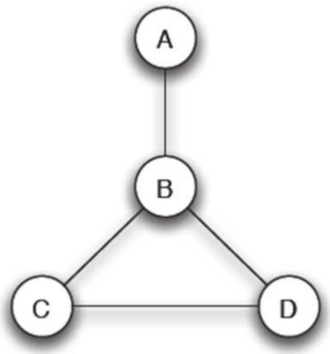
Graphs

Objectives

- Study of network structure
- Basic ideas behind graph theory
- Formulate basic network properties in a unifying language
- Some fundamental applications of the definitions

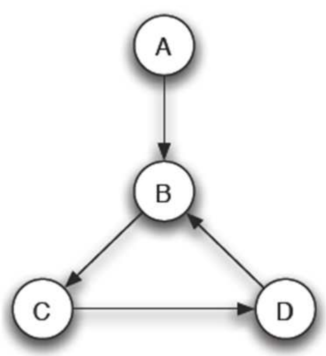
Definitions

- A graph is a way of specifying relationships among a collection of items
- A graph consists of a:
  - set of objects, called nodes,
  - with certain pairs of these objects connected by links called edges

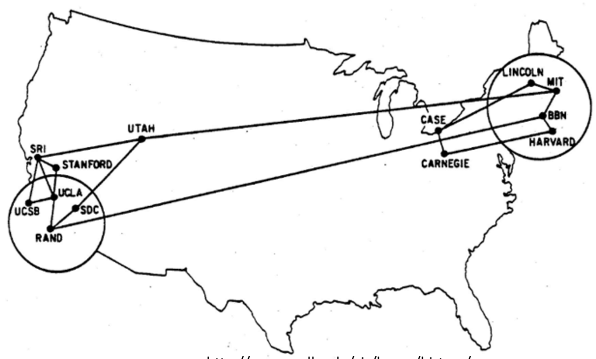


Definitions

- In many settings we want asymmetric relationships
  - Ex: *A* points to *B* but not vice versa
- A directed graph consists of:
  - a set of nodes (as before),
  - together with a set of directed edges
  - each directed edge is a link from one node to another (direction being important)



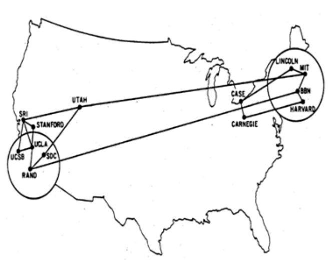
Graphs as Models of Networks

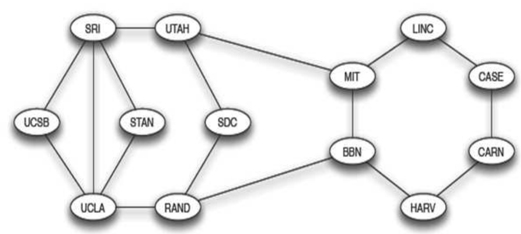


http://som.csudh.edu/cis/lpress/history/arpamaps

- The network structure of the Internet (then called the Arpanet) in 1970
  - 13 nodes represent computing hosts, edges joining when direct communication link

Graphs as Models of Networks





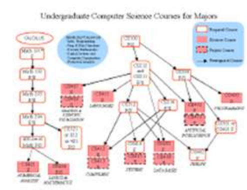
- The actual placement of nodes is immaterial
- All that matters is which nodes are linked to which others

Graphs as Models of Networks

- Graphs appear in many domains
  - In previous lecture we saw examples from social networks, information networks (Web pages), etc.

(a) *Airline routes*

(b) *Subway map*



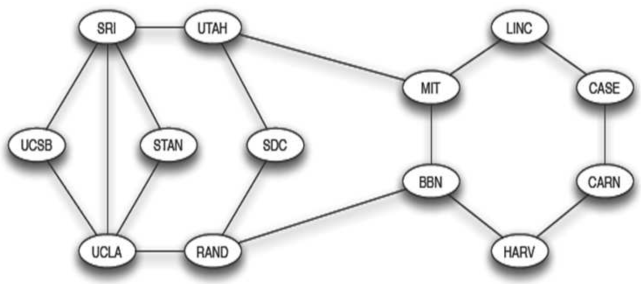
(c) *Flowchart of college courses*

(d) *Tank Street Bridge in Brisbane*

Paths

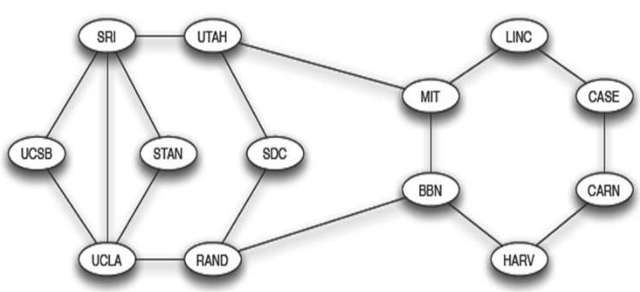
- Travel across the edges of a graph, moving from
- node to node in sequence
  - Ex: a passenger taking a sequence of airline flights
  - a piece of information being passed from person to person in a social network,
  - a computer user or piece of software visiting a sequence of Web pages by following links
- A path is simply a sequence of nodes with the property that each consecutive pair in the sequence is connected by an edge

Paths



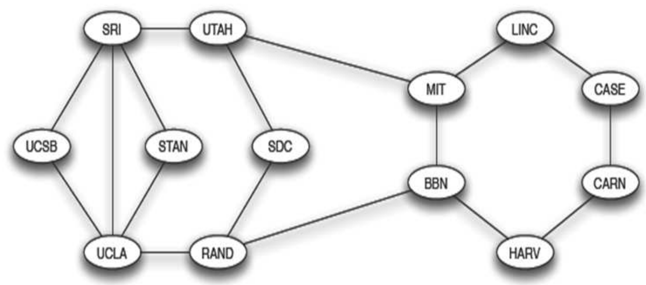
- MIT, BBN, RAND, UCLA is a (simple) path
- SRI, STAN, UCLA, SRI, UTAH, MIT is a path with repeated nodes
- SRI, STAN, UCLA, SRI, UCSB, UCLA, SRI is a path with repeated nodes and edges

Cycles



- A “ring” structure
  - Ex: LINC, CASE, CARN, HARV, BBN, MIT, LINC
- A cycle is:
  - a path with at least three edges,
  - in which the first and last nodes are the same,
  - but otherwise all nodes are distinct

Cycles



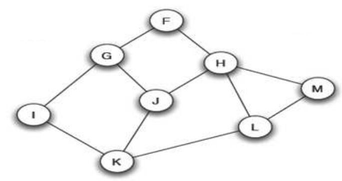
- Notice: every edge in 1970 Arpanet belongs to a cycle
- This was by design:
  - if any edge were to fail there would still be a way to get from any node to any other node

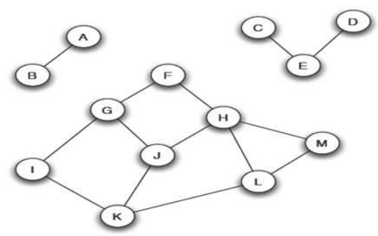
Cycles

- Communication and transportation networks often contain (by design) cycles
  - allow for redundancy and provide for alternate routings that go the "other way" around the cycle
- In the social network of friendships we notice cycles in everyday life
- Ex: your wife's cousin's close friend from high school is in fact someone who works with your brother
  - This is a cycle: you, your wife, her cousin, his high-school-friend, your brother), you

Connectivity

- Does every node can reach every other node by a path?
- A graph is connected, if for every pair of nodes, there is a path between them
- Otherwise is called disconnected



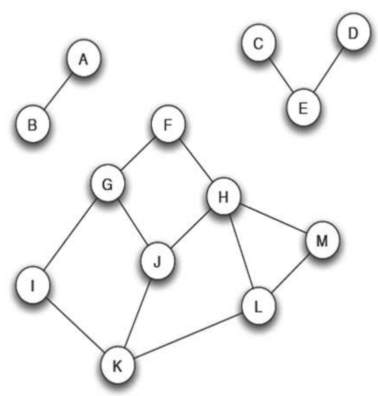


Connectivity

- the 13-node Arpanet graph is connected
- Most communication and transportation networks are connected
  - their goal is to move traffic from one node to another
- In a social network, there might exist people for which it's not possible to construct a path from one to the other

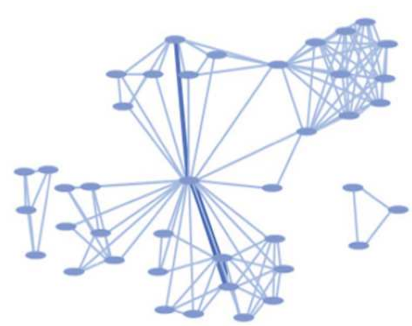
Components

- Connected component of a graph: a subset of the nodes such that
  - (i) every node in the subset has a path to every other; and
  - (ii) the subset is not part of some larger set with the property that every node can reach every other
- Ex:
  - 3 components in figure
  - {F, G, H, J} not a component, violates part (ii)



Components

- Graph components: a first, global way of describing its structure
- Within a given component, there may be richer internal structure that is important
- Ex (figure):
  - a prominent node (gatekeeper) at the center (the largest connected component would break apart into three distinct components if this node were removed)
  - tightly-knit groups linked to this node but not to each other
  - More in details in next lecture

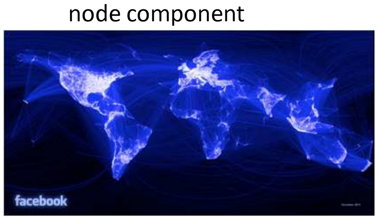


The collaboration (co-authorship) graph of the biological research center *Structural Genomics of Pathogenic Protozoa*

Giant Components

You are in one of many components

- Thought experiment
  - social network of the entire world
  - a link between two people if they know eachother
- Is this graph connected?
  - Presumably not (we don't have this graph)
  - a single person with no living friends would constitute a one-node component



Your component is giant

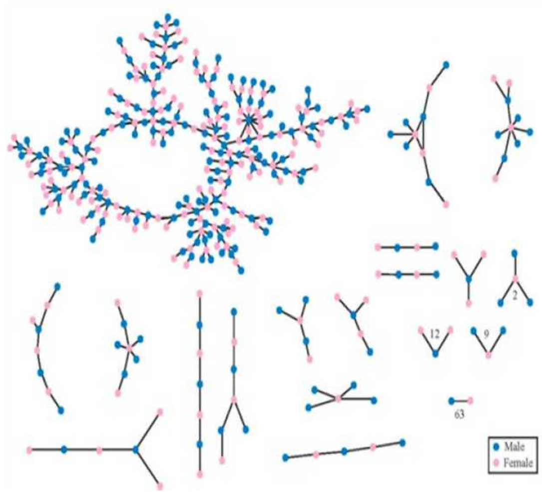
- You have friends in other countries
- Consider the parents of these friends, their friends, their friends etc.
- All these people in the same component
  - People who have never heard of you,
  - may well not share a language with you,
  - May have never traveled anywhere near where you live, and
  - may have had enormously different life experiences

Giant Components

- Large, complex networks often have what is called a giant component
  - contains a significant fraction of all the nodes
  - almost always contains only one giant component
    - All it would take is a single edge and the two giant components would merge
- Examples of merging:
  - half a millenium ago European explorers began arriving in America
    - technology and diseases of one quickly and disastrously overwhelmed the other

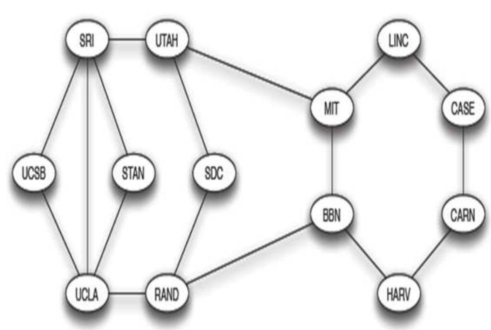
Giant Components

- Ex (figure):
  - Nodes are students in a large American high school
  - Edge joins two with romantic relationship at some point during a 18-month period (not all present at once)
- Graph contains large component
  - spread of sexually transmitted diseases
  - many paths of potential transmission



Distance

- In addition to asking whether two nodes are connected by a path, it is also interesting to ask how long such a path is
  - transportation, Internet communication, or the spread of news and diseases
- Length of a path:
  - number of steps it contains from beginning to end
  - Ex (figure):
    - (MIT, BBN, RAND, UCLA) has length 3
- Distance between two nodes:
  - the length of the shortest path between them
  - Ex (figure):
    - the distance between LINC and SRI is 3

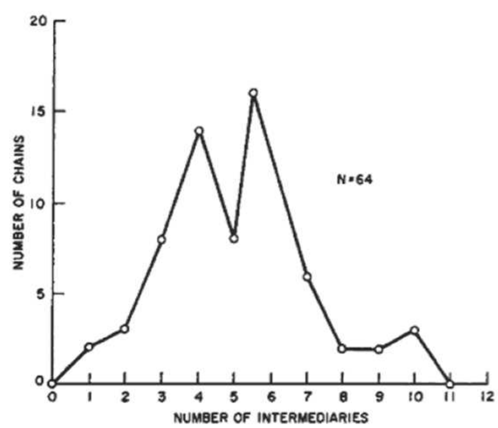


The Small-World Phenomenon

- Return to thought experiment (global social network)
  - not only do you have paths of friends connecting you to a large fraction of the world's population (giant component),
  - but these paths are surprisingly short (small distance)
  - Ex:
    - friends in another country, their parents, their friends
    - in three steps ended up in a different part of the world, in a different generation
- The small-world phenomenon:
  - the idea that the world looks "small" when you think of how short a path of friends it takes to get from you to almost anyone else
  - social networks tend to have very short paths between essentially arbitrary pairs of people

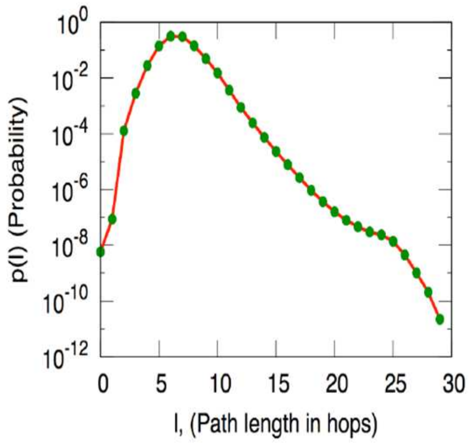
"Six degrees of separation"

- Experiment by by Stanley Milgram (1960)
  - Budget $680
  - 296 randomly chosen "starters"
- Figure: distribution of path lengths (64 chains succeeded)
- Caveats:
  - small set of people
  - only 64 made it
  - even if you can reach someone, is this useful to you?
    - But: potential speed with which information, diseases, and other kinds of contagion can spread through society

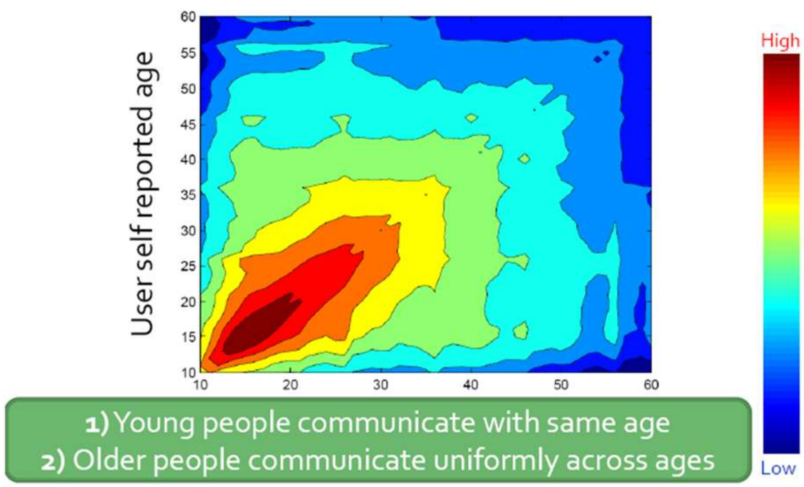


Instant Messaging

- Experiment by Leskovec and Horvitz
  - 240 million active user accounts on Microsoft Instant Messenger
  - edge between two users if they engaged in a two-way conversation at any point during a month-long observation period
  - Graph had a giant component
  - Distances within this giant component were very small (average distance of 6.6)

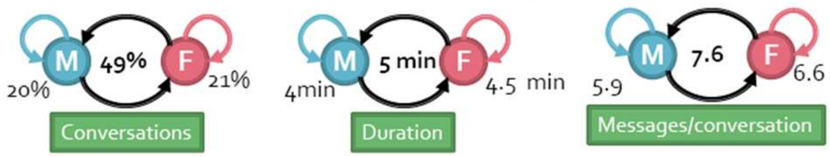


Instant Messaging

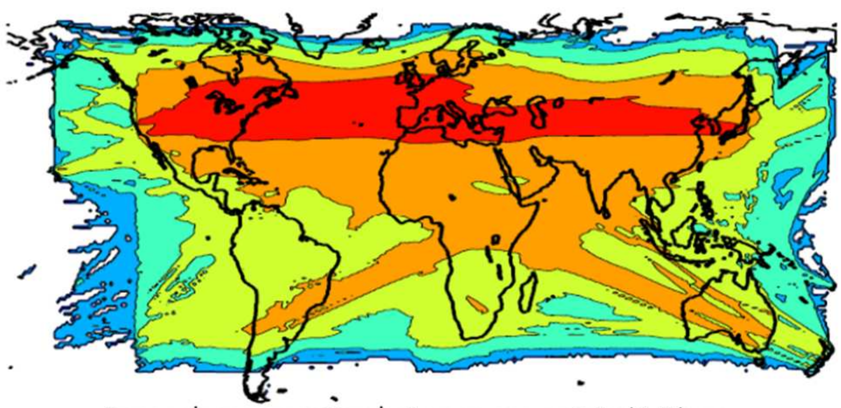


Instant Messaging

- Is gender communication biased?
  - **Homophily:** Do female talk more among themselves?
  - **Heterophily:** Do male-female conversations took longer?
- **Findings:**
  - Num. of. conversations is not biased (follows chance)
  - Cross-gender conversations take longer and are more intense (more attention invested)

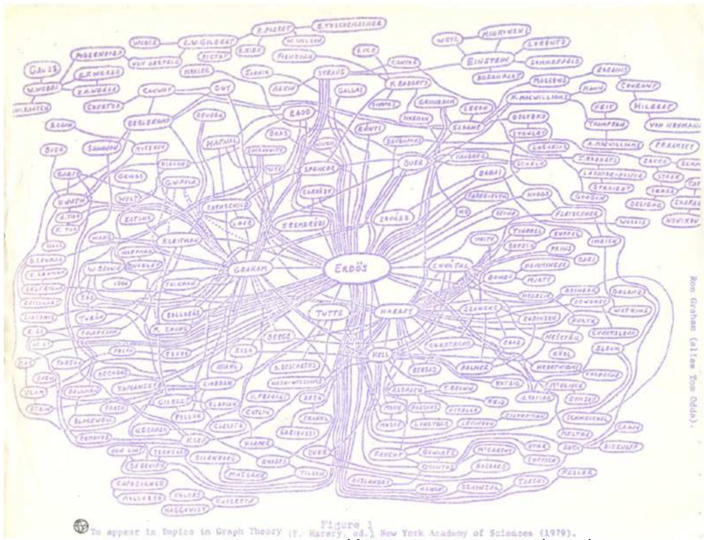


Instant Messaging



For each conversation between geo points (A,B) we increase the intensity on the line between A and B

Erdös Number



http://www.oakland.edu/enp/cgraph.jpg