Graph Approximation using MST

Zhengyu Jin

Department of Computer Science and Technology
Zhejiang University

2025 年 11 月 26 日

Overview

1. Steiner Tree Problem
2. TSP Problem
3. Shortest Hamiltonian Path

1. Steiner Tree Problem

Introduction

定义 (Steiner 树问题)

给定一个带权图 $G = (V, E)$ 和顶点子集 $R \subseteq V$（称为终端点），寻找一棵权重最小的树，使其连接 R 中的所有顶点。该树可以包含 $V \setminus R$ 中的其他顶点（称为 Steiner 点）以降低总权重。

图: Steiner 树问题示例

Introduction

为了简单起见，我们考虑完全图上满足度量的斯坦纳树问题。

给定无向完全图 $G=(V,E)$（任意两个顶点之间都存在边），并为边赋予非负权值 $c:E\to\mathbb{R}^+$；此外，这个权值需要满足度量要求：设 $(i,j)\in E$ 代表顶点 i 到 j 的边，并记 $c(i,j)$ 等价于该边对应权值，那么 c 应该满足度量性质（类比于泛函分析中的定义）：对称、非负、有限，且满足三角不等式：对任意的 $x,y,z\in V$:

$$c(x,y)\leq c(x,z)+c(z,y)$$

度量斯坦纳树问题的表述

那么 G 上的斯坦纳树问题可以表述为：给定 V 的一个子集 $S \subseteq V$，求 G 的连通子图 $T = (S', E')$，使得 S 中的所有点都在 T 中（即 $S \subseteq S' \subseteq V$），且 T 的边权之和

$$c(T) := \sum_{e \in E'} c(e)$$

最小。S 中的点便是" 终端"，$S' \setminus S$ 中的点便是斯坦纳点。

不过即使加了这么多的限制条件，它仍然是一个 NP-Hard 问题.

我们可以直接利用 MST 设计近似算法。

定理

设完全图 G 关于 S 的导出子图为 $\widetilde{G}$；直接在图 $\widetilde{G}$ 中求最小生成树，就是斯坦纳树问题的一个近似解。在满足三角不等式的前提下，这是一个 2-近似的算法，且数字 2 是紧的。

证明.

假设给定 S 后子图 $\widetilde{G}$ 上 MST 问题的解为 T、最优解斯坦纳树问题的最优解为 T^*。

证明（续）.

我们将 T^* 中的边都复制一次（在原来的边基础上再添一条相同的边），得到图 T_2^*。T_2^* 中每个顶点的度数都是偶数，因此存在一条欧拉回路 L，易知 $c(L) = c(T_2^*) = 2c(T^*)$。

在图 G 中，从 S 中的某一个点出发沿着回路 L 前进，如果遇到了已经访问过的点或者非 S 中的点（斯坦纳点），则将其跳过。由于 G 是完全图，这种" 跳跃" 或称 short-cutting 总是能做到的。通过这种方式就得到了一条 S 上的路径 P。

算法设计与分析（续）

证明（续）.

例如，如下图所示是 L 的一部分，红色代表 S 中的点，蓝色则是斯坦纳点。由于节点 a 之后遇到了斯坦纳点，我们要 short-cutting 到 b；在到达 c 之后，由于 a 已经走过了，我们要直接跳到 d。由于三角不等式成立，因此通过这种方法得到的路径 P 的长度不会超过 L。

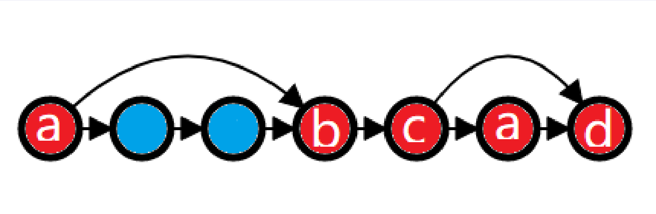


图: Short-cutting 示例

算法设计与分析（续）

证明（续）.

也可以这样理解：给定 T^*，考虑其 DFS 遍历得到的顶点序列，其中每个顶点都被访问了两次；我们仅保留该序列中的终端节点，并且仅保留第一次访问的顶点而去掉第二次访问的节点，构成的序列就得到了路径 P。

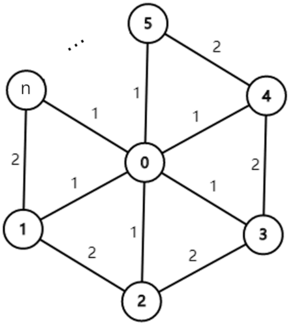
注意到 P 是 $\widetilde{G}$ 的一个生成树，显见

$$c(T) \leq c(P) \leq c(T_2^*) = 2c(T^*)$$

因此 T 便是一个 2-近似解。 $\square$

紧性分析

而且"2" 是紧的，考虑下面的轮图。



令 S 代表序号 $1, 2, \dots, n$ 对应的点，那么利用中间的 0 号点作为斯坦纳点，可以得到最优解 n；但显然利用前面的算法会得到 $2(n-1)$ 的结果。令 n 充分大即可得到 2 的近似比。
(另一方面，这里的 2 也是不能提升的，因为要保证三角不等式成立。)

2. TSP Problem

旅行商问题 (Travelling Salesman Problem, TSP)

我们在之前已经介绍了旅行商 (Travelling Salesman Problem, TSP) 问题，并给出了 ILP 建模。

考虑完全图上的 TSP 问题：给定无向完全图 $G = (V, E)$ 以及边权 $c : E \to \mathbb{R}^+$，求权重最小的哈密顿回路。设 $|V| = n$，那么该问题也就是要确定一个顶点顺序 π 使得

$$\sum_{i=1}^{n} c\left(v_{\pi(i)}, v_{\pi(i+1)}\right) \text{ 最小，其中} \pi(n+1) = \pi(1)$$

TSP 的不可近似性

完全图上的 TSP 是 NP-难问题，在 P $\neq$ NP 的假设下，有下述定理：

定理

完全图上的 *TSP* 不存在近似比为 $2^{poly(n)}$ 的多项式算法，其中 $poly(n)$ 表示关于 n 的多项式，用作 *TSP* 输入的完全图规模为 n（即 $|V| = n$）。

TSP 的不可近似性：证明

证明.

考虑简单无向图 $G' = (V, E')$，我们使用它的顶点构造完全图 $G = (V, E)$，$e \in E$ 的边权 $c(e)$ 定义如下：

$$c(e) = \begin{cases} 1 & e \in E' \\ 2^{\mathrm{poly}(n)} n & e \notin E' \end{cases}$$

G 的规模仍然是 n 的多项式。

TSP 的不可近似性：证明（续）

证明（续）.

用反证法，假设存在 TSP 的 $2^{\mathrm{poly}(n)}$ 的多项式近似算法，利用其求解 G 上的 TSP 问题。若 TSP 的最优解为 n，则 G' 中一定存在一个 Hamilton 回路，近似算法的目标函数值一定不超过 $2^{\mathrm{poly}(n)}n$；否则 G' 中无 Hamilton 回路，最优解一定不小于 $2^{\mathrm{poly}(n)}n+n-1$，近似算法的解也不会更好。

这样一来，由这种近似算法求解后，根据近似解与 $2^{\mathrm{poly}(n)}n$ 的大小关系，就可以解决 G' 的 Hamilton 回路问题。但 Hamilton 回路是一个 NP-难问题，不能被归约到多项式时间求解，矛盾！ $\square$

度量 TSP (Metric TSP)

既然普通完全图上的旅行商问题这么难，我们给它加一点限制：在 c 满足度量性质（主要是三角不等式）的完全图上（所谓 Metric TSP），TSP 就有很好的近似比，类似于斯坦纳树问题，用 MST 松弛的思想可以立即得到一个 2-近似算法。证明与斯坦纳树问题完全相同：

1. 先求图 G 的 MST，得 T^*；
2. 将 T^* 的边加倍，得到欧拉图 T_2^*；
3. 沿着 T_2^* 的欧拉回路 L 跳过所有已访问过的点，构造一个 Hamilton 回路 H；
4. $c(H) \leq c(T_2^*) = 2c(T^*) \leq 2c(H^*)$。

紧性分析

其中寻找欧拉回路可以用 DFS 或者 Fleury 算法。在一些资料中，上述算法被称作 Double-tree algorithm。下面构造一个例子说明数字"2" 是紧的：

例

考虑完全图 $G=(V,E)$，$|V|=2n+1$，其中

- v_0 与任何一个顶点之间的距离都是 1;
- v_i 与 v_{i+1} 之间的距离为 1，$i=0,\ldots,2n-1$;
- 其余所有两点间距离都是 2;

紧性分析（续）

例（续）

那么算法构造出的欧拉回路 L 可以是 $v_0 v_1 v_0 v_3 \dots v_0 v_{2n-1} v_0 v_2 v_0 v_4 \dots v_{2n} v_0$，进行 short-cutting 得到的哈密顿回路 H 则是 $v_0 v_1 v_3 \dots v_{2n-1} v_2 v_4 \dots v_{2n} v_0$；

$c(H) = 2(2n-1)+2 = 4n$。而最优解则是按照相邻选择，即 $v_0 v_1 v_2 \dots v_{2n} v_0$，最优权值和为 $2n+1$。令 n 充分大即可得到 2 的近似比。

Christofides 算法

不过，这个算法其实还可以改进。注意到数字"2" 来自于将 MST 的边加倍这一操作，而加倍是为了保证欧拉回路存在。但是，并不一定要这样才能得到一个欧拉图：对于给定的树 T^*，其上度数为奇数的点（奇点）一定有偶数个，我们将这些点两两配对，随之在每一对匹配的点之间增加一条边，那么构成的图中一定只有偶数度数的点（偶点），也就是一张欧拉图了。换言之：

- 求 T^* 中所有奇点的最小权完美匹配 M^*，将 M^* 中的边加入 T^* 中得到 T_e^*，则可以用 T_e^* 上的欧拉回路构造哈密顿回路。

Christofides 算法（续）

注意，我们允许 T_e^* 是一个重图 (multigraph)，也就是说：M^* 和 T^* 中的边都要保留，这可能会造成重边 (multiple edges, 也称平行边)。其中寻找完美匹配可以用带花树算法在多项式时间内求解。

定理 (Christofides 算法)

上述算法的近似比为 1.5。

Christofides 算法：证明

证明.

不妨设 T^* 中有 $2k$ 个奇点 $v_1, v_2, \ldots, v_{2k}$，则

$$\{(v_1, v_2), (v_2, v_3), \ldots, (v_{2k-1}, v_{2k})\} \text{ 和} \{(v_2, v_3), (v_3, v_4), \ldots, (v_{2k}, v_1)\}$$

是两组完美匹配，我们将其记作 M_1 和 M_2。如下图所示，M_1 和 M_2 中的边是交错的。

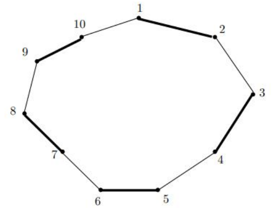


Figure 2: Illustrative instance the two matchings M_1 (thick edges), and M_2 (thin edges)

图: 两个完美匹配 M_1（粗边）和 M_2（细边）

Christofides 算法：证明（续）

证明（续）.

那么

$$c(H^*) \geq c(M_1) + c(M_2) \geq 2c(M^*) \implies c(M^*) \leq \frac{1}{2}c(H^*)$$

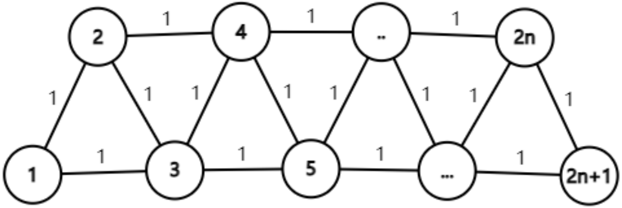
其中第一个不等式里用到了边权的度量性质（相当于在最优的 Hamilton 回路中进行 short-cutting，从而只保留奇点，这些奇点通过取交错边可以得到两个匹配）。另一方面，根据度量性质，对 T_e^* 进行 short-cutting 后得到的边权更小，因此

$$c(H) \leq c(T_e^*) = c(T^*) + c(M^*) \leq \frac{3}{2}c(H^*)$$

□

Christofides 算法的紧性

这就证明了近似比不超过 1.5。下面再给出例子证明 1.5 是紧的，如下图，



Christofides 算法的紧性（续）

紧性分析

这是一个完全图，未画出的顶点间距离由所给边下的最短距离确定，例如 $c(v_2, v_{2n}) = n - 1$、$c(v_1, v_{2n+1}) = n$。易知 $c(H^*) = 2n + 1$（按照给边走一圈即可）；而 $c(T^*) = 2n$（取路径 $v_1 v_2 \ldots v_{2n+1}$）、$c(M^*) = n$（奇点只有 v_1、v_{2n+1}，其间距离为 n），而且 T^* 与 M^* 二者的边正好构成回路 H。令 $n \to \infty$，于是

$$\frac{c(H)}{c(H^*)} = \frac{c(T) + c(M^*)}{c(H^*)} = \frac{3n}{2n+1} \to \frac{3}{2}$$

3. Shortest Hamiltonian Path

最短 Hamilton 路径问题

Hamilton 路径问题指在赋边权的无向完全图 $G = (V, E)$ 上求边权和最小的 Hamilton 路径（而非回路）。同样，我们也只考虑边权 c 满足度量的情形。

在该问题中，我们可以指定路径的端点（endpoints, 包括起点和终点），也可以不指定。这就出现了三种情形：无固定起点或终点、固定其中一个点、固定起点和终点。

相比于 TSP 问题，该问题不要求路径闭合，因此也称其 TSP Path 问题；我们总可以将其转换为 TSP 问题：

转换为 TSP 问题

- 两端点都不指定的情况。我们增加一个虚设顶点 (dummy vertex)，其到其它所有顶点距离都是 0，即可将其转换成 TSP 问题。在 TSP 得到的回路中删除虚设顶点，即可得到一条 Hamilton 路径。
- 指定一个端点的情况。假设指定起点为 s，那么可以增加两个虚设顶点 v, w，其中 v 仅与 w 和 s 相连，而 w 与所有点相连；这些边权值都是 0。在新图上求解 TSP，得到的回路中一定包含"$w - v - s$"，因此将 v, w 删除后，s 就成了端点。
- 指定起止点为 s 和 t。可以仅增加一个虚设顶点，其到 s, t 的距离为 0，而到其余顶点距离为无穷大。

转换为 TSP 问题

但是，前面的这种转换（或称" 规约"）会破坏度量性质，因此我们也没法再使用 Christofides 算法求解。因此这种思路行不通。

无固定情形

首先尝试模仿前面 Christofides 算法里利用 MST 松驰的思路。

1. 求 G 的 MST，得到 T^*；
2. 求 T^* 的奇点（除两个点之外）的最小权完美匹配 M^*，未匹配的两个点指定为起点和终点；
3. 将 M^* 的边加入 T^* 中得到欧拉图 T_e^*，对其中欧拉路径进行 short-cutting（跳过所有已访问的顶点），得到 Hamilton 路径 P。

该算法也是 1.5 近似的，证明思路与 TSP 基本相同：

无固定情形：证明

证明.

假设最优解 P^* 为，MST 上有 $2k$ 个奇点，那么可以拆成两个去除了两点的完美匹配：

$\{(v_1, v_2), (v_3, v_4), \ldots, (v_{2k-3}, v_{2k-2})\}$ （v_{2k-1} 和v_{2k} 没有匹配），以及 $\{(v_2, v_3), (v_4, v_5), \ldots, (v_{2k-2}, v_{2k-1})\}$ （v_1 和v_{2k} 没有匹配），记为M_1 和M_2。于是 $2c(M^*) \leq c(M_1) + c(M_2) \leq c(P^*)$，进而有

$$c(P) \leq c(M^*) + c(T^*) \leq \frac{3}{2}c(P^*)$$

□

无固定情形：求解完美匹配

那么，如何求解去掉任意两点的最小权完美匹配问题呢？直观上可以暴力搜索，枚举所有可能的起止点对，在 $O(n^2)$ 个匹配中寻找最好的结果。也可以用这种方法：引入两个虚设顶点 a, b，这两个点与所有奇点连接，且边权为 0，则在最小权完美匹配中 a, b 一定各自匹配到一个奇点，这相当于去掉了任意两点后的完美匹配最优解。

固定一个端点

接下来增加条件：要求固定某个点 $s \in V$ 为起点。注意到：在构造欧拉回路时，只要 s 是奇点，它就一定会变成端点。因此，如果在 G 的 MST T^* 中 s 本身就是一个奇点，那么只要让它不被匹配即可；否则 s 是一个偶点，我们要想办法让它变成奇点。

为了后续叙述简单，定义 T^* 中的" 误点（vertices of wrong degree）" 是满足下面两个条件之一的点：

- $V \setminus \{s\}$ 中的奇点；
- 若 s 是偶点，则 s 也是误点。

固定一个端点（续）

这里的奇点和偶点都是相对于图 G 的 MST T^* 而言的。显然误点一定有奇数个：若 s 在 T^* 中是奇点，那么 $V\setminus\{s\}$ 中的奇点有奇数个，这些奇点都是误点；若 s 在 T^* 中是偶点，那么 $V\setminus\{s\}$ 中的偶数个奇点与 s 都是误点，加起来仍为奇数个。

我们在所有误点中求除一个顶点（不包含 s）外的最小权完美匹配 M^*。将 M^* 中的边加到 T^* 中得到图 T_e^*。考虑其中起点 s 的两种情况：

- 若 s 在 T^* 中为奇点，则 s 不是误点，其在 T_e^* 中也是奇点（此时 s 和求完美匹配时去掉的点直接分别构成起点和终点）；
- 如果 s 在 T^* 中的度数是偶数，则 s 是误点，

固定一个端点（续）

- 若 s 与其它一个误点匹配，会被多添一条边从而变成奇点；
- 若 s 没有与其它点匹配，说明没有其它误点，我们只需删去与 s 相连的任一条边即可将 s 变成奇点。

因此，我们总能让 s 在 T_e^* 中变成奇点，这样一来它在最终路径里一定会变成端点了。对 T_e^* 中的欧拉路径进行 short-cutting，即可得到 Hamilton 路径 P。该算法仍是 1.5 近似的，如下：

固定一个端点：证明

证明.

设最优解为 P^*，G 的 MST 中有 $2k+1$ 个误点，$v_1, v_2, \dots, v_{2k+1}$，有

$$M_1 = \{(v_1, v_2), (v_2, v_3), \dots, (v_{2k-1}, v_{2k})\}$$ （去除了 v_{2k+1}）和

$$M_2 = \{(v_2, v_3), (v_3, v_4), \dots, (v_{2k}, v_{2k+1})\}$$ （去除了 v_1）是两个去除一个点后的完美匹配，因此 $c(P^*) \geq c(M_1) + c(M_2) \geq 2c(M^*)$，进而有

$$c(P) \leq c(M^*) + c(T^*) \leq \frac{3}{2}c(P^*)$$

□

固定两个端点

接下来，我们设计算法求固定起止点 s, t 的最大权哈密顿路径，并定义此时 T^* 中的"误点" 为

- $V \setminus \{s, t\}$ 中的奇点；
- $\{s, t\}$ 中的偶点。

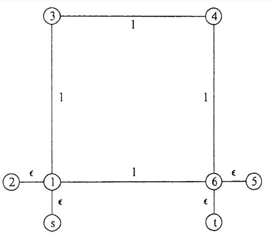
并采用与之前同样的算法。容易知道此时误点一定有偶数个：在误点中做最小权完美匹配得到 M^*，那么在 T^*_e 中 s, t 的度数均为奇数，其余点的度数为偶数，于是存在从 s 出发到 t 的欧拉路径。此外，为了保证 t 是终点，我们要把欧拉路径中所有（可能会）出现的 t 都使用 short-cutting 跳过。

固定两个端点（续）

不过，分析算法的时候与之前出现了一些不同。下面这个例子可以说明近似比至少是 $\frac{5}{3}$：

例

下图是一个 8 个点的完全图，其中未画出的顶点间距离由所给边下的最短距离长确定。ε 是个足够小的正值。



固定两个端点：例子分析

分析

$$\begin{aligned} P^* &= \{(s,1),(1,2),(2,3),(3,4),(4,5),(5,6),(6,t)\} \\ T^* &= \{(s,1),(1,2),(1,3),(1,6),(4,6),(5,6),(6,t)\} \\ M^* &= \{(2,3),(4,5)\} \end{aligned}$$

此时 $c(P^*) = 3 + 6\varepsilon$，但 $c(T^*) = 3 + 4\varepsilon$、$c(M^*) = 2 + 2\varepsilon$，算法会给出 $c(P) = 5 + 6\varepsilon$ 的结果，那么

$$\frac{c(P)}{c(P^*)} = \frac{5+6\varepsilon}{3+6\varepsilon} \to \frac{5}{3}$$

固定两个端点：近似比

定理

对于固定 s, t 的情况，该算法的近似比为 $5/3$。

证明.

设最优解为 P^*，我们将 P^* 中的边加入到 T^* 得到图 Q（这可能是个重图）。由于 P^* 是一条 Hamilton 路径，因此在 Q 中只有起点和终点的奇偶性发生了变化。容易看出，某个顶点在 T^* 是误点当且仅当其在 Q 中是奇点。

固定两个端点：证明（续）

证明（续）.

在路径 P^* 上通过 short-cutting 仅保留所有（偶数多个）误点，其上存在一个误点的完美匹配 E，有 $c(E) \geq c(M^*)$。将 Q 里所有 E 中的边删除，得到图 $Q \setminus E$，这个图中所有点的度数都是偶数；且由于 $E \subseteq P^*$，有 $T^* \subseteq (Q \setminus E)$，即 $Q \setminus E$ 是连通图。故 $Q \setminus E$ 中存在一条欧拉回路，这条回路包含两组误点的完美匹配，即 $c(Q \setminus E) \geq 2c(M^*)$。

于是，$c(T^*) + c(P^*) \geq c(Q) = c(E) + c(Q \setminus E) \geq 3c(M^*)$，进而

$$c(P) \leq c(T^*) + c(M^*) \leq c(T^*) + \frac{1}{3}(c(T^*) + c(P^*)) \leq \frac{5}{3}c(P^*)$$

因此近似比不超过 5/3。而前例说明 5/3 是紧的，因此该算法近似比正是 5/3。 $\square$

The End