Social Network Analysis

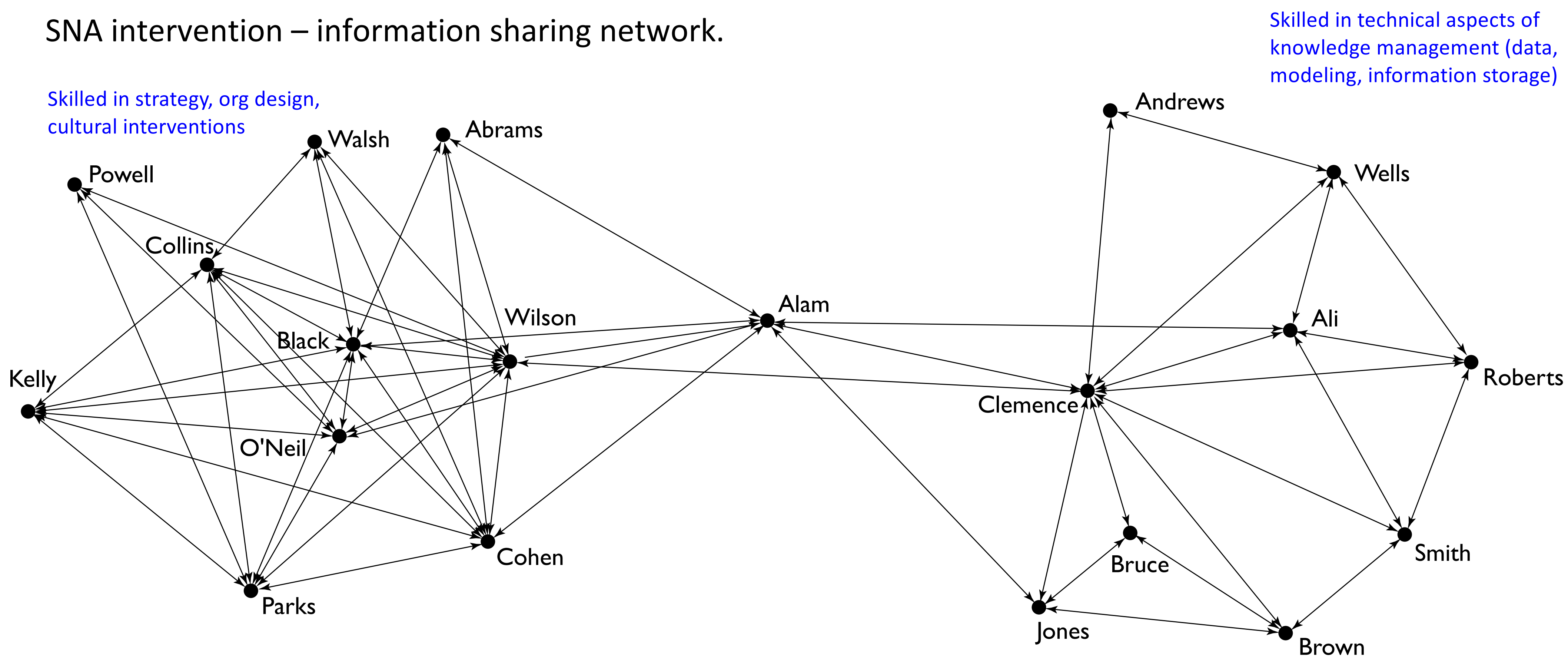
Day 2

Social Capital, Brokerage & Equivalence

Information & Success

Global consulting organization had group dedicated to provide thought leadership and specialized support to to the organization's knowledge management consultants. Group was composed of people with industry experience in **(1) organizational design (soft-skills)** and **(2) technical fields (data warehousing)**.
USP: **holistic knowledge management solution.** However, they were not delivering. Why?

SNA intervention – information sharing network.

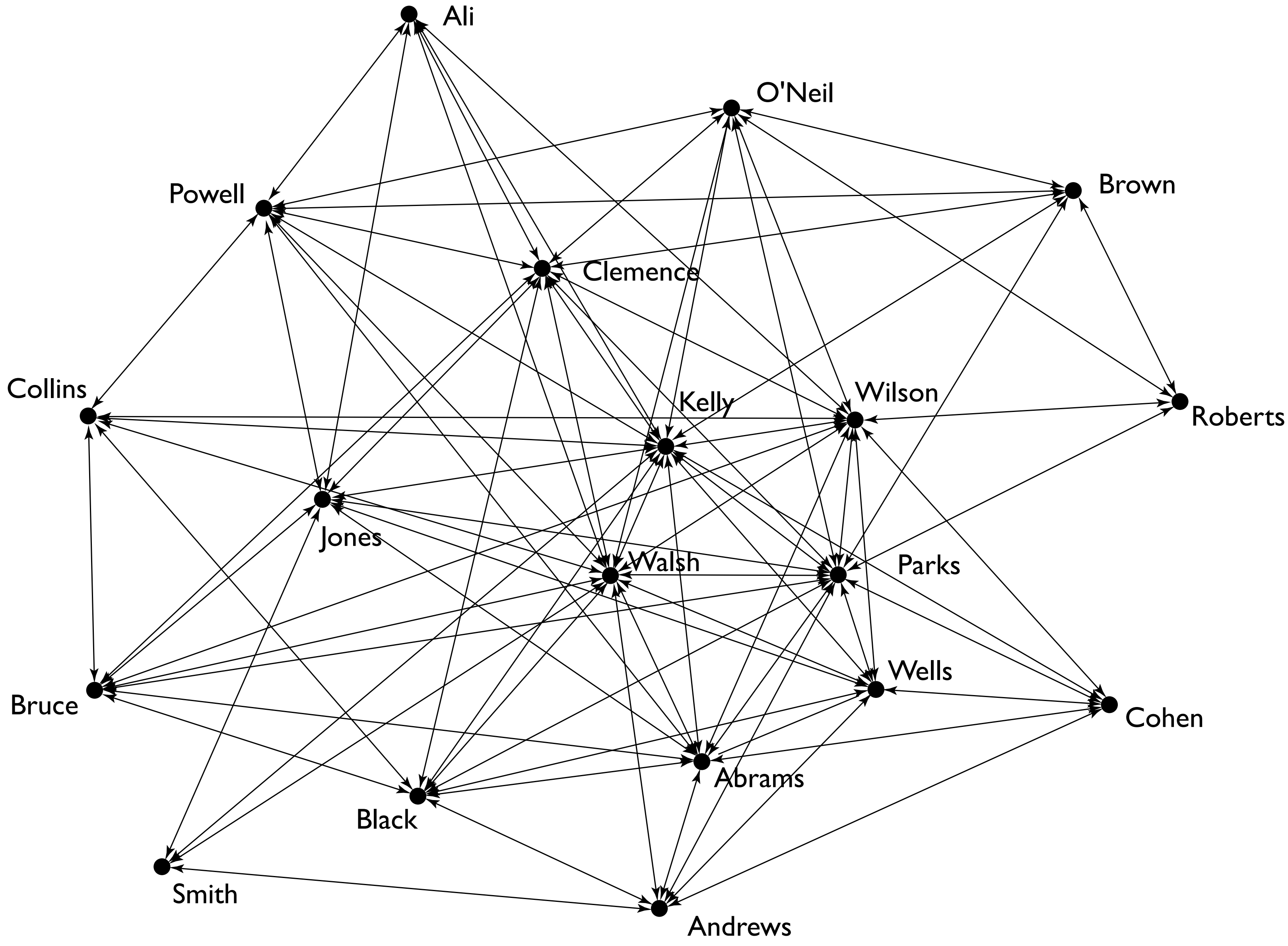


Cross, Borgatti & Parker (2002), "Making Invisible Work Visible: using social network analysis to support strategic collaboration".

Changes Made

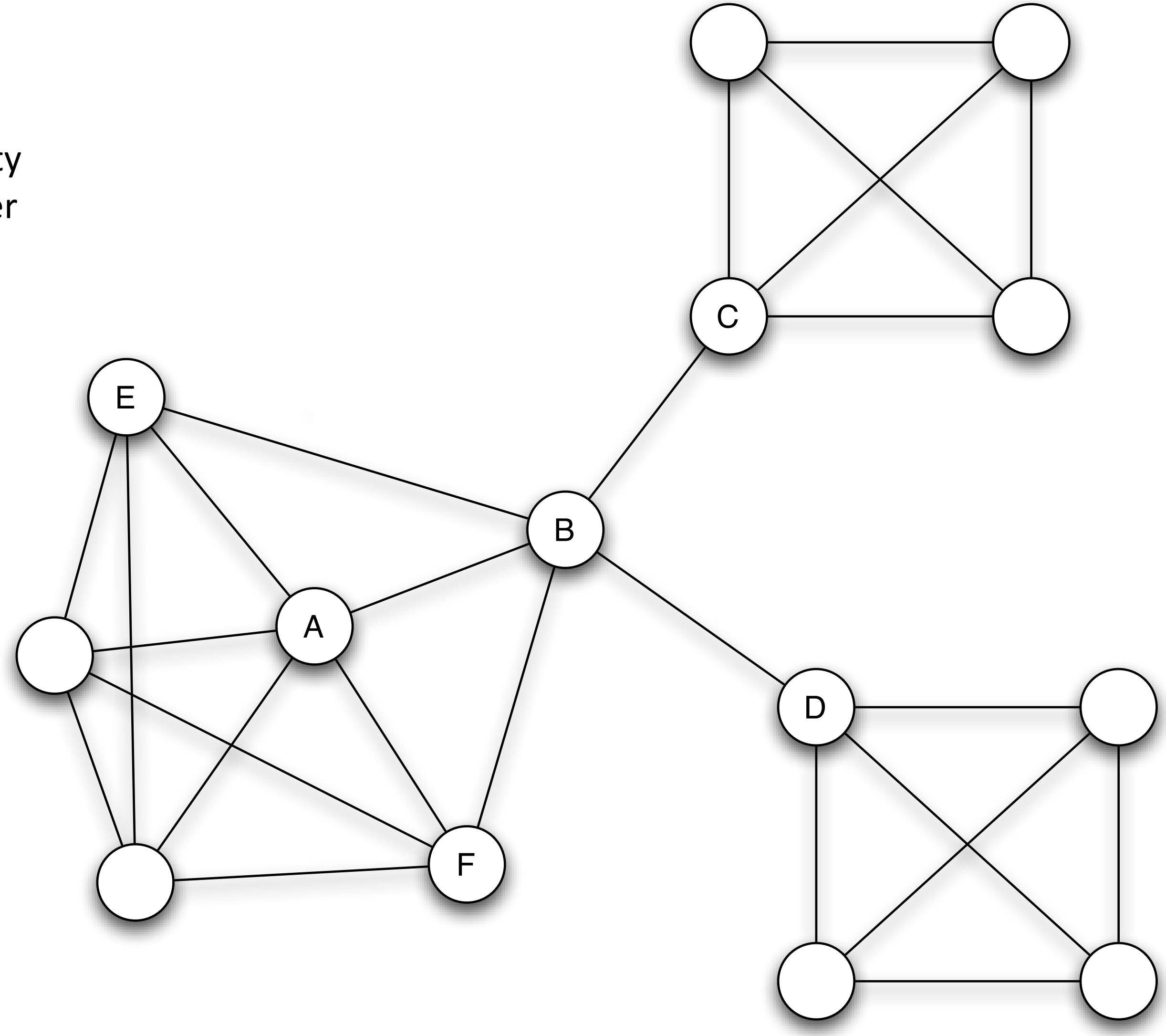
- Cross-staffed new internal projects
  - white papers, database development
- Established cross-selling sales goals
  - managers accountable for selling projects with both kinds of expertise (forced people to integrate their approaches to addressing client problems)
- New communication vehicles
  - project tracking db; weekly email update
- Personnel changes

9 Months Later



a contrasting experience: embeddedness and bridging

Bridging:
- riskier;
- Brokerage
- amplifies creativity
- Gatekeeper power



Embeddedness (# of common neighbours of edge):
- greater trust
- old, repackaged information?

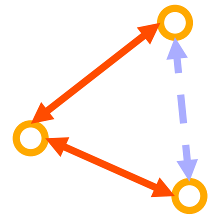
strength of weak ties

Granovetter thesis that, under many circumstances, strong ties are less useful than weak ties:

- interviewed people in Amherst, MA across professions to determine how they found out about their jobs;
- recorded whether they used social contacts and strength of the relationship;
- surprising proportion (~20%) of jobs were found through “weak ties”

Why?

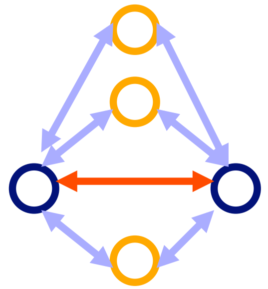
- individuals involved in weak ties less likely to overlap in their neighborhoods;
- weak ties form bridges across groups that have fewer connections to each other (plays role in disseminating information).
- weak ties hold communities together;



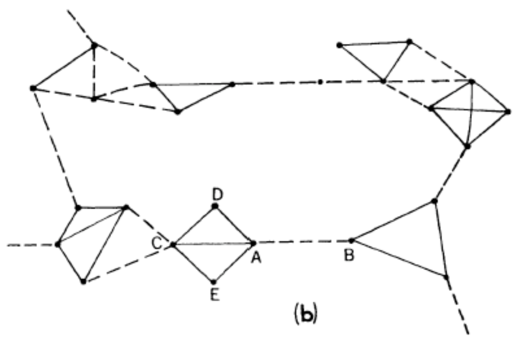
“forbidden triad”:

strong ties are likely to “close”

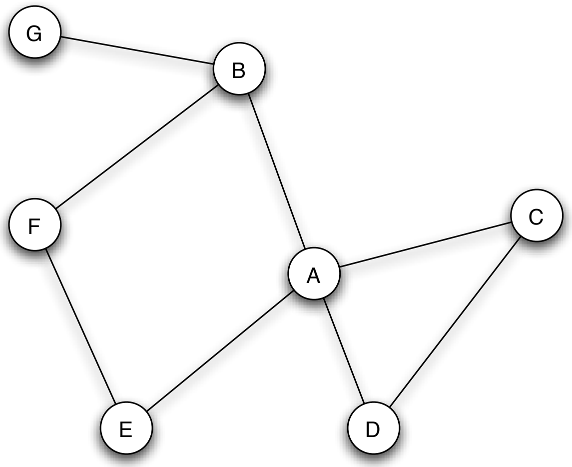
- A strong tie
  - frequent contact
  - affinity
  - many mutual contacts



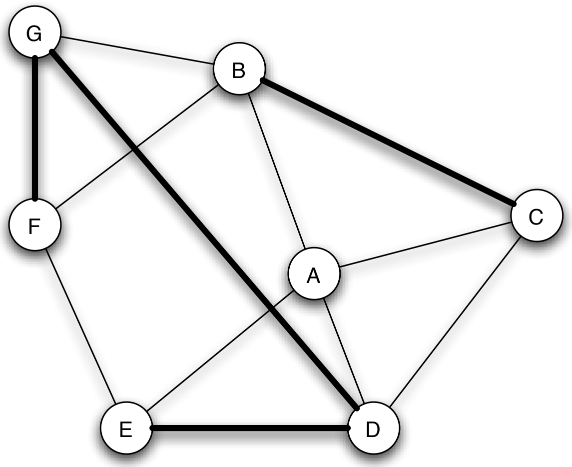
- Less likely to be a bridge (or a local bridge)



triadic closure, local bridges and weak ties



(a) *Before new edges form.*



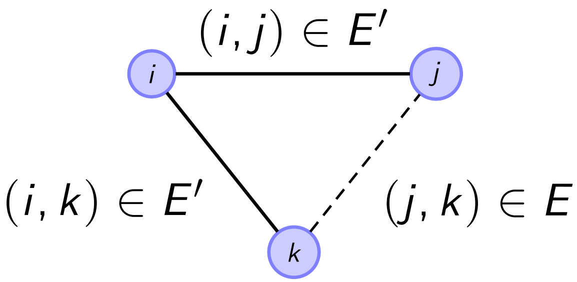
(b) *After new edges form.*

Why are we likely to observe a tie forming between B and C ?

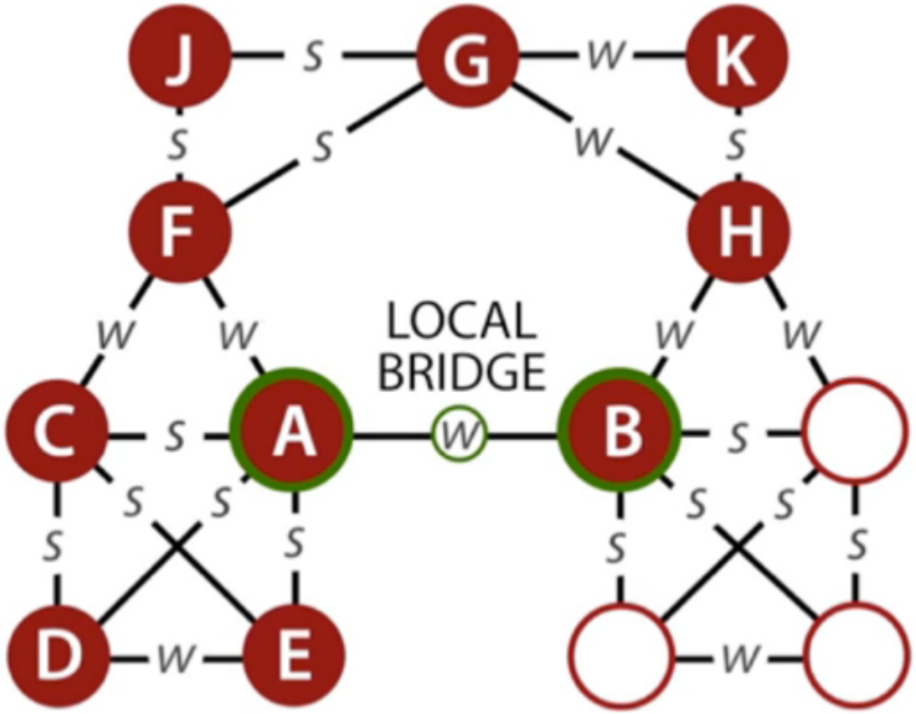
- Opportunity;
- Similarity;
- Incentive

Strong Triadic Closure Property

if $(i, j) \in E'$ and $(i, k) \in E'$, then $(j, k) \in E$.



Any local bridge will necessarily be a weak tie…
[**proof by contradiction**]



edges are either **embedded** or **bridging** (Social Capital)

“the ability of actors to secure benefits by virtue of membership in social networks or other social strucures”

Social capital is viewed as **property of a group** (favorable structures contribute to higher social capital) or as **property of an individual** (depends on position of the individual in the network). Different approaches highlight different aspects:

- **Coleman** values **embedded** edges (enable enforcement of norms, have reputational effects, enhance trusting mechanisms)
- **Burt** sees it as a tension between ***closure*** (as in Coleman’s embeddedness) and ***brokerage*** (ability to broker interactions between different groups).
- **Putnam** harmonizes both views when he discusses ***bonding*** *capital* and ***bridging*** *capital*.

“Social capital is at once the **resources contacts hold** and **the structure of contacts in a network**. The first term describes *whom* you reach. The second describes *how* you reach.”

1. **“Who you reach”** – network provides an actor with access to people with specific resoruces and functions as a conduit; establishes a correlation between your resources and theirs. Relates to concept of *power* and *prestige*.

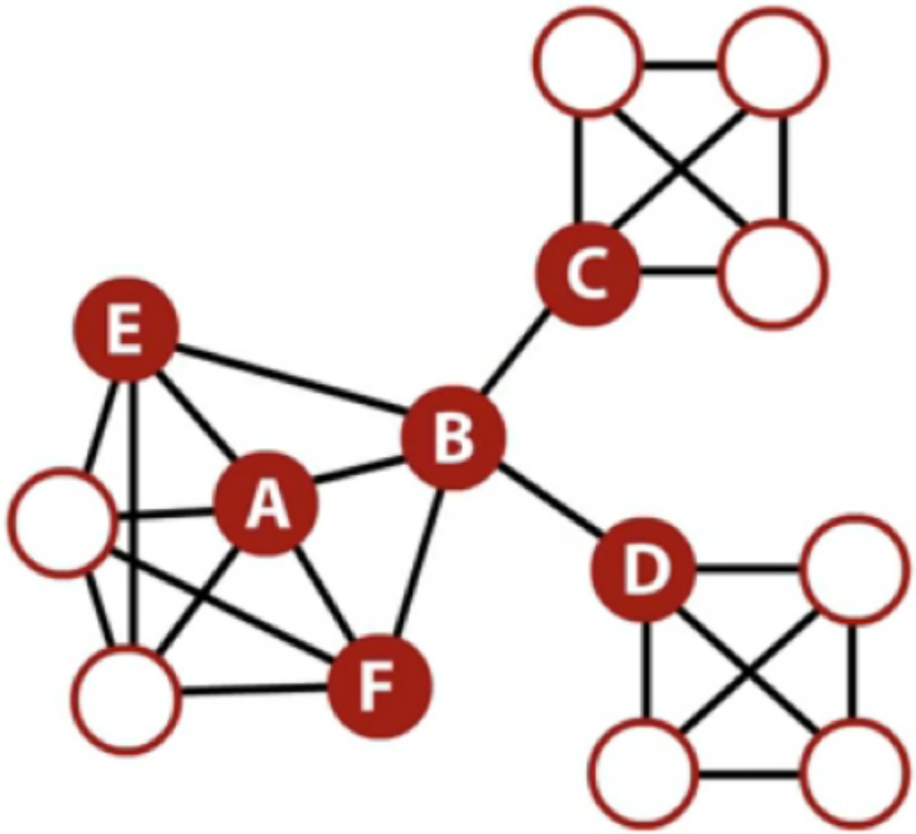
2. **“How you reach”** – social structure is capital itself that is meaured in terms of network **range** and **size**. The value of the “rate of return” can be boosted given the structure of the network and the location of the actor’s contacts within that structure. The benefits include: **information** and **control**.

Approaches to Social Capital

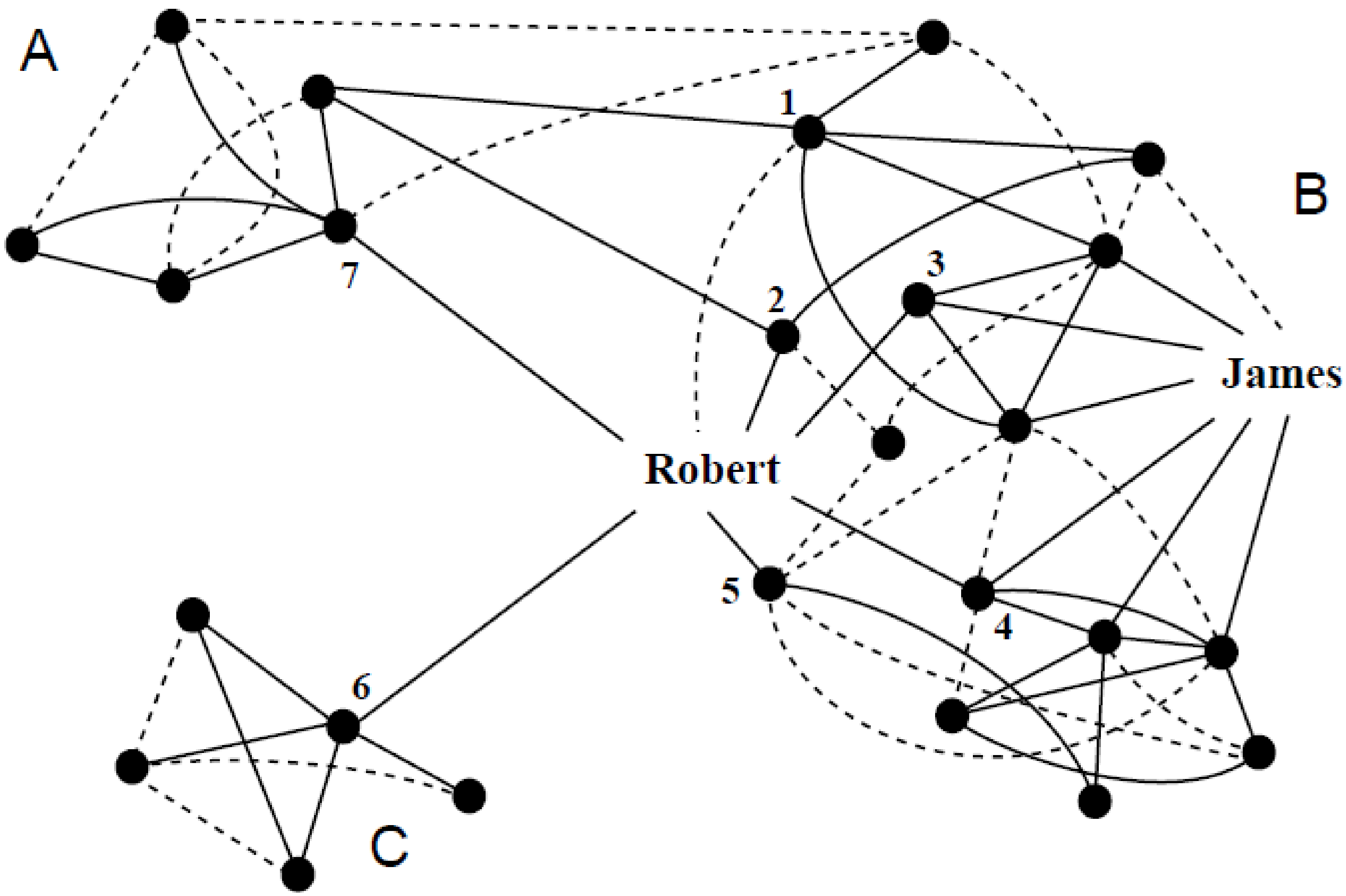
- Topological (shape-based)
  - Burt (structural holes)
  - Coleman, Putnam (connectivity/embeddedness)
- Connectionist (attribute-based)
  - Lin
- Combination of shape-based and attribute-based
  - Gould & Fernandez

Structural Holes (Ron Burt)

The distribution of bridging edges among the nodes is unequal in a network…



A and B have different sources of relative advantages. B spans structural holes in the network.



Who is better off Robert or James?

Burt study

- Managers asked to come up with an idea to improve the supply chain
- Then asked:
  - whom did you discuss the idea with?
  - whom do you discuss supply-chain issues with in general
  - do those contacts discuss ideas with one another?

  - 673 managers (455 (68%) completed the survey)
  - ~ 4000 relationships (edges)

Structural Holes (Ron Burt)

- Managers asked to come up with an idea to improve the supply chain
- Then asked:
  - whom did you discuss the idea with?
  - whom do you discuss supply-chain issues with in general
  - do those contacts discuss ideas with one another?

  - 673 managers (455 (68%) completed the survey)
  - ~ 4000 relationships (edges)

Hypotheses:

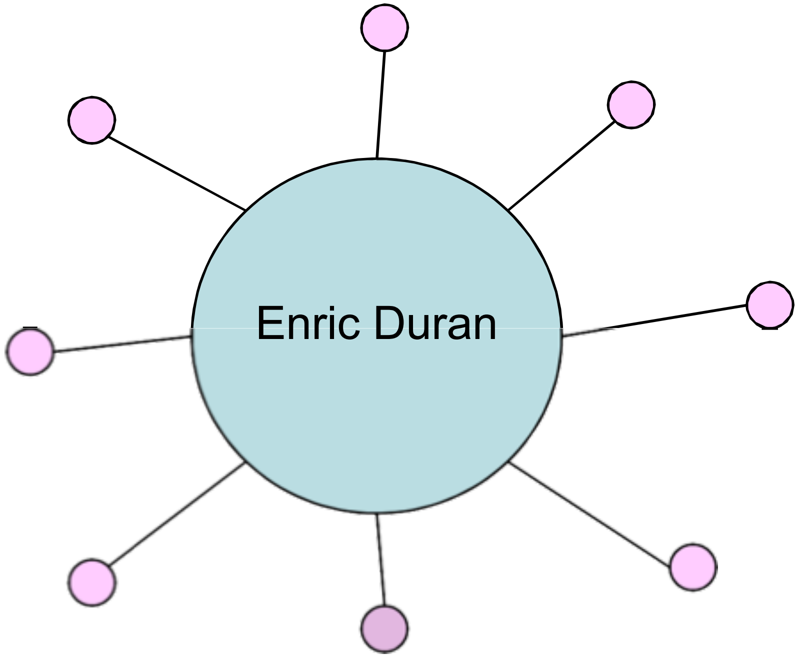
1. Opinions within groups are homogenous;
2. People who extend themselves across structural holes are exposed to new information
3. New ideas emerge from having diverse pool of options

after intervention...

results

- people whose networks bridge structural holes have
  - higher compensation
  - positive performance evaluations
  - more promotions
  - more good ideas
- these brokers are
  - more likely to express ideas
  - less likely to have their ideas dismissed by judges
  - more likely to have their ideas evaluated as valuable

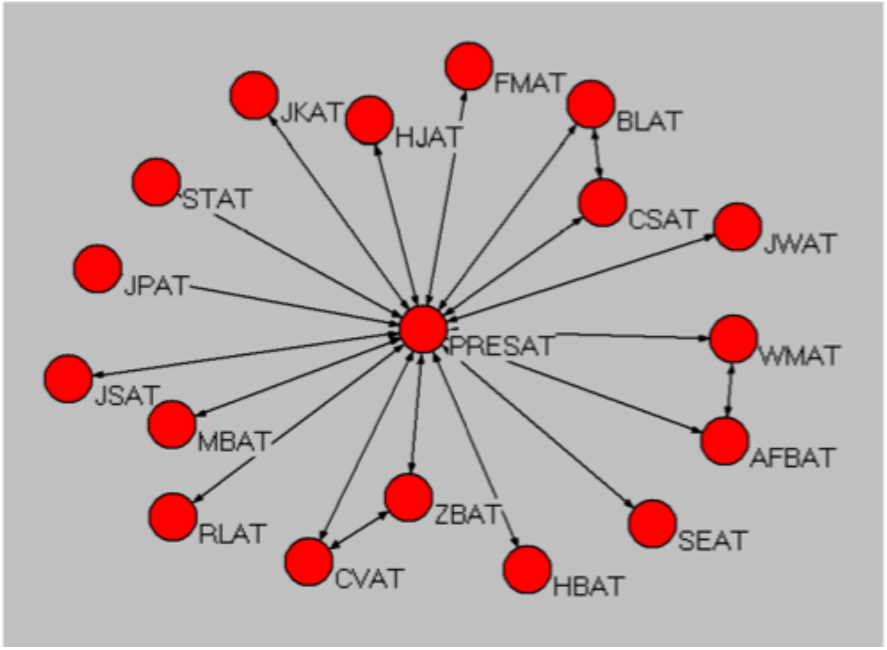
Structural Holes: the ego-net perspective



- Lack of ties among alters may benefit ego
- Benefits
  - Autonomy
  - Control
  - Information

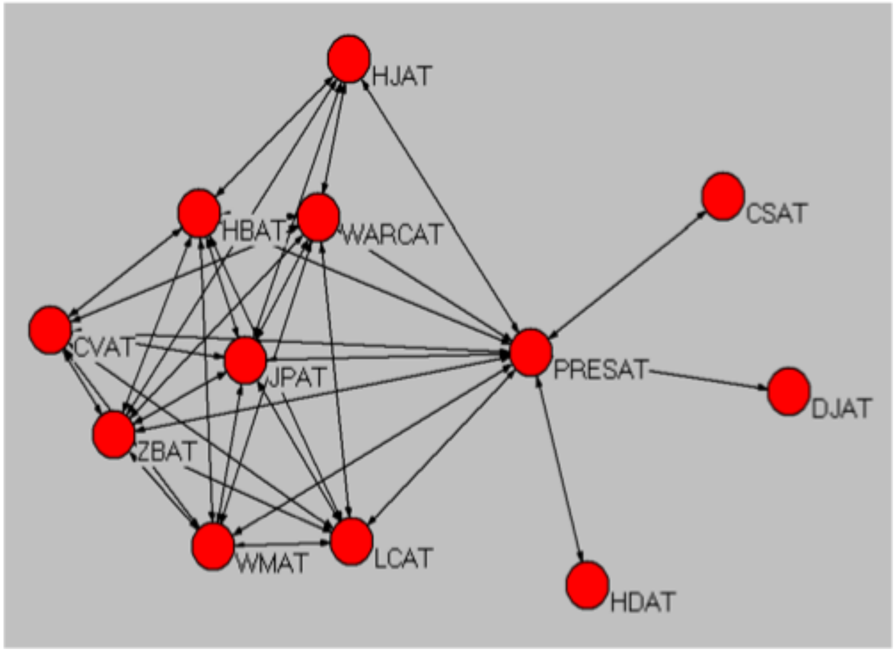
Control Benefits of Structural Holes

White House Diary Data, Carter Presidency



Year 1

Data courtesy of Michael Link

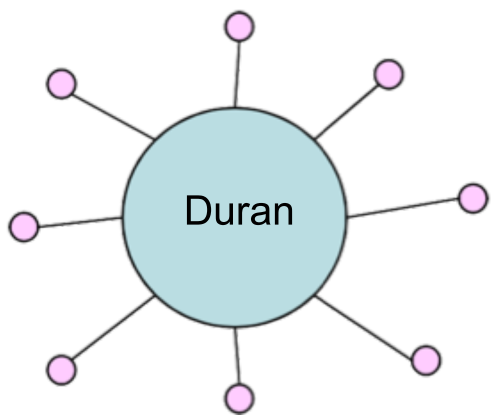


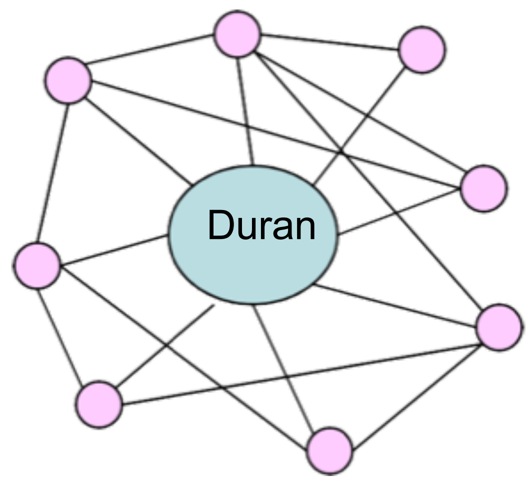
Year 4

Measures of Structural Holes

- **Effective size**;
- **Efficiency**;
- **Constraint**;
- **Hierarchy**;

Redundancy: dyadic redundancy calculates, for each actor in ego's neighborhood, how many of the other actors are also tied to the other. What % of Ego's network is redundant? Correlates with embeddedness.





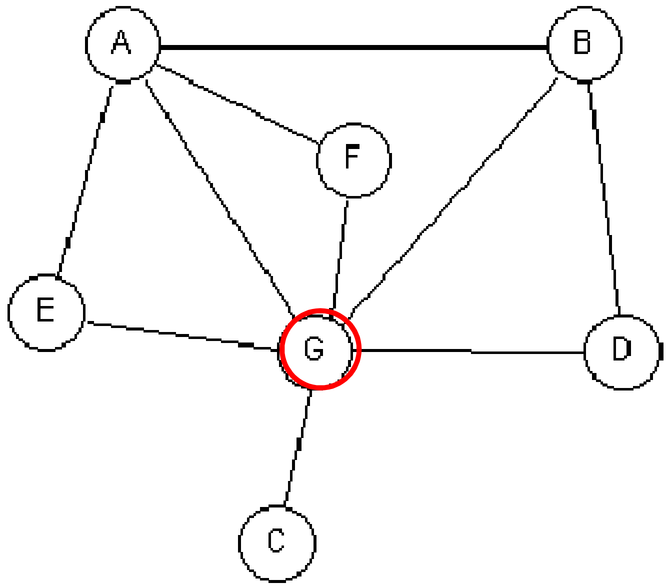
Effective Size

- Effective size is ego-network size (N) minus redundancy in network

$m_{jq} = \text{j's interaction with q divided by j's strongest relation with anyone}$
$p_{iq} = \text{proportion of i's energy invested in relation with q}$

$$ES_i = \sum_j \left[1 - \sum_q p_{iq} m_{jq} \right], \quad q \neq i, j$$
$$ES_i = \sum_j 1 - \sum_j \sum_q p_{iq} m_{jq}, \quad q \neq i, j$$

Effective Size



Node "G" is EGO	A	B	C	D	E	F	Total
Redundancy with EGO's other Alters:	3/6	2/6	0/6	1/6	1/6	1/6	1.33

Effective Size of G = Number of G's Alters – Sum of Redundancy of G's alters
= 6 – 1.33 = **4.67**

Effective Size formula

- M_{jq} = j's interaction with q divided by j's strongest tie with anyone
  - So this is always 1 if j has tie to q and 0 otherwise
- P_{iq} = proportion of i's energy invested in relationship with q
  - So this is a constant 1/N where N is ego's network size

$$ES_i = \sum_j \left[1 - \sum_q p_{iq} m_{jq} \right], \quad q \neq i, j$$

The quantity $p_{iq} m_{jq}$ is the level of redundancy between ego and a particular alter *j*.

$$ES_i = \sum_j \left[1 - \frac{1}{n} \sum_q m_{jq} \right], \quad q \neq i, j$$

$$ES_i = \sum_j 1 - \sum_j \frac{1}{n} \sum_q m_{jq}, \quad q \neq i, j$$

$$ES_i = n - \frac{1}{n} \sum_j \sum_q m_{jq}, \quad q \neq i, j$$

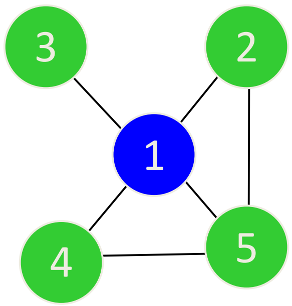
This is equivalent to n minus average degree:
n – 2t/n

Effective Size formula

Effective Size:

$$\sum_j \left[1 - \sum_q \boxed{p_{iq}} m_{jq} \right]$$

P_{iq} is the proportion of actor i's relations that are spent with q.



Adjacency

	1	2	3	4	5
1	**0**	**1**	**1**	**1**	**1**
2	**1**	**0**	**0**	**0**	**1**
3	**1**	**0**	**0**	**0**	**0**
4	**1**	**0**	**0**	**0**	**1**
5	**1**	**1**	**0**	**1**	**0**

P

	1	2	3	4	5
1	**.00**	**.25**	**.25**	**.25**	**.25**
2	**.50**	**.00**	**.00**	**.00**	**.50**
3	**1.0**	**.00**	**.00**	**.00**	**.00**
4	**.50**	**.00**	**.00**	**.00**	**.50**
5	**.33**	**.33**	**.00**	**.33**	**.00**

Effective Size formula

Effective Size:

$$\sum_j \left[1 - \sum_q p_{iq} \boxed{m_{jq}} \right]$$

m_{jq} is the marginal strength of contact j's relation with contact q. Which is j's interaction with q divided by j's strongest interaction with anyone. For a binary network, the strongest link is always 1 and thus m_{jq} reduces to 0 or 1 (whether j is connected to q or not - that is, the adjacency matrix).

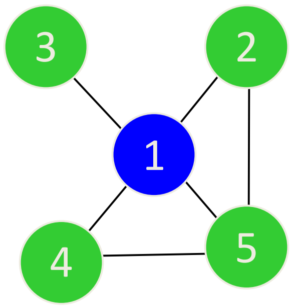
The sum of the product $p_{iq}m_{jq}$ measures the portion of i's relation with j that is redundant to i's relation with other primary contacts.

Effective Size formula

Effective Size:

$$\sum_j \left[1 - \boxed{\sum_q p_{iq} m_{jq}} \right]$$

Working with 1 as ego, we get the following <u>redundancy</u> levels:



P

	1	2	3	4	5
1	**.00**	**.25**	**.25**	**.25**	**.25**
2	**.50**	**.00**	**.00**	**.00**	**.50**
3	**1.0**	**.00**	**.00**	**.00**	**.00**
4	**.50**	**.00**	**.00**	**.00**	**.50**
5	**.33**	**.33**	**.00**	**.33**	**.00**

PM_{1jq}

	1	2	3	4	5
1	**---**	**---**	**---**	**---**	**---**
2	**---**	**.00**	**.00**	**.00**	**.25**
3	**---**	**.00**	**.00**	**.00**	**.00**
4	**---**	**.00**	**.00**	**.00**	**.25**
5	**---**	**.25**	**.00**	**.25**	**.00**

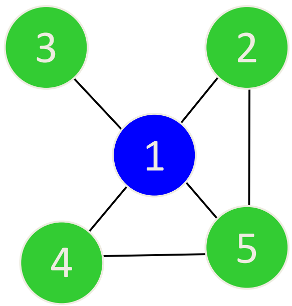
Sum=1, so

Effective size = 4-1 = 3.

Efficiency

Efficiency is the observed size divided by the observed size:

degree/effective size



Node	Size	Effective Size:	Efficiency
1	4	3	.75
2	2	1	.5
3	1	1	1.0
4	2	1	.5
5	3	1.67	.55

Constraint

refers to how much room one has to negotiate or exploit potential structural holes in the network

m_{jq}= j's interaction with q divided by j's strongest relationship with anyone So this is always 1 if j has tie to q and 0 otherwise

p_{iq} = proportion of i's energy invested in relationship with q
So this is a constant 1/N where N is network size

$$c_{ij} = p_{ij} + \sum_{q} p_{iq} m_{qj}\,,\ \ q \neq i, j$$

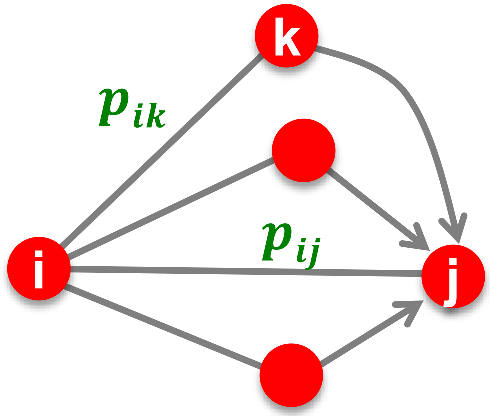
- Alter j constrains i to the extent that
  - i has invested in j
  - i has invested in people (q) who have invested heavily in j. That is, i's investment in q leads back to j.
- Even if i withdraws from j, everyone else in i's network is still invested in j

Constraint – crude idea

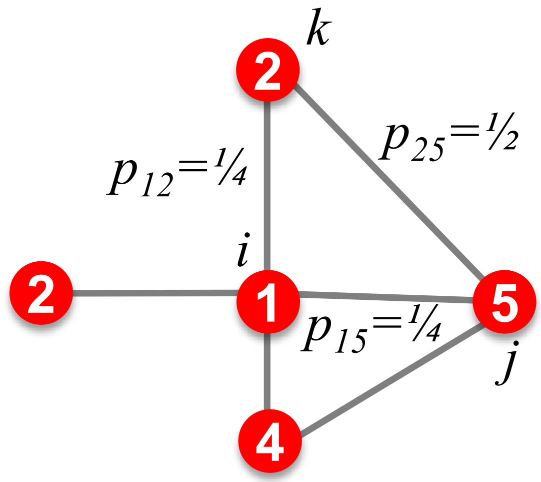
- Constraint is a summary measure that taps the extent to which ego's connections are to others who are connected to one another.
- If ego's potential trading partners all have one another as potential trading partners, ego is highly constrained. If ego's partners do not have other alternatives in the neighborhood, they cannot constrain ego's behavior. (Hanneman & Riddle, 2005)

Constraint – formula

- **To what extent are person's contacts redundant**



$$p_{uv} = 1/d_u$$



  - **Low**: disconnected contacts
  - **High**: contacts that are close or strongly tied

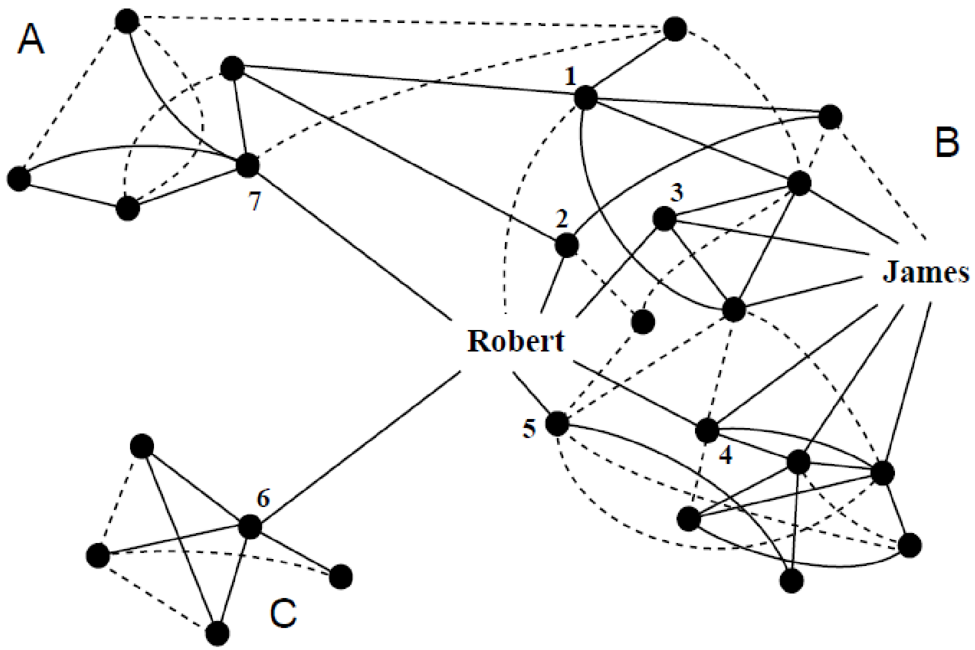
p_{uv}

	1	2	3	4	5
1	.00	.25	.25	.25	.25
2	.50	.00	.00	.00	.50
3	1.0	.00	.00	.00	.00
4	.50	.00	.00	.00	.50
5	.33	.33	.00	.33	.00

$$c_i = \sum_j c_{ij} = \sum_j \left[p_{ij} + \sum_k \left(p_{ik} p_{kj} \right) \right]^2$$

p_{uv} prop. of u's "energy" invested in relationship with v

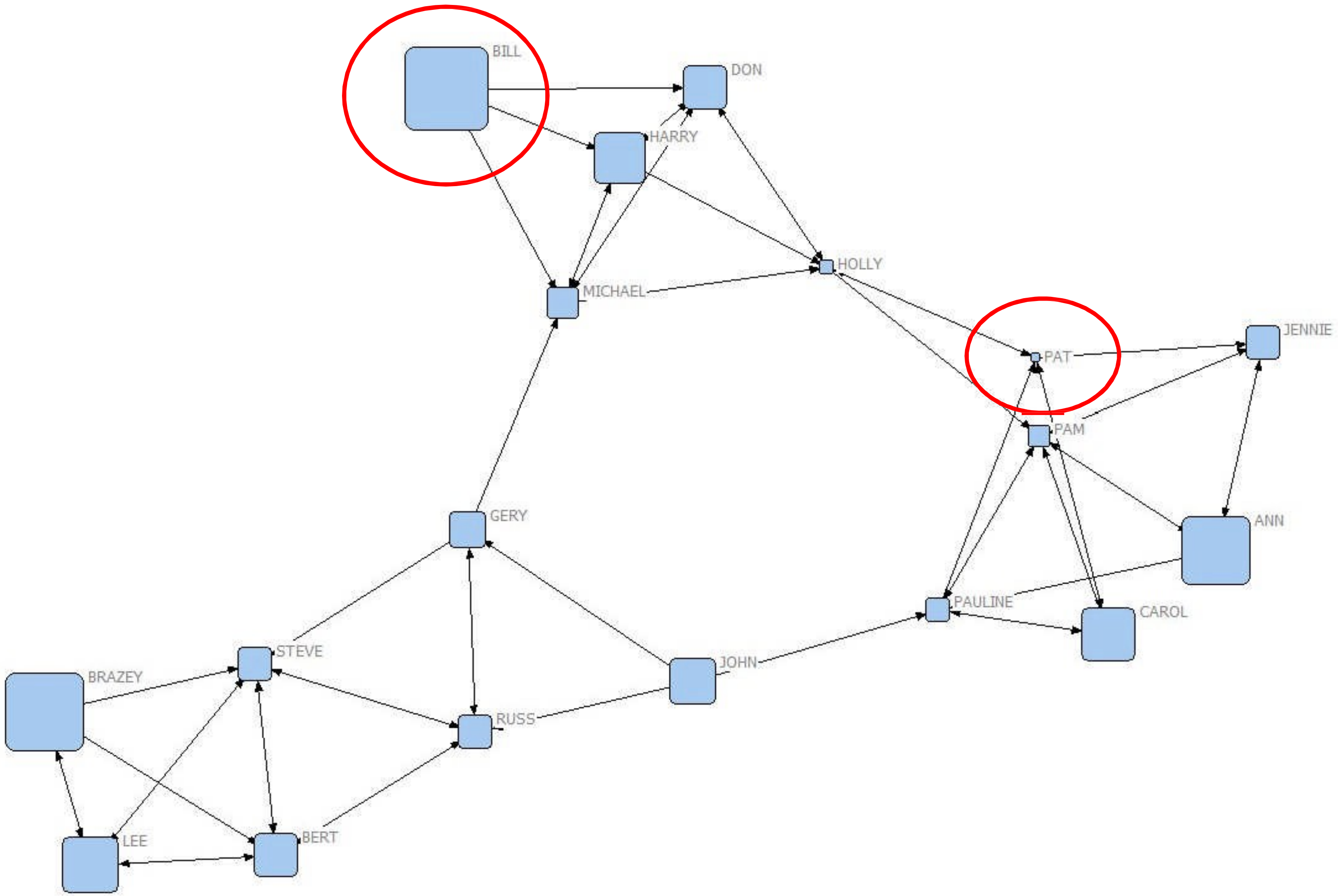
Constraint – formula



- **Network constraint:**
  - James: $c_J = 0.309$
  - Robert: $c_R = 0.148$

- **Constraint:** To what extent are person's contacts redundant
  - **Low**: disconnected contacts
  - **High**: contacts that are close or strongly tied

Sized by Constraint

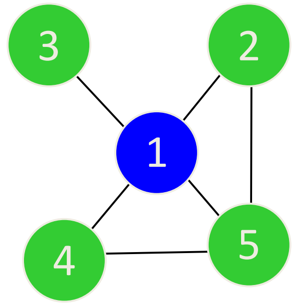


Hierarchy

Conceptually, hierarchy (for Burt) is really the extent to which **constraint** is concentrated in a single actor. It is calculated as:

$$H = \frac{\sum_j \left(\frac{C_{ij}}{C/N}\right) \ln\left(\frac{C_{ij}}{C/N}\right)}{N \ln(N)}$$

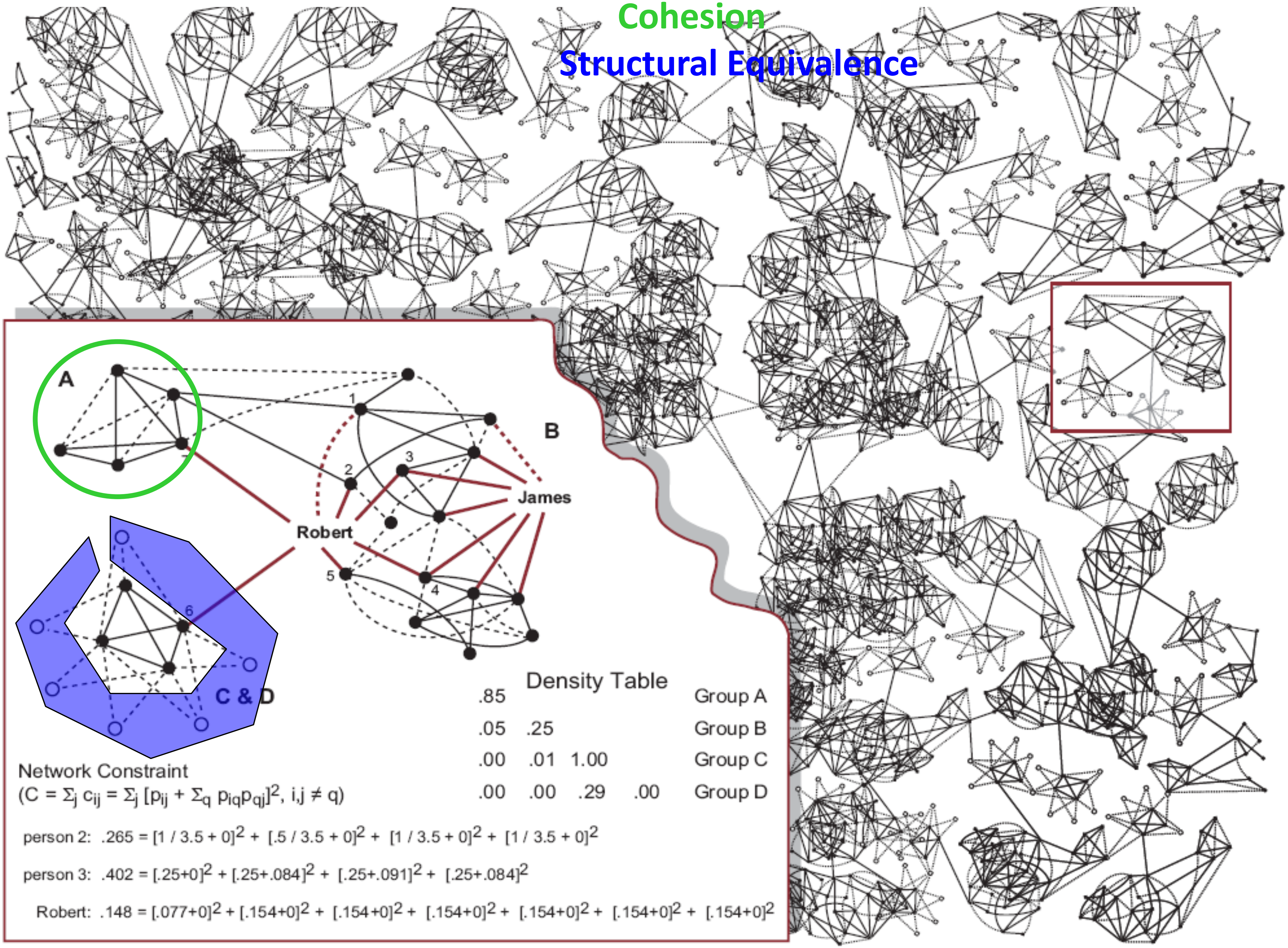
	2	3	4	5	C
C:	**0.11**	**0.06**	**0.11**	**.25**	**0.53**
$\left(\frac{C_{ij}}{C/N}\right)$	**.83**	**.46**	**.83**	**1.9**	



H=.514

Cohesion

Structural Equivalence



Density Table

.85				Group A
.05	.25			Group B
.00	.01	1.00		Group C
.00	.00	.29	.00	Group D

Network Constraint

$(C = \Sigma_j c_{ij} = \Sigma_j [p_{ij} + \Sigma_q p_{iq} p_{qj}]^2,\ i,j \neq q)$

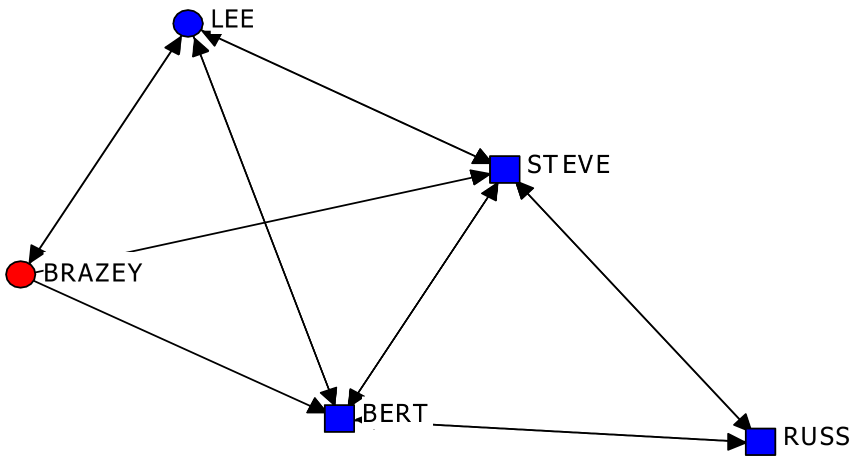
person 2: $.265 = [1 / 3.5 + 0]^2 + [.5 / 3.5 + 0]^2 + [1 / 3.5 + 0]^2 + [1 / 3.5 + 0]^2$

person 3: $.402 = [.25+0]^2 + [.25+.084]^2 + [.25+.091]^2 + [.25+.084]^2$

Robert: $.148 = [.077+0]^2 + [.154+0]^2 + [.154+0]^2 + [.154+0]^2 + [.154+0]^2 + [.154+0]^2 + [.154+0]^2$

Nan Lin (Social Resource Theory)

- Lin's view
  - Valued resources in societies represented by wealth, power and status;
  - Social capital is analysed by the amount or variety of such characteristics of others with whom an individual has ties to;
  - In short, it is the **attributes** of those you are connected to that matters.



•We can look at the composition of an ego-net in terms of **heterogeneity** in **attributes** of the alters.

Lin – social capital as assets in networks

- Variety of heterogeneity of resources – measures of **heterogeneity** in **attributes** of the alters.

Focus	Measurements	Indicators
Embedded resources	Network resources	Range of resources, best resources, variety of resources, composition (average resources), contact resources
	Contact statuses	Contacts' occupation, authority, sector
Network locations	Bridge to access to bridge	Structural hole, structural constraint
	Strength of tie	Network bridge, or intimacy, intensity, interaction & reciprocity

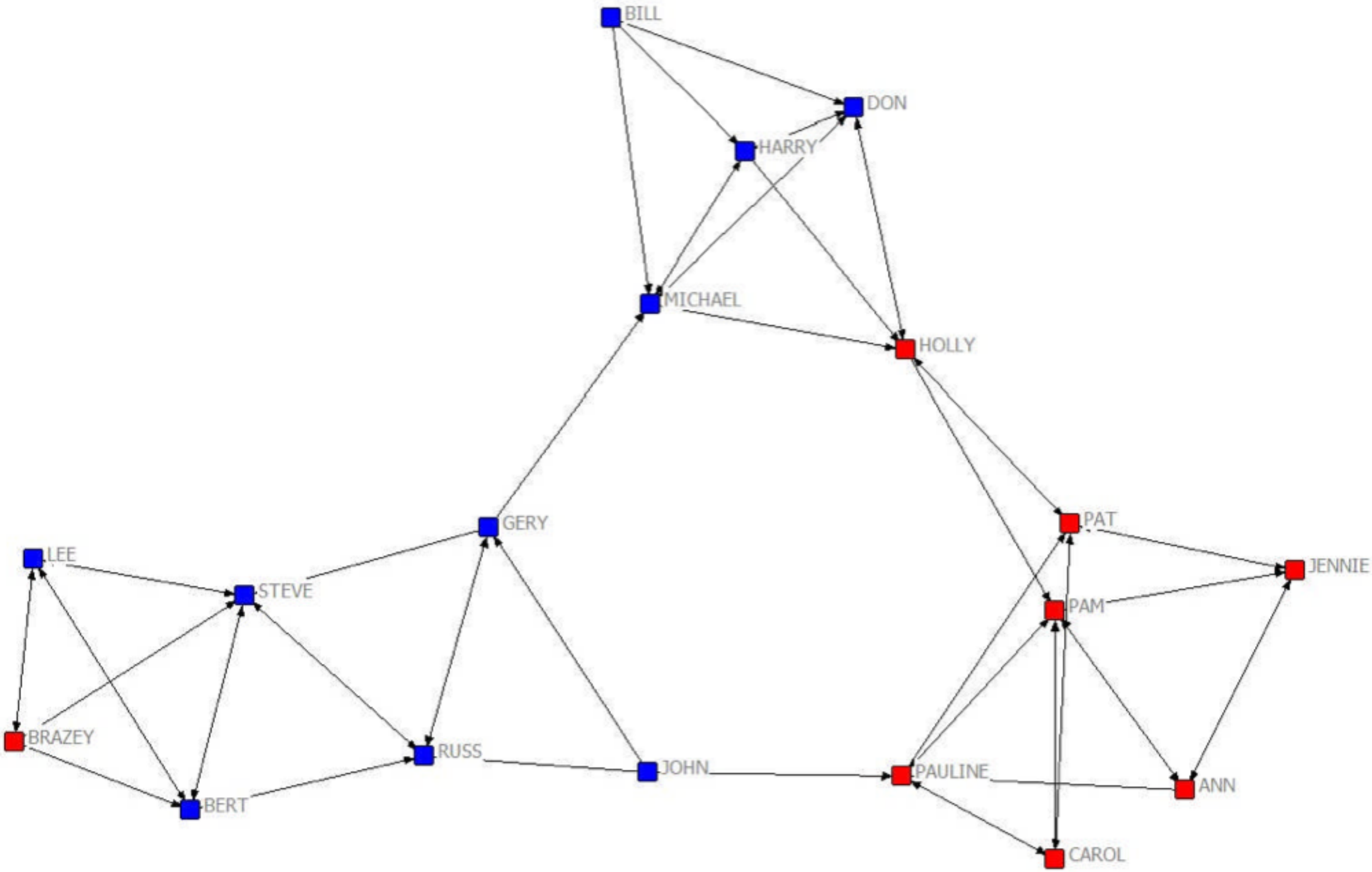
E-I Index

- Krackhardt and Stern

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

- E is number of ties between groups, I is number of ties within groups
- Varies between -1 (homophily) and +1 (heterophily)

Colored by Gender



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

	External	Internal	EI
HOLLY	3	2	0.2
BRAZEY	3	0	1
CAROL	0	3	-1
PAM	0	4	-1
PAT	0	3	-1
JENNIE	0	3	-1
PAULINE	1	4	-0.6
ANN	0	3	-1
MICHAEL	1	4	-0.6
BILL	0	3	-1
LEE	1	2	-0.333
DON	1	3	-0.5
JOHN	1	2	-0.333
HARRY	1	3	-0.5
GERY	0	4	-1
STEVE	1	4	-0.6
BERT	1	3	-0.5
RUSS	0	4	-1

Perfect heterophily

perfect homophily

Homophily & Heterogeneity

- **Homophily** is all about comparing EGO to the ALTERS
  - Complete homophily is a woman who has all women for friends
  - Heterophily is a man who has all women for friends
- **Heterogeneity** is about the diversity of ALTERS only.
  - Either man or a women with all men for friends has a Homogeneous network
  - But with half man and half women has heterogeneity

computing Heterogeneity

- **Blau Index**

$$\textbf{B} = 1 - \sum_{i=1}^{k} p_i^2,$$

  - Where where *pi* corresponds to the proportion of group members in *i*th category and *k* denotes the number of categories for an attribute of interest.
  - This index quantifies the probability that two members randomly selected from a population will be in different categories if the population size is infinite or if the sampling is carried out with replacement. Hence, **if *B* equals its minimum value** (i.e., zero), all members of the group are classified in the same category and there is no variety. In contrast, the **higher *B* is**, the more dispersed group members are over the categories.
  - Not comparable if number of categories is not identical across diversity variables

- **Index of Qualitative Variation**
  - Normalizes **B index** by dividing it by its maximum;
  - this controls for the number of categories and yields **<u>IQV index</u>** (Agresti & Agresti, 1978).

limitation of E-I index

It does not take into account the composition of the whole group:

a) people *i* connects to;

b) people *i* does not connect to.

Alternative measure: **Yule's Q**, to assess the degree to which ties (and non-ties) tend to correspond with being similar or different. 0 = no association; -1,+1 strong association.

$$Q = \frac{ad - bc}{ad + bc}$$

	Same	Different
Tie	a	b
No Tie	c	d

Alter-Ego Similarity

Ego	Same	Different
Tie	9	1
No Tie	27	3

Yule's Q

- Example: Gender and Political Party Affiliation

	Women	Men
Dem	a 27	b 10
Rep	c 16	d 15

Calculate "bc"

bc = (10)(16) = 160

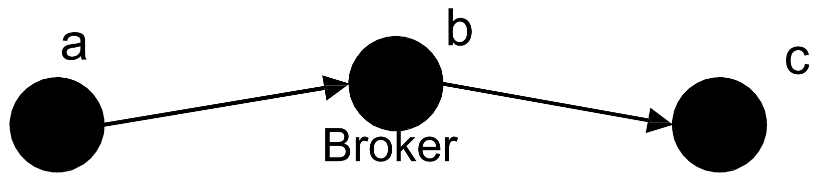
Calculate "ad"

ad = (27)(15) = 405

$$Q = \frac{bc - ad}{bc + ad} = \frac{160 - 405}{160 + 405} = \frac{-245}{505} = -.48$$

- -.48 = "weak association", almost "moderate"

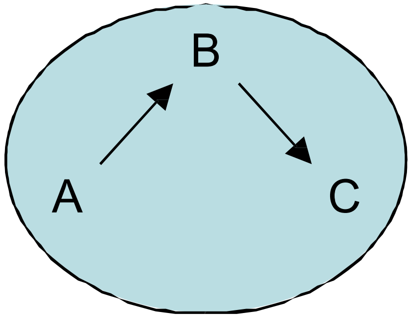
Brokerage Roles



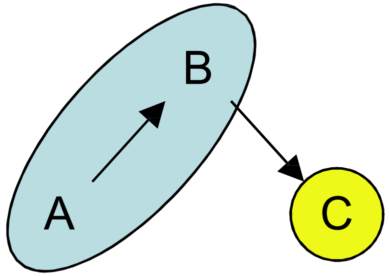
- Gould & Fernandez
- Broker is middle node of directed triad (note: a is NOT connected to c)
- What if nodes belong to different organizations?

Brokerage Roles

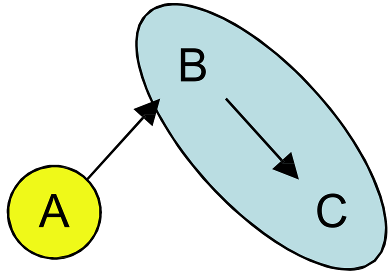
(with respect to B)



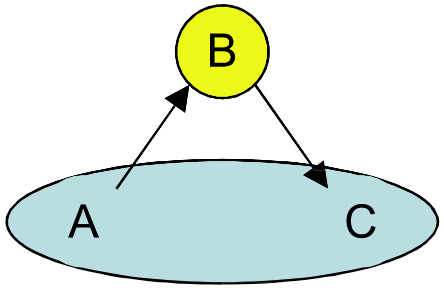
Coordinator



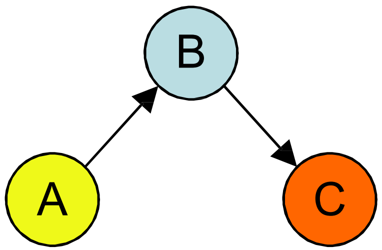
Representative



Gatekeeper



Consultant

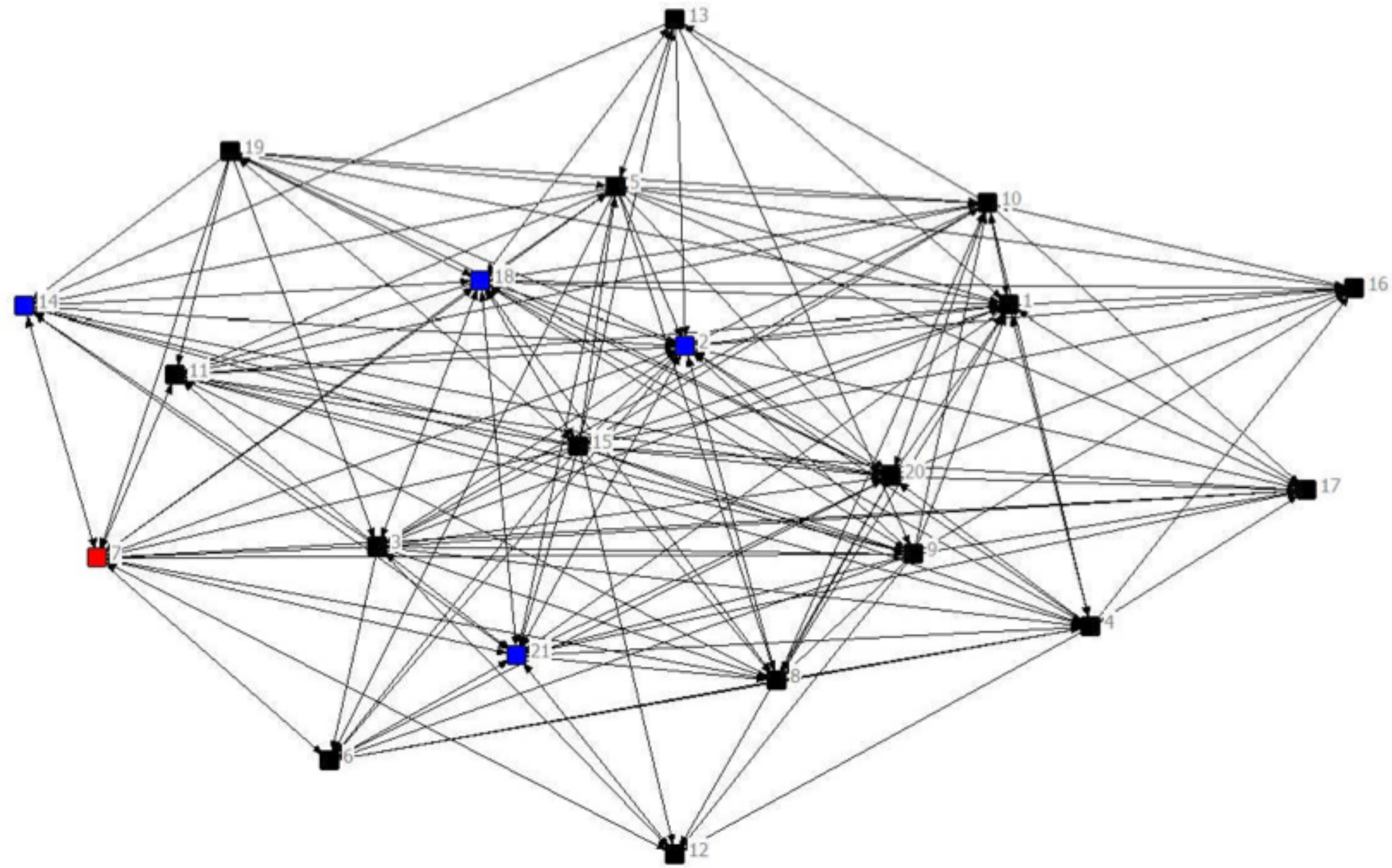


Liaison

- We can count how often a node enacts each kind of brokerage role

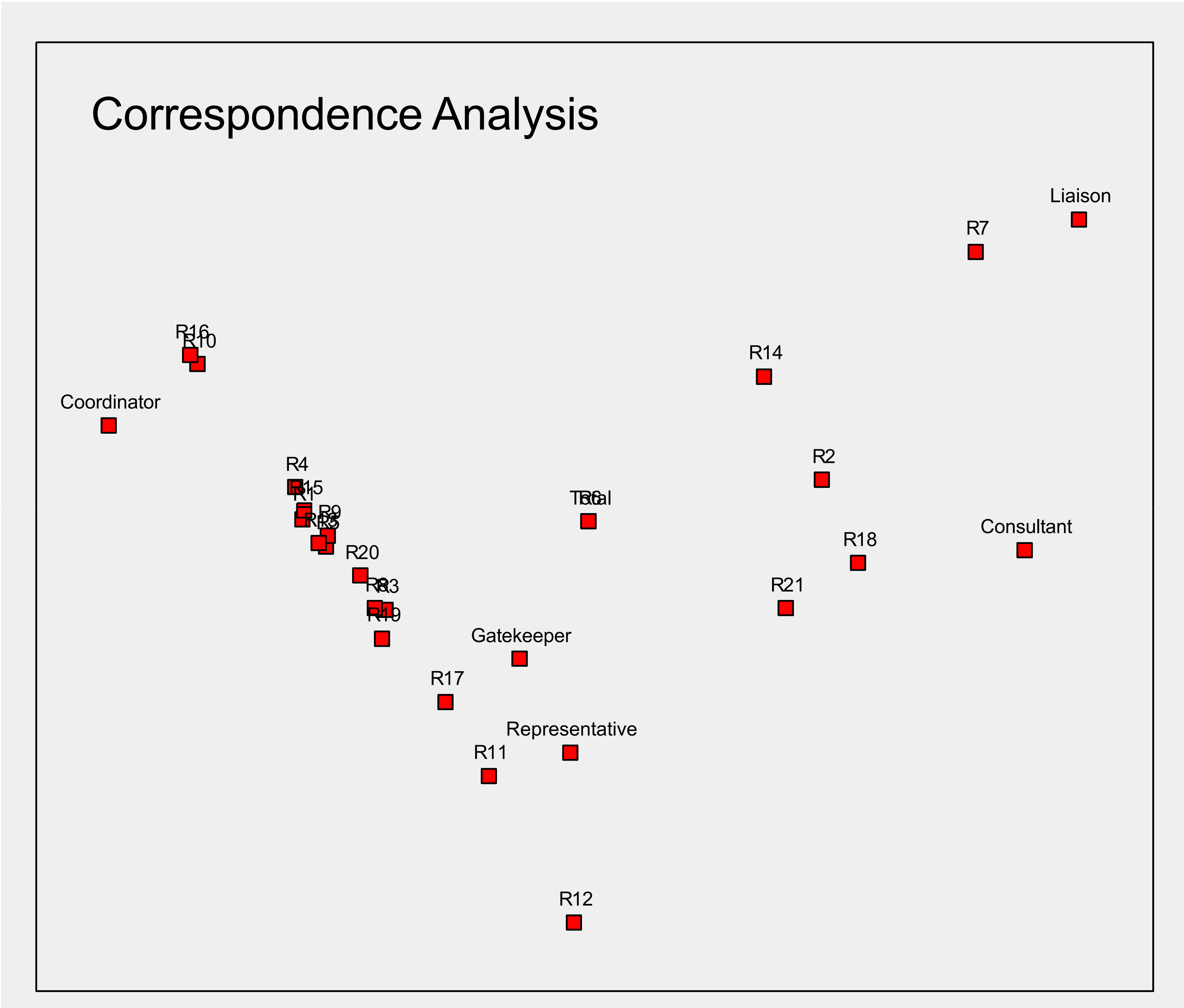
Advice Network:

Nodes Colored by Level (CEO / Manager/ Line Staff)



Counting of Role Structures

ID	Coordinator	Gatekeeper	Representative	Consultant	Liaison
7 (CEO)	0	0	0	17	21
21(Mgr)	2	11	16	35	8
18(Mgr)	0	9	22	72	18
14(Mgr)	0	2	0	0	2
2	0	5	2	7	6
6	0	0	0	0	0
5	14	2	6	0	0
3	9	7	4	0	0
9	8	3	2	0	0
10	44	1	0	0	0
1	17	0	7	0	0
12	0	0	2	0	0
13	2	0	1	0	0
4	21	7	2	0	0
15	18	3	5	0	0
16	2	0	0	0	0
17	3	3	4	0	0
8	8	3	5	0	0
19	2	0	2	0	0
20	12	7	4	0	0
11	1	1	3	0	0


Correspondence Analysis
Liaison
R7
R16
R10
Coordinator
R14
R2
R4
R15
R9
Consultant
R18
R20
R21
Gatekeeper
R17
Representative
R11
R12

SUMMARY

Name:	Description:	Relation to Social Capital:
Effective Size (Burt, 1992)	The number of alters, weighted by strength of tie, that an ego is directly connected to, minus a "redundancy" factor.	Positive. The more different regions of the network an actor has ties with, the greater the potential information and control benefits.
Constraint (Burt,1992)	The extent to which all of ego's relational investments directly or indirectly involve a single alter	Negative. The more constrained the actor, the fewer opportunities for action.
Compositional Quality (e.g., Lin)	The number of alters with high levels of needed characteristics (e.g., total wealth or power or expertise or generosity of alters)	Positive. The more connected to useful others, the more social capital.
Heterogeneity (e.g., Burt, 1983)	The variety of alters with respect to relevant dimensions (e.g., sex, age, race, occupation, talents).	Positive (except when it conflicts with compositional quality)
Brokerage Roles (Gould & Fernandez, 1989)	There are different roles that ego can play depending on network structure and composition	Depends on the situation

Equivalences

Imagine a hotel employee serving drinks at the general convention of the Episcopal church. All the delegates are in casual clothes, and at first he finds it difficult to identify the people who hold the most influential positions within the church. Eventually he notices that a group of delegates is treated with deference by everyone—call them the archbishops. Another group is treated with deference by everyone except the archbishops call them the bishops. Just by observing the delegates mingle, he might be able to guess the seniority of the office each person holds.

. . . by looking at a set of relationships within a community, we might discover that we can divide them in groups where people in the same group behave in a similar way with people of the other groups. In SNA, we say that people in these groups hold the same ***role***.

The Dream

- Formalizing hallowed notions of position, role and structure
- Society as concrete network of relationships among individuals
  - And social structure is underlying network of positions structuring observed pattern among individuals
- Role freed from essentialist and culturalist definitions and defined in terms of characteristic relations among incumbents of positions, often reciprocally defined
  - Like functional role of species in ecosystem

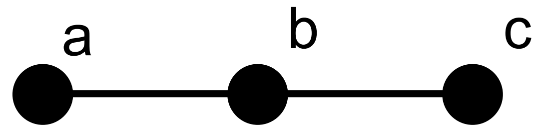
Positional Perspective

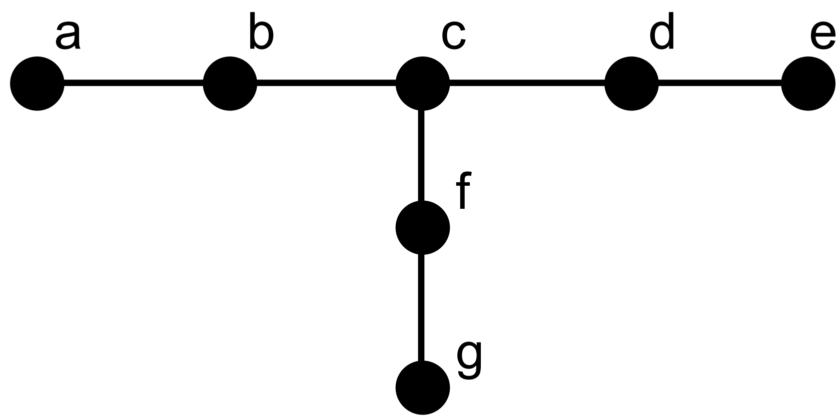
- Centrality measures one aspect of position
- But there are other aspects
  - Creating groupings of classes of nodes based on similar patterns of ties within the groups
  - Identified from their relational patterns to the groups being defined
  - Unlike the cohesive perspective where groups were based on specific ties, these are based on generalized or abstracted inter-group ties

Experimental Exchange Nets

Experimental exchange networks is an area of study in which subjects are brought into a lab and asked to play a game in which they must try to get as many points as possible. In each round, they are given 24 points to divide up with an exchange partner. The experimenter arranges the people into networks of who can exchange with whom. Initially, it was thought that centrality would determine point-getting, but this is not true.
How can we get there?

- Divvy up 24 points
- Who has what kinds of outcomes?





Implicit Hypothesis

- **Structurally similar nodes have similar outcomes**
  - Occupy same position, then same results
  - This is the distinct from Burt's brokerage
    - By occupying a position others do not, I get a competitive advantage

- **Networks with similar structures will also have similar outcomes**
  - Similarly structured teams will have similar performance outcomes

- **Role equates to Position** – we can express *role* through a relation (or set of relations). We only need to identify particular aspects of a positions.
  - *WHAT aspects?*

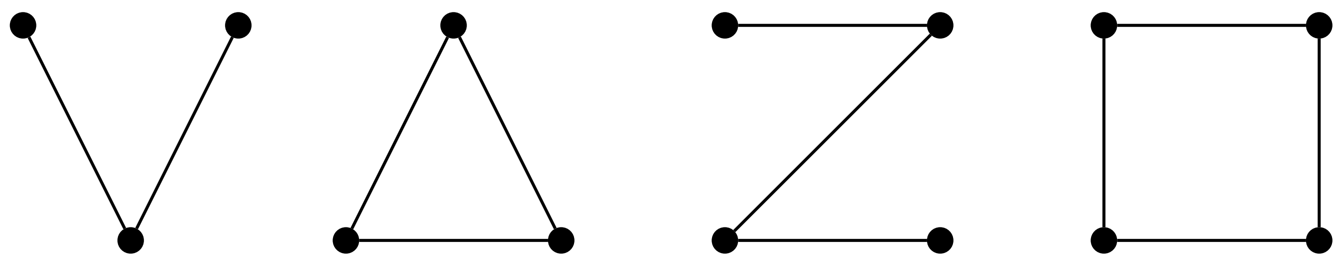
Some definitions

- Equivalences use some terms we have not yet encountered:
  - **Colorations**
  - **Neighborhoods**

When studying any type of mathematical object it is often useful to have some notion for when two objects of this type are equal. For example, when dealing with fractions we want to be able to say things like $\frac{6}{3} = 2 \neq \frac{7}{5}$. What allows us this freedom was the following condition:

$$\frac{a}{b} = \frac{c}{d} \iff ad = bc. \tag{1}$$

Since there is no shortage of different ways to draw a particular graph, we'd like to come up with a similar type of condition which would allow us to say if two graphs are the "same" or "different." On an intuitive level it may seem clear that the four graphs drawn below are "different," but it is not clear how do we go about turning this intuition into mathematics with the hope of obtaining sometype of a checkable procedure akin to (1).

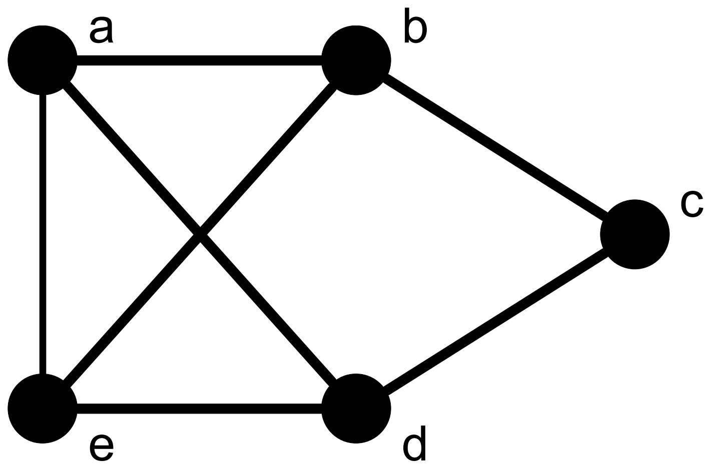


Colorations

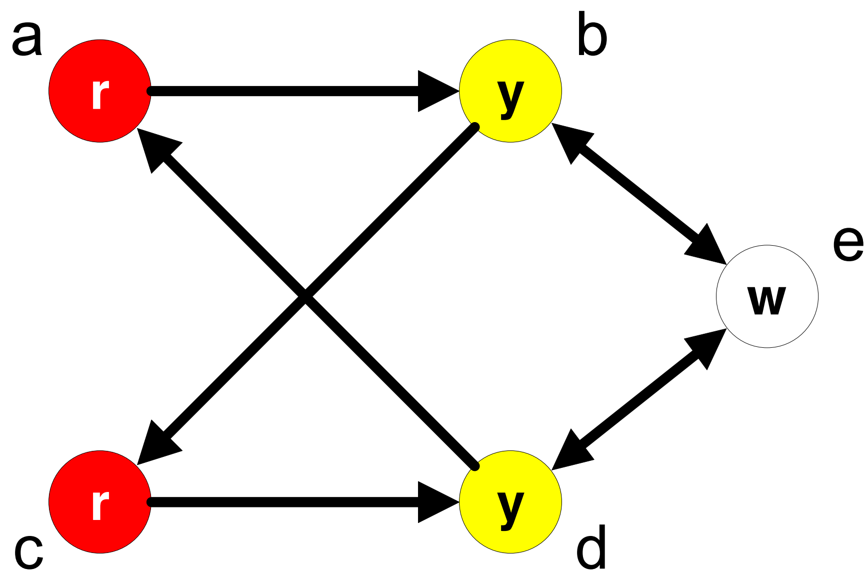
- A coloration C is just a partition of nodes. The color of a node v, written C(v) is just the equivalence class it belongs to;
- An equivalence is just the relation E induced by a partition;
- Different equivalences (E) result in different colorations (partitionings) based on different rules.
  - Structural
  - Automorphic
  - Regular

Neighborhoods

- Neighborhood of v, written N(v) is just the set of nodes adjacent to v.
- In digraphs, have
  - In-neighborhood $N^i(v)$: nodes sending arcs to v
  - Out-neighborhood $N^o(v)$: Nodes receiving arcs from v

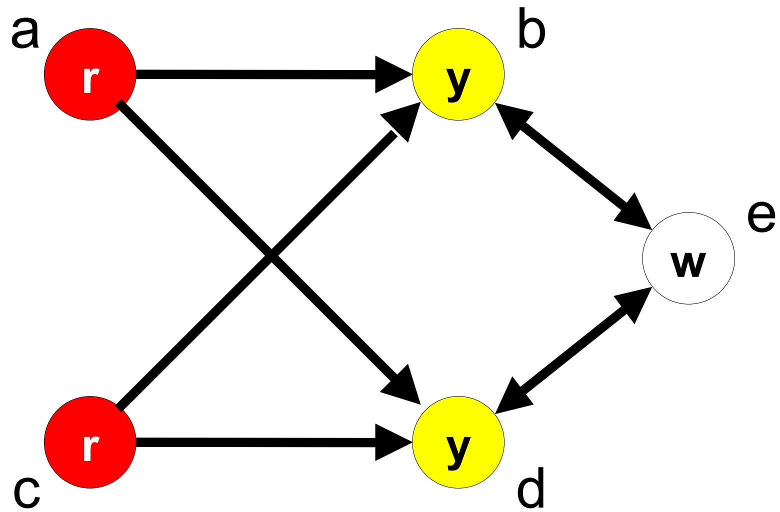


Coloration



Structural Equivalence

- u ≡ v if, for any w, whenever u→w then v→w, and whenever w→u then w→v

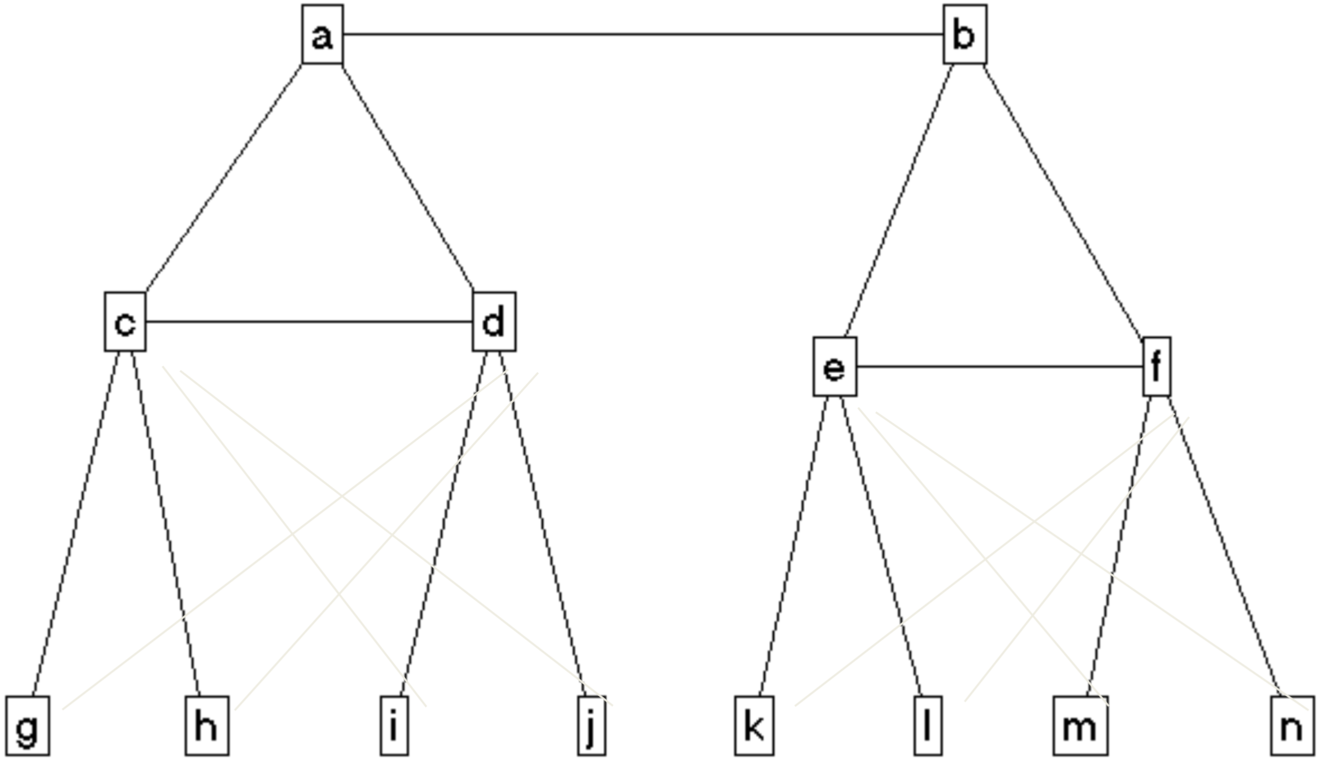


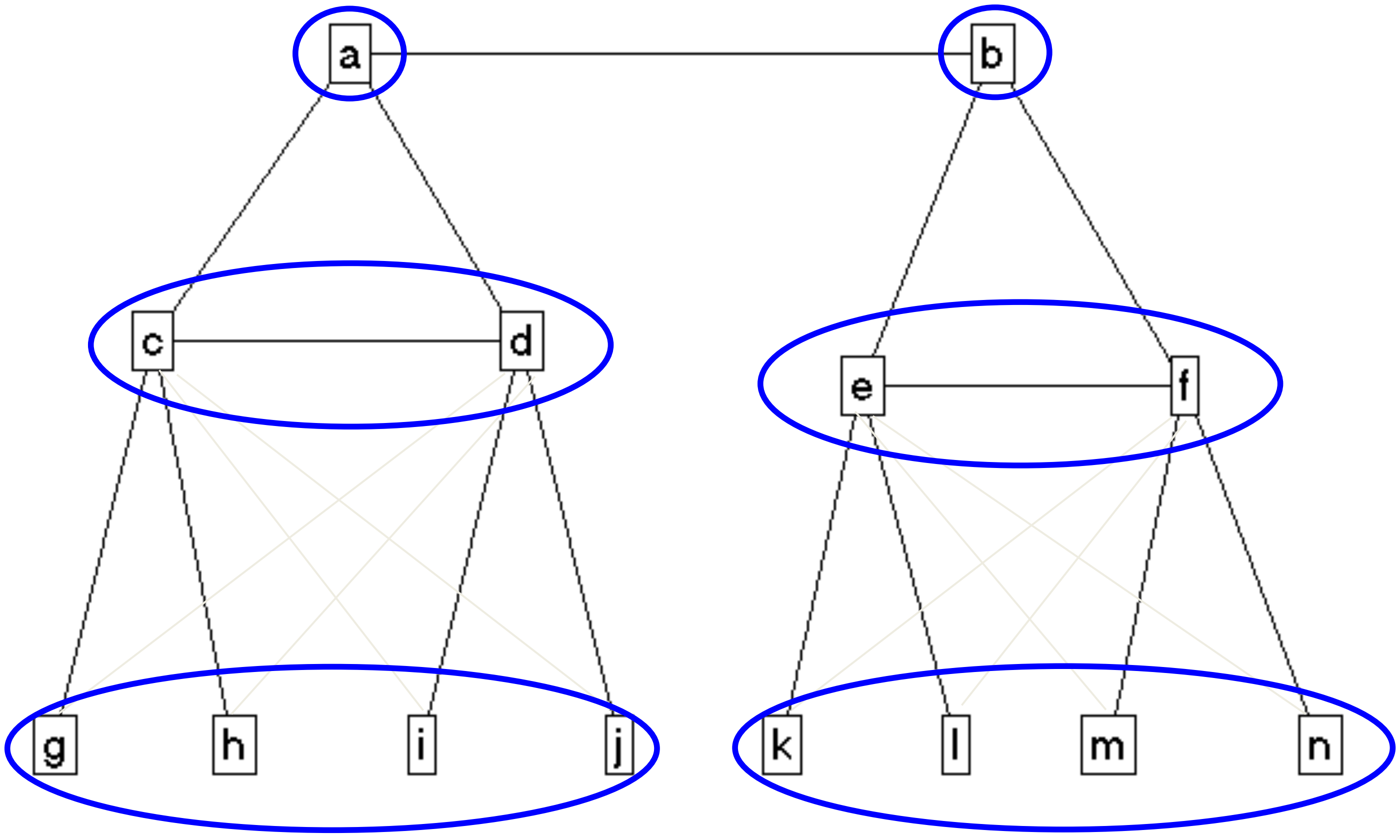
Note: Equivalent nodes have been colored the same.

- C(u) = C(v) if N(u) = N(v)
- C(u) = C(v) if $N^{out}(u) = N^{out}(v)$ and $N^{in}(u) = N^{in}(v)$

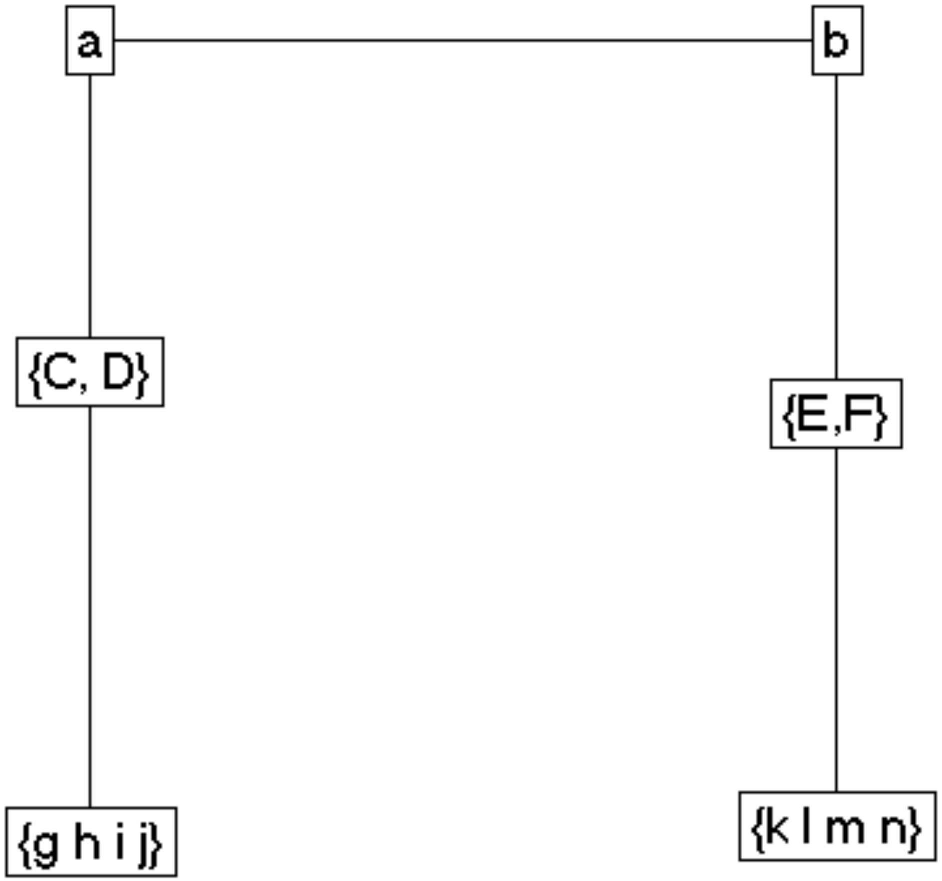
Structural Equivalence

- Structurally indistinguishable
  - Same degree, centrality, belong to same number of cliques, etc.
  - Only the label on the nodes themselves can distinguish them from those equiv to it.
  - Perfectly substitutable: same contacts, resources
  - Might identify who would take over if someone left
- Face the same social environment
  - Same forces affecting them
  - Expect, therefore, same outcomes


a
b
c
d
e
f
g
h
i
j
k
l
m
n


a
b
c
d
e
f
g
h
i
j
k
l
m
n

reducing graph to positions

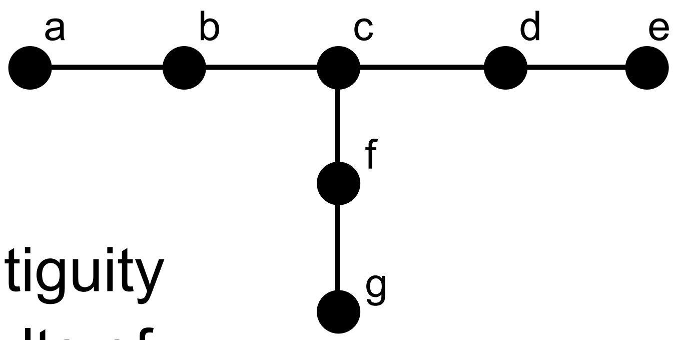


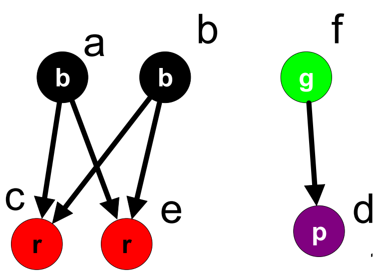
Issues with Structural Equivalence

- Location AND position
  - You are your friends
  - Confounds location with role
- In real data, few perfectly structural equivalent nodes
  - So we often calculate the degree of structural equivalence between nodes

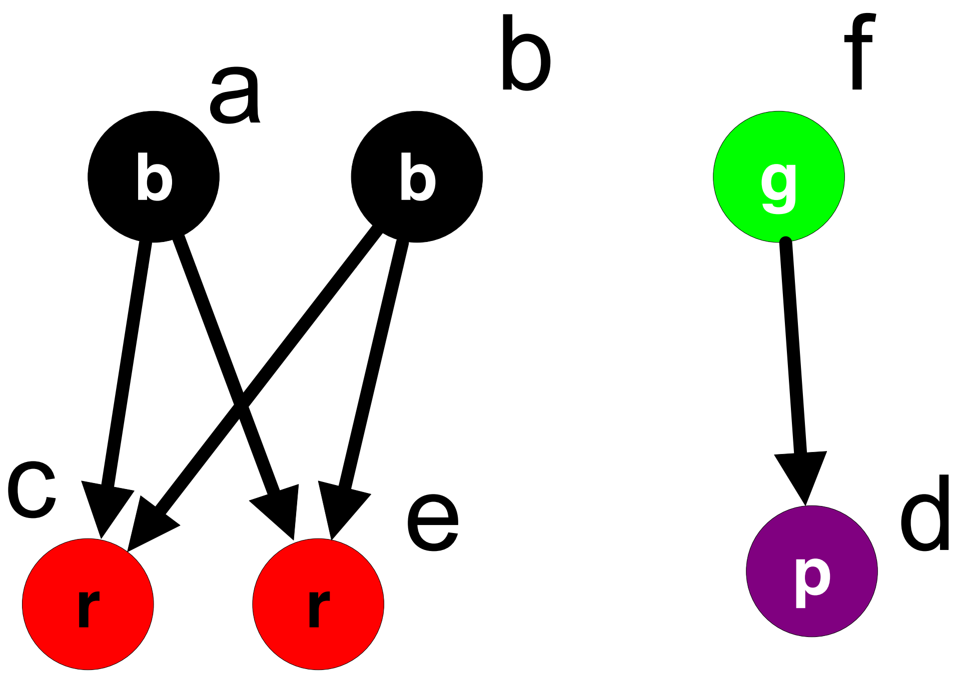
Structural Equivalence

- Pros
  - Captures notions like niche
  - Location or position
    - You are your friends
- Cons
  - Confounds similarity with contiguity
  - Not helpful for explaining results of exchange experiments
  - Not a good formalization of social role
    - Mother & father play same role to their kids, but not other parents
    - Can't use in disconnected graphs





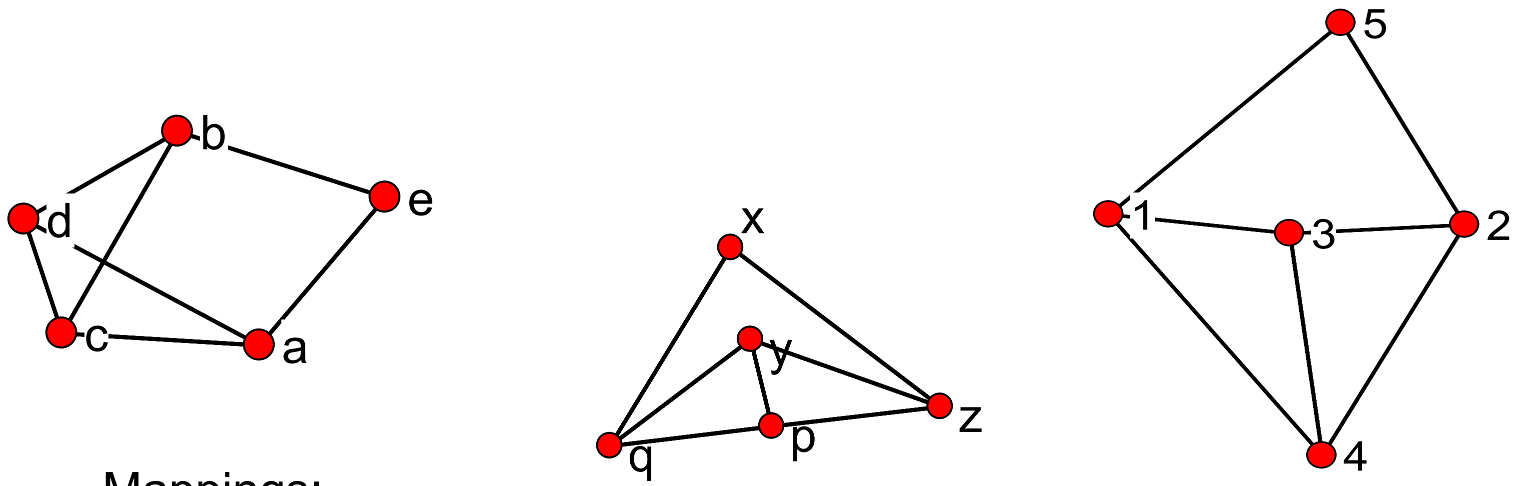
Only parents of same children are playing the same role



Automorphic Equivalence

- Aka structural isomorphism
- Two graphs G(V,E) and G'(V',E') are same if you can find a 1:1 mapping of nodes of one to the other that preserves adjacency structure
  - The mapping p is an isomorphism if $(u,v) \in E$ iff $(p(u),p(v)) \in E'$
  - P is called an automorphism when G=G'
    - Automorphisms also called symmetries of a graph

Isomorphisms



Mappings:

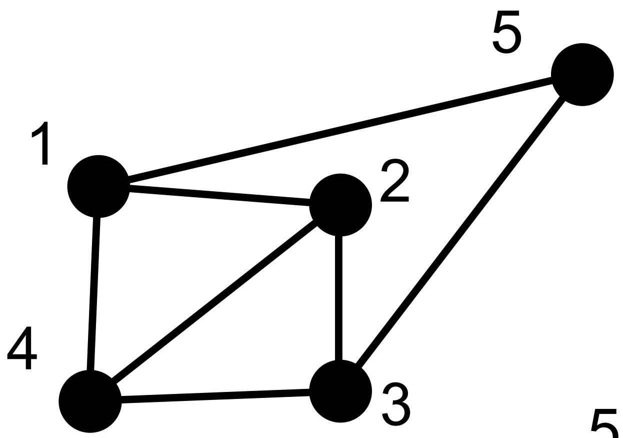
Fig 1	Fig 2	Fig 3
a	q	1
b	z	2
c	y	3
d	p	4
e	x	5

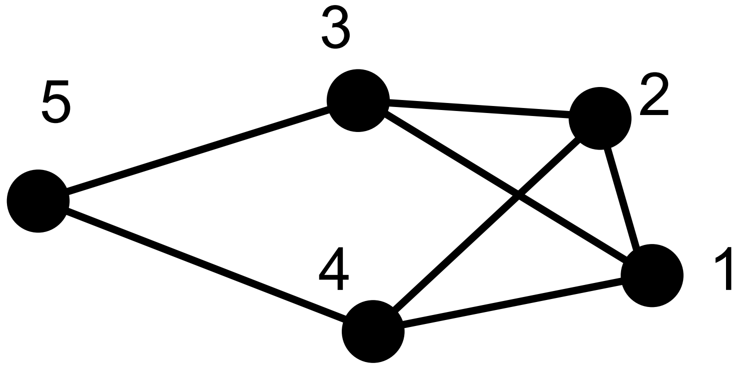
A mapping *p* from one graph to another is an isomorphism if whenever u is tied to v, p(u) is tied to p(v).

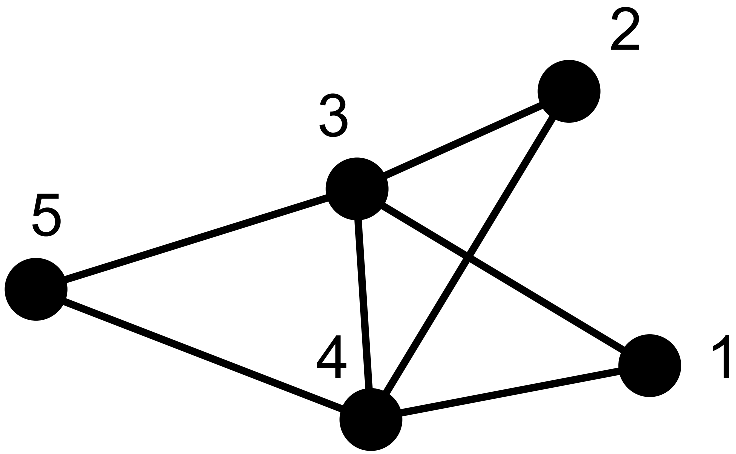
Isomorphisms are mappings that preserve structure

Isomorphism

Maps between objects that preserve structure – it can consider different target and domain.

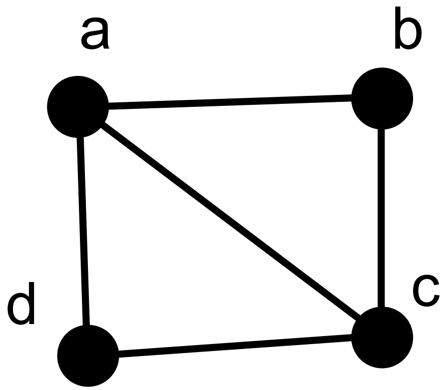




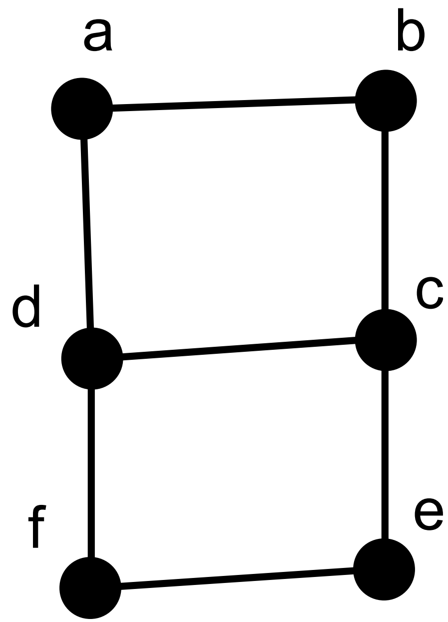


Automorphisms

Automorphism is an isomorphism from G to G – source and target match.

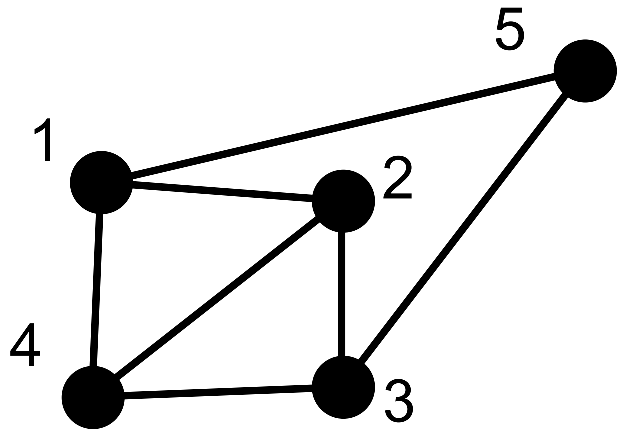


1. Is (a b) (d c) an automorphism?
2. Is (a d) (b c) an automorphism?
3. Is (a c) (b d) an automorphism?



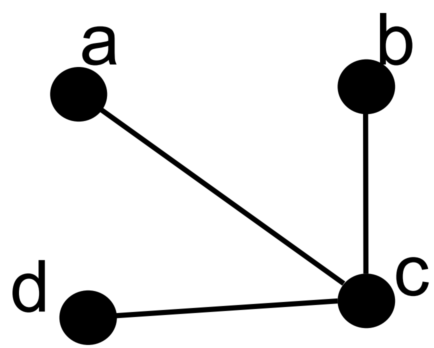
Cycle Notation

- (1 3) (2 4) (5)



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

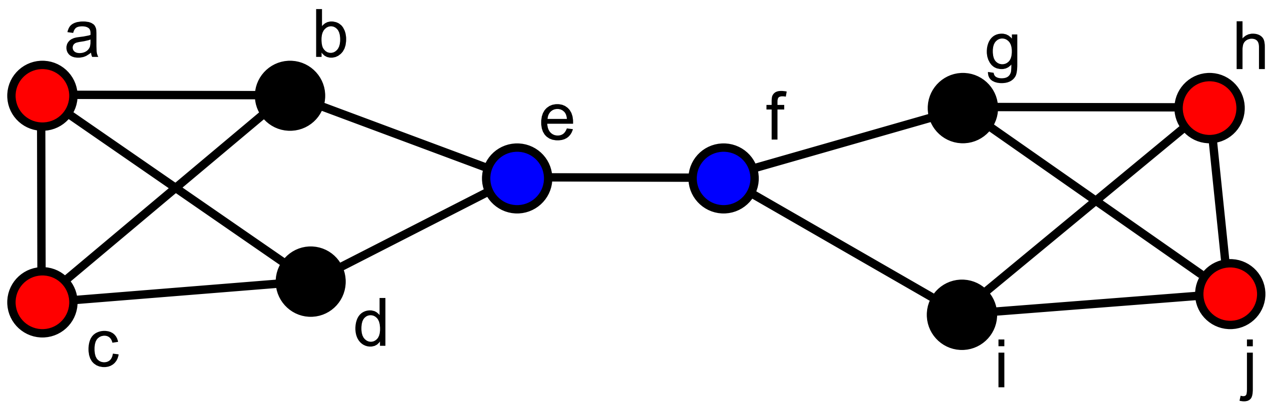
- (a b d) (c)



v	p(v)
a	b
b	d
c	c
d	a

Automorphic Equivalence

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

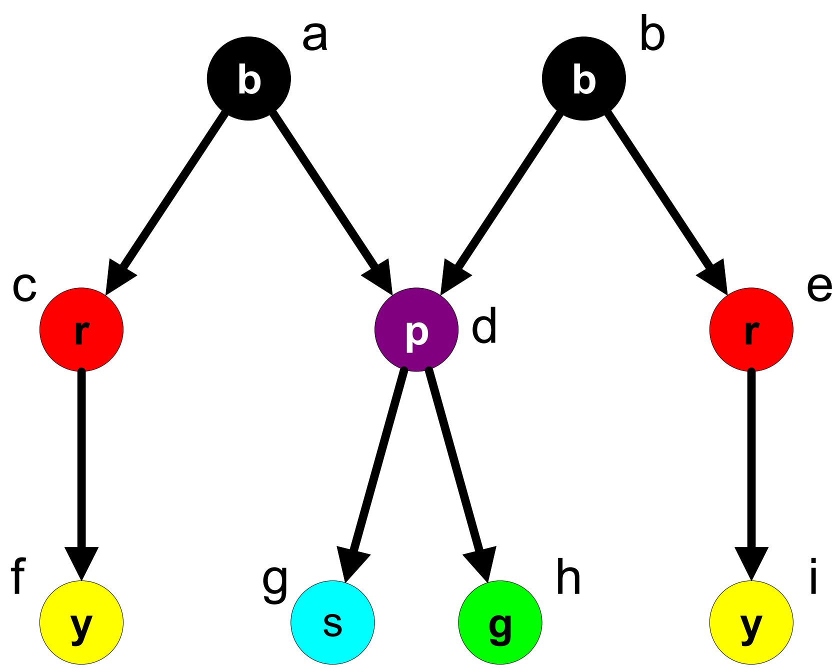


Automorphic Equivalence

- Powerful, fundamental intuitive concept
- Truly structural/positional, not confounded by contiguity
- Captures results of exchange experiments
- Captures essentials of the role concept
- Generalization of structural equivalence

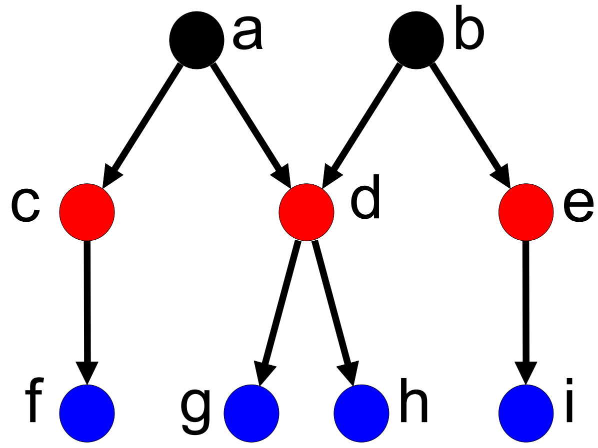
Problems with Automorphic Equivalence

- A parent with 2 children does not play the same role as one with 3 children
- Extremely difficult to compute
- No obvious way to relax the concept for application to real world data
- Does not work well with asymmetric data



Regular Equivalence

- Two nodes u and v are regularly equivalent if
  - Whenever u→c, there exists a node d such that v→d and c and d are regularly equivalent, and
  - Whenever c→u, there exists a node d such that d→v and c and d are regularly equivalent
- C(u)=C(v) implies C(N(u)) = C(N(v))
- Actually, C(u)=C(v) implies $C(N^{out}(u)) = C(N^{out}(v))$ and $C(N^{in}(u)) = C(N^{in}(v))$
- Regularly equivalent nodes have the same colors in their neighborhood (not necessarily in the same quantity)



Regularly equivalent nodes are not necessarily connected to the same third parties, but they are connected to equivalent third parties (though not necessarily in the same quantity)

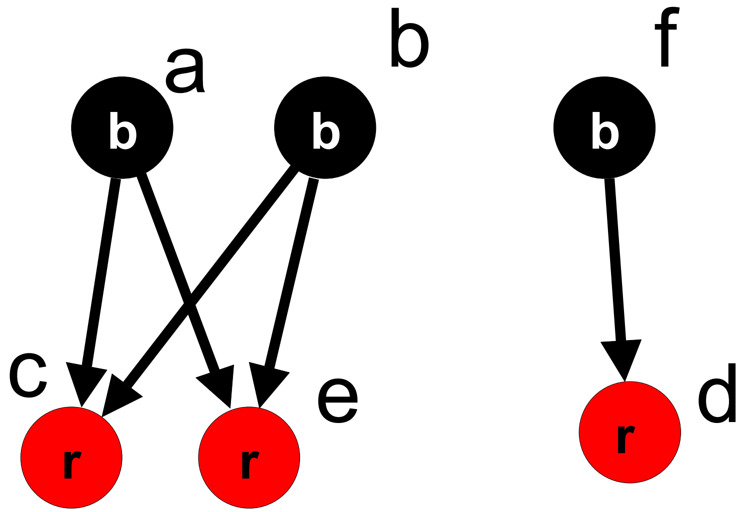
Regular Equivalence

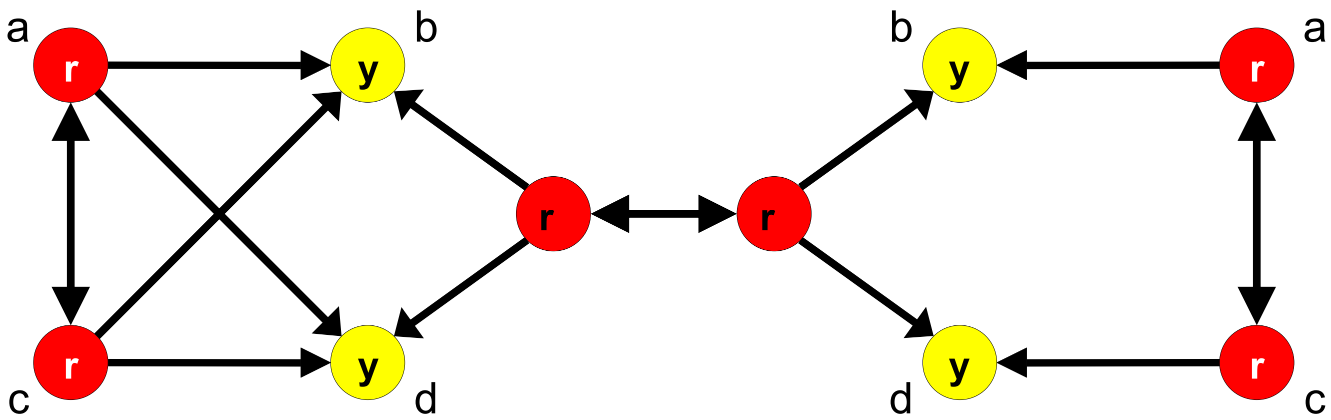
- Captures role concept really well
  - Two actors are equivalent if they have the same relations with equivalent others
  - You call American airlines and talk to clerk about booking flight, while I call USAIR and do same with their clerk
    - You and I equivalent because the clerks are equivalent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

Regular Equivalence

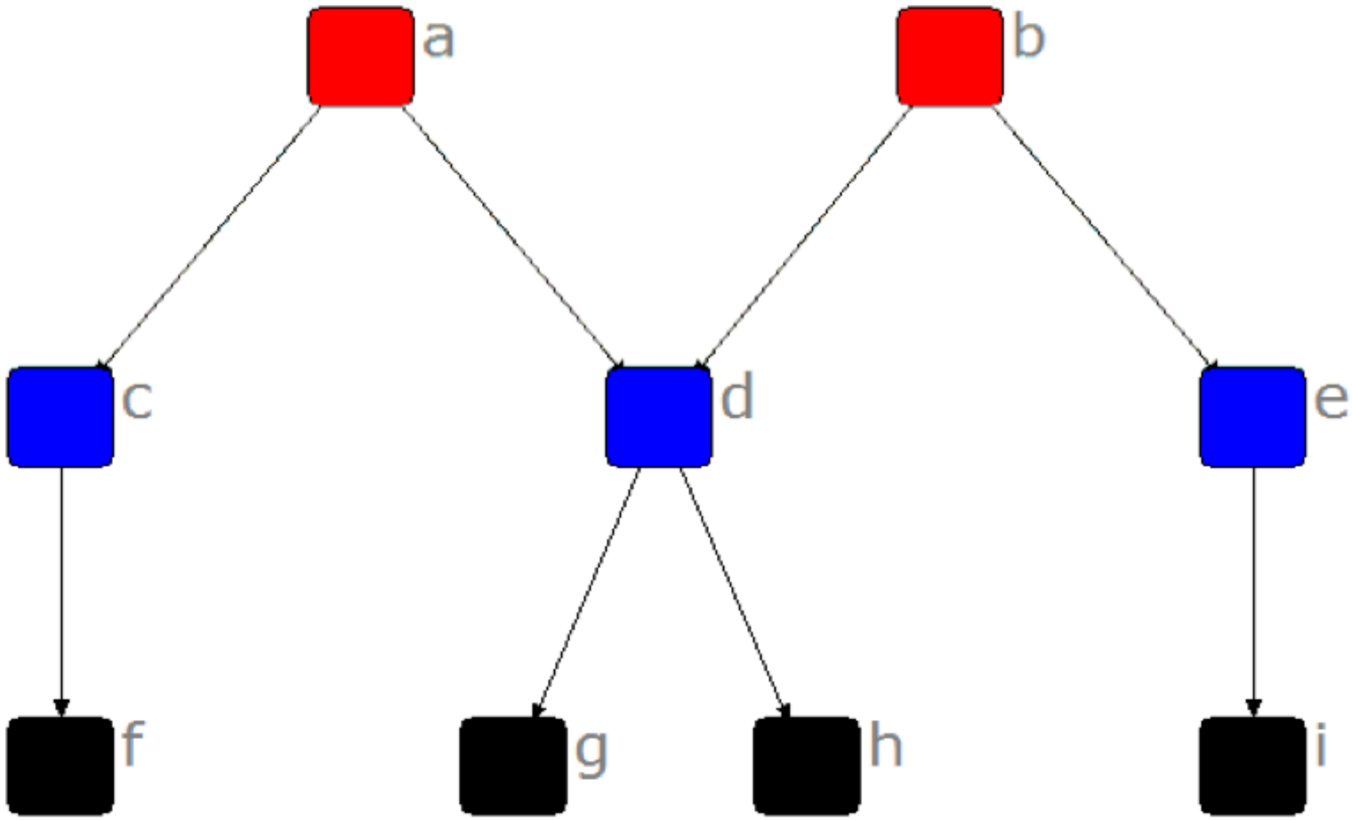
- Also captures position in hierarchies well
  - Including trophic group
- Relatively easy to compute (and to relax)
- Easy to generalize to 2-mode data
  - Consumers & brands
    - Segments & positions
    - identifying category boundaries
- Works well with multiple relations, valued, & directed data, and disconnected graphs

Regular Equivalence

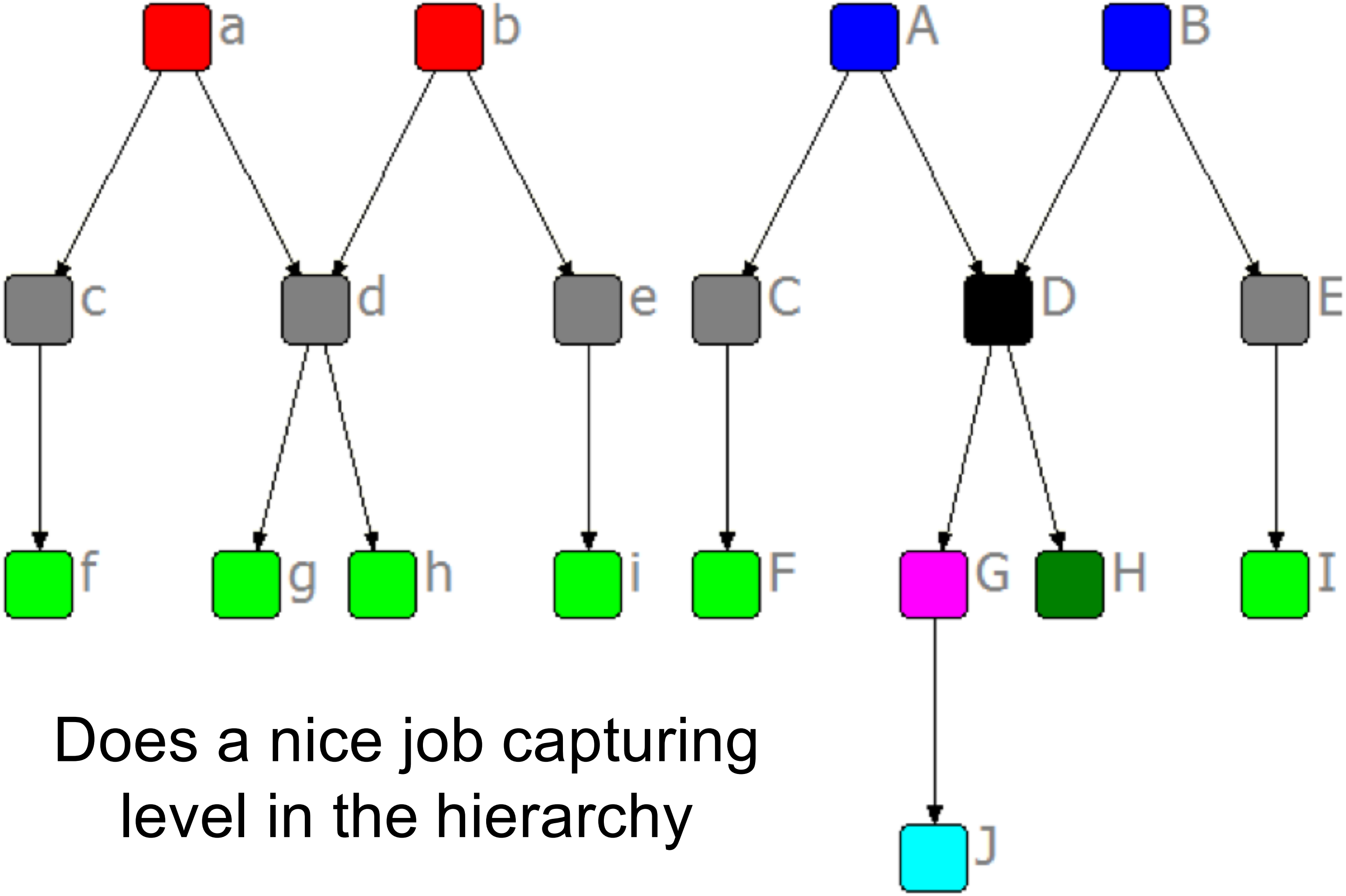




Regular Equivalences



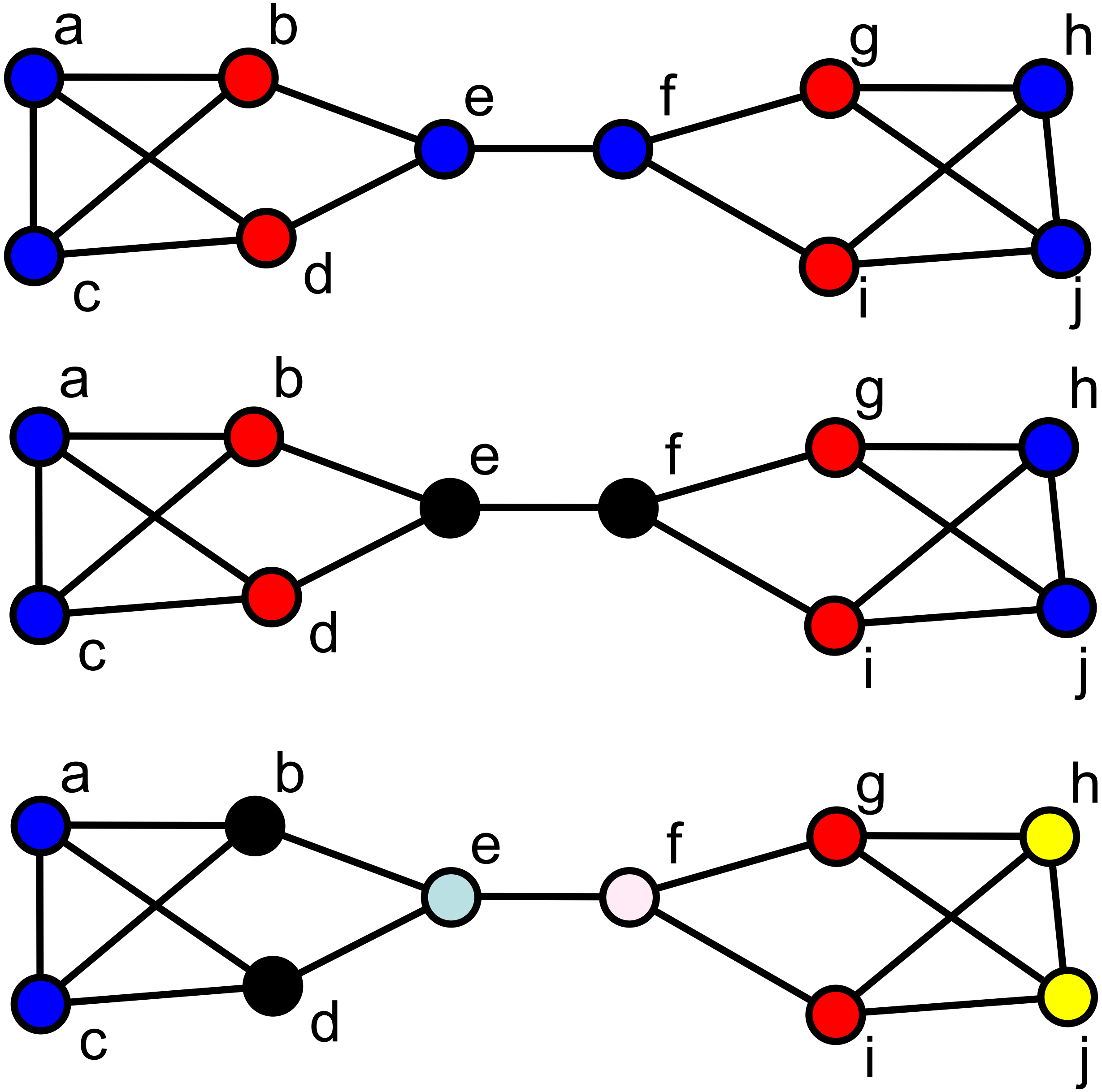
Regular Equivalence in a Disconnected Network



Does a nice job capturing level in the hierarchy

Problems with Regular Equivalence

- Often hard to interpret
  - Humans good at understanding pattern similarities, but not in the context of social ties
  - Data sets often inappropriate for R.E. analysis
    - Too small, no real roles
  - Do not work well with undirected data
- A graph may have multiple colorations that are regular

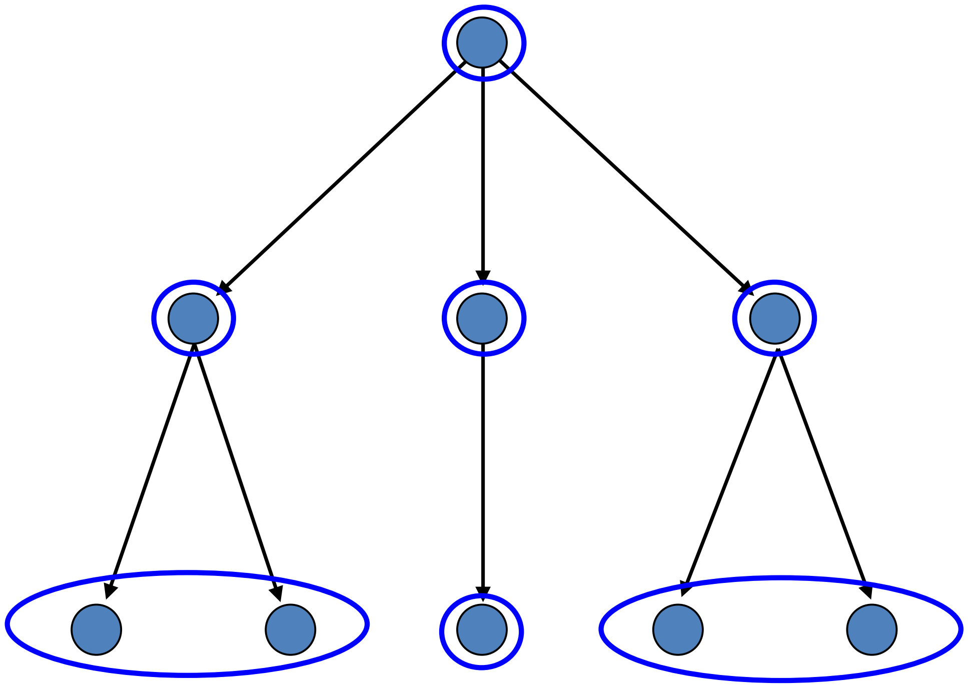


a
b
c
d
e
f
g
h
i
j
a
b
c
d
e
f
g
h
i
j
a
b
c
d
e
f
g
h
i
j

recap

Structural Equivalence

Two actors are equivalent if they have the same type of ties to the *same people*.

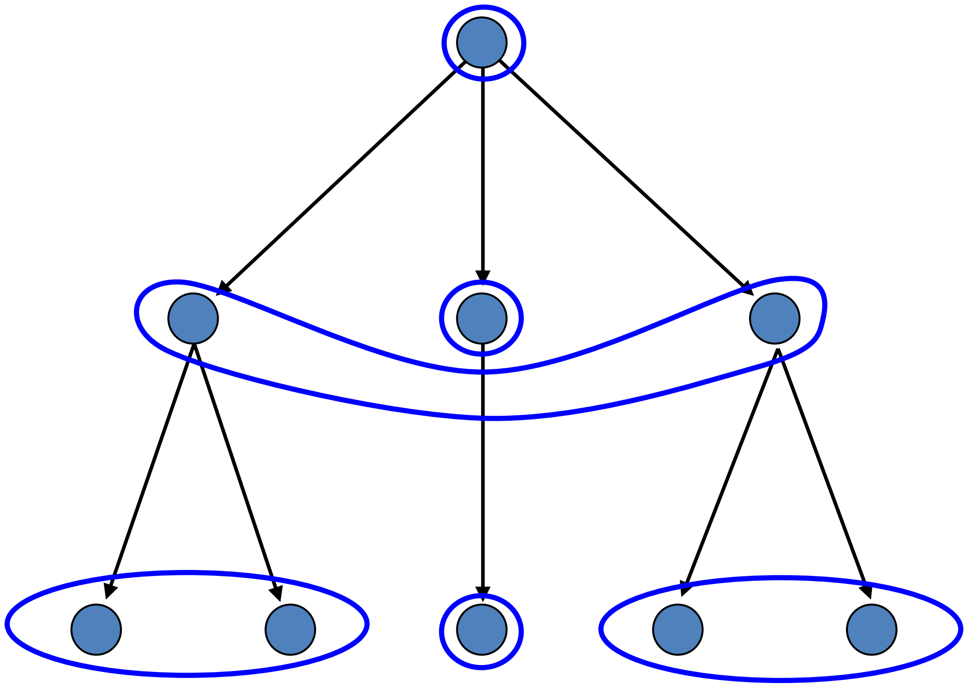


recap

Automorphic Equivalence

Actors occupy indistinguishable structural locations in the network. That is, that they are in *isomorphic* positions in the network.

In general, automorphically equivalent nodes are equivalent with respect to *all graph theoretic properties* (I.e. degree, number of people reachable, centrality, etc.)



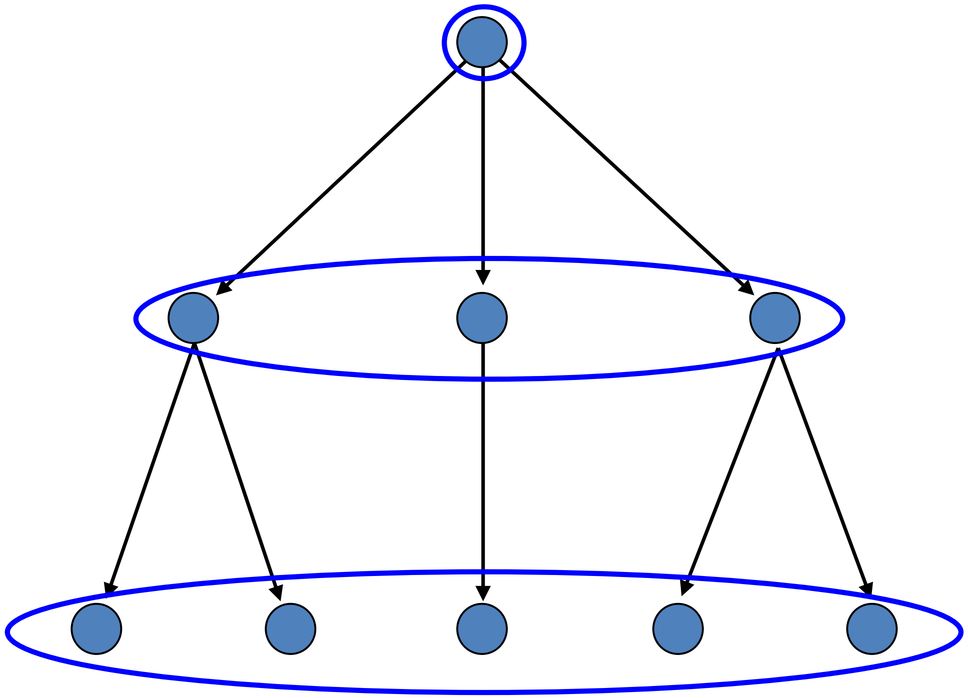
recap

Regular Equivalence

Does not require actors to have identical ties to identical actors or to be structurally indistinguishable.

Actors who are regularly equivalent have identical ties to and from *equivalent* actors.

If actors i and j are regularly equivalent, then for all relations and for all actors, if $i \longrightarrow k$, then there exists some actor m such that $j \rightarrow m$ and k is regularly equivalent to m.



There may be multiple regular equivalence partitions in a network, and thus we tend to want to find **the maximal regular equivalence position**, the one with the fewest positions.

Note that:

1. Structurally equivalent actors are automorphically equivalent;

1. Automorphically equivalent actors are regularly equivalent.

1. Structurally equivalent and automorphically equivalent actors are regular equivalent.

In practice, we tend to ignore some of these fine distinctions, as they get blurred quickly once we have to operationalize them in real graphs. It turns out that few people are ever *exactly* equivalent, and thus we approximate the links between the types.

The process of identifying positions is called ***blockmodeling***, and requires identifying a measure of similarity among nodes.

Computation

- Relaxing concepts for real world data
- Two approaches
  - Discrete or **blockmodel**
    - Partition nodes into mutually exclusive classes such that departures from equivalence model are minimized
  - **Profile similarity**
    - For each pair of nodes, calculate the **degree** to which each pair is equivalent

structural equivalence | profile similarity

- Profile similarity method [sna::sedist & blockmodeling::sedist]
  - Compute similarity/distance between rows of adjacency matrix
    - Product-Moment Correlation
    - Gamma Correlation
    - Euclidean distance
    - Hamming distance
  - Much argument over handling of diagonals
  - Can then MDS or cluster the resulting proximity matrix

computing equivalences | profile similarity

Because structural equivalence requirements of perfectly identical tie patterns almost **never occurs** in real networks, we relax criterion to find "**approximately structurally equivalent**" actors.

Several continuous measures can be computed on pairs of rows-columns. Distance measures:

- **Euclidean distance [0,1]** (number of neighbors that differ between two vertices): distance between rows *i,j* and columns *i,j* in the adjacency matrix. If two actors are structurally equivalent, then their entries in the rows and columns will be identical, and the Euclidean distance will be 0.

It also has the properties of a distance metric:

- distance from an object to itself is 0 ($d_{ii} = 0$)
- symmetric ($d_{ij} = d_{ji}$)
- distances are greater than or equal to zero ($d_{ij} \geq 0$)

$$d_{ij} = \sqrt{\sum_{k=1}^{g-2}[(x_{ik} - x_{jk})^2 + (x_{ki} - x_{kj})^2]}$$

computing equivalences | profile similarity

Because structural equivalence requirements of perfectly identical tie patterns almost **never occurs** in real networks, we relax criterion to find “**approximately structurally equivalent**” actors.

Several continuous measures can be computed on pairs of rows-columns. Distance measures:

- **Correlation coefficient [-1,1]** (normalized count of common neighbors; compares number of common neighbors with expected value of that count would take in a random network);

Permuted

	A	B	D	C	E
A	0	1	0	0	0
B	1	0	0	0	0
D	0	1	0	0	0
C	1	1	1	0	0
E	1	1	1	0	0

Euclidean

	A	B	C	D	E
A	0	0	0	1	1
B	0	0	1	1	1
C	0	1	0	1	1
D	1	1	1	0	0
E	1	1	1	0	0

Correlations

	A	B	D	C	E
A	1.0	1.0	1.0	0.5	0.5
B	1.0	1.0	0.0	0.5	0.5
D	1.0	0.0	1.0	0.5	0.5
C	0.5	0.5	0.5	1.0	1.0
E	0.5	0.5	0.5	1.0	1.0

computing equivalences | representing positions

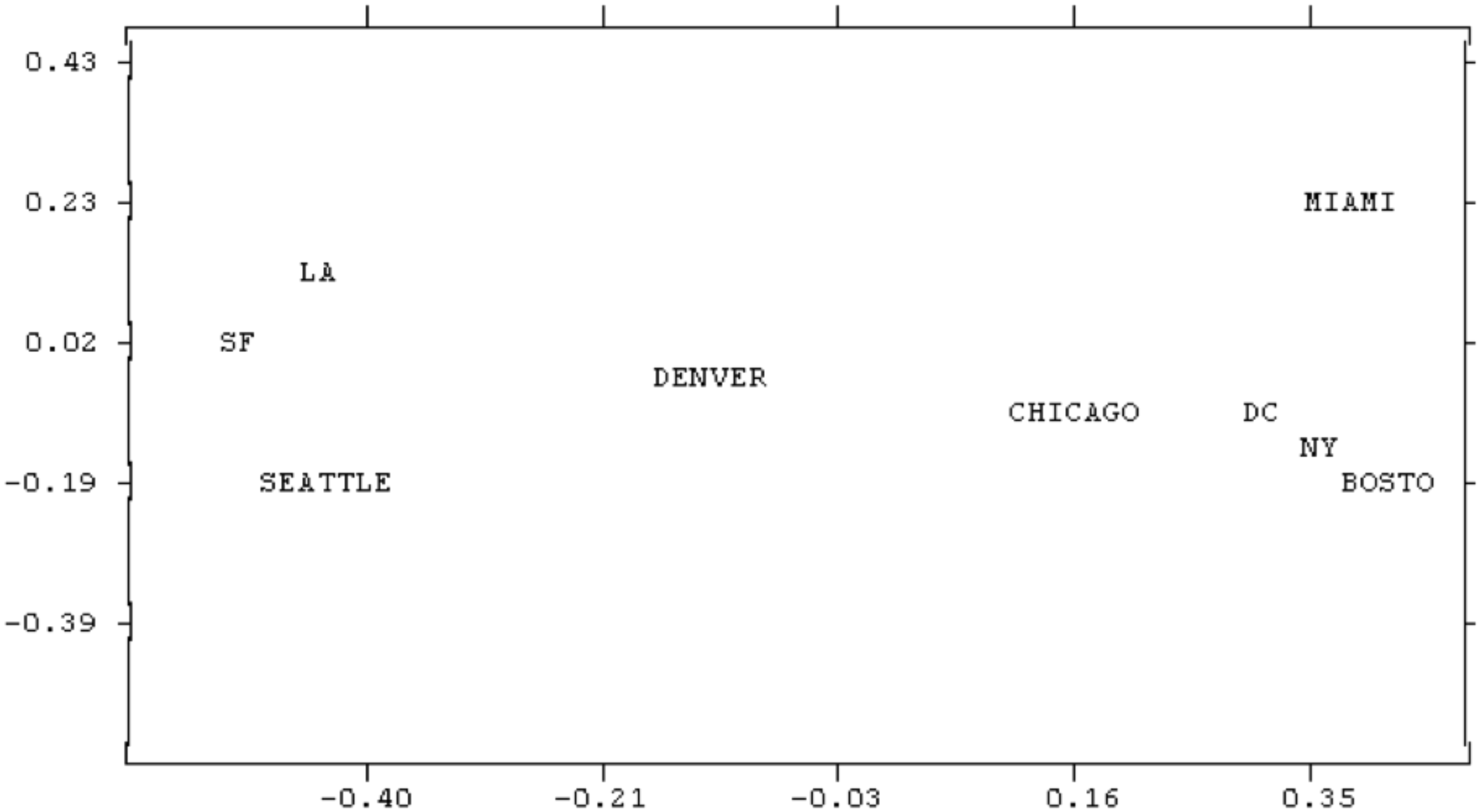
Multidimensional Scaling on similarities (Correlations) or dissimilarities (Euclidean distance) matrices. Plots on n-dimensional plane.

		1 BOST	2 NY	3 DC	4 MIAM	5 CHIC	6 SEAT	7 SF	8 LA	9 DENV
1	BOSTON	0	206	429	1504	963	2976	3095	2979	1949
2	NY	206	0	233	1308	802	2815	2934	2786	1771
3	DC	429	233	0	1075	671	2684	2799	2631	1616
4	MIAMI	1504	1308	1075	0	1329	3273	3053	2687	2037
5	CHICAGO	963	802	671	1329	0	2013	2142	2054	996
6	SEATTLE	2976	2815	2684	3273	2013	0	808	1131	1307
7	SF	3095	2934	2799	3053	2142	808	0	379	1235
8	LA	2979	2786	2631	2687	2054	1131	379	0	1059
9	DENVER	1949	1771	1616	2037	996	1307	1235	1059	0

Distances in miles between US cities

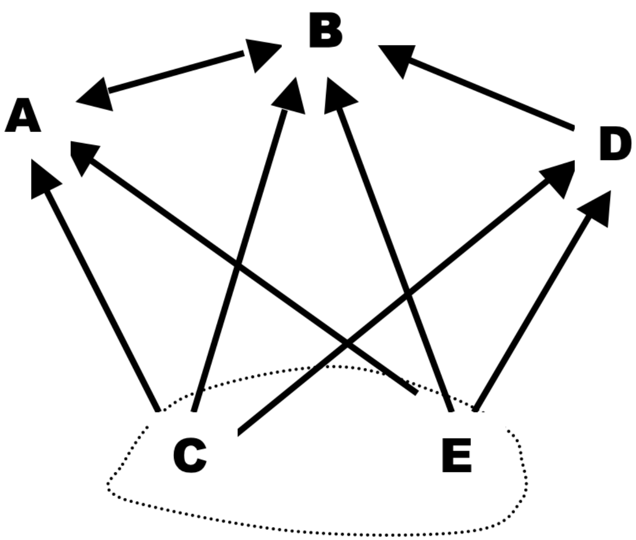
computing equivalences | representing positions

Multidimensional Scaling on similarities (Correlations) or dissimilarities (Euclidean distance) matrices. Plots on n-dimensional plane.



computing equivalences | blockmodel intuition

Equivalences within a graph are revealed by **permuting** (rearranging) the rows and columns of the adjacency matrix, to show adjacent actors that have **identical rows and columns vector entries**.



BEFORE

	A	B	C	D	E
A	0	1	0	0	0
B	1	0	0	0	0
C	1	1	0	1	0
D	0	1	0	0	0
E	1	1	0	1	0

AFTER

	A	B	D	C	E
A	0	1	0	0	0
B	1	0	0	0	0
D	0	1	0	0	0
C	1	1	1	0	0
E	1	1	1	0	0

structural equivalence | blockmodeling

The goal is to reduce a large, incoherent network to a smaller comprehensible structure that can be interpreted more readily;

it is based on the idea that units can be grouped according to the extent to which they are equivalent, according to some *meaningful* definition of equivalence

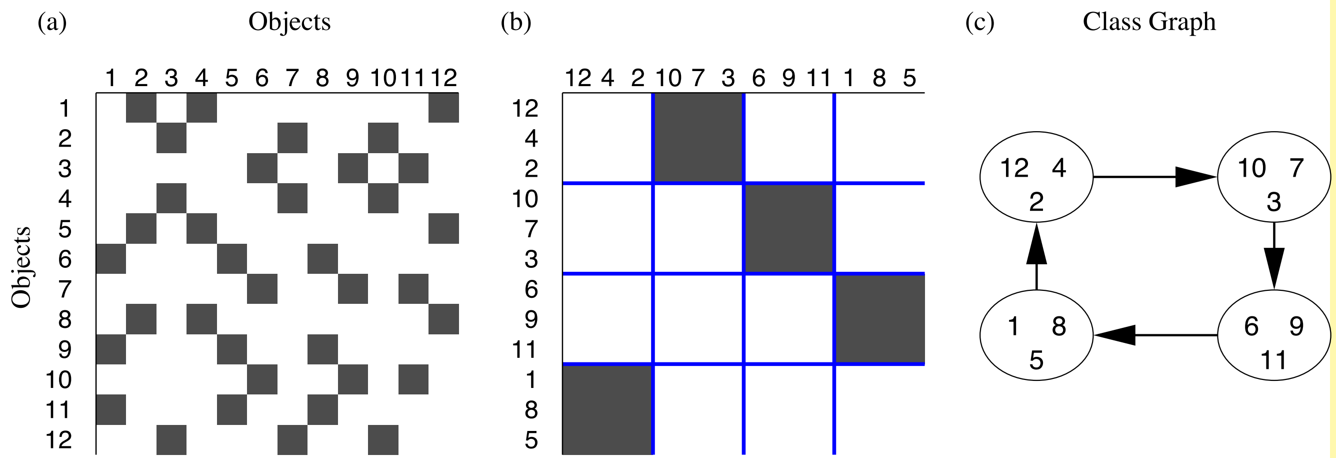
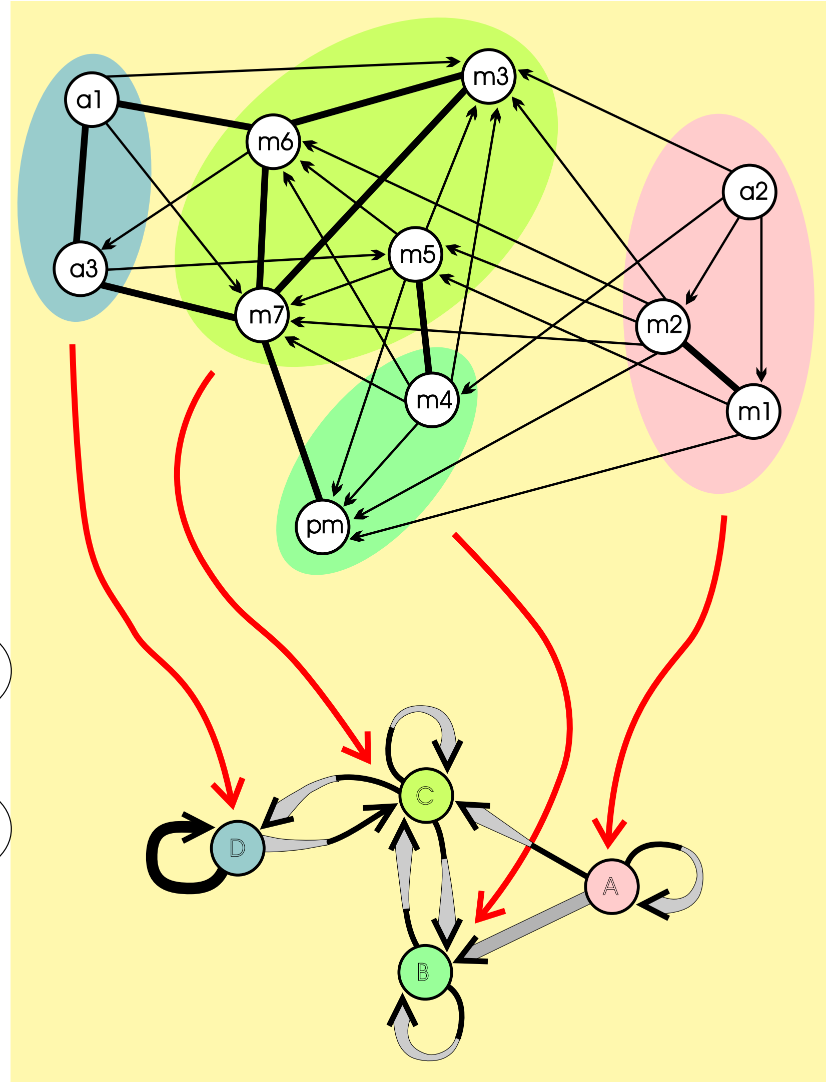


Figure 1: (a) An adjacency matrix representing a relation over a domain with twelve objects (b) The adjacency matrix permuted according to z, a vector of class assignments. The blue lines separate the four classes. (c) A class graph showing relations between the four latent classes. Each class sends links to one class and receives links from another class.



structural equivalence | blockmodeling

Blockmodel is a partition of the set of g actors into B discrete positions, with permuted and blocked matrices showing the presence of absence of ties within and between positions.

Positions in blocked matrix rarely exhibit strict structural equivalence; In real social data, block models often find positions that contains mixtures of 1s and 0s. The analyst must decide then on some criterion for assigning either 0 or 1 to each cell of the block model image.

Suppose a 4x4 blocking finds these submatrix proportions, where the overall mean network density = 0.30:

	Block I	Block II	Block III	Block IV
Block I	0.70	0.48	0.27	0.19
Block II	0.33	0.40	0.31	0.11
Block III	0.37	0.30	0.29	0.08
Block IV	0.32	0.29	0.02	0.12

Then using 0.30 as the α density criterion, the blockmodel image is:

1	1	0	0
1	1	1	0
1	1	0	0
1	0	0	0

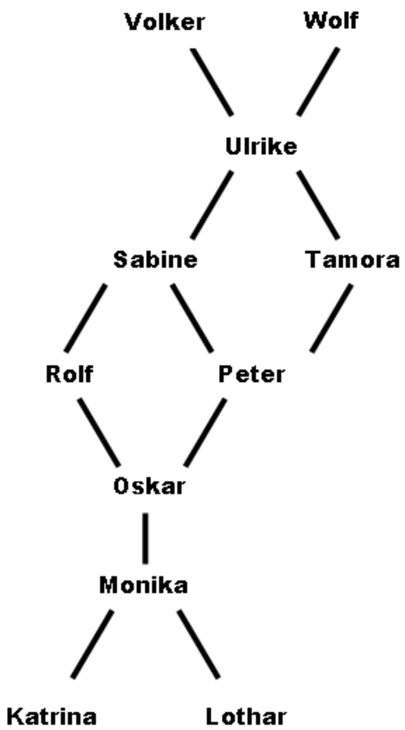
computing equivalences | representing positions

CONvergence of iterated **COR**relations [**CONCOR**] routine:

- repeatedly correlates the row and column vectors of one (or more) adjacency matrices until all entries either become +1 or -1.

- an initial partition into two structurally equivalence submatrices (or “**blocks**”); subsequent bifurcations continue until every actors occupies a solo position;

- as an analyst, you **MUST decide** which level of split to report, typically using substantive knowledge of the network.

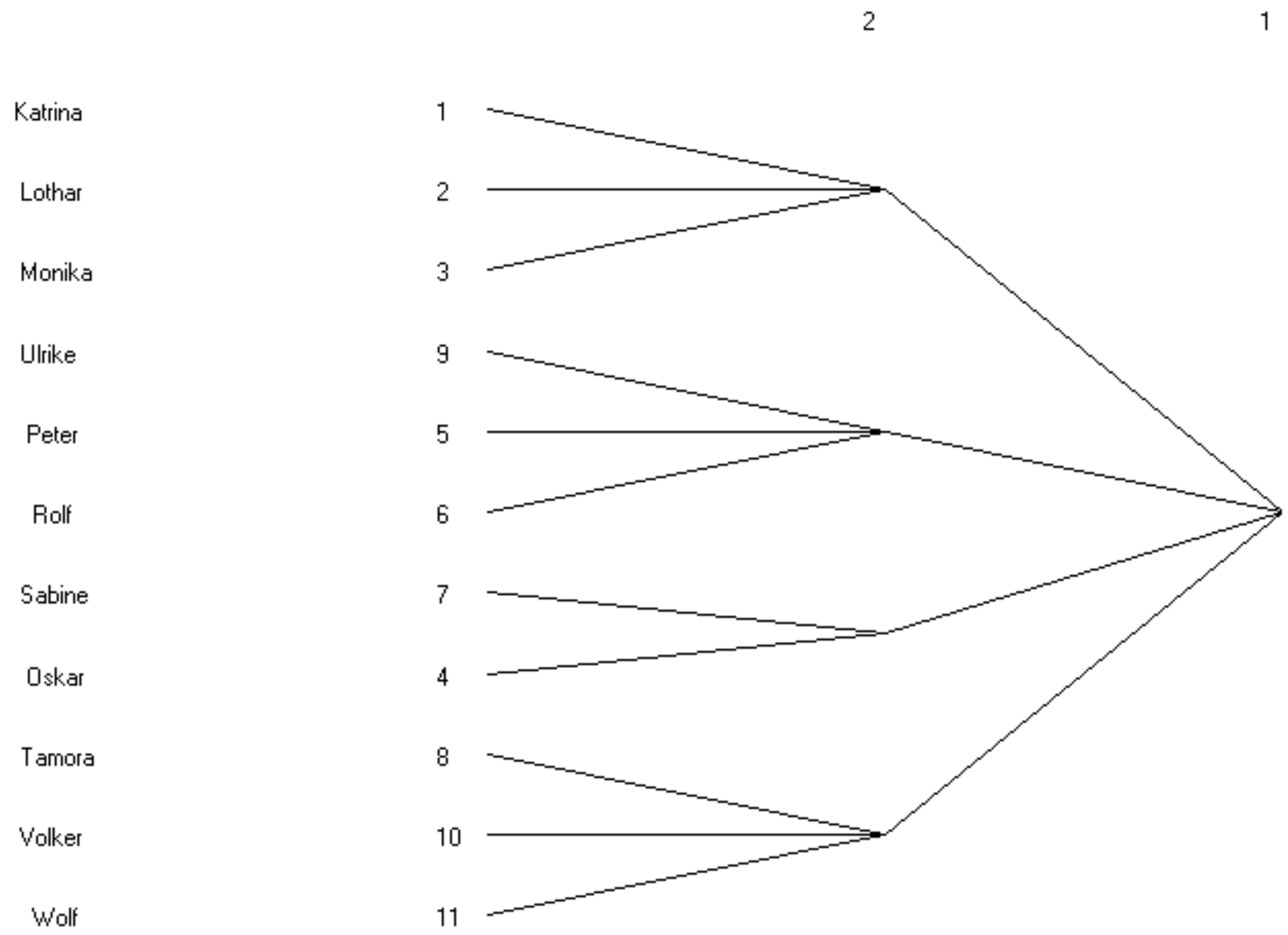


Here’s the four-block partition of nondirected binary ties in a “German Network” (courtesy Ulrich Brandes):

		K	L	M	U	P	R	S	O	T	V	W
1	Katrina			1								
2	Lothar			1								
3	Monika	1	1						1			
9	Ulrike							1		1	1	1
5	Peter							1	1	1		
6	Rolf							1	1			
7	Sabine				1	1	1					
4	Oskar			1		1	1					
8	Tamora				1	1						
10	Volker				1							
11	Wolf				1							

Densities in each/between blocks differ from 0 or 1 levels required by SE.

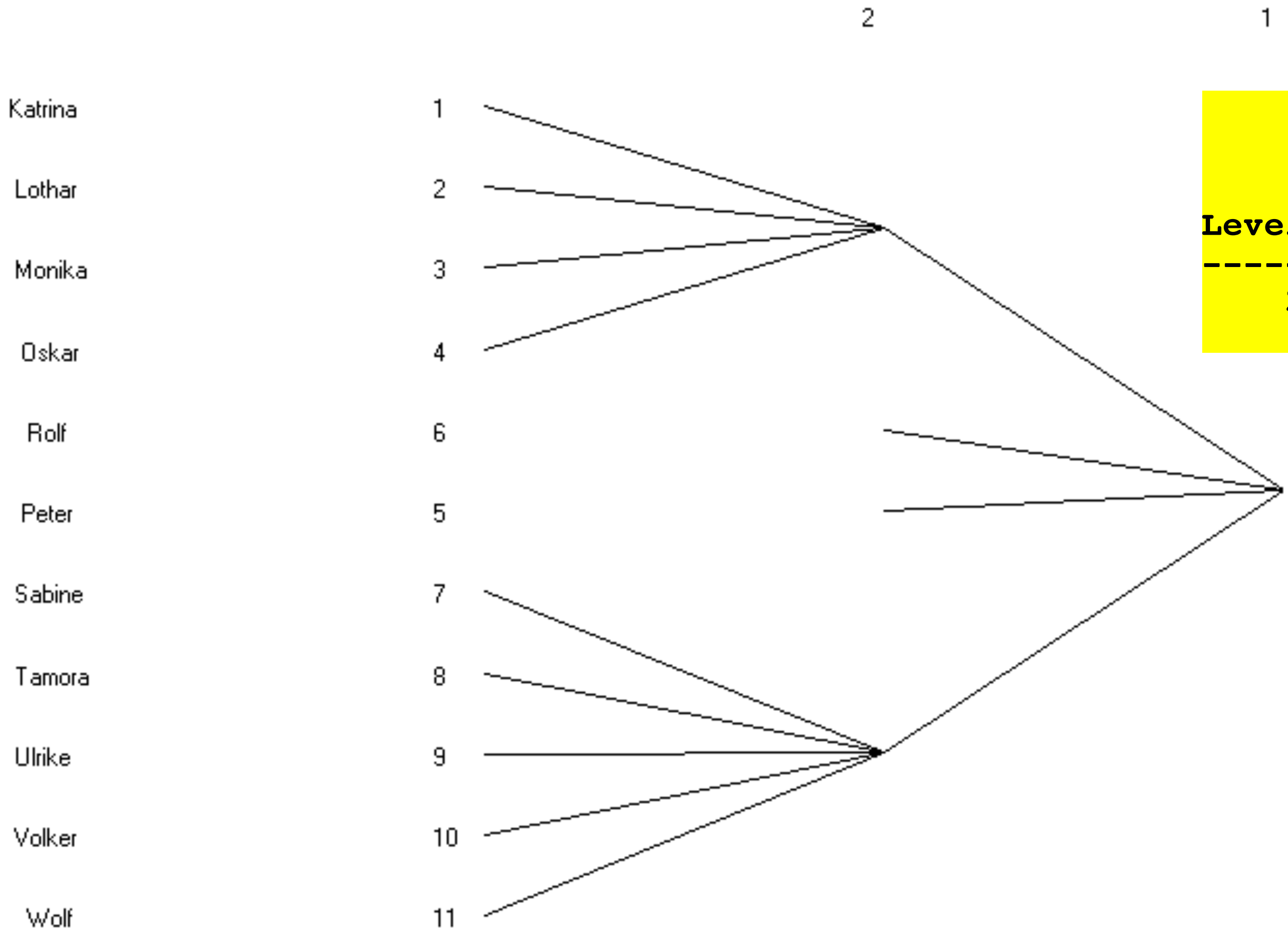
computing equivalences | representing positions



CONCOR dendogram reveals each actor joins one of 4 positions in this german network, with 0 sub-positions.

computing equivalences | representing positions

Hierarchical Clustering on Euclidean distance matrices




```
        K L M O R P S T U V W
        a o o s o e a a l o o
        t t n k l t b m r l l
Level   r h i a f e i o i k f
-----   - - - - - - - - - - -
    2   XXXXXXX . . XXXXXXXXX
    1   XXXXXXXXX XXXXXXXXXXX
```

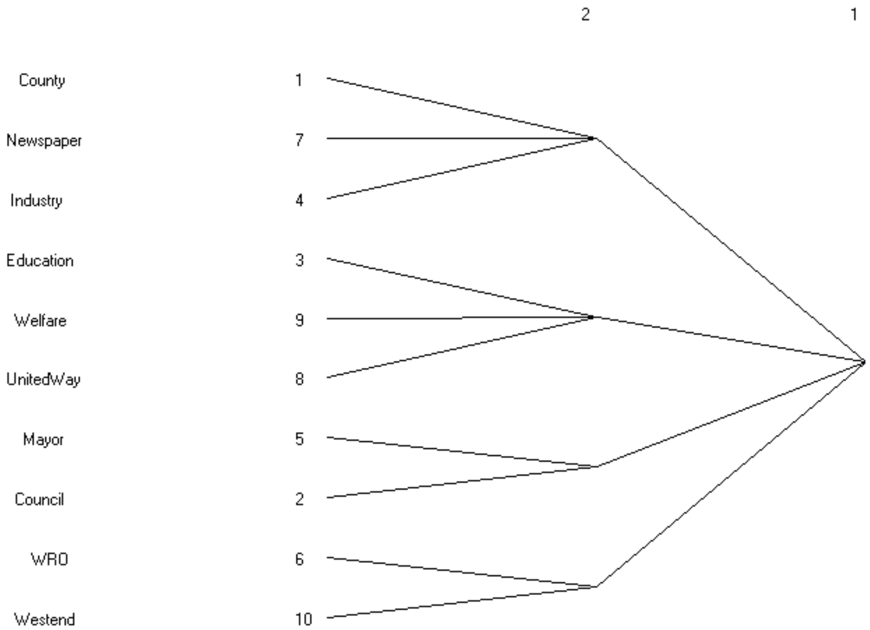
CONCOR dendogram reveals 2 multi-actor position and two solo-actor positions.

structural equivalence | blockmodeling example

(money and information exchange among 10 Indianapolist organizations)

		1 County	2 Counci	3 Educat	4 Indust	5 Mayor	6 WRO	7 Newspa	8 United	9 Welfar	10 Westen
1	County	1.000	0.142	0.150	0.451	0.278	0.105	0.298	0.257	0.341	0.107
2	Council	0.142	1.000	-0.061	0.142	0.404	0.350	0.297	0.143	0.142	0.207
3	Education	0.150	-0.061	1.000	0.043	-0.041	-0.102	0.316	0.375	0.471	0.171
4	Industry	0.451	0.142	0.043	1.000	0.383	0.105	0.298	0.150	0.341	0.358
5	Mayor	0.278	0.404	-0.041	0.383	1.000	0.317	0.323	-0.041	0.068	0.153
6	WRO	0.105	0.350	-0.102	0.105	0.317	1.000	-0.086	0.068	-0.070	0.419
7	Newspaper	0.298	0.297	0.316	0.298	0.323	-0.086	1.000	0.000	0.406	0.077
8	UnitedWay	0.257	0.143	0.375	0.150	-0.041	0.068	0.000	1.000	0.257	0.293
9	Welfare	0.341	0.142	0.471	0.341	0.068	-0.070	0.406	0.257	1.000	0.358
10	Westend	0.107	0.207	0.171	0.358	0.153	0.419	0.077	0.293	0.358	1.000

correlation matrix



Clustering informs row-partition/blocking to use in separate blockmodels

```
                                            1
                  1 7 4   3 9 8   5 2   6 0
                  C N I   E W U   M C   W W
                 ---------------------------
 1     County |       | 1 1 1 | 1   |   1 |
 7  Newspaper |       |     1 |   1 |     |
 4   Industry |   1   | 1 1 1 |   1 |     |
              -------------------------------
 3  Education |       |     1 |     |     |
 9    Welfare |       | 1   1 |     |     |
 8  UnitedWay |       |   1   |     |   1 |
              -------------------------------
 5      Mayor |       | 1 1 1 |   1 |     |
 2    Council |       | 1     |     |     |
              -------------------------------
 6        WRO |       |       |     |     |
10    Westend |       |       |     |     |
```

Reduced BlockMatrix

	1	2	3	4
1	0.167	0.778	0.500	0.167
2	0.000	0.667	0.000	0.167
3	0.000	0.667	0.500	0.000
4	0.000	0.000	0.000	0.000

>= 0.5 criterion

Reduced BlockMatrix

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

Generalized equivalence / block types

$$\begin{array}{c|ccccc|} & & & Y & & \\ \hline & 1 & 1 & 1 & 1 & 1 \\ X & 1 & 1 & 1 & 1 & 1 \\ & 1 & 1 & 1 & 1 & 1 \\ & 1 & 1 & 1 & 1 & 1 \\ \hline \end{array}$$

complete

$$\begin{array}{c|ccccc|} & & & Y & & \\ \hline & 0 & 1 & 0 & 0 & 0 \\ X & 1 & 1 & 1 & 1 & 1 \\ & 0 & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 1 & 0 \\ \hline \end{array}$$

row-dominant

$$\begin{array}{c|ccccc|} & & & Y & & \\ \hline & 0 & 0 & 1 & 0 & 0 \\ X & 0 & 0 & 1 & 1 & 0 \\ & 1 & 1 & 1 & 0 & 0 \\ & 0 & 0 & 1 & 0 & 1 \\ \hline \end{array}$$

col-dominant

$$\begin{array}{c|ccccc|} & & & Y & & \\ \hline & 0 & 1 & 0 & 0 & 0 \\ X & 1 & 0 & 1 & 1 & 0 \\ & 0 & 0 & 1 & 0 & 1 \\ & 1 & 1 & 0 & 0 & 0 \\ \hline \end{array}$$

regular

$$\begin{array}{c|ccccc|} & & & Y & & \\ \hline & 0 & 1 & 0 & 0 & 0 \\ X & 0 & 1 & 1 & 0 & 0 \\ & 1 & 0 & 1 & 0 & 0 \\ & 0 & 1 & 0 & 0 & 1 \\ \hline \end{array}$$

row-regular

$$\begin{array}{c|ccccc|} & & & Y & & \\ \hline & 0 & 1 & 0 & 1 & 0 \\ X & 1 & 0 & 1 & 0 & 0 \\ & 1 & 1 & 0 & 1 & 1 \\ & 0 & 0 & 0 & 0 & 0 \\ \hline \end{array}$$

col-regular

$$\begin{array}{c|ccccc|} & & & Y & & \\ \hline & 0 & 0 & 0 & 0 & 0 \\ X & 0 & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 & 0 \\ \hline \end{array}$$

null

$$\begin{array}{c|ccccc|} & & & Y & & \\ \hline & 0 & 0 & 0 & 1 & 0 \\ X & 0 & 0 & 1 & 0 & 0 \\ & 1 & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 1 & 0 \\ \hline \end{array}$$

row-functional

$$\begin{array}{c|cccc|} & & Y & & \\ \hline & 1 & 0 & 0 & 0 \\ & 0 & 1 & 0 & 0 \\ X & 0 & 0 & 1 & 0 \\ & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 1 \\ \hline \end{array}$$

col-functional

theory informing block partitions

Types of pre-specified blockmodels

The pre-specified blockmodeling starts with a blockmodel specified, in terms of substance, ***prior to an analysis***. Given a network, a set of ideal blocks is selected, a family of reduced models is formulated, and partitions are established by minimizing the criterion function.

The basic types of models are:

*	*	*
*	0	0
*	0	0

core - periphery

*	0	0
*	*	0
?	*	*

hierarchy

*	0	0
0	*	0
0	0	*

clustering

structural equivalence | blockmodeling

- Blockmodeling approach
  - Optimization method

    [sna::blockmodel]

    [blockmodeling::opt.random.par]

    - you tell it how many classes to create and it reports how well it did
    - similar algorithm available in R; but models for different *k* can then be assessed using "blockmodel" package.

- Older Concor method

  [devtools::install_github("aslez/concoR")]

  - **CON**vergence of iterated **COR**relations
  - Actually based on profile method, uses convergence of iterative correlation calculations
  - Not as accurate as Profile method

computing | regular equivalence

- REGE (Algorithms for computing (dis)similarities in terms of regular equivalence)

  [blockmodeling::REGE]

  – Creates a similarity matrix based on the data in the original matrix, so if there's not much variety in your original data, it's likely to clump all (most) of your nodes together.
  – Converting your data to geodesic distances may help with this, but even then if the patterns of distances are very similar, it may still produce only the trivial regular equivalence (coloration) of all nodes in one color.

REGE

- Getting around the trivial partition in data without much variation (e.g., binary, symmetric)
  - The best way around this, is to create a new matrix that has more variety (like maximum flow) and then run REGE on that
  - Or, simply use the OPTIMIZE routine, and specify the number of colors you want.
    - But this is a combinatorial optimization and can be very sensitive to number of nodes in the network.