CCSU
DEPARTMENT OF MATHEMATICAL SCIENCES

COLLOQUIUM

Friday, September 6
2:00 – 3:00 PM
Maria Sanford, Room 101

ANALYZING TEXT VARIABILITY WITH METRICS

ROGER BILISOLY

CENTRAL CONNECTICUT STATE UNIVERSITY

Abstract: Researchers in the 2000s were able to extend many statistical ideas to data constrained to a manifold. For example, the mean and variance were generalized by recasting them as solution(s) of minimizing a sum of a metric on a data space. Although this was originally done with Riemannian metrics, any type can be used. We apply this approach to text metrics and explore two applications. First, we consider the variability of English spelling over the last 1000 years using Levenshtein's edit distance. It turns out that Middle English has the least standardization, and this can be quantified. Second, we consider metrics on words in the sense of group theory. Using the word metric, the variability of a group can be defined allowing us to answer questions such as "Which is more variable: Z_3 x Z_5 or Z_{15}?" However, there are many other group metrics, and we'll finish by examining the connection between nonparametric statistical methods and metrics on the symmetric group. For example, Spearman's rank correlation is equivalent to the L^2 norm on S_n.

For further information:
gotchevi@ccsu.edu 860-832-2839
http://www.math.ccsu.edu/gotchev/colloquium/