# 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 $v \in V \setminus \{r\}$.

**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 $F$. $\square$

**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 $p \in V$ such that $v \rightsquigarrow p$ has a path $p \rightsquigarrow r$.

- If $v$ is a root then we are done.

- By induction hypothesis $v$ has a path to a root $r \in V$ such that $\mathcal{R}_F(r)$. Since $v$ is not a root it must have a parent $p \in V$ such that $(v, p) \in E$. Since $v \rightsquigarrow r$, $v \rightsquigarrow p$ and $v$ only has one parent it follows that $p \rightsquigarrow r$. Hence $p \rightsquigarrow r$ for some root $r \in V$ such that $\mathcal{R}_F(r)$.

$\square$

**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. $\square$

**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:

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

**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. $\square$

**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:

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

**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 $v \rightsquigarrow r$. 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 $\bigcup_{a \in P} a = V$.

$\square$

**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 $\sim$ 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 $V = V$ and

- 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** (Change Parent). The function for changing a parent of an element $i$ to $p$ in a forest $F = (V, E)$ is defined as:

$$\mathcal{S}_F(i, p) := \begin{cases} (V, (E \setminus \{(i, p')\}) \cup \{(i, p)\}) & \text{if } (i, p') \in E \\ (V, E \cup \{(i, p)\}) & \text{if } (i, p') \notin E \end{cases}$$

**Definition 1.12** (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 $w \in V \setminus (\mathcal{E}_F(v) \cup \mathcal{E}_F(u))$.

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

$$F' := \begin{cases} F & \text{if } p = q \\ \mathcal{S}_F(q, p) & \text{if } p \neq q \end{cases}$$

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 $v$.

- If $p = q$ hence we already have $u \sim_F v$ in $F$ and since the tree sets are equivalent $F' \cong F$ then the properties are satisfied.

- If $p \neq q$ then by definition of changing parent then $q$ will have parent $p$ in $F'$ and since $q$ is a root it has no parent and $F' = (V, E \cup \{(q, p)\})$. 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 $w \in V \setminus (\mathcal{E}_F(v) \cup \mathcal{E}_F(u))$ be an arbitrary element. Since $w \not\rightsquigarrow p$ and $w \not\rightsquigarrow q$ it follows that $w$ has the same representative in $F'$ as in $F$ hence $\mathcal{E}_{F'}(w) = \mathcal{E}_F(w)$.

$\square$

**Definition 1.13** (Multiple Tree Unions). Let forest $F = (V, E)$ and $X \subseteq V \times V$ be a subset of pairs of elements in $V$ that should be unioned. We now define the $\{(\rho_F(u), \rho_F(v)) : (u, v) \in X \wedge \rho_F(u) \neq \rho_F(v)\}$

**Definition 1.14** (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 $v$.