index.G3(clusterSim)

G3 Hubert & Levine internal cluster quality index

$$G3(u) = \frac{D(u) - r \cdot D_{min}}{r \cdot D_{max} - r \cdot D_{min}}, D_{min} \neq D_{max},$$
$$G3(u) \in (0, 1),$$

where: $D(u)$ – all within-cluster dissimilarities,

r – number of within-cluster dissimilarities,

D_{min} – smallest within-cluster dissimilarity,

D_{max} – largest within-cluster dissimilarity,

u – number of clusters ($u = 2, ..., n - 2$).

The value of u, which minimizes $G3(u)$, is regarded as specifying the number of clusters.

References

Gatnar, E., Walesiak, M. (Eds.) (2004), *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych [Multivariate statistical analysis methods in marketing research]*, Wydawnictwo AE, Wroclaw, p. 339.

Gordon, A.D. (1999), *Classification*, Chapman & Hall/CRC, London, p. 62.

Milligan, G.W., Cooper, M.C. (1985), *An examination of procedures of determining the number of cluster in a data set*, "Psychometrika", vol. 50, no. 2, 159-179.