

GRÁFOK SPEKTRUMA

SZAKDOLGOZAT

DUDÁS-MARX LÁSZLÓ

TÉMAVEZETŐ: BENCZÚR A. ANDRÁS PH.D.
OPERÁCIÓKUTATÁS TANSZÉK



EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR

Tartalomjegyzék

1. Bevezetés	4
2. Szinguláris felbontás	4
2.1. Vektorok és mátrixok normája	5
2.1.1. Vektornormák	5
2.1.2. Mátrixok normái	7
2.2. Szinguláris felbontás létezése	10
2.3. HITS-algoritmus	11
2.3.1. Az algoritmus	12
2.3.2. Az algoritmus elemzése	13
3. A spektrum	15
3.1. A Laplace-mátrix és a normalizált Laplace-mátrix	15
3.2. A spektrum néhány egyszerű tulajdonsága	17
3.3. Néhány speciális gráf spektruma	18
4. Vágások és partíciók	19
4.1. Vágások	20
4.2. Izoperimetrikus feladatok	23
4.3. Bipartíciók	25
4.3.1. A Cheeger-konstans	25
4.3.2. A Cheeger-egyenlőtlenség	26
4.4. Klaszterezés	29
4.4.1. Reprezentánsok keresése QP programmal	29
4.4.2. Spektrálklaszterezés	31
4.5. Spektrálklaszterezés több sajátvektorral	32

5. Spektráلكlaszterezés a gyakorlatban	33
5.1. A spektráلكlaszterezés egy gyengesége	33
5.1.1. A TKC-hatás	33
5.2. Lehetséges módszerek a TKC-hatás semlegesítésére	34
5.2.1. Csomópontok eltávolítása	34
5.2.2. Csápok összehúzása	34
5.3. Spektráلكlaszterezés egy konkrét példán bemutatva	35
5.3.1. Az adat: livejournal.com	35
5.3.2. A kísérlet menete	36
5.3.3. Eredmények	36
6. Összefoglalás	38

Kivonat

Jelen dolgozat célja, hogy a spektrálgráfelméletbe betekintést nyújtson. Bemutatjuk a téma előzményeit, olyan főbb fogalmait, mint a Laplace-mátrix, a Fiedler-vektor, vagy a Cheeger-konstans.

A szakirodalomban bemutatott módszerekhez képest újdonság, hogy a HITS algoritmus segítségével bemutatjuk a sajátvektorok nagy abszolút értékkel reprezentált csúcsai és az utakat koncentráló sűrű részgráfok összefüggéseit. Valós hálózatokon kísérletekkel igazoljuk, hogy ha egy gráf nem rendelkezik „nagyléptékű struktúrákkal”, hanem a csúcsok kis-közepes csoportokba rendeződnek. Ezeknek a csoportoknak mindegyike egy-egy különböző, 1-hez közeli sajátértékű sajátvektorban adja a nagy abszolút értékű helyeket. Így a spektrum nagyon lassan csökken és nagyon nehéz jó kiegyensúlyozott vágást találni ezekben a gráfokban.

1. Bevezetés

A gráfok különböző mátrixainak karakterisztikus függvényeivel, sajátvektoraival és sajátértékeivel a spektálgráfelmélet foglalkozik. Ez a téma az 1950-es közepétől vált kutatott területté [11, 10]. A spektrálgráfelmélet egyik fő kérdésköre, hogy mit tudunk mondani egy gráf spektrumáról. A másik témakör pedig, hogy hogyan függ össze a spektrum a gráf egyéb tulajdonságaival.

A spektrálgráfelmélet a matematikán belül számos más területhez kapcsolódik. Szoros kapcsolatban van a differenciálgeometriával, a véletlen bolyongásokkal, valamint a Markov-láncokkal is. A gyakorlatban is sok haszna van a spektrumnak. Először a VLSI-áramkörök tervezésénél használták [6, 2, 1], de azóta számos más területen is sikeresen alkalmazták, mint a fehérjehálózatok, szociális hálók struktúrájának felderítése, vagy képszegmentáció.

A dolgozat elején egy általánosabb eszközt, a mátrixok szinguláris felbontását mutatjuk be. Bizonyítjuk ennek létezését, valamint megmutatjuk Kleinberg HITS-algoritmusával [17] való kapcsolatát. Ezt követően definiáljuk a legfontosabb fogalmakat, mint egy gráf Laplace-mátrixa, annak súlyozott változata és utóbbi sajátértékei, a spektrum. Bemutatjuk a spektrummal kapcsolatos alapösszefüggéseket. A következő fejezetben a spektrumot különböző bi- és multipartíciók értékeinek becslésére, és maguknak az optimálist közelítő partícióknak a meghatározására használjuk. A dolgozat végén egy valós példán mutatjuk be az elmélet egy lehetséges gyakorlati alkalmazását.

2. Szinguláris felbontás

Legelőször egy általános keretek közt alkalmazható módszert, a szinguláris felbontást szeretnénk bemutatni.

A mátrixok szinguláris felbontásának (*Singular value decomposition, SVD*) létezését már a XIX. század végén bizonyította egymástól függetlenül Beltrami [3] és Jordan [15]. A szinguláris felbontást azóta számos fontos területen alkalmazzák. Ilyenek például a homogén lineáris egyenletek megoldása, pszeudinverz számolása, kis rangú mátrixszal való közelítés, a legkisebb

négyzetek módszere, vagy a főkomponens analízis.

Bár az SVD mátrixok felbontását adja meg, könnyen és gyorsan tudjuk gráfokra is alkalmazni. Ezt úgy tehetjük meg, hogy vesszük a gráf valamelyik mátrix alakú reprezentációját. Egy ilyen alkalmazásra mutat jó példát majd a HITS-algoritmus.

2.1. Vektorok és mátrixok normája

Hogy a szinguláris felbontással kapcsolatos tételeket bizonyítani tudjuk, szükségünk lesz a mátrixnormák, főként a *Frobenius-norma* ismeretére.

A $\nu: \mathbb{R}^{n \times m} \rightarrow \mathbb{R}$ leképezést *mátrixnormának* nevezzük, ha tetszőleges $A, B \in \mathbb{R}^{n \times m}$ mátrixszal és $\alpha \in \mathbb{R}$ számmal kielégíti az alábbi három összefüggést:

$$\begin{aligned} A \neq 0 &\implies \nu(A) > 0 \\ \nu(\alpha A) &= |\alpha| \nu(A) \\ \nu(A + B) &\leq \nu(A) + \nu(B) \end{aligned}$$

A mátrixnorma $m = 1$ speciális eseteként adódik a *vektornorma*. A következőkben néhány szót ejtünk erről, azt követően pedig a mátrixnormákról.

2.1.1. Vektornormák

Vektorok *2-normáját* az alábbi egyenlőséggel definiáljuk.

2.1. definíció (2-norma). $v \in \mathbb{R}^n$ *2-normája*

$$\|v\|_2 \stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^n |v_i|^2} = \sqrt{v^T v}.$$

Az első két normaaxióma nyilvánvalóan teljesül, a háromszög-egyenlőtlenség pedig a Cauchy-egyenlőtlenség felhasználásával igazolható. A Cauchy-egyenlőtlenség vektorok skaláris szorzata és 2-normája közti kapcsolatot írja le:

2.2. tétel (Cauchy-egyenlőtlenség). *Ha $x, y \in \mathbb{R}^n$, akkor*

$$|x^T y| \leq \|x\|_2 \|y\|_2,$$

ahol egyenlőség pontosan akkor van, ha x és y lineárisan összefüggők.

Vektornormákat más módon is definiálhatnánk, például négyzetgyökvonás és négyzetre emelés helyett tekinthetnénk $1/p$ -edik illetve p -edik hatványt is. Azonban a 2-normát egy fontos tulajdonság emeli ki a lehetséges vektornormák közül.

2.3. definíció (Unitér-invariancia vektornormákra). *Egy vektornormát unitér-invariánsnak nevezünk, ha tetszőleges x vektor és U ortogonális mátrix esetén*

$$\|x\| = \|Ux\| .$$

Mielőtt megmutatjuk, hogy a 2-norma rendelkezik a fenti tulajdonsággal, megadjuk az ortogonális, vagy unitér mátrixok definícióját és egy fontos tulajdonságát.

2.4. definíció. *Egy mátrixot ortogonálisnak, vagy unitérnek nevezük, ha $U^T U = I$, azaz oszlopai egységheosszúak és egymásra páronként merőlegesek.*

2.5. lemma. *Egy ortogonális mátrixra $U U^T = I$.*

Bizonyítás: Mivel $U^T U = I$, $U^{-1} = U^T$. Az állítás adódik felhasználva, hogy a mátrix-szorítás esetében a bal és jobb inverz azonos. \square

Most megmutatjuk, hogy a 2-norma unitér invariáns:

$$\|x\|_2^2 = x^T x = x^T U^T U x = (Ux)^T (Ux) = \|Ux\|_2^2 ,$$

tetszőleges x vektorral és U ortogonális mátrixszal, ahonnan adódik az állításunk helyessége. Másrészt a fordított irányhoz legyen $\|\cdot\|$ egy tetszőleges unitér-invariáns norma, és x_1 egy olyan vektor, melyre $\|x_1\| = 1$. Ekkor az $\{x \mid \|x\| = 1\}$ halmaz elemei felírhatóak Ux_1 alakban valamely U ortogonális mátrixszal. Az előző egyenlőséglánc felhasználásával e halmaz elemeinek 2-normája megegyezik a $\|x_1\|_2$ mennyiséggel. Ezzel igazoltuk a következő állítást:

2.6. állítás. *Egy $\|\cdot\|$ vektornorma pontosan akkor unitér-invariáns, ha valamely $c \geq 0$ konstansra $\|\cdot\| = c\|\cdot\|_2$, azaz ha a 2-norma konstans többszöröse.*

2.1.2. Mátrixok normái

Most áttérünk mátrixok normáinak tárgyalására. Két különböző mátrixnormát fogunk bevezetni és használni. Először a vektorok 2-normájához hasonló módon definiálunk mátrixnormát, amelyet *Frobenius-normának* nevezünk:

2.7. definíció (Frobenius-norma).

$$\|A\|_F \stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^n \sum_{j=1}^m A_{ij}^2},$$

A Frobenius-norma egy a Cauchy-egyenlőtlenséghez hasonló összefüggést is kielégít.

2.8. állítás. *Tetszőleges $A \in \mathbb{R}^{n \times m}$, $B \in \mathbb{R}^{m \times p}$ mátrixok esetén*

$$\|AB\|_F \leq \|A\|_F \|B\|_F.$$

A bizonyítás elemi lépésekkel végezhető a Cauchy-egyenlőtlenség felhasználásával. Egy mátrixnormát *konzisztensnek* nevezünk, ha összeszorozható mátrixok és szorzatuk kielégíti a fenti egyenlőtlenséget. Ezzel a tulajdonsággal nem minden mátrixnorma rendelkezik, pedig a konzisztenciára nagy szükség van iterációs algoritmusok hibabecslésénél.

A második általunk használt mátrixnormát úgy definiáljuk, hogy az konzisztens legyen a vektorok 2-normájával, azaz szeretnénk, ha minden x vektorra és A mátrixra teljesülne a következő egyenlőtlenség:

$$\|Ax\|_2 \leq \|A\| \|x\|_2.$$

Az egyenlőtlenségből alsó becslés adódik $\|A\|$ értékére. Ezzel az alsó becsléssel definiáljuk mátrixok *2-normáját*, más néven *spektrálnormáját*.

2.9. definíció (Spektrálnorma).

$$\|A\|_2 \stackrel{\text{def}}{=} \max_{\|x\|_2=1} \|Ax\|_2.$$

A képletet röviden úgy értelmezhetjük, hogy egy A mátrix esetén $\|A\|_2$ egy x vektor A mátrixszal történő szorzásával elérhető „legnagyobb nyújtás” mértékét határozza meg. A normaaxiómák teljesülése mindkét esetben egyszerűen ellenőrizhető a vektorok 2-normájának tulajdonságaiból.

Most néhány olyan összefüggést mutatunk be, melyek szintén meghatározzák az eddig megismert mátrixnormákat. Tetszőleges $A \in \mathbb{R}^{n \times m}$ mátrix esetén az $A^T A$ mátrix szimmetrikus és pozitív szemidefinit, ezért sajátértékei nemnegatív valós számok, melyeket $\lambda_m \geq \dots \geq \lambda_1 \geq 0$ jelöl.

2.10. állítás.

$$\|A\|_F = \sqrt{\text{tr}(A^T A)} \quad (1a)$$

$$\|A\|_F = \sqrt{\lambda_1 + \lambda_2 + \dots + \lambda_m} \quad (1b)$$

$$\|A\|_2 = \sqrt{\lambda_m} \quad (1c)$$

$$\|A\|_2 = \max_{\|x\|_2=1, \|y\|_2=1} |y^T Ax| \quad (1d)$$

$$\|A\|_2 = \|A^T\|_2 \quad (1e)$$

Bizonyítás: Az $A^T A$ mátrix főátlójában szereplő j -edik elem megegyezik az A mátrix j -edik oszlopában előforduló elemek négyzetösszegével. Így $\text{tr}(A^T A)$ nem más mint A összes elemének négyzetösszege. Ezzel igazoltuk az (1a) egyenlőséget.

Az (1b) összefüggés pedig (1a) következménye, hiszen egy mátrix nyoma megegyezik sajátértékeinek összegével.

Az (1c) összefüggést a következő módon kapjuk meg:

$$\|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2 = \max_{\|x\|_2=1} \sqrt{|x^T A^T Ax|} = \sqrt{\lambda_1}.$$

A (1d) egyenlőség esetén a \geq irány a Cauchy-egyenlőtlenségből adódik, mert

$$\|Ax\|_2 = \|y\|_2 \|Ax\|_2 \geq |y^T Ax|,$$

tetszőleges y és x vektorra, melynél $\|x\|_2 = \|y\|_2 = 1$. Legyen x_1 az $\|Ax\|_2$ maximumát beállító vektor, ezt behelyettesítve az előző egyenlőtlenségbe kapjuk a \geq irányt. Másrészt x_1 és $y_1 := \frac{1}{\|Ax_1\|_2} Ax_1$ választással

$$|y_1^T Ax_1| = \|Ax_1\|_2 = \|A\|_2$$

adódik, amivel a \leq irányt is beláttuk. Végül az (1e) összefüggés (1d) következménye, hiszen a kifejezésben egy skalár szerepel, amelynek transzponáltja önmaga, $|x^T A^T y|$, azaz az egyenlőségben A lecserélhető A^T -ra. \square

Mátrixok normájához is társítható a vektorokéhoz hasonló geometriai jelentés. Tetszőleges A mátrix esetén az

$$E_A \stackrel{\text{def}}{=} \{ Ax \mid \|x\|_2 = 1 \}$$

egy ellipszoidot határoz meg, melynek dimenziója $\text{rk}(A)$. A mátrix normája szemléletesen ennek az ellipszoidnak a „nagyságát” jelenti. Az ellipszoidot körbevehetjük egy olyan $\text{rk}(A)$ -dimenziós téglatesttel, melynek élei párhuzamosak a fő tengelyekkel, és az élek hossza megegyezik a fő tengelyek hosszával. Ekkor a Frobenius-norma megegyezik a téglatest leghosszabb testátlójával, a 2-norma pedig annak leghosszabb élével, vagyis a legnagyobb fő tengely hosszával. Ez (1b) és az (1c) egyenlőségből adódik, felhasználva, hogy az ellipszoid fő tengelyeinek hossza megegyezik $A^T A$ sajátértékeivel [24].

Ha egy A mátrixot az U^T és a V ortogonális mátrixokkal szorzunk, akkor E_A -ból forgatással nyerjük az $E_{U^T A V}$ ellipszoidot. A mátrix normájától azt várnánk, hogy ne változzon e művelet hatására. E tulajdonság azonban nem igaz tetszőleges normára.

2.11. definíció (Unitér-invariancia mátrixnormákra). *Egy $\|\cdot\|$ mátrixnormát unitér-invariánsnak mondunk, ha tetszőleges U és V ortogonális mátrixokkal*

$$\|U^T A V\| = \|A\|.$$

2.12. állítás. *A Frobenius- és a 2-normák unitér-invariánsak.*

Bizonyítás: A 2-normára már korábban beláttuk, a Frobenius-normára pedig nyilvánvaló, hogy egy mátrix és transzponáltja esetén megegyezik, ezért elég azt megmutatni, hogy egy ortogonális mátrixszal balról szorozva nem változik a norma. Utóbbi a 2-norma esetén abból adódik, hogy a maximalizálandó $\|Ax\|_2 = \|U^T Ax\|_2$ bármely x esetén. A Frobenius-norma unitér-invarianciája szintén

visszavezethető vektorok 2-normájára, hiszen az U mátrixszal szorozhatunk oszloponként, és a normát is számíthatjuk az oszlopokéból. \square

A dolgozat hátralevő részében vektorok 2-normájára egyszerűen a $\|\cdot\|$ jelölést, mátrixokéra pedig a $\|\cdot\|_2$ jelölést, Frobenius-normára a $\|\cdot\|_F$ jelölést használjuk.

2.2. Szinguláris felbontás létezése

Most már minden eszközünk megvan, hogy a szinguláris felbontást definiáljuk és létezését bizonyítsuk.

2.13. definíció (Szinguláris felbontás). Egy $A \in \mathbb{R}^{n \times m}$ mátrix szinguláris felbontásán az olyan

$$A = U\Sigma V^T$$

szorzattá bontást értjük, ahol $U \in \mathbb{R}^{n \times n}$ és $V \in \mathbb{R}^{m \times m}$ ortogonális mátrixok. A $\Sigma \in \mathbb{R}^{n \times m}$ mátrix pedig diagonális, és a „főátlójában” a $\sigma_1 \geq \dots \geq \sigma_r > 0$ pozitív számokat csupa 0 követi.

2.14. definíció (Szinguláris értékek). A σ_i értékeket az A mátrix szinguláris értékeinek hívjuk.

2.15. definíció (Szinguláris vektorok). A U és V oszlopait az A mátrix bal, illetve jobboldali szinguláris vektorainak hívjuk.

2.1. Tétel (SVD létezése). Tetszőleges $A \in \mathbb{R}^{n \times m}$ r rangú mátrixnak létezik szinguláris felbontása.

Bizonyítás: Teljes indukció $\min(m, n)$ -re. Ha $\min(m, n) = 1$, akkor az A mátrix egyedül az $a := A$ vektorból áll. Feltehetjük, hogy a oszlopvektor. Az a -hoz található olyan ortogonális U mátrix, melyre $U^T a = \|a\|e_1$. Így $\Sigma = \|a\|e_1$ és $V = (1)$ választással adódik a $\min(m, n) = 1$ eset.

Ha $\min(m, n) > 1$, akkor a problémát visszavezetjük egy $(n-1) \times (m-1)$ méretű mátrix felbontására. Feltehető, hogy $A \neq 0$, különben triviális a felbontás.

$$\sigma_1 := \|A\|_2 = \max_{\|v\|=1} \|Av\| \quad v_1 := \operatorname{argmax}_{\|v\|=1} \|Av\| \quad u_1 := \frac{1}{\sigma_1} Av_1$$

Egészítsük ki továbbá az u_1 és v_1 vektorokat $U_1 := (u_1 u_2 \dots u_n)$ és $V_1 := (v_1 v_2 \dots v_m)$ ortogonális mátrixokká. Ekkor a w , z és A_2 jelöléseket bevezetve

$$U_1^T A V_1 = \begin{pmatrix} \sigma_1 & w^T \\ z & A_2 \end{pmatrix}, \quad U_1^T A v_1 = \begin{pmatrix} \sigma_1 \\ z \end{pmatrix}, \quad V_1^T A^T u_1 = \begin{pmatrix} \sigma_1 \\ w \end{pmatrix}.$$

Az (1a) állítás szerint $\sigma_1 = \|A\|_2 = \|A^T\|_2$, ezért az $\|u_1\| = \|v_1\| = 1$ vektorok esetén $\|A u_1\| \leq \sigma_1$ és $\|A^T v_1\| \leq \sigma_1$. Innen a vektornorma unitér-invarianciáját felhasználva kapjuk, hogy

$$\left\| \begin{pmatrix} \sigma_1 \\ z \end{pmatrix} \right\| \leq \sigma_1 \quad \text{és} \quad \left\| \begin{pmatrix} \sigma_1 \\ w \end{pmatrix} \right\| \leq \sigma_1.$$

Ez viszont csak úgy lehetséges, ha $z = w = 0$, vagyis igazoltuk, hogy

$$U_1^T A V_1 = \begin{pmatrix} \sigma_1 & 0 \\ 0 & A_2 \end{pmatrix}.$$

Indukciós feltevésünk alapján az $(n-1) \times (m-1)$ méretű A_2 mátrixhoz található U_2 és V_2 ortogonális mátrixok, melyekkel $U_2^T A_2 V_2 = \Sigma_2$. Az

$$U^T := \begin{pmatrix} 1 & 0 \\ 0 & U_2^T \end{pmatrix} U_1^T \quad \text{és} \quad V^T := V_1 \begin{pmatrix} 1 & 0 \\ 0 & V_2 \end{pmatrix}$$

választással kapjuk a kívánt felbontását A -nak:

$$U^T A V = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix} = \Sigma.$$

□

2.3. HITS-algoritmus

Jon Kleinberg [17] 1998-ban született HITS-algoritmus (*Hyperlink-Induced Topic Search*) az internetes keresések találatait rangsorolja. A rangsort az oldalak közt levő hiperlinkek alapján állítja fel.

Az algoritmus ötlete, hogy két fontos csoportba osztja az oldalakat, központi (*hub*) és mértékadó (*authority*) oldalakra. Az algoritmus iteratív módon

minden oldalra frissíti a központiság és mértékadóság mértékét. Kleinberg meghatározásában egy oldal annál központibb, minél mértékadóbb oldalakra mutat, és annál mértékadóbb, minél központibb oldalak mutatnak rá.

A feladat megfogalmazható az irányított gráfok nyelvén is. A weboldalak a csúcsok, az éleket pedig a hiperlinkek határozzák meg. Az $auth, hub : V \rightarrow \mathbb{R}$ függvényeket használjuk majd az egyes oldalak mértékadóságának és központiságának jelzésére.

2.3.1. Az algoritmus

A HITS heurisztikája szerint egy oldal központiságát az határozza meg, hogy mennyire mértékadó oldalakra mutat rá. Tehát a központiságot a ki-élek határozzák meg.

Feltevésünk szerint egy oldal akkor lesz mértékadóbb, ha a központibb oldalak mutatnak rá. Így a mértékadóság a be-élekből fog származni.

Az algoritmus a következőképpen működik.

HITS-algoritmus

1. Tetszőleges $auth$, és hub kezdőértéket választunk.
2. Frissítjük minden csúcs $auth$ értékekét a beélek mentén a szomszédok aktuális hub értéke szerint.

$$auth^{(k+1)}(v) = \sum_{(w,v) \in E} hub^{(k)}(w)$$

3. Frissítjük minden csúcs hub értékét a kiélek mentén a szomszédok aktuális $auth$ értéke szerint.

$$hub^{(k+1)}(v) = \sum_{(v,w) \in E} auth^{(k+1)}(w)$$

4. Normáljuk az $auth$ és hub értékeket.
5. Ellenőrizzük, hogy az értékek változása a bemenetként megadott küszöb alatt van-e. Ha igen, megállunk.

6. Visszatérünk a 2. pontra

A normalizációs lépés segítségével elkerülhetjük a numerikus túlsordulást. A normálás, mint konstanssal való szorzás, a központiság és mértékadóság sorrendjén nem változtat, tehát a leállási feltétellel összhangban teszőleges módon végrehajtható. Látni fogjuk, hogy az *auth* és *hub* értékek normálás nélkül általában végtelenhez tartanak.

2.3.2. Az algoritmus elemzése

Az algoritmus egy lépését az adjacenciamátrix segítségével írjuk fel, majd használjuk a szinguláris felbontást.

Készítsünk az *auth* és *hub* értékekből egy *a* és *h* sorvektort!

$$a^{(k)} = (\text{auth}^{(k)}(v_1), \dots, \text{auth}^{(k)}(v_n))$$

$$h^{(k)} = (\text{hub}^{(k)}(v_1), \dots, \text{hub}^{(k)}(v_n))$$

Az így definiált vektorokkal és az *A* adjacenciamátrixszal a következőképp írhatjuk fel az iterációs lépést:

$$a^{(k+1)} = h^{(k)} A$$

$$h^{(k+1)} = a^{(k+1)} A^T.$$

A két egyenletet egymásba helyettesítve az alábbiakat kapjuk.

$$a^{(k+1)} = a^{(k)} A^T A = a^{(1)} (A^T A)^k$$

$$h^{(k+1)} = h^{(k)} A A^T = h^{(1)} (A A^T)^k$$

Alkalmazzuk $A = U \Sigma V^T$ szinguláris felbontását. Mivel *A* négyzetes mátrix, ezért Σ is az, így $\Sigma = \Sigma^T$.

$$a^{(k+1)} = a^{(1)} (A^T A)^k = a^{(1)} V \Sigma U^T U \Sigma V^T V \Sigma \dots U \Sigma V^T = a^{(1)} V \Sigma^{2k} V^T$$

$$h^{(k+1)} = h^{(1)} (A A^T)^k = h^{(1)} U \Sigma V^T V \Sigma U^T U \Sigma \dots V \Sigma U^T = h^{(1)} U \Sigma^{2k} U^T \quad (2)$$

Most már jól látható, hogy a mértékadóságot és a központiságot a szinguláris értékek és vektorok határozzák meg. Normalizáljuk a Σ mátrixot σ_1 -gyel.

$$\Sigma_1 = \begin{pmatrix} \frac{\sigma_1}{\sigma_1} & 0 & \dots & 0 \\ 0 & \frac{\sigma_2}{\sigma_1} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \frac{\sigma_n}{\sigma_1} \end{pmatrix}, \quad \Sigma_1^{2k} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & (\frac{\sigma_2}{\sigma_1})^{2k} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & (\frac{\sigma_n}{\sigma_1})^{2k} \end{pmatrix}$$

Tehát az első elemet kivéve a mátrix összes többi eleme az iterációs számban exponenciális sebességgel tart 0-hoz.

Vizsgáljuk meg $A^T A$ hatását a $v(i)$ szinguláris vektorokra! $A^T A v_{(i)}$ -t

$$A^T A v_i = V \Sigma^T U^T U \Sigma V^T v_i = V \Sigma^2 V^T v_i = V \Sigma^2 \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = V \begin{pmatrix} 0 \\ \vdots \\ \sigma_i \\ \vdots \\ 0 \end{pmatrix} = \sigma_i v_i$$

Látható, hogy a szinguláris értékek az $A^T A$ sajátértékei. A jobb-szinguláris vektorok a sajátvektorok.

Ha $h^{(1)}$ -t úgy választjuk, hogy merőleges legyen U első oszlopára, akkor a numerikus hibáktól eltekintve az iterációk során végig a merőleges altérben maradunk, azaz elimináljuk az első szinguláris vektor hatását. Ilyenkor a második lesz a hangsúlyos szinguláris vektor, és a megfelelő normalizációs konstans σ_2 lesz.

A fenti gondolatmenet alapján a HITS-algoritmus alkalmas a szinguláris felbontás meghatározására. Ha már ismert az első k szinguláris vektor, akkor választhatunk rájuk merőleges kiinduló vektort. Ekkor a HITS normalizálás után a $(k+1)$ -edik szinguláris vektorpárhoz fog konvergálni.

Szemléletes jelentését is társíthatunk az A^k , valamint az $(A^T A)^k$ mátrixokhoz. a_{ij} -re felfoghatjuk úgy is, mint az i és j csúcsok közötti 1 hosszú séták számát. Most nézzük A^2 -t!

$$A_{ij}^2 = \sum_{k=1}^n a_{ik} a_{kj}$$

Látható, hogy A_{ij}^2 az i és j közötti 2, hosszú séták számával lesz egyenlő. Indukcióval bizonyítható, hogy ez magasabb hatványokra is igaz.

$(AA^T)^k$ esetében hasonló eredményre jutunk, de most minden második lépésünk a megfordított gráfon történik. Látható tehát, hogy az *auth* és *hub* értékek is az adott csúcsból induló és érkező oda-vissza utak számával arányosak.

3. A spektrum

Az előző fejezetben láttuk, hogy egy gráf adjacenciamátrixának sajátértékei és sajátvektorai különleges jelentéssel bírnak, és információt hordoznak a gráf struktúrájáról.

Ebben a fejezetben áttekintjük a gráfok spektrálgráfelméletben leggyakrabban használt mátrixreprezentánsait. Bevezetjük a spektrum fogalmát, és megvizsgáljuk annak alapvető tulajdonságait.

Majd a fejezet végén megadjuk néhány konkrét gráfcsalád spektrumát.

3.1. A Laplace-mátrix és a normalizált Laplace-mátrix

A dolgozat hátralevő részében olyan egyszerű súlyozott gráfokkal foglalkozunk, melyekhez adva van egy $w : V \times V \rightarrow \mathbb{R}$ súlyfüggvény, amelyre igazak az alábbiak:

$$w(u, v) = w(v, u)$$

$$w(u, v) \geq 0$$

$$uv \text{ nem él} \Leftrightarrow w(u, v) = 0$$

A $w(u, v)$ értékeket mátrixba rendezve kapjuk a W mátrixot. Vegyük észre, hogy a W önmagában meghatározza a gráfot, így a súlyozott gráfokat a V csúcshalmazzal és a $W \geq 0$ súlymátrixszal fogjuk megadni.

A súlyozott gráfok esetében definiálhatunk egy általánosabb fokszámfogalmat.

3.1. definíció (Általánosított fokszám). $G = (V, W)$ súlyozott gráf egy v csúcsának általánosított fokszáma

$$d_v = \sum_{w(u,v)>0} w(u, v).$$

Most már definiálhatjuk egy súlyozott gráf Laplace-mátrixát.

3.2. definíció (Laplace-mátrix). *Egy $G = (V, W)$ gráf Laplace-mátrixa*

$$L(u, v) = \begin{cases} d_v - w(u, v) & \text{ha } u = v \\ -w(u, v) & \text{ha } u \text{ és } v \text{ szomszédos} \\ 0 & \text{egyebkent} \end{cases}$$

Jelölje $D \in \mathbb{R}^{n \times n}$ a fokszámmátrixot, vagyis azt a mátrixot, amiben minden főátlón kívüli elem 0, a főátlóban pedig rendre d_v áll. Ezt a jelölést használva a Laplace-mátrixot a következő formában is felírhatjuk:

$$L = D - W.$$

A Laplace-mátrix sok előnyös tulajdonsággal rendelkezik. Szimmetrikus és pozitív-szemidefinit, emiatt a sajátértékei nemnegatív, valósak. Hogy a sajátértékeket korlátok közt tartsuk, érdemes a Laplace-mátrixot az általánosított fokszámokkal súlyozni. Így kapjuk a normalizált Laplace-mátrixot.

3.3. definíció (Normalizált Laplace-mátrix). *Egy $G = (V, W)$ gráf normalizált Laplace mátrixa*

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

3.4. definíció (Spektrum). *A $G = (V, W)$ gráfhoz tartozó \mathcal{L} normalizált Laplace-mátrix sajátértékeit*

$$0 = \lambda_0 \leq \lambda_1 \leq \lambda_2 \cdots \leq \lambda_{n-1}$$

a G spektrumának hívjuk.

A definícióból látható, hogy a spektrum nem függ a csúcsok sorrendjétől, tehát két izomorf gráf spektruma megegyezik. Fordítva azonban ez nem igaz, két azonos spektrumú gráf nem feltétlenül lesz izomorf.

3.2. A spektrum néhány egyszerű tulajdonsága

A következőkben megmutatjuk a Laplace mátrix legfontosabb tulajdonságait. Legelőször bebizonyítjuk, hogy \mathcal{L} szimmetrikus pozitív-szemidefinit, azaz az összes sajátértéke valós nemnegatív. A szimmetria abból következik, hogy a gráf irányítatlan.

A pozitív szemidefinittség bizonyításához vegyünk egy tetszőleges $g : V \rightarrow \mathbb{R}$ függvényt és tekintsük, mint oszlopvektort. Definiáljuk az ún. *Rayleigh-hányadost*:

$$\frac{\langle g, \mathcal{L}g \rangle}{\langle g, g \rangle} = \frac{\langle g, D^{-\frac{1}{2}}LD^{-\frac{1}{2}}g \rangle}{\langle g, g \rangle} = \frac{\langle f, Lf \rangle}{\langle D^{\frac{1}{2}}f, D^{\frac{1}{2}}f \rangle} = \frac{\sum_{u \sim v} (f(u) - f(v))^2 w(u, v)}{\sum_v f(v)^2 d_v}, \quad (3)$$

ahol $g = D^{\frac{1}{2}}f$ és $u \sim v$ jelöli azokat a rendezetlen csúcspárokat, amikre u és v szomszédosak, valamint $\langle \cdot, \cdot \rangle$ jelöli az \mathbb{R}^n -beli skalárszorzatot.

A (3) egyenlőségből következik, hogy az összes sajátérték nemnegatív és a 0 is sajátérték. A 0-hoz tartozó sajátvektor a $\sqrt{\underline{d}} = (\sqrt{d_1}, \dots, \sqrt{d_n})^T$ vektor. A 0 akkor és csak akkor egyszeres sajátérték, ha a gráf összefüggő. Ha a gráf nem összefüggő, akkor a spektrum az egyes összefüggő részek spektrumának uniója. Ezt abból is láthatjuk, hogy ilyenkor az adjacenciamátrix blokkosított alakra hozható. Ha $g \in \mathcal{L}$ sajátvektora, akkor a Rayleigh-hányados lesz a hozzá tartozó sajátérték.

Továbbá,

$$\lambda_1 = \inf_{f \perp D\mathbf{1}} \frac{\sum_{u \sim v} (f(u) - f(v))^2 w(u, v)}{\sum_v f(v)^2 d_v} = \min_{\substack{\|g\|=1 \\ g^T \sqrt{\underline{d}}=0}} g^T \mathcal{L}g \quad (4)$$

és

$$\lambda_{n-1} = \max_{\|g\|=1} g^T \mathcal{L}g. \quad (5)$$

Az előző *min* és *max* egyenlőség hasonlóképp igaz a Laplace-mátrixra és annak sajátértékeire. Érdekes az imént vizsgált kifejezést magasabb dimenzióra is kiterjeszteni.

$X \in \mathbb{R}^{n \times k}$ adott szubortogonális mátrix. X oszlopait az x_1, \dots, x_k vektorokkal,

sorait pedig az r_1^T, \dots, r_n^T vektorokkal jelöljük.

$$\begin{aligned} Q_k = \text{tr}(X^T L X) &= \sum_{l=1}^k x_l^T (D - W) x_l = \sum_{i=1}^n d_i \|r_i\|^2 - \sum_{i=1}^n \sum_{j=1}^n w_{ij} r_i^T r_j \\ &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} \|r_i - r_j\|^2. \end{aligned} \quad (6)$$

Vegyük észre, hogy a $k = 1$ esetben a Rayleigh-hányadost kapjuk vissza.

Tudjuk, hogy egy mátrix nyoma megegyezik a sajátértékeinek összegével. Emiatt

$$\sum_{i=0}^n \lambda_i = \text{tr}(\mathcal{L}) = n.$$

Így ki tudjuk számolni a spektrum átlagát, amivel becsülni tudjuk a legnagyobb és legkisebb pozitív sajátértéket.

$$\lambda_1 \leq \frac{\sum_{i=1}^{n-1} \lambda_i}{n-1} = \frac{n}{n-1} \leq \lambda_{n-1} \quad (7)$$

Látni fogjuk majd, hogy teljes gráfra egyenlőséggel teljesül mindkét egyenlőtlenség. Nem teljes gráfra azonban már az is igaz, hogy $\lambda_1 \leq 1$, lásd [7].

3.3. Néhány speciális gráf spektruma

K_n . Az n csúcsú teljes gráf spektruma

$$\lambda_0 = 0, \lambda_1 = \dots = \lambda_{n-1} = \frac{n}{n-1},$$

mivel a gráf teljesen szimmetrikus és mert $\text{tr}\mathcal{L} = n$ [4]. Ebből látható, hogy a (7) becslés éles.

$K_{m,n}$. Az m és n elemű osztályokból álló teljes páros gráf spektruma [4] szerint

$$\lambda_0 = 0, \lambda_1 = \dots = \lambda_{n+m-2} = 1, \lambda_{n+m-1} = 2.$$

S_n . Az n ágú, azaz $n+1$ csúcsú csillag spektruma, mivel $S_n = K_{n,1}$:

$$\lambda_0 = 0, \lambda_1 = \dots = \lambda_{n-1} = 1, \lambda_n = 2.$$

\mathbf{P}_n . Az n hosszú út spektruma [8] szerint

$$\lambda_k = 1 - \cos \frac{\pi k}{n-1} \quad k = 0, \dots, n-1.$$

\mathbf{C}_n . Az n hosszú kör spektruma [7] szerint:

$$\lambda_k = 1 - \cos \frac{2\pi k}{n} \quad k = 0, \dots, n-1.$$

\mathbf{Q}_n . A 2^n csúcsú n dimenziós hiperkocka sajátértékei [21] szerint

$$\frac{2k}{n} \binom{n}{k} \text{ multiplicitással, } k = 0, \dots, n.$$

4. Vágások és partíciók

Gráfok vizsgálatának egyik fontos kérdése, hogy hogyan tudjuk egy gráf csúcsait jól csoportosítani, partícionálni. A csoportosítás által egyszerűsödik, jobban átlátható lesz a gráf szerkezete. Előkerülhetnek rejtett kapcsolatok, ami a gráf struktúrájának jobb megértéséhez vezet. Newman [23] szerint a hálózatok kutatásának kiemelten fontos témája a közösségek, azaz sűrű részgráfok vizsgálata.

A gráfok klaszterei olyan csúcshalmazok, amelyek valamilyen jellemző tulajdonság alapján mérhetően elkülönülnek. Ilyenre példa a szociális hálókból az azonos érdeklődésű emberek csoportja, a fehérjehálózatokban az egymással reagáló vegyületek vagy képek esetében az alakzathoz tartozó pixelek.

Partícionálás során minden csúcst be kell sorolnunk egy klaszterbe. Tehát a klaszterezés egy k -részesre vágás, vagyis a $G = (V, W)$ gráf csúcsainak egy olyan $P_k = (V_1, \dots, V_k)$ felosztása, ahol V_i -k diszjunktak és

$$\bigcup_{i=1}^k V_i = V.$$

$k = 2$ esetén bipartícionálásról beszélünk.

A megfelelő partícionálás matematikailag nehezen megfogható fogalom. Az általunk használt mértékekhez sok esetben könnyen találunk olyan példát,

amikor az optimális partícionálás messze van az elvárttól. Egy jó partícionálástól elvárjuk, hogy az egy csoportban levő elemek hasonlítsanak egymásra, míg a különböző csoportokban levők különbözzenek. Valamint elvárt az is, hogy a klaszterek hasonló méretűek legyenek.

Az alábbiakban bevezetjük a vágás fogalmát, és bemutatunk különböző mértékeket vágásokra és partíciókra. Ezt követően megvizsgáljuk a bipartíciók és multipartíciók kapcsolatát a spektrummal.

4.1. Vágások

Intuíció alapján két csúcshalmazt különbözőnek tekintünk, ha könnyen el tudjuk választani őket egymástól. Érdekes tehát megvizsgálni, hogy hány élt kell elhagyni ahhoz, hogy kettéválasszuk őket. Az ilyen élhalmazokat vágásoknak hívjuk.

4.1. definíció (Vágás). *Egy gráf vágása az élek olyan részhalmaza, amely elhagyása után a gráf már nem összefüggő.*

$\mathcal{C}(G)$ -vel jelöljük a G gráf vágásainak halmazát. Ahhoz, hogy a vágásokat össze tudjunk hasonlítani, bevezetjük a vágás értékét.

4.2. definíció (Vágás értéke). *Egy $G = (V, W)$ gráf $C \subset E(G)$ vágásának értéke:*

$$w(C) = \sum_{(u,v) \in C} w(u,v),$$

ahol (u, v) rendezetlen párokat jelöl.

Egy bipartíció rögtön meghatároz egy vágást is.

$$\partial S = \{\{u, v\} \in E(G) : u \in S, v \notin S\}$$

Világos, hogy $\partial S = \partial \bar{S}$.

Tetszőleges két csúcshalmazhoz is meghatározhatjuk az általuk feszített részgráfban az elválasztásukhoz szükséges vágás értékét.

4.3. definíció (Vágás értéke). A $G = (V, W)$ élsúlyozott gráfhoz tartozó $S, T \in V$ nem üres csúcshalmazok által meghatározott vágás súlya [12] alapján

$$w(S, T) = \sum_{s \in S} \sum_{t \in T} w_{st}.$$

A fenti definíciót kiterjeszthetjük tetszőleges számú halmazra.

4.4. definíció (k-vágás értéke).

$$cut(P_k, G) = \sum_{a=1}^{k-1} \sum_{b=a+1}^k w(V_a, V_b)$$

Láttuk, hogy a spektrum erős összefüggésben van a gráfok szerkezetével, így nem meglepő, hogy a minimális vágás költsége sem független a spektrumtól. Felhívnanánk a figyelmet, hogy a következő állításban a sima Laplace-mátrix spektrumát használjuk fel.

4.5. állítás. Az n csúcsú $G = (V, W)$ összefüggő gráf L Laplace-mátrixának sajátértékeit jelölje $0 = \lambda_0 < \lambda_1 \leq \dots \leq \lambda_{n-1}$. Ekkor

$$\min_{C \in \mathcal{C}(G)} w(C) \geq \frac{n-1}{n} \lambda_1.$$

Bizonyítás:

Legyen U^* egy minimum vágás. Az U segítségével definiáljuk az alábbi egydimenziós reprezentánsokat.

$$r_i = \begin{cases} n - |U^*| & \text{ha } i \in U^* \\ -|U^*| & \text{ha } i \in \overline{U^*} \end{cases}$$

Legyen \mathbf{r} az r_i értékeket összefogó vektor. Normáljuk \mathbf{r} -t.

$$\|\mathbf{r}\| = \sum_{i=1}^n r_i^2 = |U^*|(n - |U^*|)^2 + (n - |U^*|)|U^*|^2 = |U^*|(n - |U^*|)n$$

$$\tilde{\mathbf{r}} = \frac{\mathbf{r}}{\sqrt{|U^*|(n - |U^*|)n}}$$

Az (5) és (6) alapján:

$$\lambda_1 \leq \tilde{\mathbf{r}}^T L \tilde{\mathbf{r}} = \sum_{i < j} w_{ij} (\tilde{r}_i - \tilde{r}_j)^2 = w(U^*, \overline{U^*}) \frac{(n - |U^*| + |U^*|)^2}{|U^*|(n - |U^*|)n}.$$

Ebből következik, hogy

$$\min_{C \in \mathcal{E}(G)} w(C) \geq \frac{|U^*|(n - |U^*|)}{n} \lambda_1 \geq \frac{n-1}{n} \lambda_1.$$

□

Az előző bizonyítás menetét követve könnyen adódik az alábbi állítás.

4.6. állítás (Max-vágás súlya). *Legyenek a $0 = \lambda_0 \leq \lambda_1 \cdots \leq \lambda_{n-1}$ a $G = (V, W)$ gráfhoz tartozó L Laplace-mátrix sajátértékei. Ekkor*

$$\max_{U \in V} w(U, \bar{U}) \leq \frac{n}{4} \lambda_{n-1}.$$

Bizonyítás: Az előző állítás bizonyítását csak annyiban módosítjuk, hogy a (5)-t használjuk, valamint, hogy nem minimális, hanem maximális vágást határozzon meg U^* . Így kapjuk végül a kívánt egyenlőtlenséget.

$$\max_{C \in \mathcal{E}(G)} w(C) \leq \frac{|U^*|(n - |U^*|)}{n} \lambda_{n-1} \leq \frac{n}{4} \lambda_{n-1}$$

□

Fiedler [11] egy egyszerű gráfra a minimum vágás értékét élösszefüggőségnek hívta, mivel ennyi élt kell eltávolítani ahhoz, hogy a gráf ne legyen összefüggő. Az élösszefüggőségre az $e(G)$ jelölést vezette be. Mivel a λ_1 sajátérték és az élösszefüggés között szoros kétoldali kapcsolat van, a λ_1 sajátértéket elnevezte algebrai összefüggőségnek. Megmutatta, hogy nem teljes gráfra

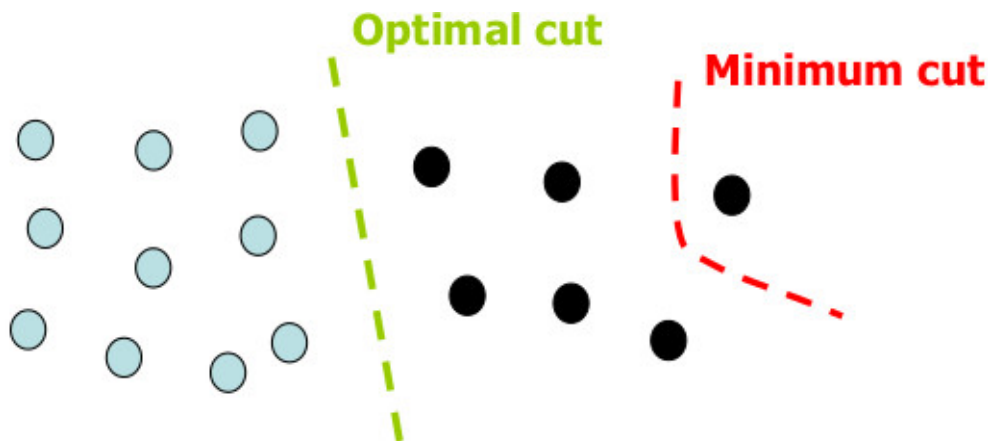
$$\lambda_1 \leq e(G).$$

A fenti bizonyítások alapján megsejthetjük, hogy hogyan keressünk optimális bipartíciót. A csúcsokat a λ_1 sajátvektorhoz tartozó sajátvektor szerint sorba rendezve megkeressük a sorrendben az optimális vágást. Ez közelíteni fogja a tényleges optimumot. Ezt a sajátvektort Fiedler-vektornak hívjuk.

4.7. definíció (Fiedler-vektor). *A $G = (V, W)$ gráf normalizált Laplace-mátrixának legkisebb pozitív sajátértékhez tartozó sajátvektort Fiedler-vektornak hívjuk.*

4.2. Izoperimetrikus feladatok

Kézenfekvő, hogy minimális vágással osszuk részekre a csúcshalmazt. Amint azonban az 1. ábrán is látható, könnyen előfordulhat, hogy egy minimális vágás csak egy csúcsot vág le a gráfról, és valójában nem ez a „legjobb” vágás.



1. ábra. Példa arra, amikor a minimális vágás nem optimális eredményhez vezet

Keresnünk kell valamilyen más mértéket, ami mentén optimalizálunk. Azokat a problémákat, ahol optimális vágásokat keresünk a halmazok méretének figyelembe vételével, izoperimetrikus problémáknak nevezik.

Célunk, hogy az egy partícióba tartozó csúcsok hasonlóak legyenek. Ennek eléréshez optimalizálhatunk arra, hogy az egy partícióba levő csúcsok ne legyenek túl messze egymástól. Másként megfogalmazva, hogy az egyes részek átmérője ne legyen túl nagy.

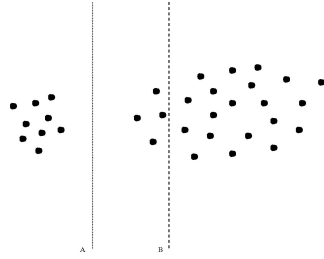
4.8. definíció (Átmérő). *Egy gráf átmérőjén a csúcsai között fellépő távolságok maximumát értjük.*

$$D(G) = \max_{u,v \in V(G)} \text{dist}(u,v)$$

Tehát van értelme vizsgálni a következő feladatot:

$$\min_{P_k \in \mathcal{P}_k} \max_{V_i \in P_k} D(V_i).$$

A 2. ábra jól demonstrálja azt az esetet, amikor az átmérővel nem a kívánt partíciókat kapjuk.



2. ábra. Példa arra, amikor a maximális átmérő nem ad optimális vágást [16].

Szeretnénk megtartani a minimum vágás azon tulajdonságát, hogy a komponensek közötti élek összszúlya kicsi. Így mértékek meghatározásához érdemes a minimum vágásból kiindulni, és normálni a csúcshalmazok valamilyen mérőszámával.

Legegyszerűbb a csúcshalmazok elemszámával normálni, így kapjuk a kiegyensúlyozottság mértékét.

4.9. definíció (Kiegyensúlyozottság).

$$\begin{aligned}
 b(P_k, G) &= \sum_{a=1}^{k-1} \sum_{b=a+1}^k \left(\frac{1}{|V_a|} + \frac{1}{|V_b|} \right) w(V_a, V_b) = \\
 &= \sum_{a=1}^k \frac{w(V_a, \bar{V}_a)}{|V_a|} = k - \sum_{a=1}^k \frac{w(V_a, V_a)}{|V_a|}.
 \end{aligned}$$

Normálhatunk az illeszkedő élek súlyával is. Később is hasznos lesz a következő fogalom.

4.10. definíció (Csúcshalmaz térfogata). A $G = (V, E)$ gráf $S \subset V$ csúcshalmazának térfogata:

$$Vol(S) = \sum_{v \in S} d_v.$$

Shi és Malik [25] az alábbi mértéket javasolja.

4.11. definíció (Normalizált k-vágás érték).

$$Ncut(P_k, G) = \sum_{a=1}^{k-1} \sum_{b=a+1}^k \left(\frac{1}{Vol(V_a)} + \frac{1}{Vol(V_b)} \right) w(V_a, V_b) =$$

$$= \sum_{a=1}^k \frac{w(V_a, \bar{V}_a)}{\text{Vol}(V_a)} = k - \sum_{a=1}^k \frac{w(V_a, \bar{V}_a)}{\text{Vol}(V_a)}$$

A kiterjesztés fogalma is gyakran használatos az optimális vágás meghatározásához. Azt fejezi ki, hogy mennyivel növekszik egy gráf, ha hozzávesszük az adott csúcshalmazt.

4.12. definíció (Kiterjesztés).

$$\Psi(S) = \frac{w(S, \bar{S})}{\min(|S|, |\bar{S}|)}.$$

A kiterjesztés hiányossága, hogy minden csúcsot egyformán súlyoz. Ezt kijavíthatjuk, ha nem elemszámot, hanem térfogatot veszünk alapul. Így a konduktancia fogalmához jutunk.

4.13. definíció (Konduktancia).

$$\Phi(S) = \frac{w(S, \bar{S})}{\min(\text{Vol}(S), \text{Vol}(\bar{S}))}.$$

4.3. Bipartíciók

Először azzal a kérdéssel foglalkozunk, amikor biparticionálni szeretnénk G -t.

4.3.1. A Cheeger-konstans

Tekintsük a következő feladatot!

1. feladat: Egy fix m -re keressük azt az S csúcshalmazt, amire $m \leq \text{Vol}(S) \leq \text{Vol}(\bar{S})$ és $w(\partial S)$ minimális.

A választ az ún. *Cheeger-konstans* adja meg.

4.14. definíció (Cheeger-konstans). Egy $G = (V, E)$ -hoz tartozó *Cheeger-konstans*

$$h_G = \min_{SCV} \Phi(S).$$

$h_G > 0$ csak akkor igaz, ha G összefüggő, ezért feltesszük, hogy G összefüggő.

Az 1. feladat megoldása egyenértékű a Cheeger-konstans meghatározásával, mivel [7] szerint

$$w(\partial S) \geq h_G \text{Vol}(S).$$

4.3.2. A Cheeger-egyenlőtlenség

A következőkben a Cheeger-konstans és a spektrum kapcsolatát fogjuk vizsgálni.

4.15. tétel (Cheeger-egyenlőtlenség). *Egy $G = (V, W)$ összefüggő élsúlyozott gráfhoz tartozó h_G Cheeger-konstansra és a spektrumának legkisebb pozitív sajátértékére igazak az alábbi egyenlőtlenségek.*

$$\frac{h_G^2}{2} < \lambda_1 \leq 2h_G$$

.

Bizonyítás (Felső becslés): Legyen $S \subset V$ Cheeger-optimális, vagyis $h_G(S) = h_G$. Tekintsük a következő $f : V \rightarrow \mathbb{R}$ függvényt:

$$f(v) = \begin{cases} \frac{1}{\text{Vol}(S)} & \text{ha } v \in S \\ \frac{1}{\text{Vol}(\bar{S})} & \text{ha } v \in \bar{S} \end{cases}$$

f -t behelyettesítve (3)-ba kapjuk a következőt:

$$\lambda_1 \leq \partial S \left(\frac{1}{\text{Vol}(S)} + \frac{1}{\text{Vol}(\bar{S})} \right) \leq \frac{2\partial S}{\min(\text{Vol}(S), \text{Vol}(\bar{S}))} = 2h_G.$$

□

A másik irányt az egyszerűség kedvéért most csak abban az esetben látjuk be, amikor nem súlyozottak az élek.

Bizonyítás (Alsó becslés): Nézzük a λ_1 -hez tartozó egység hosszú sajátvektort, f -t és $g = D^{-1/2}f$ vektorokat, mint $V \rightarrow \mathbb{R}$ függvényeket. Mivel f merőleges a \sqrt{d} -re, ezért igaz, hogy

$$\sum_v g(v)d_v = 0. \tag{8}$$

Feltehető, hogy a csúcsok f szerint csökkenő sorrendbe vannak rendezve vagyis $f(v_i) \geq f(v_j)$, ha $i < j$. Legyen $S_i = \{v_1, \dots, v_i\}$, valamint

$$\alpha_G = \min_i h_{S_i}.$$

Legyen r a legnagyobb egész, amire $Vol(S_r) < Vol(G)/2$. (8) miatt

$$\sum_v g(v)^2 d_v = \min_c \sum_v (g(v) - c)^2 d_v \leq \sum_v (g(v) - g(v_r))^2 d_v.$$

Jelölje g_+ és g_- g pozitív ill. negatív részét.

$$\begin{aligned} \lambda_1 &= \frac{\sum_{u \sim v} (g(u) - g(v))^2}{\sum_v g(v)^2 d_v} \geq \frac{\sum_{u \sim v} (g(u) - g(v))^2}{\sum_v (g(v) - g(v_r))^2 d_v} \\ &\geq \frac{\sum_{u \sim v} \left((g_+(u) - g_+(v))^2 + (g_-(u) - g_-(v))^2 \right)}{\sum_v (g_+(v)^2 + g_-(v)^2) d_v}. \end{aligned}$$

Mivel $\frac{a+b}{c+d} \geq \min(\frac{a}{c}, \frac{b}{d})$, feltehető, hogy

$$\begin{aligned} \lambda_1 &\geq \frac{\sum_{u \sim v} (g_+(u) - g_+(v))^2}{\sum_v g_+(v)^2 d_v} \\ &= \frac{\left(\sum_{u \sim v} (g_+(u) - g_+(v))^2 \right) \left(\sum_{u \sim v} (g_+(u) + g_+(v))^2 \right)}{\left(\sum_v g_+(v)^2 d_v \right) \left(\sum_{u \sim v} (g_+(u) + g_+(v))^2 \right)}. \end{aligned}$$

A Cauchy-Schwarz-Bunyakovszkij egyenlőtlenség miatt tovább

$$\begin{aligned} \lambda_1 &\geq \frac{\left(\sum_{u \sim v} (g_+(u)^2 - g_+(v)^2) \right)^2}{2 \left(\sum_v g_+^2(v) d_v \right)^2} \\ &= \frac{\left(\sum_i |g_+(v_i)^2 - g_+(v_{i+1})^2| |\partial(S_i)| \right)^2}{2 \left(\sum_v g_+(v)^2 d_v \right)^2}. \end{aligned}$$

Bevezetve a

$$\tilde{Vol}(S) = \min(Vol(S), Vol(G) - Vol(S))$$

jelölést és használva az α_g definícióját, azt kapjuk, hogy

$$\begin{aligned}
\lambda_1 &\geq \frac{\left(\sum_i |g_+(v_i)^2 - g_+(v_{i+1})^2| \alpha_G |\tilde{Vol}(S_i)|\right)^2}{\left(\sum_v g_+(v)^2 d_v\right)^2} \\
&= \frac{\alpha_g^2 \left(\sum_i g_+(v_i)^2 (|\tilde{Vol}(S_i) - \tilde{Vol}(S_{i+1})|)\right)^2}{2 \left(\sum_v g_+(v)^2 d_v\right)^2} \\
&= \frac{\alpha_g^2 \left(\sum_i g_+(v_i)^2 d_{v_i}\right)^2}{2 \left(\sum_v g_+(v)^2 d_v\right)^2} = \frac{\alpha_G^2}{2}.
\end{aligned}$$

Beláttuk tehát, hogy $\lambda_1 \geq h_G^2/2$. Ahhoz, hogy belássuk, hogy egyenlőség nem állhat fent, szükségünk lesz arra, hogy $1 - \sqrt{1 - x^2} > x^2/2$, ha $x > 0$, valamint az alábbi lemmára.

4.16. lemma. *Ha G összefüggő gráf és λ_1 a spektrumának legkisebb eleme, akkor*

$$\lambda_1 \geq 1 - \sqrt{1 - h_G^2}.$$

Bizonyítás: Láttuk, hogy

$$\lambda_1 \geq \frac{\sum_{u \sim v} (g_+(u) - g_+(v))^2}{\sum_v g_+(v)^2 d_v} = W.$$

Valamint azt is tudjuk már, hogy

$$\begin{aligned}
W &= \frac{\left(\sum_{u \sim v} (g_+(u) - g_+(v))^2\right) \left(\sum_{u \sim v} (g_+(u) + g_+(v))^2\right)}{\left(\sum_v g_+(v)^2 d_v\right) \left(\sum_{u \sim v} (g_+(u) + g_+(v))^2\right)} \\
&\geq \frac{\left(\sum_{u \sim v} |g_+(u)^2 - g_+(v)^2|\right)^2}{\left(\sum_v g_+(v) d_v\right) \left(2 \sum_v g_+(v)^2 d_v - W \sum_v g_+(v)^2 d_v\right)} \\
&\geq \frac{\left(\sum_i |g_+(v_i)^2 - g_+(v_{i+1})^2| |\partial S_i|\right)^2}{(2 - W) \left(\sum_v g_+(v)^2\right)^2 d_v} \\
&\geq \frac{\left(\sum_i (g_+(v_i)^2 - g_+(v_{i+1})^2) \alpha_G \sum_{j \leq i} d_j\right)^2}{(2 - W) \left(\sum_v g_+(v)^2\right)^2 d_v} \\
&\geq \frac{\alpha_G^2}{2 - W}.
\end{aligned}$$

Ebből következik, hogy

$$0 \geq W^2 - 2W + \alpha_G^2.$$

A másodfokú egyenlet megoldóképletét alkalmazva kapjuk a kívánt egyenlőtlenséget.

$$\lambda_1 \geq W \geq 1 - \sqrt{1 - \alpha_G^2} \geq 1 - \sqrt{1 - h_G^2}$$

□

A lemma bizonyításával beláttuk a Cheeger-egyenlőtlenség másik irányát is. □

A k -dimenziós kocka Cheeger-konstansa $2/n$, ugyanannyi, mint a spektrumának λ_1 eleme. Tehát a Cheeger-egyenlőtlenség felső becslése konstans szorzó erejéig pontos.

A n -hosszú út spektrumának első pozitív eleme $1 - \cos \frac{\pi}{n-1} \approx \pi^2/2(n-1)^2$, a Cheeger-konstansa pedig $1/\lfloor \frac{n-1}{2} \rfloor$. Megmutattuk tehát, hogy a Cheeger-egyenlőtlenség alsó iránya is a lehető legjobb becslés konstans szorzó erejéig.

4.4. Klaszterezés

A következőkben áttérünk a multipartíciók vizsgálatára. Gráfok több részre való partícionálásának az egyik kedvelt heurisztikus megoldása a spektrál módszer [9, 22, 27]. A másik kedvelt módszer az agglomeratív klaszterezés [1], amelyről azonban ismert, hogy instabil megoldást ad [14], ezért ezzel az iránnyal nem foglalkozunk.

4.4.1. Reprezentánsok keresése QP programmal

A klaszterezést megnehezíti adataink sokdimenziós volta. Ezért érdemes az egyes csúcsokat kisdimenziós vektorokkal reprezentálni. Ehhez tekintsük a következő kvadratikus programozási feladatot!

Adott $G = (V, W)$ gráfhoz és adott $1 \leq k \leq n$ egészhez keresünk olyan k -dimenziós r_1, \dots, r_n reprezentáns vektorokat, hogy a

$$Q_k = \sum_{i < j} w_{ij} \|r_i - r_j\| \geq 0 \tag{9}$$

célfüggvény érték minimális legyen a

$$\sum_{i=1}^n r_i r_i^T = I_k \quad (10)$$

feltétel mellett (I_k jelöli a k dimenziós egységmátrixot). Látható, hogy a célfüggvény értéke csökken, ha a nagy hasonlóságú csúcsok közelebb kerülnek egymáshoz.

Átalakítjuk a célfüggvényt és a feltételeket egy többet mondó formába. Legyen X az a mátrix, aminek a sorai r_1^T, \dots, r_n^T . Legyenek $x_1, \dots, x_k \in \mathbb{R}^n$ X oszlopai. A (10) feltételek miatt X oszlopai ortonormált rendszert alkotnak, tehát X egy szubortogonális mátrix. Így a feltételeket felírhatjuk $X^T X = I_k$ formában is. Ezek után a célfüggvényt tovább alakíthatjuk (6) egyenlet szerint.

$$\min \operatorname{tr}(X^T L X) \quad (11)$$

$$X^T X = I_k \quad (12)$$

4.17. tétel (Reprezentációs tétel súlyozott gráfokra). *Tekintsük a $G = (V, W)$ összefüggő gráfot és a hozzá tartozó L Laplace-mátrixot! Jelölje $0 = \lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{n-1}$ az L sajátértékeit, valamint u_0, u_1, \dots, u_{n-1} a hozzájuk tartozó egység-hosszú sajátvektorokat! Legyen $k < n$ olyan, hogy $\lambda_{k-1} < \lambda_k$! Ekkor a (9) célfüggvény értéke a (10) feltételek mellett:*

$$\sum_{i=0}^{k-1} \lambda_i = \sum_{i=1}^{k-1} \lambda_i.$$

Az optimális reprezentánsok r_1^, \dots, r_n^* az $X^* = (u_0, u_1, \dots, u_{k-1})$ mátrix sorainak transzponáltjai.*

Bizonyítás: A (11) alak szerint:

$$Q_k = \operatorname{tr}(X^T L X).$$

Így a tétel következik az alábbi általánosabb állításból.

4.18. állítás. *Legyen A egy valós $n \times n$ méretű szimmetrikus mátrix. A sajátértékei $\lambda_1 \geq \dots \geq \lambda_n$. $k > 0$ olyan, hogy $\lambda_k > \lambda_{k+1}$ Ekkor*

$$\max_{\substack{X \in \mathbb{R}^{n \times k} \\ X^T X = I_k}} \operatorname{tr}(X^T A X) = \max_{\substack{x_i \in \mathbb{R}^n (i=1, \dots, k) \\ x_i^T x_j = \delta_{ij}}} \sum_{i=1}^k x_i^T A x_i = \sum_{i=1}^k \lambda_i.$$

Az optimum az $X = (u_1, \dots, u_k)$ szubortogonális mátrix, ahol u_i az λ_i -hez tartozó egység hosszú sajátvektor [4].

□

A klaszterek meghatározásához alkalmazhatunk általános geometriai klaszterező algoritmust a jóval kisebb dimenziós reprezentáns vektorokon. Most bemutatjuk a legismertebb ilyen klaszterező eljárást, a k-means algoritmust.

k-means algoritmus

Adott l db \mathbb{R}^n -beli vektor, és i iterációszám.

1. Választunk k db tetszőleges klaszterközpontot, *számláló* := 0.
2. Minden csúcsot beosztunk abba a klaszterbe, amelyiknek a középpontja legközelebb esik hozzá.
3. Minden klaszter új középpontja legyen a benne levő pontok átlaga!
4. Ha *számláló* = i , akkor megállunk, egyébként visszalépünk 2. pontra és *számláló*-t növeljük 1-gyel.

4.4.2. Spektráلكlaszterezés

Már láttuk a 4.1 fejezetben, hogy a Fiedler-vektor segítségével hogyan határozhatunk meg vágásokat. Klaszterezés esetén alkalmazhatjuk szubrutinként a vágási eljárást.

Hagen algoritmus [13]

1. Vegyük a λ_1 -hez tartozó sajátvektort
2. A Fiedler-vektor szerint sorba rendezzük a csúcsokat
3. Megkeressük a minimális konduktanciájú vágást a sorrendben
4. A két részre külön-külön alkalmazzuk az algoritmust

Kannan, Vempala és Vetta [16] becslést is adtak az előző algoritmus által talált klaszterek minőségére. A becsléshez az alábbi mérték használták.

4.19. definíció ((α, ϵ)-partíció). Egy $\{V_1, \dots, V_k\}$ partíció (α, ϵ)-partíció, ha

- C_i konduktanciája legalább α
- $i \neq j$, C_i és C_j között menő élek súlya az összes él súlyának legfeljebb ϵ része.

4.20. állítás (Kannan-Vempala-Vetta). Ha a G gráfnak létezik (α, ϵ)-partíciója, akkor Hagen algoritmusával talál

$$\left(\frac{\alpha^2}{72 \log^2 \frac{n}{\epsilon}}, 20\sqrt{\epsilon} \log \frac{n}{\epsilon} \right) - \text{partíciót.}$$

4.5. Spektrálklaszterezés több sajátvektorral

Az előző részben látott algoritmus hátránya, hogy nem stabil és nem hatékony. Shi és Malik [25] egy másik megközelítést javasol a spektrálklaszterezésre. Nem csak a Fiedler-vektort, hanem több sajátvektort is érdemes figyelembe venni. Az általuk használt algoritmus a következő:

Shi és Malik algoritmus

1. Keressük meg a normalizált Laplace-mátrix első k pozitív sajátértékéhez tartozó sajátvektort.
2. A vektorokból képzett mátrix sorait klaszterezzük pl. $k - \text{means}$ algoritmussal.
3. A v csúcs a v -nek megfelelő sor klaszterébe kerül.

Shi és Malik megmutatta, hogy minél több sajátértéket veszünk, annál jobban közelítjük az optimális normalizált k -vágást. Brand és Huang [5] megmutatta, hogy több sajátvektort véve a klaszterek eloszlása egyenletesebb lesz, a hasonló pontok távolsága csökken, a különbözőké pedig növekszik.

5. Spektrálkласzterezés a gyakorlatban

A következő fejezetben a spektrálkласzterezés gyakorlatban való alkalmazhatóságát vizsgáljuk meg. A gyakorlatban azt tapasztaljuk, hogy a nagy gráfokon a spektrumon alapuló algoritmusok gyengébb eredményeket érnek el, mint más megközelítést használó módszerek. Az alábbiakban ennek okait, valamint néhány lehetséges megoldást mutatunk be egy konkrét példán keresztül.

5.1. A spektrálkласzterezés egy gyengesége

5.1.1. A TKC-hatás

Lempel és Moran [20] az internetgráf vizsgálatok során derítették fényt a HITS-algoritmus egy sebezhetőségére. Azt vették észre, hogy a HITS-algoritmus gyakran felértékel a valóságban nem mértékadó oldalakat. A hiba okaként a kis sűrű részgráfokat, csomópontokat azonosították. Ezért a jelenséget TKC-hatásnak (*tightly knit communities effect*) nevezték el.

A TKC jelenség elméleti alapját az adja, hogy a mátrix spektruma a 2.3. fejezet végén látott út-leszámlálás mennyiséggel függ össze. A következő állítások szerint a sűrű részgráfokban exponenciálisan gyorsan nő az utak száma, így a sajátvektorokat „magukhoz vonzzák”.

5.1. lemma. *Legyen $S = U\Lambda U^T$ szimmetrikus pozitív szemidefinit mátrix, ahol*

$$\Lambda = \begin{pmatrix} \lambda_0 & \dots & 0 \\ 0 & \ddots & 0 \\ 0 & \dots & \lambda_{n-1} \end{pmatrix}.$$

Ha az x kiinduló vektor merőleges u_1, \dots, u_{i-1} -re, de nem merőleges u_i -re, és $\lambda - i + 1 < \lambda_i$, akkor

$$\lim_{k \rightarrow \infty} \frac{S^k x}{\lambda_i^k} = u_i. \quad (13)$$

Bizonyítás: A HITS-nél látott (2) egyenlethez hasonlóan $S^k u_i = \lambda_i^k u_i$. Az állítás következik, ha felírjuk x -t az u_i -k, mint bázisok által meghatározott térben. \square

5.2. lemma. *Legyen S egy súlyozott gráf élmátrixa. Ekkor $S^k x$ az x -el súlyozott kezdőcsúcsokból induló k hosszú séták összűlya.*

A fenti két lemmát alkalmazhatjuk a HITS algoritmusnál látott $S = AA^T$ mátrixokra, a Laplace, illetve a súlyozott Laplace mátrixokra is. A 5.2. Lemma szerint azokra a j csúcsokra lesz magas az $[S^k x]_j$ értéke, amelyekből nagyon sok súlyozott út indul. A 5.1. Lemma szerint ezeknek a j csúcsoknak lesz kiemelkedő értéke az első olyan sajátvektorban, amely x -re nem merőleges.

5.2. Lehetséges módszerek a TKC-hatás semlegesítésére

Ebben a részben két heurisztikát szeretnénk ismertetni, melyek segítségével javítható a spektrálklaszterezés eredménye.

5.2.1. Csomópontok eltávolítása

A TKC-hatást okozó csomópontok megtalálására alkalmazhatjuk a SCAN algoritmust [26]. Ha megtaláltuk a gráf méretéhez képest elhanyagolható csomópontokat, azokat elhagyhatjuk.

5.2.2. Csápok összehúzása

Csápoknak nevezzük egy előre rögzített d_t fokszámnál kisebb fokú csúcsokból álló részgráfot. Ezek eltávolítására a következő eljárást alkalmazhatjuk.

1. Keresünk egy v csúcsot, amire $d_v < d_t$, ha nincs ilyen, akkor megállunk.
2. v -t összehúzzuk a legnagyobb fokszámú szomszédjával.
3. Felírjuk, hogy melyik két csúcsot húztuk össze.

4. Újra kezdjük az 1. lépéstől.

Az eljárás véges, mert minden lépésben csökken a csúcsszám.

5.3. Spektráلكlaszterezés egy konkrét példán bemutatva

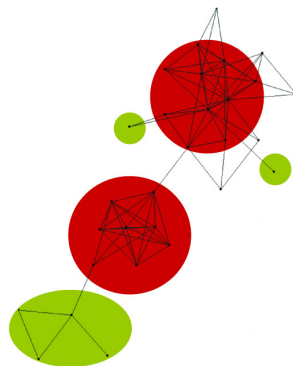
Most pedig egy valós példán demonstráljuk, hogy hogyan alkalmazható a spektráلكlaszterezés a fenti heurisztikák használatával.

5.3.1. Az adat: livejournal.com

Példánkban [18] cél a LiveJournal (www.livejournal.com) oldal felhasználóinak klaszterezése. A LiveJournal egy közösségi blog oldal, vagyis a felhasználók nem csak tartalmakat közölhetnek, hanem lehetőségük van egymással való kapcsolatukat (barátságukat) is kifejezni. A kapcsolat szimmetrikus, így a szociális háló élei irányítatlanok.

Célkitűzés volt, hogy az orosz nyelvű felhasználókat geológia elhelyezkedésük szerint szeparálják. Az orosz nyelvű felhasználók jelentős része valójában valamelyik környező ország lakosa (Ukrajna, Fehéroroszország, Észtország, ...).

A 3. ábrán a gráf egy részlete látható, amely jól szemlélteti, hogy a gráf struktúrájában hogy helyezkednek el a csápok(zöld részek) és csomók(piros részek).



3. ábra. A livejournal.com barátsággráfjának egy részlete

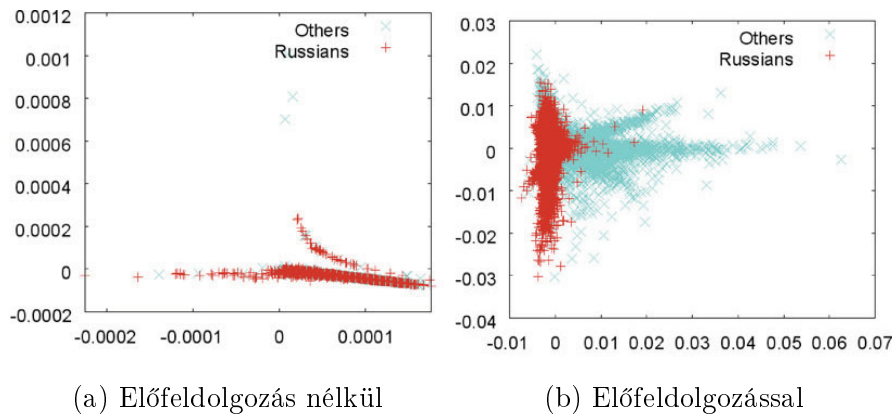
A gráfban 2.379.267 csúcs és 14.286.827 él szerepelt. A felhasználók több mint 80% amerikai és körülbelül 5% orosz.

5.3.2. A kísérlet menete

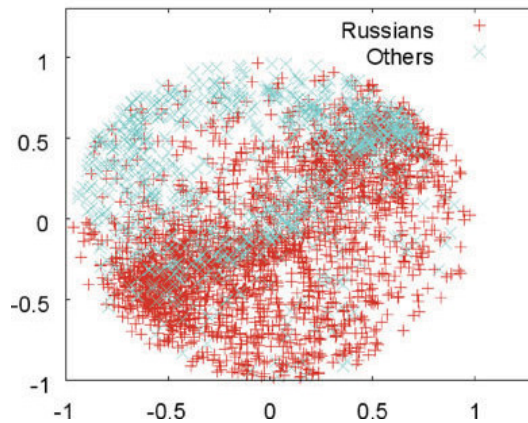
A gráfból először eltávolították a csomópontokat, majd alkalmazták a csápok összehúzását. Innen kétféleképp haladtak tovább. Egyrészt Shi és Malik algoritmusát alkalmazták. Másrészt Lang [19] algoritmusát is megvalósították, aki a 4.17. Tételben ismertetett reprezentáció meghatározására egy szemidefinit relaxált programot old meg. Végül a kivett csomókat viszarakva, és a csápokat visszánövesztve kapták a végső klasztereket. A spektráلكlaszterezést lefuttatták előfeldolgozás nélkül is, az összehasonlításhoz.

5.3.3. Eredmények

Az eredményeket a 4-5. ábrák mutatják be.



4. ábra. A csúcsok eloszlása 4. és 5. sajátirányok mentén



5. ábra. A szemidefinit program kimenete

Látható, hogy az előfeldolgozás nélkül a két csoportot nem lehet megkülönböztetni. A heurisztikák alkalmazása után a klaszterek azonban eltávolodnak egymástól.

Az egyes klaszterek minőségét az entrópiával és tisztasággal mérték.

5.3. definíció (Entrópia). Egy X halmaz X_1, \dots, X_j felbontásának entrópiája

$$h(X) = - \sum_i p_i \log p_i,$$

ahol $p_i = |X_i|/|X|$.

5.4. definíció (Tisztaság). Egy X halmaz X_1, \dots, X_j felbontásának tisztasága

$$t(X) = \max_i p_i,$$

ahol $p_i = |X_i|/|X|$.

	Futási idő(perc)	Entrópia	Tisztaság
SVD előfeldolgozás nélkül	1980	0,105	0,812
SVD előfeldolgozással	150	0,073	0,853
SDP előfeldolgozás nélkül	1755	0,111	0,857
SDP előfeldolgozással	675	0,072	0,854

1. táblázat. A spektrál heurisztikák minősége és futási ideje a LiveJournal orosz részgráfján.

Amint az 1. táblázatban látható, szemidefinit programmal összehasonlítható lett az eredmény, míg a spektrálklasszterezés sokkal gyorsabban futott.

6. Összefoglalás

Jelen dolgozatban betekintést nyújtottunk a spektrálgráfelmélet alapvető eredményeibe, bár korántsem érintettünk minden kapcsolódó témakört. A vágásokon felül nem beszéltünk a további sajátértékeket érintő becslésekről, más gráfparaméterekkel való összefüggésekről. A jelenleg ismertebb véletlen gráf modellek, mint a Barabási-Albert-modell, rosszul modellezik a valóságban létező nagy gráfokat a spektrálklasszterezés nehézségeinek tekintetében.

A fentiek tükrében további kutatási területként merül fel olyan egyszerű véletlengráfmodell felállítása, amely jól modellezi a fenti problémát. Későbbi kutatásaink során szeretnénk olyan előfeldolgozási lépéseket találni, amelyekre bizonyítható, hogy a kívánt módon módosítják a spektrumot.

Hivatkozások

- [1] Charles J. Alpert – Andrew B. Kahng: Recent directions in netlist partitioning: a survey. *Integr. VLSI J.*, 19. évf. (1995) 1-2. sz., 1–81. p. ISSN 0167-9260.
- [2] Charles J. Alpert – So-Zen Yao: Spectral partitioning: the more eigenvectors, the better. In *DAC '95: Proceedings of the 32nd ACM/IEEE conference on Design automation* (konferenciaanyag). New York, NY, USA, 1995, ACM Press, 195–200. p. ISBN 0-89791-725-1.
- [3] Eugenio Beltrami: Sulle funzioni bilineari. *Giornale di Matematiche ad Uso degli Studenti Delle Università*, 11. évf. (1873), 98–106. p.
- [4] Marianna Bolla: *Spectral Clustering and Biclustering: Learning Large Graphs and Contingency Tables*. 2013, John Wiley & Sons.

- [5] Matthew Brand–Kun Huang: A unifying theorem for spectral embedding and clustering. In *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics* (konferenciaanyag). 2003.
- [6] Pak K. Chan–Martine D. F. Schlag–Jason Y. Zien: Spectral k-way ratio-cut partitioning and clustering. In *DAC '93: Proceedings of the 30th international conference on Design automation* (konferenciaanyag). New York, NY, USA, 1993, ACM Press, 749–754. p. ISBN 0-89791-577-1.
- [7] Fan RK Chung: *Spectral graph theory*. 92. köt. 1997, American Mathematical Soc.
- [8] Dragos M Cvetkovic–Michael Doob–Horst Sachs: *Spectra of graphs: Theory and application*. 413. köt. 1980, Academic press New York.
- [9] Chris H. Q. Ding–Xiaofeng He–Hongyuan Zha–Ming Gu–Horst D. Simon: A min-max cut algorithm for graph partitioning and data clustering. In *ICDM '01: Proceedings of the 2001 IEEE International Conference on Data Mining* (konferenciaanyag). Washington, DC, USA, 2001, IEEE Computer Society, 107–114. p. ISBN 0-7695-1119-8.
- [10] W. E. Donath–A. J. Hoffman: Lower bounds for the partitioning of graphs. *IBM Journal of Research and Development*, 17. évf. (1973. szeptember) 5. sz., 420–425. p.
- [11] Miroslav Fiedler: Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 23. évf. (1973) 2. sz., 298–305. p.
- [12] LR Ford–Delbert Ray Fulkerson: *Flows in networks*. 3. köt. 1962, Princeton Princeton University Press.
- [13] Lars Hagen–Andrew B Kahng: *Fast spectral methods for ratio cut partitioning and clustering*. 1991, Computer Science Department, University of California.
- [14] John Hopcroft–Omar Khan–Brian Kulis–Bart Selman: Natural communities in large linked networks. In *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery*

- and data mining* (konferenciaanyag). New York, NY, USA, 2003, ACM, 541–546. p. ISBN 1-58113-737-0.
- [15] Camille Jordan: Mémoire sur les formes bilinéaires. *Journal de mathématiques pures et appliquées*, 1874., 35–54. p.
- [16] Ravi Kannan – Santosh Vempala – Adrian Vetta: On clusterings: Good, bad and spectral. *Journal of the ACM (JACM)*, 51. évf. (2004) 3. sz., 497–515. p.
- [17] Jon Kleinberg: Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46. évf. (1999) 5. sz., 604–632. p.
URL citeseer.nj.nec.com/article/kleinberg98authoritative.html.
- [18] Miklós Kurucz – András A Benczúr: Geographically organized small communities and the hardness of clustering social networks. In *Data Mining for Social Network Data*. 2010, Springer, 177–199. p.
- [19] Kevin J Lang: Fixing two weaknesses of the spectral method. *balance*, 2. évf. (2005) 1. sz., 2. p.
- [20] Ronny Lempel – Shlomo Moran: The stochastic approach for link-structure analysis (salsa) and the tkc effect. *Computer Networks*, 33. évf. (2000) 1. sz., 387–401. p.
- [21] László Lovász: *Combinatorial problems and exercises*. 361. köt. 2007, American Mathematical Soc.
- [22] M. Meila – J. Shi: A random walks view of spectral segmentation. In *AISTATS* (konferenciaanyag). 2001.
URL <http://citeseer.ist.psu.edu/364366.html>.
- [23] M. Newman: Detecting community structure in networks. *The European Physical Journal B - Condensed Matter*, 38. évf. (2004. March) 2. sz., 321–330. p. ISSN 1434-6028.
URL <http://dx.doi.org/10.1140/epjb/e2004-00124-y>.
- [24] Pál Rózsa: *Lineáris algebra és alkalmazásai*. 1991, Tankönyvkiadó.

- [25] Jianbo Shi–Jitendra Malik: Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22. évf. (2000) 8. sz., 888–905. p.
- [26] Xiaowei Xu–Nurcan Yuruk–Zhidan Feng–Thomas AJ Schweiger: Scan: a structural clustering algorithm for networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining* (konferenciaanyag). 2007, ACM, 824–833. p.
- [27] Hongyuan Zha–Xiaofeng He–Chris H. Q. Ding–Ming Gu–Horst D. Simon: Spectral relaxation for k-means clustering. In Thomas G. Dietterich–Suzanna Becker–Zoubin Ghahramani (szerk.): *NIPS* (konferenciaanyag). 2001, MIT Press, 1057–1064. p.