

Posudek oponenta disertační práce

Název práce: Anonymization of Social Network Datasets
Autor práce: Mgr. Jana Medková
Vedoucí práce: prof. RNDr. Josef Hynek, MBA, Ph.D.
Oponent práce: doc. Ing. Jaromír Kukul, Ph.D.

Předložená práce se zabývá velmi aktuálním tématem anonymizace databází v sociálních sítích. Téma je pojato exaktně s využitím teorie grafů a práce přináší nové metody anonymizace dat. Práce je psána v anglickém jazyce s promyšlenou strategií výkladu základních pojmů, pokročilých partií a nově vyvinutých algoritmů.

V prvních dvou kapitolách nalezneme seznámení s problematikou sociálních sítí, cílů a metod anonymizace, metody deanonymizace a analýzy systémů s využitím genetických algoritmů. Tak je formou rešerše zachycen aktuální stav oboru anonymizace sociálních sítí. Ve třetí kapitole jsou jasně formulovány a motivovány cíle výzkumu a disertační práce, které se posléze podařilo naplnit, o čemž svědčí i bohatá publikační činnost autorky. Nezbytné základy teorie grafů, maticové reprezentace dat, problematiky NP-těžkých úloh a teorie genetických algoritmů jsou shrnuty ve čtvrté kapitole, což usnadňuje porozumění novým metodám v navazujícím textu. Pátá kapitola je věnována kompozičnímu útoku a jeho analýze. Zde je vidět jasný autorský přínos při návrhu a počítačové implementaci algoritmu, studiu jeho časové náročnosti, efektivity a přesnosti a to jak teoreticky, tak pomocí počítačových simulací v prostředí Matlabu. Jiný úhel pohledu nalezneme v šesté kapitole, kde je navržena a popsána nová heuristická metoda založená na náhodném ovlivňování k -stupňového rozhodovacího procesu. Jde o modifikaci klasického chamtivého algoritmu, kde přidání šumu narušuje prvoplánovou posloupnost jednotlivých kroků, což zvyšuje šance na nalezení řešení. Exaktní popis heuristického algoritmu je opět doplněn počítačovými simulacemi, vyhodnocením časové náročnosti a efektivity. V sedmé kapitole je navržen, popsán a implementován nový hybridní algoritmus spojující obecné principy anonymizace a genetický algoritmus. Přitom je věnována pozornost určení ceny anonymizace, teoretické analýze složitosti jednotlivých dílčích úloh a podrobnému popisu hybridizace s genetickou optimalizační metodou. Simulační experimenty i zde demonstrují efektivitu nově navrženého postupu.

V rámci diskuse při obhajobě práce se nabízejí následující otázky:

- Je možné shrnout rozdíly v efektivitě mezi klasickými metodami anonymizace a novými postupy uvedenými v práci? Jde především o zobecnění poznatků uvedených v práci.
- Při kombinaci chamtivého přístupu s vneseným náhodným šumem lze s nadsázkou říci, že jde o jistou stochastickou formu zdvořilosti, kdy v daném kroku dostane přednost jiné dílčí

řešení. Bylo by možné uvedenou zdvořilost modelovat nějakým jiným způsobem, který by případně šel parametrizovat?

- Pro jak velké sociální sítě je možné implementovat klasické, respektive nově navržené metody?

Vzhledem k výše uvedeným faktům doporučuji předloženou disertační práci k obhajobě před příslušnou komisí v souvislosti s udělením titulu Ph.D.

V Praze dne 30.5.2023

doc. Ing. Jaromír Kukal, Ph.D.