

Comparing Partitions from Non-equal Sets of Units

Marjan Cugmas and Anuška Ferligoj

University of Ljubljana

Rand (1971) proposed well-known index for comparing two partitions obtained on the same set of units. The index takes the value on the interval between 0 and 1, where a higher value indicates more similar partitions. The index is symmetric in the sense that both splitting and merging of clusters lower the value of the index. Sometimes, e.g. when the units are observed in two time periods, the splitting and merging of clusters should be considered differently, according to the operationalization of stability of clusters. In that case the Wallace index B' (Wallace 1983) can be used. Furthermore, there are several cases when one would want to compare two partitions obtained on the different sets of units, where one is a subset of another. In that case, the new units and units which leave the clusters from the first partition can be considered as a factor lowering the value of the index. Therefore, two modified Rand indices will be presented, along with its correction for chance, which allow comparing different values.