SNA

Measures of similarity and structural equivalence

Similarity/distance

- Local
  - Adjacent, geodesic
    - Clique, co-authorship, direct citation
- Structure
  - Consider overall pattern
    - Co-citation, bibliographic coupling
  - Many variations
    - Co-occurrence
    - Co-relation
    - Euclidean distance

Similarity and equivalent

- Features-based similarity
  - Similar in terms of sharing attributes/features
  - Two mode-one mode
- Relational similarity
  - Who you interact with
  - Similar in terms of sharing network neighbors

Similar in terms of sharing attributes/features based

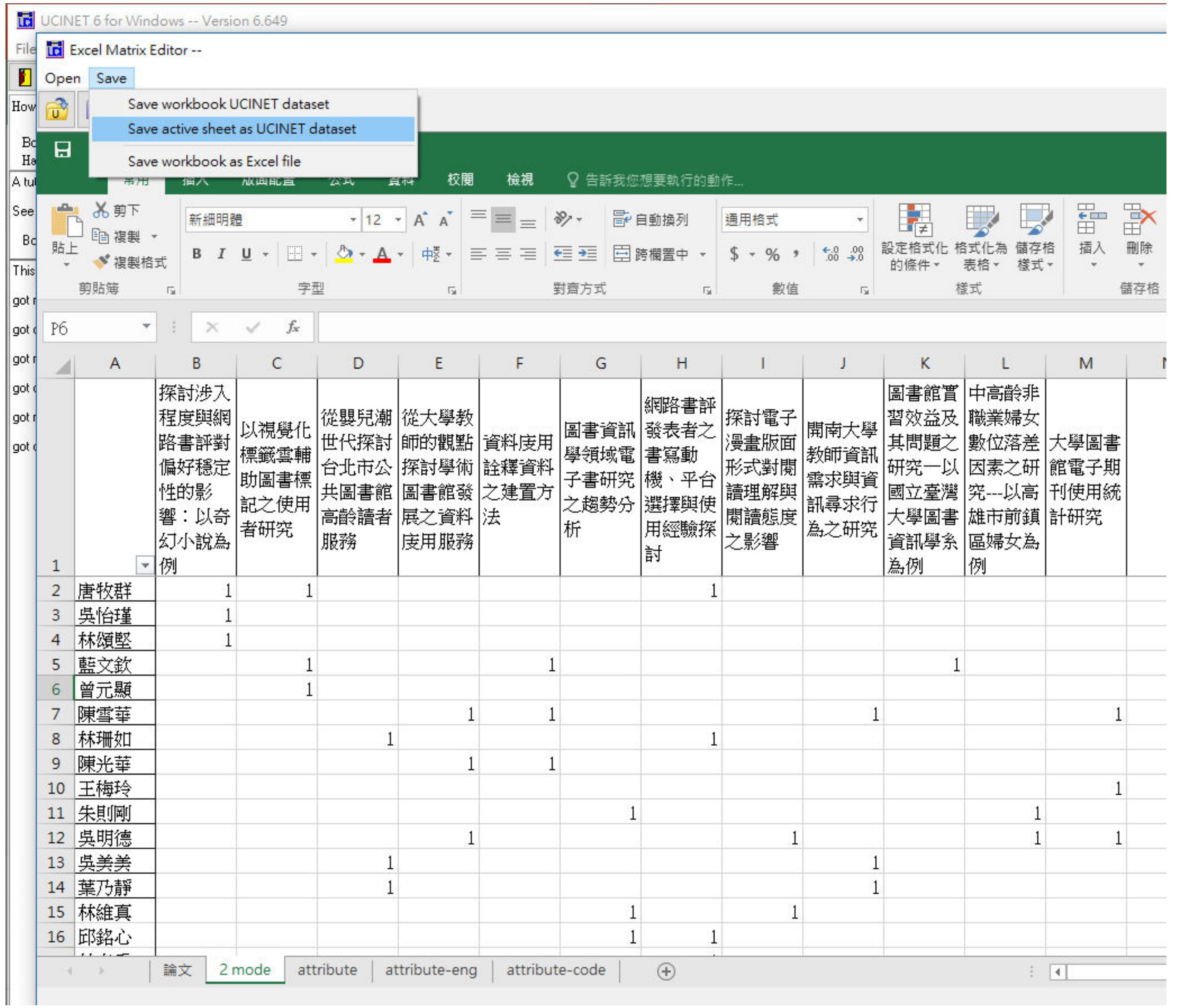
Movies

Customers

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
a					4	5			4		3					2			4		2				4
b			4							3							5	1							
c		5		4			4						3		5										
d												5							4		2				
e							5			4	5														
f												4													5
g	2	4									5							2							
h					1																				
i																								3	3
j			4										1		5										
k		5				4			2		5		1	5						4					
l					3			3				4	1		4		4							3	4
m	5		3						3			4			5						5	4			
n			1			5				4			1					3					3		4
o			4			4				5		4		5			4	2						3	
p							5														5				
q													1									4		3	4
r		4				4							2												
s			2				4						1			4							5		
t		1												5			4			4				3	3
u			2						3				1			4						4			
v					4					4			5			2								5	4
w									3			5			4			4			3				4
x	4									3										1					1
y			1							2	3						3	3							

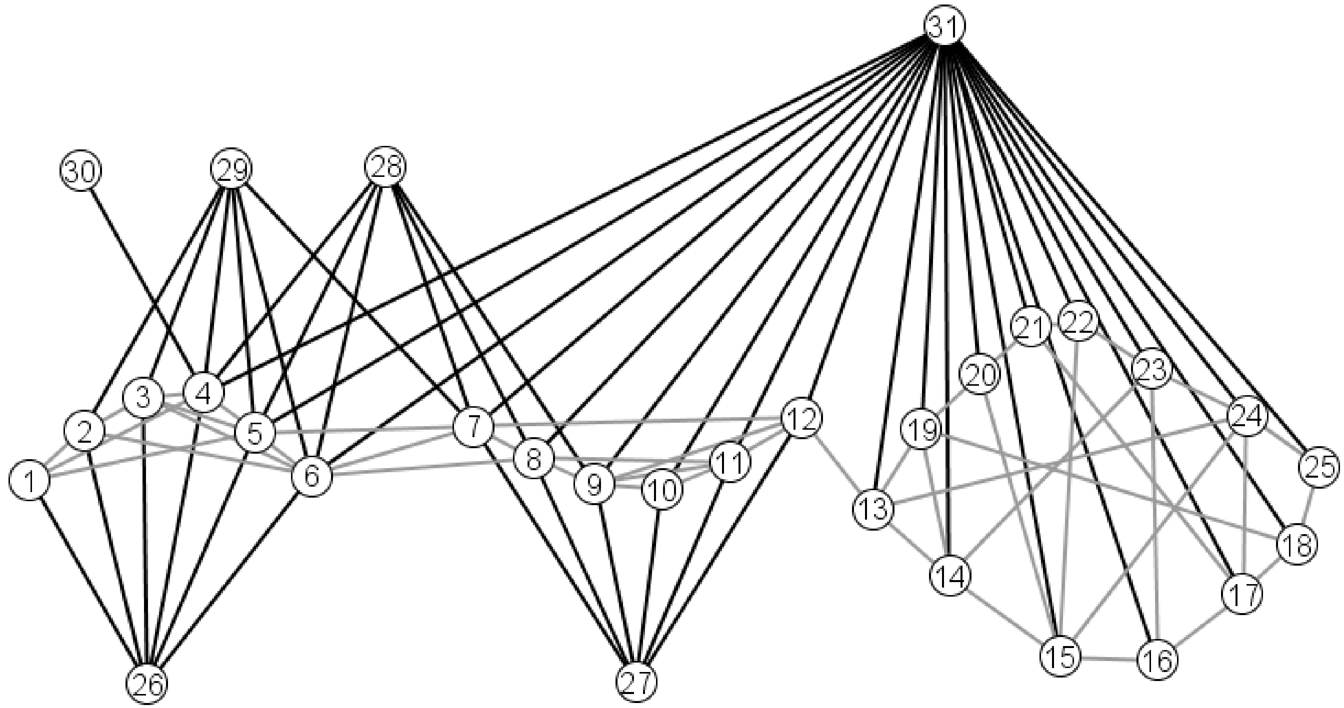
Two-mode-one mode with UCINET

- NTU thesis data



Similarity clustering

- Similarity networks
  - [Thesaurus]
  - [Music discovery]
- As opposed to adjacency matrix
- Two steps
  - Determine similarity methods
    - Correlation, distance, matches
  - Determine clustering/visualization methods
    - E.g. HAC/dendrogram, multidimensional scaling



For #26, who would you suggest to be her Facebook friends?
If using Jaccard coefficient

Source: Bruce Hoppe "Introduction to Network Mathematics"

Node y	N(y) = Neighborhood of y	SE(x,y) = J(N(x),N(y))
27	{7,8,9,10,11,12}	J(N(x),N(y)) = |{1,2,3,4,5,6} ∩ {7,8,9,10,11,12}| |{1,2,3,4,5,6} ∪ {7,8,9,10,11,12}|= 0/12; Not at all similar
28	{4,5,6,7,8,9}	J(N(x),N(y)) = |{1,2,3,4,5,6} ∩ {4,5,6,7,8,9}| |{1,2,3,4,5,6} ∪ {4,5,6,7,8,9}| = 3/9; Somewhat similar
29	{2,3,4,5,6,7}	J(N(x),N(y)) = 5/7; Quite similar
30	{4}	J(N(x),N(y)) = 1/6
31	{4,5,6,7, ... , 25}	J(N(x),N(y)) = 3/25

Position analysis

- In contrast to relational or cohesive approach, which finds subsets of actors who are strongly or closely related to each other
- Partition actors into mutually exclusive classes of equivalent actors who have similar relational patterns

Position and structural equivalence

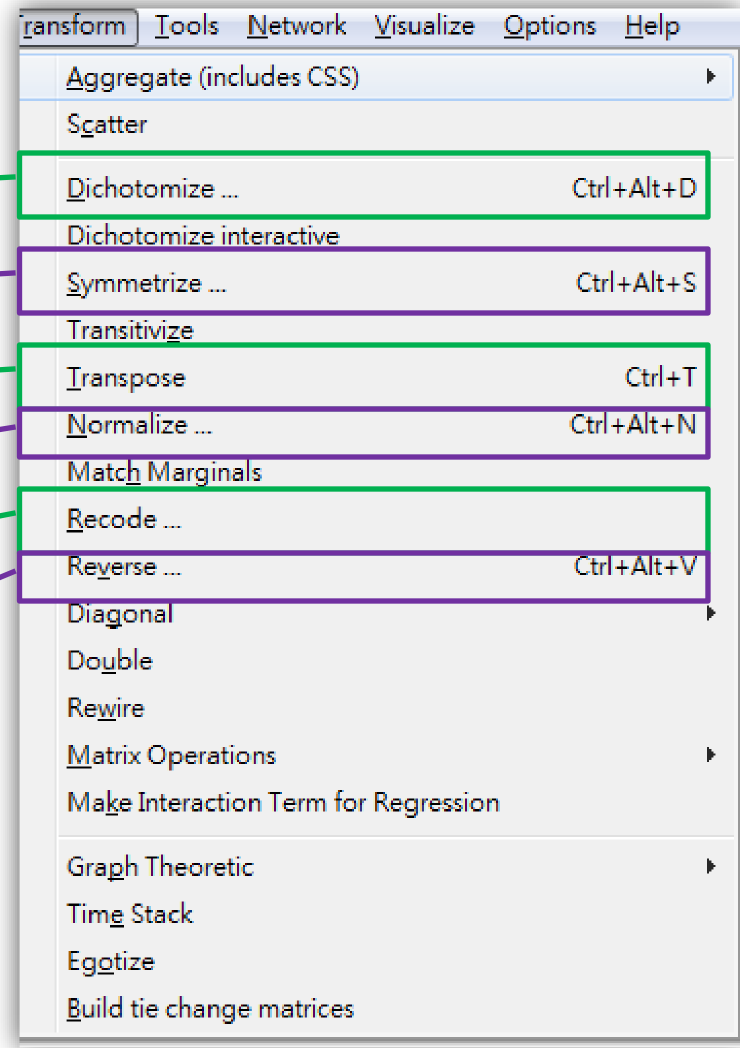
- The relationship approach to social structure, which emphasizes that what defines a role, such as that of nurse, is precisely the characteristic set of relationships that actors who are nurses have with actors who are doctors, patients, suppliers, secretaries, other nurses, and so on...
- Your social role is defined by whom you interact with

Similarity/Structural equivalence

- The notion of structural equivalence focuses on the *profile of relations* that actors have with other actors in the network.
- Two actors are said to be structurally equivalent to the extent that their relational profiles with all other actors in the network are identical. This applies both to outgoing ties and to incoming ones.
- In most real cases, it is unrealistic to expect full structural equivalence between actors. Rather, the idea is to measure the *extent of structural equivalence*.

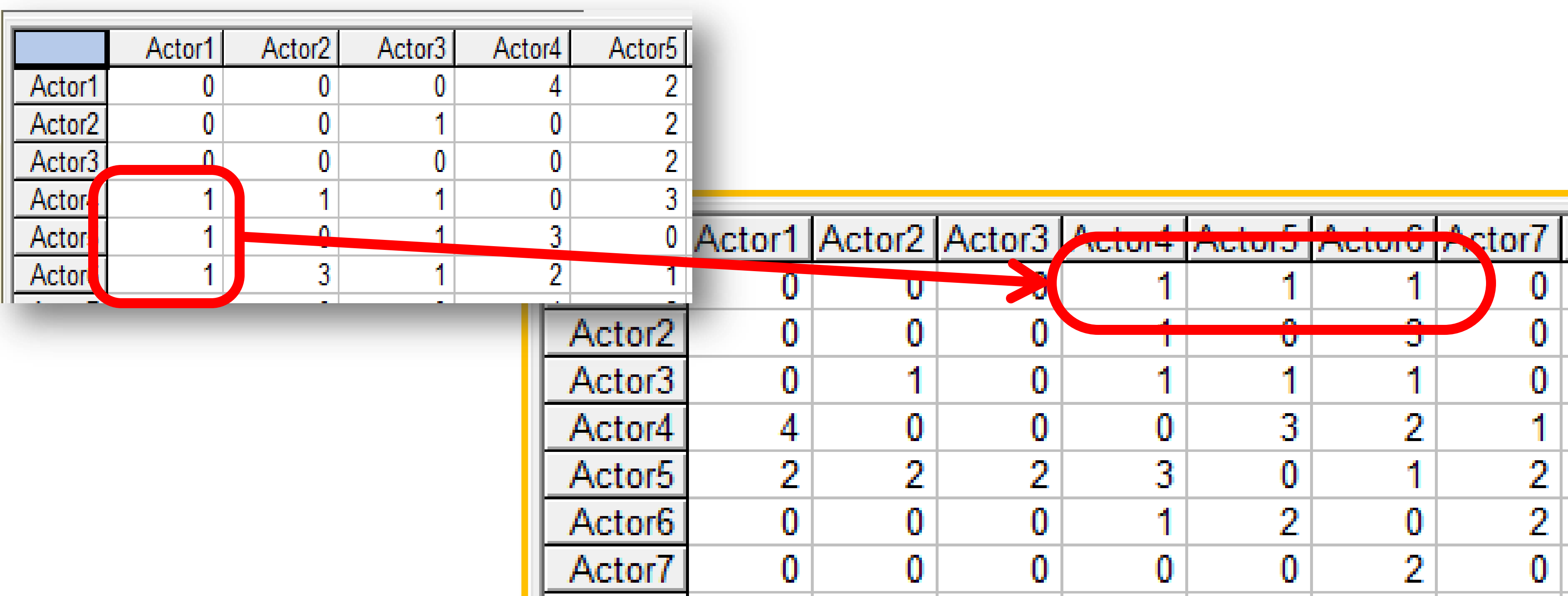
Transforming a matrix

- Transforming data values
  - 二元化
  - 對稱
  - 互換
  - 標準化
  - 代換
  - 相似-距離反轉



Transposing (for directional network)

- Interchanges the rows and columns of a matrix.
- Choose input dataset and output dataset

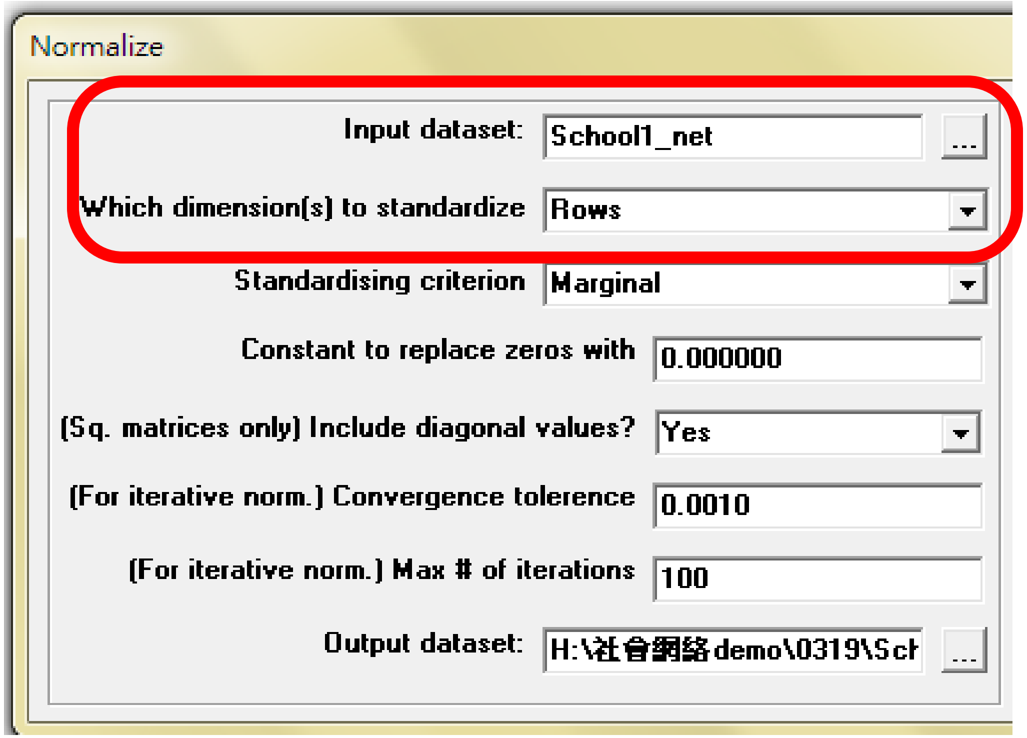


	Actor1	Actor2	Actor3	Actor4	Actor5
Actor1	0	0	0	4	2
Actor2	0	0	1	0	2
Actor3	0	0	0	0	2
Actor	1	1	1	0	3
Actor	1	0	1	3	0
Actor	1	3	1	2	1

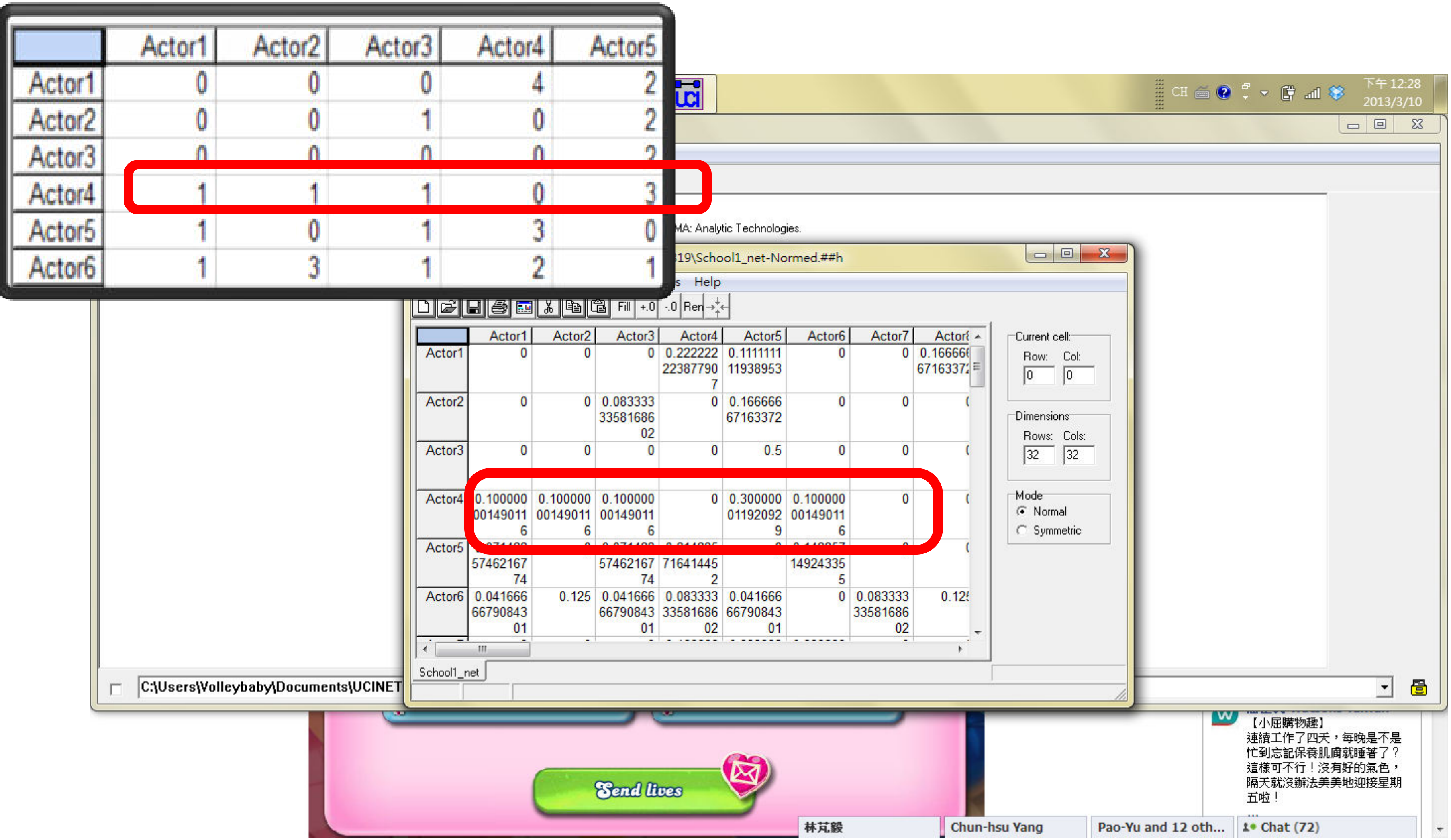
	Actor1	Actor2	Actor3	Actor4	Actor5	Actor6	Actor7
	0	0	0	1	1	1	0
Actor2	0	0	0	1	0	3	0
Actor3	0	1	0	1	1	1	0
Actor4	4	0	0	0	3	2	1
Actor5	2	2	2	3	0	1	2
Actor6	0	0	0	1	2	0	2
Actor7	0	0	0	0	0	2	0

Normalize

- Normalize the values in a matrix.
- Choose dimension
  - Rows or
  - Columns
- Choose criterion
  - Marginal
- Choose output



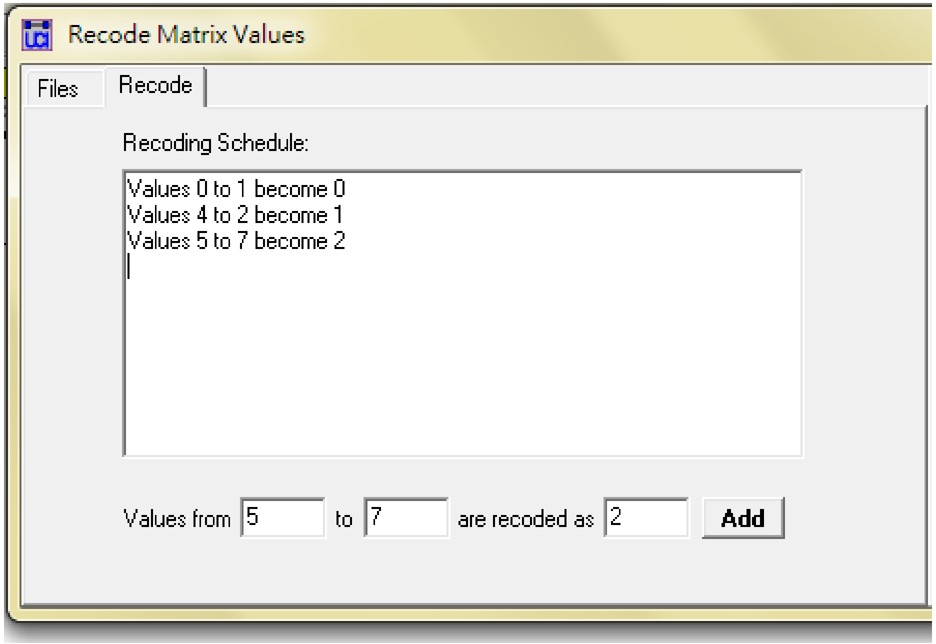
Normalize



	Actor1	Actor2	Actor3	Actor4	Actor5
Actor1	0	0	0	4	2
Actor2	0	0	1	0	2
Actor3	0	0	0	0	2
Actor4	1	1	1	0	3
Actor5	1	0	1	3	0
Actor6	1	3	1	2	1

Record

- The routine allows the user to change values or a range of values in a matrix to a new value.
- Click “transform” then
- “Record”, set rules
- Choose output
- Suggest: save the text file



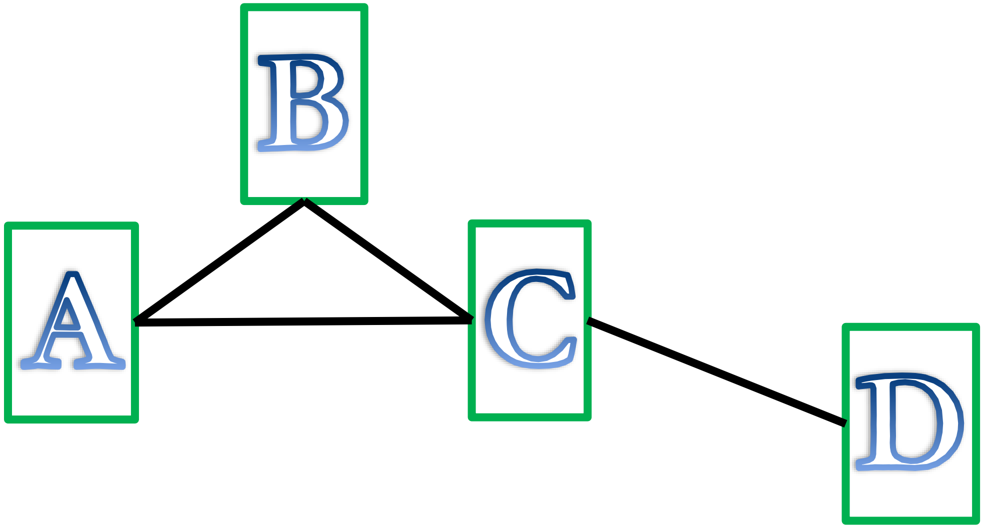
Record

```
Output dataset:                    School1_net-Rec (H:\���|���demo\0319\School1_net-Rec)

Recoding Schedule:

Values 0 to 0 become 0
Values 1 to 1 become 0
Values 2 to 2 become 1
Values 3 to 3 become 1
Values 4 to 7 become 2
```

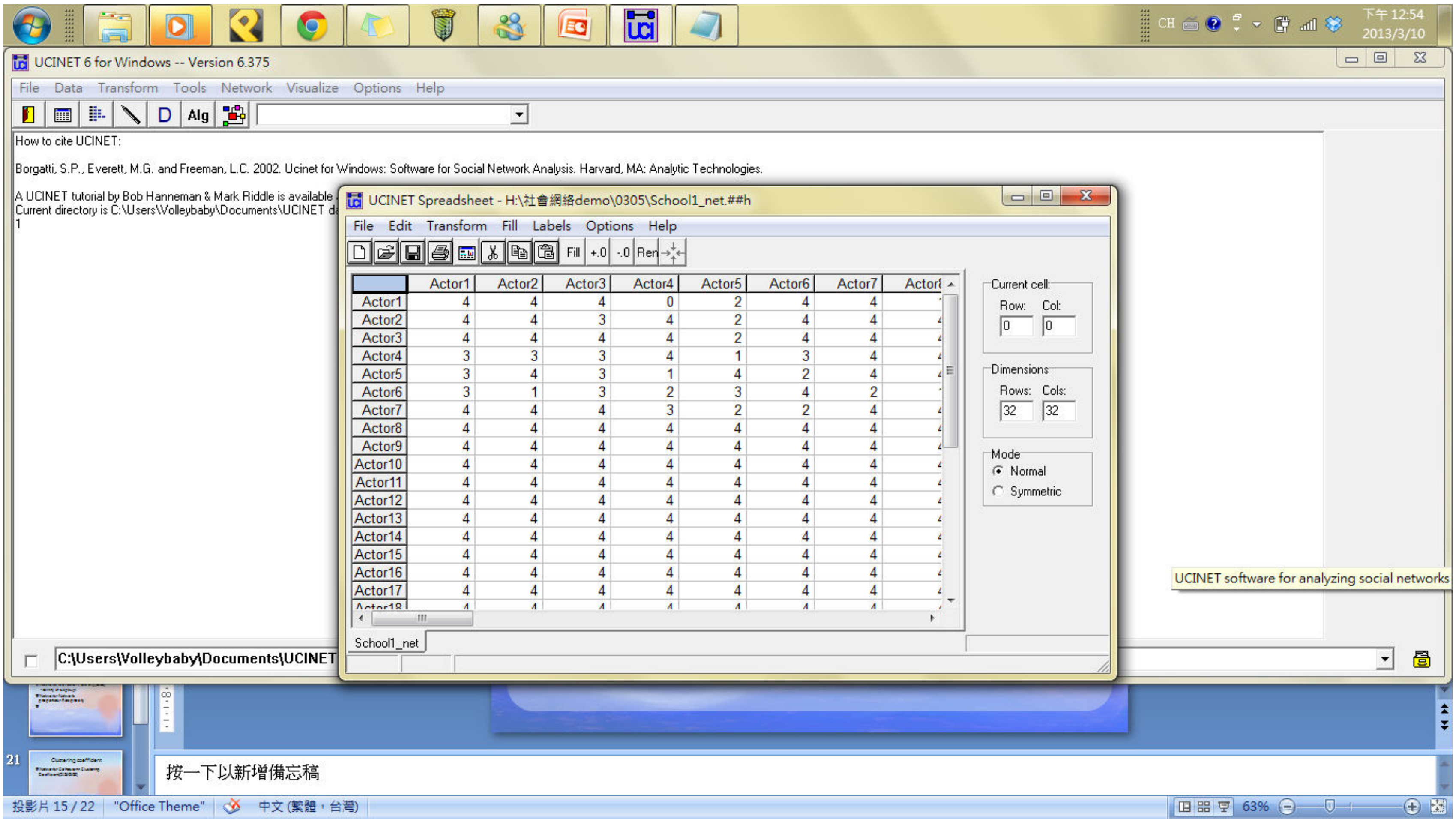
	Actor1	Actor2	Actor3	Actor4	Actor5	Actor6	Actor7	Actor8
Actor1	0	0	0	2	1	0	0	1
Actor2	0	0	0	0	1	0	0	0
Actor3	0	0	0	0	1	0	0	0
Actor4	0	0	0	0	1	0	0	0
Actor5	0	0	0	1	0	1	0	0
Actor6	0	1	0	1	0	0	1	1
Actor7	0	0	0	0	1	1	0	0
Actor8	0	0	0	0	0	0	0	0

Reverse

- Convert similarity data to distance data, or distance to similarity by a linear transformation.
  - A→B=1 step
  - B→D=2 steps
  - A & B are more similar than B & D
- Choose input
- Choose output



Reserve


UCINET 6 for Windows -- Version 6.375
File Data Transform Tools Network Visualize Options Help
How to cite UCINET:
Borgatti, S.P., Everett, M.G. and Freeman, L.C. 2002. Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.
UCINET Spreadsheet - H:\社會網絡demo\0305\School1_net.##h
File Edit Transform Fill Labels Options Help
Actor1 Actor2 Actor3 Actor4 Actor5 Actor6 Actor7
Actor1 4 4 4 0 2 4 4
Actor2 4 4 3 4 2 4 4
Actor3 4 4 4 4 2 4 4
Actor4 3 3 3 4 1 3 4
Actor5 3 4 3 1 4 2 4
Actor6 3 1 3 2 3 4 2
Actor7 4 4 4 3 2 2 4
Actor8 4 4 4 4 4 4 4
Actor9 4 4 4 4 4 4 4
Actor10 4 4 4 4 4 4 4
Actor11 4 4 4 4 4 4 4
Actor12 4 4 4 4 4 4 4
Actor13 4 4 4 4 4 4 4
Actor14 4 4 4 4 4 4 4
Actor15 4 4 4 4 4 4 4
Actor16 4 4 4 4 4 4 4
Actor17 4 4 4 4 4 4 4
Current cell:
Row: Col:
0 0
Dimensions
Rows: Cols:
32 32
Mode
Normal
Symmetric
School1_net
UCINET software for analyzing social networks
C:\Users\Volleybaby\Documents\UCINET
按一下以新增備忘稿
投影片 15 / 22
"Office Theme"
中文 (繁體，台灣)
63%
下午 12:54
2013/3/10

Exercise

- File: KNOKBUR.##h
  - First run Geodesic distance
  - Output > KNOKBUR-geo
  - Visualize it, then
  - Try transform KNOKBUR-geo
    - Record
    - Reverse

Average: 1.5
Std Dev: 0.6

	1 COUN	2 COMM	3 EDUC	4 INDU	5 MAYR	6 WRO	7 NEWS	8 UWAY	9 WELF	10 WEST
1 COUN	0	1	2	2	1	3	1	2	1	2
2 COMM	1	0	1	1	1	2	1	1	1	2
3 EDUC	2	1	0	1	1	1	1	2	2	1
4 INDU	1	1	2	0	1	3	1	2	2	2
5 MAYR	1	1	1	1	0	2	1	1	1	1
6 WRO	3	2	1	2	2	0	1	3	1	2
7 NEWS	2	1	2	1	1	3	0	2	2	2
8 UWAY	1	1	2	1	1	3	1	0	1	2
9 WELF	2	1	2	2	1	3	1	2	0	2
10 WEST	1	1	1	2	1	2	1	2	2	0

For a directed graph

Two actors are structurally equivalent to the extent that the profile of scores in their rows and columns are similar

Figure 13.2. Adjacency matrix for Knoke information network

Sending

	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	---

Receiving

Figure 13.3. Concatenated row and column adjacencies for Knoke information network

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

Coun's interaction profile, including both sending and receiving ties (transpose)
Or you can look either sending or receiving relationship, instead of both
Can only be done by Profile analysis with UCINET

Distance measure (directed graph)

Figure 13.3. Concatenated row and column adjacencies for Knoke information network

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

Exercise

- Use KNOKBUR.##h
  - Transform the data with "transpose" command
  - With the new data, run similarity analysis under TOOL
  - Again, under tool, run clustering analysis, use Q value to determine the proper clustering level
  - Perform also Newman community detection and Modularity (Leuven method)
  - Save the file as excel/csv and open it with Gephi so you can perform modularity analysis and visualize it.
  - Compare the results with actor by actor shared clique method we performed last week (first you need to transform the directional data into reciprocal, i.e. using minimal or product method).

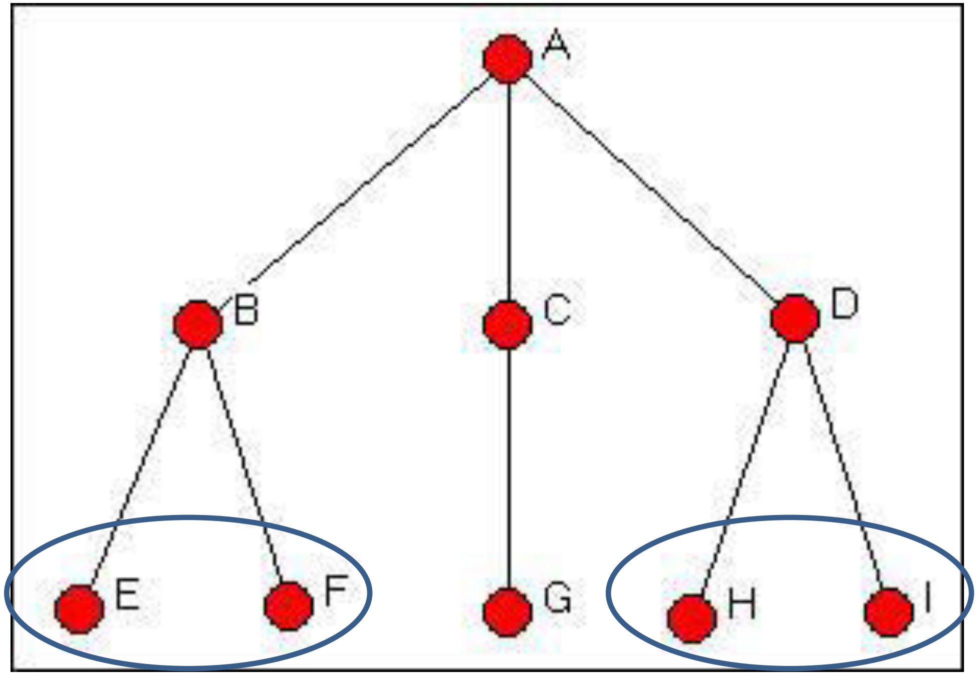
Structural equivalence

- A set of nodes connected by the same relations to exactly the same people
- An actor's position is defined by who s/he is connected to

“Perfect” structurally equivalent

- In a directed binary graph, two actors are perfectly structurally equivalent in that specific relation if they have exactly identical patterns of ties sent to and received from all other network actors.

How many perfect structurally equivalent classes?



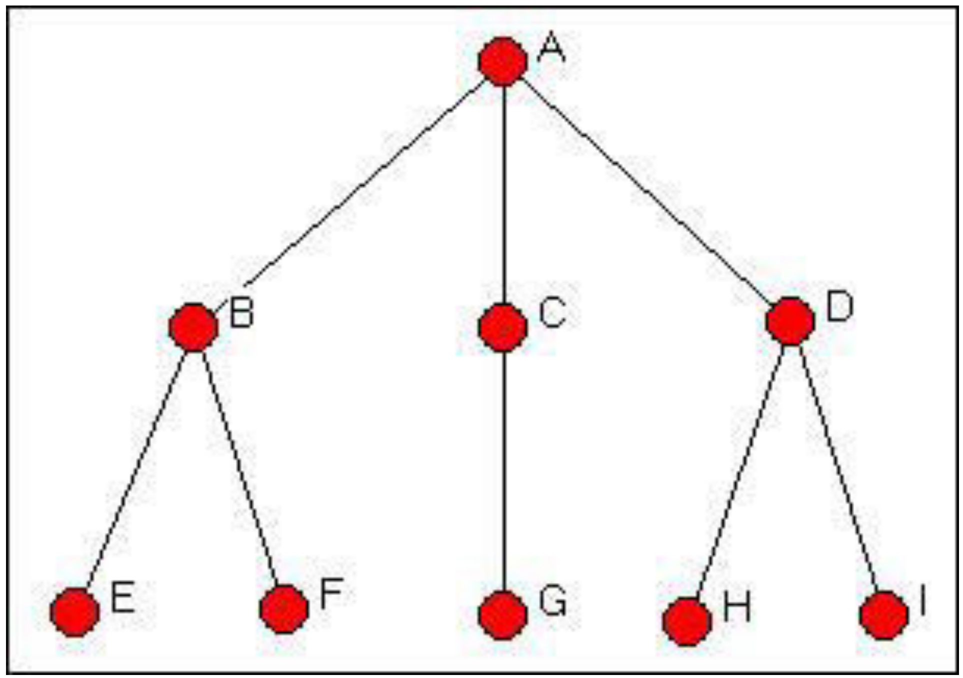
Structural equivalence

- The definition is too strict to be useful, thus we use measure of "relations similarity" to capture the equivalence between any two nodes
- Two actors are considered structural equivalent if they share many of the same network neighbors

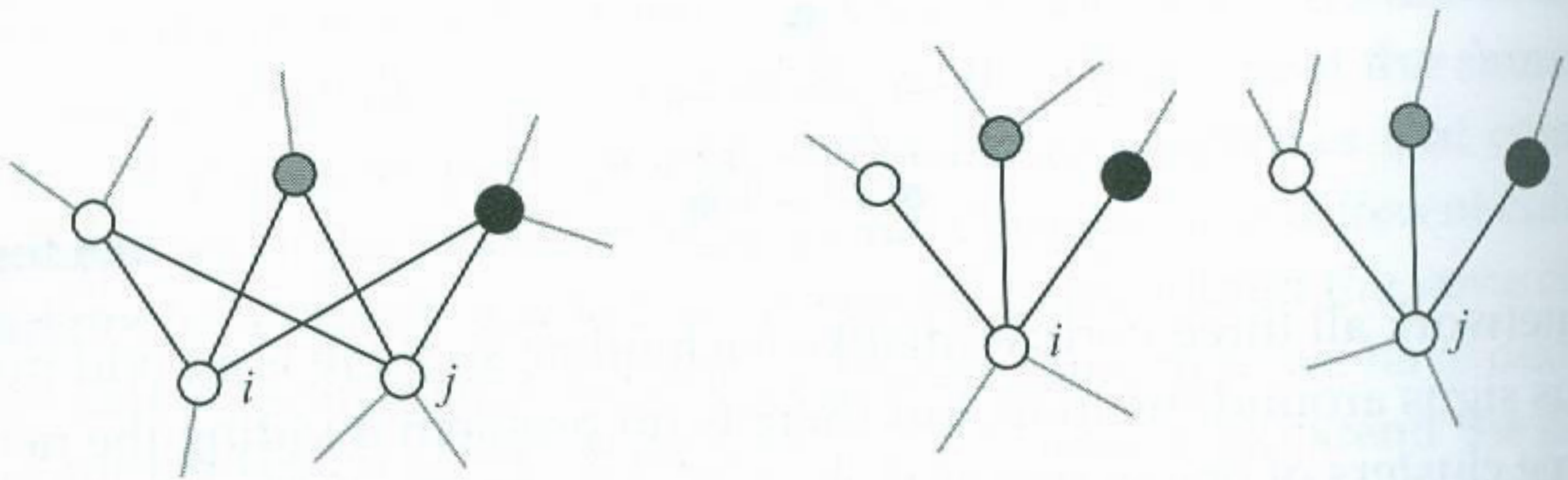
Regular equivalence

- Two nodes are said to be regularly equivalent if they have the same profile of ties with members of other sets of actors that are also regularly equivalent.
- Actors that are regularly equivalent do not necessarily fall in the same network position with respect to other individual actors; rather, they have the same kinds of relationships with different set of actors

How many regularly equivalent classes?



Position analysis



(a) Structural equivalence

(b) Regular equivalence

Operationalization

- Actors are then classified based on how similar their interaction profiles are
- Transform the original adjacency data into actor-by-actor matrix of similarity/distance measures

Tools>Similarities

*Correlation, row or column

Each cell stands for strength of correlation between two pairs

Pearson correlations of rows (sending) for Knoke Information Network

	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

Similarity measures

- Cosine measure
- Correlation
  - E.g. Correlation Coefficients, covariance
- Distance (dissimilarity measure)
  - E.g. Euclidean distance
- Association (matching-type measures of similarity)
  - Concurrence count
  - E.g. The Jaccard coefficient

Similarity measures

- Valued relations
  - Pearson correlation/ covariance
  - Cosine/cross-product
  - Euclidian distances (as opposed to geodesic)
  - Actually they can also be applied to binary relations
- Binary relations
  - Exact matches
  - Jaccard coefficient
  - Hamming distance/similarity

Tools>Similarities>Matches

Measuring the similarity of two actors by counting the shared connections as the percentage of the possible total (proportion of matches)

COUN and COMM shared 62.5 percent of their links

		1 COUN	2 COMM	3 EDUC	4 INDU	5 MAYR	6 WRO	7 NEWS	8 UWAY	9 WELF	10 WEST
1	COUN	1.000	0.625	0.625	0.625	0.625	0.250	0.625	0.750	0.625	0.500
2	COMM	0.625	1.000	0.250	0.625	1.000	0.125	0.875	0.250	0.375	0.375
3	EDUC	0.625	0.250	1.000	0.500	0.250	0.625	0.500	0.750	0.625	0.750
4	INDU	0.625	0.625	0.500	1.000	0.625	0.500	0.500	0.750	0.500	0.625
5	MAYR	0.625	1.000	0.250	0.625	1.000	0.125	0.875	0.250	0.375	0.250
6	WRO	0.250	0.125	0.625	0.500	0.125	1.000	0.125	0.625	0.375	0.875
7	NEWS	0.625	0.875	0.500	0.500	0.875	0.125	1.000	0.250	0.625	0.250
8	UWAY	0.750	0.250	0.750	0.750	0.250	0.625	0.250	1.000	0.750	0.750
9	WELF	0.625	0.375	0.625	0.500	0.375	0.375	0.625	0.750	1.000	0.375
10	WEST	0.500	0.375	0.750	0.625	0.250	0.875	0.250	0.750	0.375	1.000

Measure: HAMMING-SIM
Variables are: COLUMNS
Diagonal: TREATED AS MISSING
Input dataset: KNOKBUR (C:\Program Files (x86)\Ana
Similarity matrix: KNOKBUR-Sim (C:\Program Files (x86)

	1 COUN	2 COMM	3 EDUC	4 INDU	5 MAYR	6 WRO	7 NEWS	8 UWAY	9 WELF	10 WEST
1 COUN	9.000	5.000	5.000	5.000	5.000	2.000	5.000	6.000	5.000	4.000
2 COMM	5.000	9.000	2.000	5.000	8.000	1.000	7.000	2.000	3.000	3.000
3 EDUC	5.000	2.000	9.000	4.000	2.000	5.000	4.000	6.000	5.000	6.000
4 INDU	5.000	5.000	4.000	9.000	5.000	4.000	4.000	6.000	4.000	5.000
5 MAYR	5.000	8.000	2.000	5.000	9.000	1.000	7.000	2.000	3.000	2.000
6 WRO	2.000	1.000	5.000	4.000	1.000	9.000	1.000	5.000	3.000	7.000
7 NEWS	5.000	7.000	4.000	4.000	7.000	1.000	9.000	2.000	5.000	2.000
8 UWAY	6.000	2.000	6.000	6.000	2.000	5.000	2.000	9.000	6.000	6.000
9 WELF	5.000	3.000	5.000	4.000	3.000	3.000	5.000	6.000	9.000	3.000
10 WEST	4.000	3.000	6.000	5.000	2.000	7.000	2.000	6.000	3.000	9.000

Cronbach's Alpha = 1.082

Hamming similarity

	1 C	2 C	3 E	4 I	5 M	6 W	7 N	8 U	9 W	10 W
1 COUN	0	3	3	3	3	6	3	2	3	4
2 COMM	3	0	6	3	0	7	1	6	5	5
3 EDUC	3	6	0	4	6	3	4	2	3	2
4 INDU	3	3	4	0	3	4	4	2	4	3
5 MAYR	3	0	6	3	0	7	1	6	5	6
6 WRO	6	7	3	4	7	0	7	3	5	1
7 NEWS	3	1	4	4	1	7	0	6	3	6
8 UWAY	2	6	2	2	6	3	6	0	2	2
9 WELF	3	5	3	4	5	5	3	2	0	5
10 WEST	4	5	2	3	6	1	6	2	5	0

Hamming distance

Correlation

- Tools>similarity/dissimilarity

	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

Figure 13.4. Pearson correlations of rows (sending) for Knoke information network

Correlation Coefficient r

The measure of linear association between 2 variables x and y
Use mostly frequency when calculating person-person similarity in collaborative filtering

$$\text{cov}(X, Y) = \sum_{i=1}^{N} \frac{(x_i - \bar{x})(y_i - \bar{y})}{N}. \quad R = \frac{\sum_i [(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{\sum_i [(x_i - \bar{x})^2] \sum_i [(y_i - \bar{y})^2]}}$$

Insensitive to differences in the magnitude of the variables used to compute the coefficient
As opposed to inner product and covariance

$R = \text{cov}/SD_x \, SD_y$

Cosine measure

- a measure of similarity between two vectors of *n* dimensions by finding the cosine of the angle between them. Given two vectors of attributes, *A* and *B*, the cosine similarity, θ, is represented using a dot product and magnitude as

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|}.$$

Cosine similarity

- Numerical

$$\sigma_{ij} = \cos\theta = \frac{\sum_k A_{ik}A_{kj}}{\sqrt{\sum_k A_{ik}^2}\sqrt{\sum_k A_{jk}^2}}.$$

- Binary

$$\sigma_{ij} = \frac{\sum_k A_{ik}A_{kj}}{\sqrt{k_i k_j}} = \frac{n_{ij}}{\sqrt{k_i k_j}}.$$

Jaccard coefficient, binary data

- The Jaccard coefficient measures similarity between sample sets, and is defined as the size of the intersection divided by the size of the union of the sample sets (Salton's method, p.80):

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}.$$

$$d_J(A,B) = 1 - J(A,B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}.$$

Jaccard distance

- Ideal when data are sparse as it ignores "joint absence"

Tools>Similarities>Jaccard

Measuring the similarity of two actors by counting the shared connections as the percentage of the possible total

Percent of Positive Matches (Jaccard coefficients)

	1 COUN	2 COMM	3 EDUC	4 INDU	5 MAYR	6 WRO	7 NEWS	8 UWAY	9 WELF	10 WEST
1	1.00									
2	0.54	1.00								
3	0.46	0.31	1.00							
4	0.60	0.54	0.42	1.00						
5	0.50	0.93	0.38	0.50	1.00					
6	0.18	0.27	0.11	0.18	0.25	1.00				
7	0.58	0.64	0.54	0.55	0.60	0.08	1.00			
8	0.67	0.46	0.50	0.67	0.43	0.20	0.38	1.00		
9	0.67	0.36	0.50	0.55	0.33	0.11	0.64	0.56	1.00	
10	0.40	0.43	0.44	0.60	0.36	0.38	0.31	0.50	0.36	1.00

Hamming similarity

- Tools>Similarities>Hamming sim
  - treat joint absence as similarity

Linked-in network?

Faculty thesis committee network?

Euclidean distance

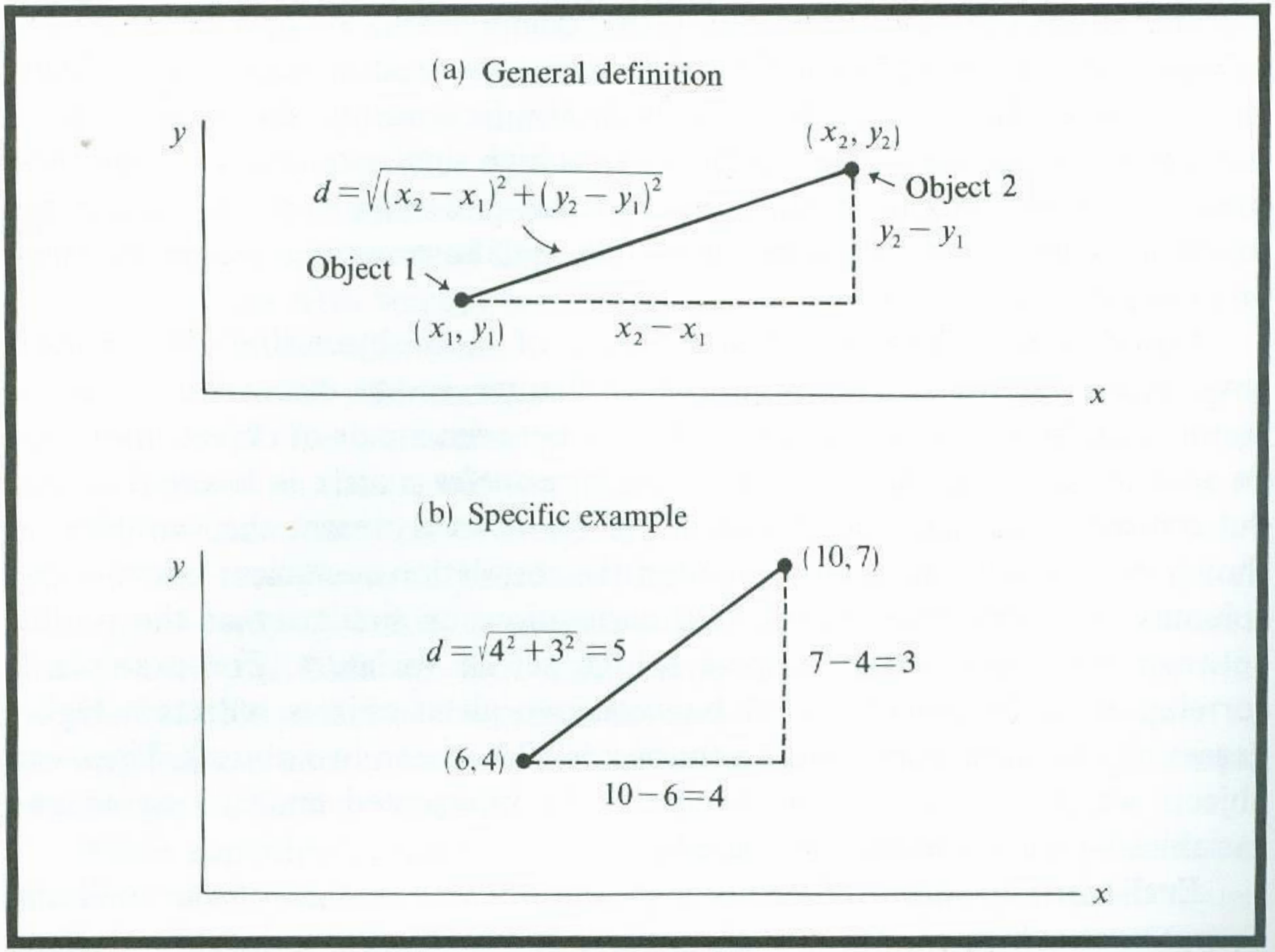


Figure 2 The Euclidean distance between two objects measured on two variables.

Euclidean distance

$P = (p_1, p_2, \ldots, p_n)$

$Q = (q_1, q_2, \ldots, q_n)$

The distance between P and Q

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}.$$

Distance measures

- Tools>Dissimilarities and Distances>Std Vector dissimilarities/distances

Euclidean distances in sending for Knoke information network

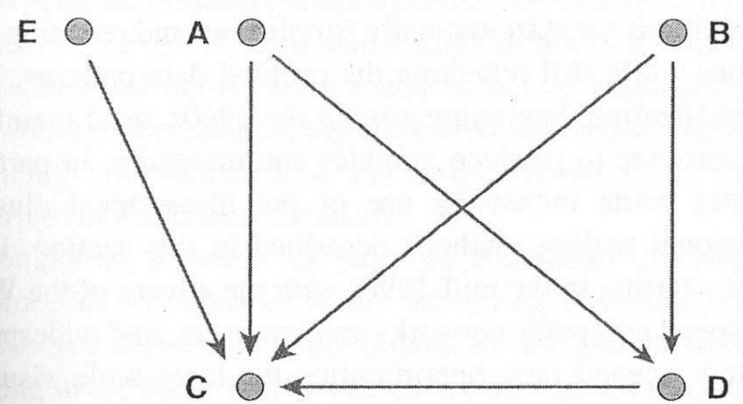
	1	2	3	4	5	6	7	8	9	10
1	0	2	2	1	2	2	1	1	0	1
2	2	0	2	2	1	2	2	1	2	2
3	2	2	0	2	2	2	1	2	2	2
4	1	2	2	0	2	2	1	1	1	1
5	2	1	2	2	0	2	2	1	2	2
6	2	2	2	2	2	0	2	2	2	2
7	1	2	1	1	2	2	0	1	1	2
8	1	1	2	1	1	2	1	0	1	2
9	0	2	2	1	2	2	1	1	0	1
10	1	2	2	1	2	2	2	2	1	0

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

Which two actors are "perfect" structural equivalent?
What's the distance between them?
Their Jaccard coefficient?
Their distance?

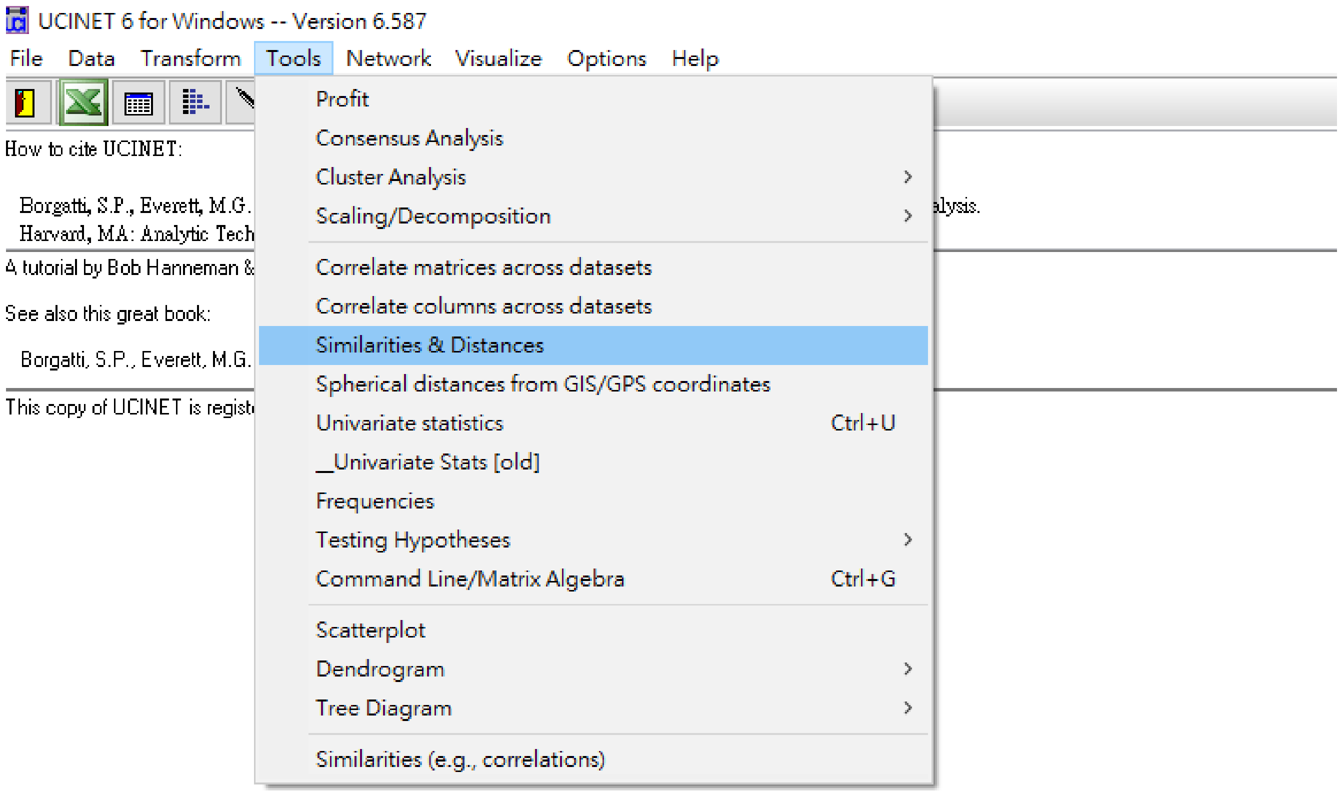
$$d_{AB} = \sqrt{\begin{array}{l}(X_{AC} - X_{BC})^2 + (X_{CA} - X_{CB})^2 + (X_{AD} - X_{BD})^2 + \\ (X_{DA} - X_{DB})^2 + (X_{AE} - X_{BE})^2 + (X_{EA} - X_{EB})^2\end{array}}$$

$$d_{12} = \sqrt{\begin{array}{c}(1-1)^2 + (0-0)^2 + (1-1)^2 \\ +(0-0)^2 + (0-0)^2 + (0-0)^2\end{array}}$$

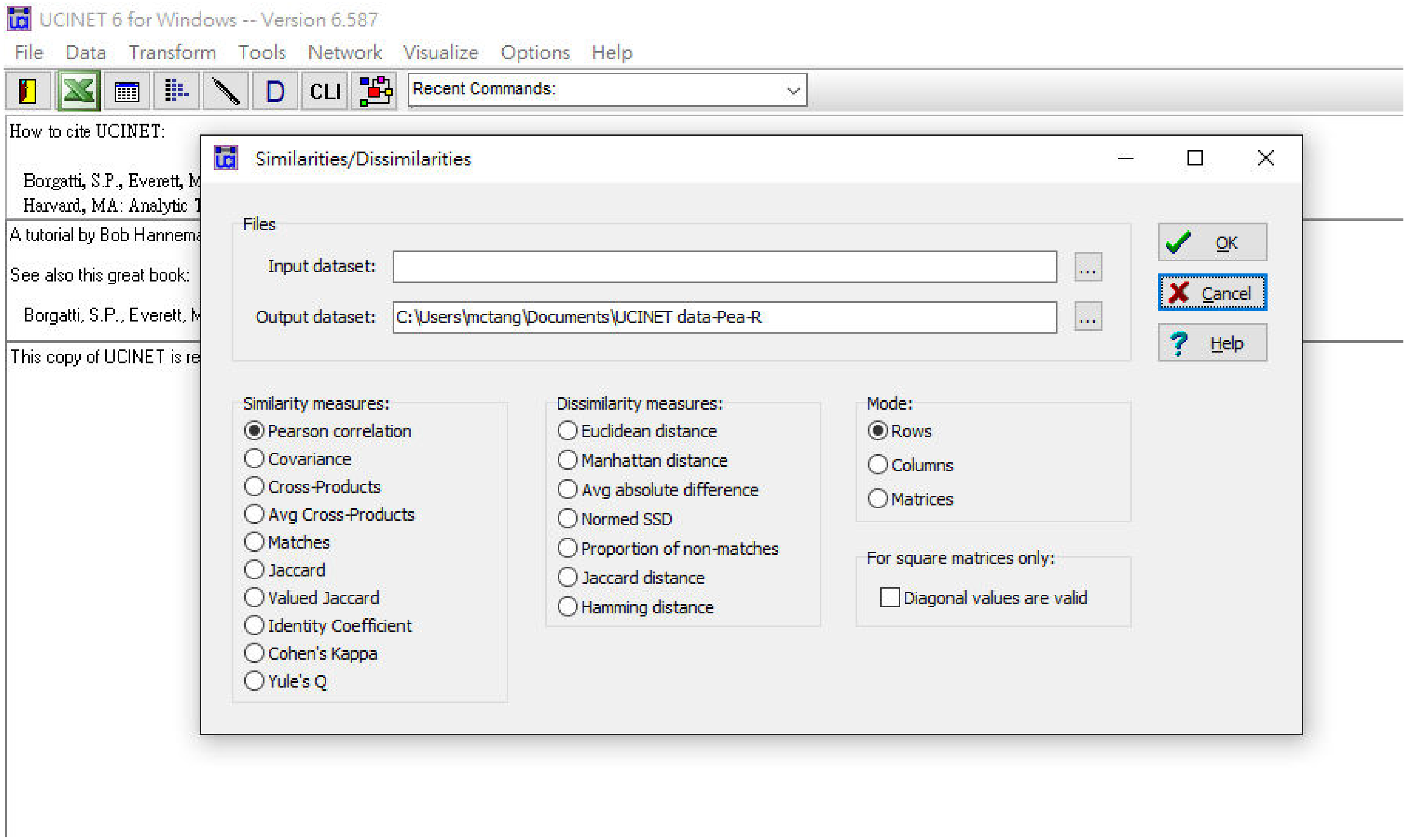


gure 4.12 Digraph of a Five-Node Network for Structural Equivalence

Dis/Similarity with UCINET



Dis/Similarity with UCINET cont.



- So we have transformed the raw data into similarity data, now what?

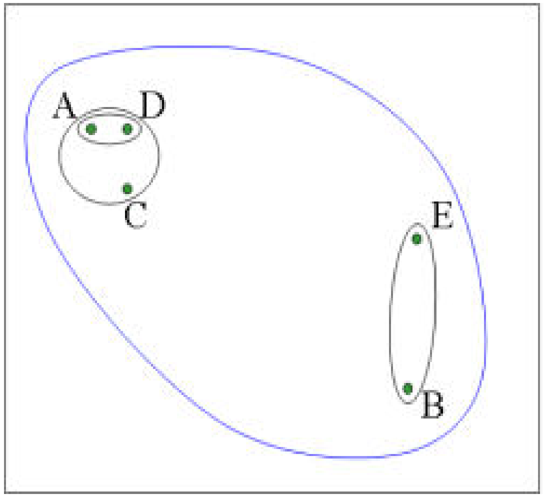
Visual displays

- Hierarchical agglomerative cluster analysis/One dimension
  - Tools>Cluster Analysis>Hierarchical
    - NTU faculty (compare “correlation” and “Jaccard”)
    - CITES
- Multi-dimensional scaling
  - Tools>MDS>Non-Metric MDS
    - CITES

HAC

- Partitions actors into subgroups whose actor is treated initially as singleton cluster, and then clusters are successively joined until all actors merge into a single cluster
- Actors within one cluster have smaller social distances (higher similarity) from one another than from the actors occupying other clusters
- A “dendrogram” (tree-like diagram) visually depicts this hierarchical sequence of merging clusters.

Hierarchical clustering



Initially placing each case in its own cluster,
The two most similar cases are then combined into a class.
The similarity of this new class to all others is then computed ,
The process is repeated until all cases are agglomerated into a single cluster

It's hierarchical because one an item has been joined into a class, it is never Re-classified. This results in clusters of increasing size that always enclose smaller clusters.

HAC cont.

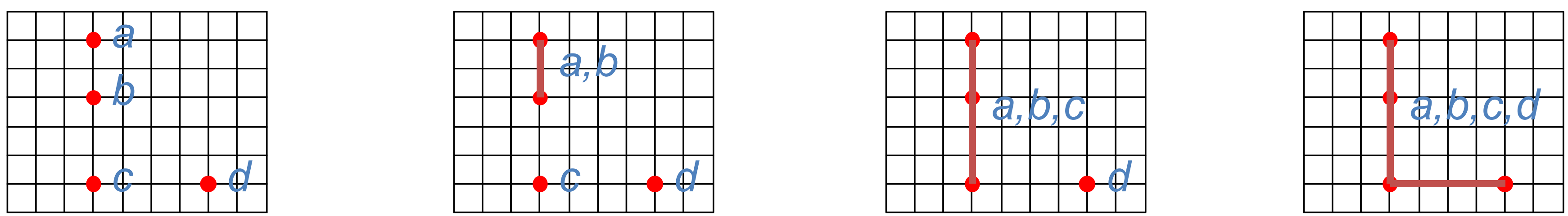
- "Hierarchical" in the sense that once an item has been joined into a cluster, it is never re-classified, which results in clusters of increasing size that always enclose smaller clusters
- Produced, nested, non-overlapping clusters
- The research must decide which level of agglomeration
  - How?
- provide the best representation
- Balance between cohesion and divisiveness

Clustering tools

- Tools>Cluster>Hierarchical
  - Single Link Clustering (Nearest)
  - Complete Linkage (Farthest)
  - Group average
  - Ward's method

Single-Link Method

Euclidean Distance



(1) (2) (3)

	b	*c*	*d*
a	2	5	6
b		3	5
c			4

	b	*c*	*d*
a	2	5	6
b		3	5
c			4

	c	*d*
a,b	3	5
c		4

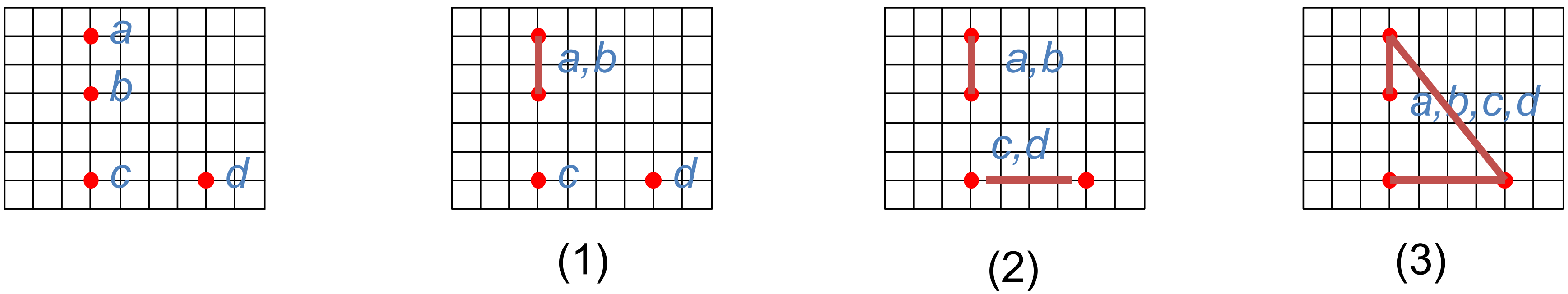
	d
a,b,c	4

Distance Matrix

This rule will, in a sense, *string* objects together to form clusters, and the resulting clusters tend to represent long "chains."

Complete-Link Method

Euclidean Distance



	b	*c*	*d*
a	2	5	6
b		3	5
c			4

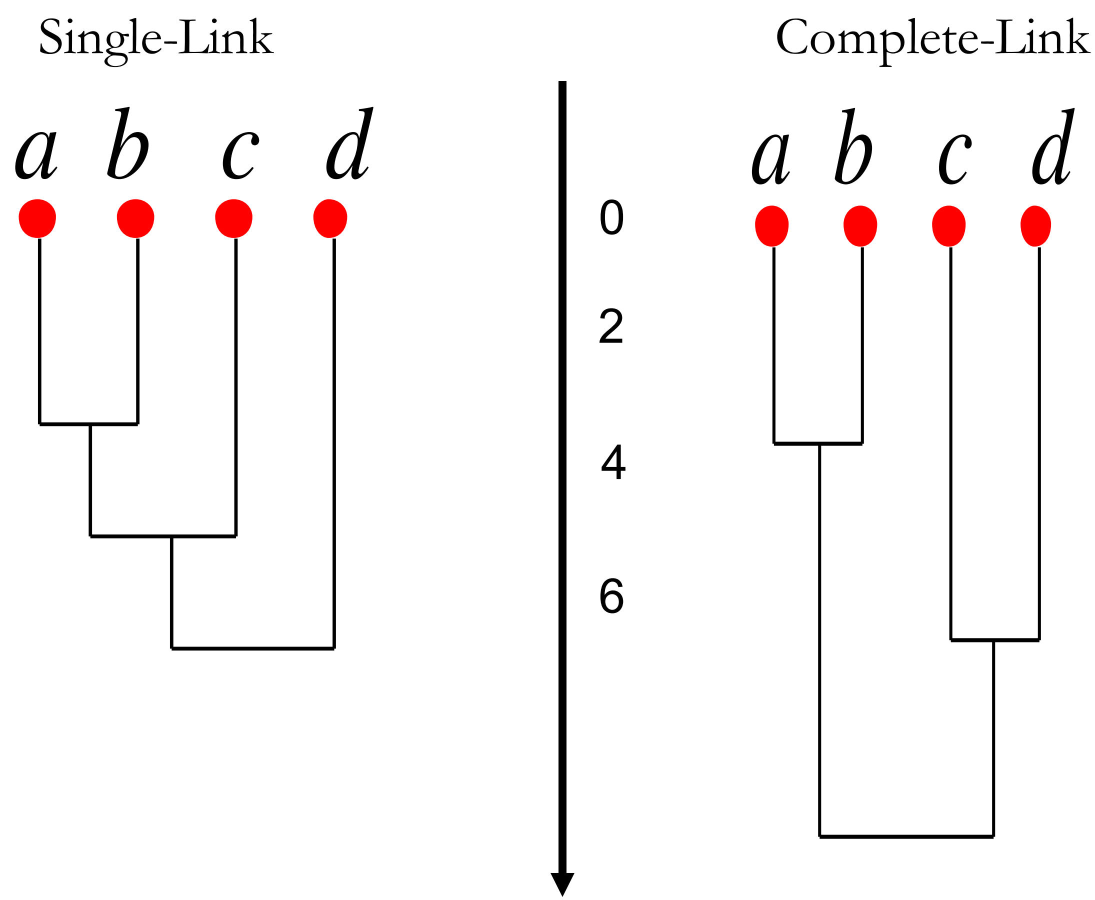
	b	*c*	*d*
a	2	5	6
b		3	5
c			4

	c	*d*
a,b	5	6
c		4

	c,d
a,b	6

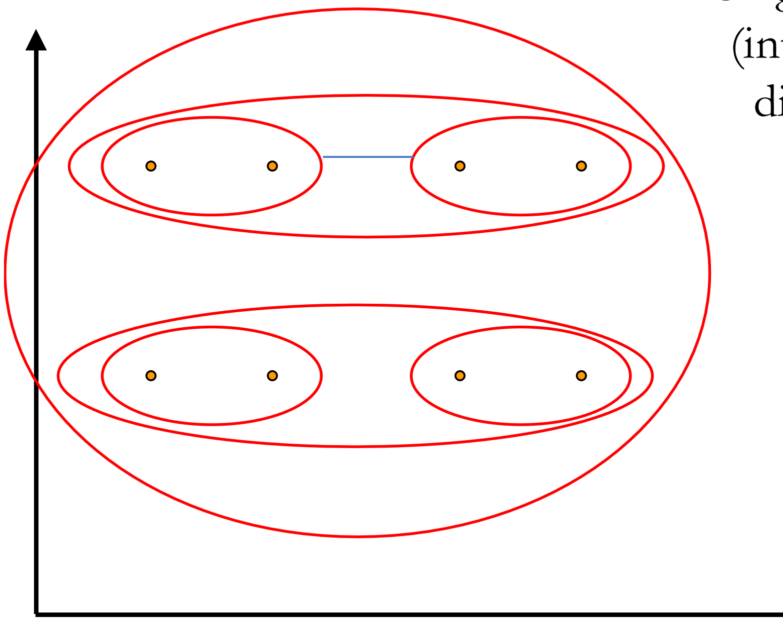
This method usually performs quite well in cases when the objects actually form naturally distinct "clumps." If the clusters tend to be somehow elongated or of a "chain" type nature, then this method is inappropriate.

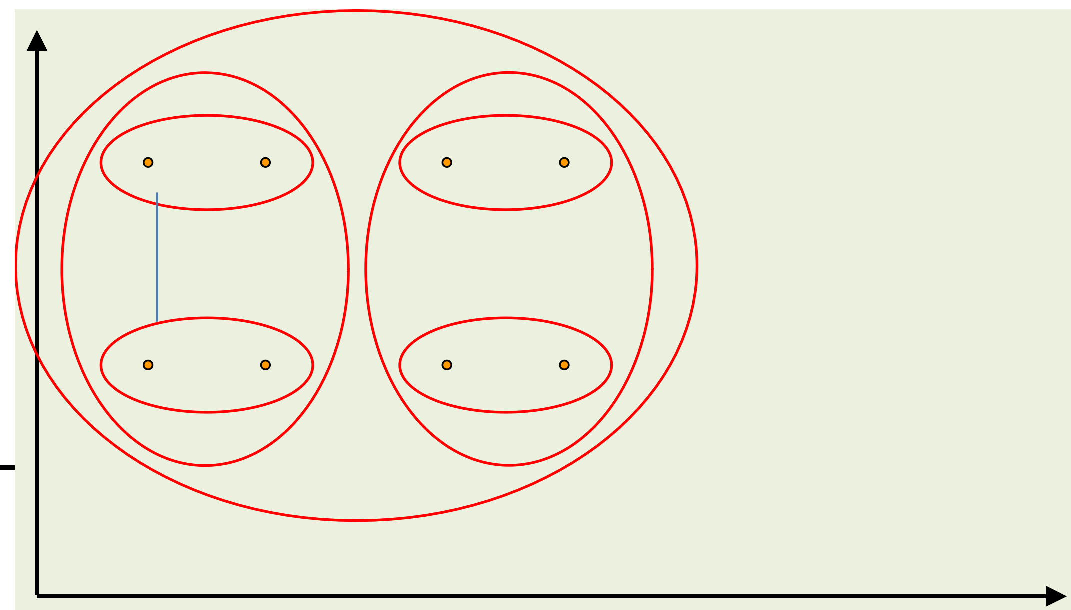
Compare Dendrograms



Which is single/complete link?

"Single-Link"
(inter-cluster distance=
distance between closest pair of points)





"Complete-Link"
(inter-cluster distance=
distance between farthest pair of points)

Average-Link Method

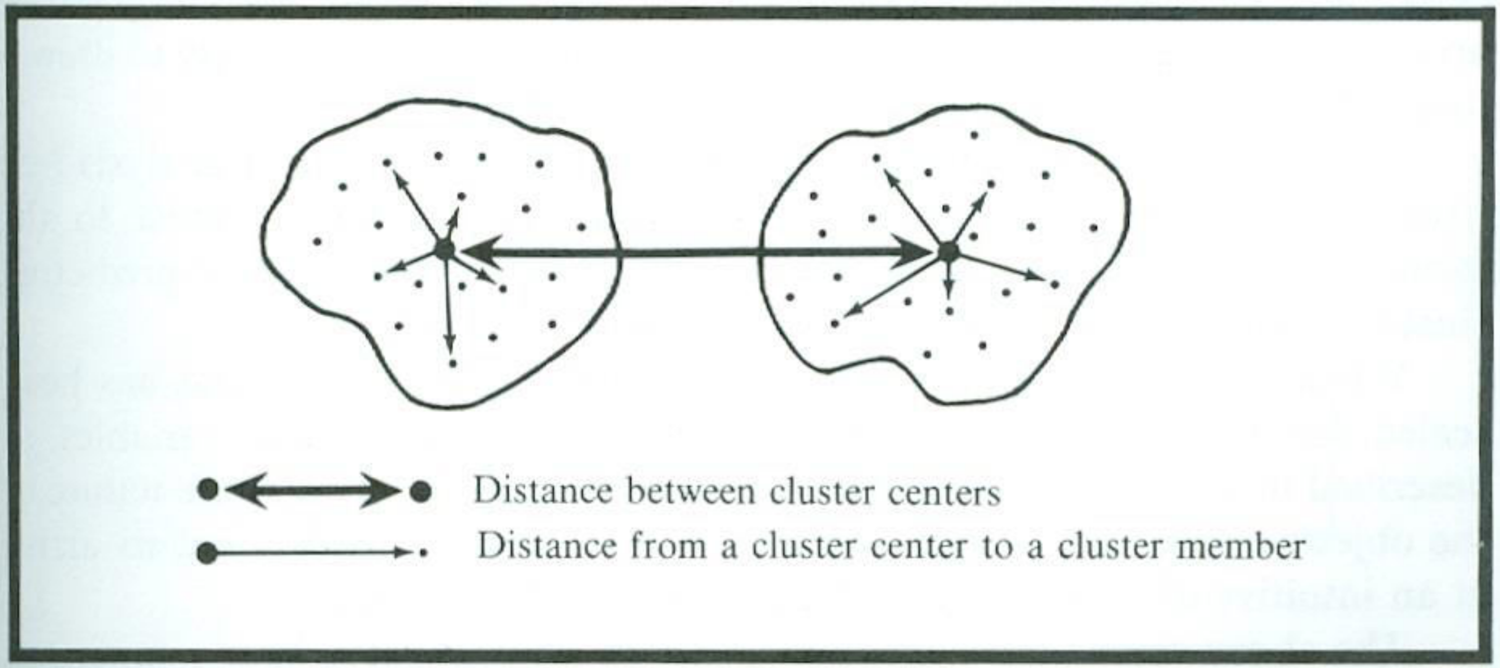


Figure 3 A schematic diagram of the variation between and within clusters. (In actuality, the clusters exist in multidimensional space.)

This method is also very efficient when the objects form natural distinct "clumps," however, it performs equally well with elongated, "chain" type clusters.

Ward's method

- Ward's minimum variance criterion minimizes the total within-cluster variance. At each step the pair of clusters with minimum between-cluster distance are merged.
  - i.e., minimum squared distance of all points from the new (?) cluster center.
- In general, this method is regarded as very efficient, however, it tends to create clusters of small size.

```
*************************************************************
Matrix KNOKI in dataset Dissim-KNOKBUR
*************************************************************
```

Hamming distance

```
HIERARCHICAL CLUSTERING

                                   1
 Level    6 3 2 5 8 4 7 1 9 0
 ------   - - - - - - - - - -
0.0000    . . . . . . . XXX .
1.0000    . . XXX . XXX XXX .
1.2222    . . XXX . XXXXXXX .
1.6667    . . XXXXX XXXXXXX .
2.1111    . . XXXXX XXXXXXXXX
2.8815    . . XXXXXXXXXXXXXXX
3.3091    . XXXXXXXXXXXXXXXXX
4.2158    XXXXXXXXXXXXXXXXXXX
```

Measures of cluster adequacy

		1	2	3	4	5	6	7	8
1	Eta	-0.308	-0.422	-0.578	-0.634	-0.687	-0.547	-0.487	
2	Q	-0.118	-0.143	-0.159	-0.173	-0.171	-0.074	-0.044	
3	Q-prime	-0.132	-0.167	-0.190	-0.216	-0.228	-0.111	-0.087	
4	E-I	1.000	0.970	0.894	0.848	0.727	0.015	-0.409	

Size of each cluster, expressed as a proportion of the total population clustered

		1	2	3	4	5	6	7	8
1	CL1	0.200	0.200	0.400	0.400	0.500	0.800	0.900	1.000
2	CL2	0.100	0.200	0.200	0.300	0.300	0.100	0.100	
3	CL3	0.100	0.100	0.100	0.100	0.100	0.100		
4	CL4	0.100	0.200	0.100	0.100	0.100			
5	CL5	0.100	0.100	0.100	0.100				
6	CL6	0.100	0.100	0.100					
7	CL7	0.100	0.100						
8	CL8	0.100							
9	CL9	0.100							
10	CL10								

E-I index measures the ratio of the numbers of tie Within the clusters to ties between clusters; not as useful in valued networks

Q index low; not meaningful grouping

Multi-dimensional scaling

- Represents visually the patterns of similarity or distance in the tie profiles among the actors as a "map" in multi-dimensional space (mostly two or three)
- Actors with smaller distance or greater similarities between them are located closer in space

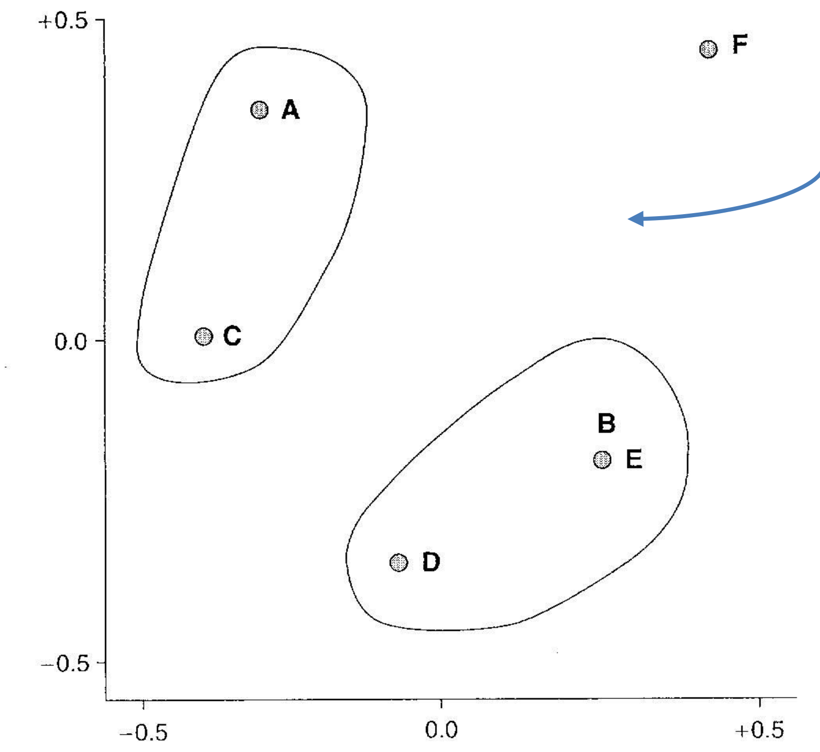
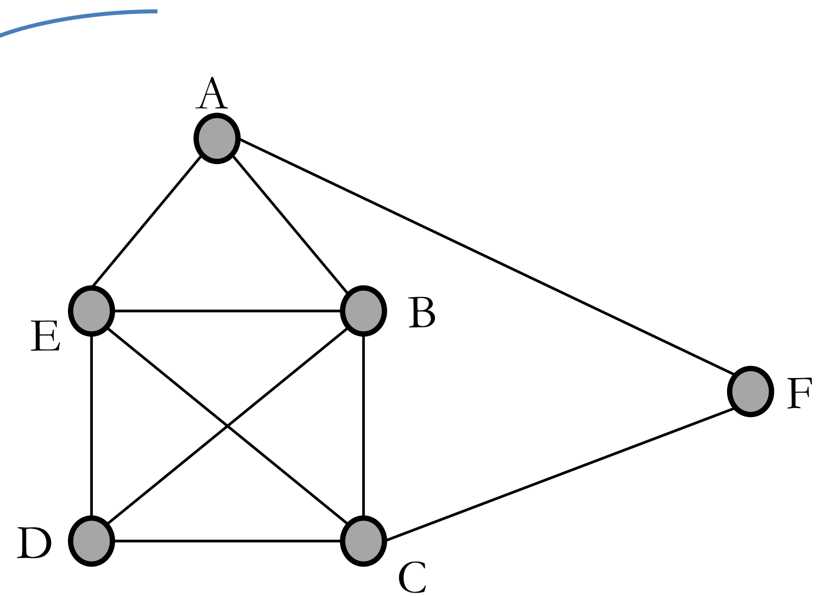


Figure 4.14 Multidimensional Scaling of Network in Figure 4.12



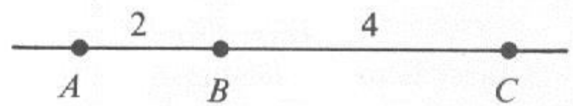
Dimensionalities

- Using cluster analysis, we are implicitly assuming that the similarity or distance among cases reflects as single underlying dimension
- Alternatives
  - MDS
  - Factor analysis

(a)

Object Pair	Inter-Object Distance
A-B	2
A-C	6
B-C	4

A perfect representation in one dimension

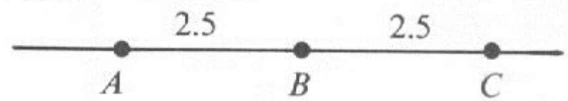


cleark libirian attorney

(b)

Object Pair	Inter-Object Distance
A-B	2
A-C	5
B-C	2

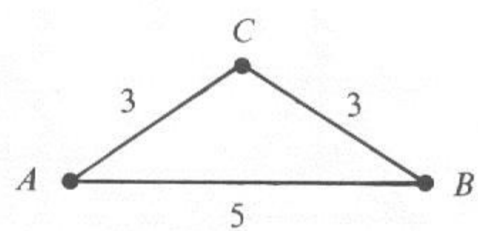
An approximate representation in one dimension



(c)

Object Pair	Inter-Object Distance
A-B	5
A-C	3
B-C	3

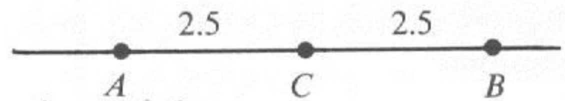
A perfect representation in two dimensions



(d)

Object Pair	Inter-Object Distance
A-B	5
A-C	3
B-C	3

An approximate representation in one dimension

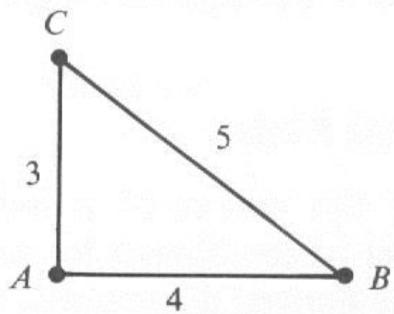


compressed sacrifying some accu

(e)

Object Pair	Inter-Object Distance
A-B	4
A-C	3
B-C	5

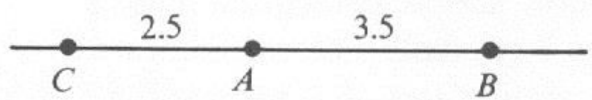
A perfect representation in two dimensions



(f)

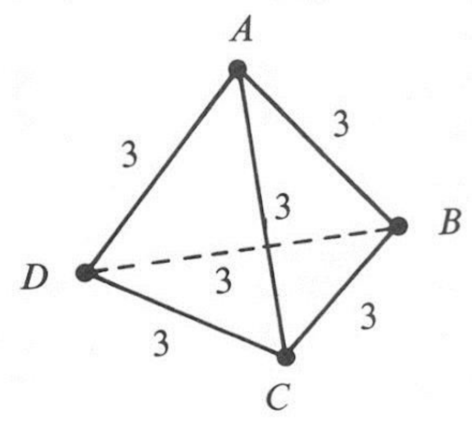
Object Pair	Inter-Object Distance
A-B	4
A-C	3
B-C	5

An approximate representation in one dimension

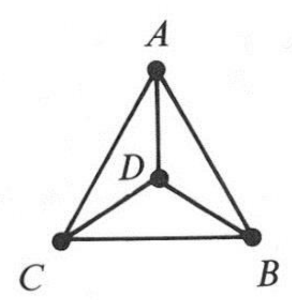


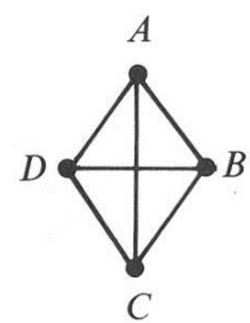
(a) Four objects represented perfectly in three dimensions.

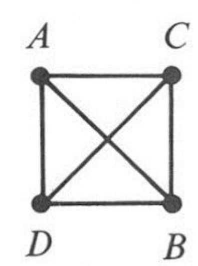
Object Pair	Inter-Object Distance
A-B	3
A-C	3
A-D	3
B-C	3
B-D	3
C-D	3



(b) Some alternative *approximate* representations in two dimensions.







Stress

- Measures loss of information in the process of arriving at a particular dimensional representation (i.e. distortion)
- A stress value of .15 or lower is usually considered satisfactory. Value over 0.2 generally considered poor fit

$$\text{Stress} = \sqrt{\frac{\sum\sum\left(f\left(X_{ij}\right) - d_{ij}\right)^2}{\text{Scale}}}$$

To keep it from 0 to 1

- Use city.##h as an example
- Run scaling on NetDraw
- Tools>with UCINET

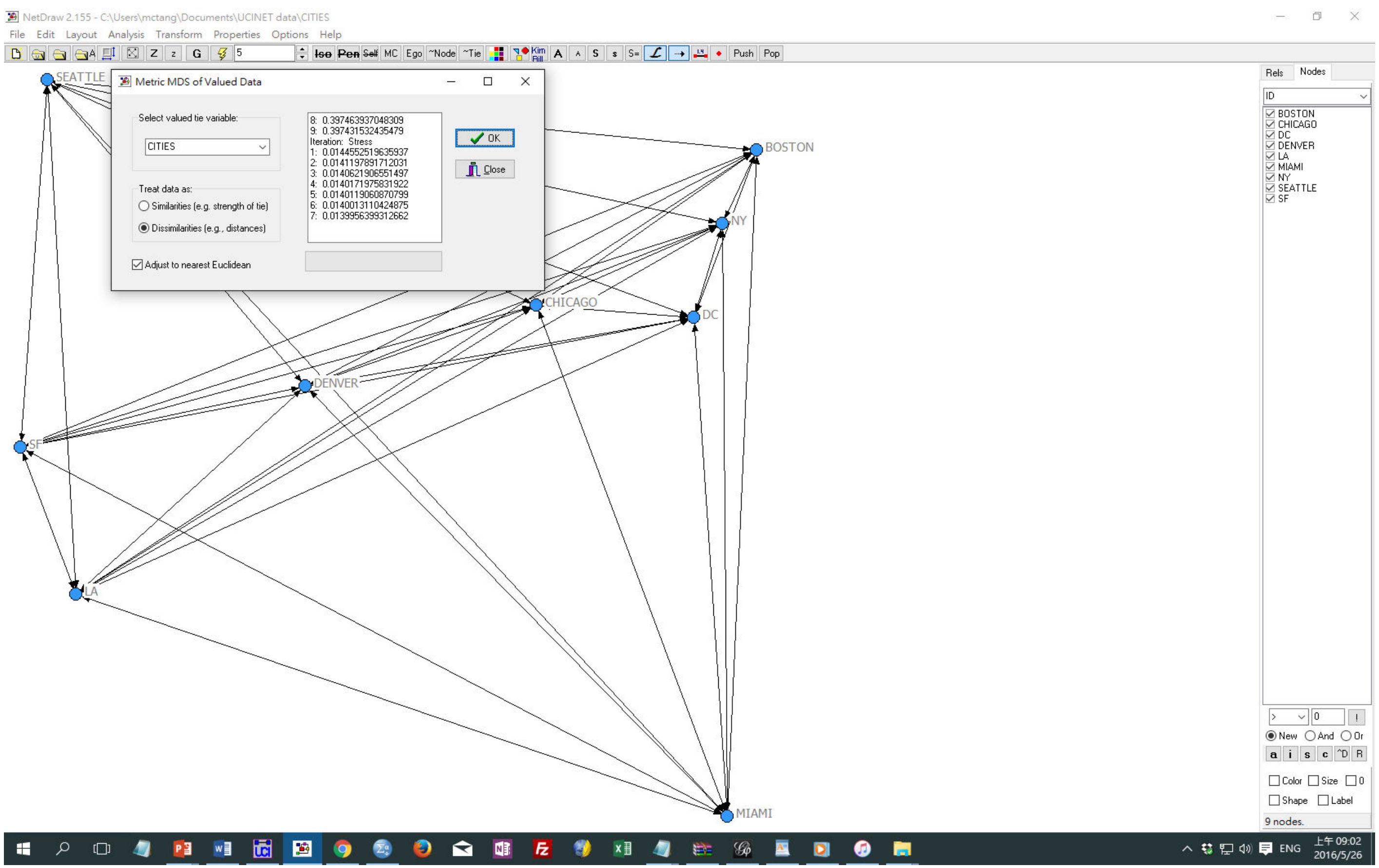
    clustering

    scaling/decomposition

    dissimilarity (distance)

Airline distances between US cities

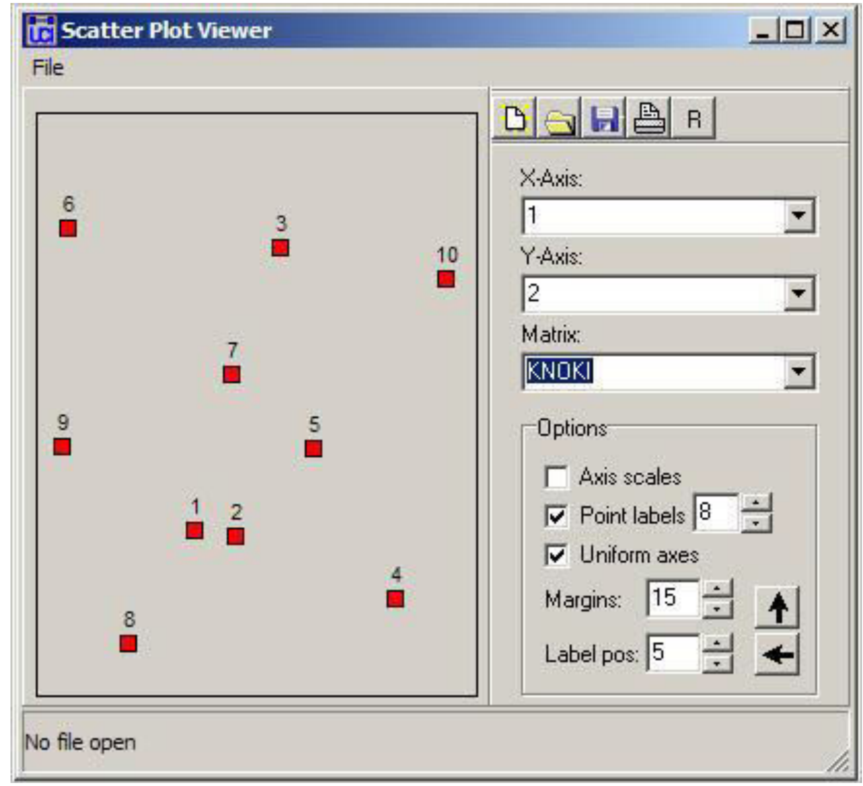
	ATLANTA	CHICAGO	DENVER	HOUSTON	LA	MIAMI	NY	SF	SEATTLE	DC
ATLANTA		587	1212	701	1936	604	748	2139	2182	543
CHICAGO	587		920	940	1745	1188	713	1858	1737	597
DENVER	1212	920		879	831	1726	1631	949	1021	1494
HOUSTON	701	940	879		1374	968	1420	1645	1891	1220
LA	1936	1745	831	1374		2339	2451	347	959	2300
MIAMI	604	1188	1726	968	2339		1092	2594	2734	923
NY	748	713	1631	1420	2451	1092		2571	2408	205
SF	2139	1858	949	1645	347	2594	2571		678	2442
SEATTLE	2182	1737	1021	1891	959	2734	2408	678		2329
DC	543	597	1494	1220	2300	923	205	2442	2329	



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.



Interpret the “meaning” of the dimensions

1. With NTNU & NTU thesis committee

Use tool to run similarity (let's try cosine and valued Jaccard)

dissimilarity (Euclid distance)

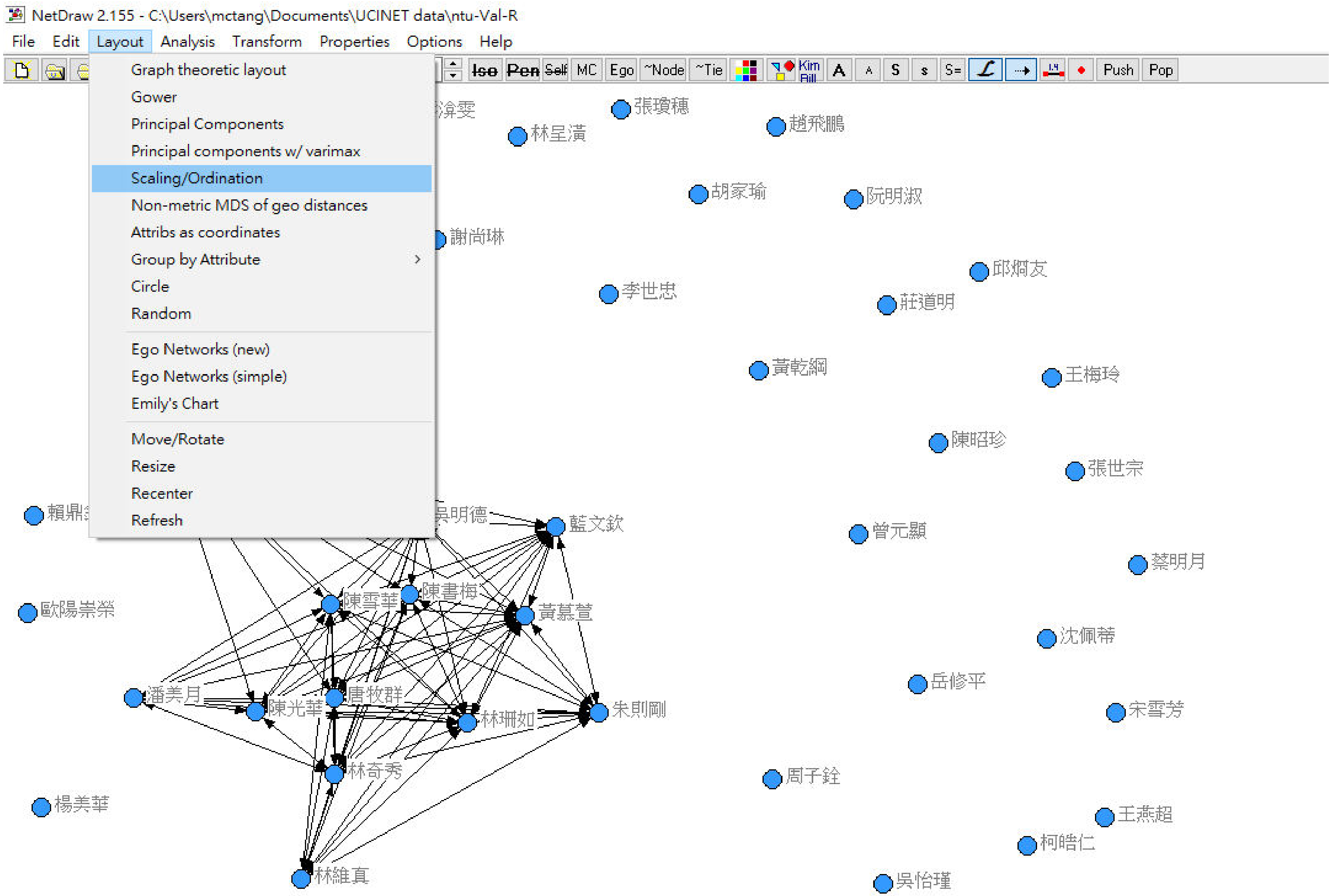
Present the result using NetDraw>layout>scaling

1. Again Run profile analysis with NTU/NTNU thesis committee data

2. Open the resulting SE with NetDraw
3. Do metric MDS

Exercise

- Use friendship data , which is a directional network
  - Run data transform> transpose
    - This is to create a sending AND receiving profile
  - Run tool similarity/dissimilarity analysis
    - Let's try Jaccard coefficient for similarity (you might want to compare different similarity methods)
    - And Euclidean distance for dissimilarity
    - Go to network draw, use LAYOUT > SCALING, use both similarity and dissimilarity with the two different file, respectively
  - Again, under tool, perform clustering analysis
    - Observe the Q value to determine the proper level of clustering
  - Run modularity and Newman community detection on the same similarity matrix and compare all three grouping methods, which make more sense to you?
  - Save it as excel file and open it with Gephi, perform modularity analysis

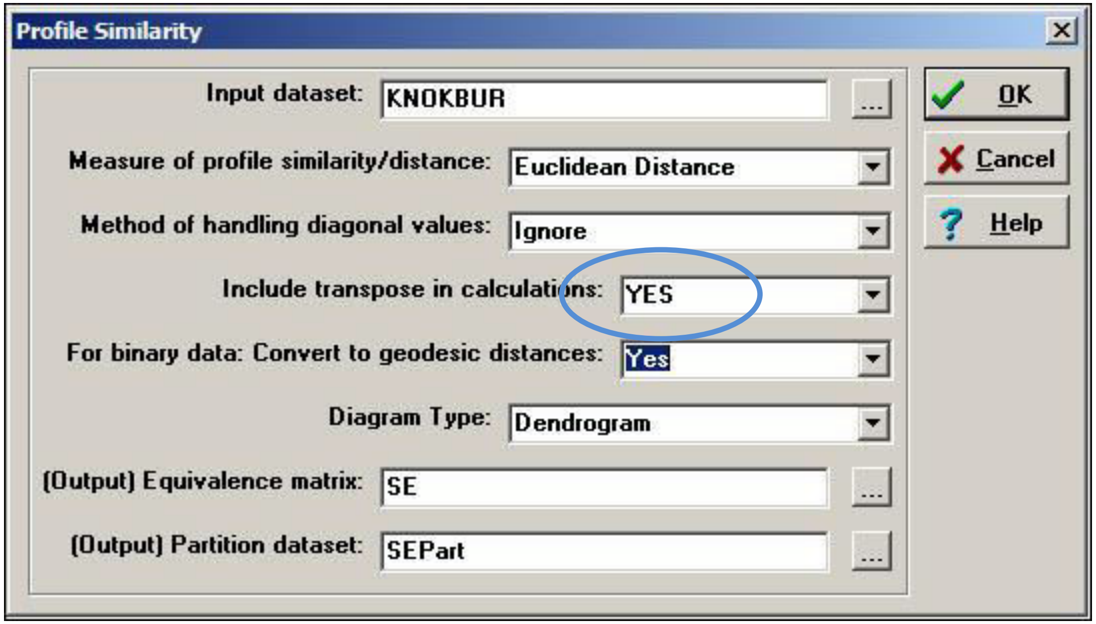


MDS with NTU valued Jaccard similarity

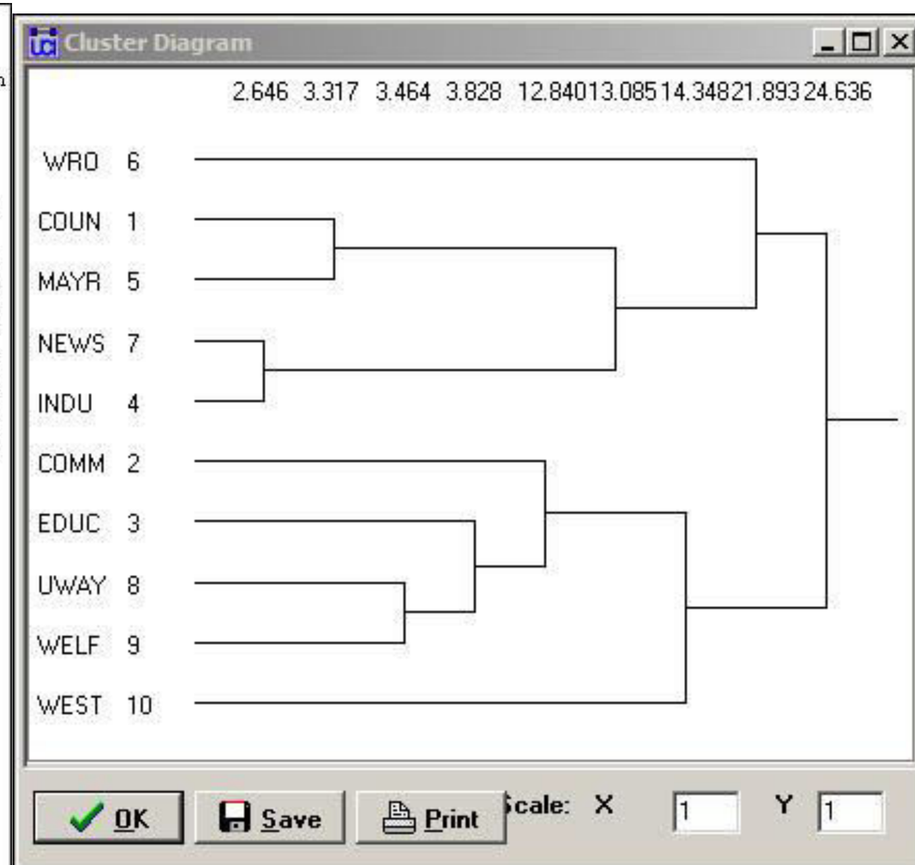
Clustering similar profiles

Use cluster analysis to identify structurally equivalent classes
In the network

Network>Roles & Positions > Structural>Profile



```
Measure:                        Euclidean Distance
Diagonal:                       Treat as Missing Values
Use geodesics?                  YES
Input dataset:                  C:\Documents and Settings\hanneman\My Documen


Structural Equivalence Matrix

             1      2      3      4      5      6      7      8      9     10
          COUN   COMM   EDUC   INDU   MAYR    WRO   NEWS   UWAY   WELF   WEST
        ------ ------ ------ ------ ------ ------ ------ ------ ------ ------
  1      0.000 18.466 24.739 13.000  3.317 22.068 13.191 24.860 23.917 29.086
  2     18.466  0.000 12.689 17.889 13.038 24.125 12.450 13.454 12.610 20.372
  3     24.739 12.689  0.000 25.000 21.517 27.514 21.863  4.000  3.742 12.570
  4     13.000 17.889 25.000  0.000 13.038 22.136  2.646 25.239 24.228 29.614
  5      3.317 13.038 21.517 13.038  0.000 22.405 13.077 21.932 20.761 26.814
  6     22.068 24.125 27.514 22.136 22.405  0.000 21.471 28.125 27.295 23.345
  7     13.191 12.450 21.863  2.646 13.077 21.471  0.000 21.954 21.024 26.533
  8     24.860 13.454  4.000 25.239 21.932 28.125 21.954  0.000  3.464 12.570
  9     23.917 12.610  3.742 24.228 20.761 27.295 21.024  3.464  0.000 13.115
 10     29.086 20.372 12.570 29.614 26.814 23.345 26.533 12.570 13.115  0.000

HIERARCHICAL CLUSTERING OF EQUIVALENCE MATRIX

           C M N I C E U W W
         W O A E N O D W E E
         R U Y W D M U A L S
         O N R S U M C Y F T

                           1
 Level   6 1 5 7 4 2 3 8 9 0
------   - - - - - - - - - -
 2.646   . . . XXX . . . . .
 3.317   . XXX XXX . . . . .
 3.464   . XXX XXX . . XXX .
 3.828   . XXX XXX . XXXXX .
12.840   . XXX XXX XXXXXXX .
13.085   . XXXXXXX XXXXXXX .
14.348   . XXXXXXX XXXXXXXXX
21.893   XXXXXXXXX XXXXXXXXX
24.636   XXXXXXXXXXXXXXXXXXX
```



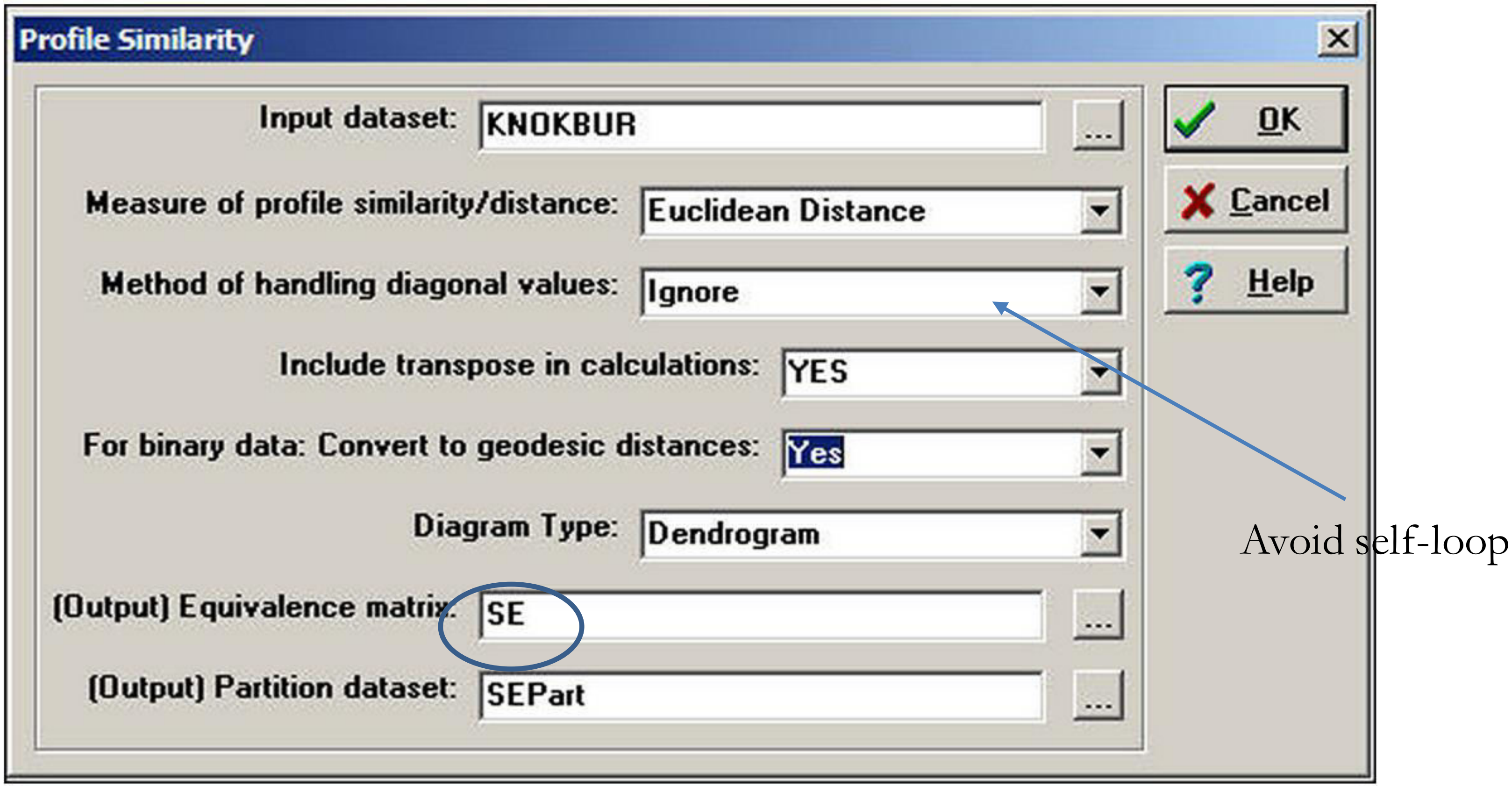


Figure 13.12. Dialog of *Network>Roles & Positions>Structural>Profile*

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.

Figure 13.10. Non-metric MDS two-dimensional coordinates of Knoke information adjacency

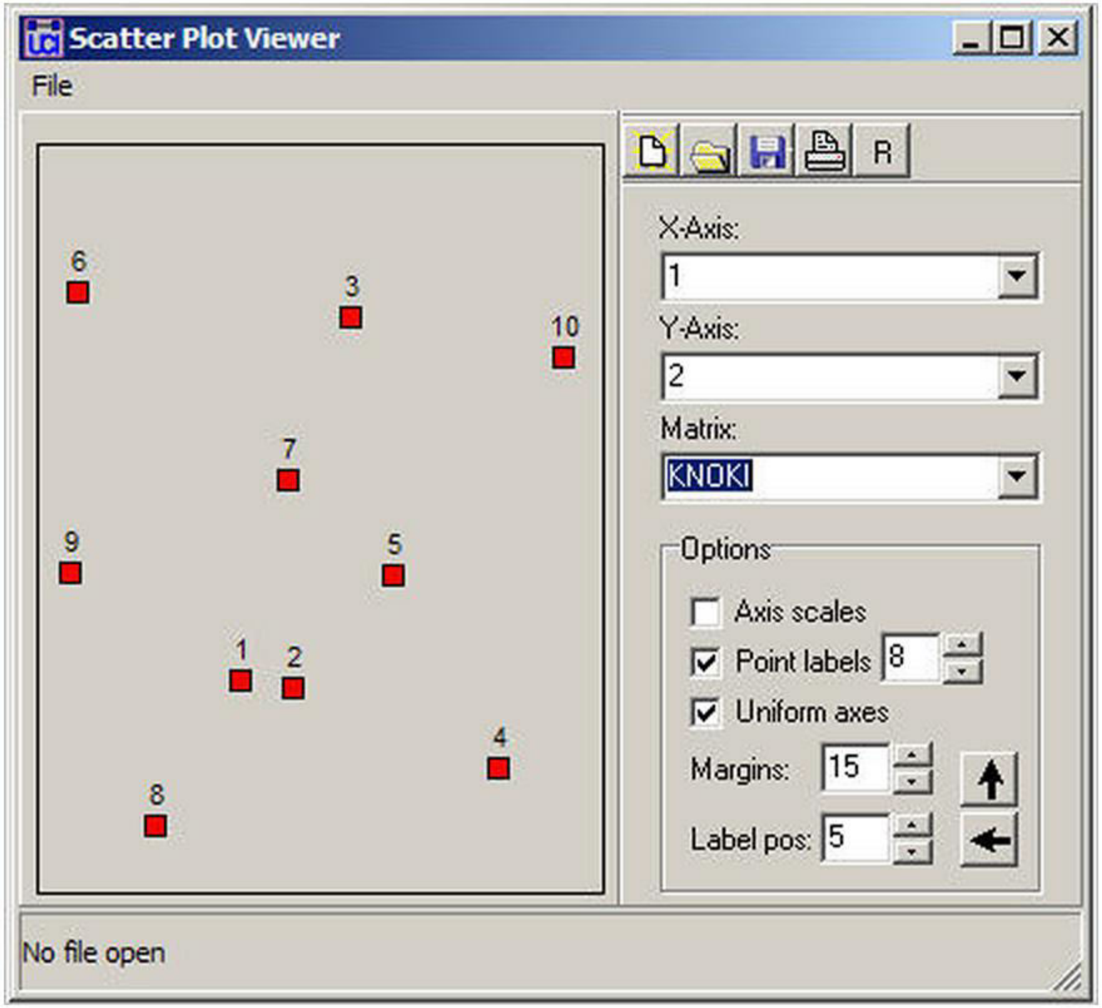


Figure 13.11. Two-dimensional map of non-metric MDS of Knoke information adjacency

```
Measure:                     Euclidean Distance
Diagonal:                    Treat as Missing Values
Use geodesics?               YES
Input dataset:               C:\Documents and Settings\hanneman\My Documen


Structural Equivalence Matrix

            1      2      3      4      5      6      7      8      9     10
         COUN   COMM   EDUC   INDU   MAYR    WRO   NEWS   UWAY   WELF   WEST
       ------ ------ ------ ------ ------ ------ ------ ------ ------ ------
  1     0.000 18.466 24.739 13.000  3.317 22.068 13.191 24.860 23.917 29.086
  2    18.466  0.000 12.689 17.889 13.038 24.125 12.450 13.454 12.610 20.372
  3    24.739 12.689  0.000 25.000 21.517 27.514 21.863  4.000  3.742 12.570
  4    13.000 17.889 25.000  0.000 13.038 22.136  2.646 25.239 24.228 29.614
  5     3.317 13.038 21.517 13.038  0.000 22.405 13.077 21.932 20.761 26.814
  6    22.068 24.125 27.514 22.136 22.405  0.000 21.471 28.125 27.295 23.345
  7    13.191 12.450 21.863  2.646 13.077 21.471  0.000 21.954 21.024 26.533
  8    24.860 13.454  4.000 25.239 21.932 28.125 21.954  0.000  3.464 12.570
  9    23.917 12.610  3.742 24.228 20.761 27.295 21.024  3.464  0.000 13.115
 10    29.086 20.372 12.570 29.614 26.814 23.345 26.533 12.570 13.115  0.000

HIERARCHICAL CLUSTERING OF EQUIVALENCE MATRIX

           C M N I C E U W W
         W O A E N O D W E E
         R U Y W D M U A L S
         O N R S U M C Y F T

                           1
 Level   6 1 5 7 4 2 3 8 9 0
------   - - - - - - - - - -
 2.646   . . . XXX . . . . .
 3.317   . XXX XXX . . . . .
 3.464   . XXX XXX . . XXX .
 3.828   . XXX XXX . XXXXX .
12.840   . XXX XXX XXXXXXX .
13.085   . XXXXXXX XXXXXXX .
14.348   . XXXXXXX XXXXXXXXX
21.893   XXXXXXXXX XXXXXXXXX
24.636   XXXXXXXXXXXXXXXXXXX
```

Figure 13.13. Profile similarity of geodesic distances of rows and columns of Knoke information network

Social network analysis

Statistical tools in UCINET

Types of relationship

- Node-level (monadic)
  - more central people tend to be happier
- Dyadic level (pairs)
  - Correlation between physical distance and frequency of communication
- One dyadic, the other monadic (node attribute; transform nodal into dyadic attributes)
  - Categorical
    - Same gender correlation with interaction frequency
  - Continuous
    - Age difference correlate to “friendship” relationship
- Network level
  - Team cohesion/diversity with performance

Table 1.1 Examples of research questions by level of analysis and type of node.

	Type of node	
Level of analysis	**Individuals**	**Collectivities**
Dyad level $O(n^2)$	Are employees whose offices are near each other more likely to develop friendships than employees whose offices are further apart?	Are firms with similar organizational cultures more likely to form joint ventures with each other?
Node level $O(n^1)$	Are employees who are more central in their organization's friendship network less likely to leave for another company?	Are firms with more diverse technology partners more likely to introduce innovative products into the market?
Network level $O(n^0)$	When a network of employees is characterized by many redundant paths between all pairs of persons, is the network less disrupted by individuals leaving the firm?	When a network of firms is densely connected, does this place the network at greater risk of catastrophic failure (because of cascade effects)?

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). *Analyzing social networks*. SAGE Publications Limited.

Table 1.4 Types of network studies classified by direction of causality and level of analysis.

	Network variables as independent/ explanatory	Network variables as dependent/ outcomes
Dyad level	Friendship between pairs of farmers to predict which pairs of farmers make the same decision about going organic	Similarity of interests (e.g., sky diving) to predict who becomes friends with each other
Node level	Centrality in organizational trust network to predict who is chosen for promotion	Extraversion to predict who becomes central in friendship network
Network level	Shortness of paths in a group's communication network to predict group's ability to solve problems	Type of organizational culture (emphasizing either cooperation or competition) to predict structure of the trust network

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). *Analyzing social networks*. SAGE Publications Limited.

Inferential statistics

- how much confidence can I have that the pattern I see in the data I've collected is actually typical of some larger population, or that the apparent pattern is not really just a random occurrence?

With relational data

- Violation of
  - 1. Independence assumption
  - 2. Normal distribution
- Alternative: boot-strapping and simulation approaches

The solution: permutation (randomization) tests

- What are the chances of observing such a large correlation even when the values of the variables are assigned independently (i.e. randomly assigned) of each other?

Node level

Performance	Centrality	Height
A		
B		
C		
D		

Correlation and regression of dyadic relationship

Friendship rating	Distance	Same town?	Same Gender	?
A-B				
B-A				
A-C				

Person	A	B	C	D	E
A	.	0	2	3	1
B	4	.	8	10	6
C	5	5	.	5	5
D	2	8	7	.	3
E	2	4	3	5	.

Correlations in a network

Person	**A**	**B**	**C**	**D**	**E**
A	**.**	**0**	**2**	**3**	**1**
B	**4**	**.**	**8**	**10**	**6**
C	**5**	**5**	**.**	**5**	**5**
D	**2**	**8**	**7**	**.**	**3**
E	**2**	**4**	**3**	**5**	**.**

Pair	Row Number	Column Number	Absolute value of age difference	Friendship Rating
AA	1	1	.	.
AB	1	2	5	0
AC	1	3	25	2
AD	1	4	35	3
AE	1	5	15	1
BA	2	1	5	4

Actor	Gender
A	1
B	2
C	2
D	1
E	1
F	2

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

Figure 5.14 Converting gender into 'same gender as'.

Actor	Age
A	14
B	67
C	34
D	33
E	56
F	45

	A	B	C	D	E	F
A	0	53	20	19	42	31
B	53	0	33	34	11	22
C	20	33	0	1	22	11
D	19	34	1	0	23	12
E	42	11	22	23	0	11
F	31	22	11	12	11	0

Figure 5.15 Converting age into 'difference in age'.

42 40 32 59 55 symmetrical

Age		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
33	1	0	−9	−7	0	1	−26	−22	−1	−29	−4	−13	−1	−15	−10	−7	6	3	0	1	−5	−3
42	2	9	0	2	9	10	−17	−13	8	−20	5	−4	8	−6	−1	2	15	12	9	10	4	6
40	3	7	−2	0	7	8	−19	−15	6	−22	3	−6	6	−8	−3	0	13	10	7	8	2	4
33	4	0	−9	−7	0	1	−26	−22	−1	−29	−4	−13	−1	−15	−10	−7	6	3	0	1	−5	−3
32	5	−1	−10	−8	−1	0	−27	−23	−2	−30	−5	−14	−2	−16	−11	−8	5	2	−1	0	−6	−4
59	6	26	17	19	26	27	0	4	25	−3	22	13	25	11	16	19	32	29	26	27	21	23
55	7	22	13	15	22	23	−4	0	21	−7	18	9	21	7	12	15	28	25	22	23	17	19
34	8	1	−8	−6	1	2	−25	−21	0	−28	−3	−12	0	−14	−9	−6	7	4	1	2	−4	−2
62	9	29	20	22	29	30	3	7	28	0	25	16	28	14	19	22	35	32	29	30	24	26
37	10	4	−5	−3	4	5	−22	−18	3	−25	0	−9	3	−11	−6	−3	10	7	4	5	−1	1
46	11	13	4	6	13	14	−13	−9	12	−16	9	0	12	−2	3	6	19	16	13	14	8	10
34	12	1	−8	−6	1	2	−25	−21	0	−28	−3	−12	0	−14	−9	−6	7	4	1	2	−4	−2
48	13	15	6	8	15	16	−11	−7	14	−14	11	2	14	0	5	8	21	18	15	16	10	12
43	14	10	1	3	10	11	−16	−12	9	−19	6	−3	9	−5	0	3	16	13	10	11	5	7
40	15	7	−2	0	7	8	−19	−15	6	−22	3	−6	6	−8	−3	0	13	10	7	8	2	4
27	16	−6	−15	−13	−6	−5	−32	−28	−7	−35	−10	−19	−7	−21	−16	−13	0	−3	−6	−5	−11	−9
30	17	−3	−12	−10	−3	−2	−29	−25	−4	−32	−7	−16	−4	−18	−13	−10	3	0	−3	−2	−8	−6
33	18	0	−9	−7	0	1	−26	−22	−1	−29	−4	−13	−1	−15	−10	−7	6	3	0	1	−5	−3
32	19	−1	−10	−8	−1	0	−27	−23	−2	−30	−5	−14	−2	−16	−11	−8	5	2	−1	0	−6	−4
38	20	5	−4	−2	5	6	−21	−17	4	−24	1	−8	4	−10	−5	−2	11	8	5	6	0	2
36	21	3	−6	−4	3	4	−23	−19	2	−26	−1	−10	2	−12	−7	−4	9	6	3	4	−2	0

Matrix 8.1 Age of each node (left) and differences in ages between all pairs of nodes (right).

Actor	Status
A	6
B	10
C	3
D	5
E	9
F	4

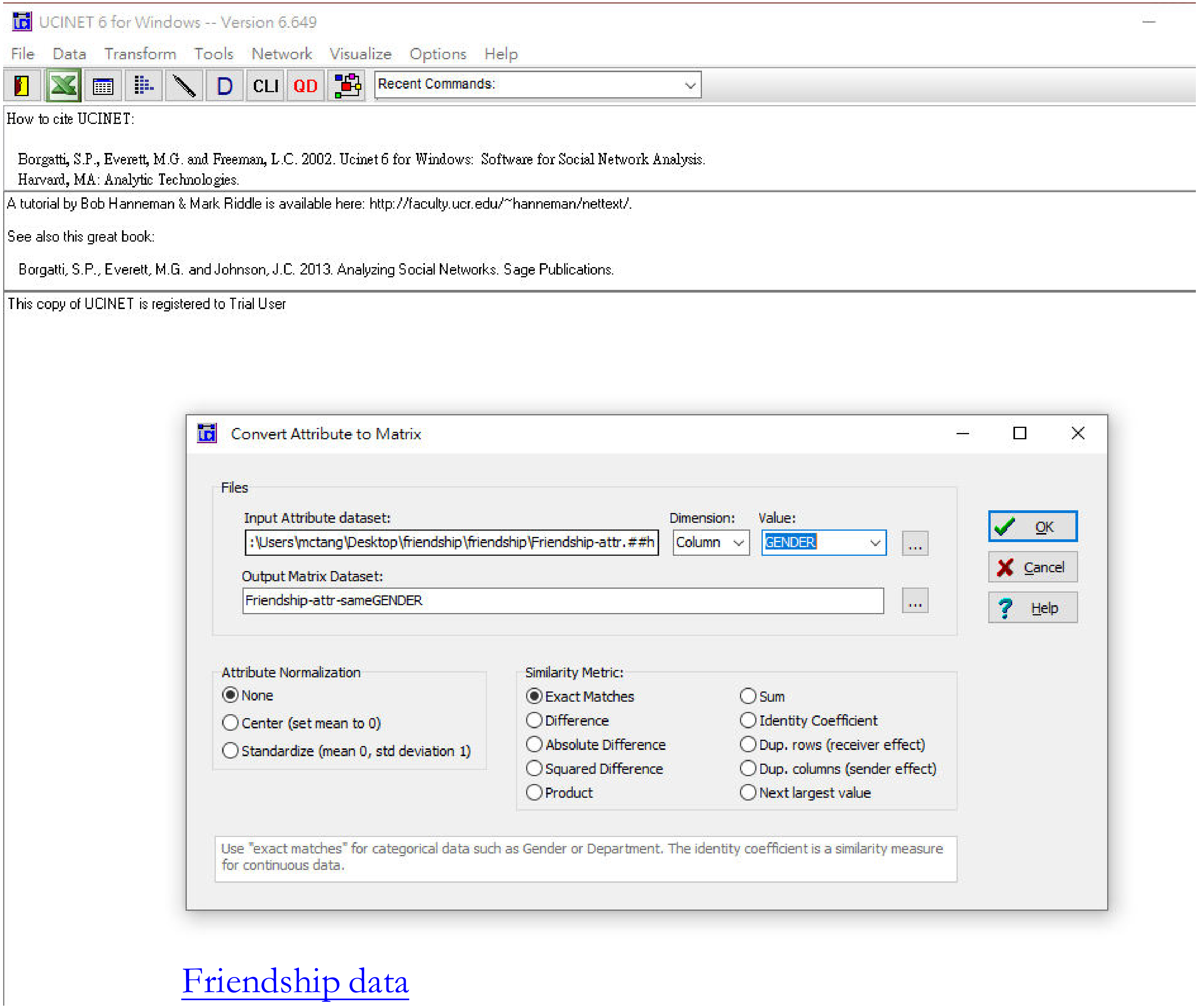
	A	B	C	D	E	F
A	0	–4	3	1	–3	2
B	4	0	7	5	1	6
C	–3	–7	0	–2	–6	–1
D	–1	–5	2	0	–4	1
E	3	–1	6	4	0	5
F	–2	–6	1	–1	–5	0

Figure 5.16 Converting status into relative status using a simple difference.

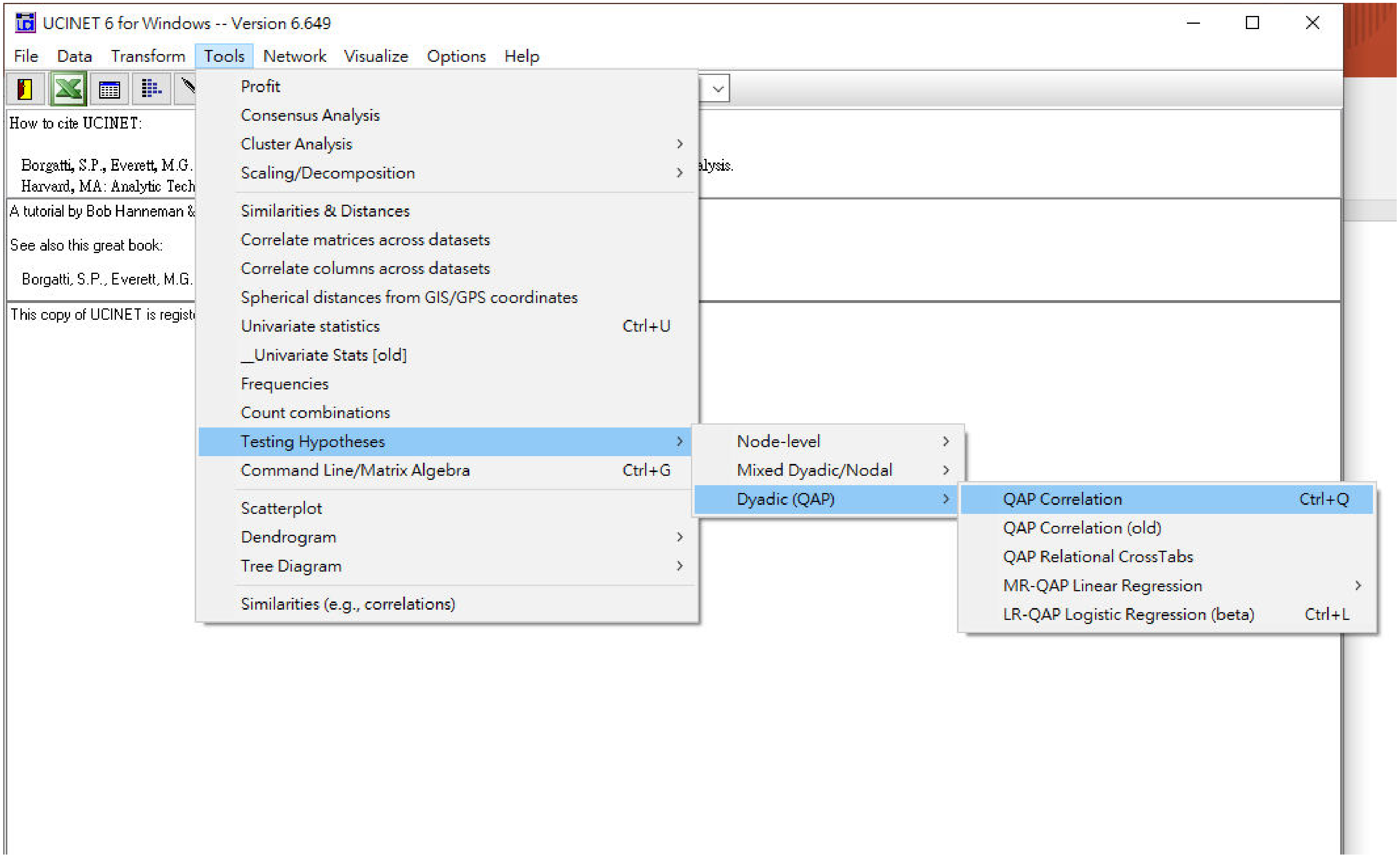
Actor	Status
A	6
B	10
C	3
D	5
E	9
F	4

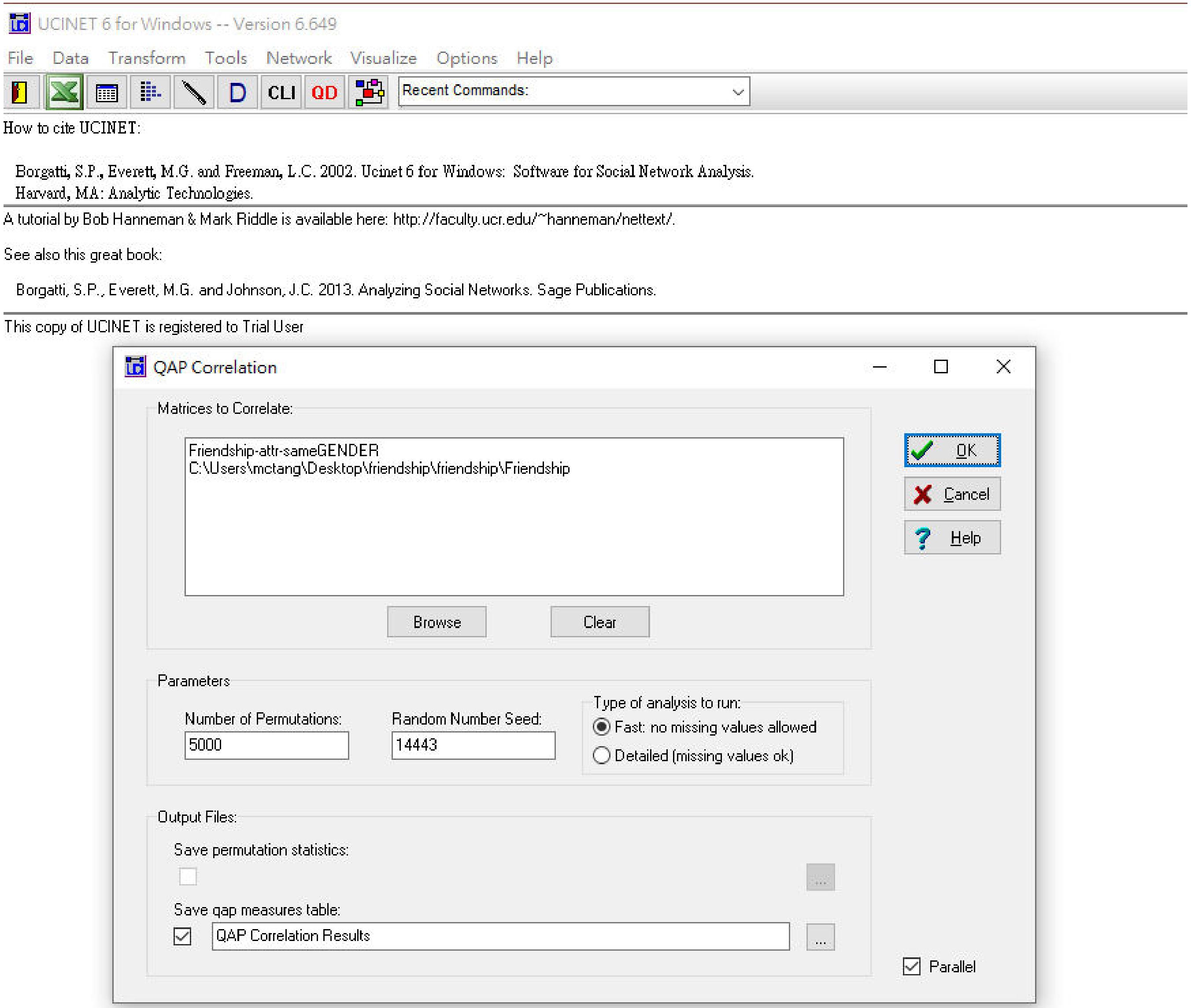
	A	B	C	D	E	F
A	6	10	3	5	9	4
B	6	10	3	5	9	4
C	6	10	3	5	9	4
D	6	10	3	5	9	4
E	6	10	3	5	9	4
F	6	10	3	5	9	4

Figure 5.17 Converting status into 'status of alter'.



[Friendship data]


UCINET 6 for Windows -- Version 6.649
File
Data
Transform
Tools
Network
Visualize
Options
Help
How to cite UCINET:
Borgatti, S.P., Everett, M.G.
Harvard, MA: Analytic Tech
alysis.
A tutorial by Bob Hanneman &
See also this great book:
Borgatti, S.P., Everett, M.G.
This copy of UCINET is regist
Profit
Consensus Analysis
Cluster Analysis
Scaling/Decomposition
Similarities & Distances
Correlate matrices across datasets
Correlate columns across datasets
Spherical distances from GIS/GPS coordinates
Univariate statistics Ctrl+U
_Univariate Stats [old]
Frequencies
Count combinations
Testing Hypotheses
Command Line/Matrix Algebra Ctrl+G
Scatterplot
Dendrogram
Tree Diagram
Similarities (e.g., correlations)
Node-level
Mixed Dyadic/Nodal
Dyadic (QAP)
QAP Correlation Ctrl+Q
QAP Correlation (old)
QAP Relational CrossTabs
MR-QAP Linear Regression
LR-QAP Logistic Regression (beta) Ctrl+L


UCINET 6 for Windows -- Version 6.649
File Data Transform Tools Network Visualize Options Help
CLI QD Recent Commands:
How to cite UCINET:
Borgatti, S.P., Everett, M.G. and Freeman, L.C. 2002. Ucinet 6 for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.
A tutorial by Bob Hanneman & Mark Riddle is available here: http://faculty.ucr.edu/~hanneman/nettext/.
See also this great book:
Borgatti, S.P., Everett, M.G. and Johnson, J.C. 2013. Analyzing Social Networks. Sage Publications.
This copy of UCINET is registered to Trial User
QAP Correlation
Matrices to Correlate:
Friendship-attr-sameGENDER
C:\Users\mctang\Desktop\friendship\friendship\Friendship
Browse
Clear
OK
Cancel
Help
Parameters
Number of Permutations:
5000
Random Number Seed:
14443
Type of analysis to run:
Fast: no missing values allowed
Detailed (missing values ok)
Output Files:
Save permutation statistics:
Save qap measures table:
QAP Correlation Results
Parallel

Testing correlation with SNA data

- The basic idea would be that you would create a data set that had the "dyad" or pair as the unit of analysis.
- The independent variables would be either attributes of each of one or both members of the pairs, or of similarities and / or matches between the pairs.

QAP (Quadratic Assignment Procedure)

- Random scramble
  - Essentially, what the QAP does is to "scramble" the dependent variable data through several permutations. By taking the data, and "scrambling" it repeatedly, resulting in multiple random datasets with the dependent variable—and then multiple analyses can be performed.

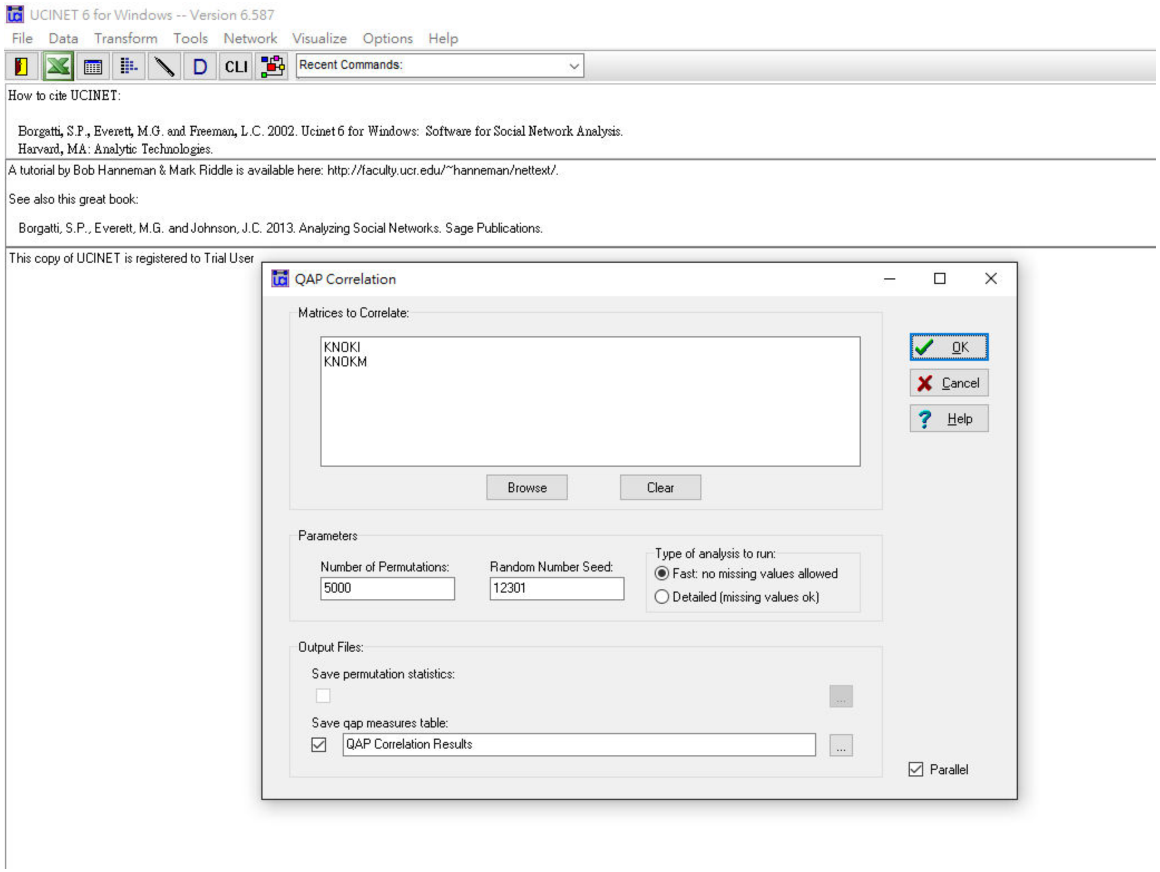
- The generation of a distribution for null-hypothesis
  - Those datasets and analyses form an empirical sampling distribution, and we can compare our coefficient with this sampling distribution of coefficients from all the permuted datasets.

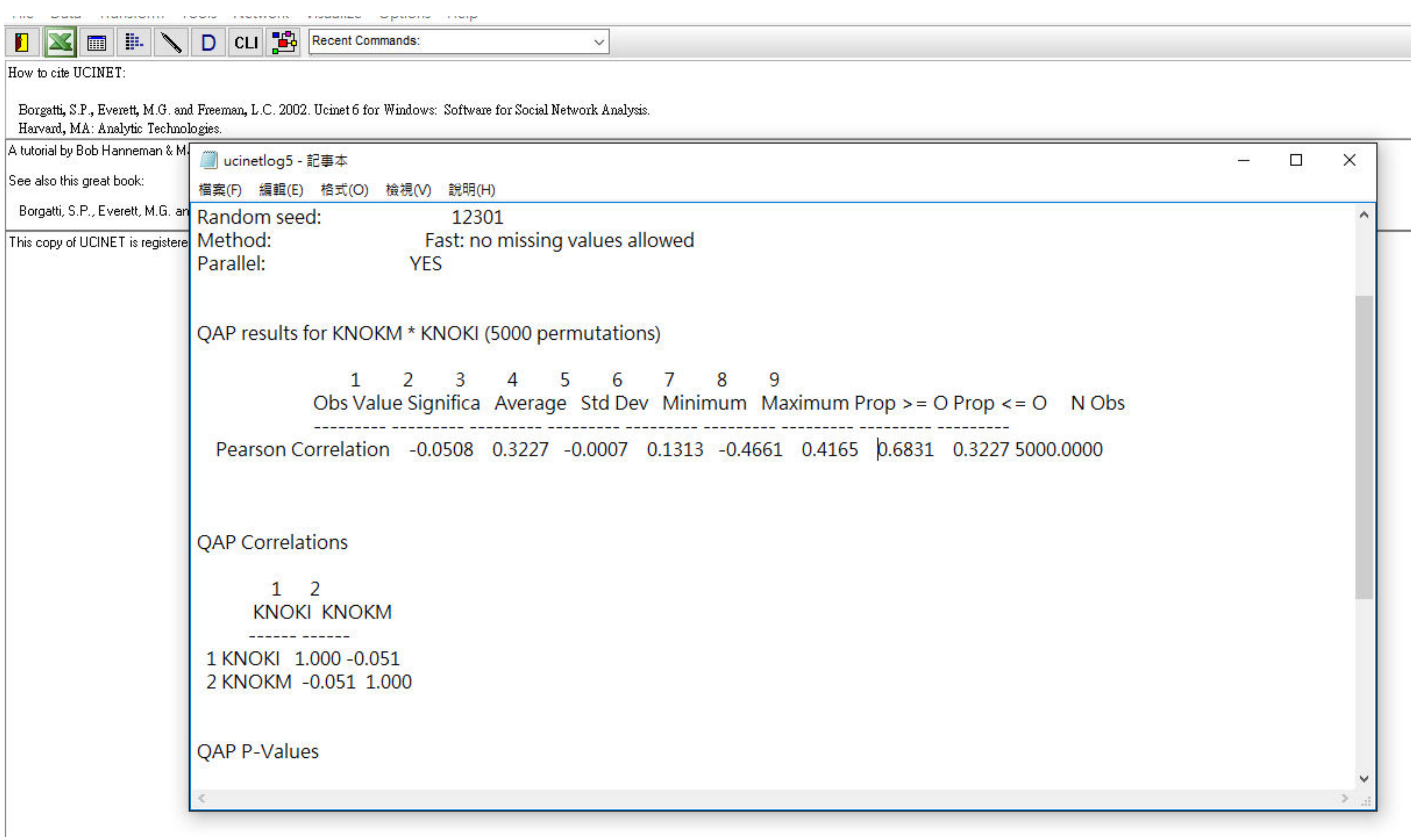
Correlation between two networks with the same actors

- When we have information about multiple relations among the same sets of actors. It is often of considerable interest whether the probability (or strength) of a tie of one type is related to the probability (or strength) of another

- Q: Pairs that engage in one type of exchange (information) are more likely to engage in the other (money)?

Correlation




Recent Commands:
How to cite UCINET:
Borgatti, S.P., Everett, M.G. and Freeman, L.C. 2002. Ucinet 6 for Windows: Software for Social Network Analysis.
Harvard, MA: Analytic Technologies.
A tutorial by Bob Hanneman & M
See also this great book:
Borgatti, S.P., Everett, M.G. an
This copy of UCINET is registere
ucinetlog5 - 記事本
檔案(F) 編輯(E) 格式(O) 檢視(V) 說明(H)
Random seed: 12301
Method: Fast: no missing values allowed
Parallel: YES
QAP results for KNOKM * KNOKI (5000 permutations)
1 2 3 4 5 6 7 8 9
Obs Value Significa Average Std Dev Minimum Maximum Prop >= O Prop <= O N Obs
Pearson Correlation -0.0508 0.3227 -0.0007 0.1313 -0.4661 0.4165 0.6831 0.3227 5000.0000
QAP Correlations
1 2
KNOKI KNOKM
1 KNOKI 1.000 -0.051
2 KNOKM -0.051 1.000
QAP P-Values

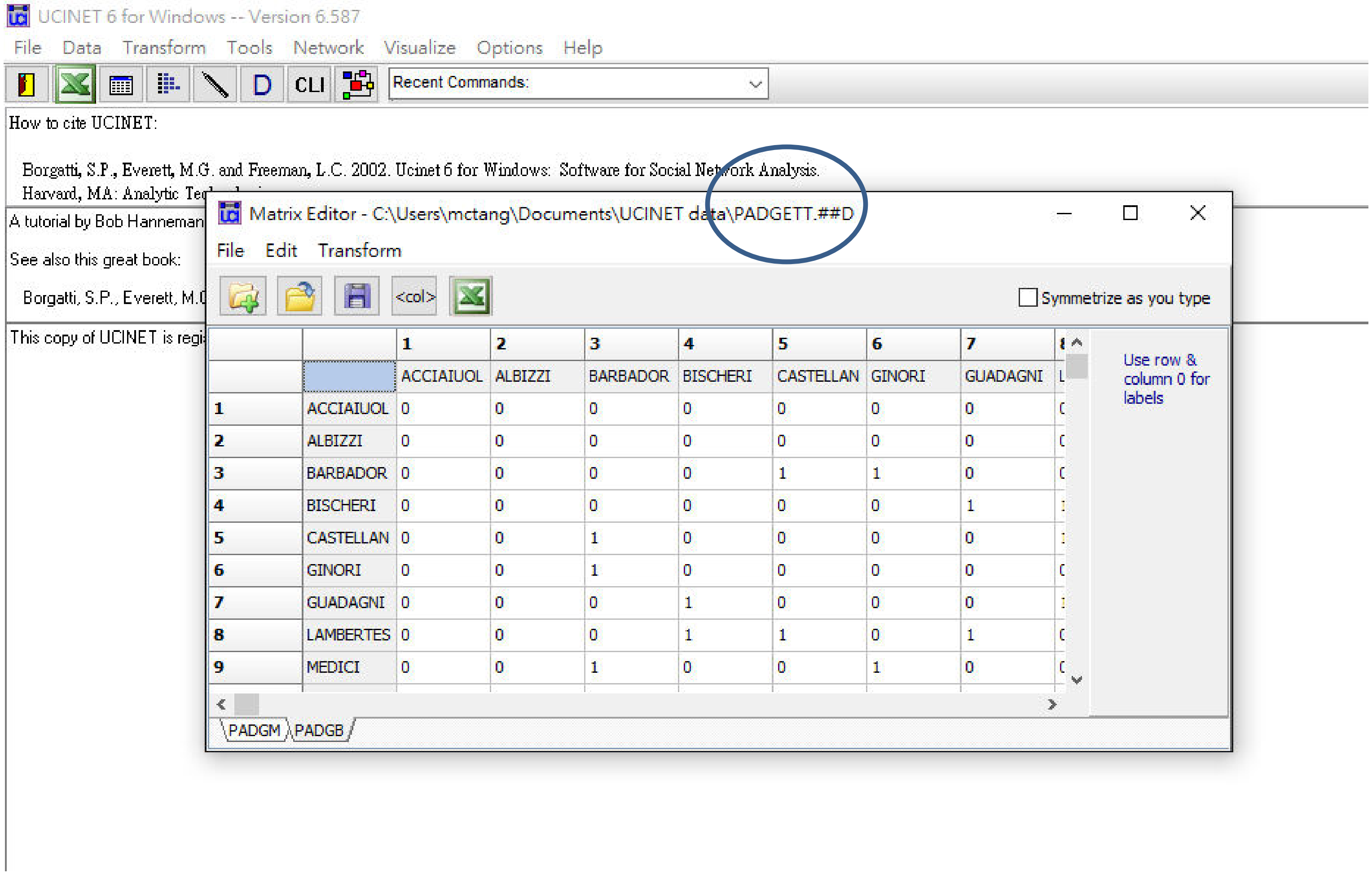
Knoi vs. Knom

Tools>Testing Hypotheses>Dyadic (QAP)>QAP Correlation

		1 Value	2 Signif	3 Avg	4 SD	5 P(Large)	6 P(Small)
1	Pearson Correlation:	-0.051	0.430	-0.004	0.130	0.721	0.430
2	Simple Matching:	0.456	0.721	0.475	0.056	0.721	0.430
3	Jaccard Coefficient:	0.183	0.721	0.203	0.051	0.721	0.430
4	Goodman-Kruskal Gamma:	-0.118	0.430	-0.005	0.288	0.721	0.430
5	Hamming Distance:	49.000	0.721	47.184	5.070	0.430	0.721

The third column (Avg) shows the average value of the measure of association across a large number of trials in which the rows and columns of the two matrices have been randomly permuted. i.e. the mean when the null hypothesis is true

TRY ALSO PADGETT Florence family marriage and business
First you need to upack


UCINET 6 for Windows -- Version 6.587
File Data Transform Tools Network Visualize Options Help
D CLI
Recent Commands:
How to cite UCINET:
Borgatti, S.P., Everett, M.G. and Freeman, L.C. 2002. Ucinet 6 for Windows: Software for Social Network Analysis.
Harvard, MA: Analytic Tec
A tutorial by Bob Hanneman
See also this great book:
Borgatti, S.P., Everett, M.G
This copy of UCINET is regi
Matrix Editor - C:\Users\mctang\Documents\UCINET data\PADGETT.##D
File Edit Transform
<col>
Symmetrize as you type
Use row & column 0 for labels
1 2 3 4 5 6 7
ACCIAIUOL ALBIZZI BARBADOR BISCHERI CASTELLAN GINORI GUADAGNI
1 ACCIAIUOL 0 0 0 0 0 0 0
2 ALBIZZI 0 0 0 0 0 0 0
3 BARBADOR 0 0 0 0 1 1 0
4 BISCHERI 0 0 0 0 0 0 1
5 CASTELLAN 0 0 1 0 0 0 0
6 GINORI 0 0 1 0 0 0 0
7 GUADAGNI 0 0 0 1 0 0 0
8 LAMBERTES 0 0 0 1 1 0 1
9 MEDICI 0 0 1 0 0 1 0
PADGM PADGB

```
QAP CORRELATION
--------------------------------------------------------------------------------

Data Matrices:                        padgm
                                      padgb
# of Permutations:                    50000
Random seed:                          24322
Method:                               Detailed (missing values ok)


QAP results for padgb * padgm (50000 permutations)

                                1         2         3         4         5         6         7         8
                        Obs Value Significa   Average   Std Dev   Minimum   Maximum Prop >= O Prop <= O
                        --------- --------- --------- --------- --------- --------- --------- ---------
  1  Pearson Correlation   0.3719    0.0007    0.0002    0.0924   -0.1690    0.5071    0.0007    0.9999
  2   Euclidean Distance   4.3589    0.0007    5.4709    0.2529    3.8730    5.9161    0.9999    0.0007
  3     Hamming Distance   0.1583    0.0007    0.2500    0.0228    0.1250    0.2917    0.9999    0.0007
  4           Match Coef   0.8417    0.0007    0.7500    0.0228    0.7083    0.8750    0.0007    0.9999
  5         Jaccard Coef   0.2963    0.0007    0.0790    0.0464    0.0000    0.4000    0.0007    0.9999
  6 Goodman-Kruskal Gamma  0.7971    0.0007   -0.0690    0.3845   -1.0000    0.9000    0.0007    0.9999
  7         Hubert Gamma   8.0000    0.0007    2.5025    1.3668    0.0000   13.0000    0.0007    0.9999

NOTE: When you have missing data, the significance of Hubert's Gamma and Euclidean Distance will differ from
that of Pearson Correlation. Otherwise, they should be the same (unless the correlation is negative).


QAP Correlations

            1     2
        padgm padgb
        ----- -----
  1 padgm 1.000 0.372
  2 padgb 0.372 1.000


QAP P-Values

            1     2
        padgm padgb
        ----- -----
  1 padgm 0.000 0.001
  2 padgb 0.001 0.000

QAP statistics saved as datafile QAP C
```

Figure 8.1 Results of QAP corre

UCINET 6 for Windows -- Version 6.587

File Data Transform Tools Network Visualize Options Help

Recent Commands:

How to cite UCINET:

Borgatti, S.P., Everett, M.G. and Freeman, L.C. 2002. Ucinet 6 for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.

A tutorial by Bob Hanneman & M

See also this great book:

Borgatti, S.P., Everett, M.G. an

This copy of UCINET is registere

ucinetlog9 - 記事本

檔案(F) 編輯(E) 格式(O) 檢視(V) 說明(H)

```
Data Matrices:             PADGB
                      PADGM
# of Permutations:           5000
Random seed:               8446
Method:                 Fast: no missing values allowed
Parallel:               YES


QAP results for PADGM * PADGB (5000 permutations)

                       1         2         3         4         5         6         7         8         9
               Obs Value Significa   Average   Std Dev   Minimum   Maximum Prop >= O Prop <= O     N Obs
               --------- --------- --------- --------- --------- --------- --------- ---------
 Pearson Correlation  0.3719    0.0008    0.0035    0.0937   -0.1690    0.3719    0.0008    0.9994 5000.0000


QAP Correlations

         1     2
     PADGB PADGM
     ----- -----
 1 PADGB 1.000 0.372
 2 PADGM 0.372 1.000
```

- **TOOLS > STATISTICS > MATRIX (QAP) > QAP-CORRELATION**

  PURPOSE Compute correlation and other similarity measures between entries of two square matrices, and assess the frequency of random measures as large as actually observed.

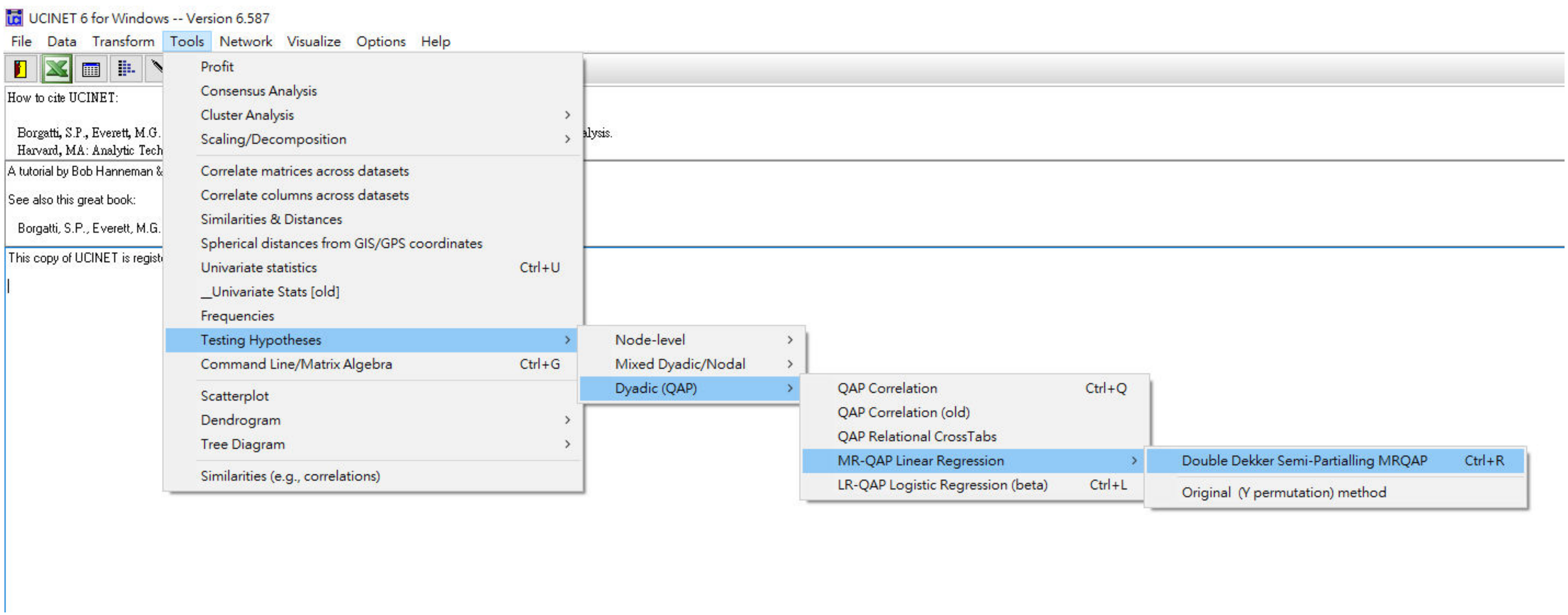
  DESCRIPTION The procedure is principally used to test the association between networks.

It randomly permutes rows and columns (synchronously) of one matrix (the observed matrix, if the distinction is relevant) and recomputes the correlation, which is carried out hundreds of times in order to compute the proportion of times that a random measure is larger than or equal to the observed measure.

A low proportion (< 0.05) suggests a strong relationship between the matrices that is unlikely to have occurred by chance.

Regression, dyad level

RQ: Can friendship (informal relationship) and report-to (formal) explain Advice seeking behaviors? Which has a higher explanatory power?



```
MULTIPLE REGRESSION QAP VIA SEMI-PARTIALLING
--------------------------------------------------------------------------

# of permutations:          10000
Diagonal valid?             NO
Random seed:                824
Dependent variable:         advice
Expected values:            F:\Data\DataFiles\mrqap-predicted
Independent variables:      REPORTS_TO
                            FRIENDSHIP
```

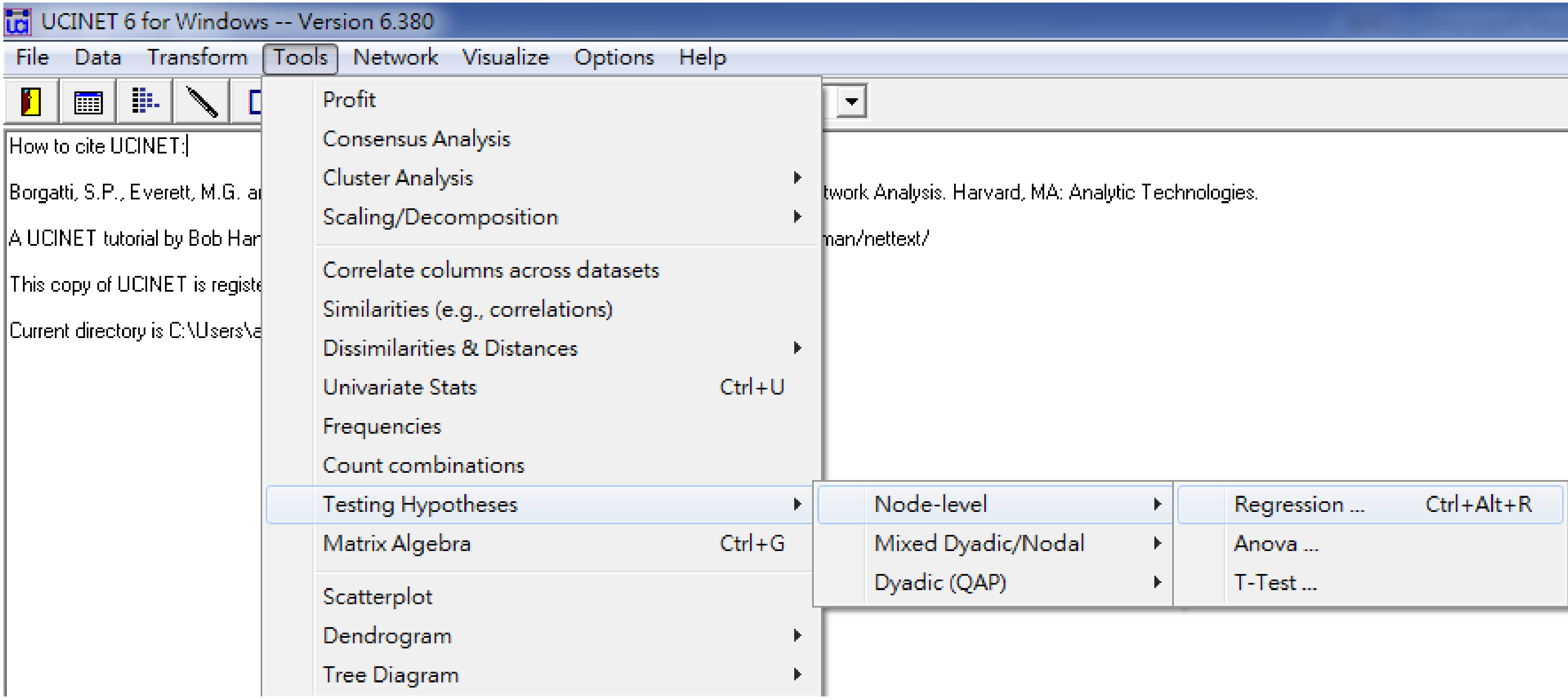
yes or no

binary variable

```
Number of permutations performed: 10000


MODEL FIT

R-square Adj R-Sqr Probability   # of Obs
-------- --------- ----------- -----------
   0.063     0.061       0.000         420


REGRESSION COEFFICIENTS
                                                          Significant
                Un-stdized      Stdized                 Proportion  Proportion
 Independent   Coefficient  Coefficient  Significance    As Large    As Small
 -----------   -----------  -----------  ------------  ----------- -----------
   Intercept      0.396942     0.000000
  REPORTS_TO      0.471569     0.201767        0.000        0.000       1.000
  FRIENDSHIP      0.135815     0.117009        0.061        0.061       0.939

----------------------------------------
Running time:  00:00:01
Output generated:  21 Nov 04 11:39:54
Copyright (c) 1999-2004 Analytic Technologies
```

Figure 8.2 Results of MR-QAP regression.

Regression, node level



Testing whether how well institutional property (socialist, capitalist , and other) and donation made influence centrality

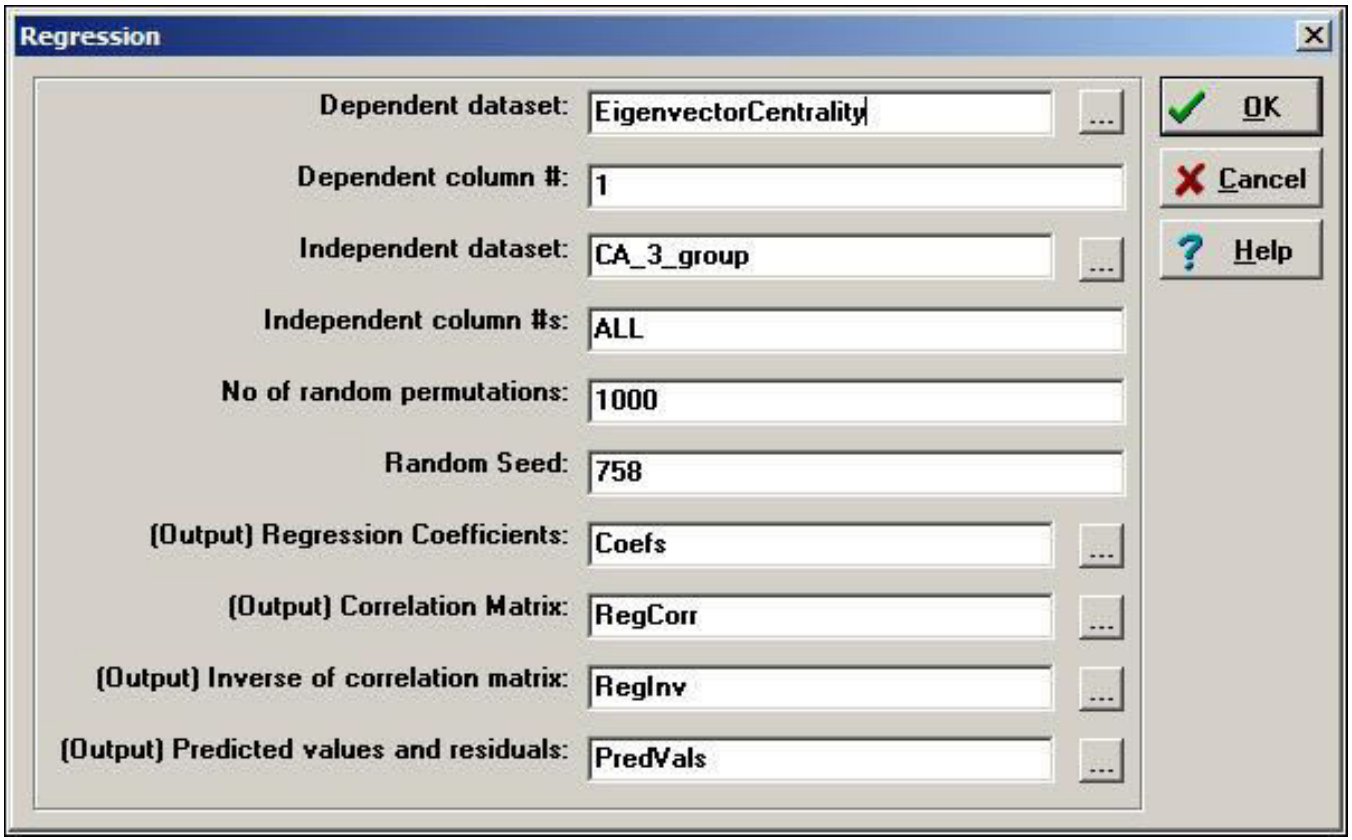


Figure 18.14. Dialog for *Tools>Testing Hypotheses>Node-level>Regression* for California donor's eigenvector centrality

```
TOOLS>STATISTICS>ANOVA
------------------------------------------------------------------------------

Dependent variable:              EigenvectorCentrality Col 1
Independent variable:            CA_3_group Col 1
# of permutations:               5000
Random seed:                     20662


        ANALYSIS OF VARIANCE

         Source               DF              SSQ   F-Statistic   Significance
 ============= ============= ============= ============= =============
      Treatment                2             0.21       34.4108         0.0002
          Error               20             0.06
          Total               22             0.28

R-Square/Eta-Square: 0.775
```

Figure 18.12. One-way ANOVA of eigenvector centrality of California political donors, with permutation-based standard errors and tests

CORRELATION MATRIX

	1	2	3	4
1	1.000	-0.483	-0.411	-0.480
2	-0.483	1.000	0.763	0.878
3	-0.411	0.763	1.000	0.970
4	-0.480	0.878	0.970	1.000

Determinant = 0.31843132

NOTE: All probabilities based on randomization tests.

MODEL FIT

R-square	Adjusted R-square	F Value	One-Tailed Probability
0.987	0.984	482.655	0.014

REGRESSION COEFFICIENTS

Independent	Un-stdized Coefficient	St'dized Coefficient	Proportion As Large	Proportion As Small	Proportion As Extreme
Intercept	0.003293	0.000000	1.000	0.000	1.000
CAP	-0.007767	-0.033723	0.555	0.445	0.911
WORK	0.075367	0.316154	0.203	0.797	0.423
POSCOAL	0.061454	0.714805	0.021	0.979	0.043

Figure 18.15. Multiple regression of eigenvector centrality with permutation based significance tests

Exercise

- Correlation between HHI and Golden melody directed network

Dependent variable: newc1D

Overall fit of the logistic regression model

	1	2	3	4	5
	Log Lik	Pseudo Rsq	Sig	Obs	Perms
1 Statistics:	-100.906	0.259	0.000	272	10000

1 rows, 5 columns, 1 levels.

LR Coefficients

		1	2	3	4	5
		Coef	OddsRat	Sig	StdErr	Avg
1	Intercept	-2.614	-8.654	0.000	0.210	-1.462
2	newc0D	2.290	9.880	0.000	0.426	-0.011
3	newc0D-Reciprocity	0.818	2.267	0.008	0.362	-0.009
4	newc0D-Transitivity (Closure)	0.598	1.818	0.071	0.398	0.010

Figure 8.3 Logistic regression results.

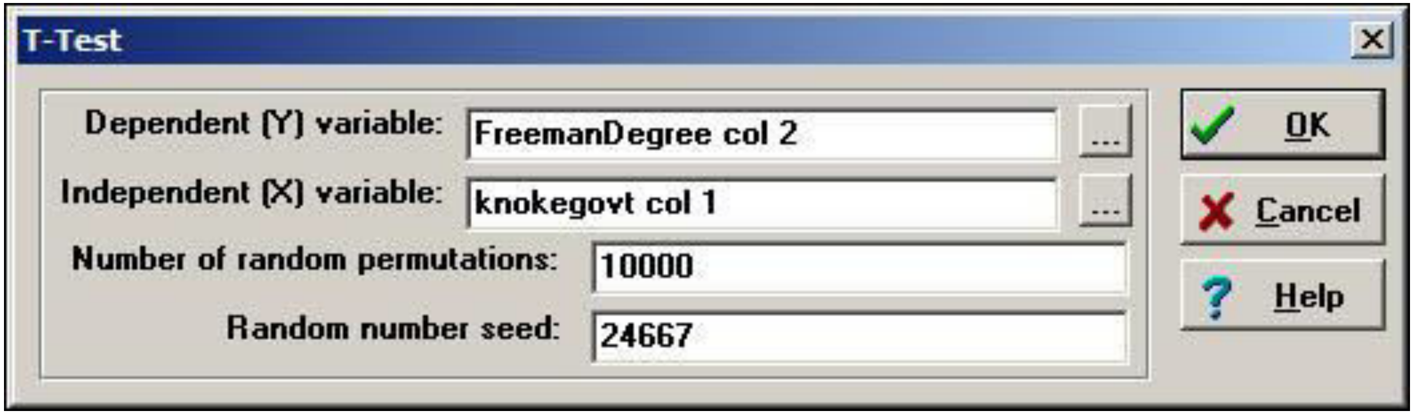
Attribute-based comparison

- Relationships between attributes and network data
  - Does an actor's seniority (gender/background) influence their betweenness centrality?
- Relationships between two attributes
  - Is there an association between actor's gender and academic performance?

Hypotheses about the means of two groups

- Is there a difference of centrality between male and female faculty members?
- Tools>Testing Hypotheses>Node-level>T-test

Figure 18.10. Dialog for *Tools>Testing Hypotheses>Node-level>T-Test*



- Test if the mean degree centrality of government organizations is lower than the mean degree centrality on non-government organization.

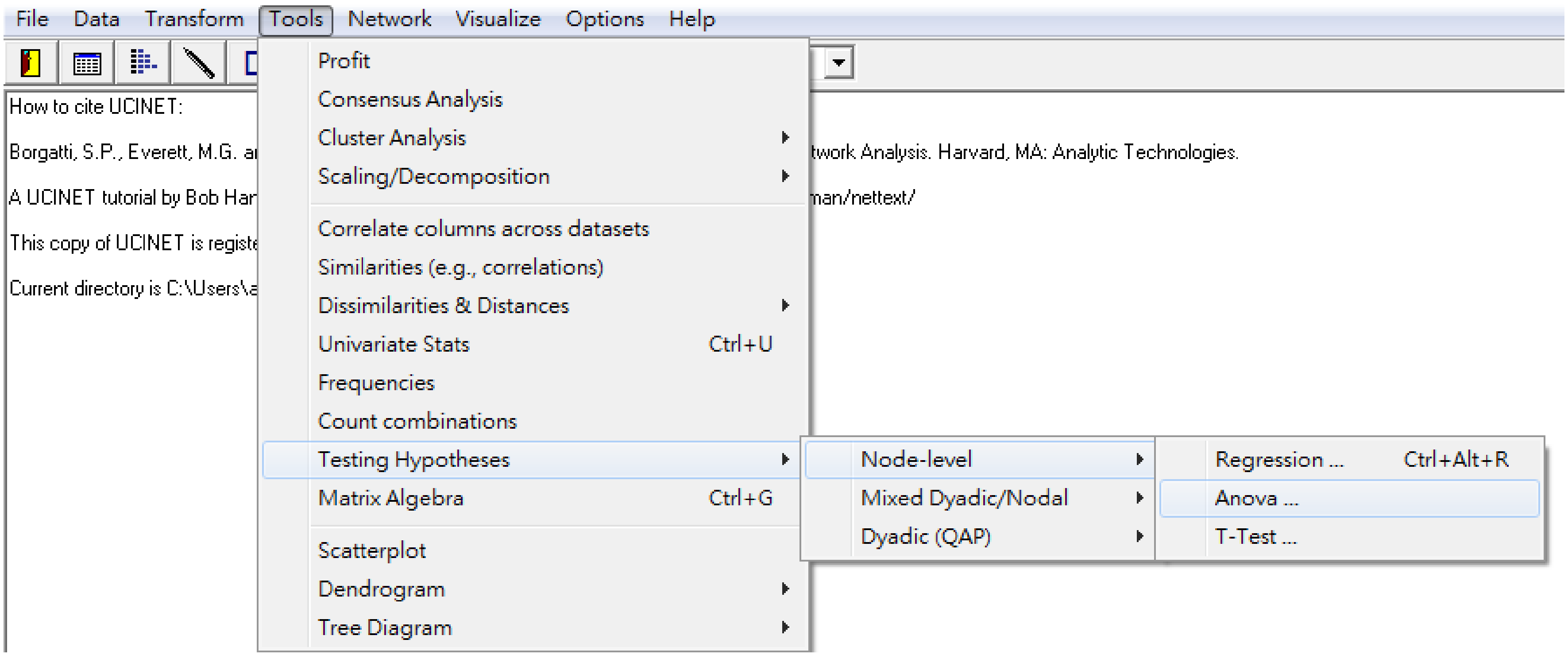
```
Dependent variable:              FreemanDegree col 2
Independent variable:            knokegovt col 1
# of permutations:               10000
Random seed:                     24667


Basic statistics on each group.

                           1         2
                     Group 1   Group 2
                   --------- ---------
  1       Mean        75.000    68.519
  2    Std Dev         9.213    21.675
  3        Sum       300.000   411.111
  4   Variance        84.877   469.822
  5        SSQ     22839.506 30987.652
  6      MCSSQ       339.506  2818.930
  7   Euc Norm       151.127   176.033
  8    Minimum        66.667    33.333
  9    Maximum        88.889   100.000
 10   N of Obs         4.000     6.000


SIGNIFICANCE TESTS

     Difference          ...One-Tailed Tests...         Two-Tailed
       in Means     Group 1 > 2     Group 2 > 1              Test
 ============== ============== ============== ==============
          6.481          0.334          0.750         0.6268
```

Figure 18.11. Test for difference in mean normed degree centrality of Knoke government and non-government organizations

Hypotheses about the means of multiple groups

- Is there a relationship between a faculty's centrality and seniority? (full, associate, assistant professor) or Is there difference of centrality among the three group of faculty?
- Tools>Testing Hypotheses>Node-level>Anova

ANOVA



Testing whether institutional property (socialist, capitalist , and other) influences degree centrality

ANOVA

ucinetlog14.txt - 記事本

檔案(F) 編輯(E) 格式(O) 檢視(V) 說明(H)

TOOLS>STATISTICS>ANOVA

Dependent variable: "I:\◆◆◆|◆◆◆demo\0416\FreemanDegree.##h" Col 1
Independent variable: "I:\◆◆◆|◆◆◆demo\0416\anova_attribute.##h" Col 1
# of permutations: 5000
Random seed: 29830

ANALYSIS OF VARIANCE

Source	DF	SSQ	F-Statistic	Significance
Treatment	2	674.31	4.9716	0.0180
Error	20	1356.30		
Total	22	2030.61		

R-Square/Eta-Square: 0.332

Running time: 00:00:01
Output generated: 22 ◆|◆◆ 13 16:51:39
Copyright (c) 2002-12 Analytic Technologies

Hypotheses about two paired means or densities

- Network>Compare densities>Paired (same node)
  - First unpack different spreadsheets in the same file

Density between money and information exchange

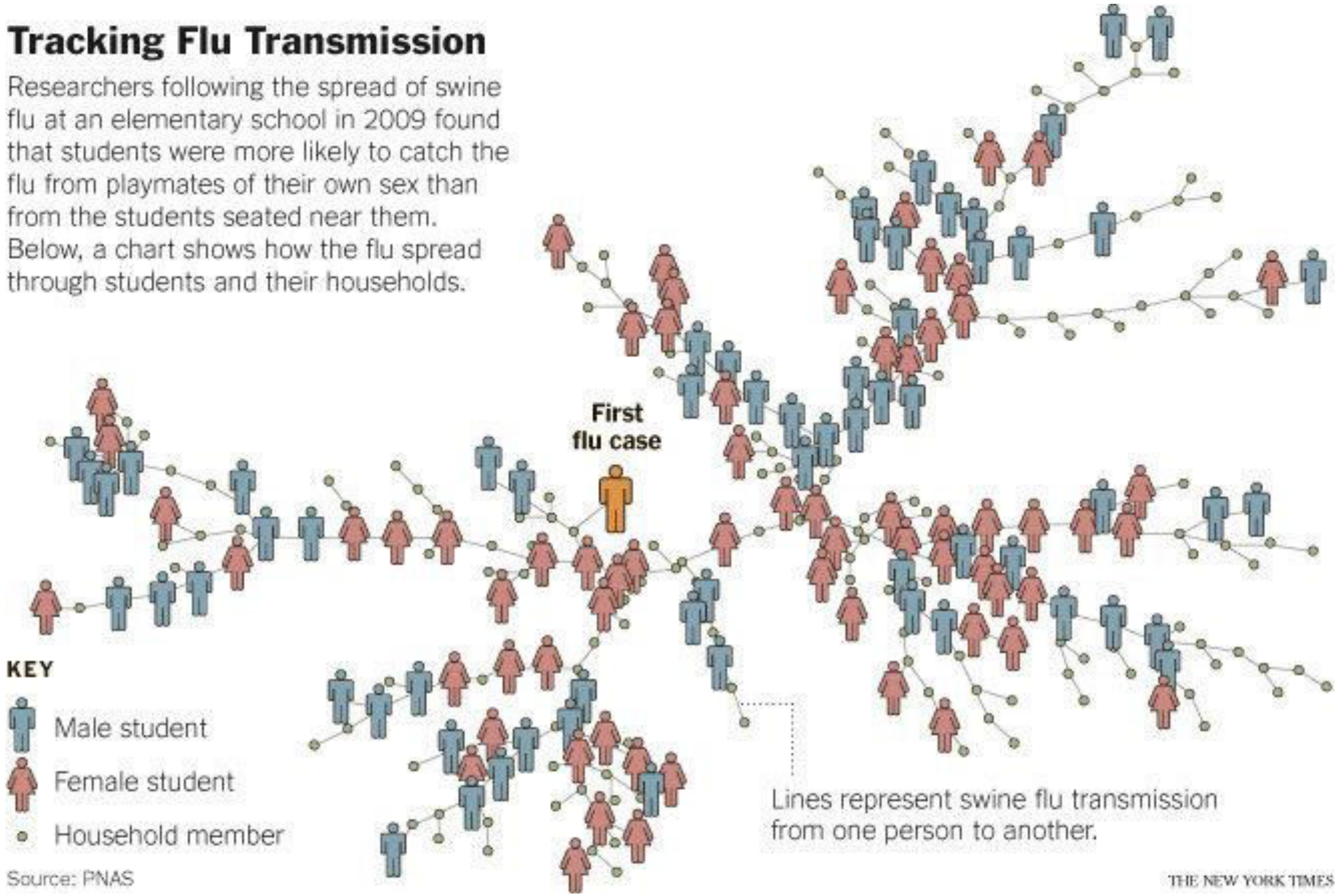
```
BOOTSTRAP PAIRED SAMPLE T-TEST
------------------------------------------------------------------------------

Density of KNOKI is: 0.5444
Density of KNOKM is: 0.2444
Difference in density is: 0.3000

Number of bootstrap samples: 10000
Variance of ties for KNOKI: 0.2508
Variance of ties for KNOKM: 0.1868
Classical standard error of difference: 0.0697
Classical t-test (indep samples): 4.3024
Estimated bootstrap standard error for density of KNOKI: 0.0965
Estimated bootstrap standard error for density of KNOKM: 0.0775
Bootstrap standard error of the difference (indep samples): 0.1237
95% confidence interval for the difference (indep samples): [0.0575, 0.5425]
bootstrap t-statistic (indep samples): 2.4247
Bootstrap SE for the difference (paired samples): 0.1259
95% bootstrap CI for the difference (paired samples): [0.0531, 0.5469]
t-statistic: 2.3820
Average bootstrap difference: 0.2587
Proportion of absolute differences as large as observed: 0.0178
Proportion of differences as large as observed: 0.0052
Proportion of differences as large as observed: 0.9949
```

Figure 18.7. Test for the difference of density in the Knoke information and money exchange relations

Social network analysis

Social contagion

Social distance trumps physical distance



“Ideas and products and message and behaviors spread just like viruses do”

(Malcolm Gladwell, The Tipping Point)

Social epidemics, like infectious diseases, are passed along by a handful of exceptional actors

Meme (“Memory” + “gene”) the mind “virus” (**Richard Dawkins**)

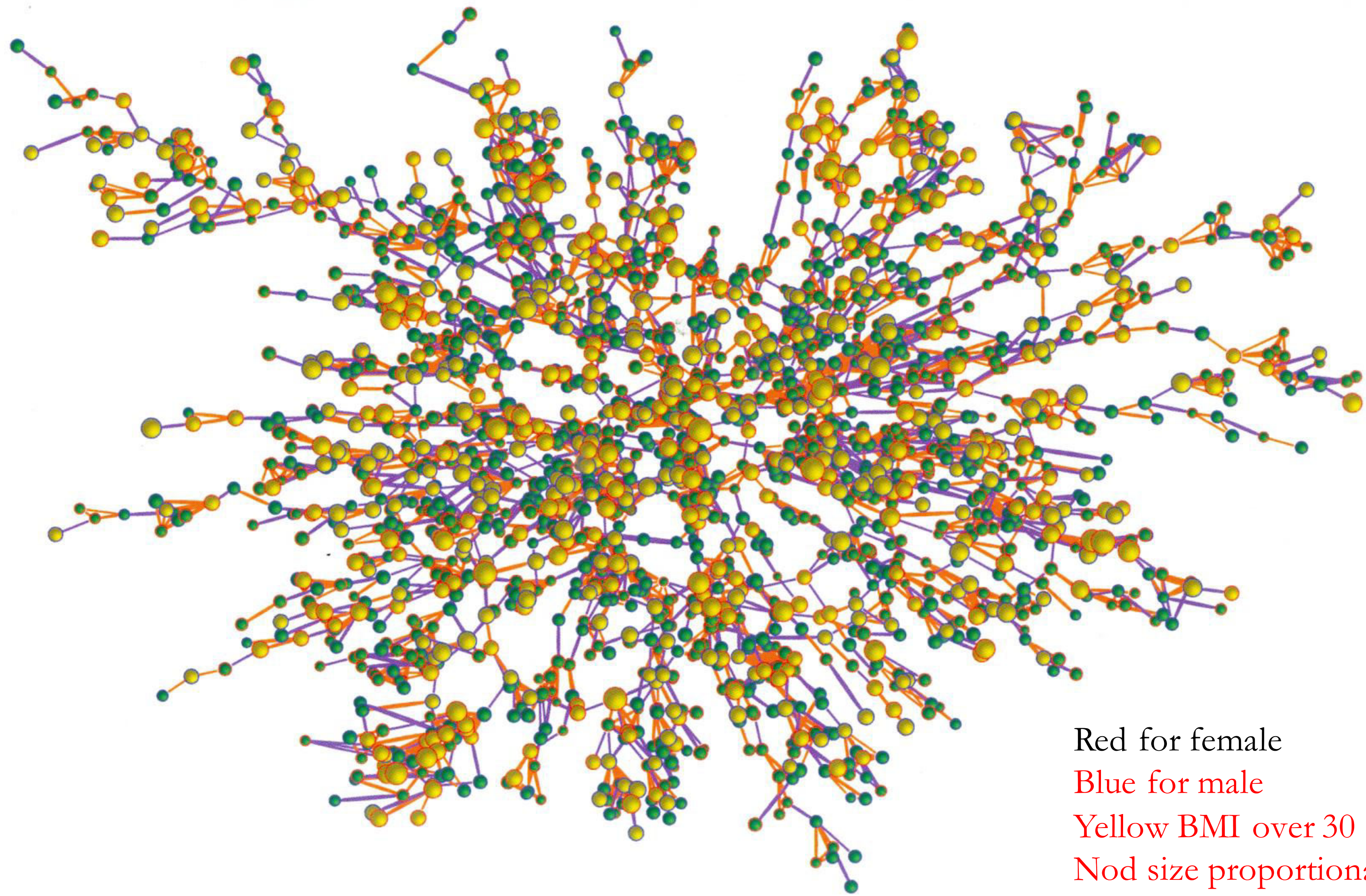
Any idea or behavior that can pass from one person to another by learning or imitation. Examples include thoughts, ideas, theories, gestures, practices, fashions, habits, songs, and dances

Obesity

- The average obese person was more likely to have friends, friends of friends, and friends of friends of friends who were obese than would be expected due to chance alone.
- The average nonobese person was, similarly, more likely to have nonobese contacts up to three degree of separation, beyond 3 degrees, the clustering stopped.

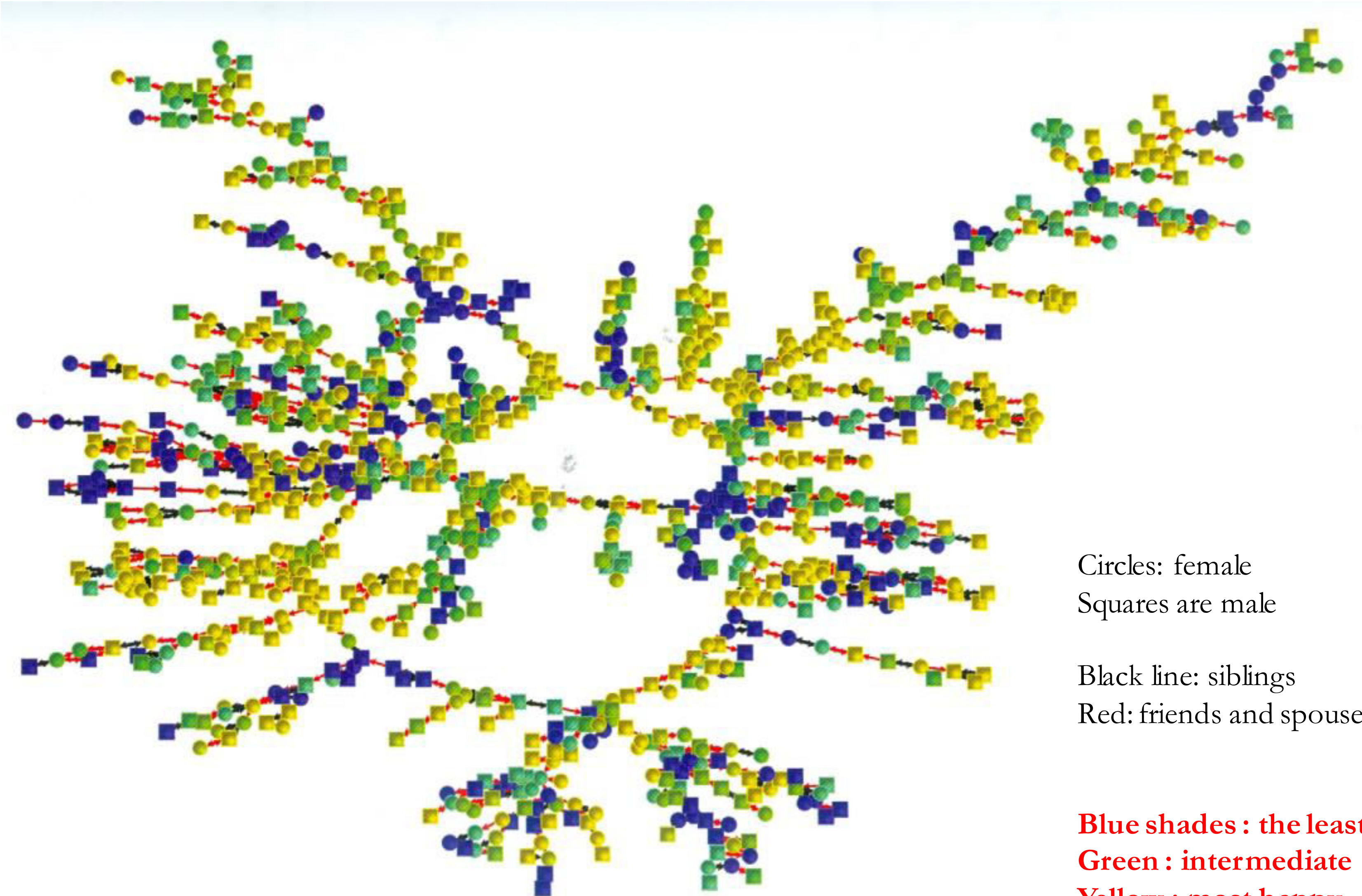
[Nicholas Christakis TED talk]

Obesity and social influences



Red for female
Blue for male
Yellow BMI over 30
Nod size proportional to BMI
Green BMI less than 30
Tie colors indicate relationships
Purple for friend or spouse, orange for family

Happiness and social influence?

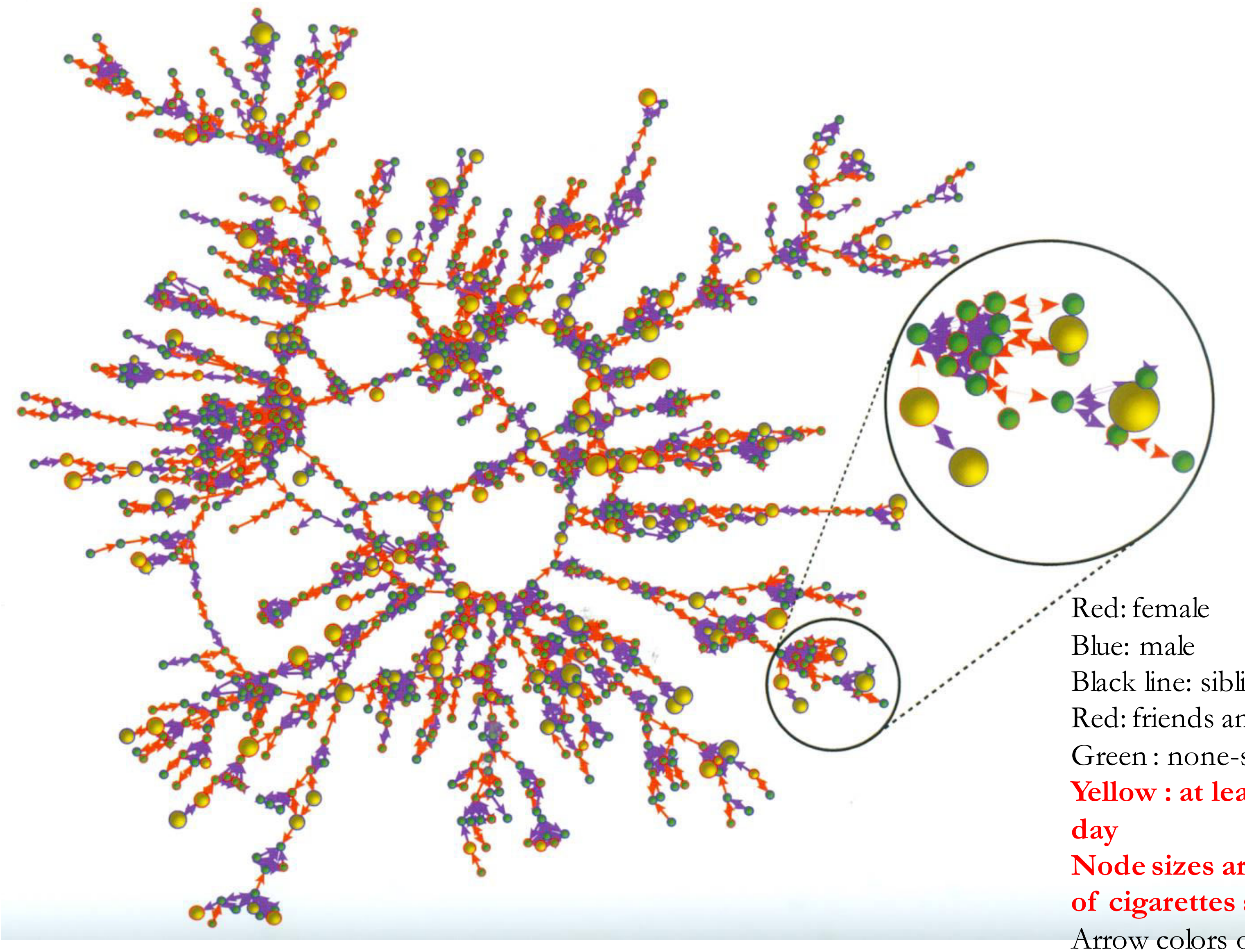


Circles: female
Squares are male

Black line: siblings
Red: friends and spouses

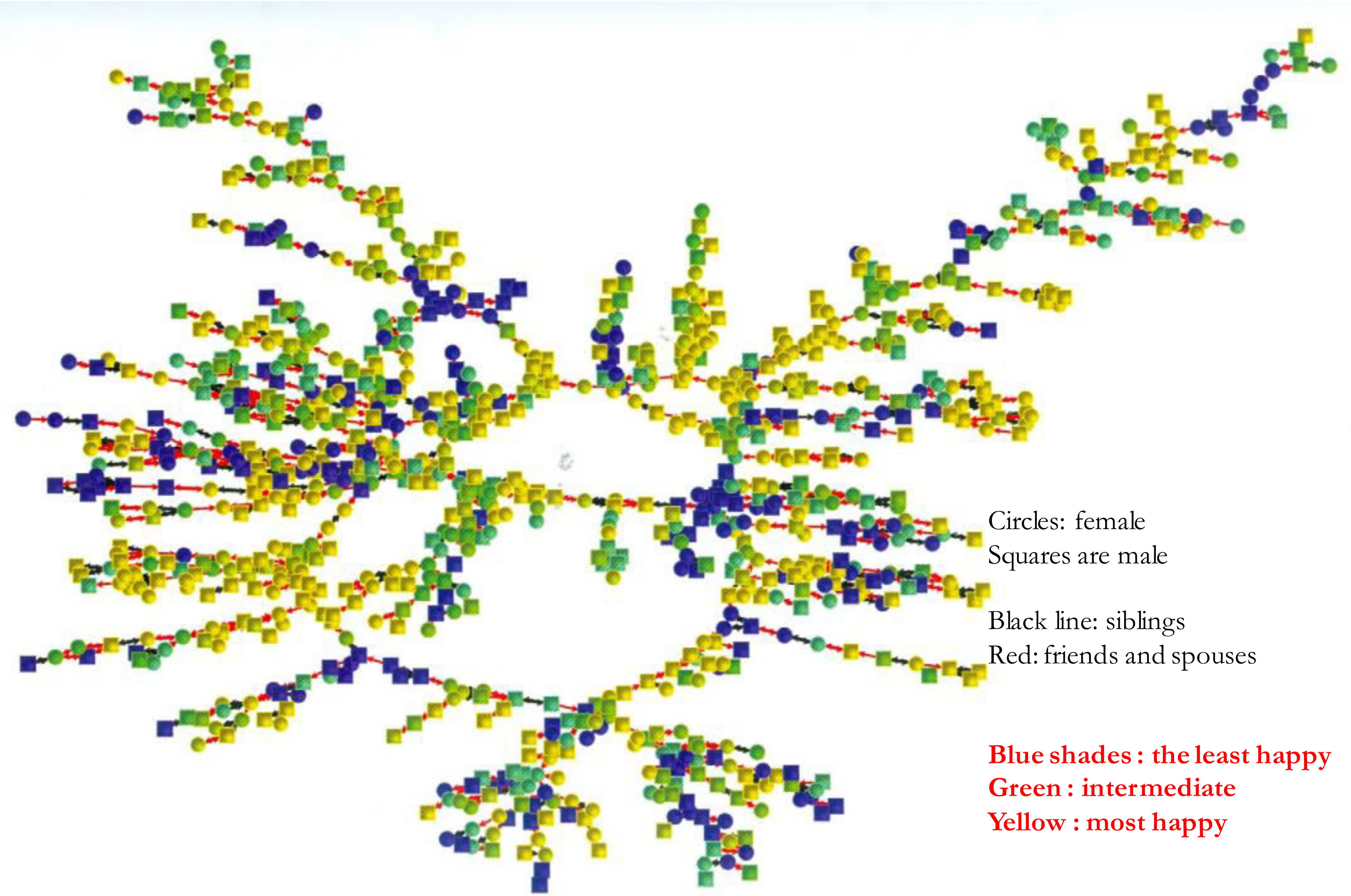
Blue shades : the least happy
Green : intermediate
Yellow : most happy

Smoking and social influence



Red: female
Blue: male
Black line: siblings
Red: friends and spouses
Green : none-smoker
Yellow : at least one cigarette a day
Node sizes are proportional to # of cigarettes smoked
Arrow colors orange : friends and spouse
Purple for family

Happiness and social influence?

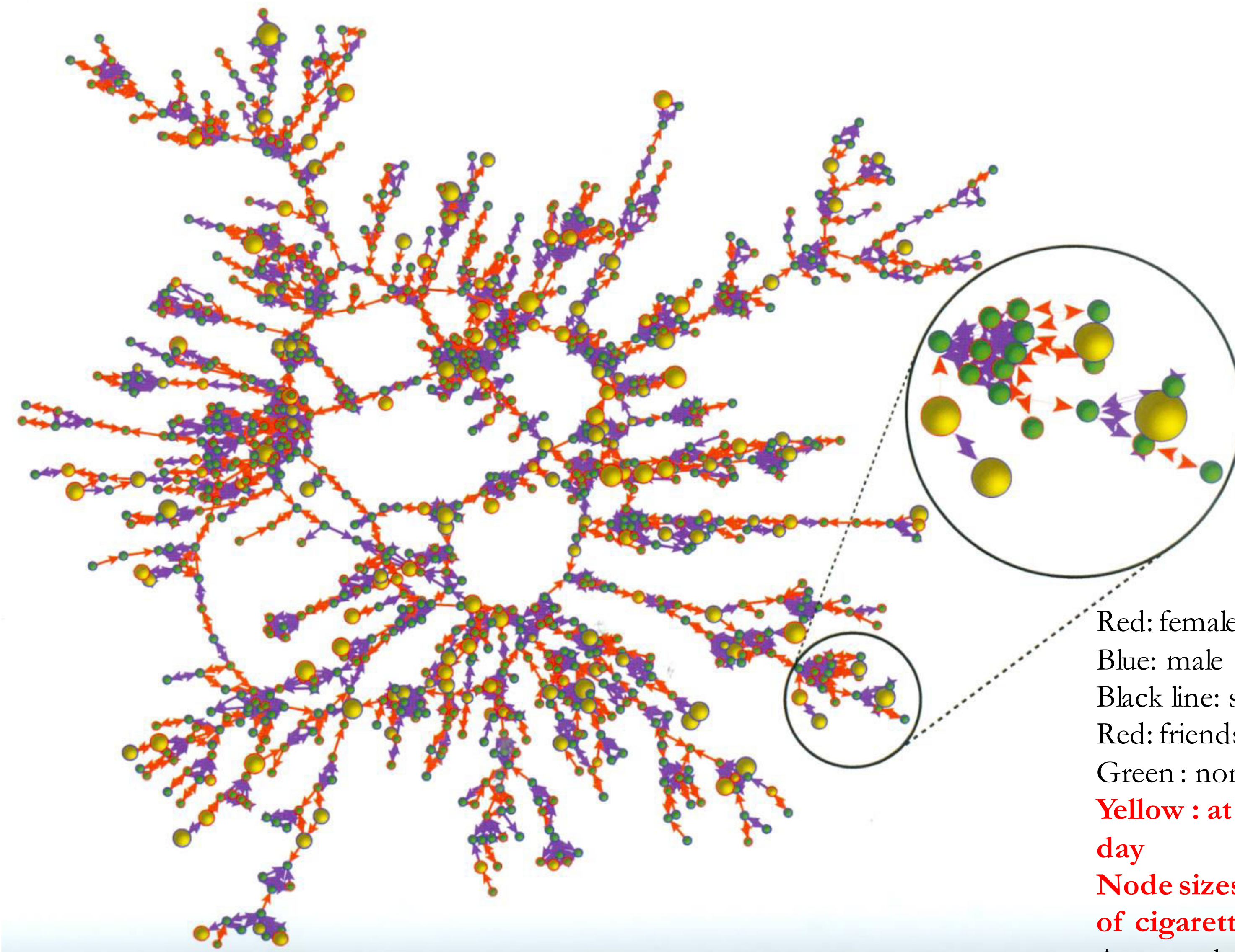


Circles: female
Squares are male

Black line: siblings
Red: friends and spouses

Blue shades : the least happy
Green : intermediate
Yellow : most happy

Smoking and social influence



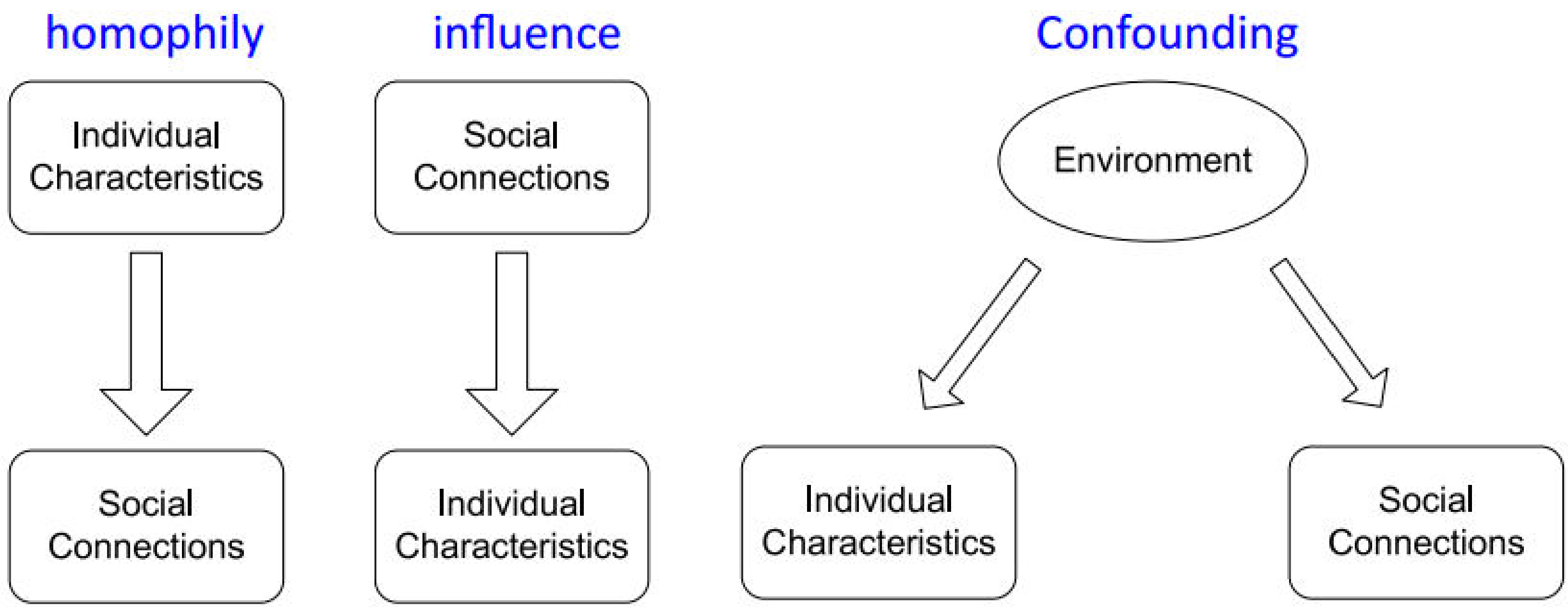
Red: female
Blue: male
Black line: siblings
Red: friends and spouses
Green : none-smoker
Yellow : at least one cigarette a day
Node sizes are proportional to # of cigarettes smoked
Arrow colors orange : friends and spouse
Purple for family

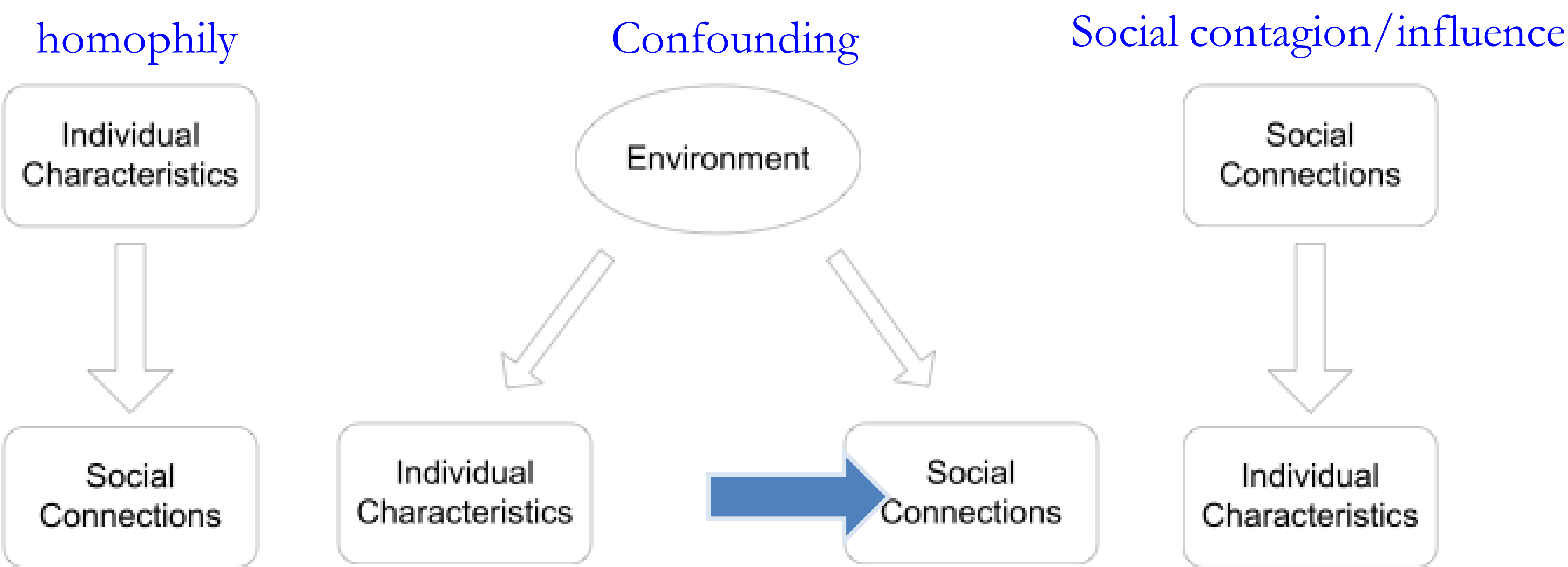
Some observations of life in the Network (Christakis and Fowler, 2009)

- We shape our network
  - Homophile or selection: “love of being alike”
- Our network shapes us
  - Social conformity, imitation, network effect
- Our friends affect us
- Our friends’ friends affect us
- The network work has a life of its own

Structure and agency

- Network analysis allows us to examine how the configuration of networks influences how individuals and groups, organizations, or systems function.


homophily
Individual Characteristics
Social Connections
influence
Social Connections
Individual Characteristics
Confounding
Environment
Individual Characteristics
Social Connections

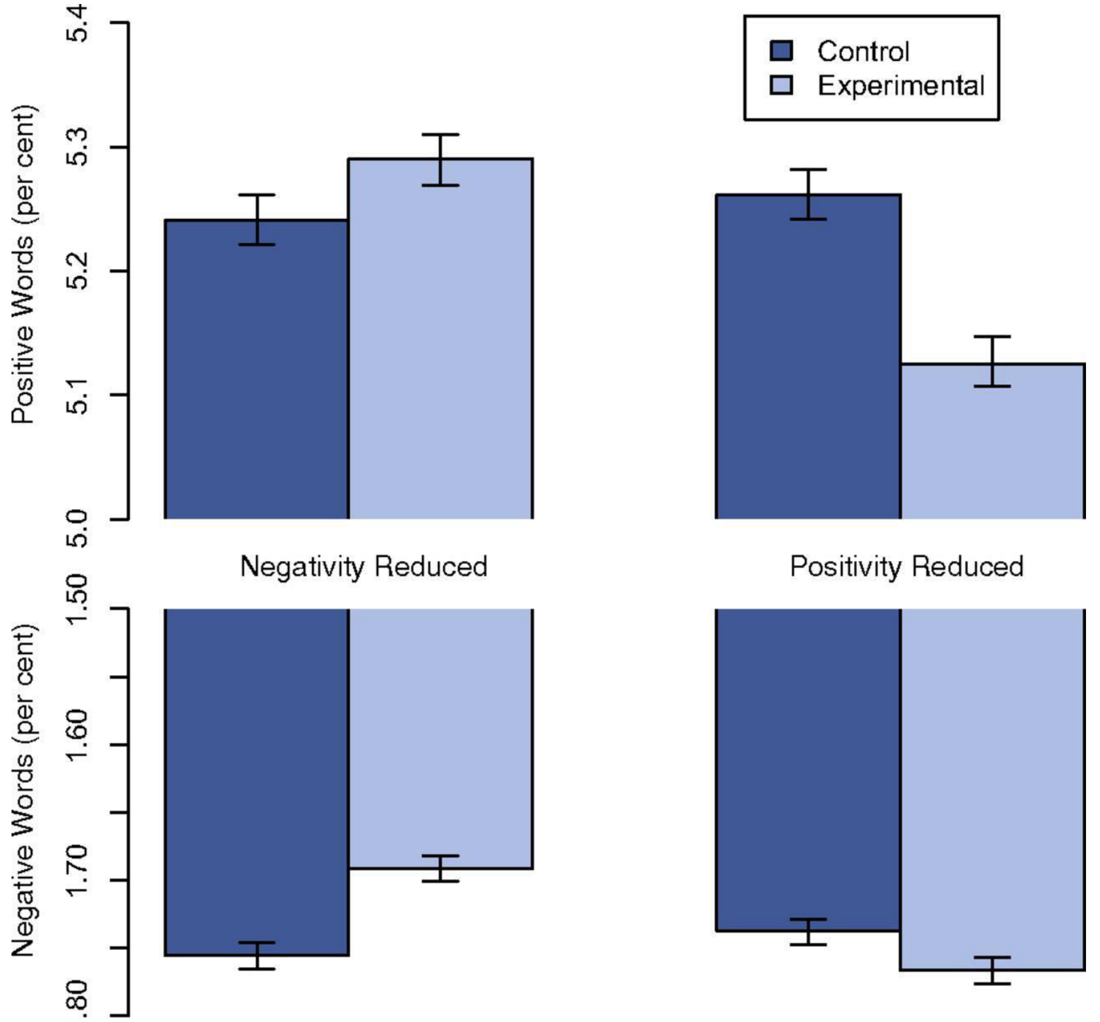


Individual characteristics
e.g. smoking, obesity
Reading profile ?

Experimental evidence of massive-scale emotional contagion through social networks (PNAS)

- Took place January 11-18, 2012, N= 689,003
- "One in which exposure to friends' positive emotional content in their News Feed was reduced, and one in which exposure to negative emotional content in their News Feed was reduced "
- And a "control condition, in which a similar proportion of posts in their News Feed were omitted entirely at random

Mean number of positive (Upper) and negative (Lower) emotion words (percent) generated people, by condition.



Kramer A D I et al. PNAS 2014;111:8788-8790

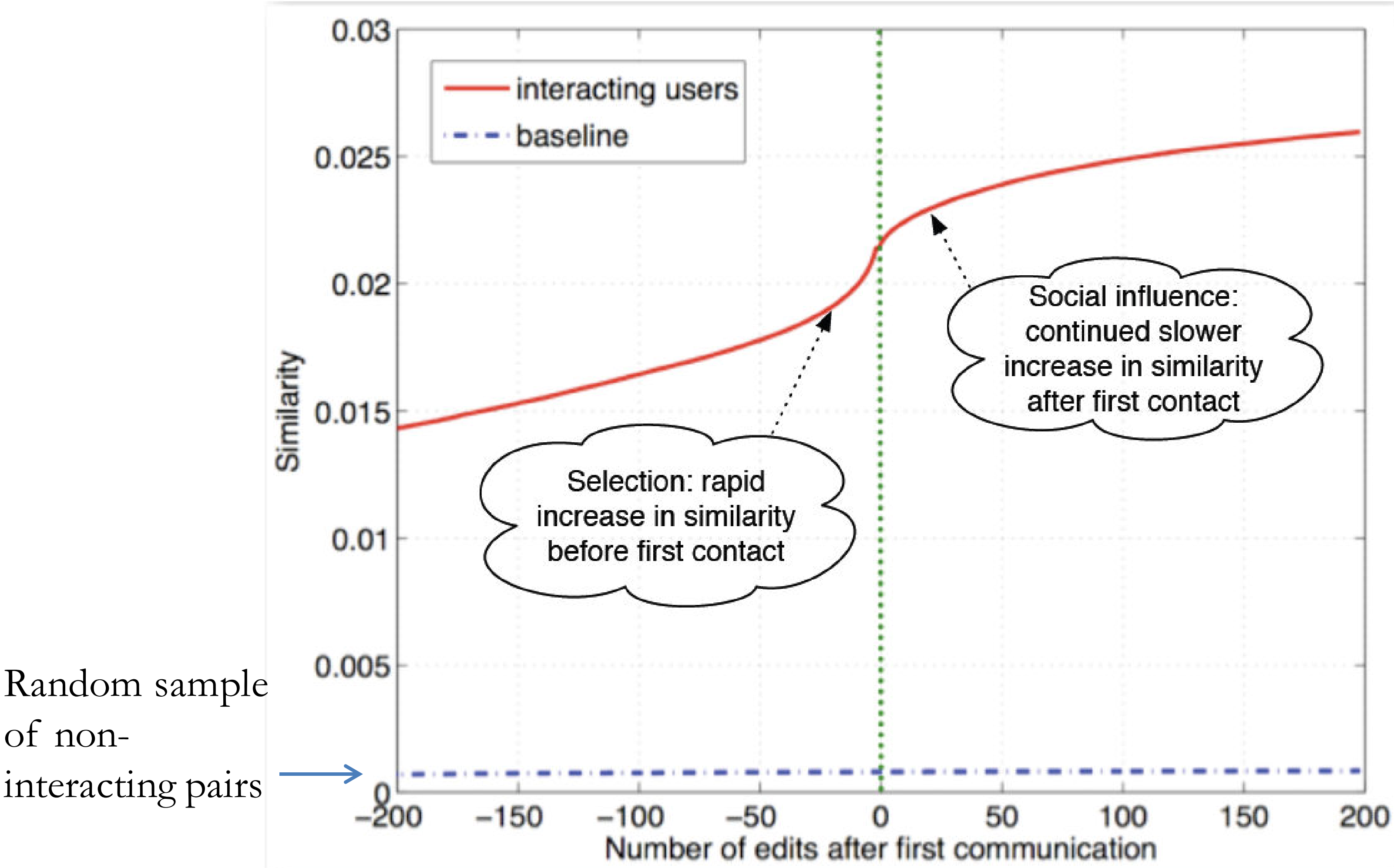


Figure 4.13: The average similarity of two editors on Wikipedia, relative to the time (0) at which they first communicated [122]. Time, on the x-axis, is measured in discrete units, where each unit corresponds to a single Wikipedia action taken by either of the two editors. The curve increases both before and after the first contact at time 0, indicating that both selection and social influence play a role; the increase in similarity is steepest just before time 0.

- Link creation and profile alignment in the aNobii social network

  Luca Maria Aiello, Alain Barrat, Ciro Cattuto,
  Giancarlo Ruffo and Rossano Schifanella
  IEEE SocialCom 2010

Outline

- Introduction of aNobii
- Static datasets analysis
  - Static properties of user's OSN and activity.
  - Geographic features.
  - Role of profile similarity
- Dynamic dataset analysis
  - How users create new relationship
    - Homophily vs. social influence
  - Network growth

aNobii

anobii
登錄 | 註冊帳號
Login
Together we find better books
書櫃
尋找
分享
書櫃
放上書櫃並展示您的藏書
尋找
藉由 aNobii 社群的協助，找尋您的下一本書籍
分享
評分並評論您的藏書，並在 Facebook 與 Twitter 上分享給大家
您現在正在閱讀哪本書？
搜尋
近期活動
哈寶寶在書架上新增了皇權禍國
14 分鐘前
const在書架上新增了視覺溝通的文法
14 分鐘前
Roxanne在書架上新增了從巴黎到巴塞隆納，慢慢走
14 分鐘前
Ariel4316 在書架上新增了談美感
14 分鐘前
voldemort在書架上新增了實用德語會話
15 分鐘前
青蛙王子在書架上新增了蒸業日記
15 分鐘前
1樓書櫃在書架上新增了龍眠
17 分鐘前
hsinyun在書架上新增了一個人的第一次
17 分鐘前
aNobii for iPhone
Enhanced Barcode Scanning
Access Expert Book Reviews
iPhone 4 Ready
aNobii 應用程式
iPhone
Android
掃描任一條碼來建立您的虛擬書櫃，並了解每個人對您藏書的看法
關於 | 翻譯 | 開發者 | Blog | 說明 | 條款 | 私隱 | 工作 | 聯絡
© 2012 aNobii
繁體中文

Social network for bookworms

- Data-driven analysis on anobii.com
  - Profile features
    - Library and wish list
    - Groups
    - Tags
  - Social network
    - Directed
    - Friendship + neighborhood

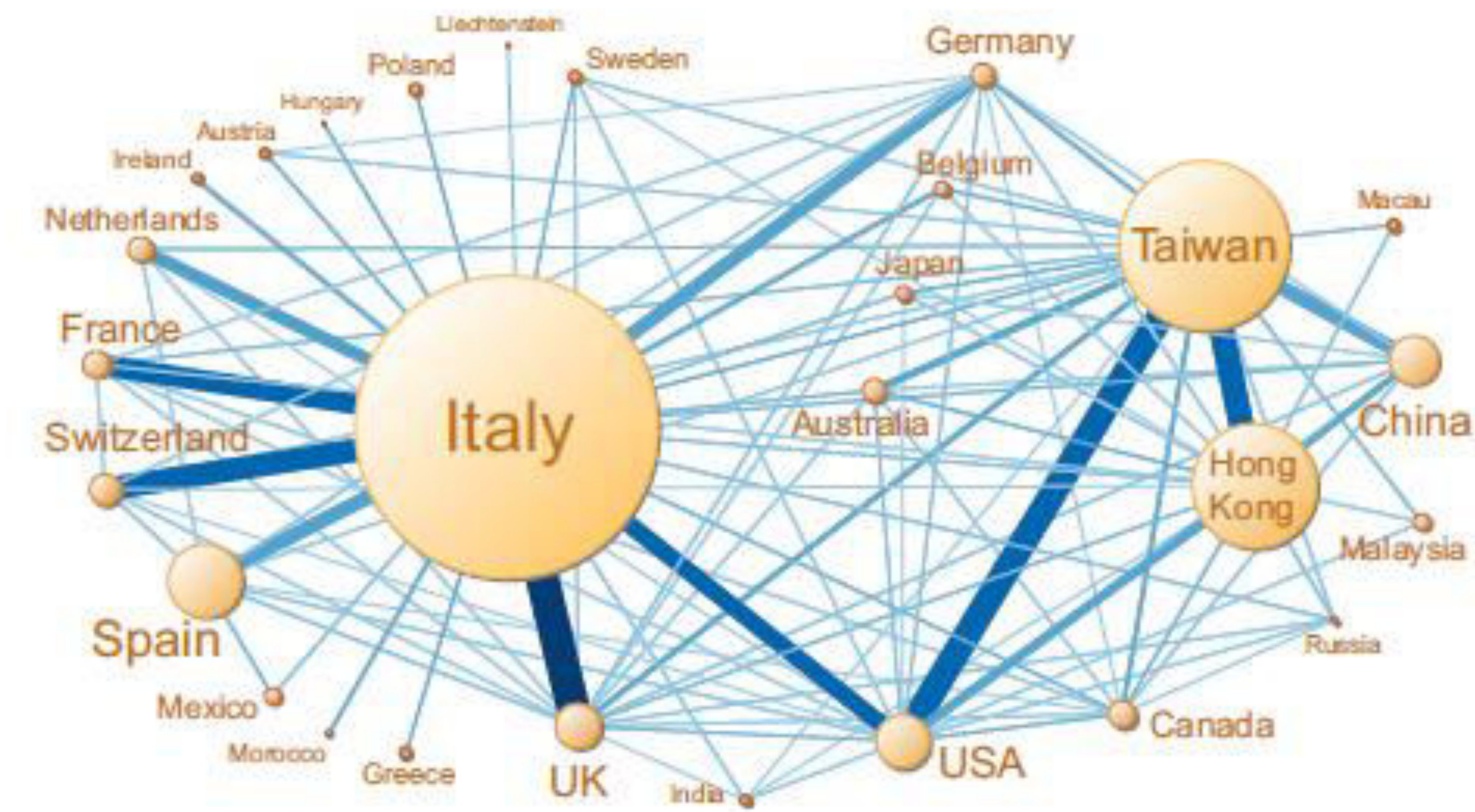
4th snapshot	Friendship	Neighborhood	Union
Nodes	74,908	54,590	86,800
Links	268,655	429,482	697,910

- 6 snapshots, 15 days apart
- *Full* giant connected component

Outline

- Introduction of aNobii
- Static datasets analysis
  - Static properties of user's OSN and activity.
  - Geographic features.
  - Role of profile similarity
- Dynamic dataset analysis
  - How users create new relationship
    - Homophily vs. social influence
  - Network growth

Geographical analysis



Dataset analysis

- Dataset crawling
  - From a random seed
  - Neighborhood network and friendship network
  - 6 snapshot starting from 11/09/2009
- Static properties
  - K-out degrees
  - SCC, WCC
  - Average shortest path length

Static Dataset analysis

	Friendship	Neighborhood	Union
Nodes	74,908	54,590	86,800
Links	268,655	429,482	697,910
Reciprocation	0.71	0.45	0.57
$< k_{out} >$	3.6	7.9	8.0
WCC size	68,624	54,246	86,800
SCC size	46,253	29,110	62,195
Density	$4.8 \cdot 10^{-5}$	$1.4 \cdot 10^{-4}$	$9.3 \cdot 10^{-5}$
Average SPL	7.3	4.7	5.3
Diameter	25	15	20
Degree centr.	0.0082	0.12	0.079

TABLE I
FRIENDSHIP, NEIGHBORHOOD AND FULL SOCIAL NETWORK STATISTICS (SPL=SHORTEST PATH LENGTH; WCC=WEAKLY CONNECTED COMPONENT; SCC=STRONGLY CONNECTED COMPONENT).

Static Dataset analysis (cont.)

- Diameter: The longest shortest path between two nodes in a network.
- Neighborhood network is denser and with a higher degree centralization
- Statistics reflects that the interest in other users' readings tends to concentrate toward a core.

Static Dataset analysis: power law distribution

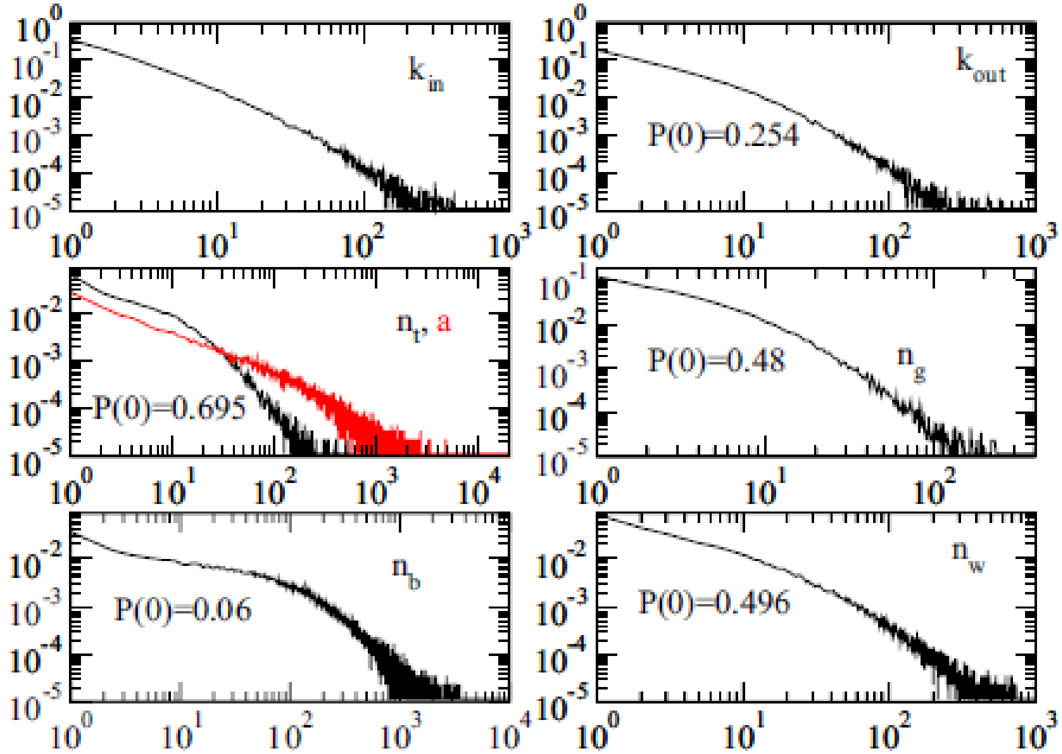
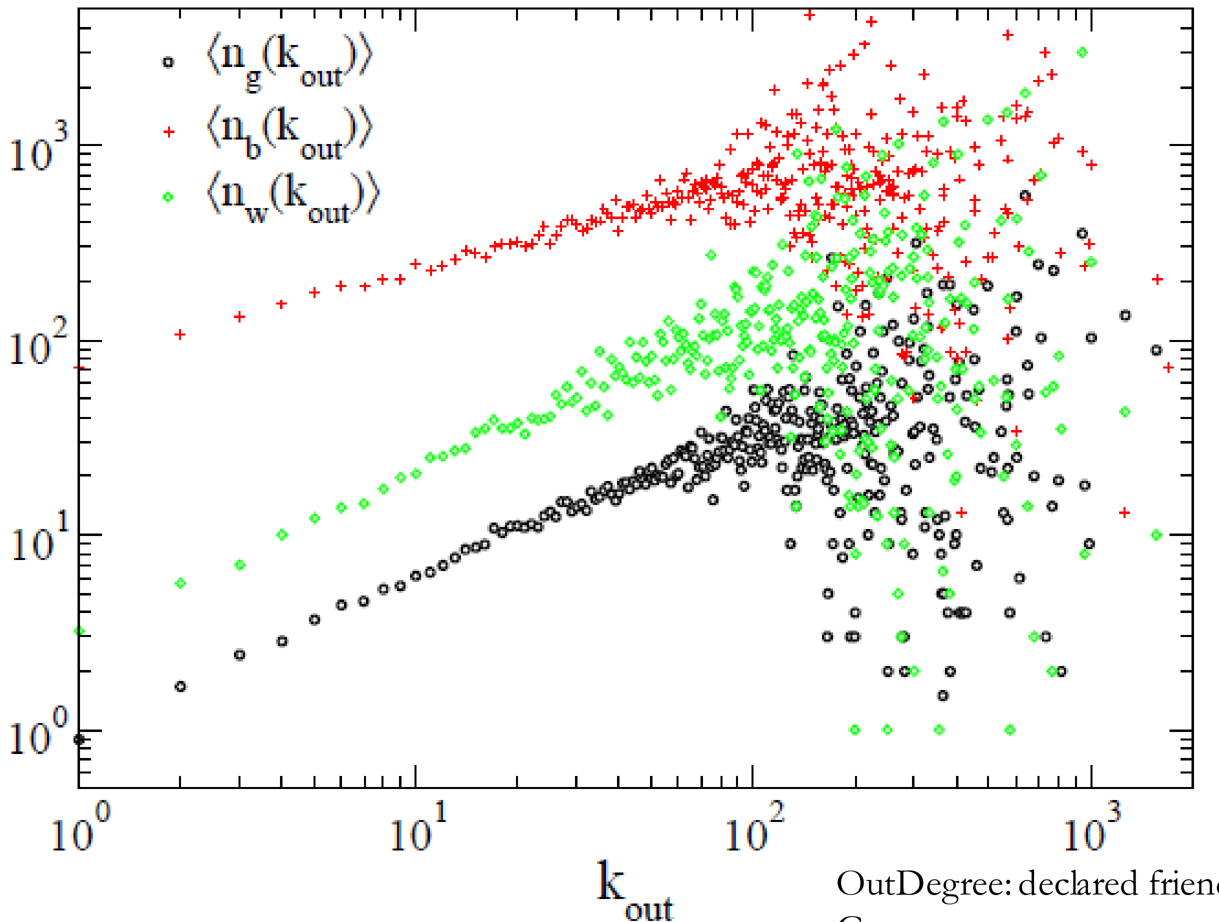


Fig. 1. Distributions of the measures of activity of aNobii users: in-degree k_{in} and out-degree k_{out} in the social network, number of distinct tags n_t and total tagging activity a (total number of tags in a user's page), number of group memberships n_g, number of books in a user library n_b and in a user wishlist n_w.

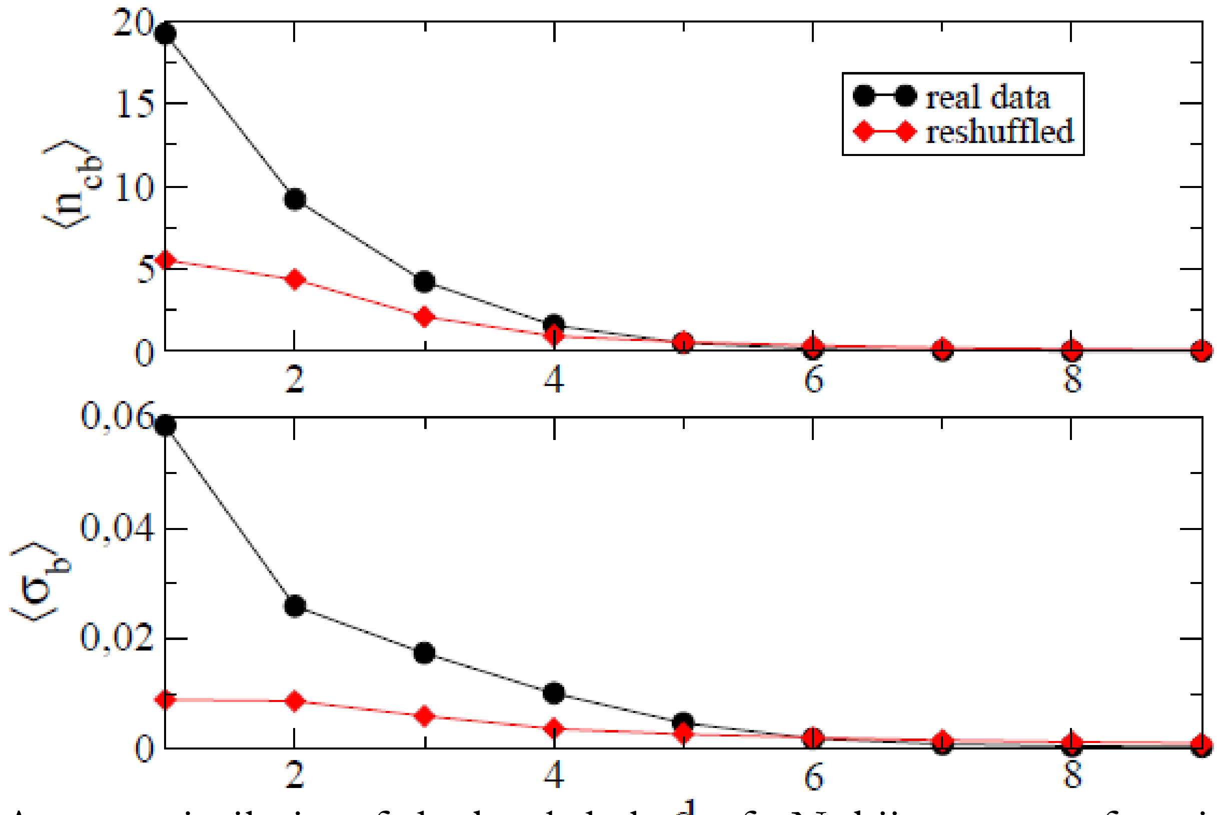
Correlations between one's outdegrees and other activities



OutDegree: declared friends and neighbors
G:groups
B:books in shelves
W:books in wish list

Profile similarity and social distance

Does similarity between user profiles depend on the social distance?



Average similarity of the bookshelves of aNobii users as a function of their Distance in the social network.

Up: Average number of shared books

Bottom: Cosine similarity of bookshelves

Reshuffled: randomization

$$\sigma_b(u,v) = \frac{\sum_b \delta_u(b)\delta_v(b)}{\sqrt{n_b(u)n_b(v)}}$$

Outline

- Introduction of aNobii
- Static datasets analysis
  - Static properties of user's OSN and activity.
  - Geographic features.
  - Role of profile similarity
- Dynamic dataset analysis
  - How users create new relationship
    - Homophily vs. social influence
  - Network growth

Dynamical analysis

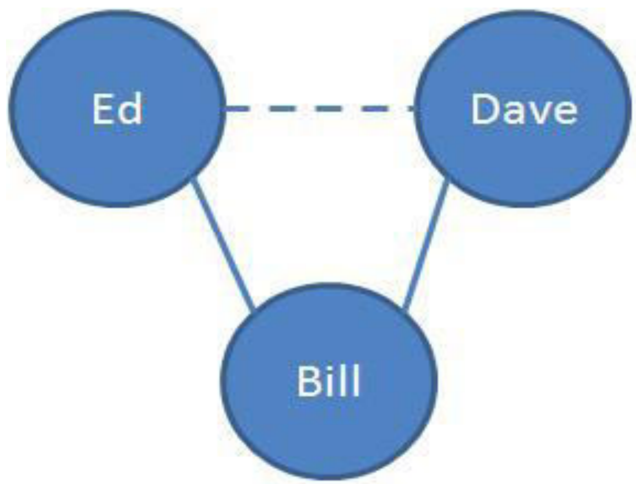
- Six snapshots in two and half months
- Dynamic analysis affords the study of
  - how people establish new relationship
  - network growth trend
  - new users' behavior

Phenomena studied

- Triad closure
- Preferential attachment
- Homophily
- Social contagion

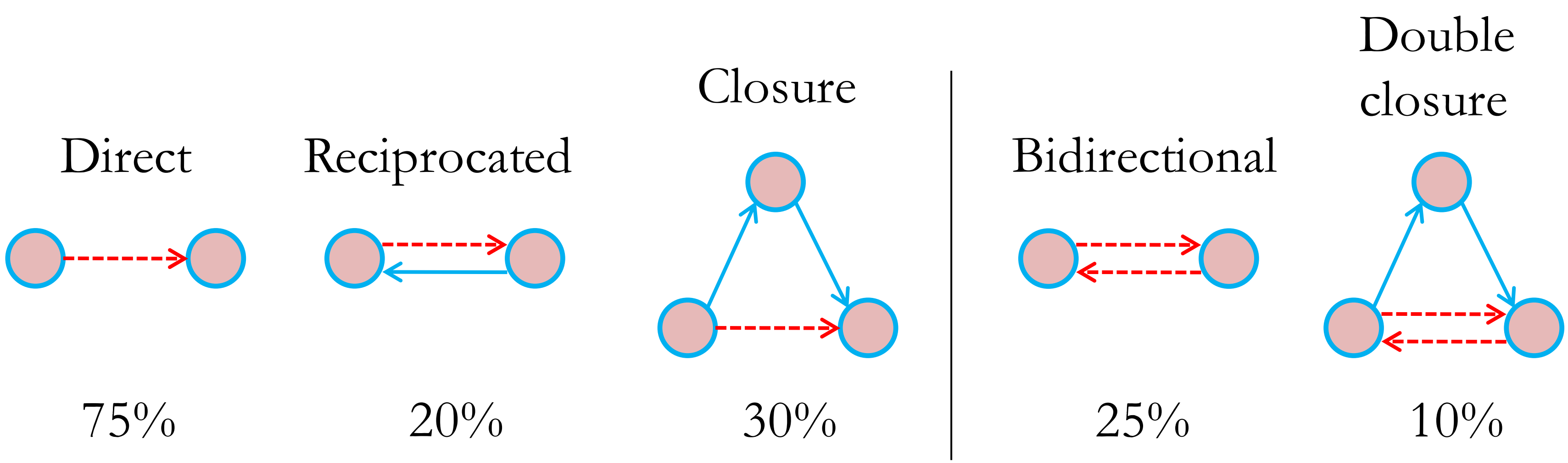
Triadic closure

- If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in future.



Triadic closure

- Classification of new links at time t+1 between nodes already present at time t (t ∈ {1,…,5})



- Reciprocation is strong
- Users tend to choose "friends of their friends" as new friends

Dynamical analysis: closure

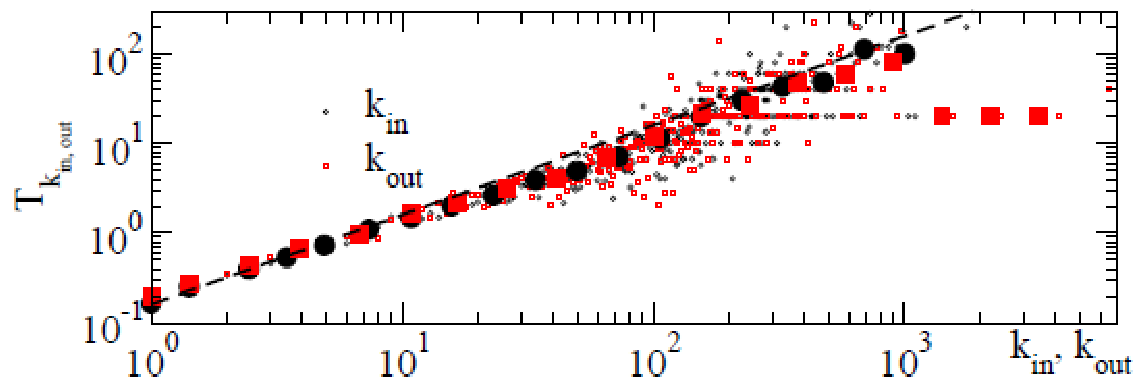
	$1 \to 2$	$2 \to 3$	$3 \to 4$	$4 \to 5$	$5 \to 6$
New nodes	2241	2121	1911	3214	3567
Removed nodes	239	222	230	220	684
New edges	19472	18324	17618	24805	26883
Died edges	642	763	713	782	700
$u \to v$	5409	4942	5259	6546	6357
Reciprocated	1016	1155	1285	1526	1688
$u \leftrightarrow v$	1809	1597	1604	1924	2235
Simple closure	2070	1976	2143	2497	2382
Double closure	955	904	877	1027	1141

TABLE III

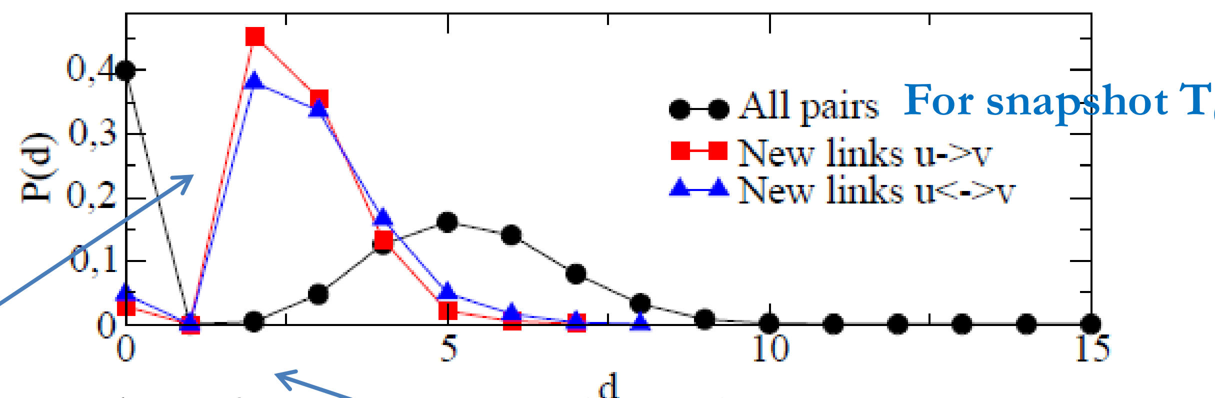
EVOLUTION OF SOME QUANTITIES FROM ONE SNAPSHOT TO THE NEXT.

Linkage bias: preferential attachment and homophily

Preferential attachment



Distance between newly created links between T4 and T5



Evidence for "selection" or "homophily"

Much shorter distance among linked pairs

Peaked at 2 degree,

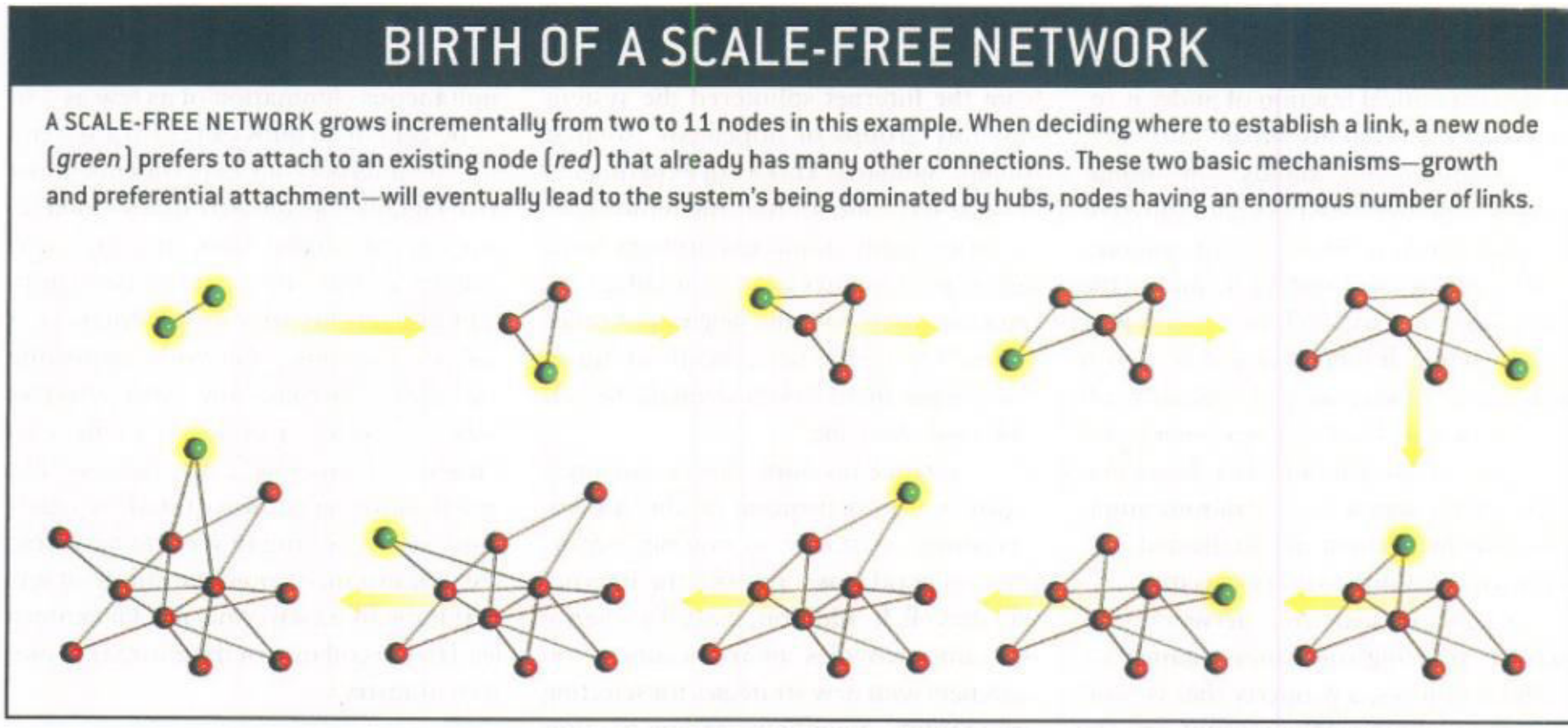
Preferential attachment

- that the more connected a node is, the more likely it is to receive new links. Nodes with higher degree have stronger ability to grab links added to the network
- New nodes are added to the network one at a time. Each new node is connected to m existing nodes with a probability that is proportional to the number of links that the existing nodes already have.
- Formally, the probability p_i that the new node is connected to node i is

$$p_i = \frac{k_i}{\sum_j k_j},$$

Preferential attachment

- Network dynamics for evolving self-organized networks



Rich gets richer

Similarity → link creation

		$\langle n_{cb} \rangle$	σ_b	$\langle n_{cg} \rangle$	σ_g
At T_0	$d_{uv} = 2$	9.5	0.02	1.12	0.05
From T_0 To T_1	$u \rightarrow v$	12.9	0.04	1.10	0.08
	$u \leftrightarrow v$	18.5	0.04	1.67	0.11
	Closure	18.2	0.04	1.81	0.10
	Dbl closure	23.4	0.05	1.20	0.12

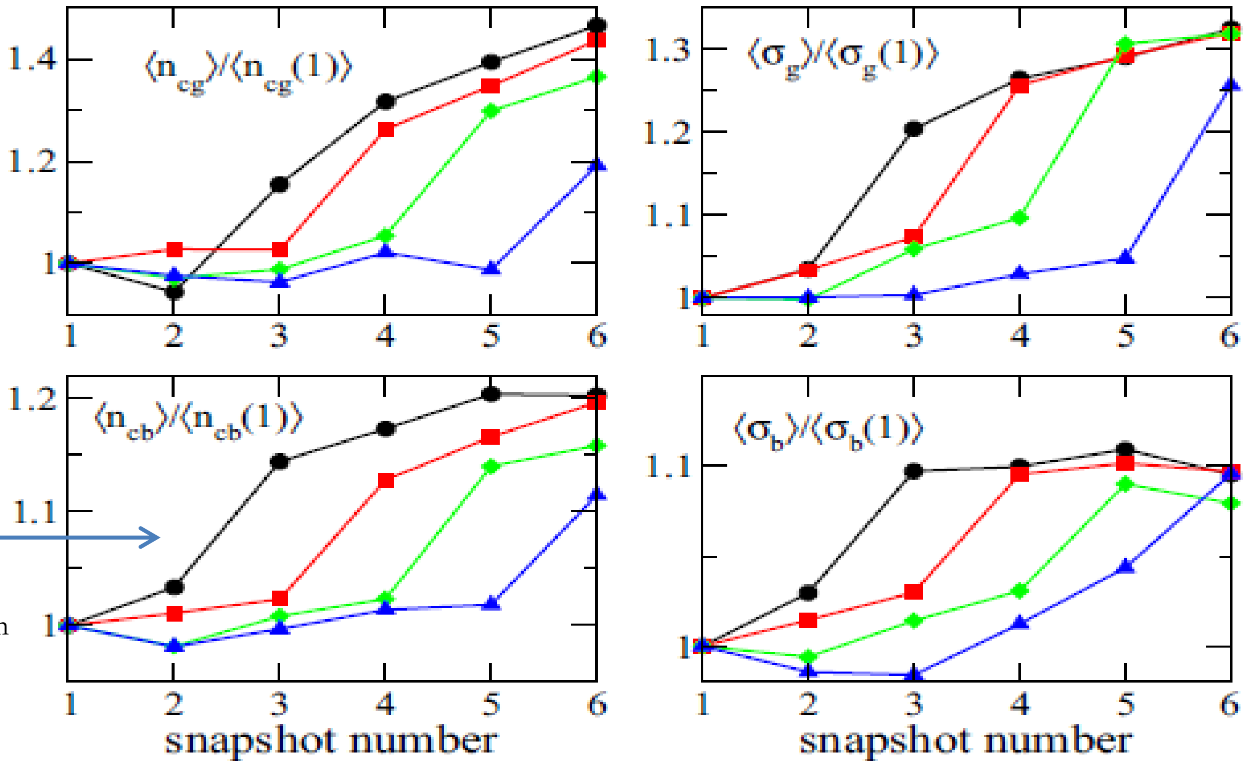
Average similarity of pairs forming new links between t_0 and t_0+1 (t_0=4), compared with average similarity of all the pairs at distance 2 at time t_0

Pairs that are going to get connected show a substantially higher similarity

Similarity of users partly drives the creation of new links

- "new links connect users who were already close, very often neighbors of neighbors; moreover, these users had more similar profiles than the average pairs of user at distance 2 "

Link creation leads to similarity



Links formed between T2 and T3

Fig. 10. Evolution of the average similarity of users' profiles, as measured by the numbers of common books or groups, and by the cosine similarities, from the first to the last snapshot, for links created between t_0 and $t_0 + 1$, for $t_0 = 2$ (black circles),3 (red squares), 4 (green diamonds), 5 (blue triangles), normalized by the average similarity in the first snapshot. Similarities are rather stationary before t_0, and clear jumps are observed between t_0 and $t_0 + 1$.

Social contagion

- "Before the creation of the links, the similarity is stationary; a large jump is observed when the links are created, and the similarity continues to grow, albeit at a slower page, after the link formation"
- "after users create links, they take inspiration from their new neighbors for new books to read and new group to join, and as a consequence align their profiles"

Aardvark and social search

- "a site which routes questions through an instant messaging bot to an appropriate user in one's extended network, comprised of friends-of-friends and strangers… prioritizes friends-of-friends for responses… When asked of the answers for feedback about the answer the user received**, 76% of the answers from within the user's extended network were rated 'good', and only 68% of answers from outside the users network were rated 'good'."**

Anyone flown Delta? Good? Bad?

March 13 at 9:17pm · Like · Comment

Pretty sure I've flown delta domestic in the US, from memory it was fine.

March 13 at 9:51pm · Like

There's about a hundred-dollar difference between them and United for three flights, two intl + one cross-country domestic...

March 13 at 9:57pm · Like

Delta is a thousand times better than united

March 13 at 9:58pm · Like

I agree, united are rubbish!

March 13 at 10:00pm · Like

Ok sweet, thanks guys! For Syd- USA, yeah?

March 13 at 10:05pm · Like

they're pretty much identical for domestic but i don't know anything about their international flights. did fly domestic or int'l? will we ever know? but yeah, generally speaking they're equivalent airlines.

also: consider the amazing AIR NEW ZEALAND. also the most important thing imo is to go to seatguru.com while you're booking your ticket so you can pick a good seat

March 13 at 10:07pm · Like

Yep, I'm all over seatguru (Expedia has it