

## Posudek oponenta bakalářské práce

Jméno a příjmení uchazeče/ky: Nikola Kalábová

Název práce: Applications of graph theory in protein function prediction

**A. Bodové hodnocení jednotlivých aspektů práce (označte právě jednu z možností)**

1. Rozsah BP a její členění	
×	A - přiměřené, odpovídají charakteru BP a významu jednotlivých částí
	B - nevyrovnané, členění není logické n. rozsah jednotlivých částí nekoresponduje s jejich významem
	C - uspokojivé, rozsah některých částí nedostačuje
	<b>N - nedostatečné</b>

2. Odborná správnost	
	A - výborná, bez závažnějších připomínek
×	B - velmi dobrá, s ojedinělými drobnými závadami (nejasnost výkladu, chyby ve vzorcích nebo chemických názvech, nedokonalý popis metod nebo výsledků)
	C - uspokojivá, s četnějšími drobnými závadami
	<b>N - nevyhovující, s hrubými chybami</b>

3. Uvedení použitých literárních a j. zdrojů	
×	A - bez připomínek, všechny převzaté údaje s citací zdroje, celkový počet citací odpovídá charakteru práce
	B - uspokojivé, s občasnými neobratnostmi zejm. v umístění odkazů, nebo s celkově nižším počtem citací
	C - s vážnějšími závadami, např. převažují "nestandardní" odkazy na učebnice, přednášky, webové stránky, nebo se ojediněle vyskytuje opominutí odkazu na zdroj převzatých dat
	<b>N - nevyhovující, velmi málo citací, ev. rysy plagiátu (časté opomíjení odkazu na zdroj převzatých dat, popř. opsání velkých částí textu)</b>

4. Jazyk práce	
	A - výborný, práce je napsána čtivě a srozumitelně, bez závažnějších gramatických n. pravopisných chyb
×	B - velmi dobrý, ojedinělé stylistické neobratnosti, gramatické n. pravopisné chyby
	C - uspokojivý, četnější slohové neobratnosti, gramatické n. pravopisné chyby, ojediněle se vyskytují obtížně srozumitelné n. nejednoznačné formulace
	<b>N - nevyhovující, s četnými hrubými chybami</b>

5. Formální a grafická úroveň práce	
	A - výborná, bez překlepů a chyb ve formátování
×	B - velmi dobrá, ojedinělé chyby formátu citací, překlepy, chybějící zkratky apod.
	C - uspokojivá, s ojedinělými většími (např. vynechání stránky) nebo četnějšími drobnými chybami
	<b>N - nevyhovující, s četnými hrubými chybami</b>

Slovní komentář k bodům 1. až 5.:

Autorka v práci prezentuje rešerši současných grafově-teoretických metod pro predikci funkce

proteinů. Objemově je práce poměrně obsáhlá, popisuje několik rámcových skupin metod pro predikci funkce, zaměřuje se na proteinové interakční sítě, a následně recenzuje několik konkrétních algoritmů. Práce obsahuje výsledky systematického porovnání několika souhrnných vlastností predikcí pomocí těchto algoritmů, tyto jsou převzaté z výzkumu Ji *et al.* (2012)

Metody a výsledky jsou prezentovány sice místy poněkud povrchně (to může být částečně způsobené značným rozsahem rešerše), ale srozumitelně a dostatečně ilustrativně. Práce neobsahuje vlastní experimenty, kód nebo výsledky, to ale odpovídá to rešeršnímu charakteru práce.

Většina nalezených nedostatků je spíše formálního charakteru a neměla by tvořit překážku obhajitelnosti práce: Angličtina práce je gramaticky v pořádku, ale na některých místech by bylo vhodné upravit “wording” a přidat kontext tak, aby předávaná informace a její význam pro výsledky práce byly zřejmé. Příklady problému lze najít v sekcích 3.4.3 a 3.4.4, občas se v anglické větě vyskytne český pořádek slov, např. na začátku sekce 1.3.1. Srozumitelnosti některých částí textu nepomáhá používání sekcí a nadpisů jako “seznamů” bez explicitního uvedení, např. v sekcích 2.3, 3.3 a 4.5.1. Nepříjemným problémem je poměrně nestandardní (a občas matoucí) vsazení obrázků a tabulek přímo do textu – tyto jsou použité jako součást toku textu bez odkázání, což často ztěžuje pochopitelnost jejich významu (především v sekci 5.1.4, nebo dále v 5.2.3, kdy čtenář vysvětlení a diskuzi výsledků *najde* až o několik stránek později). Několik drobností by pravděpodobně zmizelo dodatečnou korekturou textu, např. nevytisknutelné HTTP odkazy v poznámkách pod čarou na stranách 7 a 40, poněkud matoucí použití “in” pro citace autorů (např. str. 5: “used for function prediction in Bartel et al.” místo výrazně méně invazivního “by Bartel et al.”), občas nevyřešené přetečení (2.1.2, tabulka 3.1), místy pak problémy v anotaci obrázků (podgraf S<sub>1</sub> v obrázku 3.6, (A) až (C) v obrázku 4.7).

## B. Obhajoba

### ***Dotazy k obhajobě***

1. Metriky “*p*-value” jsou většinou používány k vyhodnocení míry platnosti hypotéz. Jaké konkrétní hypotéze, případně jakému statistickému testu, by odpovídala definice *p*-hodnoty použitá v rovnici 5.8? (str. 40) Má smysl porovnávat *p*-hodnoty menší než 10<sup>-10</sup>? (tab. 5.3 a 5.5)
2. Precision, recall, F-measure a ostatní metriky jsou poměrně souhrnné a nepodávají zajímavou informaci o tom, proč (a kde) konkrétní metody fungovaly dobře nebo špatně. Je z dostupných zdrojů možné určit např. užší kategorie nebo příklady proteinů, na kterých algoritmy produkovaly výrazně lepší/horší/jiné výsledky než průměrné? Bylo by pro porovnání kvality algoritmů možné použít nějakou metodu nevyžadující sumarizaci (např. ROC), případně jak?

Stanovisko k opravě chyb v práci:

opravný lístek/oprava v textu **NENÍ** podmínkou přijetí práce

## C. Celkový návrh

Práci **doporučuji** k přijetí k dalšímu řízení: **ANO**

Navrhovaná celková klasifikace: 1 – 2

Datum vypracování posudku: 4.6.2021

Jméno a příjmení, podpis oponenta (SIS):

Miroslav Kratochvíl, RNDr., PhD.