PHYS 7337: Statistical Physics of Complex Networks

Instructor: Prof. Dima Krioukov, DA 124, x2934, dima@neu.edu
Office Hours: 3 hours weekly, T 3-6pm, and by appointment
Academic Term: Fall 2017
Credits: 4
Course Schedule: TF 9:50-11:30am
Course Location: TBD

Course Description: Covers applications of statistical physics to network science. Focuses on maximum-entropy ensembles of networks, and on applicability of network models to real networks. Main covered topics include microcanonical, canonical, and grand canonical ensembles of networks, exponential random graphs, latent variable network models, graphons, random geometric graphs and other geometric network models, and statistical inference methods using these models. Covers applications of maximum-entropy geometric network models to efficient navigation in real networks, link prediction and community structure inference.

Course Prerequisites: PHYS 5116

Textbooks and Other Resources: Reading assignments will include chapters and sections from "Maximum Entropy Models in Science and Engineering" by J. N. Kapur, "Maximum Entropy and Bayesian Methods in Applied Statistics" by J.H. Justice (ed.), "Random Graphs and Complex Networks: Volume 1" by R. van der Hofstad, "Complex Networks: Structure, Robustness and Function" by R. Cohen and S. Havlin, "Networks – An Introduction" by M. Newman, "Thermal Physics" by C. Kittel and H. Kroemer, and from original research papers that will be available via Blackboard.

Course Objectives and Learning Outcomes: A great challenge in network science is to identify an appropriate network model for given real network data, capable of yielding reliable predictions or other forms of knowledge concerning the data. Stochastic network models tend to define highly intractable ensembles of random networks, so that it is usually impossible to tell if a given real network is a typical element in the ensemble. Another source of difficulties is that measurements of real networks are usually incomplete and prune to errors and noise. Statistical physics methods have proven useful addressing these challenges as they provide a principled way to define "maximally unbiased" maximum-entropy network models based on inherently limited and imperfect observations.

The main objective of this network-theory course is to learn the general methodology behind the application of the maximum entropy principle to network data analysis. At the beginning of the course the required concepts from equilibrium statistical mechanics will be reviewed, so that a prior course in statistical physics, although desired, is not required. The course will then focus on popular models of complex networks analyzing them from the maximum-entropy perspective. Strong clustering (high probability that two friends of the same person know each other) is a ubiquitous feature of many real networks. Since random networks with strong clustering have latent geometry, the second part of the course will focus on geometric network models. Statistical inference methods using these models will then be introduced and studied. Applications of such methods to navigation in networks, link prediction and community structure inference will be considered in the last part of the course. At the end of the course, students will be able to reason what network models are applicable to what real networks and for what applications, and will be able to apply statistical inference methods based on these models to real networks.

Course Outline:

- Introduction: network models versus real networks, and the maximum entropy principle
- Microcanonical, canonical and grand canonical ensembles as ensembles with maximum entropy
- Exponential random graphs as canonical ensembles
- Erdős-Rényi random graphs $\mathcal{G}_{n,m}$ and $\mathcal{G}_{n,p}$ as microcanonical and canonical ensembles
- Random networks with hidden variables as exponential random graphs with hyperparameters
- Configuration model: random graphs with a given degree distribution
    - Microcanonical ensemble: sharp degree constraints
    - Canonical ensemble: soft degree constraints
    - Grand canonical ensemble: degree distribution constraints
- Clustering in Erdős-Rényi graphs and configuration model
- Graphons, exchangeability and projectivity
- Random geometric graphs
- Random networks with strong clustering as random geometric graphs
- Random hyperbolic networks and their properties
- Statistical inference methods using random geometric network models
- Applications of maximum-entropy models of geometric networks to real networks
    - Navigation and biased spreading in networks
    - Prediction of missing and future links
    - Inference of soft communities

Course Organization:

<u>Lectures.</u> Lectures will be given by Prof. Krioukov. Occasionally a guest lecturer may be invited to talk on a specific topic. Reading assignments will be communicated via Blackboard.

<u>Homework.</u> There will be two written homework assignments with due dates posted at Blackboard. The assignment will be graded within a week from the due dates, and the scores will be posted at the grade center at Blackboard.

<u>Final Examination.</u> The take-home final exam will consist of theoretical and computational miniprojects due within two weeks from the assignment.

Grading: Final course grades will be determined as follows:

25% Written homework assignment 1

25% Written homework assignment 2

30% Final examination

10% Participation and attendance

10% Teacher evaluation

Grade	**Score**
A	92-100
A-	88-92
B+	84-88
B	79-84
B-	71-79
C+	68-71
C	63-68
C-	60-63

Course Policies: Be sure to review and follow the College of Science Academic Course Policies available at http://www.northeastern.edu/cos/wp-content/uploads/2014/11/Northeastern-COS-Policies-Template.pdf