

Računanje mjera centralnosti u velikim mrežama

Fabulić, Bruno

Master's thesis / Diplomski rad

2023

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:217:076208>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-11-04**



Repository / Repozitorij:

[Repository of the Faculty of Science - University of Zagreb](#)



SVEUČILIŠTE U ZAGREBU
PRIRODOSLOVNO–MATEMATIČKI FAKULTET
MATEMATIČKI ODSJEK

Bruno Fabulić

RAČUNANJE MJERA CENTRALNOSTI
U VELIKIM MREŽAMA

Diplomski rad

Voditelj rada:
prof. dr.sc. Zlatko Drmač

Zagreb, 2023.

Ovaj diplomski rad obranjen je dana _____ pred ispitnim povjerenstvom u sastavu:

1. _____, predsjednik
2. _____, član
3. _____, član

Povjerenstvo je rad ocijenilo ocjenom _____.

Potpisi članova povjerenstva:

1. _____
2. _____
3. _____

Ovaj diplomski rad posvećujem svojoj obitelji koja je uvijek vjerovala u mene i bila mi bezuvjetna podrška kroz cijelo obrazovanje. Zahvaljujem se mentoru prof. dr. sc. Zlatku Drmaču na svim uputama, savjetima, znanju kojeg mi je prenio te vremenu koje je odvojio tijekom pisanja ovog diplomskog rada. Zahvaljujem se i svim prijateljima bez kojih bi ovaj period obrazovanja bio puno teži.

Sadržaj

Sadržaj	iv
1 Uvod	1
1.1 Struktura rada	2
2 Osnovni pojmovi	3
2.1 Pojmovi iz teorije grafova	3
2.2 Računanje s matricama	5
2.2.1 Moore–Penroseov inverz	5
2.2.2 Matrična eksponencijalna funkcija	6
2.2.3 Korijen matrice	7
2.2.4 Sinus i kosinus hiperbolni matrice	7
3 Vrste centralnosti	9
3.1 <i>Page Rank</i>	9
3.2 HITS Algoritam	12
3.2.1 Primjena HITS algoritma	13
3.3 Podgrafovska centralnost	14
3.3.1 Podgrafovska centralnost za digrafe	15
3.3.2 Usporedba s HITS algoritmom	17
3.4 Totalna podgrafovska komunikativnost	18
3.4.1 Usporedba podgrafovske centralnosti i totalne podgrafovske komunikativnosti	18
3.4.2 Totalna mrežna komunikativnost	19
3.5 Katzova centralnost	21
3.5.1 Katzova centralnost u digrafima	21
4 Usporedba metoda	23
4.1 Networkx	23
4.2 Primjena na manjim mrežama	24

4.3	Primjena na skupu podataka	28
4.3.1	Opis podataka	28
4.3.2	Rezultati	30
5	Zaključak	37
	Bibliografija	38

Poglavlje 1

Uvod

Mreže se nalaze svuda u svijetu oko nas. Možemo ih pronaći u obliku karata, komunikacijskih mreža, planova letova aviona, interneta i mnogim drugim oblicima. Tu, naravno, pripadaju i društvene mreže, koje su vjerojatno većini ljudi prva asocijacija na riječ mreža. Općenito gledajući, mreže su skupovi objekata (vrhova) i poveznica (bridova) među njima. Budući da su mreže toliko rasprostranjene nameće se zaključak kako je analiza mreža vrlo važan zadatak. Analizom mreža možemo dobiti puno informacija o mrežama kao što su određivanje strukture i grupacija unutar mreža te otkrivanja raznih interakcija među članovima mreže. Navest ćemo neke stvarne primjere zašto su analize mreža bitne.

Primjer 1. Ako neka firma treba otvoriti nova skladišta kako bi što bolje opskrbila svoje trgovine te zna kolike su cijene putovanja od lokacije do lokacije, koliko često putuju na različite lokacije te koliki su troškovi samog skladišta, firma može pronaći najbolju lokaciju za svoje skladište kako bi troškovi do krajnjih destinacija bili minimalni.

Primjer 2. Ako postoji proizvod za neku ciljanu skupinu, moguće je pronaći osobu u društvenoj mreži koja ima najviše utjecaja na našu predviđenu ciljanju skupinu te se putem nje oglašavati.

Jedna od stvari koje možemo analizirati u mrežama je mjera centralnosti. Mjera centralnosti temeljno je svojstvo mreže koje rangira čvorove prema njihovoj strukturnoj važnosti. Računanje mjera centralnosti ujedno je i glavna tema ovog rada.

U ovom radu baviti ćemo se raznim načinima određivanja mjera centralnosti. Cilj je detaljno obraditi matematičke aspekte metoda te testirati numeričku realizaciju na odabranim primjerima.

1.1 Struktura rada

Diplomski je rad podijeljen u pet poglavlja. U prvom poglavlju dajemo cilj diplomskog rada te motivaciju zašto se isplati baviti mjerama centralnosti. U drugom poglavlju definiramo osnovne pojmove koji su nam potrebni za razumijevanje rada te pokazujemo potrebne metode računanja s matricama. U trećem poglavlju dajemo detaljan pregled različitih metoda mjerenja centralnosti. Metode mjerenja centralnosti koje ćemo obrađivati su:

- *Page Rank*
- Podgrafovska centralnost
- Totalna podgrafovska komunikativnost
- Katzova centralnost

U četvrtom poglavlju napraviti ćemo usporedbu različitih mjera centralnosti na par jednostavnih primjera te na primjeru iz stvarnog svijeta. Peto poglavlje namijenjeno je za zaključak.

Poglavlje 2

Osnovni pojmovi

2.1 Pojmovi iz teorije grafova

Za razumijevanje gradiva iz područja analize mreža potrebno je razumjeti pojmove iz raznih područja kao što su linearna algebra, teorija grafova te računarske znanosti. Na početku valja prikazati osnovne definicije iz teorije grafova.

Graf G definiramo kao uređen par skupova (V, E) , gdje je V skup vrhova, a E je skup bridova $E = \{(i, j) | i, j \in V\}$. Za dani graf $G = (V, E)$, $|V| = n$ je broj vrhova u grafu, a $|E| = m$ broj bridova. Za graf kažemo da je *usmjeren* ako bridovi imaju smjer, tj. određeno je koji je početni vrh, a koji krajnji. Graf s usmjerenim bridovima zovemo i *digraf*, u suprotnom kažemo da je graf *neusmjeren*. *Stupanj vrha* i neusmjerenog grafa G definiramo kao broj bridova grafa G koji sadrže vrh i . Kod usmjerenih grafova razlikujemo *ulazni* i *izlazni* stupanj vrha i . Ulazni stupanj nam govori koliko bridova ulazi u vrh i dok nam izlazni stupanj govori koliko bridova izlazi iz vrha i . Graf s višestrukim bridovima je graf koji sadrži vrhove i i j tako da postoje dva brida $e, f \in E, e \neq f$ koji oba povezuju vrh i sa vrhom j . Težina brida je 'cijena' koju treba platiti za prelazak preko tog brida; ako se ne naglasi drugačije, uzimamo da graf nema težina tj. težina svakog brida je jednaka jedan. *Jednostavni graf* jest graf bez petlji (vrh koji je povezan sam sa sobom), višestrukih bridova i bez težina. *Šetnja* grafom dužine k je niz vrhova v_1, v_2, \dots, v_{k+1} takav da vrijedi $(v_i, v_{i+1}) \in E$ za sve $1 \leq i \leq k$. *Zatvorena šetnja* jest šetnja koja započinje i završava u istom vrhu. *Ciklus* je zatvorena šetnja u kojoj se ne ponavljaju vrhovi. Ako se svi vrhovi mogu dokučiti iz bilo kojeg drugog vrha tada kažemo da je graf *povezan*.

Svaki graf možemo prikazati kao matricu pomoću grafove matrice susjedstva. Za graf $G = (V, E)$ definiramo *matricu susjedstva* $A \in \mathbb{R}^{|V| \times |V|}$ kao

$$A = (a_{ij}); \quad a_{ij} = \begin{cases} 1 & \text{ako je } (i,j) \text{ u } E \\ 0 & \text{inače} \end{cases} .$$

Zahtjev da grafovi nemaju petlje uzrokuje da u matrici A imamo nule na dijagonali. Ako je mreža neusmjerena, matrica susjedstva A biti će simetrična. Važno svojstvo simetričnih matrica, koje ćemo koristiti kasnije, jest da imaju realne svojstvene vrijednosti. Za matricu A kažemo da je operator ako zadovoljava svojstva aditivnosti i homogenosti.

Za matricu A kažemo da je hermitska, tj vrijedi $A^* = A$, ako je za sve $i, j, 1 \leq i, j \leq n$ kompleksno konjugirana vrijednosti elementa $a_{i,j}$ jednaka elementu $a_{j,i}$ dakle: $\bar{a}_{i,j} = a_{j,i}$, gdje $\bar{\cdot}$ označava konjugirano kompleksnu vrijednost. U ovom radu susretat ćemo se samo sa realnim matricama tako da je zahtjev da matrica bude hermitska zapravo jednak zahtjevu da matrica bude simetrična.

Prije nego što se krenemo baviti metodama mjerenja centralnosti koje su i tema rada prvo ćemo ukratko predstaviti neke osnovne metode mjerenja centralnosti. To su:

- Stupanjska centralnost
- *Betweenness centrality*
- Blizinska centralnost
- Centralnost svojstvenog vektora.

Stupanjska centralnost

Stupanjska centralnost najjednostavnija je mjera centralnosti. Govori koliko bridova izlazi iz pojedinog vrha. U digrafovima postoje ulazne i izlazne mjere centralnosti. Za vrh i u neusmjerenom grafu stupanjsku centralnost vrha i , d_i , računamo:

$$d_i = \sum_{j=1}^n a_{ij} = \mathbf{e}_i^T \mathbf{A} \mathbf{e},$$

gdje je \mathbf{e} vektor jedinica, a A matrica susjedstva.

Betweenness centrality

Betweenness centrality mjeri koliko puta dani vrh leži na najkraćem putu između dva vrha. Ova mjera govori koji su vrhovi poveznice između vrhova. Radi na način da identificira sve najkraće puteve u mreži te zatim pobroji koliko puta se koji vrh nalazi na najkraćim putevima.

Blizinska centralnost

Blizinska centralnost mjeri koliko je dani vrh 'blizu' svim preostalim vrhovima mreže. Prvo računamo najkraće puteve između svih vrhova te potom svakom vrhu dodjeljujemo vrijednost tako da gledamo prosječnu udaljenost najkraćim putem do svih preostalih vrhova. Računamo ju na sljedeći način:

$$c_i = \frac{N - 1}{e_i^T D e},$$

gdje je e vektor jedinica, a D matrica udaljenosti.

Centralnost svojstvenih vektora

Slično kao stupanjska centralnost, utjecaj vrha određuje se brojem poveznica, ali se također uzima u obzir koliko su dobro povezani vrhovi s kojima je originalni vrh povezan i tako kroz cijelu mrežu. Relativne vrijednosti su pridružene svim vrhovima na temelju koncepta da poveznice s vrhovima koje imaju veću vrijednost vrijede više nego poveznice s vrhovima niže vrijednosti. Ovim se načinom mogu identificirati koji vrhovi imaju najviše utjecaja na cijelu mrežu.

2.2 Računanje s matricama

Za shvaćanje rada potrebno je objasniti pojedine operacije nad matricama. Konkretno, zanimaju nas funkcije:

- Moore–Penroseov inverz
- Matrična eksponencijalna funkcija e^A
- Korijen simetrične pozitivno semidefinitne matrice \sqrt{A}
- Sinus i kosinus hiperbolni matrice $\sinh A, \cosh A$

2.2.1 Moore–Penroseov inverz

Za matricu A definiramo matricu A^+ koju nazivamo *Moore–Penroseova* matrica ako zadovoljava sljedeće Moore-Penroseove uvjete:

- $AA^+A = A$

- $A^+AA^+ = A^+$
- AA^+ je hermitska matrica
- A^+A je hermitska matrica

Ukoliko A ima linearno nezavisne retke, tj. AA^* je inverzabilna, tada A^+ možemo izraziti kao $A^+ = A^*(AA^*)^{-1}$, no ukoliko A ima linearno nezavisne stupce, tj. A^*A je inverzabilna tada A^+ možemo izraziti kao $A^+ = (A^*A)^{-1}A^*$. *Moore–Penroseova* matrica je pseudo inverz matrice A pomoću kojeg računamo rješenje jednadžbe $Ax = b$ kada je matrica A singularna, što zapravo znači da tražimo rješenje sustava $Ax \approx b$, tj. tražimo x tako da je $\|Ax - b\|_2$ minimalan.

2.2.2 Matrična eksponencijalna funkcija

Sljedi postupak raspisivanja načina na koji računamo matričnu eksponencijalnu funkciju. Taylorov red, e^z , jednak je $e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}$ te konvergira za svaki z . Matričnu eksponencijalnu funkciju za matricu A definiramo s $e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!}$. Kako bi izračunali matričnu eksponencijalnu funkciju za simetričnu matricu A prvo je potrebno napraviti dekompoziciju $A = U\Lambda U^T$ gdje je Λ dijagonalna matrica koja sadrži svojstvene vrijednosti matrice A na dijagonali, a matrica U je matrica sastavljena od svojstvenih vektora za pripadnu svojstvenu vrijednost matrice A . Pogledajmo kako izgleda A^2 .

$$A^2 = AA = U\Lambda U^T U\Lambda U^T = U\Lambda^2 U^T,$$

jer vrijedi $UU^T = I$. Sada za matricu A^3 vrijedi:

$$A^2 = A^2A = U\Lambda^2 U^T U\Lambda U^T = U\Lambda^3 U^T.$$

Zaključujemo za proizvoljni $k \in \mathbb{N}$ vrijedi $A^k = U\Lambda^k U^T$. Finalno, imamo

$$e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!} = \sum_{n=0}^{\infty} \frac{1}{n!} U\Lambda^n U^T = U \sum_{n=0}^{\infty} \frac{1}{n!} \Lambda^n U^T.$$

Pogledajmo sada zapis $\sum_{n=0}^{\infty} \frac{1}{n!} \Lambda^n$. Imamo:

$$\sum_{n=0}^{\infty} \frac{1}{n!} \Lambda^n = \sum_{n=0}^{\infty} \frac{1}{n!} \begin{bmatrix} \lambda_1^n & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n^n \end{bmatrix} = \begin{bmatrix} \sum_{n=0}^{\infty} \frac{1}{n!} \lambda_1^n & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sum_{n=0}^{\infty} \frac{1}{n!} \lambda_n^n \end{bmatrix} = \begin{bmatrix} e^{\lambda_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & e^{\lambda_n} \end{bmatrix}.$$

Primjećujemo ako nam je poznato da su $\lambda_1 \dots \lambda_n$ svojstvene vrijednosti matrice A , tada su poznate i svojstvene vrijednosti matrice e^A te one iznose $e^{\lambda_1} \dots e^{\lambda_n}$.

Neka od svojstava matične eksponencijalne funkcije su $e^0 = I$ gdje je 0 nul matrica, a I matrica identiteta. Nadalje vrijedi $e^{A^T} = (e^A)^T$. Neka su $A, B \in M_{n \times n}$, ukoliko vrijedi $AB = BA$, tada vrijedi $e^A e^B = e^{A+B}$ iz čega slijedi $e^A e^{-A} = e^{A-A} = e^0 = I$ za sve matrice koje imaju jedinstveni inverz. Za više informacija o matičnoj eksponencijalnoj funkciji možemo pogledati na primjer ovdje ([14]).

2.2.3 Korijen matrice

Ako je potrebno izračunati korijen matrice A , zapravo treba pronaći matricu M za koju vrijedi $M^2 = A$ te koja je i sama pozitivno semidefinitna.

Korijen matrice možemo računati za matricu A ukoliko je matrica pozitivno semidefinitna tj. ako vrijedi $A = A^T$ i $x^T A x \geq 0, x \in \mathbb{R}^n$. Tada za matricu A vrijedi $Au_i = \lambda_i u_i, i \in \{1, \dots, n\}$ gdje su u_i svojstveni vektori, a λ_i svojstvene vrijednosti matrice A te za svojstvene vrijednosti vrijedi $\lambda_i \geq 0, i \in \{1, \dots, n\}$.

Definiramo

$$U = [u_1, \dots, u_n] \quad \Lambda = \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{bmatrix}.$$

Imamo:

$$M = \sqrt{A} = U \sqrt{\Lambda} U^T,$$

gdje je $\sqrt{\Lambda}$ dijagonalna matrica koja na dijagonali sadrži vrijednosti korijena dijagonale matrice Λ . Primjećujemo da za matricu Λ imamo 2^n matrica $\sqrt{\Lambda}$ jer je na svakoj poziciji dijagonale moguće odabrati i pozitivni i negativni korijen vrijednosti dijagonale iz matrice Λ , mi ćemo uvijek odabirati pozitivne korijene svojstvenih vrijednosti.

2.2.4 Sinus i kosinus hiperbolni matrice

Sinus hiperbolni definiran je kao $\sinh x = \frac{e^x - e^{-x}}{2}$, dok je kosinus hiperbolni definiran kao $\cosh x = \frac{e^x + e^{-x}}{2}$. Analogno definiramo sinus i kosinus hiperbolni za matricu A kao $\sinh A = \frac{e^A - e^{-A}}{2}$ i $\cosh A = \frac{e^A + e^{-A}}{2}$. Sinus hiperbolni ima razvoj

$$\sinh x = \sum_{n=0}^{\infty} \frac{x^{2n+1}}{(2n+1)!},$$

dok razvoj za kosinus hiperbolni izgleda

$$\cosh x = \sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!}.$$

Razvoji slijede iz derivacija sinusa i kosinusa hiperbolnog ($\frac{d}{dx} \cosh x = \sinh x$ i $\frac{d}{dx} \sinh x = \cosh x$), te raspisom Maclaurinovog razvoja ($\sum_{n=0}^{\infty} \frac{x^n}{n!} f^{(n)}(0)$).

Primjetimo, u (2.2.2) smo pokazali kako se računa e^A . Za e^{-A} vrijedi $e^{-A} = (e^A)^{-1}$, budući da vrijedi $e^A e^{-A} = e^0 = I$, što ponovo znamo izračunati.

Poglavlje 3

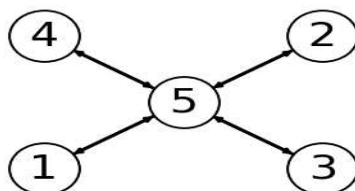
Vrste centralnosti

U sljedećem poglavlju predstaviti ćemo ranije spomenute metode mjerenja centralnosti. Proći ćemo kroz ideje iza svake metode, njihove osnovne karakteristike, prednosti i mane.

3.1 *Page Rank*

Prva ideja koja se pojavila za mjerenje centralnosti, ujedno je možda i najpoznatija metoda za mjerenje centralnosti. Riječ je, naravno, o *Page Ranku*. Razlog zašto bi *Page Rank* mogao biti poznat čitatelju je to što su ga osmislili osnivači *Googla*. Danas *Page Rank* nije jedini algoritam koji *Google* koristi kako bi sortirao rezultate pretraživanja, ali je i dalje najpoznatiji algoritam u *Googlu*.

Osnovna ideja *Page Rank* algoritma je dati svakom vrhu njegovu vrijednost na temelju poveznica između preostalih vrhova mreže i odabranog vrha. Na taj način mreža postaje svojevrsna demokracija u kojoj vrhovi glasaju za vrijednosti preostalih vrhova svojim poveznicama. No, ovdje se ne radi o pravoj demokraciji budući da svojstvo koje želimo da ovaj algoritam ima jest da glasovi vrhova veće vrijednosti vrijede više nego glasovi niže rangiranih vrhova. Također, još jedno svojstvo koje želimo za algoritam jest da pojedini vrh ne može glasati sam za sebe.



Slika 3.1: Primjer usmjerene mreže s pet vrhova u kojoj je vrh 5 središnji vrh te je povezan sa svim preostalim vrhovima, dok su preostali vrhovi povezani s vrhom 5.

Pogledamo li graf sa slike 3.1, vrh v_5 bi glasao za sve preostale vrhove, dok bi svi preostali vrhovi glasali za vrh v_5 . Rezultate primjene svih metoda centralnosti na pokazani graf mogu se pronaći u (4.2).

Page Rank je iterativni algoritam koji nam omogućava ponalaženje autoritativnih vrhova u digrafu. Dakle, možemo ga koristiti i za usmjerene i ne usmjerene grafove. U ovoj metodi važnost pojedinog vrha i određujemo pomoću važnosti vrhova koji ulaze u i . Najjednostavnije iskazano, rang zadanog vrha i definiran je sa:

$$r(i) = \sum_{u \in B_i} \frac{r(u)}{|u|}, \quad (3.1)$$

gdje je $B_i = \{u: \text{postoji izlazni brid iz } u \text{ u } i\}$, a $|u|$ je izlazni stupanj vrha u .

Iz (3.1) primjećujemo da će rangovi pojedinog vrha ovisiti o vrijednostima koje uzmemo za početnu vrijednost pojedinog vrha i . U praksi najčešće se uzima jednaka početna vrijednost svih vrhova i on iznosi $\frac{1}{n}$ gdje je n broj vrhova u mreži. Nakon postavljanja početnih vrijednosti krećemo s iterativnim računanjem sve dok vrijednosti ne krenu konvergirati. Prethodno napisano možemo napisati na sljedeći način:

$$\pi_k = P\pi_{k-1}, \quad k = 1, 2, \dots, \quad (3.2)$$

gdje π_k predstavlja vektor autoriteta u mreži u k -toj iteraciji, dok je P matrica zadana na slijedeći način:

$$p_{ij} = \begin{cases} \frac{1}{|u|} & \text{postoji izlazni brid iz vrha } i \text{ do vrha } j \\ 0 & \text{inače} \end{cases}.$$

Matricu P možemo shvatiti kao vjerojatnosnu matricu prijelaza gdje je p_{ij} vjerojatnost prelaska sa vrha i u vrh j preko jednog brida, a iteracije možemo shvatiti kao evoluciju markovljevih lanaca koji modeliraju nasumičnu šetnju po mreži. Za kvadratnu matricu P vrijedi da je ona stupčasto stohastična, što znači da su joj vrijednosti unutar svakog stupca ne negativne i imaju sumu jednaku jedan.

Propozicija 3.1.1. *Svaka stupčasto - stohastična matrica ima $\mathbf{1}$ kao svojstvenu vrijednost.*

Dokaz. Neka je A kvadratna stupčasto - stohastična matrica dimenzije n . Neka je \mathbf{e} vektor koji na svakoj poziciji ima vrijednost 1. Za matricu A vrijedi da ima iste svojstvene vrijednosti kao i matrica A^T . Budući da je A stupčasto - stohastična matrica lako provjerimo da vrijedi:

$$A^T \mathbf{e} = \mathbf{e}.$$

Dakle, $\mathbf{1}$ je svojstvena vrijednost od A^T , slijedi da je također svojstvena vrijednost od A . \square

Označimo nadalje s $V_1(A)$ svojstveni podprostor pridružen svojstvenoj vrijednosti **1**.

Postoji par nedostataka *Page Rank* algoritma. To su: mreže s nejedinstvenim rangovima i mreže s nepovezanim vrhovima.

Unutar zadane mreže M , s pripadnom matricom susjedstva A , bilo bi poželjno kada bi dimenzija svojstvenog potprostora $\dim(V_1(A)) = 1$. To bi osiguralo da postoji jedinstveni svojstveni vektor \mathbf{x} za koji vrijedi $\sum_i x_i = 1, x_i \geq 0$, te bi ga mogli koristiti za računanje vrijednosti vrhova. Ukoliko za danu mrežu M vrijedi $\dim(V_1(A)) > 1$, to znači da postoji više vektora koji mogu činiti bazu tog svojstvenog potprostora. Također, svaka linearno nezavisna kombinacija baze od $V_1(A)$ nalazi se unutar tog svojstvenog potprostora. Postavlja se pitanje koji je od tih vektora najbolje za koristiti prilikom određivanja vrijednosti pojedinih vrhova, ako je uopće ijedan dobar. Dimenzija potprostora $V_1(A)$ veća od jedan, posljedica je nepovezane mreže. Ukoliko se naša mreža M sastoji od k nepovezanih podmreža M_1, \dots, M_k tada vrijedi $\dim(V_1(A)) \geq k$, te kao posljedica ne postoji jedinstveni vektor $\mathbf{x} \in V_1(A)$ tako da vrijedi $\sum_i x_i = 1, x_i \geq 0$. To objašnjavamo na sljedeći način: ukoliko se mreža M sastoji od k podmreža, teško je uspoređivati vrijednosti vrhova pojedine podmreže $M_i, 1 \leq i \leq k$, s vrijednostima pojedinih vrhova podmreže $M_j, 1 \leq i \leq k, i \neq j$. Za mrežu M koja se sastoji od k nepovezanih podmreža M_1, \dots, M_k , označimo s n_i broj vrhova u A_i . Vrhove u mreži M_1 označimo indeksima od 1 do k_1 , u mreži M_2 s $k_1 + 1$ do $k_1 + k_2$ i tako dalje. Primjećujemo da matricu susjedstva originalne mreže možemo zapisati kao:

$$\begin{bmatrix} A_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & A_k \end{bmatrix},$$

gdje A_i označava matricu susjedstva za mrežu M_i . Svaku pojedinu podmrežu M_i možemo shvatiti kao zasebnu mrežu dimenzija $k_i \times k_i$ za koju je pripadna matrica susjedstva A_i stupčasto-stohastična. Za svaki i između 1 i k konstruiramo vektor $\mathbf{v}^i \in \mathbb{R}^n$ koji sadrži nule na svim pozicijama osim na pozicijama koje odgovaraju i bloku. Lako se pokaže da su vektori $\mathbf{v}^i, i \leq i \leq k$ linearno nezavisni svojstveni vektori za matricu A . Zaista,

$$A\mathbf{v}^i = A \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \mathbf{v}^i \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{v}^i.$$

Drugi problem s *Page Rankom* je, što za proizvoljnu mrežu, ne postoji garancija da će vrijednosti autoriteta konvergirati. Ukoliko imamo mrežu koja sadrži vrh koji nije povezan

nit i s jednim drugim vrhom tj. neka postoji vrh i koji ima izlazni stupanj jednak nula, tada matrica P neće biti stupčasto - stohastična. Kako bi doskočili tome, koristimo matricu \bar{P} gdje je svaki redak jednak $\mathbf{0}^T$ zamjenjen s $\frac{\mathbf{e}^T}{n}$. Iako ova zamjena rješava problem konvergencije, ne garantira jedinstveno rješenje te se, ako je matrica reducibilna, mogu pojaviti *Rang Sinkovi*. Rang sinkovi su vrhovi u kojima nasumične šetnje zaglave te kao posljedica ti vrhovi dobiju nerazmjerno veliki rang. No, ako matrica nije reducibilna, rang sinkovi se tada ne javljaju.

Standardni način zadavanje stohastične, ireducibilne matrice \bar{P} je uvođenjem Rank-1 matrice $E = \frac{\mathbf{e}\mathbf{e}^T}{n}$ te umjesto \bar{P} uzimamo konveksnu kombinaciju

$$\bar{\bar{P}} = \alpha \bar{P} + (1 - \alpha)E,$$

gdje je α konstanta takva da vrijedi $0 < \alpha < 1$. U praksi preporučena vrijednost je $\alpha = 0.85$.

U radu ([8]) pokazano je da ukoliko primijenimo *Page Rank* na matricu dobivenu zamjenom smjerova svih bridova dobivamo način za rangiranje svih hubova u mreži.

3.2 HITS Algoritam

U sljedećem odjeljku ćemo predstaviti **HITS**, *Hypertext Induced Topics Search*, algoritam kojeg je originalno predstavio J. Kleinberg u radu ([11]). Iako ovaj algoritam nije metoda za mjerenje centralnosti vrhova, uvodimo ga jer će nam biti koristan u kasnijim poglavljima. Konkretno, ovaj algoritam daje nam motivaciju za primjenu podgrafovske centralnosti u digrafovim. HITS algoritam zasnovan je na ideji da u svim digrafovim postoje dvije vrste važnih vrhova, to su *hubovi* i *autoriteti*. Hubovi su vrhovi koji pokazuju na vrhove koje smatramo važnima, autoriteti su ti važni vrhovi. Na ovaj način dobivamo kružnu definiciju:

Dobri hubovi su oni koji pokazuju na puno dobrih autoriteta, a dobri autoriteti su oni na koje pokazuju dobri hubovi.

Iz takve definicije slijedi da HITS rangovi ovise o iterativnoj metodi koja konvergira u finalne vrijednosti. Svakom vrhu i pridružene su dvije ne-negativne težine, autoritet težina x_i i hub težina y_i . U početnoj iteraciji svakom vrhu su dane proizvoljne vrijednosti težina. Težine se zatim mjenjaju na slijedeći način:

$$x_i^{(k)} = \sum_{j:(j,i) \in E} y_j^{(k-1)} \quad y_i^{(k)} = \sum_{j:(j,i) \in E} x_j^{(k-1)} \quad k = 1, 2, 3, \dots \quad (3.3)$$

Dobivene težine su zatim normalizirane tako da vrijede $\sum_i (x_i^{(k)})^2 = 1$ i $\sum_j (y_j^{(k)})^2 = 1$.

Navedene iteracije se izvršavaju slijedno te iako se može pokazati da vektori \vec{x}^k \vec{y}^k konvergiraju kako se k približava beskonačnosti, u praksi se postupak zaustavlja kada nema značajne promjene između iteracija.

U matricnom zapisu jednadžbe (3.3) postaje $\mathbf{x}^{(k)} = A^T \mathbf{y}^{(k-1)}$ i $\mathbf{y}^{(k)} = A^T \mathbf{x}^{(k-1)}$ te su popraćene normalizacijom u 2-normi. Iterativni proces možemo zapisati

$$\mathbf{x}^{(k)} = c_k A^T A \mathbf{x}^{(k-1)} \quad \mathbf{y}^{(k)} = c_{k'} A^T A \mathbf{y}^{(k-1)}, \quad (3.4)$$

gdje su c_k i $c_{k'}$ faktori normalizacije. Iz toga slijedi da je HITS metoda kojom računamo svojstvene vektore za $A^T A$ i AA^T . Vrijednosti autoriteta određene su vrijednostima dominantnih svojstvenih vektora matrice $A^T A$, koju nazivamo *matrica autoriteta*. Vrijednosti hubova određene su vrijednostima dominantnih svojstvenih vektora matrice AA^T , koju nazivamo *matrica hubova*. Detaljnije informacije o HITS algoritmu možemo pronaći u ([5]).

3.2.1 Primjena HITS algoritma

U digrafovima matrica susjedstva u pravilu nije simetrična, ipak matrice korištene u HITS algoritmu $A^T A$ i AA^T su simetrične. Definirajmo matricu

$$\mathcal{A} = \begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix}. \quad (3.5)$$

Tako dobivamo i

$$\mathcal{A}^2 = \begin{pmatrix} AA^T & 0 \\ 0 & A^T A \end{pmatrix}; \quad \mathcal{A}^3 = \begin{pmatrix} 0 & AA^T A \\ A^T AA^T & 0 \end{pmatrix}.$$

Općenito za $k \in \mathbb{N}$

$$\mathcal{A}^{2k} = \begin{pmatrix} (AA^T)^k & 0 \\ 0 & (A^T A)^k \end{pmatrix}; \quad \mathcal{A}^{2k+1} = \begin{pmatrix} 0 & A(A^T A)^k \\ (A^T A)^k A^T & 0 \end{pmatrix}.$$

Primjenom HITS algoritma na \mathcal{A} za koju vrijedi $\mathcal{A} = \mathcal{A}^T$, to jest $\mathcal{A}^T \mathcal{A} = \mathcal{A} \mathcal{A}^T = \mathcal{A}^2$ te uvođenjem vektora $\mathbf{u}^k = \begin{pmatrix} \mathbf{x}^k \\ \mathbf{y}^k \end{pmatrix}$ za $k = 1, 2, 3, \dots$ jednadžba (3.4) postaje

$$\mathbf{u}^{(k)} = \mathcal{A}^2 \mathbf{u}^{(k-1)} = \begin{pmatrix} AA^T & 0 \\ 0 & A^T A \end{pmatrix} \mathbf{u}^{(k-1)}, \quad (3.6)$$

popraćena s normalizacijom dvije vektorske komponente od $\mathbf{u}^{(k)}$ tako da je njihova 2-norma jednaka 1. Kao što je ranije spomenuto, kako se k približava beskonačnosti $\{\mathbf{u}^{(k)}\}$ konvergira ka dominantnom svojstvenom vektoru matrice \mathcal{A} koji daje željene rangove za hubove i autoritete. Naime, prvih n vrijednosti u \mathbf{u} daje vrijednosti rangova hubova, dok zadnjih n vrijednosti daje rangove autoriteta.

Jedno od ograničenja HITS algoritma jest potencijalna ovisnost o početnim vrijednostima za $\mathbf{x}^{(0)}$ i $\mathbf{y}^{(0)}$, zato se najčešće kao početne vrijednosti uzima

$$\mathbf{x}^{(0)} = \mathbf{y}^{(0)} = \left[\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}} \right],$$

ove početne vrijednosti uzimamo kako ni jednom vrhu na početku ne bi dali prednost te da bi 2-norma bila jednaka jedan. Drugi je nedostatak velika ovisnost o ulaznim i izlaznim stupnjevima pojedinih vrhova.

3.3 Podgrafovska centralnost

Sljedeći odjeljak započinje propozicijom koja nam daje motivaciju za korištenje matrične eksponencijalne funkcije za računanje podgrafovske centralnosti.

Propozicija 3.3.1. *Neka je $G = (V, E)$ neusmjereni graf gdje je $V = \{v_1, \dots, v_n\}$ skup vrhova i E skup bridova te $A = (a)_{ij}$ pripadna matrica susjedstva. Neka je $k \in \mathbb{N} \setminus 0$ te $1 \leq i, j \leq n$. Broj šetnji duljine k između vrhova i i j jednak je vrijednosti $(a)_{ij}$ u matrici A^k .*

Dokaz. Dokaz ide indukcijom po k .

Baza: $k = 1$, trivijalno.

Pretpostavka: Neka za proizvoljni $k = m$ vrijedi da je broj šetnji duljine m između vrhova i i j jednak je vrijednosti $(a)_{ij}$ u matrici A^m

Korak: $k = m + 1$

Neka je $B = (b)_{ij} = A^m$, $A^{m+1} = BA = (c)_{ij}$.

Tada je na poziciji (i, j) u matrici A^{m+1} jednak

$$c_{ij} = \sum_{l=1}^n b_{il} a_{lj}.$$

Po pretpostavci indukcije, za sve $1 \leq l \leq n$, b_{il} je broj šetnji duljine m između vrhova v_i i v_l . Te vrijednosti dodamo c_{ij} **akko** $v_l, v_j \in E$. Zaključujemo, c_{ij} je broj šetnji duljine $m + 1$ između vrhova v_i i v_j □

U ([7]), uveden je koncept *podgrafovska centralnost* kao mjera centralnosti vrhova u mreži. U podgrafovskoj centralnosti vrijednosti vrhova dobivamo na temelju dijagonalnih vrijednosti dobivenih primjenom matričnih funkcija na matricu susjedstva. Matrične funkcije koje koristimo mogu biti različite, no najučestalija je matrična eksponencijalna funkcija koju smo obradili u (2.2.2).

Podgrafovski centralnost vrha i dana je sa $[e^A]_{ii}$. *Podgrafovski komunikativnost* između vrhova i i j dana je sa $[e^A]_{ij}$. Ukoliko je graf neusmjeren, tada je komunikativnost između vrhova i i j jednaka komunikativnosti između vrhova j i i tj. $[e^A]_{ij} = [e^A]_{ji}$. Vrh sa većom podgrafovskom mjerom centralnosti smatra se važnijim vrhom od vrhova sa manjom podgrafovskom mjerom centralnosti. Veća mjera podgrafovski komunikativnosti između vrhova i i j nam govori da je protok informacija lakši nego pomoću vrhova s manjom mjerom komunikativnosti. Odnosno, mala mjera komunikativnosti između dva vrha nam kaže da ta dva vrha ne mogu lagano izmjenjivati informacije.

Razlog zašto koristimo vrijednosti dijagonale kao mjeru centralnosti jest to što u raspisu (2.2.2) imamo zapise oblika A^k . Takvi zapisi su nam zanimljivi jer u teoriji grafova je poznato da ukoliko je A matrica susjedstva za graf G tada nam A^k_{ij} broji koliko ima šetnji od vrha i do vrha j duljine k , što smo pokazali u gore spomenutoj propoziciji.

Dakle, podgrafovski centralnost vrha i , koja je jednaka $[e^A]_{ii}$, zapravo broji sve zatvorene šetnje centralizirane u vrhu i svih mogućih duljina k s kaznom $\frac{1}{k!}$. Uočimo da na ovaj način brojimo i "nelogične" šetnje, na primjer $v_i \rightarrow v_j \rightarrow v_i$, no svejedno kratkim šetnjama dajemo veću važnost te nam podgrafovski centralnost vrha i daje informaciju koliko je taj vrh blizu svemu ostalom u mreži. Za ovu metodu digrafovi mogu biti problematični. Uzmimo graf čiji su bridovi $E = \{(1, 2), (2, 3), \dots, (n-1, n)\}$, tada su elementi matrice e^A jednaki

$$e_{ij} = \begin{cases} \frac{1}{(j-i)!} & i \leq j \\ 0 & \text{inače} \end{cases}, \quad (3.7)$$

iz čega slijedi da su sve vrijednosti dijagonale jednake 1. Zaključujemo da su svi vrhovi jednako važni za protok informacija, no uočavamo da su očito rubni vrhovi manje uključeni u prijenos informacija nego središnji. Ovdje je dio problema to što takav digraf ne sadrži zatvorene šetnje.

3.3.1 Podgrafovski centralnost za digrafe

Ranije smo spomenuli da ćemo za digrafe koristiti matricu \mathcal{A} , (3.2.1), zbog njenog svojstva simetričnosti. Tada možemo računati pojedine komponente od $e^{\mathcal{A}}$. U sljedećoj propoziciji povežujemo $e^{\mathcal{A}}$ sa maticama hubova i autoriteta originalne matrice A .

Propozicija 3.3.2. *Neka je \mathcal{A} matrica opisana u (3.5), tada*

$$e^{\mathcal{A}} = \begin{pmatrix} \cosh(\sqrt{AA^T}) & A(\sqrt{A^T A})^+ \sinh(\sqrt{A^T A}) \\ \sinh(\sqrt{A^T A})(\sqrt{A^T A})^+ A^T & \cosh(\sqrt{A^T A}) \end{pmatrix}.$$

Dokaz. Neka je $A = U\Sigma V^T$ SVD dekompozicija originalne, nesimetrične, matrice A . Tada \mathcal{A} se može raspisati kao $\mathcal{A} = \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} \begin{pmatrix} 0 & \Sigma \\ \Sigma & 0 \end{pmatrix} \begin{pmatrix} U^T & 0 \\ 0 & V^T \end{pmatrix}$. Slijedi:

$$e^{\mathcal{A}} = \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} e^{\begin{pmatrix} 0 & \Sigma \\ \Sigma & 0 \end{pmatrix}} \begin{pmatrix} U^T & 0 \\ 0 & V^T \end{pmatrix}.$$

Sada,

$$e^{\begin{pmatrix} 0 & \Sigma \\ \Sigma & 0 \end{pmatrix}} = \cosh \begin{pmatrix} 0 & \Sigma \\ \Sigma & 0 \end{pmatrix} + \sinh \begin{pmatrix} 0 & \Sigma \\ \Sigma & 0 \end{pmatrix} = \begin{pmatrix} \cosh \Sigma & 0 \\ 0 & \cosh \Sigma \end{pmatrix} + \begin{pmatrix} 0 & \sinh \Sigma \\ \sinh \Sigma & 0 \end{pmatrix}.$$

Dakle,

$$e^{\begin{pmatrix} 0 & \Sigma \\ \Sigma & 0 \end{pmatrix}} = \begin{pmatrix} \cosh \Sigma & \sinh \Sigma \\ \sinh \Sigma & \cosh \Sigma \end{pmatrix}.$$

Sada imamo:

$$\begin{aligned} e^{\mathcal{A}} &= \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} \begin{pmatrix} \cosh \Sigma & \sinh \Sigma \\ \sinh \Sigma & \cosh \Sigma \end{pmatrix} \begin{pmatrix} U^T & 0 \\ 0 & V^T \end{pmatrix} = \dots = \\ &= \begin{pmatrix} \cosh(\sqrt{AA^T}) & A(\sqrt{A^T A})^+ \sinh(\sqrt{A^T A}) \\ \sinh(\sqrt{A^T A})(\sqrt{A^T A})^+ A^T & \cosh(\sqrt{A^T A}) \end{pmatrix}. \end{aligned}$$

□

Ostaje interpretirati dobiveni rezultat. Kako bi smo to napravili, raspíšimo prvo još par stvari o matrici \mathcal{A} . Matricu \mathcal{A} možemo promatrati na slijedeći način:

Uzmemo sve vrhove originalne mreže M te napravimo dvije kopije V i V' i označimo ih sa $v(\mathcal{A}) = V \cup V'$, zatim konstruiramo skup bridova $E' = \{(i, j') \mid \text{postoji usmjereni brid od } i \text{ do } j \text{ u originalnoj mreži}\}$ te tako kreiramo bipartitivni graf sa $2n$ vrhova. Svaki vrh ima dvije uloge, ulogu huba i ulogu autoriteta te je malo vjerojatno da je vrh i istovremeno dobar hub i dobar autoritet. U mreži \mathcal{A} te uloge su rastavljene, prvih n vrhova u $v(\mathcal{A})$ predstavljaju originalne vrhove u ulozi hubova dok zadnjih n predstavlja iste te vrhove u ulozi autoriteta.

U usmjerenom grafu, *alternirajuća šetnja* duljine k koja počinje izlaznim bridom, od v_1 do v_{k+1} je niz vrhova v_1, v_2, \dots, v_{k+1} ako postoji brid (v_{i+1}, v_i) ako je i neparan i brid (v_i, v_{i+1}) ako je i paran:

$$v_1 \rightarrow v_2 \leftarrow v_3 \rightarrow \dots$$

Alternirajuća šetnja duljine k koja počinje ulaznim bridom, od v_1 do v_{k+1} je niz vrhova v_1, v_2, \dots, v_{k+1} ako postoji brid (v_i, v_{i+1}) ako je i neparan i brid (v_{i+1}, v_i) ako je i paran:

$$v_1 \leftarrow v_2 \rightarrow v_3 \leftarrow \dots$$

U ([4]) pokazano je da $[AA^T A \dots]_{ij}$, gdje imamo k množenja matrica, broji alternirajuće šetnje duljine k koje počinju sa izlaznim vrhom od vrha i do vrha j . Slično, $[A^T AA^T \dots]_{ij}$ broji alternirajuće šetnje koje počinju ulaznim bridom od vrha i do vrha j . Dakle, $[(AA^T)^k \dots]_{ij}$ i $[(A^T A)^k \dots]_{ij}$ broje alternirajuće šetnje duljine $2k$.

Gledajući šetnje duljine $2k$ s kaznom $\frac{1}{(2k)!}$, te šetnje možemo prebrojati gledajući vrijednosti dijagonale matrice

$$I + \frac{AA^T}{2!} + \frac{AA^T AA^T}{4!} + \dots + \frac{(AA^T)^k}{(2k)!} + \dots$$

Neka je $A = U\Sigma V^T$ SVD kompozicija matrice A , tada matrica postaje

$$U \left(I + \frac{\Sigma^2}{2!} + \frac{\Sigma^4}{4!} + \dots + \frac{\Sigma^{2k}}{(2k)!} + \dots \right) U^T = U \cosh(\Sigma) U^T = \cosh \sqrt{AA^T}.$$

Sada možemo izraziti *centralnost huba* vrha i iz originalne mreže A kao

$$[e^{\mathcal{A}}]_{ii} = [\cosh \sqrt{AA^T}]_{ii}.$$

Centralnost huba nam govori koliko dobro pojedini vrh i šalje informacije prema autoritetima u mreži.

Analogno, *centralnost autoriteta* vrha i iz originalne mreže A možemo izraziti kao

$$[e^{\mathcal{A}}]_{n+i, n+i} = [\cosh \sqrt{A^T A}]_{ii}.$$

Centralnost autoriteta nam govori koliko dobro vrh i prima informacije od hubova unutar mreže.

Vratimo li se sada na primjer digrafa (3.7), gledajući dijagonale od $e^{\mathcal{A}}$, vidimo da svi vrhovi imaju jednaku vrijednost. Dijagonale u $e^{\mathcal{A}}$ daju vrhu 1 najmanju vrijednost autoriteta, a vrhu n najmanju vrijednost huba dok su ostali izjednačeni. Dakle, korištenjem $e^{\mathcal{A}}$ dobili smo razumne vrijednosti hubova i autoriteta za sve vrhove.

3.3.2 Usporedba s HITS algoritmom

U HITS algoritmu rangiranje vrhova kao hubova i autoriteta koristi jedino informaciju dominantnog svojstvenog vektora matrice A , dok pri korištenju matrice $e^{\mathcal{A}}$ iskoristavamo vrijednosti svih svojstvenih vektora matrice \mathcal{A} .

Svojstvene vrijednosti matrice \mathcal{A} možemo poredati $\lambda_1 > \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_{2n}$ jer iz definicije matrice \mathcal{A} znamo da je ona simetrična. Tada matricu \mathcal{A} možemo zapisati kao $\mathcal{A} = \sum_{i=1}^{2n} \lambda_i \mathbf{u}_i \mathbf{u}_i^T$, gdje su $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{2n}$ normalizirani svojstveni vektori za pojedinu svojstvenu vrijednost. Uzimajući eksponent dobivamo

$$e^{\mathcal{A}} = \sum_{i=1}^{2n} e^{\lambda_i} \mathbf{u}_i \mathbf{u}_i^T = e^{\lambda_1} \mathbf{u}_1 \mathbf{u}_1^T + \sum_{i=2}^{2n} e^{\lambda_i} \mathbf{u}_i \mathbf{u}_i^T.$$

Rangovi hubova i autoriteta dobivamo promatrajući vrijednosti dijagonala matrice $e^{\mathcal{A}}$

$$\text{diag}(e^{\mathcal{A}}) = e^{\lambda_1} \text{diag}(\mathbf{u}_1 \mathbf{u}_1^T) + \sum_{i=2}^{2n} e^{\lambda_i} \text{diag}(\mathbf{u}_i \mathbf{u}_i^T).$$

Podijelimo dobiveni rezultat sa e^{λ_1} te dobivamo

$$\text{diag}(e^{-\lambda_1} e^{\mathcal{A}}) = \text{diag}(e^{\mathcal{A}-\lambda_1 I}) = \text{diag}(\mathbf{u}_1 \mathbf{u}_1^T) + \sum_{i=2}^{2n} e^{\lambda_i - \lambda_1} \text{diag}(\mathbf{u}_i \mathbf{u}_i^T).$$

Nadalje, vrijednosti na dijagonali matrice $\mathbf{u}_1 \mathbf{u}_1^T$ su zapravo kvadrati vrijednosti dominantnog svojstvenog vektora matrice \mathcal{A} . Dakle, u raspisu matrice $e^{\mathcal{A}}$ u svojstvenoj bazi od \mathcal{A} vrijednost na prvom mjestu je upravo ona dobivena korištenjem HITS algoritma.

Zaključujemo ukoliko vrijedi $\lambda_1 \gg \lambda_2$, tada za rangove dobivene *podgrafovskom centralnošću* matrice e^A je mala vjerojatnost da odstupaju od rangova dobivenih uporabom HITS algoritma budući da su vrijednosti $e^{\lambda_i - \lambda_1}$ blizu nule za $i = 2, 3, \dots, 2n$. S druge strane, ukoliko je razlika između λ_1 i λ_2 mala, $\lambda_1 \approx \lambda_2$, tada doprinos preostalih svojstvenih vektora, $\sum_{i=2}^{2n} e^{\lambda_i - \lambda_1} \text{diag}(\mathbf{u}_i \mathbf{u}_i^T)$, može biti nezanemariv relativno na vrijednost $\text{diag}(\mathbf{u}_1 \mathbf{u}_1^T)$ i samim time dobiveni rangovi se mogu značajno razlikovati od rezultata dobivenih korištenjem HITS algoritma.

3.4 Totalna podgrafovska komunikativnost

Matrični ekponent matrice susjedstva, e^A , nije zanimljiv samo zbog vrijednosti dijagonale u dobivenoj matrici, već su od interesa i sume pojedinih redova. Suma reda od e^A za vrh i dana je s $\sum_{j=0}^n [e^A]_{ij}$. Suma reda broji broj šetnji između vrha i i ostalih vrhova mreže, uključujući i gdje svakoj šetnji duljine k dodajemo kaznu $\frac{1}{k!}$. Sumu reda i u matrici e^A nazivamo još i *totalna podgrafovska komunikativnost*. Ona nam govori kolika je važnost vrha i u mreži. Što je vrijednost totalne podgrafovske komunikativnosti veća za pojedini vrh i , to je taj vrh važniji za protok informacija u mreži.

3.4.1 Usporedba podgrafovske centralnosti i totalne podgrafovske komunikativnosti

Iako su metode za računanje podgrafovske centralnosti i totalne podgrafovske komunikativnosti slične, rangovi koje dobijemo primjenom tih metoda u pravilu nisu isti. Razliku između ta dva rangiranja za zadani vrh i možemo izraziti s

$$\sum_{j=1}^n [e^A]_{ij} - [e^A]_{ii} = \sum_{\substack{j=1 \\ j \neq i}}^n [e^A]_{ij} = \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{k=1}^n e^{\lambda_k} \mathbf{v}_{ki} \mathbf{v}_{kj}.$$

gdje je v_{ik} i -ti element normaliziranog svojstvenog vektora \mathbf{v}_k matrice susjedstva A pridruženog svojstvenoj vrijednosti λ_k . Važno je primijetiti da u e^A su vrijednosti na dijagonali često veće nego vrijednosti koje nisu na dijagonali. Ako vrijednosti dijagonale od e^A poprimaju vrijednosti iz širokog raspona, dok su sume redova bez dijagonala u užem rasponu vrijednosti, tada se rangovi dobiveni ovim metodama ne razlikuju mnogo. Za neke jednostavnije grafove, kao što su potpuni grafovi i zvjezdani grafovi¹, ove dvije metode dat će nam jednako rangiranje. Rezultate primjene svih metoda na potpuni i zvjezdani graf mogu se pogledati u (4.2).

Jedan od slučajeva kada bi ove dvije metode mogle dati slične rezultate je kada imamo veliku razliku, $\lambda_1 - \lambda_2$, između prve dvije svojstvene vrijednosti matrice susjedstva A . Imamo

$$[e^A]_{ii} = e^{\lambda_1} \mathbf{v}_{1i}^2 + \sum_{k=2}^n e^{\lambda_k} \mathbf{v}_{ki}^2$$

i

$$[e^A \mathbf{1}]_i = e^{\lambda_1} (\mathbf{v}_1^T)_i \mathbf{v}_{1i}^2 + \sum_{k=2}^n e^{\lambda_k} (\mathbf{v}_i^T)_k \mathbf{v}_{ki}.$$

Ukoliko podijelimo oba izraza sa e^{λ_1} i uočimo da vrijedi $\mathbf{v}_1^T = \|\mathbf{v}_1\|_1$ tada, za vrijednosti $\lambda_1 \gg \lambda_2$, oba rangiranja su uvelike određena sa vrijednostima \mathbf{v}_{1i}^2 i $\|\mathbf{v}_1\|_1 \mathbf{v}_1$, dakle određena je s v_{1i} dominantnog svojstvenog vektora od A . Dakle, ako je razlika $\lambda_1 - \lambda_2$ dovoljno velika, ove dvije metode svode se na centralnost svojstvenih vektora predstavljenu u ([2]), te možemo očekivati slično rangiranje. Ipak, teško je unaprijed odrediti kolika bi trebala biti razlika $\lambda_1 - \lambda_2$ da bi rangovi dobiveni na ovaj način bili identični ili barem približno isti.

3.4.2 Totalna mrežna komunikativnost

Totalna podgrafovska komunikativnost govori nam koliko je pojedini vrh važan za protok informacija u mreži. Ukoliko nas zanima kakav je protok informacija kroz cijelu mrežu možemo koristiti *totalnu mrežnu komunikativnost* kao mjeru centralnosti. Kako bih odredili protočnost informacija kroz cijelu mrežu možemo pozabrajati zasebne totalne podgrafovske centralnosti pojedinih vrhova. Za mrežu, gdje je A pripadna matrica susjedstva, *totalnu komunikativnost mreže* računamo

$$C(A) = \sum_{i=1}^n [e^A \mathbf{1}]_i = \sum_{i=1}^n \sum_{k=1}^n e^{\lambda_k} (\mathbf{v}_i^T)_k \mathbf{v}_{ki} = \mathbf{1}^T e^A \mathbf{1}, \quad (3.8)$$

¹Zvjezdani graf sa n vrhova je graf u kojemu je jedan vrh centralni i on je povezan sa preostalim $n - 1$ vrhova i to su svi vrhovi u grafu

gdje λ_k je k -ta svojstvena vrijednost matrice A i v_{ik} je i -ti element normaliziranog svojstvenog vektora \mathbf{v}_k pridruženog svojstvenoj vrijednosti λ_k . Primjetimo da, iako je $C(A)$ definirana kao suma svih totalnih podgrafovskih komunikativnosti pojedinih vrhova, $[e^A]_i$, za i između 1 i n , nije potrebno znati pojedine vrijednosti kako bi izračunali $C(A)$. Postoje vrlo efikasne metode za računanje kvadratnih formi tipa $\mathbf{v}^T f(A) \mathbf{v}$ za danu funkciju $f(x)$, matricu A i vektor \mathbf{v} opisane u ([1]).

Zanimljivo je usporediti totalnu mrežnu komunikativnost sa Estradinim indeksom. *Estradin indeks* definira se kao suma svih podgrafovskih centralnosti

$$EE(A) = \sum_{i=1}^n [e^A]_{ii} = \sum_{i=1}^n e^{\lambda_i} = \text{Tr}(e^A). \quad (3.9)$$

Sljedećom propozicijom dajemo gornju i donju ogradu za vrijednost totalne mrežne komunikativnosti u okvirima Estradinog indeksa te drugim vrijednostima vezanim uz promatranu mrežu.

Propozicija 3.4.1. *Neka je A matrica susjedstva jednostavne mreže sa n vrhova. Tada,*

$$EE(A) \leq C(A) \leq ne^{\|A\|_2},$$

gdje $\|A\|_2$ označava 2-normu. Za neusmjerenu mrežu imamo

$$EE(A) \leq C(A) \leq ne^{\lambda_1}.$$

Dokaz. Donja granica je trivijalna. Naime,

$$EE(A) = \sum_{i=1}^n [e^A]_{ii} \leq \sum_{i=1}^n \sum_{j=1}^n [e^A]_{ij} = \sum_{i=1}^n [e^A \mathbf{1}] = C(A).$$

Gornja granica slijedi iz

$$C(A) = \mathbf{1}^T e^A \mathbf{1} = (e^A \mathbf{1})^T \mathbf{1} = \langle e^A \mathbf{1}, \mathbf{1} \rangle,$$

te primjenom Cauchy-Swartzove nejednakosti na kvadratnu formu $\langle e^A \mathbf{1}, \mathbf{1} \rangle$. Imamo

$$|\langle e^A \mathbf{1}, \mathbf{1} \rangle| \leq \|e^A \mathbf{1}\|_2 \|\mathbf{1}\|_2 \leq \|e^A\|_2 \|\mathbf{1}\|_2 \|\mathbf{1}\|_2 = ne^{\|A\|_2}.$$

Za neusmjereni graf vrijedi da je simetričan te $\lambda_1 = \|A\|_2$ iz čega slijedi tvrdnja propozicije. \square

Primjetimo da smo donju granicu dobili promatrajući graf u kojem ni jedna dva vrha nisu povezana, tj. matrica susjedstva izgleda $A = 0$, dok smo za gornju granicu gledali potpuni graf čija matrica susjedstva izgleda $A = \mathbf{1}\mathbf{1}^T - I$.

Rezultat prethodne propozicije vrijedi i za $e^{\beta A}, \beta > 0$. Za sve povezane grafove s pripadnim matricama susjedstva A , granice se sužavaju kako se β približava 0. Parametar β možemo shvatiti kao odraz vanjskih smetnji na mrežu. (Detaljnije informacije se mogu pronaći u ([6]))

Kada je primjereno normalizirana, $C(A)$ možemo koristiti kako bi usporedili lakoću prijenosa informacija u različitim mrežama što nam može biti od koristi prilikom dizajniranja komunikacijskih mreža.

3.5 Katzova centralnost

Katzova centralnost iskazuje relativni utjecaj svakog vrha zadane mreže A uzimajući u obzir njegove neposredne susjede kao i vrhove povezane s neposrednim susjedima. Katzovu centralnost, predstavljenu u ([10]), za cijelu mrežu računamo

$$R(A; \alpha) = (I - \alpha A)^{-1} = I + \alpha A + \alpha^2 A^2 + \dots + \alpha^k A^k + \dots,$$

gdje je A matrica susjedstva promatrane mreže, a α parametar ograničen sa $0 \leq \alpha \leq \frac{1}{\lambda_{\max}(A)}$. $\lambda_{\max}(A)$ je dominantna svojstvena vrijednost matrice susjedstva A . Parametar α osigurava da je matrica $R(A; \alpha)$ dobro definirana, $(I - \alpha A)$ je invertabilna, i nenegativna.

Katzovu centralnost pojedinog vrha i možemo izraziti

$$C_{Katz}(i) = \sum_{k=1}^{\infty} \sum_{j=1}^n \alpha^k (A^k)_{ji}.$$

Budući da Katzova centralnost ovisi o parametru α , zanimljivo je promatrati kako se Katzova centralnost ponaša ovisno o mijenjanju parametra unutar granica $0 \leq \alpha \leq \frac{1}{\lambda_{\max}(A)}$. U ([12]) , pokazano je kako mijenjanje parametra α mijenja Katzovu centralnost.

Ukoliko za parametar α vrijedi $\alpha \rightarrow 0$ s desna, tada se Katzova centralnost svodi na stupanjsku centralnost. Stupanjska centralnost vrha i daje najveću važnost poveznicama koje su udaljene jedan korak od i .

Kada se parametar α približava gornjoj granici $\alpha \rightarrow \frac{1}{\lambda_{\max}(A)}$ s lijeva, tada se Katzova centralnost svodi na centralnost svojstvenih vektora.

3.5.1 Katzova centralnost u digrafima

Za digrafe ponovo ćemo koristiti matricu \mathcal{A} za bipartitivni graf predstavljenu u (3.5). Katzovu centralnost za matricu \mathcal{A} dobivamo

$$R(\mathcal{A}, \alpha) = \begin{pmatrix} (I - \alpha^2 AA^T)^{-1} & \alpha A(I - \alpha^2 A^T A)^{-1} \\ \alpha (I - \alpha^2 AA^T)^{-1} A^T & (I - \alpha^2 A^T A)^{-1} \end{pmatrix}. \quad (3.10)$$

gdje je A matrica susjedstva u originalnoj mreži, a α parametar između 0 i $\frac{1}{\sigma_1}$; $\sigma_1 = \|A\|_2$ najveća svojstvena vrijednost od A . S parametrom α osiguravamo da je matrica (3.10) dobro definirana i nenegativna s pozitivnim dijagonalama. Dijagonale matrice $(I - \alpha^2 AA^T)^{-1}$ daju vrijednosti hubova matrice A dok dijagonale matrice $(I - \alpha^2 A^T A)^{-1}$ daju vrijednosti autoriteta.

Nedostatak ove metode, kao i u slučaju neusmjerenog grafa, je to što ovisi o parametru α . Ukoliko se parametar α približava gornjoj granici, dobivamo rezultate koji ne odstupaju previše od rezultata dobivene gledanjem dijagonala $e^{-\mathcal{A}}$, a ako se α približava nuli, tada su rezultati strogo korelirani s ulaznim i izlaznim stupnjevima pojedinih vrhova. Treba napomenuti ukoliko za α uzmemo premalu vrijednost, tada nam se vrijednosti hubova i autoriteta približavaju identiteti te nije moguće izvući nikakve zaključke o vrhovima. Budući da Katzova metoda više težine stavlja na kraće šetnje nego podgrafovska centralnost i duge šetnje doprinose relativno malo, može se reći da je podgrafovska centralnost manje pristrana što se tiče stupnjeva vrhova. Također, podgrafovska centralnost ne ovisi o parametru pa samim time daje finalne rangove.

Dodajmo još da Katzove vrijednosti autoriteta i hubova također možemo dobiti gledajući redove i stupce nesimetrične matrice $(I - \alpha A)^{-1}$ gdje je A originalni digraf i $\alpha > 0$, ali dovoljno malen da pripadni Neumannov niz² konvergira. Suma reda $(I - \alpha A)^{-1}$ broji šetnje koje izlaze iz vrha, dok stupci broje šetnje koje ulaze u pojedini vrh. Označimo li sa $\mathbf{1}$ vektor jedinica, rangiranje hubova i autoriteta možemo dobiti rješavanjem sljedećih jednadžbi

$$(I - \alpha A)\mathbf{y} = \mathbf{1}; \quad (I - \alpha A^T)\mathbf{x} = \mathbf{1},$$

gdje α zadovoljava sljedeći uvjet: $0 < \alpha < \frac{1}{\rho(A)}$; $\rho(A)$ je spektralni radijus matrice A (maksimum apsolutne vrijednosti po svim svojstvenim vrijednostima od A).

²Neumannov niz je niz definiran s $\sum_{k=0}^{\infty} T^k$ gdje je T kvadratna matrica i T^k k -ta potencija

Poglavlje 4

Usporedba metoda

U slijedećem poglavlju kratko ćemo predstaviti *pythonov* paket *networkx*, usporedit ćemo proučavane metode na par jednostavnih primjera te za kraj ćemo napraviti usporedbu metoda centralnosti na skupu podataka likova iz Marvelovog svijeta.

4.1 Networkx

Numeričku realizaciju proučavanih metoda ostvariti ćemo u *pythonu*. *Python* je snažan programski jezik koji omogućuje jednostavnu i fleksibilnu reprezentaciju mreža te jasne i koncizne izraze mrežnih algoritama. *NetworkX* je *Pythonov* paket za stvaranje, rukovanje i proučavanje strukture, dinamike i funkcija složenih mreža. Neke od osnovnih funkcionalnosti su:

- alati za proučavanje strukture i dinamike društvenih, bioloških i infrastrukturnih mreža
- standardno programsko sučelje i implementacija grafova koja je prikladna za mnoge aplikacije
- sučelje za postojeće numeričke algoritme i kod napisan u C, C++ i FORTRAN
- sposobnost bezbolnog rada s velikim nestandardnim skupovima podataka

S *NetworkX* bibliotekom moguće je učitavati i pohranjivati mreže u standardnim i nestandardnim formatima podataka, generirati mnoge vrste slučajnih i klasičnih mreža, analizirati mrežnu strukturu, graditi mrežne modele, dizajnirati nove mrežne algoritme, crtati mreže i još mnogo toga. Za više informacija o paketu *Networkx* može se posjetiti stranica dokumentacije¹.

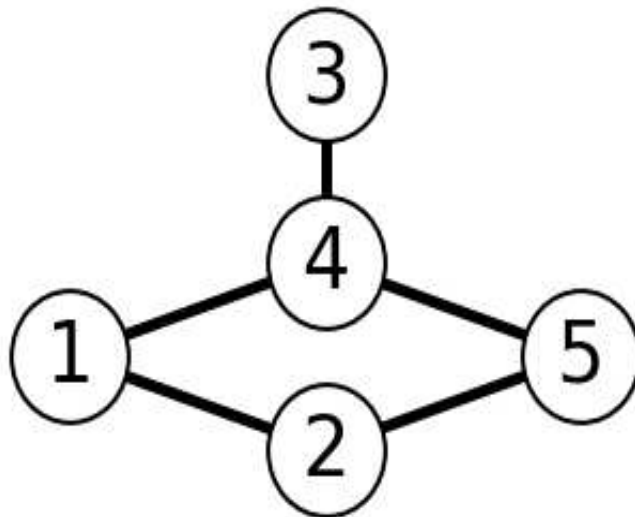
¹<https://networkx.org/documentation/stable/>

4.2 Primjena na manjim mrežama

Iako je svrha ovog rada usporedba predstavljenih metoda u velikim mrežama pokažimo prvo kako metode funkcioniraju na manjim mrežama. Proći ćemo kroz par jednostavnijih primjera.

Primjer 1

Za prvi primjer poslužit će nam slijedeći graf:



Slika 4.1: Primjer grafa koji sadrži 5 vrhova, za rezultate dobivene korištenjem svih proučavanih metoda centralnosti pogledati u tablicu 4.1.

Kao što je predstavljeno u (2), prvi je korak kreirati matricu susjedstva. Dobivamo simetričnu matricu

$$A_1 = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix}.$$

Na matricu susjedstva primjenjujemo predstavljene metode. Dobivamo sljedeće rezultate:

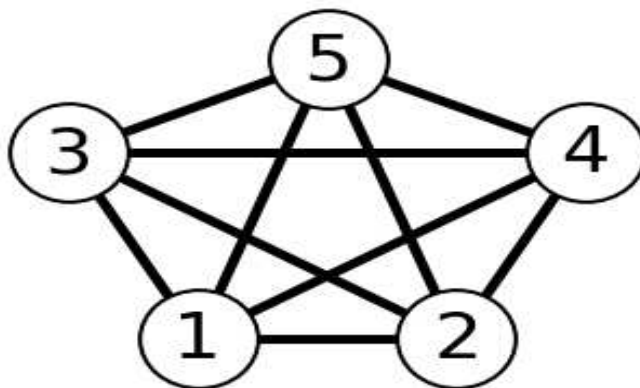
Id	Page Rank rang	Page Rank vrijednost	Podgrafovski centralnost rang	Podgrafovski centralnost vrijednost	Totalna komunikativnost rang	Totalna komunikativnost vrijednost	Katz rang	Katz vrijednost
v_1	3	0.197389	2	2.436826	2	8.497891	2	0.463712
v_2	2	0.197780	4	2.387614	4	7.968071	4	0.435160
v_3	5	0.113332	5	1.644596	5	5.177843	5	0.267889
v_4	1	0.294111	1	3.130633	1	10.331902	1	0.555710
v_5	3	0.197389	2	2.436826	2	8.497891	2	0.463712

Tablica 4.1: Vrijednosti centralnosti vrhova za sve vrhove te njihov rang u pojedinoj metodi.

Primjećujemo, sve metode za najcentralniji vrh odabiru vrh v_4 te kao najmanje važan vrh odabire se vrh v_3 . Također, primjetimo da vrhovi v_1 i v_5 imaju iste vrijednosti u svakoj promatranoj metodi. Ako pogledamo detaljnije u zadani graf, vidimo da, uistinu, vrhovi v_1 i v_5 imaju istu ulogu u smislu oba su povezana i sa vrhom v_4 i v_2 te nisu povezani međusobno niti sa vrhom v_3 .

Primjer 2

Za sljedeći primjer pogledajmo jedan potpuni graf.



Slika 4.2: Primjer potpunog grafa s pet vrhova, za ovakav graf očekujemo jednake rezultate za sve vrhove.

Dobiveni rezultati nalaze se u tablici ispod.

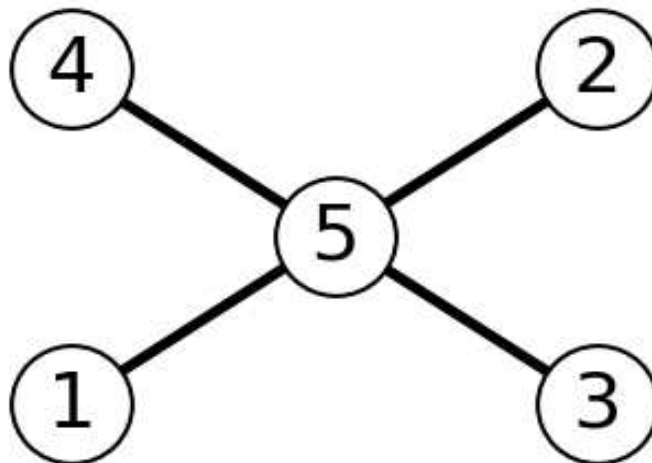
Id	Page Rank rang	Page Rank vrijednost	Podgrafovski centralnost rang	Podgrafovski centralnost vrijednost	Totalna komunikativnost rang	Totalna komunikativnost vrijednost	Katz rang	Katz vrijednost
v_1	1	0.2	1	11.213934	1	54.59815	1	0.447214
v_2	1	0.2	1	11.213934	1	54.59815	1	0.447214
v_3	1	0.2	1	11.213934	1	54.59815	1	0.447214
v_4	1	0.2	1	11.213934	1	54.59815	1	0.447214
v_5	1	0.2	1	11.213934	1	54.59815	1	0.447214

Tablica 4.2: Vrijednosti centralnosti vrhova za sve vrhove te njihov rang u pojedinoj metodi.

Za potpuni graf primjećujemo da svi vrhovi imaju jednaku vrijednost neovisno o tom koju metodu odaberemo. Dakle, u potpunom grafu ne postoji najcentralniji vrh budući da svi vrhovi mogu dosegnuti bilo koji drugi vrh u samo jednom koraku.

Primjer 3

Za sljedeći primjer pogledajmo jedan zvjezdani graf.



Slika 4.3: Zvjezdani graf s pet vrhova. Za ovaj graf očekujemo da će vrijednosti centralnog vrha v_5 biti najveće, dok za preostale vrhove neće biti razlike u rangiranju.

Pogledajmo dobivene rezultate.

Id	Page Rank rang	Page Rank vrijednost	Podgrafovska centralnost rang	Podgrafovska centralnost vrijednost	Totalna komunikativnost rang	Totalna komunikativnost vrijednost	Katz rang	Katz vrijedost
v_1	2	0.131081	2	1.690549	2	5.575626	2	0.434122
v_2	2	0.131081	2	1.690549	2	5.575626	2	0.434122
v_3	2	0.131081	2	1.690549	2	5.575626	2	0.434122
v_4	2	0.131081	2	1.690549	2	5.575626	2	0.434122
v_5	1	0.475677	1	3.762196	1	11.015917	1	0.496139

Tablica 4.3: Vrijednosti centralnosti vrhova za sve vrhove te njihov rang u pojedinoj metodi.

Za zvjezdani graf vidimo da su sve metode kao najcentralniji vrh odabrale središnji vrh, vrh v_5 . Svi preostali vrhovi imaju jednake vrijednosti centralnosti, tj. ne postoje razlike između njih. Ovakvi rezultati su očekivani budući da iz vrha v_5 možemo doći do bilo kojeg drugog vrha u jednom koraku, a ako se nalazimo u nekom drugom vrhu do svih ostalih vrhova koji nisu v_5 možemo doći jedino preko vrha v_5 .

4.3 Primjena na skupu podataka

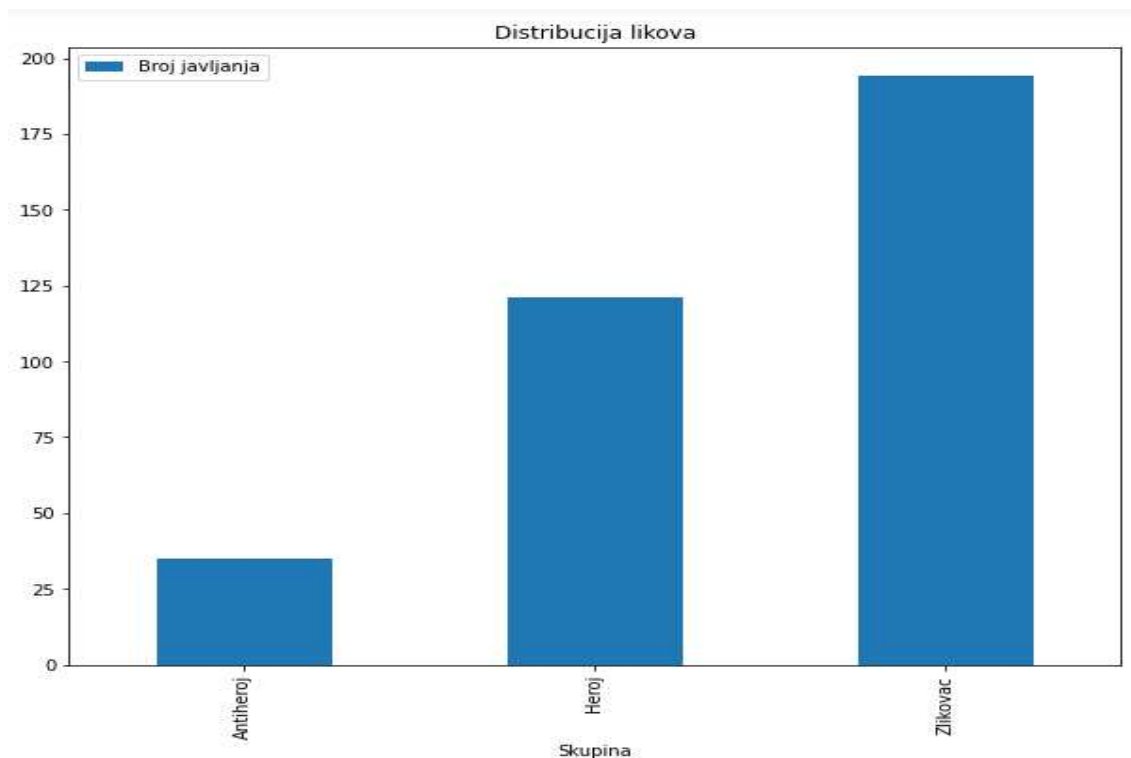
Primjenu na stvarnim podacima napraviti ćemo koristeći skup podataka koji sadrži informacije o likovima iz Marvelovog svijeta. Set podataka koji ćemo koristiti preuzet je sa <https://www.kaggle.com/>², točan link nalazi se u napomeni. U skupu podataka pojavljuju se razni likovi iz Marvelovih filmova koji su bazirani na stripovima ili dolaze iz samih stripova. Razlog zašto bi se netko mogao zapitati koji je od likova iz Marvelovog svijeta najcentralniji, bio bi, primjerice: Želimo napraviti novi film i zanima nas koji je lik najviše povezan s drugim likovima, računajući da bi na taj način mogli pogoditi najširu publiku za naš potencijalni novi film.

4.3.1 Opis podataka

Ukupno postoji 350 različitih likova, koji su podijeljeni u tri kategorije: heroji, zlikovci i antiheroji. Heroji su, uglavnom, hrabri likovi koji stoje na strani dobra te štite ljude od zlih ljudi, raznih čudovišta ili nepogoda. Zlikovci su okrutne, zlonamjerne osobe koje su upletene ili posvećene zloći ili zločinu. Antiheroji su likovi koji pokazuju znakove i heroja i zlikovca, često prave moralno korektne odluke ali su im razlozi zbog kojih čine krivi.

Pogledajmo kako izgleda distribucija likova ovisno o grupi kojoj pripadaju.

²Poveznica na skup podataka: <https://www.kaggle.com/datasets/trnguyen1510/the-marvel-comic-characters-partnerships>.



Slika 4.4: Distribucija likova po skupinama.

Vidimo da imamo najviše zlikovaca (55.43%), zatim heroja (34.57%) te najmanje antiheroja (10%).

Sljedeće što bi bilo zanimljivo pogledati je distribucija stupnjeva vrhova u mreži, tj koliko pojedini lik ima poveznica sa drugim likovima u mreži.

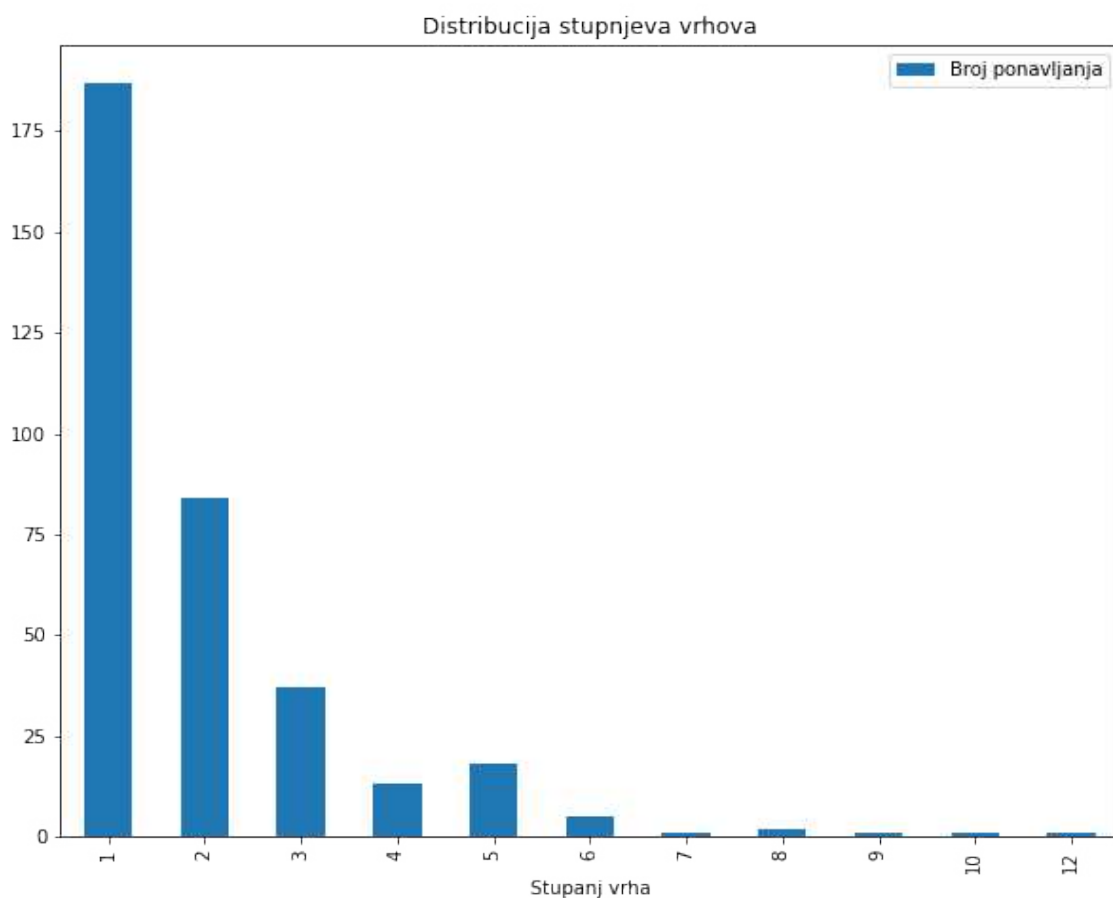
53.42% likova ima poveznicu samo s još jednim likom, a po jedan lik ima poveznicu sa sedam, devet, deset ili čak 12 drugih likova. Primjećujemo da je mreža poprilično rijetka. Pogledajmo sljedeće kako su povezani likovi unutar pojedinih kategorija.

Na slici 4.6a vidimo da 55.37% heroja ima po jednu poveznicu sa drugim likovima. Jedan heroj ima poveznicu sa 12 likova, primjećujemo da je to ujedno i najpovezaniji čvor u mreži.

54.28% antiheroja ima po jednu poveznicu. Najpovezaniji antiheroj ima 4 poveznice (slika 4.6b).

Što se tiče zlikovaca, 53.42% ima poveznicu samo s jednim likom. Najpovezaniji zlikovac ima 9 poveznica s drugim likovima i postoji samo jedan takav (slika 4.7).

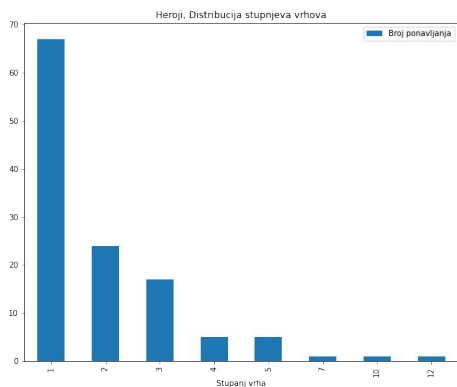
Vidimo da je povezanost likova unutar pojedinih grupa, kao i u cijeloj mreži, poprilično rijetka to jest većina likova, preko 50%, unutar pojedine grupe ima po jednu poveznicu.



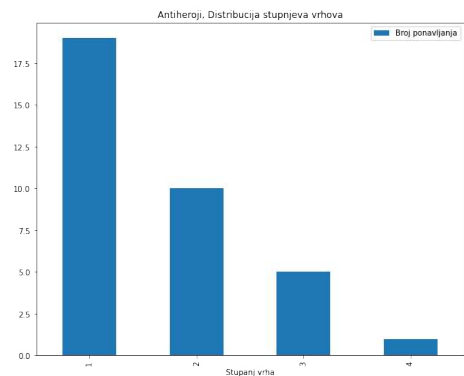
Slika 4.5: Distribucija povezanosti likova.

4.3.2 Rezultati

Nakon što smo napravili uvodnu opisnu analizu naših podataka pogledajmo kakve su mjere centralnosti poprimili likovi u pojedinoj metodi. Za svaku metodu ćemo ispisati deset najbolje rangiranih likova. Prva spomenuta metoda je *Page Rank* pa pogledajmo kakve rezultate dobivamo dok primjenimo *Page Rank* algoritam na naše podatke.

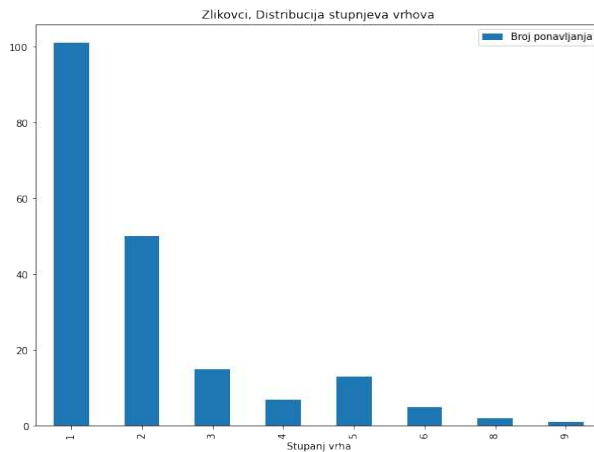


(a) Heroji



(b) Antiheroji

Slika 4.6: Distribucija povezanosti heroja i antiheroja.



Slika 4.7: Distribucija povezanosti zlikovaca.

Na vrhu liste nalaze se 2 heroja: *Spiderman* i *Captain America*. *Spiderman* ima 12 poveznica s drugim likovima dok *Captain America* ima 10 poveznica s drugim likovima. Zatim slijedi 8 zlikovaca. Zanimljivo je primijetiti da se na 4 poziciji nalazi *Selene* koja ima samo 5 poveznica, s obzirom da svi ostali u top 10, osim *Dormammu*-a, imaju više poveznica. Visoka pozicija *Selene* je zanimljiva jer su njeni susjedi povezani samo s njom, točnije rečeno, čine zasebnu podmrežu.

Pogledajmo rezultate dobivene korištenjem podgrafovske centralnosti.

Id	Lik	Skupina	Page Rank
194	Spiderman	Heroj	0.010886
302	Captain America	Heroj	0.010576
306	Red Skull	Zlikovac	0.009779
12	Selene (Comics)	Zlikovac	0.008117
220	Unicorn (Comics)	Zlikovac	0.007717
276	Grim Reaper (Comics)	Zlikovac	0.007564
249	Chameleon (Comics)	Zlikovac	0.007347
239	Attuma	Zlikovac	0.007218
311	Loki (Comics)	Zlikovac	0.007098
211	Dormammu	Zlikovac	0.006837

Tablica 4.4: Top 10 rezultata dobivenih korištenjem *Page Rank* algoritma.

Id	Lik	Skupina	Page Rank
194	Spiderman	Heroj	23.348845
111	Demogoblin	Zlikovac	20.819471
234	Shierk (Comics)	Zlikovac	20.819471
338	Doppelganger (Comics)	Zlikovac	20.819471
97	Carrion (Comics)	Zlikovac	20.819471
296	Carnage (Comics)	Zlikovac	17.295826
302	Captain America	Heroj	15.748518
229	Cletus Kasady	Zlikovac	15.285853
239	Attuma	Zlikovac	15.173122
220	Unicorn (Comics)	Zlikovac	14.092879

Tablica 4.5: Top 10 rezultata dobivenih korištenjem podgrafovske centralnosti.

Ponovo se na vrhu liste nalazi *Spiderman*, Zanimljivo je primijetiti da svi likovi od 2. do 5. pozicije imaju po 5 poveznica. Gledanjem koje su to poveznice primjećujemo da su svih 5 zlikovaca, *Demogoblin*, *Shriek*, *Doppelganger*, *Carrion* i *Carnage*, usko povezana. Točnije rečeno, povezani su svaki sa svakim, jedina je iznimka povezanost između *Carnagea* i *Shrieka*. *Carnage* za svoju petu poveznicu ima *Venoma*, koji je direktno povezan sa *Spider Manom*, najutjecajnijim članom istraživane mreže na temelju podgrafovske centralnosti. Također zanimljivo je primijetiti da se u top 10 likova po podgrafovskoj centralnosti nalazi i *Cletus Kasady* koji ima samo četiri poveznice. Razlog tome je što je *Cletus Kasady* povezan sa već ranije spomentom skupinom, *Demogoblin*, *Shriek*, *Doppelganger* i *Carrion*.

Što se tiče totalne podgrafovske komunikativnosti, rezultati su sljedeći:

Id	Lik	Skupina	totalna podgrafovka komunikativnost
194	Spiderman	Heroj	184.537606
296	Carnage (Comics)	Zlikovac	136.699878
111	Demogoblin	Zlikovac	132.750018
97	Carrion (Comics)	Zlikovac	132.750018
234	Shierk (Comics)	Zlikovac	132.750018
338	Doppelganger (Comics)	Zlikovac	132.750018
10	Venom (Marvel Comics character)	Zlikovac	123.191531
229	Cletus Kasady	Zlikovac	109.322832
308	Silk (Comics)	Heroj	100.533177
306	Red Skull	Zlikovac	95.325169

Tablica 4.6: Top 10 rezultata dobivenih korištenjem totalne podgrafovke komunikativnosti.

Ponovo se na vrhu liste nalazi *Spiderman*. Također opet imamo ranije spomenutu petorku, *Demogoblin*, *Shriek*, *Doppelganger*, *Carrion* i *Carnage*, visoko na ljestvici, od druge do pete pozicije. Za razliku od podgrafovke centralnosti vidimo da je *Carnage* ovaj put na drugoj poziciji, dok je po podgrafovskoj centralnosti bio na šestoj poziciji. Kada se malo razmisli o tome što naše metode za mjerenje centralnosti zapravo rade, ovakav rezultat ima smisla. Naime, u podgrafovskoj centralnosti gledamo kolika je šansa da se u danoj nasumičnoj šetnji nalazimo upravo u zadanom vrhu, dok u totalnoj podgrafovskoj komunikativnosti gledamo koliko smo povezani sa svim čvorovima u mreži. Dakle, možemo se pronaći u danom vrhu, *Carnage*, na više različitih načina budući da je usko vezan uz svoje susjede, no ipak totalna podgrafovka centralnost mu daje veću vrijednost budući da ima poveznicu s *Venomom* koji je kasnije povezan sa *Spidermanom*, najpovezanijim čvorom u mreži.

Pogledajmo još rezultate dobivene Katzovom centralnošću. Katzova centralnost ovisi o parametru α za koji mora vrijediti $0 \leq \alpha \leq \frac{1}{\lambda_{\max}(A)}$. Uzmimo slučajeve dok je $\alpha = \frac{1}{\lambda_{\max}(A)} * 0.1$ i $\alpha = \frac{1}{\lambda_{\max}(A)} * 0.9$.

Id	Lik	Skupina	Katz
194	Spiderman	Heroj	0.065002
302	Captain America	Heroj	0.062457
306	Red Skull	Zlikovac	0.061446
239	Attuma	Zlikovac	0.060366
220	Unicorn (comics)	Zlikovac	0.060295
163	Iron Fist (Comics)	Zlikovac	0.059134
10	Venom (Marvel Comics character)	Zlikovac	0.058439
249	Chameleon (comics)	Zlikovac	0.058091
113	Scarecrow (Marvel Comics)	Zlikovac	0.058080
311	Loki (comics)	Zlikovac	0.058018

Tablica 4.7: Top 10 rezultata dobivenih korištenjem Katzove centralnosti s parametrom $\alpha = \frac{1}{\lambda_{\max}(A)} * 0.1$.

Id	Lik	Skupina	Katz
194	Spiderman	Heroj	0.237385
97	Carrion (comics)	Zlikovac	0.223714
111	Demogoblin	Zlikovac	0.223714
234	Shriek (comics)	Zlikovac	0.223714
338	Doppelganger (comics)	Zlikovac	0.223714
296	Carnage (comics)	Zlikovac	0.220675
229	Cletus Kasady	Zlikovac	0.188543
10	Venom (Marvel Comics character)	Zlikovac	0.169477
308	Silk (comics)	Heroj	0.134745
306	Red Skull	Zlikovac	0.134073

Tablica 4.8: Top 10 rezultata dobivenih korištenjem Katzove centralnosti s parametrom $\alpha = \frac{1}{\lambda_{\max}(A)} * 0.9$.

Gledajući 4.7 ponovo se na vrhu pronalazi *Spidermana*. Zanimljivo je primijetiti, gledajući povezanost likova, da se najpovezaniji likovi nalaze na vrhu ljestvice dok se slabije povezani likovi nalaze bliže dnu ljestvice.

Gledajući 4.8 na vrhu ljestvice nalazi se *Spiderman*, te uz njega, sad već dobro poznata šestorka, *Carrion*, *Demogoblin*, *Shriek*, *Doppelganger*, *Carnage* i *Cletus Kasady*. Dobiveni rezultati Katzove centralnosti poklapaju se sa teoretskim očekivanjima. Naime, ranije je spomenuto za Katzovu centralnost ukoliko vrijedi $\alpha \rightarrow 0$ tada se Katzova centralnost svodi na stupanjsku centralnost, a kada se parametar α približava gornjoj granici

$\alpha \rightarrow \frac{1}{\lambda_{\max}(A)}$ tada se Katzova centralnost svodi na centralnost svojstvenih vektora. U tablici ispod 4.9 možemo vidjeti poredak likova poredanih po vrijednostima dobivenim gledajući centralnost svojstvenih vektora.

Id	Lik	Skupina	Centralnost svojstvenih vektora
97	Carrion (comics)	Zlikovac	0.418439
111	Demogoblin	Zlikovac	0.418439
234	Shriek (comics)	Zlikovac	0.418439
338	Doppelganger (comics)	Zlikovac	0.418439
296	Carnage (comics)	Zlikovac	0.378393
229	Cletus Kasady	Zlikovac	0.352596
10	Venom (Marvel Comics character)	Zlikovac	0.122438
194	SpiderMan	Heroj	0.074863
306	Red Skull	Zlikovac	0.043044
308	Silk (comics)	Heroj	0.033852

Tablica 4.9: Top 10 rezultata dobivenih korištenjem centralnosti svojstvenih vektora

Također, primjećujemo dosta preklapanja između poredaka dobivenih korištenjem centralnosti svojstvenih vektora i Katzovom centralnošću s parametrom $\alpha = \frac{1}{\lambda_{\max}(A)} * 0.9$. Sažmimo sada sve dobivene rezultate.

Id	Lik	Skupina	Page Rank rang	Podgrafovska centralnost rang	Totalna podgrafovska komunikativnost rang	Katz, 0.1 rang	Katz, 0.9 rang	prosječni rang
194	Spiderman	Heroj	1	1	1	1	1	1
302	Captain America	Heroj	7	12	2	2	12	7
306	Red Skull	Zlikovac	11	10	3	3	10	7.4
239	Attuma	Zlikovac	9	11	8	4	11	8.6
220	Unicorn (comics)	Zlikovac	10	13	5	5	13	9.2
10	Venom (Marvel Comics character)	Zlikovac	17	7	18	7	8	11.4
163	Iron Fist (comics)	Heroj	12	19	12	6	18	13.4
113	Scarecrow (Marvel Comics)	Zlikovac	13	18	16	9	19	15
308	Silk (comics)	Heroj	14	9	35	13	9	16
249	Chameleon (comics)	Zlikovac	24	27	7	8	23	17.8

Tablica 4.10: Top 10 rezultata dobivenih uzimanjem prosječnog rangiranja za svaku promatranu metodu računanja centralnosti

U tablici 4.10 vidimo rangove dobivene svakom proučavanom metodom, a u zadnjem stupcu vidimo prosječni rang kroz sve metode. Podatci su poredani po njihovom prosječnom rangom. Svaka je metoda, dakle, dobila jednaku važnost. Uvjerljivo najbolji rang ima *Spiderman* s prosječnim rangom **1**. Prvi pratioci su *Captain America* s prosječnim rangom **7** i *Red Skull* s prosječnim rangom **7.4**. Često spominjana petorka, *Demogoblin*, *Shriek*, *Doppelganger*, *Carnage* i *Carrion* nalazi se na mjestima od 13. do 19. mjesta, gdje ih predvodi *Carnage* kao poveznica sa *Venomom* i samim time *Spidermanom*. Druga stvar koju ističemo su već spomenuta odstupanja u Katzovoj centralnosti ovisno o parametru α . Ukoliko vrijedi $\alpha \rightarrow 0$ tada se Katzova centralnost svodi na stupanjsku centralnost, a kada se parametar α približava gornjoj granici $\alpha \rightarrow \frac{1}{\lambda_{\max}(A)}$ tada se Katzova centralnost svodi na centralnost svojstvenih vektora. Također, ostaje pitanje koja Katzova centralnost je bolja, kada je $\alpha = \frac{1}{\lambda_{\max}(A)} * 0.1$ ili $\alpha = \frac{1}{\lambda_{\max}(A)} * 0.9$. No ta pitanje su predmet brojnih istraživanja koja su motivirana sve većim brojem primjena. Finalno, pogledajmo kakva je korelacija³ između rangiranja u pojedinim metodama i finalnog ranga. Očekujemo da korelacije budu visoke s obzirom da smo finalni rang dobili kao prosječna vrijednost svih pojdinih rangiranja, no zanima nas koje su metode najkoreliranije i najmanje korelirane s prosječnom vrijednosti rangova. Računajući Pearsonove koeficijente korelacije dobivamo da su najkoreliraniji Katzova centralnost s parametrom $\alpha = \frac{1}{\lambda_{\max}} * 0.1$ s koeficijentom korelacije 0.996, sljedeća je podgrafovska centralnost s koeficijentom 0.993, dok je najmanje koreliran *Page Rank* s koeficijentom 0.438.

³Pogledati na primjer u <https://web.math.pmf.unizg.hr/~rus/nastava/stat/SLAJD/7-korelacija.pdf>.

Poglavlje 5

Zaključak

U radu smo opisali metode centralnosti za proučavanje mreža koje su reprezentirane grafovima. Iznjeli smo koje su važnosti proučavanja metoda centralnosti i proučavanja mreža općenito. Predstavili smo neke od osnovnih metoda te potom detaljnije objasnili metode za mjerenje centralnosti: *Page Rank*, podgrafovska centralnost, totalna podgrafovska komunikativnost te Katzova centralnost. Za svaku metodu iznjeli smo ideju koja stoji u pozadini, pokazali kako se računaju centralnosti te nabrojali pojedine nedostatke metoda te su dani prijedlozi kako doskočiti tim nedostacima. Finalno, demonstrirali smo metode mjerenja centralnosti na par jednostavnih primjera mreža te primjeru povezanosti likova unutar Marvelovog svijeta. Za sve metode dobili smo sličan poredak na vrhu ljestvice. Vidjeli smo razliku između podgrafovske centralnosti i totalne komunikativnosti na primjeru *Carangea*, poveznice grupe sa ostatkom svijeta. Također, mijenjenjem parametra α u Katzovoj centralnosti primjetili smo očekivano ponašanje centralnosti; što je parametar α bio bliži donjoj granici 0, to je Katzova centralnost bila bliža stupanjskoj centralnosti, a što je bila bliže gornjoj granici $\frac{1}{\lambda_{\max}(A)}$ to je Katzova centralnost bila sličnija centralnosti svojstvenih vektora. Za sve metode *Spiderman* je bio na prvoj poziciji. *Spiderman* je ujedno bio i najpovezaniji lik u mreži. Odabir *Spidermana* kao lika nad kojim bi se radile neke akcije na mreži vjerojatno je dobra odluka, budući da je u vrijeme pisanja ovog rada *Marvel* najavio novi nastavak *Spiderman* sage koji bi trebao izaći ove godine.

Bibliografija

- [1] Michele Benzi i Paola Boito, *Quadrature rule-based bounds for functions of adjacency matrices*, *Linear Algebra and its Applications* **433** (2010), 637–652, <https://www.sciencedirect.com/science/article/pii/S002437951000162X>.
- [2] P. Bonacich, *Power and Centrality: A Family of Measures*. *American Journal of Sociology*, 92(5), 1170–1182., 1987, <http://www.jstor.org/stable/2780000>.
- [3] Kurt Bryan i Tanya Leise, *The \$25,000,000,000 Eigenvector: The Linear Algebra behind Google*, *Siam Review - SIAM REV* **48** (2006), 569–581, <https://www.rose-hulman.edu/~bryan/googleFinalVersionFixed.pdf>.
- [4] Jonathan Crofts, Ernesto Estrada, Desmond Higham i Alan Taylor, *Mapping directed networks*, *ETNA - Electronic Transactions on Numerical Analysis* **37** (2010), 337–350, ISSN 1068-9613, <https://strathprints.strath.ac.uk/29072/>.
- [5] Chris Ding, Hongyuan Zha, Xiaofeng He i Horst Simon, *Link Analysis: Hubs and Authorities on the World Wide Web*, *SIAM Review* **46** (2002), https://www.researchgate.net/publication/2543139_Link_Analysis_Hubs_and_Authorities_on_the_World_Wide_Web.
- [6] Michele Benzi Ernesto Estrada, Naomichi Hatano, *The physics of communicability in complex networks*, *Physics Reports, Volume 514, Issue 3*, 2012, <https://www.sciencedirect.com/science/article/pii/S0370157312000154>.
- [7] E. Estrada i J. A. Rodriguez-Velazquez, *Subgraph centrality in complex networks*, 2005, <https://journals.aps.org/pre/pdf/10.1103/PhysRevE.71.056103>.
- [8] D Fogaras, *Where to Start Browsing the Web?*, 2003, https://doi.org/10.1007/978-3-540-39884-4_6.
- [9] Aric A. Hagberg, Daniel A. Schult i Pieter J. Swart, *Exploring Network Structure, Dynamics, and Function using NetworkX*, Proceedings of the 7th Python in Science Conference (Pasadena, CA USA) (Gaël Varoquaux, Travis Vaught i Jarrod Millman, ur.), 2008, str. 11 – 15.

- [10] Leo Katz, *A new status index derived from sociometric analysis*, 1953, <https://www.sciencedirect.com/science/article/pii/S0370157312000154>.
- [11] J. Kleinberg, *Authoritative sources in a hyperlinked environment*, 1999, <https://dl.acm.org/doi/pdf/10.1145/324133.324140>.
- [12] Christine Klymko Michele Benzi, *On the limiting behavior of parameter-dependent network centrality measures*, 2013, <https://arxiv.org/abs/1312.6722>.
- [13] Ernesto Estrada Michele Benzi i Christine Klymko, *Ranking hubs and authorities using matrix functions*, 2012, <https://arxiv.org/pdf/1201.3120.pdf>.
- [14] Dr. V.C.Borkar Mohammed Abdullah Saleh Salman, *Exponential Matrix and Their Properties*, 2016, <https://www.arcjournals.org/pdfs/ijsimr/v4-i1/8.pdf>.
- [15] Ivica Nakić, *Diskretna matematika*, (2011), <https://web.math.pmf.unizg.hr/nastava/komb/predavanja/predavanja.pdf>.

Sažetak

U svijetu gdje smo svi povezani svakim danom sve više i više, nameće se zaključak da je proučavanje mreža vrlo bitan zadatak. Jedan od aspekata proučavanja mreža je određivanje mjera centralnosti koje pronalazi svakojake primjene. Za mjere centralnosti ne postoji jedna jedinstvena vrijednost koja će nam reći koji je vrh najbitniji u mreži, nego postoji više različitih metoda koje možemo koristiti. Svaka metoda ima svoje karakteristike te na temelju njih odabiremo metodu koja nam najviše odgovara te zatim proučavamo dobivene rezultate. Predstavili smo četiri metode mjerenja centralnosti, to su: *Page Rank*, *podgrafovska centralnost*, *Totalna komunikativnost* i *Katzova centralnost*. Za svaku metodu naveli smo ideju iza nje, osnovne karakteristike, kako ju računamo te njene nedostatke. Za kraj, primjenili smo navedene metode na primjeru Marvelovog svijeta te dobili smislene rezultate.

Summary

In a world where we are connected with each other more and more by each day, the conclusion is that studying networks is a very important task. One of the aspects of studying networks is determining measures of centrality that find all kinds of applications in real world use-cases. For measures of centrality there does not exist one unique value that will tell us which vertex is the most important in the network, rather there are several different methods that we can use. Each method has its own characteristics and based on them we choose the method that suits us best and then we study the results obtained. We have presented four methods of measuring centrality, namely: *Page rank*, subgraph centrality, total subgraph communicability and Katz's centrality. For each method we stated general ideas behind them, basic characteristics, how it's calculated and their shortcomings. Finally, we applied mentioned methods to example of the Marvel world and got meaningful results.

Životopis

Rođen sam 29. rujna 1995. godine u Osijeku. Osnovnu školu sam pohađao u Osijeku nakon koje sam upisao III. gimnaziju u Osijeku. Nakon završene srednje škole upisujem preddiplomski studij matematike na Matematičkom odsjeku Prirodoslovno-matematičkog fakulteta u Zagrebu. Nakon završenog preddiplomskog studija, upisujem diplomski studij Matematike i računarstva na istom odsjeku.

Ljeto 2019. godine proveo sam na praksi u firmi mStart radeći na projektu strojnog učenja. Od 2021. radim u firmi Nanobit kao analitičar podatka unutar igara.