组合优化

浙江大学数学系　谈之奕

浙江大学
ZheJiang University

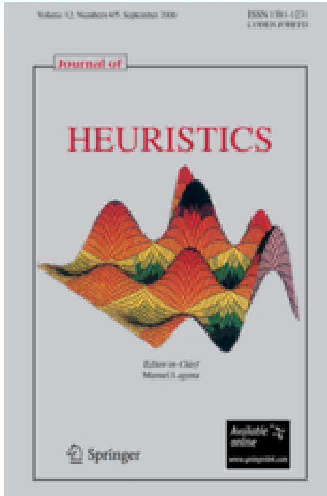
组合优化问题的求解方法

近似算法

浙江大学 ZheJiang University

组合优化

- 可给出$\mathcal{NP}$－难问题任一实例最优解的算法总需要指数时间。较为现实的思路是"用精度换时间"，在多项式时间或可接受的实际运行时间内得到一个目标值与最优值较为接近的可行解
  - 近似算法（**approximation algorithm**）：算法的时间复杂性可通过分析确定（一般要求多项式时间），算法给出的近似解与最优解目标值之间的差距可通过证明严格估计
  - 启发式算法（**heuristic**）：无法说明算法的时间复杂性，或无法估计算法给出的近似解与最优解目标值之间的差距



Journal of Heuristics

(ευρισκειν)

to find

最坏情况比

组合优化

- 设Π是一个极小化优化问题，A是它的一个算法。若对Π的任意实例I，算法A给出一个可行解，其目标值为$C^A(I)$，实例I的最优目标值为$C^*(I)$
- 称 $r_A = \inf\{r \ge 1 \mid C^A(I) \le rC^*(I), \forall I\}$或 $r_A = \sup_I \left\{ \frac{C^A(I)}{C^*(I)} \right\}$ 为算法A的最坏情况比（**worst-case ratio**）
  - 若算法A的最坏情况比为r_A，则对该问题的任意实例I，均有 $C^A(I) \le r_A C^*(I)$
  - 最坏情况比越接近于1，说明算法给出的可行解目标值越接近于最优值，算法近似性能越好

最坏情况比

- 最坏情况比的证明方法
  - （上界）A 的最坏情况比不超过 r：对任意实例 I，$C^A(I)\le rC^*(I)$
    - 对很多优化问题，不可能得到 $C^*(I)$的简单表达式。为此，给出$C^*(I)$的一些易计算的下界$C^*_{LB}(I)$，即 $C^*(I)\ge C^*_{LB}(I)$，并证明 $C^A(I)\le rC^*_{LB}(I)$
  - （下界）A 的最坏情况比至少为 r：构造实例 I或一族实例 I_n，使得 $\dfrac{C^A(I)}{C^*(I)}=r$ 或 $\dfrac{C^A(I_n)}{C^*(I_n)}\to r(n\to\infty)$
  - 若上界与下界相等，称界为紧（**tight**）的。若两者不相等或只得到其中一个，算法近似性能无法完全确定

背包问题的近似算法

组合优化

- 对极大化目标的优化问题，定义 A的最坏情况比为 $r_A = \inf\{r \ge 1 \mid C^*(I) \le rC^A(I), \forall I\}$
- 基于贪婪（**Greedy**）思想的算法
  - 将物品按价值密度非增的顺序排列，即有 $\frac{p_1}{w_1} \ge \frac{p_2}{w_2} \ge \cdots \ge \frac{p_n}{w_n}$
  - 按上述顺序将物品依次放入背包，直至第一个不能放入的物品为止
- 记 $j = \min\left\{k \mid \sum_{i=1}^{k} w_i > C\right\}$，算法将前 $j-1$个物品放入背包 $C^G = \sum_{i=1}^{j-1} p_j$

物品	1	2
价值	2	$2M$
大小	1	$2M$
背包容量 $2M$		

$C^G(I_1) = 2$，$C^*(I_1) = 2M$

$$\frac{C^*(I_1)}{C^G(I_1)} = M \to \infty (M \to \infty)$$

证明最坏情况比之前可通过构造实例对最坏情况比作出估计

背包问题的近似算法

- 算法改进
  - 将物品 j 之后可以放入背包的物品放入背包

改进无效

- 基于复合思想的改进算法
  - 运行基于贪婪思想的算法 G
  - 将 G 所得目标值与最大物品价值 $p_{\max}$ 进行比较，取优者作为输出

$$C^C(I) = \max\left\{\sum_{i=1}^{j-1} p_j, p_{\max}\right\}$$

物品	1	2
价值	2	$2M$
大小	1	$2M$
背包容量 $2M$		

$$C^G(I_1) = 2，\ C^*(I_1) = 2M$$

$$\frac{C^*(I_1)}{C^G(I_1)} = M \to \infty (M \to \infty)$$

复合算法

- **复合算法的最坏情况比为 2**
  - 用松弛线性规划的最优值作为最优值的上界

$$C^*(I) \le \sum_{i=1}^{j-1} p_i + \frac{p_j}{w_j}\left(C - \sum_{i=1}^{j-1} w_i\right)$$
$$\le \sum_{i=1}^{j-1} p_i + \frac{p_j}{w_j} w_j \le \sum_{i=1}^{j-1} p_i + p_{\max}$$

$$\frac{C^*(I)}{C^C(I)} \le \frac{\sum_{i=1}^{j-1} p_i + p_{\max}}{\max\left\{\sum_{i=1}^{j-1} p_i, p_{\max}\right\}} \le 2$$

物品	1	2	3
价值	1	M	M
大小	1	M	M
背包容量 $2M$			

$$C^C(I) = M+1, C^*(I) = 2M$$

$$\frac{C^*(I)}{C^C(I)} = \frac{2M}{M+1} \to 2 (M \to \infty)$$

$$C^C(I) = \max\left\{\sum_{i=1}^{j-1} p_j, p_{\max}\right\} \quad j = \min\left\{k \mid \sum_{i=1}^{k} w_i > C\right\}$$

最坏情况分析

- 最坏情况比的特点
  - 对问题的不同实例，算法给出的近似解目标值与最优值的比值不全相同，最坏情况界仅是这些比值的最大值，无法充分反映算法性能的全貌
  - 若算法的某种改变仅能改进部分实例的可行解，而不能改进最坏情况实例的可行解，算法的最坏情况比不会减小
- 算法设计与最坏情况比证明思路
  - 算法给出的可行解目标值的恰当描述与最优值的合适估计
  - 最坏情况比上界和下界证明互为借鉴
  - 善于构造典型实例，重视其在算法设计与改进中的作用
  - 精心思考、勇于尝试、反复比较、不断修正

平均情况比

组合优化

- 假设实例中的数据服从一定概率分布，$\frac{C^A(I)}{C^*(I)}$为一随机变量，其期望$\mathbb{E}\left(\frac{C^A(I)}{C^*(I)}\right)$即为算法 A 的平均情况比（**average-case ratio**）
  - 平均情况比依赖于所假设的实例中数据所服从的概率分布
  - 在多数情况下，缺乏足够的依据来判断一个实例中的数据来自何种分布
  - 即便已得到某一算法在某个分布下的平均情况比，对来自该分布的一个实例 I，也无法对$\frac{C^A(I)}{C^*(I)}$的值作出估计
  - 平均情况比的证明比最坏情况比更为复杂

TSP的难近似性

组合优化

- 对**TSP**问题，不存在最坏情况比为有限常数的多项式时间近似算法，除非$\mathcal{P}=\mathcal{NP}$
  - 设**TSP**问题存在最坏情况比小于M的多项式时间算法 A
  - 任给$\mathcal{NP}$–完全问题**HC**的实例I_{HC}：图 $G=(V,E)$。构造**TSP**实例I_{TSP}如下：城市数 $n=|V|$，城市之间距离

$$c_{ij}=\begin{cases}1, & 若 v_iv_j\in E,\\ nM-n+1, & 若 v_iv_j\notin E。\end{cases}$$

    - 若 G 中存在**Hamilton**圈，则 $C^*(I_{TSP})=n$
    - 若 G 中不存在**Hamilton**圈，则任一**TSP**环游中至少有一对距离为 $nM-n+1$ 的相邻城市， $C^*(I_{TSP})\geq(n-1)+(nM-n+1)=nM$

TSP的难近似性

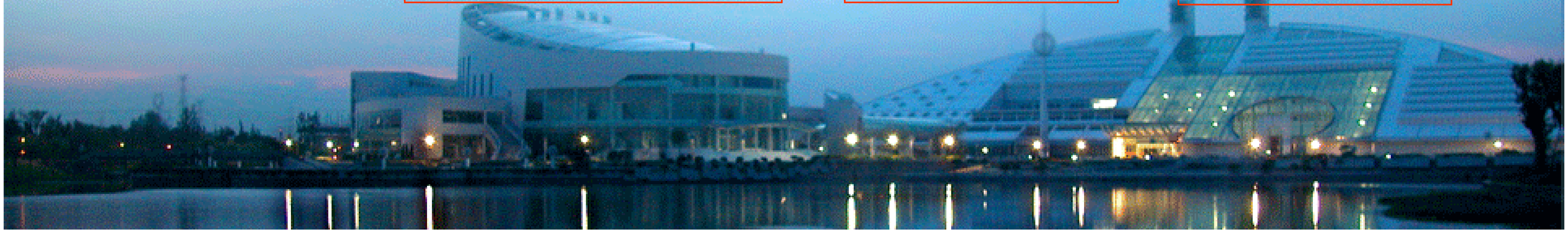
组合优化

- 用算法 A 求解实例 I_{TSP}，则 $\dfrac{C^A(I_{TSP})}{C^*(I_{TSP})} < M$
  - 若 $C^A(I_{TSP}) \ge nM$，则 $C^*(I_{TSP}) > \dfrac{C^A(I_{TSP})}{M} \ge n$，$I_{HC}$的答案为“否”
  - 若 $C^A(I_{TSP}) < nM$，则 $C^*(I_{TSP}) \le C^A(I_{TSP}) < nM$，$I_{HC}$的答案为“是”
  - 由 A 可给出**HC**问题的多项式时间算法

矛盾

$\boxed{C^*(I_{TSP}) = n}$ $\Leftarrow$ $\boxed{I_{HC}\text{ 的答案为“是”}}$ $\Leftarrow$ $\boxed{C^*(I_{TSP}) < nM}$ $\Leftarrow$ $\boxed{C^A(I_{TSP}) < nM}$

$\boxed{C^*(I_{TSP}) \ge nM}$ $\Leftarrow$ $\boxed{I_{HC}\text{ 的答案为“否”}}$ $\Leftarrow$ $\boxed{C^*(I_{TSP}) > n}$ $\Leftarrow$ $\boxed{C^A(I_{TSP}) \ge nM}$

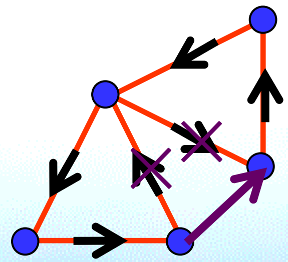


度量TSP

组合优化

- 度量TSP（metric TSP）：城市之间的距离满足三角不等式，即 $c_{ij} \leq c_{ik} + c_{kj}$
- 基于相关的多项式可解问题算法设计近似算法

经过所有城市	长度最短	环游
Hamilton圈	$\mathcal{NP}$-完全	×
最小生成树	不构成环游	×
↓ ？		
Euler环游	顶点经过多次	×
↓ 抄近路		
TSP环游		



最小生成树加倍算法

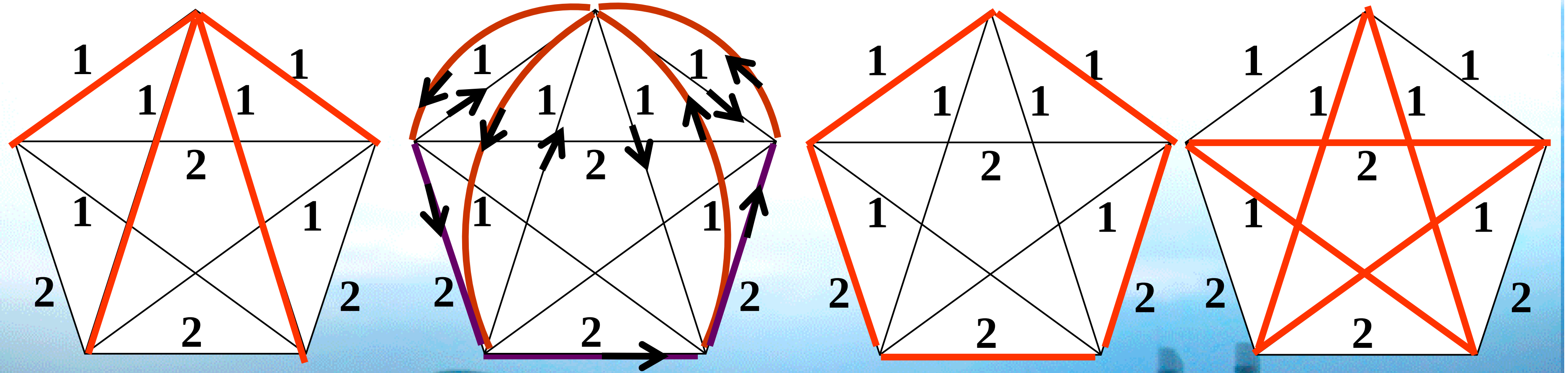
组合优化

- 最小生成树加倍算法
  - 构造赋权完全图 K_n，其中 n 为城市数，边 $v_i v_j$ 的权为相应的城市 i 和 j 之间的距离
  - 求 K_n 的最小生成树 T
  - 将 T 中的每条边变为两条有相同起点和终点，权相同的边，得到一**Euler**图 G_1
  - 求 G_1 的一条**Euler**环游 C_1
  - 用“抄近路”的方法将 C_1 改造为一条**TSP**环游

最小生成树加倍算法

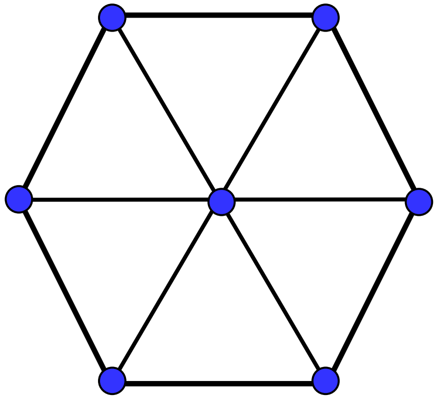
- **最小生成树加倍算法的最坏情况比不超过 2**
  - 记 l_T 为 K_n 的最小生成树长度，则 $C^*(I) \geq l_T$，

$$C^{DM}(I) \leq 2l_T \leq 2C^*(I)$$

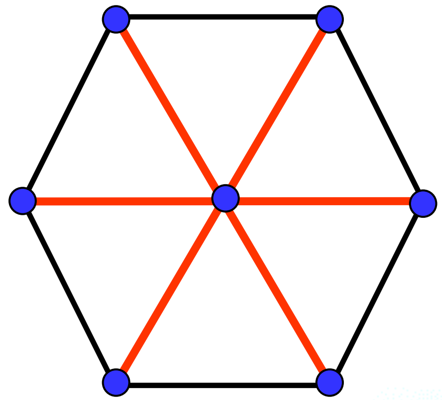


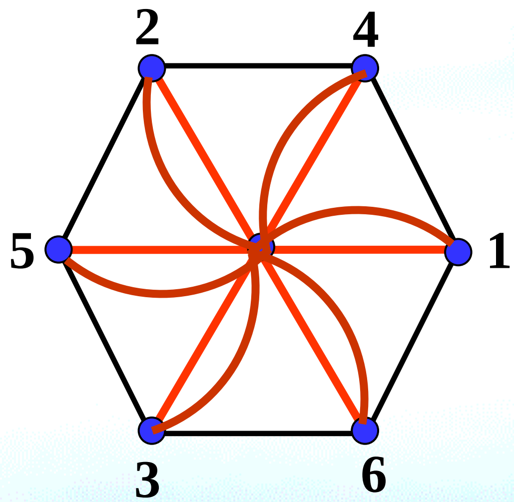
最小生成树加倍算法

- 最小生成树加倍算法的最坏情况比至少为 **2**



城市间距离在图中显示的为 **1**，未显示的为 **2**

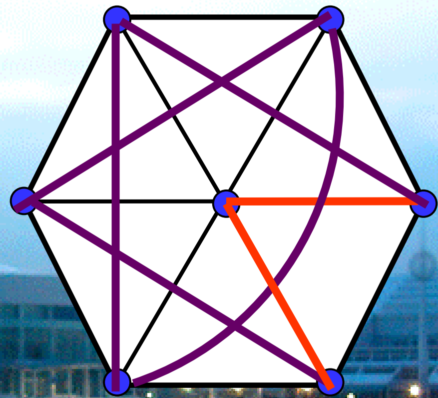


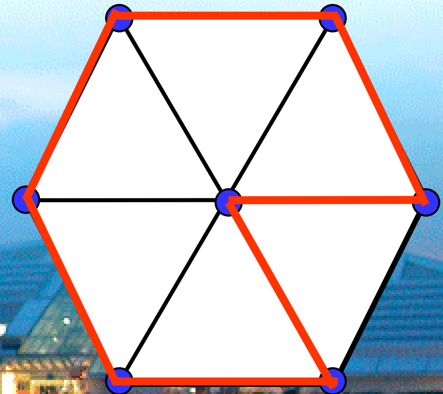


$$C^{DM}(I)=10 \qquad C^{DM}(I_n)=2n-2$$

$$C^{*}(I)=7 \qquad C^{*}(I_n)=n$$

$$\frac{C^{DM}(I_n)}{C^{*}(I_n)}=\frac{2n-2}{n}\rightarrow 2(n\rightarrow\infty)$$





最小权完美匹配

- 为将 K_n 的最小生成树 T 改造为**Euler**图，只需将 T 的奇度顶点两两相连
- 匹配
  - 给定图 $G=(V,E)$，E 的非空子集 M 称为 G 的一个匹配（**matching**），若 M 中任意两条边在 G 中均没有公共端点
  - 若 G 中所有顶点都与匹配 M 中某条边关联，则称 M 为完美匹配（**perfect matching**）
  - 偶数个顶点的完全图必存在完美匹配，任一完美匹配将图中顶点两两相连
  - 赋权图中总权和最小的完美匹配称为最小权完美匹配
- 任意图的最小权完美匹配 $\in \mathcal{P}$

Edmonds, J., Paths, trees and flowers. *Canadian Journal of Mathematics*, 17, 449–467, 1965.

Edmonds, J., Maximum matching and a polyhedron with (0, 1) vertices. *Journal of Research of the National Bureau of Standards* B, 69, 125–130, 1965.

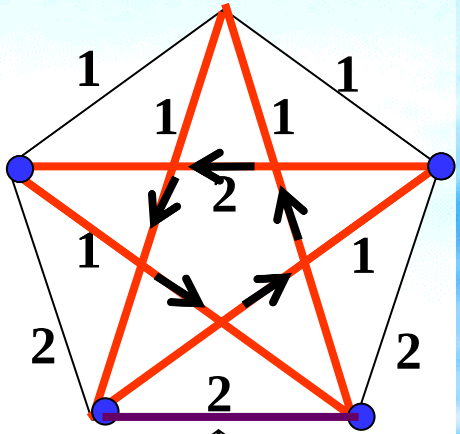
Christofides算法

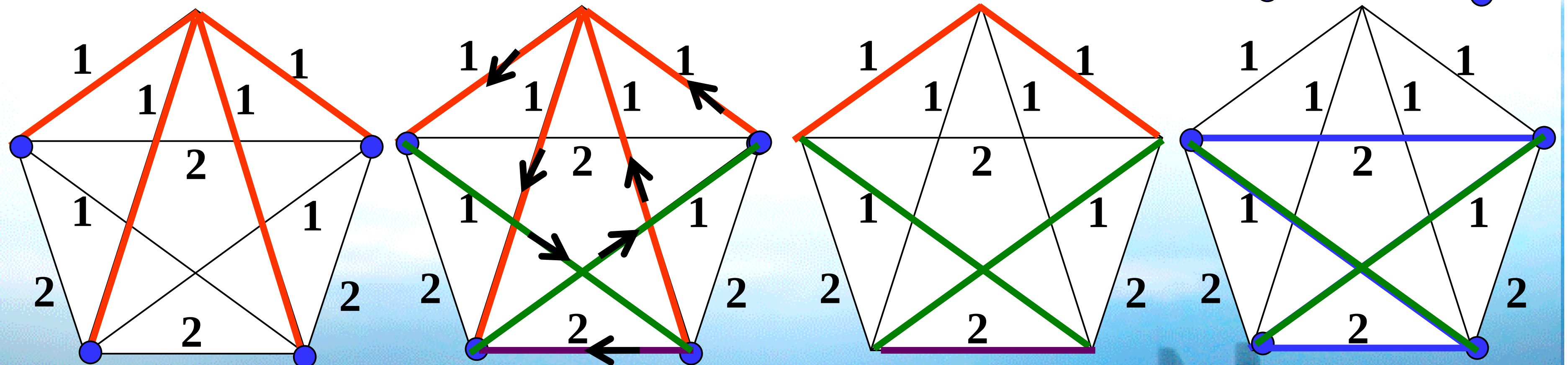
组合优化

- **Christofides算法**
  - 构造赋权完全图 K_n，其中 n 为城市数，边 $v_i v_j$ 的权为相应的城市 i 和 j 之间的距离
  - 求 K_n 的最小生成树 T
  - 求以 T 的奇度顶点集为顶点集，边的权与 K_n 中对应边相同的赋权完全图 H 的最小权完美匹配，将该匹配中的边加入 T，得到一**Euler**图 G_2
  - 求 G_2 的一条**Euler**环游 C_2
  - 用**“**抄近路**”**的方法将 C_2 改造为一条**TSP**环游

Christofides算法

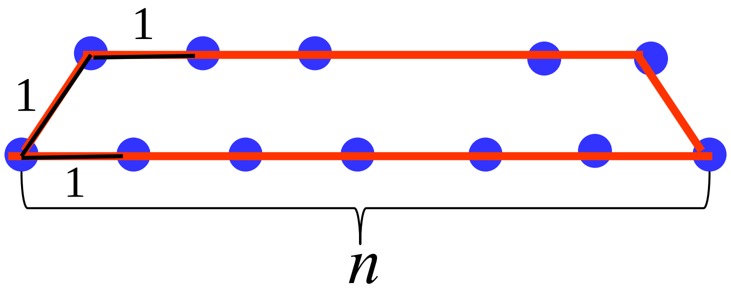
- **Christofides算法的最坏情况比不超过** $\frac{3}{2}$
  - 记 l_M 为 H 的最小权完美匹配的总权和，则 $C^*(I) \geq 2l_M$
  - $C^{Chr}(I) \leq l_M + l_T \leq \frac{3}{2} C^*(I)$



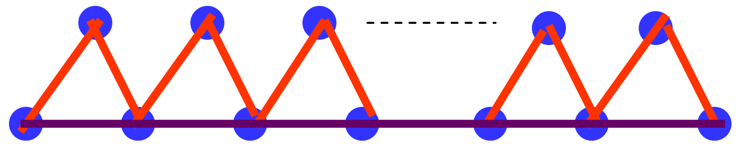


Christofides算法

- **Christofides算法的最坏情况比不小于** $\frac{3}{2}$



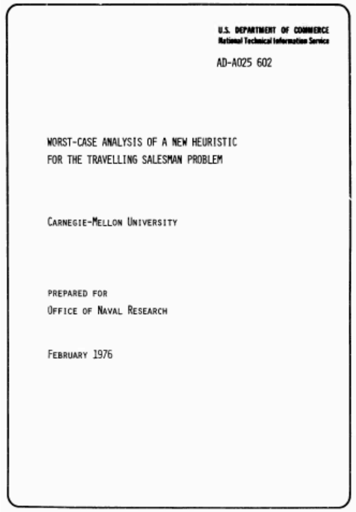
$$C^{*}=2n+1$$



$$C^{Chr}=3n$$

$$\frac{C^{Chr}(I_n)}{C^{*}(I_n)}=\frac{3n}{2n+1}\to\frac{3}{2}(n\to\infty)$$

Christofides算法求解Euclidean TSP的最坏情况比也为 3/2。它仍是至目前为止，度量**TSP最坏**情况比最小的多项式时间近似算法

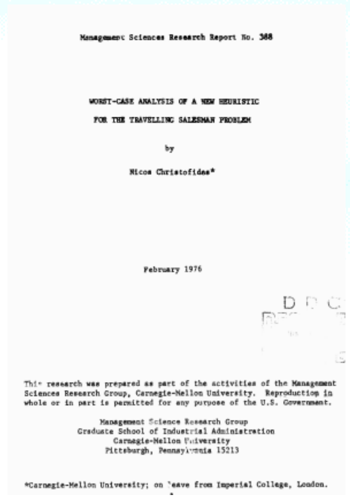

U.S. DEPARTMENT OF COMMERCE
National Technical Information Service
AD-A025 602

WORST-CASE ANALYSIS OF A NEW HEURISTIC
FOR THE TRAVELLING SALESMAN PROBLEM

CARNEGIE-MELLON UNIVERSITY

PREPARED FOR
OFFICE OF NAVAL RESEARCH

FEBRUARY 1976


Management Sciences Research Report No. 388

WORST-CASE ANALYSIS OF A NEW HEURISTIC
FOR THE TRAVELLING SALESMAN PROBLEM

by

Nicos Christofides*

February 1976

This research was prepared as part of the activities of the Management Sciences Research Group, Carnegie-Mellon University. Reproduction in whole or in part is permitted for any purpose of the U.S. Government.

Management Science Research Group
Graduate School of Industrial Administration
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

*Carnegie-Mellon University; on leave from Imperial College, London.

Christofides N., Worst-case analysis of a new heuristic for the travelling salesman problem. Report No. 388. Management Sciences Research Group, Graduate School of Industrial Administration, Carnegie-Mellon University, 1976.

顶点覆盖

浙江大学 ZheJiang University

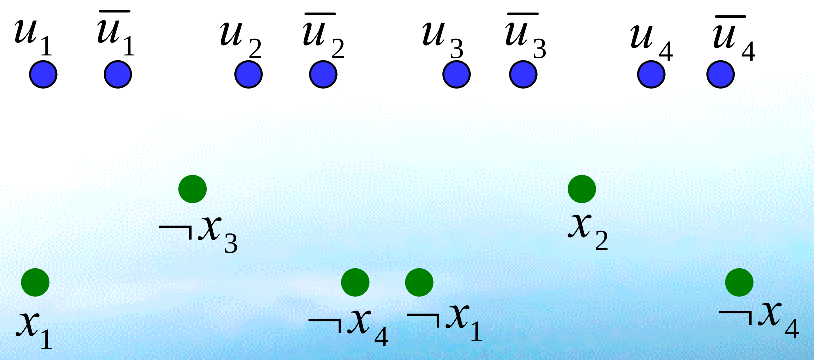
组合优化

- 给定图 $G=(V,E)$，顶点集 V 的子集 V' 称为 G 的顶点覆盖（**vertex cover**），若 E 中每条边至少有一个端点在 V' 中
- 最小权顶点覆盖问题（**WVC**）：给定图 $G=(V,E)$，其中顶点 v_j 的权为 $w_j>0$，求 G 的总权和最小的顶点覆盖
  - 最小顶点覆盖问题（**VC**）：给定图 $G=(V,E)$，求 G 的最小顶点覆盖，即包含顶点数最少的顶点覆盖
- **VC** $\in \mathcal{NP}$-难

$3\mathrm{SAT} \le_m^p \mathrm{VC}$

- 任取**3SAT**问题实例 I_{3SAT}： n 个**Boolean**变量 $x_1, x_2, \cdots x_n$的**CNF** $\bigwedge_{i=1}^{m} c_i$，其中 $c_i = l_{i_1} \vee l_{i_2} \vee l_{i_3}$
- 构造**VC**判定问题实例 I_{VC}：图 $G=(V,E)$，阈值 $M = n + 2m$
  - $V = V_1 \cup V_2, E = E_1 \cup E_2 \cup E_3$
  - $V_1 = \bigcup_{j=1}^{n} \{u_j, \bar{u}_j\}$，其中 $u_j, \bar{u}_j$ 分别对应文字 $x_j, \neg x_j$
  - $V_2 = \bigcup_{i=1}^{m} \{w_{i_1}, w_{i_2}, w_{i_3}\}$，其中 $w_{i_1}, w_{i_2}, w_{i_3}$ 分别对应 c_i 中的**3**个文字 $l_{i_1}, l_{i_2}, l_{i_3}$

$$(x_1 \vee \neg x_3 \vee \neg x_4) \wedge (\neg x_1 \vee x_2 \vee \neg x_4)$$

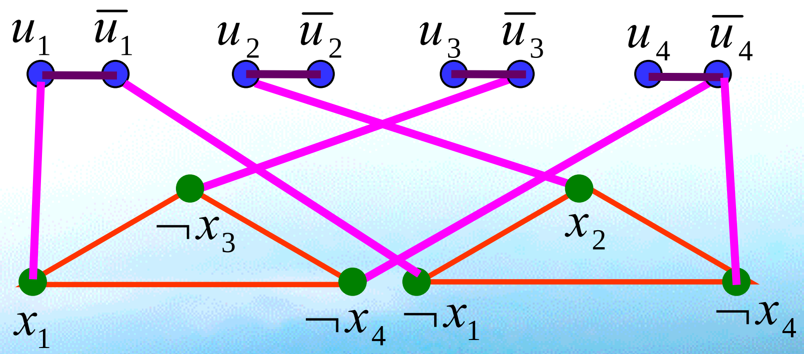


$3SAT \le_m^p VC$

组合优化

- 图 $G=(V,E)$，$E=E_1\bigcup E_2\bigcup E_3$
  - $E_1=\{e_1,e_2,\cdots,e_n\}$，其中 e_j 连接 V_1 中两个顶点 $u_j,\overline{u}_j$
  - $E_2=\bigcup_{j=1}^{m}K_i$， 其中 $K_i=\bigcup_{j=1}^{m}\{w_{i_1}w_{i_2},w_{i_2}w_{i_3},w_{i_3}w_{i_1}\}$
  - $E_3=\bigcup_{j=1}^{m}F_i$， 其中 $F_i=\{f_{i_1},f_{i_2},f_{i_3}\}$，$(x_1\vee\neg x_3\vee\neg x_4)\wedge(\neg x_1\vee x_2\vee\neg x_4)$

    f_{i_k} 连接 V_2 中顶点 w_{i_k} 与 V_1 中对应相同文字的顶点



- G 含有 $2n+3m$ 个顶点和 $n+6m$ 条边，构造可在多项式时间内完成

3SAT $\le_m^p$ VC

浙江大学 ZheJiang University

组合优化

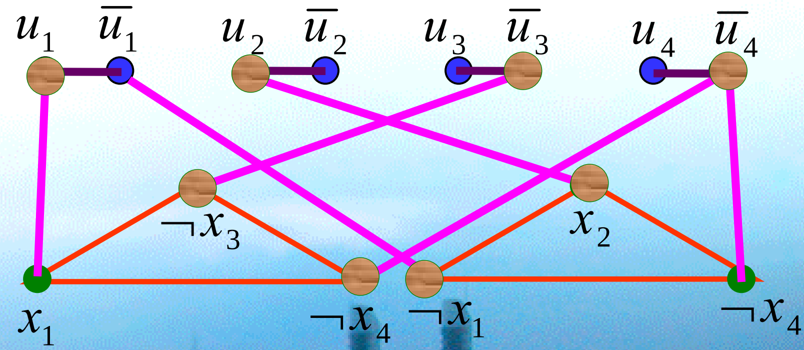
- 若 x_j 取真，则令 $u_j \in V'$，若 x_j 取假，则令 $\bar{u}_j \in V'$
- $\bigwedge_{i=1}^{m} c_i$ 可满足，c_j 可满足，c_j 中至少一文字为真，不妨设 l_{i_1} 为真，则令 $w_{i_2}, w_{i_3} \in V'$
- $u_j, \bar{u}_j$ 中至少有一个顶点在 V' 中，故 e_j 可覆盖
- 由于 $w_{i_2}, w_{i_3} \in V'$, K_i 中任一边可覆盖
- f_{i_1} 在 V_1 中的端点在 V' 中, f_{i_2}, f_{i_3} 在 V_2 中的端点 w_{i_2}, w_{i_3} 在 V' 中

若 I_{3SAT} 答案为“是”，则 I_{VC} 答案也为“是”

若存在变量的一组赋值，使 $\bigwedge_{i=1}^{m} c_i$ 可满足，则 G 存在一顶点覆盖 V', 且 $|V'| \le n+2m$

$$(x_1 \vee \neg x_3 \vee \neg x_4) \wedge (\neg x_1 \vee x_2 \vee \neg x_4)$$

1 $x_1=1$，$x_2=1$，$x_3=0$，$x_4=0$ **1**



$3SAT \le_m^p VC$

组合优化

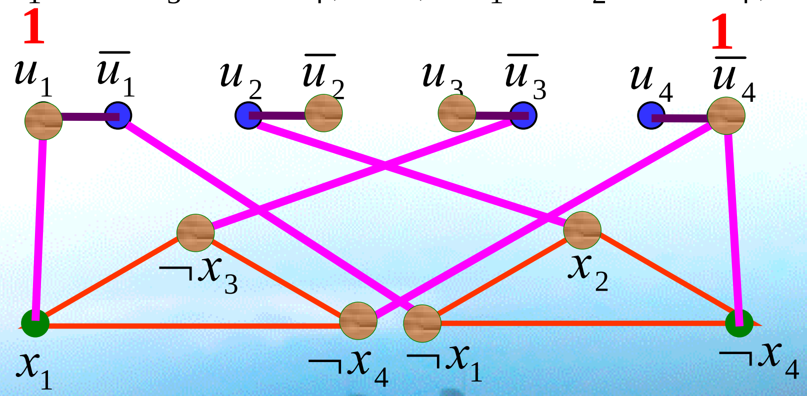
- E_1 中任一边 e_j 可覆盖，e_j 两端点 $u_j, \bar{u}_j$ 至少之一在 V' 中
- E_2 子集 K_i 中边均可覆盖，$w_{i_1}, w_{i_2}, w_{i_3}$ 三个顶点中至少两个在 V' 中
- 由于 $|V'| \le n+2m$，故 $|V' \cap V_1| = n$，$|V' \cap V_2| = 2m$
- E_3 子集 F_i 的三条边中，至多有两条边被 $V' \cap V_2$ 中顶点覆盖，至少有一条边被 $V' \cap V_1$ 中顶点覆盖
- 若 $u_j \in V'$，则令 x_j 取真，若 $\bar{u}_j \in V'$，则令 x_j 取假
- F_i 中被 $V' \cap V_1$ 中顶点覆盖的边的端点对应文字取真，子句 c_i 中至少有一个文字取真，c_i 可满足，$\bigwedge_{i=1}^{m} c_i$ 可满足

若 I_{VC} 答案为“是”，则 I_{3SAT} 答案也为“是”

若 G 存在一顶点覆盖 V'，

且 $|V'| \le n+2m$，则存在变量的一组赋值，使 $\bigwedge_{i=1}^{m} c_i$ 可满足

$$(x_1 \vee \neg x_3 \vee \neg x_4) \wedge (\neg x_1 \vee x_2 \vee \neg x_4)$$



$x_1=1$，$x_2=0$，$x_3=1$，$x_4=0$

整数规划

- 最小权顶点覆盖问题的整数规划
  - 决策变量 $x_i = \begin{cases} 1, & 若 v_i \in V' \\ 0, & 其它 \end{cases}$
  - 整数规划（**IP**）

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{n} w_i x_i \\
s.t. \quad & x_j + x_k \geq 1, \quad \forall v_j v_k \in E \\
& 0 \leq x_i \leq 1
\end{aligned}
$$

最小权顶点覆盖的最优值即为（IP）的最优值

E 中任一条边的两个端点中至少有一个在 V' 中

  - 松弛线性规划（**LP**）

线性规划松弛算法

浙江大学 ZheJiang University

组合优化

- 记（IP）的最优解为 $\mathbf{x}^* = (x_1^*, x_2^*, \cdots, x_n^*)^T$，（LP）的最优解为 $\mathbf{x}^{LP} = (x_1^{LP}, x_2^{LP}, \cdots, x_n^{LP})^T$
- 线性规划松弛算法
  - 求解（LP）并令 $V' = \left\{ v_i \mid x_i^{LP} \geq \frac{1}{2} \right\}$
  - 线性规划松弛算法的最坏情况比不超过2
  - V'是G的一个顶点覆盖
    - 对E中任意边$v_j v_k$，若$v_j \notin V', v_k \notin V'$，则 $x_j^{LP} < \frac{1}{2}, x_k^{LP} < \frac{1}{2}$，故 $x_j^{LP} + x_k^{LP} < 1$，$\mathbf{x}^{LP}$ 不是（LP）的可行解
  - $\sum\limits_{v_i \in V'} w_i \leq 2 \sum\limits_{v_i \in V'} w_i x_i^{LP} \leq 2 \sum\limits_{i=1}^{n} w_i x_i^{LP} \leq 2 \sum\limits_{i=1}^{n} w_i x_i^*$

V'定义　　　　松弛性质

可近似性与难近似性

浙江大学 ZheJiang University

组合优化

- WVC的算法
  - Local ratio算法：$2-\frac{\log_2\log_2|V|}{2\log_2|V|}$
  - SDP（semidefinite programming）松弛算法：$2-\Theta\left(\frac{1}{\sqrt{\log_2|V|}}\right)$
- VC的难近似性
  - 若 $\mathcal{P}\neq\mathcal{NP}$，不存在最坏情况比小于 $10\sqrt{5}-21\approx1.3606$ 的多项式时间近似算法

Bar-Yehuda R, Even S. A local-ratio theorem for approximating the weighted vertex cover problem. *Annals of Discrete Mathematics*, 25, 27-45, 1985.

Bar-Yehuda R, Bendel K, Freund A, Rawitz D. Local ratio: A unified framework for approximation algorithms, in memoriam: Shimon even 1935-2004. *ACM Computing Surveys*, 2004, 36, 422-463.

Karakostas G. A better approximation ratio for the vertex cover problem. *ACM Transactions on Algorithms*, 2009, 5, Article No. 41.

Dinur I., Safra S., On the hardness of approximating vertex cover, *Annals of Mathematics*, 162, 439-485, 2005.


浙江大学
ZheJiang University
谢 谢