COM SCI C121 Week 5

Aidan Jan

May 2, 2024

De Brujin Graphs Review

To make from read:

1. Sample every 3-mer from the read.
2. Sample Left and Right 2-mers from each 3-mer.
3. On the graph, each 2-mer is a node, and 3-mers are the links between the nodes of the left and right 2-mers

We cannot go back from the De Brujin to the aligned genome. However, it is important to note that some Eulerian path (path that uses every link exactly once) would produce the original read.

Eulerian Walk Definitions and Statements

- Node is *balanced* if indegree equals outdegree
- Node is *semi-balanced* if indegree differs from outdegree by 1
- Graph is *connected* if each node can be reached by some other node
- *Eulerian walk* visits each edge exactly once
- Not all graphs hae Eulerian walks. Graphs that do are *Eulerian*.
- A directed, connected graph is Eulerian if and only if it has at most 2 semi-balanced nodes and all other nodes are balanced.

Attempt 2: Build the T-DBG

Consider the example:

```
         ACATACAT---ACA
RED      ########---###
GREEN    ########---###
BLUE     #####------###
```

Where # denotes the strand having the base, and − denoting the base is absent on that strand.
For the example last week, we built a graph that was straightforward.
However, notice that some nodes were repeated this time.
ACA CAT ATA TAC