Social network analysis 3 + 4

Petr Ocelík

ESS418 Research Methods in Social Science

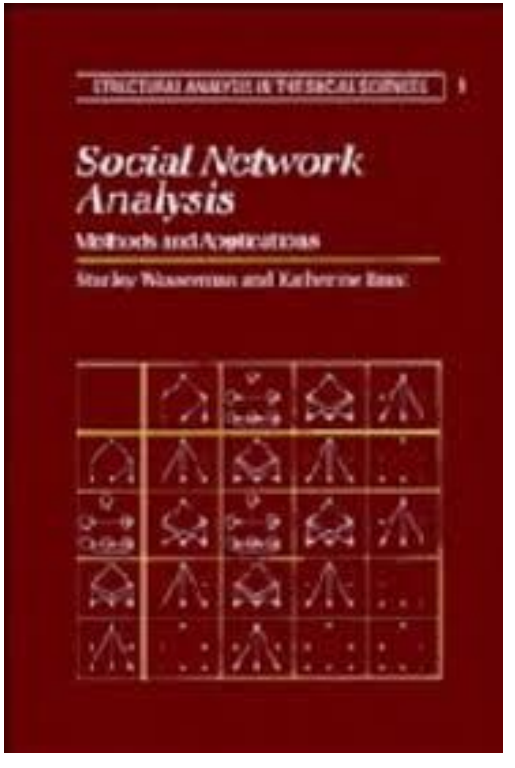
16th October 2015

Outline

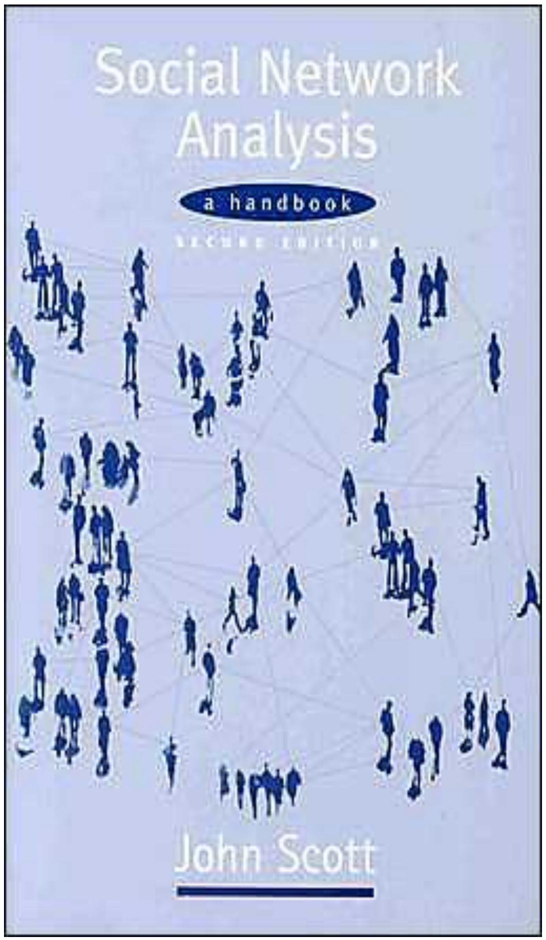
- Centrality and centralization
- Dyads and reciprocity
- Triads and transitivity
- Segments
- R: SNA mini-case

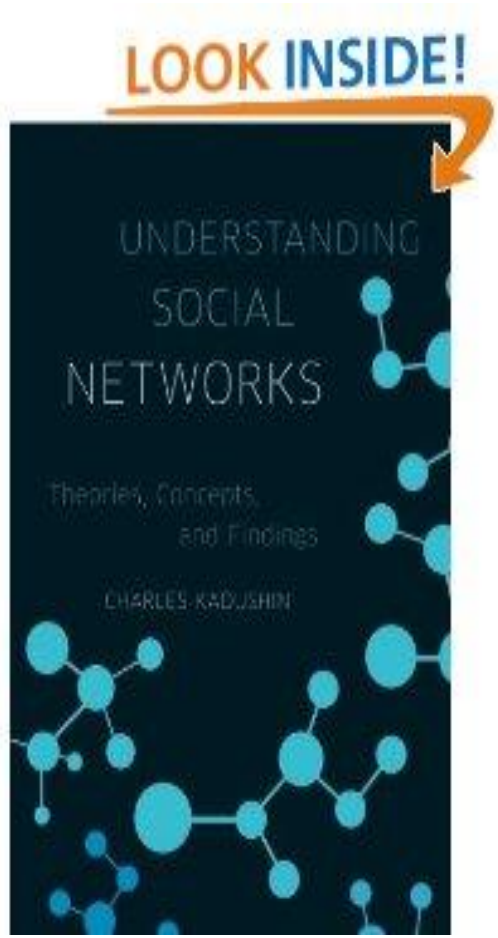
“Red bible”

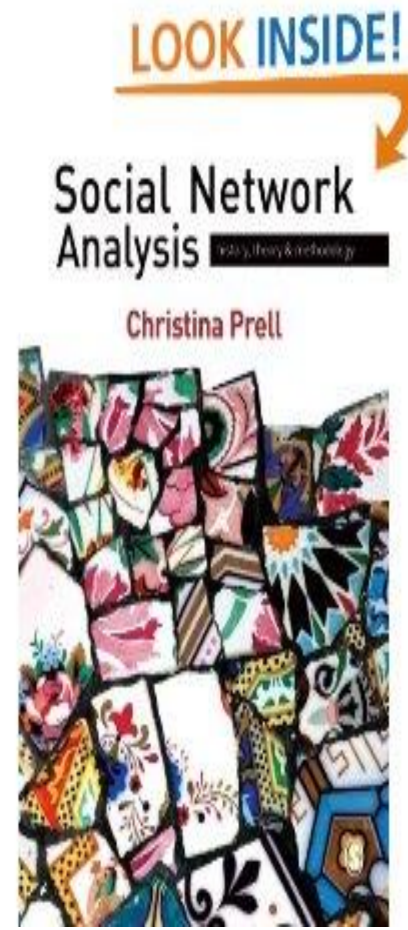
- Stanley Wasserman and Katherine Faust (1994): *Social Network Analysis: Methods and Applications*. Cambridge University Press.



Introductory sources







Graph theory

- **Network topology** is defined by two main concepts: connectivity and centrality.
- **Connectivity** describes interconnectedness of nodes in network (focus on **flows**).
- **Centrality** describes location of nodes in network (focus on **positions**).

Graph theory: notation

- G = graph/network
- N = # of nodes in network, n = individual node
- e = edge, g = geodesic
- i, j, … = indices (labels for selected elements)
- gij = geodesic between nodes i and, ni = node i
- k = # of selected elements (typically nodes)
- Cd'(G) = ' indicates that the measured value is standardized
- Upper case: global measures
- Lower case: local measures
- cd(ni) = node i degree centrality
- Cd(G) = graph G degree centralization

Centrality

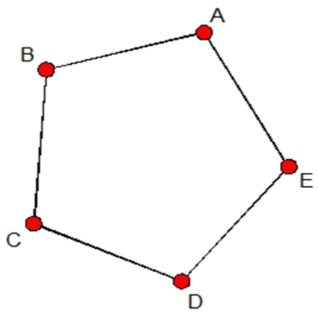
- **Local measure:** characterizes position of a particular node within a network.
- Different measures for different network classes! (e.g. bipartite networks, directed or weighted ties).
- Simplest case: undirected binary network.
- Different types of centralities:
  - Degree centrality
  - Closeness centrality
  - Betweenness centrality

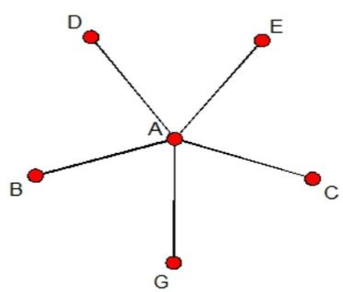
Centralization

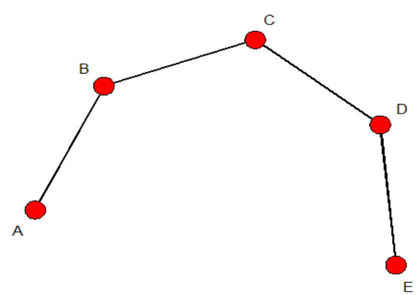
- **Global measure:** characterizes extent to which a certain local feature prevails in a network.
- **Centralization** is measured as a difference of all centrality values from the highest centrality value.
- Again, we need to consider the network class.
- Analogically, there are different types of centralizations.
- **centralization =/= centrality**

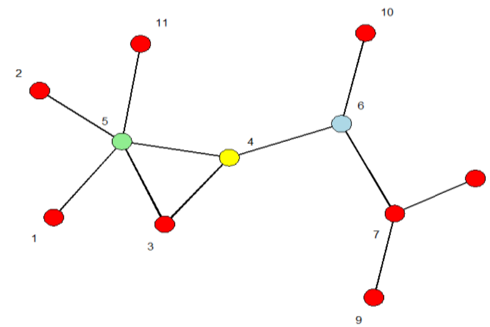
Degree centrality

- Nodes with most connections.
- **Theoretical importance:**
  - Most powerful actors in network.
  - Indication of prestige (in-degree centrality).
  - Indication of influence (out-degree centrality).
  - *Depends on tie conceptualization*.
  - *Limited only to adjacent nodes (neighborhood)*!









Degree centrality

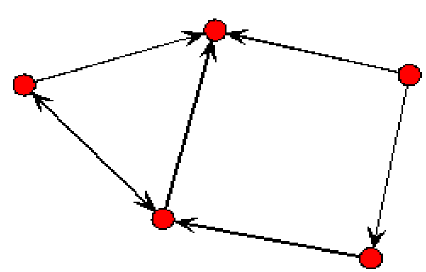
- **Undirected graph:** # of connections of a node.

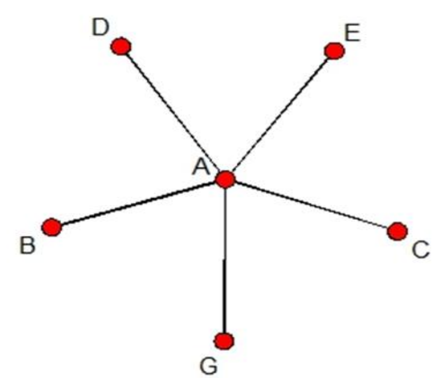
$$c_D(n_i) = d(n_i)$$

- **Standardization:** division by # of all possible connections (# of all nodes - 1).

$$c_D'(n_i) = d(n_i) / (N - 1)$$

- **Directed graph:** # of connections from/to a node.
  - In-degree centrality
  - Out-degree centrality





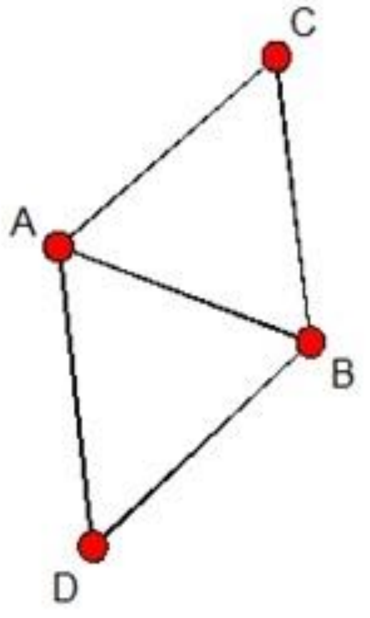
Degree centrality

$c_D(n_i) = d(n_i)$

$c_D(n_A) = d(n_A) = 3$

$c_D'(n_i) = d(n_i) / (N - 1)$

$c_D'(n_A) = 3 / 3 = 1$



Degree centralization

1. # of connections of each node (degree centrality).
2. Differences of centralities from the highest centrality value.
3. Summation of differences.

$$C_D(G) = \sum (c_D(n_{MAX}) - c_D(n_i))$$

4. Standardization: division by (N – 1) * (N- 2)

$$C_D'(G) = \sum (c_D(n_{MAX}) - c_D(n_i)) / (N - 1) * (N - 2)$$

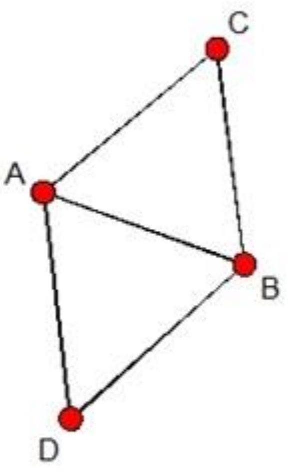
Degree centralization

$C_D(G) = \sum (c_D(n_{MAX}) - c_D(n_i))$

$C_D(G) = (0 + 0 + 1 + 1) = 2$

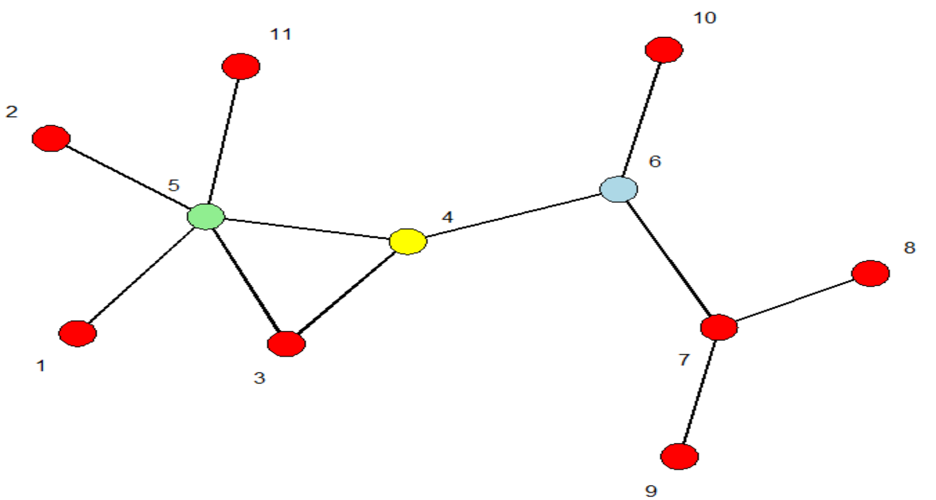
$C_D'(G) = \sum (c_D(n_{MAX}) - c_D(n_i)) / (N - 1) * (N - 2)$

$C_D'(G) = 2 / (4 - 1) * (4 - 2) = 2 / 6 = 0.33$



Closeness centrality

- Nodes with shortest average path length.
- Node is closer to more nodes than any other node.
- **Theoretical importance:**
  - Shorter path lengths = quicker and more efficient access to sources.



Closeness centrality

- **Distance:** ∑ geodesics of a node to all other nodes.
- **Closeness:** inverse concept → 1 / distance:

$$c_C(n_i) = 1 / \sum g(n_i, n_j) \;\; \mathbf{=} \;\; c_C(n_i) = [\, \sum g(n_i, n_j) \,]^{-1}$$

- Standardization: multiplication by # of nodes - 1.

$$c_C'(n_i) = [\, \sum g(n_i, n_j) * (N - 1) \,]^{-1}$$

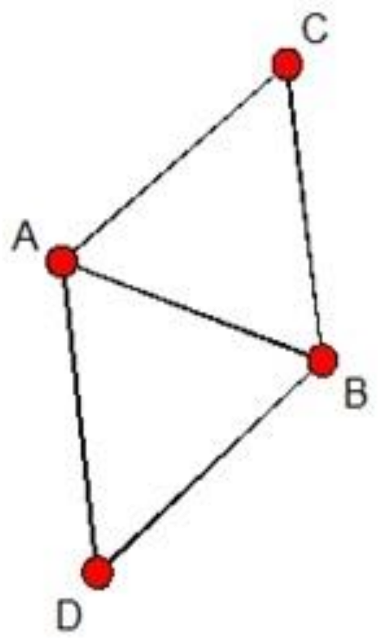
Closeness centrality

$c_C(n_i) = [\ \sum g(n_i, n_j)\ ]^{-1}$

$c_C(n_A) = [\ (1 + 1 + 1)\ ]^{-1} = 0.33$

$c_C'(n_i) = [\ \sum g(n_i, n_j) * (N - 1)\ ]^{-1}$

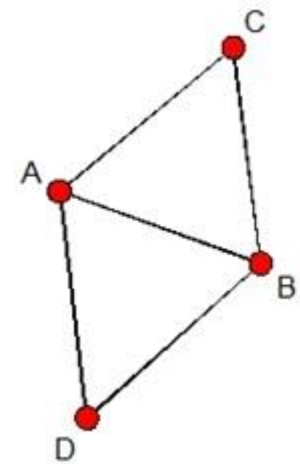
$c_C'(n_A) = 0.33 * (4 - 1) = 1$



Closeness centralization

1. **Standardized** closeness centrality values for each node.
2. Differences of centralities from the highest centrality value.
3. Summation of differences.

$$C_C'(G) = \sum (c_C'(n_{MAX}) - c_C'(n_i))$$



4. Standardization: (N – 1) * (N- 2) / (2N – 3)

$$C_C'(G) = \sum (c_C'(n_{MAX}) - c_C'(n_i)) / [(N-1) * (N-2) / (2N-3)]$$

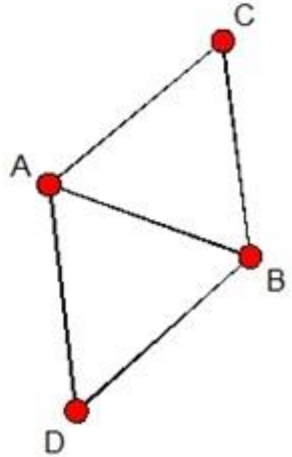
Closeness centralization

$C_C'(G) = \sum (c_C'(n_{MAX}) - c_C'(n_i))$

$C_C'(G) = \sum (c_C'(n_{MAX}) - c_C'(n_i)) / [(N - 1) * (N - 2) / (2N - 3)]$

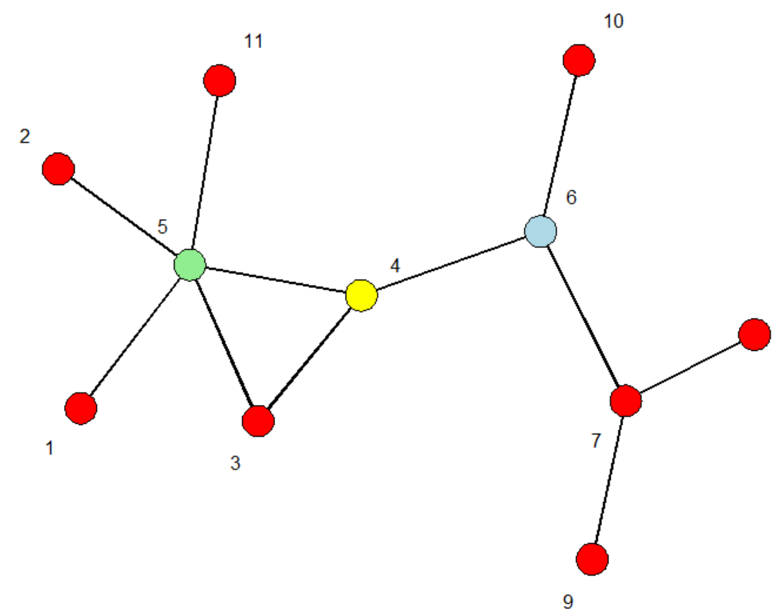
$C_C'(G) = \sum (1 - 1) + (1 - 1) + (1 - 0.75) + (1 - 0.75) / [(4 - 1) * (4 - 2) / (8 - 3)]$

$C_C'(G) = 0.5 / (12 / 5) = 0.5 / 1.2 = 0.42$



Betweenness centrality

- Nodes which are most in-between other nodes.
- **Theoretical importance:**
  - Crucial for flow control (gatekeepers, brokers).
  - Bridges to otherwise weakly connected parts of network (access to sources).



Betweenness centrality

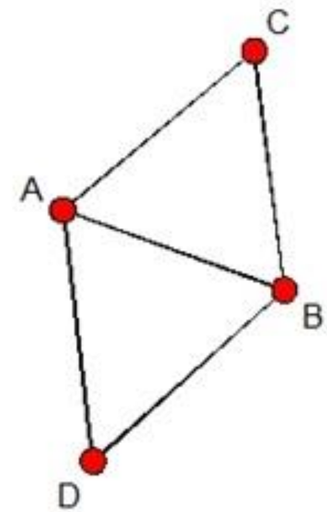
- **Betweenness:** ratio of geodesics upon which a node lies to all geodesics between remaining nodes in network.

$$c_B(n_i) = \sum g_{jk}(n_i) / g_{jk}$$

- Standardization: division by # of all possible geodesics upon which node can lie.

$$c_B'(n_i) = [\sum g_{jk}(n_i) / g_{jk}] / [(N - 1) * (N - 2) / 2]$$

Betweenness centrality



$c_B(n_i) = \sum g_{jk}(n_i) / g_{jk}$

$c_B(n_A) = 0.5 / 3 = 0.17$

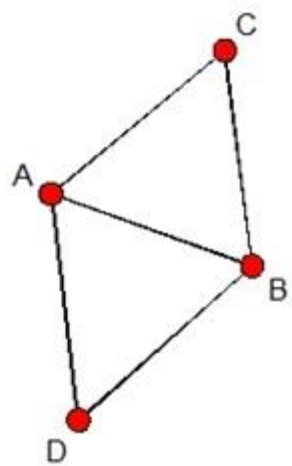
$c_B'(n_i) = [\sum g_{jk}(n_i) / g_{jk}] / [(N - 1) * (N - 2) / 2]$

$c_B'(n_A) = 0.17 / [((4 - 1) * (4 - 2)) / 2] = 0.17 / 3 = 0.06$

Betweenness centralization

1. Betweenness centrality values for each node.
2. Differences of centralities from the highest centrality value.
3. Summation of differences.

$$C_B'(G) = \sum (c_B(n_{MAX}) - c_B(n_i))$$



4. Standardization: (N – 1)^2 * (N- 2) / 2

$$C_B'(G) = \sum (c_B(n_{MAX}) - c_B(n_i)) \,/\, [\,(N-1)^2 * (N-2) / 2\,]$$

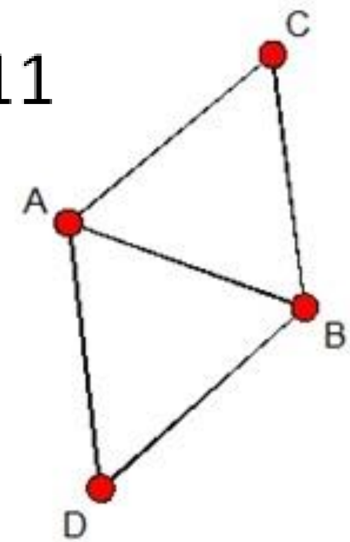
Betweenness centralization

$C_B{'}(G) = \sum (c_B(n_{MAX}) - c_B(n_i))$

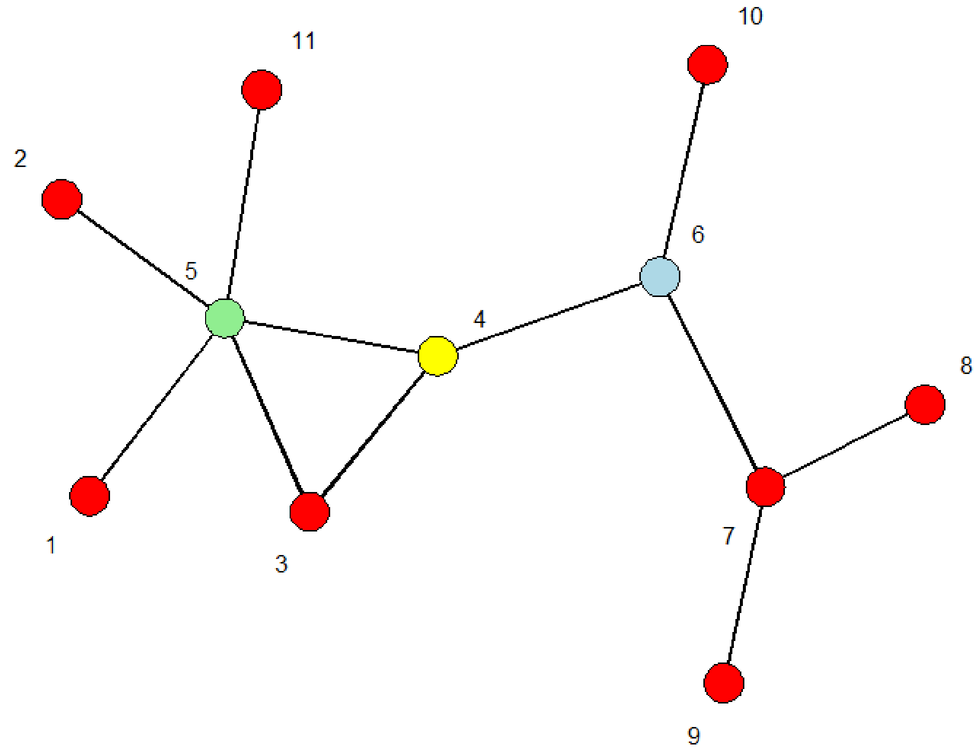
$C_B(G) = \sum (0.5 - 0.5) + (0.5 - 0.5) + (0.5 - 0) + (0.5 - 0) = 1$

$C_B{'}(G) = \sum (c_B(n_{MAX}) - c_B(n_i)) / [(N - 1)^2 * (N - 2) / 2]$

$C_B(G){'} = 1 / [(4 - 1)^2 * (4 - 2) / 2] = 1 / 9 = 0.11$



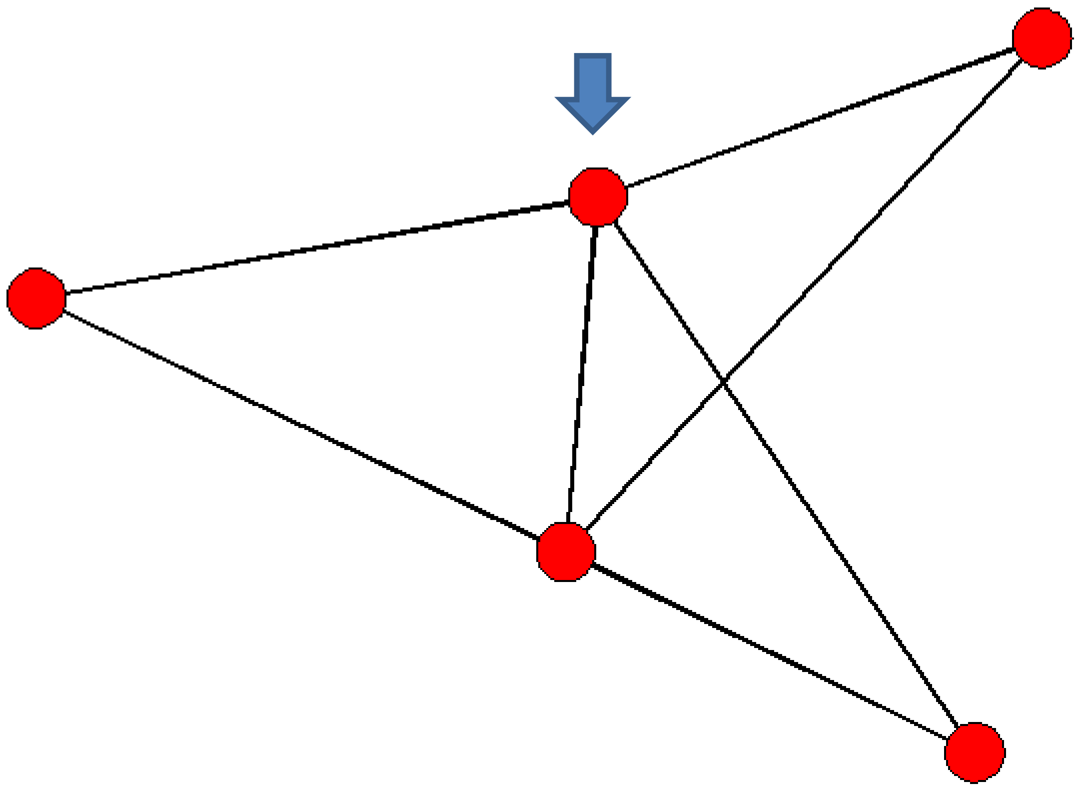
Degree vs. closeness vs. betweenness



	1	2	3	4	5	6	7	8	9	10	11
degree	1	1	2	3	5	3	3	1	1	1	1
closeness	0.33	0.33	0.42	0.53	0.48	0.5	0.4	0.29	0.29	0.34	0.33
betweenness	0	0	0	50	48	54	34	0	0	0	0

Exercise

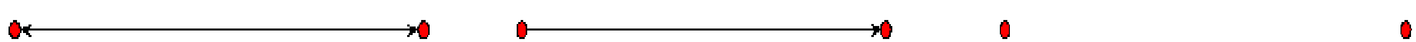
- Calculate degree and closeness centrality of a given node. What is degree centralization?



Dyads

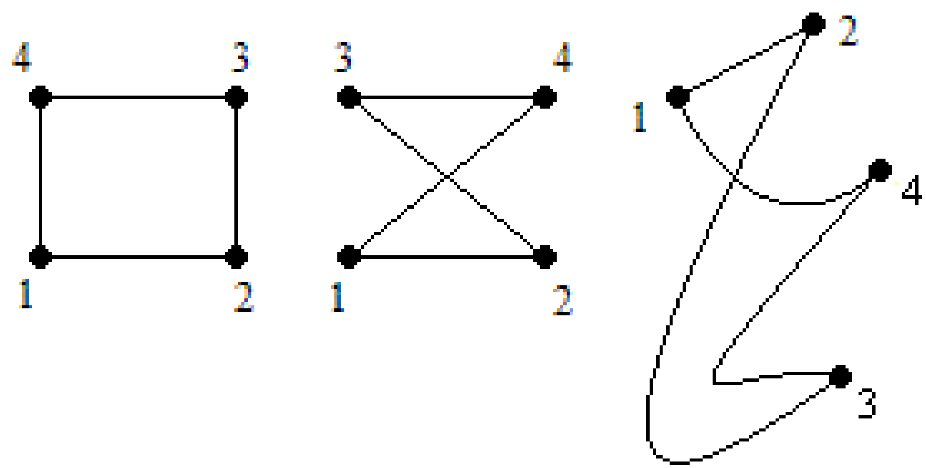
- Dyad: a most basic relational unit.
- 2 dyads for undirected graphs:

- **Isomorphism:** structural interchangeability of nodes, edges or their configurations.
- 3 isomorphic dyads for directed graphs.
- **MAN: M:** mutual, **A:** asymmetric, **N:** null:



Isomorphism

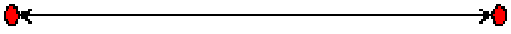
- Isomorphic representations of graphs preserve adjacency.



stackexchange.com

Reciprocity (Wasserman & Faust 2009)

- **Reciprocity:** there is a bidirectional (mutual) tie between two nodes.

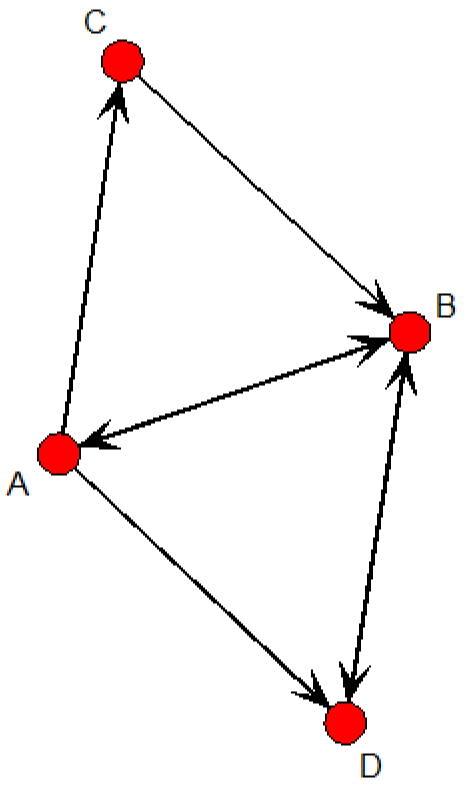


- Reciprocity in graph is given as a ratio of reciprocal (**m**utual) dyads to total number of connections (mutual + **a**symmetrical dyads).

$$R(G) = 2M / (2M + A)$$

Reciprocity

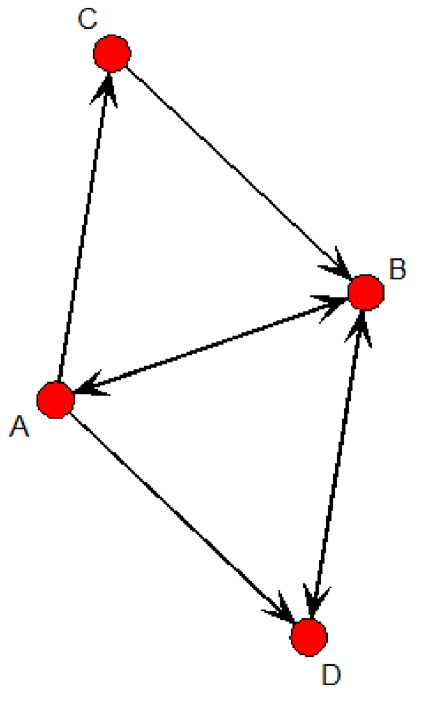
R(G) = 2M / (2M + A)



Reciprocity

R(G) = 2M / (2M + A)

R(G) = 2*2 / (2*2 + 3) = 4 / 7 = 0.57



- R default

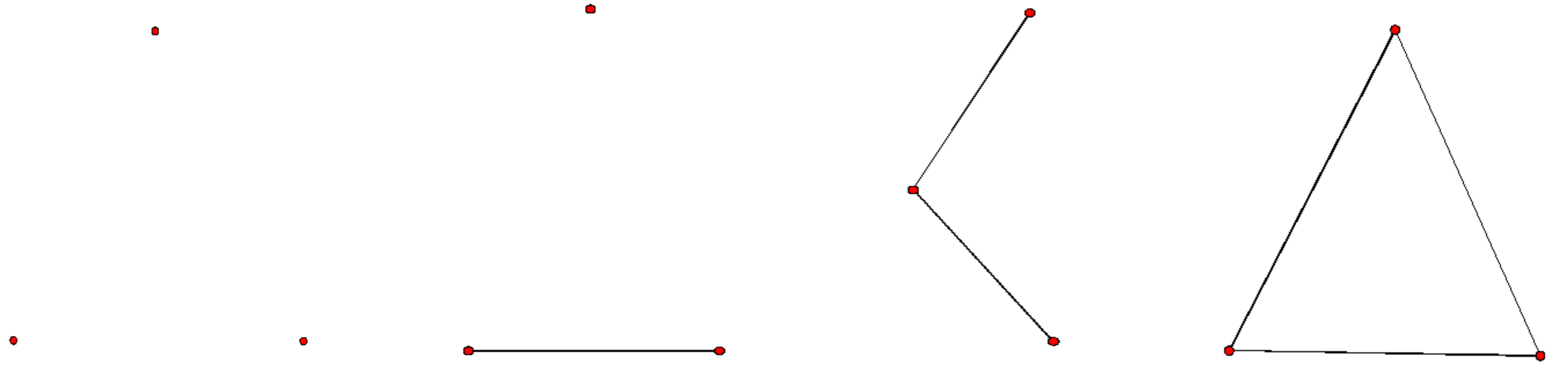
R(G) = M / (M + A + N) = (1 + 1) / ((1 + 1) + (1 + 1 + 1) + 1)) = 2 / 6 = 0.33

- R dyadic.nonnul

R(G) = M / (M + A) = (1 + 1) / ((1 + 1) + (1 + 1 + 1)) = 2 / 5 = 0.4

Triplets and triads in undirected graph

- Triad as **a basic unit of social organization**.
- triplet = empty triad
- In undirected graph: 2^3 = 8 combinations of triads with preserved identity (anisomorphic).
- 4 isomorphic triads:



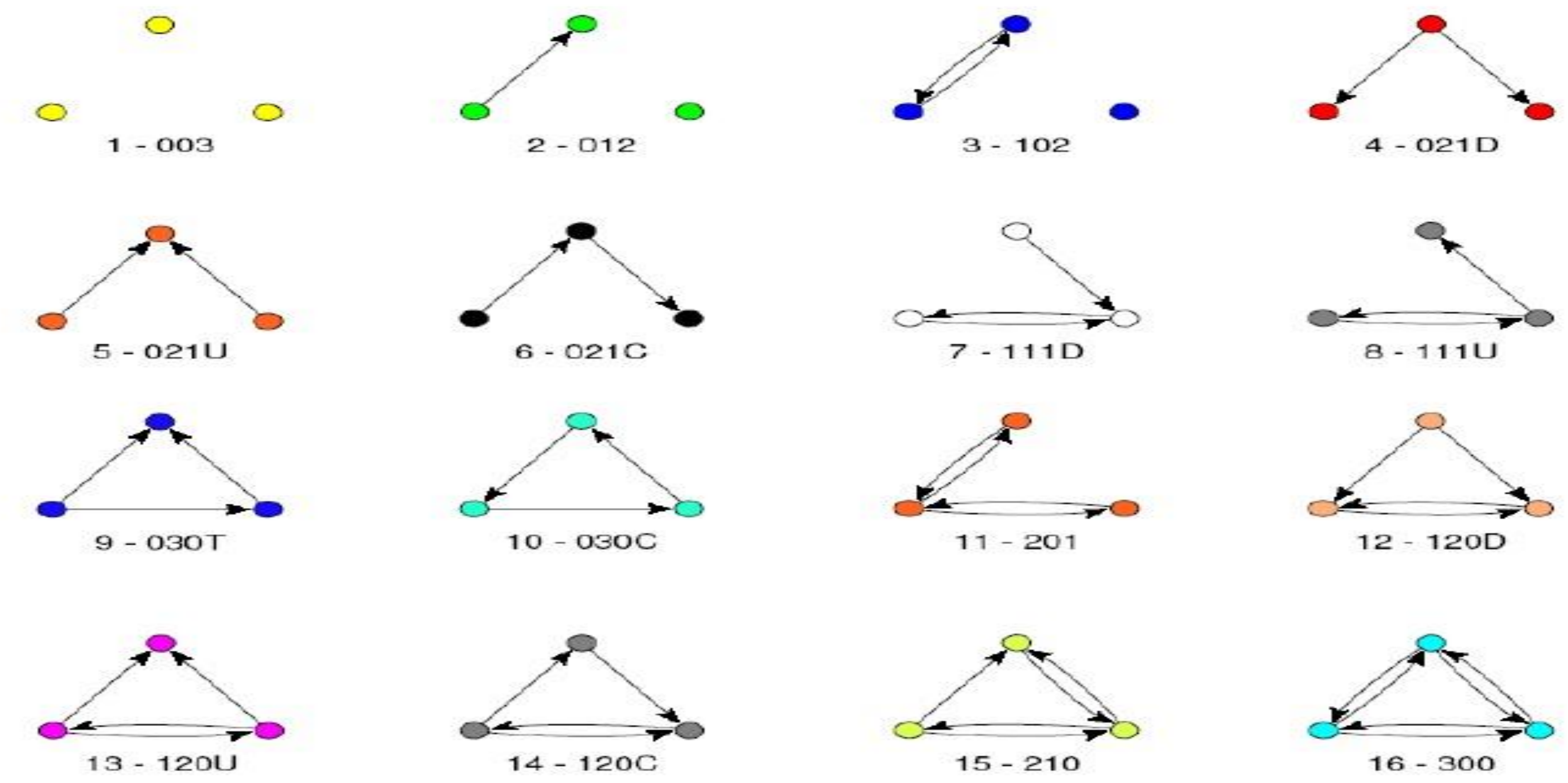
Triplets and triads in directed graph

- There is a 6 combinations of triplets: (a, b, c); (a, c, b); (b, a, c); (b, c, a); (c, a, b); (c, b, a).
- Thus: 2^6 = 64 combinations of anisomorphic triads.
- If we neglect node identities, we get 2^4 = **16 isomorphic triads = triadic census**.
- # of isomorphic triads in directed graph is given by:

$$\sum T_u = N! / (k! * (N - k)!)$$

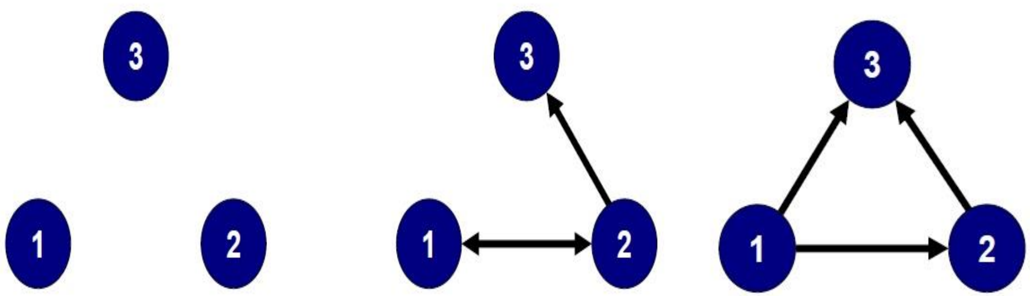
Triadic census

- MAN (**M**utual-**A**symmetric-**N**ull), tie direction (**U**p/**D**own), transitivity (**T**) vs. cycle (**C**)

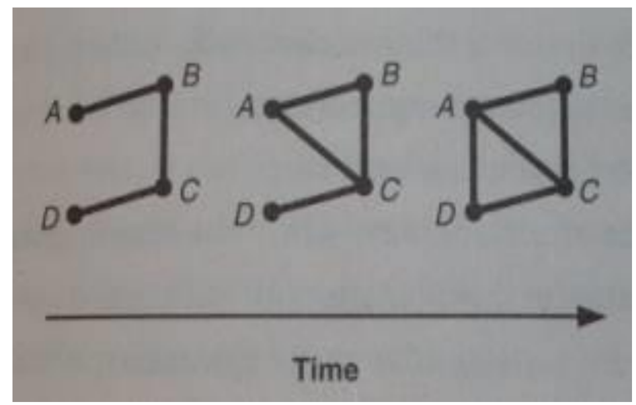


Transitivity

- Transitivity: *a friend of my friend is my friend*.
- Triadic closure:



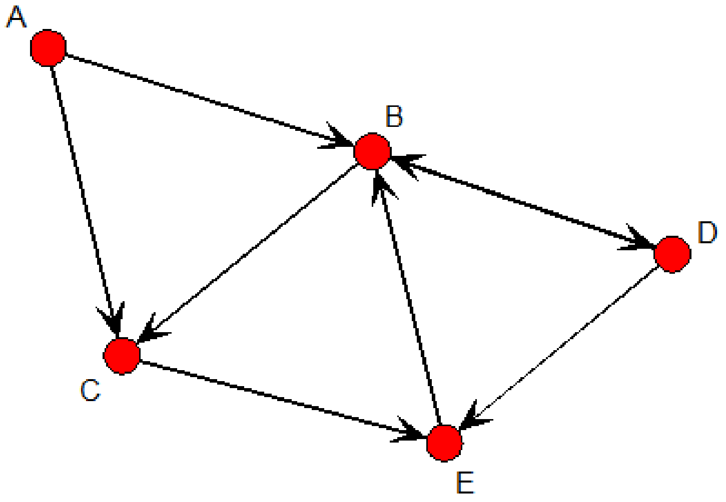
Thurner 2012



- **Measure:** # of transitive triads over # of all triads. $T(G) = \sum_{ijk} e_{ij}, e_{jk}, e_{ik} / (N! / (k! * (N - k)!))$

Transitivity

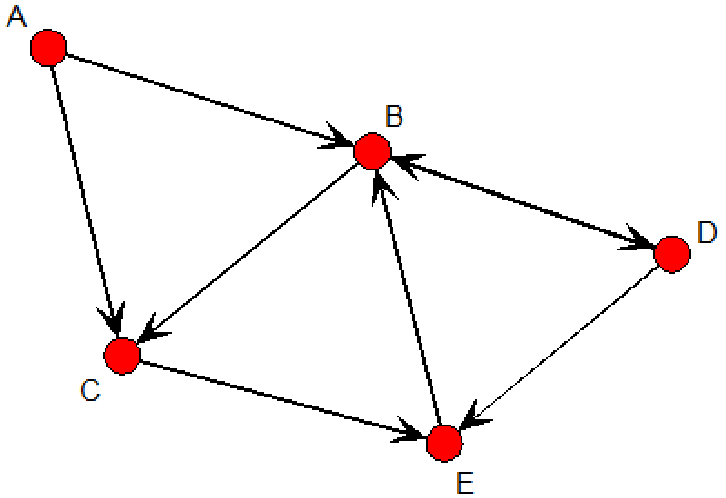
$$T(G) = \sum_{ijk} e_{ij}, e_{jk,} e_{ik} / (N! / (k! * (N - k)!))$$



Transitivity

$$T(G) = \sum_{ijk} e_{ij}, e_{jk,} e_{ik} / (N! / (k! * (N - k)!))$$

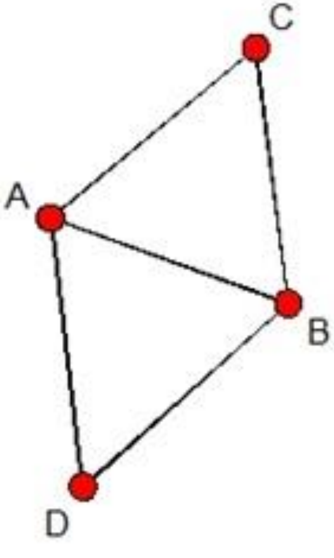
$$T(G) = 1 / (120 / 6 * 2) = 1 / 10 = 0.1$$



Local clustering coefficient

- Measures level of a given node's neighborhood's interconnectedness.
- Measure: # of interconnections of adjacent nodes over # of all possible interconnections of adjacent nodes.

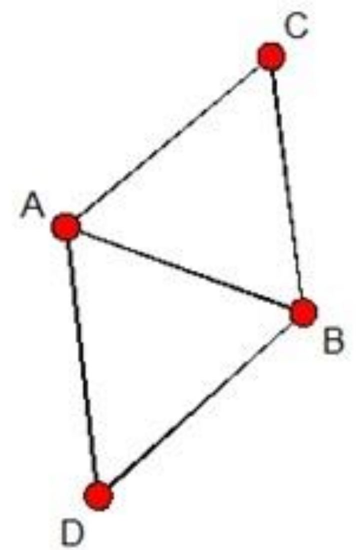
$$cc_i = e_{jk} / (n_{jk} * (n_{jk} - 1) / 2)$$



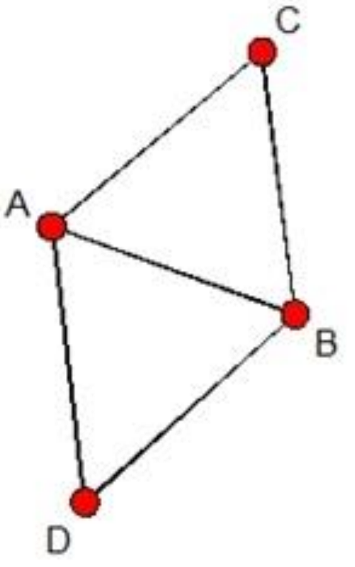
Local clustering coefficient

$cc_i = e_{jk} / (n_{jk} * (n_{jk} - 1) / 2)$

$cc_A = 2 / ((3 * 2)/2) = 2 / 3 = 0.67$



Clustering coefficient



- Global measures:
- (1) **global cluster coefficient (GCC):**
  - # of closed triads / # all connected triads
- (2) **average cluster coefficient (ACC):**
  - Arithmetic mean of local cluster coefficient values.

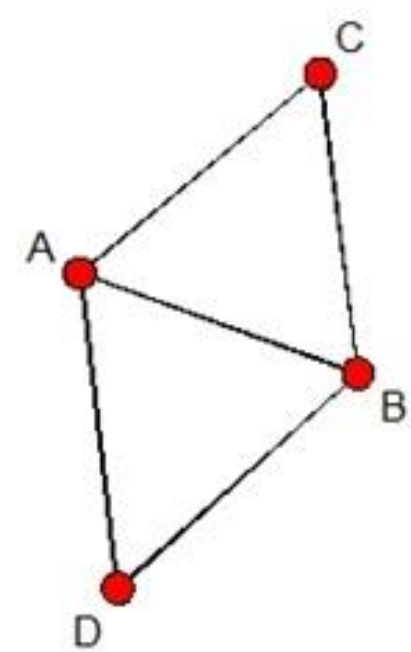
$$ACC(G) = \sum cc_i / N$$

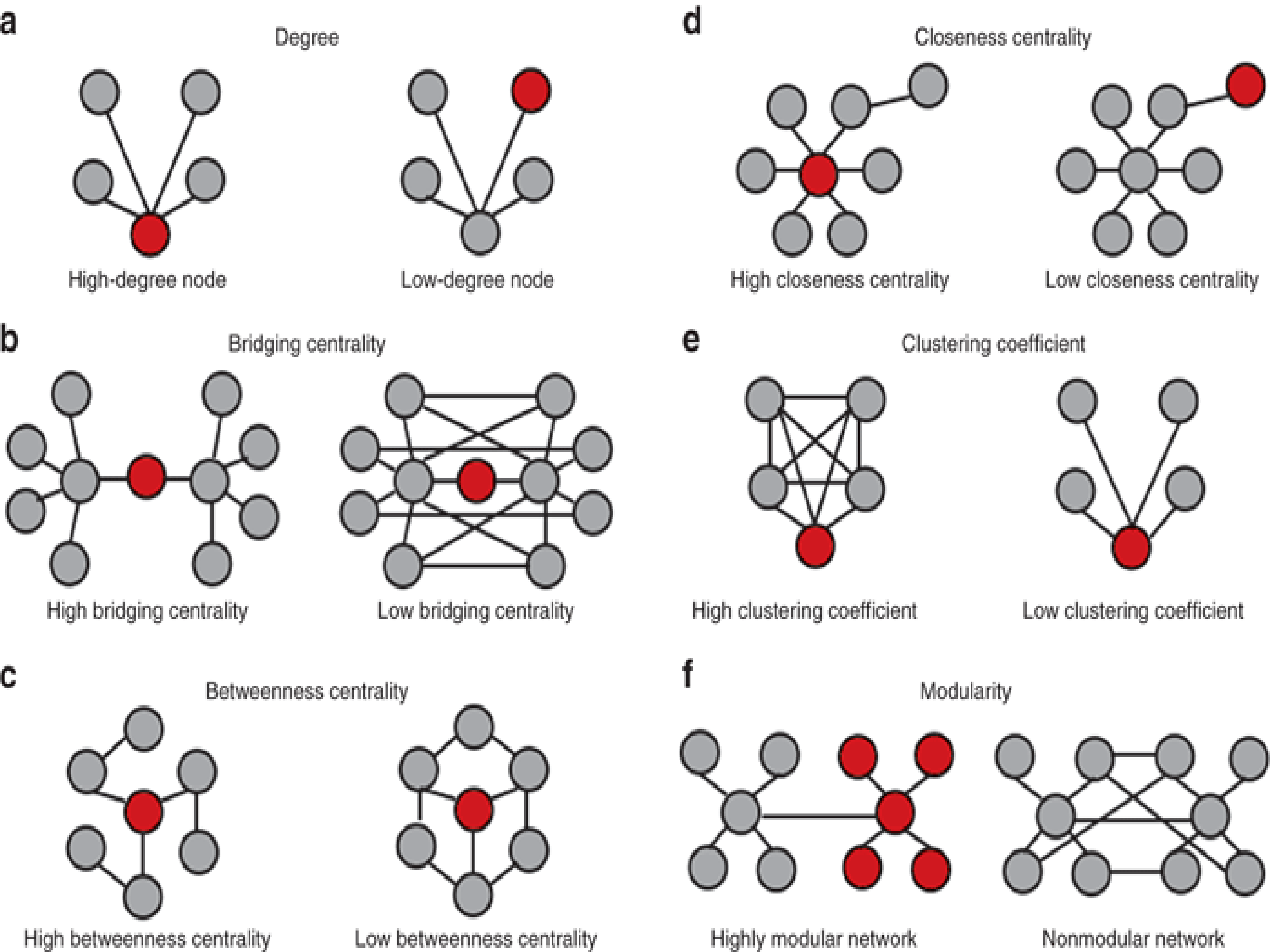
Clustering coefficient

$ACC(G) = \sum cc_i / N$

$ACC(G) = \sum(0.67 + 0.67 + 1 + 1) / 4 = 3.34 / 4 = 0.835$

$GCC(G) = 2 / 4 = 0.5$





a
Degree
High-degree node
Low-degree node
b
Bridging centrality
High bridging centrality
Low bridging centrality
c
Betweenness centrality
High betweenness centrality
Low betweenness centrality
d
Closeness centrality
High closeness centrality
Low closeness centrality
e
Clustering coefficient
High clustering coefficient
Low clustering coefficient
f
Modularity
Highly modular network
Nonmodular network

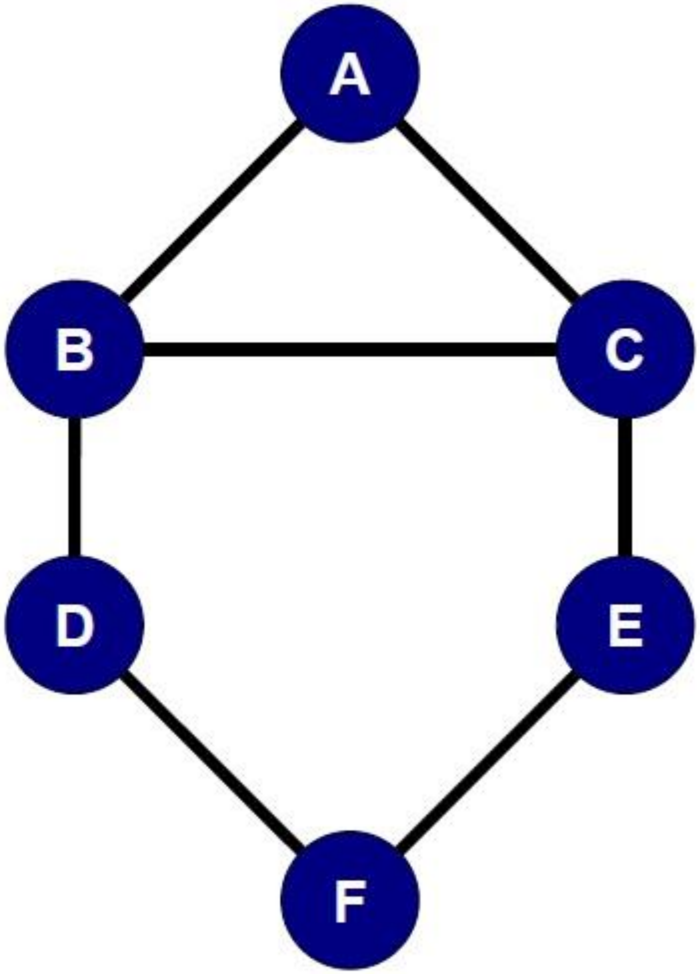
Segments

- SNA toolkit allows to describe larger parts of the graph than triads.
- The most used concepts for network segmentation measurement:
  - cliques / n-cliques
  - k-cores
  - … and many others

Cliques and n-cliques

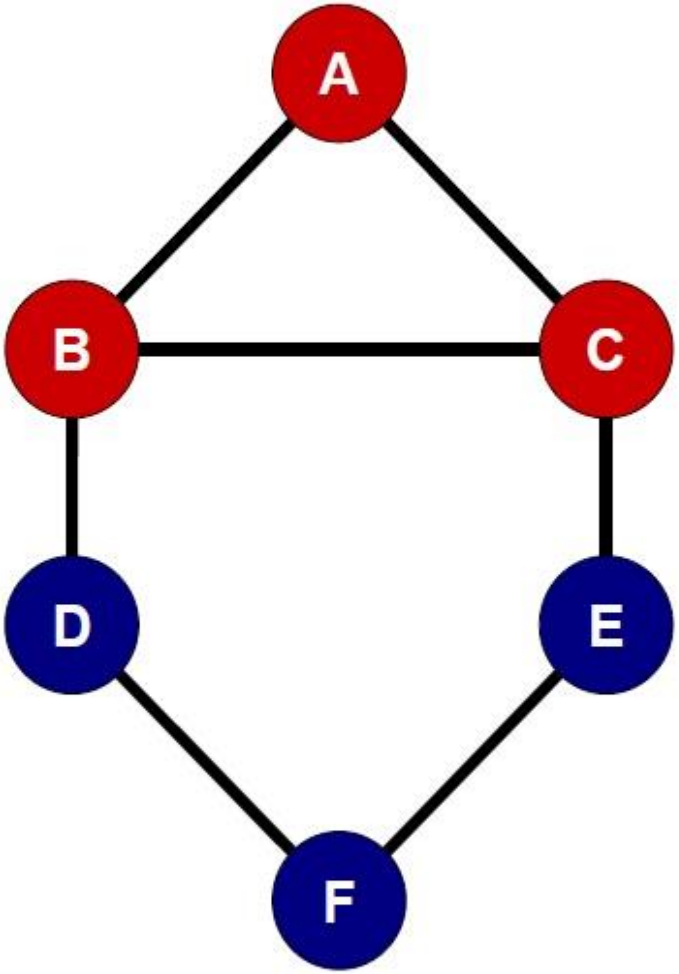
- **Clique** is a **maximal complete subgraph** that consists from three and more nodes.
- → each member of the clique has to be connected to all other members of the clique.
- Strong assumption → **n-clique**.
- N-clique is a maximal subgraph where **longest geodesic** between any two members is not greater than **n**.
- Thus in 2-clique every members is connected to all other members in 1 or 2 steps.

Cliques and n-cliques



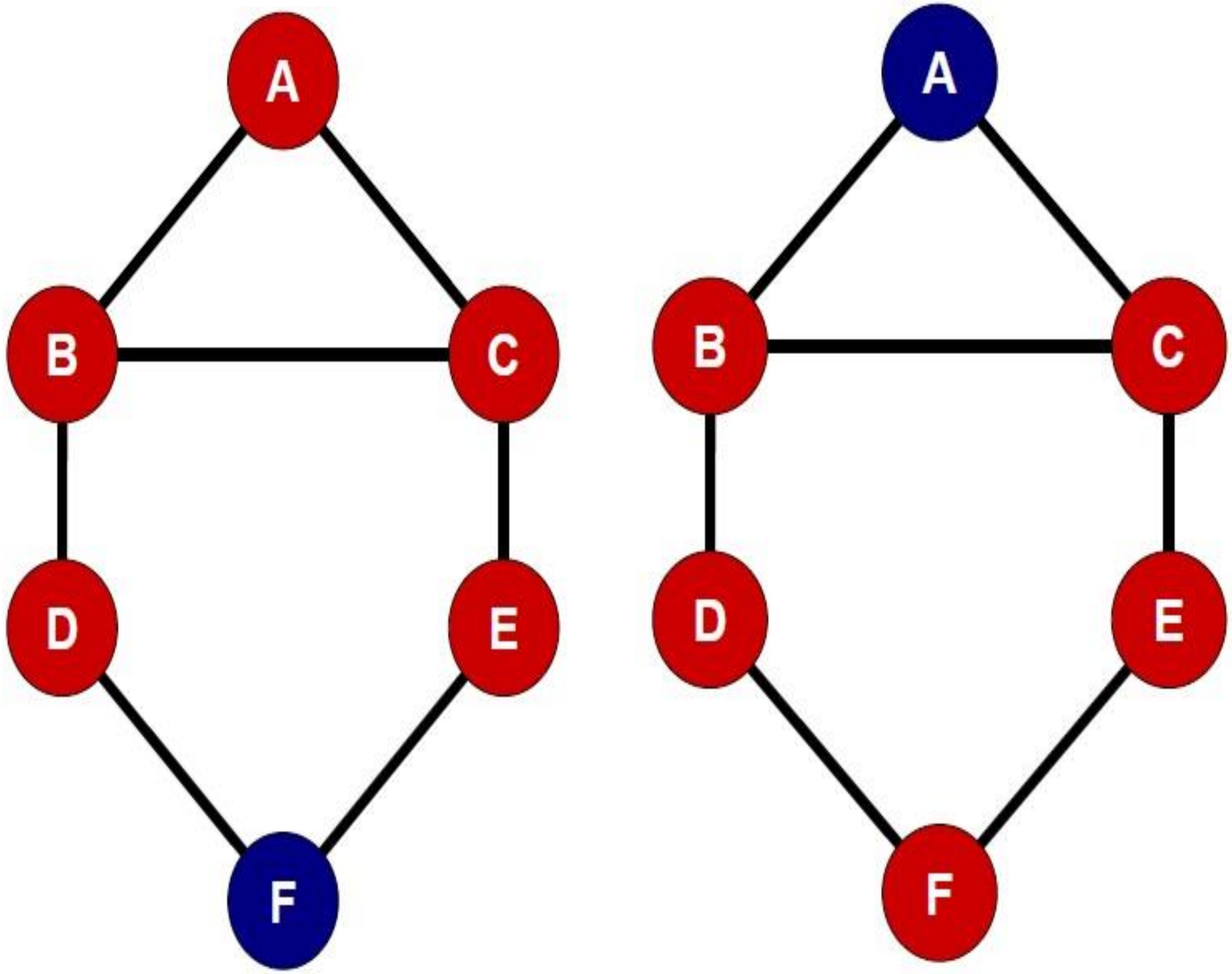
Thurner 2012

Cliques and n-cliques



Thurner 2012

Cliques and n-cliques



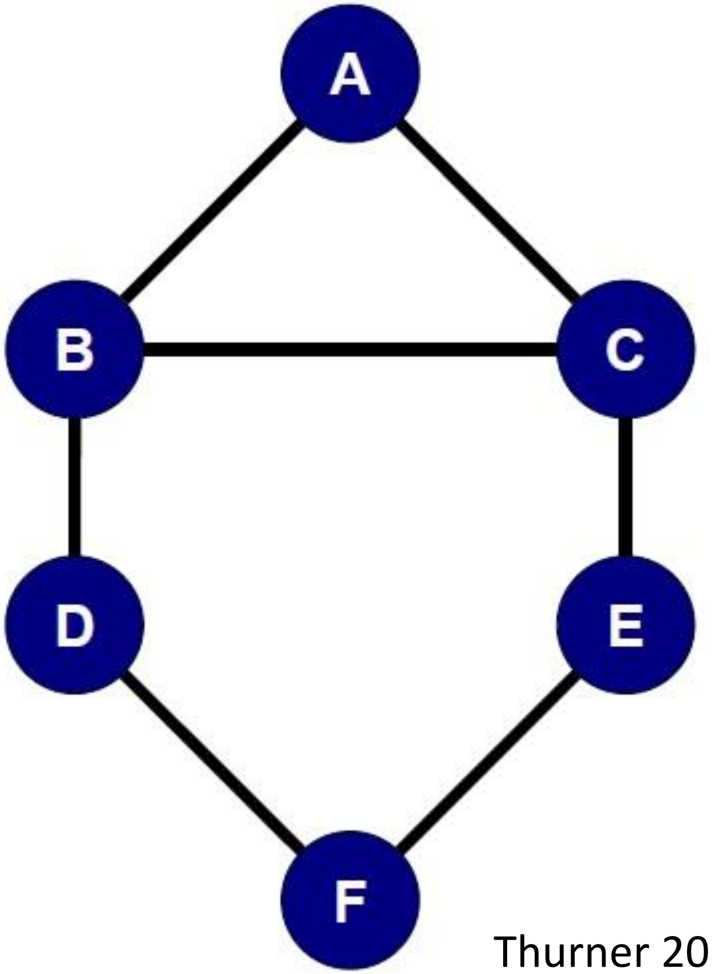
Thurner 2012

K-core

- **K-core** is maximal subgraph where all nodes are connected with specific (*k*) minimal # of nodes in the subgraph.
- Thus: ***k* indicates how many connections each member of the subgraph has to have.**
- Therefore: it is not important how many connections to other members is missing.

K-core

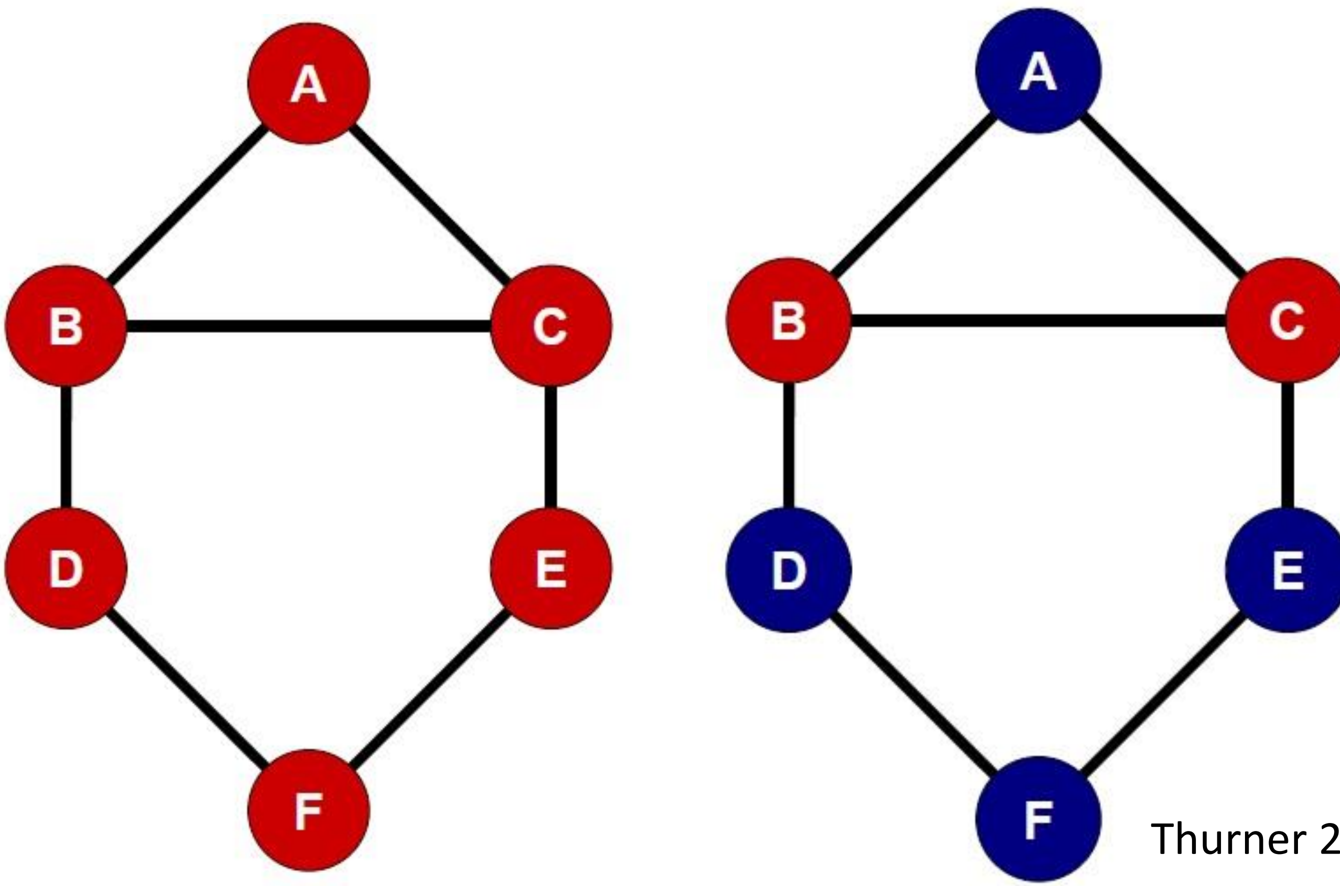
- Find 2-core and 3-core



Thurner 2012

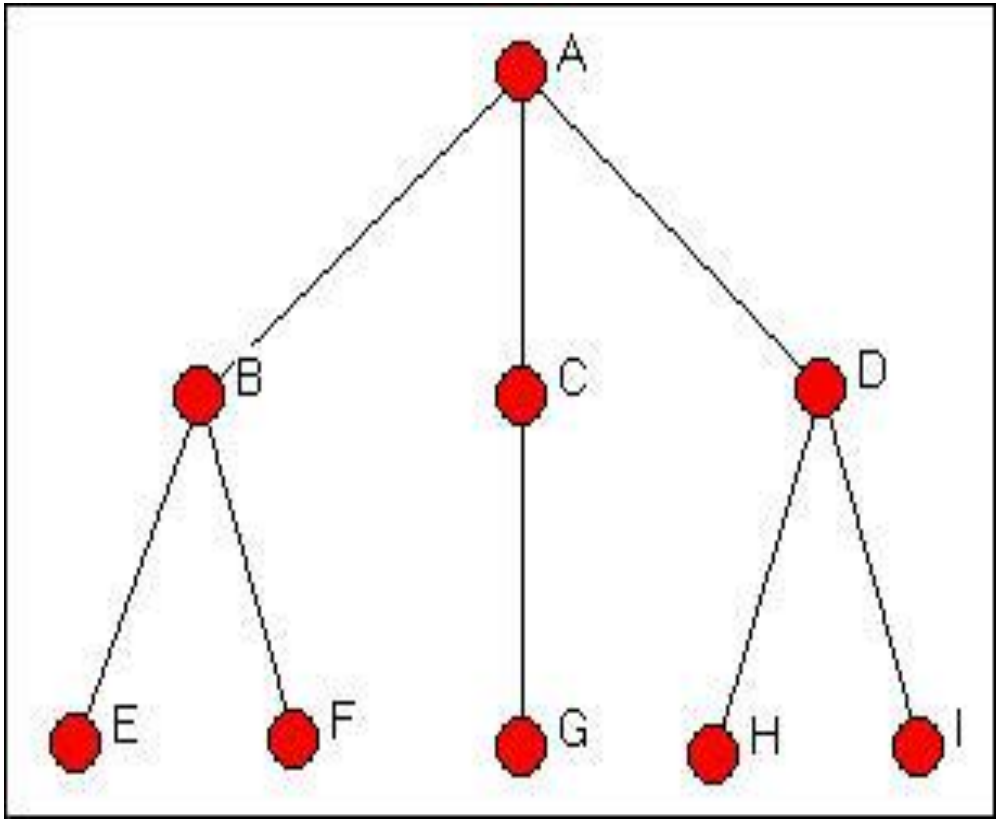
K-core

- Find 2-core and 3-core:



Thurner 2012

Structural equivalence



Hanneman & Riddle 2005

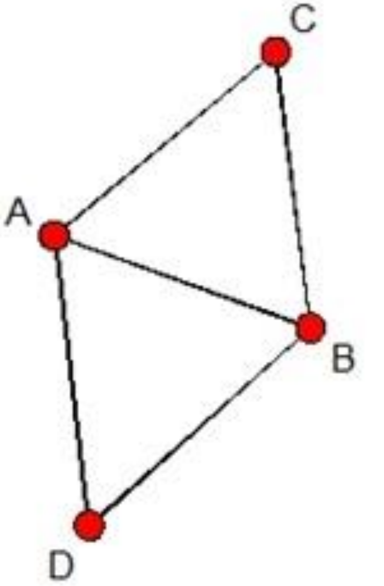
Structural equivalence

- **Social position:** similar ties to other nodes.
  - E.g. Ph.D. students at our department have similar ties to others (under/graduates, department members, supervisors, etc.) as Ph.D. students at other departments.
- **Social role:** pattern of ties to other positions.
  - E.g. professional ties with department members, competitive ties with other Ph.D., friendship ties with other students, etc.

Euclidean distance

- **Euclidean distance (ED):** is a distance of nodes i and j in relation to all other nodes in graph.
- ED of structurally equivalent nodes = 0.
- It is possible to classify nodes based on their ED.

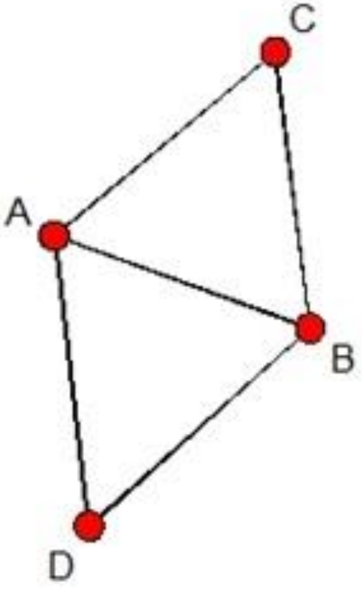
$$d_{ij} = sqrt[\ \sum\sum [\ (n_{ik} - n_{jk})^2 + (n_{ki} - n_{kj})^2\ ]\ ]$$



Euclidean distance

- **(1):** differences of distances of i and j to all other nodes.
- **(2):** sum of squares of the differences (SSD).
- **(3):** square root of the result (SSD).

$$d_{ij} = \text{sqrt}[\ \sum\sum\ [\ (n_{ik} - n_{jk})^2 + (n_{ki} - n_{kj})^2\ ]\ ]$$



Euclidean distance

$$d_{ij} = \text{sqrt}[\ \sum\sum [\ (n_{ik} - n_{jk})^2 + (n_{ki} - n_{kj})^2\ ]\ ]$$

$$d_{ac} = \text{sqrt}[\ \sum\sum [\ (n_{ak} - n_{ck})^2 + (n_{ka} - n_{kc})^2\ ]\ ]\ ;\ k = \{b, d\}$$

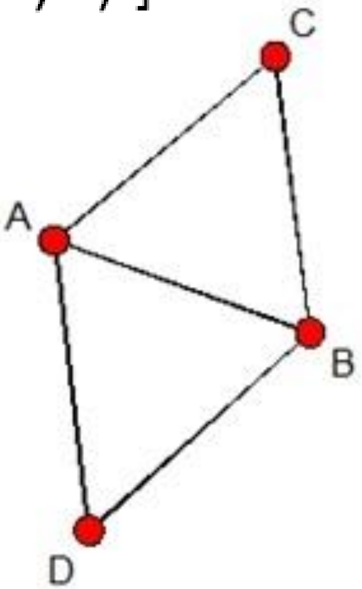
$$d_{ac} = \text{sqrt}[\ (\ (ab - cb)^2 + (ba - bc)^2\ ) + (\ (ad - cd)^2 + (da - dc)^2\ )\ ]$$

$$d_{ac} = \text{sqrt}[\ (\ (1 - 1)^2 + (1 - 1)^2\ ) + (\ (1 - 2)^2 + (1 - 2)^2\ )\ ]$$

$$d_{ac} = \text{sqrt}(0 + 0 + 1 + 1) = \text{sqrt}(2) = 1.41$$

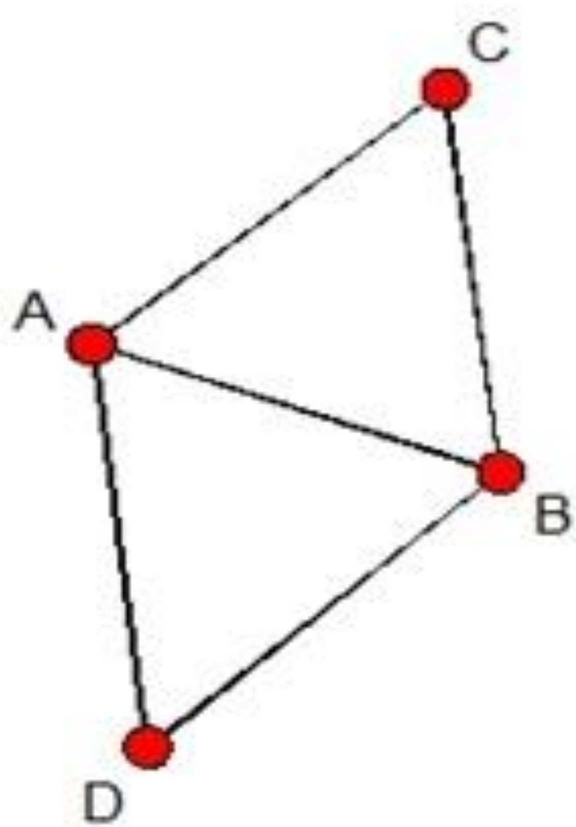
0	**1**	**1**	**1**
1	**0**	**1**	**1**
1	**1**	**0**	**0**
1	**1**	**0**	**0**

0	**0**	**1.41**	**1.41**
0	**0**	**1.41**	**1.41**
1.41	**1.41**	**0**	**0**
1.41	**1.41**	**0**	**0**



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

0	0	1.41	1.41
0	0	1.41	1.41
1.41	1.41	0	0
1.41	1.41	0	0



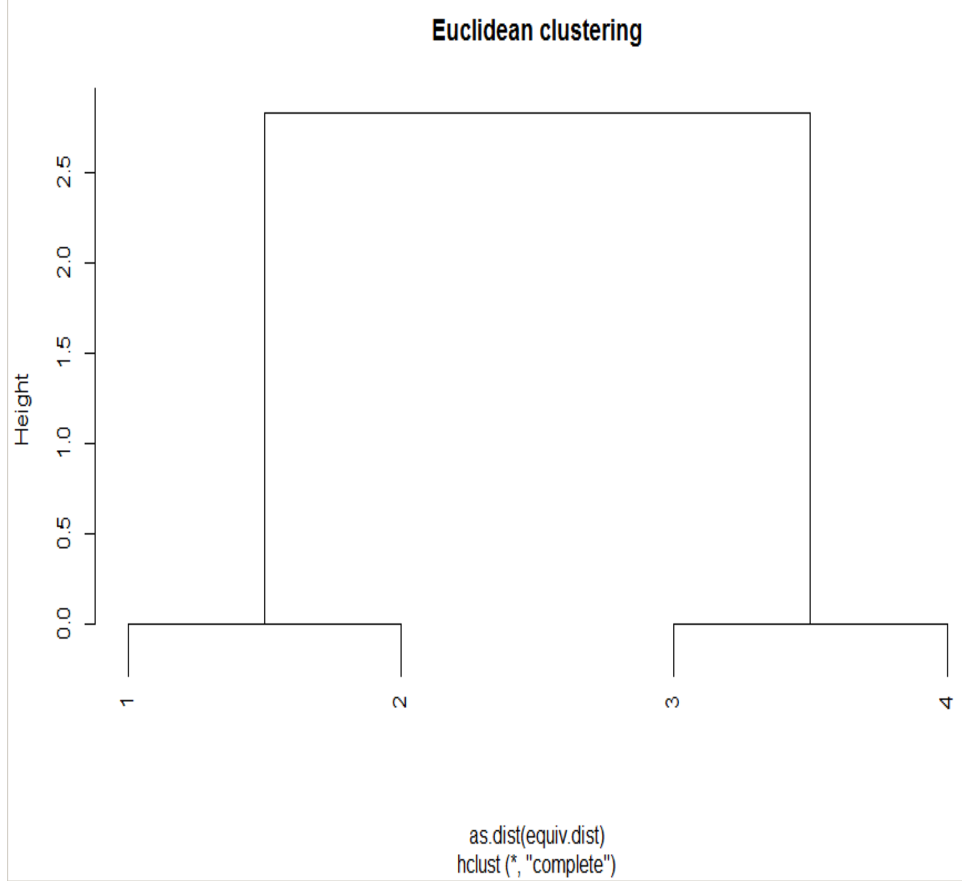


Illustration: energy interdependence

- **Research objective:**
  – Mapping energy interdependence relations at the European natural gas market (exploratory objective).
- **Research question:**
  – What is the level of energy interdependence at the European market?
- **Research importance:**
  – Collection of data and exploration.
  – Necessary step for explanatory research (liberal peace hypotheses).

Network border delineation

- positional strategy of border delineation:
  - European NG consumers and their suppliers.
- sample ~ population :
  - EU28 + Norway, FYROM, Ukraine, Belarus, Turkey and its suppliers.

Conceptualization / operationalization

- Operationalization (Barbieri 1996):
  - Interdependence has two dimensions: saliency and symmetry.
  - **Trade share (TS):** bilateral trade flow over total trade flow
    - TS = tradeAB/tradeAW ; tradeBA/tradeBW
  - **Saliency (S):** trade shares product of A and B
    - S = sqrt(tradeAB/tradeAW * tradeBA/tradeBW)
  - **Symmetry (Y):** difference of trade shares of A and B
    - Y = 1 – abs(tradeAB/tradeAW – tradeBA/tradeBW)
  - **Interdependence (I):**
    - **I = sqrt(V * S)**
- Attribute variables:
  - exporter / importer
  - Composite Index of National Capability (CINC)

Interdependence: calculation

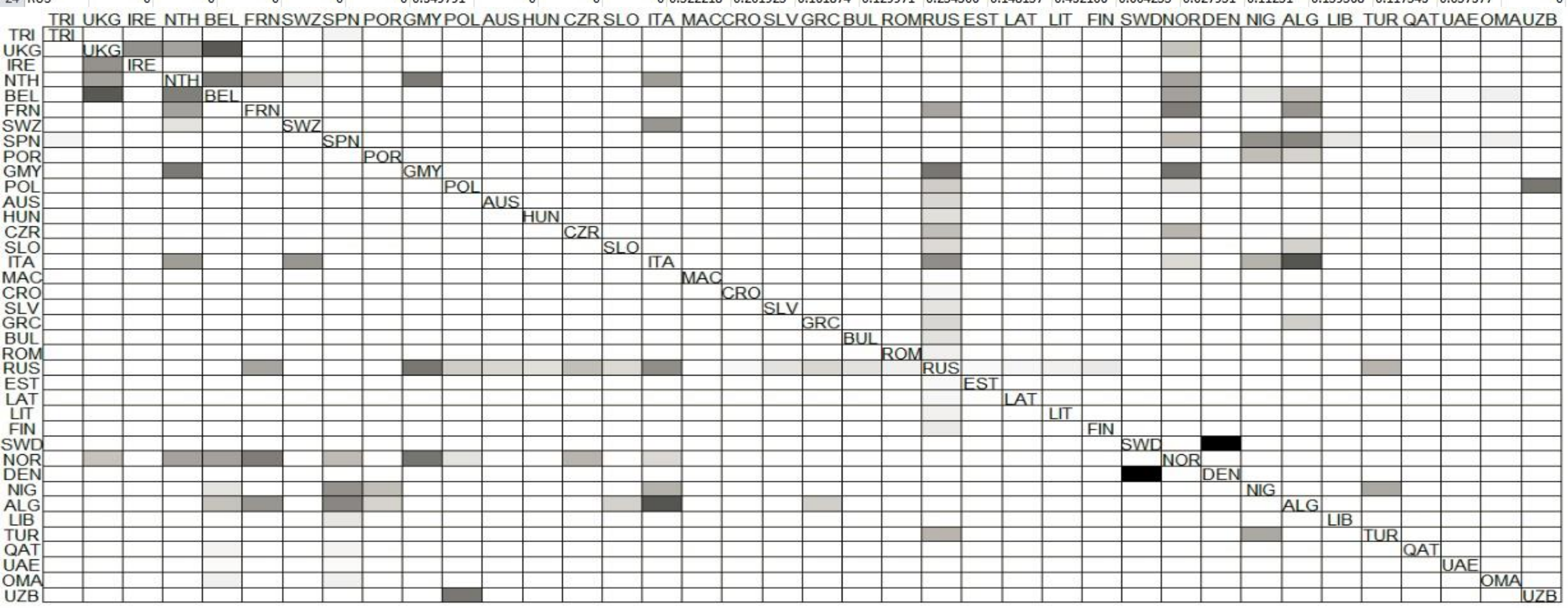
CZE-RUS	
share (bcm/y)	CZE: 7/9 = 0.77 (**77** %); RUS: 7/150 = 0.05 (**5** %)
saliency	sqrt(0.77*0.05) = 0.20 (**20** %)
symmetry	1 – abs(0.05 - 0.77) = 1 – 0.72 = 0.28 (**28** %)
interdependence	sqrt(0.20*0.28) = 0.24 (**24** %)
weighted interdependence	sqrt(0.24*sqrt(0.20*0.04)) = 0.15 (**15** %)

GER-RUS	
share (bcm/y)	GER: 40/85 = 0.47 (**47** %); RUS: 40/150 = 0.27 (**27** %)
saliency	sqrt(0.47*0.27) = 0.36 (**36** %)
symmetry	1 – abs(0.47 - 0.27) = 1 – 0. 2 = 0.8 (**80** %)
interdependence	sqrt(0.36*0.80) = 0.54 (**54** %)
weighted interdependence	sqrt(0.54*sqrt(0.24*0.04)) = 0.23 (**23** %)

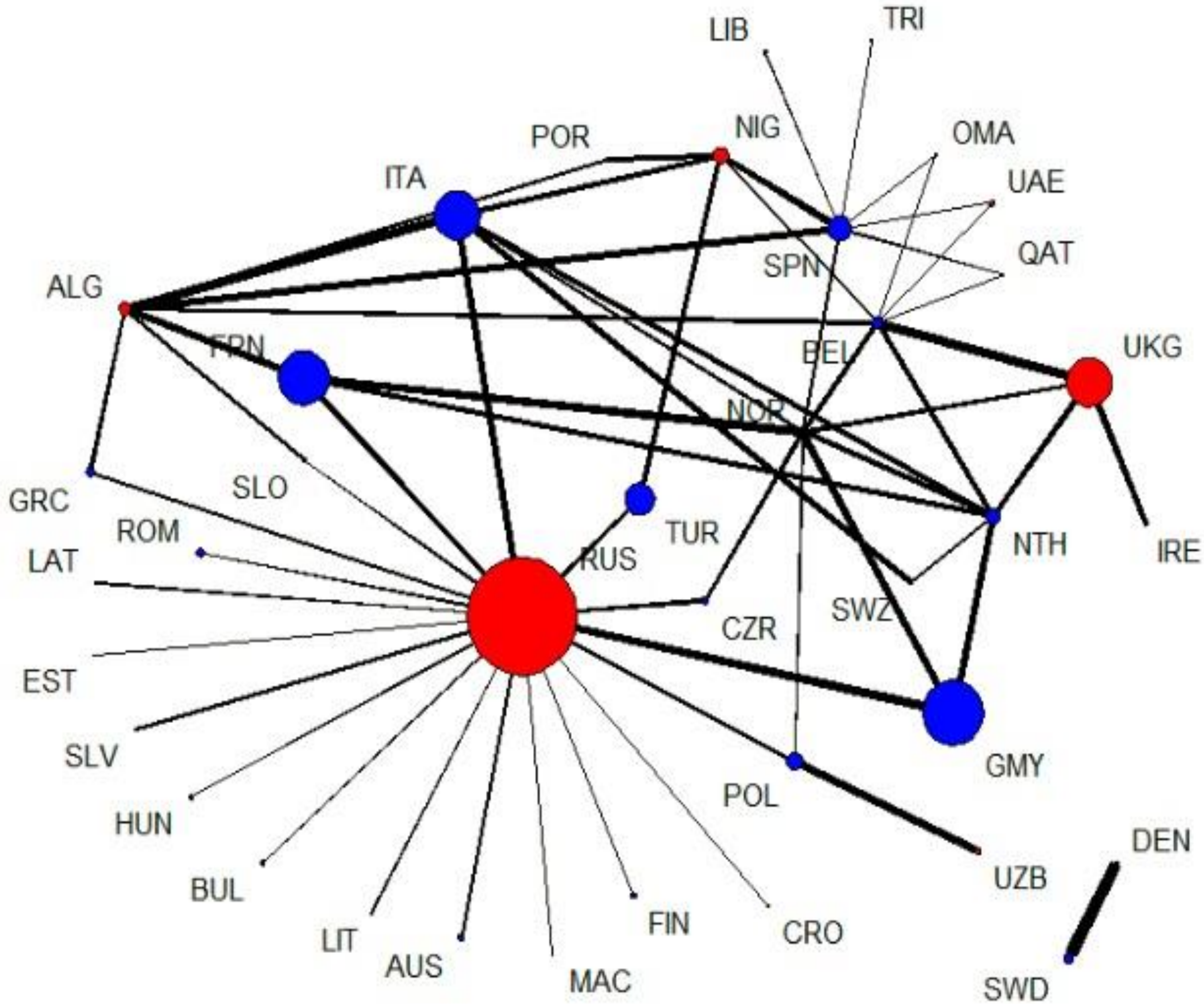
Attribute variables

variable	operationalization (bcm/y)	CZE	GER
NG consumption / TPES	NG consumption / TPES	0.20 (20 %)	0.24 (24 %)
import / consumption	import NG / consumption NG	98 %	90/95 = 0.95 (95 %)
diversification (concentration)	HH = ∑shares^2	0.7^2 + 0.3^2 = 0.58 (58 %)	0.23^2 + 0.33^2 + 0.40^2 = 0.32 (32 %)
substitutability	N - 1	0.2 (20 %)	0.6 (60 %)
storage capacity / total consumption	storage capacity / consumption NG	3/9 = 0.3 (30 %)	24/95 = 0.25 (25 %)
political regime	Freedom House Index	0.82 (82 %)	0.84 (84 %)
...			

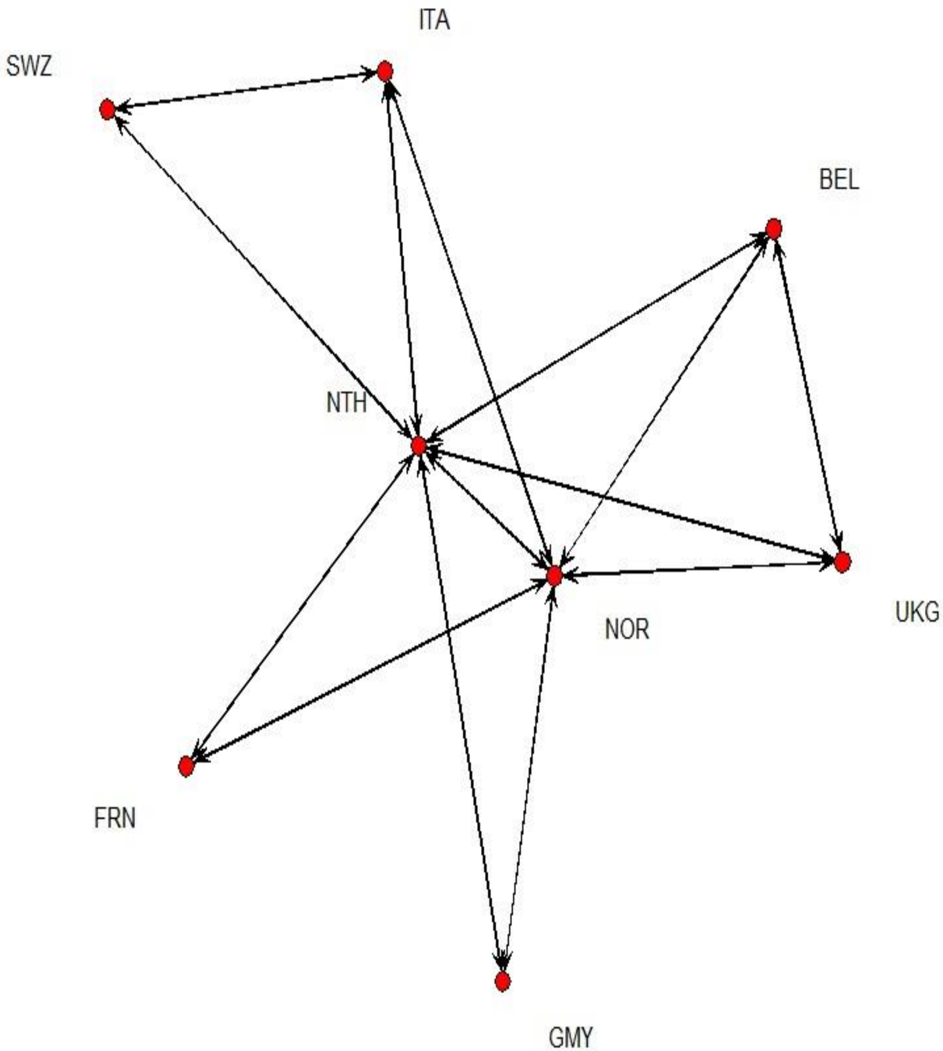
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1		TRI	UKG	IRE	NTH	BEL	FRN	SWZ	SPN	POR	GMY	POL	AUS	HUN	CZR	SLO	ITA	MAC	CRO	SLV	GRC	BUL	ROM	RUS
2	TRI	0	0	0	0	0	0	0	0.077603	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	UKG	0	0	0.414797	0.358292	0.64829	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	IRE	0	0.414797	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	NTH	0	0.358292	0	0	0.520007	0.355999	0.113207	0	0	0.509085	0	0	0	0	0	0.378856	0	0	0	0	0	0	0
6	BEL	0	0.64829	0	0.463071	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	FRN	0	0	0	0.355999	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.349791
8	SWZ	0	0	0	0.113207	0	0	0	0	0	0	0	0	0	0	0	0.40387	0	0	0	0	0	0	0
9	SPN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	POR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	GMY	0	0	0	0.509085	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.522218
12	POL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.201925
13	AUS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.161874
14	HUN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.129971
15	CZR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.254506
16	SLO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.148157
17	ITA	0	0	0	0.378856	0	0	0.40387	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.432106
18	MAC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.004255
19	CRO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.027951
20	SLV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.11251
21	GRC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.159568
22	BUL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.117543
23	ROM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.057377
24	RUS	0	0	0	0	0	0.349791	0	0	0	0.522218	0.201925	0.161874	0.129971	0.254506	0.148157	0.432106	0.004255	0.027951	0.11251	0.159568	0.117543	0.057377	0

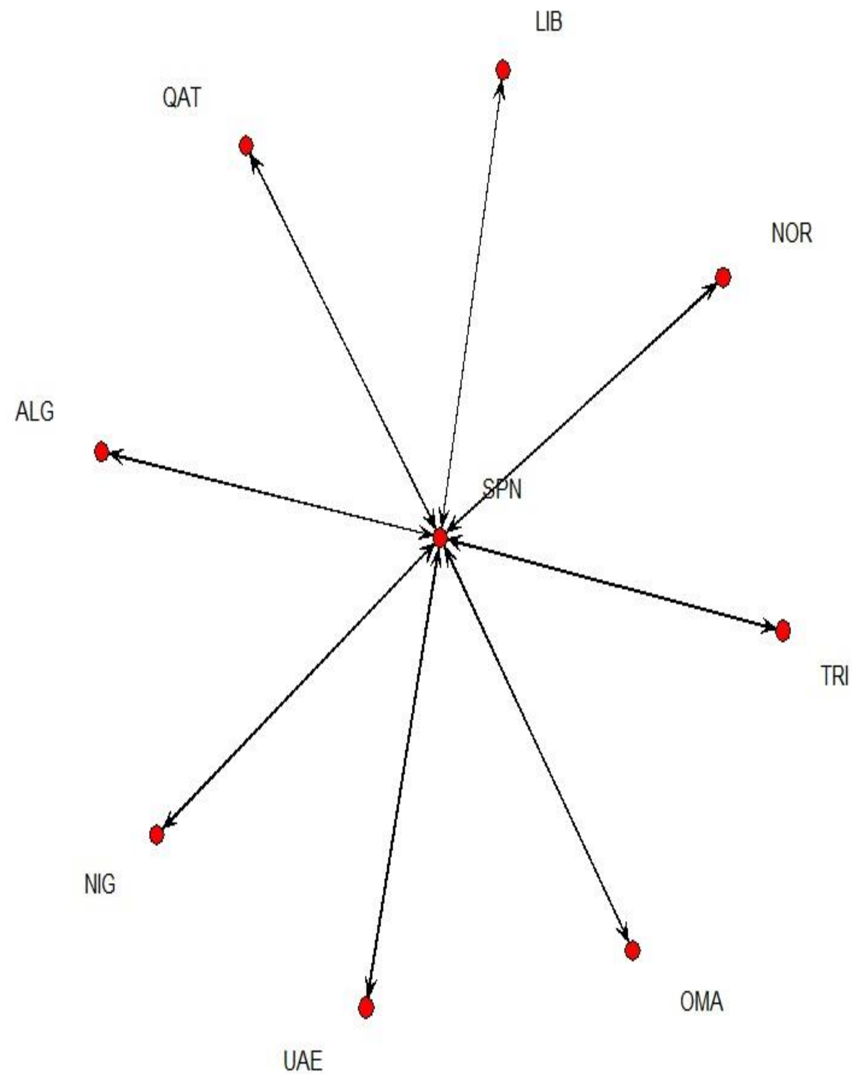


Interdependence network (European natural gas market (2001))

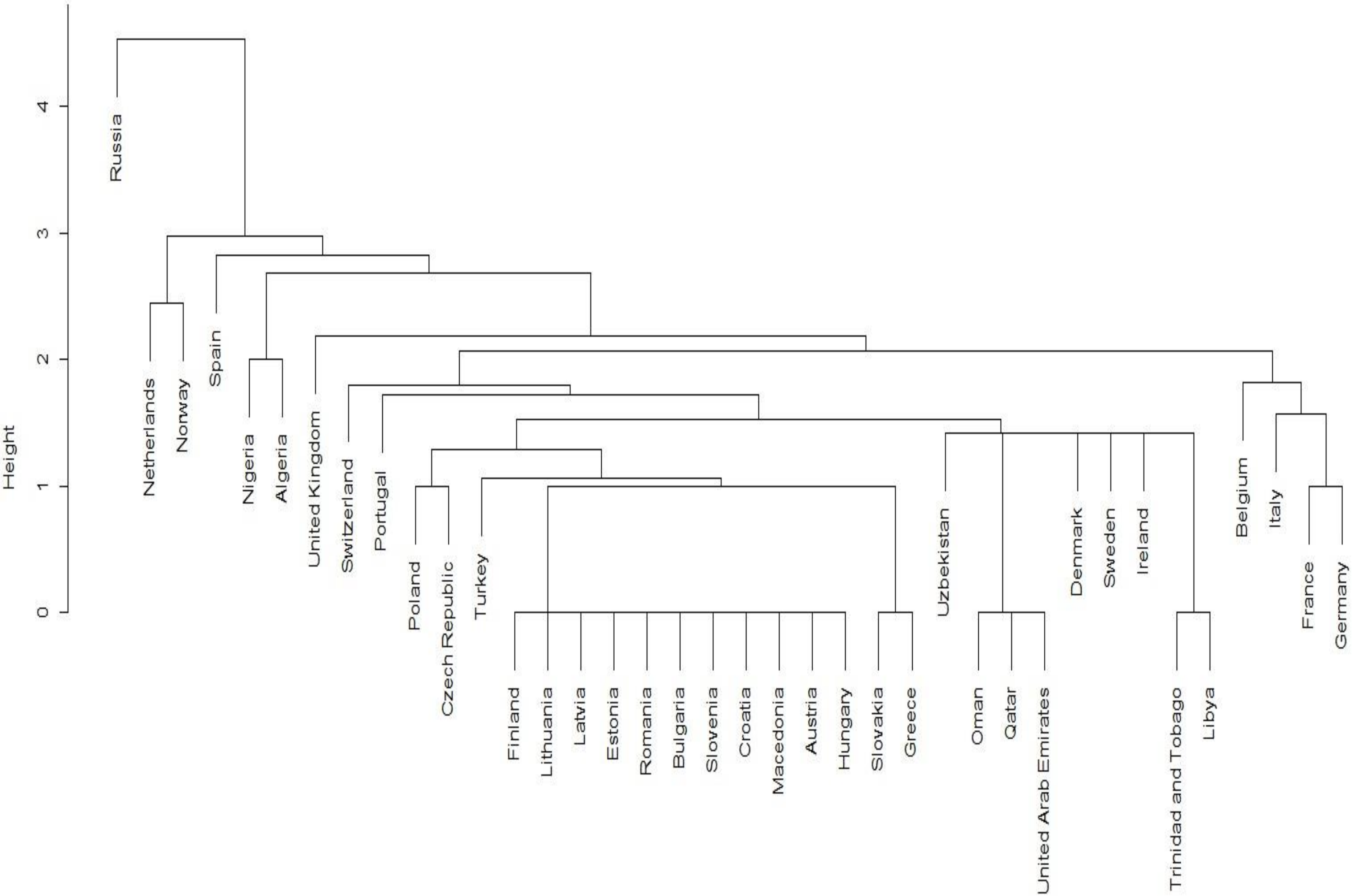


edge width: interdependence
node size: national capabilities
exporter (red), importer (blue)


ITA
SWZ
BEL
NTH
NOR
UKG
FRN
GMY


LIB
QAT
NOR
ALG
SPN
TRI
NIG
OMA
UAE

Cluster Dendrogram



Assignment 2

- Create a new script.
- Generate a random graph with 5 nodes.
- Calculate degree, closeness, and betweenness centrality.
- Calculate degree centralization.
- Visualize graph and report the centrality / centralization results.
- Bonus: display node size as a function of its degree centrality.