CS 6505, Spring 2016
Bloom Filters Project
due: Friday, April 1 (by 11:59pm EST).

The homework should be done independently.

In this assignment you will implement a Bloom filter and then analyze the false positive rate that your implementation achieves versus the theoretical claim. You need to turn in your report (as a PDF), source code, and make file via TSquare. You can use any standard programming language (C, C++, Java, or Python), but you cannot use any libraries or any code from any other source. It is easy to code it up so just do it yourself. You can use a built-in random number generator, such as drand48().

Let N denote the size of the universe, so the items are coming from the set $\{0, 1, \ldots, N-1\}$ where N is a large number. We are adding m items into the subset S that our Bloom filter is maintaining. And our Bloom filter has a table of size n. Let $c = n/m$. Let k denote the number of hash functions used. Set $k = c \ln 2$.

For a particular choice of n and c calculate the false positive rate. Vary n and c and plot the results of the false positive rate compared to the theoretical claims.

Turn in a report with a figure (or a few figures) showing your results, an explanation of what exactly you did in your simulation, and a short conclusions section. Your report needs to be at most 1 page (double-sided). Part of your grade will be on the quality of your presentation and explanations in your report.