

Katedra informatiky  
Přírodovědecká fakulta  
Univerzita Palackého v Olomouci

# BAKALÁŘSKÁ PRÁCE

Algoritmy pro problém obchodního cestujícího



2023

Vedoucí práce:  
Mgr. Petr Osička, Ph.D.

Kateřina Sáňková

Studijní program: Informatika,  
Specializace: Programování a vývoj  
software

## **Bibliografické údaje**

Autor: Kateřina Sáňková  
Název práce: Algoritmy pro problém obchodního cestujícího  
Typ práce: bakalářská práce  
Pracoviště: Katedra informatiky, Přírodovědecká fakulta, Univerzita Palackého v Olomouci  
Rok obhajoby: 2023  
Studijní program: Informatika, Specializace: Programování a vývoj software  
Vedoucí práce: Mgr. Petr Osička, Ph.D.  
Počet stran: 23  
Přílohy: 1 CD/DVD  
Jazyk práce: český

## **Bibliographic info**

Author: Kateřina Sáňková  
Title: Algorithms for the travelling salesman problem  
Thesis type: bachelor thesis  
Department: Department of Computer Science, Faculty of Science, Palacký University Olomouc  
Year of defense: 2023  
Study program: Computer Science, Specialization: Programming and Software Development  
Supervisor: Mgr. Petr Osička, Ph.D.  
Page count: 23  
Supplements: 1 CD/DVD  
Thesis language: Czech

## **Anotace**

*Anotace - jeden odstavec*

## **Synopsis**

*Anotace anglicky*

**Klíčová slova:** problém obchodního cestujícího;

**Keywords:** travelling salesman problem;

Děkuji, děkuji, děkuji.

*Místopřísežně prohlašuji, že jsem celou práci včetně příloh vypracoval/a samostatně a za použití pouze zdrojů citovaných v textu práce a uvedených v seznamu literatury.*

datum odevzdání práce

podpis autora

# Obsah

<b>1</b>	<b>Teorie</b>	<b>7</b>
1.1	Graf . . . . .	7
1.1.1	Cestování v grafech . . . . .	7
1.2	Problém . . . . .	8
1.2.1	Složitost . . . . .	10
<b>2</b>	<b>Problém obchodního cestujícího</b>	<b>11</b>
2.1	Složitost . . . . .	11
2.2	Metrický problém obchodního cestujícího . . . . .	13
2.3	Euklidovský problém obchodního cestujícího . . . . .	13
<b>3</b>	<b>Algoritmy</b>	<b>15</b>
3.1	Nearest addition . . . . .	15
3.1.1	Aproximační faktor . . . . .	15
3.2	Double-tree . . . . .	17
3.2.1	Aproximační faktor . . . . .	18
3.3	Christofides' . . . . .	18
3.3.1	Aproximační faktor . . . . .	19
3.4	Keringhan - Lin . . . . .	20
	<b>Závěr</b>	<b>21</b>
	<b>Conclusions</b>	<b>22</b>
	<b>Seznam zkratk</b>	<b>23</b>

## Seznam tabulek

# 1 Teorie

Problém obchodního cestujícího je úzce spjatý s *teorií grafů* a proto je potřeba si na úvod zavést některé z jejich základních pojmů.

## 1.1 Graf

Graf je jedna ze základních reprezentací prvků množiny objektů a jejich vzájemných propojení. Takovým objektům budeme říkat *vrcholy* (někdy také *uzly*) a propojením *hrany*. Uvažujeme-li orientaci hran, pak říkáme, že je graf *orientovaný*, jinak *neorientovaný*.

### Definice 1 (Neorientovaný graf)

*Neorientovaný graf* je dvojice  $\langle V, E \rangle$ , kde  $V$  je neprázdná množina vrcholů a  $E \subseteq \{\{u, v\} \mid u, v \in V, u \neq v\}$  je množina (*neorientovaných*) hran.

$u, v$  nazýváme *koncové uzly* hrany.

V některých situacích nás bude zajímat počet hran, kterým je jistý uzel koncovým. Tomuto číslu budeme říkat *stupeň vrcholu*  $u$  a budeme ho značit  $\deg(u)$ . Důležité bude také následující tvrzení.

### Věta 2

*V každém grafu  $G = \langle V, E \rangle$  platí, že  $\sum_{v \in V} \deg(v) = 2|E|$ .*

U problému obchodního cestujícího také chceme, aby hrany vstupních grafů měly nějakou váhu. Tu jim přiřazuje tzv. *hranové ohodnocení* definované následovně:

$$w : E \rightarrow D$$

kde  $w$  je *hranové ohodnocení*,  $E$  je množina hran příslušného grafu a  $D$  je nějaká množina hodnot.

Dále je vstupem tzv. *úplný* graf. V takovém grafu platí, že každé dva jeho vrcholy jsou spojeny hranou.

### 1.1.1 Cestování v grafech

Důležitou oblastí práce s grafy je cestování v nich. Vychází se z toho, že se z jednoho uzlu můžeme přemístit k druhé. právě když mezi nimi existuje hrana (v případě orientovaných grafů musí mít ještě hrana správný směr). Jednou z úloh o cestování je právě problém obchodního cestujícího.

Základním pojmem, od kterého budeme odvozovat další, je *sled*.

### Definice 3

*Sledem* v grafu  $G = \langle V, E \rangle$  rozumíme posloupnost  $v_0, e_1, v_1, e_2, v_2, \dots, e_n, v_n$ , kde  $\forall i \in \{0, \dots, n\} v_i \in V$  a  $\forall j \in \{1, \dots, n\} e_j \in E$ .

Sled nazýváme

- *uzavřený*, pokud  $v_0 = v_n$
- *tah*, pokud  $\forall k, l \in \{1, \dots, n\}$ , kde  $k \neq l$ ,  $e_k \neq e_l$  (neopakují se hrany)
- *cesta*, pokud  $\forall k, l \in \{0, \dots, n\}$ , kde  $k \neq l$ ,  $v_k \neq v_l$  (neopakují se vrcholy)
- *kružnice*, pokud je *cestou*, ale  $v_0 = v_n$

Speciální druh tahu je tzv. *eulerovský tah*. Pro něj platí, že vede přes všechny vrcholy a každá hrana se v něm vyskytuje právě jednou.

#### Věta 4

*Pokud je neorientovaný graf souvislý a všechny jeho vrcholy mají sudý stupeň, pak v něm existuje uzavřený eulerovský tah.*

Pro TSP je ještě klíčový termín *hamiltonovská kružnice*. Tou rozumíme takovou kružnici v grafu, která vede přes všechny jeho vrcholy.

Některé z algoritmů v knihovně jsou založené na hledání tzv. *minimální kostry grafu*. Před zavedením tohoto pojmu je ještě nutné si definovat, co je to *souvislý graf* a *podgraf* grafu.

#### Definice 5 (Souvislý graf)

Neorientovaný graf  $G = \langle V, E \rangle$  nazýváme *souvislý*, pokud  $\forall u, v \in V$  existuje sled (viz později) z  $u$  do  $v$ .

#### Definice 6 (Podgraf)

Graf  $G_2 = \langle V_2, E_2 \rangle$  nazýváme *podgraf* grafu  $G = \langle V, E \rangle$ , právě když  $V_2 \subseteq V$  a  $E_2 \subseteq E$ .

#### Definice 7 (Kostra grafu)

Kostra neorientovaného grafu je jeho souvislý podgraf, který obsahuje všechny jeho vrcholy a nevyskytují se v něm žádné kružnice (viz později).

Pokud mají hrany původního grafu přiřazené váhy příslušným hranovým ohodnocením  $w$ , potom můžeme uvažovat o *minimální kostře grafu*. Tou budeme rozumět právě takovou kostru  $MSP = \langle V, E' \rangle$ , která bude mít mezi ostatními minimální součet  $\sum_{e \in E'} w(e)$ .

## 1.2 Problém

Obecně lze problémy, u kterých je rozumné chtít pro řešení použít počítač, definovat pomocí množiny vstupů  $In$ , množiny možných výstupů  $Out$  a funkce



$p : In \rightarrow Out$ , která každému vstupu přiřazuje odpovídající výstup. Tedy  $P = \langle In, Out, p \rangle$ .

*Optimalizační problémy* jsou takové problémy, kde pro vstup existuje víc možných řešení a jejich úlohou je najít mezi nimi najít které bude mezi nimi buď minimální nebo maximální vzhledem k nějaké předem dané funkci.

Tyto problémy se dají charakterizovat

- množinou vstupů  $In$
- funkcí  $sol : In \rightarrow 2^{Out}$ , která každému vstupu přiřadí množinu *přípustných řešení*
- funkcí  $cost : In \times Out \rightarrow \mathbb{Q}$ , která vstupu a jeho přípustnému řešení přiřazuje *cenu* toho řešení
- *goal*, které je buď min nebo max

Podle hodnoty *goal* se problému říká *minimalizační* nebo *maximalizační*. *Optimálním řešením* pro vstup  $x \in In$  pak označujeme takové řešení  $y \in sol(x)$ , pro které platí, že  $cost(x, y) = goal\{cost(x, y') \mid y' \in sol(x)\}$ . Cenu takového řešení značíme  $OPT(x)$ .

Při hledání algoritmu, které řeší takový problém zvažujeme jeho *správnost* a *optimalitu*. Řekneme, že algoritmus  $A$  pro problém  $P$  je *správný*, pokud  $\forall x \in In, A(x) \in sol(x)$ . *Optimální* je navíc pokud  $\forall x \in In$  je  $A(x)$  optimální řešení.

Některé optimální algoritmy bývají ale časově náročné, často se proto hledají algoritmy, které nevydají pokaždé optimální řešení, ale pouze přibližné. Těmto algoritmům se říká *aproximační*.

Mějme takový algoritmus  $A$ . Pro každý vstup  $x \in In$  můžeme definovat *aproximační faktor* jako

$$R_A(x) = \max\left\{\frac{cost(x, A(x))}{OPT(x)}, \frac{OPT(x)}{cost(x, A(x))}\right\}$$

Pro samotný algoritmus můžeme také definovat jeho aproximační faktor, vzhledem k nějaké funkci  $F$  mapující vstupy na přirozená čísla, a to následovně

$$R_A(n) = \max\{R_A(x) \mid x \in In, F(x) = n\}$$

$A$  můžeme označit za  $f(n)$ -aproximační algoritmus, pokud  $\forall n \in \mathbb{N}$  platí  $R_A(n) \leq f(n)$ . Jinými slovy, pokud u algoritmu známe jeho aproximační faktor, pak se můžeme spolehnout, že pro každý vstup velikosti  $n$  dostaneme v nejhorším případě  $f(n)$ -krát horší výsledek vůči optimu.

Další významnou množinou problémů jsou *rozhodovací problémy*. Ty se vyznačují tím, že množina výstupů je omezena jen na ANO a NE.

Každý optimalizační problém má svoji *rozhodovací verzi*. Místo, aby na vstupu byla pouze instance problému  $x \in In$ , přidá se navíc ještě číslo  $k \in \mathbb{Q}$ . Pokud je  $k \geq OPT(x)$ , u minimalizačního problému, nebo  $k \leq OPT(x)$ , u maximalizačního, pak je výsledkem ANO, jinak NE.

### 1.2.1 Složitost

Na začátku je třeba si zavést třídy  $\mathbf{P}$  a  $\mathbf{NP}$ . Algoritmy, které spadají do třídy  $\mathbf{P}$  jsou řešitelné deterministickými stroji v polynomiálním čase a bereme je za praktický zvládnutelné. Na nedeterministických strojích můžeme řešit v polynomiálním čase  $\mathbf{NP}$  problémy. U těchto strojů předpokládáme, že z možných kroků provede vždy ten žádaný a dostane se tak ke správnému výsledku bez nutnosti zkoušet všechny možnosti. Jistě platí  $\mathbf{P} \subseteq \mathbf{NP}$ , jelikož deterministické stroje jsou speciálním případem nedeterministických, které pouze nevyužívají možnosti nedeterministického výběru.

Některé rozhodovací problémy mají navíc tu vlastnost, že jsou na ně polynomiálně redukovatelné všechny rozhodovací problémy v  $\mathbf{NP}$ . To znamená, že pro každý vstup libovolného problému  $P_1$  můžeme v polynomiálním čase najít vstup pro  $P_2$ , který bude vracet stejnou odpověď, jakou by vrátil  $P_1$  na původní vstup. Problémům s touto vlastností se říká  *$\mathbf{NP}$ -těžké*.

Pokud jsou navíc  $\mathbf{NP}$ , říká se jim  *$\mathbf{NP}$ -úplné*. Tyto problémy jsou důležité, protože kdyby se podařilo najít algoritmus řešící kterýkoli z nich v polynomiálním čase, pak bychom jistě zvládli vyřešit i všechny  $\mathbf{NP}$  problémy v polynomiálním čase, tedy  $\mathbf{NP} \subseteq \mathbf{P}$  a z toho  $\mathbf{NP} = \mathbf{P}$ . Naopak, pokud bychom zjistili, že  $\mathbf{NP} \neq \mathbf{P}$ , pak pro žádný  $\mathbf{NP}$ -úplný problém jistě nemůže existovat polynomiální algoritmus, který ho řeší.

Z  $\mathbf{P} \neq \mathbf{NP}$  jde vyvodit ještě jedno důležité tvrzení. Nechť  $P_O$  je optimalizační problém a  $P_D$  je jeho rozhodovací verze. Pokud dokážeme, že  $P_D$  je  $\mathbf{NP}$ -těžký, pak pro  $P_O$  neexistuje polynomiální algoritmus. Kdyby totiž takový algoritmus existoval, stačilo by ho využít pro nalezení optimálního řešení, porovnat jeho cenu s se zadaným číslem  $k$  pro  $P_D$  a na základě toho, jestli by bylo větší nebo menší v závislosti na povaze  $P_O$  vydat ANO nebo NE. Tímto bychom tedy získali polynomiální algoritmus řešící také  $P_D$ . Jenže protože  $P_D$  je  $\mathbf{NP}$ -těžký, tak všechny  $\mathbf{NP}$  problémy jsou na něj polynomiálně redukovatelné a zvládli bychom je tedy i v polynomiálním čase řešit a tedy  $\mathbf{P} = \mathbf{NP}$ , což je spor.

## 2 Problém obchodního cestujícího

Problém obchodního cestujícího (TSP - *Travelling Salesman Problem*) je jedním z nejstudovanějších optimalizačních problémů. Podstatou je mezi zadanými městy nalézt cestu, která začíná i končí ve stejném místě a zbytek měst navštíví právě jedenkrát. Jelikož seznam měst můžeme brát jako množinu vrcholů grafu, můžeme o TSP uvažovat jako o úloze o cestování v grafu. Cestou, kterou hledáme je pak *hamiltonovská kružnice*.

Formálně bychom mohli TSP definovat následně:

### Definice 8 (Problém obchodního cestujícího)

NÁZEV: TSP

VSTUP: Úplný neorientovaný graf  $G = \langle V, E \rangle$ ,

hranové ohodnocení  $c : E \leftarrow R^+$

VÝSTUP: Hamiltonovská kružnice v  $G$  minimální délky

Ceny hran budeme nazývat také *délky* a pro jednoduchost budeme místo  $c(\{u, v\})$  psát  $c_{u,v}$ .

### 2.1 Složitost

Jedním z důvodů, proč je TSP tak zkoumaným problémem, je jeho složitost. Samotný problém je jistě v **NP**. Nedeterministický stroj, který by měl za úkol ho vyřešit, by prostě vybral  $|V|$  hran z  $E$  a jelikož by vždy volil správně, vrátil by nejkratší hamiltonovskou kružnici v grafu. Tohle by samozřejmě zvládnul provést v polynomiálním čase.

Zavedeme si nyní rozhodovací verzi k TSP.

### Definice 9 (Rozhodovací verze TSP)

NÁZEV: TSP

VSTUP: Úplný neorientovaný graf  $G = \langle V, E \rangle$ ,

hranové ohodnocení  $c : E \leftarrow R^+$ ,

číslo  $k$

OTÁZKA Existuje v  $G$  hamiltonovská kružnice délky nejvýš  $k$ ?

Ví se, že tato verze problému je **NP**–úplná. Tedy platí, že pokud nalezneme polynomiální algoritmus řešící TSP, pak dokážeme **P** = **NP**.

Nyní si dokážeme ještě zajímavější tvrzení.

**Věta 10**

*Pokud by existoval  $2^n$ -aproximační polynomiální algoritmus pro TSP, pak  $P = NP$ .*

Něž začneme s důkazem, musíme ji ještě zavést problém nalezení hamiltonovské kružnice (*HC - Hamiltonian Cycle*) v grafu.

**Definice 11 (Rozhodovací verze HC)**

NÁZEV: HC

VSTUP: Neorientovaný graf  $G = \langle V, E \rangle$

OTÁZKA Existuje v  $G$  hamiltonovská kružnice?

O tomto problému je známo, že je **NP**-úplný. Navíc se dá snadno převést na TSP. Jak jsme již zmínili, TSP má ve své klasické podobě na vstupu graf a jeho příslušné hranové ohodnocení. Vstup  $x$  pro HC bychom mohli redukovat na vstup  $r(x)$  pro TSP tak, že bychom vytvořili úplný graf mezi uzly z původní instance a ohodnotili bychom jeho hrany následovně:

$$c_e = \begin{cases} 1 & \text{pro } e \in E \\ |V| \cdot 2^{|V|} + 1 & \text{pro } e \notin E \end{cases}$$

Snadno můžeme dojít k tomu, že

$$OPT_{TSP}(r(x)) = \begin{cases} = |V| & \text{pro } x \in HC \\ > |V| \cdot 2^{|V|} & \text{pro } x \notin HC \end{cases}$$

Uvědomme si, že hrany s cenou  $|V| \cdot 2^{|V|}$  budou obsaženy v optimálním řešení pouze v případě, kdy nelze sestavit hamiltonovskou kružnici z hran původního grafu. Pokud by byla musela být vybrána pouze jedna taková hrana, pak bude cena nalezené cesty

$$(|V| - 1) + (|V| \cdot 2^{|V|} + 1) > |V| \cdot 2^{|V|}$$

Díky takovému ohodnocení existuje tedy exponenciálně velká mezera mezi pozitivními a negativními instancemi HC. Pokud bychom měli  $2^n$ -aproximační polynomiální algoritmus  $A$  pro TSP, pak pro pozitivní instance HC by našel cestu délky nejhůře  $|V| \cdot 2^{|V|}$ , pro negativní instance zase nejlépe délky  $|V| \cdot 2^{|V|} + 1$ .

HC bychom mohli vyřešit tak, že bychom spustili  $A$  na  $r(x)$  a pokud by délka vrácené cesty byla menší nebo rovna  $|V| \cdot 2^{|V|}$ , pak bychom vrátili ANO, v opačném případě NE. Vzhledem k tomu, že převod na  $r(x)$ ,  $A$  i porovnání na konci běží v polynomiálním čase, pak bychom celý HC mohli rozhodnout v polynomiálním čase. Uvedli jsme navíc, že HC je **NP**-úplný. V kapitole o složitosti jsme řešili, že nalezení polynomiálního algoritmu pro kterýkoli z problémů této množiny by znamenalo, že  $P = NP$ .

Jedním z důvodů, proč je TSP tak zkoumaným problémem, je jeho složitost. Abychom našli správné (optimální) řešení, museli bychom projít všechna přípustná řešení a porovnat je mezi sebou. Takových řešení je ale vůči počtu měst exponenciálně. Pokud bychom ale měli k dispozici nedeterministický stroj, pak bychom se mohli spolehnout na to, že vždy najde nejkratší cestu hned. Samotné nalezení cesty je pak zvládnutelné v polynomiálním čase a můžeme o TSP říct, že je to nedeterministický polynomiální problém (*NP - problém*). Tohle platí i pro jeho *rozhodovací verzi*.

*Rozhodovací problémy* se vyznačují tím, že jako výstupem je pouze *Ano* nebo *Ne*. TSP můžeme upravit následovně.

### Definice 12 (Rozhodovací verze TSP)

NÁZEV: TSP

VSTUP: Úplný neorientovaný graf  $G = \langle V, E \rangle$ ,

hranové ohodnocení  $c : E \leftarrow R^+$ ,

číslo  $k$

OTÁZKA: Existuje v  $G$  hamiltonovská kružnice délky nejvýš  $k$ ?

Tato verze navíc tu vlastnost, že libovolný rozhodovací NP problém je na něj polynomiálně redukovatelný. Čili pro každý vstup problému můžeme v polynomiálním čase najít vstup pro TSP, který bude vracet stejnou odpověď, jakou by vrátil původní problém na původní vstup. Problémům s touto vlastností se říká *NP-těžké*. Pokud jsou navíc NP, říká se jim *NP-úplné*. Tyto problémy jsou důležité, protože kdyby se podařilo najít algoritmus řešící kterýkoli z nich v polynomiálním čase (takové problémy považujeme za prakticky zvládnutelné), pak bychom jistě zvládli vyřešit i všechny NP problémy v polynomiálním čase.

## 2.2 Metrický problém obchodního cestujícího

TSP může být žádané také omezit jiným způsobem. Pokud přidáme sna hranové ohodnocení podmínku, že musí platit trojúhelníková nerovnost, tedy  $\forall u, v, w \in V$  platí

$$c_{u,v} + c_{v,w} \geq c_{w,u}$$

pak dostaneme *metrický TSP*.

## 2.3 Euklidovský problém obchodního cestujícího

Přísnější podmínkou pak může být definice délky jako euklidovské vzdálenosti. Ta je pro vrcholy  $u, v \in V$  se souřadnicemi  $u = (u_1, u_2, \dots, u_n)$ ,  $v = (v_1, v_2, \dots, v_n)$  definována

$$c_{u,v} = \sqrt{\sum_{i=1}^n (u_i - v_i)^2}$$



## 3 Algoritmy

Jak bylo již zmíněno, TSP je řešitelný v polynomiálním čase pouze nedeterministickými stroji. U deterministických (klasických) strojů je složitost exponenciální. Z toho ale vyplývá, že bychom se pro větší vstupy nemuseli v reálném čase dočkat odpovědi. Z tohoto důvodu se pro hledání řešení používají tzv. *heuristiky*. Při jejich použití se nemůžeme spolehnout, že najdeme optimální řešení. Jsme ale schopni nalézt přibližné řešení v reálném čase.

U některých heuristik se navíc můžeme bavit o *aproximačním faktoru*. Mějme libovolný vstup problému  $I$ , pro který je cena optimálního řešení  $OPT$  a heuristiku s aproximačním faktorem  $\alpha$ . Potom se můžeme spolehnout, že cena řešení, které dostaneme použitím dané heuristiky na  $I$  bude při nejhorším  $\alpha * OPT$ .

### 3.1 Nearest addition

Tento algoritmus spadá do skupiny hladových algoritmů a je poměrně intuitivní. Vychází z myšlenky, že k vytvářené cestě vždy přidáme nejbližší nenavštívené město. Nejprve najdeme taková dvě města  $i, j$ , která jsou si v grafu nejbližší a vytvoříme cestu  $path$ , která povede z  $i$  do  $j$  a zpět. Následuje cyklus, ve kterém budeme hledat takovou hranu  $(k, l)$ , že  $k \in path$ ,  $l \notin path$  a  $c_{kl}$  je všemi takovými hranami minimální. Předpokládejme, že město  $m$  je aktuálně v  $path$  za  $k$ . Vložíme  $l$  do  $path$  mezi  $k$  a  $m$  a následuje další iterace. Cyklus skončí v momentu, kdy jsme navštívili všechny města.

---

**Algorithm 1:** Nearest addition algorithm

---

```
vytvoř seznam  $path$ ;  
vytvoř seznam  $unvisitedCities$ ;  
 $unvisitedCities \leftarrow nodes$ ;  
 $(i, j) \leftarrow$  nejkratší hrana v grafu;  
 $path.Add(i, j, i)$ ;  
 $unvisitedCities.Remove(i, j)$ ;  
while  $unvisitedCities.Count > 0$  do  
     $(k, l) \leftarrow$  nejkratší hrana v grafu, kde  $k \in path$  a  $l \in unvisitedCities$ ;  
     $path.AddAfter(k, l)$ ;  
     $unvisitedCities.Remove(l)$ ;  
return  $path$ 
```

---

#### 3.1.1 Aproximační faktor

##### Věta 13

*Nearest addition je 2-aproximační algoritmus.*

Abychom si tvrzení dokázali, musíme si nejdříve ukázat vztah k Primově algoritmu, který hledá minimální kostru grafu.

---

**Algorithm 2:** Primův algoritmus

---

```
vytvoř seznam spanningTree;  
vytvoř seznam unvisitedCities;  
unvisitedCities  $\leftarrow$  nodes;  
(i, j)  $\leftarrow$  nejkratší hrana v grafu;  
spanningTree.Add((i, j));  
unvisitedCities.Remove(i, j);  
while unvisitedCities.Count > 0 do  
    (k, l)  $\leftarrow$  nejkratší hrana v grafu, kde k  $\notin$  unvisitedCities a  
        l  $\in$  unvisitedCities;  
    spanningTree.Add((k, l));  
    unvisitedCities.Remove(l);  
return spanningTree
```

---

Z pseudokódu je jasné, že hrany nalezené Primovým algoritmem i Nearest addition algoritmem budou totožné.

**Lemma 14**

*Pro libovolný vstup TSP je cena jeho optimální cesty alespoň tak vysoká, jako je cena minimální kostry pro tentýž vstup.*

*Důkaz*

Vezmeme optimální cestu pro libovolný vstup s více než jedním vrcholem a odstraníme některou z hran. Takto vzniklý graf jistě nebude mít vyšší cenu než optimální cesta. Navíc bude také kosterou původního vstupu. To si můžeme snadno ověřit. Graf stále obsahuje všechny vrcholy, jeho souvislost jsme nijak neporušili a jedinou kružnici kterou obsahoval jsme otevřeli odstraněním jedné z hran.

To, že tato kostra nemůže mít nižší cenu než minimální kostra, je triviální a lemma jsme dokázali.  $\square$

**Věta 15**

*Nearest addition algoritmus pro metrický problém obchodního cestujícího je 2-aproximační algoritmus.*

*Důkaz*

\*\*\*to se musí prepsat\*\*\* Mějme podmnožiny  $V$   $S_2, S_3, \dots, S_n$ , kde  $S_2$  je množina vrcholů určených první iterací,  $S_3$  druhou atd. Jistě platí, že  $|S_l| = l$  – začínáme s dvěma městy a poté přidáváme v každé iteraci jedno další. Dále řekněme, že  $F = \{(i_2, j_2), (i_3, j_3), \dots, (i_n, j_n)\}$  je množina nejkratších hran určených každým průchodem cyklu. Pro hranu  $(i_l, j_l)$  platí, že  $i_l \in S_{l-1}$ , pro  $l \in 3, \dots, n$ . Již bylo zmíněno, že graf  $\langle V, F \rangle$  je minimální kostra vstupu a optimální cena cesty, označme  $OPT$ , nemůže stát méně, tedy  $OPT \geq \sum_{l=2}^n c_{i_l j_l}$ .



Uvědomme si, že cena cesty pro první dva uzly  $i_2$  a  $j_2$  je přesně  $2c_{i_2j_2}$ , protože tuto hranu vkládáme na začátku do cesty dvakrát.

Když vkládáme nově určené město  $j$  do cesty mezi  $i$  a  $k$ , tak vlastně hrany  $(i, j)$  a  $(j, k)$  přidáváme a  $(i, k)$  odstraňujeme. Tedy celková cesta se může nejvíce prodloužit o  $c_{ij} + c_{jk} - c_{ik}$ . Z trojúhelníkové nerovnosti snadno odvodíme, že  $c_{jk} - c_{ik} \leq c_{ij}$ . Rozdíl cen hran  $c_{jk}$  a  $c_{ik}$  je tedy shora ohraničený cenou  $c_{ij}$ . Dosadíme-li tuto hodnotu do předchozí rovnice, zjistíme, že celková cesta se může v nejhorším případě prodloužit o  $2c_{jk}$ . Celková cesta nalezená Nearest addition algoritmem bude tedy maximálně  $2 \sum_{l=2}^n c_{i_lj_l}$ .

Upravíme-li poslední nerovnost, dostaneme  $\sum_{l=2}^n c_{i_lj_l} \leq 2OPT$ .

To znamená, že naše nalezená cesta bude v nejhorším případě stát dvakrát tolik, co optimální cesta pro stejný vstup, což je přesně to, co jsme chtěli dokázat.  $\square$

## PŘÍKLAD 16

### 3.2 Double-tree

Jádrem algoritmu je zdvojení hran minimální kostry vstupního grafu a následné nalezení uzavřeného eulerovského tahu. To, že se v takovémto grafu takový tah nachází je snadné dokázat. Z Věty xxx víme, že stačí, aby byl graf spojitý a všechny jeho uzly měly sudý stupeň. Zdvojíme-li každou hranu, vycházející z libovolného uzlu, pak bude mít jistě uzel sudý stupeň. Jelikož vycházíme z kostry grafu, spojitost je zřejmá.

K vytvoření cesty už potřebujeme pouze zaručit, že nenavštívíme žádný uzel vícekrát. Toho dosáhneme pomocí zkracování. Máme-li eulerovský průchod  $(i_0, i_1), (i_1, i_2), \dots (i_{k-1}, i_k), (i_k, i_0)$ , pak vezmeme posloupnost  $i_0, i_1, \dots, i_k$  a ponecháme pro každý uzel pouze jeho první výskyt. Nakonec ještě vrátíme  $i_0$  na konec cesty.

---

#### Algorithm 3: Double-tree algoritmus

---

**Input:**  $G$

$doubleTree \leftarrow FindMinimalSpanningTree(G);$

**foreach**  $edge \in doubleTree$  **do**

$doubleTree.Add(edge);$

$eulerCircuit \leftarrow FindEulerCircuit(doubleTree);$

vytvoř seznam  $path$ ;

**foreach**  $node \in eulerCircuit$  **do**

**if**  $node \notin path$  **then**

$path.Add(node);$

**return**  $path$

---

### 3.2.1 Aproximační faktor

#### Věta 17

*DoubleTree je 2-aproximační algoritmus.*

Z Lemma 14 víme, že optimální cesta stojí alespoň tolik, co minimální kostra. Jelikož každou její hranu máme dvakrát, pak tato cena bude dvojnásobek její ceny.

Při hledání eulerovského průchodu pouze hrany seřazujeme.

Zbývá pouze zkracování. Zde opět vycházíme z trojúhelníkové nerovnosti. Uvažujeme, že  $i, j, k$  jsou po sobě jdoucí města a  $j$  jsme již navštívili, tedy bychom je měli přeskočit. Je zřejmé, že  $c_{ij} + c_{jk} \geq c_{ik}$ , což dokazuje, cena tohoto úseku zůstane při nejhorším stejná.

#### PŘÍKLAD 18

### 3.3 Christofides'

Christofidesův algoritmus je velice podobný Double-tree algoritmu. Opět začneme s minimální kostrou vstupního grafu. Existenci eulerovského průchodu ale zajistíme chytřeji. Připomeňme si, že eulerovský průchod se v grafu nachází právě tehdy, když mají všechny jeho stupně sudý stupeň. V Double-tree algoritmu jsme tuto podmínku splnili tak, že jsme zdvojili každou hranu kostry, to je ovšem redundantní. Postačí, když spárujeme uzly, které v ní mají lichý stupeň.

Mějme úplný graf, jehož množinou vrcholů budou právě ty vrcholy minimální kostry s lichým stupněm, označme ji  $O$ . Z množiny jeho hran se budeme snažit vybrat takovou podmnožinu, ve které bude  $\forall v \in O$  právě jedna hrana, pro kterou je  $v$  koncový uzel. Taková množina se nazývá *perfektní párování*. Takových množin může být více, přirozeně budeme hledat takovou z nich, která budeme mít minimální součet délek všech svých hran.

Uvědomme si ale, že aby takové párování pro  $O$  existovalo, tak musí být  $|O|$  sudé. Víme, že součet stupňů všech uzlů v neorientovaném grafu je sudé číslo (viz Věta 2). Označme si množinu všech vrcholů grafu  $V$  a množinu vrcholů se sudým stupněm  $E$ . Pak  $\sum_{v \in V} \deg(v) = \sum_{v \in E} \deg(v) + \sum_{v \in O} \deg(v)$ .  $\sum_{v \in V} \deg(v)$  je ale sudé a to samé jistě platí pro  $\sum_{v \in E}$ , tedy  $\sum_{v \in O} \deg(v)$  musí být sudý. Vzhledem k tomu, že každý vrchol přispívá do součtu lichým číslem, pak jejich počet musí být sudý.

Přidáním minimálního perfektního párování k minimální kostře zvýšíme stupeň každého vrcholu s lichým stupněm právě o jedna, tudíž zajistíme, že všechny uzly budou mít sudý stupeň a tudíž v grafu existuje uzavřený eulerovský tah. Po jeho nalezení už jen zkracujeme cestu stejně jako u Double-tree algoritmu.

---

**Algorithm 4:** Christofidesův algoritmus

---

**Input:**  $G$ 

```
spanningTree  $\leftarrow$  FindMinimalSpanningTree( $G$ );  
oddDegreeNodes  $\leftarrow$  FindNodesWithOddDegrees(spanningTree);  
perfectMatching  $\leftarrow$  FindPerfectMatching(oddDegreeNodes);  
eulerCircuit  $\leftarrow$  FindEulerCircuit(doubleTree + perfectMatching);  
vytvoř seznam  $path$ ;  
foreach  $node \in eulerCircuit$  do  
    if  $node \notin path$  then  
         $path.Add(node)$ ;  
return  $path$ 
```

---

### 3.3.1 Aproximační faktor

#### Věta 19

*Christofidesův algoritmus je  $3/2$ -aproximační algoritmus.*

\*\*\*nějaké obrázky\*\*\*

Nejdřív si musíme rozmyslet, co vlastně potřebujeme dokázat. Potřebujeme, aby eulerovský průchod měl při nejhorším délku  $\frac{3}{2}OPT$ . Z lemmatu 14 víme, že minimální kostra, která je v něm obsažená, má v nejhorším případě cenu  $OPT$ . Stačí tedy dokázat, že perfektní párování má maximálně cenu  $\frac{1}{2}OPT$ .

Pracujeme s množinou lichých uzlů v MST  $O$ . Budeme v ní teď hledat cestu. Vydeme-li z nejkratší cesty nad všemi vrcholy vstupu, pak je zřejmé, že její délka bude  $OPT$ . Tuto cestu teď budeme pouze zkracovat. Uvažujeme-li dvě města  $i$  a  $j$ , pro která platí, že na cestě mezi nimi jsou pouze města nenáležící do  $O$ , pak je jisté můžeme vynechat a z trojúhelníkové nerovnosti víme, že v nejhorším případě bude hrana  $(i, j)$  stejně dlouhá jako původní cesta. V nejhorším případě tedy po tomto zkracování zůstane délka cesty rovna  $OPT$ .

Nyní začneme střídavě obarvovat hrany nalezené cesty, řekněme modrou a červenou. Množina červených i množina modrých hran je perfektním párováním na  $O$ , platí totiž, že pokrývají všechny vrcholy a žádné dvě hrany nesdílí vrchol. Víme, že dohromady množiny dávají  $OPT$ , pak jistě jedna z množin bude mít délku menší nebo rovnu  $\frac{1}{2}OPT$ , což jsme chtěli dokázat.

Z analýz aproximačních faktorů Christofidesova a Double-tree algoritmu ještě vyplývá následující tvrzení.

#### Lemma 20

*Sestrojíme-li pro vstup eulerovský podgraf délky  $\alpha \cdot OPT$ , pak můžeme odvodit  $\alpha$ -aproximační algoritmus.*

#### PŘÍKLAD 21

### 3.4 Keringhan - Lin

Tento algoritmus není specifický pro TSP, ale je možné jím řešit obecně problémy, kde máme nějakou podmínku  $C$ , kterou musejí splňovat jejich řešení, a funkci  $f$ , která jim přiřazuje určitou hodnotu. Zadanou pak máme množinu  $S$ , ve které hledáme podmnožinu  $T$ , která bude nejen splňovat podmínku  $C$ , ale její hodnota  $f$  bude mezi všemi vyhovujícími podmnožinami minimální. Takové problémy mívají často exponenciální složitost. Jedním z takových problémů, kde se Kernighan - Lin využívá je například *dělení grafů*.

Základem je postupné zlepšování řešení. Hledá se vždy nějaké lokální optimum, které se, pokud přinese nějaký zisk (zmenšení hodnoty  $f$ ), použije pro další iteraci.

Na začátku se zvolí pseudonáhodné přípustné řešení  $T \subseteq S$ . Následně začne cyklus, kde se snažíme  $T$  transformovat na jiné platné řešení  $T'$ . Pokud by platilo  $f(T') < f(T)$ , pak se  $T'$  použije pro další iteraci. V momentě, kdy už nenacházíme řešení, které by přinášely nějaký zisk, tak bylo nalezeno lokálně optimální řešení. V tuto chvíli se buď může skončit nebo se vygeneruje nové počáteční řešení a proces se opakuje.

Samotná transformace probíhá tak, že se hledají takové dvě množiny  $\{x_1, x_2, \dots, x_k\} \in T$  a  $\{y_1, y_2, \dots, y_k\} \in S - T$ , pro které platí, že když nahradíme prvky první množiny v  $T$  prvky druhé množiny, pak dostaneme platné řešení  $T'$ . O tom, jak přijdeme k číslu  $k$  (už konkrétně u TSP) bude ještě řeč později.

Myšlenka algoritmu tedy spočívá v tom, že pro nějaké počáteční řešení najdeme vždy lokální optimum a mezi nimi by se snad mělo objevit i optimum globální (nebo alespoň řešení se schůdnou hodnotou pro  $f$ ). Přirozeně budeme považovat algoritmus za tím kvalitnější, čím menší počet různých lokálních optim dostaneme a čím větší počet náhodných počátečních řešení nám dá dostatečně dobrý výsledek.

---

**Algorithm 5:** Keringhan - Lin algoritmus - obecně

---

```
1  $T \leftarrow$  pseudonáhodné řešení;  
2  $T' \leftarrow$  transformace  $T$  taková, že  $f(T') < f(T)$ ;  
3 if  $T' \neq T$  then  
4    $T \leftarrow T'$ ;  
5   go to 2;  
6 else  
7   return nebo, pokud je to žádané go to 1;
```

---

## **Závěr**

Závěr práce v „českém“ jazyce.

## Conclusions

Thesis conclusions in “English”.

