Course title and number: ECEN 748 Data Stream Algorithms and Applications
Term: Spring 2022
Meeting times and location: TR 12:45pm-2:00pm; ZACH 160

Instructor Information
Name: Dr. Nicholas Duffield
Telephone number: 845-7328
Email address: duffieldng@tamu.edu
Office hours: To be announced
Office location: WEB 332D

Course Description and Prerequisites
This course will study statistical and algorithmic methods for summarizing streams of Big Data. These streams are generically found to be far from uniform. This property has led to the introduction of many ingenious algorithms to sample, sketch, or otherwise summarize massive streaming datasets, by devoting limited processing and storage resources to their dominant features. In many cases a detailed understanding of the relation between summary size and accuracy can be derived. This course will present the development of such algorithms up to the present and include applications to streams of transactional data from measurement of internet services and the applications that use them, and recent developments in summarizing streaming graph data.

Prerequisite: graduate standing; ECEN 303 RANDOM SIGNALS & SYSTEMS or equivalent course involving probability

Learning Outcomes
Acquiring knowledge of statistical and algorithmic methods in data streaming and their application in network measurement and analysis. Understanding the design issues and trade-offs between statistical, computation and implementation goals. The course will prepare students to conduct their own research in streaming algorithms, or their use in other research areas.

Grading Policies
Homework: 50%
Data Project: 15%
Student Presentation: 15%
Final Exam: 20%

Grading Scale: 90-100 A, 80-89 B, 70-79 C, 60-69 D, below 60 F.

Discussion of homework assignments is encouraged, but homework must be executed independently and copying is not allowed. Assignments must be typeset and handed in on time to receive full credit. No late homework and project proposals will be accepted unless an official document (e.g., doctor's note) justifies the absence.

Textbook and/or Resource Material
The materials for this course will comprise research literature and review articles that together cover the material. These will be communicated during the course.

Additional background material is as follows:

Stream Sampling and Summarization Background
Cormode & Duffield: Sampling for Big Data
http://nickduffield.net/download/papers/Tutorial_KDD_2014.ppsx

Duffield: Sampling for Passive Internet Measurement: A Review
http://projecteuclid.org/euclid.ss/1110999311

Probability and Statistics

Computer Networking

Course Topics

<table>
<thead>
<tr>
<th>Weeks</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Data stream models, examples &amp; applications; resource limitations &amp; analysis goals</td>
</tr>
<tr>
<td>2</td>
<td>Probabilistic analysis: estimation and bounds: Markov, Chernoff, Hoeffding</td>
</tr>
<tr>
<td>3</td>
<td>Approximate data structures for keys; Bloom Filters, Cuckoo Hashing &amp; Filtering</td>
</tr>
<tr>
<td>4</td>
<td>Approximate and distinct counting algorithms; Morris, majority, hyperloglog; Flajolet</td>
</tr>
<tr>
<td>5</td>
<td>Frequency moments, heavy hitters &amp; range queries: Misra-Gries algorithm, count &amp; count-min sketches</td>
</tr>
<tr>
<td>6</td>
<td>Inverse probability estimation; uniform, weighted, reservoir &amp; priority sampling.</td>
</tr>
<tr>
<td>7</td>
<td>IPPS sampling and optimality; stream aggregation: sample and hold, rate adaptation</td>
</tr>
<tr>
<td>8</td>
<td>Sampling vs. sketching vs. aggregation. applications in Internet measurement</td>
</tr>
<tr>
<td>9</td>
<td>Hashing &amp; coordination: consistent sampling; min-hash, k-mins, L_0 and L_p sampling</td>
</tr>
<tr>
<td>10</td>
<td>Sketching, random projection, sparse recovery. Similarity. Locality sensitive hashing</td>
</tr>
<tr>
<td>11</td>
<td>Sketching graph properties: connectivity, sparsification, spanners, matching</td>
</tr>
<tr>
<td>12</td>
<td>Graph stream sampling; subgraphs counts, queries and other features</td>
</tr>
<tr>
<td>13 &amp; 14</td>
<td>Student presentations</td>
</tr>
</tbody>
</table>

Americans with Disabilities Act (ADA)
The Americans with Disabilities Act (ADA) is a federal anti-discrimination statute that provides comprehensive civil rights protection for persons with disabilities. Among other things, this legislation requires that all students with disabilities be guaranteed a learning environment that provides for reasonable accommodation of their disabilities. If you believe you have a disability requiring an accommodation, please contact Disability Services, in Cain Hall, Room B118, or call 845-1637. For additional information visit http://disability.tamu.edu

Academic Integrity
For additional information please visit: http://www.tamu.edu/aggiehonor

“An Aggie does not lie, cheat, or steal, or tolerate those who do.”