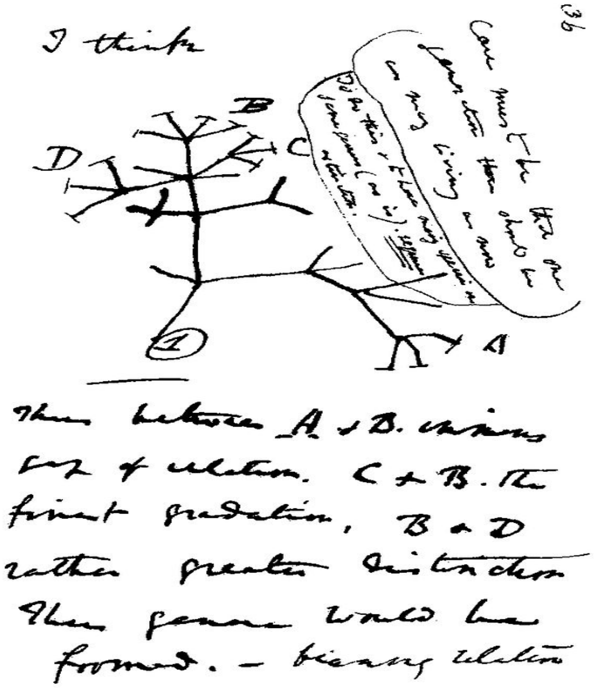
Fitting Tree Metrics with Minimum Disagreements

Evangelos Kipouridis

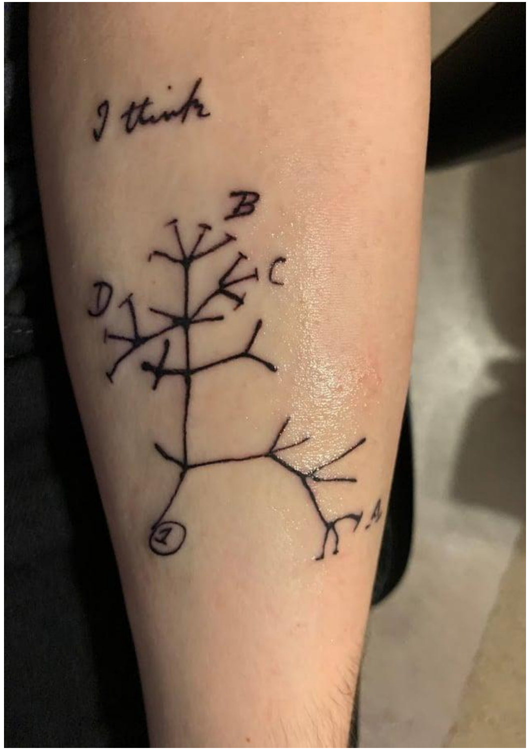
UNIVERSITÄT
DES
SAARLANDES

max planck institut
informatik

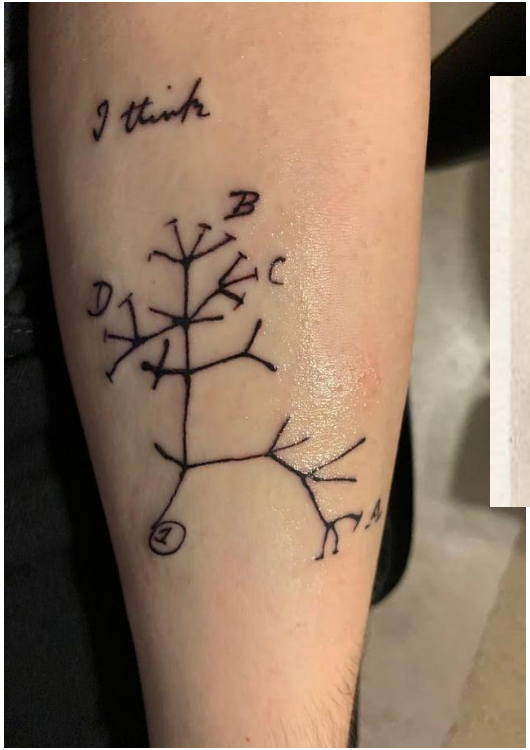
Any guess?

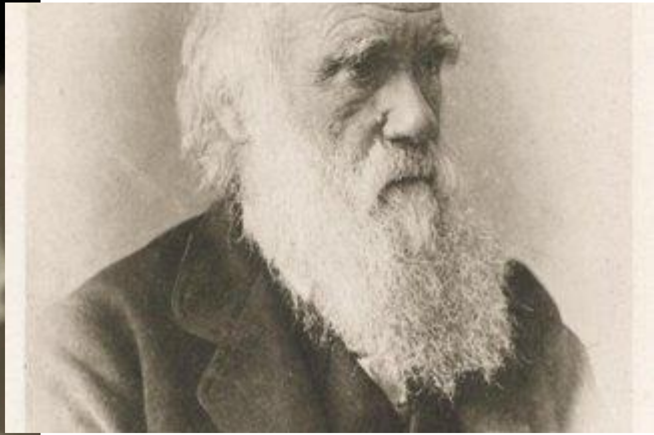

I think
B
C
D
1
A
Thus between A & B. immense gap of relation. C & B. the finest gradation, B & D rather greater distinction Thus genera would be formed. — bearing relation

Any guess?

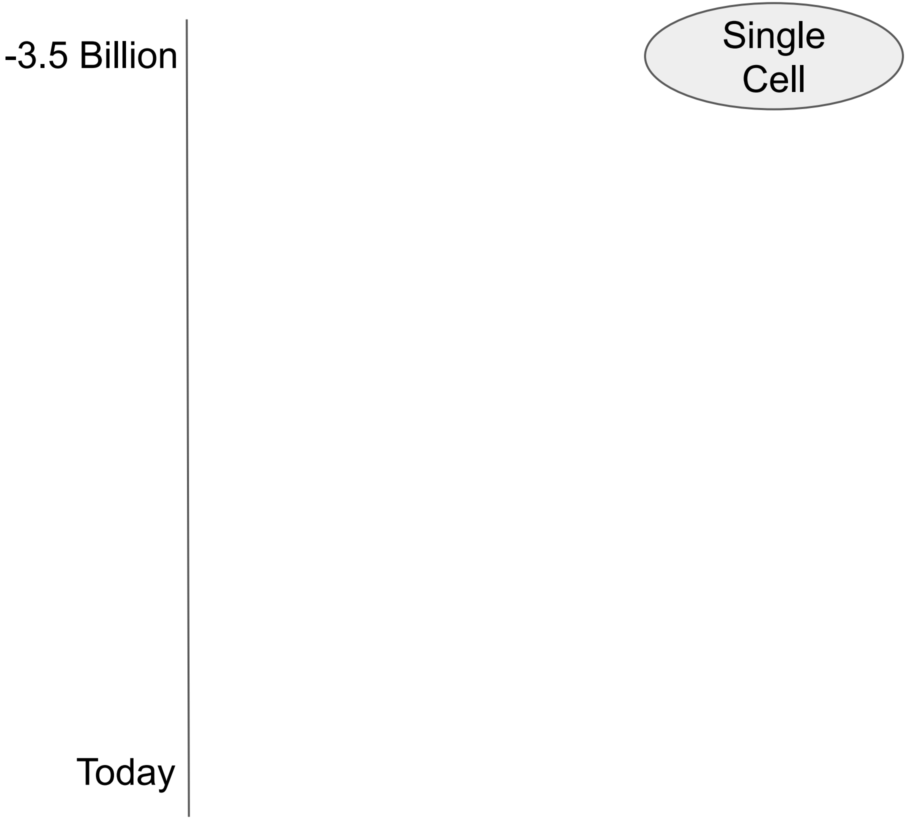


Darwin’s notes

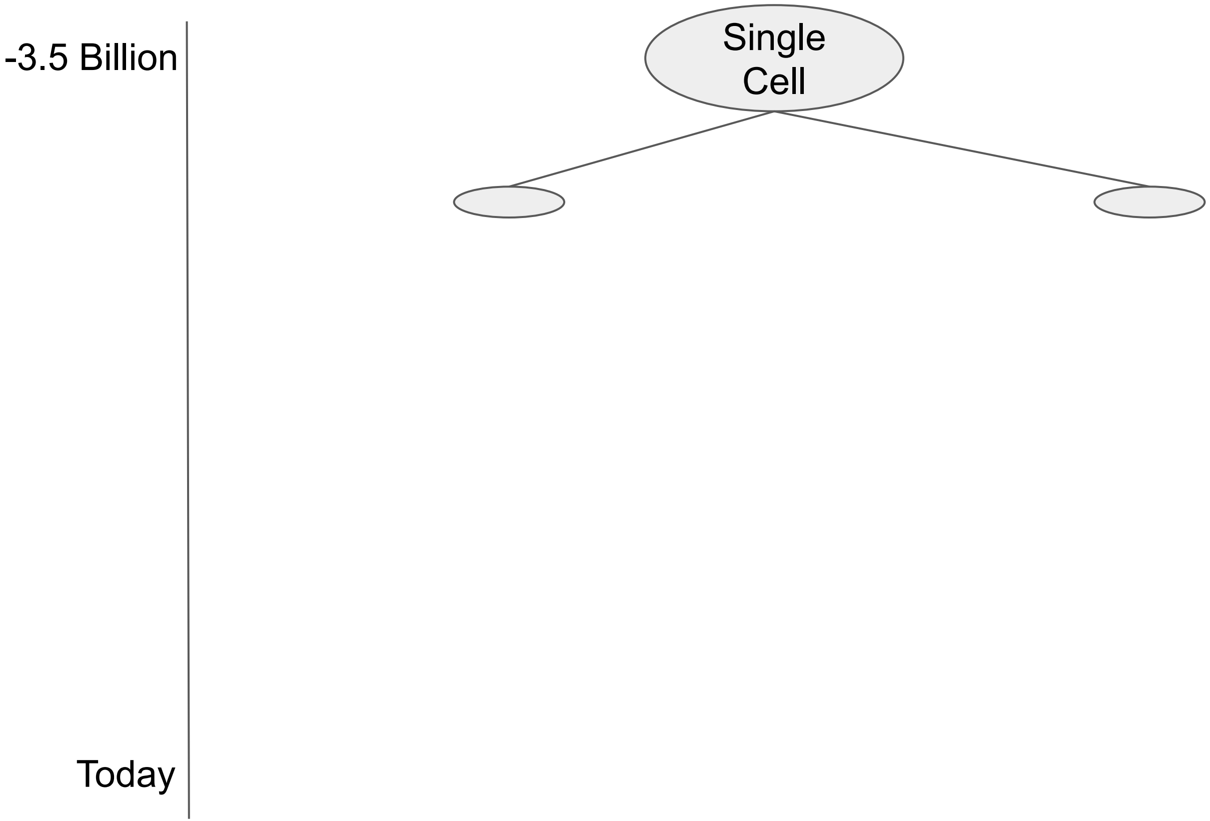




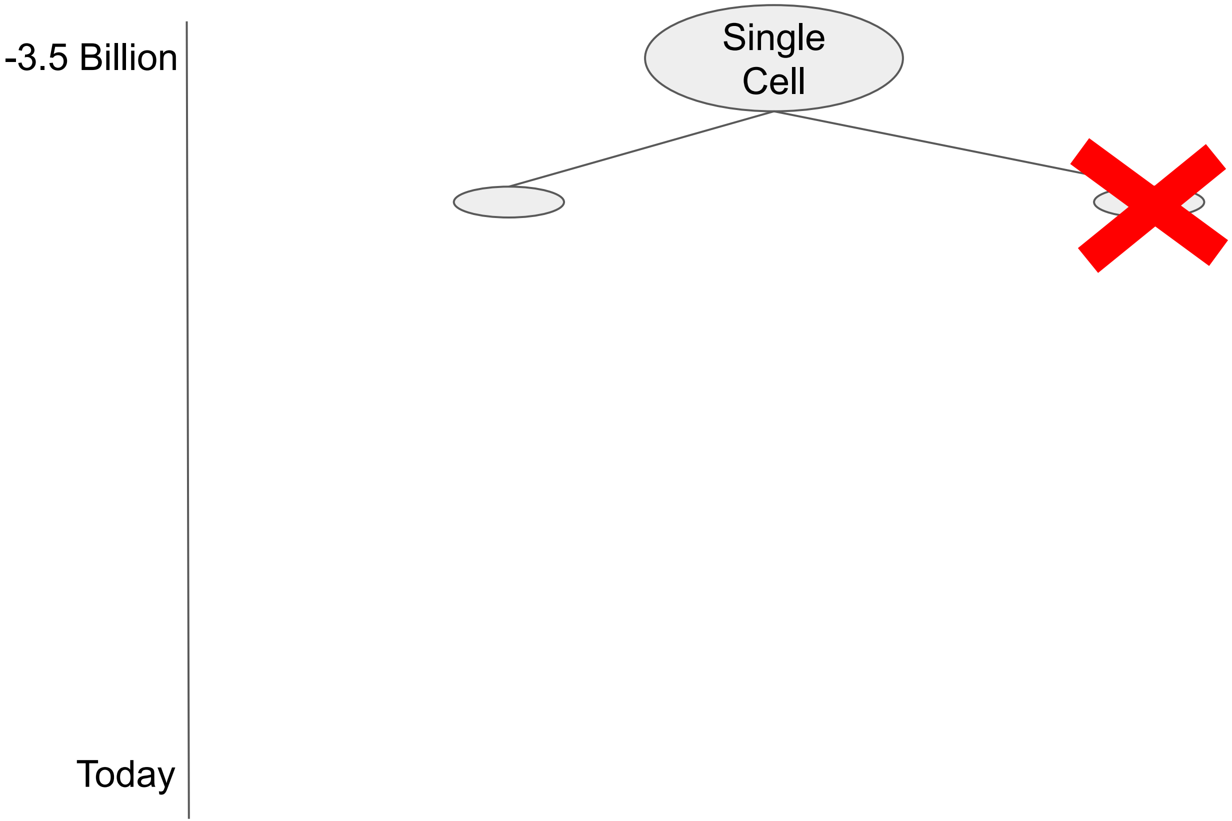
Tree of life



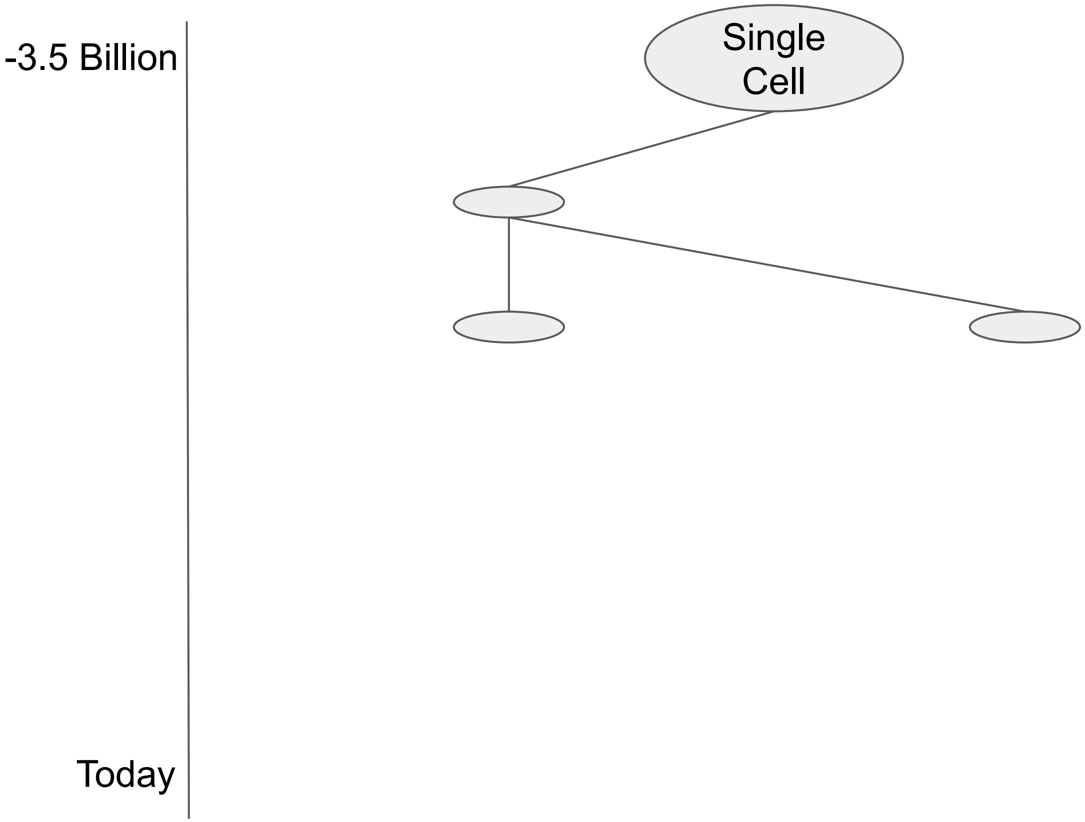
Tree of life



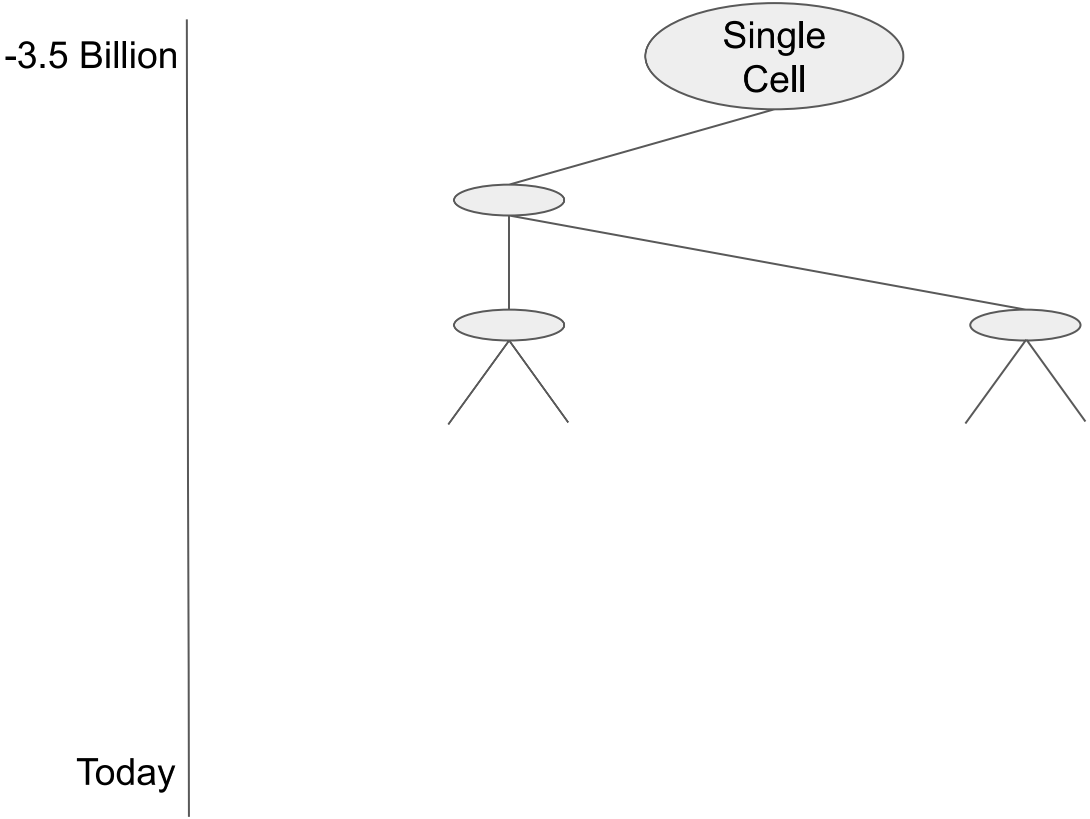
Tree of life



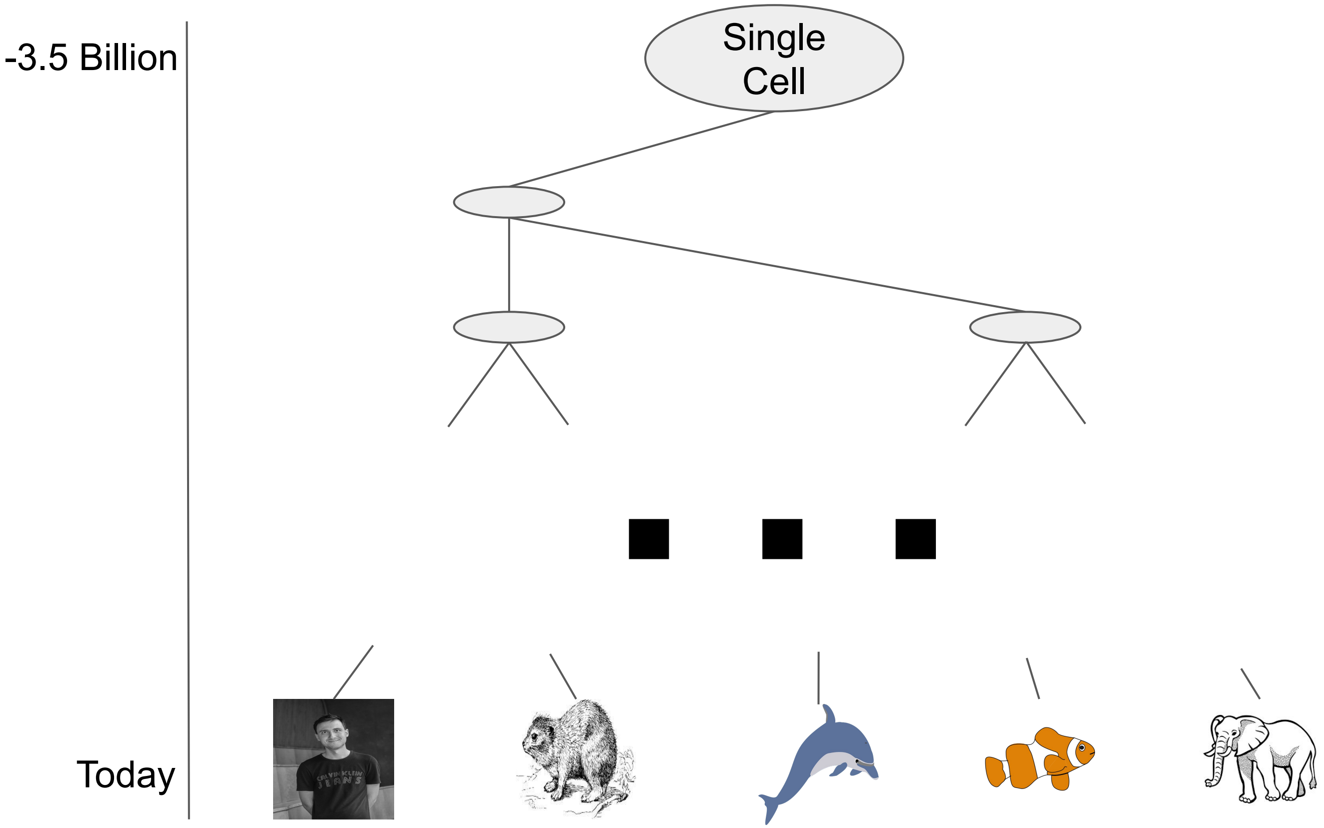
Tree of life



Tree of life

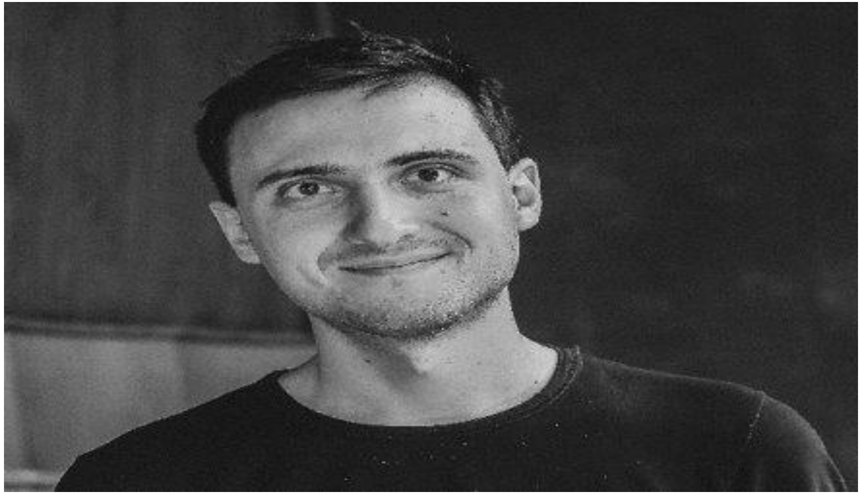


Tree of life

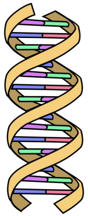


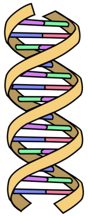
First to discuss tree reconstruction

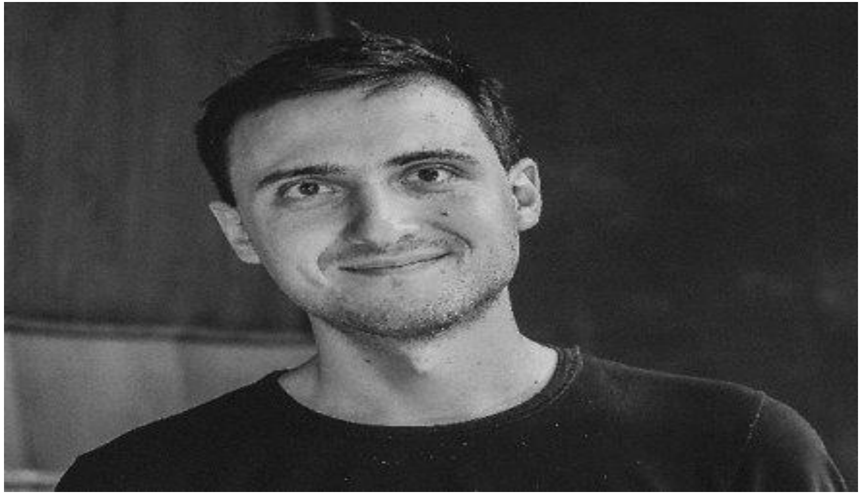
They're the same picture…



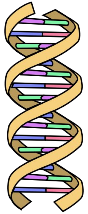
They're the same picture…



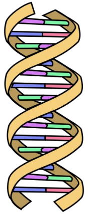


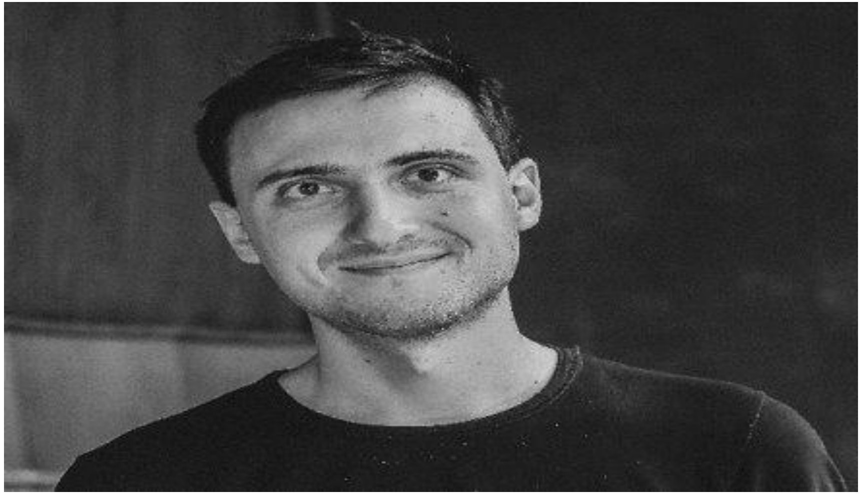


They're the same picture…



Hmm…
15 million years?





Reconstruct from distance matrix

				
	0	15	435	432
	15	0	436	433
	435	436	0	63
	432	433	63	0

Reconstruct from distance matrix

				
	0	15	435	432
	15	0	436	433
	435	436	0	63
	432	433	63	0

If, due to noise,
no matching tree?

Reconstruct from distance matrix

				
	0	15	435	432
	15	0	436	433
	435	436	0	63
	432	433	63	0

If, due to noise,
no matching tree?

Minimize
disagreements!

Reconstruct from distance matrix

				
	0	15	435	432
	15	0	436	433
	435	436	0	63
	432	433	63	0

If, due to noise,
no matching tree?

Minimize disagreements!

Can also minimize total error, max error, L2 error…

The New York Times

TRILOBITES

A Battle Is Raging in the Tree of Life

Which came first, the sponge or the comb jelly?

POPULAR MECHANICS HOME LATEST STORIES SCIENCE MILITARY POP MECH PRO SUBSCRIBE

Science > Animals

Scientists Have Found the First Branch on the Tree of Life

Something had to diverge from the trunk eventually.

BY JACKIE APPEL PUBLISHED: JUN 8, 2023

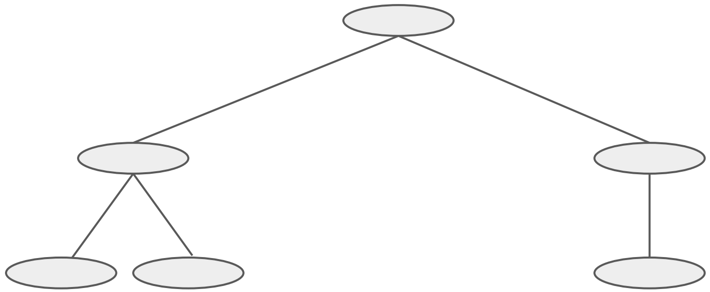
What we know…

APX - Hard

What we know…

APX - Hard

O(1) approximation for ultrametrics (structured trees) - even under mild constraints

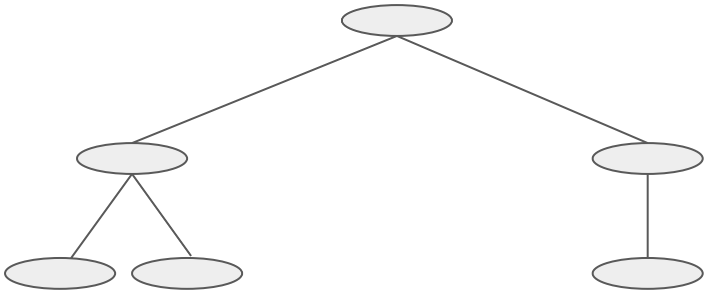


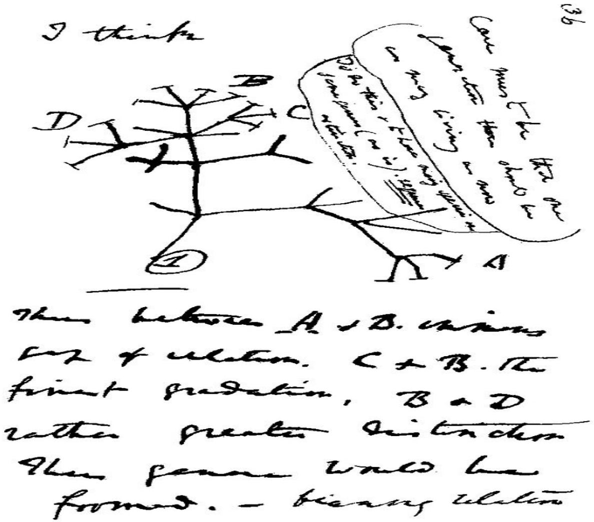
What we know…

APX - Hard

O(1) approximation for ultrametrics (structured trees) - even under mild constraints

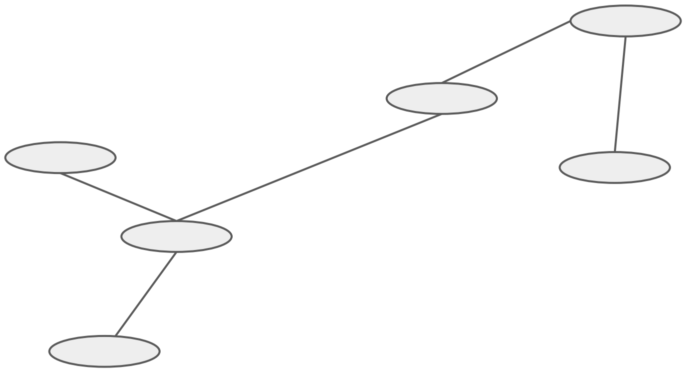
- What about unstructured trees?





Structuring the unstructured

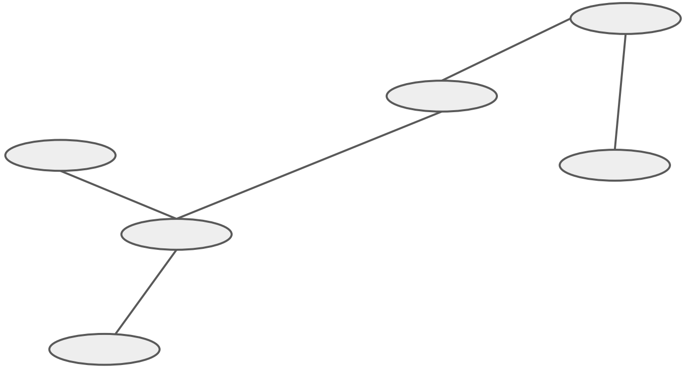
OPT



Structuring the unstructured

1) Find a root.
2) Find depths of leaves.

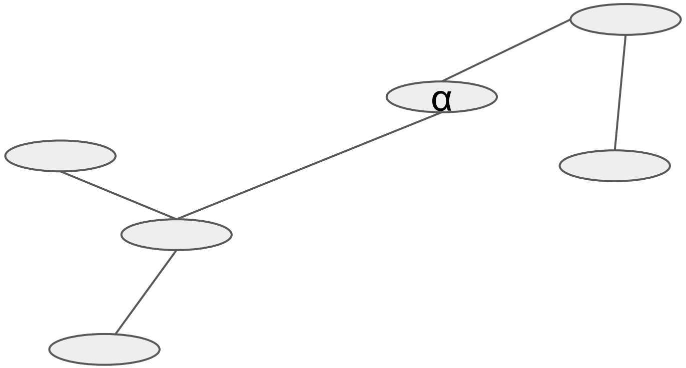
OPT



Structuring the unstructured

1) **Find a root.**
2) Find depths of leaves.

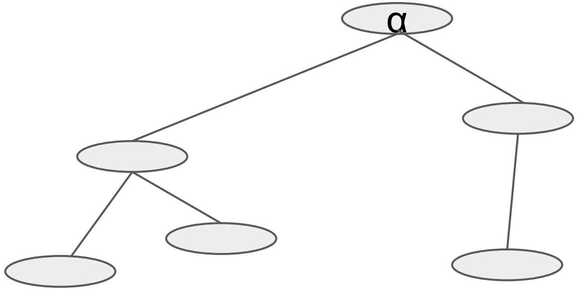
OPT



Structuring the unstructured

1) **Find a root.**
2) Find depths of leaves.

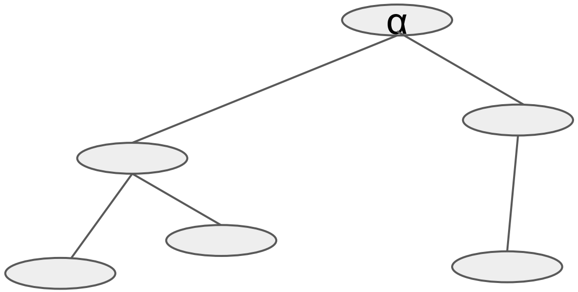
OPT



Structuring the unstructured

1) Find a root.
2) **Find depths of leaves.**

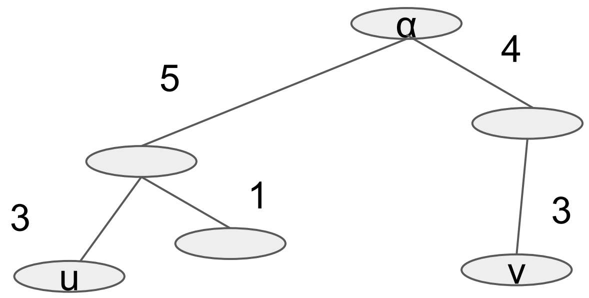
OPT



Structuring the unstructured

1) Find a root.
2) **Find depths of leaves.**

OPT

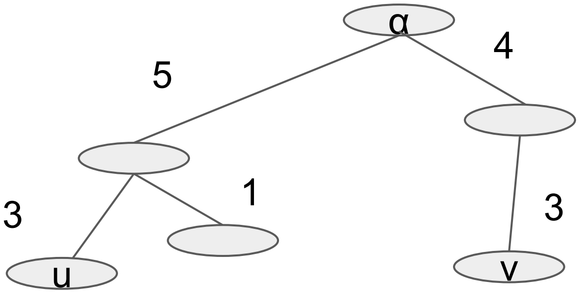


Structuring the unstructured

1) Find a root.
2) **Find depths of leaves.**

OPT

Input(α,u) = 12, but OPT(α,u) = 8

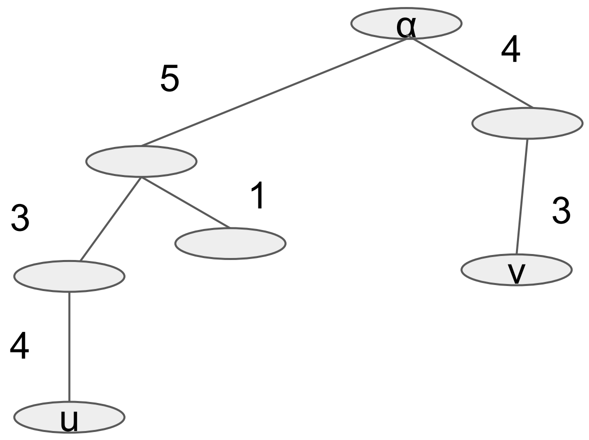


Structuring the unstructured

1) Find a root.
2) **Find depths of leaves.**

OPT’

Input(α,u) = 12, but OPT(α,u) = 8

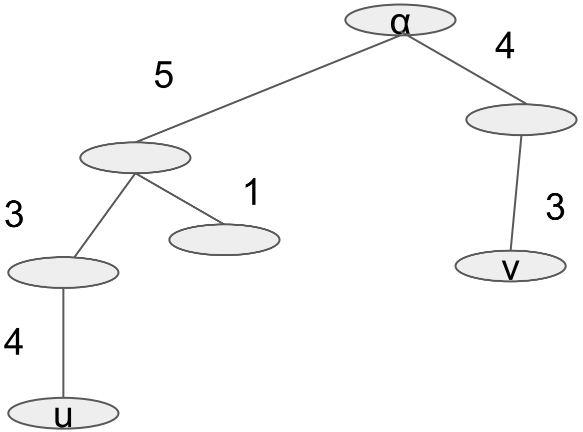


Structuring the unstructured

1) Find a root.
2) **Find depths of leaves.**

OPT'

Input(α,v) = 6, but OPT(α,v) = 7

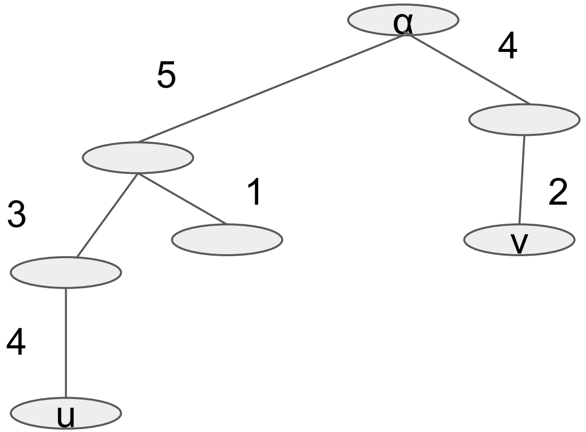


Structuring the unstructured

1) Find a root.
2) **Find depths of leaves.**

OPT'

Input(α,v) = 6, but OPT(α,v) = 7

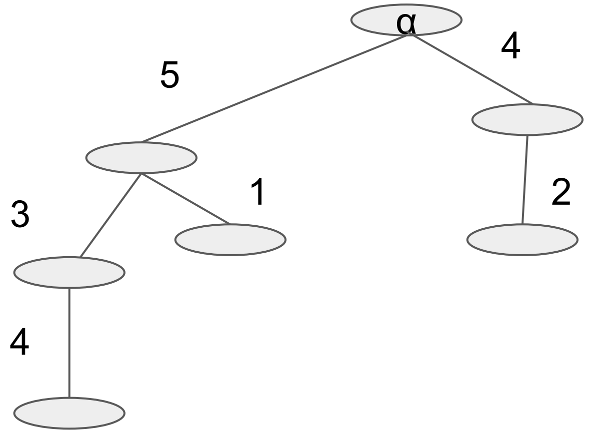


Structuring the unstructured

1) Find a root.
2) **Find depths of leaves.**

OPT'

Now for all u we know depth(u).
depth(u) = OPT'(α,u) = Input(α,u)

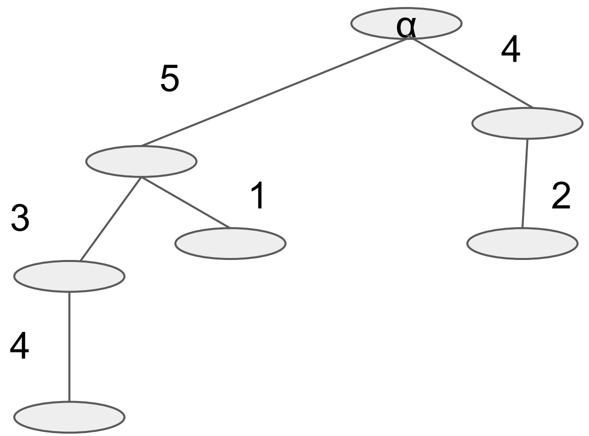


D() denotes disagreements in OPT

Structuring the unstructured

1) Find a root.
2) **Find depths of leaves.**

OPT'



Now for all u we know depth(u).
depth(u) = OPT(α,u) = Input(α,u)

How much did we pay?

- We moved exactly D(α) nodes, each introduced at most (n-1) disagreements.
- D(OPT') ≤ D(OPT) + D(α) (n-1)

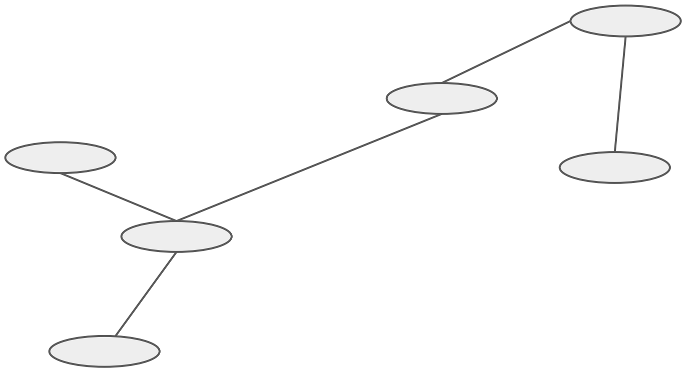
D() denotes disagreements in OPT
D(OPT') ≤ D(OPT) + D(α) (n-1)

Structuring the unstructured

1) **Find a root.**
2) Find depths of leaves.

$$D(OPT) = \frac{1}{2}\sum_u D(u)$$

OPT



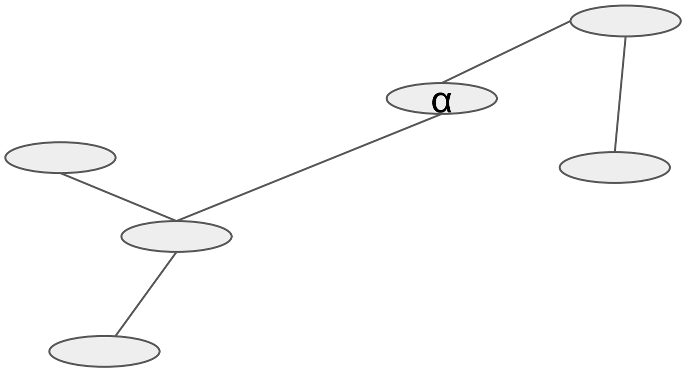
Structuring the unstructured

D() denotes disagreements in OPT
$D(OPT') \leq D(OPT) + D(\alpha)\ (n-1)$
α minimizes disagreements

1) **Find a root.**
2) Find depths of leaves.

$$D(OPT) = \frac{1}{2} \sum_u D(u)$$

OPT



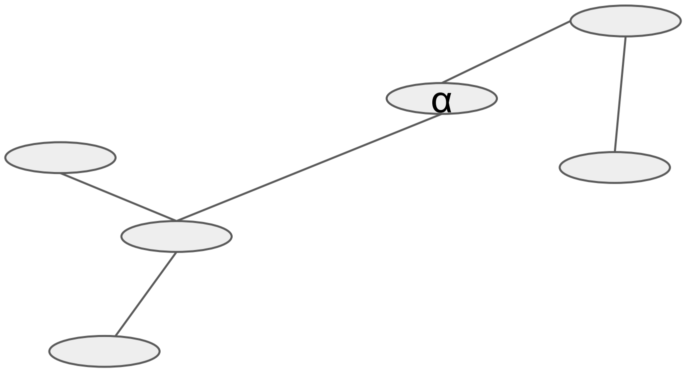
Structuring the unstructured

D() denotes disagreements in OPT
D(OPT') ≤ D(OPT) + D(α) (n-1)
α minimizes disagreements

1) **Find a root.**
2) Find depths of leaves.

$$D(OPT) = \frac{1}{2} \sum_u D(u)$$
$$\geq \frac{1}{2} \cdot n \cdot D(\alpha)$$

OPT



Structuring the unstructured

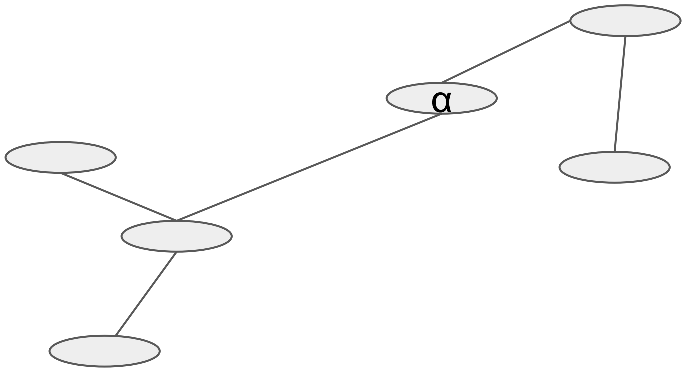
D() denotes disagreements in OPT
D(OPT') ≤ D(OPT) + D(α) (n-1)
α minimizes disagreements

1) **Find a root.**
2) Find depths of leaves.

$$D(OPT) = \frac{1}{2}\sum_u D(u)$$
$$\geq \frac{1}{2} \cdot n \cdot D(\alpha)$$

$$D(\alpha) \leq 2D(OPT)/n$$

OPT



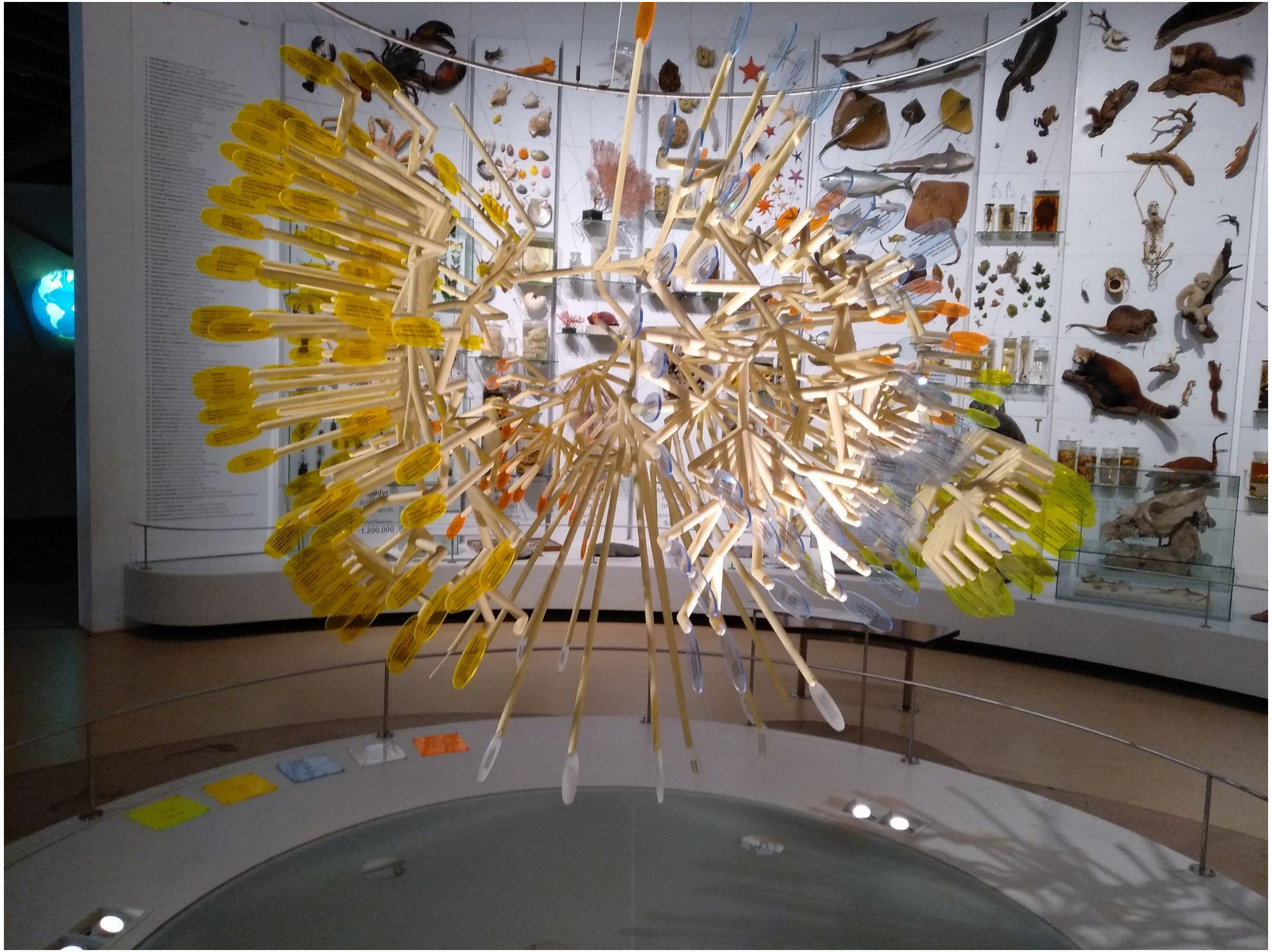
Structuring the unstructured

D(OPT') ≤ D(OPT) + D(α) (n-1)

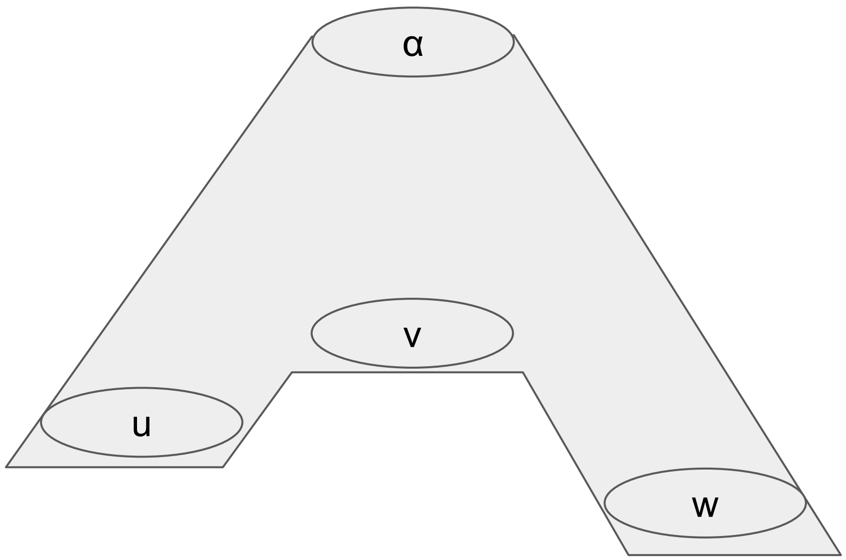
1) **Find a root.**
2) Find depths of leaves.

D(OPT') ≤ 3D(OPT)

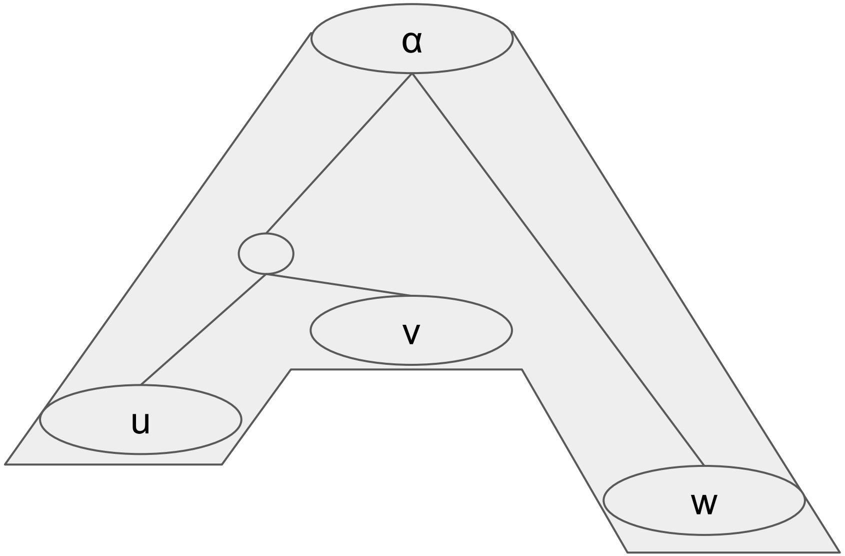
$$D(\alpha) \leq 2D(OPT)/n$$



Where do we stand?

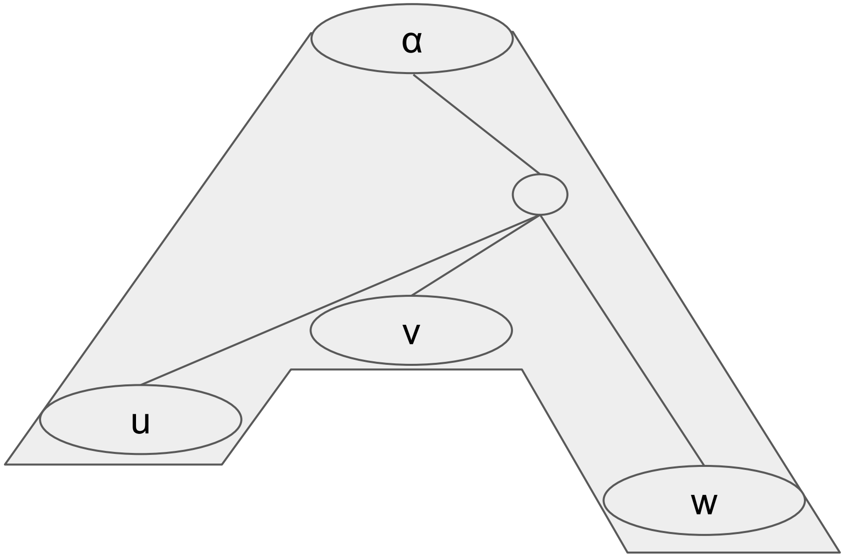


Where do we stand?



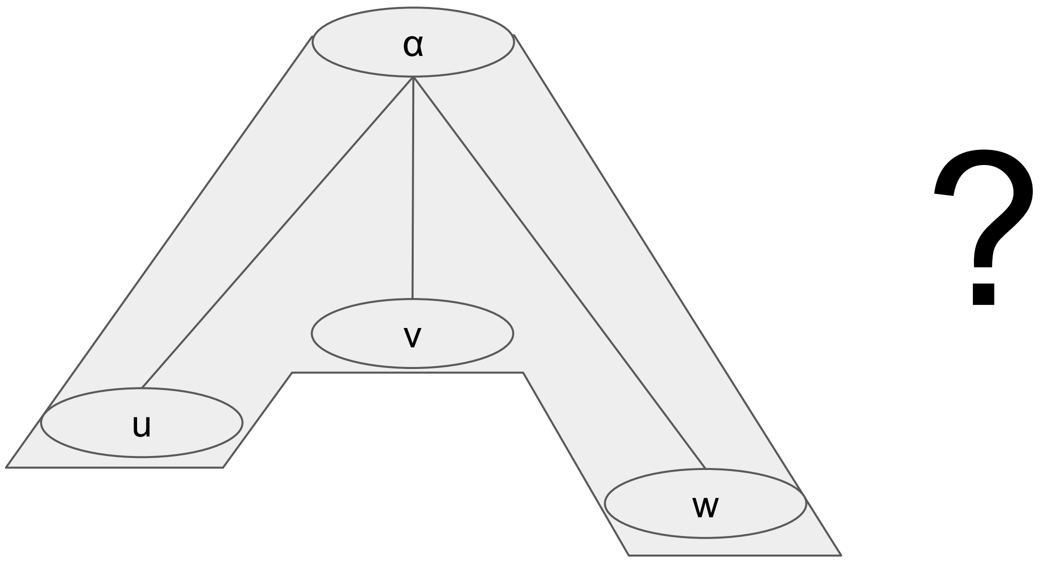
?

Where do we stand?

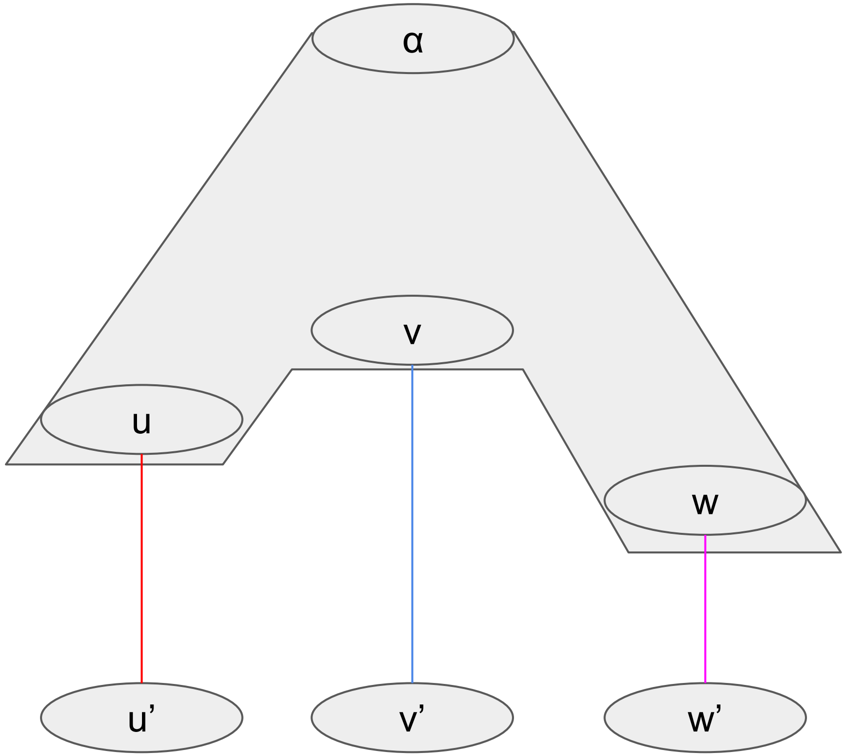


?

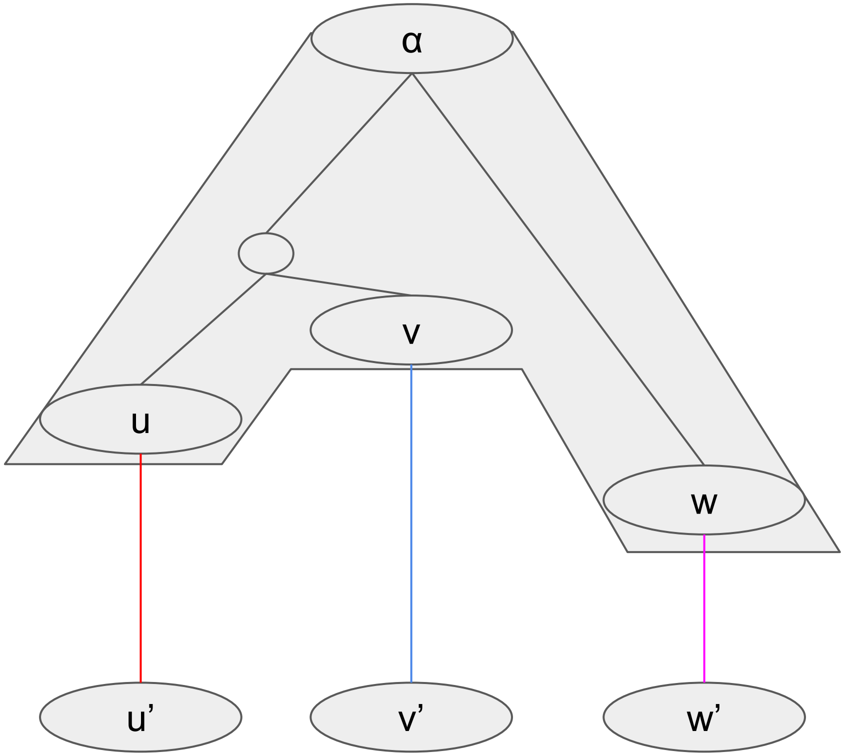
Where do we stand?



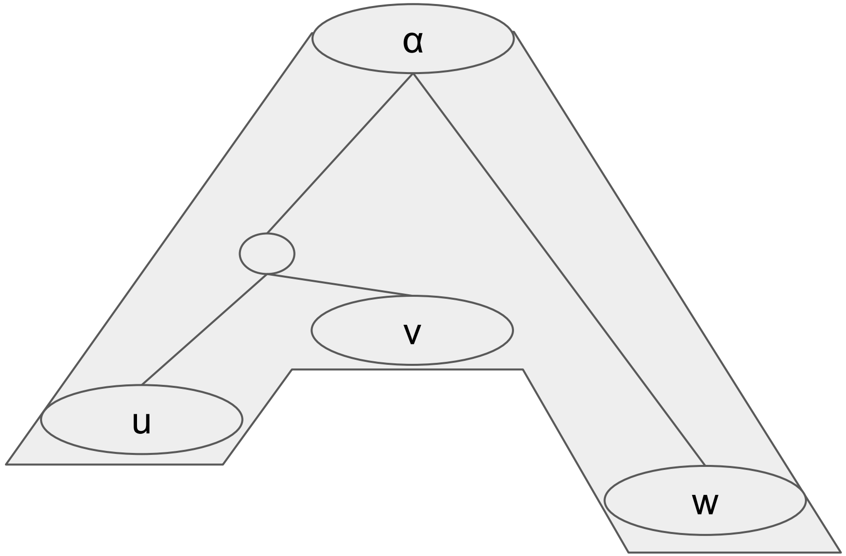
Reduce to ultrametric (all leaves same depth)



Reduce to ultrametric (all leaves same depth)



Reduce to ultrametric (all leaves same depth)



So… who is the “Oldest Sister”?

Was it the sponge or the comb jelly that diverged first?

So… who is the “Oldest Sister”?

Was it the sponge or the comb jelly that diverged first?