香港浸會大學理學院
HKBU Faculty of Science

HKBU
Computer Science Department
計算機科學系

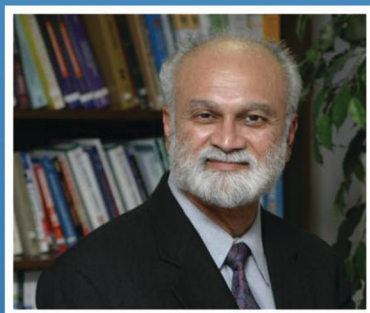# Department of Computer Science
Distinguished Lecture 2012 Series

Hypotheses Generation as Supervised Link Discovery with Automated Class Labeling on Large-Scale Biomedical Concept Networks

Prof. Vijay Raghavan

**Distinguished Professor in Computer Science
University of Louisiana, USA**

Dr. Vijay Raghavan is the Distinguished Professor of Computer Science at the Center for Advanced Computer Studies and a co-director of the Laboratory for Internet Computing. His research interests are in information retrieval and extraction, data and web mining, multimedia retrieval, data integration, and literature-based discovery. He has published over 200 peer-reviewed research papers (https://www.researchgate.net/profile/Vijay_Raghavan10/). These and other research contributions cumulatively accord him an h-index of 28, based on citations to his publications. He has served as major adviser for 23 doctoral students and has garnered $8 million in external funding. Dr. Raghavan brings substantial technical expertise, interdisciplinary collaboration experience, and management skills to his projects.

His service work at the university includes coordinating the Louis Stokes-Alliance for Minority Participation (LS-AMP) program since 2001. Raghavan has served as PC Chair, PC Co-chair or PC member in countless ACM and IEEE sponsored conferences. He received the ICDM 2005 Outstanding Service Award. Raghavan was an ACM National Lecturer from 1993 to 2006. He chaired the ICDM Awards Committee in 2006 and served as a committee member until 2008. He was a member of the Advisory Committee of the NSF Computer and Information Science and Engineering (CISE) directorate, from 2008-2010. Since 2010, he serves on the Executive Committee of the IEEE -TC on Intelligent Informatics.

From 1997 to 2003, he worked closely with the USGS National Wetlands Research Center and with the Department of Energy's Office of Science and Technical Information on a digital library with data mining capabilities incorporated. He is the Founding Director of the NSF-sponsored Industry University Cooperative Research Center, on the theme of Visual and Decision Informatics, since February 2012.

Date:	**December 12, 2012 (Wed)**
Time:	**3:30 p.m.**
Venue:	**RRS905, Sir Run Run Shaw Building**
	Ho Sin Hang Campus, HKBU

Abstract:

Computational approaches to generate hypotheses from biomedical literature have been studied intensively in recent years. Nevertheless, it still remains a challenge to automatically discover novel, cross-silo biomedical hypotheses from large-scale literature repositories. In order to address this challenge, we first model a biomedical literature repository as a comprehensive network of biomedical concepts and formulate hypotheses generation as a process of link discovery on the concept network. We extract the relevant information from the biomedical literature corpus and generate a concept network and concept-author map on a cluster using Map-Reduce frame-work. We extract a set of heterogeneous features such as random walk based features, neighborhood features and common author features. The potential number of links to consider for the possibility of link discovery is large in our concept network and to address the scalability problem, the features from a concept network are extracted using a cluster with Map-Reduce framework. We further model link discovery as a classification problem carried out on a training data set automatically extracted from two network snapshots taken in two consecutive time duration. A set of heterogeneous features, which cover both topological and semantic features derived from the concept network, have been studied with respect to their impacts on the accuracy of the proposed supervised link discovery process. A case study of hypotheses generation based on the proposed method has been presented in the paper.

ALL ARE WELCOME

Tel: 3411 2782
Email: comp@comp.hkbu.edu.hk
Website: http://www.comp.hkbu.edu.hk