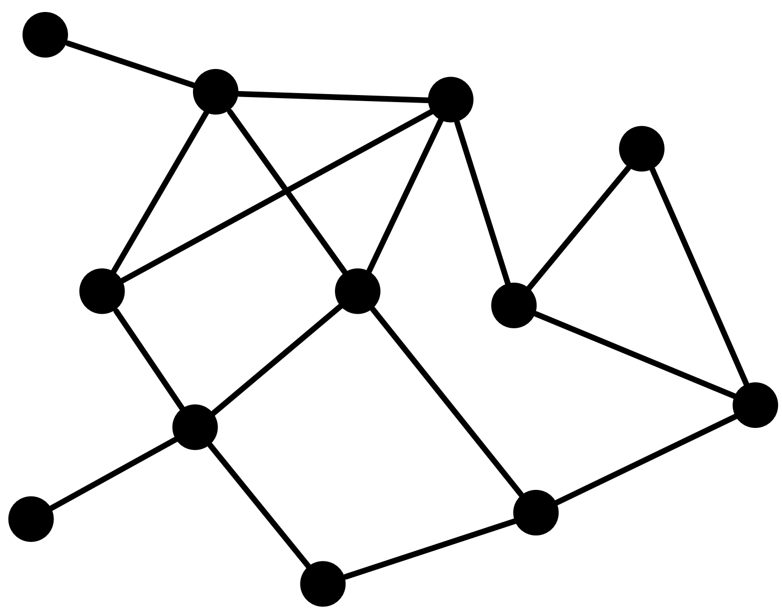
Graph Invariants

Order & Size

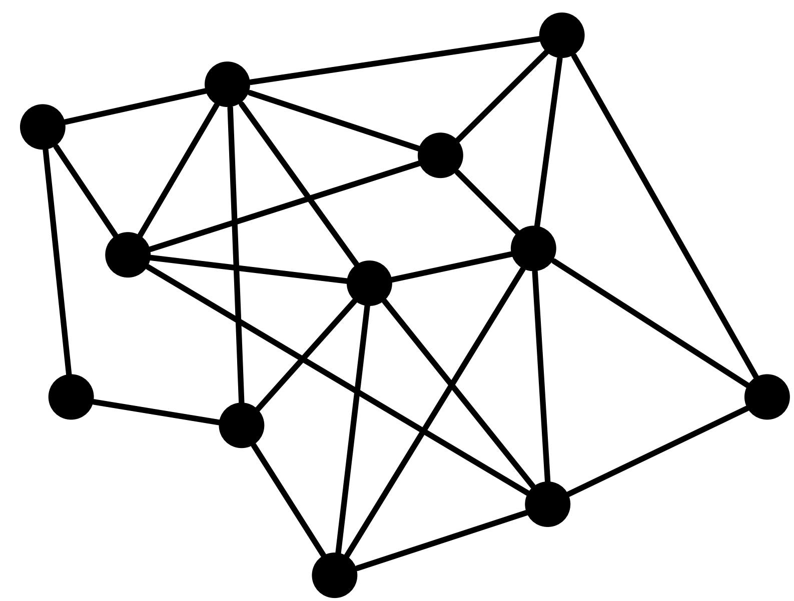
- Order: number of nodes in network
- Size: number of ties in network

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

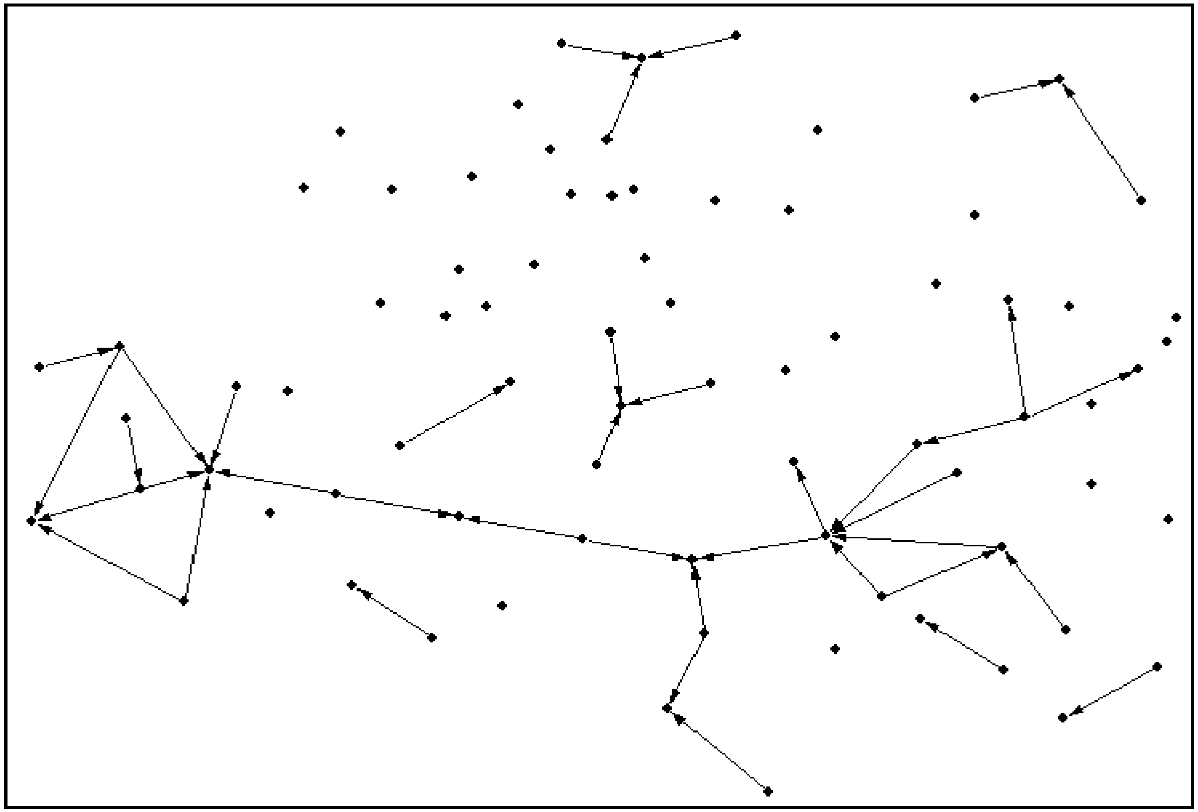
Density

Number of ties divided by number possible

	Reflexive	Non-Reflexive
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

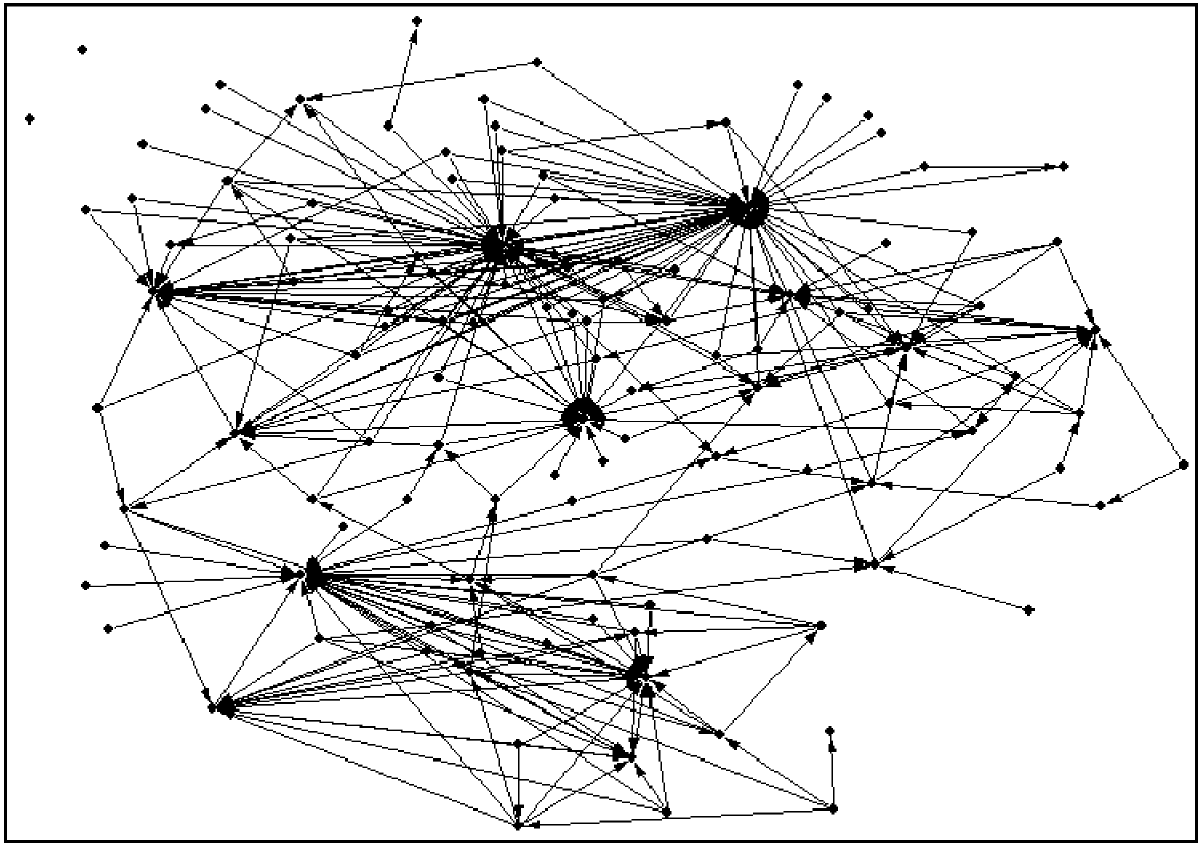
Help With the Rice Harvest



Village 1

Data from Entwistle et al

Help With the Rice Harvest



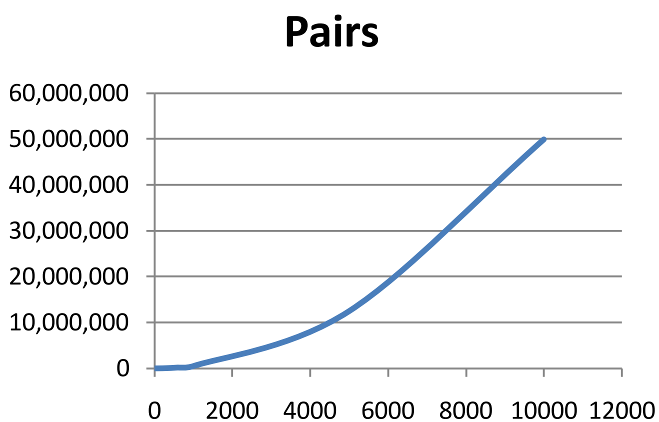
Village 2

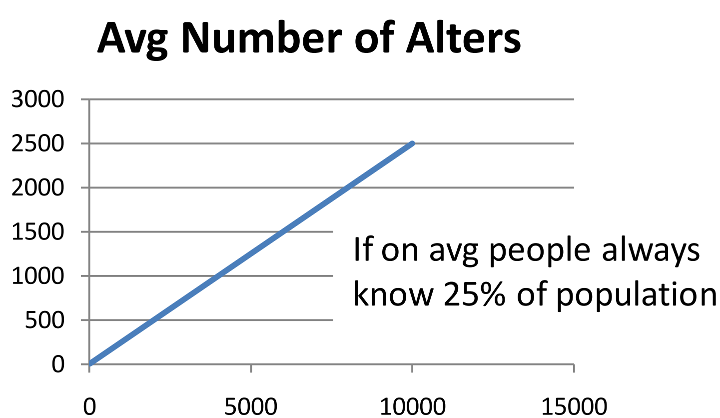
Which village is more likely to survive?

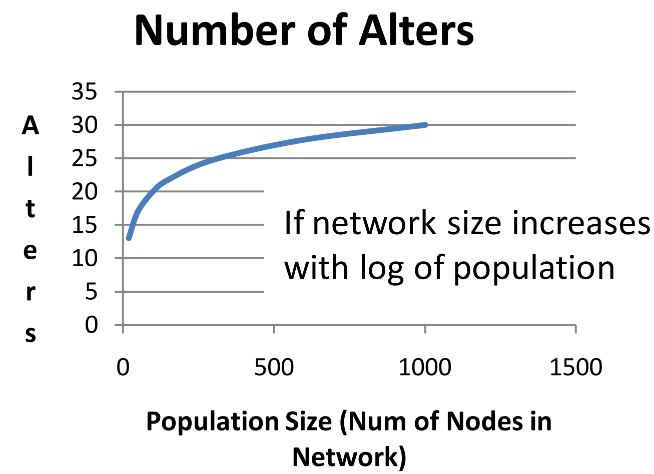
Data from Entwistle et al

Density in large networks

$$\frac{\#\ ties}{\#\ possible} \approx \frac{\#\ ties}{\#\ node\ pairs}$$







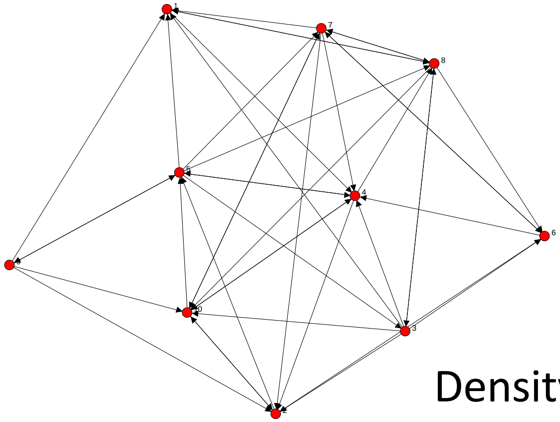
Network Density as a function of Number of Alters

Typical # of Alters

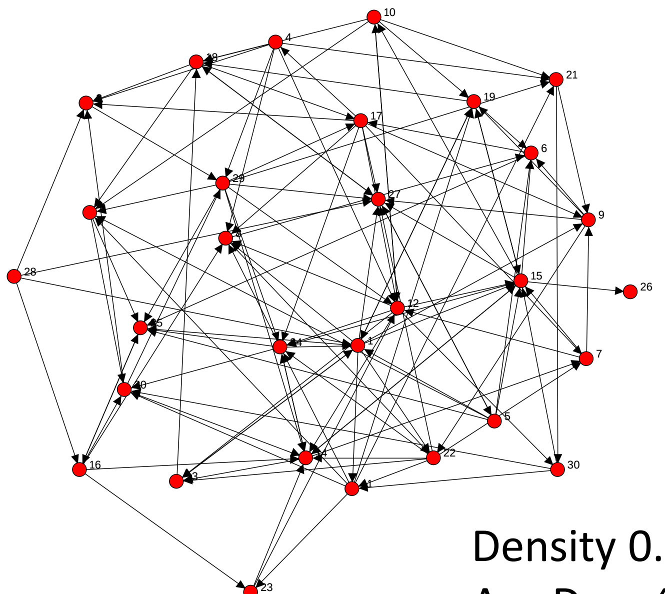
Population	.25P	.25p+10	log(p)
20	0.263	0.789	0.685
50	0.255	0.459	0.347
100	0.253	0.354	0.202
150	0.252	0.319	0.146
300	0.251	0.284	0.083
600	0.250	0.267	0.046
1000	0.250	0.260	0.030
5000	0.250	0.252	0.007
10000	0.250	0.251	0.004

Average Degree

- Average number of links per person
- Is same as density*(n-1), where n is size of network
  - Density is just normalized avg degree – divide by max possible
- Often more intuitive than density

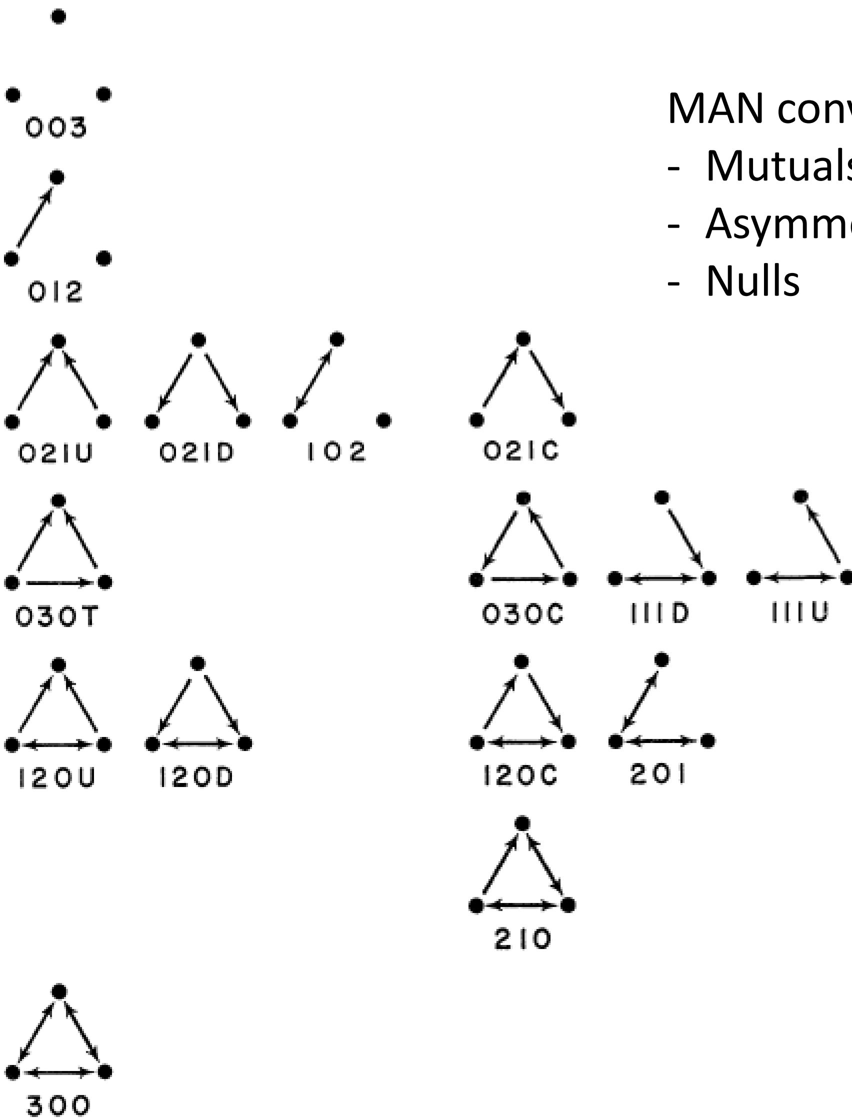


Density 0.47
Avg Deg 4



Density 0.14
Avg Deg 4

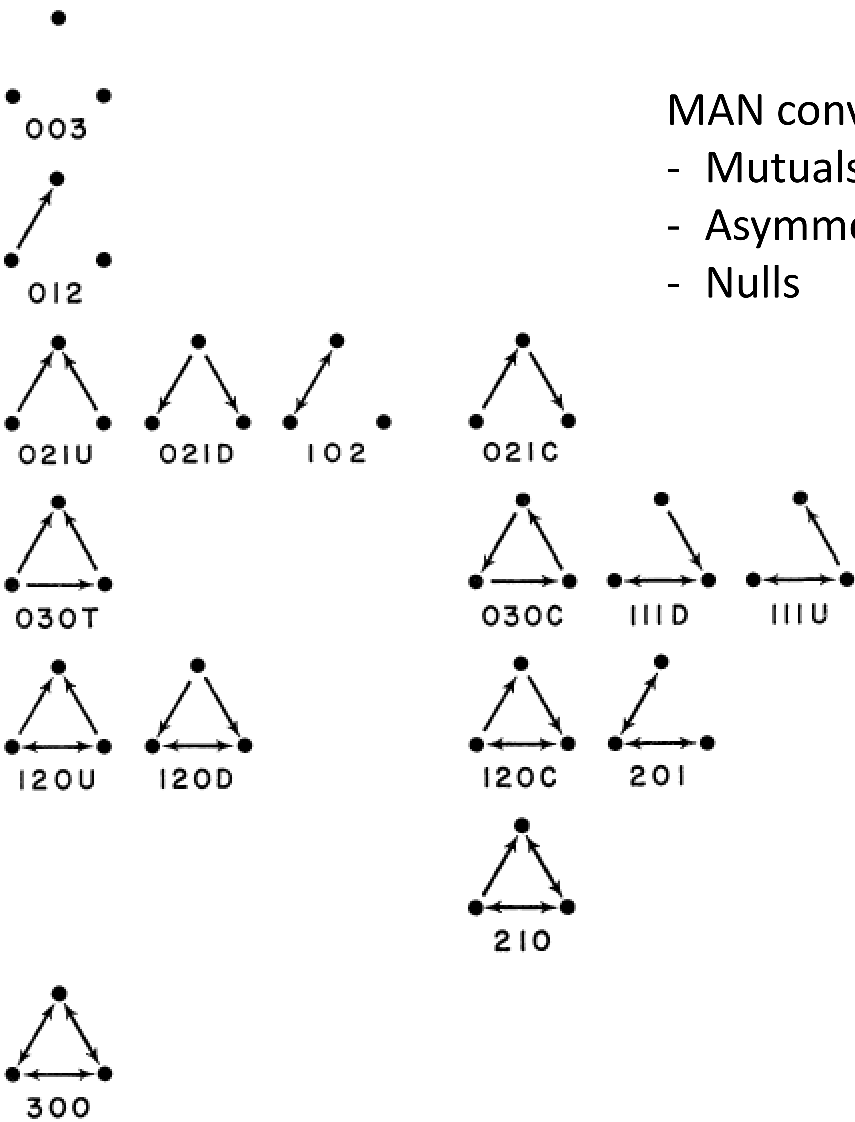
Triad Census



MAN convention:
- Mutuals
- Asymmetrics
- Nulls

	campnet
003	345
012	177
102	223
021D	3
021U	4
021C	9
111D	21
111U	6
030T	0
030C	0
201	10
120D	5
120U	1
120C	0
210	8
300	4

Triad Census

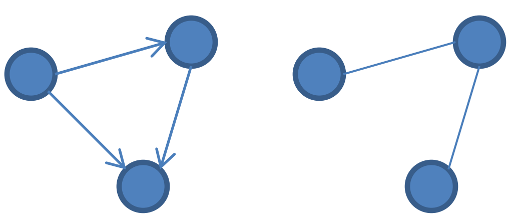


MAN convention:
- Mutuals
- Asymmetrics
- Nulls

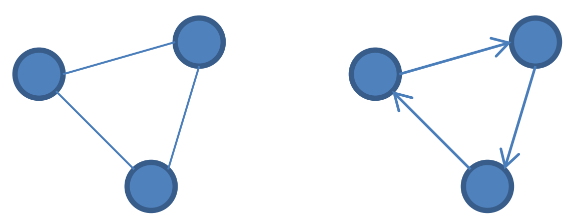
	Undirected Data	
	PADGM	PADGB
	---------	---------
003	324	381
012	0	0
102	195	153
021D	0	0
021U	0	0
021C	0	0
111D	0	0
111U	0	0
030T	0	0
030C	0	0
201	38	21
120D	0	0
120U	0	0
120C	0	0
210	0	0
300	3	5
Sum:	560	560

Transitivity

- If $x \rightarrow y$, $y \rightarrow z$, and $x \rightarrow z$ then x,y,z form a transitive triple
- Transitivity = no. of transitive triples divided by no. of triples in which $x \rightarrow y$ and $y \rightarrow z$
  - Aka, weighted clustering coefficient
- Measure of local density and clumpiness



Non-transitive triples



Transitive triples

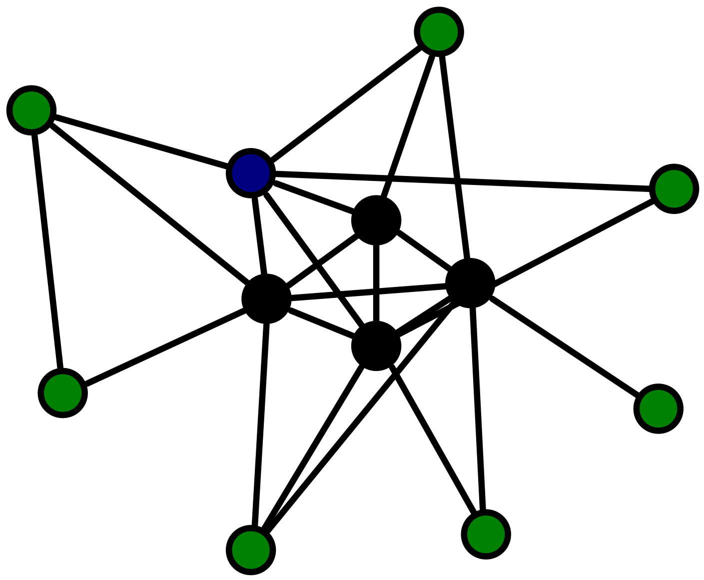
Graph cohesion measures derived from geodesic distances among pairs

	HOLLY	BRAZEY	CAROL	PAM	PAT	JENNIE	PAULINE	ANN	MICHAEL	BILL	LEE	DON	JOHN	HARRY	GERY	STEVE	BERT	RUSS
HOLLY	0	4	2	1	1	2	2	2	1	2	4	1	3	1	2	3	4	3
BRAZEY	4	0	5	5	5	6	4	5	3	4	1	4	3	4	2	1	1	2
CAROL	2	5	0	1	1	2	1	2	3	4	5	3	2	3	3	4	4	3
PAM	1	5	1	0	2	1	1	1	2	3	5	2	2	2	3	4	4	3
PAT	1	5	1	2	0	1	1	2	2	3	5	2	2	2	3	4	4	3
JENNIE	2	6	2	1	1	0	2	1	3	4	6	3	3	3	4	5	5	4
PAULINE	2	4	1	1	1	2	0	1	3	4	4	3	1	3	2	3	3	2
ANN	2	5	2	1	2	1	1	0	3	4	5	3	2	3	3	4	4	3
MICHAEL	1	3	3	2	2	3	3	3	0	1	3	1	2	1	1	2	3	2
BILL	2	4	4	3	3	4	4	4	1	0	4	1	3	1	2	3	4	3
LEE	4	1	5	5	5	6	4	5	3	4	0	4	3	4	2	1	1	2
DON	1	4	3	2	2	3	3	3	1	1	4	0	3	1	2	3	4	3
JOHN	3	3	2	2	2	3	1	2	2	3	3	3	0	3	1	2	2	1
HARRY	1	4	3	2	2	3	3	3	1	1	4	1	3	0	2	3	4	3
GERY	2	2	3	3	3	4	2	3	1	2	2	2	1	2	0	1	2	1
STEVE	3	1	4	4	4	5	3	4	2	3	1	3	2	3	1	0	1	1
BERT	4	1	4	4	4	5	3	4	3	4	1	4	2	4	2	1	0	1
RUSS	3	2	3	3	3	4	2	3	2	3	2	3	1	3	1	1	1	0

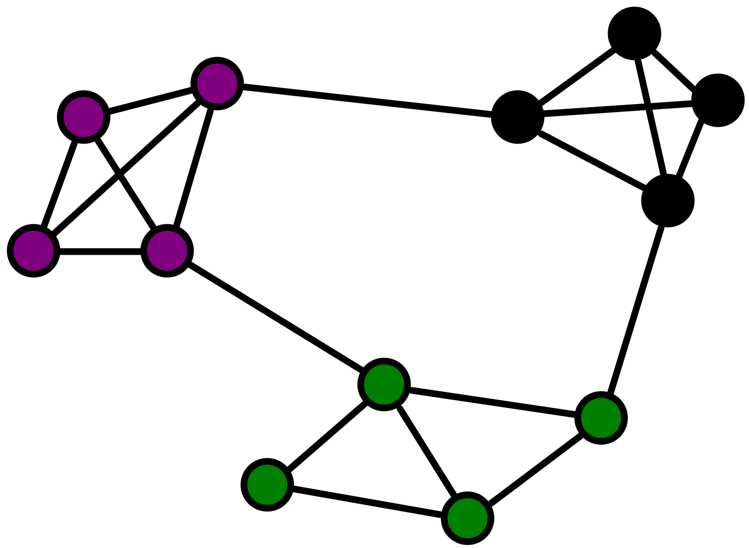
	Geo Dist
Mean	2.66
Std Dev	1.26
Sum	814
Variance	1.60
SSQ	2654
MCSSQ	488.65
Euc Norm	51.52
Minimum	1
Maximum	6
N of Obs	306

Average Distance

- Average geodesic distance between all pairs of nodes



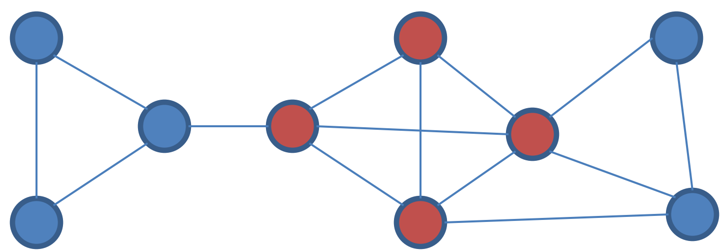
avg. dist. = 1.9



avg. dist. = 2.4

Radius and diameter

- The **eccentricity** ε of a vertex *v* is the greatest distance between *v* and any other vertex.
- The **radius** of a graph is the minimum eccentricity of any vertex.
- The **diameter** of a graph is the maximum eccentricity of any vertex in the graph. That is, it is the greatest distance between any pair of vertices.



Breadth

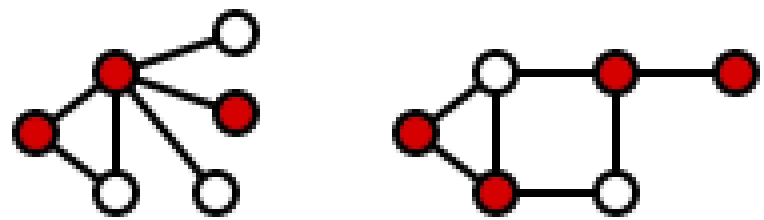
- Distance-Weighted Fragmentation
- Use average of the reciprocal of distance
  - letting $1/\infty = 0$

$$B = 1 - \frac{\sum_{i,j} \frac{1}{d_{ij}}}{n(n-1)}$$

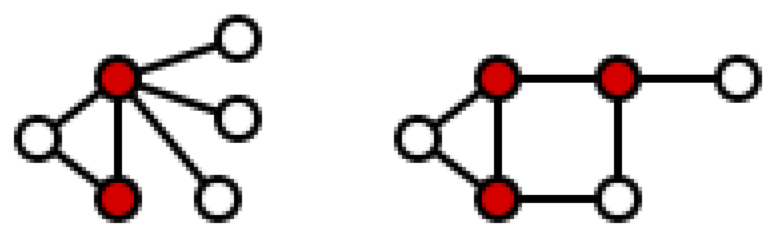
- Bounds
  - lower bound of 0 when every pair is adjacent to every other (entire network is a clique) – minimum breadth
  - upper bound of 1 when graph is all isolates

Vertex Covering Number

- Vertex cover of a graph *G* is a set of vertices *C* such that each edge of *G* is incident to at least one vertex in *C*

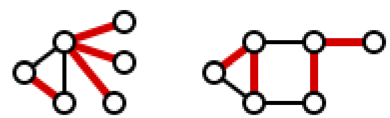


- **minimum vertex cover** is a vertex cover of smallest possible size.
- The **vertex cover number** τ is the size of a minimum vertex cover

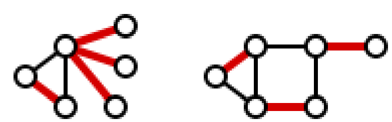


Edge Covering Number

- edge cover of a graph *G* is a set of edges *C* such that each vertex is incident with at least one edge in *C*.

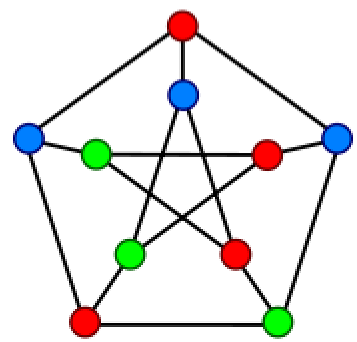


- **minimum edge covering** is an edge covering of smallest possible size
- **edge covering number** ρ(*G*) is the size of a minimum edge covering



Chromatic number

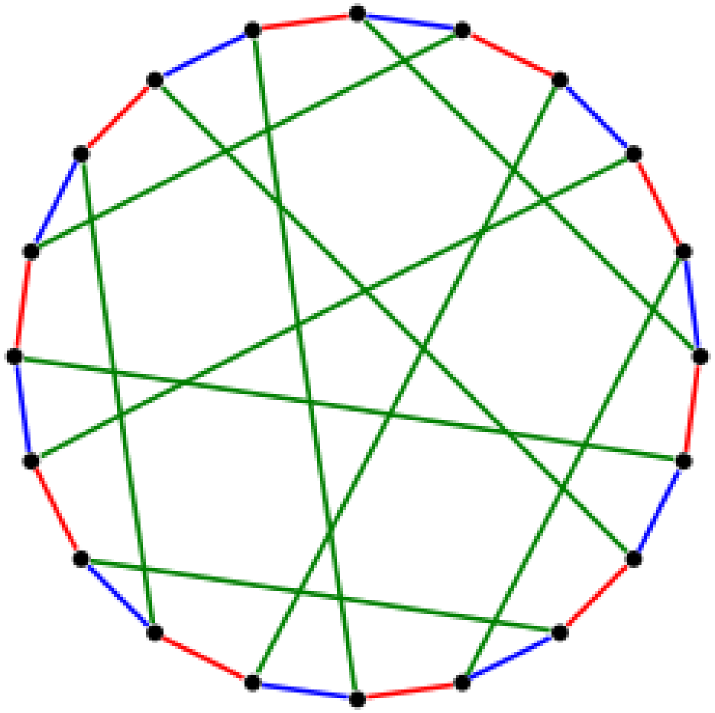
- coloring the vertices of a graph such that no two adjacent vertices share the same color



Smallest coloring has 3 colors

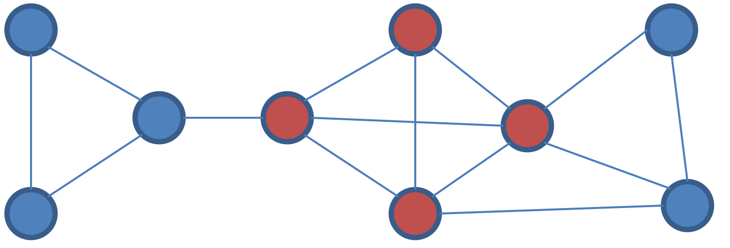
Chromatic index (edge chromatic number)

- coloring the edges of a graph such that no two adjacent edges share the same color



Clique Number

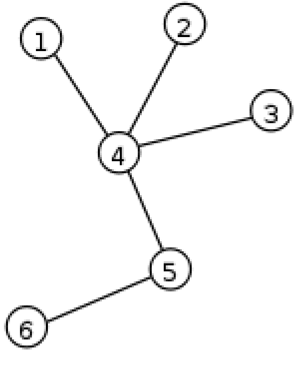
- The **clique number** $\omega(G)$ of a graph G is the number of vertices in the largest clique in G
  - *Clique is a maximal set of nodes that is complete (every member has tie to every other member)*
    - *Maximality means no other node could be added without violating the completeness condition*



Clique number is 4

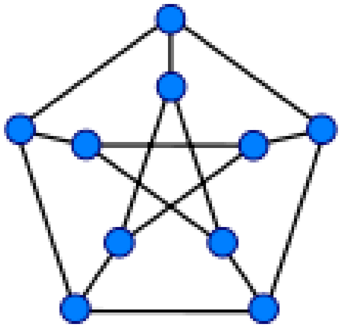
Independent sets

- An **independent set**, or *coclique*, is a set of vertices of which no pair is adjacent
- The **independence number** $\alpha(G)$ of a graph G is the size of a largest independent set of G



Girth

- **girth** of a graph is the length of a shortest cycle contained in the graph
  - If the graph does not contain any cycles, its girth is defined as infinite



Girth is 5

Reciprocity & Symmetry

- Reciprocity is the proportion of outgoing ties that are matched by an incoming tie from the same person

$$r = \frac{a}{a+b+c}$$

i to j / j back to i	1	0
1	a	b
0	c	d

- Symmetry is the proportion of cells in the adjacency matrix such that $x_{ij} = x_{ji}$

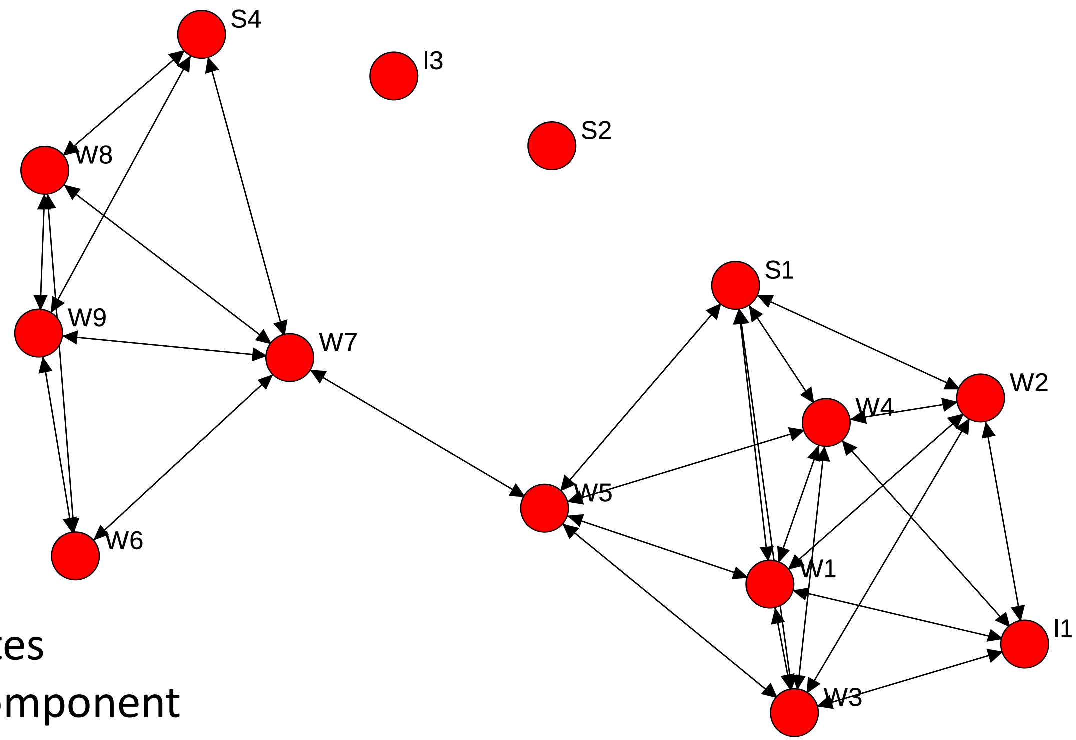
$$s = \frac{a+d}{a+b+c+d}$$

Fragmentation Measures

- Component ratio
- F measure of fragmentation
  - Same as 1-connectivity
- Breadth (Distance-weighted fragmentation) B

Wholeness
(aka Component Ratio (CR))

- No. of components minus 1 divided by number of nodes minus 1



CR is 1 when all nodes are isolates
CR is 0 when all nodes in one component

CR = (3-1)/(14-1) = 0.154

F Measure of Fragmentation

- (undirected formula) Proportion of pairs of nodes that are unreachable from each other

$$F = 1 - \frac{2\sum_{i>j} r_{ij}}{n(n-1)}$$

Krackhardt's connectedness

r_{ij} = 1 if node i can reach node j by a path of any length
r_{ij} = 0 otherwise

- If all nodes reachable from all others (i.e., one component), then F = 0
- If graph is all isolates, then F = 1

Computational Formula for F Measure

- No ties across components, and all reachable within components, hence can express in terms of size of components

$$F = 1 - \frac{\sum_k s_k(s_k - 1)}{n(n-1)}$$

S_k = size of k^{th} component

Heterogeneity/Concentration

- Sum of squared proportion of nodes falling in each component, where s_k gives size of kth component:

$$H = 1 - \sum_k \left(\frac{s_k}{n} \right)^2$$

- Maximum value is 1-1/n
- Can be normalized by dividing by 1-1/n. If we do, we obtain the F measure

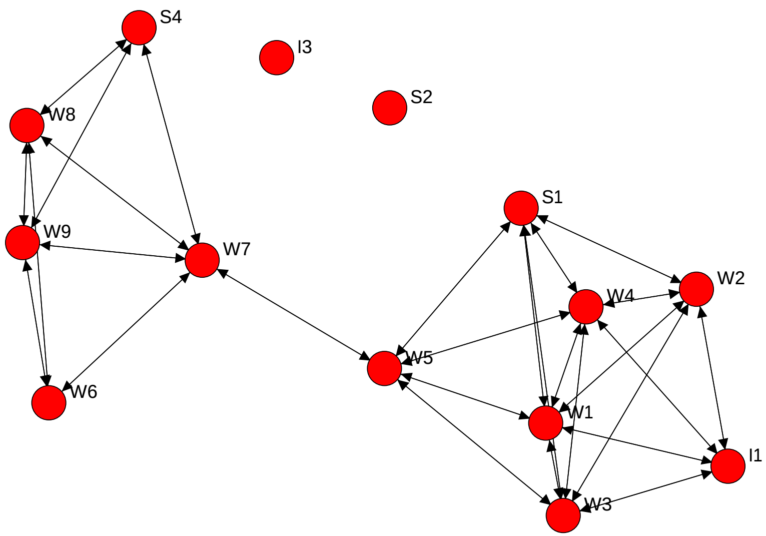
$$F = 1 - \frac{\sum_k s_k (s_k - 1)}{n(n-1)}$$

Heterogeneity Example

Games Data

Comp	Size	Prop	Prop^2
1	1	0.0714	0.0051
2	1	0.0714	0.0051
3	12	0.8571	0.7347
	14	1.0000	0.7449

Heterogeneity = 0.255

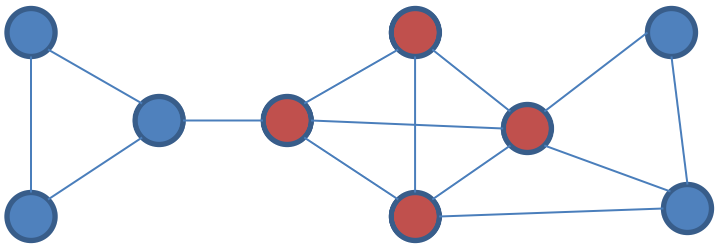


ATTRIBUTE SENSITIVE PROPERTIES

E-I homophily index

- Krackhardt & Stern
  - Number of External ties minus number of Internal ties as a proportion of all ties

$$\frac{E-I}{E+I}$$



$$\frac{4-12}{4+12} = -\frac{8}{16} = -0.5$$

Set-up for homophily measures

- Given
  - A social relation R
  - A categorical attribute vector **a**
- Construct
  - Similarity relation S in which $s_{ij} = 1$ if $a_i = a_j$, and $s_{ij} = 0$ otherwise

R

	f	f	f	f	f	f	f	f	m	m	m	m	m	m	m	m	m	m
f	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0
f	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0
f	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0
f	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
f	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
f	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0
f	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
f	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0
m	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
m	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0
m	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
m	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
m	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1
m	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
m	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1
m	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1
m	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1
m	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0

a

1
1
1
1
1
1
1
1
2
2
2
2
2
2
2
2
2
2

S

	f	f	f	f	f	f	f	f	m	m	m	m	m	m	m	m	m	m
f	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
f	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
f	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
f	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
f	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
f	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
f	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
f	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
m	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
m	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
m	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
m	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
m	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
m	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
m	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
m	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
m	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
m	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1

Homophily set-up - cont

- Construct “relational contingency table”

		S = 1	S = 0
R	1	a	b
	0	c	d

R = the data – the social relation
S = similarity – is 1 if same attrib value

- Campnet dataset:

		S = 1	S = 0
R	1	45	9
	0	101	151

E-I index

- Krackhardt & Stern
  - Number external ties minus number of internal ties as a proportion of all ties

		S: 1	S: 0
R	1	a	b
R	0	c	d

$$EI = \frac{b-a}{b+a}$$

Negative values indicated greater homophily

- Campnet

		S: 1	S: 0
R	1	45	9
R	0	101	151

EI = -0.667

Pct homophilous matches (H%)

- H% = a/(a+b)

S

		1	0
R	1	a	b
	0	c	d

- Campnet dataset

S

		1	0
R	1	45	9
	0	101	151

H% = 0.83

Point bi-serial correlation (pbsc) approach

- Take into account non-choices as well:

R \ S	1	0
1	a	b
0	c	d

$$r(R,S) = \frac{ad - bc}{\sqrt{(a+c)(b+d)(a+b)(c+d)}}$$

- Campnet dataset:

R \ S	1	0
1	45	9
0	101	151

r(R,S) = 0.33
H% = 0.83

R \ S	1	0
1	45	9
0	90	18

r(R,S) = 0.00
H% = 0.83

Matrix Correlation

- The pbsc measure is the same as a QAP correlation of the two dyadic variables R and S

r(R,S) = 0.33

	f	f	f	f	f	f	f	f	m	m	m	m	m	m	m	m	m	m
f	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0
f	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0
f	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0
f	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
f	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
f	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0
f	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
f	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0
m	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
m	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0
m	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
m	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
m	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1
m	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
m	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1
m	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1
m	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1
m	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0

R

	f	f	f	f	f	f	f	f	m	m	m	m	m	m	m	m	m	m
f	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
f	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
f	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
f	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
f	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
f	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
f	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
f	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
m	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
m	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
m	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
m	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
m	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
m	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
m	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
m	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
m	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
m	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1

S

Density Tables

- Concept
  - Number of ties within and between groups
  - Called density when expressed as a function of the number possible

	BHS	CCG	DCL	ES	HEW	IS	MS	SRG	STAT	TAS	N
BHS	1613	356	239	1601	717	74	862	231	576	239	178
CCG	356	42	272	329	206	59	228	32	58	231	15
DCL	239	272	2117	521	616	844	1005	61	79	1541	177
ES	1601	329	521	968	445	117	712	119	326	416	146
HEW	717	206	616	445	374	64	380	67	161	245	89
IS	74	59	844	117	64	188	159	16	15	397	52
MS	862	228	1005	712	380	159	666	84	263	845	134
SRG	231	32	61	119	67	16	84	15	49	32	20
STAT	576	58	79	326	161	15	263	49	62	88	19
TAS	239	231	1541	416	245	397	845	32	88	1397	130
N	178	15	177	146	89	52	134	20	19	130	960

Tie Frequencies

	BHS	CCG	DCL	ES	HEW	IS	MS	SRG	STAT	TAS
BHS	0.10	0.13	0.01	0.06	0.05	0.01	0.04	0.06	0.17	0.01
CCG	0.13	0.40	0.10	0.15	0.15	0.08	0.11	0.11	0.20	0.12
DCL	0.01	0.10	0.14	0.02	0.04	0.09	0.04	0.02	0.02	0.07
ES	0.06	0.15	0.02	0.09	0.03	0.02	0.04	0.04	0.12	0.02
HEW	0.05	0.15	0.04	0.03	0.10	0.01	0.03	0.04	0.10	0.02
IS	0.01	0.08	0.09	0.02	0.01	0.14	0.02	0.02	0.02	0.06
MS	0.04	0.11	0.04	0.04	0.03	0.02	0.07	0.03	0.10	0.05
SRG	0.06	0.11	0.02	0.04	0.04	0.02	0.03	0.08	0.13	0.01
STAT	0.17	0.20	0.02	0.12	0.10	0.02	0.10	0.13	0.36	0.04
TAS	0.01	0.12	0.07	0.02	0.02	0.06	0.05	0.01	0.04	0.17

Densities