

Grafovi i Covid-19

Rabar, Karmen

Master's thesis / Diplomski rad

2023

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Rijeka / Sveučilište u Rijeci**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:196:484919>

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

Download date / Datum preuzimanja: **2024-09-08**



Repository / Repozitorij:

[Repository of the University of Rijeka, Faculty of Mathematics - MATHRI Repository](#)



Sveučilište u Rijeci
Fakultet za matematiku

Diplomski studij Diskretna matematika i primjene

Karmen Rabar

Grafovi i Covid-19

Diplomski rad

Rijeka, 2023.

Sveučilište u Rijeci
Fakultet za matematiku

Diplomski studij Diskretna matematika i primjene

Karmen Rabar

Grafovi i Covid-19

Mentor: izv. prof. dr. sc. Andrea Švob

Diplomski rad

Rijeka, 2023.

Sadržaj

Uvod	1
1 Osnovni pojmovi	2
2 Tolerancijski grafovi	6
2.1 Povijest nastajanja i motivacija	6
2.2 Grafovi presjeka i intervalni grafovi	9
2.3 Tolerancijski grafovi	11
2.4 Savršeni grafovi	15
3 Primjena tolerancijskih grafova	15
3.1 Organizacija letova	16
3.2 Korištenje učionica pri nastavi uživo	20
3.3 Posjeti u muzej	20
3.4 Traženje kritičnih događaja	21
4 Algoritmi	22
4.1 Algoritam za organizaciju letova	22
4.2 Algoritam za korištenje učionica u nastavi uživo	23
4.3 Algoritam za raspored posjeta u muzej	23
4.4 Algoritam za traženje kritičnih događaja	24
5 Grafovi sličnosti	24
5.1 Uvod u grafove sličnosti	24
5.2 Grafovi sličnosti	25
5.3 Vrste grafova sličnosti	26
6 Spektralno grupiranje	27
6.1 Laplaceove matrice	27
6.2 Algoritam k-srednjih vrijednosti	29
6.3 Vrste spektralnog grupiranja	30

7 Primjena temeljena na grafovima sličnosti	32
7.1 Obrada podataka	32
7.2 Konstrukcija grafa sličnosti	33
7.3 Spektralno grupiranje	34
Zaključak	35
Literatura	36
Popis slika	37
Popis tablica	37

Sažetak

U ovom je radu predstavljena primjena teorije grafova u borbi protiv pandemije COVID-19. Opisani su tolerancijski grafovi, koje su 1982. godine predstavili Golubic i Monma kao generalizaciju intervalnih grafova. Navedene primjene uključuju pronalaženje klika određene veličine i izračunavanje kromatskog broja grafa. Također, uvedeni su grafovi sličnosti koji igraju glavnu ulogu u metodologiji za praćenje širenja zaraze koja se sastoji od prikupljanja i obrade podataka, modeliranja pomoću grafova sličnosti, spektralnog grupiranja i vizualizacije podataka.

Ključne riječi

COVID-19, pandemija, graf, klika, kromatski broj, bojenje grafa, graf presjeka, intervalni graf, tolerancijski graf, savršen graf, graf sličnosti, spektralno grupiranje.

Uvod

Od svog prvog pojavljivanja u prosincu 2019. godine, pandemija COVID-19 je postala široko polje znanstvenog istraživanja. Neka područja u matematici, poput umjetne inteligencije služe kao moćan alat za praćenje i predviđanje izbijanja virusa koristeći različite vrste modela. Također, teorija grafova ima važnu ulogu u rješavanju problema optimizacije i rasporeda koji mogu biti od velike pomoći. Grafovi se mogu koristiti za modeliranje određenih odnosa između objekata te za modeliranje različitih problema povezanih s pandemijom COVID-19. Oni su također jedan od osnovnih dijelova umjetne inteligencije. Na primjer, mogu služiti za praćenje kontakata zaraženih ljudi, provođenje medicinskih analiza za razumijevanje dinamike virusa i bolesti i predviđanje dinamike pandemije općenito. Cilj je ovog rada istaknuti važnost algoritama i metoda temeljenih na grafovima, točnije tolerancijskim grafovima i grafovima sličnosti, u praćenju širenja virusa, modeliranju različitih situacija koje se javljaju tijekom pandemije i u borbi protiv pandemije te usporavanju procesa širenja virusa.

1 Osnovni pojmovi

Definicija 1.1 **Graf** G je uređena trojka $G = (V(G), E(G), \psi(G))$, gdje je $\emptyset \neq V = V(G)$ skup vrhova, $E = E(G)$ (skup disjunktan s $V(G)$) skup bridova od G , a $\psi(G)$ funkcija incidencije koja svakom bridu od G pridružuje neuređeni par (ne nužno različitih) vrhova od G .

Definicija 1.2 Funkcija incidencije $\psi(G) : E(G) \rightarrow \left(\binom{V(G)}{2}\right)$ svakom bridu $e \in E(G)$ pridružuje 2-člani multiskup vrhova $\psi(G)(e) = \{u, v\}$, gdje su $u, v \in V(G)$. Kažemo da brid e **spaja** vrhove u i v , koje zovemo **krajevi** brida e , pišemo $\psi(G)(e) = uv$. Kažemo da su vrhovi u i v **susjedni** vrhovi, **incidentni** s e , u oznaci $e = \{u, v\}$ ili $e = uv$.

Definicija 1.3 Brid koji je incidentan samo s jednim vrhom zove se **petlja**. Dva ili više bridova s istim parom krajeva zovu se **višestruki bridovi**.

Definicija 1.4 Neka je G graf sa n vrhova, $V(G) = \{v_1, \dots, v_n\}$. **Matrica susjedstva** grafa G je $n \times n$ matrica $A = [a_{ij}]$ gdje je $a_{ij} \in \mathbb{N}_0$ broj bridova koji spajaju vrhove v_i i v_j , $i, j \in \{1, \dots, n\}$.

Napomena 1.1 Matrica susjedstva je simetrična matrica.

Definicija 1.5 **Stupanj ili valencija** vrha v u grafu G je broj bridova od G incidentnih sa v pri čemu svaku petlju brojimo dvaput.

Definicija 1.6 **Matrica stupnjeva** grafa G sa skupom vrhova $V(G) = \{v_1, \dots, v_n\}$ je dijagonalna matrica $D = [d_{ij}]$ $i, j \in \{1, \dots, n\}$ u kojoj svaki element dijagonale označava stupanj odgovarajućeg vrha. Odnosno,

$$d_{ij} = \begin{cases} \text{stupanj vrha } v_i, & \text{ako } i = j \\ 0, & \text{inače.} \end{cases}$$

Definicija 1.7 Graf je **jednostavan** ako ne sadrži petlju i višestruke bridove.

Definicija 1.8 Jednostavan graf G je **potpun** ako su svaka dva različita vrha u G susjedna.

Definicija 1.9 Neka je $G = (V(G), E(G), \psi(G))$ graf. **Težinski graf** je uređeni par (G, ω) gdje je $w : E(G) \rightarrow \mathbb{R}_0^+$ funkcija koju nazivamo težinska funkcija te ona svakom bridu od G pridružuje nenegativan realan broj.

Definicija 1.10 Neka je $G = (V(G), E(G), \psi(G))$ težinski graf sa skupom vrhova $V = \{v_1, \dots, v_n\}$ te neka svaki brid između dva vrha v_i i v_j nosi nenegativnu težinu $w_{ij} \geq 0$. **Težinska matrica susjedstva** grafa G je matrica $W = [w_{ij}]$, $i, j = 1, \dots, n$. Ako je $w_{ij} = 0$ znači da vrhovi v_i i v_j nisu povezani bridom.

Definicija 1.11 **Stupanj vrha** $v_i \in V(G) = \{v_1, \dots, v_n\}$ **težinskog grafa** G definiran je kao

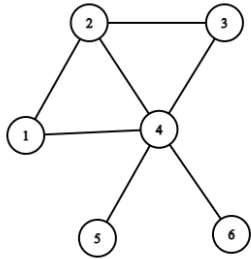
$$d_i = \sum_{j=1}^n w_{ij}.$$

Matrica stupnjeva D je definirana kao dijagonalna matrica sa stupnjevima d_1, \dots, d_n na glavnoj dijagonali.

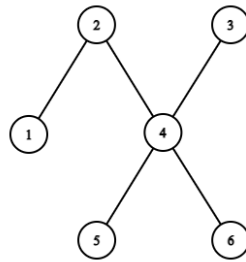
Definicija 1.12 Neka su G i H grafovi. Graf H je **podgraf** grafa G , u oznaci $H \subseteq G$, ako je $V(H) \subseteq V(G)$, $E(H) \subseteq E(G)$, a svaki brid iz H ima iste krajeve u H kao što ih ima u G . Ako je $H \subseteq G$ i $H \neq G$, pišemo $H \subset G$ i H zovemo **pravi podgraf** od G .

Definicija 1.13 Podgraf H grafa G za koji vrijedi $V(G) = V(H)$ je **razapinjući podgraf** od G .

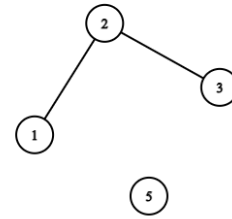
Definicija 1.14 Podgraf od G čiji je skup vrhova V' , $\emptyset \neq V' \subseteq V(G)$, a skup bridova je podskup od $E(G)$ kojeg čine bridovi s oba kraja u V' zove se **podgraf induciran** s V' i označava se s $G[V']$.



Slika 1: Graf G_1



Slika 2: Razapinjući podgraf od G_1



Slika 3: Inducirani podgraf od G_1

Definicija 1.15 *Klika* K u grafu G je potpuni podgraf grafa G . **Veličina klike** je broj vrhova koje klika sadrži.

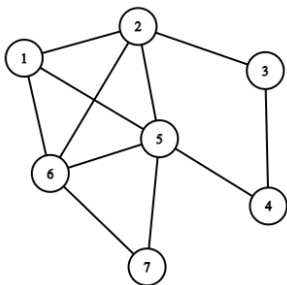
Dakle, klika K u grafu G je podskup skupa vrhova V od G takav da su svaka dva različita vrha povezana bridom.

Definicija 1.16 Za kliku K kažemo da je **neproširiva** ako joj se ne može dodati još jedan susjedan vrh.

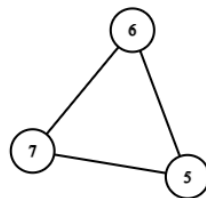
Drugim riječima, neproširiva klika je klika koja nije sadržana u niti jednoj većoj kliku.

Definicija 1.17 Za kliku K kažemo da je **najveća** u grafu G ako ne postoji klika s više vrhova.

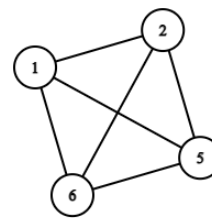
Napomena 1.2 Označimo s $\omega(G)$ broj vrhova najveće klike u grafu G .



Slika 4: Graf G_2



Slika 5: Neproširiva klika



Slika 6: Najveća klika

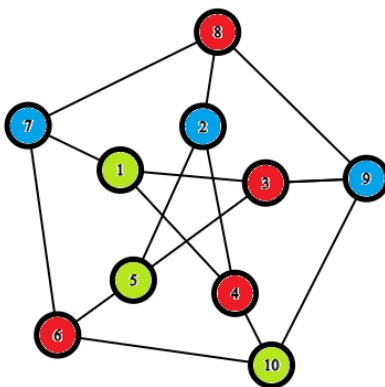
Definicija 1.18 Za zadani prirodni broj k , **k -bojenje** vrhova grafa G je preslikavanje $c : V(G) \rightarrow \{1, 2, \dots, k\}$, koje svakom vrhu iz G pridružuje jednu od k boja. Ako je $c(v) = i$, kažemo da je vrh v obojen bojom i .

Definicija 1.19 **Pravilno k -bojenje vrhova** grafa G , $k \in \mathbb{N}$, je preslikavanje $c : V(G) \rightarrow \{1, 2, \dots, k\}$, takvo da $c(v) \neq c(w)$ kad god su v i w povezani bridom.

Napomena 1.3 Samo grafovi bez petlji dopuštaju pravilno bojenje vrhova.

Definicija 1.20 Kažemo da je graf G **k -obojuv** ako dopušta pravilno k -bojenje vrhova.

Definicija 1.21 Najmanji broj k , takav da je graf G k -obojuv, nazivamo **kromatski broj** i pišemo $\chi(G) = k$. Ako za G vrijedi $\chi(G) = k$, kažemo da je G **k -kromatski**.



Slika 7: Pravilno 3-bojenje Petersenovog grafa

Definicija 1.22 Neka je A kvadratna matrica reda n . Kažemo da je vektor \vec{x} , $\vec{x} \neq \vec{0}$, $\vec{x} \in \mathbb{R}^n$ **svojstveni vektor** matrice A ako vrijedi

$$A \cdot \vec{x} = \lambda \cdot \vec{x}$$

za neki $\lambda \in \mathbb{R}$. Pritom se skalar λ naziva **svojstvena vrijednost**. Skup svih svojstvenih vrijednosti naziva se **spektar**.

Primijetimo da se izraz $A\vec{x} = \lambda\vec{x}$ može zapisati u obliku $(A - \lambda I)\vec{x} = 0$, gdje je I jedinična matrica reda n , iz čega proizlazi da vrijedi $\det(A - \lambda I) = 0$. Nadalje, $\det(A - \lambda I)$ je polinom po varijabli λ kojeg nazivamo **karakterističnim polinomom** i pišemo

$$p(\lambda) = \det(A - \lambda I).$$

Korijeni karakterističnog polinoma, ili drugim riječima rješenja jednadžbe $\det(A - \lambda I) = 0$, su svojstvene vrijednosti koje odgovaraju pripadnom svojstvenom vektoru.

Definicija 1.23 *Slučajni proces* je familija slučajnih varijabli koje su definirane na nekom vjerojatnosnom prostoru:

$$\{X_t : t \in T\}.$$

Skup indeksa T možemo shvatiti kao vremensku komponentu podataka te onda govorimo o **vremenskom nizu**. Obično promatramo $T = [0, t]$, $T = [0, +\infty)$ ili $T = \{t_i : i \in I\}$. Pri mjerenju se opaža realizacija dijela vremenskog niza

$$\{X_t(\omega) : t \in T_0\} = \{x_t : t \in T_0\},$$

gdje je $T_0 \subseteq T$ konačan.

2 Tolerancijski grafovi

2.1 Povijest nastajanja i motivacija

Na 13. Jugoistočnoj konferenciji o kombinatorici, teoriji grafova i računarstvu (13th Southeastern Conference on Combinatorics, Graph Theory and Computing) u gradu Boca Raton 1982. godine, Golumbic i Monma su predstavili matematički model tolerancije, nazvan tolerancijski grafovi. Motivacija za uvođenje tolerancijskih grafova bila je rješavanje problema rasporeda u kojima resursi (kao što su primjerice prostorije, vozila, pomoćno osoblje i slično)

mogu biti ograničeni, ali postoji dopuštena mjera fleksibilnosti ili tolerancije za dijeljenje tih resursa, u slučaju kada je zbog ograničenja nemoguće doći do rješenja.

Tijekom narednih godina proučavana su svojstva tolerancijskih grafova, a u literaturi se pojavio veliki broj varijacija, uključujući biotolerancijske grafove, ϕ -tolerancijske grafove, NeST grafove, tolerancijske digrafove i druge. Ovo je i dalje zanimljivo i aktivno područje istraživanja s brojnim primjerima.

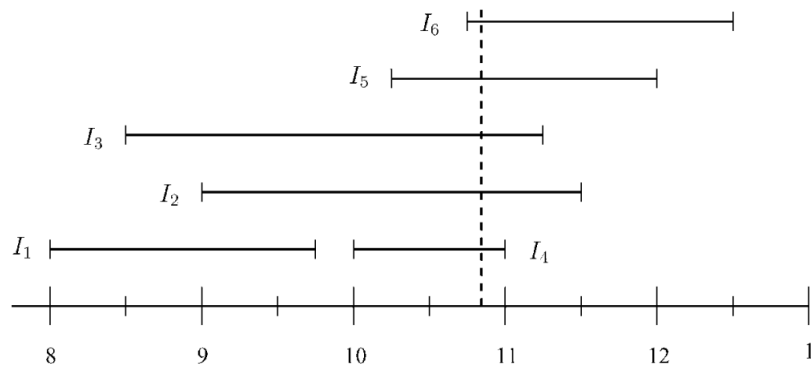
Kod tolerancijskih grafova, promatra se kolekcija intervala na realnoj osi. Intervali mogu predstavljati trajanje skupa događaja ili pak fragmente DNK na genomu. Neki od intervala mogu se međusobno sjeći, a drugi mogu biti disjunktni. Postoje mnoge veze između intervala koje bismo mogli proučavati, no u ovom će radu naglasak biti na presjecima tih intervala koje možemo protumačiti na više načina. Može značiti da imaju nešto zajedničko, poput prilike za razmjenu informacija. Na primjer, ako je svaki interval predstavljao vremenski period tijekom kojeg bi skupina školske djece posjetila muzej, tada bi dvije grupe čiji se intervali sijeku mogle sudjelovati u zajedničkoj aktivnosti. S druge strane, presjek intervala može značiti sukob, kao što je natjecanje za resurs koji se ne može dijeliti. Primjerice, u kućanstvu s jednom televizijom, kada roditelj želi gledati vijesti, a u isto vrijeme tinejdžer želi gledati film na drugom kanalu, dolazi do vremenskog sukoba.

U ovom će radu presjeci intervala biti povezani s virusom COVID-19. Odnosno, predstavljat će vremenske intervale u kojima može postojati opasnost od potencijalne zaraze. Promatrat ćemo ih u cilju pronalaženja rasporeda određenih aktivnosti kako bi se daljnje širenje i mogućnost zaraze spriječili.

Primjer 2.1 *Ponedjeljkom ujutro se šest sastanaka treba održati prema fiksnom rasporedu, pri čemu je održavanje sastanka s_i predviđeno za vremenski interval $I_i = [a_i, b_i]$. Svakom sastanku mora biti dodijeljena prostorija.*

$$I_1 = [8 : 00 - 9 : 45], I_2 = [9 : 00 - 11 : 30], I_3 = [8 : 30 - 11 : 15],$$

$$I_4 = [10 : 00 - 11 : 00], I_5 = [10 : 15 - 12 : 00], I_6 = [10 : 45 - 12 : 30].$$



Slika 8: Vremenski intervali trajanja sastanaka (preuzeto iz [2])

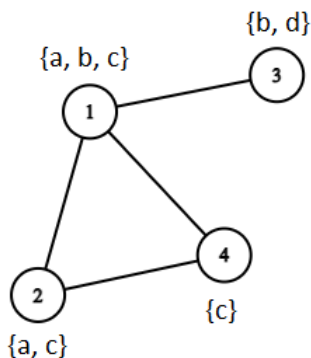
Vidimo da bi sastanak s_1 mogao biti u istoj prostoriji kao s_4 ili s_5 ili s_6 jer se njegov pripadni vremenski interval I_1 ne preklapa s vremenskim intervalima I_4 , I_5 ili I_6 . Budući da vrlo strogo promatramo intervale, vidimo da je u 10:50 potrebno pet prostorija istovremeno. No, pretpostavimo sada da postoje samo četiri prostorije za sastanke. Umjesto da se otkáže neki od sastanaka, pokušajmo naći određenu fleksibilnost u vremenskim intervalima koja bi mogla omogućiti prihvatljivu dodjelu prostorija. Pomoću tolerancijskih grafova, kojeg ćemo formalno definirati u nastavku, možemo pridružiti toleranciju svakom sastanku. Tolerancija indicira stupanj fleksibilnosti u dopuštanju veličine presjeka s drugim intervalima te bi nam mogla pomoći u pronalasku dodjele prostorija. Primjerice, kada bi sastanci s_4 i s_6 dopuštali preklapanje dulje od 15 minuta, bilo bi moguće održati sve sastanke u četiri prostorije.

Problemi s dodjelom resursa ove prirode javljaju se u mnogim kontekstima: dostava hrane ili pošte, korištenje pećnice u restoranima, raspored medicinskih sestara u operacijskim sobama i slično. U stvarnoj situaciji, neki sastanci ili dostave mogu doista imati stroge rokove koji se moraju poštovati, dok drugi mogu biti fleksibilniji. Uzimajući u obzir pripadne tolerancije, često se mogu pronaći rješenja koja inače ne bi postojala pod strogim ograničenjima.

2.2 Grafovi presjeka i intervalni grafovi

Definicija 2.1 Neka je \mathcal{F} kolekcija skupova. **Graf presjeka** je graf dobiven dodjeljivanjem različitog vrha svakom skupu u \mathcal{F} i spajanjem dvaju vrhova bridom onda kada njihovi odgovarajući skupovi imaju neprazan presjek.

Sljedeća slika prikazuje primjer grafa presjeka G , gdje je skup vrhova jednak $V(G) = \{1, 2, 3, 4\}$, a kolekcija skupova je $\mathcal{F} = \{\{a, b, c\}, \{a, c\}, \{b, d\}, \{c\}\}$. Vrh 1 je pridružen skup $\{a, b, c\}$, vrhu 2 skup $\{a, c\}$, vrhu 3 skup $\{b, d\}$, a vrhu 4 skup $\{c\}$. Vidimo da je presjek skupova pridruženih vrhovima 1 i 2 jednak $\{a, c\}$, odnosno, neprazan je, $I_1 \cap I_2 = \{a, b, c\} \cap \{a, c\} = \{a, c\} \neq \emptyset$. Stoga vrhove 1 i 2 povežemo bridom. Nadalje, skupovi pridruženi vrhovima 3 i 4 su disjunktni iz čega slijedi da oni neće biti susjedni, $I_3 \cap I_4 = \{b, d\} \cap \{c\} = \emptyset$. Slično se dobiva da će vrhovi 1 i 3, 1 i 4, 2 i 4 biti susjedni, a ostali parovi vrhova neće.

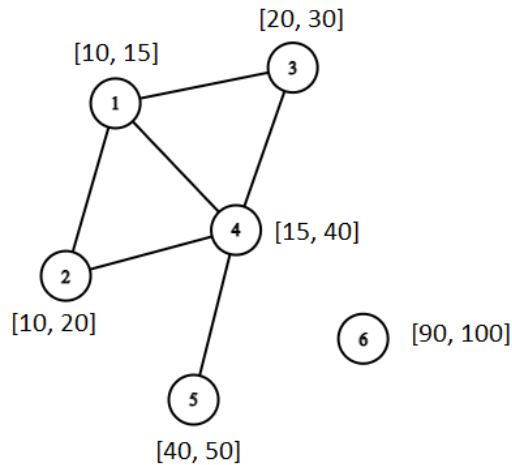


Slika 9: Graf presjeka

Možemo promatrati razne vrste skupova u \mathcal{F} te tako dobiti zanimljive klase grafova. Najvažniji će nam biti intervalni grafovi koji nastaju kada su skupovi u \mathcal{F} intervali na realnoj osi koje želimo proučavati.

Definicija 2.2 Graf $G = (V(G), E(G), \psi(G))$ je **intervalni graf** ako se svakom vrhu $v \in V$ može dodijeliti realan interval I_v tako da $xy \in E \iff I_x \cap I_y \neq \emptyset$, za sve $x, y \in V$. Skup intervala $\{I_v \mid v \in V\}$ predstavlja reprezentaciju intervalnog grafa G .

Na sljedećoj slici prikazan je primjer intervalnog grafa G čiji je skup vrhova $V(G) = \{1, 2, 3, 4, 5, 6\}$. Reprezentacija intervalnog grafa je jednaka $\{[10, 15], [10, 20], [20, 30], [15, 40], [40, 50], [90, 100]\}$. Odnosno, vrhu 1 pridružen je interval $[10, 15]$, vrhu 2 interval $[10, 20]$, vrhu 3 interval $[20, 30]$, vrhu 4 interval $[15, 40]$, vrhu 5 interval $[40, 50]$, a vrhu 6 interval $[90, 100]$. Primijetimo da su intervali pridruženi vrhovima 1 i 6 disjunktni, dakle, oni neće biti susjedni, $I_1 \cap I_6 = [10, 15] \cap [90, 100] = \emptyset$. Također, vidimo da se intervali pridruženi vrhovima 1 i 2 sijeku, $I_1 \cap I_2 = [10, 15] \cap [10, 20] = [10, 15] \neq \emptyset$, što implicira da će biti povezani bridom. Na isti način dobijemo da će vrhovi 1 i 4, 1 i 3, 2 i 4, 3 i 4, 4 i 5 biti susjedni te ostali parovi vrhova neće.

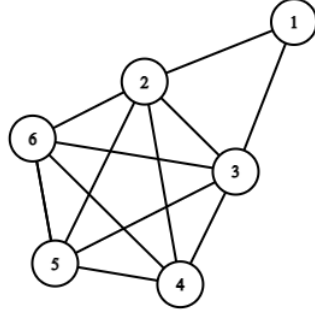


Slika 10: Intervalni graf

Ako je graf $G = (V(G), E(G), \psi(G))$ graf presjeka (intervalni graf), onda je svaki inducirani podgraf H od G također graf presjeka (intervalni graf), gdje je $V(H) = X \subseteq V(G)$, $E(H) = \{uv \in E(G) | u, v \in X\}$.

Intervalni grafovi važni su zbog njihove primjene na probleme planiranja. U primjeru 2.1., intervali su predstavljali fiksne vremenske intervale za skup sastanaka kojima je trebalo dodijeliti prostorije. Pripadni intervalni graf prikazan je na slici 11. Pronalaženje odgovarajuće dodjele prostorija može se promatrati kao problem pravilnog bojanja intervalnog grafa, gdje su prostorije za sastanke boje, a susjednim vrhovima moraju biti dodijeljene različite boje. U navedenom primjeru ne postoji pravilno 4-bojenje zbog toga što in-

tervalni graf ima kliku veličine 5. Doista, jedini podskupovi koji bi u ovom primjeru mogli biti obojeni istom bojom su $\{1, 4\}$ ili $\{1, 5\}$ ili $\{1, 6\}$.



Slika 11: Pripadni intervalni graf za primjer 2.1.

2.3 Tolerancijski grafovi

Tolerancijske grafove možemo shvatiti kao generalizaciju intervalnih grafova te ih slično i definiramo, ali sada će veličina presjeka intervala imati ključnu ulogu. Naime, hoće li dva vrha biti susjedna ili ne ovisi o veličini presjeka njihovih pripadnih intervala. Drugim riječima, ako se presjek može "tolerirati" ili zanemariti, onda oni neće biti susjedni. Formalno ih definiramo u nastavku.

Definicija 2.3 Graf $G = (V(G), E(G), \psi(G))$ je **tolerancijski graf** ako se svakom vrhu $v \in V$ može pridružiti zatvoreni interval I_v i tolerancija $t_v \in \mathbb{R}^+$ tako da je $xy \in E$ ako i samo ako $|I_x \cap I_y| \geq \min\{t_x, t_y\}$.

Kolekcija intervala i tolerancija, $\langle I, t \rangle$, naziva se **tolerancijska reprezentacija** gdje je $I = \{I_x \mid x \in V\}$ i $t = \{t_x \mid x \in V\}$.

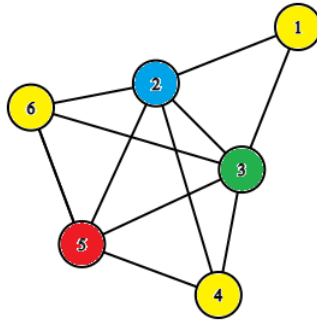
Definicija 2.4 Ako u grafu G vrijedi $t_v \leq |I_v|$ za sve $v \in V$, tada se G naziva **ograničeni tolerancijski graf**, a pripadna reprezentacija se naziva **ograničena tolerancijska reprezentacija**.

Razmotrimo još jednom primjer 2.1., ako bi svaka od tolerancija bila 5 minuta, onda bi tolerancijski graf bio isti kao intervalni graf jer su svi

neprazni presjeci veći od 5 minuta. Međutim, ako su tolerancije od I_4 i I_6 20 minuta (ili nešto veće od 15 minuta), a svaka od ostalih 5 minuta, tada tolerancijski graf ne bi imao brid između vrhova 4 i 6, kao što je prikazano na slici 12;

$$\begin{aligned} I_4 \cap I_6 &= [10 : 00 - 11 : 00] \cap [10 : 45 - 12 : 30] = [10 : 45 - 11 : 00] \\ \Rightarrow |I_4 \cap I_6| &= |10 : 45 - 11 : 00| = 00 : 15 \leq \min\{00 : 20, 00 : 20\} = 00 : 20 \\ \Rightarrow \text{vrhovi 4 i 6} &\text{ nisu susjedni.} \end{aligned}$$

U ovom slučaju, tolerancijski graf je 4-obojev, što osigurava dosljednu dodjelu četiri prostorijske za sastanke.



Slika 12: Pravilno bojenje tolerancijskog grafa iz primjera 2.1. gdje I_4 i I_6 imaju toleranciju od 20 minuta, a ostali intervali 5 minuta

Napomena 2.1 Za danu (ograničenu) tolerancijsku reprezentaciju $\langle I, t \rangle$ grafa G , za bilo koji podskup vrhova $W \subseteq V(G)$ intervali $\{I_w \mid w \in W\}$ i tolerancije $\{t_w \mid w \in W\}$ daju reprezentaciju G_W . Dakle, inducirani podgrafovi tolerancijskih grafova su također tolerancijski grafovi, a inducirani podgrafovi ograničenih grafova tolerancije također su ograničeni grafovi tolerancije.

U tolerancijskoj reprezentaciji grafa G možemo imati intervale oblika $I_x = [a_x, a_x]$, za neki $x \in V(G)$. U ovom slučaju, za bilo koji drugi interval I_y , za neki $y \in V(G)$, vrijedi $|I_x \cap I_y| = 0 < \min\{t_x, t_y\}$. Budući da su tolerancije elementi iz \mathbb{R}^+ , nejednakost je stroga. Dakle, $xy \in E(G)$. Međutim, u ograničenoj tolerancijskoj reprezentaciji nemoguće je da će bilo kakav interval I_v biti točka jer mora vrijediti $|I_v| \geq t_v > 0$.

Napomena 2.2 (i) Ako je $\langle I, t \rangle$ tolerancijska reprezentacija grafa G i I_v točka, tada je odgovarajući vrh v izolirani vrh u G , tj. vrh koji nije incidentan ni s jednim bridom.

(ii) Ako je $\langle I, t \rangle$ ograničena tolerancijska reprezentacija grafa G , tada nijedan interval I_v nije točka, odnosno graf G nema izoliranih vrhova.

Često je potrebno da tolerancijska reprezentacija grafa G zadovoljava jedno ili više dodatnih svojstava:

- (1) Ako za vrh $x \in V(G)$ vrijedi $|I_x| < t_x$, onda $t_x = \infty$,
(svaka tolerancija veća od duljine odgovarajućeg intervala jednaka je beskonačno).
- (2) $\forall x, y \in V(G), x \neq y$ t.d. $t_x, t_y \neq \infty$ vrijedi $t_x \neq t_y$,
(sve tolerancije su različite, osim onih koje su jednake beskonačno).
- (3) $\forall I_x = [a_x, b_x], I_y = [a_y, b_y], I_x \neq I_y$ vrijedi $a_x \neq a_y, a_x \neq b_y, b_x \neq a_y, b_x \neq b_y$,
(ne postoje dva različita intervala koja imaju istu krajnju točku).

Definicija 2.5 Tolerancijska reprezentacija koja zadovoljava sva tri prethodno navedena svojstva naziva se **regularna reprezentacija**.

Lema 2.1 Svaki graf tolerancije ima regularnu reprezentaciju.

Dokaz:

Neka je $G = (V(G), E(G), \psi(G))$ graf tolerancije i neka je $\langle I, t \rangle$ njegova tolerancijska reprezentacija. Pozivat ćemo se na krajeve intervala u tolerancijskoj reprezentaciji grafa. Lijevi kraj intervala I_v označit ćemo s $L(v)$, a desni kraj s $R(v)$. Dakle, neka je $I_v = [L(v), R(v)]$ za svaki $v \in V(G)$.

Dokaz od (1):

Ako je $t_x > |I_x|$ za neki vrh $v \in V(G)$, onda vrijedi $xy \in E(G) \iff |I_x \cap I_y| \geq t_y$. Navedeno vrijedi i u slučaju kada je $t_x = \infty$.

Dokaz od (2) i (3):

Definirajmo sljedeće skupove:

- (i) $\{|L(x) - L(y)| : x, y \in V(G)\}$,
- (ii) $\{|R(x) - R(y)| : x, y \in V(G)\}$,
- (iii) $\{|L(x) - R(y)| : x, y \in V(G)\}$,
- (iv) $\{t_x : x \in V(G)\}$,
- (v) $\{|t_x - t_y| : x, y \in V(G)\}$,
- (vi) $\{t_x - |I_x \cap I_y| : x, y \in V(G)\}$.

Neka je ϵ najmanji pozitivan broj koji se pojavljuje u uniji skupova (i)-(vi). Ako su x i y različiti vrhovi iz $V(G)$ za koje vrijedi $t_x = t_y$, onda uzmimo jedan od njih. Bez smanjenja općenitosti, neka to bude x . Zamijenimo t_x sa $t'_x = t_x - \frac{\epsilon}{2}$ i ostavimo t_y nepromijenjen. Pokazat ćemo da se time dobiva reprezentacija grafa G s jednom tolerancijom koja se ponavlja manje. Ako je $xz \in E(G)$, onda vrijedi $|I_x \cap I_z| \geq \min\{t_x, t_z\} \geq \min\{t'_x, t_z\}$. Ako pak vrijedi $xz \notin E(G)$, onda vrijedi $|I_x \cap I_z| < \min\{t_x, t_z\}$ te zbog tako odabranog ϵ , znamo $\min\{t_x, t_z\} - |I_x \cap I_z| \geq \epsilon$. Dakle, slijedi $|I_x \cap I_z| \leq \min\{t_x, t_z\} - \epsilon < \min\{t'_x, t_z\}$, što se i trebalo pokazati. Ako još postoje dva različita vrha s istom tolerancijom, ponovno se izračuna ϵ te se ponovi navedeni postupak dok sve tolerancije nisu različite.

Pretpostavimo sada da dva intervala imaju jednak neki od krajeva. Neka je $S = \{L(v), R(v) : v \in V(G)\}$ skup svih krajeva intervala u tolerancijskoj reprezentaciji grafa G . Označimo sa s najmanji kraj u S koji se pojavljuje barem dva puta, odnosno najmanji kraj koji se ponavlja. Neka su x i y različiti vrhovi u $V(G)$ za koje je s neki od krajeva intervala I_x i I_y . Ako postoje $x \in V(G)$ tako da je $R(x) = s$, uzmimo onaj čiji je interval I_x najduži i zamijenimo I_x s $I'_x = [L(x) + \frac{\epsilon}{2}, R(x) + \frac{\epsilon}{2}]$. Ako ne postoji $x \in V(G)$ tako da je $R(x) = s$, onda zbog načina na koji smo odabrali x sigurno postoje vrhovi $x \in V(G)$ tako da je $L(x) = s$. Uzmimo onaj za koji je $|I_x|$ najveća moguća i zamijenimo I_x s $I'_x = [L(x) + \frac{\epsilon}{2}, R(x) + \frac{\epsilon}{2}]$. Vidimo da na ovaj način dobivamo tolerancijsku reprezentaciju grafa G s jednim parom intervala, koji imaju jednak neki od krajeva manje. Sve tolerancije su ostale različite. Ako i dalje postoji par intervala takav da svi krajevi nisu različiti, potrebno je ponovno

izračunati ϵ i ponoviti navedeni postupak sve dok krajevi svih intervala nisu međusobno različiti. \square

Napomena 2.3 *Transformacija u dokazu Leme 2.3. ne mijenja duljinu intervala.*

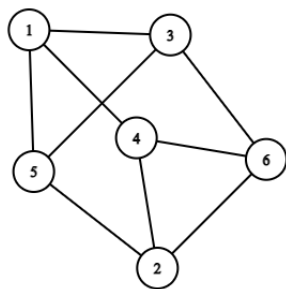
2.4 Savršeni grafovi

Jedno od bitnih svojstava tolerancijskih grafova je to da su savršeni grafovi.

Definicija 2.6 *Za graf G kažemo da je **savršen** ako je, za svaki inducirani podgraf H od G , kromatski broj od H jednak broju vrhova u najvećoj kliku u H . Odnosno, ako vrijedi $\chi(H) = \omega(H)$.*

Nadalje, L. Lovász je dokazao da je graf savršen ako i samo ako je njegov komplement također savršen. Navedeni rezultat je izvorno pretpostavio C. Berge, a poznat je kao Teorem o savršenom grafu (The perfect graph theorem).

Savršene grafove koristimo zato što imaju svojstvo da, za probleme koji su generalno NP-potpuni, dopuštaju algoritme za rješavanje u polinomnom vremenu.



Slika 13: Savršen graf

3 Primjena tolerancijskih grafova

U cilju suzbijanja širenja zaraze virusom COVID-19, u ovom su poglavlju predstavljene različite primjene tolerancijskih grafova. Svaka od navedenih

primjena dana je s algoritmom implementiranim u programu SageMath. Algoritmi se nalaze u 4. poglavlju, a preuzeti su iz [3].

3.1 Organizacija letova

Način na koji se mogu organizirati letovi tijekom COVID-19 pandemije jedan je od problema u stvarnom životu koji se može modelirati pomoću tolerancijskih grafova.

Problem je sljedeći:

Zračna luka Z ima pravilo da oni putnici koji koriste domaće letove moraju doći na ulazna vrata barem 2 sata prije polaska, a oni putnici koji koriste međunarodne letove moraju doći barem 3 sata prije. Kako bi se smanjila mogućnost širenja zaraze među putnicima, zračna luka želi rasporediti letove a i b na različite ulaze ako se vremenski period u kojemu putnici moraju biti na ulaznim vratima poklapa za više od 30 minuta za domaće letove, a više od 15 minuta za međunarodne. Dakle, potrebno je naći najmanji mogući broj različitih ulaza koji se mogu koristiti svaki dan u zračnoj luci Z .

Neka s_j označava vrijeme predviđeno za polazak leta j . Ako je let j međunarodan, onda neka je $I_j = [s_j - 3, s_j]$, zbog toga što putnici moraju biti 3 sata prije polaska na ulazu. Ako je j domaći let, putnici moraju doći 2 sata prije pa pišemo $I_j = [s_j - 2, s_j]$. Nadalje, neka je $I = \{I_1, \dots, I_x, I_{x+1}, \dots, I_{x+y}\}$ skup intervala za svaki let planiran za određeni dan s x domaćih i y međunarodnih letova. Svaki interval predstavlja vrh u grafu G . Svakom vrhu dodajemo toleranciju t_i , $i = 1, \dots, v$, ($v = x + y$), koja ovisi o broju putnika predviđenih za i -ti let. Ako je $|I_a \cap I_b| \leq \min\{t_a, t_b\}$ za letove a i b , onda ista ulazna vrata mogu biti korištena. Najmanji broj vratiju koja se moraju koristiti je jednak kromatskom broju pripadnog tolerancijskog grafa, gdje svaka boja predstavlja jedna ulazna vrata.

Primjer 3.1 *Promotrimo raspored odlaznih letova, međunarodnih i domaćih, iz zračne luke Venecija Marco Polo, Italija, na dan 17. ožujka 2022. godine. Oznaka M znači da je let međunarodan, a D da je let domaći. Uvedimo sada tolerancije; Neka je tolerancija za međunarodne letove jednaka 15*

minuta. Kako je općenito broj putnika na domaćim letovima manji, tolerancija će biti veća pa neka ona iznosi 30 minuta. Dakle, tolerancija nam označava vremenski period u kojemu je dozvoljeno da se putnici za dva različita leta nalaze na istim vratima za ulazak. Podaci o letovima su prikazani u tablici 1.

Broj leta	Odredište	Vrijeme polaska	Međunarodni/domaći
0	München	6:15	M
1	Paris	6:25	M
2	Amsterdam	6:30	M
3	Bruxelles	6:45	M
4	Manchester	7:00	M
5	Madrid	7:15	M
6	Palermo	8:50	D
7	Bari	9:55	D
8	Istanbul	10:20	M
9	Frankfurt	10:40	M
10	Rim	11:10	D
11	Barcelona	13:15	M
12	Berlin	13:30	M
13	Catania	13:40	D
14	Zurich	14:55	M
15	Dubai	15:25	M
16	Naples	17:00	D
17	Catania	17:15	D
18	Lisabon	17:40	M
19	London	18:10	M
20	Vienna	19:15	M
21	Casablanca	19:30	M
22	Dublin	19:45	M
23	Palermo	22:20	M

Tablica 1: Letovi iz Venecije 17.3.'22.

Želimo naći najmanji mogući broj različitih ulaza koji se moraju koristiti toga dana uz navedene pretpostavke. Odnosno, problem se svodi na pronalazak kromatskog broja pripadnog tolerancijskog grafa. Drugim riječima, kromatski broj je najmanji broj vratiju koja se moraju koristiti pri čemu svaka boja predstavlja jedna ulazna vrata.

Promotrimo intervale nekoliko letova. Za let 0 za München, čiji je polazak u 6:15 vrijedi da je međunarodan pa će stoga njegov pripadni interval biti $I_0 = [3 : 15 - 6 : 15]$, a tolerancija će iznositi 00:15. Let 6 za Palermo kreće u 8:50 te je domaći pa će njegov interval biti $I_6 = [6 : 50 - 8 : 50]$, a tolerancija 00:30. Pitamo se možemo li koristiti ista ulazna vrata za navedena dva leta. Pogledajmo presjek njihovih intervala:

$$I_0 \cap I_6 = [3 : 15 - 6 : 15] \cap [6 : 50 - 8 : 50] = \emptyset.$$

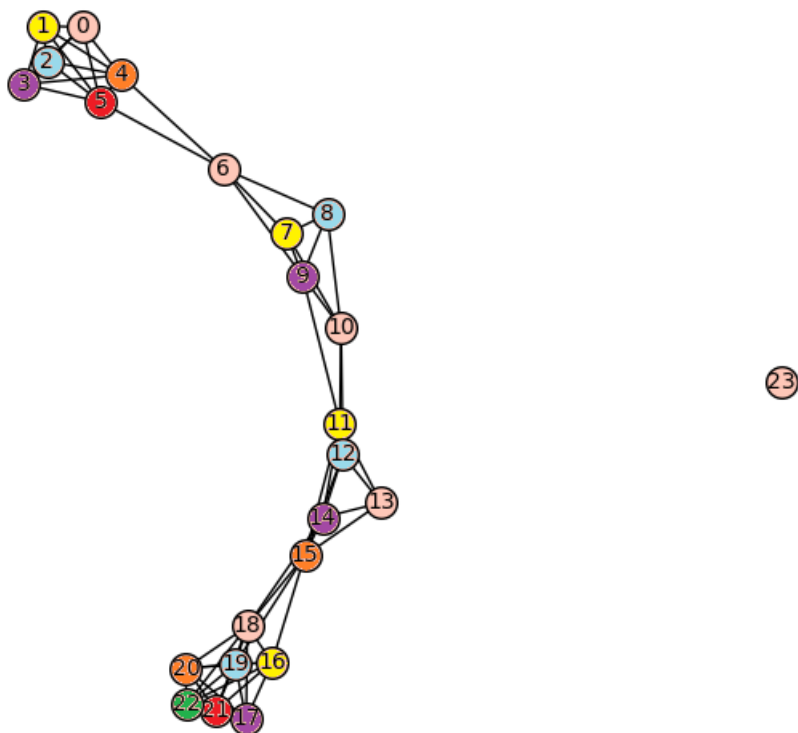
Dakle, imamo:

$$|I_0 \cap I_6| = 0 \leq \min\{t_0, t_6\} = \min\{00 : 15, 00 : 30\} = 00 : 15,$$

Zaključujemo da letovi 0 i 6 mogu koristiti ista ulazna vrata. Nadalje, za let 5 za Madrid, čiji je polazak u 7:15, vrijedi $I_5 = [4 : 15 - 7 : 15]$ i $t_5 = 00 : 15$. Provjermimo trebaju li nam nova ulazna vrata za let za Madrid.

$$\begin{aligned} I_0 \cap I_5 &= [3 : 15 - 6 : 15] \cap [4 : 15 - 7 : 15] = [4 : 15 - 6 : 15], \\ |I_0 \cap I_5| &= |[4 : 15 - 6 : 15]| = 02 : 00 \geq \min\{t_0, t_5\} = 00 : 15, \\ I_5 \cap I_6 &= [4 : 15 - 7 : 15] \cap [6 : 50 - 8 : 50] = [6 : 50 - 7 : 15], \\ |I_5 \cap I_6| &= |[6 : 50 - 7 : 15]| = 00 : 25 \geq \min\{t_5, t_6\} = 00 : 15. \end{aligned}$$

Dobivamo da su presjeci pripadnih intervala letova 0 i 5 te 5 i 6 veći od minimuma pripadnih dviju tolerancija pa slijedi da se dobiveno vremensko preklapanje ne može "tolerirati". Dakle, za let 5 moramo koristiti različita ulazna vrata od vratiju za letove 0 i 6.



Slika 14: Tolerancijski graf za raspored letova u primjeru 3.1

Nakon pokretanja programa s danim podacima, dobivamo da najmanji broj vratiju koja se moraju koristiti jest 7. Kao što je prikazano na slici 14, graf je 7-kromatski.

Nadalje, dobivamo sljedeće rješenje:

Vrata 0 korištena su za letove 0, 6, 10, 13, 18, 23;

vrata 1 korištena su za letove 1, 7, 11, 16;

vrata 2 korištena su za letove 2, 8, 12, 19;

vrata 3 korištena su za letove 3, 9, 14, 17;

vrata 4 korištena su za letove 4, 15, 20;

vrata 5 korištena su za letove 5, 21;

vrata 6 korištena su za let 22.

3.2 Korištenje učionica pri nastavi uživo

Tijekom pandemije moguće je neka predavanja ili sastanke organizirati uživo, uz poštivanje epidemioloških mjera. Jedan od glavnih problema koji se pojavljuje je korištenje učionica za održavanje nastave iz različitih kolegija. To jest, pojavljuje se problem izrade rasporeda tako da se koristi minimalan broj učionica za različite kolegije, a da se pri tome mjere poštuju.

Problem glasi:

Neka zgrada sveučilišta ima k učionica te neka je v broj kolegija čija bi se predavanja trebala održati u jednom danu. Za svaki kolegij je točno određeno vremensko razdoblje u kojem će se nastava održati. Nakon svakog predavanja, dodan je još jedan sat za čišćenje učionice između dvije upotrebe. Pod tim uvjetima tražimo minimalan broj različitih učionica koje se moraju koristiti svaki dan u zgradi sveučilišta.

Neka je $I_j = [s_j, e_j + 1]$, gdje su s_j i e_j vrijeme početka i završetka nastave iz kolegija j predviđene za taj dan, respektivno. Pretpostavimo da je za čišćenje dodano 60 min (1 h). Nadalje, neka je $I = \{I_1, \dots, I_v\}$ skup intervala za v kolegija. Svaki interval predstavlja vrh grafa G . Svakom vrhu dodajemo toleranciju t_i , $i = 1, \dots, v$, gdje tolerancija t_i predstavlja broj studenata koji su upisali kolegij. Ako je $|I_x \cap I_y| \leq \min\{t_x, t_y\}$, za kolegije x i y može se koristiti ista učionica. Minimalni broj učionica koje se mora koristiti jednak je kromatskom broju odgovarajućeg tolerancijskog grafa gdje svaka boja predstavlja jednu učionicu. Važno je istaknuti da se navedeni pristup može koristiti i za organizaciju sastanaka i mnogih drugih događanja.

3.3 Posjeti u muzej

Dok je pandemija prisutna u našim životima, obilasci muzeja, izložbe i mnoga druga slična događanja obično se organiziraju u skladu s propisanim mjerama i odgovarajućim rasporedom. Jedan od glavnih ciljeva epidemiologa je otkrivanje kontakata osobe zaražene virusom, a tolerancijski grafovi mogu nam pomoći da epidemiolozima damo popis mogućih kontakata.

Imamo sljedeći problem:

Neka je $I_j = [s_j, e_j]$, gdje su s_j i e_j vrijeme početka i završetka posjeta osobe (ili grupe ljudi) određenom muzeju, respektivno. Neka je v broj ljudi (ili različitih grupa ljudi) predviđenih za posjet muzeju i neka je $I = \{I_1, \dots, I_v\}$ skup navedenih intervala. Svaki interval predstavlja vrh grafa G . Svakom vrhu dodajemo toleranciju t_i , $i = 1, \dots, v$, gdje tolerancija t_i predstavlja vremenski period koji im je epidemiolog definirao kao trajanje kontakta koje bi moglo dovesti do infekcije. Ako je $|I_x \cap I_y| \geq \min\{t_x, t_y\}$, za osobe (grupe) x i y , a osoba (grupa) x je zaražena virusom, tada epidemiolozi mogu propisati odgovarajuće mjere za osobu (grupu) y . Isti pristup može se koristiti za druge slične događaje koji uključuju različite skupine ljudi koji posjećuju isto mjesto.

3.4 Traženje kritičnih događaja

Jedan od najvažnijih koraka u sprječavanju širenja virusa jest otkrivanje događaja na kojima postoji mogućnost od zaraze velikog broja ljudi. Zbog toga ih nazivamo kritičnima. Postoje dvije mogućnosti; jedna je da se oni u potpunosti zabrane, a druga je da se ograniči broj sudionika tog događaja.

Model navedenog problema je sljedeći:

Neka je G graf čiji skup vrhova odgovara skupu intervala $I = \{I_1, \dots, I_v\}$. Svaki je vrh predstavljen intervalom $I_j = [s_j, e_j]$, pri čemu s_j predstavlja inicijalno vrijeme kada je osoba j zaražena, a e_j predstavlja vrijeme prestanka zaraze. Svakom vrhu dodana je tolerancija t_i , $i = 1, \dots, v$ određena od strane epidemiologa. Ako vrijedi da je $|I_a \cap I_b| \geq \min\{t_a, t_b\}$ za osobe a i b , onda su vrhovi koji odgovaraju tim osobama susjedni. U navedenom grafu tolerancije tražimo neproširive klike ili maksimalne klike. Promatrajući dobivene klike, epidemiolozi mogu ispitati jesu li odgovarajući ljudi bili prisutni na nekom događaju. Ako pronađu takav događaj koji odgovara velikoj kliku, odnosno veliki broj zaraženih osoba koja su povezane, mogu ga obilježiti kritičnim. Tada, potrebne mjere se mogu uvesti i tako smanjiti širenje zaraze na takvom i sličnim događajima.

4 Algoritmi

U ovom poglavlju navedeni su algoritmi implementirani u programu Sage-Math koji su korišteni u svrhu pronalaska rješenja problema iz prethodnog poglavlja. Algoritmi su preuzeti iz literature [3].

4.1 Algoritam za organizaciju letova

ULAZ: uređena trojka (s_j, e_j, t_j) gdje (s_j, e_j) predstavlja segment (odnosno vrh grafa) na realnoj osi, a t_j pripadnu toleranciju.

IZLAZ: podaci o broju vrhova grafa i pravilno bojenje vrhova

```
1 int_tol = [(s_j,e_j,t_j) za j u [1,...,v]]
2
3 def schedule_flights(int_tol):
4     g = graphs.ToleranceGraph(int_tol)
5     from sage.graphs.graph_coloring import chromatic_number
6     cronum = chromatic_number(g)
7     from sage.graphs.graph_coloring import all_graph_colorings
8     GraphColoring = all_graph_colorings(g,cronum)
9     return [g, cronum, next(GraphColoring)]
```

U primjeru opisanom u potoglavlju 3.1. imamo sljedeće podatke:

```
1 int_tol = [(3.15,6.15,0.15), (3.25,6.25,0.15), (3.30,6.30,0.15), (3.45,6.45,0.15),
2 (4.00,7.00,0.15), (4.15,7.15,0.15), (6.50,8.50,0.3), (7.55, 9.55, 0.3), (7.20,10.20,0.15),
3 (7.40,10.40,0.15), (9.10,11.10,0.3), (10.15,13.15,0.15), (10.30,13.30,0.15),
4 (11.40,13.40,0.3), (11.55,14.55,0.15), (12.25,15.25,0.15), (15.00, 17.00,0.3),
5 (15.15,17.15,0.3), (14.40,17.40, 0.15), (15.10, 18.10,0.15), (16.15, 19.15, 0.15),
6 (16.30,19.30,0.3), (16.45, 19.45,0.15), (20.20,22.20, 0.3)]
```

Nakon pokretanja programa dobivamo:

```
1 [Graph on 24 vertices,
2 7,
```

```

3  {0: [0, 6, 10, 13, 18, 23],
4  1: [1, 7, 11, 16],
5  2: [2, 8, 12, 19],
6  3: [3, 9, 14, 17],
7  4: [4, 15, 20],
8  5: [5, 21],
9  6: [22]}}

```

4.2 Algoritam za korištenje učionica u nastavi uživo

ULAZ: uređena trojka $(s_j, e_j + 1, t_j)$ gdje (s_j, e_j) predstavlja segment (odnosno vrh grafa) na realnoj osi, a t_j pripadnu toleranciju.

IZLAZ: podaci o broju vrhova grafa i pravilno bojenje vrhova

```

1  int_tol = [(s_j, e_j + 1, t_j) za j u [1, ..., v]]
2
3  def classroom(int_tol):
4      g = graphs.ToleranceGraph(int_tol)
5      from sage.graphs.graph_coloring import chromatic_number
6      cronum = chromatic_number(g)
7      from sage.graphs.graph_coloring import all_graph_colorings
8      GraphColoring = all_graph_colorings(g, cronum)
9      return [g, cronum, next(GraphColoring)]

```

4.3 Algoritam za raspored posjeta u muzej

ULAZ: uređena trojka $(s_j, e_j + 1, t_j)$ gdje (s_j, e_j) predstavlja segment (odnosno vrh grafa) na realnoj osi, a t_j pripadnu toleranciju.

IZLAZ: popis susjednih vrhova

```

1  int_tol = [(s_j, e_j, t_j) za j u [1, ..., v]]
2
3  def museum_vis(int_tol):

```

```

4     g = graphs.ToleranceGraph(int_tol)
5     ver = g.vertices()
6     neigh = [g.neighbors(v) for v in range(len(ver))]
7     return neigh

```

4.4 Algoritam za traženje kritičnih događaja

ULAZ: uređena trojka (s_j, e_j, t_j) gdje (s_j, e_j) predstavlja segment (odnosno vrh grafa) na realnoj osi, a t_j pripadnu toleranciju.

IZLAZ: uređen par (num_c, max_c) gdje je num_c broj vrhova najveće klike, a max_c najveća klika u pripadnom grafu.

```

1 int_tol = [(s_j, e_j, t_j) za j u [1, ..., v]]
2
3 def critical_events(int_tol):
4     g = graphs.ToleranceGraph(int_tol)
5     max_c = g.cliques_maximum()
6     num_c = g.clique_number()
7     return [num_c, max_c]

```

5 Grafovi sličnosti

5.1 Uvod u grafove sličnosti

U praćenju i predviđanju širenja pandemije vrlo je važna umjetna inteligencija koja koristi različite skupove podataka. Primjerice, rendgenske slike prsnog koša ili vremenski niz podataka. Primjenom klasifikacijskih tehnika na rendgenske slike prsnog koša dobivaju se dobri rezultati u otkrivanju je li osoba preboljela zarazu Covidom-19. Klasifikacijske tehnike se koriste za klasifikaciju slika u unaprijed definirani skup klasa, koji se također nazivaju izlazni slojevi u neuronskim mrežama. S druge strane, tehnike grupiranja (eng. clustering) se koriste za podjelu podatka u skupine tako da se u istoj

skupini nalaze podaci koji dijele slična svojstva prema kojima ih razvrstavamo. Primjerice, moguće je grupirati zemlje prema sličnom širenju zaraze Covidom-19.

Mnoge primjene grafova nastale su zahvaljujući analitici grafova, nedavnom području istraživanja. Analitika grafova povezuje matematičku definiciju grafa s njegovom primjenom u računalnoj znanosti, gdje se na graf gleda kao na složenu podatkovnu strukturu. Jedna od važnije analitike grafova je otkrivanje zajednica - inteligentno i nenadzirano grupiranje skupa podataka modeliranih grafovima korištenjem njihove sličnosti. Dakle, iz toga proizlazi motivacija za uvođenje i korištenje grafova sličnosti. Važan algoritam za otkrivanje zajednica je spektralno grupiranje.

5.2 Grafovi sličnosti

Jedan od načina na koji se podaci mogu predstaviti, odnosno modelirati je pomoću grafa sličnosti.

Definicija 5.1 *Neka je dan skup podatkovnih točaka x_1, \dots, x_n i nenegativnih sličnosti s_{ij} između svakog para točaka x_i i x_j . Neka je $G = (V(G), E(G), \psi(G))$ graf sa skupom vrhova $V = \{v_1, \dots, v_n\}$ gdje svaki vrh v_i predstavlja podatkovnu točku x_i . Dva su vrha v_i i v_j povezana ako je sličnost s_{ij} između odgovarajućih podatkovnih točaka x_i i x_j pozitivna ili veća od određenog praga, a težina tog brida iznosi s_{ij} . Takav graf G nazivamo **grafom sličnosti** te je on neusmjeren težinski graf.*

Intuitivno, cilj grupiranja je podatkovne točke podijeliti u nekoliko skupina tako da točke u istoj skupini dijele slična svojstva, a da se one koje su u različitim skupinama razlikuju. S obzirom na to da su težine bridova u grafu zapravo sličnosti između odgovarajućih krajnjih točaka, grupiranje u kontekstu grafa sličnosti je sljedeće: treba pronaći particiju grafa tako da bridovi između različitih skupina imaju vrlo male težine, a bridovi unutar iste skupine imaju velike težine.

5.3 Vrste grafova sličnosti

Postoji nekoliko načina za konstrukciju grafa sličnosti, odnosno transformaciju skupa podatkovnih točaka s danom sličnošću ili udaljenošću između svakog para. Cilj je modelirati susjedne odnose između točaka. Razlikujemo sljedeće vrste grafova sličnosti:

- **Graf ϵ - susjedstva:** Dvije su točke povezane (pa time i odgovarajući vrhovi) ako je udaljenost između njih manja od ϵ . Graf ϵ - susjedstva nije težinski.
- **Graf k-najbližih susjeda:** Vrhovi v_i i v_j će biti susjedni ako je v_j među k-najbližih susjeda od v_i ili v_i među k-najbližih susjeda od v_j .

Postoji još jedna vrsta grafa k-najbližih susjeda, a to je da su vrhovi v_i i v_j susjedni ako je istovremeno v_j među k-najbližih susjeda od v_i i v_i među k-najbližih susjeda od v_j . Time se dobiva **graf međusobnih k-najbližih susjeda**.

U oba slučaja, nakon povezivanja odgovarajućih vrhova bridom, uvodimo težine bridova ovisno o sličnosti između njihovi krajeva.

- **Potpuno povezani graf:** Dva su vrha v_i i v_j susjedna ako je sličnost s_{ij} između njih pozitivna ili veća od nekog određenog praga, a svakom bridu dodana je težina s_{ij} .

Prije nego se krene s konstrukcijom grafa sličnosti, potrebno je definirati funkciju sličnosti na podacima. Budući da će se konstruirati graf susjedstva, lokalna susjedstva inducirana ovom funkcijom sličnosti moraju imati smisla. Dakle, moramo biti sigurni da su točke koje funkcija sličnosti smatra "vrlo sličnim" također blisko povezane u skupu podataka iz kojih dolaze. U uobičajenom slučaju kada podatkovne točke žive u euklidskom prostoru, dobar izbor za funkciju sličnosti je Gaussova funkcija sličnosti

$$s(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}},$$

gdje parametar σ kontrolira širinu susjedstva. Ovaj parametar igra sličnu ulogu kao parametar ϵ u slučaju grafa ϵ -susjedstva.

Svi navedeni grafovi redovito se koriste u spektralnom grupiranju.

6 Spektralno grupiranje

Spektralno grupiranje je kompatibilno s podacima koji su modelirani grafovima te je jedno od glavnih tehnika grupiranja grafova. Sastoji se od izračunavanja Laplaceove matrice ulaznog grafa pomoću njegove matrice stupnjeva i težinske matrice susjedstva, ili matrice susjedstva ako bridovima nisu dane težine, primjerice kod grafa ϵ -susjedstva. Zatim se svojstvene vrijednosti i svojstveni vektori dobivaju iz Laplaceove matrice. Konačno, grupiranjem algoritmom k-srednjih vrijednosti dobivaju se izlazne skupine.

Postoji više vrsta spektralnog grupiranja, obzirom na odabir tipa Laplaceove matrice.

6.1 Laplaceove matrice

Definicija 6.1 *Neka je zadan graf sličnosti G sa skupom vrhova $V = \{v_1, \dots, v_n\}$ te njegova matrica stupnjeva $D = [d_{ij}]$ i težinska matrica susjedstva $W = [w_{ij}]$, $i, j \in \{1, \dots, n\}$. **Ne-normalizirana Laplaceova matrica grafa** definira se kao*

$$L = D - W.$$

Svojstva navedene matrice daje sljedeća propozicija, a njezin se dokaz može pronaći u [5].

Propozicija 6.1 *Ne-normalizirana Laplaceova matrica L redan n zadovoljava sljedeća svojstva:*

1. *Za svaki vektor $f = (f_1, \dots, f_n) \in \mathbb{R}^n$ vrijedi*

$$f^T L f = \frac{1}{2} \sum_{i,j=1}^n w_{ij} (f_i - f_j)^2.$$

2. L je simetrična ($L = L^T$) i pozitivno semi-definitna ($x^T Lx \geq 0, \forall x \in \mathbb{R}^n$)
3. Najmanja svojstvena vrijednost od L je 0, a odgovarajući svojstveni vektor je konstantni jedinični vektor.
4. L ima n nenegativnih, realnih svojstvenih vrijednosti $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$.

Nadalje, razlikujemo dvije vrste normalizirane Laplaceove matrice.

Definicija 6.2 Neka je zadan graf sličnosti G te njegova matrica stupnjeva D , težinska matrica susjedstva W i ne-normalizirana Laplaceova matrica L . **Simetrična normalizirana simetrična Laplaceova matrica grafa** definira se kao

$$L_{sym} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}} W D^{-\frac{1}{2}}.$$

Normalizirana Laplaceova matrica grafa definira se kao

$$L_{norm} = D^{-1} L = I - D^{-1} W.$$

U sljedećoj propoziciji navedena su neka njihova svojstva. Dokaz propozicije može se pronaći u [5].

Propozicija 6.2 Normalizirane Laplaceove matrice L_{norm} i L_{sym} zadovoljavaju sljedeća svojstva:

1. Za svaki $f = (f_1, \dots, f_n) \in \mathbb{R}^n$ vrijedi

$$f' L_{sym} f = \frac{1}{2} \sum_{i,j=1}^n w_{ij} \left(\frac{f_i}{\sqrt{d_i}} - \frac{f_j}{\sqrt{d_j}} \right)^2,$$

pri čemu je d_i element matrice stupnjeva D na poziciji (i, i) .

2. λ je svojstvena vrijednost od L_{norm} pridružena svojstvenom vektoru u ako i samo ako je λ svojstvena vrijednost od L_{sym} pridružena svojstvenom vektoru $w = D^{\frac{1}{2}} u$.

3. λ je svojstvena vrijednost od L_{norm} pridružena svojstvenom vektoru u ako i samo ako λ i u zadovoljavaju jednadžbu $Lu = \lambda Du$.
4. 0 je svojstvena vrijednost od L_{norm} s pridruženim svojstvenim vektorom $\mathbf{1}$. 0 je svojstvena vrijednost od L_{sym} s pridruženim svojstvenim vektorom $D^{\frac{1}{2}}\mathbf{1}$.
5. L_{sym} i L_{norm} su pozitivno semi-definitne i imaju n ne-negativnih realnih svojstvenih vrijednosti $0 = \lambda_1 \leq \dots \leq \lambda_n$.

6.2 Algoritam k-srednjih vrijednosti

Jedan od koraka u spektralnom grupiranju je algoritam k -srednjih vrijednosti. Cilj ovog algoritma je podijeliti n podataka u k skupina, tako da se svaki podatak pridruži najbližem težištu skupine. Svaka je skupina predstavljena svojim težištem. Pretpostavlja se da je broj skupina, k , unaprijed poznat. Međutim, moguće je deducirati k metodom pokušaja i pogreške. Algoritam k -srednjih vrijednosti je iterativan, počinje s proizvoljnim rješenjem, tj. odabirom k točaka te se u svakom sljedećem koraku pokušava pronaći bolje rješenje, dok god se bolje rješenje više ne može naći ili dok se ne dosegne maksimalan broj iteracija. Srce algoritma je petlja u kojoj razmatramo svaku točku i dodjeljujemo ju skupini čijemu je težištu najbliža.

Koraci algoritma:

1. Odaberi k početnih točaka t_1, \dots, t_k , za koje je vjerojatno da će biti unutar različitih skupina. Neka te točke čine težište svoje skupine, S_1, \dots, S_k respektivno.
2. Za svaku preostalu točku pronaći težište koje joj je najbliže i dodati ju skupini čijem je težištu najbliža. Odnosno točku x_i , $i = 1, \dots, k$ dodijeliti skupini S_l ako i samo ako

$$\|x_i - t_l\| < \|x_i - t_j\|, \quad j = 1, \dots, k, \quad j \neq l.$$

3. Izračunati nova težišta t_1^*, \dots, t_k^* tako da

$$t_j^* = \frac{\sum_{x_i \in S_j} x_i}{n_j}, j = 1, \dots, k,$$

pri čemu je n_j broj točaka u skupini S_j .

4. Zaustaviti algoritam ako se težišta skupina više ne mijenjaju kroz iteracije ili je dosegnut maksimalan broj iteracija. U suprotnom, postaviti $t_j = t_j^*$, $j = 1, \dots, k$ i vratiti se na korak 2.

6.3 Vrste spektralnog grupiranja

Pretpostavimo da imamo n podatkovnih točaka x_1, \dots, x_n . Sličnost između svakog para točaka $s_{ij} = s(x_i, x_j)$ mjeri se nekom funkcijom sličnosti koja je simetrična i nenegativna, a odgovarajuća matrica sličnosti označava se sa $S = [s_{ij}]$, $i, j = 1, \dots, n$. Ovisno o tome koja se Laplaceova matrica koristi, razlikujemo tri tipa spektralnog grupiranja:

i) Ne-normalizirano spektralno grupiranje; Neka je dana matrica sličnosti $S \in \mathbb{R}^{n \times n}$, broj skupina k koje treba konstruirati. Koraci su sljedeći:

1. Konstruirati graf sličnosti i naći težinsku matricu W .
2. Izračunati ne-normaliziranu Laplaceovu matricu L .
3. Izračunati svojstvene vrijednosti od L i k najmanjih označiti s u_1, \dots, u_k .
4. Neka je $U \in \mathbb{R}^{n \times k}$ matrica čiji su stupci vektori u_1, \dots, u_k .
5. Za svaki $i = 1, \dots, n$ neka je $y_i \in \mathbb{R}^k$ vektor koji odgovara i -tom retku od U .
6. Grupirati točke y_i , $i = 1, \dots, n$ algoritmom k -srednjih vrijednosti u skupine C_1, \dots, C_k .

Na kraju, konstruirati skupine A_1, \dots, A_k takve da $A_i = \{j | y_j \in C_i\}$.

ii) Shi-Malik normalizirano spektralno grupiranje; Neka je dana matrica sličnosti $S \in \mathbb{R}^{n \times n}$, broj skupina k koje treba konstruirati. Koraci su sljedeći:

1. Konstruirati graf sličnosti i naći težinsku matricu W .
2. Izračunati normaliziranu Laplaceovu matricu L_{norm} .
3. Izračunati svojstvene vrijednosti od $L_{norm}u = \lambda Du$ i k najmanjih označiti s u_1, \dots, u_k .
4. Neka je $U \in \mathbb{R}^{n \times k}$ matrica čiji su stupci vektori u_1, \dots, u_k .
5. Za svaki $i = 1, \dots, n$ neka je $y_i \in \mathbb{R}^k$ vektor koji odgovara i -tom retku od U .
6. Grupirati točke $y_i, i = 1, \dots, n$ algoritmom k -srednjih vrijednosti u skupine C_1, \dots, C_k .

Na kraju, konstruirati skupine A_1, \dots, A_k takve da $A_i = \{j | y_j \in C_i\}$.

iii) Ng, Jordan, Weiss normalizirano spektralno grupiranje; Neka je dana matrica sličnosti $S \in \mathbb{R}^{n \times n}$, broj skupina k koje treba konstruirati. Koraci su sljedeći:

1. Konstruirati graf sličnosti i naći težinsku matricu W .
2. Izračunati normaliziranu Laplaceovu matricu L_{sym} .
3. Izračunati svojstvene vrijednosti od L_{sym} i k najmanjih označiti s u_1, \dots, u_k .
4. Neka je $T \in \mathbb{R}^{n \times k}$ matrica dobivena iz U normalizacijom redaka, tj. $t_{ij} = \frac{u_{ij}}{(\sum_k u_{ik}^2)^{\frac{1}{2}}}$.
5. Za svaki $i = 1, \dots, n$ neka je $y_i \in \mathbb{R}^k$ vektor koji odgovara i -tom retku od T .
6. Grupirati točke $y_i, i = 1, \dots, n$ algoritmom k -srednjih vrijednosti u skupine C_1, \dots, C_k .

Na kraju, konstruirati skupine A_1, \dots, A_k takve da $A_i = \{j | y_j \in C_i\}$.

Sva tri navedena algoritma za spektralno grupiranje izgledaju prilično slično, osim činjenice da koriste tri različite Laplaceove matrice. U svakom je glavni trik promijeniti reprezentaciju podatkovnih točaka x_i u točke $y_i \in \mathbb{R}^k$. Ova promjena poboljšava grupiranje podataka, tako da se skupine mogu trivijalno otkriti. Konkretno, algoritam k-srednjih vrijednosti bez poteškoća otkriva skupine točaka nove reprezentacije.

7 Primjena temeljena na grafovima sličnosti

U ovom poglavlju, predstaviti će se metodologija temeljena na grafovima za praćenje pandemije, koristeći vremenski niz podataka i spektralno grupiranje kao algoritam za otkrivanje zajednice s obzirom na sličnost između podataka modeliranim grafovima. Odnosno, metodologiju učenja bez nadzora inteligentnog grupiranja država prema sličnom širenju Covida-19, uzimajući u obzir skup čimbenika kao što su stanovništvo i drugi čimbenici koji utječu na ponašanje pandemije.

Metodologija kreće od obrade podataka gdje se opisuje način na koji su podaci formatirani i skalirani. Nakon toga se podaci modeliraju pomoću grafa sličnosti. Treće, na dobiveni graf primjenjuje se spektralno grupiranje kako bi se izdvojile zajednice. Na kraju, grupirani podaci mogu se vizualizirati pomoću programa *R Studio* koji sadrži pakete prilagođene za podatke modelirane grafovima, kao na primjer paket *igraph*, *Neo4j* i *Gephi*.

7.1 Obrada podataka

Podaci na koje je navedena primjena fokusirana su vremenski nizovi podataka o broju oboljelih, umrlih i oporavljenih od zaraze Covid-19 u svakoj državi, broju napravljenih testova u svakom danu za svaku državu, broju stanovnika u svakoj državi i slično.

Važno je imati odgovarajuće ulazne podatke kako bi se dobili dobri izlazni podaci, zbog toga što je s lošim ulaznim podacima teško dobiti ispravne izlazne, unatoč tome koliko je model dobar.

Podaci se formatiraju tako da odgovaraju ulazu modela. Važno je riješiti

se suvišnih podataka kako bi model bio brži, a to se može postići objektno-orijentiranom reprezentacijom u kojoj se svaka država spominje kao pojedinačni primjer s različitim vektorima značajki koji predstavljaju vremenski niz podataka.

Značajke se mogu skalirati tako da se normaliziraju srednje vrijednosti kako bi raspon vrijednosti bio blizu raspona $[-1, 1]$. Proces normalizacije pomaže značajkama da imaju slične težine, a potom i sličan utjecaj na konvergenciju modela. Kako bi utjecaj nekih značajki bio jači od drugih, može se dodati pozitivan koeficijent. Primjerice, bilo bi dobro da je utjecaj broja oboljelih, oporavljenih i umrlih jači u usporedbi s utjecajem države. Normalizacija srednje vrijednosti se definira kao:

$$v'_i = \frac{v_i - \text{mean}(V)}{\max(V) - \min(V)},$$

gdje je:

v_i i -ta vrijednost vektora značajki v ,

v'_i normalizirana vrijednost od v_i ,

V skup vrijednosti vektora značajki v ,

$\max(V)$, $\min(V)$, $\text{mean}(V)$ su maksimalna, minimalna i srednja vrijednost značajki, respektivno.

7.2 Konstrukcija grafa sličnosti

Nakon obrade, koristeći konačnu verziju podataka može se dobiti graf sličnosti. U ovom slučaju, vrhovi predstavljaju države pa se računa sličnost između svake dvije države. Kao mjera sličnosti, koristi se Gaussova jezgra sličnosti koja se računa na sljedeći način:

$$s_{ij} = e^{-\frac{d_{ij}^2}{2\sigma^2}},$$

pri čemu je:

d_{ij} Euklidska udaljenost između država i i j ,

σ^2 veličina susjedstva korištena kao parametar za skaliranje.

Dakle, graf sličnosti je graf $G = (V(G), E(G))$, gdje je $V(G)$ skup vrhova i svaka država je prikazana vrhom, $E(G)$ je skup bridova, a težine bridova su vrijednosti sličnosti između značajki dvije države.

7.3 Spektralno grupiranje

Bilo koji od tri navedena algoritma za spektralno grupiranje se može koristiti, a izbor ovisi o tome koji je algoritam najkompatibilniji s podacima te najjednostavniji za implementirati. Prije odabira, dobro je pogledati distribuciju stupnjeva grafa sličnosti. Ako većina vrhova ima približno isti stupanj, tada su sve Laplaceove matrice vrlo slične jedna drugoj pa sva tri algoritma rade jednako dobro. Međutim, ako je razlika u stupnjevima velika, Laplaceove matrice se znatno razlikuju. Dodatno objašnjenje o tome koju matricu pa time i koji algoritam koristiti nalazi se u [5]

Zaključak

U ovom su radu objašnjeni grafovi presjeka, intervalni grafovi te posebice tolerancijski grafovi, njihova povijest nastajanja i svojstva. Predstavljene su primjene tolerancijskih grafova u borbi protiv širenja zaraze virusa Covid-19. Navedene primjene svode se na pronalazak kromatskog broja pripadnog grafa te najvećih i maksimalnih klika u grafu. Važnost tolerancijskih grafova jest da se problemi, koji su inače NP-potpuni, mogu riješiti u polinomnom vremenu. Nadalje, uvedeni su grafovi sličnosti, njihove vrste i konstrukcija. Iznesena je metodologija temeljena na grafovima sličnosti koja koristi spektralno grupiranje, a služi za praćenje širenja pandemije. Navedena metodologija još je predmet proučavanja, kao i područje analitike grafova, no od velike je koristi, kako u praćenju i predviđanju dinamike pandemije, tako i šire.

Literatura

- [1] D. Veljan, *Kombinatorna i diskretna matematika*, Algoritam, Zagreb, 2001.
- [2] M. C. Golumbic, A. N. Trenk, *Tolerance graphs*, Cambridge University Press, 2004.
- [3] D. Crnković, A. Švob, *Application of Tolerance Graphs to Combat COVID-19 Pandemic*, SN Comput. Sci. 2 (2021), 83, 8 pages
- [4] Z. Aitt El Mouden, M. Hajar, A. Jakimi, R. M. Taj *A Graph-based Methodology for Tracking Covid-19 in Time Series Datasets*, 2020. IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), 2020., pp. 5 pages
- [5] U. von Luxburg, *A Tutorial on Spectral Clustering*, Max Planck Institute for Biological Cybernetics, Statistics and Computing, 17 (4), 2007.
- [6] J. Leskovec, A. Rajaraman, J. D. Ullman, *Mining of Massive Datasets*, Millway Labs, Stanford University, 2014.

Popis slika

1	Graf G_1	4
2	Razapinjujući podgraf od G_1	4
3	Inducirani podgraf od G_1	4
4	Graf G_2	4
5	Neproširiva klika	4
6	Najveća klika	4
7	Pravilno 3-bojenje Petersenovog grafa	5
8	Vremenski intervali trajanja sastanaka (preuzeto iz [2])	8
9	Graf presjeka	9
10	Intervalni graf	10
11	Pripadni intervalni graf za primjer 2.1.	11
12	Pravilno bojenje tolerancijskog grafa iz primjera 2.1. gdje I_4 i I_6 imaju toleranciju od 20 minuta, a ostali intervali 5 minuta	12
13	Savršen graf	15
14	Tolerancijski graf za raspored letova u primjeru 3.1	19

Popis tablica

1	Letovi iz Venecije 17.3.'22.	17
---	--------------------------------------	----