Cohesion

Relational and Group

Relational vs Group

- Relational or dyadic cohesion refers to pairwise social closeness
- Network cohesion refers to the cohesion of an entire group

Ways to Approach This

- Many ways to define or theoretically conceive of cohesion
  - cohesion → outcome
  - What is the mechanism that would relate cohesion to the outcome of interest?
  - Define cohesion consistent with this mechanism
- For each way, we can then devise an operational measurement
  - Don't confuse the measure with the construct

Adjacency & Strength of Tie

- Raw dyadic data
- Positive ties
- Guttman scale of social closeness or obligation
- Valued relations
  - Frequency of interaction
  - Duration of relation
  - Intensity

Multiplexity

- Multiplexity is often what is meant by "relational embeddedness"
  - As in economic ties being embedded in social ties
- Combination of (the right set of) ties can be seen as yielding greater closeness than just one tie

Embedded Ties

- A tie (u,v) is structurally embedded if there exists node p (possibly several nodes) such that $(u,p) \in E$ and $(v,p) \in E$
  - I.e, then endpoints u and v have “friends” in common

Simmelian Ties

- Krackhardt's definition:
- A dyad has a simmelian tie if it is reciprocal ties to each other and to third parties
- The value of a simmelian tie is the number of third parties they have in common
  - Ideally, it is the number of cliques they have in common

Reachability

- If there exists a path from u to v of any length, then v is said to be reachable from u
- The reachability matrix R in which rij = 1 if I can reach j records the relational cohesions in the graph
- Is a weak form of cohesion – minimal in fact
- Can define a weak form of Simmelian ties on the reachability graph

Geodesic Distance

Adjacency

	a	b	c	d	e	f	g	h	i	j
a	0	1	1	1	0	0	0	0	0	0
b	1	0	1	0	1	0	0	0	0	0
c	1	1	0	1	0	0	0	0	0	0
d	1	0	1	0	1	0	0	0	0	0
e	0	1	0	1	0	1	0	0	0	0
f	0	0	0	0	1	0	1	0	1	0
g	0	0	0	0	0	1	0	1	0	1
h	0	0	0	0	0	0	1	0	1	1
i	0	0	0	0	0	1	0	1	0	1
j	0	0	0	0	0	0	1	1	1	0

Geodesic Distance

	a	b	c	d	e	f	g	h	i	j
a	0	1	1	1	2	3	4	5	4	5
b	1	0	1	2	1	2	3	4	3	4
c	1	1	0	1	2	3	4	5	4	5
d	1	2	1	0	1	2	3	4	3	4
e	2	1	2	1	0	1	2	3	2	3
f	3	2	3	2	1	0	1	2	1	2
g	4	3	4	3	2	1	0	1	2	1
h	5	4	5	4	3	2	1	0	1	1
i	4	3	4	3	2	1	2	1	0	1
j	5	4	5	4	3	2	1	1	1	0

More nuance in the representation of non-connection

Reciprocal Distance

	a	b	c	d	e	f	g	h	i	j
a	0.00	1.00	1.00	1.00	0.50	0.33	0.25	0.20	0.25	0.20
b	1.00	0.00	1.00	0.50	1.00	0.50	0.33	0.25	0.33	0.25
c	1.00	1.00	0.00	1.00	0.50	0.33	0.25	0.20	0.25	0.20
d	1.00	0.50	1.00	0.00	1.00	0.50	0.33	0.25	0.33	0.25
e	0.50	1.00	0.50	1.00	0.00	1.00	0.50	0.33	0.50	0.33
f	0.33	0.50	0.33	0.50	1.00	0.00	1.00	0.50	1.00	0.50
g	0.25	0.33	0.25	0.33	0.50	1.00	0.00	1.00	0.50	1.00
h	0.20	0.25	0.20	0.25	0.33	0.50	1.00	0.00	1.00	1.00
i	0.25	0.33	0.25	0.33	0.50	1.00	0.50	1.00	0.00	1.00
j	0.20	0.25	0.20	0.25	0.33	0.50	1.00	1.00	1.00	0.00

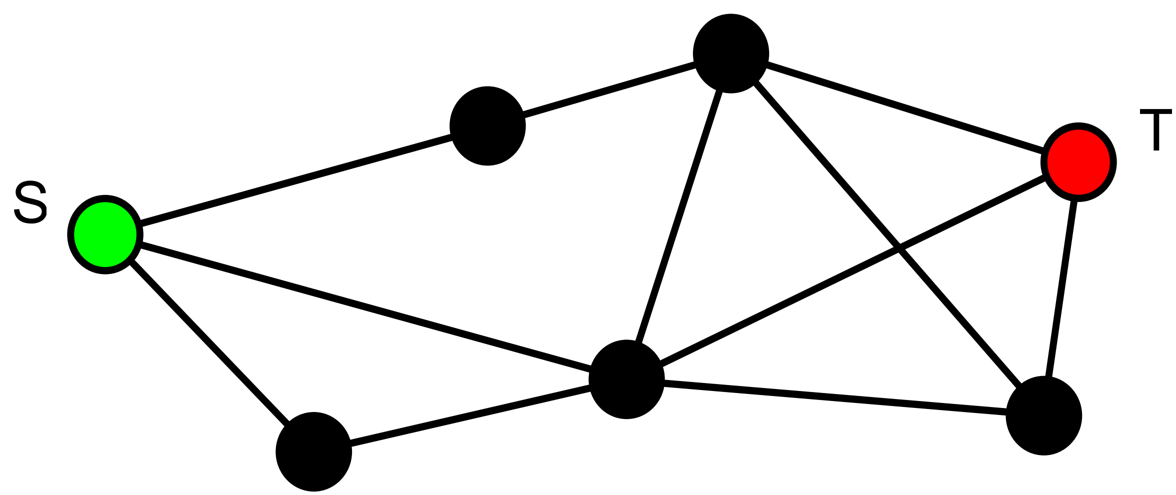
Number of Walks*

		1 a	2 b	3 c	4 d	5 e	6 f	7 g	8 h	9 i	10 j
1	a	194	167	195	167	154	50	30	12	30	12
2	b	167	188	167	188	115	82	22	30	22	30
3	c	195	167	194	167	154	50	30	12	30	12
4	d	167	188	167	188	115	82	22	30	22	30
5	e	154	115	154	115	150	59	82	50	82	50
6	f	50	82	50	82	59	150	115	154	115	154
7	g	30	22	30	22	82	115	188	167	188	167
8	h	12	30	12	30	50	154	167	194	167	195
9	i	30	22	30	22	82	115	188	167	188	167
10	j	12	30	12	30	50	154	167	195	167	194

*Of length of length 6 or less

Independent Paths

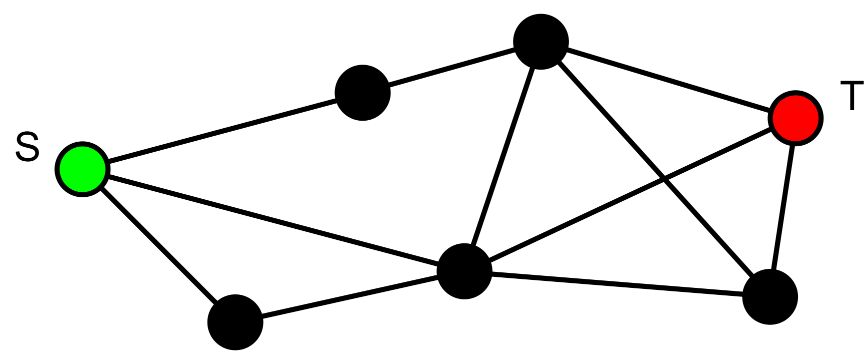
- A set of paths is node-independent if they share no nodes (except beginning and end)
  - They are line-independent if they share no lines



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

Connectivity

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

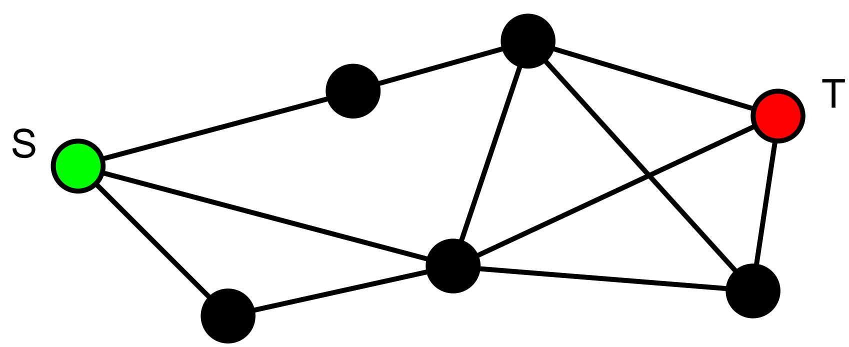


Menger's Theorem

- 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



Group Cohesion

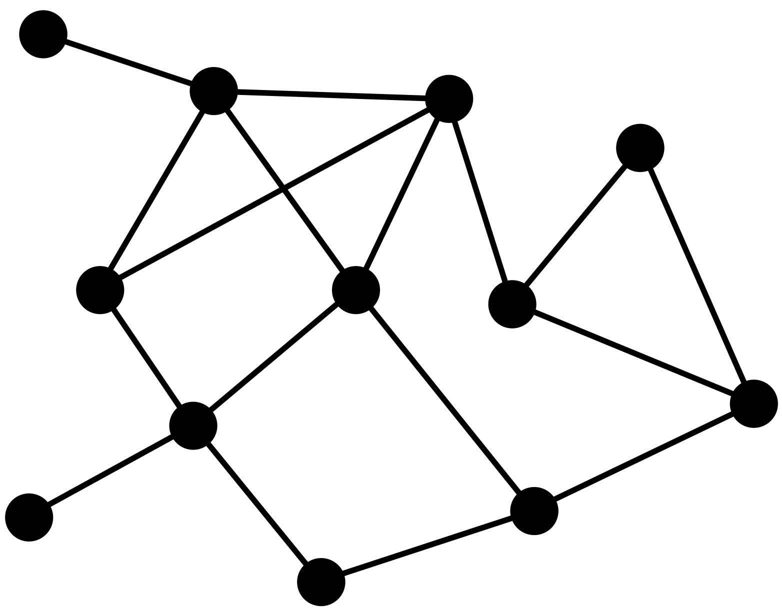
- Whole network measures can be
  - Averages of dyadic cohesion
  - Measures not easily reducible to dyadic measures

Measures of Group Cohesion

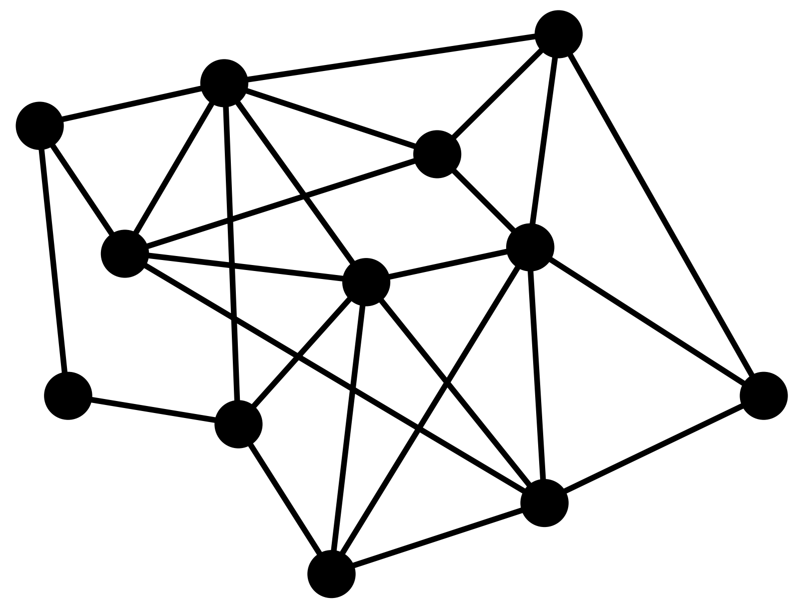
- Density & Average degree
- Average Distance and Diameter
- Number of components
- Fragmentation
- Distance-weighted Fragmentation
- Cliques per node
- Connectivity
- Centralization
- Core/Peripheriness
- Transitivity (clustering coefficient)

Density

- Number of ties, expressed as percentage of the number of ordered/unordered pairs

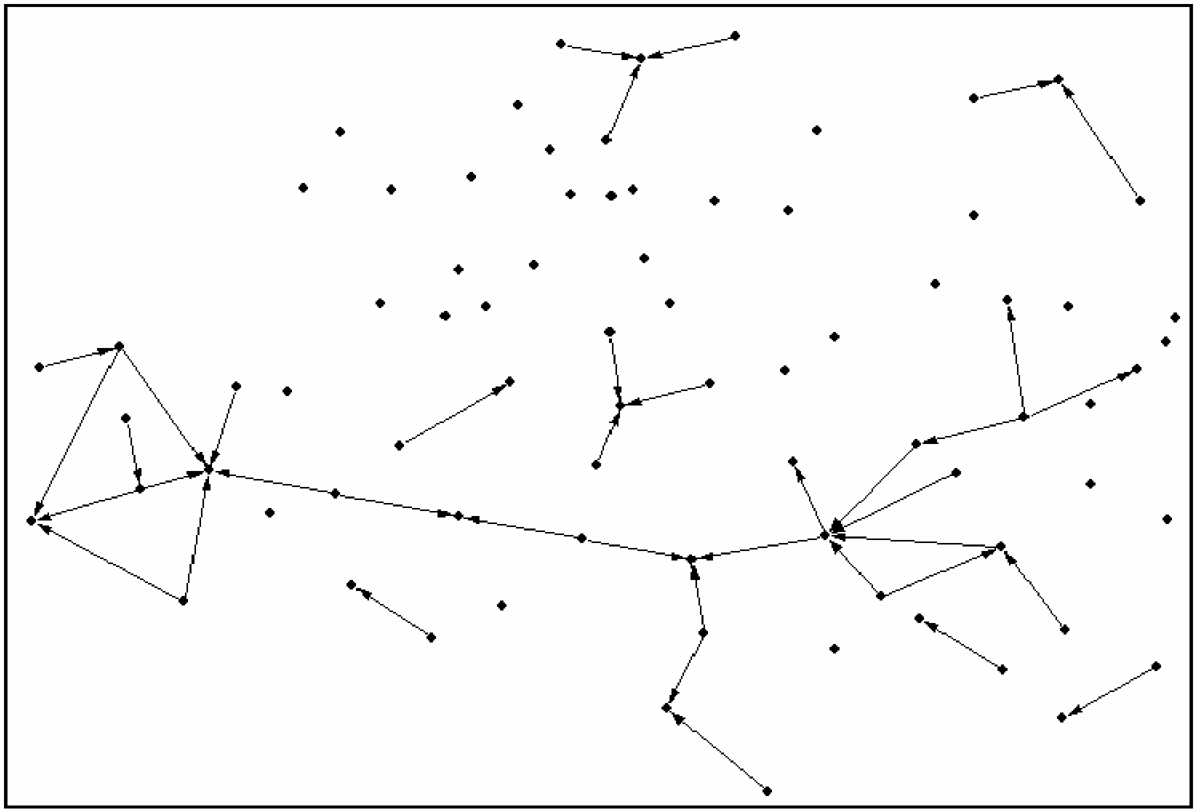


Low Density (25%)
Avg. Dist. = 2.27



High Density (39%)
Avg. Dist. = 1.76

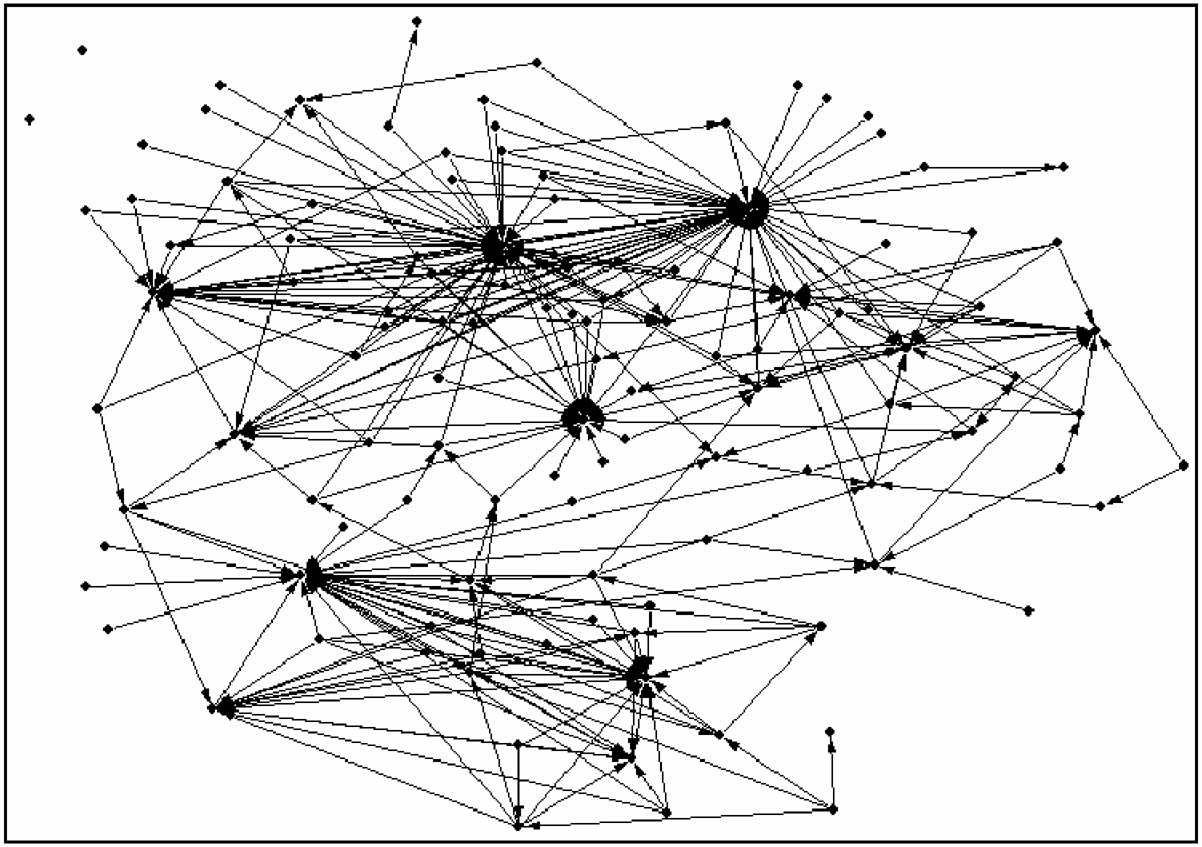
Help With the Rice Harvest



Village 1

Data from Entwistle et al

Help With the Rice Harvest



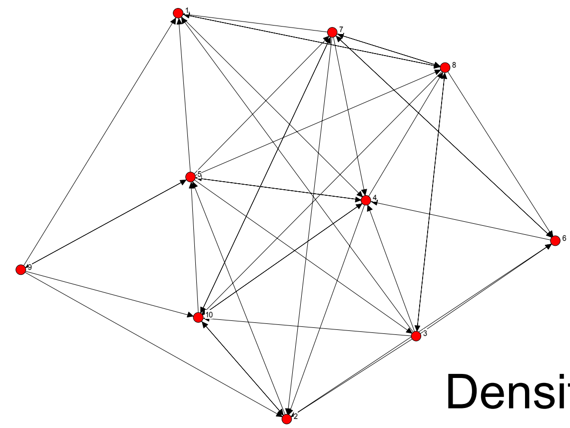
Village 2

Which village is more likely to survive?

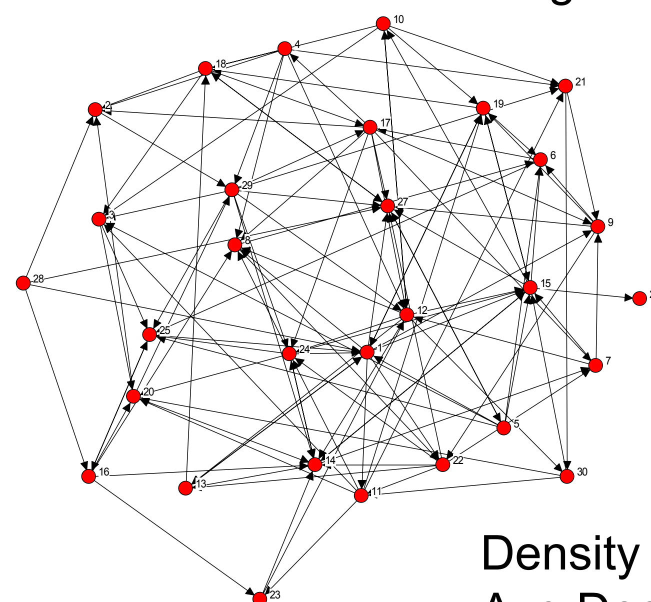
Data from Entwistle et al

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



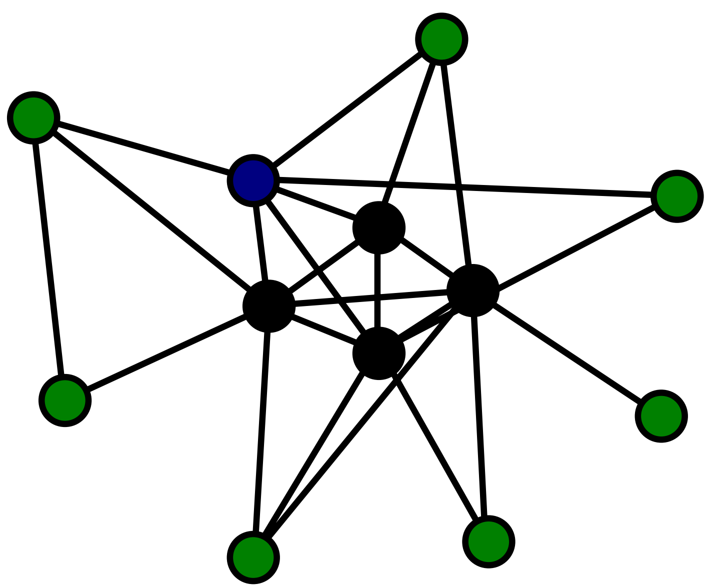
Density 0.47
Avg Deg 4

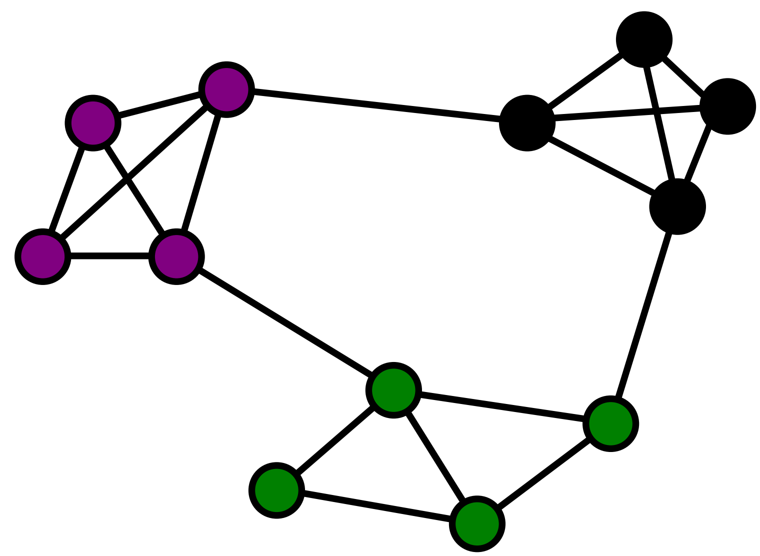


Density 0.14
Avg Deg 4

Average Distance

- Average geodesic distance between all pairs of nodes



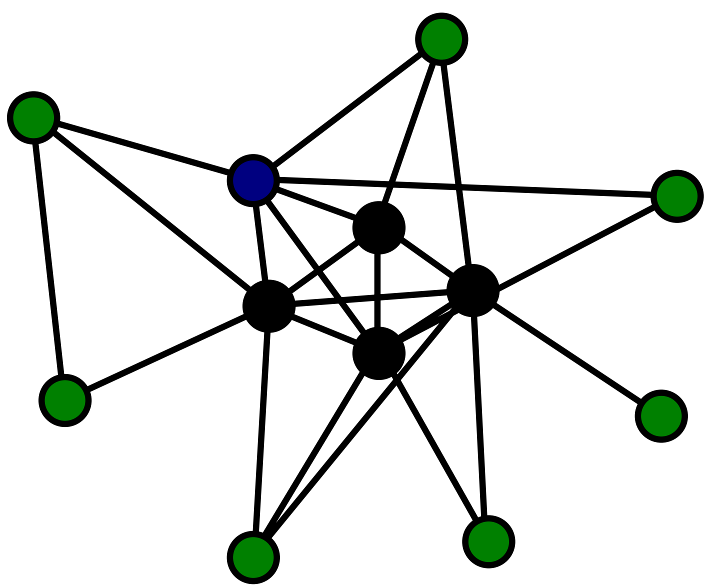


avg. dist. = 1.9

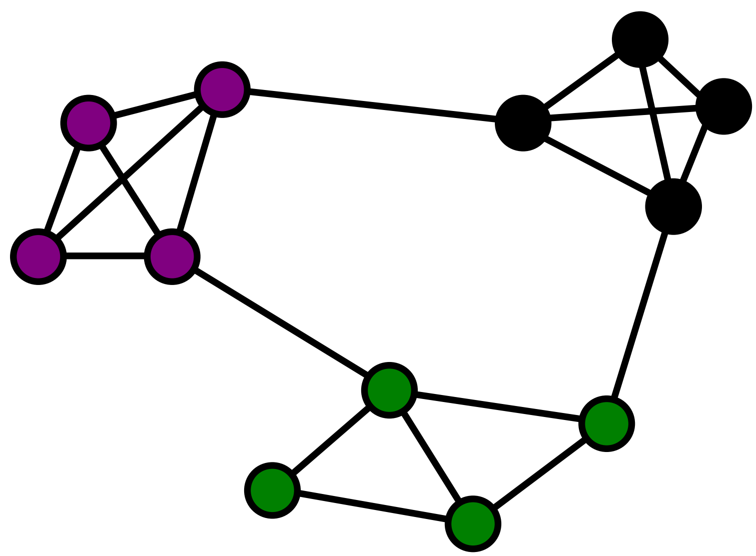
avg. dist. = 2.4

Diameter

- Maximum distance



Diameter = 3



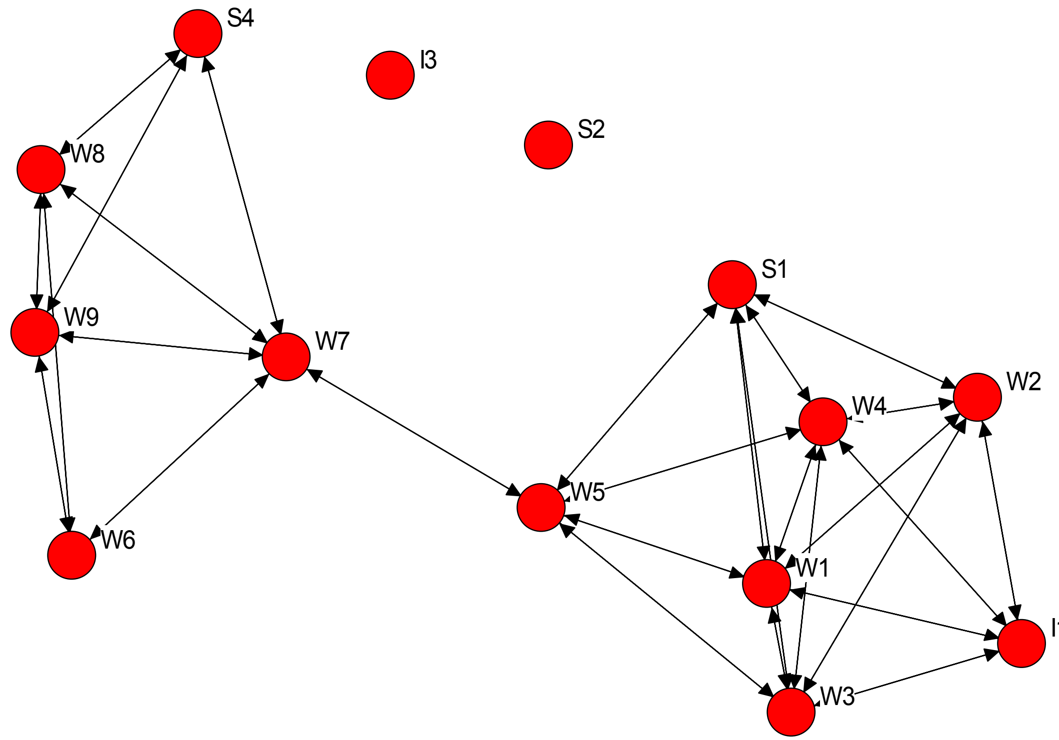
Diameter = 3

Fragmentation Measures

- Component ratio
- F measure of fragmentation
- Breadth (Distance-weighted fragmentation) B

Component Ratio

- No. of components divided by number of nodes



Component ratio = 3/14 = 0.21

F Measure of Fragmentation

- Proportion of pairs of nodes that are unreachable from each other

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

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

Computation Formula for F Measure

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

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

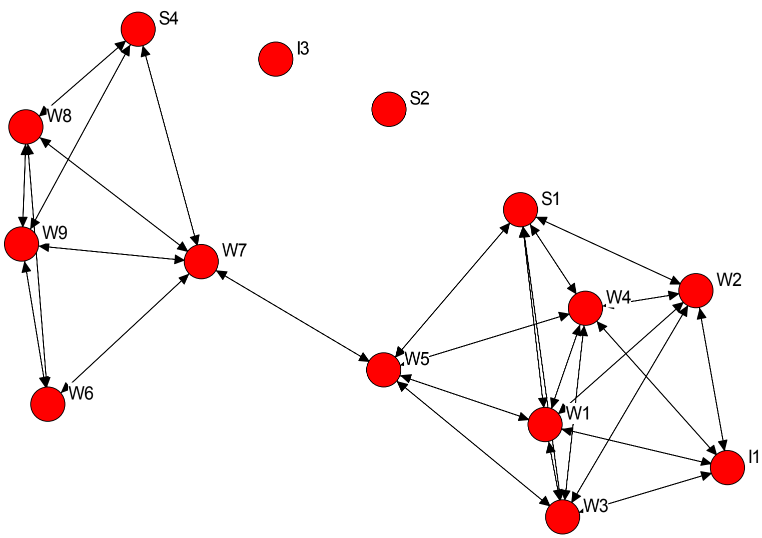
Sk = size of kth component

Computational Example

Games Data

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

0.2747 = 14/(132*131) = F



Heterogeneity/Concentration

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

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

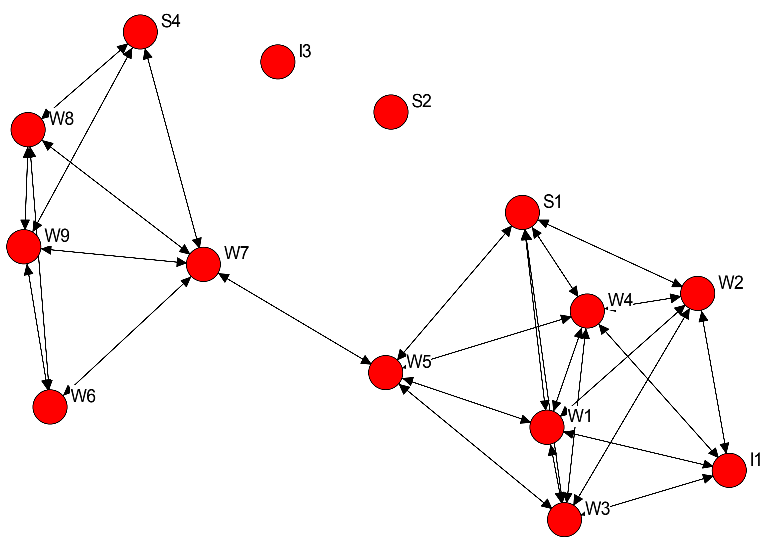
- Maximum value is 1-1/n
- Can be normalized by dividing by 1-1/n. If we do, we obtain the F measure

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

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

Breadth

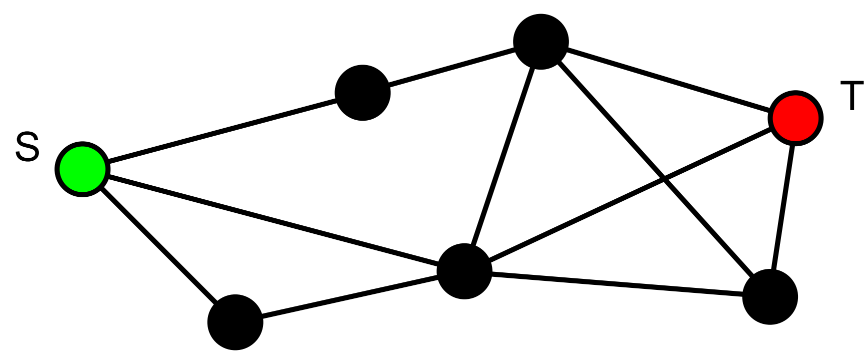
- Distance-Weighted Fragmentation
- Use average of the reciprocal of distance
  - letting $1/\infty = 0$

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

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

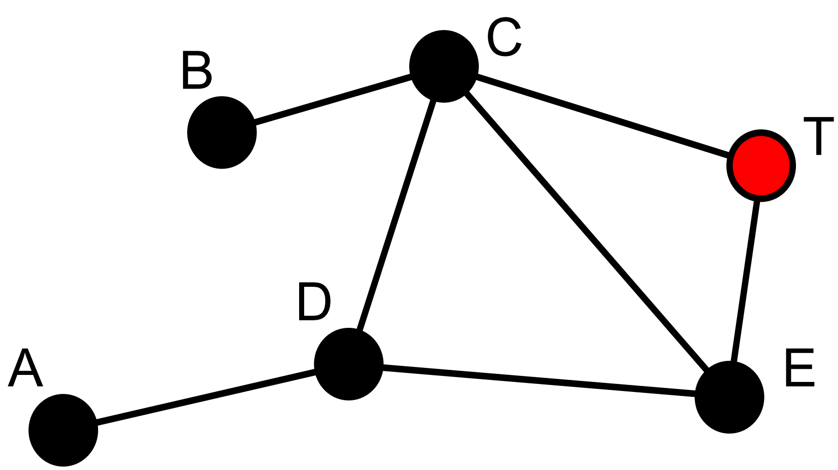
Connectivity

- 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



Transitivity

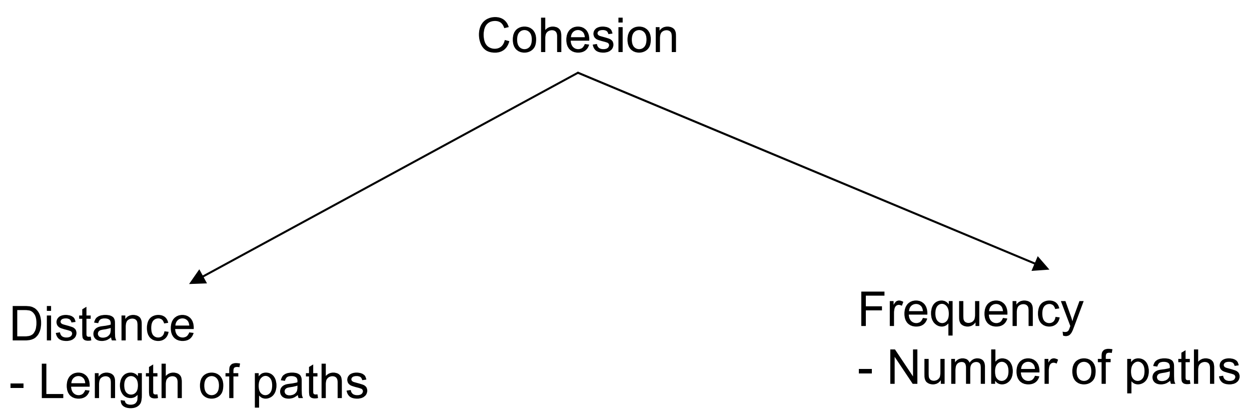
- Proportion of triples with 3 ties as a proportion of triples with 2 or more ties
  - Aka the clustering coefficient



cc = 12/26 = 46.15%

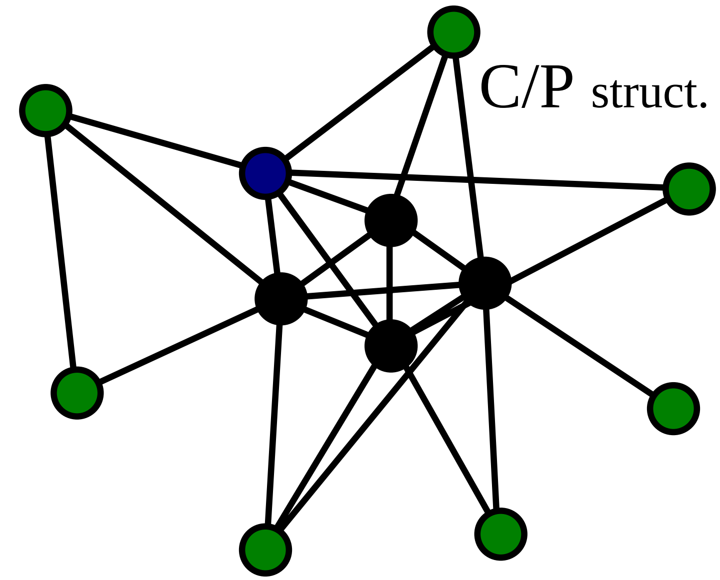
{C,T,E} is a transitive triple, but {B,C,D} is not. {A,D,T} is not counted at all.

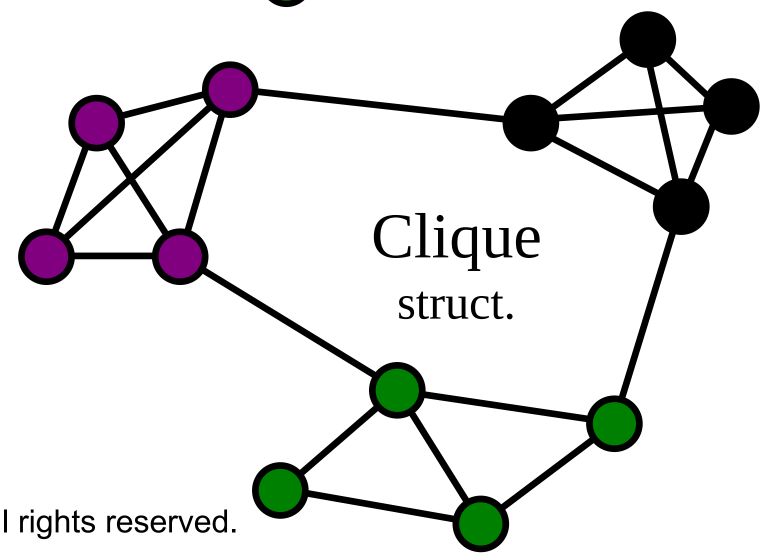
Classifying Cohesion



Core/Periphery Structures

- Does the network consist of a single group (a core) together with hangers-on (a periphery), or
- are there multiple sub-groups, each with their own peripheries?

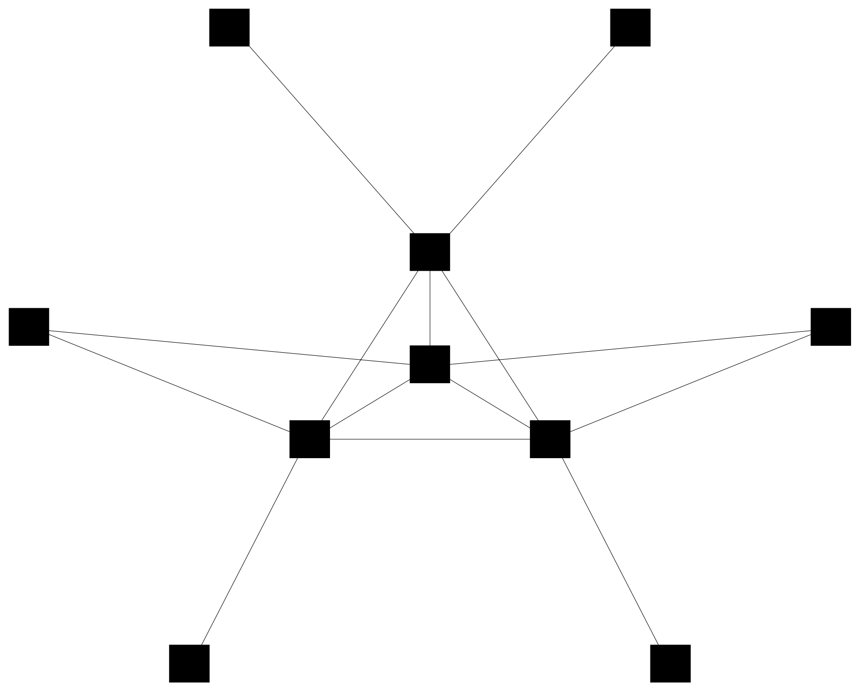




Kinds of CP/Models

- Partitions vs. subgraphs
  - just as in cohesive subgroups
- Discrete vs. continuous
  - classes, or
  - coreness

A Core/Periphery Structure



Blocked/Permuted Adjacency Matrix

	CORE				PERIPHERY					
	-	1	1	1	1	0	0	1	0	0
	1	-	1	1	0	1	1	0	0	0
CORE	1	1	-	1	0	0	0	1	1	0
	1	1	1	-	1	0	0	0	0	1
	1	0	0	1	-	0	0	0	0	0
	0	1	0	0	0	-	0	0	0	0
PERIPHERY	0	1	0	0	0	0	-	0	0	0
	1	0	1	0	0	0	0	-	0	0
	0	0	1	0	0	0	0	0	-	0
	0	0	0	1	0	0	0	0	0	-

- Core-core is 1-block
- Core-periphery are (imperfect) 1-blocks
- Periphery-periphery is 0-block

Idealized Blockmodel

	CORE				PERIPHERY					
CORE	-	1	1	1	1	1	1	1	1	1
	1	-	1	1	1	1	1	1	1	1
	1	1	-	1	1	1	1	1	1	1
	1	1	1	-	1	1	1	1	1	1
PERIPHERY	1	1	1	1	-	0	0	0	0	0
	1	1	1	1	0	-	0	0	0	0
	1	1	1	1	0	0	-	0	0	0
	1	1	1	1	0	0	0	-	0	0
	1	1	1	1	0	0	0	0	-	0
	1	1	1	1	0	0	0	0	0	-

c_i = class (core or periphery) that node *i* is assigned to

$$\delta_{ij} = \begin{Bmatrix} 1 \ if \ \ c_i = CORE \ \ or \ \ c_j = CORE \\ 0 \ otherwise \end{Bmatrix}$$

Partitioning a Data Matrix

- Given a graphmatrix, we can randomly assign nodes to either core or periphery
- Search for partition that resembles the ideal

Assessing Fit to Data

a_{ij} = cell in data matrix

c_i = class (core or periphery) that node *i* is assigned to

$$\delta_{ij} = \begin{cases} 1 \;\; if \;\; c_i = CORE \;\; or \;\; c_j = CORE \\ 0 \;\; otherwise \end{cases}$$

$$\rho = \sum_{i,j} a_{ij}\delta_{ij}$$

- A Pearson correlation coefficient r(A,D) is bett

Alternative Images

	Core					Periphery				
C	-	1	1	1	1	0	0	0	0	0
o	1	-	1	1	1	0	0	0	0	0
r	1	1	-	1	1	0	0	0	0	0
e	1	1	1	-	1	0	0	0	0	0
	1	1	1	1	-	0	0	0	0	0
P	0	0	0	0	0	-	0	0	0	0
e	0	0	0	0	0	0	-	0	0	0
r	0	0	0	0	0	0	0	-	0	0
i	0	0	0	0	0	0	0	0	-	0
	0	0	0	0	0	0	0	0	0	-

Alternative Images

	Core					Periphery				
C	-	1	1	1	1	-	-	-	-	-
o	1	-	1	1	1	-	-	-	-	-
r	1	1	-	1	1	-	-	-	-	-
e	1	1	1	-	1	-	-	-	-	-
	1	1	1	1	-	-	-	-	-	-
P	-	-	-	-	-	-	0	0	0	0
e	-	-	-	-	-	0	-	0	0	0
r	-	-	-	-	-	0	0	-	0	0
I	-	-	-	-	-	0	0	0	-	0
	-	-	-	-	-	0	0	0	0	-

Continuous Model

- $X_{ij} \sim C_i C_j$
  - Strength or probability of tie between node i and node j is function of product of coreness of each
  - Central players are connected to each other
  - Peripheral players are connected only to core

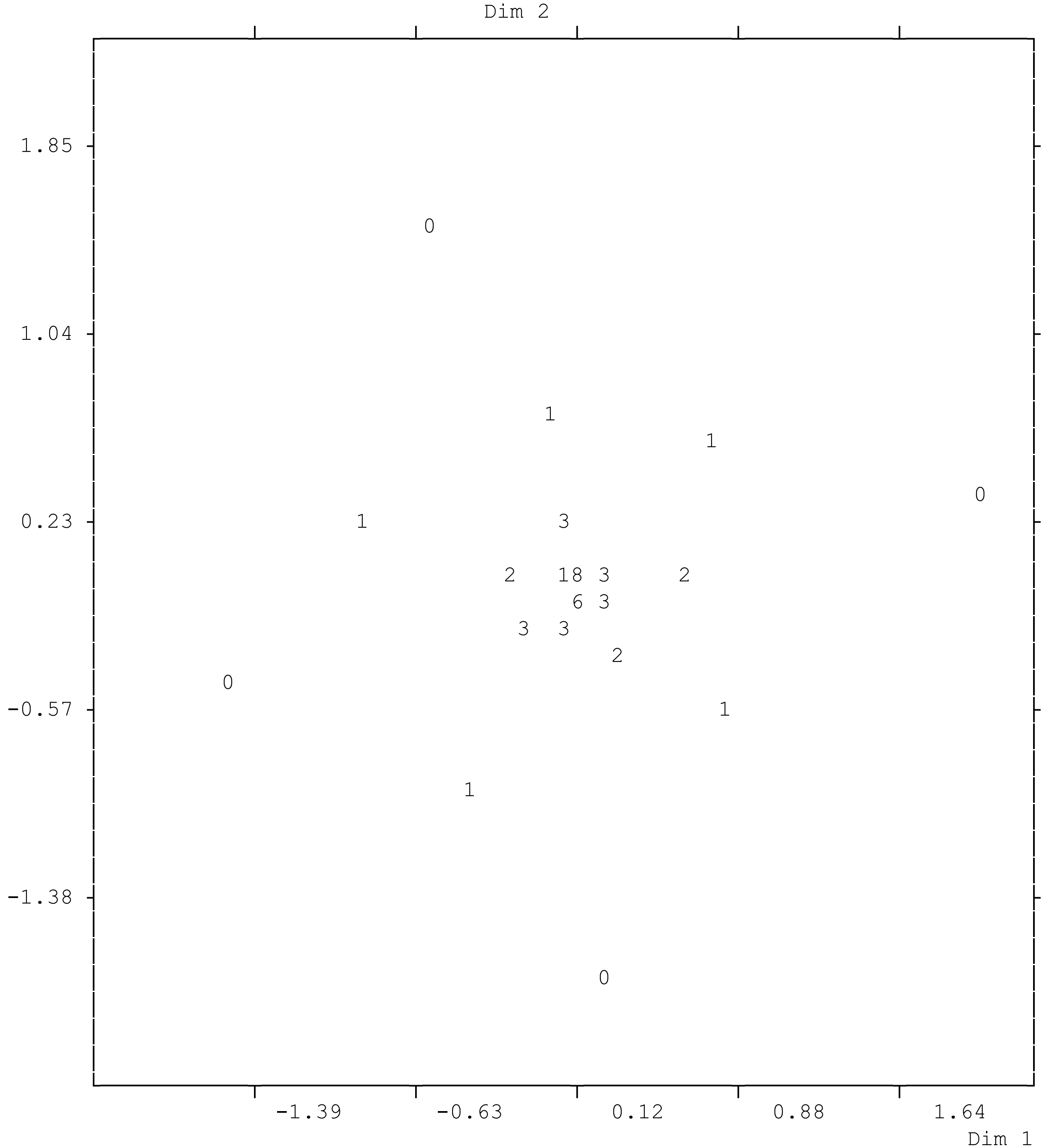
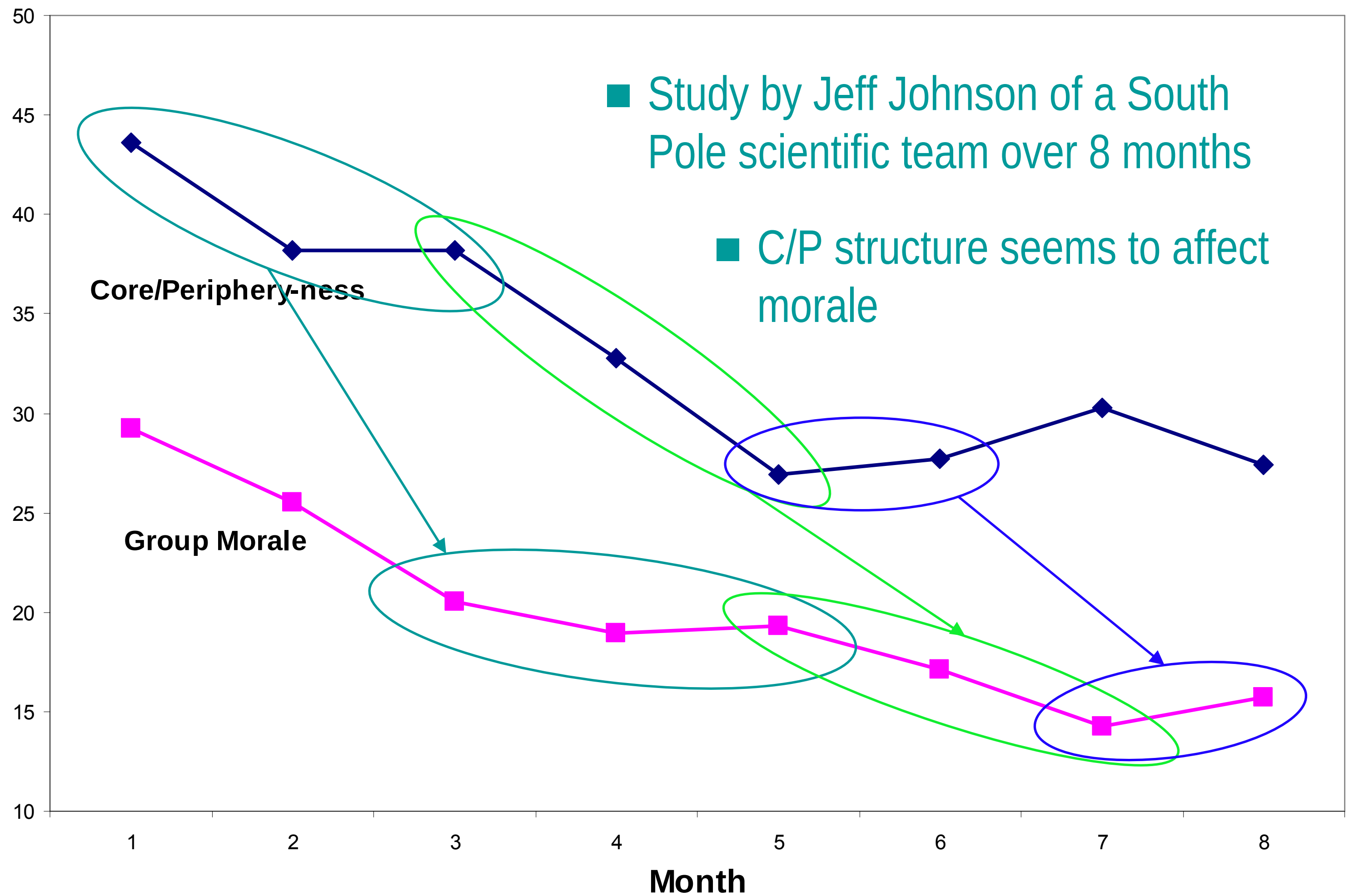


Figure 4. MDS of core/perip

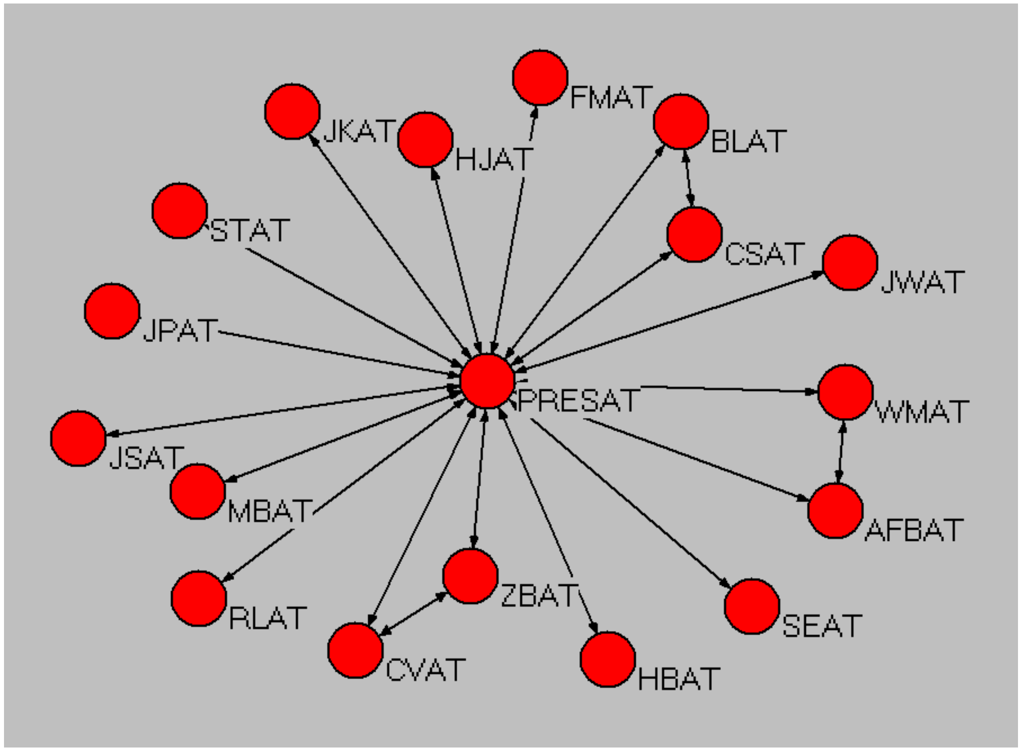
CP Structures & Morale

- Study by Jeff Johnson of a South Pole scientific team over 8 months
  - C/P structure seems to affect morale



Centralization

- Degree to which network revolves around a single node



Carter admin.
Year 1