CSC 225

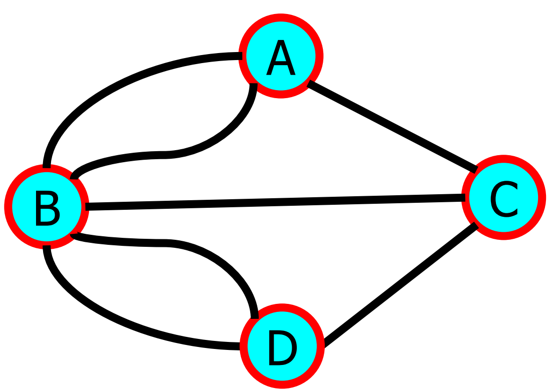
Algorithms and Data Structures I

Rich Little
rlittle@uvic.ca
ECS 516

Abstract Meaning of the Term Graph

- A ***graph*** $G = (V, E)$ is a set V of ***vertices*** (*nodes*) and a collection E of pairs from V, called ***edges*** (*arcs*).

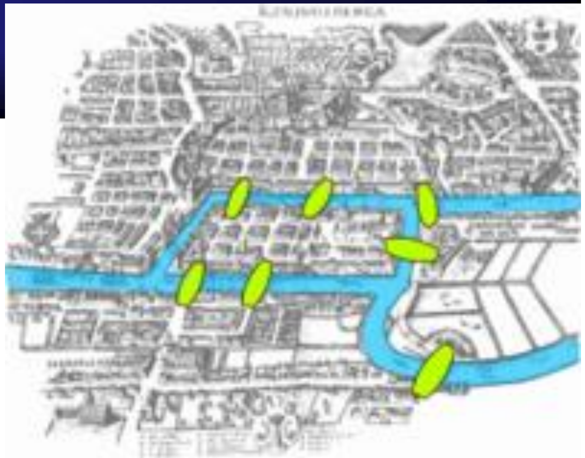
- **Graph Example:**



$$V = \{A, B, C, D\}$$

$$E = \left\{ \begin{array}{l} \{A, B\}, \{A, B\}, \{A, C\}, \\ \{B, C\}, \{B, D\}, \{B, D\}, \{C, D\} \end{array} \right\}$$

History

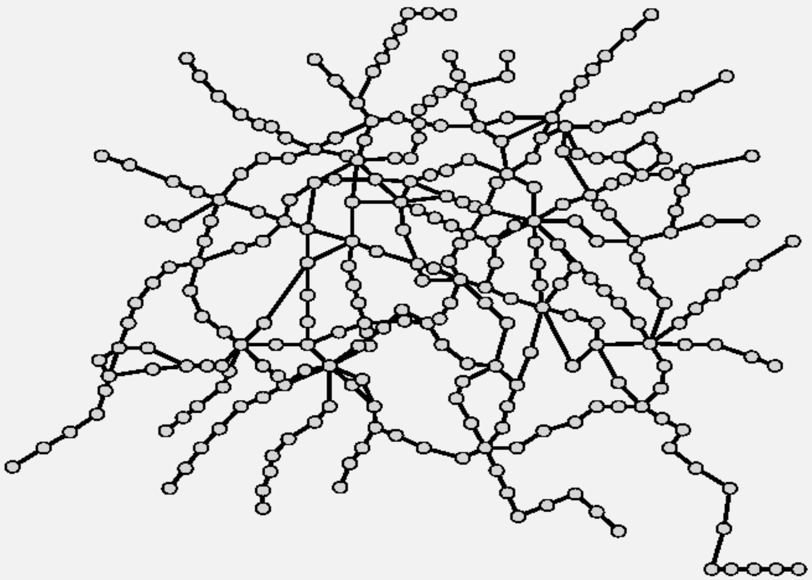


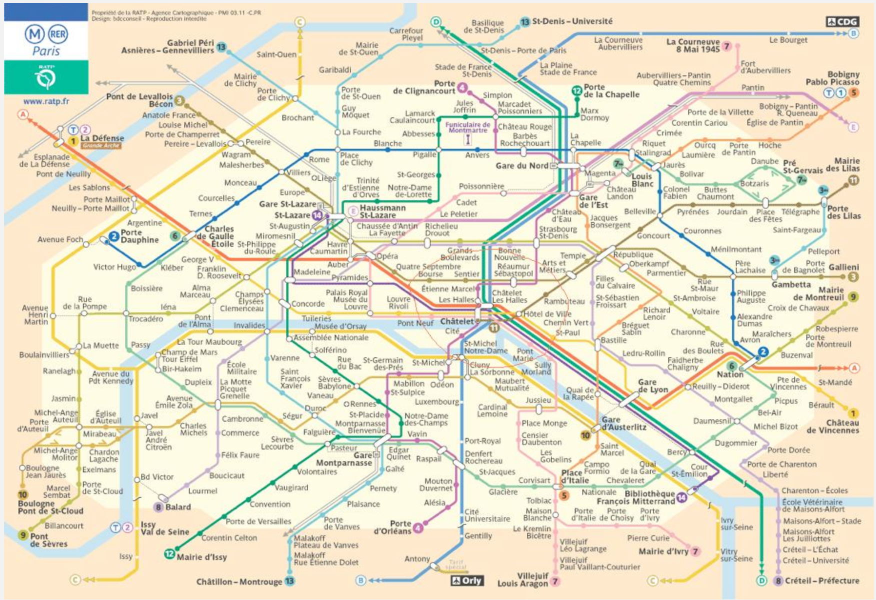
- 1736
  - Euler (Swiss)
  - Seven Bridges of Königsberg
- 1878
  - The term graph was introduced by Sylvester in a Nature article
- 1930
  - Graph theory evolved into an organized brand of Mathematics
- 1959
  - Dijkstra's Shortest Path Algorithm
- 1972
  - Tarjan and Hopcroft: depth first search; Turing Award 1986

http://en.wikipedia.org/wiki/Graph_theory

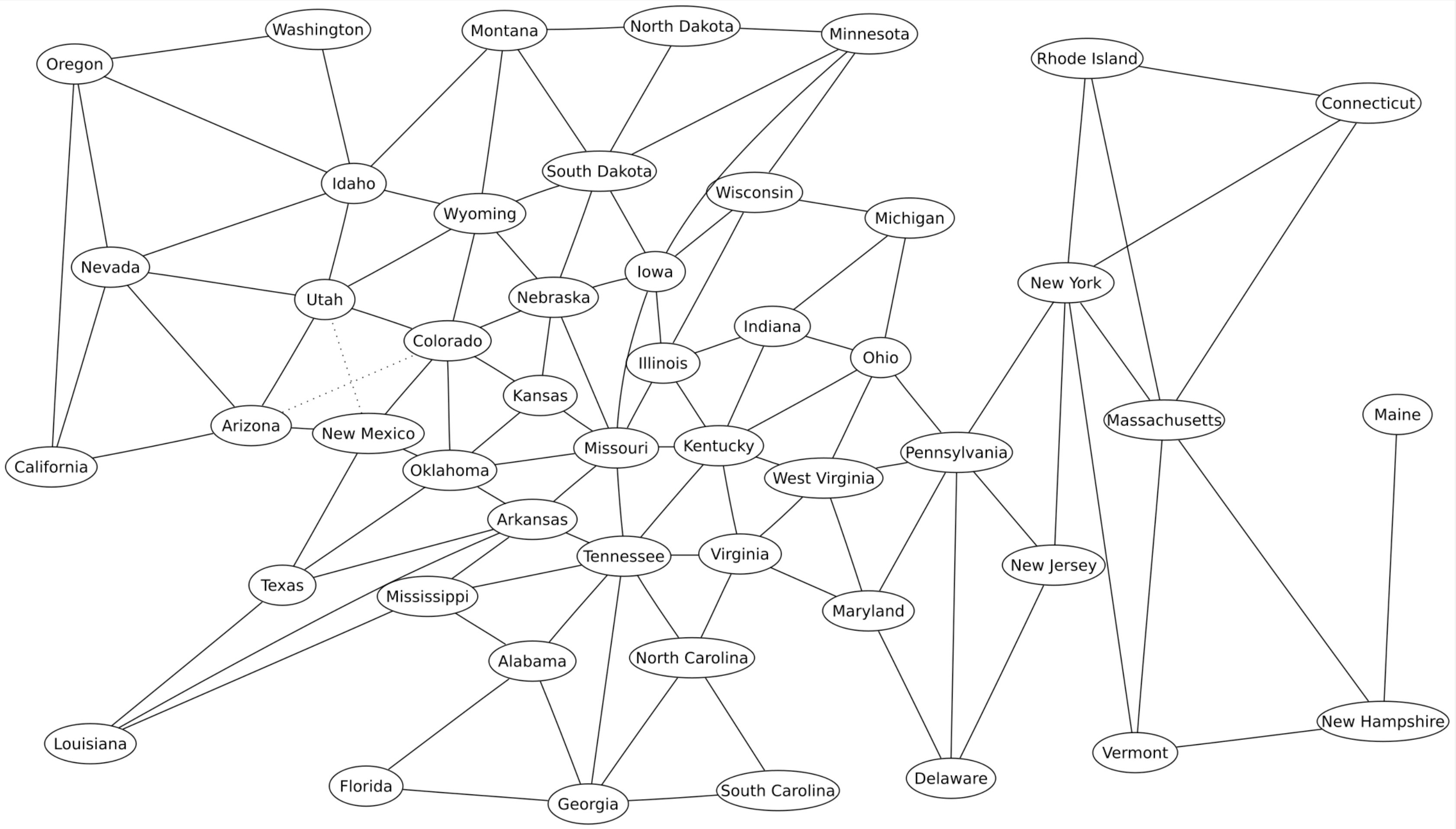
Why study graph algorithms?

- Thousands of practical applications.
- Hundreds of graph algorithms known.
- Interesting and broadly useful abstraction.
- Challenging branch of computer science and discrete math.

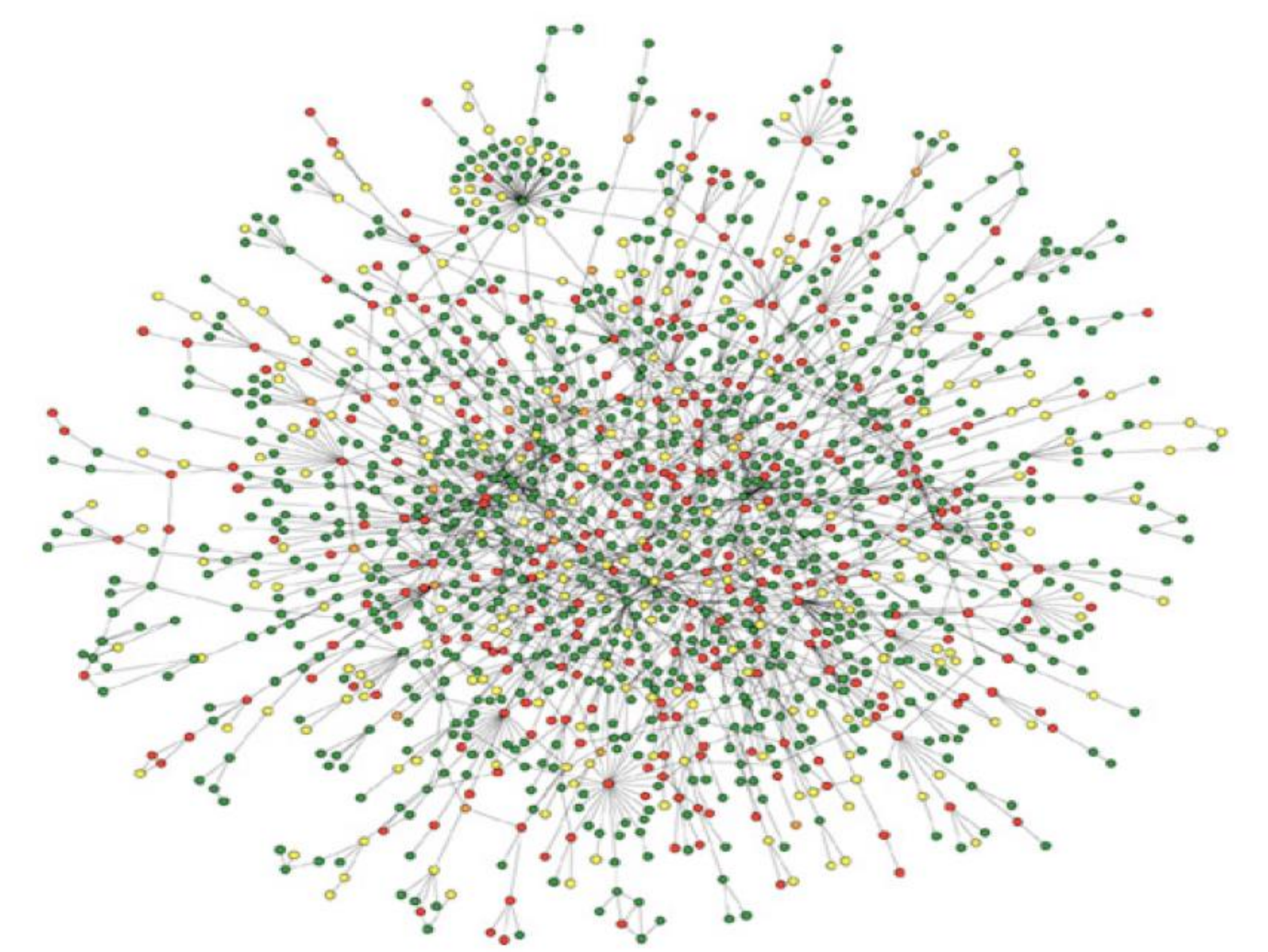




Border graph of 48 contiguous United States

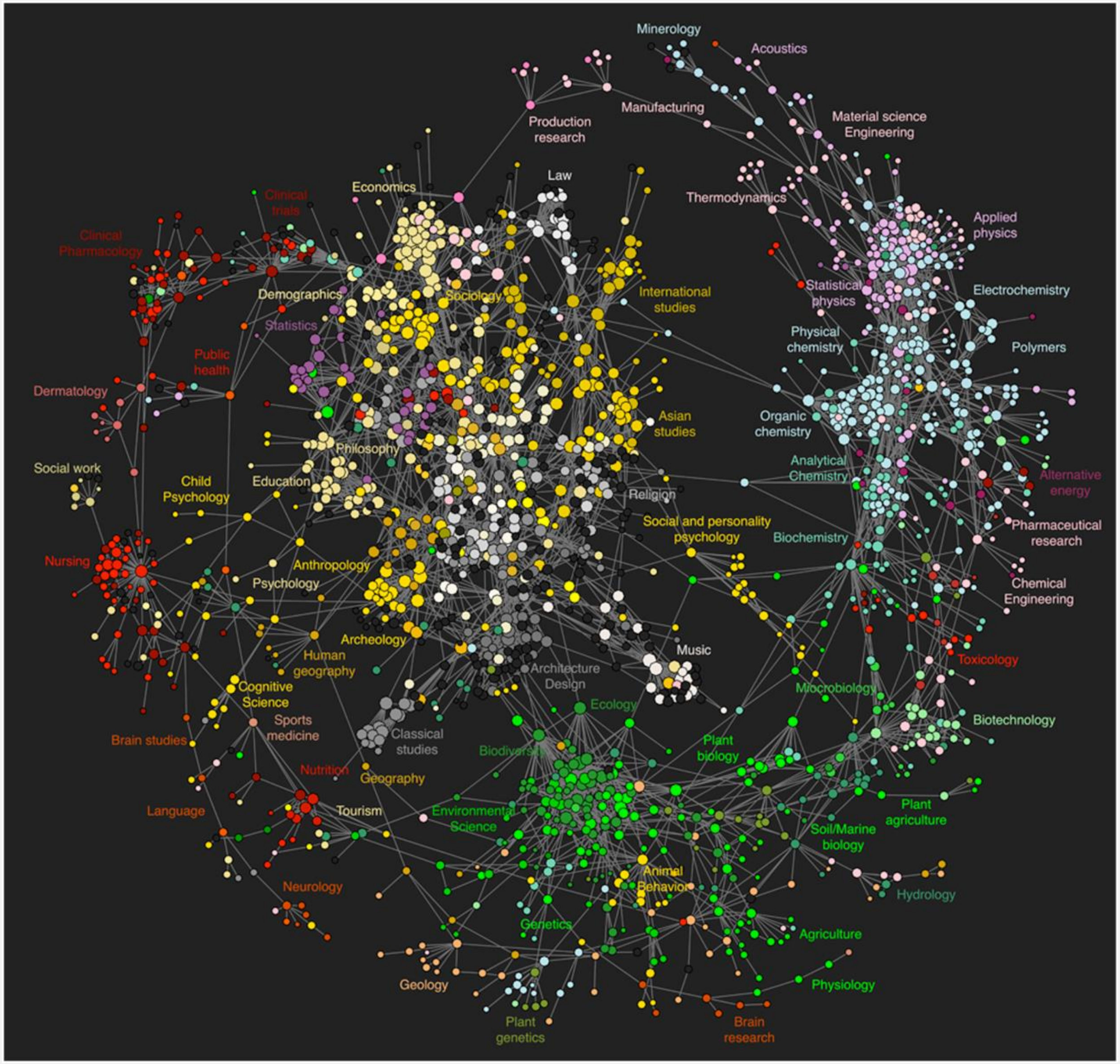


Protein-protein interaction network



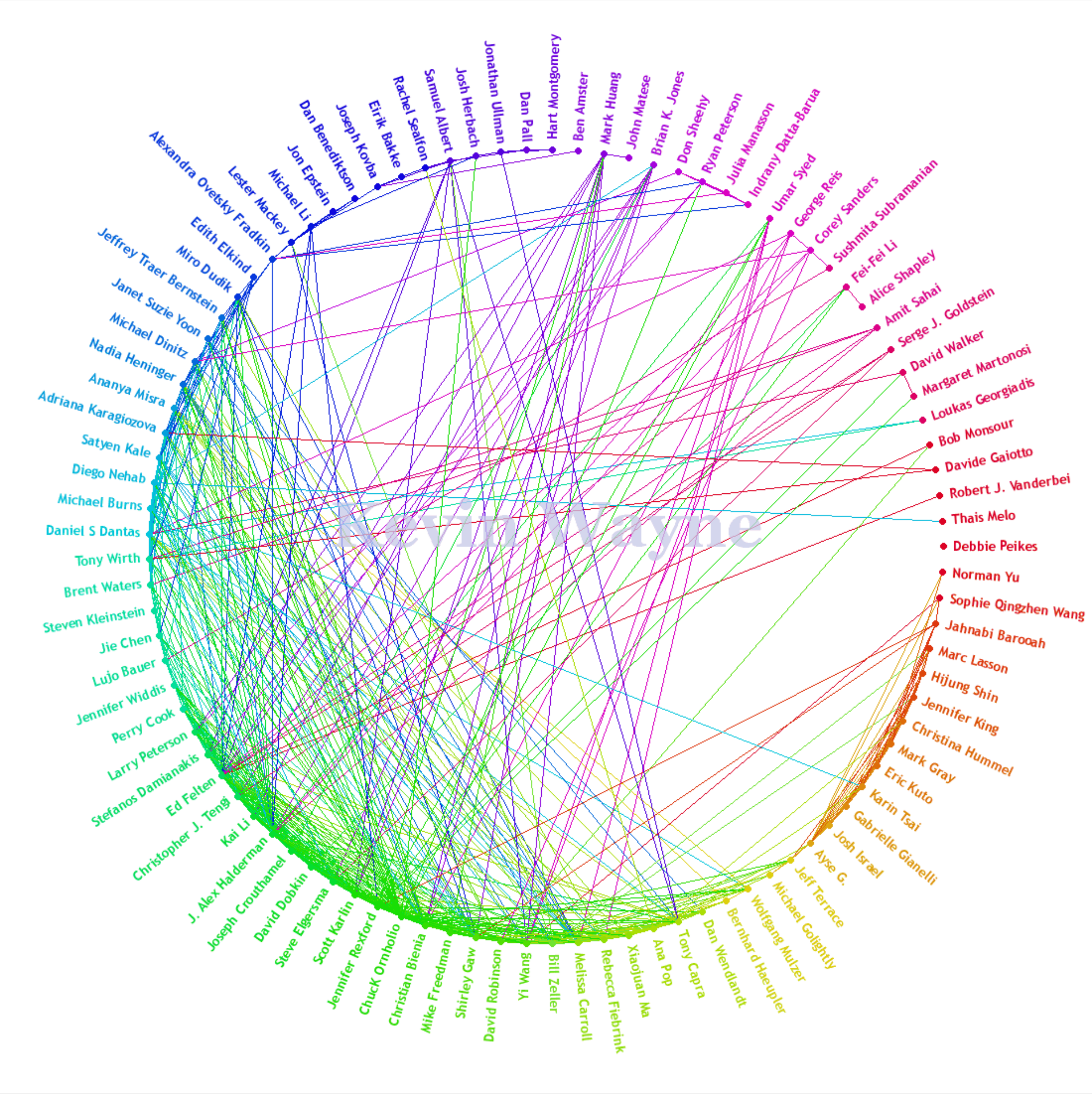
Reference: Jeong et al, Nature Review | Genetics

Map of science clickstreams



http://www.plosone.org/article/info:doi/10.1371/journal.pone.0004803

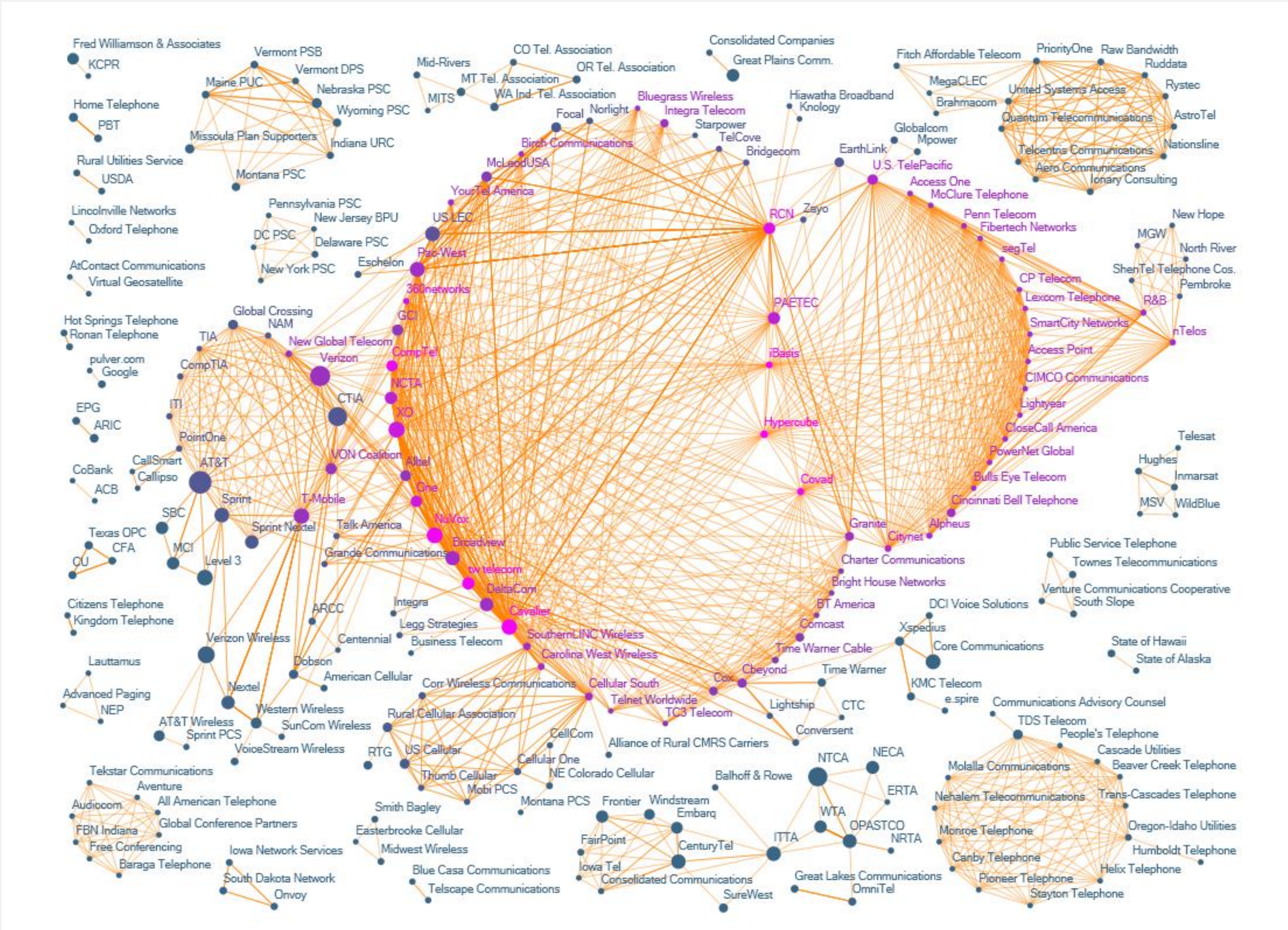
Kevin's facebook friends (Princeton network, circa 2005)



10 million Facebook friends

"Visualizing Friendships" by Paul Butler

The evolution of FCC lobbying coalitions



"The Evolution of FCC Lobbying Coalitions" by Pierre de Vries in JoSS Visualization Symposium 2010

Framingham heart study

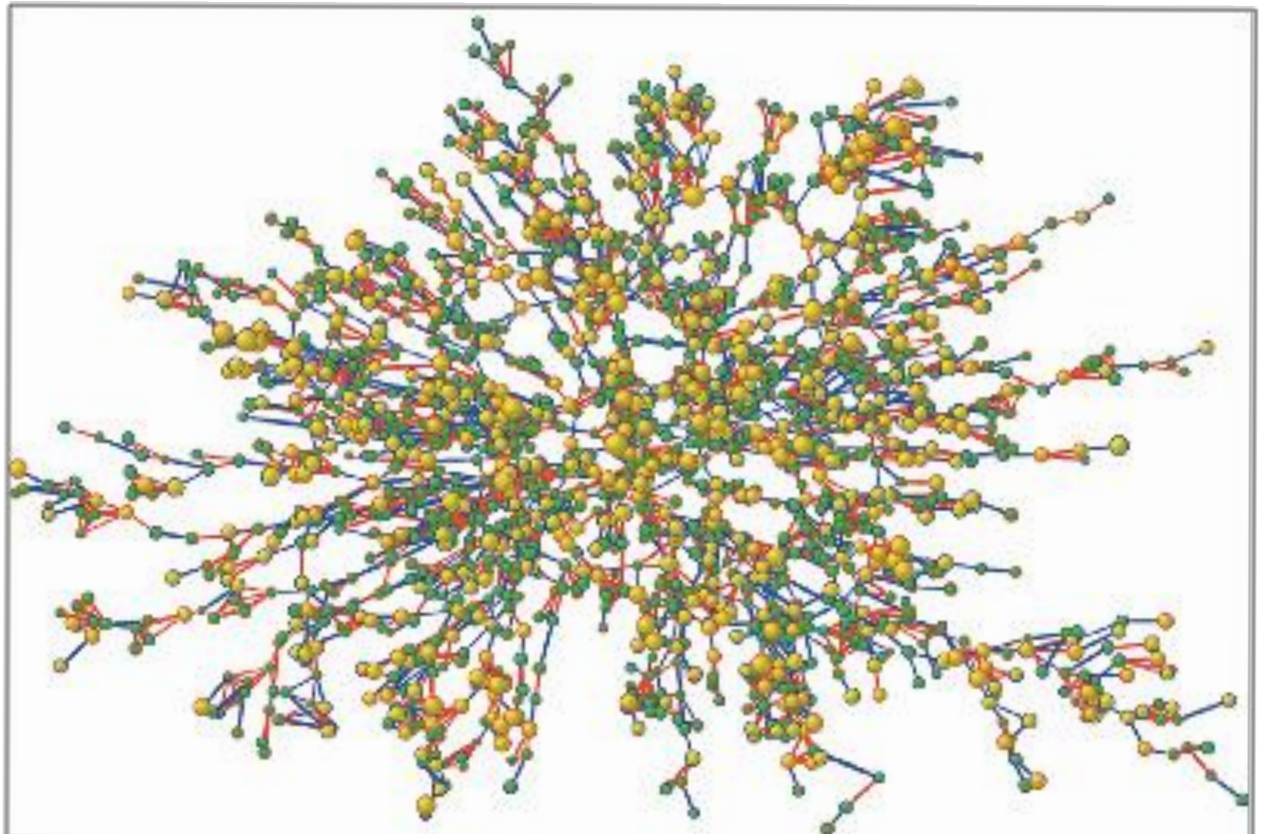
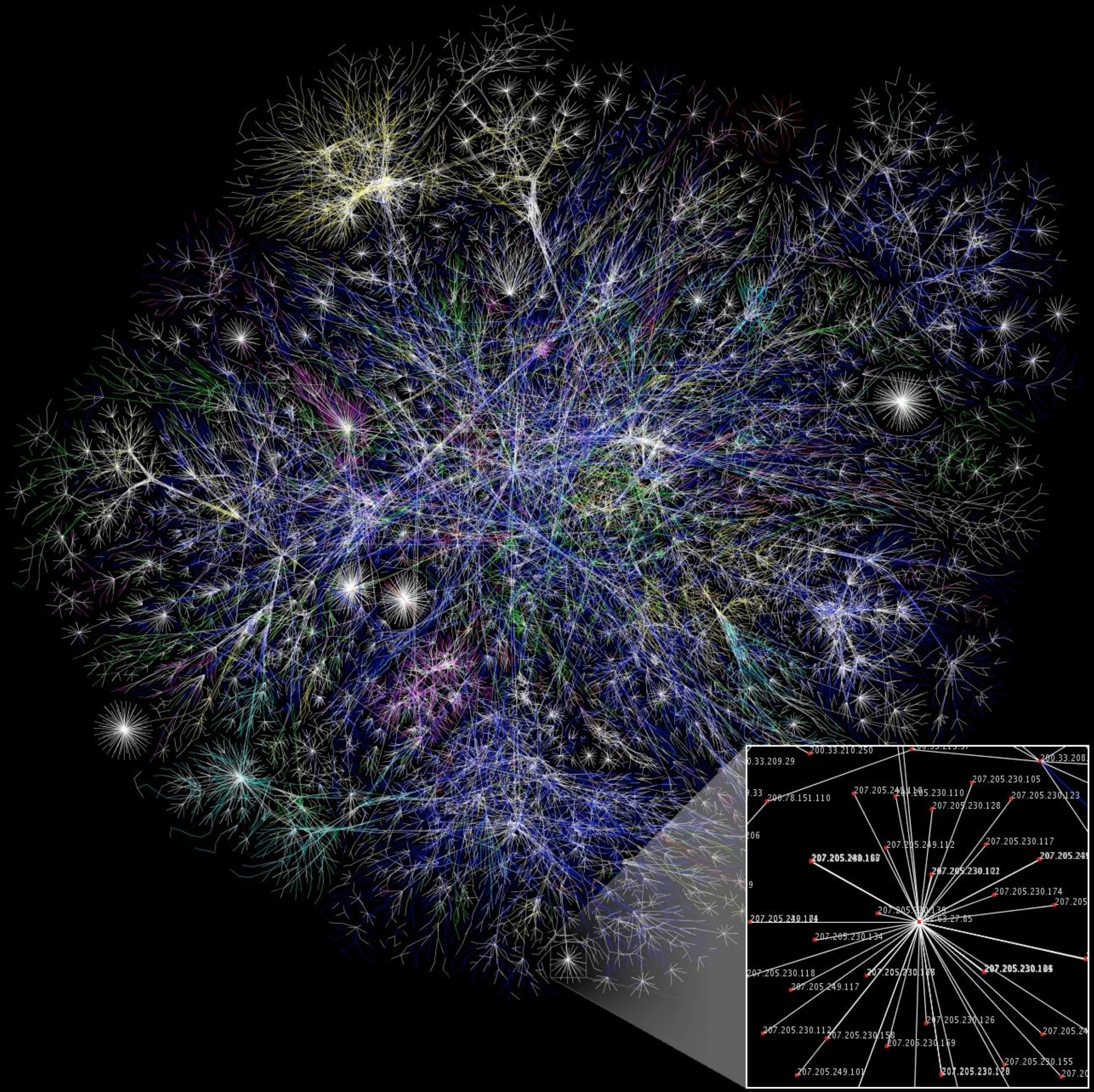


Figure 1. **Largest Connected Subcomponent of the Social Network in the Framingham Heart Study in the Year 2000.**
Each circle (node) represents one person in the data set. There are 2200 persons in this subcomponent of the social network. Circles with red borders denote women, and circles with blue borders denote men. The size of each circle is proportional to the person's body-mass index. The interior color of the circles indicates the person's obesity status: yellow denotes an obese person (body-mass index, ≥30) and green denotes a nonobese person. The colors of the ties between the nodes indicate the relationship between them: purple denotes a friendship or marital tie and orange denotes a familial tie.

"[The Spread of Obesity in a Large Social Network over 32 Years]" **by Christakis and Fowler in New England Journal of Medicine, 2007**



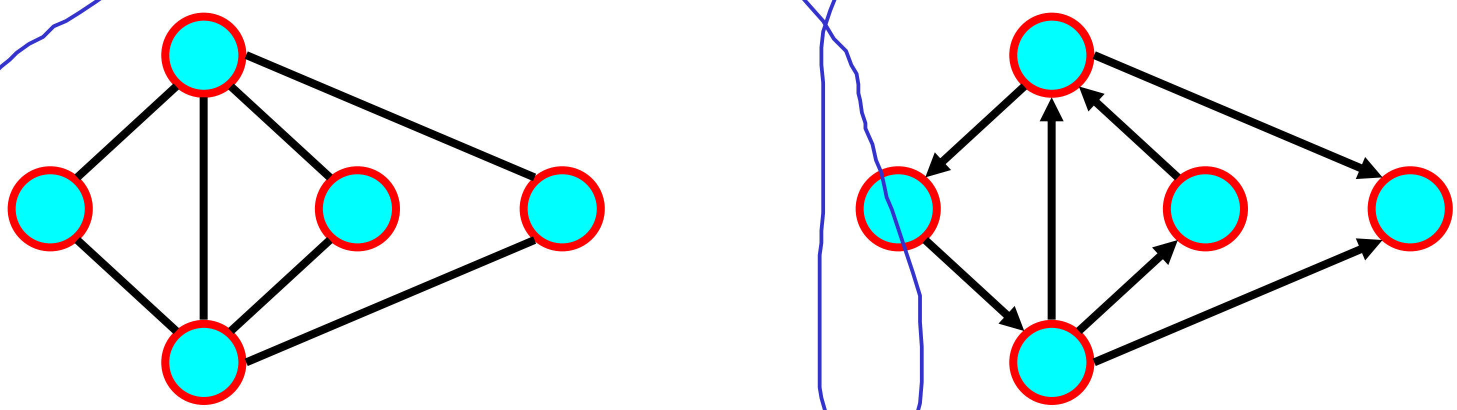
200.33.210.250
200.78.151.110
207.205.230.105
207.205.230.110
207.205.230.123
207.205.230.128
207.205.249.112
207.205.230.117
207.205.230.174
207.205.230.134
207.205.230.118
207.205.249.117
207.205.230.126
207.205.230.112
207.205.230.169
207.205.230.155
207.205.249.101

Different types and special cases of graphs

- *Directed* graphs *(Digraphs):* The edges of the graph are *directed*.
- *Undirected* graphs: The edges of the graph are *undirected*.
- *Simple* graphs
- *Complete* graphs
- *Connected* graphs
- *Acyclic* graphs
- *Bipartite* graphs
- *Weighted* graphs
- *Trees*

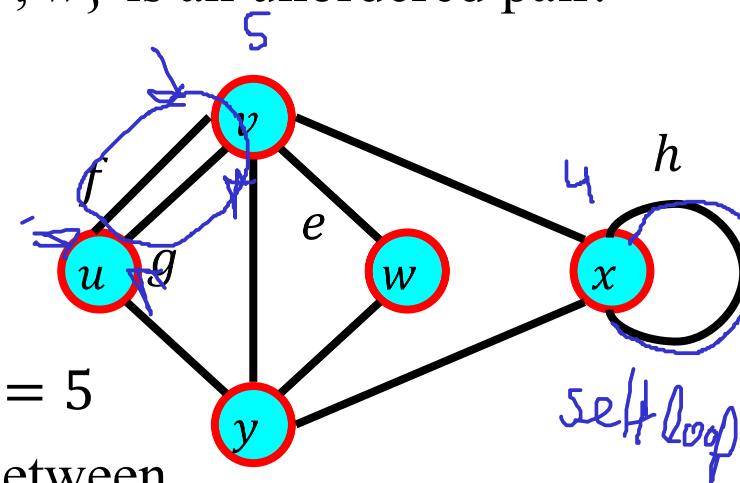
Graph Terminology

- Much of the terminology for graphs is applicable to undirected graphs and directed graphs



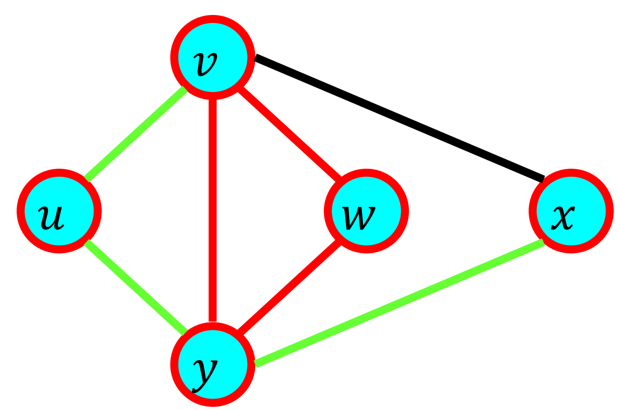
Undirected Edges

- An *undirected edge e* represents a *symmetric* relation between two vertices *v* and *w* represented by the vertices.
  - We usually write $e = \{v, w\}$, where $\{v, w\}$ is an unordered pair.
  - *v*, *w* are the *endpoints* of the edge
  - *v* is *adjacent* to *w*
  - *e* is *incident* upon *v* and *w*
  - The *degree* of a vertex is the number of incident edges, eg. $\deg(v) = 5$
  - *parallel* edges – more than one edge between a pair of vertices, eg. *f* and *g*
  - *self-loop* – edge that connects a vertex to itself, eg. *h*
  - Typically, the number of vertices is denoted by *n* and the number of edges by *m*.



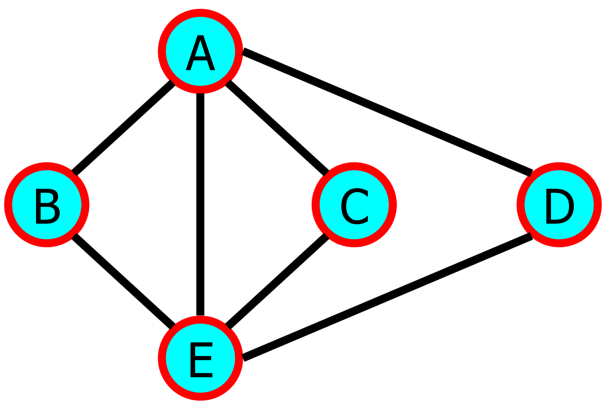
Undirected Paths

- A ***walk*** in a graph is a sequence of vertices $v_1, v_2, \ldots, v_n$ such that there exist edges $\{v_1, v_2\}$, $\{v_2, v_3\}$, …, $\{v_{n-1}, v_n\}$
- The ***length*** of a walk is the number of edges
  - if $v_1 = v_n$, ***closed***, otherwise ***open***
- If no edge is repeated, it's a ***trail***
- A closed trail is a ***circuit***
- If no vertex is repeated, it's a ***path***
- A ***cycle*** is a path with the same start and end vertices



Connected Graphs

- A graph is ***connected*** if every pair of vertices is connected by a path.
- Example
  - ➢ Two *connected components* of a graph
  - ➢ *Unconnected* graph

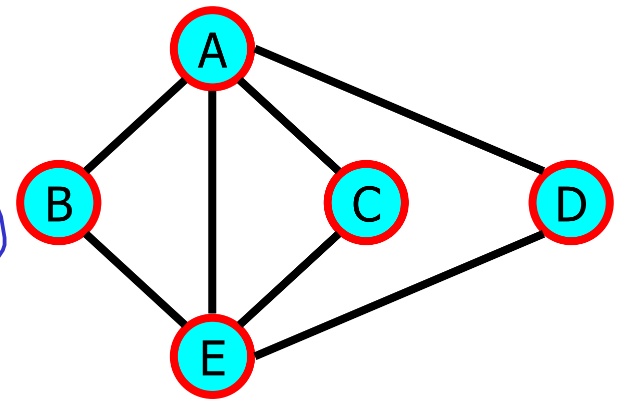


Simple Graphs

- A ***simple graph*** is a graph with no self-loops and no parallel or multi-edges

- **Theorem:** If $G = (V, E)$ is a graph with m edges, then

$$\boxed{\sum_{v \in V} \deg(v) = 2m} \leq n(n-1)$$



- **Theorem:** Let G be a simple graph with n vertices and m edges. Then,

$$m \leq \frac{n(n-1)}{2}$$

$$\frac{n(n-1)}{2} \geq m$$

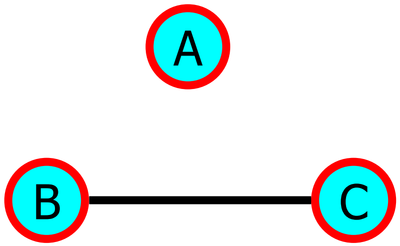
- **Corollary:** A simple graph with n vertices has $O(n^2)$ edges.

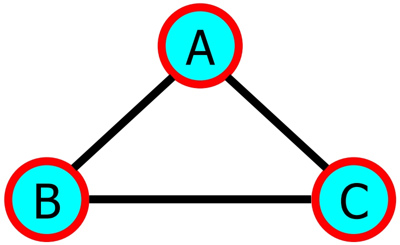
Complete Graphs

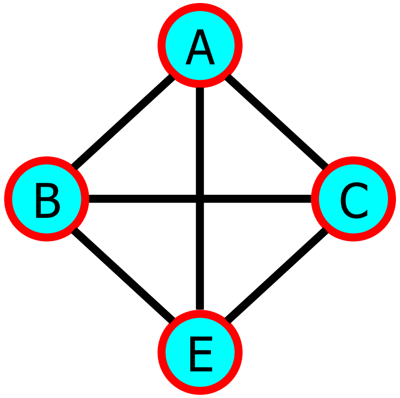
- A ***complete graph*** is a simple graph where an edge connects every pair of vertices

- The complete graph on n vertices has exactly $n(n-1)/2$ edges

- A complete graph with at most one self loop per vertex on n vertices has exactly $n(n+1)/2$ edges



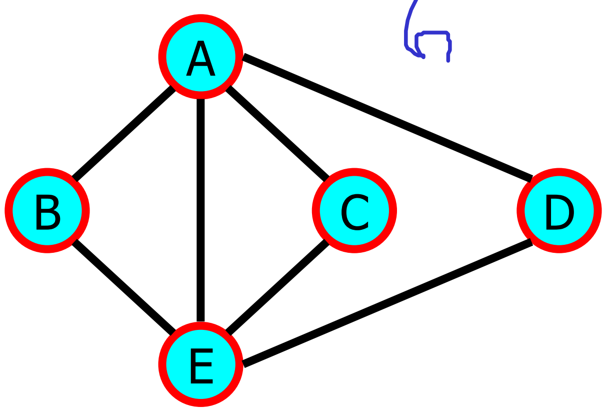




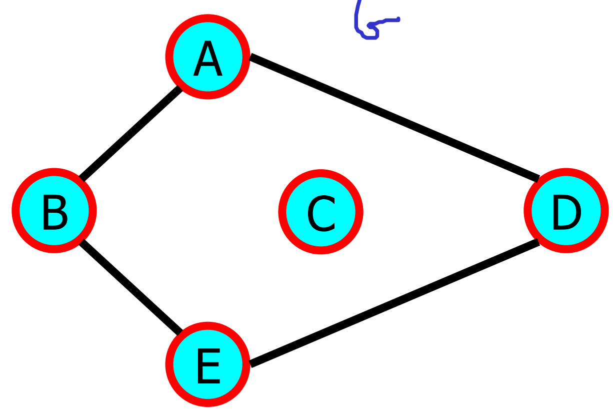
Subgraphs

- A ***subgraph*** of $G = (V, E)$ *is a graph* $G' = (V', E')$ where
  - V' is a subset of V
  - E' consists of edges $\{v, w\}$ in E such that both v and w are in V'
- A ***spanning subgraph*** of G contains all the vertices of G

G



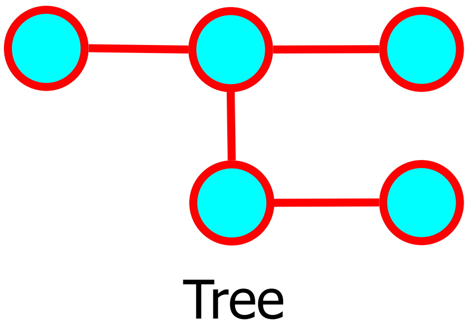
G'



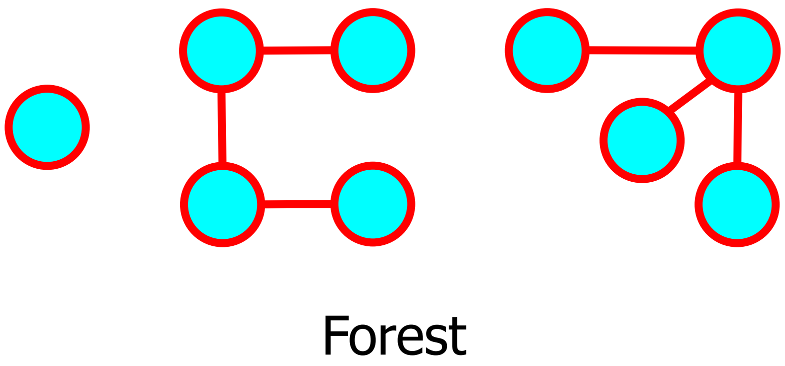
Trees and Forests

- A (*free*) *tree* is an undirected graph T such that
  - T is connected
  - T has no cycles

  This definition of tree is different from the one of a rooted tree

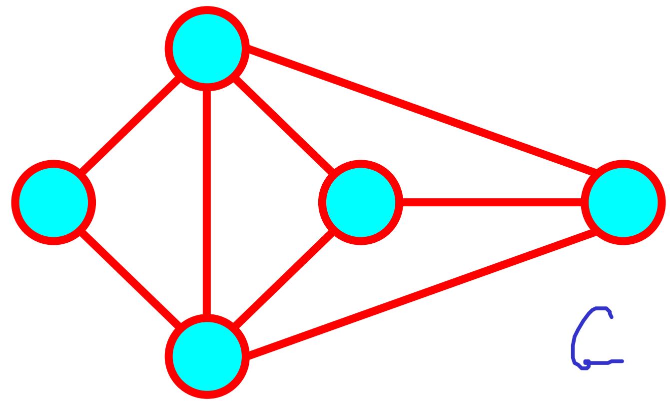


- A *forest* is an undirected graph without cycles
- The connected components of a forest are trees

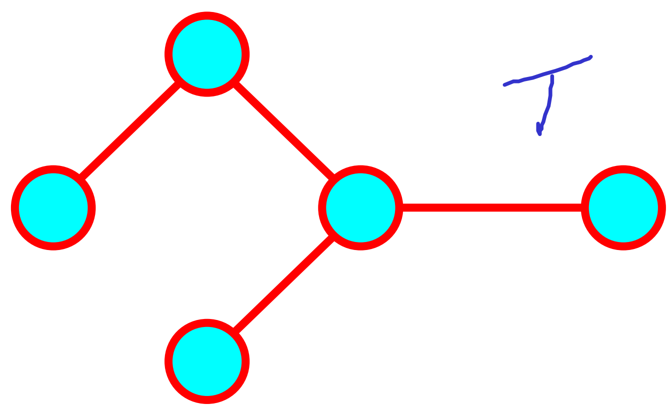


Spanning Trees and Forests

- A *spanning tree* of a connected graph is a spanning subgraph that is a tree
- A spanning tree is not unique unless the graph is a tree
- Spanning trees have applications to the design of communication networks
- A *spanning forest* of a graph is a spanning subgraph that is a forest



Graph



Spanning tree

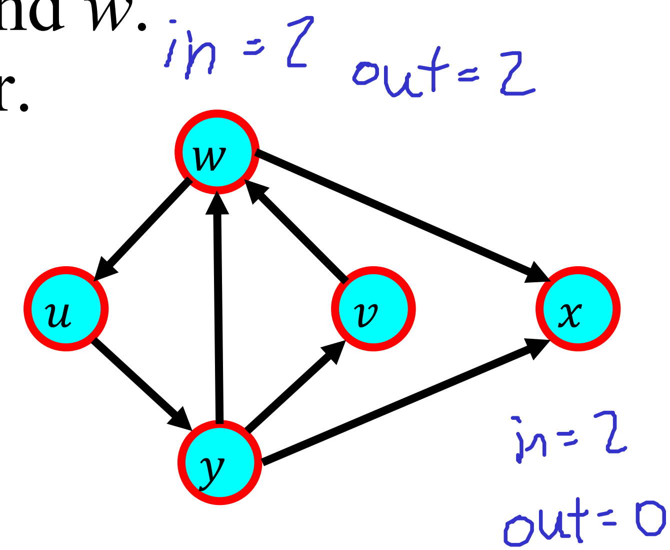
Properties of Trees, Forests and Graphs

$m \leq \frac{n(n-1)}{2}$

- **Theorem:** Let G be an undirected graph with n vertices and m edges. Then we have the following:
  - If G is connected, then $m \geq n - 1$.
  - If G is a tree, the $m = n - 1$.
  - If G is a forest, then $m \leq n - 1$.

Directed Edges or Arcs

- A *directed edge* (*or arc*) *e* represents an *asymmetric* relation between two vertices *v* and *w*. *e* = (*v*, w) denotes an ordered pair.
  - *v*, *w* are the endpoints of the edge
  - *v* is *adjacent* to *w*
  - *e* is *incident* upon *v* and *w*
  - The arc goes from the *source* vertex *v* to the *destination* vertex *w*
- The *indegree* of a vertex is the number of incoming arcs
- The *outdegree* of a vertex is the number of outgoing arcs



Simple Graphs

- A ***simple digraph*** is a graph with no self-loops and no parallel or multi-edges

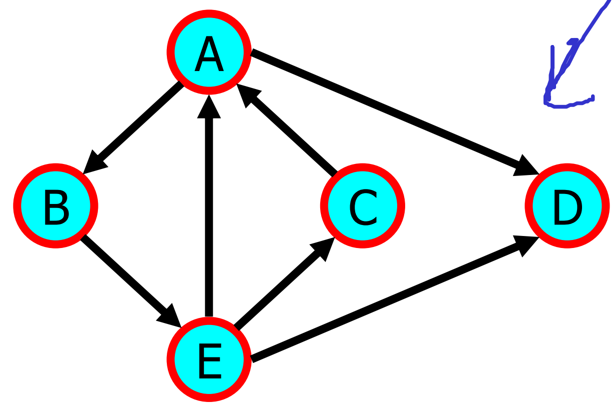
- **Theorem:** If $G = (V, E)$ is a digraph with m edges, then

$$\sum_{v \in V} \mathrm{indeg}(v) = \sum_{v \in V} \mathrm{outdeg}(v) = m$$

- **Theorem:** Let G be a simple digraph with n vertices and m edges. Then,

$$m \leq n(n-1)$$

- **Corollary:** A simple digraph with n vertices has $O(n^2)$ edges.



Con. but not strongly

Connected Digraphs

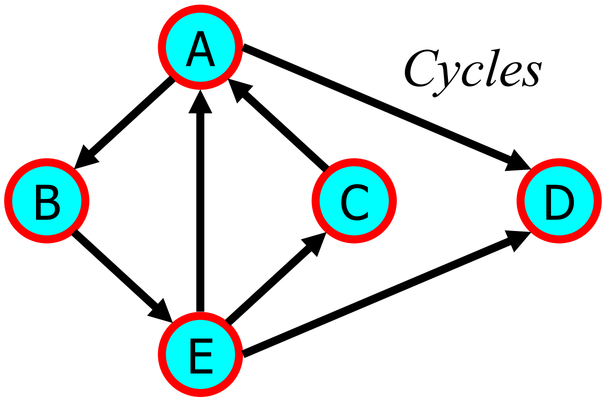
- Given vertices u and v of a digraph G, we say v is ***reachable*** from u if G has a directed path from u to v.

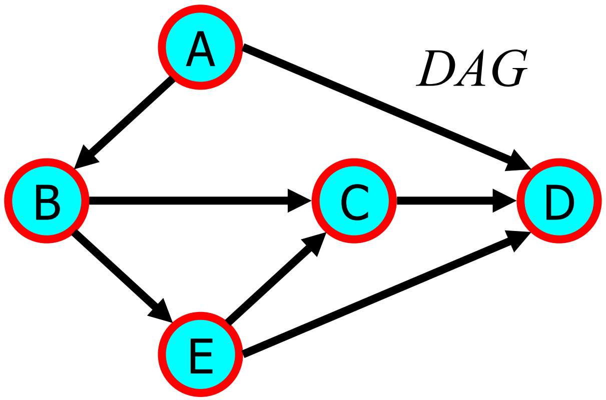
- A digraph G is ***connected*** if every pair of vertices is connected by an undirected path.

- A digraph G is ***strongly connected*** if for every pair of vertices u and v of G, u is reachable from v and v is reachable from u.

Directed Acyclic Graphs (DAGs)

- A ***directed acyclic graph*** (DAG) is a directed graph with no cycles.
- DAGs are more general than trees, but less general than arbitrary directed graphs.





Weighted Digraphs

- A graph with edge labels is called a *labeled graph* (or *weighted graph*)
- Edge labels denote edge attributes such as distance, throughput, capacity, time, etc.
- A vertex can have both a name and a label
- These will be studied in CSC226