

Oponentský posudek na disertační práci Mgr. Jana Medková: Anonymization of social network datasets

Předložená disertační práce je zaměřena na důležitý problém anonymizace databází sociálních sítí. Práce je dobře strukturovaná.

První kapitola je úvodem do zkoumané problematiky a seznámení se zkoumanými oblastmi v disertaci.

Druhá kapitola je přehledem současného stavu problematiky, který obsahuje základní pojmy, definice, metody anonymizace, deanonymizační útoky, hodnotící nástroje, genetické algoritmy v analýze sociálních sítí.

Ve třetí kapitole jsou formulovány 3 výzkumné otázky. První otázka se týká kompozičního útoku na databáze sociálních sítí. Druhá otázka je věnována vlivům na efektivnost k -DA algoritmu. Třetí otázka se zaměřuje na modifikace pro zlepšení k -automorfnní anonymizační metody.

Čtvrtá kapitola shrnuje základní potřebné teoretické poznatky zejména z oblastí teorie grafů, NP-obtížnosti, genetických algoritmů. Další kapitoly jsou strukturovány podle výzkumných otázek.

Pátá kapitola je věnována kompozičnímu útoku, kde je ukázáno, že tento útok může být aplikován i na databáze sociálních sítí.

Šestá kapitola je věnována druhé výzkumné otázce a vyústí v návržení heuristické metody přidávání šumu do k -DA algoritmu.

Sedmá kapitola je řešením třetí výzkumné otázky, které navrhuje hybridní k -automorfnní anonymizační algoritmus s využitím principů genetických algoritmů.

Osmá otázka je shrnutím, které obsahuje přehled dosažených výsledků řešení výzkumných otázek.

V další části jsou uvedeny komentáře k hodnotícím kritériím disertační práce:

1. Aktuálnost tématu disertační práce

Téma disertační práce je velice aktuální a nesmírně důležité a zajímavé. V současné době rozšiřujících se sociálních sítí nabývá tento problém na stále větším významu. Roste i problém zneužívání dat ze sociálních sítí. Anonymizace databází sociálních sítí je tedy velice aktuální téma.

2. Splnění cílů disertační práce

Doktorandka formulovala 3 základní výzkumné otázky v oblasti anonymizace databází sociálních sítí. Cílem disertace bylo zodpovězení těchto otázek, což se doktorandce bezesbýtku podařilo návržením algoritmů a jejich ověřením na datech. Základní cíl práce byl tedy splněn.

3. Metody a postupy řešení

Doktorandka klasicky vycházela z přehledu literatury v této oblasti a zpracovala více než 160 předních zahraničních publikací. Vytvořila přehled dosud publikovaných anonymizačních a deanonymizačních metod. Z tohoto přehledu vyplynuly 3 díleí problémy, ze kterých doktorandka formulovala 3 výzkumné otázky (popsané výše). Doktorandka analyzovala tyto otázky pomocí matematického aparátu (viz čtvrtá kapitola). Analýza byla vyvážená od teoretické analýzy, přes návržení metod, implementaci algoritmů v MATLABu, experimentování na datech, srovnání s jinými algoritmy, použití ohodnocovacího nástroje SecGraph k měření zranitelnosti metod vůči útokům.

4. Výsledky disertační práce a konkrétní přínosy disertanta

Hlavními výsledky disertační práce byly odpovědi na tři výzkumné otázky navržením nových metod a algoritmů, jejich implementací v MATLABu a testováním na syntetických a reálných datech. Výsledky vycházely ze známých ověřených metod, které byly modifikovány, kombinovány a obohaceny. U některých výsledků by mohla být zvýrazněna přesnější hranice mezi známými a nově dosaženými výsledky.

5. Význam pro praxi a rozvoj studijního oboru Aplikovaná informatika

Navržená řešení vybraných problémů anonymizace databází sociálních sítí má zásadní význam pro praxi vzhledem ke stále rostoucím sociálním sítím a nebezpečí zneužívání dat. Zvolené téma jistě patří do studijního oboru Aplikovaná informatika a je přínosem pro rozvoj daného studijního oboru vytvořením přehledu publikovaných metod, stanovením a řešením výzkumných otázek a navržením nových postupů.

6. Formální úprava disertační práce a její jazyková úroveň

Disertační práce je psána v anglickém jazyku a má více než 130 stran. Práce má velmi dobrou formální úpravu textu, matematických vzorců, popisu algoritmů, obrázků i tabulek a má rovněž dobrou jazykovou úroveň. Práce je velmi přehledná, orientaci v práci usnadňují seznamy obrázků, tabulek a zejména seznamy použitých symbolů a zkratk.

7. Publikační úroveň doktorandky

Doktorandka publikovala v souvislosti s tématem disertace 6 kvalitních výstupů v zahraničních časopisech nebo sbornících konferencí. Publikovala další výstupy, které nejsou přímo z oblasti anonymizace databází sociálních sítí. Rovněž se účastnila výzkumných projektů a mezinárodní spolupráce.


8. Připomínky a závěrečné zhodnocení disertační práce

Doktorandka předložila návrhy na řešení vytypovaných výzkumných otázek a tím splnila základní cíl práce. Domnívám se, že předložená práce splňuje požadavky na disertační práci. Doporučuji práci k obhajobě a po úspěšné obhajobě udělit titul Ph.D.

Otázky pro diskusi při obhajobě:

- 1) Jaký je váš postoj k názoru, že perfektní anonymizace ztěžuje další zpracování údajů v databázích? Jak volit „stupeň“ anonymizace?
- 2) Který z vašich přínosů považujete za nejdůležitější?

Praha, 31. 5. 2023


Prof. RNDr. Ing. Petr Fiala, CSc., MBA
Vysoká škola ekonomická v Praze