BI

Nordic Centre
for Internet and Society

Visualisations and Network Theory

Social Networks: Advanced Concepts

Lecture 4

February 10th and 11th 2020

Agenda

Embedding Measures
Reciprocity, Transitivity, Clustering, Homophily

Positions and Roles
Structural Equivalence

Network Mechanisms

Exercise 3
Clustering and Layout Algorithms

1

Embedding Measures

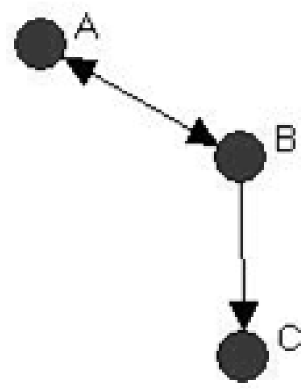
Reciprocity, Transitivity, Clustering, Homophily

Reciprocity

- With undirected data two actors are either connected or not.
- With directed data there are four possible dyadic relationships:
  - A and B are not connected
  - A sends to B
  - B sends to A
  - A and B send to each other.

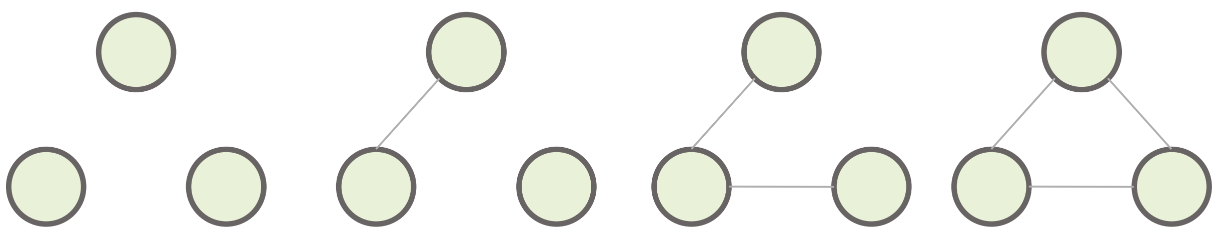
Reciprocity II

- What is the reciprocity in this network?
  - Answer 1: % of pairs that have reciprocated ties / all possible pairs
    - AB of {AB,AC,BC} = 0.33
  - Answer 2: % of pairs that have reciprocated ties / existing pairs
    - AB of {AB,BC} = 0.5
  - Answer 3: % directed ties / all directed ties
    - {AB,BA} of {AB, BA, AC, CA, BC, CB} = 0.33



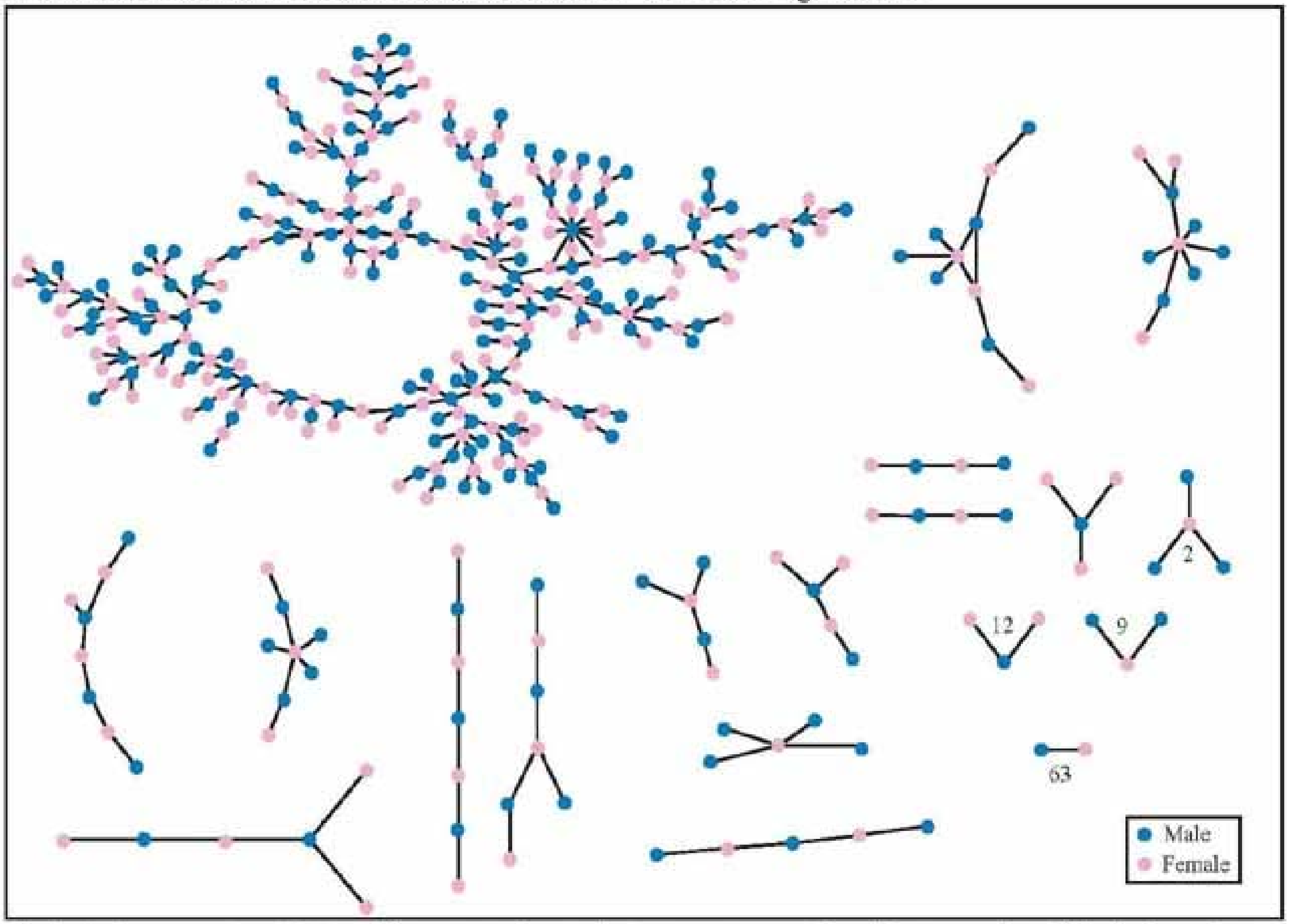
Triad Census I: Undirected Networks

- With undirected ties, there are four possible types of triadic relations
  - No ties
  - One tie
  - Two ties
  - Three ties
- The count of the relative prevalence of these four types of relations is called **triad census.** A population can be characterized by:
  - isolation
  - couples only
  - structural holes (one actors is connected to two others, who are not connected to each other)
  - clusters



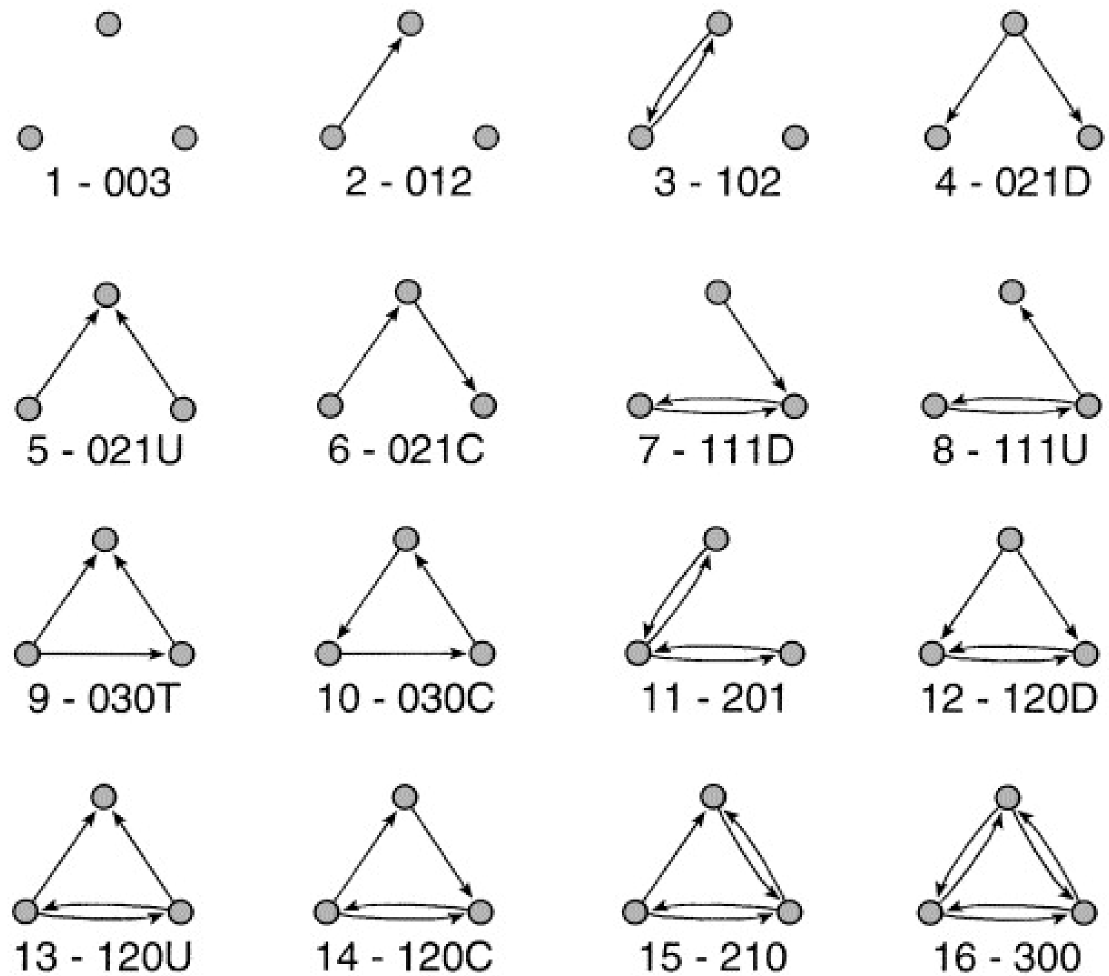
Sexual Networks

The Structure of Romantic and Sexual Relations at "Jefferson High School"



Each circle represents a student and lines connecting students represent romantic relations occuring within the 6 months preceding the interview. Numbers under the figure count the number of times that pattern was observed (i.e. we found 63 pairs unconnected to anyone else).

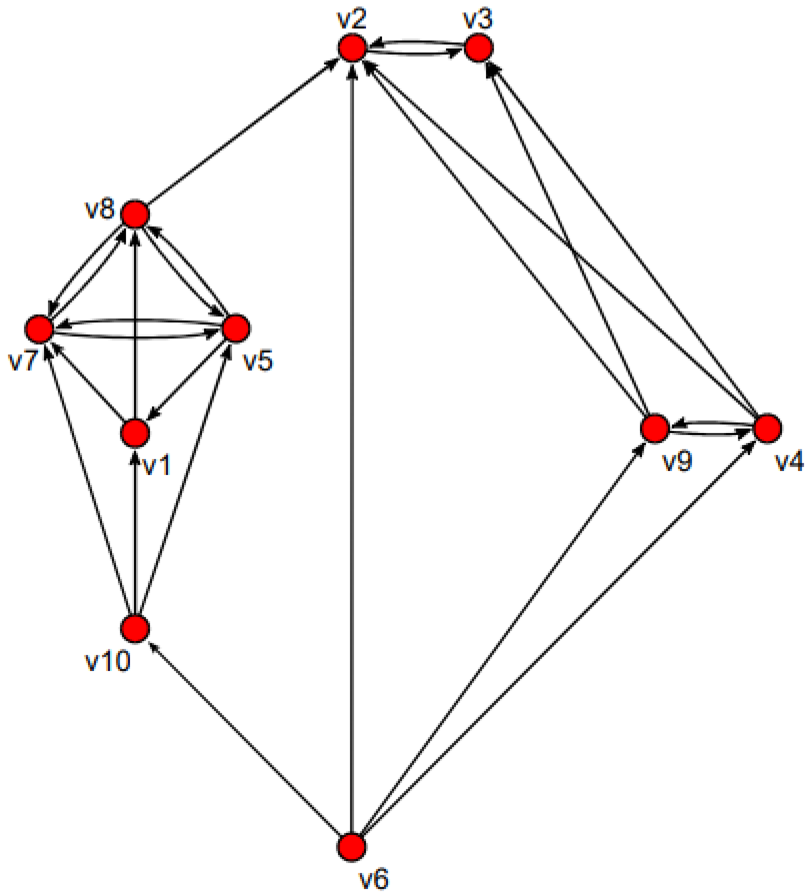
Triad Census II: Directed Networks



M-A-N number:
M: # of mutual dyads
A: # of asymmetric dyads
N: # of null dyads

D =Down, U = Up, C = Cyclic, T= Transitive

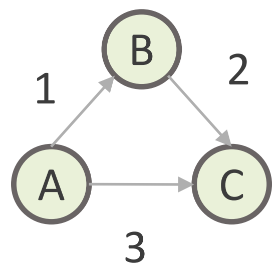
Example



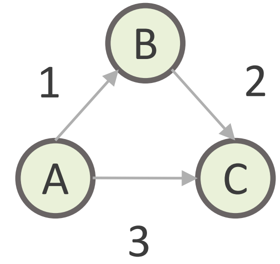
	Type	Number of triads	Expected
3	102	**22**	7.56
16	300	1	0.06
1	003	**7**	17.03
4	021D	**3**	7.56
5	021U	**3**	7.56
9	030T	4	5.81
12	120D	**5**	1.12
13	120U	2	1.12
2	012	**58**	39.3
14	120C	2	2.24
15	210	0	0.86
6	021C	**7**	15.12
7	111D	4	5.81
8	111U	**2**	5.81
10	030C	0	1.94
11	201	0	1.12
Total		120	

Transitivity

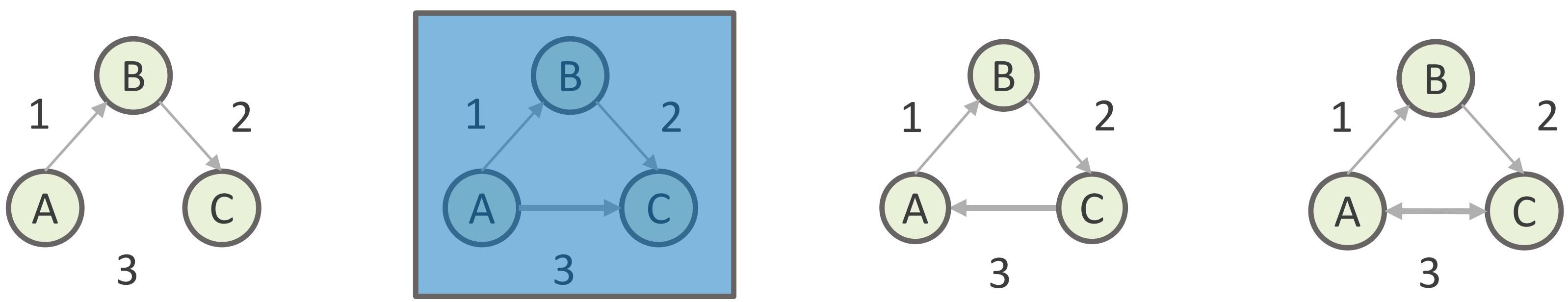
In particular, we may be interested in the proportion of triads that are transitive, that is, display a type of balance where, if A directs a tie to B, and B directs a tie to C, then A also directs a tie to C.



Transitivity



- How to measure transitivity?
  - A) Divide the number of found transitive triads by the total number of **triplets** of all kinds
  - B) Norm the number of transitive triads by the number of cases where a link **could** complete the triad. Norm {AB, BC, AC} by {AB, BC, anything} (for 3 nodes there are 4 possibilities; see below)



Transitivity

```
TRANSITIVITY
--------------------------------------------------------------------------------

Type of transitivity:         ADJACENCY
Input dataset:                C:\Program Files\Ucinet 6\DataFiles\KNOKBUR

Relation: KNOKI
---------------

Number of non-vacuous transitive ordered triples: 146
Number of triples of all kinds: 720
Number of triples in which i-->j and j-->k: 217

Percentage of all ordered triples: 20.28%
Transitivity: % of ordered triples in which i-->j and j-->k that are transitive: 67.28%
```

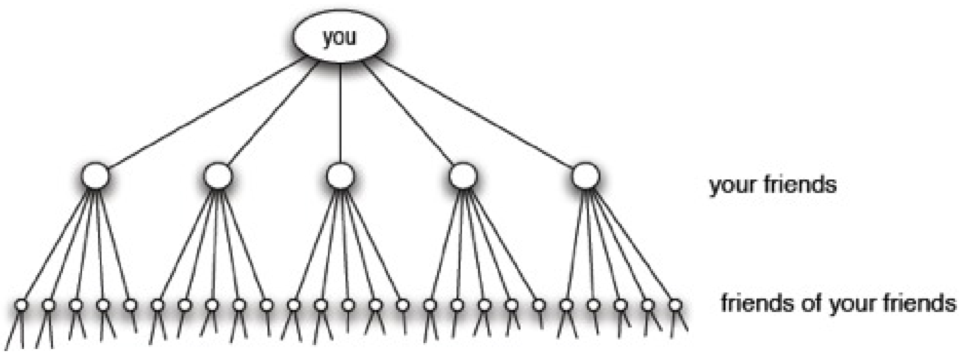
146/720

146/217

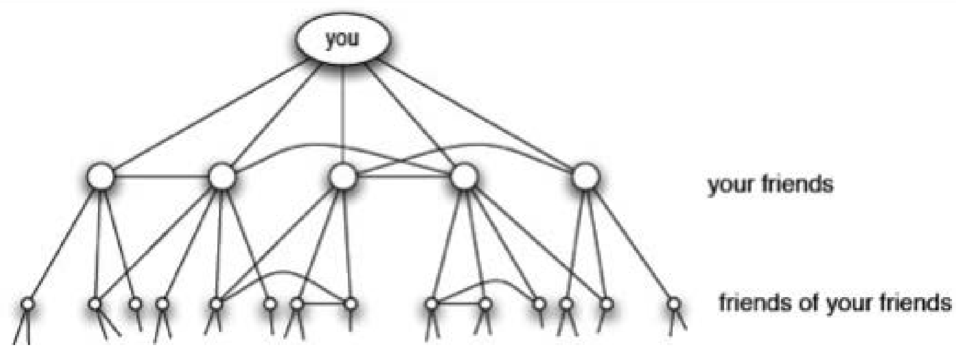
Network>Cohesion>Transitivity

Clustering

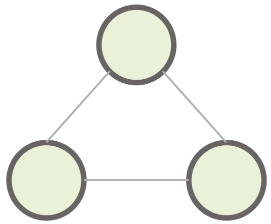
Most actors live in local neighborhoods and are connected to one another. A large proportion of the total number of ties is highly clustered into local neighborhoods.



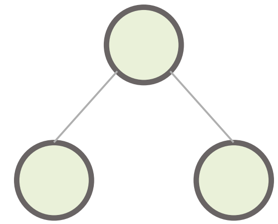
vs.



Global Clustering Coefficient



Closed triplet

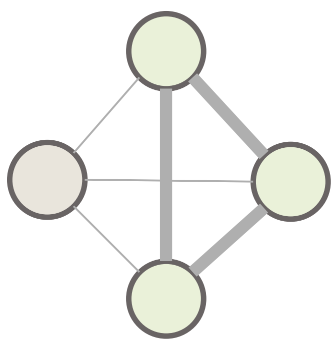


Triplet

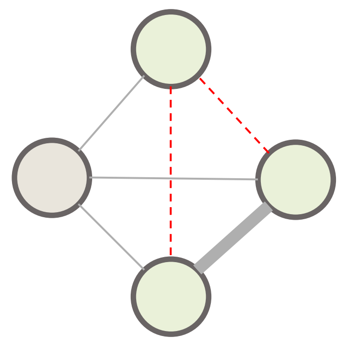
$$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triples of vertices}} = \frac{\text{number of closed triplets}}{\text{number of connected triples of vertices}}.$$

Average Local Clustering Coefficient

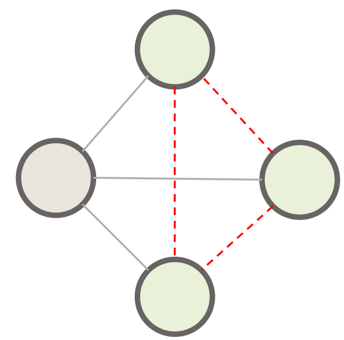
A measure to calculate how clustered the graph is to examine the local neighborhood of an actor (all actors who are directly connected to ego) and calculate the density in this neighborhood (leaving out the ego). After doing this for all actors, we can characterize the degree of clustering as an average of all the neighborhoods.



C = 1



C = 1/3



C = 0

Individual Local Clustering Coefficient

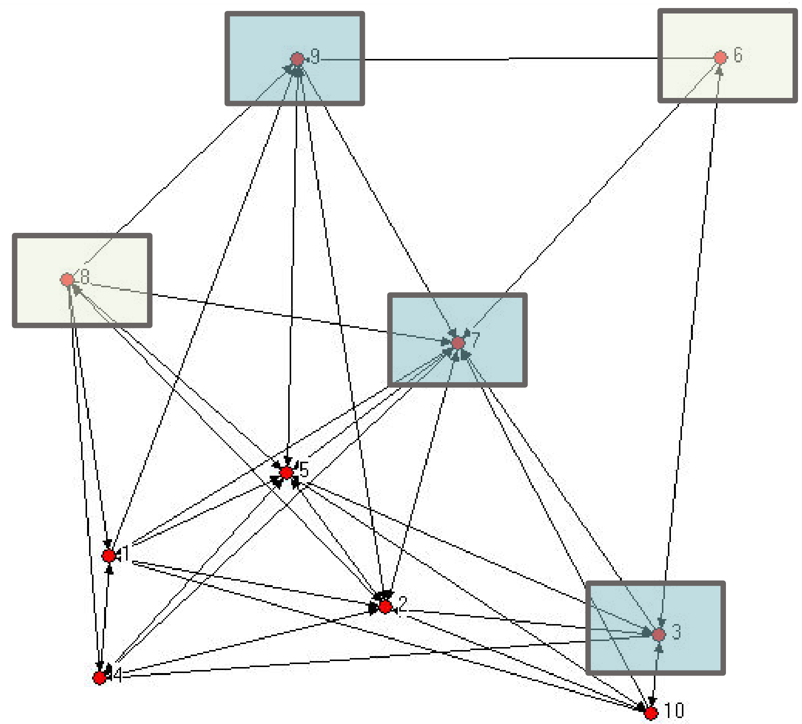
Clustering can also be examined for each actor:

- **Actor 6** has 3 neighbors and hence 6 possible directed ties between them are possible. Only 2 are present, so actor 6 is not highly clustered.
- **Actor 8** has 6 neighbors and they are thightly connected. Actor 8's network is highly clustered.

Node Clustering Coefficients

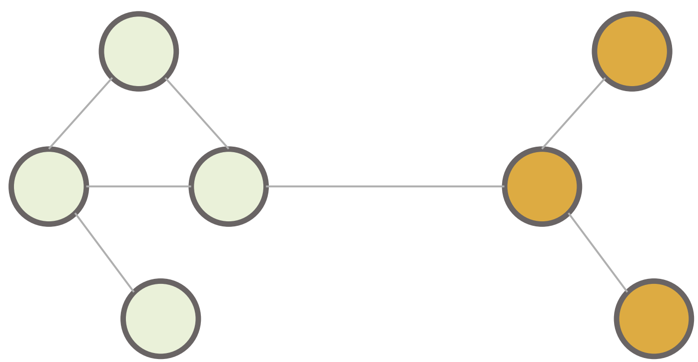
	1 Clus C	2 nPairs
1	0.667	21.000
2	0.536	28.000
3	0.567	15.000
4	0.733	15.000
5	0.518	28.000
6	0.333	3.000
7	0.514	36.000
8	0.800	15.000
9	0.600	15.000
10	0.800	10.000

2 edges out of 6 edges



E-I Index

- The E-I (external – internal) index takes the number of **ties of group members to outsiders**, **subtracts the number of ties to other group members**, and divides by **the total number of ties.**



(1-4)/7 = -3/7   (1-2)/7 = -1/7

E-I Index

- The resulting E-I index ranges from -1 (all ties internal) to +1 (all ties external).
- The E-I index can be applied at three levels:
  - entire population
  - each group
  - each individual

Notice: The relative size of sub populations (e.g. 10 vs. 1000) has dramatic consequences for the degree of internal and external contacts, even when individuals may choose contacts at random.

Permutation Tests

To assess whether the E-I index value is significantly different from what would be expected by **random** mixing a **permutation test** is performed.

```
Max possible E-I given density & group sizes: 1.000
Min possible E-I given density & group sizes: 0.250

Re-scaled E-I index: -0.167


Permutation Test
Number of iterations = 5000
```

	1 Obs	2 Min	3 Avg	4 Max	5 SD	6 P >= Ob	7 P <= Ob
1 Internal	0.219	0.625	0.733	0.844	0.039	1.000	0.000
2 External	0.781	0.156	0.267	0.375	0.039	0.000	1.000
3 E-I	0.563	0.250	0.467	0.688	0.078	0.203	0.953

Notice: Under random distribution, the E-I Index would be expected to have a value of .467 which is not much different from .563, especially given the standard error .078 (given the result, the difference of .10 could be just by chance)

E-I Index for Individuals

Individual Level E-I Index

	1 Inter	2 Exter	3 Total	4 E-I
1	1.000	6.000	7.000	0.714
2	2.000	6.000	8.000	0.500
3	1.000	5.000	6.000	0.667
4	2.000	4.000	6.000	0.333
5	2.000	6.000	8.000	0.500
6	1.000	2.000	3.000	0.333
7	2.000	7.000	9.000	0.556
8	1.000	5.000	6.000	0.667
9	2.000	4.000	6.000	0.333
10	0.000	5.000	5.000	1.000

Notice: Several actors (4,6,9) tend toward **closure**, while others (10,1) tend toward creating ties **outside** their groups.

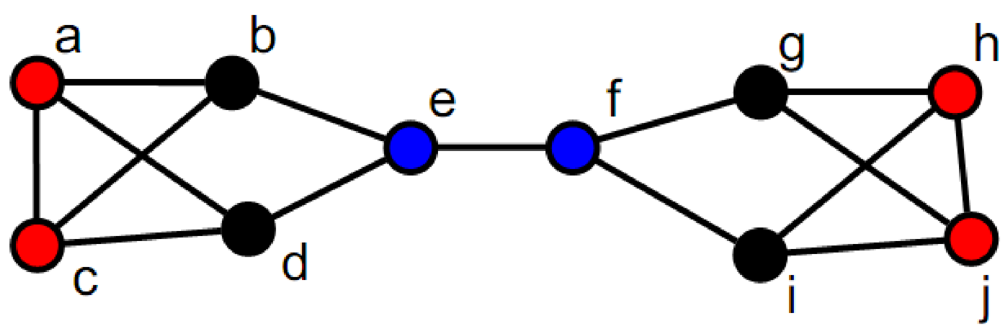
2 Position and Roles

Chinese Kinship Relations

First cousins			[hide]
Relation	**Term**	**English equivalent**	**Degree of mourning (duration)**
children of father's brother's son	堂姪兒女	first cousin once removed	
all other grandchildren of father's sibling	表姪兒女	"	
grandchildren of mother's sibling	表甥兒女	"	
son of paternal grandfather's brother who is older than ego's father	堂伯	"	
son of paternal grandfather's brother who is younger than ego's father	堂叔	"	
daughter of paternal grandfather's brother	堂姑	"	
son of maternal grandfather's brother	堂舅	"	
daughter of maternal grandfather's brother	堂姨	"	
son of paternal grandfather's sister who is older than ego's father	表伯	"	
son of paternal grandmother's sibling who is older than ego's father	表伯	"	
son of paternal grandfather's sister who is younger than ego's father	表叔	"	
son of paternal grandmother's sibling who is younger than ego's father	表叔	"	
daughter of paternal grandfather's sister	表姑	"	
daughter of paternal grandmother's sibling	表姑	"	
son of maternal grandfather's sister	表舅	"	
son of maternal grandmother's sibling	表舅	"	
daughter of maternal grandfather's sister	表姨	"	
daughter of maternal grandmother's sibling	表姨	"	
son of paternal (maternal) great-grandfather's brother who is older than ego's grandfather	(外)族伯祖父	first cousin twice removed	
son of paternal (maternal) great-grandfather's brother who is younger than ego's grandfather	(外)族叔祖父	"	
daughter of paternal great-grandfather's brother	(外)族祖姑	"	
children of father's brother's son's son	堂姪孫兒女		

Positions and Roles

- **Positions**: Actors that show a similar structure of relationships and are thus similarly embedded into the network. (e.g., b and g)
- **Roles**: The pattern of relationships of members of same or different positions.
- Note: Many of the category systems used by sociologists are based on **attributes** of individual actors that are common across actors.



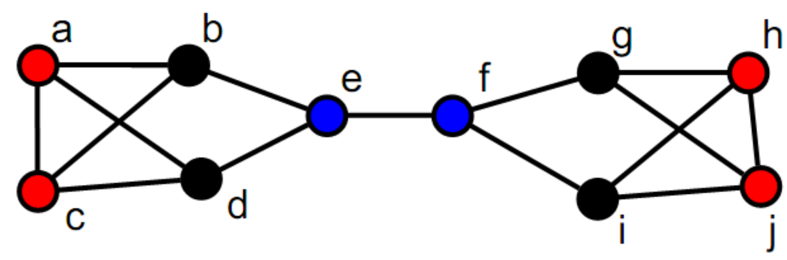
Similarity

- The idea of similarity has to be rather precisely defined
- Nodes are similar if they fall in the same **equivalence class**
  - We could come up with a equivalence class of out-degree of zero for example

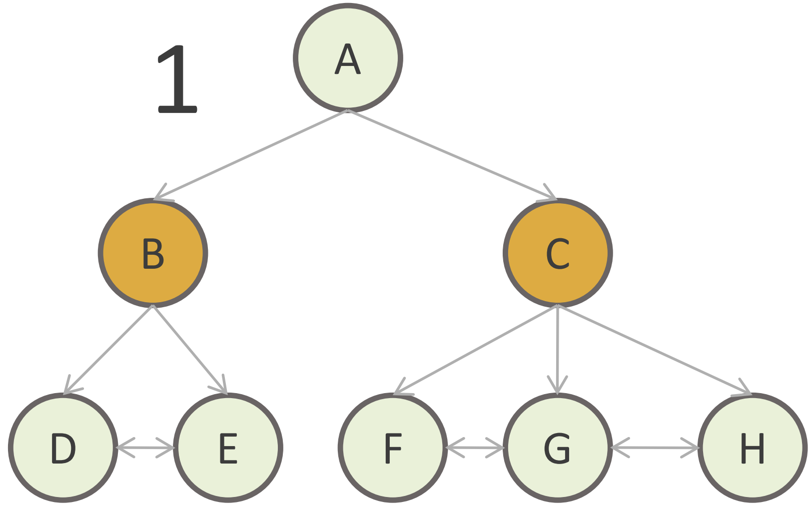
- There are **three particular definitions of equivalence**:
  - Structural Equivalence
  - Automorphic Equivalence (rarely used)
  - Regular Equivalence

Equivalence

- **Structural Equivalence:** Two structural equivalent actors could exchange their positions in a network **without changing their connections to the other actors** in the network.
- Structural equivalence is the strongest form of equivalence.
- Problem: Imagine two teachers in Toronto and Oslo. Rather than looking for connections to exactly the same persons we would like to find connection to similar persons but not exactly the same ones.
- **Regular Equivalence:** Two positions are considered as similar, if every **important** aspect of the observed structure applies (or does not apply) for both positions.

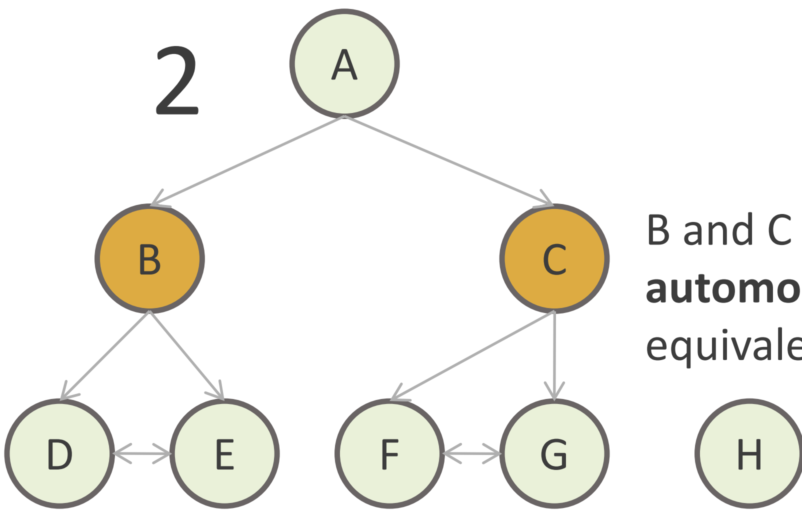


1



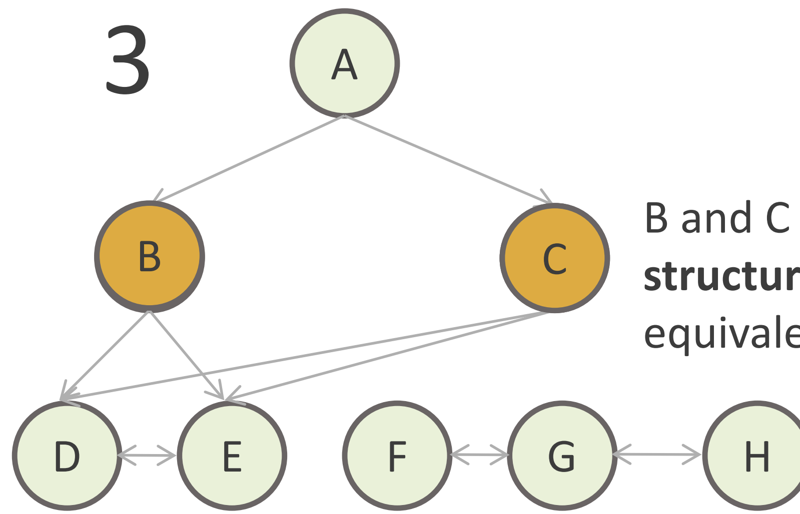
B and C are **regular** equivalent

2



B and C are **automorph** equivalent

3



B and C are **structural** equivalent

Measuring Similarity
Adjacency Matrix

	1 Coun	2 Comm	3 Educ	4 Indu	5 Mayr	6 WRO	7 News	8 UWay	9 Welf	10 West
1 Coun	---	1	0	0	1	0	1	0	1	0
2 Comm	1	---	1	1	1	0	1	1	1	0
3 Educ	0	1	---	1	1	1	1	0	0	1
4 Indu	1	1	0	---	1	0	1	0	0	0
5 Mayr	1	1	1	1	---	0	1	1	1	1
6 WRO	0	0	1	0	0	---	1	0	1	0
7 News	0	1	0	1	1	0	---	0	0	0
8 UWay	1	1	0	1	1	0	1	---	1	0
9 Welf	0	1	0	0	1	0	1	0	---	0
10 West	1	1	1	0	1	0	1	0	0	---

Pearson Correlation Coefficients, Covariances and Cross-Products (here for rows)

- Person correlation (ranges from -1 to +1) summarize pair-wise structural equivalence.

	1	2	3	4	5	6	7	8	9	10
1	1.000	0.447	-0.000	0.775	0.293	0.258	0.467	0.775	1.000	0.500
2	0.447	1.000	-0.447	0.447	0.655	0.293	0.333	0.745	0.333	0.378
3	-0.000	-0.447	1.000	0.258	-0.293	-0.149	0.600	-0.333	0.447	0.258
4	0.775	0.447	0.258	1.000	0.293	-0.258	0.745	0.775	0.775	0.775
5	0.293	0.655	-0.293	0.293	1.000	0.000	0.218	0.488	0.218	0.378
6	0.258	0.293	-0.149	-0.258	0.000	1.000	-0.447	-0.149	0.149	0.067
7	0.467	0.333	0.600	0.745	0.218	-0.447	1.000	0.600	0.745	0.258
8	0.775	0.745	-0.333	0.775	0.488	-0.149	0.600	1.000	0.600	0.149
9	1.000	0.333	0.447	0.775	0.218	0.149	0.745	0.600	1.000	0.600
10	0.500	0.378	0.258	0.775	0.378	0.067	0.258	0.149	0.600	1.000

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{N}\right)\left(\sum Y^2 - \frac{(\sum Y)^2}{N}\right)}}$$

Euclidean Squared Distances

Number of vertices that differ between two vertices. Euclidean or squared Euclidean distances are not sensitive to the linearity of association and can be used with valued or binary data.

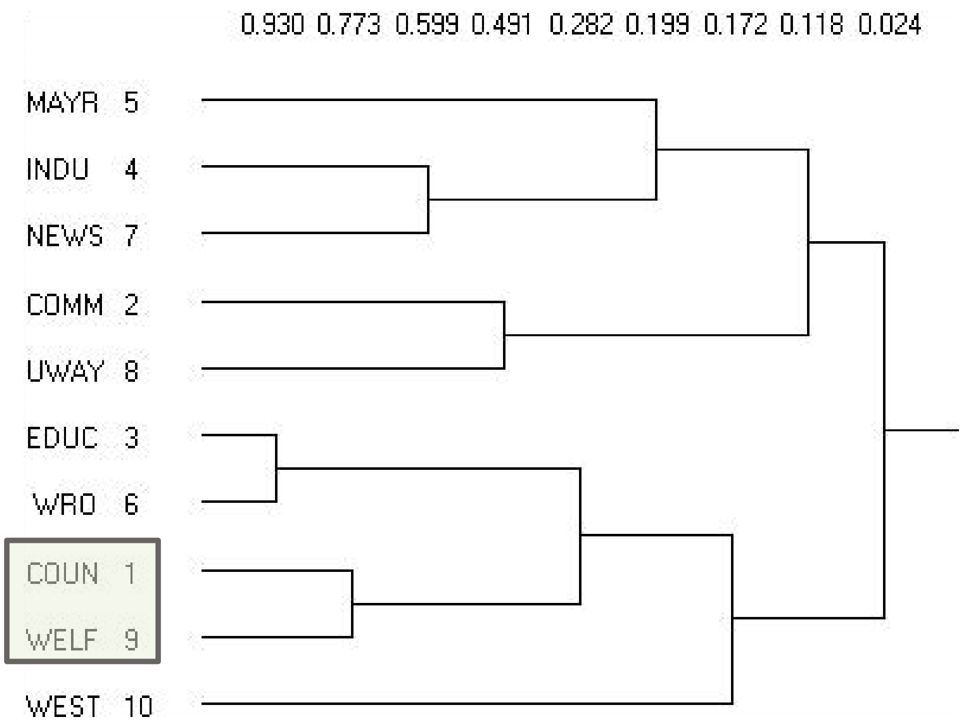
	1	2	3	4	5	6	7	8	9	10
1	0	2	2	1	2	2	1	1	0	1
2	2	0	2	2	1	2	2	1	2	2
3	2	2	0	2	2	2	1	2	2	2
4	1	2	2	0	2	2	1	1	1	1
5	2	1	2	2	0	2	2	1	2	2
6	2	2	2	2	2	0	2	2	2	2
7	1	2	1	1	2	2	0	1	1	2
8	1	1	2	1	1	2	1	0	1	2
9	0	2	2	1	2	2	1	1	0	1
10	1	2	2	1	2	2	2	2	1	0

Other similar measures can be Jaccard or Hamming distance.

$$d(x,y) = \|x-y\|_2 = \sqrt{(x_1-y_1)^2 + \cdots + (x_n-y_n)^2} = \sqrt{\sum_{i=1}^{n}(x_i-y_i)^2}$$

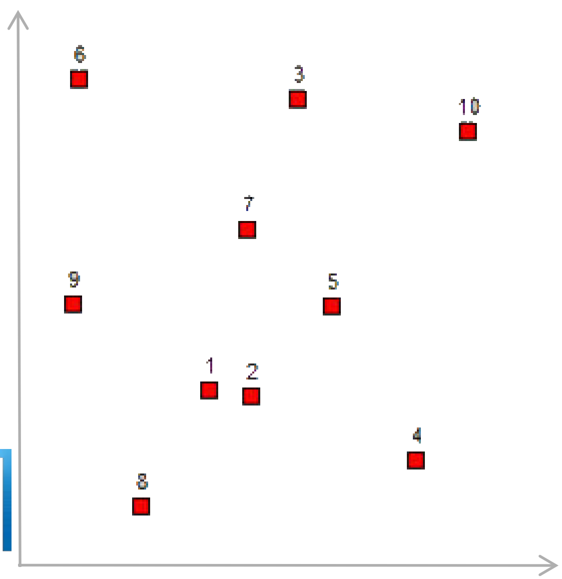
Hierarchical Clustering

- Hierarchical Clustering:
  - Initially places each case in its own cluster
  - The two most similar cases are then combined
  - This process is repeated until all cases are agglomerated into a single cluster (once a case has been joined it is never re-classsified)



Multi Dimensional Scaling

- MDS represents the patterns of similarity or dissimilarity in the profiles among the actors as a "map" in a multi-dimensional space. This map lets us see how "close" actors are and whether they "cluster".
  - Stress is a measure of badness of fit
  - The author has to determine the meaning of the dimensions



Non-metric MDS coordinates (stress = 0.161)

	1	2
1	-0.255	-0.452
2	0.004	-0.480
3	0.283	0.864
4	0.992	-0.774
5	0.478	-0.074
6	-1.038	0.962
7	-0.028	0.277
8	-0.667	-0.981
9	-1.070	-0.068
10	1.302	0.725

Stress = 0.161 in 22 iterations.

3 Network Mechanisms

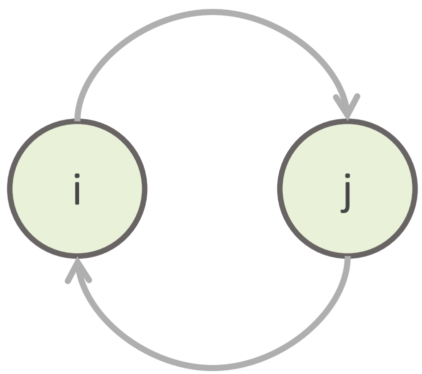
Outdegree Effect

- The most basic effect is defined by the outdegree of actor i. It represents the **basic tendency to have ties** at all.

- In a decision-theoretic approach this effect can be regarded as the **balance of benefits and costs of an arbitrary tie**.
  - Most networks are sparse (i.e., they have a density well below 0.5) which can be represented by saying that for a tie to an arbitrary other actor – arbitrary meaning here that the other actor has no characteristics or tie pattern making him/her especially attractive to i –, the costs will usually outweigh the benefits.

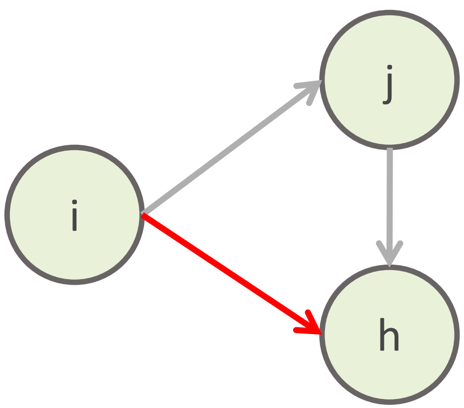
Reciprocity Effect

- Another quite basic effect is the tendency toward reciprocity, represented by the number of reciprocated ties of actor i. This is a basic feature of most social networks (e.g.. Wasserman and Faust, 1994, Chapter 13)

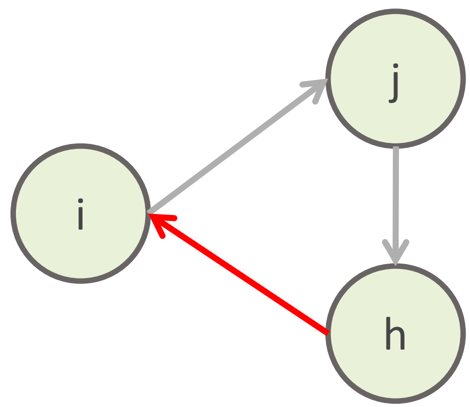


Transitivity and Other Triadic Effects

- Next to reciprocity, an essential feature in most social networks is the tendency toward transitivity, or transitive closure (**clustering**): friends of friends become friends, or in graph-theoretic terminology: two-paths tend to be, or to become, **closed** (e.g., Davis 1970, Holland and Leinhardt 1971).



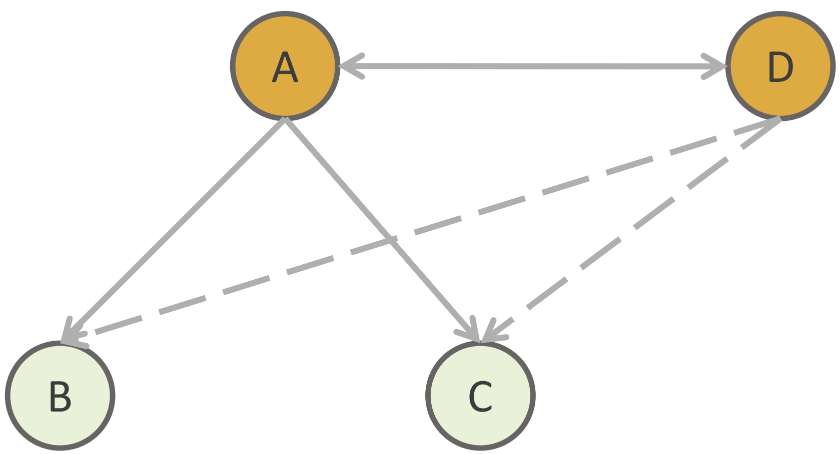
Transitive triplet



Three cycle

Balance Effect

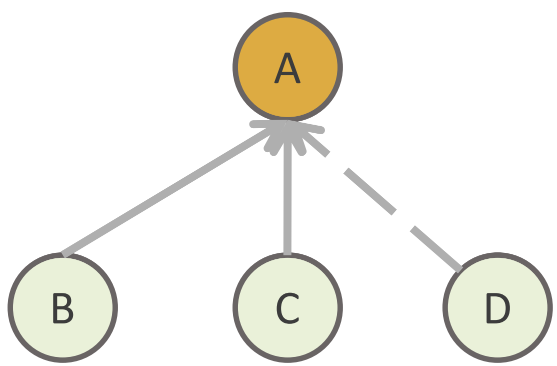
- An effect closely related to transitivity is balance (Newcomb, 1962), which is the same as **structural equivalence with respect to out-ties** (Burt, 1982), is the tendency to have and create ties to other actors who make the same choices as ego.



A and D become friends

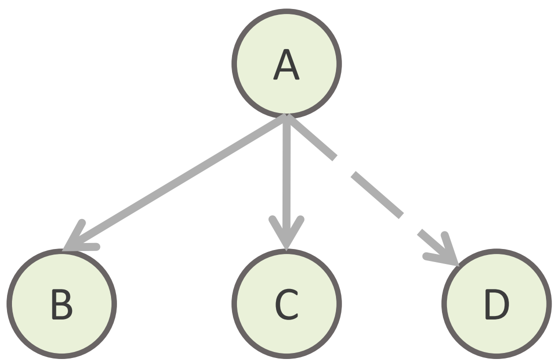
In/Out Popularity Effect

- The degree-related popularity effect is based on indegree or outdegree of an actor. Nodes with higher indegree, or higher outdegree, are more **attractive** for others to send a tie to.
- That implies that **high indegrees reinforce themselves**, which will lead to a relatively high dispersion of the indegrees (a Matthew effect in popularity as measured by indegrees, cf. Merton, 1968 and Price, 1976).



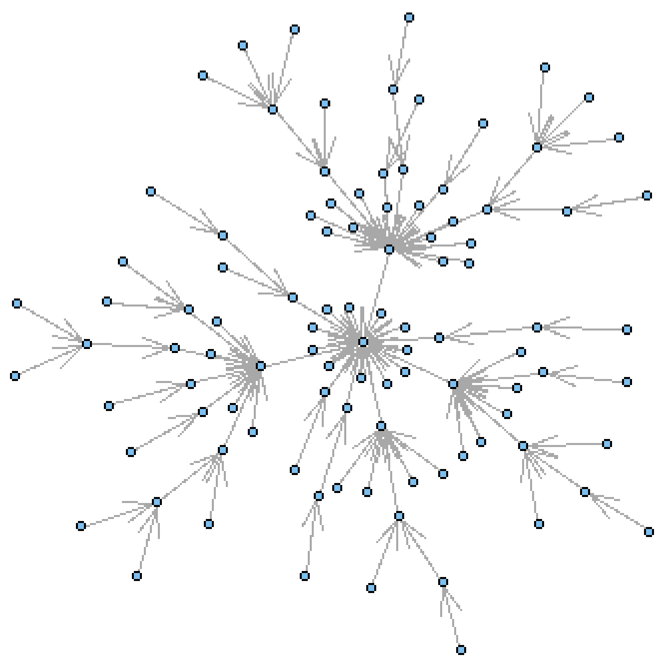
In/Out Activity Effect

- Nodes with higher indegree, or higher outdegree respectively, will have an extra propensity to **form ties** to others.
- The outdegree-related activity effect again is a self-reinforcing effect: when it has a positive parameter, the dispersion of outdegrees will tend to increase over time, or to be sustained if it already is high.



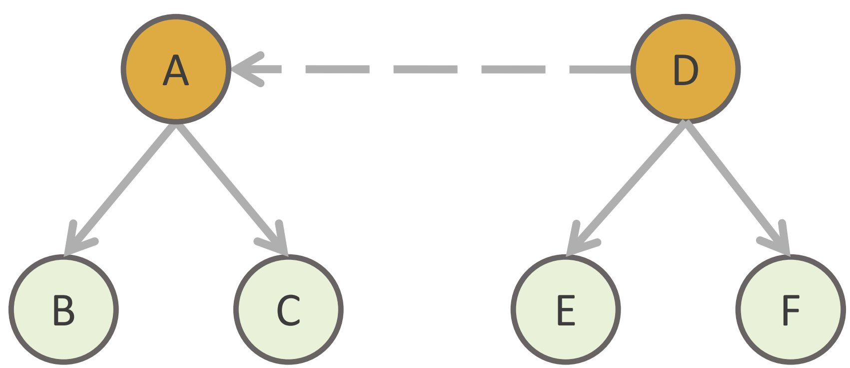
Preferential Attachment

- Notice: These four degree-related effects can be regarded as the analogues in the case of directed relations of what was called **cumulative advantage** by Price (1976) and **preferential attachment** by Barabasi and Albert (1999) in their models for dynamics of non-directed networks: a self-reinforcing process of degree differentiation.



In/Out Assortativity Effect

- Preferences of actors dependent on their degrees. Depending on their own out- and in-degrees, actors can have differential preferences for ties to others with also high or low out- and in-degrees (Morris and Kretzschmar 1995; Newman 2002)



Leaders become friends

Covariate Similarity Effect

- The covariate similarity effect, describes **whether ties tend to occur more often between actors with similar values on a value** (homophily effect). Tendencies to homophily constitute a fundamental characteristic of many social relations, see McPherson, Smith-Lovin, and Cook (2001).

- **Example**: Ipad Owners tend to be friends with other Ipad owners.

Application: ERGM

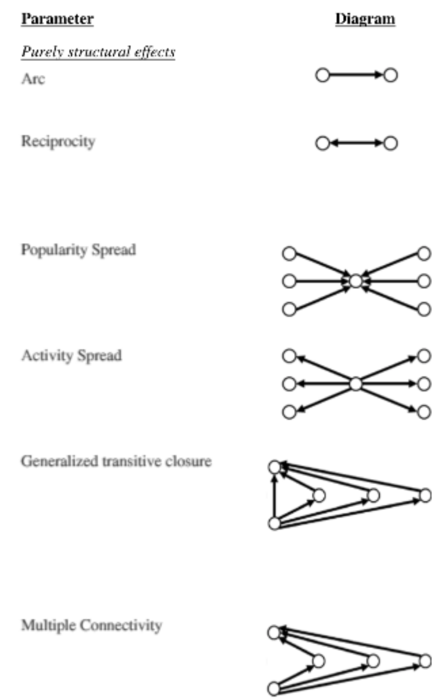
- Allow to test network mechanisms in a regression-type fashion

UNDERSTANDING NETWORK FORMATION IN STRATEGY RESEARCH: EXPONENTIAL RANDOM GRAPH MODELS

JI YOUN (ROSE) KIM,[1] MICHAEL HOWARD,[2]* EMILY COX PAHNKE,[3] and WARREN BOEKER[3]

Parameter	Model 1 co-variates–only (no structural term)	Model 2 co-variates and structural term
Purely structural effects		
Arc	−13.23***	−10.39***
Reciprocity	—	2.72***
Popularity spread	—	−0.07
Activity spread	—	−0.57
Generalized transitive closure	—	0.83***
Multiple connectivity	—	−0.12+
Actor relation effects		
Sender (firm size)	0.49***	0.35***
Sender (profitability)	8.46***	6.25**
Sender (prior alliance activities)	0.002	0.002
Sender (market uncertainty)	1.56	1.61
Receiver (firm size)	0.23**	0.15+
Receiver (profitability)	4.81*	3.42
Receiver (prior alliance activities)	0.001	0.001
Receiver (market uncertainty)	−3.17+	−3.01+
Homophily (state)	1.08***	0.79***
Homophily (firm size)	−0.09	−0.05
Co-variate network		
Prior interlocks	0.74**	0.59*
Akaike information criterion (AIC) goodness of fit	1,264	1,196

$+p < 0.1$; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.



thank you for your time

Norwegian Business School (BI)
Nordic Center for Internet & Society
Nydalsvn. 37 / N-0442 Oslo
bi.edu/ncis @BI_NCIS