Course title and number: ECEN689 Special Topics in Data Science for Communications Networking
Term: Fall 2015
Meeting times and location: MW 03:00pm-04:15pm, HELD 111

Instructor Information

Name: Dr. Nicholas Duffield
Telephone number: 845-7328
Email address: duffieldng@tamu.edu
Office hours: MW 11:00am-12:00pm
Office location: WEB 332D

Course Description and Prerequisites

This course will study statistical and algorithmic methods for acquiring and analysing massive, complex, and incomplete datasets, with application to measurement and analysis of operational data in ISP communication networks, routers and protocols. Topics include network measurement, sampling, sketching, network probing, network tomography and graph sampling.

Prerequisite: graduate standing; approval of instructor. Students should have working knowledge of the basics of probability and statistics, and of computer networking.

Learning Outcomes

Acquiring knowledge of statistical and algorithmic methods in data science 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 this area.

Grading Policies

Homework: 50%
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

Background references
Baron: Probability and Statistics for Computer Scientists (2nd Edition)

Peterson & Davie: Computer Networks (5th Edition)

Detailed references: selections from
Leskovec, Rajaraman & Ullman: Mining of Massive Data Sets
http://www.mmds.org

Kolaczyk: Statistical Analysis of Network Data: Methods and Models

Background review articles and tutorials
Duffield: Sampling for Passive Internet Measurement: A Review
http://projecteuclid.org/euclid.ss/1110999311

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

Research literature references: will be communicated in class notes

Course Topics

<table>
<thead>
<tr>
<th>Weeks</th>
<th>Topic</th>
<th>References &amp; Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-3</td>
<td>Introduction; Passive Traffic Measurement in ISPs Sampling in Traffic Measurement; Reservoir Sampling, Weighted Sampling</td>
<td>Peterson Ch. 3,5 and Research Literature</td>
</tr>
<tr>
<td>4-6</td>
<td>Order Sampling. Database Applications. Stateful stream sampling. Counting Samples. Sample and Hold. Performance trade-offs.</td>
<td>Ullman Ch. 4, and Research Literature</td>
</tr>
<tr>
<td>10-11</td>
<td>Network Tomography</td>
<td>Kolaczyk, Ch. 7, 9</td>
</tr>
<tr>
<td>12</td>
<td>Student Presentations</td>
<td></td>
</tr>
<tr>
<td>13-14</td>
<td>Graph Sampling and Estimation</td>
<td>Kolaczyk, Ch. 4</td>
</tr>
</tbody>
</table>

Additional references to the technical literature will be provided in class.

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.”