

Dr Linda Tan

Assistant professor, Department of Statistics and Applied Probability, National University of Singapore

**Title:**

Topic-Adjusted Visibility Metric for Scientific Articles

**Abstract:**

Measuring the impact of scientific articles is important for evaluating the research output of individual scientists, academic institutions and journals. While citations are raw data for constructing impact measures, there exist potential issues if factors affecting citation patterns are not accounted for. In this seminar, we introduce a visibility metric for individual articles that address the problem of field variation in citation rates. This measure derives from joint probabilistic modelling of the text in the articles and the citations among them using latent Dirichlet allocation and the mixed membership stochastic blockmodel. Our proposed model also provides a structural understanding of citation behaviour in different fields and article recommendations which factor in article visibility and citation patterns. We develop an efficient algorithm for model fitting using variational methods. To scale up to large networks, an online variant is developed using stochastic gradient methods and case-control likelihood approximation. We apply our methods to the benchmark KDD Cup 2003 dataset with approximately 30,000 high energy physics papers.