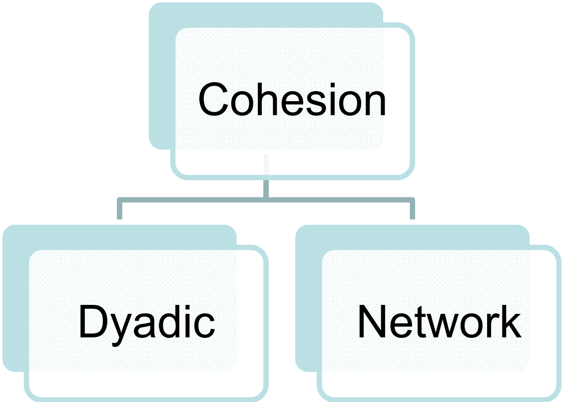
Graph Cohesion

Steve Borgatti
MGT 780

Relational vs Group

- Relational or dyadic cohesion refers to pairwise social closeness
- Network or graph cohesion refers to the cohesion of an entire group
  - Group could be whole population (the network)
  - Or a subset
    - Marketing dept
    - Cluster/clique



Graphs and Subgraphs

- Subgraphs
  - Given graph G(V,E), a graph S is called a subgraph of G if $V(S) \subseteq V(G)$ and $E(S) \subseteq E(G)$.
    - i.e. all nodes in S are also in G and all edges in S are also in G
    - Note that G is a subgraph of itself
- Induced subgraph

Measures of Group Cohesion

- Density & variations
  - Average tie strength
  - Average degree
- Number of components
- Fragmentation / connectedness
- Size of largest clique
- Cliques per node
- Connectivity / independent paths / max flow

- Average Distance (cpl)
- Diameter
- Breadth

Measures of Graph Cohesion

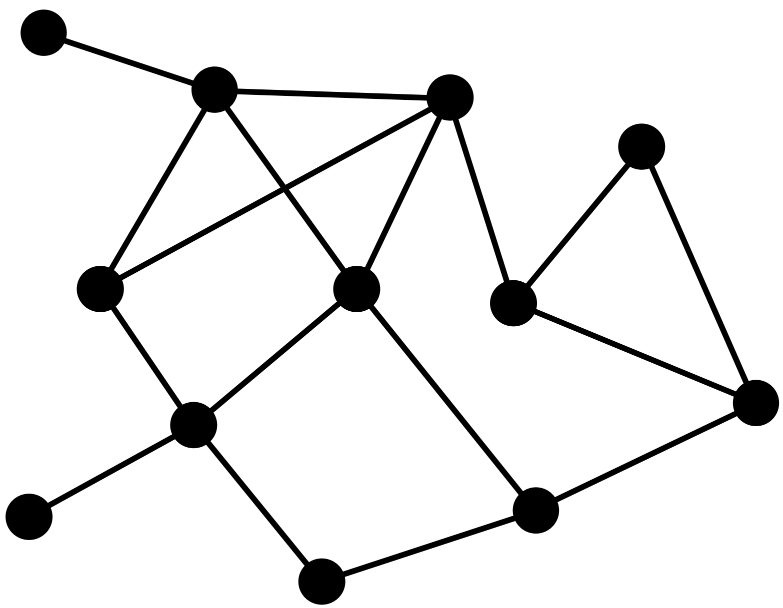
	Frequency-based	Length-based

Network only	• Density & variations
	  • Average tie strength
	  • Average degree
	• Component ratio
	• Fragmentation / connectedness
	• Connectivity / independent paths / max flow
	• Size of largest clique
	• Cliques per node
		• Average Distance (cpl)
		• Diameter
		• Breadth

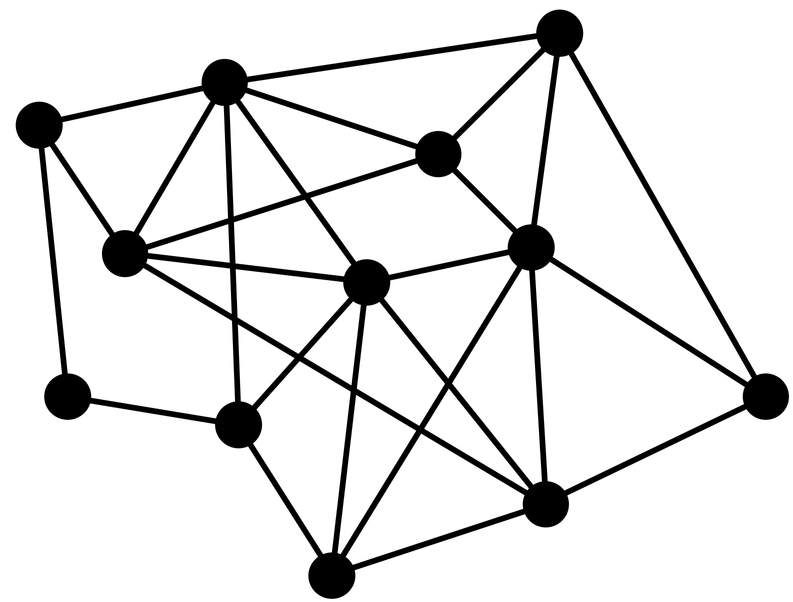
Network + Node Attribs	• Anova density models • EI index	• Valued density tables

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

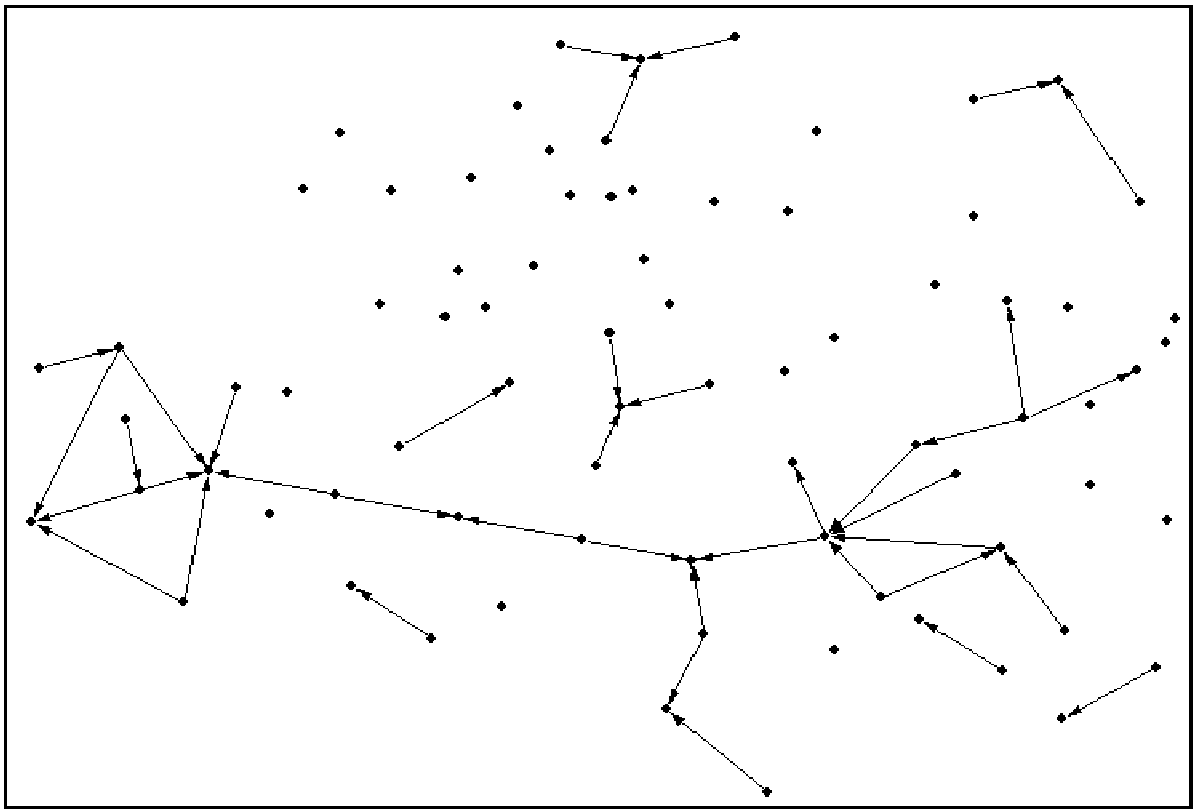
Density as aggregated dyadic cohesion

	HOLLY	BILL	DON	HARRY	MICHAEL	PAM	JENNIE	ANN	PAULINE	PAT	CAROL	LEE	JOHN	BRAZEY	GERY	STEVE	BERT	RUSS	Avg
HOLLY		0	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0.294
BILL	0		1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0.176
DON	1	1		1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0.235
HARRY	1	1	1		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0.235
MICHAEL	1	1	1	1		0	0	0	0	0	0	0	0	0	1	0	0	0	0.294
PAM	1	0	0	0	0		1	1	1	0	1	0	0	0	0	0	0	0	0.294
JENNIE	0	0	0	0	0	1		1	0	1	0	0	0	0	0	0	0	0	0.176
ANN	0	0	0	0	0	1	1		1	0	0	0	0	0	0	0	0	0	0.176
PAULINE	0	0	0	0	0	1	0	1		1	1	0	1	0	0	0	0	0	0.294
PAT	1	0	0	0	0	0	1	0	1		1	0	0	0	0	0	0	0	0.235
CAROL	0	0	0	0	0	1	0	0	1	1		0	0	0	0	0	0	0	0.176
LEE	0	0	0	0	0	0	0	0	0	0	0		0	1	0	1	1	0	0.176
JOHN	0	0	0	0	0	0	0	0	1	0	0	0		0	1	0	0	1	0.176
BRAZEY	0	0	0	0	0	0	0	0	0	0	0	1	0		0	1	1	0	0.176
GERY	0	0	0	0	1	0	0	0	0	0	0	0	1	0		1	0	1	0.235
STEVE	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1		1	1	0.294
BERT	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1		1	0.235
RUSS	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1		0.235
Avg	0.29	0.18	0.24	0.24	0.29	0.29	0.18	0.18	0.29	0.24	0.18	0.18	0.18	0.18	0.24	0.29	0.24	0.24	0.229

Four Views

- Number of ones divided by number of valid cells
- Sum of all values divided by number of valid cells
- Avg of all values
- Avg of the row averages

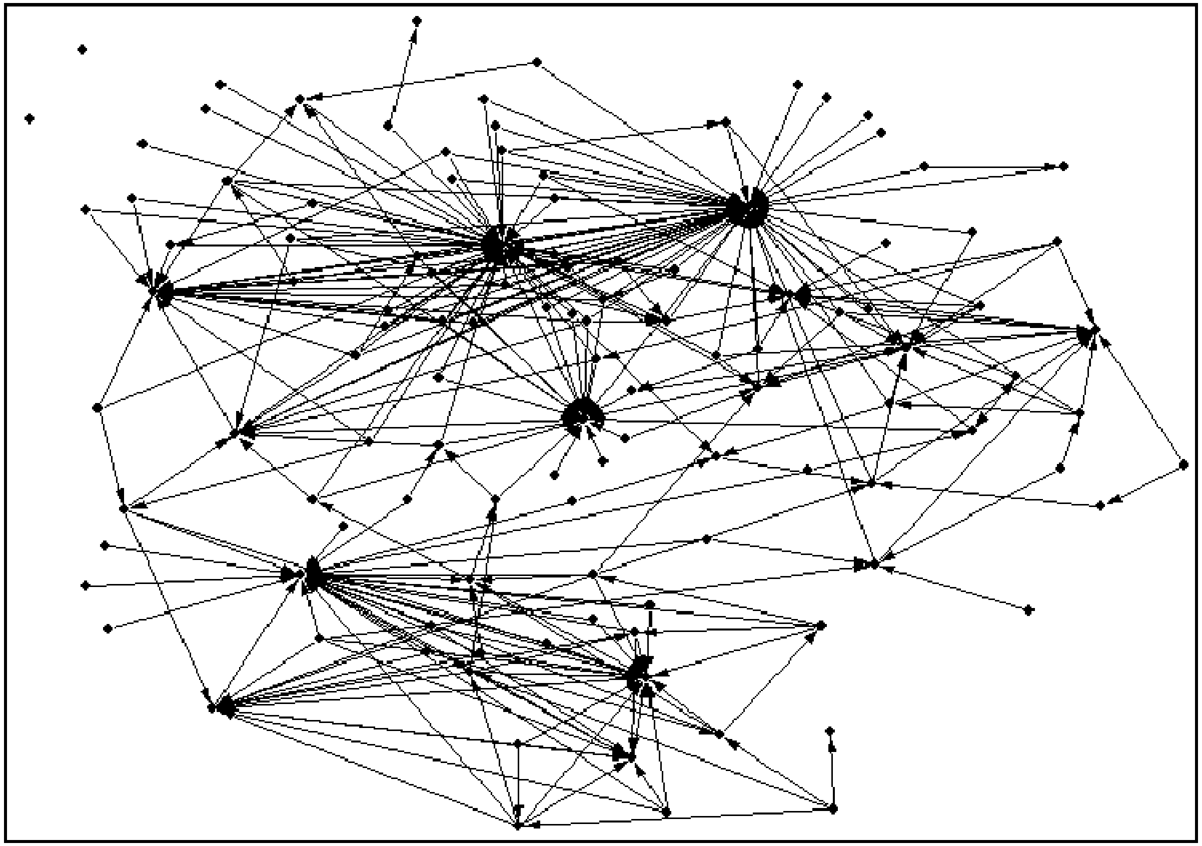
Help With the Rice Harvest



Village 1

Data from Entwistle et al

Help With the Rice Harvest



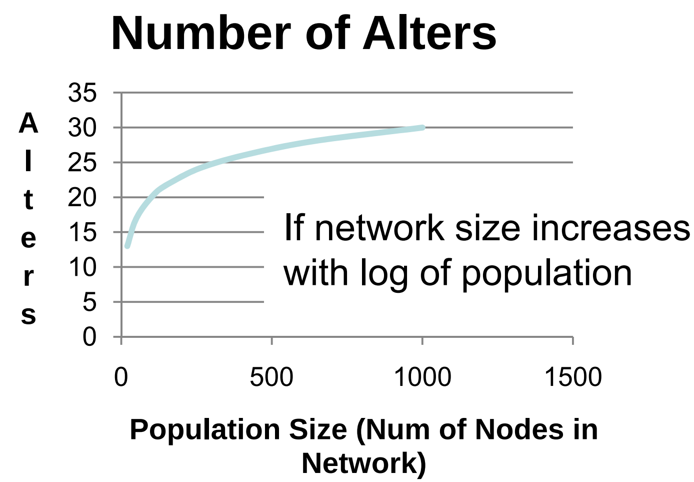
Village 2

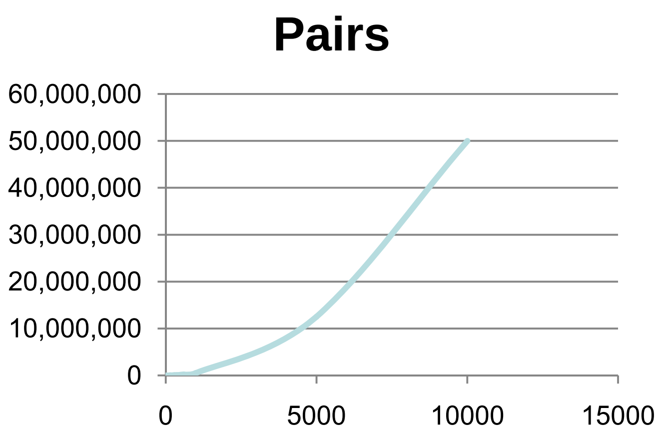
Which village is more likely to survive?

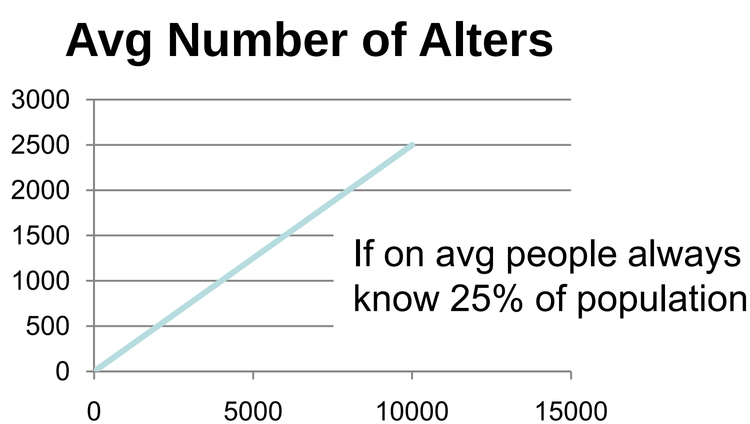
Data from Entwistle et al

Density in large networks

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





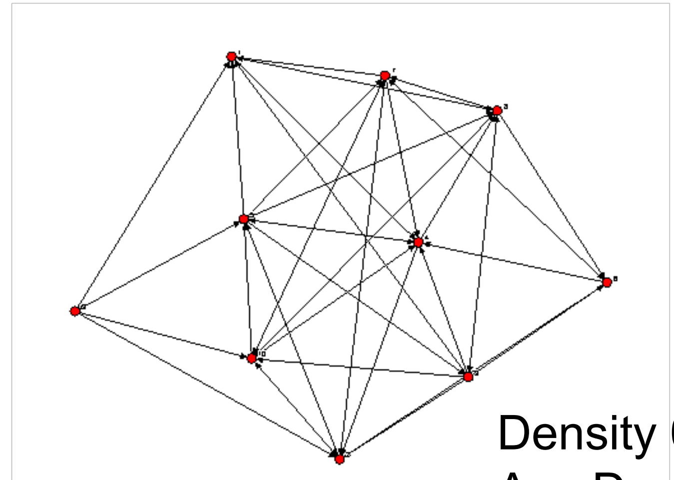


Network Density as a function of Number of Alters

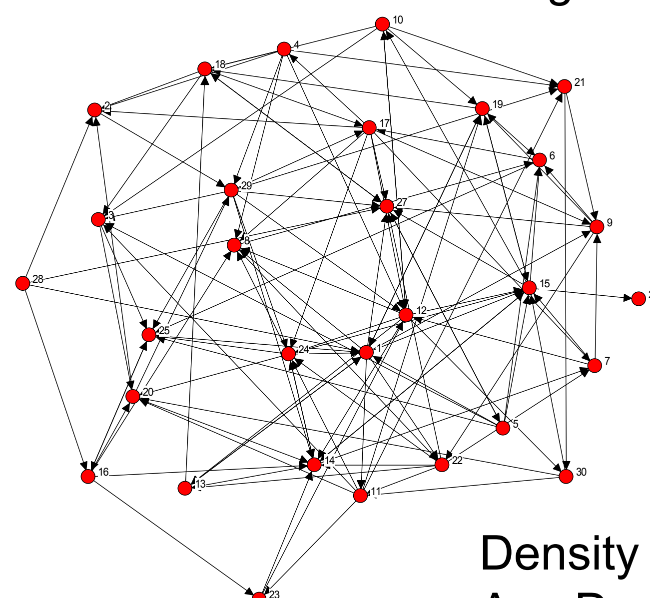
	Typical # of Alters		
Population	.25P	.25p+10	log(p)
20	0.263	0.789	0.685
50	0.255	0.459	0.347
100	0.253	0.354	0.202
150	0.252	0.319	0.146
300	0.251	0.284	0.083
600	0.250	0.267	0.046
1000	0.250	0.260	0.030
5000	0.250	0.252	0.007
10000	0.250	0.251	0.004

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



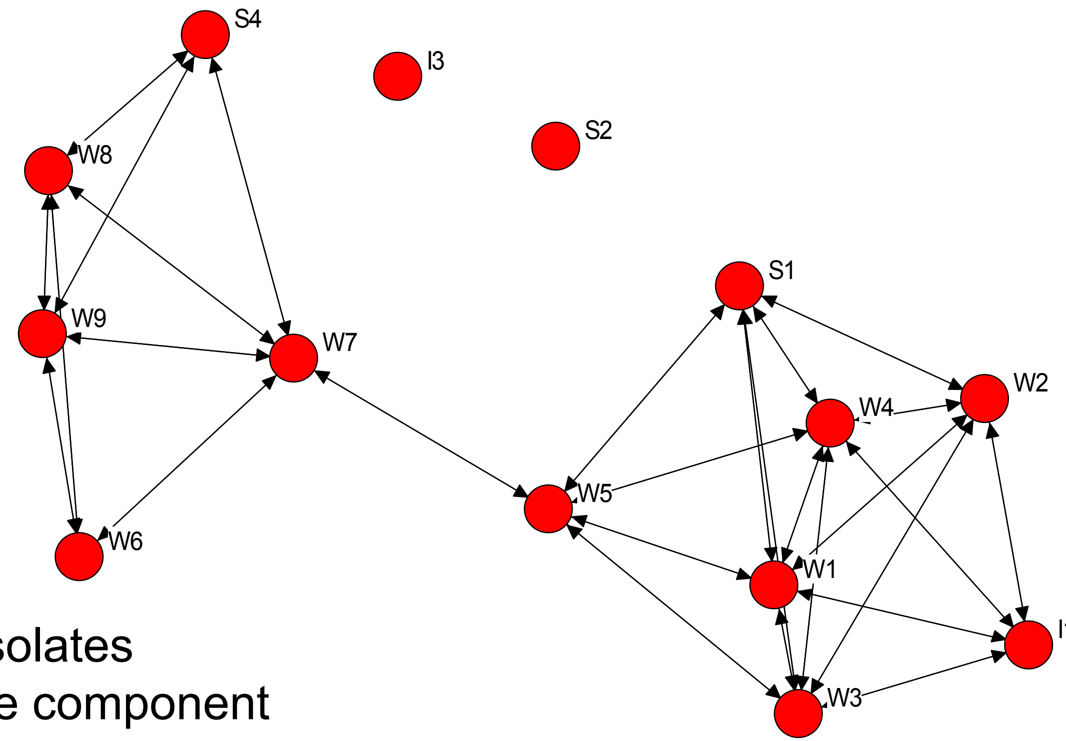
Density 0.14
Avg Deg 4

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

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

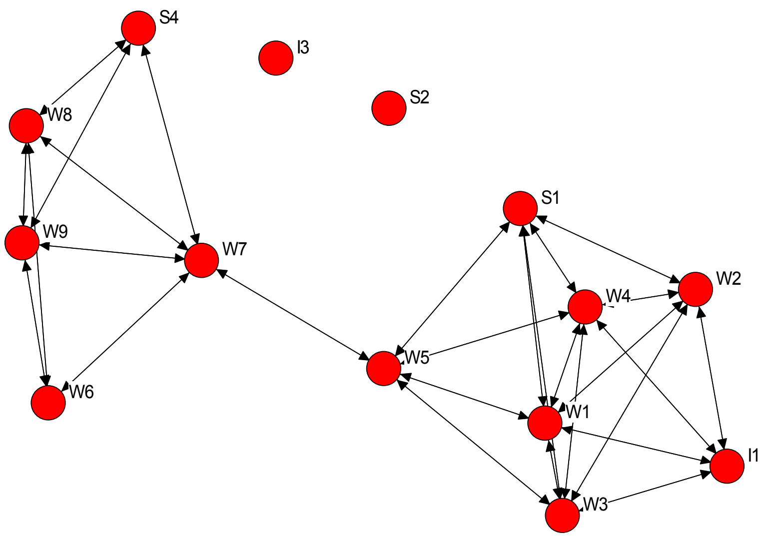
S_k = size of k^{th} component

Computational Example

Games Data

Comp	Size	Sk(Sk-1)
1	1	0
2	1	0
3	12	132
	14	132
		0.2747

0.2747 = 132/(14*13) = F



Heterogeneity/Concentration

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

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

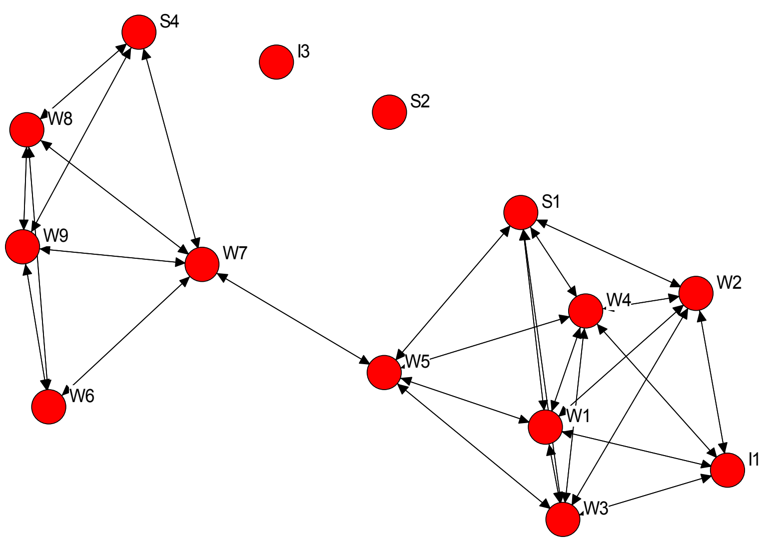
- Maximum value is 1-1/n
- Can be normalized by dividing by 1-1/n. If we do, we obtain the F measure

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

Heterogeneity Example

Games Data

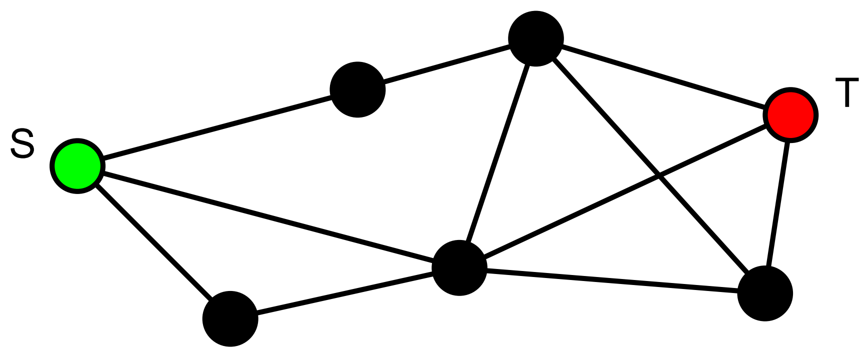
Comp	Size	Prop	Prop^2
1	1	0.0714	0.0051
2	1	0.0714	0.0051
3	12	0.8571	0.7347
	14	1.0000	0.7449



Heterogeneity = 0.255

Connectivity

- Line connectivity λ is the minimum number of lines that must be removed to discon-nect network
- Node connectivity κ is minimum number of nodes that must be removed to discon-nect network

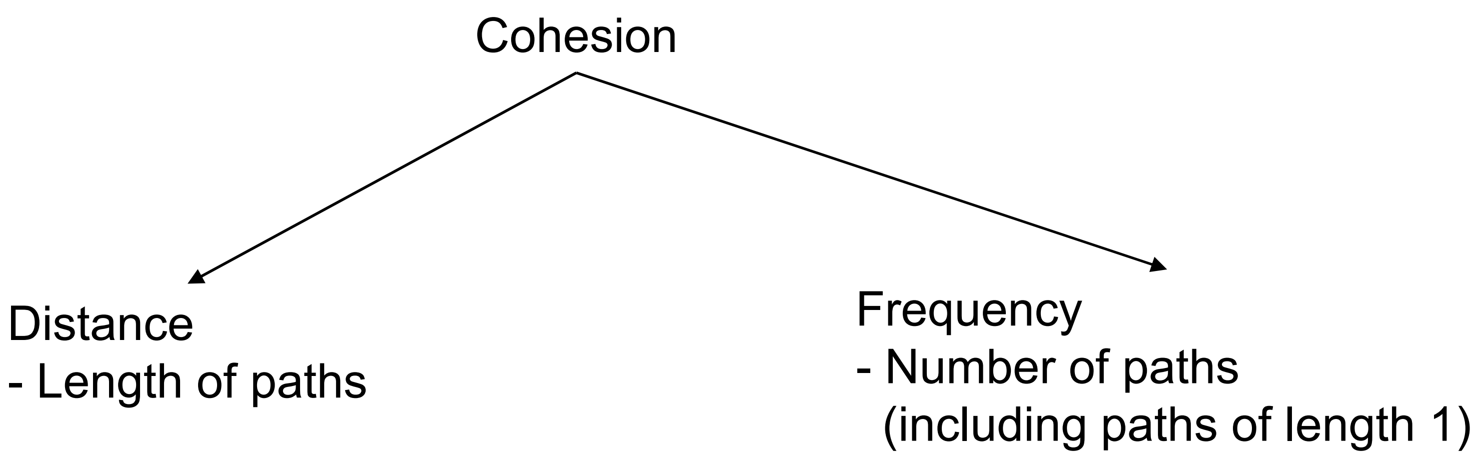


Graph Connectivity as aggregate dyadic cohesion measure

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

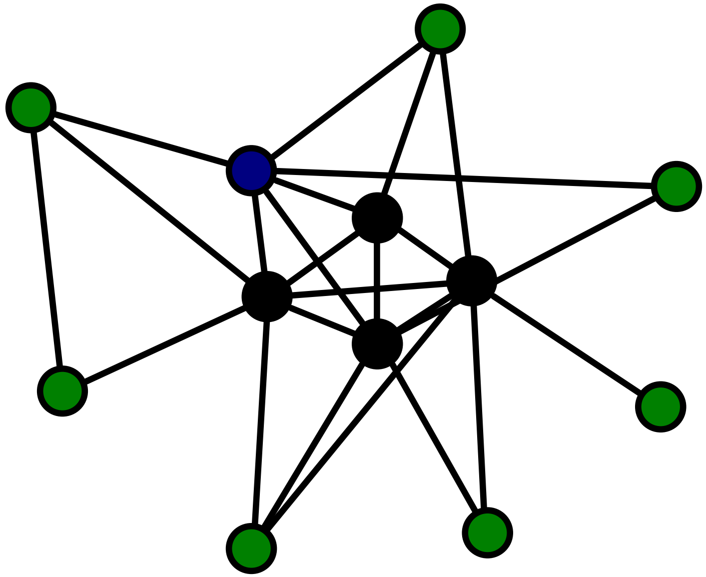
	MaxFlow
Mean	2.57
Std Dev	0.63
Sum	786
Variance	0.40
SSQ	2142
MCSSQ	123.06
Euc Norm	46.28
Minimum	2
Maximum	5
N of Obs	306

Classifying Cohesion

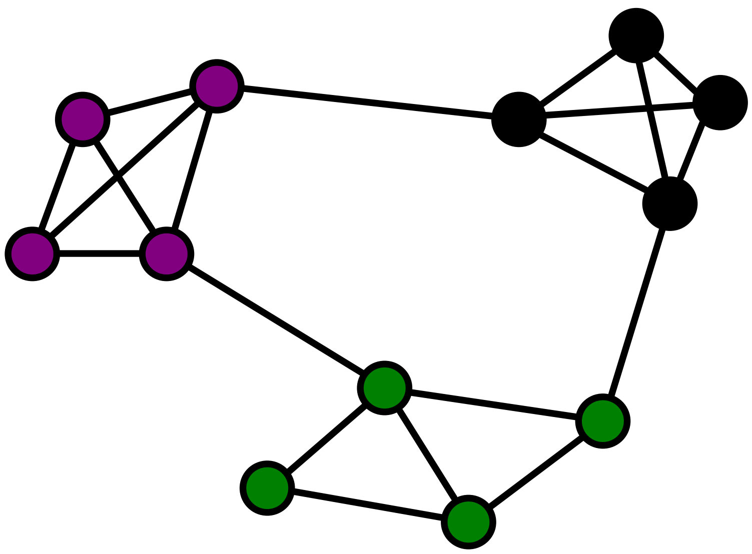


Average Distance

- Average geodesic distance between all pairs of nodes



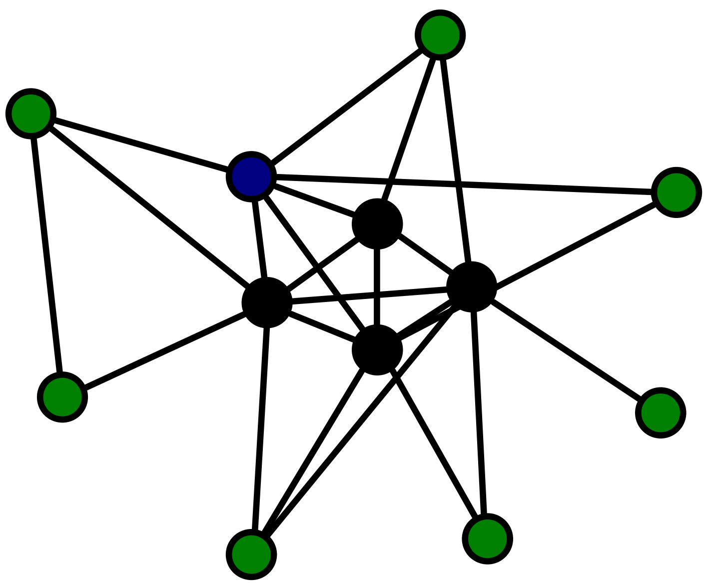
avg. dist. = 1.9



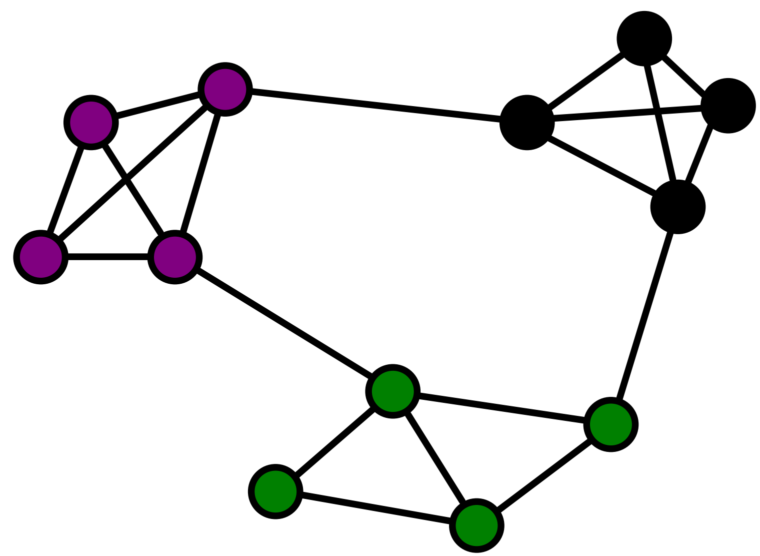
avg. dist. = 2.4

Diameter

- Maximum distance
  - Length of the longest shortest path



Diameter = 3



Diameter = 3

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

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

Cohesion family of concepts

	HO LLY	BIL L	DO N	HA RR Y	MI CH AEL	PA M	JEN NIE	AN N	PA ULI NE	PAT	CA RO L	LEE	JO HN	BR AZ EY	GE RY	STE VE	BE RT	RUSS	Sum
HOLLY			1	1	1	1				1									5
BILL			1	1	1														3
DON	1	1		1	1														4
HARRY	1	1	1		1														4
MICHAEL	1	1	1	1											1				5
PAM	1						1	1	1		1								5
JENNIE						1		1		1									3
ANN						1	1		1										3
PAULINE						1		1		1	1		1						5
PAT	1						1		1		1								4
CAROL						1			1	1									3
LEE														1		1	1		3
JOHN									1						1			1	3
BRAZEY												1				1	1		3
GERY					1								1			1		1	4
STEVE												1		1	1		1	1	5
BERT												1		1		1		1	4
RUSS													1		1	1	1		4
Sum	5	3	4	4	5	5	3	3	5	4	3	3	3	3	4	5	4	4	70

0.229 (average)

Concept:

- Relational
- Centrality
- Subgroups
- Network