

Posudek diplomové práce **Bc. Jana Kohouta:**

„Graph Clustering by Means of Evolutionary Algorithms“

Problematika analýzy sociálních sítí představuje jeden z nejprogresivnějších směrů v oblasti dobývání znalostí. Formálně lze sociální síť chápat jako množinu jedinců a jejich vzájemných vztahů a reprezentovat ji ve formě orientovaného grafu. Strukturu sociální sítě lze studovat mj. pomocí klastrovacích metod. Pojem klastr se ovšem v oblasti sociálních sítí používá ve dvou různých významech. Jednak jako skupina jedinců, kteří jsou spolu navzájem dobře propojeni, zatímco s jedinci mimo skupinu mají vazeb méně. Jedinci daného klastru jsou si tak podobní tím, v jaké sociální skupině se vyskytují. Další alternativou pojmu klastr je pak skupina jedinců, kteří mají v rámci sítě stejnou (anebo podobnou) roli či pozici (např. roli "otce"). V tomto případě už spolu uzly uvnitř daného klastru žádnou vzájemnou vazbu mít nemusí.

Cílem předkládané diplomové práce proto bylo navrhnout nový evoluční algoritmus použitelný pro klastrování grafů. Posлуhač měl zrekapitulovat relevantní přístupy (standardní i přístupy založené na evolučních algoritmech) a navrhnout a otestovat původní algoritmus pro klastrování grafů. Stěžejní roli přitom měla hrát vhodná reprezentace a evoluční operátory. Algoritmus měl být otestován na reálné úloze (z oblasti www anebo sociálních sítí) a získané výsledky měly být porovnány s dalšími přístupy.

Rešeršní část práce obsahuje zevrubný popis genetických algoritmů a metod používaných pro klastrování grafů. Problematikou sociálních sítí se rešerše nezabývá vůbec. Těžiště práce proto vidím především v návrhu nových přístupů pro evoluční klastrování grafů, které jsou zpracované v Kapitole 3. Pro experimentální ověření vlastností vyvinutých metod posluchač použil vlastní serverovou aplikaci „Graph Clusterizer,“ kterou implementoval v programovacím jazyce Java. Program, který je k práci přiložen na CD, představuje dobrý základ pro vývoj a implementaci dalších algoritmů ze studované oblasti. Filosofii vyvinutého systému i možnými směry jeho budoucího vývoje se podrobně zabývá Kapitola 4.

Formulace i analýza navržených klastrovacích technik by mohla být přesnější (viz např. klastrové křížení na str. 23 a 24) a měla by vzít v úvahu i strategie navrhované přímo pro analýzu sociálních sítí (např. PageRank, blokové metody aj.). Jisté rezervy má předkládaná práce, bohužel, i v experimentální analýze popisovaných metod. Prováděné testy zahrnovaly pouze uměle generovaná data, navíc poměrně omezené velikosti (do 75 uzlů) i rozsahu propojení. Adekvátní, resp. skutečně nalezený počet shluků pro řešenou úlohu je v práci uváděn jen sporadicky, podobně jako jejich velikost, složení aj. Vzhledem ke specifické volbě klastrovacích kritérií i ohodnocovací funkce by pak měla být správnost navrhovaných metod ověřena především na reálných datech. Takové testy však v práci chybí úplně.

Práce sama je psaná anglicky a až na řídké výjimky (např. na str. 15) má hezkou grafickou úpravu. Bohužel, obsahuje značné množství gramatických chyb a překlepů, které její čitelnost zhoršují (např. by namísto be na str. 11 a be namísto by na str. 21, preciously namísto precisely na str. 19, on namísto one na str. 24, is namísto if na str. 50, thought namísto though na str. 47, can not namísto cannot na str. 34, outputted namísto output na str. 38 a další). Čitelnosti textu příliš nepřispívá ani neadekvátní členění do příliš dlouhých vět a odstavců.

I přes výše uvedené nedostatky lze ovšem konstatovat, že práce z větší části splňuje svůj původní cíl. Diplomant pronikl dostatečně hluboko do problematiky evolučních algoritmů a prokázal schopnost samostatně navrhnout nové techniky pro klastrování grafů a implementovat je. Při obhajobě předkládané práce by se uchazeč, nicméně, měl zaměřit na detailnější porovnání vlastností testovaných metod. Zhodnocení získaných výsledků a zkušeností by měl podpořit podstatně rozsáhlejší testováním, a to zejména na reálných datech z oblasti webu anebo sociálních sítí.

Práci Bc. Jana Kohouta tedy doporučuji k obhajobě a po jejím úspěšném obhájení doporučuji uznat ji jako práci diplomovou.

V Praze, 27. 8. 2012

Doc. RNDr. Iveta Mrázová, CSc.
KTIML MFF UK