MATRIČNE I TENZORSKE METODE U ANALIZI PODATAKA

# Multiclass Spectral Clustering

Alen Softić

Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
prosinac 2022.

# Sadržaj

# 1 Uvod

Klasteriranje ili grupiranje poodataka je jedan od osnovnih alata u analizi podataka. Uz klasifikaciju, regresiju i rangiranje, klasteriranje smatramo osnovnom metodom strojnog, odnosno statističkog učenja.

Klasteriranje predstavlja problem nenadziranog učenja, što znači da su podaci bez ciljne, odnosno izlazne vrijednosti *(label)*. Ponekad se nazivaju i neoznačenim podacima *(unlabeled)*. Klasteriranju pristupamo na način da podatke unutar grupe smatramo međusobno *sličnim*, dok elemente različitih grupa smatramo *nesličnima*, što god pojam "sličnosti" predstavljao u danom kontekstu. Kako bi taj pristup imao smisla razumno bi bilo pretpostaviti da u pozadini zaista postoji nekakva struktura u podacima, tj. prirodne grupacije koje su nama apriori nepoznate. Ova pretpostavka nije nužna kako bi se nadolazeći algoritmi klasteriranja koristili te će oni funkcionirati i na nestrukturiranim podacima. Samim tim, algoritmi klasteriranja će vrlo vjerojatno detektirati činjenicu kako podaci nemaju strukturu, što će se odraziti na grupacije.

## 1.1 Primjene klasteriranja

Neke od primjena klasteriranja podataka su:

- Društvene mreže: prijedlozi prijateljstava
- Marketing: segmentacija potrošača u grupe
- Biologija: grupiranje organizama prema njihovim značajkama
- Text mining: grupiranje sličnih dokumenata
- Bioinformatika: grupiranje DNA-mikropolja
- Obrada slike: sažimanje slike grupiranjem istobojnih piksela
- Segmentacija slike: prepoznavanje objekata na slici.

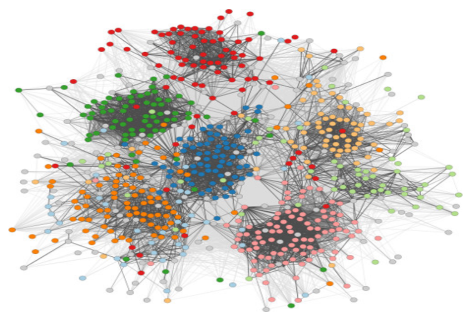
Rad se fokusira na progresivno poboljšanje algoritama spektralnog klasteriranja baziranih na rezovima u grafu. Koriste se alati vektorskog modela podataka, linearne algebre i spektralne teorije, teorije grafova i optimizacije.

# 2 Uvod u problematiku

Mnogi se problemi, uključujući i neke od navedenih u uvodu, mogu prikazati kao problemi *particioniranja grafa*. Particioniranjem grafa možemo smatrati razdvajanje grafa na disjunktne grupacije. Ono što zaista particioniramo u tom slučaju je skup vrhova.

Jasno je kako se grafovi pojave u modeliranju društvenih mreža:
U najjednostavnijem modelu korisnici $A$ i $B$ predstavljaju vrhove u grafu te su povezani bridom ukoliko su u relaciji *prijateljstva*. Prisjetimo se kako je početna stranica društvene mreže Facebook prikazivala upravo tu situaciju.


(a) Društvena mreža Facebook



(b) Stvarniji prikaz mreže

Manje je jasno kako sliku prevesti u jezik teorije grafova. Primjerice, pikseli mogu predstavljati vrhove grafa. Vrhove ćemo smatrati *sličnima* ukoliko su blizu jedan drugom te imaju slične karakteristike poput boje, intenziteta itd. Na osnovu toga konstruiramo *matricu sličnosti* i dobivamo težinski neusmjeren graf. Nadamo se iz njegovih svojstava, odnosno svojstava težinske matrice, doći do traženih rezultata. Zanimljivo je kako spektar težinske matrice igra veliku ulogu u particioniranju grafa, odakle i slijedi ime *Spektralno klasteriranje*.

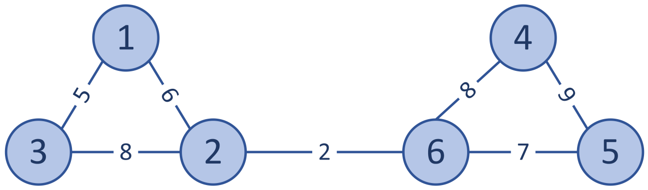
Ukoliko su dobiveni podaci točke u $\mathbb{R}^n$ one će predstavljati vrhove grafa, matricu sličnosti najčešće konstruiramo po nekom kriteriju koji uzima u obzir udaljenost točaka. Najčešće se koriste *Gaussove okoline* odnosno *jezgre* gdje je sličnost između $x_i$, $x_j \in \mathbb{R}^n$ definirana kao $exp(-\frac{\|x_i - x_j\|_2^2}{\sigma^2})$ za parametar okoline $\sigma > 0$. Alternativno, postoje i $\epsilon$-okoline ili okoline bazirane na najbližim susjedima.

## 2.1 Particioniranje grafa

Obzirom da je graf osnovni objekt koji izučavamo promotrimo sljedeću sliku.

Graf na slici 1 je potpuno povezan težinski graf, međutim prirodno se pojavljuju dvije grupacije vrhova. Vrhove 1, 2, 3 proglasili bismo prvom, a 4, 5, 6 drugom grupacijom. Grupacije su posljedica analize povezanosti vrhova, odnosno težina bridova koji ih povezuju.

Ovaj pristup ima gornju interpretaciju sa sličnošću, gdje snažno povezane vrhove smatramo sličnima. Naglasimo kako grafovi ne rade pretpostavke na oblike grupacija. Ovo svojstvo pokazati će se kao prednost u odnosu na određene metode klasifikacija za specifične probleme.

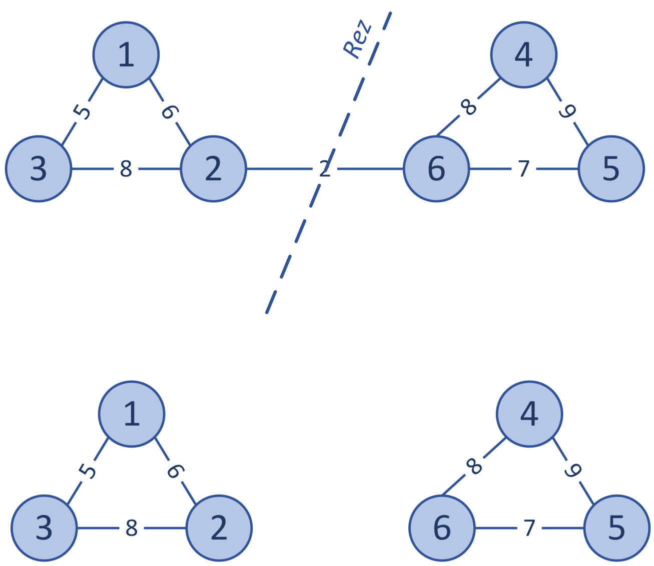



Slika 1: Primjer grafa

### 2.2 Biparticioniranje grafa

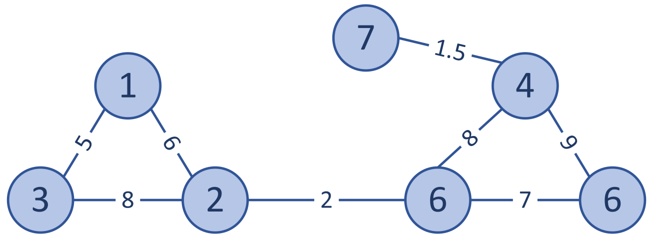
Particioniranje grafa na dva disjunktna dijela, odnosno podgrafa, nazivamo *biparticioniranjem grafa*. Razlog zašto je biparticioniranja grafa osnovna metoda particioniranja je činjenica kako ju rekurzivno možemo primijeniti kako bismo graf dodatno particionirali na proizvoljan broj grupacija.

Jedna ideja biparticioniranja grafa je pronalazak grupacija dobivenih eliminacijom bridova koje ih povezuju, na način da je zbroj težina tih bridova minimalan. Ova metoda, koju ćemo kasnije formalizirati, davala bi zadovoljavajuć rezultat za graf na slici 1. Eliminaciju bridova nazivamo *rezom*, a naziv je opravdan sljedećom slikom.



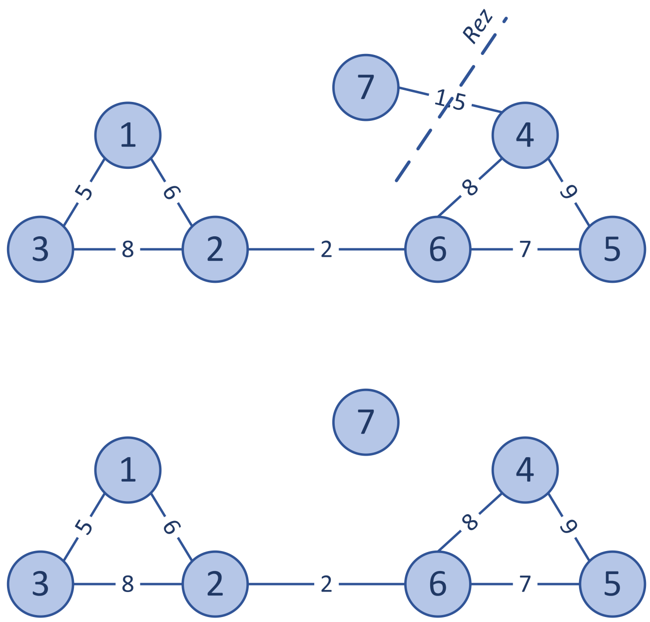

Slika 2: Particioniranje grafa rezom

Dodamo li grafu sa slike 1 novi vrh 7 (slika 3) povezan jednim bridom s težinom manjom od 2, njegov rez bio bi optimalan. Ovime pokazujemo kako gornja metoda, koju ćemo nazivati *metodom minimalnog reza*, ne daje nužno optimalne rezultate. Primjer je dan slikom 4.




Slika 3: Novi graf

Ponovno imamo dvije grupacije (eventualno tri, no još uvijek možemo tražiti najbolju biparticiju grafa), ali metoda ne daje očekivani rezultat. Prođemo li po svim biparticijama grafa te zbrojimo težine odsječenih bridova najbolji rezultat dan je slikom 4.




Slika 4: Particioniranje novog grafa rezom

Grupacije su u ovom primjeru vrlo neprirodne. Jedna opcija bila bi penalizirati malu veličinu grupacije, odnosno njen kardinalitet. Time dobivamo metodu *razmjernog reza*. Pokazuje se kako je to napredak, no ipak nije najbolje rješenje. Kod težinskog grafa kardinalitet grupacije nije najbolji pokazatelj njenog udjela u cijelom grafu, a razlog je upravo što ne uzimamo u obzir koliko su te grupacije "teške". Uzimajući u obzir težinu bridova unutar grupacija dolazimo do *normalziranih rezova*. Normalizirani rezovi su završna tema ovog rada. Imaju vrlo pogodna svojstva i značajan su napredak obzirom na prvi algoritam minimalnog reza, kao i algoritam razmjernog reza.

# 3 Osnovni pojmovi

Uvodimo neke od osnovnih pojmova koji su potrebni za daljnji razvoj teorije. Cilj je konstrukcija algoritama spektralnog klasteriranja.

Za početak naglašavamo da unutar vektorskog modela podataka vektore smatramo vektorima stupcima i označavamo ih malim slovima latinice. Matrice označavamo velikim slovima latinice.

**Definicija 3.1.** *Uređeni par $(V, W)$, gdje je $V = \{v_1, v_2, \dots v_n\}$ i $W$ nenegativna simetrična matrica sa nazivamo neusmjeren težinski graf sa vrhovima $v_1, v_2 \dots v_n$ i težinskom matricom $W$.*

Vrhove možemo smatrati i brojevima $1, 2 \dots n$, odnosno $V = \{1, 2, \dots, n\}$. Iz težinske matrice $W$ čitamo jesu li dva vrha povezana te je iz tog razloga klasični skup bridova $E$ izostavljen. Naime, $w_{ij} > 0$ ekvivalntno je činjenici da su vrhovi $v_i$ i $v_j$ povezani. Elementi na dijagonali težinske matrice govore nam radimo li sa grafovima bez petlji.

**Definicija 3.2.** *Stupanj vrha $v_i$ je zbroj svih težina bridova koji iz njega izlaze, a označavamo ga sa $d_i$.*

Težine bridova koje izlaze iz vrha $v_i$ su $w_{ij}$ za $j = 1, 2, \dots, n$. To je upravo $i$-ti redak matrice $W$. Stupnjeve vrhova biti će pogodno prikazati kao elemente dijagonalne matrice. Označimo n-dimenzionalni vektor jedinica sa $\mathbb{1}_n$. Primijetimo kako vrijedi

$$d_i = \sum_{j=1}^{n} w_{ij} = \sum_{j=1}^{n} w_{ij} \cdot 1 = \sum_{j=1}^{n} w_{ij} (\mathbb{1}_n)_j = (W \mathbb{1}_n)_i.$$

.

**Definicija 3.3.** *Matricu $D = diag(W\mathbb{1}_n)$ nazivamo matricom stupnjeva.*

**Definicija 3.4.** *$\Gamma_V^K$ nazivamo $K$-particijom skupa vrhova $V$ ako je $\Gamma_V^K = \{V_1, V_2, \dots V_K\}$, gdje su $V_i \cap V_j = \phi$ $\forall i \neq j$, $\bigcup_{i=1}^{K} V_i = V$.*

Ova definicija ostavlja mogućnost praznih elemenata particije, a time i praznih grupacija podataka. U tom slučaju broj *stvarnih* grupacija je manji od pretpostavljanog.

Fokusirajmo se na slučaj biparticioniranja grafa ($K = 2$).

## 3.1 Minimalan rez

Za proizvoljne $A$,$B \subseteq V$ definiramo mjeru $C(A, B)$ kao

$$C(A, B) := \sum_{\substack{i \in A \\ j \in B}} w_{ij}.$$

**Definicija 3.5.** *Rez biparticije $\Gamma_V^2$ definiramo kao $C(V_1, V_2)$.*

Pronalazak najbolje biparticije $(V_1^*, V_2^*)$ formuliramo kao diskretni optimizacijski problem:

$$(V_1^*, V_2^*) = \arg\min\{C(V_1, V_2) : V_1 \cap V_2 = \phi, V_1 \cup V_2 = V\}$$

**Napomena 3.6.** *Često ćemo u oznaku ubaciti simbol * kao pokazatelj optimalnosti.*

**Napomena 3.7.** *Korektnije bi bilo pisati* $(V_1^*, V_2^*) \in \arg\min\{C(V_1, V_2) : V_1 \cap V_2 = \phi, V_1 \cup V_2 = V\}$ *kako je moguće da dvije particije imaju jednak rez koji je ujedno i minimalan.*

Metoda minimalnog reza dobro je istražena te postoje efikasni algoritmi za njeno računanje, a koriste dinamičko programiranje. Ipak, ona favorizira rezanje manjih skupova izoliranih vrhova u grafu.

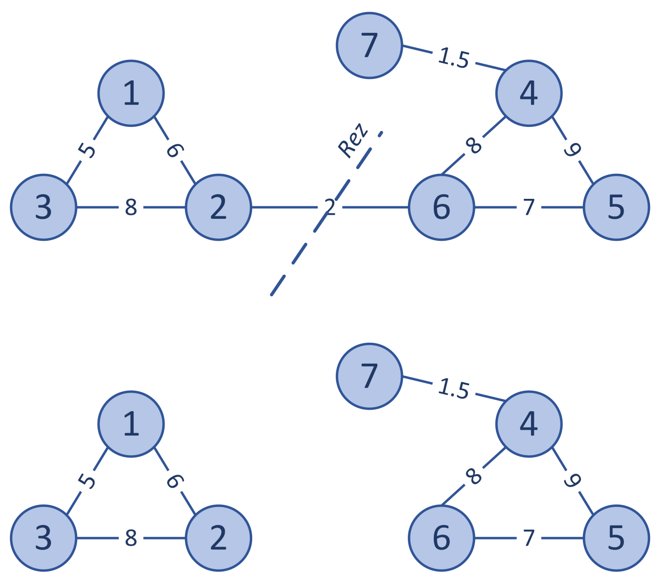
## 3.2 Razmjerni rez

Uvodimo penalizaciju na kardinalitet particija.

**Definicija 3.8.** *Razmjerni rez biparticije* $\Gamma_V^2$ *definiramo kao*

$$R(V_1, V_2) := \frac{C(V_1, V_2)}{|V_1|} + \frac{C(V_1, V_2)}{|V_2|}.$$

Razmjerni rez riješiti će problem iz uvoda, a dan je slikom 5. Vrijedi:
$R(\{1,2,3\},\{4,5,6,7\}) = \frac{2}{3} + \frac{2}{4} < R(\{1,2,3,4,5,6\},\{7\}) = \frac{1.5}{6} + \frac{1.5}{1}$. Ostale particije daju veće razmjerne rezove, dakle pronašli minimum. Najbolja biparticija u smislu razmjernog reza dana je sa $V_1 = \{1,2,3\}$, $V_2 = \{4,5,6,7\}$.




Slika 5: Rješenje uvodnog problema razmjernim rezom

Postoji bolje rješenje od razmjernog reza, a to je normalizirani rez. Shvatimo li podatke kao slučajni uzorak iz neke distribucije, možemo se pitati hoće li uzorkovanjem sve više podataka rezultati algoritma grupiranja konvergirati ka *pravoj* particiji? Za normalizirane rezove odgovor je da, što ne možemo reći za metode minimalnog i razmjernog reza. Za njih se može pokazati kako je moguće da konvergiraju trivijalnim rješenjima kao što su izolirane točke ili da ne konvergiraju uopće.

Ponovno, pronalazak najbolje biparticije formuliramo kao diskretni optimizacijski problem.

$$(V_1^*, V_2^*) = \arg\min\{R(V_1, V_2) : V_1 \cap V_2 = \phi, V_1 \cup V_2 = V\}$$

## 3.3 Normalizirani rez

Za razliku od razmjernog reza koji kao veličinu skupa vrhova uzima njegov kardinalitet, normalizirani rez uzima u obzir težine bridova unutar skupa vrhova. Također, normalizirani rez ima i svojstvo dualnosti, a kasnije ćemo dualnu i primarnu zadaću implementirati različitim algoritmima radi usporedbe.

**Definicija 3.9.** *Za $A \subseteq V$ stupanj skupa vrhova $A$ definiramo kao*

$$D(A) := \sum_{\substack{i \in A \\ j \in V}} w_{ij}.$$

**Definicija 3.10.** *Normalizirani rez biparticije $\Gamma_V^2$ definiramo kao*

$$N(V_1, V_2) := \frac{C(V_1, V_2)}{D(V_1)} + \frac{C(V_1, V_2)}{D(V_2)}.$$

Pronalazak optimalne biparticije postaje već viđen optimizacijski problem.

$$(V_1^*, V_2^*) = \arg\min\{N(V_1, V_2) : V_1 \cap V_2 = \phi, V_1 \cup V_2 = V\}$$

Normalizirani rez također rješava uvodni problem koji je motivirao pronalazak boljih algoritama od minimalnog reza. Razmjerni i normalizirani rezovi traže balansirane grupacije. U slučaju razmjernog reza, balansiranim grupacijama smatramo one sličnog kardinaliteta particija, a u slučaju normaliziranog reza one sličnog stupnja. Drugim riječima, funkcije cilja postižu minimum ukoliko su kardinaliteti, odnosno stupnjevi slični tako da sume

$$\frac{1}{|V_1|} + \frac{1}{|V_2|} \text{ i } \frac{1}{D(V_1)} + \frac{1}{D(V_2)}$$

nemaju premale sumadne. Ovo možemo vizualizirati promatrajući funkciju $x \longmapsto \frac{1}{x} + \frac{1}{1-x}$ na intervalu $(0, 1)$, koja izgledom podsjeća na parabolu te ima minimum upravo u $\frac{1}{2}$.

## 3.4 $K$-particioniranje grafa

Proširimo dosadašnje pojmove na $K$-particije.

Neka je $\{V_1, \ldots, V_K\}$ $K$-particija $\Gamma_V^K$ skupa $V$ te $V_i^c$ komplement skupa $V_i$.

**Definicija 3.11.** *Rez $K$-particije $\Gamma_V^K$ definiramo kao*

$$C(V_1, \dots, V_K) := \sum_{i=1}^{K} C(V_i, V_i^c).$$

Ovo je zaista poopćenje reza kako je u slučaju biparticije $V_1^c = V_2$.

**Definicija 3.12.** *Razmjerni rez $K$-particije $\Gamma_V^K$ definiramo kao*

$$R(V_1, \dots, V_k) := \sum_{i=1}^{K} \frac{C(V_i, V_i^c)}{|V_i|}.$$

**Definicija 3.13.** *Normalizirani rez $K$-particije $\Gamma_V^K$ definiramo kao*

$$N(V_1, \dots, V_K) := \sum_{i=1}^{K} \frac{C(V_i, V_i^c)}{D(V_i)}.$$

Optimizacijski problemi pronalaska najbolje $K$-particije minimalnim, odnosno razmjernim rezom potpuni su analogon pronalasku optimalne biparticije i ovdje izostavljamo njihov formalni zapis. Ono što ćemo prikazati su primarna i dualna zadaća normaliziranog reza.

Problem pronalaska optimalne $K$-particije metodom normaliziranog reza formuliramo na sljedeći način.

$$(V_1^*, \dots, V_K^*) = \arg\min\{N(V_1, \dots, V_K) : V_i \cap V_j = \phi \ \forall i \neq j, \bigcup_{i=1}^{K} V_i = V\}$$

Minimizacijski problem smatrati ćemo primarnom zadaćom normaliziranog reza $K$-particije $\Gamma_V^K$.

Prisjetimo se da je ono što minimiziramo nesličnost među podacima u različitim grupacijama. Pokazuje se kako je u slučaju normaliziranih rezova ovaj problem ekvivalentan maksimizaciji sličnosti podataka unutar grupacija. Vrijedi:

$$N(V_1, \dots, V_K) = \sum_{i=1}^{K} \frac{C(V_i, V_i^c) \pm C(V_i, V_i)}{D(V_i)} = \sum_{i=1}^{K} \frac{D(V_i) - C(V_i, V_i)}{D(V_i)} = K - \sum_{i=1}^{K} \frac{C(V_i, V_i)}{D(V_i)}.$$

Minimizacija normaliziranog reza $K$-particije ekvivalentna je maksimizaciji sume

$$\sum_{i=1}^{K} \frac{C(V_i, V_i)}{D(V_i)}$$

po svim $K$-particijama.

**Definicija 3.14.** *Asocijativnost $K$-particije $\Gamma_V^K$ definiramo kao*

$$A(V_1, \dots, V_K) = \sum_{i=1}^{K} \frac{C(V_i, V_i)}{D(V_i)}.$$

Ekvivalentan problem pronalaska optimalne $K$-particije, odnosno dualna zadaća normaliziranog reza $K$-particije dana je sa

$$(V_1^*, \dots, V_K^*) = \arg\max\{A(V_1, \dots, V_K) : V_i \cap V_j = \phi \ \forall i \neq j, \bigcup_{i=1}^{K} V_i = V\}$$

Iako znamo da optimalna $K$-particija zaista postoji (iz diskretne prirode problema), metode razmjernog reza i normaliziranog reza su NP-teški optimizacijski problemi. Spektralno klasteriranje predstavlja neprekidnu relaksaciju prethodno diskretnog optimizacijskog problema. Relaksacijom razmjernog reza dolazimo do *nenormaliziranog spektralnog klasteriranja*, a normaliziranog reza do *normaliziranog spektralnog klasteriranja*.

## 3.5 Laplacijan grafa i njegova svojstva

Prije nego što opišemo algoritme nenormaliziranog i normaliziranog spektralnog klasteriranja potrebni su nam još neki pojmovi iz teorije grafova. Neka je $(V, W)$ neusmjeren težinski graf.

**Definicija 3.15.** *Matricu $L := D - W$ nazivamo Laplacijanom grafa.*

Kod spektralne analize unutar teorije grafova svojstvene vrijednosti često su poredane u rastućem poretku. Razlog je taj što su one najmanje upravo najbitnije i njih stavljamo na početak. Situacija sa SVD dekompozicijom je iz istog razloga obratna. Prvih $k$ svojstvenih vrijednosti značiti će $k$ najmanjih svojstvenih vrijednosti.

Neka od osnovnih svojstava Laplacijana grafa navodimo bez dokaza.

**Propozicija 3.16.** *Laplacijan grafa $L \in M_n(\mathbb{R})$ zadovoljava sljedeća svojstva:*

1. *Za svaki vektor $x \in \mathbb{R}^n$ vrijedi $x^\tau L x = \frac{1}{2}\sum_{i,j}^{n} w_{ij}(x_i - x_j)^2$.*
2. *L je simetrična i pozitivno semi-definitna matrica.*
3. *svojsveni par pridružen najmanjoj svojstvenoj vrijednosti je $(0, \mathbb{1}_n)$.*
4. *L ima n nenegativnih realnih svojstvenih vrijednosti $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$.*

$\square$

Laplacijan grafa ne ovisi o dijagonalnim elementima težinske matrice $W$. Oni bi se poništili sa stupnjevima vrhova iz matrice $D$.

**Propozicija 3.17.** *Neka je sa $k$ označena kratnost svojstvene vrijednosti $0$ Laplacijana $L$ grafa $(V, W)$. Tada je $k$ jednak broju povezanih komponenata grafa $(V, W)$. Označimo li povezane komponente sa $A_1$, $A_2, \ldots, A_k$ svojstveni vektori su tada indikatorski vektori $\mathbb{1}_{A_i}$*[1] *za $i = 1, \ldots, k$.*

$\square$

Uvodimo još dva Laplacijana grafa koja se prirodno pojave kod izučavanja problema particioniranja grafa.

**Definicija 3.18.** *Matricu $L_{sym} := D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$ nazivamo simetričnim Laplacijanom grafa.*

**Definicija 3.19.** *Matricu $L_{rw} := D^{-1} L$ nazivamo normaliziranim Laplacijanom grafa.*

$L_{sym}$ je simetrična matrica, dok je $L_{rw}$ usko povezana sa slučajnom šetnjom. Njihova svojstva slična su onima Laplacijana grafa.

**Propozicija 3.20.** *Matrice $L_{sym}$, $L_{rw} \in M_n(\mathbb{R})$ imaju sljedeća svojstva:*

1. *Za svaki vektor $x \in \mathbb{R}^n$ vrijedi $x^\tau L_{sym} x = \frac{1}{2} \sum_{i,j}^{n} w_{ij} (\frac{x_i}{\sqrt{d_i}} - \frac{x_j}{\sqrt{d_j}})^2$.*

2. *$(\lambda, u)$ je svojstveni par od $L_{rw}$ ako i samo ako je $(\lambda, D^{\frac{1}{2}} u)$ svojstveni par od $L_{sym}$.*

3. *$(\lambda, u)$ je svojstveni par od $L_{rw}$ ako i samo ako je rješava generalzirani svojsveni problem $Lu = \lambda D u$.*

4. *$(0, \mathbb{1}_n)$ je svojstveni par od $L_{rw}$. $(0, D^{\frac{1}{2}} \mathbb{1}_n)$ je svojstveni par od $L_{sym}$.*

5. *$L_{sym}$ i $L_{rw}$ su pozitivno semi-definitne i imaju $n$ realnih svojstvenih vrijednosti $0 = \lambda_1, \ldots, \lambda_n$.*

$\square$

Lako se pokaže da su matrice $L_{sym}$ i $L_{rw}$ slične, što znači da imaju jednak spektar. Kao i u slučaju sa Laplacijanom grafa, kratnost svojstvene vrijednost $0$ u predstavlja broj povezanih komponenata grafa.

**Propozicija 3.21.** *Neka je sa $k$ označena kratnost svojstvene vrijednosti $0$ normaliziranog Laplacijana $L_{sym}$ i $L_{rw}$ grafa $(V, W)$. Tada je $k$ jednak broju povezanih komponenata grafa $(V, W)$. Označimo li povezane komponente sa $A_1$, $A_2, \ldots, A_k$ svojstveni vektori matrice $L_{rw}$ su tada dani sa $\mathbb{1}_{A_i}$ za $i = 1, \ldots, k$, a matrice $L_{sym}$ dani su sa $D^{\frac{1}{2}} \mathbb{1}_{A_i}$*

$\square$

[1] $\mathbb{1}_{A_i} \in M_{n,1}$ te $(\mathbb{1}_{A_i})_j = 1$ ukoliko je vrh $j$ unutar $A_i$, $0$ inače

# 4 Algoritmi klasteriranja

Kako bismo predstavili algoritme spektralnog klasteriranja bazirane na matricama $L$, $L_{sym}$ i $L_{rw}$ potrebni su nam i osnovni algoritmi klasteriranja podataka. Jedan od njih je $k$-means algoritam.

## 4.1 Iterativni $k$-means algoritam

Neka su $a_1, \ldots, a_n \in \mathbb{R}^m$ podaci (dokumenti) koje želimo klasterirati. Tražimo optimalnu particiju $\pi = (\pi_1, \ldots \pi_k)$. Neka je $s_i$ kardinalitet od $\pi_i = \{a_1^{(i)}, \ldots, a_{s_i}^{(i)}\}$. Optimalna particija minimizira funkciju cilja

$$Q(\pi) = \sum_{i=1}^{k} \sum_{s=1}^{s_i} \|a_s^{(i)} - m_i\|_2^2 \,, \quad \text{gdje je } m_i = \frac{1}{s_i} \sum_{s=1}^{s_i} a_s^{(i)}.$$

Dakle, minimiziramo $Q$ po svim $k$-particijama skupa $\{\, a_1, \ldots a_n \,\}$. Predstavimo iterativni $k$-means algoritam za klasteriranje .

**Iterativni $k$-means**

**Ulaz:** Dokumenti $\{\, a_1, \ldots a_n \,\}$, broj klastera k

1. Zadaj početnu particiju $\pi_j^{(0)}, j = 1, \ldots, k$ i izračunaj početne centroide $m_j^{(0)}, j = 1, \ldots, k.t = 0$.
2. Svaki dokument $a_i$, $(i = 1, \ldots, n)$ neka potraži sebi najbliži centroid i pridruži se novom skupu
$$\pi_j^{(t+1)} = \left\{ a \in \{a_1, \ldots, a_n\} : \left\| a - m_j^{(t)} \right\|_2 = \min_{\ell = 1, \ldots, k} \|a - m_\ell\|_2 \right\}$$
3. Izračunaj centroide $m_j^{(t+1)}$ skupova $\pi_j^{(t+1)}, j = 1, \ldots, k$.
4. Provjeri kriterij zaustavljanja. Ako je zadovoljen, $\pi_j^{(t+1)}, j = 1, \ldots, k$ je tražena particija; STOP. Inače, $t = t + 1$ i idi na Korak 2.

**Izlaz:** Završna particija $\pi$ i pripadni skup centroida $m = \{\, m_1, \ldots, m_k \,\}$

Može se pokazati kako je funkcija cilja manja, odnosno optimalnija, svakom novom iteracijom. Novi podatak pridružuje se klasteru čijem je centroidu najbliži.

Problem se može formulirati i matrično pomoću vektorskog modela podataka.

## 4.2 Matrična formulacija $k$-means algoritma

Dokumnte organiziramo kao stupce matrice $A$, dakle $A = (a_1, \ldots, a_n)$. Podatke ćemo permutirati matricom permutacija $\Pi$ tako da je prvi klaster u prvih $s_1$ stupaca matrice $A$, drugi u idućih $s_2$ stupaca itd.

$$A\Pi = (A_1, \ldots, A_k), \quad A_i = (a_1^{(i)}, \ldots, a_{s_i}^{(i)}), \quad i = 1, \ldots, k.$$

Centroid $m_i$ tada možemo dobiti tako da zbrojimo sve stupce u pripadnom bloku $A_i$ te pripadni vektor stupac podijelimo sa brojem elemenata unutar $A_i$. Dakle, $m_i = \frac{1}{s_i} A_i \mathbb{1}_{s_i}$. Funkciju cilja želimo izraziti pomoću matričnih operacija. Vrijedi

$$\sum_{s=1}^{s_i} \left\| a_s^{(i)} - m_i \right\|_2^2 = \left\| A_i - m_i \mathbb{1}_{s_i}^\tau \right\|_F^2 = \left\| A_i - \frac{1}{s_i} A_i \mathbb{1}_{s_i} \mathbb{1}_{s_i}^\tau \right\|_F^2 = \left\| A_i \left( I - \frac{1}{s_i} \mathbb{1}_{s_i} \mathbb{1}_{s_i}^\tau \right) \right\|_F^2,$$

gdje je

$$\left( I - \frac{1}{s_i} \mathbb{1}_{s_i} \mathbb{1}_{s_i}^\tau \right)^2 = \left( I - \frac{1}{s_i} \mathbb{1}_{s_i} \mathbb{1}_{s_i}^\tau \right) = \left( I - \frac{1}{s_i} \mathbb{1}_{s_i} \mathbb{1}_{s_i}^\tau \right)^\tau$$

ortogonalni projektor. Nadalje,

$$\begin{aligned} \left\| A_i \left( I - \frac{1}{s_i} \mathbb{1}_{s_i} \mathbb{1}_{s_i}^\tau \right) \right\|_F^2 &= \operatorname{tr} \left( A_i \left( I - \frac{1}{s_i} \mathbb{1}_{s_i} \mathbb{1}_{s_i}^\tau \right) \left( I - \frac{1}{s_i} \mathbb{1}_{s_i} \mathbb{1}_{s_i}^\tau \right) A_i^\tau \right) \\ &= \operatorname{tr} \left( A_i \left( I - \frac{1}{s_i} \mathbb{1}_{s_i} \mathbb{1}_{s_i}^\tau \right) A_i^\tau \right) = \operatorname{tr} \left( \left( I - \frac{1}{s_i} \mathbb{1}_{s_i} \mathbb{1}_{s_i}^\tau \right) A_i^\tau A_i \right) \\ &= \operatorname{tr} \left( A_i^\tau A_i \right) - \frac{1}{s_i} \operatorname{tr} \left( \mathbb{1}_{s_i} \mathbb{1}_{s_i}^\tau A_i^\tau A_i \right) = \operatorname{tr} \left( A_i^\tau A_i \right) - \frac{1}{s_i} \operatorname{tr} \left( e^\tau A_i^\tau A_i \mathbb{1}_{s_i} \right) \\ &= \operatorname{tr} \left( A_i^\tau A_i \right) - \left( \frac{\mathbb{1}_{s_i}}{\sqrt{s_i}} \right)^\tau A_i^\tau A_i \left( \frac{\mathbb{1}_{s_i}}{\sqrt{s_i}} \right) \end{aligned}$$

.

Ostaje sumirati izraz po svim klasterima i time dobivamo izraz za funkciju cilja:

$$Q(\pi) = \sum_{i=1}^{k} \left( \operatorname{tr} \left( A_i^\tau A_i \right) - \left( \frac{\mathbb{1}_{s_i}}{\sqrt{s_i}} \right)^\tau A_i^\tau A_i \left( \frac{\mathbb{1}_{s_i}}{\sqrt{s_i}} \right) \right)$$

Prvi izraz u sumi je zbroj tragova dijagonalnih blokova Gramove matrice $(A\Pi)^\tau (A\Pi)$. Međutim, to je upravo i trag matrice $(A\Pi)^\tau (A\Pi)$ kako trag uzima u obzir samo dijagonalne elemente. Drugi izraz je kvadratna forma koju bismo također htjeli zapisati u obliku matričnog množenja. To je moguće uz blok dijagonalnu matricu $X$ definiranu kao

$$X = \begin{bmatrix} \frac{1}{\sqrt{(s_1)}} \mathbb{1}_{s_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{(s_k)}} \mathbb{1}_{s_k} \end{bmatrix}.$$

Definiramo familiju $\mathcal{F}_{n,k}$ kao familiju svih matrica tipa $n \times k$ čiji su elementi 0 ili 1 te svaki redak sadrži točno jedan netrivijalan element. Ovakve matrice nazivati ćemo particijskim matricama. Primijetimo da je matrica $X$ particijska matrica nakon normiranja stupaca. Obe matrice na jedinstveno određuju particiju. Vrijedi

$$A\Pi X = \left( \frac{1}{\sqrt(s_1)} A_1 \mathbb{1}_{s_1} \dots \frac{1}{\sqrt(s_k)} A_k \mathbb{1}_{s_k} \right)$$

te Gramova matrica $(A\Pi X)^\tau A\Pi X$ ponovno na dijagonali sadrži tražene elemente. Ovoga puta će to biti brojevi kako su blokovi vektor stupci te se umnožak svede na kvadratnu formu. Svakako je trag gornje matrice drugi sumand u izrazu za funkciju cilja. Označimo sa $\tilde{X}$ matricu $\Pi X$. $\tilde{X} = \Pi X$ je ortonormalna $\left(\tilde{X}^\tau \tilde{X} = I_k\right)$ te u svakom retku ima točno jedan netrivijalan element. Ona jednistveno određuje particiju kao i matrica $X$, dakle permutacijom redaka nismo izgubili ovo svojstvo. Izraz za funkciju cilja postaje

$$\sum_{i=1}^{k} \left( \frac{e}{\sqrt{s_i}} \right)^\tau A_i^\tau A_i \left( \frac{e}{\sqrt{s_i}} \right) = \operatorname{tr}\left(X^\tau \Pi^\tau A^\tau A \Pi X\right) = \operatorname{tr}\left(\tilde{X}^\tau A^\tau A \tilde{X}\right)$$

$$Q(\pi) = \operatorname{tr}\left(A^\tau A\right) - \operatorname{tr}\left(\tilde{X}^\tau A^\tau A \tilde{X}\right)$$

i minimiziramo ga po svim $\tilde{X}$ sa strukturom particijske matrice nakon normiranja stupaca.

## 4.3 Relaksirani $k$-means algoritam

U primjenama ukupan broj dokumenata $n$ može biti jako velik te su dokumentiv nekom $m$-dimenzionalnom prostoru s potencijalno velikim $m$, tako da i za relativno mali $k$ trivijalna provjera po svim $k$-particijama nije praktično izvediva. Metoda $k$ sredina ima dosta nedostataka, primjerice problema sa sporom konvergencijom. Iako imamo kombinatornu optimizaciju te *samo* konačno mnogo slučajeva za provjeriti, može se pokazati kako je problem NP-težak. U tu svrhu problem ćemo relaksirati na način da umjesto matrica $\tilde{X}$, koje su imale specifičnu diskretnu strukturu, promatramo proizvoljnu ortonormalnu matricu $X \in M_{n,k}(\mathbb{R})$. Razlog je što neprekidna verzija optimizacijskog problema ima eksplicitno rješenje.

**Teorem 4.1.** ***(Ky Fan)*** *Neka je $H \in \mathbb{R}^{n \times n}$ simetrična matrica $(H = H^\tau)$ sa svojstvenim vrijednostima $\lambda_1 \geq \dots \geq \lambda_n$ i pripadnom matricom svojstvenih vektora $U = (u_1, \dots, u_n)\,, Hu_i = \lambda_i u_i, i = 1, \dots, n; U^\tau U = I_n$. Tada je*

$$\max_{\substack{X \in \mathbb{R}^{n \times k} \\ X^\tau X = I_k}} \operatorname{tr}\left(X^\tau H X\right) = \lambda_1 + \ldots + \lambda_k.$$

*Za $\lambda_k > \lambda_{k+1}$ se optimalne matrice $X$ mogu opisati kao elementi skupa*

$$\{(u_1, \dots, u_k)Q, Q \text{ ortogonalna}\}$$

□

Donja ograda za funkciju cilja slijedi iz Ky Fanovog teorema.

$$\min_{\pi\text{-part.}} Q(\pi) \geq \operatorname{tr}(A^\tau A) - \max_{X \in \mathbb{R}^{n\times k} X^\tau X = I_k} \operatorname{tr}(X^\tau A^\tau A X)$$

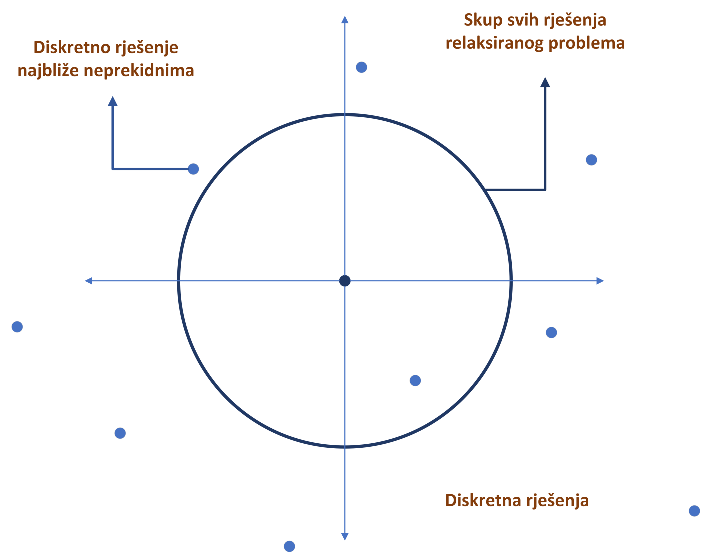
$$= \sum_{i=1}^{\min(m,n)} \sigma_i^2 - \sum_{i=1}^{k} \sigma_i^2 = \sum_{i=k+1}^{\min(m,n)} \sigma_i^2 = \|A\|_F^2 - \sum_{i=1}^{k} \sigma_i^2,$$

gdje su $\sigma_1 \geq \cdots \geq \sigma_{\min(m,n)}$ singularne vrijednosti od $A$.

Simetrična matrica koju promatramo je $A^\tau A$. Srećom, njene svojstvene vrijednosti upravo su singularne vrijednosti matrice $A = U\Sigma V^\tau$, a svojstveni vektori dani su matricom $V$ što je moguće direktno provjeriti. Označimo li matricu prvih k desnih singularnih vektora $(v_1, \dots, v_k)$ matrice $A$ sa $V_k$ po teoremu 4.1 sva rješenja relaksiranog problema pronalaska optimalne particije dana su sa $V_k Q$, $Q \in M_k(\mathbb{R})$ ortogonalna.

Bitno za naglasiti je da smo računanje svojstvenih vrijednosti matrice $A^\tau A$ sveli na računanje singularnih vrijednosti matrice $A$, odnosno na problem SVD dekompozicije. Samo računanje matrice $A^\tau A$ zadaje poteškoće ukoliko primjerice $A$ nije rijetka matrica. Također, za računanje (dijela) SVD dekompozicije postoje efikasni algoritmi.

Skup rješenja relaksiranog problema (vrlo vjerojatno) ne sadrži matricu sa diskretnom struktorom iz koje bismo iščitavali particije. Ipak, razumno je pretpostaviti skupu svih rješenja neprekidnog problema postoji vrlo bliska matrica particije $\tilde{X}$. Kao vizualizaciju diskretna rješenja možemo smatrati točkama u ravnini, dok neprekidna rješenja smatramo kružnicom kako su ortogonalne matrice reprezentacije rotacija (i refleksija). Primjer je dan slikom 6.




Slika 6: Diskretna i neprekidna rješenje

Ostaje riješiti problem pronalaska najbliže diskretne matrice onoj dobivenoj rješavanjem relaksiranog problema kako bismo mogli iščitati particije.
Zahtjevi su:

Pstoje dva prstupa rješavanju ovog problema:

1. Pronaći klastere među retcima dobivene matrice (primjerice, pomoću $k$-meansa)

2. $QR$ faktorizacija s pivotiranjem

Koristiti ćemo prvi pristup kod spektralnog klasteriranja. Razlog je što će podaci biti dovoljno dobro separirani i $k$-means algoritam će brzo i lako detektirati klastere.

# 5 Algoritmi spektralnog klasteriranja

Matrično formulirajmo problem razmjernog reza. Pronalazak optimalne $K$-particije ekvivalentan je pronalasku optimalne particijske matrice.

## 5.1 Razmjerni rez

Neka je $X \in \mathcal{F}_{n,k}$ neka particijska matrica, a $X_1, \ldots, X_k$ njeni stupci. Iz svojstava Laplacijana grafa slijedi da je

$$X_l^\tau L X_l = \frac{1}{2}\sum_{i,j}^{n} w_{ij}(X_{li} - X_{lj})^2 = \frac{1}{2}C(V_l, V_l^c).$$

Primijetimo da normiranjem stupaca particijske matrice $X$ gornjima računom dobivamo $i$-ti sumand u razmjernom rezu $K$-particije, do na kardinalitet particije. Neka je $\tilde{X}$ particijska matrica nakon normiranja stupaca. Tada je

$$\tilde{X_l}^\tau L \tilde{X_l} = \frac{1}{2}\sum_{i,j}^{n} w_{ij}(\tilde{X_{li}} - \tilde{X_{lj}})^2 = \frac{1}{2}\frac{C(V_l, V_l^c)}{|V_l|}.$$

Razmjerni rez $K$-particije moguće je prikazari kao

$$R(V_1, \ldots, V_k) = \frac{1}{2}\operatorname{tr}(\tilde{X}^\tau L \tilde{X}).$$

**Napomena 5.1.** *Broj $\frac{1}{2}$ neće utjecati optimizacijske zadaće i njega kao i do sada ispuštamo. Ovdje napominjemo da je moguće eliminirati $\frac{1}{2}$ iz izraza definirajući mjeru $C(A, B)$ kao*

$$C(A, B) := \frac{1}{2}\sum_{\substack{i\in A\\ j\in B}} w_{ij}.$$

*Ovo je prirodnija definicija kako ne brojimo bridove dvostruko i korištena je u [1].*

Problem pronalaska optimalne particijske matrice $X^*$, zadrživši gornje oznake, formuliramo kao:

$$X^* = \underset{X\in\mathcal{F}_{n,k}}{\arg\min}\ \operatorname{tr}(\tilde{X}^\tau L \tilde{X})$$

Relaksacija ovog problema znači eliminirati zahtjev na specifičnu strukturu matrice $\tilde{X}$ koja je ortonormalna. Dakle, relaksacijom problema dozvoljavamo sve ortonormalne matrice. Ovaj problem rješava varijanta Ky Fanovog teorema, a oboje su posljedica Rayleight-Ritzovog teorema. Također, teoremi garantiraju da je optimizacijski problem dobro postavljen.

Algoritam vezan uz Laplacijan grafa $L$ nazivati ćemo nenormalizirano spektralno klasteriranje, a ostale normalizirano spektralno klasteriranje.

**Nenormalizirano spektralno klasteriranje**

**Ulaz:** Težinska matrica $W$, broj klastera $k$

1. Izračunaj Laplacijan grafa $L$.
2. Izračunaj $u_1, \ldots, u_k$, prvih $k$ svojstvenih vektora od $L$.
3. Neka matrica $U \in M_{n,k}(\mathbb{R})$ sadrži redom $u_1, \ldots, u_k$ kao stupce.
4. Za $i = 1, \ldots, n$ neka $y_i$ bude $i$-ti redak od matrice $U$
5. Klasteriraj vektore $y_i \in \mathbb{R}^k$, $i = 1, \ldots, n$ u klastere $C_1, \ldots, C_k$

**Izlaz:** Skupovi $V_1, \ldots, V_k$ td. $V_i = \{\, j : y_j \in C_i \,\}$.

Alternativno, kao izlaze algoritama možemo uzimati i particijsku matricu ili vektor koji indicira klaster podatka na pojedinom indeksu brojem 1-$k$, a potrebno je samo modificirati zadnji korak.

## 5.2 Normalizirani rez

Pokušajmo sličnu svar izvesti za normalizirane rezove. Neka je $X \in \mathcal{F}_{n,k}$ ponovno particijska matrica.

$$N(V_1, \ldots, V_K) = \sum_{i=1}^{K} \frac{C(V_i, V_i^c)}{D(V_i)} = \sum_{i=1}^{K} \frac{X_i^\tau L X_i}{X_i^\tau D X_i}.$$

Normalizirani rez $K$-particije želimo prikazati pomoću traga neke matrice $H^\tau L H$. Ukoliko $H$ ima dovoljno dobru strukturu, relaksirani problem rješavamo biti će rješiv alatima spektralne teorije. Zapis $H^\tau L H$ je moguće postići dijeljenjem $i$-tog stupca particijske matrice $X$ sa $\sqrt{D(V_i)}$. Označimo takvu matricu sa $\tilde{X}$. Vrijedi:

$$\sum_{i=1}^{K} \frac{X_i^\tau L X_i}{X_i^\tau D X_i} = \sum_{i=1}^{K} \frac{X_i^\tau L X_i}{D_i} = \sum_{i=1}^{K} \frac{X_i^\tau}{\sqrt{D_i}} L \frac{X_i}{\sqrt{D_i}} = \sum_{i=1}^{K} \tilde{X_i}^\tau L \tilde{X}_i.$$

Ovakva matrica $\tilde{X}$ nema strukturu nalik ortonormalnoj:

$$\tilde{X}^\tau \tilde{X} = \begin{bmatrix} D(V_1) & & \\ & \ddots & \\ & & D(V_k) \end{bmatrix}.$$

Rješenje ovog problema postaje očito nakon što izrazimo matricu $\tilde{X}$ pomoću particijske matrice i matrice stupnjeva.

Naime, kako je

$$\tilde{X} = X(X^\tau DX)^{-\frac{1}{2}},$$

vrijedi da je

$$\tilde{X}^\tau D\tilde{X} = (X^\tau DX)^{-\frac{1}{2}} X^\tau DX(X^\tau DX)^{-\frac{1}{2}} = I_k$$

Optimizacijski problem pronalaska najbolje particijske matrice postaje

$$X^* = \underset{X\in\mathcal{F}_{n,k}}{\arg\min}\ \mathrm{tr}\left((X^\tau DX)^{-\frac{1}{2}})X^\tau LX(X^\tau DX)^{-\frac{1}{2}}\right).$$

Relaksacija problema pronalaska optimalne particijske matrice ima formu:

$$X^* = \underset{H^\tau DH=I_k}{\arg\min}\ \mathrm{tr}(H^\tau LH)$$

Gornji uvijet predstavlja ortonormiranost vektora stupaca matrice $H$ u uniratnom prostoru $(\mathbb{R}^n, \langle\ ,\ \rangle_D)$ , gdje je $\langle x, y\rangle_D = y^\tau Dx$. Cilj je pokazati kako je to ekvivalentno sa poopćenim problemom svojstvenih vrijednosti te povezati problem optimizacije sa simetričnim i normaliziranim Laplacijanom.

Primijetimo kako izraz $\mathrm{tr}(H^\tau LH)$ možemo zapisati kao

$$\mathrm{tr}(H^\tau D^{\frac{1}{2}} D^{-\frac{1}{2}} LD^{-\frac{1}{2}} D^{\frac{1}{2}} H)$$

te uz zamijenu varijabli $T = D^{\frac{1}{2}}H$ kao izraz koji minimiziramo dobivamo

$$\mathrm{tr}(T^\tau L_{sym}T),$$

uz uvjet $T^\tau T = I_k$.

Sada je optimizacijski problem moguće relaksirati, a bit će dobro zadan kao posljedica varijacije Rayleight-Ritzovog teorema. Iz propozicija 3.16 i 3.20 slijedi kako Laplacijan grafa i normalizirani Laplacijan grafa kao svojstvene parove imaju $(0, \mathbb{1}_n)$, dok je on za simetrični Laplacijan $(0, D^{\frac{1}{2}}\mathbb{1}_n)$. Iz ovog razloga sljedeći algoritam zahtijeva dodatni korak normiranja redaka i on je posebno naglašen.

**Normalizirano spektralno klasteriranje (Ng, Jordan, Weiss)**

**Ulaz:** Težinska matrica $W$, broj klastera k

1. Izračunaj normalizirani Laplacijan grafa $L_{sym}$.
2. Izračunaj $u_1, \ldots, u_k$, prvih $k$ svojstvenih vektora matric $L_{sym}$.
3. Neka matrica $U \in M_{n,k}(\mathbb{R})$ sadrži $u_1, \ldots, u_k$ kao stupce.
4. **Formiraj matricu $Z \in \mathbb{R}^{n \times k}$ normiranjem redaka matrice U**
5. Za $i = 1, \ldots, n$ neka $y_i$ bude $i$-ti redak od matrice $Z$
6. Klasteriraj vektore $y_i \in \mathbb{R}^k$, $i = 1, \ldots, n$ u klastere $C_1, \ldots, C_k$

**Izlaz:** Skupovi $V_1, \ldots, V_k$ td. $V_i = \{ j : y_j \in C_i \}$.

Razlog uvođenja matrice $Z$ krije se i u teoriji perturbacije. U idealnom slučaju svojstveni vektori matrica $L$ i $L_{rw}$ su indikatorski vektori, što nije slučaj za matricu $L_{sym}$. Čak i u idealnoj situaciji svojsveni vektori su $D^{\frac{1}{2}}\mathbb{1}_{A_i}$ i ovo stvara poteškoće koje se mogu značajno poboljšati normiranjem redaka matrice $U$. Više o ovom može se pronaći u [5].

**Napomena 5.2.** *k-means algoritam može se koristiti u koraku klasteriranja vektora $y_i$ te zbog svojstava Laplacijana grafa klastere ćemo lako detektirati.*

Iz propozicije 3.20 vidimo kako su matrice $L_{sym}$ i $L_{rw}$ usko povezane. Pokazuje se kako dodatan korak normiranja redaka nije potreban kod matrice $L_{rw}$. Predstavimo i algoritam normaliziranog spektralnog klasteriranja baziran na poopćenom problemu svojstvenih vrijednosti.

**Normalizirano spektralno klasteriranje (Shi, Malik)**

**Ulaz:** Težinska matrica $W$, broj klastera k

1. Izračunaj Laplacijan grafa $L$.
2. Izračunaj $u_1, \ldots, u_k$, prvih $k$ generaliziranih svojstvenih vektora generalziranog svojsvenog problem $Lu = \lambda Du$.
3. Neka matrica $U \in M_{n,k}(\mathbb{R})$ sadrži $u_1, \ldots, u_k$ kao stupce.
4. Za $i = 1, \ldots, n$ neka $y_i$ bude $i$-ti redak od matrice $U$
5. Klasteriraj vektore $y_i \in \mathbb{R}^k$, $i = 1, \ldots, n$ u klastere $C_1, \ldots, C_k$

**Izlaz:** Skupovi $V_1, \ldots, V_k$ td. $V_i = \{\, j : y_j \in A_i \,\}$.

### 5.3 Dualna zadaća normaliziranog reza

Kod dualne zadaće mjeru $C(V_i, V_i)$ možemo prikazati kao

$$C(V_i, V_i) = X_i^\tau W X_i.$$

Time asocijativnost postaje

$$A(V_1, \ldots, V_K) = \sum_{i=1}^{K} \frac{C(V_i, V_i)}{D(V_i)} = \sum_{i=1}^{K} \frac{X_i^\tau W X_i}{X_i^\tau D X_i} = \mathrm{tr}((X^\tau D X)^{-\frac{1}{2}} X^\tau W X (X^\tau D X)^{-\frac{1}{2}})$$

Kao i ranije, preferiramo zapis u obliku $\mathrm{tr}(H^\tau W H)$ za neku matricu $H$ koja ima pogodnu strukturu. Ona je ponovno $H = X(X^\tau D X)^{-\frac{1}{2}}$ i u skalarnom produktu generiranom matricom $D$ vrijedi da su stupci matrice $H$ ortonormirani, tj. $H^\tau D H = I_k$.

Relaksacija problema pornalaska optimalne particijske matrice ima formu

$$X^* = \underset{H^\tau D H = I_k}{\arg\max} \ \mathrm{tr}(H^\tau W H).$$

Neka je i $T$ kao ranije matrica $T = D^{\frac{1}{2}} H$. Time optimizacijski problem postaje

$$X^* = \underset{T^\tau T = I_k}{\arg\max} \ \mathrm{tr}(T^\tau D^{-\frac{1}{2}} W D^{-\frac{1}{2}} T).$$

Ranije smo sličan minimizacijski problem rješavali pomoću Rayleigh-Ritzovog teorema, odnosno poopćenih svojstvenih vrijednosti. Sada Ky Fanov teorem daje rješenje relaksiranog problema. Uzevši u obzir činjenicu da je $\mathrm{tr}(Q^\tau A Q) = \mathrm{tr}(A)$ za sve ortogonalne matrice

$Q$, možemo opisati skup rješenje relaksiranog problema dualne zadaće normaliziranih rezova kao

$$S = \{\ (v_1, \dots, v_k)Q : Q^\tau Q = QQ^\tau = I_k\ \},$$

pri čemu su $v_1, \dots, v_k$ svojstveni vektori pridruženi najvećim svojstvenim vrijednostima $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$ matrice $D^{-\frac{1}{2}}WD^{-\frac{1}{2}}$.

Nakon što odaberemo neko rješenje $V \in S$, odnosno pripadni $Q$, iz njega trebamo iščitati particijsku matricu. Ovaj problem bio je rješavan klasteriranjem redaka matrice $V$ (uz eventualno normiranje), iako smo mogli koristiti i neki alternativan način poput $QR$ faktorizacije s pivotiranjem kao u relaksiranom $k$-means algoritmu. Sada želimo pronaći neprekidno rješenje u skupu $S$ koje će diskretizacijom biti dovoljno blizu globalnom optimalnom rješenju originalnog diskretnog optimizacijskog problema.

Za početak matricu $T$ dobili smo kao $T = D^{\frac{1}{2}}H$, a matrica $H$ je jedinstveno određivala particiju. $V$ možemo smatrati istog oblika kao $T$ pa je prirodno na nju djelovati sa $D^{-\frac{1}{2}}$ u prvom koraku diskretizacije.

Takvoj matrici ćemo normirati retke u drugom koraku diskretizacije kako i particijska matrica $X$ ima normirane retke, a njoj se želimo približiti. Razlog je također opravdan ranijim komentarom na teoriju pertutbacije. Neka je $f$ matrična funkcija koja normira retke matrice. Konkretno, za matricu $A$ vrijedi

$$f(A) = diag(Diag(AA^\tau))^{-\frac{1}{2}}A.$$

Funkcija $f$ zadovoljava svojstvo $f(AQ) = f(A)Q$ za ortogonalnu matricu $Q$ kako je $(AQ)(AQ)^\tau = AA^\tau$.

Označimo sa $\tilde{X}$ matricu $D^{-\frac{1}{2}}(v_1, \dots, v_k)$ nakon normiranja redaka. $\tilde{X}Q$ predstavlja skup rješenja neprekidnog problema nakon prethodna dva koraka diskretizacije. $\tilde{X}Q$ nema strukturu koja jedinstveno određuje particiju, ali očekujemo da ima blisku i da ćemo za dobar odabir ortogonalne matrice $Q$ pronaći rješenje koje je pretvorbom u particijsku matricu dovoljno blizu optimalnom.

Pitanje je kako pronaći takav najbolji $Q^*$ i koju particijsku matricu smatrati bliskom transformiranom neprekidnom rješenju $\tilde{X}Q^*$? Zahtijevati ćemo da je za najbližu particijsku matricu $X^*$ Frobeniusova norma $\|X - \tilde{X}Q^*\|_F^2$ minimalna po svim particijskim matricama $X$ za fiksiranu matricu $Q^*$. Dakle, ona matrica koja minimizira problem najmanjih kvadrata u Frobeniusovoj normi.

Problem pronalaska matrica $X^*$ i $Q^*$ rješavamo alternirajućim najmanjim kvadratima. U prvoj iteraciji fiksiramo matricu $Q^*$ i optimiramo po svim particijskim matricama $X$ u svrhu pronalaska one najbolje, tj. $X^*$. U drugom koraku fiksiramo optimalnu particijsku matricu $X^*$ iz prvog koraka i optimiramo po svim $Q$, odnosno rješavamo isti problem najmanjih kvadrata $\|X - \tilde{X}Q^*\|_F^2$, sada minimizirajući po svim ortogonalnim matricama $Q$.

Ukoliko neki kriterij zaustavljanja nije zadovoljen ponavljamo iteracije. Kriterij zaustavljanja može biti broj iteracija nakon kojeg nema značajnog poboljšanja u odnosu na prethodni optimum (moguće vizualno reprezentirati *elbow plotom*), vremensko ograničenje, broj iteracija i slično.

Ostaje riješiti same probleme najmanjih kvadrata u prvoj, odnosno drugoj iteraciji. Koja matrica $X^*$ zadovoljava

$$X^* = \underset{X \in \mathcal{F}_{n,k}}{\arg\min} \|X - \tilde{X}Q^*\|_F^2$$

za neki fiksni $Q^*$?

**Propozicija 5.3.** *Neka je $Y \in \mathbb{R}^{n \times k}$ prizvoljna matrica sa normiranim retcima. Particijska matrica $X^*$ takva da je $X^* = \underset{X \in \mathcal{F}_{n,k}}{\arg\min} \|X - Y\|_F^2$ dana je sa*

$$X^*_{ij} = \begin{cases} 1, \ j = \underset{l=1,\ldots,k}{\arg\max} Y_{il} \\ 0, \ \textit{inače} \end{cases}$$

Dokaz: Neka je $X$ proizvoljna permutacijska matrica. Označimo sa $X_{i\cdot}$, odnosno $Y_{i\cdot}$ $i$-ti redak matrice $X$ odnosno $Y$. Primijetimo da je $Y_{ij} \leq 1$ kako norma svakog retka mora biti 1.

$$\|X - Y\|_F^2 = \sum_{i=1}^{n} \|X_{i\cdot} - Y_{i\cdot}\|_2^2.$$

Kako je $Y_{i\cdot} - X^*_{i\cdot} = (Y_{i1}, \ldots, Y_{ij} - 1, \ldots Y_{ik})^\tau$ za upravo onaj $j$ gdje je $Y_{ij}$ najbliži jedinici, odnosno najveći, jasno da ne možemo dobiti manju Euklidsku normu oduzevši jedinicu na nekom drugom indeksu. Dakle

$$\|X - Y\|_F^2 = \sum_{i=1}^{n} \|X_{i\cdot} - Y_{i\cdot}\|_2^2 \leq \sum_{i=1}^{n} \|X^*_{i\cdot} - Y_{i\cdot}\|_2^2 = \|X^* - Y\|_F^2.$$

□

Drugi korak u alternirajućim najmanjim kvadratima naziva se Prokrustov problem.

**Propozicija 5.4.** ***(Prokrustov problem)*** *Neka su $X$, $Y \in M_{n,k}(\mathbb{R})$ i $U\Sigma V^\tau$ SVD dekompozicija matrice $Y^\tau X$. Tada je*

$$UV^\tau = \arg\min\{\|X - YQ\|_F^2 : Q^\tau Q = QQ^\tau = I_k\}.$$

Dokaz: Frobeniusova norma generirana je sklarnim produktom, odnosno vrijedi

$$\|X - YQ\|_F^2 = \mathrm{tr}((X - YQ)(X - YQ)^\tau) = \mathrm{tr}(XX^\tau) - \mathrm{tr}(X(YQ)^\tau) - \mathrm{tr}(YQX^\tau) + \mathrm{tr}(YQQ^\tau Y^\tau)$$

$$= \mathrm{tr}(XX^\tau) - 2\,\mathrm{tr}((Q^\tau Y^\tau X) + tr(YY^\tau).$$

U posljednjoj jednakosti koristimo komutativnost realnog skalarnog produkta i činjenicu da je $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ za obostrano ulančane matrice. Obzirom da izraz minimiziramo po svim ortogonalnim matricama $Q$, optimizacijski problem ekvivalentan je s

$$\max_Q \operatorname{tr}(Q^\tau Y^\tau X).$$

Neka je $SVD$ dekompozicija matrice $Y^\tau X$ dana sa $U\Sigma V^\tau$. Vrijedi

$$\operatorname{tr}(Q^\tau Y^\tau X) = \operatorname{tr}(Q^\tau U\Sigma V^\tau) = \operatorname{tr}(\Sigma V^\tau Q^\tau U) = \sum_{i=1}^{k} \sigma_i (V^\tau Q^\tau U)_{ii} \leq \sum_{i=1}^{k} \sigma_i = \operatorname{tr}(\Sigma).$$

Nejednakost slijedi iz činjenica da ortogonalna matrica $V^\tau Q^\tau U$ ima sve elemente manje ili jednake 1. Kako matrica $Q^\tau = VU^\tau$ poništava $U$ i $V^\tau$ nejednakost postaje jednakost, odnosno postiže se gornja ograda pa time i maksimum.

Ovime smo dokazali kako je rješenje dano sa $Q = UV^\tau$.

□

Propozicije 5.3 i 5.4 omogućavaju nam sljedeći algoritam.

**Normalizirano spektralno klasteriranje (Yu, Shi)**

**Ulaz:** Težinska matrica $W$, broj klastera $k$

1. Izračunaj matricu $V_k = (v_1, \dots, v_k)$, gdje su $v_1, \dots, v_k$ svojstveni vektori pridruženi najvećim svojstvenim vrijednostima $\lambda_1 \geq \cdots \geq \lambda_k$ matrice $D^{-\frac{1}{2}} W D^{-\frac{1}{2}}$.
2. Izračunaj matricu $\tilde{X}$ normirajući retke matrici $D^{-\frac{1}{2}} V_k$.
3. Inicijaliziraj početni $Q^* = I_k$.
4. Odredi optimalnu particijsku matricu $X^*$ minimizirajući $\|X - \tilde{X} Q^*\|_F^2$ po svim particijskim matricama $X$.
5. Odredi optimalnu ortogonalnu matricu $Q^*$ minimizirajući $\|X^* - \tilde{X} Q\|_F^2$ po svim ortogonalnim matricama $Q$.
6. Ukoliko kriterij zaustavljanja nije ispunjen vrati se na korak 4.

**Izlaz:** Particijska matrica $X^*$

Primijetimo kako je

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

Ovo je ekvivalentno dualnosti normaliziranih rezova i asocijativnosti $K$-particije.

# Literatura


[1] Ulrike von Luxburg (2007) A Tutorial on Spectral Clustering, Statistics and Computing, 17 (4), 2007.

[2] Stella X. Yu, Jianbo Shi (2003) Multiclass Spectral Clustering, Proceedings Ninth IEEE International Conference on Computer Vision

[3] Jianbo Shi and Jitendra Malik (2000) Normalized Cuts and Image Segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence, VOL. 22, NO. 8,

[4] Hongyuan Zha & Xiaofeng He, Chris Ding & Horst Simon, Ming Gu (2001) Spectral Relaxation for $K$-means Clustering, Advances in Neural Information Processing Systems 14 (NIPS 2001)

[5] Andrew Ng, Michael Jordan, Yair Weiss (2001) On Spectral Clustering: Analysis and an algorithm, Advances in Neural Information Processing Systems 14 (NIPS 2001)