22BIO211: Intelligence of Biological Systems - 2

Lab Sheet 3

1. Solve the following Rosalind Challenge : https://rosalind.info/problems/ba3e/and upload the 'strand' page.

   Given an arbitrary collection of k-mers Patterns (where some k-mers may appear multiple times), we define CompositionGraph(Patterns) as a graph with |Patterns| isolated edges. Every edge is labeled by a k-mer from Patterns, and the starting and ending nodes of an edge are labeled by the prefix and suffix of the k-mer labeling that edge. We then define the de Bruijn graph of Patterns, denoted DeBruijn(Patterns), by gluing identically labeled nodes in CompositionGraph(Patterns), which yields the following algorithm.

   DEBRUIJN(Patterns)
   represent every k-mer in Patterns as an isolated edge between its prefix and suffix
   glue all nodes with identical labels, yielding the graph DeBruijn(Patterns)
   return DeBruijn(Patterns)

   De Bruijn Graph from k-mers Problem

   Construct the de Bruijn graph from a collection of k-mers.

   Given: A collection of k-mers Patterns.

   Return: The de Bruijn graph DeBruijn(Patterns), in the form of an adjacency list.

   Sample Dataset

   GAGG

   CAGG

   GGGG

   GGGA

   CAGG

   AGGG

GGAG

Sample Output

AGG -> GGG

CAG -> AGG,AGG

GAG -> AGG

GGA -> GAG

GGG -> GGA,GGG

2. Given any DNA Sequence, generate the the K-mers and Construct the de Bruijn Graph. Visualize the de Bruijn Graph using 'Networkx'
3. Solve the following Rosalind Challenge : https://rosalind.info/problems/ba3f/and upload the 'strand' page.

Find Eulerian Cycle in a graph

A cycle that traverses each edge of a graph exactly once is called an Eulerian cycle, and we say that a graph containing such a cycle is Eulerian. The following algorithm constructs an Eulerian cycle in an arbitrary directed graph.

```
EULERIANCYCLE(Graph)
    form a cycle Cycle by randomly walking in Graph (don't visit the same edge twice!)
    while there are unexplored edges in Graph
        select a node newStart in Cycle with still unexplored edges
        form Cycle' by traversing Cycle (starting at newStart) and then randomly walking
        Cycle ← Cycle'
    return Cycle
```

Eulerian Cycle Problem

Find an Eulerian cycle in a graph.

Given: An Eulerian directed graph, in the form of an adjacency list.

Return: An Eulerian cycle in this graph.

Sample Dataset

0 -> 3

1 -> 0

2 -> 1,6

3 -> 2

4 -> 2

5 -> 4

6 -> 5,8

7 -> 9

8 -> 7

9 -> 6

Sample Output

6->8->7->9->6->5->4->2->1->0->3->2->6