Posudek vedoucího bakalářské práce

Název práce: **Heuristiky v dolování dat z grafů pomocí vnoření uzlů**
Autor: **Adeliia Gataullina**
Vedoucí práce: **Ing. Matej Mojzeš, Ph.D.**

Autorka se v textu bakalářské práce drží cíle, kterým je studie využití heuristik při ladění nastavení algoritmu node2vec, tedy reprezentace uzlů grafu.

Základní pojmy jsou definovány v první kapitole spolu s rešerší algoritmu word2vec, ze kterého node2vec vychází. V druhé kapitole se autorka věnuje implementaci algoritmu node2vec. Tady si pochvalu zaslouží použití úpravy původní implementace, díky které je implementace node2vec schopna konzistentně reprodukovat výsledky na nezměněných datech – vlastnost nezbytná pro další experimenty, resp. výzkum obecně. Ve třetí kapitole je definována účelová funkce, popsány vybrané heuristiky a jejich výkon změřen na testovacích datech. Nastavení s nejlepšími výsledky je posléze spuštěno na větší datové sadě a výsledky diskutovány, zejména z hlediska kvality výsledku s použitím heuristik. Dle závěru autorky použití heuristik v tomto případě přináší pozitivní výsledky.

Pochvalu zaslouží text práce v anglickém jazyce a přiložený zdrojový kód v jazyku Python. Na druhé straně, výtku bych měl k nedůsledné matematické typografii a místy nejasné, či chybějící citace. V textu práce nejsou jednotlivé pojmy vždy dostatečně formálně popsány, v pseudokódu se vyskytují nedefinované funkce, resp. pro jejich pochopení je nutná znalost např. knihovny NumPy. Čitelnosti práce neprospívají také konstanty použité přímo v textu místo tradičních proměnných, ideálně s vysvětlením motivace pro jejich vybrané hodnoty.

K vysvětlení při obhajobě bych doporučoval následující otázky:

1. Z tabulek v kapitole 4.5 plyne, že s vhodně nastavenou heuristikou Shoot & Go je možné vylepšit hodnotu účelové funkce (kvalitu shlukové analýzy) oproti referenční metodě. Došlo při experimentech i k opačnému jevu, tj. že nejlepší řešení nalezené heuristickým prohledáváním bylo horší? Jestli ano, za jakých podmínek?
2. Jaký by byl postup k rozšíření množiny heuristicky optimalizovaných parametrů algoritmu node2vec? Které parametry byste doporučovala jako vhodné kandidáty k optimalizaci?

Závěrem mohu potvrdit, že cíl práce byl naplněn a předložená práce splňuje jak obsahové, tak formální požadavky bakalářské práce a navrhuji hodnocení známkou **C (dobře)**.

V Praze dne 17. 8. 2020 Ing. Matej Mojzeš, Ph.D.