

Primjene teorije grafova

Prkačin, Barbara

Master's thesis / Diplomski rad

2021

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:217:110930>

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

Download date / Datum preuzimanja: **2024-09-13**



Repository / Repozitorij:

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



SVEUČILIŠTE U ZAGREBU
PRIRODOSLOVNO–MATEMATIČKI FAKULTET
MATEMATIČKI ODSJEK

Barbara Prkačin

PRIMJENE TEORIJE GRAFOVA

Diplomski rad

Voditelj rada:
doc. dr. sc. Slaven Kožić

Zagreb, rujan, 2021.

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. _____

Zahvaljujem se mentoru doc. dr. sc. Slavenu Kožiću na pomoći i vodstvu tijekom izrade ovog diplomskog rada.

Hvala mojoj obitelji na strpljenju i podršci kroz sve ove godine.

Posebno hvala mojim predivnim drugovima i drugaricama koji su vjerovali u mene cijelo vrijeme. Bez vaše pomoći i poticaja ovo ne bi bilo moguće.

I'm grateful to my sister and brother-in-law, your help and motivation pushed me across the finish line.

Na kraju, hvala svim feministkinjama i feministima, vaša borba dala nam je priliku.

Sadržaj

Sadržaj	iv
Uvod	1
1 Teorija grafova	2
1.1 Osnovni pojmovi	2
1.2 Neke specifične topologije grafova	4
2 Najkraći putevi u grafu	7
2.1 Dijkstrin algoritam	8
2.2 Neke primjene algoritama za pronalazak najkraćeg puta	10
3 Bojanje grafova	12
3.1 Osnovni pojmovi	12
3.2 Kromatski brojevi za neke topologije grafova	15
3.3 Pohlepni algoritam za bojanje grafova	17
3.4 Neke primjene bojanja grafova	21
3.5 Dodjeljivanje kodova baznim stanicama telekomunikacijske mreže	23
4 Mjere centralnosti	28
4.1 Pregled mjera centralnosti	28
4.2 Primjer računanja PageRanka	34
4.3 Neke primjene mjera centralnosti	41
4.4 Analiza mreže citata znanstvenih članaka	44
Bibliografija	53

Uvod

Teorija grafova grana je diskretne matematike koja modelira složene probleme matematičkim objektima koje nazivamo grafovi. Početkom razvoja teorije grafova smatra se članak iz 1741. u kojem je Leonhard Euler opisao rješenje problema Sedam mostova Königsberga. Bilo je potrebno pronaći šetnju u kojoj se svaki od mostova u gradu Königsbergu pređe točno jednom. Euler je, modelirajući problem grafom, 1735. pokazao da rješenje ne postoji. Graf možemo promatrati kao matematičku strukturu koja predstavlja skup povezanih objekata koje nazivamo vrhovi, a poveznice između dva vrha nazivamo bridovi. Teorija grafova danas je iznimno raznoliko područje sa širokim primjenama. Pokazalo se da su grafovi izvrstan alat za modeliranje sustava koji naglasak stavljaju na veze i odnose među objektima. Ako obratimo pozornost, primijetiti ćemo da su problemi koje proučava teorija grafova zapravo posvuda oko nas. U ovom radu dat ćemo pregled nekoliko takvih problema koji su riješeni primjenom teorije grafova te opisati neke zanimljive primjere.

U prvom poglavlju upoznat ćemo se s osnovnim pojmovima teorije grafova koji su neophodni za razumijevanje preostalih poglavlja. U drugom poglavlju obradit ćemo problem najkraćeg puta, koji se kao jedan od najpoznatijih problema teorije grafova, često pojavljuje u telekomunikacijama i logistici. Opisat ćemo kako funkcionira Dijkstrin algoritam, jedan od najpoznatijih algoritama za pronalazak najkraćeg puta te navesti dvije zanimljive primjene. Sljedeća dva poglavlja glavni su fokus ovog rada i u skladu s time problemi u njima nešto su detaljnije obrađeni. Treće poglavlje bavi se jednim od najzanimljivijih problema u ovom području, bojanjem grafova. Povijest ovog problema kreće 1852. godine kada je Francis Guthrie iznio slutnju da su dovoljne četiri boje kako bi se geografska karta obojala na način da susjedna područja na karti budu različite boje. Iako je problem naizgled jednostavan, teorem četiri boje dokazan je tek 1976. godine uz pomoć računala. U ovom poglavlju pokazat ćemo kako je problem bojanja grafova moguće riješiti takozvanim pohlepnim algoritmom. U posljednjem poglavlju razmotrit ćemo netrivialan problem određivanja važnosti, tj. centralnosti nekog vrha u grafu. Brojne različite mjere centralnosti pokušavaju dati što bolji odgovor na pitanje: *Što određeni vrh u grafu čini centralnim, a neki drugi perifernim?* Mjere centralnosti nalaze važnu primjenu u analizi različitih društvenih mreža, pa se stoga često interpretiraju u tom kontekstu. Međutim, primjene mjera centralnosti uistinu su raznolike što ćemo pokazati kroz nekoliko različitih primjera.

Poglavlje 1

Teorija grafova

U ovom poglavlju ćemo iznijeti i objasniti osnovne pojmove teorije grafova koji su neophodni za razumijevanje ostalih poglavlja.

Glavni izvori ovog poglavlja su [8] i [30].

1.1 Osnovni pojmovi

Kako bismo mogli započeti izlaganje moramo najprije formalno definirati koncept grafa sljedećom definicijom.

Definicija 1.1.1. *Graf je uređeni par $G = (V, E)$, gdje je $V \neq \emptyset$ skup čije elemente zovemo **vrhovi**, a E familija dvočlanih podskupova od V koje zovemo **bridovi**.*

Prethodnu definiciju često proširujemo tako da dopustimo **petlje** (bridove koji spajaju vrh sa samim sobom), **višestruke bridove** (više bridova između para vrhova) i **usmjerene bridove** (bridovi koji imaju orijentaciju, tj. jedan vrh je početak, a drugi kraj brida). Petlje reprezentiramo dvočlanim multiskupom koji sadrži dva jednaka elementa. Usmjerene bridove reprezentiramo uređenim parovima, a ne dvočlanim podskupovima. Na primjer za usmjereni brid (u, v) , vrh u je početak, a vrh v kraj brida. Ako graf sadrži višestruke bridove, E postaje multiskup.

Definicija 1.1.2. *Graf koji ima usmjerene bridove nazivamo **usmjereni graf** ili **digraf**, a graf koji ima višestruke bridove nazivamo **multigraf**.*

Definicija 1.1.3. *Težinski graf $G = (V, E)$ je graf s funkcijom $f : E \rightarrow \mathbb{R}$, koja svakom bridu iz E pridružuje realan broj kojeg nazivamo **težina**. Funkciju f nazivamo **težinska funkcija**.*

Definicija 1.1.4. Za vrhove $u, v \in V$ grafa $G = (V, E)$ kažemo da su **susjedni** ako postoji $e = \{u, v\} \in E$, tj. $e = (u, v) \in E$ ako je G digraf. Za e kažemo da je **incidentan** s vrhovima u i v te pišemo $e = uv$.

Definicija 1.1.5. Za graf $G' = (V', E')$ kažemo da je **podgraf** grafa $G = (V, E)$ ako je $V' \subseteq V$ i $E' \subseteq E$. Podgraf oblika $G' = (V, E')$ nazivamo **razapinjući podgraf**.

Za svaki brid $e \in E'$ vrijedi da su oba njegova vrha u V' .

Definicija 1.1.6. **Stupanj vrha** v , u oznaci $\deg(v)$, u neusmjerenom grafu G je broj bridova koji su incidentni s v . Svaka petlja se računa kao dva brida.

Ulazni stupanj vrha v u digrafu D , u oznaci $\deg_{\text{in}}(v)$, je broj bridova oblika $e = (x, v)$.

Analogno, **izlazni stupanj vrha** $\deg_{\text{out}}(v)$ je broj bridova oblika $e = (v, x)$.

Za vrh stupnja nula kažemo da je **izolirani vrh**, dok je vrh stupnja jedan **list**.

Definicija 1.1.7. Za graf $G = (V, E)$ kažemo da je **regularan** ako su svi njegovi vrhovi istog stupnja. Za G kažemo da je **k -regularan** ako za svaki vrh $v \in V$ vrijedi $\deg(v) = k$. Prirodni broj k nazivamo **stupanj regularnosti** grafa G .

Definicija 1.1.8. Niz $W := (v_0, e_1, v_1, e_2, v_2, \dots, e_n, v_n)$ u grafu $G = (V, E)$, gdje je brid $e_i = v_{i-1}v_i$ za $i = 1, \dots, n$, nazivamo **šetnja**. Šetnju koja kreće iz vrha v_0 i završava u vrhu v_n označavamo s v_0Wv_n , a za vrhove v_0 i v_n kažemo da su **krajevi** šetnje.

Niz $W := (v_0, e_1, v_1, e_2, v_2, \dots, e_n, v_n)$ u digrafu $D = (V, E)$, gdje je brid $e_i = v_{i-1}v_i$ usmjereni brid s početkom u vrhu v_{i-1} i krajem u vrhu v_i za $i = 1, \dots, n$ nazivamo **usmjerena šetnja**.

Duljina šetnje je broj bridova u nizu (tj. broj vrhova u nizu manje jedan), odnosno suma vrijednosti težinske funkcije svih bridova u nizu za težinske grafove. Kažemo da je šetnja zatvorena ukoliko je $v_n = v_0$.

Šetnju u kojoj su svi vrhovi različiti (osim eventualno početnog i završnog vrha) nazivamo **put**, a zatvoreni put nazivamo **ciklus**.

Šetnju u kojoj su svi bridovi različiti nazivamo **staza**.

Definicija 1.1.9. Za vrhove u i v grafa G kažemo da su **povezani** ako postoji put uWv .

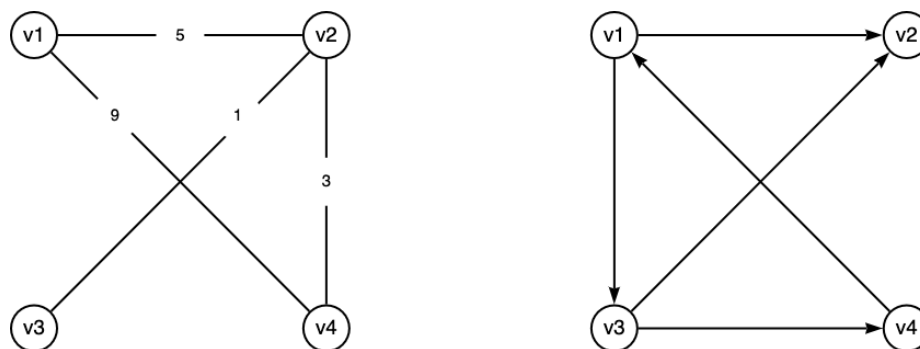
Graf je **povezan** ako su svi njegovi vrhovi povezani.

Digraf $D = (V, E)$ je **jako povezan** ako između svaka dva vrha $u, v \in V$ postoje usmjerene šetnje uWv i vWu .

Definicija 1.1.10. Neka je $G = (V, E)$ graf, $n = |V|$ broj vrhova te $V = \{v_1, v_2, \dots, v_n\}$. Matricu $\mathbf{A}_G \in M^{n \times n}$ koja na mjestu (i, j) ima vrijednost jednaku broju bridova incidentnih vrhovima v_i i v_j nazivamo **matrica susjedstva** grafa G .

Za digraf D , matrica susjedstva \mathbf{A}_D na mjestu (i, j) ima vrijednost jednaku broju bridova s početkom u v_i i krajem u v_j .

Grafovi se često opisuju njihovim grafičkim prikazom, koji treba biti takav da iz njega možemo rekonstruirati formalni zapis grafa oblika (V, E) . Općenito, smatramo da je graf zadan, ukoliko su zadani njegovi vrhovi, te ukoliko znamo koji su vrhovi međusobno povezani [30]. U grafičkom prikazu usmjerene bridove označavamo strelicom.



Slika 1.1: Težinski graf G lijevo i digraf D desno.

Primjer 1.1.11. Na slici 1.1 su neusmjereni težinski graf G i digraf D . Pripadajuće matrice susjedstva \mathbf{A}_G i \mathbf{A}_D su:

$$\mathbf{A}_G = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{A}_D = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Možemo primijetiti da je matrica susjedstva \mathbf{A}_G za neusmjeren graf simetrična, dok za matricu \mathbf{A}_D pridruženu digrafu to ne vrijedi.

1.2 Neke specifične topologije grafova

U ovom odjeljku opisati ćemo neke specifične tipove grafova koji će nam trebati u ostalim poglavljima.

Definicija 1.2.1. Graf $G = (V, E)$ za kojeg vrijedi $E = \emptyset$ nazivamo **prazan graf**. Prazan graf koji sadrži n vrhova označavamo s N_n .

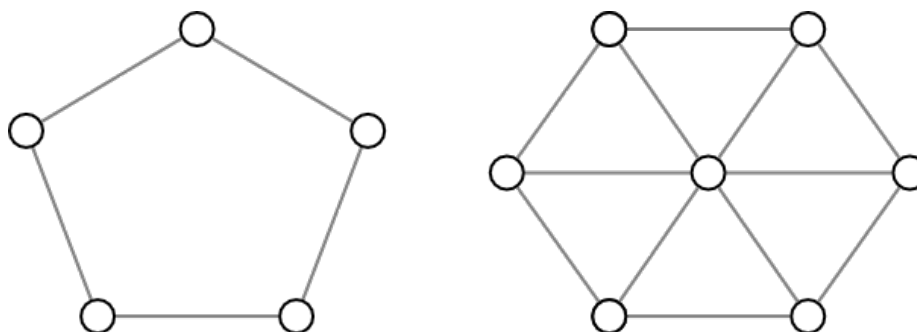
Definicija 1.2.2. Graf u kojem je svaki par vrhova spojen bridom nazivamo **potpuni graf**. Potpuni graf koji sadrži n vrhova označavamo s K_n .

Definicija 1.2.3. Podgraf $G' = (V', E')$ grafa $G = (V, E)$ nazivamo **klika** ako su svi vrhovi $v \in V'$ međusobno susjedni, tj. ako je G' potpuni graf.

Definicija 1.2.4. Skup vrhova $V' \subseteq V$ grafa $G = (V, E)$ nazivamo **nezavisni skup** ako za sve parove vrhova $u, v \in V'$ vrijedi $u, v \in V', \{u, v\} \notin E$, tj. niti jedan par vrhova u V' nisu susjedni.

Definicija 1.2.5. 2-regularan povezan graf nazivamo **ciklički graf** ili **ciklus**. Ciklus koji ima n vrhova označavamo sa C_n . Kada je n paran broj kažemo da je to paran ciklus, a u suprotnom kažemo da je ciklus neparan.

Definicija 1.2.6. Graf koji dobijemo iz ciklusa C_{n-1} tako da svaki njegov vrh spojimo s jednim novim vrhom nazivamo **kotač** s n vrhova i označavamo ga s W_n .



Slika 1.2: Ciklus C_5 i kotač W_7 .

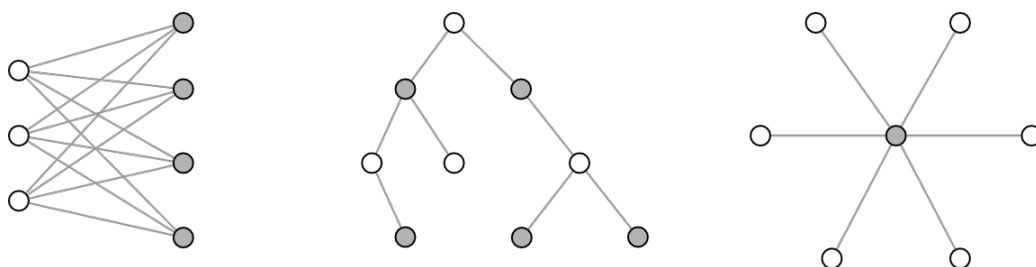
Definicija 1.2.7. Povezan graf koji ne sadrži cikluse zovemo **stablo**.

Definicija 1.2.8. Povezani graf sa n vrhova takav da jedan vrh ima stupanj $n - 1$, a ostali vrhovi su svi stupnja 1 nazivamo **zvijezda**. Zvijezdu s n vrhova označavamo S_n .

Definicija 1.2.9. Graf $G = (V, E)$ je **bipartitan** ako se skup vrhova V može particionirati u dva nezavisna skupa X i Y , tj. za svaki brid $e = \{u, v\} \in E$ vrijedi $u \in X$ i $v \in Y$. Particija (X, Y) se tada zove **biparticija** grafa. Ako je $|X| = x$ i $|Y| = y$, takav graf označavamo sa $K_{x,y}$.

Graf s biparticijom (X, Y) u kojoj je svaki vrh iz X spojen sa svakim vrhom iz Y zovemo **potpuni bipartitni graf**.

Na slici 1.3 prikazana su tri bipartitna grafa s bojom istaknutim particijama. Prvi graf $K_{3,4}$ je potpuni bipartitni graf i lagano je uočiti dva nezavisna skupa vrhova. Drugi graf $K_{4,5}$ je stablo i bipartitan je iako to nije očito na prvi pogled. Graf $K_{1,6}$ je zvijezda. Sva stabla su bipartitni grafovi, a sve zvijezde su potpuni bipartitni grafovi oblika $K_{1,k}$.



Slika 1.3: Potpuni bipartitni graf $K_{3,4}$, stablo $K_{4,5}$ i zvijezda $K_{1,6}$.

Definicija 1.2.10. Za graf kažemo da je **planaran** ako se može nacrtati (smjestiti) u ravnini \mathbb{R}^2 tako da mu se bridovi sijeku samo u vrhovima.

Može se pokazati da su grafovi koji pripadaju klasama: ciklusa, kotača, zvijezdi i stabla uvijek planarni. Potpuni bipartitni graf $K_{3,3}$ nije planaran [30], pa isto možemo zaključiti i za graf $K_{3,4}$ sa slike 1.3 kojem je $K_{3,3}$ podgraf.

Poglavlje 2

Najkraći putevi u grafu

Razmotrimo sljedeći problem: nalazimo se u Dubrovniku pored Velike Onofrijeve fontane i želimo doći do Porporele u najkraćem vremenu. Najdirektnija ruta bila bi preko Straduna pa ispod zvonika i na kraju kroz Porat do Porporele. Trajanje putovanja obično je proporcionalno njegovoj duljini, ali neki faktori kao gužve ili reljef mogu utjecati na to. Ljeti je Grad najčešće prepun turista, pa se šetnja Stradunom može pretvoriti u probijanje kroz rijeke ljudi. Ako nam cilj nije prijeći najkraću udaljenost, nego doći do odredišta u što kraćem vremenu, moramo razmotriti alternativne puteve. Okolne uličice, koje zaobilaze Stradun i popularne turističke atrakcije, dovest će nas do Porporele u kraćem vremenu iako ćemo prijeći veću udaljenost. Pronalaženje najkraćeg, najbržeg ili najjeftinijeg puta od točke A do točke B problem je s kojim se susrećemo gotovo svakodnevno, a u nastavku ćemo ga promotriti iz perspektive teorije grafova.

U teoriji grafova, *problem najkraćeg puta* opisujemo na sljedeći način: za težinski graf $G = (V, E)$ i vrhove $u, v \in V$ potrebno je pronaći put uWv najmanje duljine (definicija 1.1.8) [25]. Ako za naš primjer križanja ulica reprezentiramo vrhovima, a same ulice bridovima, neki putevi koje vode od Velike Onofrijeve fontane do Porporele prikazani su grafom na slici 2.1.¹ Bridovima kao težinu možemo pridružiti vrijeme potrebno za prolazak kroz neku ulicu u ljetnoj gužvi ili njenu duljinu. Put označen crvenom bojom na slici 2.1 predstavlja najkraću udaljenost dok je zeleni najbrži u slučaju gužve.

Postoje brojni algoritmi za pronalaženje najkraćih puteva u grafu, ali najpoznatiji je Dijkstrin algoritam koji ćemo opisati u sljedećem odjeljku. U zadnjem odjeljku ovog poglavlja opisati ćemo dvije primjene algoritma.

¹Kod kreiranja slike korištena je karta preuzeta s ©OpenStreetMap. Više informacija o ©OpenStreetMap može se pronaći na <https://www.openstreetmap.org/>.



Slika 2.1: Prikaz grafom nekih puteva koje vode od Velike Onofrijeve fontane do Porporele.

2.1 Dijkstrin algoritam

Nizozemski znanstvenik Edsger Wybe Dijkstra svojim pionirskim radom postavio je temelje suvremenoj računarskoj znanosti i tehnologiji, a mnogi od njegovih članaka su pokrenuli cijela nova područja istraživanja. Među njegovim brojnim doprinosima je i algoritam za pronalaženje najkraćeg puta u grafu, opisan članku [15] objavljenom 1959. godine [6]. U nastavku ćemo opisati kako funkcioniše **Dijkstrin algoritam**.

Dijkstrin algoritam koristi se za težinske grafove koji imaju nenegativne težine bridova. Za odabrani vrh u grafu, algoritam pronalazi najkraće puteve između toga vrha i svih ostalih vrhova [26]. Vrh koji se obrađuje u nekom koraku nazivamo *aktivni vrh*. U svakom koraku se računa nova udaljenost za sve susjede aktivnog vrha i pridružuje se vrhu ako je manja od trenutne. Među preostalim vrhovima izabere se onaj s najmanjom vrijednosti trenutne udaljenost, dodaje se u skup N obrađenih vrhova i postavlja se za novi aktivni vrh. Postupak se ponavlja dok svi vrhovi ne budu obrađeni. Ako je potrebno pronaći najkraći put samo za određeni par vrhova onda se jedan od njih odabere kao početni i algoritam se

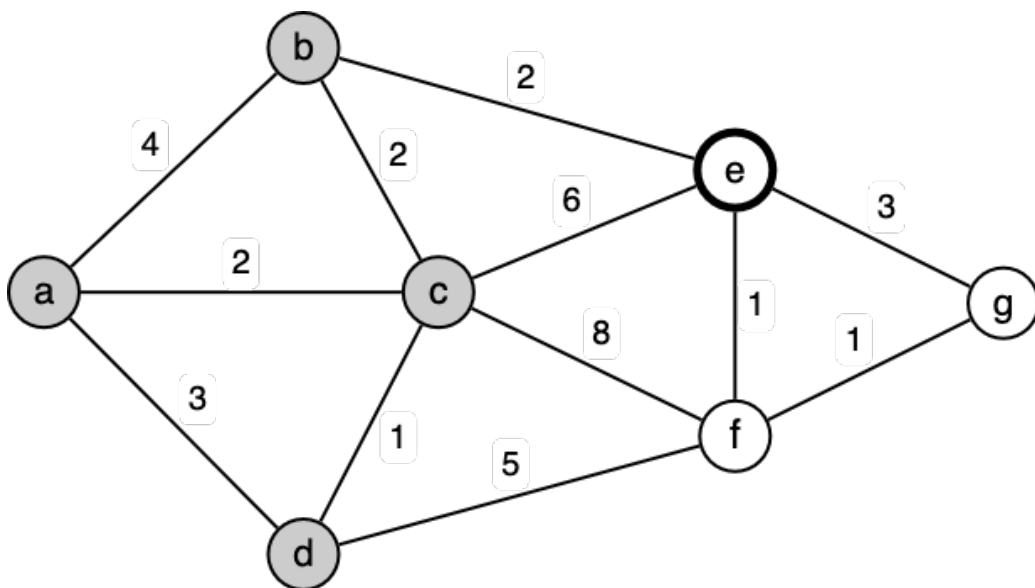
izvodi dok drugi vrh ne bude dodan u skup obrađenih vrhova.

Za neki graf $G = (V, E)$ Dijkstrin algoritam funkcionira na sljedeći način.

Odaberemo **početni vrh** v , a svim ostalim vrhovima u grafu udaljenost označimo s ∞ . Skup vrhova za koje je određen najkraći put od početnog vrha nazvat ćemo N . Početni vrh ima udaljenost 0 u odnosu na samog sebe. Dodamo ga u N i označimo kao **aktivni vrh**.

- Za sve susjede aktivnog vrha izračunamo novu udaljenost na način da udaljenosti aktivnog vrha dodamo težinu brida koji ih spaja (ili 1 ako graf nije težinski);
- ako je trenutna udaljenost susjednog vrha veća od nove, vrhu pridružimo novu udaljenost;
- od preostalih vrhova izaberemo onaj koji ima pridruženu najmanju udaljenost, dodamo ga u N i označimo kao aktivnog;

koraci se ponavljaju dok svi vrhovi ne budu dodani u N .



Slika 2.2: Vrhovi grafa G označeni sivom bojom imaju određen najkraći put od početnog vrha a .

Konkretan primjer izvođenja algoritma promotrit ćemo za graf G prikazan na slici 2.2.

Vrh a je izabran kao početni, a vrhovi označeni sivom bojom imaju već određen najkraći put. Slika prikazuje korak algoritma u kojem je skup $N = \{a, c, d, b, e\}$, a aktivni vrh je e . U ovom koraku računamo nove udaljenosti za sve susjede od e . Nova udaljenost za f je 7 što je manje od trenutne vrijednosti pa vrhu f pridružujemo novu vrijednost. Za g je nova udaljenost jednaka trenutnoj pa se ništa ne mijenja. U sljedećem koraku u skup N dodajemo vrh f zato što ima najmanju trenutnu vrijednost udaljenosti od svih vrhova preostalih u $V \setminus N$. Vrh f postaje aktivni vrh, a nova udaljenost do jedinog preostalog susjeda g iznosi 8, što je manje od trenutne vrijednosti. Vrh g pridružujemo novu udaljenost i dodajemo ga u N , a kako nema više vrhova za obraditi, algoritam staje. Svi koraci izvođenja Dijkstrinog algoritma za graf G su prikazani u tablici 2.3.

korak	a	b	c	d	e	f	g
1.	0	∞	∞	∞	∞	∞	∞
2.	0	4	2	3	∞	∞	∞
3.	0	4	2	3	8	10	∞
4.	0	4	2	3	8	8	∞
5.	0	4	2	3	6	8	∞
6.	0	4	2	3	6	7	9
7.	0	4	2	3	6	7	8

Tablica 2.3: Koraci izvođenja Dijkstrinog algoritma za graf G .

2.2 Neke primjene algoritama za pronalazak najkraćeg puta

Procjena očekivanog vremena dolaska na odredište

Google Maps je digitalna platforma koja omogućuje pregled geografskih karata i satelitskih snimki. Od nastanka 2005. godine do danas, Google Maps je dobio brojne nove funkcionalnosti i postao alat koji svakodnevno koriste milijuni ljudi. Google Maps korisnicima nudi informacije o sadržajima dostupnim u blizini lokacije na kojoj se nalaze, informacije o dostupnosti javnog prijevoza i trenutnom stanju u prometu, kao i mogućnost pronalaska najbrže rute do odabranog odredišta. Procjena očekivanog vremena dolaska na odredište do 2007. godine računala se samo na osnovu udaljenosti između dvije točke i prosječne brzine kretanja korisnika. Za računanje najkraće udaljenosti koristi se graf i algoritam za pronalaženje najkraćeg puta. Algoritmi kao Dijkstrin se ne uspijevaju efikasno nositi s količinom podataka koju Google Maps analizira pa se koristi heuristički algoritam A^* koji je sličan Dijkstrinom, ali provjerava udaljenosti za manji broj vrhova pa se u praksi

pokazao bržim iako zahtjeva više memorije i operacija po obrađenom vrhu [26].² Danas Google Maps uzima u obzir cijeli niz dodatnih faktora koji utječu na trajanje putovanja kao što su ograničenja brzine, prometne gužve, prosječnu brzinu korisnika u nekom vremenskom periodu i mnoge druge. Za više informacija zainteresiranog čitatelja se upućuje na [26].

Tablice usmjeravanja

Mrežni komunikacijski protokol predstavlja skup pravila koja omogućavaju prijenos podataka kroz komunikacijske kanale. Usmjernik (engl. router) je uređaj koji usmjerava podatkovne pakete koji putuju mrežom. Za provjeru odredišne adrese paketa i preusmjeravanje na odgovarajuće sučelje, usmjernici koriste *tablice usmjeravanja*. U tablicama usmjeravanja pohranjena je topologija mreže, tj. za svaku adresu odredišta, spremljena je mrežna ruta koja do njega vodi. Na efikasnost usmjeravanja podataka kroz mrežu mogu utjecati postupci kreiranja i održavanja ovih tablica. Za određivanje najbolje rute prijenosa paketa algoritmi usmjeravanja koriste različite metrike, kao što su duljina, pouzdanost, opterećenje mreže i druge. Najčešće korištena metrika je duljina rute koja predstavlja broj usmjerivača kroz koje paket mora proći ili sumu troškova mrežnih veza na određenoj ruti. Kako bi mreža funkcionirala optimalno, poželjno je u tablicama usmjeravanja imati najkraće puteve između parova usmjernika. Mrežni protokoli IS-IS (engl. Intermediate System to Intermediate System) i OSPF (engl. Open Shortest Path First) za izračunavanje ruta koriste Dijkstrin algoritam [19].

²Više o algoritmu A^* može se pronaći u [26].

Poglavlje 3

Bojanje grafova

Problem bojanja grafova pojavljuje se u 19. stoljeću kada je Francis Guthrie formulirao *problem četiri boje*. Bojajući kartu Engleskih pokrajina, tako da susjedne pokrajine budu obojane različitom bojom, Guthrie je primijetio da su dovoljne samo četiri boje. Pokušao je pronaći primjer karte koja se ne bi mogla obojati s najviše četiri boje i to mu nije pošlo za rukom. Od 1852. godine, kada je prvi put formuliran, do 1976. kada je uz pomoć računala dokazan *teorem četiri boje*, ovim problemom su se bavili neki od najznačajnijih matematičara devetnaestog i dvadesetog stoljeća. Problem četiri boje, tj. problem bojanja karte možemo promotriti kao problem bojanja grafa čiji vrhovi predstavljaju područja na karti (na primjer pokrajine ili države), a bridovi povezuju dva vrha ako odgovarajuća područja dijele granicu. Da bi graf bio ispravno obojan susjedni vrhovi moraju biti različite boje i ukupan broj iskorištenih boja treba biti minimalan.

Proučavanje bojanja grafova je krenulo s bojanjem geografskih karata ali primjene ovog polja teorije grafova su moguće u izrazito puno, ponekad ne odmah očitih, područja. U ovom poglavlju najprije ćemo uvesti osnovne pojmove i navesti primjere. Kratko ćemo opisati pohlepni algoritam za bojanje grafova ali nećemo razmatrati detalje implementacije, složenosti i ograničenja algoritama. Navest ćemo značajne primjere i posebne slučajeve te opisati neke zanimljive primjene. Na kraju ćemo promotriti problem dodjele identifikacijskih kodova baznim stanicama na primjeru telekomunikacijske mreže baznih stanica grada Zagreba. U ovom poglavlju uglavnom pratimo [24].

3.1 Osnovni pojmovi

Najprije ćemo formalno definirati bojanje grafova i pojmove koji su potrebni da bismo opisali određena svojstva koja želimo da bojanje ima.

Definicija 3.1.1. *Neka je $G = (V, E)$ neusmjereni graf. Funkciju $\phi : V \rightarrow \mathbb{N}$ koja svakom*

vrhu $v \in V$ pridružuje prirodan broj nazivamo **bojanje grafa**, a broj $\phi(v)$ nazivamo **boja vrha**.

Za neki $k \in \mathbb{N}$, funkciju $\phi : V \rightarrow \{1, \dots, k\}$ nazivamo **k -bojanje grafa G** .

Primijetimo da *bojanje* zapravo predstavlja označavanje vrhova, tj. pridruživanje prirodnih brojeva koji mogu predstavljati odgovarajuću boju iz određenog skupa ili neku drugu vrstu oznake. Bojanje grafova postaje zanimljiv i kompleksan problem tek kada uvedemo određena ograničenja. U praktičnim primjenama tražimo *dopustivo* bojanje. Također, najčešće je broj boja koje su na raspolaganju ograničen ili ga želimo minimizirati.

Definicija 3.1.2. *k -bojanje grafa $G = (V, E)$ je **dopustivo** ako svaka dva susjedna vrha imaju pridružene različite boje. Preciznije, za sve vrhove $v, u \in V$ vrijedi: ako postoji brid $\{v, u\} \in E$ tada mora vrijediti $\phi(v) \neq \phi(u)$.*

*Kažemo da je graf **k -obojev** ako postoji dopustivo k -bojanje grafa.*

Bitno je napomenuti da dopustivo bojanje grafa nije moguće ako graf sadrži petlje. Ako u vrhu $v \in V$ postoji petlja, tj. brid $\{v, v\} \in E$ tada je nemoguće obojati taj vrh i zadovoljiti uvjete za dopustivo bojanje pa ćemo u nastavku razmatrati samo grafove bez petlji.

Definicija 3.1.3. *Brid koji spaja dva vrha iste boje nazivamo **konfliktni brid**.*

*Dva vrha spojena konfliktnim bridom nazivamo **konfliktni vrhovi**.*

Svaki graf s n vrhova je n -obojev zato što svaki vrh možemo obojati jednom od n boja i biti sigurni da ne postoje konfliktni vrhovi bez obzira na to koji su vrhovi susjedni. Ključni problem bojanja grafova je pitanje – Koji je najmanji broj boja potrebnih za dopustivo bojanje zadanog grafa?

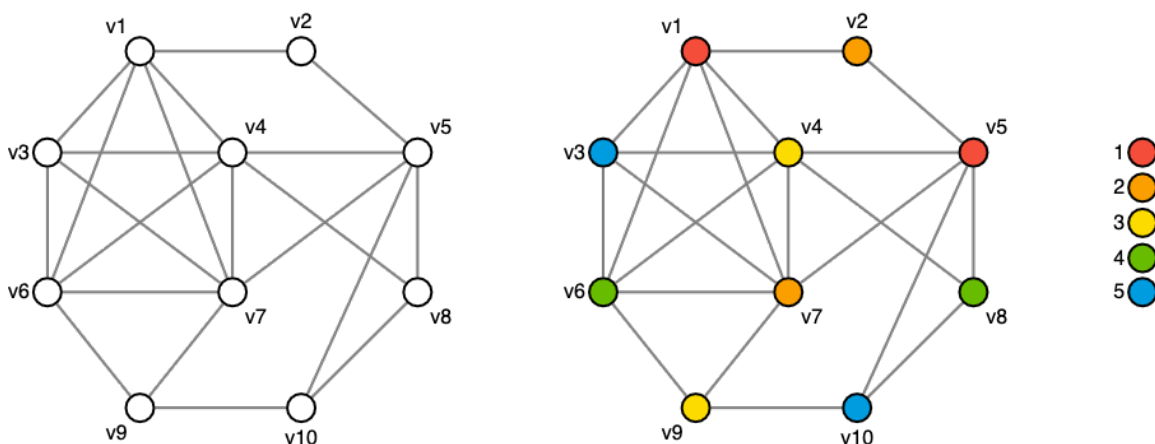
Definicija 3.1.4. *Najmanji broj boja potrebnih za dopustivo bojanje grafa G nazivamo **kromatski broj** i označavamo ga sa $\chi(G)$. Ako za graf G vrijedi $\chi(G) = k$, kažemo da je G **k -kromatski**. Bojanje grafa G koje koristi točno $\chi(G)$ boja naziva se **optimalnim bojanjem**.*

Određivanje kromatskog broja zadanog grafa je općenito netrivialan problem,¹ ali u ovom radu nećemo razmatrati njegovu složenost. Za neke topologije grafova lagano je odrediti kromatski broj i to ćemo pokazati u sljedećem dijelu.

Definicija 3.1.5. *Za neko bojanje grafa $G = (V, E)$ definiramo **klasu boja** kao skup koji sadrži sve vrhove kojima je dodijeljena određena boja u tom bojanju. Preciznije, za konkretnu boju $i \in \{1, \dots, k\}$ u nekom k -bojanju grafa G , klasa boje i je skup $\{v \in V : \phi(v) = i\}$.*

Ako promatramo dopustivo k -bojanje za neki graf, klase boja će u tom slučaju nužno biti nezavisni skupovi vrhova.

¹Problem k -obojevnosti grafa spada u klasu NP-potpunih problema. Više o klasama složenosti i NP-potpunim problemima može se pronaći na primjer u [35].

Slika 3.1: Primjer jednog dopustivog 5-bojanja grafa G .

Primjer 3.1.6. Promotrimo graf G na slici 3.1. G se sastoji od 10 vrhova i 21 brida. Na slici je prikazano jedno moguće 5-bojanje grafa G . Ako konkretnim bojama pridružimo oznake od 1 do 5 kao na slici, onda bojanje možemo zapisati na sljedeći način:

$$\begin{aligned} \phi(v_1) = 1, & \quad \phi(v_2) = 2, & \quad \phi(v_3) = 5, & \quad \phi(v_4) = 3, & \quad \phi(v_5) = 1, \\ \phi(v_6) = 4, & \quad \phi(v_7) = 2, & \quad \phi(v_8) = 4, & \quad \phi(v_9) = 3, & \quad \phi(v_{10}) = 5. \end{aligned} \quad (3.1)$$

Možemo vidjeti da ne postoje konfliktne vrhove pa je ovo bojanje dopustivo. Očito je da je kromatski broj ovog grafa $\chi(G) \leq 5$. Ako pozornije promotrimo graf G , možemo uočiti da su vrhovi v_1, v_3, v_4, v_6 i v_7 svi međusobno susjedni, tj. čine kliku. Za dopustivo bojanje klike od n vrhova potrebno je točno n boja. Kako ovaj graf sadrži kliku od 5 vrhova, dopustivo bojanje će uvijek imati barem 5 boja. Sada možemo zaključiti da je $\chi(G) = 5$ pa je bojanje na slici 3.1 optimalno.

Važno je napomenuti da su oznake, tj. brojevi koje pridružujemo vrhovima proizvoljni i više nas zanima ukupan broj boja potrebnih za dopustivo bojanje grafa, a ne nužno koje su to boje. Tako je na primjer bojanje:

$$\begin{aligned} \phi(v_1) = 1, & \quad \phi(v_2) = 5, & \quad \phi(v_3) = 2, & \quad \phi(v_4) = 3, & \quad \phi(v_5) = 1, \\ \phi(v_6) = 4, & \quad \phi(v_7) = 5, & \quad \phi(v_8) = 4, & \quad \phi(v_9) = 3, & \quad \phi(v_{10}) = 2; \end{aligned} \quad (3.2)$$

identično bojanju 3.1 jer su vrhovi grupirani na jednak način samo su oznake (boje) drugačije. Klase boja u oba rješenja su isti skupovi samo pridruženi različitim bojama (u rješenju (3.2) boje 2 i 5 su zamijenjene u odnosu na (3.1)).

3.2 Kromatski brojevi za neke topologije grafova

Određivanje kromatskog broja proizvoljnog grafa nije jednostavno ako je graf jako velik. Međutim, za određene topologije grafova kromatski su brojevi unaprijed poznati.

Promotrimo najjednostavniji slučaj, prazan graf. Prazan graf $G = (V, E)$ ima $|V| = n$ vrhova, a $E = \emptyset$. Očito niti jedan par vrhova nije susjedan pa ako sve vrhove obojamo istom bojom dobijemo dopustivo bojanje. Dakle, kromatski broj praznog grafa je $\chi(G) = 1$.

U nastavku navodimo neke manje trivijalne topologije grafova i njihove kromatske brojeve.

Potpuni graf

Potpuni graf s n vrhova, K_n ima brid između svakog para vrhova, tj. svaki vrh je susjedan svim ostalim vrhovima. Očito je da, kako bi dobili dopustivo bojanje, svakom vrhu moramo dodijeliti vlastitu boju pa kromatski broj potpunog grafa K_n mora biti $\chi(K_n) = n$.

Bipartitni graf

Vrhove bipartitnog grafa uvijek možemo podijeliti u dva nezavisna skupa. Nezavisni skupovi vrhova ne sadrže niti jedan par vrhova koji su susjedni pa svi vrhovi u nekom nezavisnom skupu mogu biti obojani istom bojom. Dakle, ako svaku particiju bipartitnog grafa obojamo vlastitom bojom dobit ćemo dopustivo bojanje. Kromatski broj svakog bipartitnog grafa $K_{x,y}$ je $\chi(K_{x,y}) = 2$. Na slici 1.3 su prikazana 3 bipartitna grafa s optimalnim bojanjima.

Ciklus

Kromatski broj cikličkog grafa C_n ovisi o broju njegovih vrhova stoga ćemo odvojeno razmatrati parne i neparne cikluse.

Prvi slučaj je kada je n paran broj, tj. kada je C_n paran ciklus. Vrhove možemo obojati tako da izaberemo nasumični vrh od kojeg ćemo krenuti (neka to bude v_1) i obojamo ga prvom bojom, neka to bude bijela. Graf obilazimo u smjeru kazaljke na satu i prvi sljedeći vrh moramo obojati drugom bojom, neka to bude crna. Ako nastavimo obilaziti graf, sve vrhove možemo naizmjenično bojati crnom i bijelom bojom. Kada dođemo do posljednjeg vrha v_n njegovi susjedni vrhovi v_1 i v_{n-1} su već obojani bijelom bojom pa v_n možemo obojati crnom. Dakle, za svaki paran ciklus C_n postoji dopustivo 2-bojanje pa je kromatski broj parnog ciklusa $\chi(C_n) = 2$. Iz ovoga možemo zaključiti da je vrhove parnog ciklusa moguće podijeliti u dva nezavisna skupa stoga su parni ciklusi bipartitni grafovi.

Za neparni ciklus C_{n+1} slijedimo isti postupak kao za parni ciklus, krenemo u vrhu v_1 i naizmjenično bojamo vrhove bijelom i crnom bojom. Kada dođemo do posljednjeg vrha v_{n+1} , njegovi susjedni vrhovi, v_1 i v_n će biti obojani različitim bojama, v_1 u bijelo, a v_n u

crno. Da bi dobili dopustivo bojanje vrh v_{n+1} moramo obojati trećom bojom. Možemo zaključiti da je za sve neparne cikle C_{n+1} kromatski broj jednak $\chi(C_{n+1}) = 3$.

Kotač

Kotač W_n konstruiramo tako da nekom ciklusu C_{n-1} dodamo jedan vrh i spojimo ga sa svim ostalim vrhovima. Jasno je da ćemo za kotač, kao i za ciklus, imati dva slučaja ovisno o tome je li broj vrhova n paran ili neparan.

Ako je n paran broj, kotač W_n je nastao od neparnog ciklusa za čije bojanje su potrebne 3 boje. Dodavanjem jednog vrha koji je susjedan svim ostalim vrhovima broj boja potrebnih za dopustivo bojanje povećava se za 1 pa je kromatski broj kotača W_n s parnim brojem vrhova jednak $\chi(W_n) = 4$.

Ako imamo kotač W_{n+1} sa neparnim brojem vrhova, on je konstruiran iz parnog ciklusa i za dopustivo bojanje trebamo jednu boju više nego za parni ciklus. Dakle, kromatski broj kotača W_{n+1} s neparnim brojem vrhova je $\chi(W_{n+1}) = 3$.

Planarni graf

Problem četiri boje, spomenut na samom početku ovog poglavlja, zapravo je problem bojanja *planarnih* grafova. Naime, svakoj geografskoj karti možemo pridružiti odgovarajući planarni graf pa problem bojanja karte postaje problem bojanja grafova. Tvrdnja da svaki planarni graf G ima kromatski broj $\chi(G) \leq 4$ jednostavno je iskazana slavnim teoremom četiri boje. [24]

Teorem 3.2.1. (Teorem četiri boje) *Za svaki planarni graf bez petlji postoji dopustivo 4-bojanje. Ekvivalentno, svaka karta se može obojati sa 4 boje.*

Iako je naizgled jednostavan, od formulacije slutnje do dokaza ovog teorema proteklo je više od 120 godina. Njime su se bavili brojni ugledni matematičari devetnaestog i dvadesetog stoljeća, a mnogi ponuđeni dokazi pokazali su se neispravnima. Teorem su 1976. godine konačno dokazali Kenneth Appel i Wolfgang Haken [5] uz pomoć računala. Više o teoremu četiri boje može se pronaći u [24]. Iako se dokaz pokazao iznimno kompleksnim, mnoge ideje i rezultati koji su proizašli iz neuspješnih pokušaja doprinijeli su konačnom rješenju problema. Matematičar Alfred Kempe objavio je 1879. godine dokaz teorema četiri boje, ali nakon 11 godina Percy Heawood je u njemu pronašao pogrešku. Heawood nije ponudio vlastiti dokaz, ali je uz pomoć Kempeovih rezultata dokazao *teorema pet boja*. [24]

Teorem 3.2.2. (Teorem pet boja) *Za svaki planarni graf bez petlji postoji dopustivo 5-bojanje.*

Dokaz ovog teorema nije pretjerano izazovan i može se pronaći na primjer u [3]. Pregled kromatskih brojeva za klase grafova navedene u ovom odjeljku nalazi se u tablici 3.2.

Klasa grafa	Kromatski broj χ
Potpuni graf (K_n)	n
Planarni graf (G_n), $n \geq 1$	≤ 4
Paran ciklus (C_n), $n \geq 1$	2
Neparan ciklus (C_n), $n \geq 1$	3
Bipartitni graf ($K_{x,y}$), $n \geq 1$	2
Kotač (W_n), $n \geq 2$ je paran	4
Kotač (W_n), $n \geq 2$ je neparan	3

Tablica 3.2: Pregled kromatskih brojeva za različite klase grafova.

3.3 Pohlepni algoritam za bojanje grafova

U prethodnom odjeljku smo promotrili različite topologije grafova za koje je jednostavno pronaći optimalno bojanje i odrediti kromatski broj. Za grafove koji imaju velik broj vrhova i bridova, a nemaju neku od pravilnih struktura kao ranije promatrani grafovi, određivanje kromatskog broja i pronalazak optimalnog bojanja je izrazito kompleksan zadatak. Naime, bojanje grafova je *NP-potpun* problem,² pa ne postoji algoritam koji, za svaki graf, pronalazi optimalno bojanje u polinomnom vremenu. U praksi nije uvijek nužno pronaći optimalno bojanje, obično je prihvatljivo neko dopustivo bojanje pa je dosta napora uloženo u razvoj heurističkih algoritama koji u razumnom vremenu najčešće generiraju dovoljno dobro, ali ne nužno optimalno rješenje [23]. Jedan od takvih algoritama koji se najčešće koristi je pohlepni algoritam i njega ćemo opisati u nastavku. Navest ćemo neke specifičnosti algoritma i promotriti nekoliko primjera.

Opis algoritma

Pohlepni algoritam (engl. greedy algorithm) je jedan od konceptualno najjednostavnijih algoritama za bojanje grafova. Nedostatak mu je da ne generira uvijek optimalno bojanje. Algoritmu je na početku potrebno predati *redoslijed kojim će obilaziti vrhove*. Algoritam će uvijek pronaći dopustivo bojanje ali broj boja koje se koriste u rješenju može značajno varirati ovisno o redoslijedu kojim se obilaze vrhovi.

²Problem k -obojivosti grafa spada u klasu NP-potpunih problema. Više o klasama složenosti i NP-potpunim problemima može se pronaći na primjer u [35].

Za neki graf G koji se sastoji od n vrhova pohlepni algoritam funkcionira na sljedeći način.

Ako vrhove grafa G poredamo nekim redoslijedom, na primjer (v_1, v_2, \dots, v_n) , pohlepni algoritam se izvršava slijedom:

- u prvom koraku vrhu v_1 pridružuje boju 1;
- u i -tom koraku za vrh v_i provjerava redom može li biti obojan nekom od j do sada korištenih boja bez konflikta te:
 - ako za v_i na popisu već korištenih boja postoji boja k koja neće narušiti dopustivost bojanja, prva (najmanja) takva se pridružuje vrhu v_i ;
 - ako svaka od j do sad korištenih boja stavlja vrh v_i u konflikt s nekim od već obojanih vrhova, dodaje se nova boja $j+1$ i pridružuje se vrhu v_i ;

koraci se ponavljaju dok svi vrhovi ne budu obojani.

Najzornije možemo razumjeti funkcioniranje algoritma promotrimo li konkretan primjer. Za graf G sa slike 3.1 i redoslijed vrhova $(v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10})$, koraci izvođenja pohlepnog algoritma dani su na slici 3.3. Pohlepni algoritam generira optimalno 5-bojanje za graf G , ali primijetimo da su vrhovi ovaj put obojani drugačije nego u (3.1) i (3.2).

Novo 5-bojanje je:

$$\begin{aligned} \phi(v_1) = 1, & \quad \phi(v_2) = 2, & \quad \phi(v_3) = 2, & \quad \phi(v_4) = 3, & \quad \phi(v_5) = 1, \\ \phi(v_6) = 4, & \quad \phi(v_7) = 5, & \quad \phi(v_8) = 2, & \quad \phi(v_9) = 1, & \quad \phi(v_{10}) = 3. \end{aligned} \quad (3.3)$$

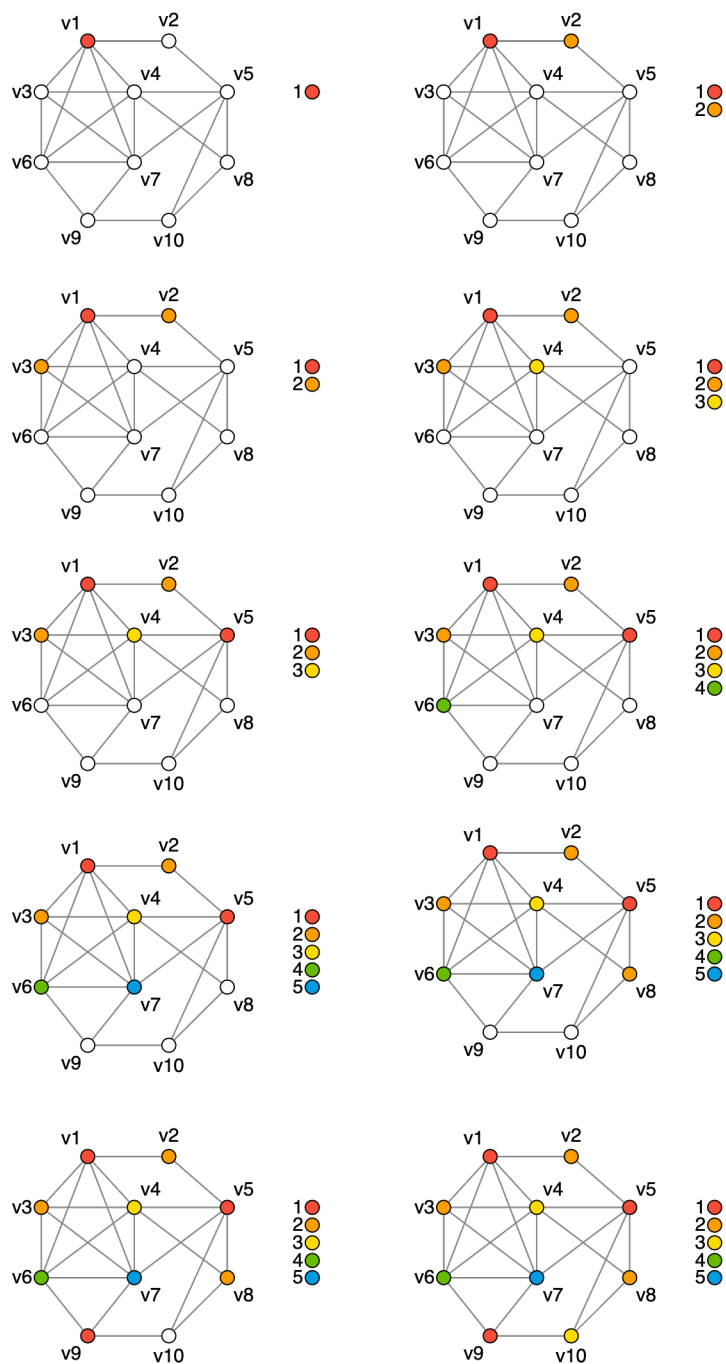
Dok smo u (3.1) i (3.2) imali vrhove grupirane u klase boja:

$$\{v_1, v_5\}, \{v_2, v_7\}, \{v_3, v_{10}\}, \{v_4, v_9\}, \{v_6, v_8\},$$

u (3.3) grupirani su na sljedeći način:

$$\{v_1, v_5, v_9\}, \{v_2, v_3, v_8\}, \{v_4, v_{10}\}, \{v_6\}, \{v_7\}.$$

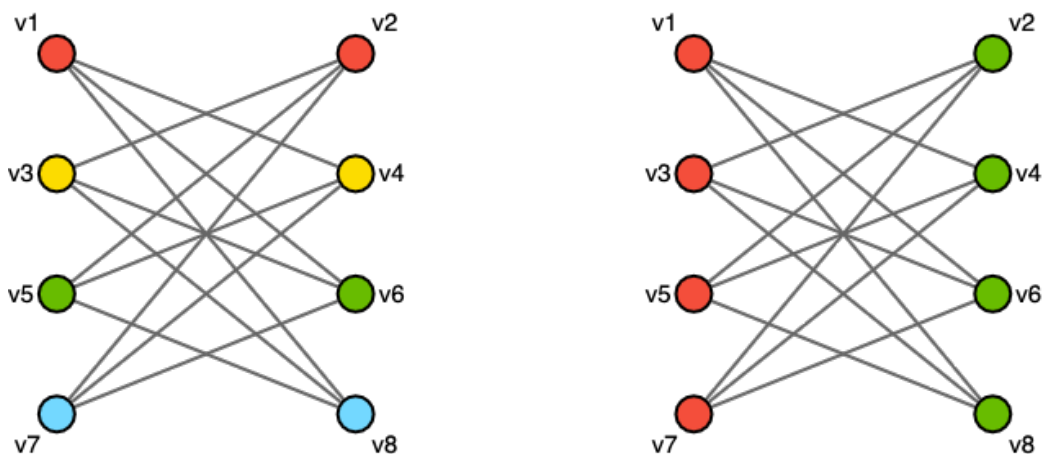
Uzrok ovih značajnih razlika u rezultatu je upravo redoslijed kojim pohlepni algoritam obilazi vrhove grafa.



Slika 3.3: Koraci izvođenja pohlepnog algoritma.

Redosljed obilaska vrhova

Redosljed obilaska vrhova značajno utječe na strukturu klasa boja i na ukupan broj boja koje se koriste. Na primjer, za 3-regularni bipartitni graf $K_{4,4}$ na slici 3.4, rezultati bojanja pohlepnim algoritmom sa dva različita redosljeda obilaska vrhova daju bitno različite rezultate.



Slika 3.4: Primjer bojanja bipartitnog grafa pohlepnim algoritmom za dva različita redosljeda obilaska vrhova.

Dok za lijevi graf redosljed obilaska vrhova $(v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8)$ rezultira 4-bojanjem, za desni graf redosljed $(v_1, v_3, v_5, v_7, v_2, v_4, v_6, v_8)$ daje optimalno 2-bojanje. Očito je da je redosljed kojim pohlepni algoritam obilazi vrhove grafa ključan za optimalnost rješenja.

Iako će pohlepni algoritam često generirati suboptimalno rješenje, moguće je pokazati da **za svaki graf uvijek postoji redosljed obilaska vrhova za koji će rješenje biti optimalno bojanje** [24]. Za graf s n vrhova postoji $n!$ mogućih redosljeda obilaska, stoga je pronalazak optimalnog obilaska jednako kompleksan zadatak kao i pronalaženje kromatskog broja ili optimalnog bojanja.

Jedna od metoda poboljšanja nekog poznatog dopustivog bojanja je poredati vrhove grafa tako da vrhovi koji pripadaju jednoj klasi boja u postojećem rješenju budu smješteni na susjedne pozicije u novom redosljedu obilaska. Ako je inicijalno rješenje koristilo k boja, onda je vrhovima koji su prije pripadali i -toj klasi moguće pridružiti i -tu ili bilo koju od već korištenih boja pa ako inicijalno rješenje nije bilo optimalno, postoji mogućnost da će svi vrhovi iz i -te klase biti smješteni u neke druge klase pa time dobivamo poboljšanje početnog rješenja koje koristi $l \leq k$ boja.

Na primjer, za graf G sa slike 3.1 možemo pokrenuti pohlepni algoritam sa redosljedom vrhova $(v_1, v_5, v_9, v_2, v_3, v_8, v_4, v_{10}, v_6, v_7)$ i novi rezultat će biti barem jednako dobar kao inicijalni, tj. dobiti ćemo 5-bojanje. Za G ne možemo dobiti bolje rješenje promjenom redosljeda vrhova zato što je kromatski broj $\chi(G) = 5$, pa je jedina moguća promjena eventualno različito grupiranje vrhova u klase boja. Ovo je upravo jedan mogući redosljed vrhova koji garantira optimalno bojanje za G . Može se dokazati da za bilo koji graf redosljed vrhova, za kojeg će pohlepni algoritam generirati optimalno bojanje, dobijemo tako da grupiramo vrhove po klasama boja nekog već poznatog optimalnog bojanja za taj graf. Formalni iskaz i dokaz iznesenih tvrdnji može se pronaći u [24].

3.4 Neke primjene bojanja grafova

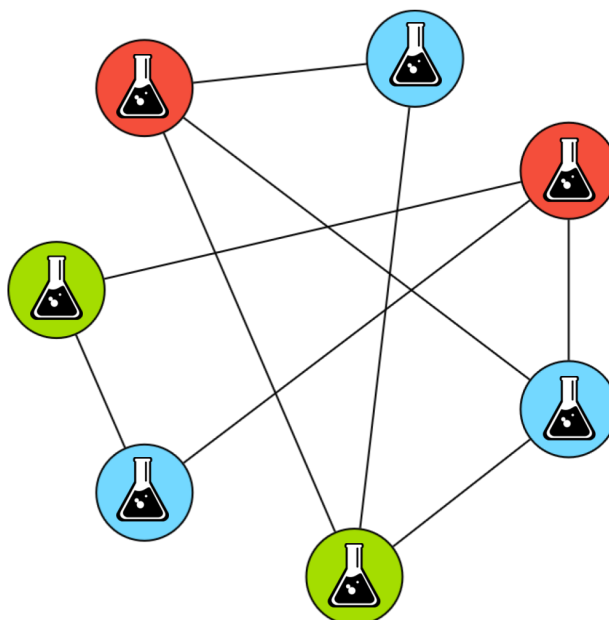
Kreiranje rasporeda

Kreiranje kvalitetnog sveučilišnog rasporeda nije jednostavan zadatak. Problem se obično dijeli u dvije kategorije: kreiranje rasporeda predavanja i kreiranje rasporeda održavanja ispita. Da bi se neki raspored održavanja ispita smatrao dovoljno dobrim, moraju biti zadovoljeni određeni uvjeti. Ključan uvjet je da svi ispiti koje polaže isti student moraju biti u različitim terminima. Cilj je kreirati zadovoljavajući raspored, ali minimizirati broj potrebnih termina. Najčešće postoje još brojni dodatni zahtjevi kao na primjer da se određeni ispit održi u specifično vrijeme ili da se određena skupina ispita održi najranije moguće u ispitnom razdoblju.

Problem kreiranja rasporeda ispita može se preslikati u problem bojanja grafova. Jednostavan pristup je da se ispitima i dostupnim terminima pridruže vrhovi grafa. Svi termini su međusobno spojeni bridovima da bi činili kliku u kojoj će svakom vrhu nužno biti pridružena različita boja. Ako za neki par ispita postoji student koji mora izaći na oba, ispiti se spajaju bridom kako bi se osiguralo da budu obojani različitom bojom. Za ispit koji se ne može održati u određenom terminu dodaje se brid koji ga spaja s tim terminom [23]. U ovako konstruiranom grafu, dopustivo bojanje će svim parovima ispita i termina koji imaju neku vrstu konflikta pridružiti različitu boju. Svaka klasa boja sadržavati će jedan termin i ispiti koji se u njemu održavaju. Ako broj dostupnih termina nije dovoljan da bi se zadovoljili svi postavljeni uvjeti, pojavit će klasa boja koja ne sadrži niti jedan termin. U tom slučaju, bez dodavanja novih termina, nemoguće je kreirati dobar raspored ispita. Mogu se postaviti još brojni zahtjevi koje bi bilo poželjno zadovoljiti, a time kompleksnost problema raste. Kroz godine su se pojavili različiti pristupi rješavanju problema te su čak održana natjecanja u kreiranju rasporeda. Temeljit pregled različitih pristupa i rješenja može se pronaći u [24].

Skladištenje kemikalija

Iako su mnoge kemikalije bezopasne i jednostavne za skladištenje i odlaganje, neke od njih, kada dođu u izravan kontakt, mogu postati iznimno opasne. Određene kombinacije kemikalija će rezultirati nastankom toksičnih spojeva dok neke reakcije mogu izazvati eksplozije i uzrokovati požare. Osim skladištenja kemikalija u istom prostoru, problem predstavlja i sigurno odlaganje otpadnih kemikalija. Kao kod skladištenja, prilikom odlaganja postoji mogućnost opasnih reakcija u slučaju miješanja nekih kemikalija, ali i potencijalno nastajanje spojeva štetnih za okoliš ili ljudsko zdravlje. Ponekada je situacija prilično jednostavna, na primjer treba osigurati da određene kemikalije ne dođu u dodir s vodom ili da se dvije nekompatibilne kemikalije ne skladište u istom prostoru, međutim postoji još mnoštvo kombinacija koje mogu dovesti do ozbiljnih posljedica.



Slika 3.5: Graf G interakcija za 7 različitih kemikalija.

Ovaj problem može se preslikati u problem bojanja grafova tako da se konstruira graf interakcija u kojem vrhovi predstavljaju pojedine kemikalije, a bridovi nesigurne kombinacije kemikalija koje mogu rezultirati nezgodama. Dopustivo bojanje ovakvog grafa predstavlja kombinacije kemikalija koje su sigurne za skladištenje u istom spremniku. Kromatski broj grafa interakcija predstavlja minimalan broj spremnika potrebnih za sigurno skladištenje ili odlaganje kemikalija. Ilustracija problema prikazana je na slici 3.5 gdje je graf interakcija G konstruiran za 7 različitih kemikalija. Graf G ima kromatski broj

$\chi(G) = 3$, što znači da su potrebna barem tri različita spremnika za skladištenje ovih kemikalija, a klase boja predstavljaju sigurne kombinacije kemikalija. Više detalja o primjeni bojanja grafova za rješavanje problema skladištenja i odlaganja kemikalija te zanimljive primjere i softversko rješenje može se pronaći u [31].

3.5 Dodjeljivanje kodova baznim stanicama telekomunikacijske mreže

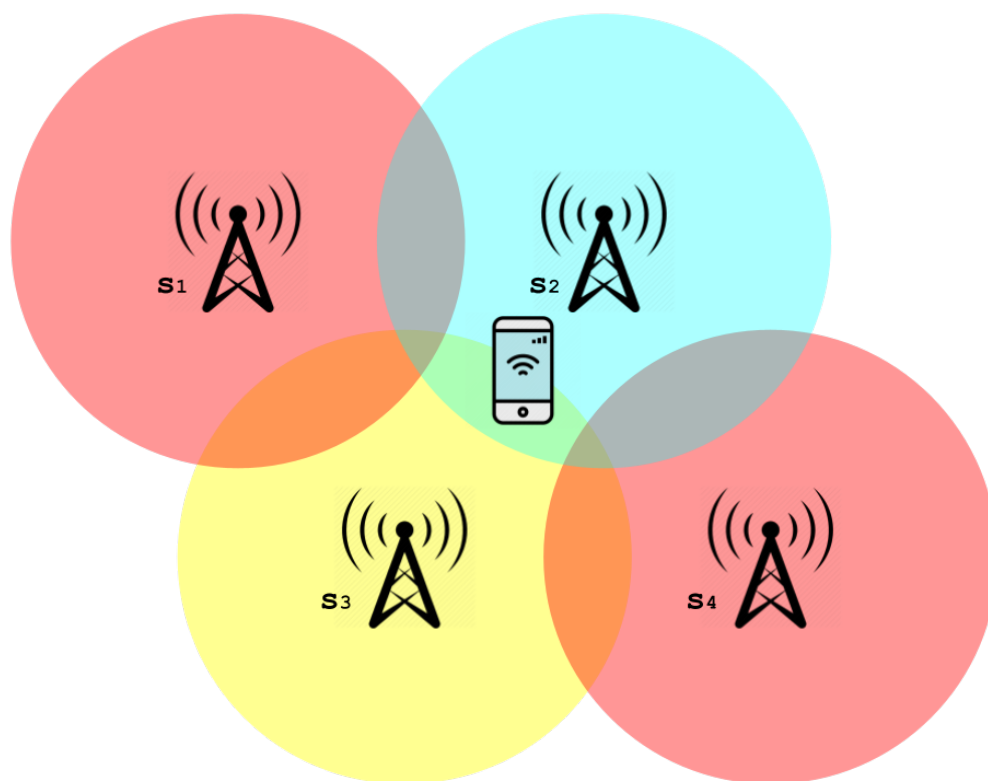
Najčešće primjene teorije grafova u telekomunikacijama obično su usmjerene na analizu i povećanje efikasnosti prijenosa podataka putem interneta ili analizu mreže u svrhu identificiranja potencijalnih ranjivosti. Jedan manje očit problem iz domene telekomunikacija, koji se može modelirati grafom, je dodjeljivanje identifikacijskih kodova baznim stanicama.

U ovom odjeljku najprije ćemo opisati ovaj problem i objasniti zašto se on pojavljuje. Zatim ćemo pokazati da, kad se modelira grafom, dodjeljivanje kodova baznim stanicama postaje problem bojanja grafa. Na kraju ćemo na primjeru konkretne telekomunikacijske mreže pokazati rezultat.

Opis problema

Da bi se za neko područje osigurala dobra pokrivenost signalom, bazne stanice, tj. antene moraju biti postavljene dovoljno gusto. Svaka bazna stanica ima svoj identifikator, odnosno kod (engl. scrambling code) i određenu jačinu signala koja opada s udaljenosti. Broj kodova je ograničen, preciznije, na raspolaganju je samo 512 kodova [37], pa je nemoguće svakoj baznoj stanici dodijeliti jedinstveni. Kako bi pristupili mreži, mobilni uređaji moraju biti spojeni na neku baznu stanicu. Dok bazne stanice imaju fiksnu lokaciju, mobilni su uređaji najčešće u pokretu pa povremeno izlaze iz dometa jedne bazne stanice i ulaze u domet neke druge. Kada uređaj izađe iz dometa signala bazne stanice na koju je spojen, veza se prekida i uređaj se spaja na novu stanicu u čijem dometu se nalazi. Da bi ove tranzicije protekle neprimjetno za korisnika, uređaj pohranjuje kod stanice u čijem je dometu i koristi ga kao identifikator za spajanje.

Kao što smo već spomenuli, bazne stanice moraju biti postavljene dovoljno gusto da bi se spriječilo gubljenje signala na nekim područjima. Posljedica ovoga je da se dometi pojedinih baznih stanica često preklapaju, stoga nerijetko postoji više različitih stanica na koje se neki uređaj može istovremeno spojiti. Primjer na slici 3.6 ilustrira preklapanja područja pokrivenosti signalom za 4 bazne stanice s oznakama s_1 , s_2 , s_3 i s_4 . Bazne stanice za koje postoji područje preklapanja signala nazivat ćemo susjednima. Na primjer, bazna stanica s_3 na slici 3.6 ima susjede s_1 , s_2 i s_4 . Ključno je da stanice čiji se dometi prekla-



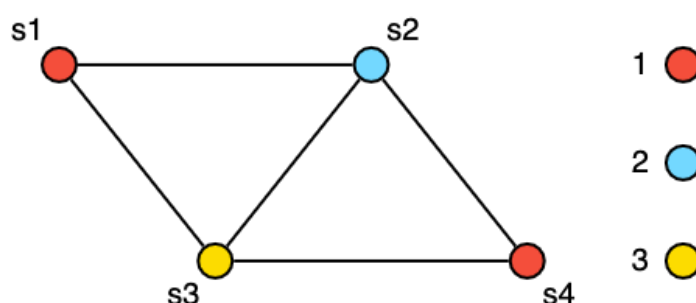
Slika 3.6: Ilustracija preklapanja područja pokrivenosti signalom za 4 bazne stanice.

paju imaju dodijeljene različite kodove kako bi uređaj koji je u dometu mogao u svakom trenutku točno identificirati stanicu na koju se spaja.

Problem koji treba riješiti je – *Kako ograničen broj kodova dodijeliti velikom broju baznih stanica tako da susjedne stanice uvijek imaju različit kod?*

Modeliranje problema grafom

Problem ćemo modelirati grafom u kojem vrhovi predstavljaju bazne stanice s bridovima među stanicama za koje postoji područje preklapanja signala, tj. vrhovi su susjedni ako predstavljaju stanice koje su susjedne. Očito je da problem dodjele kodova baznim stanicama tada postaje problem bojanja odgovarajućeg grafa. Ako postoji k dostupnih kodova koji mogu biti dodijeljeni baznim stanicama u zadanoj telekomunikacijskoj mreži, problem dodjele kodova je ekvivalentan pronalasku dopustivog k -bojanja za graf pridružen toj mreži. Za mrežu baznih stanica na slici 3.6, pripadni graf G i optimalno bojanje nalaze se na slici 3.7. Ako je dostupan skup kodova $\{1, 2, 3\}$, baznim stanicama s_1 i s_4 možemo



Slika 3.7: Pripadni graf G i optimalno bojanje za opisanu mrežu 4 bazne stanice.

dodijeliti kod 1, baznoj stanici s_2 kod 2 i baznoj stanici s_3 kod 3.

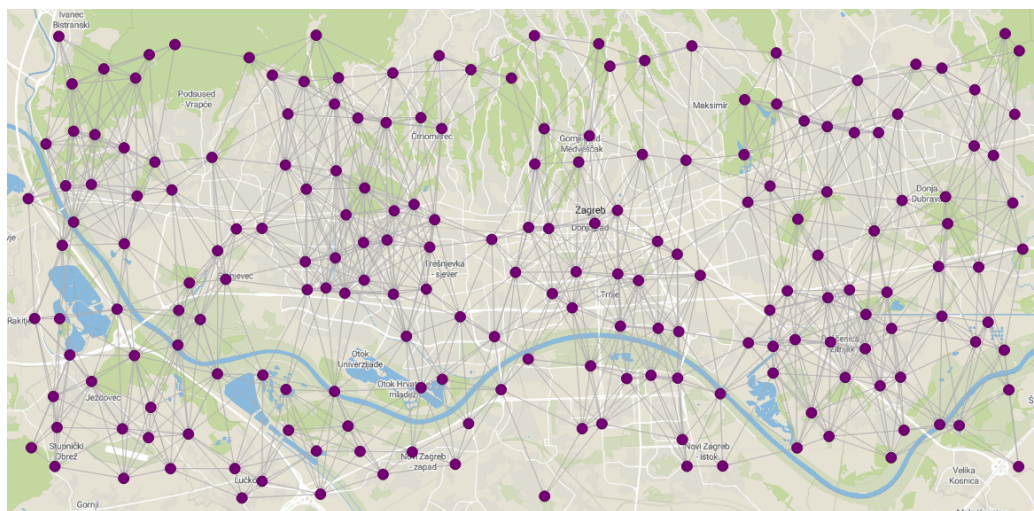
Telekomunikacijska mreža grada Zagreba

Za manje trivijalnu ilustraciju problema iskoristiti ćemo skup podataka koji sadrži mrežu baznih stanica grada Zagreba. Skup podataka je dostupan u razvojnoj okolini *Memgraph Lab* koja služi za analizu i vizualizaciju podataka pohranjenih u graf-bazi podataka *Memgraph*.³ Skup ne sadrži stvarne podatke o telekomunikacijskoj mreži, već su bazne stanice i lokacije umjetno generirane. Pri generiranju preklapanja signala, geografski i ostali elementi, koji utječu na jačinu i prostornu raspodjelu signala, modelirani su nasumično. Kod vizualizacije podataka na geografskoj karti, *Memgraph Lab* koristi ©OpenStreetMap.⁴

Graf se sastoji od 200 vrhova koji predstavljaju bazne stanice. Svaka bazna stanica ima naziv te podatke o geografskoj širini i dužini koji se koriste za prikaz podataka na karti. Dvije stanice su spojene bridom ako postoji preklapanje u područjima koja pokrivaju signalom, tj. ako su susjedne. Zbog specifičnosti pohrane podataka, na raspolaganju su nam samo usmjereni bridovi pa se efekt neusmjerenog grafa postiže na način da se, umjesto svakog neusmjerenog brida, koriste dva usmjerena brida suprotne orijentacije. U promatranoj mreži postoji 2388 usmjerenih bridova koji predstavljaju 1194 neusmjerena. U nastavku ćemo zanemariti specifičnosti implementacije i govoriti samo o neusmjerenim bridovima. Vizualizacija podataka nalazi se na slici 3.8. Vrhovi koji predstavljaju bazne stanice su prikazani kružićima. Graf je prikazan na geografskoj karti tako da su vrhovi smješteni na odgovarajućim lokacijama uz pomoć podataka o geografskoj širini i dužini. Susjedni vrhovi su povezani bridovima.

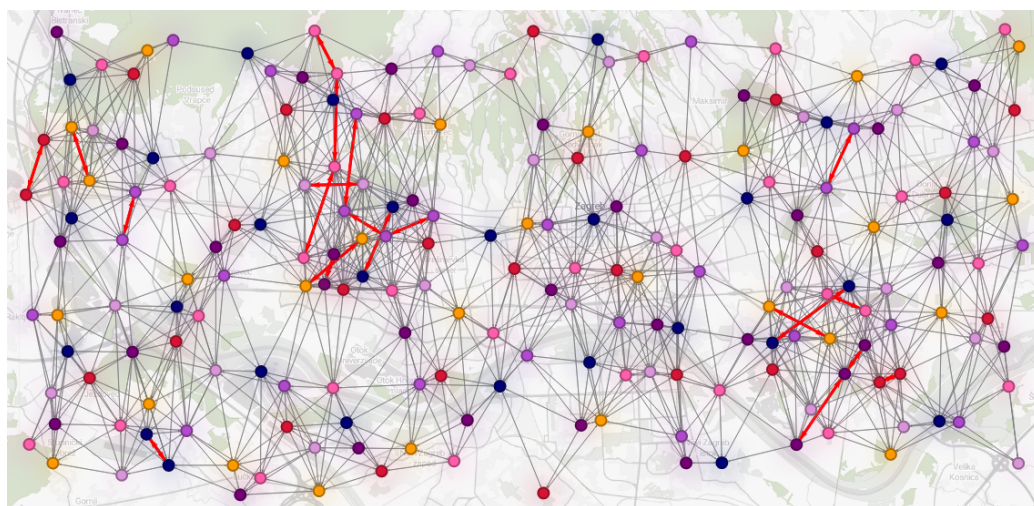
³Više o bazi podataka *Memgraph* i razvojnoj okolini *Memgraph Lab* te korištenom skupu podataka može se pronaći na https://memgraph.com/blog/optimizing_telco_networks_with_graph_coloring_and_memgraph_mage.

⁴Više informacija o ©OpenStreetMap može se pronaći na <https://www.openstreetmap.org/>.



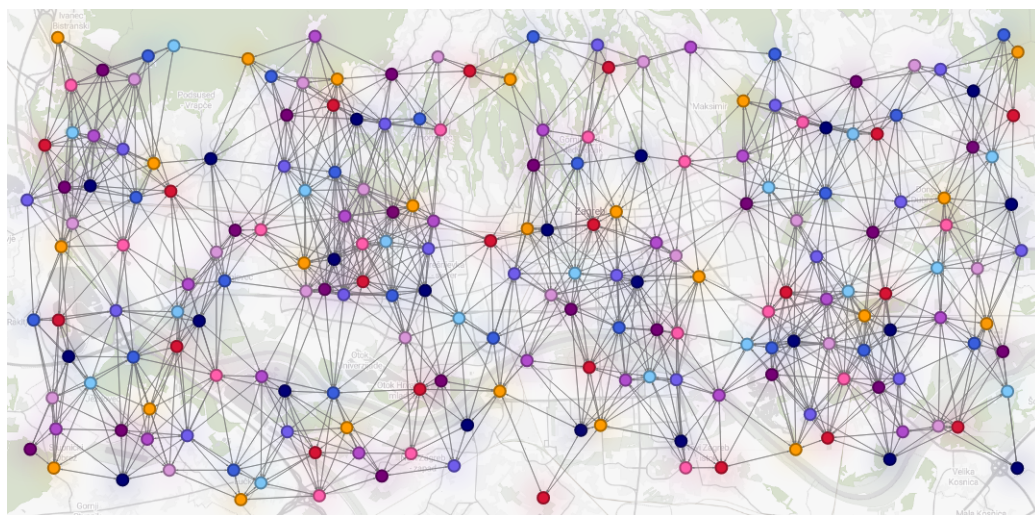
Slika 3.8: Prikaz mreže baznih stanica grada Zagreba.

Ako ovoj mreži baznih stanica pokušamo dodijeliti 7 različitih kodova, tj. ako pokušamo pripadni graf obojati sa 7 boja, uvijek će postojati konfliktni bridovi. Na slici 3.9 prikazano je 7-bojanje grafa, a konfliktni bridovi su naglašeni crvenom bojom i ukupno ih je 21. Za 8-bojanje, broj konfliktnih bridova je jedanaest, a za 9-bojanje imamo dva konfliktna brida. Ovakav rezultat je posljedica postojanja klika u grafu. Postoji točno jedanaest 9-klika i



Slika 3.9: Prikaz 7-bojanja mreže baznih stanica grada Zagreba s konfliktnim bridovima naglašenim crvenom bojom.

dvije 10-klike. Zbog postojanja 10-klike možemo zaključiti da je kromatski broj promatranog grafa najmanje 10. Zaista, ako iskoristimo deset boja dobijemo dopustivo 10-bojanje prikazano na slici 3.10. Sada slijedi da je kromatski broj ovog grafa 10, a prikazano je bojanje optimalno.



Slika 3.10: Prikaz optimalnog 10-bojanja mreže baznih stanica grada Zagreba.

Poglavlje 4

Mjere centralnosti

Grafove često koristimo za modeliranje mreža. Identificiranje najvažnijeg ili najutjecajnijeg vrha obično je ključni korak u analizi svojstava kompleksne mreže. Mjeru važnosti nekog vrha u kontekstu određene mreže nazivamo *centralnost*.

Postoji više različitih mjera centralnosti, svaka sa specifičnim karakteristikama koje proizlaze iz načina na koji se definira “važnost” vrha [12]. Bilo je mnoštvo pokušaja da se uvede reda u područje mjera centralnosti, uključujući radove koje su objavili Sabidussi [32], Koschützki i suradnici [21] i Borgatti i Everett [11]. Daleko najutjecajniji je rad kojeg je objavio Freeman 1979. godine [16], od kada se centralnosti stupnja, blizine i međupoloženosti smatraju mjerama koje obuhvaćaju najvažnije aspekte centralnosti. Još jedna podjednako prepoznata mjera centralnosti, uz tri prethodno navedene, je svojstvena centralnost koju uvodi Bonacich u svom članku [7] iz 1972. godine.

Uz ove klasike potrebno je još spomenuti i PageRank koji su razvili osnivači Googlea, Larry Page i Sergey Brin [13] 1998. godine.

U ovom poglavlju upoznat ćemo se поблише sa najvažnijim mjerama centralnosti i navesti neke njihove primjene, najčešće u analizi različitih društvenih mreža. Detaljno ćemo opisati postupak računanja PageRanka na jednostavnom primjeru i na kraju primijeniti PageRank na mrežu citata znanstvenih članaka i usporediti rezultat sa vrijednostima koje se dobiju za druge mjere centralnosti na istom skupu podataka.

4.1 Pregled mjera centralnosti

Kod odabira centralnosti koja će se koristiti u analizi konkretne mreže treba uzeti u obzir vrstu i strukturu mreže te pitanja na koja se želi odgovoriti. *Što određeni vrh čini centralnim, a neki drugi perifernim?* U ovom poglavlju ćemo pokušati odgovoriti na to pitanje.

Stupanj vrha

Konceptualno najjednostavnija i najintuitivnija mjera centralnosti je *stupanj vrha* (engl. degree centrality).

Definicija 4.1.1. U neusmjerenom grafu, **centralnost stupnja** c_d vrha v jednaka je stupnju tog vrha:

$$c_d(v) = \deg(v).$$

U usmjerenom grafu definiramo **ulaznu centralnost stupnja** v kao $c_{in}(v) = \deg_{in}(v)$ i analogno, **izlaznu centralnost stupnja** kao $c_{out} = \deg_{out}(v)$.

Dakle, vrh je važan ako ima velik stupanj. Najčešće mjere centralnosti interpretiramo u kontekstu komunikacije unutar društvenih mreža kao pokazatelj važnosti ili utjecaja nekog pojedinca. Za osobu koja je u direktnom kontaktu s velikim brojem pojedinaca možemo pretpostaviti da je u žarištu komunikacije i da je na poziciji koja se vjerojatno nalazi na glavnom toku informacija, tj. da će informacije koje teku mrežom doći do nje. Ta osoba je vjerojatno bitan akter u komunikaciji koja se odvija. S druge strane, osoba koja je u direktnom kontaktu s malim brojem pojedinaca vjerojatno je relativno izolirana od ostatka mreže i sporedan je akter u komunikacijskim procesima koji se odvijaju u mreži.

Centralnost stupnja djelomično ovisi o veličini grafa. Ponekad je korisno imati mjeru koja je neovisna o veličini grafa da bi mogli uspoređivati relativne centralnosti vrhova iz različitih grafova. U grafu koji ima n vrhova, svaki vrh može imati najviše $n - 1$ susjeda pa utjecaj veličine grafa možemo eliminirati tako da definiramo normaliziranu centralnost [16]:

$$c_d(v) = \frac{\deg(v)}{n - 1}. \quad (4.1)$$

Centralnost stupnja je *lokalna* mjera jer uzima u obzir samo susjedne vrhove, zanemarujući indirektne veze koje mogu postojati među vrhovima. Vrh s velikim brojem susjeda je važan lokalno, u svom neposrednom okružju, ali ne mora biti globalno centralan. Vrh je *globalno* centralan kada se nalazi na značajnoj poziciji u kontekstu stukture cjelokupnog grafa.

Centralnost blizine

Jedna globalna mjera centralnosti koja se temelji na međusobnoj udaljenosti vrhova je *centralnost blizine* (engl. closeness centrality). Udaljenost vrha je formalno definirana kao suma udaljenosti od ostalih vrhova, a blizina vrha je definirana kao inverz njegove udaljenosti. Prema tome, što je manja njegova ukupna udaljenost od ostalih vrhova, to će konkretni vrh biti centralniji.

Definicija 4.1.2. Neka je $G = (V, E)$ graf te $v \in V$ neki vrh. Za $u \in V$, $u \neq v$, s $\text{dist}(v, u)$ označimo duljinu najkraćeg puta između vrhova v i u , a sa n označimo $|V|$. Tada **centralnost blizine** za vrh v definiramo kao:

$$c_c(v) = \frac{n - 1}{\sum_{u \neq v} \text{dist}(v, u)}. \quad (4.2)$$

Freeman [16] iznosi dvije interpretacije centralnosti blizine u kontekstu društvenih mreža, kao mjeru pristupačnosti ili kao mjeru neovisnosti. Prva interpretacija promatra efikasnost, tj. brzinu širenja informacija. Centralna točka u mreži je ona iz koje će se informacije najbrže proširiti po ostatku mreže, tj. ona osoba koja može najefikasnije komunicirati s ostatkom mreže. Druga interpretacija promatra u kojoj mjeri je osoba neovisna o utjecajima drugih pojedinaca na protok informacija u mreži. Osoba koja se mora osloniti na mnogo posrednika u komunikaciji nije neovisna i samim time nije centralna.

Centralnost blizine je korisna za analizu povezanih grafova, međutim za nepovezane grafove nam ne daje nikakvu informaciju o važnosti vrhova. Naime, u nepovezanom grafu svaki vrh je beskonačno udaljen od barem jednog vrha pa će za svaki vrh v vrijediti $c_c(v) = 0$.

Međupovezanost vrhova

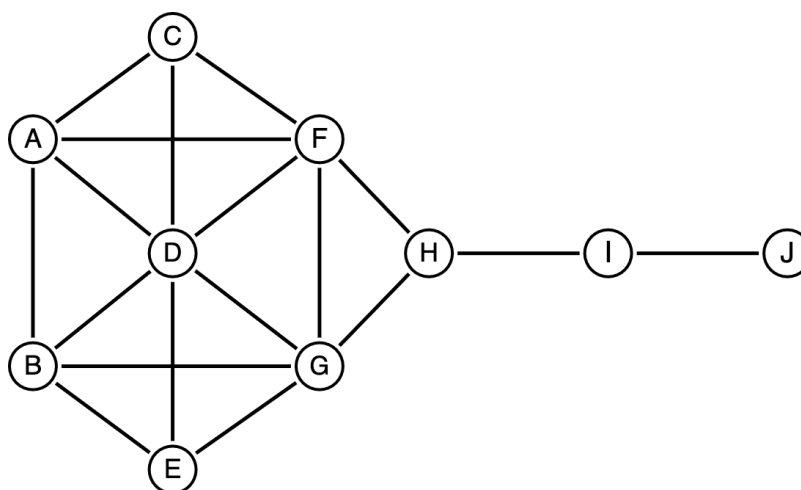
Kao i centralnost blizine, *međupovezanost vrhova* (engl. betweenness centrality) se zasniva na udaljenostima među vrhovima. Međupovezanost vrha je broj pojavljivanja toga vrha na najkraćim putevima među parovima ostalih vrhova u grafu.

Definicija 4.1.3. Neka je $G = (V, E)$ graf te $u, v, k \in V$, $u \neq v$, $u \neq k$, $v \neq k$ neki vrhovi. Označimo sa g_{uv} broj najkraćih puteva od vrha u do vrha v , a s g_{ukv} broj najkraćih puteva od vrha u do vrha v koji prolaze vrhom k . **Međupovezanost** $c_b(k)$ vrha k definiramo kao:

$$c_b(k) = \sum_{u \neq k} \sum_{v \neq k, v \neq u} \frac{g_{ukv}}{g_{uv}}. \quad (4.3)$$

Ponekad postoji više različitih najkraćih puteva među vrhovima u i v , pa $\frac{g_{ukv}}{g_{uv}}$ predstavlja vjerojatnost izbora puta koji prolazi vrhom k , tj. za svaki par vrhova sumi se dodaje 1 samo ako se k nalazi na svim mogućim najkraćim putevima među njima [10].

Dok se centralnost blizine interpretira kao mjera neovisnosti nekog vrha od ostatka mreže, međupovezanost možemo interpretirati kao mjeru ovisnosti ostatka mreže o nekom vrhu, tj. potencijal za kontrolu protoka informacija kroz mrežu. Obje interpretacije imaju pretpostavku da se informacije ili utjecaj kroz mrežu šire na najefikasniji način, tj. po najkraćim putevima. Centralnost blizine za svaki vrh u nepovezanom grafu iznosi 0 i ne daje



Slika 4.1: Krackhardtov graf zmaja.

nam nikakvu informaciju o svojstvima grafa, a međupovezanost nema taj problem. U nepovezanom grafu svaki par vrhova između kojih ne postoji najkraći put samo ne doprinosi međupovezanosti ostalih vrhova.

Primjer 4.1.4. U svom radu [22] iz 1990. godine, David Krackhardt predstavlja “graf zmaja” prikazan na slici 4.1. Ovaj jednostavni graf se sastoji od samo 10 vrhova, ali za svaku od tri prethodno opisane mjere centralnosti ima maksimalnu vrijednost u različitim vrhovima.

- Vrh D ima najveću centralnost stupnja.
- Vrhovi F i G imaju najveću centralnost blizine.
- Vrh H ima najveću vrijednost međupovezanosti.

Svojstvena centralnost

1972. godine Bonacich [7] je predložio da bi svojstveni vektor matrice susjedstva (definicija 1.1.10) pridružen najvećoj svojstvenoj vrijednosti mogao biti dobra mjera centralnosti. *Svojstvena centralnost* (engl. *eigenvector centrality*) je proširenje centralnosti stupnja gdje osim broja susjeda ulogu igra i koliko su ti susjedi važni. Centralnost vrha je proporcionalna sumi centralnosti njegovih susjeda, tj. važnost vrha raste ako je povezan sa drugim važnim vrhovima.

Definicija 4.1.5. Neka je \mathbf{A}_G matrica susjedstva pridružena grafu G . Svojstvenu centralnost x_i vrha i računamo na sljedeći način:

$$\lambda x_i = \sum_{j=1}^n a_{ij} x_j, \quad i = 1, \dots, n; \quad (4.4)$$

gdje je λ najveća svojstvena vrijednost matrice \mathbf{A}_G , a n broj vrhova grafa G . Ako (4.4) zapišemo u matričnom obliku:

$$\mathbf{A}_G x = \lambda x, \quad (4.5)$$

svojstvena centralnost vrha i je jednaka i -toj komponenti svojstvenog vektora x pridruženog najvećoj svojstvenoj vrijednosti matrice \mathbf{A}_G .

Primjer 4.1.6. Promotrimo graf G na slici 4.2. Pripadna matrica susjedstva \mathbf{A}_G je

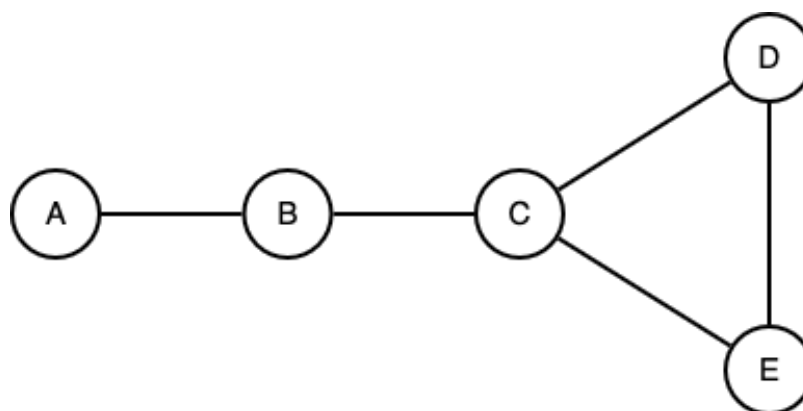
$$\mathbf{A}_G = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}.$$

Svojstvene vrijednosti matrice \mathbf{A}_G su: $\lambda_1 = 1, \lambda_2 = -1, \lambda_3 \approx 1.6751, \lambda_4 \approx 0.5391$ i $\lambda_5 \approx 2.2143$. Najveća svojstvena vrijednost je λ_5 i njoj je pridružen svojstveni vektor:

$$x \approx \begin{bmatrix} 0.3111 \\ 0.6888 \\ 1.2143 \\ 1 \\ 1 \end{bmatrix},$$

čiji elementi odgovaraju vrijednostima svojstvene centralnosti odgovarajućih vrhova grafa G .¹ Očekivano, vrh C ima najveću vrijednost svojstvene centralnosti, a vrh A najmanju. Zanimljivo je usporediti vrijednosti svojstvene centralnosti i centralnosti stupnja u tablici 4.3.

¹Kako ne bi došlo do zabune s vektorima koji su označeni malim slovima, u nastavku ćemo vrhove označavati velikim slovima. Treba pripaziti da se graf također označava velikim slovom, međutim iz konteksta bi trebalo biti očito na što se oznaka odnosi.

Slika 4.2: Graf G .

Vrh	Svojstvena centralnost	Centralnost stupnja
A	0.3111	1
B	0.6888	2
C	1.2143	3
D	1	2
E	1	2

Tablica 4.3: Vrijednosti svojstvene centralnosti i centralnosti stupnja za vrhove grafa G .

Vrhovi s najvećim i najmanjim vrijednostima se podudaraju. Vrhovi B, D i E svi imaju jednaki stupanj, ali različite vrijednosti svojstvene centralnosti. Ovakav rezultat ima smisla jer važniji susjedi doprinose više rezultatu od manje važnih susjeda. Sva tri vrha imaju po dva susjeda od kojih je jedan vrh C. Bitna razlika je to što je vrhu B drugi susjed vrh A koji je najmanje bitan, dok su D i E međusobno susjedni. Kada promotrimo graf G odmah možemo zaključiti da D i E imaju jednaku centralnost i da bi trebali biti centralniji od B, dok je C najcentralniji vrh.

Svojstvena centralnost se interpretira kao mjera utjecaja ili statusa pojedinca u nekoj društvenoj mreži. Pretpostavka je da osoba koja ima malo prijatelja ali je povezana s utjecajnim ljudima ima veći utjecaj ili status od neke osobe koja ima mnogo prijatelja bez utjecaja. – “Nije bitno koliko ljudi poznaješ nego koga poznaješ.”

Varijacija svojstvene centralnosti je Googleov PageRank koji ćemo opisati u nastavku.

PageRank

PageRank je algoritam kojeg Google koristi za rangiranje web-stranica. Algoritam su razvili osnivači Googlea, Larry Page i Sergey Brin [13] 1998. godine. Oni daju interpretaciju PageRanka kao modela ponašanja korisnika kojega nazivaju “*slučajni posjetitelj*” (engl. random surfer). Pretpostavka je da slučajni posjetitelj kojemu je dana nasumično odabrana početna web-stranica nastavlja klikati poveznice (bez povratka natrag) dok mu ne postane dosadno. U trenutku kad mu je postalo dosadno, slučajni posjetitelj, umjesto da klikne na neku od poveznica na trenutnoj stranici, započinje cijeli proces ponovno sa nove nasumično odabrane stranice. Vjerojatnost da slučajni posjetitelj posjeti neku web-stranicu je upravo vrijednost njezinoga PageRanka.

Možda još intuitivnija interpretacija je da web-stranica ima visoki rang ako se na nju referira mnogo drugih stranica ili neke od stranica koje se referiraju na nju i same imaju visoki rang. Logika iza ove pretpostavke je da ako web-stranicu citiraju mnoge druge web-stranice onda je vjerojatno vrijedna pažnje. Također, ako na web-stranicu vodi poveznica sa neke bitne i pouzdane stranice, onda je vjerojatno da je njen sadržaj relevantan. Lagano je uočiti sličnost sa interpretacijom svojstvene centralnosti. Precizna definicija uz konkretan primjer i objašnjenje nalazi se u sljedećem nastavku.

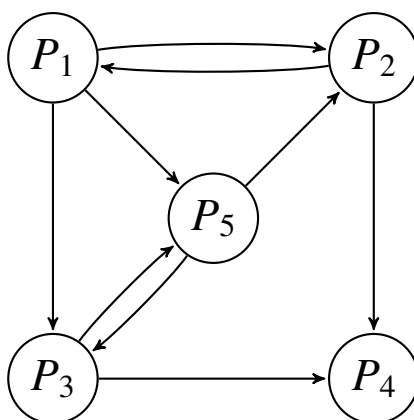
4.2 Primjer računanja PageRanka

Internet možemo promatrati kao jedan veliki digraf gdje su vrhovi web-stranice, a svaka poveznica usmjereni brid između web-stranica među kojima postoji poveznica. PageRank je još jedna metoda računanja važnosti vrhova u grafu, tj. mjera centralnosti. U nastavku pratimo [29] i [18].

Da bi lakše demonstrirali kako funkcionira PageRank, promotrimo jednostavan model interneta koji se sastoji od samo 5 web-stranica prikazan na slici 4.4. Vrhovi usmjerenoga grafa D predstavljaju web-stranice P_1, P_2, P_3, P_4 i P_5 , a usmjereni brid od vrha P_i do P_j postoji samo ako na stranici P_i postoji poveznica koja vodi na stranicu P_j . Izlazni stupanj vrha P_i predstavlja ukupan broj poveznica na odgovarajućoj stranici. Na primjer, stranica P_2 ima $\deg_{\text{out}}(P_2) = 2$ zato što su na njoj poveznice na P_1 i P_4 . Sada možemo definirati **rang** stranice P , tj. mjeru PageRank centralnosti, kao realni broj [29]:

$$r(P) = \sum_{Q \in B_P} \frac{r(Q)}{\deg_{\text{out}}(Q)}, \quad (4.6)$$

pri čemu $r(P)$ predstavlja rang stranice P , a B_P je skup svih stranica koje imaju poveznicu prema P . Ako je $B_P = \emptyset$ tada je $r(P) = 0$. Definicija je rekurzivna, rang stranice P ovisi o rangovima svih stranica $Q \in B_P$. Doprinos svake stranice ovisi o broju $\deg_{\text{out}}(Q)$ poveznica koje se na njoj nalaze. Ako stranice P i Q imaju poveznice jedna na drugu onda rang od P

Slika 4.4: Digraf D je model interneta koji se sastoji od 5 web-stranica.

ovisi o rangu stranice Q i obrnuto. Stoga definiramo iterativni postupak za računanje ranga stranice P kao [29]:

$$r_{k+1}(P) = \sum_{Q \in B_P} \frac{r_k(Q)}{\deg_{\text{out}}(Q)}, \quad (4.7)$$

gdje je $r_k(P)$ rang stranice P nakon k iteracija. U prvom koraku svim stranicama dajemo jednak početni rang pa je $r_0(P) = \frac{1}{n}$ gdje je n ukupan broj web-stranica. Početne rangove stranica iz primjera sa slike 4.4 možemo zapisati kao vektor:

$$z_0 = (r_0(P_1), r_0(P_2), r_0(P_3), r_0(P_4), r_0(P_5)) = \left(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5} \right).$$

Sada možemo izračunati prvu iteraciju (4.7):

$$\begin{aligned} r_1(P_1) &= \frac{r_0(P_2)}{\deg_{\text{out}}(P_2)} = \frac{\frac{1}{5}}{2} = \frac{1}{10}, \\ r_1(P_2) &= \frac{r_0(P_1)}{\deg_{\text{out}}(P_1)} + \frac{r_0(P_5)}{\deg_{\text{out}}(P_5)} = \frac{\frac{1}{5}}{3} + \frac{\frac{1}{5}}{2} = \frac{1}{6}, \\ r_1(P_3) &= \frac{r_0(P_1)}{\deg_{\text{out}}(P_1)} + \frac{r_0(P_5)}{\deg_{\text{out}}(P_5)} = \frac{\frac{1}{5}}{3} + \frac{\frac{1}{5}}{2} = \frac{1}{6}, \\ r_1(P_4) &= \frac{r_0(P_2)}{\deg_{\text{out}}(P_2)} + \frac{r_0(P_3)}{\deg_{\text{out}}(P_3)} = \frac{\frac{1}{5}}{2} + \frac{\frac{1}{5}}{2} = \frac{1}{5}, \\ r_1(P_5) &= \frac{r_0(P_1)}{\deg_{\text{out}}(P_1)} + \frac{r_0(P_3)}{\deg_{\text{out}}(P_3)} = \frac{\frac{1}{5}}{3} + \frac{\frac{1}{5}}{2} = \frac{1}{6}. \end{aligned}$$

Nakon prve iteracije vektor ranga jest:

$$z_1 = \left(\frac{1}{10}, \frac{1}{6}, \frac{1}{6}, \frac{1}{5}, \frac{1}{6} \right).$$

Sljedeća iteracija nam daje:

$$\begin{aligned} r_2(P_1) &= \frac{r_1(P_2)}{\deg_{\text{out}}(P_2)} = \frac{\frac{1}{6}}{2} = \frac{1}{12}, \\ r_2(P_2) &= \frac{r_1(P_1)}{\deg_{\text{out}}(P_1)} + \frac{r_1(P_5)}{\deg_{\text{out}}(P_5)} = \frac{\frac{1}{10}}{3} + \frac{\frac{1}{6}}{2} = \frac{7}{60}, \\ r_2(P_3) &= \frac{r_1(P_1)}{\deg_{\text{out}}(P_1)} + \frac{r_1(P_5)}{\deg_{\text{out}}(P_5)} = \frac{\frac{1}{10}}{3} + \frac{\frac{1}{6}}{2} = \frac{7}{60}, \\ r_2(P_4) &= \frac{r_1(P_2)}{\deg_{\text{out}}(P_2)} + \frac{r_1(P_3)}{\deg_{\text{out}}(P_3)} = \frac{\frac{1}{6}}{2} + \frac{\frac{1}{6}}{2} = \frac{1}{6}, \\ r_2(P_5) &= \frac{r_1(P_1)}{\deg_{\text{out}}(P_1)} + \frac{r_1(P_3)}{\deg_{\text{out}}(P_3)} = \frac{\frac{1}{10}}{3} + \frac{\frac{1}{6}}{2} = \frac{7}{60}, \end{aligned}$$

pa je novi vektor ranga:

$$z_2 = \left(\frac{1}{12}, \frac{7}{60}, \frac{7}{60}, \frac{1}{6}, \frac{7}{60} \right).$$

Ovo nije konačni rang stranica iz primjera koji promatramo. Konačan rang ćemo dobiti daljnjim ponavljanjem postupka (4.7). Postavlja se pitanje, kako izgleda konačan vektor ranga i postoji li uopće. Odgovore na ova pitanja lakše ćemo pronaći ako gornji postupak zapišemo u matičnom obliku.

Moramo mapirati internet u matricu koja reprezentira strukturu međusobno povezanih web-stranica. Za graf D , koji predstavlja naš jednostavan model interneta, konstruirat ćemo pripadnu **matricu poveznica** \mathbf{H} na sličan način kao što to radimo za matricu susjedstva. Općenito, $\mathbf{H} = [h_{i,j}]$ je kvadratna matrica reda n pri čemu je n ukupan broj web-stranica, a $h_{i,j} = \frac{1}{\deg_{\text{out}}(P_j)}$ vjerojatnost da korisnik koji se nalazi na stranici P_j klikne na poveznicu koja vodi na stranicu P_i . Ako na stranici P_j ne postoji poveznica koja vodi na stranicu P_i , tada je $h_{i,j} = 0$. Matrica poveznica \mathbf{H} pridružena grafu D je:

$$\mathbf{H} = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{2} & 0 & 0 \end{bmatrix}.$$

Na primjer, sa stranice P_5 postoji poveznica na P_3 i $\deg_{\text{out}}(P_5) = 2$ pa je $h_{3,5} = \frac{1}{2}$. Isto tako na stranici P_5 ne postoji poveznica na P_4 pa je $h_{4,5} = 0$. Ako je z_k vektor ranga nakon k iteracija:

$$z_k = \begin{bmatrix} r_k(P_1) \\ r_k(P_2) \\ \vdots \\ r_k(P_n) \end{bmatrix},$$

tada (4.7) možemo zapisati u matričnom obliku na sljedeći način

$$z_{k+1} = \mathbf{H}z_k, \quad k = 0, 1, 2, \dots \quad (4.8)$$

gdje je \mathbf{H} matrica poveznica. Prve dvije iteracije našeg primjera u novom matričnom zapisu izgledaju ovako:

$$z_1 = \mathbf{H}z_0 = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{5} \\ \frac{1}{5} \\ \frac{1}{5} \\ \frac{1}{5} \\ \frac{1}{5} \end{bmatrix} = \begin{bmatrix} \frac{1}{10} \\ \frac{1}{6} \\ \frac{1}{6} \\ \frac{1}{5} \\ \frac{1}{6} \end{bmatrix},$$

$$z_2 = \mathbf{H}z_1 = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{10} \\ \frac{1}{6} \\ \frac{1}{6} \\ \frac{1}{5} \\ \frac{1}{6} \end{bmatrix} = \begin{bmatrix} \frac{1}{12} \\ \frac{7}{60} \\ \frac{7}{60} \\ \frac{1}{6} \\ \frac{7}{60} \end{bmatrix}.$$

Uzastopne iteracije (4.8) se može zapisati i preko početnog vektora z_0 i potencija matrice \mathbf{H} u obliku:

$$z_k = \mathbf{H}^k z_0, \quad k = 1, 2, 3, \dots \quad (4.9)$$

Ovaj postupak se naziva **metoda potencija** i koristi se za računanje svojstvenog vektora pridruženog najvećoj (po apsolutnoj vrijednosti) svojstvenoj vrijednosti matrice. Množimo početni vektor z_0 i potencije matrice \mathbf{H} dok se vrijednosti rezultata ne prestanu mijenjati i dobijemo konačni (ako postoji) vektor ranga koji je zapravo svojstveni vektor matrice susjedstva. Dakle, ako je:

$$\lim_{k \rightarrow \infty} \mathbf{H}^k z_0 = z, \quad (4.10)$$

tada su komponente vektora z redom rangovi stranica P_1, P_2, \dots, P_n , tj. nakon određenog broja iteracija vektor z_k predstavlja dovoljno dobru aproksimaciju vektora z . Egzistencija i

jedinstvenost vektora z ovisi o svojstvima matrice \mathbf{H} . U trenutnom obliku \mathbf{H} ne garantira konvergenciju metode potencija, ali uz nekoliko modifikacija možemo osigurati konvergenciju. Želimo da \mathbf{H} bude stohastička, ireducibilna matrica.

Definicija 4.2.1. Za kvadratnu matricu kažemo da je *stohastička po stupcima* ako su svi njezini elementi nenegativni i suma elemenata u svakom stupcu je jednaka 1.

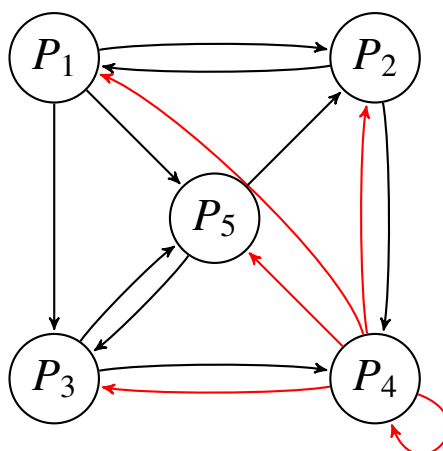
Definicija 4.2.2. Za kvadratnu matricu pridruženu usmjerenom grafu G kažemo da je *ireducibilna* ako je G jako povezan graf.

Ireducibilne nenegativne matrice imaju jedinstvenu, pozitivnu najveću svojstvenu vrijednost kojoj je pridružen jedinstveni nenegativni svojstveni vektor [27]. Po stupcima stohastička matrica je nenegativna i ima najveću svojstvenu vrijednost 1. Ako je \mathbf{H} ireducibilna i stohastička matrica tada će postojati (4.10) i metoda potencija će konvergirati. Za više informacija čitatelja se usmjerava na izvore [4] i [27].

Svi elementi matrice \mathbf{H} su nenegativni. Broj nenegativnih elemenata u svakom stupcu jednak je izlaznom stupnju vrha pridruženog odgovarajućoj stranici P_i pa će suma elemenata stupca biti $\deg_{\text{out}}(P_i) \cdot \frac{1}{\deg_{\text{out}}(P_i)} = 1$. Problem predstavljaju stranice koje nemaju poveznice na druge stranice, tj. vrhovi bez izlaznih bridova koje nazivamo **viseći vrhovi** [18]. Da bi \mathbf{H} postala stohastička, moramo eliminirati nulstupce koji predstavljaju viseće vrhove. Jednostavna modifikacija je zamijeniti nulstupce vektorom $v = (\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$ kojemu je suma elemenata 1. Ovime dobivamo novu matricu \mathbf{S} koja je stohastička. Ova modifikacija implicira da smo visećem vrhu umjetno dodali izlazne bridove prema svim ostalim vrhovima, tj. na stranicu bez poveznica smo dodali poveznice prema svim ostalim stranicama. Ovo možemo opravdati ako se prisjetimo da smo element $h_{i,j}$ matrice poveznica \mathbf{H} definirali kao vjerojatnost da korisnik koji se nalazi na stranici P_j klikne na poveznicu koja vodi na stranicu P_i . Po modelu slučajnog posjetitelja, ako korisnik dođe na stranicu koja nema poveznice onda svoje pretraživanje može nastaviti tako da upiše web-adresu bilo koje stranice na internetu, tj. tako da krene opet od neke nasumične web-stranice kao što je to slučaj kad mu postane dosadno. Vjerojatnost da korisnik na ovakav način skoči na neku stranicu je $\frac{1}{n}$, pa modifikacija ima smisla.

U matrici poveznica \mathbf{H} pridruženoj grafu D sa slike 4.4 problematičan je je stupac pridružen vrhu P_4 . Ako taj stupac zamijenimo vektorom $v = (\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5})$ dobivamo matricu:

$$\mathbf{S} = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{5} & 0 \\ \frac{1}{3} & 0 & 0 & \frac{1}{5} & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & \frac{1}{5} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{5} & 0 \\ \frac{1}{3} & 0 & \frac{1}{2} & \frac{1}{5} & 0 \end{bmatrix}.$$



Slika 4.5: Modificirani usmjereni graf kojemu je pridružena matrica S .

Matrica S je stohastička po stupcima i odgovara joj modificirani digraf na slici 4.5. Da bismo osigurali konvergenciju metode potencija, matrica S mora biti ireducibilna, tj. usmjereni graf kojem je pridružena mora biti jako povezan. Graf D na slici 4.4 nije jako povezan jer, na primjer ne postoji put između vrhova P_4 i P_1 pa inicijalna matrica H nije ireducibilna. Digraf pridružen matrici S je jako povezan jer smo dodavanjem izlaznih bridova vrhu P_4 dobili puteve među svim vrhovima grafa pa je S ireducibilna matrica. Međutim, modifikacijom H da bi dobili stohastičku matricu ne dobijemo uvijek ireducibilnu matricu S kao što se dogodilo u našem primjeru. Da bi osigurali ireducibilnost u svim slučajevima, svakom vrhu usmjerenoga grafa kojemu je pridružena matrica S dodamo izlazne bridove prema svim ostalim vrhovima, bez obzira postoje li već. Na taj način smo osigurali da je dobiveni graf nužno jako povezan.

Matricu pridruženu novo-dobivenom usmjerenom grafu dobijemo tako da matrici S pribrojimo matricu E kojoj su stupci jednaki $v = (\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$, gdje je v **vektor personalizacije** kojemu su svi elementi pozitivni i suma elemenata je jednaka 1 [18]. Ova modifikacija predstavlja umjetno dodavanje poveznica između svih stranica. Implikacija je opet da korisnik u svakom trenutku ima opciju upisati web-adresu nasumične stranice i skočiti na nju bez obzira postoje li poveznice na trenutnoj stranici ili ne. Izbor matrice E , tj. vektora personalizacije v može značajno utjecati na konačnu vrijednost vektora ranga. Pristup koji se često koristi, i koji smo mi izabrali, je uzeti v tako da predstavlja situaciju u kojoj je vjerojatnost skoka na bilo koju stranicu interneta, neovisno o trenutnoj, jednaka. Tada je PageRank uistinu globalna mjera centralnosti vrhova [17]. Ponekad više smisla ima pristup gdje se v konstruira na način da prednost daje nekom posebno izabranom skupu stranica ili jednoj stranici. Takav izbor može za rezultat imati da slučajni posjetitelj nikad ne ode

predaleko od inicijalnog skupa stranica na koji se iznova vraća svaki put kad napravi skok. Tada će najveće rangove imati stranice koje se nalaze u susjedstvu preferiranog skupa, a ostatak grafa će imati male vrijednosti. Ovakav pristup može otkriti neka zanimljiva svojstva nekog manjeg dijela grafa koja bi se možda “izgubila” iz perspektive cijelog grafa. U našem primjeru nova matrica je:

$$\mathbf{S} + \mathbf{E} = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{5} & 0 \\ \frac{1}{3} & 0 & 0 & \frac{1}{5} & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & \frac{1}{5} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{5} & 0 \\ \frac{1}{3} & 0 & \frac{1}{2} & \frac{1}{5} & 0 \end{bmatrix} + \begin{bmatrix} \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{bmatrix} = \begin{bmatrix} \frac{1}{5} & \frac{7}{10} & \frac{1}{5} & \frac{2}{5} & \frac{1}{5} \\ \frac{8}{15} & \frac{1}{5} & \frac{1}{5} & \frac{2}{5} & \frac{7}{10} \\ \frac{8}{15} & \frac{1}{5} & \frac{1}{5} & \frac{2}{5} & \frac{7}{10} \\ \frac{1}{5} & \frac{7}{10} & \frac{7}{10} & \frac{2}{5} & \frac{1}{5} \\ \frac{8}{15} & \frac{1}{5} & \frac{7}{10} & \frac{2}{5} & \frac{1}{5} \end{bmatrix}.$$

Elementi matrice $\mathbf{S} + \mathbf{E}$ su nenegativni i ona je ireducibilna ali sume elemenata po stupcima više nisu 1 pa $\mathbf{S} + \mathbf{E}$ nije stohastička po stupcima. Ovaj problem rješavamo na način da odaberemo neki $\alpha \in (0, 1)$ i definiramo novu matricu \mathbf{G} na sljedeći način [18]:

$$\mathbf{G} = \alpha \mathbf{S} + (1 - \alpha) \mathbf{E}. \quad (4.11)$$

Sada (4.10) možemo zapisati:

$$\lim_{k \rightarrow \infty} \mathbf{G}^k z_0 = \mathbf{z}. \quad (4.12)$$

Metoda potencija konvergira jer je matrica \mathbf{G} ireducibilna i stohastička [27]:

$$z_k = \mathbf{G}^k z_0, \quad k = 1, 2, 3, \dots \quad (4.13)$$

Broj α u (4.11) nazivamo **faktor prigušenja**. U modelu slučajnog posjetitelja, α interpretiramo kao vjerojatnost da u svakom koraku korisnik nastavi nasumično slijediti poveznice, tj. da će kliknuti na neku od poveznica na stranici na kojoj se nalazi. S druge strane, $(1 - \alpha)$ je vjerojatnost u svakom koraku da je korisniku dosadilo klikanje poveznica i da će skočiti na neku nasumičnu stranicu na internetu. Ovdje do izražaja dolazi izbor matrice \mathbf{E} . Možemo reći da α predstavlja važnost inicijalne strukture interneta u procesu izračunavanja ranga stranica, tj. u kojoj mjeri početna struktura reprezentirana matricom \mathbf{S} doprinosi konačnom rezultatu. U originalnoj implementaciji za Google, pri računanju ranga stranica korištena je vrijednost $\alpha = 0.85$ [13]. U našem primjeru \mathbf{G} je:

$$\mathbf{G} = 0.85 \mathbf{S} + 0.15 \mathbf{E} = \begin{bmatrix} \frac{3}{100} & \frac{91}{200} & \frac{3}{100} & \frac{1}{5} & \frac{3}{100} \\ \frac{47}{150} & \frac{3}{100} & \frac{3}{100} & \frac{1}{5} & \frac{91}{200} \\ \frac{47}{150} & \frac{3}{100} & \frac{3}{100} & \frac{1}{5} & \frac{91}{200} \\ \frac{3}{100} & \frac{91}{200} & \frac{91}{200} & \frac{1}{5} & \frac{3}{100} \\ \frac{47}{150} & \frac{3}{100} & \frac{91}{200} & \frac{1}{5} & \frac{3}{100} \end{bmatrix}.$$

$z_1 =$	0.149 0.2056666666 0.2056666666 0.234 0.2056666666	$z_2 =$	0.1571883333 0.199405 0.199405 0.2445966666 0.199405
$z_3 =$	0.15632855833 0.20086525277 0.20086525277 0.24107568333 0.20086525277	$z_4 =$	0.15635059859 0.20064369012 0.20064369012 0.24171833102 0.20064369012
$z_5 =$	0.15636568457 0.20066502084 0.20066502084 0.24163925288 0.20066502084	$z_6 =$	0.15636130685 0.20066491748 0.20066491748 0.24164394071 0.20066491748

Tablica 4.6: Vrijednosti vektora z_k u prvih 6 iteracija metode potencija.

Prvih šest iteracija metode potencija (4.13) nalazi se u tablici 4.6. U posljednje četiri iteracije prve tri decimale elemenata se ne mijenjaju i suma im je približno jednaka 1, pa je za ovaj primjer to dovoljno dobra aproksimacija. Vektor ranga $z = (r(P_1), r(P_2), r(P_3), r(P_4), r(P_5))$ je:

$$z = (0.156, 0.201, 0.201, 0.242, 0.201).$$

Stranica P_1 je najlošije rangirana što ima smisla jer na nju vodi samo jedna poveznica. P_2, P_3 i P_5 sve imaju jednaki rang, a P_4 je najbolje rangirana iako sve četiri stranice imaju po dvije poveznice koje vode na njih. Ovaj poredak je također očekivan jer na P_1 , koja je najlošije rangirana, ne postoji poveznica prema P_4 dok postoji prema preostalim stranicama.

4.3 Neke primjene mjera centralnosti

Prognoziranje pobjednika u sportu

PageRank se može koristiti u sportu za rangiranje ekipa koje sudjeluju na turniru i za prognoziranje pobjednika u konkretnim utakmicama. Govan i Meyer [18] su iskoristili

PageRank za rangiranje ekipa NFL lige američkog nogometa, a Mundar i Horvat [28] za prognoziranje ishoda Europskog rukometnog prvenstva. Analizira se takozvana mreža pobjednika, usmjereni težinski graf u kojem je svaki vrh predstavlja jednog sudionika turnira, a bridovi predstavljaju odigrane utakmice. Smjer brida ide od tima koji je izgubio prema timu koji je pobijedio, a težina je razlika u rezultatu (na primjer, razlika u broju postignutih golova). Poredak po vrijednosti PageRanka je dobar pokazatelj relativne snage timova, ali nije uvijek najbolja predikcija konačnog poretka. Rezultati turnira često ovise i o drugim faktorima pa najsposobnije ekipe ne završe uvijek na vrhu. Za prognoziranje pobjednika konkretne utakmice točnost PageRanka je dosta velika. U analizi koju su napravili Mundar i Horvat [28], PageRank točno prognozira pobjednika u 71% slučajeva, a ako se izuzmu utakmice koje su odigrane izjednačeno, preciznost raste na 79% [17].

Citiranje akademskih članaka

Možemo konstruirati graf u kojem vrhovi predstavljaju akademske članke, a usmjereni bridovi iz jednog vrha pokazuju na sve članke koji su citirani u promatranom članku. Ovakvi grafovi mogu imati tisuće vrhova i nekoliko milijuna bridova. Prilagođena verzija PageRanka može se iskoristiti u analizi ovakvih grafova za pronalazak “prikrivenih dragulja”, članaka koji nisu citirani veliki broj puta, ali su značajni. Ključna modifikacija je kod izbora vektora personalizacije i faktora prigušenja. Stariji članci su obično favorizirani u sličnim analizama zato što kroz veći period mogu skupiti više citata nego nedavno objavljeni članci. Ako vektor personalizacije konstruiramo tako da vrijednosti budu recipročne starosti članka, onda će noviji članci biti preferirani. U analizi članaka iz područja fizike objavljenih između 1893. i 2003. koju su napravili Chen i suradnici [14], pokazalo se da su broj citata i rang članka pozitivno korelirani. Međutim, pojavile su se i neke iznimke s visokim rangom unatoč relativno malom broju citata i to su upravo “skriveni dragulji”, tj. dobro poznati članci među fizičarima, ali sa malim brojem citata [17].

Epidemiologija

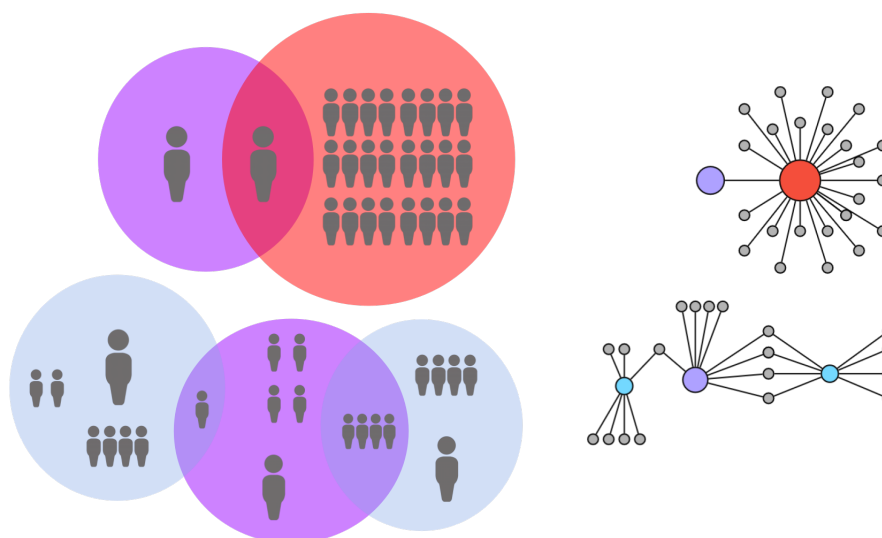
Razumijevanje dinamike širenja infekcija u populaciji ključ je prevencije i kontrole epidemije. Standardni epidemiološki modeli često se baziraju na hipotezi homogenog miješanja koje pretpostavlja jednaku vjerojatnost prijenosa infekcije među svim jedinkama u populaciji, bez obzira na dob, zaposlenje, lokaciju i različite uzorke ponašanja [34]. Međutim, u stvarnim populacijama broj interakcija svake jedinice značajno je manji od ukupnog broja jedinica pa u takvim uvjetima ne dolazi do homogenog miješanja. Interakcija među jedinkama u kojoj je moguć prijenos određene infekcije naziva se **kontakt**. Pretpostavku o homogenom miješanju moguće je izbjeći ako se širenje infekcije modelira grafom. U

ovakvom modelu vrhovi grafa predstavljaju jedinke, a bridovi kontakte. Uglavnom je, u interakcijama, prijenos infekcije moguć u oba smjera pa se kontakt modelira neusmjerenim bridom. Međutim, u nekim slučajevima, kao na primjer kod transfuzije krvi, infekcija je moguća samo u jednom smjeru pa je odgovarajući graf usmjeren. Analiza strukture ovih grafova može pomoći u razumijevanju i predviđanju dinamike širenja infekcije i pronalaženju efikasnih mjera suzbijanja. Prioritet je identificiranje pojedinaca koji, ako su zaraženi, imaju potencijal proširiti infekciju na relativno velik dio populacije. To su obično jako povezani i utjecajni pojedinci s velikim brojem kontakata, a nazivamo ih **superširitelji**. Superširitelji igraju ključnu ulogu u širenju i održavanju infekcije u populaciji, ali su i u većoj opasnosti da budu izloženi infekciji [20]. Za identificiranje potencijalnih superširitelja ili najugroženijih skupina moguće je iskoristiti mjere centralnosti.

Međupovezanost vrhova u kontekstu društvene mreže interpretirala se kao potencijal za kontrolu protoka informacija kroz mrežu. U kontekstu širenja infekcije, međupovezanost bi se mogla iskoristiti kao alat u nastojanjima suzbijanja epidemije. Naime, ako iz odgovarajućeg grafa uklonimo vrh s visokom vrijednosti međupovezanosti, brzina i vjerojatnost širenja infekcije do nekog proizvoljnog vrha u grafu će se značajnije smanjiti nego što bi to bio slučaj da smo uklonili neki vrh s nižom vrijednosti međupovezanosti [9]. Ovaj pristup mogao bi se upotrijebiti za ciljanu imunizaciju u samim počecima epidemije kako bi se malim brojem cjepiva postigao značajniji efekt u suzbijanju epidemije [20].

Centralnost stupnja možemo tumačiti kao potencijal neke osobe za direktno širenje infekcije, ali i kao mjeru izloženosti infekciji, tj. vjerojatnosti da se osoba zarazi [9]. S brojem kontakata koje osoba ima raste i vjerojatnost infekcije, no ovakav je pristup dosta ograničen jer uzima u obzir samo direktne kontakte. Bolju procjenu izloženosti pojedinca dobit ćemo ako iskoristimo svojstvenu centralnost.

Od početka pandemije virusa SARS-CoV-2 različitim se mjerama, kao što su praćenje kontakata i fizičko distanciranje, pokušava kontrolirati širenje virusa u populaciji. Jedna od preporuka je i formiranje takozvanih “društvenih mjehurića”, tj. ograničenog broja bliskih osoba s kojima se ne prakticira fizičko distanciranje. Cilj “društvenih mjehurića” je omogućiti određenu razinu kontakata među ljudima, a istovremeno smanjiti rizik od prijenosa virusa. Ova ideja ilustrirana je na slici 4.7. Mjehurići se obično formiraju na razini kućanstva ili među prijateljima i kolegama ali najčešće nisu izolirani sustavi nego se preklapaju. Osoba koja u svom mjehuriću ima veći broj ljudi koji se ne ponašaju rizično ili imaju jako mali broj kontakata vjerojatno će biti manje izložena od pojedinca koji ima samo jedan visoko-rizični kontakt. Slično kao za utjecaj u društvenoj mreži, rizik od infekcije, tj. izloženost nekog pojedinca, proporcionalan je sumi izloženosti njegovih kontakata i možemo je mjeriti svojstvenom centralnosti odgovarajućeg vrha. Stoga ovaj pristup daje realističniju procjenu rizika od centralnosti stupnja.



Slika 4.7: Ilustracija “društvenih mjehurića”.

4.4 Analiza mreže citata znanstvenih članaka

Razvojem nekog znanstvenog područja i novim otkrićima u tom polju raste broj objavljenih akademskih članaka koji obrađuju teme iz tog polja. Relevantnost unutar nekog polja se mjeri različitim indeksima koji imaju svoje prednosti i mane. Relevantnost članka ili autora često se procjenjuje po broju tj. frekvenciji citata i to se koristi kao mjera utjecaja ili kvalitete.

Manje trivijalna mjera produktivnosti i citiranosti autora (ili znanstvenog časopisa) je **h-indeks** koji se definira kao maksimalna vrijednost h takva da je autor objavio barem h članaka od kojih je svaki citiran najmanje h puta. Očito ograničenje je što h-indeks nekog autora ne može biti veći od broja članaka koji je objavio. Jedan zanimljiv primjer znanstvenika sa niskim h-indeksom ali izrazito velikim doprinosom znanosti je nobelovac Peter Higgs. Kvaliteta znanstvenih časopisa najčešće se kvantificira **faktorom odjeka** koji predstavlja omjer broja citata i broja radova objavljenih u tom časopisu u posljednje dvije godine. Faktor odjeka je ponekad kritiziran u znanstvenoj zajednici jer često ne reflektira istinsku kvalitetu časopisa i članaka objavljenih u njemu zbog postojanja upitnih uredničkih praksi koje pokušavaju maksimizirati faktor odjeka te brojnih drugih razloga [1].

Konstrukcija i analiza **grafa citata** je još jedan od načina na koji se mogu identificirati relevantni autori ili znanstveni radovi. Ovisno o cilju analize, grafove, tj. mreže citata moguće je konstruirati među autorima, člancima ili znanstvenim časopisima. Graf se kons-

truire tako da vrhovi predstavljaju autore ili članke, a usmjereni bridovi idu iz vrha koji je citirao prema citiranom vrhu. Za pronalazak relevantnih vrhova u tim grafovima mogu se iskoristiti mjere centralnosti. Jedna od mjera utjecaja koja koristi graf citata i mjere centralnosti za rangiranje znanstvenih časopisa je **SCImago Journal Rank (SJR)**. SJR časopise rangira uz pomoć PageRanka i upotrebljava se kao alternativa faktoru odjeka. Osim za rangiranje časopisa, SJR se koristi i za rangiranje institucija i država u različitim kategorijama [33].

U nastavku ćemo iskoristiti mjere centralnosti za identificiranje značajnih članaka iz skupa podataka CORD-19 te usporediti rezultate za različite mjere.

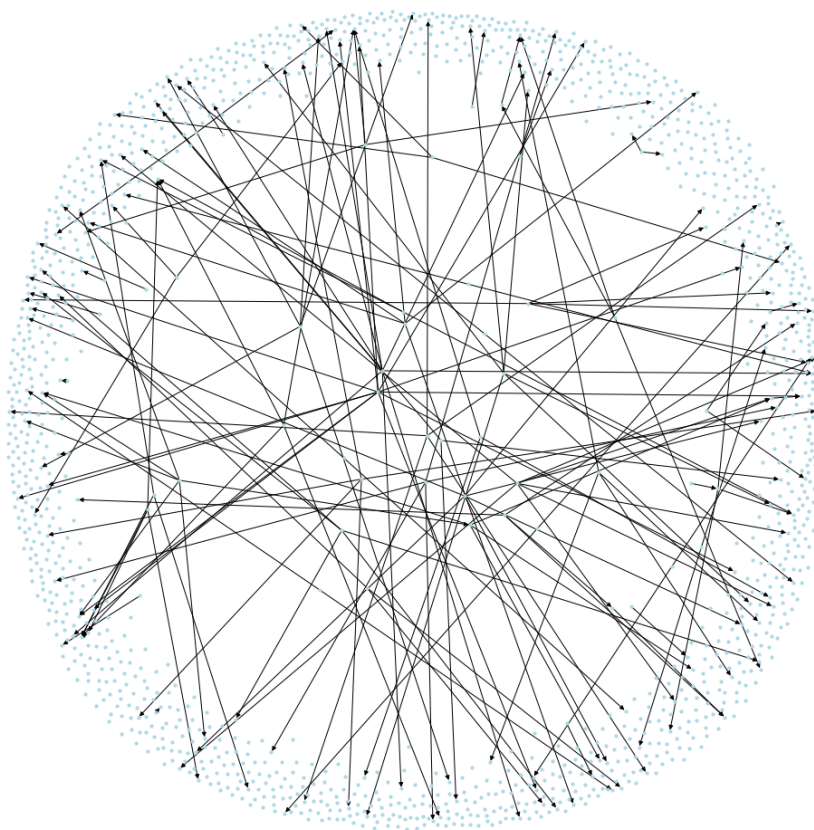
CORD-19 skup podataka

2020. godina je obilježena globalnom borbom protiv pandemije virusa SARS-CoV-2. Znanstvenici diljem svijeta su naporno radili na otkrivanju lijekova i cjepiva u cilju suzbijanja pandemije i kao rezultat toga rada objavljene su tisuće znanstvenih radova i istraživanja. CORD-19 [36] je otvorena baza podataka koja objedinjuje znanstvene članke i istraživanja povezana s virusom SARS-CoV-2. Inicijalni set podataka je sadržavao preko 28 tisuća znanstvenih članaka, a taj broj je narastao na preko 500 tisuća od čega preko 200 tisuća sa punim tekstom. U setu podataka se nalaze članci koji u naslovu, sažetku ili tekstu imaju neku od ključnih riječi navedenih u tablici 4.8.

Ključne riječi
COVID
COVID-19
Coronavirus
Corona virus
2019-nCoV
SARS-CoV
MERS-CoV
Severe Acute Respiratory Syndrome
Middle East Respiratory Syndrome

Tablica 4.8: Ključne riječi korištene za dohvaćanje relevantnih znanstvenih članaka.

Prepoznati uistinu relevantne podatke u tako velikom skupu nije jednostavno. Mi ćemo pokušati identificirati najvažnije znanstvene radove korištenjem mjera centralnosti u analizi grafa citata. Graf, odnosno mreža citata, preuzeta sa [2], konstruirana je tako da svaki vrh mreže predstavlja članak iz CORD-19 skupa podataka. Usmjereni bridovi idu od članka prema svim člancima koji su u njemu citirani. Mreža sadrži 1398451 vrh i 2554291 brid. Njen manji dio, samo 1500 vrhova i pripadnih bridova, prikazan je na slici 4.9.



Slika 4.9: Podgraf mreže citata koji sadrži 1500 nasumičnih vrhova i pripadne bridove.

Rezultati

Najjednostavnije je krenuti od centralnosti stupnja. U slučaju mreže citata, ulazni stupanj vrha je broj citata tog članka. Ovo je standardna mjera relevantnosti i koristan podatak za dobivanje početne slike o tome u kojem rasponu se kreću osnovne vrijednosti u ovom skupu podataka. Petnaest najcitiranijih članaka navedno je u tablici 4.12.

Najcitiraniji članak u ovom skupu podataka ima ulazni stupanj 1151, tj. citiran je 1151 put. Ta vrijednost se zamjetno smanjuje za članke kako se spuštamo prema dnu tablice, te posljednji članak u tablici ima ulazni stupanj 440. Ovako značajna razlika najveće i najmanje vrijednosti, za iznimno malen raspon rezultata koje promatramo, mogla bi upućivati na to da centralnost stupnja uistinu identificira relevantne članke. Međutim, ona će izdvojiti

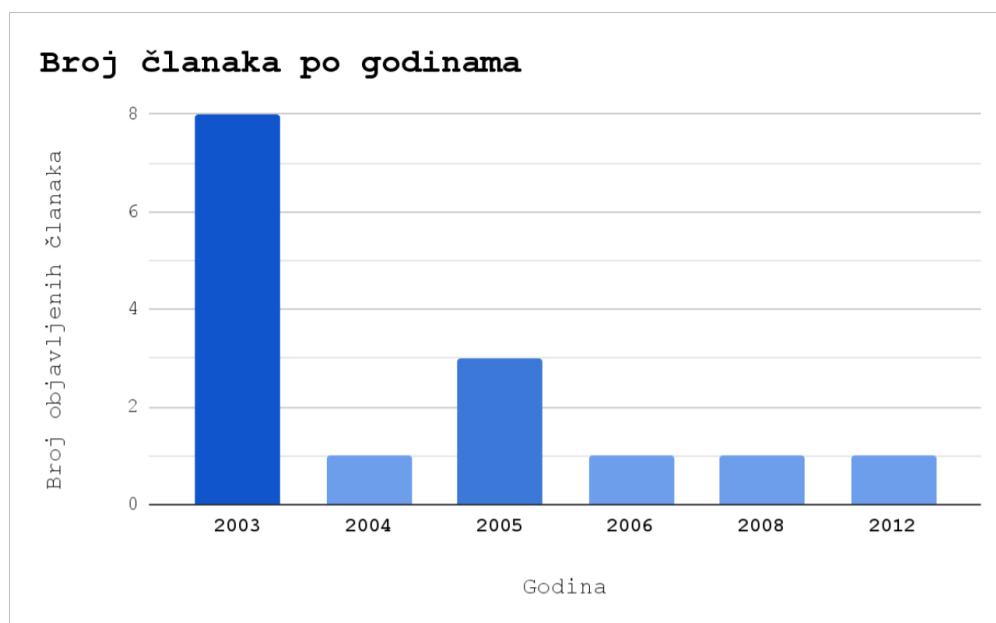
samo nekolicinu najznačajnijih članaka koji bitno odstupaju od ostatka. Za većinu članaka, centralnost stupnja nam neće otkriti mnogo o njihovoj relevantnosti ili poslužiti kao dobra mjera za usporedbu. Potvrdu ove slutnje možemo dobiti ako promotrimo vrijednosti za neke raspone niže rangiranih članaka. U tablici 4.10 nalaze se vrijednosti ulaznog stupnja za različite raspone rangova. Lagano je uočiti da centralnost stupnja ne uspijeva naglasiti razlike za većinu članaka i to postaje sve očitije za niže rangove. Na primjer, za 1000 članaka koji se nalaze u rasponu rezultata od 1001 do 2000, vrijednosti ulaznog stupnja su između 58 i 41. Ovakav rezultat sugerira da je broj citata vjerojatno dobar alat za identificiranje članaka koji postavljaju temelje novih ideja ili istraživanja u nekom području te su baza za daljnji razvoj i napredak. Međutim, iznimno veliki doprinosi znanosti ili revolucionarne ideje koje pokreću istraživanja u novim područjima su rijetki događaji, pa je očekivano da je najveći broj citata koncentriran na relativno mali skup članaka.

Raspon rangova	Raspon vrijednosti centralnosti stupnja
1001 – 2000	58 – 41
10 001 – 20 000	17 – 11
50 001 – 100 000	6 – 4
100 001 – 1 000 000	4 – 3

Tablica 4.10: Rasponi vrijednosti centralnosti stupnja.

Također je bitno spomenuti da broj citata, kao mjera relevantnosti, favorizira starije članke koji imaju prilike biti citirani kroz duži vremenski period. Ako promotrimo najbolje rangirane članke po ovom kriteriju možemo primijetiti da se u samom vrhu ne nalazi niti jedan članak koji je objavljen u razdoblju od početka pandemije virusa SARS-CoV-2. Za 15 najcitiranijih članaka raspodjelu po godini objavljivanja možemo vidjeti na slici 4.11. Razumna je pretpostavka da će se neki od članaka objavljenih 2020. i 2021. godine naći na vrhu liste zbog činjenice da je najviše istraživanja i novih otkrića upravo iz toga razdoblja. Međutim, od 15 najcitiranijih članaka u ovom skupu podataka, najveći broj je objavljen 2003. godine. Ovaj naizgled čudan rezultat postaje jasan kada se prisjetimo da je 2002. godine izbila epidemija virusa SARS-CoV-1 koja je trajala do 2004. godine. Kad rezultat interpretiramo u tom kontekstu, pretpostavka da je broj citata dobar alat za identificiranje članaka, koji su baza za istraživanja koja će uslijediti, pokazuje se opravdanom. U nastavku ćemo promotriti kakve rezultate će producirati svojstvena centralnost i PageRank na istom skupu podataka.

Deset članaka s najvećim vrijednostima svojstvene centralnosti nalazi se u tablici 4.13. U posljednjem stupcu tablice nalaze se vrijednosti izlaznog stupnja za odgovarajuće članke. Iako nisu u istom redoslijedu, možemo primijetiti da je većina članaka koja je rangirana



Slika 4.11: Raspodjela najcitiranijih članaka po godini objavljivanja.

visoko po vrijednosti svojstvene centralnosti također na vrhu i po centralnosti stupnja. Pojavile su se samo dvije iznimke, tj. dva članka koja se nisu našla između petnaest najboljih po centralnosti stupnja. Članci koji odstupaju po broju citata (ne pojavljuju se u tablici 4.12) su članak 8 koji ima ukupno 122 citata i članak 10 koji ima 331. Moglo bi se pretpostaviti da će članci možda odudarati po godini objavljivanja od skupine članaka koju smo ranije promatrali, ali to je samo djelomično točno. Članak 8 je objavljen 1987. godine, međutim članak 10 je objavljen 2003. godine, kao i većina članaka koji su rangirani visoko u prošlom slučaju. Zanimljivu poveznicu pronaći ćemo ako promotrimo čime se članci bave. Fokus oba članka je na sekvenciranju i svojstvima genoma virusa iz skupine koronavirusa. Podudaranje u tematici je zanimljiv rezultat koji implicira da bi problematika kojom se članci bave mogla biti značajna. Zaista nije teško povjerovati da su raniji rezultati sekvenciranja genoma virusa iz ove skupine bili referentna točka i temelj za brojna istraživanja koja su provedena u posljednje dvije godine.

Promotrimo sada tablicu 4.14 u kojoj je nalaze rezultati rangiranja PageRankom. Pojavila su se značajnija odstupanja od centralnosti stupnja nego je to bio slučaj kod svojstvene centralnosti. U PageRank rangiranju odstupaju čak četiri članka: 5, 6, 9 i 10. Zbog sličnosti svojstvene centralnosti i PageRanka razumljivo je pretpostaviti da će članci koje favoriziraju biti na neki način povezani, međutim, ispostavilo se da to nije slučaj. Ipak, postoji sličnost. Kao kod svojstvene centralnosti, favorizirani članci imaju međusobnu poveznicu

i obrađuju temu koja je bitna za razumijevanje pandemije bolesti COVID-19. Sva četiri članka koja odudaraju zbog relativno manjeg broja citata su objavljena na početku 2020. godine. Štoviše, sva četiri članka fokusirana su na istu temu, rane stadije pandemije i prve slučajeve zaraze koji su se pojavili. Očekivano je da će članci objavljeni na samom početku pandemije koji se bave ovom tematikom biti značajni za brojna istraživanja koja su uslijedila.

Možemo zaključiti da su mjere centralnosti koristan alat u analizi mreže citata i dobar indikator važnosti nekog članka. Ovisno o odabranoj mjeri poredak poprilično varira stoga je u analizi poželjno upotrijebiti više različitih mjera i usporediti dobivene rezultate. Rezultati koje smo dobili su zanimljivi i mogu se smisljeno interpretirati kao što je pokazano u ovom odjeljku, no valja naglasiti kako kompletna interpretacija iziskuje dulje vrijeme proučavanja i sveobuhvatno istraživanje problematike obrađene u radovima.

Rang	Naslov	Normalizirani stupanj ×10000	Stupanj
1	isolation of a novel coronavirus from a man with pneumonia in saudi arabia (2012)	8.231	1151
2	identification of a novel coronavirus in patients with severe acute respiratory syndrome (2003)	7.723	1080
3	a novel coronavirus associated with severe acute respiratory syndrome (2003)	7.58	1060
4	coronavirus as a possible cause of severe acute respiratory syndrome (2003)	5.842	817
5	characterization of a novel coronavirus associated with severe acute respiratory syndrome (2003)	5.37	751
6	bats are natural reservoirs of sars-like coronaviruses (2005)	4.362	610
7	angiotensin-converting enzyme 2 is a functional receptor for the sars coronavirus (2003)	4.319	604
8	isolation and characterization of viruses related to the sars coronavirus from animals in southern china (2003)	4.198	587
9	global trends in emerging infectious diseases (2008)	3.711	519
10	a major outbreak of severe acute respiratory syndrome in hong kong (2003)	3.704	518
11	the molecular biology of coronaviruses (2006)	3.64	509
12	the genome sequence of the sars-associated coronavirus (2003)	3.633	508
13	identification of a new human coronavirus (2004)	3.368	471
14	severe acute respiratory syndrome coronavirus-like virus in chinese horseshoe bats (2005)	3.254	455
15	characterization and complete genome sequence of a novel coronavirus, coronavirus hku1, from patients with pneumonia (2005)	3.146	440

Tablica 4.12: Prvih petnaest članaka sa najvećim vrijednostima centralnosti stupnja.

Rang	Naslov	Svojstvena centralnost $\times 100$	Stupanj
1	a novel coronavirus associated with severe acute respiratory syndrome	16.127	1060
2	identification of a novel coronavirus in patients with severe acute respiratory syndrome	15.065	1080
3	characterization of a novel coronavirus associated with severe acute respiratory syndrome	14.4	751
4	the molecular biology of coronaviruses	13.738	509
5	coronavirus as a possible cause of severe acute respiratory syndrome	12.81	817
6	the genome sequence of the sars-associated coronavirus	11.463	508
7	isolation and characterization of viruses related to the sars coronavirus from animals in southern china	9.651	587
8	completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus	9.641	122
9	angiotensin-converting enzyme 2 is a functional receptor for the sars coronavirus	8.901	604
10	unique and conserved features of genome and proteome of sars-coronavirus, an early split-off from the coronavirus group 2 lineage	7.588	331

Tablica 4.13: Prvih deset članaka sa najvećim vrijednostima svojstvene centralnosti.

Rang	Naslov	PageRank ×100000	Stupanj
1	isolation of a novel coronavirus from a man with pneumonia in saudi arabia	2.299	1151
2	identification of a novel coronavirus in patients with severe acute respiratory syndrome	2.268	1080
3	a novel coronavirus associated with severe acute respiratory syndrome	2.202	1060
4	coronavirus as a possible cause of severe acute respiratory syndrome	1.777	817
5	early transmission dynamics in wuhan, china, of novel coronavirus-infected pneumonia	1.574	390
6	a novel coronavirus from patients with pneumonia in china	1.516	351
7	characterization of a novel coronavirus associated with severe acute respiratory syndrome	1.469	751
8	a major outbreak of severe acute respiratory syndrome in hong kong	1.361	518
9	clinical features of patients infected with 2019 novel coronavirus in wuhan	1.157	284
10	epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in wuhan, china: a descriptive study	1.145	322

Tablica 4.14: Prvih deset članaka sa najvećim vrijednostima PageRanka.

Bibliografija

- [1] *Time to remodel the journal impact factor*, Nature **535** (2016), br. 446, <https://doi.org/10.1038/535466a>.
- [2] *Simple citation network and Pagerank score*, <https://www.kaggle.com/baptistemetge/simple-citation-network-and-pagerank-score?select=pagerank.csv>, posjećena 07-08-2021.
- [3] M. Aigner i G. M. Ziegler, *Proofs from THE BOOK*, 4th., Springer Publishing Company, Incorporated, 2009, ISBN 3642008550.
- [4] H. Anton i C. Rorres, *Elementary Linear Algebra: Applications Version*, 11., Wiley, 2014.
- [5] K. Appel i W. Haken, *Every planar map is four colorable. Part I: Discharging*, Illinois Journal of Mathematics **21** (1977), br. 3, 429 – 490, <https://doi.org/10.1215/ijm/1256049011>.
- [6] K. R. Apt, *Edsger Wybe Dijkstra (1930 - 2002): A Portrait of a Genius*, CoRR **cs.GL/0210001** (2002), <https://arxiv.org/abs/cs/0210001>.
- [7] P. Bonacich, *Factoring and weighting approaches to status scores and clique identification*, The Journal of Mathematical Sociology **2** (1972), br. 1, 113–120, <https://doi.org/10.1080/0022250X.1972.9989806>.
- [8] J.A. Bondy i U.S.R. Murty, *Graph Theory*, 1st., Springer Publishing Company, Incorporated, 2008, ISBN 1846289696.
- [9] S. P. Borgatti, *Centrality and AIDS*, Connections (1995), 112–114.
- [10] ———, *Centrality and network flow*, Social Networks **27** (2005), br. 1, 55–71, ISSN 0378-8733, <https://www.sciencedirect.com/science/article/pii/S0378873304000693>.

- [11] S. P. Borgatti i M. G. Everett, *A Graph-theoretic perspective on centrality*, Social Networks **28** (2006), br. 4, 466–484, <https://doi.org/10.1016/j.socnet.2005.11.005>.
- [12] U. Brandes, S. P. Borgatti i L. C. Freeman, *Maintaining the duality of closeness and betweenness centrality*, Social Networks **44** (2016), 153–159, <https://doi.org/10.1016/j.socnet.2015.08.003>.
- [13] S. Brin i L. Page, *The anatomy of a large-scale hypertextual Web search engine*, Computer Networks and ISDN Systems **30** (1998), br. 1, 107–117, [https://doi.org/10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X).
- [14] P. Chen, H. Xie, S. Maslov i S. Redner, *Finding scientific gems with Google's PageRank algorithm*, Journal of Informetrics **1** (2007), br. 1, 8–15, ISSN 1751-1577, <https://www.sciencedirect.com/science/article/pii/S1751157706000034>.
- [15] E. W. Dijkstra, *A note on two problems in connexion with graphs*, Numerische Mathematik **1** (1959), 269–271.
- [16] L. C. Freeman, *Centrality in social networks conceptual clarification*, Social Networks (1979), br. 3, 215–239, [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7).
- [17] D. F. Gleich, *PageRank Beyond the Web*, SIAM Review **57** (2015), br. 3, 321–363, <https://doi.org/10.1137/140976649>.
- [18] A. Y. Govan i C. D. Meyer, *Ranking National Football League teams using Google's PageRank*, <https://projects.ncsu.edu/crsc//reports/ftp/pdf/crsc-tr06-19.pdf>.
- [19] L. Havaš, D. Keček i K. Knez, *Usporedba i primjena "Distance Vector" i "Link State" mrežnih protokola*, Tehnički glasnik (Varaždin) **7** (2013), 108–115, <https://hrcak.srce.hr/105589>.
- [20] M. J. Keeling i K. T. D. Eames, *Networks and epidemic models*, Journal of The Royal Society Interface **2** (2005), br. 4, 295–307, <https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2005.0051>.
- [21] D. Koschützki, K. A. Lehmann, L. Peeters, S. Richter, D. Tenfelde-Podehl i O. Zlotowski, *Centrality Indices*, Springer, Berlin, Heidelberg, 2005, https://doi.org/10.1007/978-3-540-31955-9_3.

- [22] D. Krackhardt, *Assessing the Political Landscape: Structure, Cognition, and Power in Organizations*, *Administrative Science Quarterly* **35** (1990), br. 2, 342–369, <http://www.jstor.org/stable/2393394>.
- [23] F. T. Leighton, *A graph coloring algorithm for large scheduling problems*, *Journal of research of the national bureau of standards* **84** (1979), br. 6, 489–506.
- [24] R. M. R. Lewis, *A Guide to Graph Colouring: Algorithms and Applications*, 1st., Springer Publishing Company, Incorporated, 2015, ISBN 3319257285.
- [25] F. Li i R. Klette, *Euclidean Shortest Paths: Exact or Approximate Algorithms*, Springer-Verlag London, 2011, ISBN 1447122550,9781447122555, <http://gen.lib.rus.ec/book/index.php?md5=8BB2951A29C1A0478872F2209F7623A1>.
- [26] H. Mehta, P. Kanani i P. Lande, *Google Maps*, *International Journal of Computer Applications* **178** (2019), 41–46.
- [27] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, Society for Industrial and Applied Mathematics, 2000, ISBN 0898714540.
- [28] D. Mundar i D. Horvat, *Rangiranje ekipa i prognoziranje ishoda u rukometu korištenjem PageRank algoritma*, *Poučak* **17** (2016), br. 67, 35–42, <https://hrcak.srce.hr/180919>.
- [29] ———, *Rangiranje web stranica*, *Osječki matematički list* **5** (2017), br. 1, 51–62, <https://hrcak.srce.hr/186508>.
- [30] I. Nakić, *Diskretna matematika*, Sveučilište u Zagrebu, Prirodoslovno matematički fakultet, Matematički odsjek, 2011./2012.
- [31] J. Ott, D. Tan, T. Loveless, W. H. Grover i P. Brisk, *ChemStor: Using Formal Methods To Guarantee Safe Storage and Disposal of Chemicals*, *Journal of Chemical Information and Modeling* **60** (2020), br. 7, 3416–3422, <https://doi.org/10.1021/acs.jcim.9b00951>.
- [32] G. Sabidussi, *The centrality index of a graph*, *Psychometrika* **31** (1966), 581–603, <https://doi.org/10.1007/BF02289527>.
- [33] (n.d.). SJR SCImago, *SCImago Journal Country Rank [Portal]*., 2021, <http://www.scimagojr.com>, posjećena 25-08-2021.
- [34] S. Y. Del Valle, J. M. Hyman i N. Chitnis, *Mathematical models of contact patterns between age groups for predicting the spread of infectious diseases*, *Mathematical Biosciences Engineering* **10** (2013), br. 56, 1475–1497.

- [35] M. Vuković, *Složenost algoritama*, PMF-Matematički odsjek, 2011.
- [36] L. L. Wang, K. Lo, Y. Chandrasekhar, R. Reas, J. Yang, D. Burdick, D. Eide, K. Funk, Y. Katsis, R. M. Kinney, Y. Li, Z. Liu, W. Merrill, P. Mooney, D. A. Murdick, D. Rishi, J. Sheehan, Z. Shen, B. Stilson, A. D. Wade, K. Wang, N. X. R. Wang, C. Wilhelm, B. Xie, D. M. Raymond, D. S. Weld, O. Etzioni i S. Kohlmeier, *CORD-19: The COVID-19 Open Research Dataset*, Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020, Association for Computational Linguistics, 2020, <https://aclanthology.org/2020.nlpcovid19-acl.1>.
- [37] D. Šantl, *Bojanje grafova prilagodljivim metaheurističkim postupcima (Diplomski rad)*, (2014), <https://urn.nsk.hr/urn:nbn:hr:168:470371>.

Sažetak

U ovom radu obrađene su tri primjene teorije grafova. Nakon pregleda osnovnih koncepata teorije grafova slijedi kratka analiza problema najkraćeg puta i primjer rješenja Dijkstrinim algoritmom te se navode neke primjene. U nastavku je detaljnije obrađen problem bojanja grafova kojem je cilj minimizirati broj boja potrebnih da susjedni vrhovi grafa budu obojani različitim bojama. Dan je pregled tri primjene bojanja grafova s naglaskom na primjeru iz telekomunikacija. Posljednje i najopširnije poglavlje obrađuje mjere centralnosti čija je svrha odrediti važnost nekog vrha u grafu. Opisuje se nekoliko različitih primjena mjera centralnosti, uključujući analizu mreže citata akademskih članaka koji obrađuju temu pandemije bolesti COVID-19.

Summary

This thesis is a study of three topics on the application of graph theory. After a review of the central concepts of graph theory, we first give a brief analysis of the shortest path problem and an example of its solution via Dijkstra's algorithm, together with some applications of this approach. We then proceed to give a more detailed analysis of the problem of graph colouring, which aims to minimise the number of colours needed to ensure that adjacent graph vertices are always coloured differently. Here we review three applications of graph colouring, focusing on an application in telecommunications. The main and final chapter is a discussion of graph centrality measures, which seeks to measure the "importance" of a vertex in a graph. We describe several different applications of centrality measures, including an analysis of citation networks for academic articles on the topic of the COVID-19 pandemic.

Životopis

Rođena sam 1.6.1991. godine. 2006. godine upisala sam Matematičku gimnaziju u Dubrovniku. 2013. godine upisala sam studij primjenjenog računarstva na Sveučilištu u Dubrovniku, na kojem sam 2015. stekla prvostupničku diplomu, te sam iste godine upisala diplomski studij na Prirodoslovno-matematičkom fakultetu Sveučilišta u Zagrebu, smjer Računarstvo i matematika.