Tyler Thatcher

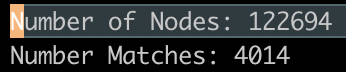
Homework 6

12/03/2018

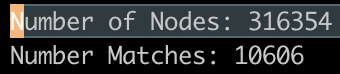
Homework 6 Write-Up

**Question A:**

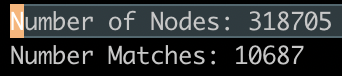
* **5k**

****

* **50k**

****

* **100k**

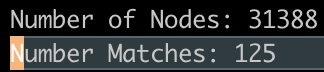
****

1. This doesn’t make sense, I expected to find all 500, but I understand that because these sequences can duplicate which means you will have less matches than the total number of sequences.

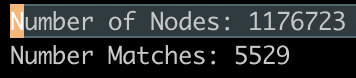
2. The number of matches also increases but levels out around 10k because that is how many unique sequences.

**Question B:**

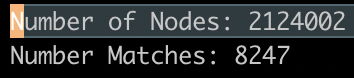
* **1k**

****

* **50k**

****

* **100k**

****

1. The size of nodes increases so drastically because of the amount of unique or not shared nodes increases when we introduce the error rates.
2. The number of matches is so little because of the 5% error rate, so a lot of them are unique sequences that will not be found because it is not in the file.