Elementary Graph Theory & Matrix Algebra

Steve Borgatti

Drawn from:
2008 LINKS Center Summer SNA Workshops
Steve Borgatti, Rich DeJordy, & Dan Halgin

Introduction

- In social network analysis, we draw on three major areas of mathematics regularly:
  - Relations
    - Branch of math that deals with mappings between sets, such as objects to real numbers (measurement) or people to people (social relations)
  - Matrix Algebra
    - Tables of numbers
    - Operations on matrices enable us to draw conclusions we couldn’t just intuit
  - Graph Theory
    - Branch of discrete math that deals with collections of ties among nodes and gives us concepts like paths

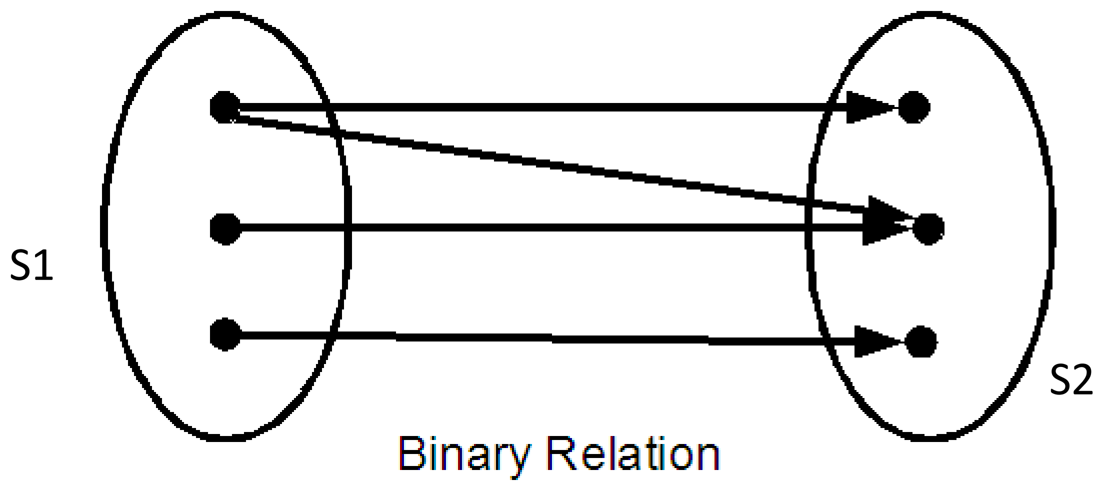
BINARY RELATIONS

Binary Relations

- The Cartesian product $S1 \times S2$ of two sets is the set of all possible ordered pairs (u,v) in which $u \in S1$ and $v \in S2$
  - Set {a,b,c,d}
  - Ordered pairs:
    - (a,a), (a,b), (a,c), (a,d)
    - (b,a),(b,b), (b,c), (b,d)
    - (c,a),(c,b), (c,c), (c,d)
    - (d,a),(d,b),(d,c),(d,d)

Binary Relations

- Given sets S1 and S2, a binary relation R is a subset of their Cartesian product



Note: S1 and S2 could be the same set

Relational Terminology

- To indicate that "u is R-related to v" or "u is mapped to v by the relation R", we write
  - $(u,v) \in R$, or
  - uRv

$$u \xrightarrow{\text{likes}} v$$

- Example: If R is "likes", then
  - uRv says u likes v
  - $(\text{jim},\text{jane}) \in R$ says jim likes jane

Functions

- A function is a relation that is many to one. If F is a function, then there can only be one v such that uFv
- Function form
  - v = F(u) means that uFv
  - So if F is “likes” then v=F(u) says that the person u likes is v. That is uFv, or u likes v

Properties of Relations

- A relation is *reflexive* if for all u, $(u,u) \in R$
  - E.g., suppose R is “is in the same room as”
  - u is always in the same room as u, so the relation is reflexive
- A relation is *symmetric* if for all u and v, uRv implies vRu
  - If u is in the same room as v, then it always true that v is in the same room as u. So the relation is symmetric
- A relation is *transitive* if for all u,v,w, the presence of uRv together with vRw implies uRw
  - If u is in the same room as v, and v is in the same room as w, then u is necessarily in the same room as w
  - So the relation is transitive
- A relation is an *equivalence* if it is reflexive, symmetric and transitive
  - The relation “is in the same room as” is reflexive, symmetric and transitive

Equivalences and Partitions

- A partition P of a set S is an exhaustive set of mutually exclusive classes such that each member of S belongs to one and only one class
    - E.g., any categorical variable like gender or cluster id
- – We use the notation p(u) to indicate the class that item u belongs to in partition P
- Equivalence relations give rise to partitions and vice-versa
    - – The relation “is in the same class as” is an equivalence relation

Operations

- The *converse* or *inverse* of a relation R is denoted R^{-1} (but we will often use R' instead)
  - For all u and v, $(u,v)\in R^{-1}$ if and only if $(v,u)\in R$
  - The converse reverses the direction of the mapping
- Example
  - If R is represents "gives advice to", then
    - uRv means u gives advice to v, and
    - $uR^{-1}v$ indicates that v gives advice to u
- If R is symmetric, then $R = R^{-1}$

Important note: In the world of matrices, the relational converse corresponds to the matrix concept of a transpose, denoted X' or X^T, and not to the matrix inverse, denoted X^{-1}. The $^{-1}$ superscript and the term "inverse" are unfortunate false cognates.

Relational Composition

- If F and E are binary relations, then their composition $F \circ E$ is a new relation such that $(u,v) \in F \circ E$ if there exists *w* such that $(u,w) \in F$ and $(w,v) \in E$.
  - i.e., u is $F \circ E$-related to v if there exists an intermediary w such that u is F-related to w and w is E-related to v
- Example:
  - Suppose F and E are friend of and enemy of, respectively
  - u $F \circ E$ v means that u has a friend who is the enemy of v
- This “right” notation* which means rightmost relations are applied first
  - start from the end and ask “what is v to u?”
  - u $F \circ E$ v means that v is the enemy of a friend of u
    - In functional notation v=E(F(u))

*Important note: Many authors reverse the meaning of $F \circ E$, writing it as $E \circ F$. This is known as “left” convention, meaning that the left relation is applied first. So $uF \circ Ev$ would mean v is the friend of an enemy of u. That is v = F(E(u))

More Relational Composition

Assume F is “likes”

- u $F \circ F$ v means u likes someone who likes v (v is liked by someone who is liked by u)
  - If uFv = u $F \circ F$ v for all u and v, we have transitivity
- u $F \circ F^{-1}$ v means u likes someone who is liked by v
  - Both u and v like w
- u $F^{-1} \circ F$ v means u is liked by someone who likes v (v is liked by someone who likes u)
  - Both u and v are liked by w

Relations can relate different kinds of items

- "is tasked with" relates persons to tasks they are responsible for
  - uTv means person u is responsible for task v
- "controls resource" relates persons to resources they control
  - uCv means person u controls resource v
- "requires resource" relates tasks to the resources needed to accomplish them
  - uRv means task u requires resource v

These kinds of relations can be composed as well

- If T is “tasked with”, C is “controls”, and R is “requires”, then
  - $uT \circ Rv$ means person u is tasked with a task that requires resource v
  - $uT \circ R \circ C^{-1}v$ means person u is tasked with a task that requires a resource that is controlled by person v
    - i.e., u is dependent on v to get something done

Relational Equations

- $F = F \circ F$ means that uFv if and only if $uF \circ Fv$, for all u and v
  - Friends of friends are always friends, and vice versa
  - Transitivity plus embeddedness
- $F = E \circ E$ means that uFv if and only if $uE \circ Ev$
  - Enemies of enemies are friends, and all friends have common enemies
- $E = F \circ E = E \circ F$ means that uEv if and only if $uF \circ Ev$ and $uE \circ Fv$
  - Both enemies of friends and friends of enemies are enemies, and vice-versa

Matrix Algebra

- In this section, we will cover:
  - Matrix Concepts, Notation & Terminologies
  - Adjacency Matrices
  - Transposes
  - Aggregations & Vectors
  - Matrix Operations
  - Boolean Algebra (and relational composition)

Matrices

- Matrices are simply tables. Sometimes multidimensional
- Symbolized by a capital letter, like A
- Each cell in the matrix identified by row and column subscripts: 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

$a_{12} = 1$

$a_{43} = 8000$

Vectors

- Each row and each column in a matrix is a vector
  - Vertical vectors are column vectors, horizontal are row vectors
- Denoted by lowercase bold letter: **y**
- Each cell in the vector identified by subscript z_i

$y_3 = 2.1$

$z_2 = 45,000$

	X	Y	Z
Mary	32	1	90,000
Bill	50	2	45,000
John	12	2.1	0
Larry	20	2	8,000

Ways and Modes

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

	Event 1	Event 2	Event 3	Event 4
EVELYN	1	1	1	1
LAURA	1	1	1	0
THERESA	0	1	1	1
BRENDA	1	0	1	1
CHARLO	0	0	1	1
FRANCES	0	0	1	0
ELEANOR	0	0	0	0
PEARL	0	0	0	0
RUTH	0	0	0	0
VERNE	0	0	0	0
MYRNA	0	0	0	0

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

1-Mode Matrices

- Item by item proximity matrices
  - Correlation matrices
    - Matrix of correlations among variables
  - Distance matrices
    - Physical distance between cities
  - Adjacency matrices
    - Actor by actor matrices that record who has a tie of a given kind with whom
    - Strength of tie

Adjacency matrix

Which 3 people did you interact with the most last week?

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	HOLLY	-	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0
2	BRAZEY	0	-	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0
3	CAROL	0	0	-	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0
4	PAM	0	0	0	-	0	1	1	1	0	0	0	0	0	0	0	0	0	0
5	PAT	1	0	1	0	-	1	0	0	0	0	0	0	0	0	0	0	0	0
6	JENNIE	0	0	0	1	1	-	0	1	0	0	0	0	0	0	0	0	0	0
7	PAULINE	0	0	1	1	1	0	-	0	0	0	0	0	0	0	0	0	0	0
8	ANN	0	0	0	1	0	1	1	-	0	0	0	0	0	0	0	0	0	0
9	MICHAEL	1	0	0	0	0	0	0	0	-	0	0	1	0	1	0	0	0	0
10	BILL	0	0	0	0	0	0	0	0	1	-	0	1	0	1	0	0	0	0
11	LEE	0	1	0	0	0	0	0	0	0	0	-	0	0	0	0	1	1	0
12	DON	1	0	0	0	0	0	0	0	1	0	0	-	0	1	0	0	0	0
13	JOHN	0	0	0	0	0	0	1	0	0	0	0	0	-	0	1	0	0	1
14	HARRY	1	0	0	0	0	0	0	0	1	0	0	1	0	-	0	0	0	0
15	GERY	0	0	0	0	0	0	0	0	1	0	0	0	0	0	-	1	0	1
16	STEVE	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	-	1	1
17	BERT	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	-	1
18	RUSS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	-

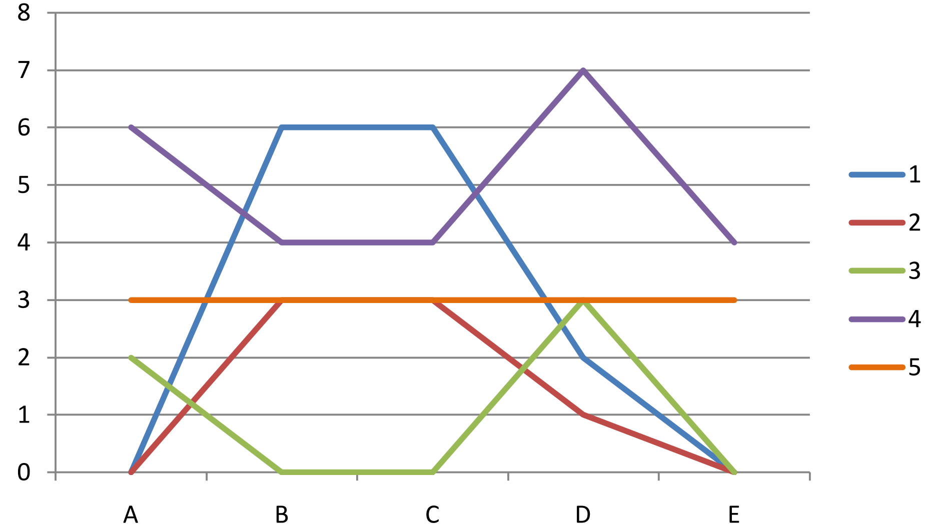
2-Mode Matrices

- Profile matrices
  - Individuals' scores on a set of personality scales
- Participation in events; membership in groups

Profile Matrices

- Typically, we use profiles to refer to the patterns of responses across a row of a matrix, generally a 2-mode matrix.
- We might then compare profiles across the rows to see which rows have the most similar or dissimilar profiles.
  - We can also conceive of this down the columns, as well. In fact, when we correlate variables in traditional OLS, we are actually comparing the profiles of each pair of variables across the respondents.

ID	A	B	C	D	E
1	0	6	6	2	0
2	0	3	3	1	0
3	2	0	0	3	0
4	6	4	4	7	4
5	3	3	3	3	3



Aggregations and Operations

- Unary (Intra-Matrix) Operations
  - Row sums/marginals
  - Column sums/marginals
  - Matrix Sums
  - Transpose
  - Normalizations
  - Dichotomization
  - Symmetrizing
- Cellwise Binary (Inter-Matrix) Operations
  - Sum
  - Cellwise multiplication
  - Boolean Operations
- Special Binary (Inter-Matrix) Operations
  - Cross Product (Matrix Multiplication)

Summations

- Row sums (aka row marginals)

$$r_i = \sum_j x_{ij} = [3,2,1,0]'$$

- Column sums

$$c_j = \sum_i x_{ij} = [1,1,2,2]$$

- Matrix sums

$$m = \sum_{i,j} x_{ij} = 6$$

	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

$r_3 = 1$ $\quad c_3 = 2$ $\quad m = 6$

Normalizing

- Converting to proportions
  - Rows $$x_{ij}^* = \frac{x_{ij}}{r_i}$$ where r_i gives the sum of row i
  - Columns $$x_{ij}^* = \frac{x_{ij}}{c_j}$$

	Mary	Bill	John	Larry	Row Sums
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

	Mary	Bill	John	Larry	Row sums
Mary	0	.33	.33	.33	1
Bill	.5	0	.5	0	1
John	0	0	0	1	1
Larry					
Column Marginals	.5	.33	.83	1.33	3

Normalizing

- Converting to z-scores (standardizing)
  - Columns

$$x_{ij}^{*} = \frac{x_{ij} - u_j}{\sigma_j}$$

where u_j gives the mean of column j, and σ_j is the std deviation of column j

	Var 1	Var 2	Var 3	Var 4
Mary	3	20	25	10
Bill	1	55	15	45
John	0	32	10	22
Larry	2	2	20	-8
Mean	1.5	27.3	17.5	17.3
Std Dev	1.1	19.3	5.6	19.3

	Var 1	Var 2	Var 3	Var 4
Mary	1.34	-0.38	1.34	-0.38
Bill	-0.45	1.44	-0.45	1.44
John	-1.34	0.25	-1.34	0.25
Larry	0.45	-1.31	0.45	-1.31
Mean	0.00	0.00	0.00	0.00
Std Dev	1.00	1.00	1.00	1.00

Transposes

- Transpose of matrix M is denoted M' or M^T
- The transpose of a matrix is created by interchanging rows and columns
  - For all i and j, $m_{ij}^{T} = m_{ji}$
  - So the transpose of an m by n matrix is an n by m matrix

	A	B	C	D	E
1	0	6	6	2	0
2	0	3	3	1	0
3	2	0	0	3	0
4	6	4	4	7	4
5	3	3	3	3	3

Matrix M

	1	2	3	4	5
A	0	0	2	6	3
B	6	3	0	4	3
C	6	3	0	4	3
D	2	1	3	7	3
E	0	0	0	4	3

Its transpose, M'

Transpose (Another Example)

- Given Matrix M, swap the rows and columns to make Matrix M^T

M	Tennis	Football	Rugby	Golf
Mike	0	0	1	0
Ron	0	1	1	0
Pat	0	0	0	1
Bill	1	1	1	1
Joe	0	0	0	0
Rich	0	1	1	1
Peg	1	1	0	1

M^T	Mike	Ron	Pat	Bill	Joe	Rich	Peg
Tennis	0	0	0	1	0	0	1
Football	0	1	0	1	0	1	1
Rugby	1	1	0	1	0	1	0
Golf	0	0	1	1	0	1	1

Dichotomizing

- X is a valued matrix, say 1 to 10 rating of strength of tie
- Construct a matrix Y of ones and zeros so that
  - $y_{ij} = 1$ if $x_{ij} > 5$, and $y_{ij} = 0$ otherwise

X

	EVE	LAU	THE	BRE	CHA
EVELYN	8	6	7	6	3
LAURA	6	7	6	6	3
THERESA	7	6	8	6	4
BRENDA	6	6	6	7	4
CHARLOTTE	3	3	4	4	4

Y

	EVE	LAU	THE	BRE	CHA
EVELYN	1	1	1	1	0
LAURA	1	1	1	1	0
THERESA	1	1	1	1	0
BRENDA	1	1	1	1	0
CHARLOTTE	0	0	0	0	0

$x_{ij} > 5$

Symmetrizing

- When matrix is not symmetric, i.e., $x_{ij} \neq x_{ji}$
- Symmetrize various ways. Set y_{ij} and y_{ji} to:
  - Maximum(x_{ij}, x_{ji}) {union rule}
  - Minimum (x_{ij}, x_{ji}) {intersection rule}
  - Average: $(x_{ij} + x_{ji})/2$
  - Lowerhalf: choose x_{ij} when $i > j$ and x_{ji} otherwise
  - etc

Symmetrizing Example

- X is non-symmetric (and happens to be valued)
- Construct matrix Y such that y_{ij} (and y_{ji}) = maximum of x_{ij} and x_{ji}

X

	ROM	BON	AMB	BER	PET	LOU
ROMUL_10	0	1	1	0	3	0
BONAVEN_5	0	0	1	0	3	2
AMBROSE_9	0	1	0	0	0	0
BERTH_6	0	1	2	0	3	0
PETER_4	0	3	0	1	0	2
LOUIS_11	0	2	0	0	0	0

	ROM	BON	AMB	BER	PET	LOU
ROMUL_10	0	1	1	0	3	0
BONAVEN_5	1	0	1	1	3	2
AMBROSE_9	1	1	0	2	0	0
BERTH_6	0	1	2	0	3	0
PETER_4	3	3	0	3	0	2
LOUIS_11	0	2	0	0	2	0

Symmetrized by Maximum

Cellwise Binary Operators

- Sum (Addition)

  $C = A + B$ where $c_{ij} = a_{ij} + b_{ij}$
- Cellwise (Element) Multiplication

  $C = A * B$ where $c_{ij} = a_{ij} * b_{ij}$
- Boolean operations

  $C = A \wedge B$ (Logical And) where $c_{ij} = a_{ij} \wedge b_{ij}$

  $C = A \vee B$ (Logical Or) where $c_{ij} = a_{ij} \vee b_{ij}$

Matrix Multiplication

- Notation: C = AB
- Definition: $$c_{ij} = \sum_k a_{ik} b_{kj}$$

Note: matrix products are not generally commutative. i.e., AB does not usually equal BA

- 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

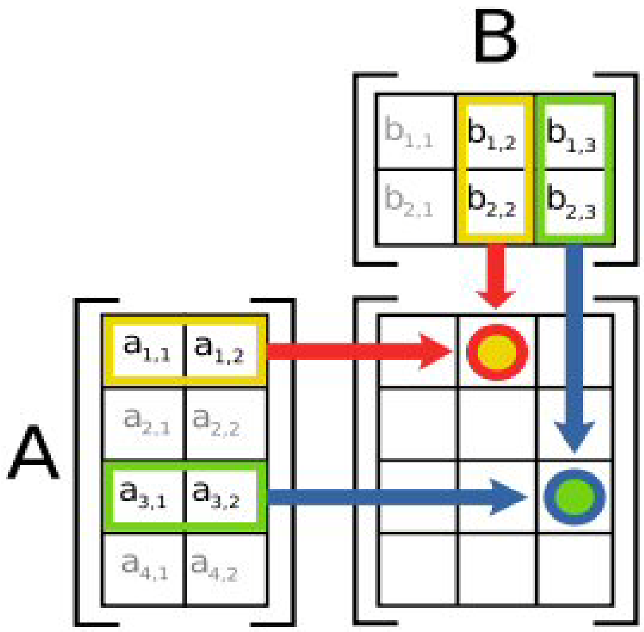
	Mary	Bill	John	Larry
Mary	0	0	1	1
Bill	1	0	1	0
John	0	0	0	1
Larry	0	1	0	0

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

Matrix Multiplication



- C = AB or C = A x B
  - Only possible when the number of columns in A is the same as the number of rows in B, as in $_mA_k$ and $_kB_n$
  - These are said to be <u>conformable</u>
  - Produces $_mC_n$
- It is calculated as:

  $c_{ij} = \Sigma\, a_{ik} * b_{kj}$ for all k

$$\begin{bmatrix} 1 & 0 & 2 \\ -1 & 3 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 & 1 \\ 2 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 \times 3 + 0 \times 2 + 2 \times 1 & 1 \times 1 + 0 \times 1 + 2 \times 0 \\ -1 \times 3 + 3 \times 2 + 1 \times 1 & -1 \times 1 + 3 \times 1 + 1 \times 0 \end{bmatrix} = \begin{bmatrix} 5 & 1 \\ 4 & 2 \end{bmatrix}$$

A Matrix Product Example

Skills	Math	Verbal	Analytic
Kev	1.00	.75	.80
Jeff	.80	.80	.90
Lisa	.75	.60	.75
Kim	.80	1.00	.85

Items	Q1	Q2	Q3	Q4
Math	.50	.75	0	.1
Verbal	.10	0	.9	.1
Analytic	.40	.25	.1	.8

- Given a Skills and Items matrix calculate the "affinity" that each person has for each question
- Kev for Question 1 is:
  = 1.00 * .5 + .75* .1 + .80 * .40
  = .5 + .075 + .32 = **0.895**
- Lisa for Question 3 is:
  = .75 * .0 + .60* .90 + .75 * .1
  = .0 + .54 + .075 = **0.615**

Affin	Q1	Q2	Q3	Q4
Kev	0.895	0.95	0.755	0.815
Jeff	0.840	0.825	0.810	0.880
Lisa	0.735	0.75	0.615	0.735
Kim	0.840	0.813	0.985	0.860

Matrix Inverse and Identity

- The inverse of a matrix X is a Matrix X^{-1} such that $XX^{-1} = I$, where I is the identity matrix
- Inverse matrices can be very useful for solving matrix equations that underlie some network algorithms

1	0	-2
4	1	0
1	1	7

X

7	-2	2
-28	9	-8
3	-1	1

X^{-1}

=

1	0	0
0	1	0
0	0	1

I

Note:

- $(XX^{-1} = X^{-1}X = I)$
- Non square matrices do not have an inverse*

Linear Combinations

- Multiply matrix X by vector **b**
  - X consists of scores obtained by persons (rows) on tests (columns)
  - **b** is a set of weights for each test
  - Matrix product **y**=X**b** gives the sum of scores for each person, with each test weighted according to b
  - The cells of **y** are constructed as follows:

$$y_i = \sum_j x_{ij} b_j = x_{i1} b_1 + x_{i2} b_2 + \dots$$

y
56.75
48.75
54.50
35.75
28.00
71.25
34.00

=

X		
80	69	39
87	90	9
17	43	79
36	93	7
67	19	13
92	93	50
53	69	7

b
0.25
0.25
0.50

Regression in matrix terms

- $y = Xb$
- $X'y = X'Xb$
- $(X'X)^{-1}X'y = B$
- $y_i = b_1 x_{i1} + b_2 x_{i2} + \dots$

Regression in matrix terms

- We have matrix X whose columns are variables, and vector Y which is an outcome, and want to build model
  - $y_i = b_1 x_{i1} + b_2 x_{i2} + \ldots$
  - Trouble is, we don't know what the values of b are
- Express regression equation as matrix product
  - $\mathbf{y} = X\mathbf{b}$
- Now do a little algebra
  - $X'\mathbf{y} = X'X\mathbf{b}$ //pre-multiply both sides by X'
  - $(X'X)^{-1}X'\mathbf{y} = \mathbf{b}$ //pre-multiply by $(X'X)^{-1}$

Products of matrices & their transposes

- X'X = pre-multiplying X by its transpose

$$(X'X)_{ij} = \sum_k a_{ki} b_{kj}$$

- Computes sums of products of each pair of columns (cross-products)
- The basis for most similarity measures

	1	2	3	4
Mary	0	1	1	1
Bill	1	0	1	0
John	0	0	0	1
Larry	0	0	0	0

	1	2	3	4
1	1	0	1	0
2	0	1	1	1
3	1	1	2	1
4	0	1	1	2

Products of matrices & their transposes

- XX' = product of matrix X by its transpose

$$(XX')_{ij} = \sum_k a_{ik} b_{jk}$$

- Computes sums of products of each pair of rows (cross-products)
- Similarities among rows

	1	2	3	4
Mary	0	1	1	1
Bill	1	0	1	0
John	0	0	0	1
Larry	0	0	0	0

	Mary	Bill	John	Larry
Mary	3	1	1	0
Bill	1	2	0	0
John	1	0	1	0
Larry	0	0	0	0

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

	EV	LA	TH	BR	CH	FR	EL	PE	RU	VE	MY	KA	SY	NO	HE	DO	OL	FL
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

	EVE	LAU	THE	BRE	CHA	FRA	ELE	PEA	RUT	VER	MYR	KAT	SYL	NOR	HEL	DOR	OLI	FLO
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

Boolean matrix multiplication

- Values can be 0 or 1 for all matrices
- Products are dichotomized to conform:

Would have been a 2 in regular matrix multiplication

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

AB

Relational Composition

- If we represent binary relations as binary adjacency matrices, boolean matrix products correspond to relational composition
  - $F \circ E$ corresponds to FE

Likes

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

F

Has conflicts with

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

E

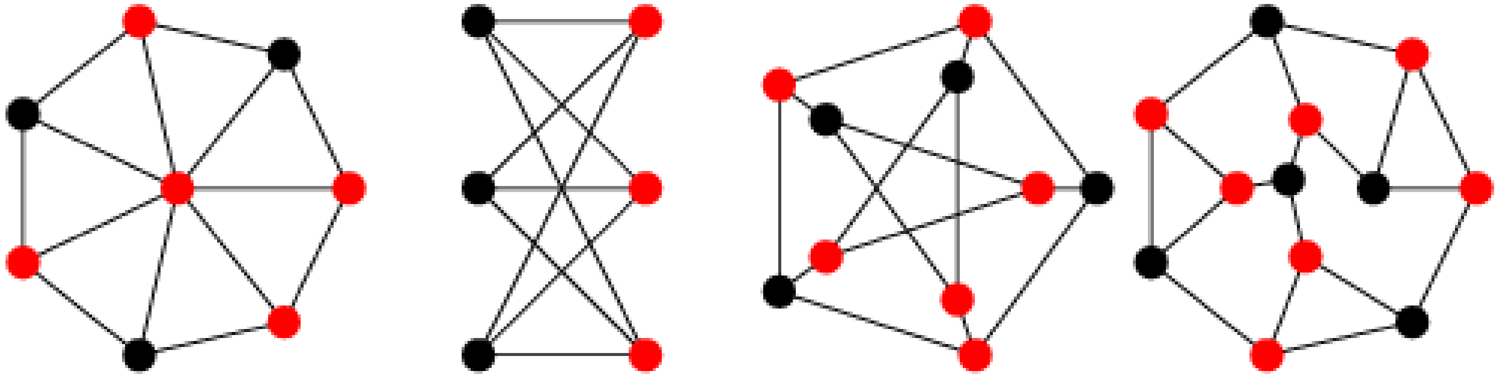
Likes someone who has conflicts with

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

FE

More Relational Composition

- Given these relations
  - A (authored). Relates persons → documents
  - P (published in). Relates docs → journals
  - K (has keyword). Relates docs → keywords
- Compositions
  - AA^{-1}. if $(i,j) \in AA^{-1}$, then *i* authors a documents that is authored by j. i.e., *i* and *j* are coauthors
  - AP. Person i authored a document that is published in journal j. so i has published in journal j
  - AK. Person i authored a doc that has keyword j. So, i writes about topic j
  - $AKK^{-1}A^{-1}$. person i authored a document that has a keyword that is in a document that was authored by j. In other words, i and j write about the same topics
  - $AKK^{-1}A^{-1}AP$. person i authored a document that has a keyword that is in a document that was authored by someone who has published in journal j. I.e., i has written about a topic that has appeared in journal j

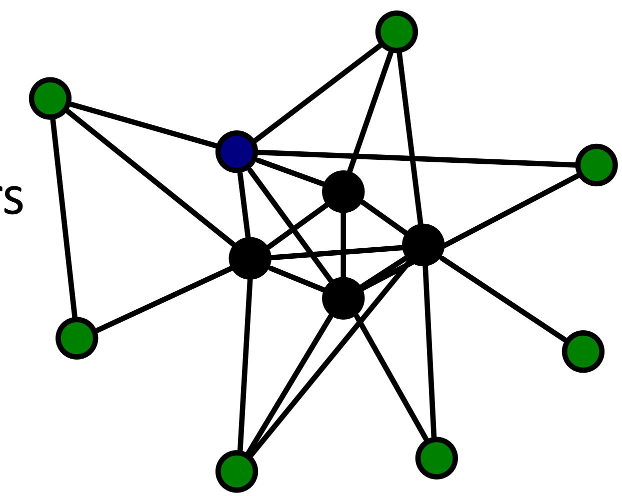


Graph Theoretic Concepts

- In this section we will cover:
  - Definitions
  - Terminology
  - Adjacency
  - Density concepts
    - E.g, Completeness
  - Walks, trails, paths
  - Cycles, Trees
  - Reachability/Connectedness
    - Connectivity, flows
  - Isolates, Pendants, Centers
  - Components, bi-components
  - Walk Lengths, distance
    - Geodesic distance
  - Independent paths
  - Cutpoints, bridges

Undirected Graphs

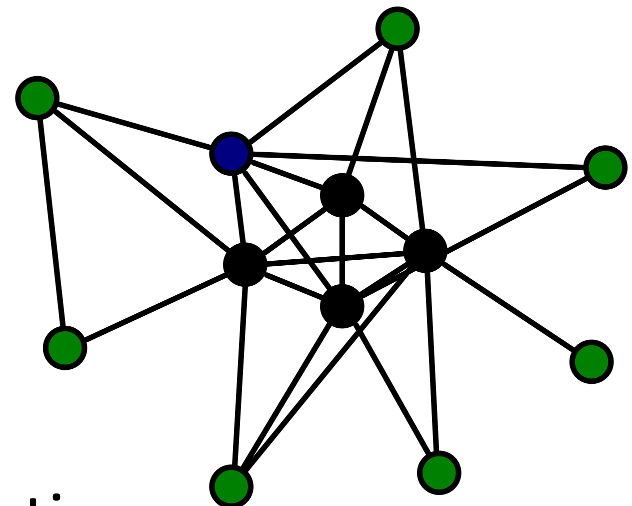
- An undirected graph G(V,E) consists of …
  - Set of nodes|vertices V representing actors
  - Set of lines|links|edges E representing ties among pairs of actors
    - An edge is an unordered pair of nodes (u,v)
    - Nodes u and v adjacent if $(u,v) \in E$
    - So E is subset of set of all pairs of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

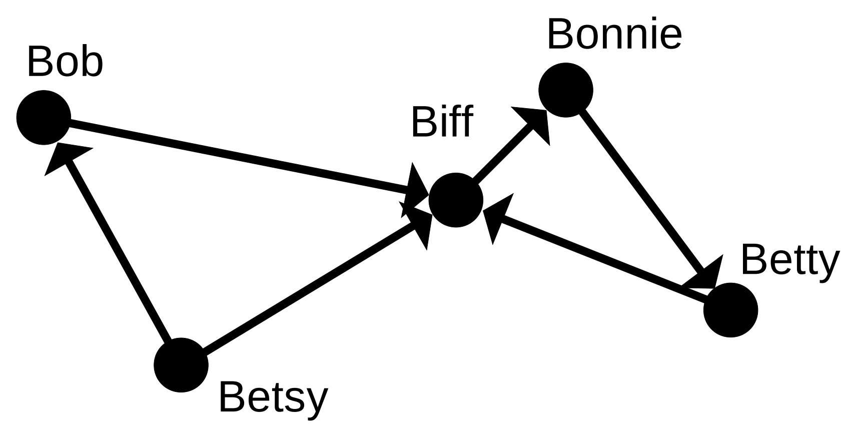


- Drawn without arrow heads
  - Sometimes with dual arrow heads
- Used to represent social relations where direction doesn't make sense, or symmetry is logically necessary
  - In communication with; attending same meeting as

Directed vs. Undirected Ties

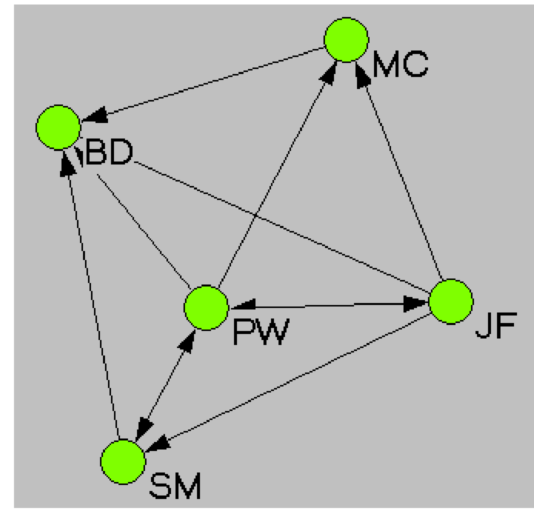
- Undirected relations
  - Attended meeting with
  - Communicates daily with
- Directed relations
  - Lent money to
- Logically vs empirically directed ties
  - Empirically, even un-directed relations can be non-symmetric due to measurement error





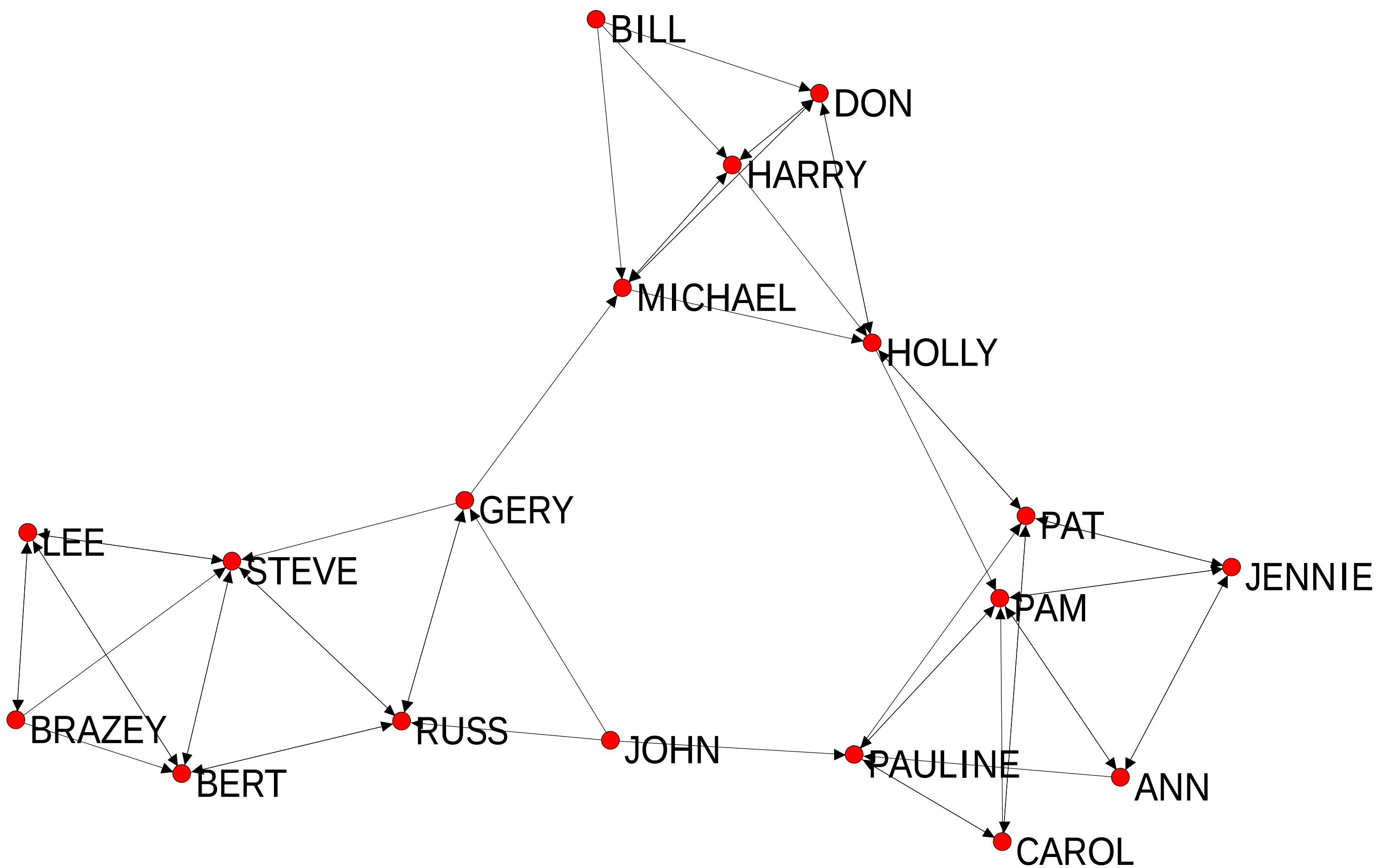
Directed Graphs (Digraphs)

- 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) $\in$ E indicates u sends arc to v
    - (u,v) $\in$ E does <u>not</u> necessarily imply that (v,u) $\in$ E (although it might happen)



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

Graphical representation of a digraph



Adjacency matrix of a digraph

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	HOLLY	1	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0
2	BRAZEY	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0
3	CAROL	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0
4	PAM	0	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0
5	PAT	1	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
6	JENNIE	0	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0
7	PAULINE	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0
8	ANN	0	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0
9	MICHAEL	1	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0
10	BILL	0	0	0	0	0	0	0	0	1	1	0	1	0	1	0	0	0	0
11	LEE	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0
12	DON	1	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0
13	JOHN	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1
14	HARRY	1	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0
15	GERY	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	1
16	STEVE	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1
17	BERT	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1
18	RUSS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1

Transposing adjacency matrix

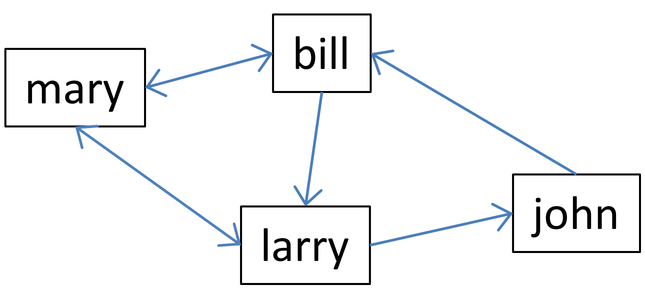
- Interchanging rows/columns of adjacency matrix effectively reverses the direction of ties

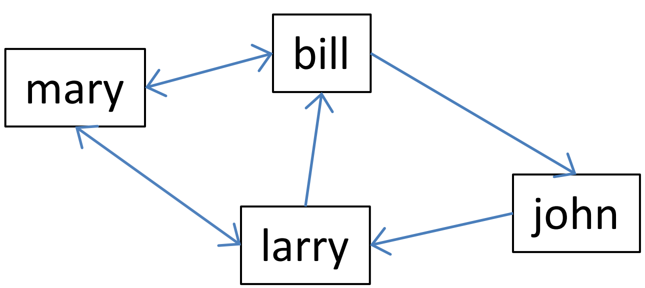
	Mary	Bill	John	Larry
Mary	0	1	0	1
Bill	1	0	0	1
John	0	1	0	0
Larry	1	0	1	0

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

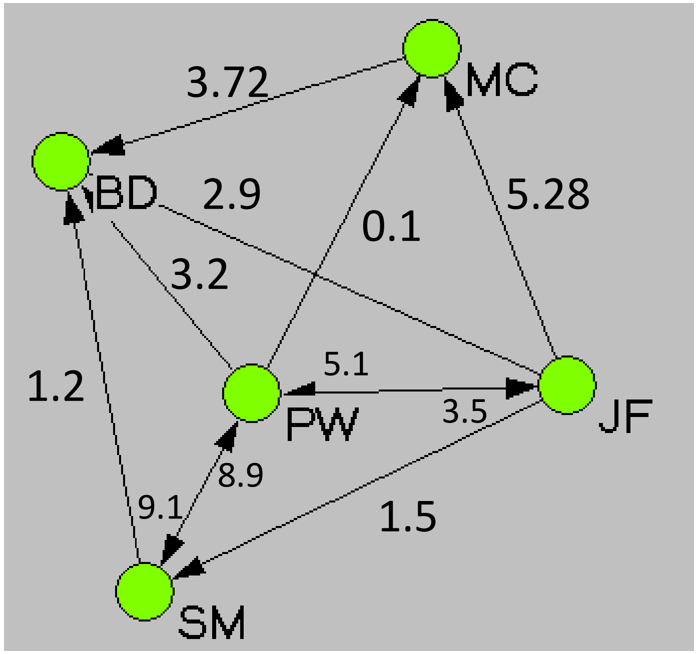
Gets money from





Valued Digraphs (vigraphs)

- A valued digraph G(V,E,W) 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 <u>not</u> imply that $(v,u) \in E$
  - Mapping W of arcs to real values



- Values can represent such things as
  - 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 Adjacency Matrix

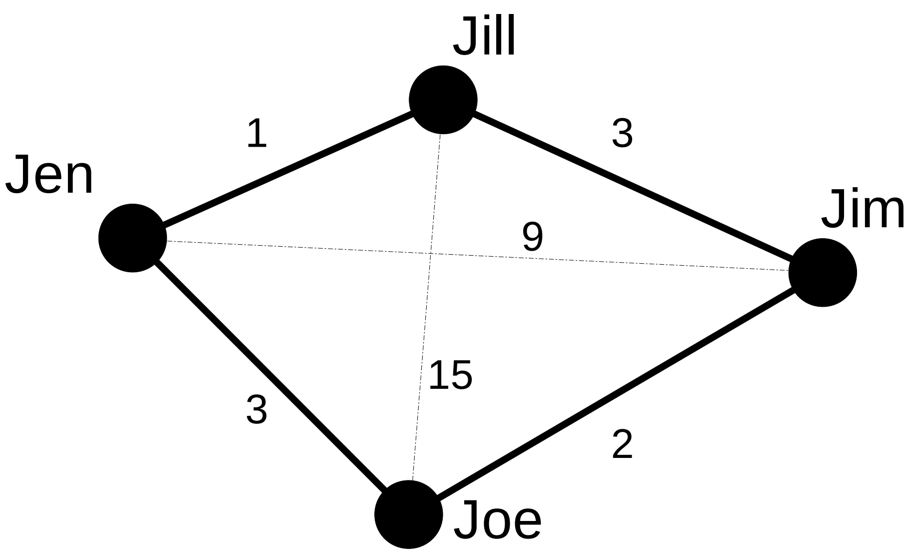
Dichotomized

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

Distances btw offices

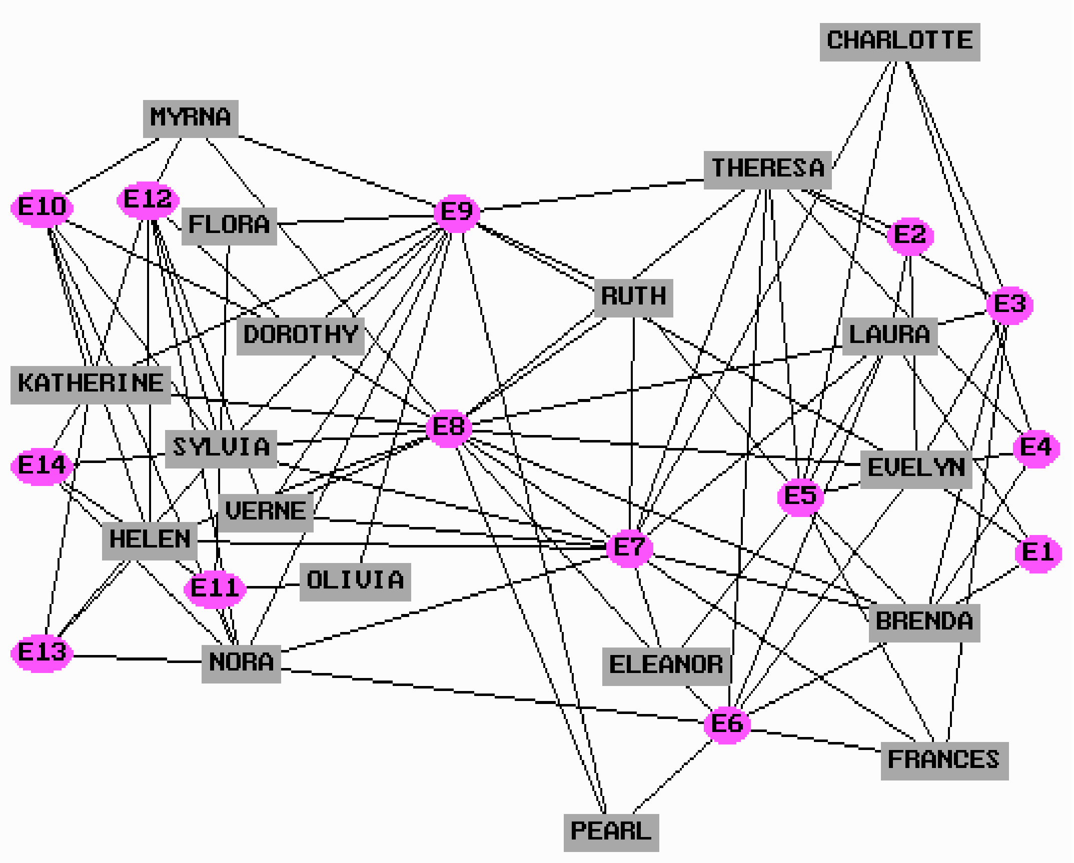
	Jim	Jill	Jen	Joe
Jim	-	3	9	2
Jill	3	-	1	15
Jen	9	1	-	3
Joe	2	15	3	-

- The diagram below uses solid lines to represent the adjacency matrix, while the numbers along the solid line (and dotted lines where necessary) represent the proximity matrix.
- In this particular case, one can derive the adjacency matrix by dichotomizing the proximity matrix on a condition of $p_{ij} <= 3$.



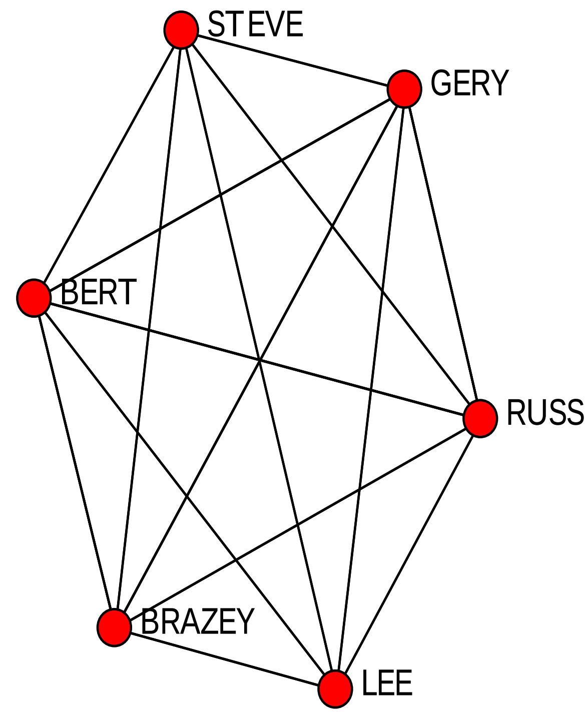
Bipartite graphs

- Used to represent 2-mode data
- Nodes can be partitioned into two sets (corresponding to modes)
- Ties occur only between sets, not within



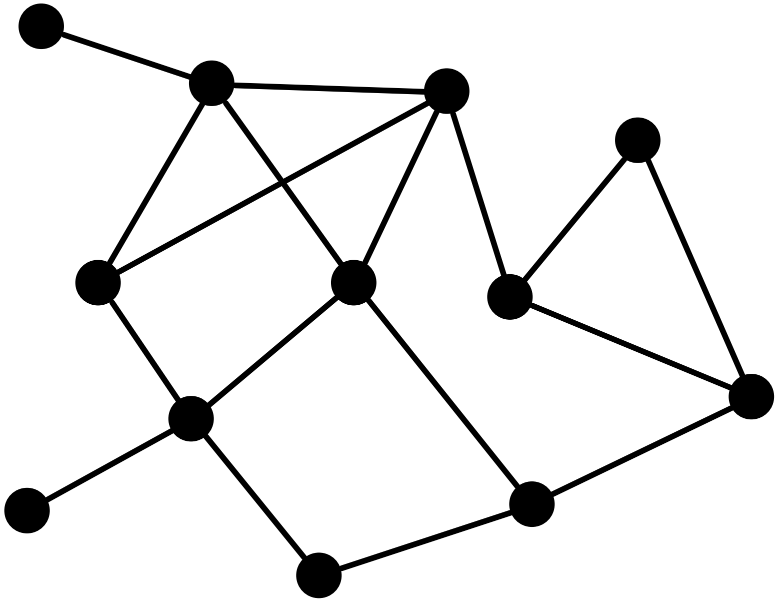
Density and Completeness

- A graph is *complete* if all possible edges are present.
- The *density* of a graph is the number of edges present divided by the number that could have been

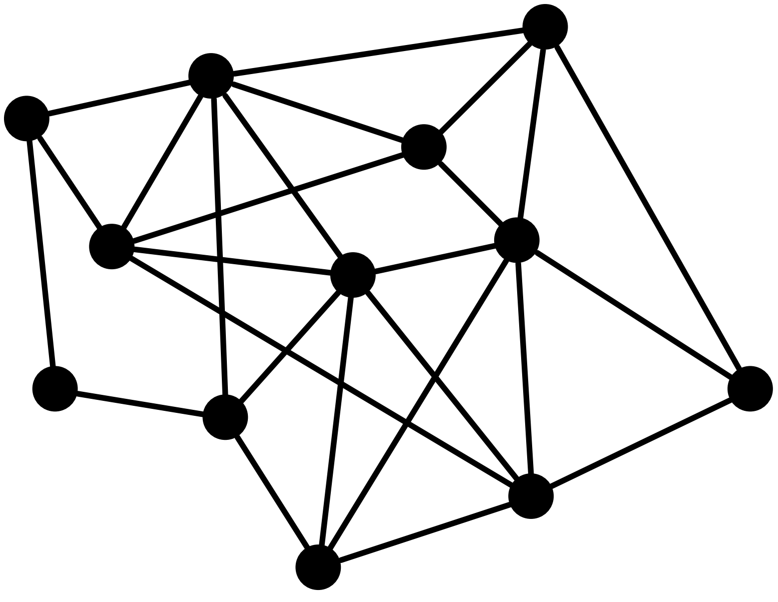


Density

- 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

Density

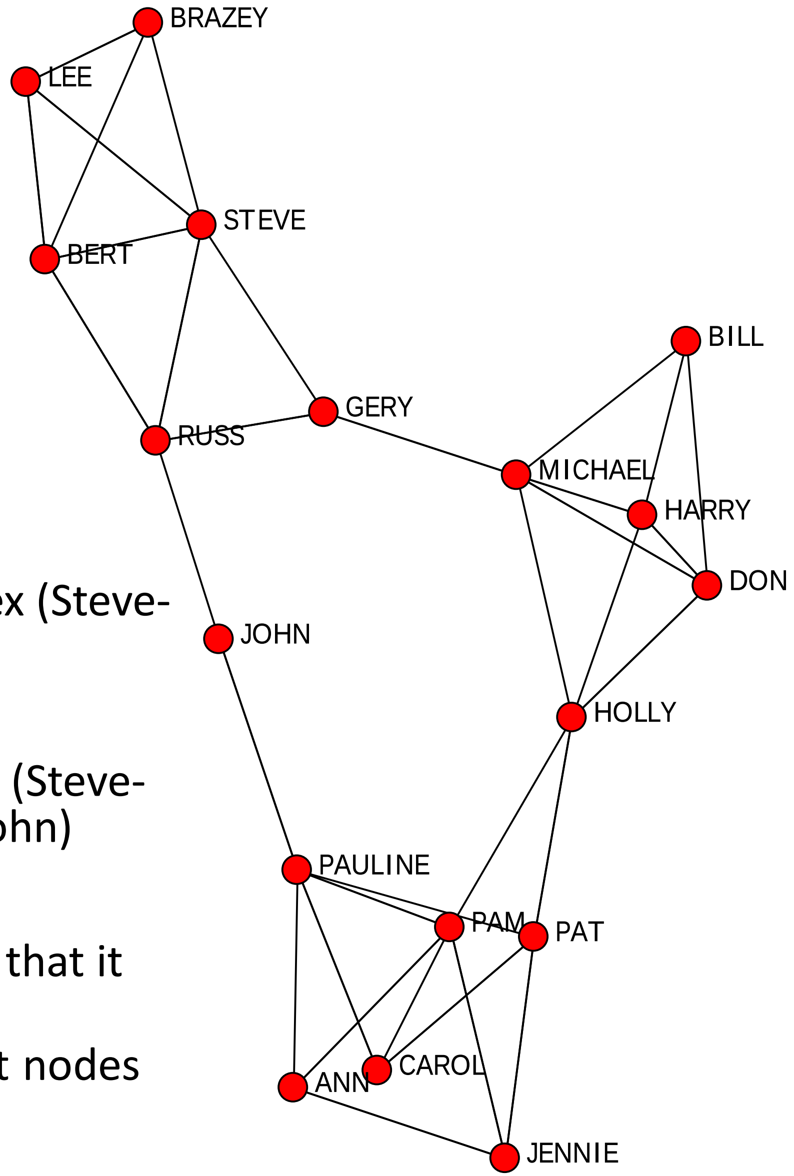
Number of ties divided by number possible

	Ties to Self Allowed	No ties to self
Undirected	$= \frac{T}{n^2/2}$	$= \frac{T}{n(n-1)/2}$
Directed	$= \frac{T}{n^2}$	$= \frac{T}{n(n-1)}$

T = number of ties in network
n = number of nodes

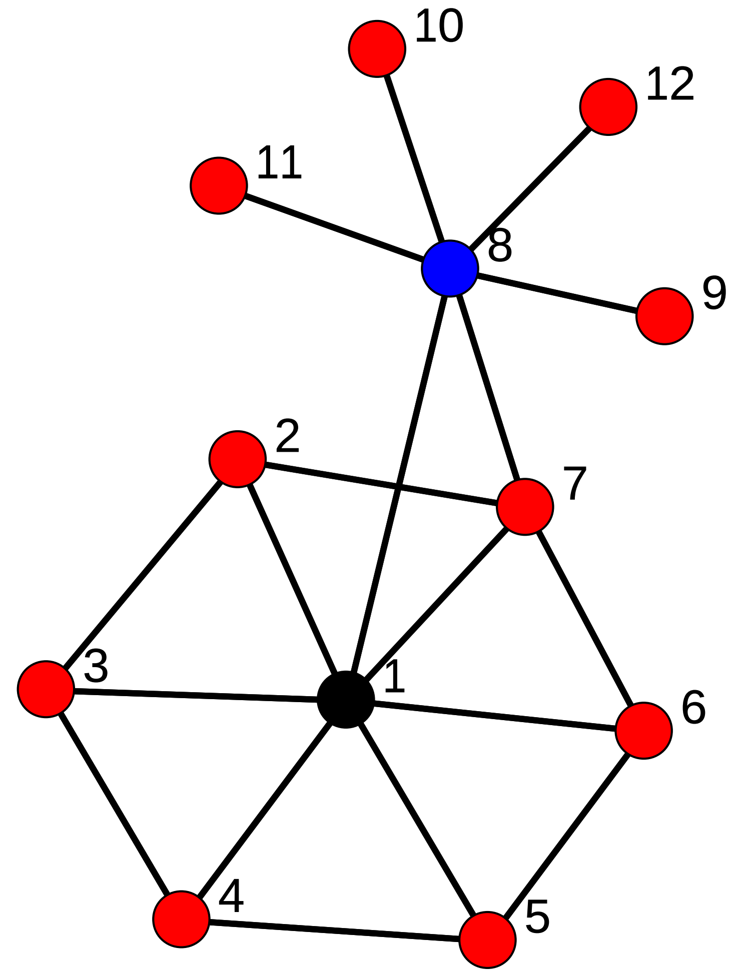
Graph traversals

- Walk
  - Any unrestricted traversing of vertices across edges (Russ-Steve-Bert-Lee-Steve)
- Trail
  - A walk restricted by not repeating an edge or arc, although vertices can be revisited (Steve-Bert-Lee-Steve-Russ)
- Path
  - A trail restricted by not revisiting any vertex (Steve-Lee-Bert-Russ)
- Geodesic Path
  - The shortest path(s) between two vertices (Steve-Russ-John is shortest path from Steve to John)
- Cycle
  - A cycle is in all ways just like a path except that it ends where it begins
  - Aside from endpoints, cycles do not repeat nodes
  - E.g. Brazey-Lee-Bert-Steve-Brazey



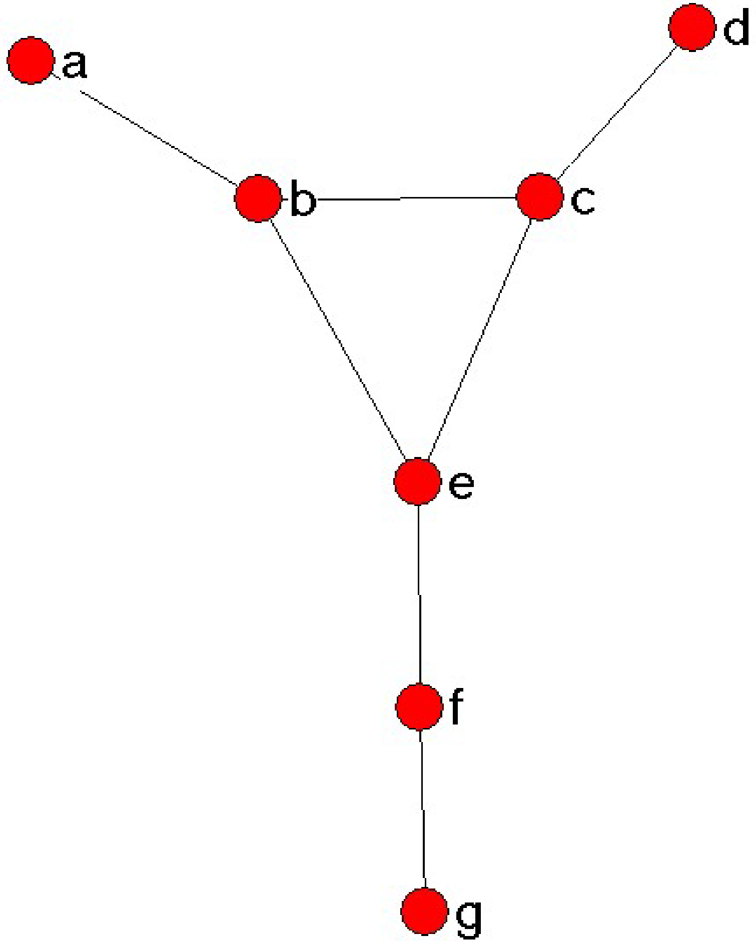
Length & Distance

- Length of a path (or any walk) is the number of links it has
- The **Geodesic Distance** (aka graph-theoretic distance) between two nodes is the length of the shortest path
  - Distance from 5 to 8 is 2, because the shortest path (5-1-8) has two links



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

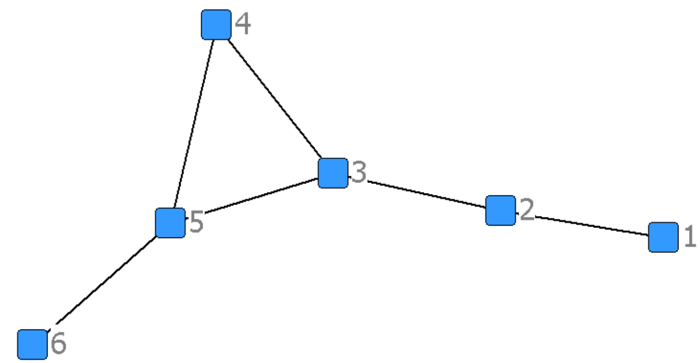


Powers of the adjacency matrix

- If you multiply an adjacency matrix X by itself, you get XX or X^2
- A given cell x^2_{ij} gives the number of walks from node i to node j of length 2
- More generally, the cells of X^k give the number of walks of length exactly k from each node to each other

Matrix powers example

Note that shortest path from 1 to 5 is three links, so $x_{1,5} = 0$ until we get to X^3



	1	2	3	4	5	6
1	0	1	0	0	0	0
2	1	0	1	0	0	0
3	0	1	0	1	1	0
4	0	0	1	0	1	0
5	0	0	1	1	0	1
6	0	0	0	0	1	0

X

	1	2	3	4	5	6
1	1	0	1	0	0	0
2	0	2	0	1	1	0
3	1	0	3	1	1	1
4	0	1	1	2	1	1
5	0	1	1	1	3	0
6	0	0	1	1	0	1

X^2

	1	2	3	4	5	6
1	0	2	0	1	1	0
2	2	0	4	1	1	1
3	0	4	2	4	5	1
4	1	1	4	2	4	1
5	1	1	5	4	2	3
6	0	1	1	1	3	0

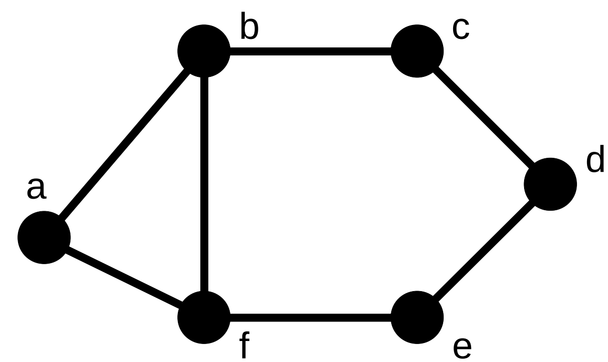
X^3

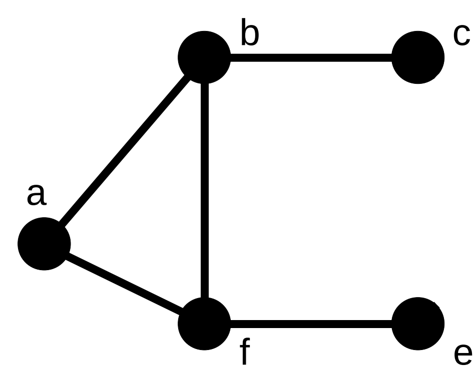
	1	2	3	4	5	6
1	2	0	4	1	1	1
2	0	6	2	5	6	1
3	4	2	13	7	7	5
4	1	5	7	8	7	4
5	1	6	7	7	12	2
6	1	1	5	4	2	3

X^4

Subgraphs

- Set of nodes
  - Is just a set of nodes
- A subgraph
  - Is set of nodes together with ties among them
- An induced subgraph
  - Subgraph defined by a set of nodes
  - Like pulling the nodes and ties out of the original graph





Subgraph induced by considering the set {a,b,c,f,e}

Components

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

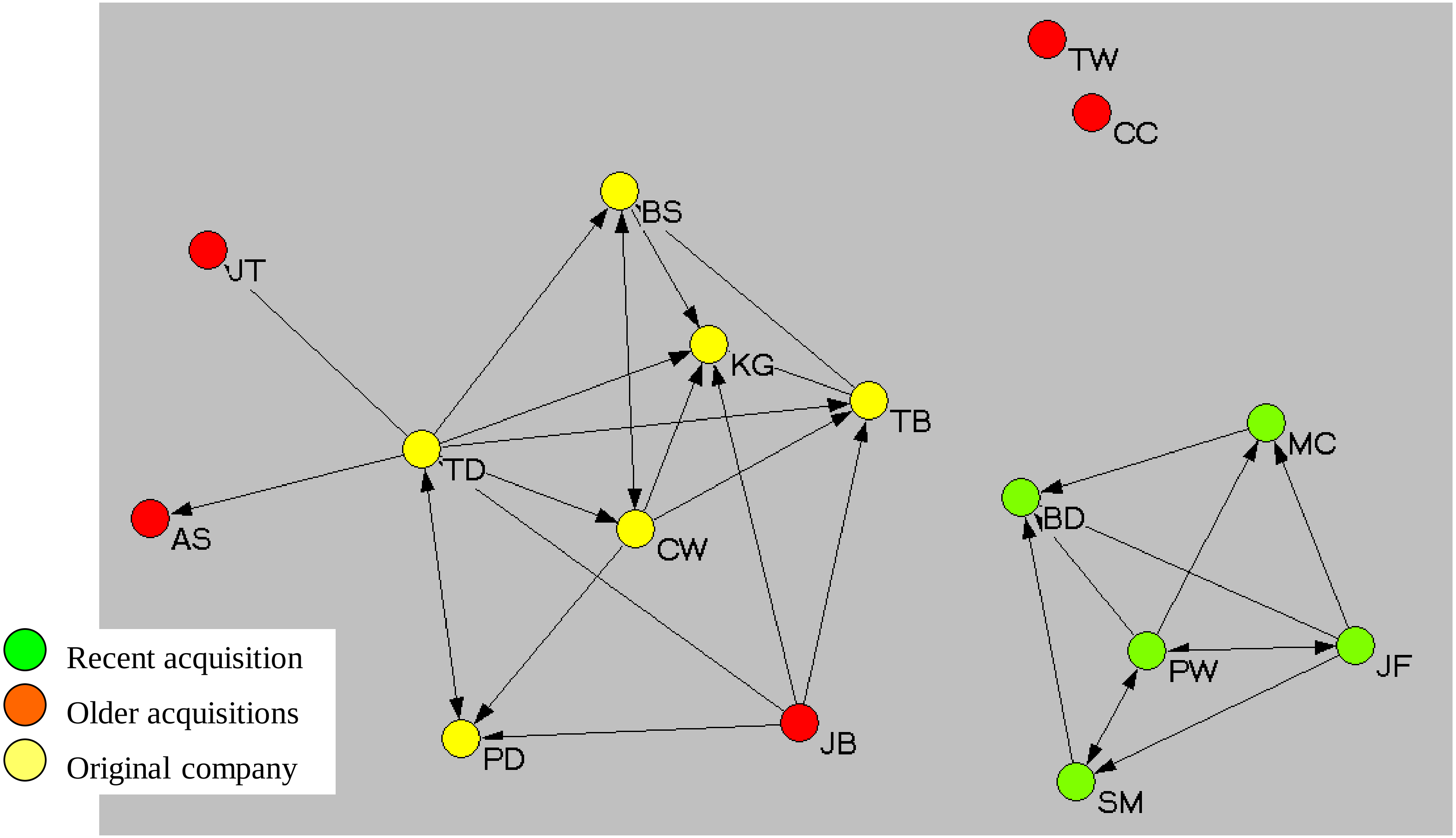
It is relations (types of tie) that define different networks, not components. A network that has two components remains one (disconnected) network.

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

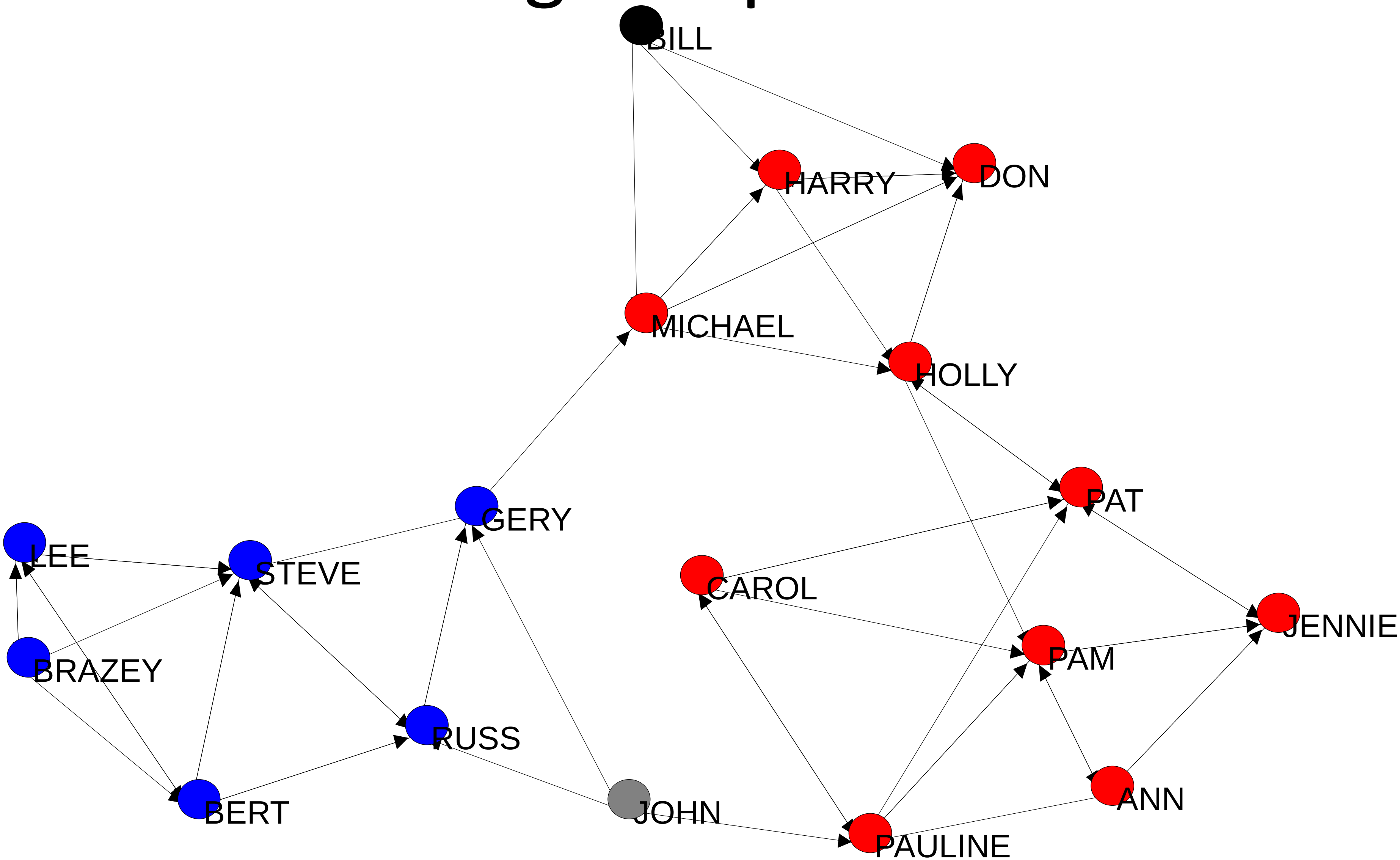
A network with 4 weak components

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

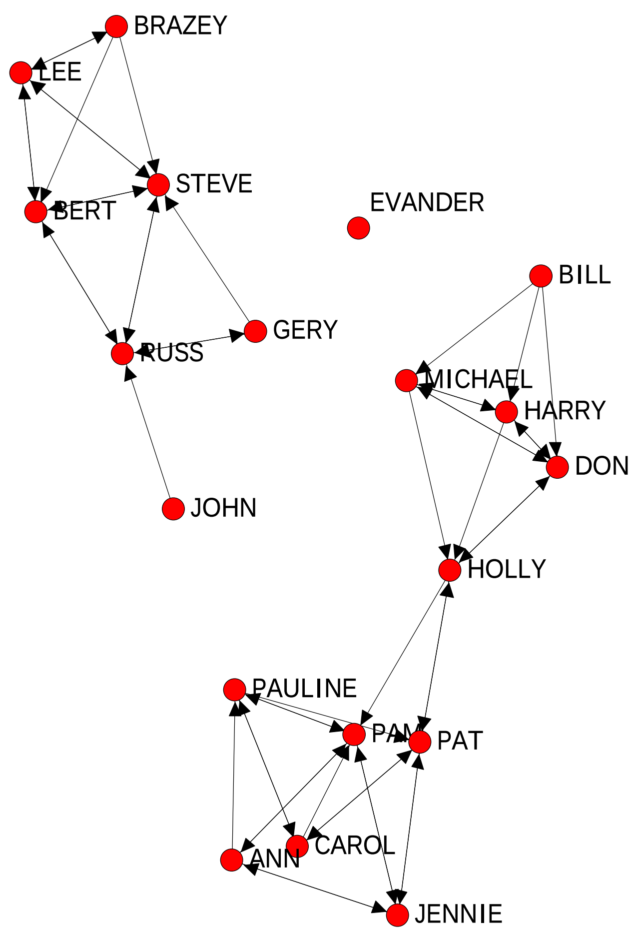


Data drawn from Cross, Borgatti & Parker 2001.

Strong components



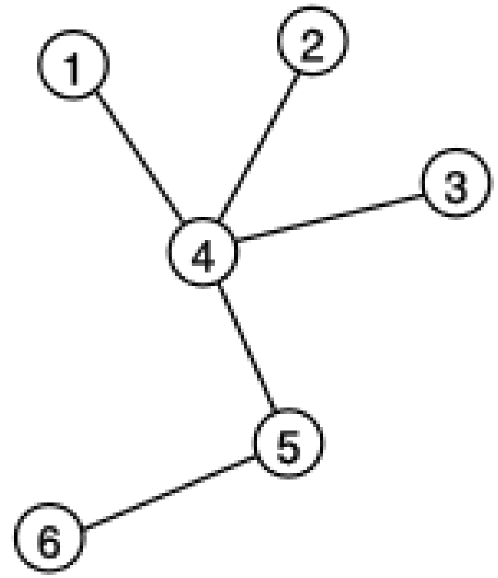
Node-related concepts



- **Degree**
  - The number of ties incident upon a node
  - In a digraph, we have indegree (number of arcs to a node) and outdegree (number of arcs from a node)
- **Pendant**
  - A node connected to a component through only one edge or arc
    - A node with degree 1
    - Example: John
- **Isolate**
  - A node which is a component on its own
    - E.g., Evander

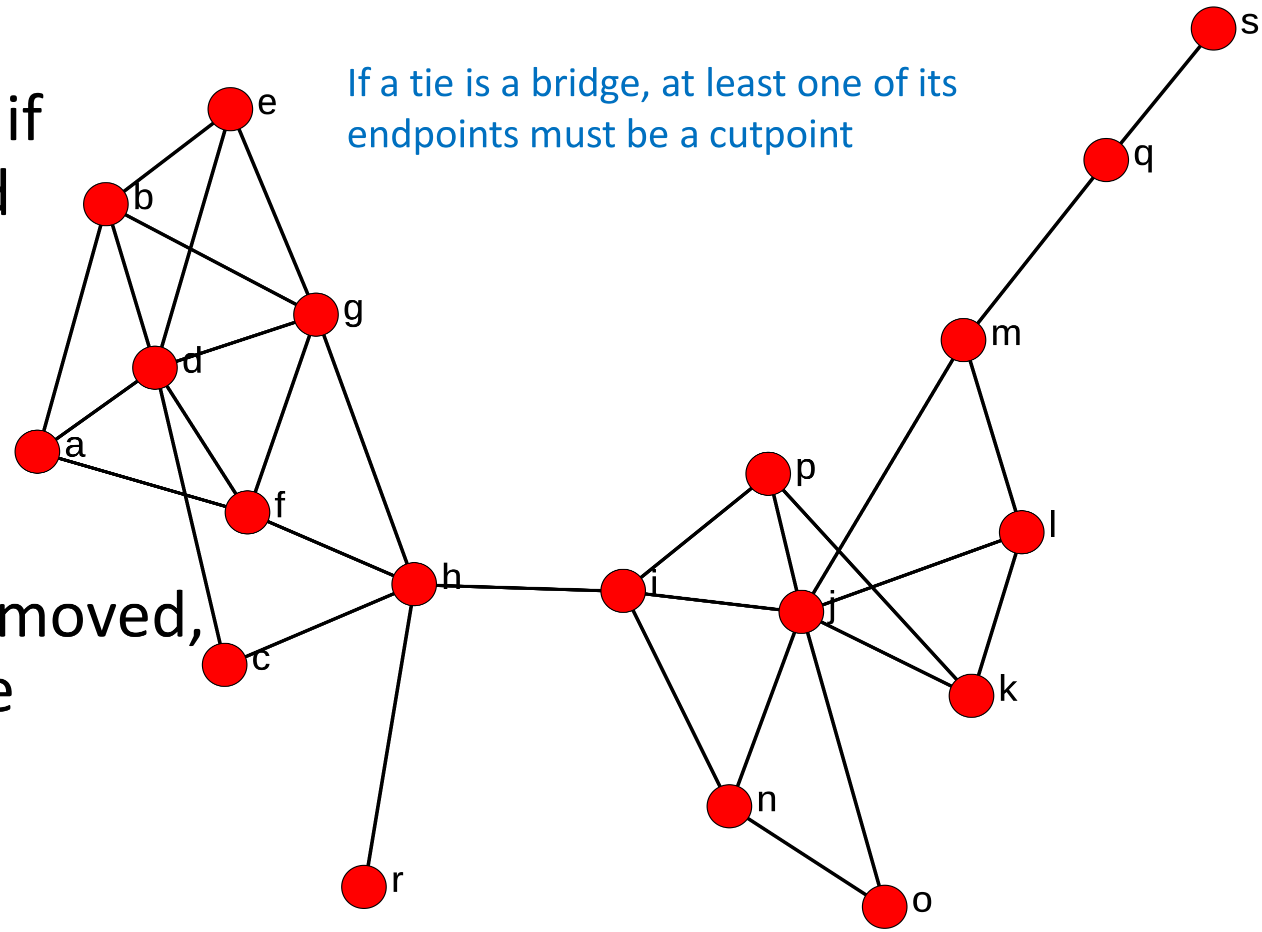
Trees

- A tree is a connected graph that contains no cycles
- In a tree, there is exactly one path from any node to any other



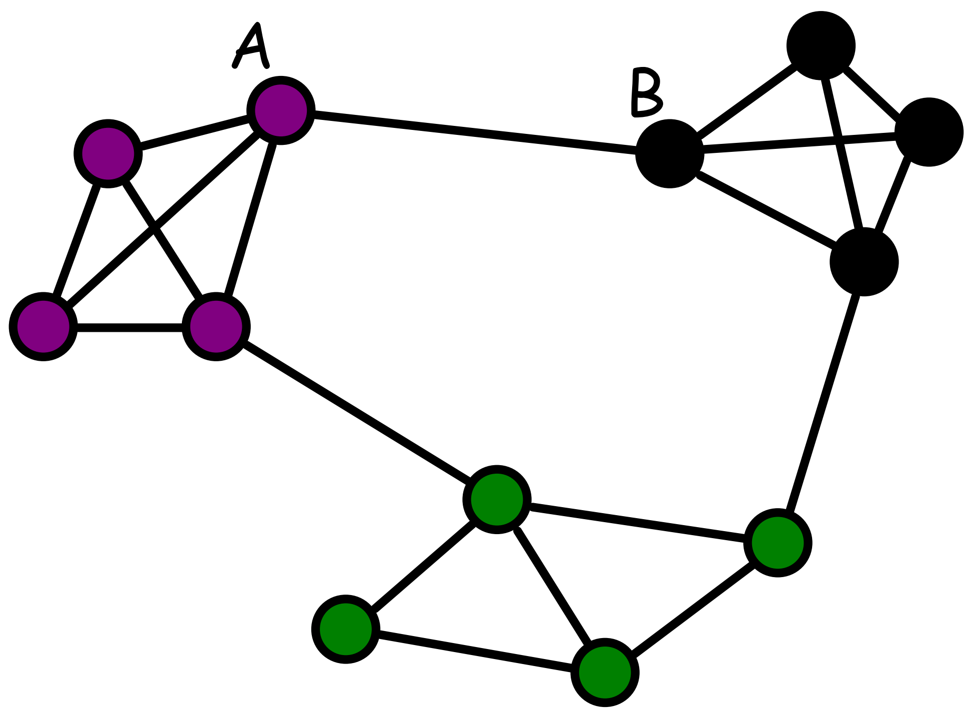
Cutpoints and Bridges

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



Local Bridge of Degree K

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

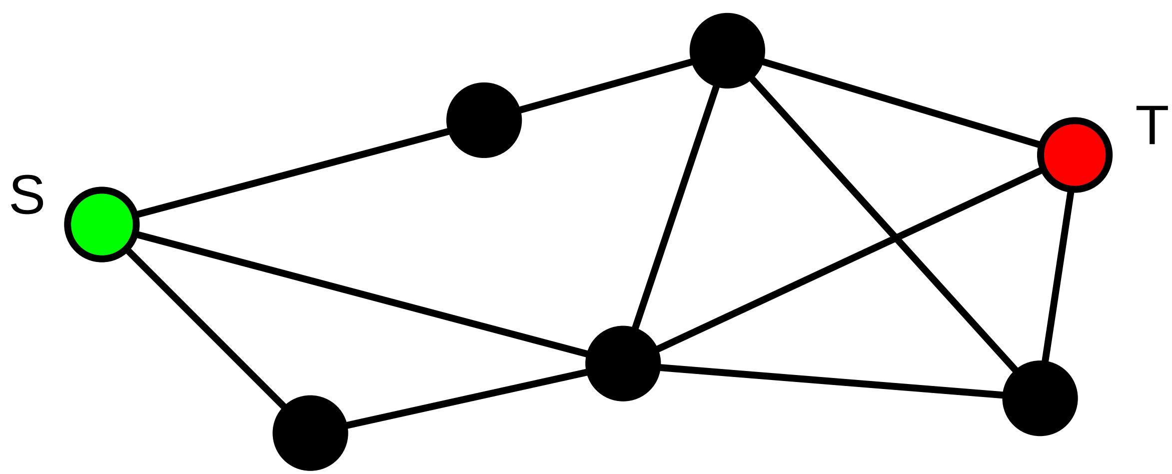


Cutsets

- 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

Independent Paths

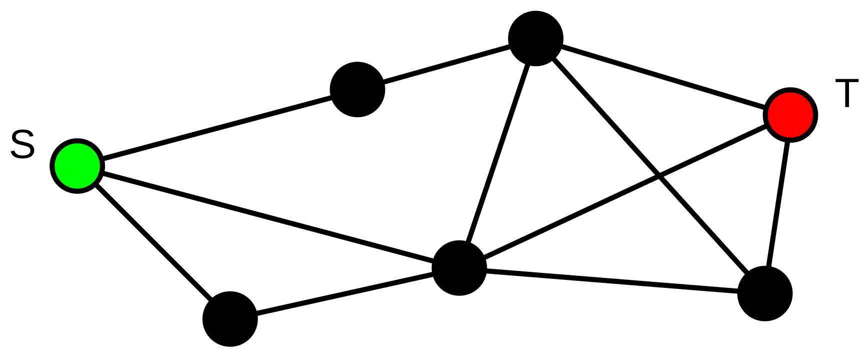
- A set of paths is node-independent if they share no nodes (except beginning and end)
  - They are line-independent if they share no lines



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

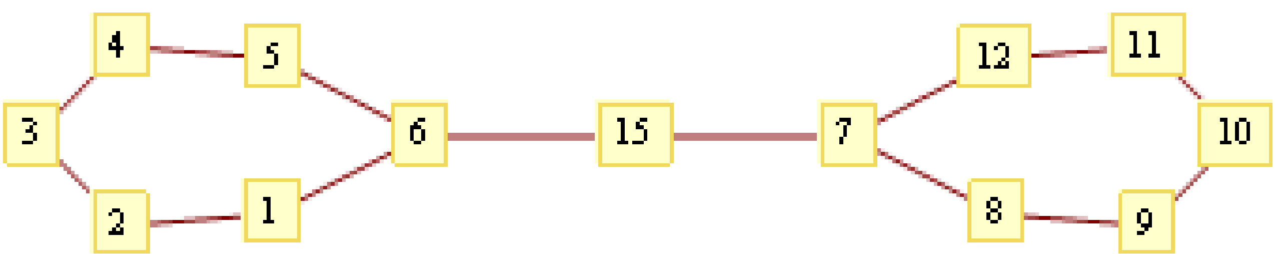
Connectivity

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



Bi-Components (Blocks)

- A bicomponent is a maximal subgraph such that every node can reach every other by at least two node-independent paths
- Bicomponents contain no cutpoints



There are four bicomponents in this graph:
{1 2 3 4 5 6}, {6 15}, {15 7}, and {7 8 9 10 11 12}

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