Marketing network 解题报告

绍兴市第一中学 任之洲

1 试题大意

给出一张n个点m条边的**随机**带权连通图，若一个子图是合法的需要满足以下条件：

- 这个子图是一个**森林**。
- 对于指定的S个点在这张图中连通。

求合法的**边权和**前k小的子图。

2 数据规模

测试点	n	m	S	k	c_i
1 ~ 5	=10	=20	=4	=10	≤100
6 ~ 10	=50	=100	=10	=1	
11 ~ 15					
16 ~ 20			=5	=20	
21 ~ 25			=7	=50	
26 ~ 30			=9		
31 ~ 35			=10		
36 ~ 40			=11		
41 ~ 45			=13		
46 ~ 50			=15		

3 算法介绍

3.1 算法一

对于测试点$1 \sim 5$，$m \leq 20$。

由于图上的边数较小，可以暴力枚举每条边是否使用，再暴力并查集判断图是否是森林，以及指定的点是否互相连通。

时间复杂度$O(m2^m)$，空间复杂度$O(n+m)$，期望得分 10分。

3.2 算法二

对于测试点$6 \sim 15$，$k = 1$，也就是说只需要求出最优解，是经典的最小斯坦纳树问题。

设$f[i][j]$为当前点在j，连通状态为i的最小费用，其中i为已经连通的点的二进制表示。

转移有两种：

- $f[i][j] = Min(f[k][j] + f[i\ xor\ k][j])\ (0 < k < i)$
- $f[i][j] = Min(f[i][k] + dist(j,k))\ (1 \leq k \leq n)$

第一种转移可以直接暴力枚举，考虑集合$A \subsetneq B \subseteq S$，每个集合$S$中的点$u$的状态共有三种可能$u \in A$、$u \in B - A$、$u \in S - B$，所以暴力枚举的复杂度为$O(n3^S)$。

第二种转移可以套用最短路模型，由于题中所给的图为随机的稀疏图，SPFA算法的期望复杂度为$O(m)$，所以这种转移复杂度为$O(m2^S)$。

时间复杂度$O(n3^S + m2^S)$，空间复杂度$O(n2^S)$，期望得分 30分（结合算法一）。

3.3 算法三

对于测试点$16 \sim 30$，$S \le 9$，$k \le 50$，可以使用A*算法。

把算法二作为估价函数，维护一个优先队列，每次从优先队列中选取（已决定费用+最小费用）最小的状态扩展，依次枚举每条边是否使用，进行$O(km)$次扩展后就能得到前k优解，瓶颈在于计算估价。

由于题中要求构成森林，所以在计算估价之前可以预判当前枚举的边是否可以使用，计算估价的次数为$O(kn)$。

总共计算的状态共$O(kn)$种，所以可以用暴力代替优先队列。

时间复杂度$O(kn^2 3^S + k^2 n^2)$，空间复杂度$O(n2^S + knm)$，期望得分 $50 \sim 60$（并不能通过这个梯度的所有数据）。

3.4 算法四

对于所有测试点，$S \le 15$，$k \le 50$，把算法二作为估价的效率是不可接受的。

在搜索过程中，子状态的S会不断缩小。考虑设定一个阈值，比如说当$S \le 7$时，使用算法二作为估价，当$S > 7$时，设计一个贪心算法作为估价。

有一个很自然的贪心：在图上找一棵最小生成树，再把多余的边删掉（不影响指定的点的连通性）。

设图G的估价为$g(G)$，最优值为$f(G)$。这个贪心的解是不小于最优解的，即$g(G) \ge f(G)$。由于图是随机生成的，做一些样本会发现，在大多数情况下$\frac{f(G)}{g(G)} \ge 0.7$。

为了平衡正确性和速度，可以做一些近似化处理：

- 设定一个阈值t_1，若当前费用$+t_1 g(G)$不能更新解，就舍弃该状态。
- 设定一个阈值t_2，扩展t_2次状态后就直接终止程序。

把边按照权值排成降序后，先确定边权大的边，可以有效减少搜索量。

期望得分 $60 \sim 80$（采用不同的估价和阈值会获得不同的效果）。

3.5 算法五

要通过所有数据需要较靠谱的估价，这里提供一种，流程如下：

- 构造一张新图，新图上只存在被指定的S个点，新图上的边是原图上的最短路。
- 在新图上做一棵最小生成树，选取最小生成树上的边所覆盖的路径。
- 这些边构成的树作为近似解。

构图时需要做S次最短路，选用SPFA算法可以做到期望$O(Sm)$，最小生成树可以选用$O(S^2)$的Prim，提出路径的复杂度为$O(Sn)$，所以这个估价的复杂度为$O(Sm)$。

这个估价对于大多数随机数据可以做到$\frac{f(G)}{g(G)} \geq 0.95$，可以把阈值设定为当前费用$+0.85g(G)$，对于随机数据可以做到不错的正确率。

由于估价很接近正确解，且边权的范围比较小，对于随机数据的平均状态扩展次数是$O(kn)$的。

期望时间复杂度$O(kSnm)$，空间复杂度$O(knm)$，期望得分 $98 \sim 100$。

3.5.1 对于该近似算法的进一步分析

设$leaf$为最小斯坦纳树的叶节点个数，这个近似算法可以做到不超过精确解的$2(1-\frac{1}{leaf})$倍。

设T为最小斯坦纳树，P为斯坦纳树的所有叶子节点按DFS序排序后的序列。

把序列P中相邻两点的路径长度相加（包括P_1到P_{leaf}的路径），相当于把整棵树的每条边遍历了2遍。减去最长的一条路径，路径长度和将不超过整棵树的$2(1-\frac{1}{leaf})$倍。

设T_H为近似算法得到的生成树，树上每条路径都是原图中的最短路，所以T_H的边权和是不大于序列P的路径和的。于是可以得到结论，T_H的边权和不超过T的$2(1-\frac{1}{leaf})$倍。

4 分数估计

- 30分：基本暴力和经典问题，不写挂都很容易拿到。
- 50 ~ 60分：类似WC2015的k小割，应该有半数以上选手能完成。
- 60 ~ 80分：近似化乱搞效果无法估计，可能不会有太多选手尝试。
- 90 ~ 100分：期待能有同学能A题，也期待更优解法。