# CPSC 319 HW5 Report

## Mitchell Sawatzky

The size of the table was determined from the load factor by using the formula for load factor:

Rearranging, we get

The *elements in table* can be found by counting every word in the input file.

While (fileScanner.hasNext()) {

fileScanner.next();

wordCount++;

}

After getting this minimum table size, all we need to do is find a prime number equal to or larger than this number. A number is **not prime** if it is even or if it is divisible by another number that isn’t 1 or itself.

if (number % 2 == 0) {

// number is not prime

}

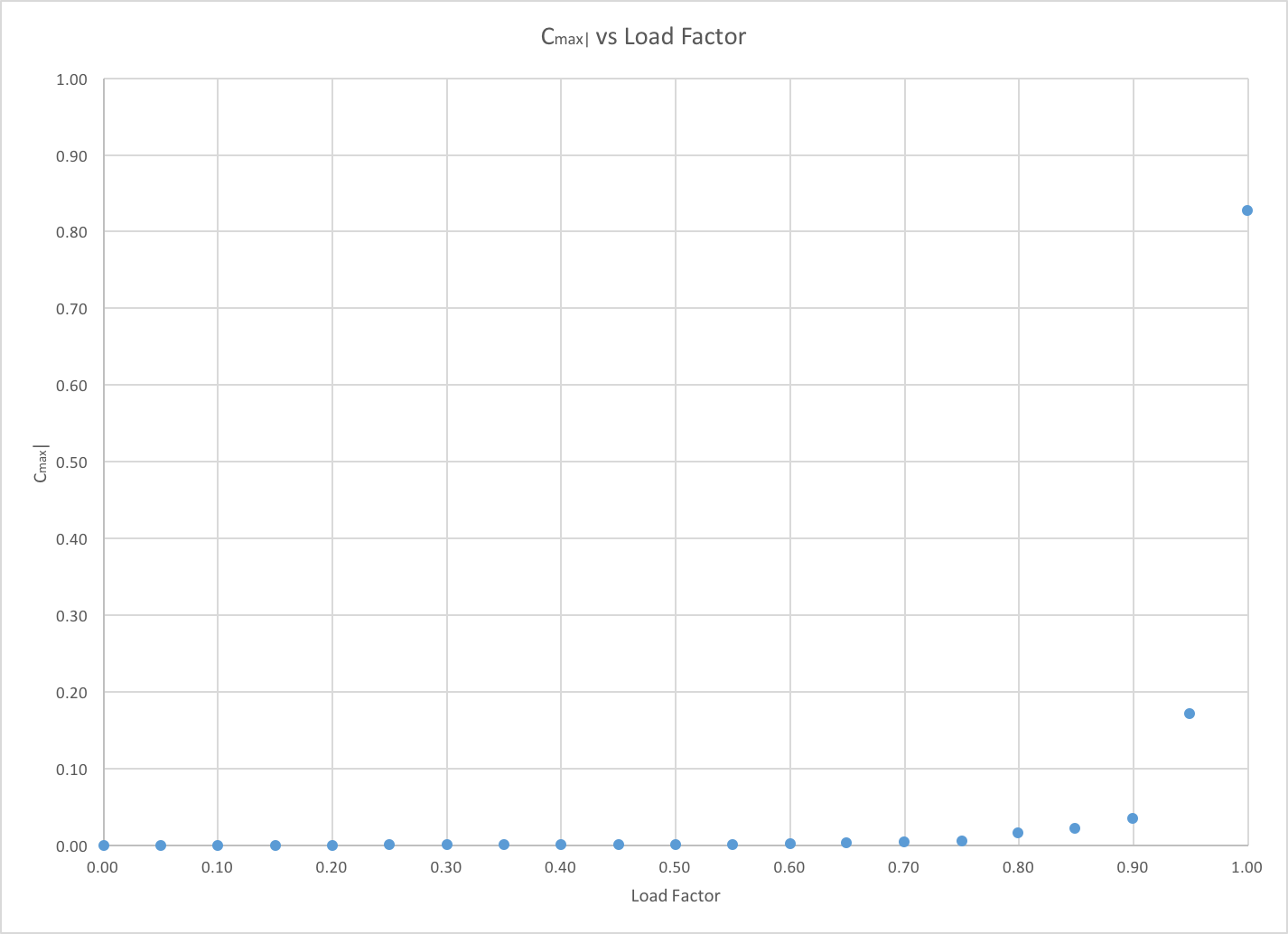
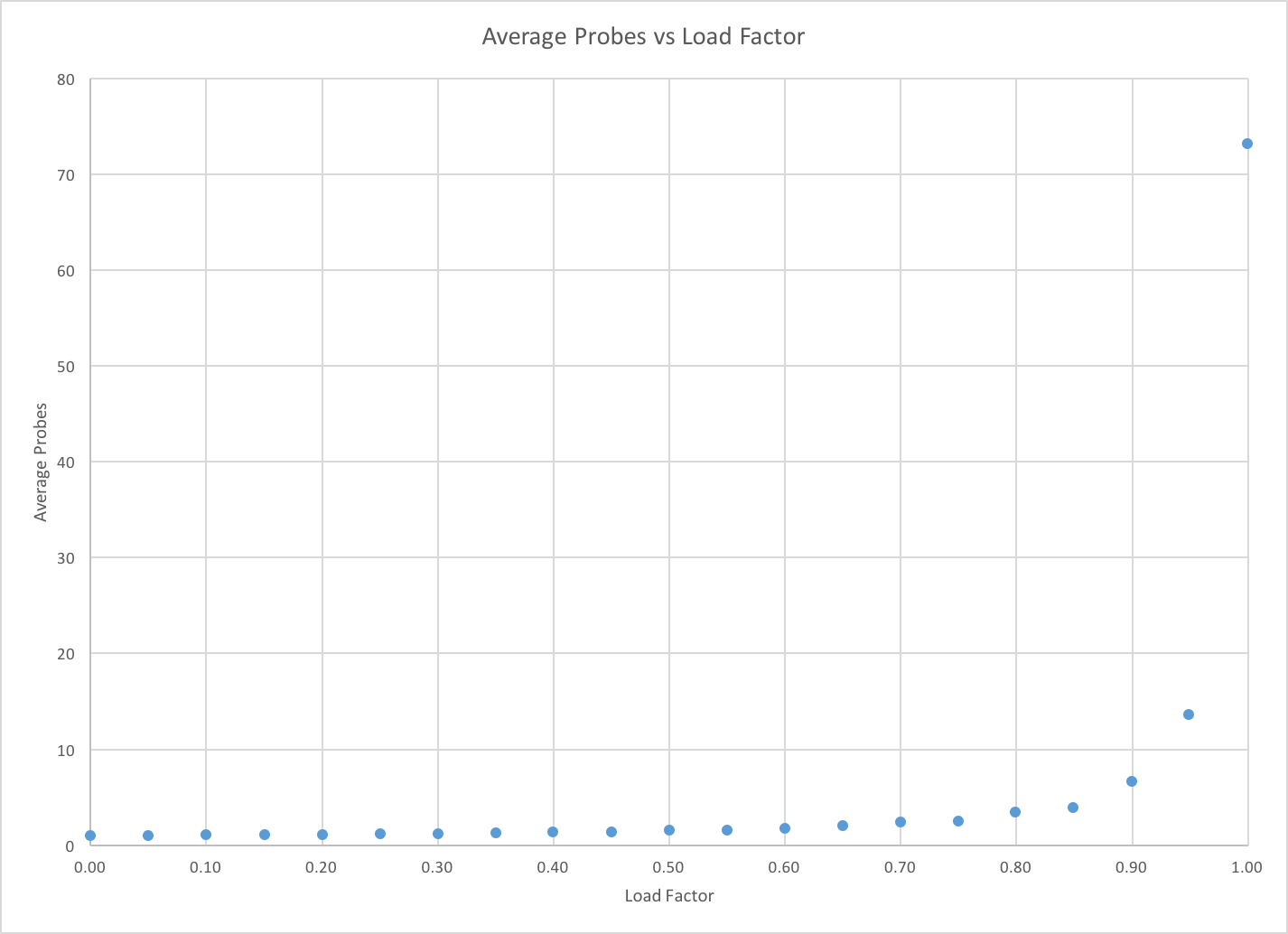
for(int i = 3; i \* i <= number; i += 2) {

if (number % i == 0) {

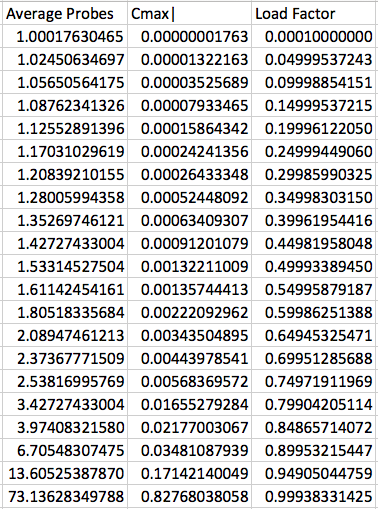
// number is not prime

}

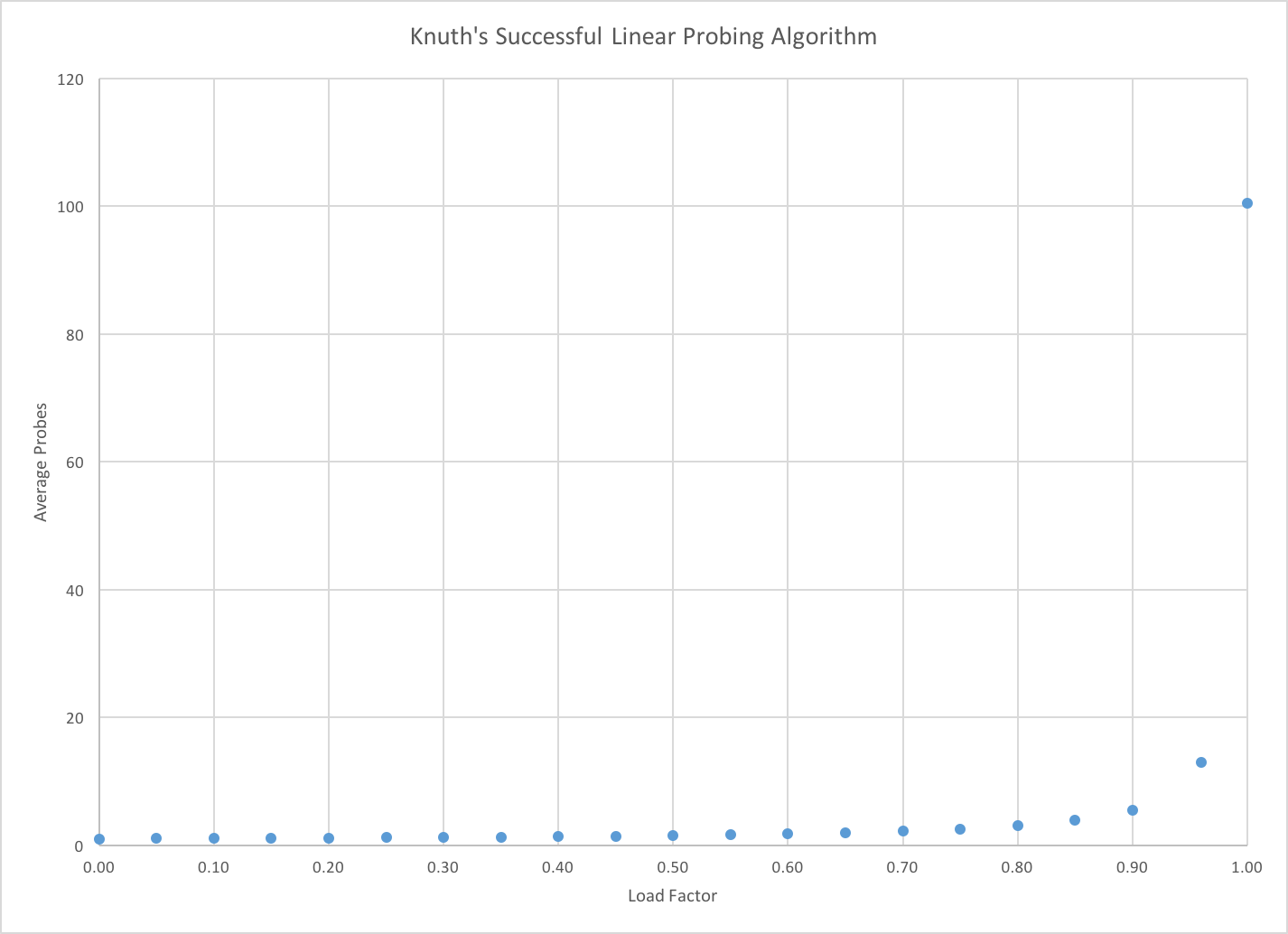
}



Data used for plots



We can see that there is a striking similarity between these and Knuth’s formula,



Based on this data, the load factor should be kept under **0.6** if the average number of probes is to be kept under 2.

Cmax| represents the maximum size of cluster as a percent of the total hash table size. As the load factor approaches 1.0, Cmax| approaches 1.0. In the plot, Cmax| only around 83% due to the fact that the table size was larger than the word count. The table size is larger because it needed to be increased in order to be prime.