CSC200: Lecture 2

Allan Borodin

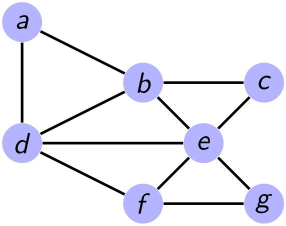
Networks and mathematics used in this course

- Two main mathematical subjects of primary relevance to this course:
  1. graph theoretic concepts
  2. probability

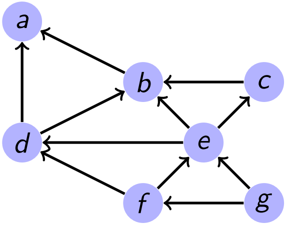
- In motivating the course, we have seen a number of examples of applications of graphs (also commonly referred to as networks) and hinted at some basic graph-theoretic concepts. We will continue that discussion (i.e. material from Chapter 2 of the text) now and for part of the next lecture before moving on to Chapter 3 in the next lecture.

Graphs: come in two varieties

1. undirected graphs (graph usually means an undirected graph.)



2. directed graphs (often called di-graphs).



Visualizing Networks as Graphs

- nodes: entities (people, countries, companies, organizations, ...)
- edges (may directed or weighted): relationship between entities
  - friendship, classmates, did business together, viewed the same web pages, ...
  - membership in a club, class, political party, ...

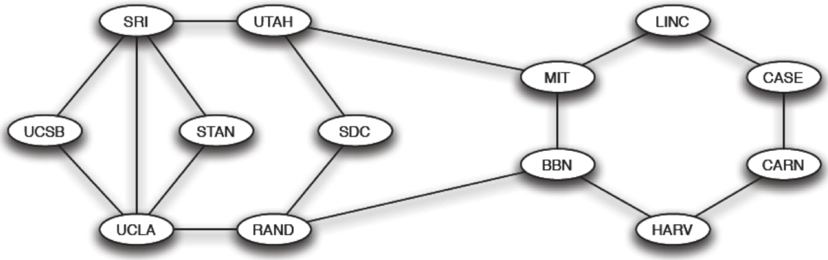


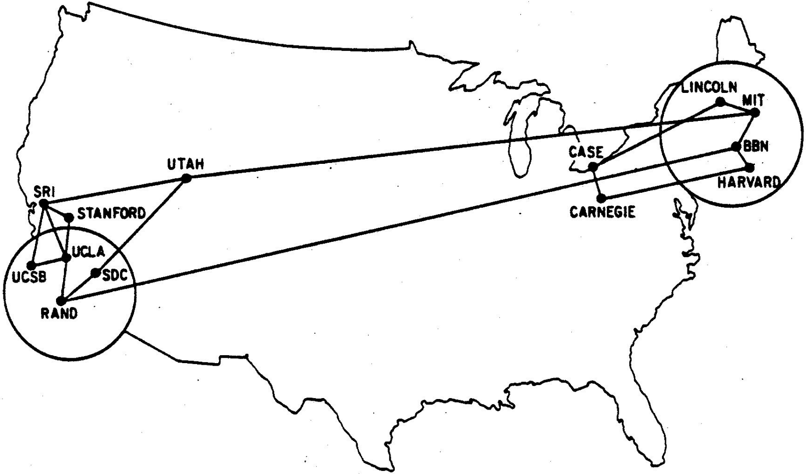
Figure : Internet: Dec. 1970 [E&K, Ch.2]

Adjacency matrix for graph induced by eastern sites (in alphabetical order) in 1970 internet graph: another way to represent a graph)

$$A(G) = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

- This node induced subgraph is a 6 node regular graph of degree 2. It is a simple graph in that there are no self-loops or multiple edges.
- Note that the adjacency matrix of an (undirected) simple graph is a symmetric matrix (i.e. $A_{i,j} = A_{j,i}$) with $\{0,1\}$ entries.
- To specify distances, we would need to give weights to the edges to represent the distances.

December 1970 internet visualized geographically [Heart et al 1978]



Kidney Exchange: Swap Chains

- Waiting list for kidney donation: 80K in US, 3K in Canada
- Live kidney donation common in N.A. to get around waiting list problems: requires donor-recipient pairs
- Exchange: supports willing pairs who are incompatible
  1. allows multiway-exchange
  2. supported by sophisticated algorithms to find matches

Kidney Exchange: Swap Chains

- Waiting list for kidney donation: 80K in US, 3K in Canada
- Live kidney donation common in N.A. to get around waiting list problems: requires donor-recipient pairs
- Exchange: supports willing pairs who are incompatible
  1. allows multiway-exchange
  2. supported by sophisticated algorithms to find matches
- But what if someone reneges? $\Rightarrow$ require simultaneous transplantation!

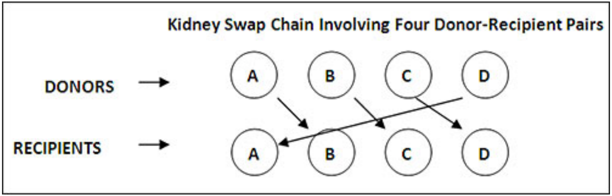
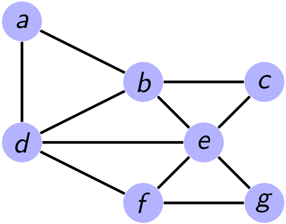
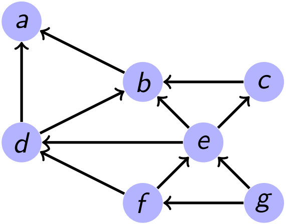


Figure : Dartmouth-Hitchcock Medical Center, NH, 2010

Recall: undirected graphs vs. directed graphs





More definitions and terminology

- In order to refer to the nodes and edges of a graph, we define graph $G = (V, E)$, where
  - V is the set of nodes (often called vertices)
  - E is the set of edges (sometimes called links or arcs)

More definitions and terminology

- In order to refer to the nodes and edges of a graph, we define graph $G = (V, E)$, where
  - V is the set of nodes (often called vertices)
  - E is the set of edges (sometimes called links or arcs)

- Undirected graph: an edge (u, v) is an unordered pair of nodes.

More definitions and terminology

- In order to refer to the nodes and edges of a graph, we define graph $G = (V, E)$, where
  - V is the set of nodes (often called vertices)
  - E is the set of edges (sometimes called links or arcs)

- Undirected graph: an edge (u, v) is an unordered pair of nodes.

- Directed graph: an edge (u, v) is an ordered pair of nodes $\langle u, v \rangle$.
  - However, we usually know when we have a directed graph and just write (u, v).

Basic definitions continued

- First start with undirected graphs $G = (V,E)$.

Basic definitions continued

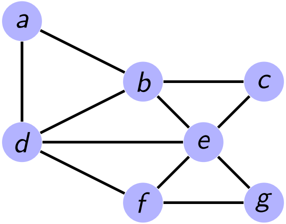
- First start with undirected graphs G = (V,E).
- A path between two nodes, say u and v is a sequence of nodes, say $u_1, u_2, \ldots, u_k$, where for every $1 \leq i \leq k-1$,
  - the pair (u_i, u_{i+1}) is an edge in E,
  - $u = u_1$ and $v = u_k$

Basic definitions continued

- First start with undirected graphs G = (V,E).
- A path between two nodes, say u and v is a sequence of nodes, say $u_1, u_2, \ldots, u_k$, where for every $1 \leq i \leq k-1$,
  - the pair (u_i, u_{i+1}) is an edge in E,
  - $u = u_1$ and $v = u_k$
- The length of a path is the number of edges on that path.

Basic definitions continued

- First start with undirected graphs G = (V,E).
- A path between two nodes, say u and v is a sequence of nodes, say $u_1, u_2, \ldots, u_k$, where for every $1 \leq i \leq k-1$,
  - the pair (u_i, u_{i+1}) is an edge in E,
  - $u = u_1$ and $v = u_k$
- The length of a path is the number of edges on that path.
- A graph is a connected if there is a path between every pair of nodes. For example, the following graph is connected.



Romantic Relationships [Bearman et al, 2004]

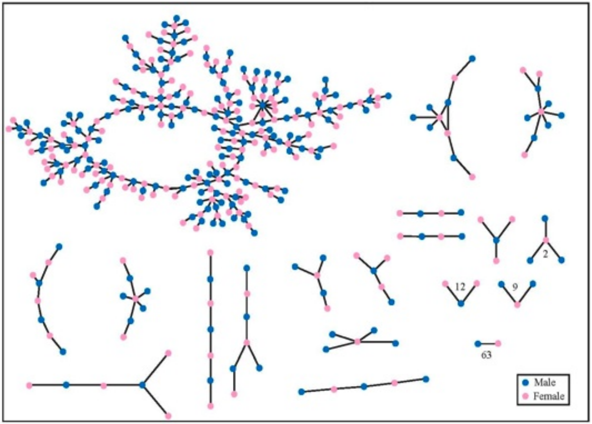
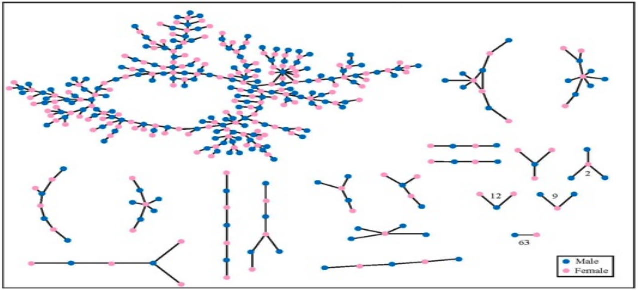


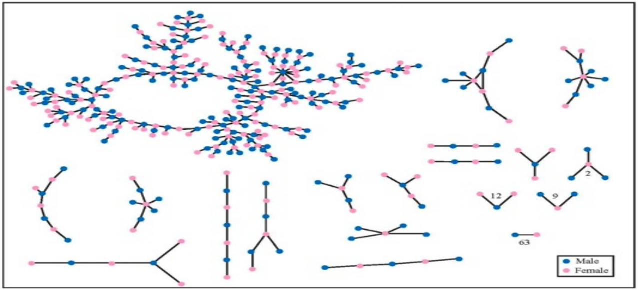
Figure : Dating network in US high school over 18 months.

- Illustrates common structural properties of many networks
- What predictions could you use this for?

More basic definitions



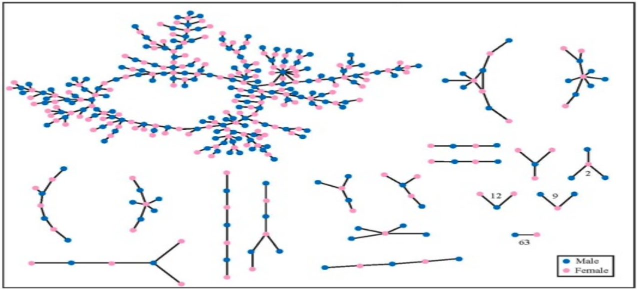
More basic definitions



Observation

Many connected components including one "giant component"

More basic definitions



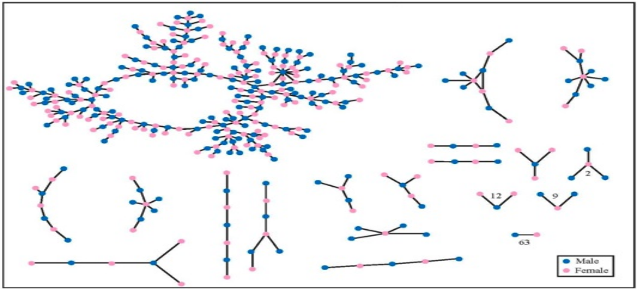
Observation

Many connected components including one "giant component"

- We will use this same graph to illustrate some other basic concepts.
- A cycle is path $u_1, u_2, \ldots, u_k$ such that $u_1 = u_k$; that is, the path starts and ends at the same node.

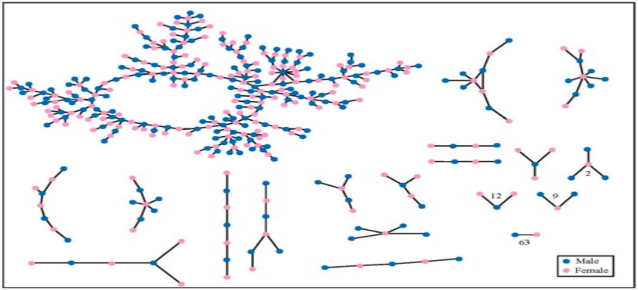
Simple paths and simple cycles

- Usually only consider simple paths and simple cycles: no repeated nodes (other than the start and end nodes in a simple cycle.)



Simple paths and simple cycles

- Usually only consider simple paths and simple cycles: no repeated nodes (other than the start and end nodes in a simple cycle.)



Observation

- There is one big simple cycle and (as far as I can see) three small simple cycles in the "giant component".
- Only one other connected component has a cycle: a triangle having three nodes. Note: this graph is "almost" bipartite and "almost" acyclic.

Example of an acyclic bipartite graph

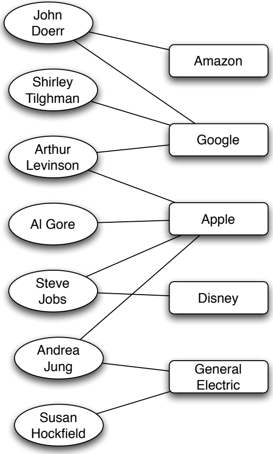


Figure : [E&K, Fig 4.4] One type of affiliation network that has been widely studied is the memberships of people on corporate boards of directors. A very small portion of this network (as of mid-2009) is shown here.

Florentine marriages and shortest paths

- Medici connected to more families, but not by much
- More importantly: lie between most pairs of families
  - shortest paths between two families: coordination, communication
  - Medici lie on 52% of all shortest paths; Guadagni 25%; Strozzi 10%

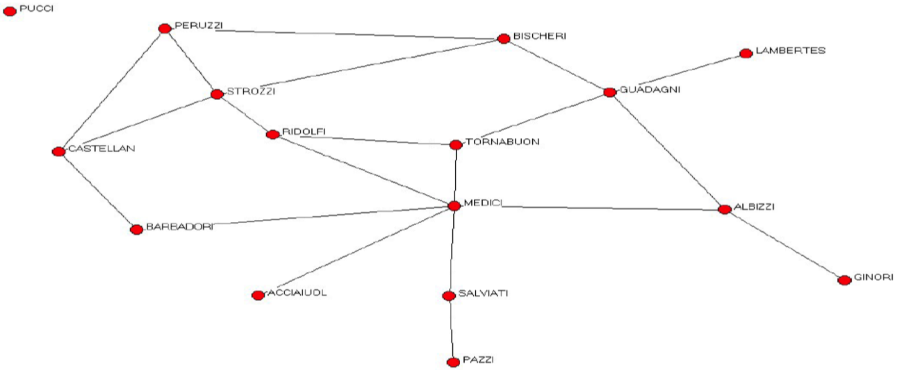


Figure : see [Jackson, Ch 1]

Breadth first search and path lengths [E&K, Fig 2.8]

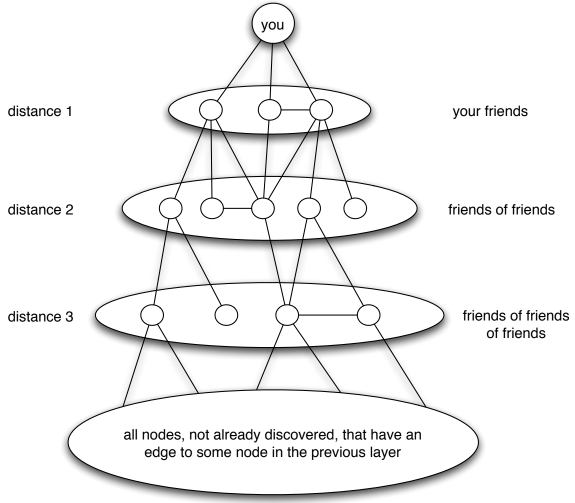


Figure : Breadth-first search discovers distances to nodes one layer at a time. Each layer is built of nodes adjacent to at least one node in the previous layer.

The Small World Phenomena

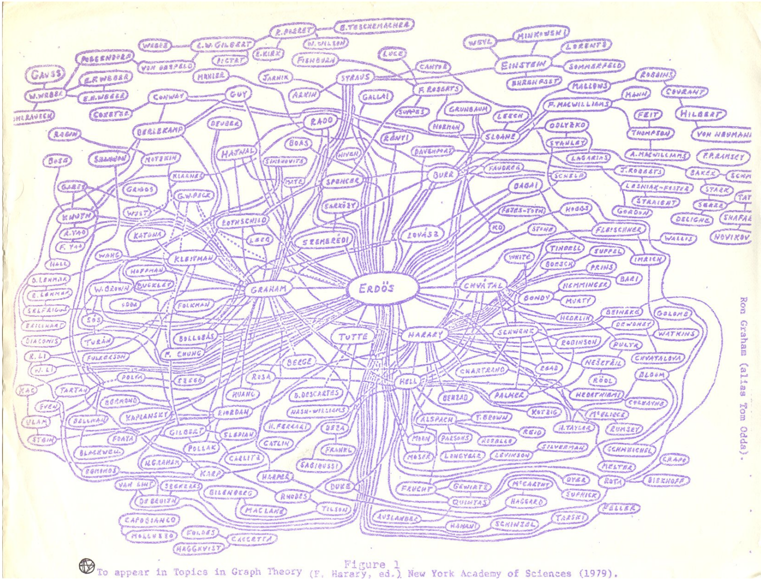
The small world phenomena suggests that in a connected social network any two individuals are likely to be connected (i.e. know each other indirectly) by a short path.

Later in the course we will study 1967 Milgram's small world experiment where he asked random people in Omaha Nebraska to forward a letter to a specified individual in a suburn of Boston which became the origin of the idea of six degrees of separation.

Small Collaboration Worlds

For now let us just consider collaboration networks like that of mathematicians or actors. For mathematicians (or more generally say scientists) we can form the collaboration network in which an edge represents co-authorhsip on a published paper. For actors, we can form a collaboration network where an edge represents actors perfoming in the same movie. For mathematicians one considers their Erdos number which is the length of the shortest path ito Paul Erdos. For actors, a popular notion is ones Bacon number, the shortest path to Kevin Bacon.

Erdos collaboration graph drawn by Ron Graham [http:/www.oakland.edu/enp/cgraph.jpg]



Analogous concepts for directed graphs

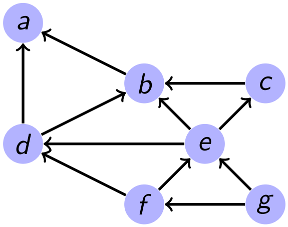
- We use the same notation for directed graphs, i.e. denoting a di-graph as $G = (V, E)$, where now the edges in E are directed.

Analogous concepts for directed graphs

- We use the same notation for directed graphs, i.e. denoting a di-graph as $G = (V, E)$, where now the edges in E are directed.
- Formally, an edge $\langle u, v \rangle \in E$ is now an ordered pair in contrast to an undirected edge (u, v) which is unordered pair.
  - However, it is usually clear from context if we are discussing undirected or directed graphs and in both cases most people just write (u, v).

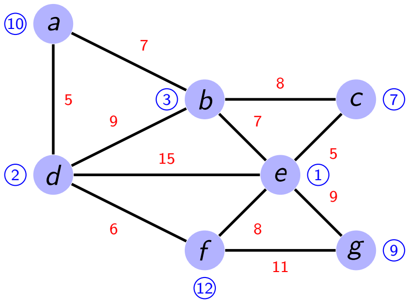
Analogous concepts for directed graphs

- We use the same notation for directed graphs, i.e. denoting a di-graph as $G = (V, E)$, where now the edges in E are directed.
- Formally, an edge $\langle u, v \rangle \in E$ is now an ordered pair in contrast to an undirected edge (u, v) which is unordered pair.
  - However, it is usually clear from context if we are discussing undirected or directed graphs and in both cases most people just write (u, v).
- We now have directed paths and directed cycles. Instead of connected components, we have strongly connected components.



Weighted graphs

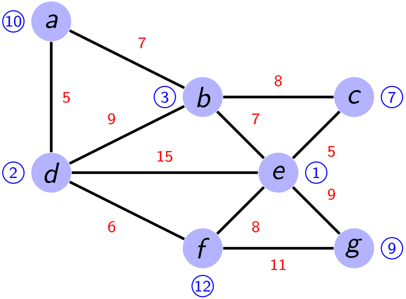
- We will often consider weighted graphs. Lets consider a (directed or undirected) graph $G = (V, E)$. Example:



  - red numbers: edge weights
  - blue numbers: vertex weights

Weighted graphs

- We will often consider weighted graphs. Lets consider a (directed or undirected) graph $G = (V, E)$. Example:

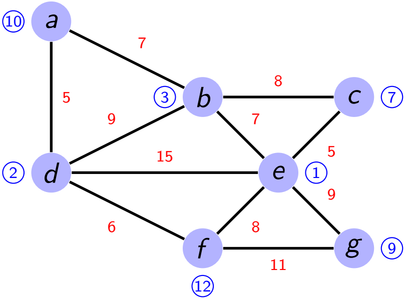


  - red numbers: edge weights
  - blue numbers: vertex weights

- We can have a weight $w(v)$ for each node $v \in V$ and/or a weight $w(e)$ for each edge $e \in E$.

Weighted graphs

- We will often consider weighted graphs. Lets consider a (directed or undirected) graph $G = (V, E)$. Example:

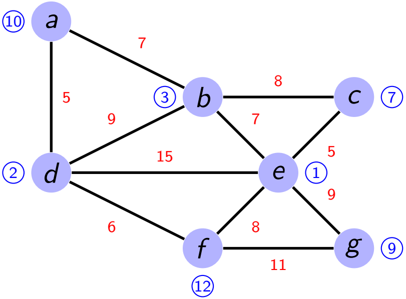


  - red numbers: edge weights
  - blue numbers: vertex weights

- We can have a weight $w(v)$ for each node $v \in V$ and/or a weight $w(e)$ for each edge $e \in E$.
- For example, in a social network whose nodes represent people, the weight $w(v)$ of node v might indicate the importance of this person.

Weighted graphs

- We will often consider weighted graphs. Lets consider a (directed or undirected) graph $G = (V, E)$. Example:



  - red numbers: edge weights
  - blue numbers: vertex weights

- We can have a weight $w(v)$ for each node $v \in V$ and/or a weight $w(e)$ for each edge $e \in E$.
- For example, in a social network whose nodes represent people, the weight $w(v)$ of node v might indicate the importance of this person.
- The weight $w(e)$ of edge e might reflect the strength of a friendship.

Edge weighted graphs

- When considering edge weighted graphs, we often have edge weights $w(e) = w(u, v)$ which are non negative (with $w(e) = 0$ meaning no edge).
- In some cases, weights can be either positive or negative. A positive (resp. negative) weight reflects the intensity of connection (resp. repulsion) between two nodes (with $w(e) = 0$ being a neutral relation).
- Sometimes (as in Chapter 3) we will only have a qualitative (rather than quantitative) weight, to reflect a strong or weak relation (tie).
- Analogous to shortest paths in an unweighted graph, we often wish to compute least cost paths, where the cost of a path is the sum of weights of edges in the path.