De Bruijn Graph (DBG) Assembly

(find and work with a partner)

1. Given the 4 bases (A, C, G, T) and a positive integer $k$, how many $k$-mers exist?

2. Given a genome of length $G$, what is the maximum number of unique $k$-mers present in the genome?

3. List all the $k$-mers of the string $S = ZABCDABEFABY$, for $k = 3$.

4. Draw the de Bruijn graph for the given $S$ and $k$ above.

5. How many Eulerian walks exist for the de Bruijn graph above? For each walk, write down the resulting string (output assembly).

6. Can I have a graph with just one semi-balanced node? Why or why not?

7. To form a graph with an Eulerian cycle, we can draw an edge between the two semi-balanced nodes so that each node is now balanced. To find an Eulerian cycle, note that if the edges of any cycle are removed from the graph, the resulting connected component(s) also have Eulerian cycles. Why is that?