University of Rochester
Department of Electrical and Computer Engineering Colloquia

Sampling Massive Datasets in the Internet and Beyond

Nick Duffield

Wednesday, March 19th
11:00AM – 12:00PM
CSB 209

Abstract: Massive graph datasets are used operationally by providers of internet, social network and search services. Sampling can reduce storage requirements as well as query execution times, while prolonging the useful life of the data for baselining and retrospective analysis. Sampling must mediate between the characteristics of the data, the available resources, and the accuracy needs of queries. This talk concerns a cost-based formulation to express these opposing priorities, and how this formulation leads to optimal sampling schemes without prior statistical assumptions. The talk concludes with a discussion of open technical problems and potential applications of the methods beyond the Internet.

Bio: Nick Duffield joined Rutgers University as a Research Professor in 2013. Previously, he worked at AT&T Labs Research, where he was a Distinguished Member of Technical Staff and an AT&T Fellow, and held faculty positions in Europe. He works on network and data science, particularly the acquisition, analysis and applications of operational network data. He was formerly chair of the IETF Working Group on Packet Sampling, and an associate editor of the IEEE/ACM Transactions on Networking. He is an IEEE Fellow and was a co-recipient of the ACM Sigmetrics Test of Time Award in both 2012 and 2013 for work in network tomography. He was recently TPC Co-Chair of IFIP Performance 2013, a keynote speaker at the 25th International Teletraffic Congress in Shanghai, China, and an invited speaker at the workshop on Big Data in the Mathematical Sciences in Warwick, UK.

Light refreshments will be provide