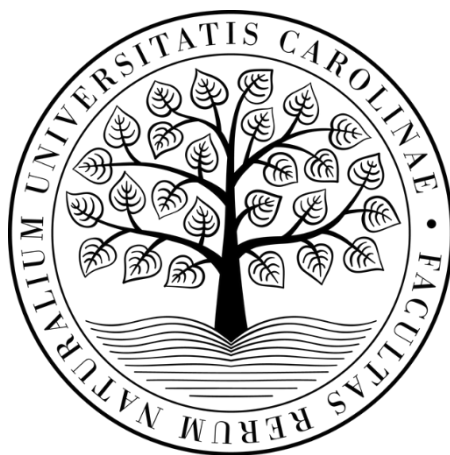


Univerzita Karlova

Přírodovědecká fakulta



Problém obchodního cestujícího

Geoinformatika

Miroslav Hruběš

1. N-GKDPZ

Světice, 2023

Zadání

Vstup: množina uzlů U reprezentující 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 Analyst 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.

Rozbor problému

Problém obchodního cestujícího (anglicky: Travelling Salesman Problem – TSP) je známý diskrétní kombinatorický problém s cílem nalézt nejkratší cestu mezi všemi zadanými body na mapě. Spadá to kategorie NP – úplných úloh. U takovýchto úloh není možné nalézt exaktní řešení v dostatečně rychlém čase. Jsme pouze schopni získat přibližné řešení, které bude akceptovatelné.

Metoda Nearest Neighbor

Metoda spočívá v prvotním náhodném či vybraném uzlu (bodu), který dále označíme jako zpracovaný. Následný uzel k propojení je vybírán podle nejmenší vzdálenosti od předešlého vybraného uzlu a zároveň se však jedná o nezpracovaný (ještě nepoužitý) uzel. Vzdálenost mezi těmito nejbližšími uzly je pak připočtena k celkové vzdálenosti cesty. Metodu

aplikujeme tak dlouho, dokud nejsou všechny uzly zpracované. Abychom na závěr získali Hamiltonovskou kružnici je potřeba propojit poslední zpracovaný uzel s počátečním uzlem.

Výhodou tohoto algoritmu je jeho deterministické chování. Nevýhodou je jeho malá účinnost. Nalezená cesta se může protínat. Výsledná Hamiltonovská kružnice je pak závislá na volbě počátečního bodu.

Metoda The Best Insertion

Oproti předchozí metodě je tato mnohem účinnější. Nejprve jsou náhodně (či záměrně) vybrány tři body jako základ uzavřené Hamiltonovské kružnice. Následně je pak každý další uzel přidán do kružnice tak, aby byl minimalizován přírůstek celkové cesty. To znamená, že uzel není připojen na konec, ale může být připojen do libovolného místa kružnice, a při jeho připojení je pak vypočítána vzdálenost mezi ním a dvěma nejbližšími uzly. Tyto vzdálenosti jsou pak přičteny k celkové délce kružnice, zatímco samotná vzájemná vzdálenost mezi těmito dvěma nejbližšími uzly je od celkové délky odečtena. V této části je zde využíváno trojúhelníkové nerovnosti. Metoda je opakována do té doby, než jsou všechny uzly součástí kružnice.

Výhodou této metody je její výpočetní složitost, která je podobná předchozí metodě, při získání mnohem lepších výsledků. Tím je myšlena kratší délka kružnice a mnohem menší pravděpodobnost, že se její segmenty budou vzájemně protínat.

Zpracování

První částí úkolu byla příprava vhodných testovacích dat. Vybrány byly dva datasety. První obsahoval souřadnice 141 měst Česka, které měli více než 6 000 obyvatel. Druhý dataset obsahoval souřadnice 125 obcí Karlovarského kraje s více než 100 obyvatel. Tyto datasety byly převzaty z volné databáze ArcČR 500 a uloženy do formátu csv.

Následovalo samotné zpracování algoritmu v jazyce Python. Ten můžeme rozdělit do čtyř částí: Metoda Nearest Neighbor, Metoda The Best Insertion, Metoda vizualizace a zbytek algoritmu, jenž zahrnuje zadání vstupů od uživatele (výběr datasetu a konkrétní metody) a aplikaci výše zmíněných metod. Pro první dvě metody figurují jako vstupy seznam souřadnic bodů a jako výstupy seznam cesty (resp. indexů daných uzlů na sebe navazujících) a celkovou délku kružnice.

Pseudokód metody Nearest Neighbor:

Inicializuj seznam statusů jednotlivých souřadnic a nastav velikost kružnice na 0.

Náhodně vyber první uzel, označ ho za uzavřený a inicializuj seznam cesty kružnice.

Dokud je nějaký uzel označen jako neuzavřený:

Nastav minimální vzdálenost w_{min} na nekonečno a proměnou indexů u na -1.

Urči souřadnice posledního uzlu na seznamu cesty kružnice

Pro každý uzel se zeptej,

zda je uzavřený a **pokud není,**

urči jeho souřadnice a spočítej vzdálenost mezi ním a posledním uzlem ze seznamu cesty.

Pokud je vzdálenost menší než minimální vzdálenost w_{min} ,

nastav tuto vzdálenost jako novou w_{min} a

index tohoto uzle jako nové u .

Uzel s indexem u označ jako uzavřený a přidej ho do cesty.

Proměnou w_{min} přičti k celkové délce.

Urči souřadnice prvního a posledního uzlu ze seznamu cesty a vzdálenost mezi nimi, kterou přičti k celkové vzdálenosti. První uzel přidej na konec seznamu cesty.

Vrať jako výstup funkce seznam cesty a celkovou vzdálenost.

Pseudokód metody The Best Insertion:

Nastav délku kružnice na 0 a Inicializuj pomocný seznam, do kterého přidej indexy jednotlivých uzlů.

Inicializuj seznam cesty, které přidáš náhodně zvolené číslo indexu uzlů, a odstraníš ho z pomocného seznamu.

Třikrát proved':

Určení souřadnic posledního uzlu (1) ze seznamu cesty.

Pokud už to uděláš potřetí,

Urči souřadnice prvního uzlu (2) seznamu cesty a přidej ho na konec tohoto seznamu.

Jinak:

Vyber náhodný uzel (2) z pomocného seznamu, přidej ho do seznamu cesty.

Odstraň ho z pomocného seznamu a urči jeho souřadnice

K celkové vzdálenosti přičti vzdálenost mezi uzly (1) a (2).

Dokud nebude pomocný seznam vyprázdněný:

Vyber náhodný uzel (1) a urči jeho souřadnice

Nastav proměnou w_{min} na nekonečno a proměnou nd na -1.

Pro každý uzel (2) ze seznamu cesty:

Urči jeho souřadnice a souřadnice uzlu následujícího (3).

Vypočítej vzájemné vzdálenosti mezi uzly (1), (2) a (3).

Vypočítej součet dvou z těchto vzdáleností, od které odečteš vzdálenost poslední.

Pokud je výsledek tohoto výpočtu menší než proměnná w_{min} :

Nastav výsledek jako novou hodnotu proměnné w_{min} .

K indexu uzle přičti 1 a číslo nastav jako novou proměnou nd .

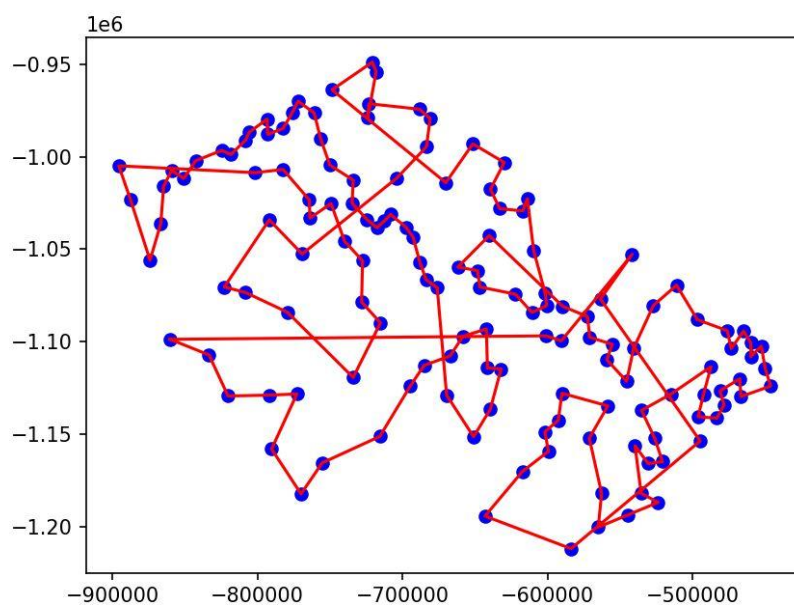
K celkové vzdálenosti přičti proměnnou w_{min} .

V seznamu cesty na indexu nd připoj uzel (1) a zároveň ho odstraň z pomocného seznamu.

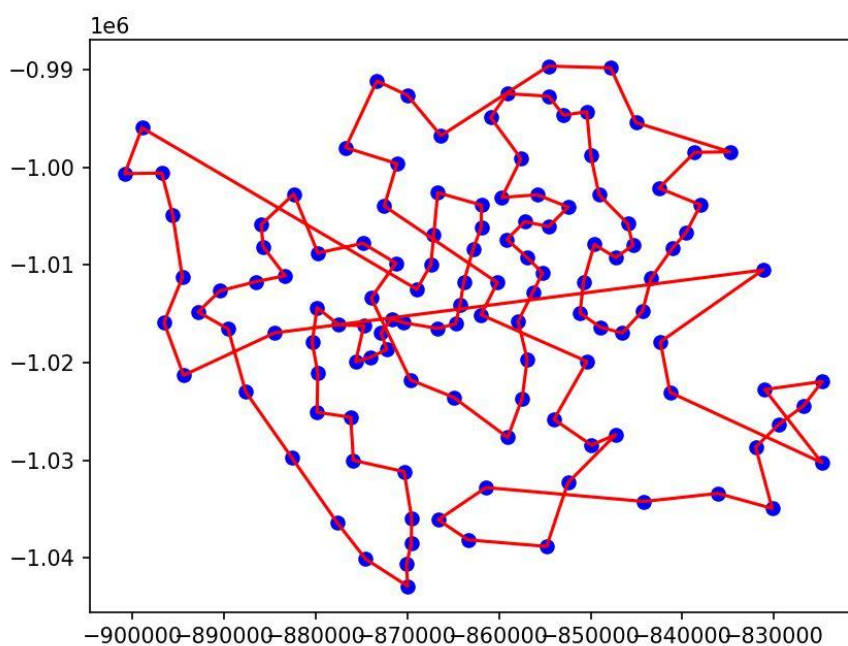
Vrať jako výstup funkce seznam cesty a celkovou vzdálenost.

Výsledky

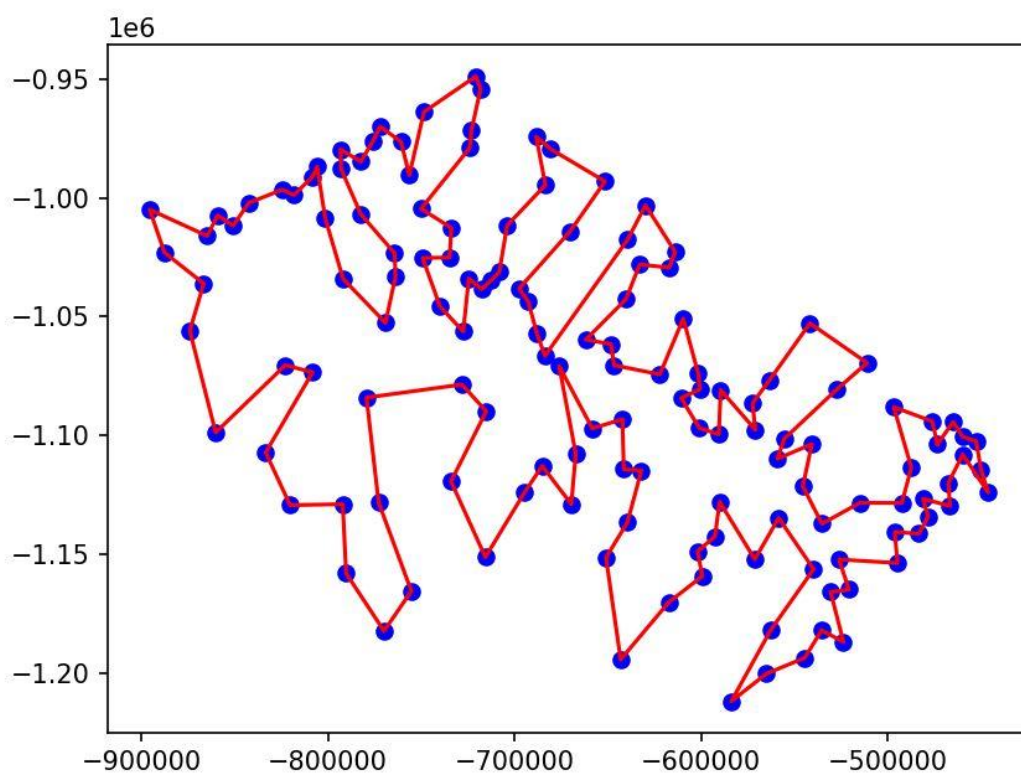
Metoda Nearest Neighbour (Obr. 1-2) byla oproti metodě The Best Insertion (Obr. 3-4) snadněji vytvářena. Negativem jsou ovšem její výsledky (Tab. 1), ve kterých v porovnání zaostává. Proto také výsledky metody The Best Insertion byly porovnány s výsledky analýzy v ArcGIS Pro (Obr. 5-6).



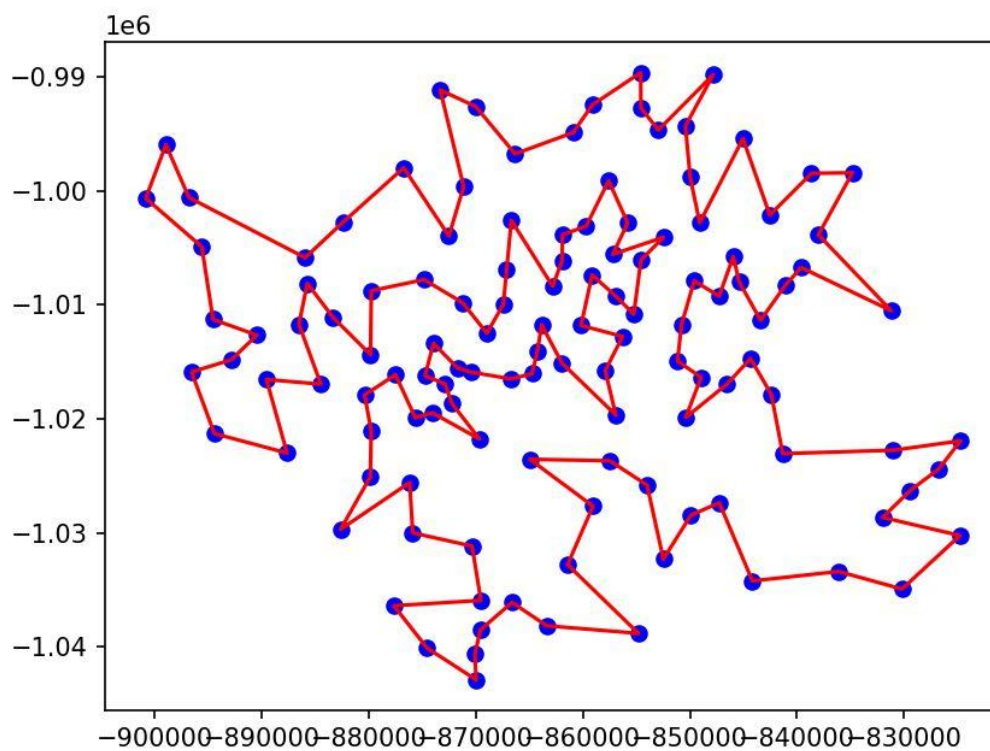
Obr. 1 – Vizualizace metody Nearest Neighbor na dataset obcí nad 6 000 obyvatel



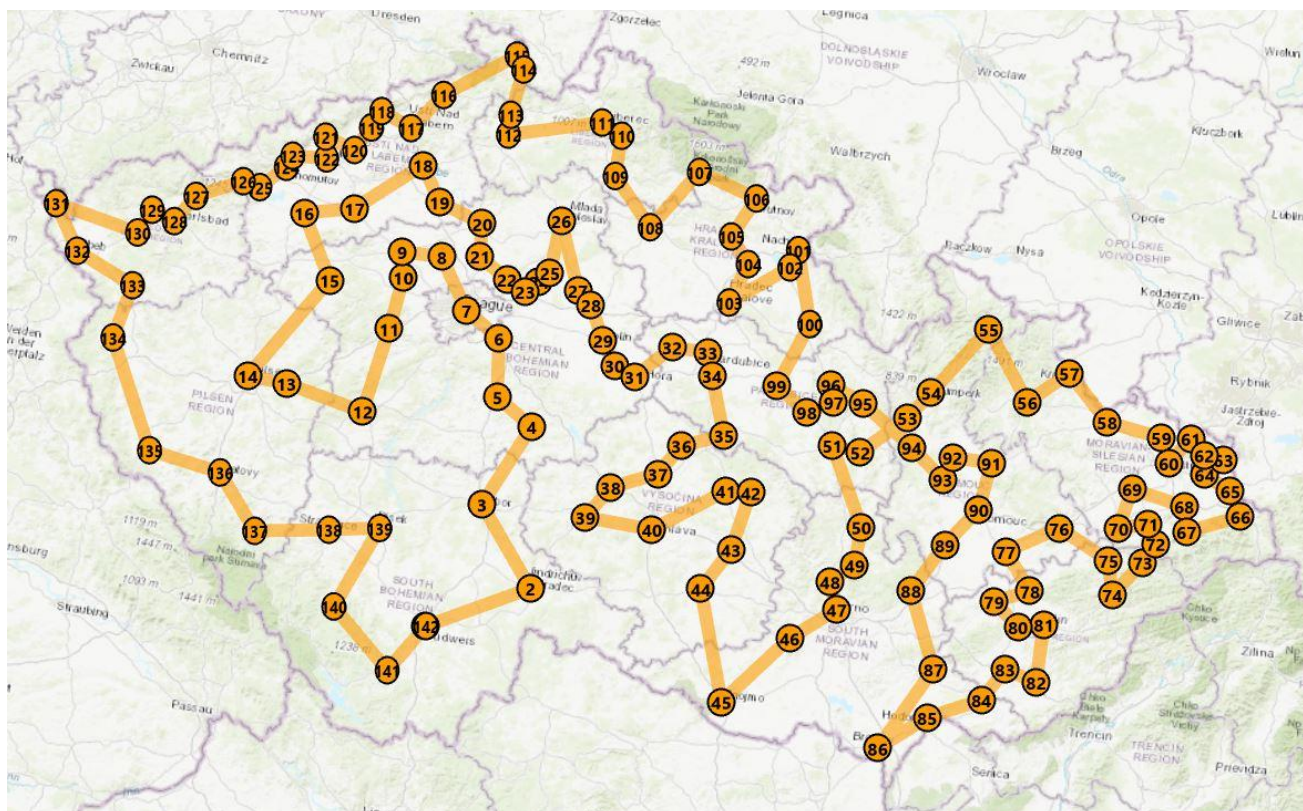
Obr. 2 – Vizualizace metody Nearest Neighbor na dataset obcí nad 100 obyvatel



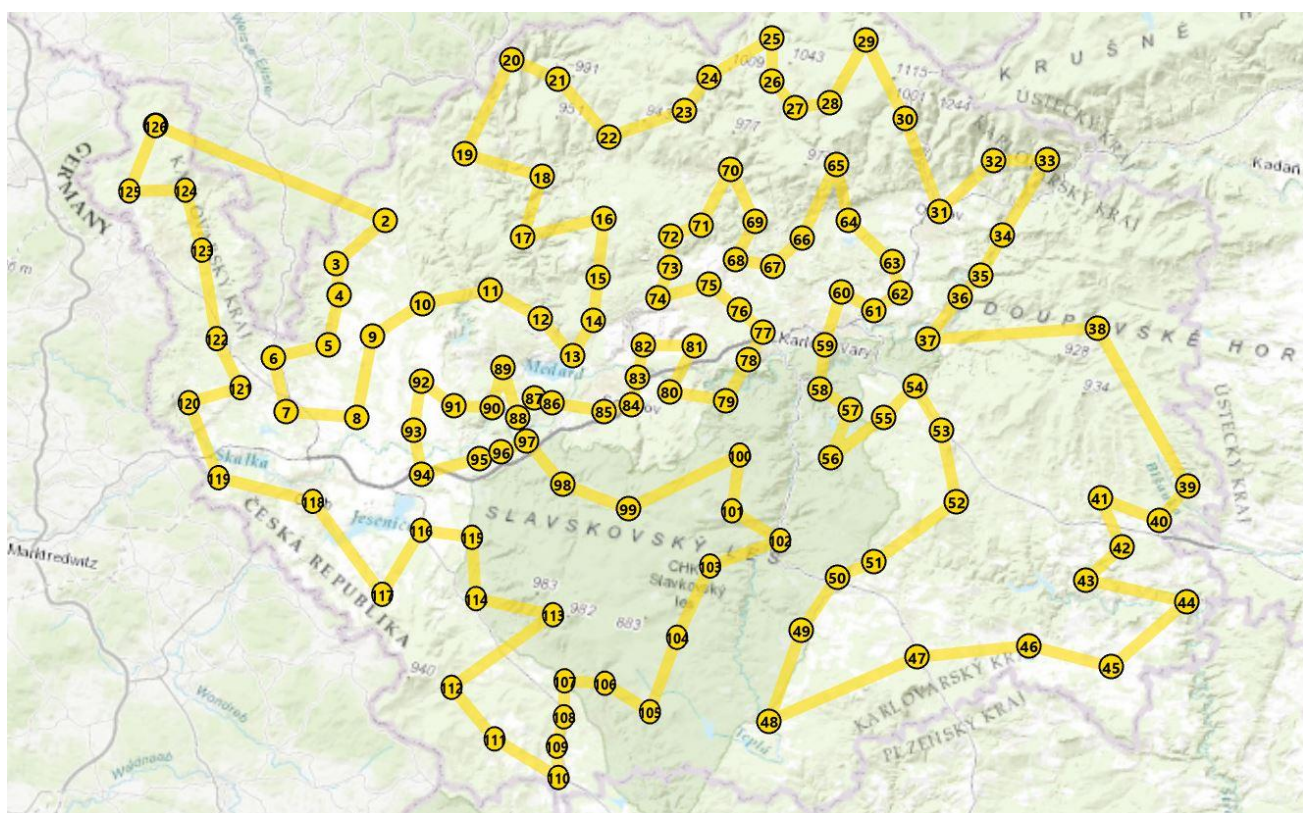
Obr. 3 – Vizualizace metody The Best Insertion na dataset obcí nad 6 000 obyvatel



Obr. 4 – Vizualizace metody The Best Insertion na dataset obcí nad 100 obyvatel



Obr. 5 – Vizualizace analýzy datasetu obcí nad 6 000 obyvatel v ArcGIS Pro



Obr. 6 – Vizualizace analýzy datasetu obcí nad 100 obyvatel v ArcGIS Pro

Obce Česka s počtem obyvatel nad 6000			Obce Karlovarska s počtem obyvatel nad 100		
Průběh	Nearest Neigh.	Best Inser.	Průběh	Nearest Neigh.	Best Inser.
1	3450,896	2494,391	1	683,062	630,183
2	3428,437	2717,037	2	695,409	666,723
3	3393,011	2762,667	3	683,462	659,326
4	3520,624	2271,317	4	702,941	624,359
5	3365,114	2150,685	5	677,452	650,916
6	3245,224	2398,255	6	699,11	587,125
7	3418,39	2382,164	7	675,579	612,073
8	3169,131	2699,523	8	676,739	564,286
9	3243,703	2380,666	9	697,846	597,475
10	3436,153	2514,621	10	665,686	614,855
Minimum	3367,068	2477,133	Minimum	685,729	620,732
ArcGIS Pro	2641,065		ArcGIS Pro	563,135	

Tab. 1 – Výsledky (v km) použitých metod na oba datasety

V případě první množiny uzlů vyšel překvapivě nejlepší výsledek při použití metody The Best Insertion (2 477,133 km). Až na druhém místě se umístil výsledek analýzy provedené v ArcGIS Pro (2 641,065 km). Nejhorší výsledek vyplynul z metody Nearest Neighbor (3 367,068 km), který se od druhé použité metody liší téměř o 900 km, což znamená rozdíl přes 35 % oproti nejlepšímu výsledku.

V případě druhé množiny uzlů už výsledky nebyly tolik překvapivé. Nejlépe dopadla analýza v ArcGIS Pro (563,135 km) za kterou se umístila metoda The Best Insertion (620,732 km) a na posledním místě opět skončila metoda Nearest Neighbor. Rozdíly mezi jednotlivými výsledky jsou v tomto případě méně extrémní. Nejhorší a nejlepší výsledek se zde liší pouze o něco lehce nad 20 %.

Zdroje

ARCDATA PRAHA (2016): ArcČR 500, databáze,
<https://www.arcdata.cz/produkty/geograficka-data/arccr-500> (cit. 7. 1. 2023)

BAYER, T. (2022): Problém obchodného cestujícího, konstrukční heuristiky – Stručný návod na cvičení. Učební text,
https://agony.natur.cuni.cz/~bayertom/images/courses/Geoinf/tsp_uloha.pdf (cit. 7. 1. 2023).