(1.3) Scalar Indices of Cluster Validity; Do you believe the outputs?

Abstract. This talk is about the non-visual way to attack cluster validity. There are literally hundreds and hundreds of cluster validity indices (CVIs) available for evaluating crisp and soft clusters. None of them work over a large cross section of input data; all of them work from time to time. I cannot cover all of the listed indices in any reasonable amount of time, but I will be able to give you an overview of how they all fit together.

1. Overview of methods for choosing a CVI.
    A. Best c-method for internal indices
    B. Best match to a reference partition for external indices
    C. Combining best c and best match methods to choose an internal CVI
2. Crisp partitions:
    A. Distance and pair-counting measures (e.g. Rand, Adjusted Rand, Jacard ...)
    B. Statistical correlation (e.g. Pearson, Spearman, Gamma, CCV ...)
    C. Information Theoretic (e.g. Mutual Information, Variation of Information ...)
    D. Graph-Theoretic (e.g. Newman-Girvan modularity, Estabrook...)
    E. Classics (e.g. Davies-Bouldin, C-index ...)
3. Soft partitions (fuzzy/probabilistic):
    A. Measures of fuzziness (e.g. partition coefficient & entropy...)
    B. Generalizations of 2A and 2C using the contingency matrix
    C. Generalizations of 2D (e.g. fuzzy modularity...)
    D. Classics: (e.g. Xie-Beni, iCOMP, MDL, AIC ...)
4. Incremental CVIs for streaming data
    A. Xie-Beni, Davies-Bouldin, Generalized Dunn Indices
    B. Incremental sequential k-means