

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY



MODELOVANIE VPLYVU V SOCIÁLNEJ SIETI
POMOCOU DEGROOTOVHO MODELU

DIPLOMOVÁ PRÁCA

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

MODELOVANIE VPLYVU V SOCIÁLNEJ SIETI
POMOCOU DEGROOTOVHO MODELU

DIPLOMOVÁ PRÁCA

Študijný program: Ekonomicko - finančná matematika a modelovanie
Študijný odbor: 1114 Aplikovaná matematika
Školiace pracovisko: Katedra aplikovanej matematiky a štatistiky
Vedúci práce: doc. RNDr. Beáta Stehlíková, PhD.



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Katarína Ivanová
Študijný program: ekonomicko-finančná matematika a modelovanie
(Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: aplikovaná matematika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Modelovanie vplyvu v sociálnej sieti pomocou DeGrootovho modelu
Modelling influence in a social network using the DeGroot model

Cieľ: Sociálna sieť je množina uzlov (členov siete), ktoré sú prepojené vzťahmi, modelovanými hranami. Na začiatku práce sa vysvetlí modelovanie vplyvu, resp. dôležitosti uzlov siete pomocou rôznych mier centrality. Potom sa zameria na DeGrootov model, popisujúci vytváranie a dynamiku názorov, otázku ich konvergencie a meranie sociálneho vplyvu jednotlivcov v sociálnej sieti. V Krackhardtovej aplikácii na sieť manažérov firmy (Krackhardt, 1987), (Jackson, 2008) sa informácie na vytvorenie siete získali pomocou dotazníkov. V diplomovej práci sa zrealizuje podobný výskum zvolenej siete, pričom sa bude študovať aj robustnosť výsledkov vzhľadom na spôsob vytvárania siete (kritéria na existenciu hrany v grafe) z dotazníkových údajov.

Vedúci: RNDr. Beáta Stehlíková, PhD.
Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry: prof. RNDr. Daniel Ševčovič, CSc.

Dátum zadania: 10.02.2015

Dátum schválenia: 11.02.2015 prof. RNDr. Daniel Ševčovič, CSc.
garant študijného programu

.....
študent

.....
vedúci práce

Pod'akovanie Touto cestou úprimne ďakujem svojej vedúcej diplomovej práce doc. RNDr. Beáte Stehlíkovej, PhD. za trpezlivé vedenie, cenné rady a usmernenia pri tvorbe tejto práce. Ďakujem tiež mojej rodine, priateľom a blízkym za podporu a trpezlivosť.

Abstrakt

IVANOVÁ, Katarína: Modelovanie vplyvu v sociálnej sieti pomocou DeGrootovho modelu [Diplomová práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; školiteľ: doc. RNDr. Beáta Stehlíková, PhD., Bratislava, 2016, 83 s.

Práca sa zaoberá problematikou týkajúcou sa modelovania sietí a popisuje základné charakteristiky tejto oblasti. Cieľom tejto práce je voviesť čitateľa do danej témy, predstaviť DeGrootov model a následne tento model aplikovať na dáta získané formou dotazníkov a interpretovať zistené informácie pomocou zrozumiteľných grafov.

Kľúčové slová: sociálna sieť, centralita, DeGrootov model, dotazník, vplyv

Abstract

IVANOVÁ, Katarína: Modelling influence in a social network using the DeGroot model [thesis], Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; Supervisor: doc. RNDr. Beáta Stehlíková, PhD., Bratislava, 2016, 83 p.

This thesis deals with problems concerning the modelling of networks and describes the essential characteristics of this area. The aim of this work is to lead the reader into the topic, introduce the DeGroot model and then apply this model to the data obtained using questionnaires and interpret the information detected by comprehensible charts.

Keywords: social network, centrality, DeGroot model, questionnaire, influence

Obsah

Úvod	9
1 Motivácia	10
1.1 Motivačné príklady	10
2 Základy, definície a charakteristika	14
2.1 Základné pojmy	14
2.2 Cesty, prechádzky a cykly	17
2.2.1 Neorientovaná sieť	17
2.2.2 Orientovaná sieť	19
2.3 Ostatné charakteristiky siete	20
3 Grafové štruktúry	23
3.1 Mosty v Königsbergu	24
3.2 Rôzne druhy sietí	25
3.3 Komponenty siete	26
4 Centralita	28
4.1 Centralita stupňa	28
4.2 Centralita blízkosti	29
4.3 Centralita stredovej medzipolohy	30
4.4 Centralite na základe charakteristiky okolia a vlastného vektora	31
4.5 Manželstvá vo Florencii - rôzne miery centrality	32
5 DeGrootov model	34
5.1 Základná charakteristika	35
5.2 Aktualizácia v modeli	35
5.3 Konvergenca matice sociálneho vplyvu	37
5.4 Konvergenca matice aktualizácií - iný prístup	42
6 Cognitive Social Structures	45
6.1 Definícia CSS	45
6.2 Zoskupenie (Aggregations)	46

6.2.1	Rezy (Slices)	46
6.2.2	LAS	46
6.2.3	Štruktúry konsenzu (CS)	47
6.3	Príklad použitia CSS	47
7	Aplikácia Degrootovho modelu	52
7.1	Kolektív učiteliek v materskej škole	52
7.1.1	Analýza otázky o problematickom dieťati	52
7.1.2	Analýza otázky o súkromí	63
7.1.3	Miery centralít	71
7.2	Kolektív študentov	72
7.3	Nekonvergentné prípady	75
7.4	Konvergentné prípady	75
7.4.1	Otázka č.1: Za kým by si išiel, keby si potreboval poradiť s preberaným učivom?	75
7.4.2	Otázka č.2: Od koho by si si išiel požičať poznámky z prednášky/cvičenia, na ktorom si nebol?	76
7.4.3	Otázka č.5: Koho by si kontaktoval v prípade, že zháňaš prácu?	78
7.4.4	Porovnanie vplyvov z otázok č.1, č.2 a č.5	79
	Záver	81
	Zoznam použitej literatúry	82

Úvod

Každý z nás je súčasťou mnohých sociálnych sietí, medzi ktoré patrí napríklad rodina, kolektív v škole či v práci, skupina priateľov a pod. V týchto sieťach existujú rôzne duhy vzťahov, ktoré je možné preskúmať. Dá sa zistiť, kto je v rámci skupiny priateľov najobľúbenejší, za kým by sa v práci prišli poradiť kolegovia v prípade nejasností, ktorý člen rodiny bude vyberať dovolenkovú destináciu a mnoho iného.

V tejto práci predstavíme, ako zistiť vplyv jednotlivca a určiť, kto má v sieti najväčší vplyv na rozhodovanie ostatných. Jeden z prístupov, ako popísať vytváranie názorov a meranie vplyvu členov siete, je popísaný pomocou DeGrootovho modelu. Cieľom tejto práce je objasniť terminológiu týkajúcu sa sietí vo všeobecnosti, predstaviť DeGrootov model a následne ho aplikovať na nami získané dáta z dotazníkov.

Práca je rozdelená do siedmich kapitol. V prvej kapitole čitateľovi v krátkosti načrtneme niekoľko príkladov, kde sa dá zužitkovať teória sietí. Druhá kapitola spolu s treťou kapitolou obsahujú potrebné definície týkajúce sa problematiky sociálnych sietí a ich štruktúry. Dôležitosť jednotlivých členov siete sa dá vysvetliť prostredníctvom rôznych mier centrality, ktoré sú prebrané v štvrtej kapitole. Piata kapitola je zameraná na spomínaný DeGrootov model. Aplikácia na sieť manažérov firmy, ktorú vo svojej práci popísal Krackhardt, je predstavená v šiestej kapitole. Posledná siedma kapitola obsahuje všetky získané a zistené údaje spolu s aplikáciou DeGrootovho modelu na dáta zozbierané pomocou dotazníkov.

1 Motivácia

Prvá otázka, ktorá nám môže napadnúť, je: Prečo sa zaujímať o siete? Odpoveď nájdeme napríklad v [13], kde sa uvádza, že z pohľadu sociológov je mnoho ekonomických, politických a sociálnych interakcií zakotvených práve v sociálnom prostredí. Štruktúra týchto vzťahov je dôležitá pri určovaní, ako sa ľudia správajú a aké sú výsledky tohto správania. Interakcie, ktoré sú formované štruktúrou vzťahov, sú napríklad:

- obchod s tovarom a službami,
- zdieľanie informácií, láskavostí, risku,
- prenos vírusov, názorov,
- prístup k informáciám o práci,
- voľby správania, vzdelania,
- politické a obchodné spojenectvá.

Existujú mnohé štúdie, ako sa uvádza v [5], ktoré sa zaoberajú sociálnymi a ekonomickými sieťami. Vďaka veľkému počtu takýchto štúdií bolo preto možné dozvedieť sa veľa o štruktúre sietí. Na druhej strane je potrebné nezabúdať, že skúmanie štruktúry nejakej sociálnej siete je naozaj náročná úloha. Táto úloha totiž obsahuje významné prekážky spojené s tým, ako definovať a merať väzby a vzťahy v sieti. Základným nástrojom na odhad sociálnej siete je zvyčajne prieskum alebo rozhovory zúčastnených strán. Vzhľadom na to, že jednotlivci môžu mať stovky až tisíce vzťahov, pripomenutie tých relevantných s akoukoľvek požadovanou presnosťou je ťažké.

1.1 Motivačné príklady

Na začiatok ukážeme niekoľko názorných príkladov, ktoré súvisia so skúmaním vzťahov v danej sociálnej sieti. Jedná sa o rôznorodé využitie a takisto rôzne typy sociálnych sietí.

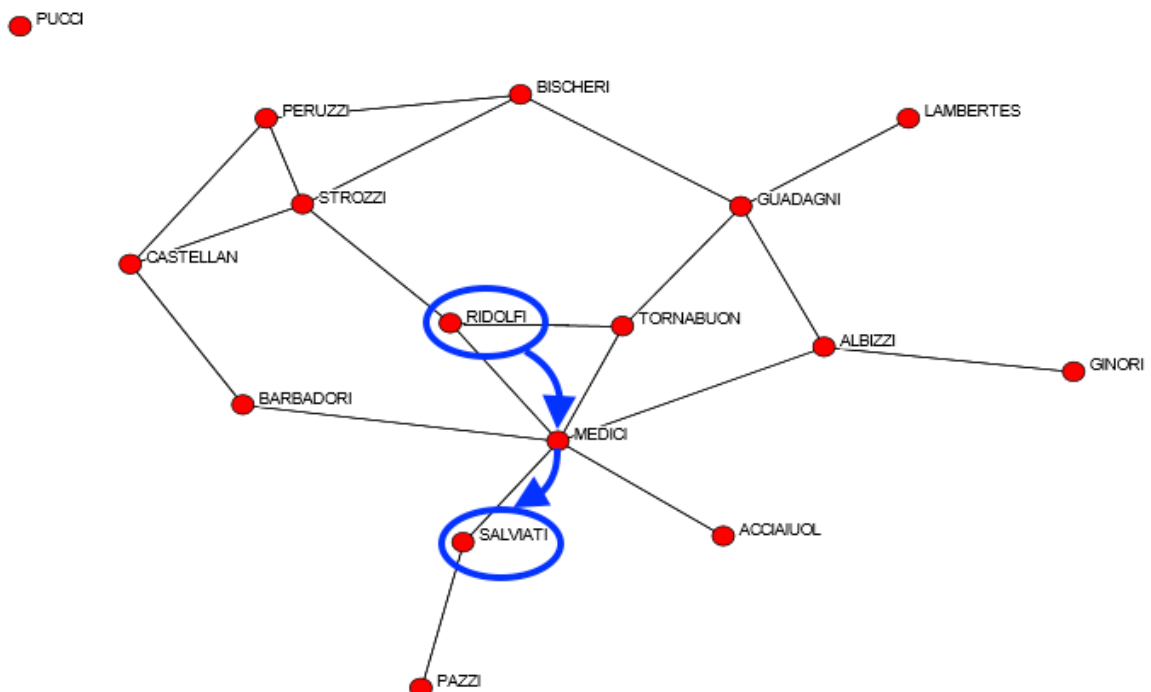
Príklad 1: Manželstvá vo Florencii

Dáta, ktoré nájdeme spracované v [10], sú zamerané na vzťahy medzi rodinami vo Florencii v renesančnom Taliansku v tridsiatych rokoch pätnásteho storočia. Autori tu analyzujú centralizáciu predchodcu politických strán a elitných sietí, ktoré stoja za

vznikom renesančného štátu. My sa zameriame na to, akú úlohu zohrávala sociálna sieť v prípade vzostupu rodiny Medici v 15. storočí. Spojenie v sociálnej sieti medzi dvoma rodinami predstavuje sobáš.

V [5] sa dočítame, že hromadenie moci bolo v tomto období vo Florencii organizované členom rodiny Medici. Cosmo de' Medici sa o to postaral, aj keď jeho rodina nebola najbohatšou spomedzi rodín vo Florencii, ani nemala najlepšie politické styky. Upevnil politickú a ekonomickú moc využitím centrálnej polohy rodiny Medici v sieťach rodinných sobášov, ekonomických vzťahov a politického patronátu. Jeho chápanie a postavenie v týchto sociálnych sieťach mu umožnili vybudovať a riadiť predchodcu politickej strany, zatiaľ čo sa ostatné významné rodiny len prizerali.

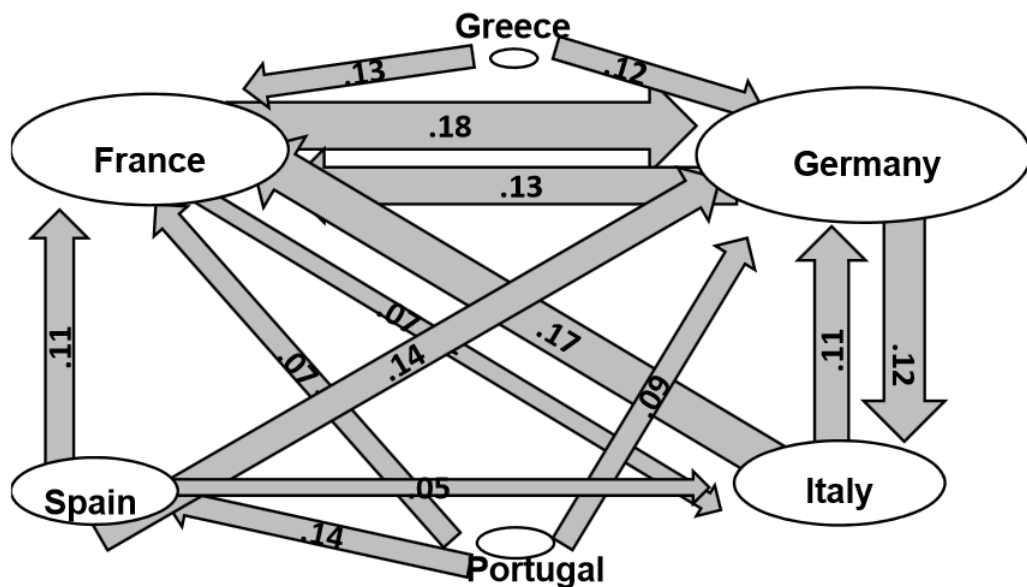
V tejto práci sa budeme zaoberať aj centralitou vrchola v rámci siete. Jednou možnosťou, prirodzenou v tomto prípade, je pod centralitou chápať množstvo ciest medzi danou rodinou a ostatnými rodinami nachádzajúcimi sa v sieti. Ak sa Ridolfi a Salviati chcú v sieti dať dokopy, musia prejsť cez rodinu Medici. Bližšie to môžeme vidieť vyznačené na obrázku 1.



Obr. 1: Príklad 1: Manželstvá vo Florencii [13]

Príklad 2: Dlhy- krajiny Európy

Tento príklad z [4] je ekonomického charakteru. Na obrázku 2 vidíme vzťahy medzi šiestimi európskymi krajinami: Nemeckom, Francúzskom, Gréckom, Talianskom, Portugalskom a Španielskom. Vzájomné vzťahy vyjadrujú, aká časť štátneho dlhu je v držbe subjektov vnútri inej krajiny. Príklad, ako rozumieť tomuto grafu na obrázku 2 je, že 18% Francúzskeho štátneho dlhu je držaný v Nemecku. Táto sieť bola východiskom k analýze toho, ako prenos finančných šokov a kríz v jednej krajine môže viesť k znehodnoteniu alebo problémom v inej krajine.



Obr. 2: Príklad 2: Dlhy- krajiny Európy [4]

Príklad 3: Trh práce

Na pochopenie fungovania trhov potrebujeme porozumieť, čo predstavujú jednotlivé väzby v sieti. Oblasť, ktorej bola venovaná pozornosť, je práve úloha sietí na trhu práce - ako sa ľudia dozvedajú o práci. Meyer a Schultz sa tejto otázke venovali vo svojej práci [9] z roku 1951, ktorá opisuje túto situáciu v textilnom priemysle. Z prieskumu vyplynulo, že 62% sa dozvedelo o danej pozícii v textilnom priemysle od niekoho z priemyslu, 23% priamou ponukou a 15% sa dozvedelo o práci cez agentúru, reklamu a pod.

Príklad 4: Predikcie

V zásade si podľa [13] môžeme klásť aj iné otázky. Ak má sieť veľa väzieb, čo to spraví s celkovou štruktúrou siete? Informácie tohto typu nám ukážu, akým smerom sa vydať pri vytváraní predikcií. Ak by sme chceli napríklad prísť s novým návrhom, ako potlačiť chrípku či epidémiu, vynára sa nám jedna podstatná otázka. Aká účinná musí byť vakcína, aby sme obmedzili rozsah epidémie? Práve na túto otázku môžeme dať odpoveď pomocou analýzy sietí.

Koľko môže byť sietí v tridsaťčlennej triede?

Na základe predchádzajúcich príkladov, v ktorých sme videli rôzne siete, nás môže zaujímať, koľko rôznych sietí vôbec môže existovať.

Pozrime sa preto, rovnako ako v [13], na malé spoločenstvo, akým je práve trieda v škole. Ak má táto trieda 30 členov, koľko možných rôznych sietí reprezentujúcich priateľstvo by tu mohlo existovať? Mohla by to byť prázdna sieť, teda nikto sa s nikým nepriatelí. Na druhej strane by to ale mohla byť aj úplná sieť, v ktorej sa priatelia každý s každým. Počet možných sietí je nesmierne veľký.

Najskôr sa pozrime na to, aký je počet rôznych priateľstiev, ktoré by mohli byť prítomné v tejto triede. Vypočítame to jednoducho, keďže každé priateľstvo je určené dvojicou spolužiakov. Týchto možných priateľstiev je teda toľko, koľkými spôsobmi môžeme vybrať dvoch spolužiakov spomedzi 30, teda

$$\binom{30}{2} = \frac{30!}{28!2!} = \frac{29 \cdot 30}{2} = 435$$

Dostávame teda 435 rôznych priateľstiev, možných väzieb v triede, z ktorých každá väzba môže byť buď prítomná alebo nie. Z toho vyplýva, že máme 2^{435} možných prítomných sietí, čo je už naozaj veľké číslo, rádovo 10^{130} .

2 Základy, definície a charakteristika

Slovo sieť pokrýva veľa rôznych aplikácií. Je preto potrebné zvoliť si jednu z nich. V tejto časti definujeme základné pojmy, s ktorými budeme ďalej pracovať. V rôznych zdrojoch môžu byť pojmy definované rôzne. Siete vzťahov sú rôznych tvarov a veľkostí, a preto existuje viac spôsobov ich reprezentácie. My sa v tejto kapitole zameriame na definície z [5] a [13], ktoré sú postačujúce na popísanie viacerých aplikácií a zároveň sú aj relatívne ľahko intuitívne pochopiteľné.

2.1 Základné pojmy

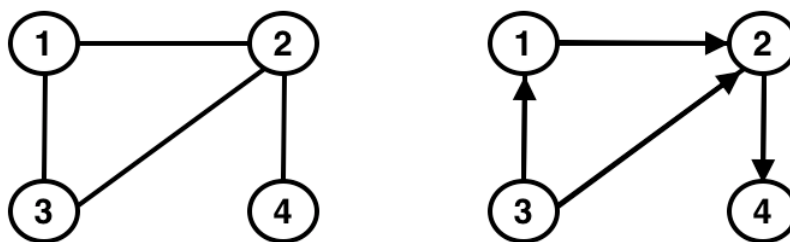
Vrchol patrí medzi základné pojmy z teórie sietí. Zvyčajne tento pojem predstavuje jednotlivca, človeka, ktorý sa rozhoduje. Vrcholom môžeme myslieť aj skupinu jednotlivcov, ako sú organizácie, krajiny či firmy. Teda množina vrcholov je konečná a definovaná ako $N = \{1, \dots, n\}$.

Medzi rôznymi vrcholmi môžu byť prítomné rôzne druhy vzťahov, napr. v príklade 1 z kapitoly 1 v sieti rodín vo Florencii predstavujú tieto väzby práve manželstvo, v ďalšom príklade to bolo priateľstvo medzi spolužiakmi v triede. Môže to byť ale aj napríklad mailová alebo telefónna sieť. Tieto väzby budeme nazývať **hrany**.

Ďalším krokom je potreba definovať, čo predstavuje sieť. **Sieť** je reprezentovaná **grafom**, ktorý je charakterizovaný množinou vrcholov a hrán. Graf a sieť teda predstavujú to isté. V tejto práci budeme používať oba pojmy podľa toho, ktorý bude v danej situácii vhodnejší.

Grafy môžu byť neorientované a orientované. **Neorientovaný graf** je taký graf, v ktorom buď sú dva vrcholy spojené alebo nie sú. V takýchto grafoch nie je možné, aby sa jeden vrchol spájal s druhým bez toho, že je druhý vrchol spojený s prvým. Takáto vlastnosť platí vo všeobecnosti pre veľa spoločenských aj ekonomických vzťahov, ako sú partnerstvá, priateľstvá, spojenectvá či známosti.

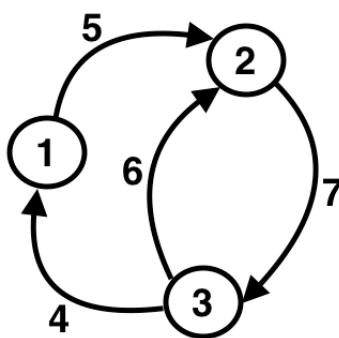
Orientovaný graf je graf, v ktorom jeden vrchol je spojený s druhým bez toho, aby to muselo platiť aj opačne. Situácie, ktoré je treba modelovať pomocou orientovaných grafov, sú napríklad citácie, sledovanie na sociálnej sieti či odkazy z jednej webovej stránky na druhú. Obidva spomínané typy grafov vidíme na obrázku 3.



Obr. 3: Neorientovaný a orientovaný graf

Rozdiel medzi orientovanými a neorientovanými grafmi nie je len formalitou. Je zásadný pre rôzne analýzy, keďže aplikácia a modelovanie týchto dvoch typov je úplne odlišná. Najmä vtedy, keď sú jednotlivé väzby vzájomné, vo všeobecnosti platí, že na vytvorenie a udržanie vzťahu je potrebný spoločný súhlas. Vidíme to na príklad pri vzniku obchodných partnerstiev, kde sa na tom musia dohodnúť obe zúčastnené strany. Na druhej strane platí, že jedna webová stránka môže odkazovať na druhú bez jej súhlasu. Tieto rozdiely majú za následok niektoré základné odlišnosti v modelovaní siete a tiež rôzne závery o tom, ktoré siete vzniknú, ktoré sú optimálne a pod.

Priblížime si ešte pojmy **vážená** a **nevážená** sieť. Vo vázenej sieti majú väzby svoju váhu, napr. koľko hodín trávajú spolu dvaja ľudia počas týždňa. Tento príklad je zobrazený na obrázku 4. V opačnom prípade majú jednotlivé vzťahy váhu len 0 alebo 1 a predstavujú to otázky typu: „Sú dvaja ľudia priateľmi na niektorej sociálnej sieti?“ [13]



Obr. 4: Vážená sieť

Graf, ktorý reprezentuje sieť je definovaný nasledovne. Graf (N, g) sa skladá z množiny vrcholov $N = \{1, \dots, n\}$ a $n \times n$ matice g reálnych čísel, kde g_{ij} reprezentuje vzťah medzi i a j . Táto matica sa nazýva **matica susednosti**, nakoľko vymenúva, ktoré vrcholy sú navzájom prepojené.

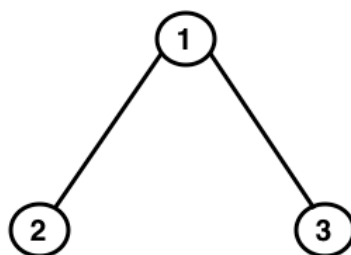
V prípade neváženej siete nadobúda matica g , iba hodnoty 0 a 1. Ak existuje hrana medzi vrcholom i a j , potom $g_{ij} = 1$, inak $g_{ij} = 0$. Pre váženú sieť platí, že v prípade existencie hrany medzi vrcholom i a j dostaneme g_{ij} rovné prislúchajúcej váhe, inak $g_{ij} = 0$.

Ak je sieť neorientovaná, potom $g_{ij} = g_{ji}$ pre všetky i a j . Ak je sieť orientovaná, tak pre nejakú dvojicu i, j môže platiť $g_{ij} \neq g_{ji}$, keďže sa môže stať, že existuje orientovaná hrana z vrcholu i do vrcholu j , ale hranu z j do i sieť neobsahuje.

Napríklad, ak $N = \{1, 2, 3\}$, potom

$$g = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

popisuje neorientovanú a neváženú sieť s väzbami medzi vrcholom 1 a 2, 1 a 3, ale so žiadnym spojením medzi vrcholmi 2 a 3, ktorú vidno na obrázku 5.



Obr. 5: Sieť s tromi vrcholmi

Matice susednosti pre grafy z obrázkov 3 a 4, ktoré sú príkladmi na neorientovaný graf, orientovaný graf a váženú sieť, sú:

$$g_1 = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, g_2 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, g_3 = \begin{pmatrix} 0 & 5 & 0 \\ 0 & 0 & 7 \\ 4 & 6 & 0 \end{pmatrix}.$$

Graf je možné popísať aj iným spôsobom. Namiesto zobrazenia g ako $n \times n$ maticu je niekedy jednoduchšie popísať graf vymenovaním všetkých jeho hrán. Môžeme teda vyjadriť našu maticu g z predchádzajúceho príkladu ako $g = \{\{1, 2\}, \{1, 3\}\}$ alebo

zjednodušene $g = \{12, 13\}$. Teda ij reprezentuje hranu spájajúcu vrchol i a j a potom $ij \in g$ znamená, že i a j sú spojené v sieti g .

2.2 Cesty, prechádzky a cykly

Veľký záujem v rámci siete vzťahov pochádza podľa [5] z toho, že jednotlivé vrcholy, napr. ľudia, majú prospech alebo trpia z nepriamych vzťahov v rámci siete. V skutočnosti totiž priatelia môžu sprostredkovať rôzne láskavosti svojim priateľom a informácie sa šíria cez väzby v sieti. Na druhej strane to platí aj v negatívnom zmysle, a teda v zamedzení prístupu k informáciám. Na zachytenie týchto nepriamych interakcií je potrebné v sieti správne modelovať jednotlivé cesty. Zvlášť sa pozrieme na definovanie pojmov pre neorientovanú a orientovanú sieť, ktorú nájdeme práve v [5].

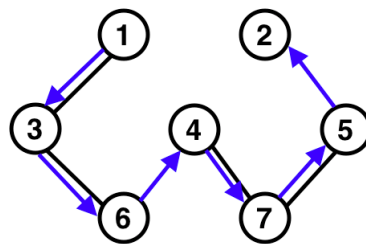
2.2.1 Neorientovaná sieť

V tejto časti definujeme, čo v neorientovanej sieti znamenajú pojmy: cesta, prechádzka a cyklus.

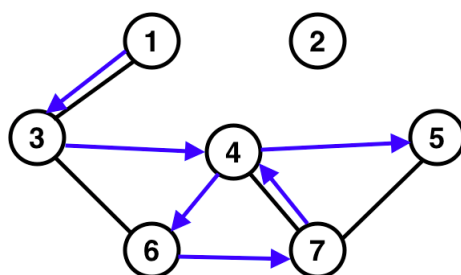
Cesta v sieti g medzi vrcholmi i a j je taký sled hrán $i_1i_2, i_2i_3, \dots, i_{K-1}i_K$, že $i_ki_{k+1} \in g$ pre každé $k \in \{1, \dots, K-1\}$, pričom $i_1 = i$ a $i_K = j$ a každý vrchol v slede i_1, \dots, i_K je odlišný.

Prechádzka v sieti g medzi vrcholmi i a j je taký sled hrán $i_1i_2, i_2i_3, \dots, i_{K-1}i_K$, že $i_ki_{k+1} \in g$ pre každé $k \in \{1, \dots, K-1\}$, pričom $i_1 = i$ a $i_K = j$.

Na obrázku 6 a 7 vidíme, vyznačené modrými šípkami, príklady na cestu a prechádzku v neorientovanej sieti. Rozdiel medzi cestou a prechádzkou je v tom, či sú všetky zahrnuté vrcholy rozdielne alebo nie. Pri prechádzke môžeme prejsť jeden vrchol viackrát, zatiaľ čo cesta je prechádzka, pri ktorej nikdy neprejdeme jedným vrcholom viac ako raz.

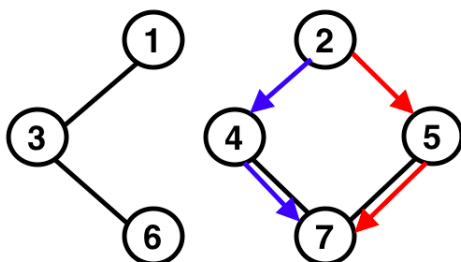


Obr. 6: Cesta (a prechádzka) z vrcholu 1 do vrcholu 2: 1, 3, 6, 4, 7, 5, 2.



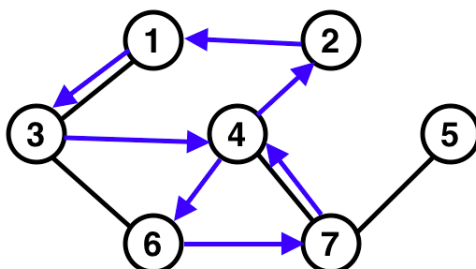
Obr. 7: Prechádzka, ktorá nie je cestou z vrcholu 1 do vrcholu 5: 1, 3, 4, 6, 7, 4, 5.

Často treba nájsť **najkratšiu cestu** medzi zadanými vrcholmi, t.j. takú, ktorá obsahuje čo najmenej hrán. Na obrázku 8 vidíme, že medzi vrcholmi 2 a 7 sú dve najkratšie cesty, a to sled vrcholov 2, 4, 7 (vyznačené modrou) a 2, 5, 7 (vyznačené červenou).



Obr. 8: Najkratšie cesty medzi vrcholom 2 a 7: 2, 4, 7 a 2, 5, 7.

Cyklus je prechádzka $i_1 i_2, \dots, i_{K-1} i_K$, ktorá začína aj končí v tom istom vrchole, teda $i_1 = i_K$ a všetky ostatné vrcholy sú rôzne, t.j. $i_k \neq i_{k'}$, keď $k < k'$. Tým pádom je cyklus prechádzka, v ktorej jediný opakujúci sa vrchol je začiatočný, resp. končiaci vrchol, ako je zobrazené na obrázku 9.



Obr. 9: Cyklus (a prechádzka) z vrcholu 1 do vrcholu 1: 1, 3, 4, 6, 7, 4, 2, 1.

Pri definovaní $g_{ii} = 0$, získame vďaka mocninám matice g informácie o počte prechádzok medzi jednotlivými vrcholmi v sieti. Zoberme ako príklad neorientovanú sieť

z obrázku 3, ktorej matica susednosti g vyzerá nasledovne

$$g = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix},$$

potom jej druhá mocnina g^2 , kde

$$g^2 = \begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 3 & 1 & 0 \\ 1 & 1 & 2 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix},$$

dáva informácie o počte prechádzok dĺžky 2 medzi vrcholmi, v matici vyjadrenými pomocou i a j . Teda medzi vrcholmi 1 a 4 je iba jedna prechádzka dĺžky 2, a to z vrcholu 1 do vrcholu 2 a odtiaľ do vrcholu 4. Podobne aj matica

$$g^3 = \begin{pmatrix} 2 & 4 & 3 & 1 \\ 4 & 2 & 4 & 3 \\ 3 & 4 & 2 & 1 \\ 1 & 3 & 1 & 0 \end{pmatrix}$$

hovorí o tom, koľko prechádzok dĺžky 3 je medzi jednotlivými vrcholmi. Medzi vrcholmi 1 a 2 sú štyri prechádzky dĺžky 3. Rovnako to platí aj pre vyššie mocniny matice g , a teda k -ta mocnina, g^k udáva počet prechádzok dĺžky k v rámci siete. Dôkaz matematickou indukciou nájdeme napr. v [8].

Ďalšiu aplikáciu mocnín matice pre špeciálny typ vážených sietí budeme mať neskôr v kapitole 5.

2.2.2 Orientovaná sieť

V prípade orientovaných sietí sa definície cesty, prechádzky a cyklu odlišujú od definícií pre neorientované siete v tom, že pri hľadaní postupnosti hrán musíme sledovať smer existujúcej hrany. Hrana môže existovať v smere z prvého vrcholu do druhého, opačne alebo obojsmerne. Platí to napríklad v prípade siete väzieb medzi webovými stránkami. Ak nás napríklad zaujíma, či sa na jednu webovú stránku môžeme dostať z inej

nasledovaním odkazov na prepojených webových stránkach, musíme sa na to pozerat' ako na orientovanú sieť.

2.3 Ostatné charakteristiky siete

V tejto časti definujeme ďalšie pojmy, ktoré popisujú rôzne vlastnosti siete. Všetky tieto definície nájdeme v [5].

Okolie vrcholu i je množina vrcholov, ktoré sú vo väzbe s vrcholom i , teda platí

$$N_i(g) = \{j : g_{ij} = 1\}.$$

Ak si vezmeme množinu vrcholov S , potom **okolie množiny** S je zjednotením okolí jej členov, t.j.

$$N_S(g) = \bigcup_{i \in S} N_i(g) = \{j : \exists i \in S, g_{ij} = 1\}.$$

V širšom kontexte môžeme hovoriť aj o rozšírenom okolí, napr. o vrchoch, ktoré sú dosiahnuteľné prechádzkou nie dlhšou ako 2. Potom **2-okolie** vrcholu i je definované

$$N_i^2(g) = N_i(g) \cup \left(\bigcup_{j \in N_i(g)} N_j(g) \right).$$

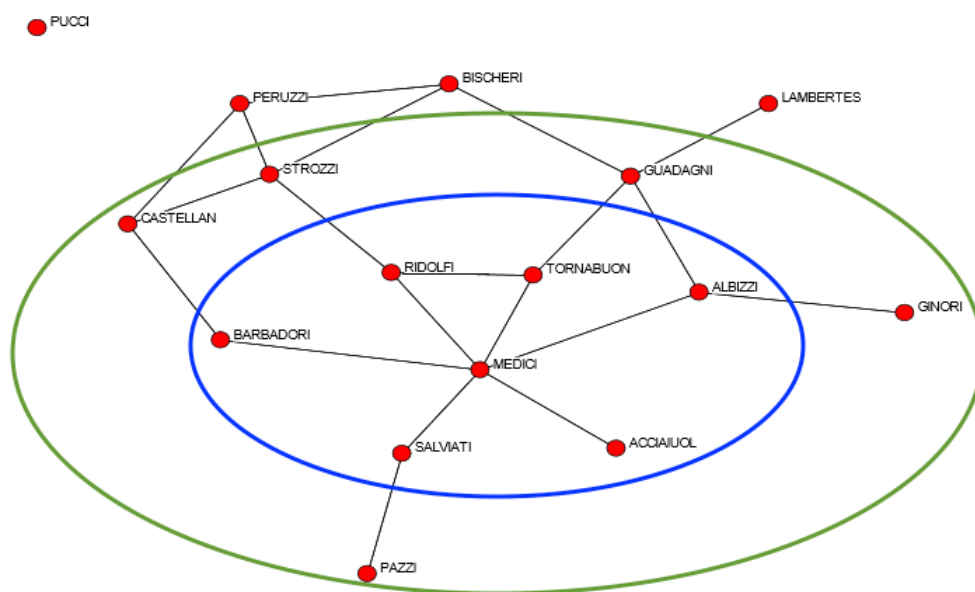
Indukciou, všetky vrcholy, ktoré sú dosiahnuteľné z i prechádzkou dĺžky nie viac ako k vytvárajú **k-okolie** vrcholu i , teda

$$N_i^k(g) = N_i(g) \cup \left(\bigcup_{j \in N_i(g)} N_j^{k-1}(g) \right).$$

Podobná definícia **k-okolia** platí pre akúkoľvek množinu S , takže $N_S^k(g) = \bigcup_{i \in S} N_i^k(g)$ je množina vrcholov, ktoré môžu byť dosiahnuteľné z niektorého vrcholu v S prechádzkou nie dlhšou ako k .

Všetky tieto spomenuté definície tiež platia pre orientované siete, kde ale vrcholy v $N_i^k(g)$ sú také, ktoré sú dosiahnuteľné z i pomocou orientovanej prechádzky.

Na príklade rodín z Florencie z kapitoly 1, môžeme vidieť, že do okolia vrcholu rodiny Medici patria rodiny Ridolfi, Tornabuon, Albizzi, Acciaiuol, Salviati a Barbadori, na obrázku 10 ohraničené modrou. 2-okolie rodiny Medici tvoria okrem spomenutých rodín ešte rodiny Pazzi, Castellan, Strozzi, Guadagni a Ginori, na obrázku 10 ohraničené zelenou.



Obr. 10: Okolie vrcholu Medici

Stupeň vrcholu je počet hrán, ktoré zahŕňajú daný vrchol, čo je mohutnosť susednosti vrcholu. Potom stupeň vrcholu i v sieti g označujeme $d_i(g)$ a je definovaný

$$d_i(g) = \#\{j : g_{ij} = 1\} = \#N_i(g).$$

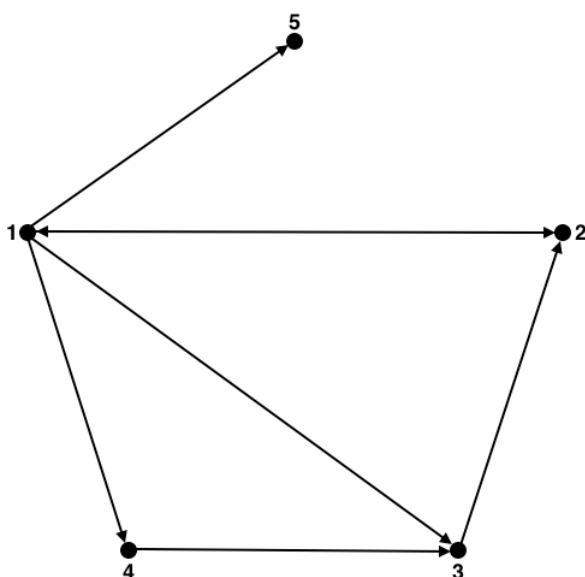
V prípade orientovanej siete, definícia vyššie je označovaná ako **in-degree** vrcholu. Pojem **out-degree** vrcholu i zodpovedá výrazu $\#\{j : g_{ji} = 1\}$.

Na obrázku 11 vidíme orientovanú sieť na piatich vrcholoch, kde vrchol 1 má 2 in-degree a 4 out-degree.

Hustota siete zachytáva relatívny podiel hrán prítomných v sieti.

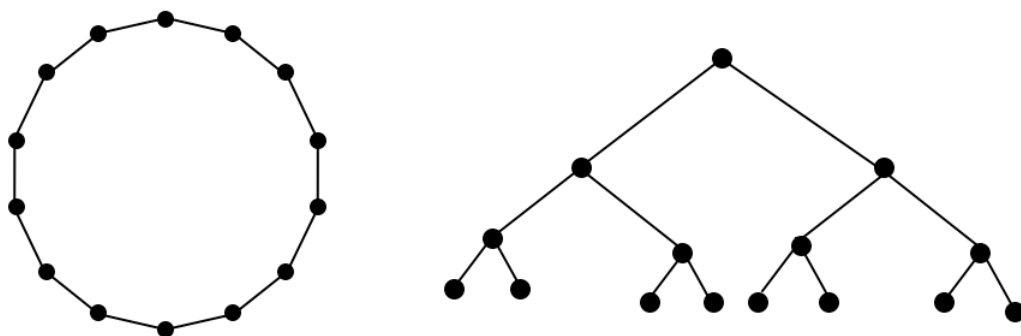
Vzdialenosť dvoch vrcholov je dĺžka (počet hrán) najkratšej cesty alebo *geodesic* medzi nimi. Ak sa medzi vrcholmi nenachádza žiadna cesta, potom je vzdialenosť medzi nimi nekonečná. Táto idea nás vedie k ďalšej dôležitej charakteristike siete, k jej priemeru. **Priemer** siete je najväčšia vzdialenosť akýchkoľvek dvoch vrcholov v sieti.

Na nasledujúcom príklade uvidíme, ako sa môže priemer líšiť v rámci siete na rovnakom počte vrcholov a hrán. Na obrázku 12 máme zobrazené dve rôzne siete, prvá predstavuje kruh a druhá strom. Obe siete majú približne priemerný stupeň 2 napriek celkom odlišnej štruktúre. V prípade kruhu má každý vrchol práve dve hrany, zatiaľ čo v binárnom strome má takmer polovica vrcholov stupeň 3 a druhá polovica má stupeň 1 (okrem začiatočného vrcholu- ten má stupeň 2). Keď sa pozrieme na priemer kruhu



Obr. 11: Orientovaná sieť- stupne

pre n vrcholov, dostávame $n/2$ alebo $(n - 1)/2$. Na druhej strane priemer binárneho stromu na n vrchoch je zhruba $2 \log_2(n + 1) - 2$.¹



Obr. 12: Kruh a strom- príklad na priemer

Priemer je jedným zo spôsobov merania dĺžky cesty, ktorý však udáva iba hornú hranicu. **Priemerná dĺžka cesty (charakteristická dĺžka cesty)** medzi dvomi vrcholmi je spôsob merania, ktorý zachytáva podobné vlastnosti. Priemerná dĺžka sa berie cez najkratšiu cestu. Pochopiteľne, priemerná dĺžka cesty je zhora ohraničená priemerom, v niektorých prípadoch to môže byť oveľa menej ako priemer. Preto je užitočné vedieť, či je priemer určený aj nejakými odľahlými hodnotami (outliermi) alebo nie.

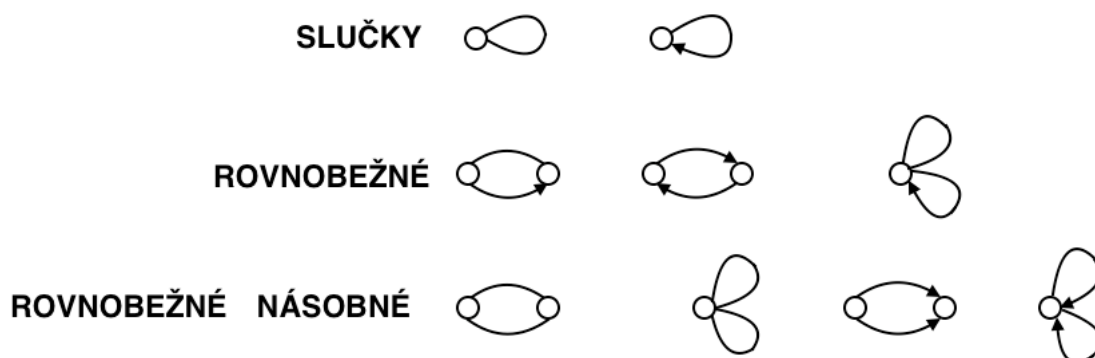
¹Toto platí práve vtedy, keď máme také celé číslo K , že $n = 2^K - 1$ v prípade binárneho stromu.

3 Grafové štruktúry

V tejto kapitole predstavíme rôzne typy hrán a grafových štruktúr definovaných v [11] a tiež popíšeme rôzne siete a ich komponenty podľa [5].

Graf je definovaný konečnou množinou vrcholov a hrán. Okrem spomínaných základných typov grafov (orientované a neorientované grafy), môžu byť grafy, ktoré majú niektoré hrany orientované a niektoré neorientované. Podobne existujú aj iné typy hrán zobrazené na obrázku 13:

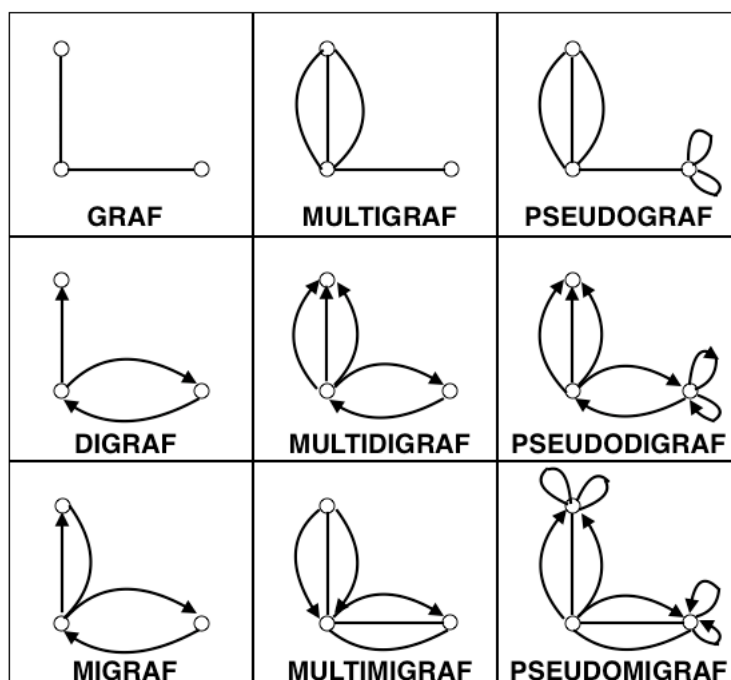
- rovnobežné - dvojice vrcholov sú spojené dvoma alebo viacerými hranami,
- slučky - vrchol je spojený sám so sebou,
- násobné - rovnobežné hrany, ktoré sú buď všetky šípy alebo rebrá a majú ten istý začiatok aj koniec.



Obr. 13: Klasifikácia hrán

Druhy grafových štruktúr podľa [11] (viď obrázok 14):

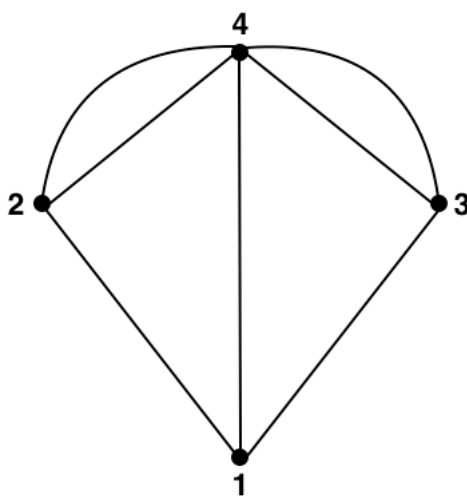
- graf - jeho súčasťou nie sú násobné hrany, slučky ani orientované hrany,
- multigraf - jeho súčasťou nie sú slučky ani orientované hrany,
- pseudograf - jeho súčasťou nie sú orientované hrany,
- digraf - neobsahuje násobné hrany, neorientované hrany ani slučky,
- multidigraf - neobsahuje slučky ani neorientované hrany,
- pseudodigraf - neobsahuje neorientované hrany,
- migraf - neobsahuje násobné hrany a slučky,
- multimigraf - neobsahuje slučky,
- pseudomigraf - môže obsahovať všetky typy hrán.



Obr. 14: Druhy grafových štruktúr, ktoré nájdeme v [11]

3.1 Mosty v Königsbergu

Jedným z príkladov v [5], na ktorých vidíme jednu z grafových štruktúr, je úloha o mostoch v Königsbergu. Königsberg bolo staré mesto v Prusku, ktoré ležalo na rieke Pregola. Mesto bolo oddelené riekou na 4 časti a malo 7 mostov. Otázkou bolo, či bolo možné prejsť cez každý most práve jedenkrát a vrátiť sa naspäť do štartovacieho bodu. Touto úlohou sa zaoberal matematik Leonhard Euler.



Obr. 15: Multigraf- mosty v Königsbergu

Pozrime sa na to ako na graf. Potom 4 časti mesta predstavujú vrcholy a 7 mostov charakterizuje hrany medzi vrcholmi v grafe. Dostaneme teda multigraf, ktorý vidíme na obrázku 15. Otázkou teda zostáva, či existuje v grafe prechádzka, ktorá obsahuje každú hranu práve raz a začína aj končí v tom istom vrchole. Takáto uzavretá prechádzka sa nazýva **Eulerovský ťah** ². Uzavretá prechádzka je taká prechádzka, ktorá začína a končí v tom istom vrchole.³ Potom, ak máme uzavretú prechádzku, ktorá obsahuje každú hranu v sieti práve raz, každý vrchol v tejto sieti musí mať párny stupeň. Vyplýva to z toho, že zakaždým, keď pri prechádzke navštívime cez určitú hranu jeden vrchol, musíme z neho vyjsť inou hranou a tiež vždy, keď navštívime daný vrchol, musí to byť cez hranu, ktorá sa ešte v rámci prechádzky nepoužila.

Veta 3.1 (Eulerova veta). *Súvislá sieť g má uzavretú prechádzku obsahujúcu každú hranu práve raz vtedy a len vtedy, keď stupeň každého vrcholu je párny.*

Táto veta je teda nutnou a postačujúcou podmienkou existencie Eulerovského ťahu.

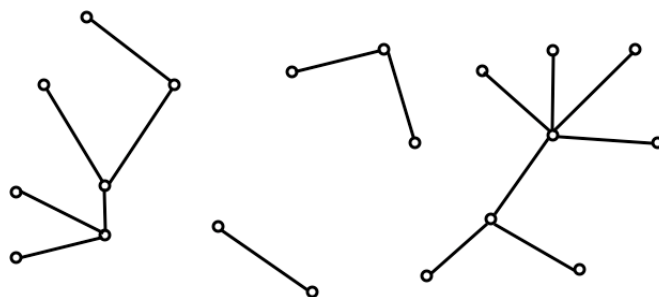
3.2 Rôzne druhy sietí

Podobne ako máme viacero druhov grafových štruktúr, poznáme aj niekoľko sieťových štruktúr, ktoré môžu byť prítomné v sieti. Zobrazené sú na obrázkoch 16 a 17 a sú to napr.:

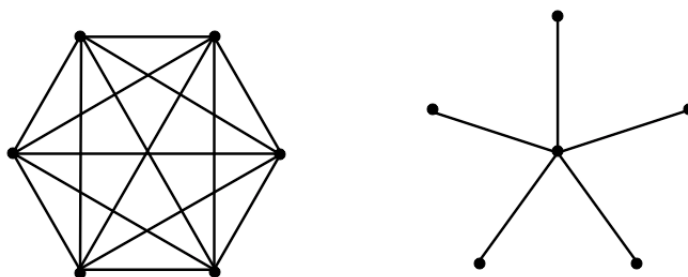
- **strom** - súvislá sieť bez cyklu,
- **les** - sieť, v ktorej je každý komponent strom (každá sieť, v ktorej sa nevyskytuje cyklus),
- **hviezda** - špeciálny typ štruktúry les; sieť, v ktorej existuje nejaký vrchol i taký, že každá hranu v sieti zahŕňa tento vrchol i (vrchol i predstavuje stred hviezdy),
- **kompletná sieť** - sieť, v ktorej sú prítomné všetky možné hrany, teda $g_{ij} = 1$ pre všetky $i \neq j$,
- **kruh** - sieť, ktorá obsahuje jediný cyklus a každý vrchol v sieti má práve dva susedné vrcholy.

²Pozri [11]

³Rozdiel medzi cyklom a uzavretou prechádzkou je ten, že v cykle nesmieme prejsť jedným vrcholom viackrát.



Obr. 16: Štyri stomy v lese



Obr. 17: Kompletná sieť a hviezda na šiestich vrcholoch

3.3 Komponenty siete

V mnohých aplikáciách je dôležité sledovať, ktoré vrcholy sú dosiahnuteľné cez cesty vrámci siete s ostatnými vrcholmi. Táto schopnosť zohráva rozhodujúcu úlohu v javoch ako nákaza, učenie či šírenie rozličného správania cez sociálnu sieť. Pozerajúc sa na vzťahy ciest v sieti, prirodzene dochádza k rozdeleniu siete na rôzne spojené podgrafy, čo nás privádza k definícii **komponentu**.

V [11] navyše nájdeme, že medzi základné a zároveň jednoduché úlohy na grafoch patrí aj úloha nájsť všetky komponenty, t.j. zistiť, či nejaké dva vrcholy v grafe súvisia. Pojem **komponent** grafu predstavuje maximálny súvislý podgraf tohto grafu.

Sieť (N, g) je **súvislá**, ak každé dva vrcholy v sieti sú spojené cestou. To znamená, že sieť (N, g) je súvislá, ak pre každé $i \in N$ a $j \in N$ existuje cesta v (N, g) medzi i a j .

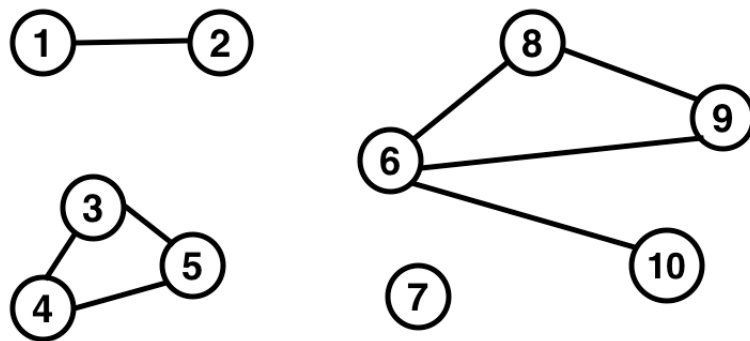
Komponent siete (N, g) je taká neprázdna podsieť (N', g') , kde $\emptyset \neq N', g' \subset g$ a platí:

- (N', g') je súvislá,
- ak $i \in N'$ a $ij \in g$, potom $j \in N'$ a $ij \in g'$.

Teda komponenty siete sú rôzne maximálne súvislé podgrafy siete.

Na obrázku 18 vidíme štyri komponenty:

- vrchol 7, ktorý nemá žiadne hrany,
- vrcholy 6, 7, 8, 9, 10 spolu s hranami $\{68, 89, 96, 610\}$,
- vrcholy 1 a 2 spolu s jednou hranou medzi nimi,
- vrcholy 3, 4, 5 spolu s hranami $\{34, 35, 45\}$.



Obr. 18: Sieť so štyrmi komponentmi

4 Centralita

Väčšina z doteraz spomínaných spôsobov merania siete opisuje všeobecné charakteristiky siete. V mnohých prípadoch nás môžu navyše zaujímať aj tie, ktoré nám dovoľujú porovnávať vrcholy a hovoria o tom, v akom vzťahu je daný vrchol k celej sieti. V príklade *Manželstvá vo Florencii* vidíme, že informácia o centralite vrcholu je dôležitá. Zvyčajne je práve predstava o tom, kde sa vrchol v sieti nachádza, celkom užitočná. Definujeme teda podľa [5] odlišné spôsoby, ako zmerať centralitu jednotlivých vrcholov v sieti.

Miery centrality môžu byť roztriedené do štyroch hlavných skupín závisiacich na type štatistiky, na ktorej sú založené:

1. **stupeň** - ako je vrchol prepojený,
2. **blízkosť** (*closeness*) - ako jednoducho môže vrchol dosiahnuť ostatné vrcholy,
3. **stredová medzipoloha** (*betweenness*) - aká je dôležitosť vrcholu v zmysle spojenia s ostatnými,
4. **charakteristika okolia a vlastného vektora** - ako sú dôležité, centrálné alebo vplyvné susedné vrcholy.

4.1 Centralita stupňa

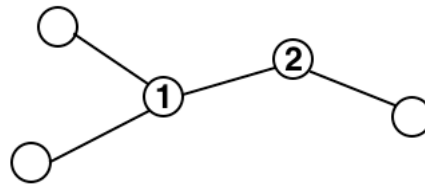
Pravdepodobne najjednoduchším spôsobom ako zmerať polohu daného vrcholu v sieti je sledovanie stupňa tohto vrcholu. Vrchol so stupňom $n - 1$ by bol priamo spojený so všetkými ostatnými vrcholmi a v konečnom dôsledku pomerne centrálny vzhľadom na sieť. Na druhej strane vrchol spojený iba s dvomi ďalšími vrcholmi, v prípade veľkého n , by bol považovaný za menej centrálny.[5]

Centralita stupňa vrcholu je definovaná

$$\frac{d_i(g)}{(n - 1)}, \quad (1)$$

kde $d_i(g)$ je stupeň vrcholu i a n je počet všetkých vrcholov v sieti g . Popisuje, ako dobre je vrchol pripojený v zmysle orientovaných vzťahov a jej hodnota sa pohybuje v rozmedzí 0 až 1.

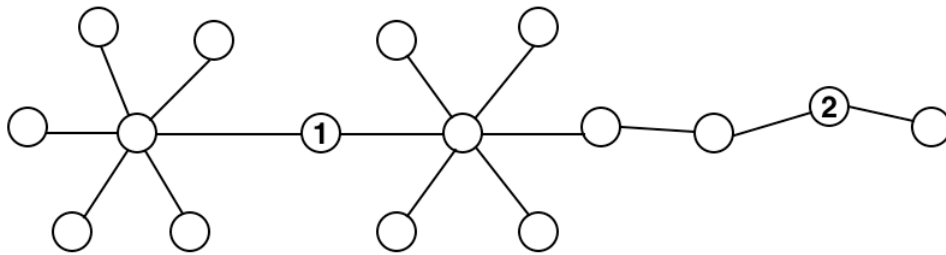
Na obrázku 19 máme 5 vrcholov, čo predstavuje naše n . Stupeň vrcholu 1 je $d_1(g) = 3$, pre vrchol 2 je to $d_2(g) = 2$. Potom centralita stupňa vrcholu 1 je rovná $d_1(g)/(n -$



Obr. 19: Príklad centrality stupňa

1) = $3/(5 - 1) = 3/4$ a pre vrchol 2 sú to $2/4$.

Očividne tento typ centrality vynecháva veľa zaujímavých aspektov siete. Konkrétne nemerá, ako dobre je vrchol v sieti umiestnený. Môže to znamenať, že vrchol má niekoľko hrán, ale leží v kritickej oblasti siete, ako je to napríklad na obrázku 20. Vrchol 1 a 2 tu majú rovnaký stupeň, ale vidíme, že vrchol 1 má v rámci siete lepšie postavenie. Pre veľa aplikácií je miera centrality, ktorá je citlivá na vplyv vrcholu v sieti, dôležitá.



Obr. 20: Centralita stupňa- rovnaký stupeň

4.2 Centralita blízkosti

Táto miera centrality sleduje, ako blízko je daný vrchol k ostatným vrcholom v sieti. Jednou z možností, ako merať centralitu blízkosti medzi vrcholom i a j , je podľa [5]

$$\frac{(n - 1)}{\sum_{j \neq i} \ell(i, j)}, \quad (2)$$

kde $\ell(i, j)$ predstavuje počet väzieb najkratšej cesty medzi vrcholom i a j a n je počet vrcholov v sieti.

Použitím siete na obrázku 19 vypočítajme centralitu blízkosti pre vrchol 1. V sieti je n rovné 5, i predstavuje vrchol 1 a $j = 2, 3, 4, 5$. Dosadením do vzorca (2) dostávame:

$$\frac{(n - 1)}{\sum_{j \neq i} \ell(i, j)} = \frac{(5 - 1)}{\sum_{j=2}^5 \ell(1, j)} = \frac{4}{1 + 1 + 1 + 2} = \frac{4}{5}.$$

Druhou možnosťou pri meraní centrality založenej na blízkosti je zohľadnenie **parametra tlmenia (decay parameter)** δ , kde $0 < \delta < 1$ a zváženie blízkosti medzi daným vrcholom a každým iným vrcholom váženým tlmením. Potom je táto centralita definovaná

$$\sum_{j \neq i} \delta^{\ell(i,j)}, \quad (3)$$

kde $\ell(i, j)$ je nekonečno, ak i a j nie sú spojené cestou.

Lepešie to vidieť na príklade k obrázku 19, kde nám pre vrchol $i = 1$, vyjde táto centralita

$$\sum_{j=2}^5 \delta^{\ell(1,j)} = \delta^1 + \delta^1 + \delta^1 + \delta^2 = 3 \cdot \delta^1 + \delta^2.$$

Ak zvolíme $\delta = 0,5$, dostaneme pre vrchol 1 centralitu rovnú 1,75. Pre $\delta = 0,25$ dostaneme hodnotu 0,8125 a pre $\delta = 0,75$ vyjde 2,8125.

Podľa hodnôt, ktoré dosahuje δ , môžeme povedať nasledovné. Ak sa δ blíži k 1, vidíme, že tento typ centrality meria, aký veľký je komponent, v ktorom leží daný vrchol. V prípade, že sa δ blíži k 0, centralita dáva nekonečne väčšiu váhu bližším vrcholom ako tým vzdialenejším. Z toho vyplýva, že sa v tomto prípade správa podobne ako centralita stupňa. Pre ostatné hodnoty δ (medzi 0 a 1) platí, že vrchol je ohodnotený podľa toho, ako blízko je k ostatným vrcholom. Veľmi vzdialený vrchol má menšiu váhu ako tie bližšie.

4.3 Centralita stredovej medzipolohy

Miera centrality, ktorá je založená na tom, ako je vrchol dobre situovaný v sieti v zmysle ciest, na ktorých leží.[5]

Nech $P_i(kj)$ označuje počet najkratších ciest (geodesic) medzi vrcholom k a j , na ktorých leží vrchol i a $P(kj)$ je celkový počet najkratších ciest medzi k a j . Môžeme odhadnúť dôležitosť i z hľadiska prepojenia k a j , keď sa pozrieme na pomer $P_i(kj)/P(kj)$. Ak je tento pomer blízko 1, potom i leží na väčšine najkratších ciest spájajúcich k s j . V opačnom prípade, keď je pomer blízko 0, je i menej rozhodujúce

pre k a j . **Centralita stredovej medzipolohy** je potom definovaná:

$$C e_i^B(g) = \sum_{k \neq j: i \neq \{k, j\}} \frac{P_i(kj)/P(kj)}{(n-1)(n-2)/2}. \quad (4)$$

4.4 Centralite na základe charakteristiky okolia a vlastného vektora

Rôznymi inými spôsobmi, ako merať centralitu, sa zaoberalo viacero autorov vo svojich dielach. Jedným z nich je aj Bonacich⁴, ktorého myšlienkou bolo, že meranie významnosti vrcholu je určované tým, ako dôležité sú jeho susedné vrcholy. To znamená, že by sme chceli počítať nie len súvislosť a blízkosť s ostatnými okolitými vrcholmi, ale v oblasti záujmu by boli aj tie „dôležité“ vrcholy. Táto myšlienka je ústredná pri takých javoch, akými sú rebríčky citácií a stránok Google. Takýto prístup je vlastne sebereferečný. Centralita vrcholu závisí od toho, ako centrálné sú susedné vrcholy, ktoré závisia na centralite ich susedov a podobne.[5]

Centralita vlastného vektora

Nech $C^e(g)$ označuje centralitu vlastného vektora priradenú sieti g . Centralita vrcholu je úmerná súčtu centralít jeho susedných vrcholov. Teda dostávame vzťah:

$$\lambda C_i^e(g) = \sum_j g_{ij} C_j^e(g), \quad (5)$$

v maticovom značení vyjadrené ako

$$\lambda C^e(g) = g C^e(g).$$

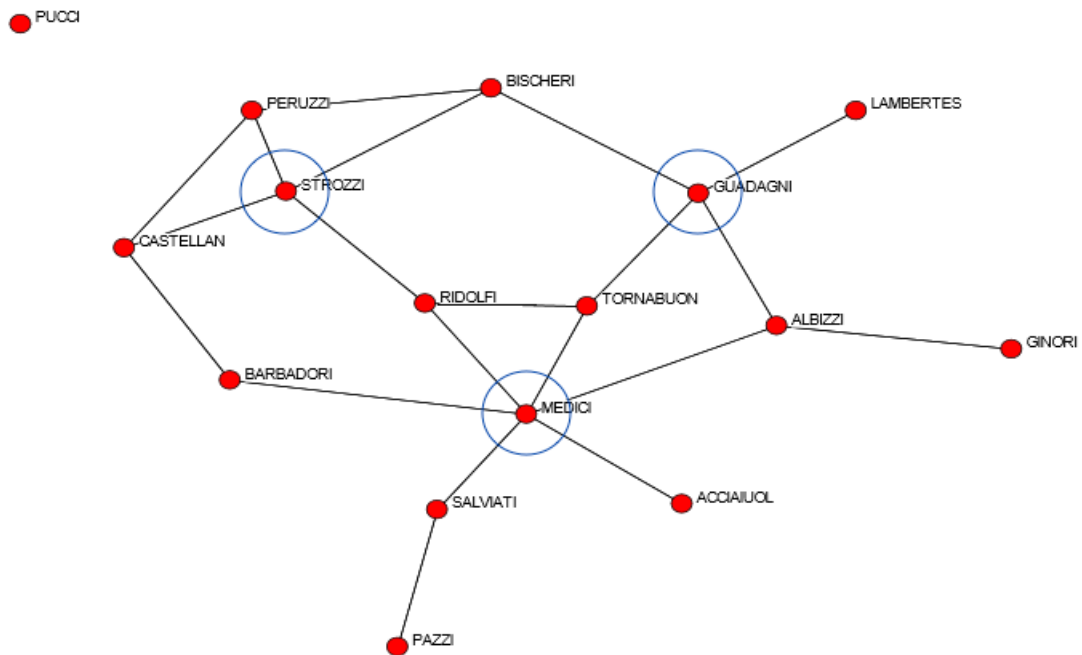
Potom $C^e(g)$ je vlastný vektor g a λ je jeho zodpovedajúce vlastné číslo. Vo všeobecnosti dáva zmysel hľadať mieru s nezápornými hodnotami, zvyčajne sa používa vlastný vektor s najväčšou prislúchajúcou vlastnou hodnotou.

Definícia tejto centrality funguje na vážených a/alebo orientovaných sieťach bez akejkoľvek zmeny.

⁴Pre bližšie informácie pozri [2].

4.5 Manželstvá vo Florencii - rôzne miery centrality

Pripomeňme si príklad z kapitoly 1, ktorý hovoril o rodinách vo Florencii, viď obrázok 21. V nasledujúcej časti ukážeme porovnanie spomínaných mier centralít práve na tejto sieti. Konkrétne sa pozrieme na rodinu Medici, Strozzi a Guadagni. Zhrnutie výsledných centralít je nájdeme v tabuľke 1.



Obr. 21: Manželstvá vo Florencii

Centralita stupňa

Na obrázku 21 vidíme, že rodina Medici je priamo spojená so šiestimi ďalšími rodinami. Dokopy je v celej sieti 16 rodín, a teda centralitu stupňa vrcholu Medici vypočítame podľa zadaného vzťahu (1) ako $d_i(g)/(n - 1) = 6/(16 - 1) = 6/15$. Podobne pre rodinu Strozzi a Guadagni platí rovnaký vzťah. Obidve tieto rodiny sú v priamom spojení práve so štyrmi rodinami a ich miera centrality stupňa je teda rovná $4/15$.

Centralita blízkosti

V tomto prípade neberieme do úvahy rodinu Pucci, ktorá nie je v spojení so žiadnou z ostatných rodín. Inak by nám totiž pre každú rodinu vyšla táto centralita rovná 0 (nekonečná vzdialenosť od rodiny Pucci). Dosadením do vzťahu (2) pre výpočet tejto

centrality, kde $n = 15$, dostaneme centralitu blízkosti pre rodinu Medici rovnú $14/25$, pre rodinu Strozzi $14/32$ a pre rodinu Guadagni $14/26$.

Centralita stredovej medzipolohy

Táto miera centrality je definovaná vzťahom (4). Pripomeňme si, že $P_i(kj)$ označuje počet najkratších ciest medzi vrcholom k a j , na ktorých leží vrchol i a $P(kj)$ je celkový počet najkratších ciest medzi k a j . Potom pre rodinu Medici vychádza táto centralita rovná $0,533$, pre rodinu Strozzi $0,125$ a pre rodinu Guadagni $0,253$.

Centralita vlastného vektora

Vzťah (5), ktorý definuje centralitu vlastného vektora, nám v prípade rodín vo Florencii dáva tieto hodnoty centralít. Rodina Medici má centralitu vlastného vektora rovnú $0,430$, rodina Strozzi $0,356$ a Guadagni $0,289$.

	Medici	Strozzi	Guadagni
Stupeň	0,400	0,267	0,267
Blížkosť	0,560	0,438	0,538
Stredová medzipoloha	0,533	0,125	0,253
Vlastný vektor	0,430	0,356	0,289

Tabuľka 1: Rôzne miery centrality pre rodiny vo Florencii

5 DeGrootov model

V [5] nájdeme, že sociálne siete zohrávajú ústrednú úlohu v zdieľaní informácií a formovaní názorov. Platí to v prípade, že sa chceme s priateľmi poradiť, ktorý film si pozrieť, tiež pri odovzdávaní informácií o schopnostiach a vhodnosti prípadného nového zamestnanca, či pri debatovaní o politikoch. Vzhľadom na úlohu sociálnych sietí pri vytváraní názorov, presvedčení a následné formovanie správania, je podľa [5] dôležité, že máme dôkladné poznatky o tom, ako štruktúra týchto sietí ovplyvňuje učenie a šírenie informácií. Niektoré základné otázky sa týkajú toho, ako sociálne siete ovplyvňujú,

- či sa jednotlivci v rámci spoločenstva pridržajú spoločnej viery alebo zostávajú rozdielny v názoroch,
- ktorí jednotlivci majú najväčší vplyv na presvedčenie ostatných v skupine,
- ako rýchlo sa jednotlivci učia,
- či spočiatku rôznorodé informácie rozptýlené po celej spoločnosti môžu byť sformované presným spôsobom.

Na tieto otázky existuje viacero odpovedí v podobe rôznych modelov. My sa zameriame na jeden z počiatočných modelov sociálneho vplyvu a imitácie - DeGrootov model, popísaný v [5]. Tento model prenosu informácií, formovania názorov a všeobecného súhlasu je jednoduchým a prirodzeným štartovacím bodom pre teóriu, ktorá nám umožňuje porozumieť, ako štruktúra siete ovplyvňuje šírenie informácií a formovanie názorov. V [3] DeGroot popísal, ako skupina jednotlivcom s rôznymi názormi na istú skutočnosť môže dospieť k vzájomnému konsenzu. Podľa neho má model tri dôležité výhody:

1. Proces, ktorý tento model popisuje, je prirodzene zaujímavý.
2. Má jednoduché podmienky na určenie, či je možné, aby skupina dospela ku konsenzu.
3. V prípade, že môže byť dosiahnutý konsenzus, použité váhy sa dajú jednoducho vyčíslieť.

5.1 Základná charakteristika

Jednotlivci v sieti začínajú s nejakým počiatočným názorom na vec. Nech sú tieto názory teda reprezentované n -rozmerným vektorom pravdepodobností $p(0)$, kde $p(0) = (p_1(0), \dots, p_n(0))$. Každé $p_i(0)$ leží v intervale $[0, 1]$ a môže predstavovať pravdepodobnosť, že daná skutočnosť je pravdivá, tiež kvalitu daného produktu, či pravdepodobnosť, že sa jednotlivec môže zapojiť do danej aktivity, a pod. Vzory vzájomného pôsobenia sú zachytené váženou a orientovanou $n \times n$ nezápornou maticou T , ktorá je (riadková) stochastická matica, teda jej zložky majú v riadku súčet rovný 1. Jedná sa o znormovanie pôsobenia tak, aby v súčte dávalo 1. Prvok T_{ij} tejto matice vyjadruje váhu (dôveru), ktorú agent i dáva súčasnému presvedčeniu agenta j pre prípad formovania presvedčania agenta i na ďalšie obdobie.

5.2 Aktualizácia v modeli

Presvedčenia sú aktualizované v priebehu času podľa vzťahu

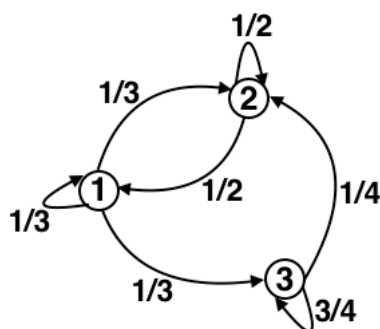
$$p(t) = Tp(t - 1) = T^t p(0). \quad (6)$$

Príklad 1

Definíciu DeGrootovho modelu si ukážeme na príklade z [5]. Na obrázku 22 máme znázornený proces aktualizácie pre troch jedincov, ktorý je vyjadrený maticou

$$T = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 0 & 1/4 & 3/4 \end{pmatrix}.$$

Jednotlivé prvky matice vyjadrujú nasledovné. Prvky matice T_{1j} vyjadrujú presvedčenie agenta číslo 1. Rovnakú váhu $1/3$ vkladá do seba a aj do agentov 2 a 3. V prípade druhého agenta je to trochu inak. Z T_{2j} vidíme, že agent 2 dáva váhu $1/2$ sebe a agentovi 1, ale ignoruje tretieho agenta. Agent 3 dáva váhu $1/4$ agentovi 2 a sebe $3/4$ a ignoruje prvého agenta.



Obr. 22: Proces aktualizácie

Predpokladajme, že počiatočný vektor presvedčení je

$$p(0) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},$$

takže agent 1 začína s presvedčením 1 ohľadom nejakej udalosti, zatiaľ čo ostatní dvaja agenti majú na začiatku nulové presvedčenie. Potom v nasledujúcom čase platí:

$$p(1) = Tp(0) = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 0 & 1/4 & 3/4 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1/3 \\ 1/2 \\ 0 \end{pmatrix}.$$

Podobne po ďalšej aktualizácii dostávame

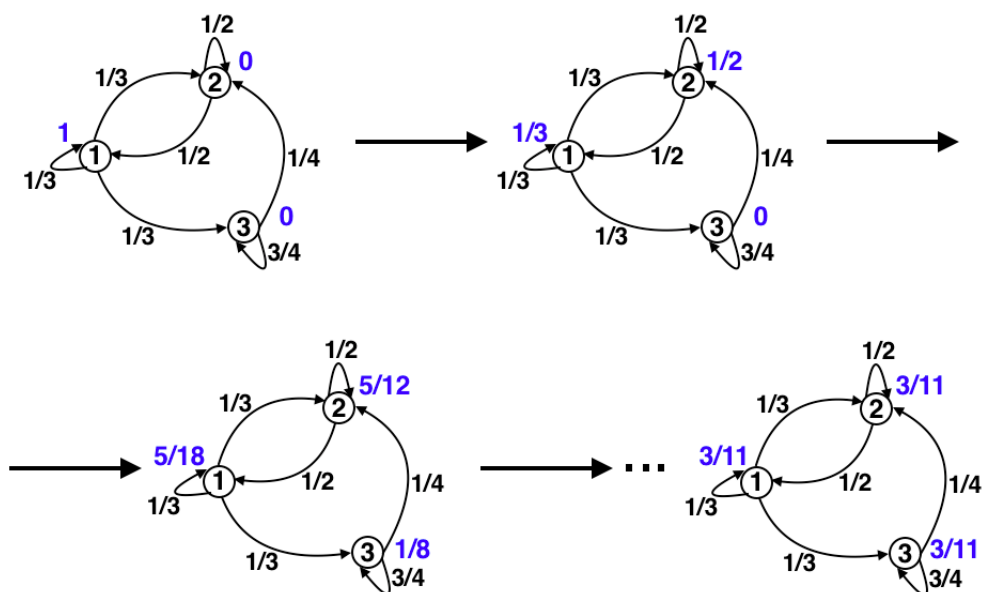
$$p(2) = Tp(1) = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 0 & 1/4 & 3/4 \end{pmatrix} \begin{pmatrix} 1/3 \\ 1/2 \\ 0 \end{pmatrix} = \begin{pmatrix} 5/18 \\ 5/12 \\ 1/8 \end{pmatrix}.$$

Celý tento proces vidíme aj na obrázku 23.

Iterovanie tohto procesu aktualizácií pre $t \rightarrow \infty$ konverguje podľa vzťahu (6) k presvedčeniu

$$p(t) = \lim_{t \rightarrow \infty} T^t p(0) = \begin{pmatrix} 3/11 & 4/11 & 4/11 \\ 3/11 & 4/11 & 4/11 \\ 3/11 & 4/11 & 4/11 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 3/11 \\ 3/11 \\ 3/11 \end{pmatrix}. \quad (7)$$

V nasledujúcej kapitole sa budeme zaoberať konvergenciou mocnín matice vplyvu a následne dokážeme aj tvrdenie o limite (7) z tohto príkladu.



Obr. 23: Príklad 1- aktualizácia procesu

5.3 Konvergencia matice sociálneho vplyvu

Predtým, ako si ukážeme konvergenciu pre maticu aktualizácií z príkladu 1, potrebujeme poznať základnú vetu týkajúcu sa konvergenencie. Špeciálny prípad vety o konvergencii, ktorý niekedy stačí, nájdeme v [12]. Uvádzame ho preto, lebo tento dôkaz je jednoduchší v porovnaní s inými.

Veta 5.1 (Perron-Frobenius). *Nech M je n -rozmerná stochastická matica. Predpokladajme, že existuje také k , že M^k má všetky zložky kladné. Potom existuje taký riadkový vektor $m = (m_j)_1^n$ s kladnými zložkami, ktorých súčet je 1, že pre každé $1 \leq i \leq n$ platí*

$$\lim_{\ell \rightarrow \infty} M_{i,j}^\ell = m_j. \quad (8)$$

Navyše platí, že $m = (m_i)_1^n$ je taký riadkový vektor, pre ktorý platí $\sum_{i=1}^n m_i = 1$ a $mM = m$.

Na dôkaz Perronovej- Frobeniovej vety využijeme nasledovnú lemu.

Lema 5.2. *Nech M je n -rozmerná stochastická matica. Predpokladajme, že pre každý pár $(i, j), 1 \leq i, j \leq n$ existuje také $k = k(i, j)$, že $M_{i,j} > 0$. Potom existuje jediný riadkový vektor $m = (m_j)_1^n$ s kladnými zložkami v súčte 1, že $mM = m$.*

Dôkaz Lemy 5.2: Podľa predpokladu stĺpcový vektor $\vec{1}$ so všetkými zložkami rovnými 1 spĺňa $M\vec{1} = \vec{1}$. Z lineárnej algebry vieme, že M^T má tiež 1 ako vlastnú hodnotu, t.j.

existuje nenulový riadkový vektor v taký, že $vM = v$.

Ďalej existuje také k , že $M_{i,j}^k > 0$ pre všetky (i, j) . Pre maticu M^k potom platí:

$$vM^k = (vM)M^{k-1} = vM^{k-1} = \dots = v \quad (9)$$

$$|v|M^k = (|v|M)M^{k-1} = |v|M^{k-1} = \dots = |v| \quad (10)$$

Dokážeme najskôr, že žiadna zložka vektora $|v|$, a teda ani v , nie je nula. Ak by i -ta zložka bola nulová:

$$|v|_i = (|v|M^k)_i = \sum_{j=1}^n |v|_j (M^k)_{ji} \geq 0,$$

lebo $(M^k)_{ji} > 0$ a $|v|_j \geq 0$, pričom rovnosť môže nastať len vtedy, keď sú všetky sčítance nulové, t.j. pre každé j

$$|v|_j \underbrace{(M^k)_{ji}}_{> 0} = 0,$$

z čoho vyplýva, že $|v|_j = 0$ a $v_j = 0$ pre všetky j , čo je spor s tým, že v ako vlastný vektor je nenulový.

Teraz dokážeme, že v všetky zložky s rovnakým znamienkom. Zo vzťahov (9) a (10) máme:

$$(|v| - v)M^k = |v|M^k - vM^k = |v| - v,$$

čiže $|v| - v$ je buď nulový vektor, vtedy $|v| = v$, t.j. $|v_j| = v_j$ pre všetky j vtedy a len vtedy, keď $v_j > 0$ pre všetky j alebo $|v| - v$ je vlastný vektor M^k . Videli sme, že vlastný vektor M^k nemôže mať žiadnu zložku nulovú. Pre všetky j je teda $|v_j| - v_j > 0$ alebo $|v_j| - v_j < 0$. Druhá možnosť nemôže nastať, lebo $|v_j| = \max\{v_j, -v_j\} \geq v_j$. Teda pre všetky j platí $|v_j| - v_j > 0$, a teda $|v_j| > 0$ vtedy a len vtedy, keď $v_j < 0$, čiže všetky zložky v sú záporné.

To znamená, že ak v je vlastný vektor M , tak buď má všetky zložky kladné alebo všetky záporné. Ak v je vlastný vektor, tak aj $-v$ je vlastný vektor, takže prípadným vynásobením (-1) dostaneme vlastný vektor v , ktorého všetky zložky sú kladné. Za vektor m z lemy teda zoberieme $m = \frac{v}{v_1 + \dots + v_n}$.

Dokážeme ešte, že taký vektor m je jednoznačne určený. Ak m, m' sú dva normalizované vlastné vektory s kladnými zložkami, potom $m - m'$ je buď nulový alebo vlastný vektor. Ak $m - m'$ nie je nulový, jeho zložky musia meniť znamienka, čo je spor. Máme teda $m = m'$. \square

Dôkaz Vety 5.1: Podľa predpokladu vety, nech je také k , že pre všetky i, j $M_{i,j}^k > 0$. Nech $m = (m_i)_1^n$ je riadkový vektor a $M_{i,j}^\infty = m_j$ tak, že M^∞ je matica s riadkami rovnými m . Všimnime si, že platí

$$MM^\infty = M^\infty M = M^\infty. \quad (11)$$

Analogicky platí $M^N M^\infty = M^\infty M^N = M^\infty$ pre ľubovoľnú mocninu N (napr. $M^N M^\infty = M^{N-1}(MM^\infty) = M^{N-1}M^\infty = \dots = M^\infty$). M je n -rozmerná stochastická matica a $M_{i,j}^k \geq cM_{i,j}^\infty$, kde $c = \min_{i,j} \{M_{i,j}^k / M_{i,j}^\infty\} > 0$.

Uvažujme maticu

$$N = \frac{1}{1-c}(M^k - cM^\infty),$$

pričom platí, že $N = 0$, ak $c = 1$, t.j. $M^k = M^\infty$ a N je stochastická matica, ak $0 < c < 1$. Tiež platí, že použitím vzťahu (11) dostaneme

$$\begin{aligned} NM^\infty &= \frac{1}{1-c}(M^k - cM^\infty)M^\infty \\ &= \frac{1}{1-c}(M^k M^\infty - cM^\infty M^\infty) \\ &= \frac{1}{1-c}(M^\infty M^k - cM^\infty M^\infty) \\ &= M^\infty \frac{1}{1-c}(M^k - cM^\infty) \\ &= M^\infty N \end{aligned}$$

a

$$\begin{aligned} NM^\infty &= M^\infty N = \frac{1}{1-c}(M^\infty M^k - cM^\infty M^\infty) \\ &= \frac{1}{1-c}(M^\infty - cM^\infty) = \frac{1}{1-c}(1-c)M^\infty \\ &= M^\infty \end{aligned}$$

V oboch prípadoch, pre $c = 1$ a $0 < c < 1$, sú zložky $(N - M^\infty)^\ell = N^\ell - M^\infty$ ohraničené 1 v absolútnej hodnote pre všetky $\ell = 1, 2, \dots$. Navyše z definície matice N dostávame

$$M^k - M^\infty = (1-c)(N - M^\infty). \quad (12)$$

Tiež platí

$$M^{k\ell} - M^\infty = (M^k - M^\infty)^\ell, \quad (13)$$

čo teraz dokážeme matematickou indukciou. Pre $\ell = 1$ tvrdenie platí:

$$M^{k*1} - M^\infty = (M^k - M^\infty)^1.$$

Indukčný predpoklad bude, že toto tvrdenie platí pre ℓ , teda

$$M^{k\ell} - M^\infty = (M^k - M^\infty)^\ell.$$

Potom pre $\ell + 1$ platí nasledovné:

$$\begin{aligned} (M^k - M^\infty)^{\ell+1} &= (M^k - M^\infty)^\ell (M^k - M^\infty) = (M^{k\ell} - M^\infty)(M^k - M^\infty) = \\ &= M^{k(\ell+1)} - M^{k\ell} M^\infty - M^\infty M^k + (M^\infty)^2 = \\ &= M^{k(\ell+1)} - M^\infty - M^\infty + M^\infty M^\infty = \\ &= M^{k(\ell+1)} - 2M^\infty + M^\infty = M^{k(\ell+1)} - M^\infty. \end{aligned}$$

Teda použitím vzťahu (12) máme

$$M^{k\ell} - M^\infty = (M^k - M^\infty)^\ell = (1 - c)^\ell (N - M^\infty)^\ell.$$

Potom

$$|M_{i,j}^{k\ell} - M_{i,j}^\infty| \leq (1 - c)^\ell. \quad (14)$$

Zoberúc do úvahy maximovú normu $\|A\|_\infty = \max_{i,j} |A_{i,j}|$ a takú funkciu

$$\ell \rightarrow \|M^\ell - M^\infty\|_\infty,$$

ktorá je nerastúca, pretože $M^{\ell+1} - M^\infty = M(M^\ell - M^\infty)$ implikuje

$$\begin{aligned} (M^{\ell+1} - M^\infty)_{i,j} &= \sum_s M_{i,s} (M^\ell - M^\infty)_{s,j} \leq \underbrace{\left(\sum_s M_{i,s} \right)}_{=1} \|M^\ell - M^\infty\|_\infty \\ &= \|M^\ell - M^\infty\|_\infty \end{aligned}$$

Preto vychádzajúc zo vzťahu (14)

$$\max_{i,j} \{|M_{i,j}^\ell - m_j|\} \leq (1 - c)^{\lfloor \ell/k \rfloor}$$

a pre ℓ idúce do nekonečna výraz $(1 - c)^{\lfloor \ell/k \rfloor}$ ide v limite k 0 tak, ako aj rozdiel $M_{i,j}^\ell - m_j$, t.j. aj $\max_{i,j} \{|M_{i,j}^\ell - m_j|\}$. Teda v konečnom dôsledku dostávame

$$\lim_{\ell \rightarrow \infty} M_{i,j}^\ell = m_j.$$

□

Príklad 2

Konvergenciu matice sociálneho vplyvu ukážeme na tej istej matici aktualizácií ako v príklade 1 (obrázok 22), teda

$$T = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 0 & 1/4 & 3/4 \end{pmatrix}.$$

Výpočtom si môžeme overiť, že platí

$$T^2 = \begin{pmatrix} 5/18 & 13/36 & 13/36 \\ 5/12 & 5/12 & 1/6 \\ 1/8 & 5/16 & 9/16 \end{pmatrix}, T^3 = \begin{pmatrix} 59/216 & 157/432 & 157/432 \\ 25/72 & 7/18 & 19/72 \\ 19/96 & 65/192 & 89/192 \end{pmatrix},$$

$$T^4 = \begin{pmatrix} 707/2592 & 1885/5184 & 1885/5184 \\ 67/216 & 325/864 & 271/864 \\ 271/1152 & 809/2304 & 953/2304 \end{pmatrix}, \dots$$

a teda podmienka kladných prvkov niektorej mocniny matice T je splnená. Teraz vypočítame limitu mocniny matice T . Vlastné hodnoty matice T sú $1, 1/2$ a $1/12$. Ľavý vlastný vektor k vlastnej hodnote 1 je $(1, 4/3, 4/3)$, resp. jeho ľubovoľný násobok. Preškálujeme teda vlastný vektor tak, aby jeho zložky v súčte boli rovné 1 :

$$\frac{1}{1 + \frac{4}{3} + \frac{4}{3}} \cdot \left(1, \frac{4}{3}, \frac{4}{3}\right) = \left(\frac{3}{11}, \frac{4}{11}, \frac{4}{11}\right).$$

Mocniny T^t konvergujú k matici, ktorá má všetky riadky rovné tomuto vektoru:

$$\lim_{t \rightarrow \infty} T^t = \begin{pmatrix} 3/11 & 4/11 & 4/11 \\ 3/11 & 4/11 & 4/11 \\ 3/11 & 4/11 & 4/11 \end{pmatrix}.$$

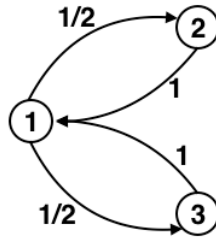
Keďže riadky limitnej matice sú rovnaké, nezáleží od toho, s akým vektorom presvedčení $p(0)$ agenti začnú, všetci skončia s limitnými hodnotami presvedčení, ktoré zodpovedajú vstupom $p(\infty) = \lim_{t \rightarrow \infty} T^t p(0)$, kde

$$p_1(\infty) = p_2(\infty) = p_3(\infty) = \frac{3}{11}p_1(0) + \frac{4}{11}p_2(0) + \frac{4}{11}p_3(0).$$

Na tomto príklade vidíme, že presvedčenia konvergujú v čase. Ďalšou zaujímavou informáciou je, že agenti dospejú ku konsenzu.

Príklad 3

V tomto príklade uvidíme, že nedospejeme ku konvergencii. Majme proces aktualizovania daný podľa obrázku 24.



Obr. 24: Príklad 3- nekonvergentný proces

Pre tento prípad máme teda maticu aktualizácií

$$T = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

jej mocniny sú

$$T^2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix}, T^3 = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, T^4 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix}, \dots$$

Vidíme, že dochádza k oscilácii matice a teda nedospejeme ku konvergencii. Zoberme do úvahy počiatočné presvedčenia $p_1(0) = 1$ a $p_2 = p_3(0) = 0$. Potom, keďže agent 1 aktualizuje svoje presvedčenie na základe presvedčení agentov 2 a 3 a agenti 2 a 3 aktualizujú podľa presvedčení agenta 1, agenti si jednoducho vymieňajú svoje presvedčenie v priebehu času.

5.4 Konvergencia matice aktualizácií - iný prístup

Podmienka z vety 5.1 nemusí platiť, je to len špeciálny prípad. Vo všeobecnom tvrdení vo vete 5.3 je slabší predpoklad ireducibility, zato však treba overovať aperiodickosť.

Veta 5.3 (Perron - Frobenius pre aperiodické matice [6]). *Predpokladajme, že A je nezáporná štvorcová matica. Ak je A ireducibilná a aperiodická, potom existuje reálna vlastná hodnota $\lambda > 0$, pre ktorú platí:*

- λ je jednoduchý koreň charakteristického polynómu,
- λ má striktne kladné ľavé a pravé vlastné vektory,
- vlastné vektory pre λ sú jedinečné až na konštantný násobok,
- $\lambda > |\mu|$, kde μ je akákoľvek iná vlastná hodnota,
- ak $0 \leq B \leq A$ (zložka po zložke) a β je vlastná hodnota B , potom $|\beta| \leq \lambda$ a rovnosť platí vtedy a len vtedy, keď $B = A$,
- $\lim_{n \rightarrow \infty} A^n / \lambda^n = r\ell$, kde r a ℓ sú ľavý a pravý vlastný vektor A normalizovaný tak, že $\ell \cdot r = 1$.

Špeciálna vlastná hodnota λ je Perronova hodnota matice A . Kladný vlastný vektor prislúchajúci λ sa nazýva Perronov vlastný vektor.

V našich prípadoch, ktoré sme analyzovali, bola častá situácia, že matica aktualizácií má jednu vlastnú hodnotu rovnú jednej ($\lambda_1 = 1$), vlastné hodnoty s absolútnou hodnotou z intervalu $(0, 1)$ majú násobnosť 1 (označme ich $\lambda_2, \dots, \lambda_k$) a prípadne viacnásobná vlastná hodnota je $\lambda_{k+1} = \dots = \lambda_n = 0$.

Máme teda vzťah, ktorý platí pre aktualizáciu:

$$p(t+1) = Tp(t). \quad (15)$$

Zoberme maticu T takú že platí $T = AJA^{-1}$, kde J je Jordanov tvar

$$J = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & \ddots & 0 & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & \lambda_k & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & J_1 & 0 & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & J_m \end{pmatrix}$$

a bloky J_1, \dots, J_m zodpovedajú nule. Potom platí aj $T^t = AJ^tA^{-1}$. Limitu J^t vieme vypočítať, lebo limita Jordanových blokov zodpovedajúcich nule je nulová matica a

$|\lambda_i| < 1$ pre $i = 2, \dots, k$, takže λ_i^k ide k nule. Preto

$$\lim_{t \rightarrow \infty} J^t = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & \ddots & \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & \ddots & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & \ddots & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \end{pmatrix} = \widehat{J}$$

a teda konverguje aj T^t . K čomu však konverguje $\lim_{t \rightarrow \infty} T^t = A\widehat{J}A^{-1}$?

Matica A má stĺpce rovné vlastným vektorom a zovšeobecneným vlastným vektorom matice T , špeciálne prvý stĺpec je vlastný vektor prislúchajúci k vlastnej hodnote $\lambda_1 = 1$, čo je $(1, \dots, 1)^T$. Podľa predpokladu je násobnosť tejto vlastnej hodnoty 1. Teda

$$\begin{aligned} \lim_{t \rightarrow \infty} T^t &= A \cdot \widehat{J} \cdot A^{-1} = \begin{pmatrix} 1 & A_{12} & \cdots & A_{1k} \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ 1 & A_{k2} & \cdots & A_{kk} \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & \cdots & \cdots \\ 0 & \cdots & \cdots & \cdots \\ \vdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & \cdots \end{pmatrix} \cdot A^{-1} = \\ &= \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix} \cdot \begin{pmatrix} A_{11}^{-1} & \cdots & \cdots & A_{1k}^{-1} \\ A_{21}^{-1} & \cdots & \cdots & A_{2k}^{-1} \\ \vdots & \vdots & \vdots & \vdots \\ A_{k1}^{-1} & \cdots & \cdots & A_{kk}^{-1} \end{pmatrix} = \begin{pmatrix} A_{11}^{-1} & \cdots & \cdots & A_{1k}^{-1} \\ A_{11}^{-1} & \cdots & \cdots & A_{1k}^{-1} \\ \vdots & \vdots & \vdots & \vdots \\ A_{11}^{-1} & \cdots & \cdots & A_{1k}^{-1} \end{pmatrix}. \end{aligned}$$

Vidíme, že výsledná matica má všetky riadky rovnaké, čiže limita $p(t)$ pre t idúce do nekonečna nezávisí od $p(0)$. Ideme teda nájsť ten prvý riadok matice A^{-1} , pričom platí:

$$A^{-1}T = JA^{-1}.$$

Nech matica A^{-1} má prvý riadok w_1 , potom platí

$$w_1^T T = w_1^T$$

$$T^T w_1 = w_1,$$

teda w_1 je vlastný vektor matice T^T prislúchajúci k vlastnej hodnote 1.

6 Cognitive Social Structures

Niektoré problémy v oblasti analýzy sietí je možné preskúmať pomocou kognitívnych sociálnych štruktúr, ktoré môžu byť modelované ako trojrozmerné sieťové štruktúry ako sa uvádza v [7]. Spomenieme teda tri typy zhukovania kognitívnych sociálnych štruktúr- Slices, Locally Aggregated Structures (LAS) a Consensus Structures (CS), ktoré sú navrhnuté tak, aby zredukovali dimenziu na dvojrozmerné, čo je potrebné pre ďalšiu analýzu. Definície jednotlivých štruktúr takisto ako konkrétny príklad na využitie tohto prístupu nájdeme v [7].

Za vznikom kognitívnych sociálnych štruktúr boli zistenia, že pri skúmaní analýzy sietí nie sú behaviorálne merania interakcií veľmi úzko späté s pohľadom účastníkov na tie isté interakcie. Autori v [1] dospeli k záveru, že respondenti sú nepresní, keďže pamäť sa časom zhoršuje a dochádza k systematickému skreslovaniu udalostí v tom, ako si ich respondenti pripomínajú.

6.1 Definícia CSS

V oblasti sociálnych sietí zohráva štruktúra špecifický význam. Štruktúra každého systému je definovaná ako množina príkazov vzťahov medzi všetkými párami zúčastnených strán v systéme. Tieto príkazy môžu byť zhrnuté v súbore R matic, jeden súbor matic pre každý vzťah, v tvare \mathfrak{R}_{ij} , kde \mathfrak{R} je vzťah definujúci štruktúru, i je „vysielač“ vzťahu, j je „prijímač“. Potom $\mathfrak{R}_{ij} = 1$, ak i je vo vzťahu s j určenom \mathfrak{R} , inak $\mathfrak{R}_{ij} = 0$. Ak je \mathfrak{R} definované ako „obrátenie sa o pomoc a radu“, $\mathfrak{R}_{3,12} = 1$ by sa dalo interpretovať tak, že osoba 3 sa obrátila na osobu 12 o pomoc a radu.

Kognitívna sociálna štruktúra tohto systému by bola vyjadrená ako $\mathfrak{R}_{i,j,k}$, kde i je „vysielač“ vzťahu, j je „prijímač“ a k je „vnímateľ“ tohto vzťahu medzi i a j . Potom $\mathfrak{R}_{3,12,8} = 1$ by sa dalo interpretovať tak, že osoba 8 si myslí, že osoba 3 by sa obrátila na osobu 12 o pomoc a radu.

Teda, ak máme N účastníkov v systéme popísanom o R vzťahmi, potom sociálna štruktúra by bola popísaná $RN \times N$ maticami, ale kognitívna sociálna štruktúra $RN \times N \times N$ maticami. Tým pádom je zjavné, že množstvo informácií v kognitívnej sociálnej štruktúre prevyšuje to v tradičnej sociálnej štruktúre.

6.2 Zoskupenie (Aggregations)

Základnou otázkou pri analýze kognitívnych sociálnych štruktúr je, ako skúmať trojrozmernú sadu dát. Jeden z prístupov je použitím nejakých pravidiel zlúčiť vzťah $\mathfrak{R}_{i,j,k}$ do dvojrozmerného vzťahu $\mathfrak{R}'_{i,j}$. Priblížime tri základné druhy takejto redukcie: Slices, Locally Aggregated Structures (LAS) a Consensus Structures (CS). Každý z týchto typov sa riadi vlastnými pravidlami motivovanými na základe otázok, na ktoré chce nájsť odpoveď.

6.2.1 Rezy (Slices)

Najjednoduchším spôsobom, ako zmenšiť rozmer, je vybratie časti z trojrozmernej matice a udržanie dimenzie „vnímateľa“ konštantné, t.j. pre $K = \text{konštanta}$

$$\mathfrak{R}'_{i,j} = \mathfrak{R}_{i,j,K}.$$

Tento prístup dáva z vybratého rezu k informáciu, či osoba k vníma všetkých svojich priateľov, že sú priatelia medzi sebou navzájom.

6.2.2 LAS

Druhým prístupom, ako znížiť dimenziu v prípade kognitívnych sociálnych štruktúr, je v zmysle diagonálnych rezov, kde $k = i$ alebo $k = j$ a teda:

- riadkovo dominantné LAS: $\mathfrak{R}'_{i,j} = \mathfrak{R}_{i,j,i}$,
- stĺpcovo dominantné LAS: $\mathfrak{R}'_{i,j} = \mathfrak{R}_{i,j,j}$.

Tento spôsob nazývame lokálne agregovaná štruktúra, pretože výsledný vzťah medzi i a j je závislý na informáciách poskytnutých najlokálnejším z členov v sieti. Potom riadkovo dominantné LAS- $\mathfrak{R}_{i,j,i}$ reprezentuje odpovede na otázku: „S kým si vo vzťahu v danom zmysle?“ a stĺpcovo dominantné LAS- $\mathfrak{R}_{i,j,j}$ obsahuje odpovede na otázku: „Kto je vo vzťahu s tebou?“. Zatiaľ čo väčšina prieskumov týkajúca sa sietí sa pýta iba jednu zo spomínaných typov otázok, často sú požadované odpovede na obe otázky za účelom zlúčiť ich v snahe zvýšiť spoľahlivosť merania. Keďže CSS dátové štruktúry automaticky obsahujú obe formy, je možné kombinovať tieto informácie ďalej buď ako prienik alebo zjednotenie:

- LAS- prienik: $\mathfrak{R}'_{i,j} = \{\mathfrak{R}_{i,j,i} \cap \mathfrak{R}_{i,j,j}\}$,
- LAS- zjednotenie: $\mathfrak{R}'_{i,j} = \{\mathfrak{R}_{i,j,i} \cup \mathfrak{R}_{i,j,j}\}$.

Jedným z prípadov využitia tejto metódy je, keď predpokladáme, že i a j sú tí, ktorí najlepšie vedia, či daný vzťah medzi nimi naozaj existuje. Ak sa na to teda pozrieme ako na definíciu priateľského vzťahu, potom by malo platiť, že i a j sú priatelia, ak obaja sú si vedomí toho, že sú priatelia.

6.2.3 Štruktúry konsenzu (CS)

Posledný z prístupov pri kognitívnych sociálnych štruktúrach je zvážiť celý vektor vnímania (i, j) informácií potrebný na určenie vzťahu (i, j) . Všeobecná forma vyzerá takto:

$$\mathfrak{R}'_{i,j} = f(\mathfrak{R}_{i,j,k_1}, \mathfrak{R}_{i,j,k_2}, \dots, \mathfrak{R}_{i,j,k_n}).$$

Praktcké v prípade CS je použitie funkcie:

$$\mathfrak{R}'_{i,j} = \begin{cases} 1, & ak \frac{1}{N} \sum_k \mathfrak{R}_{i,j,k} \geq T \\ 0, & inak, \end{cases}$$

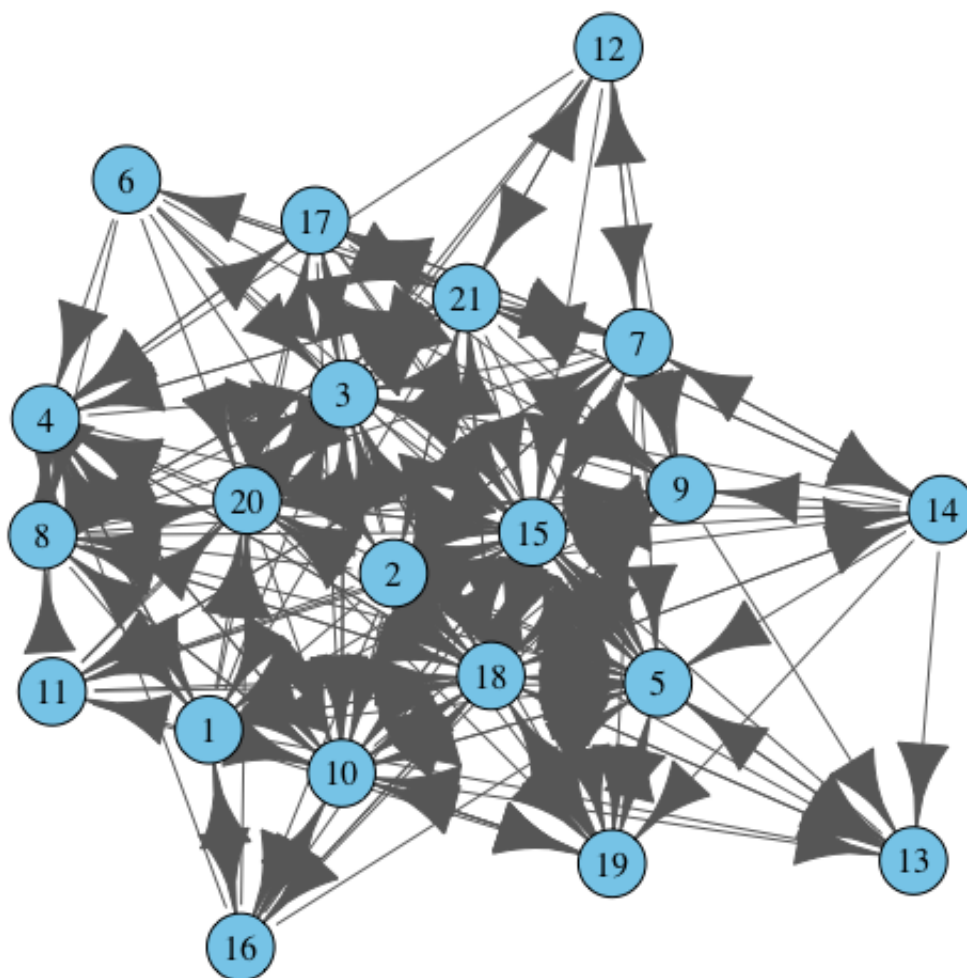
kde T nadobúda hodnotu od 0 do 1. Ak napr. $T = 0,5$, znamená to, že vzťah medzi i a j existuje vtedy a len vtedy, ak väčšina členov siete vníma existenciu tohto vzťahu.

6.3 Príklad použitia CSS

Krackhardt vo svojej práci [7] ukázal spomínané typy zoskupovania na dátach zozbieraných v malej podnikateľskej firme, ktorá sa zameriavala na výrobu strojov pre ostatné spoločnosti. Táto firma zamestnávala 100 ľudí, z toho 21 bolo súčasťou manažmentu. Riaditeľ firmy poskytol príležitosť zozbierať dáta od manažérskeho tímu s cieľom vyhodnotiť dopad nedávneho zásahu do organizačného rozvoja.

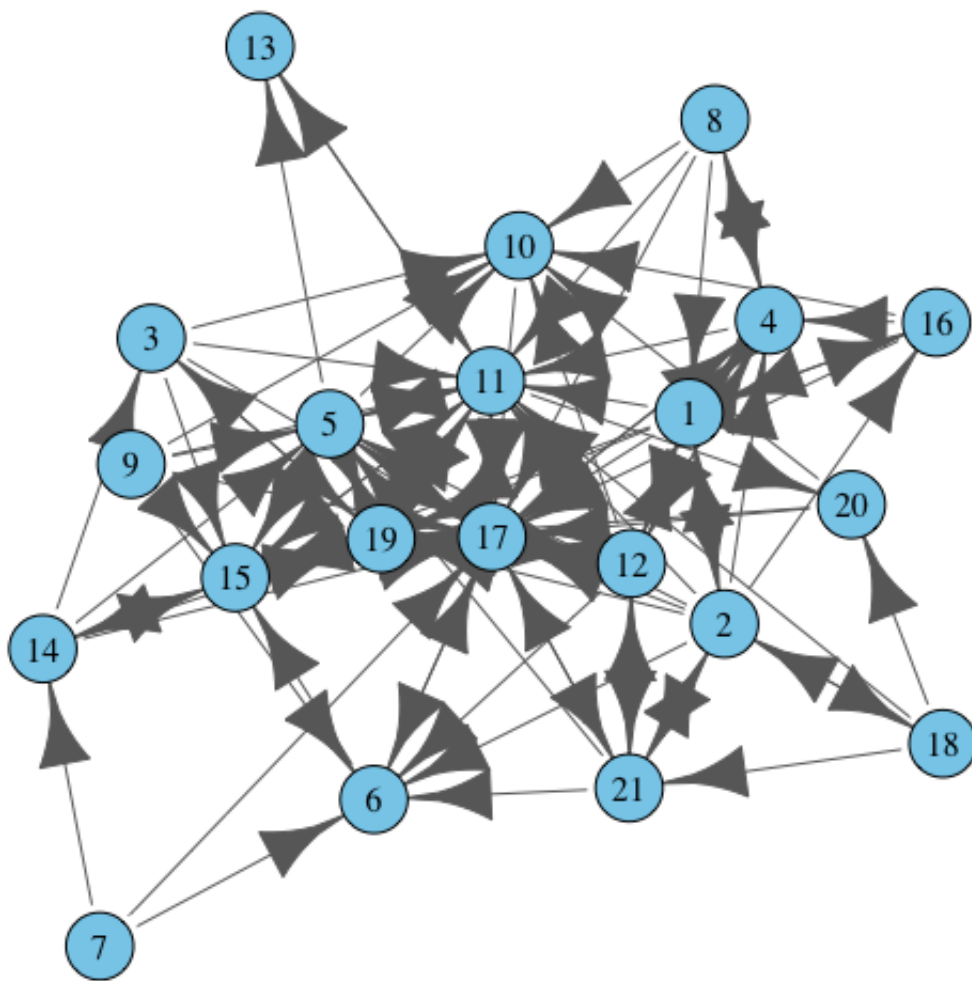
Každý z týchto 21 dotyčných vyplnilo dotazník, ktorého súčasťou boli otázky o tom, kto by sa išiel s kým poradiť ohľadom problémov týkajúcich sa práce. Jedna z otázok bola: „Za kým by XY išiel po pomoc alebo radu ohľadom práce?“, druhá bola: „Kto je priateľ s XY?“. Respondenti boli požiadaní, aby označili mená všetkých ľudí zo zoznamu, za ktorými by XY išiel po radu a tiež s kým je v priateľskom vzťahu.

Maticice, ktoré zachytávajú vnímanie siete každého účastníka prieskumu, nájdeme v [14]. Na obrázku 25 vidíme sieť z Krackhardtovho výskumu, ktorá popisuje odpovede na otázku- „Za kým by XY išiel po pomoc alebo radu ohľadom práce?“, a teda celkové vnímanie siete ohľadom práce. Obrázok 26 popisuje priateľstvá v sieti manažérov firmy, ktoré boli zistené otázkou „Kto je priateľ s XY?“.



Obr. 25: Krackhardt - rada ohľadom práce

Takto zozbierané údaje vytvorili kognitívnu sociálnu štruktúru definovanú vyššie. Z týchto dát bolo vybraných 21 rezov (Slices) pre $k = 1, \dots, 21$, jedna LAS a CS. Metóda LAS bola použitá ako prienik $\mathfrak{R}_{i,j,i} \cap \mathfrak{R}_{i,j,j}$, čo znamená, že vzťah ohľadom rady bol považovaný za existujúci práve vtedy, keď sa obe strany vo vzťahu zhodli, že existuje.

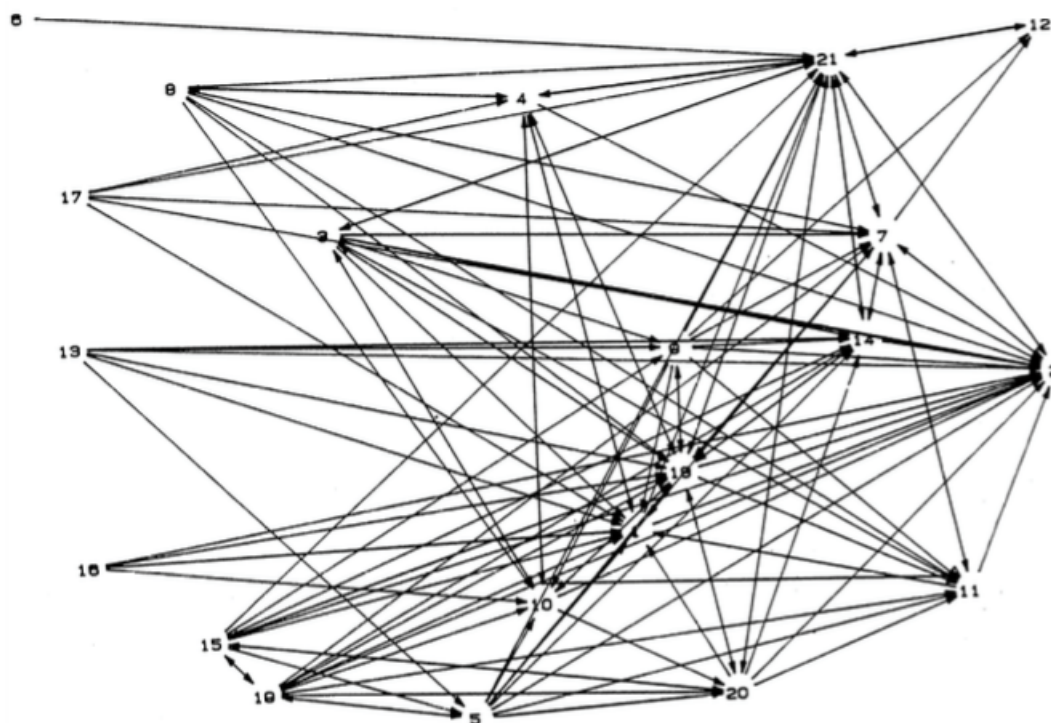


Obr. 26: Krackhardt - priateľstvá

Pri CS bola použitá táto funkcia:

$$\mathfrak{R}'_{i,j} = \begin{cases} 1, & \text{ak } \frac{1}{21} \sum_{k=1}^{21} \mathfrak{R}_{i,j,k} \geq 0,5 \\ 0, & \text{inak.} \end{cases}$$

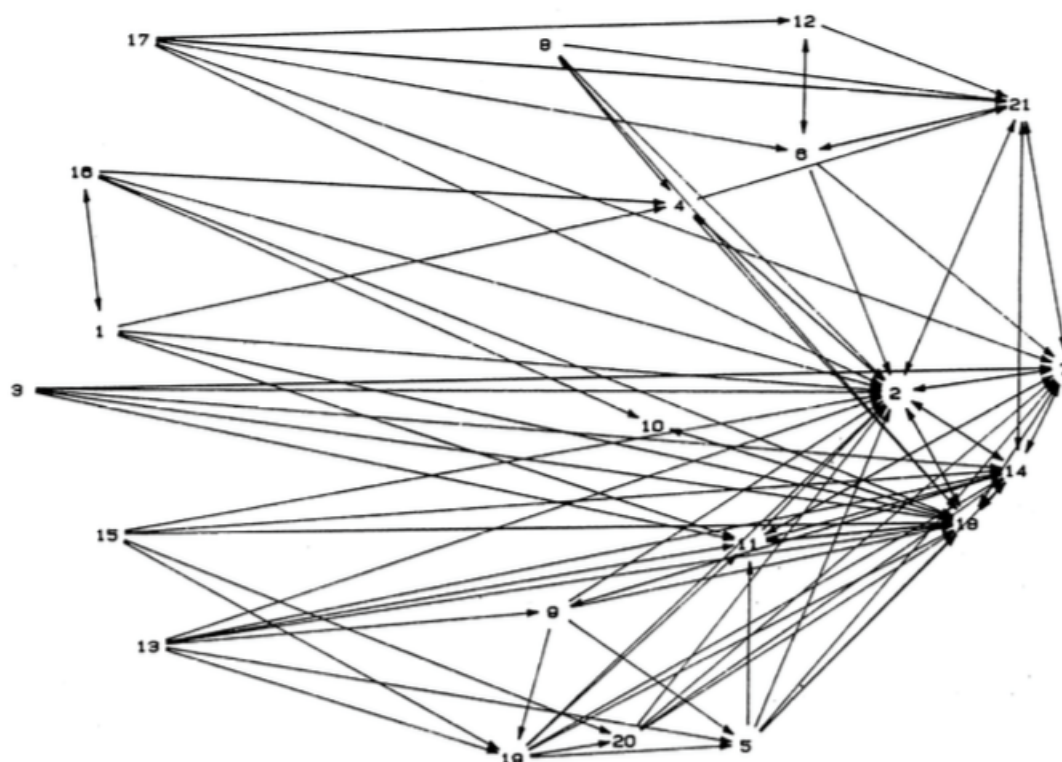
Na obrázku 27 máme zobrazený tzv. sociogram štruktúry LAS. Jasne sú vidno dvaja členovia, ku ktorým by si išlo po radu najviac opýtaných- čísla 2 a 21, ktoré reprezentujú viceprezidentov. Tiež je viditeľná istá hierarchia v rámci vzťahov.



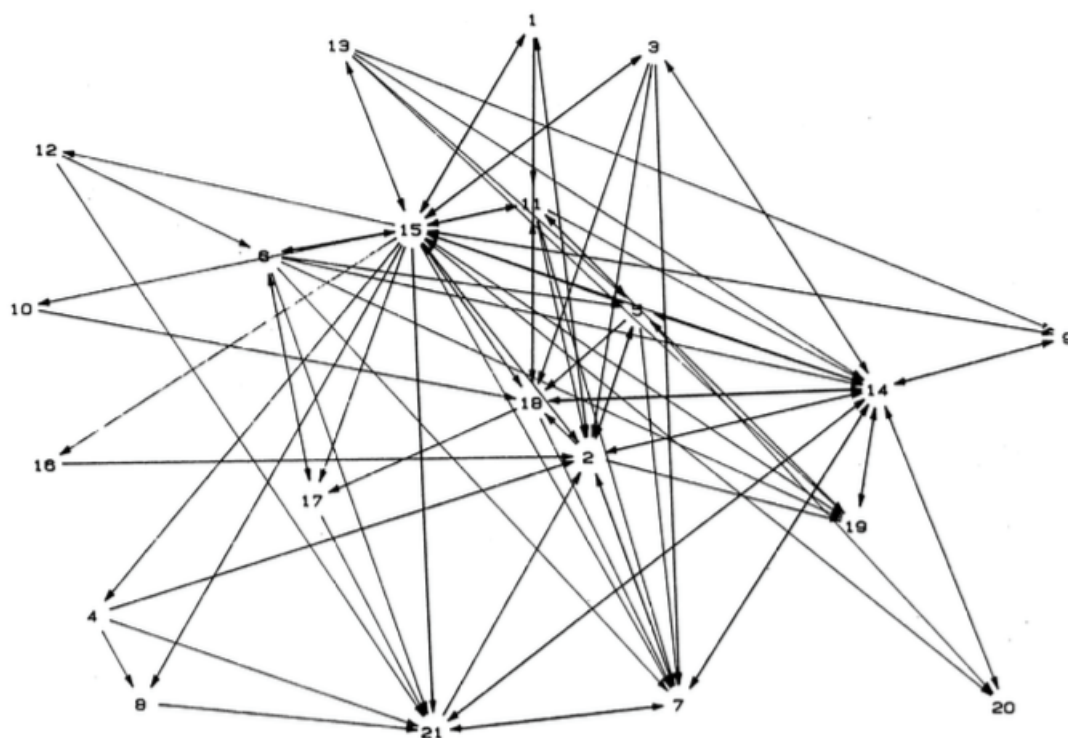
Obr. 27: Sociogram LAS [7]

Graf štruktúry konsenzu (CS), vid' obrázok 28, tiež zachytáva istú hierarchiu v rámci siete vzťahov, kde opäť vrchol 2 dosiahol najväčší počet „nominácií“ ohľadom rady. Vidíme však zmenu oproti LAS štruktúre, keďže 21 stráca dôležitosť a vrcholy 18, 14 a 7 sa zdajú byť centrálnejšie.

Individuálny pohľad osoby s číslom 15 je na obrázku 29, kde seba samého táto osoba umiestnila do stredu s tým, že poskytuje radu a sám ju aj hľadá od mnohých iných. Zo štruktúry LAS však vidno, že o jeho radu stálo málo ľudí a podľa CS dokonca nikto. Tiež podľa osoby 15 takmer každý v sieti dáva a pýta si radu, čím vylučuje jasnú hierarchiu, ktorá je zjavná z obidvoch obrázkov 27 a 28.



Obr. 28: Sociogram CS [7]



Obr. 29: Sociogram rezu osoby 15 [7]

7 Aplikácia Degrootovho modelu

Podobne ako Krackhardt vo svojej práci [7], údaje potrebné na testovanie modelu sme získali z dotazníkov. Zvolili sme dve skupiny ľudí- kolektív učiteliek v materskej škole a študentov druhého ročníka magisterského štúdia odboru Ekonomicko-finančná matematika a modelovanie. Obidva kolektívy boli požiadané o vyplnenie dotazníkov prostredníctvom internetu. V spomínanej terminológii teda učiteľky a spolužiaci predstavujú vrcholy a priznaný vzťah medzi týmito vrcholmi predstavuje hranu v sieti.

7.1 Kolektív učiteliek v materskej škole

Súčasťou kolektívu v materskej škole bolo 12 učiteliek. Dotazník, ktorí vyplňali, obsahoval dve základné otázky:

- Za kým by **XY** išla po radu ohľadom problematickeho dieťaťa?
- Za kým by **XY** išla po radu týkajúcu sa súkromia?

Učiteľky boli požiadané, aby označili mená všetkých kolegyň, o ktorých si myslia, že by sa s nimi osoba XY išla poradiť ohľadom problematickeho dieťaťa, príp. súkromia. V konečnom dôsledku tak každá z vyplňajúcich odpovedala dokopy na 24 otázok. Dostali sme teda 24 matíc popisujúcich odpovede každej z opýtaných, z čoho 12 sa týkalo problematickeho dieťaťa a zvyšných 12 súkromia.

7.1.1 Analýza otázky o problematickom dieťati

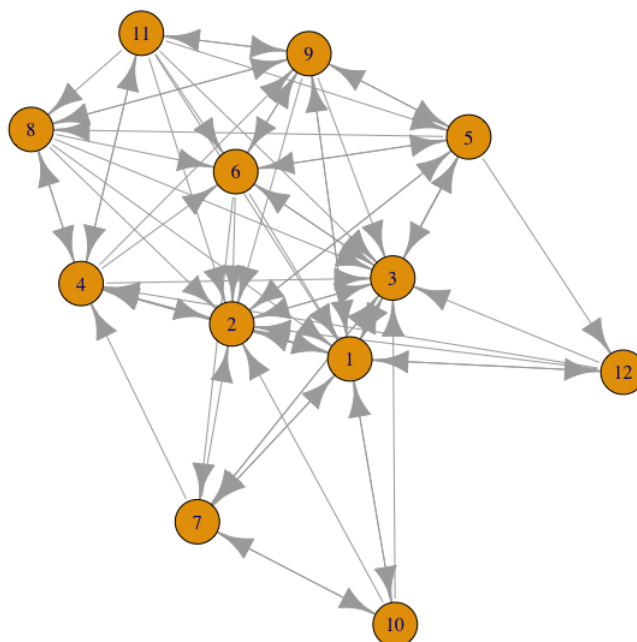
Sčítaním všetkých 12 matíc, ktoré zaznamenali odpovede na otázku o problematickom dieťati, sme dostali maticu D . Prvok matice D_{ij} hovorí o tom, koľko osôb si myslí, že by osoba i išla po radu ohľadom problematickeho dieťaťa za osobou j . Následne sme z tejto matice vytvorili stochastickú riadkovú maticu D_s , kde je súčet každého riadku rovný 1. Pre túto maticu teda platí, že váha je úmerná hodnote v matici D , t.j. čím viac ľudí si myslí, že existuje vzťah, tým tomu dáva väčšiu váhu. Otestovali sme podmienku konvergenzie matice a zistili sme, že existuje už tretia mocnina matice D , ktorej všetky prvky sú kladné, čím je splnená táto podmienka. Postupným iterovaním sme sa dostali aj k vektoru vplyvov.

$$D = \begin{pmatrix} 0 & 8 & 9 & 6 & 1 & 0 & 4 & 0 & 1 & 4 & 0 & 2 \\ 5 & 0 & 12 & 1 & 7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 8 & 0 & 0 & 1 & 5 & 1 & 0 & 0 & 0 & 0 & 0 \\ 9 & 6 & 5 & 0 & 0 & 1 & 0 & 7 & 1 & 0 & 1 & 0 \\ 1 & 10 & 5 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 4 & 4 & 12 & 0 & 4 & 0 & 2 & 0 & 2 & 0 & 0 & 0 \\ 9 & 1 & 9 & 1 & 0 & 0 & 0 & 0 & 0 & 8 & 0 & 0 \\ 4 & 2 & 8 & 8 & 0 & 1 & 0 & 0 & 9 & 0 & 0 & 0 \\ 1 & 6 & 5 & 0 & 1 & 1 & 0 & 9 & 0 & 0 & 1 & 0 \\ 8 & 2 & 8 & 0 & 0 & 0 & 9 & 0 & 0 & 0 & 0 & 0 \\ 1 & 8 & 2 & 1 & 1 & 1 & 0 & 2 & 12 & 0 & 0 & 0 \\ 12 & 11 & 3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

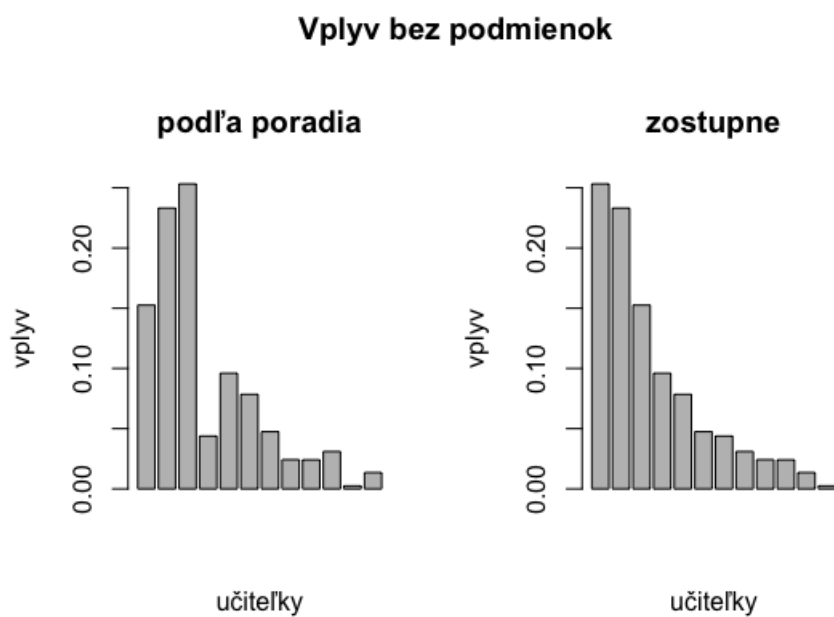
$$D_s = \begin{pmatrix} 0 & 0,23 & 0,26 & 0,17 & 0,03 & 0 & 0,11 & 0 & 0,03 & 0,11 & 0 & 0,06 \\ 0,20 & 0 & 0,48 & 0,04 & 0,28 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0,17 & 0,44 & 0 & 0 & 0,06 & 0,28 & 0,06 & 0 & 0 & 0 & 0 & 0 \\ 0,30 & 0,20 & 0,17 & 0 & 0 & 0,03 & 0 & 0,23 & 0,03 & 0 & 0,03 & 0 \\ 0,05 & 0,50 & 0,25 & 0 & 0 & 0,05 & 0 & 0,05 & 0,05 & 0 & 0 & 0,05 \\ 0,14 & 0,14 & 0,43 & 0 & 0,14 & 0 & 0,07 & 0 & 0,07 & 0 & 0 & 0 \\ 0,32 & 0,04 & 0,32 & 0,04 & 0 & 0 & 0 & 0 & 0 & 0,28 & 0 & 0 \\ 0,13 & 0,06 & 0,25 & 0,25 & 0 & 0,03 & 0 & 0 & 0,28 & 0 & 0 & 0 \\ 0,04 & 0,25 & 0,21 & 0 & 0,04 & 0,04 & 0 & 0,38 & 0 & 0 & 0,04 & 0 \\ 0,30 & 0,07 & 0,30 & 0 & 0 & 0 & 0,33 & 0 & 0 & 0 & 0 & 0 \\ 0,04 & 0,29 & 0,07 & 0,04 & 0,04 & 0,04 & 0 & 0,07 & 0,43 & 0 & 0 & 0 \\ 0,44 & 0,41 & 0,11 & 0,04 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Sieť reprezentovaná stochastickou maticou, ktorá vznikla upravením matice D , je zobrazená na obrázku 30. Vidíme, že k vrcholom 1, 2 a 3 smeruje najviac šípok, t.j. predpokladáme, že by mali byť v sieti najvplyvnejšie. Grafy na obrázku 31 ukazujú jednotlivé vplyvy a potvrdzujú, že medzi najvplyvnejšie vrcholy patria spomínané vrcholy 1, 2 a 3. Tiež sa zdá, že by sa učiteľky dali zaradiť do 2 alebo 3 skupín podľa vplyvu, ktorý majú v tejto sieti. Logicky by sme jednu zo skupín vytvorili práve z vrcholov 1,

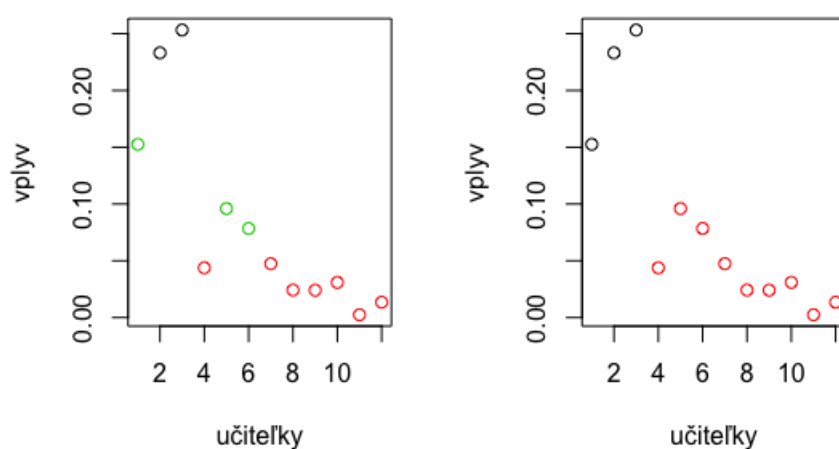
2, 3. Toto platí iba v prípade, že rozdelíme všetky vrcholy do 2 skupín. Keby sme chceli vytvoriť 3 skupiny, vrchol 1 bude súčasťou inej skupiny ako vrcholy 2 a 3, čo môžeme vidieť na obrázku 32.



Obr. 30: Sieť k otázke o problematickom dieťati



Obr. 31: Vplyv v sieti k otázke o problematickom dieťati



Obr. 32: Zhluky v sieti k otázke o problematickom dieťati

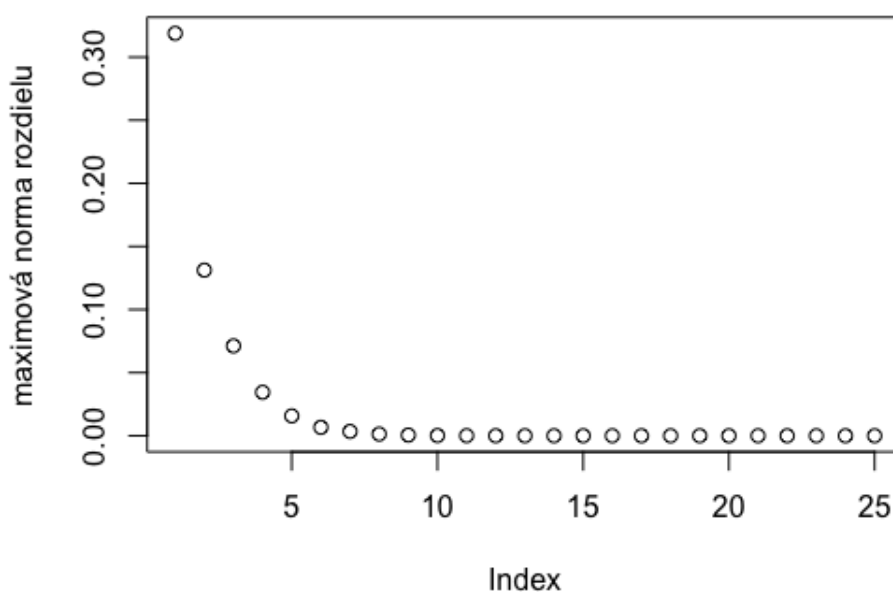
Ďalej sa pozrieme, ako sa zmení sieť a jednotlivé vplyvy, ak stanovíme podmienku na existenciu hrany medzi vrcholmi. Povedzme teda, že budeme akceptovať iba ten vzťah medzi dvoma učiteľkami, ktorý bol označený ako existujúci viac ako 1 učiteľkou. To znamená, že v matici D berieme do úvahy iba tie prvky D_{ij} , ktoré sa rovnajú 2 a viac. Týmto priradíme číslo 1, t.j. existuje vzťah a ostatným priradíme 0. Dostaneme nasledovnú maticu

$$D' = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Opäť je potrebné, otestovať, či táto matica spĺňa podmienky konvergenie a privedie

nás k vektoru vplyvov. Maticu D' sme teda upravili, aby sme dostali stochastickú maticu a hľadali sme mocninu a , pri ktorej budú všetky prvky stochastickej matice kladné. Mocnina matice však nespĺňala postačujúcu podmienku konvergence kladnosti prvkov. Numericky sa však zdá, že mocniny aj napriek tomu konvergujú k matici získanej z vlastného vektora prislúchajúceho k vlastnej hodnote 1. Ukážeme tento numerický výpočet a potom sa vrátíme k možnosti analytického dôkazu.

Zo získaného vektora vplyvov, ktorý je totožný s vlastným vektorom danej stochastickej matice, vytvoríme štvorcovú maticu D'_{ss} (v našom prípade 12×12), ktorej riadky sú rovné tomuto vektoru. Porovnáme túto maticu D'_{ss} s a -tou mocninou stochastickej matice D' , pričom chceme, aby norma rozdielu $D'^a - D'_{ss}$ išla k nule. Použijeme maximovú normu, teda sa pozeráme na rozdiel $\max |D'^a - D'_{ss}|$. Na obrázku 33 vidíme, že norma rozdielu ide k nule, čo ilustruje konvergenciu k matici s riadkami rovnými vektoru vplyvov.

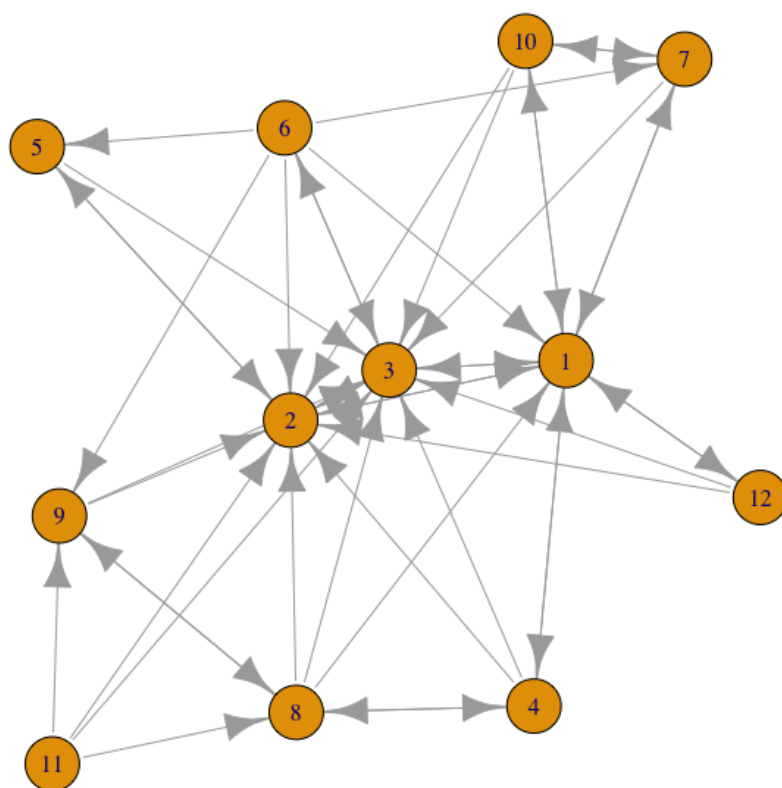


Obr. 33: Norma rozdielu ilustrujúca konvergenciu

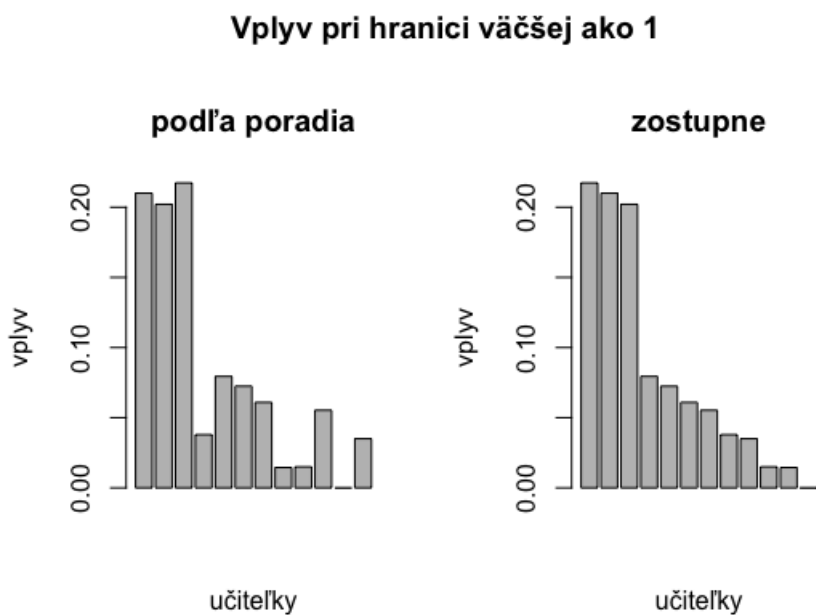
Všeobecnejším tvrdením je veta 5.3, museli by sme však overovať aperiodickosť. Jednoduchšou možnosťou bolo overovanie podmienky z kapitoly 5.4. Tá nám hovorí o konvergencii presvedčení. V tomto prípade máme takéto vlastné hodnoty: $\lambda_1 = 1$, $\lambda_2 = -0.296 + 0.267i$, $\lambda_3 = -0.296 - 0.267i$, $\lambda_4 = -0.395 + 0.047i$, $\lambda_5 = -0.395 - 0.047i$, $\lambda_6 =$

0.347 , $\lambda_7 = -0.311$, $\lambda_8 = 0.302$, $\lambda_9 = 0.139$, $\lambda_{10} = -0.093$, $\lambda_{11} = 6.239 \cdot 10^{-17}$, $\lambda_{12} = 0$. Takto je to aj v iných prípadoch, ale nebudeme to uvádzať pri každom, keďže postup overovania je rovnaký.

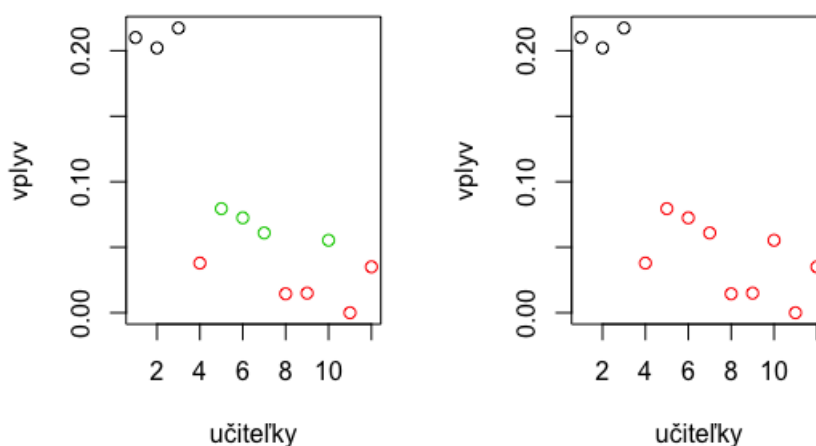
Matica D' charakterizuje sieť na obrázku 34. Vidno na nej, že najvplyvnejšími stále zostávajú vrcholy 1, 2 a 3, rovnako ako to ukazuje obrázok 35. Rozdiel oproti prvému príkladu bez stanovenej hranice je, že pri vytváraní zhlukov vplyvu je viditeľne odčlenená trojica vrcholov 1, 2 a 3 v oboch prípadoch, viď obrázok 36.



Obr. 34: Sieť k otázke o problematickom dieťati pri hranici viac ako 1



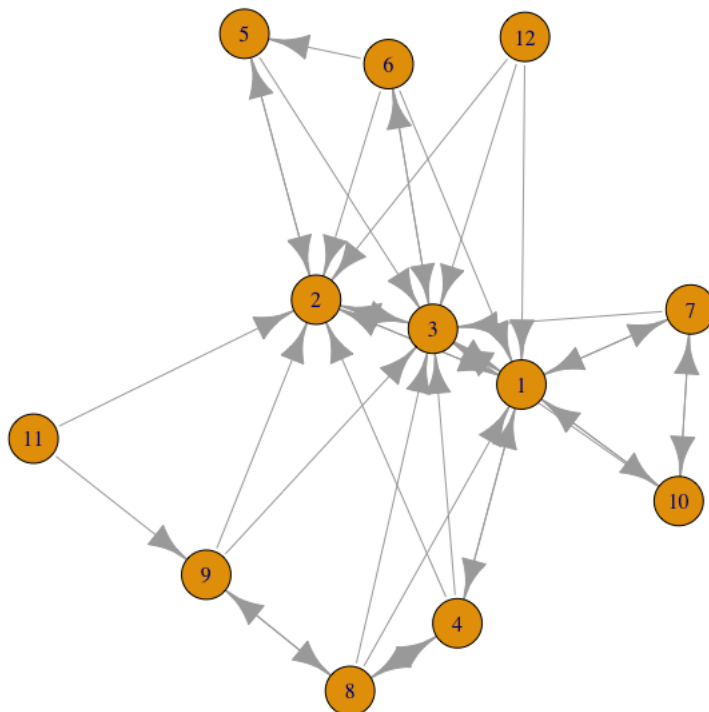
Obr. 35: Vplyv v sieti k otázke o problematickom dieťati pri hranici viac ako 1



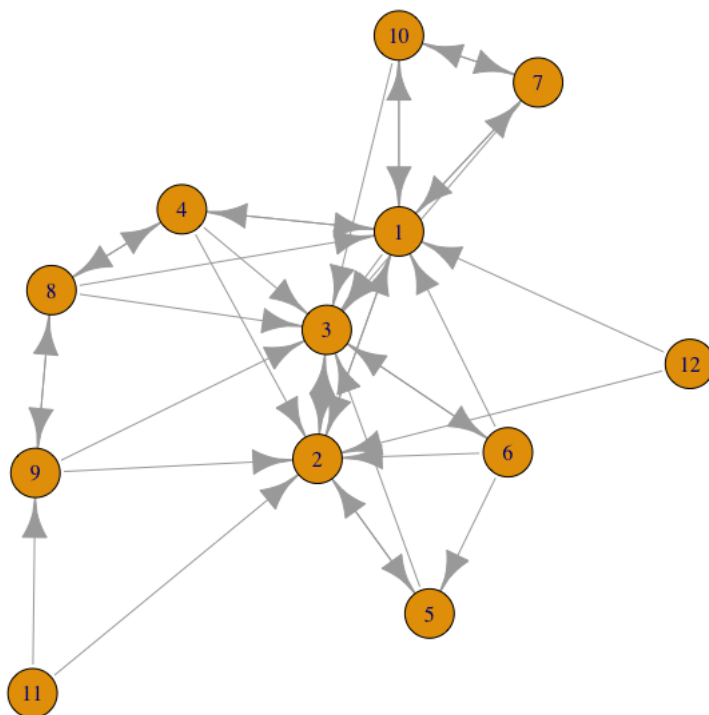
Obr. 36: Zhluky v sieti k otázke o problemtickom dieťati pri hranici viac ako 1

Pri stanovených hraniciach viac ako 2, viac ako 3 a viac ako 4 mocniny vytvorených stochastických matíc nespĺňali postačujúcu podmienku na konvergenciu, zopakovali sme teda postup z predchádzajúceho prípadu a overili sme, či norma rozdielu matice skladajúcej sa z vlastného vektoru vplyvov a mocnina matice ide k nule. Vo všetkých prípadoch sa potvrdilo, že norma rozdielu ide k nule a dostali sme požadované vektory vplyvov. Jednotlivé siete vidíme na obrázkoch 37, 38 a 39. Na obrázkoch 40, 41 a 42

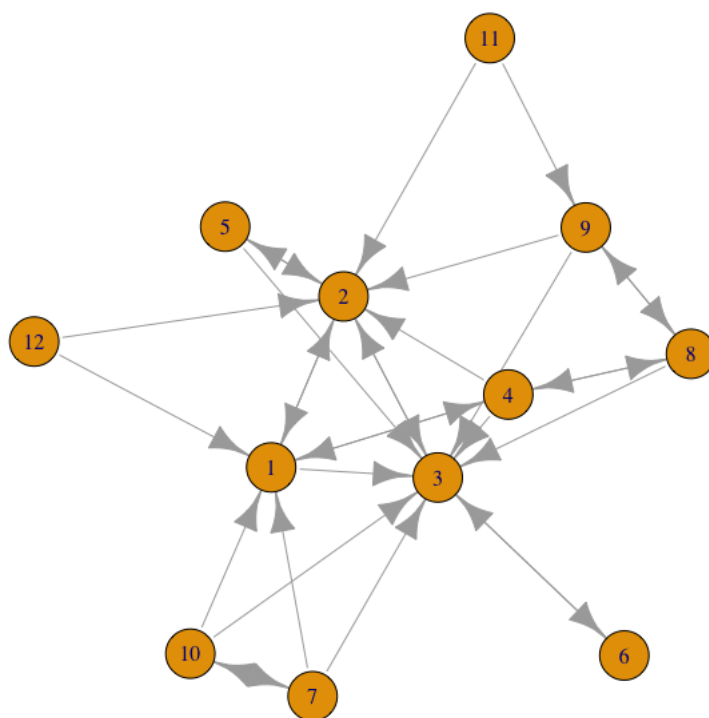
vidíme porovnanie vektorov vplyvov. Zhlukovanie pri jednotlivých hraniciach máme na obrázkoch 43, 44 a 45.



Obr. 37: Sieť k otázke o problematickom dieťati pri hranici viac ako 2

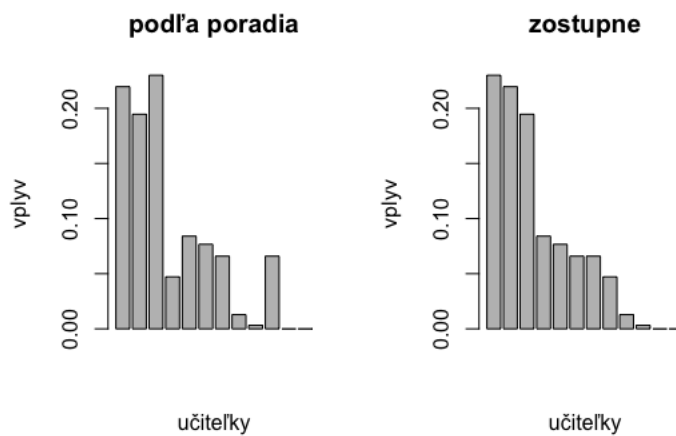


Obr. 38: Sieť k otázke o problematickom dieťati pri hranici viac ako 3

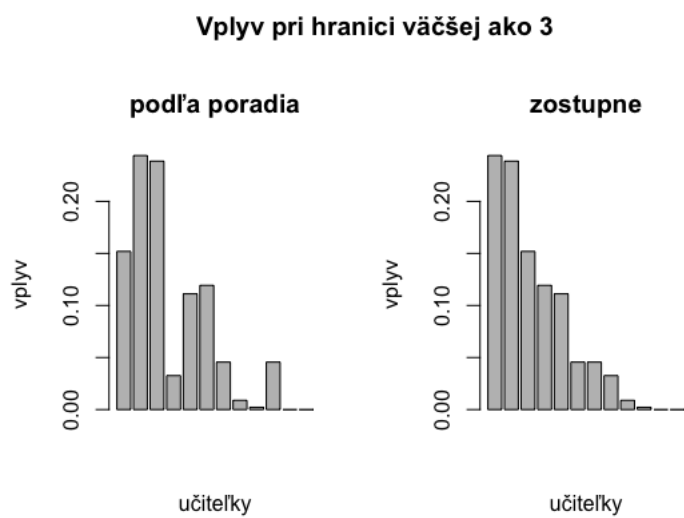


Obr. 39: Sieť k otázke o problematickom dieťati pri hranici viac ako 4

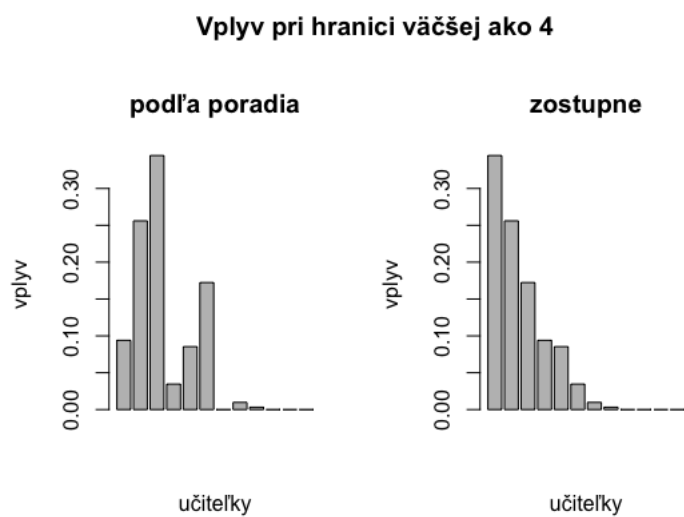
Vplyv pri hranici väčšej ako 2



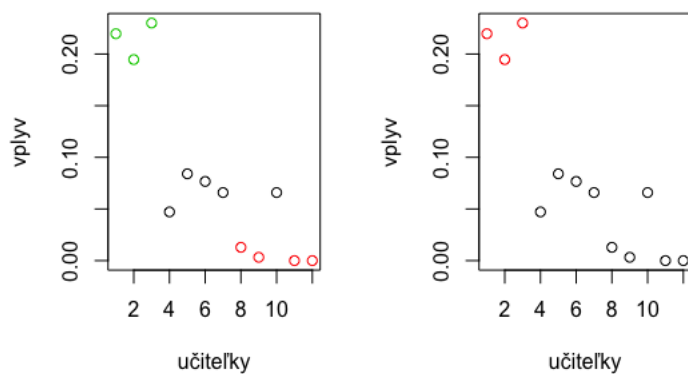
Obr. 40: Vplyv k otázke o problematickom dieťati pri hranici viac ako 2



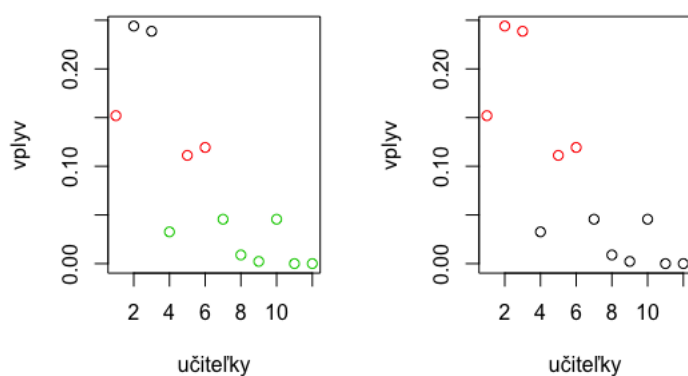
Obr. 41: Vplyv k otázke o problematickom dieťati pri hranici viac ako 3



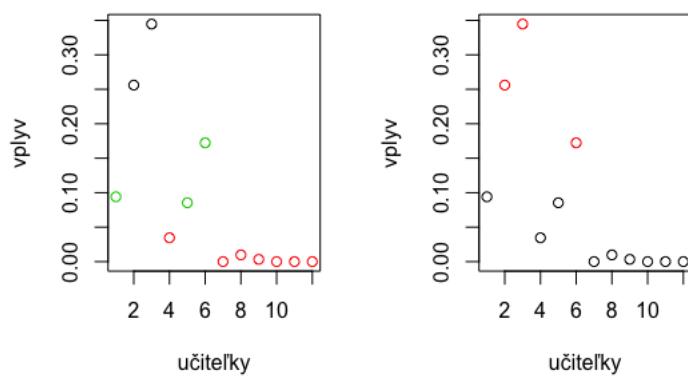
Obr. 42: Vplyv k otázke o problematickom dieťati pri hranici viac ako 4



Obr. 43: Zhluky k otázke o problematickom dieťati pri hranici viac ako 2



Obr. 44: Zhluky k otázke o problematickom dieťati pri hranici viac ako 3



Obr. 45: Zhluky k otázke o problematickom dieťati pri hranici viac ako 4

Prehľadné porovnanie vplyvov pri jednotlivých hraniciach pre každý vrchol v sieti vidno v tabuľke 2.

Vrchol	bez hranice	hranica > 1	hranica > 2	hranica > 3	hranica > 4
1	0.15255	0.21007	0.21964	0.15194	0.09400
2	0.23315	0.20211	0.19464	0.24402	0.25608
3	0.25326	0.21734	0.23000	0.23865	0.34468
4	0.04382	0.03790	0.04714	0.03261	0.03458
5	0.09601	0.07945	0.08405	0.11117	0.08536
6	0.07846	0.07245	0.07667	0.11933	0.17234
7	0.04744	0.06091	0.06589	0.04558	0.00000
8	0.02424	0.01447	0.01286	0.00889	0.00972
9	0.02410	0.01497	0.00321	0.00222	0.00324
10	0.03099	0.05532	0.06589	0.04558	0.00000
11	0.00246	0.00000	0.00000	0.00000	0.00000
12	0.01352	0.03501	0.00000	0.00000	0.00000

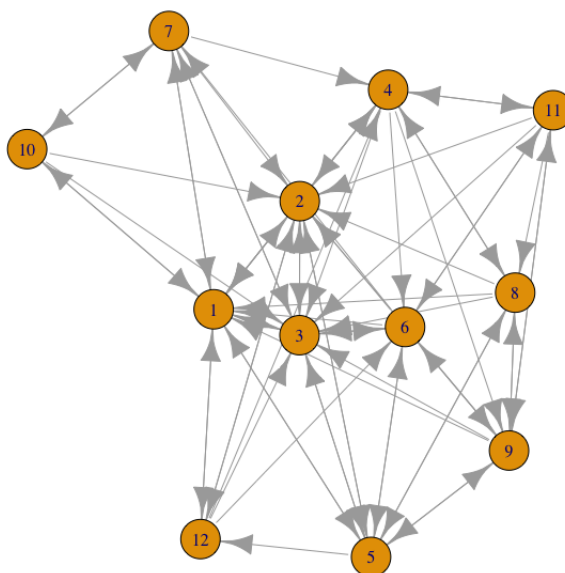
Tabuľka 2: Vplyvy pri stanovených hraniciach pre každý vrchol

7.1.2 Analýza otázky o súkromí

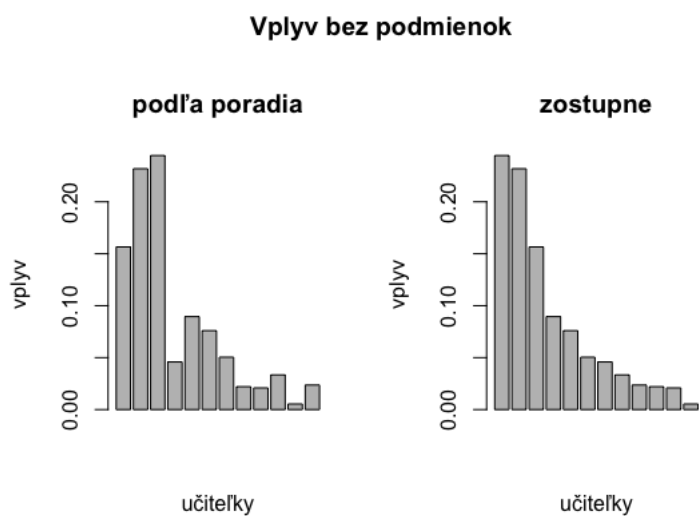
Rovnako ako v predchádzajúcej časti sme sčítali všetkých 12 matíc, ktoré teraz obsahovali odpovede na otázku „Za kým by XY išla po radu týkajúcu sa súkromia?“. Dostali sme maticu S , v ktorej S_{ij} reprezentuje koľko osôb si myslí, že by osoba i išla po radu týkajúcu sa súkromia za osobou j .

$$S = \begin{pmatrix} 0 & 8 & 8 & 6 & 1 & 0 & 4 & 0 & 0 & 4 & 0 & 2 \\ 5 & 0 & 11 & 1 & 6 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 3 & 8 & 0 & 0 & 1 & 5 & 1 & 0 & 0 & 0 & 0 & 0 \\ 8 & 6 & 5 & 0 & 0 & 1 & 0 & 6 & 1 & 0 & 1 & 0 \\ 1 & 10 & 5 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 3 & 3 & 11 & 0 & 4 & 0 & 2 & 0 & 2 & 0 & 1 & 0 \\ 9 & 1 & 9 & 1 & 0 & 0 & 0 & 0 & 0 & 8 & 0 & 0 \\ 5 & 3 & 7 & 7 & 1 & 0 & 0 & 0 & 9 & 0 & 0 & 0 \\ 2 & 6 & 5 & 0 & 1 & 1 & 0 & 9 & 0 & 0 & 1 & 0 \\ 7 & 2 & 7 & 0 & 0 & 0 & 9 & 0 & 0 & 0 & 0 & 0 \\ 0 & 7 & 2 & 1 & 0 & 1 & 0 & 1 & 12 & 0 & 0 & 0 \\ 12 & 10 & 3 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

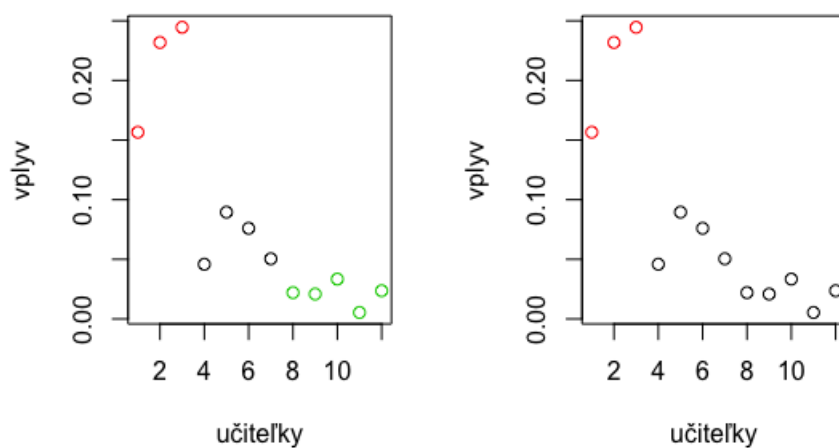
Zopakovali sme postup z prvej otázky a dostali sme sieť, viď obrázok 46. Rovnako aj vektor vplyvov, ktorý je zobrazený na obrázku 47 a zaradenie do skupín s podobným vplyvom vidíme na obrázku 48. Medzi najvplyvnejších patria opäť učiteľky s číslami 1, 2 a 3, tiež vytvorili jeden samostatný zhluk.



Obr. 46: Sieť k otázke o súkromí

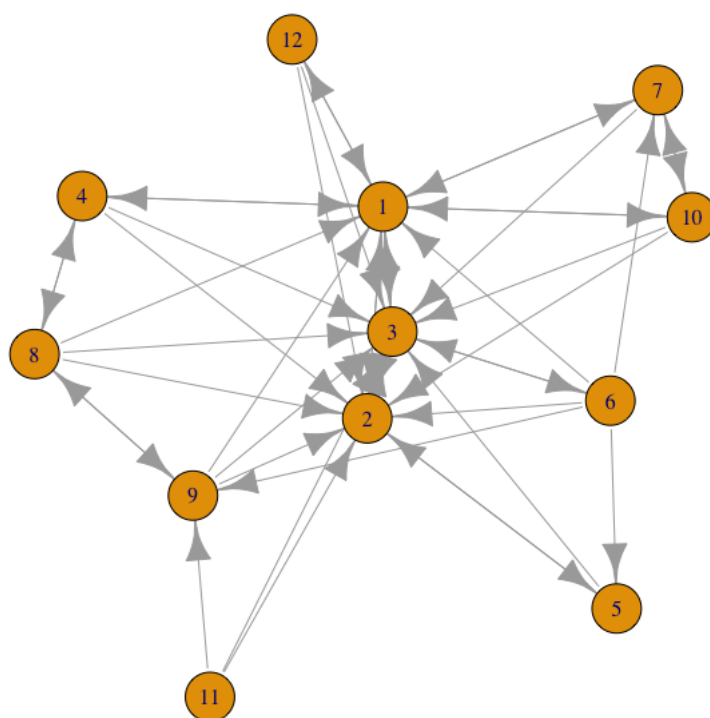


Obr. 47: Vplyv v sieti k otázke o súkromí

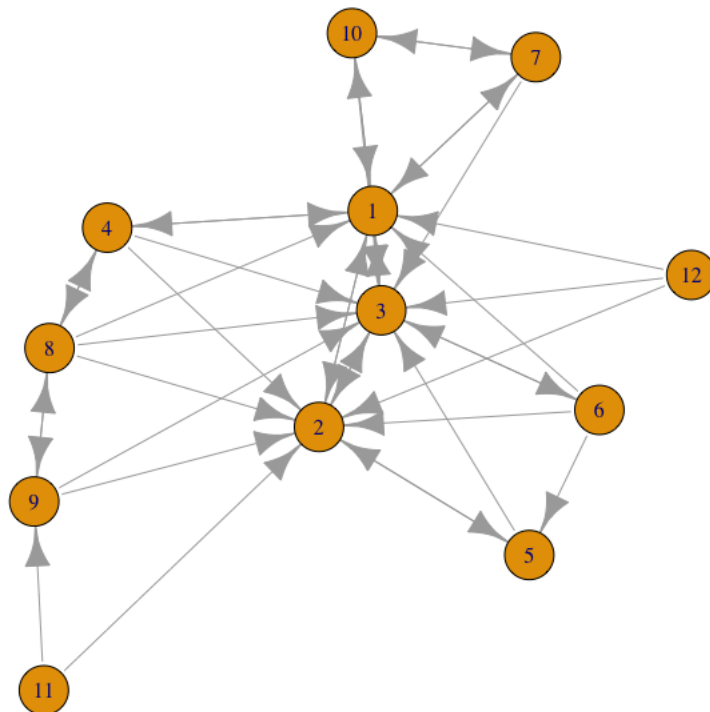


Obr. 48: Zhluky v sieti k otázke o súkromí

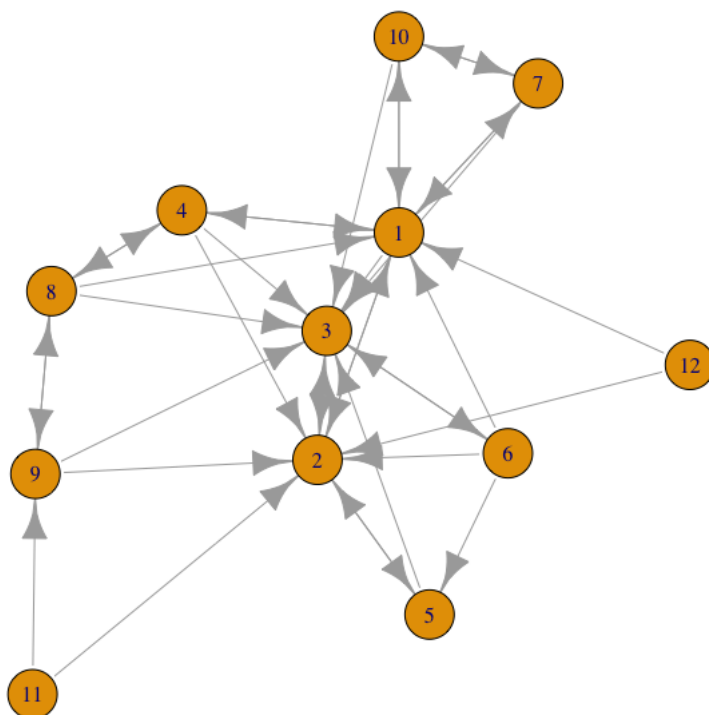
Stanovili sme si hranice na existenciu hrany v sieti. Nižšie vidíme porovnanie sietí pri rôznych hraniciach (obrázky 49, 50, 51 a 52), takisto jednotlivé vplyvy (obrázky 53, 54, 55 a 56) a zhluky (obrázky 57, 58, 59 a 60).



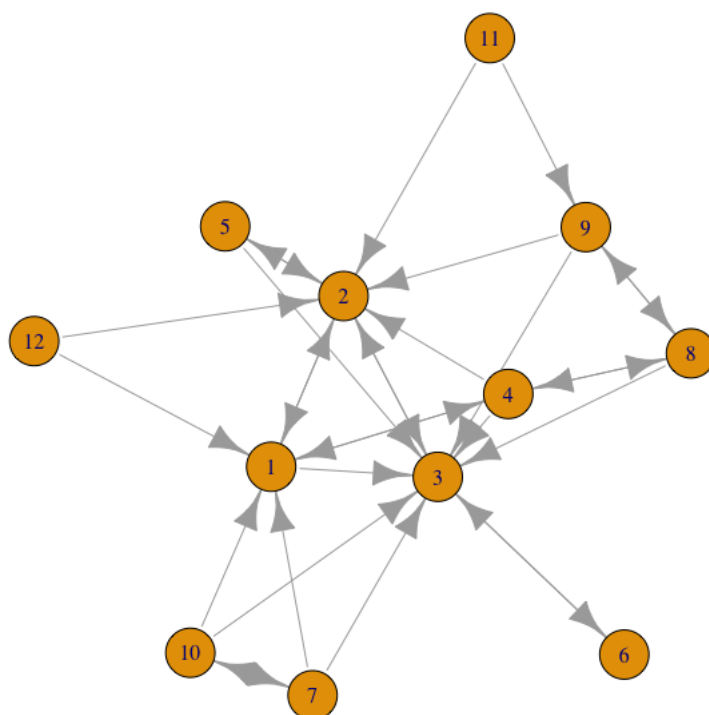
Obr. 49: Sieť k otázke o súkromí pri hranici viac ako 1



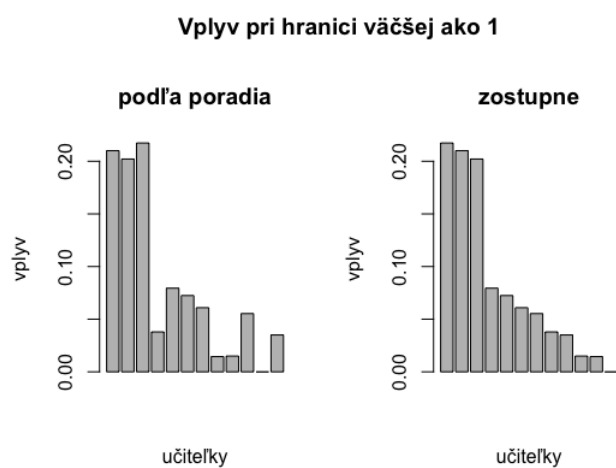
Obr. 50: Sieť k otázke o súkromí pri hranici viac ako 2



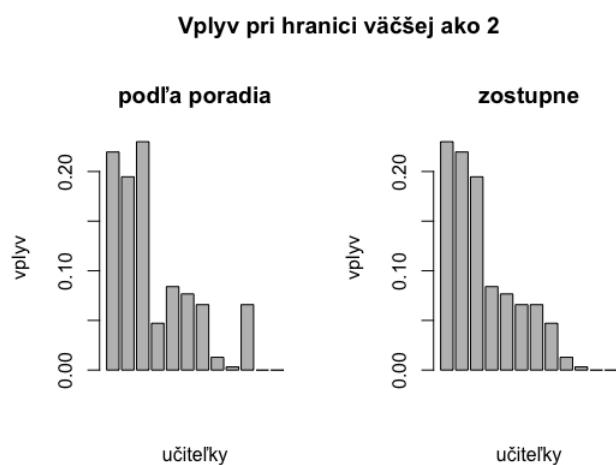
Obr. 51: Sieť k otázke o súkromí pri hranici viac ako 3



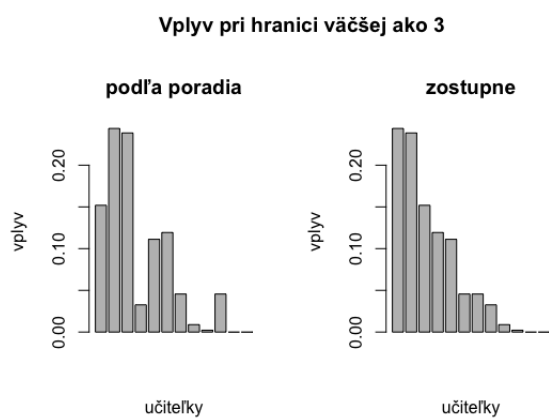
Obr. 52: Sieť k otázke o súkromí pri hranici viac ako 4



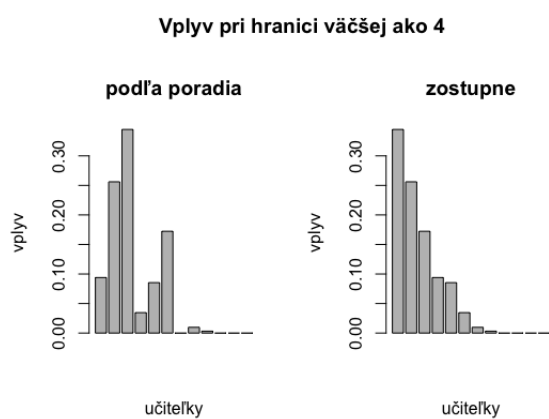
Obr. 53: Vplyv k otázke o súkromí pri hranici viac ako 1



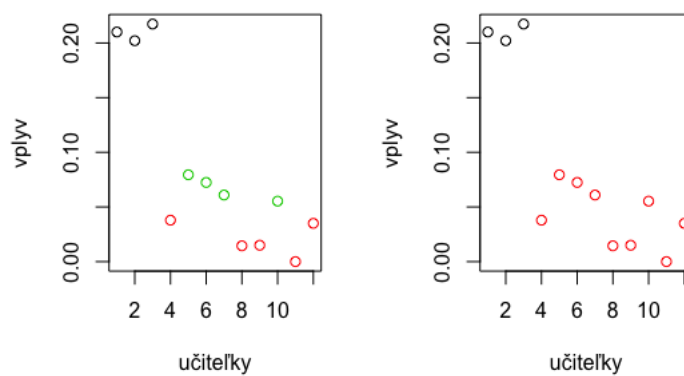
Obr. 54: Vplyv k otázke o súkromí pri hranici viac ako 2



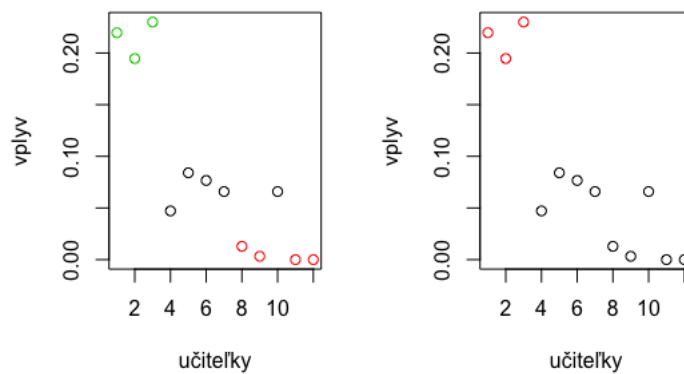
Obr. 55: Vplyv k otázke o súkromí pri hranici viac ako 3



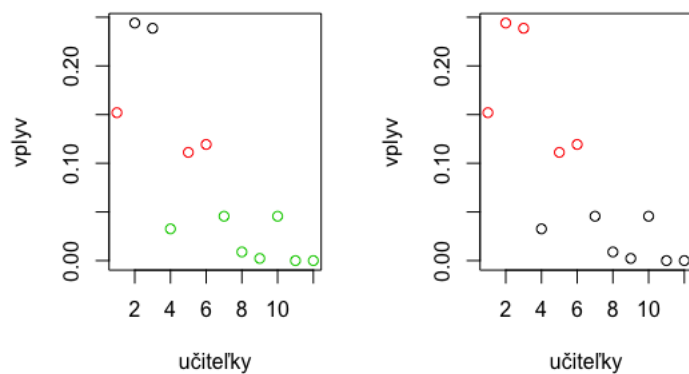
Obr. 56: Vplyv k otázke o súkromí pri hranici viac ako 4



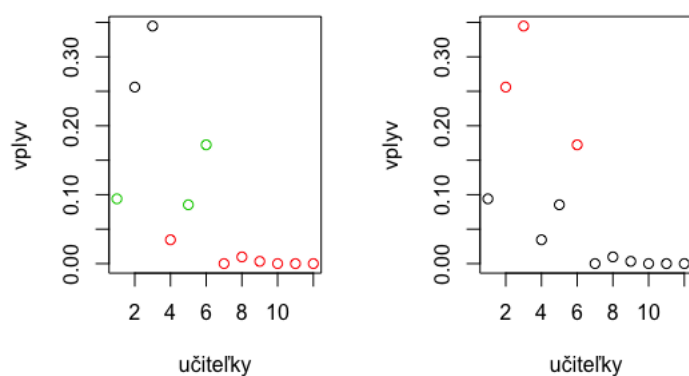
Obr. 57: Zhluky k otázke o súkromí pri hranici viac ako 1



Obr. 58: Zhluky k otázke o súkromí pri hranici viac ako 2



Obr. 59: Zhluky k otázke o súkromí pri hranici viac ako 3



Obr. 60: Zhluky k otázke o súkromí pri hranici viac ako 4

Tabuľka 3 zobrazuje porovnanie vplyvov pri jednotlivých hraniciach pre každý vrchol v sieti.

Vrchol	bez hranice	hranica > 1	hranica > 2	hranica > 3	hranica > 4
1	0.15657	0.21315	0.21926	0.10855	0.09636
2	0.23185	0.20071	0.19686	0.23827	0.25617
3	0.24455	0.21609	0.22975	0.27279	0.34391
4	0.04585	0.03816	0.04634	0.02330	0.03447
5	0.08950	0.07891	0.08476	0.14762	0.08539
6	0.07598	0.07203	0.07658	0.13640	0.17195
7	0.05043	0.06154	0.06578	0.03256	0.00000
8	0.02203	0.01320	0.01241	0.00635	0.00940
9	0.02085	0.01465	0.00248	0.00159	0.00235
10	0.03339	0.05604	0.06578	0.03256	0.00000
11	0.00539	0.00000	0.00000	0.00000	0.00000
12	0.02362	0.03552	0.00000	0.00000	0.00000

Tabuľka 3: Vplyvy pri stanovených hraniciach pre každý vrchol

7.1.3 Miery centralít

V kapitole 4 sme definovali niekoľko druhov centralít, ktoré v tejto časti porovnáme pre jednotlivé vrcholy v sieti popisujúcej vplyv. Pozrieme sa na centralitu zistených troch najvplyvnejších vrcholov pri stanovenej hranici viac ako 2.

Analýza otázky o problematickom dieťati nás dovedla k záveru, že pri hranici viac ako 2 medzi tri najvplyvnejšie vrcholy patria tie s prideleným číslom 1, 2 a 3. Za pravdu nám dávajú výsledky v tabuľke 3. Porovnáme teda tento výsledok s tým, ako centrálné sú tieto vrcholy pri použití rôznych mier centralít, vid' tabuľka 4.

Vrchol	In- degree	Out- degree	Blížkosť	Betweenness	Vlastný vektor
1	0.727	0.455	0.786	0.748	0.923
2	0.818	0.273	0.846	0.297	0.836
3	0.909	0.273	0.917	0.297	1.000

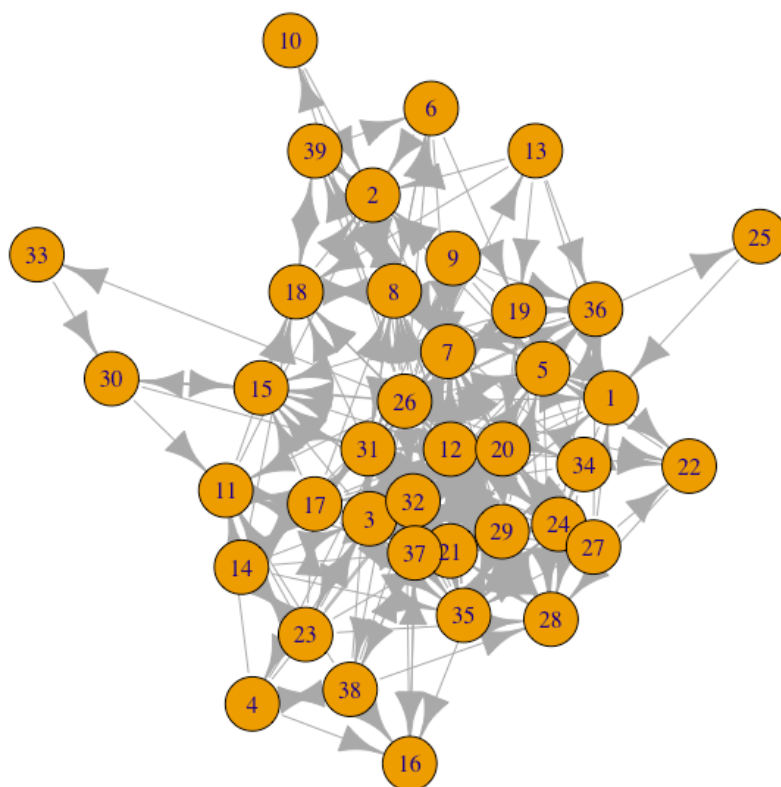
Tabuľka 4: Rôzne miery centralít v sieti k otázke o problematickom dieťati

7.2 Kolektív študentov

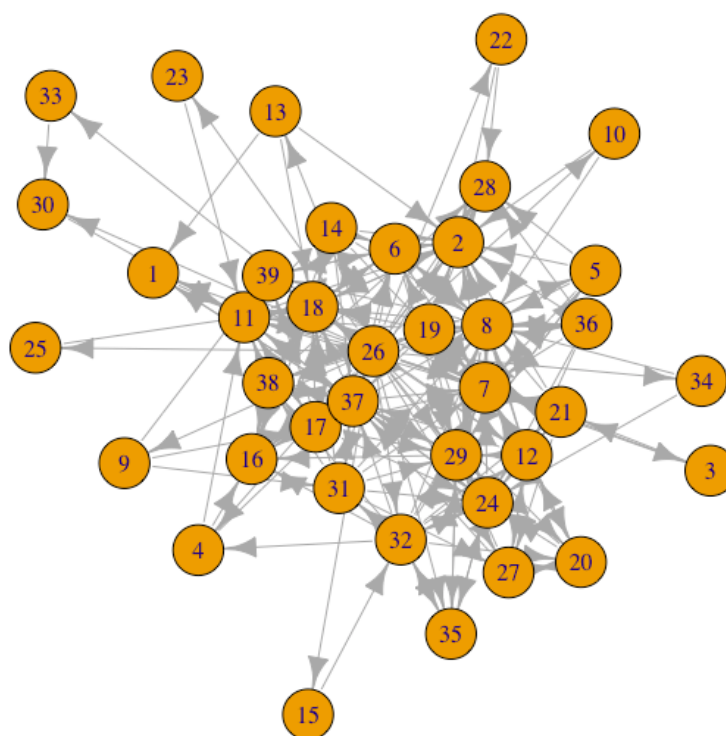
V prípade študentov bol počet vyplňajúcich rovný 39. Spolužiaci odpovedali na tieto otázky:

- č.1: Za kým by si išiel, keby si potreboval poradiť s preberaným učivom?
- č.2: Od koho by si si išiel požičať poznámky z prednášky/cvičenia, na ktorom si nebol?
- č.3: Koho by si oslovil, keď by si potreboval poradiť nejaký dobrý podnik?
- č.4: S kým by si sa riešil svoje osobné problémy?
- č.5: Koho by si kontaktoval v prípade, že zháňaš prácu?

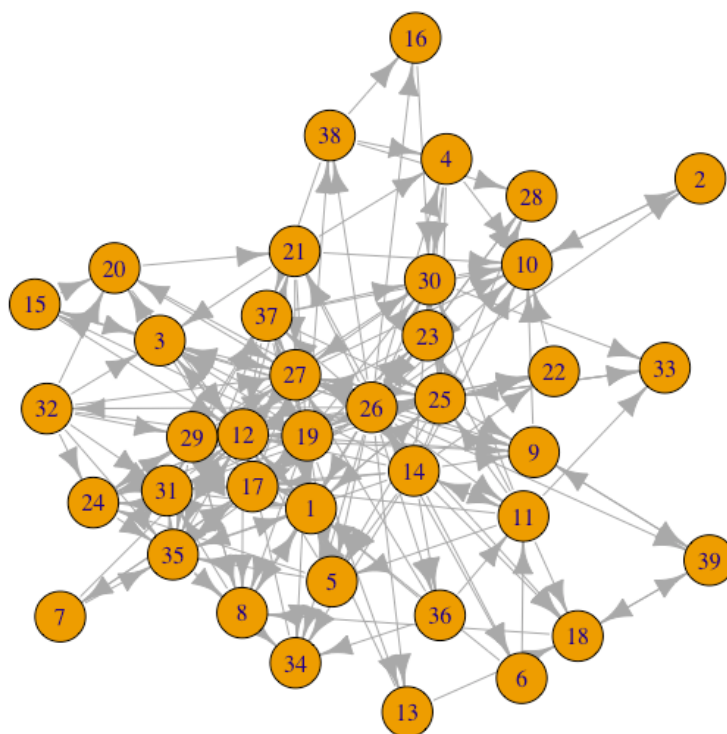
Výsledkom bolo teda 5 štvorcových matíc rozmeru 39×39 , ktorých prvky boli 1 a 0. Na základe týchto údajov sme vytvorili ku každej otázke zodpovedajúcu sieť, vid' obrázky 61, 62, 63, 64 a 65.



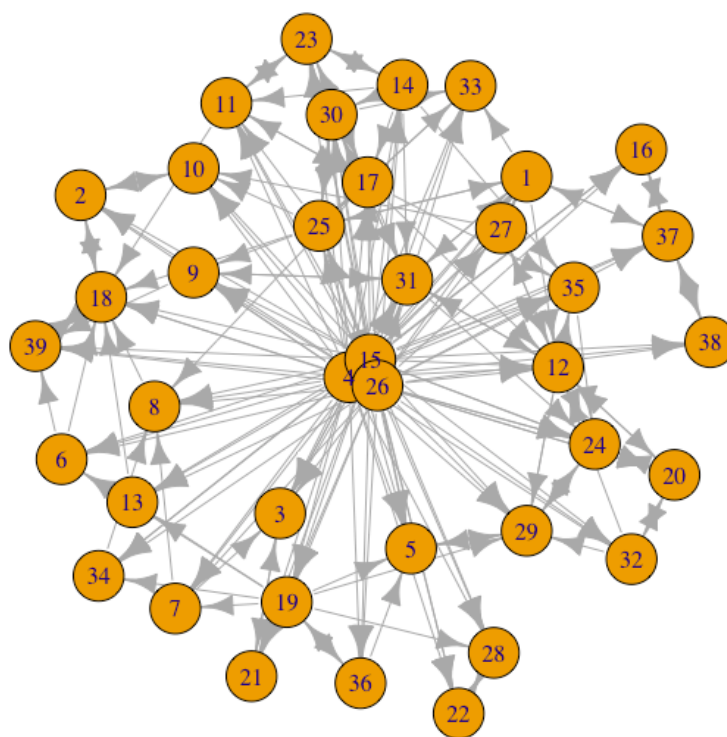
Obr. 61: Sieť popisujúca vzťahy podľa otázky č.1



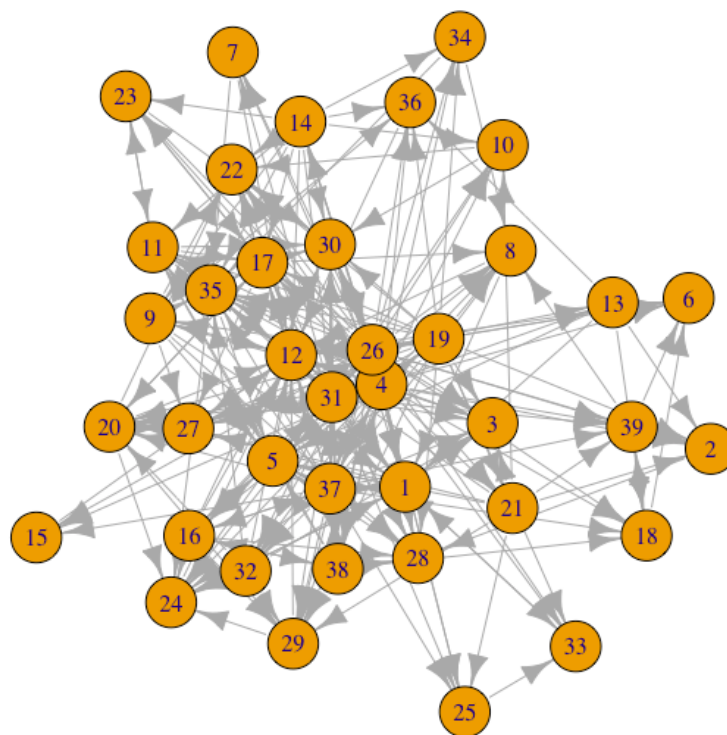
Obr. 62: Sieť popisujúca vzťahy podľa otázky č.2



Obr. 63: Sieť popisujúca vzťahy podľa otázky č.3



Obr. 64: Sieť popisujúca vzťahy podľa otázky č.4



Obr. 65: Sieť popisujúca vzťahy podľa otázky č.5

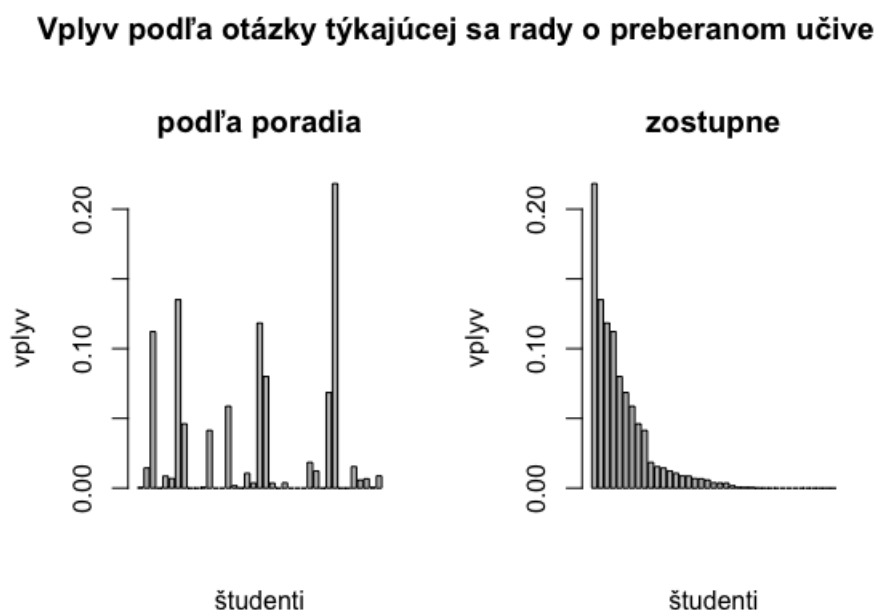
7.3 Nekonvergentné prípady

Pri otázke č.3: Koho by si oslovil, keď by si potreboval poradiť nejaký dobrý podnik? a otázke č.4: S kým by si sa riešil svoje osobné problémy? matice susednosti nekonvergovali. Dôvodom, prečo nedošlo ku konvergencii je, že v oboch prípadoch sa našli jedinci, ktorí by sa poradili iba medzi sebou. V otázke č.3 to bola dvojica spolužiakov s prideleným číslom 2 a 10, v otázke č.4 dvojica s číslami 22 a 28.

7.4 Konvergentné prípady

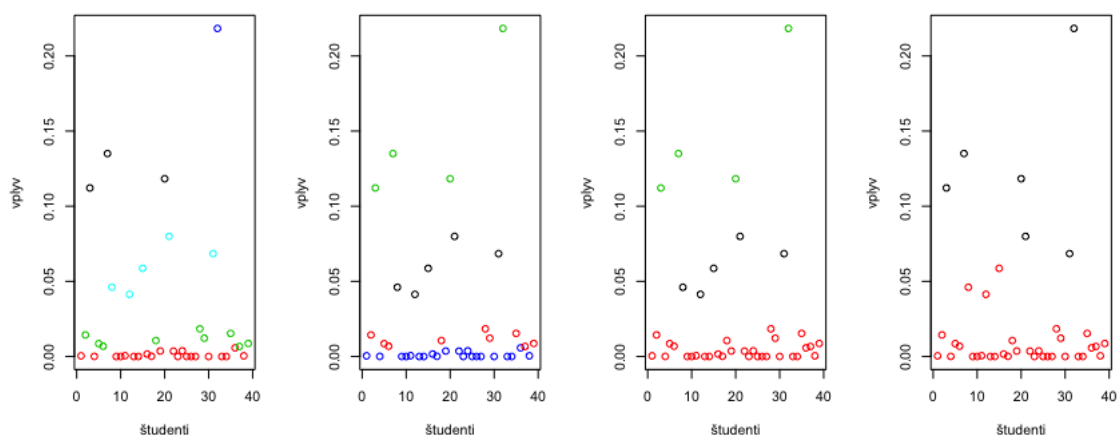
7.4.1 Otázka č.1: Za kým by si išiel, keby si potreboval poradiť s preberaným učivom?

Na obrázku 66 je zobrazený vplyv spolužiakov týkajúci sa rady ohľadom preberaného učiva.



Obr. 66: Vplyv podľa otázky č.1

Ďalej na obrázku 67 vidíme, ako by sme spolužiakov mohli zaradiť do skupín podľa vplyvu.

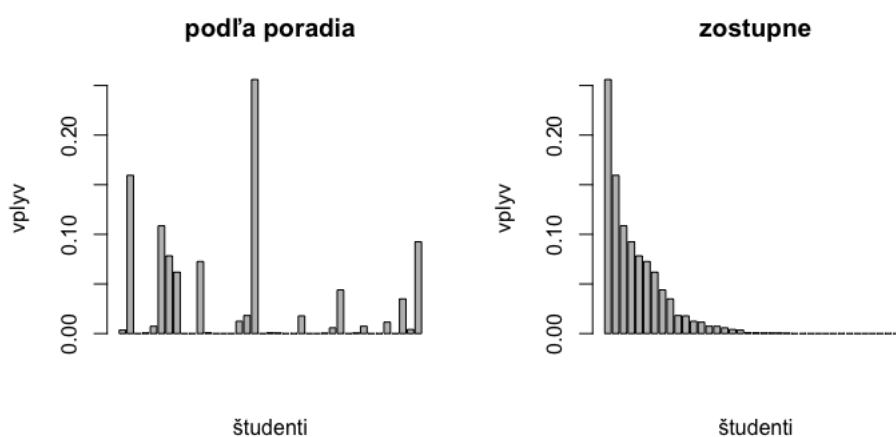


Obr. 67: Zadelenie spolužiakov do skupín podľa vplyvu

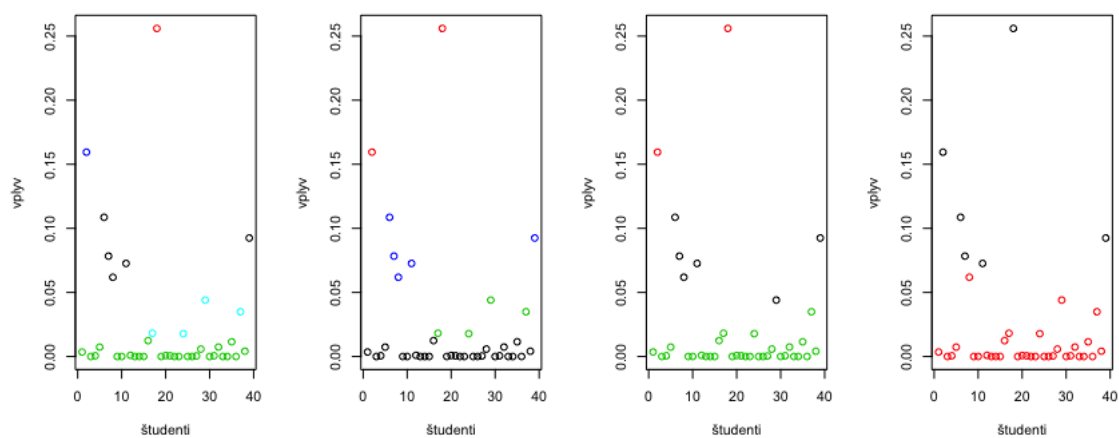
7.4.2 Otázka č.2: Od koho by si si išiel požičať poznámky z prednášky/-cvičenia, na ktorom si nebol?

Podobne ako pri v predchádzajúcej časti, aj pri otázke ohľadom požičania poznámok vidíme z obrázkov 68 a 69, že spolužiakov môžeme podľa vplyvu v sieti začleniť do viacerých skupín, pričom jeden je výrazne odčlenený od zvyšku.

Vplyv podľa otázky týkajúcej sa požičania poznámok z prednášky/cvičenia



Obr. 68: Vplyv podľa otázky č.2



Obr. 69: Zadelenie spolužiakov do skupín podľa vplyvu

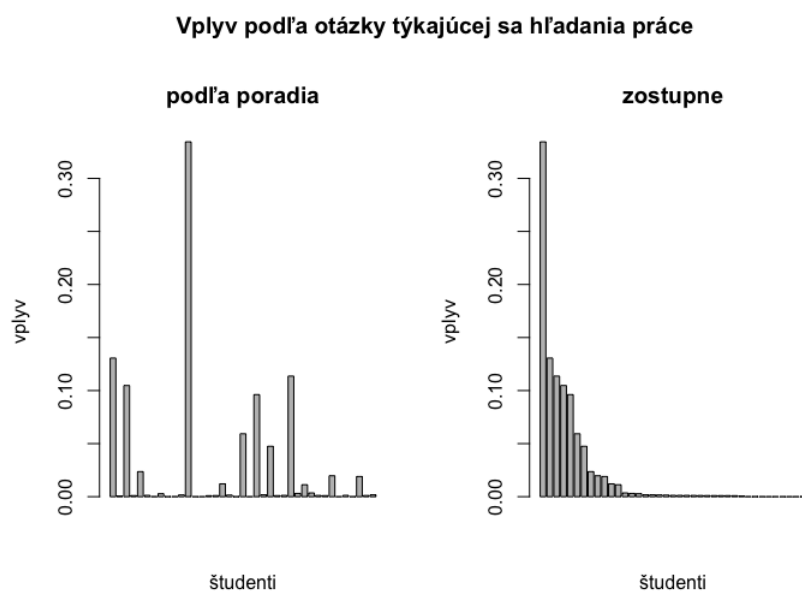
Za zamyslenie stojí, či spolužiaci, za ktorými by sa ostatní išli poradiť ohľadom učiva, sú rovnako vplyvní aj pri požičaní poznámok. Podľa obrázku 70 však vidíme, že platí opak, korelácia medzi týmito vplyvmi je záporná. Teda spolužiaci, za ktorými by išli ostatní po radu ohľadom učiva nie sú tí, od ktorých by si požičiavali poznámky, práve naopak.



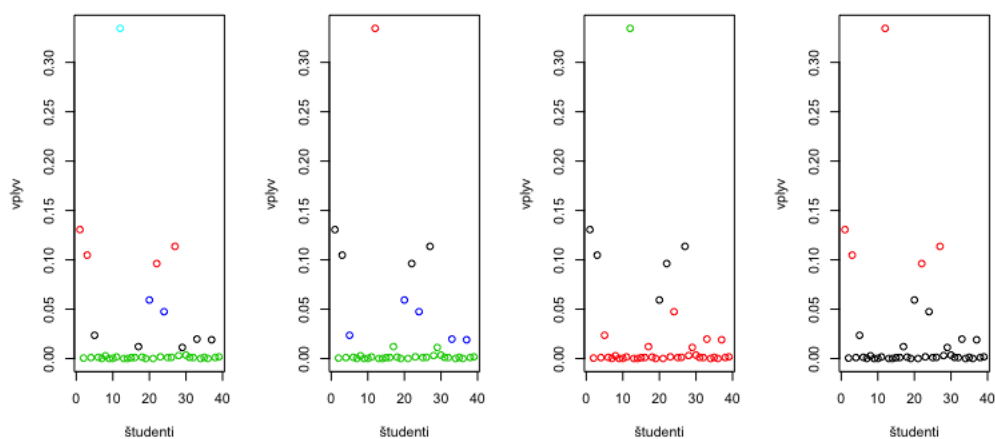
Obr. 70: Závislosť vplyvov medzi otázkami č.1 a č.2

7.4.3 Otázka č.5: Koho by si kontaktoval v prípade, že zháňaš prácu?

Pri otázke, kto by bol kontaktovaný kvôli pracovným príležitostiam, je viditeľne odlišný od zvyšku práve jeden spolužiak. Tohto najvplyvnejšieho spolužiaka vidíme jasne odlúčeného od ostatných na obrázku 71. Pri zoskupovaní spolužiakov do rovnako vplyvných skupín vidíme na obrázku 72, že jedna zo vzniknutých skupín je tvorená len touto jednou osobou v troch zo štyroch prípadov.



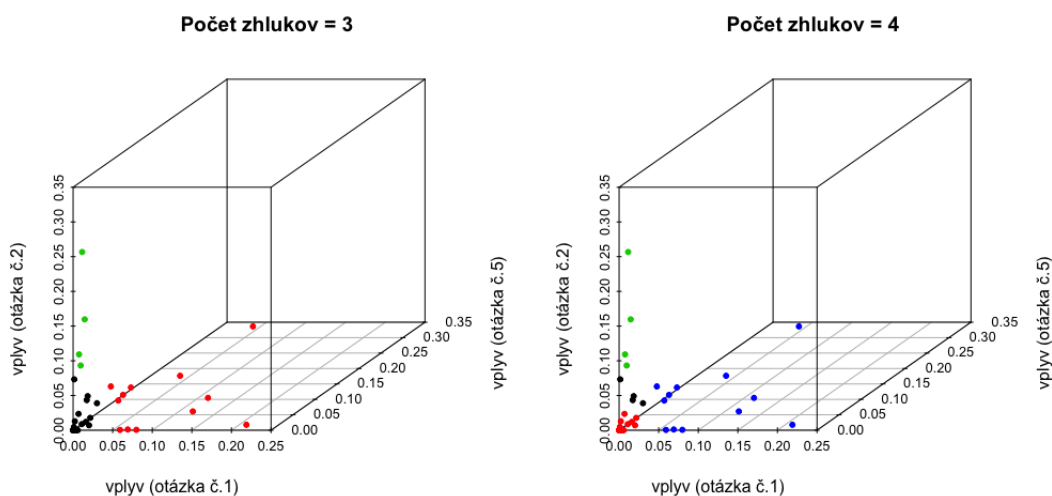
Obr. 71: Vplyv podľa otázky č.5



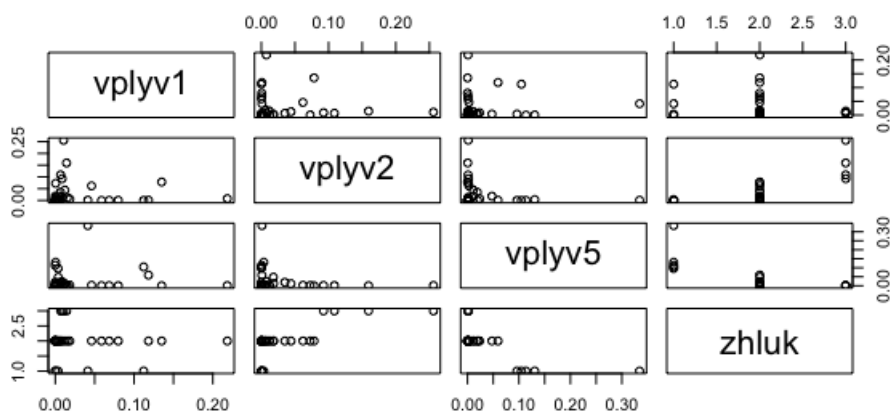
Obr. 72: Zadenie spolužiakov do skupín podľa vplyvu

7.4.4 Porovnanie vplyvov z otázok č.1, č.2 a č.5

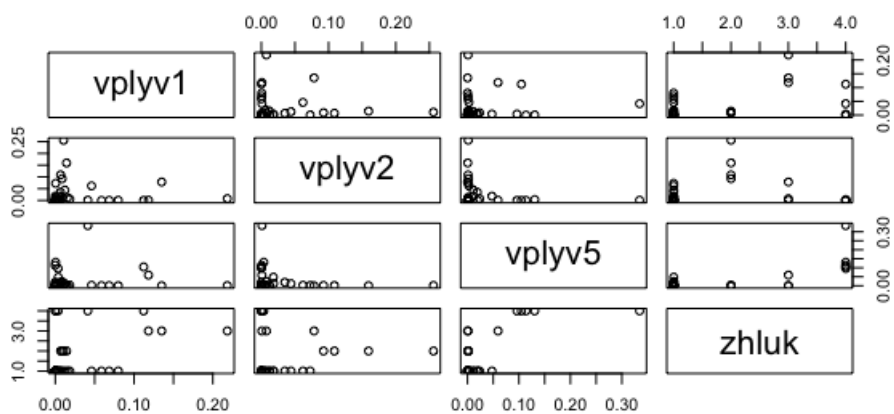
Zlúčením všetkých troch vektorov vplyvov získaných z otázok č.1, 2 a 5 sa pozrieme, aké skupiny sa vytvoria na základe zistených vplyvov jednotlivcov. Zvolili sme na porovnanie rozčlenenie do 3 a 4 zhlukov, ako je zobrazené na obrázku 73. Vzájomnú závislosť medzi jednotlivými vplyvmi vidíme aj spolu s rozdelením do 3 zhlukov na obrázku 74, podobne pre zadelenie do 4 zhlukov viď obrázok 75.



Obr. 73: Zadenie spolužiakov do skupín podľa vplyvu pri zlúčení 3 otázok



Obr. 74: Vzájomná závislosť medzi vplyvmi z otázok č.1, č.2 a č.5 a rozdelenie do 3 zhlukov



Obr. 75: Vzájomná závislosť medzi vplyvmi z otázok č.1, č.2 a č.5 a rozdelenie do 4 zhlukov

Záver

V tejto práci sme čitateľovi sme podrobnejšie popísali a vysvetlili základné pojmy a načrtli sme problematiku sietí. Vysvetlili sme modelovanie vplyvu v sieti pomocou rôznych mier centralít a tým aj dôležitosť jednotlivých členov siete. Predstavili sme DeGrootov model, ktorý popisuje dynamiku vytvárania názorov a meranie sociálneho vplyvu. Tento model sme následne aplikovali na nami získané dáta z dotazníkov na dvoch sociálnych skupinách. Pokračovali sme popísaním dôležitosti jednotlivých členov v týchto sieťach, čo sme názorne doplnili grafmi. Prínosom práce bolo zrozumiteľné vysvetlenie modelovania sietí pomocou príkladov a obrázkov, ktoré čitateľovi poskytnú jasný prehľad o danej problematike spolu s analýzou konkrétnych sietí a vplyvu kritéria na existenciu hrany a tiež zaradovanie ľudí do skupín na základe ich vplyvu.

Zoznam použitej literatúry

- [1] Bernard, H.R., Killworth, P., Kronenfeld, D., Sailer, L.: *The problem of informant accuracy: The validity of retrospective data*, Annual Review of Anthropology 13 (1984), 495-517
- [2] Bonacich, P.: *Power and Centrality: A Family of Measures*, The American Journal of Sociology 92 (1987), 1170-1182
- [3] DeGroot, M.H.: *Reaching a Consensus*, Journal of the American Statistical Association 69 (1974), 118-121
- [4] Elliott, M., Golub, B., Jackson M. O.: *Financial Networks and Contagion*, American Economic Review 104 (2014), 3115-3153
- [5] Jackson, M. O.: *Social and economic networks*, Princeton University Press, Princeton, 2008
- [6] Kitchens, B.P.: *Symbolic Dynamics: One-sided, Two-sided and Countable State Markov Shifts*, Springer Science & Business Media, 2012
- [7] Krackhardt, D.: *Cognitive Social Structures*, Social Networks 9 (1987), 109-134
- [8] Matoušek, J., Nešetřil, J.: *Kapitoly z diskrétní matematiky*, Karolinum, Praha, 2000
- [9] Myers, C. A., Shultz G. P.: *The Dynamics of a Labor Market*, New York: Prentice-Hall, 1951
- [10] Padgett, J. F., Ansell, C. K.: *Robust Action and the Rise of the Medici*, The American Journal of Sociology 98 (1993), 1400-1434
- [11] Plesník, J.: *Grafové algoritmy*, VEDA, vydavateľstvo Slovenskej akadémie vied, Bratislava, 1983
- [12] Saloff-Coste, L.: *Lectures on finite Markov chains (Lectures on probability theory and statistics)*, Springer Berlin Heidelberg, 1997, 309-313

- [13] Social and Economic Networks: Models and Analysis, Matthew O. Jackson, online kurz, dostupné na internete (16.6.2015):
<https://www.coursera.org/course/networksonline?action=enroll&sessionId=975770>
- [14] UCINET IV Datasets, dáta k výskumu Davida Krackhardta, dostupné na internete (20.3.2016):
<http://vlado.fmf.uni-lj.si/pub/networks/data/ucinet/ucidata.htm#krackoff>