

Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
Matematički odsjek

Višeklasno spektralno klasteriranje

Dorotea Rajšel, Iva Sokač

19. veljače 2018.

seminarski rad iz kolegija Uvod u složeno pretraživanje podataka
mentor: prof. dr. sc. Zlatko Drmač

Sažetak

Cilj ovog rada je opisati metodu klasteriranja pomoću spektralne relaksacije. Prvo postavljamo diskretnu formulaciju problema. Zatim rješavamo relaksirani problem u neprekidnoj domeni pomoću dekompozicije singularnih vrijednosti. Potom tražimo diskretno rješenje najbliže kontinuiranom optimumu. Diskretizacija se efikasno računa na iterativni način pomoću SVD-a i ne-maksimalne supresije. Tako dobivena diskretna rješenja su gotovo globalno optimalna. Ovakva metoda je robusna s obzirom na inicijalizacije i konvergira brže nego druge metode klasteriranja. Metodu ćemo testirati na skupovima točaka i slikama.

1 Uvod

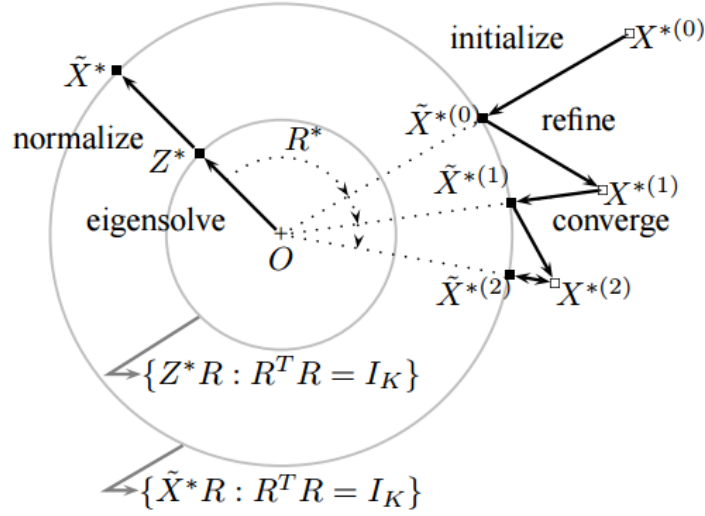
Metoda spektralnog klasteriranja grafa se uspješno primjenjuje na balansiranje opterećenja, nacрте strujnih krugova i segmentaciju slika. Ne pretpostavlja se ništa o globalnoj strukturi podataka, već se prvo prikupljaju lokalni dokazi o tome koliko dva slična podatka pripadaju istom klasteru i onda se donosi globalna odluka o tome kako podijeliti podatke u disjunktne skupove po nekom kriteriju.

Ono što čini spektralno klasteriranje privlačnom metodom je to što je globalni optimum u relaksiranom kontinuiranom području dobiven preko spektralne dekompozicije. No, da bismo dobili nazad diskretno rješenje iz dekompozicije, potrebno je riješiti još jedan problem klasteriranja, samo što je u nižedimenzionalnom prostoru.

Način za otkrivanje diskretnog optimuma se bazira na činjenici da je kontinuirani optimum cijela familija razapeta svojstvenim vektorima kroz ortonormirane transformacije.

U ovoj metodi se nalazi prava ortonormirana transformacija koja nas vodi do diskretnog rješenja.

Algoritam se sastoji od dva temeljna koraka. Prvo se rješava relaksirani optimizacijski problem. Globalni optimum je dan svojstvenim vektorima koje podliježemo ortonormiranim transformacijama. Na iterativni način tražimo diskretno rješenje koje je najbliže kontinuiranom optimumu koristeći alternirajući optimizacijski postupak. Alterniramo sljedeće: kontinuirani optimum najbliži diskretnom rješenju se pronalazi računajući najbolju ortonormiranu transformaciju, a diskretno rješenje najbliže kontinuiranom se nalazi ne-maksimalnom supresijom (ako ima više maksimuma, samo jedan od njih, ali bilo koji, možemo izabrati tako zadovoljava uvjete diskretnog rješenja). Takve iteracije monotonno smanjuju udaljenost između kontinuiranog i diskretnog optimuma. Tako dobivamo gotovo globalno optimalnu particiju.



Slika 1: Shematski dijagram algoritma: Prvo dobijemo svojstvene vektore Z^* . Pri-
kazano kao unutarnji krug, Z^* generira cijelu familiju globalnih optimuma kroz
ortonormiranu transformaciju R . Poslije normiranja duljine, svaki optimum odgovara
rješenju particioniranja na kontinuiranom području (vanjski krug). Onda dobijemo
diskretno rješenje $X^{*(0)}$, nađemo $\tilde{X}^{*(0)}$ računajući R^* koji približava $\tilde{X}^{*(0)}$ najbliže
 $X^{*(0)}$. Preko neprekidnog optimuma $X^{*(0)}$ izračunamo diskretnu verziju i tako da-
lje. Algoritam konvergira prema paru rješenja $(X^{*(2)}, \tilde{X}^{*(2)})$ koji su najbliži jedno
drugome. Optimalnost $\tilde{X}^{*(2)}$ garantira da je $X^{*(2)}$ blizu globalnog optimuma.

2 Višeklasni normalizirani rezovi

Težinski graf je dan s $\mathbb{G} = (\mathbb{V}, \mathbb{E}, W)$, gdje je \mathbb{V} skup svih vrhova grafa, \mathbb{E} skup
bridova, a W matrica susjedstva s težinama koje predstavljaju kolika je vjerojatnost
da dva čvora pripadaju istoj grupi. W je po pretpostavci nenegativna i simetrična
matrica.

Neka $[n]$ označava skup prirodnih brojeva od 1 do n . Neka $\mathbb{V} = [N]$ označava skup
svih elemenata koje trebamo grupirati. Za grupiranje N točaka u K grupa trebamo
naći particiju \mathbb{V} na K disjunktnih skupova. To K -particioniranje označavamo s
 $\Gamma_{\mathbb{V}}^K = \{\mathbb{V}_1, \dots, \mathbb{V}_K\}$.

2.1 Kriterij višeklasnog particioniranja

Neka su $\mathbb{A}, \mathbb{B} \subset \mathbb{V}$. Definiramo $links(\mathbb{A}, \mathbb{B})$ kao zbroj ukupnih težina bridova koji
izlaze iz \mathbb{A} i ulaze u \mathbb{B} :

$$links(\mathbb{A}, \mathbb{B}) = \sum_{i \in \mathbb{A}, j \in \mathbb{B}} W(i, j)$$

Definiramo *stupanj skupa* kao zbroj težina svih bridova s početkom u tom skupu:

$$degree(\mathbb{A}) = links(\mathbb{A}, \mathbb{V})$$

Koristeći stupanj kao pojam normalizacije, definiramo:

$$linkratio(\mathbb{A}, \mathbb{B}) = \frac{links(\mathbb{A}, \mathbb{B})}{degree(\mathbb{A})}$$

Dva omjera su od posebnog značaja, $linkratio(\mathbb{A}, \mathbb{A})$ koji govori koliko bridova ostaje unutar skupa $linkratio(\mathbb{A}, \mathbb{V} \setminus \mathbb{A})$ koji nam govori koliko bridova izlazi iz skupa. Dobro grupiranje zahtijeva dobru povezanost podataka unutar komponenti particije i labavu povezanost podataka iz različitih komponenti. Oba cilja postizemo K-normaliziranim vezanjem i normaliziranim rezovima:

$$knassoc(\Gamma_{\mathbb{V}}^K) = \frac{1}{K} \sum_{t=1}^{\infty} linkratio(\mathbb{V}_t, \mathbb{V}_t)$$

$$kncut(\Gamma_{\mathbb{V}}^K) = \frac{1}{K} \sum_{t=1}^{\infty} linkratio(\mathbb{V}_t, \mathbb{V} \setminus \mathbb{V}_t)$$

Budući da je $knassoc(\Gamma_{\mathbb{V}}^K) + kncut(\Gamma_{\mathbb{V}}^K) = 1$ i te funkcije su nenegativne, maksimiziranje veza između komponenti particije i minimiziranje rezova između komponenti se postiže simultano. Uvažavajući tu činjenicu, bavit ćemo se samo jednim problemom; minimizirat ćemo funkciju reza:

$$\epsilon(\Gamma_{\mathbb{V}}^K) = kncut(\Gamma_{\mathbb{V}}^K)$$

Svaki način raspodjele podataka na K particija ima svoju ϵ vrijednost. Pokazat ćemo da se gornja granica od ϵ monotonno smanjuje kako se K povećava.

2.2 Reprezentacija

Koristimo $N \times K$ matricu particije da bismo prikazali $\Gamma_{\mathbb{V}}^K$. Neka je $X = [X_1, X_2, \dots, X_K]$, gdje je X_l binarni indikator za V_l :

$$X(i, l) = \langle i \in \mathbb{V}_l \rangle, i \in \mathbb{V}, l \in [K]$$

Budući da je svaki čvor pridružen isključivo jednoj particiji, stupci matrice X su međusobno isključivi, tj. $X1_K = 1_K$, gdje je 1_d vektor jedinica dimenzije $d \times 1$. Za simetričnu težinsku matricu W definiramo matricu stupnja:

$$D = \text{Diag}(W1_N)$$

gdje je $\text{Diag}(x)$ dijagonalna matrica generirana svojim argumentom (vektorom x). Sada vezu i stupanj možemo zapisati kao:

$$links(\mathbb{V}_l, \mathbb{V}_l) = X_l^T W X_l$$

$$degree(\mathbb{V}_l) = X_l^T D X_l$$

Kriterij K-particioniranih normaliziranih rezova izražavamo u optimizacijskom problemu s varijablom X , kojeg zovemo $PNCX$:

$$maksimizirati \epsilon(X) = \frac{1}{K} \sum_{t=1}^K \frac{X_t^T W X_t}{X_t^T D X_t}, X \in \{0, 1\}^{N \times K}, X1_K = 1_N$$

Ovaj problem je NP-potpun čak i za $K=2$ i za planarne grafove. Zbog toga ovaj algoritam pronalazi njemu najbliži optimum, tj. gotovo globalno rješenje ovog problema.

3 Rješavanje K-particioniranih normaliziranih rezova

PNCX problem se rješava u 2 koraka:

1. Transformirana formulacija problema se relaksira u problem svojstvenih vrijednosti. Pokazuje se da globalni optimum nije jedinstven te je specijalno rješenje generalizirani svojstveni vektori matrice (W, D) . Transformacijom svojstvenih vektora u prostor matrice particije dobiva se skup neprekidnih globalnih optimuma.
2. Rješavamo diskretan problem u kojem tražimo diskretnu matricu particije najbližu onoj neprekidnoj. Takvo diskretno rješenje je stoga gotovo globalno optimalno.

3.1 Pronalazak optimalnih rješenja relaksiranog problema

Pojednostavimo li jednadžbu $\epsilon(X) = \frac{1}{K} \sum_{t=1}^K \frac{X_t^T W X_t}{X_t^T D X_t}$ dobivamo:

$$\epsilon(X) = \frac{1}{K} \text{tr}(Z^T W Z),$$

gdje je

$$Z = X(X^T D X)^{-1/2}$$

Vrijedi: $Z^T D Z = (X^T D X)^{-1/2} X^T D X (X^T D X)^{-1/2} = I_K$, gdje je I_K $K \times K$ jedinična matrica. Odbacujući uvjete iz *PNCX*, dolazimo do novog problema s varijablom Z te ga zovemo *PNCZ*:

$$\text{maksimizirati } \epsilon(X) = \frac{1}{K} \text{tr}(Z^T W Z)$$

uz zahtjev

$$Z^T D Z = I_K$$

Na ovaj način prevodimo diskretan problem u rješivi neprekidni optimizacijski problem.

Propozicija 3.0.1 (Ortonormirana invarijantnost). *Neka je R $K \times K$ matrica. Ako je Z moguće rješenje za *PCNZ*, tada je to i $ZR : R^T R = I_K$. Nadalje, ova rješenja imaju istu ciljanu vrijednost: $\epsilon(ZR) = \epsilon(Z)$.*

Prema propoziciji, dopustivo rješenje je jednako dobro ako na njega djelujemo proizvoljnom rotacijom i refleksijom. Sljedeća propozicija pokazuje da se među svim optimumima nalaze svojstveni vektori od (W, D) te da su oni ekvivalentni svojstvenim vektorima normalizirane težinske matrice P :

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

Budući da je P stohastička matrica, pripada joj trivijalni vektor 1_N koji ima i najveću svojstvenu vrijednost koja iznosi 1.

Propozicija 3.0.2 (Optimalno svojstveno rješenje). *Neka je (V, S) svojstvena dekompozicija matrice P : $PV = VS$, gdje je $V = [V_1, \dots, V_N]$ i $S = \text{Diag}(s)$ sa svojstvenim vrijednostima poredanim nerastuće: $s_1 \geq \dots \geq s_N$. (V, S) dobijemo iz ortonormalnog svojstvenog rješenja (\bar{V}, S) simetrične matrice $D^{-1/2}WD^{-1/2}$, gdje je*

$$V = D^{-1/2}\bar{V}$$

$$D^{-1/2}WD^{-1/2}\bar{V} = \bar{V}S, \bar{V}^T\bar{V} = I_N$$

V i S su realne i bilo kojih K različitih svojstvenih vektora lokalnog optimuma su kandidati za PNCZ, s

$$\epsilon([V_{\pi_1}, \dots, V_{\pi_K}]) = \frac{1}{K} \sum_{l=1}^K s_{\pi_l}$$

gdje je π vektor indeksa K različitih prirodnih brojeva iz $[N]$. Globalni optimum PNCZ problema se tada postiže za $\pi = [1, \dots, K]$:

$$Z^* = [V_1, \dots, V_K]$$

$$\Lambda^* = \text{Diag}([s_1, \dots, s_K])$$

$$\epsilon(Z^*) = \frac{1}{K} \text{tr}(\Lambda^*) = \max_{Z^T DZ = I_K} \epsilon(Z)$$

Dakle, globalni optimum PNCZ problema nije jedinstven. To je prostor razapet s prvih K najvećih svojstvenih vektora matrice P kroz ortonormirane matrice:

$$Z^*R : R^T R = I_K, PZ^* = Z^*\Lambda^*$$

U slučaju da sve svojstvene vrijednosti nisu jednake, Z^*R više nisu svojstveni vektori od P .

Korolar 3.0.1 (Monotonost gornje međe). *Za bilo koji K vrijedi:*

$$\max \epsilon(\Gamma_K^V) \leq \max_{Z^T DZ = I_K} \epsilon(Z) = \frac{1}{K} \sum_{l=1}^K s_l$$

$$\max_{Z^T DZ = I_{K+1}} \epsilon(Z) \leq \max_{Z^T DZ = I_K} \epsilon(Z)$$

Sada transformiramo Z natrag u prostor matrica particije. Ako je f preslikavanje sa X u Z , tada je f^{-1} normalizacija koja vraća Z natrag u X :

$$Z = f(X) = X(X^T DX)^{-1/2}$$

$$X = f^{-1}(Z) = \text{Diag}(\text{diag}^{-1/2}(ZZ^T))Z$$

gdje diag vraća dijagonalnu matricu argumenata (koji je matrica) kao vektor-stupac. Ako uzmemo retke matrice Z kao koordinate K -dimenzionalnih točaka, tada f^{-1} zapravo normalizira njihove duljine tako da leže na jediničnoj hipersferi centriranoj u ishodištu. Sa f^{-1} transformiramo neprekidni optimum Z^*R iz Z -prostora u X -prostor, budući da je $R^T R = I_K$ vrijedi:

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

Ovo pojednostavljenje omogućava da optimume neprekidnog problema direktno opisujemo s $f^{-1}(Z^*)$ u X -prostoru:

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

Sada je jasno da nam za podjelu skupa na K particija treba točno K svojstvenih vektora (umjesto 2^K). To je iz razloga što indikatori grupa moraju biti ortogonalni, oni ne mogu biti proizvoljni.

3.2 Pronalaženje optimalnih diskretnih rješenja

Optimumi programa $PNCZ$ generalno nisu ostvarivi u programu $PNCX$. Međutim, možemo ih iskoristiti za nalaženje približnog diskretnog rješenja. To rješenje možda ne maksimizira $PNCX$, ali je gotovo globalno-optimalno zbog neprekidnosti funkcije cilja. Stoga, cilj je naći diskretno rješenje koje zadovoljava binarna ograničenja originalnog $PNCX$ programa, a najbliže je neprekidnom optimumu danom jednačbom $\tilde{X}^* = f^{-1}(Z^*)$.

Teorem 3.1 (Optimalna diskretizacija). . *Neka je $\tilde{X}^* = f^{-1}(Z^*)$. Optimalna diskretna particija od X^* je ona koja rješava problem POD , koji je definiran s:*

$$\text{minimizirati} \quad \phi(X, R) = \|X - \tilde{X}^* R\|^2$$

uz

$$X \in \{0, 1\}^{N \times K}, X 1_K = 1_N$$

$$R^T R = I_K$$

gdje s $\|M\|$ označavamo Frobeniusovu normu matrice M : $\|M\| = \sqrt{\text{tr}(MM^T)}$. Lokalni optimum (X^*, R^*) ovog bilinearnog optimizacijskog problema se može riješiti iterativno. S poznatim R^* , POD se reducira na problem $PODX$:

$$\text{minimizirati} \quad \phi(X) = \|X - \tilde{X}^* R\|^2$$

$$\text{uz } X \in \{0, 1\}^{N \times K}, X 1_K = 1_N$$

Neka je $\tilde{X} = \tilde{X}^* R^*$. Optimalno rješenje je dano metodom "stanjivanja rubova" (ako ima više maksimuma, bira se jedan i samo jedan maksimum):

$$X^*(i, l) = \langle l = \text{argmax}_{k \in [K]} \tilde{X}(i, k) \rangle, i \in \mathbb{V}$$

S poznatim X^* , POD reduciramo na $PODR$ u R :

$$\text{minimizirati} \quad \phi(R) = \|X - \tilde{X}^* R\|^2$$

$$R^T R = I_K$$

i rješenje je dano preko singularnih vektora:

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

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

gdje je (U, Ω, \tilde{U}) SVD matrice $(X^*)^T \tilde{X}^*$, uz $U^T U = I_K$, $\tilde{U}^T \tilde{U} = I_K$ i $\omega_1 \geq \dots \geq \omega_K$

Dokaz. Definiramo $\phi(X; R) = \|X\|^2 + \|\tilde{X}^*\|^2 - \text{tr}(X R^T \tilde{X}^{*T} + X^T \tilde{X}^* R) = 2N - 2\text{tr}(X R^T \tilde{X}^{*T})$. Iz toga slijedi da je minimiziranje funkcije $\phi(X, R)$ ekvivalentno maksimiziranju izraza $\text{tr}(X R^T \tilde{X}^{*T})$. Budući da se svaki element od $\text{diag}(X R^T \tilde{X}^{*T})$ može zasebno maksimizirati, rješenje problema $PODX$, uz $R = R^*$, slijedi $X^*(i, l) = \langle l = \text{argmax}_{k \in [K]} \tilde{X}(i, k) \rangle, i \in \mathbb{V}$. Za $PODR$, uz $X = X^*$, koristimo Lagrangeovu funkciju koristeći Λ , matricu simetričnu s obzirom na matrično množenje:

$$L(R, A) = \text{tr}(X^* R^T \tilde{X}^{*T}) - \frac{1}{2} \text{tr}(\Lambda^T (R^T R - I_K))$$

Rješenje (R^*, Λ^*) mora zadovoljavati

$$L_R = \tilde{X}^{*T} X^* - R\Lambda = 0, t_j \Lambda^* = R^{*T} \tilde{X}^{*T} X^*$$

Slijedi $\Lambda^{*T} \Lambda^* = U\Omega^2 U^T$. Budući da je $\Lambda = \Lambda^T$, vrijedi $\Lambda^* = U\Omega U^T$. Nadalje, iz jednačbe $L(R, A) = \text{tr}(X^* R^T \tilde{X}^{*T}) - \frac{1}{2} \text{tr}(\Lambda^T (R^T R - I_K))$ slijedi $R^* = \tilde{U} U^T$ i $\phi(R^*) = 2N - 2\text{tr}(\Omega)$. Što je veći $\text{tr}(\Omega)$, to je X^* bliže $\tilde{X}^* R^*$. \square

Zbog ortonormirane invarijantnosti rješenja neprekidnog problema, naša metoda ne ovisi o inicijalizaciji, bilo X ili R . No, dobra inicijalizacija svejedno može pridonijeti brzini konvergencije. Dobra heuristička metoda je metoda K-sredina s K priližno ortogonalnih točaka kao centrima. S poznatim X^* , mi rješavamo *PODR* da bi našli najbliži neprekidni optimum $\tilde{X}^* R^*$. Za ovaj optimum, rješavamo *PODX* da bi našli najbliže diskretno rješenje. Svakim korakom reduciramo ciljnu ϕ . Možemo garantirati da će tim iteracijama program završiti u lokalnom optimumu, koji može varirati sa inicijalnom procjenom. Međutim, jer su $\tilde{X}^* R^*$ svi globalni optimumi neovisno o R^* , kojem god $\tilde{X}^* R^*$ *POD* konvergira, diskretno rješenje X neće biti daleko od optimalnog rješenja.

3.3 Algoritam

Uz danu matricu težina W i broj klasa K , provodimo sljedeći algoritam:

1. Računamo matricu $D = \text{Diag}(W1_N)$
2. Tražimo optimalne svojstvene vektore Z^* preko:
$$D^{-1/2} W D^{-1/2} \bar{V}_{[K]} = \bar{V}_{[K]} \text{Diag}(s_K), \bar{V}_{[K]}^T \bar{V}_{[K]} = I_K$$

$$Z^* = D^{-1/2} \bar{V}_{[K]}$$
3. Normaliziramo Z^* : $\tilde{X}^* = \text{Diag}(\text{diag}^{-1/2}(Z^* Z^{*T})) Z^*$
4. Inicijaliziramo X^* računajući R^* kao: $R_1^* = [\tilde{X}^*(I, 1), \dots, \tilde{X}^*(i, K)]^T$, za random $i \in [N]$
 $c = 0_{Nx1}$
For $k = 2, \dots, K$, do:
 $\dots c = c + \text{abs}(\tilde{X}^* R_{k-1}^*)$
 $\dots R_k^* = [\tilde{X}^*(i, 1), \dots, \tilde{X}^*(i, K)]^T, i = \text{argmin} c$
5. Inicijaliziramo parametar $\bar{\phi}^* = 0$ za praćenje konvergencije
6. Tražimo optimalno diskretno rješenje X^* :

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

$$X^*(i, l) = \langle l = \text{argmax}_{k \in [K]} \tilde{X}(i, k) \rangle, i \in \mathbb{V}, l \in [K]$$

7. Tražimo optimalnu ortonormiranu matricu R^* :
 $X^{*T} \tilde{X}^* = U\Omega \tilde{U}^T, \Omega = \text{Diag}(\omega)$
 $\bar{\phi} = \text{tr}(\Omega)$
if $|\bar{\phi} - \bar{\phi}^*| < \text{machine precision}$, then stop and output X^*
 $\bar{\phi}^* = \bar{\phi}$
 $R^* = \tilde{U} U^T$

8. Vratimo se na 6. korak

U 2. koraku $\bar{V}_{[K]}$ označava $[\bar{V}_1, \dots, \bar{V}_K]$, slično za $s_{[K]}$. Može se pokazati da je složenost algoritma $O(N^{\frac{3}{2}}K + NK^2)$

4 Eksperimenti

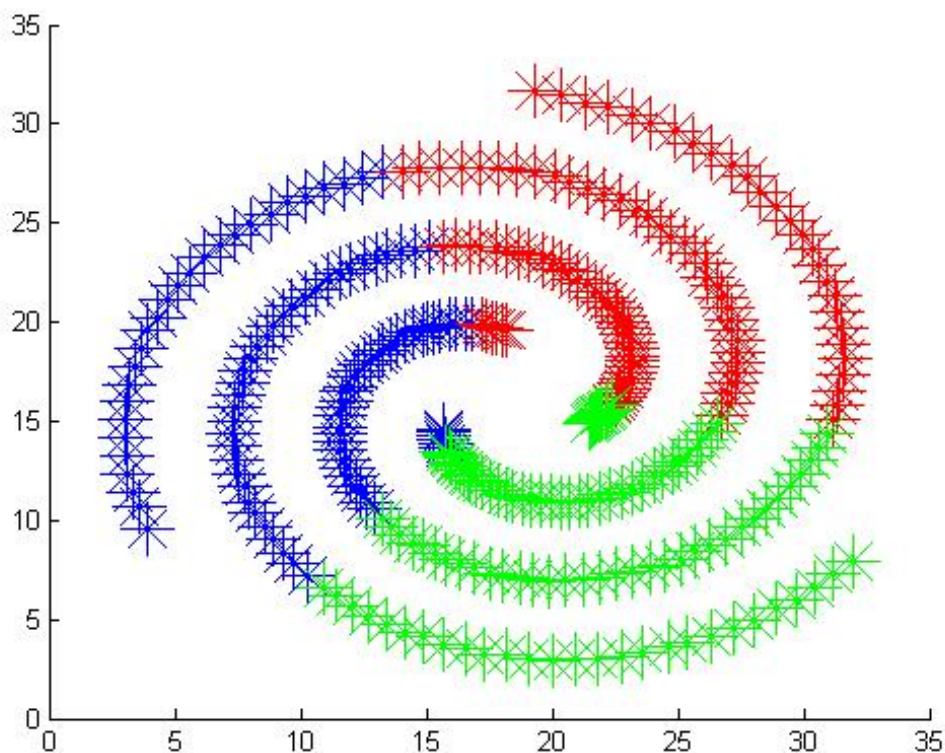
4.1 Model

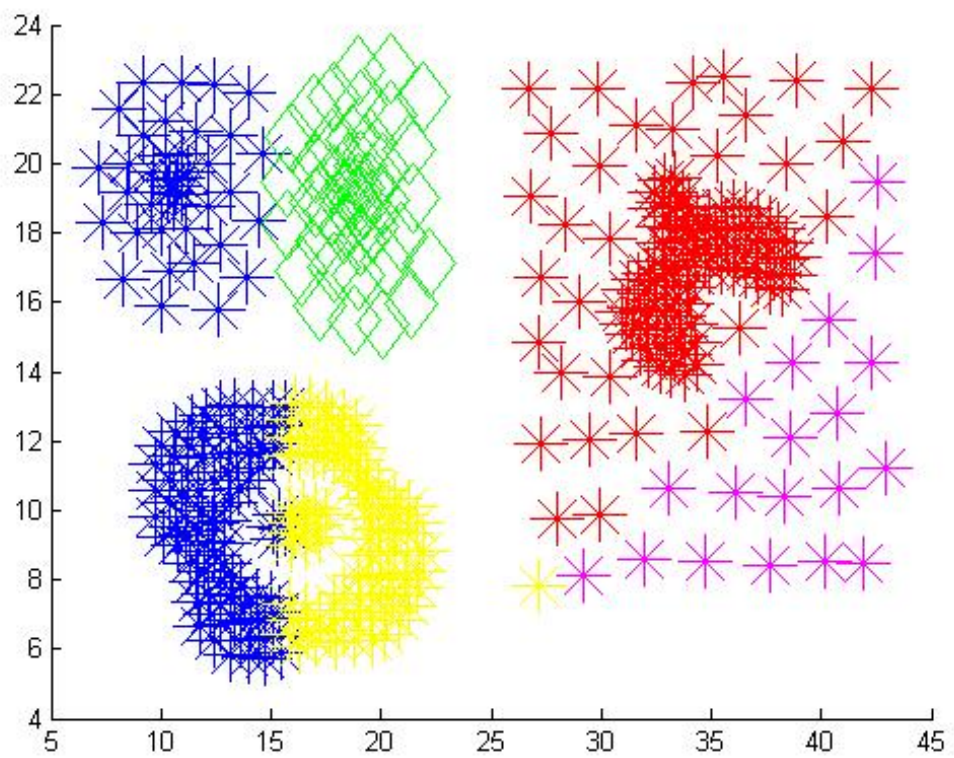
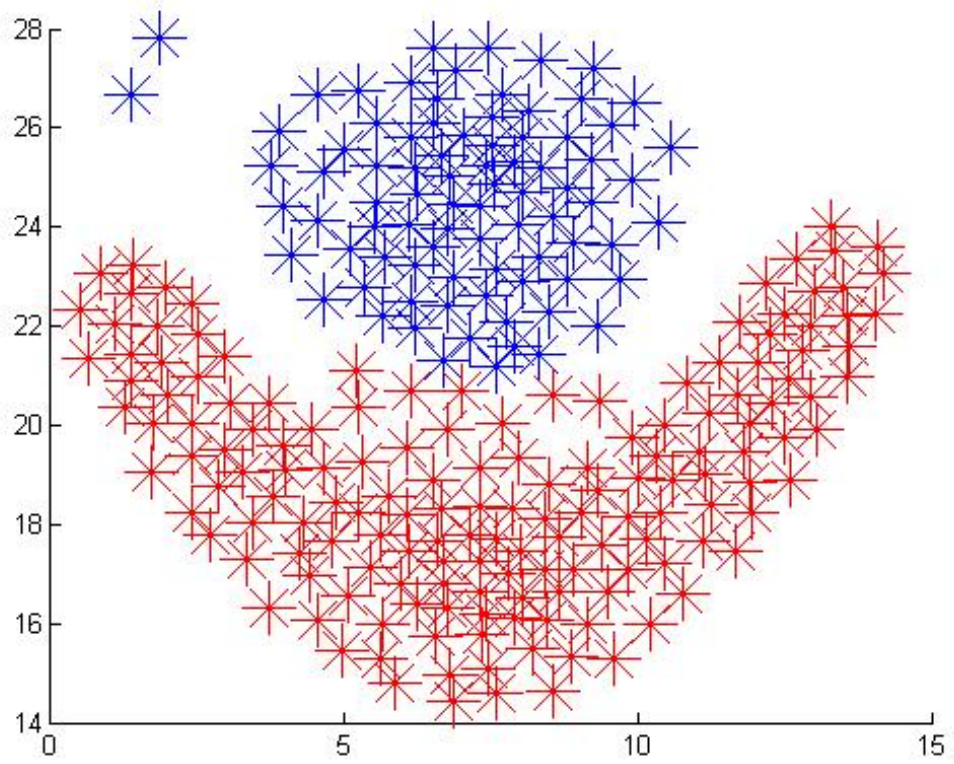
Algoritam smo primijenili na dva problema: klasteriranje točaka po udaljenosti i segmentaciju slika. Prvi korak potreban za primjenu algoritma je modeliranje problema pomoću grafa.

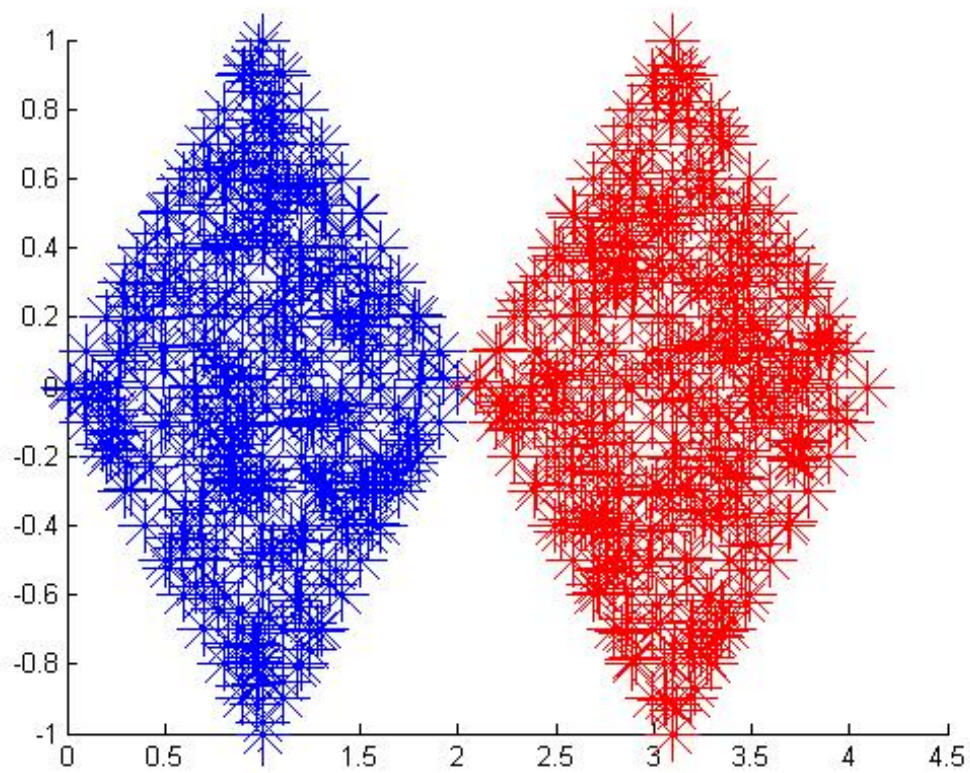
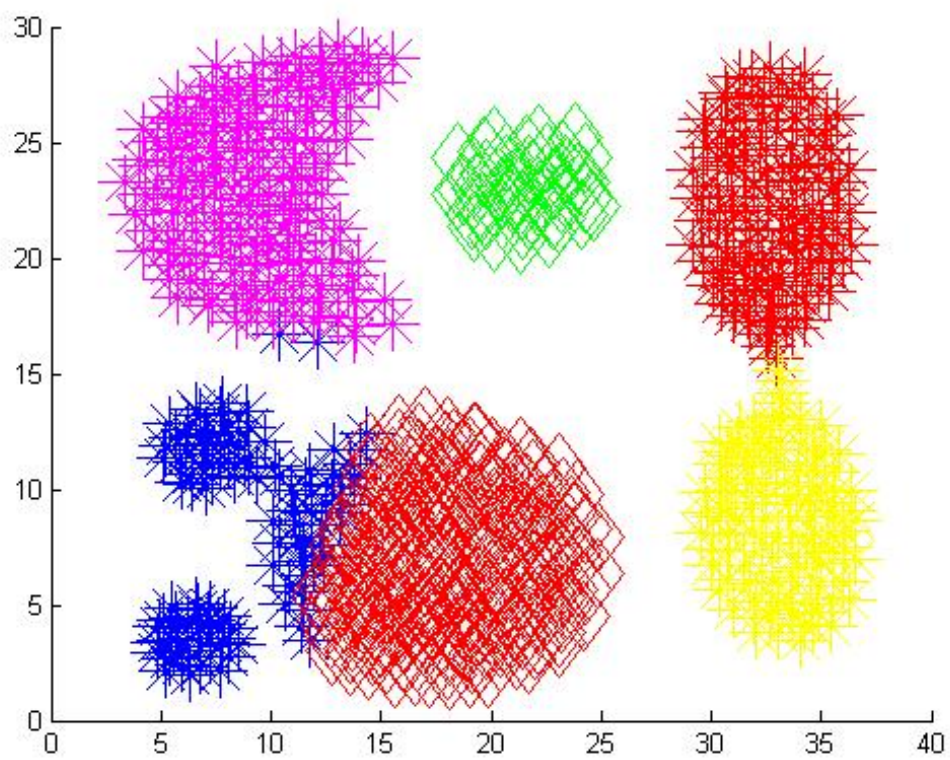
U slučaju klasteriranja točaka to je vrlo jednostavno. Koristimo potpuni graf gdje svaki čvor predstavlja jednu točku, a težina brida između dvije točke je Gaussova funkcija njihove udaljenosti.

Kod segmentacije slike je situacija nešto kompleksnija. Svaki čvor u grafu će predstavljati jedan piksel (uočimo da je već za male slike dimenzija problema vrlo velika). Prvo se metodama matematike i računanja pronađu "jakosti" rubova (engl. magnitude od edge response) na slici. Zatim težinu brida između neka dva piksela postavljamo da bude obrnuto proporcionalna najvećoj jakosti ruba koji siječe dužina koja spaja ta dva piksela. Ova mjera udaljenosti nema smisla za piksele koji nisu blizu pa postavljamo težinu na nula za piksele udaljene više od neke fiksne udaljenosti.

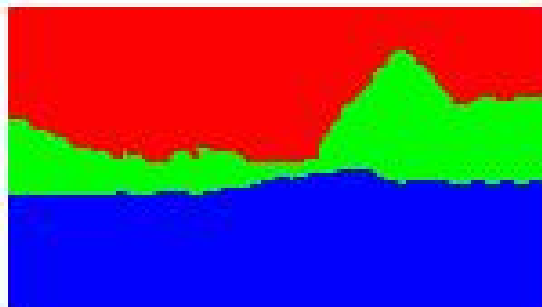
4.2 Rezultati klasterizacije točaka

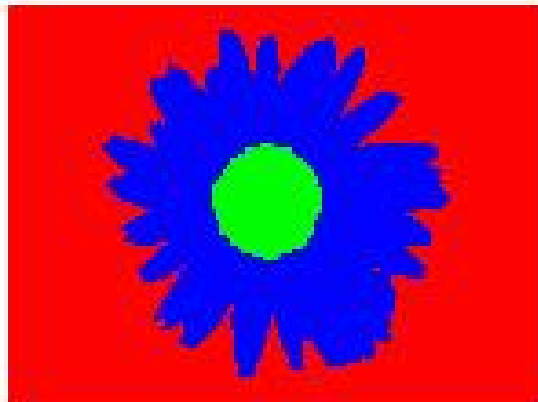


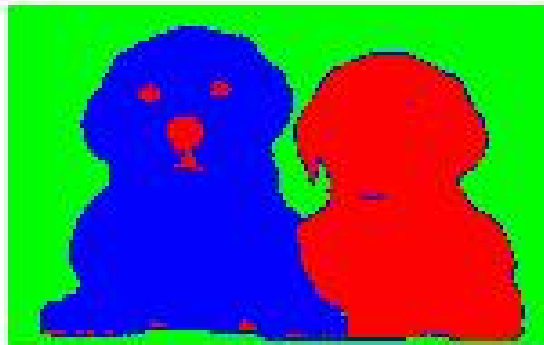




4.3 Rezultati segmentacije slika









Literatura

[SY] Jianbo Shi and Stella X. Yu. Multiclass spectral clustering.