



Posudek oponenta na diplomovou práci

Bc. T. Zikmund: Detekce překrývajících se komunit v bipartitních grafech.

Předkládaná diplomová práce se věnuje problému, který patří do oblasti analýzy struktur v datech – detekci komunit v (bipartitním) grafu, a to jak komunit bez překryvu, tak komunit překrývajících se. Jde o zajímavý problém s ještě zajímavější motivací, která pochází z matematického modelování společnosti ve starověkém Egyptě.

Student bezezbytku naplnil cíle práce stanovené v jejím zadání, úspěšně se vypořádal jak se stránkou teoretickou, tak se stránkou praktickou – implementační. V tomto ohledu mám drobné výtky snad pouze je splnění posledního bodu zadání, tj. „Proveďte diskuzi získaných výsledků“. Jako výsledky testování jednotlivých detekčních metod autor předkládá 4 tabulky a 8 obrázků, jejich textový popis/rozběr ale zabírá méně než stranu textu (z toho ještě část tvoří technický popis grafů). Diskuzi výsledků bych prostě očekával obsáhlejší.

Zásadnější výtky mám k vlastnímu textu práce. Zhruba řečeno – matematický úvod je místy dost vágní, větší přednost / formální důslednost by zde neuškodila (vizte poznámky na konci posudku – během obhajoby by se měl student vyjádřit především k poznámkám 3+5 a 14). Další kapitoly týkající se oblastí, které mi nejsou tak blízké se také špatně čtou, protože autor velice málo vysvětluje (konec konců jde o diplomovou práci, ne o článek do odborného časopisu psaný pro úzce profilovanou skupinu vědců), často používá odborné termíny, které nevysvětluje.

Doporučuji předkládanou diplomovou práci přijmout k obhajobě a ohodnotit ji známkou

C (dobře).

V Praze dne 17.5.2020

Petr Ambrož



Poznámky

1. *Poznámka 1.1.3, bod 2:* Tvrzení je sice pravdivé, ale mírně zavádějící, i nejednoduché neorientované grafy mají symetrickou adjacenci matici.
2. *Poznámka 1.1.3, bod 3:* Proč by měla přítomnost smyčky nutně znamenat $A_{ii} = 2$. Vždyť smyček může být na vrcholu víc.
3. *Str 18, ř. 5:* Hypergraf je možné reprezentovat bipartitním grafem. Jak? Proč tam ta poznámka je, když to není vysvětleno/naznačeno. Navíc bipartitní graf je definován až v následující sekci.
4. *Definice 1.1.5:* toto není – jak autor připouští v poznámce pod čarou – definice incidenční matice (ta vyjadřuje incidenci, tj. které hrany a vrcholy spolu souvisí). Nápodvedou studentovi mohlo být to, že vztah (1.1) identický vztahu (1.2) používá k definici matice adjacenci.
5. *Poznámka 1.1.6:* nechápu, co tím chce autor říct, pravděpodobně se týká výše naznačeného a nevysvětleného vztahu bipartitních grafů a hypergrafů.
6. *Definice 1.1.10:* Co tam dělá ta 2. věta? To je část definice 1.1.11, která se tam patrně dostala nějakou neopatrnou editací textu.
7. *Str 20, začátek sekce 1.3.2:* Co myslíte větou „je maticovou normou, která je blízká euklidovskému prostoru“?
8. *Poznámka 2.2.1:* Co jsou matice dokumenty a slova? To jsou nějaké standardní problémy v data miningu? I pokud tomu tak je, zasloužilo by si to alespoň stručné vysvětlení.
9. *Sekce 2.3, 2. odstavec (popis různého chápání pojmu komunita):* Vzhledem k tomu, že jde o stěžejní pojem celé práce čekal bych minimálně pár příkladů.
10. *Sekce 2.3.1, 3. odst:* Dendrogram je definovaný jako kořenový strom. V kapitole o teorii grafů není vysvětleno, co to je strom, natož strom kořenový.
11. *Popisek Obr. 2.3:* „bipartitní graf s ohodnocenými hranami“ – čím? „hrany v původním grafu jako projekci“ – čeho? kam? (případně podle čeho?)
12. *Definice 3.0.1:* Formální definice používá nedefinovaný pojem „stochastický blokový model“. Krátké vysvětlení na začátku Kapitoly 3 v tomto ohledu nepovažují za dostatečné.
Navíc ač v té definici není ani slovo o grafu (potažmo jeho vztahu k v definici použitým objektům), pracuje se tam s hranami a vrcholy a s jejich počty.
13. *Věta 3.0.2:* Co je Λ ?
14. *Sekce 3.1.3, popis Kernighanova-Linova algoritmu:*
 - Co je počáteční stav? Jak se volí výchozí nastavení?
 - Jak ze zvolí/mění počet komunit?

Dále se v práci nachází několik překlepů (např. komunt str. 25/–1), problémů ve shodě podmětu s přísudkem (grafy byli, modely byli,...), či nekonzistence ve značení (vektory opakovaně se šipkou i bez).