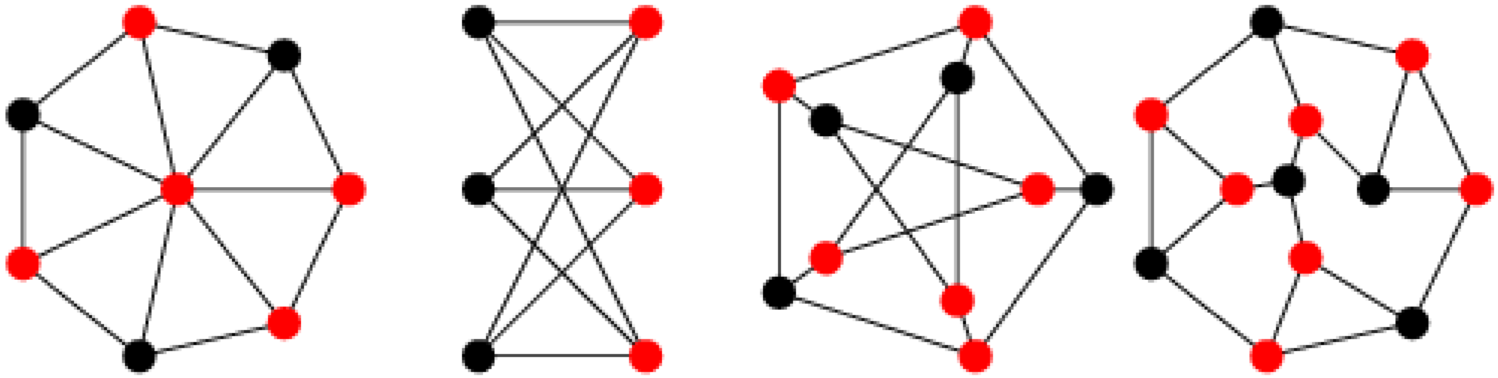


Mathematical Foundations of Social Network Analysis

Steve Borgatti

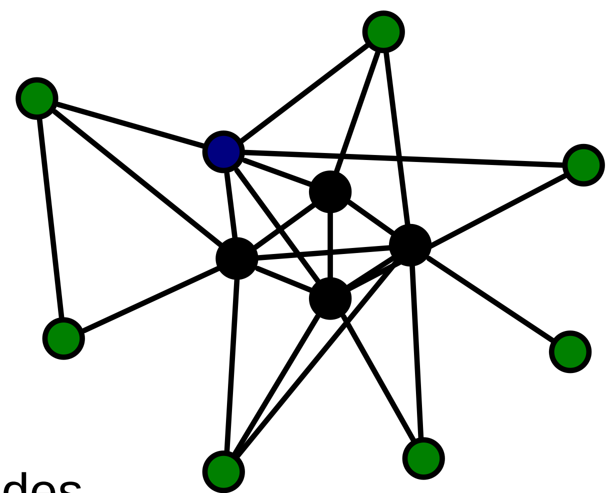
Revised Jan 2008 for MGT 780

Three Representations

- Network/relational data typically represented in one of three ways
  - Graphs
    - Graphs vs digraphs
  - Matrices
  - Relations on sets

(proper) Graphs

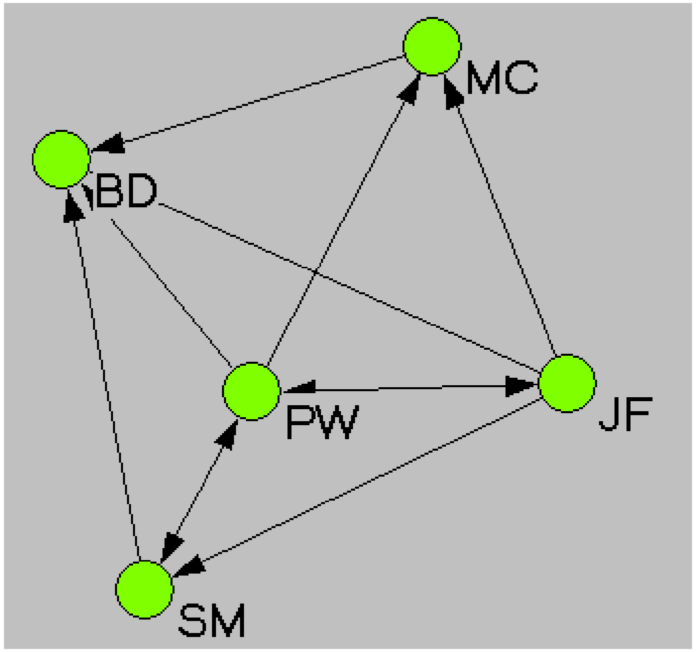
- A graph G(V,E) consists of …
  - Set of nodes|vertices V representing actors
  - Set of lines|edges E representing ties
    - An edge is an unordered pair of nodes (u,v)
    - Nodes u and v adjacent if $(u,v) \in E$
    - So E is subset of set of all pairs of nodes



- Drawn without arrow heads
  - Sometimes with dual arrow heads
- Represent logically symmetric social relations
  - In communication with; attending same meeting as

Digraphs

- Digraph G(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) ∈ E indicates u sends arc to v
    - (u,v) ∈ E does not imply that (v,u) ∈ E



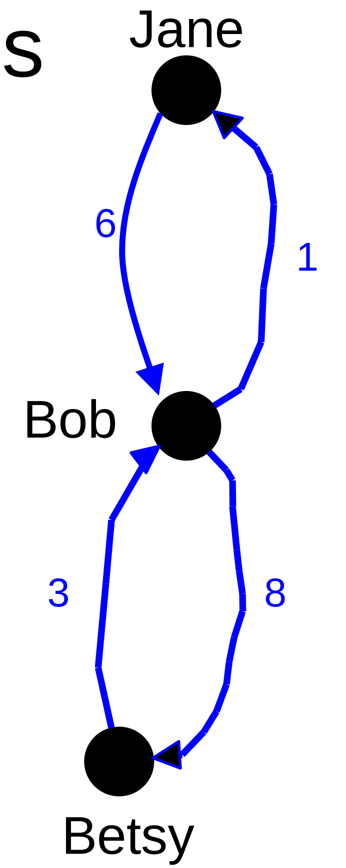
- Ties drawn with arrow heads, which can be in both directions
- Represent logically non-symmetric or anti-symmetric social relations
  - Lends money to

Logical vs Empirical direction

- Empirically, even un-directed relations can be non-symmetric due to measurement error
- Friendship

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 G(V,E,F)
  - F maps E to real numbers



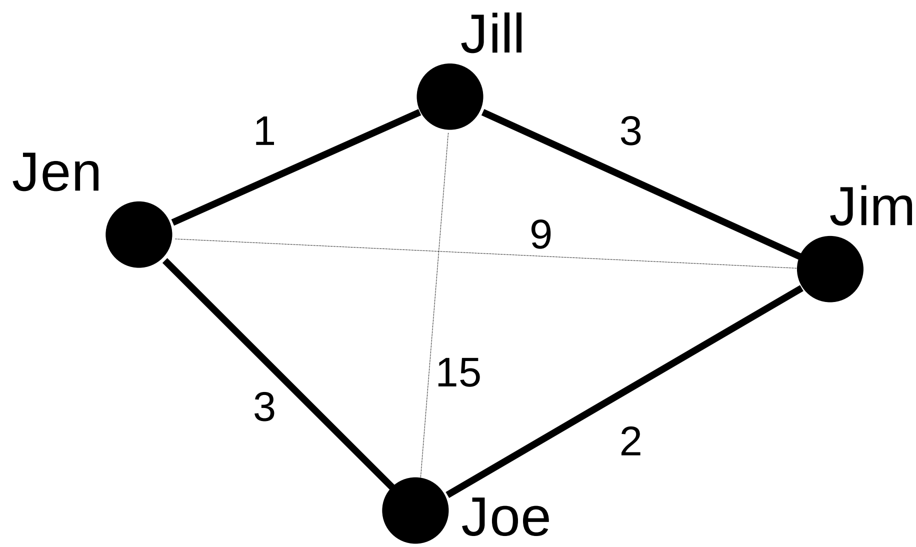
Adjacency Matrices

Friendship

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

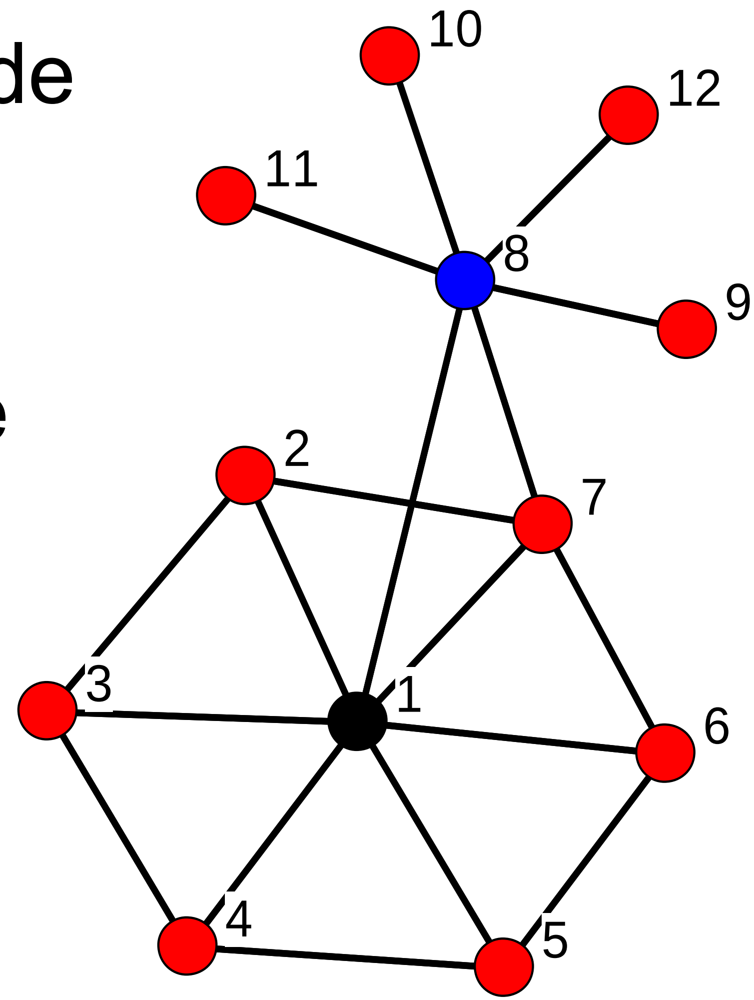
Proximity

	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



Modeling flow processes

- Paths
  - Virus. Hosts become immune or die, so can never return to previous node
  - Con man working a group
- Trails
  - Gossip
  - Used paperback passed along – may be unknowingly be given to someone who already had it
- Walks
  - Dollar bill moving through economy

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

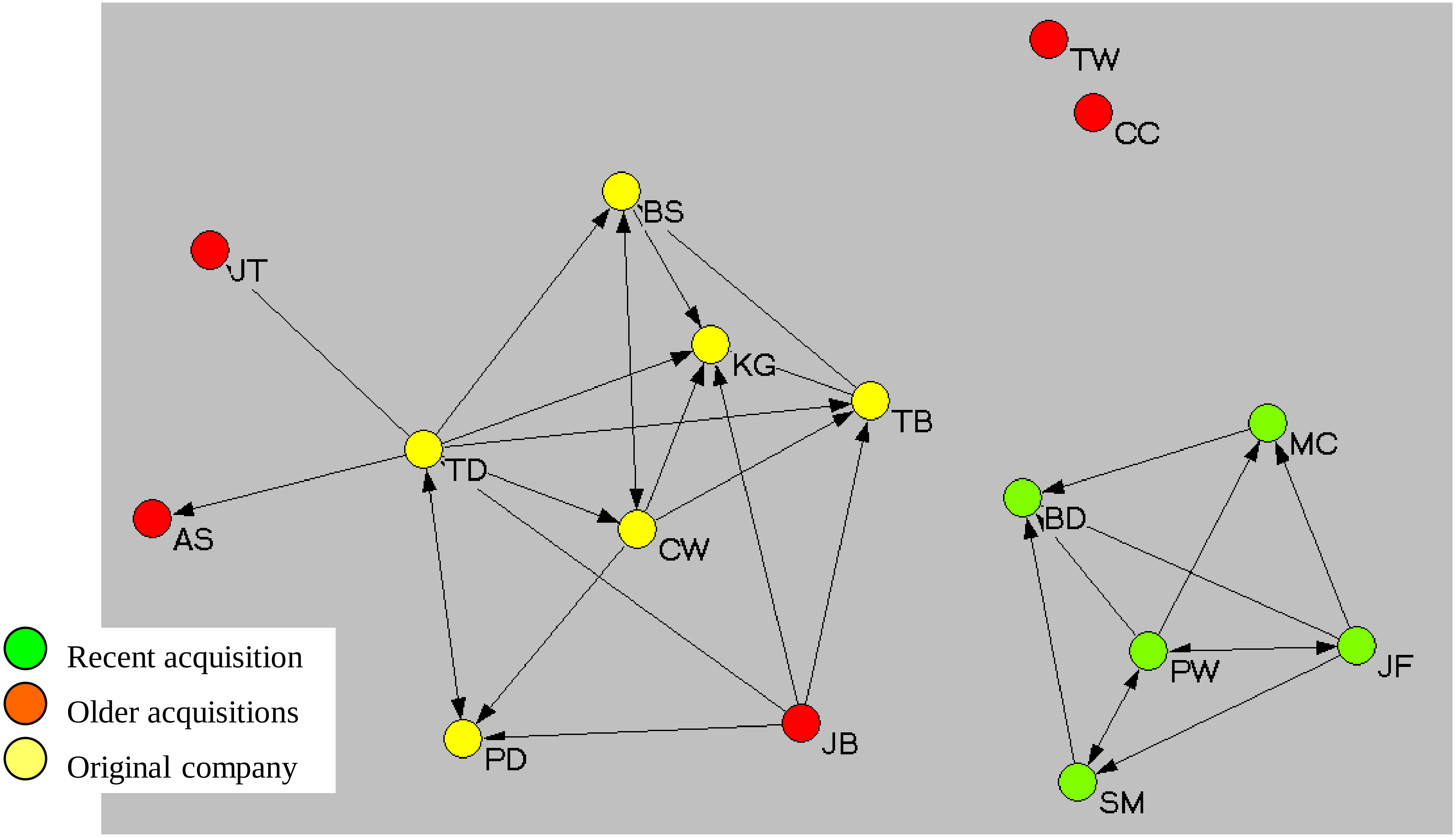
It is relations (types of tie) that define different networks, not components. A network that has two components remains one (disconnected) network.

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 ignoring the direction of ties – driving the wrong way if you have to

A network with 4 components

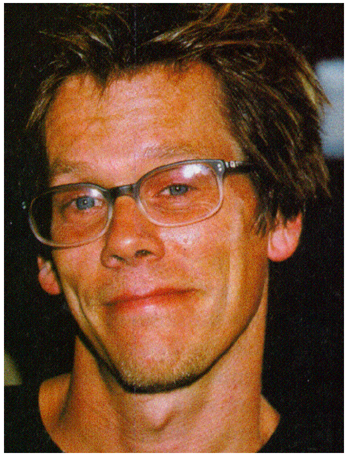
Who you go to so that you can say 'I ran it by ____, and she says ...'



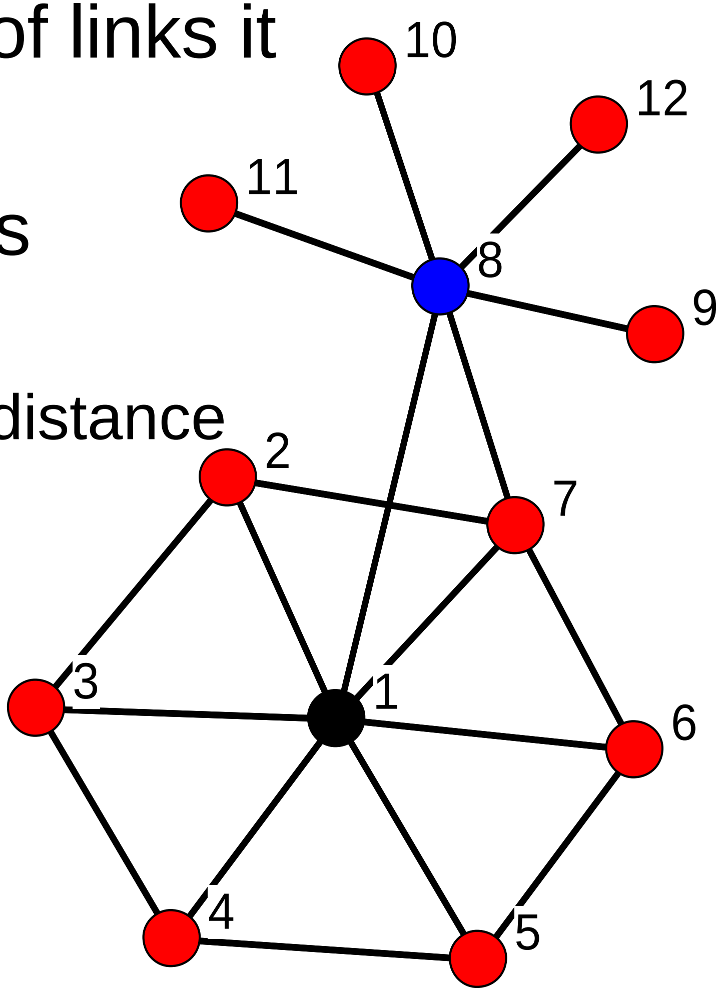
Data drawn from Cross, Borgatti & Parker 2001.

Length & Distance

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

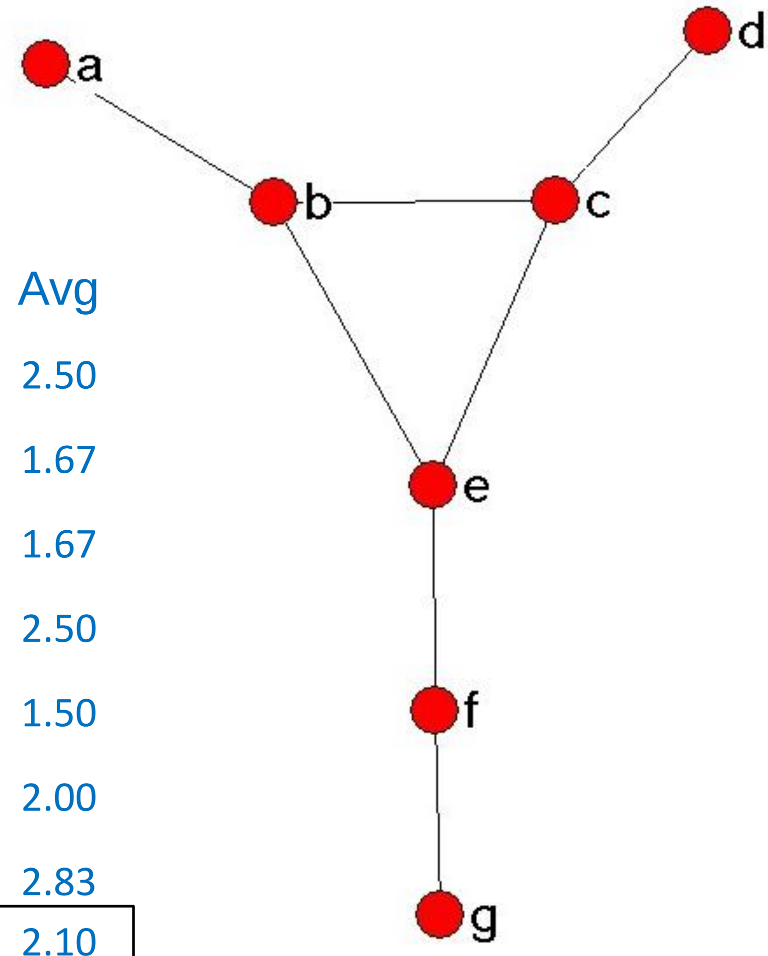


http://oracleofbacon.org/



Geodesic Distance Matrix

	a	b	c	d	e	f	g	Avg
a		1	2	3	2	3	4	2.50
b	1		1	2	1	2	3	1.67
c	2	1		1	1	2	3	1.67
d	3	2	1		2	3	4	2.50
e	2	1	1	2		1	2	1.50
f	3	2	2	3	1		1	2.00
g	4	3	3	4	2	1		2.83
								2.10

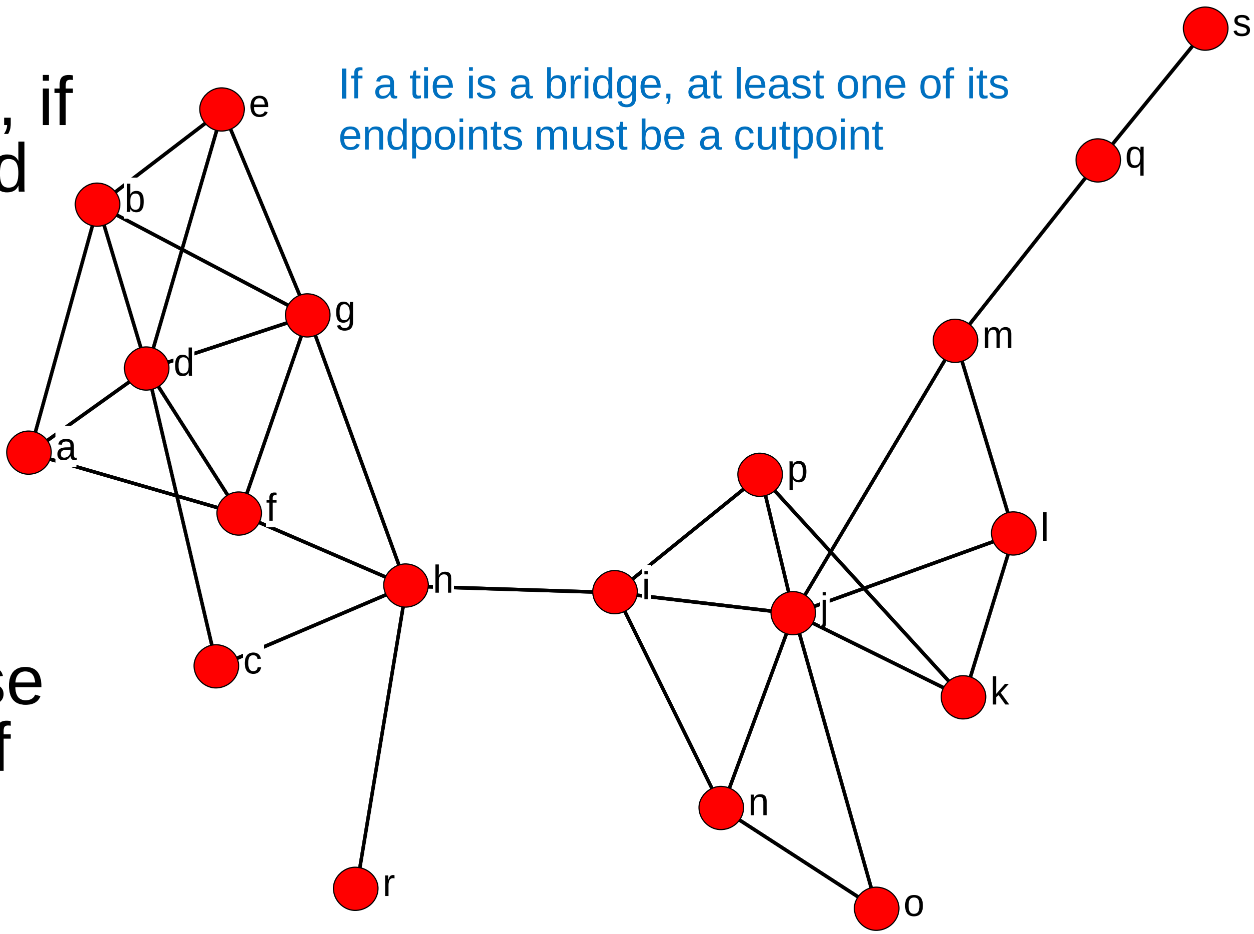


Implications of Distance

- For something flowing across links, expect distance to correlate …
  - Inversely with probability of arrival
  - Inversely with time until arrival
  - Positively with amount of distortion or change
- Averaging to node level we have
  - Typical distance to other nodes in network
- Averaging to network level we have
  - A type of cohesion

Cutpoints and Bridges

- Cutpoint
  - A node which, if deleted, would increase the number of components
- Bridge
  - A tie that, if removed, would increase the number of components

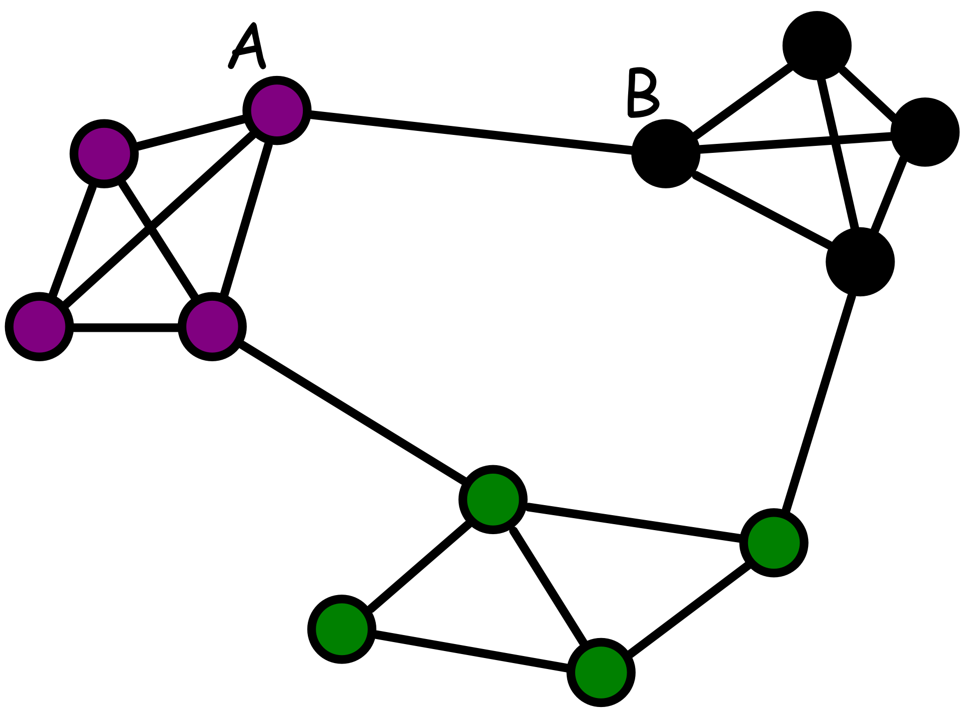


Cutsets

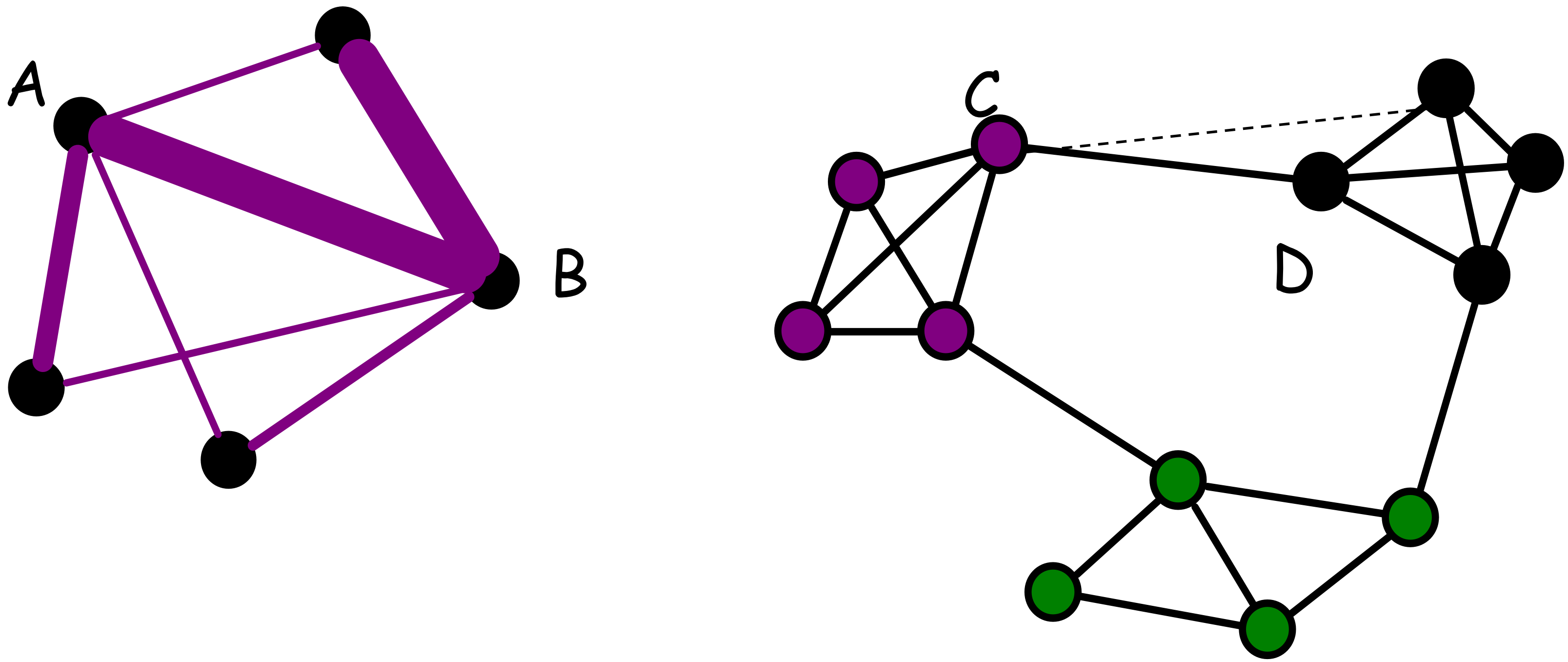
- Vertex cut sets (aka cutsets)
  - A set of vertices S = {u,v,…} of minimal size whose removal would increase the number of components in the graph
- Edge cut sets
  - A set of edges S = {(u,v),(s,t)…} of minimal size whose removal would increase the number of components in the graph

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

MATRICES

Notation

- Matrix A
- Cell a_{ij}
  - First subscript is row, second is column

	Age	Gender	Income
Mary	32	1	90,000
Bill	50	2	45,000
John	12	2	0
Larry	20	2	8,000

A

$a_{12} = 1$
$a_{43} = 8K$

Ways and Modes

- Ways are the dimensions of a matrix.
- Modes are the sets of entities indexed by the ways of a matrix

	Age	Gender	Income
Mary	32	1	90,000
Bill	50	2	45,000
John	12	2	0
Larry	20	2	8,000

2-way, 2-mode

	Mary	Bill	John	Larry
Mary	0	1	0	1
Bill	1	0	0	1
John	0	1	0	0
Larry	1	0	1	0

2-way, 1-mode

Mainstream Logical Data Structure

- 2-mode rectangular matrices
  - Rows (cases) are entities, e.g., persons
  - Columns (variables) are attributes of the cases
- Analysis consists of correlating columns
  - Typically identify one column as the thing to be explained
  - We explain one attribute as a function of the others

Variables (attributes)

Cases (entities)

	Age	Sex	Education	Income
1001				
1002				
1003				
1004				
1005				
…				

Network Logical Data Structures

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	-

Incidence matrix

	Friendship	Proximity
Jim - Jill	1	3
Jim - Jen	0	9
Jim - Joe	1	2
Jill - Jen	1	1
Jill - Joe	0	15
Jen - Joe	1	3

- Multiple relations for same set of actors
- Each relation is a (dyadic) variable
  - But can also be aggregated to node/group level
- Cases are pairs of actors
- Some hypotheses can be phrased in terms of correlations between relations

Transpose

- Interchange rows and columns

	Age	Gender	Income
Mary	32	1	90,000
Bill	50	2	45,000
John	12	2	0
Larry	20	2	8,000

	Mary	Bill	John	Larry
Age	32	50	12	20
Gender	1	2	2	2
Income	90,000	45,000	0	8,000

Transpose Adjacency matrix

- Interchange rows and columns
- If matrix represents network, transpose effectively reverses the direction of ties

	Mary	Bill	John	Larry
Mary	0	1	0	1
Bill	1	0	0	1
John	0	1	0	0
Larry	1	0	1	0

Gives money to

	Mary	Bill	John	Larry
Mary	0	1	0	1
Bill	1	0	1	0
John	0	0	0	1
Larry	1	1	0	0

Gets money from

Marginals

	Mary	Bill	John	Larry	Row Marginals
Mary	0	1	1	1	3
Bill	1	0	1	0	2
John	0	0	0	1	1
Larry	0	0	0	0	0
Column Marginals	1	1	2	2	6

Matrix marginal

Matrix Product

- Notation: C = AB

- Definition: $$c_{ij} = \sum_k a_{ik} b_{jk}$$

- Example:

	Mary	Bill	John	Larry
Mary	0	1	1	1
Bill	1	0	1	0
John	0	0	0	1
Larry	0	0	0	0

A

	Mary	Bill	John	Larry
Mary	0	0	1	1
Bill	1	0	1	0
John	0	0	0	1
Larry	0	1	0	0

B

	Mary	Bill	John	Larry
Mary	1	1	1	1
Bill	0	0	1	2
John	0	1	0	0
Larry	0	0	0	0

C=AB

Multiplying a matrix by itself

- $A^2 = AA$
- Cell a_{ij}^2 gives the number of walks of length *2* from *i* to *j*
- In general a_{ij}^k gives the number of walks of length *k* from *i* to *j*

	Mary	Bill	John	Larry
Mary	0	0	1	1
Bill	1	0	1	0
John	0	0	0	1
Larry	0	1	0	0

A

×

	Mary	Bill	John	Larry
Mary	0	0	1	1
Bill	1	0	1	0
John	0	0	0	1
Larry	0	1	0	0

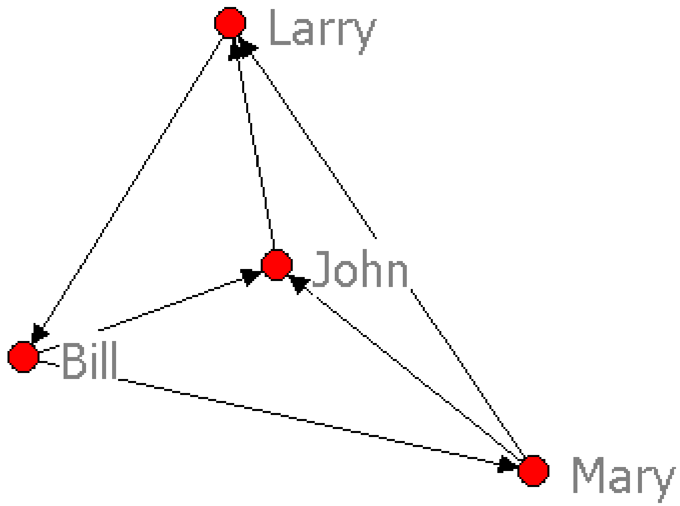
A

=

	Mary	Bill	John	Larry
Mary	0	1	0	1
Bill	0	0	1	2
John	0	1	0	0
Larry	1	0	1	0

A^2

Powers of a Matrix



	Mary	Bill	John	Larry
Mary	0	0	1	1
Bill	1	0	1	0
John	0	0	0	1
Larry	0	1	0	0

A

	Mary	Bill	John	Larry
Mary	0	1	0	1
Bill	0	0	1	2
John	0	1	0	0
Larry	1	0	1	0

A^2

	Mary	Bill	John	Larry
Mary	1	1	1	0
Bill	0	2	0	1
John	1	0	1	0
Larry	0	0	1	2

A^3

	Mary	Bill	John	Larry
Mary	1	0	2	2
Bill	2	1	2	0
John	0	0	1	2
Larry	0	2	0	1

A^4

Multiplying a matrix by its transpose

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14
EVELYN	1	1	1	1	1	1	0	1	1	0	0	0	0	0
LAURA	1	1	1	0	1	1	1	1	0	0	0	0	0	0
THERESA	0	1	1	1	1	1	1	1	1	0	0	0	0	0
BRENDA	1	0	1	1	1	1	1	1	0	0	0	0	0	0
CHARLOTTE	0	0	1	1	1	0	1	0	0	0	0	0	0	0
FRANCES	0	0	1	0	1	1	0	1	0	0	0	0	0	0
ELEANOR	0	0	0	0	1	1	1	1	0	0	0	0	0	0
PEARL	0	0	0	0	0	1	0	1	1	0	0	0	0	0
RUTH	0	0	0	0	1	0	1	1	1	0	0	0	0	0
VERNE	0	0	0	0	0	0	1	1	1	0	0	1	0	0
MYRNA	0	0	0	0	0	0	0	1	1	1	0	1	0	0
KATHERINE	0	0	0	0	0	0	0	1	1	1	0	1	1	1
SYLVIA	0	0	0	0	0	0	1	1	1	1	0	1	1	1
NORA	0	0	0	0	0	1	1	0	1	1	1	1	1	1
HELEN	0	0	0	0	0	0	1	1	0	1	1	1	0	0
DOROTHY	0	0	0	0	0	0	0	1	1	0	0	0	0	0
OLIVIA	0	0	0	0	0	0	0	0	1	0	1	0	0	0
FLORA	0	0	0	0	0	0	0	0	1	0	1	0	0	0

	EVELYN	LAURA	THERESA	BRENDA	CHARLOTTE	FRANCES	ELEANOR	PEARL	RUTH	VERNE	MYRNA	KATHERINE	SYLVIA	NORA	HELEN	DOROTHY	OLIVIA	FLORA
E1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E3	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
E4	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
E5	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0
E6	1	1	1	1	0	1	1	1	0	0	0	0	0	1	0	0	0	0
E7	0	1	1	1	1	0	1	0	1	1	0	0	1	1	1	0	0	0
E8	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	0	0
E9	1	0	1	0	0	0	0	1	1	1	1	1	1	1	0	1	1	1
E10	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0
E11	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1
E12	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0
E13	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0
E14	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0

Woman by Woman Matrix

	EVELYN	LAURA	THERESA	BRENDA	CHARLOTTE	FRANCES	ELEANOR	PEARL	RUTH	VERNE	MYRNA	KATHERINE	SYLVIA	NORA	HELEN	DOROTHY	OLIVIA	FLORA
EVELYN	8	6	7	6	3	4	3	3	3	2	2	2	2	2	1	2	1	1
LAURA	6	7	6	6	3	4	4	2	3	2	1	1	2	2	2	1	0	0
THERESA	7	6	8	6	4	4	4	3	4	3	2	2	3	3	2	2	1	1
BRENDA	6	6	6	7	4	4	4	2	3	2	1	1	2	2	2	1	0	0
CHARLOTTE	3	3	4	4	4	2	2	0	2	1	0	0	1	1	1	0	0	0
FRANCES	4	4	4	4	2	4	3	2	2	1	1	1	1	1	1	1	0	0
ELEANOR	3	4	4	4	2	3	4	2	3	2	1	1	2	2	2	1	0	0
PEARL	3	2	3	2	0	2	2	3	2	2	2	2	2	2	1	2	1	1
RUTH	3	3	4	3	2	2	3	2	4	3	2	2	3	2	2	2	1	1
VERNE	2	2	3	2	1	1	2	2	3	4	3	3	4	3	3	2	1	1
MYRNA	2	1	2	1	0	1	1	2	2	3	4	4	4	3	3	2	1	1
KATHERINE	2	1	2	1	0	1	1	2	2	3	4	6	6	5	3	2	1	1
SYLVIA	2	2	3	2	1	1	2	2	3	4	4	6	7	6	4	2	1	1
NORA	2	2	3	2	1	1	2	2	2	3	3	5	6	8	4	1	2	2
HELEN	1	2	2	2	1	1	2	1	2	3	3	3	4	4	5	1	1	1
DOROTHY	2	1	2	1	0	1	1	2	2	2	2	2	2	1	1	2	1	1
OLIVIA	1	0	1	0	0	0	0	1	1	1	1	1	1	2	1	1	2	2
FLORA	1	0	1	0	0	0	0	1	1	1	1	1	1	2	1	1	2	2