Dynamic Network Analysis
Groups, Clustering, Structure

Dr. Kathleen M. Carley
June 16, 2006

Center for Computational Analysis of Social and Organizational Systems

http://www.casos.cs.cmu.edu/

Block Modeling

- A block model is a reduced form representation such that nodes are divided into a set of mutually exclusive groups
- The resulting groups can then be analyzed as a network such that
  - The group's connection to itself is the density of the connections among members
  - For each pair of groups, the inter-group connection is the density of the connections of group 1 (row) to group 2 (column)
  - The resulting block matrix can be turned into a binary matrix by simply comparing the level of connections in the block to the overall density of the original matrix such that there if the value of the cell is >= to the overall density then we replace it with a 1, else 0

Carnegie Mellon

Example

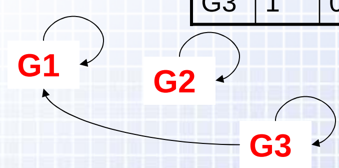
	A	B	C	D	E	F	G	H	I	J
A	0	1	1							
B	1	G1		1						
C	1		0					1		
D	1	1		0						
E					0	1				
F					1	0				
G	1						0	1		1
H		1					1	0	1	
I								1	0	1
J	1			1			1			0

G2 G3

Density = 21/90 = .22

	G1	G2	G3
G1	.58	0	.06
G2	0	1	0
G3	.25	0	.58

	G1	G2	G3
G1	1	0	0
G2	0	1	0
G3	1	0	1



CASOS

Carnegie Mellon

Common Blockmodels

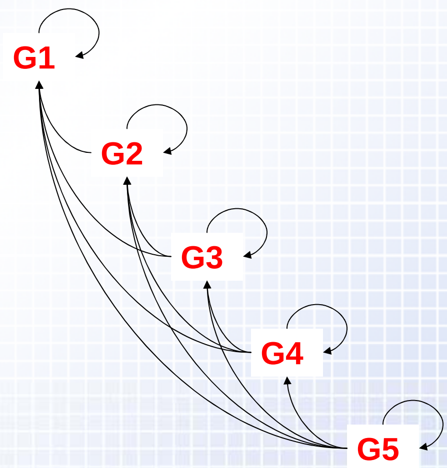
- Completely connected → 11 11
- Opposing groups → 10 01
- Supporters and Supporting → 01 10
- Central Core → 10 00
- Hierarchy → 10 10
- Core with Outreach → 11 00
- Core-periphery → 11 10
- Isolates → 00 00

CASOS

Carnegie Mellon

Illustrative Hierarchy

	G1	G2	G3	G4	G5
G1	1	0	0	0	0
G2	1	1	0	0	0
G3	1	1	1	0	0
G4	1	1	1	1	0
G5	1	1	1	1	1

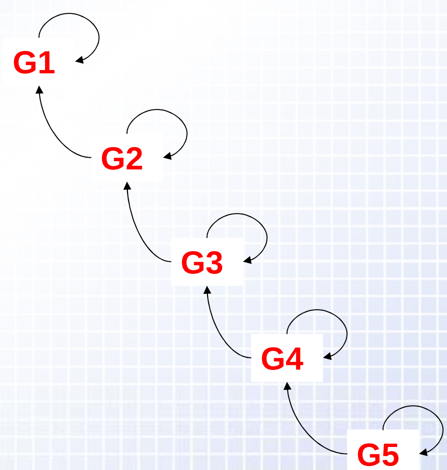


CASOS

Carnegie Mellon

Illustrative Alternative Hierarchy

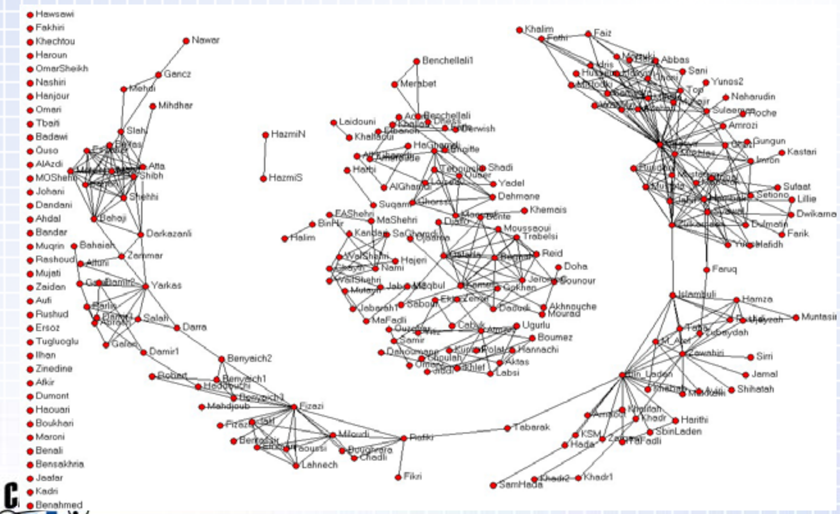
	G1	G2	G3	G4	G5
G1	1	0	0	0	0
G2	1	1	0	0	0
G3	0	1	1	0	0
G4	0	0	1	1	0
G5	0	0	0	1	1



CASOS

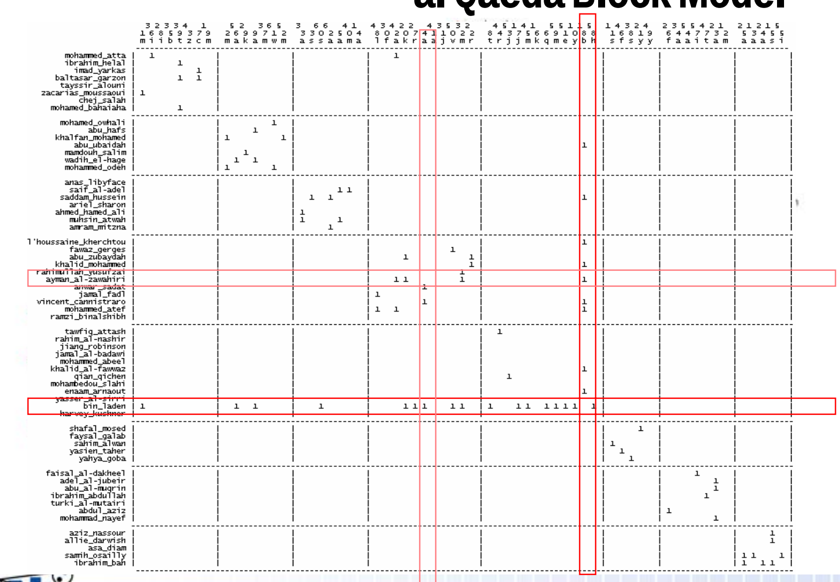
Carnegie Mellon

ORA Demonstration
Output



Carnegie Mellon

al Qaeda Block Model



Carnegie Mellon

Terminology

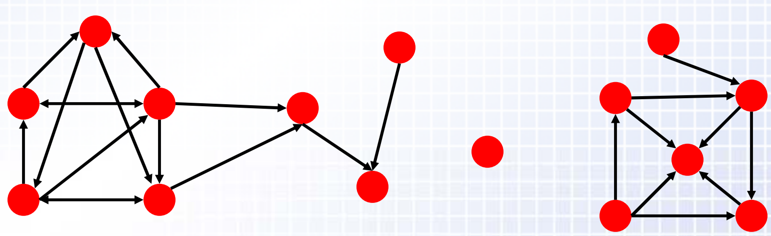
- Graph - (V,E)
  - consists of a set of ***nodes*** V(G) and a set of ***links*** E(G)
- Alpha operator
  - Let α(S1,S2) be the number of ties from members of set S1 to members of the set S2
  - α(*u*,S) is number of ties node *u* has with members of set S
  - α(S) is number of ties from members of set S to members of V-S (i.e., all other nodes)

CASOS

Carnegie Mellon

Terminology: Components

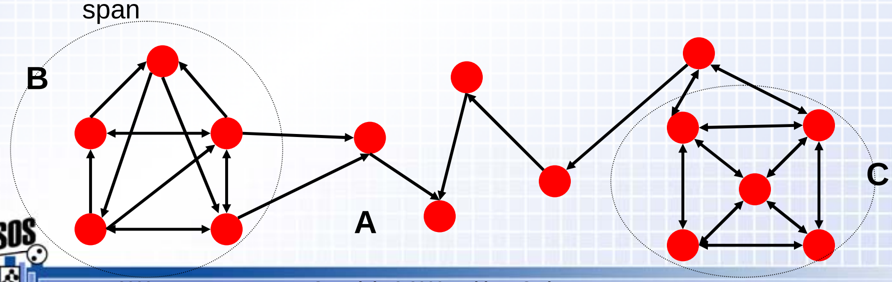
- A subgraph S of a graph G is a component if S is maximal and connected
- If G is a digraph, then
  - S is a weak component if it is a component of the underlying (undirected) graph
  - S is a strong component if for all dyads u,v in S, there is a path from u to v
- Finding components is the first step in analysis of large graphs
  - Analyze each component separately, or discard very small components



CASOS

Terminology: K-Cores

- K-CORE
  - A maximal subgraph S such that for all u in S, $\alpha(u,S) >= k$
    - S=A is 1-core & 2-core; B and C 3-core
    - There is no 4-core or higher
  - Finds large regions within which cohesive subgroups may be found
  - Identifies fault lines across which cohesive subgroups do not span



Groups

- Set of nodes that meet some criteria – a node set
- Goal is to extract these automatically based on node properties (such as – how they are connected)
- Finding groups is pattern analysis
- 2 types of approaches mechanistically
  - Bottom up – combine
    - E.g., Clustering nodes
    - E.g., Cluster "dyads" or "links"
  - Top down – split entire set into subsets
    - E.g. break up groups (Concor)
    - E.g. segregate set of links
- 2 types of approaches based on need
  - Locate members, locate anomalies
  - Break the network (locate components, sub-cells, …), segregate links

CarnegieMellon

Group Rationales

3 conceptual reasons for why groups matter

- Cohesion
  - Because the nodes have the same kind of position – relations to same type of other nodes
  - Network region might contain cohesive subgroups
- Equivalence
  - Because the nodes have the same linkages = relationships to the same other nodes
- Distinction
  - Because the nodes are different from other nodes around them, anomalies

NOTE: A group may or may not be a ***component or a K-Cores***

CASOS

CarnegieMellon

Canonical Hypothesis

- Similar nodes have similar outcomes
  - If two nodes occupy the same position, then they will get the same results, even if unconnected to each other
    - Even if only connected to similar others – cohesion
    - Only if connected to same others – equivalence
- Networks with similar structures will have similar outcomes
  - Similar structures = similar topology
  - E.g., Similarly structured teams will have similar performance outcomes
- Members of group will have similar outcomes
  - Ideas, attitudes, illnesses, behaviors
  - Due to interpersonal transmission
    - Transference
    - Influence / persuasion
    - Co-construction of beliefs & practices
      - As in communities of practice

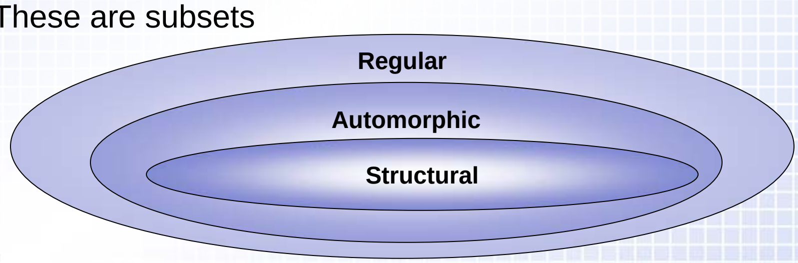
CASOS

Generic Mechanisms

- Structural substitutability
  - Structural processes affect structurally similar nodes similarly
  - Two nodes connected to the same other nodes can be substituted with no loss – equivalence
  - Two nodes connected to similar other nodes play the same role in a group – cohesion
- Environmental determinism
  - Location, location, location
  - Nodes with similar environments are similarly affected by the environment
    - Important when environment is important
  - Nodes connected to the same others get the same “stuff” through identical paths – equivalence
  - Nodes connected to similar others get the same “stuff” through equivalent paths

Groups and Equivalences

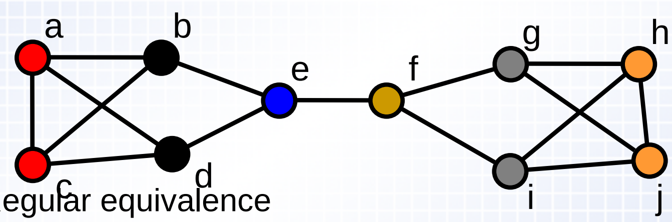
- Many grouping mechanisms are based on equivalences
- Common ones:
  - Structural
  - Regular
  - Automorphic *At least as defined in JMS paper in 1994.
- These are subsets



Carnegie Mellon

Estimates of Equivalence

- Structural equivalence
  - Two nodes are equivalent if connected to same others
  - Concor



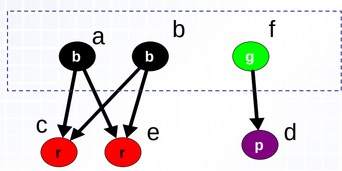
- Regular equivalence
  - Two nodes are equivalent if connected to others who are connected in the same way
  - rege

CASOS

Carnegie Mellon

Issues with Structural Equivalence

- Is there a Mechanism? What's the mechanism leading to similarity?
  - Confounds similarity with contiguity



Parents?

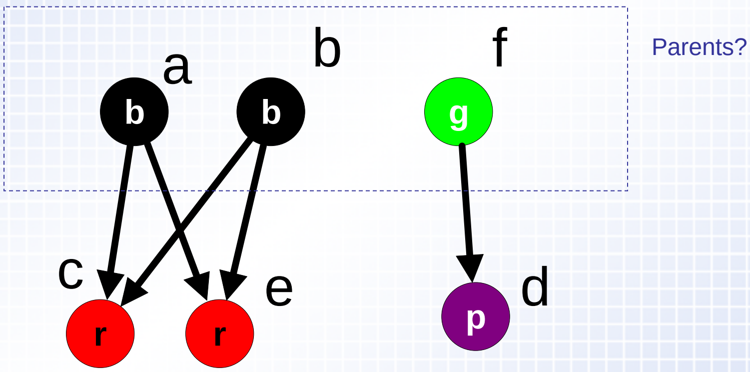
According to structural equivalence, only parents of the same children are playing the same role

- What kinds of entities should exhibit structural equivalence?
  - Aren't humans too unique???
  - Approximation – vs – Actual
- Tools for finding structural equivalent groups
  - Concor
  - Heuristic based mechanisms

CASOS

Carnegie Mellon

Issue: SE is not necessarily a Social Role



According to structural equivalence, only parents of the same children are playing the same role

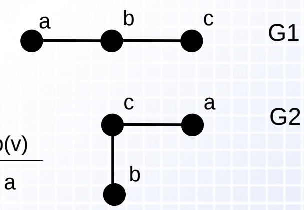
Carnegie Mellon

Structural Equivalence

- Compute similarity/distance between rows of adjacency matrix
  - Correlation
  - Euclidean distance
- Much argument over handling of diagonals
- Can then MDS or cluster the resulting proximity matrix
  - Bottom-up
  - Problem – stopping algorithm
- Or use Concor
  - Correlation – iteratively
  - Problem – top-down, and so imposes structure

Terminology: Isomorphism

- Two graphs are isomorphic if you can find a 1:1 mapping of nodes from one to the other that preserves adjacency structure
- G(V,E) is isomorphic to G'(V',E') if there exists mapping p:V→V' such that (u,v) $\in$ E iff (p(u),p(v)) $\in$ E'
  - Such a mapping p is called an isomorphism

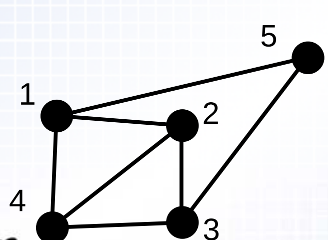


v	p(v)
a	a
b	c
c	b

(a,b) ←→ (a,c)
(b,c) ←→ (c,b)

Automorphism

- Aka structural isomorphism
- An automorphism is an isomorphism of graph to itself

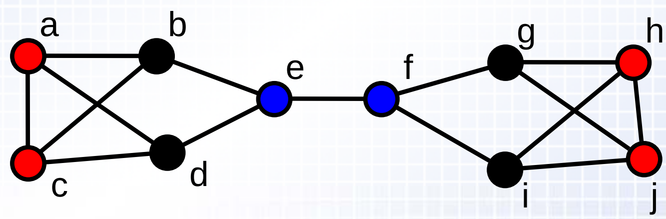


v	p(v)
1	3
2	4
3	1
4	2
5	5

Carnegie Mellon

Automorphic Equivalence

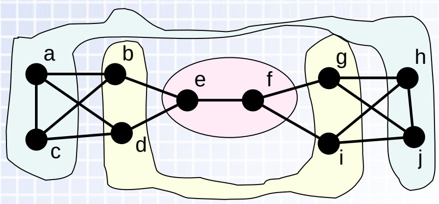
- A coloration C is automorphic if C(u)=C(v) iff there exists automorphism p such that u=p(v)



CASOS

Carnegie Mellon

Automorphic Equivalence



G	1	2	3	4	5	6	7	8	…	20
A	H	C	J	C	A	C	A	J	…	A
B	G	D	I	D	B	B	D	G	…	B
C	J	A	H	A	C	A	C	H	…	C
D	I	B	G	B	D	D	B	I	…	D
E	F	E	F	E	E	E	E	F	…	E
F	E	F	E	F	F	F	F	E	…	F
G	B	I	D	G	I	G	G	B	…	G
H	A	J	C	H	J	H	H	C	…	H
I	D	G	B	I	G	I	I	D	…	I
J	C	H	A	J	H	J	J	A	…	J

CASOS

Carnegie Mellon

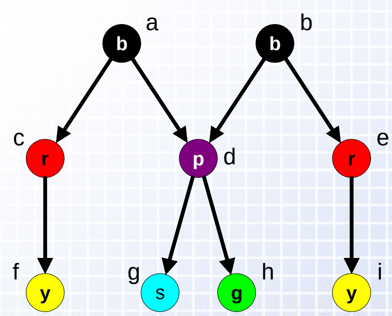
Automorphic Equivalence

- Fits with “role” mechanisms
- Powerful, fundamental intuitive concept
- Truly structural/positional, not confounded by contiguity
- Captures essentials of the role concept
- Generalization of structural equivalence

Carnegie Mellon

Issues with Automorphic Equivalence

- Very Strict mechanisms
  - A parent with 2 children does not play the same role as one with 3 children
- What kinds of entities should exhibit automorphic equivalence?
  - Aren’t humans too unique???
  - Approximation – vs – Actual
- Tools for finding structural equivalent groups
  - Rege – heuristic based routine
  - Extremely difficult to compute
  - No obvious way to relax the concept for application to real world data



Copyright © 2006 Kathleen Carley

Carnegie Mellon

Regular Equivalence

- Captures role concept really well
  - Two actors are equivalent if they have the same relations with equivalent others
  - You call American airlines and talk to clerk about booking flight, while I call USAIR and do same with their clerk
    - You and I equivalent because the clerks are equivalent (and they are equivalent because you and I are equivalent)
- Less strict than automorphiic
  - Not concerned with quantities of colors
  - Finds more equivalent nodes

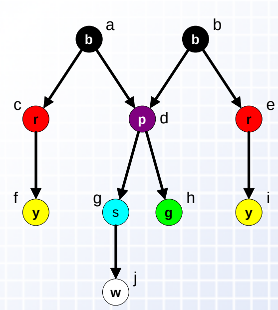
CASOS

Compliments of Steve Borgatti

Carnegie Mellon

Regular Equivalence

- Also captures position in hierarchies well, if no cycles
  - Including trophic group
- Relatively easy to compute (and to relax)
- Easy to generalize to 2-mode data
  - Consumers & brands
    - Segments & positions
    - identifying category boundaries
- Works well with multiple relations

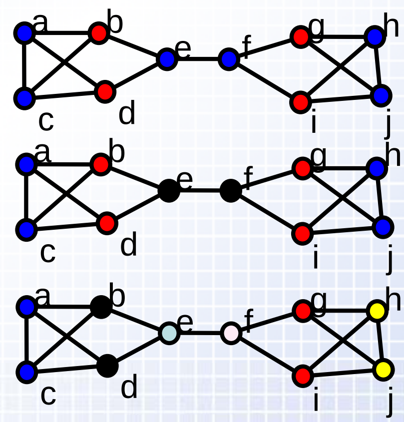


CASOS

Carnegie Mellon

Issues with Regular Equivalence

- Often hard to interpret
  - Humans good at understanding pattern similarities, but not in the context of social ties
  - Data sets inappropriate for R.E. analysis
    - Too small, no real roles
- A graph may have multiple colorations that are regular – especially undirected graphs
- Heuristic tools can vary widely in results and have poor scaling properties



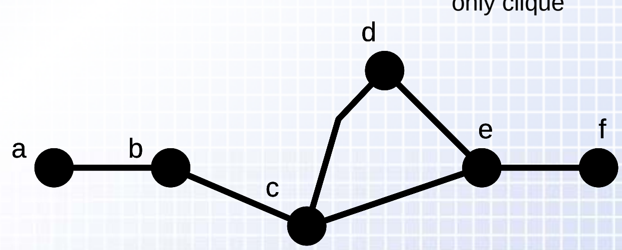
CASOS

Carnegie Mellon

Cliques

- Definition
  - Maximal, complete subgraph
  - Set S s.t. for all u,v in S, (u,v) in E
- Properties
  - Maximum density (1.0)
  - Minimum distances (all 1)
  - overlapping
  - Strict

{c,d,e} is the only clique



CASOS

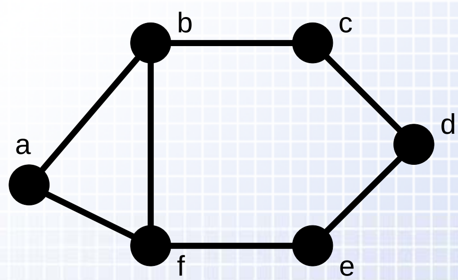
 Compliments of Steve Borgatti

Types of Relaxations

- Distance (length of paths)
  - N-clique, n-clan, n-club
- Density (number of ties)
  - K-plex, ls-set, lambda set, k-core, component
- Both Distance and Density
  - Factions
  - Look at ratio of within to without ties – optimize groups to maximize this ration

N-cliques

- Definition
  - Maximal subset s.t. for all u,v in S, d(u,v) <= n
  - Distance among members less than specified maximum
  - When n = 1, we have a clique
- Properties
  - Relaxes notion of clique
    - Avg distance can be greater than 1



Is {a,b,c,f,e} a 2-clique?

Compliments of Steve Borgatti

Carnegie Mellon

Issues with N-Cliques

- Overlapping
  - {a,b,c,f,e} and {b,c,d,f,e} are both 2-cliques
- Membership criterion satisfiable through non-members
- Even 2-cliques can be fairly non-cohesive
  - Red nodes belong to same 2-clique but none are adjacent