ČESKÉ VYSOKÉ
UČENÍ TECHNICKÉ
V PRAZE

Fakulta jaderná
a fyzikálně
inženýrská

Trojanova 13, 120 00 Praha 2

Posudek vedoucího bakalářské práce

Autor práce:	Daniel Khol
Název práce:	Laplaceova matice a její aplikace
Vedoucí práce:	doc. Jan Vybíral

Téma předkládané práce:
Předkládaná práce se zabývá Laplaceovou maticí grafu, jejími vlastnostmi a jejím použitím v tzv. spectral clusteringu grafu. Práce je členěna do čtyř kapitol. První kapitola definuje základní pojmy z teorie grafů, zavádí Laplaceovu matici neorientovaného grafu, dokazuje některé její vlastnosti, podává několik příkladů grafů (úplný graf, kružnice, cesta, hvězda, strom) a jejich Laplaceových matic. Tato kapitola pak končí definicí algebraické konektivity a Fiedlerova vektoru. Druhá kapitola studuje možnost rozšířit pojem Laplaceovy matice na orientované grafy. Třetí a čtvrtá kapitola se zabývají souvislostí mezi Laplaceovou maticí a náhodnými procházkami na grafu či metodou tzv. spectral clusteringu.

Obsahová úroveň práce:
Úroveň předkládané práce je velmi rozporuplná. Mezi její pozitiva jistě patří fakt, že zadání práce bylo vesměs splněno a vytyčené úkoly jsou více či méně úspěšně pokryty. Taktéž je třeba ocenit snahu studenta o vlastní vědecký přínos pokusem definovat novým způsobem Laplaceovu matici pro orientované grafy. I celkový rozsah práce je na bakalářskou práci překvapivý.

Na druhou stranu nelze nepoznamenat, že práce vykazuje i značné nedostatky, mezi které patří nejen řada překlepů, nepřesností v matematickém vyjadřování a gramatické chyby, ale i řada koncepčních nejasností.

Zásadní komentáře:

- Řada formulací je jen obtížně srozumitelná. Např. na str. 12: “Jinými slovy, graf můžeme rozdělit jen v případě, že nám nezbude nějaká hrana bez vrcholu.”, “Hrany poté fungují jako váhy.”, apod.
- Věta 1.2.1 na str. 15: Autor zjevně váhá, zda jde o větu nebo definici. Pokud je objekt $(Lg)_i$ již definován, je komentář za větou (“$(Lg)_i$ nazveme Diskrétním Laplaceovým operátorem” a “jedná se o ekvivalentní definici”) matoucí. Použití symbolu $\sum\limits_{i\sim j}^{k}$ navíc slouží jako ukázkový

příklad zmatečného použití matematické notace v této práci. Zatímco v (1.1) označuje tato suma součet přes všechna j sousedící s daným pevným i, na konci str. 15 je stejný symbol použit pro součet přes všechny sousedící dvojice i, j. Stejně tak na str. 16 v definici 1.2.4 není zřejmé, zda se v $\sum_{i \neq j}$ jedná o součet přes i, přes j, nebo přes dvojice i, j.

- Str. 17: Definice 1.3.2 a definice 1.4.2 zavádí dva charakteristické polynomy. Jsou tyto pojmy v práci někde později použity?

- Na řadě míst v textu chybí interpunkce mezi větami, zejména končí-li věta matematickou formulí (viz. např. definici 1.4.3). To značně znesnadňuje čtení celého textu.

- Dalším příkladem obtížné čitelnosti textu je věta "Jako první si uvědomíme, že naše definice není úplně jednoznačná, z důvodu, že $m < n$ není dobře definována." na str. 18 dole.

- Jakým způsobem plyne rovnice $Lg = H(g)g$ z (1.3) a proč je tato rovnice rovnicí pro vlastní čísla Laplaceovy matice L?

- Str. 23 dole: Opravdu by autor ověřoval souvislost grafu výpočtem spektra Laplaceovy matice?

- Kapitoly 1.5 a 4.3 jsou typograficky velmi nešťastně řešeny, což vede k velkým prázdným místům.

- Definice 1.5.6 zavádí pojem Laplaceovy energie. Zavedení tohoto pojmu není motivováno a navíc se později vyskytuje jen ve velmi neformální diskuzi u věty 1.5.6, což činí celou diskuzi spektrálních vlastností Laplaceovy matice stromu poněkud nepřehlednou.

- Pojmy "algebraická konektivita" a "Fiedlerův vektor" jsou v Definici 1.6 sice zavedeny, ale bez jakéhokoliv kontextu. Formulace "algebraická konektivita funguje" je nejasná.

- Věta 1.6.2 zjevně neplatí. Množina $\{x : \sum_i x_i = 0 \;\&\; x \neq 0\}$ obsahuje prvky libovolně blízké nule, takže pravá strana identity pro λ_2 je rovna nule (nahradíme-li minimum infimem). "Důkaz" této věty je navíc jen obtížně čitelný, význam symbolu $\displaystyle\sum_{(i,j)=1}$ není dobře definován. Stejně tak není jasné, odkud plyne podmínka $x^T x = 1$, nebo co je to "kolmé X".

- Celková struktura kapitoly 2 je nedotažená. Pokus o definici Laplaceovy matice pro orientované grafy je veden snahou, aby "měla podobné vlastnosti jako klasická Laplaceova matice na neorientovaných grafech". Nicméně dosažené vlastnosti (viz. věta 2.3.1) platí vesměs pro jakoukoliv pozitivně semidefinitní matici, takže není jasné, proč bychom měli definovat L^a právě pomocí (2.2).

- Str. 37: Ve formuli (2.1) mělo být zřejmě uvedeno MM^T?

- Str. 39: Důkaz věty 2.3.2 plyne triviálně z faktu $(1, \ldots, 1)M = 0$.

- Obecně lze říci, že autor nešetřil definicemi a zavedl mnoho sobě vzájemně velmi podobných pojmů (Laplaceova matice pro neorientované grafy, Laplaceova matice pro orientované grafy, abstraktní Laplaceova matice, normalizovaná Laplaceova matice, Laplaceova matice náhodné procházky), což velmi znesnadňuje čtení celého textu. Jako důsledek je pak řada vyplývajících vět či tvrzení prakticky triviálních.

- Str. 47: Důkaz věty 3.2.5 je obtížně čitelný. Pomineme-li nešťastný obrat “Pro druhou hranici”, tak není zřejmý “přepis do sumačního zápisu”, znaménko “-” v následující formuli ani význam symbolu $\deg_i$.
- Str. 48: Kapitola 3.3 je celkově velmi špatně pojata a pro čtenáře téměř nesrozumitelná. Na ca. jedné straně textu autor nashromáždil čtyři definice a jednu větu. Pojem “distribuce na V” v definici 3.3.1 je použit nepřesně, jedná se spíš o míry na V. Význam symbolu $\max_{v \in V}$ v definici $d(n, p)$ je nejasný. V definici 3.3.3 patrně označují symboly $d(n)$ a $d(n, p)$ jednu a tutéž věc.
- Souvislost mezi Markovskými řetězci a Laplaceovou maticí je odbyta, věta 3.3.1 je uvedena nejen bez důkazu, ale i bez jakékoliv motivace, komentáře či aplikace. Autor se tedy vcelku správně ptá “Nabízí se tedy otázka, proč to vůbec zahrnujeme do práce o Laplaceově matici.”
- Str. 51: Definici pojmu cluster lze v literatuře úspěšně nalézt na mnoha místech, například v R. F. Ling: A Probability Theory of Cluster Analysis.

Drobnější komentáře:

- Str. 11: “připojené” vrcholy se obvykle nazývají sousedící.
- V celém textu: Autorovi se mnohdy nepodařilo správně vyjádřit výčet prvků a místo logického $v_1, \dots, v_j$ nalezneme nejrůznější kombinace typu $v_1, ... v_j$ (Str. 11),
- Průběžně v celé práci používá autor nevhodně slovo “první” místo “nejprve” (např. na str. 11: “Na to první potřebujeme znát pojem kružnice.” nebo na str. 17: “V této kapitole první zadefinujeme ...”)
- Str. 11: Hranám nepřiřazujeme “různé pravděpodobnosti” ale různé váhy.
- Str. 16: “To je z definice slabě diagonální matice.” není věta.
- Str. 17: “To by se nám pro spektrální analýzu velmi hodilo.”: Proč?
- Str. 18: Proč je někdy uváděn index matice v závorce a jindy bez ní? Tedy čím se liší $M_{n,j}$ a $M_{i,(j)}$?
- Str. 18: Je “náš graf” grafem z obr. 1.1 nebo obr. 1.2?
- Str. 19, důkaz věty 1.4.2: “Začneme prvky, když $i = j$.” je gramaticky velmi kostrbaté. Je-li n počet hran (který byl dříve označován spíše l), pak by sumy v (1.6) a (1.7) měly jít až do $2n$. Význam symbolu $\sum_{e \in V} M^2_{e,v_i}$ je nejasný.
- Str. 20: Tvrzení “Pro pozitivně semi-definitní matice totiž máme efektivní numerické algoritmy ...” by bylo dobré podpořit referencí.
- Str. 20: důsledku, nikoli následku.
- Str. 20, důkaz věty 1.4.4: Vlastní vektory jsou nejprve označeny $X_i = (x_{i,1}, \dots, x_{i,k})$ ale posléze přeznačeny na $x = (x_1, \dots, x_k)$.
- Str. 21: Co je x v (1.10)?

- Str. 25: V důkazu věty 1.5.1 se objevuje rovnice $\sum_j x_j = 0$, aniž by byly x_j definovány.
- Str. 26: Důkaz věty 1.5.2: Matice $L(C_k)$ je cirkulantní matice, její spektrum je tedy možné snadno najít pomocí Fourierovy transforamce, což by byl asi přehlednější způsob odvození, než jej "odhadovat". Důkaz se nezmiňuje o algeraické násobnosti uváděné ve větě. Formule na začátku str. 27 platí pro $z_m(j)$. Závěr důkazu by měl být "$m \in \{1, \dots, k\}$."
- Str. 28: V definici cesty by mělo být $i = 2, \dots, k$.
- Str. 28, důkaz věty 1.5.3: Není zřejmé, jak souvisí odebrání jedné hrany s důkazem věty. Jediná uvedená formule navíc opět platí pro $z_m(j)$, ne $z_k(j)$.
- Str. 31: Formule po "…pak máme stupeň vrcholu" se netýká stupně vrcholu.
- Str. 31: Strom je graf s danými vlastnostmi. Sousloví "graf stromu" tedy nedává příliš smysl.
- Str. 37: Má-li matice M příslušející ke grafu na obr. 2.1 čtenáři osvětlit definici 2.2.1, pak je nutné uvést i očíslování hran.
- Str. 38: Pojem symetrizace grafu ve větě 2.2.1 není jasně zaveden.
- Str. 43: Pojem "matice přechodu" je použit poněkud nejasně. Na jednu stranu je zaveden hned v úvodu kapitoly 3, ale poté je opět matice přechodu definována v definici 3.1.1.
- Str. 46: V důkazu věty 3.2.2 není zřejmý vztah mezi druhou a třetí uvedenou formulí.
- Str. 46: Rovnice ve větě 3.2.3 se zřejmě týká L^n, ne L.
- Str. 47: Věta 3.2.4 plyne triviálně už z definice 3.1.2:
  $x^T L^n x = x^T D^{-1/2} L D^{-1/2} x = (D^{-1/2}x)^T L (D^{-1/2}x) \geq 0$.
- Str. 48: Pojem "aperidockého grafu" na úvod kapitoly 3.3 není zaveden.
- Str. 49: článek [4] je ve skutečnosti kniha.
- Str. 50: Definice 3.4.1 nedává z gramatického hlediska smysl a konstrukt "Nechť …. Pak …" odpovídá spíše větě než definici.
- Str. 50: Důkaz věty 3.4.1 je spíše souborem neokomentovaných rovnic než důkazem.
- Str. 51: "abstratní"
- Str. 53: Nahrazení nuly aritmetickým průměrem hodnot Fiedlerova vektoru je tautologií, tento průměr je opět nula.
- Str. 53: "Úmyslné ponechání nulových sloupců" v incidenční matici je ve sporu s její definicí.
- Kapitola 4.3: Matice L_1^s je zbytečně uváděna na str. 55 a poté opět na str. 56, na str. 58 chybí uvést získaný clustering, "logický způsob" na str. 61 je příliš subjektivní měřítko úspěchu. Na str. 62 a 63 by bylo možno snadno ušetřit spoustu místa, pokud by vektory X_2 byly uváděny jako řádkové, tedy např. $X = [0.3383, \dots]^T$. Překlepy "$L_2 ='$" na str. 63 či číslice "3" za rovnící pro X_2 na str. 65 jen dokládají zoufalou nedoladěnost této kapitoly. Celkově v této kapitole chybí několik desítek (!) interpunkčních znamének.

Vlastní práce studenta:
I přes četné výtky považuji práci za obhajitelnou a to zejména díky snaze o vlastní matematický přínos v sekci o Laplaceově matici pro orientované grafy a za rozsah pojednané látky.

Závěr: Bakalářskou práci doporučuji k obhajobě a navrhuji klasifikovat známkou

dostatečně (E).

V Praze, 25. 7. 2022

doc. Jan Vybíral