

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY



CENTRALITA VRCHOLOV V SOCIÁLNYCH SIEŤACH A
JEJ SUBJEKTÍVNE VNÍMANIE

DIPLOMOVÁ PRÁCA

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

**CENTRALITA VRCHOLOV V SOCIÁLNYCH SIEŤACH A
JEJ SUBJEKTÍVNE VNÍMANIE**

DIPLOMOVÁ PRÁCA

Študijný program: Ekonomicko - finančná matematika a modelovanie
Študijný odbor: 9.1.9. 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. Adriána Kromková
Študijný program: ekonomicko-finančná matematika a modelovanie
(Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: matematika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Centralita vrcholov v sociálnych sieťach a jej subjektívne vnímanie
Centrality od nodes in social networks and its subjective perception

Anotácia: Sociálna sieť obsahuje vrcholy (napríklad ľudia alebo štáty) a hrany (určité väzby, napr. 'A a B sú priatelia na Facebooku). Meranie centrality vrcholov - ich určitej dôležitosti - bolo témou bakalárskej práce Zuzany Ondrejákovskej, ktorú obhájila v roku 2017. Pracuje sa v nej s niekoľkými mierami centrality aj s AHP metódou, ktorá ich kombinuje s určitými váhami. Pri pohľade na grafické znázornenie sociálnej siete človek aj intuitívne vidí, ktoré vrcholy sú viac a ktoré menej centrálné. V diplomovej práci sa zisťuje, či toto subjektívne intuitívne vnímanie zodpovedá nejakej konkrétnej miere centrality známej z teórie alebo či sa dá vysvetliť AHP metódou - so štandardnými Saatyho váhami alebo inými. Obsahuje výsledky zrealizovaného experimentu vyhodnoteného vhodnými štatistickými metódami.

Vedúci: doc. RNDr. Mgr. Beáta Stehlíková, PhD.
Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry: prof. RNDr. Marek Fila, DrSc.
Dátum zadania: 24.01.2018

Dátum schválenia: 25.01.2018
prof. RNDr. Daniel Ševčovič, DrSc.
garant študijného programu

.....
študent

.....
vedúci práce

Podakovanie Touto cestou by som sa rada podakovala svojej vedúcej diplomovej práce doc. RNDr. Beáte Stehlíkovej, PhD. za jej odbornú pomoc, cenné rady a pripomienky a za množstvo času, ktorý mi bola ochotná venovať pri konzultovaní práce. Taktiež chcem poďakovať svojej rodine, najmä svojmu manželovi, za veľkú dávku trpezlivosti, pochopenia a podpory, ktorú mi pri písaní práce prejavili.

Abstrakt

KROMKOVÁ, Adriána: Centralita vrcholov v sociálnych sieťach a jej subjektívne vnímanie. [Diplomová práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; školiteľ: doc. RNDr. Beáta Stehlíková, PhD., Bratislava, 2020, 80 s.

V našej práci sa zaoberáme rôznymi mierami centrality sociálnych sietí aplikovaných na vážené sociálne siete. Cieľom práce je vybrať z už definovaných mier centrality takú, ktorá by sa najlepšie zhodovala s intuitívnym vnímaním centrality. Vychádzame z článku [24], v ktorom je navrhnutý nový pohľad na centralitu vážených sociálnych sietí.

Prácu môžeme rozdeliť na teoretickú a praktickú časť. V teoretickej časti uvádzame základné pojmy a definície z teórie grafov, ktorá úzko súvisí s analýzou sociálnych sietí. Sústredíme sa pri tom najmä na také pojmy, ktoré sú kľúčové pre našu prácu. V praktickej časti práce robíme jednoduchšie vizualizácie váženej sociálnej siete a pozeráme sa na to, ktoré uzly siete sú Freemanovými mierami centrality vyhodnotené ako najviac centrálné. Ďalej v praktickej časti spracovávame výsledky z našich experimentov a porovnávame ich s odpoveďami z dotazníka, ktorý sme vytvorili pre potreby zistenia subjektívnych názorov na centrálnosť vrcholov.

Kľúčové slová: Sociálne siete, Analýza sociálnych sietí, Centralita vrcholov, Vizualizácia, Subjektívne vnímanie centrality

Abstract

KROMKOVÁ, Adriána: Centrality of nodes in social networks and its subjective perception. [Diploma Thesis], Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; Supervisor: doc. RNDr. Beáta Stehlíková, PhD., Bratislava, 2020, 80 s.

In our work we deal with different network measures applied to weighted social networks. The aim of the work is to choose from the already defined measures of centrality one, that would best coincide with the intuitive perception of centrality. We start from the article [24], which proposes a new view of the centrality of weighted social networks.

The work can be divided into theoretical and practical part. In the theoretical part we present the basic concepts and definitions of graph theory, which is closely related to the social network analysis. We focus mainly on such concepts that are key to our work. In the practical part of the work, we make simple visualizations of the weighted social network and look at which nodes of the network are evaluated as the most central by Freeman's measures of centrality. Furthermore, in the practical part, we process the results of our experiments and compare them with the answers from the questionnaire, which we created for the purpose of finding out subjective opinions on the centrality of nodes.

Keywords: Social networks, Social network analysis, Node centrality, Visualization, Subjective perception of centrality

Obsah

Úvod	9
1 Čo sú sociálne siete?	11
1.1 Základné definície a pojmy	13
1.2 Centralita vrcholov v sociálnych sieťach	16
1.2.1 Centralita stupňa	18
1.2.2 Centralita blízkosti	19
1.2.3 Centralita stredovej medzipolohy	19
1.2.4 Ilustračný príklad	20
2 Vážené sociálne siete	23
2.1 Centralita vrcholov vo vážených sieťach	23
2.1.1 Centralita stupňa pre vážené siete	24
2.1.2 Centralita blízkosti a stredovej medzipolohy pre vážené siete . .	24
2.2 Váhy vs. počet väzieb	27
2.2.1 Váhy a počet väzieb pri centralite stupňa	28
2.2.2 Váhy a počet väzieb pri centralite blízkosti a centralite stredovej medzipolohy	33
3 Aplikácia Freemanových a zovšeobecnených mier centrality	35
3.1 Popis a vizualizácia dát	35
3.2 Aplikácia Freemanových mier centrality na váženú sociálnu sieť	38
3.3 Aplikácia zovšeobecnenej centrality stupňa na sociálnu sieť	40
4 Výsledky dotazníka	46
4.1 Metódy porovnania poradí	46
4.1.1 Jarova - Winklerova podobnosť	47
4.1.2 Levenshteinova vzdialenosť	49
4.1.3 Spearmanov korelačný koeficient	51
4.1.4 Kendallov korelačný koeficient	52
4.2 Vnímanie dôležitosti postáv na základe filmu	53
4.2.1 Výber 9 hlavných postáv	54

4.2.2	Priradenie poradia postavám	59
4.3	Vnímanie centrálnosti uzlov v zobrazenej sieti	66
4.3.1	Vyhodnotenie metódou sčítania poradí	68
4.3.2	Vyhodnotenie metódami porovnávania poradí	69
4.3.3	Spojenie metód sčítania a porovnávania poradí	70
4.3.4	Porovnanie troch najlepších centralít	72
Záver		76
Zoznam použitej literatúry		78

Úvod

Žijeme vo svete, kde online sociálne siete ako Facebook či Twitter sú neodmysliteľnou súčasťou spoločenského života miliónov ľudí. Výrazne uľahčujú komunikačné procesy, napomáhajú k zveľaďovaniu vzťahov a umožňujú jednoduchšie nadväzovanie nových konaktov. Napriek tomu, tieto prostriedky komunikácie sú len jedným príkladom toho, čo sa vo všeobecnosti chápe ako sociálna sieť. Súčasťou nejakej sociálnej siete je každý z nás, či už ide o rodinu, spoločenstvo, do ktorého patríme alebo partiu kamarátov, v ktorej sa pohybujeme. Spoločnou charakteristikou všetkých týchto sietí je, že sú tvorené skupinou aktérov, ktorí vytvárajú medzi sebou systém vzťahov. Vzťahy pritom môžu mať k sebe pridruženú určitú prirodzenú pevnosť vychádzajúcu z ich trvania alebo intenzity.

Záujem o analýzu sociálnych sietí výrazne vzrástol v posledných desaťročiach, čoho dôkazom je aj vznik viacerých inštitúcií a časopisov zameraných na sociálne siete. Bolo publikovaných množstvo štúdií a vedeckých prác, ktoré s využitím súboru metód a algoritmov na analyzovanie vzťahov hľadajú vysvetlenia rôznych sociálnych javov, pričom sa snažia v maximálnej miere využiť poznatky o väzbách. Cieľom týchto analýz zvyčajne býva objasňovať dynamické procesy vo vnútri siete a identifikovať centrálnych jedincov, ktorí bývajú často spájaní s mocou a vplyvom v sieti. Takéto poznatky sa následne dajú využiť napríklad pri kontrole šírenia epidémie v sieti, maximalizovaní rýchlosti rozšírenia správy do celej siete alebo na účely znefunkčnenia siete.

Pri určovaní centrálnych jedincov siete zvykne byť prvým krokom výpočet niektorej z veľkého množstva existujúcich mier centrality, pričom výber vhodnej centrality závisí od typu sociálnej siete. Často sa však už pri grafickom znázornení siete intuitívne dajú určiť niektoré uzly, ktoré majú vo vnútri siete dôležité postavenie. V našej práci sa teda pokúsime nájsť mieru centrality, ktorá najlepšie zodpovedá subjektívnemu názoru na dôležitosť uzlov v sieti. Zameriame sa pritom na tri najznámejšie miery centrality - centralitu stupňa, centralitu blízkosti a centralitu stredovej medzipolohy - a ich zovšeobecnenia pomocou dodatočného parametra. Zároveň nás bude zaujímať aj to, do akej miery pevnosť väzieb v sieti ovplyvňuje vnímanie centrálnosti uzlov.

Práca je rozdelená na dve hlavné časti - teoretickú a praktickú časť. Celkovo obsahuje štyri kapitoly, z ktorých prvé dve sú venované základným pojmom a definíciám z

teórie sociálnych sietí. V prvej kapitole vysvetľujeme čo sú sociálne siete, ako ich definujeme a aké typy sociálnych sietí poznáme. Taktiež v nej zavádzame pojmy ako cesta a najkratšia cesta medzi vrcholmi siete a uvádzame tri najznámejšie miery centrality, ktoré sa používajú pri identifikácii centrálnych vrcholov. V druhej kapitole sa zaoberáme zovšeobecnením pojmov a definícií aj na vážené sociálne siete, čiže také, v ktorých sú väzby od seba odlišené. Posledné dve kapitoly sú už praktické. V tretej kapitole sa aplikujú poznatky z teórie sociálnych sietí na váženú sociálnu sieť vytvorenú z postáv z filmu *Hviezdne vojny: Sila sa prebúdzá*, v ktorej sú vzťahy definované podľa toho, či a koľkokrát sa postavy objavujú spolu na scéne. Posledná štvrtá kapitola tvorí hlavnú časť práce. Vyhodnocujeme v nej odpovede z dotazníka, v ktorom sme zisťovali, ako ľudia intuitívne vnímajú centralitu vrcholov v sieti. Uzly, ktoré sa určili výpočtami pri viacerých mierach centrality ako najviac centrálné sa porovnávajú s tými, ktoré označili ľudia po predložení grafického znázornení siete ako najviac centrálné. Nakoniec sa vyhodnocuje, pri ktorej centralite nastala najväčšia zhoda s názormi z odpovedí.

1 Čo sú sociálne siete?

V dnešnej dobe sa pojem sociálnej siete zvyčajne spája s masívnym využívaním webových služieb fungujúcich pomocou webovej stránky ako sú Facebook či Twitter, ktoré slúžia na nadväzovanie priateľstiev, zdieľanie zážitkov či udržiavanie kontaktov medzi ľuďmi. To vôbec nie je prekvapivé, nakoľko sme svedkami čoraz väčšieho vplyvu týchto webových prostriedkov na každodenný život a komunikáciu miliónov užívateľov. Napriek tomu, keď vo vedeckom prostredí hovoríme o sociálnych sieťach, máme tým na mysli všeobecnejší koncept sociálnej siete. O sociálnej sieti uvažujeme ako o sociálnom konštrukte tvorenom množinou subjektov (ako sú organizácie, krajiny alebo osoby) a množinou vzťahov (napr. obchodné vzťahy, priateľstvá či rodinné väzby), ktorými sú prepojené, čím vytvárajú celkovú sieť vzťahov. Dané subjekty sú teda uzly siete a hrany medzi uzlami predstavujú vzťahy medzi subjektami. Za sociálnu sieť sa potom považujú sociálne štruktúry, v ktorých sa vytvárajú komunikačné procesy či procesy prenosu (napr. informácií alebo choroby).

V súvislosti s takto chápanými sociálnymi sieťami uvádzame na ukážku niekoľko príkladov z odbornej literatúry, v ktorých môžeme analýzy sociálnych sietí nájsť. Staršou, no z hľadiska štúdia sociálnych sietí pomerne dôležitou prácou je článok *An Information Flow Model for Conflict and Fission in Small Groups* [28], v ktorom autor, Wayne W. Zachary, predstavil koncept štruktúry spoločenstva. V článku analyzoval sociálnu sieť univerzitného karate klubu, ktorý sa počas sledovaného obdobia troch rokov kvôli nezhodám vo vedení klubu rozdelil na dva kluby. To spôsobilo, že aj členovia pôvodného klubu sa museli rozdeliť na dva tábory. Zachary si všimol, že hoci v rámci každej skupiny sa členovia kontaktujú často aj mimo klubového prostredia, medziskupinové kontakty sú veľmi zriedkavé. Na základe toho navrhol model, ktorý opisuje štiepenie sociálnej siete na podskupiny. Vrcholy v jednotlivých podskupinách sú medzi sebou husto prepojené, avšak medzi skupinami sú prepojenia len veľmi slabé. Vďaka tomu, že je známe ako sa klub rozdelil, Zacharyho dáta o kontaktoch medzi členmi sú doteraz používané na testovanie nových algoritmov zhľukovania vrcholov sietí.

V článku *Social Network Analysis of Study Environment* [9] autorky Blaženka Div-

jak a Petra Peharda skúmajú, či existuje korelácia medzi akademickými výsledkami študenta a jeho postavením v sociálnej sieti študentov na jednej z fakúlt Univerzity v Záhrebe. Práca rozoberá viacero hypotéz, ku ktorým bolo potrebné zostaviť aj rôzne sociálne siete. Uzly sietí znázorňovali študentov fakulty a hrany medzi uzlami vznikli podľa toho čo sieť predstavovala, čiže napríklad v sieti s výmenami študijných materiálov bola medzi dvoma uzlami hrana práve vtedy keď si títo študenti medzi sebou posúvali študijné materiály. Hoci bola analýza robená na relatívne malom počte študentov, výsledky sú pomerne zaujímavé. Hlavným zistením je, že z postavenia študenta v sieti sa síce nedajú predpovedať jeho úspechy v štúdiu ale úspešnejší študenti majú v sieti zvyčajne lepšie postavenie. Je viacero faktorov, ako napríklad zodpovedosť či komunikatívnosť študenta, ktoré vplývajú na jeho postavenie v sieti. Pri porovnaní dvoch sietí, jedna zostavená za základe tímovej spolupráce a druhá podľa toho ako si študenti vymieňali študijné materiály, si autorky všimli, že druhá sieť je značne hustejšia. Z toho sa dá dedukovať, že pokiaľ ide o výmenu materiálov, študenti medzi sebou viac komunikujú, avšak pri tímových spoluprákach, preferujú pracovať s tými istými študentami. Okrem týchto zistení, analýza sociálnej siete ukázala aj to, že externí študenti tvoria samostatnú skupinu, ktorá je s dennými študentami len slabo prepojená. To môže byť jeden z hlavných dôvodov, prečo majú externí študenti vo všeobecnosti slabšie študijné výsledky.

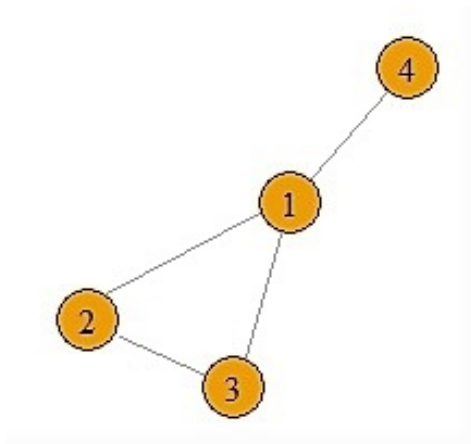
Za zmienku určite stojí aj článok *Detecting criminal organizations in mobile phone networks* [10]. V posledných desaťročiach sa ukazuje, že analýza sociálnych sietí môže byť cenným pomocníkom pri vyšetrovaní teroristických alebo kriminálnych organizácií. S jej využitím je možné určiť centrálnych členov siete, slabých členov, styky medzi jednotlivcami, prípadne vzťahy, ktorých odstránenie by viedlo k rozloženiu siete. V tomto kontexte je analýza hovorov rozhodujúca na získanie základných informácií o prepojeniach a komunikácii medzi kriminálnikmi. Autori v článku [10] predstavujú systém, ktorý umožní odhalenie kriminálnych spoločenstiev vďaka sociálnej sieti zostrojenej zo záznamov o telefónnych hovoroch. Opisujú, ako sa tento systém dá použiť pri skúmaní kriminálnej siete, čo umožňuje odhaliť niektoré jej primárne charakteristiky.

1.1 Základné definície a pojmy

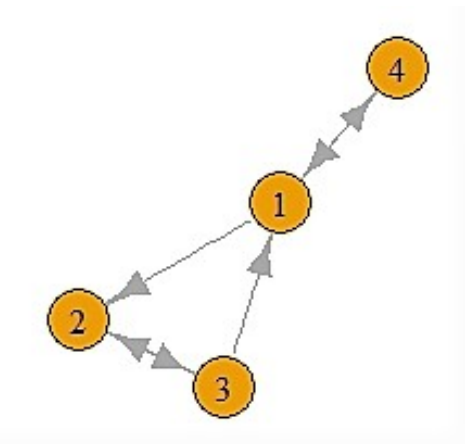
Existuje niekoľko spôsobov, ako formálne zadefinovať sociálnu sieť v závislosti od vedného odboru, ktorý sa štúdiom tejto siete zaoberá. Najčastejšie sa pri jej definícii využíva teória grafov, ktorá sa na sieť pozerá cez dve množiny: množinu vrcholov a množinu hrán reprezentujúcich väzby medzi vrcholmi.

Definícia 1.1. [15] Definujeme sociálnu sieť $G = (N, E)$ jednoznačne určenú množinou vrcholov $N = \{n_1, \dots, n_p\}$ a množinou hrán $E = \{e_1, \dots, e_r\}$, a jej maticu susednosti A typu $p \times p$. Prvky matice $a_{ij} \in \{0, 1\}$, kde $a_{ij} = 1$ ak existuje hrana, ktorá spája vrchol n_i s vrcholom n_j , a $a_{ij} = 0$ ak taká hrana neexistuje.

Definícia 1.2. [15] Hovoríme, že sociálna sieť $G = (N, E)$ je neorientovaná, ak pre všetky dvojice vrcholov (n_i, n_j) platí $a_{ij} = a_{ji}$. Ak existuje aspoň jedna dvojica vrcholov (n_i, n_j) , pre ktorú $a_{ij} \neq a_{ji}$, hovoríme o orientovanej sociálnej sieti.



Obr. 1: Neorientovaná sociálna sieť.



Obr. 2: Orientovaná sociálna sieť

Príklad 1.1. Pre ilustráciu pojmu matice susednosti a lepšie pochopenie konceptov neorientovanej a orientovanej sociálnej siete uvádzame na Obr. 1 resp. Obr. 2 jednoduché príklady takýchto sietí. Obidve siete sú tvorené rovnakou množinou vrcholov $N = \{n_1, \dots, n_4\}$. Nech teda $n_1 = 1$, $n_2 = 2$, $n_3 = 3$ a $n_4 = 4$. Matice susednosti pre

tieto siete potom vyzerajú nasledovne

$$A_N = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad A_O = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix},$$

kde A_N je matica susednosti neorientovanej siete z Obr. 1 a A_O je matica susednosti orientovanej siete z Obr. 2. Vidíme, že tak pre orientovanú ako aj pre neorientovanú sieť majú tieto matice na diagonále nuly, avšak zatiaľ čo A_N je symetrická, A_O symetrická nie je.

Definícia 1.3. [15] *Stupeň vrcholu n_i v neorientovanej sieti $G = (N, E)$ je definovaný ako počet priamych väzieb vrcholu n_i s ostatnými vrcholmi siete. Matematicky zapísané*

$$\deg(n_i) = \sum_{j=1}^p a_{ij} = \sum_{j=1}^p a_{ji},$$

kde a_{ij} sú prvky matice susednosti A , ktorá prislúcha k sieti G a $p = |N|$ je celkový počet uzlov v sieti G .

V orientovanej sociálnej sieti rozlišujeme dva typy stupňa:

1. vstupný stupeň vrcholu n_i (In-degree) reprezentujúci počet hrán vchádzajúcich do vrcholu n_i , ktorý môžeme formálne zapísať

$$\text{Ideg}(n_i) = \sum_{j=1}^p a_{ji},$$

2. výstupný stupeň vrcholu n_i (Out-degree) reprezentujúci počet hrán vychádzajúcich z vrcholu n_i , ktorý môžeme matematicky zapísať ako

$$\text{Odeg}(n_i) = \sum_{j=1}^p a_{ij}.$$

Definícia 1.4. [15] *Prechádzkou z vrcholu v_0 do vrcholu v_k nazveme postupnosť hrán $L = \{l_1, l_2, \dots, l_k\}$, pre ktorú existuje postupnosť vrcholov $V = \{v_0, v_1, \dots, v_k\}$, kde l_i je hrana spájajúca vrcholy v_{i-1} a v_i pre $i = 1, \dots, k$. Prechádzka, pri ktorej sa v postupnosti V vrcholy v_i neopakujú sa nazýva cestou medzi vrcholmi v_0 a v_k .*

Definícia 1.5. [15] Najkratšou cestou medzi vrcholmi n_i a n_j v sieti $G = (N, E)$ nazveme takú, ktorej dĺžka meraná počtom hrán medzi týmito vrcholmi je najmenšia.

Poznámka: Uzly v_i pre $i = 1, \dots, k - 1$ z definície 1.4 budeme v tejto práci nazývať *medziuzly* alebo *prostredné uzly*.

Príklad 1.2. Uvažujme opäť siete z Obr. 1 resp. Obr. 2. Obidve siete obsahujú štyri vrcholy $n_1 = 1$, $n_2 = 2$, $n_3 = 3$ a $n_4 = 4$. Formulky na výpočet stupňa vrcholu aplikujeme tak v orientovanej ako aj neorientovanej sieti iba na vrchol s číslom 1, čo by pre pochopenie pojmov malo byť postačujúce. Keďže k neorientovanej sieti z Obr. 1 prislúcha matica susednosti A_N rozmeru 4×4 , stupeň vrcholu n_1 (t.j. vrcholu označeného číslom 1) tejto siete vypočítame s využitím matice A_N nasledovne

$$\deg(n_1) = \sum_{j=1}^4 a_{1j} = 0 + 1 + 1 + 1 = 3.$$

Takýto výsledok je ľahko pozorovateľný už z Obr. 1, na ktorom vidíme, že vrchol označený číslom 1 je hranami spojený so všetkými troma zostávajúcimi vrcholmi siete. Pre orientovanú sieť z Obr. 2 vypočítame vstupný (In-degree) aj výstupný (Out-degree) stupeň vrcholu s číslom 1 s využitím matice susednosti A_O jednoducho so vzorcov

$$Ideg(n_1) = \sum_{j=1}^4 a_{j1} = 0 + 0 + 1 + 1 = 2,$$

$$Odeg(n_1) = \sum_{j=1}^4 a_{1j} = 0 + 1 + 0 + 1 = 2.$$

Opäť vidíme, že výsledky sa zhodujú s tým, čo pozorujeme aj na Obr. 2. Vstupný stupeň (In-degree) vrcholu n_1 je 2, nakoľko do tohoto vrcholu vstupujú dve hrany. Podobne výstupný stupeň (Out-degree) vrcholu n_1 je 2, keďže dve hrany z neho aj vychádzajú.

Teraz sa bližšie pozrieme na pojmy cesta a najkratšia cesta medzi dvoma vrcholmi. Znova použijeme príklady sociálnych sietí z Obr. 1 a Obr. 2. V obidvoch prípadoch budeme hľadať najkratšiu cestu z vrcholu s číslom 4 do vrcholu s číslom 3, čiže $v_0 = 4$ a $v_k = 3$. V neorientovanej sieti z Obr. 1 existujú dve cesty, teda dve postupnosti hrán a k nim prislúchajúce postupnosti vrcholov, ktoré majú začiatok vo vrchole $v_0 = 4$ a končia vo vrchole $v_k = 3$:

1.

$$L = \{l_1, l_2, l_3\} = \{v_0v_1, v_1v_2, v_2v_3\},$$

$$|L| = 3,$$

kde $v_0 = 4, v_1 = 1, v_2 = 2, v_3 = 3$,

2.

$$L = \{l_1, l_2\} = \{v_0v_1, v_1v_2\},$$

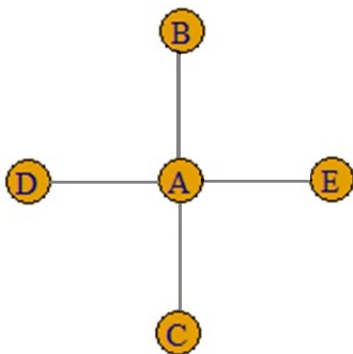
$$|L| = 2,$$

kde $v_0 = 4, v_1 = 1, v_2 = 3$.

Z toho už vieme určiť, že najkratšou cestou meranou počtom hrán medzi vrcholmi $v_0 = 4$ a $v_k = 3$ je druhá v poradí. Nakoľko táto cesta obsahuje iba jeden medziuzol, treba pri nej prejsť len po dvoch hranách, zatiaľ čo prvá cesta má medziuzly až dva a je potrebné pri nej prejsť až po troch hranách. Na druhej strane, v prípade orientovanej siete z Obr. 2 existuje iba jedna cesta začínajúca vo vrchole $v_0 = 4$ a končiaca vo vrchole $v_k = 3$, je ňou prvá cesta z vyššie spomenutých. Cesta, ktorá vyšla v neorientovanej sieti ako najkratšia, už v orientovanej sieti neexistuje. Aj napriek tomu, že z vrcholu s číslom 3 do vrcholu s číslom 1 existuje priama hrana, v opačnom smere to už neplatí, a teda vrchol 3 sa z vrcholu 1 dá dosiahnuť len cez vrchol s číslom 2. Keďže v tomto prípade existuje len jedna cesta, ktorá vychádza z vrcholu $v_0 = 4$ a končí vo vrchole $v_k = 3$, táto cesta je zároveň aj najkratšou cestou meranou počtom hrán.

1.2 Centralita vrcholov v sociálnych sieťach

Centralita vrcholov sociálnej siete, t. j. identifikácia vrcholov, ktoré sú v danej sieti centrálnnejšie, je jedným z hlavných nástrojov pri štúdiu organizačných štruktúr sociálnych sietí. Centrálna pozícia jedincov bola vždy spájaná s vodcovstvom, mocou, popularitou či dobrou reputáciou. Takíto ľudia potom spravidla majú možnosť napríklad ovplyvňovať tok informácií v rámci siete, vďaka čomu majú kľúčovú úlohu pri rozhodovacích procesoch.



Obr. 3: Sociálna sieť tvorená piatimi uzlami.

Myšlienka centrality bola známa už v štyridsiatych rokoch 20. storočia. Jej pôvodom bol Bavelas [3], ktorý predpokladal, že centralita priamo súvisí s vplyvom určitých osôb v skupine. Počas nasledujúcich rokov sa centralita stala predmetom záujmu viacerých odborníkov z rôznych oblastí vedy. Bolo navrhnutých niekoľko mier centrality, avšak väčšina z nich bola nejasná alebo zbytočne komplikovaná. Detailnejšie tento koncept spracoval až v roku 1978 L. C. Freeman vo svojom článku *Centrality in social networks: Conceptual clarification* [12]. Zameral sa v ňom na tri podľa neho najrozumnejšie prístupy k meraniu centrality, pričom na ich vysvetlenie použil príklad sociálnej siete pozostávajúcej z piatich uzlov ako je na Obr. 3. Tvrdil, že uzol, ktorý sa nachádza uprostred siete, má oproti ostatným tri výhody:

- má najväčší počet hrán, čiže najväčší počet priamych kontaktov s ostatnými uzlami,
- všetky ostatné vrcholy môže dosiahnuť rýchlejšie,
- môže kontrolovať tok medzi zvyšnými vrcholmi.

Na základe toho následne uviedol tri rôzne miery centrality vrcholov, v ktorých boli tieto vlastnosti zahrnuté. Boli to:

- centralita stupňa,
- centralita blízkosti,

- centralita stredovej medzipolohy.

1.2.1 Centralita stupňa

Základným a najjednoduchším indikátorom centrálnosti vrcholu v sociálnej sieti je centralita stupňa. Tá je založená na počte priamych väzieb vrcholu k ostatným vrcholom. Uzol s najväčšou centralitou stupňa, čiže najväčším počtom väzieb, má oproti ostatným uzlom siete výhodnejšiu pozíciu.

Aj napriek tomu, že bolo viacero pokusov nájsť vhodnú a výstižnú definíciu centrality stupňa, väčšina z týchto definícií bola nejasná, priveľmi komplikovaná alebo obmedzená len na určité druhy prípadov. Prvú jednoduchú a úplne všeobecnú mieru centrality stupňa navrhol až v roku 1974 Nieminen [22]. Tá sledovala koľko „susedov“ má východiskový vrchol a teda sa zhodovala s tým, čo sme definovali ako stupeň vrcholu:

$$C_d(n_i) = \deg(n_i).$$

Ako poznamenáva Freeman vo svojom článku [12], takáto miera centrality je čiastočne funkciou veľkosti siete, na ktorej sa počíta. Pre niektoré aplikácie je táto informácia relevantná, zatiaľčo pre iné by mohlo byť žiaduce mať takú mieru, ktorá je nezávislá od veľkosti siete. Aby bolo možné porovnať centrality stupňa pre rôzne siete, Freeman navrhol definovať centralitu stupňa vrcholu n_i v sieti $G = (N, E)$ ako

$$C'_d(n_i) = \frac{\deg(n_i)}{p - 1},$$

kde $\deg(n_i)$ je stupeň vrcholu n_i v sieti G a $p = |N|$ je celkový počet uzlov v sieti G . Vidíme, že platí $0 \leq C'_d(n_i) \leq 1$.

V orientovaných sieťach rozlišujeme dva typy centrality stupňa:

- In-degree centralita: založená na počte priamych väzieb vstupujúcich do vrcholu,
- Out-degree centralita: založená na počte priamych väzieb vystupujúcich z vrcholu.

Pre jednoduchosť výpočtu býva centralita stupňa zvyčajne prvým krokom pri analýze sociálnej siete. Zameriava sa však iba na lokálnu štruktúru siete okolo hlavného uzla

pričom ignoruje celkový charakter siete. Aby sa pri analýze sociálnych sietí zachytil aj tento aspekt, boli definované ďalšie miery centrality: centralita blízkosti a centralita stredovej medzipolohy.

1.2.2 Centralita blízkosti

Hlavnou motiváciou pre výpočet centrality blízkosti je identifikácia uzlov, ktoré môžu najjednoduchšie dosiahnuť všetky ostatné uzly, inak povedané, takých uzlov, ktoré majú najmenší súčet minimálnych vzdialeností k ostatným uzlom siete. Pod minimálnou vzdialenosťou sa rozumie najkratšia cesta medzi párami vrcholov, ktorou sme sa zaoberali v definíciách 1.4 a 1.5. V sociálnej sieti $G = (N, E)$ sa potom centralita blízkosti vrcholu n_i definuje ako obrátená hodnota tohoto súčtu, čo sa dá matematicky vyjadriť nasledovne

$$C_c(n_i) = \frac{1}{\sum_{j=1}^p l(i, j)},$$

kde $l(i, j)$ predstavuje dĺžku najkratšej cesty medzi uzlami n_i a n_j meranú počtom hrán a $p = |N|$ je celkový počet uzlov v sieti G . Podobne ako pri centralite stupňa, aj pri centralite blízkosti je často výhodnejšie nemať mieru závislú od veľkosti siete, preto sa zvykne štandardizovať a uvádza sa v tvare

$$C'_c(n_i) = \frac{p-1}{\sum_{j=1}^p l(i, j)}.$$

Tu si treba uvedomiť, že mieru centrality založenú na vzdialenostiach medzi uzlami má zmysel počítať iba v sieťach, ktoré neobsahujú izolované komponenty. Dva uzly, ktoré patria k rôznym komponentom, nemajú medzi sebou konečnú vzdialenosť. Takže centralita blízkosti je vo všeobecnosti obmedzená na uzly v rámci najväčšej zložky siete.

1.2.3 Centralita stredovej medzipolohy

Centralita stredovej medzipolohy sa zaoberá otázkou aký potenciálny vplyv má uzol na prípadné ďalšie rozšírenie alebo zadržanie informácie či choroby k zvyšným uzlom siete, inak povedané, ako dôležitý je uzol z hľadiska prepojenia siete. Určuje, koľko najkratších ciest medzi všetkými párami vrcholov prechádza daným vrcholom. Ak je uzol umiestnený tak, že jediná cesta medzi ostatnými uzlami je práve cez tento uzol, potom je v sieti veľmi dôležitý a jeho centralita by mala byť vysoká. Formálne môžeme

centralitu stredovej medzipolohy vrcholu n_i v sieti $G = (N, E)$ zapísať ako

$$C_b(n_i) = \sum_{k \neq j, i \notin \{k, j\}} \frac{\mathcal{P}_i(kj)}{\mathcal{P}(kj)},$$

kde $\mathcal{P}(kj)$ je celkový počet najkratších ciest z vrcholu n_k do vrcholu n_j a $\mathcal{P}_i(kj)$ je celkový počet najkratších ciest z vrcholu n_k do vrcholu n_j prechádzajúcich cez vrchol n_i . Niekedy sa uvádza aj v štandardizovanom tvare

$$C'_b(n_i) = \frac{2}{(p-1)(p-2)} \sum_{k \neq j, i \notin \{k, j\}} \frac{\mathcal{P}_i(kj)}{\mathcal{P}(kj)},$$

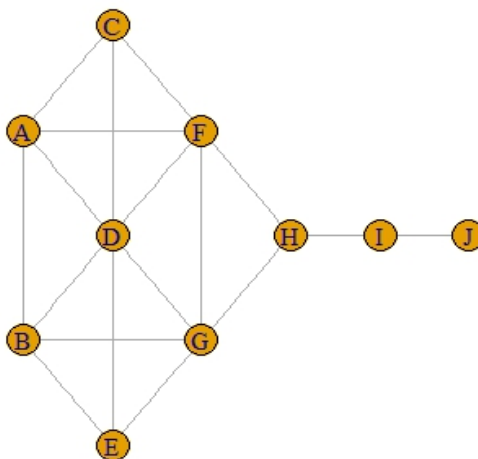
kde p predstavuje celkový počet uzlov v sociálnej sieti G . Aj tu musíme poznamenať, že nakoľko je centralita stredovej medzipolohy založená na najkratších vzdialenostiach medzi všetkými párami vrcholov siete, opäť platí, že je možné ju použiť len na prípady súvislých sietí.

1.2.4 Ilustračný príklad

Rozdiel medzi tromi najznámejšími mierami centrality najlepšie uvidíme na príklade sociálnej siete nazvanej *Kite Network*. Ide o sieť vymyslenú profesorom Davidom Krackhardtom, ktorý ju v práci [18] používa na ilustrovanie mier centrality z kapitol 1.2.1, 1.2.2, 1.2.3.

Sieť je tvorená 10 uzlami A - J a jej vizualizáciu vidíme na Obr. 4. Pod uzlami si môžeme predstaviť 10 ľudí pracujúcich v jednej firme a hrany medzi uzlami potom znázorňujú pracovné kontakty, inak povedané dva uzly sú spojené ak sú v pravidelnom osobnom kontakte alebo spolu komunikujú iným spôsobom. Tak napríklad uzol A pravidelne spolupracuje s uzlom C zatiaľ čo s uzlom I v kontakte nie je. Na základe vizualizácie siete je možné vidieť, že niektoré uzly majú v sieti silnejšie postavenie ako iné. To súvisí s ich centrálnosťou v danej sieti a teda s ich mierou centrality. Každý stĺpec v Tab. 1 obsahuje uzly A - J usporiadané na základe hodnoty centrality pri troch Freemanových mierach centrality. Hodnoty centrality pre konkrétne uzly sú uvedené v zátvorkách.

Všimnime si, že pri každej z uvedených mier centrality vyšiel ako najcentrálnejší iný vrchol. Centralita stupňa určila ako najcentrálnejší vrchol D nakoľko má najviac priamych väzieb s ostatnými vrcholmi. Osoba s najväčším počtom väzieb sa zvykne



Obr. 4: Krackhardtov príklad sociálnej siete.

považovať za najviac aktívnu či populárnu v sieti. Všeobecne vládne presvedčenie, že čím viac väzieb vrchol má, tým má výhodnejšie postavenie v sieti.

V prípade centrality blízkosti skončili najlepšie uzly F a G. Hoci majú obidva uzly menej priamych väzieb ako D, všetky ostatné vrcholy dosiahnu rýchlejšie ako ktorýkoľvek iný vrchol, inými slovami, súčet najkratších vzdialeností k všetkým ostatným vrcholom majú najmenší. Centralita blízkosti priamo súvisí s tým, ako rýchlo sa informácia dokáže rozšíriť z vychádzajúceho uzlu k všetkým ostatným uzlom, čiže z hľadiska rýchleho rozšírenia informácie do celej siete majú vrcholy F a G najvýhodnejšiu polohu. Platí to zároveň aj naopak, a to, že tieto dva vrcholy sa tiež vedú k informáciám veľmi rýchlo dostať.

Centralita stredovej medzipolohy má najvyššiu hodnotu pri vrchole H. Uzol H má len tri väzby ale strategicky veľmi silnú pozíciu, keďže predstavuje jediné spojenie medzi dvoma oddelenými zhlukmi uzlov. Na jednej strane takto výhodné umiestnenie dáva vrcholu H moc kontrolovať tok informácií, no na druhej strane má v prípade potreby zodpovednosť za to, aby sa informácie dostali aj k vrcholom I a J.

	Centralita stupňa	Centralita blízkosti	Centralita stredovej medzipolohy
1	D (0.6667)	F (0.6429)	H (0.3889)
2	F (0.5556)	G (0.6429)	F (0.2315)
3	G (0.5556)	D (0.6000)	G (0.2315)
4	A (0.4444)	H (0.6000)	I (0.2222)
5	B (0.4444)	A (0.5294)	D (0.1019)
6	C (0.3333)	B (0.5294)	A (0.0231)
7	E (0.3333)	C (0.5000)	B (0.0231)
8	H (0.3333)	E (0.5000)	C (0.0000)
9	I (0.2222)	I (0.4286)	E (0.0000)
10	J (0.1111)	J (0.3103)	J (0.0000)

Tabuľka 1: Usporiadanie uzlov a hodnoty centrality pri rôznych mierach centrality.

Na záver ešte uvedieme, že na výpočty centralít sme používali program `R` a v ňom konkrétne funkcie `degree()`, `closeness()` a `betweenness()` z knižnice `igraph` [7]. Krackhardtova sieť je jednou zo sietí, ktoré sú v programe `R` priamo dostupné prostredníctvom knižnice `igraphdata` [6], načítame ju pomocou príkazu `data(kite)`.

2 Vážené sociálne siete

V mnohých sociálnych sieťach okrem existencie, resp. neexistencie väzieb medzi vrcholmi pozorujeme aj ich prirodzenú pevnosť, ktorá jednotlivé väzby od seba odlišuje. Pevnosť väzby môže byť daná napríklad množstvom spoločne stráveného času, silou vzťahu či počtom interakcií medzi subjektami sociálnej siete. Na to, aby sme v rámci siete vedeli zachytiť aj informáciu o pevnosti väzieb, boli vytvorené tzv. vážené sociálne siete, v ktorých sa väzbám priradia váhy na základe intenzity vzťahu medzi uzlami prepojenými danou väzbou. Matica susednosti v prípade váženej sociálnej siete obsahujúcej vrcholy $N = \{n_1, n_2, \dots, n_p\}$ potom bude vyzeráť nasledovne

$$W = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1p} \\ w_{21} & w_{22} & \cdots & w_{2p} \\ \vdots & \ddots & & \\ w_{p1} & w_{p2} & \cdots & w_{pp} \end{pmatrix},$$

kde $p = |N|$ je počet vrcholov v sieti, $w_{ij} = 0$ ak neexistuje väzba, ktorá by spájala vrcholy n_i a n_j , a w_{ij} je kladné číslo predstavujúce váhu väzby medzi vrcholmi n_i a n_j , ak táto väzba existuje. Podobne ako pri nevážených sieťach ak existuje aspoň jedna dvojica (n_i, n_j) , pre ktorú platí $w_{ij} \neq w_{ji}$, hovoríme, že vážená sieť je orientovaná. Ak pre všetky dvojice vrcholov (n_i, n_j) platí $w_{ij} = w_{ji}$, potom ide o neorientovanú sieť.

Poznámka: Nevážené sociálne siete, v ktorých väzby medzi vrcholmi buď existujú alebo neexistujú, pričom každá existujúca väzba má váhu 1, budeme v tejto práci tiež alternatívne nazývať *binárne siete*.

2.1 Centralita vrcholov vo vážených sieťach

Freemanove miery centrality boli navrhnuté pre binárne sociálne siete, a teda brali do úvahy iba počty existujúcich hrán, pričom nezohľadňovali celkovú štruktúru siete. Ich použitím na vážené siete sa strácala podstatná časť informácie obsiahnutej v dátach. Vznikla tak potreba takých mier centrality, ktoré by zohľadňovali pri výpočte aj váhy jednotlivých väzieb. Existuje viacero pokusov zovšeobecnenia Freemanových mier centrality na vážené sociálne siete, avšak všetky tieto pokusy sa zameriavajú výlučne na váhy a nie na počet väzieb, ako to bolo navrhnuté v pôvodných Freemanových mierach.

2.1.1 Centralita stupňa pre vážené siete

Centralita stupňa bola vo vážených sociálnych sieťach definovaná ako súčet váh prislúchajúcich väzbám, ktoré sú pripojené k vrcholu [2],[21],[23]. Formálne môžeme takto definovanú centralitu stupňa vrcholu n_i vo váženej sociálnej sieti zapísať nasledovne

$$C_d^W(n_i) = \sum_{j=1}^p w_{ij},$$

kde p predstavuje celkový počet uzlov v sieti a w_{ij} sú váhy z matice susednosti W . Túto centralitu stupňa budeme nazývať aj *pevnosťou* uzla. Všimnime si, že pevnosť uzla n_i je ekvivalentná stupňu vrcholu $\deg(n_i)$ v prípade binárnych sociálnych sietí, teda takých, kde všetky väzby majú váhu 1. Jednoduchosť výpočtu je jej hlavnou výhodou, keďže stačí poznať lokálnu štruktúru siete okolo vrcholu na to aby sa dala vypočítať. Pre vážené sociálne siete je takáto miera centrality preferovaná, avšak, ako sme už spomenuli, jej nevýhodou je, že nezohľadňuje počty väzieb vrcholov.

2.1.2 Centralita blízkosti a stredovej medzipolohy pre vážené siete

Centralita blízkosti a stredovej medzipolohy v nevážených sociálnych sieťach je založená na myšlienke najkratšej vzdialenosti medzi vrcholmi. Jej výpočet pre siete, kde väzba buď existuje alebo neexistuje je pomerne jednoduchý: stačí spočítať počet hrán pozdĺž najkratšej cesty. Komplikovanejšia situácia nastáva, keď sú v rámci siete väzby rozlíšené, ako je to vo vážených sociálnych sieťach. Napríklad, ak je medzi vrcholmi n_i a n_j dvakrát väčšia váha ako medzi n_i a n_k , mohlo by to znamenať, že uzol n_i má dvakrát častejší kontakt s uzlom n_j ako s uzlom n_k . To následne implikuje, že z vrcholu n_i do vrcholu n_j by sa informácia či choroba rozšírila dvakrát pravdepodobnejšie ako z vrcholu n_i do vrcholu n_k . Ak by sme sa teda pozerali ako sa v sieti rozširuje informácia alebo choroba, rýchlosť a cesta rozšírenia sú priamo ovplyvnené váhami väzieb medzi vrcholmi.

V roku 1956 Edsger W. Dijkstra navrhol algoritmus na hľadanie najkratšej cesty medzi vrcholmi grafu [8]. V tomto algoritme sa váhy chápu ako určité straty alebo náklady na cestu, ktorých súčet je potrebné minimalizovať. V roku 2001 Newman [20] aplikoval Dijkstrov algoritmus na vážené sociálne siete použitím obrátených hodnôt

kladných váh, čím sa váhy transformovali na náklady. Pri vážených sieťach budeme pod najkratšou cestou rozumieť „cestu najmenšieho odporu“, teda takú, pri ktorej sú straty alebo náklady na cestu najmenšie. Inými slovami povedané, najkratšia cesta z vrcholu n_i do vrcholu n_j je taká, na ktorej je súčet obrátených hodnôt váh väzieb, cez ktoré cesta vedie, minimálny. Centralitu blízkosti vrcholu n_i vo váženej sieti možno potom vypočítať ako obrátenú hodnotu súčtu najkratších ciest z vrcholu n_i ku všetkým ostatným vrcholom vo váženej sieti. Matematicky to môžeme zapísať

$$C_c^W(n_i) = \left[\sum_{j=1}^p l^W(i, j) \right]^{-1},$$

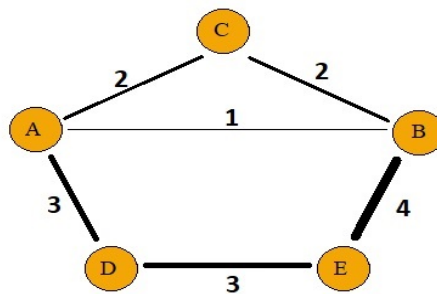
kde p je celkový počet vrcholov v sieti a $l^W(i, j)$ reprezentuje dĺžku najkratšej cesty medzi vrcholmi n_i a n_j vo váženej sociálnej sieti vypočítanú pomocou Dijkstrovho algoritmu. Centralitu stredovej medzipolohy pre vrchol n_i počítame rovnako ako pri nevážených sieťach zo vzorca

$$C_b^W(n_i) = \sum_{i \notin \{k, j\}} \frac{\mathcal{P}_i^W(j, k)}{\mathcal{P}^W(j, k)},$$

kde $\mathcal{P}^W(j, k)$ je počet najkratších ciest z vrcholu n_j do vrcholu n_k vo váženej sociálnej sieti vypočítaných pomocou Dijkstrovho algoritmu a $\mathcal{P}_i^W(j, k)$ je počet tých najkratších ciest medzi vrcholmi n_j a n_k , ktoré prechádzajú vrcholom n_i .

Príklad 2.1. Tento príklad je založený na podobnom príklade z článku [24]. Využíva sa v ňom rovnaká sociálna sieť, avšak váhy boli mierne upravené pre potreby nášho príkladu. Cieľom bude názorne ukázať ako sa s použitím Dijkstrovho algoritmu určuje cesta najmenšej vzdialenosti medzi dvoma uzlami vo váženej sociálnej sieti. Zároveň poukážeme aj na fakt, že najkratšia cesta z hľadiska počtu hrán medzi uzlami sa nemusí zhodovať s cestou, ktorú určil Dijkstrov algoritmus ako najrýchlejšiu.

Uvažujme sociálnu sieť z Obr. 5 tvorenú piatimi uzlami A, B, C, D, E, ktoré sú spojené váženými väzbami. Pevnosť väzby je vyjadrená číselnou hodnotou váhy pri hrane a zároveň aj graficky hrúbkou hrany. V tomto príklade budeme hľadať najkratšiu cestu z vrcholu A do vrcholu B. Je zrejmé, že ak by táto sieť bola binárna, najkratšou cestou by bolo priame spojenie po hrane $(\{A, B\})$. Vo váženej sieti však môže nastať situácia, že najkratšia cesta z hľadiska počtu hrán nemusí byť aj najkratšia resp. najrýchlejšia z



Obr. 5: Sociálna sieť tvorená 5 uzlami a 6 váženými väzbami, v ktorej existujú tri cesty z A do B.

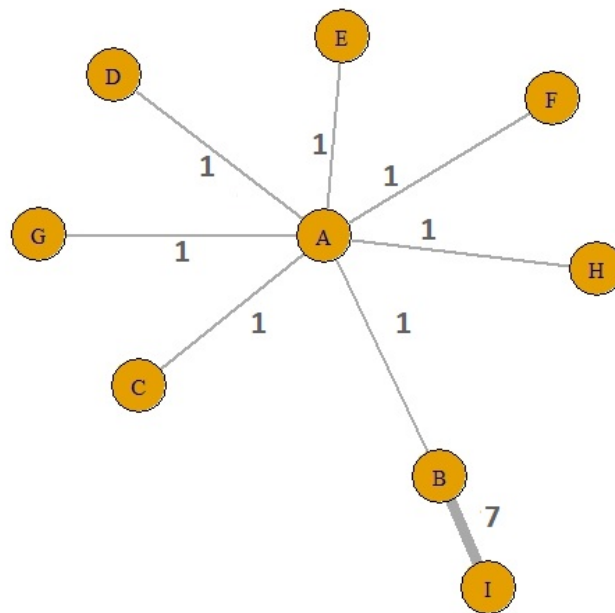
hľadiska rýchlosti prenosu. Inak povedané, v niektorých prípadoch sa napr. informácia vďaka silným väzbám dokáže dostať k adresátovi rýchlejšie po ceste s väčším počtom medziuzlov. Ako už Obr. 5 naznačuje, z A do B existujú tri cesty:

1. $\{A, B\}$,
2. $\{A, C, B\}$,
3. $\{A, D, E, B\}$.

Každá z nich má iný počet prostredných uzlov, cez ktoré je potrebné na ceste prejsť. Ako už bolo spomenuté, Dijkstrov algoritmus bol navrhnutý pre siete, kde váhy predstavovali isté straty na hrane. Z toho dôvodu je nutné pri výpočte pracovať s obrátenými hodnotami váh. Hľadáme teda cestu, na ktorej je súčet obrátených hodnôt váh najmenší. Počítajme dĺžku najkratšej cesty Dijkstrovým algoritmom medzi vrcholmi A a B v sieti z Obr. 5 ako

$$l^W(A, B) = \min \begin{cases} \frac{1}{1} = 1, & \text{pre cestu } \{A, B\} \\ \frac{1}{2} + \frac{1}{2} = 1, & \text{pre cestu } \{A, C, B\} \\ \frac{1}{3} + \frac{1}{3} + \frac{1}{4} = \frac{11}{12}, & \text{pre cestu } \{A, D, E, B\} \end{cases}$$

Všimnime si, že počet uzlov resp. počet hrán nie je žiadnym spôsobom do výpočtu zahrnutý. Dijkstrov algoritmus implicitne predpokladá, že počet medziuzlov predstavuje len bezvýznamnú stratu a jeho efekt je preto zanedbateľný. Z výpočtov je už jasne



Obr. 6: Vážená sociálna sieť zostrojená z 9 uzlov a 8 vážených väzieb, na ktorej sa ilustruje konflikt vo vnímaní centrálnosti vrcholov z hľadiska počtu väzieb a z hľadiska váh väzieb.

vidieť, že ako najlepšia cesta z hľadiska rýchlosti prenosu vyšla tretia v poradí, a to aj napriek tomu, že táto cesta obsahuje najväčší počet medziuzlov. Taktiež je dobré si všimnúť, že hoci na druhej ceste je potrebné prejsť po väčšom počte hrán ako na prvej, na základe výpočtov sú obe cesty rovnocenné, no obidve horšie ako tretia.

2.2 Váhy vs. počet väzieb

V predchádzajúcich kapitolách sme videli, že tak centralita stupňa ako aj centralita blízkosti či centralita stredovej medzipolohy zovšeobecnené na vážené siete neberú do úvahy počet väzieb napojených na vrchol, resp. počet hrán na najkratšej ceste. Všetky tri Freemanove miery centrality však boli pôvodne založené na počte väzieb, čo môže niekto považovať za dôležitejší indikátor centrálnosti vrcholu a teda by mal mať aj väčší vplyv na centralitu. Uvedieme jednoduchý príklad.

Príklad 2.2. Na Obr. 6 je znázornená vážená sociálna sieť pozostávajúca z deviatich uzlov A - I, ktoré sú spojené hranami predstavujúcimi väzby. Pevnosť väzieb je vyjadrená číselnou váhou pri každej z hrán a zároveň je naznačená aj graficky hrúbkou

Uzol	Nevážené centrality		
	Stupeň	Blízkosť	Medzipoloha
A	7.00	7.50	27.00
B	2.00	5.00	7.00
C	1.00	4.33	0.00
D	1.00	4.33	0.00
E	1.00	4.33	0.00
F	1.00	4.33	0.00
G	1.00	4.33	0.00
H	1.00	4.33	0.00
I	1.00	3.50	0.00

Tabuľka 2: Hodnoty centrality stupňa, blízkosti a medzipolohy (Freemanove centrality), ak by sme nebrali do úvahy váhy väzieb.

Uzol	Vážené centrality		
	Pevnosť	Blízkosť	Medzipoloha
A	7.00	4.50	27.00
B	8.00	6.29	7.00
C	1.00	2.55	0.00
D	1.00	2.55	0.00
E	1.00	2.55	0.00
F	1.00	2.55	0.00
G	1.00	2.55	0.00
H	1.00	2.55	0.00
I	7.00	6.10	0.00

Tabuľka 3: Hodnoty zovšeobecnených centralít stupňa (pevnosť), blízkosti a medzipolohy, ak neberieme do úvahy počet väzieb.

hrany. Ak by sme sa na túto sieť pozerali cez Freemanove miery centrality, ako najcentrálnejší by vo všetkých troch mierach vyšiel vrchol A (Tab. 2). Tento vrchol má najviac väzieb a strategicky najvýhodnejšiu pozíciu lebo spája najviac dvojíc ostatných uzlov. Na druhej strane, keby sme si v sieti pri výpočte centralít vrcholov všímali iba váhy väzieb a ich počet by sme ignorovali, výsledky by boli značne odlišné (Tab. 3). Pozrime sa napríklad na pevnosť uzlov, t. j. súčty váh väzieb pripojených k uzlu, ktorá je analógiou stupňa vrcholu pri nevážených sieťach. Tu je zaujímavé všimnúť si dva vrcholy: vrchol A a vrchol I. Obidva vrcholy majú rovnakú mieru centrality napriek tomu, že vrchol A má výrazne väčší počet väzieb ako vrchol I, ktorý má väzbu len jednu. V mnohých prípadoch by sa však práve vyšší počet väzieb mohol považovať za lepší indikátor centrálnosti vrcholu.

V snahe skombinovať stupeň vrcholu a váhy väzieb vrcholu autori v článku [24] zaviedli ladiaci parameter $\alpha \geq 0$, prostredníctvom ktorého je možné nastaviť relatívnu dôležitosť počtu väzieb vzhľadom na váhy. Parameter α sa dá nastaviť ľubovoľne v závislosti od toho, aké máme dáta a čo chceme skúmať.

2.2.1 Váhy a počet väzieb pri centralite stupňa

Prvá miera centrality navrhnutá v [24], ktorá kombinuje počet väzieb pripojených k vrcholu a priemernú váhu väzieb vrcholu využitím parametra α , je odvodená z centrality

stupňa. Jej matematický zápis vyzerá nasledovne

$$C_d^{\alpha W}(n_i) = \deg(n_i) \times \left(\frac{s_i}{\deg(n_i)} \right)^\alpha, \quad (1)$$

kde $\deg(n_i)$ je stupeň uzla n_i a $s_i = C_d^W(n_i)$ je súčet váh prislúchajúcich väzbám, ktoré sú spojené s uzlom n_i . Po úprave dostávame vzorec (1) v tvare

$$C_d^{\alpha W}(n_i) = \deg(n_i)^{1-\alpha} \times s_i^\alpha. \quad (2)$$

Z tvaru (2) sa dajú vypočítavať dve dôležité hodnoty pre parameter α : $\alpha = 0$ a $\alpha = 1$.

Keď $\alpha = 0$, platí, že

$$C_d^{\alpha W}(n_i) = \deg(n_i).$$

Takto definovaná miera centrality sa potom zhoduje so stupňom vrcholu, čiže sa pozerá výlučne na počet väzieb pripojených k vrcholom. Podobne ak nastavíme $\alpha = 1$ potom

$$C_d^{\alpha W}(n_i) = s_i = C_d^W(n_i),$$

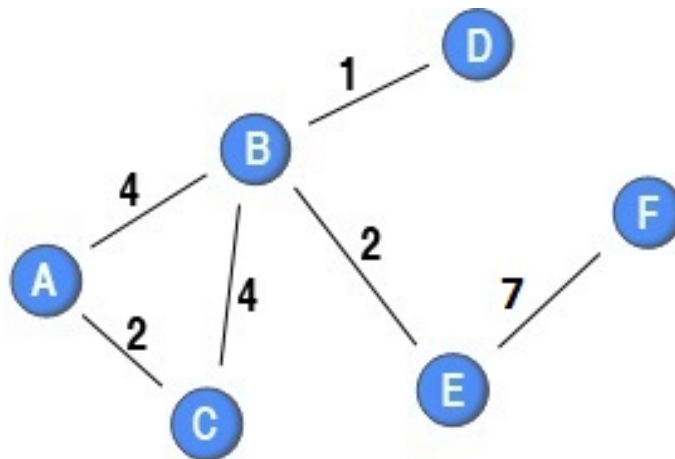
a teda je ekvivalentná pevnosti uzla.

Uvažujme teraz parameter $\alpha < 1$ a predpokladajme, že pevnosť uzla $C_d^W(n_i) = s_i$ je zafixovaná. So vzrastajúcim počtom väzieb, na ktoré je rozložený súčet váh, sa zväčšuje aj hodnota miery centrality. Väčší počet väzieb tak kladne vplýva na centralitu, čím sa pri výpočte zabezpečí väčšie či menšie zohľadnenie stupňa vrcholu. Na druhej strane, keď parameter $\alpha > 1$ a pevnosť uzla $C_d^W(n_i) = s_i$ je fixovaná, dá sa pozorovať, že vzrastajúci počet väzieb znižuje mieru centrality.

Keby sme naopak predpokladali, že stupeň vrcholu je fixný, a teda, že vrchol má pevne daný počet väzieb, vidíme, že so vzrastajúcou pevnosťou uzla sa zväčšuje aj hodnota miery centrality s výnimkou prípadu keď $\alpha = 0$.

Poznámka: Ak zvolíme $\alpha = 1$, vrcholy A a I z príkladu 2.2 majú rovnakú centralitu. Pre $\alpha = 0.5$ je však centralita vrchola A (7) väčšia ako centralita vrchola I (2.646), kým pre $\alpha = 1.5$ je centralita vrchola A (7) menšia ako centralita vrchola I (18.520).

Príklad 2.3. Vplyv ladiaceho parametra α ukážeme na príklade neorientovanej váženej sociálnej siete. Tento príklad je založený na podobnom príklade z článku [24] s malou úpravou váh. Na Obr. 7 je zobrazená vážená sociálna sieť pozostávajúca zo šies-



Obr. 7: Neorientovaná vážená sociálna sieť z príkladu 2.3.

tich vrcholov A - F. Tie sú spojené hranami, ktoré reprezentujú väzby medzi uzlami. Číslo uvedené pri hrane predstavuje váhu väzby medzi vrcholmi, ktoré hrana spája. V prípade, keď dva vrcholy siete nie sú spojené hranou, neexistuje väzba medzi týmito vrcholmi a váha takejto väzby je potom nulová.

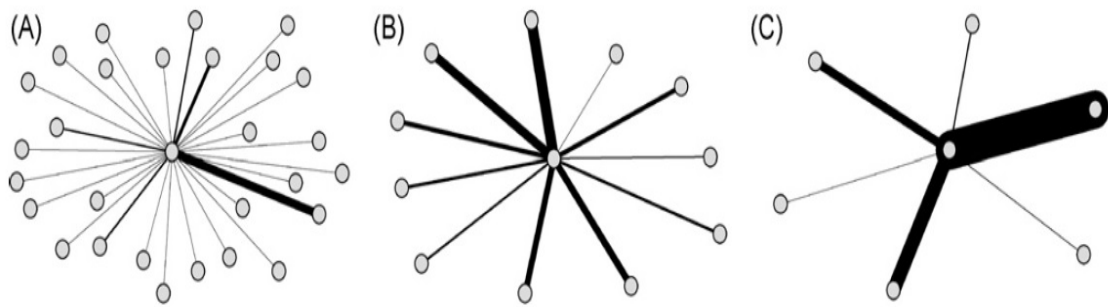
	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$
1	B (4)	B (6.63)	B (11)	E (19.09)
2	A (2)	E (4.24)	E (9)	F (18.52)
3	C (2)	A (3.46)	F (7)	B (18.24)
4	E (2)	C (3.46)	A (6)	A (10.39)
5	D (1)	F (2.65)	C (6)	C (10.39)
6	F (1)	D (1)	D (1)	D (1)

Tabuľka 4: Hodnoty miery centrality pre jednotlivé vrcholy pri rôznych nastaveniach parametra α z príkladu 2.3.

V Tab. 4 sú usporiadané vrcholy sociálnej siete z Obr. 7 na základe hodnoty centrality stupňa pri rôznych nastaveniach parametra α . Skúmajúc výsledky v tabuľke si môžeme všimnúť, že takmer vo všetkých prípadoch vyšiel ako najviac centrálny vrchol B. Dôvodom je, že má najviac prepojení na ostatné vrcholy siete a zároveň tieto väzby sú pomerne silné. Vďaka týmto dvom faktom vrchol B vychádza ako najcentrálnejší tak

v prípade, že nás zaujíma iba jeho postavenie v rámci siete (t. j. počet priamych hrán s ostatnými vrcholmi) ako aj v prípade, že sa pozeráme iba na súčet váh väzieb, kým ich počet je pre nás irelevantný. Až pri nastavení α na hodnotu 1.5 vrchol B klesá z hľadiska centrálnosti na tretie miesto a ako najviac centrálny sa zdá byť vrchol E. Dôvodom je veľký súčet váh pri vrchole E rozložený na menší počet väzieb, čo pri $\alpha > 1$ zvyšuje hodnotu centrality. Ako najmenej centrálny vo všetkých prípadoch vyšiel uzol D. Keďže spolu s uzlom F má najmenší stupeň vrcholu zo všetkých vrcholov siete a zároveň aj najmenšiu pevnosť uzla, ani so znižovaním dôrazu na počet hrán sa jeho pozícia z hľadiska centrálnosti nezlepšila. Efekt ladiaceho parametra α je najviac viditeľný pri vrchole F. Keby sme sa na túto sociálnu sieť pozerali ako na binárnu a pri výpočte centrality by sme nastavením $\alpha = 0$ váhy úplne ignorovali, uzol F by z hľadiska centrality stupňa vyšiel ako najhorší. To znamená, že vrchol F má v sieti s ostatnými vrcholmi najhoršie vzťahy. Avšak, akonáhle do výpočtu zahrnieme aj váhu väzby, ktorou je spojený s vrcholom E, jeho pozícia sa zlepší. Je to najmä preto, že s rastúcim α sa táto vysoká hodnota váhy oveľa viac zohľadnila a pri $\alpha = 1.5$ bol menší počet väzieb dokonca výhodou.

Príklad 2.4. Tento príklad je prebratý z článku [24] a ilustruje vplyv rôznych hodnôt α na centralitu uzlov v sieti. V príklade sa využíva všeobecne známy EIES (Electronic Information Exchange System) súbor dát, ktorý obsahuje tri rôzne sociálne siete. Prvé dve sa zaoberajú intenzitou priateľstva medzi skúmanými vedcami na začiatku a konci štúdie a tretia sa zameriava na elektronickú komunikáciu medzi 32 vedcami, pričom v príklade sa využíva práve tretia menovaná. Ide o váženú sociálnu sieť pozostávajúcu z 32 uzlov (32 vedcov) a väzieb definovaných ako počty elektronických správ, ktoré si medzi sebou vedci poslali. Na základe hodnoty centrality stupňa pri nastavení parametra α postupne na hodnoty 0, 0.5, 1 a 1.5 bolo ku každej hodnote zostavené poradie od najviac centrálnych uzlov po najmenej centrálnu (podobne ako sme to urobili v príklade 2.3). Skúmajúc poradia jednotlivcov sa ako najviac centrálny vo všetkých prípadoch ukázal Linn Freeman, nakoľko bez ohľadu na α sa v každom poradí nachádzal na prvom mieste. Znamená to, že Linn Freeman je spojený s najväčším počtom ľudí v sieti a zároveň celkovo im rozposlal najväčší počet správ. Rovnaký počet spojení ako Linn Freeman mali aj ďalší dvaja vedci Sue Freeman a Nick Mullins, avšak akonáhle



Obr. 8: Znázornenie individuálnych sociálnych sietí troch vedcov z Freemanovej EIES sociálnej siete: Phipps Arabie (A), John Boyd (B), a Maureen Hallinan (C). Šírka väzby zodpovedá počtu správ poslaných od ústredného uzla k jeho kontaktom.

sa α zvýšila z 0 na 0.5 ich postavenie z hľadiska centrálnosti sa zhoršilo lebo poslali výrazne menej správ ako iní vedci, aj keď väčšiemu počtu z nich.

Väčšina vedcov si aj pri zmene α zachovala relatívne stabilnú pozíciu. Zaujímavé výsledky vyšli najmä pri porovnaní troch mien: Phipps Arabie, John Boyed a Maureen Hallinan, ktorých individuálne sociálne siete sú znázornené na Obr. 8. Ako je vidieť, Phipps Arabie poslal správy veľkému počtu vedcov (takmer všetkým), vďaka čomu mal pri $\alpha = 0$ popredné postavenie. Avšak priemerný počet správ, ktoré poslal, je v porovnaní napr. s Johnom Boydom relatívne malý. V dôsledku toho sa jeho umiestnenie pri vyššom α zhoršuje a stáva sa menej centrálnym ako John Boyd. Úplne opačný prípad je Maureen Hallinan, ktorá má celkovo približne rovnaký počet poslaných správ ako Phipps Arabie a John Boyed, ale tieto boli poslané iba šiestim ľuďom. Keďže počet jej kontaktov je pomerne malý, pri nastavení $\alpha = 0$ je jej postavenie z hľadiska centrálnosti slabé. Na druhej strane, vďaka menšiemu počtu kontaktov je priemerný počet odoslaných správ každému z nich pomerne vysoký, čoho dôsledkom potom je, že zvýšením α na hodnotu 1.5 sa Maureen stáva podstatne centrálnjšou osobou ako sú Phipps Arabie a John Boyed.

2.2.2 Váhy a počet väzieb pri centralite blízkosti a centralite stredovej medzipolohy

Pri binárnych sociálnych sieťach sa ako najkratšia cesta medzi dvoma vrcholmi chápala tá, ktorá mala pozdĺž cesty najmenší počet hrán. Takýto prístup k identifikácii optimálnej cesty pre vážené siete však nebol úplne vhodný lebo nebral do úvahy váhy jednotlivých hrán. Na základe toho sa navrhol spôsob s využitím Dijkstrovho algoritmu a transformáciou váh w_{ij} na náklady $\frac{1}{w_{ij}}$, v ktorom už identifikácia najkratšej cesty závisela výlučne od váh.

Vhodnou transformáciou váh na náklady vstupujúce do Dijkstrovho algoritmu by mohla byť aj transformácia $\left(\frac{1}{w_{ij}}\right)^\alpha$ s parametrom $\alpha \geq 0$, ktorú navrhujú autori v článku [24]. Pri takejto transformácii voľbou $\alpha = 0$ vieme dosiahnuť, že sa na dĺžku najkratšej cesty aj s použitím Dijkstrovho algoritmu pozeráme cez počet hrán na ceste, ako to bolo pri binárnych sieťach. Zároveň pre $\alpha = 1$ dostávame rovnakú transformáciu ako bola pôvodne navrhnutá Newmanom pre vážené sociálne siete. Tu je zaujímavé pozrieť sa, aký vplyv môže mať zmena α pri určovaní optimálnej cesty. Tam, kde po pôvodnej transformácii $\frac{1}{w_{ij}}$ Dijkstrov algoritmus určil rovnakú vzdialenosť dvom cestám s rôznym počtom medziuzlov (ako v príklade 2.1 pre cesty $\{A, B\}$ a $\{A, C, B\}$), nastavenie $\alpha < 1$ priradí ceste s väčším počtom medziuzlov väčšiu vzdialenosť, inak povedané uprednostní sa cesta so slabšími váhami a menším počtom medziuzlov. Naopak, keď $\alpha > 1$, vplyv dodatočných uzlov je relatívne nepodstatný vzhľadom na silné väzby a uprednostňuje sa cesta s väčším počtom medziuzlov [24]. Zovšeobecnená miera centrality blízkosti pre vrchol n_i sa tak upraví na tvar

$$C_c^{\alpha W}(n_i) = \left[\sum_{j=1}^p l^{\alpha W}(i, j) \right]^{-1},$$

kde p je celkový počet vrcholov v sieti a $l^{\alpha W}(i, j)$ reprezentuje dĺžku najkratšej cesty medzi vrcholmi n_i a n_j vo váženej sociálnej sieti vypočítanú pomocou Dijkstrovho algoritmu s transformáciou váh w_{ij} na náklady $\left(\frac{1}{w_{ij}}\right)^\alpha$.

Centralitu stredovej medzipolohy pre vrchol n_i vypočítame v tomto prípade zo vzorca

$$C_b^{\alpha W}(n_i) = \sum_{i \notin \{k, j\}} \frac{\mathcal{P}_i^{\alpha W}(j, k)}{\mathcal{P}^{\alpha W}(j, k)},$$

kde $\mathcal{P}^{\alpha W}(j, k)$ je počet najkratších ciest z vrcholu n_j do vrcholu n_k určených pomocou Dijkstrovho algoritmu s úpravou nákladov pomocou parametra α a $\mathcal{P}_i^{\alpha W}(j, k)$ je počet tých najkratších ciest medzi vrcholmi n_j a n_k , ktoré prechádzajú vrcholom n_i .

3 Aplikácia Freemanových a zovšeobecnených mier centrality

Vplyv ladiaceho parametra α na mieru centrality stupňa sme čiastočne naznačili v kapitole 2 v príkladoch 2.3 a 2.4. V tejto kapitole aplikujeme podobný postup na váženú sociálnu sieť, ktorá zachytáva stretnutia postáv z filmu *Hviezdne vojny: Sila sa prebúda*. Získané výsledky sa pokúsime interpretovať vzhľadom na dej filmu.

3.1 Popis a vizualizácia dát

Dáta, ktoré budeme používať, sú verejne dostupné na internetovej stránke [13]. Sú vo formáte JSON (JavaScript Object Notation) [17] a na prácu s týmito dátami v programe R použijeme príkaz `fromJSON` z knižnice `rjson` [5], ktorý ich prevedie na zodpovedajúci R objekt. Dáta popisujú váženú sociálnu sieť postáv siedmeho dielu známej americkej série *Hviezdne vojny: Sila sa prebúda* z roku 2015 [26]. Pre lepšiu celkovú obraz si v krátkosti priblížime dej filmu. Príbeh sa odohráva 30 rokov po tom čo bola zničená Hviezdna smrti a následne obnovená Republika. Nová Republika však čelí novej hrozbe, Prvému rádu, ktorý zostrojil zbraň oveľa nebezpečnejšiu ako bola Hviezdna smrti. V týchto ťažkých časoch by Republika opäť potrebovala pomoc rytierov Jedi, avšak Luke, posledný z rytierov Jedi, už pred mnohými rokmi zmizol a zanechal iba mapu k miestu kde sa nachádza. Dej filmu sa začína na planéte Jakku, kde sa jednému zo spojencov Republiky podarilo časť mapy vypátrať. Armáda Odboja brániaca Republiku, je však pri preberaní mapy napadnutá rytiermi Prvého rádu. Poe, najlepší pilot Odboja, ukryje mapu do svojho droida BB-8, ktorému sa podarí utiecť, avšak on sám sa stáva zajatcom Prvého rádu. Uniknúť sa mu podarí s pomocou Finna, vzbúrenca Prvého rádu. Ich cesty sa však rozdelia po tom, čo stroj, na ktorom leteli, havaruje. Druhý deň ráno nájde Poeovho droida mladé miestne dievča Rey, ktorá ho zachráni pred rozpredaním na súčiastky. Šťastná náhoda zariadi, že Finn a Rey sa stretnú a spoločne sa snažia dopraviť droida BB-8 do centrality Odboja. Podarí sa im to aj vďaka pomoci Han Sola a jeho legendárnej vesmírnej lode Millennium Falcon.

Sociálna sieť, s ktorou budeme pracovať, obsahuje 27 uzlov predstavujúcich 27 po-

```

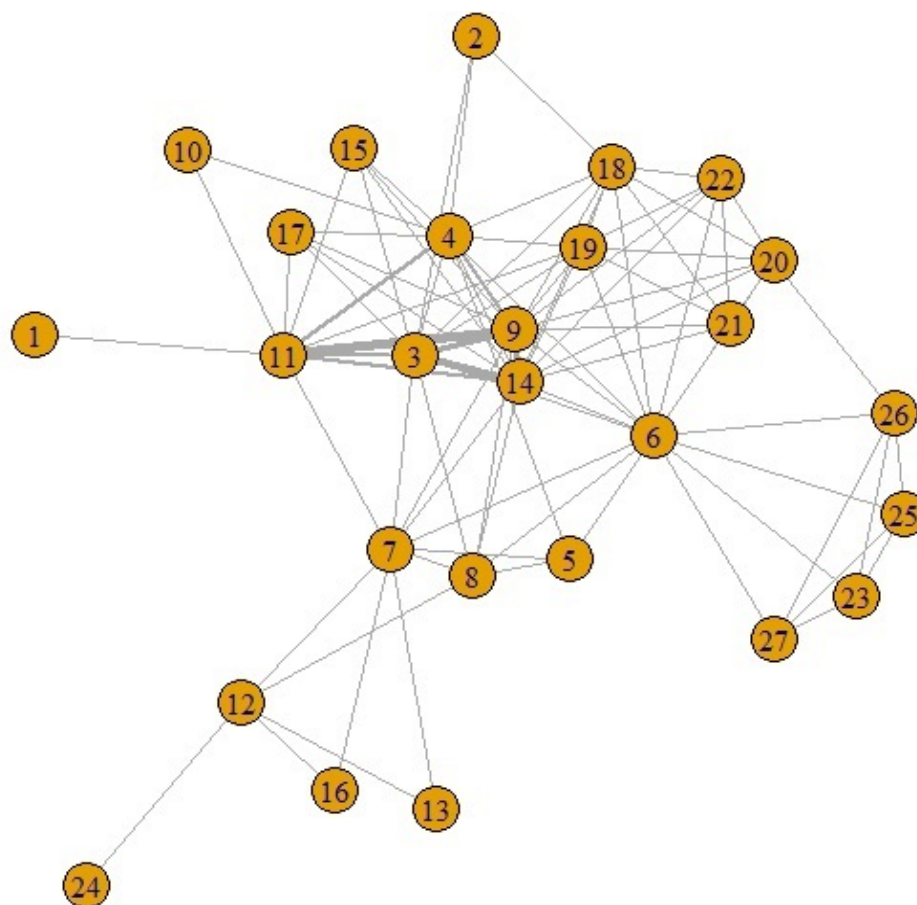
"nodes": [
  {
    "name": "LUKE",
    "value": 2,
    "colour": "#3881e5"
  },
  {
    "name": "R2-D2",
    "value": 5,
    "colour": "#bde0f6"
  },
  {
    "name": "CHEWBACCA",
    "value": 43,
    "colour": "#A0522D"
  },
  {
    "name": "BB-8",
    "value": 40,
    "colour": "#eb5d00"
  },
],
"links": [
  {
    "source": 2,
    "target": 1,
    "value": 1
  },
  {
    "source": 3,
    "target": 1,
    "value": 2
  },
  {
    "source": 3,
    "target": 2,
    "value": 9
  },
  {
    "source": 17,
    "target": 1,
    "value": 2
  },
],

```

Obr. 9: Ukážka z pôvodných dát.

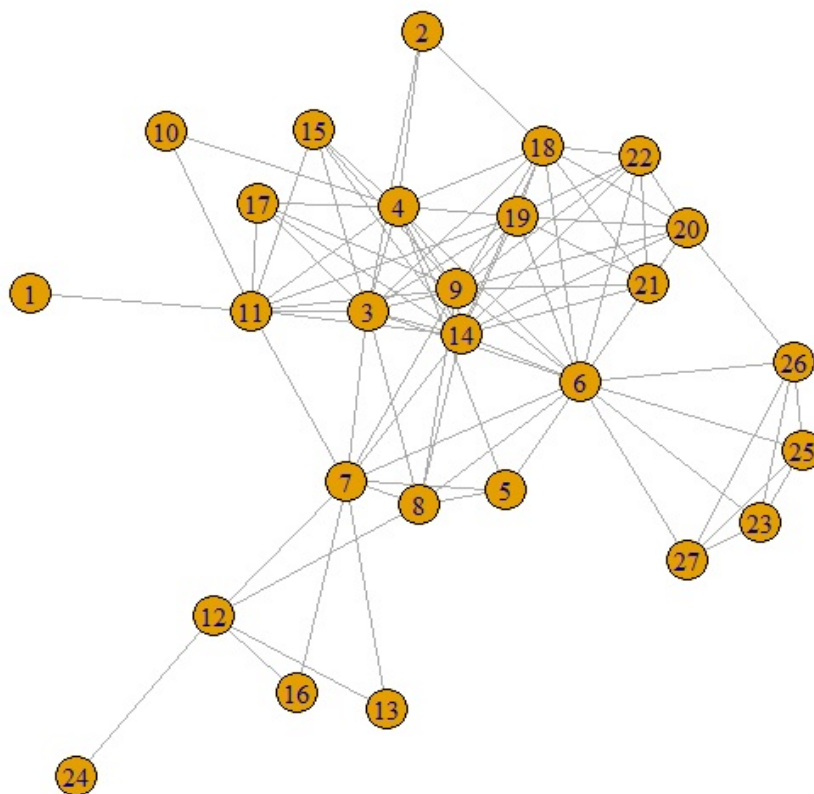
stáv filmu. Ukážku pôvodných dát, z ktorých bola táto sieť vytvorená, môžeme vidieť na Obr. 9. Informácie o uzloch siete sa nachádzajú na Obr. 9 vľavo, kde je okrem mena postavy, ktorú daný uzol reprezentuje, uvedený aj celkový počet scén, v ktorých sa táto postava objavila a farba, ktorá by mala byť použitá pri vizualizácii siete. Posledné dva údaje sme však pri výpočtoch ani pri vizualizácii siete nevyužívali. Na Obr. 9 vpravo sú zaznamenané informácie o hranách siete, a to konkrétne: čísla uzlov (počiatočný a koncový), ktoré hrana spája, a váha tejto hrany. Váha hrany je určená na základe počtu scén, v ktorých sa dané dve postavy objavili spolu. Sociálna sieť je teda v tomto prípade definovaná pomocou dvojstĺpcového zoznamu párov vrcholov spojených hranou, ku ktorým je pripojená aj informácia o váhe hrany. Dôležité je uvedomiť si, že táto sociálna sieť je neorientovaná. Ak sa totiž jedna postava objaví v rovnakej scéne s druhou, potom aj druhá s prvou je v tej istej scéne.

Pri vizualizácii dát nebudeme vrcholom priradovať konkrétne mená postáv ale ich necháme zakódované pod číslami 1 - 27. Jedným z dôvodov je, že výsledkom vizuali-



Obr. 10: Vizualizácia sociálnej siete postáv z filmu *Hviezdne vojny: Sila sa prebúda*.

zácie bude graf, ktorý tak bude oveľa prehľadnejší. Druhým dôvodom je, že tieto grafy sa následne využijú pri ďalšej časti práce. V nej sa budeme snažiť na základe odpovedí z dotazníka zistiť, ktoré miery centrality najlepšie korešpondujú so subjektívnym vnímaním centrality, prípadne či zmenou hodnoty α vieme dostať takú mieru, ktorá by lepšie opisovala subjektívny pohľad na centrálnosť vrcholov. Použitie konkrétnych mien v grafoch by tak mohlo konečné výsledky znehodnotiť. Do vizualizácie dát je tiež potrebné zahrnúť aj informáciu o váhach väzieb. Tu je najjednoduchší spôsob pretransformovať váhu väzby do hrúbky hrany, ktorá ju reprezentuje. Prvotnú vizualizáciu našej siete, v ktorej je hrúbka každej hrany prispôsobená váhe väzby, môžeme vidieť na Obr. 10. Už pri prvom pohľade na graf znázorňujúci túto sociálnu sieť je možné intuitívne určiť niektoré vrcholy (napr. vrchol 3 či vrchol 9), ktoré sú zrejme v sieti veľmi dôležité. Postavy, ktoré tieto vrcholy znázorňujú, majú dobré postavenie v sieti lebo sú v častom kontakte s veľkým počtom ďalších postáv. Očakávame preto, že táto



Obr. 11: Vizualizácia sociálnej siete postáv z filmu *Hviezdne vojny: Sila sa prebúdzá* bez ohľadu na váhy väzieb v sieti.

výhodná pozícia sa prejaví aj na vyšších hodnotách centrality. Naopak, vrcholy 1 a 24 majú väzbu iba na jeden ďalší vrchol v sieti, ktorá je ešte vzhľadom na iné väzby pomerne slabá. Pravdepodobne tak tieto vrcholy nebudú mať v rámci sociálnej siete veľký vplyv na to, čo sa v sieti deje.

3.2 Aplikácia Freemanových mier centrality na váženú sociálnu sieť

Na začiatok sa pozrieme, ktoré postavy v našej Star Wars sieti sa z hľadiska Freemanových mier centrality zdajú byť dôležité a nakoľko tieto zistenia korešpondujú s dejom filmu. Keďže Freemanove centrality boli vymyslené pre binárne siete, v ktorých sa neprihliadalo na váhu väzby, v prvom kroku urobíme pre lepšie pochopenie výsledkov vizualizáciu siete bez toho, aby sme hrúbky hrán prispôbovali váham daných väzieb. Takáto sieť je zobrazená na Obr. 11.

3 APLIKÁCIA FREEMANOVÝCH A ZOVŠEOBECNENÝCH MIER

3.2 Aplikácia Freemanových mier centrality na váženú sociálnu sieť CENTRALITY

	Centralita stupňa	Centralita blízkosti	Centralita stredovej medzipolohy
1	POE (16)	POE (0.0263)	POE (97.2109)
2	FINN (14)	FINN (0.0256)	KYLO REN (71.9044)
3	HAN (14)	HAN (0.0256)	REY (38.7449)
4	CHEWBACCA (12)	CHEWBACCA (0.0244)	BB-8 (29.1528)
5	BB-8 (12)	KYLO REN (0.0238)	FINN (26.7843)
6	KYLO REN (10)	BB-8 (0.0222)	HAN (26.7843)
7	REY (10)	CAPTAIN PHASMA (0.0213)	GENERAL HUX (26.5)
8	C-3PO (10)	REY (0.0213)	CHEWBACCA (23.862)
9	LEIA (10)	LEIA (0.0213)	CAPTAIN PHASMA (18.2623)
10	SNAP (8)	C-3PO (0.0208)	C-3PO (7.4083)
11	CAPTAIN PHASMA (7)	LOR SAN TEKKA (0.02)	LEIA (6.1278)
12	ADMIRAL ACKBAR (7)	SNAP (0.0196)	SNAP (4.5833)
13	ADMIRAL STATURA (7)	ADMIRAL ACKBAR (0.0192)	JESS (1.5)
14	GENERAL HUX (5)	ADMIRAL STATURA (0.0192)	LOR SAN TEKKA (1.1747)
15	BALA-TIK (5)	BALA-TIK (0.0179)	LUKE (0)
16	MAZ (5)	MAZ (0.0179)	R2-D2 (0)
17	JESS (5)	R2-D2 (0.0169)	UNKAR PLUTT (0)
18	LOR SAN TEKKA (4)	JESS (0.0169)	LIEUTENANT MITAKA (0)
19	YOLO ZIFF (4)	YOLO ZIFF (0.0167)	BALA-TIK (0)
20	ELLO ASTY (4)	ELLO ASTY (0.0167)	SNOKE (0)
21	NIV LEK (4)	NIV LEK (0.0167)	MAZ (0)
22	R2-D2 (3)	GENERAL HUX (0.0161)	ADMIRAL ACKBAR (0)
23	UNKAR PLUTT (2)	UNKAR PLUTT (0.0159)	ADMIRAL STATURA (0)
24	LIEUTENANT MITAKA (2)	LIEUTENANT MITAKA (0.0154)	YOLO ZIFF (0)
25	SNOKE (2)	SNOKE (0.0154)	COLONEL DATOO (0)
26	LUKE (1)	LUKE (0.0139)	ELLO ASTY (0)
27	COLONEL DATOO (1)	COLONEL DATOO (0.0115)	NIV LEK (0)

Tabuľka 5: Usporiadania postáv z filmu podľa hodnoty ich centrality vypočítanej tromi Freemanovými mierami.

V Tab. 5 môžeme vidieť v jednotlivých stĺpcoch usporiadania postáv zostavené podľa základných Freemanových mier centrality. V zátvorke pri mene postavy sa nachádza hodnota centrality vypočítaná pri miere, ktorá je uvedená ako názov stĺpca. Hodnoty centralít v tomto prípade nie sú normalizované. Pri našich Star Wars dátach nenormalizovaná centralita stupňa konkrétneho vrcholu, teda postavy reprezentovanej týmto vrcholom, vlastne hovorí, s koľkými ďalšími postavami sa počas deja aspoň raz stretla v rovnakej scéne. Neberie sa pritom do úvahy koľko tých vzájomných stretnutí bolo. Centralita stupňa čiastočne súvisí aj s centralitou blízkosti, ktorá sa pozerá na to, aký je súčet najkratších vzdialeností z hľadiska počtu hrán ku všetkým ostatným vrcholom. Ak sa vrcholu zvýši počet väzieb, vrchol má viac priamych spojení a tým sa znižuje súčet najkratších vzdialeností. Tento súčet vystupuje pri výpočte v menovateli, vďaka čomu sa zvyšuje hodnota centrality. Vidíme, že až na pár výnimiek sú usporiadania podľa centrality stupňa a centrality blízkosti veľmi podobné. Spomínanými výnimkami sú postavy LOR SAN TEKKA, ktorý je v sieti reprezentovaný uzlom 5, a R2-D2, ktorý má v sieti číslo 2. Ak si všimneme, ich postavenie sa pri centralite blízkosti výrazne

zlepšilo, až o sedem resp. päť priečok. Tieto dva uzly majú síce pomerne málo väzieb, vďaka čomu sa v prvom stĺpci nachádzajú až v poslednej desiatke, avšak sú napojené na viacero významných uzlov s veľkým počtom väzieb, čo im dopomáha k lepšiemu umiestneniu v druhom stĺpci. Čo sa týka centrality stredovej medzipolohy, tam je poradie o niečo iné ako v prvých dvoch prípadoch. Táto centralita sleduje, koľko najkratších ciest medzi všetkými párami vrcholov prechádza konkrétnym vrcholom. Prvé dva uzly, POE a KYLO REN, majú centralitu výrazne vyššiu ako ostatné uzly. POE má v sieti uzol číslo 6 a KYLO REN uzol číslo 7. Ak sa pozrieme na sociálnu sieť z Obr. 11, dá sa pozorovať, že tak uzol 6 ako uzol 7 tvoria akési premostenie medzi menšou skupinkou uzlov a zvyškom siete. Hoci k týmto uzlom existuje aj iná cesta, väčšina najkratších ciest k nim bude prechádzať práve cez uzly 6 a 7 lebo tieto majú zároveň veľký počet spojení i s uzlami vo zvyšku siete. Taktiež sa dá pozorovať, že až trinásť uzlov siete má centralitu stredovej medzipolohy nulovú. Z nich sú prekvapivé najmä dve mená: ADMIRAL ACKBAR a ADMIRAL STATURA reprezentovaní vrcholmi 21 a 22. Pri centralite stupňa sa so siedmimi väzbami tieto dve postavy umiestnili pomerne vysoko, no aj napriek vyššiemu počtu väzieb cez vrcholy 21 a 22 neprechádza žiadna najkratšia cesta. Dôvodom je, že všetky väzby vrcholov 21 a 22 sú len s takými vrcholmi, ktoré už navzájom medzi sebou spojenie majú a to priamou hranou.

Vzhľadom na dej filmu nás mená v prvej desiatke určite neprekvapujú, keďže ide o hlavné postavy, čo by však mohlo byť prekvapivé je ich poradie. Príčinám takéhoto usporiadania sa bližšie poverujeme v ďalšej podkapitole.

3.3 Aplikácia zovšeobecnenej centrality stupňa na sociálnu sieť

Nakoľko naša sociálna sieť nie je binárna ale vážená, Freemanove miery centrality pravdepodobne nebudú najvhodnejším nástrojom na analyzovanie centrálnosti jej vrcholov. V kapitole 2 sme definovali určité zovšeobecnenia týchto mier pre vážené sociálne siete, ktoré nám teraz môžu lepšie poslúžiť na rozbor vzťahov a dôležitosti jednotlivých postáv v sieti. V Tab. 6 a Tab. 7 sa nachádzajú výsledky z výpočtov zovšeobecnenej centrality stupňa uzlov s rôznym nastavením parametra α . Uzly sú v každom stĺpci tabuľky usporiadané podľa hodnoty centrality, aká im bola vypočítaná pri danom α .

	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$
1	6 (16)	9 (39.77)	9 (113)	9 (321.04)
2	9 (14)	14 (37.04)	14 (98)	14 (259.28)
3	14 (14)	3 (31.18)	11 (87)	11 (256.61)
4	3 (12)	4 (30.2)	3 (81)	3 (210.44)
5	4 (12)	11 (29.5)	4 (76)	4 (191.26)
6	7 (10)	6 (27.13)	6 (46)	6 (78)
7	11 (10)	19 (16.43)	19 (27)	19 (44.37)
8	18 (10)	7 (15.81)	7 (25)	7 (39.53)
9	19 (10)	18 (15.49)	18 (24)	18 (37.18)
10	20 (8)	20 (8.49)	17 (13)	17 (20.96)
11	8 (7)	8 (8.37)	12 (11)	12 (16.32)
12	21 (7)	17 (8.06)	8 (10)	8 (11.95)
13	22 (7)	22 (7.94)	20 (9)	25 (11.31)
14	12 (5)	12 (7.42)	22 (9)	22 (10.21)
15	15 (5)	21 (7)	25 (8)	26 (10.12)
16	17 (5)	26 (6.32)	26 (8)	20 (9.55)
17	26 (5)	25 (5.66)	5 (7)	5 (9.26)
18	5 (4)	15 (5.48)	21 (7)	27 (9.26)
19	23 (4)	5 (5.29)	27 (7)	16 (7.91)
20	25 (4)	27 (5.29)	15 (6)	21 (7)
21	27 (4)	23 (4)	2 (5)	15 (6.57)
22	2 (3)	2 (3.87)	16 (5)	2 (6.45)
23	10 (2)	16 (3.16)	23 (4)	23 (4)
24	13 (2)	10 (2.45)	10 (3)	10 (3.67)
25	16 (2)	13 (2.45)	13 (3)	13 (3.67)
26	1 (1)	1 (1)	1 (1)	1 (1)
27	24 (1)	24 (1)	24 (1)	24 (1)

Tabuľka 6: Hodnoty zovšeobecnenej centrality stupňa pri rôznych nastaveniach α .

Tieto hodnoty sú uvedené v zátvorkách.

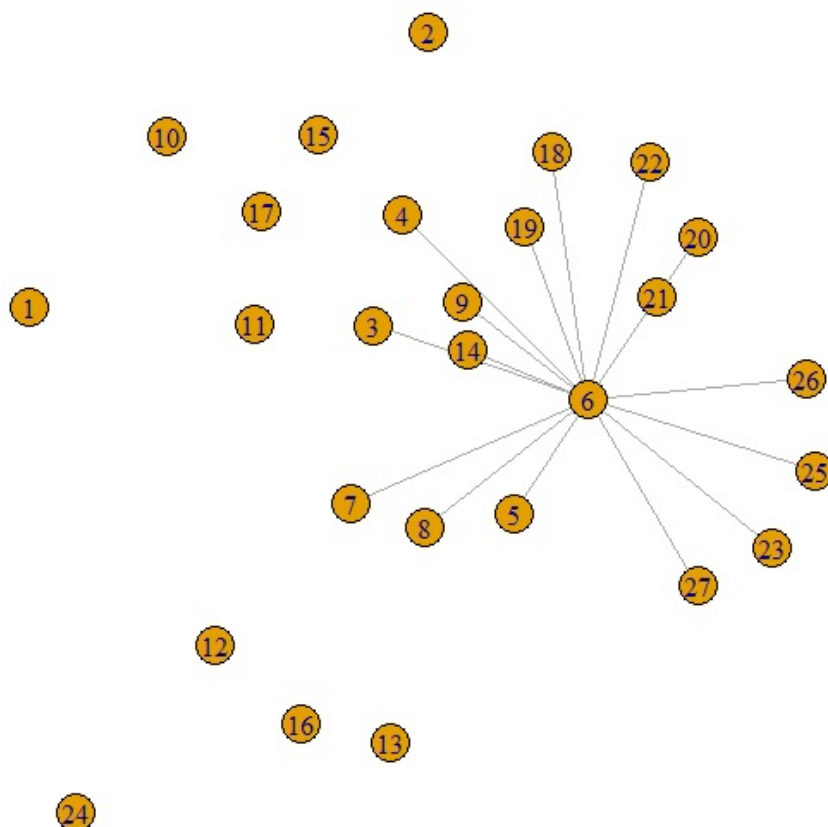
Všimnime si, že usporiadanie vrcholov na základe miery centrality je na prvých piatich miestach veľmi podobné pre všetky α . Ako najcentrálnejšie sa skoro všade ukazujú vrcholy 9 (FINN) a 14 (HAN SOLO) nasledované vrcholmi 3 (CHEWBACCA), 4 (BB-8) a 11 (REY) v rôznom poradí. Výnimkou je iba prípad, keď miera centrality nezohľadňuje váhy, teda keď $\alpha = 0$, kde sa na prvom mieste objavil vrchol 6 (POE) s počtom väzieb 16. Poe je v tomto filme jedným z bojovníkov armády Odboja brániaceho Novú Republiku. Hoci v dejovej línii filmu zohráva dôležitú úlohu, na obrazovke sa nevyskytuje tak pravidelne ako napríklad Rey či Finn, o čom svedčia aj relatívne slabé väzby. Mohlo by sa dokonca zdať, že je v celku nevýraznou postavou. To, čo z neho robí dôležitú postavu, sú jeho stretnutia s veľkým počtom iných postáv, ktoré vznikli ako dôsledok jeho presunov počas deja. Na začiatku bol zajatý armádou Prvého rádu, vďaka čomu sa objavil v rovnakej scéne s viacerými zápornými postavami. Následne sa mu podarilo ujsť a vrátil sa k armáde Odboja, kde sa zasa stretáva s mnohými ďalšími

členmi armády. Okrem toho predstavuje aj určité spojenie medzi armádou a vrchnými predstaviteľmi Republiky. V mnohých scénach sa tak objavuje buď s väčším počtom iných postáv (najmä bojovníkov Odboja) alebo s niekoľkými hlavnými postavami filmu. Vďaka tomu má najväčší počet kontaktov, čím získava v sieti špeciálne postavenie.

$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$
1 POE (16)	FINN (39.77)	FINN (113)	FINN (321.04)
2 FINN (14)	HAN (37.04)	HAN (98)	HAN (259.28)
3 HAN (14)	CHEWBACCA (31.18)	REY (87)	REY (256.61)
4 CHEWBACCA (12)	BB-8 (30.2)	CHEWBACCA (81)	CHEWBACCA (210.44)
5 BB-8 (12)	REY (29.5)	BB-8 (76)	BB-8 (191.26)
6 KYLO REN (10)	POE (27.13)	POE (46)	POE (78)
7 REY (10)	LEIA (16.43)	LEIA (27)	LEIA (44.37)
8 C-3PO (10)	KYLO REN (15.81)	KYLO REN (25)	KYLO REN (39.53)
9 LEIA (10)	C-3PO (15.49)	C-3PO (24)	C-3PO (37.18)
10 SNAP (8)	SNAP (8.49)	MAZ (13)	MAZ (20.96)
11 CAPTAIN PHASMA (7)	CAPTAIN PHASMA (8.37)	GENERAL HUX (11)	GENERAL HUX (16.32)
12 ADMIRAL ACKBAR (7)	MAZ (8.06)	CAPTAIN PHASMA (10)	CAPTAIN PHASMA (11.95)
13 ADMIRAL STATURA (7)	ADMIRAL STATURA (7.94)	SNAP (9)	ELLO ASTY (11.31)
14 GENERAL HUX (5)	GENERAL HUX (7.42)	ADMIRAL STATURA (9)	ADMIRAL STATURA (10.21)
15 BALA-TIK (5)	ADMIRAL ACKBAR (7)	ELLO ASTY (8)	JESS (10.12)
16 MAZ (5)	JESS (6.32)	JESS (8)	SNAP (9.55)
17 JESS (5)	ELLO ASTY (5.66)	LOR SAN TEKKA (7)	LOR SAN TEKKA (9.26)
18 LOR SAN TEKKA (4)	BALA-TIK (5.48)	ADMIRAL ACKBAR (7)	NIV LEK (9.26)
19 YOLO ZIFF (4)	LOR SAN TEKKA (5.29)	NIV LEK (7)	SNOKE (7.91)
20 ELLA ASTY (4)	NIV LEK (5.29)	BALA-TIK (6)	ADMIRAL ACKBAR (7)
21 NIV LEK (4)	YOLO ZIFF (4)	R2-D2 (5)	BALA-TIK (6.57)
22 R2-D2 (3)	R2-D2 (3.87)	SNOKE (5)	R2-D2 (6.45)
23 UNKAR PLUTT (2)	SNOKE (3.16)	YOLO ZIFF (4)	YOLO ZIFF (4)
24 LIEUTENANT MITAKA (2)	UNKAR PLUTT (2.45)	UNKAR PLUTT (3)	UNKAR PLUTT (3.67)
25 SNOKE (2)	LIEUTENANT MITAKA (2.45)	LIEUTENANT MITAKA (3)	LIEUTENANT MITAKA (3.67)
26 LUKE (1)	LUKE (1)	LUKE (1)	LUKE (1)
27 COLONEL DATOO (1)	COLONEL DATOO (1)	COLONEL DATOO (1)	COLONEL DATOO (1)

Tabuľka 7: Hodnoty zovšeobecnenej centrality stupňa pri rôznych nastaveniach α po priradení mien k jednotlivým uzlom.

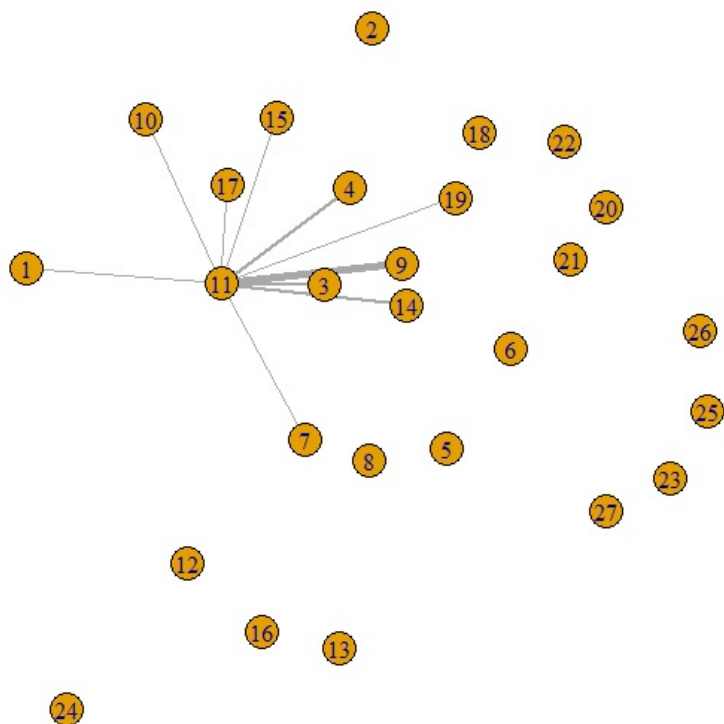
Pre mnohých, ktorí film Star Wars 7 videli, by mohlo byť prekvapivé, že jedna z hlavných postáv filmu, Rey, sa v prvom stĺpci, spolu s ďalšími dvoma postavami, umiestnila až na siedmom mieste. Rey sa, na rozdiel od Poe, na obrazovke vyskytuje prakticky stále, no väčšinou vždy len s tými istými osobami. V sieti preto nemá až taký veľký počet väzieb ako Poe, čo je dôvod, že ak by sa centralita stupňa zameriavala iba na počty väzieb uzla, skončila by Rey ďaleko za Poeom. Z Tab. 7 tiež vidíme, že Finn (vrchol 9) a Han Solo (vrchol 14) majú zhodne po 14 väzieb, vďaka čomu sa pri ignorovaní váh zaradili na druhé, resp. tretie miesto. Keď zväčšíme α z 0 na 0.5 a zahrnieme do výpočtu centrality aj váhy, lepšie je na tom Finn lebo má v sieti silnejšie väzby. Zároveň sa o dve miesta zlepšilo aj umiestnenie Rey, ktorá má taktiež veľmi silné väzby. Sociálnu sieť, v ktorej sú znázornené iba väzby pre individuálne vrcholy



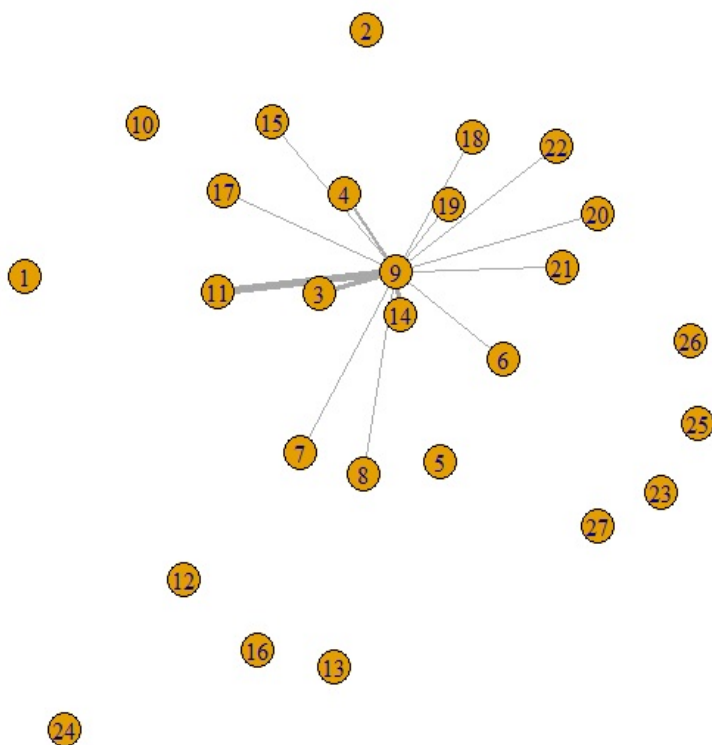
Obr. 12: Vizualizácia sociálnej siete v okolí vrcholu 6.

6 (POE), 11 (REY), 9 (FINN) a 14 (HAN SOLO) sa dajú bližšie pozrieť na Obr. 12, Obr. 13, Obr. 14 a Obr. 15.

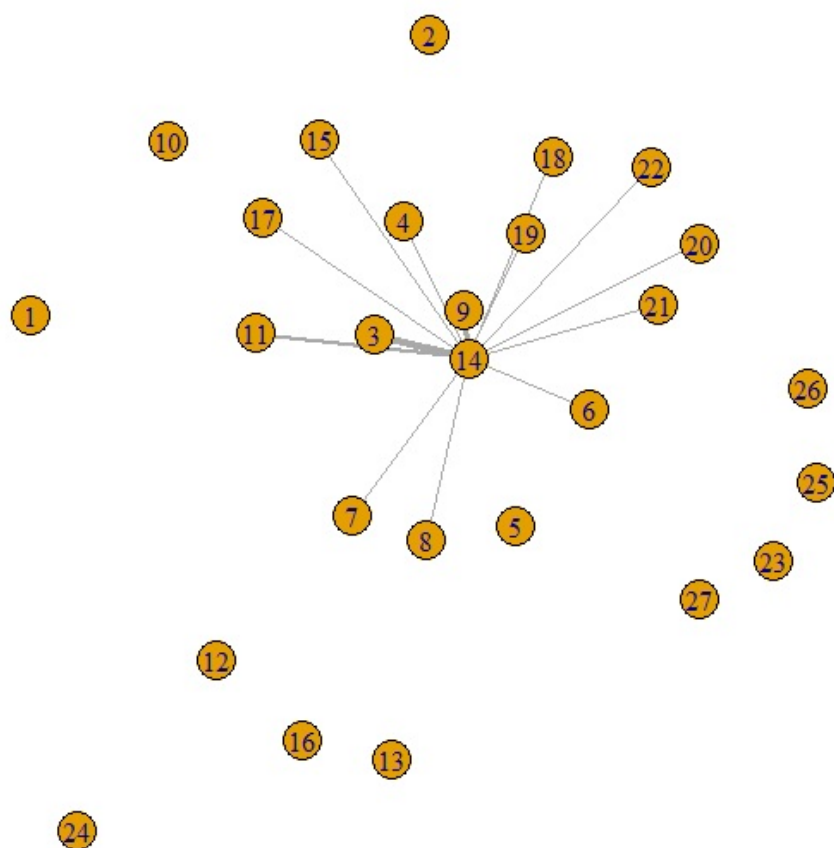
Ako sme na začiatku predpokladali najhoršie skončili vrcholy 1 a 24 a to pri všetkých nastaveniach α . Obidva mali stupeň vrcholu 1 a keďže aj váha jedinej väzby bola veľmi malá v porovnaní s ostatnými, ani zmenou α sa ich pozícia nevylepšila.



Obr. 13: Vizualizácia sociálnej siete v okolí vrcholu 11.



Obr. 14: Vizualizácia sociálnej siete v okolí vrcholu 9.



Obr. 15: Vizualizácia sociálnej siete v okolí vrcholu 14.

4 Výsledky dotazníka

Cieľom našej práce je zistiť, ako ľudia intuitívne vnímajú centralitu vrcholov v sociálnej sieti, ktoré faktory najviac ovplyňujú rozhodnutie o tom koho považujú v sieti za centrálnu jedinu a na základe toho určiť, ktorá miera najlepšie vystihuje subjektívne vnímanie centrality. Pre potreby tohoto výskumu sme vytvorili anonymný dotazník a v tejto kapitole budeme vyhodnocovať odpovede. Dotazník je rozdelený na dve hlavné časti, ktoré budeme analyzovať samostatne. Prvá časť sa zameriava na subjektívne vnímanie dôležitosti postáv na základe deja filmu *Hviezdne vojny: Sila sa prebúdzá*. V druhej časti boli otázky spojené s vizualizáciou sociálnej siete postáv tohoto filmu a s vnímaním centrality uzlov siete bez toho, aby odpovedajúci vedel, akú postavu ktorý uzol reprezentuje.

4.1 Metódy porovnania poradí

Hodnota centrality nám sama o sebe veľa nenapovie o postavení vrcholu v sieti. Okrem prípadu nenormalizovanej centrality stupňa ju ako samostatnú hodnotu nevieme dobre interpretovať. Na hodnotu centrality vrcholu je potrebné sa pozerať v kontexte porovnania s ostatnými vrcholmi, inak povedané, nie je až tak dôležité aká je hodnota centrality ale či je vyššia ako hodnoty centrality iných vrcholov. Preto aj otázky v dotazníku boli koncipované tak, aby odpovedajúci musel podľa svojho subjektívneho názoru vytvoriť poradie postáv z hľadiska ich dôležitosti pre dej filmu a z hľadiska ich centrality v predloženej sociálnej sieti. Pri vyhodnocovaní odpovedí z dotazníka potom budeme skúmať mieru zhody poradia vytvoreného pomocou výpočtu centrality s poradím, ktoré určil odpovedajúci v dotazníku. Na toto porovnávanie usporiadaní budeme využívať štyri metódy: Jarova - Winklerova podobnosť, Levenshteinova vzdialenosť, Kendallov korelačný koeficient a Spearmanov korelačný koeficient. Prvé dve metódy sú založené na podobnosti reťazcov, zatiaľ čo Kendallov a Spearmanov korelačný koeficient určujú mieru korelácie medzi dvoma skupinami poradí priradených rovnakej množine objektov.

4.1.1 Jarova - Winklerova podobnosť

Jarova - Winklerova podobnosť (ozn. JWS) vyjadruje mieru podobnosti medzi dvoma reťazcami. Ide o spojenie pôvodnej Jarovej podobnosti, ktorú navrhol v roku 1989 Matthew A. Jaro [16] na porovnávanie reťazcov, a Winklerovej modifikácie, ktorá dávala priaznivejšie hodnotenia reťazcom zhodujúcim sa v znakoch na začiatku reťazcov [27]. Po normalizácii Jarov - Winklerov porovnávací algoritmus dával hodnoty z intervalu $\langle 0, 1 \rangle$, kde 1 znamenalo perfektnú zhodu a 0 žiadne zhodné znaky, a teda čím väčšia bola podobnosť (JWS) medzi reťazcami, tým viac sa dané reťazce zhodovali.

Výpočet JWS sa podľa [4] urobí v dvoch krokoch. V prvom kroku sa vypočíta Jarova podobnosť a následne sa urobí Winklerova korekcia. Uvažujme dva reťazce s_1 a s_2 , ktoré chceme porovnať. Jarova podobnosť medzi týmito reťazcami je definovaná nasledovne

$$sim_j = \begin{cases} 0, & \text{ak } m = 0 \\ \frac{1}{3} \left(\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m-t}{m} \right), & \text{inak} \end{cases} \quad (3)$$

kde

- m je počet zhodných znakov medzi reťazcami s_1 a s_2 , pričom za zhodné sa považujú len tie znaky, pri ktorých vzdialenosť medzi ich pozíciami v reťazcoch nie je väčšia ako $\left\lfloor \frac{\max(|s_1|, |s_2|)}{2} \right\rfloor - 1$,
- $|s_1|$ a $|s_2|$ sú dĺžky reťazcov,
- t je polovica počtu transpozícií, kde transpozícia je definovaná ako počet zhodných znakov v zhodných podsekvenciách, ktoré sú však v inom poradí.

Winklerova korekcia modifikuje Jarov algoritmus podľa toho, či je medzi reťazcami s_1 a s_2 identická zhoda na prvých niekoľko miestach. Jarova - Winklerova podobnosť potom vyzerá nasledovne

$$JWS = sim_j + lp(1 - sim_j), \quad (4)$$

kde

- sim_j je Jarova podobnosť,
- $l \in \{0, 1, 2, 3, 4\}$ je počet zhodných znakov v rovnakom poradí na začiatku reťazcov s_1 a s_2 , môže nadobudnúť najviac hodnotu 4,

- p je škálovacia konštanta, ktorá by nemala presiahnuť hodnotu 0.25, štandardne sa nastavuje na hodnotu $p = 0.1$.

Príklad 4.1. Pre lepšie pochopenie Jarovej - Winklerovej podobnosti sme vytvorili príklad, na ktorom ukážeme, aký postup sa uplatňuje pri jej výpočte. Uvažujme nasledovné dva reťazce $s_1 = ABBBWXYZ$ a $s_2 = ABCBWYXZM$. Ideme vypočítať Jarovu - Winklerovu podobnosť týchto dvoch reťazcov. Najprv vypočítame maximálnu vzdialenosť pozícií rovnakých znakov, aby znaky mohli byť započítané ako zhoda: $\left\lfloor \frac{\max(|s_1|, |s_2|)}{2} \right\rfloor - 1 = \left\lfloor \frac{\max(8, 9)}{2} \right\rfloor - 1 = \left\lfloor \frac{9}{2} \right\rfloor - 1 = 3$. Vytvoríme tzv. zónu zhody (viď. Tab. 8).

	A	B	B	B	W	X	Y	Z
A	1	0	0	0				
B	0	1	0	0	0			
C	0	0	0	0	0	0		
B	0	0	1	0	0	0	0	
W		0	0	0	1	0	0	0
Y			0	0	0	0	1	0
X				0	0	1	0	0
Z					0	0	0	1
M						0	0	0

Tabuľka 8: Zóna zhody. V tabuľke sú to tie políčka, v ktorých sa nachádzajú čísla. Keďže maximálna vzdialenosť medzi znakmi aby mohli byť považované za zhodné je 3, zónu zhody tvorí hlavná diagonála, tri diagonály nad hlavnou a tri diagonály pod hlavnou. Číslo 1 znamená zhodu, číslo 0 znamená, že zhoda nenastala. Prázdne políčka sú mimo zóny zhody.

Všimnime si, že tretie B z reťazca s_1 už nemá žiadnu zhodu lebo reťazec s_2 obsahuje iba dva znaky B. Vytvoríme podreťazce ss_1 a ss_2 zo znakov, v ktorých sme našli zhodu, t. j. z tých, pri ktorých sa v Tab. 8 v príslušnom stĺpci resp. riadku objavilo číslo 1:

$$ss_1 = ABBW \underline{XYZ}$$

$$\uparrow \uparrow \uparrow \uparrow \uparrow \uparrow \uparrow$$

$$ss_2 = ABBWY \underline{XZ}.$$

Následný výpočet je už pomerne jednoduchý:

- celkový počet zhodných znakov $m = 7$, čo je súčasne dĺžka podreťazcov ss_1 a ss_2 ,
- t je polovica počtu znakov, ktoré majú v ss_1 inú pozíciu ako v ss_2 (vyznačené 2 znaky X a Y), čiže $t = \frac{2}{2} = 1$

- $l = 2$, lebo s_1 a s_2 sa identicky zhodujú v prvých dvoch znakoch
- $p = 0.1$.

Hodnoty dosadíme do vzorcov (3) a (4)

$$sim_j = \frac{1}{3} \left(\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m-t}{m} \right) = \frac{1}{3} \left(\frac{7}{8} + \frac{7}{9} + \frac{6}{7} \right) = 0.83664,$$

$$JWS = sim_j + lp(1 - sim_j) = 0.83664 + 2 \times 0.1 \times (1 - 0.83664) = 0.869312.$$

4.1.2 Levenshteinova vzdialenosť

Levenshteinova vzdialenosť (ozn. LD) je taktiež mierou podobnosti (resp. rozdielnosti) medzi dvoma reťazcami navrhnutou v roku 1965 ruským vedcom Vladimirom Levenshteinom. Princíp tejto miery je založený na myšlienke určenia minimálneho počtu zmien (nahradenie, odstránenie alebo vloženie znaku), ktoré je potrebné vykonať aby sme prvý reťazec pretransformovali na druhý [4]. Napríklad, minimálny počet zmien pri transformácii reťazca $\mathbf{a} = TEN$ na reťazec $\mathbf{b} = TEST$ je 2 (náhrada N za S a pridaanie T), a teda Levenshteinova vzdialenosť $LD = 2$. Logicky teda platí, že čím väčšia je LD , tým rozdielnejšie dané dva reťazce sú. Algoritmus na výpočet Levenshteinovej vzdialenosti je nasledovný [4]

$$lev_{a,b}(i, j) = \begin{cases} \max(i, j), & \text{ak } \min(i, j) = 0 \\ \min \begin{cases} lev_{a,b}(i-1, j) + 1 \\ lev_{a,b}(i, j-1) + 1 \\ lev_{a,b}(i-1, j-1) + \mathbf{1}_{(a_i \neq b_j)} \end{cases} & , \text{ inak} \end{cases}$$

kde $\mathbf{1}_{(a_i \neq b_j)}$ je funkcia, ktorá nadobúda hodnotu 0 ak $(\mathbf{a}_i = \mathbf{b}_j)$ a hodnotu 1 ak $(\mathbf{a}_i \neq \mathbf{b}_j)$, $lev_{a,b}(i, j)$ zodpovedá počtu zmien potrebných na transformáciu prvých i znakov reťazca $\mathbf{a}$ na reťazec pozostávajúci z prvých j znakov reťazca $\mathbf{b}$. Levenshteinova vzdialenosť LD medzi reťazcami $\mathbf{a}$ a $\mathbf{b}$ je potom hodnota $lev_{a,b}(|\mathbf{a}|, |\mathbf{b}|)$ z tohoto algoritmu, kde $|\mathbf{a}|, |\mathbf{b}|$ sú dĺžky reťazcov $\mathbf{a}$ a $\mathbf{b}$.

Príklad 4.2. Algoritmus na výpočet LD vysvetlíme na jednoduchom príklade porovnania dvoch podobných reťazcov, pričom pri vysvetľovaní postupu sme sa inšpirovali podobnými príkladmi z [4] a [19]. Ideme vypočítať Levenshteinovu vzdialenosť reťazcov $\mathbf{a} = POST$ a $\mathbf{b} = POSOL$. Postup pri výpočte bude sledovať nasledovné kroky:

1. Položíme $n = \text{dĺžka reťazca } \mathbf{a}$, $m = \text{dĺžka reťazca } \mathbf{b}$. Ak $m = 0$, potom $LD = n$; ak $n = 0$, tak $LD = m$.

Zostrojíme tabuľku $\mathbf{T}$ obsahujúcu $m + 1$ riadkov a $n + 1$ stĺpcov.

2. Prvky v nulťom riadku a j -tom stĺpci tabuľky $\mathbf{T}$ ($\mathbf{T}_{0j}$, $j = 0, \dots, m$) sa vyplnia hodnotami j . Prvky $\mathbf{T}_{i0}$ v riadku i a stĺpci 0 tabuľky $\mathbf{T}$ sa vyplnia hodnotami i ($i = 0, \dots, n$).

3. Prvky ij ($i = 1, \dots, n$; $j = 1, \dots, m$) tabuľky $\mathbf{T}$ budú rovné minimu z troch hodnôt :

$$T_{ij} = \min \begin{cases} T_{i-1,j} + 1, \\ T_{i,j-1} + 1, \\ T_{i-1,j-1} + \mathbf{1}_{(a_i \neq b_j)}. \end{cases}$$

Ak $\mathbf{a}_i = \mathbf{b}_j$ potom funkcia $\mathbf{1}_{(a_i \neq b_j)} = 0$, ak $\mathbf{a}_i \neq \mathbf{b}_j$ potom $\mathbf{1}_{(a_i \neq b_j)} = 1$.

4. Po ukončení všetkých iterácií, LD je mn -tý prvok tabuľky $\mathbf{T}$: $LD = T_{mn}$

V našom prípade $m = 5$, $n = 4$. Tabuľka $\mathbf{T}$ bude obsahovať šesť riadkov a päť stĺpcov. Do nulťého riadku tabuľky vložíme hodnoty $0, \dots, 4$ a do nulťého stĺpca tabuľky vložíme hodnoty $0, \dots, 5$ (viď. Tab. 9).

		P	O	S	T
	0	1	2	3	4
P	1				
O	2				
S	3				
O	4				
L	5				

Tabuľka 9: Tabuľka $\mathbf{T}$.

Túto tabuľku následne vyplňame postupom z kroku 3, čím dostávame Tab. 10. Požadovaná LD sa nachádza v pravom dolnom rohu tabuľky, čiže $LD = 2$. Na transformáciu reťazca $\mathbf{a} = POST$ na reťazec $\mathbf{b} = POSOL$ je teda potrebné urobiť dve zmeny, konkrétne sú to výmena T za O a pridanie L.

		P	O	S	T
	0	1	2	3	4
P	1	0	1	2	3
O	2	1	0	1	2
S	3	2	1	0	1
O	4	3	2	1	1
L	5	4	3	2	2

Tabuľka 10: Tabuľka **T** vyplnená hodnotami z algoritmu. Levenshteinova vzdialenosť medzi reťazcami $\mathbf{a} = POST$ a $\mathbf{b} = POSOL$ je číslo v pravom dolnom rohu tabuľky.

4.1.3 Spearmanov korelačný koeficient

Spearmanov korelačný koeficient (ozn. ρ_s) je neparametrickou verziou Pearsonovho korelačného koeficientu, ktorú je možné použiť v prípade, že dáta nie sú z normálneho rozdelenia alebo existuje riziko znehodnotenia výsledkov spôsobené odľahlými hodnotami. Vyjadruje mieru asociácie medzi dvoma veličinami, pričom na to využíva korelácie v poradiach. Myšlienka spočíva v zámene realizácie náhodnej premennej za jej poradie v usporiadaní podľa veľkosti vzhľadom k zvyšným hodnotám výberu [25].

Nech $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ je náhodný výber z dvojrozmerného spojitého rozdelenia. Odhad Spearmanovho korelačného koeficientu dostaneme zo vzťahu [25]

$$\hat{\rho}_s = \frac{\sum_{i=1}^n (R_i - \bar{R})(Q_i - \bar{Q})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2} \sqrt{\sum_{i=1}^n (Q_i - \bar{Q})^2}}, \quad (5)$$

kde R_i a Q_i sú poradia pridelené realizáciám x_i resp. y_i po ich usporiadaní podľa veľkosti, n je rozsah výberu a $\bar{R} = \bar{Q} = \frac{1}{n} \sum_{i=1}^n i$ je priemer vypočítaný zo všetkých poradových hodnôt. Ak nastane situácia, že sa viackrát opakujú hodnoty x_i alebo y_i , v usporiadaní sa im pridelí priemerná hodnota z poradových pozícií, ktoré obsadia. Odhad Spearmanovho korelačného koeficientu sa častejšie zvykne uvádzať v tvare [14]

$$\hat{\rho}_s = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (R_i - Q_i)^2, \quad (6)$$

ktorý sa však nedá použiť v prípade, že nastanú zhody v realizáciách. Z tvaru (5) sa dá vidieť, že Spearmanov korelačný koeficient je vlastne Pearsonov korelačný koeficient

aplikovaný na poradia. Kým Pearsonov korelačný koeficient popisuje iba mieru lineárnej závislosti medzi párovými dátami, Spearmanov korelačný koeficient zachytáva vo všeobecnosti ich monotónne závislosti, či už rastúce alebo klesajúce. Výhodou je jeho odolnosť voči odľahlým hodnotám, ktorá vzniká ako dôsledok práce s poradiami realizácií namiesto samotných realizácií. Spearmanov korelačný koeficient je vhodný na použitie aj v prípadoch, keď nemáme k dispozícii hodnoty samotných realizácií ale iba ich poradia, prípadne poznáme usporiadané hodnotenie kvalitatívnych premenných. Ako príklad sa v [11] uvádza možnosť takéhoto využitia pri zisťovaní zhody v dvoch hodnoteniach, kde každý hodnotiteľ mal usporiadať a ohodnotiť zoznam alternatív v preferovanom poradí. Spearmanov korelačný koeficient ρ_s nadobúda hodnoty z intervalu $\langle -1, 1 \rangle$. Ak majú pozorovania veľmi podobné alebo rovnaké poradie, $\hat{\rho}_s$ bude nadobúdať hodnoty blízke alebo rovné 1, ak majú poradie veľmi rozdielne alebo úplne protichodné, $\hat{\rho}_s$ nadobúda hodnoty blízke -1.

4.1.4 Kendallov korelačný koeficient

Kendallov korelačný koeficient (ozn. τ) určuje stupeň podobnosti medzi dvoma množinami hodnôt (poradí) pridelených rovnakej skupine objektov, ide teda o mieru závislosti v usporiadaniach podľa dvoch premenných. Nadobúda hodnoty z uzavretého intervalu $\tau \in \langle -1, 1 \rangle$. Pre úplnú pozitívnu závislosť v poradiach nadobúda hodnotu 1, pre úplnú negatívnu závislosť nadobúda hodnotu -1 a pri nezávislosti poradí nadobúda hodnoty blízke 0.

Je logické predpokladať, že Kendallova korelácia medzi dvoma premennými X a Y bude väčšia, keď ich realizácie majú podobné alebo identické poradie a menšia v prípade nízkej podobnosti v poradiach, resp. ich úplnej odlišnosti [25]. Nech teda $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ je náhodný výber náhodných premenných (X, Y) zo spojitého rozdelenia. Hovoríme, že páry $(x_i, y_i), (x_j, y_j)$, kde $i \neq j$ sú konkordantné páry ak platí jedna z nasledovných možností

- $(x_i > x_j) \wedge (y_i > y_j)$,
- $(x_i < x_j) \wedge (y_i < y_j)$.

Hovoríme, že páry $(x_i, y_i), (x_j, y_j)$, kde $i \neq j$ sú diskordantné páry ak nastane jedna z možností

- $(x_i > x_j) \wedge (y_i < y_j)$,
- $(x_i < x_j) \wedge (y_i > y_j)$.

Odhad teoretickej hodnoty Kendallovho koeficientu korelácie τ je definovaný vzťahom [25]

$$\hat{\tau} = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \text{sgn}(x_i - x_j) \text{sgn}(y_i - y_j). \quad (7)$$

Ak sú náhodné premenné X a Y nezávislé, odhad Kendallovho τ bude približne 0. Dôvodom je, že v pároch $(x_i, y_i), (x_j, y_j)$ nastáva približne v polovici prípadov konkordancia a v polovici prípadov diskordancia, čo znamená, že v sume z definície $\hat{\tau}$ sa pripočíta a odpočíta približne rovnaký počet jednotiek. Ak sú X a Y pozitívne závislé, nastávať bude skôr konkordancia vďaka čomu sa viac jednotiek pripočíta ako odpočíta, čiže $\hat{\tau} \approx 1$. Podobne ak sú X a Y negatívne závislé, nastáva častejšie diskordancia a teda $\hat{\tau} \approx -1$.

Rovnako ako pri Spearmanovom koeficiente korelácie, aj v tomto prípade na výpočet odhadovanej hodnoty Kendallovho korelačného koeficientu postačí, ak poznáme usporiadania premenných X a Y náhodného výberu. Ich poradia môžeme potom vo vzorci (7) ekvivalentne použiť namiesto samotných realizácií.

4.2 Vnímanie dôležitosti postáv na základe filmu

Ako už bolo spomenuté, v prvej časti nášho dotazníka boli otázky späté priamo s dejom filmu *Hviezdne vojny: Sila sa prebúda*. To implicitne predpokladá, že odpovedajúci film už niekedy predtým videli. Z celkového počtu 37 prijatých odpovedí sme preto museli vylúčiť tie, pri ktorých respondent odpovedal záporne na otázku, či už tento film niekedy videl. Takáto odpoveď bola jedna. Okrem toho nás zaujímalo aj to, či odpovedajúci videl aj iné časti zo série *Hviezdných vojen*. V tejto časti 30 ľudí uviedlo, že okrem siedmeho dielu videli tiež aspoň jednu z predchádzajúcich častí. S týmito dátami sme potom mohli pristúpiť k analýze prvej polovice dotazníka, ktorej základ tvorili dve otázky:

```
#Hlavné postavy: označte 9 postáv, ktoré pokladáte
#vo filme Hviezdne vojny 7: Sila sa prebúdzá za hlavné
#
CHEWBACCA,BB-8,POE,KYLO REN,FINN,REY,GENERAL HUX,HAN,SNOKE
BB-8,POE,KYLO REN,FINN,REY,GENERAL HUX,SNOKE,MAZ,LEIA
R2-D2,KYLO REN,FINN,REY,HAN,SNOKE,MAZ,LEIA,ADMIRAL ACKBAR
R2-D2,CHEWBACCA,BB-8,POE,KYLO REN,FINN,REY,HAN,LEIA
LUKE,R2-D2,CHEWBACCA,BB-8,KYLO REN,FINN,HAN,C-3PO,LEIA
CHEWBACCA,BB-8,POE,KYLO REN,FINN,REY,HAN,C-3PO,LEIA
R2-D2,CHEWBACCA,BB-8,POE,FINN,REY,HAN,C-3PO,LEIA
LUKE,R2-D2,BB-8,POE,KYLO REN,FINN,REY,HAN,C-3PO
CHEWBACCA,BB-8,POE,KYLO REN,FINN,REY,GENERAL HUX,HAN,LEIA
LUKE,R2-D2,LOR SAN TEKKA,KYLO REN,FINN,REY,GENERAL HUX,HAN,LEIA
```

Obr. 16: Ukážka dát: Každý riadok predstavuje jednu odpoveď. Odpovedajúci si z ponúkaného zoznamu postáv vyberal 9, ktoré pokladal vo filme *Hviezdne vojny: Sila sa prebúdzá* za hlavné.

1. Označte 9 postáv, ktoré pokladáte vo filme *Hviezdne vojny 7: Sila sa prebúdzá* za hlavné.
2. Postavy uvedené nižšie zoradte podľa dôležitosti. Na základe Vami vytvoreného poradia priradte postavám hodnoty od 1 do 9, kde 1 je najdôležitejšia (najdôležitejšia dostane hodnotu 1, druhá najdôležitejšia hodnotu 2, atď. až po 9).

Každú otázku budeme rozoberať samostatne.

4.2.1 Výber 9 hlavných postáv

Na Obr. 16 môžeme vidieť ukážku z dát, ktoré sme dostali ako odpovede na prvú z vyššie uvedených otázok. Okrem spomínanej odpovede, kde odpovedajúci uviedol, že film *Hviezdne vojny: Sila sa prebúdzá* nevidel, sme museli v tejto časti vylúčiť aj odpovede, kde bol označený menší počet postáv ako bolo v zadaní. Celkovo nám tak ostalo 30 odpovedí, ktoré sa nachádzajú uložené v súbore *Vyber_9_hlavnych_postav.txt*. V každom riadku súboru je zaznamenaných 9 mien oddelených čiarkou, ktoré označil jeden odpovedajúci ako hlavné postavy filmu. Na prácu s dátami sme použili program R a kód v ňom vytvorený pre vyhodnotenie týchto dát nájdeme v priloženom súbore *Kap4_2_1_Vyber_9_hl_postav.R*.

```

> ScoreMatrix_Stupen
      alpha = 0 alpha = 0.5 alpha = 1 alpha = 1.5
odpoved_1      182.5      186.0      190.5      193.0
odpoved_2      166.5      172.0      179.5      182.0
odpoved_3      151.0      157.0      160.0      160.0
odpoved_4      192.5      194.0      194.5      194.0
odpoved_5      167.0      169.5      168.0      167.5
odpoved_6      207.0      207.0      207.0      207.0
odpoved_7      192.5      193.0      193.5      193.0
odpoved_8      170.5      168.5      170.0      169.5
odpoved_9      199.0      202.0      205.0      205.0
odpoved_10     141.0      147.0      154.0      154.0
odpoved_11     184.0      190.0      191.5      194.0
odpoved_12     182.5      186.0      190.5      193.0

```

Obr. 17: *ScoreMatrix_Stupen*: matica vyplnená "score" hodnotami, čo sú súčty hodností pridelené postvám z odpovede. Hodnosti boli pridelené podľa hodnoty centrality stupňa vypočítanej pri štyroch nastaveniach α : $\alpha = 0$, $\alpha = 0.5$, $\alpha = 1$, $\alpha = 1.5$.

Myšlienku postupu pri spracovaní dát a tvorbe kódu vysvetlíme len s výpočtami centrality stupňa, keďže pri centralite blízkosti a centralite stredovej medzipolohy je princíp rovnaký, mení sa pri nich iba typ centrality. Prvým krokom je vytvorenie väzenej sociálnej siete z verejne dostupných dát [13] a následné vypočítanie centralít stupňa pri nastaveniach α postupne na hodnoty 0, 0.5, 1, 1.5. Použijeme na to príkaz `degree_w()` s príslušným nastavením α . Pri každom z týchto štyroch nastavení urobíme usporiadanie postáv podľa vypočítanej centrality a určíme im poradové hodnoty (niekedy budeme používať výraz hodnoti) podľa poradového miesta, na ktorom sa umiestnili. Postavy s najvyššou centralitou dostanú najvyššie hodnoti. Na to nám poslouží príkaz `rank()`. Ďalej načítame všetky naše odpovede a vytvoríme maticu, v programe nazvanú *ScoreMatrix_Stupen*, ktorá má štyri stĺpce zodpovedajúce štyrom nastaveniam α a počet riadkov zhodný s počtom odpovedí. Maticu vyplníme tzv. **score** hodnotami, čo sú súčty poradí postáv, ktoré vybral odpovedajúci ako hlavné. Čiže napríklad číslo v prvom riadku a prvom stĺpci matice hovorí, aký je súčet hodností (poradí) pridelených postavám z prvej odpovede, keď sa centralita stupňa počítala s $\alpha = 0$. Ukážku tejto matice môžeme vidieť na Obr. 17.

V ďalšom kroku bude potrebné určiť, aké zo štyroch porovnávaných nastavení α pri centralite stupňa najlepšie vystihuje názor ľudí na to, ktoré postavy sú vo filme hlavné. Predpokladajme, že hlavné postavy vo filme sú tie, ktoré majú najviac kontaktov alebo sa najčastejšie objavujú na scéne. Z hľadiska centrality by potom mali mať najvyššie

hodnoty, čo sa následne prejaví na vyšších pridelených poradiach a tým aj na vyššej *score* hodnote. V každom riadku matice *ScoreMatrix_Stupen* nás preto bude zaujímať, pri ktorom α je vypočítané *score* najvyššie. Potom môžeme povedať, že centralita stupňa popisuje odpoveď v danom riadku najlepšie, keď je počítaná s tým α , ktoré dáva najvyššie *score*. V niektorých riadkoch matice sa najvyššia *score* hodnota objavila vo viacerých stĺpcoch. Vtedy sa ako najlepšie α zoberú všetky nastavenia, pri ktorých vyšlo najvyššie *score*. Takýmto spôsobom vyhodnotíme celú maticu a pozrieme sa aké kombinácie nastavenia α sa vyhodnotili v riadkoch matice ako najlepšie pre odpoveď a s akými počtami. Zhrnutie môžeme vidieť v Tab. 11.

									$\alpha = 0$ $\alpha = 0.5$ $\alpha = 1$ $\alpha = 1.5$
	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$	$\alpha = 0$ $\alpha = 1$	$\alpha = 0$ $\alpha = 1.5$	$\alpha = 0.5$ $\alpha = 1.5$	$\alpha = 1$ $\alpha = 1.5$	
POČET	1	3	3	11	1	1	2	3	5

Tabuľka 11: Počty pre všetky rôzne kombinácie α , ktoré sa pre centralitu stupňa aspoň raz pri odpovediach vyhodnotili ako najlepšie.

Ukazuje sa, že najvhodnejšie zo štyroch porovnávaných nastavení α pri centralite stupňa, tak aby čo najlepšie vystihovalo odpovede je $\alpha = 1.5$, ktoré sa buď samostatne alebo v kombinácii s iným nastavením α vyhodnotilo ako najlepšie pri väčšine odpovedí.

Rovnaký postup pri hľadaní najvhodnejšieho nastavenia α sme použili aj s centralitou blízkosti a centralitou stredovej medzipolohy. Všetky kombinácie najvhodnejších α pri centralite blízkosti aj s ich počtami nájdeme v Tab. 12. Zdá sa, že v tomto prípade má $\alpha = 1.5$ tiež najväčšie zastúpenie, aj keď rozdiel s $\alpha = 1$ nie je až taký markantný ako pri centralite stupňa.

						$\alpha = 0$ $\alpha = 0.5$ $\alpha = 1$ $\alpha = 1.5$
	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$	$\alpha = 1$ $\alpha = 1.5$	$\alpha = 0.5$ $\alpha = 1$ $\alpha = 1.5$	
POČET	2	8	11	3	1	5

Tabuľka 12: Počty pre všetky rôzne kombinácie α , ktoré sa pre centralitu blízkosti aspoň raz pri odpovediach vyhodnotili ako najlepšie.

Čo sa týka centrality stredovej medzipolohy, postup pri spracovaní dát bol taktiež identický ako pri centralite stupňa. Výsledky vidíme v Tab. 13. V tomto prípade sa ako najvhodnejšie nastavenie pre α javí $\alpha = 0$. Pri všetkých troch typoch centrality sa ešte

	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1.5$	$\alpha = 1$ $\alpha = 1.5$
POČET	22	1	6	1

Tabuľka 13: Počty pre všetky rôzne kombinácie α , ktoré sa pre centralitu stredovej medzipolohy aspoň raz pri odpovediach vyhodnotili ako najlepšie.

pozrieme aj na celkové početnosti pre každé α , čiže koľkokrát sa ktoré α vyhodnotilo ako najlepšie pri danom type centrality, bez ohľadu na to, či samostatne alebo v kombinácii s iným α . Celkové zhrnutie nájdeme v Tab. 14.

	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$
Centralita Stupňa	8	10	12	22
Centralita Blízkości	5	8	17	20
Centralita stredovej medzipolohy	22	1	1	7

Tabuľka 14: Celkové počty všetkých skúmaných nastavení α pri všetkých troch centralitách - koľkokrát sa ktoré α vyhodnotilo ako najlepšie bez ohľadu na to či samostatne alebo v kombinácii s iným.

Aj z tejto tabuľky jasne vidieť, že nastavenie $\alpha = 1.5$ jednoznačne dominuje pri centralite stupňa a centralite blízkosti. Pre centralitu stredovej medzipolohy je zjavné, že s výrazným náskokom sa najčastejšie vyhodnotilo $\alpha = 0$ ako to nastavenie, pri ktorom sa objavuje najväčšie *score*, čiže dáva najlepšie zhody s odpoveďami.

Teraz, keď už vieme, ktoré α je pri každej miere centrality najvhodnejšie aby sa čo najviac zhodovali s názorom odpovedajúcich, porovnáme ešte tieto tri miery centrality medzi sebou. Použijeme teda centralitu stupňa počítanú s nastavením $\alpha = 1.5$, centralitu blízkosti taktiež s $\alpha = 1.5$ a centralitu stredovej medzipolohy počítanú s $\alpha = 0$. Postup pri porovnávaní je veľmi podobný tomu, ktorý sme robili pri hľadaní najlepšieho α . Vytvoríme opäť "*score*" maticu v programe nazvanú **ScoreMat**. Tentokrát budú stĺpce zodpovedať konkrétnej miere, čiže napríklad číslo v prvom riadku a prvom stĺpci je *score* priradené prvej odpovedi vypočítané pri centralite stupňa s

nastavením $\alpha = 1.5$. Ukážku tejto matice vidíme na Obr. 18.

```
> ScoreMat
      C.stupna(alpha=1.5) C.Blízkosti(alpha=1.5) C.Medzipolohy(alpha=0)
odpoved_1      193.0      196.0      195.0
odpoved_2      182.0      186.0      176.5
odpoved_3      160.0      163.0      141.0
odpoved_4      194.0      197.0      191.0
odpoved_5      167.5      171.0      164.0
odpoved_6      207.0      207.0      202.0
odpoved_7      193.0      198.0      183.0
odpoved_8      169.5      174.0      179.0
odpoved_9      205.0      204.0      205.0
odpoved_10     154.0      161.0      162.0
```

Obr. 18: ScoreMat: matica *score* hodnôt. Prvý stĺpec obsahuje *score* vypočítané z poradia pri centralite stupňa s $\alpha = 1.5$, druhý stĺpec obsahuje *score* pri centralite blízkosti s $\alpha = 1.5$ a tretí stĺpec pri centralite stredovej medzipolohy s $\alpha = 0$.

Celkové vyhodnotenie už potom prebieha identicky ako predtým. Pre každý riadok matice sa zistí, ktorá z týchto troch centralít dáva najvyššie *score* a teda najlepšie vystihuje danú odpoveď. Nakoniec sa určia počty pre každú kombináciu centralít, ktoré sa pri vyhodnocovaní objavili ako najlepšie. Zaznamenané sú v Tab. 15.

	Cen. Stupňa ($\alpha = 1.5$)	Cen. Blízkosti ($\alpha = 1.5$)	Cen. Medzipolohy ($\alpha = 0$)	Cen. Stupňa ($\alpha = 1.5$) Cen. Blízkosti ($\alpha = 1.5$)	Cen. Stupňa ($\alpha = 1.5$) Cen. Medzipolohy ($\alpha = 0$)	Cen. Blízkosti ($\alpha = 1.5$) Cen. Medzipolohy ($\alpha = 0$)
POČET	2	19	2	5	1	1

Tabuľka 15: Počty pre každú kombináciu centralít, ktoré sa pri celkovom vyhodnocovaní objavili ako najlepšie.

V Tab. 16 sa nachádza celkové zastúpenie každej centrality s najlepším α . Tabuľka hovorí, koľkokrát sa centralita vyhodnotila ako najlepšia v porovnaní so zvyšnými dvoma, bez ohľadu na to, či samostatne alebo spolu s inou centralitou.

Vidíme, že centralita blízkosti má výrazne vyššie zastúpenie ako zvyšné dve centrality. Z celkového hľadiska tak prichádzame k záveru, že centralita blízkosti počítaná s nastavením $\alpha = 1.5$ sa ukazuje ako tá, ktorá najlepšie vystihuje subjektívny názor ľudí na dôležitosť postavy vo filme. Pre úplnosť treba ešte dodať, že v odpovediach na prvú otázku odpovedajúci mal iba vybrať skupinu postáv, ktoré pokladal za dôležité

	Cen. Stupňa ($\alpha = 1.5$)	Cen. Blízosti ($\alpha = 1.5$)	Cen. Medipolohy ($\alpha = 0$)
POČET	8	25	4

Tabuľka 16: Celkové počty, koľkokrát sa každá z troch centralít - stupňa s $\alpha = 1.5$, blízkosti s $\alpha = 1.5$ a stredovej medzipolohy s $\alpha = 0$ - vyhodnotila ako tá, čo dáva najlepšiu zhodu s odpoveďou.

pričom nemal možnosť pri nich určiť poradie dôležitosti pre dej. Preto sme pracovali s vybratými menami ako so skupinou postáv, ktoré boli pre odpovedajúceho rovnocenné. V otázke, ktorú budeme analyzovať v ďalšej časti takáto možnosť bola, bude preto potrebné pristupovať k dátam spôsobom, v ktorom sa stupeň dôležitosti zohľadní .

4.2.2 Priradenie poradia postavám

V druhej z položených otázok bolo úlohou odpovedajúceho na základe vlastnej intuície alebo subjektívneho vnímania určiť poradie od 1 do 9 predloženým deviatim postavám, pričom hodnotu 1 dostala najdôležitejšia postava a hodnotu 9 najmenej dôležitá z nich. Postavy, ku ktorým mali odpovedajúci vytvoriť poradie podľa dôležitosti, sme vybrali v súlade s výsledkami našich výpočtov centralít. Pri použití rôznych mier centrality a následnom zoradení postáv podľa vypočítaných hodnôt sme si všimli, že mená postáv na prvých miestach v usporiadaní sa opakujú, mierne sa líši iba ich poradie. Vybrali sme preto práve tieto postavy: POE, FINN, HAN SOLO, CHEWBACCA, BB-8, KYLO REN, REY, LEIA a C-3PO. Každému odpovedajúcemu sa mená zobrazovali v náhodnom poradí. Aj v tomto prípade sme museli niektoré odpovede vylúčiť. Dôvodom bolo, že odpovedajúci, pravdepodobne z nepozornosti alebo kvôli zle pochopenému zadaniu, určil viacerým postavám rovnaké hodnotenie. Z pôvodných 37 nám tak ostalo 31 odpovedí, ktoré sú uložené v súbore *UsporiadaniePostav.txt*. Ukážku z týchto dát vidíme na Obr. 19. Každý riadok v súbore predstavuje jednu odpoveď a čísla v riadku znamenajú aké poradie priradil odpovedajúci danej postave. Na prácu s dátami sme použili program R a kód k tejto časti práce sa nachádza v súbore *kap4_2_2_Usporiadanie_postav.R*. Postup pri spracovaní dát vysvetlíme aj v tomto prípade iba na centralite stupňa, keďže pri zvyšných dvoch centralitách je rovnaký, na

#POE,	FINN,	HAN	SOLO,	CHEWBACCA,	BB-8,	KYLO REN,	REY,	LEIA,	C-3PO
5	2	6	7	4	3	1	9	8	
5	3	6	8	4	2	1	7	9	
4	8	5	6	3	2	1	7	9	
3	2	5	7	4	8	1	6	9	
5	1	3	7	4	6	2	8	9	
6	2	4	9	3	5	1	8	7	
9	2	3	4	5	6	1	8	7	
6	1	3	5	8	2	4	7	9	
6	4	2	5	9	3	8	1	7	
6	5	3	8	4	2	1	7	9	

Obr. 19: Ukážka dát: Pridelenie hodnotenia postavám na základe vlastného vnímania ich dôležitosti pre dej filmu.

začiatku sa iba zmení druh centrality, s ktorou robíme následné porovnania.

Tak ako sme to robili v kapitole 4.2.1, potrebujeme najskôr vytvoriť váženú sociálnu sieť z dát [13], na ktorej sa potom vypočítajú centrality stupňa s nastavenými hodnotami α postupne na 0, 0.5, 1, 1.5. S použitím príkazu `degree_w()` (v prípade centrality blízkosti `closeness_w()` a centrality stredovej medzipolohy `betweenness_w()`) a príslušnou hodnotou ladiaceho parametra α zadanou priamo v príkaze vypočítame centrality stupňa všetkých uzlov siete pri každom nastavení. Dostávame štyri zoznamy s menami postáv a hodnotami ich centrality pri danom α . V kóde sa nachádzajú pod názvami `Stupen_0`, `Stupen_0.5`, `Stupen_1` a `Stupen_1.5`. Zo zoznamov vyberieme iba tých deväť mien, ktoré boli predložené odpovedajúcim, a podľa hodnoty centrality im určíme poradie tak, aby ten s najvyššou centralitou dostal hodnotenie 1 a ten s najnižšou centralitou hodnotenie 9. Keďže ideme porovnávať poradie z výpočtu s poradiami z dotazníka, je dôležité myslieť aj na to, že niekedy sa hodnoty centrality viacerých uzlov zhodujú. V takom prípade, všetky poradie, kde si tieto uzly vymenia miesta sú ekvivalentné. Pre našich 9 postáv takáto situácia nastáva pri centralite stupňa s $\alpha = 0$, centralite blízkosti s $\alpha = 0$ a centralite stredovej medzipolohy s $\alpha = 0$ a $\alpha = 0.5$. Na tento účel sa v programe vytvoria matice permutácií poradí, ktoré sú ekvivalentné a všetky tieto poradie sa porovnávajú s odpoveďou, pričom na konci sa vyberie to poradie, ktoré sa zhoduje s odpoveďou najviac.

Teraz môžeme pristúpiť k samotnému porovnávaniu poradí. Využijeme na to štyri

metódy, ktoré sme opísali na začiatku kapitoly. Prvé dve metódy, Jarova - Winklerova podobnosť a Levenshteinova vzdialenosť, porovnávajú podobnosť resp. rozdielnosť dvoch reťazcov. Ak sa teda na poradia z výpočtov a z odpovedí pozeráme ako na reťazce, tieto dve metódy sú vhodné na ich porovnanie. V programe R sú na to implementované funkcie `jarowinkler(str1, str2)` a `levenshteinSim(str1, str2)`. Tie počítajú mieru podobnosti v intervale $\langle 0, 1 \rangle$ pre reťazce `str1` a `str2`, kde 1 znamená najväčší stupeň podobnosti (zvyčajne úplnú zhodu) a 0 najmenší stupeň podobnosti. Funkcia `jarowinkler(str1, str2)` dáva Jarovu - Winklerovu podobnosť, kým funkcia `levenshteinSim(str1, str2)` vracia hodnotu $1 - \frac{LD}{\max(|str1|, |str2|)}$, kde LD je Levenshteinova vzdialenosť medzi reťazcami `str1`, `str2` definovaná v kapitole 4.1.2 a $|str1|$, $|str2|$ sú dĺžky týchto reťazcov. S použitím týchto dvoch funkcií vytvoríme dve matice zobrazené na Obr. 20 a Obr. 21, v ktorých jeden riadok vyjadruje ako sa poradie z dotazníkovej odpovede podobá na poradia vytvorené centralitou stupňa pri jednotlivých nastaveniach α . Napríklad čísla v prvom riadku matice hovoria, ako sa poradie z prvej odpovede podobá na poradia vytvorené podľa centrality stupňa s $\alpha = 0$, centrality stupňa s $\alpha = 0.5$, centrality stupňa s $\alpha = 1$ a centrality stupňa s $\alpha = 1.5$. Platí pritom, čím je číslo bližšie k hodnote 1, tým je podobnosť s touto centralitou väčšia. Zvyšné dve použité metódy, Kendallov a Spearmanov korelačný koeficient, určujú mieru podobnosti medzi dvoma množinami poradí pridelených rovnakej skupine objektov. Vychádzajú z myšlienky, že ak sú poradia veličín priradených tej istej skupine objektov veľmi podobné, svedčí to o istej závislosti medzi nimi. Pri porovnaní poradí cez Kendallov a Spearmanov korelačný koeficient dostávame dve korelačné matice, kde čísla v riadku vyjadrujú mieru poradovej korelácie medzi odpoveďou a poradiami z centrality stupňa pri nastaveniach $\alpha = 0$, $\alpha = 0.5$, $\alpha = 1$ a $\alpha = 1.5$. Ukážku týchto matíc vidíme na Obr. 22 a Obr. 23.

Posledným krokom je vyhodnotenie podobností. Najskôr sa na každú zo vzniknutých matíc pozeráme samostatne. Pre každý riadok matice sa vyhodnotí, na ktorú centralitu sa odpoveď podobá najviac, inak povedané, pri ktorej α je podobnosť centrality stupňa s odpoveďou najväčšia. V niektorých riadkoch matice sa môžu nachádzať dve, tri alebo aj štyri rovnaké čísla, čo znamená, že poradie z odpovede je rovnako podobné s poradiami vytvorenými pri rôznych nastaveniach α . V takom prípade sa ako najlepšie

```
> Jar_Stupen
      α = 0      α = 0.5      α = 1      α = 1.5
odp_1 0.8009259 0.6111111 0.6666667 0.6666667
odp_2 0.8009259 0.7566138 0.7222222 0.7222222
odp_3 0.8009259 0.6111111 0.6666667 0.6666667
odp_4 0.8009259 0.7566138 0.7222222 0.7222222
odp_5 0.7566138 0.6370370 0.7566138 0.7566138
odp_6 0.8842593 0.6370370 0.6370370 0.6370370
```

Obr. 20: Ukážka matice podobnosti poradí vytvorenej Jarovou - Winklerovou metódou porovnávania reťazcov: na škále $\langle 0, 1 \rangle$, kde 1 znamená zhodu a 0 úplnú rozdielnosť, vyjadruje podobnosť poradí z odpovede a z výpočtu centrality stupňa s príslušným α .

```
> Lev_Stupen
      α = 0      α = 0.5      α = 1      α = 1.5
odp_1 0.4444444 0.1111111 0.2222222 0.2222222
odp_2 0.4444444 0.3333333 0.2222222 0.2222222
odp_3 0.4444444 0.1111111 0.2222222 0.2222222
odp_4 0.4444444 0.3333333 0.2222222 0.2222222
odp_5 0.2222222 0.2222222 0.2222222 0.2222222
odp_6 0.6666667 0.3333333 0.2222222 0.2222222
```

Obr. 21: Ukážka matice podobnosti poradí vytvorenej Levenshteinovou metódou porovnávania reťazcov: na škále $\langle 0, 1 \rangle$, kde 1 znamená zhodu a 0 úplnú rozdielnosť, vyjadruje podobnosť poradí z odpovede a z výpočtu centrality stupňa s príslušným α .

```
> kendall_Stupen
      α = 0      α = 0.5      α = 1      α = 1.5
odp_1 0.3333333 0.2222222 0.3333333 0.3333333
odp_2 0.2222222 0.1666667 0.2777778 0.2777778
odp_3 0.3333333 0.2222222 0.3333333 0.3333333
odp_4 0.2222222 0.1666667 0.2777778 0.2777778
odp_5 0.1666667 -0.0555556 0.0555556 0.0555556
odp_6 0.5555556 0.4444444 0.5555556 0.5555556
```

Obr. 22: Ukážka matice korelácií poradí vytvorenej pomocou odhadu Kendallovho korelačného koeficientu.

```
> Spearman_Stupen
      α = 0      α = 0.5      α = 1      α = 1.5
odp_1 0.4166667 0.3333333 0.4833333 0.4833333
odp_2 0.2833333 0.1833333 0.3500000 0.3500000
odp_3 0.4166667 0.3333333 0.4833333 0.4833333
odp_4 0.2833333 0.1833333 0.3500000 0.3500000
odp_5 0.2000000 -0.0333333 0.0833333 0.0833333
odp_6 0.6833333 0.5666667 0.7166667 0.7166667
```

Obr. 23: Ukážka matice korelácií poradí vytvorenej pomocou odhadu Spearmanovho korelačného koeficientu.

vyhodnotia všetky, pri ktorých je podobnosť najvyššia. Takýmto spôsobom dostaneme z matíc štyri zoznamy, v ktorých je pri každej dotazníkovej odpovedi uvedené, na ktorú z centralít stupňa sa odpoveď podobá najviac. V programe sú tieto zoznamy nazvané

max.m1 , max.m2 , max.m3 a max.m4 . Nakoľko nie je dôvod predpokladať, že by niektorá z použitých metód dávala presnejšie výsledky ako iná, považujeme všetky štyri za rovnocenné. Zoznamy, ktoré vznikli z rôznych metód tak môžeme spojiť do jedného, v ktorom pri každej odpovedi bude uvedené, aké α vyhodnotili jednotlivé metódy ako najlepšie. Následne ho vyhodnocujeme ako celok tak, že pri každej odpovedi sa pozeráme, ktorá α sa pri danej odpovedi objavuje najčastejšie ako tá, čo dáva najlepšiu zhodu. Ak je početnosť viacerých α pri jednej odpovedi rovnaká a vyššia ako zvyšné, všetky tieto α sa celkovo pri danej odpovedi vyhodnotia ako najlepšie. Nakoniec sa pozrieme aké kombinácie α a s akými počtami v celom zozname nastali. Zhrnutie môžeme vidieť v Tab. 17.

Cent. stupňa		$\alpha = 0$	$\alpha = 1$	$\alpha = 0$ $\alpha = 1$	$\alpha = 0.5$ $\alpha = 1$
	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1.5$	$\alpha = 1.5$	$\alpha = 1.5$
POČET	16	1	5	8	1

Tabuľka 17: Počty rôznych kombinácií α , ktoré sa pri centralite stupňa vyhodnotili ako tie čo dávajú najväčšiu zhodu s odpoveďami.

Po aplikácii rovnakého postupu aj na centralitu blízkosti a centralitu stredovej medzipolohy sa dopracujeme k Tab. 18 resp. Tab. 19, ktoré sumarizujú, aké kombinácie α a koľkokrát sa vyhodnotili pri takomto postupe ako najlepšie.

Cent. blízkosti			$\alpha = 0$	$\alpha = 1$	$\alpha = 0$ $\alpha = 1$
	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 0.5$	$\alpha = 1.5$	$\alpha = 1.5$
POČET	3	19	5	2	2

Tabuľka 18: Počty rôznych kombinácií α , ktoré sa pri centralite blízkosti vyhodnotili ako tie čo dávajú najväčšiu zhodu s odpoveďami.

Cent. stredovej medzipolohy					$\alpha = 0$	$\alpha = 0$	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$
	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$	$\alpha = 1.5$	$\alpha = 1.5$
POČET	4	1	3	12	2	1	2	1	5

Tabuľka 19: Počty rôznych kombinácií α , ktoré sa pri centralite stredovej medzipolohy vyhodnotili ako tie čo dávajú najväčšiu zhodu s odpoveďami.

	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$
Stupeň	25	2	14	14
Blízkosť	10	24	4	4
Medzipoloha	9	4	9	20

Tabuľka 20: Celkový počet prípadov, kedy sa dané α pri každej centralite vyhodnotilo ako to, ktoré vytvorilo poradie najlepšie sa zhodujúce s odpoveďou.

Celkové počty pre všetky hodnoty α pri všetkých mierach centrality, čiže počty koľkokrát sa konkrétne α pri každom type centrality vyhodnotilo ako najlepšie, uvádzame v Tab. 20. Z nej sa dá ľahko vidieť, že pre centralitu stupňa sa ako najlepšie ukazuje $\alpha = 0$, pre centralitu blízkosti $\alpha = 0.5$ a pre centralitu stredovej medzipolohy $\alpha = 1.5$. Zároveň vidíme, že každé α vyhodnotené ako najlepšie má pri svojej centralite výrazne vyššie zastúpenie ako zvyšné tri.

```
> Jar_Celkovo
```

	c. Stupňa($\alpha=0$)	c. Blízkosti($\alpha=0.5$)	c. Medzipolohy($\alpha=1.5$)
odp_1	0.8009259	0.7566138	0.6370370
odp_2	0.8009259	0.7566138	0.6666667
odp_3	0.8009259	0.7566138	0.6370370
odp_4	0.8009259	0.7566138	0.6666667
odp_5	0.7566138	0.6666667	0.7089947

Obr. 24: Ukážka: matica s Jaro - Winklerovými mierami podobnosti poradia z odpovede a toho, ktoré bolo vytvorené centralitami.

```
> Lev_Celkovo
```

	c. Stupňa($\alpha=0$)	c. Blízkosti($\alpha=0.5$)	c. Medzipolohy($\alpha=1.5$)
odp_1	0.4444444	0.2222222	0.2222222
odp_2	0.4444444	0.3333333	0.3333333
odp_3	0.4444444	0.2222222	0.2222222
odp_4	0.4444444	0.3333333	0.3333333
odp_5	0.2222222	0.2222222	0.3333333

Obr. 25: Ukážka: matica s Levenshteinovými mierami podobnosti poradia z odpovede a toho, ktoré bolo vytvorené centralitami.

S týmito zisteniami môžeme prejsť k záverečnému kroku. Porovnáme tri miery, ktoré vyšli ako najlepšie, aj vzájomne medzi sebou. Spôsob ich porovnávania bude rovnaký ako sme opísali pri centralite stupňa, teraz však namiesto najlepšieho α hľadáme najlepšiu centralitu. Zoberiem centralitu stupňa s nastavením $\alpha = 0$, centralitu blízkosti s $\alpha = 0.5$ a centralitu stredovej medzipolohy s $\alpha = 1.5$. S využitím Jarovej - Winklerovej podobnosti a Levenshteinovej vzdialenosti zostrojíme dve matice, v ktorých sa nachá-

```
> kendall_celkovo
```

	C. Stupňa($\alpha=0$)	C. Blízkości($\alpha=0.5$)	C. Medzipolohy($\alpha=1.5$)
odp_1	0.33333333	0.3888889	0.61111111
odp_2	0.22222222	0.3333333	0.66666667
odp_3	0.33333333	0.3888889	0.61111111
odp_4	0.22222222	0.3333333	0.66666667
odp_5	0.16666667	0.1111111	0.33333333

Obr. 26: Ukážka - Kendallove korelácie: matica korelácií poradí z odpovede a tých, ktoré boli vytvorené centralitami.

```
> spearman_celkovo
```

	C. Stupňa($\alpha=0$)	C. Blízkości($\alpha=0.5$)	C. Medzipolohy($\alpha=1.5$)
odp_1	0.4166667	0.5333333	0.8333333
odp_2	0.2833333	0.4166667	0.8333333
odp_3	0.4166667	0.5333333	0.8333333
odp_4	0.2833333	0.4166667	0.8333333
odp_5	0.2000000	0.1333333	0.3833333

Obr. 27: Ukážka - Spearmanove korelácie: matica korelácií poradí z odpovede a tých, ktoré boli vytvorené centralitami.

dzajú hodnoty vyjadrujúce mieru podobnosti medzi poradiami z odpovedí a poradiami vytvorenými týmito tromi centralitami (Obr. 24, Obr. 25). Matice tak majú tri stĺpce zodpovedajúce trom rôznym centralitám a počet riadkov zodpovedajúci počtu odpovedí. Rovnakým spôsobom sa zostroja aj dve korelačné matice, v ktorých sa ale nachádzajú poradové korelácie medzi odpoveďami a tromi vybratými centralitami (Obr. 26, Obr. 27).

Následný postup vyhodnotenia najlepšej centrality je už identický ako sme opísali pri centralite stupňa. V každej matici v každom riadku sa vyhodnotí najlepšia centralita, čím dostaneme štyri zoznamy. Tie spojíme do jedného a pozrieme sa pri každej odpovedi, ktorá centralita sa objavuje najčastejšie. Keď sú počty dvoch alebo aj všetkých troch centralít rovnaké, ako najlepšie sa vyhodnotia obidve, resp. všetky tri centrality. Celkové výsledky s počtami všetkých kombinácií centralít sú zaznamenané v Tab. 21 a počty pre každú centralitu samostatne v Tab. 22. Vidíme, že jednoznačne najčastejšie sa v zozname určila centralita stupňa s nastavením $\alpha = 0$ ako najvhodnejšia. Takýto výsledok hovorí, že pri ohodnocovaní hlavných postáv na základe filmu, ľudia pokladali veľký počet kontaktov za najdôležitejšie kritérium.

	Cen. Stupňa ($\alpha = 0$)	Cen. Blízkości ($\alpha = 0.5$)	Cen. Medzipol. ($\alpha = 1.5$)	Cen. Stupňa ($\alpha = 0$) Cen. Blízkości ($\alpha = 0.5$)	Cen. Stupňa ($\alpha = 0$) Cen. Medzipol. ($\alpha = 1.5$)	Cen. Blízkości ($\alpha = 0.5$) Cen. Medzipol. ($\alpha = 1.5$)	Cen. Stupňa ($\alpha = 0$) Cen. Blízkości ($\alpha = 0.5$) Cen. Medzipol. ($\alpha = 1.5$)
POČET	7	1	3	8	10	1	1

Tabuľka 21: Centrality dávajúce najlepšie zhody s poradiami z odpovedí a k nim ich početnosť.

	Cent. Stupňa ($\alpha = 0$)	Cent. Blízkości ($\alpha = 0.5$)	Cent. Medzipolohy ($\alpha = 1.5$)
POČET	26	11	15

Tabuľka 22: Koľkokrát sa ktorá centralita vyhodnotila ako centralita najlepšej zhody v poradiach s odpoveďami.

4.3 Vnímanie centrálnosti uzlov v zobrazenej sieti

V druhej časti dotazníka boli otázky zamerané najmä na subjektívne vnímanie centrality uzlov v sociálnej sieti na základe vizualizácie siete. Úlohou bolo z predloženého grafického znázornenia Star Wars sociálnej siete určiť, ktoré uzly siete sa odpovedajúcemu zdajú najviac centrálné. Hlavnú časť v druhej polovici dotazníka tvorili nasledovné úlohy

1. Na základe vizualizácie sociálnej siete sa pokúste určiť, ktorých 7 uzlov je v tejto sieti najviac centrálnych (t. j. sú v sieti dôležité, čiže majú najvyššiu centralitu).
2. Vami vybrané uzly usporiadajte od (podľa Vás) najviac centrálnych uzlov (prvé miesto) po najmenej centrálné (posledné siedme miesto).

Pre potreby dotazníka bola vytvorená aj stránka, na ktorej si odpovedajúci mali možnosť pozrieť vizualizácie celej siete s viacerými rozmiestneniami uzlov a aj vizualizácie jednotlivých uzlov samostatne.

Celkovo sme z dotazníka získali 37 odpovedí, z ktorých sme pri predošlých otázkach museli niektoré vylúčiť, keďže odpovedajúci buď nespĺňali podmienku, že videli film alebo pre nesprávne pochopené zadanie odpovedali nevhodným spôsobom. Pri druhej časti dotazníka sme tieto odpovede z analýzy vylúčiť nemuseli, nakoľko otázky tu

#1. miesto	2. miesto	3. miesto	4. miesto	5. miesto	6. miesto	7. miesto
3	14	9	11	6	4	19
9	14	3	4	11	6	7
9	21	14	11	3	18	22
8	19	6	3	10	1	9
11	9	14	18	3	15	21
9	3	14	11	6	18	8
6	9	14	7	4	3	18
9	3	11	14	4	6	19
3	4	9	11	14	6	18
9	14	11	3	6	4	19

Obr. 28: Ukážka dát: Výber a usporiadanie uzlov podľa vizualizácie siete.

neboli viazané priamo na dej filmu. Avšak aj v tomto prípade prišli dve odpovede, ktoré nebolo možné použiť lebo odpovedajúci v nich vybral menší počet uzlov ako bolo potrebné. V dátach je tak zaznamenaných iba 35 odpovedí. Uložené sú v súbore *UsporiadanieUzlov.txt*. Každý riadok súboru predstavuje jedného odpovedajúceho, ktorý vybral sedem uzlov sociálnej siete, o ktorých si myslel, že majú významné postavenie v sieti. Zároveň tieto uzly aj usporiadal podľa svojho subjektívneho názoru od najviac centrálnych po menej centrálnu. Čiže v každom riadku sa nachádzajú čísla siedmych, podľa respondenta najcentrálnejších uzlov, pričom tieto sú na základe jeho subjektívneho vnímania zoradené od najviac po najmenej centrálnu. Ukážku z týchto dát vidíme na Obr. 28

Na spracovanie dát sme opäť použili program R a zdrojové kódy k ich spracovaniu sú uložené v súboroch *Najlepsie_alpha_stupen.R*, *Najlepsie_alpha_Blizkost.R*, *Najlepsie_alpha_Medzipoloha.R* a *Najlepsia_centralita_celkovo.R*. V tomto prípade sa na dáta budeme pozeráť z dvoch uhlov pohľadu. Najprv skupinu vybraných uzlov v odpovedi budeme brať ako celok a ich poradie budeme ignorovať. Tu použijeme rovnakú *score* metódu ako sme použili v kapitole 4.2.1. Následne sa zameriame aj na poradie, ktoré bolo v odpovedi uzlom priradené a porovnáme ho s poradím, ktoré určili výpočty. Na to sa využijú rovnaké štyri metódy porovnania poradia aké sme použili v kapitole 4.2.2. Obidva tieto prístupy nakoniec spojíme a dáta vyhodnotíme ako celok.

```

> ScoreMatrix_Stupen
      α = 0      α = 0.5      α = 1      α = 1.5
odp_1    166.0    168.0    168.0    168.0
odp_2    166.0    167.0    167.0    167.0
odp_3    147.5    148.0    145.5    143.0
odp_4    118.0    117.0    115.0    115.0
odp_5    144.0    143.0    139.0    136.0
odp_6    158.5    159.0    159.0    159.0

```

Obr. 29: Ukážka `ScoreMatrix_Stupen`: matica *score* hodnôt.

4.3.1 Vyhodnotenie metódou sčítania poradí

Postup pri spracovaní vysvetíme iba s jedným zdrojovým kódom, kde sa hľadá najlepšia α hodnota k centralite stupňa. Zvyšné sú koncipované rovnako, s tým rozdielom, že sa v nich pracuje s centralitami blízkosti a stredovej medzipolohy. Na začiatku programu sa ako vždy vytvorí vážená sociálna sieť z verejne dostupných dát [13] a s použitím príkazu `degree_w()` (v prípade centrality blízkosti `closeness_w()` a centrality stredovej medzipolohy `betweenness_w()`) sa vypočítajú centrality stupňa všetkých uzlov siete. Parameter α , ktorý určuje ako sa majú pri výpočte zohľadniť váhy, sa nastaví priamo v príkazoch. Opäť použijeme štyri hodnoty $\alpha = 0$, $\alpha = 0.5$, $\alpha = 1$, $\alpha = 1.5$. Zoradenie uzlov podľa ich hodnoty centrality stupňa zabezpečí príkaz `rank()`, ktorý určí každému uzlu poradové miesto v usporiadaní. Ak má viac uzlov rovnakú hodnotu centrality, všetkým sa priradí priemerné poradie vypočítané z poradových miest, ktoré by obsadili ak by nasledovali za sebou.

V ďalšej časti kódu sa už pracuje s odpoveďami z dotazníka. Vytvoríme maticu `ScoreMatrix_Stupen`, ktorá má štyri stĺpce zodpovedajúce štyrom nastaveniam α a počet riadkov zhodný s počtom odpovedí. Vyplníme ju *score* hodnotami podobne ako sme to robili v kapitole 4.2.1. Pre uzly uvedené v odpovedi zoberieme poradia, ktoré im boli určené podľa centrality stupňa pri jednotlivých nastaveniach α . Keď tieto poradia sčítame, dostávame *score* hodnoty pre príslušné α . Ukážku matice vidíme na Obr. 29. Následne sa podľa najvyššej *score* hodnoty v každom riadku matice určí, pri ktorej α dáva centralita stupňa najlepšiu zhodu s odpoveďou. Keď sa najvyššia *score* hodnota nachádza pri viacerých α , tieto všetky sa vyhodnotia spolu ako ekvivalentne najlepšie pre danú odpoveď. Dostaneme tak zoznam (v priloženom kóde má

názov `Zoznam_zo_score_metody`), v ktorom je ku každej odpovedi pripojená informácia o týchto najlepších hodnotách α pre danú odpoveď. Nakoniec sa zosumarizuje s akými početnosťami sa vyskytli rôzne prípady ekvivalentných nastavení α vyhodnotených *score* metódou ako najlepšie nastavenie pre centralitu stupňa aby sa čo najviac zhodovala s odpoveďami. Tento sumár pre centralitu stupňa nájdeme v Tab. 23.

					$\alpha = 0$	$\alpha = 0.5$
			$\alpha = 0$	$\alpha = 1$	$\alpha = 1$	$\alpha = 1$
	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$	$\alpha = 1.5$	$\alpha = 1.5$
POČET	9	3	1	4	15	3

Tabuľka 23: Početnosti výskytu rôznych prípadov ekvivalentných nastavení α vyhodnotených *score* metódou ako najlepšie nastavenie pre centralitu stupňa, aby sa čo najviac zhodovala s odpoveďami.

4.3.2 Vyhodnotenie metódami porovnávania poradí

V prostrednej časti programu *Najlepsie_alpha_stupen.R* sa využívajú na porovnanie výpočtov s odpoveďami metódy porovnania poradí z kapitol 4.1.1, 4.1.2, 4.1.3 a 4.1.4. Túto časť tvorí `for{}` cyklus, v ktorom sa vytvárajú štyri matice vyjadrujúce mieru zhody medzi poradiami z odpovede a poradiami z výpočtov. V každom kroku cyklu sa načíta jedna odpoveď a uzly z odpovede sa usporiadajú podľa hodnoty centrality stupňa od najviac po najmenej centrálne, a to pri všetkých nastaveniach α . Usporiadanie uzlov z odpovede vytvorené na základe subjektívneho názoru odpovedajúceho sa potom porovná s usporiadaním rovnakých uzlov, ktoré bolo vytvorené podľa vypočítaných hodnôt ich centrality stupňa. Zároveň sa v každom kroku myslí aj na prípady, kedy existuje viacero ekvivalentných usporiadaní uzlov podľa centrality, čiže tie, keď dva alebo viac uzlov vybratých odpovedajúcim má rovnakú hodnotu centrality. Vtedy sa vyberá poradie, ktoré dáva najväčšiu zhodu s poradím odpovedajúceho. Porovnania sa robia všetkými štyrmi metódami, ktoré sme opísali na začiatku kapitoly, z čoho potom vznikajú spomínané štyri matice vyjadrujúce podobnosť odpovede s výpočtom. Každá matica obsahuje štyri stĺpce prislúchajúce štyrom nastaveniam α a 35 riadkov, čo je počet odpovedí. Čím sú hodnoty v maticiach bližšie k 1, tým je zhoda medzi odpoveďou a centralitou s daným α väčšia. Ďalší postup je už totožný s tým, ktorý sme

použili v kapitole 4.2.2. Na jeho konci dostávame zoznam, kde je ku každej odpovedi priložená informácia o tom, pri ktorých so štyroch porovnávaných nastavení α vykazuje centralita stupňa najväčšiu zhodu s touto odpoveďou. V programe je tento zoznam nazvaný `zoznam_z_porov_retazcov`. Zhrnutie so všetkými kombináciami α , ktoré sa v zozname vyskytli ako ekvivalentne najlepšie nastavenia pre centralitu stupňa, aby sa čo najviac zhodovala v poradí s odpoveďou, aj s počtami, koľkokrát sa v zozname tieto kombinácie nachádzajú, nájdeme v Tab. 24.

					$\alpha = 0$	$\alpha = 0.5$	$\alpha = 0$
			$\alpha = 0$	$\alpha = 1$	$\alpha = 1$	$\alpha = 1$	$\alpha = 0.5$
	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 0.5$	$\alpha = 1.5$	$\alpha = 1.5$	$\alpha = 1.5$	$\alpha = 1$
POČET	8	7	1	9	1	6	3

Tabuľka 24: Početnosti výskytu rôznych prípadov ekvivalentných nastavení α vyhodnotených metódami porovnania poradí ako najlepšie nastavenie pre centralitu stupňa, aby sa čo najviac zhodovala s odpoveďami.

4.3.3 Spojenie metód sčítania a porovnávania poradí

V poslednej časti programu `Najlepsie_alpha_stupen.R` spájame `score` metódu z prvej časti kódu s metódou porovnávania poradí z druhej časti kódu. Konkrétne je na to v programe vytvorená funkcia `Spoj_zoznamy(zoznam1, zoznam2)`, do ktorej vkladáme finálne zoznamy (`Zoznam_zo_score_metody` a `zoznam_z_porov_retazcov`) z prvých dvoch častí kódu. Zoznamy musia mať rovnaké dĺžky, pretože vo funkcii sa robia porovnania po úrovniach a hľadá sa prienik v hodnotách na týchto úrovniach. Výsledkom je nový zoznam rovnakej dĺžky, ktorý má na i -tej úrovni buď všetky hodnoty z prieniku i -tých úrovní vstupných zoznamov, alebo zjednotenie i -tých úrovní ak prienikom bola prázdna množina. Z našich dvoch zoznamov sa teda vytvorí tretí, ktorý po zahrnutí obidvoch spôsobov porovnania uvádza ku každej odpovedi z dotazníka tie nastavenia α pre centralitu stupňa, pri ktorých centralita stupňa najlepšie vystihuje danú odpoveď. Nakoniec sa zosumarizuje aké rôzne kombinácie nastavení α pre centralitu stupňa sa vyhodnotili ako najlepšie a pri koľkých odpovediach každá takáto kombinácia nastala. Konečný sumár zo spojených zoznamov pre centralitu stupňa nájdeme v Tab. 25. Keď zopakujeme rovnaký postup od `score` metódy cez metódy porovnania poradí až

po ich konečné spojenie aj pre centralitu blízkosti a centralitu stredovej medzipolohy, dostaneme na konci celkové vyhodnotenia, ktoré sú zaznamenané v Tab. 26 a Tab. 27.

				$\alpha = 0$	$\alpha = 0.5$	$\alpha = 0$
			$\alpha = 1$	$\alpha = 1$	$\alpha = 1$	$\alpha = 0.5$
	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1.5$	$\alpha = 1.5$	$\alpha = 1.5$	$\alpha = 1$
POČET	8	8	9	1	5	4

Tabuľka 25: Počty výskytu rôznych kombinácií α vyhodnotených pre centralitu stupňa po spojení dvoch prístupov ako (ekvivalentne) najlepšie vystihujúce odpoveď.

								$\alpha = 0$	$\alpha = 0.5$	$\alpha = 0$
				$\alpha = 0$	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 0.5$
	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$	$\alpha = 1.5$	$\alpha = 1$	$\alpha = 1$	$\alpha = 1$
POČET	3	8	1	1	2	1	6	5	5	3

Tabuľka 26: Počty výskytu rôznych kombinácií α vyhodnotených pre centralitu blízkosti po spojení dvoch prístupov ako (ekvivalentne) najlepšie vystihujúce odpoveď.

				$\alpha = 0$	$\alpha = 0$	$\alpha = 0$	$\alpha = 0$	$\alpha = 0$	$\alpha = 0$
	$\alpha = 0$	$\alpha = 1$	$\alpha = 1.5$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$	$\alpha = 1.5$	$\alpha = 0.5$	$\alpha = 1$
POČET	5	1	2	7	1	16	2	$\alpha = 1.5$	$\alpha = 1.5$

Tabuľka 27: Počty výskytu rôznych kombinácií α vyhodnotených pre centralitu stredovej medzipolohy po spojení dvoch prístupov ako (ekvivalentne) najlepšie vystihujúce odpoveď.

Nakoniec sa ešte pozrieme na celkový výskyt každého so štyroch nastavení parametra α , bez ohľadu či sa objavilo v zoznamoch samostatne alebo v kombinácii, pri obidvoch opísaných metódach aj po ich spojení. Pre centralitu stupňa ho nachádzame v Tab. 28, pre centralitu blízkosti v Tab. 29 a pre centralitu stredovej medzipolohy v Tab. 30. Z Tab. 28 pre centralitu stupňa vidíme, že po spojení obidvoch metód porovnania sa ako najvhodnejšie hodnoty α s najvyšším počtom výskytov vyhodnotili $\alpha = 1$ a $\alpha = 1.5$. Metódou porovnávania poradí sa tieto dve hodnoty taktiež vyhodnotili ako rovnako dobré avšak pri score metóde o niečo lepšie skončila $\alpha = 1$. Vybrali sme preto pre stupeň nastavenie $\alpha = 1$ ako najvhodnejšie. Podobne z Tab. 29 pre centralitu

blízkosti vyšlo $\alpha = 1$ ako najlepšia hodnota. Centralita stredovej medzipolohy sa podľa Tab. 30 najviac zhoduje s odpoveďami pri nastavení $\alpha = 0$.

CENTRALITA STUPŇA	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$
Počty pri score metóde	13	21	23	22
Počty pri porovnaní reťazcov	13	17	19	19
Po spojení	13	17	19	19

Tabuľka 28: Celkový výskyt každého so štyroch nastavení parametra α pre centralitu stupňa, bez ohľadu či sa objavilo v zoznamoch samostatne alebo v kombinácii, pri oboch opísaných metódach aj po ich spojení.

CENTRALITA BLÍZKOSTI	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$
Počty pri score metóde	12	16	20	18
Počty pri porovnaní reťazcov	8	20	20	20
Po spojení	14	18	22	20

Tabuľka 29: Celkový výskyt každého so štyroch nastavení parametra α pre centralitu blízkosti, bez ohľadu či sa objavilo v zoznamoch samostatne alebo v kombinácii, pri oboch opísaných metódach aj po ich spojení.

CENTRALITA STREDOVEJ MEDZIPOLOHY	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$
Počty pri score metóde	31	1	2	4
Počty pri porovnaní reťazcov	6	12	7	23
Po spojení	32	9	3	21

Tabuľka 30: Celkový výskyt každého so štyroch nastavení parametra α pre centralitu stredovej medzipolohy, bez ohľadu či sa objavilo v zoznamoch samostatne alebo v kombinácii, pri oboch opísaných metódach aj po ich spojení.

4.3.4 Porovnanie troch najlepších centralít

Pre každú z troch centralít - centralita stupňa, centralita blízkosti, centralita stredovej medzipolohy - bol v programe R vytvorený samostatný súbor s kódom. V nich sme sa vyššie opísaným postupom dopracovali k najviac vyhovujúcim hodnotám α pre všetky typy centrality, aby bola zhoda medzi subjektívnym vnímaním centrality odpovedajúcich a výpočtami centrality čo najväčšia. V poslednom kroku ešte porovnáme všetky tri typy centrality medzi sebou. Konkrétne hľadáme najlepšiu z týchto

troch: centralita stupňa s $\alpha = 1$, centralita blízkosti s $\alpha = 1$ a centralita stredovej medzipolohy s $\alpha = 0$. Na toto konečné porovnanie je v programe R vytvorený štvrtý súbor s názvom *Najlepsia_centralita_celkovo.R*. Postup a teda aj kód sú zostavené veľmi podobne ako pri hľadaní najlepšieho α . V prvej časti sa metódou *score* hodnotí vytvorí *zoznam_zo_score_metody*, v ktorom je ku každej odpovedi vybratá najlepšia z troch spomínaných centralít. Najlepšia centralita pre odpoveď sa určí podľa toho, že sa pri nej vypočíta najvyššie *score* (šúčet poradí uzlov) pre uzly z odpovede. V prípade zhodných *score* sa ako najlepšie vyhodnocujú všetky centrality, kde je *score* najvyššie. Bilanciu z tohoto zoznamu zaznamenáva Tab. 31.

	Cen. Stupňa ($\alpha = 1$)	Cen. Blízkosti ($\alpha = 1$)	Cen. Medzipolohy ($\alpha = 0$)	Cen. Stupňa ($\alpha = 1$) Cen. Blízkosti ($\alpha = 1$)	Cen. Stupňa ($\alpha = 1$) Cen. Medzipolohy ($\alpha = 0$)
POČET	10	9	2	12	2

Tabuľka 31: Bilancia zo *score* metódy pre najlepšie centrality.

V druhej časti kódu prebieha `for{}` cyklus, v ktorom sa vytvárajú dve korelačné matice a dve matice s mierami zhody pre poradia z odpovede a poradia z jednotlivých centralít. Tie sa použijú na vytvorenie druhého zoznamu *zoznam_z_porov_retazcov*, v ktorom je ku každej odpovedi vybratá najlepšia z troch centralít podľa najväčších korelácií a zhôd v poradiach. Niekedy sa pri jednej odpovedi viacero centralít vyhodnotí ako najlepšie, vtedy sú pri tejto odpovedi zaznamenané všetky. Súpis všetkých kombinácií centralít vyhodnotených spolu ako najlepšie aj s počtami koľkokrát každá nastala je v Tab. 32.

	Cen. Stupňa ($\alpha = 1$)	Cen. Blízkosti ($\alpha = 1$)	Cen. Medzipolohy ($\alpha = 0$)	Cen. Stupňa ($\alpha = 1$) Cen. Blízkosti ($\alpha = 1$)	Cen. Stupňa ($\alpha = 1$) Cen. Blízkosti ($\alpha = 1$) Cen. Medzipolohy($\alpha = 0$)
POČET	9	5	7	13	1

Tabuľka 32: Bilancia z metód porovnávaví poradi pre najlepšie centrality.

V poslednej časti kódu sa spájajú cez funkciu *Spoj_zoznamy(zoznam1, zoznam2)* dva vytvorené zoznamy - *Zoznam_zo_score_metody* a *zoznam_z_porov_retazcov* -

do tretieho, ktorý k odpovediam z dotazníka uvádza tie centrality, ktoré ich najlepšie vystihujú. Počty jednotlivých prípadov, ktoré nastali vidíme v Tab. 33.

	Cen. Stupňa ($\alpha = 1$)	Cen. Blízkości ($\alpha = 1$)	Cen. Medzipolohy ($\alpha = 0$)	Cen. Stupňa ($\alpha = 1$) Cen. Blízkości ($\alpha = 1$)	Cen. Stupňa ($\alpha = 1$) Cen. Medzipolohy ($\alpha = 0$)	Cen. Blízkości ($\alpha = 1$) Cen. Medzipolohy ($\alpha = 0$)
POČET	13	6	1	8	3	4

Tabuľka 33: Bilancia po spojení obidvoch prístupov k hľadaniu najlepšej centrality.

	Cen. Stupňa ($\alpha = 1$)	Cen. Blízkości ($\alpha = 1$)	Cen. Medzipolohy ($\alpha = 0$)
Počty pri score metóde	24	21	4
Počty pri porovnaní poradí	23	19	8
Po spojení	24	18	8

Tabuľka 34: Celkové počty pri každej centralite

Teraz už ostáva iba pozrieť sa na celkový výskyt každej centrality vo všetkých troch zoznamoch - `zoznam_zo_score_metody`, `zoznam_z_porov_retazcov` a `zoznam_vzniknutý` z ich spojenia - pričom zaujímať nás budú najmä počty v poslednom. Nájdeme ich zhrnuté v Tab. 34.

Podľa Tab. 34, z troch porovnávaných mier centrality, tak pri *score* metóde ako aj pri metódach porovnávania poradí, centralita stupňa s nastavením $\alpha = 1$ najlepšie vystihuje názor ľudí na centrálnosť uzlov v sieti. Avšak pri oboch metódach je z počtu prípadov, kedy bola táto centralita vyhodnotená ako najlepšia vidieť, že rozdiel oproti centralite blízkosti s $\alpha = 1$ je len veľmi malý. Dá sa tiež vidieť, že obidve tieto centrality pri obidvoch metódach dosiahli výrazne vyššie počty ako má centralita stredovej medzipolohy s $\alpha = 0$. Hoci centralita stredovej medzipolohy úzko súvisí s celkovým prepojením siete, zdá sa, že tento faktor pri našej Star Wars sieti vplýval na výber centrálnych uzlov len veľmi málo. Je logické, že pri identifikácii centrálnych uzlov v husto prepojenej sieti ľudia nebudú skúmať, ktoré cesty sú medzi pármí vrcholov najkratšie a koľko ich prechádza cez daný vrchol. Na druhej strane, dá sa predpokladať, že ľudia by si prepojenie siete všímali viac, keby existovali v sieti uzly, ktoré predstavujú jediné spojenie medzi dvoma samostatnými zhlukmi jednej siete. Vtedy aj intuícia hovorí, že takéto vrcholy sú pre sieť dôležité.

Nakoľko obidve metódy určili ako najvhodnejšiu centralitu stupňa s $\alpha = 1$, logicky sa dalo očakávať, že aj ich spojením sa centralita stupňa vyhodnotí ako najlepšia. Mierne sa pritom zvýšil aj jej náskok pred centralitou blízkosti pri konečných počtoch. Na základe týchto výsledkov vieme povedať, že v našej Star Wars sieti si ľudia pri identifikácii dôležitých uzlov najviac všímali ich silné väzby. Treba ešte poznamenať, že pri konečnom porovnaní troch centralít sme síce zobrali tie centrality s najlepšími alfami, ale pri ich identifikácii boli rozdiely minimálne. Preto, aby sme mohli vyvodiť záver, že zo všetkých centralít, ktorým sme sa v tejto práci venovali, centralita stupňa s nastavením $\alpha = 1$ najlepšie vystihuje subjektívny názor ľudí na centrálnosť vrcholov, by sme potrebovali ešte viac dát. Aj napriek tomu, výsledky v tejto práci jasne naznačujú, že váhy väzieb v sieti u ľudí do značnej miery vnímanie centrálnosti vrcholov ovplyvňujú.

Záver

Naša diplomová práca sa zaoberá analýzou sociálnych sietí ako nástroja na posúdenie vzťahov v sieti a vytváranie grafických znázornení siete, pričom teoretické poznatky z tejto oblasti prepája so subjektívnym vnímaním siete. Hlavným cieľom analýzy sietí zvyčajne býva identifikácia kľúčových uzlov siete za účelom využitia ich vplyvu či strategického postavenia. Niektoré takéto uzly je možné určiť už pri prvých pohľadoch na vizualizáciu siete, avšak analýza sociálnych sietí pracuje najmä s rôznymi mierami centrality, ktoré istým spôsobom vyjadrujú dôležitosť uzlov.

Na identifikáciu centrálnych uzlov bolo navrhnutých množstvo mier centrality. Len balík **CINNA** [1] implementovaný v štatistickom softvéri R ich obsahuje niekoľko desiatok a stále vznikajú nové. Každá pritom vychádza z iného pohľadu na centrálnosť uzlov. Väčšina z nich, a platí to aj pre tri Freemanove centrality (stupňa, blízkosti a stredovej medzipolohy), sa zameriava len na jedno kritérium, podľa ktorého hodnotí jednotlivé uzly. Avšak, pre mnoho aplikácií z reálneho sveta, žiadne samostatné kritérium nemusí byť adekvátne. Navrhujú sa preto metódy viackriteriálneho rozhodovania, ktoré cez vhodné kombinácie váh zahrnú do výpočtov viacero prístupov k meraniu centrálnosti, pričom sa využívajú aj známe metódy AHP, Borda či SIMOS.

Cieľom našej práce bolo nájsť takú mieru centrality, ktorá by čo najlepšie odrážala subjektívny pohľad na centralitu vrcholov viditeľnú už pri grafickom znázornení siete. V práci sme sa zamerali práve na tri najznámejšie prístupy k meraniu centrality - Freemanove miery centrality. Pracovali sme pri tom s ich pôvodnými verziami vysvetlenými v [12], s ich zovšeobecnenými verziami aj pre vážené sociálne siete [2], [21] a kombináciu týchto dvoch s použitím dodatočného parametra α , ktorá bola navrhnutá v [24].

Názor ľudí na centralitu sme zisťovali prostredníctvom anonymného dotazníka, v ktorom im boli predložené vizualizácie sociálnej siete. Na ich základe bolo potom potrebné určiť poradie dôležitosti uzlov tak, ako ho vnímali odpovedajúci. S použitím vhodných metód porovnávania poradí sme následne vyhodnocovali mieru zhody odpovede a poradia vytvoreného výpočtom podľa viacerých centralít. Celkovo sme skúmali dvanásť mier centrality - pri každom z troch druhov centrality štyri nastavenia ladiaceho parametra α . Všetky čiastkové výsledky sme zaznamenávali do prehľadných tabuliek aj s krátkymi popismi a vysvetleniami.

Podľa konečných výsledkov našej práce sa ukázalo, že subjektívny názor ľudí na dôležitosť uzlov siete najlepšie zo skúmaných mier vystihuje centralita stupňa s nastavením ladiaceho parametra $\alpha = 1$. To by znamenalo, že pri rozhodovaní o tom, či je konkrétny uzol v sieti dôležitý alebo nie, sa ľudia najviac nechávajú ovplyvniť tým, či má uzol silné väzieb. Jedným dychom ale treba tiež dodať, že rozdiely oproti centralite blízkosti s nastavením $\alpha = 1$ boli len minimálne. Preto, aby sme mohli s určitosťou potvrdiť záver, že centralita stupňa s $\alpha = 1$ je na opísanie subjektívneho vnímania ľudí zo skúmaných centralít najlepšia, potrebovali by sme k tomu viac dát a dodatočný výskum.

Zoznam použitej literatúry

- [1] Ashtiani, M.: *CINNA: Deciphering Central Informative Nodes in Network Analysis*, R package version 1.1.53, (2019)
<https://CRAN.R-project.org/package=CINNA>
- [2] Barrat, A. et al.: *The architecture of complex weighted networks*, Proceedings of the National Academy of Sciences 101 (2004), 3747-3752
- [3] Bavelas, A.: *A mathematical model for group structures*, Human Organization 7 (1948), 16–30
- [4] Christen, P.: *Data matching: concepts and techniques for record linkage, entity resolution, and duplicate detection*, Springer Science & Business Media, Canberra, 2012.
- [5] Couture-Beil, A.: *rjson: JSON for R*, R package version 0.2.20, (2018)
<https://CRAN.R-project.org/package=rjson>
- [6] Csardi, G.: *igraphdata: A Collection of Network Data Sets for the 'igraph' Package*, R package version 1.0.1, (2015),
<https://CRAN.R-project.org/package=igraphdata>
- [7] Csardi, G., Nepusz, T.: *The igraph software package for complex network research*, InterJournal, Complex Systems 1695, 2006, <http://igraph.org>
- [8] Dijkstra, E. W.: *A note on two problems in connexion with graphs*, Numerische Mathematik 1 (1959), 269-271
- [9] Divjak, B., Peharda, P.: *Social network analysis of study environment*, Journal of Information and Organizational Sciences 34 (2010), 67-80
- [10] Ferrara, E. et al.: *Detecting criminal organizations in mobile phone networks*, Expert Systems with Applications 41 (2014), 5733-5750
- [11] Finkelstein, M. O., Levin, B.: *Statistics for lawyers*, Springer, New York, 2001

- [12] Freeman, L. C.: *Centrality in social networks: Conceptual clarification*, Social Networks 1 (1978), 215-239, dostupné na internete:
- [13] Gabasova, E.: *Star Wars social network*, (2016), dáta dostupné na internete (12.01.2019): <https://zenodo.org/record/1411479#.XDyda1xKjDc>
- [14] HENDL, J.: *Přehled statistických metod: analýza a metaanalýza dat*, Páté, rozšířené vydání, Portál, Praha, 2015.
- [15] Jackson, M. O.: *Social and Economic Networks*, Princeton University Press, New Jersey, 2008
- [16] Jaro, M. A.: *Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida*, Journal of the American Statistical Association, 1989 , 84.406 : 414-420.
- [17] JSON: <https://www.json.org/json-en.html>
- [18] Krackhardt, D.: *Assessing the political landscape: Structure, cognition, and power in organizations*, Administrative science quarterly, 342-369, 1990.
- [19] *Levenshtein Distance*: <http://www.w3.org/1999/xhtml>
- [20] Newman, M. E. J.: *Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality*, Physical Review E 64, 2001, dostupné na internete:
- [21] Newman, M. E. J.: *Analysis of weighted networks*, Physical Review E 70, 2004
- [22] Nieminen, J.: *On centrality in a graph*, Scandinavian Journal of Psychology 15 (1974), 322-336.
- [23] Opsahl, T. et al.: *Prominence and control: the weighted rich-club effect*, Physical Review Letters 101, 2008
- [24] Opsahl, T., Agneessens, F., Skvoretz, J.: *Node centrality in weighted networks: Generalizing degree and shortest paths*, Social Networks (2010), 245-251
- [25] Somorčík, J.: *Neparametrická štatistika*, prednášky, FMFI UK, Bratislava, 2018

- [26] *Star Wars: Episode VII The Force Awakens* / *StarWars.com*:
<https://www.starwars.com/films/star-wars-episode-vii-the-force-awakens>
- [27] Winkler, W. E.: *String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage*, Washington, DC , 1990.
- [28] Zachary, W.: *An information flow model for conflict and fission in small groups*,
Journal of Anthropological Research 4 (1997), 452-473