

Univerzita Karlova
Přírodovědecká fakulta



GEOINFORMATIKA

Problém obchodního cestujícího

Bc. Markéta Růžičková
Geoinformatika, kartografie a DPZ, 1. ročník
leden 2023

Zadání:

Úloha: Řešení problému obchodního cestujícího

Vstup: množina uzlů U reprezentujících body.

Výstup: nalezení nejkratší Hamiltonovské kružnice mezi těmito body. Nad množinou U naleznete 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 Analyst v SW ArcMap. Otestování proveďte nad dvěma zvolenými dataseťmi, které by měly obsahovat alespoň 100 uzlů. Jako vstup použijte existující geografická data (např. města v ČR s více než 100 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 10), a zhodnoťte je.

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

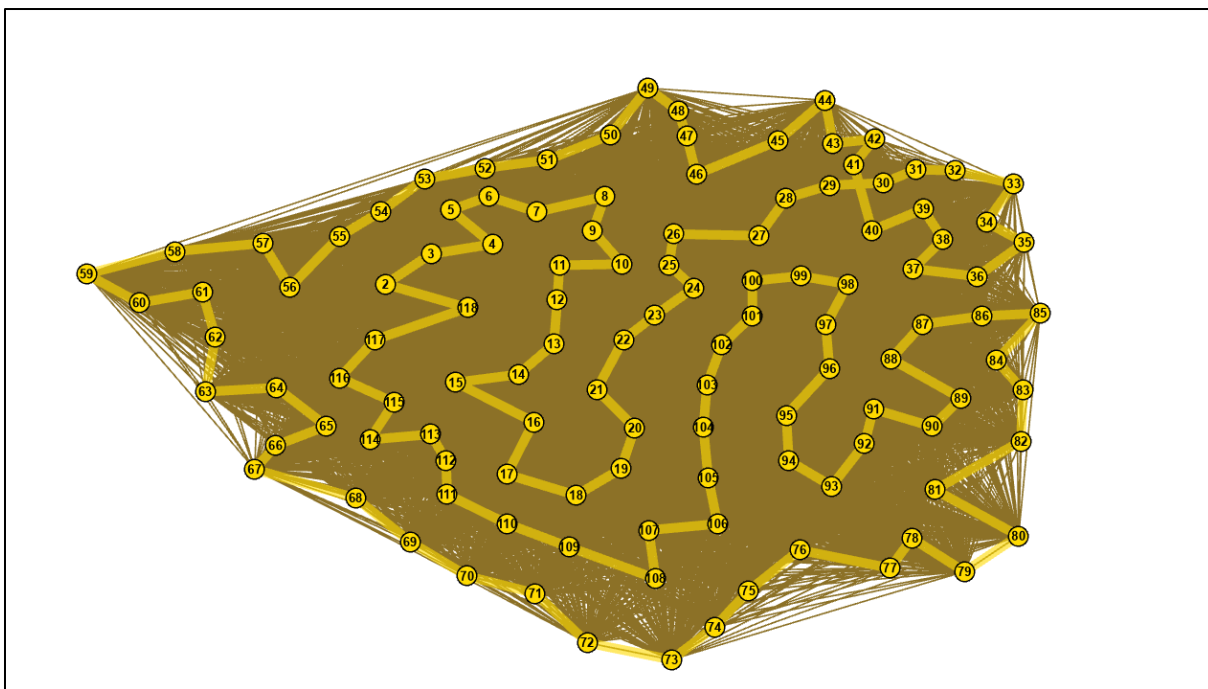
Zpracování:

Jako vstupní data pro tuto úlohu byla využita data z databáze ArcČR, konkrétně dataset (téměř) všech ORP v Čechách (`orp_sourad`) se 117 uzly, a dataset obcí v okrese Benešov (`benesov_sourad`) se 114 uzly.

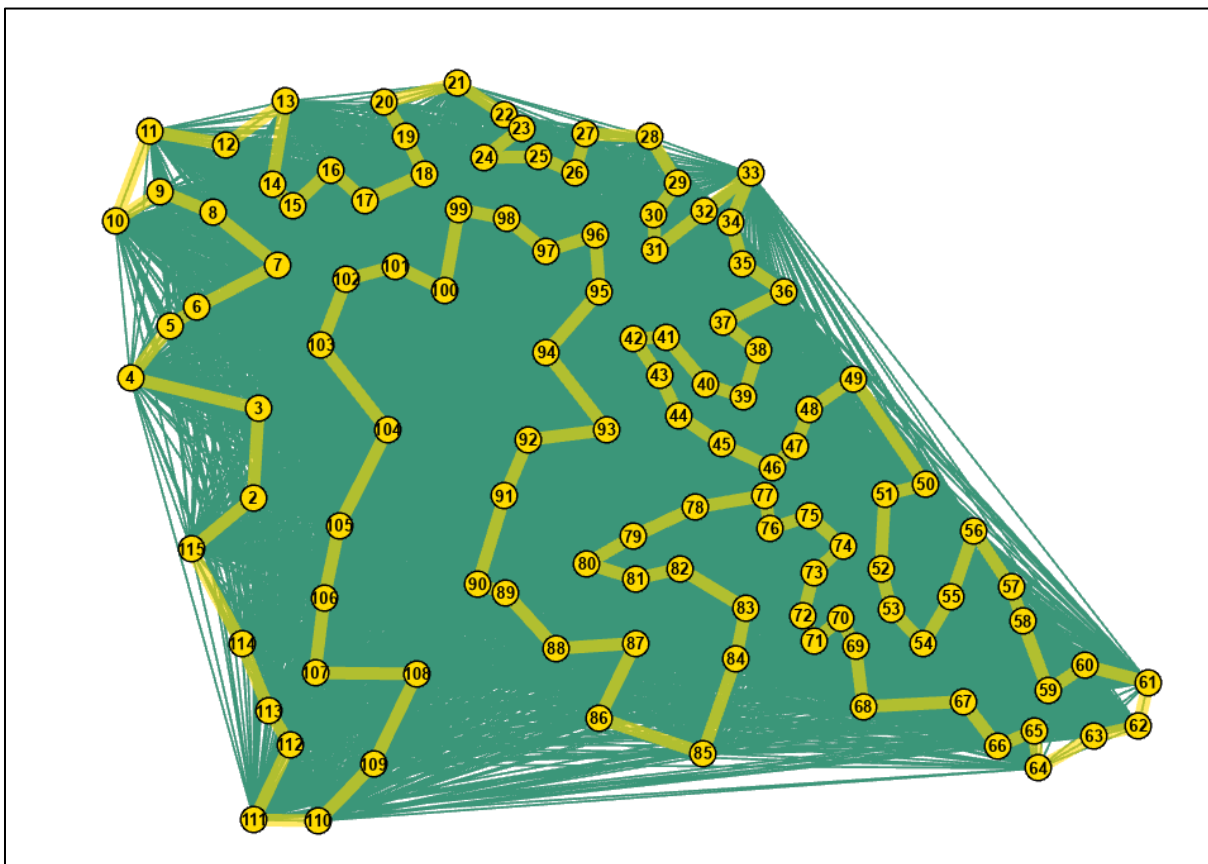
ArcGIS Pro

Nejprve byla nalezena optimální trasa pomocí SW ArcGIS Pro. Do SW byla nahrány bodové dataseťmi obcí, dále bylo potřeba vygenerovat hrany mezi nimi. Byly generovány veškeré existující hrany mezi uzly. Nejprve byla vytvořena tabulka souřadnic všech dvojic uzlů (*Generate Near Table*), dále byl vytvořen linie spojující všechny uzly pomocí funkce *XY To Line*. Tato liniová vrstva byla vložena do samostatného Feature datasetu. Dále byly využity funkce *Create Network Dataset* a *Build Network*. Jakožto Stops (zastávky) byly využity bodové vrstvy obcí. Síťová analýza *Route* byla spuštěna nejprve s parametrem *Find Best*, který najde optimální výchozí bod. Aby bylo možné spočítat celou Hamiltonovskou kružnici (a nikoli jen Hamiltonovskou cestu), bylo zapotřebí vytvořit jeden další bod na stejné poloze, kde leží právě nalezený optimální výchozí bod, a tomuto nově nalezenému bodu přepsat atribut *Sequence* tak, aby se stal posledním uzlem v pořadí. Následně byla *Route* spuštěna ještě jednou, ale s parametrem *Preserve First and Last Stop*. Výsledkem této analýzy bude již kružnice grafu.

Výstupy ze SW ukazují Obrázek 1 a Obrázek 2.



Obrázek 1: Network Analysis – Route v SW ArcGIS Pro, ORP v Čechách



Obrázek 2: Network Analysis – Route v SW ArcGIS Pro, obce okresu Benešov

Nearest Neighbour

Konstrukční heuristika Nearest Neighbour (NN) pracuje následujícím způsobem: Všechny uzly zpočátku nastaví jako otevřené. Počáteční uzel je buď vybrán náhodně, nebo jeho volbu zadá uživatel. Tento uzel se nastaví jako uzavřený. Následně se z okolních otevřených uzlů vybírá uzel u , který je zkoumanému uzlu u_i nejbližší, tj. ten který minimalizuje přírůstek ohodnocení celé cesty Δw , a zároveň ještě není součástí Hamiltonovské kružnice K_h .

$$\Delta w = \min w(u_i, u) \wedge u \notin K_h$$

Uzel u se přidá do kružnice a její délka W (která je na počátku algoritmu nastavena na 0) se prodlouží o vzdálenost mezi zkoumaným uzlem a jeho nejbližším sousedem.

$$W = W + \Delta w$$

Tento nalezený soused se označí jako uzavřený a jako jeho předchůdce se označí zkoumaný uzel. Kružnici je po proběhnutí algoritmu možné zrekonstruovat pomocí seznamu předchůdců uzlů. Metoda generuje tolik řešení, kolik je uzlů.

V tomto cvičení byla metoda NN zpracována tímto způsobem:

Pseudokód NN:

Založ prázdný seznam předchůdci

Založ prázdný seznam delky

Pro každý uzel pu v seznamu uzlů:

 všechny předchůdce P *nastav* na -1

 Všechny stavy S *nastav* na 'O'

Přiřad' $u = pu$

Nastav $S[u] = 'C'$

Inicializuj $delkaKruznice = 0$

Pro každý uzel kromě počátečního:

Nastav $dk = \inf$

Pro každý uzel $v0$:

Pokud $S[v0] == 'O'$ A ZÁROVEŇ $dk > \text{vzdálenost } u \text{ od } v0$:

$dk = \text{vzdálenost } u \text{ a } v0$

$v0$ *ulož* do nové proměnné v

Nastav $S[v] = 'C'$

Nastav $P[v] = u$

K *delkaKruznice* *přičti* dk

Přiřad' $u = \text{nově nalezený uzel } v$

Nastav $P[pu] = v$

 Do delky *přidej* ($delkaKruznice + \text{vzdálenost } pu \text{ a } v$)

 Do předchůdci *přidej* seznam nalezených předchůdců P

Vypiš minimum z delky

Vykresli

Algoritmus tedy pracuje tak, že postupně bere všechny uzly ze seznamu jako počáteční, a vytvoří tolik kružnic, kolik je zadáno uzlů, a z délek těchto kružnic je vybrána ta nejkratší. Tím je zajištěno, že tento algoritmus nám pro jakýkoli dataset vydá nejlepší řešení, jakého je metoda NN schopna. Z tohoto důvodu nebyla funkce testována 10x, ale pouze jednou (všechny výsledky by byly totožné). Funkce vrací seznam délek, seznam předchůdců. V seznamu délek najde minimální délku a tu vykreslí.

Best insertion

Metoda Best Insertion je účinnější než NN. Na počátku jsou všechny uzly označené jako otevřené. Následně jsou vybrány tři uzly jakožto výchozí kružnice, a tyto uzly se označí jako uzavřené, spočítá se délka této kružnice W . Z ostatních otevřených uzlů jsou postupně náhodně vybírány všechny ostatní uzly. Vybraný uzel u je potřeba přidat do kružnice, ale na takovou pozici (napojit do takové hrany), aby výsledná délka kružnice byla co nejkratší, tzn. aby vzniklá cesta u_i, u, u_{i+1} přes uzel u co nejméně prodloužila stávající kružnici.

$$\Delta w = \min w(u_i, u) + w(u, u_{i+1}) - w(u_i, u_{i+1}) \wedge u \notin K_h$$

Tento uzel se následně označí jako uzavřený a délka kružnice W je aktualizována.

$$W = W + \Delta w$$

Tímto způsobem se postupně přiřadí všechny uzly grafu. Opakované spuštění této metody dává různé výsledky, vzhledem k náhodnému výběru bodů.

V tomto cvičení byla metoda Best Insertion zpracována tímto způsobem:

Pseudokód Best Insertion:

Založ výchozí kružnici kružnice z prvních tří bodů seznamu uzlů

Spočítej délku kružnice

Stavy S všech uzlů *nastav* jako 'O'

Stavy S uzlů v kružnice *nastav* jako 'C'

Založ frontu o délce seznamu uzlů mínus 3

Dokud fronta není prázdná:

Nastav $dk_bi = \inf$

Vygeneruj náhodný uzel $rand$

Pokud $S[rand] = 'O'$:

Nastav $S[rand] = 'C'$

Pro každou stranu kružnice:

Pokud $dk_bi > \text{délka kružnice vzniklé napojením uzlu } rand \text{ k této straně}$:

Nastav $dk_bi = \text{délka kružnice vzniklé napojením uzlu } rand \text{ k této straně}$

 Pozici, kam má být do kružnice vložen uzel, *ulož* do nové proměnné

 Do kružnice *vlož* tento uzel

Nastav $delka_bi = dk_bi$

 Z fronty *odeber* jeden prvek

Vykresli

V tomto řešení Best Insertion jsou jakožto výchozí kružnice zadány tři první body ze seznamu uzlů. Další body jsou pak vybírány náhodně, čímž je zajištěno nedeterministické chování algoritmu.

Testování výsledků

Jakožto optimální řešení pro tuto úlohu budeme považovat výsledky ze SW ArcGIS Pro.

Poměr k můžeme vypočítat jako podíl délky kružnice nalezené jinou metodou, a optimální délky kružnice.

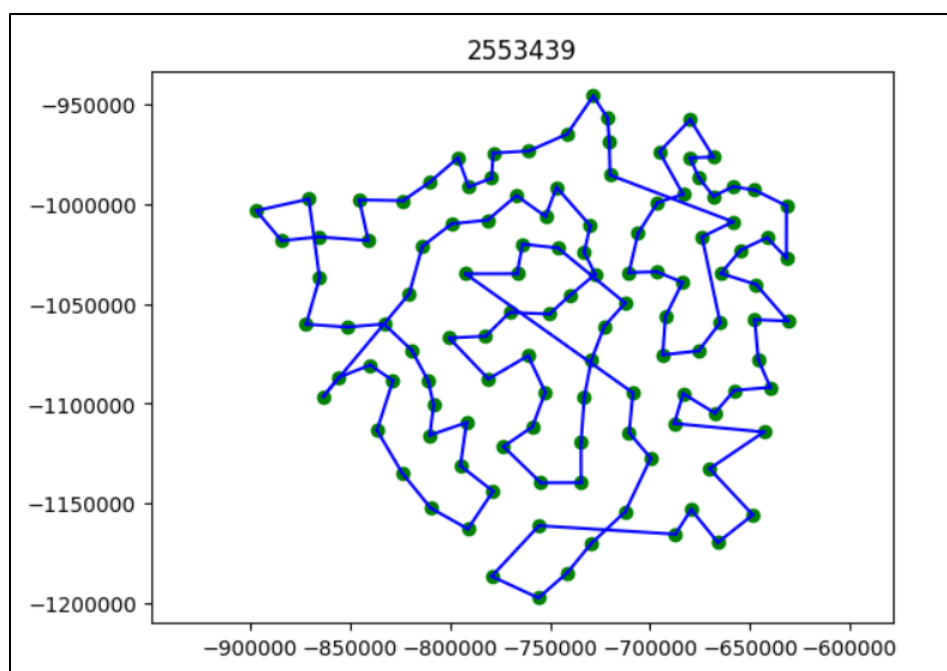
Vzhledem k tomu, že vzdálenosti mezi uzly byly počítány v metrech, byly výsledky zaokrouhleny na celá čísla. Zachování desetinných míst v tomto případě nedává smysl.

NN

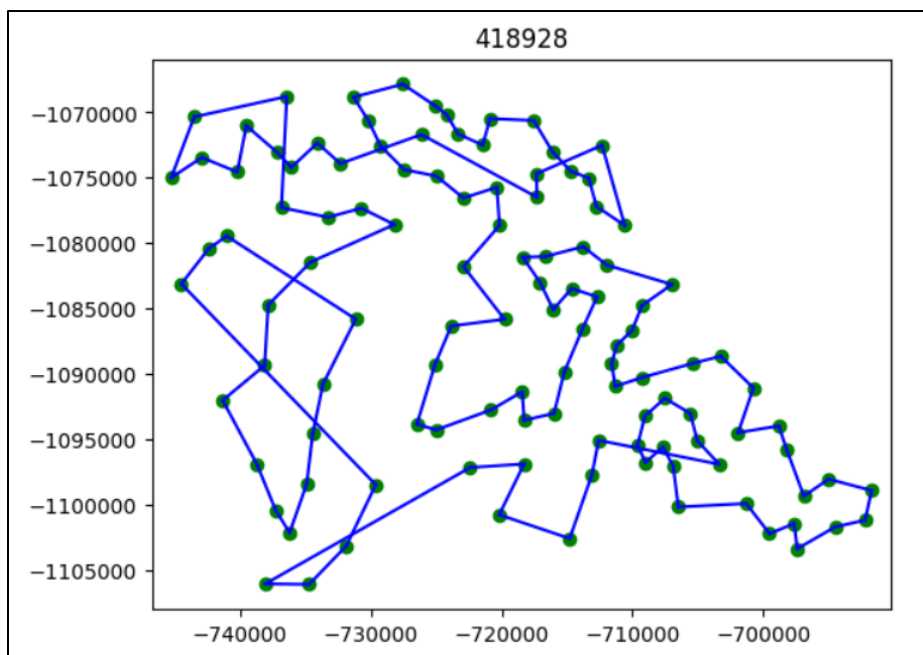
Pro metodu NN, jak již bylo řečeno, bylo nalezeno vždy nejoptimálnější řešení ze všech možných, výsledky v Tabulce 1. Na Obrázku 3 a 4 jsou vykresleny nalezené kružnice.

Tabulka 1: Výsledky analýzy NN a ArcGIS Pro analýzy pro oba datasety

Dataset	NN (m)	ArcGIS Pro (m)	k (%)
ORP v Čechách	2553439	2280407	11,97
Obce Benešov	418928	362229	15,65



Obrázek 3: Kružnice nalezená metodou NN, dataset ORP v Čechách



Obrázek 4: Kružnice nalezená metodou NN, dataset obcí okresu Benešov

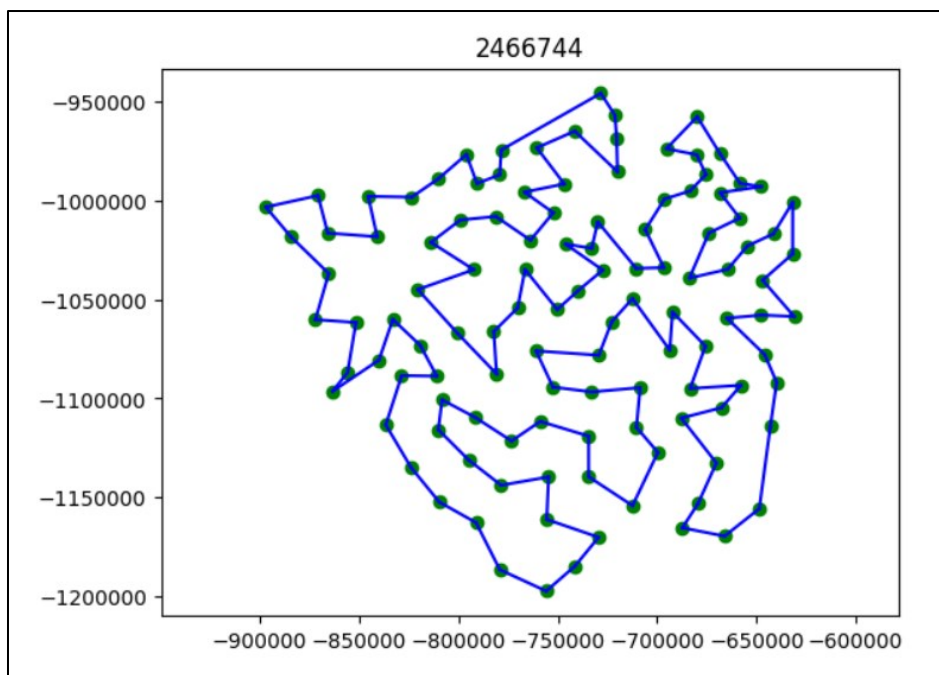
Best Insertion

V tomto případě byla metoda proběhnuta 10x, výsledky jsou zaznamenány v Tabulkách 2 a 3.

Tabulka 2: Výsledky analýzy Best Insertion a ArcGIS pro analýzy pro dataset ORP

Pokus č.	BI (m)	ArcGIS Pro (m)	k (%)
1	2536702	2280407	11,24
2	2478639	2280407	8,69
3	2457028	2280407	7,75
4	2507959	2280407	9,98
5	2484823	2280407	8,96
6	2484969	2280407	8,97
7	2489981	2280407	9,19
8	2480521	2280407	8,78
9	2469363	2280407	8,29
10	2524526	2280407	10,71
průměr	2491451	2280407	9,25

Nejkratší délka kružnice, která v rámci 10 pokusů této analýzy vyšla, byla 2 457 028 m, s prodloužením oproti optimu o 7,75 % (Obrázek 5). Průměrné prodloužení bylo o 9,25 %.

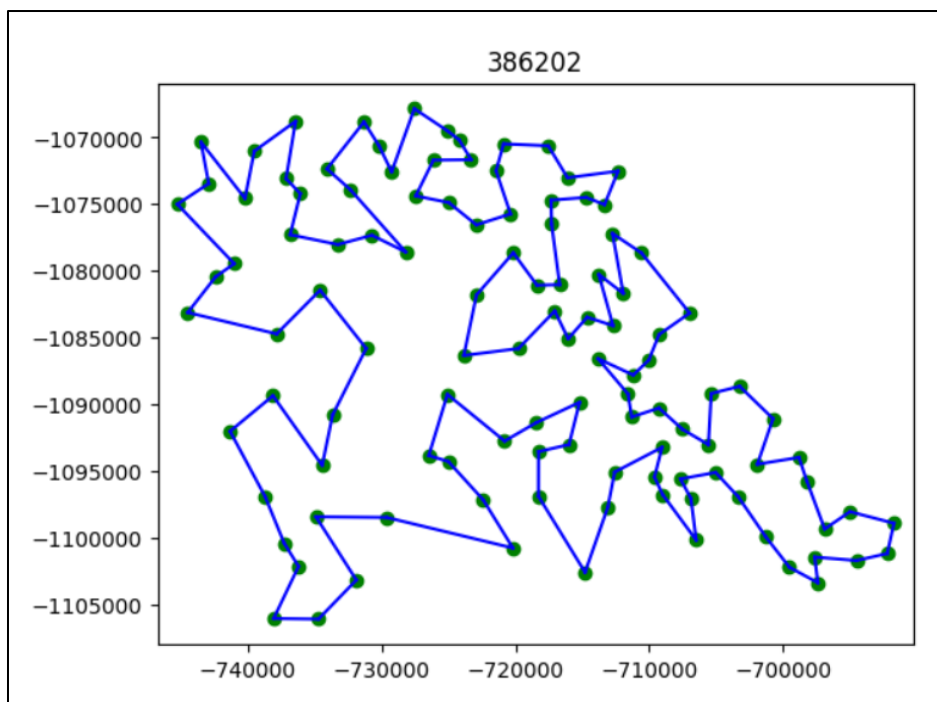


Obrázek 5: Kružnice nalezená metodou Best Insertion, dataset ORP v Čechách

Tabulka 3: Výsledky analýzy Best Insertion a ArcGIS Pro analýzy pro dataset obcí okresu Benešov

Pokus č.	BI (m)	ArcGIS Pro (m)	k (%)
1	395538	362229	9,20
2	381265	362229	5,26
3	393195	362229	8,55
4	383148	362229	5,78
5	384384	362229	6,12
6	396332	362229	9,41
7	395686	362229	9,24
8	390925	362229	7,92
9	378310	362229	4,44
10	390455	362229	7,79
průměr	388934	362229	7,37

Nejkratší délka kružnice, která v rámci 10 pokusů této analýzy vyšla, bylo 378 310 m, s prodloužením oproti optimu o 4,44 %. Průměrně byla trasa prodloužena o 3,37 %.



Obrázek 6: Kružnice nalezená metodou Best Insertion, dataset obcí okresu Benešov

Hodnocení

Podle očekávání byla neoptimálnější (nejkratší) Hamiltonovská kružnice nalezena pomocí SW ArcGIS Pro. Metoda Nearest Neighbour byla v rámci tohoto úkolu aplikována tak, aby našla vždy nejkratší řešení ze všech možných, které metoda pro daný soubor uzlů generuje. Prodloužení vzdálenosti (zhoršení efektivity) oproti analýze v ArcGISu bylo bezmála 12 % pro dataset ORP, a přes 15,5 % pro dataset obcí okresu Benešov – pro jejich nejkratší možná řešení. Výsledek již nelze dále vylepšovat, protože při volbě stejného počátečního uzlu bude výsledná kružnice vždy stejná (deterministické chování metody). Kružnice nalezené touto metodou se také často protínají, jinak tomu bylo v tomto případě.

Metoda Best Insertion dosahovala o poznání lepších výsledků. Nejkratší z 10 pokusů pro dataset ORP byl horší než optimum o 7,75 %, co se týče obcí okresu Benešov, bylo prodloužení oproti optimu u nejúspěšnějšího pokusu jen 4,44 %. Díky náhodnému výběru uzlů dává metoda pokaždé jiné řešení (nedeterministické chování metody). Existují způsoby, jak efektivitu této metody zvýšit – například aproximací výchozí kružnice konvexní obálkou, což vede ke zkrácení cesty o dalších cca 5 %.

Použité zdroje:

BAYER, T. (2022): Problém obchodního cestujícího, konstrukční heuristiky: stručný návod na cvičení, Výukový materiál k předmětu Geoinformatika, https://web.natur.cuni.cz/bayertom/images/courses/Geoinf/tsp_uloha.pdf

BAYER, T. (2022): Úvod do grafových algoritmů. Prezentace k předmětu Geoinformatika.

ArcČR 500 3.3, ARCDATA PRAHA, s.r.o., ZÚ, ČSÚ, 2016 (cit. 16. 12. 2022) Dostupné z: <https://www.arcdata.cz/produkty/geograficka-data/arccr-500>