

# JELANI NELSON

ASSOCIATE PROFESSOR OF COMPUTER SCIENCE

JOHN A. PAULSON SCHOOL OF ENGINEERING AND APPLIED SCIENCES



**WEDNESDAY, OCTOBER 25, 2017 @7:00PM**

HARVARD UNIVERSITY, SCIENCE CENTER LECTURE HALL C  
ONE OXFORD STREET, CAMBRIDGE, MA

## “SKETCHING BIG DATA”

A “sketch” is a data structure supporting some pre-specified set of queries and updates to a database while consuming space substantially (often exponentially) less than the information theoretic minimum required to store everything seen, and thus can also be seen as some form of functional compression. The advantages of sketching include less memory consumption, faster algorithms, and reduced bandwidth requirements in distributed computing environments.

This talk will touch on some of the magic made possible by sketching techniques, such as:

- \* (Approximately) counting up to an integer  $N$  in exponentially less memory than what’s required to actually write the digits of  $N$  down.
- \* (Approximately) computing the number of distinct words ever appearing in any of Shakespeare’s works, via a method that reads through them all once while only remembering 3 lines’ worth of text in memory at any given time.
- \* Detecting trending keywords queried to a search engine, such as ‘bigly’, or ‘Irma’, while never remembering more than a negligible fraction of the query stream seen thus far.

