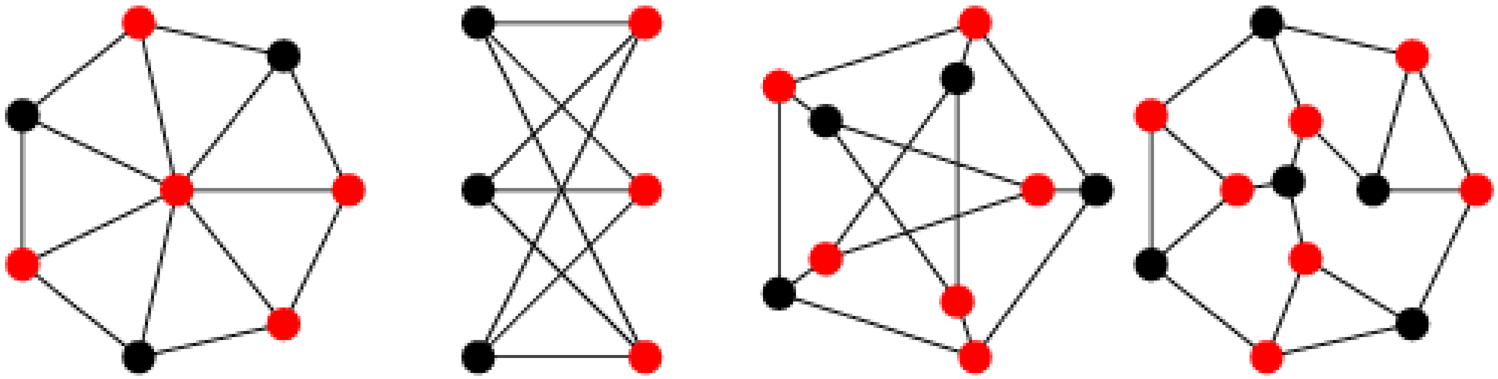


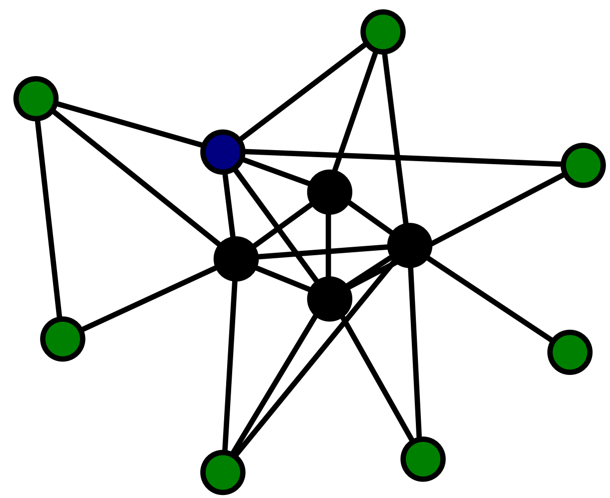
Mathematical Foundations

Steve Borgatti

Revised 7/04 in Colchester, U.K.

Graphs

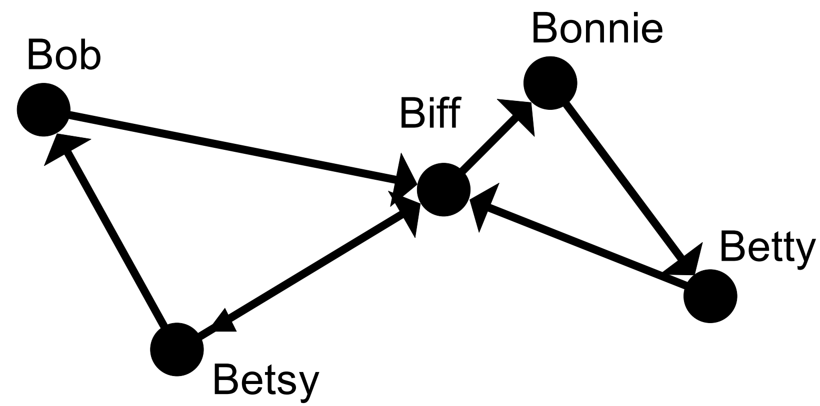
- Networks represented mathematically as graphs
- A graph G(V,E) consists of …
  - Set of nodes|vertices V representing actors
  - Set of lines|edges E representing ties
    - An edge is an unordered pair of nodes (u,v)
    - Nodes u and v adjacent if $(u,v) \in E$
    - So E is subset of set of all pairs of nodes



- Typically drawn without arrow heads

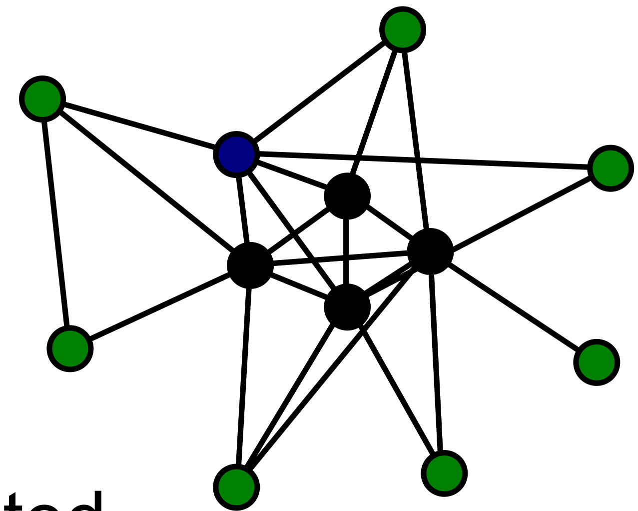
Digraphs

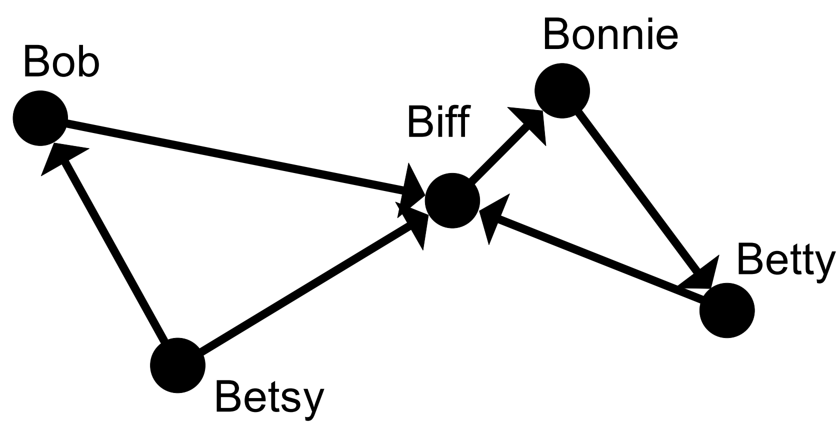
- Digraph D(V,E) consists of …
  - Set of nodes V
  - Set of directed arcs E
    - An arc is an ordered pair of nodes (u,v)
    - (u,v) $\in$ E indicates u sends arc to v
    - (u,v) $\in$ E does not imply that (v,u) $\in$ E
- Ties drawn with arrow heads, which can be in both directions



Directed vs undirected graphs

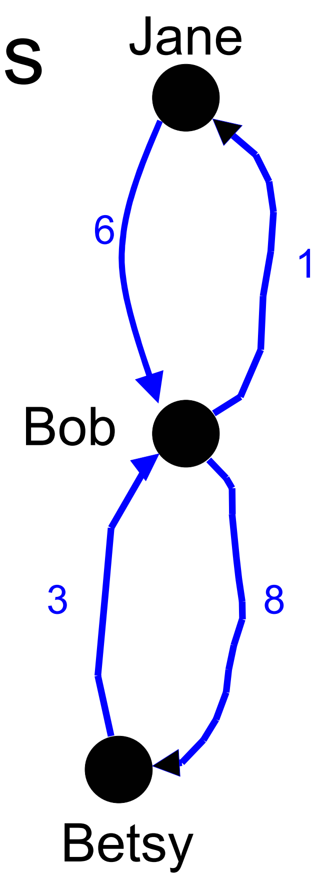
- Undirected relations
  - Attended meeting with
  - Communicates daily with
- Directed relations
  - Lent money to
- Logically vs empirically directed ties
  - Empirically, even un-directed relations can be non-symmetric due to measurement error





Strength of Tie

- We can attach values to ties, representing quantitative attributes
  - Strength of relationship
  - Information capacity of tie
  - Rates of flow or traffic across tie
  - Distances between nodes
  - Probabilities of passing on information
  - Frequency of interaction
- Valued graphs or vigraphs



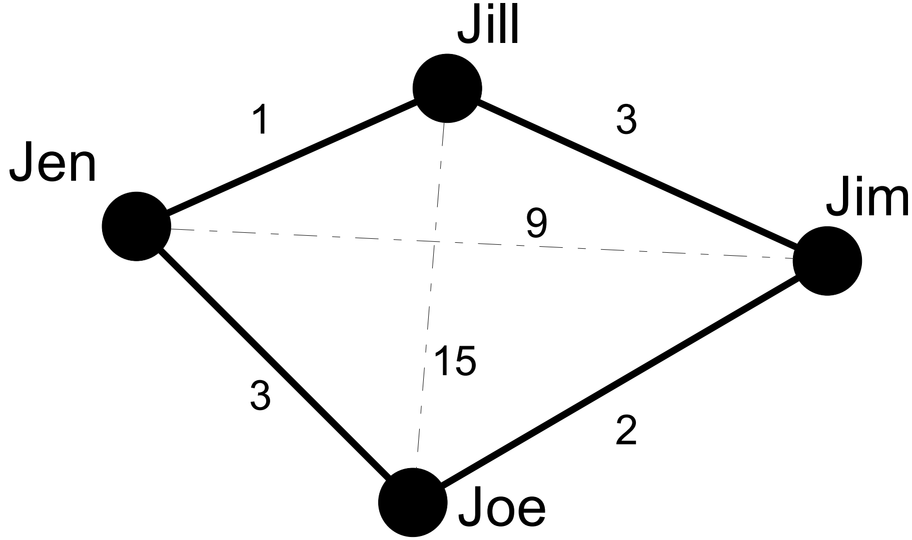
Adjacency Matrices

Friendship

	Jim	Jill	Jen	Joe
Jim	-	1	0	1
Jill	1	-	1	0
Jen	0	1	-	1
Joe	1	0	1	-

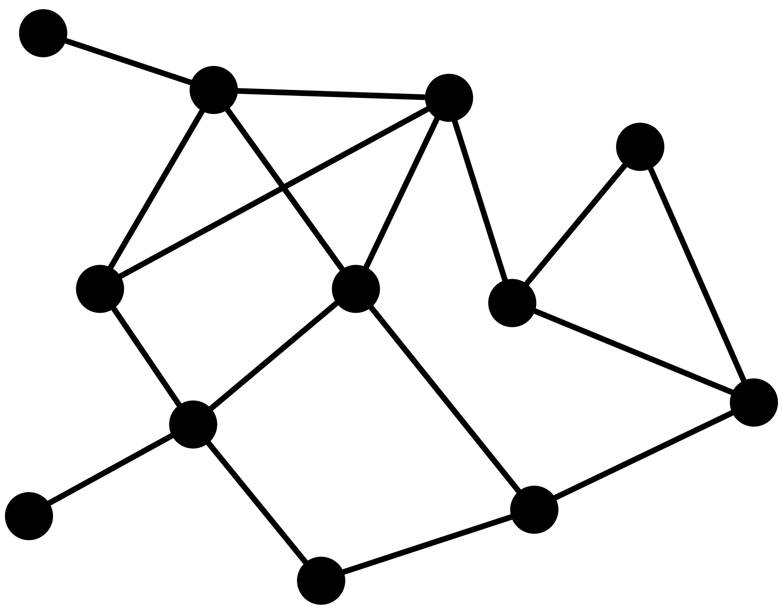
Proximity

	Jim	Jill	Jen	Joe
Jim	-	3	9	2
Jill	3	-	1	15
Jen	9	1	-	3
Joe	2	15	3	-

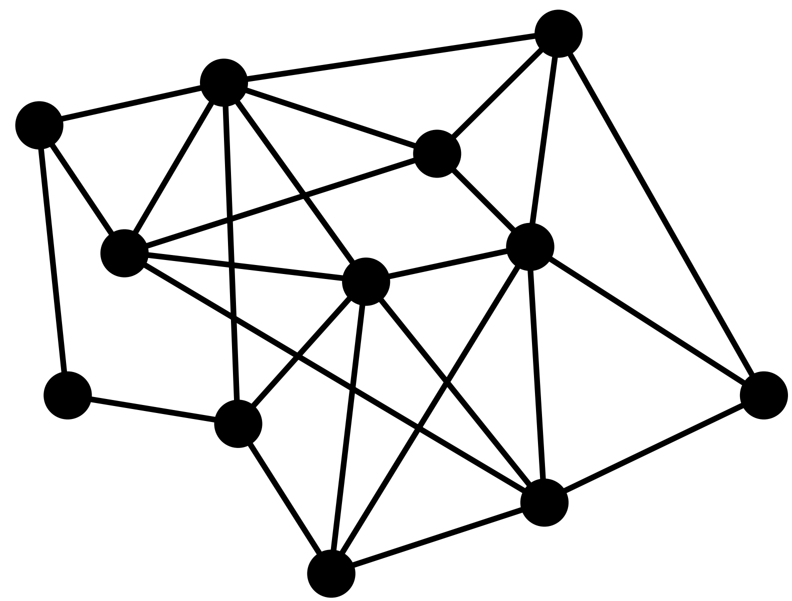


Density

- Number of ties, expressed as percentage of the number of ordered/unordered pairs

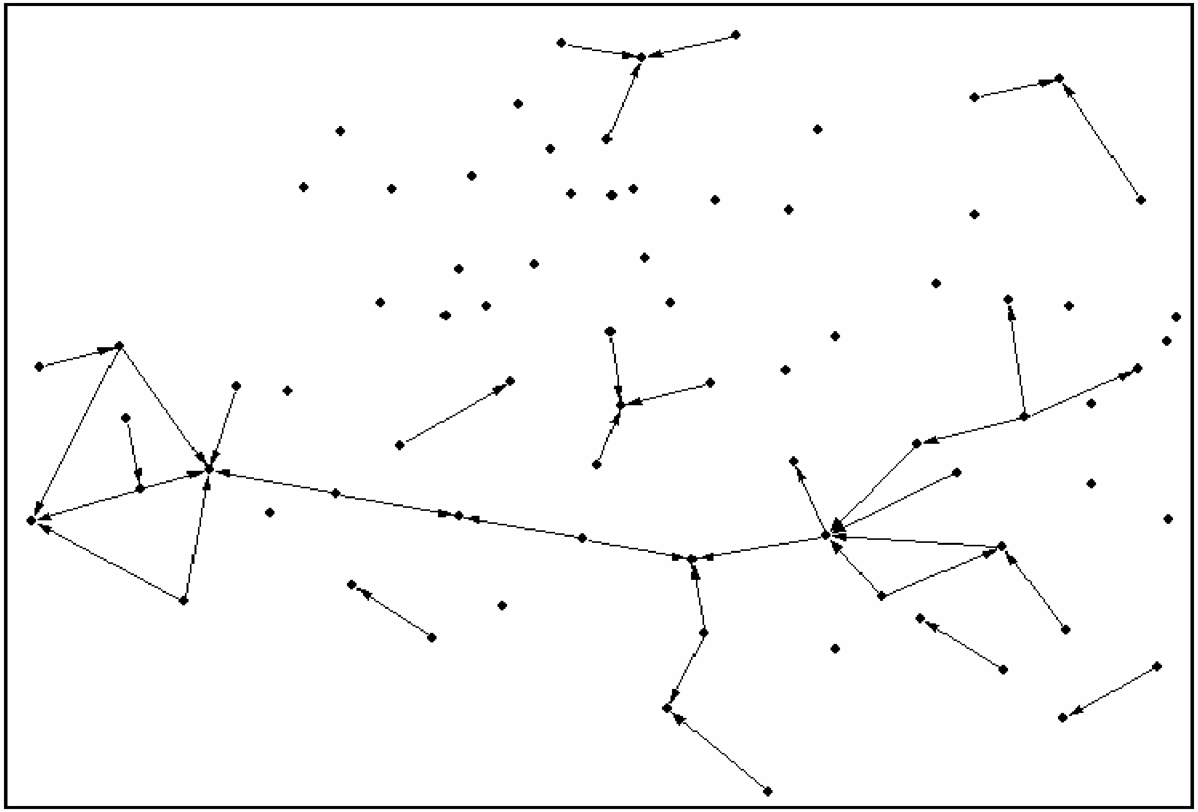


Low Density (25%)
Avg. Dist. = 2.27



High Density (39%)
Avg. Dist. = 1.76

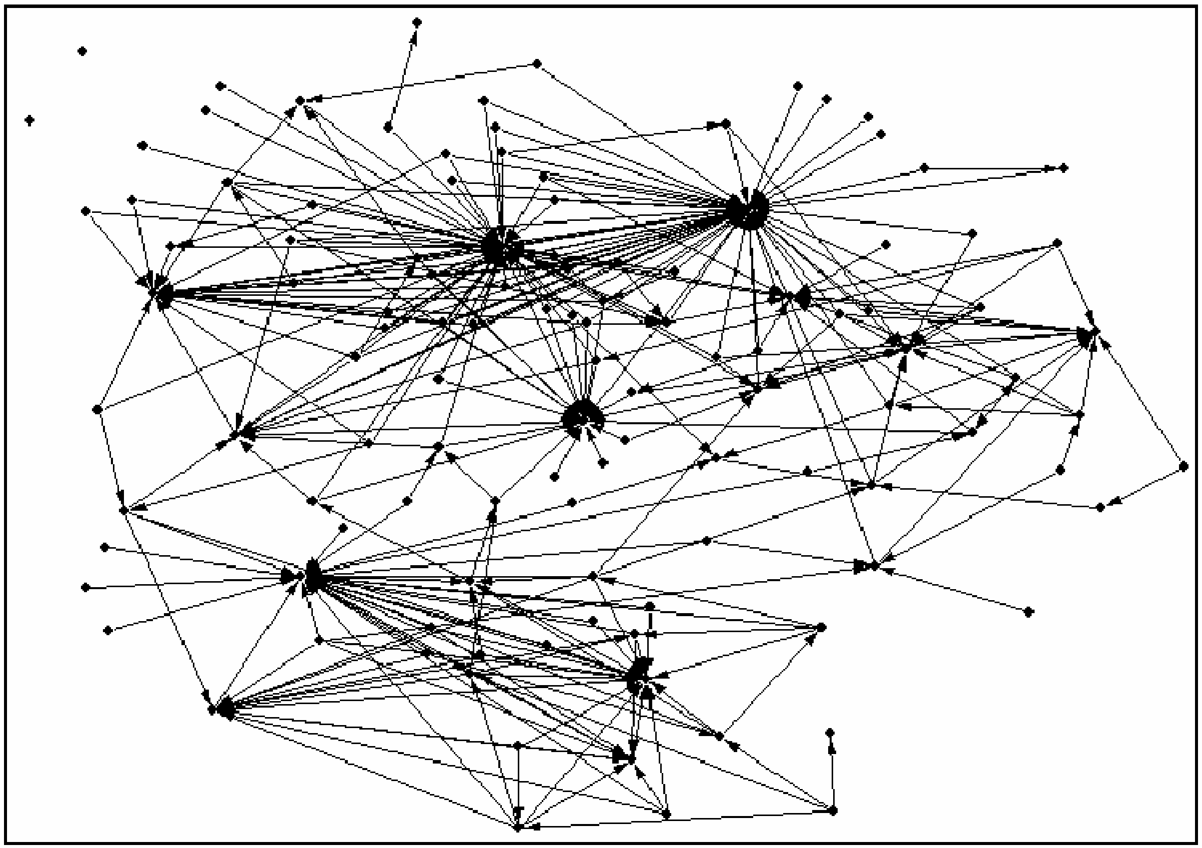
Help With the Rice Harvest



Village 1

Data from Entwistle et al

Help With the Rice Harvest



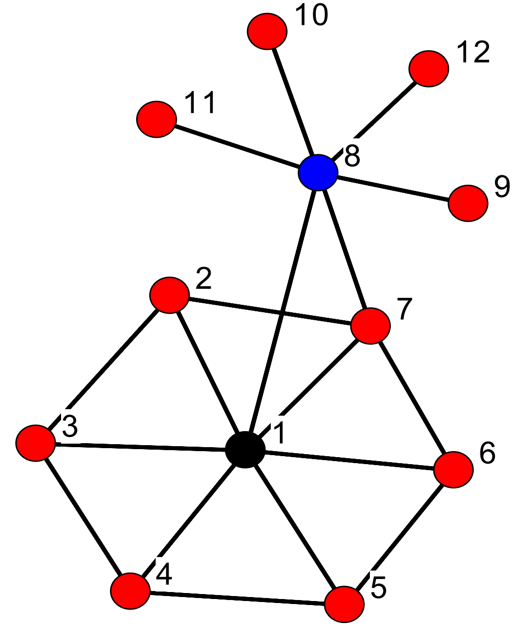
Village 2

Which village is more likely to survive?

Data from Entwistle et al

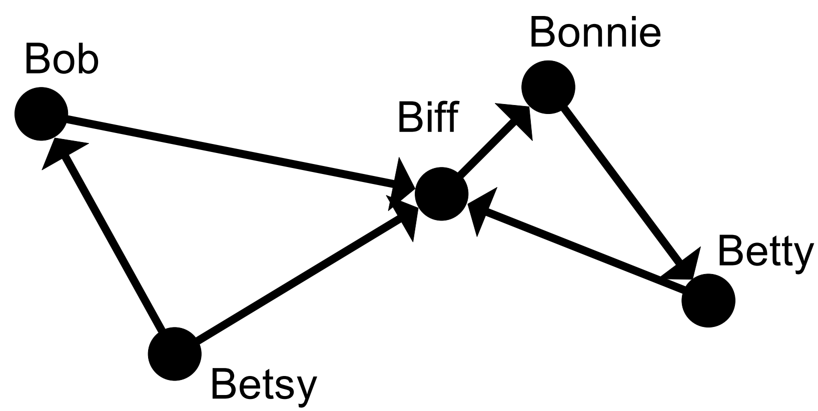
Degree

- Number of edges incident upon a vertex
  - $d_8 = 6$, while $d_{10} = 1$
- Sum of degrees of all nodes is twice the number of edges in graph
- Average degree = density times (n-1)



InDegree & OutDegree
(Directed graphs only)

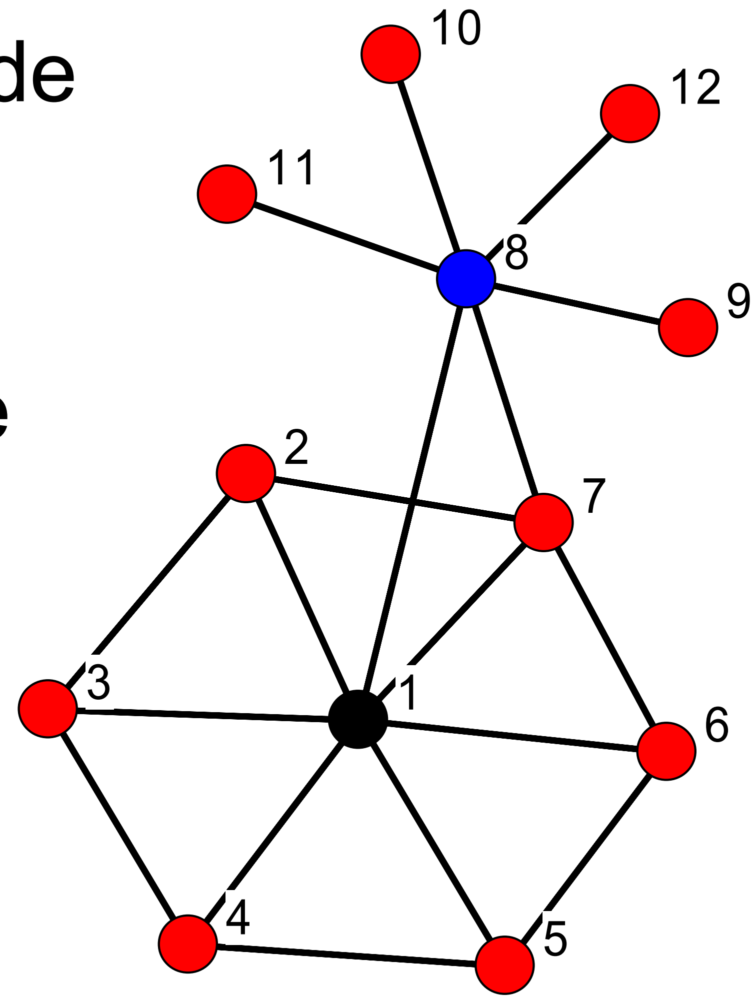
- <u>Indegree</u> is number of arcs that terminate at the node (incoming ties)
  - Indeg(biff) = 3
- <u>Outdegree</u> is number of arcs that originate at the node (outgoing ties)
  - Outdeg(biff) = 1



Average indegree always equals average outdegree

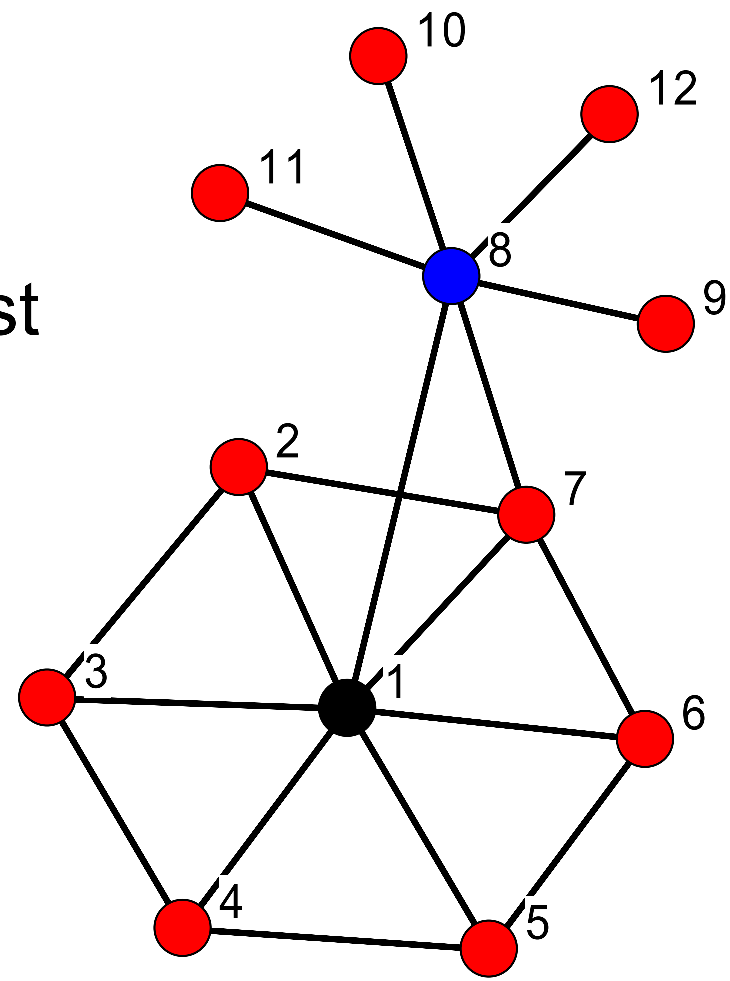
Walks, Trails, Paths

- Path: can't repeat node
  - 1-2-3-4-5-6-7-8
  - Not 7-1-2-3-7-4
- Trail: can't repeat line
  - 1-2-3-1-7-8
  - Not 7-1-2-7-1-4
- Walk: unrestricted
  - 1-2-3-1-2-7-1-7-1



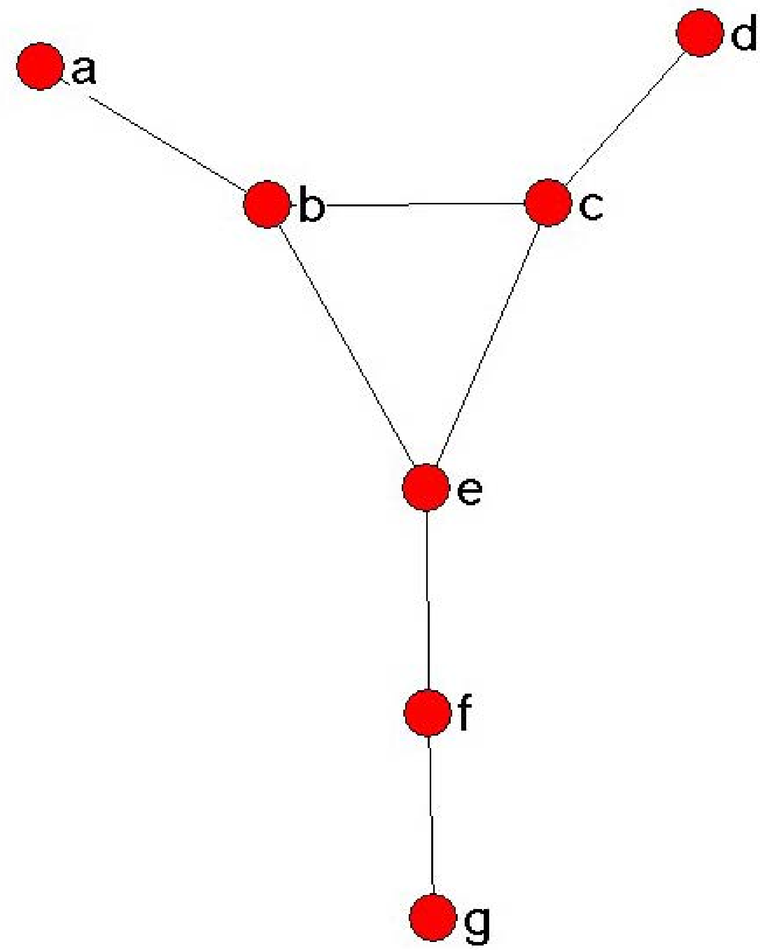
Length & Distance

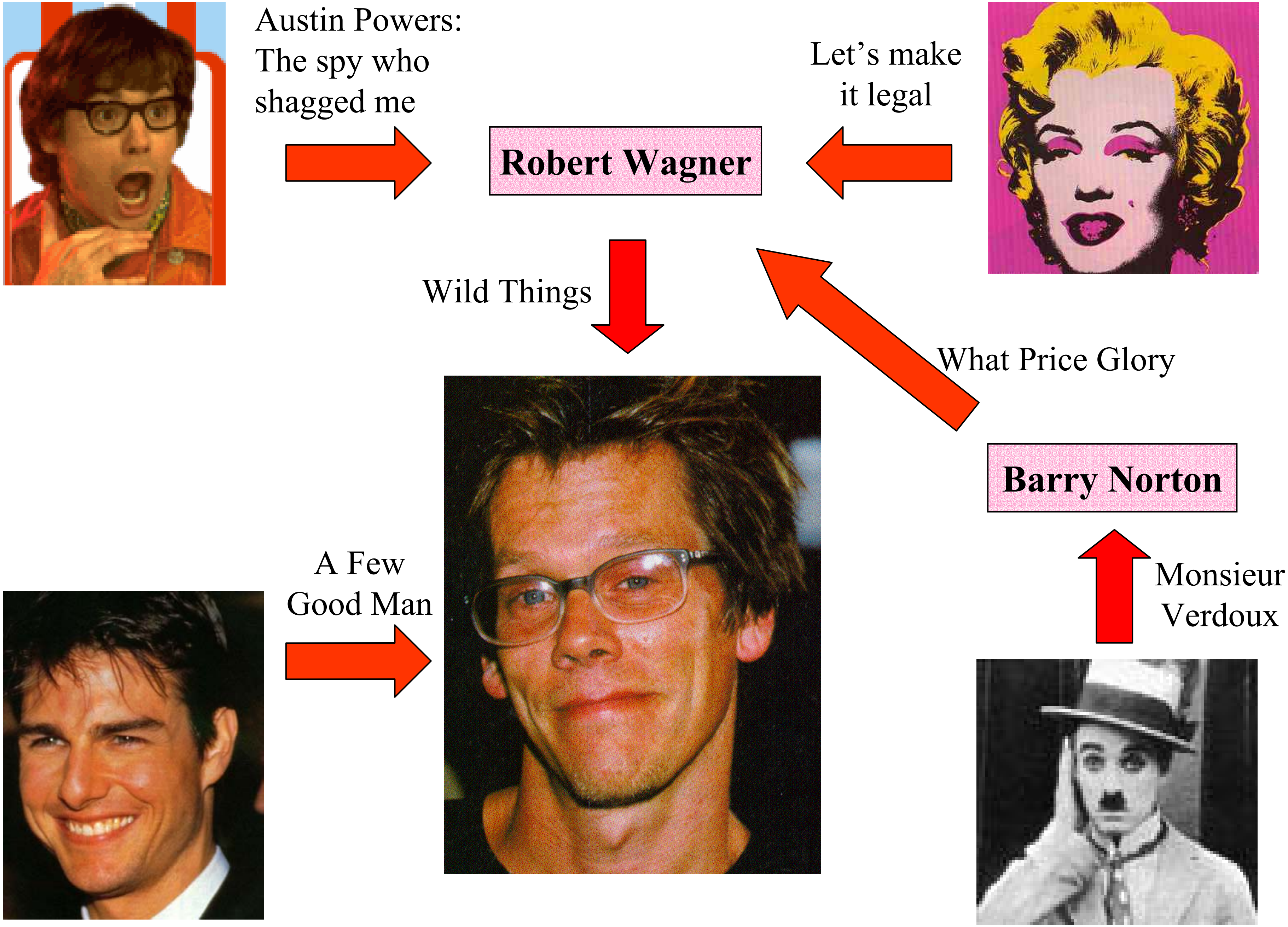
- Length of a path is number of links it has
- Distance between two nodes is length of shortest path (aka geodesic)



Geodesic Distance Matrix

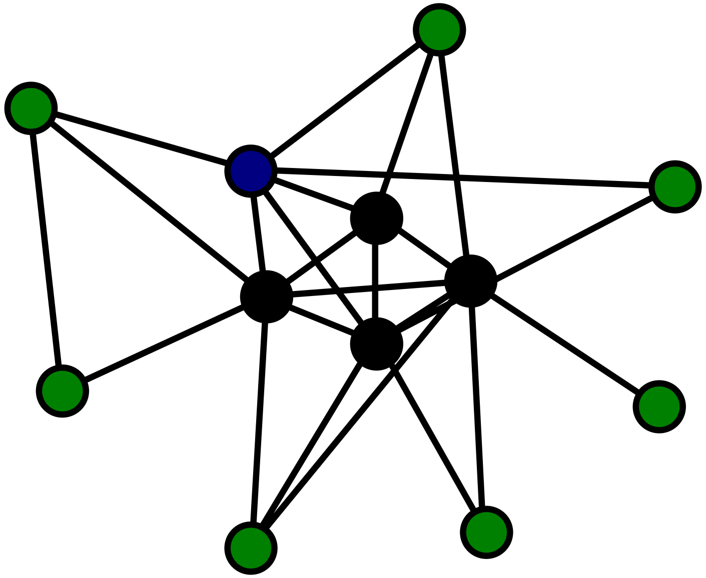
	a	b	c	d	e	f	g
a	0	1	2	3	2	3	4
b	1	0	1	2	1	2	3
c	2	1	0	1	1	2	3
d	3	2	1	0	2	3	4
e	2	1	1	2	0	1	2
f	3	2	2	3	1	0	1
g	4	3	3	4	2	1	0



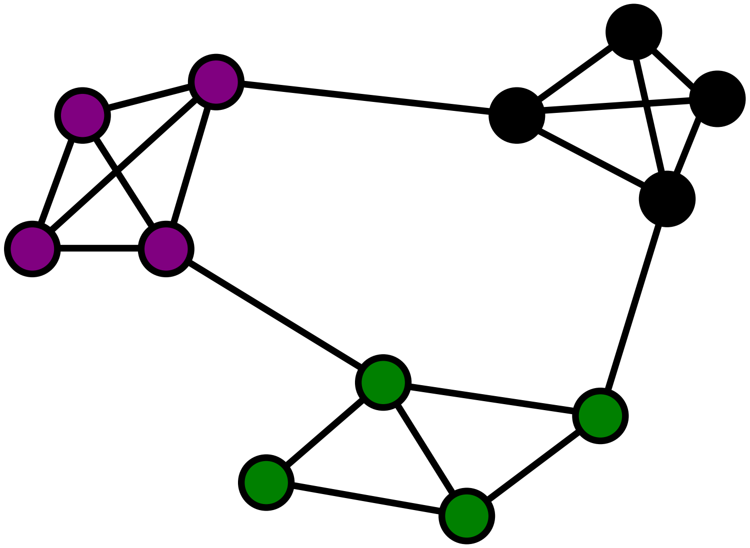

Austin Powers:
The spy who
shagged me
Robert Wagner
Let's make
it legal
Wild Things
What Price Glory
Barry Norton
A Few
Good Man
Monsieur
Verdoux

Diameter

- Maximum distance



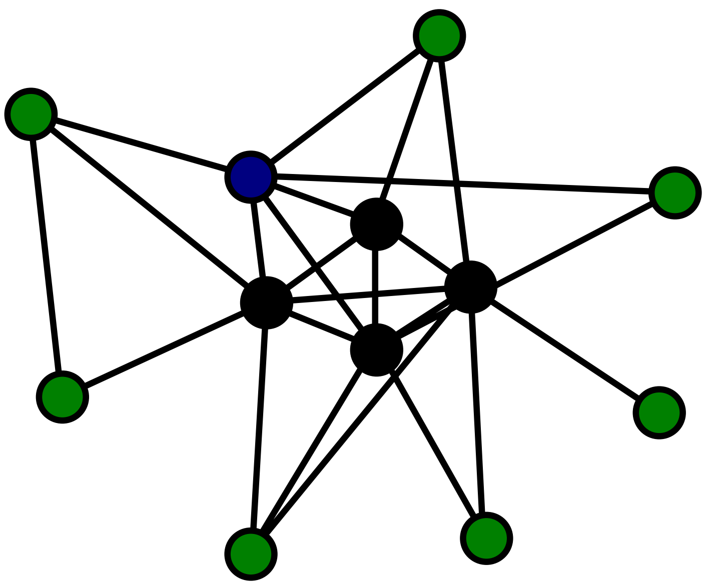
Diameter = 3



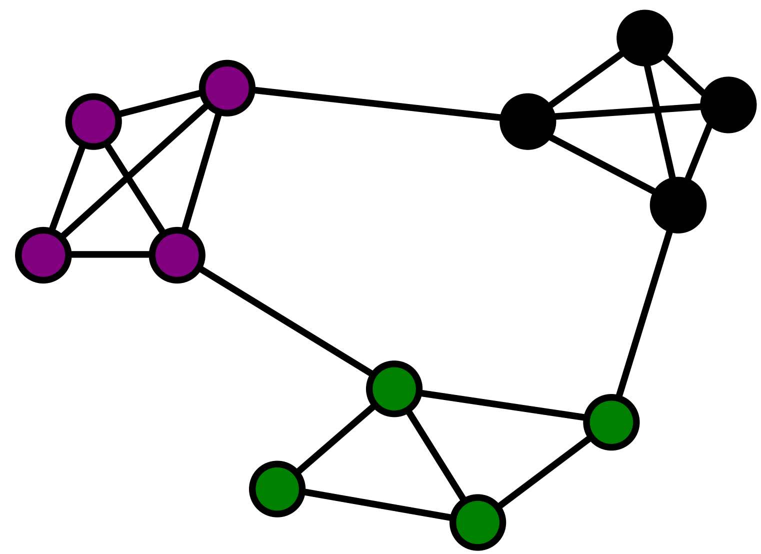
Diameter = 3

Average Distance

- Average geodesic distance between all pairs of nodes



Core/Periphery
c/p fit = 0.97, avg. dist. = 1.9



Clique structure
c/p fit = 0.33, avg. dist. = 2.4

Types of Flow Processes

- Gift process
- Currency process
- Transport process
- Postal process

- Gossip process
- E-mail process
- Infection process
- Influence process

(several others)

Monetary Exchange Process

- Canonical example:
  - specific dollar bill moving through the economy
- Single object in only one place at a time
- Can travel between same pair more than once
    - A--B--C--B--C--D--E--B--C--B--C ...

Gossip Process

- Example:
  - juicy story moving through informal network
- Multiple copies exist simultaneously
- Person tells only one person at a time*
- Doesn’t travel between same pair twice
- Can reach same person multiple times

* More generally, they tell a very limited number at a time.

Infection Process

- Example:
  - virus which activates effective immunological response
- Multiple copies may exist simultaneously
- Cannot revisit a node
    - A--B--C--E--D--F...

Three Kinds of Flows

Type of Flow	Type of Trajectory
Virus	Path
Gossip	Trail
Dollar bill	Walk

Typology

goods

information

	parallel duplication	serial duplication	transfer
geodesics			delivery
paths	nameserver	infection	moocher
trails	sending e-mail	gossiping	hand-me-down
walks	influence		monetary exchange

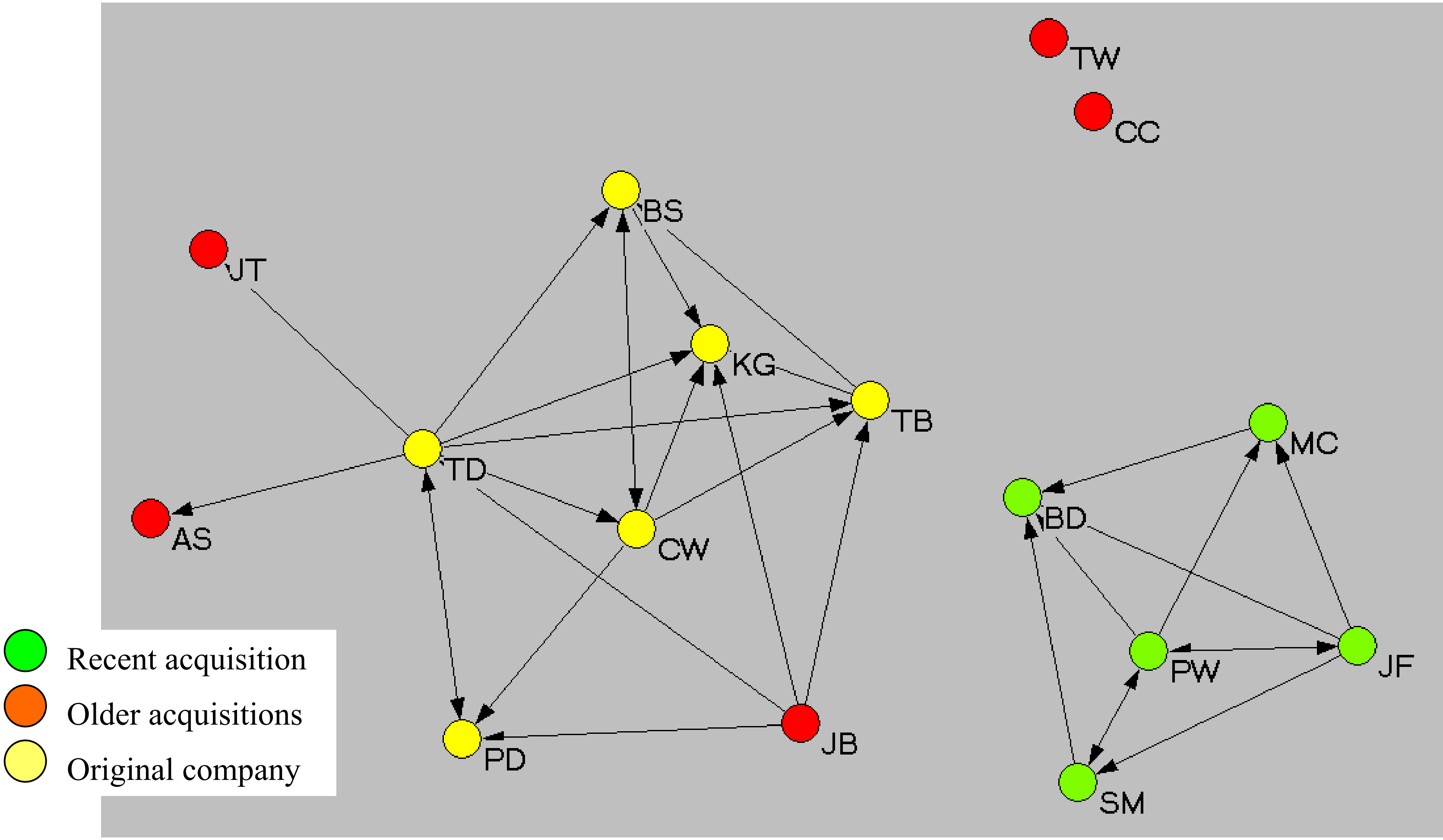
Components

- Maximal sets of nodes in which every node can reach every other by some path (no matter how long)
- A connected graph has just one component

It is relations (types of tie) that define different networks, not components. *A graph that has two components remains one (disconnected) graph.*

A network with 4 components

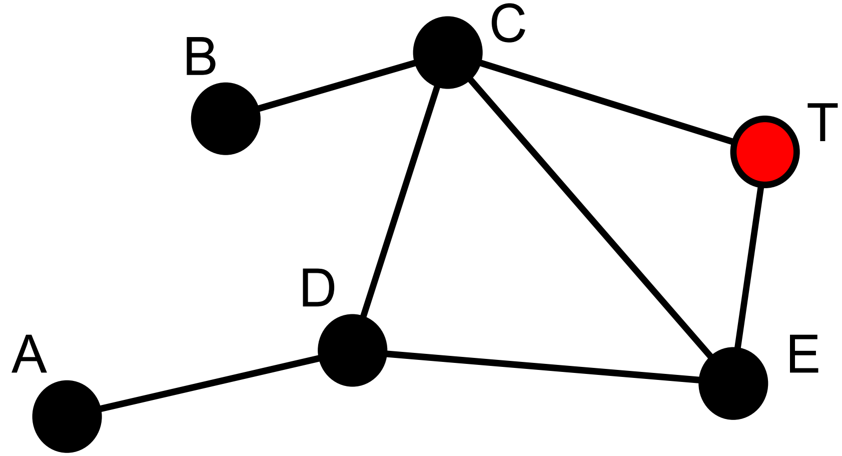
Who you go to so that you can say 'I ran it by ____, and she says ...'



Data drawn from Cross, Borgatti & Parker 2001.

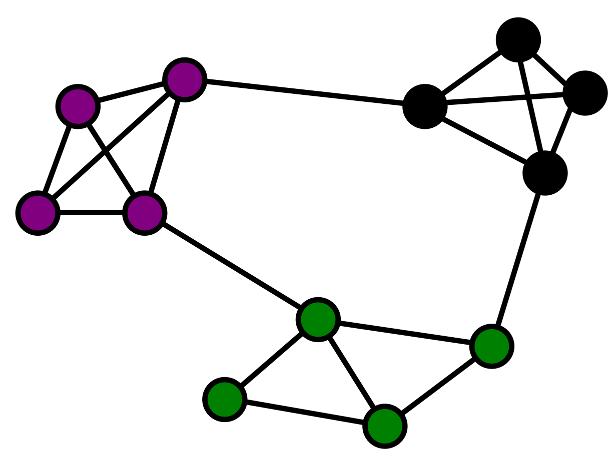
Transitivity

- Number of triples with 3 ties expressed as a proportion of triples with 2 or more ties
  - Aka the clustering coefficient



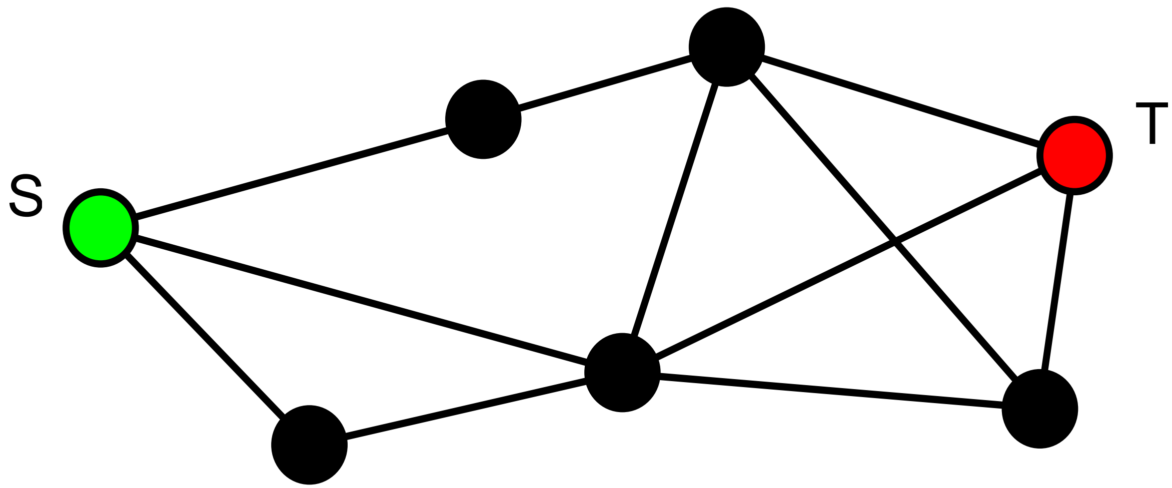
cc = 2/6 = 33%

{C,T,E} is a transitive triple, but {B,C,D} is not



Independent Paths

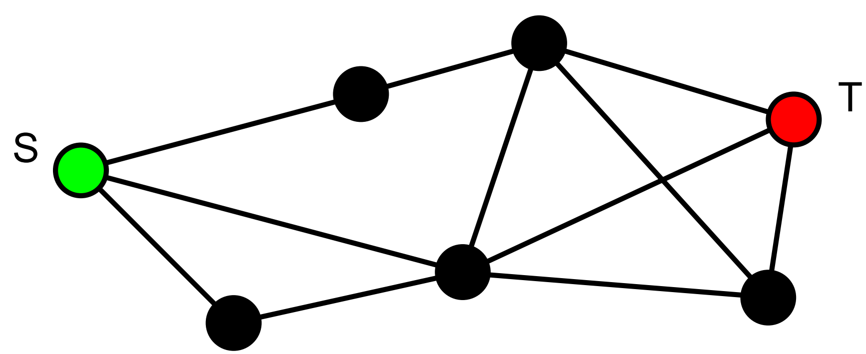
- A set of paths is node-independent if they share no nodes (except beginning and end)
  - They are line-independent if they share no lines



- 2 node-independent paths from S to T
- 3 line-independent paths from S to T

Connectivity

- Line connectivity $\lambda(s,t)$ is the minimum number of lines that must be removed to disconnect s from t
- Node connectivity $\kappa(s,t)$ is minimum number of nodes that must be removed to disconnect s from t

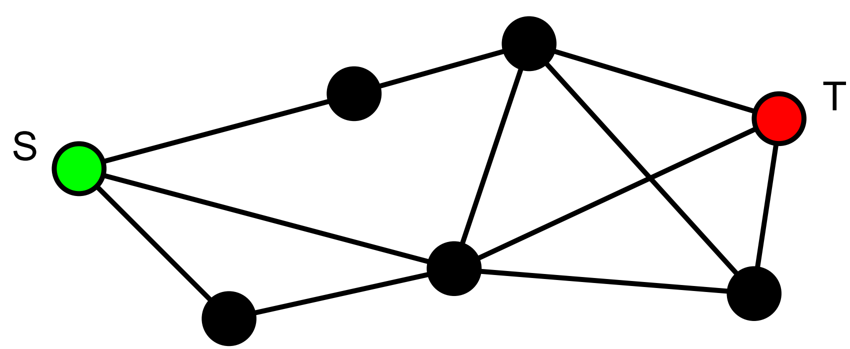


Menger’s Theorem

- Menger proved that the number of line independent paths between s and t equals the line connectivity $\lambda(s,t)$
- And the number of node-independent paths between s and t equals the node connectivity $\kappa(u,v)$

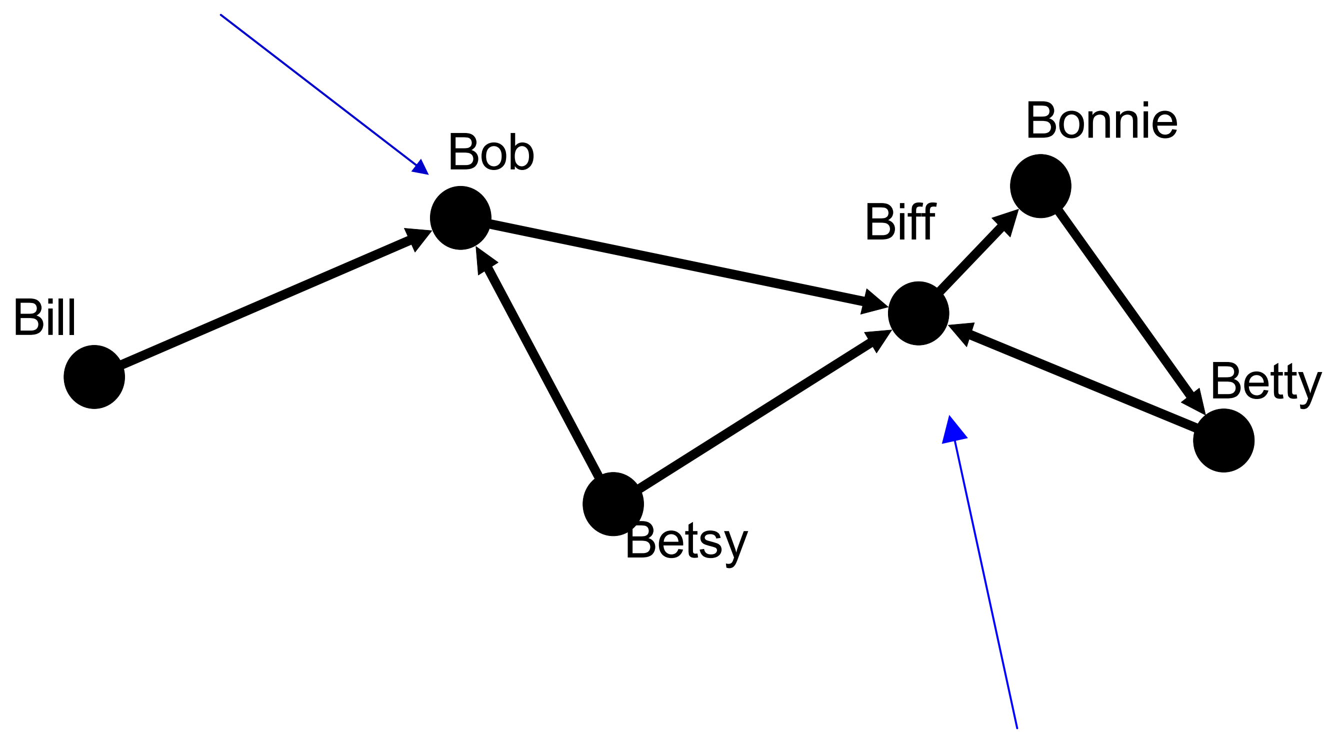
Maximum Flow

- If ties are pipes with capacity of 1 unit of flow, what is the maximum # of units that can flow from s to t?
- Ford & Fulkerson show this was equal to the number of line-independent paths



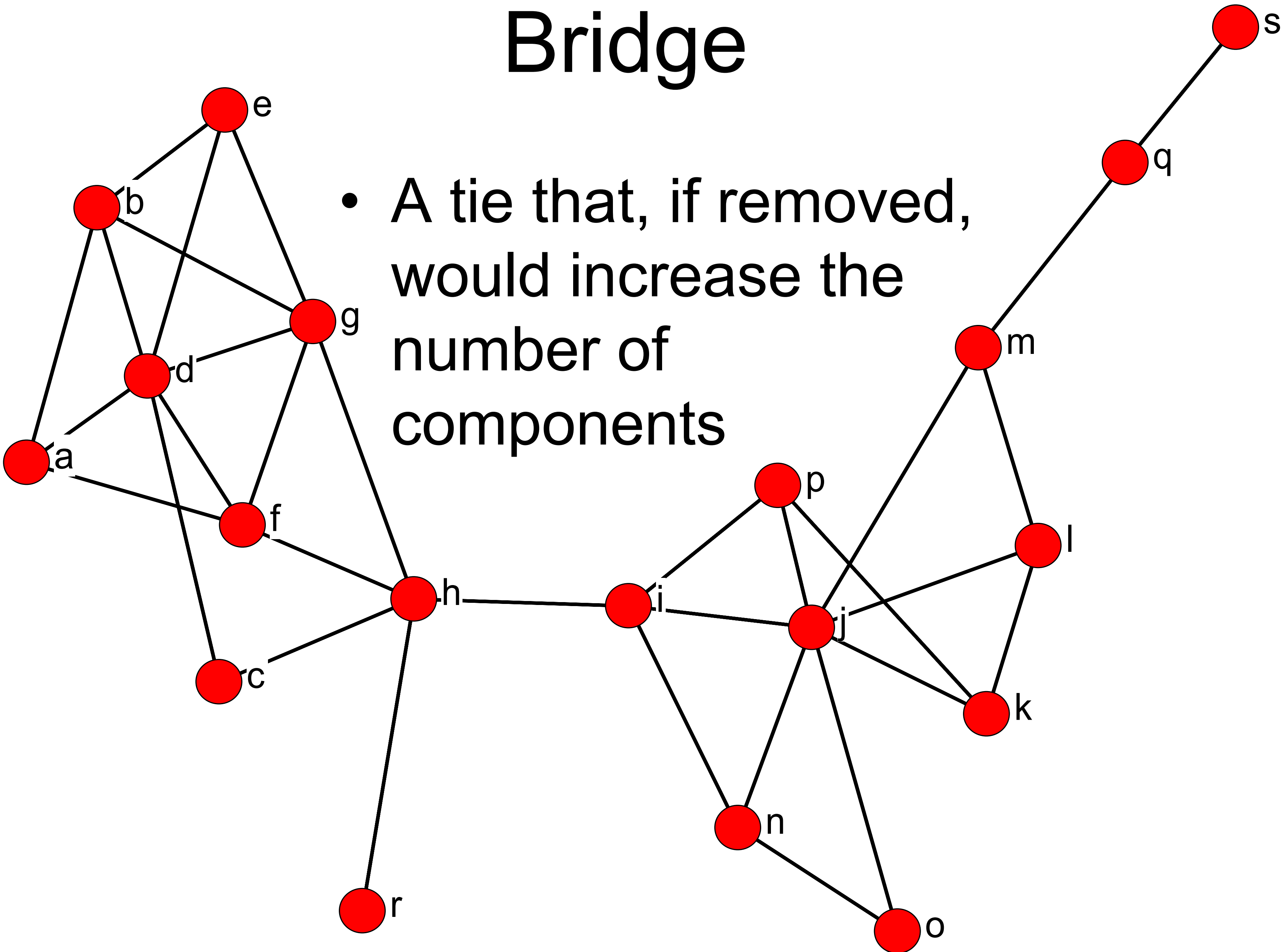
Cutpoint

- A node which, if deleted, would increase the number of components



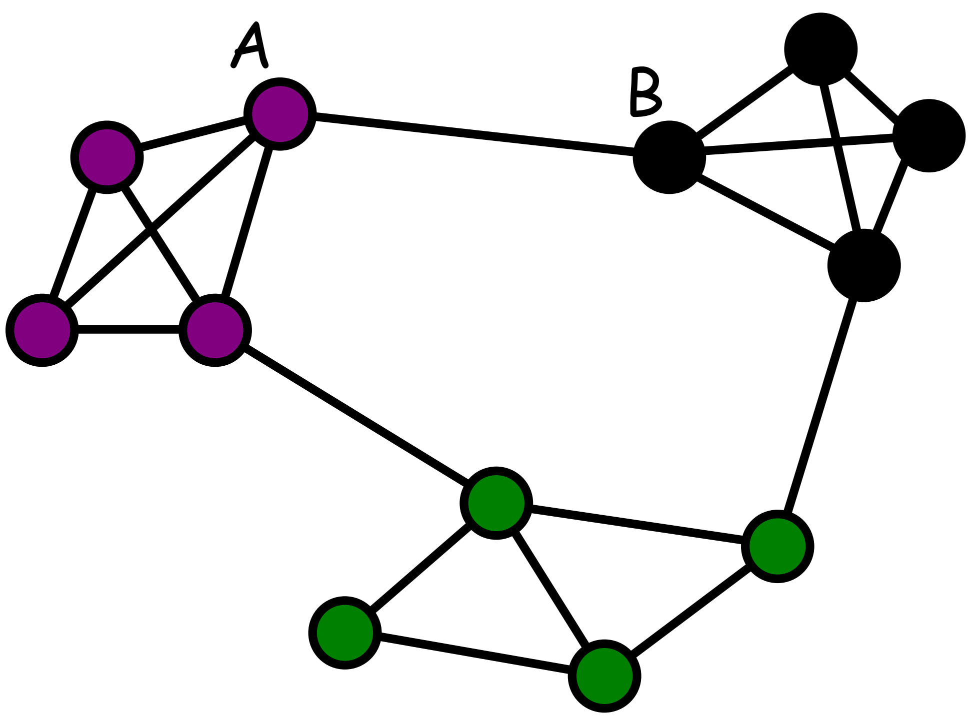
Bridge

- A tie that, if removed, would increase the number of components

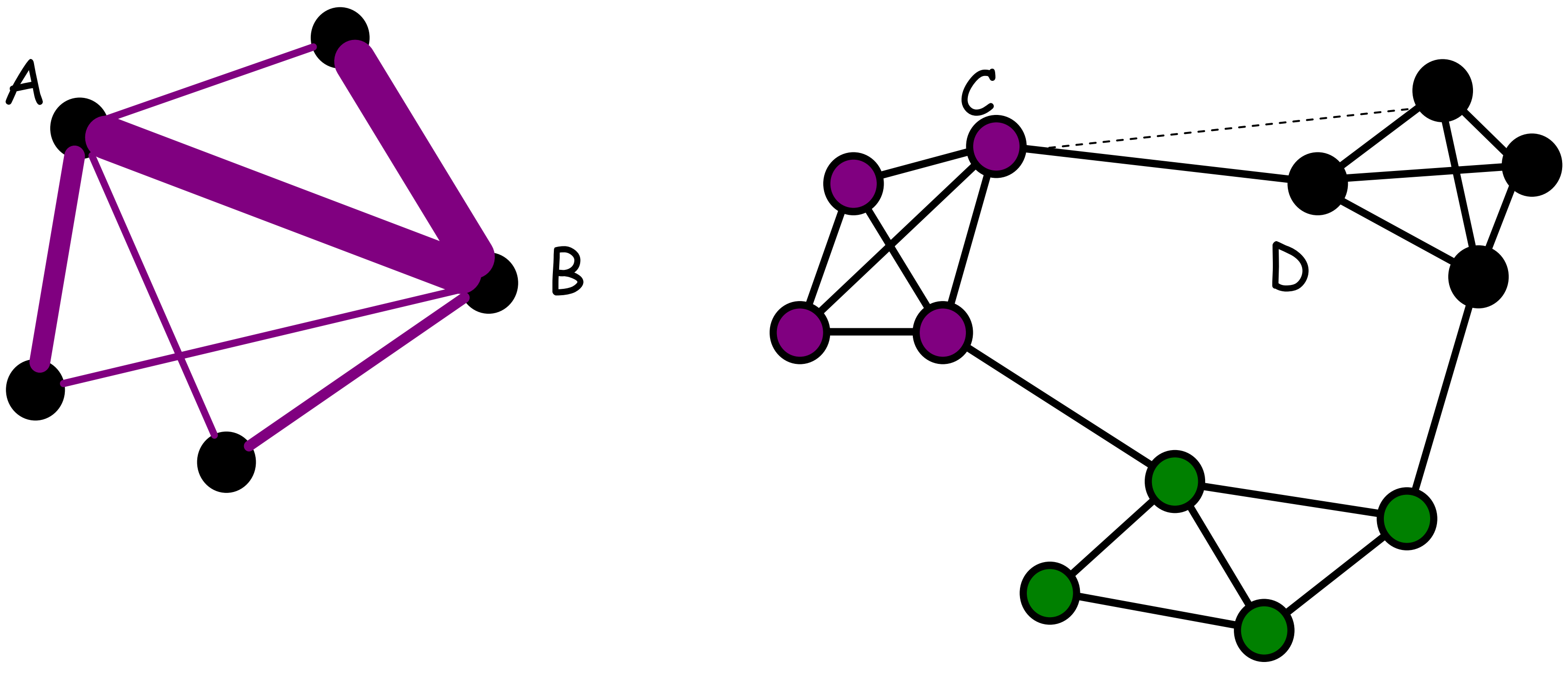


Local Bridge of Degree K

- A tie that connects nodes that would otherwise be at least *k* steps apart



Granovetter Transitivity



Granovetter's SWT Theory

- Strong ties create transitivity
  - Two nodes connected by a strong tie will have mutual acquaintances (ties to same 3rd parties)
- Ties that are part of transitive triples cannot be bridges or local bridges
- Therefore, only weak ties can be bridges
  - Hence the value of weak ties

Granovetter’s SWT

- Strong ties are embedded in tight homophilous clusters,
- Weak ties connect to diversity
- Weak ties a source of novel information