

Posudek bakalářské práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce	František Szczepanik	
Název práce	Centralities in computation of approximate symmetries	
Rok odevzdání	2024	
Studijní program	Informatika	
Specializace	Obecná informatika	
Autor posudku	David Mikšaník	Oponent
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	X		
<p>Práce se zabývá možností vylepšení metody simulovaného žíhání pro výpočet aproximativních symetrií v komplexních sítích. Hrubě řečeno, výpočtem aproximativní symetrie rozumíme nalezení permutace vrcholů s „malým“ počtem pevných bodů a co s nejmenším počtem chybně zobrazených hran, tj. počet hran které se zobrazí na nehrany (a opačně). V jednotlivých krocích výpočtu přecházíme mezi permutacemi, které se liší transpozicí dvou vrcholů. Přesněji, z jedné permutace přejdeme (s určitou pravděpodobností) do druhé permutace náhodnou transpozicí dvou vrcholů, které vybíráme uniformně náhodně. Existuje hypotéza, že výpočet lze vylepšit, pokud budeme vrcholy vybírat dle distribuce definované pomocí centrality vrcholů (tj. budeme preferovat permutace takové, ve které budou podobné vrcholy ve smyslu centrality budou zobrazeny na sebe). Cílem práce je analyzovat toto vylepšení.</p> <p>Autor v první polovině práce nejprve definuje všechny potřebné pojmy, včetně různých pravděpodobnostních modelů komplexních sítí na kterých jsou v druhé polovině práce experimentálně vyhodnocena výše zmíněná vylepšení metody simulovaného žíhání. Autor nejenže zadání splnil, ale rozšířil ho následujícím způsobem. Dřívější práce při výpočtu aproximativních symetrií uvažovali pouze permutace s maximálně k pevnými body pro fixní k ($k = 0$ nebo $k = n/2$, kde n je počet vrcholů grafu). I když dává smysl mít horní hranici na počet pevných bodů, volba $k = 0$ nebo $k = n/2$ vypadá nahodile. Autor navrhuje způsob jak se zbavit této pevné hranice tím, že penalizuje permutace podle počtu pevných bodů. Výsledky ukazují, že tento přístup vede na obdobné výsledky jako pro $k = n/2$ a lepší než pro $k = 0$. Tento přístup bez pevné hranice je také přirozenější, a proto byl zvolen pro všechny experimenty.</p>				

Textová část práce

lepší OK horší nevyhovuje

	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		
Analýza	X	X		

Práce je napsána v anglickém jazyce na velmi dobré úrovni s minimálním počtem gramatických chyb (většinou se jedná o chybějící člen). Typografická úroveň práce je v pořádku s následujícími výhradami:

- V Kapitole 5 některé matematické výrazy nejsou správně vysázeny (např. „k“ místo „k“ v Obrázku 5.5, „N/2FP-SA“ místo „N/2FP-SA“ v kontextu KFP-SA).
- Dvojitý popis obrázků—jeden jako součást samotného obrázku a druhý jako součást textového pole v bloku 'Figure'. Pro pochopení je často potřeba dát oba popisy dohromady. Bylo by přehlednější uvádět popis pouze jeden.
- Odkazování na různé elementy textu probíhá někdy nestandardně (např. „section 5.1.1“ místo „Section 5.1.1“ na str. 46, „in the attachment A.8“ místo „in Appendix A.8“ na str. 38).

Samotný text je srozumitelný až na několik matematických nepřesností, které uvádím níže. Pozitivně hodnotím, že autor průběžně vysvětluje svůj postup a záměr, což přispívá k celkové přehlednosti práce. Práce je smysluplně strukturována s jasně definovanými cíli.

- Množina $Neighbors(P)$ v Algoritmu 1 je definován až v Sekci 2.3 ačkoliv Algoritmus 1 je uveden v Sekci 2.1. (str. 14)
- Popis DD modelu není správně. Za prvé, není uvedeno jestli mezi původním a zkopírovaným vrcholem je hrana. Za druhé, proces by měl začínat se souvislým grafem se dvěma vrcholy, nikoliv s jedním vrcholem—v opačném případě by proces pro $\sigma = 1$ vygeneroval buď prázdný nebo úplný graf, ale autor tvrdí, že proces vygeneruje úplný bipartitní graf. (str. 22)
- Z textu jsem pochopil, že definice $S_F(A)$ je poprvé zavedené v této práci. Bohužel definici doprovází nesrovnalosti. Za prvé, počet pevných bodů v permutační matici P by se měl značit jako $F(P)$, nikoliv F . Za druhé, ve jmenovateli se vyskytuje F , resp. $F(P)$, ale P není definována (P je definována pouze v čitateli). Domnívám se, že by definice mohla vypadat následovně

$$S_F(A) = \min_P \frac{\|A - PAP^T\|_1}{n(n-1) - 2F(P)(F(P) - 1)}.$$

Za třetí, není zřejmé přes jakou množinu se minimalizuje. **Myslím si, že je vhodné, aby autor definici na obhajobě objasnil.** (str. 24)

Výše jsem uvedl nejzásadnější nepřesnosti. V práci se dále vyskytují i další drobné nepřesnosti (např. tvrzení „which has 2475 edges“ na str. 35 není pravdivé, množina přes kterou se sčítá v definici $B(v_i)$ na str. 18 není správně dle obvyklé definice). V práci se vyskytuje pár překlepů, kterých je ale minimální množství.

Prezentace a analýza výsledků probíhá přehledným způsobem s přiměřeným množstvím detailů—výsledky experimentů jsou znázorněny ve formě grafů, které jsou řádně okomentovány a doplněny o autorovo možné vysvětlení. Statistické testy jsou dále použity na rozhodnutí zda výsledky jsou statisticky signifikantní.

Implementační část práce

lepší OK horší nevyhovuje

Nebyla součástí práce.

Celkové hodnocení Výborně
Práci navrhuji na zvláštní ocenění Ne

Datum 26. 8. 2024