Posudek vedoucího diplomové práce:

Diplomová práce: Detekce překrývajících se komunit v bipartitních grafech
Autor diplomové práce: Bc. Tomáš Zikmund

Diplomová práce studenta Tomáše Zikmunda se zabývá problematikou detekce překrývajících se komunit ve složitých sítích. Motivačně tato práce navázala na vstupní studie do problematiky aplikace metod detekce komunit na data o společnosti lidí Staré Říše Egypta v rámci spolupráce s egyptology. Jednalo se o GAČR projekt, v jehož rámci se vyvíjely a používaly matematické modely k hodnocení archeologických nálezů a formalizovaného zdůvodnění tehdejších společenských procesů.

Na počátku studie student vyšel z několika článků a knih, své bakalářské práce a svého výzkumného úkolu. Student se zaměřil na problematiku detekce nepřekrývajících i překrývajících se komunit, které lze aplikovat nejenom na unipartitní grafy, ale upravit i na bipartitní sítě. Cílem diplomové práce bylo nalezení postupu identifikující parametry modelu, který umožňuje detekci překrývajících komunit na bipartitních grafech, která není dosud v odborné literatuře řešena. Specificky se pak zabýval metodami založenými na stochastickém blokovém modelu, při kterém se pozorované grafy generují jako projevy blokového modelu komunit. Postupně student sestavil sadu nedávno rozvíjených metod, které typicky vycházejí ze zobecněného stochastického blokového modelu s Poisonovou distribucí pravděpodobností hran a ze kterých v rámci svého výzkumného úkolu sestavil zobecněný model, který umožňuje generovat grafy jak unipartitní, tak bipartitní, a to jak s nepřekrývajícími se, tak i překrývajícími se komunitami. Netriviální detekční algoritmus modelu student vyvinul až ve své diplomové práci. Ačkoliv student navazuje na principy známých algoritmů identifikující stochastické blokové modely, lze prohlásit, že jejich výsledná kombinace včetné unikátního použití Kroneckerových symbolů v navržených a provedených důkazech je čistě práce studenta a jeho cíleného intenzivního úsilí. Řešení z dnešního pohledu je přímočaré a jasné, je to však výsledek velmi intenzivního hledání, mnohastránkových odvozování a zkoušení různých variant, kterým se diplomant trpělivě věnoval po mnoho měsíců. Práce obsahuje ale i návrh implementační strategie, která vede k výpočtu s rozumnou komplexitou. Výsledky byly rovněž experimentálně ověřovány na syntetických tak i reálných datech. Student v diskusi výsledků upozorňuje na řadu zajímavých jevů.

Student chodil na konzultace pravidelně a vesměs dobře připraven. Přiznám se, že nebylo jednoduché studenta navigovat rozumným směrem, neboť problematika je značně široká a student se svým velmi dobrým matematicky přesným přístupem měl tendenci řešit a vyčerpávajícím způsobem porozumět všem nových problémů. Výsledkem je však vyvážená diplomová práce, ve které se student snaží o stručnou, ale jednotnou formalizací vytvořit přehled základních principů a taxonomii. Rigorózní přehled by však vydal na rozsah dizertační práce, ke které však předložená práce nemá daleko. Těžiště práce je a po celou dobu jejího řešení bylo v netriviální kombinaci metod, která unifikuje řadu přístupů a umožňuje řízeným způsobem generovat a detekovat jejich varianty. Výsledek byl získán po mnoha experimentech a dílčích odvozování. Student model a detekci jeho parametrů ověřil poměrně vyčerpávajícími experimenty. Zajímavým výsledkem je například to, že postup dokáže detekovat parametry modelu, který z pohledu věrohodnosti vysvětluje vstupní syntetická data lépe než model, podle kterého byl vstupní graf generován.

Po formální stránce se jedná o vysoce nadstandardní diplomovou práci rozsahu 74 stran. Práce má velmi dobrou grafickou úroveň se zanedbatelným počtem překlepů v textů. Postupy svědčí o schopnosti studenta využít všech znalostí a dovedností získaných během bakalářského a magisterského studia. Práce patří určitě mezi nadstandardní práce velmi dobrých studentů. Byl jsem potěšen, že se našel student, který se nebojí v rámci své diplomové práce řešit takto složitou tématiku, kdy apriori není znám správný postup aplikace vybrané matematické teorie. Měl bych podotknout, že jsme navržený model bez detekce již publikovali na konferenci CNA 2018 (indexace SCOPUS), kde vzbudil značnou pozornost a řada výzkumníků byla zvědava, zda se pro model podaří vytvořit detekční algoritmus jeho parametrů. Student algoritmus zřejmě odhalil již mnohem dříve než v listopadu 2019. Vzhledem k velmi pomalé počáteční konvergenci, kdy se její projevy dostavují až po mnohahodinovém výpočtu, student však postupu řadu měsíců nevěřil. Student pracoval velmi samostatně, trpělivě sdílel mou víru v dobrý konec a snažil se vytvořit sjednocující nadhled.

Studentovi jsem doporučil pokračovat ve výzkumné dráze jako PhD student. Výsledek zamýšlím se studentem publikovat jak na nejbližší konferenci CNA, tak i v časopisu s impakt faktorem. Dovoluji si proto komisi navrhnout, aby uvážila případné další ocenění této diplomové práce.

Do případné diskuse bych navrhoval následující témata:

1. Jaké jsou možnosti urychlení pomalé počáteční konvergence algoritmu?
2. Je možné zobecnit postup pro obecné n-partitní grafy?

Vzhledem k výše uvedeným bodům a k celkovému přístupu studenta k řešení diplomové práce si výsledek dovoluji ocenit známkou

Výborně (A).

V Řeži 17. července 2020

Ing. Radek Mařík, CSc.
Katedra telekomunikace, FEL ČVUT Praha