Elijah Andrushenko

CPTS 453

Graph Theory

11-27-2019

<div align="center">Project Description</div>

Title

Graph Theory in Bioinformatics

Abstract

Genome Sequencing is an important subject in bioinformatics. Being able to match sequences and recognize patterns is an important tool and it gives researchers valuable information in biology. There are 24 chromosomes in the human body having around 3 million base pairs in total from those chromosomes. Due to the high number of base pairs it is computationally expensive in both memory storage and time complexity. To make comparing genes more efficient applications in graph theory are used to sequence and compare genomes. Some applications include the Needleman-Wunsch Algorithm which is a matrix that allows two sequences to be scored based on how similar they may be. Another application would be suffix trees which is a way to compare two sequences in a memory efficient way. How these algorithms work will be discussed in the paper.

Brief Outline

The introduction will cover basic graph theory concepts that will be discussed later such as graphs, matrices, and trees. The paper will go over important terms and how graph theory tools apply to genome sequencing and bioinformatics. Section 2 will go over matrices in bioinformatics specifically the Needleman-Wunsch Algorithm. How to construct the matrix and how to use the matrix to compare two genome sequences. The section will go over how the Needleman-Wunsch Algorithm applies in fields outside of bioinformatics and the problems it solves. Section 3 will go over suffix trees and other variants of suffix trees and how they are used in genome sequencing and in other fields. The conclusion will summarize the algorithms used in bioinformatics thanks to graph theory and will suggest other graph theory models that can be applied to bioinformatics and genome sequencing.