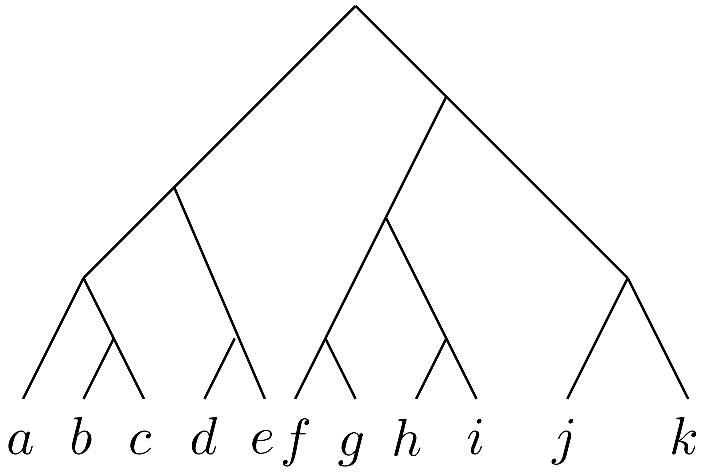
The Age of a Vertex

Tanja Gernhard

15.11.2006

The Age of a Vertex

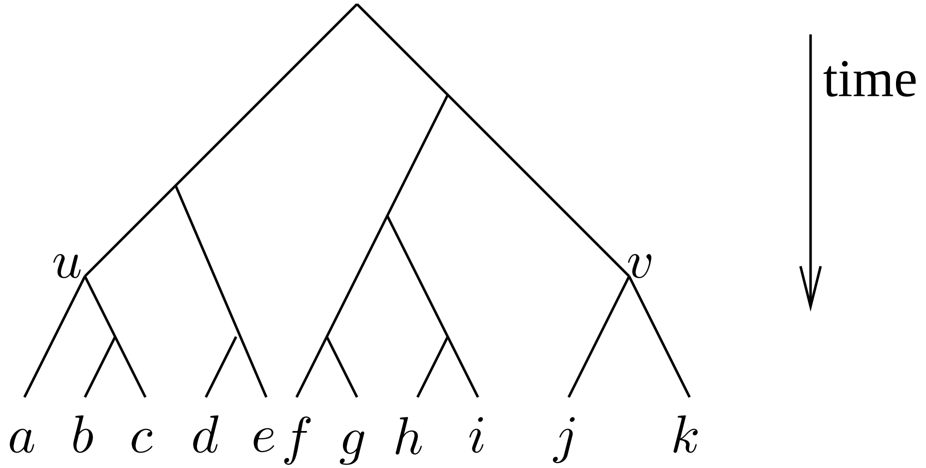
A tree displays the relationships between species.



The Age of a Vertex

A tree displays the relationships between species.

Can we say something about the relative time?

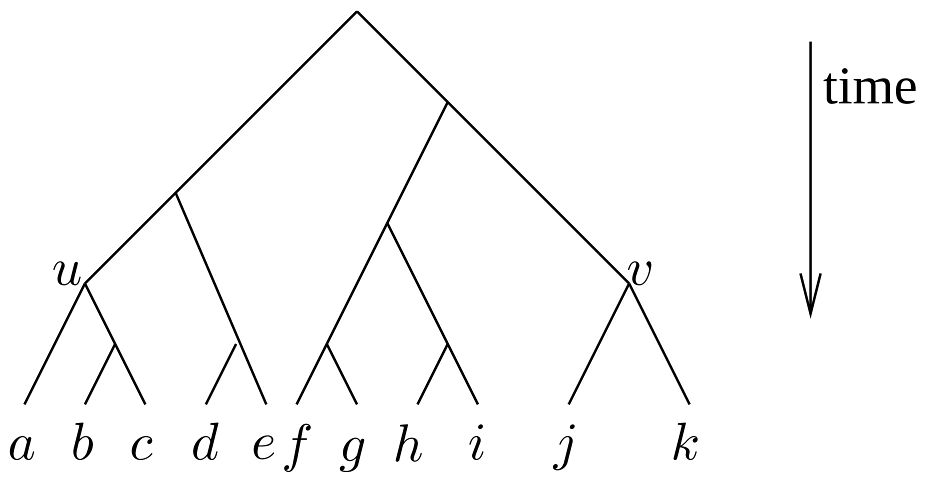


The Age of a Vertex

A tree displays the relationships between species.

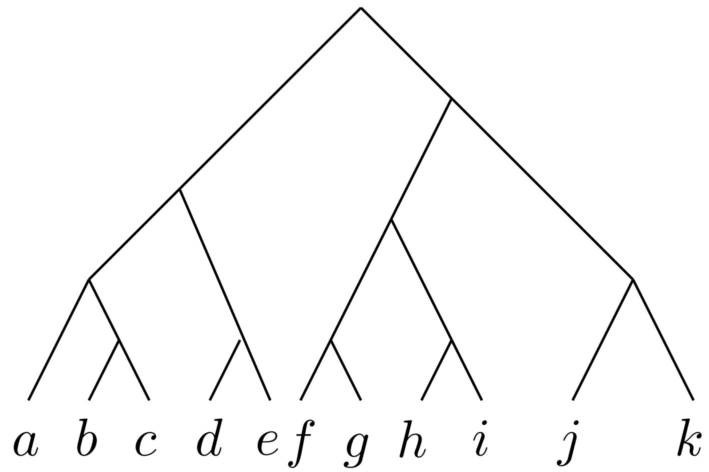
Can we say something about the relative time?

Is an inner vertex an early or late event compared to the other vertices?

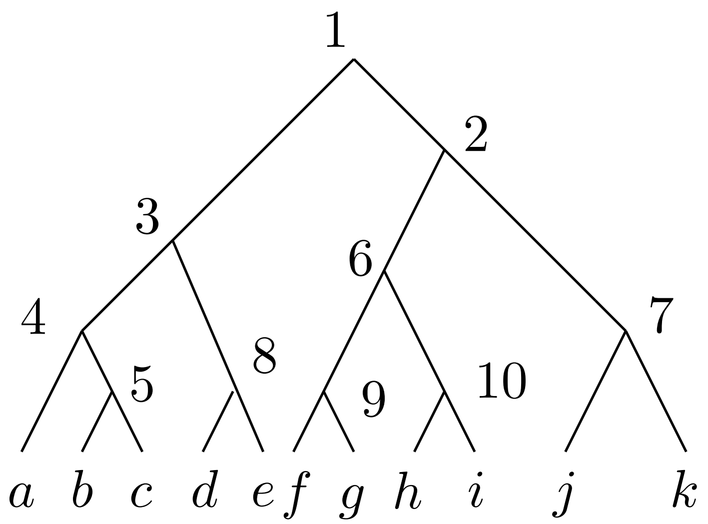


The Rank Function

We want to put the inner vertices in a linear order.



We want to put the inner vertices in a linear order.



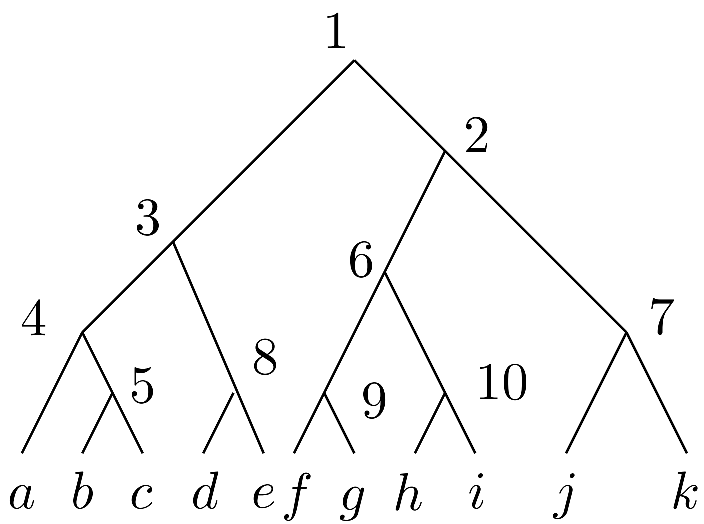
The Rank Function

We want to put the inner vertices in a linear order.

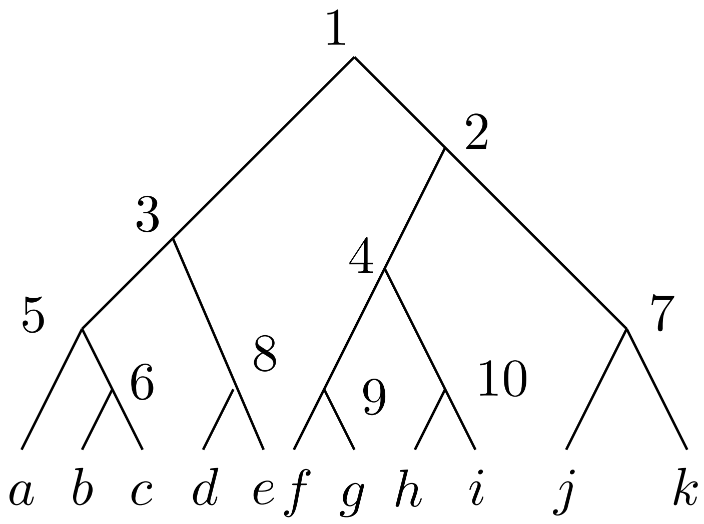
A rank function r on a tree induces a linear order on the inner vertices:

$$r : \mathring{V} \to \{1, \ldots, |\mathring{V}|\}$$

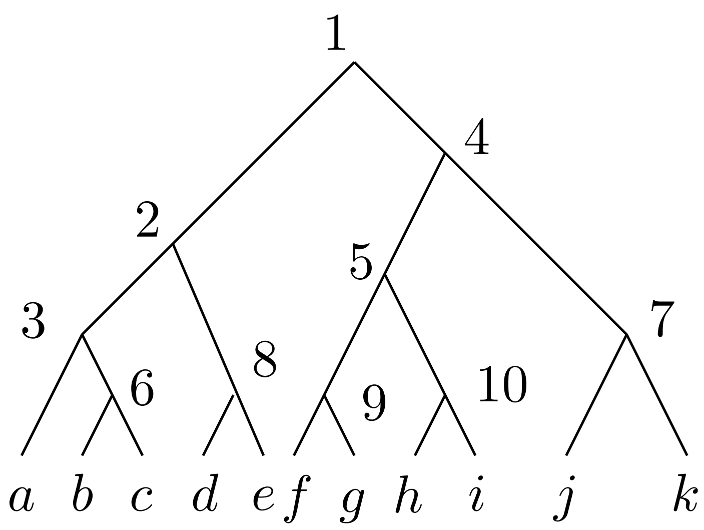
r bijection with $r(v) < r(w)$ if vertex v is an ancestor of w.



The Rank Function



The Rank Function



A neutral Model

We need a stochastic model for speciation - we assume a pure chance model (neutral model).

A neutral Model

We need a stochastic model for speciation - we assume a pure chance model (neutral model).

Under a pure chance model, each pendant edge is equally likely to split.

A neutral Model

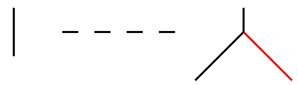
We need a stochastic model for speciation - we assume a pure chance model (neutral model).

Under a pure chance model, each pendant edge is equally likely to split.

A neutral Model

We need a stochastic model for speciation - we assume a pure chance model (neutral model).

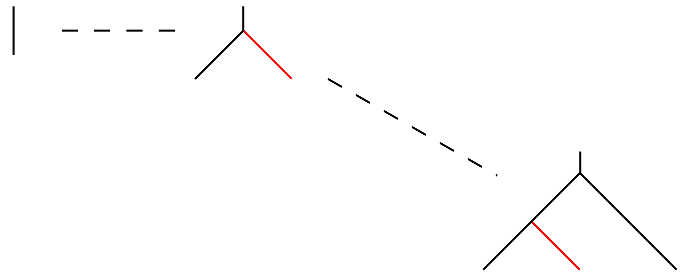
Under a pure chance model, each pendant edge is equally likely to split.



A neutral Model

We need a stochastic model for speciation - we assume a pure chance model (neutral model).

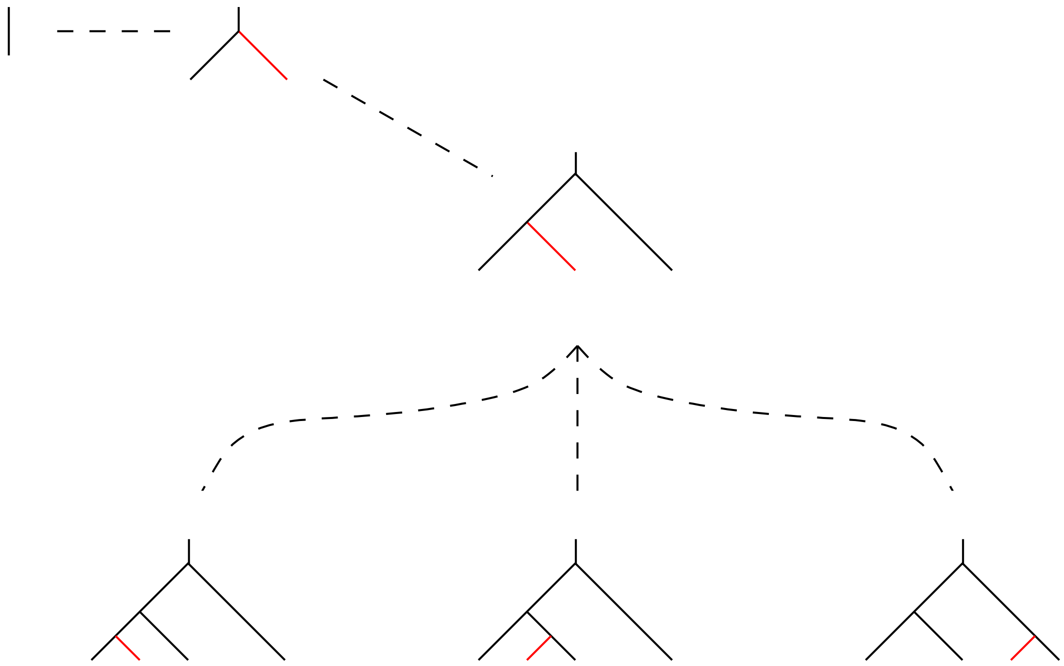
Under a pure chance model, each pendant edge is equally likely to split.



A neutral Model

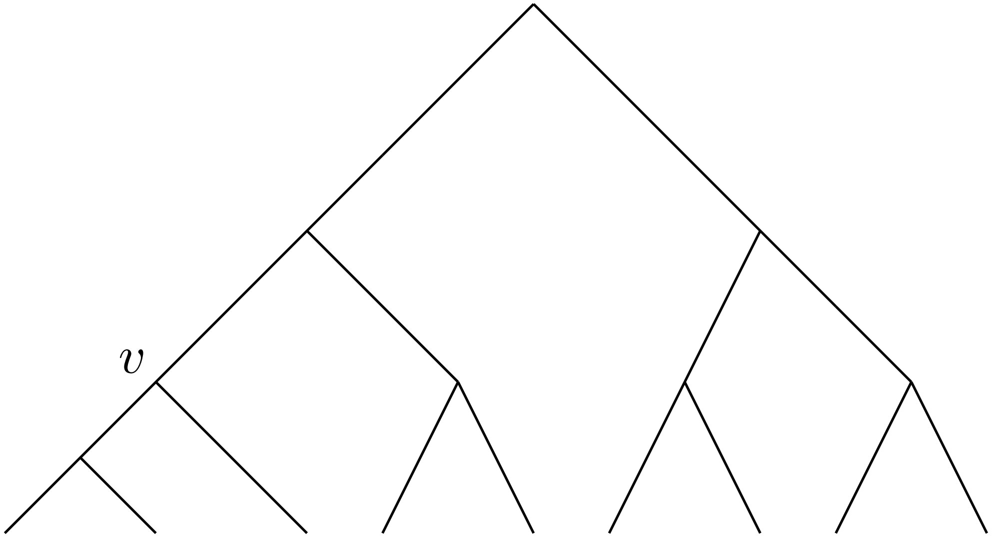
We need a stochastic model for speciation - we assume a pure chance model (neutral model).

Under a pure chance model, each pendant edge is equally likely to split.



Rank of a Vertex

Under a neutral model, how likely is it that vertex v has rank 5, i.e. $r(v) = 5$?



Rank of a Vertex

- Under a neutral model, each ranked phylogenetic tree is equally likely.

Rank of a Vertex

- Under a neutral model, each ranked phylogenetic tree is equally likely.
- That means, $\mathbb{P}[r(v) = i] = \frac{|\{r : r(v) = i, r \in r(\mathcal{T})\}|}{|r(\mathcal{T})|}$
  where $r(\mathcal{T})$ is the set of all rank functions on $\mathcal{T}$.

Rank of a Vertex

- Under a neutral model, each ranked phylogenetic tree is equally likely.
- That means, $\mathbb{P}[r(v) = i] = \frac{|\{r : r(v) = i, r \in r(\mathcal{T})\}|}{|r(\mathcal{T})|}$ where $r(\mathcal{T})$ is the set of all rank functions on $\mathcal{T}$.
- It is known that $|r(\mathcal{T})| = \frac{|\mathring{V}|!}{\prod_{v \in \mathring{V}} \lambda_v} =: R_{\mathcal{T}}$.

Rank of a Vertex

- Under a neutral model, each ranked phylogenetic tree is equally likely.
- That means, $\mathbb{P}[r(v) = i] = \frac{|\{r : r(v) = i, r \in r(\mathcal{T})\}|}{|r(\mathcal{T})|}$ where $r(\mathcal{T})$ is the set of all rank functions on $\mathcal{T}$.
- It is known that $|r(\mathcal{T})| = \frac{|\mathring{V}|!}{\prod_{v \in \mathring{V}} \lambda_v} =: R_{\mathcal{T}}$.
- We developed a quadratic algorithm to calculate $|\{r : r(v) = i, r \in r(\mathcal{T})\}| =: \alpha_{\mathcal{T},v}(i)$.

Rank of a Vertex

- Under a neutral model, each ranked phylogenetic tree is equally likely.
- That means, $\mathbb{P}[r(v) = i] = \frac{|\{r : r(v)=i, r \in r(\mathcal{T})\}|}{|r(\mathcal{T})|}$ where $r(\mathcal{T})$ is the set of all rank functions on $\mathcal{T}$.
- It is known that $|r(\mathcal{T})| = \dfrac{|\mathring{V}|!}{\prod_{v \in \mathring{V}} \lambda_v} =: R_{\mathcal{T}}$.
- We developed a quadratic algorithm to calculate $|\{r : r(v) = i, r \in r(\mathcal{T})\}| =: \alpha_{\mathcal{T},v}(i)$.

Algorithm RankCountBin

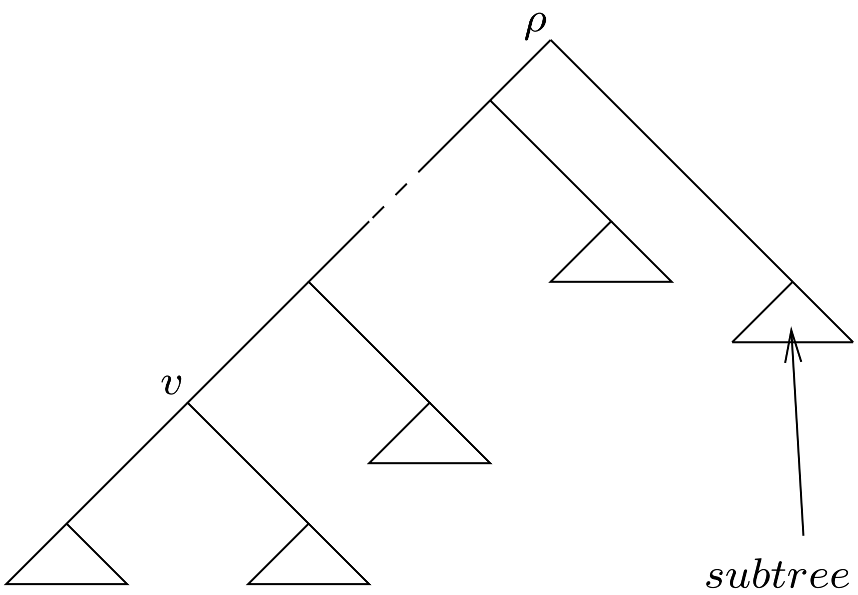
- **Input**: A rooted binary phylogenetic tree $\mathcal{T}$ and an interior vertex v.

Algorithm RankCountBin

- **Input**: A rooted binary phylogenetic tree $\mathcal{T}$ and an interior vertex v.
- **Output**: The values of $\alpha_{\mathcal{T},v}(i)$ for $i = 1, \ldots, |\mathring{V}|$.

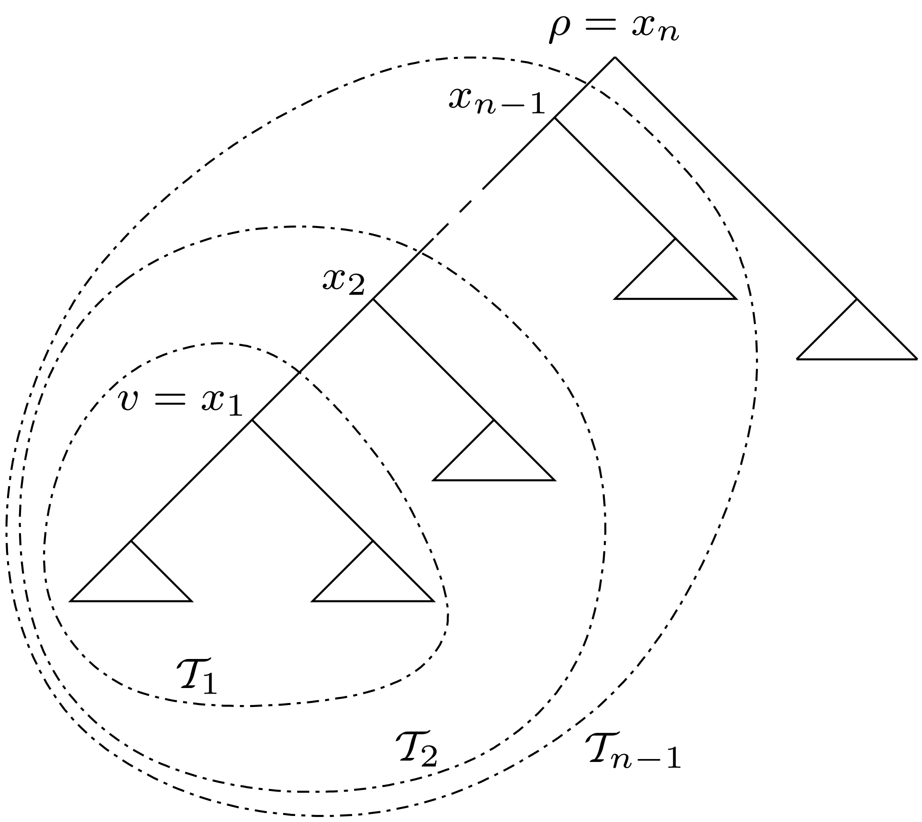
Algorithm RankCountBin

- **Input**: A rooted binary phylogenetic tree $\mathcal{T}$ and an interior vertex v.
- **Output**: The values of $\alpha_{\mathcal{T},v}(i)$ for $i = 1, \ldots, |\mathring{V}|$.

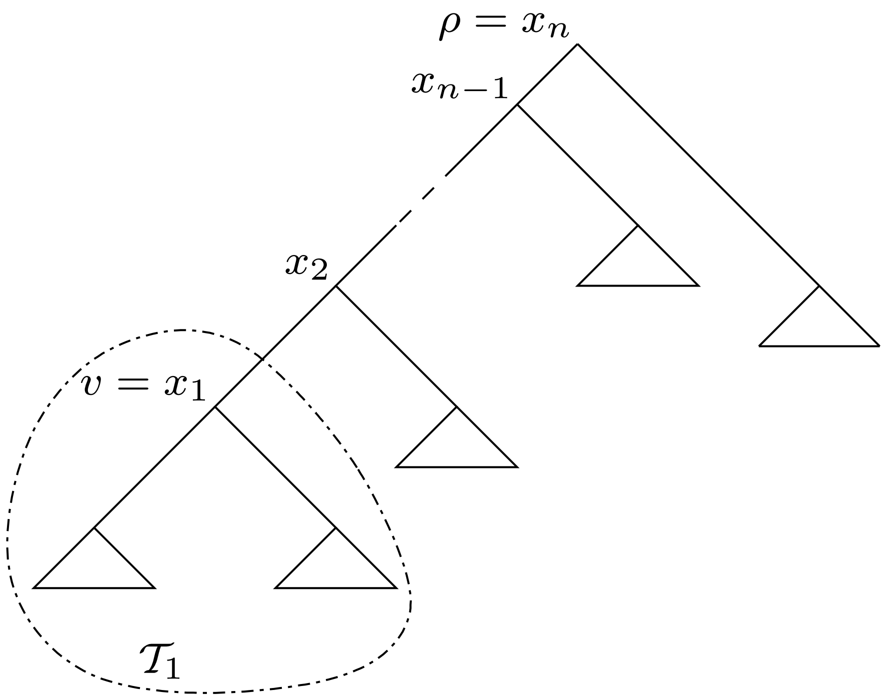


Algorithm RankCountBin

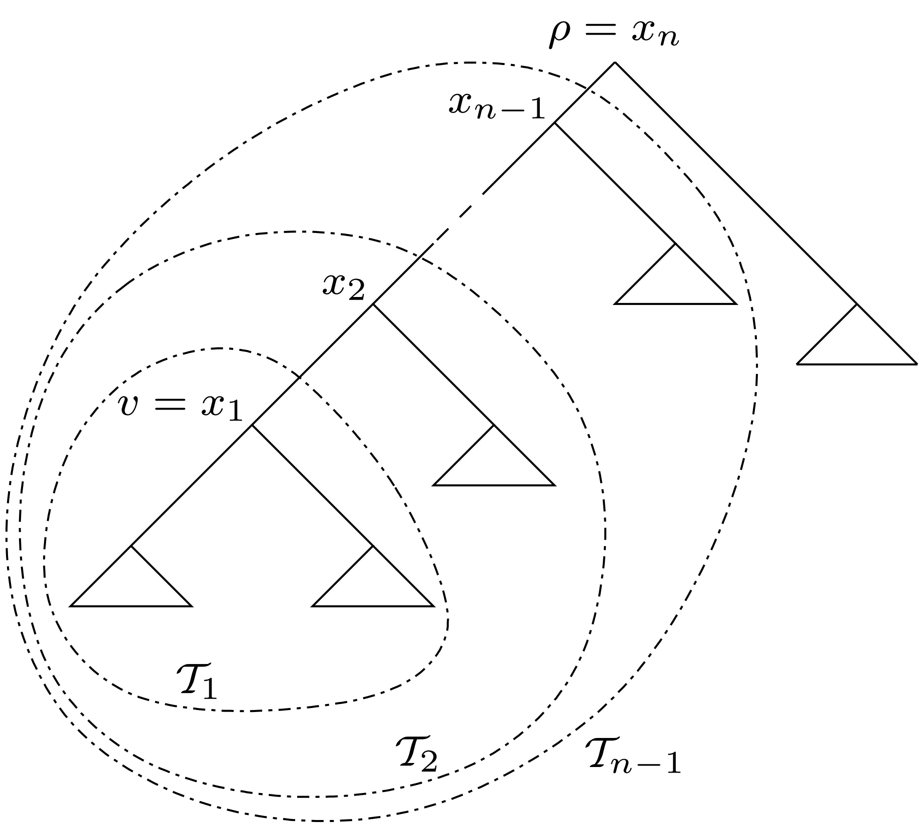
- **Input**: A rooted binary phylogenetic tree $\mathcal{T}$ and an interior vertex v.
- **Output**: The values of $\alpha_{\mathcal{T},v}(i)$ for $i = 1, \ldots, |\mathring{V}|$.



$$\alpha_{\mathcal{T}_1,v}(1) := \frac{|\mathring{V}_{\mathcal{T}_1}|!}{\prod_{v \in \mathring{V}_{\mathcal{T}_1}} \lambda_v}$$

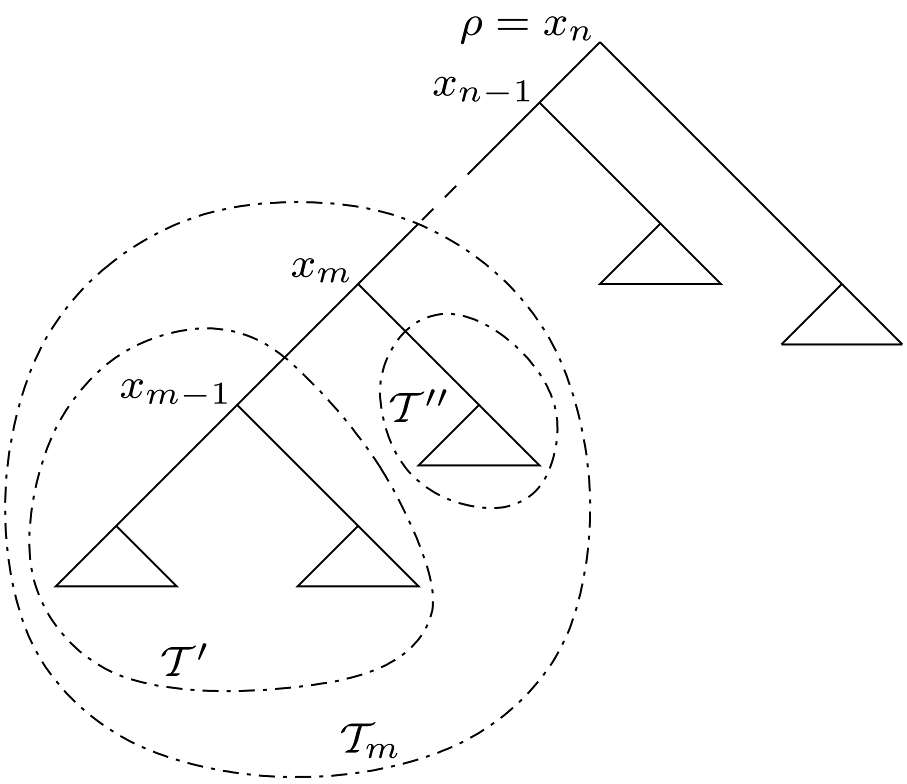


$$\alpha_{\mathcal{T}_m,v}(i) := \sum_{j=0}^{\min\{i-2,|\mathring{V}_{\mathcal{T}''}|\}} \alpha_{\mathcal{T}',v}(i-j-1) R_{\mathcal{T}''} \binom{|\mathring{V}_{\mathcal{T}'}| + |\mathring{V}_{\mathcal{T}''}| - (i-1)}{|\mathring{V}_{\mathcal{T}''}| - j} \binom{i-2}{j}$$



Algorithm RankCountBin - *The Recursion*

$$\alpha_{\mathcal{T}_m,v}(i) := \sum_{j=0}^{\min\{i-2,|\mathring{V}_{\mathcal{T}''}|\}} \alpha_{\mathcal{T}',v}(i-j-1) R_{\mathcal{T}''} \binom{|\mathring{V}_{\mathcal{T}'}| + |\mathring{V}_{\mathcal{T}''}| - (i-1)}{|\mathring{V}_{\mathcal{T}''}| - j} \binom{i-2}{j}$$



Probability of $r(v) = i$

Our goal was to calculate $\mathbb{P}[r(v) = i]$.

Probability of $r(v) = i$

Our goal was to calculate $\mathbb{P}[r(v) = i]$.

Since each ranked tree is equally likely, we have

$$\mathbb{P}[r(v) = i] = \frac{\alpha_{\mathcal{T},v}(i)}{R_{\mathcal{T}}}.$$

Probability of $r(v) = i$

Our goal was to calculate $\mathbb{P}[r(v) = i]$.

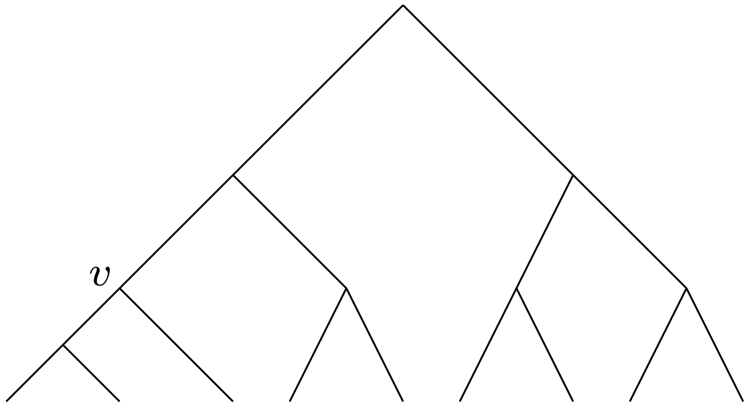
Since each ranked tree is equally likely, we have

$$\mathbb{P}[r(v) = i] = \frac{\alpha_{\mathcal{T},v}(i)}{R_{\mathcal{T}}}.$$

Example

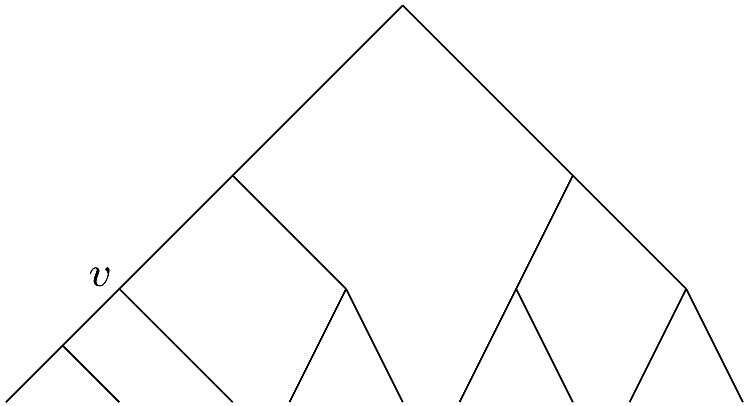
Example

Consider the following tree:



Example

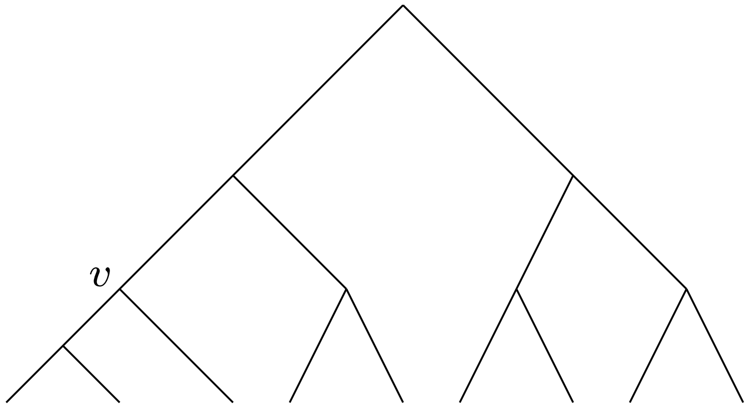
Consider the following tree:



$$\begin{aligned}
\mathbb{P}[r(v)=1] &= 0\\
\mathbb{P}[r(v)=2] &= 0\\
\mathbb{P}[r(v)=3] &= \frac{20}{93}\\
\mathbb{P}[r(v)=4] &= \frac{16}{93}\\
\mathbb{P}[r(v)=5] &= \frac{27}{93}\\
\mathbb{P}[r(v)=6] &= \frac{20}{93}\\
\mathbb{P}[r(v)=7] &= \frac{10}{93}\\
\mathbb{P}[r(v)=8] &= 0
\end{aligned}$$

Example

Consider the following tree:



$$\mu_{r(v)} = \sum_{i=1}^{8} i\mathbb{P}[r(v) = i) \approx 4.83$$

$$\sigma_{r(v)}^2 = \sum_{i=1}^{8} i^2\mathbb{P}[r(v) = i) - \mu_{r(v)}^2 \approx 1.65$$

Extensions of RankCountBin

- RankCountBin can be generalized to RankCount for non-binary trees.

Extensions of RankCountBin

- RankCountBin can be generalized to RankCount for non-binary trees.
  The runtime is $O(|V|^2)$.

Extensions of RankCountBin

- RankCountBin can be generalized to RankCount for non-binary trees.
  The runtime is $O(|V|^2)$.

- Algorithm Compare compares two inner vertices - which one is more likely to be earlier?

Extensions of RankCountBin

- RankCountBin can be generalized to RankCount for non-binary trees.
  The runtime is $O(|V|^2)$.

- Algorithm Compare compares two inner vertices - which one is more likely to be earlier?
  Compare runs in $O(|V|^2)$ and uses RankCountBin and RankCount.

Expected edge lengths

Let X be the random variable ‘length of edge $e = (u, v)$’ with an exp(1)-distribution (Yule model).

Expected edge lengths

Let X be the random variable 'length of edge $e = (u, v)$' with an exp(1)-distribution (Yule model).

$$\mathbb{E}[X] = \sum_{i,j} \mathbb{E}[X | r(u) = i, r(v) = j] \mathbb{P}[r(u) = i, r(v) = j]$$

Expected edge lengths

Let X be the random variable 'length of edge $e = (u, v)$' with an exp(1)-distribution (Yule model).

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i,j} \mathbb{E}[X | r(u) = i, r(v) = j] \mathbb{P}[r(u) = i, r(v) = j] \\
&= \dots \\
&= \frac{\sum_{i<j} \left[\left(\sum_{k=1}^{j-i} \frac{1}{i+k} \right) \cdot \mathbb{P}[r(u) = i] \cdot \prod_{k=0}^{|\mathring{V}_{\mathcal{T}_v}|-2} (|\mathring{V}_{\mathcal{T}}| - j - k) \right]}{\sum_{i<j} \left[\mathbb{P}[r(u) = i] \cdot \prod_{k=0}^{|\mathring{V}_{\mathcal{T}_v}|-2} (|\mathring{V}_{\mathcal{T}}| - j - k) \right]}
\end{aligned}$$

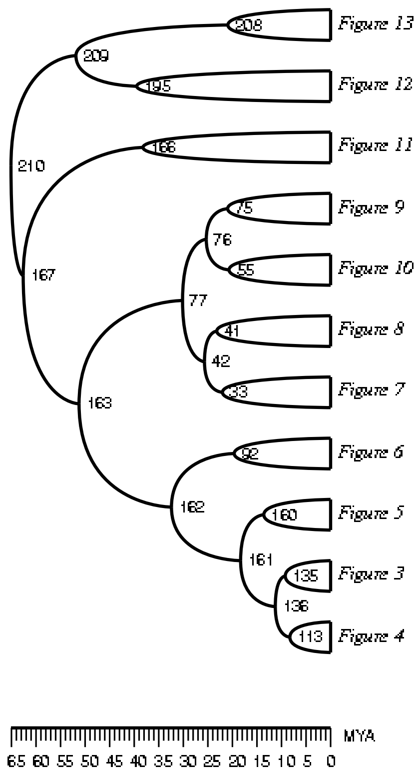
Expected edge lengths

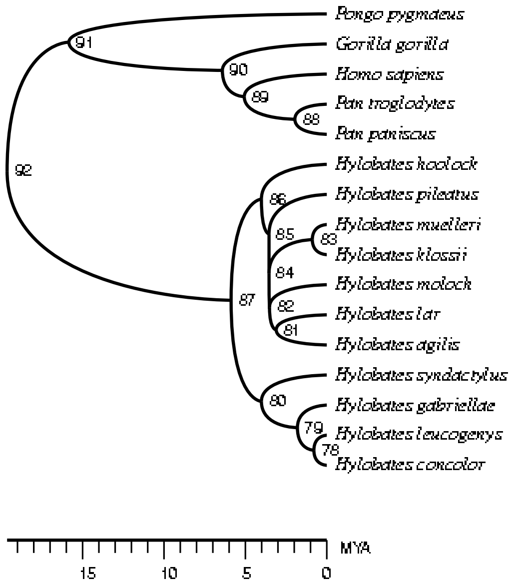
Let X be the random variable 'length of edge $e = (u, v)$' with an exp(1)-distribution (Yule model).

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i,j} \mathbb{E}[X | r(u) = i, r(v) = j] \mathbb{P}[r(u) = i, r(v) = j] \\
&= \dots \\
&= \frac{\sum_{i<j} \left[\left(\sum_{k=1}^{j-i} \frac{1}{i+k} \right) \cdot \mathbb{P}[r(u) = i] \cdot \prod_{k=0}^{|\mathring{V}_{\mathcal{T}_v}|-2} (|\mathring{V}_{\mathcal{T}}| - j - k) \right]}{\sum_{i<j} \left[\mathbb{P}[r(u) = i] \cdot \prod_{k=0}^{|\mathring{V}_{\mathcal{T}_v}|-2} (|\mathring{V}_{\mathcal{T}}| - j - k) \right]}
\end{aligned}$$

This expected value can be calculated with RANKCOUNT.

Edge lengths for a primate tree:





Further projects

- Include extinction into the neutral model (still uniform distribution on ranked trees for the neutral case)
- Model selection
- Parsimony network reconstruction