CALCULATING THE EXPECTED AUTOMORPHISM GROUP FOR RRTS

DAVID MATTHEWS

1. Random recursive trees

A *random recursive tree* (RRT) is a labelled, rooted tree obtained by assigning a root vertex and adding $n-1$ vertices one by one such that each new vertex is joined by an edge to a randomly and uniformly chosen existing vertex. It is natural to consider RRTs as nested sequences of rooted, labelled trees

$$T_1 \subset T_2 \subset \cdots \subset T_n$$

Where each T_t has precisely t vertices (and $(t-1)$ edges). At time t vertex v is chosen uniformly at random from $V(T_{t-1})$ and a new vertex v_t is attached to T_{t-1} via the edge $\{(v, v_t)\}$. Furthermore, we use the notation $\{T_i\}_{i=1}^{n}$ to mean a RRT on n vertice and we denote the set of all RRTs on n vertices by $\mathcal{T}$

Let $T = (V(T), E(T))$ be a labelled tree (not necessarily a RRT) and $d(v, w)$ be the length of the (unique) shortest path between any pair of vertices $v, w \in V(T)$. Every vertex $v \neq 1$ has a well defined *father*: the unique vertex v' adjacent to v such that $d(v', 1) < d(v, 1)$. Let $\mathbb{N}_n = \{1, 2, 3, \ldots, n\}$.

Lemma 1.1. *Let $\mathcal{F}_n$ be the set of functions $f : \mathbb{N}_n \longrightarrow \mathbb{N}_n$ such that $f(1) = 1$ and $f(i) < i$ for $i = 2, 3, \ldots n$. There is a bijection between T_n and $\mathcal{F}_n$.*

Proof. Since any vertex $1 \neq v \in V(T)$ is adjacent to exactly one vertex with a lesser label if T is a RRT, one can associate a function $f \in \mathcal{F}$ to T by assigning $f(1) = 1$ and $f(i)$ the father of i. For the converse, take any $f \in F_n$ and build $\{T_i\}_{i=1}^{n}$ by setting T_1 to be the graph with one vertex and no edges and subsequent T_i to be the graph built from T_{i-1} by attaching vertex i to $f(v)$ for $i = 2, 3, \ldots, n$. □

Corollary 1. $|\mathcal{T}_n| = (n-1)!$

Proof. Since $|\mathcal{T}_n| = |\mathcal{F}_n|$ it is enough to enumerate $\mathcal{F}_n$. One can write any $f \in \mathcal{F}_n$ as:

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 & \ldots & n \\ 1 & f(2) & f(3) & f(4) & \ldots & f(n) \end{pmatrix}$$

Subject to $f(1) = 1$ and $f(i) < i$ for $i = 1, 2, \ldots n$. Note that f has 1 *choice* for $f(2)$ (i.e. $f(2) = 1$), two choices for $f(3)$ and, more generally, $i-1$ choices for $f(i-1)$. Therefore, $|\mathcal{F}_n| = (n-1)!$ □

Let $\tilde{\mathcal{T}}_n$ be the set of labelled rooted tree on n vertices. The symmetric group, S_n, can act on $\tilde{\mathcal{T}}_n$ by permuting the non-root vertices of any rooted, labelled tree. Given a permutation $\sigma \in S_n$ and a tree $T \in \tilde{\mathcal{T}}_n$ we write the action of σ on T as $\sigma \cdot T$. Figure 1 shows that this action does not restrict to RRTs. This begs the question: Given $T \in T_n$ and $\sigma \in S_n$ under what conditions is $\sigma \cdot T \in T_n$?

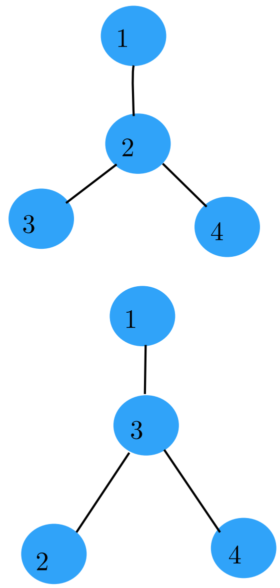

FIGURE 1. The top tree, T, is a RRT on n vertices. The bottom tree, $(2,3)\cdot T$ is obviously not a RRT.

Lemma 1.2. *Let $T \in T_n$ correspond to $f \in \mathcal{F}_n$ then $\sigma \cdot T$ corresponds to the following function:*

$$f' = \begin{pmatrix} 1 & \sigma(2) & \sigma(3) & \sigma(4) & \dots & \sigma(n) \\ 1 & \sigma(f(2)) & \sigma(f(3)) & \sigma(f(4)) & \dots & \sigma(f(n)) \end{pmatrix}$$

Proof. Let $T' = \sigma \cdot T$, there exists some function g corresponding to T' such that:

$$g = \begin{pmatrix} 1 & \sigma(2) & \sigma(3) & \sigma(4) & \dots & \sigma(n) \\ 1 & g(\sigma(2)) & g(\sigma(3)) & g(\sigma(4)) & \dots & g(\sigma(n)) \end{pmatrix}$$

Where $g(\sigma(i))$ is the father of $\sigma(i)$ but it is clear that the father of $\sigma(i)$ is $\sigma(f(i))$ hence $g(i) = \sigma(f(i))$ for $i = 2, 3, \dots, n$. □

Corollary 2. *Let $T \in \mathcal{T}_n$ and $\sigma \in S_n$. Then $\sigma \cdot T \in \mathcal{T}_n$ if and only if $\sigma(f(i)) < \sigma(i)$.*

Remark 1. *If i and j are adjacent vertices in a RRT, T, and T is acted upon by the transposition (i,j) then $\sigma \cdot T \notin \mathcal{T}_n$. Without loss of generality assume that $i < j$. Since i and j are adjacent $f(j) = i$, hence:*

$$\sigma(j) = i < j = \sigma(i) = \sigma(f(j))$$

The result follows from corollary 2.

We define an indicator function for any $\sigma \in S_n$ and $T \in \mathcal{T}_n$ as follows:

$$I(\sigma, T) = \begin{cases} 1 & \text{if } \sigma \cdot T \in \mathcal{T}_n \\ 0 & \text{otherwise} \end{cases}$$

1.1. **Transpositions.** In order to understand the effect of permutations of vertices on RRTs we shall examine $\sigma \cdot T$ where $\sigma = (p, q)$ is a transposition such that (without loss of generality) $p < q$.

By Corollary 2 if $\sigma \cdot T \in T_n$ the corresponding function, f, satisfies $\sigma(f(i)) < \sigma(i)$ for $i = 2, 3, \ldots, n$.

Lemma 1.3. *Given a RRT $T = \{T_i\}_{i=1}^n$ and a transposition $\sigma = (p, q)$ the labelled tree $\sigma \cdot T$ is a RRT if and only if $f(q) < p$ and p is a leaf in T_q .*

Remark 2. *The proof of Lemma 1.3 relies on the fact that that any $f \in \mathcal{F}_n$ can be split up into 5 parts as follows:*

$$f = \left(\begin{array}{ccc|c|ccc|c|ccc} 1 & \ldots & p-1 & p & p+1 & \ldots & q-1 & q & q+1 & \ldots & n \\ f(1) & \ldots & f(p-1) & f(p) & f(p+1) & \ldots & f(q-1) & f(q) & f(q+1) & \ldots & f(n) \end{array} \right)$$

Notice that the first and fifth parts (with domain $i < p$ and $i > q$ respectively) are irrelevant to whether or not $\sigma \cdot T$ is a random recursive tree. It remains to find necessary and sufficient conditions for the second third and fourth parts such that $\sigma \cdot T \in \mathcal{T}_n$.

Proof. [of Lemma 1.3] Let f correspond to a RRT. We can partition the domain of f into 5 sets as follows:

Case 1 $(i < p)$**.** Since T is a RRT $f(i) < i < p$ therefore $\sigma(i) = i$ and $\sigma(f(i)) = f(i)$ so we trivially have $\sigma(f(i)) < \sigma(i)$.

Case 2 $(i = p)$**.** Since T is a RRT $f(p) < p$ so $\sigma(f(p)) = f(p)$. Therefore $\sigma(f(p)) = f(p) < p < q = \sigma(p)$ is always satisfied.

Case 3 $(p < i < q)$**.** Since $i \neq p$ and $i \neq q$, $\sigma(i) = i$. Also note that since T is a RRT $f(i) < i < q$. Therefore, $\sigma \cdot T \in \mathcal{T}_n$ if and only if $\sigma(f(i)) < i$ which is the case if and only if $f(i) \neq p$.

Case 4 $(i = q)$**.** By Remark 1 if $f(q) = p$ then $\sigma \cdot T$ is *not* an RRT. Furthermore, $\sigma \cdot T$ is a RRT if and only if:

$$\begin{aligned} & \sigma(f(q)) < \sigma(q) \\ \iff & \sigma(f(q)) < p \end{aligned}$$

This is the case if and only if $f(q) < p$.

Case 5 $(i > q)$**.** Since $i \neq p$ and $i \neq q$ it is always the case that $\sigma(i) = i$ hence

$$\sigma(f(i)) = \begin{cases} f(i) & \text{if } f(i) \neq p, q \\ p & \text{if } f(i) = q \\ q & \text{if } f(i) = p \end{cases}$$

Since $f(i), p, q < i$ it is always the case that $\sigma(i) < \sigma(f(i))$.

Therefore $\sigma \cdot T \in \mathcal{T}_n$ if and only if $f(q) < p$ *and* $f(i) \neq p$ for $i = p+1, p+2, \ldots, q-1$. Equivalently we could say $\sigma \cdot T \in \mathcal{T}_n$ if and only if $f(q) < p$ and p is a leaf in T_q. $\square$

Lemma 1.4. *For a fixed $\sigma \in S_n$, we write $P_n(\sigma) = \sum_{T \in T_n} I(\sigma, T)$. Then $P_n(p, q) = \frac{(p-1)^2}{(q-1)(q-2)}$*

Proof. Note that $P_n(p,q)$ is the number of trees $T \in \mathcal{T}_n$ such that $(p,q)\cdot T \in \mathcal{T}_n$. By Lemma 1.3, $\sigma = (p,q)\cdot T \in T_n$ if and only if p is a leaf in T_q and $f(q) < p$. Therefore, $P_n(p,q)$ is the number of trees $T \in \mathcal{T}_n$ such that p is a leaf in T_q and $f(q) < p$.

For every $T \in \mathcal{T}_n$ the associated function f can be split up into 5 parts as described in Remark 2, in particular the following matrix shows the number of possible values of $f(i)$ such that $\sigma \cdot T \in \mathcal{T}_n$ given each i:

$$f = \left(\begin{array}{cccc|c|ccc|c|ccc} 1 & 2 & 3 & \dots & p & p+1 & \dots & q-1 & q & q+1 & \dots & n \\ 1 & 1 & 2 & \dots & p-1 & p-1 & \dots & q-2 & p-1 & q & \dots & n-1 \end{array}\right)$$

Therefore,

$$P_n(p,q) = \frac{(p-1)^2}{(q-1)(q-2)}(n-1)!$$

□

2. n-CYCLES

In this section we will generalise Lemma 1.3 to permutations $\sigma = (p_1, p_2, \dots, p_m)$ such that $p_1 < p_2 < \dots < p_m$.

Lemma 2.1. *Given a RRT $\{T = T_i\}_{i=1}^n$ and a transposition $\sigma = (p_1, p_2, \dots, p_m)$ the labelled tree $\sigma \cdot T$ is a RRT if and only if p_l is a leaf in $T_{p_{l+1}}$ for $l = 1, 2, \dots, n-1$ and $f(p_m) < p_1$.*

Proof. The proof of Lemma 2.1 follows closely the proof of Lemma 1.3. Let $T \in \mathcal{T}_n$ and $f \in \mathcal{F}_n$ be the corresponding function. Using the same argument from the proof of Lemma 1.3 we can see immediately that there are no conditions on $f(i)$ for $i < p_1$ and $i > p_m$ for $\sigma \cdot T \in T_n$. Similarly we can immediately see that there is no condition on $f(p_1)$ order for $\sigma \cdot T \in T_n$.

For all $p_l < i < p_{l+1}$ such that $l = 1, 2, \dots, n-1$, if $f(i) = p_l$ then $\sigma(i) = i$ and $\sigma(f(i)) = p_{l+1}$. This means that:

$$\sigma(i) = i < p_{l+1} = \sigma f(i),$$

so $\sigma \cdot T$ is *not* a RRT.

We now need only consider $i = p_l$ for $l = 2, 3, \dots, m$. If $\sigma \cdot T \in T_n$ then $\sigma(f(i)) < \sigma(i)$. Therefore, $\sigma \cdot T \in T_n$ if $\sigma(f(i)) < p_{l+1} = \sigma f(i)$ for $l = 1, 2, \dots, n-1$, this is clearly always satisfied.

Finally we consider p_m. Note that if $\sigma \cdot T \in \mathcal{T}_n$ then $\sigma f(p_m) < \sigma(p_m) = p_1$.

We conclude that $\sigma \cdot T \in T_n$ if and only if p_l is *not* either a leaf in $T_{p_{l+1}}$ or there exists the edge $(p_l p_{l+1})$ then $\sigma \cdot T$ is not a RRT for $l = 1, 2, \dots, n-1$ and $f(p_m) < p_1$. □

Corollary 3. *Let $\sigma \in S_n$ have the form $\sigma = (p_1, p_2, p_3, \dots p_m)$ such that $p_1 < p_2 < \dots < p_m$.*

$$P_n(\sigma) = (n-1)!\frac{(p_1-1)^2(p_2-1)(p_3-1)\dots(p_m-1)}{(p_2-2)(p_3-2)\dots(p_m-2)(p_m-1)}.$$

Proof. We use a similar method to the proof of Lemma 1.4. Recall that we can think of $P_n(\sigma)$ as the number of trees $T \in \mathcal{T}_n$ such that $\sigma \cdot T \in \mathcal{T}_n$. By Lemma 2.1, given a RRT $\{T = T_i\}_{i=1}^n$ and a transposition $\sigma = (p_1, p_2, \dots, p_m)$ the labelled tree $\sigma \cdot T$ is a RRT if and only if p_l is a leaf in $T_{p_{l+1}}$ for $l = 1, 2, \dots, n-1$ and $f(p_m) < p_1$.

We therefore know the number of possible values that $f(i)$ can take for each i. The following matrix shows the number of possible values of $f(i)$ for $p_1 < i < p_m$.

$$f = \begin{pmatrix} \dots & p_1 & p_1+1 & \dots & p_2-1 & p_2 & p_2+1 & \dots & p_m-1 & p_m & \dots \\ \dots & p_1-1 & p_1-1 & \dots & p_2-3 & p_2-1 & p_2-1 & \dots & p_m-3 & p_1-1 & \dots \end{pmatrix}$$

□

Remark 3. *Let $T \in T_n$ and a permutation $\sigma = (p_1, p_2, \dots, p_m)$ such that $p_1 < p_2 < \dots p_m$ and denote transpositions $\sigma_l = (p_1, p_l)$ for $l = 1, 2, \dots, m$. It is interesting to note that given a random recursive tree T such that $\sigma \cdot T \in \mathcal{T}_n$ it is* not *necessarily the case that $\sigma_l \cdot T \in T_n$. If we let T be the RRT given in Figure 2 it is clear that $(234) \cdot T \in \mathcal{T}_4$ but $(23) \cdot T, (24) \cdot T \notin \mathcal{T}_4$.*

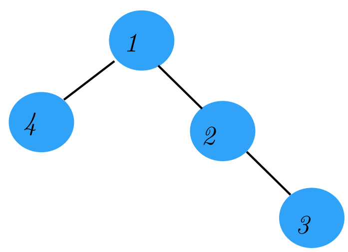

FIGURE 2

On the other hand we can make the following corollary.

Corollary 4. *Let $T \in T_n$ and a permutation $\sigma = (p_1, p_2, \dots, p_m)$ such that $p_1 < p_2 < \dots p_m$ and $\sigma \cdot T \in \mathcal{T}_n$. Let $\sigma' = (p_l, p_l + 1, \dots, p_m)$, then $\sigma' \cdot T \in \mathcal{T}_n$ for $l = 1, 2, \dots m-1$.*

Remark 4. *The converse of Corollary 4 is not true; take Figure 2 for example.*

3. Enumeration of legal moves part 2

In this section we will calculate $Q_n = \sum_{\sigma \in S_n} \sum_{T \in \mathcal{T}_n}$ for certain specific cases of σ beginning with transpositions. We can write out all transpositions $(p, q) \in S_n$ in a grid form as follows:

$$\begin{array}{c|c|c|c|c} (2,3) & (2,4) & (2,5) & \dots & (2,n) \\ & (3,4) & (3,5) & \dots & (3,n) \\ & & (4,5) & \dots & (4,n) \\ & & & & \vdots \\ & & & & (n-1,n) \end{array}$$

By Lemma 1.4, for any transposition $\sigma(p, q)$

$$P_n = |\mathcal{T}_n| \frac{(p-1)^2}{(q-1)(q-2)}$$

To calculate Q_n we simply sum P_n over the columns of the grid above:

$$\begin{aligned}
Q_n &= |\mathcal{T}_n| \sum_{i=1}^{n-2} \left(\frac{\sum_{j=1}^{i} j^2}{i(i+1)} \right) \\
&= |\mathcal{T}_n| \frac{1}{6} \sum_{i=1}^{n-2} \frac{i(i+1)(2i+1)}{i(i+1)} \\
&= |\mathcal{T}_n| \frac{1}{6} \sum_{i=1}^{n-2} 2i+1 \\
&= \frac{|\mathcal{T}_n|}{6} n(n-2)
\end{aligned}$$

4. Notation

$\{T_i\}_{i=1}^n$	A random recursive tree process on n vertices.
$\mathcal{T}_n$	The set of of random recursive tree processes on n vertices.
$\tilde{\mathcal{T}}_n$	The set of labelled rooted tree on n vertices.
S_n	the symmetric group on n elements.
$I(\sigma, T)$	$\begin{cases} 1 & \text{if } \sigma \cdot T \in \mathcal{T}_n \\ 0 & \text{otherwise} \end{cases}$
$P_n(\sigma)$	$\sum_{T \in T_n} I(\sigma, T)$
Q_n	$\sum_{\sigma \in S_n} \sum_{T \in \mathcal{T}_n}$