

Analýza sociálních sítí (ASS) 1 + 2

Petr Ocelík

MEB421 25. 9. 2015

Osnova

- Úvod
- Teorie grafů
- Organizace dat
- Úvod do R

Úvod

- Společenskovední disciplíny se – z definice – zabývají sociálními **interakcemi**.
- ASS umožňuje sbírat a analyzovat **relační data**.

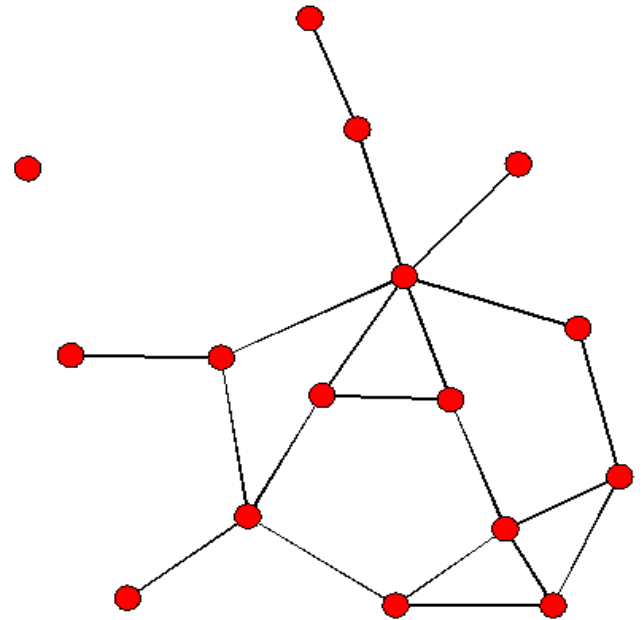
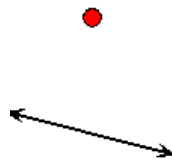
Úvod

- Hlavní předpoklad: sociální svět je organizován **relačně**.

“...transactions, interactions, social ties, and conversations constitute central stuff of social life.”

(Tilly 2008: 7)

- uzel
- spojnice



Terminologie (Guclu 2012)

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

Úvod: historie ASS

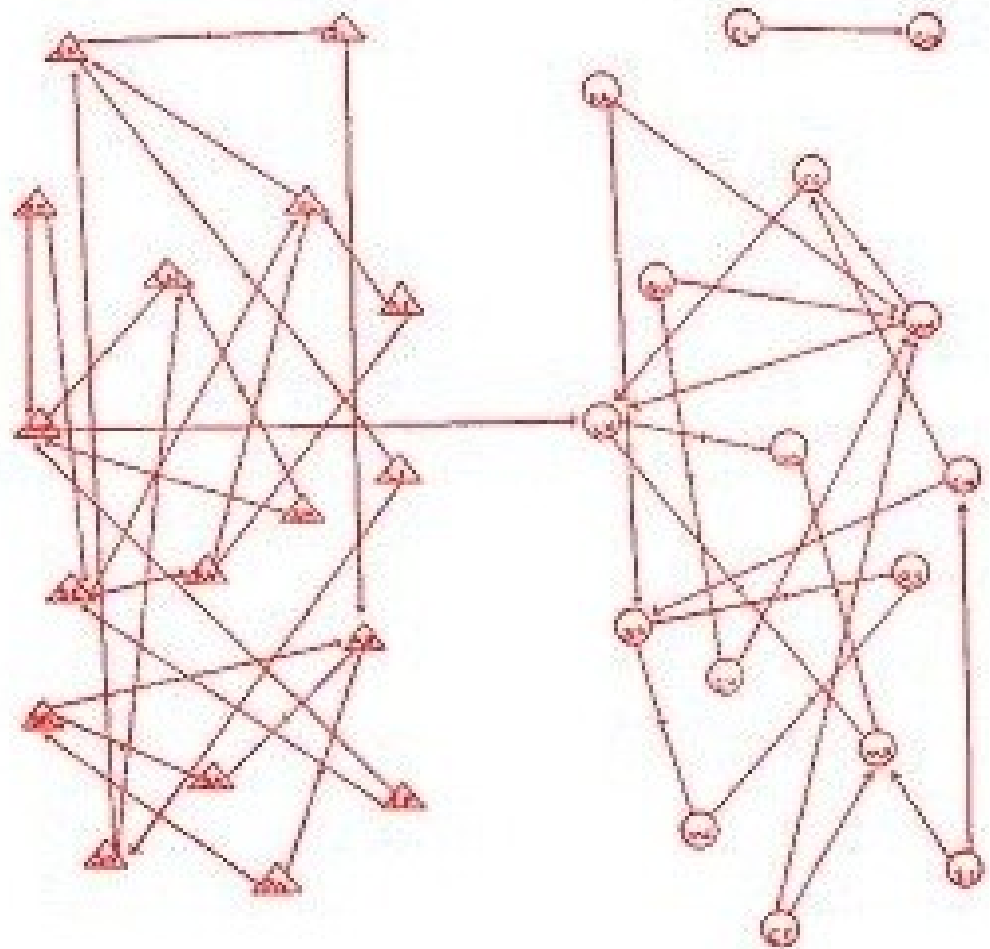
- Počátky spadají do 30. let 20. století (Jacob Moreno).
- ASS rozvíjena spíš na ad hoc základě v různých výzkumných centrech.
- ASS se coby disciplína etabluje v 70. letech (Mark Granovetter 1973).
- Revoluce sociální fyziky v 90. letech:
 - Watts and Strogatz (1998): síť malého světa
 - Barabasi and Albert (1999): bezškálové síť

Jacob Moreno

EMOTIONS MAPPED BY NEW GEOGRAPHY

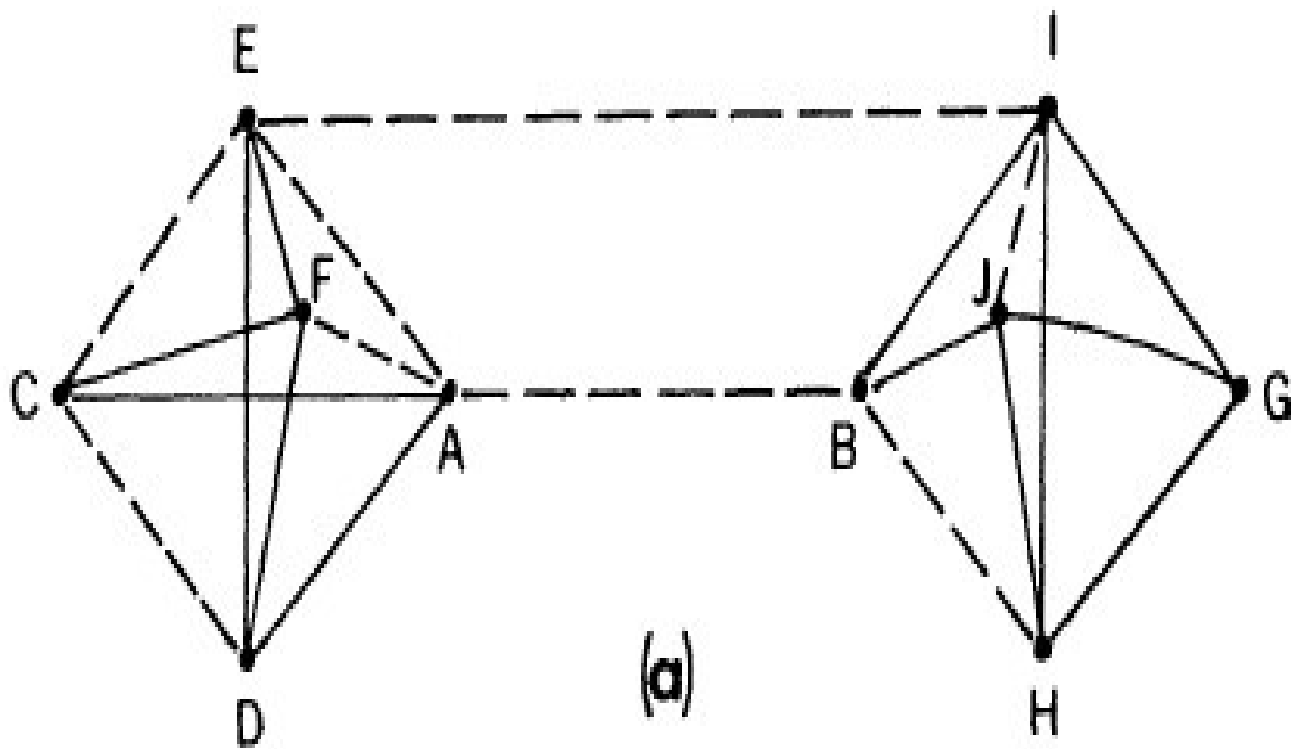
Charts Seek to Portray the
Psychological Currents of
Human Relationships.

New York Times
April 3, 1933

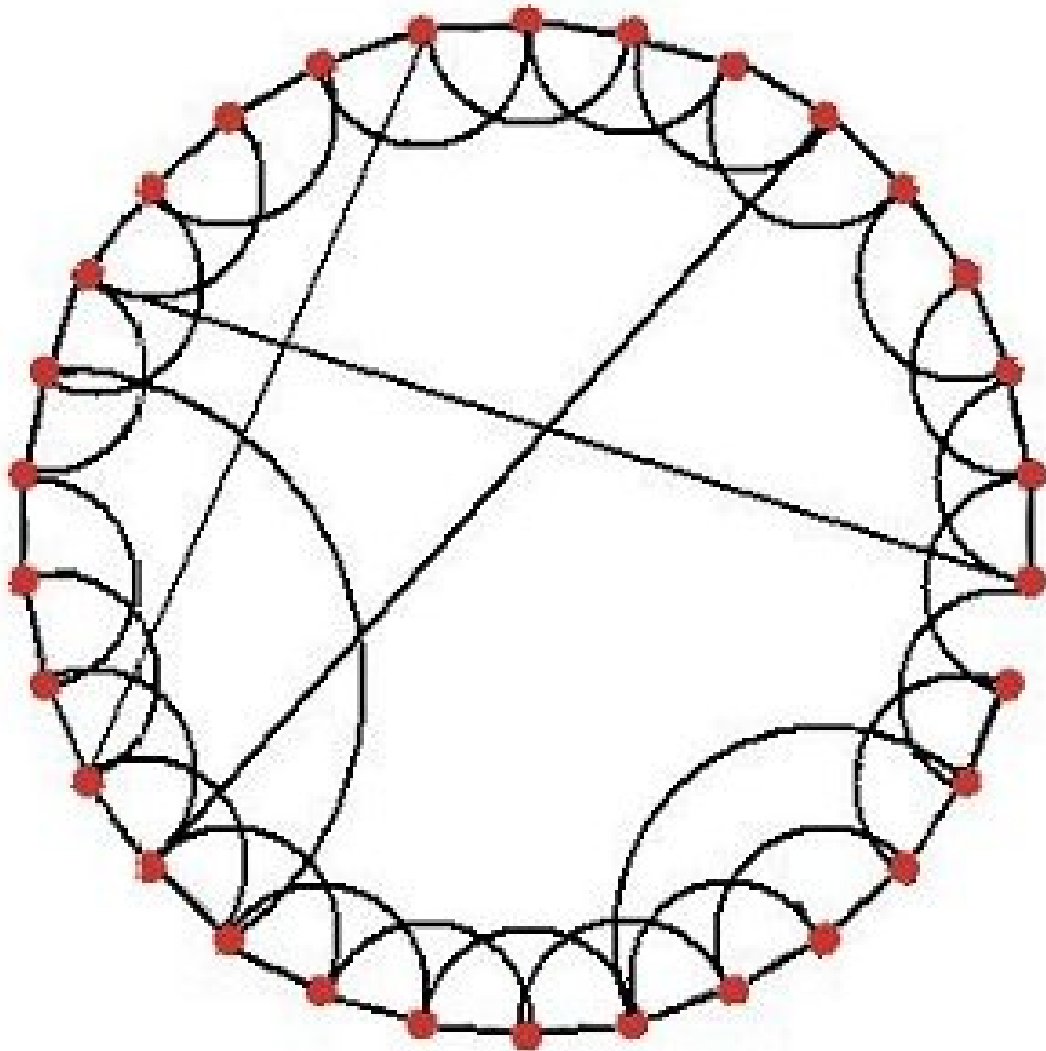


Mark Granovetter

The Strength of Weak Ties

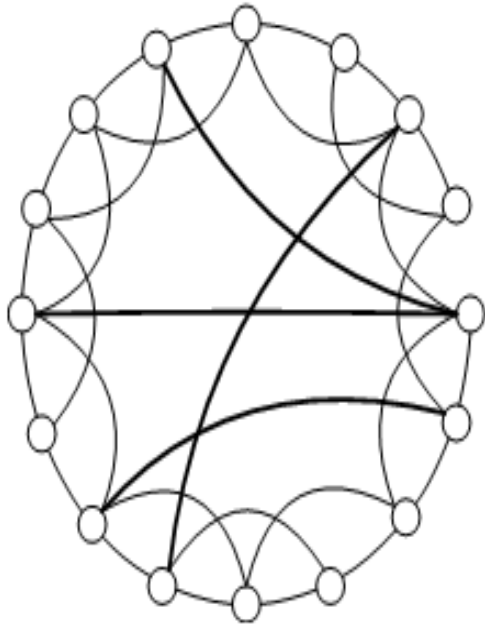


Small-world network

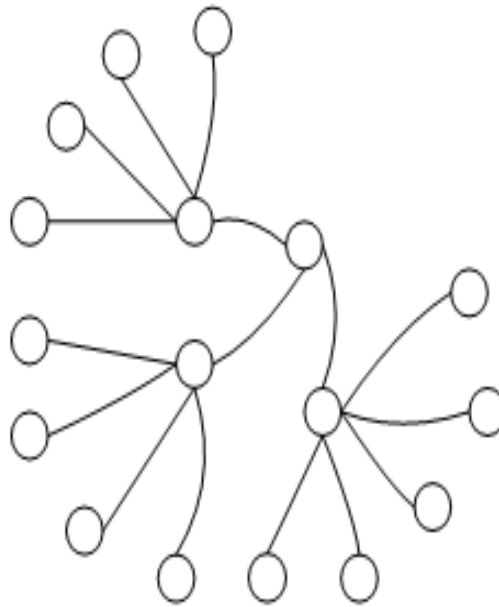


Scale-free network

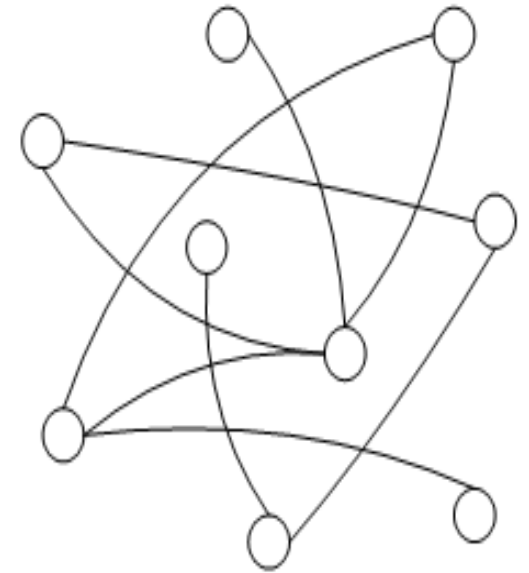
(a) Small-World Network (SWN)



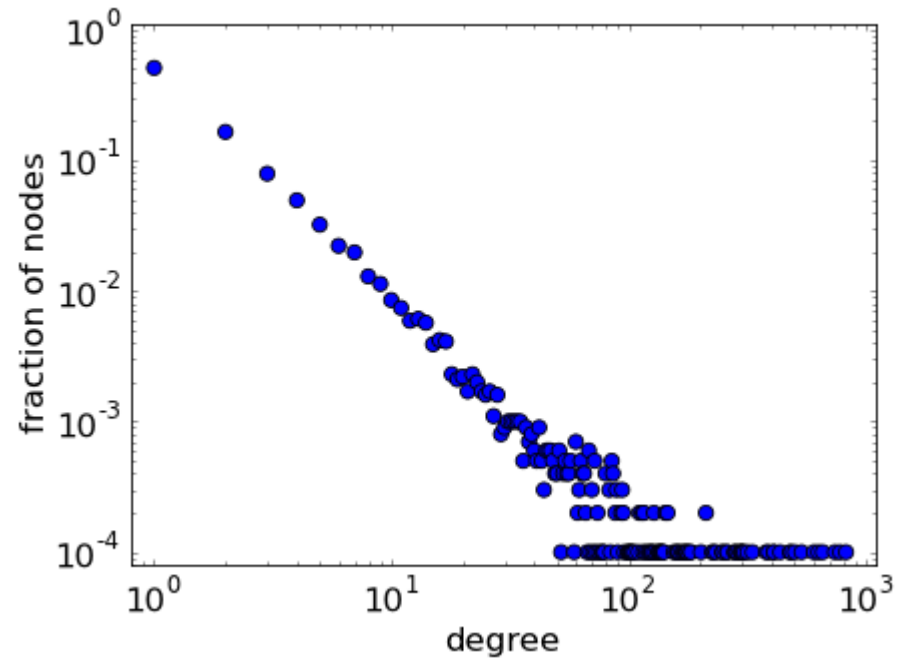
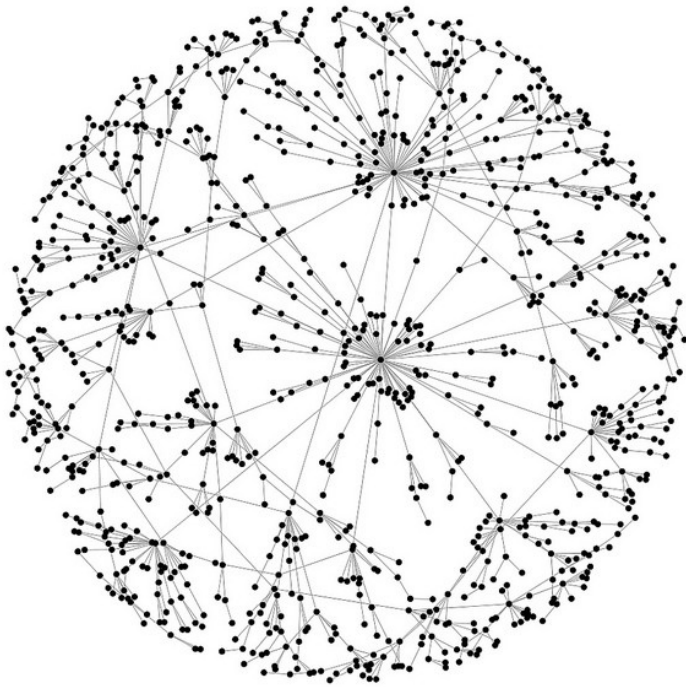
(b) Scale-Free Network (SFN)



(c) Random Network (RN)



Scale-free network



Teorie grafů

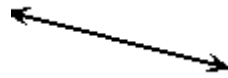
- Graf = matematická struktura, která modeluje **párové vztahy mezi objekty**.
- **Graf** (G) je uspořádaný pár sestávající se z množiny vrcholů (V) a množiny (neorientovaných) hran (E) nebo (orientovaných) oblouků (A).
- **$G = (V, E \text{ v } A)$**

Teorie grafů

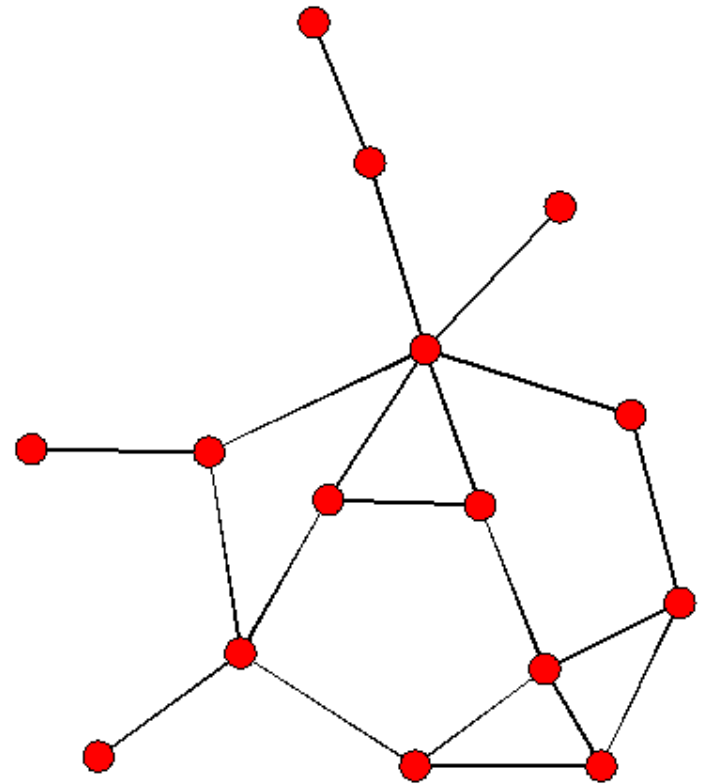
- Síť je složena z množiny **uzlů** a množiny **spojnic**.

uzel

spojnice

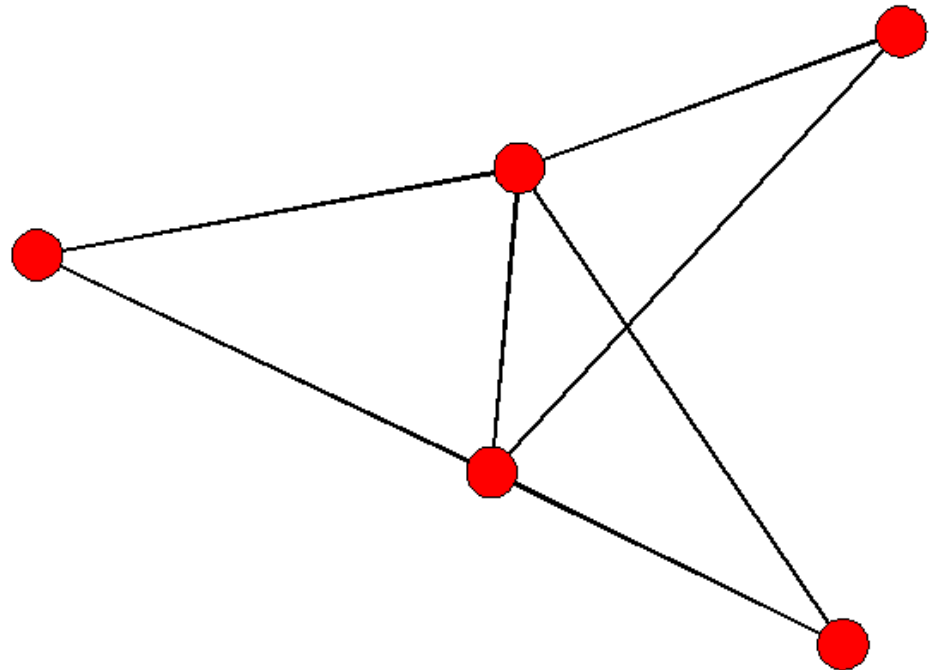


- **síť = graf**



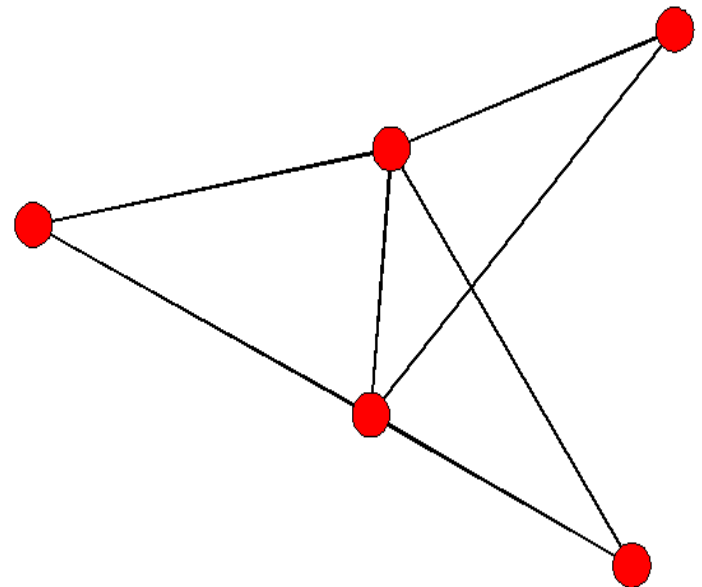
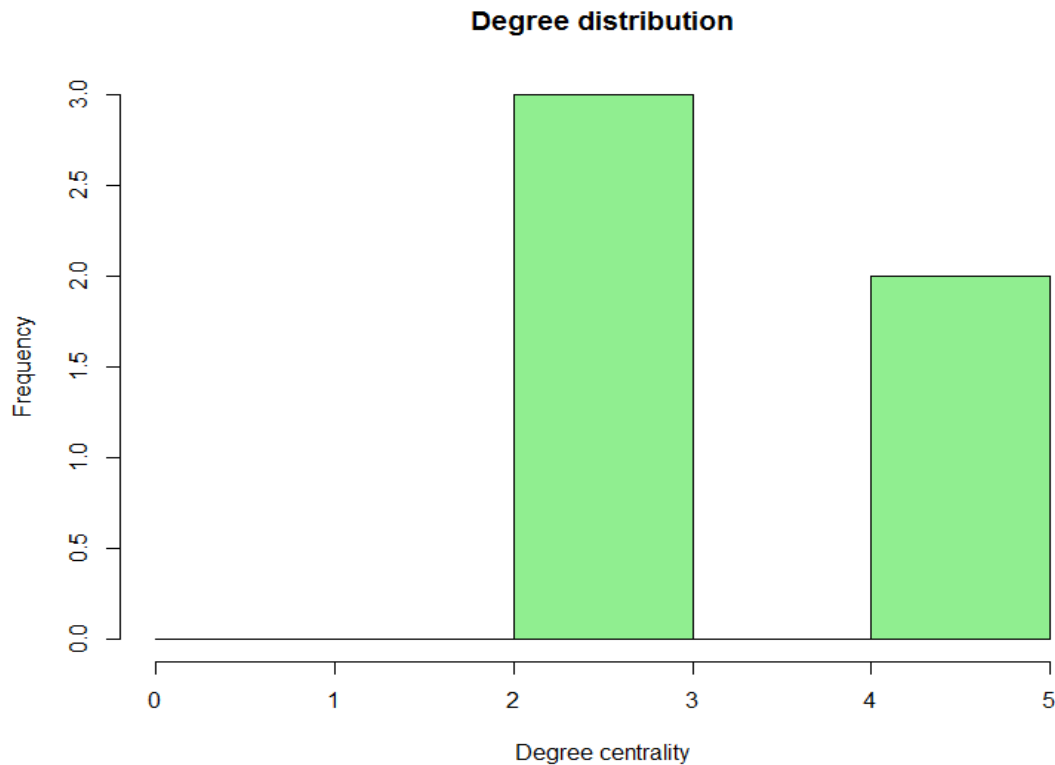
Teorie grafů

- **řád (order)** = # uzlů
- **velikost (size)** = # hran
- **stupeň (degree)** = # spojení jednotlivých uzlů



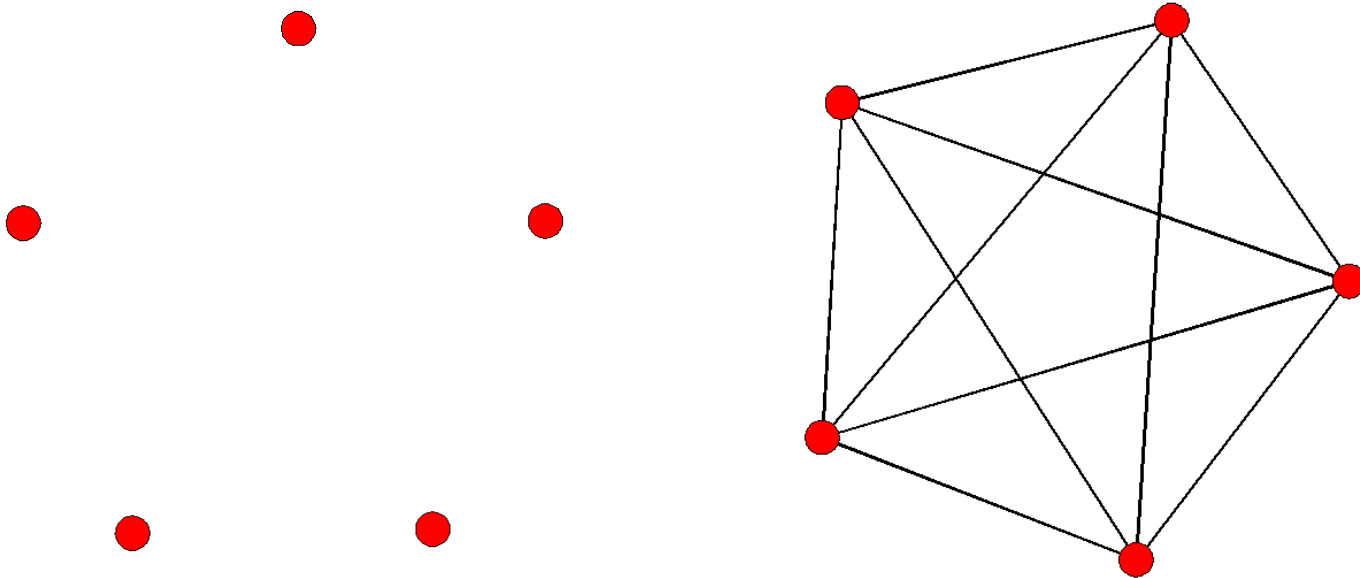
Teorie grafů

- řád (order) = 5
- velikost (size) = 7



Teorie grafů

- **Úplný graf** je graf, jehož všechny vrcholy jsou vzájemně přilehlé.
- **Prázdný graf** je graf, v němž není ani jedna spojnice.

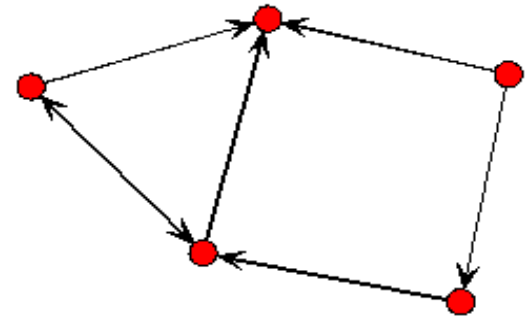
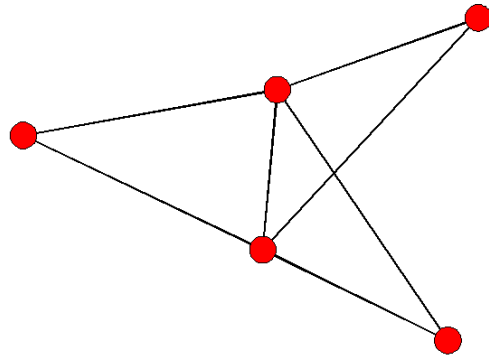


Teorie grafů

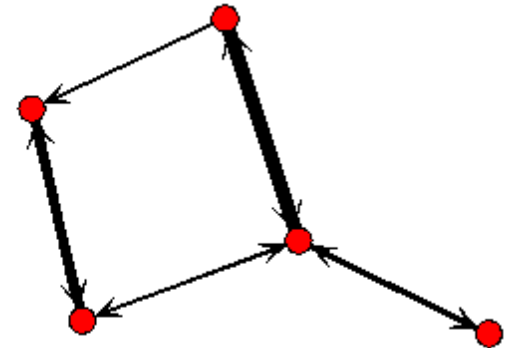
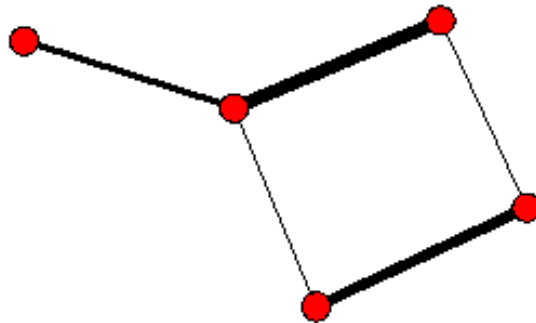
neorientované

orientované

binární



vážené



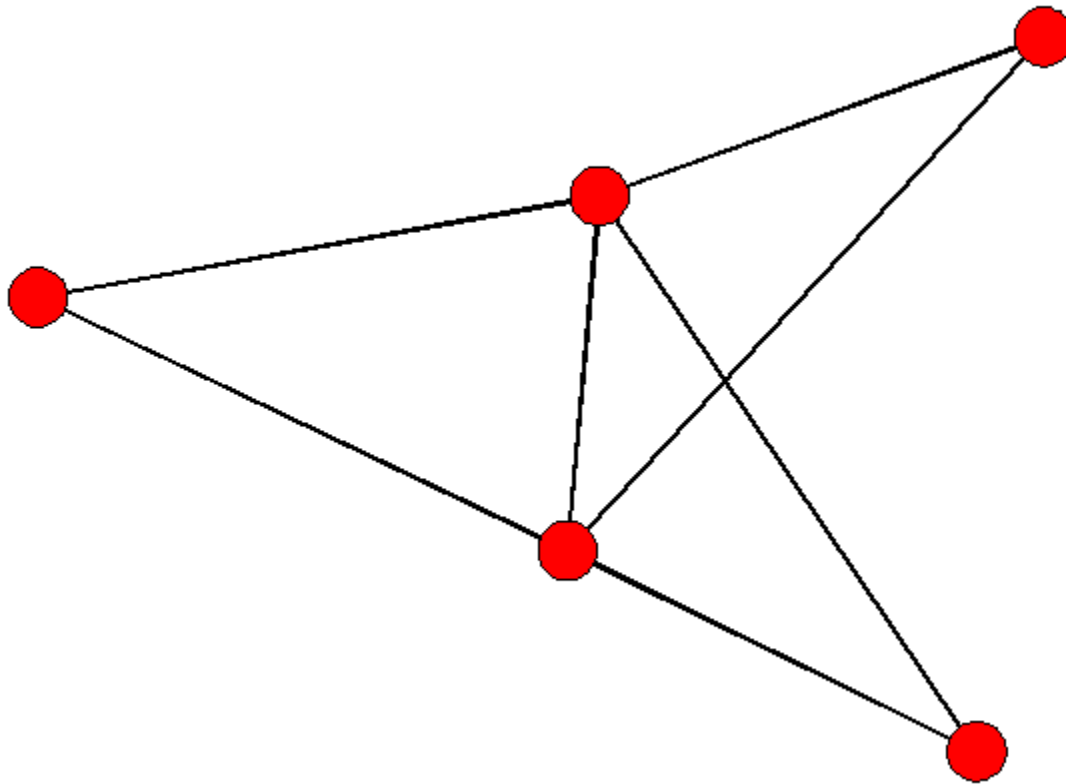
Teorie grafů

- **Topologie sítě** je definována dvěma koncepty: konektivitou a centralitou.
- **Konektivita** popisuje propojenost uzlů v síti (zaměřuje se na **toky**).
- **Centralita** popisuje umístění uzlů v síti (zaměřuje se na **pozice**).

Teorie grafů

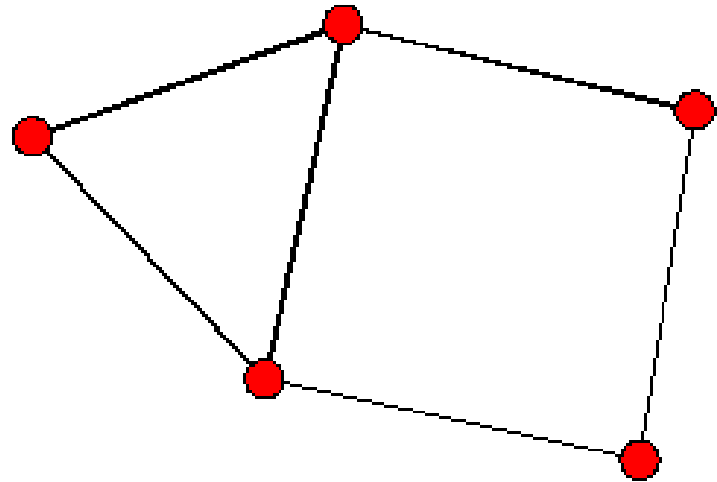
- **Krok (step):** pohyb podél jedné spojnice začínající a končící uzlem.
- **Sled (walk):** posloupnost kroků, která začíná a končí uzlem.
- **Cesta (path):** sled, kde se žádný uzel ani spojnice neopakují.
- **Geodetika:** nejkratší cesta spojující dva různé uzly.
- **Vzdálenost** dvou uzlů = geodetika.
- **Diametr:** nejdelší vzdálenost jakýchkoli dvou uzlů v síti.

Teorie grafů



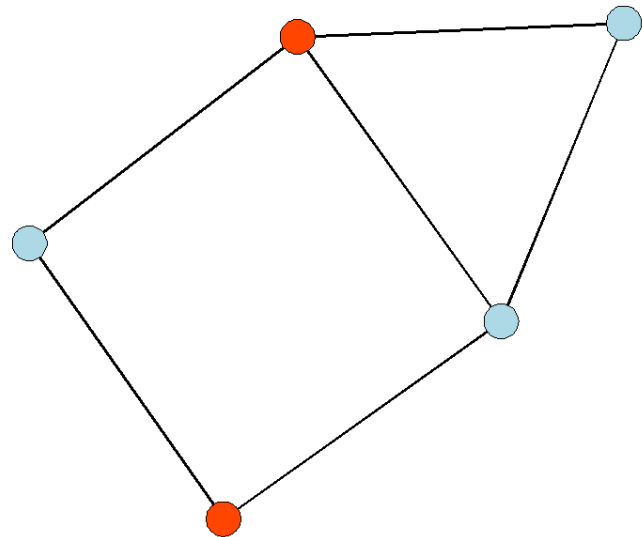
Teorie grafů

- Přímou spojený uzel je **přilehlý (adjacent)**.
- Spojnice přímo spojená s uzlem je **incidenční (incident)**.
- Všechny přímo spojené uzly vytvářejí **sousedství (neighbourhood)**.



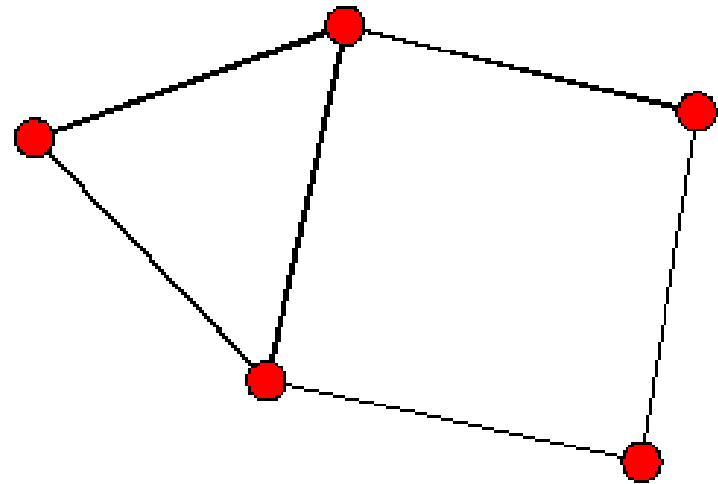
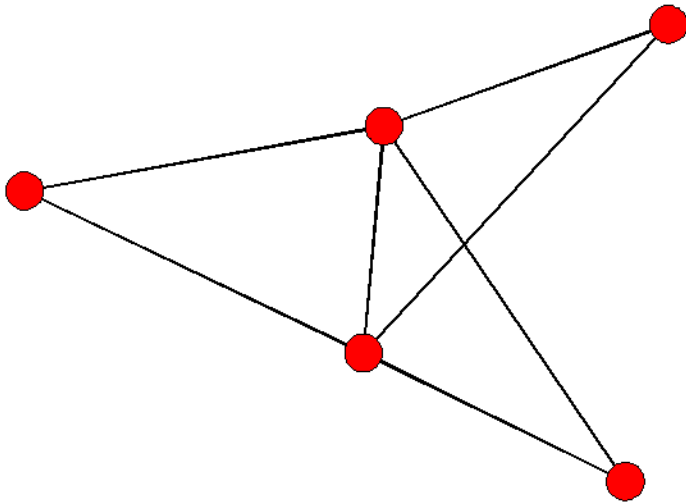
Teorie grafů

- Přímou spojený uzel je **přilehlý (adjacent)**.
- Spojnice přímo spojená s uzlem je **incidenční (incident)**.
- Všechny přímo spojené uzly vytvářejí **sousedství (neighbourhood)**.



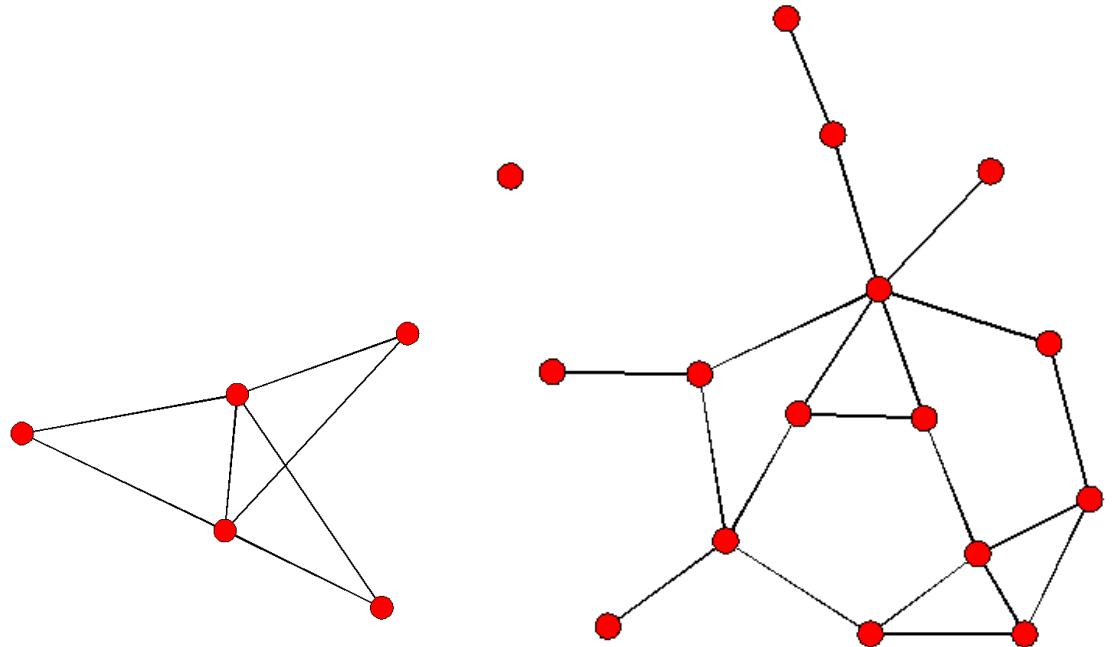
Teorie grafů

- **Subgraf** je jakákoli podmnožina uzlů a spojnic grafu.
- **Komponent** je souvislý subgraf.



Teorie grafů

- **Dosažitelnost (reachability)** je dána existencí cesty mezi uzly.
- **Izolát** je uzel bez jediného spojení, tj. uzel se stupněm 0.

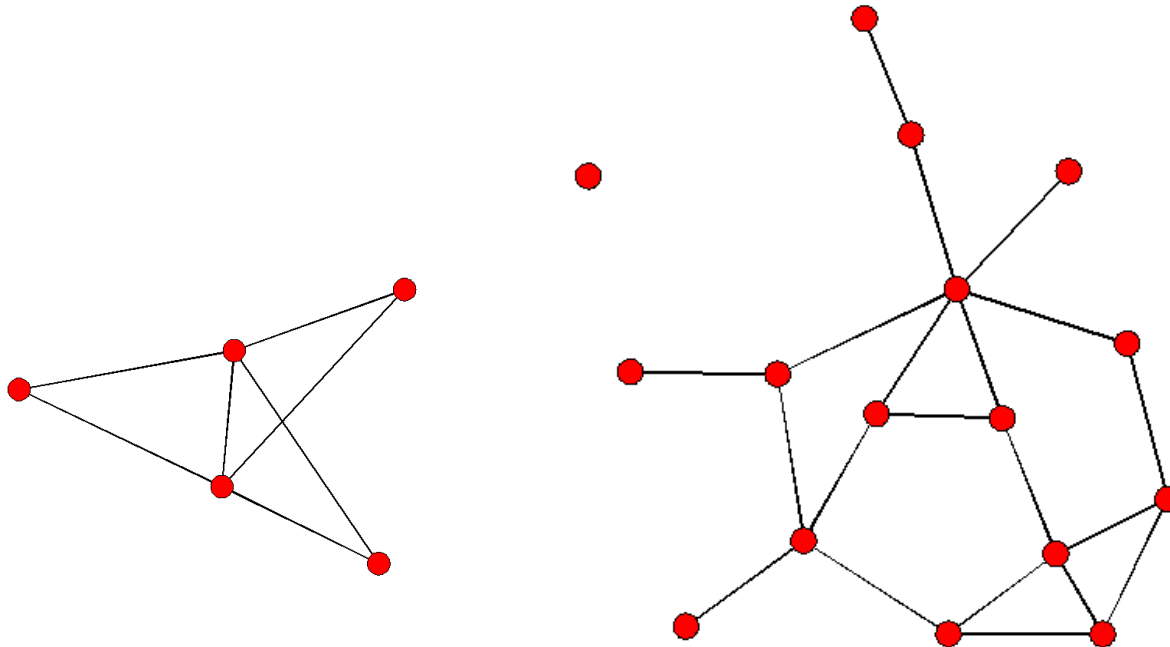


Teorie grafů

- **Strukturální mezera (hole)** is a lack of connection between two nodes or subgraphs.
- **Křižovatka (cutpoint)** je uzel, jehož odnětím vznikne strukturální mezera.
- **Most (bridge)** je spojnice, jejímž odnětím vznikne strukturální mezera.

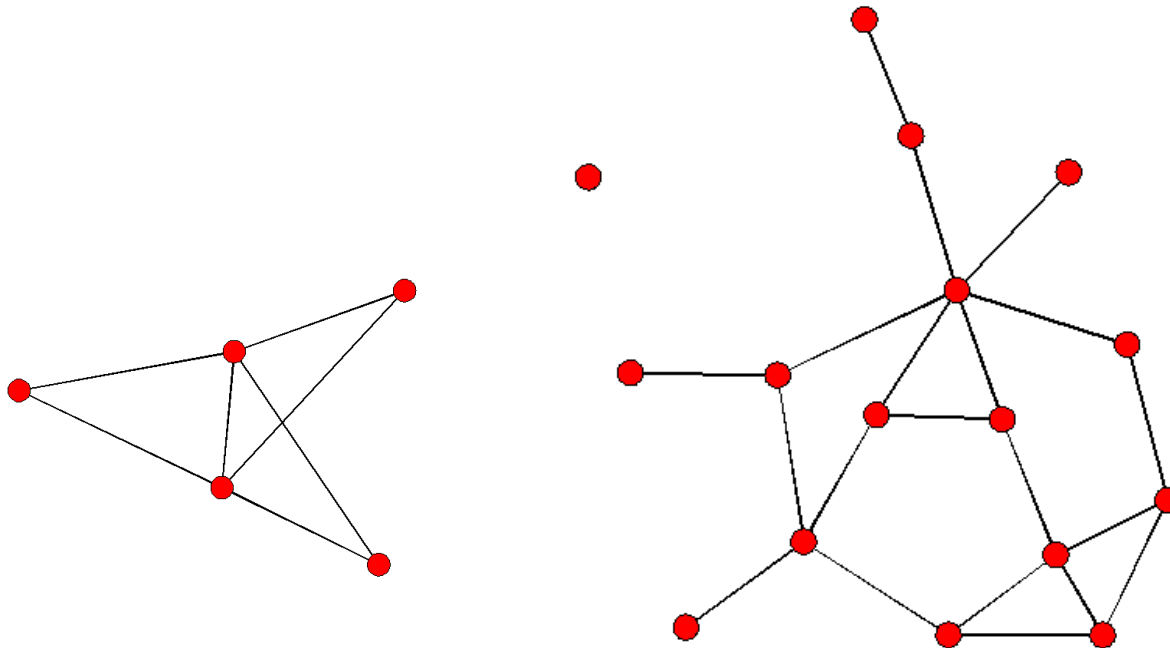
Teorie grafů

- **Strukturální mezera (hole)** je absence spojení mezi dvěma uzly nebo subgrafy.



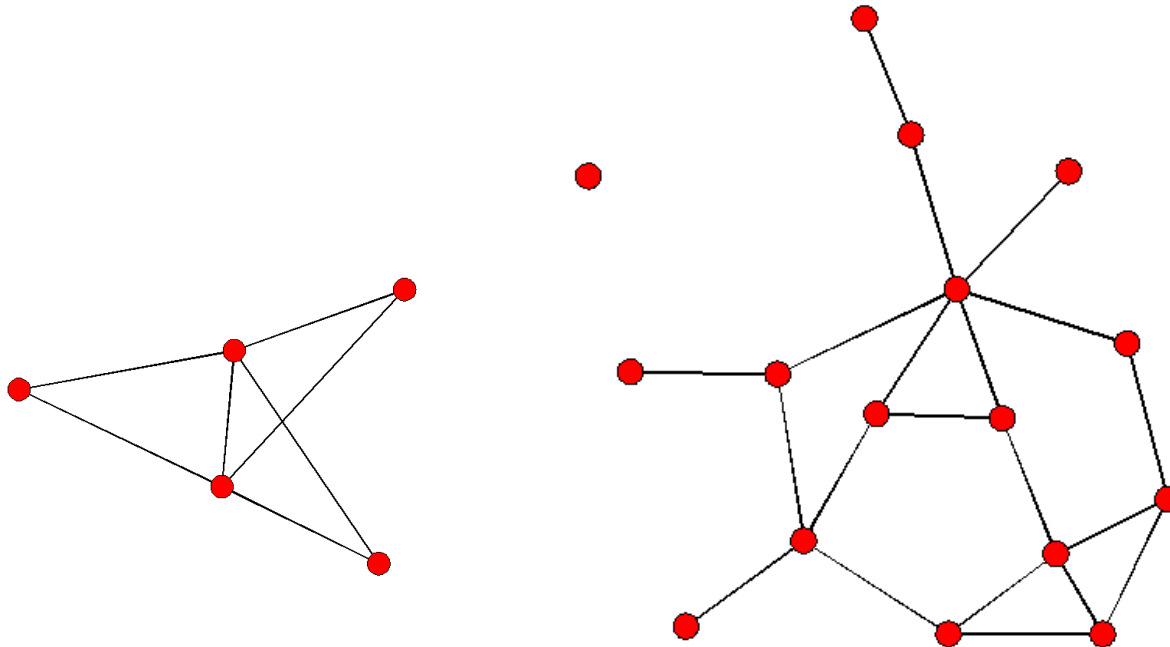
Teorie grafů

- **Křižovatka (cutpoint)** je uzel, jehož odnětím vznikne strukturální mezera.



Teorie grafů

- **Most (bridge)** je spojnice, jejímž odnětím vznikne strukturální mezera.

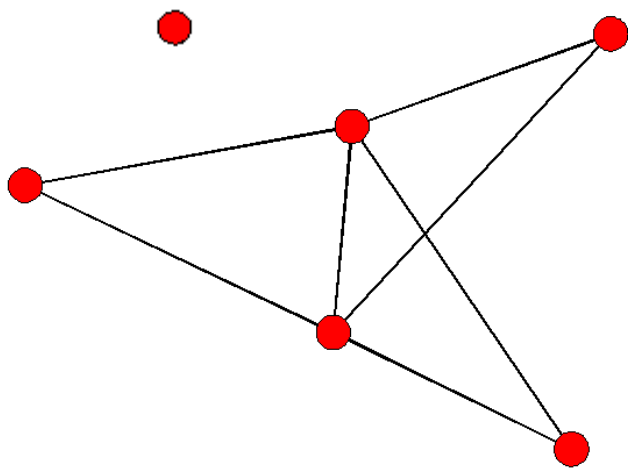


Teorie grafů

- **Inkluzivita** je dána # propojených uzlů vzhledem k celkovému # uzlů v síti.
- **Hustota** je podíl # pozorovaných spojníc na maximálním možném # spojníc v síti.

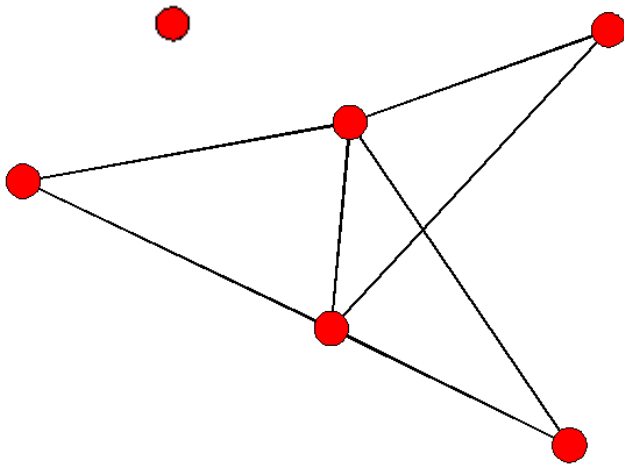
Teorie grafů

- **Inkluzivita (inclusivity)** je dána # propojených uzlů vzhledem k celkovému # uzlů v síti.



Teorie grafů

- **Inkluzivita (inclusivity)** je dána # propojených uzlů vzhledem k celkovému # uzlů v síti.
- **Inkluzivita = $5 / 6 = 0.83$**



Teorie grafů: notace

- G = graf/síť
- N = # uzlů v síti, n = jednotlivý uzel
- e = spojnice, g = geodetika
- i, j, \dots = indexy (označují vybrané prvky)
- g_{ij} = geodetika spojující uzly i a j , n_i = uzel i
- k = # vybraných prvků (typicky uzlů)

- Velká písmena: globální indikátory
- Malá písmena: lokální indikátory
- $cd(n_i)$ = stupňová centralita uzlu i
- $Cd(G)$ = stupňová centralizace grafu G

Teorie grafů

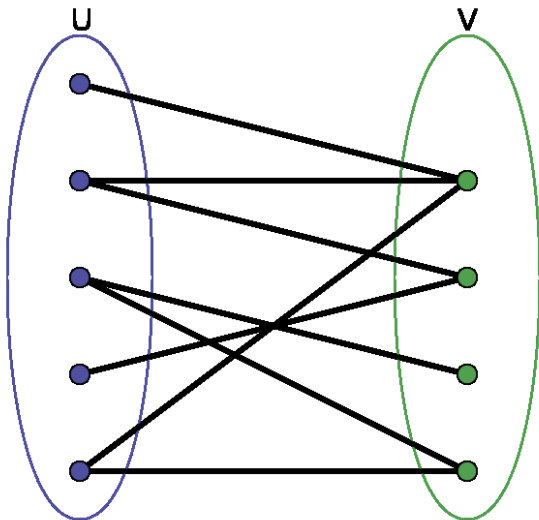
- **Hustota (density)** dána # pozorovaných spojníc ($\sum e$) na celkovém možném # spojení v síti.
- # všech možných spojení v neorientované síti = $(N * (N - 1)) / 2$
- # všech možných spojení v orientované síti = $(N * (N - 1))$
- Hustota (neoreint.): $\sum e / ((N * (N - 1)) / 2)$

Cvičení

- Hustota (neorient.): $\sum e / ((N * (N - 1)) / 2)$
- Předpokládejme, že vytváříte síť. Spojnice vzniká, pokud dva studenti sedí vedle sebe nebo „našikmo“.
 - Jaká je inkluzivita a hustota této sítě?

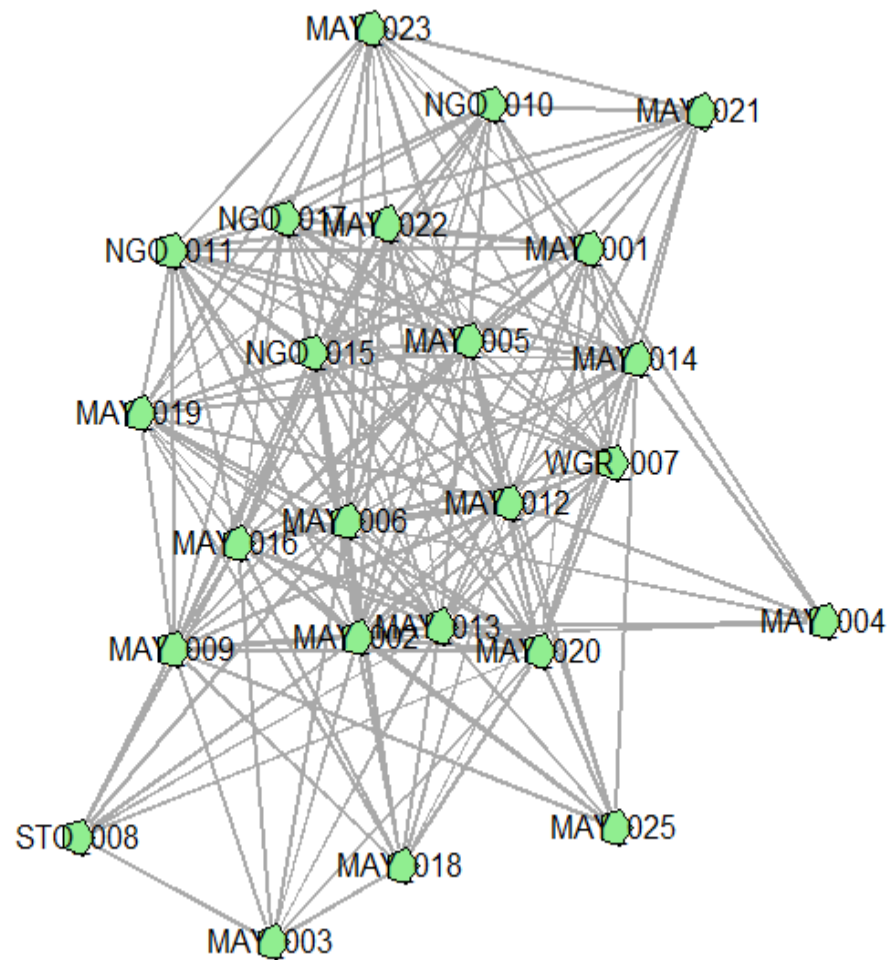
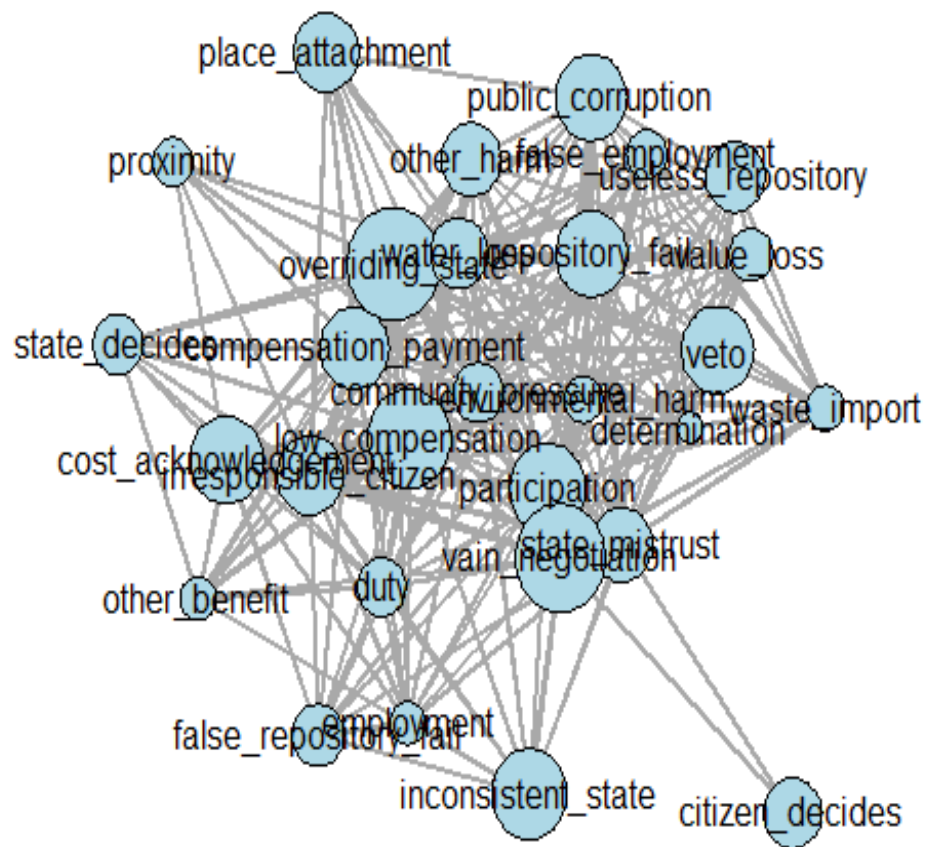
Teorie grafů

- **Bipartitní (někdy two-mode) síť** obsahuje dvě oddělené množiny uzlů (U a V).
- **Spojnice** jsou možné pouze **mezi těmito dvěma množinami uzlů**, nikoli uvnitř těchto množin.



Teorie grafů

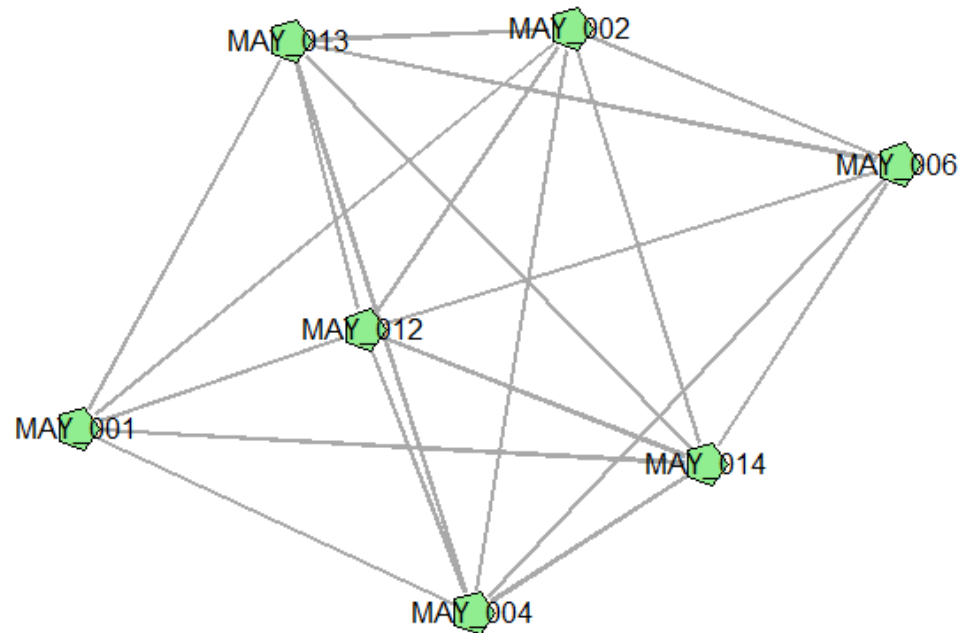
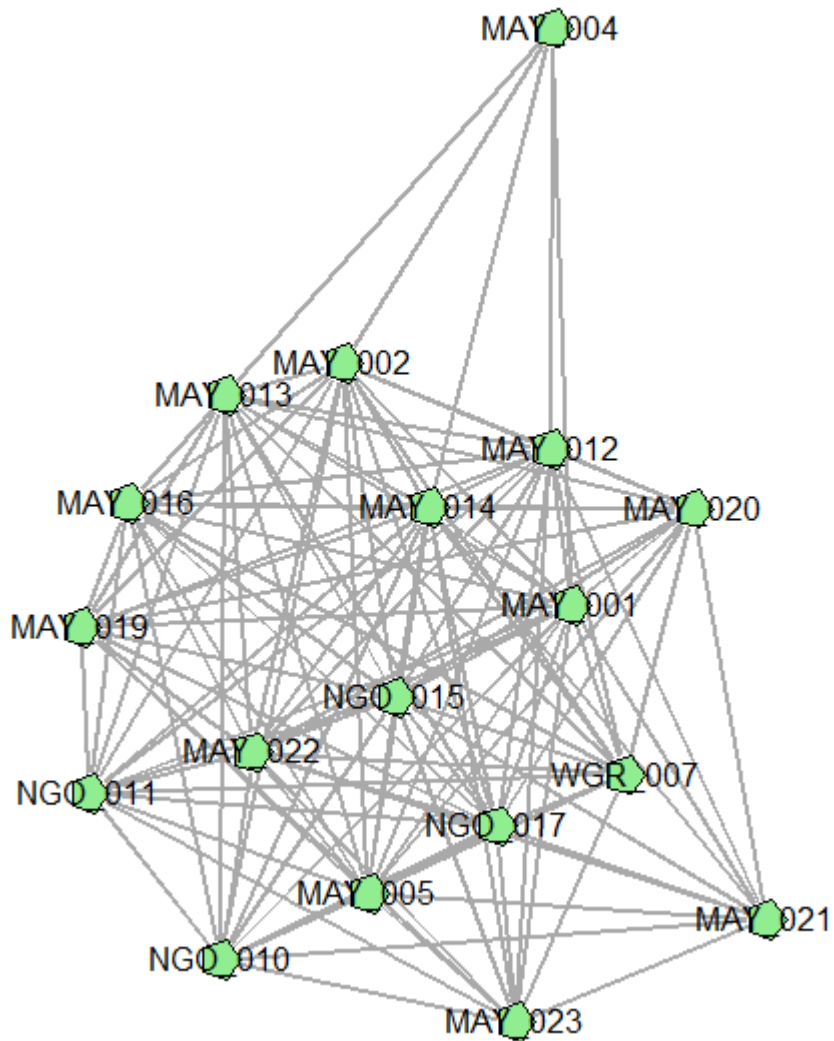
- **Jednorozměrné projekce** bipartitní sítě ukazují sítě aktérů a konceptů.
- Aktéři jsou propojeni, pokud sdílejí alespoň jeden koncept.
- Koncepty jsou propojeny, pokud jsou sdíleny alespoň jednou dvojicí aktérů.



Teorie grafů

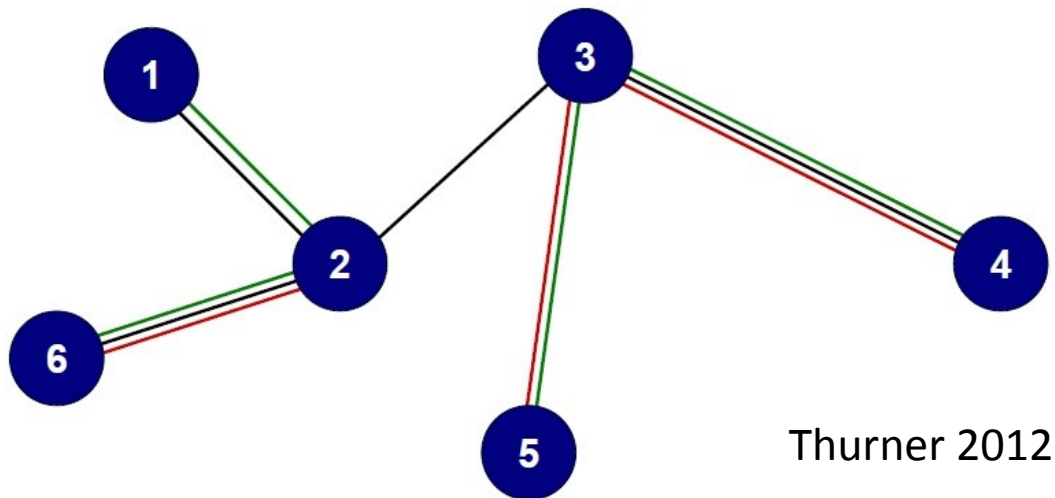
- **Egocentrická síť** je osobní síť daného jednotlivce (ega).
- Počet kroků spojující uzel s egem klasifikuje uzel do příslušné zóny.
- **Zóna prvního řádu** zahrnuje všechny přímo spojené uzly (alteri), zóna druhého řádu všechny uzly spojené dvěma kroky atd.

Teorie grafů



Teorie grafů

- **Multiplexní síť** se skládá z jedné množiny uzlů a více než jedné množiny spojnic.
- Např. tatáž skupina lidí propojená skrze různé online sociální sítě (Facebook, Twitter, LinkedIn atd.).



Organizace dat

- **Atribuční data:** individuální charakteristiky.
 - Např.: věk, příjem, vzdělání, HDP, TPES, atd.
- **Relační data:** charakteristiky vztahů.
 - Např.: rodinné vztahy, obchodní toky, konflikty, atd.

Organizace dat: hranice sítě

- **Hranice sítě:** vymezení zpravidla problematické.
- Často žádné „přirozené hranice“.
- Různé strategie vymezení hranic:
 - nominální (např. všichni členové EU)
 - poziční (např. všechny demokratické státy)
 - realistická (např. všechny státy prezentující se jako lidové demokracie)
 - relační (např. všechny státy, které jsou ostatními označovány za demokracie)
 - založená na událostech (např. všechny státy, které se účastnily války v Iráku)

Organizace dat: vzorkování

- Často nemožné získat přístup k celé populaci.
- Náhodný výběr není vhodný - proč?

Organizace dat: vzorkování

- Často nemožné získat přístup k celé populaci.
- Náhodný výběr není vhodný - proč?
- Burtova formule ztráty informace = $(100 - k)\%$.
- Metody výběru vzorku:
 - Výběr nabalováním
 - Výběr založený na attributech

Organizace dat: sběr dat

- Dotazníky / rozhovory
- Generátor jmen (dotazník)

“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)

- Pozorování / experiment
- Archivní data

Cvičení

- Definujte výzkumnou otázku.
- Definujte populaci a hranice sítě.
- Definujte metodu výběru vzorku a sběru dat.

Organizace dat

- **(Sociální) data**, atribuční stejně jako relační, jsou typicky organizována v **datových maticích**.
- **Case-by-variable matice** je standardní způsob organizace dat v kvantitativním výzkumu.
- **Není vhodné pro relační data.**

Case-by-variable matrix

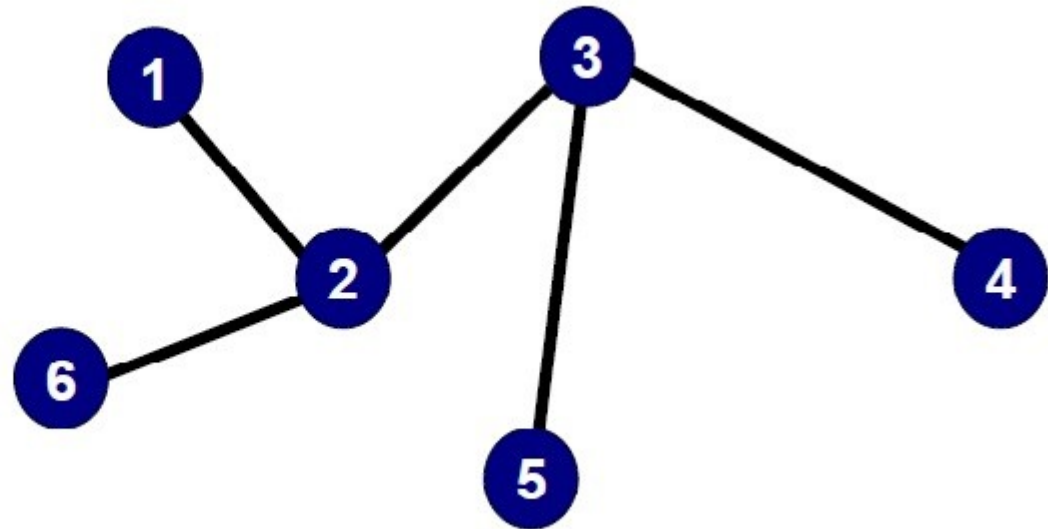
	A	B	C	D	E
1	id	G	P	PPG	NAT
2	Wayne Gretzky	1487	2857	1.92131809	CAN
3	Mark Messier	1756	1887	1.074601367	CAN
4	Gordie Howe	1767	1850	1.046972269	CAN
5	Ron Francis	1731	1798	1.03870595	CAN
6	Marcel Dionne	1348	1771	1.31379822	CAN
7	Steve Yzerman	1514	1755	1.159180978	CAN
8	Mario Lemieux	915	1723	1.883060109	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	1.114151681	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

Organizace dat: datové matice

- **Maticе sousednosti (adjacency matrix)**
ukazuje, zda spolu uzly sousedí, či nikoli.
- **Incidenční matice (incidence matrix)**
zaznamenává vztahy mezi dvěma typy uzlů.
 - Řádky reprezentují jeden typ uzlů.
 - Sloupce reprezentují druhý typ uzlů.

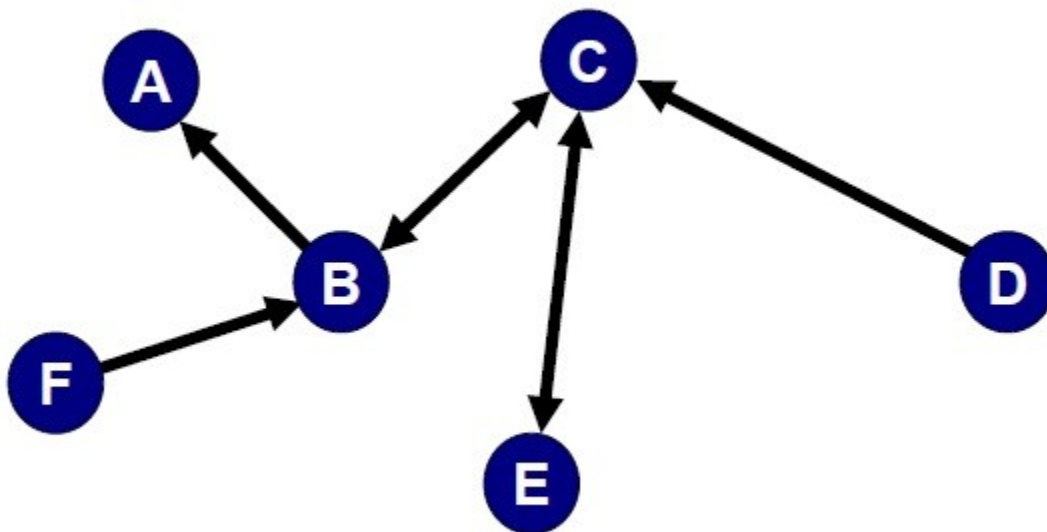
Neorientovaná binární síť

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



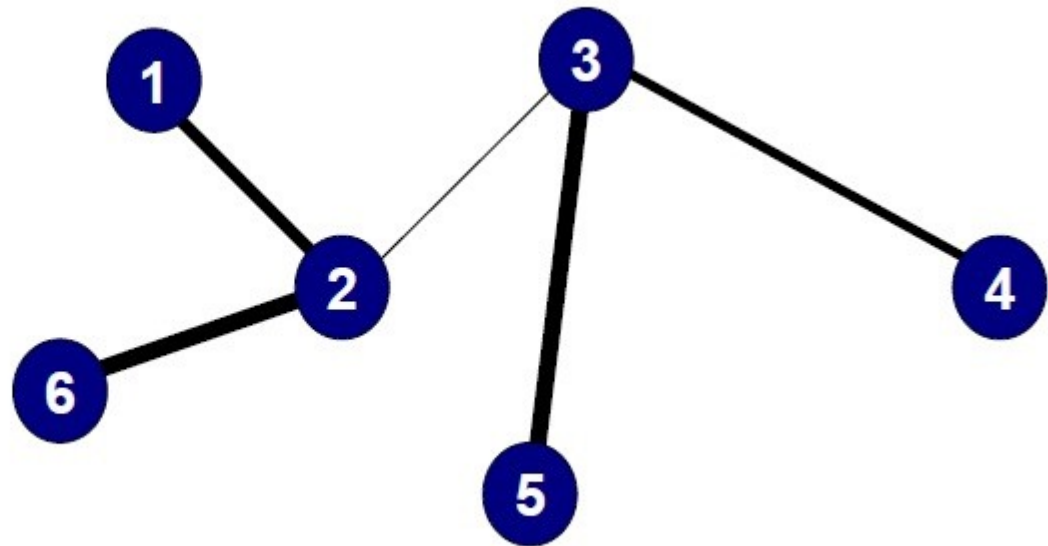
Orientovaná binární síť

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



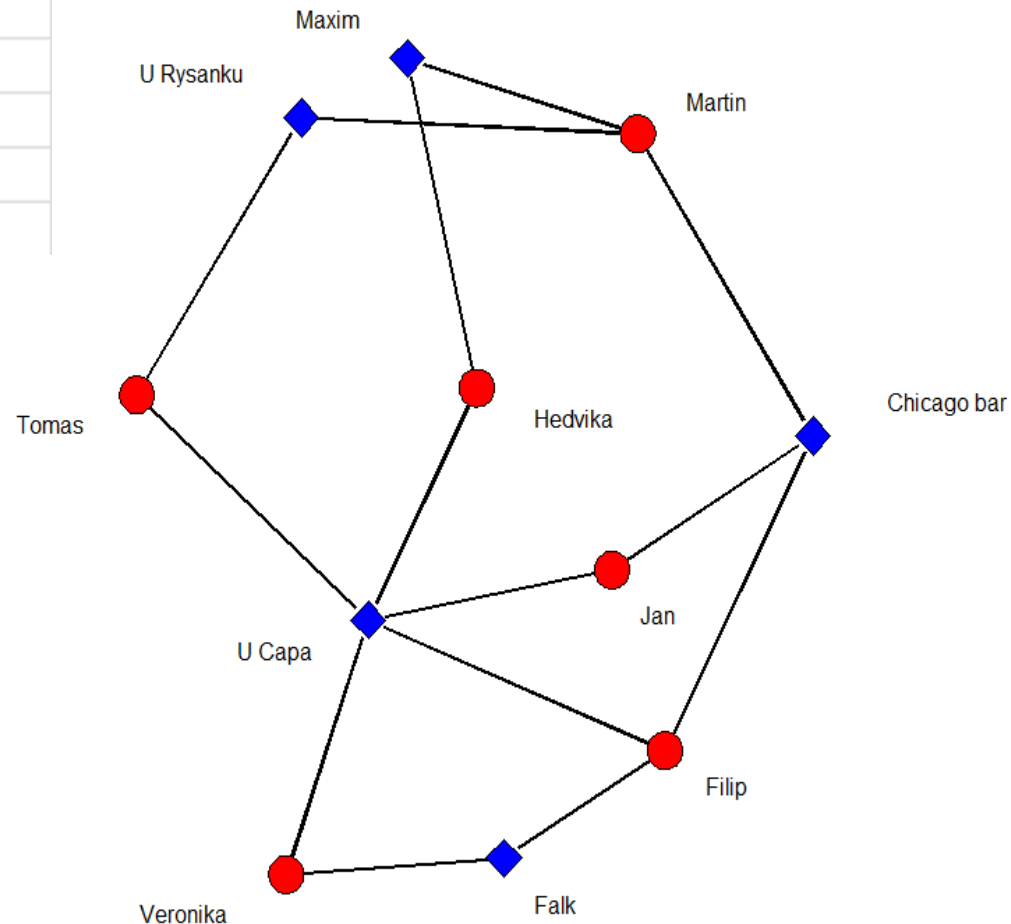
Neorientovaná vážená síť

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



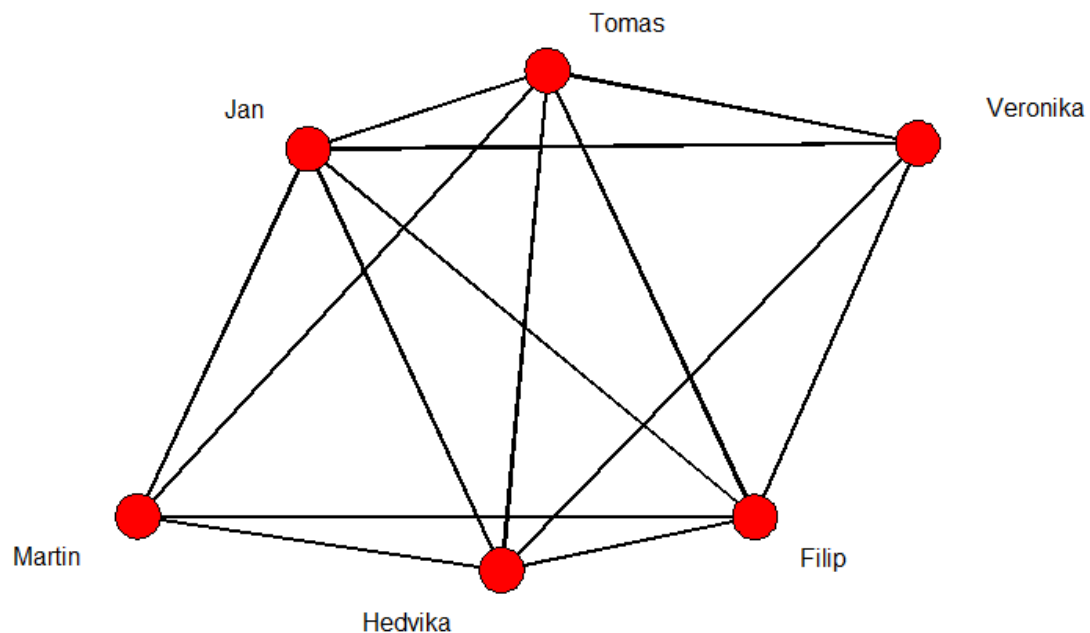
Incidenční matice (case-by-event)

	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



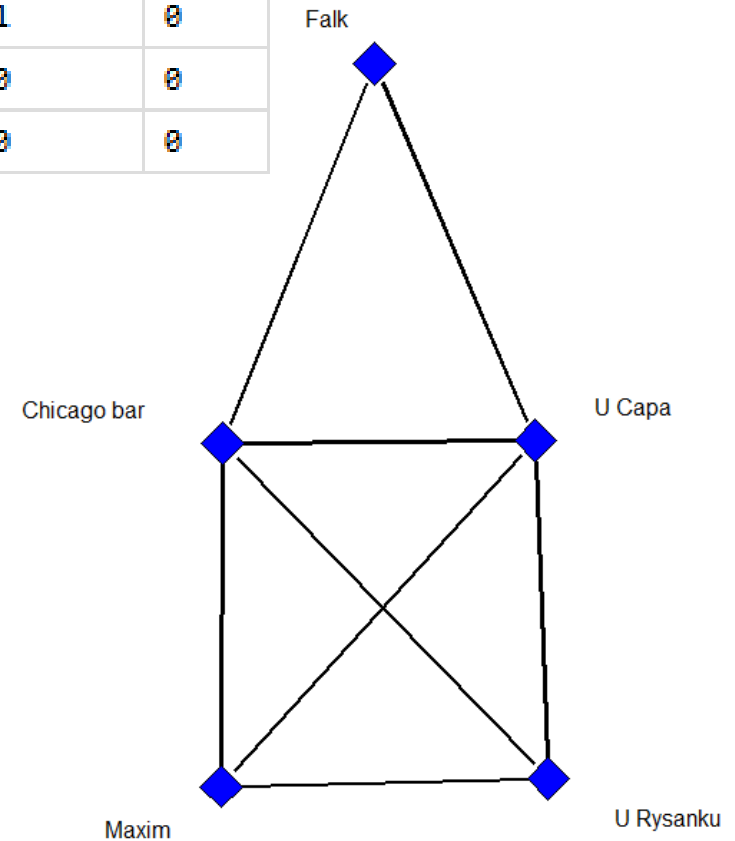
Matice sousednosti (case-by-case)

	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



Matrice susednosti (event-by-event)

	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



Operace s maticemi: jednorozměrná projekce

- Jednorozměrnou projekci incidenční matice získáme, pokud tuto matici vynásobíme její transpozicí.
 - Transponovaná matice: řádky v pozici sloupců a naopak.
- Pro případy (**řádky**) musí být transpozice na druhém místě.
 - `matrix %*% t(matrix)`
- Pro události (**sloupce**) musí být transpozice na prvním místě.
 - `t(matrix) %*% matrix`

Transpozice matice

- Incidenční matice

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

- Transpozice

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

Násobení matic (případy)

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

$\%*\%$

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

Tečkový součin: nejprve první řádek a první sloupec

$(1, 0, 1, 1)$ a $(1, 0, 1, 1)$, vynásobíme korespondující prvky a sečteme jejich produkty.

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

Násobení matic (případy)

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

Násobení matic (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

Tečkový součin: nejprve první řádek a první sloupec

(1, 0, 1, 1) a (1, 0, 1, 1), vynásobíme korespondující prvky a sečteme jejich produkty.

$$(1, 0, 1, 1) * (1, 0, 1, 1) = 1*1 + 0*0 + 1*1 = \mathbf{2}$$

Násobení matic (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

Cvičení

- Definujte jednorozměrné projekce této incidenční matice:

	Jan	Petr	Hedvika
Introduction	1	0	1
Methodology	1	1	0

Mini-case study

- Příprava projektu hlubinného úložiště zahájena v roce 1990.
- Úložiště projektováno tak, aby umožnilo bezpečně uskladnit kontejnery s radioaktivním odpadem v řádu stovek tisíc let.
- V současnosti 7 kandidátských lokalit.
- Projekt je od počátku provázen projevem lokální opozice.
- **Cíl výzkumu:** zmapovat, jak je téma rámováno ze strany lokální opozice a akceptace.

Čertovka

lokality na území obcí Blatno a Lubenec
v Středočeském kraji a na území obcí
Tis u Blatna a Žihle v Plzeňském kraji



Březový potok

lokality na území obcí Pačejov, Kvěšňovice,
Člšany, Maňovice, Chanovice a Velký Bor
v Plzeňském kraji



Magdaléna

lokality se nachází na území obcí Jistebnice,
Nadějkov a Božetice v Jihočeském kraji

Čihadlo

lokality se nachází na území města Dešná
a obcí Světlá, Loháňkov a Pluhův Žďár
v Jihočeském kraji

Hrádek

lokality na území obcí Rohozná, Dolní Čerekev,
Čejle, Hojkov, Mlýkov a městyse Nový Rychov
v kraji Vysočina

Kraví Hora

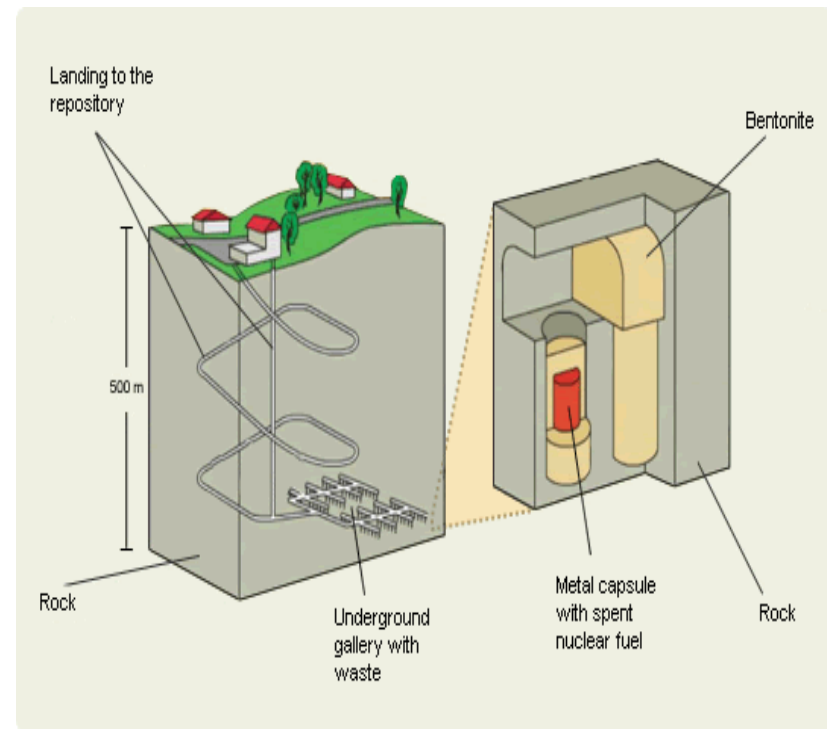
lokality na území obcí Bukov, Věžná, Střítež,
Moravské Pavlovice, Drahonín, Okří a Sejtek
spadající pod kraj Vysočina a Jihomoravský kraj

Horka

lokality na území obcí Hodov, Rohy, Oslavička,
Budíšov, Nárámec, Vlčatín, Osová, Rudíkov
a Oslavice v kraji Vysočina

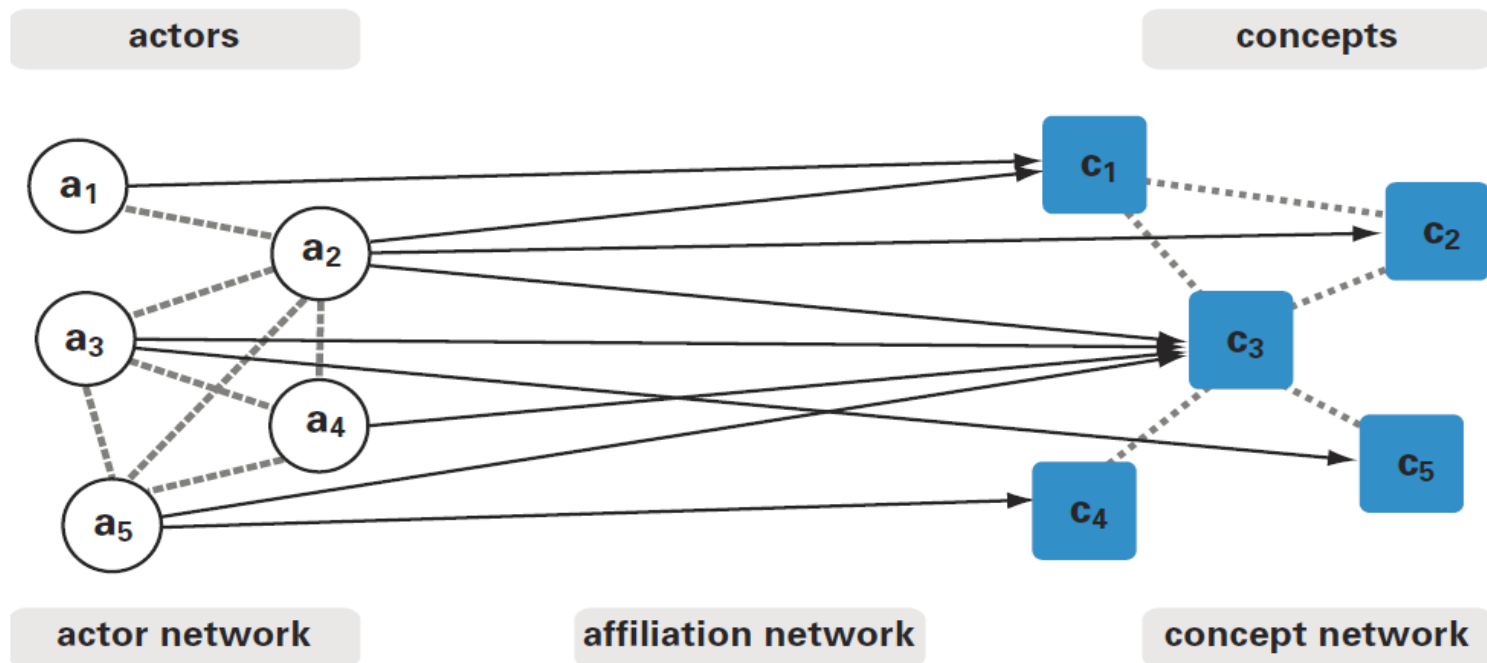
zdroj: SÚRAO

ČTK



Diskursivní síť

- Bipartitní síť sestávající se z aktérů a konceptů.



Rám

- **Sdílené interpretační schéma**, skrze které aktéři chápou a prosazují určitou verzi reality (např. Benford a Snow 2000).
- Cílem je prosazení vlastní interpretace tématu.
- Aktéři – prostřednictvím rámu – *strategicky* zdůrazňují či potlačují určité prvky tématu.
- **Operacionalizace**: skupina kódů, které jsou v podobné pozici vůči zbytku sítě.

(Diskursivní) koalice

- **(Diskursivní) koalice** je chápána jako “...*skupina aktérů, kteří sdílejí sociální konstrukt [rám].*” (Hajer 1995: 43)
- **Operacionalizace:** hustě propojený segment (**komunita**) v síti aktérů.

+ 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

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 nedodrželi, 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 Čeříňku, takže...

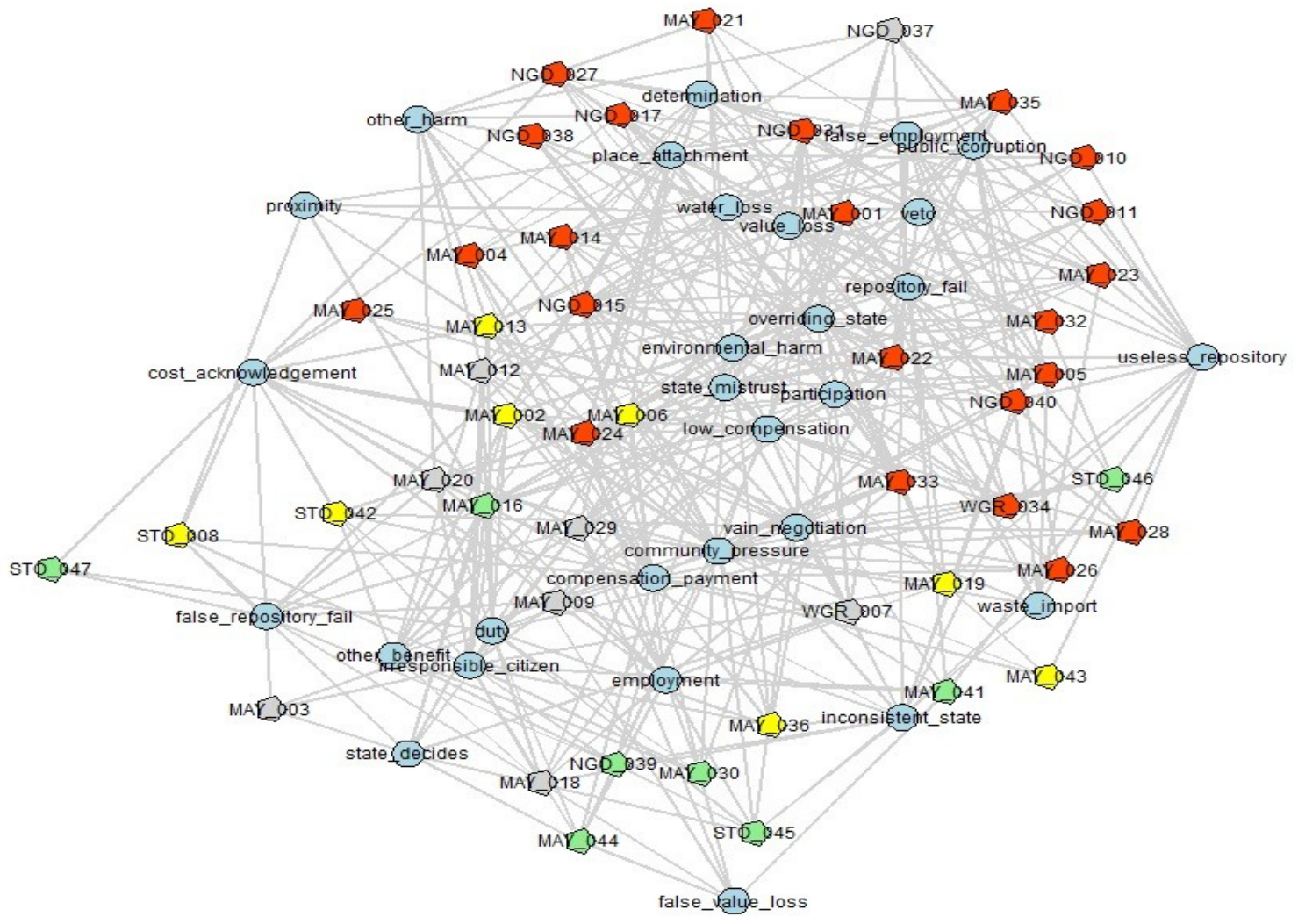
Incidenční matice

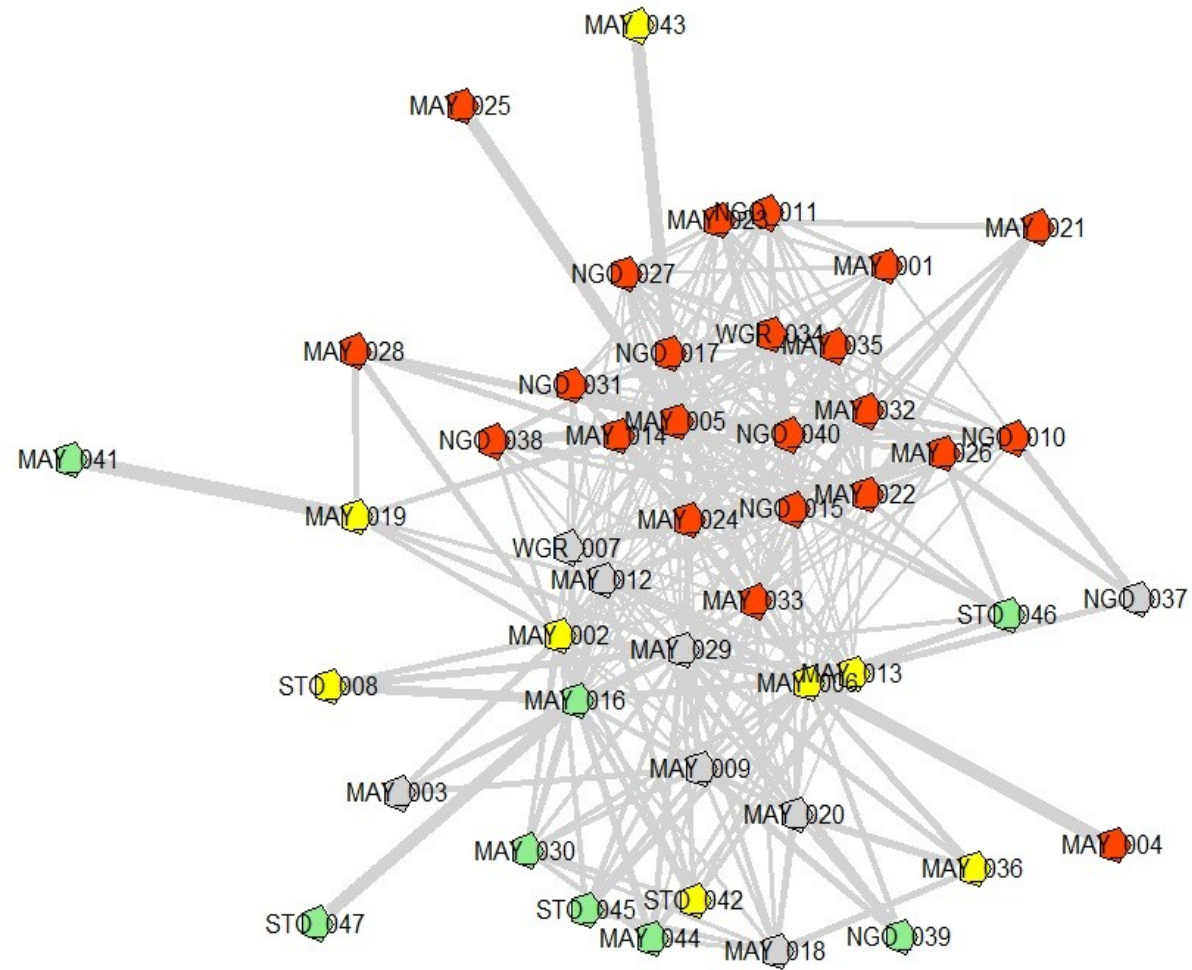
koncept 1

.

koncept j

aktér 1 (rozhovor)	buňka ij ukazuje kolikrát aktér i užil koncept j	0	3
.	.	.	.
.	.	.	.
.	.	.	.
aktér i	0	2	1

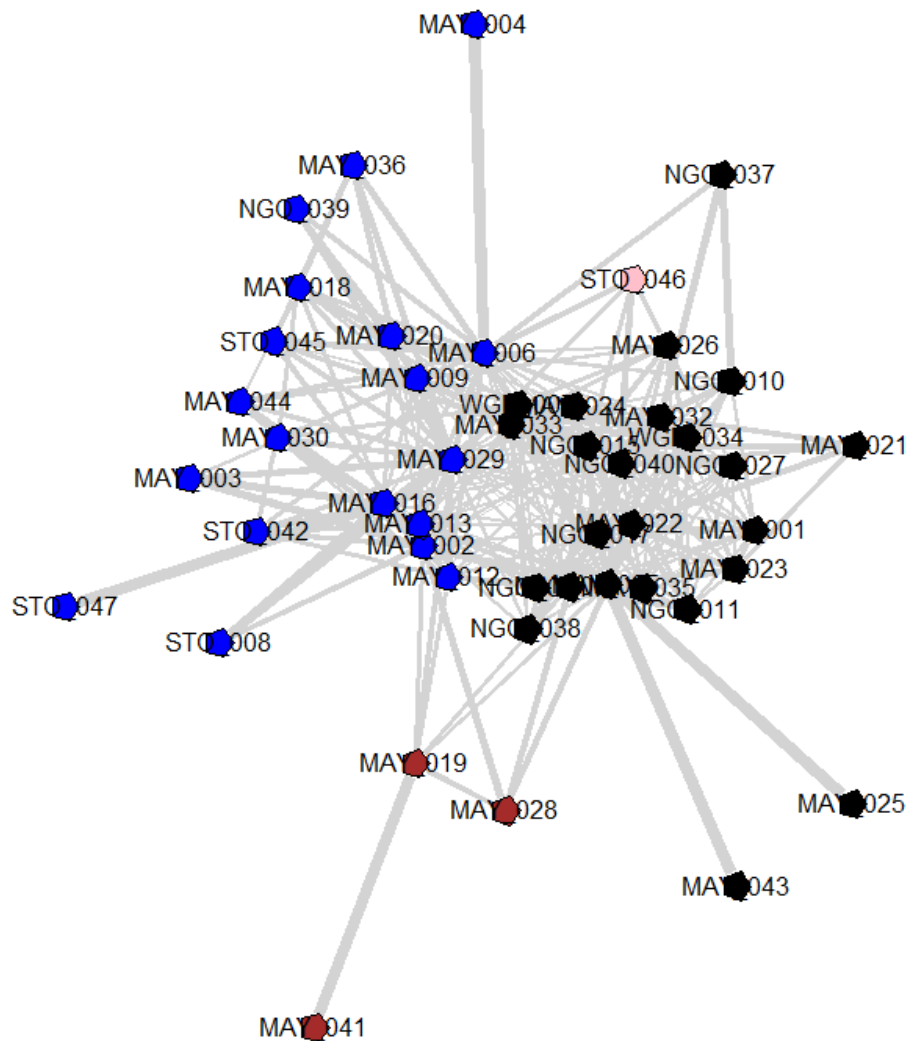




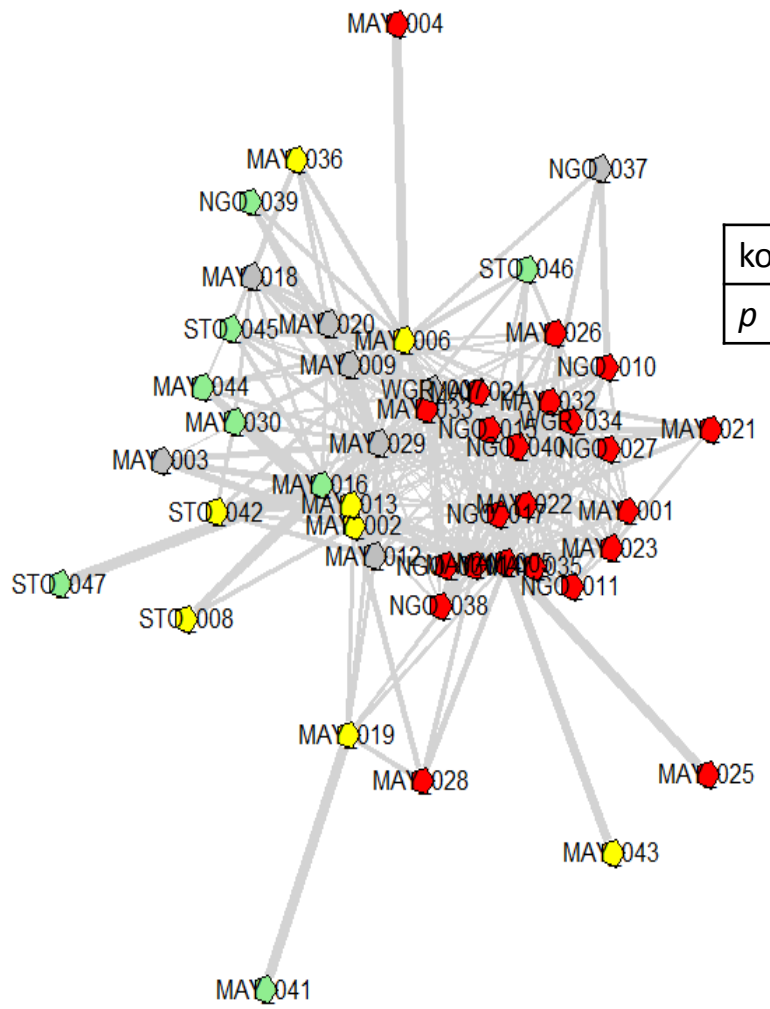
hustota	0.38
stupeň (centralizace)	0.36
mezilehlost (centralizace)	0.08

Komunita

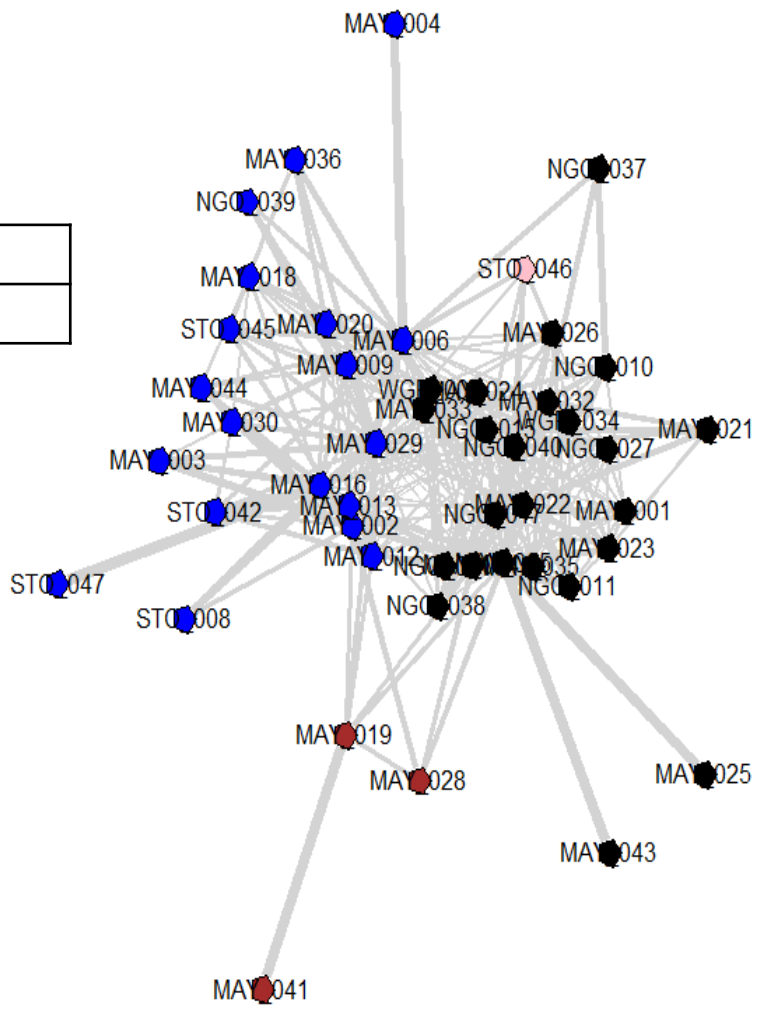
- **Komunita** je segment sítě tvořený množinou uzlů (členů), které jsou propojeny více navzájem nežli se zbytkem sítě (tj. s ne-členy komunity).

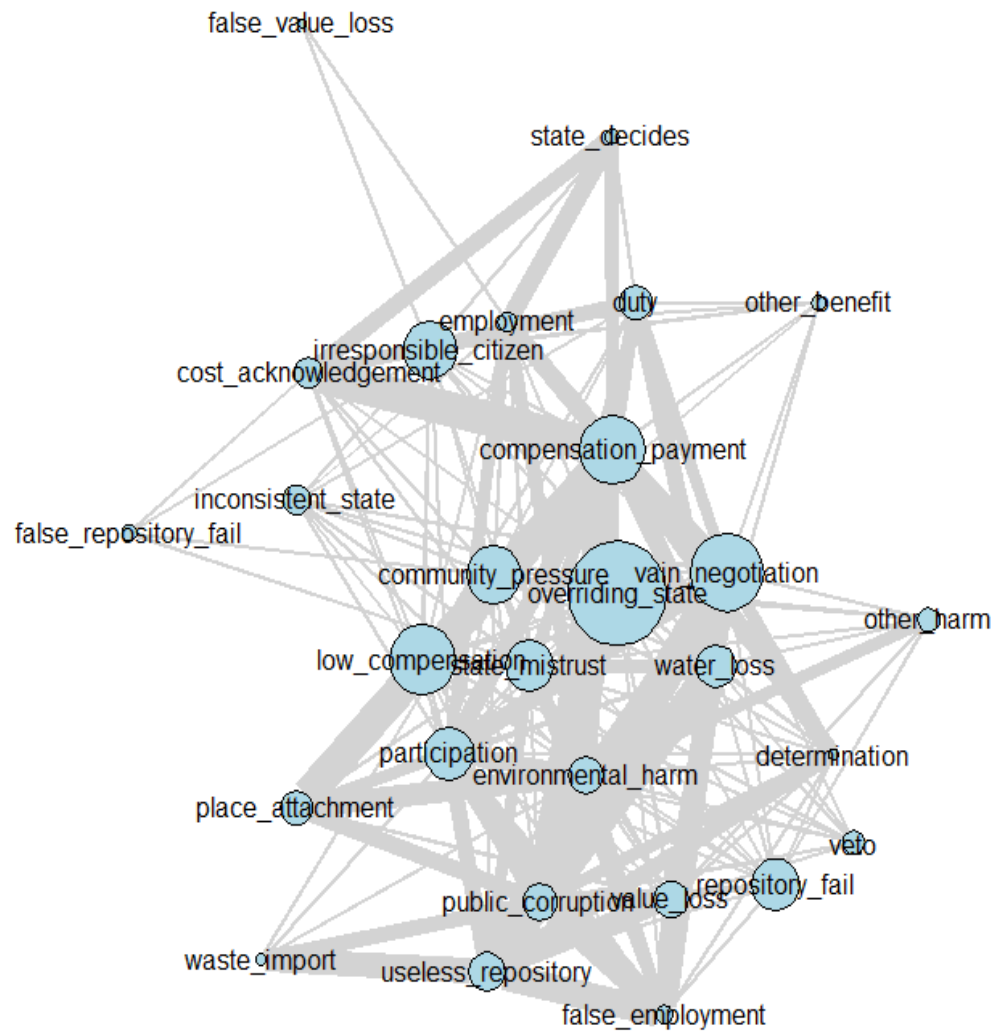


komunita 1	14 MAYs, 1 NGO, 4 STOs
komunita 2	15 MAYs, 9 NGOs
komunita 3	3 MAYs
"komunita" 4	STO_046
$p \leq 0.001$	

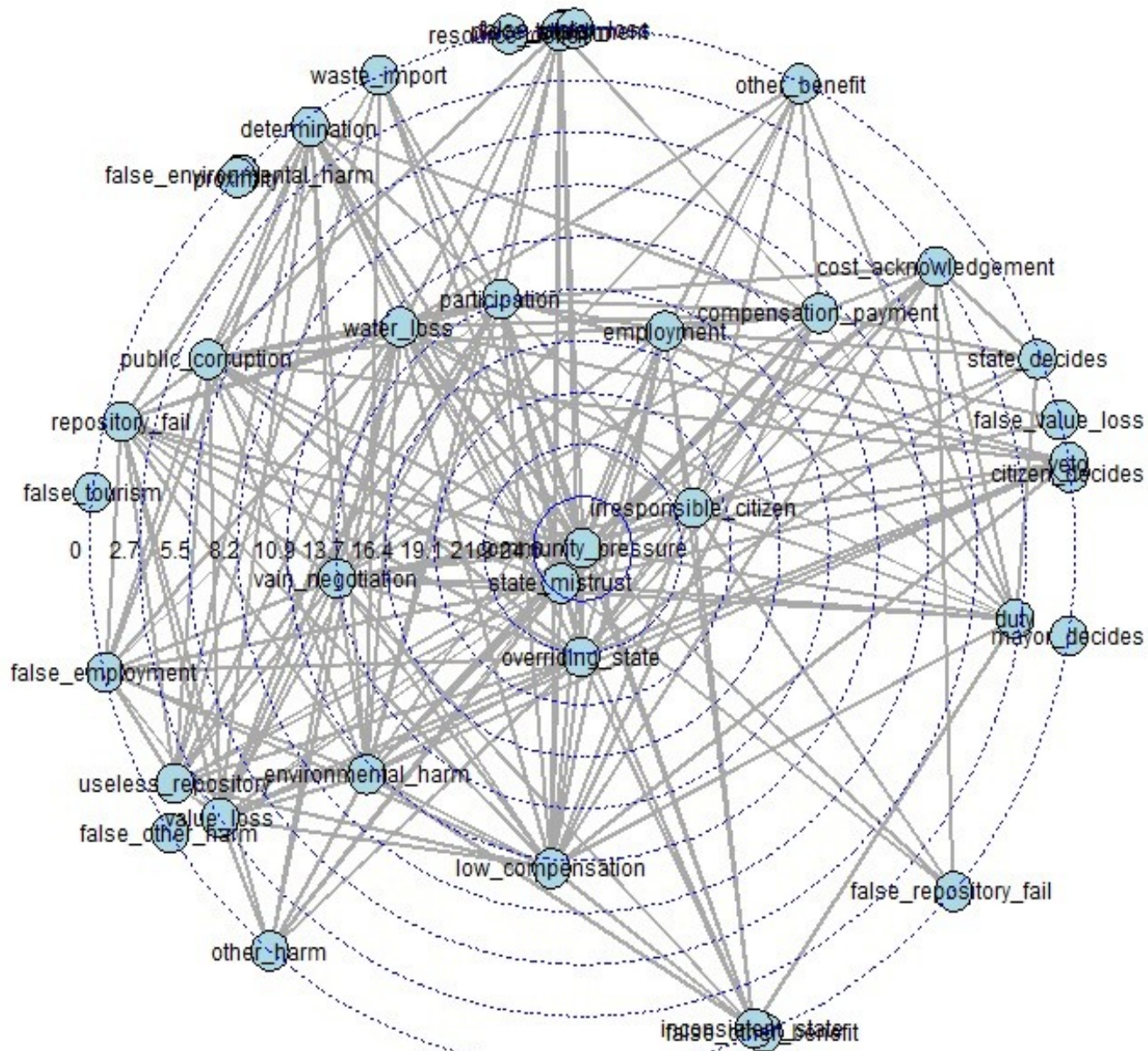


korelace	0.37
p	0.013

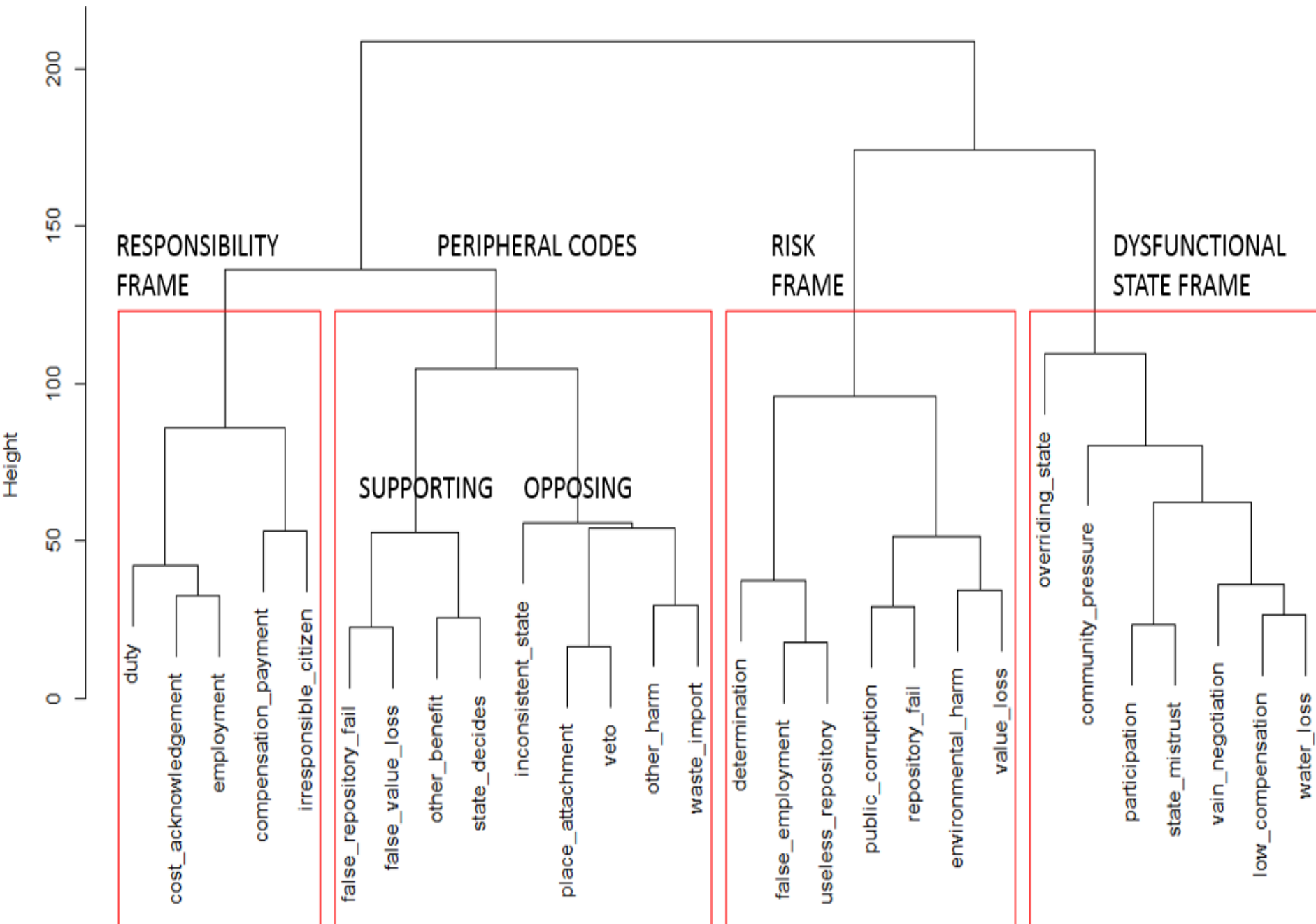




hustota	0.31
stupeň (centralizace)	0.43
mezilehlost (centralizace)	0.12

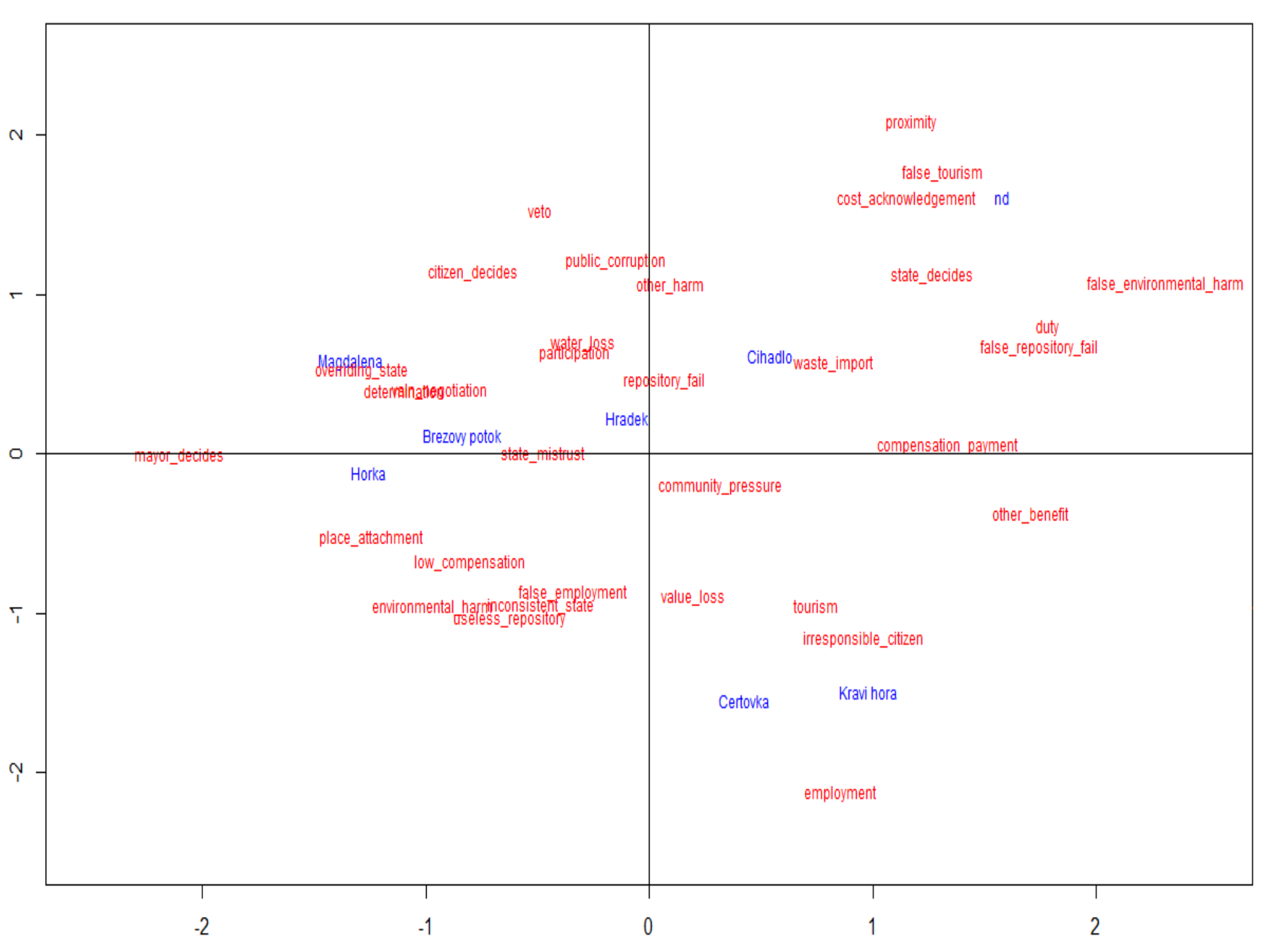


Euclidean distances (concepts)



Rekonstruované rámy

Odpovědnost	<ul style="list-style-type: none">• Spotřebováváme elektřinu z jaderných elektráren a vytváříme radioaktivní odpad.• Máme proto morální (a rovněž právní) závazek se s touto zátěží vyrovnat.• Hlubinné úložiště je jediným ekonomicky a technologicky proveditelným řešením.• Odpor vůči projektu tedy toto břemeno pouze přenáší na další generace.• Opozice vůči projektu je proto nezodpovědná a založená na iracionálních a emocionálních argumentech.• Stát má (legitimně) poslední slovo; lokality budou finančně kompenzovány.
Riziko	<ul style="list-style-type: none">• Proces výběru lokality a potenciální stavba a provoz úložiště je spojena s řadou rizik (environmetálních, ekonomických, sociálních, zdravotních).• Máme odpovědnost zachovat lokality pro další generace.• Projekt úložiště proto musí být zastaven či alespoň zpomalen, dokud nebude k dispozici alternativní (pravděpodobně technologické) řešení.
Dysfunkční stát	<ul style="list-style-type: none">• Stát není schopen projekt úložiště kompetentně a legitimně řídit.• Lokality nejsou efektivně zapojeny do vyjednávacího procesu.• Pracovní skupina je fasáda; konečné rozhodnutí bude náležet pouze státu.• Mezi stakeholdery je nedostatek důvěry a proces výběru lokality postrádá legitimitu.



R: výhody

- Freeware
- Open source
- Celosvětová aktivní komunita
- Flexibilní a rozvinuté



R komunita / zdroje

- Obrovské množství volně dostupných zdrojů
- R balíky (packages) / manuály
- Homepage: <http://cran.r-project.org>
- Komunitní fóra:
 - <http://stackoverflow.com>
 - <http://www.statmethods.net>
 - <http://www.r-bloggers.com>
- Youtube tutoriály:
<https://www.youtube.com/watch?v=qHfSTRNg6jE>
- Googlování (často nejrychlejší)

R knihovny / balíky

- Knihovna (library) / balík (package):
 - Rozšíření, které dodává novou funkcionalitu.
 - Knihovny musí být instalovány (pouze před prvním použitím) a nahrány.
 - Mezi knihovnami mohou vznikat konflikty (např. Různé funkce se stejnými názvy) – knihovny je možné deaktivovat.
 - Mezi knihovnami jsou často závislosti (pokud určitá knihovna využívá funkce z jiné knihovny).

R: nevýhody

- Méně přístupné než „klikací“ programy.
- Příprava dat může být náročná / zdlouhavá.
- R je pomalejší pro větší datové soubory.

R jazyk

- objektově orientované programování
 - **objekt:** případ určité datové třídy, s nímž může být dle určitých pravidel (metod) manipulováno.
- funkčně orientované programování
 - **funkce:** vztah spojující vstup(y) s výstupem(y).
- V R můžeme definovat objekty, aplikovat na ně funkce a naopak.

Typy dat (data classes)

- **Numeric:** spojitá numerická data (-1, 0.5, 10.49)
- **Integer:** diskrétní numerická data (-1, 0, 1, ...)
- **Character:** řetězce (strings) = “cokolivzávorkách+ 9”
- **Logical:** výstup logické operace
5 > 10 = **FALSE**
5 < 7 | 7 > 10 = **TRUE**

Typy dat: factor

- **Factor:** proměnná mající omezený počet diskrétních hodnot/úrovní (levels). Kategorická proměnná.
- Funkce faktorů mění vektor hodnot na vektor **faktorových hodnot** (vždy mají podobu **character**).
- Faktory mohou být neuspořádané (**unordered**) – nominální p., nebo uspořádané (**ordered**) – ordinální

```
p. > 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
```


R: objekty a funkce

- **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 funkce

- *word()* indikuje funkci
- `mean(vector)`

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

- Základní funkce (součást základního balíku R)
- Funkce balíků (součást určitého balíku)
- Uživatelské funkce (definované uživatelem)

R objekty

- **Vector**
 - Posloupnost (1-rozměrná) prvků **stejného typu dat.**
- **Matrix**
 - 2-rozměrné čtvercové uspořádání prvků **stejného typu dat.**
 - Pole (array): n-rozměrná matice.
- **List**
 - Vector který může obsahovat **různé typy dat.**
- **Data frame**
 - List vektorů o stejné délce.
 - Tabulární 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
```

Matrix

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

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
```

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  ...
.....
```