COLLOQUIUM

School of Computing and Software Engineering (CSE)

Subject: Exploiting Hierarchical Domain Structure to Compute Similarity

Speaker: **Dr. Jennifer Vandenbussche, Math Department, SPSU**

Date: **Thursday, May 21, 2009**

Time: **3:00 PM – 4:00PM**

Place: **J-381**

ABSTRACT

There are a variety of contexts in which one wishes to systematically evaluate the "similarity" of various collections of objects: Analysis of consumer buying patterns, categorization of web pages and other documents, stock market risk analysis, and many more. Traditional measures of similarity have focused on simply counting the number of items two collections have in common. In many contexts, this measure of similarity is far too naive. For example, consider three customers at a grocery store. Bob buys oranges and bread, Alice buys apples and pretzels, and Joe buys ice cream and cookies. Intuitively, Bob and Alice are more "similar" customers than Bob and Joe; however, an algorithm which simply looks for identical items in their grocery cart will miss this similarity.

In this talk, we present a paper by Ganesan, Garcia-Molina, and Widom from 2003 that introduces several measures of similarity that may correct this problem. Given a categorization of items via a rooted tree, the authors incorporate information from the tree in their measures of similarity in simple yet effective ways. In many contexts, these new measures of similarity seem to agree with our intuitive idea of similarity much more closely.

This talk will be accessible (and hopefully interesting!) to a general audience.

Biographical Sketch

Dr. Jennifer Vandenbussche is an assistant professor in the Department of Mathematics at SPSU. She received her Ph.D. in Mathematics from the University of Illinois at Urbana-Champaign in 2008. Dr. Vandenbussche's primary area of research is graph theory, with a focus on extremal problems in graphs - that is, finding the largest or smallest value a certain graph parameter may take on under certain conditions. Specific areas of interest include matching theory, coloring, independent sets, and connectivity parameters.

MARK YOUR CALENDAR **PLEASE POST**

Visit http://cise.spsu.edu/research/seminar.asp for information on CISE Brown Bag series and CISE Technical Reports.