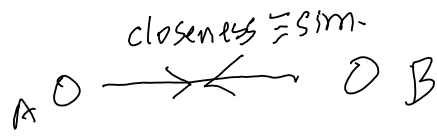


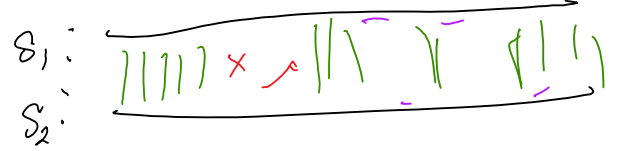
Distance Measures

Friday, February 19, 2021 10:56 AM

Distance vs. Similarity:

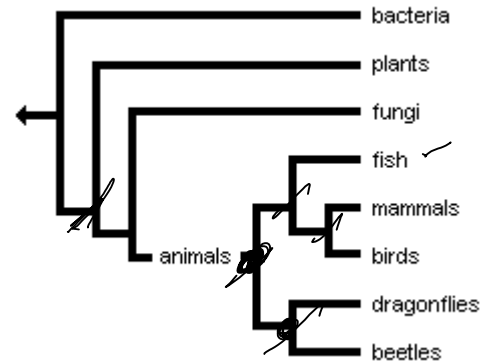


Alignment:

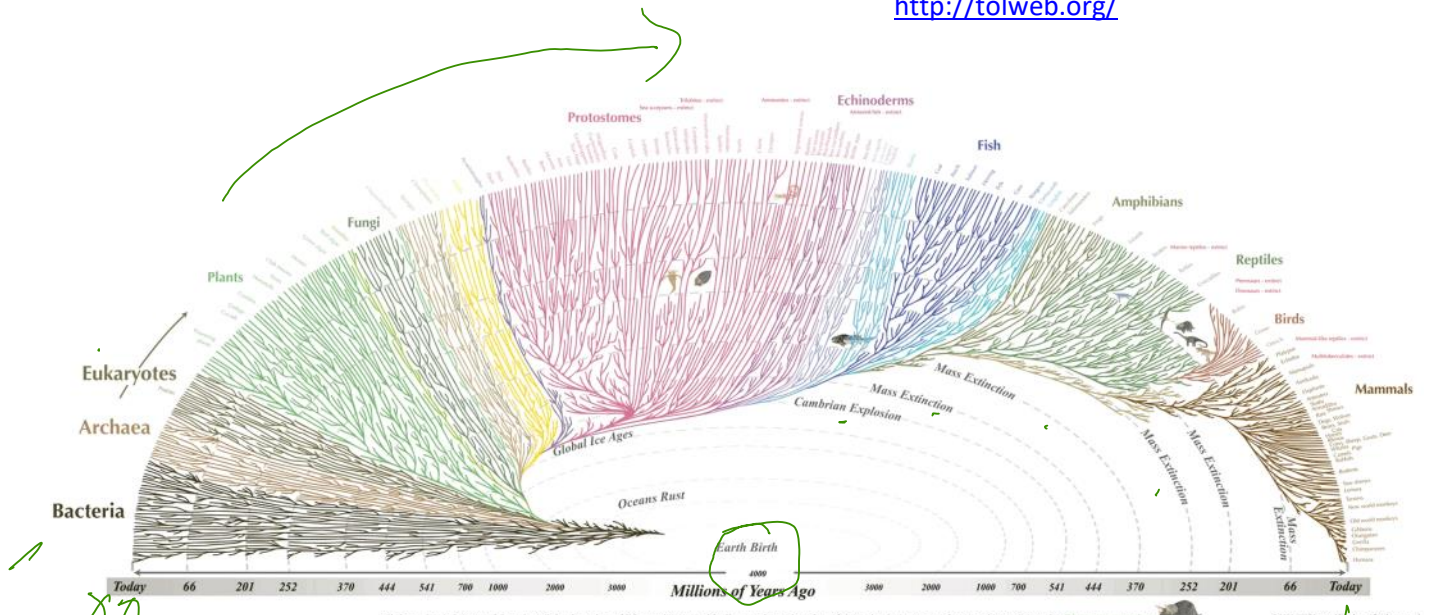


score \approx sim.
dist. \approx distance
distance increases

phylogenetic trees



<http://tolweb.org/>



<https://flowingdata.com/2018/10/25/tree-of-life/>

Distance measures

Friday, February 19, 2021 11:00 AM

Distance

↓
1) **Hamming Distance.**

s_1 & s_2 } $\Rightarrow H(s_1, s_2) = \# \text{ diff positions}$
 $|s_1| = |s_2|$ } (only allow substitution)

s_1 : a c c g t a t
 s_2 : a c g t a t g

0 0 1 1 1 1 1 = 5 ←

2) **Edit Distance**

"Edit" $\left\{ \begin{array}{l} \text{substitution} \\ \text{gap (deletion)} \\ \text{insert a gap} \end{array} \right.$

$E(s_1, s_2) = \# \text{ minimum edits to } s_1 \text{ to convert it to } s_2$

→ s_1 : a c c g t a t
 ✓ ✓ | ✓ ✓ ✓ ✓ |

s_1 | | |
 ↓ x - t
 s_2 | | |

s_2 : a c g t a t g

$E(s_1, s_2) = 2$ ←

$E(s_1, s_2) \leq H(s_1, s_2)$

smaller $E(s_1, s_2)$
 no more similar.

Edit distance algorithm

Friday, February 19, 2021 11:00 AM

Algorithm:

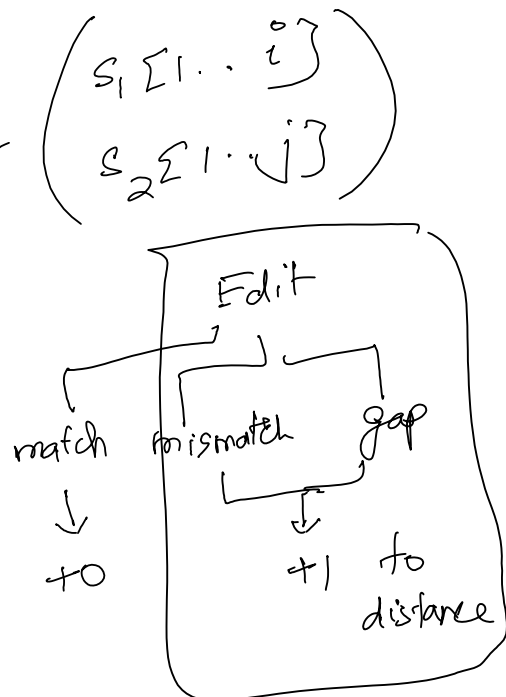
Let $E(i, j)$ ← Edit distance for

for $(s_1 \Sigma 1 \dots i)$
 $(s_2 \Sigma 1 \dots j)$

Output: $E(m, n) \equiv \text{Edit}(s_1, s_2)$

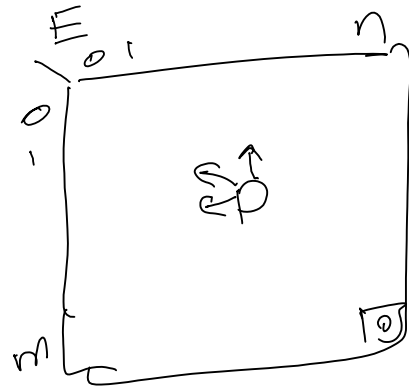
Base:
 $E(0, 0) = 0$
 $E(i, 0) = i$
 $E(0, j) = j$

for all $1 \leq i \leq m$
for all $1 \leq j \leq n$



$$E(i, j) = \min \left\{ \begin{array}{l} E(i-1, j-1) + \begin{cases} 0, & \text{if } a_i = b_j \\ 1, & \text{o/w} \end{cases} \\ E(i-1, j) + 1 \\ E(i, j-1) + 1 \end{array} \right.$$

$O(mn)$ time



Edit distance properties

Friday, February 19, 2021 11:41 AM

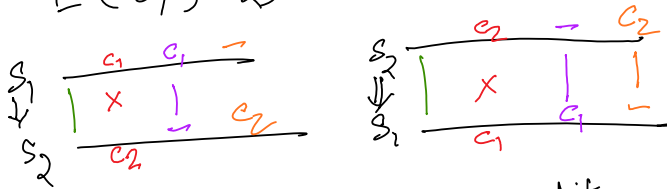
Properties of Edit Distance:

1) Reflexive!

$$E(S, S) = 0$$

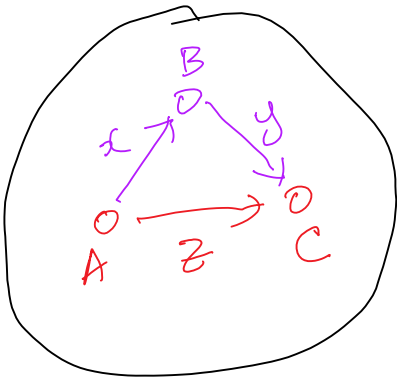
2) Symmetry:

$$E(S_1, S_2) = E(S_2, S_1)$$



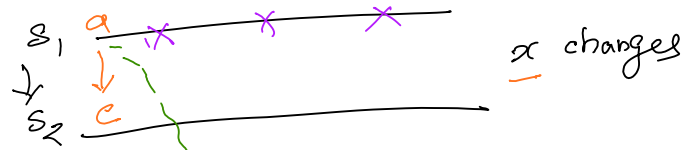
3) Transitivity / Triangular inequality:

$$E(S_1, S_2) + E(S_2, S_3) \geq E(S_1, S_3)$$

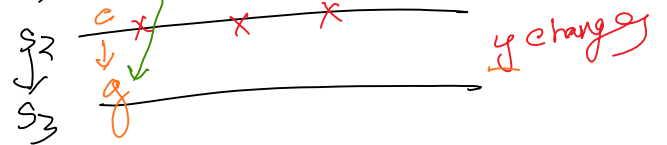


$$z \leq x + y$$

$E(S_1, S_2)$:



$E(S_2, S_3)$:



$$E(S_1, S_3) \leq E(S_1, S_2) + E(S_2, S_3)$$

Distance Measure

- ✓ Reflexive
 - ✓ Symmetry
 - ✓ Δ inequality
- \Rightarrow Metric

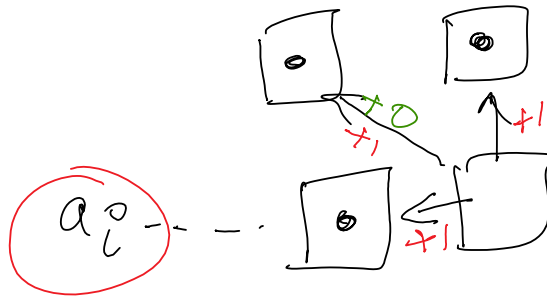
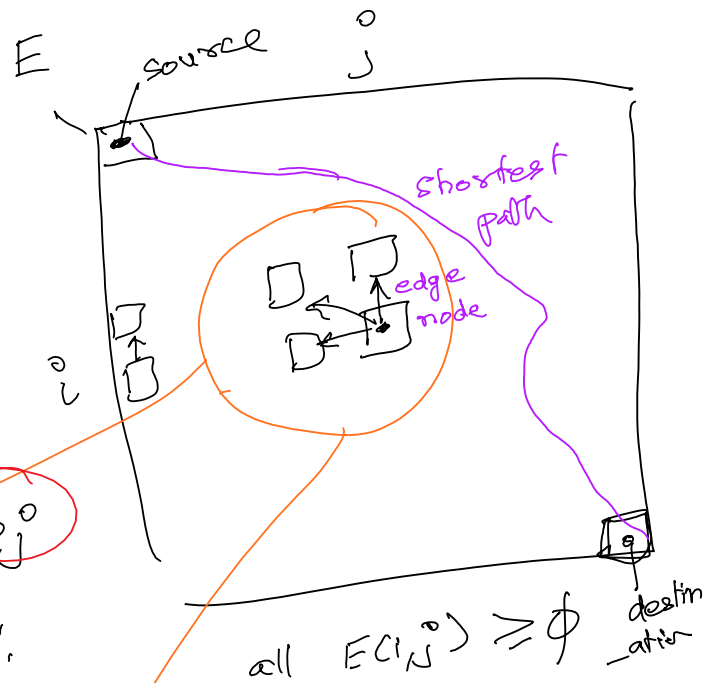
Graph-theoretic view

Friday, February 19, 2021 11:51 AM

$$\#nodes = (m+1)(n+1)$$

$$\#edges \leq 3 \times (m+1)(n+1)$$

Graph



$$|E| \lg(|V|)$$

$$mn \lg(mn)$$

m

