Wed. Jan 13th: Announcements & Corrections

- Currently seeking volunteer note takers (details on Quercus)
- Correction: Adjacency *Matrix* of a Graph, *not* Adjacency Graph

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

- Recording of Monday's lecture & tentative slides are up

Wed. Jan 13th: Announcements & Corrections

- Currently seeking volunteer note takers (details on Quercus)
- Correction: Adjacency *Matrix* of a Graph, *not* Adjacency Graph

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

- Recording of Monday's lecture & tentative slides are up
- Happy 24rth (29th?) Birthday Hal!

Today's agenda

- Last class we saw examples of various social & information networks and began a combination of review, and a sneak-preview of topics we'll be going into more detail later in the course.
- Today we'll finish this review & preview, and start motivations for the strength of edges

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

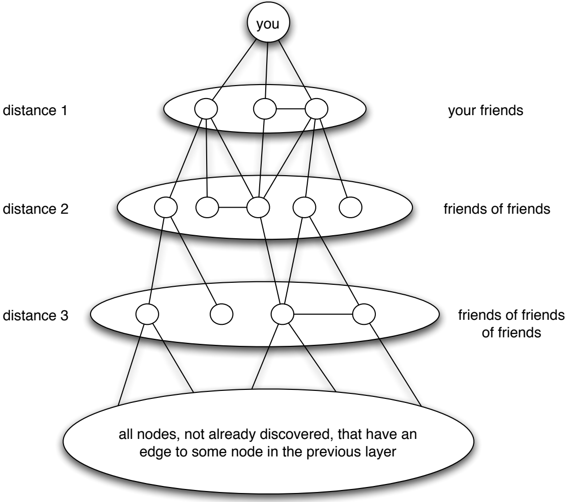


Figure: Breadth-first search discovers distances to nodes one layer at a time. Each layer is built of nodes adjacent to at least one node in the previous layer.

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 suburb 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 co-authorship on a published paper. For actors, we can form a collaboration network where an edge represents actors performing in the same movie. For mathematicians one considers their Erdos number which is the length of the shortest path to 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]

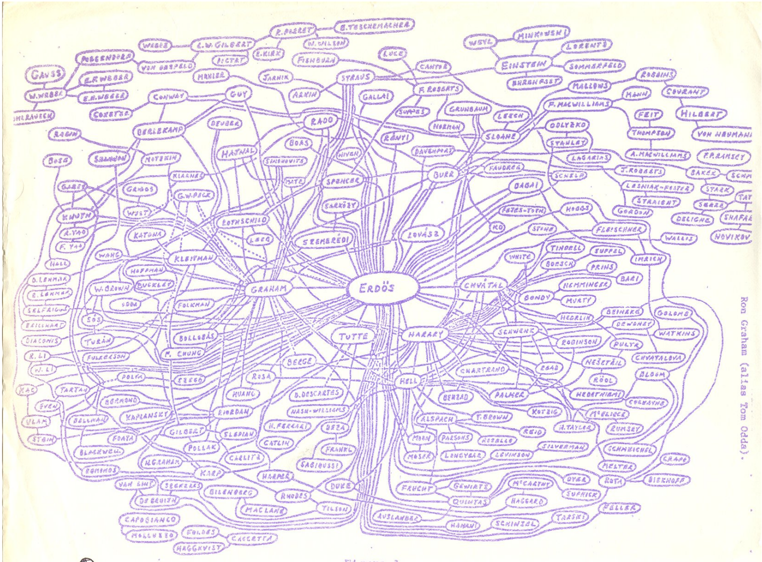


Figure 1
To appear in Topics in Graph Theory (F. Harary, ed.) New York Academy of Sciences (1979).

Analogous concepts for directed graphs

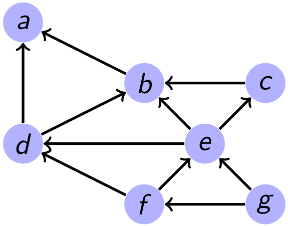
- We use the same notation for directed graphs, i.e. denoting a di-graph as $G = (V, E)$, where now the edges in E are directed.

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).

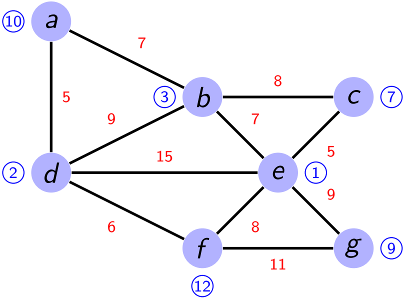
Analogous concepts for directed graphs

- 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

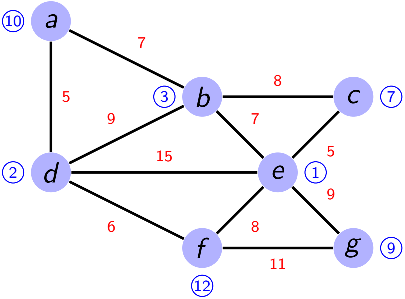
- We will often consider weighted graphs. Lets consider a (directed or undirected) graph $G = (V, E)$. Example:



  - 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:

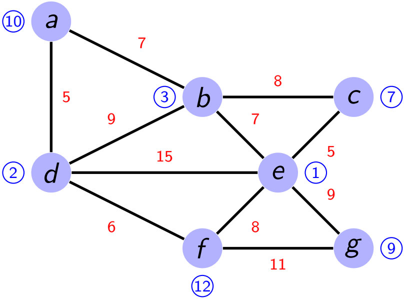


  - red numbers: edge weights
  - blue numbers: vertex weights

- 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:

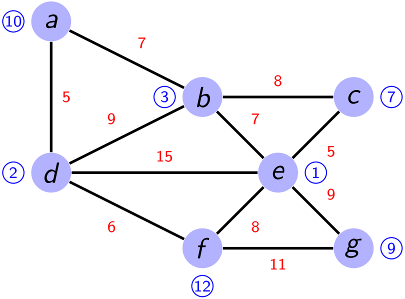


  - red numbers: edge weights
  - blue numbers: vertex weights

- 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

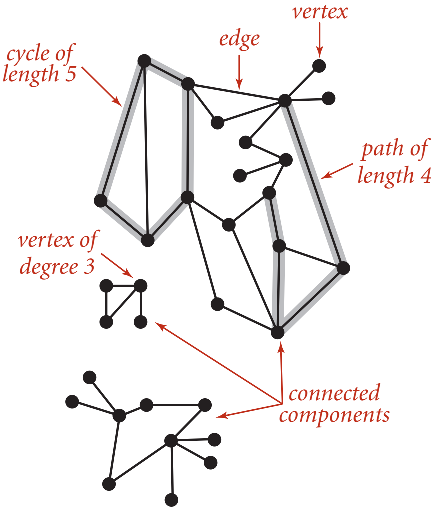
- When considering edge weighted graphs, we often have edge weights $w(e) = w(u, v)$ which are non negative (with $w(e) = 0$ or $w(e) = \infty$ meaning no edge depending on the context).

- 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.

Graph anatomy: summary thus far

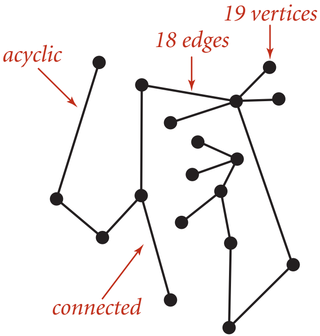


[from Algorithms, 4th Edition by Sedgewick and Wayne]

Acyclic graphs (forests)

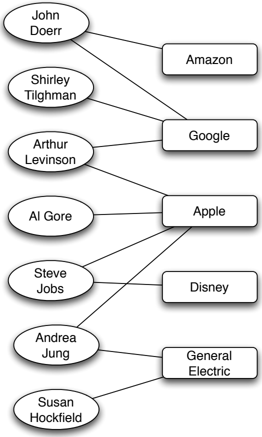
- A graph that has no cycles is called a forest.
- Each connected component of a forest is a tree.

  - A tree is a connected acyclic graph.
  - Question: Why are such graphs called trees?
  - Fact: There are always $n - 1$ edges in an n node tree.



- Thus, a forest is simply a collection of trees.

Another tree [E&K Figure 4.4]



- The bipartite graph from last class (depicting membership on corporate boards) is also an example of a tree.
- In general, bipartite graphs can have cycles.
- Question: is an acyclic graph always bipartite?

Facts

- It is computationally easy to decide if a graph is acyclic or bipartite.
- However, we (in CS) strongly "believe" it is not easy to determine if a graph is tripartite (i.e. 3-colourable).

Analogous concepts for directed graphs

- We now have directed paths and directed cycles.
- Instead of the degree of a node, we have the in-degree and out-degree of a node.

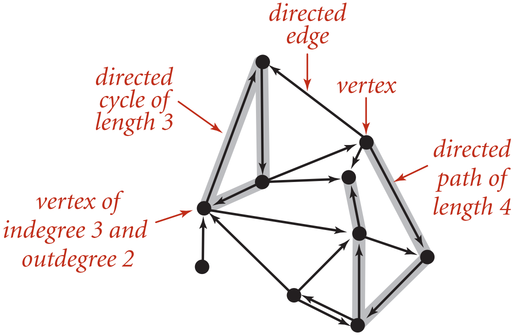
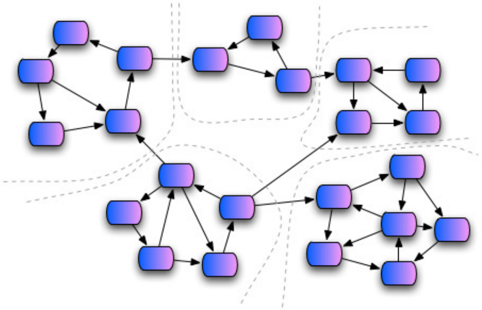


Figure: Directed graph antonomy [from Sedgewick and Wayne]

More analogous concepts for directed graphs

- Acyclic mean no directed cycles.
- Instead of connected components, we have strongly connected components.

[from `http://scientopia.org/blogs/goodmath/`]



- Instead of trees, we have directed (i.e. rooted) trees which have a unique root node with in-degree 0 and having a unique path from the root to every other node.
- Question: What is a natural example of a rooted tree?

Detecting the romantic relation in Facebook: Course motivation and a lead in to Chapter 3 of text

- There is an interesting paper by Backstrom and Kleinberg (http://arxiv.org/abs/1310.6753) on detecting "the" romantic relation in a subgraph of Facebook users who specify that they are in such a relationship.
- Backstrom and Kleinberg construct two datasets of randomly sampled Facebook users: (i) an extended data set consisting of 1.3 million users declaring a spouse or relationship partner, each with between 50 and 2000 friends and (ii) a smaller data set extracted from neighbourhoods of the above data set (used for the more computationally demanding experimental studies).
- The main experimental results are nearly identical for both data sets.
- Question: How would you go about identifying someone's spouse given their Facebook profile & feed?

Detecting the romantic relation (continued)

- They consider various "interaction features" including
  1. the number of photos in which both A and B appear.
  2. the number of profile views within the last 90 days.
- Their focus was various graph structural features of edges, including
  1. the *embeddedness* of an edge (A, B) which is the number of mutual friends of A and B.
  2. various forms of a new *dispersion* measure of an edge (A, B) where high dispersion intuitively means that the mutual neighbours of A and B are not "well-connected" to each other (in the graph without A and B).
  3. One definition of dispersion given in the paper is the number of pairs (s, t) of mutual friends of A and B such that $(s, t) \notin E$ and s, t have no common neighbours except for A and B.

Embeddedness and dispersion example from paper

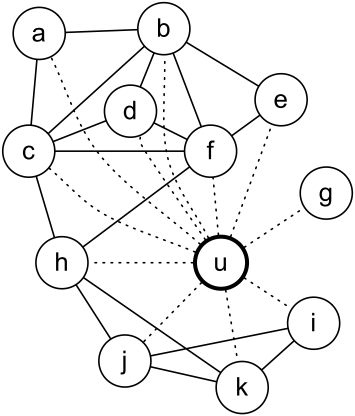


Figure 2. A synthetic example network neighborhood for a user u; the links from u to b, c, and f all have embeddedness 5 (the highest value in this neighborhood), whereas the link from u to h has an embeddedness of 4. On the other hand, nodes u and h are the unique pair of intermediaries from the nodes c and f to the nodes j and k; the u-h link has greater dispersion than the links from u to b, c, and f.

Qualitative results from Backstrom and Kleinberg

- The goal is to predict (for each user in the data set) which of their friendship edges is the romantic relation. Note that each user has between 50 and 2000 friends and assuming say a median of 200 friends, a random guess would have prediction accuracy of $1/200 = .5\%$

Qualitative results from Backstrom and Kleinberg

- The goal is to predict (for each user in the data set) which of their friendship edges is the romantic relation. Note that each user has between 50 and 2000 friends and assuming say a median of 200 friends, a random guess would have prediction accuracy of $1/200 = .5\%$
- Various disperson measures do better than the embeddedness measure in its ability to predict the correct romantic relationship. Why would high dispersion be a better measure than high embeddedness?