

## POSUDEK OPONENTA BAKALÁŘSKÉ PRÁCE

**Název:** Node-attributed community detection

**Autor:** Kateřina Vokálová

### **Shrnutí obsahu práce**

Cílem práce bylo srovnat algoritmy pro detekci komunit v grafech s atributy uzlů. Konkrétně autorka uvažuje model grafu s předem definovanými komunitami (tj. se znalostí "ground truth"), jejichž velikosti jsou popsány mocninným rozdělením (power law), tzv. X-Mark benchmark. V závislosti na parametrech modelu autorka srovnává výsledky detekce komunit při použití různých algoritmů pro detekci komunit a různých metrik. Výsledky detekce komunit oproti ground truth hodnotí pomocí tzv. normalized mutual information.

V úvodu práce autorka pečlivě provede čtenáře nutnými znalostmi, s četnými odkazy jak na zdroje prezentovaných tvrzení, tak na rozšiřující literaturu. Tato část práce je velmi čtivá a výborně zpracovaná. Oceňuji pečlivost, s jakou autorka k práci přistoupila.

V druhé části práce autorka prezentuje vlastní výsledky srovnání úspěšnosti různých algoritmů za použití různých metrik. Nejlepší výsledky byly dosaženy pomocí algoritmů Eva a Louvain algoritmu s metrikou "mixed similarity". Algoritmus Eva pro detekci komunit používá jak informace o hranách, tak attributech uzlů (tzv. simultaneous fusion). Louvain algoritmus byl použit na transformovaný graf bez atributů, jehož váhy nesou informaci o původních attributech uzlů (tzv. early fusion), přičemž autorka testovala různé metriky definující tento vztah. Dalším výsledkem bylo pozorování, že zvyšující se "mixing parameter" modelu negativně ovlivňuje výsledky všech algoritmů a "noise parameter" negativně ovlivňuje výsledky algoritmů, které používají atributy. I tato část je přehledně členěná a algoritmy roztřízeny do skupin podle typu. Autorka pečlivě diskutuje výběr parametrů a metrik a poskytuje odkazy na alternativní přístupy. Rozumím, že výsledky jsou z podstaty otázky komplexní a není jednoduché je prezentovat. Přesto si myslím, že by práci prospěl nějaký obrázek, který by konsolidoval znalosti z dílčích grafů, opatřený statistickým testem týkajícím se hlavních tvrzení. V závěru, kde se diskutují hlavní výsledky, by pak na tento obrázek měl být odkaz. Vzhledem k již tak velkému rozsahu práce by s tímto dodatkem práce pravděpodobně šla kvalifikovat jako magisterská, a proto tento nedostatek nepromítám do známky.

Celkově práci hodnotím jako vynikající - text je psán srozumitelně, je dobře členěný, obsahuje vlastní výsledky a rozsah silně převyšuje obvyklý rozsah bakalářských prací.

### **Celkové hodnocení práce**

**Téma práce.** Téma bylo zvoleno vhodně, přiměřené bakalářské práci, a zadání splněno.

**Vlastní příspěvek.** Práce obsahuje vlastní příspěvek autora, konkrétně kvantitativní srovnání výsledků různých algoritmů na detekci komunit v grafech s atributy uzlů, v závislosti na parametrech modelu. Z textu je zřejmé, co je převzato (a řádně odcitováno) a co je vlastní výsledek.

**Matematická úroveň.** Práce obsahuje řádně zformulované definice.

**Práce se zdroji.** Zdroje jsou řádně odcitovány a je jich nadstandardní množství.

**Formální úprava.** Formální úprava je vynikající, text je přehledně členěn, angličtina na vysoké úrovni. Obrázky jsou řádně anotovány.

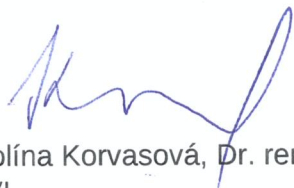
### **Připomínky a otázky**

1. V práci navrhuje, že by bylo vhodné provést rozsáhlou studii, která ověří nebo vyvrátí tvrzení, že velikosti komunit v reálných sítích jsou rozdělené podle mocninného rozdělení. Neočekávali bychom, že v některých sítích to tak bude a v jiných ne? Můžete uvést příklady reálných sítí, kde bychom očekávali mocninné rozdělení, a kde naopak ne?

## **Závěr**

Práci považuji za vynikající a doporučuji ji uznat jako bakalářskou práci.

*Návrh klasifikace vedoucí/oponent sdělí předsedovi zkušební (sub)komise.*



Karolína Korvasová, Dr. rer. nat.

KSVI

28. 8. 2023