University of Copenhagen

Union-Find

Author: William Henrich Due

1 Theory

Definition 1.1 (Reachability)**.** A node v is *reachable* from a node u in a directed graph $G = (V, E)$ if there exists a sequence of directed edges $e_1, e_2, \ldots, e_m \in E$ where $m \geq 1$ and $e_i = (v_{i-1}, v_i)$ for $1 \leq i \leq m$, such that $v_0 = u$ and $v_m = v$. We denote this by $u \rightsquigarrow v$.

Definition 1.2 (Cycle)**.** A cycle in a directed graph $G = (V, E)$ has a cycle if there exists $v \in V$ such that $v \rightsquigarrow v$.

Definition 1.3 (Forest)**.** A forest is a directed graph $F = (V, E)$ where V is a set of vertices and $E \subseteq V \times V$ is a set of directed edges such that:

1. There are no cycles $v \not\rightsquigarrow v$ for all $v \in V$, and
2. each node has at most one parent i.e. for all $(u, v_1), (u, v_2) \in E$ it holds that $v_1 = v_2$.

Definition 1.4 (Root)**.** A node $v \in V$ in a forest $F = (V, E)$ is a root if it has no parent. This is defined as the predicate:

$$\mathcal{R}_F(v) : v \not\rightsquigarrow u \text{ for all } u \in V$$

Definition 1.5 (Tree)**.** A tree is a forest $T = (V, E)$ where there exists a unique root $r \in V$ such that $v \rightsquigarrow r$ for all $v \in V \backslash \{r\}$.

Proposition 1.1 (Forest Root Count)**.** A forest $F = (V, E)$ where $|V| = n$ and $|E| = n - k$ has k roots.

Proof. Let $F = (V, E)$ be a forest where $|V| = n$ and $|E| = n - k$. By the second property of a forest then $n - k$ vertices must have a parent. Since there are n vertices in total it follows that there are exactly k vertices $r_1, r_2, \ldots, r_k \in V$ that has no parent. Hence there are exactly k roots in F. □

Proposition 1.2 (Roots Path Exist)**.** In a forest $F = (V, E)$ for each element $v \in V$ there exists at least one root $r \in V$ such that $\mathcal{R}_F(r)$ and either $v \rightsquigarrow r$ or $v = r$.

Proof. Let $F = (V, E)$ be a forest and let $v \in V$ be an arbitrary element in V. By proposition 1.1 there exists at least one root $r \in V$ such that $\mathcal{R}_F(r)$. This can be shown by structural induction on a vertex $v \in V$ that any $(v, p) \in E$ then either $p = r$ or p has a path to some root $r \in V$.

- If $p = r$ then p is a root and v has a path to a root r by $(v, r) \in E$.
- By induction hypothesis p has a path to a root $r \in V$ such that $\mathcal{R}_F(r)$. Since $(v, p) \in E$ it follows that $v \rightsquigarrow p \rightsquigarrow r$ so .

□

Proposition 1.3 (Forest Edge Limit)**.** A forest $F = (V, E)$ where $|V| = n$ and $|E| > n - 1$ is not a forest.

Proof. Let $F = (V, E)$ be a forest where $|V| = n$ and $|E| > n - 1$. By proposition 1.1 a forest with $n - 1$ has exactly one root, so it is a tree. By adding one more edge to the tree it must make one vertex have two parents. Or since every vertex $v \in V$ has a path to the root $r \in V$ it must create a cycle $v \rightsquigarrow v$ for some $v \in V$. In both cases it contradicts the properties of a forest. □

Definition 1.6 (Representative)**.** The representative of an element $v \in V$ in a forest $F = (V, E)$ is the root $r \in V$ such that there is a path from v to r. This is defined as the function:

$$\rho_F(v) := r \text{ where } r \in V \text{ such that } \mathcal{R}_F(r) \wedge (v \rightsquigarrow r \vee v = r)$$

Proposition 1.4 (Unique Representative)**.** In a forest $F = (V, E)$ each element $v \in V$ has a unique representative $\rho_F(v)$.

Proof. Let $F = (V, E)$ be a union-find structure and let $v \in V$ be an arbitrary element in V. By proposition 1.2 there exists at least one root $r \in V$ such that $\mathcal{R}_F(r)$ and $v \rightsquigarrow r$. Now assume that there exists another root $r' \in V$ such that $\mathcal{R}_F(r')$ and $v \rightsquigarrow r'$. Since F is a forest it follows by the second property of a forest that $r = r'$ hence the representative is unique. □

Definition 1.7 (Tree Set)**.** The set of vertices of the same tree $\mathcal{E}_F(v)$ in a forest $F = (V, E)$ is defined as:

$$\mathcal{E}_F(v) := \{u : u \in V \text{ where } \rho_F(u) = \rho_F(v))\}$$

Definition 1.8 (Partition)**.** The set $P \subseteq \mathbb{P}(S)$ is a partition of a set S if:

1. $a \neq \emptyset$ for all $a \in P$
2. $a \cap b = \emptyset$ for all $a, b \in P$ where $a \neq b$
3. $\bigcup_{a \in P} a = S$

Proposition 1.5 (Forest Partition)**.** A forest $F = (V, E)$ is a partition of V for the following set:

$$\{\mathcal{E}_F(v) : v \in V\}$$

Proof. Let $F = (V, E)$ be a forest. We will show that the set in the proposition is a partition of V by showing that it satisfies the three properties in definition 1.8.

1. By definition of $\mathcal{E}_F(v)$ it can not be empty since for $\mathcal{E}_F(v)$ then $\rho_F(v) = \rho_F(v)$. Hence $\mathcal{E}_F(v) \neq \emptyset$ for all $v \in V$.

2. Let a and b be two arbitrary elements in the set such that $a \neq b$. By definition of a and b there exists $v_1, v_2 \in V$ such that $a = \{u : u \in V \wedge \rho_F(u) = \rho_F(v_1)\}$ and $b = \{u : u \in V \wedge \rho_F(u) = \rho_F(v_2)\}$. Since $a \neq b$ it follows that $\rho_F(v_1) \neq \rho_F(v_2)$ since otherwise $a = b$, hence $a \cap b = \emptyset$.

3. Let v be an arbitrary element in V. By proposition 1.2 there exists a root $r \in V$ such that $\mathcal{R}_F(r)$ and $v \rightsquigarrow r$ or $v = r$. By definition of the representative it follows that $\rho_F(v) = r$. Now let $a = \{u : u \in V \wedge \rho_F(u) = \rho_F(v)\}$. By definition of a it follows that $v \in a$. Since v was arbitrary it follows that $\bigcup_{a \in P} a = V$.

□

Definition 1.9 (Same Tree Relation)**.** The relation $\sim_F$ on a forest F is defined as:

$$u \sim_F v :\iff u \in \mathcal{E}_F(v)$$

Corollary 1.1 (Same Tree Relation is an Equivalence Relation)**.** The relation $\sim_F$ on a forest F is an equivalence relation due to $\{\mathcal{E}_F(v) : v \in V\}$ being a partition of V. by proposition 1.5.

Definition 1.10 (Forests with Equivalent Tree Sets)**.** Two forests $F = (V, E)$ and $F' = (V', E')$ have equivalent tree sets $F \cong F'$ if:

- Vertices are the same $V = V'$.
- The tree sets are equivalent $\mathcal{E}_F(v) = \mathcal{E}'_F(v)$ for all $v \in V$.

Definition 1.11 (Tree Union)**.** The tree union of two elements v and u for a forest $F = (V, E)$ is such that $v \sim_{F'} u$ in a new forest $F' = (V', E')$ and F' satify the following properties:

1. $\mathcal{E}_{F'}(v) = \mathcal{E}_{F'}(u) = \mathcal{E}_F(v) \cup \mathcal{E}_F(u)$ and
2. $\mathcal{E}_{F'}(w) = \mathcal{E}_F(w)$ for all $w \in V \backslash (\mathcal{E}_F(v) \cup \mathcal{E}_F(u))$.

Proposition 1.6 (Tree Union)**.** Let forest $F = (V, E)$, $p = \rho_F(u)$ be the representative of u and let $q = \rho_F(v)$ be the representative of v where $q \neq p$. Then defined F' as:

$$F' := (V, E \cup \{(q, p)\})$$

Then $u \sim_{F'} v$ in F' and F' will satisfy the properties of a tree union.

Proof. Let $F = (V, E)$, $p = \rho_F(u)$ be the representative of u and let $q = \rho_F(v)$ be the representative of v. By definition q will have parent p in F' and since q is a root it has no parent then F' is a forest. Now for all $w \in \mathcal{E}_F(q)$ it holds that $w \rightsquigarrow q$ or $q = w$ and since $q \rightsquigarrow p$ it follows that $w \rightsquigarrow p$. Hence $w \in \mathcal{E}_{F'}(p)$ for all $w \in \mathcal{E}_F(q)$ and trivially $w \in \mathcal{E}_{F'}(p)$ for all $w \in \mathcal{E}_F(p)$ so it follows that $\mathcal{E}_{F'}(v) = \mathcal{E}_{F'}(u) = \mathcal{E}_F(v) \cup \mathcal{E}_F(u)$. Now let $w \in V \backslash (\mathcal{E}_F(v) \cup \mathcal{E}_F(u))$ be an arbitrary element. Since $w \not\rightsquigarrow p$, $w \neq p$, $w \not\rightsquigarrow q$, and $w \neq q$ it follows that w has the same representative in F' as in F hence $\mathcal{E}_{F'}(w) = \mathcal{E}_F(w)$. $\square$

Definition 1.12 (Conflict-free Set)**.** Let F be a forest, $X \subseteq \{(\rho_F(v), \rho_F(u)) : (v, u) \in V \times V\}$ be a set of root pairs. Then X is a conflict-free set in F if (V, Y) is a forest.

Proposition 1.7. Let forest $F = (V, E)$ be a forest and let $X \subseteq V \times V$ be a conflict-free set in F where $|X| = n$. Then defining the following forests:

$$F_0 := F$$
$$F_i := (V, E_{i-1} \cup \{(v_i, u_i)\}) \text{ for } (v_i, u_i) \in X \text{ and } 1 \leq i \leq n$$

Then F_n is a forest.

Proof. Let forest $F = (V, E)$ be a forest, $X \subseteq V \times V$ be a conflict-free set in F. We will show that F_n is a forest by induction on i.

- Base case: If $i = 0$ then $F_i = F_0 = F$ which is a forest.
- Induction hypothesis: Assume that F_{i-1} is a forest for all $1 \leq i < n$. Let $(v_i, u_i) \in X$, we know that $v_i \neq v_j$ for all $(v_j, u_j) \in Y \backslash \{(v_i, u_i)\}$ since otherwise (V, X) would not be a forest and X would not be a conflict-free set in F. So v_i will only have one parent in F_i since it only appears once as a child in (V, X). By definition all of the edges in X consists of roots in F, and since (V, X) is a forest there are no cycles $y \not\rightsquigarrow y$ for all $y \in V$ in F_i. Hence F_i is a forest.

Thus by induction F_n is a forest. $\square$

Proposition 1.8. Let forest F be a forest and let $X \subseteq V \times V$ be a conflict-free set in F where $|X| = n$. Then defining the following forests:

$$\begin{aligned}
F_0 &:= F \\
F_i &:= (V, E_{i-1} \cup \{(v_i, u_i)\}) \text{ for } (v_i, u_i) \in X \text{ and } 1 \leq i \leq n \\
G_0 &:= F \\
G_j &:= (V, E_{i-1} \cup \{(\rho_{G_{j-1}}(v_j), \rho_{G_{j-1}}(u_j))\}) \text{ for } (v_j, u_j) \in X \text{ and } 1 \leq j \leq n
\end{aligned}$$

Then $F_n \cong G_n$.

Proof. Let forest F be a forest, $X \subseteq V \times V$ be a conflict-free set in F. We will show that $F_n \cong G_n$. We know that for some $(v_i, u_i) \in X$ then $v_i \neq y$ for all $(y, w) \in X \backslash \{(v_i, u_i)\}$ since otherwise (V, X) would not be a forest and X would not be a conflict-free set in F. So all edge set unions will only give a root v_i a new parent u_i once. So $\rho_{F_n}(v_i) = \rho_{F_n}(u_i)$ and $\rho_{G_n}(u_i) = \rho_{G_n}(v_i)$ hence v_i remains in the same tree in both F_n and G_n. Since this holds for all $(v_i, u_i) \in X$ it follows that all elements in V remains in the same tree in both F_n and G_n. Hence $F_n \cong G_n$. $\square$

Definition 1.13 (Split)**.** A split of a set $Z \subseteq V \times V$ into two sets $X, Z' \subseteq Y$ such that X is a conflict-free set in forest F and $X = Y \backslash Z'$ where:

$$Y = \{(a, b) : (v, u) \in X \text{ where } \{a, b\} = \{\rho_F(v), \rho_F(u)\} \wedge a \neq b\}$$

Algorithm 1.1 (Conflict-free Tree Union)**.** Let forest F be a tree and let $Z \subseteq V \times V$ be a set of pairs of elements in V that will be unioned in parallel. The conflict-free tree union algorithm is defined as:

```
while Z ≠ ∅ do
    E ← π₂(F)
    split Z into X, Z'
    parfor (u, v) ∈ X do
        E ← E ∪ {(u, v)}
    F ← (V, E)
    Z ← Z'
```

Definition 1.14 (Strongly Connected Graph)**.** A strongly connected graph $G = (V, E)$ is a directed graph where for if $G' = (V, E \cup \{(u, v) : (v, u) \in E\})$. Then for all $u, v \in V$ it holds that $u \rightsquigarrow v$.

Proposition 1.9 (Maximum Edges in Conflict-free Set)**.** Let V be a set and $Z \subseteq V \times V$ be a set that will be split into a conflict-free set X and a remainder set Z'. Then it is possible to find the maximum conflict-free set X such that $|Z| = 0$ and $|X| = |V| - k$ where k is the number of strongly connected components in the graph $G = (V, Z)$.

Proof. By definition of a split then we can swap the components of any pair in Z. So if we consider the undirected graph $G' = (V, A)$ where $A = \{\{u, v\} : (u, v) \in Z\}$ then it has k strongly connected components. Now by choosing some arbitrary tree which connects all vertices in each strongly connected component and directing the edges such that the tree has some arbitrary root in each component it will create a conflict-free set X where $|X| = |V| - k$ since each tree will have exactly one root that will not have a parent. And since all other edges in Z will already be in some tree in (V, X) it follows that $Z' = \emptyset$. $\square$

Definition 1.15 (Isolated Vertex)**.** An *isolated vertex* in a directed graph $G = (V, E)$ is a $v \in V$ such that $(u, v) \notin E$ and $(u, v) \notin E$ for some $u \in V$.

Proposition 1.10 (Edge Swap Bound)**.** Let $G = (V, E)$ be a directed graph without isolated vertices then:

$$|\{v : (v, u) \in E\}| < \frac{|V|}{2} \implies |\{u : (v, u) \in E\}| > \frac{|V|}{2}$$

Proof. Let $G = (V, E)$ without any isolated vertices and $|\{v : (v, u) \in E\}| < \frac{|V|}{2}$. Let $A = \{v : (v, u) \in E\}$, $B = \{u : (v, u) \in E\}$ and $C = B \backslash A$. By definition of C we have $A \cap C = \emptyset$ so $|A| + |C| = |V|$ and since $|B| \geq |C|$ we can conclude that:

$$|B| \geq |C| = |V| - |A| > |V| - \frac{|V|}{2} = \frac{|V|}{2}$$

Hence $|\{u : (v, u) \in E\}| > \frac{|V|}{2}$. $\square$

Definition 1.16 (Union-Find Structure)**.** The union-find structure U is a forest $U = (V, E)$ where the vertices $V = S$ for some set of elements S. The edges $E \subseteq V \times V$ represent parent relations between elements in S such that $(u, v) \in E$ means that u has parent v.