香港中文大學
The Chinese University of Hong Kong

Institute of Theoretical Computer Science and Communications

ITCSC-CSE Joint Seminar

The Power of Adaptivity in Data Analysis

By
Mr. Thomas Steinke
Harvard University

January 21, 2016, Thursday

11:00 am – 12:00 noon

Room 121, 1/F, Ho Sin Hang Engineering Building, CUHK

Abstract:

If a dataset is only used once, a rich theory exists for ensuring that the conclusions are valid. But what happens if the same dataset is reused for multiple analyses? Since each analysis may now depend on the outcome of previous analyses, the danger of overfitting the dataset is increased. For example, if the same dataset is used to select a learning model and then fit that model, the resulting model may appear to explain the data better than it should.

In this talk, I will discuss a recent line of research on adaptive data analysis. I will show that there are sophisticated techniques -- using tools from information theory and differential privacy – that enable us to ensure that adaptive analysis provides sound conclusions. I will also discuss how adaptive data analysis is inherently more powerful than non-adaptive data analysis.

Biography:

Thomas Steinke is a PhD candidate in computer science at Harvard University advised by Prof. Salil Vadhan. His research focuses include pseudorandomness in complexity theory and data privacy, particularly its connections to information theory and learning theory. Thomas is originally from New Zealand where he completed BSc(Hons) and MSc degrees at the University of Canterbury.

***** ALL ARE WELCOME *****

Hosted & Enquiries : Institute of Theoretical Computer Science and Communications   Tel: 3943 1257