(see Felsenstein, J. (1973). Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Systematic Biology*, *22*(3), 240-249.)

Rooted tree $T = (V, E)$ with root $r \in V$. A leaf of T is a node with no children and we call the set of all leaves $L(T)$. Consider a discrete, finite alphabet Σ and let $d := |\Sigma|$.

Let $\mathcal{D}$ be a data matrix $\in \mathbb{N}^{|L(T)| \times d}$ that encodes the symbol that has been observed at each leaf.

An *evolutionary tree* is a tuple (T, θ) with a rooted tree $T = (V, E)$ where each node $V_i \in V$ is a random variable with values in Σ and parameters $\theta = (\tau, Q, \pi)$. For each $e \in E$ τ_e is the evolutionary time along the tree edge, $Q \in \mathbb{R}^{d \times d}$ is a rate matrix and π is the equilibrium distribution at the root (also see: probabilistic model of evolution).
Let T_θ be a tree parameterized by θ.

Goal: Estimate $P(\mathcal{D}|T, \theta)$.

Let $\mathcal{D}_{|u}$ for any $u \in V \setminus L(T)$ denote the data restricted to leaves below u.

Algorithm (dynamic programming)

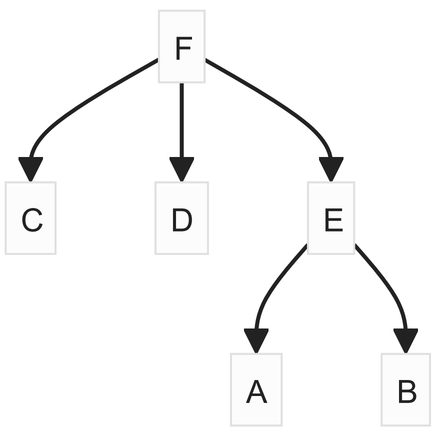
Input: T_σ, $\mathcal{D}$
Output: $\alpha(u, v) = P(\mathcal{D}_{|u}|V_u = v, T)$ for all $u \in V \setminus L(T)$

The $\alpha(u, v)$ are computed dynamically starting with leaf edges. This dependings on the $P_{a,b}$ of a substitution model (see: probabilistic model of evolution).

Then we have $P(\mathcal{D}|T) = \Sigma_v \alpha(r, v) * \pi_v$

Example

T



$\mathcal{D}$

	A	B	C	D
1	1	0	0	0
2	0	1	0	0
3	0	0	0	0
4	0	0	1	1

$d = 4$

Let $\tau = 1$ for all edges.

Model: Jukes-Cantor
$P^{\tau}_{i,i} = \frac{1}{4} + \frac{3}{4}\exp(-\frac{4}{3}\tau)$
$P^{\tau}_{i,j} = \frac{1}{4} - \frac{1}{4}\exp(-\frac{4}{3}\tau)$

Compute $P(D|T) = P(A = 1, B = 2, C = D = 4|T)$ with the following steps:

1. $P(A = 1|E) = P^{\tau_{E,A}} D_{;A} \approx (\frac{7}{16}, \frac{3}{16}, \frac{3}{16}, \frac{3}{16})$
2. $P(B = 2|E) = P^{\tau_{E,B}} D_{;B} \approx (\frac{3}{16}, \frac{7}{16}, \frac{3}{16}, \frac{3}{16})^T$
3. $P(C = 4|F) = P^{\tau_{F,C}} D_{;C} \approx (\frac{3}{16}, \frac{3}{16}, \frac{3}{16}, \frac{7}{16})^T$
4. $P(D = 4|F) = P^{\tau_{F,D}} D_{;D} \approx (\frac{3}{16}, \frac{3}{16}, \frac{3}{16}, \frac{7}{16})^T$

5. $P(A=1,B=2|E) = P(A|E)P(B|E) \approx (\frac{21}{256}, \frac{21}{256}, \frac{9}{256}, \frac{9}{256})^T$
6. $P(A=1,B=2|F) = P^{\tau_{E,D}} P(A=1,B=2|E) \approx (0.064, 0.064, 0.053, 0.053)^T$
7. $P(A=1,B=2,C=4,D=4|F) = P(A=1,B=2|F)P(C=4|F) = P(D=4|F) \approx (0.002, 0.002, 0.0018, 0.01)^T$
   Symbol 4 is more likely at the root than 1,2,3, since we observed it 2 times at C and D and require 2 substitutions for A and B, whereas for any other symbol, we require at least 3 mutations.