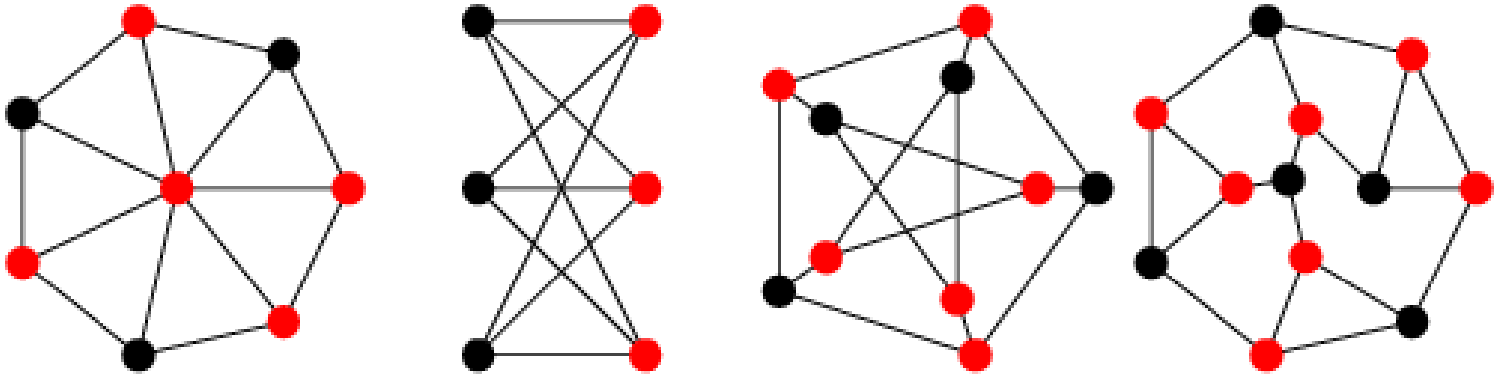


Mathematical Foundations of Social Network Analysis

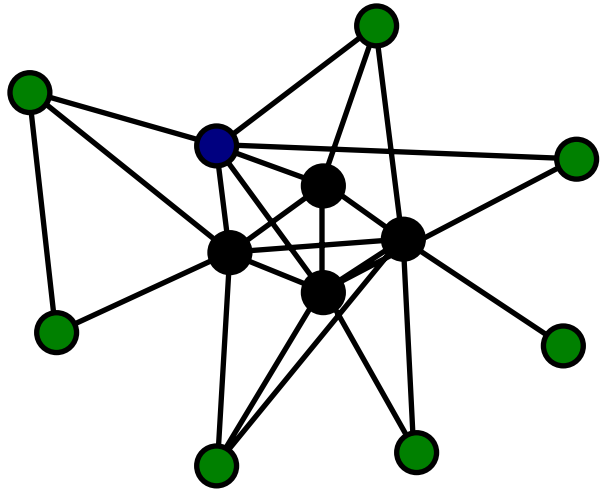
Steve Borgatti

Three Representations

- Graphs
- Matrices
- Relations

Graphs

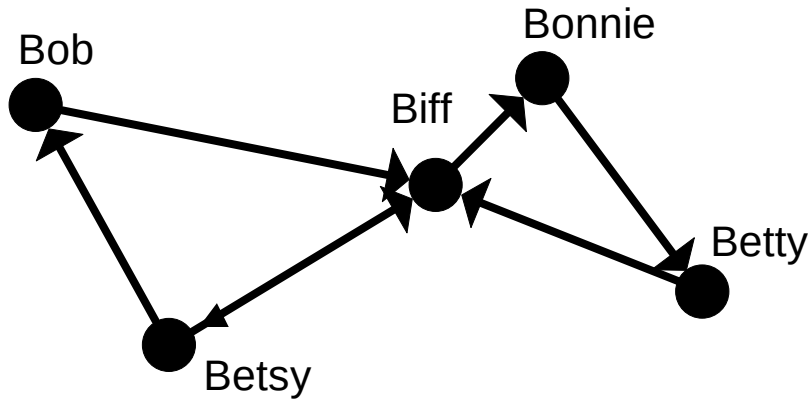
- Networks represented mathematically as graphs
- A graph G(V,E) consists of …
  - Set of nodes|vertices V representing actors
  - Set of lines|edges E representing ties
    - An edge is an unordered pair of nodes (u,v)
    - Nodes u and v adjacent if (u,v) $\in$ E
    - So E is subset of set of all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



- Typically drawn without arrow heads

Digraphs

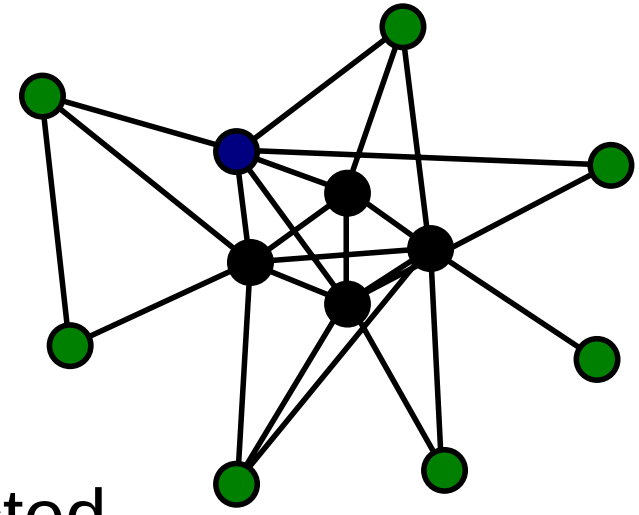
- Digraph D(V,E) consists of …
  - Set of nodes V
  - Set of directed arcs E
    - An arc is an ordered pair of nodes (u,v)
    - (u,v) $\in$ E indicates u sends arc to v
    - (u,v) $\in$ E does not imply that (v,u) $\in$ E

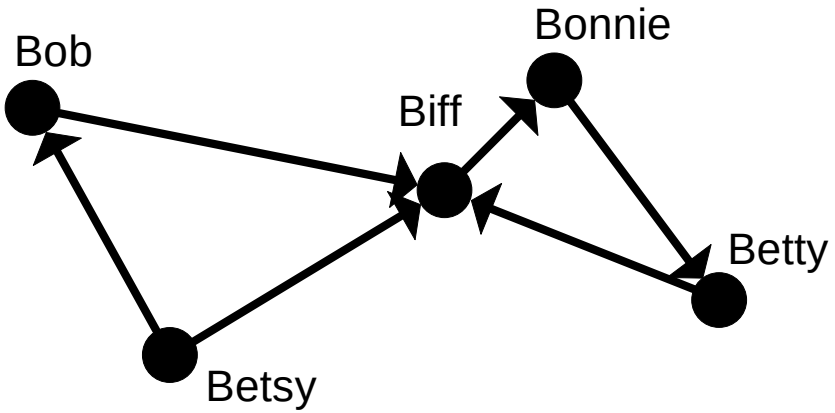


- Ties drawn with arrow heads, which can be in both directions

Directed vs undirected graphs

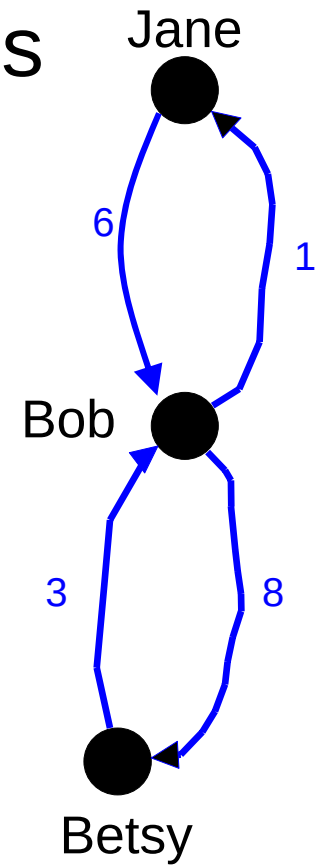
- Undirected relations
  - Attended meeting with
  - Communicates daily with
- Directed relations
  - Lent money to
- Logically vs empirically directed ties
  - Empirically, even un-directed relations can be non-symmetric due to measurement error





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



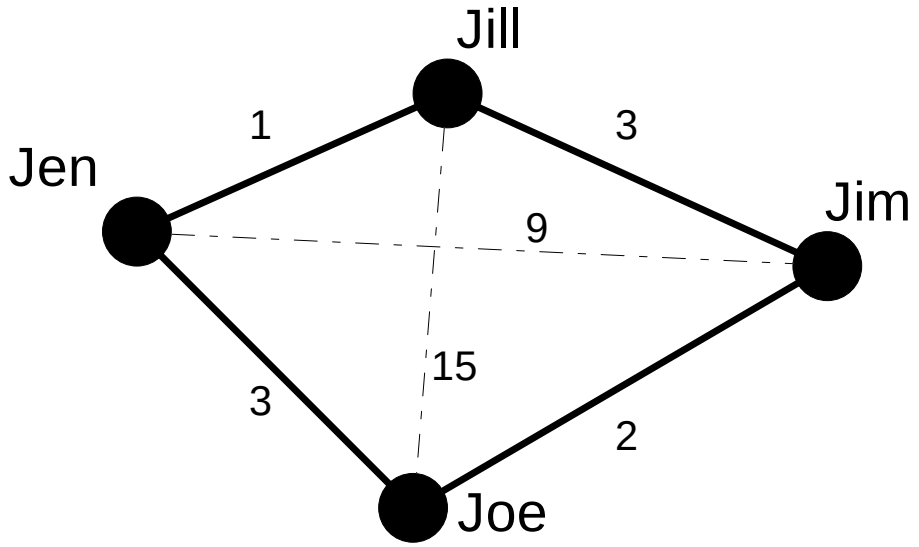
Adjacency Matrices

Friendship

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

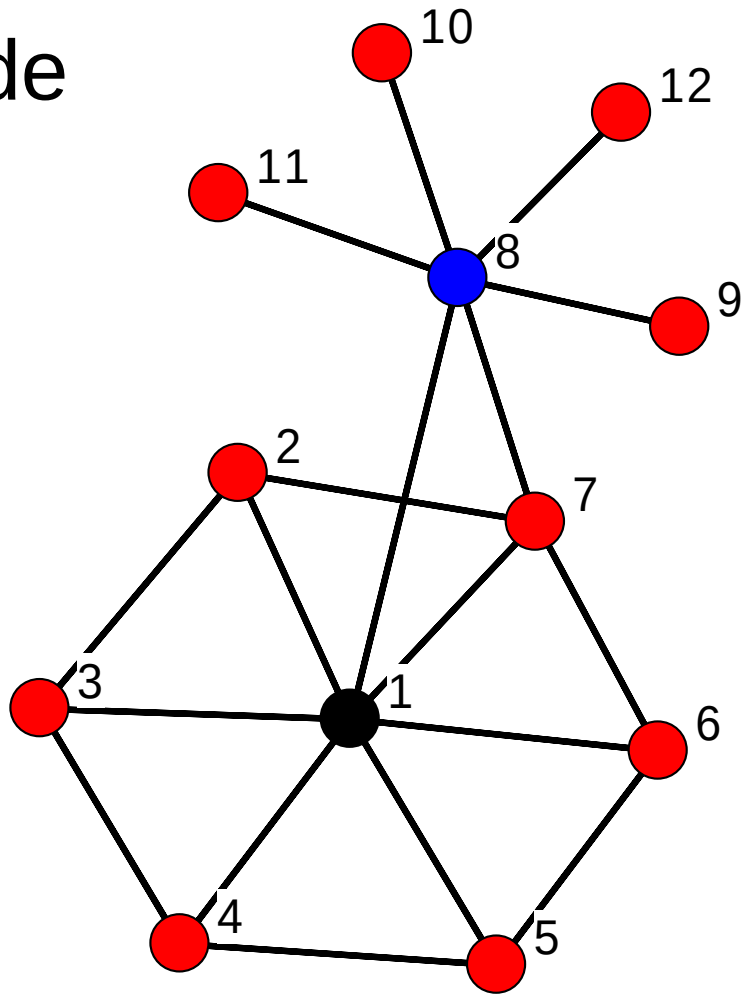
Proximity

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



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



Things that move thru networks

- Used goods
- Money
- Packages
- Personnel
- Orders
- Innovations
- Practices

- Gossip / information
- E-mail
- Infections
- Attitudes
- Influence?

Borgatti, S.P. 2005. Centrality and network flow. *Social Networks*. 27(1): 55-71.

Used Goods Process

- Canonical example:
  - passing along used paperback novel
- Single object in only one place at a time
- Doesn’t (usually) travel between same pair twice
- Could be received by the same person twice
    - A--B--C--B--D--E--B--F--C ...
    - Travels along graph-theoretic trails

Package Delivery Process

- Example:
  - package delivered by postal service
- Single object at only one place at one time
- Map of network enables the intelligent object to select only the shortest paths to all destinations
  - Travels along shortest paths (geodesics)

Mooch Process

- Examples
  - Obnoxious homeless relative who visits for six months until kick out and moves to next relative
  - Personnel flows between firms
- In just one place at a time
- Doesn’t repeat a node (bridges burned)
  - Travels along paths

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.

E-Mail Process

- Example:
  - forwarded jokes and virus warnings
  - e-mail viruses themselves
- Multiple copies exist simultaneously
- All (or many) connected nodes told simultaneously
  - Except, perhaps, the immediate source

Influence Process

- Example:
  - attitude formation
- Multiple “copies” exist simultaneously
- Multiple simultaneous transmission, even between the same pairs of nodes

Viral Infection Process

- Example:
  - virus which activates effective immunological response or which kills host
- Multiple copies may exist simultaneously
- Cannot revisit a node
  - A--B--C--E--D--F...
  - Travels along graph-theoretic paths

Package Delivery Process

- Example:
  - package delivered by postal service
- Single object at only one place at one time
- Map of network enables the intelligent object to select only the shortest paths to all destinations

Two Properties of Flow Processes

- Sequence type: path, trail, walk
  - path: can't revisit node nor edge (tie)
  - trail: can revisit node but not edges
  - walk: can revisit edges & nodes
- Deterministic vs non-deterministic
  - blind vs guided
  - always chooses best route; aware of map
- Combine into 4-way "traversal type" property:
  - geodesics, paths, trails, walks

Traversal Dimension of Flow

- Sequence type: path, trail, walk
  - path: can’t revisit node nor edge (tie)
  - trail: can revisit node but not edges
  - walk: can revisit edges & nodes
- Deterministic vs non-deterministic
  - blind vs guided
  - always chooses best route; aware of map
- Combine into 4-way “traversal type” property:
  - geodesics, paths, trails, walks

Flow Method Property

- Duplication vs transfer (copy vs move)
  - transfer/move: only one place at one time
  - duplication/copy: multiple copies exist
- Serial vs parallel duplication
  - serial: only one transmission at a time
  - parallel: broadcast to all surrounding nodes
- Combine into “method” 3-way property:
  - parallel dup., serial dup., transfer

Simplified Typology

goods

information

	parallel duplication	serial duplication	transfer
geodesics	<no process>	mitotic reproduction	package delivery
paths	internet name-server	viral infection	mooch
trails	e-mail broadcast	gossip	used goods
walks	attitude influencing	emotional support	money exchange

*Note: Names not to be taken too seriously.

Markov

Implications for SNA

- Measures like centrality are constructed by counting things like paths or walks
- Which measure is sensible for a given application depends on how things flow
  - Different measures for different processes
- Most common flow processes do not have measures designed for them
  - Currently misapply std measures to all situations

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

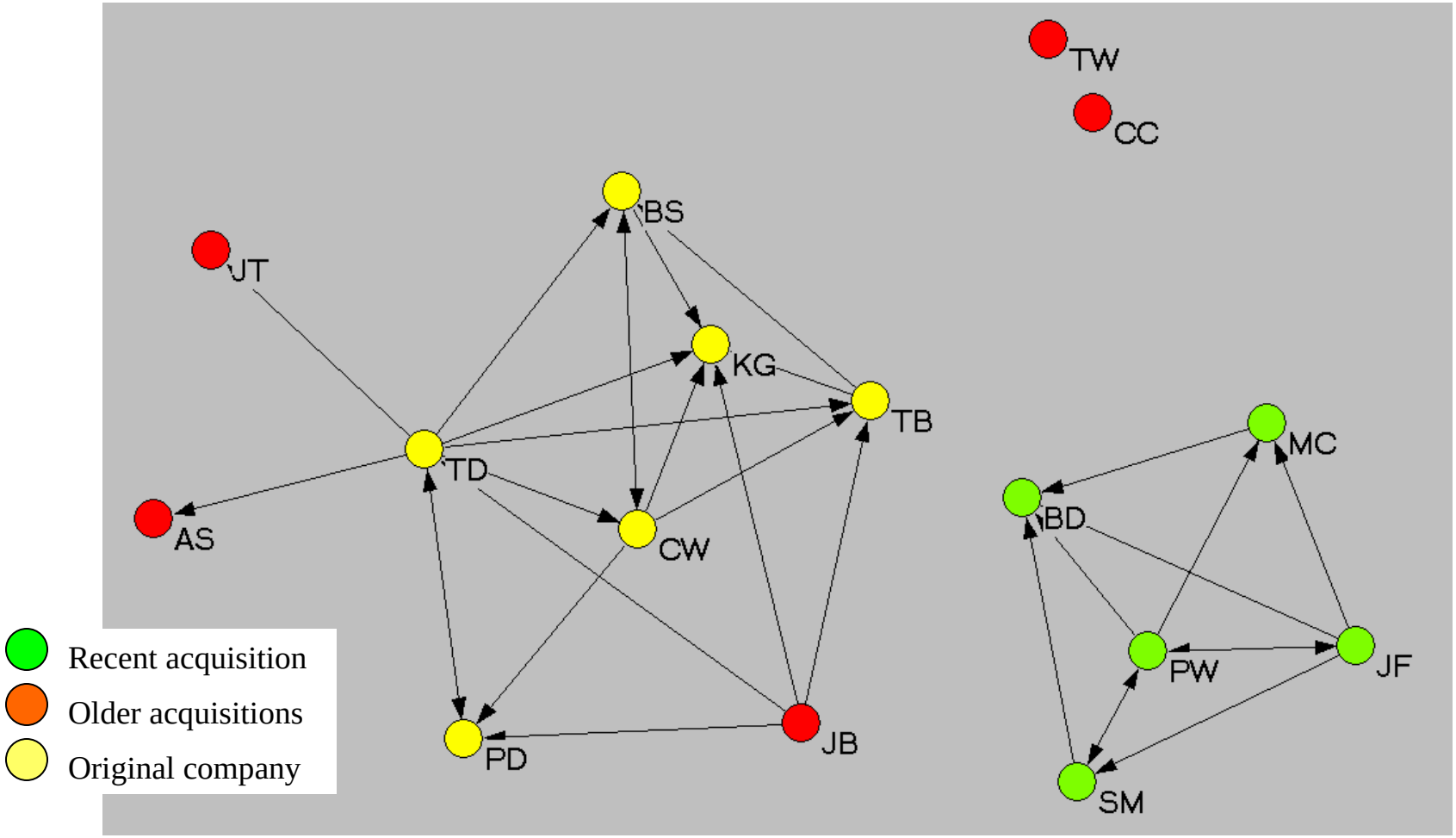
It is relations (types of tie) that define different networks, not components. A graph that has two components remains one (disconnected) graph.

Components in Directed Graphs

- Strong component
  - There is a directed path from each member of the component to every other
- Weak component
  - There is an undirected path (a weak path) from every member of the component to every other
  - Is like symmetrizing via maximum method

A network with 4 components

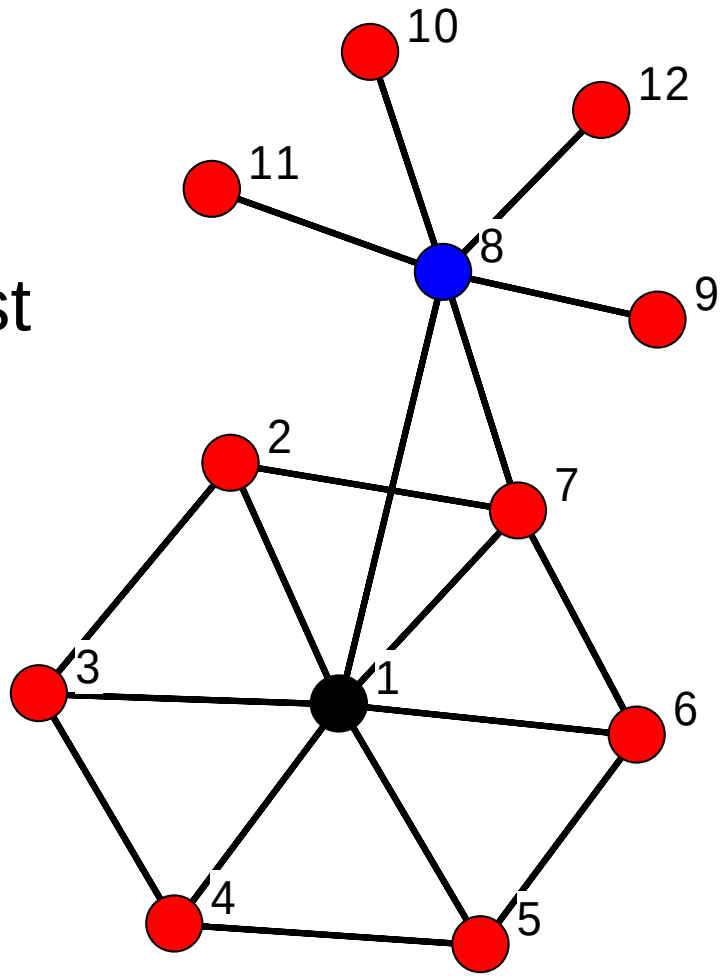
Who you go to so that you can say ‘I ran it by ____, and she says ...’

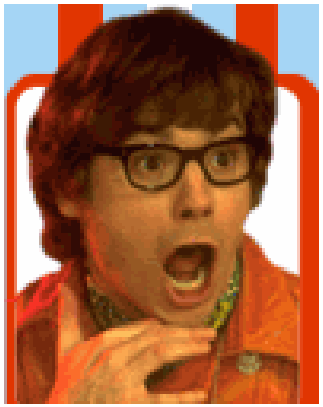


Data drawn from Cross, Borgatti & Parker 2001.

Length & Distance

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





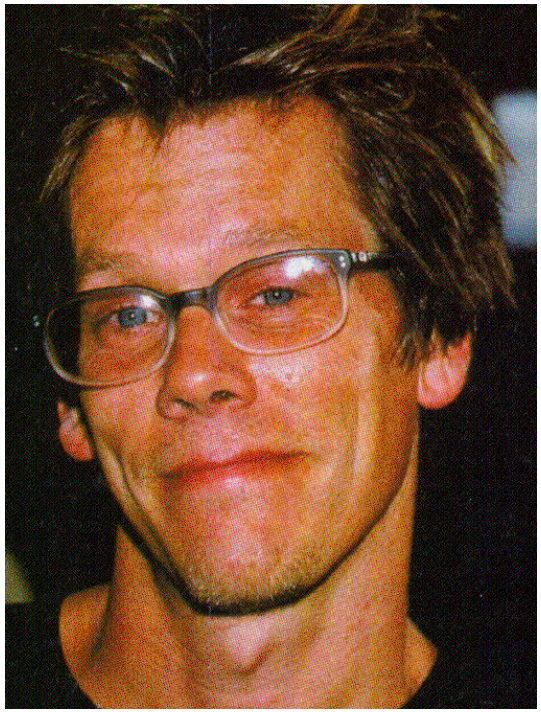
Austin Powers:
The spy who
shagged me

Let's make
it legal

Robert Wagner

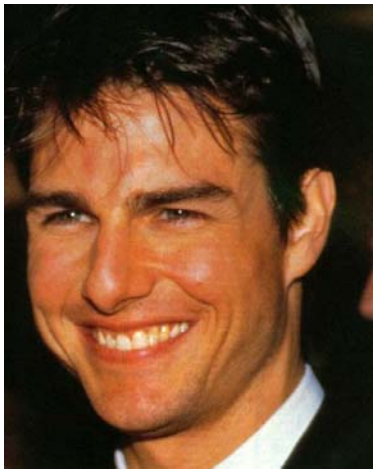
Wild Things

What Price Glory



Barry Norton

A Few
Good Man



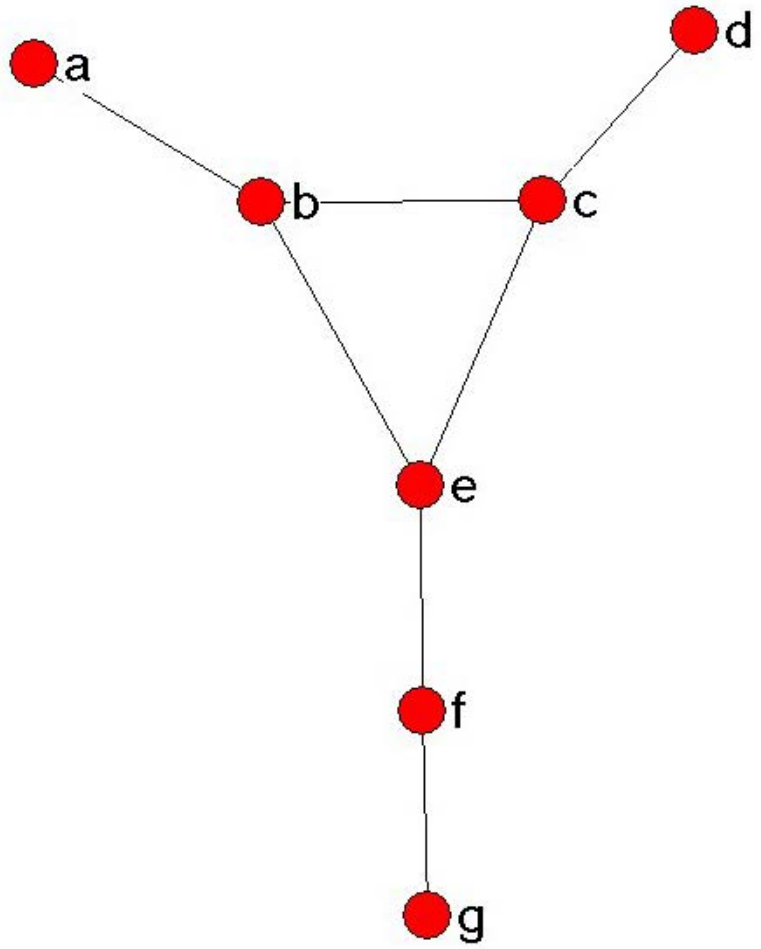
Monsieur
Verdoux

Implications of Distance

- For something flowing across links, distance correlates …
  - Inversely with probability of arrival
  - Inversely with time until arrival
  - Positively with amount of distortion or change

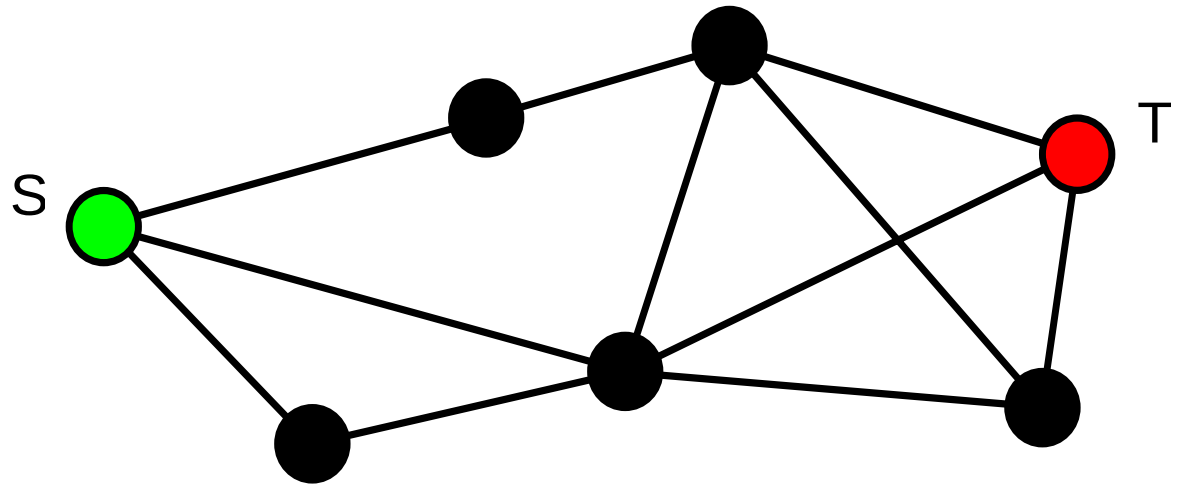
Geodesic Distance Matrix

	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



Independent Paths

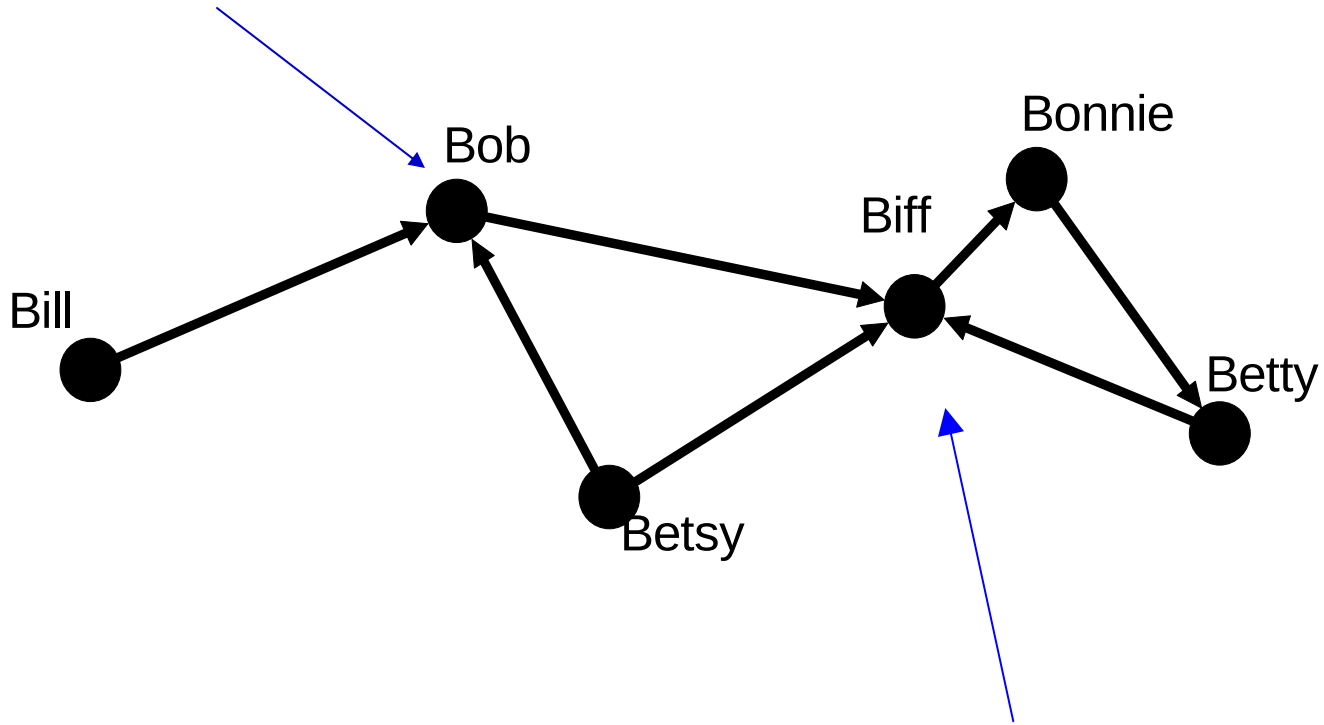
- A set of paths is node-independent if they share no nodes (except beginning and end)
  - They are line-independent if they share no lines



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

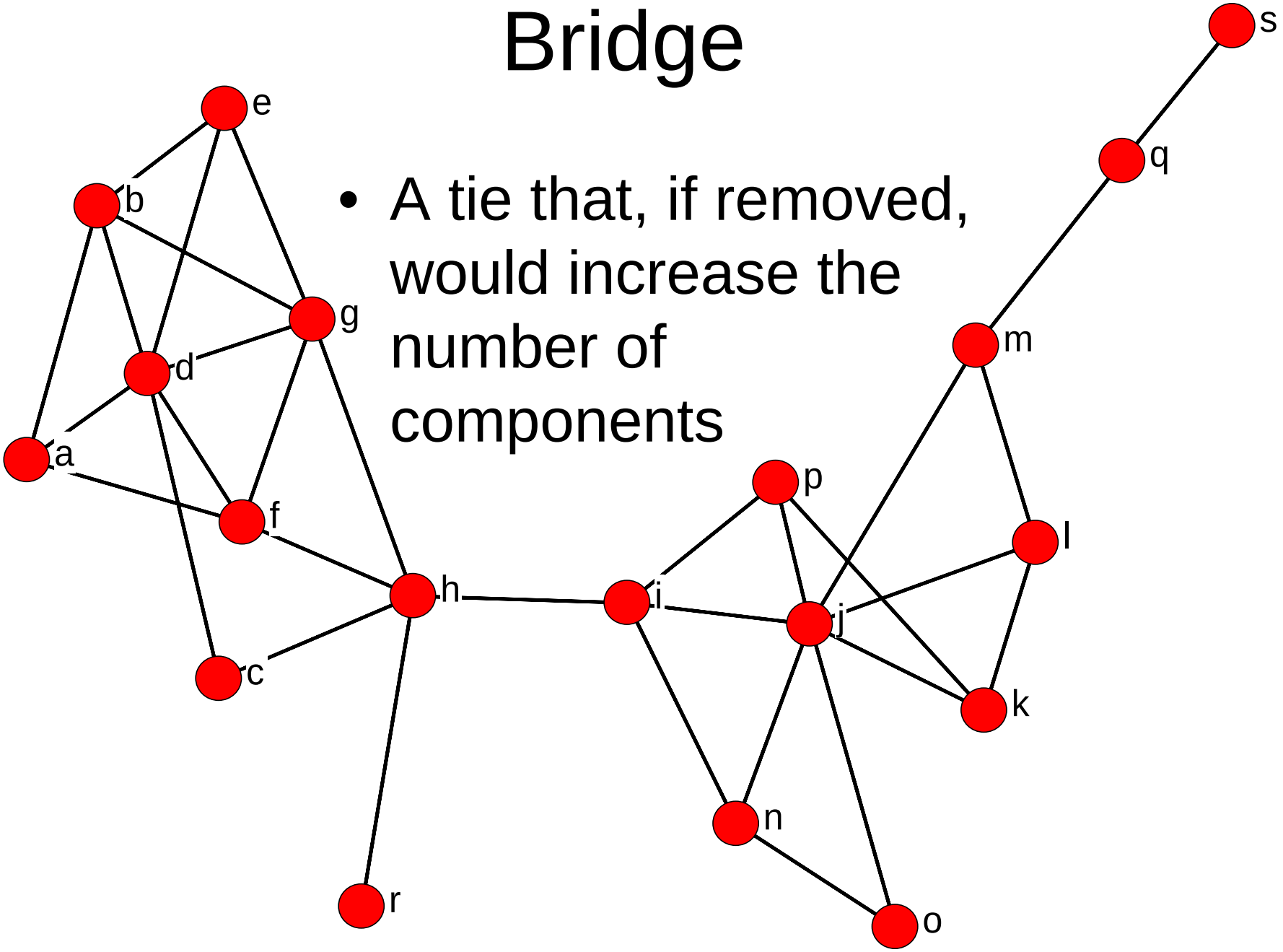
Cutpoint

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



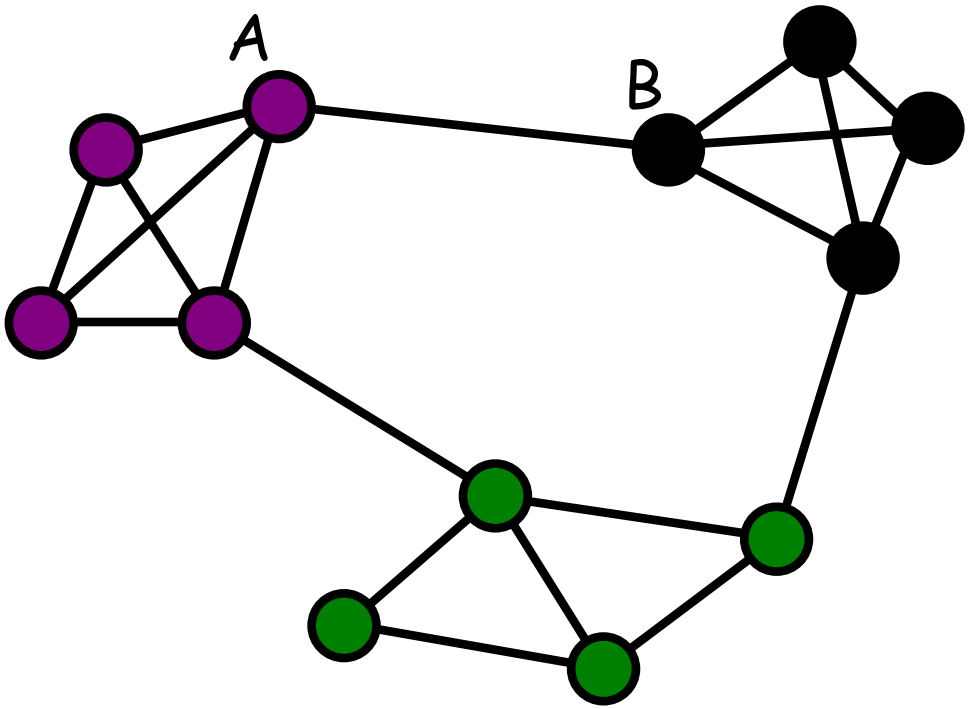
Bridge

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

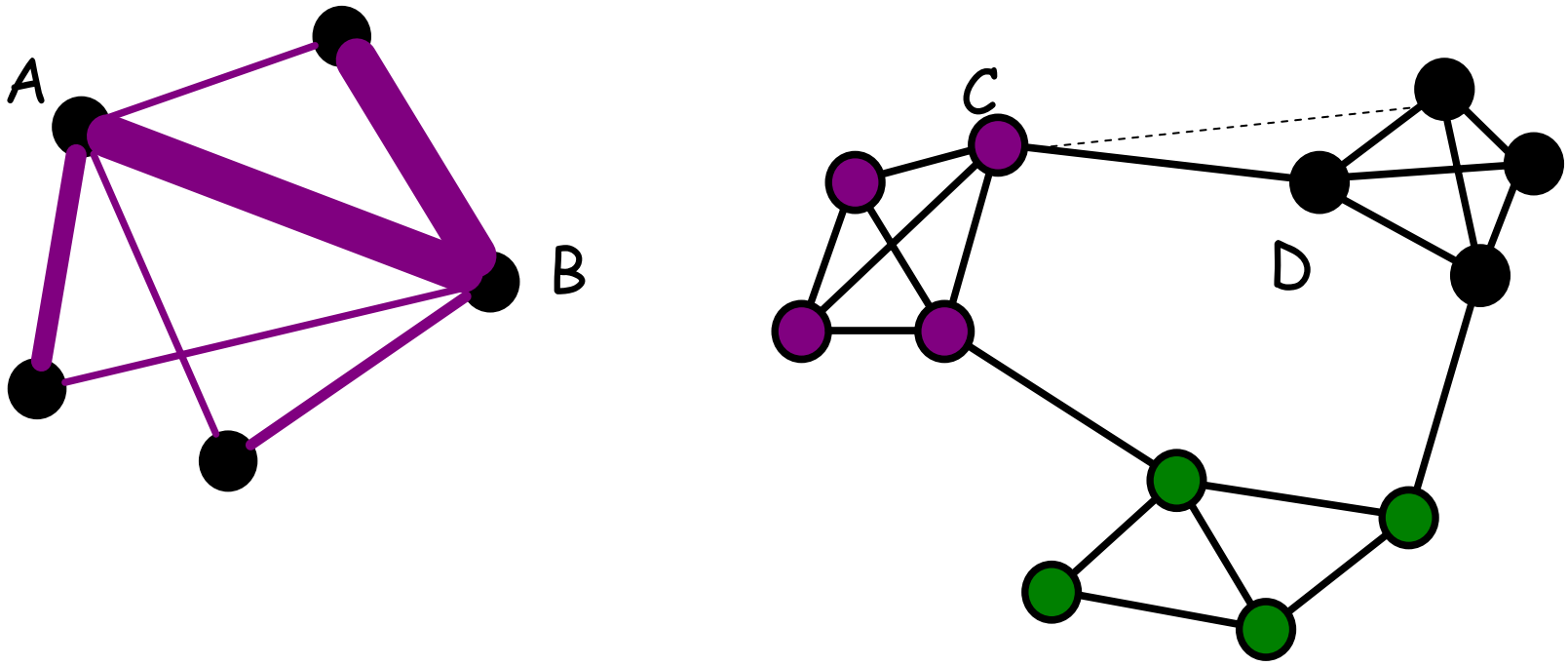


Local Bridge of Degree K

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



Granovetter Transitivity



Granovetter's SWT Theory

- Bridges are sources of novel information
- Only weak ties can be bridges
  - Strong ties create g-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
- Weak ties are sources of novel information

In other words …

- Strong ties are embedded in tight homophilous and homogeneous clusters
- You don’t often hear new stuff from your strong ties because
  - Strong ties embedded in tight homogeneous and homophilous clusters that recirculate same old stories
  - You know all the same people as your strong ties
- Weak ties a source of novel information
- Strong ties are locally cohesive but weak ties are globally cohesive