

1. Data Structure and Algorithm

(a.) Undirected graph:

題目要求要 no cycle，因此對於 undirected graph 來說就要要建構 tree，最多只能有 $V-1$ 個 edges，同時又要 connected，只能是 $V-1$ 個 edge，找拔掉最少 cost，也就等於找最大使之連起來的 $V-1$ 個 nodes，因此使用 Kruskal's Algorithm 尋找 Minimum Spanning Tree (MST)，但是是從最大開始找。

使用 disjoint-set data structure 來確認已經連在一起的成員，如果已經連在一起那將會找到共同的祖先，就不能再連 edge(否則形成 cycle)，反之則可以，並利用 Union by rank(結合時，淺的 set 連到深的 set，rank 就不會增長太快)和 Path compression(結合後，之後搜尋到的 node 的祖先都直接連到 root 的祖先，這樣之後就不用重複多層找這個 node 的祖先)的技巧讓 Time complexity 幾乎可以壓在 $O(E \lg V + V)$ 上。

(b.) Directed graph:

1. 先把負 weight 的 edge 取出。 $O(E)$

2. 利用正 weight 的 edge 去做各種排列，並將 nodes 編號。 $O(V^2 + E)$

3. 將正 weight 的 edge 中，node 由小到大的組，大到小的一組，選 weight 小的那組拔除。 $O(E)$

4. 將在 graph 中的 nodes 用剩下來的 edges，建立 disjoint-set data structure(同上)。 $O(E + V)$

5. 將沒有放入的 edges 由 weight 由大到小排列，執行 Kruskal's Algorithm，直到所有 nodes 屬於同一個 set。 $O(E \lg V)$

Total Time complexity: $O(V^2 + E \lg V)$

如果直接執行(省略步驟 2.): Time complexity: $O(E \lg V + V)$

步驟 3. 和 5. 可以確保留在 graph 中的 edge 屬於 DAG(沒有 cycle)，因為 3.

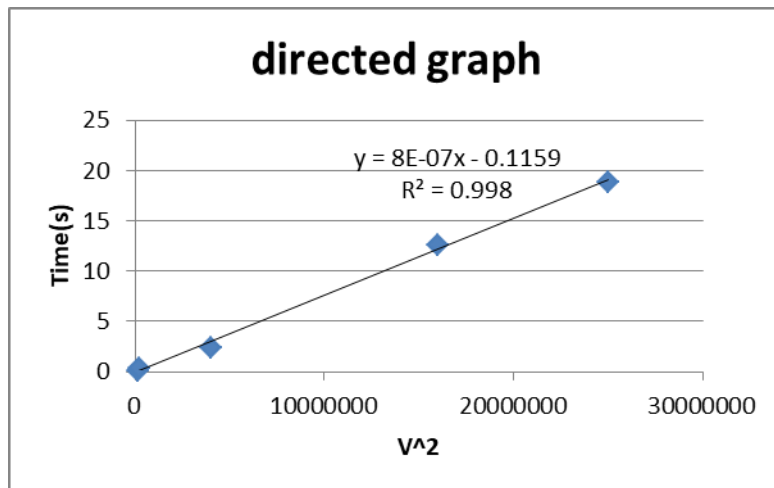
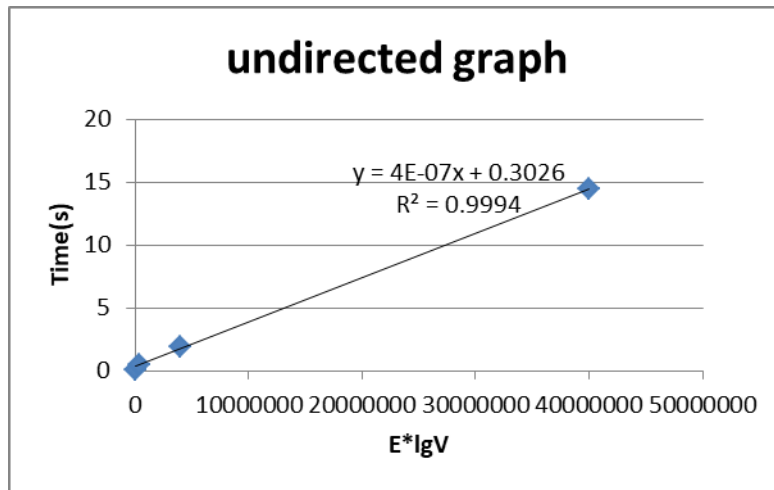
只能有單一方向，5.的 Kruskal's Algorithm 確保有連過的在同一個 set，也不會有 cycle。而步驟 5.利用 set 確認所有 nodes 連接起來，為 connected。

因為 minimum feedback arc set 目前是 NP-Hard problem，所以沒辦法用 polynomial time 解完最佳解，但可以確保其 cost 的 worse bound 在最佳 cost (C) 的 $C + 100 \times \frac{E}{2} - (-100 \times (V - 1))$ 以下，因為步驟 3.可以確保超過 1/2 的總正 weight 不被選到，步驟 5.最壞只要放 V-1 個-100，即可讓 graph connect，而 average case 的狀況假設正 weight 的 edge 平均是 50，負 weight 的 edge 平均是-50，且數量一樣，worse bound 在 $C + 50 \times \frac{E}{2} - (-17 \times (V - 1))$ 之下。另外在步驟 2.，利用每個 nodes 的輸出 edge weight 和輸入 edge weight 乘上不同的權重(-1 到 1)，來做選擇，從中挑選最佳解輸出。從步驟 2.的不同選擇中，平均會有 10%-30%的變化量，可以藉此選到比較好的 cost。

2. Complexity Analysis

undirected graph				
# of node	1000	10000	10000	10000
# of edge	10000	100000	1000000	10000000
$E \cdot \lg V$	30000	400000	4000000	40000000
time(s)	<0.1	0.5	1.9	14.5

directed graph					
# of node	400	500	2000	4000	5000
# of edge	5000	10000	10000	10000	10000
V^2	160000	250000	4000000	16000000	25000000
time(s)	<0.1	0.3	2.4	12.6	18.9



Undirected graph 的花費時間與 $E \cdot \lg V$ 成正比關係，符合 Kruskal's Algorithm 的 Time complexity，而 Directed graph 的花費時間跟 V^2 成正比，代表在挑選不同 model 的權重上花費較多時間，但在控制好參數的數量下，依然可以在時間內完成 5000/10000 的 case。