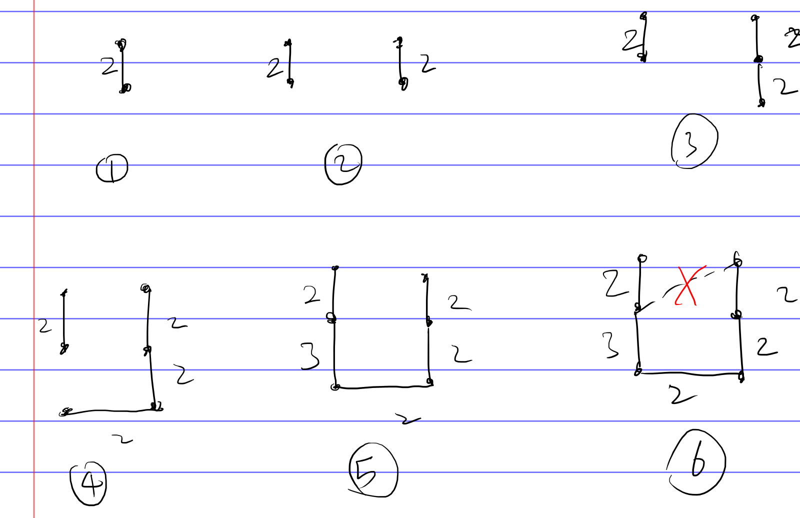
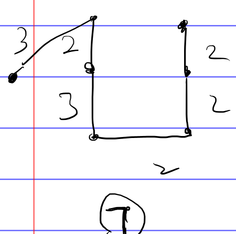
HW6

1. 边按边长排序：

2, 2, 2, 2, 3, 3, 3, 5, 6, 7, 8



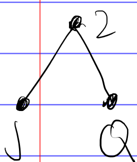


至此最小生成树已经构造完成。

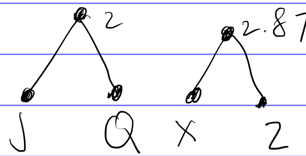
注：上面的构造过程不是唯一的，可以不同。

2. Huffman 编码

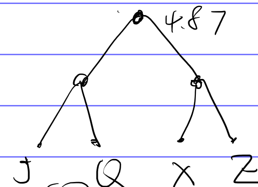
① 以下用字母的权重构造Huffman编码，用百分比也可以。



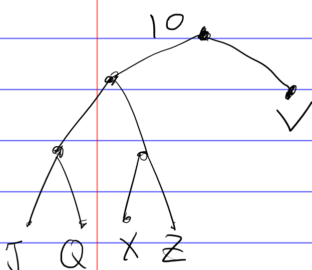
①



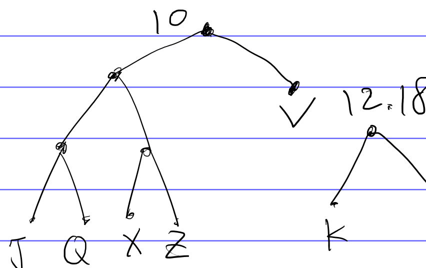
②



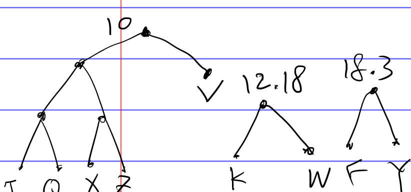
③



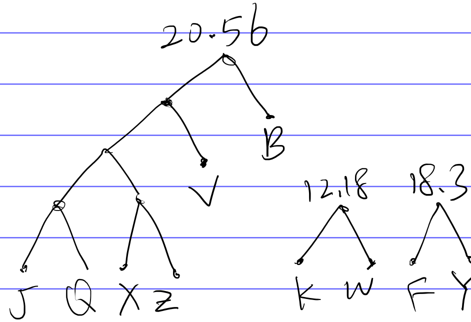
④



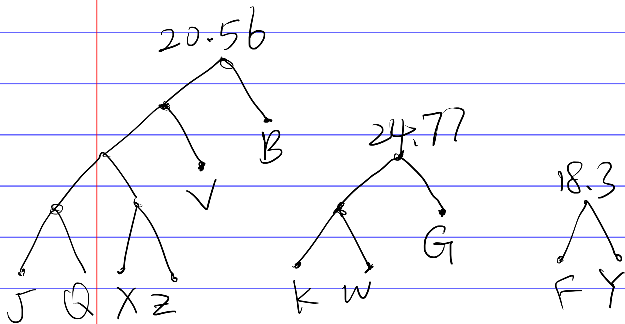
⑤



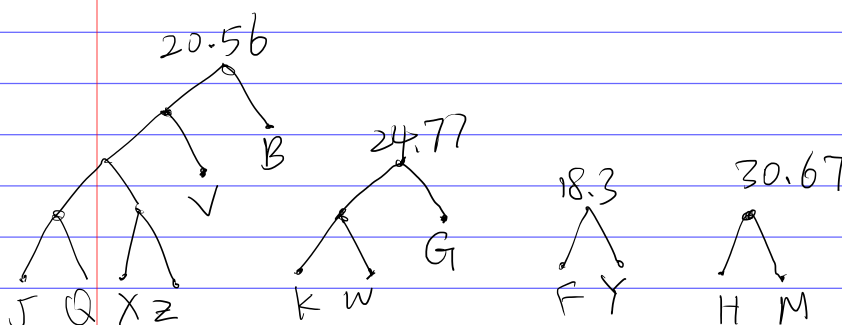
⑥



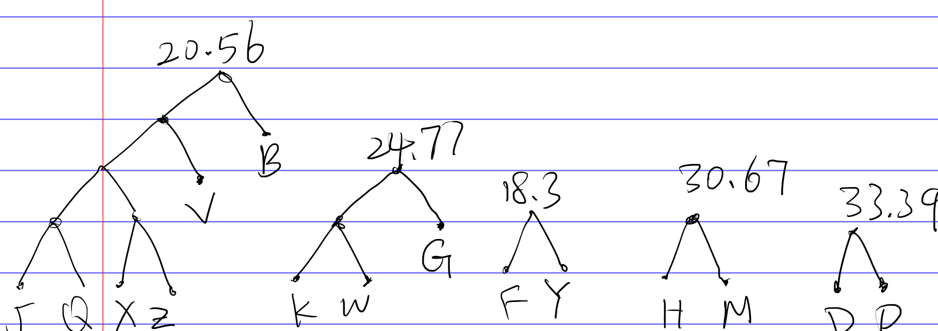
⑦


20.56
B
V
J Q X Z
24.77
G
K W
18.3
F Y

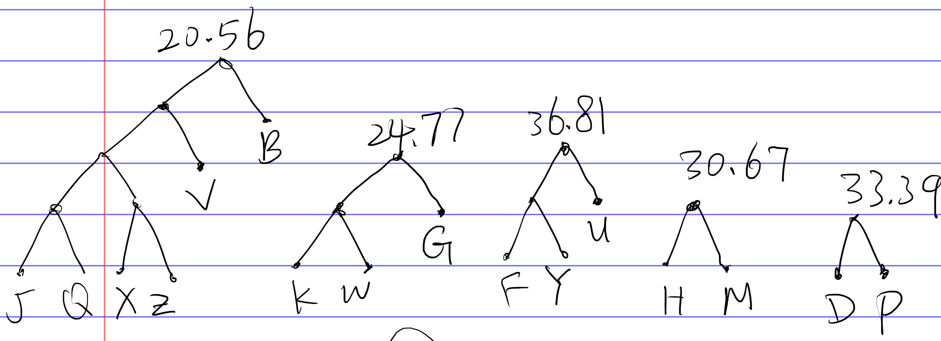
8


20.56
B
V
J Q X Z
24.77
G
K W
18.3
F Y
30.67
H M

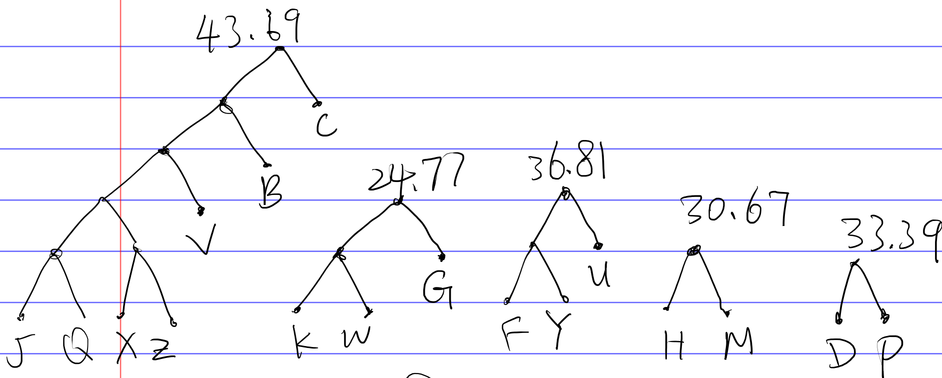
9


20.56
B
V
J Q X Z
24.77
G
K W
18.3
F Y
30.67
H M
33.39
D P

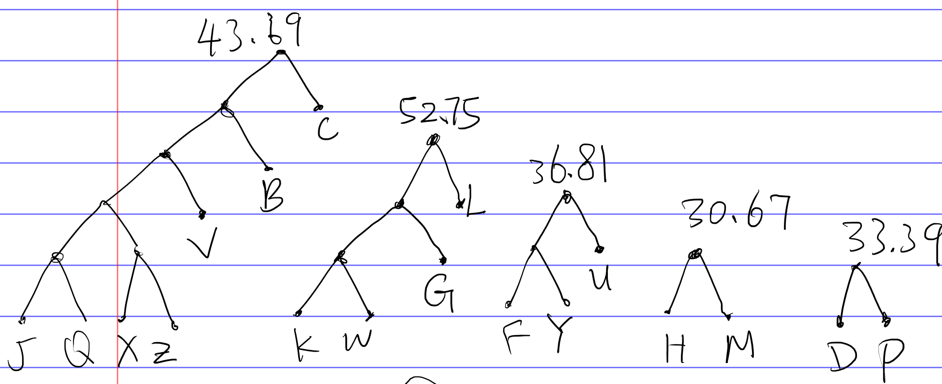
10


20.56
24.77
36.81
30.67
33.39
B
V
G
U
J Q X Z
K W
F Y
H M
D P

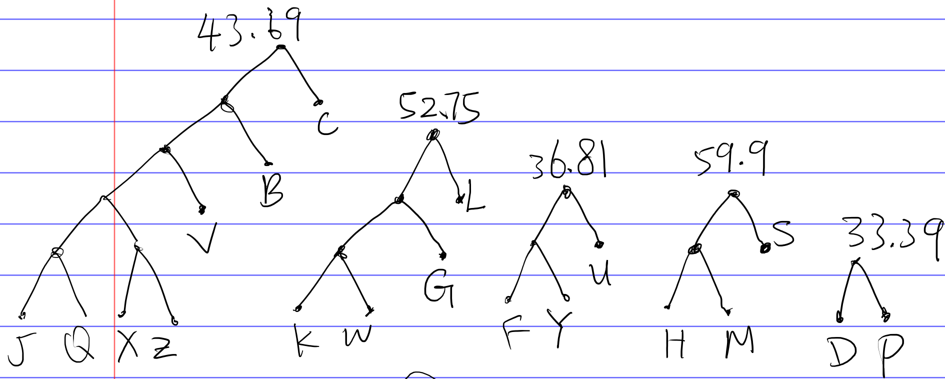
11


43.69
C
B
V
24.77
36.81
30.67
33.39
G
U
J Q X Z
K W
F Y
H M
D P

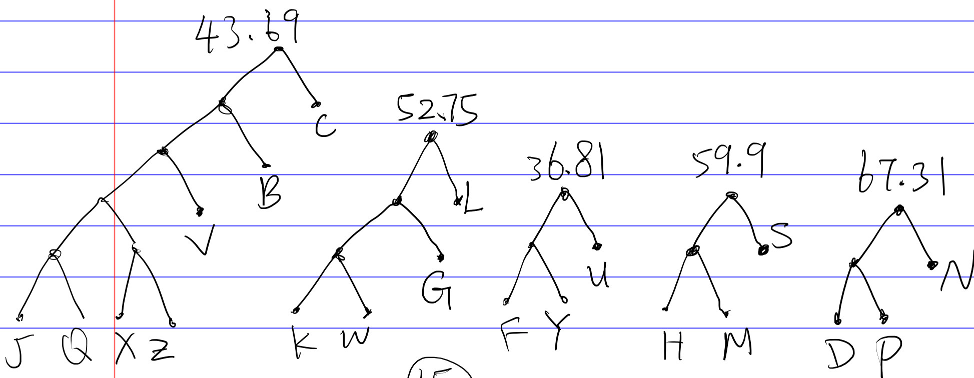
12


43.69
C
52.75
B
L
V
36.81
30.67
33.39
G
U
J Q X Z
K W
F Y
H M
D P

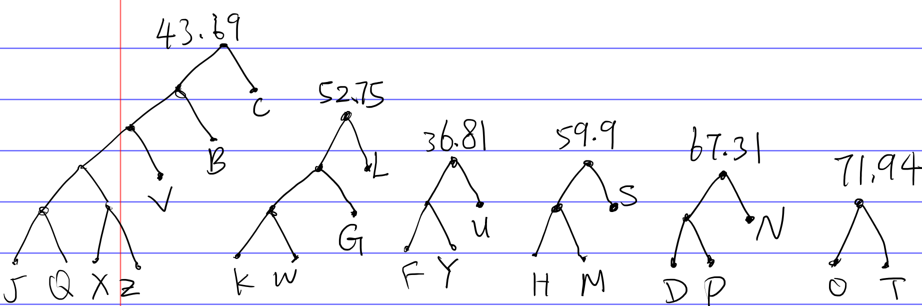
13


43.69
52.75
36.81
59.9
33.39
C B V L G U S
J Q X Z K W F Y H M D P

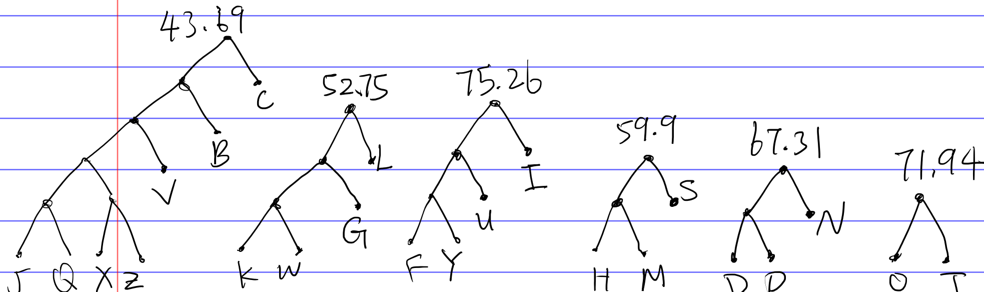
14


43.69
52.75
36.81
59.9
67.31
C B V L G U S N
J Q X Z K W F Y H M D P

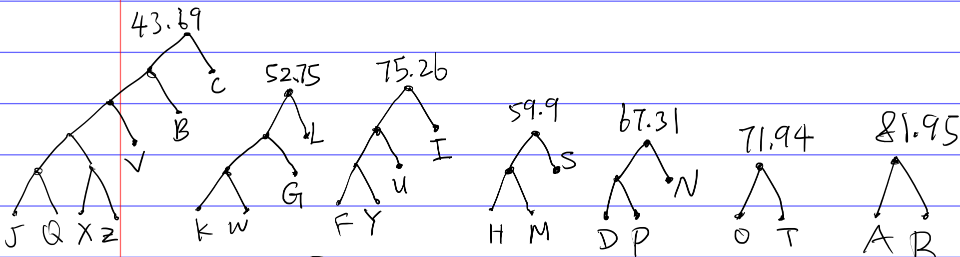
15


43.69
52.75
36.81
59.9
67.31
71.94
C B V L G U S N
J Q X Z K W F Y H M D P O T

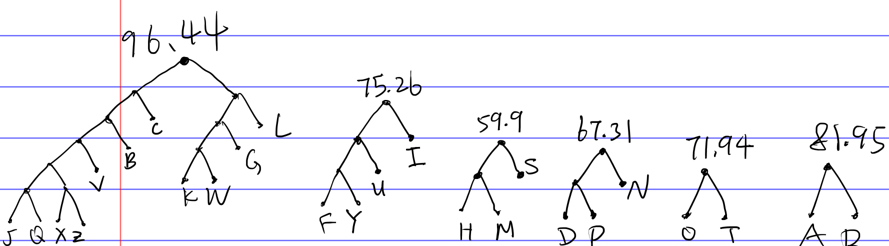
16


43.69
C
B
V
J Q X Z
52.75
L
G
K W
75.26
I
U
F Y
59.9
S
H M
67.31
N
D P
71.94
O T

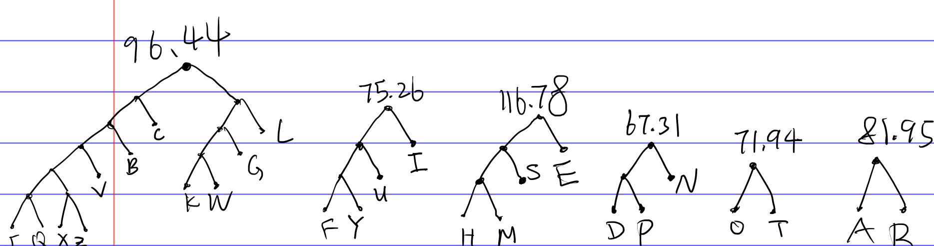
17


43.69
C
B
V
J Q X Z
52.75
L
G
K W
75.26
I
U
F Y
59.9
S
H M
67.31
N
D P
71.94
O T
81.95
A R

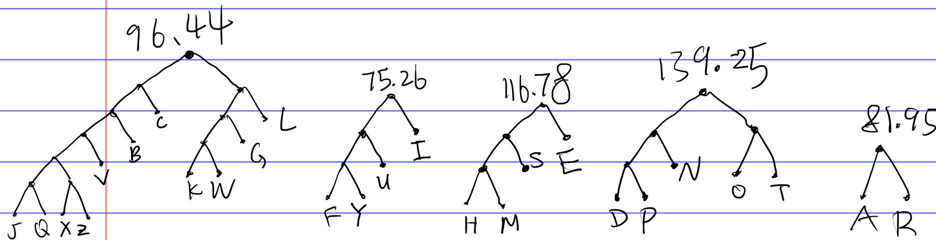
18


96.44
C
B
V
J Q X Z
L
G
K W
75.26
I
U
F Y
59.9
S
H M
67.31
N
D P
71.94
O T
81.95
A R

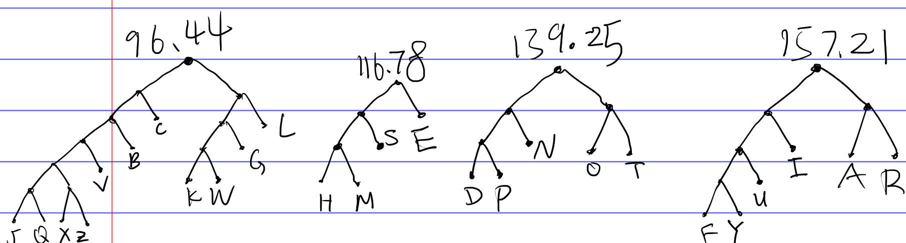
19


96.44
C
B
V
J Q X Z
L
G
K W
75.26
I
U
F Y
116.78
S E
H M
67.31
N
D P
71.94
O T
81.95
A R

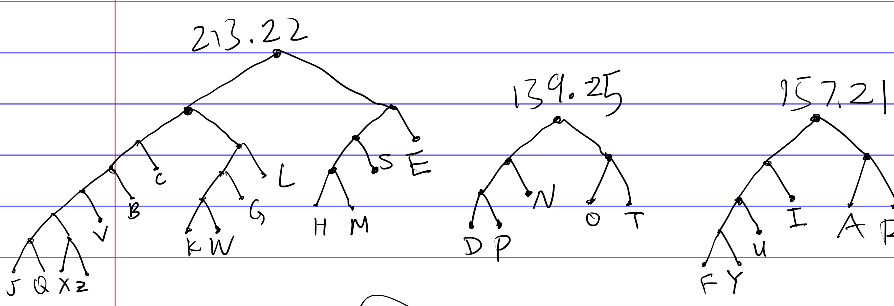
20


96.44
J Q X Z V B C K W G L
75.26
F Y U I
116.78
H M S E
139.25
D P N O T
81.95
A R

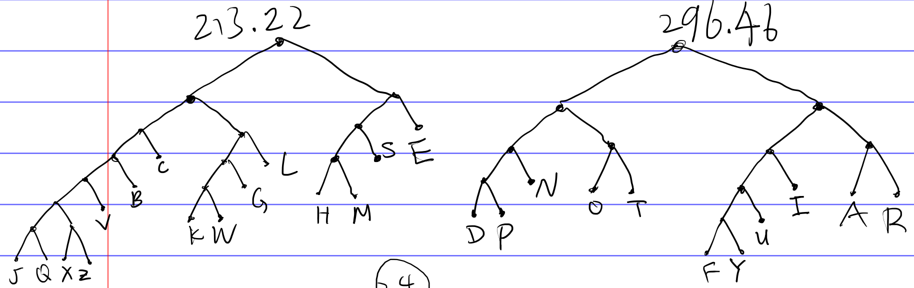
21


96.44
J Q X Z V B C K W G L
116.78
H M S E
139.25
D P N O T
157.21
F Y U I A R

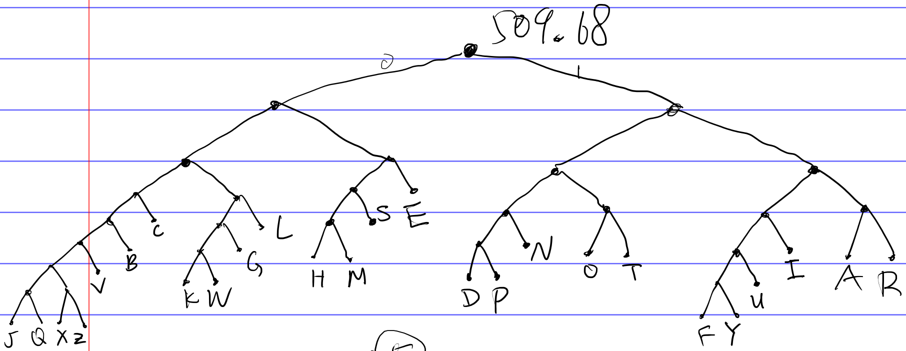
22


213.22
J Q X Z V B C K W G L H M S E
139.25
D P N O T
157.21
F Y U I A R

23



24



25

Huffman coding: [编码树不是唯一的]

		length	%			length	%
E	011	3	0.112	M	01001	5	0.030
A	1110	4	0.085	H	01000	5	0.030
R	1111	4	0.076	G	00101	5	0.025
I	1101	4	0.075	B	00001	5	0.021
O	1010	4	0.072	F	110000	6	0.018
T	1011	4	0.070	Y	110001	6	0.018
N	1001	4	0.067	W	001001	6	0.013
S	0101	4	0.057	K	001000	6	0.011
L	0011	4	0.055	V	000001	6	0.010
C	0001	4	0.045	X	00000010	8	0.003
U	11001	5	0.036	Z	00000011	8	0.003
D	10000	5	0.034	J	00000000	8	0.002
P	10001	5	0.032	Q	00000001	8	0.002

平均码长 $= 3 \times 0.111607 + 4 \times 0.084966 + \cdots$
$+ 8 \times 0.001965 + 8 \times 0.001962$
$= 4.27$

(2) 香农熵：

$$= 0.111607 \times \log_2 \frac{1}{0.111607} + 0.084966 \times \log_2 \frac{1}{0.084966}$$

$$+ \cdots$$

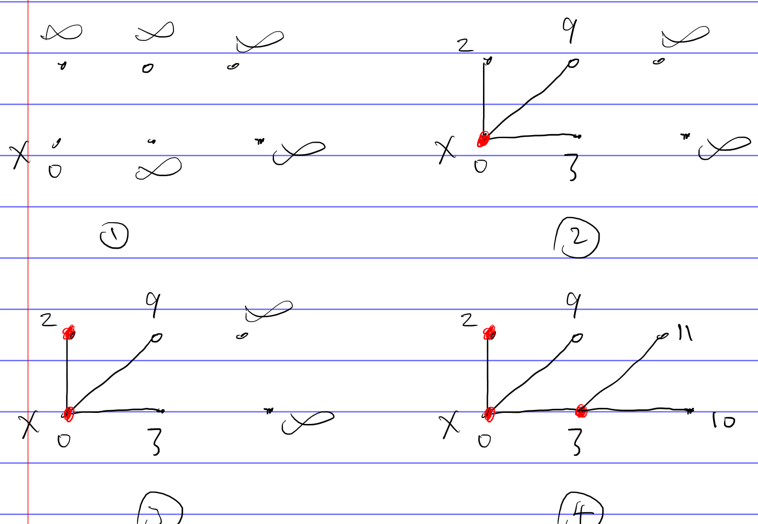
$$+ 0.001965 \times \log_2 \frac{1}{0.001965} + 0.001962 \times \log_2 \frac{1}{0.001962}$$

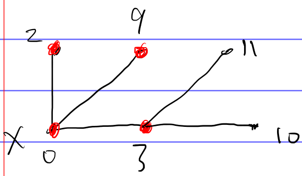
$$\approx 4.24$$

香农熵与Huffman编码的平均码长之差

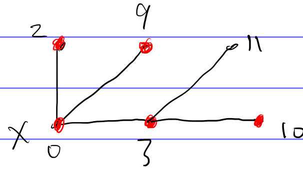
$$= 4.27 - 4.24 = 0.03$$

3.

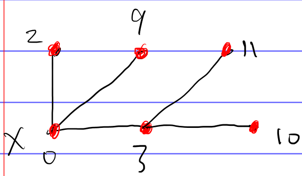




⑤

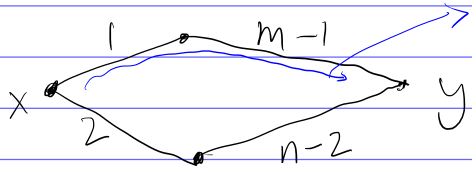


⑥



⑦

4.



负以箭头给出的从x到y的路.

注：图的~~的~~构造不是唯一的

5. 注：此题有点难，不感兴趣的可以忽略.

证明：对图中任意两点 a, b. 用 $d(a,b)$ 表示 a 与 b 之间的最短距离.
题目要求证明，当 $v = curvtx$ 时，

$$d(v) = d(x, v).$$

设v是Dijkstra算法执行过程中的第k个curvtx。
用记号
F_k 表示在第k个curvtx时已经Finished的所有顶点集合
U_k 表示在第k个curvtx时还没有Finished的所有顶点集合
我们证明如下结论：
① 对每一个顶点 $a \in F_k$，
有 $d(a) = d(x, a)$
② 若v是第k个curvtx，则
$d(v) = d(x, v)$。

对k用归纳法。
基础情形：k=1时，
此时 curvtx = x，$d(x)=0$
$F_1 = \varnothing$
$U_1 = V$
因为 $F_1=\varnothing$，所以①自动成立。
第1个curvtx是x，
$d(x)=0 = d(x, x)$，
所以，②也成立。

归纳假设：设①,②对k=t时成立
即：
- 对每一个顶点 $a \in F_t$，
有 $d(a) = d(x, a)$
- 若u是第t个curvtx，则
$d(u) = d(x, u)$。

欲证明①,②对k=t+1时也成立。
根据Dijkstra算法，有
$F_{t+1} = F_t \cup \{u\}$。
因此，根据归纳假设，①对于 F_{t+1} 也成立

设 v 是第 t+1 个 vertex，我们需要证明②，即要证明

$$d(v)=d(x,v)$$

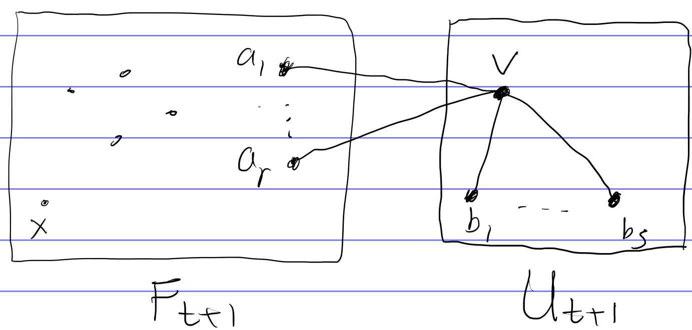
首先，根据 Dijkstra 算法，有：对于任意的 $y\in U_{t+1}$，有

$$d(v)\le d(y) \qquad (\Delta)$$

假设 v 的邻居顶点为

$$\begin{aligned} a_1,\ \dots,\ a_r &\in F_{t+1} \\ b_1,\ \dots,\ b_s &\in U_{t+1} \end{aligned} \qquad (*)$$

如下图所示



根据 Dijkstra 算法的定义，有

$$d(v)=\min_{a_i}\{d(a_i)+d(a_i,v)\}$$

因为 $a_i\in F_{t+1}$，所以，

$$d(a_i)=d(x,a_i)$$

所以：

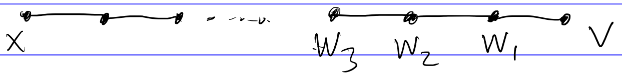
$$d(v)=\min_{a_i}\{d(x,a_i)+d(a_i,v)\}$$

现证明 $d(v) = d(x,v)$

用反证法 假设 $d(v) \neq d(x,v)$. (※)

不妨假设 $d(v) > d(x,v)$

设从 x 到 v 的某一条最短路如下



(□)

用 w_1 表示在这条路上 v 的邻居，w_2 表示在这条路上 w_1 的邻居，等等.

则：

$$d(x,v) = d(x,w_1) + d(w_1,v)$$

$$d(x,w_1) = d(x,w_2) + d(w_2,w_1)$$

等等

根据(※)，w_1 要么是某个顶点 a_i，要么是某个顶点 b_i

下面证明 w_1 不可能是某个 a_i，

这是因为

$$d(x,w_1) + d(w_1,v)$$
$$= d(x,v)$$
$$< d(v) = \min_{a_i}\{d(x,a_i) + d(a_i,v)\}.$$

因此，w_1 只能是某个 b_i，即

$$w_1 \in U_{t+1}.$$

因此，根据(△)，有

$$\begin{aligned} d(w_1) &\geq d(v) \\ &> d(x,v) \\ &= d(x,w_1) + d(w_1,v) \\ &> d(x,w_1) \end{aligned} \qquad (☆)$$

即：

$$\begin{aligned} d(w_1) &> d(x,w_1) \\ &= d(x,w_2) + d(w_2,w_1) \end{aligned}$$

下面证明 $w_2 \in U_{t+1}$. 用反证法.

假设 $w_2 \notin U_{t+1}$，则 $w_2 \in F_{t+1}$

那么由 $w_2 \in F_{t+1}$ 有，

$$d(w_2) = d(x,w_2)$$

因此，根据 Dijkstra 算法 有

$$d(w_1) \le d(w_2) + d(w_2, w_1) = d(x, w_2) + d(w_2, w_1) < d(w_1)$$

矛盾. 故假设不成立. 即必有 $w_2 \in U_{t+1}$

从而由 (△) 有:

$$d(w_2) \geq d(v) > d(x, w_1) = d(x, w_2) + d(w_2, w_1) > d(x, w_2). \quad (☆')$$

注意 (☆') 与 (☆) 类似; 因此 类似地 又可以证明 $w_3 \in U_{t+1}$, 等等. 由此可知 那么 逐个地 证明 在 最短路 (□) 中, 所有 顶点 都属于 U_{t+1}. 特别地,

$$x \in U_{t+1}$$

然而 我们知道 $x \in F_{t+1}$.

即 $x \notin U_{t+1}$.

这是矛盾.

因此, 假设 (☆) 不成立, 即 必有

$$d(v) = d(x, v).$$

□

6.

(1) 设边的排序中每次比较算一步

用冒泡法要最多 $O(|E|^2)$ 步.

设 while 循环中的每一次执行也算一步,

则 while 循环要 $O(|V|)$ 步.

因此, Kruskal 算法 最多用

$O(|E|^2) + O(|V|) = O(|E|^2 + |V|)$ 步.

(2) 设第一个 for 循环中每一次执行算一步,

则第一个 for 循环用 $O(|V|)$ 步.

设 while 里的 for 循环中的每一次执行算一步, 则 while 循环用

$$O(|V| \cdot |V|) = O(|V|^2)$$

因此, Dijkstra 算法 最多用

$$O(|V|) + O(|V|^2) = O(|V|^2) \text{ 步}$$

□

7. 略.