Graph wavefront algorithm

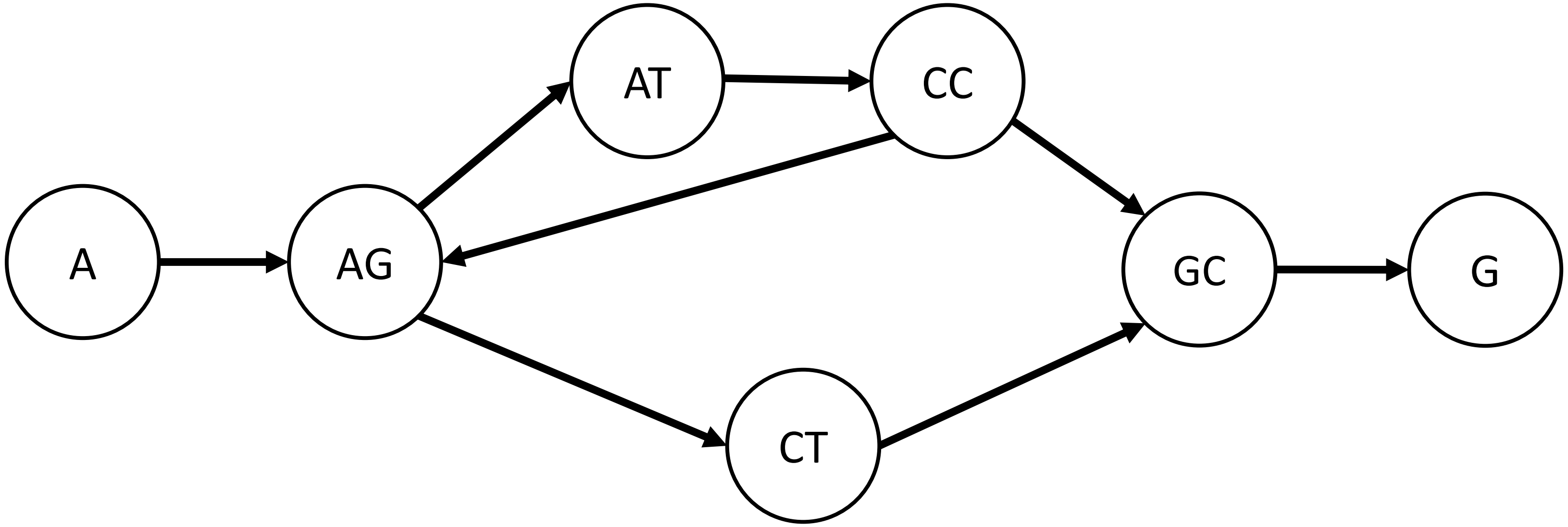
Haowen Zhang

Georgia Institute of Technology

July 7, 2022

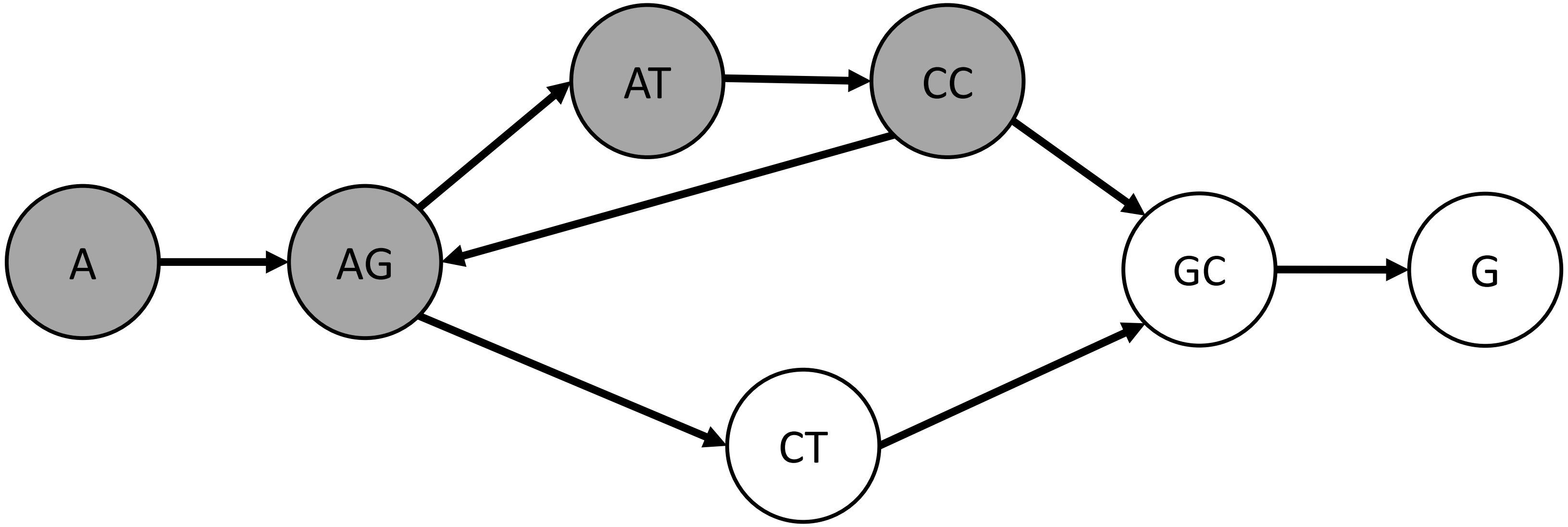
Sequence graph

- A sequence graph $G(V, E, \sigma)$ is a directed labeled graph
- Function $\sigma: V \rightarrow \mathbf{\Sigma}^{+}$ labels each vertex with a string over the alphabet $\mathbf{\Sigma}$={A, C, G, T}



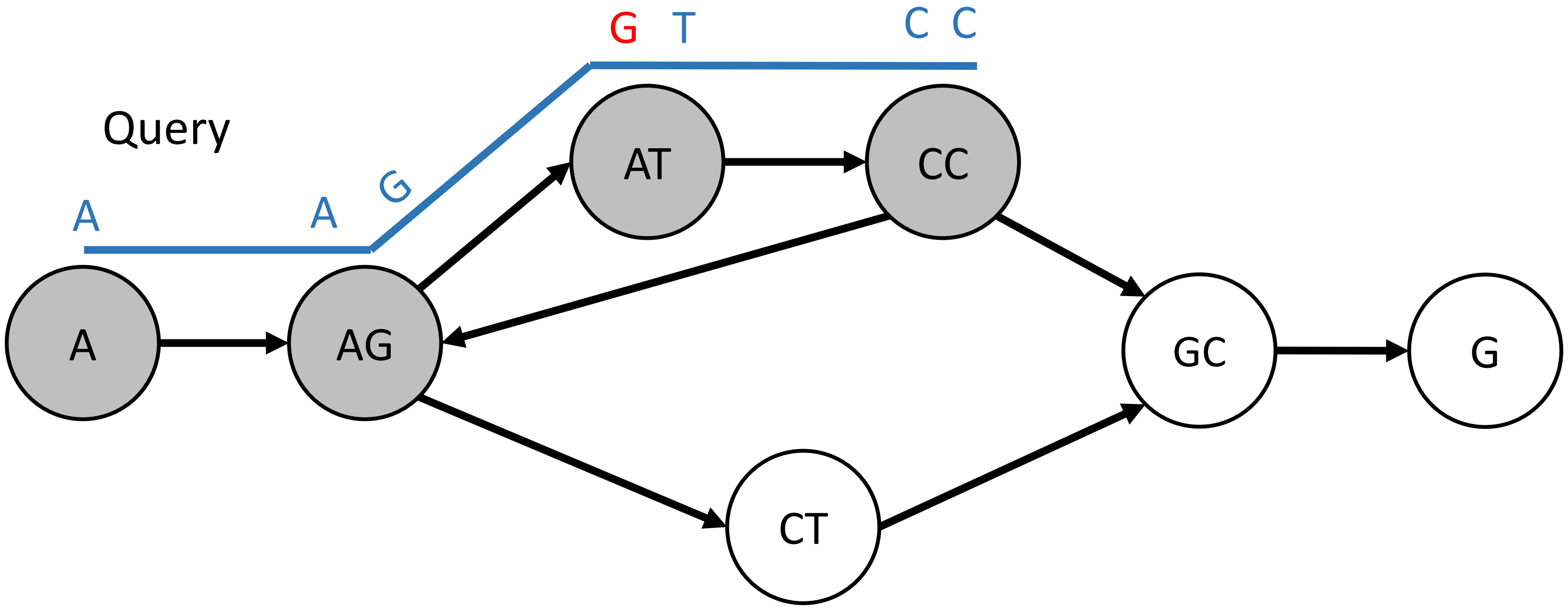
Sequence graph

- A sequence graph $G(V, E, \sigma)$ is a directed labeled graph
- Function $\sigma: V \rightarrow \mathbf{\Sigma}^+$ labels each vertex with a string over the alphabet $\mathbf{\Sigma}$={A, C, G, T}



Sequence to graph alignment

- Given a query and a sequence graph, find a walk in the graph so that the edit distance between the query and the walk is minimized.



Edit distance between two sequences

- DP recurrence to compute edit distance between two strings

$$H_{i,j} = \min \begin{cases} H_{i-1,j} + 1 \\ H_{i,j-1} + 1 \\ H_{i-1,j-1} + \Delta_{i,j} \end{cases} \tag{1}$$

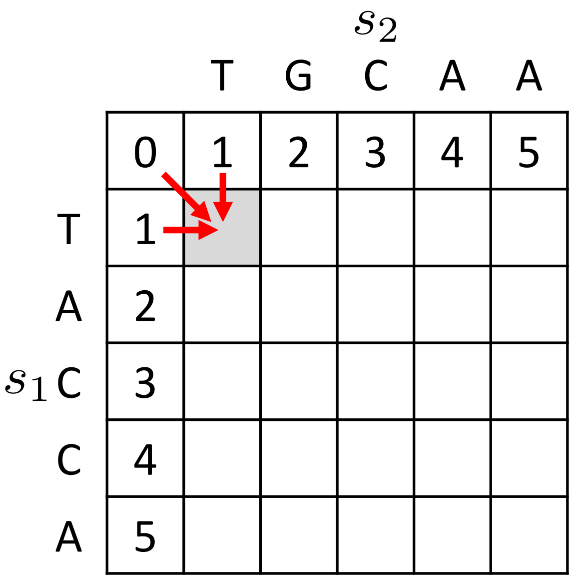
s_2

		T	G	C	A	A
	0	1	2	3	4	5
T	1					
A	2					
s_1 C	3					
C	4					
A	5					

Edit distance between two sequences

- DP recurrence to compute edit distance between two strings

$$H_{i,j} = \min \begin{cases} H_{i-1,j} + 1 \\ H_{i,j-1} + 1 \\ H_{i-1,j-1} + \Delta_{i,j} \end{cases} \qquad (1)$$

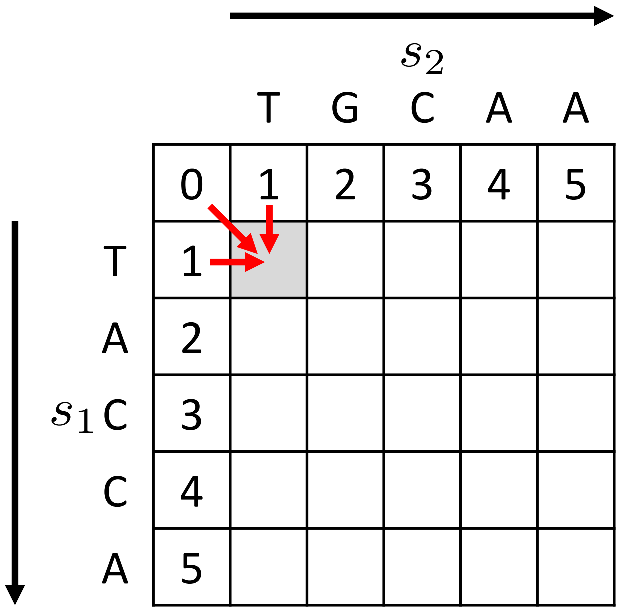


		T	G	C	A	A
	0	1	2	3	4	5
T	1					
A	2					
C	3					
C	4					
A	5					

Edit distance between two sequences

- DP recurrence to compute edit distance between two strings

$$H_{i,j} = \min \begin{cases} H_{i-1,j} + 1 \\ H_{i,j-1} + 1 \\ H_{i-1,j-1} + \Delta_{i,j} \end{cases} \tag{1}$$

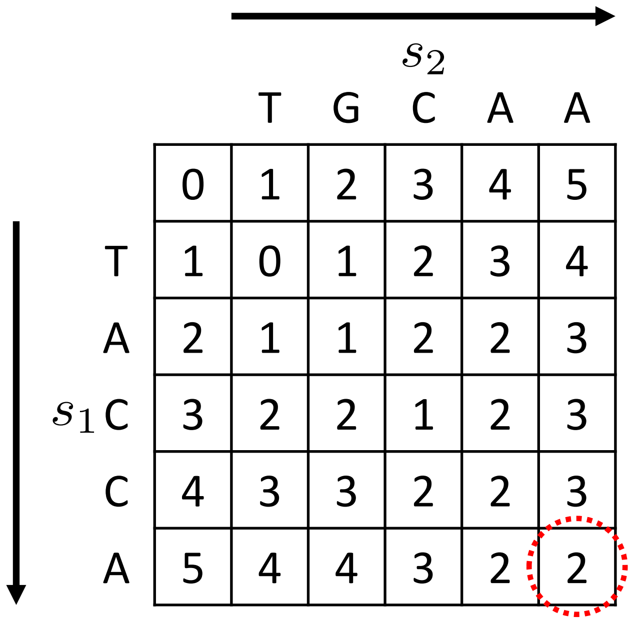


		T	G	C	A	A
	0	1	2	3	4	5
T	1					
A	2					
C	3					
C	4					
A	5					

Edit distance between two sequences

- DP recurrence to compute edit distance between two strings

$$H_{i,j} = \min \begin{cases} H_{i-1,j} + 1 \\ H_{i,j-1} + 1 \\ H_{i-1,j-1} + \Delta_{i,j} \end{cases} \tag{1}$$

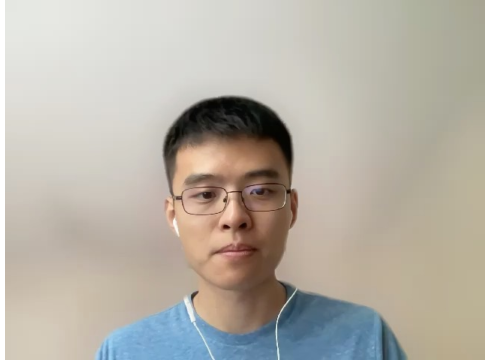


s_2

		T	G	C	A	A
	0	1	2	3	4	5
T	1	0	1	2	3	4
A	2	1	1	2	2	3
s_1 C	3	2	2	1	2	3
C	4	3	3	2	2	3
A	5	4	4	3	2	2

Time: $O(N^2)$

Edit distance between two sequences



- DP recurrence to compute edit distance between two strings

$$H_{i,j} = \min \begin{cases} H_{i-1,j} + 1 \\ H_{i,j-1} + 1 \\ H_{i-1,j-1} + \Delta_{i,j} \end{cases} \tag{1}$$

s_2

s_1

		T	G	C	A	A
	0	1	2	3	4	5
T	1	0	1	2	3	4
A	2	1	1	2	2	3
C	3	2	2	1	2	3
C	4	3	3	2	2	3
A	5	4	4	3	2	2

Time: $O(N^2)$

Can we do better in practice?

Edit distance between two sequences

- DP recurrence to compute edit distance between two strings

$$H_{i,j} = \min \begin{cases} H_{i-1,j} + 1 \\ H_{i,j-1} + 1 \\ H_{i-1,j-1} + \Delta_{i,j} \end{cases} \tag{1}$$

s_2

s_1

		T	G	C	A	A
	0	1	2	3	4	5
T	1	0	1	2	3	4
A	2	1	1	2	2	3
C	3	2	2	1	2	3
C	4	3	3	2	2	3
A	5	4	4	3	2	2

Time: $O(N^2)$

Can we do better in practice?
Yes. Only compute DP cells with values ≤2.

Edit distance between two sequences

- DP recurrence to compute edit distance between two strings

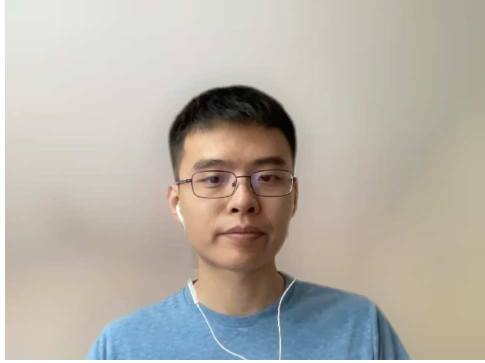
$$H_{i,j} = \min \begin{cases} H_{i-1,j} + 1 \\ H_{i,j-1} + 1 \\ H_{i-1,j-1} + \Delta_{i,j} \end{cases} \tag{1}$$

s_2

		T	G	C	A	A
	0	1	2			
T	1	0	1	2		
A	2	1	1	2	2	
s_1 C		2	2	1	2	
C				2	2	
A					2	2

Time: ~~$O(N^2)$~~ $O(DN)$

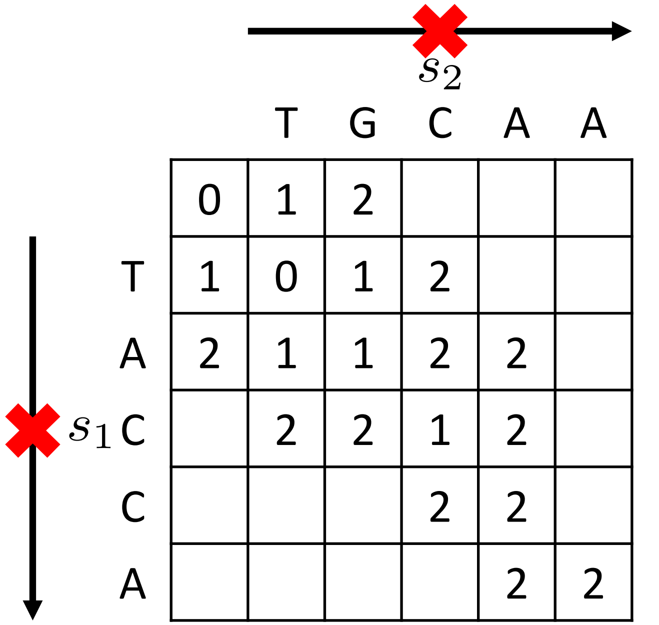
Can we do better in practice?
Yes. Only compute DP cells with values ≤2.



Edit distance between two sequences

- DP recurrence to compute edit distance between two strings

$$H_{i,j} = \min \begin{cases} H_{i-1,j} + 1 \\ H_{i,j-1} + 1 \\ H_{i-1,j-1} + \Delta_{i,j} \end{cases} \tag{1}$$



		T	G	C	A	A
	0	1	2			
T	1	0	1	2		
A	2	1	1	2	2	
C		2	2	1	2	
C				2	2	
A					2	2

Time: $\cancel{O(N^2)}\ O(DN)$

Can we do better in practice?
Yes. Only compute DP cells with values ≤2.

Edit distance between two sequences

- DP recurrence to compute edit distance between two strings

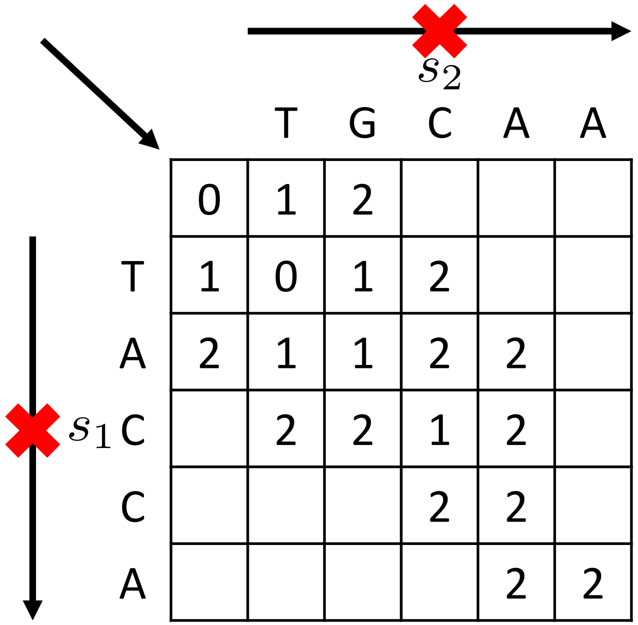
$$H_{i,j} = \min \begin{cases} H_{i-1,j} + 1 \\ H_{i,j-1} + 1 \\ H_{i-1,j-1} + \Delta_{i,j} \end{cases} \qquad (1)$$

$$\tilde{J}_{d,k} = \max \begin{cases} \tilde{H}_{d-1,k-1} \\ \tilde{H}_{d-1,k+1} + 1 \\ \tilde{H}_{d-1,k} + 1 \end{cases} \qquad (2)$$

$$\tilde{H}_{d,k} = j + LCP\left(s_1[i+1, |s_1|], s_2[j+1, |s_2|]\right),$$

$$j = \tilde{J}_{d,k},\ i = k + j$$

Observation: DP cells with d are always adjacent to DP cells with d-1 or d



		T	G	C	A	A
	0	1	2			
T	1	0	1	2		
A	2	1	1	2	2	
C		2	2	1	2	
C				2	2	
A					2	2

Time: $\cancel{O(N^2)}\ O(DN)$

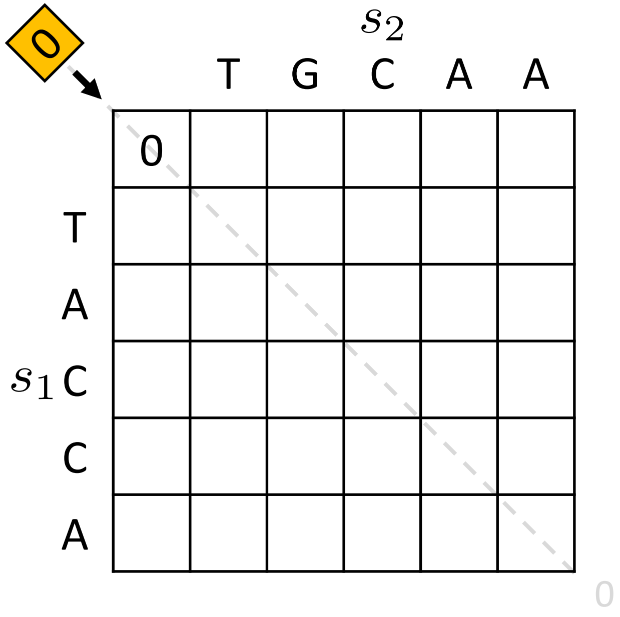
Can we do better in practice? Yes. Only compute DP cells with values ≤2.

Edit distance between two sequences

- DP recurrence to compute edit distance between two strings

$$H_{i,j} = \min \begin{cases} H_{i-1,j} + 1 \\ H_{i,j-1} + 1 \\ H_{i-1,j-1} + \Delta_{i,j} \end{cases} \tag{1}$$

$$\begin{aligned} \tilde{J}_{d,k} =& \max \begin{cases} \tilde{H}_{d-1,k-1} \\ \tilde{H}_{d-1,k+1} + 1 \\ \tilde{H}_{d-1,k} + 1 \end{cases} \\ \tilde{H}_{d,k} =& j + LCP\left(s_1[i+1, |s_1|], s_2[j+1, |s_2|]\right), \\ & j = \tilde{J}_{d,k},\ i = k + j \end{aligned} \tag{2}$$



d=0

Initial state

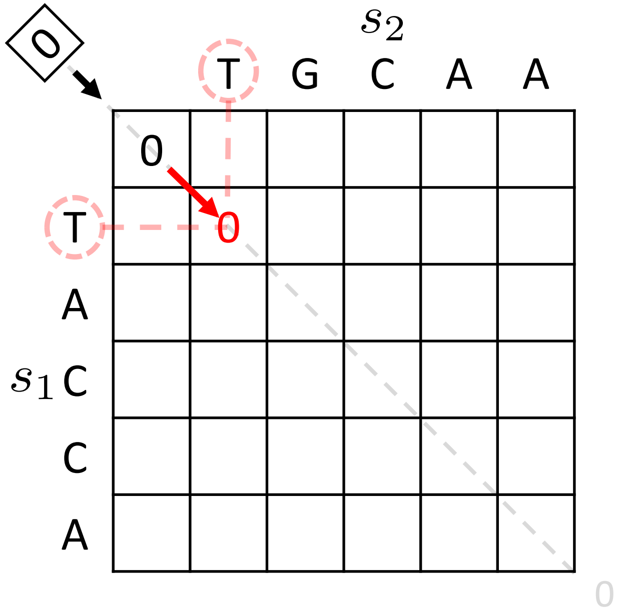
Edit distance between two sequences

- DP recurrence to compute edit distance between two strings

$$H_{i,j} = \min \begin{cases} H_{i-1,j} + 1 \\ H_{i,j-1} + 1 \\ H_{i-1,j-1} + \Delta_{i,j} \end{cases} \qquad (1)$$

$$\tilde{J}_{d,k} = \max \begin{cases} \tilde{H}_{d-1,k-1} \\ \tilde{H}_{d-1,k+1} + 1 \\ \tilde{H}_{d-1,k} + 1 \end{cases} \qquad (2)$$

$$\begin{aligned} \tilde{H}_{d,k} =& j + LCP\left(s_1[i+1, |s_1|], s_2[j+1, |s_2|]\right), \\ & j = \tilde{J}_{d,k},\ i = k + j \end{aligned}$$



d=0

Extension step

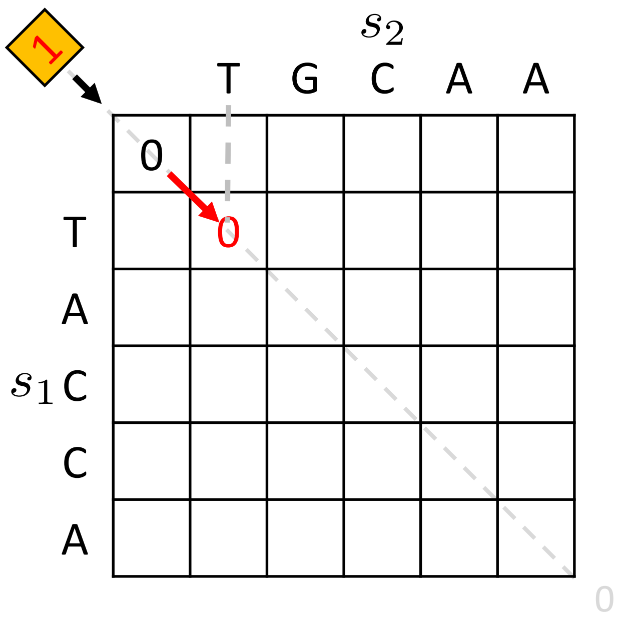
Edit distance between two sequences

- DP recurrence to compute edit distance between two strings

$$H_{i,j} = \min \begin{cases} H_{i-1,j} + 1 \\ H_{i,j-1} + 1 \\ H_{i-1,j-1} + \Delta_{i,j} \end{cases} \tag{1}$$

$$\begin{aligned} \tilde{J}_{d,k} &= \max \begin{cases} \tilde{H}_{d-1,k-1} \\ \tilde{H}_{d-1,k+1} + 1 \\ \tilde{H}_{d-1,k} + 1 \end{cases} \\ \tilde{H}_{d,k} &= j + LCP\left(s_1[i+1,|s_1|], s_2[j+1,|s_2|]\right), \\ & j = \tilde{J}_{d,k},\ i = k + j \end{aligned} \tag{2}$$

Offset on diagonal <=> cell column index



d=0

Extension step

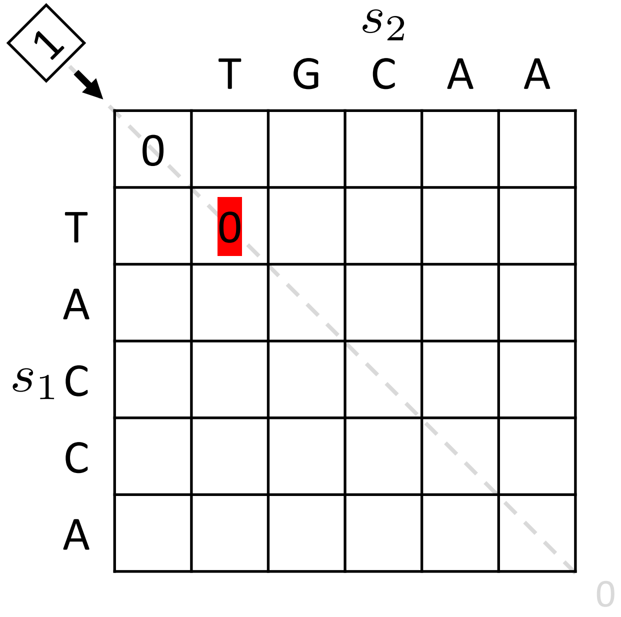
Edit distance between two sequences

- DP recurrence to compute edit distance between two strings

$$H_{i,j} = \min \begin{cases} H_{i-1,j} + 1 \\ H_{i,j-1} + 1 \\ H_{i-1,j-1} + \Delta_{i,j} \end{cases} \tag{1}$$

$$\begin{aligned} \tilde{J}_{d,k} =& \max \begin{cases} \tilde{H}_{d-1,k-1} \\ \tilde{H}_{d-1,k+1} + 1 \\ \tilde{H}_{d-1,k} + 1 \end{cases} \\ \tilde{H}_{d,k} =& j + LCP\left(s_1[i+1, |s_1|], s_2[j+1, |s_2|]\right), \\ & j = \tilde{J}_{d,k},\ i = k + j \end{aligned} \tag{2}$$

Offset on diagonal <=> cell column index



d=0

Edit distance between two sequences

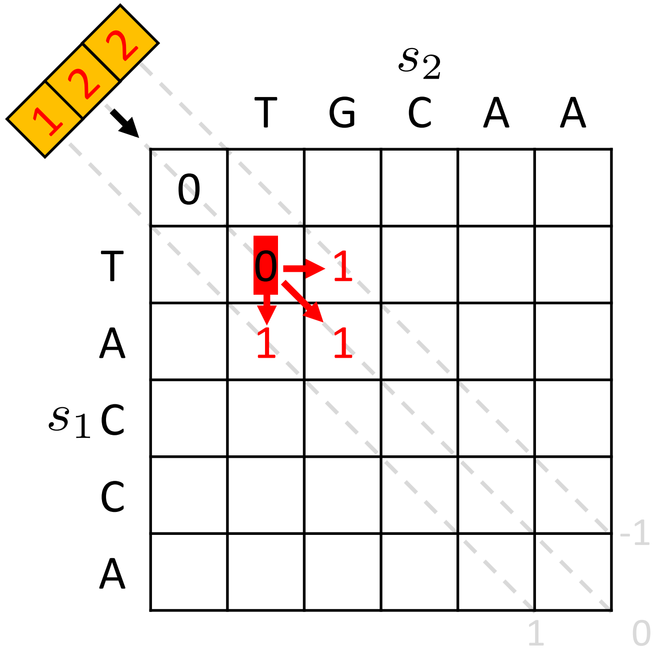
- DP recurrence to compute edit distance between two strings

$$H_{i,j} = \min \begin{cases} H_{i-1,j} + 1 \\ H_{i,j-1} + 1 \\ H_{i-1,j-1} + \Delta_{i,j} \end{cases} \tag{1}$$

$$\tilde{J}_{d,k} = \max \begin{cases} \tilde{H}_{d-1,k-1} \\ \tilde{H}_{d-1,k+1} + 1 \\ \tilde{H}_{d-1,k} + 1 \end{cases} \tag{2}$$

$$\tilde{H}_{d,k} = j + LCP\left(s_1[i+1,|s_1|], s_2[j+1,|s_2|]\right),$$

$$j = \tilde{J}_{d,k},\ i = k + j$$



d=1

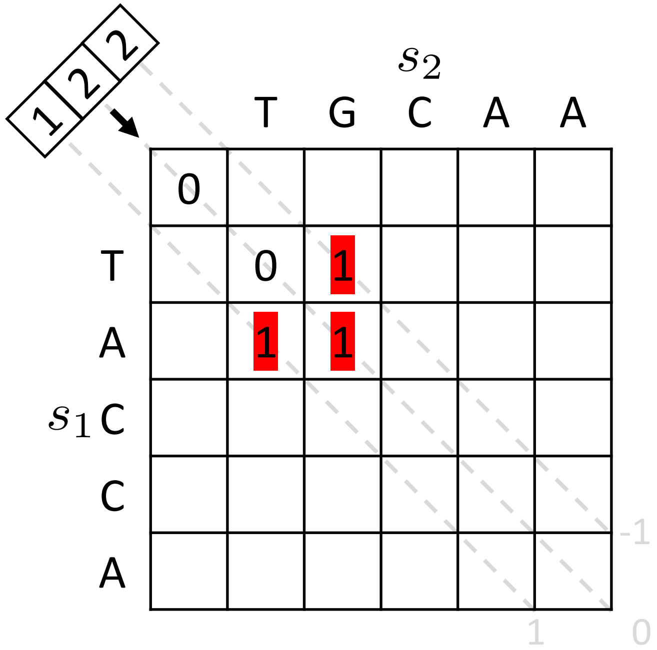
Expansion step

Edit distance between two sequences

- DP recurrence to compute edit distance between two strings

$$H_{i,j} = \min \begin{cases} H_{i-1,j} + 1 \\ H_{i,j-1} + 1 \\ H_{i-1,j-1} + \Delta_{i,j} \end{cases} \tag{1}$$

$$\begin{aligned} \tilde{J}_{d,k} &= \max \begin{cases} \tilde{H}_{d-1,k-1} \\ \tilde{H}_{d-1,k+1} + 1 \\ \tilde{H}_{d-1,k} + 1 \end{cases} \\ \tilde{H}_{d,k} &= j + LCP\left(s_1[i+1, |s_1|], s_2[j+1, |s_2|]\right), \\ & j = \tilde{J}_{d,k},\ i = k + j \end{aligned} \tag{2}$$



		T	G	C	A	A
	0					
T		0	1			
A		1	1			
C						
C						
A						

d=1

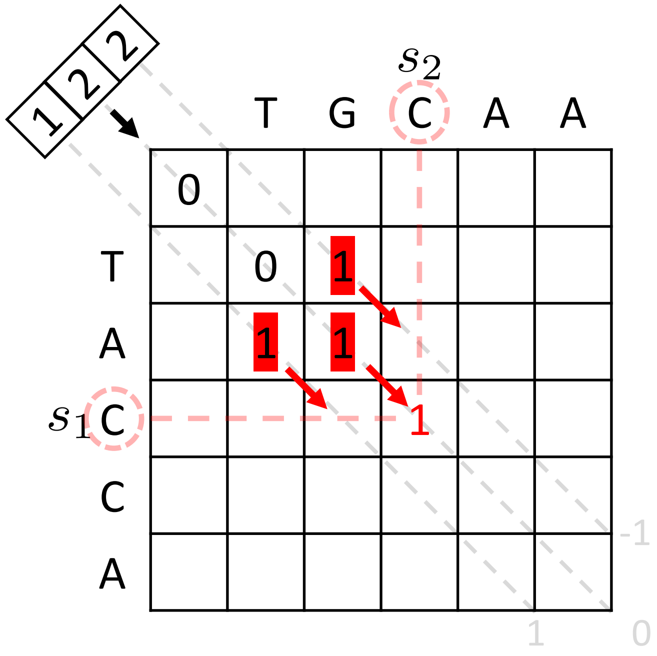
Edit distance between two sequences

- DP recurrence to compute edit distance between two strings

$$H_{i,j} = \min \begin{cases} H_{i-1,j} + 1 \\ H_{i,j-1} + 1 \\ H_{i-1,j-1} + \Delta_{i,j} \end{cases} \tag{1}$$

$$\tilde{J}_{d,k} = \max \begin{cases} \tilde{H}_{d-1,k-1} \\ \tilde{H}_{d-1,k+1} + 1 \\ \tilde{H}_{d-1,k} + 1 \end{cases} \tag{2}$$

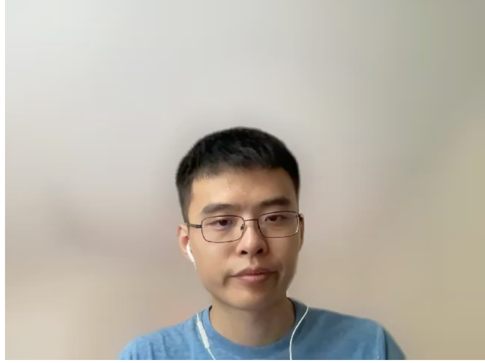
$$\begin{aligned} \tilde{H}_{d,k} =& j + LCP\left(s_1[i+1, |s_1|], s_2[j+1, |s_2|]\right), \\ & j = \tilde{J}_{d,k}, \ i = k + j \end{aligned}$$



d=1

Extension step

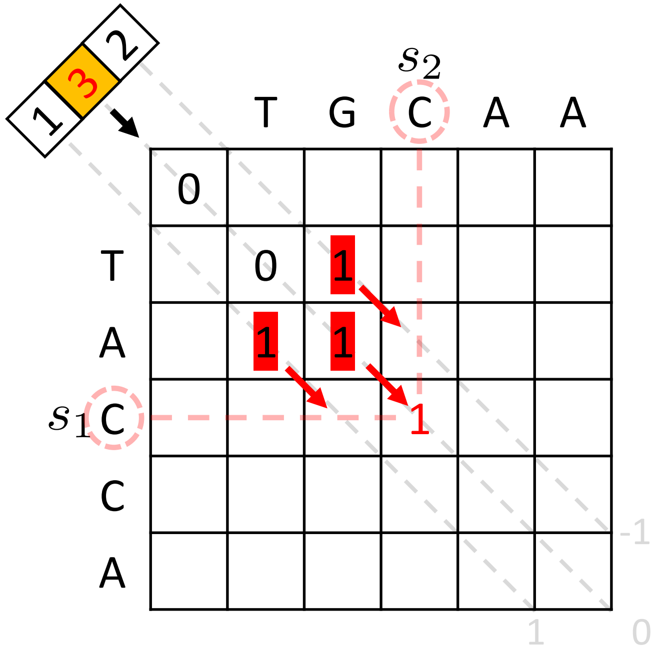
Edit distance between two sequences



- DP recurrence to compute edit distance between two strings

$$H_{i,j} = \min \begin{cases} H_{i-1,j}+1 \\ H_{i,j-1}+1 \\ H_{i-1,j-1}+\Delta_{i,j} \end{cases} \tag{1}$$

$$\begin{aligned} \tilde{J}_{d,k} =& \max \begin{cases} \tilde{H}_{d-1,k-1} \\ \tilde{H}_{d-1,k+1}+1 \\ \tilde{H}_{d-1,k}+1 \end{cases} \\ \tilde{H}_{d,k} =& j + LCP\left(s_1[i+1,|s_1|], s_2[j+1,|s_2|]\right), \\ & j=\tilde{J}_{d,k},\ i=k+j \end{aligned} \tag{2}$$



d=1

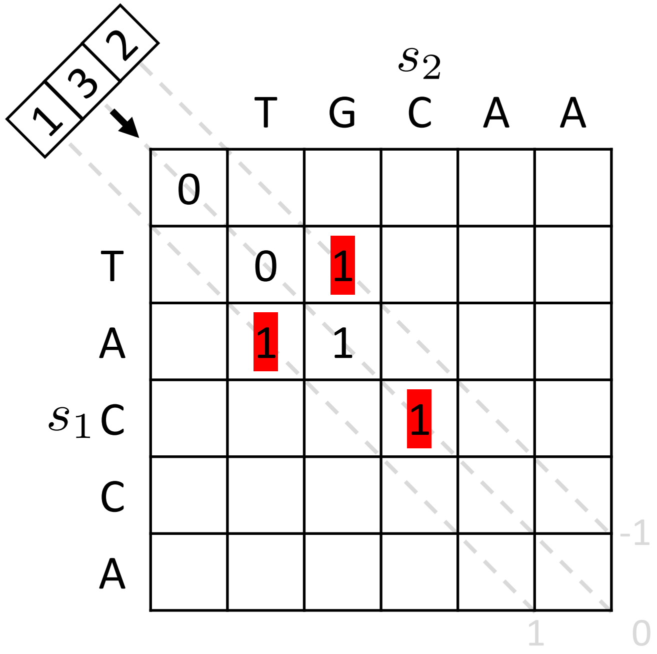
Extension step

Edit distance between two sequences

- DP recurrence to compute edit distance between two strings

$$H_{i,j} = \min \begin{cases} H_{i-1,j} + 1 \\ H_{i,j-1} + 1 \\ H_{i-1,j-1} + \Delta_{i,j} \end{cases} \tag{1}$$

$$\begin{aligned} \tilde{J}_{d,k} =& \max \begin{cases} \tilde{H}_{d-1,k-1} \\ \tilde{H}_{d-1,k+1} + 1 \\ \tilde{H}_{d-1,k} + 1 \end{cases} \\ \tilde{H}_{d,k} =& j + LCP\left(s_1[i+1, |s_1|], s_2[j+1, |s_2|]\right), \\ & j = \tilde{J}_{d,k},\ i = k + j \end{aligned} \tag{2}$$



d=1

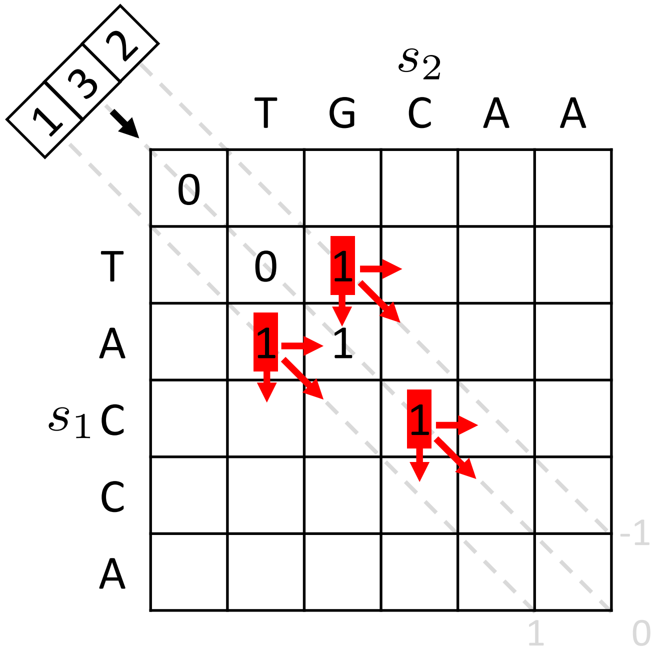
Edit distance between two sequences

- DP recurrence to compute edit distance between two strings

$$H_{i,j} = \min \begin{cases} H_{i-1,j} + 1 \\ H_{i,j-1} + 1 \\ H_{i-1,j-1} + \Delta_{i,j} \end{cases} \tag{1}$$

$$\tilde{J}_{d,k} = \max \begin{cases} \tilde{H}_{d-1,k-1} \\ \tilde{H}_{d-1,k+1} + 1 \\ \tilde{H}_{d-1,k} + 1 \end{cases} \tag{2}$$

$$\begin{aligned} \tilde{H}_{d,k} =& j + LCP\left(s_1[i+1, |s_1|], s_2[j+1, |s_2|]\right), \\ & j = \tilde{J}_{d,k},\ i = k + j \end{aligned}$$



d=2

Expansion step

Edit distance between two sequences

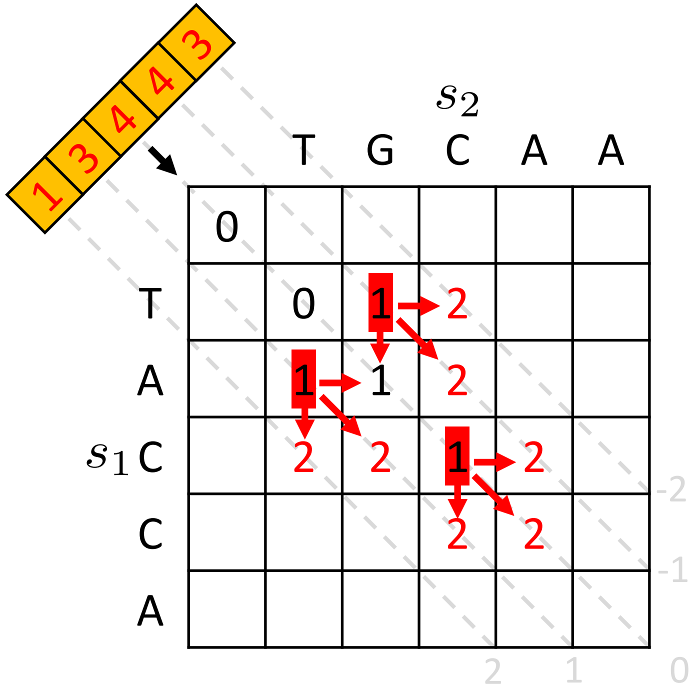
- DP recurrence to compute edit distance between two strings

$$H_{i,j} = \min \begin{cases} H_{i-1,j} + 1 \\ H_{i,j-1} + 1 \\ H_{i-1,j-1} + \Delta_{i,j} \end{cases} \qquad (1)$$

$$\tilde{J}_{d,k} = \max \begin{cases} \tilde{H}_{d-1,k-1} \\ \tilde{H}_{d-1,k+1} + 1 \\ \tilde{H}_{d-1,k} + 1 \end{cases} \qquad (2)$$

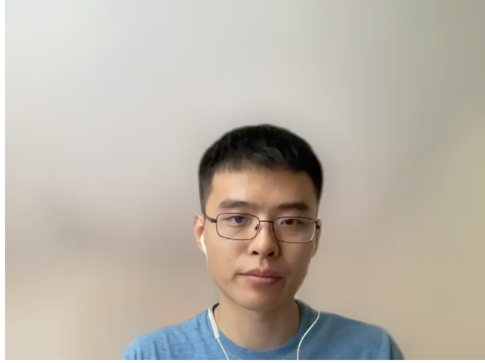
$$\tilde{H}_{d,k} = j + LCP\left(s_1[i+1, |s_1|], s_2[j+1, |s_2|]\right),$$

$$j = \tilde{J}_{d,k}, \; i = k + j$$



d=2

Expansion step



Edit distance between two sequences

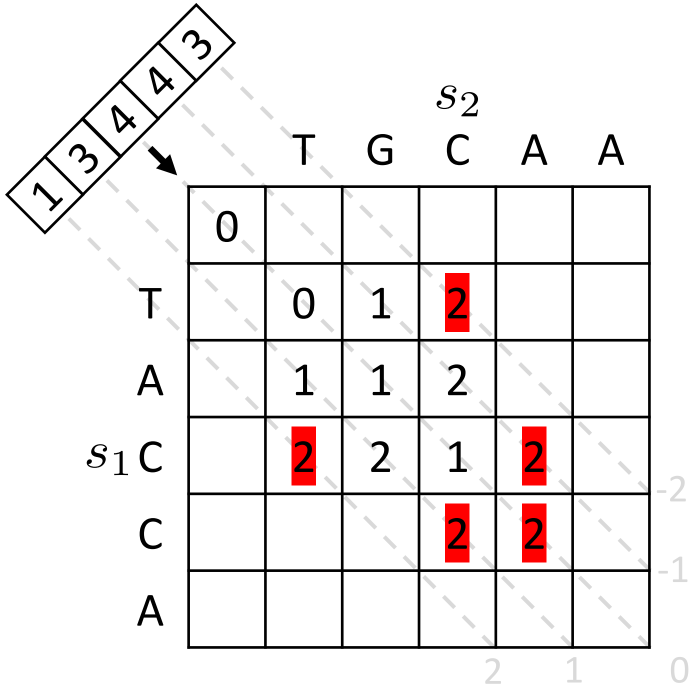
- DP recurrence to compute edit distance between two strings

$$H_{i,j} = \min \begin{cases} H_{i-1,j} + 1 \\ H_{i,j-1} + 1 \\ H_{i-1,j-1} + \Delta_{i,j} \end{cases} \tag{1}$$

$$\tilde{J}_{d,k} = \max \begin{cases} \tilde{H}_{d-1,k-1} \\ \tilde{H}_{d-1,k+1} + 1 \\ \tilde{H}_{d-1,k} + 1 \end{cases} \tag{2}$$

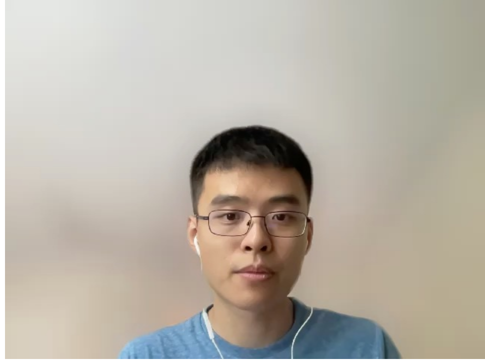
$$\tilde{H}_{d,k} = j + LCP\left(s_1[i+1, |s_1|], s_2[j+1, |s_2|]\right),$$

$$j = \tilde{J}_{d,k},\ i = k + j$$



d=2

Edit distance between two sequences



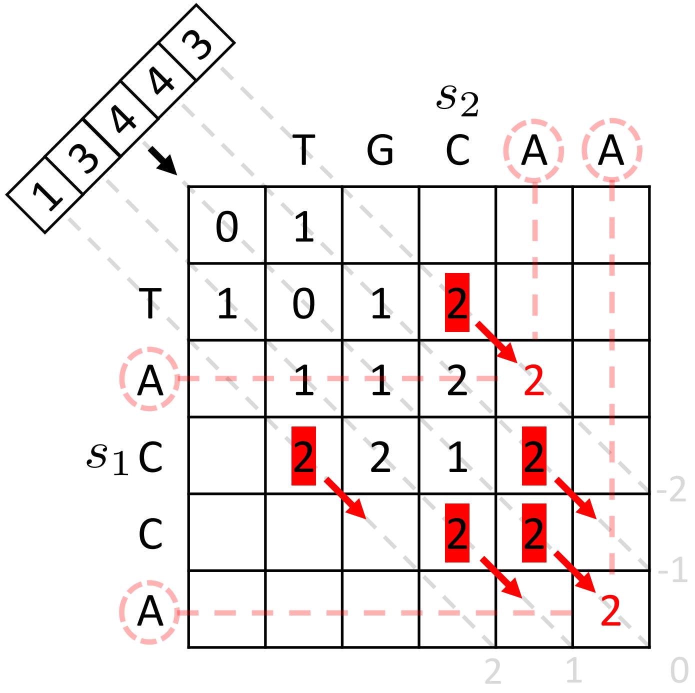
- DP recurrence to compute edit distance between two strings

$$H_{i,j} = \min \begin{cases} H_{i-1,j} + 1 \\ H_{i,j-1} + 1 \\ H_{i-1,j-1} + \Delta_{i,j} \end{cases} \tag{1}$$

$$\tilde{J}_{d,k} = \max \begin{cases} \tilde{H}_{d-1,k-1} \\ \tilde{H}_{d-1,k+1} + 1 \\ \tilde{H}_{d-1,k} + 1 \end{cases} \tag{2}$$

$$\tilde{H}_{d,k} = j + LCP\left(s_1[i+1, |s_1|], s_2[j+1, |s_2|]\right),$$

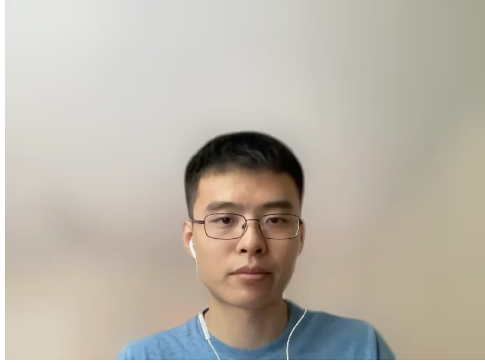
$$j = \tilde{J}_{d,k},\ \ i = k + j$$



		T	G	C	A	A
	0	1				
T	1	0	1	2		
A		1	1	2	2	
C		2	2	1	2	
C				2	2	
A						2

d=2

Extension step



Edit distance between two sequences

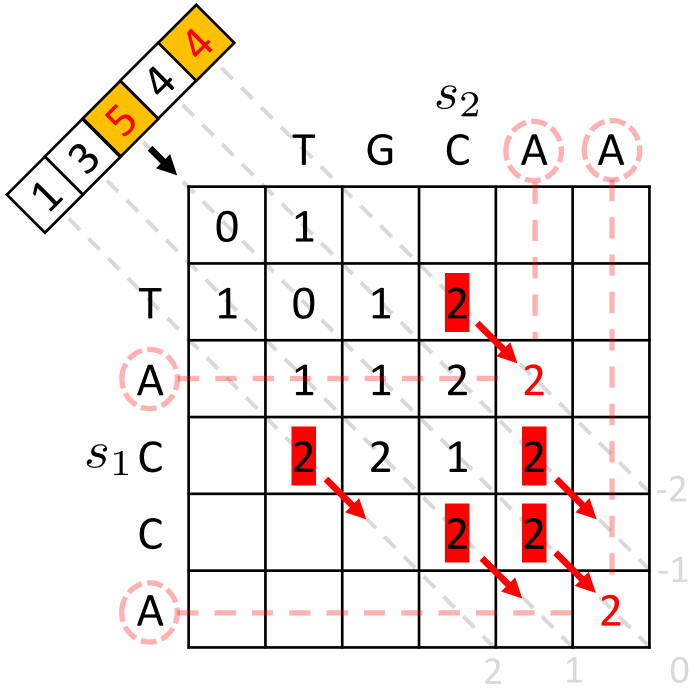
- DP recurrence to compute edit distance between two strings

$$H_{i,j} = \min \begin{cases} H_{i-1,j} + 1 \\ H_{i,j-1} + 1 \\ H_{i-1,j-1} + \Delta_{i,j} \end{cases} \tag{1}$$

$$\tilde{J}_{d,k} = \max \begin{cases} \tilde{H}_{d-1,k-1} \\ \tilde{H}_{d-1,k+1} + 1 \\ \tilde{H}_{d-1,k} + 1 \end{cases} \tag{2}$$

$$\tilde{H}_{d,k} = j + LCP\left(s_1[i+1, |s_1|], s_2[j+1, |s_2|]\right),$$

$$j = \tilde{J}_{d,k},\ i = k + j$$



d=2

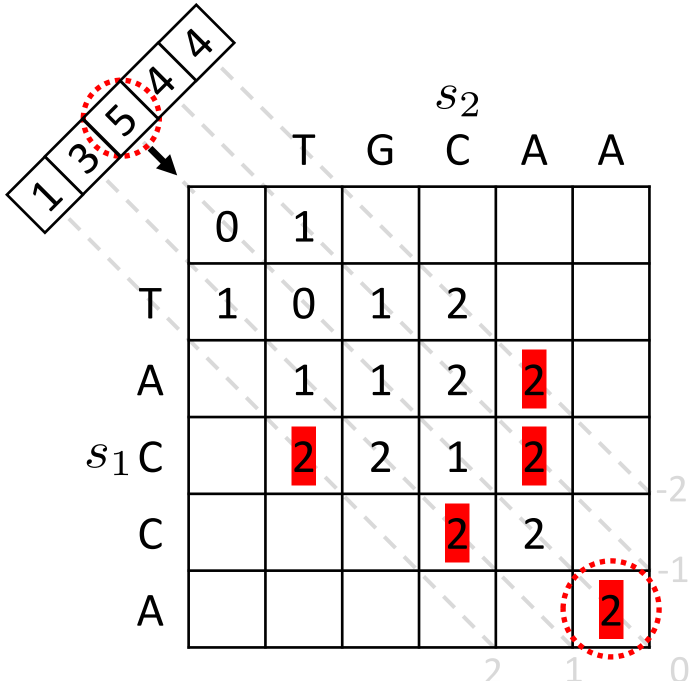
Extension step

Edit distance between two sequences

- DP recurrence to compute edit distance between two strings

$$H_{i,j} = \min \begin{cases} H_{i-1,j} + 1 \\ H_{i,j-1} + 1 \\ H_{i-1,j-1} + \Delta_{i,j} \end{cases} \tag{1}$$

$$\begin{aligned} \tilde{J}_{d,k} =& \max \begin{cases} \tilde{H}_{d-1,k-1} \\ \tilde{H}_{d-1,k+1} + 1 \\ \tilde{H}_{d-1,k} + 1 \end{cases} \\ \tilde{H}_{d,k} =& j + LCP\left(s_1[i+1, |s_1|], s_2[j+1, |s_2|]\right), \\ & j = \tilde{J}_{d,k},\ i = k + j \end{aligned} \tag{2}$$



	T	G	C	A	A
0	1				
T 1	0	1	2		
A	1	1	2	2	
C	2	2	1	2	
C			2	2	
A					2

d=2

Return min cost

Edit distance between two sequences

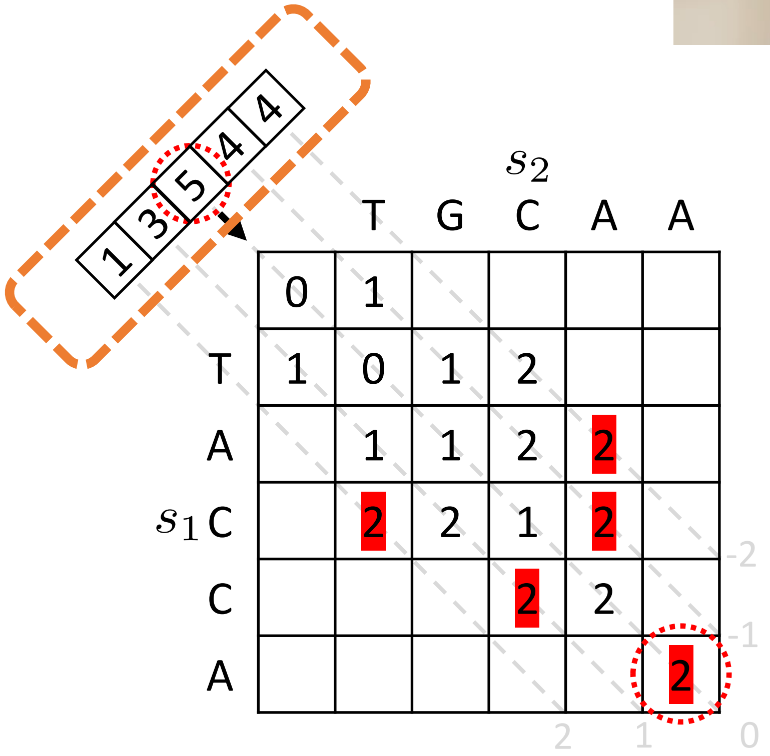
- DP recurrence to compute edit distance between two strings

$$H_{i,j} = \min \begin{cases} H_{i-1,j} + 1 \\ H_{i,j-1} + 1 \\ H_{i-1,j-1} + \Delta_{i,j} \end{cases} \qquad (1)$$

$$\tilde{J}_{d,k} = \max \begin{cases} \tilde{H}_{d-1,k-1} \\ \tilde{H}_{d-1,k+1} + 1 \\ \tilde{H}_{d-1,k} + 1 \end{cases} \qquad (2)$$

$$\tilde{H}_{d,k} = j + LCP\left(s_1[i+1, |s_1|], s_2[j+1, |s_2|]\right),$$

$$j = \tilde{J}_{d,k}, \;\; i = k + j$$



		T	G	C	A	A
	0	1				
T	1	0	1	2		
A		1	1	2	2	
s_1 C		2	2	1	2	
C				2	2	
A						2

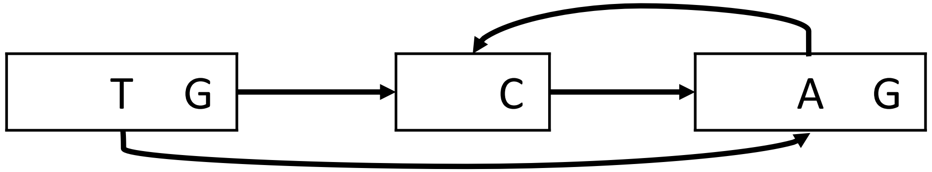
d=2

Return min cost

Edit distance of sequence to graph alignment

- DP recurrence to compute sequence-to-graph alignment

$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j} + 1, & i \geq 1 \\ H_{i,v,j-1} + 1, & j \geq 1 \\ H_{i-1,v,j-1} + \Delta_{i,v,j}, & i \geq 1, j \geq 1 \\ H_{i,u,|\sigma(u)|}, & j = 0, \forall u, (u,v) \in E \end{cases} \tag{1}$$



T

A

A

G

C

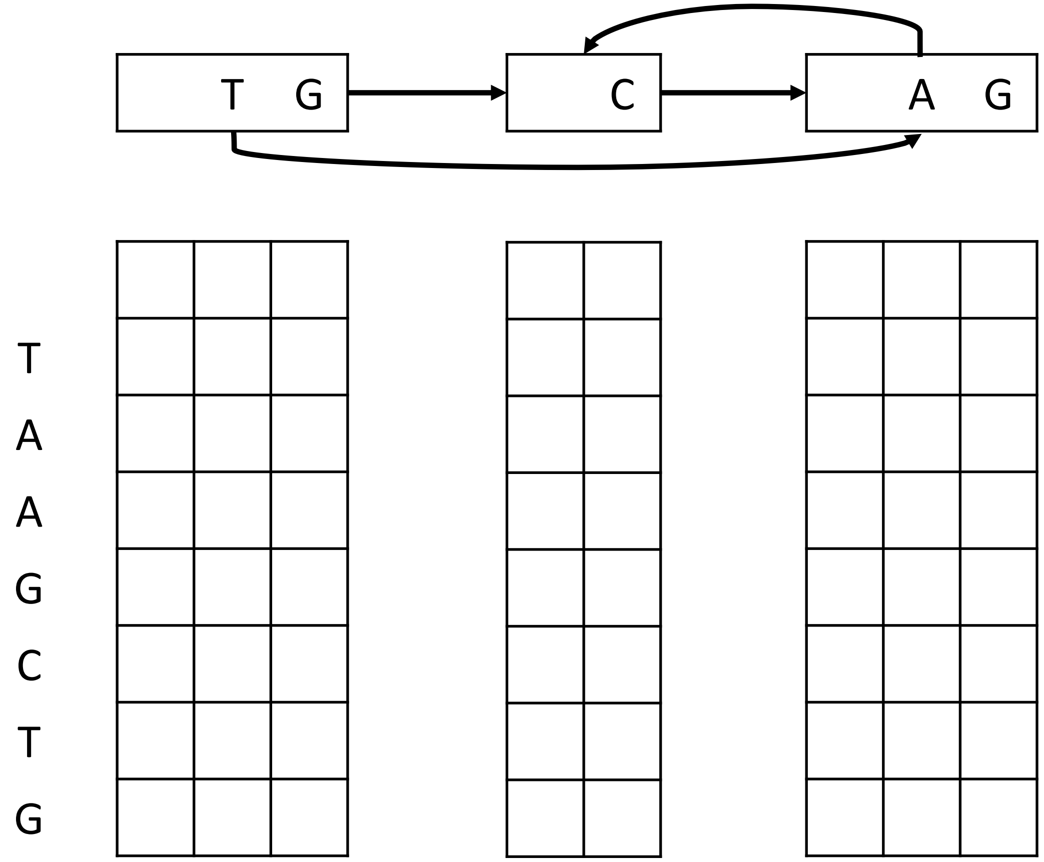
T

G

Edit distance of sequence to graph alignment

- DP recurrence to compute sequence-to-graph alignment

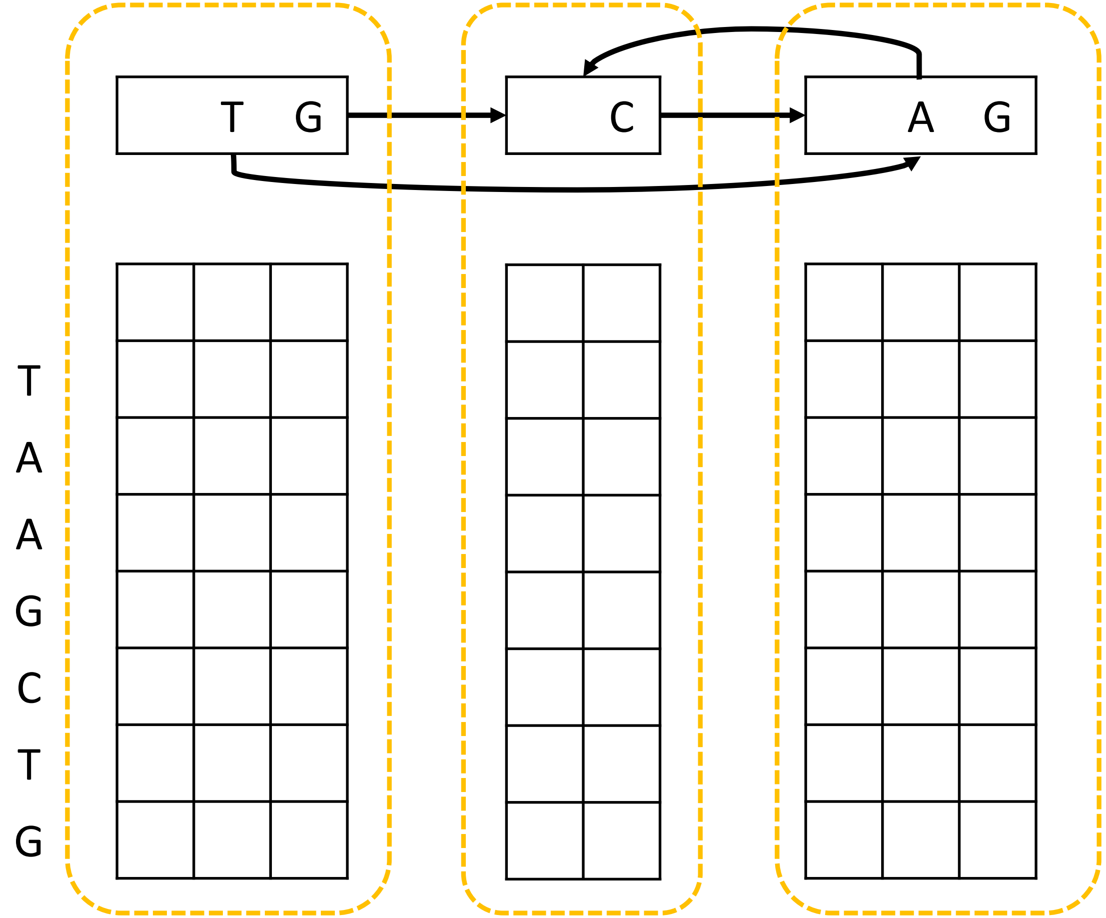
$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j} + 1, & i \geq 1 \\ H_{i,v,j-1} + 1, & j \geq 1 \\ H_{i-1,v,j-1} + \Delta_{i,v,j}, & i \geq 1, j \geq 1 \\ H_{i,u,|\sigma(u)|}, & j = 0, \forall u, (u,v) \in E \end{cases} \tag{1}$$



Edit distance of sequence to graph alignment

- DP recurrence to compute sequence-to-graph alignment

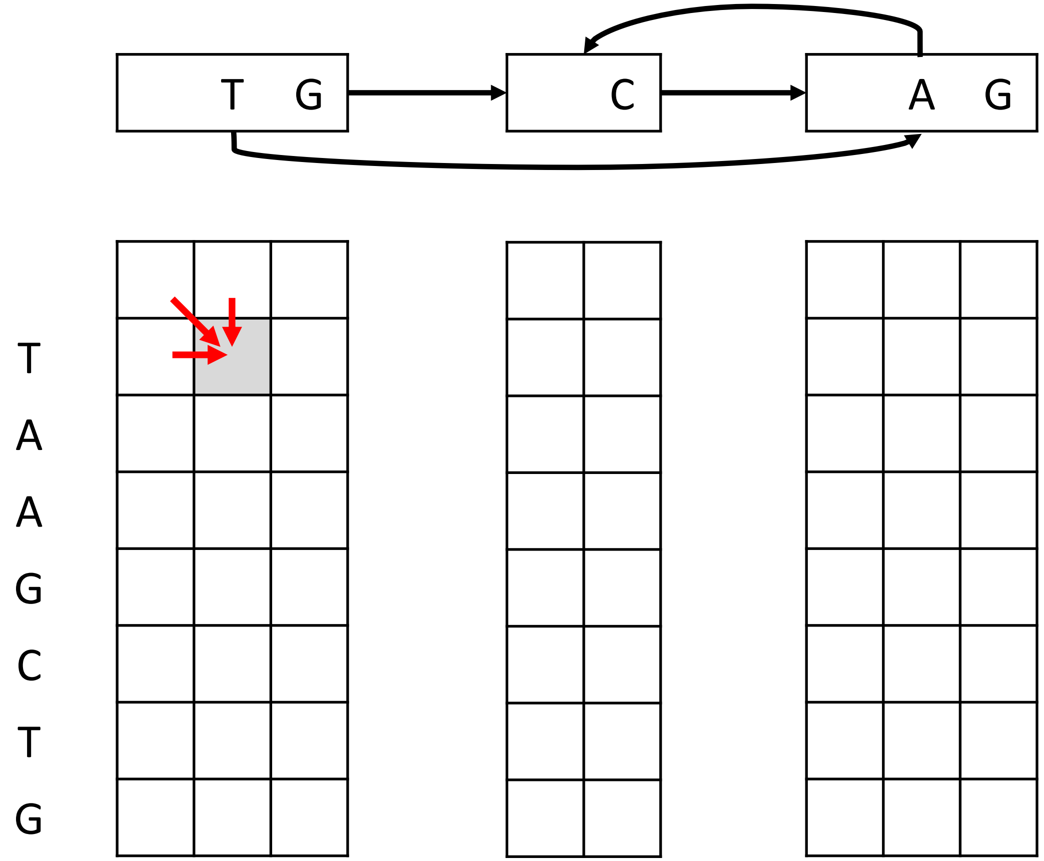
$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j} + 1, & i \geq 1 \\ H_{i,v,j-1} + 1, & j \geq 1 \\ H_{i-1,v,j-1} + \Delta_{i,v,j}, & i \geq 1, j \geq 1 \\ H_{i,u,|\sigma(u)|}, & j = 0, \forall u, (u,v) \in E \end{cases} \tag{1}$$



Edit distance of sequence to graph alignment

- DP recurrence to compute sequence-to-graph alignment

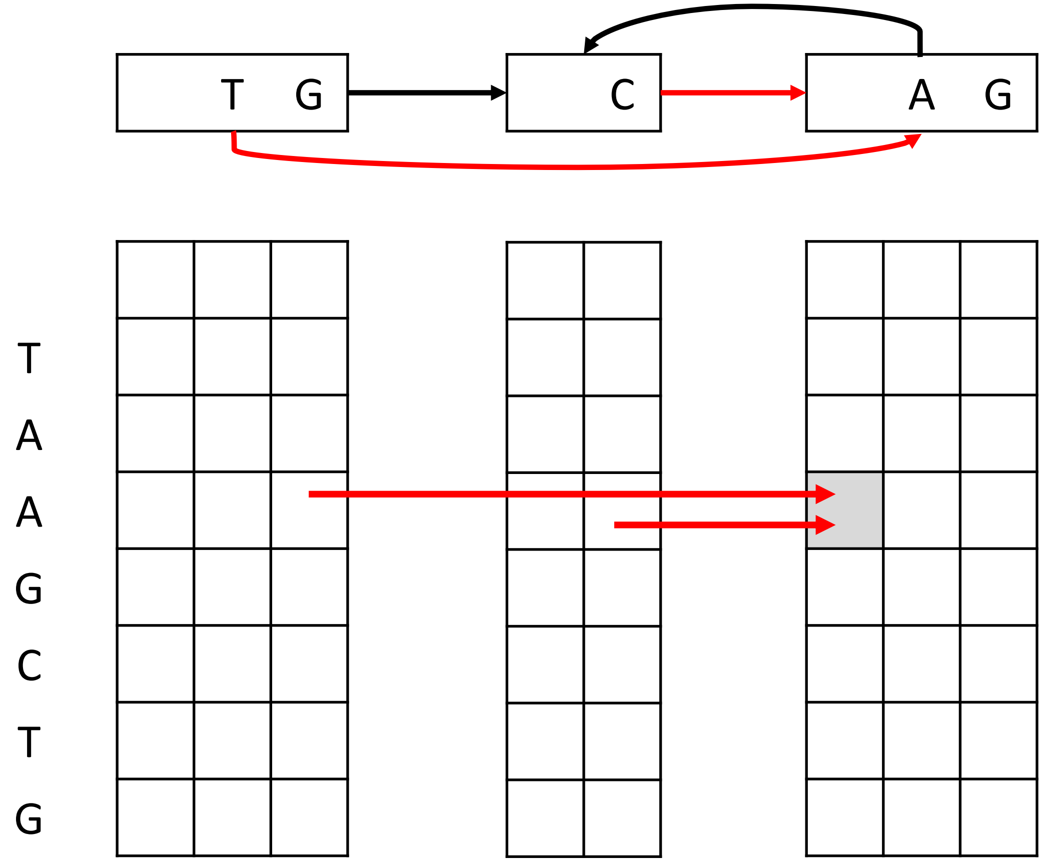
$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j} + 1, & i \geq 1 \\ H_{i,v,j-1} + 1, & j \geq 1 \\ H_{i-1,v,j-1} + \Delta_{i,v,j}, & i \geq 1, j \geq 1 \\ H_{i,u,|\sigma(u)|}, & j = 0, \forall u, (u,v) \in E \end{cases} \tag{1}$$



Edit distance of sequence to graph alignment

- DP recurrence to compute sequence-to-graph alignment

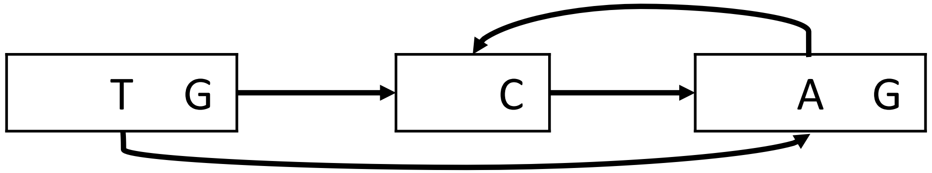
$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j} + 1, & i \geq 1 \\ H_{i,v,j-1} + 1, & j \geq 1 \\ H_{i-1,v,j-1} + \Delta_{i,v,j}, & i \geq 1, j \geq 1 \\ H_{i,u,|\sigma(u)|}, & j = 0, \forall u, (u,v) \in E \end{cases} \quad (1)$$



Edit distance of sequence to graph alignment

- DP recurrence to compute sequence-to-graph alignment

$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j} + 1, & i \geq 1 \\ H_{i,v,j-1} + 1, & j \geq 1 \\ H_{i-1,v,j-1} + \Delta_{i,v,j}, & i \geq 1, j \geq 1 \\ H_{i,u,|\sigma(u)|}, & j = 0, \forall u, (u,v) \in E \end{cases} \tag{1}$$



	0	1	2
T	1	0	1
A	2	1	1
A	3	2	2
G	4	3	2
C	5	4	3
T	6	5	4
G	7	6	5

2	3
1	2
1	2
2	2
1	3
2	1
3	2
4	3

2	3	4
1	3	4
1	1	2
2	1	2
2	2	1
1	2	2
2	2	3
3	3	2

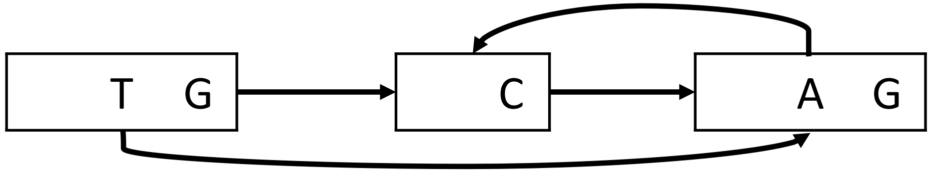
Edit distance of sequence to graph alignment

- DP recurrence to compute sequence-to-graph alignment

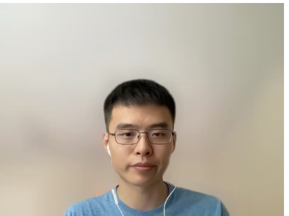
$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j}+1, & i \geq 1 \\ H_{i,v,j-1}+1, & j \geq 1 \\ H_{i-1,v,j-1}+\Delta_{i,v,j}, & i \geq 1, j \geq 1 \\ H_{i,u,|\sigma(u)|}, & j=0, \forall u, (u,v) \in E \end{cases} \tag{1}$$

$$\tilde{J}_{d,v,k} = \max \begin{cases} \tilde{H}_{d-1,v,k-1} \\ \tilde{H}_{d-1,v,k+1}+1 \\ \tilde{H}_{d-1,v,k}+1 \\ 0, \exists u, (u,v) \in E, \tilde{H}_{d,u,k-|\sigma(u)|} = |\sigma(u)| \end{cases} \tag{2}$$

$$\begin{aligned} \tilde{H}_{d,v,k} = & j + LCP\left(q[i+1,|q|], \sigma(v)[j+1,|\sigma(v)|]\right), \\ & j = \tilde{J}_{d,v,k},\ i = k+j \end{aligned}$$



	0							
T		0	1	1	2	1		
A		1	1	1	2	1	1	2
A		2	2	2	2		1	2
G			2	1				1
C					1	1	2	2
T					2	2	2	
G								2

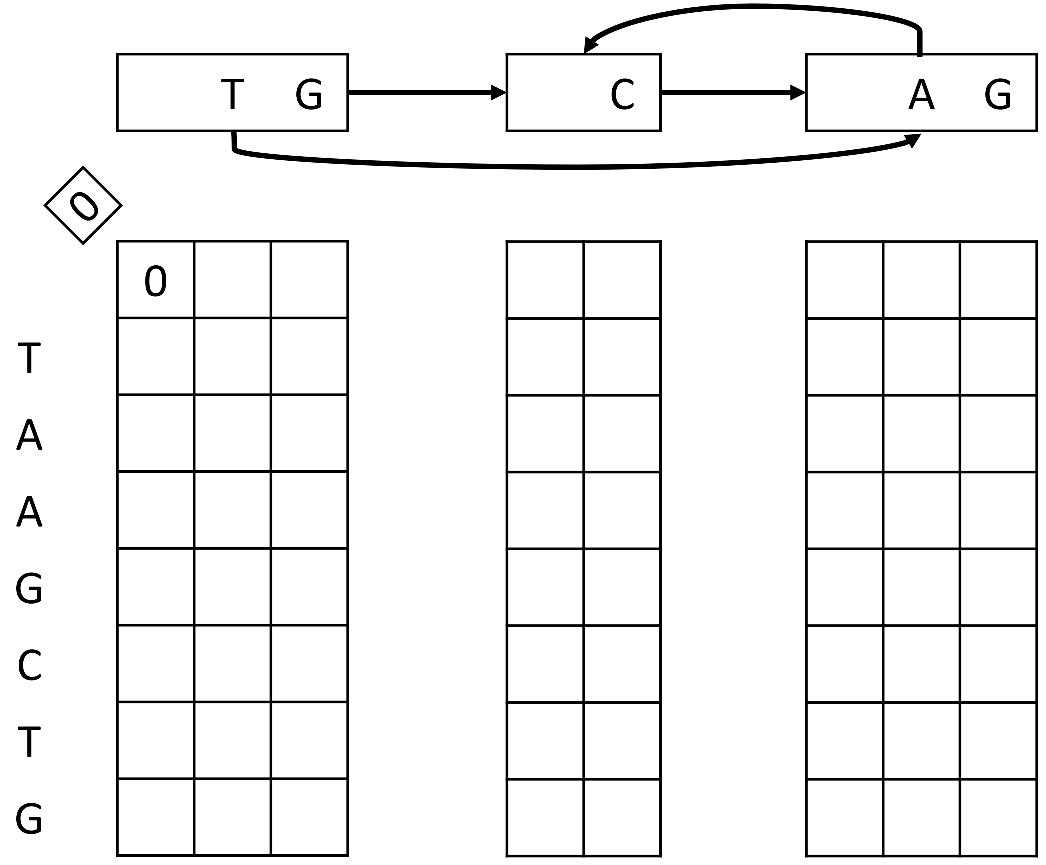


Edit distance of sequence to graph alignment

- DP recurrence to compute sequence-to-graph alignment

$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j} + 1, & i \geq 1 \\ H_{i,v,j-1} + 1, & j \geq 1 \\ H_{i-1,v,j-1} + \Delta_{i,v,j}, & i \geq 1, j \geq 1 \\ H_{i,u,|\sigma(u)|}, & j = 0, \forall u, (u,v) \in E \end{cases} \tag{1}$$

$$\begin{aligned} \tilde{J}_{d,v,k} =& \max \begin{cases} \tilde{H}_{d-1,v,k-1} \\ \tilde{H}_{d-1,v,k+1} + 1 \\ \tilde{H}_{d-1,v,k} + 1 \\ 0, \exists u, (u,v) \in E, \tilde{H}_{d,u,k-|\sigma(u)|} = |\sigma(u)| \end{cases} \\ \tilde{H}_{d,v,k} =& j + LCP\left(q[i+1,|q|], \sigma(v)[j+1,|\sigma(v)|]\right), \\ & j = \tilde{J}_{d,v,k},\ i = k + j \end{aligned} \tag{2}$$



Initial state, d=0

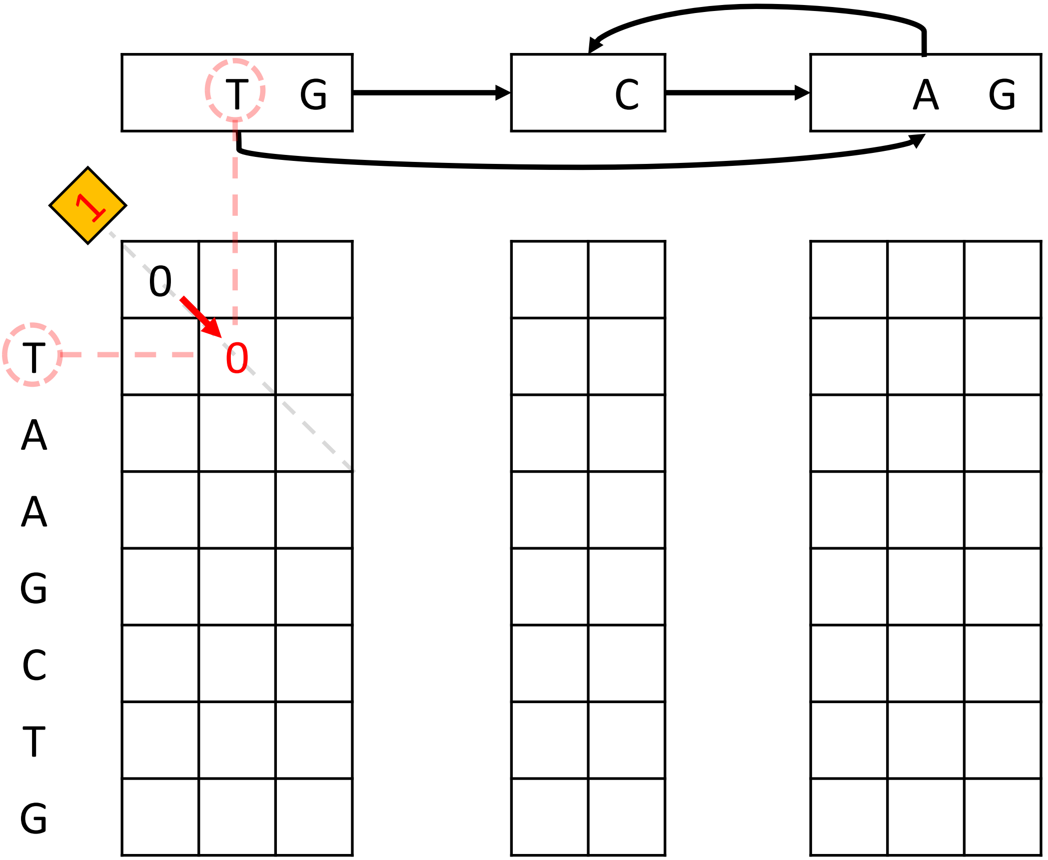
Edit distance of sequence to graph alignment

- DP recurrence to compute sequence-to-graph alignment

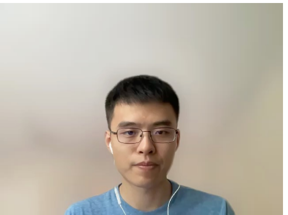
$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j} + 1, & i \geq 1 \\ H_{i,v,j-1} + 1, & j \geq 1 \\ H_{i-1,v,j-1} + \Delta_{i,v,j}, & i \geq 1, j \geq 1 \\ H_{i,u,|\sigma(u)|}, & j = 0, \forall u, (u,v) \in E \end{cases} \qquad (1)$$

$$\tilde{J}_{d,v,k} = \max \begin{cases} \tilde{H}_{d-1,v,k-1} \\ \tilde{H}_{d-1,v,k+1} + 1 \\ \tilde{H}_{d-1,v,k} + 1 \\ 0, \exists u, (u,v) \in E, \tilde{H}_{d,u,k-|\sigma(u)|} = |\sigma(u)| \end{cases} \qquad (2)$$

$$\begin{aligned} \tilde{H}_{d,v,k} =& j + LCP\left(q[i+1, |q|], \sigma(v)[j+1, |\sigma(v)|]\right), \\ & j = \tilde{J}_{d,v,k}, \ i = k + j \end{aligned}$$



Extension step, d=0

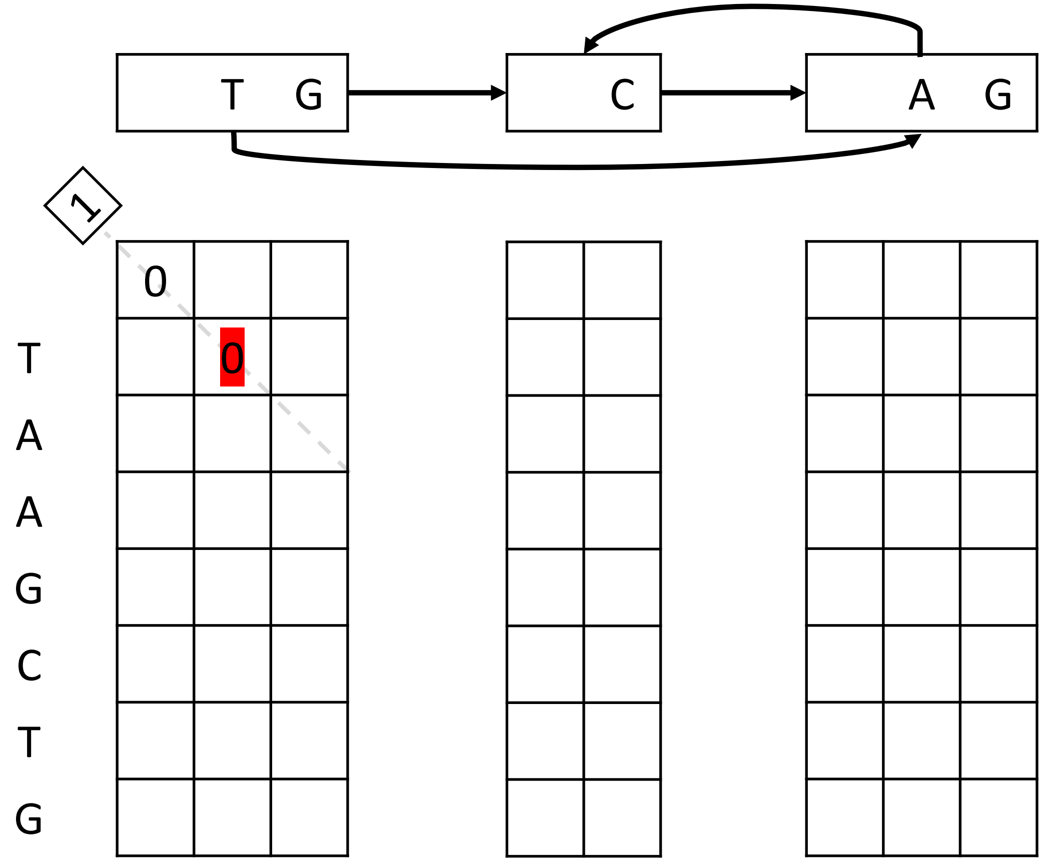


Edit distance of sequence to graph alignment

- DP recurrence to compute sequence-to-graph alignment

$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j}+1, & i \geq 1 \\ H_{i,v,j-1}+1, & j \geq 1 \\ H_{i-1,v,j-1}+\Delta_{i,v,j}, & i \geq 1, j \geq 1 \\ H_{i,u,|\sigma(u)|}, & j = 0, \forall u, (u,v) \in E \end{cases} \tag{1}$$

$$\begin{aligned} \tilde{J}_{d,v,k} &= \max \begin{cases} \tilde{H}_{d-1,v,k-1} \\ \tilde{H}_{d-1,v,k+1}+1 \\ \tilde{H}_{d-1,v,k}+1 \\ 0, \exists u, (u,v) \in E, \tilde{H}_{d,u,k-|\sigma(u)|} = |\sigma(u)| \end{cases} \\ \tilde{H}_{d,v,k} &= j + LCP\left(q[i+1,|q|], \sigma(v)[j+1,|\sigma(v)|]\right), \\ & \quad j = \tilde{J}_{d,v,k},\ i = k+j \end{aligned} \tag{2}$$



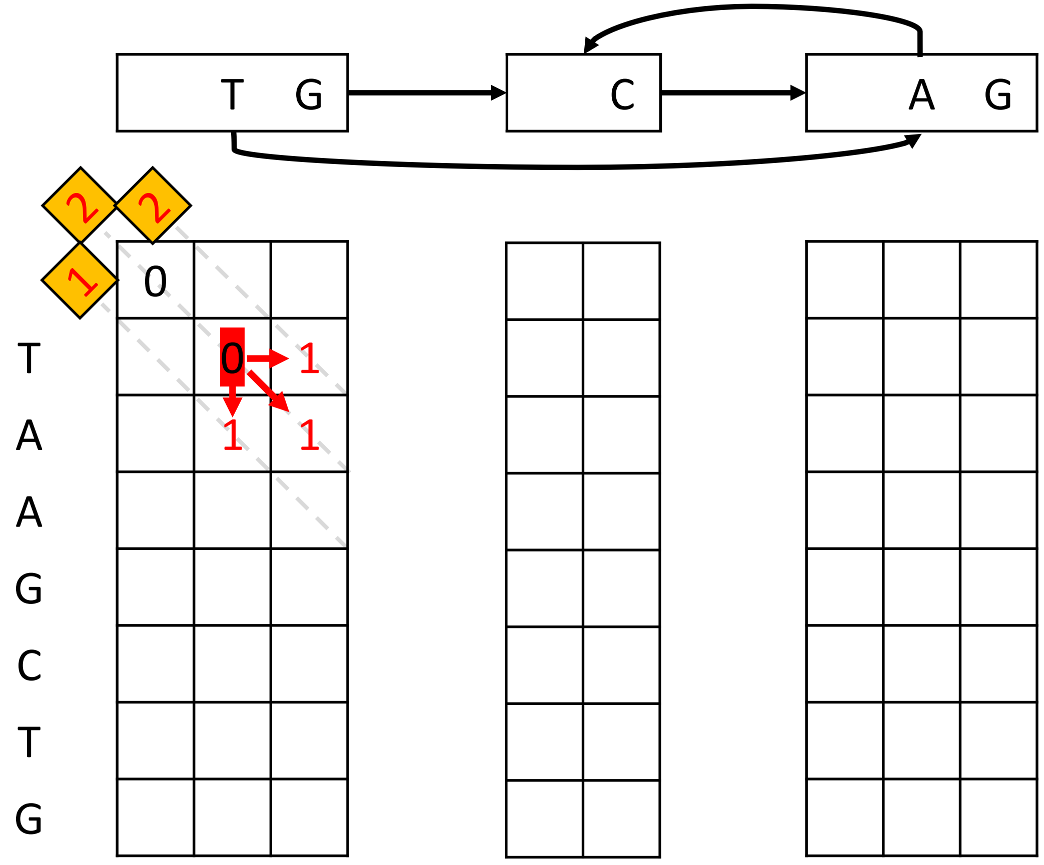
d=0

Edit distance of sequence to graph alignment

- DP recurrence to compute sequence-to-graph alignment

$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j}+1, & i \geq 1 \\ H_{i,v,j-1}+1, & j \geq 1 \\ H_{i-1,v,j-1}+\Delta_{i,v,j}, & i \geq 1, j \geq 1 \\ H_{i,u,|\sigma(u)|}, & j = 0, \forall u, (u,v) \in E \end{cases} \tag{1}$$

$$\begin{aligned} \tilde{J}_{d,v,k} &= \max \begin{cases} \tilde{H}_{d-1,v,k-1} \\ \tilde{H}_{d-1,v,k+1}+1 \\ \tilde{H}_{d-1,v,k}+1 \\ 0, \exists u, (u,v) \in E, \tilde{H}_{d,u,k-|\sigma(u)|} = |\sigma(u)| \end{cases} \\ \tilde{H}_{d,v,k} &= j + LCP\left(q[i+1,|q|], \sigma(v)[j+1,|\sigma(v)|]\right), \\ & j = \tilde{J}_{d,v,k},\ i = k+j \end{aligned} \tag{2}$$



Expansion step, d=1

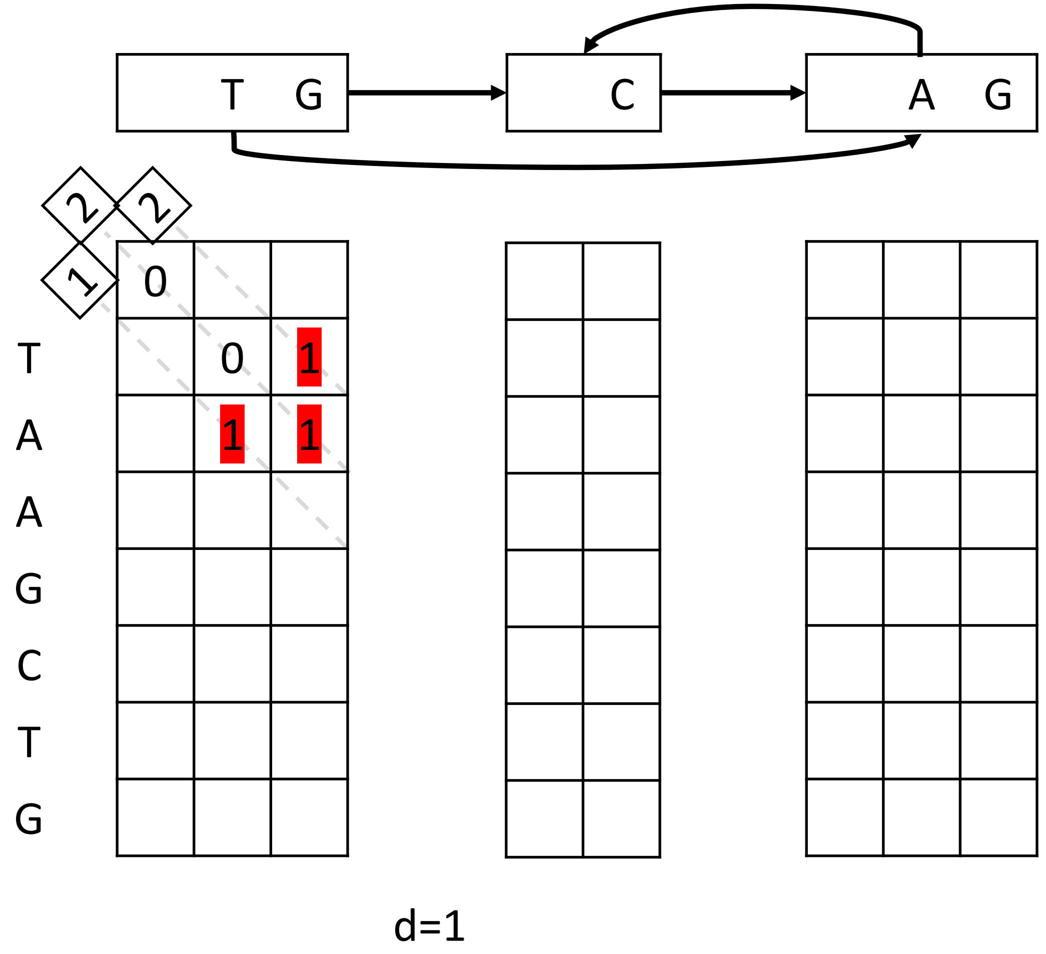
Edit distance of sequence to graph alignment

- DP recurrence to compute sequence-to-graph alignment

$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j}+1, & i \geq 1 \\ H_{i,v,j-1}+1, & j \geq 1 \\ H_{i-1,v,j-1}+\Delta_{i,v,j}, & i \geq 1, j \geq 1 \\ H_{i,u,|\sigma(u)|}, & j = 0, \forall u, (u,v) \in E \end{cases} \qquad (1)$$

$$\tilde{J}_{d,v,k} = \max \begin{cases} \tilde{H}_{d-1,v,k-1} \\ \tilde{H}_{d-1,v,k+1}+1 \\ \tilde{H}_{d-1,v,k}+1 \\ 0, \exists u, (u,v) \in E, \tilde{H}_{d,u,k-|\sigma(u)|} = |\sigma(u)| \end{cases} \qquad (2)$$

$$\tilde{H}_{d,v,k} = j + LCP\left(q[i+1,|q|], \sigma(v)[j+1,|\sigma(v)|]\right), \quad j = \tilde{J}_{d,v,k},\ i = k+j$$



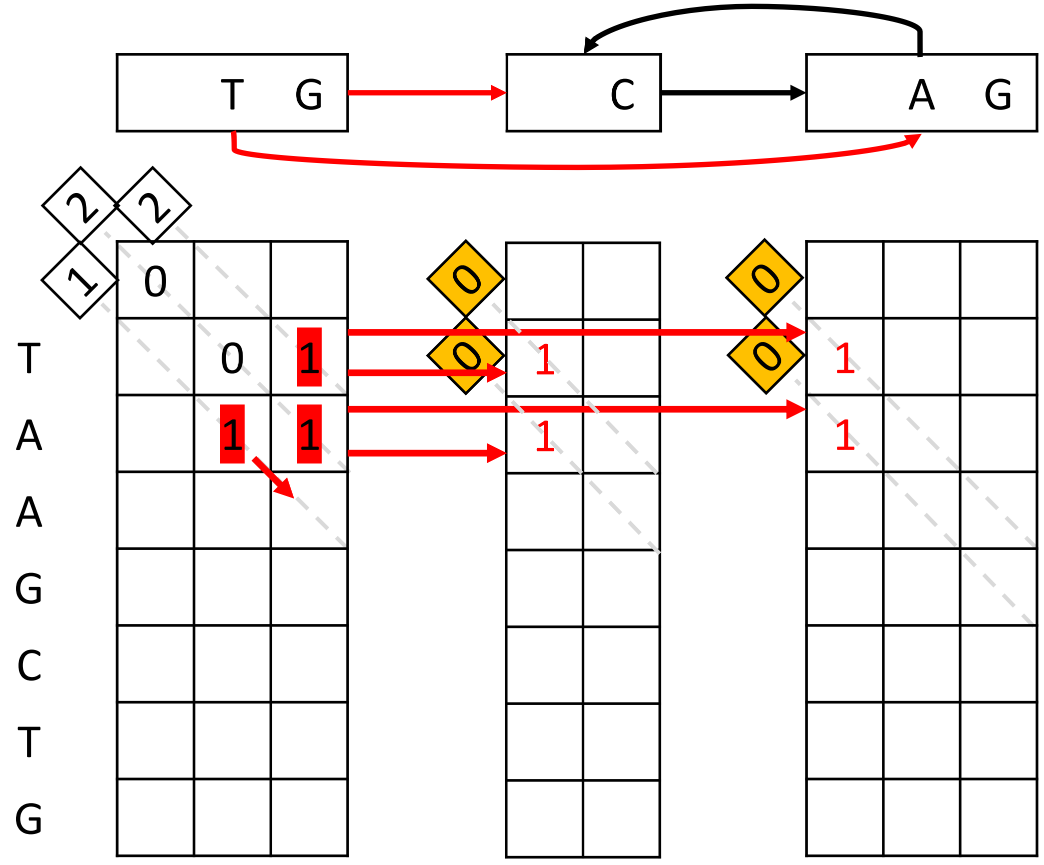
Edit distance of sequence to graph alignment

- DP recurrence to compute sequence-to-graph alignment

$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j} + 1, & i \geq 1 \\ H_{i,v,j-1} + 1, & j \geq 1 \\ H_{i-1,v,j-1} + \Delta_{i,v,j}, & i \geq 1, j \geq 1 \\ H_{i,u,|\sigma(u)|}, & j = 0, \forall u, (u,v) \in E \end{cases} \qquad (1)$$

$$\tilde{J}_{d,v,k} = \max \begin{cases} \tilde{H}_{d-1,v,k-1} \\ \tilde{H}_{d-1,v,k+1} + 1 \\ \tilde{H}_{d-1,v,k} + 1 \\ 0, \exists u, (u,v) \in E, \tilde{H}_{d,u,k-|\sigma(u)|} = |\sigma(u)| \end{cases} \qquad (2)$$

$$\tilde{H}_{d,v,k} = j + LCP\left(q[i+1, |q|], \sigma(v)[j+1, |\sigma(v)|]\right),$$
$$j = \tilde{J}_{d,v,k}, \; i = k + j$$



Extension step, d=1

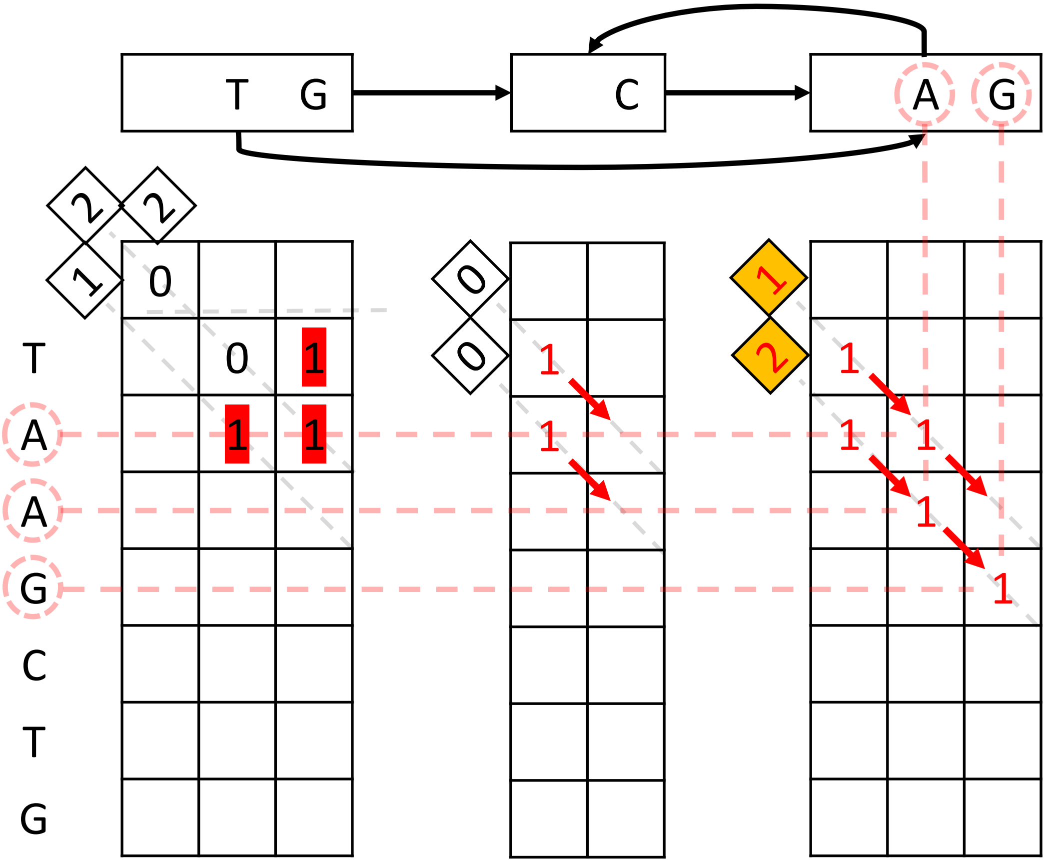
Edit distance of sequence to graph alignment

- DP recurrence to compute sequence-to-graph alignment

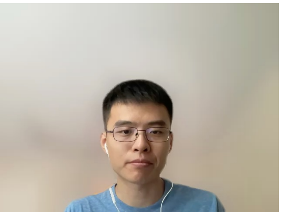
$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j} + 1, & i \geq 1 \\ H_{i,v,j-1} + 1, & j \geq 1 \\ H_{i-1,v,j-1} + \Delta_{i,v,j}, & i \geq 1, j \geq 1 \\ H_{i,u,|\sigma(u)|}, & j = 0, \forall u, (u,v) \in E \end{cases} \tag{1}$$

$$\tilde{J}_{d,v,k} = \max \begin{cases} \tilde{H}_{d-1,v,k-1} \\ \tilde{H}_{d-1,v,k+1} + 1 \\ \tilde{H}_{d-1,v,k} + 1 \\ 0, \exists u, (u,v) \in E, \tilde{H}_{d,u,k-|\sigma(u)|} = |\sigma(u)| \end{cases} \tag{2}$$

$$\tilde{H}_{d,v,k} = j + LCP\left(q[i+1,|q|], \sigma(v)[j+1,|\sigma(v)|]\right), \quad j = \tilde{J}_{d,v,k}, \; i = k + j$$



Extension step, d=1



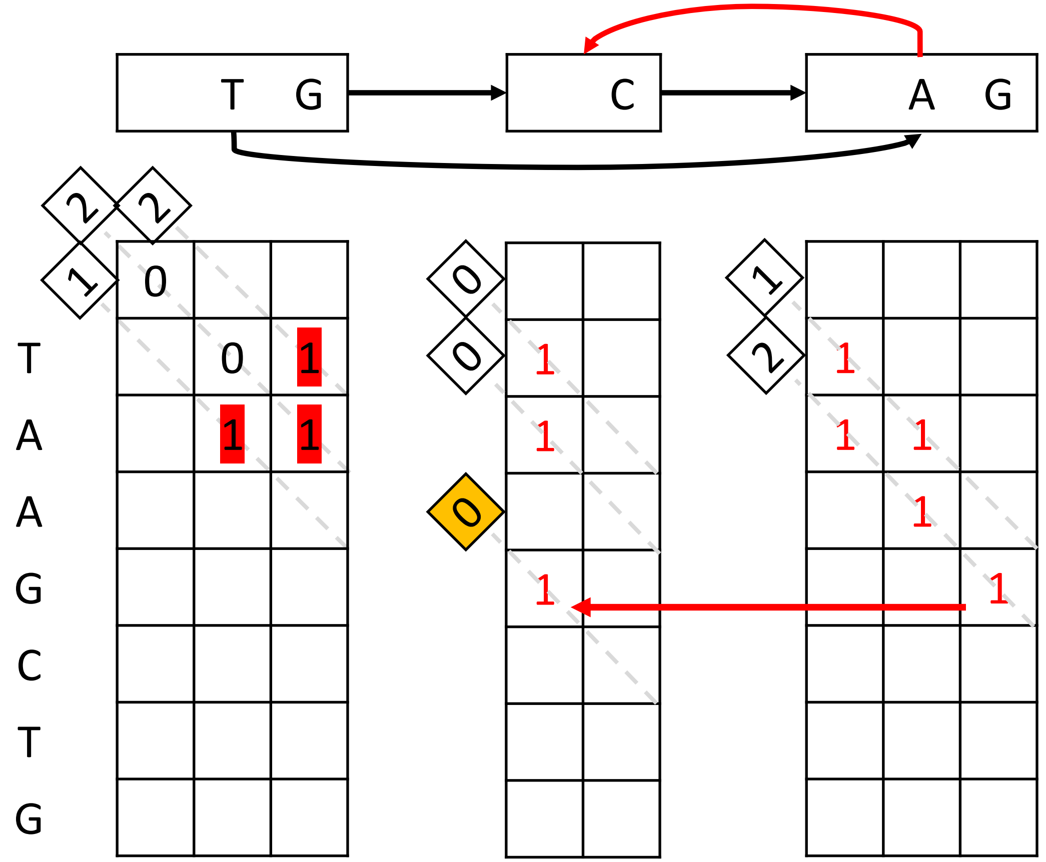
Edit distance of sequence to graph alignment

- DP recurrence to compute sequence-to-graph alignment

$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j}+1, & i \geq 1 \\ H_{i,v,j-1}+1, & j \geq 1 \\ H_{i-1,v,j-1}+\Delta_{i,v,j}, & i \geq 1, j \geq 1 \\ H_{i,u,|\sigma(u)|}, & j=0, \forall u,(u,v) \in E \end{cases} \qquad (1)$$

$$\tilde{J}_{d,v,k} = \max \begin{cases} \tilde{H}_{d-1,v,k-1} \\ \tilde{H}_{d-1,v,k+1}+1 \\ \tilde{H}_{d-1,v,k}+1 \\ 0, \exists u,(u,v) \in E, \tilde{H}_{d,u,k-|\sigma(u)|} = |\sigma(u)| \end{cases}$$

$$\tilde{H}_{d,v,k} = j + LCP\left(q[i+1,|q|], \sigma(v)[j+1,|\sigma(v)|]\right),$$
$$j = \tilde{J}_{d,v,k}, \ i = k+j \qquad (2)$$



Extension step, d=1

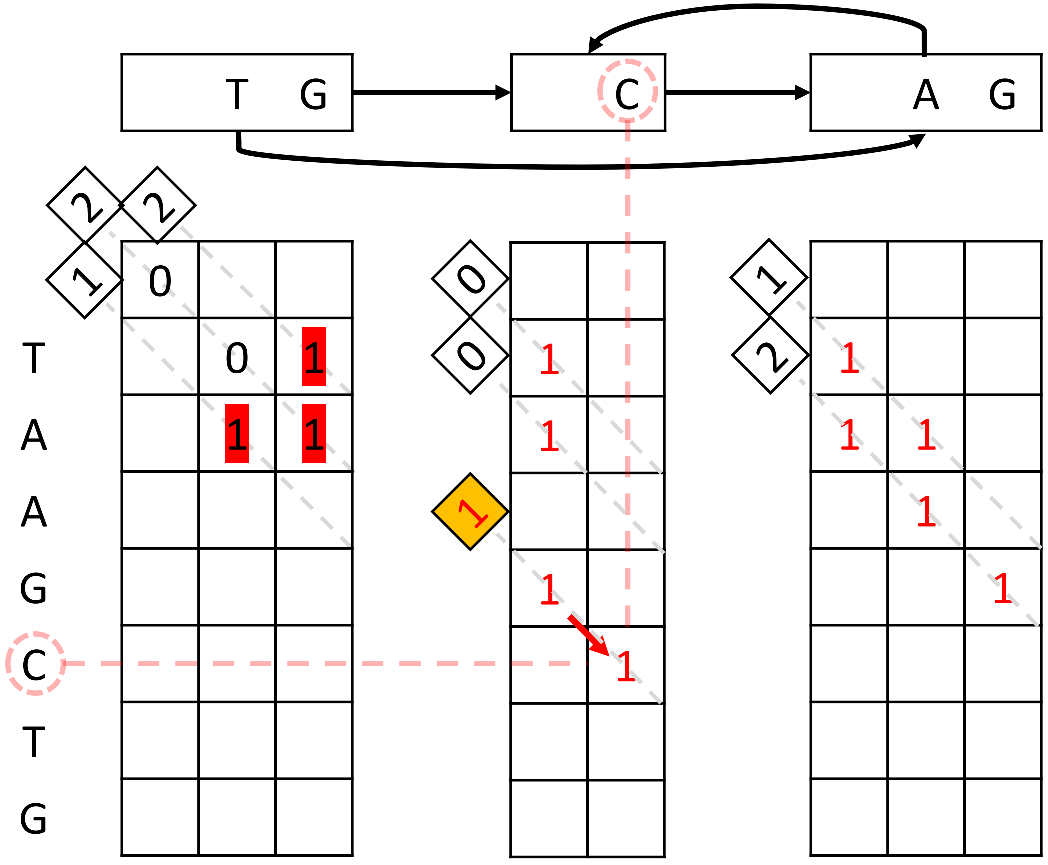
Edit distance of sequence to graph alignment

- DP recurrence to compute sequence-to-graph alignment

$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j} + 1, & i \geq 1 \\ H_{i,v,j-1} + 1, & j \geq 1 \\ H_{i-1,v,j-1} + \Delta_{i,v,j}, & i \geq 1, j \geq 1 \\ H_{i,u,|\sigma(u)|}, & j = 0, \forall u, (u,v) \in E \end{cases} \tag{1}$$

$$\tilde{J}_{d,v,k} = \max \begin{cases} \tilde{H}_{d-1,v,k-1} \\ \tilde{H}_{d-1,v,k+1} + 1 \\ \tilde{H}_{d-1,v,k} + 1 \\ 0, \exists u, (u,v) \in E, \tilde{H}_{d,u,k-|\sigma(u)|} = |\sigma(u)| \end{cases} \tag{2}$$

$$\begin{aligned} \tilde{H}_{d,v,k} =& j + LCP\left(q[i+1,|q|], \sigma(v)[j+1,|\sigma(v)|]\right), \\ & j = \tilde{J}_{d,v,k}, \ i = k + j \end{aligned}$$



Extension step, d=1

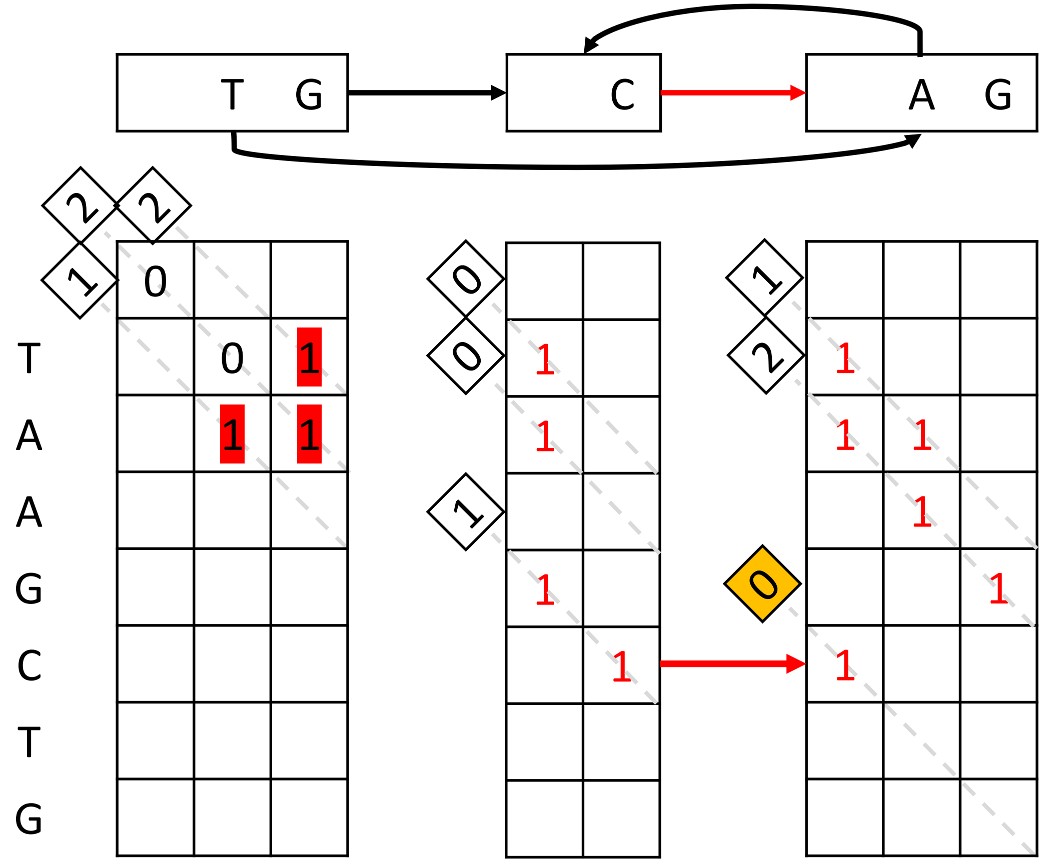
Edit distance of sequence to graph alignment

- DP recurrence to compute sequence-to-graph alignment

$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j} + 1, & i \geq 1 \\ H_{i,v,j-1} + 1, & j \geq 1 \\ H_{i-1,v,j-1} + \Delta_{i,v,j}, & i \geq 1, j \geq 1 \\ H_{i,u,|\sigma(u)|}, & j = 0, \forall u, (u,v) \in E \end{cases} \tag{1}$$

$$\tilde{J}_{d,v,k} = \max \begin{cases} \tilde{H}_{d-1,v,k-1} \\ \tilde{H}_{d-1,v,k+1} + 1 \\ \tilde{H}_{d-1,v,k} + 1 \\ 0, \exists u, (u,v) \in E, \tilde{H}_{d,u,k-|\sigma(u)|} = |\sigma(u)| \end{cases} \tag{2}$$

$$\tilde{H}_{d,v,k} = j + LCP\left(q[i+1, |q|], \sigma(v)[j+1, |\sigma(v)|]\right), \quad j = \tilde{J}_{d,v,k}, \; i = k + j$$



Extension step, d=1

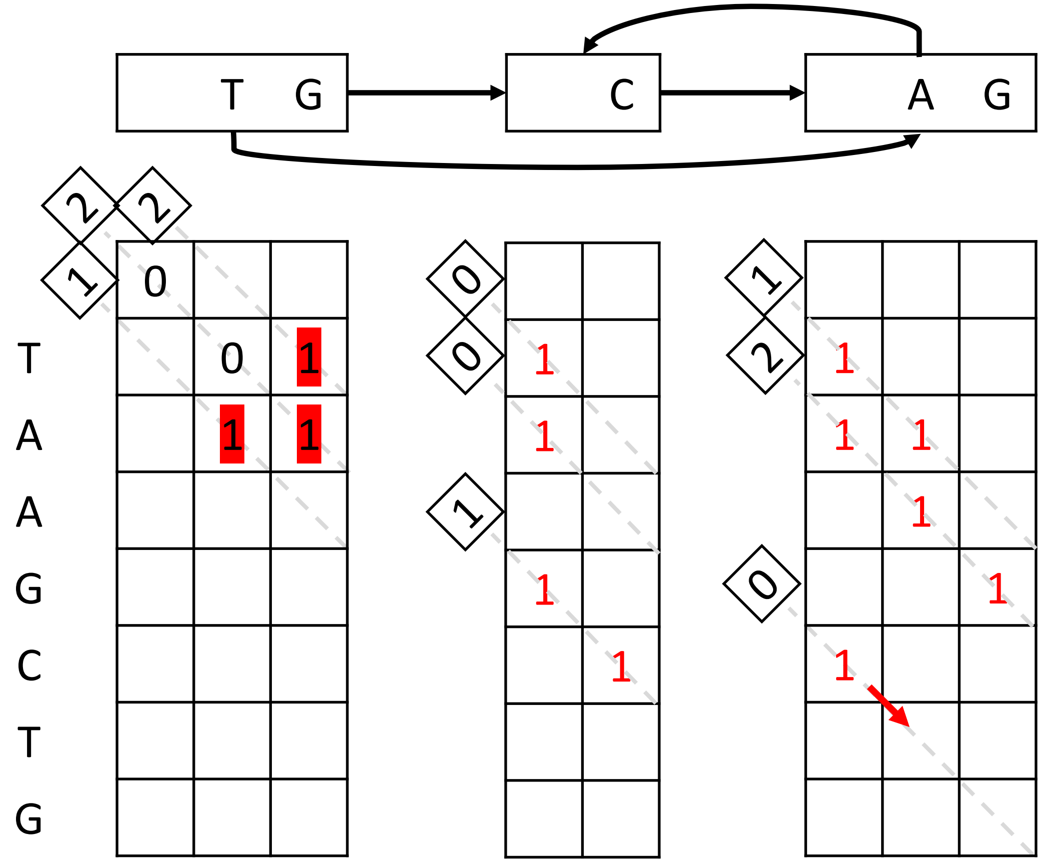
Edit distance of sequence to graph alignment

- DP recurrence to compute sequence-to-graph alignment

$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j}+1, & i \ge 1 \\ H_{i,v,j-1}+1, & j \ge 1 \\ H_{i-1,v,j-1}+\Delta_{i,v,j}, & i \ge 1, j \ge 1 \\ H_{i,u,|\sigma(u)|}, & j = 0, \forall u, (u,v) \in E \end{cases} \qquad (1)$$

$$\tilde{J}_{d,v,k} = \max \begin{cases} \tilde{H}_{d-1,v,k-1} \\ \tilde{H}_{d-1,v,k+1}+1 \\ \tilde{H}_{d-1,v,k}+1 \\ 0, \exists u, (u,v) \in E, \tilde{H}_{d,u,k-|\sigma(u)|} = |\sigma(u)| \end{cases} \qquad (2)$$

$$\tilde{H}_{d,v,k} = j + LCP\left(q[i+1,|q|], \sigma(v)[j+1,|\sigma(v)|]\right), \quad j = \tilde{J}_{d,v,k}, \; i = k + j$$



Extension step, d=1

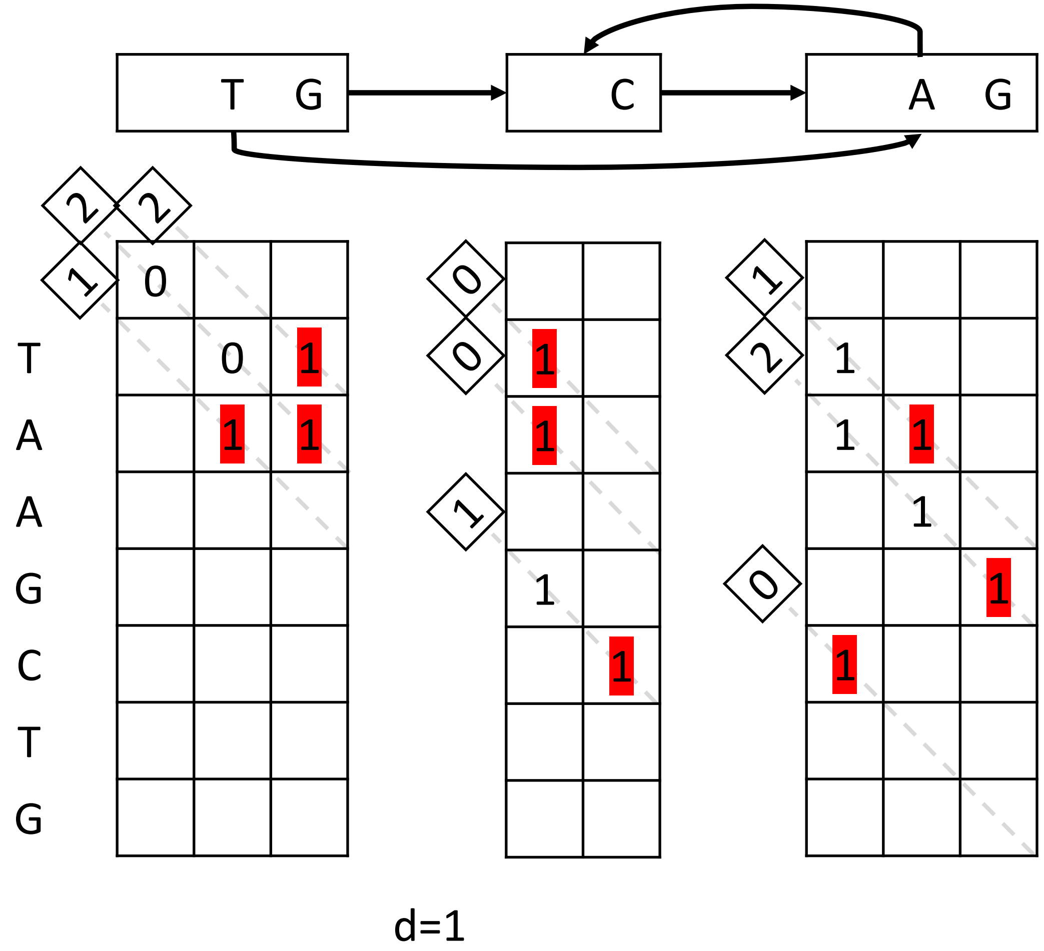
Edit distance of sequence to graph alignment

- DP recurrence to compute sequence-to-graph alignment

$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j}+1, & i\geq 1 \\ H_{i,v,j-1}+1, & j\geq 1 \\ H_{i-1,v,j-1}+\Delta_{i,v,j}, & i\geq 1, j\geq 1 \\ H_{i,u,|\sigma(u)|}, & j=0, \forall u, (u,v)\in E \end{cases} \qquad (1)$$

$$\tilde{J}_{d,v,k} = \max \begin{cases} \tilde{H}_{d-1,v,k-1} \\ \tilde{H}_{d-1,v,k+1}+1 \\ \tilde{H}_{d-1,v,k}+1 \\ 0, \exists u, (u,v)\in E, \tilde{H}_{d,u,k-|\sigma(u)|} = |\sigma(u)| \end{cases} \qquad (2)$$

$$\begin{aligned} \tilde{H}_{d,v,k} =& j + LCP\left(q[i+1,|q|], \sigma(v)[j+1,|\sigma(v)|]\right), \\ & j=\tilde{J}_{d,v,k},\ i=k+j \end{aligned}$$



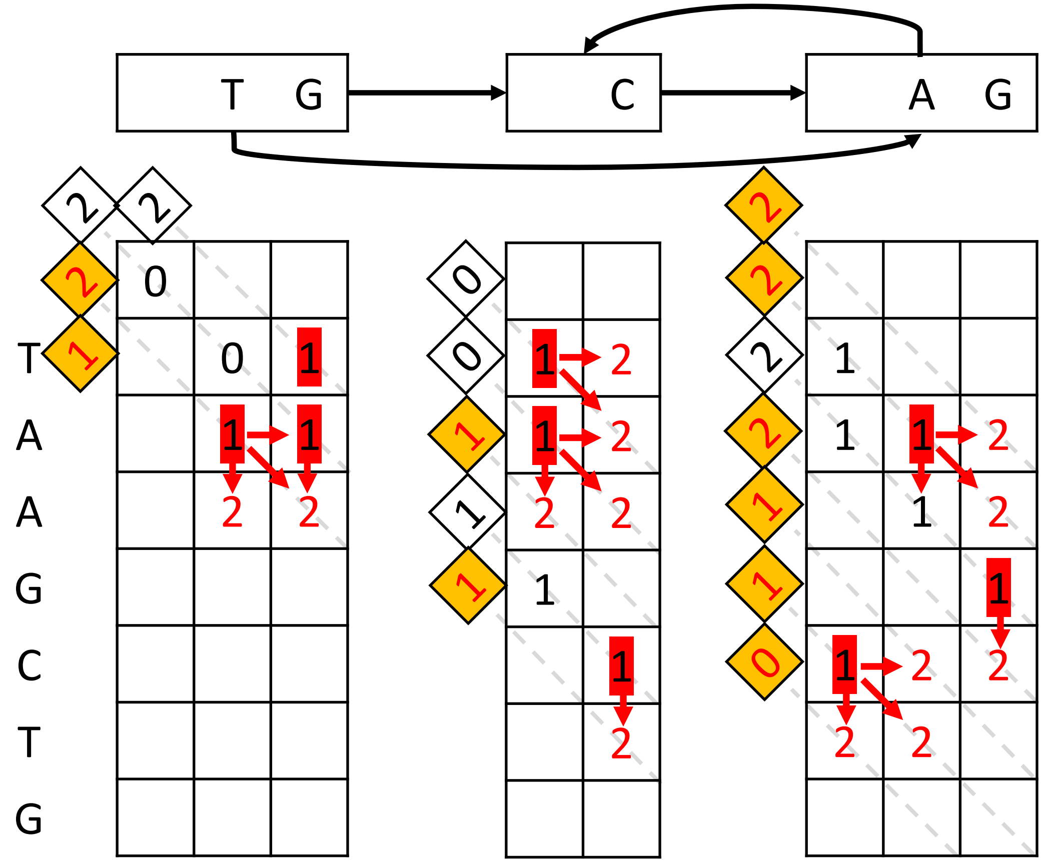
Edit distance of sequence to graph alignment

- DP recurrence to compute sequence-to-graph alignment

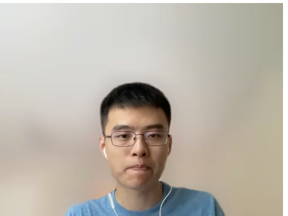
$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j} + 1, & i \geq 1 \\ H_{i,v,j-1} + 1, & j \geq 1 \\ H_{i-1,v,j-1} + \Delta_{i,v,j}, & i \geq 1, j \geq 1 \\ H_{i,u,|\sigma(u)|}, & j = 0, \forall u, (u,v) \in E \end{cases} \tag{1}$$

$$\tilde{J}_{d,v,k} = \max \begin{cases} \tilde{H}_{d-1,v,k-1} \\ \tilde{H}_{d-1,v,k+1} + 1 \\ \tilde{H}_{d-1,v,k} + 1 \\ 0, \exists u, (u,v) \in E, \tilde{H}_{d,u,k-|\sigma(u)|} = |\sigma(u)| \end{cases} \tag{2}$$

$$\begin{aligned} \tilde{H}_{d,v,k} =& j + LCP\left(q[i+1,|q|], \sigma(v)[j+1,|\sigma(v)|]\right), \\ & j = \tilde{J}_{d,v,k}, \; i = k + j \end{aligned}$$



Expansion step, d=2

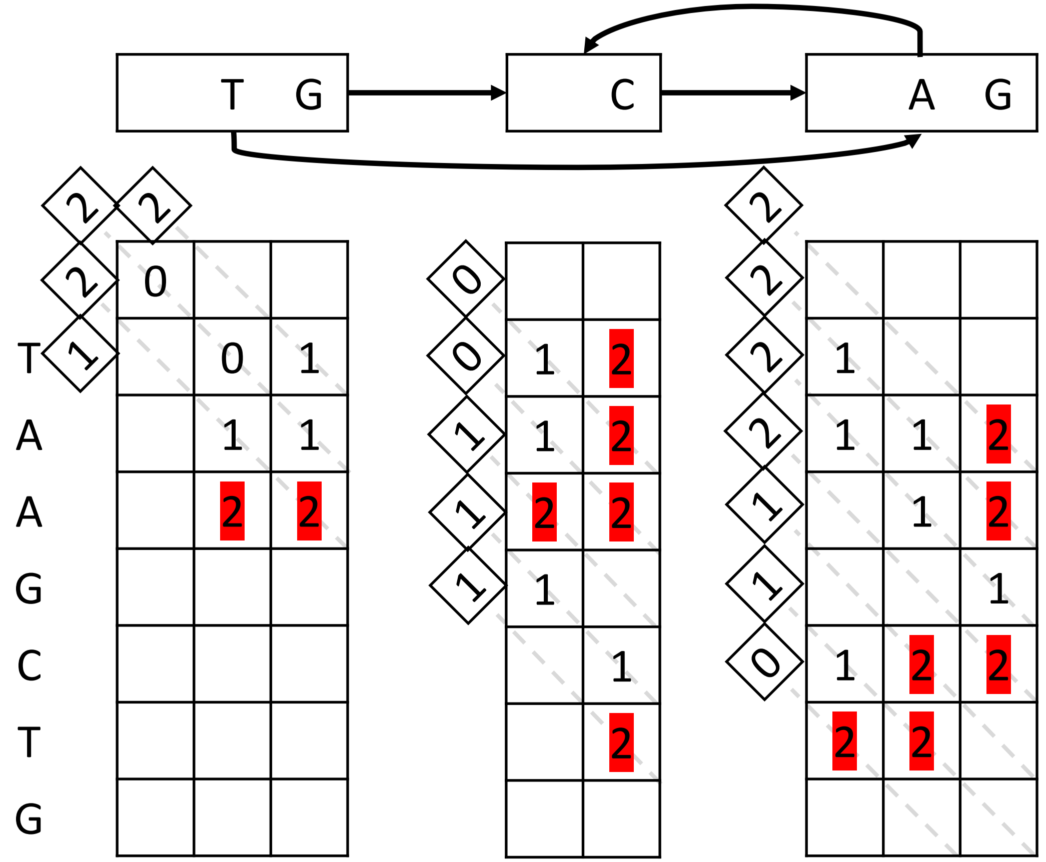


Edit distance of sequence to graph alignment

- DP recurrence to compute sequence-to-graph alignment

$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j}+1, & i \geq 1 \\ H_{i,v,j-1}+1, & j \geq 1 \\ H_{i-1,v,j-1}+\Delta_{i,v,j}, & i \geq 1, j \geq 1 \\ H_{i,u,|\sigma(u)|}, & j = 0, \forall u, (u,v) \in E \end{cases} \tag{1}$$

$$\begin{aligned} \tilde{J}_{d,v,k} &= \max \begin{cases} \tilde{H}_{d-1,v,k-1} \\ \tilde{H}_{d-1,v,k+1}+1 \\ \tilde{H}_{d-1,v,k}+1 \\ 0, \exists u, (u,v) \in E, \tilde{H}_{d,u,k-|\sigma(u)|} = |\sigma(u)| \end{cases} \\ \tilde{H}_{d,v,k} &= j + LCP\left(q[i+1,|q|], \sigma(v)[j+1,|\sigma(v)|]\right), \\ & j = \tilde{J}_{d,v,k},\ i = k+j \end{aligned} \tag{2}$$



d=2

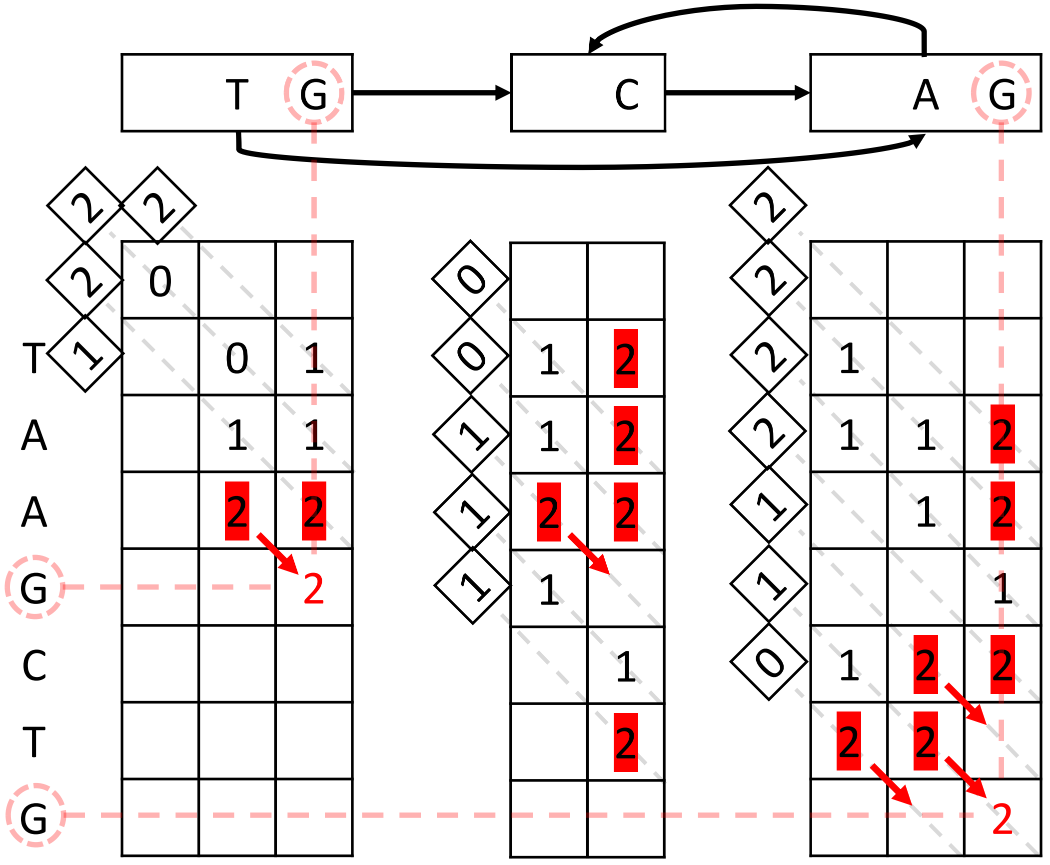
Edit distance of sequence to graph alignment

- DP recurrence to compute sequence-to-graph alignment

$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j} + 1, & i \geq 1 \\ H_{i,v,j-1} + 1, & j \geq 1 \\ H_{i-1,v,j-1} + \Delta_{i,v,j}, & i \geq 1, j \geq 1 \\ H_{i,u,|\sigma(u)|}, & j = 0, \forall u, (u,v) \in E \end{cases} \tag{1}$$

$$\tilde{J}_{d,v,k} = \max \begin{cases} \tilde{H}_{d-1,v,k-1} \\ \tilde{H}_{d-1,v,k+1} + 1 \\ \tilde{H}_{d-1,v,k} + 1 \\ 0, \exists u, (u,v) \in E, \tilde{H}_{d,u,k-|\sigma(u)|} = |\sigma(u)| \end{cases} \tag{2}$$

$$\begin{aligned} \tilde{H}_{d,v,k} =& j + LCP\left(q[i+1,|q|], \sigma(v)[j+1,|\sigma(v)|]\right), \\ & j = \tilde{J}_{d,v,k},\ i = k + j \end{aligned}$$



Extension step, d=2

Edit distance of sequence to graph alignment

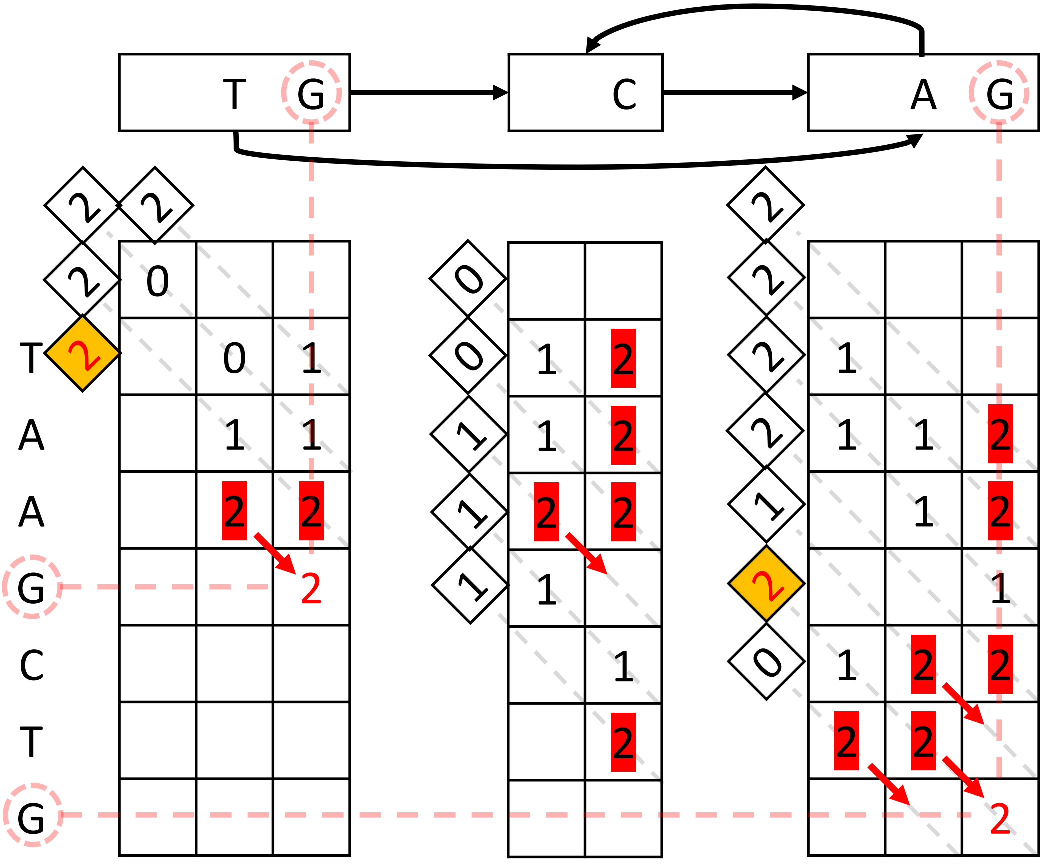
- DP recurrence to compute sequence-to-graph alignment

$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j}+1, & i \ge 1 \\ H_{i,v,j-1}+1, & j \ge 1 \\ H_{i-1,v,j-1}+\Delta_{i,v,j}, & i \ge 1, j \ge 1 \\ H_{i,u,|\sigma(u)|}, & j = 0, \forall u, (u,v) \in E \end{cases} \qquad (1)$$

$$\tilde{J}_{d,v,k} = \max \begin{cases} \tilde{H}_{d-1,v,k-1} \\ \tilde{H}_{d-1,v,k+1}+1 \\ \tilde{H}_{d-1,v,k}+1 \\ 0, \exists u, (u,v) \in E, \tilde{H}_{d,u,k-|\sigma(u)|} = |\sigma(u)| \end{cases} \qquad (2)$$

$$\tilde{H}_{d,v,k} = j + LCP\left(q[i+1,|q|], \sigma(v)[j+1,|\sigma(v)|]\right),$$

$$j = \tilde{J}_{d,v,k},\ i = k + j$$



	0									
T		0	1		1	2		1		
A		1	1		1	2		1	1	2
A		2	2		2	2			1	2
G			2		1					1
C						1		1	2	2
T						2		2	2	
G										2

Extension step, d=2

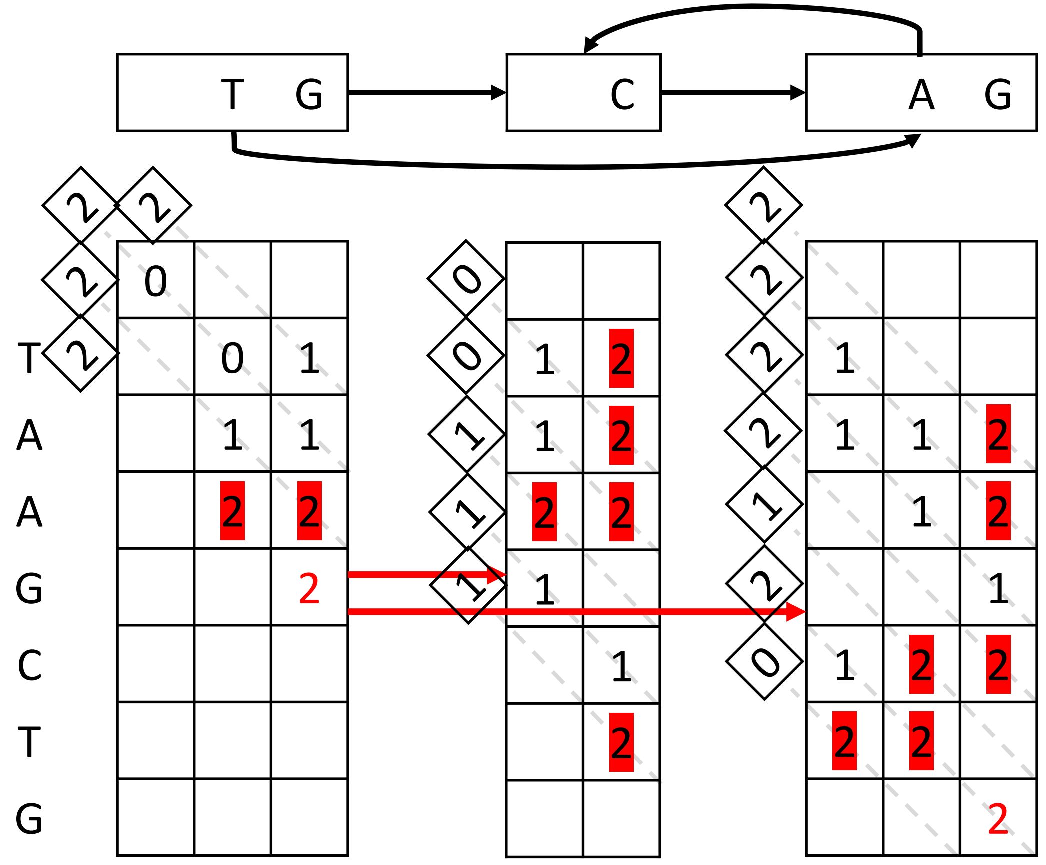
Edit distance of sequence to graph alignment

- DP recurrence to compute sequence-to-graph alignment

$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j}+1, & i \geq 1 \\ H_{i,v,j-1}+1, & j \geq 1 \\ H_{i-1,v,j-1}+\Delta_{i,v,j}, & i \geq 1, j \geq 1 \\ H_{i,u,|\sigma(u)|}, & j = 0, \forall u, (u,v) \in E \end{cases} \qquad (1)$$

$$\tilde{J}_{d,v,k} = \max \begin{cases} \tilde{H}_{d-1,v,k-1} \\ \tilde{H}_{d-1,v,k+1}+1 \\ \tilde{H}_{d-1,v,k}+1 \\ 0, \exists u, (u,v) \in E, \tilde{H}_{d,u,k-|\sigma(u)|} = |\sigma(u)| \end{cases} \qquad (2)$$

$$\tilde{H}_{d,v,k} = j + LCP\left(q[i+1,|q|], \sigma(v)[j+1,|\sigma(v)|]\right), \quad j = \tilde{J}_{d,v,k},\ i = k+j$$



Extension step, d=2

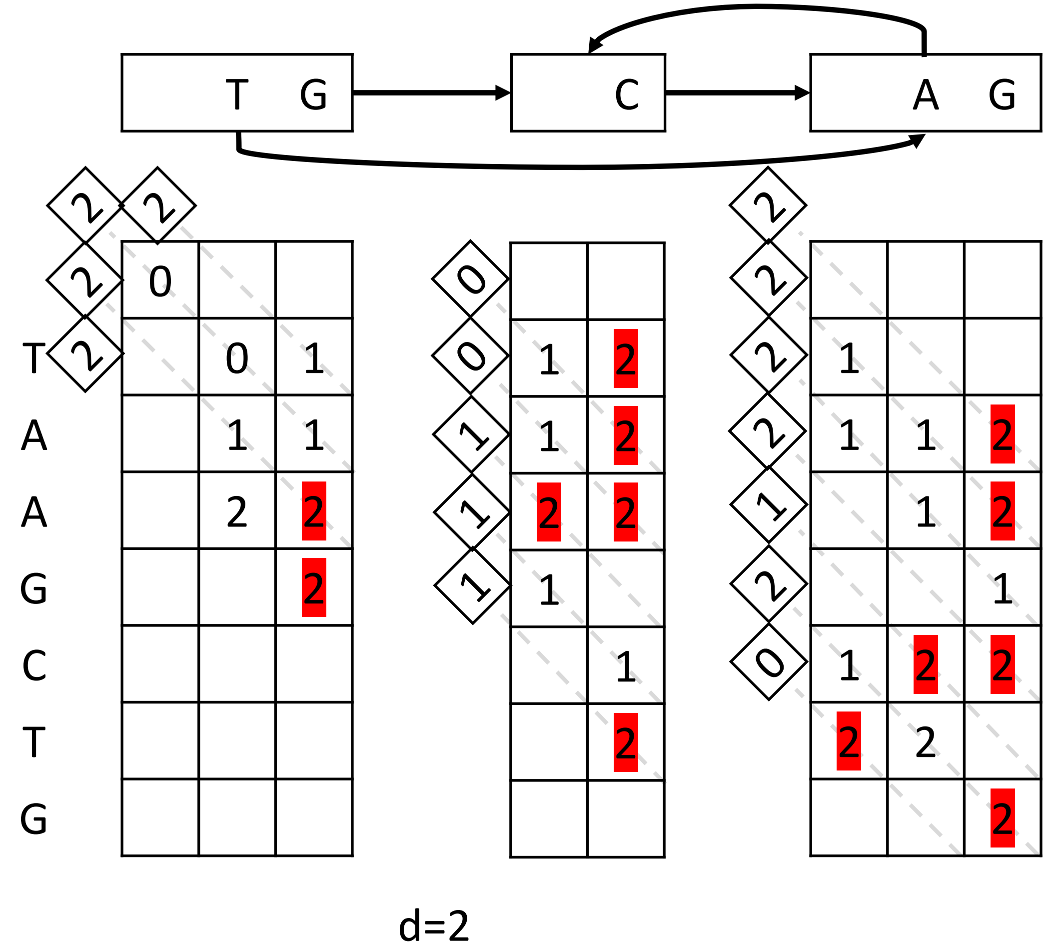
Edit distance of sequence to graph alignment

- DP recurrence to compute sequence-to-graph alignment

$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j} + 1, & i \geq 1 \\ H_{i,v,j-1} + 1, & j \geq 1 \\ H_{i-1,v,j-1} + \Delta_{i,v,j}, & i \geq 1, j \geq 1 \\ H_{i,u,|\sigma(u)|}, & j = 0, \forall u, (u,v) \in E \end{cases} \tag{1}$$

$$\tilde{J}_{d,v,k} = \max \begin{cases} \tilde{H}_{d-1,v,k-1} \\ \tilde{H}_{d-1,v,k+1} + 1 \\ \tilde{H}_{d-1,v,k} + 1 \\ 0, \exists u, (u,v) \in E, \tilde{H}_{d,u,k-|\sigma(u)|} = |\sigma(u)| \end{cases} \tag{2}$$

$$\tilde{H}_{d,v,k} = j + LCP\left(q[i+1,|q|], \sigma(v)[j+1,|\sigma(v)|]\right), \\ j = \tilde{J}_{d,v,k}, \; i = k + j$$



	0		
T		0	1
A		1	1
A		2	2
G			2
C			
T			
G			

T	1	2
A	1	2
A	2	2
G	1	
C		1
T		2
G		

T	1		
A	1	1	2
A		1	2
G			1
C	1	2	2
T	2	2	
G			2

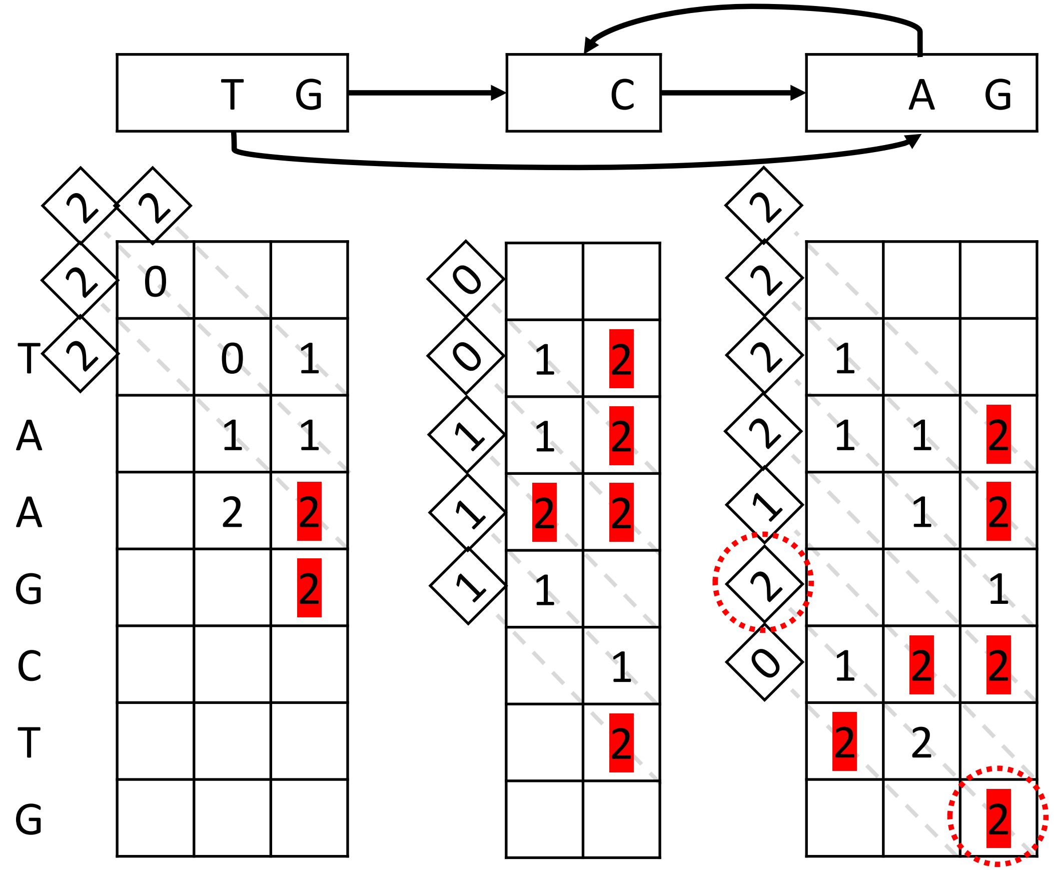
Edit distance of sequence to graph alignment

- DP recurrence to compute sequence-to-graph alignment

$$H_{i,v,j} = \min \begin{cases} H_{i-1,v,j} + 1, & i \geq 1 \\ H_{i,v,j-1} + 1, & j \geq 1 \\ H_{i-1,v,j-1} + \Delta_{i,v,j}, & i \geq 1, j \geq 1 \\ H_{i,u,|\sigma(u)|}, & j = 0, \forall u, (u,v) \in E \end{cases} \tag{1}$$

$$\tilde{J}_{d,v,k} = \max \begin{cases} \tilde{H}_{d-1,v,k-1} \\ \tilde{H}_{d-1,v,k+1} + 1 \\ \tilde{H}_{d-1,v,k} + 1 \\ 0, \exists u, (u,v) \in E, \tilde{H}_{d,u,k-|\sigma(u)|} = |\sigma(u)| \end{cases} \tag{2}$$

$$\tilde{H}_{d,v,k} = j + LCP\left(q[i+1, |q|], \sigma(v)[j+1, |\sigma(v)|]\right),$$
$$j = \tilde{J}_{d,v,k}, \; i = k + j$$



	0		
T		0	1
A		1	1
A		2	2
G			2
C			
T			
G			

T	1	2
A	1	2
A	2	2
G	1	
C		1
T		2
G		

T	1		
A	1	1	2
A		1	2
G			1
C	1	2	2
T	2	2	
G			2

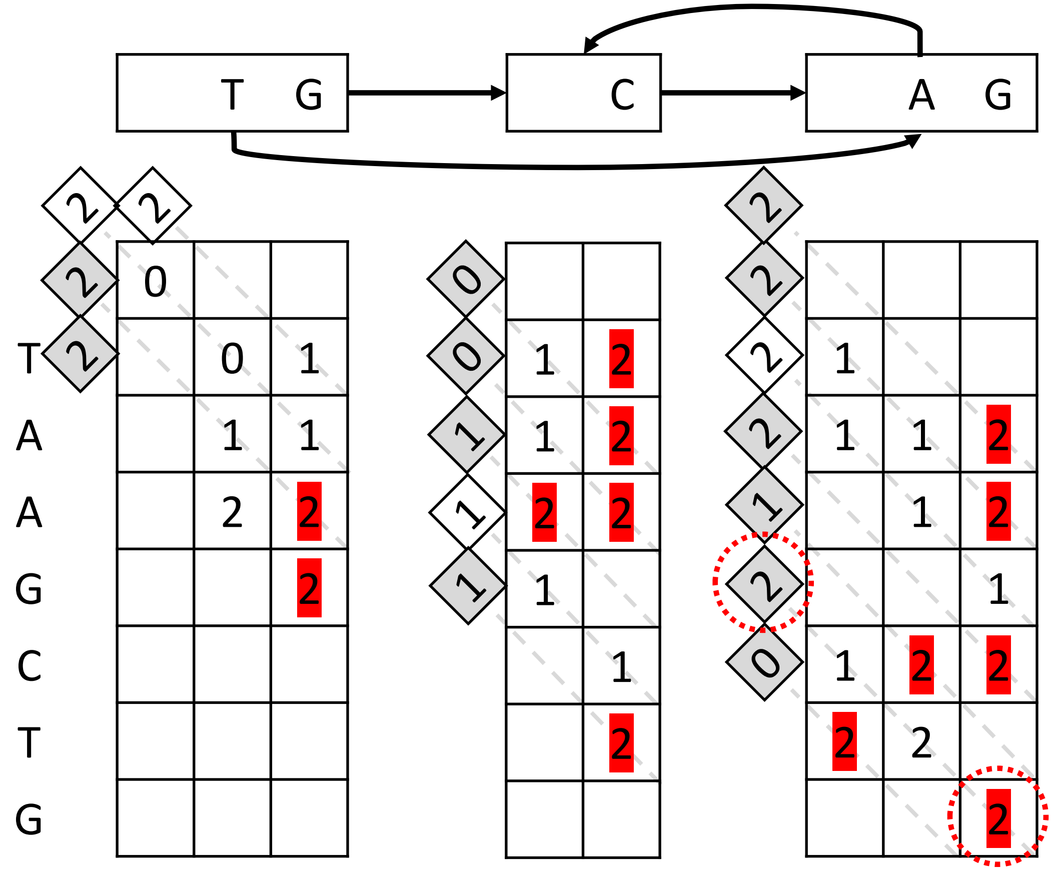
Return min cost, d=2

Edit distance of sequence to graph alignment

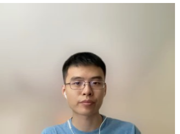
- We use an array to keep the offset of the furthest cell and a queue to keep the diagonals that contain the furthest cells with d

Algorithm 1: Graph wavefront algorithm to find the optimal global sequence to graph alignment

```
Input: Query sequence $q$, sequence graph $G = (V, E, \sigma)$, start vertex $v_s \in V$ and
       end vertex $v_e \in V$.
function GwfEditDist($q, G, v_s, v_e$) begin
    $k_e \leftarrow |q| - |v_e|$
    $\tilde{H}_{v_s,0} \leftarrow 0$
    $Q \leftarrow [(v_s, 0)]$
    $d \leftarrow 0$
    while true do
        GwfExtend($q, G, Q, \tilde{H}$)
        if $\tilde{H}_{v_e,k_e} = |v_e|$ then
            return $d$
        $d \leftarrow d + 1$
        GwfExpand($q, G, Q, \tilde{H}$)
```



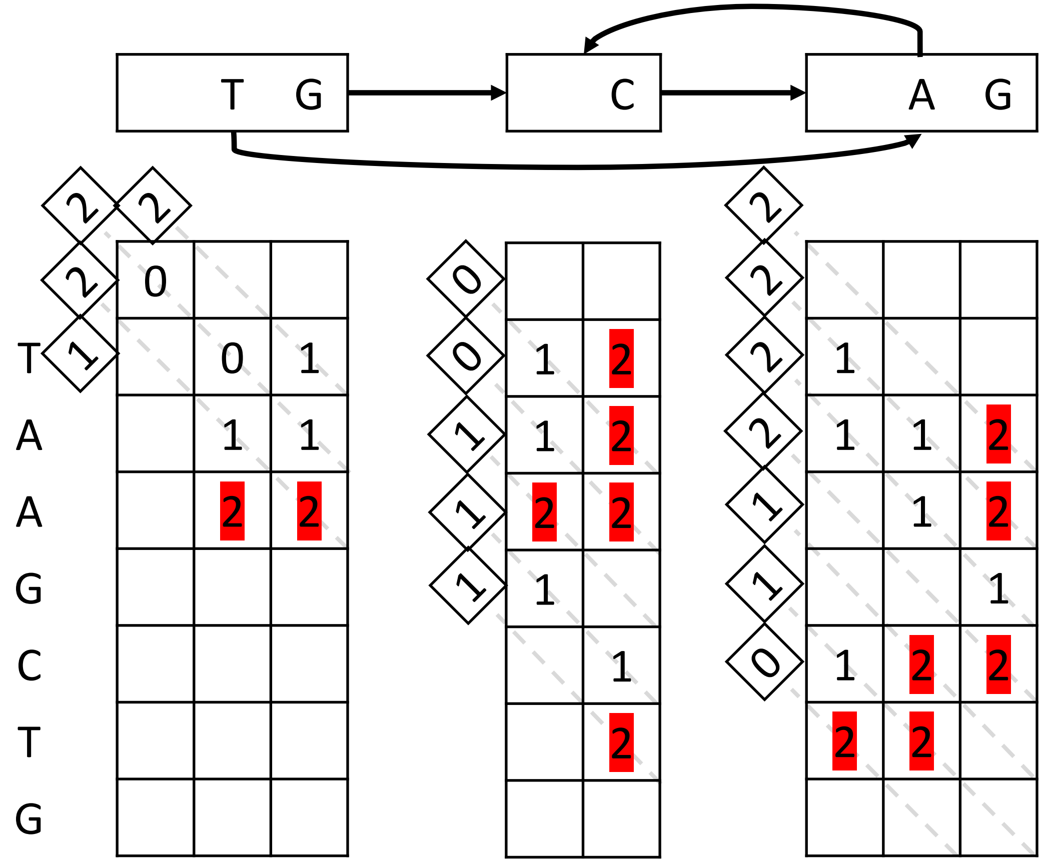
Return min cost, d=2



Graph wavefront pruning

- Find the max sum of aligned query length and graph walk length

- Prune the wavefront left behind



	0		
T		0	1
A		1	1
A		2	2
G			
C			
T			
G			

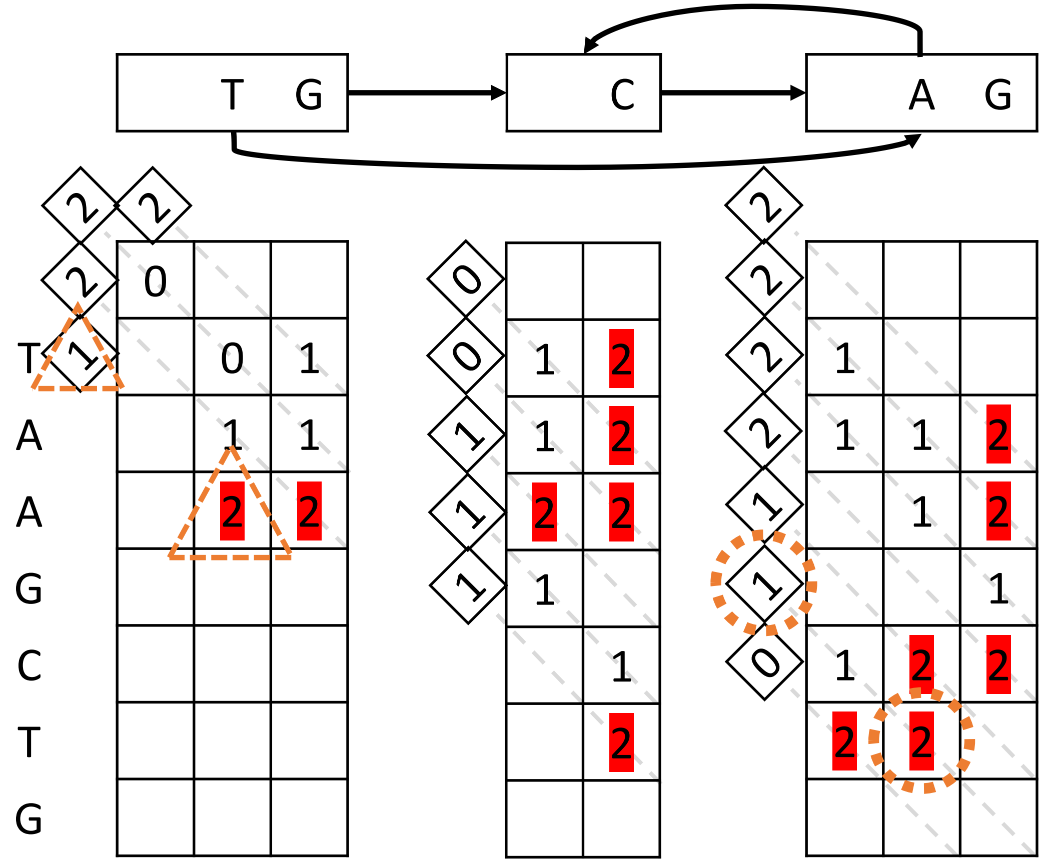
T	1	2
A	1	2
A	2	2
G	1	
C		1
T		2
G		

T	1		
A	1	1	2
A		1	2
G			1
C	1	2	2
T	2	2	
G			

Pruning, d=2

Graph wavefront pruning

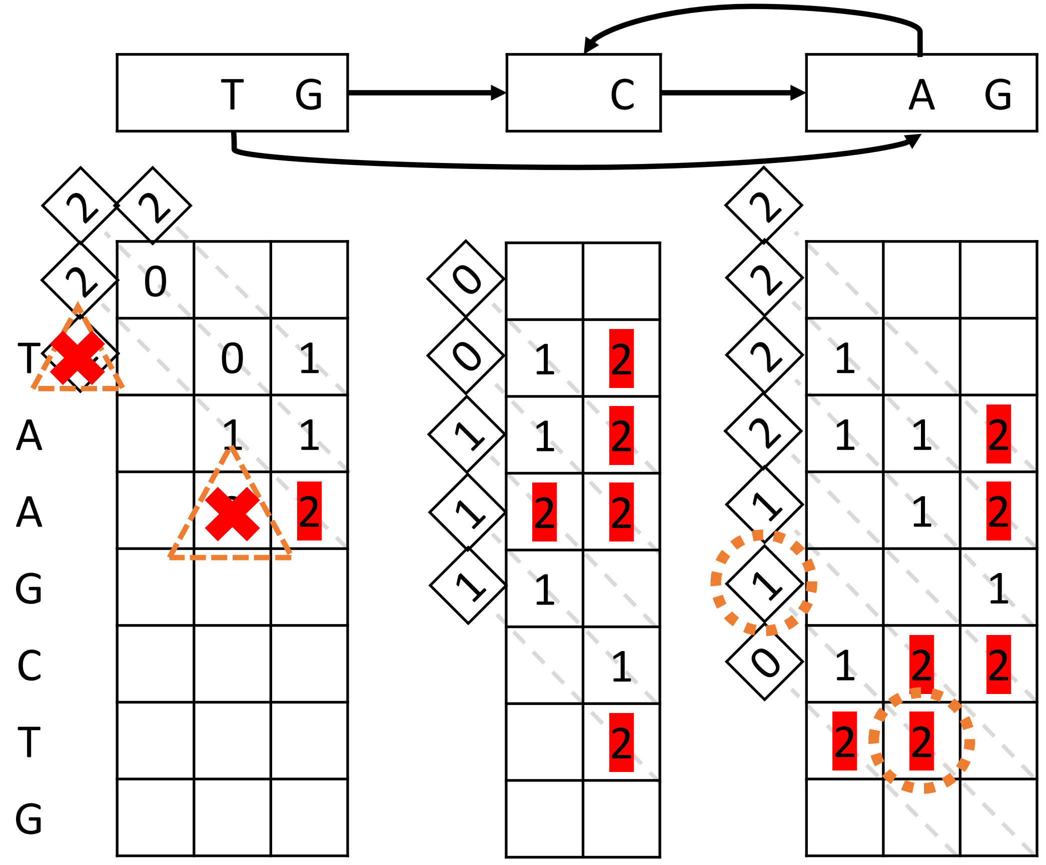
- Find the max sum of aligned query length and graph walk length
  - ○ 6 + 6 = 12
- Prune the wavefront left behind
  - △ 1 + 3 = 4 << 12



Pruning, d=2

Graph wavefront pruning

- Find the max sum of aligned query length and graph walk length
  - ○ 6 + 6 = 12
- Prune the wavefront left behind
  - △ 1 + 3 = 4 << 12



Pruning, d=2