Univerzita Karlova

Přírodovědecká fakulta

Řešení problému obchodního cestujícího

Geoinformatika

Lucie Peterková

1. ročník, N-GKDPZ

Praha 2022

Zadání

Vstup: množina uzlů U reprezentujících body.

Výstup: nalezení nejkratší Hamiltonovské kružnice mezi těmito uzly.

Nad množinou U nalezněte nejkratší cestu, která vychází z libovolného uzlu, každý z uzlů navštíví pouze jedenkrát, a vrací se do uzlu výchozího. Využijte níže uvedené metody konstrukčních heuristik:

• Nearest Neighbor,

• Best Insertion.

Výsledky porovnejte s výstupem poskytovaným nástrojem Network Analysis v SW ArcMap.

Otestování proveďte nad dvěma zvolenými datasety, které by měly obsahovat alespoň 100 uzlů. jako vstup použijte existující geografická data (např. města v ČR s více než 10 000 obyvateli, evropská letiště, ...), ohodnocení hran bude představovat vzdálenost mezi uzly (popř. vzdálenost měřenou po silnici); pro tyto účely použijte vhodný GIS.

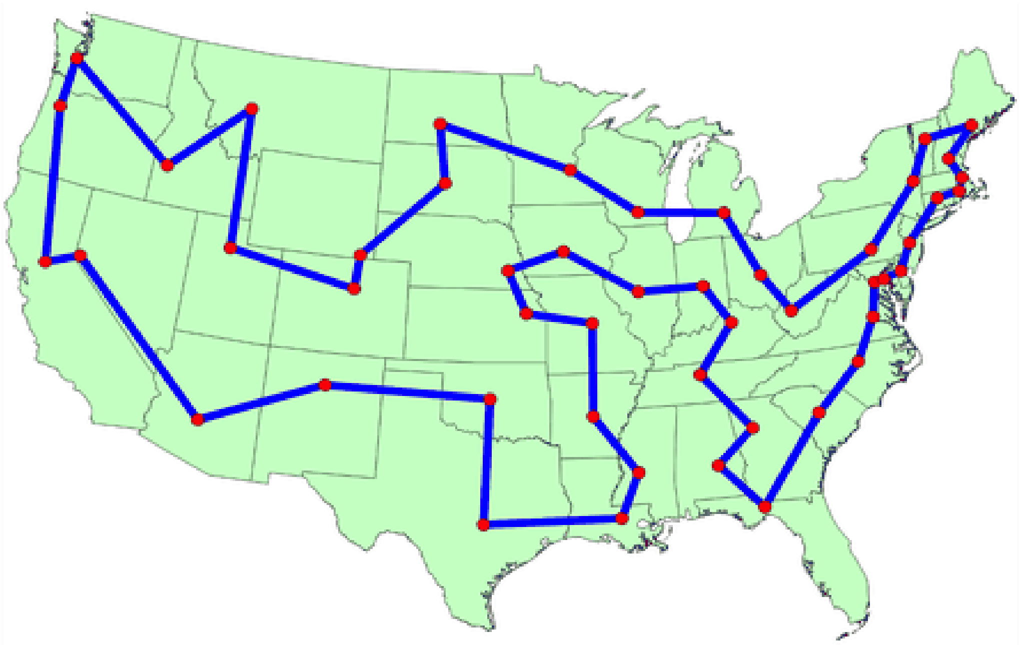
Výsledky s uvedením hodnot W,k, uspořádejte do přehledné tabulky (obě metody nechte proběhnout alespoň 10x), a zhodnoťte je.

Pro implementaci obou konstrukčních heuristik použijte programovací jazyk Python, vizualizaci výstupů proveďte ve vhodné knihovně, např. matplotlib.

Popis problému

TSP

Problém obchodního cestujícího (v angličtině Travelling Salesman Problem, dále TSP) patří mezi tzv. NP–úplné úlohy, tedy úlohy, u kterých nejsme schopni efektivně nalézt exaktní řešení. Řešení, které nalézáme, je často pouze přibližné a jehož kvalitu považujeme za akceptovatelnou (Bayer 2022). Během řešení TSP (obr. 1) z hlediska terminologie teorie grafů hledáme co nejkratší Hamiltonovskou kružnici Kh pro Hamiltonovský graf Gh. Obecný popis problému může být následovný: Obchodní cestující chce navštívit n různých měst a poté se vrátit domů. Známe vzdálenosti mezi jednotlivými městy. Cestující chce stanovit pořadí cesty, tak aby cestovní vzdálenost byla minimalizovaná, když navštíví každé město pouze jedno a výchozí a koncový bod cesty je ve stejném městě. Náročnost výpočtu se enormně optimální cesty se zvyšuje s celkovým počtem bodů (tab. 1) (Pokorná 2008).



Obr. 1: TSP řešení pro státy USA.
Zdroj: https://optimization.mccormick.northwestern.edu/

n	Počet různých cest
3	1
5	12
7	360
10	181440
15	43589145600
20	60822550204416000
25	310224200866620000000000
30	4420880996869850000000000000000
35	147616399519802000000000000000000000000
40	10198941040598700000000000000000000000000000000
45	1329135787394220000000000000000000000000000000000000000
50	304140932017134000000000000000000000000000000000000000000000000

Tab. 1: Počet uzlů (n) x počet různých cest.
Zdroj: Pokorná (2008)

Nearest Neighbor (NN)

Princip metody NN je následující:

1. Vybereme počáteční město náhodně nebo dle libosti.
2. Procházíme zbývající města a vždy přidáme nejbližší město k aktuální trase.
3. Opakujeme krok 2, dokud nenavštívíme všechna města.
4. Přidáme vzdálenost mezi posledním navštíveným městem a počátečním městem k celkové vzdálenosti a přidáme počáteční město do trasy.

Metoda NN je rychlá a jednoduchá na implementaci. Její nevýhodou je, že nemusí vést k optimálnímu řešení a její výsledky mohou být od optimálního řešení vzdáleny i o více než 50 %, jelikož algoritmus „neuznává“ možné důsledky pouhého zaměření na jedno město. Toto řešení umožňuje rychlý průběh algoritmu, ale snižuje se optimálnost konečné délky trasy (Bryan 2021).

Best Insertion (BI)

Jedná se o rozšíření metody Nearest Neighbor. BI funguje tak, že nejprve vybere některé město jako začátek trasy a přidá ho do seznamu navštívených měst. Poté zbývající města postupně vkládá do trasy tak, aby se co nejvíce zkrátila celková délka trasy. K tomu se pro každé zbývající město zkoumají všechna místa, kam by se mohlo vložit, a vybere se takové místo, kde dojde k největšímu zkrácení trasy. Tento postup se opakuje, dokud nejsou všechna města v trase.

Řešení problému

Problém obchodního cestujícího je řešen v jazyce Python. Jako vývojové prostředí byl zvolen software Visual Studio Code. Výsledky metod NN a BI jsou porovnány s výsledky ze softwaru ArcGIS Pro. Předpokladem je získání nejkratší cesty ze softwaru ArcGIS Pro, následuje Best Insertion a Neirest Neighbor.

Vstupní datasety

- orp.txt (orp.shp): obce s rozšířenou působností v Libereckém, Královehradeckém, Ústeckém, Středočeském (+ Praha), Karlovarském, Plzeňském a Jihočeském kraji. Celkem se jedná o 102 obcí.
- Zeleznice.txt (zeleznice.shp): vybrané železniční zastávky ve Slezsku a na Moravě. Celkem se jedná o 100 železničních zastávek.

Data byla získána z digitální vektorové databáze České republiky ArcČR 500. Data byla vybírána tak, aby byla zachována podmínka minimálně 100 uzlů.

Pro správné fungování scriptu bylo nutné naimportovat knihovny csv, random a matplotlib. Na začátku scriptu byla načtena data ve formátu .txt se souřadnicemi bosů a byl vytvořen prázdný seznam pro vkládání souřadnic.

Nearest Neighbor

Vstupem Funkce je textový soubor se souřadnicemi. Dále metoda funguje následovně:

1. Vybere náhodné město jako začátek trasy a přidá jej do seznamu navštívených měst.
2. Pro každé zbývající město zkoumá vzdálenost k ostatním městům a vybírá nejbližší nenavštívené město. Přidá toto město do seznamu navštívených měst a současně do celkové délky trasy připočítá vzdálenost od předchozího města.
3. Opakuje krok 2, dokud nejsou navštívena všechna města.
4. Přidá do seznamu navštívených měst začáteční město a do celkové délky trasy připočítá vzdálenost od posledního navštíveného města.
5. Vykreslí graf s body reprezentujícími města a spojnicemi reprezentujícími trasu.
6. Vrátí celkovou délku trasy jako výsledek.

Best Insertion

Pro účely této funkce je nejdříve definována funkce vracející vzdálenost mezi dvěma body.

Vstupem funkce Best Insertion je textový soubor se souřadnicemi. Dále metoda funguje následovně:

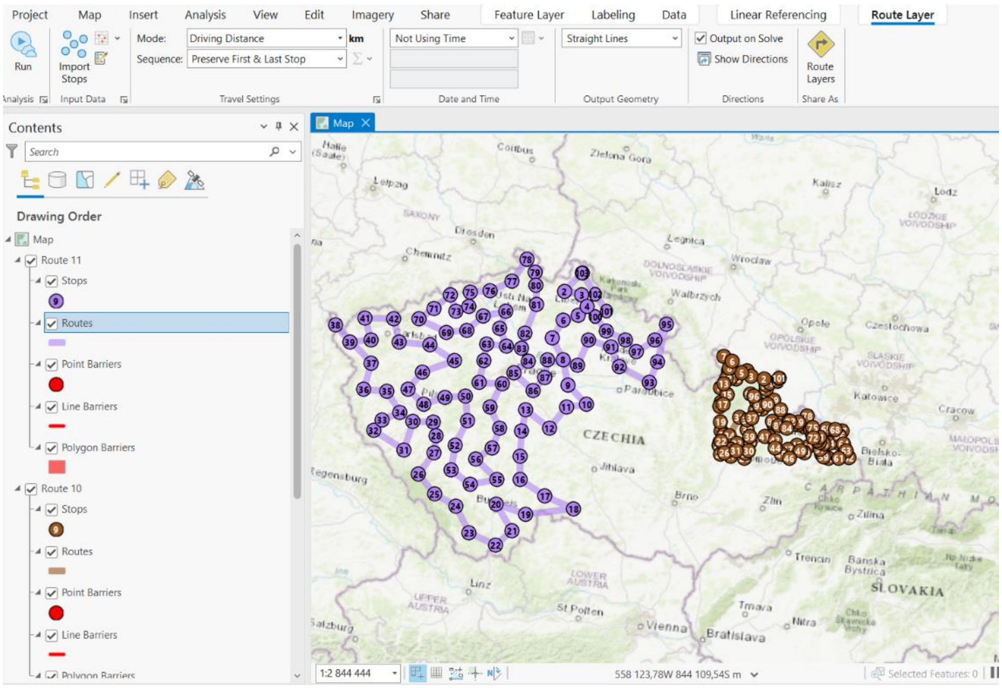
1. Vybere náhodně město jako začátek cesty a uloží jej do proměnné start.
2. Vybere druhé město náhodně zbývajících měst a přidá jej do seznamu path.
3. Nastaví celkovou vzdálenost na vzdálenost mezi prvním a druhým městem.
4. Projde seznam souřadnic měst pomocí cyklu for.
5. Pokud je souřadnice města rovna souřadnici města start, přeskočí toto město.
6. Vybere nejlepší místo pro vložení města do cesty tak, aby se minimalizovala celková vzdálenost.

7. Pro každé místo v cestě pomocí cyklu for zkontroluje, zda vložení města v daném místě způsobí nejmenší celkovou vzdálenost.
8. Pokud vložení města v daném místě způsobí nejmenší celkovou vzdálenost, aktualizuje nejmenší celkovou vzdálenost a nejlepší index pro vložení města.
9. Vloží město do cesty na nejlepší index a aktualizuje celkovou vzdálenost.
10. Po dokončení procházení seznamu měst přidá město, kde cesta začala, do cesty a aktualizuje celkovou vzdálenost o vzdálenost mezi posledním a předposledním městem v cestě.
11. Vypíše graf znázorňující všechna města a cestu mezi nimi.
12. Vrátí celkovou vzdálenost cesty

Nakonec program vypíše vypočítanou celkovou vzdálenost mezi všemi body v metrech.

Network Analysis v ArcGIS Pro

V softwaru ArcGIS Pro byla pomocí nástroje Network Analysis a funkce Route spočítána nejkratší cesta. Jelikož software počítá pouze Hamiltonovskou cestu, bylo nutné vstupní dataset upravit. Aby výchozí bod odpovídal koncovému, bylo nutné původní bod zkopírovat a vložit jako nový do datasetu. Jako startovní a cílový bod byl zvolen bod na okraji Hamiltonovské cesty, která se zjistila pomocí Network Analysys a funkce Route, kde byla zvolena sekvence Find Best. Startovním a cílovým bodem pro dataset ORP byl zvolen Frýdlant a pro dataset Železnice Osoblaha. Následně byla analýza provedena znovu, ale se zvolenou frekvencí Preserve First And Last Stop. Jelikož součástí datasetů nejsou součástí cesty mezi body, byla nastavena výstupní geometrie na Straight Lines, čím se body propojí vzdušnou čarou. Nastavené parametry pro výpočet Hamiltonovské cesty jsou zobrazeny na obr. 2.



Obr. 2: ArcGIS Pro Network Route Analysis.
Zdroj: vlastní zpracování

Testování výsledků

Jednou z metod testování přesnosti je poměr (k) nalezené délky W Hamiltonovské kružnice vůči optimu Wo, který ukazuje, ke kolikanásobnému prodloužení délky Hamiltonovské kružnice zvolené řešení vede.

$$k = \frac{W}{Wo}$$

Dle zadání byl jako optimální výsledek určen výstup z Network Analysis, který by měl mít přibližně o 10 % lepší výsledek než výstupy z Pythonu.

Výsledky

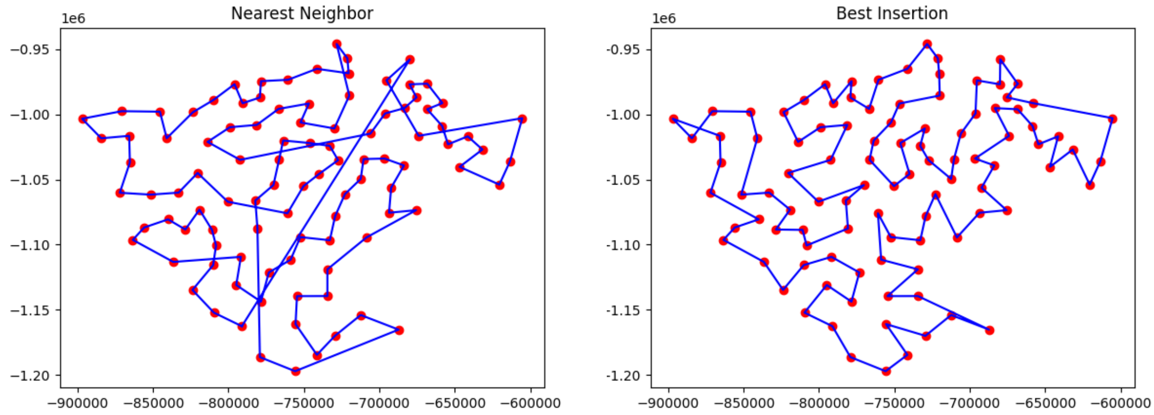
Každá z metod byla v Pythonu spuštěná 10x. Tab. 2 představuje jednotlivé výsledky s porovnáním s výsledkem z Network Analysis. Z tabulky je možné vyčíst mírně lepší výsledek u všech hodnot pro metodu BI oproti metodě NN u datasetu železnice. Pro dataset orp je u větší část výstupů lepší výsledek lepší výsledek pro metodu BI, nicméně objevují se i lepší hodnoty pro metodu NN. Tab. 3 nabízí porovnání nejlepších výsledků (tedy nejnižších délek Hamiltonovské kružice) z 10 výpočtů pro každou z metod. Pro oba datasety nabízí nejnižší délku metoda BI. Obr. 3 a 4 vizualizuje nejkratší a nejdelší cestu pro dataset orp z metod z Pythonu. Obr. 5 a 6. Nejhoršího výsledku pro oba datasety dosáhla metoda Network Analysis. Vizualizaci zobrazuje obr. 7.

	Nearest Neighbor				Best Insertion			
	železnice W (m)	k (%)	orp W(m)	k (%)	železnice W (m)	k (%)	orp W(m)	k (%)
1	825002,056	91,1	2591663,814	91,2	749593,622	82,7	2290359,811	80,6
2	769824,848	85,0	2359856,062	83,0	716147,4	79,1	2258960,605	79,5
3	776506,379	85,7	2293895,356	80,7	711700,785	78,6	2430985,199	85,5
4	774243,606	85,5	2395413,027	84,3	760255,502	83,9	2315706,305	81,5
5	837043,325	92,4	2391167,891	84,1	690807,105	76,3	2403785,224	84,6
6	785517,874	86,7	2415467,117	85,0	780924,292	86,2	2349233,704	82,7
7	809920,549	89,4	2561077,405	90,1	723469,477	79,9	2469001,455	86,9
8	772313,654	85,3	2487227,864	87,5	762654,348	84,2	2379268,469	83,7
9	849329,36	93,8	2439628,237	85,8	756412,214	83,5	2509166,559	88,3
10	776506,379	85,7	2454079,535	86,4	766372,584	84,6	2390899,627	84,1
průměr	797620,803	88,1	2438947,631	85,8	741833,7329	81,9	2379736,696	83,7

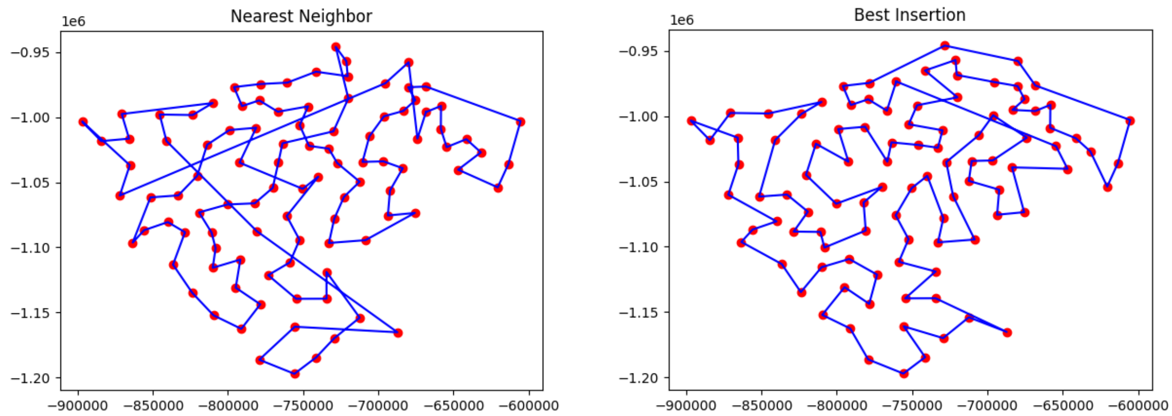
Tab. 2: Výsledky Nearest Neighbor a Best Insertion.
Zdroj: vlastní zpracování

	orp		železnice	
	W (km)	k (%)	W (km)	k (%)
Nearest Neighbor	2293,895	80,7	769,824	85,0
Best Insertion	2258,960	79,5	690,807	76,3
Network Analyst	2841,886	100,0	905,856	100,0

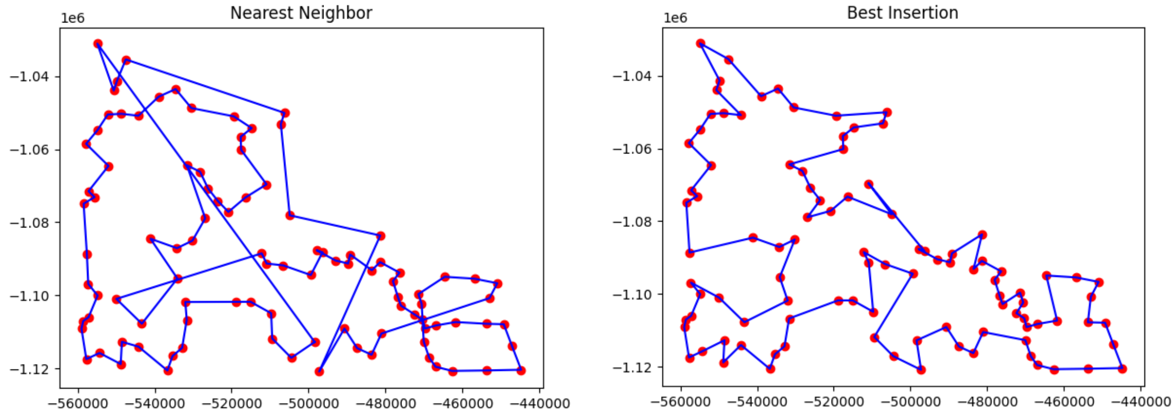
Tab.3: Porovnání nejlepších výsledků.
Zdroj: vlastní zpracování



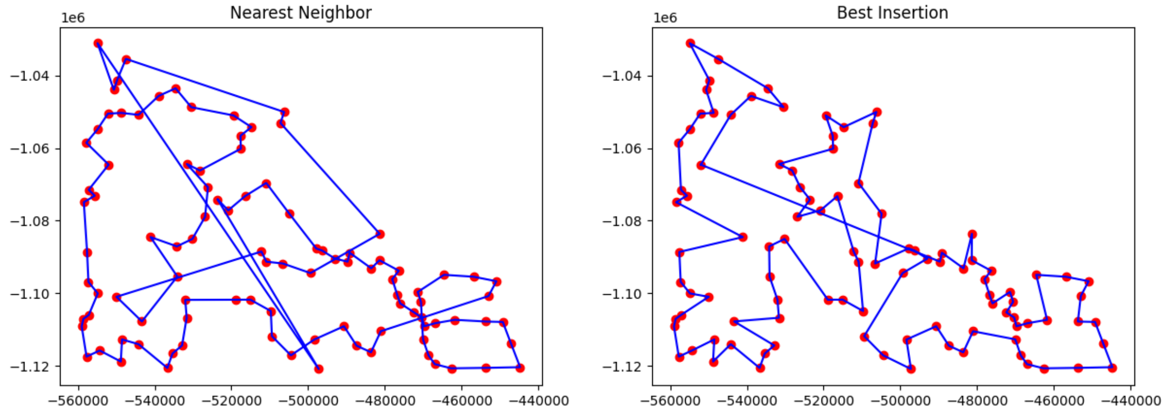
Obr. 3: Nejlepší výsledky pro dataset orp.
Zdroj: vlastní zpracování



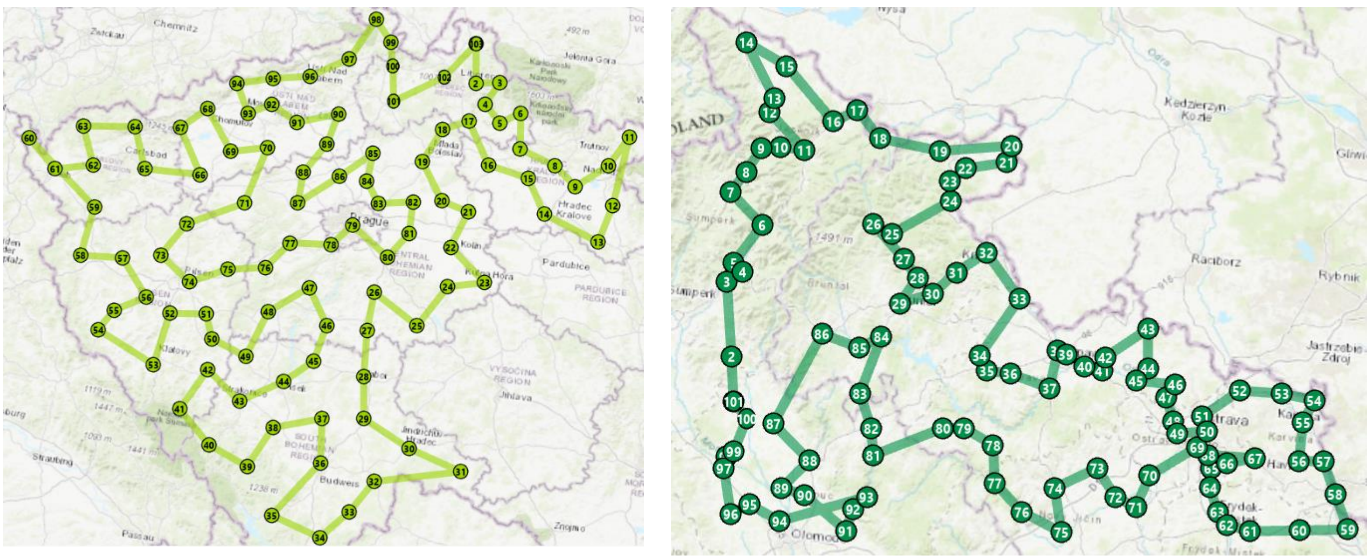
Obr. 4: Nejhorší výsledek pro dataset orp.
Zdroj: vlastní zpracování



Obr. 5: Nejlepší výsledek pro dataset železnice.
Zdroj: vlastní zpracování



Obr. 6: Nejhorší výsledky pro dataset železnice.
Zdroj: vlastní zpracování



Obr. 7: Network Analysis v ArcGIS Pro. Vlevo: dataset orp. Vpravo: dataset: železnice.
Zdroj: vlastní zpracování

Závěr

Dle očekávání dosáhla metoda Best Insertion lepších výsledků než metoda Nearest Neighbor. Nicméně rozdíl o tolik větší nebyl. Je možné, že po provedení více než 10 pokusů se výsledky ustálí s větším rozdílem. Pro dataset železnice nabízí metoda v průměru o 14 % lepší výsledek než metoda Nearest Neighbor. Pro dataset orp se jedná o lepší výsledek pouze o 4 % (z 10 výsledků).

Velkým překvapením je výrazně nejhorší výsledek pro metodu Network Analysis. Problém může být ve spočítání vzdáleností mezi jednotlivými body, kde v této práci do analýzy vstupovaly pouze body a software v rámci analýzy spočítal vzdálenosti vzdušnou čarou mezi nimi. Další možností získání cest mezi body by bylo vytvoření liniové vrstvy, kde každý bod bude propojen se všemi zbývajícími. Do analýzy by nevstupovaly pouze body, ale již i vytvořené cesty, kde by v rámci analýzy byly vybrány ty nejvhodnější.

Zpřesnění výsledků nabízí i úprava vytvořeného kódu, kde by bylo např. možné přesněji počítat vzdálenosti mezi jednotlivými body. Další možností je i rozšíření funkcí. Například u metody Best Insertion by bylo možné metodu derandomizovat a vybrat v každém kroku takový uzel u, který minimalizuje hodnotu Δw, což by ale způsobilo výrazný vzrůst výpočetní náročnosti.

Zdroje

Literatura:
BAYER, T. (2022): Problém obchodního cestujícího, konstrukční heuristiky: stručný návod na cvičení. https://agony.natur.cuni.cz/~bayertom/images/courses/Geoinf/tsp_uloha.pdf

BRYAN, K. (2021): Different Heuristic Approaches to Solving Traveling Salesman Problem. International Journal of Social Sciences and Humanities Research, 5, 1,1–16.

POKORNÁ, P. (2008): Problém obchodního cestujícího pomocí metody Mravenčí kolonie. Bakalářská práce. Univerzita Pardubice, fakulta Ekonomicko-správní.

Data:
ARCČR (2020). Vektorová databáze České republiky ARCČR 500.

Obrázky:
Obr. 1: https://optimization.mccormick.northwestern.edu/index.php/File:48StatesTSP.png