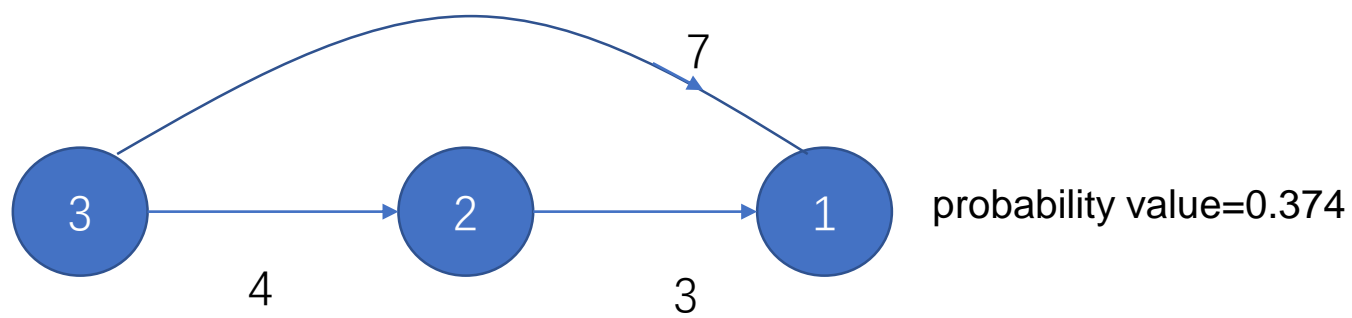
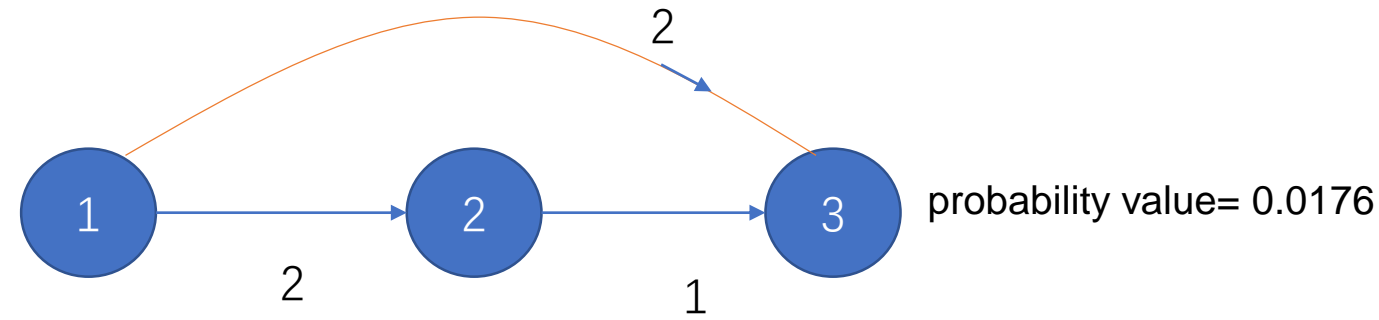


Most likely Order

For each order, a probability can be calculated, the path with highest likelihood is the most likely order (MLO)..

	1	2	3
1	0	2	2
2	3	0	1
3	7	4	0

Every order can form a DAG



$$P(2 \text{ spliced before } 1) = \frac{3}{3+2} = 0.6$$
$$P(1 \text{ spliced before } 2) = \frac{2}{3+2} = 0.4$$
$$P(3 \text{ spliced before } 1) = \frac{7}{7+2} = 0.78$$
$$P(1 \text{ spliced before } 3) = \frac{2}{7+2} = 0.22$$
$$P(3 \text{ spliced before } 2) = \frac{4}{4+1} = 0.8$$
$$P(2 \text{ spliced before } 3) = \frac{1}{4+1} = 0.2$$

p-value of in order splicing
Likelihood ratio test
 $-2 \cdot \log(0.0176 / 0.374) \sim X, df=n \cdot (n-1) / 2$

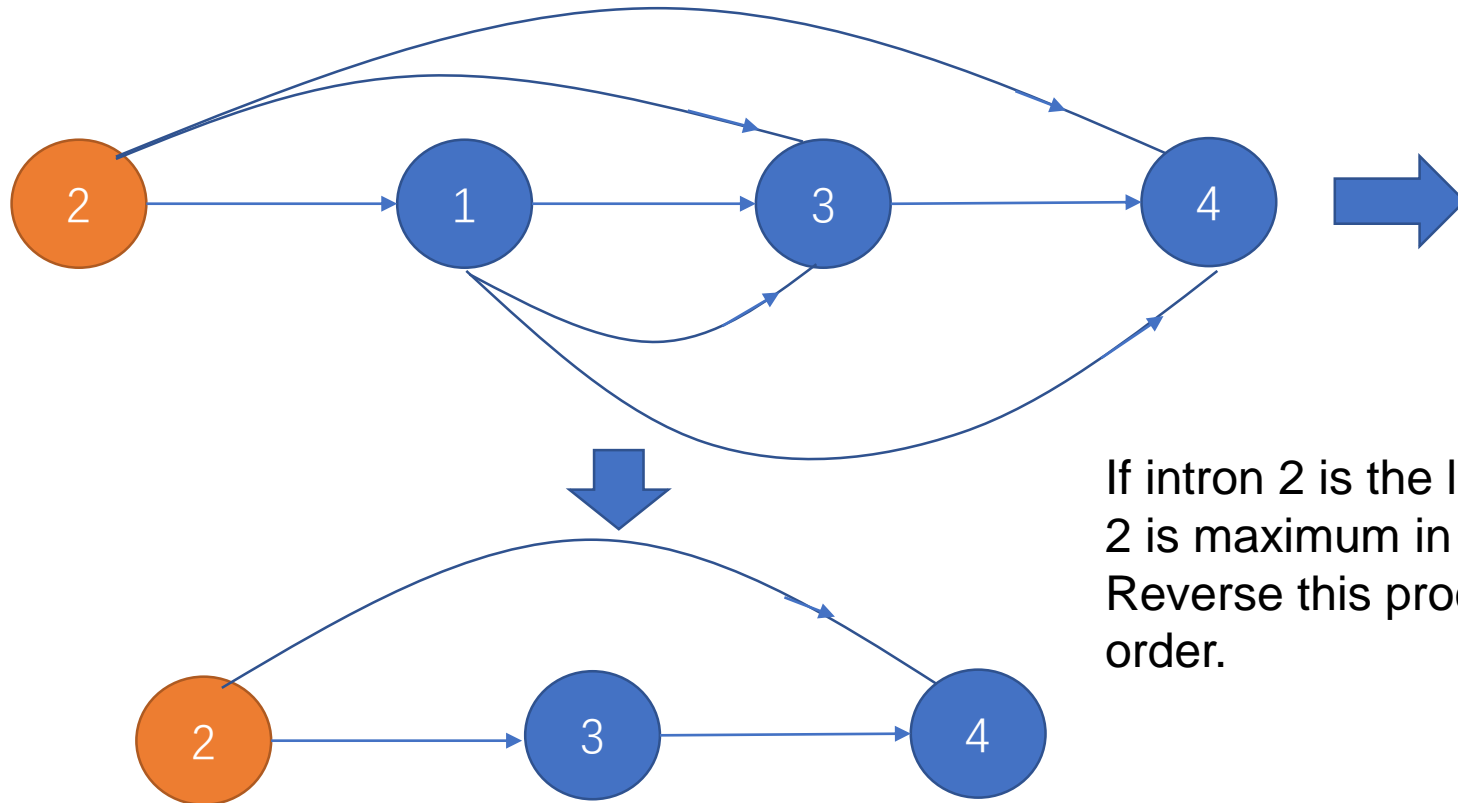
P-value=0.89

Most likely Order

Find MLO is $N!$, difficult for intron number > 10

*Row i represent the frequency of each intron spliced after i .

*Column i represent the frequency of each intron spliced before i .



	1	2	3	4
1	0	0	1	1
2	1	0	1	1
3	0	0	0	1
4	0	0	0	0

If intron 2 is the left node (first spliced), then the sum of Row 2 is maximum in the right.

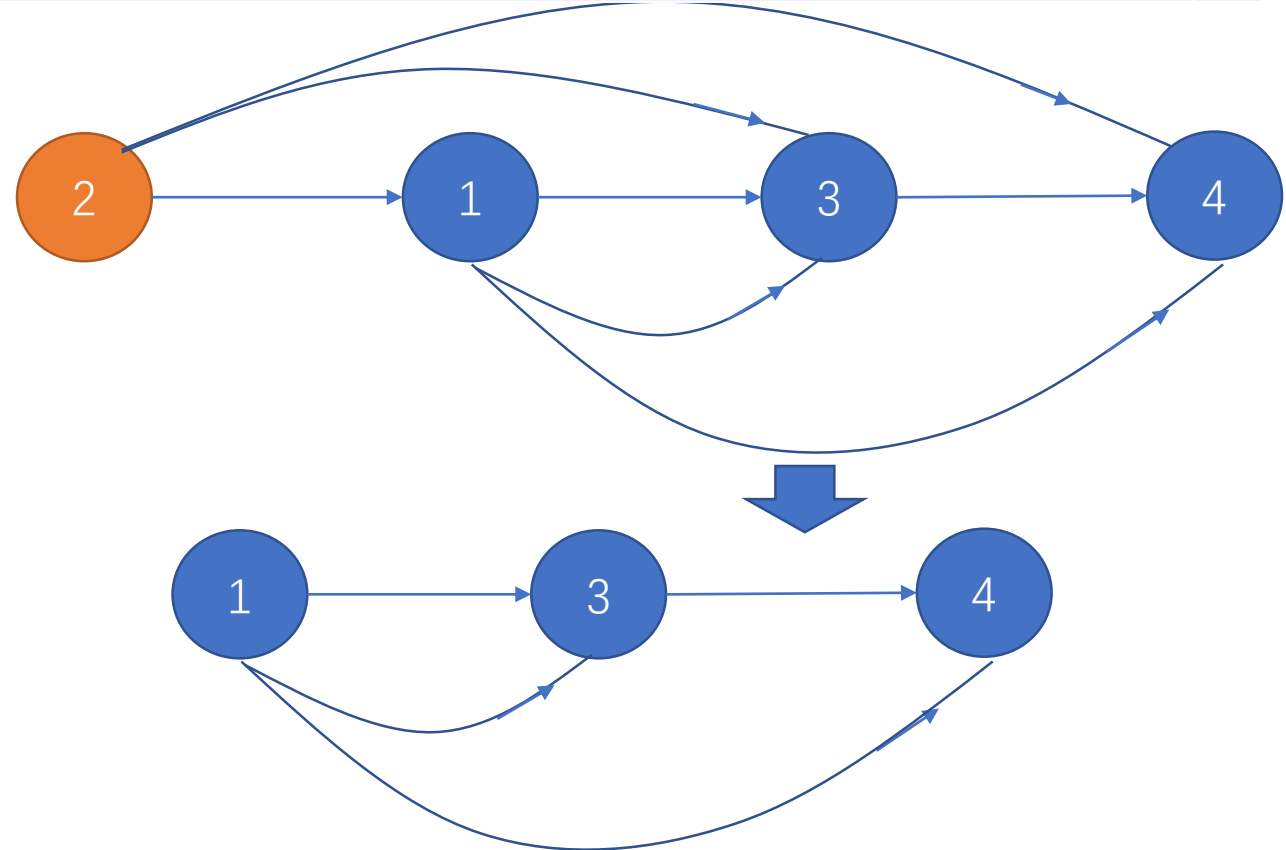
Reverse this process we can get an algorithm to estimate the order.

Remove one node doesn't change the order of the rest.

Correspond to remove row i and column i doesn't change the order when opti rest in matrix,
Thus make iteration possible

Most likely Order

For each isoform got the frequency matrix D as above described.	1
Find potential most left intron using D, by which $\max(\text{rowsum}) = q_1$	2
Find 2nd, 3rd, 4th maximum rowsum row number: q_2, q_3, q_4 .	3
For i in vector (q_1, q_2, q_3, q_4)	4
Remove column i and row i in D and repeat step 1.	5
Stop if the number of orders calculated reach 5,000.	6
After got all the potential best orders, calculate probability for each order and return the best.	7



D

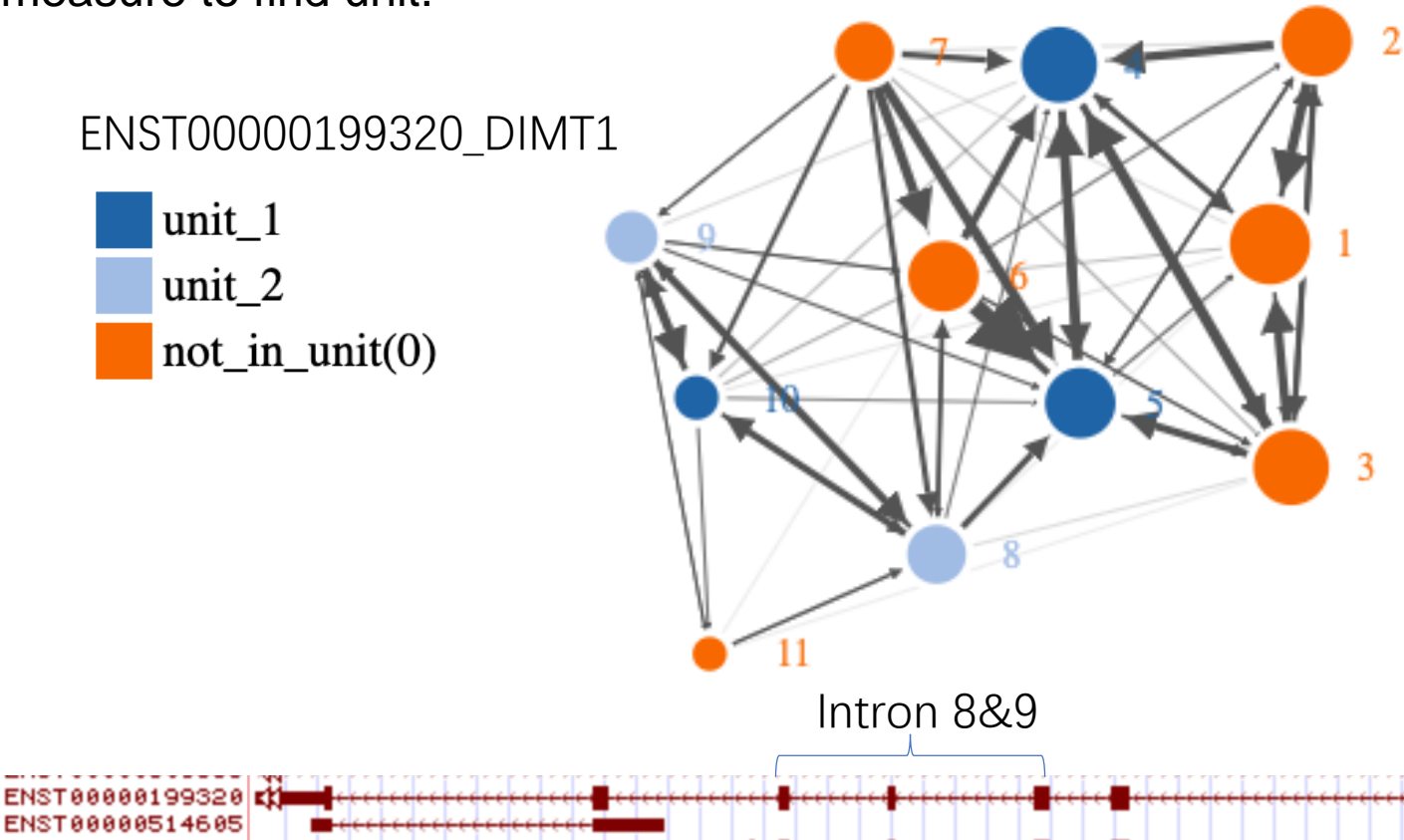
	1	2	3	4
1	0	0.1	0.6	0
2	0.9	0	0.4	0.9
3	0.4	0.6	0	0
4	0	0.1	0	0

remove intron 2

	1	3	4
1		0.6	
3	0.4		
4			

Splicing unit

Some introns tends to spliced simultaneously.
Two introns are spliced together, if their spliced rate relative to other introns are similar.
i.e. column i and column j are correlated or row j and row i are correlated.
We used DBscan algorithm with pearson correlation as distance measure to find unit.



row 2 and row 3 are similar, means their splicing rate are similar relative to others.
Thus they have higher probability tends to spliced together.
Diagonal filled with 0.5

	1	2	3	4
1	0.5	0.1	0.2	0.9
2	0.9	0.5	0.4	0.3
3	0.8	0.6	0.5	0.3
4	0.1	0.7	0.7	0.5