Social network analysis 1 + 2

Petr Ocelík

ESS418 Research Methods in Social Science

9th October 2015

Outline

- Empirical instances of networks
- History of SNA
- Graph theory
- Data organization
- Mini-case study
- R: working with network data

Introduction

- Networks are everywhere.

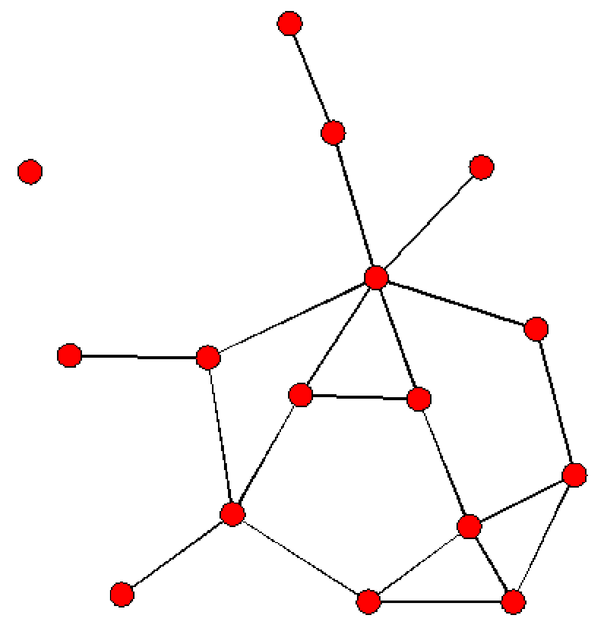
- Social disciplines are – by definition – dealing with social **inter**actions.

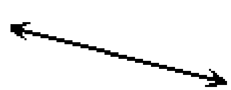
- SNA allows us to collect and analyze **relational** data.

Introduction

- The main assumption: world is organized **relationally**.

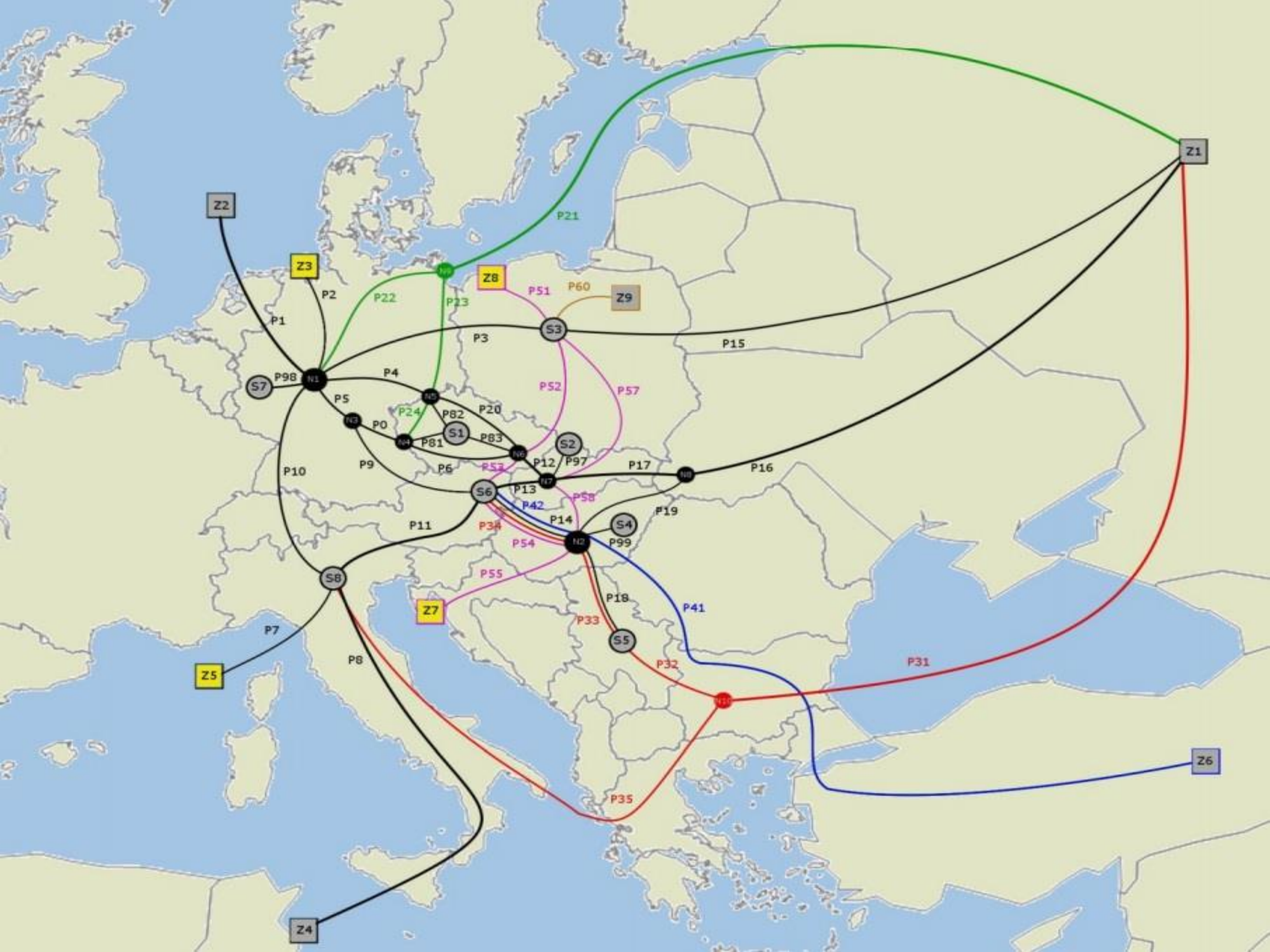
“...transactions, interactions, social ties,
and conversations constitute
central stuff of social life.”
(Tilly 2008: 7)

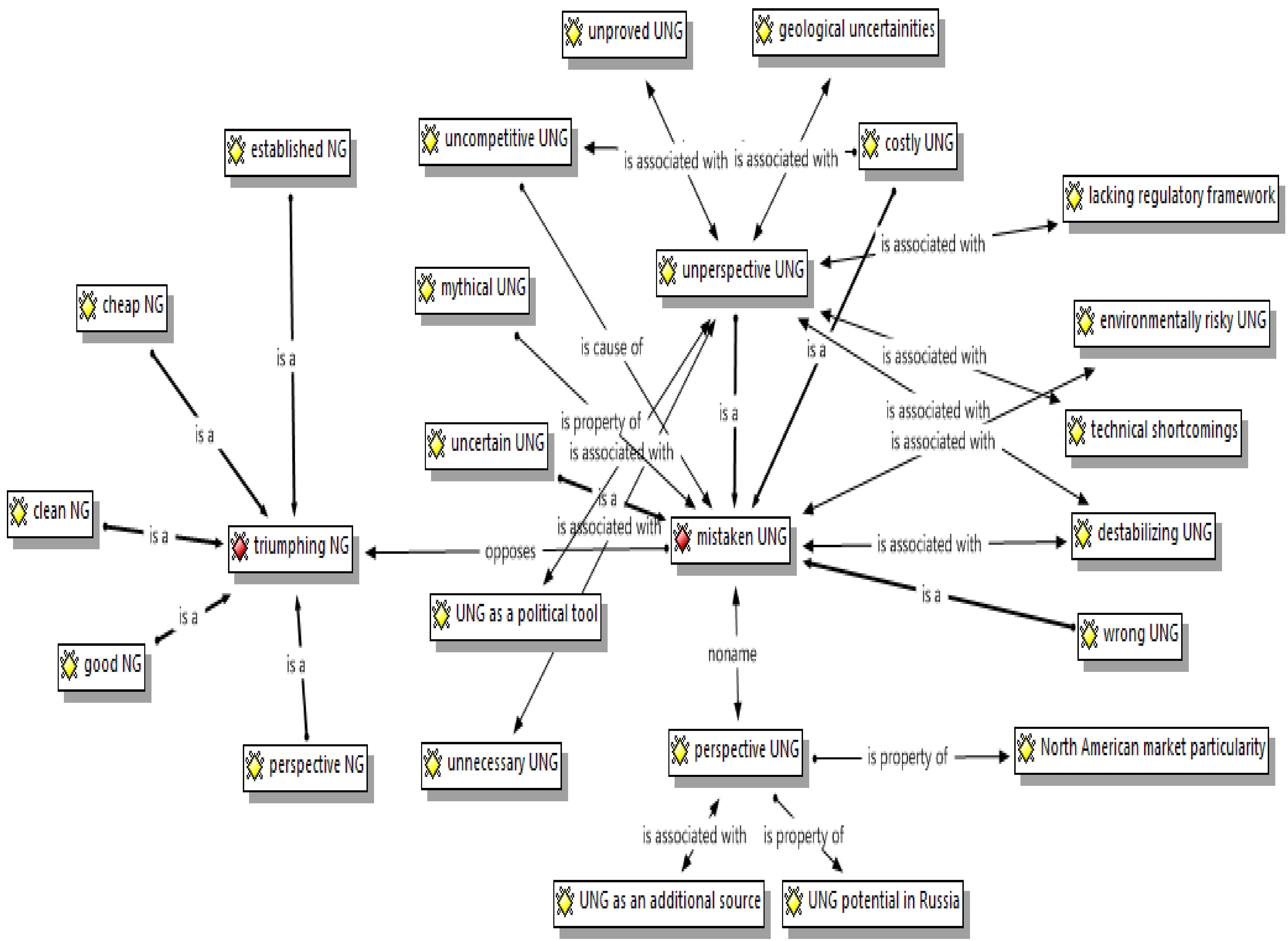


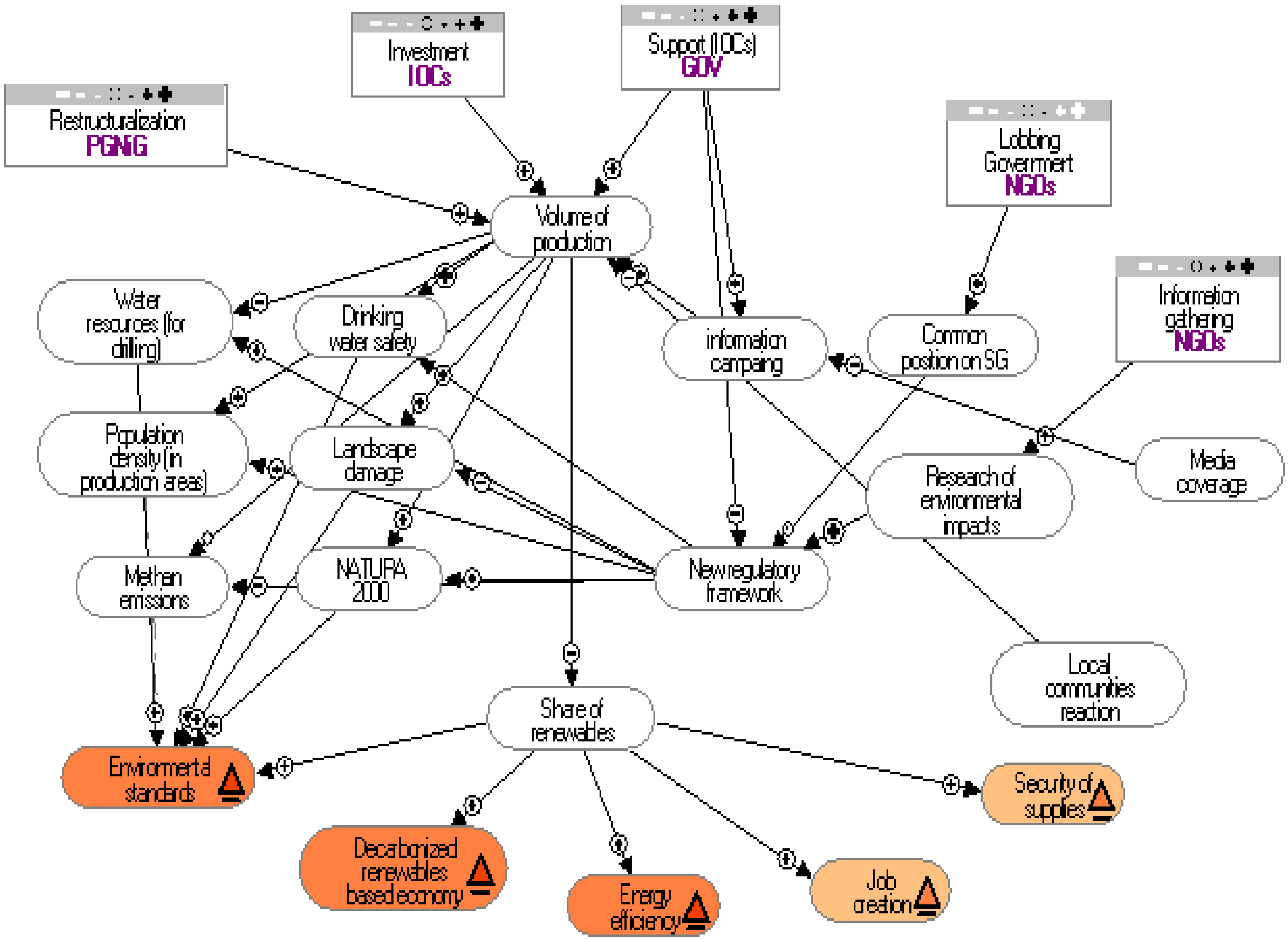
- node 
- edge 

Terminology (Guclu 2012)

points	lines	
vertices	edges, arcs	math
nodes	links	computer science
sites	bonds	physics
actors	ties, relations	sociology


Z1
Z2
Z3
Z4
Z5
Z6
Z7
Z8
Z9
S1
S2
S3
S4
S5
S6
S7
S8
N1
N2
N3
N4
N5
N6
N7
N8
N9
N10
P0
P1
P2
P3
P4
P5
P6
P7
P8
P9
P10
P11
P12
P13
P14
P15
P16
P17
P18
P19
P20
P21
P22
P23
P24
P31
P32
P33
P34
P35
P41
P42
P51
P52
P53
P54
P55
P57
P58
P60
P81
P82
P83
P97
P98
P99


unproved UNG
geological uncertainities
uncompetitive UNG
costly UNG
established NG
lacking regulatory framework
unperspective UNG
mythical UNG
cheap NG
environmentally risky UNG
technical shortcomings
uncertain UNG
clean NG
triumphing NG
mistaken UNG
destabilizing UNG
wrong UNG
good NG
UNG as a political tool
perspective NG
unnecessary UNG
perspective UNG
North American market particularity
UNG as an additional source
UNG potential in Russia
is associated with
is a
is cause of
is property of
opposes
noname


Restructuralization
PGNiG
Investment
IOCs
Support (IOCs)
GOV
Lobbing Goverment
NGOs
Information gathering
NGOs
Volume of production
Water resources (for drilling)
Drinking water safety
information campaing
Common position on SG
Population density (in production areas)
Landscape damage
Research of enviromental impacts
Media coverage
Methan emissions
NATURA 2000
New regulatory framework
Local communities reaction
Share of renewables
Enviromental standards
Decarbonized renewables based economy
Energy efficiency
Job creation
Security of supplies

History of SNA

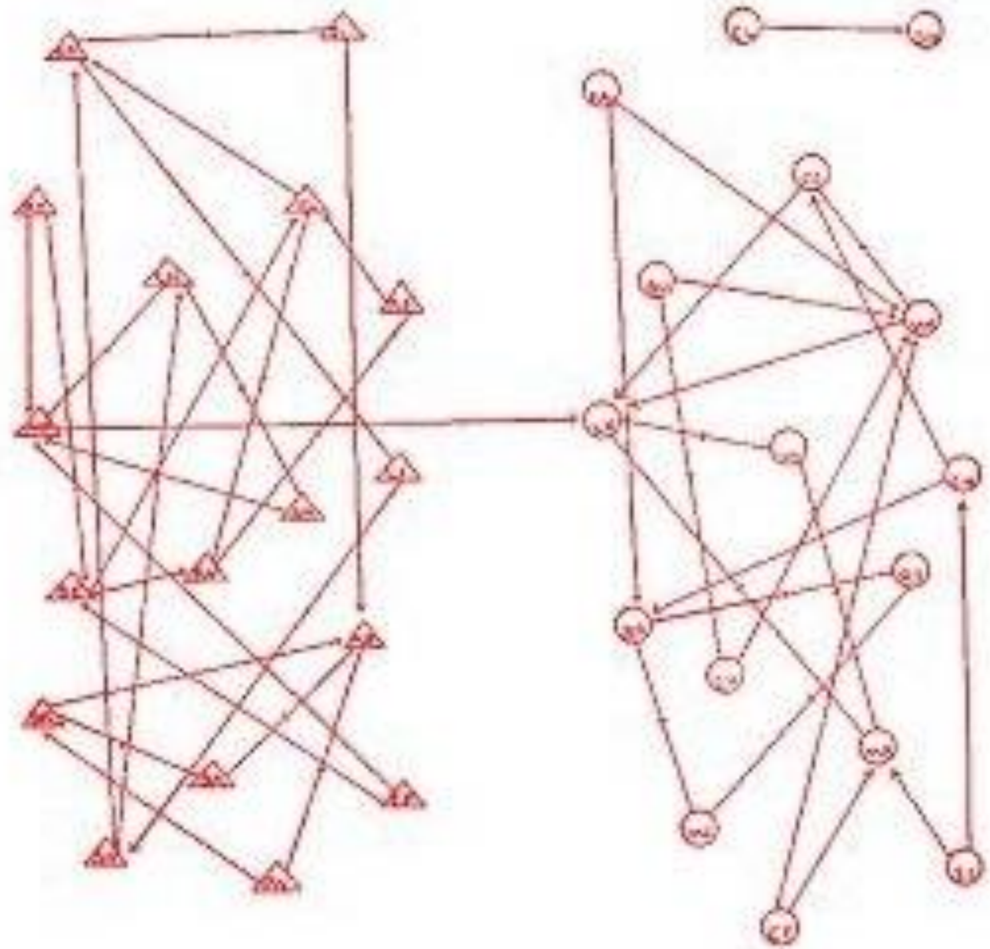
- The beginnings of SNA fall into 1930s.
- Mostly connected with work of Jacob Moreno.
- SNA had been developing on ad hoc basis in separate research centers.
- The structural approach has been widely recognized in 1970s (Mark Granovetter 1973).
- The revolution of social physics in 1990s:
  - Watts and Strogatz (1998): Small-world networks
  - Barabasi and Albert (1999): Scale-free networks

Jacob Moreno

EMOTIONS MAPPED BY NEW GEOGRAPHY

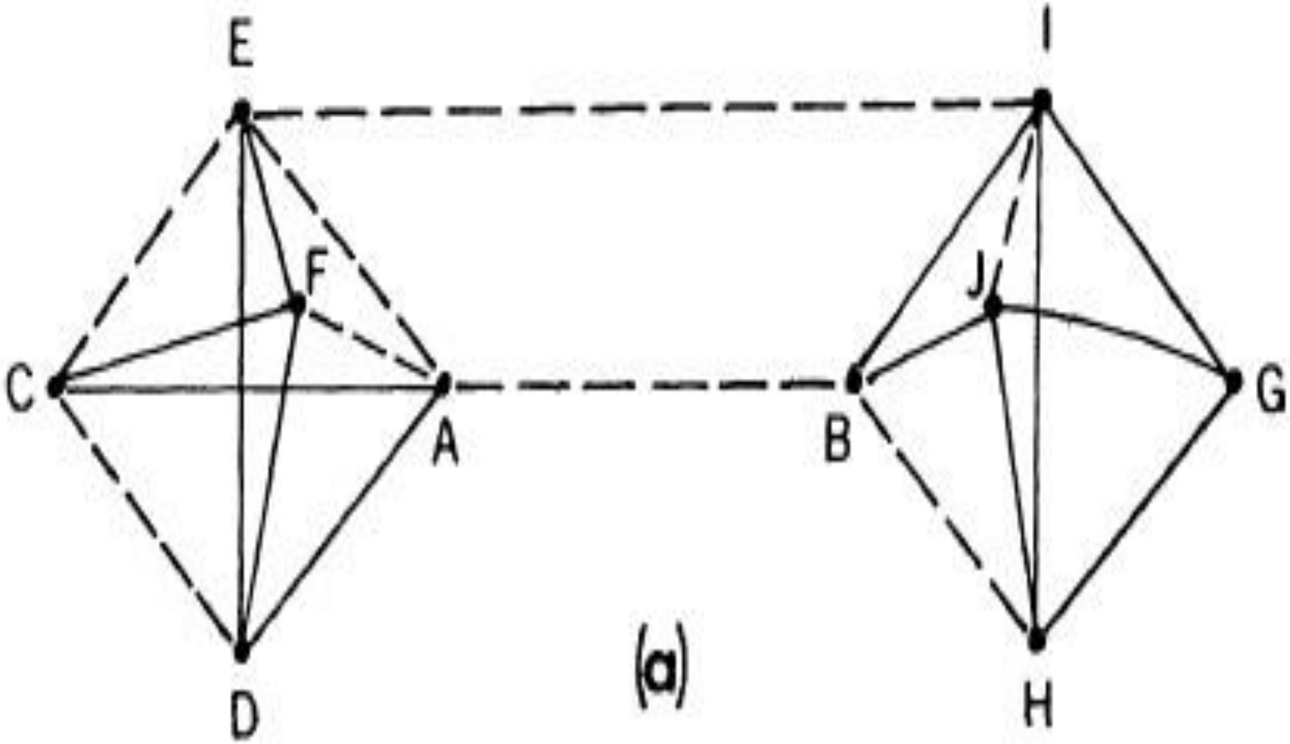
Charts Seek to Portray the Psychological Currents of Human Relationships.

New York Times
April 3, 1933

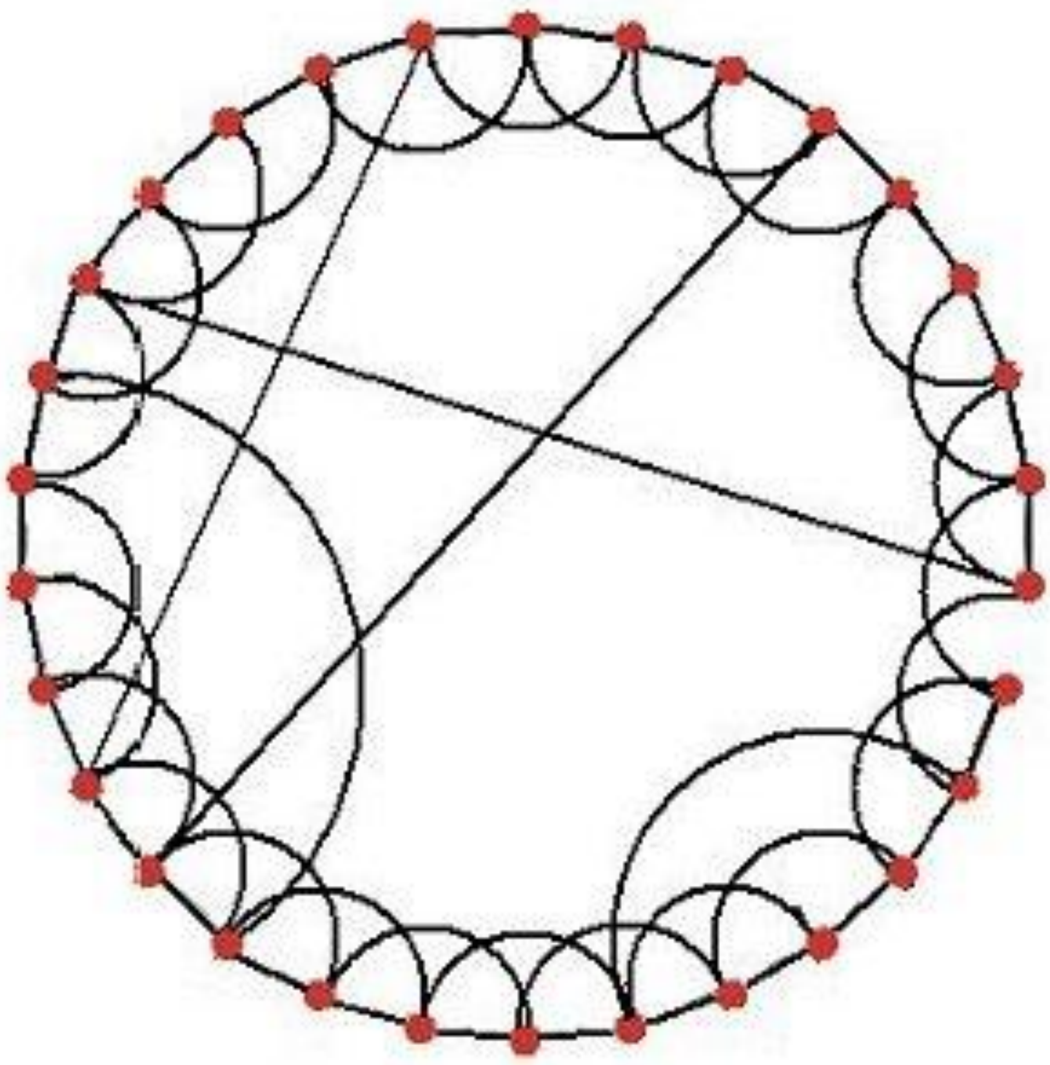


Mark Granovetter

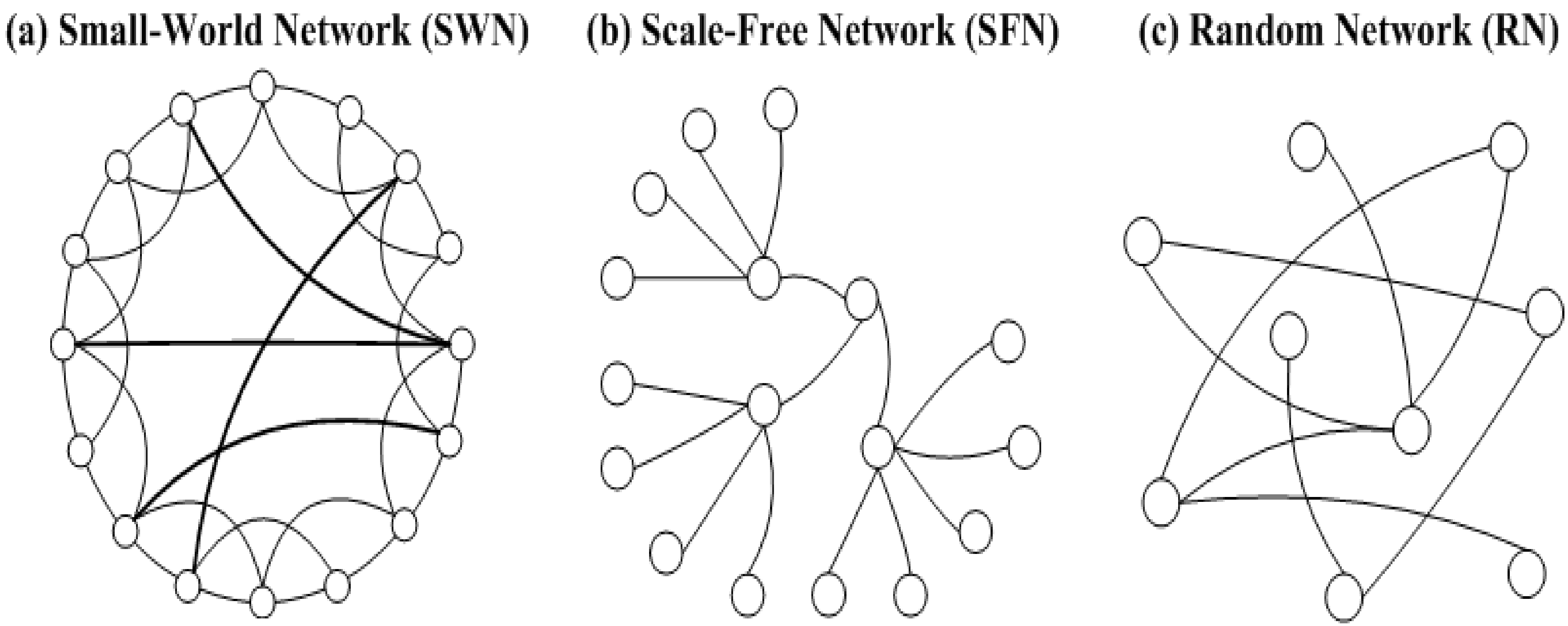
The Strength of Weak Ties



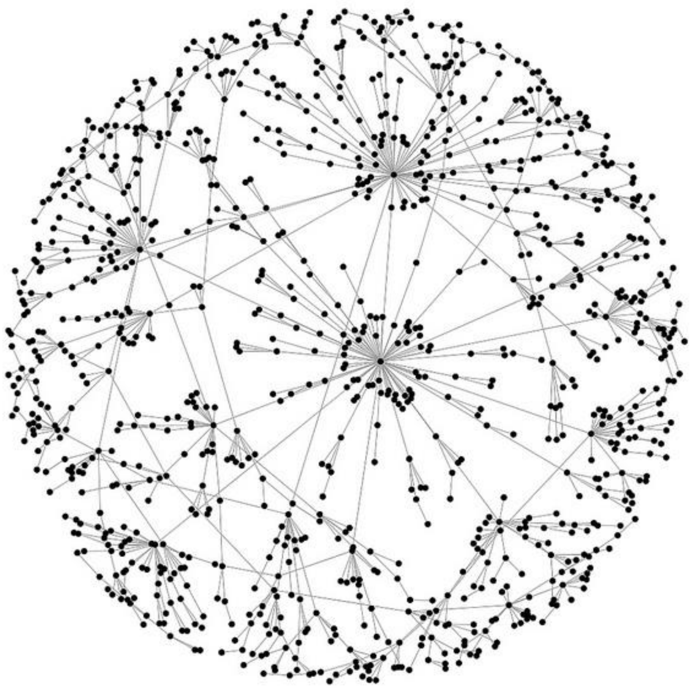
Small-world network

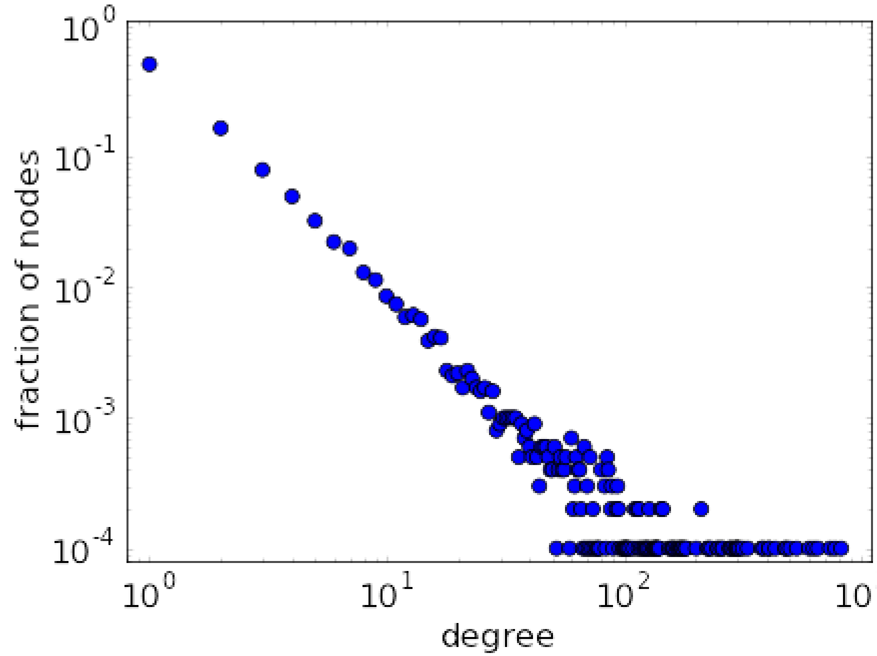


Scale-free network



Scale-free network





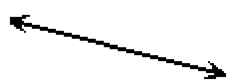
Graph theory

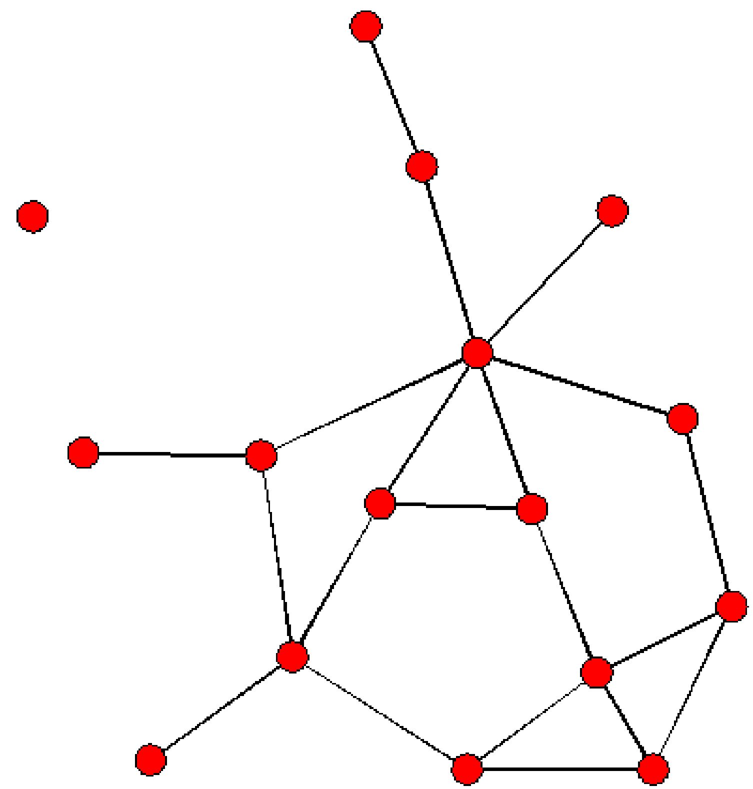
- Graph theory investigates graphs, i.e. mathematical structures that **model pairwise relations between objects**.
- A **graph** (G) is an ordered pair consisting from a set of vertices (V) and a set of (undirected) edges (E) or (directed) arcs (A).
- **G = (V, E v A)**

Graph theory

- Network consists from a set of **nodes** and a set of **edges**.

node 

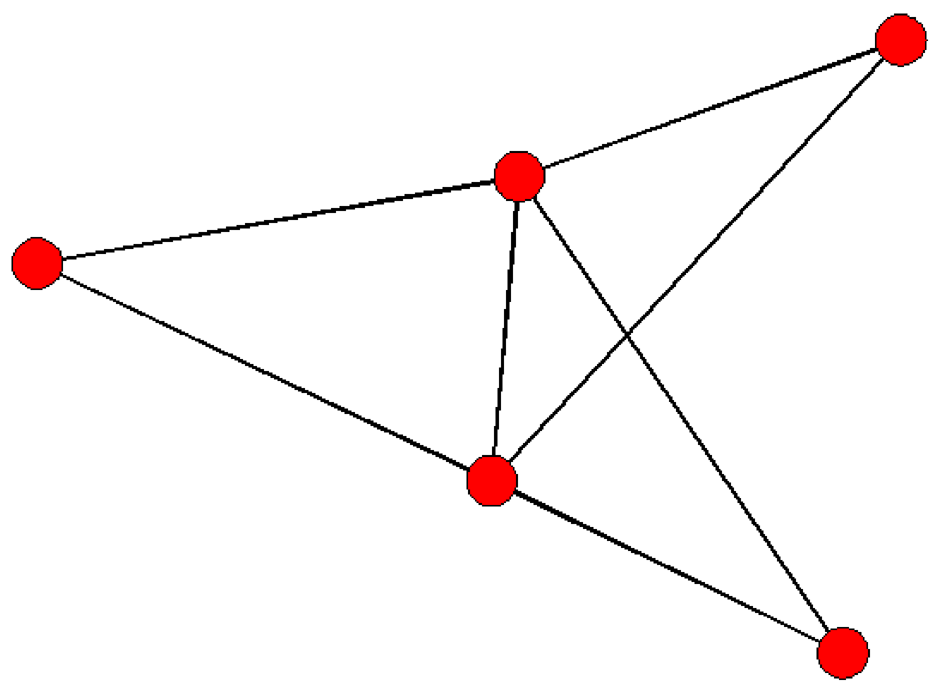
edge 



- **network = graph**

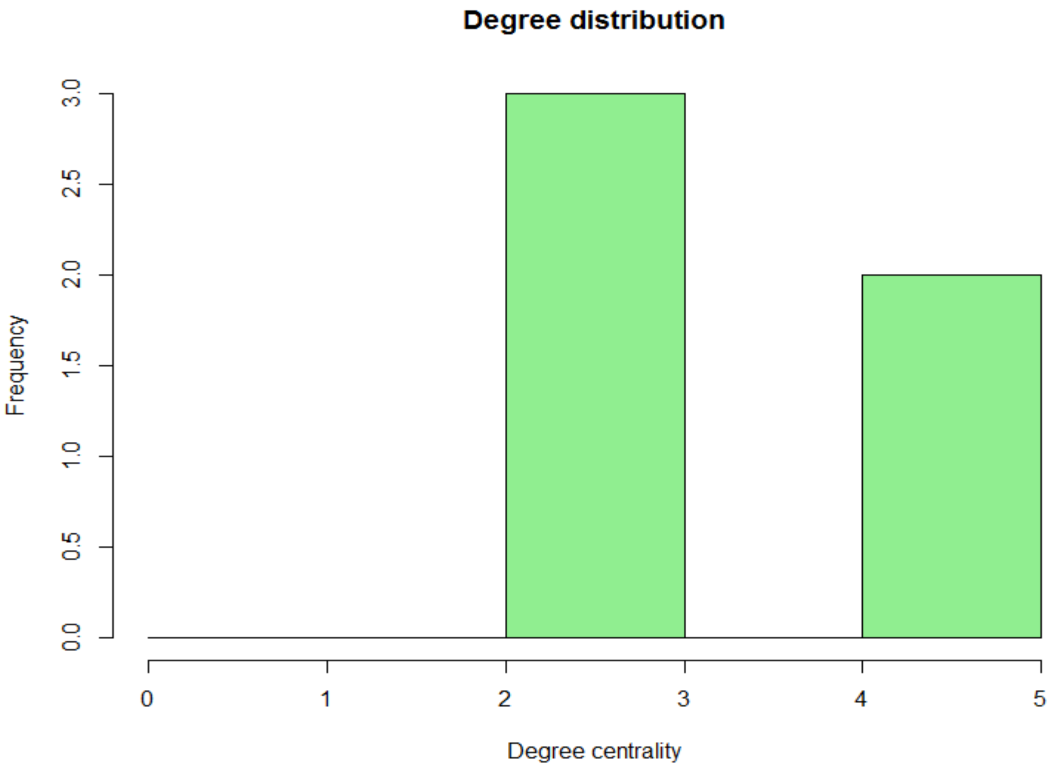
Graph theory

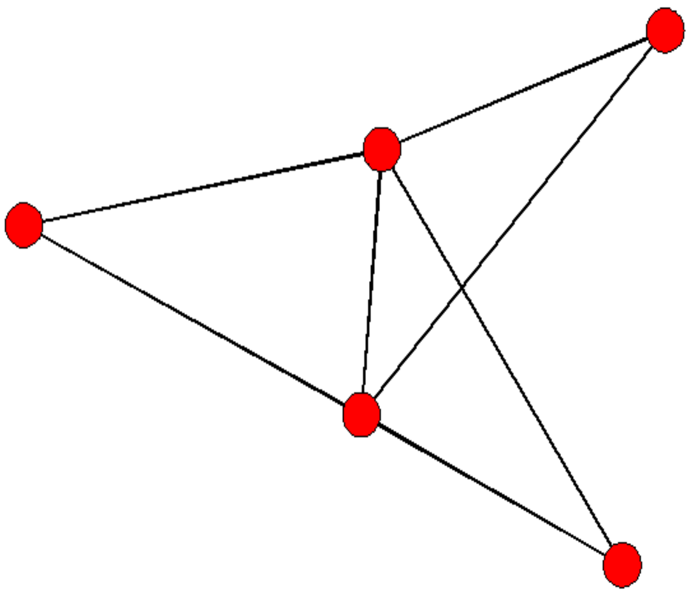
- **order** = # nodes
- **size** = # edges
- **degree** = # connections of individual nodes



Graph theory

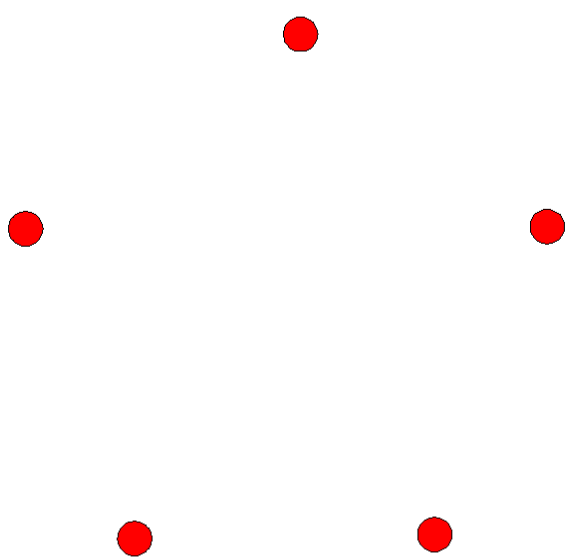
- order = 5
- size = 7

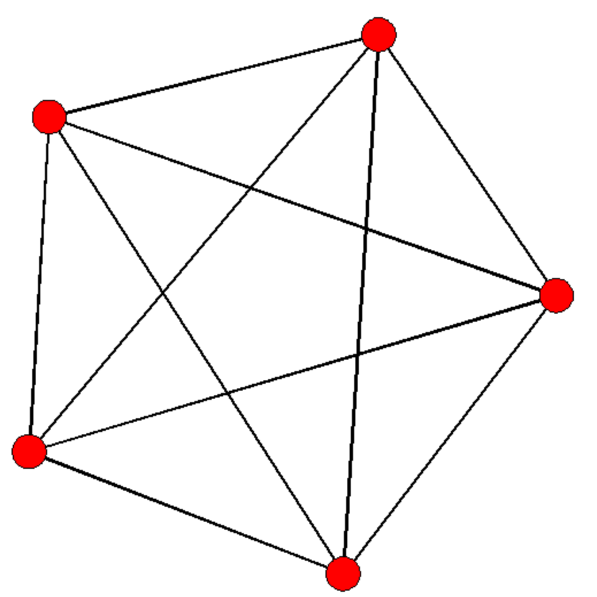




Graph theory

- **Complete graph** is maximally connected graph.
- **Empty graph** is graph with no edges.



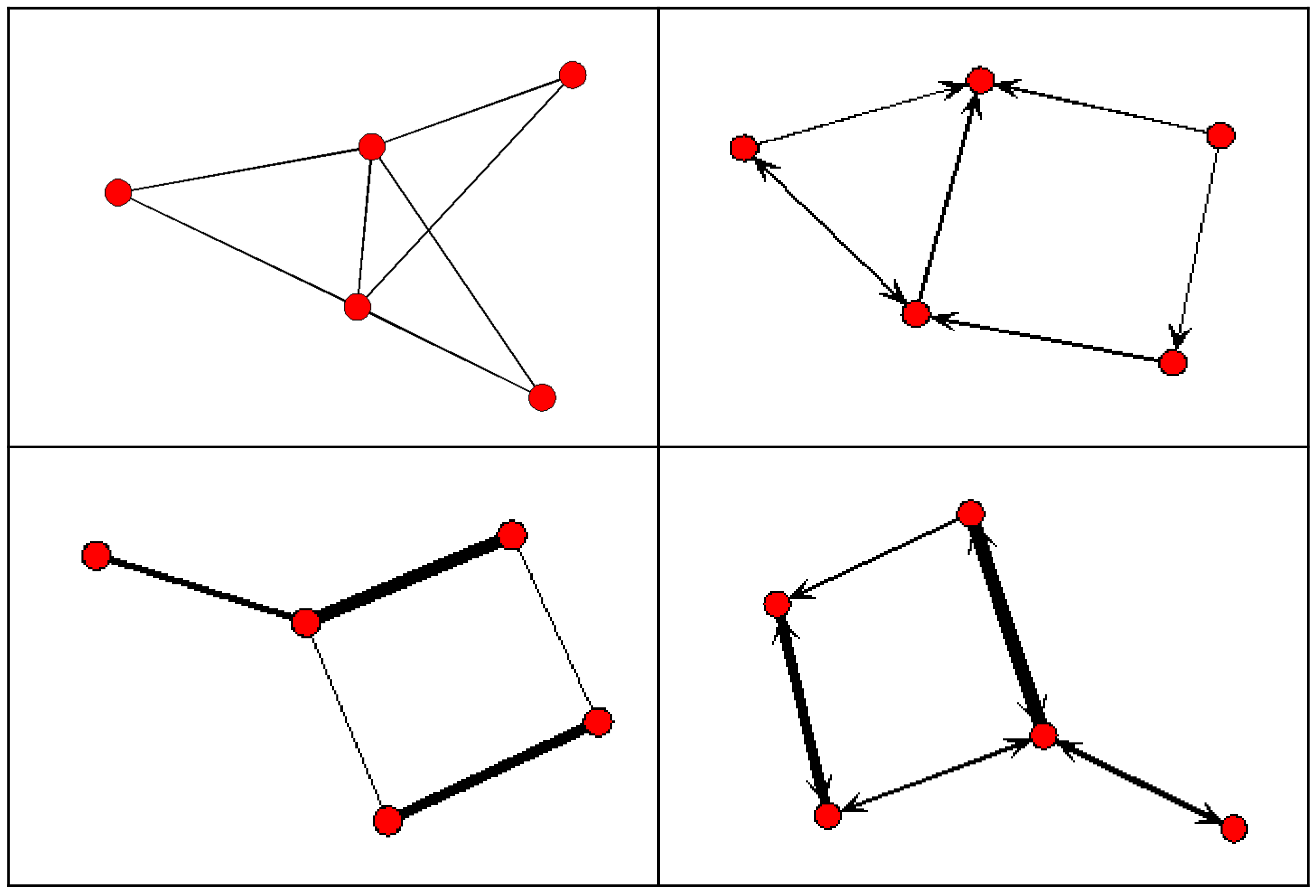


Graph theory: relations

undirected directed

binary

weighted



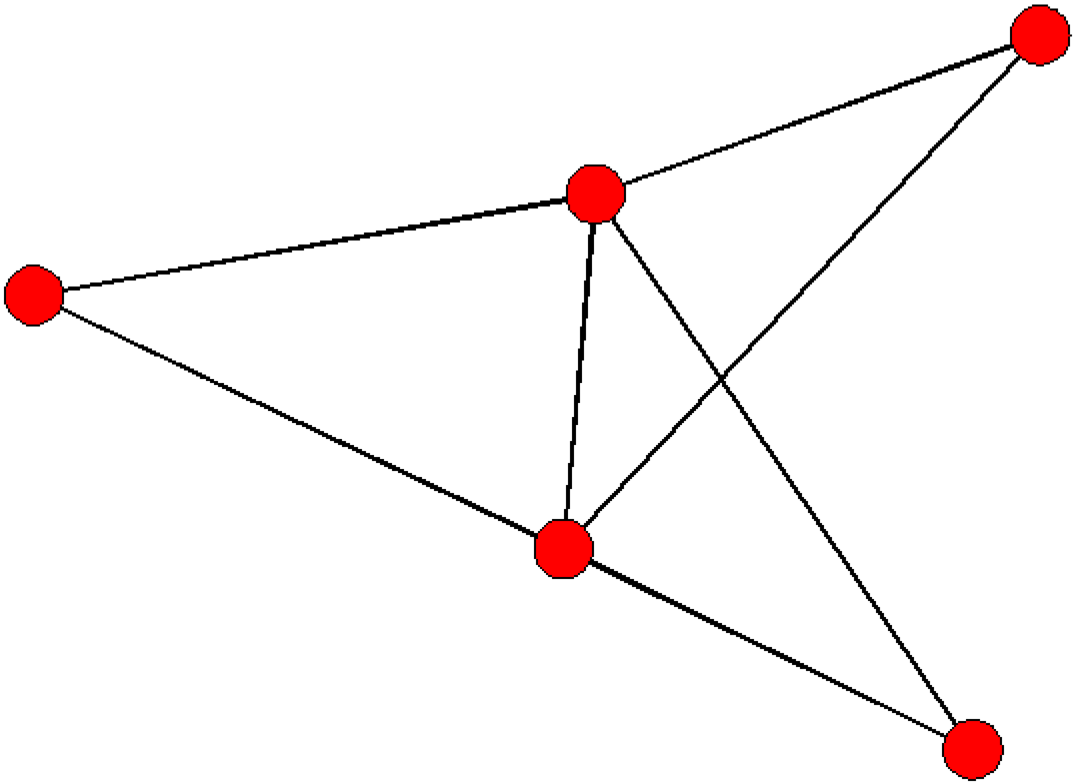
Graph theory

- **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

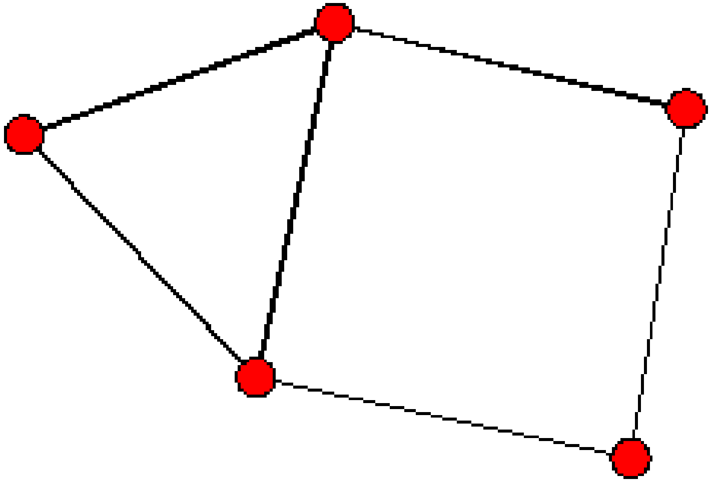
- **Step:** move along one edge.
- **Walk:** sequence of steps in network.
- **Path:** walk in which no node as well as no edge occurs more than once.
- **Geodesic:** shortest path that connects two nodes.
- **Distance** of two nodes = geodesic.
- **Diameter:** longest distance between any two nodes in graph.

Graph theory



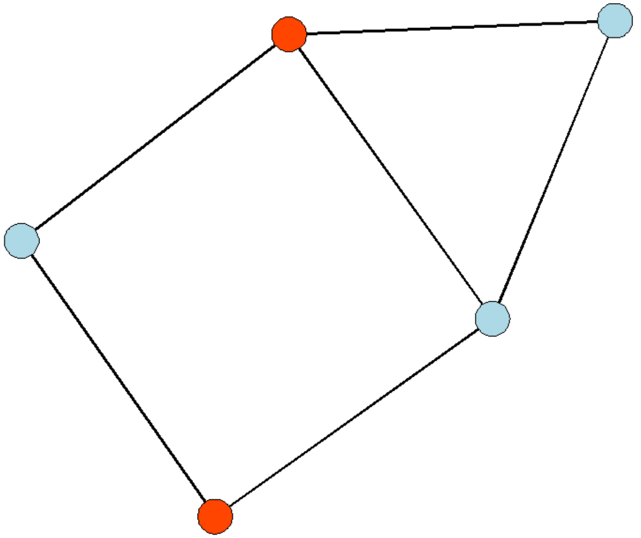
Graph theory

- A directly connected node is called **adjacent.**
- An edge linked to a node is called **incident.**
- All directly connected nodes create **neighbourhood.**



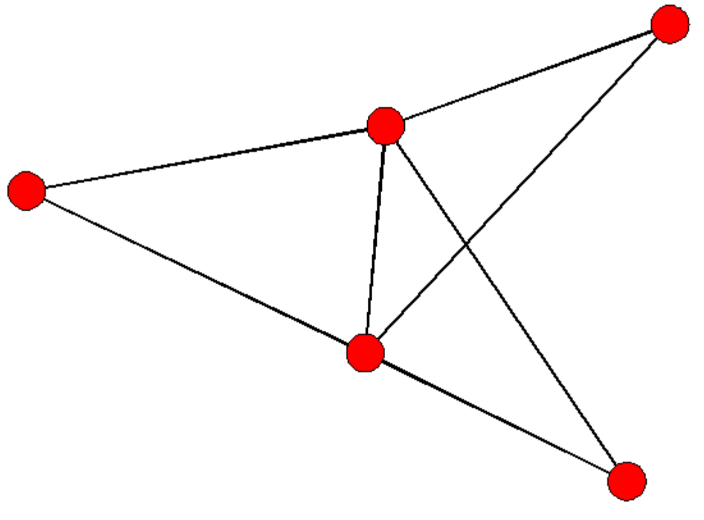
Graph theory

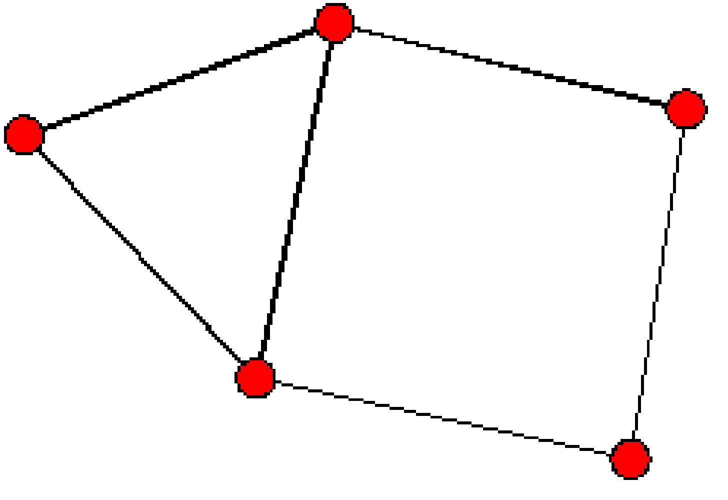
- A directly connected node is called **adjacent.**
- An edge linked to a node is called **incident.**
- All directly connected nodes create **neighbourhood.**



Graph theory

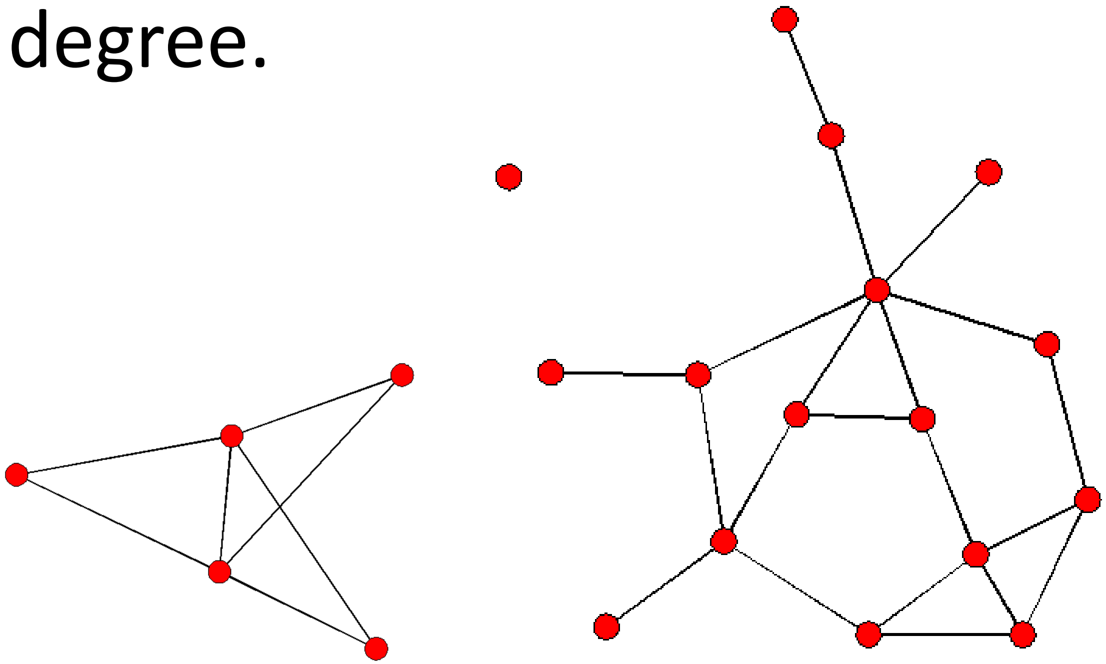
- **Subgraph** is any subset of nodes and edges of graph.
- **Component** is connected subgraph.





Graph theory

- **Reachability** is given by the existence of the path between the nodes.
- **Isolate** is a node without any connection, i.e. node with zero degree.

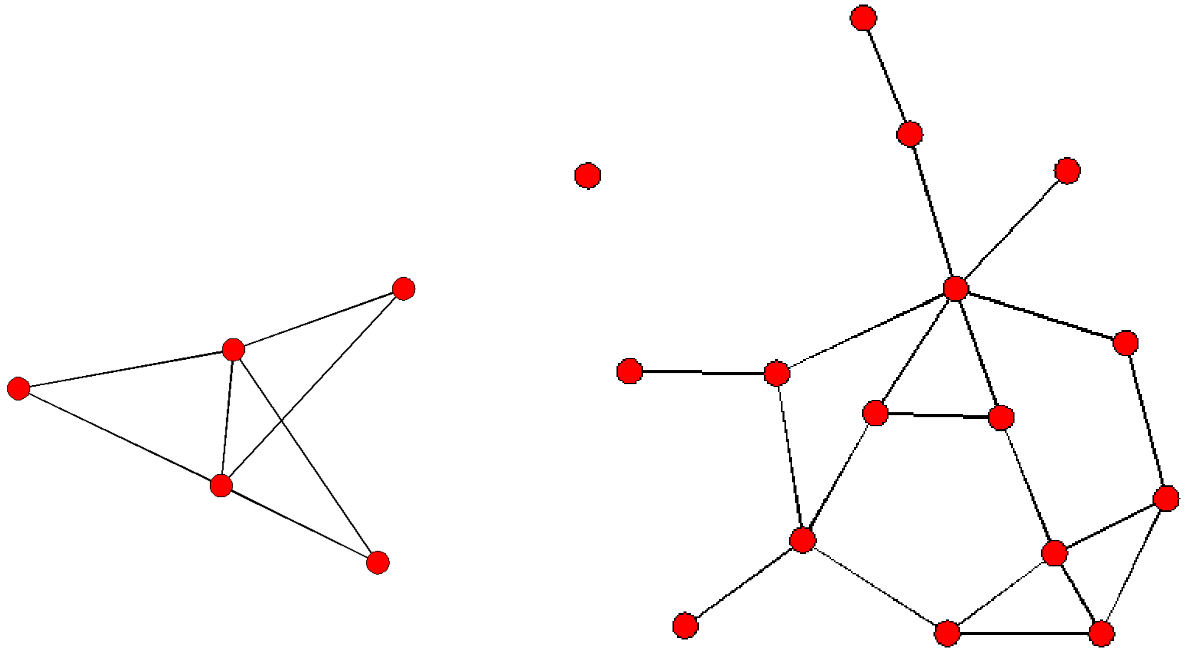


Graph theory

- **Structural hole** is a lack of connection between two nodes or subgraphs.
- **Cutpoint** is a node whose removal creates structural hole.
- **Bridge** is an edge whose removal creates structural hole.

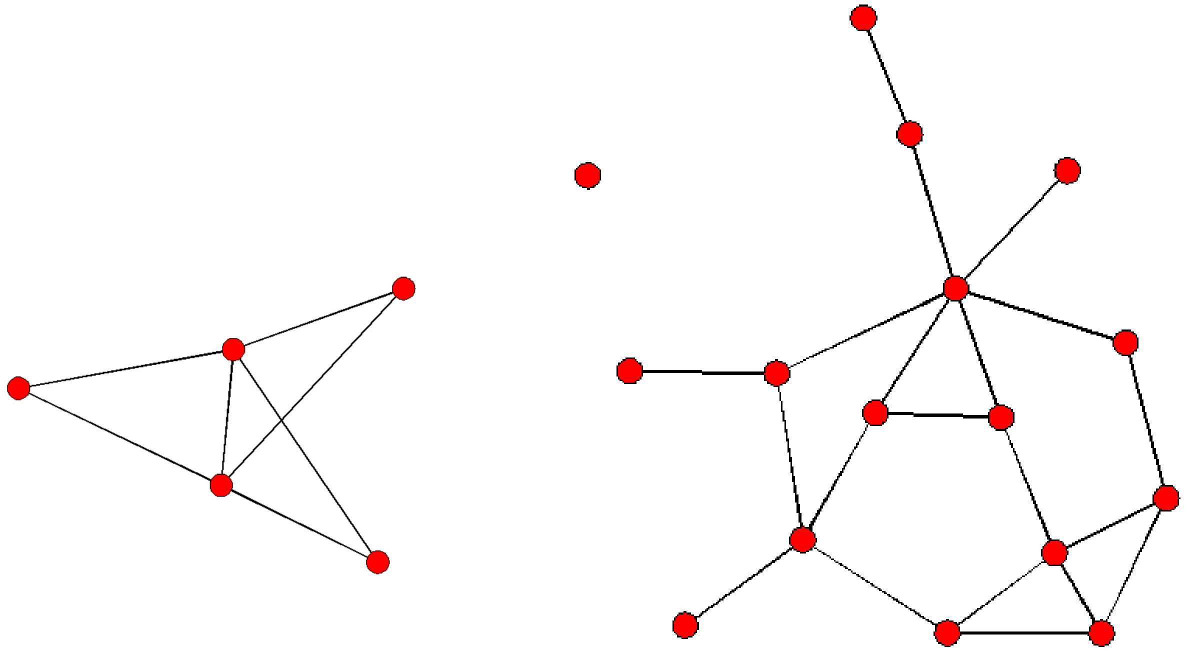
Graph theory

- **Structural hole** is a lack of connection between two nodes or subgraphs.



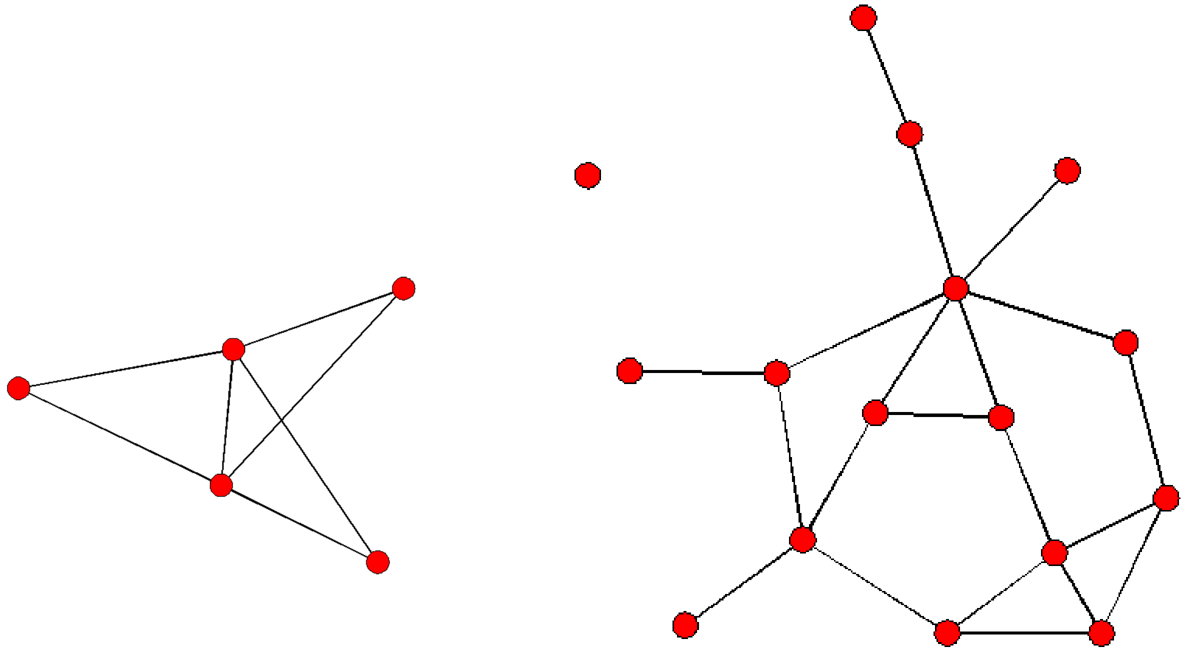
Graph theory

- **Cutpoint** is a node whose removal creates a structural hole.



Graph theory

- **Bridge** is an edge whose removal creates a structural hole.

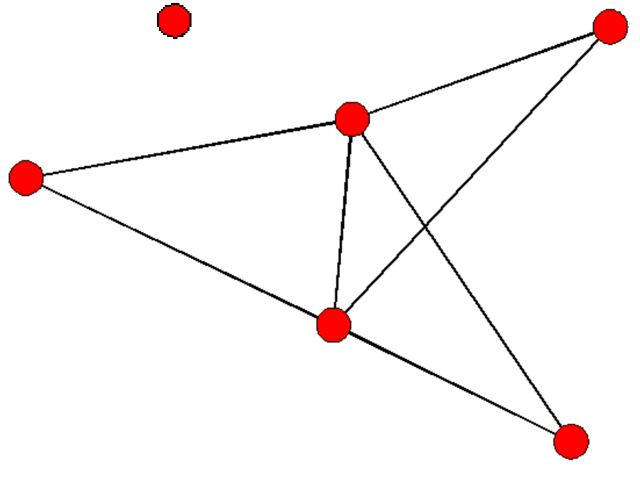


Graph theory

- **Inclusivity** is given by # of connected nodes over total # of nodes in network.
- **Density** is given by # of observed connections over total # of all possible connections in network.

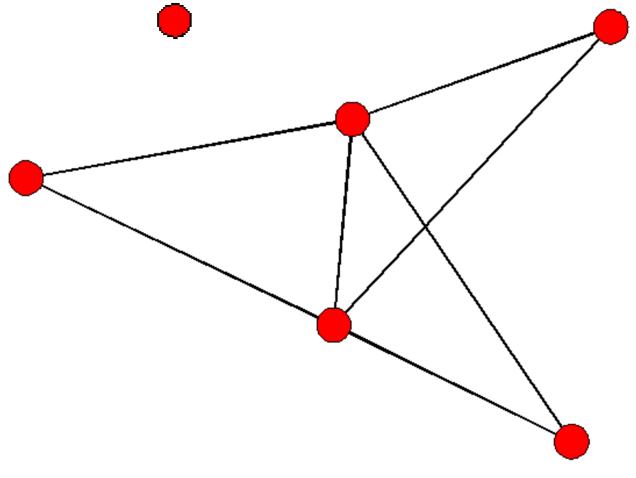
Graph theory

- **Inclusivity** is given by # of connected nodes over total # of nodes in network.



Graph theory

- **Inclusivity** is given by # of connected nodes over total # of nodes in network.
- **Inclusivity** = 5 / 6 = **0.83**



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)

- Upper case: global indicators
- Lower case: local indicators
- cd(ni) = node i degree centrality
- Cd(G) = graph G degree centralization

Graph theory

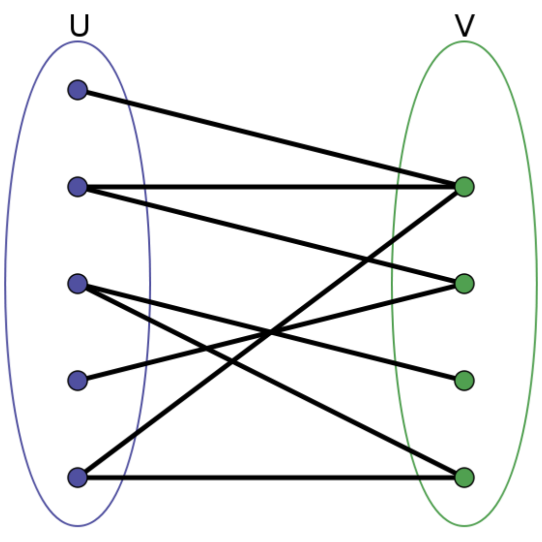
- **Density** is given by # of observed connections ($\sum e$) over total # of all possible connections in network.
- # of all possible connections in undirected network = $(N * (N – 1)) / 2$
- # of all possible connections in directed network = $(N * (N – 1))$
- Density (undirected): $\boldsymbol{\sum e / ((N * (N - 1)) / 2)}$

Graph theory

- Density (undirected): **$\sum e / ((N * (N - 1)) / 2)$**
- **Quiz:** assume that you create a network. Edge exists if you sit next or askew to each other.
  - What is inclusivity and density of this network?

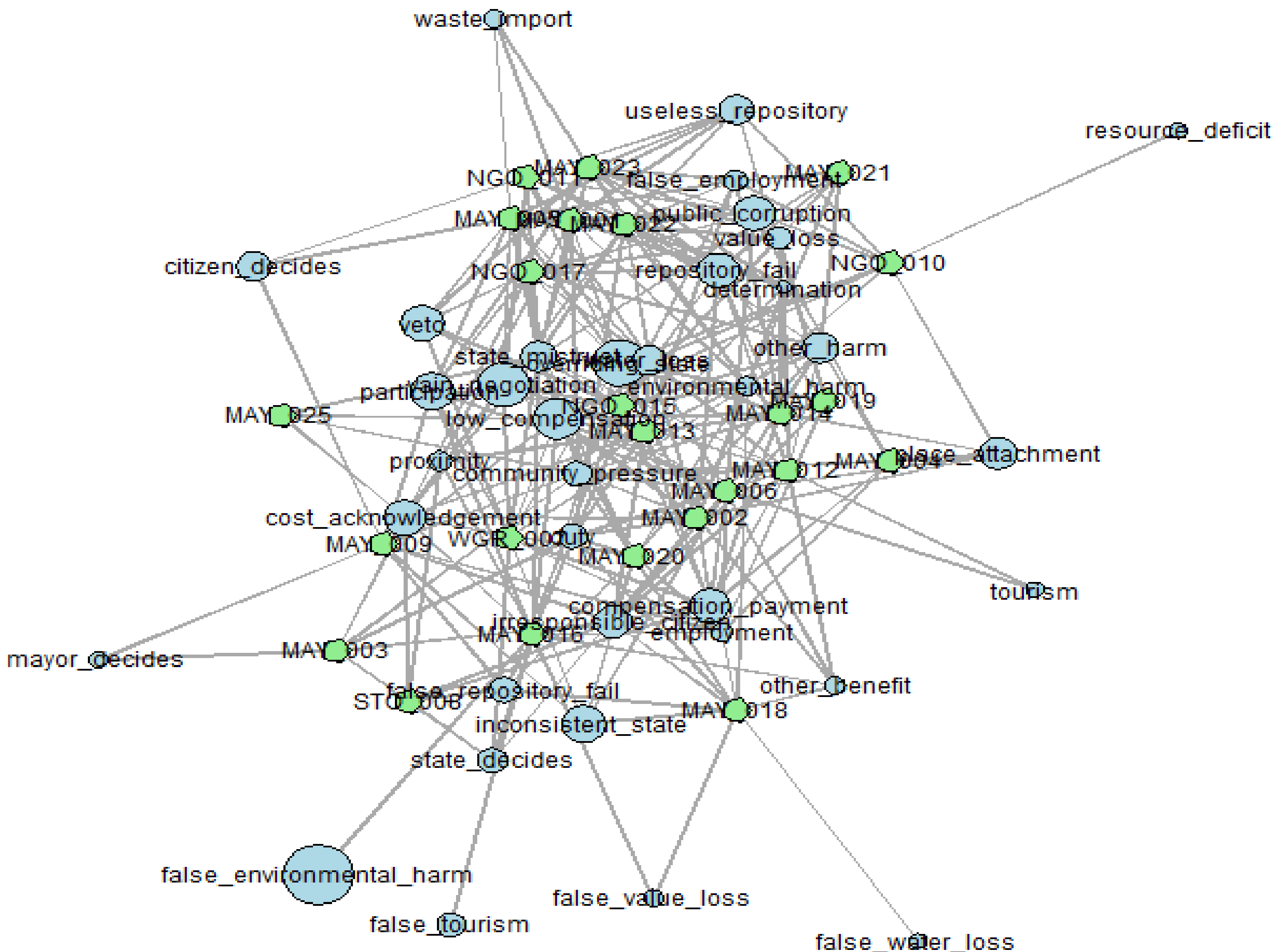
Graph theory

- **Bipartite (or two-mode) network** consists from two disjoint sets of vertices (U and V).
- The **connections** are allowed **only between** these two sets of vertices, not within them.



Graph theory

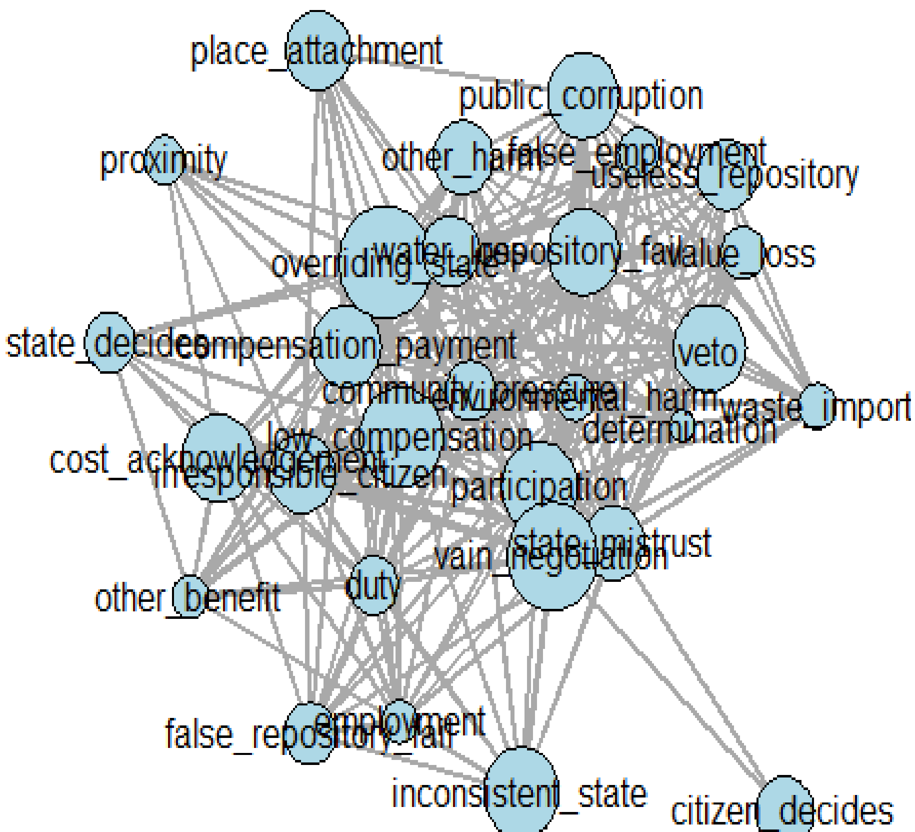
- Example: two types of nodes: (1) individuals and (2) concepts.
- The edges are between individuals and concepts.

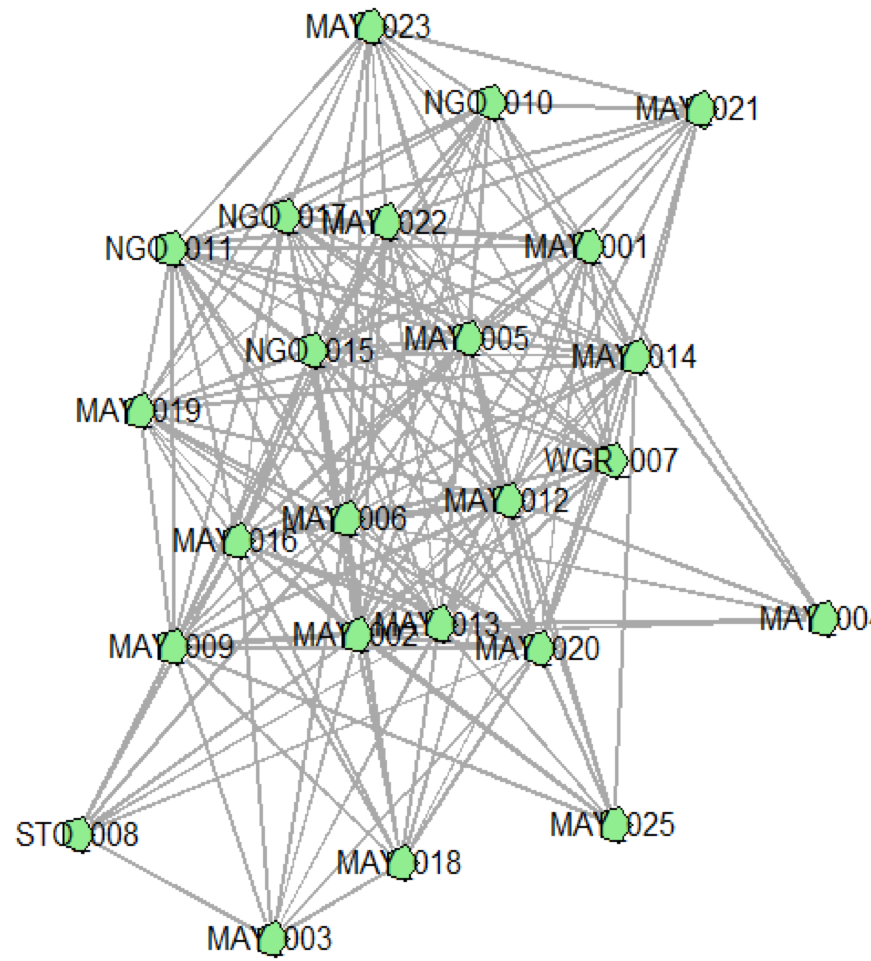

waste_import
useless_repository
resource_deficit
MAY_023
NGO_011
false_employment
MAY_021
MAY_008
MAY_022
public_corruption
value_loss
citizen_decides
NGO_017
repository_fail
NGO_010
determination
veto
other_harm
state_mistrust
overruling_state
participation
fair_negotiation
environmental_harm
MAY_025
NGO_015
MAY_019
low_compensation
MAY_014
MAY_013
proximity
place_attachment
MAY_004
community_pressure
MAY_012
MAY_006
cost_acknowledgement
MAY_002
MAY_009
duty
MAY_020
tourism
compensation_payment
irresponsible_citizen
employment
MAY_016
MAY_003
mayor_decides
other_benefit
false_repository_fail
MAY_018
inconsistent_state
state_decides
false_environmental_harm
false_value_loss
false_tourism
false_water_loss

Graph theory

- We can do **one-mode projections**; i.e. we can reconstruct one-mode networks of individuals and concepts.
- Individuals are connected if they share at least one concept.
- Concepts are connected if they are shared by at least one individual.

Graph theory

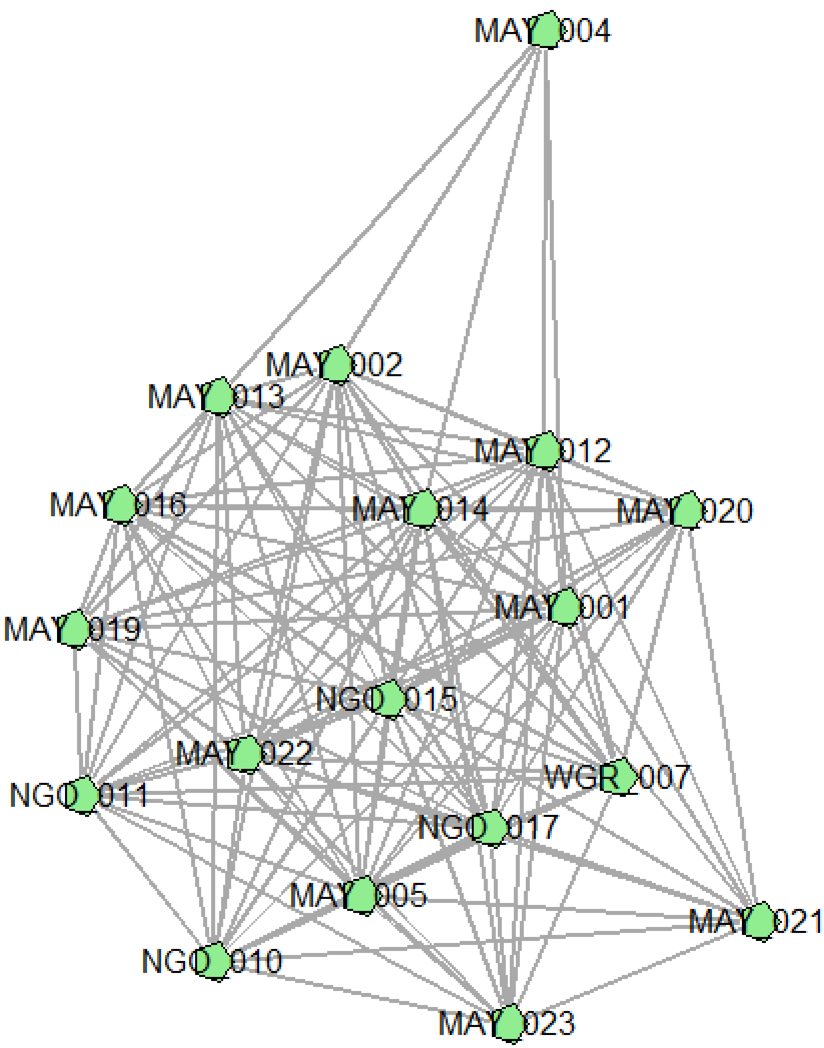


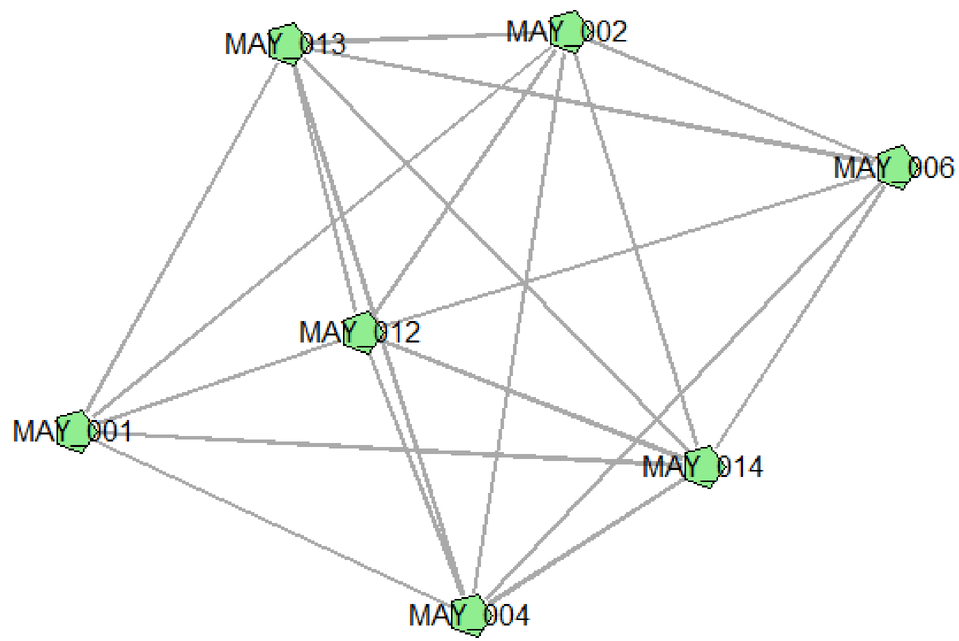


Graph theory

- **Egocentric network** is a personal network of a given individual (ego).
- Number of steps that connect a given node to ego classify the node into zones.
- **First-order zone** includes all directly connected nodes (alteri), second-order zone includes all nodes connected by two steps etc.

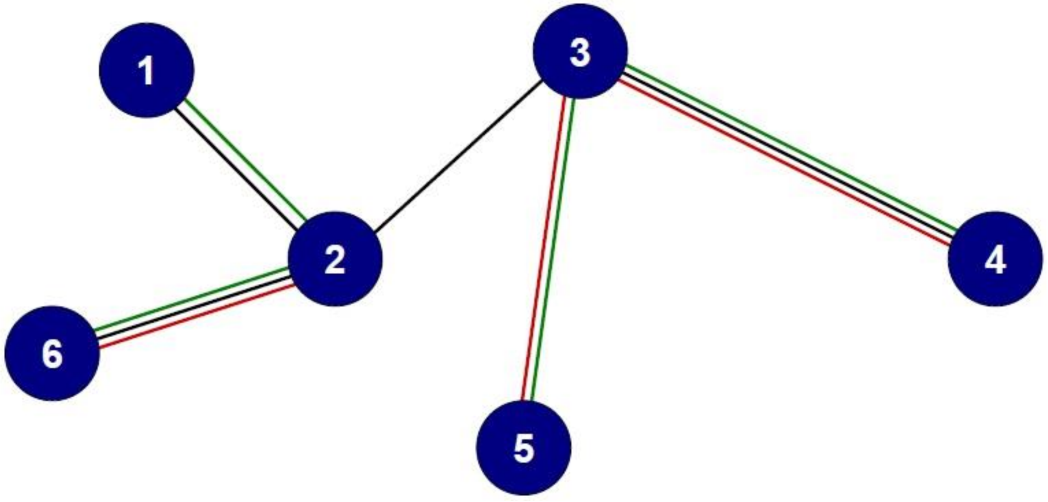
Graph theory





Graph theory

- **Multiplex network** consists from one set of vertices and more than one sets of edges.
- E.g.: imagine group of people and how they are interconnected through various social media (Facebook, Twitter, Linkedin etc.).



Data organization

- **Attributional data:** individual characteristics.
  - E.g.: age, income, education, GDP, TPES, etc.
- **Relational data:** characteristics of ties.
  - E.g.: kinship ties, trade flows, conflicts, etc.

Data organization: network borders

- **Network border** delineation usually problematic.
- Often no natural borders.
- Different strategies for border delineation:
  - nominal (e.g. all EU member states)
  - positional (e.g. all democratic states)
  - realistic (e.g. all states that present themselves as democracies)
  - relational (e.g. all states that are referred by others as democracies)
  - event-based (e.g. all states that participated in Iraqi war)

Data organization: network sampling

- Typically it is impossible to have whole population.
- Random sampling is not appropriate – why?

Data organization: network sampling

- Typically it is impossible to have population data.
- Random sampling is not appropriate – why?
- Burt's formula of information loss = (100 - k)%.
- Sampling methods:
  - Snowballing
  - Attribute based selection

Data organization: data collection

- Questionnaires / interviews
- Name generator (questionnaires)

“Looking back over the last 6 months – who are the people with whom you discussed matters important to you?” (since 1984 in the US General Social Survey)

- Observation / experiment
- Archival data

Exercise

- Define your research question.
- Define network borders and population.
- Define sampling method and data collection technique.

Data organization

- (Social) data, attributional as well as relational, are organized in **data matrices**.
- Case-by-variable matrix is a standard way of data organization in quantitative research.
- **Not appropriate for relational data.**

Case-by-variable matrix

	A	B	C	D	E
1	id	G	P	PPG	NAT
2	Wayne Gretzky	1487	2857	1.92131809	CAN
3	Mark Messier	1756	1887	1.074601367	CAN
4	Gordie Howe	1767	1850	1.046972269	CAN
5	Ron Francis	1731	1798	1.03870595	CAN
6	Marcel Dionne	1348	1771	1.31379822	CAN
7	Steve Yzerman	1514	1755	1.159180978	CAN
8	Mario Lemieux	915	1723	1.883060109	CAN
9	Jaromir Jagr	1346	1653	1.22808321	CZ
10	Joe Sakic	1378	1641	1.190856313	CAN
11	Phil Esposito	1282	1590	1.24024961	CAN
12	Ray Bourque	1612	1579	0.979528536	CAN
13	Mark Recchi	1652	1533	0.927966102	CAN
14	Paul Coffey	1409	1531	1.086586231	CAN
15	Stan Mikita	1394	1467	1.052367288	CAN
16	Bryan Trottier	1279	1425	1.11415168	US
17	Adam Oates	1337	1420	1.062079282	CAN
18	Doug Gilmour	1474	1414	0.959294437	CAN
19	Dale Hawerchuk	1188	1409	1.186026936	CAN
20	Teemu Selanne	1341	1406	1.04847129	FIN

Adjacency (case-by-case) matrix

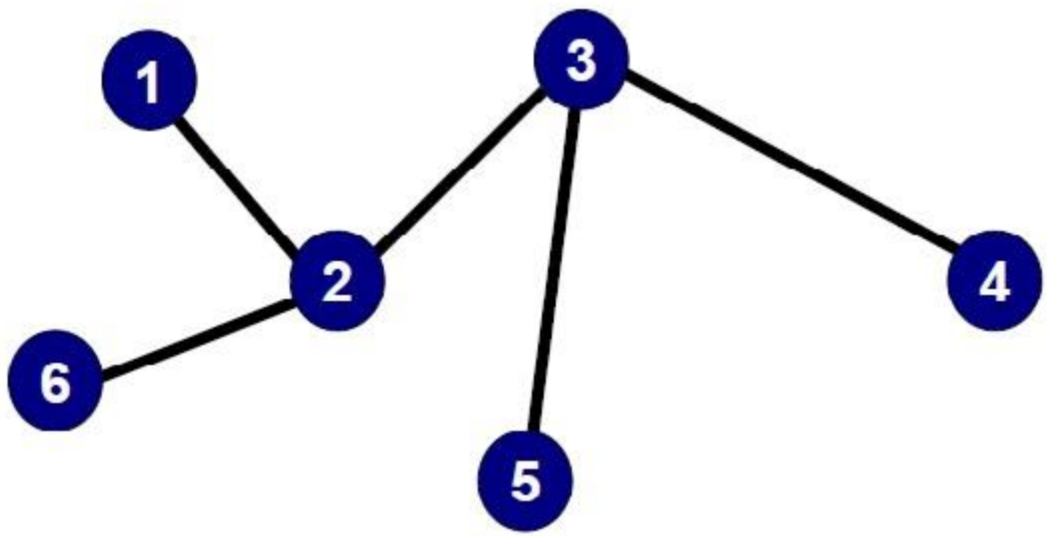
	A	B	C	D	E	F	G	H	I	J
1		Wayne Gr	Mark Mes	Gordie Ho	Ron Franc	Marcel Di	Steve Yze	Mario Len	Jaromir Ja	Joe Sakic
2	Wayne Gr	0	9	0	0	0	0	0	0	0
3	Mark Mes	9	0	0	0	0	0	0	1	0
4	Gordie Ho	0	0	0	0	0	0	0	0	0
5	Ron Franc	0	0	0	0	0	0	7	7	0
6	Marcel Di	0	0	0	0	0	0	0	0	0
7	Steve Yze	0	0	0	0	0	0	0	0	0
8	Mario Len	0	0	0	7	0	0	0	8	0
9	Jaromir Ja	0	1	0	7	0	0	8	0	0
10	Joe Sakic	0	0	0	0	0	0	0	0	0
11	Phil Espos	0	0	0	0	0	0	0	0	0
12	Ray Bourq	0	0	0	0	0	0	0	0	2
13	Mark Recc	0	0	0	2	0	0	4	2	0
14	Paul Coffe	8	6	0	2	0	3	3	2	0
15	Stan Mikit	0	0	0	0	0	0	0	0	0
16	Bryan Trot	0	0	0	4	0	0	4	4	0
17	Adam Oat	0	0	0	0	0	4	0	0	0
18	Doug Gilm	0	0	0	0	0	0	0	0	0
19	Dale Hawe	0	0	0	0	0	0	0	0	0

Data organization

- **Adjacency matrix** represents which nodes are adjacent to other nodes.
- **Incidence matrix** represents the relations between two classes of nodes.
  – Rows represent one class of nodes.
  – Columns represent second class of nodes.

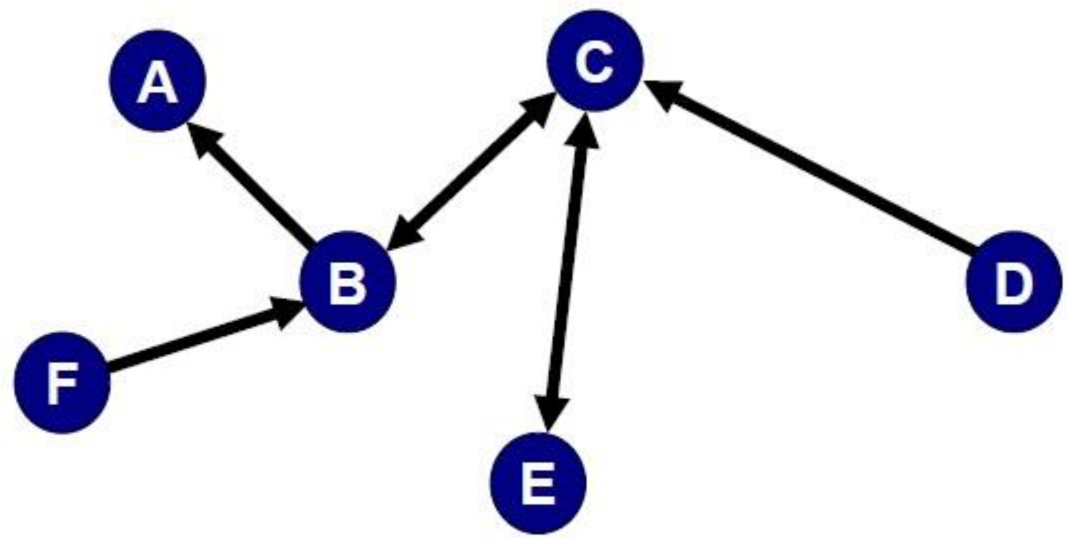
Undirected one-mode network

	1	2	3	4	5	6
1		1				
2	1		1			1
3		1		1	1	
4			1			
5			1			
6		1				



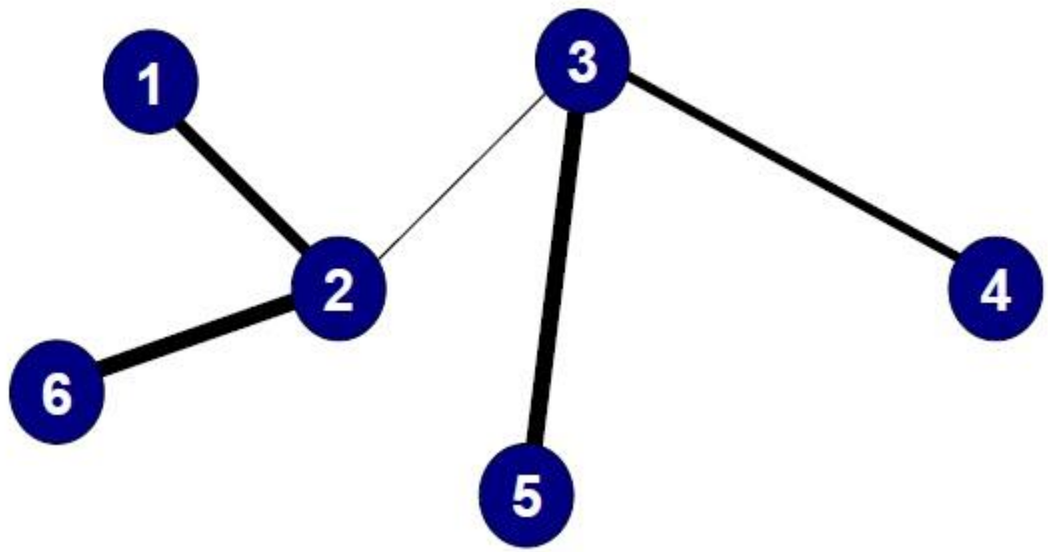
Directed one-mode network

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



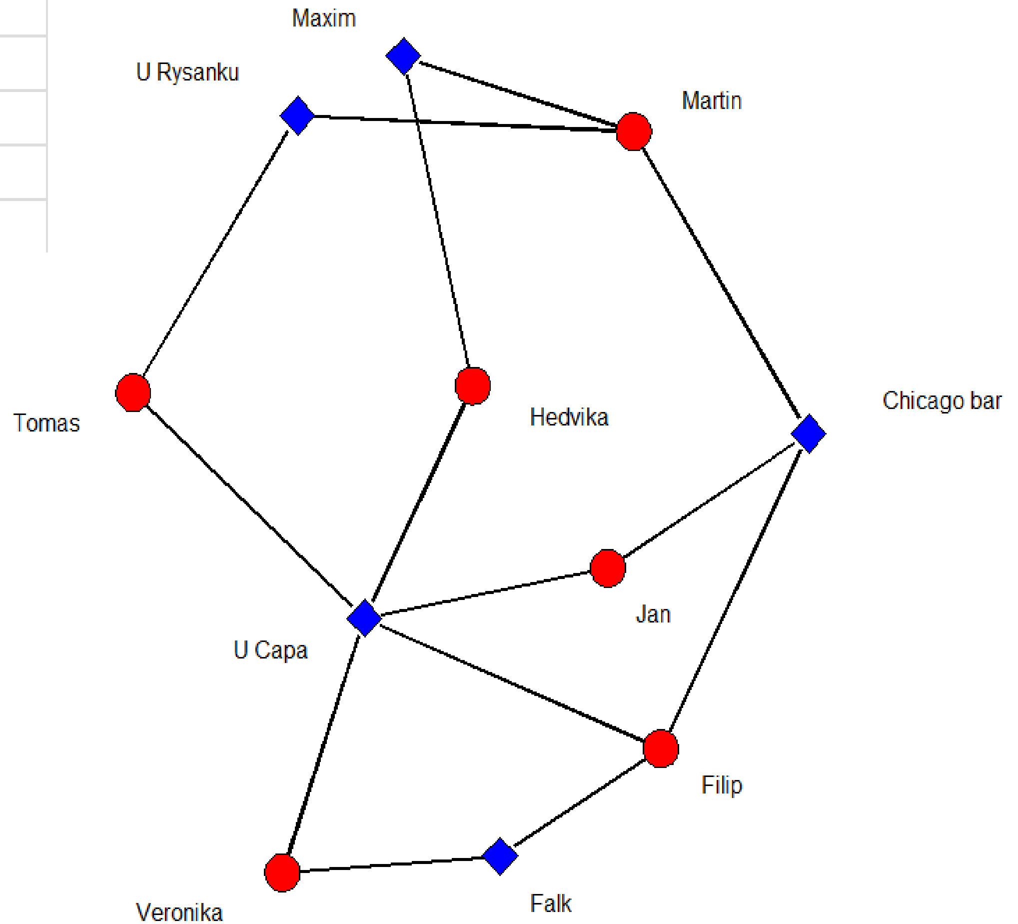
Undirected weighted one-mode network

	1	2	3	4	5	6
1		5				
2	5		1			7
3		1		4	7	
4			4			
5			7			
6		7				



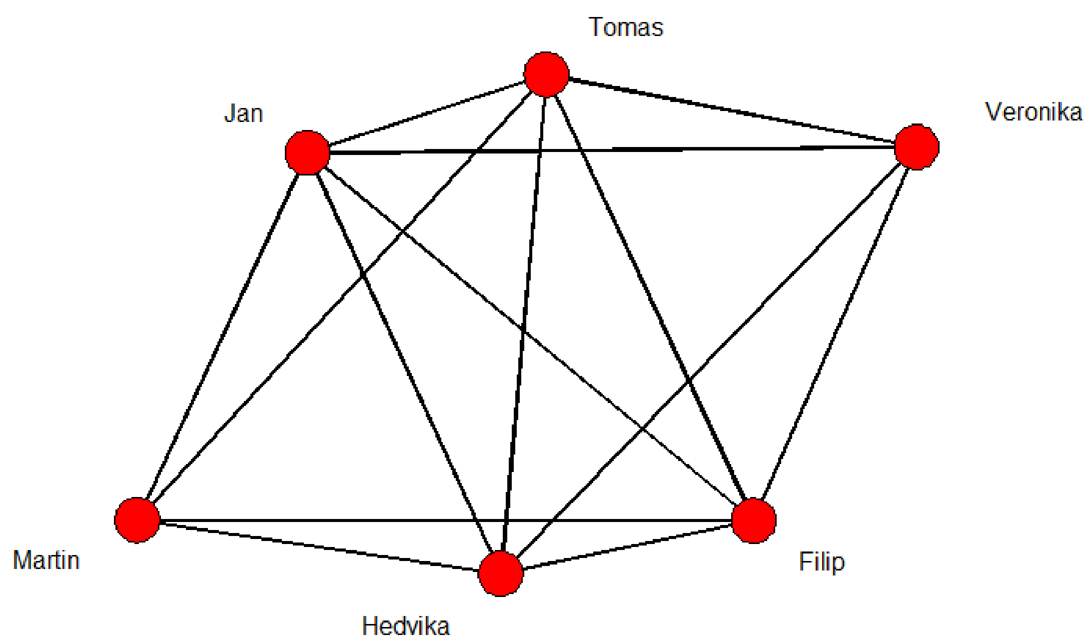
Incidence (case-by-event) matrix

	row.names	U Capa	Chicago bar	U Rysanku	Maxim	Falk
1	Filip	1	1	0	0	1
2	Hedvika	1	0	0	1	0
3	Jan	1	1	0	0	0
4	Veronika	1	0	0	0	1
5	Tomas	1	0	1	0	0
6	Martin	0	1	1	1	0



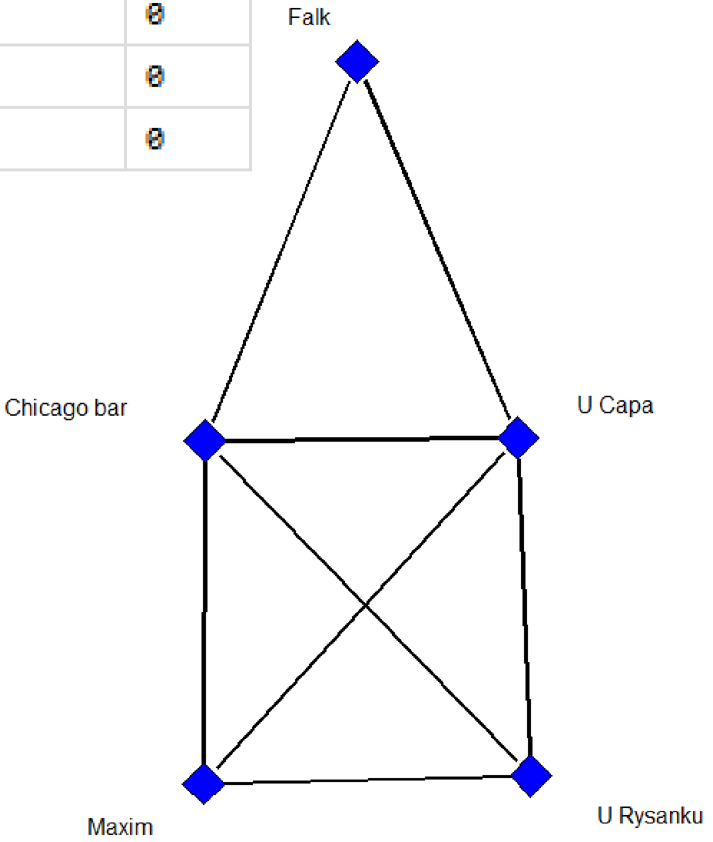
Adjacency (case-by-case) matrix

	row.names	Filip	Hedvika	Jan	Veronika	Tomas	Martin
1	Filip	0	1	2	2	1	1
2	Hedvika	1	0	1	1	1	1
3	Jan	2	1	0	1	1	1
4	Veronika	2	1	1	0	1	0
5	Tomas	1	1	1	1	0	1
6	Martin	1	1	1	0	1	0



Adjacency (event-by-event) matrix

	row.names	U Capa	Chicago bar	U Rysanku	Maxim	Falk
1	U Capa	0	2	1	1	2
2	Chicago bar	2	0	1	1	1
3	U Rysanku	1	1	0	1	0
4	Maxim	1	1	1	0	0
5	Falk	2	1	0	0	0



Matrix operations: one-mode projection

- We get one-mode projection (adjacency) matrix by multiplying incidence matrix by its **transposition**.
  - Transposed matrix: rows turned to columns and vice versa.
- For **cases (rows)** we put transpose on the second place.
  - matrix %*% t(matrix)
- For **events (columns)** we put transpose on the first place.
  - t(matrix) %*% matrix

Matrix transposition

- Incidence matrix

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

- Transposition

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

Matrix by matrix multiplication (cases)

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

%*%

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

Dot product: first we take first row and first column

(1, 0, 1, 1) and (1, 0, 1, 1), second we multiply corresponding elements and sum up the products.

(1, 0, 1, 1) * (1, 0, 1, 1) = 1*1 + 0*0 + 1*1 + 1*1 = **3**

Matrix by matrix multiplication (cases)

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

%*%

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

=

3	2	2
2	2	1
2	1	3

Matrix by matrix multiplication (events)

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

%*%

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

Dot product: first we take first row and first column

(1, 0, 1) and (1, 0, 1), second we multiply corresponding elements and sum up the products.

(1, 0, 1) * (1, 0, 1) = 1*1 + 0*0 + 1*1 = **2**

Matrix by matrix multiplication (events)

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

%*%

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

=

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

Exercise

- Do one-mode projections of incidence matrix:

	Jan	Petr	Hedvika
Introduction	1	0	1
Methodology	1	1	0

Mini-case study

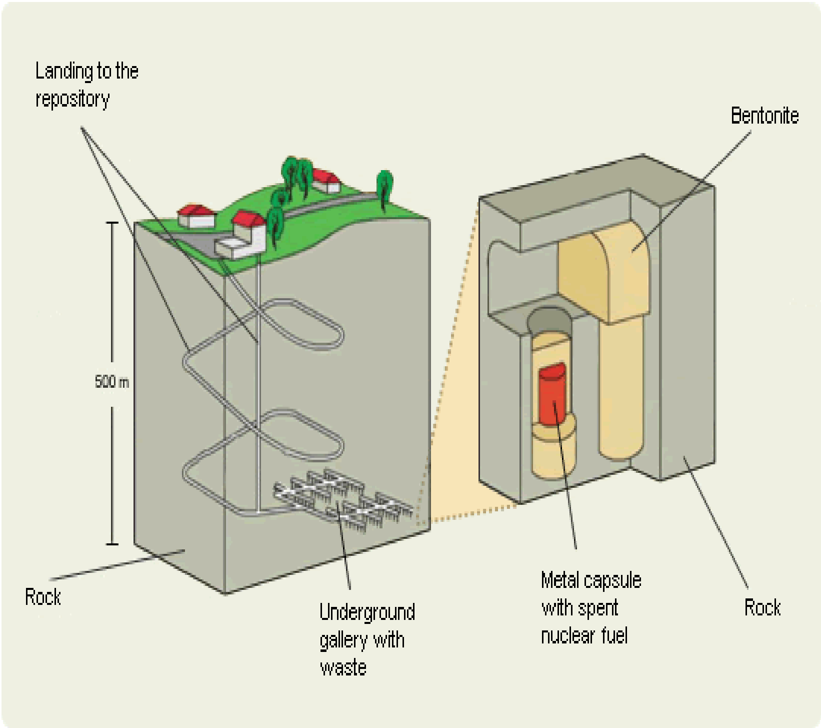
- **Deep geological repository** designed to contain nuclear waste for hundreds of thousands of years.

- There are 7 pre-selected (candidate) localities in the Czech republic.

- Since the beginning there are repeated occurrences of local opposition.

- **Research objective:** to map how the issue is framed by the local opposition / acceptance opinion leaders.

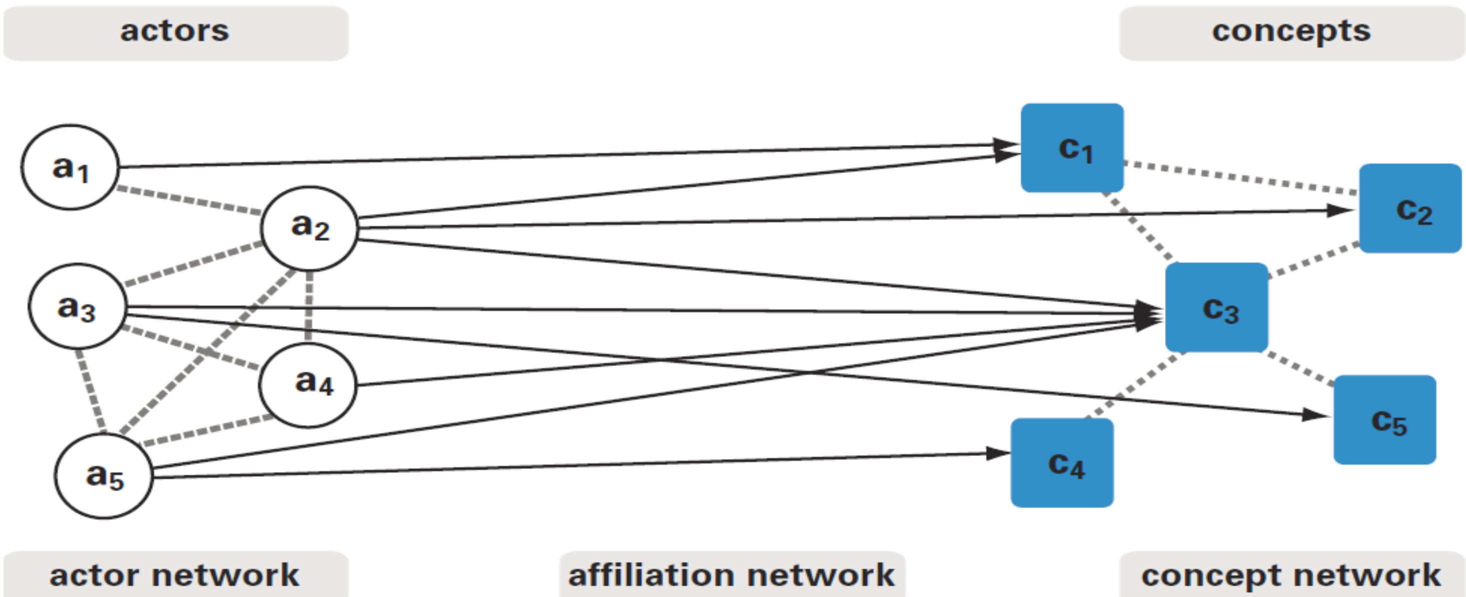
Čertovka
lokalita na území obcí Blatno a Lubenec v Severočeském kraji a na území obcí Tis u Blatna a Žihle v Plzeňském kraji
Hrádek
lokalita na území obcí Rohozná, Dolní Cerekev, Cejle, Hojkov, Milíčov a městyse Nový Rychnov v kraji Vysočina
PRAHA
Březový potok
lokalita na území obcí Pačejov, Kvášňovice, Olšany, Maňovice, Chanovice a Velký Bor v Plzeňském kraji
Kraví Hora
lokalita na území obcí Bukov, Věžná, Střítež, Moravecké Pavlovice, Drahonín, Olší a Sejřek spadající pod kraj Vysočina a Jihomoravský kraj
Magdaléna
lokalita se nachází na území obcí Jistebnice, Nadějkov a Božetice v Jihočeském kraji
Čihadlo
lokalita se nachází na území města Deštná a obcí Světce, Lodhéřov a Pluhův Žďár v Jihočeském kraji
Horka
lokalita na území obcí Hodov, Rohy, Oslavička, Budišov, Náramč, Vlčatín, Osová, Rudíkov a Oslavice v kraji Vysočina
zdroj: SÚRAO
ČTK



Landing to the repository
Bentonite
500 m
Rock
Underground gallery with waste
Metal capsule with spent nuclear fuel
Rock

Discourse network

- A **bipartite network** that consists of actors and concepts.



Haunss, Dietz & Nullmeier 2013: 13

Frame

- **Frame** defined as *shared interpretative scheme through which actors understand and promote a particular version of reality* (see e.g. Benford & Snow 2000).

- Actors – strategically – use frames to emphasize or suppress particular aspects of the contested issue.

- Intuition: group of nodes (**cluster**) in the concept network which are in a similar position to the rest of the network.

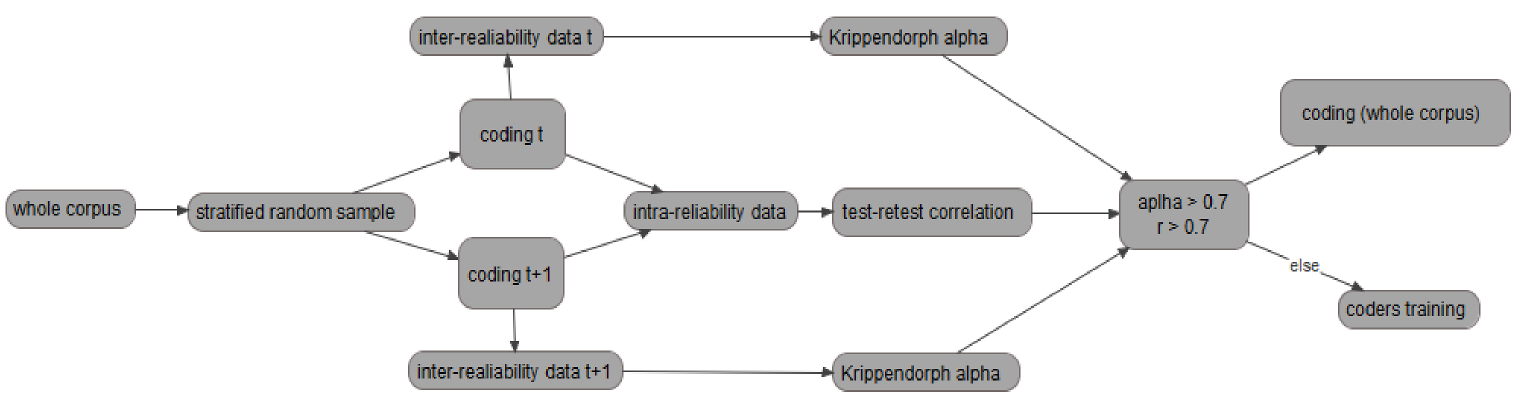
(Discourse / frame) coalition

- **Frame coalition** is understood as a "...*group of actors that share social construct* [frame]*.*" (Hajer 1995: 43)

- Intuition: densely connected segment (**community**) of the actor network.

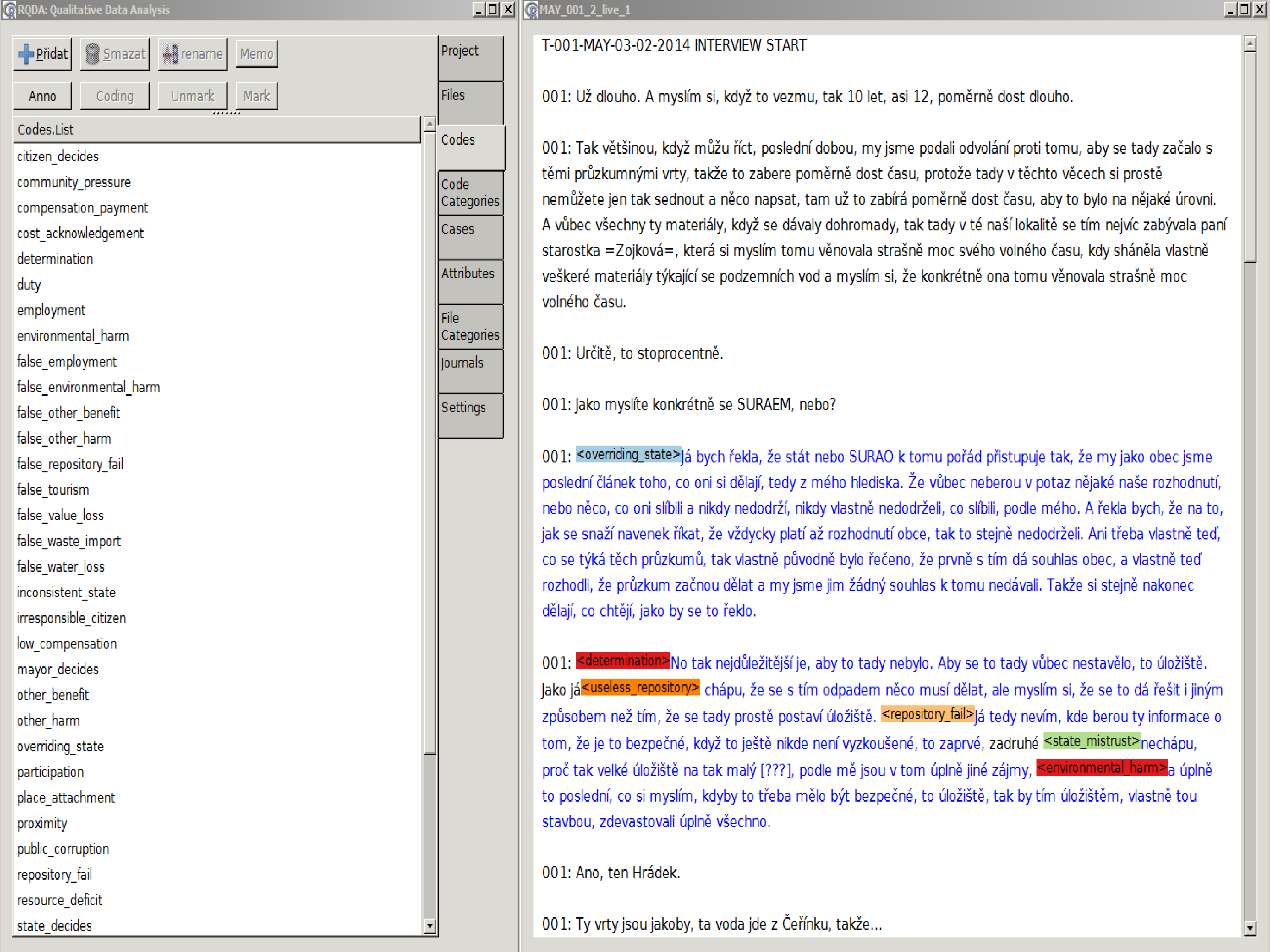
Data and coding

- **Data:**
  - 47 semi-standardized interviews (mayors, activists and state officials).
- **Coding:**
  - Corpus coded by 2 independent coders with Krippendorff's alpha = 0.81 (inter-rater), r = 0.79 (intra-rater reliability).
  - The coded corpus contains 634 observations (38 codes).



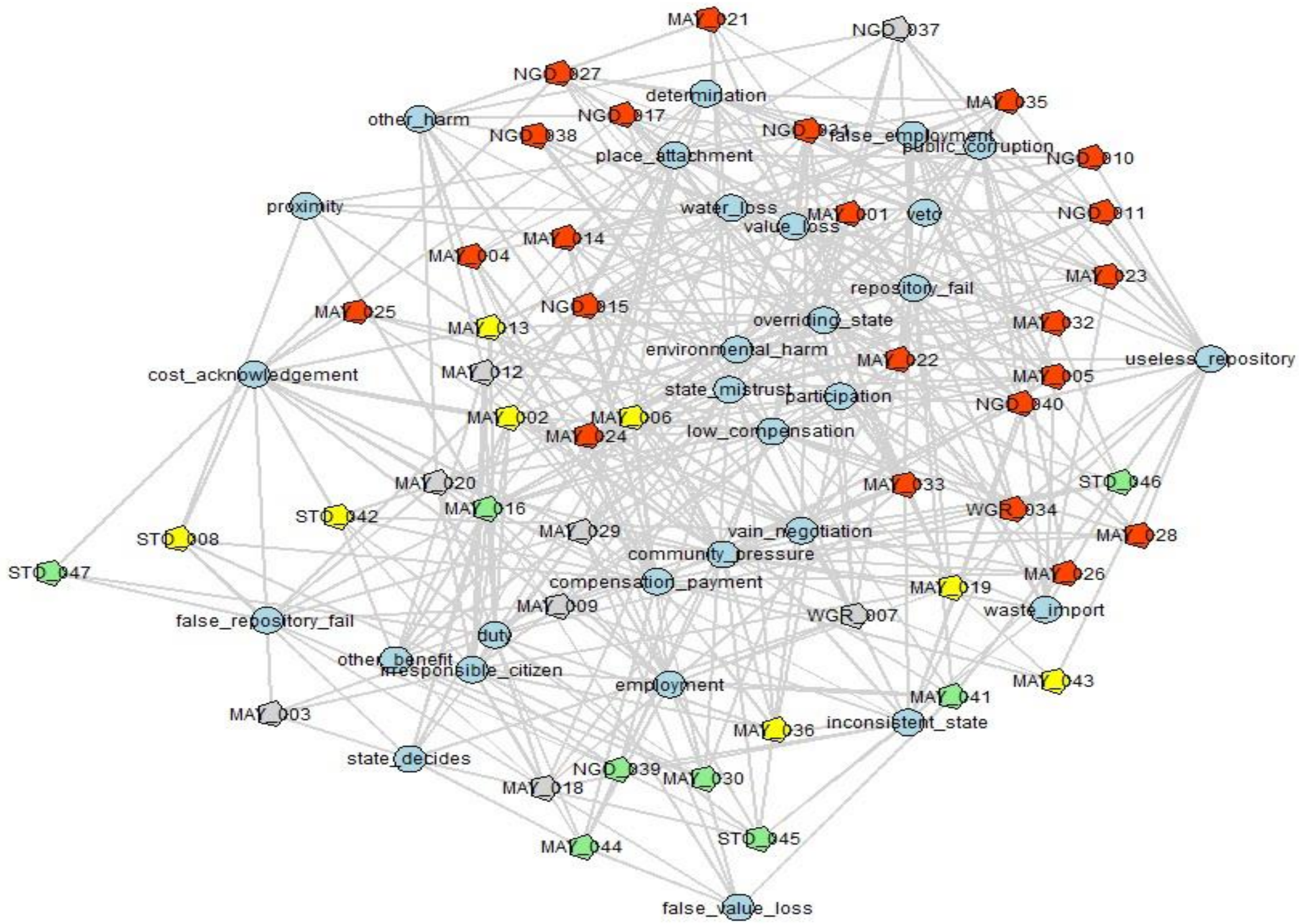
RQDA package

- Corpus has been coded in RQDA package (Ronggui Huang 2014).
- R package for qualitative data analysis with GUI.
- Provides basic functions of CAQDAS.


RQDA: Qualitative Data Analysis
Přidat
Smazat
rename
Memo
Anno
Coding
Unmark
Mark
Codes.List
citizen_decides
community_pressure
compensation_payment
cost_acknowledgement
determination
duty
employment
environmental_harm
false_employment
false_environmental_harm
false_other_benefit
false_other_harm
false_repository_fail
false_tourism
false_value_loss
false_waste_import
false_water_loss
inconsistent_state
irresponsible_citizen
low_compensation
mayor_decides
other_benefit
other_harm
overriding_state
participation
place_attachment
proximity
public_corruption
repository_fail
resource_deficit
state_decides
Project
Files
Codes
Code Categories
Cases
Attributes
File Categories
Journals
Settings
MAY_001_2_live_1
T-001-MAY-03-02-2014 INTERVIEW START
001: Už dlouho. A myslím si, když to vezmu, tak 10 let, asi 12, poměrně dost dlouho.
001: Tak většinou, když můžu říct, poslední dobou, my jsme podali odvolání proti tomu, aby se tady začalo s těmi průzkumnými vrty, takže to zabere poměrně dost času, protože tady v těchto věcech si prostě nemůžete jen tak sednout a něco napsat, tam už to zabírá poměrně dost času, aby to bylo na nějaké úrovni. A vůbec všechny ty materiály, když se dávaly dohromady, tak tady v té naší lokalitě se tím nejvíc zabývala paní starostka =Zojková=, která si myslím tomu věnovala strašně moc svého volného času, kdy sháněla vlastně veškeré materiály týkající se podzemních vod a myslím si, že konkrétně ona tomu věnovala strašně moc volného času.
001: Určitě, to stoprocentně.
001: Jako myslíte konkrétně se SURAEM, nebo?
001: <overriding_state>Já bych řekla, že stát nebo SURAO k tomu pořád přistupuje tak, že my jako obec jsme poslední článek toho, co oni si dělají, tedy z mého hlediska. Že vůbec neberou v potaz nějaké naše rozhodnutí, nebo něco, co oni slíbili a nikdy nedodrží, nikdy vlastně nedodrželi, co slíbili, podle mého. A řekla bych, že na to, jak se snaží navenek říkat, že vždycky platí až rozhodnutí obce, tak to stejně nedodrželi. Ani třeba vlastně teď, co se týká těch průzkumů, tak vlastně původně bylo řečeno, že prvně s tím dá souhlas obec, a vlastně teď rozhodli, že průzkum začnou dělat a my jsme jim žádný souhlas k tomu nedávali. Takže si stejně nakonec dělají, co chtějí, jako by se to řeklo.
001: <determination>No tak nejdůležitější je, aby to tady nebylo. Aby se to tady vůbec nestavělo, to úložiště. Jako já<useless_repository> chápu, že se s tím odpadem něco musí dělat, ale myslím si, že se to dá řešit i jiným způsobem než tím, že se tady prostě postaví úložiště. <repository_fail>Já tedy nevím, kde berou ty informace o tom, že je to bezpečné, když to ještě nikde není vyzkoušené, to zaprvé, zadruhé <state_mistrust>nechápu, proč tak velké úložiště na tak malý [???], podle mě jsou v tom úplně jiné zájmy, <environmental_harm>a úplně to poslední, co si myslím, kdyby to třeba mělo být bezpečné, to úložiště, tak by tím úložištěm, vlastně tou stavbou, zdevastovali úplně všechno.
001: Ano, ten Hrádek.
001: Ty vrty jsou jakoby, ta voda jde z Čeřínku, takže...

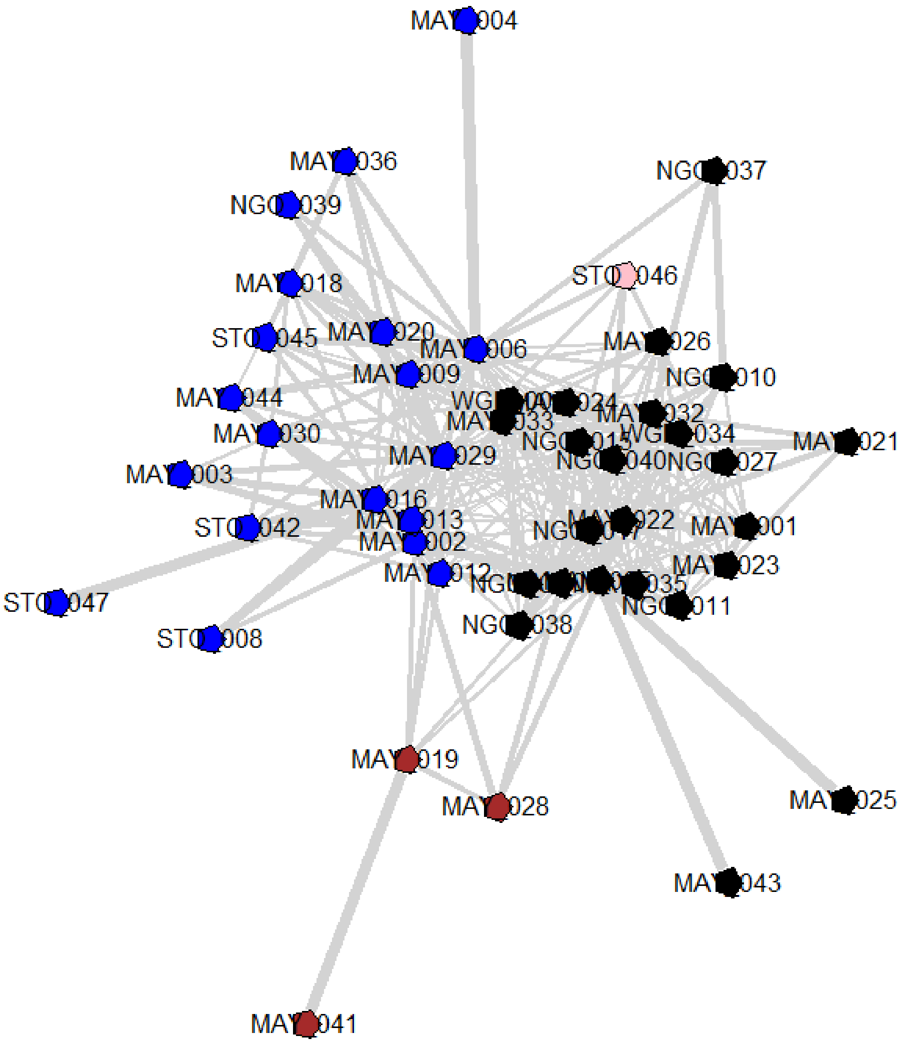
Incidence matrix

	concept 1	. . . . . .	concept *j*
actor 1 (interview)	*ij* cell indicates how many times actor *i* uses concept *j*	0	3
. . . .	. . . .	. . . .	. . . .
actor *i*	0	2	1

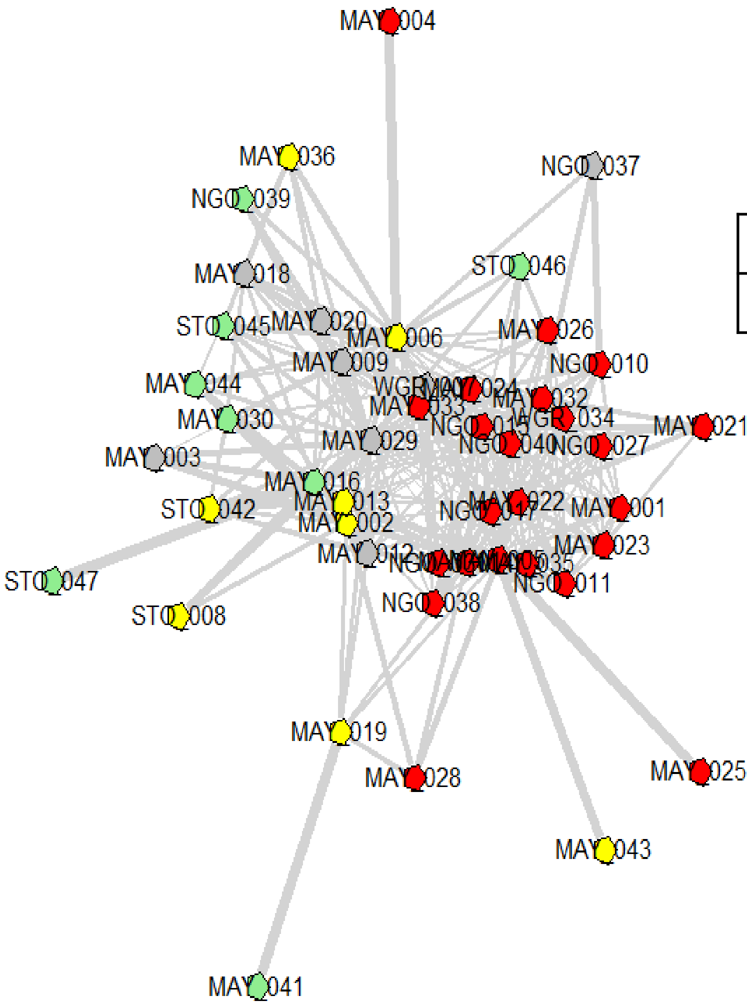

MAY_021
NGO_037
NGO_027
determination
MAY_035
other_harm
NGO_017
NGO_038
NGO_031
false_employment
public_corruption
NGO_010
place_attachment
proximity
water_loss
MAY_001
veto
NGO_011
value_loss
MAY_014
MAY_004
MAY_023
repository_fail
NGO_015
MAY_025
MAY_013
overriding_state
MAY_032
environmental_harm
MAY_022
useless_repository
cost_acknowledgement
MAY_012
MAY_005
state_mistrust
participation
NGO_040
MAY_002
MAY_006
MAY_024
low_compensation
MAY_020
MAY_033
STO_046
MAY_016
WGR_034
STO_042
STO_008
MAY_029
vain_negotiation
MAY_028
community_pressure
STO_047
MAY_026
compensation_payment
MAY_019
MAY_009
false_repository_fail
WGR_007
waste_import
duty
other_benefit
irresponsible_citizen
employment
MAY_043
MAY_041
MAY_003
MAY_036
inconsistent_state
state_decides
NGO_039
MAY_030
MAY_018
STO_045
MAY_044
false_value_loss

Network communities

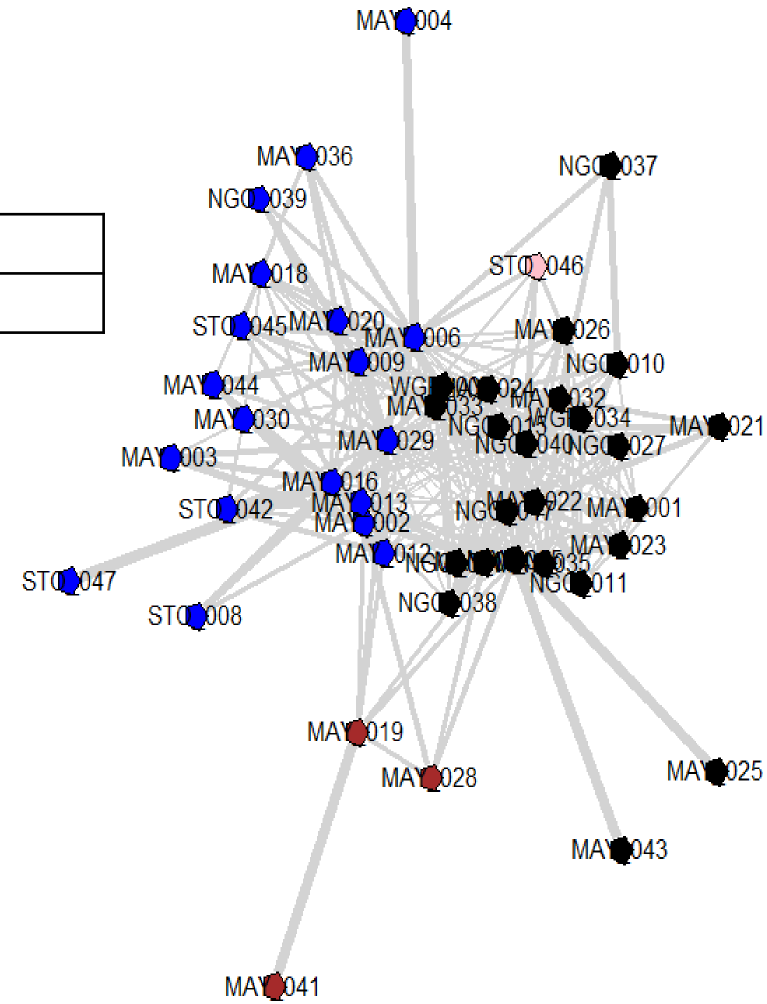
- **Network community** is a segment of the network created by a set of nodes (members) that are more densely connected internally than with the non-members of the community.

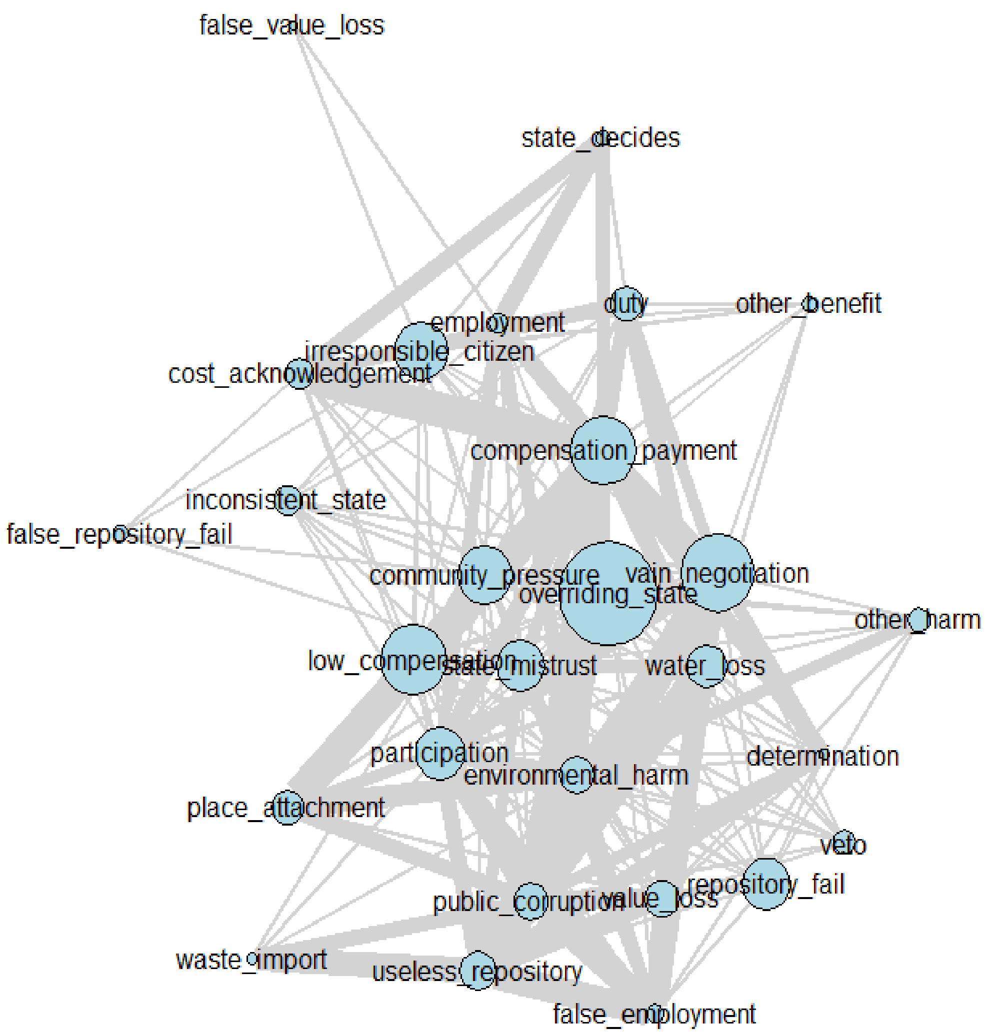


community 1	14 MAYs, 1 NGO, 4 STOs
community 2	15 MAYs, 9 NGOs
community 3	3 MAYs
"community" 4	STO_046
$p \leq 0.001$	

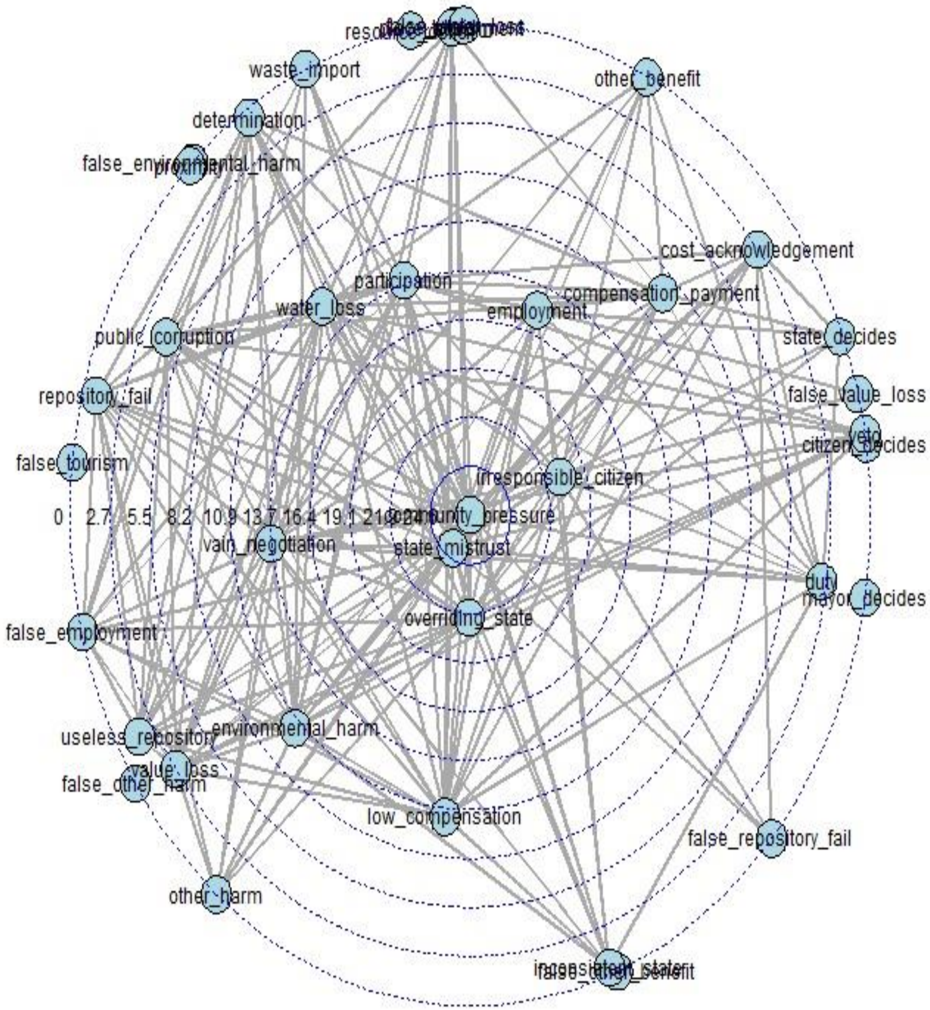


correlation	0.37
p-value	0.013

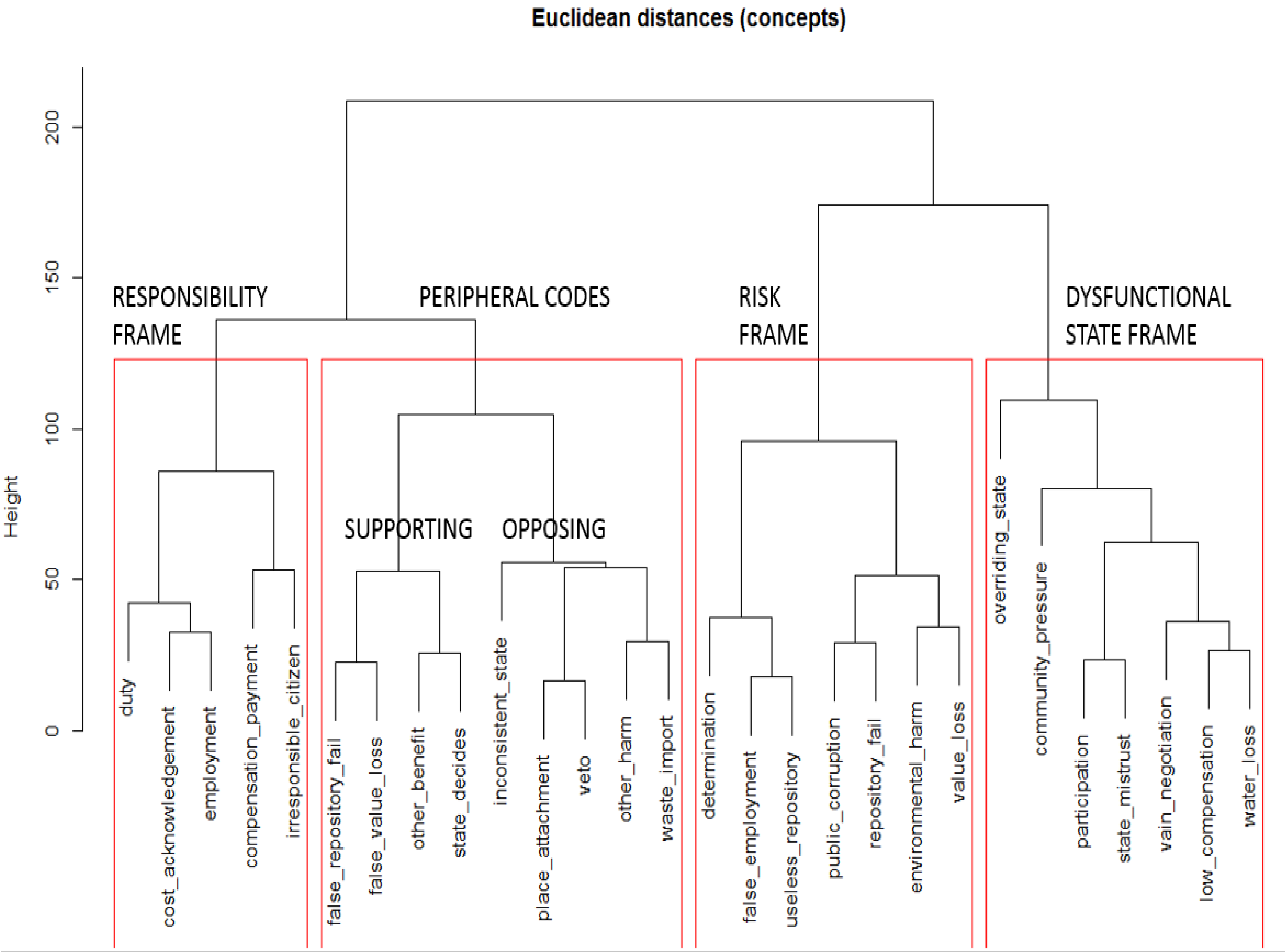




density	0.31
deg. centralization	0.43
bet. centralization	0.12



state mistrust: *"Man cannot believe… that this project, or anything else about the project, is gonna be proper … the rules are not set, they are changing as it goes … they are changing according to current political situation, how sirs need them, so they can move towards their goals."* (NGO_038: 82-87)


Euclidean distances (concepts)
Height
0
50
100
150
200
RESPONSIBILITY FRAME
PERIPHERAL CODES
RISK FRAME
DYSFUNCTIONAL STATE FRAME
SUPPORTING
OPPOSING
duty
cost_acknowledgement
employment
compensation_payment
irresponsible_citizen
false_repository_fail
false_value_loss
other_benefit
state_decides
inconsistent_state
place_attachment
veto
other_harm
waste_import
determination
false_employment
useless_repository
public_corruption
repository_fail
environmental_harm
value_loss
overriding_state
community_pressure
participation
state_mistrust
vain_negotiation
low_compensation
water_loss

Reconstructed frames

Responsibility	• We consume electricity and thus produce radioactive waste. • We (as well as state) have a moral and legal obligation to deal with this burden. • The repository is the only viable solution. • By delays and opposition we transfer this responsibility to further generations. • Opposition is irresponsible and based on emotional/irrational argumentation. • State has (legitimately) the last word; localities will be financially compensated.
Risk	• The siting process as well as potential construction of the repository is accompanied by number of risks (environmental, economic, social, health etc.). • We have responsibility to further generations to preserve the localities. • It is necessary to stop or to slow down the project till another (technological) solution is available.
Dysfunctional state	• The state is not able to deal with the issue competently and legitimately. • The localities are not effectively involved in the selection process. • The Working group is just a facade; the final decision depends solely on the state. • There is a lack of trust among stakeholders and the whole process lacks legitimacy.

R: advantages

- Freeware
- Open source
- Worldwide active community
- Flexible and developed

R community / sources

- There is huge number of free resources
- R package / library manuals
- R site: http://cran.r-project.org
- Community forums:
  - http://stackoverflow.com
  - http://www.statmethods.net
  - http://www.r-bloggers.com
- Youtube videos: https://www.youtube.com/watch?v=qHfSTRNg6jE
- Googling (often fastest)

R libraries / packages

- Library / package:
  - Can be though of as an extension that adds new functionality.
  - Libraries must be installed (just before the first use) and loaded.
  - Sometimes there can be conflicts among libraries (e.g. different functions with same names) – we can unload them.
  - Often there are dependencies among libraries (some libraries use functions from other libraries).

R: disadvantages

- Not as easily accessible as “clicking-programs”
- Data preparation could be demanding
- Could be slower for large datasets

R language

- object-oriented programming
  - **object:** instance of certain data class that can be manipulated according set of procedures (methods)
- functional-oriented programming
  - **function:** relation that associates input(s) with output(s)

- We can define certain objects and apply functions on them and vice versa.

Data types

- **Numeric:** continuous numeric data (**-1**, **0.5**, **10.49**)

- **Integer:** discrete numeric data (**-1**, **0**, **1**, …)

- **Character:** string values = **“anythingwithinquotes"**

- **Logical:** output of logical operation

  5 > 10 = **FALSE**

  5 < 7 | 7 > 10 = **TRUE**

Data types: factor

- **Factor:** variable that take limited number of discrete values - levels (categorical variable).
- Factor function converts vector of values into vector of **factor values** (always have form of **character**).
- Factors can be **unordered** (nominal variable) or **ordered** (ordinal variable).

```
> data = c(1,2,2,3,1,2,3,3,1,2,3,3,1)
> fdata = factor(data)
> fdata
 [1] 1 2 2 3 1 2 3 3 1 2 3 3 1
Levels: 1 2 3
```

http://www.r-tutor.com/

R: object and function

- **Object:**

  `vector <- c(1,2,3,4,5)`

- **Function:**

  `fun <- function(x) { x^2 }`

- **Output:**

  `fun(vector) = 1, 4, 9, 16, 25`

- **Nesting:**

  `fun_2 <- function(x) { fun(x) + 1 }`

R functions

- *word()* indicates function
- mean(vector)

- *function(argument_1, argument_2, …)*
- sample(0:100, 10, rep=FALSE)

- basic functions (part of the basic R package)
- package functions (part of the particular package)
- user functions (user-defined functions)

R objects

- **Vector**
  - Sequence (1-dimensional) of elements of **same data type**
- **Matrix**
  - 2-dimensional rectangular collection of elements of **same data type**
  - Array: n-dimensional matrix.
- **List**
  - Vector that can contain elements of **different data types**
- **Data frame**
  - List of vectors of equal length
  - Table data

Vector

```
> c(2, 3, 5)
[1] 2 3 5
```

```
> c("aa", "bb", "cc", "dd", "ee")
[1] "aa" "bb" "cc" "dd" "ee"
```

```
> c(TRUE, FALSE, TRUE, FALSE, FALSE)
[1]  TRUE FALSE  TRUE FALSE FALSE
```

http://www.r-tutor.com/

Matrix

```
     [,1] [,2] [,3]
[1,]    2    4    3
[2,]    1    5    7
```

http://www.r-tutor.com/

List

```
> n = c(2, 3, 5)
> s = c("aa", "bb", "cc", "dd", "ee")
> b = c(TRUE, FALSE, TRUE, FALSE, FALSE)
> x = list(n, s, b, 3)   # x contains copies of n, s, b
```

```
> x[c(2, 4)]
[[1]]
[1] "aa" "bb" "cc" "dd" "ee"

[[2]]
[1] 3
```

http://www.r-tutor.com/

Data frame

```
> mtcars
               mpg cyl disp  hp drat   wt ...
Mazda RX4     21.0   6  160 110 3.90 2.62 ...
Mazda RX4 Wag 21.0   6  160 110 3.90 2.88 ...
Datsun 710    22.8   4  108  93 3.85 2.32 ...
               ............
```

http://www.r-tutor.com/