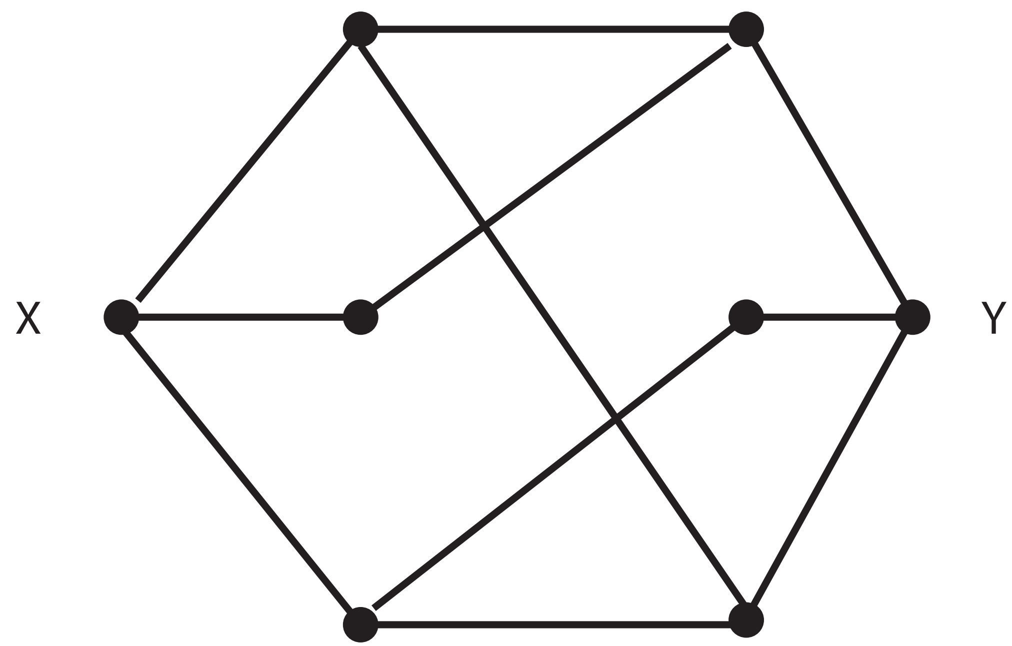
Graph Theory Fundamentals

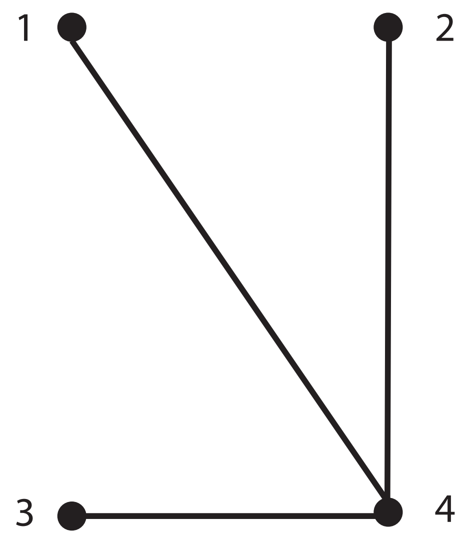
Undirected Graphs

Edges have no direction



Graphs and Edge Lists

Throughout the semester, let n = number of nodes and m = number of edges



n=4
m=3

Edge list: (1,4), (2,4), (3,4)

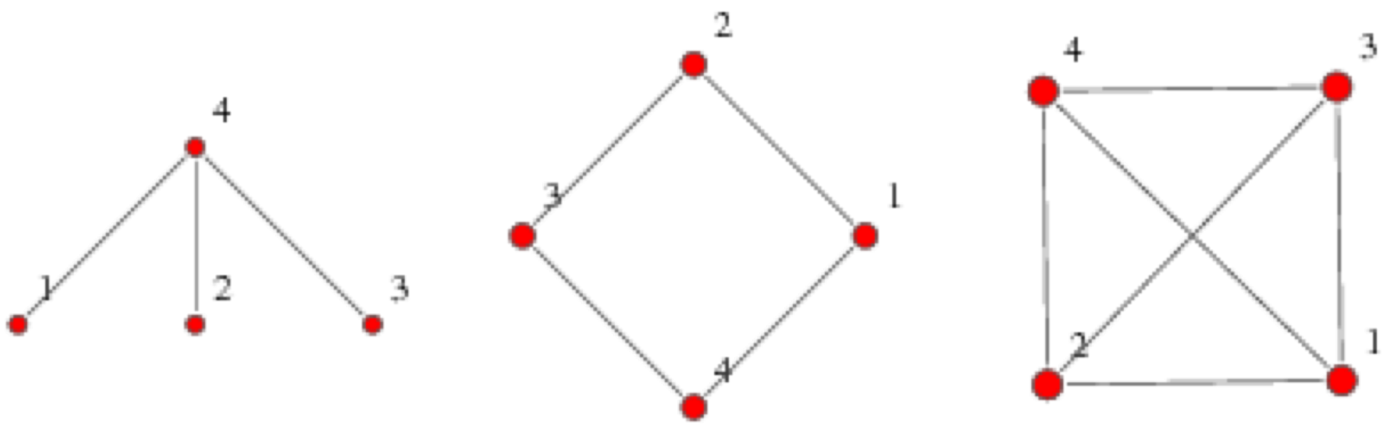
Two nodes connected with an edge are called neighbors

Edge lists can get really long for big graphs

The Adjacency Matrix (A)

A weighted graph has weights on the edges. In an unweighted graph all edges have weight of 1.

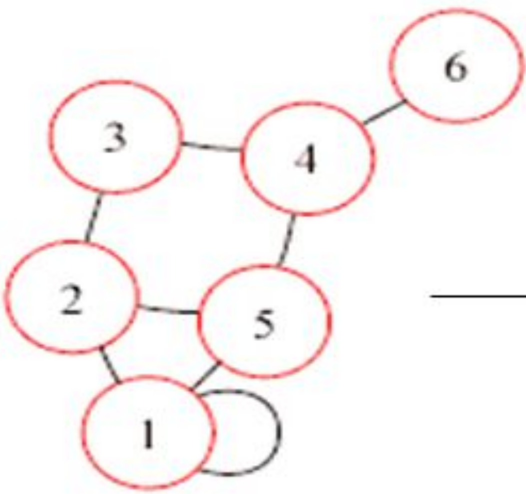
The adjacency matrix for a graph is n X n and each element contains 0 for non-neighbors and the edge weight for neighbors.



$$A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

Properties of the Adjacency Matrix

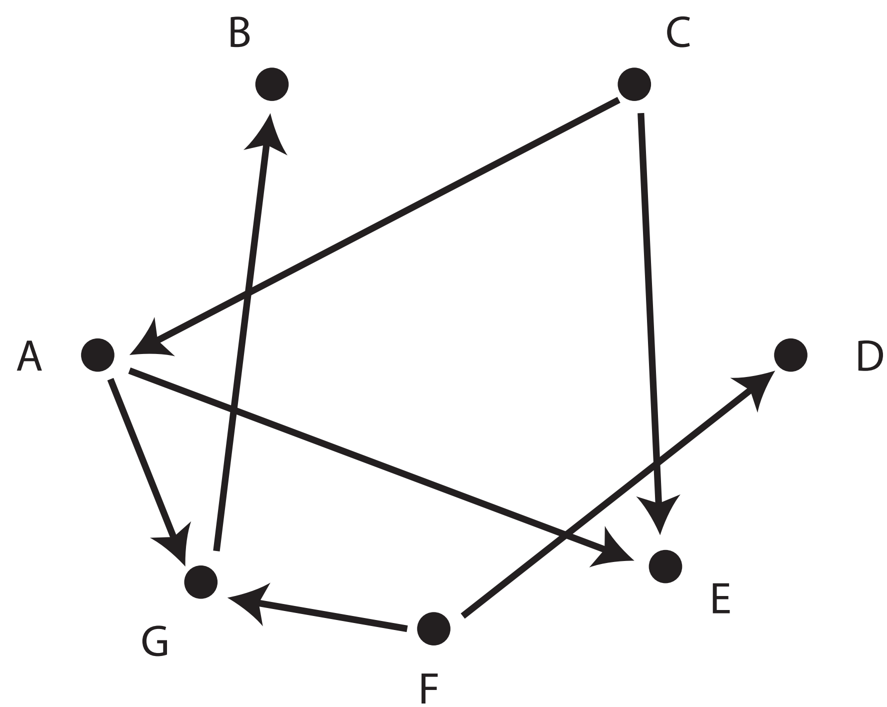
1. It is symmetric ($A = A^T$)
2. For a simple graph, with no self-edges, elements on main diagonal are 0
3. If the graph is not simple, then the matrix element for a node with a self-edge is represented by 2*(edge weight).



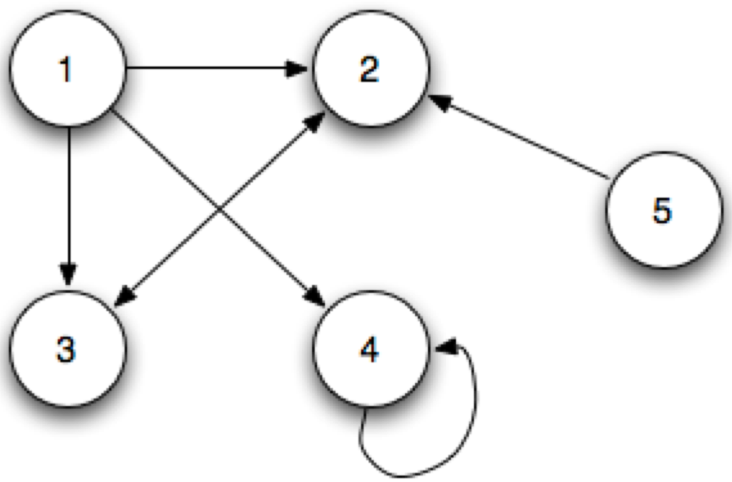
$$\begin{pmatrix} 2 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Directed Graphs

Edges have arrows giving direction



Directed Graph with Adjacency Matrix



To

From

A =

0	1	1	1	0
0	0	1	0	0
0	1	0	0	0
0	0	0	1	0
0	1	0	0	0

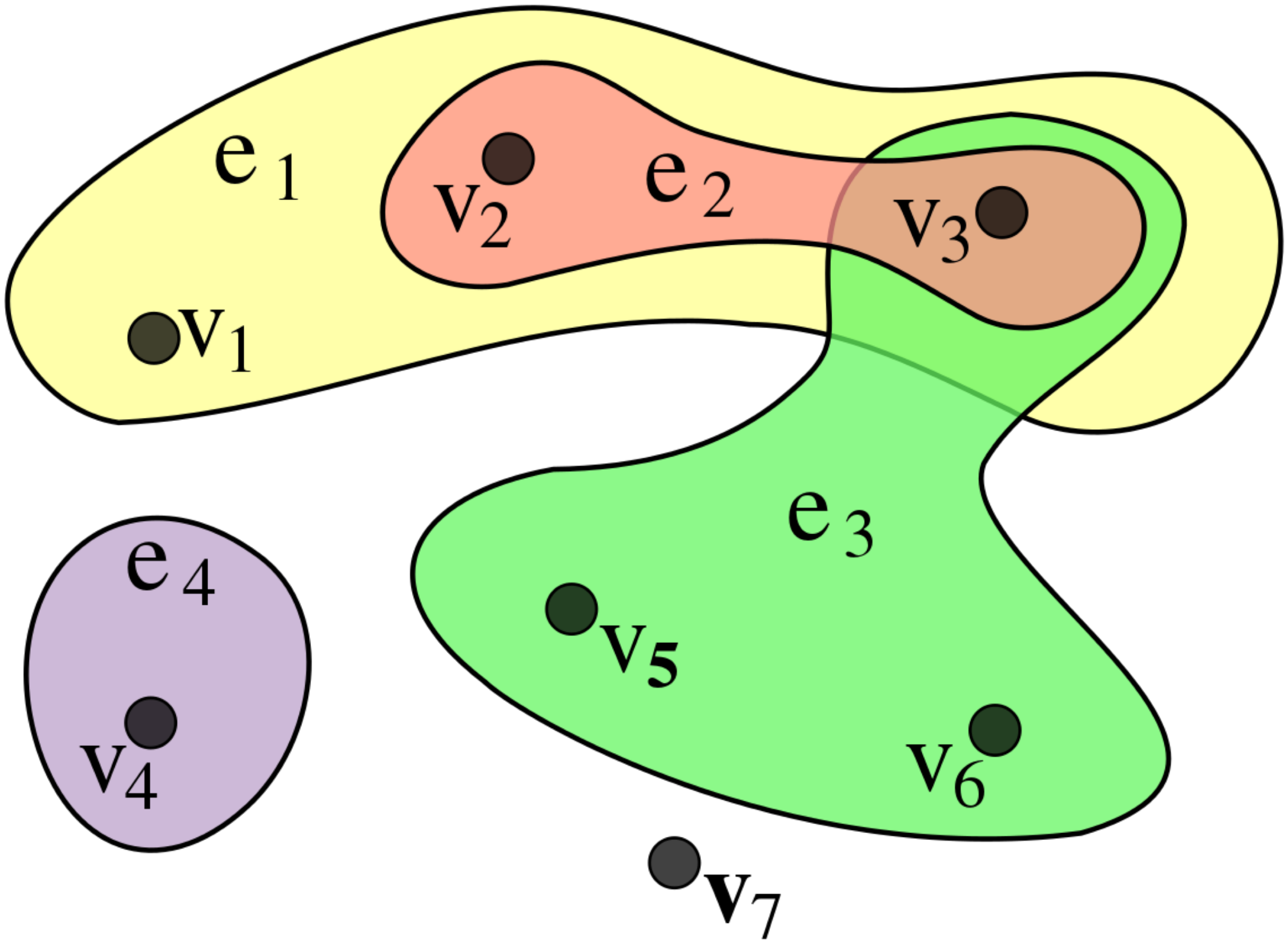
$a_{ij} = 1$ if there is an edge from node i to node j

(this convention is not universal)

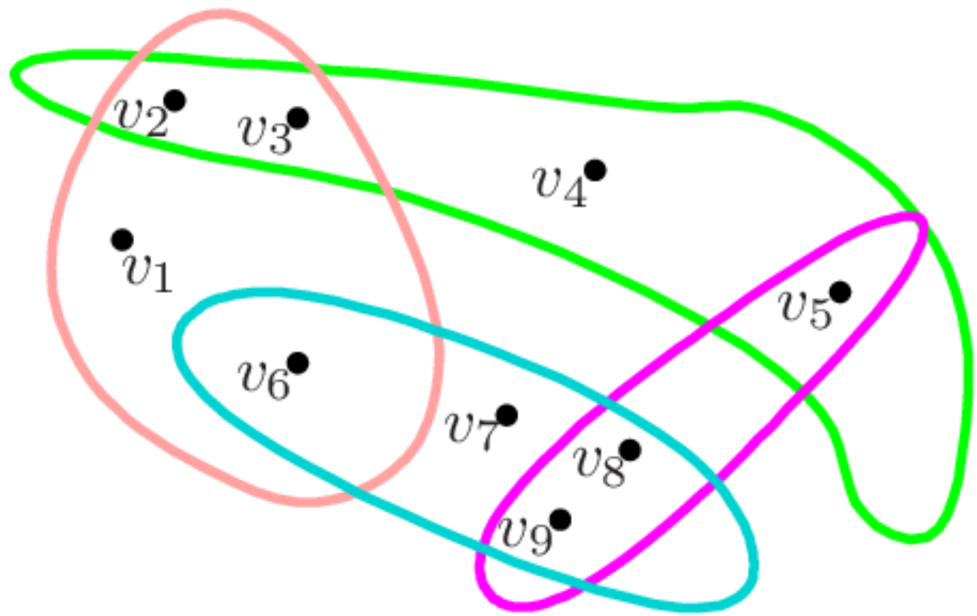
A self-edge just gets the weight of the edge

The adjacency matrix of a directed graph is usually not symmetric

Hypergraphs and Bipartite Graphs



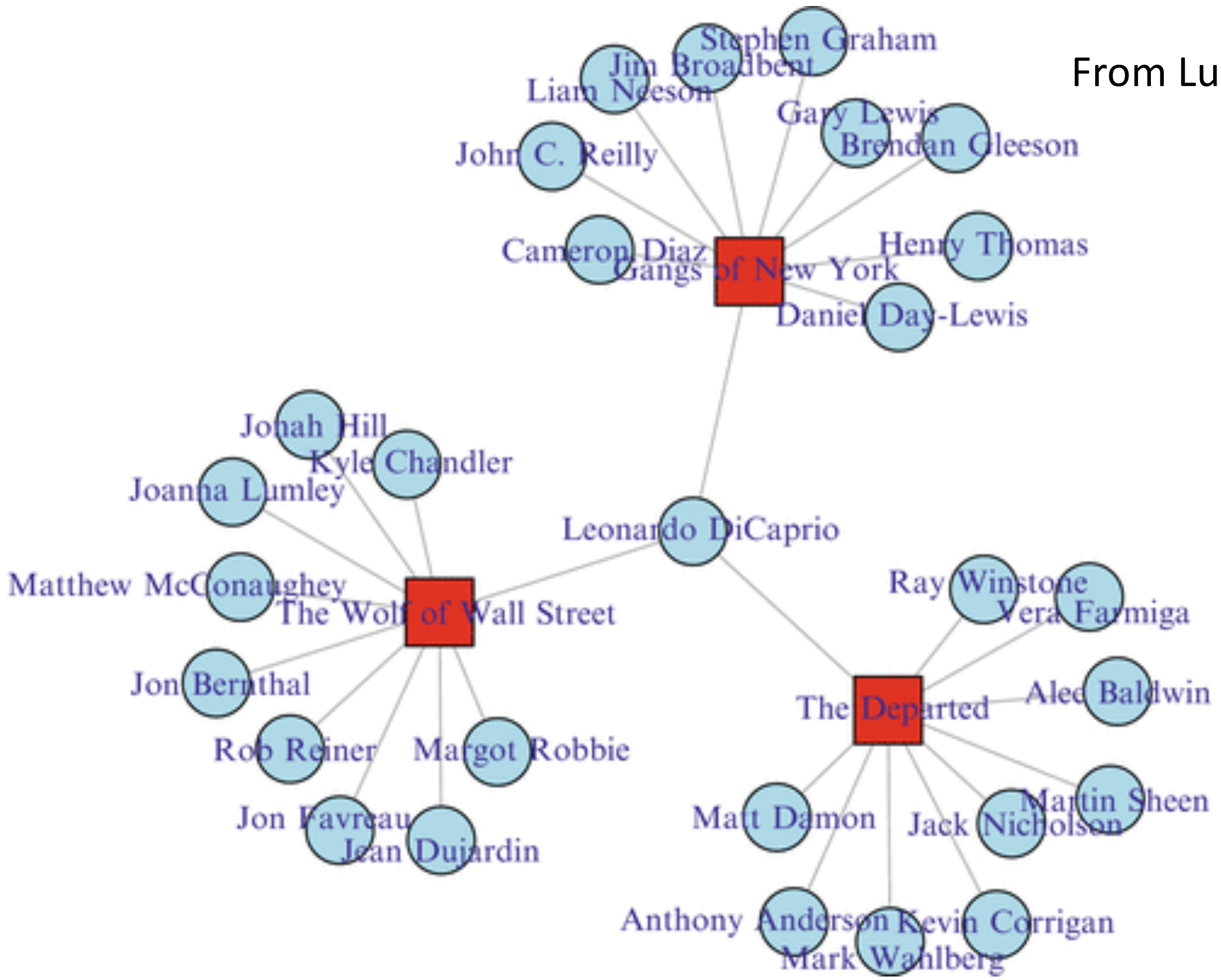
A Hypergraph Indicates Group Inclusion



The nodes could represent actors and the groups could represent movies. There are no standard edges in a hypergraph. Each closed curve is called a hyperedge.

Another Way: Bipartite Graphs

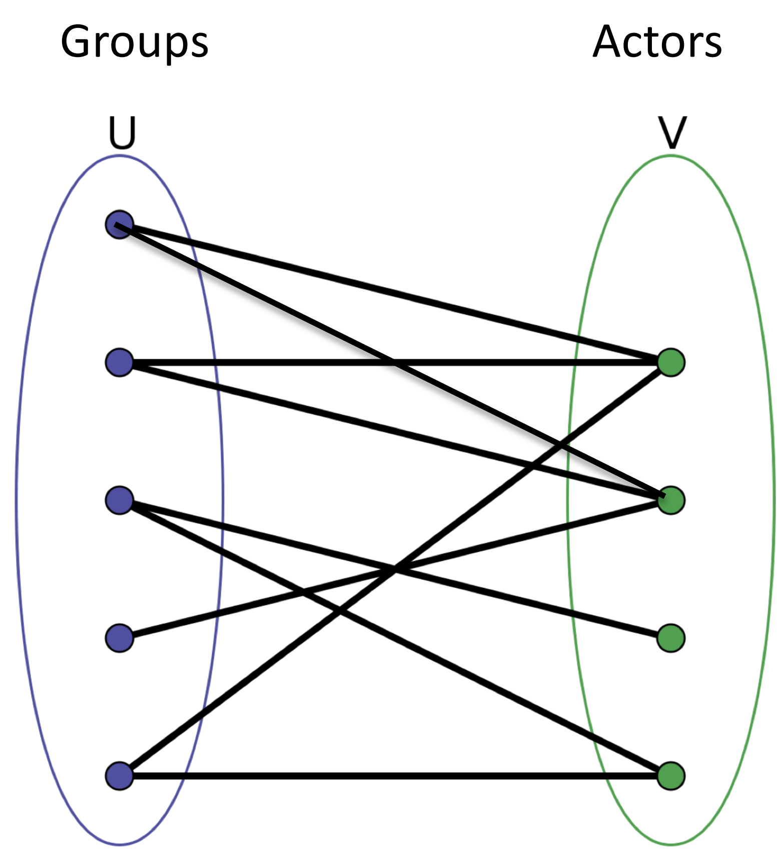
From Luke (2015)



Example of a *bipartite graph* in which there are two types of vertices: The "actors" connect to the "groups", but there are no actor-actor or group-group connections

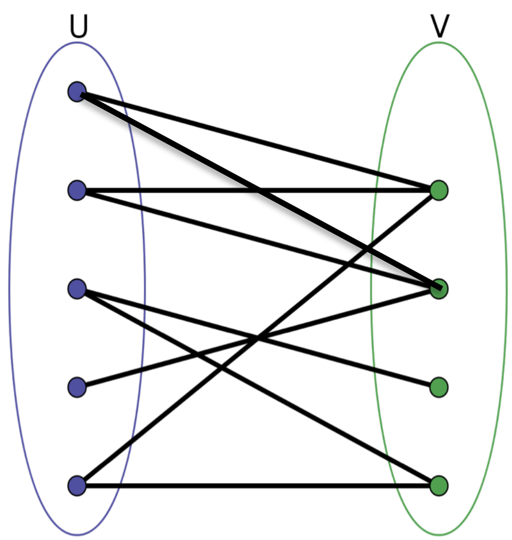
Bipartite Graphs are an Alternative to Hypergraphs

Two types of nodes: one represents groups and the other actors.
No edge between nodes of the same type.
This is an example of a two-mode graph.



The Incidence Matrix (B) Contains the Bipartite Graph Structure

This is a g X r matrix where g = number of groups and r = number of actors. Then $b_{ij} = 1$ if actor j belongs to group i.



Actors

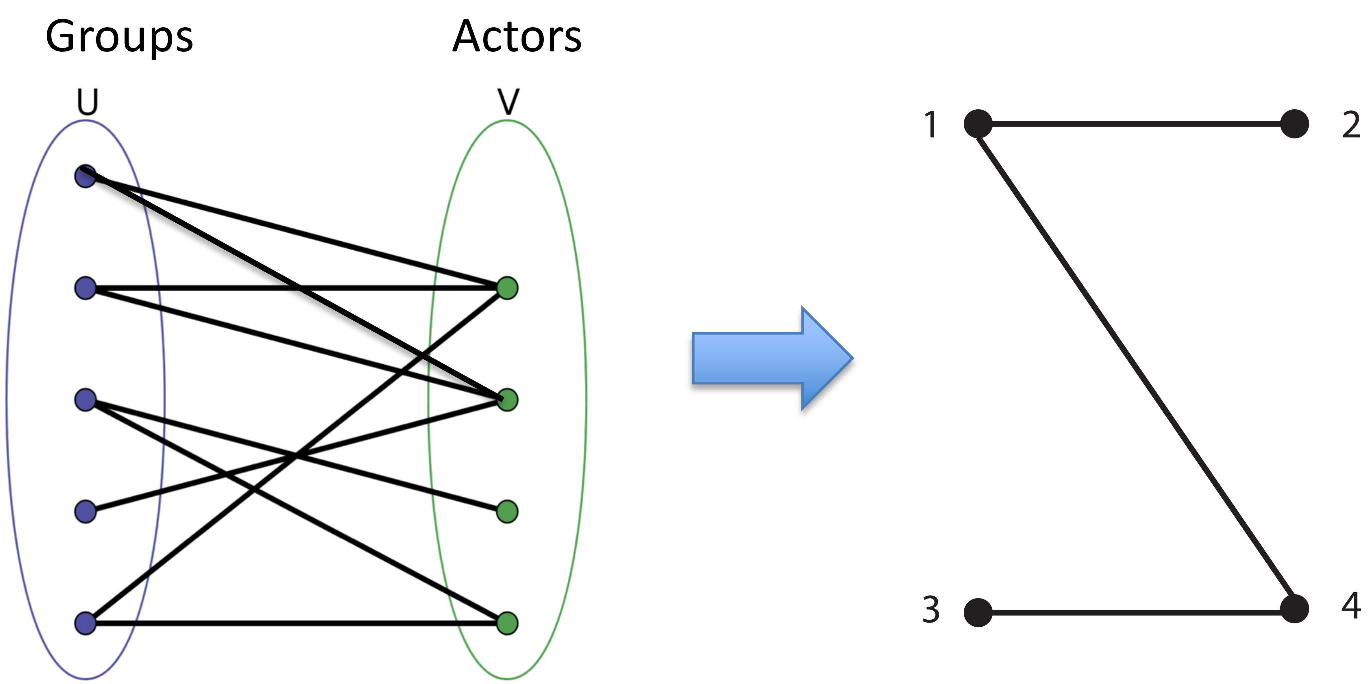
Groups

$$B = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

The incidence matrix is typically **not square**

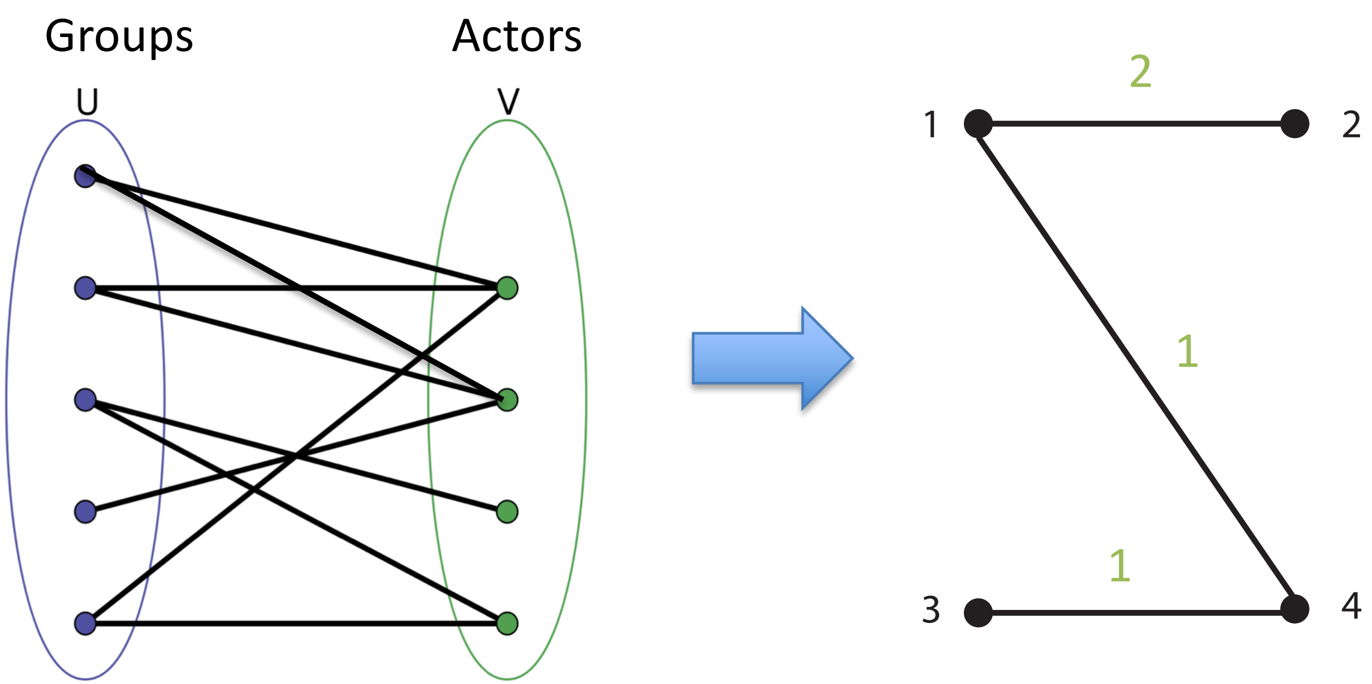
One-Mode Projections of a Bipartite Graph

Suppose we want to indicate which actors are related by appearing in the same movie (or group)? We can just connect such actors by edges. This is an example of a one-mode projection of the bipartite graph according to actors.

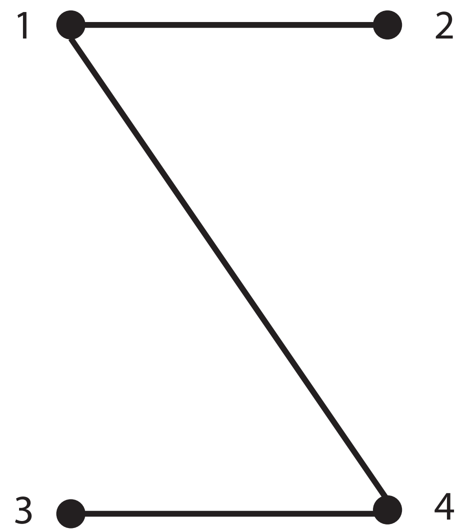


Weighted One-Mode Projection According to Actors

The edge weights indicate the number of times neighbors appeared together in a movie.



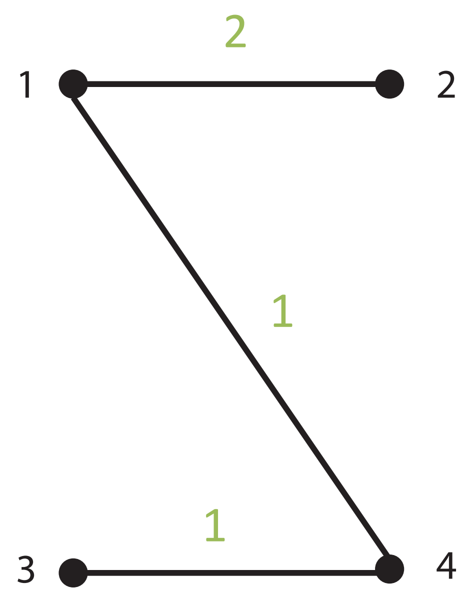
Adjacency Matrix for a One-Mode Projection



This symmetric matrix is r X r, where r is the number of actors

$$A_r = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

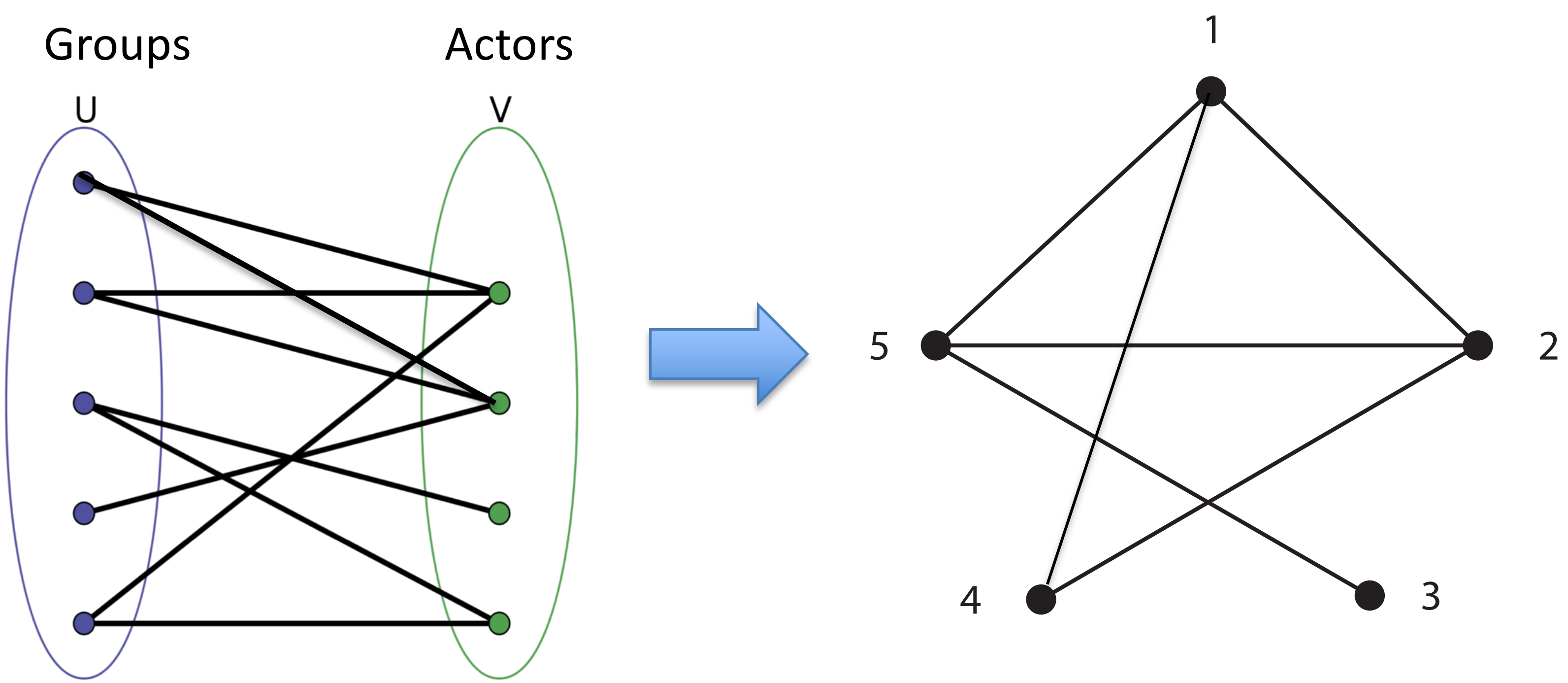
Weighted Adjacency Matrix for a One-Mode Projection



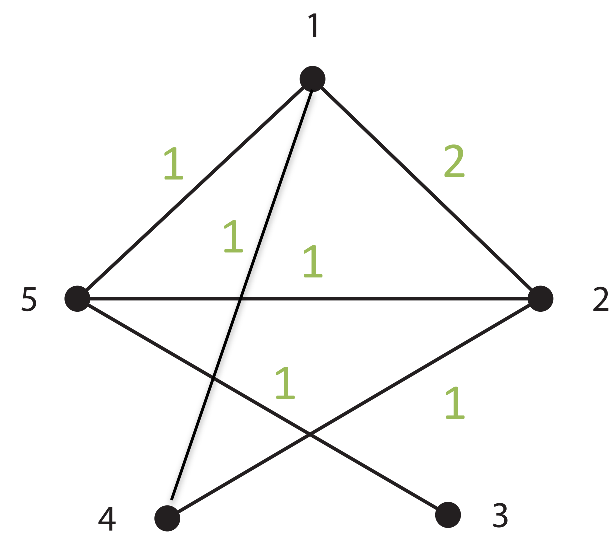
$$A_r^w = \begin{bmatrix} 0 & 2 & 0 & 1 \\ 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

There is a Second One-Mode Projection

This one-mode projection is done according to the groups (e.g., which movies have a common actor).



And a Second Adjacency Matrix



This symmetric matrix is g X g, where g is the number of groups.

$$A_g^w = \begin{bmatrix} 0 & 2 & 0 & 1 & 1 \\ 2 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{bmatrix}$$

How are B, A_r^w, and A_g^w Related?

Actors *i* and *j* are both in group *k* if B_{ki}=1 and B_{Kj}=1. That is, if $B_{Ki}B_{kj}$=1. The number of groups they share is then

$$P_{ij} = \sum_{k=1}^{g} B_{ki}B_{kj} = \sum_{k=1}^{g} B_{ik}^{T}B_{kj}$$

The r X r matrix P is then $P = B^T B$. This has the right dimensions as A_r^w, but is it the same?

No! Since A_r^w has 0 down the main diagonal, while

$$P_{ii} = \sum_{k=1}^{g} B_{ki}B_{ki} = \sum_{k=1}^{g} B_{ki}^{2}$$

This is the square sum down column *i* of B, and is related to the number of groups that actor *i* is in, which is usually non-zero. So set diagonal elements of P to 0.

How are B, A_r^w, and A_g^w Related?

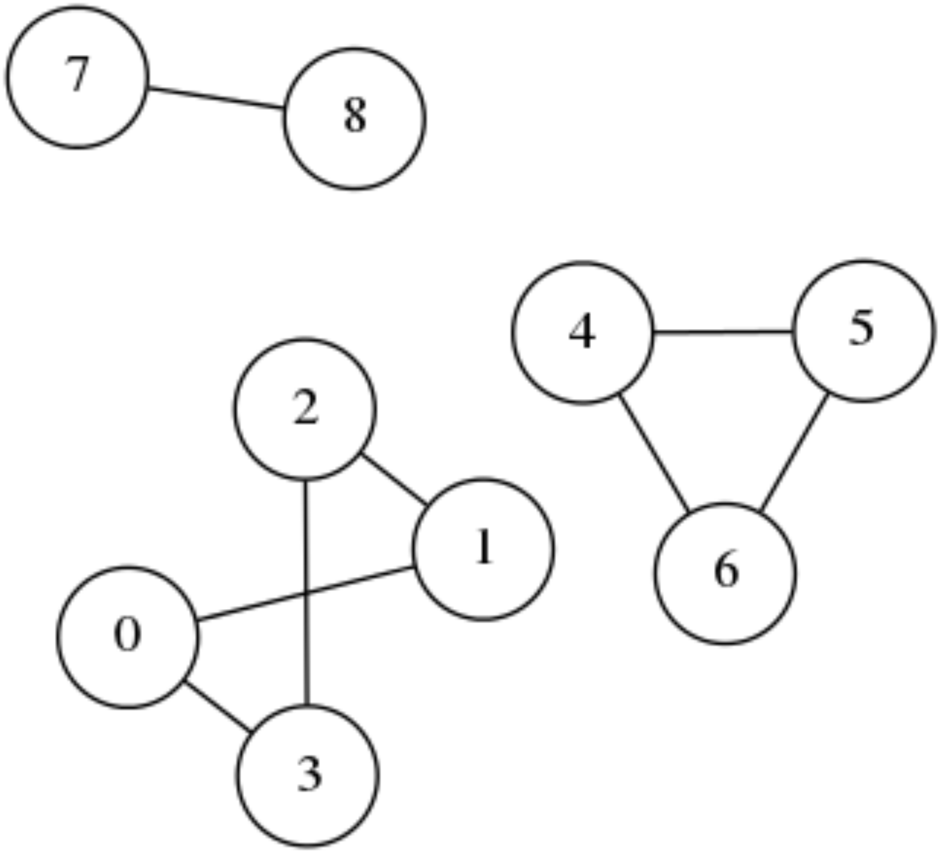
$$\boxed{A_r^w = P = B^T B \text{ with } P_{ii} \equiv 0}$$

And if we define the g X g matrix $P' = BB^T$ then P'_{ii} = the square row sum across row *i* of B, which is the number of actors in group *i* (not usually 0, as are diagonal elements of A_g^w). So set diagonal elements of P' to 0.

$$\boxed{A_g^w = P' = BB^T \text{ with } P'_{ii} \equiv 0}$$

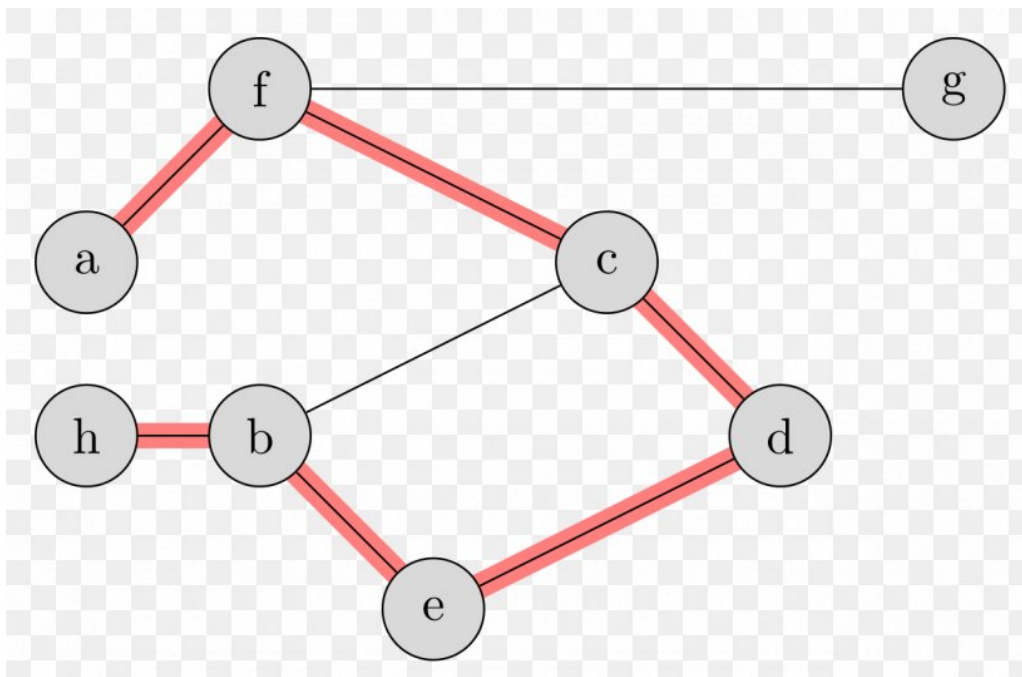
Thus, the adjacency matrices corresponding to the two one-mode projections can be calculated directly from the incidence matrix, without having to construct the network diagrams.

Paths and Connectivity



Path Through a Graph

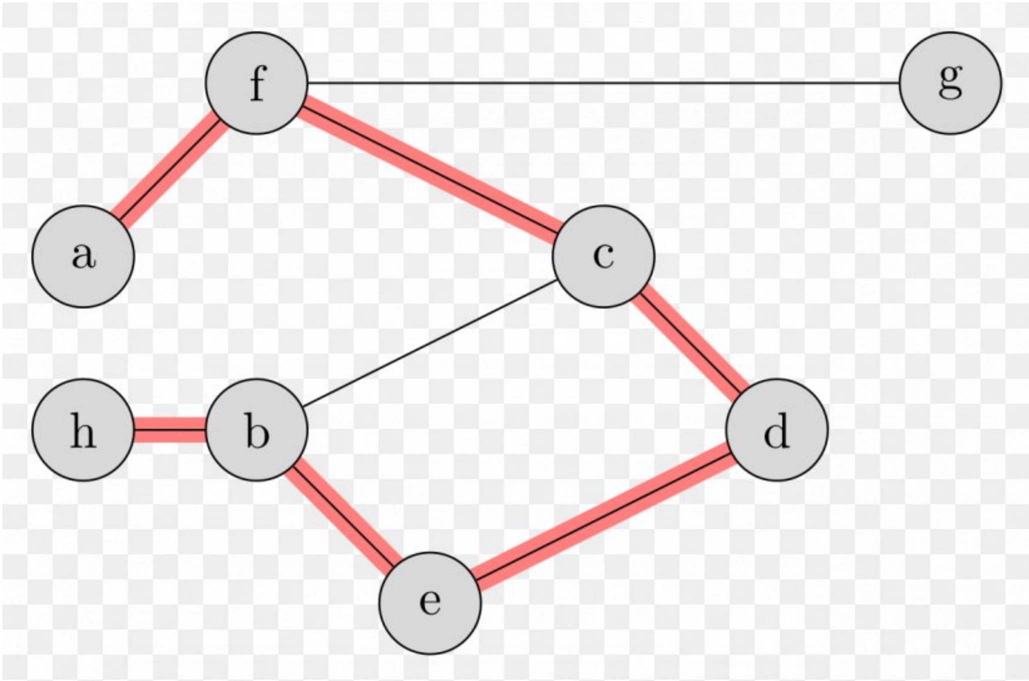
A **path** from node *A* to node *B* is just a sequence of nodes that starts at *A* and ends at *B* such that each consecutive node pair is connected by an edge. There may be many paths between two nodes, or there may not be any. A path is called simple or self-avoiding if it does not repeat any nodes.



Example of a simple path between *a* and *h*

How Many Paths Are There?

There may be many paths between any two nodes. In the graph below, here is a simple path (a,f,c,b,h) from node *a* to *h* and another simple path (a,f,c,d,e,b,h), but also the non-simple path (a,f,c,d,e,b,c,d,e,b,h). There are an infinite number of non-simple paths from a to h due to the cycle c, d, e, b.



Example of a simple path from node *a to h*

A Cool Theorem on Path Cardinality

Theorem: The *i,j*-entry of A^k equals the number of paths of length *k* from node *i* to node *j*.

Note: these paths include both simple and non-simple ones.

Proof: The proof proceeds by induction.

The $k = 1$ case comes directly from the definition of adjacency matrix A.

Assume that the statement is true for A^k and look at the *i,j*-entry of A^{k+1}. By the definition of matrix multiplication,

$$A^{k+1} = A^k A$$

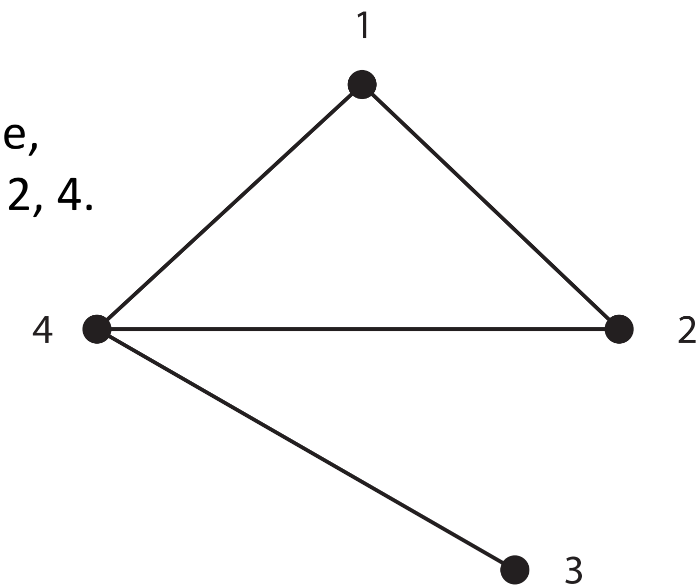
so

$$\left(A^{k+1}\right)_{ij} = \left(A^k\right)_{i1} A_{1j} + \left(A^k\right)_{i2} A_{2j} + \cdots + \left(A^k\right)_{in} A_{nj}$$

Since $\left(A^k\right)_{ip}$ is the number of paths of length *k* from node *i* to node *p* and $A_{pj} = 1$ if and only if there is an edge from node *p* to node *j* (and 0 otherwise), then $\left(A^k\right)_{ip} A_{pj}$ is the number of paths of length $k + 1$ from *i* to *j* that go through *p*. The sum on the right adds these up over all nodes *p*, and is therefore the total number of paths of length $k + 1$ from node *i* to *j* as desired.

Another Cool Theorem on Triangles

Graph with one triangle, comprised of nodes 1, 2, 4.



Theorem: The number of triangles of a graph equals $\frac{1}{6}\mathrm{Tr}(A^3)$.

Another Cool Theorem on Triangles

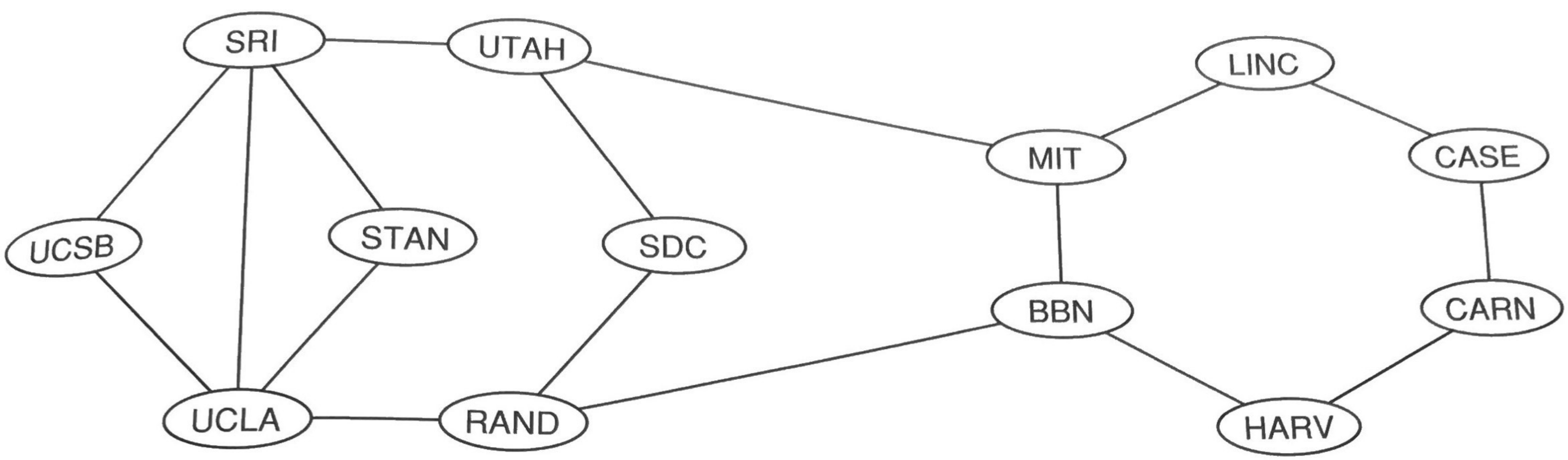
Theorem: The number of triangles of a graph equals $\frac{1}{6}\mathrm{Tr}(A^3)$.

Recall that $\mathrm{Tr}(A) = A_{11} + A_{22} + \cdots + A_{nn}$

Proof: A path of length 3 from a node to itself is a triangle, and that triangle actually yields two paths, one in each direction. Therefore, if node *i* is contained in a triangle then $(A^3)_{ii} = 2$. Then $\mathrm{Tr}(A^3)$ equals twice the number of nodes contained in triangles. However, since each triangle contains 3 nodes, it follows that $\mathrm{Tr}(A^3)$ equals six times the number of triangles. Thus, the number of triangles is $\frac{1}{6}\mathrm{Tr}(A^3)$.

Length of a Path and Distance

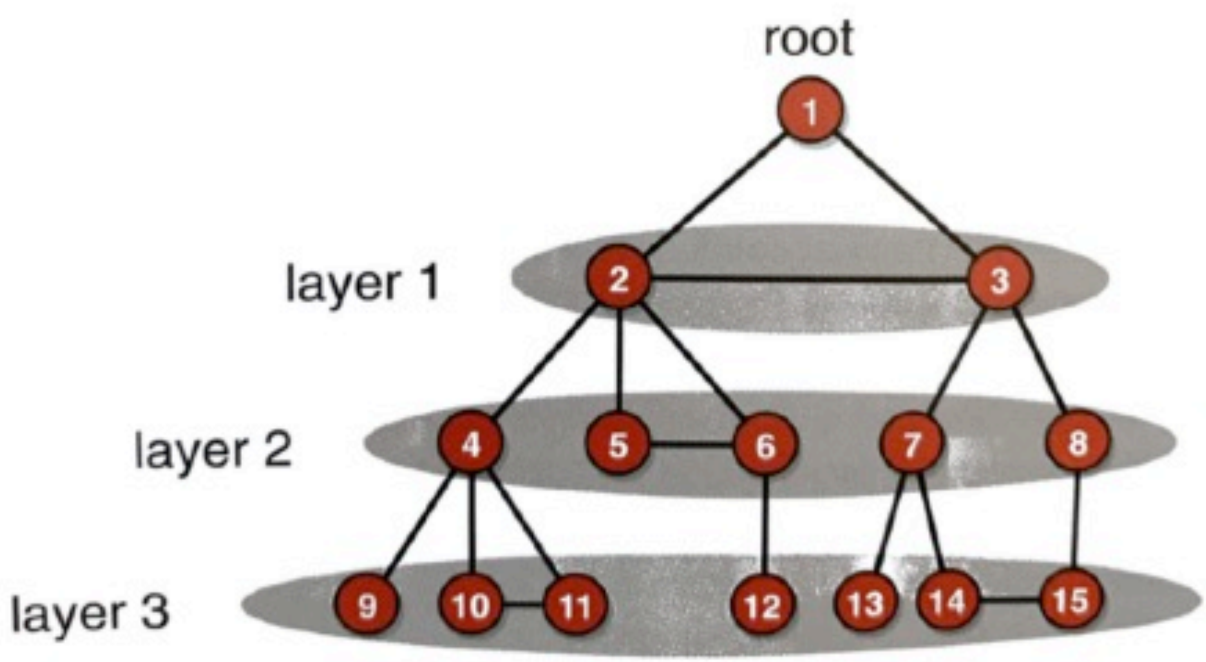
The **length** of a path is the number of edges from beginning to end.
The distance between two nodes *A* and *B* is the length of the shortest path between them.



Graph of the ARPANET from 1970. What are the simple paths between Univ. California Santa Barbara (UCSB) and the Rand Corporation (RAND)? How long are these paths? What is the distance between UCSB and RAND?

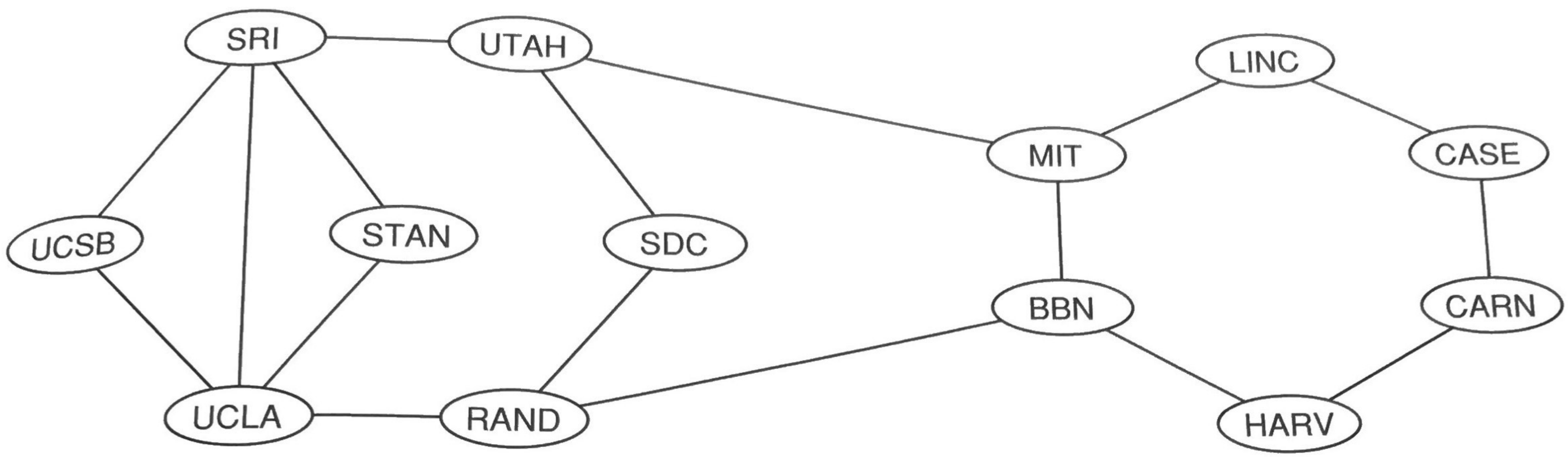
Breadth-First Search

In a small graph finding the distance between two nodes is easy. It is not so trivial for a large network, where visualization does not work. A useful algorithm is called a breadth-first search, which is illustrated below for a 15-node undirected graph.



Cycles

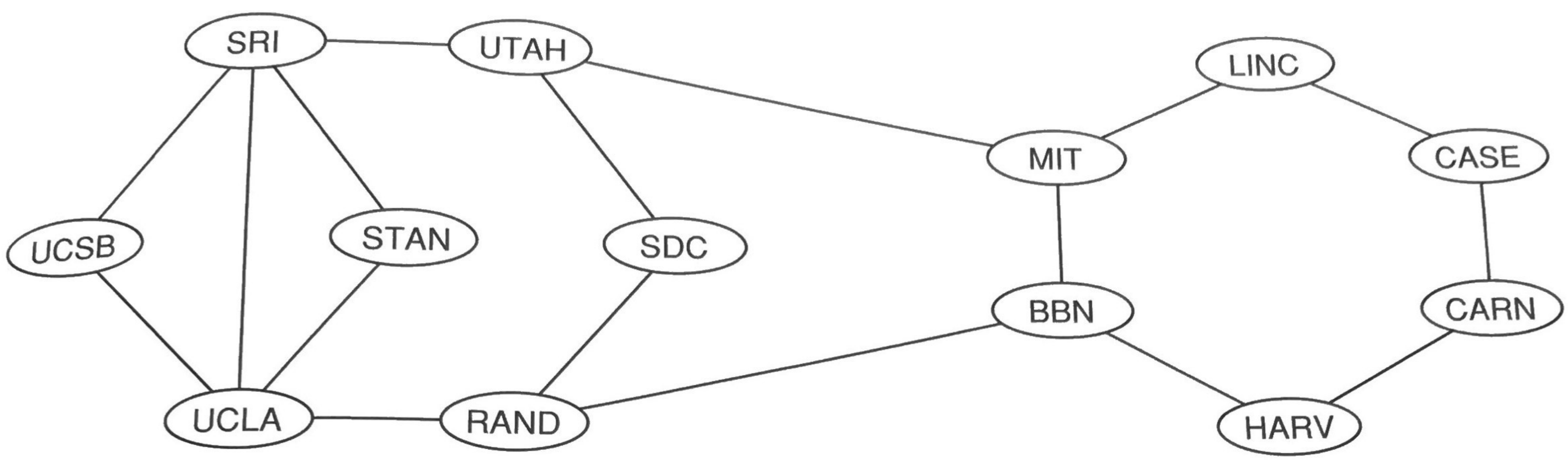
A cycle is a path with at least three edges, in which the first and last node are the same and no node (except the first/last) is repeated. This gives a ring structure to the nodes in a cycle.



What are some of the cycles involving UCSB in the ARPANET graph?

Cycles and Redundancy

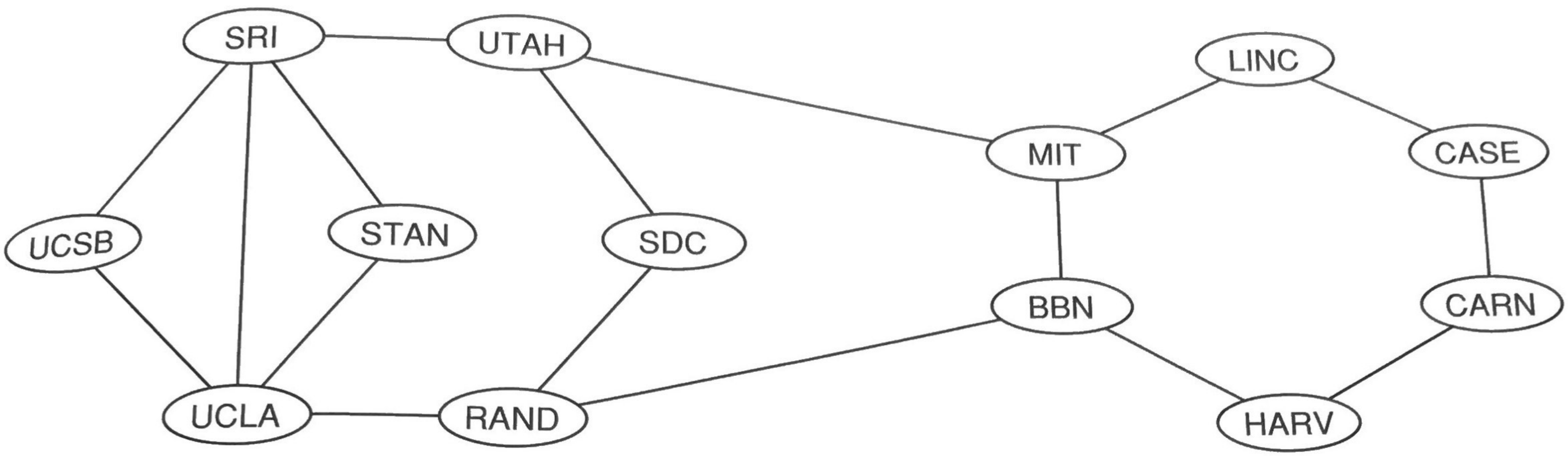
If a sequence of nodes form a cycle, then any connecting edge can be removed and there will still be a path between any pair of nodes. In terms of a communication network or transportation network, this is an example of redundancy. If one of the communication lines (or roads) breaks it won't leave anyone stranded. Alternate routes could be used to get between any pair of nodes.



If the link between UCSB and UCLA goes down, there are still paths between routers at these universities.

Connected Graphs

A graph is connected if there is a path between every pair of nodes.



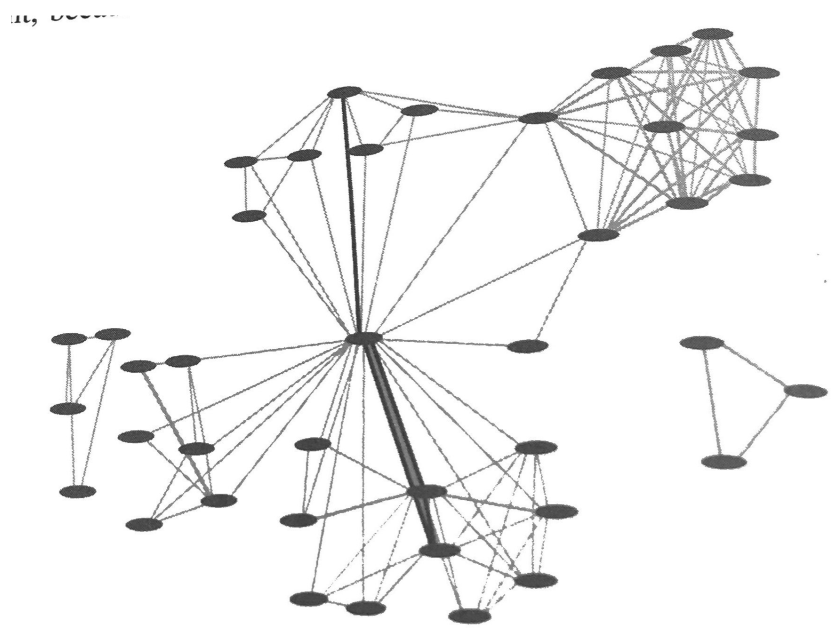
The 1970 ARPANET network is connected. Why should we expect most communication and transportation networks to be connected?

Should we expect social networks to be connected? How about biological networks?

Components

A connected component (or just component) of a graph is a subset of the nodes such that:

(1) every node in the subset has a path to every other node
(2) the subset is not part of some larger set with the property that every node can reach every other. That is, it is a maximal subset.

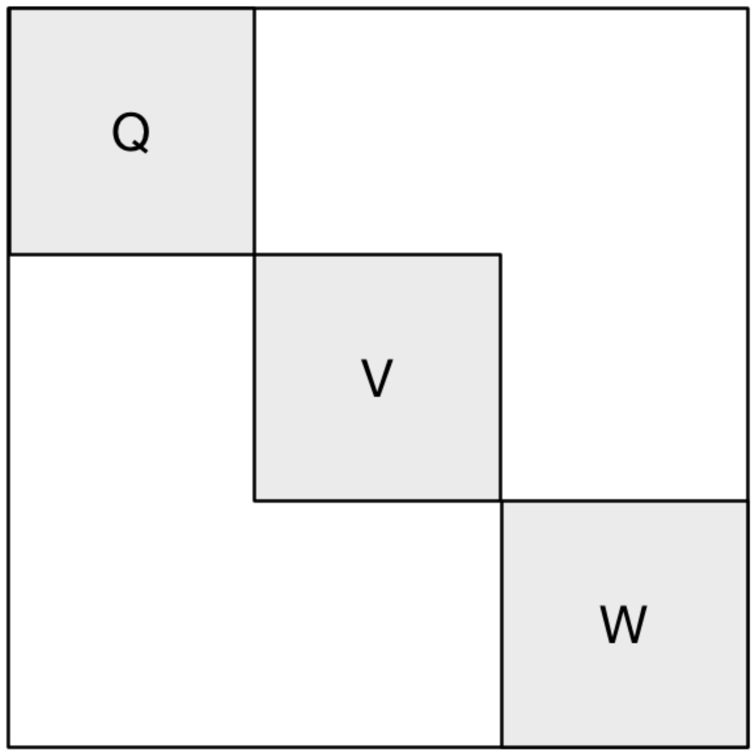


Easely and Kleinberg, 2010

How many components are there in this collaboration network of the biological research center Structural Genomics of Pathogenic Protozoa?

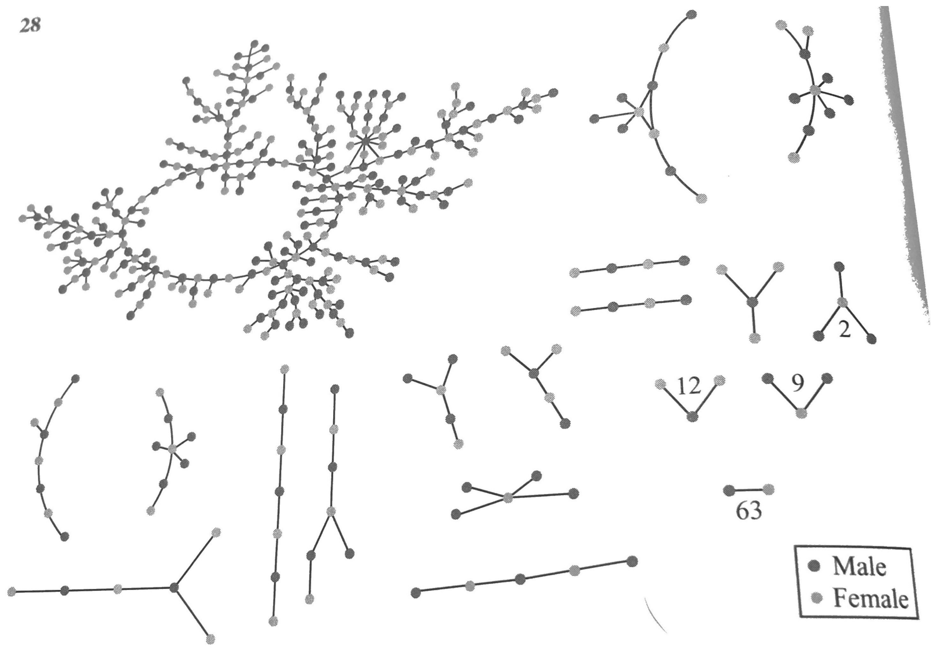
Block Diagonal Adjacency Matrix

The nodes in a graph with *p* components can be numbered so that the adjacency matrix has a block diagonal form with *p* blocks. That is, *A* is a matrix with smaller square matrices along the main diagonal, and off-diagonal elements of 0.



Adjacency matrix A
Sparse / block-diagonal

Giant Components



Easely and Kleinberg, 2010

The network above describes dating patterns among students over an 18-month period. There are many components, but one is much larger than the others. This is an example of a giant component, which is just a really large component. Such things are typical in networks with more than one component.

Why is there a giant component in this network? Why would a typical communication network or transportation network have a giant component?

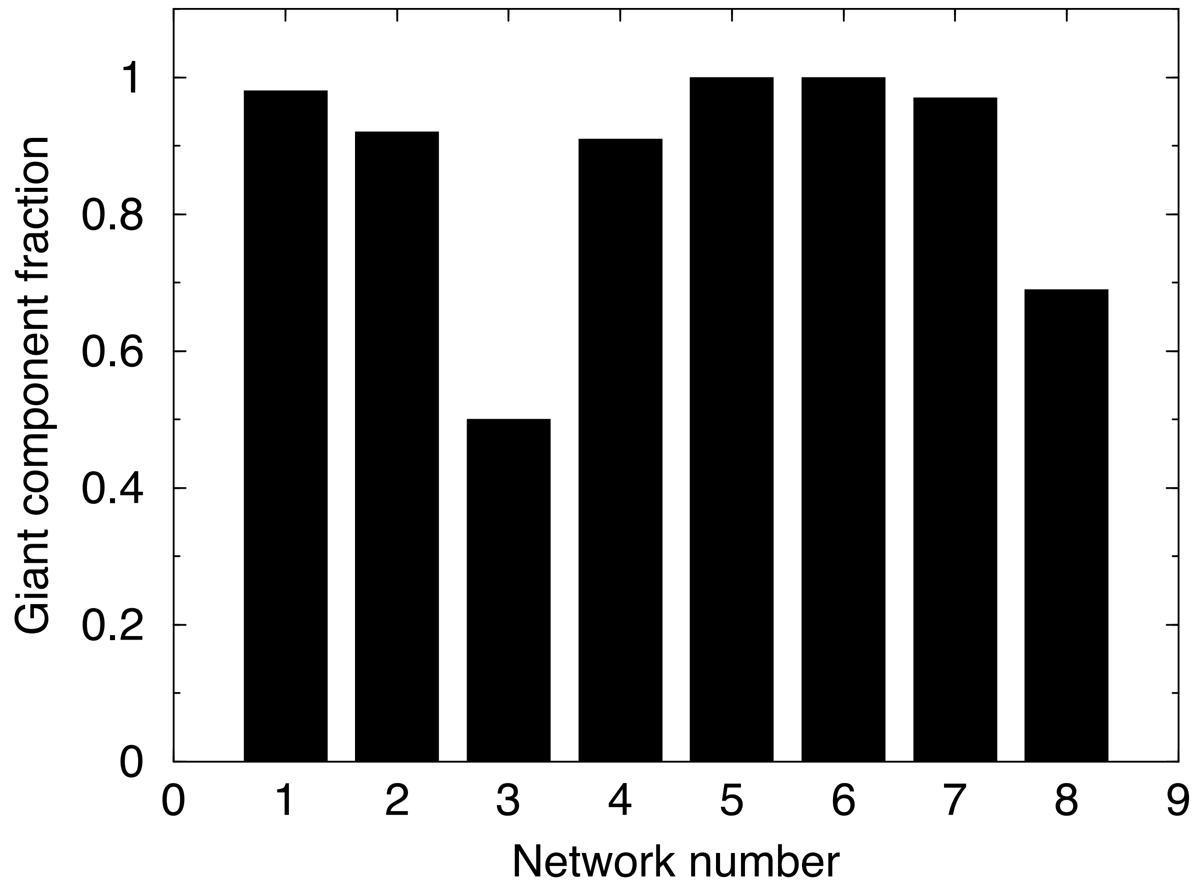
Can There Be Multiple Giant Components?

Suppose a network with 250 nodes has two giant components, each with 100 nodes. How many ways are there for this to collapse into a single giant component?

$$100^2=10,000$$

Having more than one giant component is very unlikely since all it takes is one connection from one giant component to another for the two giant components to collapse into one component, and since there are lots of nodes in the components there are lots of ways to connect them.

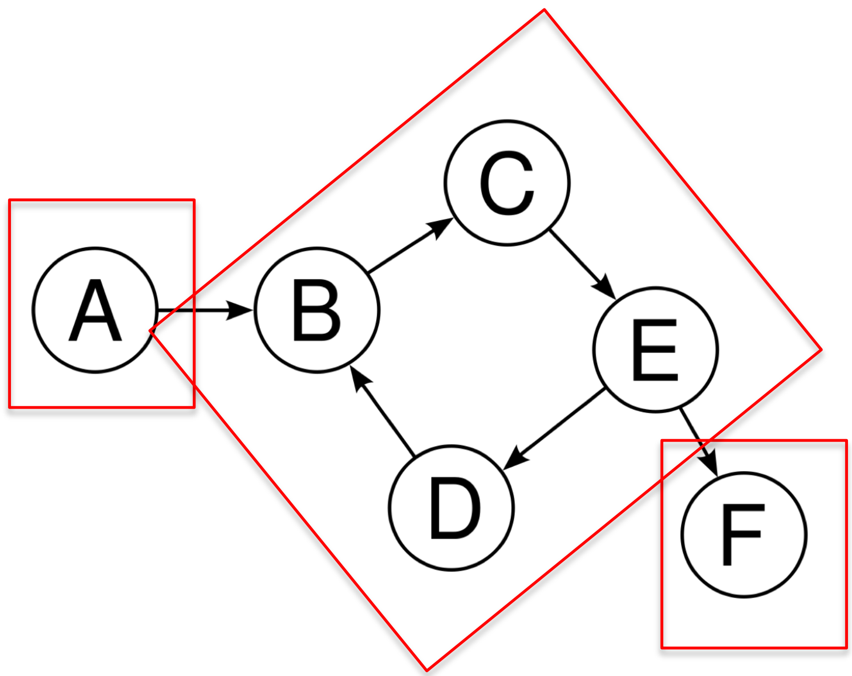
Some Examples of Giant Components in Networks



1. Film actors
2. Math coauthorship
3. Student dating
4. WWW (weakly connected)
5. Internet
6. Power grid
7. Metabolic network
8. Protein interactions

Connectedness of a Directed Graph

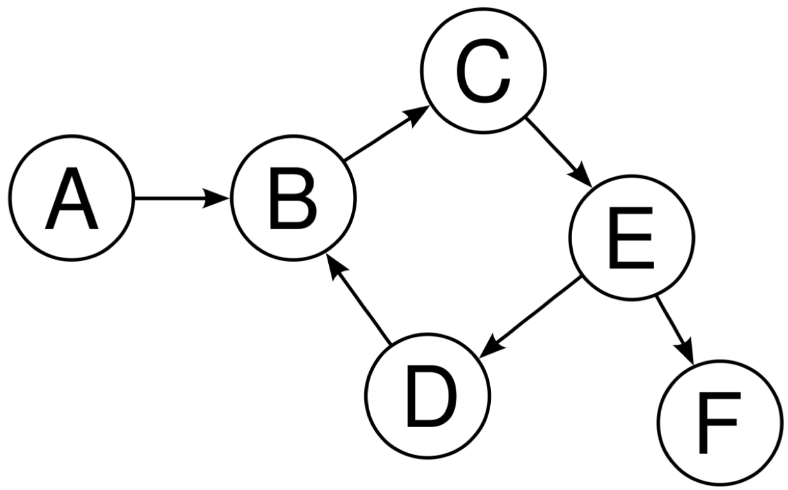
A directed graph is strongly connected if you can get from one node to any other by following the arrows.



This graph has three strongly connected components. What are they?

Connectedness of a Directed Graph

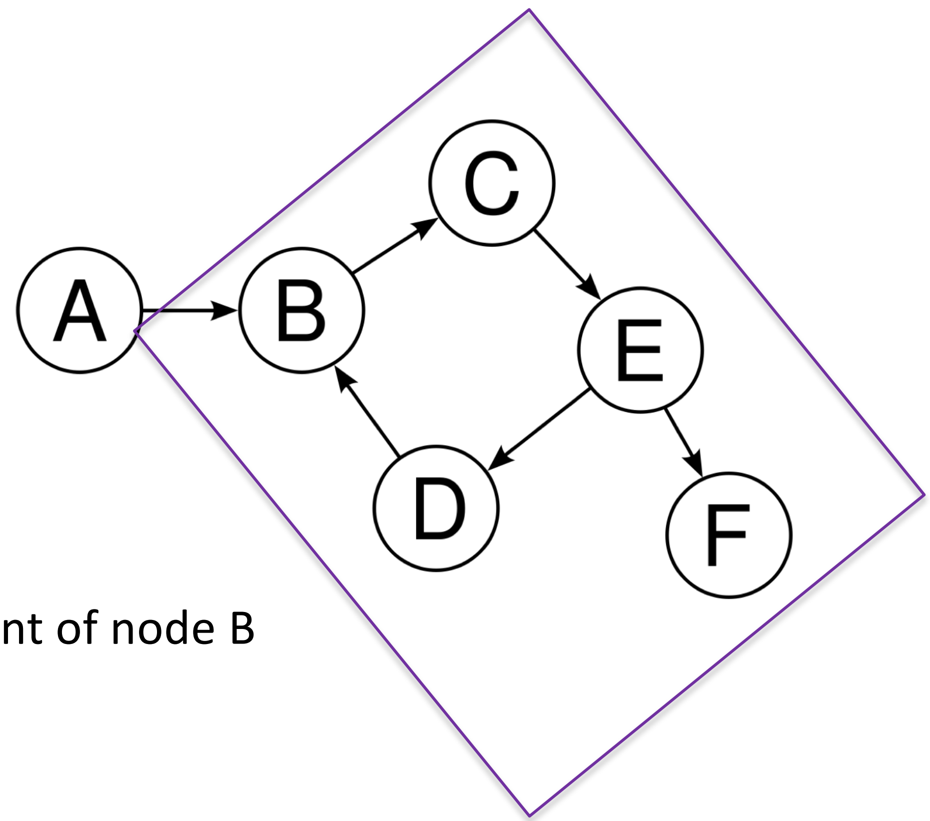
A directed graph is weakly connected if when you remove the arrow from the edges the resulting undirected graph is connected.



This graph is weakly connected.

Out-Component of a Node in a Directed Graph

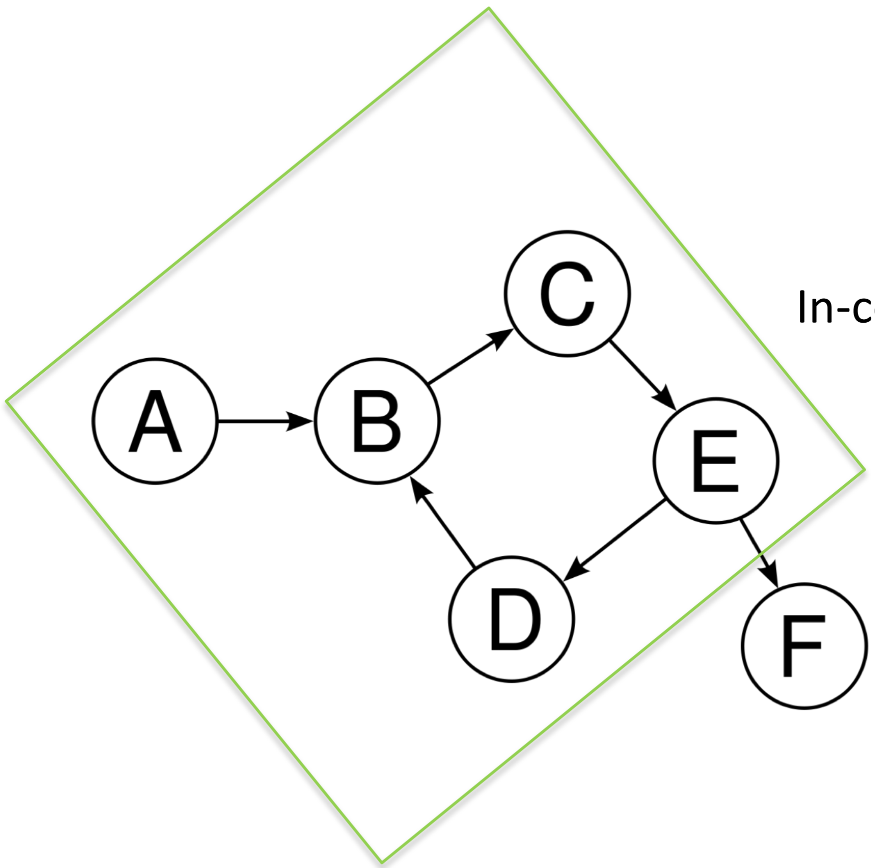
The set of nodes you can reach from node *N* by following the arrows is the out-component of node *N*.



Out-component of node B

In-Component of a Node in a Directed Graph

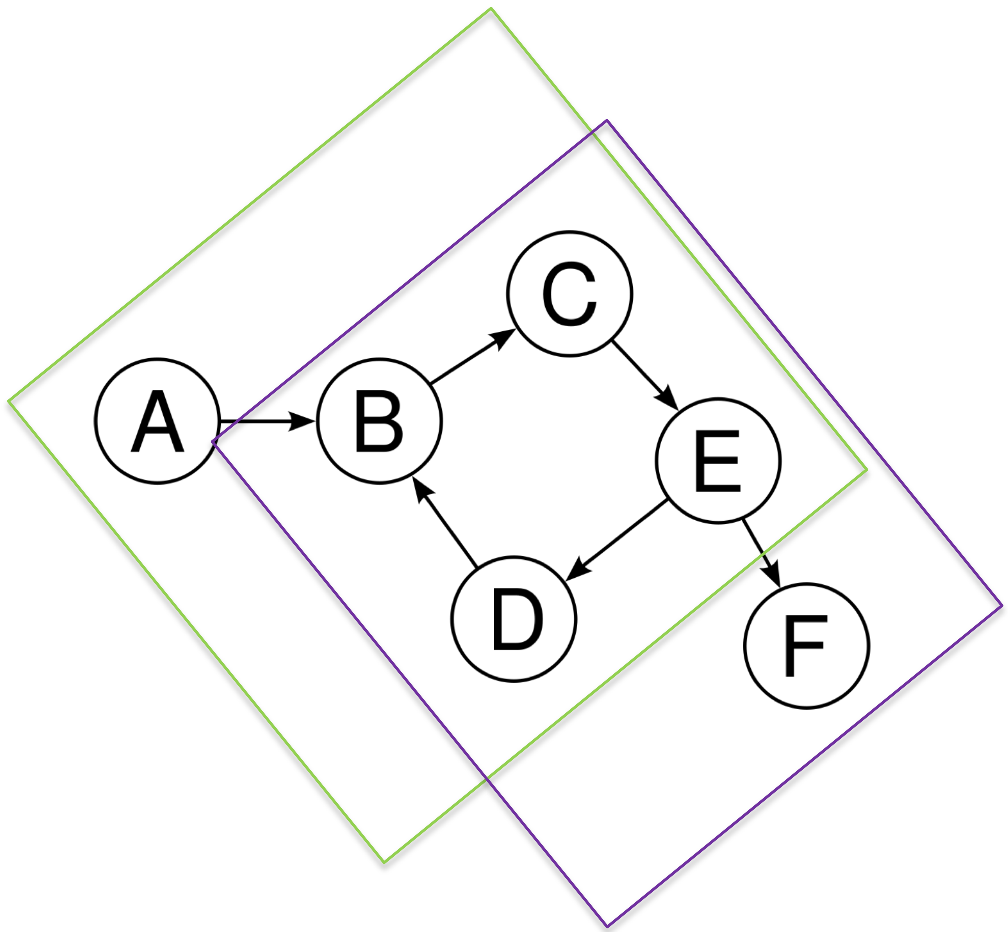
The set of nodes that can reach node *N* by following the arrows is the in-component of node *N*.



In-component of node B

In-Component/Out-Component Intersection

The strongly connected component of a directed graph that contains node *N* is the intersection of the node's in-component and out-component.

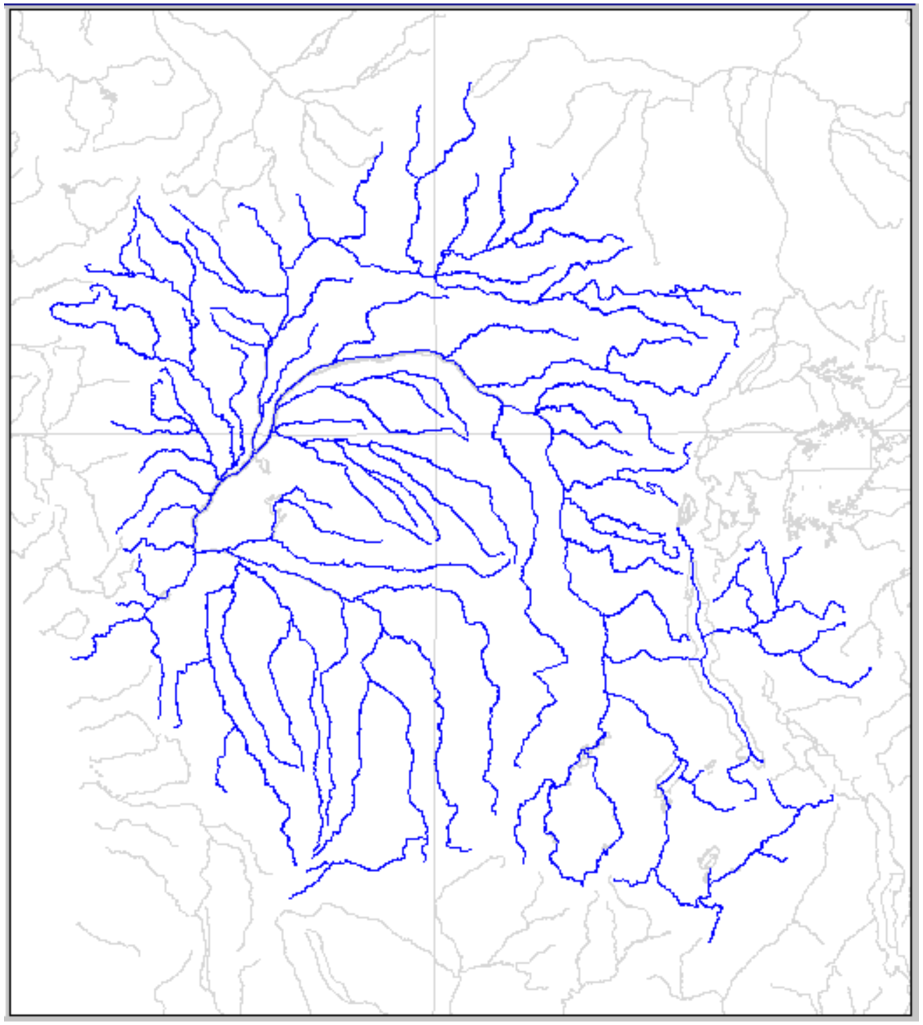


Nodes B-C-D-E are the strongly connected component containing B.

Planar Graphs and Trees

River Networks

The Congo River

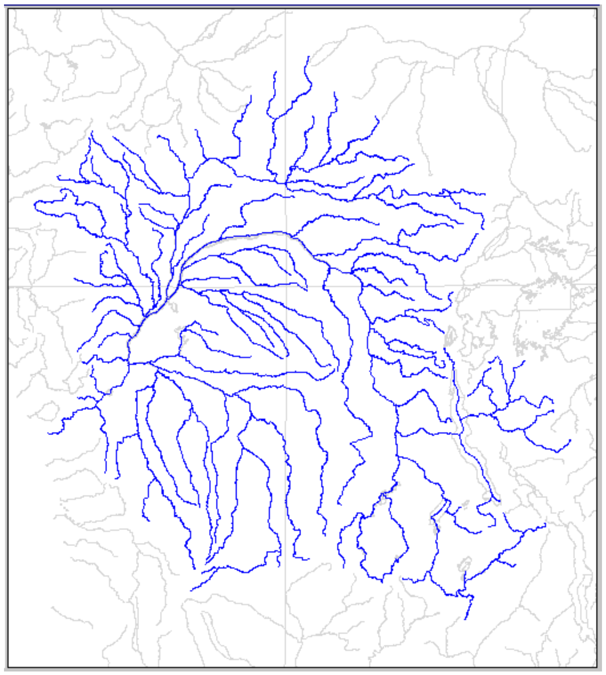


Asante & Maidment, 1999

Nodes are bifurcation points and endpoints

Edges are spans of the river

What are Some Properties?



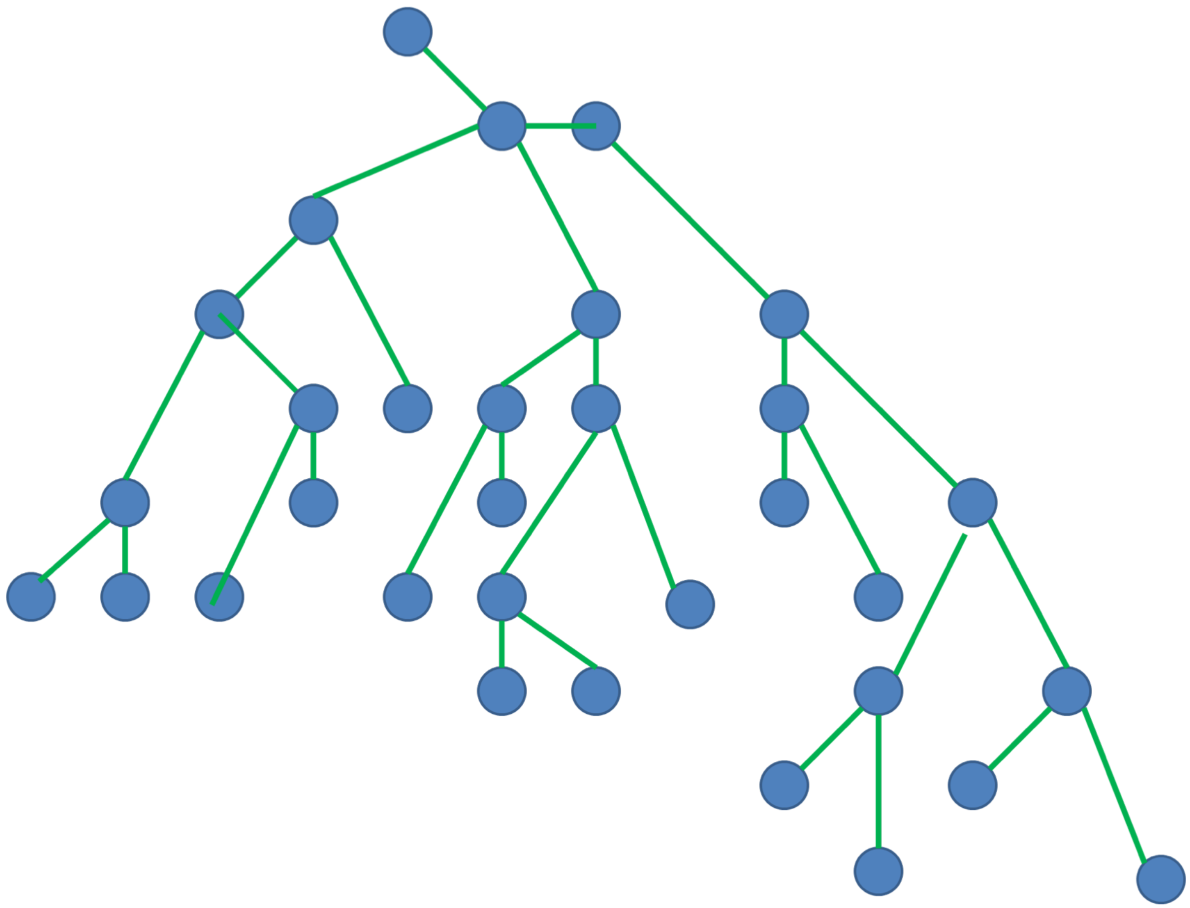
The graph is connected

Edges don't cross – such a graph is called a planar graph

There are no cycles – such a graph is called a tree

A disconnected graph with the latter two properties is just a collection of trees … called a forest.

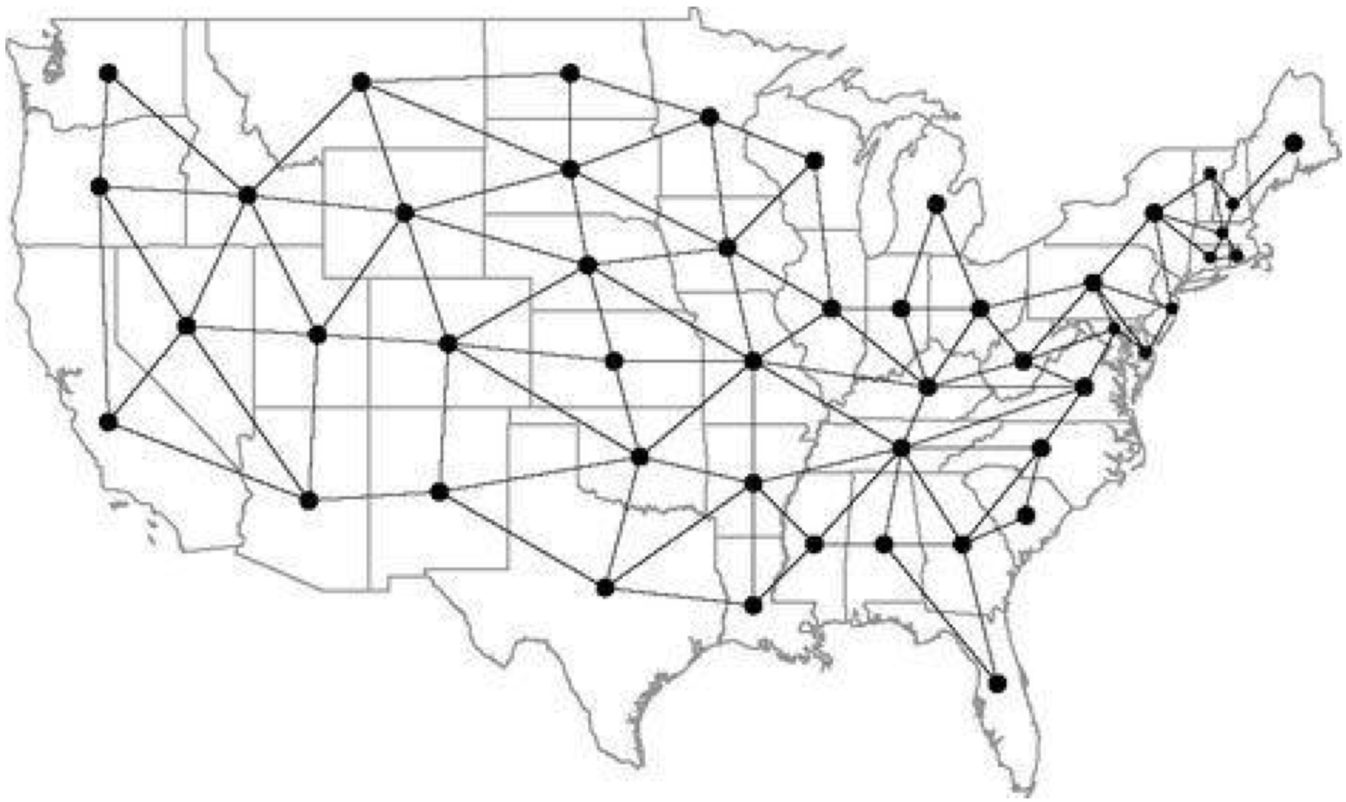
Are All Trees Planar Graphs?



Yes! If an edge crosses another just move one of the connecting nodes.

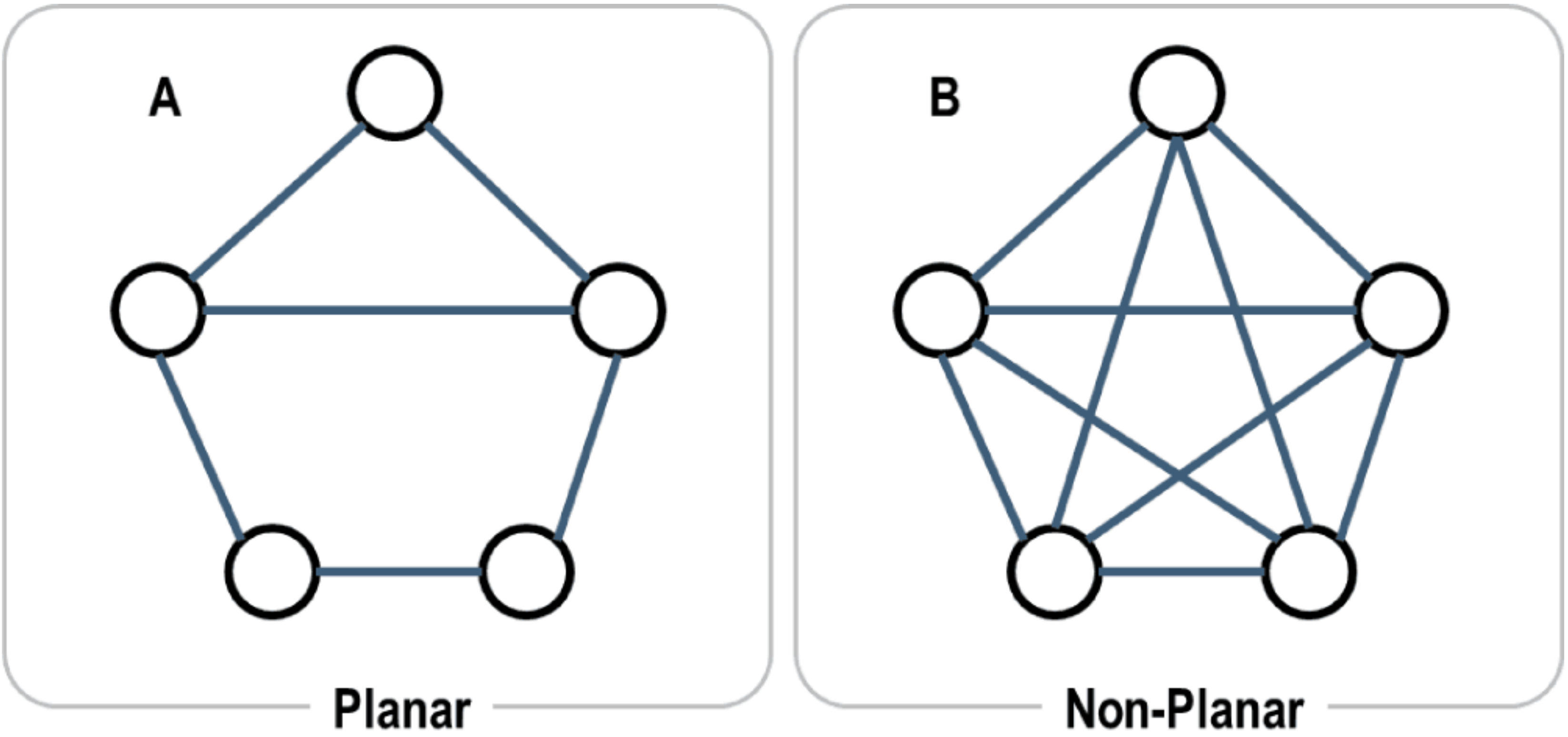
Are All Planar Graphs Trees?

Nodes are midpoints of states
Edges are connections between adjacent states



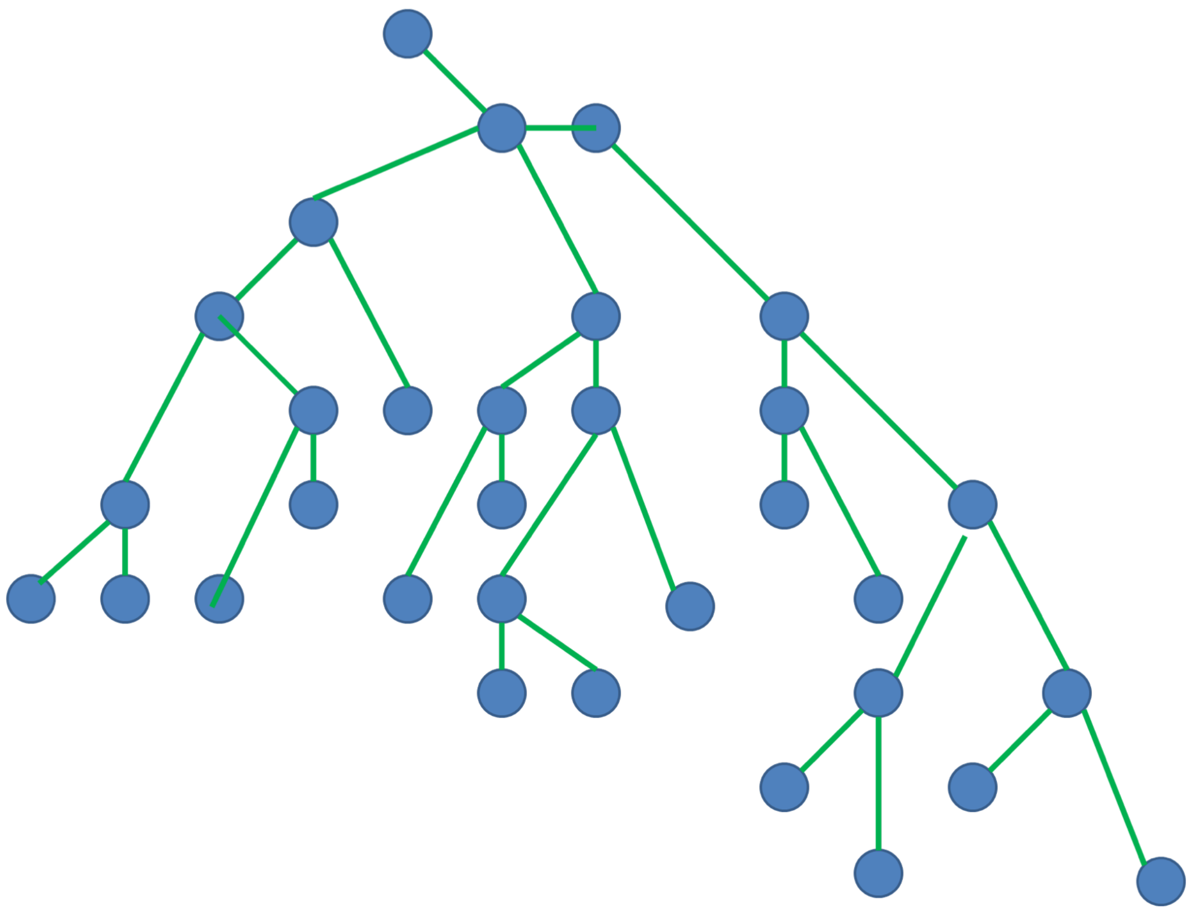
No! Lots of cycles in this planar graph

What Does a Non-Planar Graph Look Like?



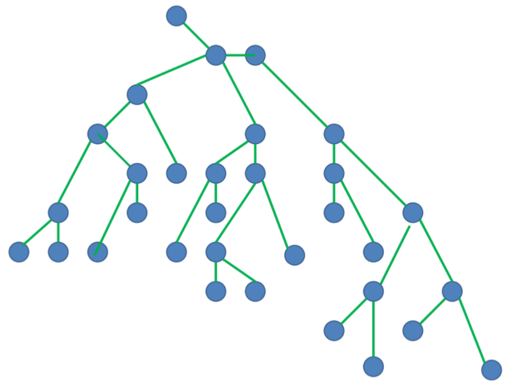
Can you redraw the edges in B so that they don't intersect?

How Many Paths Are There Between Any Two Nodes in a Tree?



Looks like exactly 1 path between any two nodes

Proof That There is Exactly One Path Between Any Two Nodes of a Tree



Let T be a tree and A and B any two nodes in T.

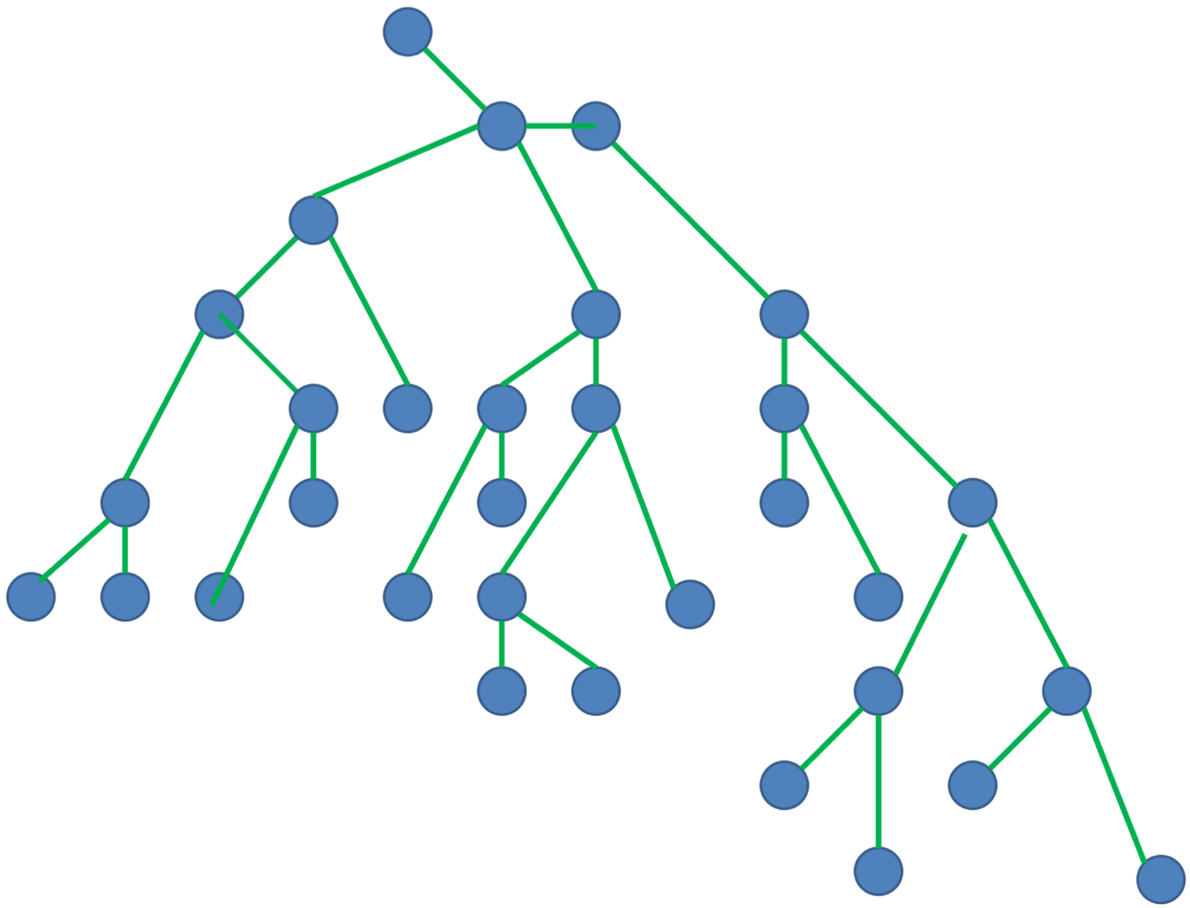
If there is no path between A and B then T is disconnected, contradicting the statement that T is a tree.

Suppose there are two or more paths between A and B, and consider path p_1 and path p_2.

Follow p_1 from A to B and then p_2 from B to A. This forms a cycle, contradicting the statement that T is a tree.

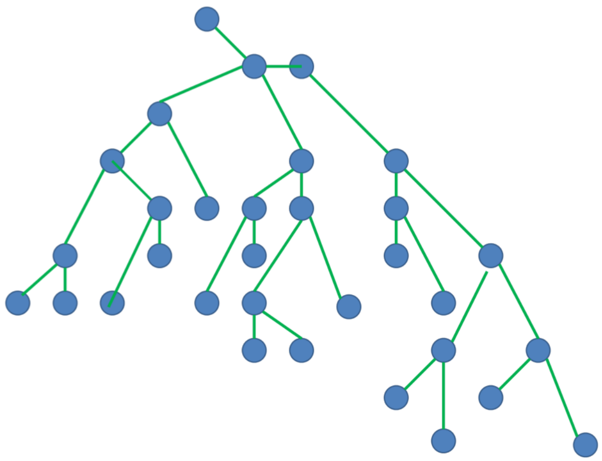
Therefore, using proof by contradiction, a tree has exactly one path between any two nodes.

How Many Edges Are There in a Tree?



Looks like edge number is 1 less than node number

Proof That a Tree With n Nodes Has n-1 edges



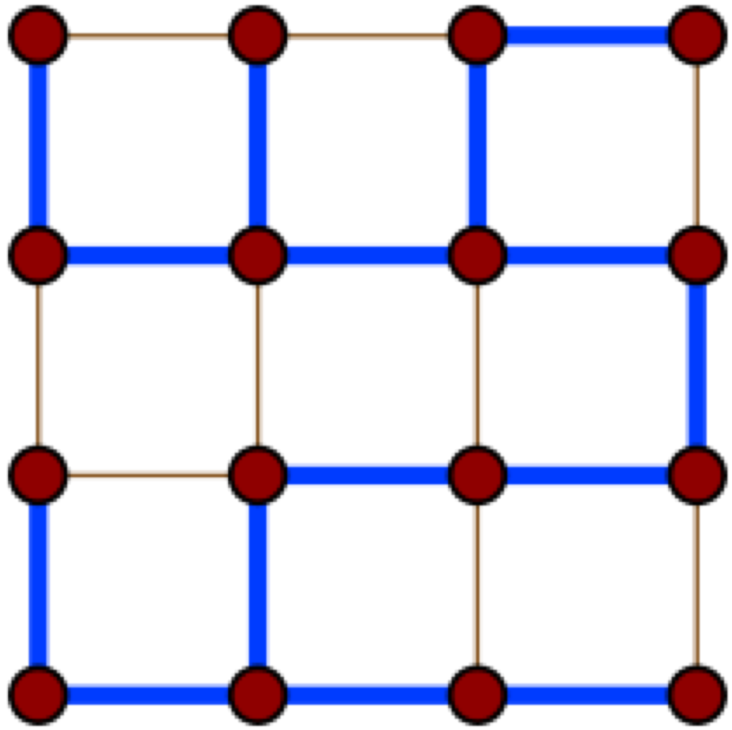
Let T be a tree with n nodes and m edges.

Suppose that m is smaller than n-1. Then there are not enough edges to connect all the nodes and T is disconnected. This contradicts the statement that T is a tree.

Suppose that m > n-1. The n-1 of these edges can be used to connect all the nodes of T, with a few left over. Connecting any two nodes with any one of these remaining edges would form a cycle, contradicting the statement that T is a tree.

Therefore, using proof by contradiction, a tree with n nodes has exactly n-1 edges.

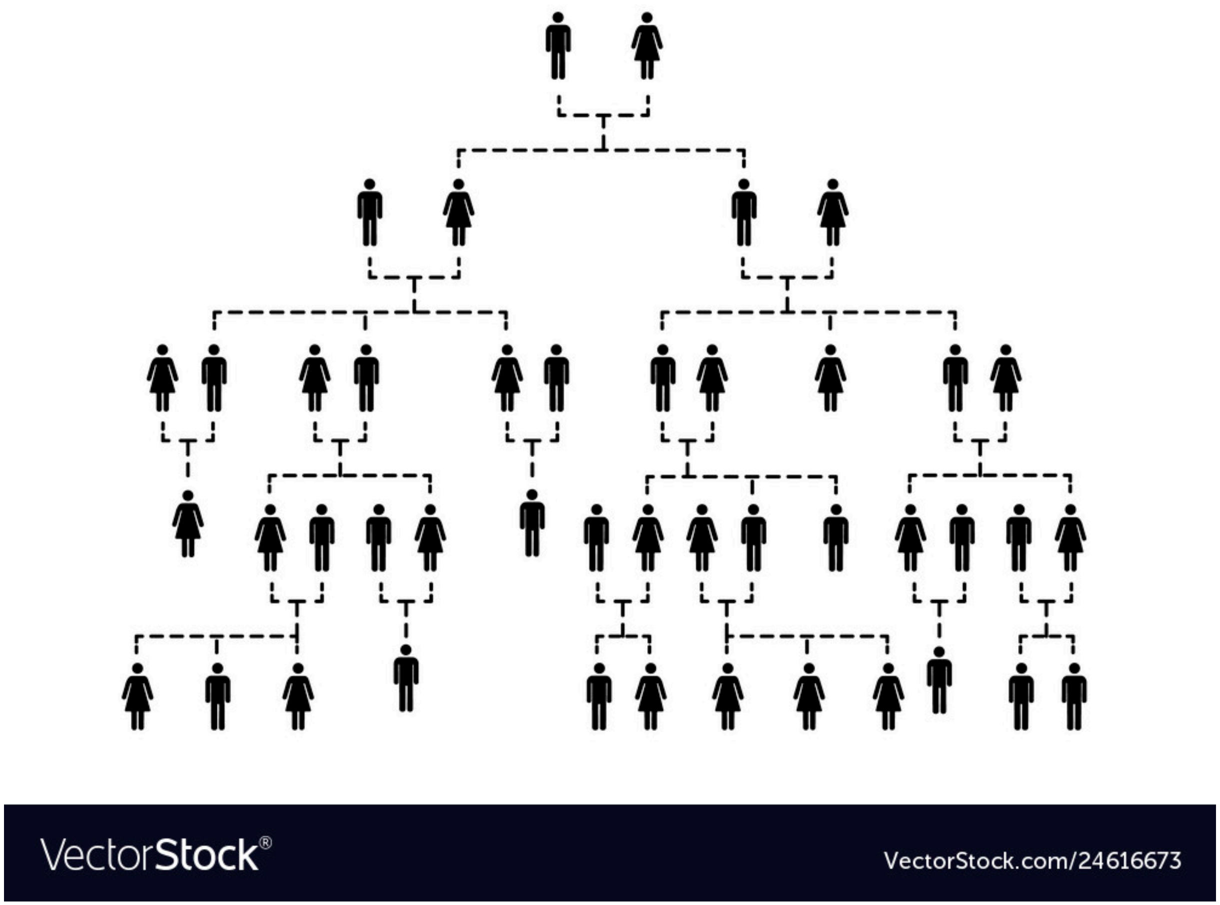
A Spanning Tree of an Undirected Graph



A spanning tree of a graph is a subgraph containing all the nodes and just enough edges so that the nodes remain connected. This subgraph is a tree.

All connected undirected graphs have at least one spanning tree.

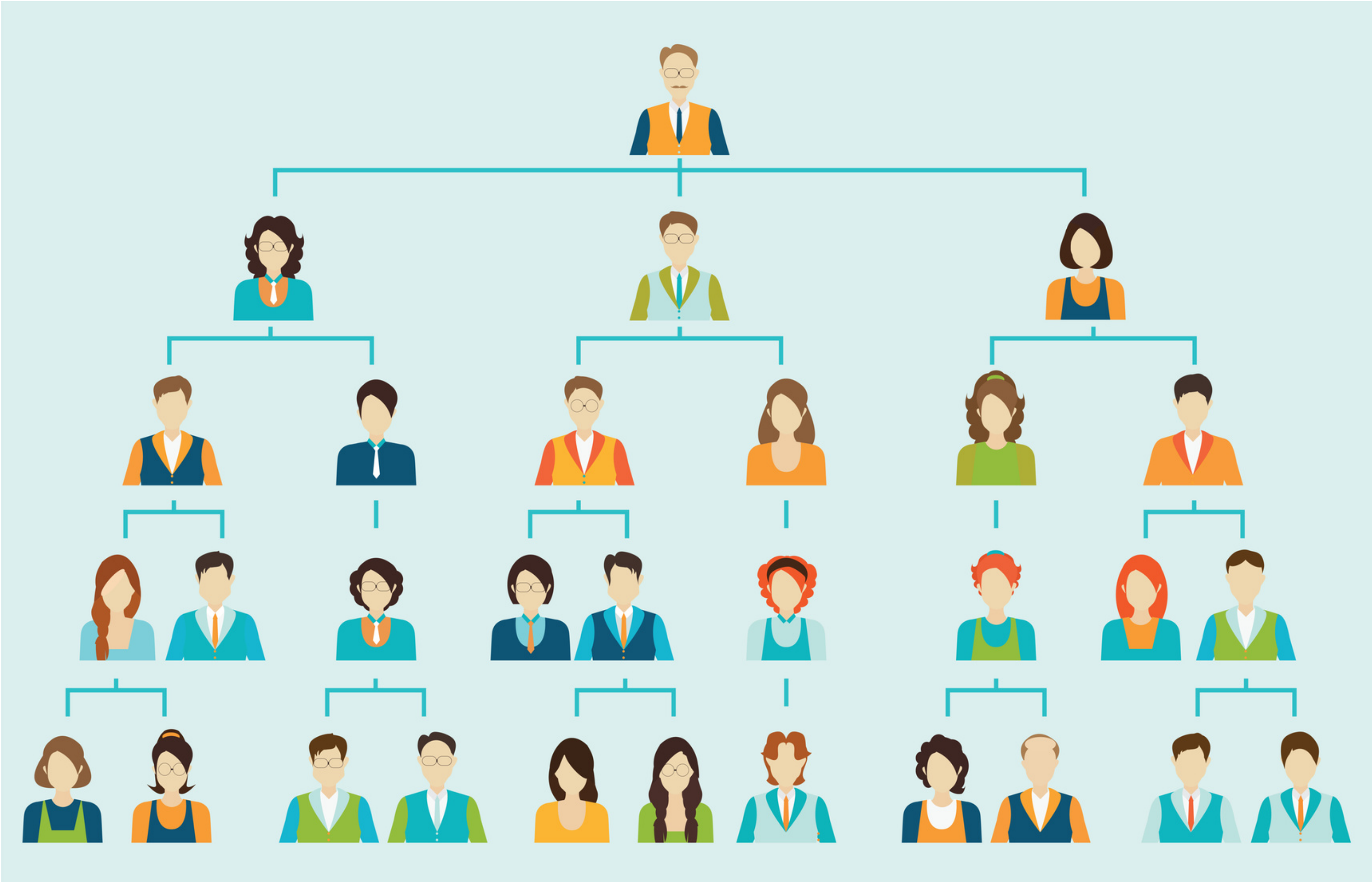
A Family Tree



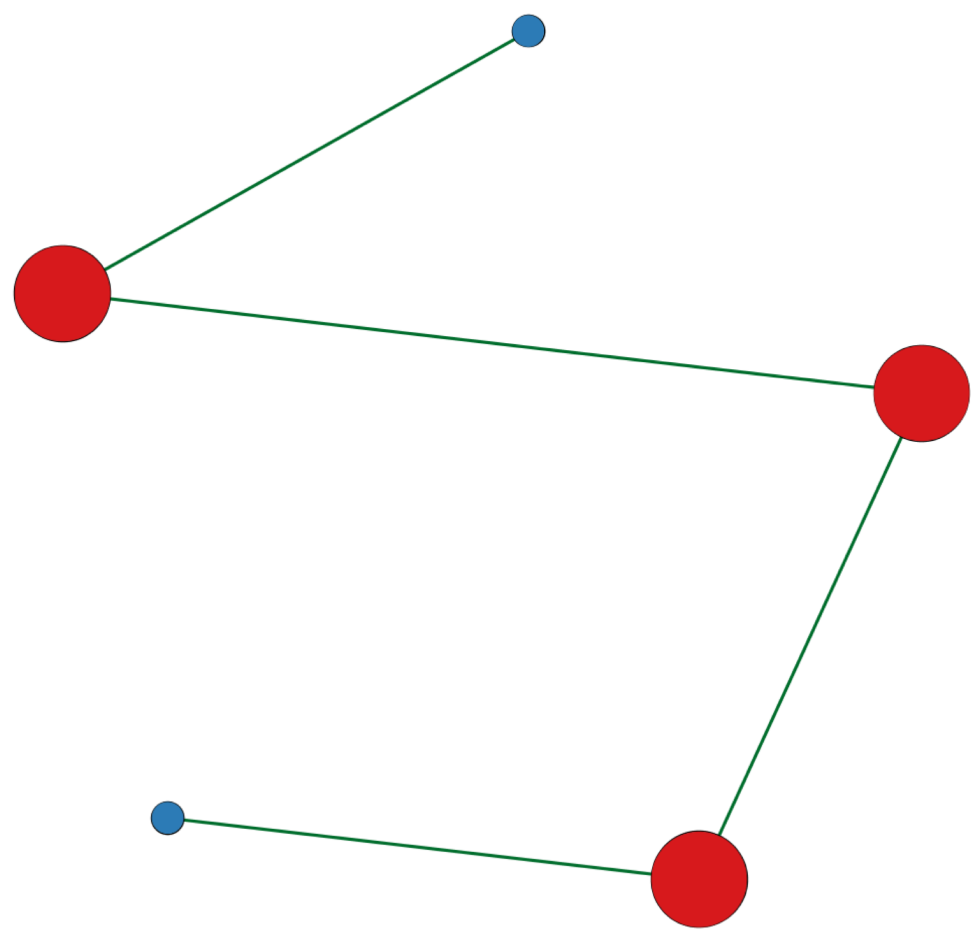
This is an example of a rooted tree, where one node (or node pair in this case) is at the top (and called the root) and others are below.

All trees can be expressed as rooted trees.

Organizational Charts are Typically Trees

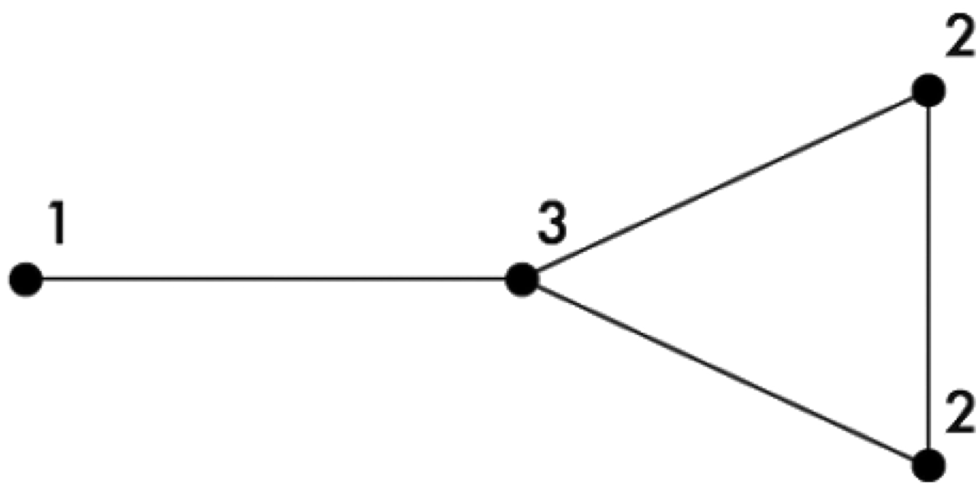


The Degrees of Nodes and Graphs



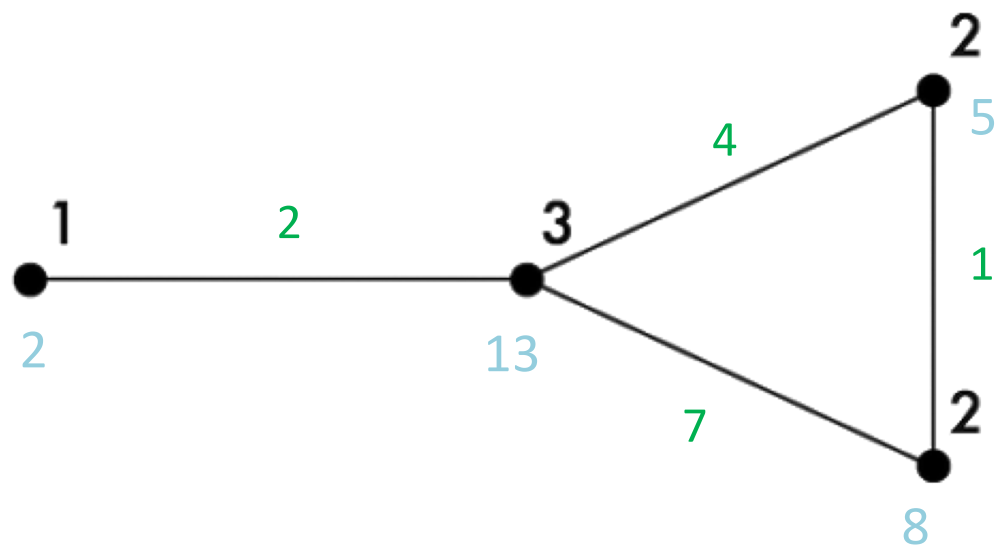
Degree of a Node in an Undirected Graph

The degree of a node is just the number of edges connected to it.



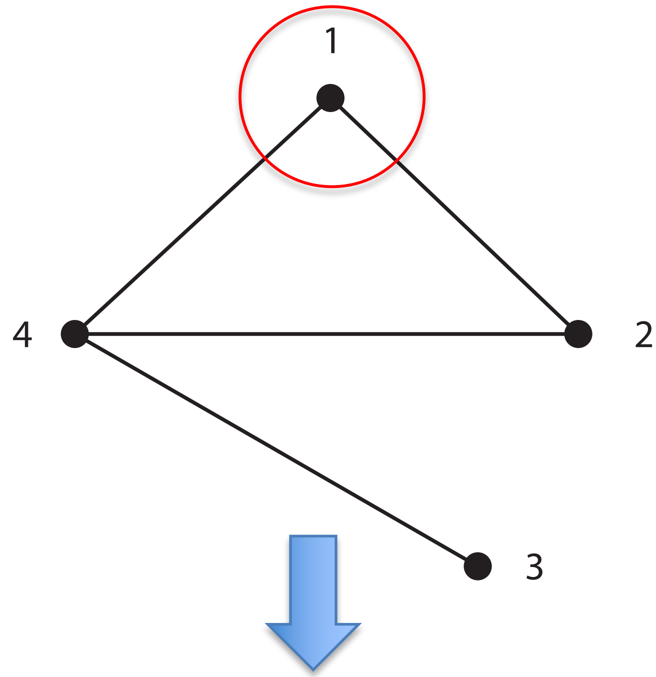
The Strength in an Weighted Graph

The strength of a node is the sum of the weights associated with all edges connected to it.



Black: degree
Green: weight
Cyan: strength

Relationship Between Degree and the Adjacency Matrix



$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

Row sum = 2

Column sum = 2

Relationship Between Degree and the Adjacency Matrix

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

Row sum = 2

Column sum = 2

d_i = degree of node i = row sum of row i = column sum of column i

$$d_i = \sum_{j=1}^{n} A_{ij} = \sum_{j=1}^{n} A_{ji}$$

Degree of an Undirected Graph (G)

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

degree(G) = sum of all elements of A

degree(G) = $\sum_{i=1}^{n} d_i$

degree(G) = 2m, since each edge is counted twice

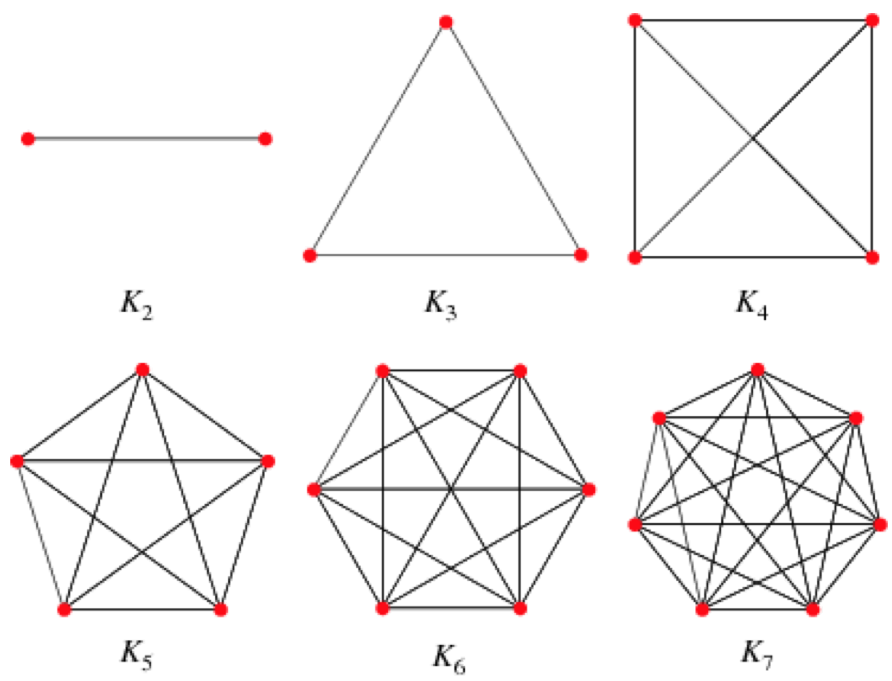
c = mean degree(G) = degree(G)/n so $c = \frac{2m}{n}$

Complete Graphs

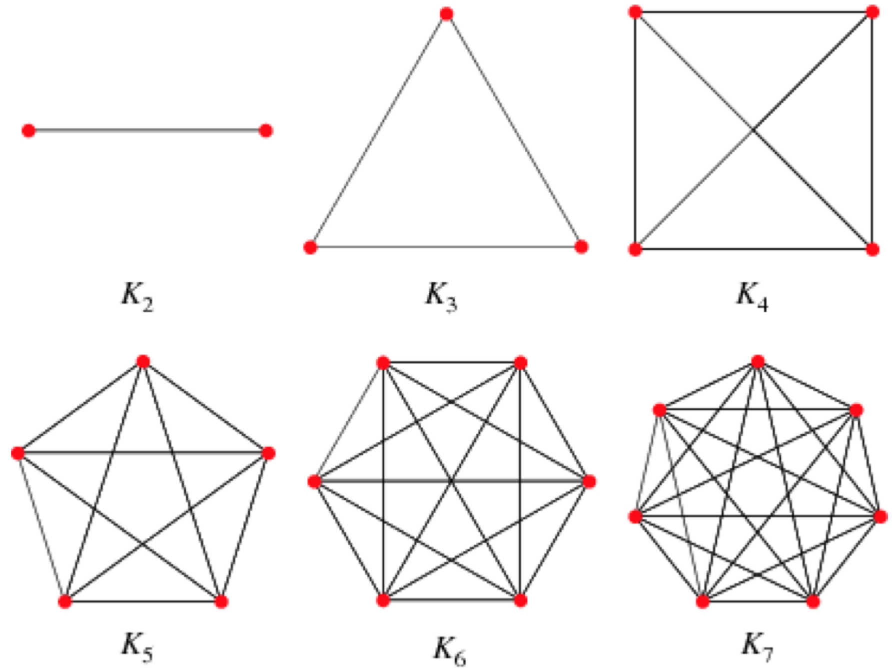
Mean degree tells us how connected the nodes are on average. But if $c = 2.7$ does that mean the graph is densely or sparsely connected?

This depends on how connected it *could be*

A complete graph (K_n) has the maximum possible number of edges (assuming the graph is simple)



Complete Graphs

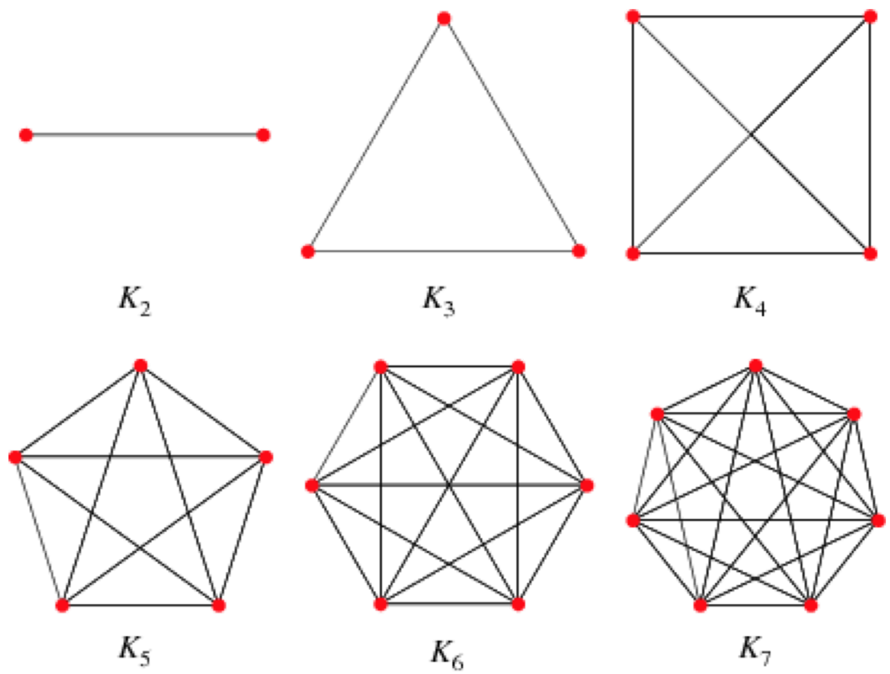


$m(K_2) = 1$
$m(K_3) = 3$
$m(K_4) = 6$

$$m(K_n) = \binom{n}{2} = \frac{1}{2}n(n-1)$$

Recall that $\binom{n}{k} \equiv \frac{n!}{k!(n-k)!}$

Degree of Complete Graphs



$$m(K_n) = \binom{n}{2} = \frac{1}{2}n(n-1)$$

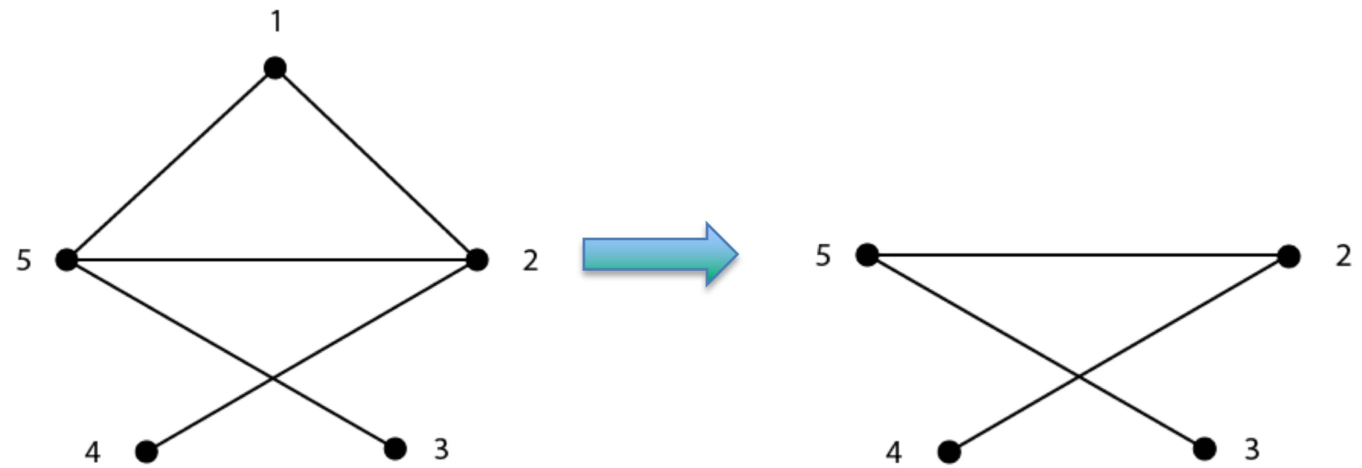
$$c = \frac{2m}{n} = n - 1$$

Each node has the same degree, which is the mean degree!

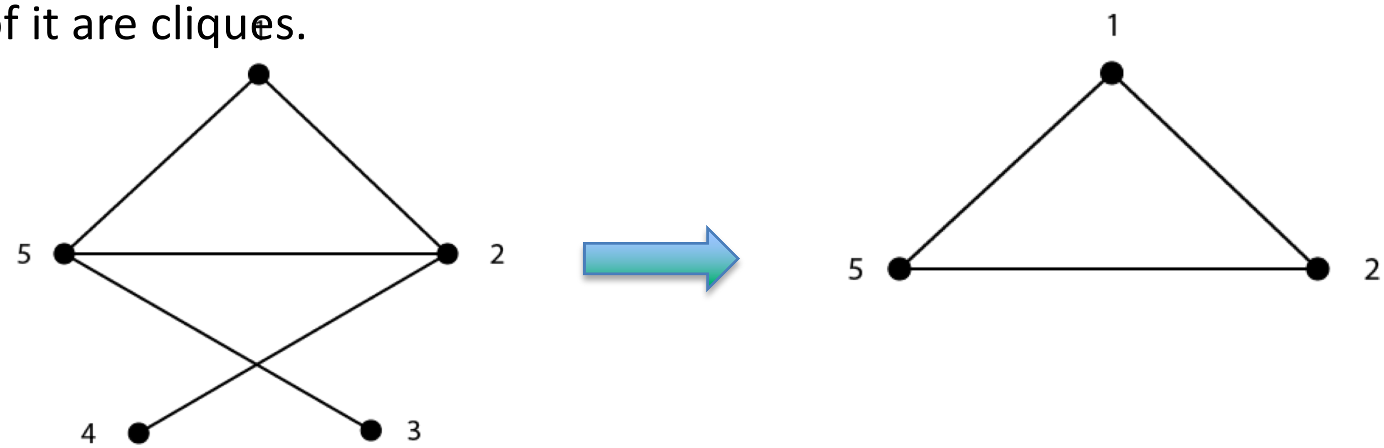
$d_i = n - 1$ for each i

Subgraphs

A subgraph or subnetwork is obtained by selecting a subset of the nodes of a graph and all of the edges among those nodes.



A clique is a fully connected subgraph of a graph (i.e., a complete subgraph). If the original graph is itself complete, then all subgraphs of it are cliques.



Graph Density

The graph density is the ratio of number of edges to the possible number of edges

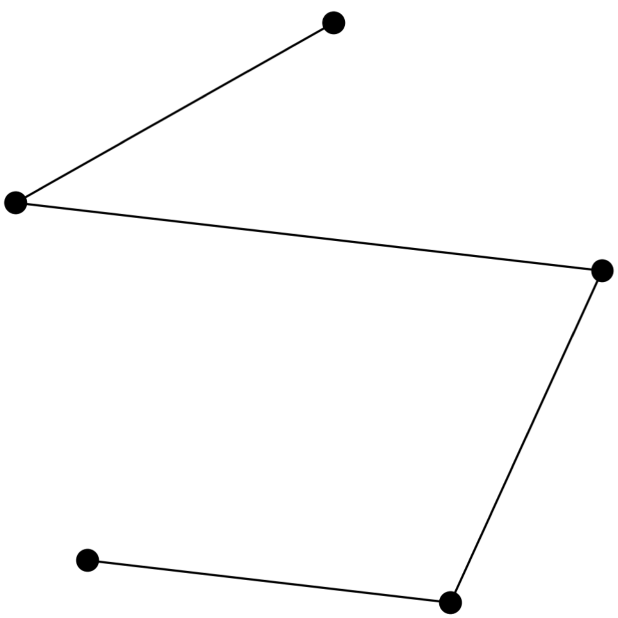
$$\rho \equiv \frac{m(G)}{m(K_n)} = \frac{m}{\binom{n}{2}} = \frac{2m}{n(n-1)}$$

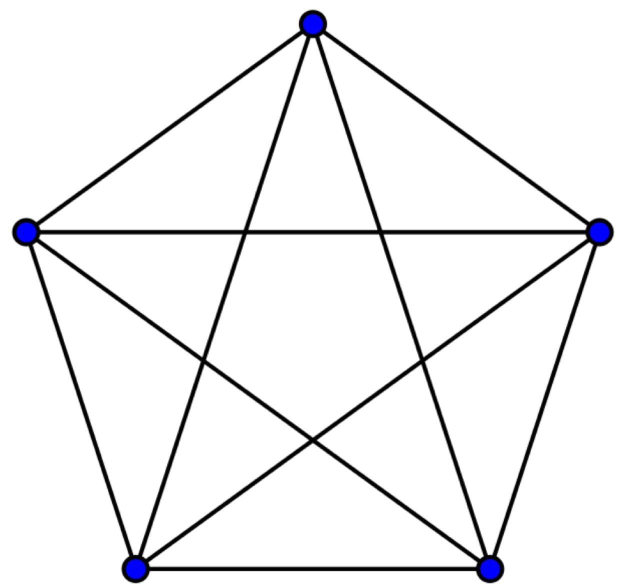
and in terms of mean degree,

$$\boxed{\rho = \frac{c}{n-1}}$$

Graph Density Examples

$$\rho = \frac{c}{n-1}$$





$$c = \frac{8}{5}$$

$$\rho = \frac{8}{5} \Big/ 4 = \frac{2}{5}$$

$$c = 4$$

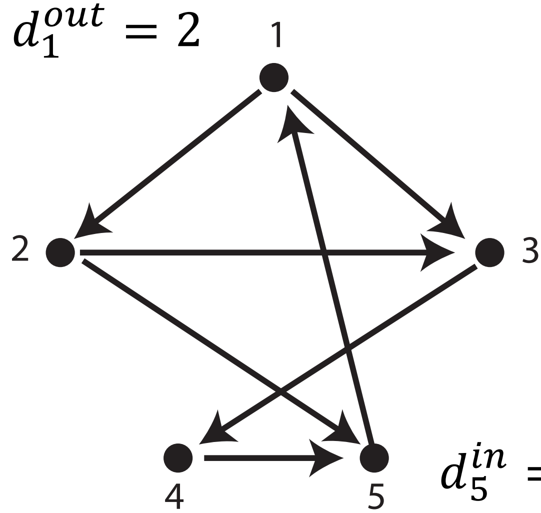
$$\rho = \frac{4}{4} = 1$$ Max density

Degree of a Node in a Directed Graph

Each node of a directed graph has an in degree (number of incoming edges) and an out degree (number of outgoing edges).

$d_1^{in} = 1$

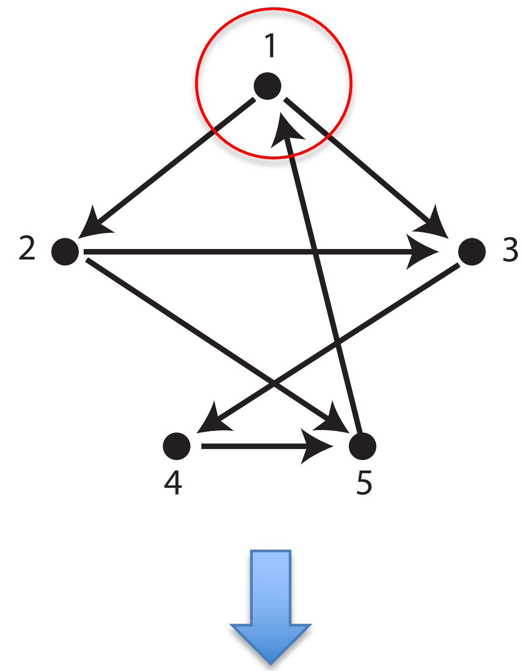
$d_1^{out} = 2$



$d_5^{in} = 2$

$d_5^{out} = 1$

Relationship to Adjacency Matrix



$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

out degree = row sum

in degree = column sum

Relationship to Adjacency Matrix

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Out degree = row sum

In degree = column sum

$$d_i^{out} = \sum_{j=1}^{n} A_{ij}$$ Row sum

$$d_i^{in} = \sum_{j=1}^{n} A_{ji}$$ Column sum

Degrees of a Directed Graph (G)

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\left.\begin{aligned} G^{out} &= \sum_{i=1}^{n} d_i^{out} = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij} \\ G^{in} &= \sum_{i=1}^{n} d_i^{in} = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ji} \end{aligned}\right\} \Rightarrow \boxed{G^{out} = G^{in} = m}$$

Mean Degree of a Directed Graph

$$c^{in} \equiv \frac{1}{n}\sum_{i=1}^{n} d_i^{in} = \frac{1}{n} G^{in}$$

$$c^{out} \equiv \frac{1}{n}\sum_{i=1}^{n} d_i^{out} = \frac{1}{n} G^{out}$$

but $G^{in} = G^{out} = m$

so $c^{in} = c^{out} = c$

and $c = \frac{m}{n}$

This is half of what it would be if edges were not directed

Graph Density of a Directed Graph

Starting with the definition of graph density as the ratio of number of edges to the possible number of edges, note that between any 2 nodes there are now 2 possible edges. So

$$\rho \equiv \frac{m(G)}{m(K_n)} = \frac{m}{2\binom{n}{2}} = \frac{m}{n(n-1)}$$

and in terms of mean degree,

$$\boxed{\rho = \frac{c}{n-1}}$$

This is the same formula as for an undirected graph.

The End