清华大学考试试题专用纸

考试课程：《算法设计与分析》

系别__________班号__________学号__________姓名____________

注意：共三大题，满分 100 分。

一、（40 分）

1. 叙述稳定匹配问题及 Gale-Shapley 算法。
2. 分析 Gale-Shapley 算法的正确性及计算复杂度。

二、（20 分）

有向图 $G=(\mathrm{V},\mathrm{E})$ 的顶点为 $v_1, v_2, \cdots, v_n$，我们称 G 是有序的，如果它满足：

（1）对于一条边 $(\mathrm{v}_i, \mathrm{v}_j)$ 都有 $i<j$。

（2）除了点 v_n，任何一个顶点都至少有一条出弧。

设计一个算法求从 v_1 到 v_n 的有向最长路，分析算法的正确性及计算复杂度。

三、（40 分）

定义 Hitting set problem：给定 m 个非空集合 $S_i, i=1,\cdots m$，令 $U=\bigcup_{i=1}^{m} S_i$，求 U 的一个最小子集 H，使得 $H \cap S_i \neq \varnothing, i=1,\cdots m$，

1. 证明 Hitting set problem 的判定问题属于 NPC。
2. 对 Hitting set problem 设计一个近似算法，并分析其近似比。

一、 1. 稳定匹配问题: 给定一个集合 $M=\{m_1,..,m_n\}$和一个集合 $W=\{w_1,..,w_n\}$, 和二分图 $G=(M,W,M\times W)$。一个匹配 S 是指有序对 M×W 的一个子集,其中每个 M 和每个 W 的元素最多出现在 S 的一个有序对中。此外,每一个 M 的元素对所有 W 的元素有一个优先度的排序，反之每一个 W 的元素对所有 M 的元素也有一个优先度的排序。一个匹配 S 是稳定的,当它满足 1) 它是完美匹配,即每个 M 和每个 W 的元素正好出现在 S 的一个有序对中, |S|=|M|=|W|=n; 2) S 中不存在不稳定对,即 S 不存在有序对(m,w)和(m',w'),其中 m 把 w'排在 w 前面，而 w'把 m 排在 m'前面。稳定匹配问题是判定给定实例是否存在一个稳定匹配，如果存在，则输出一个。

Gale-Shapley 算法:

```
Initially all m∈M and w∈W are free
While there is a man m who is free and hasn't proposed to
every woman
   Choose such a man m
   Let w be the highest-ranked woman in m's preference list
      to whom m has not yet proposed
   If w is free then
      (m, w) become engaged
   Else w is currently engaged to m'
      If w prefers m' to m then
         m remains free
      Else w prefers m to m'
         (m, w) become engaged
         m' becomes free
      Endif
   Endif
Endwhile
Return the set S of engaged pairs
```

2. 正确性: 先证明 S 是完美匹配，即在 Gale-Shapley 算法返回的匹配中，所有 M 和 W 的元素都在匹配中。

利用反证法，假设有 M 中有一个元素不在 S 中，那么因为|M|=|W|=n，必有 W 中有一个元素也不在 S 中，由于算法中 M 向每一个元素都有提出匹配，矛盾。

接下来证明 S 是稳定匹配,即 Gale-Shapley 算法返回的匹配中没有不稳定对。

考虑不在算法返回的匹配中的一个有序对(m,w)。如果 m 从未向 w 提出匹配，则 m 对它的匹配对象一定优于 w；反之，如果 m 向 w 提出匹配，那一定被 w 拒绝，从而 w 对它的匹配对象一定优于 m，得证。

计算复杂度：证明算法的迭代最多运行 n^2 次。

每个 M 的元素每次都向一个新的 W 的元素提出匹配，最多有 n^2 个可能。

二、题意即有向图是有向无圈图，因此可以应用最短路的算法，在 $O(n^2)$内求得要求的最长路。此外，注意到条件 1 即图已进行拓扑排序，可以写出 Bellman 方程

$$\begin{cases} u_s = u_1 = 0, \\ u_j = min_{i<j}\{u_i + w_{ij}\}. \end{cases}$$

(min 应为 max), u_j 表示从起点到 v_j 的最短路。采用递推计算求解上述方程，每个节点和每条弧均被考虑一次，因此这一最短路算法的复杂度为 O(m+n)=O(m).

三、1. 先写出 Hitting set 问题的判定问题，即。用顶点覆盖规约，构造如下：给定顶点覆盖的任一实例，令 U 对应每一个顶点，S_i 对应每一条边，包含该边中的两个顶点。可证明顶点覆盖问题有一个不大于 K 的顶点覆盖，当且仅当 Hitting set 问题存在一个不大于 K 的 Hitting set。

2. 可以用线性规划舍入，原始对偶，或 local-ratio 法，近似比都为$b_i = |S_i|$，其中$b = \max_{i=1,\dots,m} b_i$。用贪婪算法的话可以得到 log(n)近似比的算法(参见 The design of approximation algorithm, Williamson and Shmoys)。

主要需要证明的有以下几点：

1) 返回的集合是一个 Hitting set（如利用线性规划松弛的性质或原始对偶算法）
2) 近似比 （舍入得到的解<=b 线性规划松弛的解<=b 最优解，由弱对偶性）

以下给出线性规划舍入的方法，其它的都和顶点覆盖的算法类似。（B_j 为集合，ai 为 U 中元素，xi 表示 U 中元素的选取）

Consider the following problem for Linear Programming:

$$\begin{array}{lll} \text{Min} & \sum_{i=1}^{n} w_i x_i & \\ \text{s.t.} & 0 \le x_i \le 1 & \text{for all } i = 1, \dots, n \\ & \sum_{i:a_i \in B_j} x_i \ge 1 & \text{for all } j = 1, \dots, m \text{ (all sets are hit)} \end{array}$$

Let x be the solution of this problem, and w_{LP} is a value of this solution (i.e. $\sum_{i=1}^{n} w_i x_i$).

Now define the set S to be all those elements where $x_i \ge 1/b$:

$$S = \{a_i \mid x_i \ge 1/b\}$$

(1) *S is a hitting set.*

Proof. We want to prove that any set B_j intersects with S. We know that the su
all x_i where $a_i \in B_j$ is at least 1. The set B_j contains at most b elements. Therefore
$x_i \geq 1/b$, for some $a_i \in B_j$. By definition of S, this element $a_i \in S$. So, B_j intersects
S by a_i. ■

(2 *The total weight of all elements in S is at most $b \cdot w_{LP}$.*

Proof. For each $a_i \in S$ we know that $x_i > 1/b$, i.e., $1 < bx_i$. Therefore

$$w(S) = \sum_{a_i \in S} w_i \leq \sum_{a_i \in S} w_i \cdot bx_i \leq b \sum_{i=1}^{n} w_i x_i = bw_{LP}$$

■

(3 *Let S^* be the optimal hitting set. Then $w_{LP} \leq w(S^*)$.*

Proof. Set $x_i = 1$ if a_i is in S^*, and $x_i = 0$ otherwise. Then the vector x satisfy const
of our problem for Linear Programming:

$0 \leq x_i \leq 1$ for all $i = 1, \ldots, n$

$\sum_{i: a_i \in B_j} x_i \geq 1$ for all $j = 1, \ldots, m$ (because all sets are hit)

Therefore the optimal solution is not worse that this particular one. That is,

$$w_{LP} \leq \sum_{i=1}^{n} w_i x_i = \sum_{a_i \in S} w_i = w(S^*)$$

■

Therefore we have a hitting set S, such that $w(S) \leq b \cdot w(S^*)$.