Oponentní posudek na disertační práci (Doctoral thesis)

Název práce: Anonymization of social network datasets (Anonymizace datových souborů sociálních sítí)

Autorka: Mgr. Jana Medková
Školitel: prof. RNDr. Josef Hynek, MBA, Ph.D., Fakulta informatiky a managementu, Univerzita Hradec Králové
Oponent: prof. RNDr. PhDr. Antonín Slabý, CSc., Fakulta informatiky a managementu, Univerzita Hradec Králové

Struktura a obsah práce

Práce napsaná v anglickém jazyce má 129 stran a 101 stran základního textu. Základní text je tvořen 8 číslovanými kapitolami včetně Úvodu a Závěru.

Kapitola 1 – Úvod (Introduction) - je stručným úvodem do tématiky práce, kterou tvoří sociální sítě zkoumané z pohledu informační bezpečnosti. Práce je věnována zásadám shromažďování dat o uživatelích a zejména způsobům anonymizace a nebezpečím deanonymizace těchto dat.
Kapitola 2 – Stav poznání v oblasti (The state of art) je věnována vybraným významným výsledkům v oblasti tématu práce a formálním definicím a označení základních pojmů, jejich vysvětlení a ilustracím. Jde např. o pojmy: Sociální sítě s uživatelskými atributy, kvazi identifikátor, anonymizace, sociální síť bez atributů. Dále se kapitola věnuje rozmanitým metodám, technikám, pojmům a výsledkům a speciálnímu dalšímu pojmovému aparátu zavedenému různými autory (např. anonymizace relačních dat, pojem k-anonymizace) a souvisejícím pojmům (suppression, l diversity, t closeness, se zdůrazněním přístupů, k jejichž zlepšení autorka přispěla. Pozornost je věnována i deanonymizačním útokům a možnosti využití genetických algoritmů v analýze sociálních sítí.
Kapitola 3 – Cíl práce a výzkumné otázky (Research questions and objectives) formuluje cíl práce a související výzkumné otázky.
Kapitola 4 – Preliminaries (Pojmový aparát z oblasti) formuluje pro práci podstatný pojmový aparát z oblastí využitých v autorčiných přístupech. Jde mj. o Teorii grafů, ROC, ekvivalence tříd v G_A., NP hard problémy a Genetické algoritmy.
Následují 3 kapitoly dokumentující hlavní přínosy disertační práce a autorčiny přístupy k řešení 3 výzkumných otázek práce.
Kapitola 5 – Composition attack (Sdružený útok) je věnována kompozičnímu útoku se věnuje autorčinu přístupu k 1. výzkumné otázce. Autorka ukazuje, že kompoziční útok (dříve zmiňovaný jako informační hrozba pro relační databáze) může být aplikován i na databáze sociálních sítí.
Kapitola 6 – Heuristic noise addition method (Heuristická metoda přidávající šum) odpovídá 2 výzkumné otázce. Autorka navrhuje novou heuristickou metodu spočívající v přidávání šumu do existující metody k-stupňového anonymizačního algoritmu.
Kapitola 7 – Hybrid algorithm for k-automorphism anonymization (Hybridní algoritmus pro anonymizaci založenou na K automorfismu) odpovídá 3 výzkumné otázce. Kapitola demonstruje využití genetických algoritmů v k autonomní anonymizační metodě a popisuje nový autorkou navržený hybridní k-autonomní anonymizační algoritmus.
Kapitola 8 – Conclusion (Závěr) představuje krátké shrnutí.
Následují obligátní kapitoly References (Literatura) - obsahující 161 položek, List of figures, List of Tables, Authors publications (Autorčiny publikace) a Research activities (přehled účastí na výzkumných projektech a zahraničních stážích) a Additional experimental results Supplementary material.

Cíl práce, výzkumné otázky

Cílem práce je přispět podstatným věrohodným příspěvkem k tématice anonymizace a vývoji anonymizačních metod, zejména k -anonymizačních metod (k -anonymization methods) a zejména metody k-DA a k-automorphism a dále k důležitému problému (hrozbě) – kterou je composition attack - deanonymization.
Autorka formuluje v této souvislosti 3 výzkumné otázky formulující nevyřešené problémy jimž se v práci věnuje, ke kterým svou prací podstatně přispěla
Otázka 1. Jaké modifikace by měly být provedeny v návrhu tzv. sdruženého útoku (composition attack) tak, aby byl útok použitelný pro data sociální sítě? (What modifications should be done to the design of the composition attack such that the attack becomes applicable to the social network data?)
Otázka 2. Jaký účinek má korekce anonymizovaných hodnot ve skupinách d^*_G na účinnost algoritmu k-DA? (What effect does the correction of the anonymized values in groups of $d*G$ have on the efficiency of the k-DA algorithm?)
Otázka 3. Jaké modifikace lze provést v metodě k-automorphism tak, že zefektivní datový nástroj / postup? (What modifications can be done to the k-automorphism method such that it preserves data utility better?) Zde jde autorkou vytvořen novátorský hybridní algoritmus pro algoritmus HAkAu (Hybrid algoritmus for k-automorphism method) se zamýšlenou redukcí složitosti (v NP hard části problému – na principu nalezení a rozšiřování izomorfních podgrafů pomocí genetických algoritmů.)

Metody a postupy použité v práci

Autorka vychází z rozsáhlé rešerše literárních zdrojů zkoumané problematiky. Definuje a využívá pojmový aparát, výsledky, algoritmy a postupy specifické aplikační oblasti, které patří věcně do oblasti aplikované diskrétní matematiky a teorie grafů (definice, pojmový aparát, formulace postupů a výsledků) , informatiky – (algoritmizace, aplikace genetických algoritmů a speciální postupy k redukci složitosti) i speciální statistické metody a další metody (ROC analýza použitá k hodnocení kvality klasifikátorů) a Secgraph k měření datových souborů a další speciální metody a přístupy. K realizaci algoritmů a experimentům s nimi autorka využila systém Matlab.
Metody a postupy použité v práci lze souhrnně hodnotit, jako adekvátní, ke splnění cílů vedoucí a částečně novátorské.

Aktuálnost tématu, disertabilita, splnění cílů práce

Tématika práce je vysoce aktuální, velmi obtížná, dosti široce pěstovaná a široce využitelná. Tématika práce je rozhodně disertabilní. Práce svojí tématikou patří do oboru doktorského studia Aplikovaná informatika.
Autorka volila tématiku, která je jejím delším výzkumným zájmem a tématiku, ve které již dosáhla kvalitních výsledků, na které může dobře navázat, a současně tématiku ve vazbě na její profesní minulost. Práce navazuje i na tradici výzkumů, metodických postupů, zkušeností, výsledků a vybavení v dané oblasti na FIM UHK, na výsledky autorčiných studijních pobytů a grantových aktivit i expertní znalosti školitele.
Cíl práce byl splněn. Výzkumné otázky jsou vhodně voleny a přispívají k dosažení hlavního cíle práce.

Přesnost práce, formální stránka práce

Práce je napsána v angličtině, jasným a přesným jazykem. Volba anglického jazyka má výhodu v možnosti používat anglické terminologie v oblasti, která není v českém jazyce plně

stabilizována. Definice a výsledky jsou přesně formulovány. Fakta v rešeršní části jsou dobře vybrána, jasně formulována a demonstrují též velmi dobře přínosy autorčiny. Formální stránka práce i přesnost vyjadřování, úprava vzorců, diagramů a grafů mají vynikající úroveň. Práce má jasnou strukturu, proto je možno se v ní dobře orientovat.

Výsledky práce a přínosy práce

Práce přináší výsledky v oblasti teoretické, metodologické i praktické.
Za 3 důležité teoreticko-metodologické výsledky lze považovat velmi pěkné, přesné a srozumitelné a komplexní a hluboké zpracování tématiky formulované ve 3 výzkumných otázkách. s použitím adekvátních metod. Je možno konstatovat, že autorka přispěla ke všem třem v práci formulovaným výzkumným otázkám kvalitním a věrohodným příspěvkem. Přístupy uvedené v práci jsou moderní, ověřené dosti rozsáhlou autorčinou pedagogickou a výzkumnou prací a mohou být opakovaně použity a dále modifikovány a dále rozvinuty.

Publikační činnost autorky

Mgr Medková je ke dni vypracování tohoto posudku autorkou nebo spoluautorkou sedmi publikací. Dva články jsou v impaktovaných časopisech Computers and Security a Social Network Analysis and Mining (oba časopisy jsou v Q2 podle WoS). Dalších pět publikovaných příspěvků má autorka na mezinárodních konferencích. Autorka tedy splňuje a překračuje publikační požadavky stanovené na k závěrečné obhajobě práce

Poznámky a otázky do diskuse při obhajobě

Diskusi navrhuji věnovat způsobu autorčina řešení podstatné otázky 3, konkrétně hybridnímu algoritmu HAkAu (Hybrid algoritmus for k-automorphism method), například inovativnímu autorčinu návrhu chromosomové representace v genetickém algoritmu.

Závěr:

Práce splňuje v nároky na disertační práce kladené. Jde o kvalitní, přesný a hluboký text poskytující pěkný vhled do problematiky řešené v práci. Rozsah, originalitu, systematičnost a náročnost aktivit studentky, tak jak je dokumentuje předložená disertační práce, je možno velmi pozitivně hodnotit. Je nutno též konstatovat, že v případě Mgr. Medkové jde již o zralou osobnost, která v práci unikátně integruje své vzdělání a zkušenosti v oblasti výuky matematických a informatických předmětů na FIM UHK a zúročuje delší výzkumné a profesní aktivity konané v oblasti tématiky práce. Doporučuji disertační práci k obhajobě a též, aby Mgr. Medkové byl po úspěšné obhajobě udělen titul Ph.D.

Hradec Králové 15. 5. 2023

Antonín Slabý