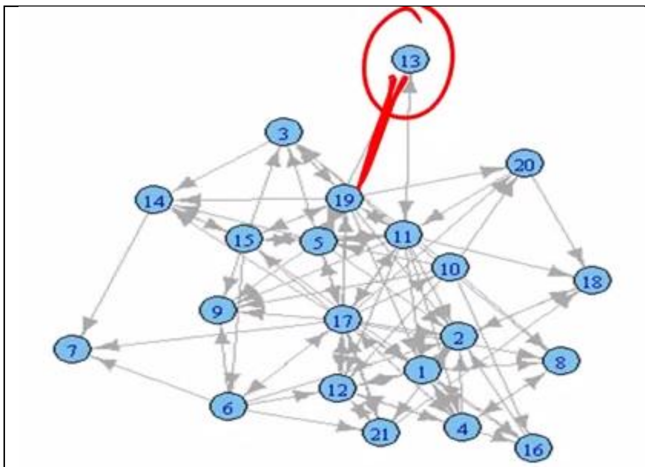




This session will focus on the
(1) the logic forming networks - how it forms,
(2) the structure – measures, and
(3) the function of the network – what it does, enables (often emergent rather than planned)
Topics include: six degrees, emergence

Book:
Networks, an Introduction
by M E J Newman

14.2 Structure of Networks



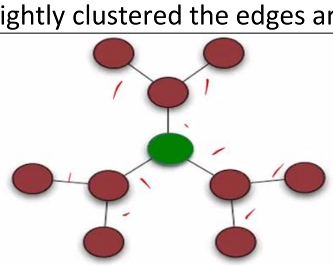
Nodes & Edges:
directed (↗) or undirected (/)

Undirected: / e.g.,
Nodes - 50 United States,
Edges - Adjacencies (two states connected by an edge if they share a border)

Directed: ↗ or ↗↙ e.g.,
Nodes -students**,**
Edges - if one person looks to other for fashion advice. Directed can be uni- or bi-directed.

Structure: Measures – (a) degree (how many edges each node has on average)**, (b) path length** (how far to another node)**, (c) connectedness** (is the entire graph connected to itself), and **(d) clustering coefficient** (how tightly clustered the edges are)

Degree (node): number of edges attached to a node
Degree (network): average degree of all nodes
$Average\ Degree\ =\ 2(Edges/Nodes)$



Example:
10 nodes, 9 edges, average degree = 2(9/10) = 1.8

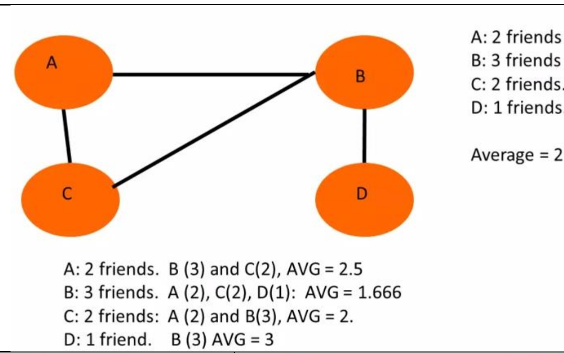
Quiz: A network has 50 nodes & 100 edges. What is the degree of the network?
Ans: 2(100/50) = 4

Neighbors of a node: all other nodes connected by an edge to the node.

Theorem: The average degree of neighbors of nodes will be at least as large as the average degree of the network.

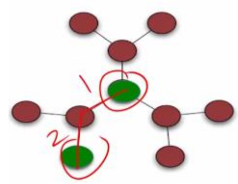
Implication: *Most people's friends are more popular than they are!!*

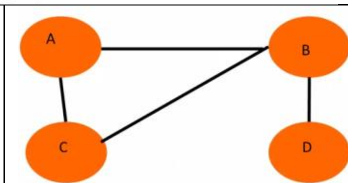
Note chart: A, B, C, D average 2 friends. BUT Their friends average (2.5, 1.67, 2.5, 3) = 2.4 is more than their average, 2.4 vs. 2.



Path length A to B: the minimal number of edges that must be traversed to go from node A to node B. See left diagram. Path Length is 2

Average path Length: Average path length between all pairs of nodes in a network. AB(1), AC(1), AD (2), BC(1), BD(1), CD(2) → 8/6 = 1.33 avg path length.

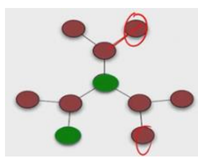




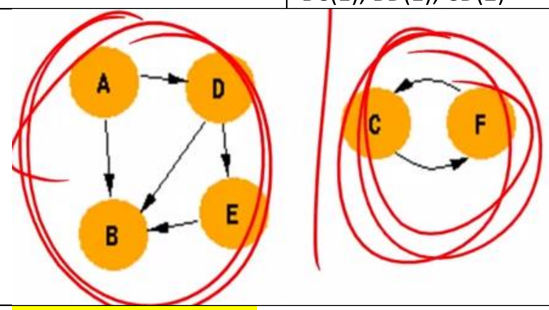
path lengths: AB(1), AC(1), AD (2), BC(1), BD(1), CD(2)

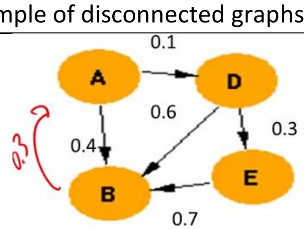
Connectedness: A graph is connected if you can get from any node to any other by way of the edges.

Connected Graph →



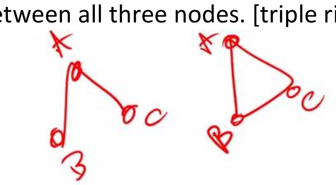
Example of disconnected graphs at right →

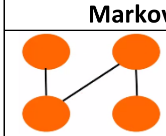




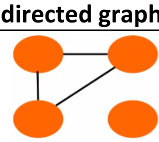
Markov Process as a directed graph

Clustering Coefficient: Percentage of triples of nodes that have edges between all three nodes. [triple right]

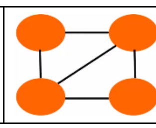




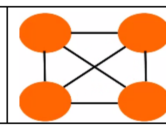
CC=0/4
= 0 , 4 is max Δ's



CC=1/4
= 0.25

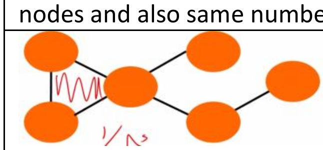


CC=2/4
= 0. 5

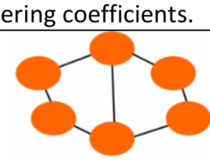


CC=4/4
= 1.0

Quiz: Draw two connected graphs each with 6 nodes and 5 edges with different clustering coefficients. Edges tell how connected the graph is but it doesn't tell how clustered it is. Possible to have two graphs with same nodes and also same number of edges but different clustering coefficients.



1Δ out of 6!/3! = 20 possible triangles if edges added → CC=1/20



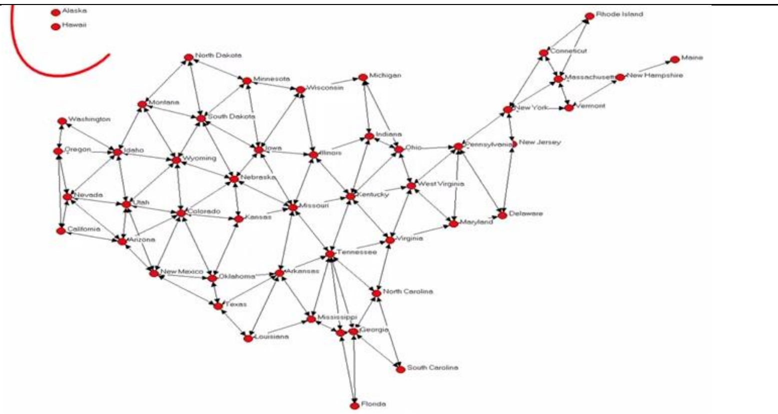
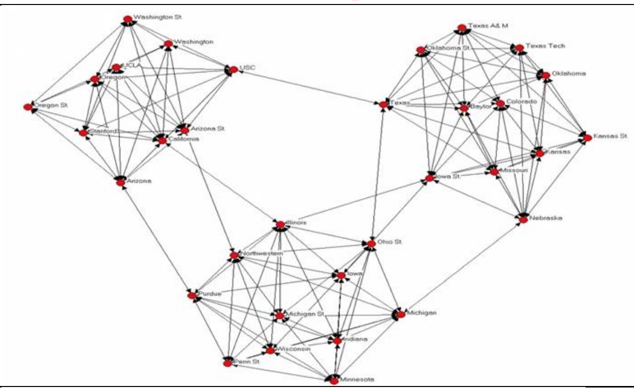
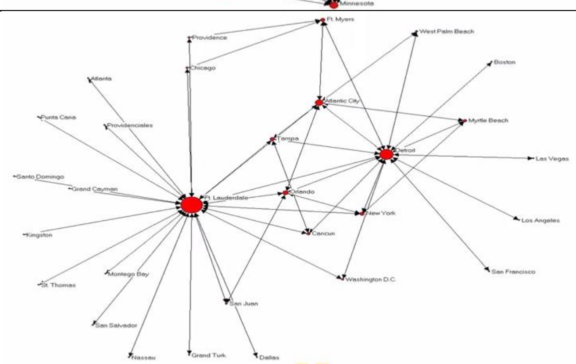
No triangles but 6 node → 20 potential triangles if edges added: hence clustering of 0.

Degree: → (a) density of connections, (b) social capital [social connectedness], (c) speed of diffusion

Path Length: (a) hurdles - e.g., # flights needed, (b) social distance [rate of info transfer], (c) likelihood of information spreading

Connectedness: (a) Markov process applicability, (b) implication for terrorist group capabilities, (c) robustness/fragility, e.g., Internet or power failure, (d) information isolation potential.

Clustering Coefficient: (a) redundancy/robustness (Δ's), (b) social capital, (c) innovation adoption(Δ's, virtuous, reinforcing feedback/cycle)

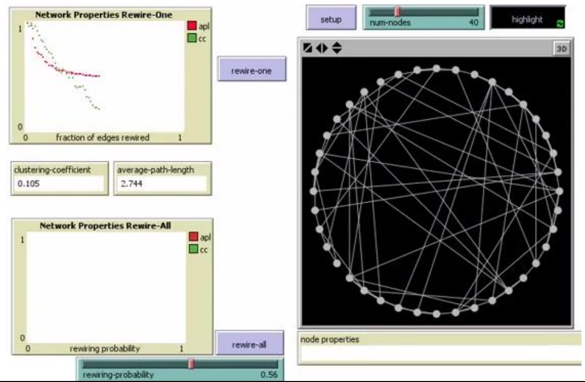
US States: Adjacency With Alaska and Hawaii isolated. Note Many clusters (Δ's) so some Cluster Coefficient **Degree = 8.6, Path = 6.1**	
Universities: Football Games With three conferences, inter and intra conference games. Note many clusters (Δ's) hence high Cluster Coefficient **Degree = 10.1, Path = 4.0**	
Airline: Flights With hubs Note few clusters (Δ's) hence low Cluster Coefficient **Degree = 10.8, Path = 4.1** **Key Point**: Degree & Path statistics of these three networks are similar but structure is different. Clustering coefficients are different. Need several metrics to distinguish structure.	

14.3 Logic of Network

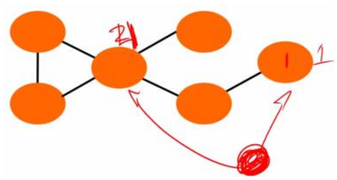
How Networks Form: **Random** – connections that form at random **Small Worlds** – cliques that connect at random to others. **Preferential Attachment** – Internet & WWW, more likely to connect to dominant nodes. Erdish Rainey Network Contextual Tipping Point 	**Random Attachment** – N nodes, P = probability that two nodes randomly connect.  Erdish Rainey Networks: Contextual Tipping Point is exhibited for large N. The network almost always becomes connected when $P > 1/(N-1)$. BUT real social networks (see 13.1 social network diagrams) don't look like random networks. Leads to Small World graph.

Small Worlds Network – People have some percentage of "local" or "clique" friends and some percentage of random friends. (Duncan Watts)

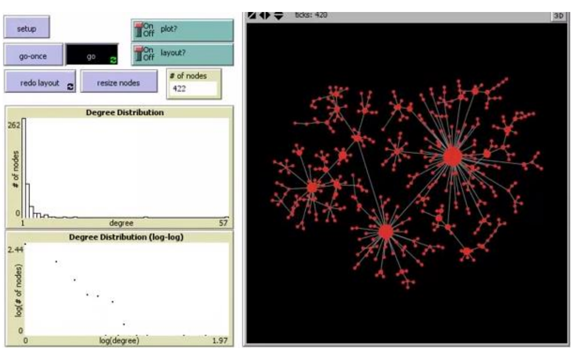
As random paths are added, the CC (chart green line) drops as well as the Average Path Length (red). As people have more random friends we get less clustering and a shorter average distance between people.

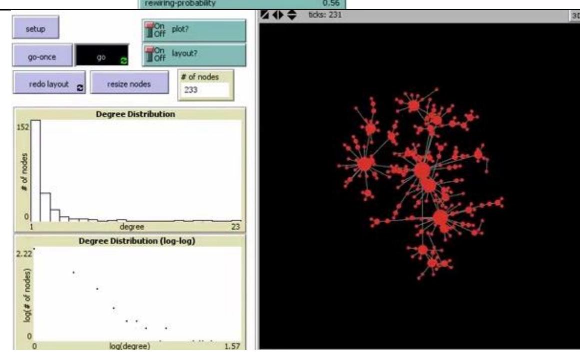


Preferential Attachment – A 'Node' arrives. The probability that 'Node' connects to an existing node is proportional to another node's **Degree**.



Below, one node is 4 times more likely to be connected to than the other shown.





Netlogo example starting from a few nodes and expanding by connections weighted by the node degree. Note that some have a large number and most have just a few (Degree distribution histogram). Degree distribution is the log-log chart. Another example to the left has the same Degree Distribution but evolves into a different graph. The exact network you get is 'path dependent'. The ***Degree Distribution <u>is not</u> path dependent.*** Don't have a path dependent equilibrium of the Degree Distribution. The Chris Anderson ***'Long Tail'*** is an example of this Degree Distribution with many nodes having a Degree of 1 and only a few with a large Degree.

Quiz: Sophia is looking for a job after she graduates from college. Which kind of network best explains why she is more likely to find a job in New York City than in Lima, OH (a small town)? (a) random, (b) small world, (c) preferential attachment, (d) path length.

Ans: (c) preferential attachment,

Explanation: *(a) (wrong) Random: if she just chose where to live randomly, she could end up anywhere- it doesn't explain why a more populated town would be more likely to attract a new resident than a less populated town. (b) (wrong) Small World: Now, this might work, but it's not the best response. Perhaps it's more likely that Sophia has small world connections to New York since there are more people there and this helps her find a job there. However, we don't know about her connections with the information given, therefore we shouldn't be making that assumption. (c) (correct) Preferential Attachment: There are already a lot of people in New York. Preferential attachment suggests that the probability of connecting to a certain node is proportional to the number of connections nodes already have. New York has far more connections than Lima therefore she is more likely to go to New York as well. That more people are there already probably means more job opportunities.*

Summary:

- **random connections** ➔ interesting tipping point phenomena.
- **Small world connect network** as people have more random friends ➔a lot less clustering, but shorter social distance between people.
- **Preferential attachment rule** ➔ long tailed degree distribution (most nodes have very, very low degree, but a handful of nodes have a really, really high degree).

14.4 Network Function

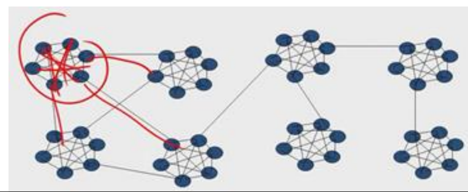
Discuss how the network structure enables certain things to occur. How network carries out Function. These functions are typically emergent from the logical process of how it forms.

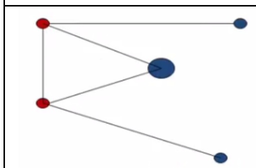
Six Degrees: Stanley Milgram (1960's) asked 296 people from Nebraska to get a letter to a stockbroker outside Boston. Could only send the letter to someone they know on a first name basis. Result ~ 6 steps.

Duncan Watts almost 40 years later did an email experiment with 48,000 senders and 19 targets in 57 countries. The average number of connections (steps) was six. Explain with random clique below.

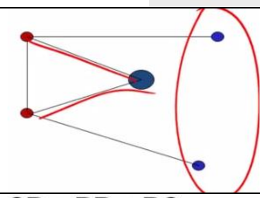
Investigate with the Random Clique Network: Each person has C clique friends and R random friends.

Add **K-neighbor:** all nodes that are of path length K to a node but not of any shorter path length (shortest distance).



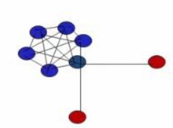


1 - neighbors to large blue node are the two red nodes.



2 – neighbors to the large blue node are the two small blue nodes

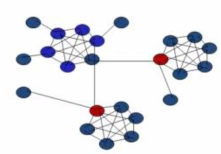
R + C



1 – Neighbors set:
Random friends (R–red)
Clique friends (C-blue)

Note: clique friends of clique friends (CC) in right diagram are just a node's clique friends (CC = C) so no CC in 2-neighbor set.

CR + RR + RC



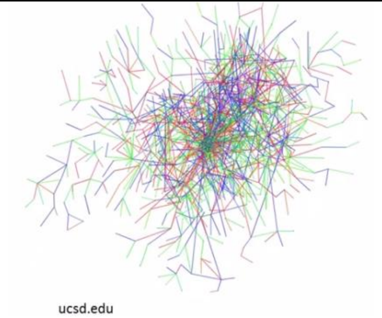
2 – Neighbors set:
CR = random friends of clique friends
RR = random friends of random friends
RC = clique friends of random friends

3 – Neighbors set:
RRR + RRC + RCR + CRR + CRC but note that RCC = RC, CCR = CR, and CCC = C so are not included.

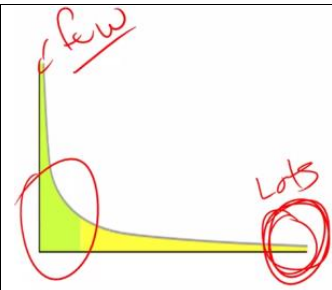
Given C = 140 & R = 10 find n-neighbors.
1: C + R = 140 + 10 = 150
2: CR + RC + RR = 1400 + 1400 + 100 = 2900
3: RRR + RRC + RCR + CRR + CRC = 1000 + 14000 + 14000 +14000 + $10(140)^2$ = 229,000

Quiz: Regina has 30 clique friends and 20 random friends. Assuming everyone has the same number of clique and random friends, ho many 2-neighbors does she have? (a) 1600, (b) 50, (c) 30,000, (d) 24,600
Analysis. CR + RC + RR = 30(20) + 20(30) + 20(20) = 600 + 600 + 400 =1600
Ans: (a) 1600
Explanation: *2-Neighbors: CR+RR+RC, (30)(20)+(20)(20)+(20)(30)=1600*

Granovetter(1973) "***The Stength of Weak Ties***" – people get jobs, find friends, partners, etc. . . through weak ties. That is your 2- and 3- neighbors; roomate's brother's friend, mother's coworker's daughter, high school friend's college roomate's dad. Why? Note we have: 150 1N's, 2900 2N's, & 229,000 3N's!



UCSD research collaboration links.
Shows how certain researchers drive the theories & research.



The Internet is relatively robust to random node failures as most have few connections. However, it is much more susceptible to failure if large connected nodes are targeted to be disabled.

Summary:
Logic: how networks form.
Structure: measures (of connectedness, degree, and path lengths, statistical properties).
Function: what a network does/enables (robustness, degrees of separation, connected or not).

Intervention:

Recall the SIS model which assumed random contacts. With the networks, we can identify high degree nodes that then enable more powerful interventions. This is an argument for being a many model thinker to mix models appropriately to gain better insights.