Geometrijske tehnike za traženje pseudo-klika u grafu

Carević, Ana

SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO–MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Ana Carević

GEOMETRIJSKE TEHNIKE ZA TRAŽENJE PSEUDO-KLIKA U GRAFU

Diplomski rad

Voditelj rada:
doc. dr. sc. Pavle Goldstein

Zagreb, veljača 2024.

Ovaj diplomski rad obranjen je dana ____________________ pred ispitnim povjerenstvom u sastavu:

1. ____________________________, predsjednik
2. ____________________________, član
3. ____________________________, član

Povjerenstvo je rad ocijenilo ocjenom _______________.

Potpisi članova povjerenstva:

1. ____________________________
2. ____________________________
3. ____________________________

Zahvaljujem mentoru doc. dr. sc. Pavlu Goldsteinu na zanimljivoj temi, uloženom trudu i pruženoj pomoći tijekom pisanja ovog rada.
Hvala mojoj obitelji i svim mojim prijateljima koji su uvijek bili uz mene.
Najveće hvala mojoj mami Sandri, baki Anici i sestri Luci na bezuvjetnoj podršci u najtežim trenucima.

Sadržaj

Uvod

U ovom radu bavit ćemo se problemom pronalaženja klike. Pristup rješavanju ovakvog problema najčešće je kombinatorni, koji zbog dugog vremena izvršavanja često nije optimalan. Stoga ovdje razvijamo geometrijski pristup, gdje će nam cilj biti u što kraćem vremenu pronaći klike čak i za velike dimenzije grafova. Prvo ćemo konstruirati regularno ulaganje grafa u euklidski prostor, što će nam omogućiti pronalazak povoljnog uređaja na skupu vrhova grafa. Takav uređaj omogućit će nam pronalaženje klika, a maksimizacijom najveće dobit ćemo traženu maksimalnu kliku.

Rad se sastoji od tri poglavlja. Prvo poglavlje sadrži pojmove koje ćemo koristiti, a to su definicije, teoremi i propozicije iz linearne algebre i teorije grafova, te definicija regularnog n-simpleksa. U drugom poglavlju je opisan problem, zatim ulaganje konačnog metričkog prostora u euklidski prostor, te ideja i kod našeg algoritma. Konačno, u trećem poglavlju ćemo opisati postupak generiranja primjera, te ih navesti i na kraju komentirati dobivene rezultate i uspješnost samog algoritma.

Poglavlje 1

Teorijska pozadina

Pojmovi iz ovog poglavlja su preuzeti iz izvora [1], [2], [3], [4], [5], [6], [7], [8], [9] i [10].

1.1 Linearna algebra

Definicija 1.1.1. *Neka je V neprazan skup na kojem su zadane binarne operacije zbrajanja $+ : V \times V \to V$ i operacija množenja skalarima iz polja $\mathbb{F}$, $\cdot : \mathbb{F} \times V \to V$. Kažemo da je uređena trojka $(V, +, \cdot)$* ***vektorski prostor nad poljem*** *$\mathbb{F}$ ako vrijedi:*

1) $a + (b + c) = (a + b) + c, \ \forall a, b, c \in V;$

2) $postoji\ 0 \in V\ sa\ svojstvom\ a + 0 = 0 + a = a, \ \forall a \in V;$

3) $za\ svaki\ a \in V, postoji\ -a \in V\ tako\ da\ je\ a + (-a) = (-a) + a = 0;$

4) $a + b = b + a, \ \forall a, b \in V;$

5) $\alpha(\beta a) = (\alpha\beta)a, \ \forall \alpha, \beta \in \mathbb{F}, \forall a \in V;$

6) $(\alpha + \beta)a = \alpha a + \beta a, \ \forall \alpha, \beta \in \mathbb{F}, \forall a \in V;$

7) $\alpha(a + b) = \alpha a + \alpha b, \ \forall \alpha \in \mathbb{F}, \forall a, b \in V;$

8) $1 \cdot a = a \cdot 1, \ \forall a \in V.$

Definicija 1.1.2. *Neka je V vektorski prostor nad poljem $\mathbb{F}$.* ***Skalarni produkt*** *na V je preslikavanje $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$ koje ima sljedeća svojstva:*

1) $\langle x, x \rangle \geq 0, \ \forall x \in V;$

2) $\langle x, x \rangle = 0 \Leftrightarrow x = 0;$

3) $\langle x_1 + x_2, y\rangle = \langle x_1, y\rangle + \langle x_2, y\rangle, \ \forall x_1, x_2, y \in V$*;*

4) $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle, \ \forall \alpha \in \mathbb{F}, \forall x, y \in V$*;*

5) $\langle x, y\rangle = \overline{\langle y, x\rangle}, \ \forall x, y \in V.$

Napomena 1.1.3. *U* $\mathbb{R}^n$ *kanonski skalarni produkt definiran je s*

$$\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n)\rangle = \sum_{i=1}^{n} x_i y_i.$$

Definicija 1.1.4. *Vektorski prostor na kojem je definiran skalarni produkt zove se* ***unitarni prostor****.*

Definicija 1.1.5. *Neka je* V *unitaran prostor.* ***Norma*** *na* V *je funkcija* $\|\cdot\| : V \to \mathbb{R}$ *definirana s*

$$\|x\| = \sqrt{\langle x, x\rangle}.$$

Propozicija 1.1.6. *Norma na unitarnom prostoru* V *ima sljedeća svojstva:*

1) $\|x\| \geq 0, \ \forall x \in V$*;*

2) $\|x\| = 0 \Leftrightarrow x = 0$*;*

3) $\|\alpha x\| = |\alpha| \, \|x\|, \ \forall \alpha \in \mathbb{F}, \forall x \in V$*;*

4) $\|x + y\| \leq \|x\| + \|y\|, \ \forall x, y \in V.$

Definicija 1.1.7. *Svaka funkcija* $\|\cdot\| : V \to \mathbb{R}$ *na vektorskom prostoru* V *sa svojstvima iz propozicije 1.1.6 naziva se norma. Tada* $(V, \|\cdot\|)$ *zovemo* ***normirani prostor****.*

Definicija 1.1.8. *Norma koja potječe od kanonskog skalarnog produkta na* $\mathbb{R}^n$*, definiranog u napomeni 1.1.3, dana je formulom*

$$\|(x_1, \ldots, x_n)\| = \sqrt{\sum_{i=1}^{n} |x_i|^2}.$$

Ova norma zove se ***euklidska norma****.*

Definicija 1.1.9. *Neka je* V *normiran prostor.* ***Metrika*** *ili* ***udaljenost*** *vektora* x *i* y *je funkcija* $d : V \times V \to \mathbb{R}$ *definirana s*

$$d(x, y) = \|x - y\|.$$

Propozicija 1.1.10. *Metrika na normiranom prostoru ima sljedeća svojstva:*

1) $d(x,y) \geq 0, \ \forall x,y \in V$;

2) $d(x,y) = 0 \Leftrightarrow x = y, \ \forall x,y \in V$;

3) $d(x,y) = d(y,x), \ \forall x,y \in V$;

4) $d(x,y) \leq d(x,z) + d(z,y), \ \forall x,y,z \in V$.

Definicija 1.1.11. *Neka je $X \neq \emptyset$. Svaka funkcija $d : X \times X \to \mathbb{R}$ sa svojstvima iz propozicije 1.1.10 naziva se metrika ili udaljenost. Tada (X,d) zovemo **metrički prostor**.*

Definicija 1.1.12. *Neka su $x = (x_1, \ldots, x_n)$ i $y = (y_1, \ldots, y_n)$ proizvoljni vektori u $\mathbb{R}^n$. Metrika na $\mathbb{R}^n$, inducirana euklidskom normom iz definicije 1.1.8, dana je s*

$$d((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}.$$

*Ova metrika naziva se **euklidska metrika**, a prostor $\mathbb{R}^n$ zajedno s tom metrikom nazivamo **euklidski prostor**.*

Definicija 1.1.13. *Za prirodne brojeve m i n, preslikavanje*

$$A : \{1, 2, \ldots, m\} \times \{1, 2, \ldots, n\} \to \mathbb{F}$$

*naziva se **matrica** tipa (m,n) s koeficijentima iz polja $\mathbb{F}$.*

Definicija 1.1.14. *Neka je $X \subseteq \mathbb{R}^n$ konačan, neprazan skup, tj. neka je: $X = \{x^1, x^2, ..., x^k\}$, za $k \in N$, pri čemu je $x^1 = (x_1^1, x_2^1, ..., x_n^1), ..., x^k = (x_1^k, x_2^k, ..., x_n^k) \in \mathbb{R}^n$ i neka je d euklidska metrika na $\mathbb{R}^n$. **Težište** skupa X je točka $t_x = (t_1, t_2, ..., t_n) \in \mathbb{R}^n$ u kojoj funkcija: $f(t) := \sum\limits_{i=1}^{k} d(t, x^i)^2$ postiže minimum, pri čemu je $t \in \mathbb{R}^n, x^1, ..., x^k \in X$ i d euklidska metrika.*

Napomena 1.1.15. *Uočimo da je težište skupa u euklidskom prostoru aritmetička sredina njegovih elemenata.*

Definicija 1.1.16. *Neka je dana matrica $A \in \mathbb{C}^{mxn}$. **Adjungirana matrica** matrice A je matrica $A^* \in \mathbb{C}^{nxm}$ za koju vrijedi $A^* = (\overline{a_{ji}})_{ij}$. Ako je $A \in \mathbb{R}^{mxn}$ tada $A^* = A^{\mathrm{T}} \in \mathbb{R}^{nxm}$ nazivamo transponirana matrica matrice A za koju vrijedi $A^{\mathrm{T}} = (a_{ji})_{ij}$.*

Definicija 1.1.17. *Matrica $A \in \mathbb{C}^{nxn}$ je **hermitska** ako vrijedi $A^* = A$. Ako je $A \in \mathbb{R}^{nxn}$ koristimo izraz simetrična matrica te je $A^* = A^{\mathrm{T}} = A$.*

Definicija 1.1.18. *Neka je A kvadratna matrica reda n. Za broj $\lambda \in \mathbb{C}$ kažemo da je* ***svojstvena vrijednost*** *matrice A, a za vektor $x \in \mathbb{C}^n \setminus \{0\}$ kažemo da je svojstveni vektor pridružen svojstvenoj vrijednosti λ matrice A, ako vrijedi $Ax = \lambda x$. Skup svih svojstvenih vrijednosti matrice A zove se spektar matrice A.*

Definicija 1.1.19. *Kažemo da je hermitska matrica $A \in \mathbb{C}^{nxn}$* ***pozitivno semidefinitna*** *ako je $x^* Ax \geq 0$ za svaki $x \in \mathbb{C}^n$.*

Teorem 1.1.20. *Matrica A je pozitivno definitna ako i samo ako su joj sve svojstvene vrijednosti pozitivne.*

Napomena 1.1.21. *Svaka pozitivno definitna matrica je ujedno i pozitivno semidefinitna.*

1.2 Teorija grafova

Definicija 1.2.1. ***Graf*** *G je uređeni par (V, E), gdje je V skup* ***vrhova****, a E skup 2-podskupova od V, koje zovemo* ***bridovi****.*

Napomena 1.2.2. *Katkada gornju definiciju proširujemo tako da dopustimo* ***petlje*** *(bridove koji spajaju vrh sa samim sobom),* ***višestruke bridove*** *(više bridova između para vrhova) i* ***usmjerene bridove*** *(bridovi koji imaju orijentaciju tako da idu od jednog vrha prema drugome). Usmjereni bridovi se reprezentiraju uređenim parovima, a ne 2-podskupovima, dok kod višestrukih bridova E postaje multiskup.*

Napomena 1.2.3. *U ovom radu od interesa će nam biti grafovi koji nemaju usmjerenih bridova, petlji niti višestrukih bridova, te ćemo u skladu s time tretirati gornju definiciju.*

Definicija 1.2.4. *Kažemo da su vrhovi $u, v \in V$, u grafu $G = (V, E)$,* ***susjedni*** *ako postoji brid $e = \{u, v\} \in E$.*

Definicija 1.2.5. *Kažemo da je graf $G = (V, E)$* ***potpun*** *ukoliko za svaki par vrhova u grafu vrijedi da su susjedni.*

Definicija 1.2.6. ***Podgraf*** *grafa $G = (V, E)$ je graf kojemu su skup vrhova i skup bridova podskupovi od V i E, respektivno.*

Definicija 1.2.7. *Neka je $G = (V, E)$ graf te neka je $|V| = n$. Definiramo* ***matricu susjedstva*** *$A = [a_{i,j}] \in M_{n,n}(\mathbb{R})$, $i, j \in \{1, 2, \ldots, n\}$ sa*

$$a_{i,j} = \begin{cases} 1 & \{i, j\} \in E \\ 0 & \textit{inače} \end{cases}.$$

Definicija 1.2.8. ***Klika u grafu*** *$G = (V, E)$ je njegov podgraf s bar dva vrha, koji je potpun (tj. postoji brid između svaka dva vrha podgrafa).*

Definicija 1.2.9. ***Maksimalna klika*** *je klika koja nije sadržana niti u jednoj većoj kliki, tj. dodavanjem nekog vrha, ona prestaje biti klika.*

Definicija 1.2.10. ***Najveća maksimalna klika*** *je maksimalna klika koja ima najveći broj vrhova.*

Definicija 1.2.11. ***Slučajan graf*** *je graf $G = (n, p)$ gdje je n broj vrhova, a svaki brid postoji s vjerojatnošću $p \in \langle 0, 1\rangle$ nezavisno od drugih bridova.*

Teorem 1.2.12. *Neka je $G = (n, p)$ slučajan graf. Tada je* ***očekivana veličina najveće maksimalne klike*** *grafa G jednaka $2\log_{1/p} n$.*

Definicija 1.2.13. *Neka je $G = (V, E)$ graf i neka su τ i γ t.d. $0 \leq \tau \leq \gamma \leq 1$. Podgraf induciran podskupom vrhova $V' \subseteq V$ je (τ, γ)* ***pseudo-klika*** *ako je*

1) $\forall v \in V'$, $deg_{V'}(v) \geq \tau \cdot (|V'| - 1)$;

2) $|E'| \geq \gamma \cdot \binom{|V'|}{2}$,

gdje je $E' = E \cap (V' \times V')$, a $deg_{V'}(v)$ je broj elemenata od V' povezanih s vrhom v.

1.3 Regularni n-simpleks

Definicija 1.3.1. *Za skup točaka S u n-dimenzionalnom prostoru kažemo da je* ***konveksan*** *ako za svake dvije točke $x, y \in S$ vrijedi $[x, y] \subseteq S$, pri čemu je $[x, y] = \{\lambda x + (1 - \lambda)y | \lambda \in [0, 1]\}$.*

Definicija 1.3.2. *Konveksna kombinacija točaka $x_1, x_2, ..., x_m \in \mathbb{R}^n$ je svaka točka x oblika*

$$x = \sum_{i=1}^{m} \lambda_i x_i,$$

gdje su $\lambda_i \geq 0, i = 1, ..., m$, realni brojevi takvi da je $\sum_{i=1}^{m} \lambda_i = 1$. Skup svih konveksnih kombinacija točaka iz skupa $S \subseteq \mathbb{R}^n$ naziva se ***konveksna ljuska*** *skupa S i označava sa $convS$.*

Definicija 1.3.3. *Za konveksnu ljusku $n+1$ točaka u generalnom položaju u n-dimenzionalnom prostoru kažemo da je* ***n-dimenzionalni simpleks****.*

Više o definiciji n-dimenzionalnog simpleksa može se pronaći u [9].

Definicija 1.3.4. *Simpleks dimenzije n kojemu su svi bridovi jednake duljine nazivamo* ***regularni n-simpleks****.*

Napomena 1.3.5. *Simpleks dimenzije 0 je točka, a 1-simpleks je dužina. Iz gornje definicije možemo vidjeti da je regularni 2-simpleks jednakostranični trokut, koji se sastoji od 3 jednake dužine, dok je regularni 3-simpleks pravilni tetraedar, koji se sastoji od 4 jednakostranična trokuta. Regularni n-simpleks za n>*3 *možemo definirati kao generalizaciju tetraedra u većim dimenzijama.*

Propozicija 1.3.6. *Radijus R opisane sfere regularnog n-simpleksa s duljinom stranice a jednak je* $R = a\sqrt{\frac{n}{2n+2}}$.

Dokaz. Simpleks dimenzije n možemo uložiti u $\mathbb{R}^n$ tako da za njegove vrhove uzmemo $n+1$ točaka s koordinatama $(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)$. Uzimanjem ovakvih točaka dobijemo da je težište simpleksa dano s $(\frac{1}{n+1}, \frac{1}{n+1}, \ldots, \frac{1}{n+1})$. Kako je težište simpleksa jednako udaljeno od svih svojih vrhova, a radijus opisane sfere prolazi kroz svaki vrh, slijedi da je duljina radijusa jednaka udaljenosti od bilo kojeg vrha simpleksa do njegovog težišta. Označimo s v_1 vrh simpleksa s koordinatama $(1, 0, \ldots, 0)$. Udaljenost vrha v_1 od težišta simpleksa t iznosi:

$$d(v_1, t) = \|v_1 - t\| = \sqrt{(1 - \frac{1}{n+1})^2 + n(\frac{-1}{n+1})^2)} = \sqrt{(\frac{n}{n+1})^2 + \frac{n}{(n+1)^2}} = \sqrt{\frac{n}{n+1}}.$$

Kako je svaki brid duljine $\sqrt{2}$, moramo još skalirati dobiveni izraz sa $\sqrt{2}$. Dakle, formula za radijus opisane sfere regularnog jediničnog n-simpleksa iznosi $\sqrt{\frac{n}{2n+2}}$, pa je formula za radijus opisane sfere regularnog n-simpleksa s duljinom stranice a jednaka $a\sqrt{\frac{n}{2n+2}}$.

□

Poglavlje 2

Opis problema i algoritam

U ovom poglavlju ćemo opisati problem, ulaganje konačnog metričkog prostora u euklidski prostor, a na kraju ćemo dati ideju i detaljno opisati naš algoritam. Koristit ćemo izvore [11], [12], [13], [14] i [15].

2.1 Motivacija

Problem pronalaženja klike je poznat u teoriji grafova, a zbog svoje kompleksnosti se često svodi na problem pronalaženja pseudo-klike, gdje umjesto potpunog podgrafa tražimo gotovo potpuni podgraf. Ovaj problem postaje sve prisutniji u današnjem svijetu, uzimajući u obzir veliku zastupljenost raznih društvenih mreža. Ukoliko promatrane ljude reprezentiramo vrhovima grafa, klikom ćemo smatrati podskup ljudi gdje se svi međusobno poznaju. Standardni pristup rješavanju ovakvog problema je kombinatorni, a jedan primjer takvog algoritma je Bron-Kerbosch algoritam koji pronalazi sve maksimalne klike zadanog grafa. Više o Bron-Kerbosch algoritmu može se pronaći u [12]. Međutim, problem kod kombinatornog pristupa nastaje prelaskom u veće dimenzije, što je vrlo čest slučaj. Predugo vrijeme izvršavanja ovakvog algoritma zahtijeva korištenje novih pristupa koji će nam u što kraćem vremenu pronaći klike čak i kod većih dimenzija grafova, stoga u ovom radu razvijamo geometrijski pristup. Prije njegovog detaljnog opisa, prvo ćemo na skupu vrhova grafa konstruirati metriku, a u sljedećem dijelu opisujemo ulaganje konačnog metričkog prostora u euklidski prostor.

2.2 Ulaganje konačnog metričkog prostora u euklidski prostor

Ovdje ćemo navesti teorem o izometričkom ulaganju u euklidski prostor i definicije i teoreme koji će nam za to biti potrebni.

Definicija 2.2.1. *Neka je* (X, d) *konačni metrički prostor gdje je* $X = \{p_0, p_1, \dots, p_n\}$. *Označimo* $D_{i,j} = d(p_i, p_j)^2$ *te* $g_{i,j} = \frac{1}{2}(D_{0,i} + D_{0,j} - D_{i,j})$. ***Grammova matrica metričkog prostora*** (X, d) *je matrica* $G = [g_{i,j}] \in M_{n,n}(\mathbb{R})$.

Definicija 2.2.2. *Neka je* V *unitaran prostor i* $\{x_1, \dots, x_k\} \subset V$. *Matrica reda* k *dana s*

$$G(x_1, \dots, x_k) = \begin{pmatrix} \langle x_1, x_1 \rangle & \langle x_1, x_2 \rangle & \cdots & \langle x_1, x_k \rangle \\ \langle x_2, x_1 \rangle & \langle x_2, x_2 \rangle & \cdots & \langle x_2, x_k \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle x_k, x_1 \rangle & \langle x_k, x_2 \rangle & \cdots & \langle x_k, x_k \rangle \end{pmatrix}$$

zove se ***Gramova matrica***.

Pogledajmo sada definiciju Grammove matrice metričkog prostora. Neka su dane točke $p_0, p_1, \dots, p_n$ i pretpostavimo da postoji izometričko ulaganje tih točaka dano s $x_0, x_1, \dots, x_n$ takvo da $x_0 \mapsto 0$. Iz definicije $D_{i,j}$ vidimo da je $D_{0,i} = d(0, x_i)^2 = \|x_i\|^2 = \langle x_i, x_i \rangle$, $D_{0,j} = d(0, x_j)^2 = \|x_j\|^2 = \langle x_j, x_j \rangle$, te $D_{i,j} = d(x_i, x_j)^2 = \|x_i - x_j\|^2 = \langle x_i - x_j, x_i - x_j \rangle$. S obzirom da su x_i i x_j realni vektori, onda vrijedi jednakost $\|x_i - x_j\|^2 = \|x_i\|^2 - 2\langle x_i, x_j \rangle + \|x_j\|^2$. Slijedi da je $g_{i,j} = \frac{1}{2}(D_{0,i} + D_{0,j} - D_{i,j}) = \frac{1}{2}(\|x_i\|^2 + \|x_j\|^2 - \|x_i - x_j\|^2) = \frac{1}{2}(2\langle x_i, x_j \rangle) = \langle x_i, x_j \rangle$. Dakle, na mjestu (i, j) Grammove matrice G se nalazi vrijednost skalarnog produkta vektora x_i i x_j. Slijedi da je Grammova matrica metričkog prostora jednaka Gramovoj matrici definiranoj u 2.2.2. Dodatno, ako Gramovu matricu G zapišemo u obliku $G = X^T X$, pri čemu je $X = \{x_1, \dots, x_k\}$, lako se vidi da je G pozitivno semidefinitna matrica. Naime, ako uzmemo proizvoljan $y \in \mathbb{R}^n$, imamo $y^T G y = y^T X^T X y = (Xy)^T (Xy) = \|Xy\|^2 \geq 0$.

Definicija 2.2.3. *Ulaganje* $f : X \to \mathbb{R}^n$ *metričkog prostora* (X, d') *u euklidski prostor* $\mathbb{R}^n$ *nazivamo* ***izometričkim*** *ukoliko ono čuva udaljenost tj.*

$$\forall a, b \in X \text{ vrijedi } d'(a, b) = d(f(a), f(b)),$$

gdje je d *euklidska metrika.*

Teorem 2.2.4. *Postoji izometričko ulaganje prostora* (X, d) *u euklidski prostor* $\mathbb{R}^n$ *ako i samo ako je Grammova matrica* G *pozitivno semidefinitna. Nadalje, ako je* G *pozitivno semidefinitna matrica, onda se* G *može napisati kao* $G = CDC^T$, *pri čemu je* C *ortogonalna matrica, a* D *dijagonalna. Tada je ulaganje dano sa:* $x_1 \mapsto 0$, *a slika vektora* $x_2, x_3, ..., x_n$ *je dana stupcima matrice* $\sqrt{D}C$.

2.3 Ulaganje grafa kao gotovo regularnog n-simpleksa

Sada ćemo objasniti ulaganje našeg grafa kao gotovo regularnog n-simpleksa, gdje ćemo koristiti prethodno spomenuti teorem 2.2.4. Prije implementacije algoritma, na skupu vrhova grafa ćemo definirati metriku, a zatim provodimo ulaganje grafa u euklidski prostor. Graf dimenzije n ćemo reprezentirati matricom susjedstva 1.2.7, koja se sastoji od nula i jedinica. Nula na mjestu (i, j) označava da između vrha i i vrha j ne postoji brid, dok jedinica označava da postoji. Da bismo izračunali Grammovu matricu, potrebno je prvo definirati matricu udaljenosti D iz definicije 2.2.1. Ideja je definirati matricu udaljenosti tako da su bliži oni vrhovi između kojih postoji brid, pa će naša matrica D izgledati ovako:

$$D_{i,j} = 1 - (t_i + t_j)10^{-8}, (i, j) \in E$$

$$D_{i,j} = 1 + (t_i + t_j)10^{-8}, (i, j) \notin E,$$

gdje je t_i = broj vrhova sa kojima je vrh i povezan. Sada koristeći teorem 2.2.4 provodimo ulaganje grafa u euklidski prostor, koje će zbog pozitivnosti svojstvenih vrijednosti Grammove matrice, biti izometričko 1.1.20 1.1.21.

Napomena 2.3.1. *Napomenimo da gornjom definicijom matrice D osiguravamo, za dovoljno mali n, pozitivnu definitnost Grammove matrice, iz koje nam po teoremu slijedi da se radi o izometričkom ulaganju. Detaljna analiza ovakvih matrica udaljenosti D može se pronaći u [16].*

Napomena 2.3.2. *Uočimo da će slika našeg ulaganja, iz definicije matrice udaljenosti D i činjenice da smo napravili izometričko ulaganje, biti gotovo regularni n-simpleks. Dakle, uzeli smo konačan metrički prostor, gdje su sve udaljenosti gotovo iste i zatim smo ga izometrički uložili, te smo time dobili n-dimenzionalni simpleks iz definicije 1.3.3, kojemu su sve stranice gotovo jednakih duljina.*

2.4 Algoritam

Ideja

U ovom poglavlju ćemo detaljno opisati implementirani kod u kojem smo na dva načina pronašli maksimalnu kliku, te ćemo pritom koristiti težište u euklidskom prostoru. Ideja algoritma je, uz korištenje uređaja, sortirati vrhove grafa po njihovoj udaljenosti od neke zadane točke. Cilj je pronaći najdulji uzastopni niz kod kojeg između svaka dva vrha postoji brid, što će biti k-klika. Najveću takvu kliku maksimiziramo, tj. provjerimo postoji li još neki vrh grafa takav da je povezan sa svim vrhovima najveće k-klike. Zadanu točku od koje polazimo ćemo odrediti na dva načina. Prvi način je perturbacija težišta, gdje će nam polazna točka biti konveksna kombinacija težišta cijelog grafa i vrha koji je povezan sa najviše drugih vrhova. Dakle, na ovaj način biramo polazne točke tako da se šetamo unutar našeg gotovo regularnog n-simpleksa po spojnici čije su krajnje točke težište cijelog grafa i vrh koji je povezan sa najviše ostalih vrhova. U drugom načinu ćemo pronaći dva vrha koja su povezana sa najviše drugih vrhova, s uvjetom da postoji brid između njih, te će nam težište takva dva vrha biti polazna točka od koje ćemo sortirati udaljenosti vrhova grafa. Slijedi detaljan opis koda kojeg smo u cijelosti implementirali u programskom jeziku *Python*.

Opis koda

1) Generiranje matrice $B = [b_{i,j}]$ tako da se klika veličine k nalazi u gornjem lijevom kutu, odnosno $b_{i,j} = 1$ za $i, j = 1, \ldots, k$, $b_{i,j} \in \{0, 1\}$ za $i, j = k + 1, \ldots, n$ s vjerojatnošću pojavljivanja jedinice p.

2) Definiranje funkcije za računanje Grammove matrice.

3) Definiranje funkcije za računanje težišta u euklidskom prostoru.

4) Definiranje funkcije **maxniz(A)**: funkcija prima niz A koji se sastoji od uređenih parova (a_i, b_i), pri čemu a_i predstavlja udaljenost točke b_i od neke zadane točke, te je A sortiran uzlazno prema navedenim udaljenostima. Funkcija vraća najdulji uzastopni niz kod kojeg su sve točke međusobno povezane, odnosno za svaki par točaka postoji brid između njih.

5) Definiranje funkcije **maxklika(N)**: funkcija prima niz N koji se sastoji od točaka za koje vrijedi da između svake dvije postoji brid. Funkcija vraća niz K koji se sastoji od niza N i točaka grafa za koje vrijedi da su povezane sa svim točkama niza N tako da između svake dvije točke niza K postoji brid između njih.

6) Definiranje funkcije **makeDist(B)** za računanje matrice udaljenosti opisane u poglavlju 2.3. Funkcija prima matricu susjedstva B, te računa elemente matrice udaljenosti D po formuli:

$$D_{i,j} = 1 - (t_i + t_j)10^{-8}, (i, j) \in E$$
$$D_{i,j} = 1 + (t_i + t_j)10^{-8}, (i, j) \notin E,$$

gdje je t_i = broj vrhova sa kojima je vrh i povezan.

7) Računanje Grammove matrice.

8) Definiranje funkcije **emb2(B)**. Funkcija prima matricu susjedstva B, te koristeći teorem 2.2.4 provodi ulaganje grafa u euklidski prostor. Zbog pozitivnosti svojstvenih vrijednosti Grammove matrice, to će ulaganje biti izometričko.

9) Funkcija **emb2** šalje prvi vektor u ishodište, stoga ga nadodajemo ručno.

10) Ovdje opisujemo dva načina za pronalazak maksimalne klike.

1. NAČIN:

- pronađemo vrh koji je povezan sa najviše drugih vrhova
- izračunamo težište cijelog grafa i računamo 19 konveksnih kombinacija tog težišta i vrha koji je povezan sa najviše drugih vrhova s korakom 0.05
- uzlazno sortiramo udaljenosti svih vrhova grafa od svake konveksne kombinacije
- za svaki od 19 uzlazno sortiranih nizova pozivamo funkciju **maxniz** koja vraća najdulji uzastopni niz kod kojeg su sve točke međusobno povezane
- pozivamo funkciju **maxklika** za svaki niz koji vrati funkcija **maxniz**
- povratne vrijednosti funkcije **maxklika** su tražene maksimalne klike

2. NAČIN:

- pronađemo dva vrha koja su povezana sa najviše drugih vrhova, takva da postoji brid između njih, te izračunamo njihovo težište
- uzlazno sortiramo udaljenosti svih vrhova grafa od tog težišta
- za dobiveni uzlazno sortirani niz pozivamo funkciju **maxniz** koja vraća najdulji uzastopni niz kod kojeg su sve točke međusobno povezane
- pozivamo funkciju **maxklika** za niz koji vrati funkcija **maxniz**
- povratna vrijednost funkcije **maxklika** je tražena maksimalna klika

Napomena 2.4.1. *Napomenimo da smo kod prvog načina za svaki graf računali 19 maksimalnih klika, odnosno imali smo 19 polaznih točaka od kojih smo sortirali udaljenosti, dok smo kod drugog načina za svaki graf računali po jednu maksimalnu kliku. Jasno je da bi kod prvog načina svih 19 klika trebale biti jednake s obzirom da se radi o istom grafu.*

Napomena 2.4.2. *Primijetimo, ukoliko je veličina maksimalne klike puno veća od očekivane 1.2.12, tada je možemo uočiti iz težišta grafa. Više o ovom slučaju ćemo vidjeti u 3.2.*

Napomena 2.4.3. *Uočimo da je polazna točka od koje sortiramo udaljenosti blizu ortogonalu maksimalnog niza (povratne vrijednosti funkcije **maxniz**), odnosno ortogonalu naše k-klike.*

Poglavlje 3

Analiza rezultata

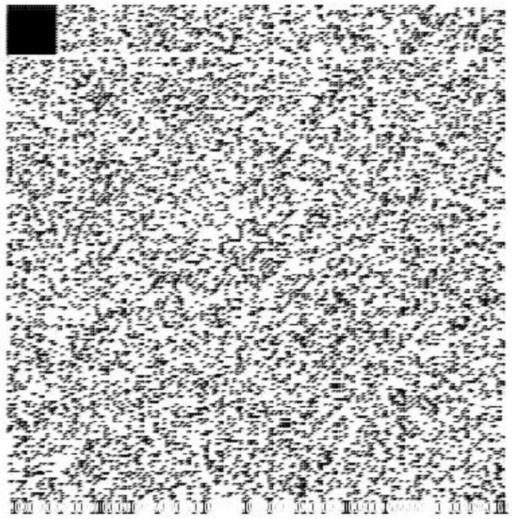
3.1 Generiranje slučajnih matrica

Sada ćemo promatrati grafove različitih veličina. Definirat ćemo matrice susjedstva tako da nam se klika veličine k nalazi u gornjem lijevom kutu, dok nam je vjerojatnost pojavljivanja jedinice na ostalim mjestima jednaka p. Dakle, promatramo slučajne grafove $G = (n, p)$ s klikama veličine k, gdje parametre n, p i k mijenjamo.

3.2 Primjeri i rezultati

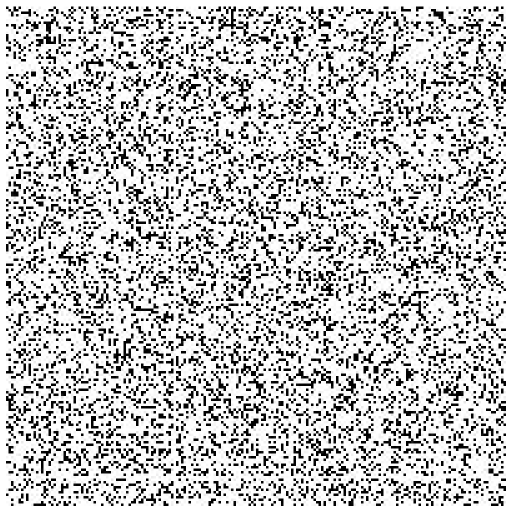
Primjere ćemo zapisivati kao uređene trojke (n, k, p), gdje nam je n veličina grafa, k veličina klike i p vjerojatnost pojavljivanja jedinice. Generirali smo 17 primjera od kojih je 6 primjera $(200, 20, 0.25)$, 5 primjera $(500, 50, 0.25)$, te po jedan primjer od uređenih trojki: $(200, 20, 0.4)$, $(500, 50, 0.4)$, $(500, 20, 0.25)$, $(1000, 50, 0.25)$, $(1000, 50, 0.4)$ i $(2000, 50, 0.25)$.

Uzmimo sada jedan primjer uređene trojke $(200, 20, 0.25)$. Prvo generiramo matricu veličine 200 koja sadrži kliku veličine 20 u gornjem lijevom kutu, a vjerojatnost pojavljivanja jedinice na ostalim mjestima je jednaka 0.25. Generiranu matricu susjedstva smo reprezentirali bijelim i crnim točkama tako da smo stavili bijelu točku na mjesto gdje ne postoji brid, odnosno crnu na mjesto gdje postoji. Na slici ispod možemo vidjeti kako izgleda matrica susjedstva na ovom primjeru.



Slika 3.1: Matrica susjedstva

Uočimo kako je gornji lijevi kut matrice potpuno zacrnjen, kako bi i trebalo biti, s obzirom da se tu nalazi generirana klika veličine 20. Nakon što permutiramo danu matricu dobijemo sljedeće:



Slika 3.2: Permutirana matrica susjedstva

Vidimo da nam sada klika nije lako uočljiva.

U sljedećem koraku računamo Grammovu matricu G i njezine svojstvene vrijednosti. Sve svojstvene vrijednosti su rastući pozitivni realni brojevi, a nekoliko prvih i posljednjih redom iznose:

```
0.49997694   0.49997732   0.49997757   0.49997840   0.49997868
0.50002311   0.50002343   0.50002399   0.50003784   100.00007267
```

Po teoremu znamo da tada postoji izometričko ulaganje točaka $p_0, \ldots, p_{199}$ tako da $p_0 \mapsto 0$, a slika vektora $p_1, \ldots, p_{199}$ je dana stupcima matrice $\sqrt{D}C$, pri čemu je D dijagonalna matrica sa svojstvenim vrijednostima Grammove matrice na dijagonali, a C je matrica čiji su stupci svojstveni vektori Grammove matrice. Jer p_0 šaljemo u nul-vektor, ručno ga nadodajemo. U sljedećem koraku pokušavamo pronaći kliku perturbacijom težišta. Prvo pronađemo vrh koji je povezan sa najviše drugih vrhova. Zatim izračunamo težište cijelog grafa i računamo 19 konveksnih kombinacija tog težišta i vrha koji je povezan sa najviše drugih vrhova s korakom 0.05, te uzlazno sortiramo udaljenosti svih vrhova grafa od svake konveksne kombinacije. Sada za svaki od 19 uzlazno sortiranih nizova pozivamo funkciju **maxniz** koja vraća najdulji uzastopni niz kod kojeg su sve točke međusobno povezane. Pozivamo funkciju **maxklika** za svaki niz koji vrati funkcija **maxniz**, te su povratne vrijednosti funkcije **maxklika** tražene maksimalne klike. Ovdje je važno napomenuti da iz svih 19 konveksnih kombinacija uspijemo pronaći cijelu kliku. Niz koji vraća funkcija **maxniz** za konveksnu kombinaciju s koeficijentima $\alpha = 0.95$ i $\beta = 0.05$ izgleda ovako:

$$[12, 18, 15, 11, 17, 4, 13, 8, 3, 2, 14, 7, 5, 16, 9, 0],$$

a njegovom maksimizacijom dobijemo:

$$[12, 18, 15, 11, 17, 4, 13, 8, 3, 2, 14, 7, 5, 16, 9, 0, 1, 6, 10, 19],$$

odnosno cijelu kliku. Pokušajmo sada pronaći kliku koristeći drugi način. Prvo pronađemo dva vrha koja su povezana sa najviše drugih vrhova, takva da postoji brid između njih, te izračunamo njihovo težište. Uzlazno sortiramo udaljenosti svih vrhova grafa od tog težišta, te za tako dobiveni niz pozivamo funkciju **maxniz**, koja nam vraća sljedeće:

$$[9, 0, 16, 5, 7, 3, 8, 4, 15, 18],$$

a njegovom maksimizacijom dobijemo:

$$[9, 0, 16, 5, 7, 3, 8, 4, 15, 18, 1, 2, 6, 10, 11, 12, 13, 14, 17, 19].$$

Dakle, koristeći oba načina uspijemo pronaći cijelu kliku, koja je u ovom slučaju duljine 20. Napomenimo samo da smo vrhove grafa označavali brojevima $0, \ldots, 199$, pa nam klika duljine 20 sadržava vrhove $0, \ldots, 19$, što smo i dobili kod oba načina. Važno je naglasiti da nam je očekivana veličina najveće maksimalne klike $2\log_{\frac{1}{0.25}} 200 \approx 7$, dok je duljina niza koji vraća funkcija **maxniz** kod prvog načina jednaka 16, a kod drugog 10, što je i dalje veće od očekivane veličine.

Analogno, u svim ostalim primjerima oblika (n, k, p), koristeći oba načina, uspijemo pronaći cijelu k-kliku. Također je važno napomenuti da kod svih primjera u prvom načinu uspijemo pronaći kliku za svaku od 19 konveksnih kombinacija. Dakle, nije nam bitno gdje se nalazimo na spojnici čije su krajnje točke težište grafa i vrh koji je povezan sa najviše drugih vrhova, jer će nam svaka od tih 19 polaznih točaka dati sortiran niz iz kojeg, pomoću funkcija **maxniz** i **maxklika**, uspijemo pronaći cijelu kliku veličine k.

Sada ćemo navesti primjer gdje nam je veličina generirane klike puno veća od očekivane. Neka je $n = 2000, k = 100$ i $p = 0.25$. Tada je očekivana veličina najveće maksimalne klike jednaka $2\log_{\frac{1}{0.25}} 2000 \approx 10$, dok je naša klika veličine 100. U ovakvim slučajevima kliku možemo uočiti već iz težišta grafa. Sortiranjem udaljenosti vrhova grafa od njegovog težišta , te pozivanjem funkcije **maxniz** dobijemo kliku veličine 80, što znači da nam je već početna klika čak 8 puta veća od očekivane, a njezinom maksimizacijom dobijemo cijelu kliku veličine 100.

Dakle, iako se bavimo problemom pronalaženja pseudo-klika, za svaki od 17 generiranih primjera smo uspjeli pronaći cijelu kliku. Generirane klike nisu bile puno veće od očekivanih, ali možemo reći da su naši primjeri bile dosta "rijetke" generirane matrice, gdje je najveća vjerojatnost pojavljivanja jedinice bila 0.4. Navedimo još da su složenosti funkcija **sort** i **maxniz** jednake i iznose $n \log n$, pa je složenost algoritma jednaka $n^2 \log^2 n$.

Na kraju, možemo zaključiti da našim geometrijskim pristupom uspješno pronalazimo cijele klike. Nažalost, to samo možemo reći za spomenutih 17 primjera, iako očekujemo jednaku uspješnost kod slično generiranih matrica. Sigurni smo da bi raznim modifikacijama ovog algoritma mogli poboljšati njegove rezultate, konkretno optimizacijom polaznih točaka. Tako smo kod prvog načina umjesto konveksne kombinacije dvije točke, mogli promatrati konveksne kombinacije težišta grafa i svih njegovih n vrhova, ali takve modifikacije ostavljamo za neka buduća istraživanja.

Bibliografija

[1] D. Bakić, *Linearna algebra*, Školska knjiga, Zagreb, 2008.

[2] Š. Ungar, *Metrički prostori*, predavanja, 2016., dostupno na: https://web.math.pmf.unizg.hr/~ungar/Analiza3_internet.pdf.

[3] J. Okopni, *Pozitivno definitne matrice*, Završni rad, Sveučilište Josipa Jurja Strossmayera u Osijeku, Odjel za matematiku, 2017.

[4] D. Veljan, *Kombinatorna i diskretna matematika*, Algoritam, Zagreb, 2001.

[5] https://cs.stackexchange.com/questions/26180/expected-number-of-maximal-cliques-in-gn-p.

[6] M. Brunato, H.H.Hoos, R.Battiti, *On Effectively Finding Maximal Quasi-Cliques in Graphs*, Universita di Trento, Trento, Italy, University of British Columbia, Vancouver, BC, Canada

[7] P. Nujić, *Konveksni skupovi*, Diplomski rad, Sveučilište Josipa Jurja Strossmayera u Osijeku, Odjel za matematiku, 2017.

[8] https://en.wikipedia.org/wiki/Simplex.

[9] https://bartoszmilewski.com/2018/12/

[10] https://math.stackexchange.com/questions/2739915/radius-of-inscribed-sphere-of-n-simplex#:~:text=By%20scaling%2C%20a%20regular%20n,)%20and%20Rr%3Dn.

[11] https://en.wikipedia.org/wiki/Bron%E2%80%93Kerbosch_algorithm

[12] K. Martinić, *Maksimalne klike u analizi sličnosti proteinskih motiva*, Diplomski rad, Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet, 2018.

[13] I. Čermak, *K-means clustering u prostoru Minkowskog*, Diplomski rad, Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet, 2011.

[14] F. Janjić, *Semantičko indeksiranje i klasifikacija dokumenata*, Diplomski rad, Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet, 2019.

[15] https://web.math.pmf.unizg.hr/~hari/LA.pdf.

[16] H. Maehara, *Regular embeddings of a graph*, Pacific Journal of Mathematics, 1983.

Sažetak

U ovom diplomskom radu bavili smo se problemom pronalaženja klika u grafovima. Implementirani algoritam rješava problem korištenjem geometrijskog pristupa. Ideja algoritma je bila sortirati udaljenosti vrhova grafa od neke određene točke iz koje iščitamo kliku, a njezinom maksimizacijom dobijemo traženu maksimalnu kliku u grafu.

Na početku smo naveli pojmove koje koristimo u radu, a zatim smo na skupu vrhova grafa definirali metriku te smo izometrički uložili graf u euklidski prostor. Naš algoritam smo proveli na 17 primjera uz različite veličine grafova, različite vjerojatnosti bridova i različite veličine klika. Za sve generirane primjere algoritam je uspješno pronašao klike, iako veličine klika nisu bile znatno veće od očekivanih.

Summary

This thesis is concerned with finding cliques in graphs. We present a solution to this problem by developing a geometric approach. Given a graph $G = (V, E)$, we construct a finite metric space (X, d) and consider an isometric embedding of X into $\mathbb{E}^n$. Using the properties of the embedded object, we construct a suitable order on V, that enables us to find large cliques efficiently, and these can easily be extended to either maximal cliques or pseudo-cliques.

After introducing the necessary mathematical objects and notions, we give a detailed description of our method. We test our algorithm on 17 random matrices of various sizes and properties. The largest maximal clique has been successfully detected in each case, although its size was only slightly larger than expected.

Životopis

Rođena sam 09.08.1997. u Makarskoj. Osnovnoškolsko obrazovanje završila sam 2012. godine u Osnovnoj školi dr. Franje Tuđmana u Brelima, te iste godine upisujem opću gimnaziju u Srednjoj školi fra Andrije Kačića Miošića u Makarskoj. Godine 2016. upisujem preddiplomski sveučilišni studij Matematika na Matematičkom odsjeku Prirodoslovno-matematičkog fakulteta u Zagrebu. Završetkom preddiplomskog studija 2021. godine stekla sam titulu sveučilišne prvostupnice Matematike, nakon čega upisujem diplomski studij Matematička statistika na istom fakultetu.