

## **AAI Advanced Analytics Seminar Series on 13/04/2012**

**Seminar Title:** Large scale data mining tasks using random walks

**Speaker:** Khoa Nguyen, School of Information Technologies, University of Sydney

**Date and Time:** 1:30pm to 3:00pm, the 13th of April 2012 (Friday)

**Seminar Room:** UTS Blackfriars Campus CC05.GD.01 (5 minutes walk from Tower Building CB01 of UTS)

**Street Address:** 2-12, Blackfriars Street, Chippendale, NSW 2008

**Seminar Chairman:** Prof Longbing Cao (longbing.cao@uts.edu.au)

**Abstract:** Commute time is a robust measure derived from random walks on graphs. Commute time has found widespread applications in personalized search, collaborative filtering, image segmentation, and making search engines robust against manipulation. The fact that the commute time is averaged over all paths (and not just the shortest path) makes it more robust to data perturbations. We present methods to use commute time as a distance measure for data mining tasks such as anomaly detection and clustering. However, the computation of commute time involves the eigen decomposition of the graph Laplacian and thus is impractical to use in large graphs. We also propose methods to efficiently and accurately compute commute time in batch and incremental fashions. Specifically, we present a novel distance-based method using commute time which can capture global, local, and group anomalies. We propose a fast and accurate approximation for spectral clustering using an approximate commute time embedding. Moreover, we propose a method to incrementally estimate the commute time in constant time and use it for online anomaly detection.

**Short biography of the speaker:** Khoa Nguyen received a B.E. degree from Ho Chi Minh City University of Technology, Vietnam, in 2002 and then became a lecturer assistant at Faculty of Computer Science and Engineering of the same university. After that he earned a M.E. degree from Ritsumeikan University, Japan, in 2007. He is currently a Ph.D. student at the University of Sydney, Australia and has just submitted his thesis. During the Ph.D. course, he also worked part-time as a data mining researcher in Insurance Australia Group (AIG) and Brain Resource Company (BRC). His research interests include data mining, machine learning, large scale anomaly detection and clustering.

### **Overview to This Seminar Series**

The Advanced Analytics Seminar Series presents the latest theoretical advancement and empirical experience in a broad range of interdisciplinary and business-oriented analytics fields. It covers topics related to data mining, machine learning, statistics, bioinformatics, behavior informatics, marketing analytics and multimedia analytics. It also provides a platform for the showcase of commercial products in ubiquitous advanced analytics. Speakers are invited from both academia and industry. It opens regularly on every Friday afternoon at the garden-like UTS Blackfriars Campus. You

are warmly welcome to attend this seminar series.

Jinyan Li, Seminar Coordinator  
Associate Professor, Advanced Analytics Institute  
Faculty of Engineering and IT  
University of Technology, Sydney  
P.O. Box 123, Broadway, NSW 2007, Australia  
<http://www-staff.it.uts.edu.au/~jinyli>