Algorithm PA3 report

李采蓉
心理五 b08207042

December 29, 2023

Algorithms

本次作業可以從兩個階段進行發想，針對問題「取出的邊集有最小權重和」和「剩下的圖為 connected 且 acyclic」，對於 undirected graph 來說，這部分的解法就在「邊集的最小權重和」的反面–「找出 Maximum Spanning Tree」，於是乎採用 Kruskal's algorithm 將 sorting 的部分改為 nondecresing，便可達成任務。因為知道 weight 的上下界，於是可以使用 sorting in linear time 的方法進一步壓縮 sorting 的時間，在此我使用的是 counting sort。Disjoint set 考量到時間成本，用的是 Union by rank 避免花過多時間在 merge 上。

第二階段是對 directed graph 的延伸處理，directed graph 我們一樣先使用 Kruskal's algorithm 找出最小權重和邊集，但因為邊有方向性的關係使的有些被挑出來的邊就算放回去也不會形成 cycle，再者我們的目標是找出「最小權重和」，講求的是負邊越多越好，正邊越少越好，所以在邊集中挑出正邊放回圖中，以 DFS 檢查如此一來是否會形成環，假若有環產生便不可被從邊集中剃除，反之則可以從邊集中拿到，權重和扣掉此邊權重。

Data Structures

接下來介紹此程式中的資料結構，首先最基礎的便是「邊」與「點」表達，所建構的 struct 很簡單，就是把測資中所給的資料整理成如下：

```
struct edge {
    int u; // begin
    int v; // end
    int w; // weight
};

struct vertex {
    int d; // point to
    int w; // weight
};
```

輸入資料時使用長度為邊數的 vector<edge> edge_set 儲存每條邊的資訊，此外也有另外兩個 vector<edge> 儲存需要被保留的邊集和需要被剃除的邊集，分別為 remain 和 remove。至於 vertex 的使用時機則被用在建構 adjacency list，將 remain 所存的邊轉換為方便操作 DFS 的形式。最後還有兩個儲存 int 的 vector，長度皆為點數，用於記錄每個點所屬的 disjoint set 和 rank。

Analysis

根據上述的演算法，因為 sorting 採用的是 counting sort，所花時間為 $O(E+k)$，至於 Union 的部分整體所花時間為 $O(E)$，因此 Kruskal's algorithm 的時間複雜度為 $O(E+k)$。找 cycle 的部分為修改版的 DFS，因此可能不會歷遍整個 graph，同時間 remain 和 remove 所存的點數加總起來必為圖上所有的邊數，設 remain 存了 e_1 個邊，remove 存了 e_2 個邊，$e_1+e_2=E$。我們可以根據這項關係約略得出在將邊放回圖中以 DFS 檢查是否產生 cycle 的過程所花時間為 $O(e_2*(e_1+V))$。基於上述的推論，undirected graph 的時間複雜度為 $O(E)$，directed graph 的時間複雜度會界在 $E+V$ 到 $E*V$ 之間，視為 $O(EV)$。

接下來對資料結構進行討論。在上面有提及了三個儲存邊的 vector，主要的 edge_set 和其餘兩個邊集 remain、remove，edge_set 的長度必為邊長，而 remain、remove 的邊長在前一段也有提過，加總起來必為邊長，因此這部分的空間複雜度為 $O(E)$。Adjacency list 僅以 remain 內有的邊為基礎，故剛建好時空間複雜度為 $O(e_1+V)$，然而這是會隨著插回檢測的進行而有所改變的，最

壞的情況是幾乎所有的邊都被插回去了，這時 adjacency list 的空間複雜度為 $O(E+V)$。最後兩個 vector，parent 跟 rank，長度皆為邊數，故空間複雜度一樣為 $O(E)$。總的來說，空間複雜度為 $O(E+V)$

下表為 public cases 的執行時間和所花記憶體空間

Case	Type	Time (ms)	Memory (KB)
0	weighted directed	0.199	6044
1	unweighted undirected	0.192	6044
2	weighted undirected	0.316	6044
3	weighted undirected	3.366	6044
4	weighted directed	0.169	6044
7	weighted directed	1.182	6044
8	weighted directed	27.238	6176