

Spektralno klasteriranje

Pelin, Helena

Master's thesis / Diplomski rad

2016

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:217:536542>

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

Download date / Datum preuzimanja: **2024-04-25**



Repository / Repozitorij:

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



SVEUČILIŠTE U ZAGREBU
PRIRODOSLOVNO–MATEMATIČKI FAKULTET
MATEMATIČKI ODSJEK

Helena Pelin

SPEKTRALNO KLASITERIRANJE

Diplomski rad

Voditelj rada:
Zlatko Drmač

Zagreb, 2016.

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. _____

Sadržaj

Sadržaj	iii
Uvod	2
1 Grupiranje kao particioniranje grafa	3
1.1 Model grafa	3
1.2 Funkcije particioniranja	4
2 Laplacijan grafa	9
2.1 Pojam i svojstva	9
2.2 Povezane komponente grafa i Laplacijan	12
2.3 Normalizirana matrica susjedstva i Laplacijan	14
3 Varijacijska karakterizacija spektra	15
3.1 Varijacijska karakterizacija svojstvenih vrijednosti realnih simetričnih matrica	15
4 Spektralna biparticija grafa	22
4.1 Model	22
4.2 Pronalaženje optimalnog reza za biparticiju	23
5 Spektralna K-particija grafa	25
5.1 Model	25
5.2 Funkcije K-particioniranja	25
5.3 Matrična reprezentacija problema	26
5.4 Nalaženje optimalnog relaksiranog rješenja	29
5.5 Nalaženje optimalnog diskretnog rješenja	30
5.6 Algoritam K-particioniranja	32
6 Implementacija algoritma i primjeri	34
6.1 Konstrukcija matrice susjedstva	34

6.2	Spektralno klasteriranje na simuliranim podacima	36
7	K-sredine i spektralno klasteriranje	40
7.1	Algoritam K-sredina	40
7.2	Usporedba sa spektralnim klasteriranjem	41
7.3	Usporedba K-sredina i spektralne rotacije u spektralnom klasteriranju . .	43
8	Klasteriranje slike	45
9	Segmentacija slika normaliziranim rezom	47
10	Dodatak	50
	Bibliografija	54

Uvod

Klasteriranje, odnosno grupiranje podataka jedna je od najčešćih tehnika analize podataka, koja svoje primjene nalazi u raznim područjima statistike, strojnog učenja, računalnim znanostima, fizici, biologiji itd. U strojnom učenju klasteriranje općenito pripada skupini nenadziranog učenja (eng. *unsupervised learning*) čiji je cilj otkriti prirodnu strukturu ulaznih podataka, bez početnih notacija ili saznanja o njima, što tu metodu razlikuje od klasifikacijskog grupiranja podataka u nadziranom učenju (eng. *supervised learning*). U procesu otkrivanja te prirodne strukture podataka, otkrivaju se podaci koji su međusobno više ili manje slični, tj. dijele neko zajedničko obilježje te se na temelju toga grupiraju u tzv. klastere.

U ovom radu koncentrirat ćemo se na metodu spektralnog klasteriranja. Ona je zadnjih godina dosta istraživana, a bazira se na definiranju sličnosti između svih točaka koje želimo klasterirati, pridruženoj matrici sličnosti te njenoj spektralnoj analizi. Glavni cilj je kao i kod svakog algoritma klasteriranja - postići veliku sličnost među podacima u istoj grupi, a malu sličnost između podataka u različitim grupama.

Za rješavanje tog problema, u Poglavlju 1 podatke ćemo prikazati u formi težinskog neusmjerenog grafa G , definirati mu težinsku matricu sličnosti, tj. susjedstva W i funkcije particioniranja čijom ćemo minimizacijom odnosno maksimizacijom smanjivati ili povećavati sličnost među podacima u grupama. Funkcije particioniranja mogu se definirati na razne načine, a pokazat ćemo kako je normalizirani rez onaj koji daje najbolje rezultate.

Podataka koje želimo grupirati ima konačno mnogo, što znači da je naš problem diskretan. Pokazuje se kako je optimizacija ciljnih funkcija u tom slučaju NP-težak problem, tj. nije riješiv egzaktnim algoritmom u polinomijalnom vremenu. Zato ćemo problem pokušati riješiti barem približno na realnoj domeni gdje će nam neprekidno rješenje dati mnogi rezultati iz područja spektralne analize, odnosno varijacijske karakterizacije spektra matrice susjedstva W kojom ćemo se baviti u Poglavlju 3.

U Poglavlju 2 definirat ćemo Laplacijan i normalizirani Laplacijan grafa te pokazati kako je nalaženje njegovih svojstvenih vektora i vrijednosti ključno za rješenje problema. U Poglavlju 4 ćemo korištenjem drugog svojstvenog vektora Laplacijana doći do rješenja biparticioniranja grafa, a u Poglavlju 5 problem ćemo generalizirati na K -particioniranje i

pronalaženje K svojstvenih vektora. U šestom poglavlju vidjet ćemo kako možemo formirati grafove i matricu susjedstva te implementirani algoritam testirati na klasteriranju simuliranih podataka. Neki klasični algoritmi spektralnog klasteriranja do diskretnog rješenja K -particioniranja dolaze upotrebom metode K -sredina na dobivene svojstvene vektore, no u ovom radu do diskretnog rješenja koje najbolje aproksimira pravo doći ćemo algoritmom spektralne rotacije te u Poglavlju 7 pokazati kako rezultira boljim rezom. Spektralno klasteriranje i normalizirani rez često se primjenjuju na segmentaciju slika pa ćemo neke primjere (rekurzivnog) biparticioniranja vidjeti u Poglavljima 8 i 9.

Poglavlje 1

Grupiranje kao particioniranje grafa

1.1 Model grafa

Neka su dane točke $x_1, \dots, x_n$ koje predstavljaju objekte koje želimo klasterirati (osobe, piksele na slici...) i neka je dana sličnost $s_{ij} \geq 0$ među svim parovima točaka. Intuitivno, cilj klasteriranja je grupirati dane točke na način da je sličnost među točkama iste grupe velika, a među točkama koje pripadaju različitim grupama mala. U ovakvom slučaju praktično je problem klasteriranja predočiti u formi grafa $\mathcal{G} = (\mathbb{V}, \mathbb{E})$.

Vrhovi grafa predstavljaju točke koje grupiramo, a svaka dva vrha povezana su bridom ukoliko je sličnost među njima pozitivna. Težina brida između vrhova v_i i v_j je upravo sličnost s_{ij} .

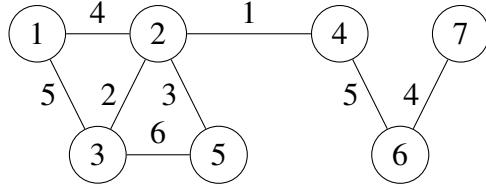
Notacija

Neka je $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ težinski neusmjereni graf sa skupom vrhova $\mathbb{V} = \{v_1, \dots, v_n\}$ i bridova $\mathbb{E}$. Težine između vrhova v_i i v_j označavat ćemo sa $w_{ij} \geq 0$.

Težinska matrica susjedstva grafa $\mathcal{G}$ je matrica $W = (w_{ij})_{i,j=1,\dots,n}$. Dva vrha nisu spojena bridom ako je $w_{ij} = 0$, a budući da je graf neusmjeren vrijedi $w_{ij} = w_{ji}$, odnosno matrica W je simetrična.

Uz matricu susjedstva definiramo i **matricu stupnjeva** $D = \text{Diag}(W\mathbb{1}_n) = \text{Diag}(d_1, \dots, d_n)$, gdje $\text{Diag}(d_1, \dots, d_n)$ označava dijagonalnu matricu sa elementima $d_1, \dots, d_n$ na dijagonali, a $d_i = \sum_{j=1}^n w_{ij}$ stupanj vrha i , $i = 1, \dots, n$.

Primjer 1.1.1. Težinski neusmjereni graf $\mathcal{G}$ sa pripadajućom matricom susjedstva W .



$$W = \begin{bmatrix} 0 & 4 & 5 & 0 & 0 & 0 & 0 \\ 4 & 0 & 2 & 1 & 3 & 0 & 0 \\ 5 & 2 & 0 & 0 & 6 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 5 & 0 \\ 0 & 3 & 6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 & 0 & 4 & 0 \end{bmatrix}.$$

Uz uvedenu notaciju, naš problem klasteriranja sada može biti preformuliran u problem particioniranja grafa $\mathcal{G}$, odnosno skupa vrhova $\mathbb{V}$. Ono što želimo postići jest grupirati skup vrhova grafa u disjunktne skupove na način da je mjera sličnosti između vrhova jednog skupa velika, a između vrhova iz različitih skupova mala.

Sljedeće ćemo razmotriti koje su to mjere sličnosti koje algoritam može koristiti prilikom particioniranja i koje su prednosti odnosno nedostaci korištenja istih.

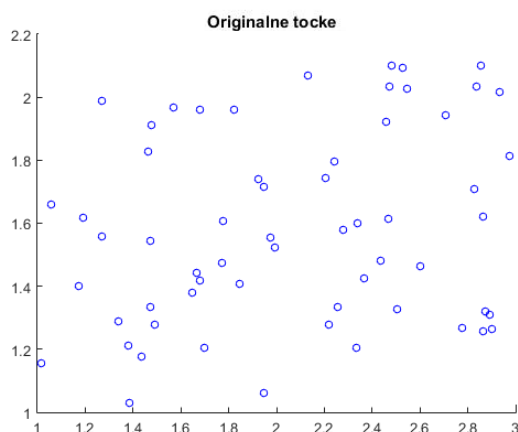
1.2 Funkcije particioniranja

Neka je dan graf $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ i pretpostavimo da želimo napraviti particiju skupa vrhova $\mathbb{V}$ na dva skupa A i B takve da $A \cap B = \emptyset$, $A \cup B = \mathbb{V}$. Jedan način na koji to možemo učiniti je da odstranimo sve bridove koji povezuju vrhove iz A sa vrhovima iz B i obratno, a mjera nesličnosti između A i B može se izračunati kao ukupna težina bridova koji su odstranjeni. Mi ćemo tu mjeru zvati **rez particije**:

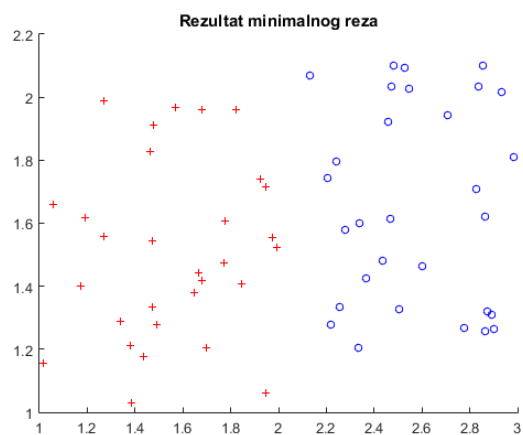
$$rez(A, B) = \sum_{i \in A, j \in B} w(i, j). \quad (1.1)$$

Problem particioniranja uz pomoć reza particije svodi se na nalaženje optimalnog reza koji će minimizirati vrijednost gore definiranog reza particije, odnosno pronaći rez sa najmanjim težinama bridova.

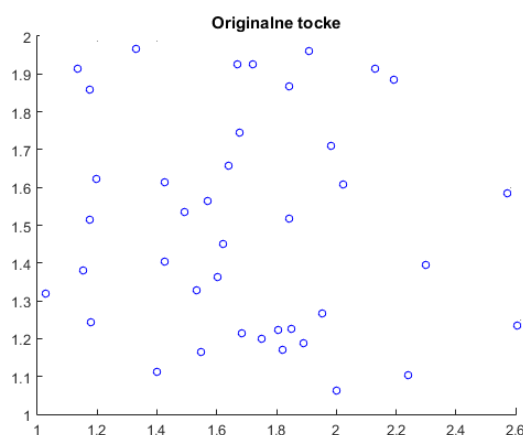
Primjer 1.2.1. Težine bridova između točaka su obrnuto proporcionalne njihovim udaljenosti, tj. elementi matrice W su: $w(i, j) = w(v_i, v_j) = \begin{cases} 0 & i = j \\ \frac{1}{\|v_i - v_j\|_2^2} & i \neq j. \end{cases}$



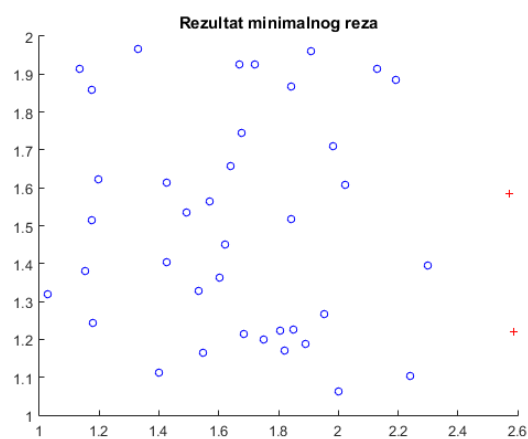
Slika 1.1: Simulirane točke.



Slika 1.2: Minimalni rez našao je dobro rješenje.



Slika 1.3: Simulirane točke.



Slika 1.4: Minimalni rez odvojio je dvije točke od ostatka, što nije dobra particija.

Kako bismo dobili bolje rješenje, razmotrimo drugačiju funkciju koja u obzir uzima kardinalitet skupova A i B, tj. broj vrhova u skupu A odnosno B ovdje u oznakama $|A|$ i $|B|$. Ovu funkciju zovemo **razmjerni rez particije**, a definiramo sa:

$$Rrez(A, B) = \frac{rez(A, B)}{|A|} + \frac{rez(A, B)}{|B|}. \quad (1.2)$$

Malo drugačiji način particioniranja predložen je u radu Shi i Malik [2] i to korištenjem normaliziranog reza particije koji se umjesto na kardinalitetu skupova particija bazira na težinama bridova te se je pokazao boljim izborom od razmjernog reza.

Ukupnu povezanost vrhova iz skupa $A \subset \mathbb{V}$ sa svim vrhovima grafa definiramo kao **vez** skupa A i $\mathbb{V}$:

$$vez(A, \mathbb{V}) = \sum_{i \in A, j \in \mathbb{V}} w_{ij}. \quad (1.3)$$

Normalizirani rez particije je sada definiran na sljedeći način:

$$Nrez(A, B) = \frac{rez(A, B)}{vez(A, \mathbb{V})} + \frac{rez(A, B)}{vez(B, \mathbb{V})}. \quad (1.4)$$

Uz $Nrez(\cdot, \cdot)$ između dvije grupe, možemo definirati i ukupnu **normaliziranu povezanost** vrhova unutar grupe, odnosno $Nvez$:

$$Nvez(A, B) = \frac{vez(A, A)}{vez(A, \mathbb{V})} + \frac{vez(B, B)}{vez(B, \mathbb{V})}, \quad (1.5)$$

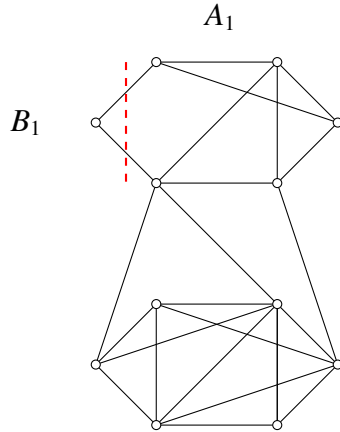
gdje $vez(A, A)$, $vez(B, B)$ označavaju ukupnu povezanost vrhova unutar A i unutar B redom.

Posljednje dvije funkcije su međusobno povezane što nam pokazuje sljedeći račun:

$$\begin{aligned} Nrez(A, B) &= \frac{rez(A, B)}{vez(A, \mathbb{V})} + \frac{rez(A, B)}{vez(B, \mathbb{V})} \\ &= \frac{vez(A, \mathbb{V}) - vez(A, A)}{vez(A, \mathbb{V})} + \frac{vez(B, \mathbb{V}) - vez(B, B)}{vez(B, \mathbb{V})} \\ &= 2 - \left(\frac{vez(A, A)}{vez(A, \mathbb{V})} + \frac{vez(B, B)}{vez(B, \mathbb{V})} \right) \\ &= 2 - Nvez(A, B). \end{aligned} \quad (1.6)$$

Algoritam klasteriranja želi minimizirati vezu među vrhovima različitih skupova particije, a maksimizirati vezu unutar vrhova istog skupa pa nam gornji račun pokazuje kako su ti postupci zapravo ekvivalentni i postižu se istovremeno, tj. normalizirani rez posjeduje svojstvo dualnosti.

Primjer 1.2.2. Usporedimo sada definirane funkcije particioniranja na primjeru grafova u kojima zbog jednostavnosti uzimamo da su težine svih bridova jednake 1.



$$\text{Rez}(A_1, B_1) = 2$$

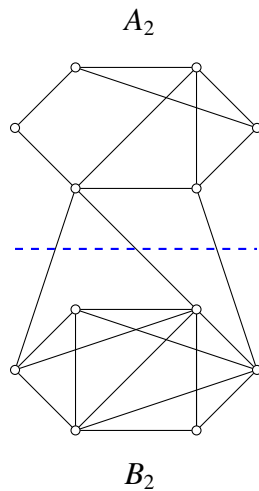
$$\text{Vez}(A_1, \mathbb{V}) = 2$$

$$\text{Vez}(B_1, \mathbb{V}) = 25$$

$$\text{Rrez}(A_1, B_1) = \frac{2}{1} + \frac{2}{11} = 2.18$$

$$\text{Nrez}(A_1, B_1) = \frac{2}{2} + \frac{2}{25} = 1.08$$

$$\text{Nvez}(A_1, B_1) = 2 - \text{Nrez}(A_1, B_1) = 0.92$$



$$\text{Rez}(A_2, B_2) = 3$$

$$\text{Vez}(A_2, \mathbb{V}) = 12$$

$$\text{Vez}(B_2, \mathbb{V}) = 15$$

$$\text{Rrez}(A_2, B_2) = \frac{3}{6} + \frac{3}{6} = 1$$

$$\text{Nrez}(A_2, B_2) = \frac{3}{12} + \frac{3}{15} = 0.45$$

$$\text{Nvez}(A_2, B_2) = 2 - \text{Nrez}(A_2, B_2) = 1.55$$

Vidimo kako je manji Rez odvojio jedan samostalni vrh i doveo do lošijeg rješenja particioniranja. Problem je riješilo uvođenje razmjernog i normaliziranog reza. Primijetimo i kako se uz minimizaciju Nreza, maksimizirao Nvez što je upravo ono što želimo postići.

Od svih promatranih rezova particija, svojstvo dualnosti ima samo normalizirani rez. Osim toga, budući da on u obzir uzima ukupnu težinu bridova koji povezuju particiju sa ostatkom skupa, problem odvajanja vrhova kojeg smo vidjeli u Primjeru 1.2.1 kod funkcije reza particije je riješen. Zbog svega navedenog algoritam spektralnog klasteriranja u ovom radu bazirat će se na normaliziranom rezu kao načinu particioniranja grafa.

NP-težak optimizacijski problem

Teorem 1.2.3. (*Papadimitrou 97*) *Problem Nreza grafa je NP-težak.*

Dakle, minimiziranje Nreza, odnosno maksimiziranje Nveza je NP težak optimizacijski problem, tj. nije riješiv egzaktnim algoritmom u polinomijalnom vremenu. Zbog toga ćemo problem relaksirati na realnu domenu te u narednim poglavljima pokazati kako se spektralnom analizom pronalazi aproksimativno diskretno rješenje. Da bismo to mogli, moramo uvesti pojmove poput Laplacijana grafa te navesti važne rezultate iz teorije varijacijske karakterizacije spektra simetrične matrice W .

Poglavlje 2

Laplacijan grafa

2.1 Pojam i svojstva

Definicija 2.1.1. Laplacijan težinskog grafa $\mathcal{G}$ definira se sa $L = D - W$, odnosno:

$$l_{ij} = \begin{cases} d_i & i = j \\ -w_{ij} & i \neq j \end{cases} \quad (2.1)$$

gdje su W i D pripadne matrice susjedstva, odnosno stupnjeva grafa $\mathcal{G}$.

Propozicija 2.1.2. Svojstva Laplacijana su:

- (1) Za svaki vektor $\mathbf{x} \in \mathbb{R}^n$ vrijedi: $\mathbf{x}^T L \mathbf{x} = \frac{1}{2} \sum_{i,j=1}^n w_{ij} (x_i - x_j)^2$
- (2) L je simetrična i pozitivno semi-definitna
- (3) Svojstveni vektor uz svojstvenu vrijednost $\lambda_1 = 0$ je konstantni vektor $\mathbb{1} = [1, \dots, 1]^T$
- (4) L ima n nenegativnih, realnih svojstvenih vrijednosti, tj. $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$.

Dokaz. (1) Iz definicije Laplacijana i matrice $D = \text{Diag}(d_i)$ imamo:

$$\begin{aligned} \mathbf{x}^T L \mathbf{x} &= \mathbf{x}^T D \mathbf{x} - \mathbf{x}^T W \mathbf{x} = \sum_{i=1}^n d_i x_i^2 - \sum_{i,j=1}^n x_i x_j w_{ij} \\ &= \frac{1}{2} \left(\sum_{i=1}^n d_i x_i^2 - 2 \sum_{i,j=1}^n x_i x_j w_{ij} + \sum_{i=1}^n d_i x_i^2 \right) = \frac{1}{2} \sum_{i,j=1}^n w_{ij} (x_i - x_j)^2, \quad \mathbf{x} = (x_1, \dots, x_n). \end{aligned}$$

(2) Simetričnost Laplacijana slijedi direktno iz simetričnosti matrica D i W , a pozitivna semidefinitnost iz (1): $\mathbf{x}^T L \mathbf{x} \geq 0$, $\mathbf{x} \in \mathbb{R}^n$.

(3) Uvrstimo vektor $\mathbb{1}$ u (1): $\mathbb{1}^T L \mathbb{1} = \frac{1}{2} \sum_{i,j=1}^n w_{ij}(1-1)^2 = 0$, pa direktno slijedi da je 0 najmanja svojstvena vrijednost.

(4) Znamo da za svojstvene vrijednosti pozitivno semi-definitne matrice vrijedi $\lambda_i \geq 0, i = 1, \dots, n$. Tvrdnja slijedi iz (1) - (3). □

Definicija 2.1.3. *Težinski normalizirani Laplacijan L_{norm} definira se kao $L_{norm} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$, tj.:*

$$l_{ij}^{norm} = \begin{cases} 1 - \frac{w_{jj}}{d_j} & i = j \\ -\frac{w_{ij}}{\sqrt{d_i d_j}} & i \neq j. \end{cases} \quad (2.2)$$

L_{norm} je simetrična matrica, a u literaturi se uz nju često veže matrica L_{rw} vezana uz slučajnu šetnju. Ona se definira kao $L_{rw} = D^{-1} L$ i vrijedi sljedeće

$$L_{norm} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = D^{\frac{1}{2}} D^{-1} L D^{-\frac{1}{2}} = D^{\frac{1}{2}} L_{rw} D^{-\frac{1}{2}} \quad (2.3)$$

Propozicija 2.1.4. *Za svaki vektor $\mathbf{x} \in \mathbb{R}^n$ vrijedi*

$$\mathbf{x}^T L_{norm} \mathbf{x} = \frac{1}{2} \sum_{i,j=1}^n w_{ij} \left(\frac{x_i}{\sqrt{d_i}} - \frac{x_j}{\sqrt{d_j}} \right)^2$$

Dokaz. Dokaz je analogan dokazu svojstva (1) Propozicije 2.1.2. □

Propozicija 2.1.5. *L_{rw} i L_{norm} povezane su na sljedeće načine:*

- (1) λ je svojstvena vrijednost od $L_{rw} = D^{-1} L$ sa svojstvenim vektorom v ako i samo ako je λ svojstvena vrijednost od L_{norm} sa pripadnim svojstvenim vektorom $w = D^{\frac{1}{2}} v$.
- (2) λ je svojstvena vrijednost od $L_{rw} = D^{-1} L$ sa svojstvenim vektorom v ako i samo ako je λ rješenje generaliziranog svojstvenog problema $Lv = \lambda Dv$.
- (3) 0 je svojstvena vrijednost od L_{rw} sa svojstvenim vektorom $\mathbb{1}$. 0 je svojstvena vrijednost od L_{norm} sa svojstvenim vektorom $D^{\frac{1}{2}} \mathbb{1}$.

Dokaz. (1) Pomnožimo izraz $L_{norm}w = \lambda w$ sa $D^{-\frac{1}{2}}$ s lijeva, iskoristimo jednakost (2.3) te supstitucijom $v = D^{-\frac{1}{2}}w$ dobijemo:

$$\begin{aligned} D^{-\frac{1}{2}}L_{norm}w &= D^{-\frac{1}{2}}\lambda w \\ D^{-\frac{1}{2}}D^{\frac{1}{2}}L_{rw}D^{-\frac{1}{2}}w &= D^{-\frac{1}{2}}\lambda w \\ L_{rw}D^{-\frac{1}{2}}w &= D^{-\frac{1}{2}}\lambda w \\ L_{rw}v &= \lambda v \end{aligned}$$

(2) Slijedi množenjem $L_{rw}v = \lambda v$ sa D s lijeva i činjenice $L_{rw} = D^{-1}L$.

(3) Prva tvrdnja slijedi iz činjenice da je 0 svojstvena vrijednost od L sa svojstvenim vektorom $\mathbb{1}$, pa onda i $L_{rw}\mathbb{1} = 0$. Druga tvrdnja slijedi iz (1).

□

Primjer 2.1.6. *Primjer Laplacijana i normaliziranog Laplacijana za graf iz 1.1.1.*

$$L = \begin{bmatrix} 9 & -4 & -5 & 0 & 0 & 0 & 0 \\ -4 & 10 & -2 & -1 & -3 & 0 & 0 \\ -5 & -2 & 13 & 0 & -6 & 0 & 0 \\ 0 & -1 & 0 & 7 & -6 & 0 & 0 \\ 0 & -3 & -6 & 0 & 9 & 0 & 0 \\ 0 & 0 & 0 & -5 & 0 & 9 & -4 \\ 0 & 0 & 0 & 0 & 0 & -4 & 4 \end{bmatrix}, \quad L_{norm} = \begin{bmatrix} 1 & -0.42 & -0.46 & 0 & 0 & 0 & 0 \\ -0.42 & 1 & -0.17 & -0.12 & -0.32 & 0 & 0 \\ -0.46 & -0.17 & 1 & 0 & -0.55 & 0 & 0 \\ 0 & -0.12 & 0 & 1 & -0.76 & 0 & 0 \\ 0 & -0.32 & -0.55 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -0.76 & 0 & 1 & -0.67 \\ 0 & 0 & 0 & 0 & 0 & -0.67 & 1 \end{bmatrix}$$

2.2 Povezane komponente grafa i Laplacijan

Definicija 2.2.1. *Povezana komponenta grafa je maksimalni povezan podgraf od $\mathcal{G}$, tj. povezani podgraf koji nije sadržan ni u jednom većem podgrafu.*

Ako graf ima samo jednu komponentu povezanosti, onda je povezan (inače je nepovezan).

Propozicija 2.2.2. *Neka je $\mathcal{G}$ težinski neusmjereni graf i L Laplacijan grafa sa svojstvenim vrijednostima λ_i , $i = 1, \dots, n$.*

Kratnost k svojstvene vrijednosti 0 jednaka je broju povezanih komponenti $A_1, \dots, A_k$ grafa $\mathcal{G}$. Pripadni prostor razapet je s pridruženim svojstvenim vektorima $\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_k}$, gdje su posljednji vektori definirani kao $\mathbb{1}_{A_l}(i, 1) = 1_{\{i \in A_l\}}$, $i = 1, \dots, n$, $l = 1, \dots, k$.

Dokaz. Neka je naprije graf povezan, odnosno $k=1$.

Neka je $\mathbf{x}$ svojstveni vektor pridružen svojstvenoj vrijednosti 0. Iz Propozicije 2.1.2 (1) znamo da:

$$0 = \mathbf{x}^T L \mathbf{x} = \sum_{i,j=1}^n w_{ij} (x_i - x_j)^2.$$

Ako su dva vrha v_i, v_j povezana, tj. $w_{ij} > 0$, gornja suma će postati 0 ako $(x_i - x_j) = 0$, tj. $x_i = x_j$. Dakle, $\mathbf{x}$ mora biti konstantan za sve međusobno povezane vrhove grafa. Općenito, svi vrhovi povezane komponente neusmjerenog grafa mogu se međusobno povezati nekim putom, pa slijedi da je $\mathbf{x}$ konstantan na cijeloj povezanoj komponenti. Dakle, ako se graf sastoji od jedne povezane komponente, imamo jedan svojstveni vektor $\mathbf{x} = \mathbb{1}$ i pripadnu svojstvenu vrijednost 0, što je očito indikator povezane komponente.

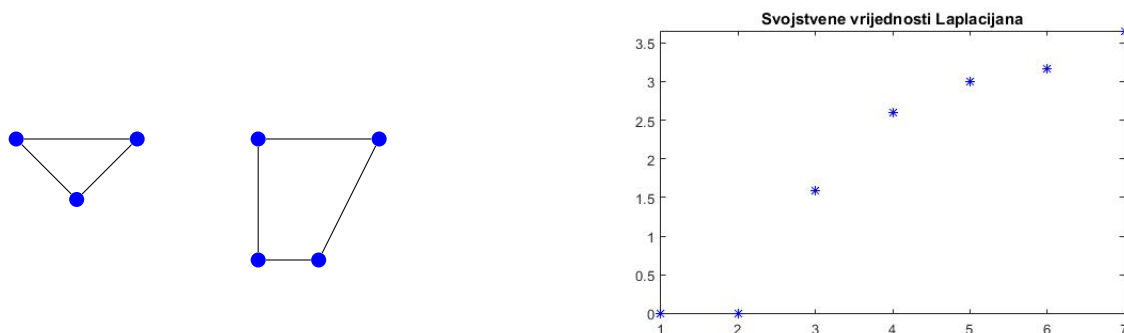
Neka se sada graf sastoji od k povezanih komponenti. Bez smanjenja općenitosti, možemo pretpostaviti da su vrhovi u redoslijedu po kojem pripadaju određenoj komponenti te su u tom slučaju W i L blok dijagonalne matrice:

$$L = \begin{bmatrix} L_1 & & & \\ & L_2 & & \\ & & \ddots & \\ & & & L_k \end{bmatrix}$$

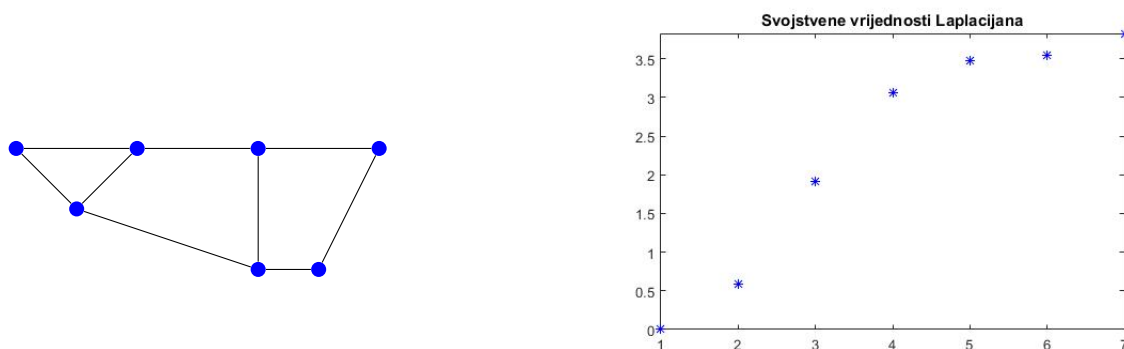
Svaka od L_i je Laplacijan matrica pridružena i -toj povezanoj komponenti. Za spektar blok matrica općenito vrijedi da se on može napisati kao unija spektara njenih blokova, pa je i u ovom slučaju $\sigma(L) = \cup_i \sigma(L_i)$, a pripadni svojstveni vektori od L su svojstveni vektori od L_i sa nulama na pozicijama koje pripadaju drugim blokovima. Budući da su L_i Laplacijani, znamo iz Propozicije 2.1.2 da svaki L_i ima svojstvenu vrijednost 0 s kratnošću 1, a pripadni svojstveni vektor je konstantan vektor $\mathbb{1}$ na i -toj povezanoj komponenti. Dakle, matrica L ima toliko svojstvenih vrijednosti jednakih 0 koliko ima i povezanih komponenti u grafu, a pripadni svojstveni vektori su njihovi indikatori, tj. $\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_k}$.

□

Primjer 2.2.3. Dana su dva grafa kao na slikama. Prvi se sastoji od dvije povezane komponente ($k=2$), a drugi graf je povezan.



Slika 2.1: Kratnost nule kao svojstvene vrijednosti Laplacijana je u grafu sa 2 povezane komponente jednaka 2.



Slika 2.2: Kratnost 0 kao svojstvene vrijednosti Laplacijana u povezanom grafu je jednaka 1.

2.3 Normalizirana matrica susjedstva i Laplacijan

Definicija 2.3.1. *Normalizirana matricu susjedstva definirana je sa*

$$P = D^{-1}W.$$

Budući da je suma svakog reda ove matrice jednaka 1, P je stohastička matrica. Iz teorije Markovljevih lanaca znamo da je p_{ij} vjerojatnost prelaska sa vrha i na vrh j .

Svojtvene vrijednosti matrice P su $1 = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq -1$, a $y_1, \dots, y_n$ su pripadni svojstveni vektori. Prvi svojstveni vektor jednak je $y_1 = \mathbb{1}$.

Normalizirana matrica susjedstva povezana je s Laplacijanom grafa na način da je spektralni problem za P ($Px = \lambda x$) ekvivalentan spektralnom problemu Laplacijana što pokazuje sljedeća propozicija:

Propozicija 2.3.2. *Ako je λ svojstvena vrijednost od P sa svojstvenim vektorom x , onda su $1 - \lambda$ i x pripadna svojstvena vrijednost i vektor Laplacijana L_{rw} .*

Dokaz. Tvrdnja slijedi odmah iz $L_{rw} = D^{-1}L = I - D^{-1}W = I - P$:

$$Px = \lambda x$$

$$(I - L_{rw})x = \lambda x$$

$$L_{rw}x = (1 - \lambda)x.$$

□

Poglavlje 3

Varijacijska karakterizacija spektra

U ovom poglavlju bavimo se teorijom vezanom uz varijacijsku karakterizaciju spektra koja nam, kao što je već spomenuto, daje glavne alate za rješavanje optimizacijskog problema normaliziranog reza.

Navedimo najprije neke osnovne pojmove i teoreme:

Definicija 3.0.3. Matrica $A = (a_{ij}) \in M_n$ je Hermitska ako vrijedi $A = A^*$, gdje je $A^* = \overline{A}^T = (\overline{a_{ji}})$.

Definicija 3.0.4. Za vektor $x \neq 0$ i Hermitsku matricu $A \in \mathbb{C}^{n \times n}$ definiramo Rayleighov kvocijent sa

$$R(A, x) = R(x) = \frac{x^T A x}{x^T x}.$$

Napomena 3.0.5. Za Hermitsku matricu $A \in \mathbb{R}^{n \times n}$ vrijedi $A^* = \overline{A}^T = A^T$, tj. matrica je simetrična. Budući da je ćemo mi problem klasteriranja relaksirati na realnu domenu, od sada se fokusiramo isključivo na realne simetrične matrice.

Teorem 3.0.6. (Spektralni teorem)

Matrica A je realna simetrična ($A \in \mathbb{R}^{n \times n}$) ako i samo ako postoji realna ortogonalna matrica $U \in M_n$ i realna dijagonalna matrica $\Lambda \in M_n$ takva da je $A = U \Lambda U^T$.

3.1 Varijacijska karakterizacija svojstvenih vrijednosti realnih simetričnih matrica

O spektru realnih simetričnih matrica možemo reći mnogo više od toga da su pripadne svojstvene vrijednosti rješenja karakteristične jednačbe. Njih naime možemo prikazati kao rješenja nekih optimizacijskih problema što ćemo pokazati kroz sljedeće teoreme.

Neka su

$$\lambda_{\min} = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{n-1} \leq \lambda_n = \lambda_{\max} \quad (3.1)$$

svojstvene vrijednosti simetrične matrice $A \in \mathbb{R}^{n \times n}$.

Teorem 3.1.1. (Rayleigh-Ritz)

Neka je $A \in \mathbb{R}^{n \times n}$ simetrična i neka su njene svojstvene vrijednosti kao u (3.1). Tada vrijedi:

$$\begin{aligned} \lambda_1 x^T x &\leq x^T A x \leq \lambda_n x^T x, \quad \forall x \in \mathbb{R}^n \\ \lambda_{\max} &= \lambda_n = \max_{x \neq 0} \frac{x^T A x}{x^T x} = \max_{x^T x = 1} x^T A x \\ \lambda_{\min} &= \lambda_1 = \min_{x \neq 0} \frac{x^T A x}{x^T x} = \min_{x^T x = 1} x^T A x \end{aligned}$$

Dokaz. Po Teoremu 3.0.6 postoji ortogonalna matrica $U \in \mathbb{R}^{n \times n}$ i dijagonalna $\Lambda \in \mathbb{R}^{n \times n}$, $\Lambda = \text{Diag}(\lambda_1, \dots, \lambda_n)$ takva da je $A = U \Lambda U^T$. Za svaki $x \in \mathbb{R}^n$ imamo:

$$x^T A x = x^T (U \Lambda U^T) x = (U^T x)^T \Lambda (U^T x) = \sum_{i=1}^n \lambda_i |(U^T x)_i|^2.$$

Svaki $|(U^T x)_i|^2$ je nenegativan, pa imamo

$$\lambda_{\min} \sum_{i=1}^n |(U^T x)_i|^2 \leq x^T A x = \sum_{i=1}^n \lambda_i |(U^T x)_i|^2 \leq \lambda_{\max} \sum_{i=1}^n |(U^T x)_i|^2.$$

Budući da je U ortogonalna i čuva normu, vrijedi:

$$\sum_{i=1}^n |(U^T x)_i|^2 = \sum_{i=1}^n |x_i|^2 = x^T x.$$

Iz toga slijedi prva tvrdnja:

$$\lambda_1 x^T x = \lambda_{\min} x^T x \leq x^T A x \leq \lambda_{\max} x^T x = \lambda_n x^T x. \quad (3.2)$$

Ako je x svojstveni vektor uz svojstvenu vrijednost λ_1 , tada $x^T A x = x^T \lambda_1 x = \lambda_1 x^T x$ te slično za λ_n . Ako je $x \neq 0$, iz 3.2 imamo

$$\frac{x^T A x}{x^T x} \leq \lambda_n.$$

Jednakost se postiže ako je x svojstveni vektor uz svojstvenu vrijednost λ_n , odnosno

$$\max_{x \neq 0} \frac{x^T A x}{x^T x} = \lambda_n. \quad (3.3)$$

Konačno, ako je $x \neq 0$,

$$\frac{x^T Ax}{x^T x} = \left(\frac{x}{\sqrt{x^T x}} \right)^T A \left(\frac{x}{\sqrt{x^T x}} \right) \quad \text{i} \quad \left(\frac{x}{\sqrt{x^T x}} \right)^T \left(\frac{x}{\sqrt{x^T x}} \right) = 1,$$

pa je uvjet (3.3) ekvivalentan uvjetu

$$\max_{x^T x=1} x^T Ax = \lambda_n. \quad (3.4)$$

Analogno dobijemo i za λ_1 .

□

Posljednji teorem dao nam je varijacijsku karakterizaciju najveće i najmanje svojstvene vrijednosti matrice A . No željeli bismo okarakterizirati i druge svojstvene vrijednosti. U tu svrhu, pretpostavimo da je $U = [u_1, u_2, \dots, u_n]$ ortonormalna matrica svojstvenih vektora od A , te $A = U \Lambda U^T$. Neka je $x \in \mathbb{R}^n$ vektor koji je okomit na u_1 . Tada imamo

$$x^T Ax = \sum_{i=1}^n \lambda_i |(U^T x)_i|^2 = \sum_{i=1}^n \lambda_i |u_i^T x|^2 = \sum_{i=2}^n \lambda_i |u_i^T x|^2.$$

Oдавде i (3.1) slijedi

$$x^T Ax = \sum_{i=2}^n \lambda_i |u_i^T x|^2 \geq \lambda_2 \sum_{i=2}^n |u_i^T x|^2 = \lambda_2 \sum_{i=1}^n |(U^T x)_i|^2 = \lambda_2 x^T x.$$

Jednakost se postiže za $x = u_2$, odnosno kada je x drugi svojstveni vektor matrice A . Dakle, dobili smo

$$\min_{\substack{x \neq 0 \\ x \perp u_1}} \frac{x^T Ax}{x^T x} = \min_{\substack{x^T x=1 \\ x \perp u_1}} x^T Ax = \lambda_2. \quad (3.5)$$

Generalizaciju ovih razmatranja daje nam sljedeći korolar:

Korolar 3.1.2. *Neka je $A \in \mathbb{R}^{n \times n}$ simetrična realna matrica te $u_1, \dots, u_n$ njezini svojstveni vektori pridruženi svojstvenim vrijednostima (3.1). Tada vrijedi*

$$\min_{\substack{x \neq 0 \\ x \perp u_1, u_2, \dots, u_{k-1}}} \frac{x^T Ax}{x^T x} = \min_{\substack{x^T x=1 \\ x \perp u_1, u_2, \dots, u_{k-1}}} x^T Ax = \lambda_k, \quad k = 2, \dots, n, \quad (3.6)$$

$$\max_{\substack{x \neq 0 \\ x \perp u_n, u_{n-1}, \dots, u_{n-k+1}}} \frac{x^T Ax}{x^T x} = \max_{\substack{x^T x=1 \\ x \perp u_n, u_{n-1}, \dots, u_{n-k+1}}} x^T Ax = \lambda_k, \quad k = 1, \dots, n-1. \quad (3.7)$$

Tvrdnje gornjeg korolara u praktičnim primjenama zahtijevaju znanja o svojstvenim vektorima, što često i nije slučaj. Zato problem treba generalizirati na proizvoljne vektore $w \in \mathbb{R}^n$. Tada imamo:

$$\begin{aligned}
 \sup_{\substack{x^T x = 1 \\ x \perp w}} x^T A x &= \sup_{\substack{x^T x = 1 \\ x \perp w}} x^T U \Lambda U^T x = \sup_{\substack{x^T x = 1 \\ x \perp w}} \sum_{i=1}^n \lambda_i |(U^T x)_i|^2 \\
 &= \sup_{\substack{z^T z = 1 \\ x = Uz \perp w}} \sum_{i=1}^n \lambda_i |z_i|^2 = \sup_{\substack{z^T z = 1 \\ z \perp U^T w}} \sum_{i=1}^n \lambda_i |z_i|^2 \\
 &\geq \sup_{\substack{z^T z = 1 \\ z \perp U^T w \\ z_1 = z_2 = \dots = z_{n-2} = 0}} \sum_{i=1}^n \lambda_i |z_i|^2 \\
 &= \sup_{\substack{|z_{n-1}|^2 + |z_n|^2 = 1 \\ z \perp U^T w}} \lambda_{n-1} |z_{n-1}|^2 + \lambda_n |z_n|^2 \geq \lambda_{n-1}.
 \end{aligned}$$

U drugom redu stavimo $z = U^T x$ i iskoristimo činjenicu da je U ortogonalna matrica pa je $z^T z = 1$ ako i samo ako je $x^T x = 1$. Prva nejednakost slijedi iz svojstva supremuma, tj. restrikcije skupa na kojem gledamo supremum, a posljednja slijedi iz (3.1), tj. rastućih svojstvenih vrijednosti. Budući da je w bio proizvoljan, ali fiksni vektor, gornja nejednakost prelazi u

$$\inf_{w \in \mathbb{R}^n} \sup_{\substack{x^T x = 1 \\ x \perp w}} x^T A x \geq \lambda_{n-1},$$

a jednakost se postiže kada je $w = u_n$, odnosno n -ti svojstveni vektor. Osnovna ideja ovog razmatranja i uzimanja maksimuma, odnosno minimuma umjesto supremuma i infimuma leži u sljedećem teoremu:

Teorem 3.1.3. (Courant-Fisher) *Neka je $A \in \mathbb{R}^{n \times n}$ simetrična matrica sa svojstvenim vrijednostima $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ i neka je k prirodan broj $1 \leq k \leq n$. Tada vrijedi*

$$\min_{w_1, w_2, \dots, w_{n-k} \in \mathbb{R}^n} \max_{\substack{x \neq 0, x \in \mathbb{R}^n, \\ x \perp w_1, w_2, \dots, w_{n-k}}} \frac{x^T A x}{x^T x} = \lambda_k, \quad (3.8)$$

$$\max_{w_1, w_2, \dots, w_{k-1} \in \mathbb{R}^n} \min_{\substack{x \neq 0, x \in \mathbb{R}^n, \\ x \perp w_1, w_2, \dots, w_{k-1}}} \frac{x^T A x}{x^T x} = \lambda_k. \quad (3.9)$$

Napomena 3.1.4. *U slučaju $k = n$ u (3.8) ili $k = 1$ u (3.9), gornji teorem svodi se na Rayleigh-Ritzov teorem.*

Dokaz. Budući da su dokazi ekvivalentni, dokazat ćemo samo (3.8).

Po Spektralnom teoremu 3.0.6 postoji ortogonalna matrica U i dijagonalna $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ takva da je $A = U\Lambda U^T$. Neka je $1 < k \leq n$. Ako je $x \neq 0$, vrijedi

$$\frac{x^T A x}{x^T x} = \frac{(U^T x)^T \Lambda (U^T x)}{x^T x} = \frac{(U^T x)^T \Lambda (U^T x)}{(U^T x)^T (U^T x)}$$

i $\{U^T x : x \in \mathbb{R}^n, x \neq 0\} = \{y \in \mathbb{R}^n : y \neq 0\}$. Neka su $w_1, w_2, \dots, w_{n-k} \in \mathbb{R}^n$. Tada imamo:

$$\begin{aligned} \sup_{\substack{x \neq 0 \\ x \perp w_1, \dots, w_{n-k}}} \frac{x^T A x}{x^T x} &= \sup_{\substack{y \neq 0 \\ y \perp U^T w_1, \dots, U^T w_{n-k}}} \frac{y^T \Lambda y}{y^T y} = \sup_{\substack{y^T y = 1 \\ y \perp U^T w_1, \dots, U^T w_{n-k}}} \sum_{i=1}^n \lambda_i |y_i|^2 \\ &\geq \sup_{\substack{y^T y = 1 \\ z \perp U^T w_1, \dots, U^T w_{n-k} \\ y_1 = y_2 = \dots = y_{k-1} = 0}} \sum_{i=1}^n \lambda_i |y_i|^2 = \sup_{\substack{|y_k|^2 + |y_{k+1}|^2 + \dots + |y_n|^2 = 1 \\ y \perp U^T w_1, \dots, U^T w_{n-k}}} \sum_{i=k}^n \lambda_i |y_i|^2 \geq \lambda_k. \end{aligned}$$

Dakle,

$$\sup_{\substack{x \neq 0 \\ x \perp w_1, \dots, w_{n-k}}} \frac{x^T A x}{x^T x} \geq \lambda_k,$$

za proizvoljnih $n - k$ vektora $w_1, \dots, w_{n-k}$.

Iz Korolara 3.1.2 slijedi da jednakost vrijedi za $w_i = u_{n-i+1}$ gdje je $U = [u_1, \dots, u_n]$. Imamo:

$$\inf_{w_1, \dots, w_{n-k}} \sup_{\substack{x \neq 0 \\ x \perp w_1, \dots, w_{n-k}}} \frac{x^T A x}{x^T x} = \lambda_k,$$

odnosno:

$$\min_{w_1, \dots, w_{n-k}} \max_{\substack{x \neq 0 \\ x \perp w_1, \dots, w_{n-k}}} \frac{x^T A x}{x^T x} = \lambda_k.$$

□

Neka je $B \in \mathbb{R}^{n \times n}$ realna simetrična matrica i $B = U\Lambda U^T$, gdje su $U = [u_1 \dots u_n]$, $\Lambda = \text{diag}(\beta_1, \dots, \beta_n)$. Rang matrice B jednaka je broju svojstvenih vrijednosti različitih od 0. Ako je rang matrice B manji ili jednak r , možemo pretpostaviti da je $\beta_{r+1} = \dots = \beta_n = 0$. Ako je rang manji od r , tada su neke od $\beta_1, \dots, \beta_r$ jednake 0. Primijetimo da vrijedi:

$$B = U\Lambda U^T = \sum_{i=1}^r \beta_i u_i u_i^T.$$

Teorem 3.1.5. *Neka je $A \in \mathbb{R}^{n \times n}$ simetrična matrica i i cijeli broj takav da $1 \leq r \leq n$. Neka je A_r $r \times r$ matrica nastala brisanjem $n - r$ redaka i stupaca iz matrice A . Za svaki k takav da $1 \leq k \leq r$ vrijedi*

$$\lambda_k(A) \leq \lambda_k(A_r) \leq \lambda_{k+n-r}(A).$$

Dokaz. Neka je A_r $r \times r$ matrica nastala brisanjem redova $i_1, \dots, i_{n-r}$ i pripadajućih stupaca matrice A . Neka je $1 \leq k \leq r$. Iz Courant Fisherovog teorema 3.1.3 imamo:

$$\begin{aligned} \lambda_{k+n-r}(A) &= \min_{w_1, \dots, w_{r-k} \in \mathbb{R}^n} \max_{\substack{x \neq 0, x \in \mathbb{R}^n \\ x \perp w_1, \dots, w_{r-k}}} \frac{x^T A x}{x^T x} \\ &\geq \min_{w_1, \dots, w_{r-k} \in \mathbb{R}^n} \max_{\substack{x \neq 0, x \in \mathbb{R}^n \\ x \perp w_1, \dots, w_{r-k} \\ x \perp e_{i_1}, \dots, e_{i_{n-r}}}} \frac{x^T A x}{x^T x} \\ &= \min_{v_1, \dots, v_{r-k} \in \mathbb{R}^r} \max_{\substack{y \neq 0, y \in \mathbb{R}^r \\ y \perp v_1, \dots, v_{r-k}}} \frac{y^T A_r y}{y^T y} = \lambda_k(A_r), \end{aligned}$$

$$\begin{aligned} \lambda_k(A) &= \max_{w_1, \dots, w_{k-1} \in \mathbb{R}^n} \min_{\substack{x \neq 0, x \in \mathbb{R}^n \\ x \perp w_1, \dots, w_{k-1}}} \frac{x^T A x}{x^T x} \\ &\leq \max_{w_1, \dots, w_{k-1} \in \mathbb{R}^n} \min_{\substack{x \neq 0, x \in \mathbb{R}^n \\ x \perp w_1, \dots, w_{k-1} \\ x \perp e_{i_1}, \dots, e_{i_{n-r}}}} \frac{x^T A x}{x^T x} \\ &= \max_{v_1, \dots, v_{k-1} \in \mathbb{R}^r} \min_{\substack{y \neq 0, y \in \mathbb{R}^r \\ y \perp v_1, \dots, v_{k-1}}} \frac{y^T A_r y}{y^T y} = \lambda_k(A_r). \end{aligned}$$

□

Iz prethodnog teorema slijedi korolar poznat u literaturi kao **Poincarov separacijski teorem**.

Korolar 3.1.6. (*Poincare*)

Neka je $A \in \mathbb{R}^{n \times n}$ simetrična matrica i r cijeli broj takav da $1 \leq r \leq n$ i neka su $u_1, \dots, u_r \in \mathbb{R}^n$ ortonormalni vektori. Neka je $B_r = [u_i^T A u_j] \in \mathbb{R}^{r \times r}$. Ako su svojstvene vrijednosti od A i B_r kao u 3.1, imamo

$$\lambda_k(A) \leq \lambda_k(B_r) \leq \lambda_{k+n-r}(A), \quad k = 1, 2, \dots, r. \quad (3.10)$$

Simetrična matrica $B_r \in \mathbb{R}^{r \times r}$ može se zapisati kao $B_r = U^T A U$, gdje je $U \in \mathbb{R}^{n \times r}$ ortonormalna matrica sa r ortonormalnih stupaca. Sumiranjem nejednakoti (3.10) i činjenice da je $\text{tr}(B_r) = \lambda_1(B_r) + \dots + \lambda_r(B_r)$, dobivamo sljedeće korolare:

Korolar 3.1.7. Neka je $A \in \mathbb{R}^{n \times n}$ simetrična matrica, svojstvene vrijednosti kao u (3.1) i r cijeli broj takav da $1 \leq r \leq n$. Tada

$$\lambda_1(A) + \dots + \lambda_r(A) = \min_{U^T U = I_r} \text{tr}(U^T A U), \quad (3.11)$$

$$\lambda_{n-r+1}(A) + \dots + \lambda_n(A) = \max_{U^T U = I_r} \text{tr}(U^T A U), \quad U \in \mathbb{R}^{n \times r}. \quad (3.12)$$

Napomena 3.1.8. Formule (3.11) i (3.12) mogu se shvatiti kao generalizacije Rayleigh-Ritz teorema. Jednakost (3.11) vrijedi ako su stupci matrice U ortonormalni svojstveni vektori pridruženi najmanjih r svojstvenih vrijednosti matrice A te analogno u (3.12).

Poglavlje 4

Spektralna biparticija grafa

U ovom poglavlju pokazat ćemo kako se NP-težak optimizacijski problem minimizacije Nreza približno rješava relaksacijom i spektralnom analizom Laplacijana matrice grafa te korištenjem važnih teorema karakterizacije spektra iz Poglavlja 3.

4.1 Model

Neka su $V_1, V_2 \subset \mathbb{V}$, $V_1 \cap V_2 = \emptyset$, $V_1 \cup V_2 = \mathbb{V}$. Neka je $N = |\mathbb{V}|$. Definirat ćemo **particijsku matricu** $X = [X_1, X_2]$ dimenzija $N \times 2$ na sljedeći način:

$$X(i, 1) = 1_{\{i \in V_1\}}, \quad X(i, 2) = 1_{\{i \in V_2\}}, \quad i = 1, \dots, N,$$

gdje $1_{\{\cdot\}}$ predstavlja indikatorsku funkciju. Budući da svaki vrh zbog disjunktnosti pripada točno jednom od skupova V_1 i V_2 , stupci matrice X imat će jedinice na različitim mjestima, tj. vrijedi $X\mathbb{1}_2 = \mathbb{1}_N$. Dakle, stupci od X su međusobno okomiti.

Primjer 4.1.1. Za graf sa skupom vrhova $\{1, 2, 3, 4, 5\}$ i bipartcijom $\{\{1, 3\}, \{2, 4, 5\}\}$, particijska matrica X izgleda:

$$X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}.$$

4.2 Pronalaženje optimalnog reza za bipartciju

Formulacija funkcija particioniranja

Želimo particionirati vrhove grafa $\mathcal{G}$ na dva skupa V_1 i V_2 , odnosno minimizirati normalizirani rez između njih. Budući da su glavni alati spektralnog klasteriranja Laplacijan grafa, njegova svojstva te teoremi kojima smo se bavili u prijašnjim poglavljima, ovdje ćemo pokazati da se problem biparticioniranja svodi na rješavanje svojstvenog problema Laplacijana grafa $\mathcal{G}$. Kako bismo to mogli, preformulirajmo definicije Nreza i Nveza pomoću nove matrice particije X , matrica W , D te Laplacijana L .

Vrijedi sljedeće:

$$rez(V_1, V_2) = \sum_{i \in V_1, j \in V_2} w_{ij} = X_1^T D X_1 - X_1^T W X_1 = X_1^T L X_1, \quad (4.1)$$

$$vez(V_1, \mathbb{V}) = X_1^T D X_1, \quad (4.2)$$

$$vez(V_2, \mathbb{V}) = X_2^T D X_2, \quad (4.3)$$

$$vez(V_l, V_l) = X_l^T W X_l, \quad l = 1, 2. \quad (4.4)$$

Sada imamo:

$$\begin{aligned} Nrez(V_1, V_2) &= \frac{rez(V_1, V_2)}{vez(V_1, \mathbb{V})} + \frac{rez(V_1, V_2)}{vez(V_2, \mathbb{V})} \\ &= \frac{X_1^T L X_1}{X_1^T D X_1} + \frac{X_2^T L X_2}{X_2^T D X_2}, \end{aligned} \quad (4.5)$$

$$\begin{aligned} Nvez(V_1, V_2) &= \frac{vez(V_1, V_1)}{vez(V_1, \mathbb{V})} + \frac{vez(V_2, V_2)}{vez(V_2, \mathbb{V})} \\ &= \frac{X_1^T W X_1}{X_1^T D X_1} + \frac{X_2^T W X_2}{X_2^T D X_2}. \end{aligned} \quad (4.6)$$

Znamo iz (1.6) da su optimizacijski problemi

$$\mu([V_1, V_2]) = Nrez(V_1, V_2) \longrightarrow \min, \quad (4.7)$$

$$\varepsilon([V_1, V_2]) = Nvez(V_1, V_2) \longrightarrow \max \quad (4.8)$$

međusobno ekvivalentni pa ćemo se ovdje fokusirati na (4.7).

Primijetimo da je svaki od sumanada u (4.5) u formi Rayleighovog kvocijenta, odnosno oblika $\frac{\mathbf{y}^T L \mathbf{y}}{\mathbf{y}^T D \mathbf{y}}$, za neki $\mathbf{y} \in \{0, 1\}^N$.

Uzmemo li $\mathbf{z}$ takav da je $\mathbf{z} = D^{\frac{1}{2}} \mathbf{y}$, imamo:

$$\frac{\mathbf{y}^T L \mathbf{y}}{\mathbf{y}^T D \mathbf{y}} = \frac{(D^{-\frac{1}{2}} \mathbf{z})^T L D^{-\frac{1}{2}} \mathbf{z}}{(D^{-\frac{1}{2}} \mathbf{z})^T D (D^{-\frac{1}{2}} \mathbf{z})} = \frac{\mathbf{z}^T L_{norm} \mathbf{z}}{\mathbf{z}^T \mathbf{z}}. \quad (4.9)$$

Sada vidimo da se minimizacija kvocijenata u (4.5) svela na optimizacijski problem minimiziranja Rayleighovog kvocijenta sa $\mathbf{y}$ relaksiranim na realnu domenu, pa nam iz rezultata Poglavlja 3 o varijacijskoj karakterizaciji spektra slijedi da za nalaženje minimuma zapravo trebamo riješiti spektralni problem simetrične matrice iz brojnika Rayleighovog kvocijenta, a to je u ovom slučaju matrica normaliziranog Laplacijana L_{norm} .

Rješenje

Spektralni problem za normalizirani Laplacijan glasi

$$L_{norm} \mathbf{z} = \lambda \mathbf{z},$$

pa iz Propozicije 2.1.5 slijedi $D^{-1} L \mathbf{y} = \lambda \mathbf{y}$ uz $\mathbf{y} = D^{-\frac{1}{2}} \mathbf{z}$.

Time dobivamo generalizirani spektralni problem za Laplacijan L , odnosno:

$$(D - W) \mathbf{y} = \lambda D \mathbf{y}, \quad \mathbf{y} = D^{-\frac{1}{2}} \mathbf{z}. \quad (4.10)$$

Iz Rayleigh Ritzovog teorema 3.1.1, Korolara 3.1.2, činjenice da su svojstveni vektori Laplacijana međusobno okomiti slijedi da će $\frac{\mathbf{z}^T L_{norm} \mathbf{z}}{\mathbf{z}^T \mathbf{z}} = \frac{\mathbf{y}^T L \mathbf{y}}{\mathbf{y}^T D \mathbf{y}}$ biti minimizirani drugim najmanjim svojstvenim vektorom, odnosno njemu pripadajućoj svojstvenoj vrijednosti λ_2 .

Dakle, uz $\mathbf{y} = D^{-\frac{1}{2}} \mathbf{z}$ imamo:

$$\lambda_2 = \min_{\mathbf{z} \perp \mathbf{z}_1} \frac{\mathbf{z}^T L_{norm} \mathbf{z}}{\mathbf{z}^T \mathbf{z}} = \min_{\mathbf{z} \perp D^{\frac{1}{2}} \mathbf{1}} \frac{\mathbf{z}^T L_{norm} \mathbf{z}}{\mathbf{z}^T \mathbf{z}}, \quad (4.11)$$

$$\lambda_2 = \min_{\mathbf{y} \perp \mathbf{y}_1} \frac{\mathbf{y}^T L \mathbf{y}}{\mathbf{y}^T D \mathbf{y}} = \min_{\mathbf{y} \perp D \mathbf{1}} \frac{\mathbf{y}^T L \mathbf{y}}{\mathbf{y}^T D \mathbf{y}}. \quad (4.12)$$

Poglavlje 5

Spektralna K-particija grafa

U ovom poglavlju problem particioniranja grafa generalizirat ćemo na način da skup vrhova $\mathbb{V}$ želimo podijeliti u K disjunktih grupa te ćemo u te svrhe definirati funkcije K-particioniranja i primijeniti sličan princip kao i kod biparticije. Međutim, do krajnjeg diskretnog rješenja doći ćemo na malo drugačiji način, korištenjem spektralne rotacije.

5.1 Model

Kao i prije, promatramo graf $\mathcal{G} = (\mathbb{V}, \mathbb{E})$, pripadajuću mu matricu susjedstva W te matricu stupnjeva D . Neka je $\mathbb{V} = [N] = \{1, 2, \dots, N\}$ skup elemenata koje grupiramo.

Da bismo grupirali N elemenata u K grupa, particioniramo skup vrhova $\mathbb{V}$ na skupove $V_1, \dots, V_K$ takve da $\mathbb{V} = \cup_{l=1}^K V_l$ i $V_k \cap V_l = \emptyset$, $\forall k \neq l$. Ovaj postupak zovemo K-particioniranje grafa i zapisujemo $\Gamma_{\mathbb{V}}^K = \{V_1, \dots, V_K\}$.

5.2 Funkcije K-particioniranja

K-normalizirani rez K-particioniranja grafa definiramo s:

$$K\text{-}N\text{rez}(\Gamma_{\mathbb{V}}^K) = \frac{1}{K} \sum_{l=1}^K \frac{\text{vez}(V_l, V \setminus V_l)}{\text{vez}(V_l, V)}. \quad (5.1)$$

K-normalizirani vez K-particinoiranja grafa definiramo s:

$$K\text{-}N\text{vez}(\Gamma_{\mathbb{V}}^K) = \frac{1}{K} \sum_{l=1}^K \frac{\text{vez}(V_l, V_l)}{\text{vez}(V_l, V)}, \quad (5.2)$$

gdje je vez definiran kao u (1.3). Kao i prije, postoji veza između K-Nreza i K-Nveza pa se na taj način minimizacija jednog, odnosno maksimizacija drugog ponovno postižu istovremeno:

$$K\text{-Nvez}(\Gamma_{\mathbb{V}}^K) + K\text{-Nrez}(\Gamma_{\mathbb{V}}^K) = 1. \quad (5.3)$$

5.3 Matrična reprezentacija problema

Kao i u slučaju biparticije, za prikaz $\Gamma_{\mathbb{V}}^K$ definirat ćemo particijsku matricu $X = [X_1, \dots, X_K]$ dimenzija $N \times K$ takvu da:

$$X(i, l) = 1_{\{i \in \mathbb{V}_l\}}, \quad i \in \mathbb{V}, l \in [K].$$

Stupci matrice X su međusobno okomiti budući da svaki vrh pripada točno jednoj particiji. Dakle, vrijedi $X\mathbb{1}_K = \mathbb{1}_N$, gdje $\mathbb{1}_d$ predstavlja vektor jedinica dimenzija $d \times 1$.

Sada možemo funkcije K-particioniranja zapisati u matričnom obliku korištenjem matrice susjedstva, stupnjeva i Laplacijana grafa.

$$\text{vez}(V_l, V_l) = X_l^T W X_l, \quad (5.4)$$

$$\text{vez}(V_l, \mathbb{V}) = X_l^T D X_l, \quad (5.5)$$

$$\begin{aligned} \text{vez}(V_l, V \setminus V_l) &= X_l^T D X_l - X_l^T W X_l \\ &= X_l^T L X_l, \quad l \in \{1, \dots, K\} \end{aligned} \quad (5.6)$$

Ciljne funkcije su:

$$\mu(\Gamma_{\mathbb{V}}^K) = K\text{-Nrez}(\Gamma_{\mathbb{V}}^K) = \frac{1}{K} \sum_{l=1}^K \frac{X_l^T L X_l}{X_l^T D X_l} \longrightarrow \min, \quad (5.7)$$

$$\varepsilon(\Gamma_{\mathbb{V}}^K) = K\text{-Nvez}(\Gamma_{\mathbb{V}}^K) = \frac{1}{K} \sum_{l=1}^K \frac{X_l^T W X_l}{X_l^T D X_l} \longrightarrow \max \quad (5.8)$$

Minimizacija $\mu(\Gamma_{\mathbb{V}}^K)$ i maksimizacija $\varepsilon(\Gamma_{\mathbb{V}}^K)$ su po (5.3) ekvivalentne, odnosno postižu se istovremeno, pa ćemo našim optimizacijskim problemom smatrati (5.8).

Optimizacijski problem po X (OPX)

$$\begin{aligned} \text{maksimiziraj} \quad & \varepsilon(\Gamma_{\mathbb{V}}^K) = \frac{1}{K} \sum_{l=1}^K \frac{X_l^T W X_l}{X_l^T D X_l} \\ \text{uz} \quad & X \in \{0, 1\}^{N \times K}, \quad X\mathbb{1}_K = \mathbb{1}_N. \end{aligned} \quad (5.9)$$

Kao i za slučaj $K=2$ (Teorem 1.2.3) i ovaj problem je NP-težak. Kako bi se riješio barem približno, potrebno je problem relaksirati na realnu domenu gdje nam na raspolaganju stoje korisni teoremi varijacijske karakterizacije spektra iz Poglavlja 3. Primijetimo da su ovdje, kao i u slučaju biparticije, sumandi u (5.7) i (5.8) oblika $\frac{\mathbf{x}^T L \mathbf{x}}{\mathbf{x}^T D \mathbf{x}}$, odnosno $\frac{\mathbf{x}^T W \mathbf{x}}{\mathbf{x}^T D \mathbf{x}}$ za neki $\mathbf{x} \in \{0, 1\}^N$.

Da bismo sumande sveli na poznati oblik za optimizaciju Rayleighovog kvocijenta, postepeno ćemo definirati **skaliranu particijsku matricu** $Z = [Z_1, \dots, Z_K]$. Neka je:

$$Z(i, l) = \begin{cases} a_l, & v_i \in V_l \\ b_l, & v_i \notin V_l, \end{cases} \quad l = 1, \dots, K, \quad i = 1, \dots, N, \quad (5.10)$$

gdje su a_l, b_l različiti realni brojevi.

Želimo a_l, b_l definirane tako da vrijedi

$$\frac{\text{rez}(V_l, \mathbb{V} \setminus V_l)}{\text{vez}(V_l, \mathbb{V})} = \frac{Z_l^T L Z_l}{Z_l^T D Z_l}, \quad l = 1, \dots, K. \quad (5.11)$$

Označimo sa $d = \mathbb{1}^T D \mathbb{1}$ zbroj svih stupnjeva grafa $\mathcal{G}$ te označimo sa $\alpha_l = \text{vez}(V_l, \mathbb{V})$ za $l = 1, \dots, K$. Tada zbog svojstva Laplacijana iz Propozicije 2.1.2 (1) te iz definicije reza 1.1 za $l = 1, \dots, K$ vrijedi

$$Z_l^T L Z_l = (a_l - b_l)^2 \text{rez}(V_l, \mathbb{V} \setminus V_l), \quad (5.12)$$

$$Z_l^T D Z_l = \alpha_l a_l^2 + (d - \alpha_l) b_l^2. \quad (5.13)$$

Uvrštavanjem u (5.11) dobivamo sljedeći uvjet:

$$2\alpha_l b_l (b_l - a_l) = d b_l^2.$$

Za $b_l \neq 0$ imamo

$$a_l = \frac{2\alpha_l - d}{2\alpha_l} b_l,$$

a za $b_l = 0$ uvjet je trivijalno ispunjen. Zbog jednostavnosti, uzet ćemo da vrijedi $b_l = 0$, $l = 1, \dots, K$. Sada u (5.13) imamo

$$Z_l^T D Z_l = \alpha_l a_l^2$$

pa ima smisla uzeti da je

$$a_l = \frac{1}{\sqrt{\alpha_l}} = \frac{1}{\sqrt{\text{vez}(V_l, \mathbb{V})}}, \quad l = 1, \dots, K.$$

Dakle, stupci skalirane particijske matrice Z sadrže iste stupce kao i početna particijska matrica X , ali skalirane sa kvadratnim korijenom stupnjeva particija $V_1, \dots, V_K$. Zato matricu Z možemo zapisati kao

$$Z = X(X^T DX)^{-\frac{1}{2}}. \quad (5.14)$$

Definirajmo još i **normalizacijsku funkciju** f koja $X \mapsto Z$, tj.

$$Z = f(X) = X(X^T DX)^{-\frac{1}{2}}. \quad (5.15)$$

Tada je njen inverz f^{-1} :

$$X = f^{-1}(Z) = \text{Diag}(\text{diag}^{-\frac{1}{2}}(ZZ^T))Z, \quad (5.16)$$

gdje diag vrati vektor stupac dijagonalnih elemenata matrice.

Ako na primjer uzmemo da su redovi matrice Z koordinate točaka u K -dimenzionalnom prostoru, onda inverz f^{-1} normalizira njihove norme tako da sve točke leže na jediničnoj sferi sa središtem u ishodištu.

Budući da je funkcija f samo skalirala stupce particijske matrice X , matrica Z također ima međusobno okomite stupce.

Osim toga vrijedi:

$$Z^T DZ = I_K.$$

Sada problem (5.9) možemo formulirati na sljedeći način:

Optimizacijski problem po Z (OPZ)

$$\begin{aligned} &\text{maksimiziraj} && \varepsilon(Z) = \frac{1}{K} \text{tr}(Z^T WZ) \\ &\text{uz} && Z^T DZ = I_K \end{aligned} \quad (5.17)$$

Napomena 5.3.1. Problemi (5.9) i (5.17) su ekvivalentni u smislu da je $(X_1, \dots, X_K)$ rješenje od OPX (5.9) ako i samo ako je $((Z_1^T DZ_1)^{-\frac{1}{2}} Z_1, \dots, (Z_K^T DZ_K)^{-\frac{1}{2}} Z_K)$ rješenje od OPZ (5.17).

Relaksiranje matrice Z na realnu domenu sada nam je omogućilo da za naš optimizacijski problem koristimo poznate algoritme i teoreme iz Poglavlja 3.

5.4 Nalaženje optimalnog relaksiranog rješenja

Propozicija 5.4.1. (Ortonormalna invarijantnost)

Neka je R ortogonalna matrica dimenzija $K \times K$. Ako je Z rješenje OPZ (5.17), tada je rješenje i skup $\{ZR : R^T R = I_K\}$ te vrijedi $\varepsilon(ZR) = \varepsilon(Z)$.

Drugim riječima, proizvoljna rotacija čuva rješenje problema OPZ.

Neka je sada P normalizirana matrica susjedstva definirana u 2.3.1, odnosno $P = D^{-1}W$. Pokazat ćemo da je za pronalaženje rješenja problema (5.17) potrebno pronaći svojstvene vektore normaliziranog Laplacijana L_{norm} , točnije s njime povezanog L_{rw} .

Neka su $V = [V_1, V_2, \dots, V_N]$ i $S = \text{Diag}(s_1, \dots, s_N)$ matrice svojstvenih vektora i svojstvenih vrijednosti matrice P . Znamo da vrijedi $s_1 \geq s_2 \geq \dots \geq s_N$. Iz Propozicije 2.3.2 slijedi da su $\lambda_1 = 1 - s_1 \leq \lambda_2 = 1 - s_2 \leq \dots \leq 1 - s_N = \lambda_N$ svojstvene vrijednosti od $L_{rw} = I - D^{-1}W$ sa svojstvenim vektorima V .

Sada iz Propozicije 2.1.5 (1) slijedi da su $1 - \lambda_i = s_i$, $i = 1, \dots, N$ svojstvene vrijednosti od $D^{-\frac{1}{2}}WD^{-\frac{1}{2}} = I - L_{norm}$ sa pripadajućim svojstvenim vektorima $\bar{V} = D^{\frac{1}{2}}V$.

Dakle, P i $D^{-\frac{1}{2}}WD^{-\frac{1}{2}}$ imaju iste svojstvene vrijednosti, no budući je $D^{-\frac{1}{2}}WD^{-\frac{1}{2}}$ simetrična, matrica njenih svojstvenih vektora, $\bar{V} = [\bar{V}_1, \dots, \bar{V}_N]$ je ortonormalna (Dodatak, Teorem 10.0.7).

Vrijedi

$$\begin{aligned} V &= D^{-\frac{1}{2}}\bar{V}, \\ D^{-\frac{1}{2}}WD^{-\frac{1}{2}} &= \bar{V}S, \quad \bar{V}^T\bar{V} = I_N. \end{aligned} \quad (5.18)$$

Sada nam korolar 3.1.7 daje rješenje problema OPZ 5.17, odnosno vrijedi:

$$\begin{aligned} \varepsilon(Z^*) &= \max_{Z^T DZ = I_K} \frac{1}{K} \text{tr}(Z^T WZ) \stackrel{??}{=} \frac{1}{K} (s_1 + \dots + s_K) = \frac{1}{K} \text{tr}(\Lambda^*), \\ Z^* &= [V_1, \dots, V_K], \\ \Lambda^* &= \text{Diag}([s_1, \dots, s_K]). \end{aligned} \quad (5.19)$$

Vidimo da rješenje OPZ nije jedinstveno već je razapeto sa prvih K najvećih svojstvenih vektora matrice P , a zbog ortonormalne invarijantnosti 5.4.1 skup rješenja izgleda:

$$\{Z^*R : R^T R = I_K, PZ^* = Z^*\Lambda^*\}. \quad (5.20)$$

Gornji skup rješenja možemo okarakterizirati i partijskom matricom X uz pomoć normalizirajuće funkcije definirane u (5.15), odnosno njenog inverza (5.16). Budući da je $R^T R = I_K$, vrijedi

$$f^{-1}(Z^*R) = f^{-1}(Z^*)R,$$

pa skup rješenja možemo zapisati kao:

$$\{\tilde{X}^* R : \tilde{X}^* = f^{-1}(Z^*), R^T R = I_K\}. \quad (5.21)$$

5.5 Nalaženje optimalnog diskretnog rješenja

Uz pomoć rješenja relaksiranog problema (5.17) (OPZ) želimo pronaći diskretno rješenje X^* za problem 5.9 (OPX). Budući da su svi $Z^* R$ iz skupa rješenja (5.20) međusobno ekvivalentni u smislu da rezultiraju istim maksimumom ciljne funkcije, ima smisla tražiti diskretno rješenje X^* koje će biti najbliže jednom od $Z^* R$. Prelaskom na (5.21), karakterizaciju skupa relaksiranih rješenja pomoću particijske matrice X , problem formuliramo kao traženje diskretnog rješenja X^* najbližeg jednom od $\tilde{X}^* R = f^{-1}(Z^*) R$. Naša nova zadaća glasi:

Diskretni optimizacijski problem (DOP)

$$\begin{aligned} &\text{minimiziraj} && \Phi(X, R) = \|X - \tilde{X}^* R\|_F^2 \\ &\text{uz} && X \in \{0, 1\}^{N \times K}, \quad X \mathbb{1}_K = \mathbb{1}_N, \quad R^T R = I_K, \end{aligned} \quad (5.22)$$

gdje $\|A\|_F$ označava Frobeniusovu normu matrice A .

Gornji problem je Prokrustov problem (Dodatak, 10.3) i znamo da vrijedi sljedeće:

$$\|X - \tilde{X}^* R\|_F^2 = \|X\|_F^2 - 2\text{tr}(R^T (\tilde{X}^*)^T X) + \text{tr}((\tilde{X}^*)^T \tilde{X}^* R R^T),$$

pa se minimizacija (5.22) svodi na minimizaciju $-2\text{tr}(R^T (\tilde{X}^*)^T X) + \text{tr}((\tilde{X}^*)^T \tilde{X}^* R R^T)$ što je nelinearan optimizacijski problem sa dvije nepoznanice.

Da bismo to pojednostavili, optimalno rješenje X^* i pripadnu matricu $\tilde{R}$ tražit ćemo alternirajućim koracima:

1. držeći R fiksnim, minimiziraj $\Phi(X, R)$ u odnosu na X ,
2. držeći X fiksnim, minimiziraj $\Phi(X, R)$ u odnosu na R .

Drugim riječima, traženje rješenja $(X^*, \tilde{R})$ u (5.22) sveli smo na sljedeća dva potproblema:

Diskretni optimizacijski problem po X (DOPX)

Neka je dana ortogonalna matrica $\tilde{R}$.

$$\begin{aligned} &\text{minimiziraj} && \Phi(X) = \|X - \tilde{X}^* \tilde{R}\|_F^2 \\ &\text{uz} && X \in \{0, 1\}^{N \times K}, \quad X \mathbb{1}_K = \mathbb{1}_N. \end{aligned} \quad (5.23)$$

Diskretni optimizacijski problem po R (DOPR)

Neka je dano X^* .

$$\begin{aligned} &\text{minimiziraj} && \Phi(R) = \|X^* - \tilde{X}^* R\|_F^2 \\ &\text{uz} && R^T R = I_K. \end{aligned} \quad (5.24)$$

Rješenja gornjih problema, a istovremeno i rješenje od (5.22) daju nam sljedeći teoremi:

Teorem 5.5.1. (DOPX)

Neka je $\tilde{X} = \tilde{X}^* R_*$. Optimalno rješenje za problem (5.23) je dano sa:

$$X^*(i, l) = 1_{\{l = \arg \max_{k \in [K]} \tilde{X}(i, k)\}}, \quad i \in \mathbb{V}. \quad (5.25)$$

Teorem 5.5.2. (DOPR)

Rješenje problema (5.24) dano je sa:

$$\tilde{R} = \tilde{U} U^T, \quad (5.26)$$

$$X^{*T} \tilde{X}^* = U \Omega \tilde{U}^T, \quad \Omega = \text{Diag}(\omega), \quad (5.27)$$

gdje je $(U, \Omega, \tilde{U})$ dekompozicija singularnih vrijednosti od $X^{*T} \tilde{X}^*$ te vrijedi $U^T U = I_K$, $\tilde{U}^T \tilde{U} = I_K$ i $\omega_1 \geq \omega_2 \geq \dots \geq \omega_K$.

Dokaz. Dokaz slijedi iz Teorema 10.0.10 za $A \equiv \tilde{X}^*$, $B \equiv X$.

□

Gornji postupak pronalaženja rješenja spektralnog klasteriranja uz pomoć matrice R zovemo **spektralna rotacija**.

5.6 Algoritam K-particioniranja

Algoritam K-particioniranja

Neka je dana matrica susjedstva W i željeni broj klastera K .

1. Izračunati matricu susjedstva $D = \text{Diag}(W \mathbb{1}_N)$

2. Pronaći rješenje Z^* zadaje OPZ:

$$D^{-\frac{1}{2}} W D^{-\frac{1}{2}} \bar{V}_{[K]} = \bar{V}_{[K]} \text{Diag}(s_{[K]}), \quad \bar{V}_{[K]}^T \bar{V}_{[K]} = I_K$$

$$Z^* = D^{-\frac{1}{2}} \bar{V}_{[K]}$$

3. Normalizirati Z^* sa $\tilde{X}^* = \text{Diag}(\text{diag}^{-\frac{1}{2}}(Z^* Z^{*T})) Z^*$

4. Izračunati $\tilde{R}$ te inicijalizirati X^* :

$$\tilde{R}_1 = [\tilde{X}^*(i, 1), \dots, \tilde{X}^*(i, K)]^T, \text{ za slučajno odabrani } i \in [N]$$

$$c = 0_{N \times 1}$$

Za $k = 2, \dots, K$:

$$c = c + \text{abs}(\tilde{X}^* \tilde{R}_{k-1})$$

$$\tilde{R}_k = [\tilde{X}^*(i, 1), \dots, \tilde{X}^*(i, K)]^T, \quad i = \arg \min c$$

5. Inicijalizirati parametar konvergencije, $\bar{\phi}^* = 0$

6. Pronaći optimalno diskretno rješenje X^* :

$$\tilde{X} = \tilde{X}^* \tilde{R}$$

$$X^*(i, l) = 1_{\{l = \arg \max_{k \in [K]} \tilde{X}(i, k)\}}, \quad i \in V, l \in [K]$$

7. Pronaći optimalnu ortonormalnu matricu $\tilde{R}$:

$$(X^*)^T \tilde{X}^* = U \Omega \tilde{U}^T, \quad \Omega = \text{Diag}(\omega)$$

$$\bar{\phi} = \text{tr}(\Omega)$$

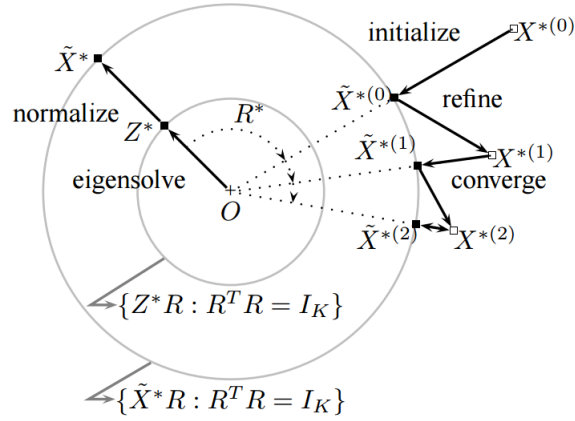
Ako $|\bar{\phi} - \bar{\phi}^*| < \text{preciznost}$, zapamti X^* i **STOP**

$$\bar{\phi} = \bar{\phi}^*$$

$$\tilde{R} = \tilde{U} U^T$$

8. Idi na korak 6.

Sljedeća slika iz [2] daje shematski dijagram gornjeg algoritma:



Slika 5.1: Shematski prikaz algoritma: (1) Najprije nalazimo svojstvene vektore od Z^* . Unutarnja kružnica predstavlja skup rješenja kojeg generira Z^* kroz ortogonalnu transformaciju R . Nakon normalizacije, rješenja partitioniranja leže na realnoj domeni koju ovdje predstavlja vanjska kružnica. (2) Iterativnim koracima nalazimo diskretno rješenje koje je najbliže dobivenom realnom rješenju. Počinjemo od diskretnog $X^{*(0)}$, pa računanjem $\tilde{R}$ nalazimo realno $\tilde{X}^{*(0)}$ koje $\tilde{X}^*$ dovodi bliže ka $X^{*(0)}$. Tada uz dani realni optimum $\tilde{X}^{*(0)}$ tražimo njemu najbliže diskretno rješenje $X^{*(1)}$ itd. Algoritam konvergira ka paru rješenja $(X^{*(2)}, \tilde{X}^{*(2)})$ koji su najbliže jedan drugom. Optimalnost od $\tilde{X}^{*(2)}$ garantira da je $X^{*(2)}$ skoro globalno optimalno rješenje

Poglavlje 6

Implementacija algoritma i primjeri

Algoritam K-particioniranja implementiran je u Matlab te proveden na primjerima klasteriranja točaka u $\mathbb{R}^2$.

Prvi korak algoritma zahtijeva definiranje matrice sličnosti, tj. susjedstva, a njeni elementi su težine bridova između vrhova grafa, tj. neke mjere sličnosti s_{ij} . Odabir mjere sličnosti ovisi o domeni naših podataka i vrsti grafa, a koje sve postoje razmotrit ćemo sljedeće.

6.1 Konstrukcija matrice susjedstva

Vrste grafova

- **Graf ε -okoline**

Povezujemo sve točke čije su međusobno udaljene za manje od ε .

- **Graf k-najbližih susjeda**

Vrh v_i je povezan sa vrhom v_j ako je v_j među k-najbližih susjeda vrha v_i ili ako je v_i među k-najbližih susjeda vrha v_j čime smo dobili neusmjeren graf.

- **Potpuno povezan graf**

Povezujemo sve vrhove za koje je mjera sličnosti između njih nenegativna. Sličnost bi trebala biti definirana na način da sličniji vrhovi imaju veće vrijednosti s_{ij} .

Funkcija sličnosti

U primjerima koji slijede, koristi ćemo potpuno povezani graf, a mjera sličnosti, da bi imala smisla, mora biti definirana na način da sama stvara lokalna susjedstva u smislu da poprima veće vrijednosti za vrhove koji su blizu i pripadaju istom klasteru, a manje za one

udaljenije. Zato ćemo kao sličnost koristiti Gaussovu funkciju sličnosti između točaka v_i, v_j definiranu s:

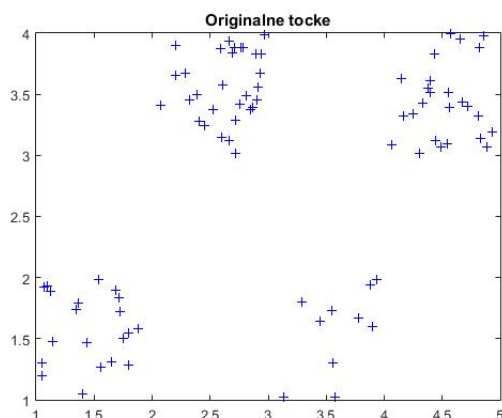
$$s(v_i, v_j) = s_{ij} = \exp\left(\frac{-\|v_i - v_j\|^2}{2\sigma^2}\right), \quad (6.1)$$

Parametar σ ovdje ima ulogu kontrole okoline svake točke u smislu da će sličnost između točaka iz manje okoline biti veća od onih koje se nalaze dalje. Primijetimo, matrica sličnosti neće biti rijetka budući da s_{ij} između udaljenijih točaka nikad neće biti nula, ali će biti dovoljno mala.

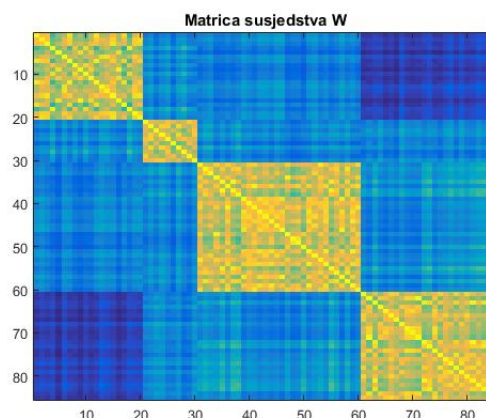
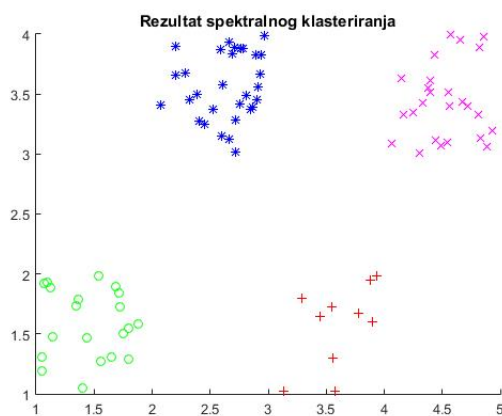
Pitanje koje se nameće jest odabir parametra σ . Naravno, ne postoji jedinstveno rješenje, već odabir uvelike ovisi o podacima te njihovoj distribuciji. No, neka istraživanja pokazala su kako σ možemo odabrati na način da izračunamo prosječnu udaljenost točke od njoj najbližih k susjednih točaka, gdje je $k = \log(n) + 1$.

6.2 Spektralno klasteriranje na simuliranim podacima

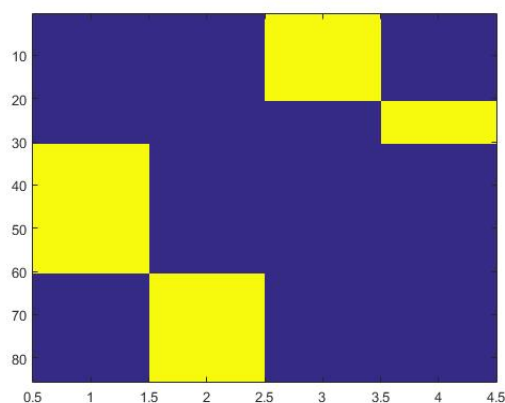
Primjer 6.2.1. Simulirano je $N=85$ točaka iz $\mathbb{R}^2$ sa očitom pripadnošću jednom od četiri klastera. Matrica susjedstva izračunata je kao $W_{ij} = s_{ij}$ gdje je s_{ij} Gaussijan funkcija sličnosti iz (6.1) uz $\sigma = 1$. Algoritam K -partitioniranja dao je točne rezultate.

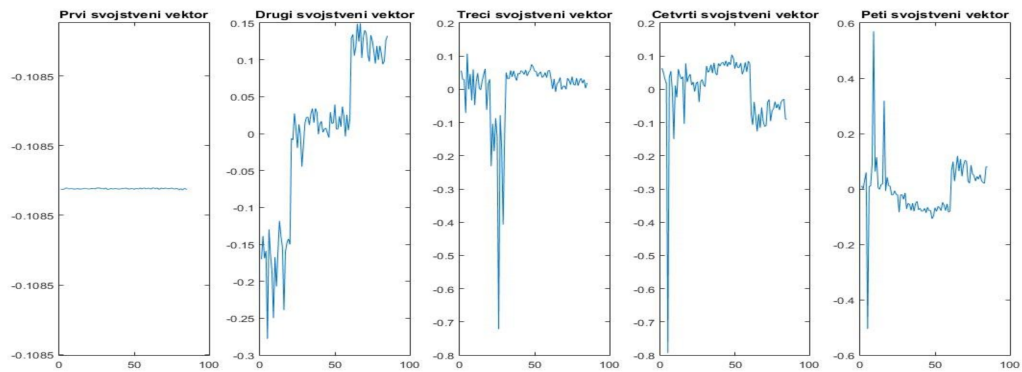


Slika 6.1: Originalni podaci

Slika 6.2: Matrica susjedstva W 

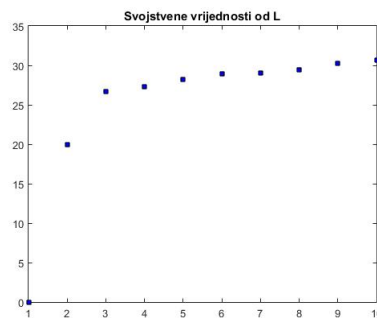
Slika 6.3: Rezultat klasteriranja

Slika 6.4: Matrica X^*



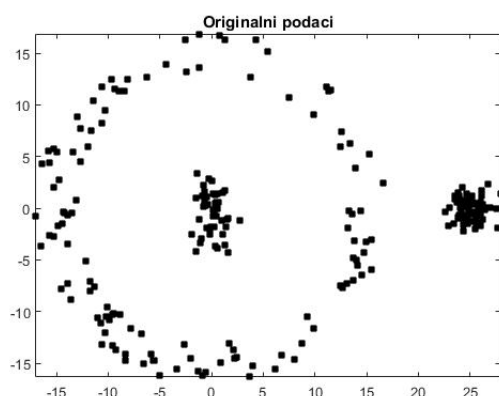
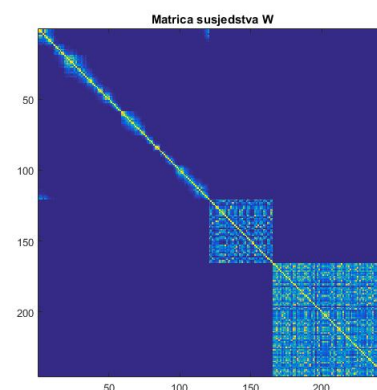
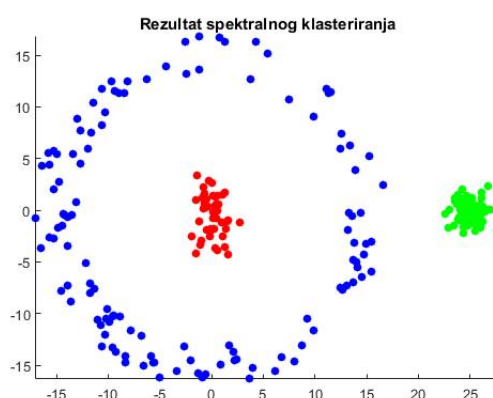
Slika 6.5: Pet najmanjih svojstvenih vektora Laplacijana

Pogledajmo sada kako izgledaju svojstvene vrijednosti i vektori Laplacijana $L = D - W$. Iz Propozicije 2.2.2 znamo da ako graf ima K povezanih komponenti, tada je kratnost 0 kao svojstvene vrijednosti Laplacijana matrice jednaka K . Budući da je graf potpuno povezan, a Gaussijan funkcija je uvijek nenegativna, graf se tretira kao jedna komponenta na što nam ukazuje i kratnost svojstvene vrijednosti 0.



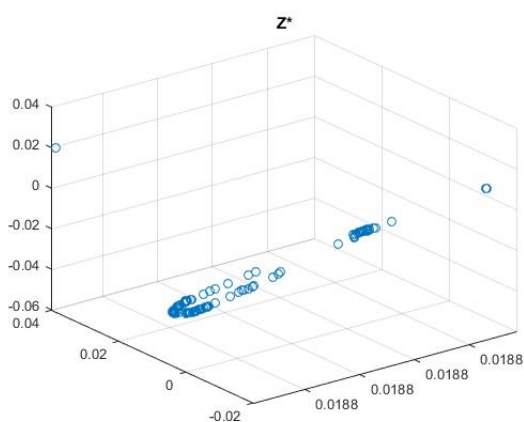
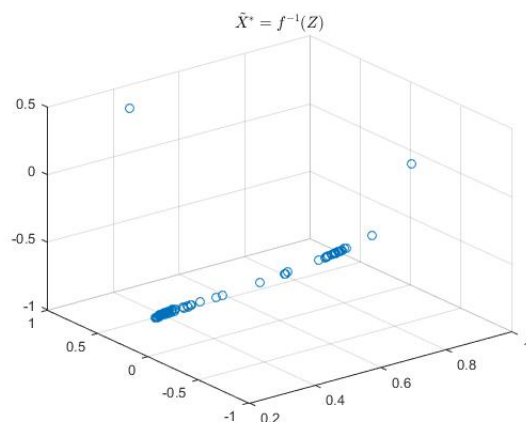
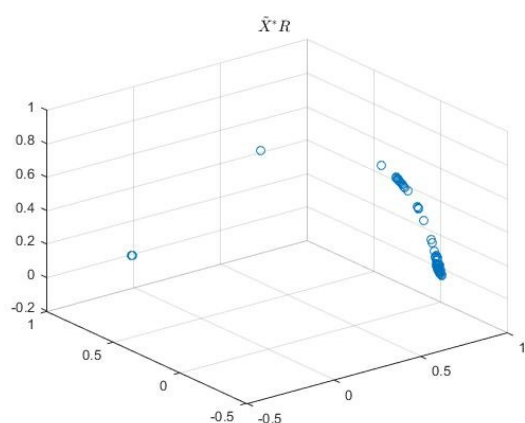
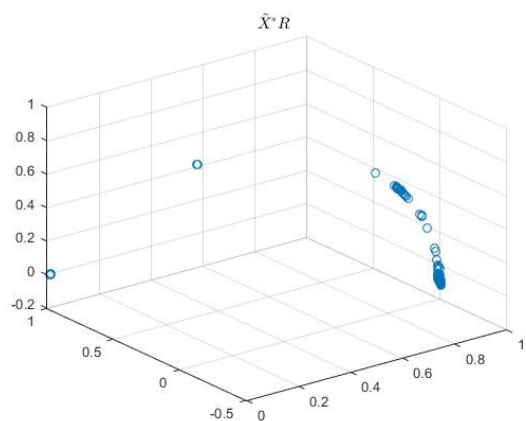
Slika 6.6: Svojstvene vrijednosti Laplacijana

Primjer 6.2.2. Generirano je $N=200$ točaka koje pripadaju jednom od 3 klastera, no oni su ovog puta drugačijeg oblika.

Slika 6.7: Originalni podaci, $N=200$ Slika 6.8: Matrica susjedstva W , definirana sa 6.1 uz $\sigma = 0.8$.

Slika 6.9: Rezultat spektralnog klasteriranja

Promotrimo na ovom primjeru proces spektralne rotacije. Budući da je dani broj klastera $K=3$, drugi korak Algoritma spektralnog klasteriranja zahtijeva računanje generaliziranog spektralnog problema te u matricu Z^* prema prvih 3 dobivenih svojstvenih vektora. Funkcija f^{-1} normalizira retke matrice Z^* te na taj način formira matricu $\tilde{X}^*$. Rotacijska matrica R u svakom koraku dovodi $\tilde{X}^*$ sve bliže ka diskretnom rješenju X^* koje u ovom primjeru algoritam pronalazi vrlo brzo, nakon svega 2 koraka.

Slika 6.10: $Z^* = [V_1, V_2, V_3]$ Slika 6.11: $\tilde{X}^* = f^{-1}(Z^*)$ Slika 6.12: Inicijalizacija $\tilde{X}^* R^{(0)}$ Slika 6.13: $\tilde{X}^* R^{(1)}$. Ova rotacija dovodi do konvergencije.

Spektralno klasteriranje u ovom primjeru dalo je dobre rezultate. Upravo na ovakvim podacima, tj. intuitivno manje kompaktnim klasterima vidi se razlika rezultata spektralnog klasteriranja i drugih metoda (npr. K-sredina), o čemu će biti riječi u sljedećem poglavlju.

Poglavlje 7

K-sredine i spektralno klasteriranje

7.1 Algoritam K-sredina

Neka je dan skup observacija $C = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ gdje je svaki $\mathbf{x}_i$ d-dimenzionalni vektor. Cilj algoritma K-sredina jest particionirati dani skup C na $K \leq N$ grupa $C_1, \dots, C_K$ takve da $C = \cup_{i=1}^N C_i$, $C_i \cap C_j = \emptyset$, $\forall i \neq j$. Algoritam to postiže na način da minimizira udaljenosti observacija unutar klastera od aritmetičke sredine klastera, tzv. centroida. Matematički,

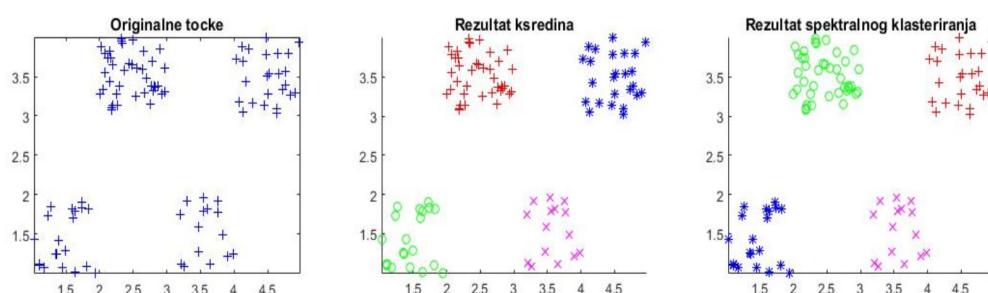
$$\arg \min_C \sum_{i=1}^K \sum_{\mathbf{x}_i \in C_i} \|\mathbf{x}_i - \mu_i\|_2^2, \quad (7.1)$$

gdje je μ_i aritmetička sredina vektora iz C_i .

Algoritam K-sredina
Neka je dana skup C i željeni broj klastera K. 1. Odabrati K centroida, po jedan za svaki klaster 2. Svaki vektor pridružiti klasteru čijem je centroidu najbliži 3. Ažurirati centroide, tj. izračunati sredine novonastalih klastera iz 2. koraka 4. Ponavljati korake 2 i 3 sve dok više nema promjene u centroidima

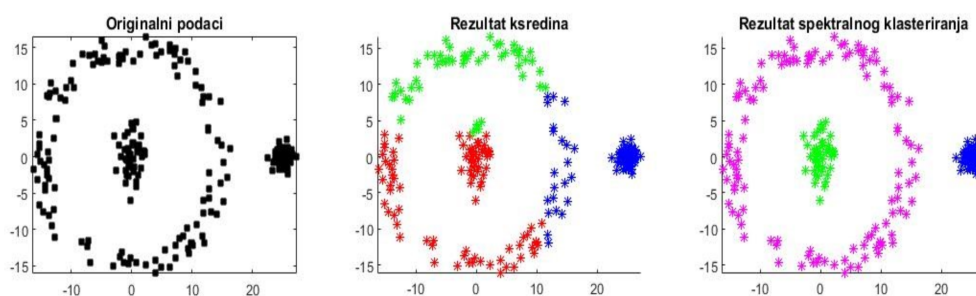
7.2 Usporedba sa spektralnim klasteriranjem

Primjer 7.2.1. Točke generirane kao u Primjeru 6.2.1 tvore četiri klastera koji su intuitivno govoreći kompaktni, odnosno točke unutar istog klastera su zbijene i unutar čvrstih granica. U ovom slučaju oba algoritma daju iste rezultate.



Slika 7.1: Na klasterima čije su točke međusobno dovoljno zbijene, K-sredine i spektralno klasteriranje daju iste rezultate.

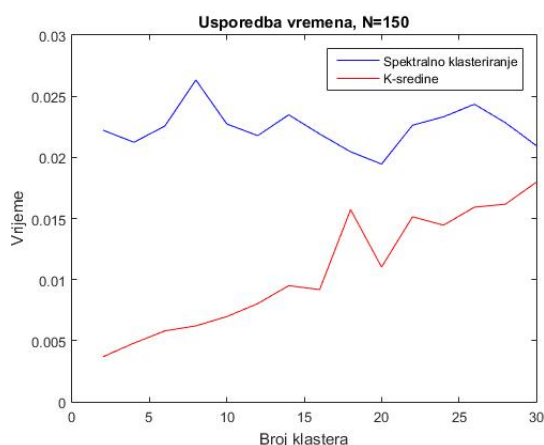
Primjer 7.2.2. Točke simulirane kao u Primjeru 6.2.2 tvore 3 klastera koji nisu nužno grupirani unutar čvrstih granica. Ovdje je, kao što vidimo, spektralno klasteriranje rezultiralo boljim, tj. točnim rezultatom.



Slika 7.2: Spektralno klasteriranje dalo je bolje rezultate.

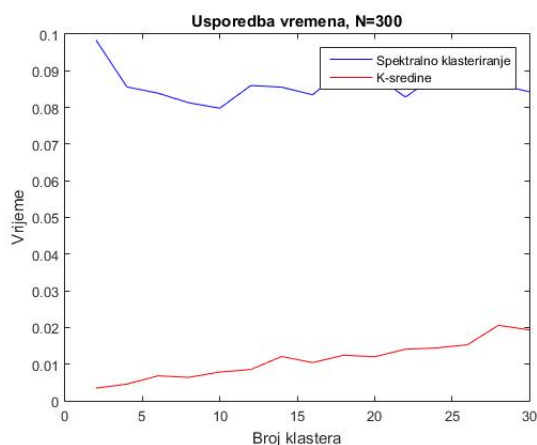
Vrijeme izvršavanja

Generirano je $N=150$ i $N=300$ točaka koje algoritmima K-sredina i spektralnog klasteriranja želimo podijeliti u $K = \{2, 4, 6, \dots, 30\}$ klastera. Algoritam K-sredina pokazuje se bržim od spektralnog klasteriranja te je njegovo vrijeme izvršavanja manje osjetljivo na promjene veličine uzorka.



K	Spektralno klasteriranje	K-sredine
2	0,022	0,004
4	0,021	0,005
6	0,023	0,006
8	0,026	0,006
10	0,023	0,007
12	0,022	0,008
14	0,023	0,010
16	0,022	0,009
18	0,020	0,016
20	0,019	0,011
22	0,023	0,015
24	0,023	0,014
26	0,024	0,016
28	0,023	0,016
30	0,021	0,018

Slika 7.3: Usporedba vremena izvršavanja spektralnog klasteriranja i K-sredina za $N=150$.



K	Spektralno klasteriranje	K-sredine
2	0,098	0,003
4	0,086	0,005
6	0,084	0,007
8	0,081	0,006
10	0,080	0,008
12	0,086	0,009
14	0,086	0,012
16	0,083	0,010
18	0,091	0,012
20	0,089	0,012
22	0,083	0,014
24	0,089	0,014
26	0,088	0,015
28	0,086	0,021
30	0,084	0,019

Slika 7.4: Usporedba vremena izvršavanja spektralnog klasteriranja i K-sredina za $N=300$.

7.3 Usporedba K-sredina i spektralne rotacije u spektralnom klasteriranju

Neki algoritmi spektralnog klasteriranja na retke matrice Z^* primjenjuju algoritam K-sredina te na taj način dolaze do konačnog rješenja. Algoritam iz ovog rada međutim za pronalaženje diskretnog rješenja koristi tzv. spektralnu rotaciju koja, kao što smo vidjeli, naizmjenično pronalazi rotacijsku matricu R i diskretno rješenje sve više približava onom pravom.

Neka je $Z^* \in \mathbb{R}^{N \times K}$ matrica K svojstvenih vektora i X^* particijska matrica. Jednadžbu (7.1) algoritma K-sredina možemo zapisati i u terminologiji našeg optimizacijskog problema:

$$\min_H \|Z^* - X^* H\|_F^2. \quad (7.2)$$

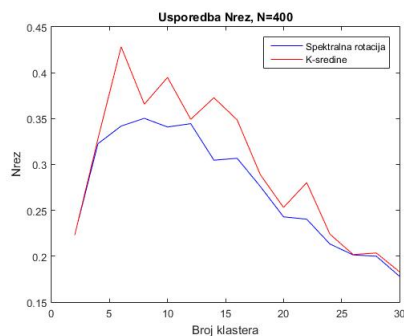
Rješenje gornjeg je matrica H takva da $X^* H$ najbolje aproksimira Z^* . Međutim, da li particijska matrica X^* dovoljno dobro opisuje Z^* ?

Prisjetimo se, diskretni optimizacijski problem po R (5.24) glasi

$$\min_{R^T R = I_K} \|X^* - \tilde{X}^* R\|_F^2 \quad (7.3)$$

što je slično formulaciji problema K-sredina (7.2) no ono što čini glavnu razliku jest da će rješenje spektralne rotacije zbog ortonormalnosti matrice R garantirati da će X^* biti najbliže pravom rješenju među svim ostalim rješenjima oblika (5.21).

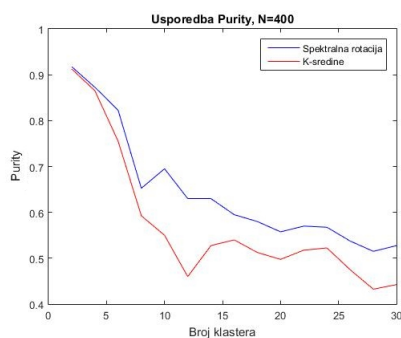
Upravo zbog bolje aproksimacije, spektralno klasteriranje sa rotacijom rezultira i sa boljom optimizacijom normaliziranog reza.



K	Spektralna rotacija	K-sredine
2	0,238	0,239
4	0,319	0,322
6	0,327	0,329
8	0,336	0,358
10	0,351	0,365
12	0,352	0,367
14	0,365	0,367
16	0,364	0,386
18	0,347	0,358
20	0,361	0,367
22	0,351	0,388
24	0,345	0,370
26	0,354	0,390
28	0,351	0,378
30	0,345	0,352

Slika 7.5: Generirano je $N=400$ točaka. Spektralna rotacija poprima manje vrijednosti normaliziranog reza, što je rezultat bolje aproksimacije stvarnog rješenja.

Osim toga, spektralna rotacija dovest će do veće "čistoće" (eng. *Purity*) klastera koja je jedna od mjera za evaluaciju algoritama klasteriranja. Računa se kao broj točno raspoređenih točaka podijeljen sa ukupnim brojem točaka N.



K	Spektralna rotacija	K-sredine
2	0,950	0,950
4	0,880	0,865
6	0,720	0,725
8	0,753	0,665
10	0,723	0,653
12	0,713	0,573
14	0,605	0,585
16	0,583	0,460
18	0,595	0,510
20	0,558	0,450
22	0,575	0,518
24	0,513	0,490
26	0,543	0,538
28	0,518	0,495
30	0,528	0,415

Slika 7.6: Pri klasteriranju N=400 točaka na $K=\{2, 4, \dots, 30\}$ klastera spektralno klasteriranje sa rotacijom ima veću čistoću od spektralnog klasteriranja sa K-sredinama.

Poglavlje 8

Klasteriranje slike

Osim na simulirane točke, Algoritam spektralnog klasteriranja možemo primijeniti i na klasteriranje slika učitane kao "grayscale". Vrhove grafa čine pikseli, tj. njihova boja, a matrica susjedstva definirana je kao i prije, sa Gaussovom funkcijom sličnosti (6.1).

Primjer 8.0.1. Učitana je slika dimenzija 580×580 piksela. Cilj je da algoritam uspije odvojiti slova od pozadine, odnosno klasterirati sliku u $K=2$ grupe.



Slika 8.1: Originalna slika je dimenzija 580×580 piksela. Zbog bržeg računanja, slika se smanjuje na dimenziju 50×50 čime slova i dalje ostaju jasna. Slika će se tretirati kao *grayscale*. Algoritam je uspio odvojiti slova od pozadine te ih shvatiti kao jedan klaster.

Primjer 8.0.2. Učitana je crno bijela fotografija 8.2 dimenzija 250×250 piksela.



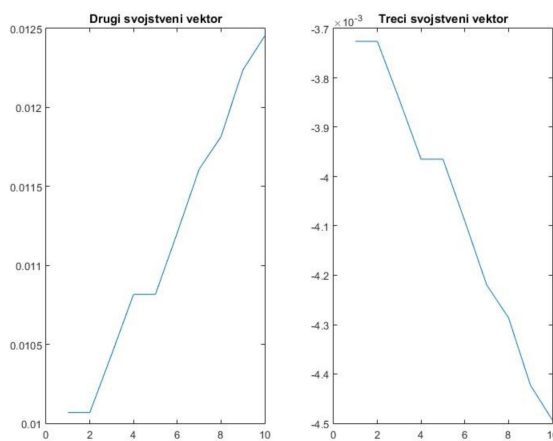
Slika 8.2: Originalna fotografija

Na sljedećoj slici vidimo rezultat klasteriranja gornje fotografije na $K=2$ klastera.



Slika 8.3: Rezultat. Potrebno vrijeme izvršavanja $t = 20.2$ sekunde.

Budući da je $K=2$, drugi svojstveni vektor Laplacijana L nosi sve potrebne informacije o klasterima.



Slika 8.4: Drugi i treći svojstveni vektor pripadnog Laplacijana.

Poglavlje 9

Segmentacija slika normaliziranim rezom

U ovom poglavlju vidjet ćemo još jedan način korištenja normaliziranog reza za segmentaciju slike. Princip je sličan kao i prije, konstruiramo graf $\mathcal{G}$, definiramo matricu susjedstva W te traženje najboljeg reza svodimo na traženje svojstvenih vrijednosti i vektora Laplacijana. Matricu W sada ćemo definirati da bude rijetka, odnosno da poprima vrijednosti 0 za vrhove koji su međusobno dovoljno udaljeni.

Konstrukcija grafa i matrica susjedstva

Graf $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ konstruirat ćemo na način da pikseli slike čine vrhove, a težine bridova su vjerojatnosti da dva piksela pripadaju istom objektu na slici. Kako bismo to postigli, koristit ćemo sljedeću funkciju:

$$w_{ij} = e^{\frac{-\|F(i)-F(j)\|_2^2}{\sigma_I^2}} * \begin{cases} e^{\frac{-\|X(i)-X(j)\|_2^2}{\sigma_X^2}}, & \|X(i) - X(j)\|_2 < r \\ 0, & \text{inače,} \end{cases} \quad (9.1)$$

gdje je F_i razina svjetlosti piksela, a r je unaprijed definirana najveća tolerirana udaljenost među pikselima.

Algoritam (rekurzivnog) biparticioniranja

Algoritam (rekurzivnog) biparticioniranja

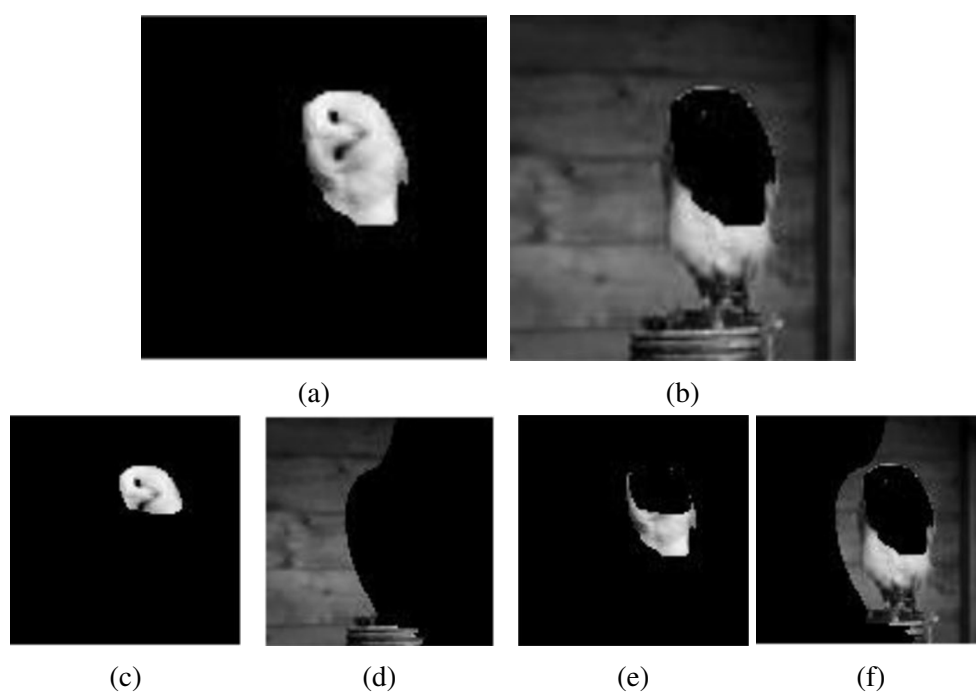
1. Konstruirati graf $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ sa pikselima kao vrhovima i matricom susjedstva W definiranom kao u (9.1)
2. Riješiti generalizirani svojstveni problem $Lx = \lambda Dx$, tj. pronaći drugu najmanju svojstvenu vrijednost normaliziranog Laplacijana koja minimizira vrijednost normaliziranog reza.
3. Ukoliko je potrebno, rekurzivno reparticioniramo dobivene segmentirane dijelove.

U idealnom slučaju, svojstveni vektor poprima dvije diskretne vrijednosti čiji nam predznaci ukazuju na pripadnost particijama. No, to ne mora biti slučaj, tj. vektor poprima neprekidne vrijednosti te zato odabiremo graničnu kojom ćemo stvoriti particije. Ta vrijednost može biti 0, medijan ili ona koja ima najbolju vrijednost Nreza.

Primjer 9.0.3. Učitati ćemo grayscale sliku te na njoj primjeniti Algoritam rekurzivnog biparticioniranja. Najprije ćemo sliku podijeliti na dva dijela prikazana na Slici 9.2 (a) i (b), a zatim dobivenu particiju na još 4 podgrafa (c)-(f).



Slika 9.1: Originalna slika dimenzija 250×250 .



Slika 9.2: (a) Originalna fotografija dimenzija 250×250 . (b) i (c) su dobivene particije uz $\sigma_I = 0.1, \sigma_x = 0.3, r = 5$. Granična vrijednost razdvajanja drugog svojstvenog vektora je 0. (c)-(f) su dobivene rekurzivnom bipartijom (a) i (b) na 4 podgraфа.

Poglavlje 10

Dodatak

Osnovne definicije

Definicija 10.0.4. (Svojstvena vrijednost i vektor)

Svojstveni vektor kvadratne matrice $A \in M^{n \times n}$ je nenul vektor $x \in \mathbb{R}^n$ za kojeg postoji skalar λ takav da vrijedi

$$A\lambda = \lambda x.$$

λ je svojstvena vrijednost matrice A , a par (λ, x) zovemo svojstveni par.

Skup svih svojstvenih vrijednosti matrice A zovemo spektar od A oznake $\sigma(A)$.

Općenito, svojstvene vrijednosti matrice A su rješenja karakteristične jednadžbe $\det(A - \lambda I) = 0$.

Definicija 10.0.5. (Frobeniusova norma)

Neka je $A \in \mathbb{R}^{m \times n}$. Frobeniusovu normu matrice A definiramo sa

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}. \quad (10.1)$$

Definicija 10.0.6. (Ortogonalna matrica)

Za realnu matricu R kažemo da je ortogonalna ako vrijedi

$$R^T R = R R^T = I.$$

Teorem 10.0.7. *Neka je $A \in \mathbb{R}^{n \times n}$ simetrična matrica. Njezini svojstveni vektori tada su međusobno okomiti.*

Dokaz. Neka je λ svojstvena vrijednost od A sa pripadnim svojstvenim vektorom x te neka je $\mu \neq \lambda$ druga svojstvena vrijednost sa pripadnim vektorom y . Imamo $Ax = \lambda x$ i $Ay = \mu y$. Odavde imamo

$$\begin{aligned} Ax \cdot y &= \lambda x \cdot y \\ x \cdot Ay &= \mu x \cdot y. \end{aligned}$$

Zbog simetričnosti od A slijedi

$$\begin{aligned} Ax \cdot y &= x \cdot Ay \\ \lambda x \cdot y &= \mu x \cdot y \\ (\lambda - \mu)x \cdot y &= 0, \end{aligned}$$

a kako je $\lambda \neq \mu$ slijedi $x \cdot y = 0$. □

Teorem 10.0.8. *(Dekompozicija singularnih vrijednosti, SVD ¹)*

Ako je $A \in \mathbb{R}^{m \times n}$ realna $m \times n$ matrica, tada postoje ortogonalne matrice $U = [u_1, \dots, u_m] \in \mathbb{R}^{m \times m}$ i $V = [v_1, \dots, v_n] \in \mathbb{R}^{n \times n}$ takve da

$$U^T A V = \Sigma = \text{Diag}(\sigma_1, \dots, \sigma_p) \in \mathbb{R}^{m \times n}, \quad p = \min\{m, n\} \quad (10.2)$$

i $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$.

Prokrustov problem

Definicija 10.0.9. *Neka su $A, B \in \mathbb{R}^{m \times n}$, $m \geq n$ zadane matrice. Optimizacijski problem*

$$\min_{R^T R = I_n} \|AR - B\|_F \quad (10.3)$$

zovemo ortogonalni Prokrustov problem.

Riječima, cilj ovog optimizacijskog problema je pronaći ortogonalnu matricu R koja će točke iz A što je najviše moguće približiti točkama iz B .

Kako bismo pronašli rješenje problema, iskoristit ćemo svojstvo Frobeniusove norme: Za

¹Eng. *Singular Value Decomposition*

$A \in \mathbb{R}^{m \times n}$ vrijedi $\|A\|_F^2 = \text{tr}(A^T A)$, gdje je $\text{tr}(\cdot)$ trag matrice. Sada problem (10.3) možemo zapisati na sljedeći način:

$$\begin{aligned} \min_{R^T R = I_n} \|AR - B\|_F^2 &= \min_{R^T R = I_n} \text{tr}((AR - B)^T (AR - B)) = \\ &= \min_{R^T R = I_n} \text{tr}(R^T A^T AR) - 2\text{tr}(R^T A^T B) + \text{tr}(B^T B) = \\ &= \min_{R^T R = I_n} \|A\|_F^2 - 2\text{tr}(R^T A^T B) + \|B\|_F^2, \end{aligned} \quad (10.4)$$

što znači da je on ekvivalentan problemu maksimiziranja traga matrice $R^T A^T B$.

Sljedeći teorem daje nam rješenje Prokrustovog problema.

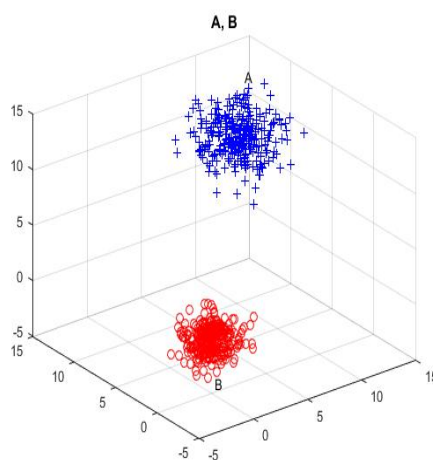
Teorem 10.0.10. *Rješenje problema (10.3) dano je sa $R = UV^T$ gdje su U i V matrice iz dekompozicije singularnih vrijednosti matrice $A^T B$.*

Dokaz. Pokazali smo da je problem (10.3) ekvivalentan problemu (10.4), odnosno maksimiziranju traga matrice $R^T A^T B$. Budući da je, po pretpostavci, $U\Sigma V^T = A^T B$ SVD matrice $A^T B$, $\Sigma = \text{Diag}(\sigma_1, \dots, \sigma_n)$, imamo:

$$\text{tr}(R^T A^T B) = \text{tr}(R^T U\Sigma V^T) = \text{tr}(V^T R^T U\Sigma) = \sum_{i=1}^n (V^T R^T U)_{ii} \sigma_i \leq \sum_{i=1}^n \sigma_{ii}$$

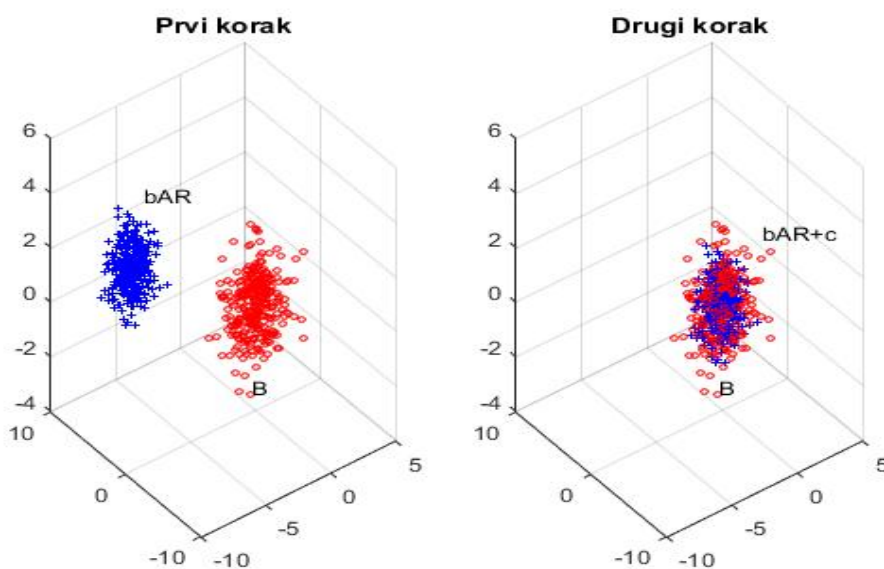
Jednakost, odnosno maksimum traga matrice $R^T A^T B$, postiže se za $V^T R^T U = I_n$, a to je ispunjeno za $R = UV^T$ čime je teorem dokazan. □

Primjer 10.0.11. *Prikazat ćemo Prokrustov problem na simuliranim točkama iz $\mathbb{R}^3$ koje izgledaju kao na Slici 10.1. Dane su matrice A i B te algoritam pokušava naći ortogonalnu matricu R takvu da se točke iz A najviše moguće približe točkama iz B .*



Slika 10.1: Početne matrice A i B. Cilj je točke iz A što više približiti točkama iz B.

Matlabova funkcija *procrust*($\cdot, \cdot$) vraća tri komponente - komponentu skaliranja b , ortogonalnu rotaciju i refleksiju R te c translacijsku komponentu. Na slici 10.2 najprije vidimo da je A skalirana parametrom b te reflektirana za matricu R , a za konačno rješenje još je translahirana za c .



Slika 10.2: Prokrustov problem. A je približena matrici B.

Bibliografija

- [1] J.Huang, F. Nie i H. Huang, *Spectral Rotation versus K-Means in Spectral Clustering*”.
- [2] Shi i Malik, *Normalized Cuts and Image Segmentation*, (2000), 888–904, <https://people.eecs.berkeley.edu/~malik/papers/SM-ncut.pdf>.
- [3] Thomas Viklands, *Algorithms for the Weighted Orthogonal Procrustes Problem and other Least Squares Problems*, (2006), <http://www8.cs.umu.se/~viklands/PhD.pdf>.
- [4] Ulrike von Luxburg, *Tutorial on Spectral Clustering*, (2007), http://www.kyb.mpg.de/fileadmin/user_upload/files/publications/attachments/Luxburg07_tutorial_4488%5b0%5d.pdf.
- [5] H. Zha, X.He, C.Ding, H.Simon i Ming Gu, *Spectral Relaxation for K-Means Clustering*, <http://ranger.uta.edu/~chqding/papers/KmeansNIPS2001.pdf>.

Sažetak

Ukratko, metoda spekralnog klasteriranja zahtijeva da naše podatke koje je potrebno grupirati prikazemo u formi neusmjerenog grafa u kojem težine bridova definiramo kao mjere sličnosti između podataka. Kako bismo particionirali graf, potrebne su nam funkcije particioniranja poput reza, razmjernog reza ili normaliziranog reza. Kao najbolja funkcija pokazuje se normalizirani rez jer jedini posjeduje svojstvo dualnosti, odnosno istovremene minimizacije sličnosti između klastera te maksimizacije sličnost unutar klastera što je upravo i cilj metode. Kako su optimizacije normaliziranog reza NP-teški problemi, metoda se iz diskretnog prostora prebacuje na neprekidnu domenu gdje nam na raspolaganju stoje sva dobra svojstva Laplacijana grafa, njegovih svojstvenih vrijednosti i vektora te teoremi iz varijacijske karakterizacije spektra simetrične matrice susjedstva. Pokazuje se da je za particioniranje grafa na K dijelova potrebno pronaći svojstvene vektore K najvećih svojstvenih vrijednosti normaliziranog Laplacijana grafa čime dobivamo neprekidno rješenje particioniranja. Da bismo dobili diskretno rješenje koje najbolje aproksimira pravo, spektralnom rotacijom naizmjenično nalazimo ortogonalnu matricu i diskretna rješenja čija će ortonormalna transformacija konvergirati ka pravom rješenju.

Summary

In this work we focused on the spectral clustering method which has become one of the most popular modern clustering algorithms. For the purpose of clustering the given objects (people, pixels, points...), we represented the data in a form of a weighted undirected graph where the nodes are the points in a feature space and the edges represent similarities between the points. In order to partition the graph, one needs partitioning functions like Cut, Ratio Cut or Normalized cut. The last one appears to perform the best because of the duality property. This means that the function simultaneously minimizes the disassociation between the groups and maximizes the association within the groups which is exactly what the clustering method is aiming to do. The optimization of the Normalized cut function is shown to be a NP-hard problem so in order to find a solution, we have to move our problem from the discrete domain to the continuous one. Based on the variational characterization of the spectrum of the associated Laplacian, an approximate solution is constructed using the eigenspace of its dominant eigenvalues. It is shown that to partition the graph into K groups, one has to find the eigenvectors of the K largest eigenvalues of the normalized Laplacian. By doing so, we obtain the relaxed solution to the problem. In order to get the discrete one, we iteratively find a sequence of orthogonal matrices and a discrete approximate solutions that tend to approximate well the true discrete solution.

Životopis

Rođena sam 7.srpnja 1992. u Zagrebu gdje trenutno i živim. Po nacionalnosti sam Hrvatica. Završila sam Osnovnu školu Augusta Cesarca u Krapini, 2007. godine upisala sam Opću gimnaziju u Krapini, a 2011. godine Prirodoslovno-matematički fakultet, smjer Matematika u Zagrebu te 2014. stekla titulu univ.bacc.math. Nakon toga, 2014. godine upisala sam Diplomski studij Matematičke statistike na PMF-u u Zagrebu.

U rujnu 2014. bila sam praktikantica u Erste banci u odjelu Upravljanje rizicima, a 2015. odradila sam praksu u Zagrebačkoj banci u odjelu Tržišne inteligencije. Trenutno radim kao pripravnica za CRM analitičara u Privrednoj banci Zagreb.

Raspolažem C1 razinom engleskog i njemačkog jezika u govoru i pismu.