

CCSU
DEPARTMENT OF MATHEMATICAL SCIENCES
COLLOQUIUM

Friday, February 24

3:00 – 4:00 PM

Maria Sanford, Room 101

**TEXT MINING WWI-ERA
CONNECTICUT NEWSPAPER STORIES
WITH GRAPH THEORY AND BEYOND**

ROGER BILISOLY

CENTRAL CONNECTICUT STATE UNIVERSITY

Abstract: Because of the wild success of the web and the increase in computing power, network theory has experienced explosive growth in the past 20 years. Moreover, tools such as NetworkX (for Python) are available to apply this methodology to data. This talk discusses my ongoing work on analyzing WWI-era newspaper articles with Christine Pittsley, who works at the Connecticut State Library. Articles are from two Connecticut newspaper corpora for dates mostly in the 1910s. Using text mining, these archives can be searched for terms of interest such as Connecticut cities or battles of WWI. Graph structures can arise in several ways. For example, a set of proper names constitutes the nodes, and an edge exists exactly when there exists a newspaper story that mentions the two proper names in question. Two sets of nodes can be analyzed at the same time by using affiliation matrices (an idea from mathematical sociology), which can be studied as bipartite graphs, and techniques such as Galois lattices can be used to study the relationship between these two sets. Finally, it has been recognized that graph theory is not general enough for certain situations. For instance, if nodes were cities in Connecticut, then the following two cases are distinct. First, let A, B, C be three specific towns. Suppose there were three newspaper stories mentioning A and B, A and C, B and C, respectively. Second, suppose there is one story that mentions all three cities. The former is a complete graph of size 3, but the latter could be modeled as a 2-simplex. That is, simplicial complexes could be used for modeling, which has been done by a few researchers.

For further information:

gotchevi@ccsu.edu 860-832-2839

<http://www.math.ccsu.edu/gotchev/colloquium/>