**Syllabus for Networking for Big Data (Tentative)**

1. Introduction: What is networking for big data?
2. Examples of computations in Data Centers
   * MapReduce/Hadoop, Apache Spark Systems
   * Examples of large-scale computations: PageRank, Logistic Regression, TextSearch, WordCount, etc.
   * Programming using Apache Spark
3. Networking servers for computations involving communication
   * Top-of-the-rack switches: bipartite graphs and matchings
   * Tradeoffs between complexity of scheduling algorithms and their performance
   * Hierarchical architecture of a data center network
4. Traffic Engineering inside a data center
   * Multicommodity flow problem formulation
   * Convex optimization: Karush-Kuhn-Tucker conditions and duality
   * Numerical solutions for constrained optimization problems: the penalty function method, the method of multipliers, and dual descent
5. Congestion control for data centers
   * Window flow control
   * Congestion feedback using packet marking and window size adjustment
   * Simple model for Datacenter TCP and parameter selection
6. Data Center as a Collection of Storage Servers
   * Poisson processes, exponential distribution, and their properties
   * Continuous-time Markov chains and the M/M/1 queue
   * Large-scale simulation using uniformization
   * Mean-field approximation, comparing different load balancing, replication and coding schemes
   * Caching: different caching strategies: LRU, LFU, multi-stage caching, etc.
   * Simple models for analyzing and comparing caching strategies
7. Load Balancing
   * Load balancing in a very large cluster of servers: power-of-two choices, task replication and killing, data locality
   * Virtual machine placement: Infrastructure as a Service, VMs, models with and without migration

Pre-requisites: An undergraduate course in probability at the level of EEE350. Programming in matlab and C. Convex optimization basics will be covered in class.

Target audience: Senior undergraduate and graduate students.

Course credit: 3 hours. Can earn an additional hour of credit by doing a course project.