

Posudek bakalářské práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Daniel Trlifaj
Název práce Graphlets in Complex Networks
Rok odevzdání 2023
Studijní program Informatika
Specializace Obecná informatika

Autor posudku David Hartman Vedoucí
Pracoviště Informatický ústav Univerzity Karlovy

K celé práci

lepší OK horší nevyhovuje

	lepší	OK	horší	nevyhovuje
Obtížnost zadání	X			
Splnění zadání	X			
Rozsah práce <i>... textová i implementační část, zohlednění náročnosti</i>	X			

Graflety jsou nástrojem, který získává na popularitě v oblasti využití komplexních sítí v datové analýze. Hlavním úskalím tohoto přístupu je absence hlubšího teoretického základu, ale i rigoróznějšího testování na množině náhodných sítí. Téma předkládané práce se zabývá oběma pohledy na danou oblast. Studuje tedy jednak teoretický pohled na tuto problematiku a možné souvislosti se známými výsledky z teorie grafů, dále pak realizuje numerické testy na náhodných sítích ukazující chování této charakteristiky na modelových příkladech.

Na počátku bych rád upozornil, že práce vznikala s cílem provést zmiňovanou numerickou analýzu vlivu graflletů a teoretická část byla později vytvořena při studiu vlastností těchto objektů. Samotná numerická analýza zahrnovala potřebu tvorby softwarového nástroje, který není součástí odevzdání. Náročnější ovšem bylo jeho následné ladění a parametrizace během analýzy potřeb datových zdrojů. Z toho plyne velká náročnost již původní numerické části a zahrnutí části teoretické lze tedy brát jako velké plus.

Celkově se dá říci, že na práci bylo vykonáno velké množství úsilí a rozhodně představuje zajímavý počín. Student na ní velmi aktivně pracoval a také dobře interagoval při jejím sestavování. Zvláště numerická část je velmi zajímavá. Zde je třeba upozornit, že získání výsledků i spolu s nastavením vhodné míry porovnání bylo opravdu náročné. Z tohoto důvodu se také ke konci mírně projevil nedostatek času způsobený původním soustředěním na numerickou část, která se z implementačních a analytických důvodů protáhla. Z tohoto důvodu některé finální úpravy práce nejsou ideální, viz níže. To ovšem nesnižuje přínos práce, který je, nejen na úrovni bakalářské, značný.

Student vytvořil práci, která si dala za nelehký úkol kombinovat oba pohledy na grafllety - jak teoretický, tak výpočetní. V tom teoretickém ustanovil (jinak nesourodou) terminologii graflletů a graflletové distribuce zobecňující a parametrizující pojem distribuce stupňů vrcholů. Následně pro tyto pojmy ukázal několik nových výsledků hlavně ve směru, který lze charakterizovat jako reprodukovatelnost, tj. kdy a pro jaké hodnoty parametrů lze z dané graflletové distribuce určit co nejpřesnější informace o grafu (ideálně graf samotný). Kromě výsledků kolem propojení různých graflletových distribucí autor ukazuje také souvislost k dimenzionálního Weisfeiler-Lehmanova testu a modifikované graflletové distribuce do velikosti k využitím myšlenek z výzkumu grafové neuronových sítí. Nakonec se autor zabývá přímo tématem rekonstrukce grafu z graflletové distribuce, která vede ke zformování prvních myšlenek pro konstrukci základní verze zobecněného konfiguračního modelu, na kterou by bylo možné navázat úspěšnou konstrukcí v budoucnu. Z výše zmíněného důvodu nedostatku času nejsou bohužel v této části všechny definice, tvrzení a důkazy dotaženy do rigorózní čistoty, nicméně myšlenky v nich jsou korektní a představují velmi pěkný přínos do oblasti.

V numerické části autor velmi důsledně probírá samotné nastavení experimentů a navrhuje různé metriky porovnání výsledných sítí a jejich graflletových distribucí, což je nelehká úloha s velkou potřebou zvolit vhodné kombinace parametrů. Nakonec provádí a prezentuje výsledky, které komentuje, vyvozuje výsledky a navrhuje možná vylepšení v budoucí práci. Tuto část lze opět jednoznačně označit za přínos do dané oblasti hlavně z hlediska promyšlenosti jednotlivých porovnání. Navíc si je třeba uvědomit, že grafllety nejsou z hlediska výpočetního jednoduchou charakteristikou, a tedy opakování pokusů při hledání vhodných parametrů není jednoduché.

Byť softwarové řešení nebylo součástí odevzdání, dovoluji si poznamenat, že samotné vyřešení efektivního počítání graflletů také není jednoduchý úkol, který ovšem kandidát velmi dobře zvládl.

Celkově hodnotím práci jako výbornou a s radostí ji doporučuji k obhajobě. Práce kvalitami rozhodně dosahuje prací, které by bylo možné navrhnout na ocenění, nicméně plánovaný publikační výstup nejspíše vznikne až v době následujícího zimního semestru.

Textová část práce

lepší OK horší nevyhovuje

Formální úprava <i>... jazyková úroveň, typografická úroveň, citace</i>	X	X		
Struktura textu <i>... kontext, cíle, analýza, návrh, vyhodnocení, úroveň detailu</i>	X	X		
Analýza	X			
<p>Struktura práce i jazyk je na velmi dobré úrovni a práce se velmi dobře čte. Text sice občas obsahuje některé malé typografické chyby a v teoretické části je pár vágnějších postupů, ale to nesnižuje čitelnost práce. Výsledky numerické části jsou velmi pěkné a propracované a kladně lze hodnotit analýzu této nelehké oblasti.</p> <p>Součástí odevzdání není související softwarový nástroj, nicméně si dovoluji podotknout, že i tento je z pohledu realizace kvalitní. Jedná se o nástroj umožňující velmi efektivně počítat grafetové distribuce pomocí některých externích nástrojů a porovnávat vzniklé výsledky.</p>				

Implementační část práce

lepší OK horší nevyhovuje

Viz komentář v předchozí části.

Celkové hodnocení Výborně
Práci navrhuji na zvláštní ocenění Na hraně Ano/Ne, viz celkové hodnocení

Datum

Podpis