

## DEPARTMENT OF STATISTICS AND BIOSTATISTICS

**Pranjal Awasthi**

Department of Computer Science  
Rutgers University

*Efficient PAC Learning from the Crowd*

**September 20, 2017**

**3:20 – 4:20pm**

Light refreshments will be served

**110 Frelinghuysen Road**

**Hill Center, Room 552**

**Abstract:** In recent years crowdsourcing has become the method of choice for gathering labeled training data for learning algorithms. Standard approaches to crowdsourcing view the process of acquiring labeled data separately from the process of learning a classifier from the gathered data. This can give rise to computational and statistical challenges. For example, in most cases there are no known computationally efficient learning algorithms that are robust to the high level of noise that exists in crowdsourced data. Furthermore, efforts to eliminate noise through majority voting often require a large number of queries per example.

In this talk I will describe a model of crowdsourcing and algorithms that interleave the process of labeling and learning to get good classifiers with minimal overhead. In particular, I will show that any function class learnable in the traditional PAC model is also learnable in the crowdsourced model where only a noticeable fraction of the labelers are correct. Furthermore, each labeler is required to label only a constant number of examples and each example is only queried a constant number of times on average.

Joint work with Avrim Blum, Nika Haghtalab and Yishay Mansour.

**Bio:** Pranjal Awasthi is an assistant professor at Rutgers in the Computer Science department, with expertise in Data Science and Machine Learning. Before joining Rutgers, he obtained his Ph.D. degree from Carnegie Mellon University and held a post-doctoral position at Princeton University.

