Adjusted Concordance Index: an Extension of the Adjusted Rand Index to Fuzzy Partitions

Antonio D’Ambrosio1 · Sonia Amodio2 · Carmela Iorio3 · Giuseppe Pandolfo3 · Roberta Siciliano3

Published online: 15 June 2020 © The Classification Society 2020

Abstract
In comparing clustering partitions, the Rand index (RI) and the adjusted Rand index (ARI) are commonly used for measuring the agreement between partitions. Such external validation indexes can be used to quantify how close the clusters are to a reference partition (or to prior knowledge about the data) by counting classified pairs of elements. To evaluate the solution of a fuzzy clustering algorithm, several extensions of the Rand index and other similarity measures to fuzzy partitions have been proposed. An extension of the ARI for fuzzy partitions based on the normalized degree of concordance is proposed. The performance of the proposed index is evaluated through Monte Carlo simulation studies.

Keywords Clustering · Cluster validity · Fuzzy partitions · Normalized degree of concordance

Antonio D’Ambrosio
antdambr@unina.it

Sonia Amodio
S.Amodio@lumc.nl

Carmela Iorio
carmela.iorio@unina.it

Giuseppe Pandolfo
giuseppe.pandolfo@unina.it

Roberta Siciliano
roberta@unina.it

1 Department of Economics and Statistics, University of Naples Federico II, Naples, Italy
2 Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, Netherlands
3 Department of Industrial Engineering, University of Naples Federico II, Naples, Italy
Cluster analysis consists of a large collection of methods that aim to find groups (clusters) in data (Everitt et al. 2011; Hennig and Meila 2015). Most of the clustering methods look for a partition of the data in which each unit (or object) belongs to a single cluster and the complete set of clusters contains all of the individuals (or objects). However, overlapping clusters may sometimes provide a better solution. In addition to the abovementioned partitioning methods, there are many clustering algorithms within the literature. In general, the major fundamental clustering methods can be classified into partitioning, hierarchical, density-based, grid-based, and model-based methods (Han et al. 2012). Here, we are interested in the so-called heuristic methods for which “no assumptions about class structure are made and the choice of the clustering procedure is often based on conventional or posterior criteria such as the interpretability of the results” (Stahl and Sallis, 2012, p. 342). We refer the reader to Everitt et al. (2011) for a complete overview.

With this clarification, broadly speaking, cluster analysis can be defined as an unsupervised method to partition a set of \( n \) objects in a \( p \)-dimensional space, \( X = \{ x_{i,j} \}_{(n \times p)} \), into a finite set of clusters, \( C = \{ C_1, ..., C_K \} \), according to the similarities among these objects such that the objects in the same cluster are as similar as possible and objects in different clusters are as dissimilar as possible (Anderberg 1973; Berkhin 2006; Böck 1974; Duran and Odell 2013; Fasulo 1999; Hartigan 1975; Kaufman and Rousseeuw 2005; Jain and Dubes 1988, 1999; Mirkin 1998; Spath 1980).

An important distinction between crisp (or hard) and fuzzy (or soft) clustering algorithms can be made. The former consider disjoint partitions, namely, an object can belong only to a single cluster. This clustering approach may be inadequate when data points are almost equally distant from two or more clusters. Crisp data partitions “might not express the difference between data points in the center and those that are rather at the boundary of a cluster” (D’Urso, 2015, p. 547). In such cases, fuzzy clustering algorithms are usually preferred.

Fuzzy clustering algorithms are aimed at determining the degree of membership of the objects to each cluster (Ruspini 1970; Bezdek et al. 1984; Höppner et al. 1999). In this way, objects on the boundary between different clusters are not forced to belong to one of them, and they have a different degree of membership for each cluster.

Cluster validation plays a very important role in cluster analysis because clustering methods tend to generate clusters even for fairly homogeneous data sets. Cluster validation can involve both internal and external validation criteria. Internal validation criteria “evaluate to what extent the clustering fits the data set based on the data used for clustering” (Halkidi, Vazirgiannis, and Hennig, 2015, p. 596). They are based only on the observation of data that are clustered. External validation criteria evaluate how well a clustering matches a predefined structure. They are defined based on some information not used in the clustering process, i.e., a golden standard cluster structure known a priori or a different clustering method.

In this work, we focus on external validation measures. In the literature, many external validation criteria have been proposed (Albatineh et al. 2006). Among them are the Rand index (Rand 1971), the Fowlkes and Mallows \( F \)-measure (Fowlkes and Mallows 1983), the Jaccard index (Downton and Brennan 1980), the Mirkin metric (Mirkin 1998) and the Dice coefficient (Dice 1945). One of the most popular indexes is the adjusted Rand index (ARI), which was proposed by Hubert and Arabie (1985) in order to correct some drawbacks of the Rand index.

In this paper, we argue that in comparing the partitions from two different clustering methods, an index should satisfy two desiderata, i.e., properties to be satisfied by an index...
in general and not for a given data set (Warrens 2008a): (i) reflexivity and (ii) a proper interpretation of correction for agreement due to chance. We emphasize that among the indexes proposed in the literature, no single one satisfies both abovementioned properties when two fuzzy partitions are compared. For this reason, we propose a fuzzy extension of the ARI, called the adjusted concordance index (ACI), that satisfies both abovementioned desiderata. The key idea is to use the normalized degree of concordance (NDC) defined by H¨ullermeier et al. (2012) and to normalize the difference between the NDC and the point estimate of its expected value in order to correct this measure for agreement that may be due to chance, as suggested in Hubert and Arabie (1985, p. 198).

The paper is organized as follows: Section 2 is devoted to a short overview of cluster analysis and to the existing extensions of the Rand index and its fuzzy extensions. In Section 3, the adjusted Rand index is recalled. In Section 4, the normalized degree of concordance is presented. Section 5 is used to introduce the adjusted concordance index. Section 6 is dedicated to the analysis of the performance of the proposed index. Concluding remarks end the paper in Section 7.

2 Rand Index and Its Fuzzy Extensions

Given an $n \times p$ data matrix $X$, where $n$ is the number of objects and $p$ the number of variables, the crisp clustering structure can be presented as a set of nonempty $K \geq 2$ subsets $\{ C_1, \ldots, C_k, \ldots, C_K \}$ such that:

$$X = \bigcup_{k=1}^{K} C_k, \quad C_k \cap C_{k'} = \emptyset, \quad \text{for} \quad k \neq k'.$$

Two objects of $X$, i.e., $(x, x')$, are paired in $C$ if they belong to the same cluster. Let $P$ and $Q$ be two partitions of $X$. The Rand index is calculated as follows:

$$RI = \frac{a + d}{a + b + c + d} = \frac{a + d}{\binom{n}{2}},$$

where

- $a$ is the number of pairs $(x, x') \in X$ that are paired in $P$ and in $Q$;
- $b$ is the number of pairs $(x, x') \in X$ that are paired in $P$ but not paired in $Q$;
- $c$ is the number of pairs $(x, x') \in X$ that are not paired in $P$ but are paired in $Q$ and
- $d$ is the number of pairs $(x, x') \in X$ that are not paired in either $P$ or in $Q$.

The Rand index takes values in $[0, 1]$, with 0 indicating that the two partitions do not agree for any pair of elements and 1 indicating that the two partitions are exactly the same. Moreover, the Rand index is reflexive, namely, $RI(P, P) = 1$. Comparison measures that can also be expressed in terms of the quantities $a, b, c,$ and $d$ include the Jaccard index (Jaccard 1901), the H-index (Hamann 1961; Hubert 1977), Kulczynski’s index (Kulczynski 1927), the GL index (Gower and Legendre 1986), the Fowlkes-Mallow $F$-index (Fowlkes and Mallows 1983), the Mirkin metric (Mirkin 1998), and the Dice coefficient (Dice 1945). See Albatineh et al. (2006) for a more extensive description of such indexes.
Fuzzy clustering methods assign the elements of \( \mathbf{X} \) into \( K \) fuzzy clusters with respect to some defined criterion, and they return a set of cluster centers and a partition matrix \( \mathbf{W} \) of the following form:

\[
\mathbf{W} = \{ w_{i,k} \}_{(n \times K)} \in [0, 1]; \quad \sum_{k=1}^{K} w_{i,k} = 1 \quad \forall i \in \{1, \cdots, n\}.
\]

In the literature, several extensions of the Rand index and other external validation criteria for fuzzy partitions have been proposed (Campello 2007; Frigui et al. 2007; Brouwer 2009; Anderson et al. 2010; Hüllermeier et al. 2012). They are based on fuzzy extensions of the abovementioned four cardinalities.

More specifically, Campello (2007) proposed a fuzzy extension of the Rand index and related indexes by defining a set-theoretic form for the calculation of the four cardinalities in order to compare a fuzzy partition with a nonfuzzy one. As noted by the author, this approach can also be used to compare two fuzzy partitions. Frigui et al. (2007) proposed a similar measure, which can be considered a special case of the index of Campello.

An extension of the Rand index based on cosine correlation as a measure of the similarity between two items with fuzzy membership vectors was proposed by Brouwer (2009).

A fuzzy generalization of the Rand index and many other measures between soft partitions (i.e., fuzzy and possibilistic partitions) based on matrix operations were proposed by Anderson et al. (2010). This proposal exhibits a clear advantage in terms of efficiency because it does not consider all pairs of objects involved in the calculation of the four cardinalities. The proposed fuzzy extensions of the Rand index present a drawback that arises with respect to fuzzy partitions. Indeed, in such cases, it is not straightforward to obtain contingency tables. Following the approach of Hubert and Arabie (1985), Anderson et al. (2010) derived a fuzzy generalization of the contingency tables, but neither the marginals nor the elements of the tables are integers, and some of the cardinalities can be negative. This implies that the use of the binomial coefficients through the Gamma function is no longer straightforward.

Hüllermeier et al. (2012) proposed a generalization of the Rand index and related measures. This proposal was based on the concept of a fuzzy equivalence relation in terms of similarity measures, producing an alternative version of the four fuzzy cardinalities. The authors proposed, among others, a fuzzy Rand index, called the normalized degree of concordance (NDC), which varies between 0 and 1.

All of the abovementioned fuzzy Rand indexes have maximum theoretical value equal to one. Except for the index proposed by Hüllermeier et al. (2012), the maximum value is attained only when the partitions are crisp, which is a special case of fuzzy data. This means that, by denoting a generic index by \( I \), none of these indexes is reflexive, a condition that requires \( I(\mathbf{P}, \mathbf{P}) = 1 \). From our perspective, reflexivity is an especially important property in the comparisons of two fuzzy partitions. Comparison of the outcome of a fuzzy clustering algorithm with a hard reference partition may be the case for many applications (Campello 2007). Indeed, in such cases, an index, although reflexive, never equals one. However, comparing two fuzzy partitions is not a matter of secondary importance. For example, comparing the output of two fuzzy clustering algorithms for the same data is an important issue (Halkidi et al. 2015), for which the use of a reflexive index is an essential condition.
3 Adjusted Rand Index

The Rand index presents several drawbacks: (i) it concentrates in a small interval close to 1, thus presenting low variability (Fowlkes and Mallows 1983); (ii) it approaches its upper limit as the number of clusters increases (Morey and Agresti 1984); and (iii) it is extremely sensitive to the number of groups considered in each partition and their density (Meilă 2007).

To overcome these problems, Hubert and Arabie (1985) proposed an adjusted version of the Rand index (ARI). This correction is equal to the normalized difference between the Rand index and its expected value under the hypothesis that the model of randomness follows the generalized hypergeometric distribution:

$$ARI = \frac{RI - E(RI)}{1 - E(RI)},$$

where $RI$ and $E(RI)$ indicate the Rand index and its expected value, respectively. In terms of the quantities $a, b, c,$ and $d$, the Hubert-Arabie formulation of the adjusted Rand index is given by:

$$ARI = \frac{2(ad - bc)}{b^2 + c^2 + 2ad + (a + d)(c + b)}.$$  

The formulation reported in Eq. 2 was derived as a simplification of:

$$ARI = \frac{\sum_{i,j} \binom{n_{ij}}{2} - \sum_i \binom{n_{i+}}{2} \sum_j \binom{n_{+j}}{2} / \binom{n}{2}}{0.5[\sum_i \binom{n_{i+}}{2} + \sum_j \binom{n_{+j}}{2}] - \sum_i \binom{n_{i+}}{2} \sum_j \binom{n_{+j}}{2} / \binom{n}{2}},$$

where $n$ is the number of objects, and $n_{i+}$ and $n_{+j}$ are the row and column marginals, respectively, of the contingency table obtained by crossing two crisp partition vectors. Simpler alternatives to Eq. 2 have been proposed in the literature (see, e.g., the equivalence of the ARI and Cohen’s kappa in Warrens, 2008b). It is worth noting that the above reported drawbacks still hold for the fuzzy extensions of the Rand index. The proposals of Campello (2007), Frigui et al. (2007), Brouwer (2009), and Anderson et al. (2010) are all inclusive of the adjusted versions of the fuzzy Rand index, for which results are “most meaningful in the fuzzy case just as the ARI is in the strictly crisp partition case” (Brouwer, 2009, p. 233). Hüllermeier et al. (2012) did not include any adjusted Rand index in their proposal.

It is worth noting that the ARI is a reflexive and chance-corrected index. By examining (1), it is clear that a fuzzy corrected index is reflexive if the corresponding raw index is reflexive as well. In other words, fuzzy ARI $(P, Q) = 1$ if and only if fuzzy RI $(P, Q) = 1$. Stating that fuzzy RI $(P, Q) = 1$ is unequivocally equivalent to stating that $P = Q$.

4 Normalized Degree of Concordance

Let $W$ be a fuzzy partition of the data matrix $X$. Each object $x \in X$ is then characterized by its membership vector:

$$w(x) = (w_1(x), w_2(x), ..., w_K(x)) \in [0, 1]^K,$$

where $w_k(x)$ is the membership degree of $x$ in the $k$-th cluster. Given any pair $(x, x') \in X$, Hüllermeier et al. (2012) defined a fuzzy equivalence relation on $X$ in terms of similarity measure as:

$$E_W = 1 - \|W(x) - W(x')\|,$$
where \( \| \cdot \| \) is the normalized \( L_1 \)-norm, yielding a value in \([0, 1] \). \( E_w \) is equal to 1 if and only if \( x \) and \( x' \) have the same membership pattern.

Given two fuzzy partitions, \( P \) and \( Q \), the basic idea behind the fuzzy extension of the Rand index is to generalize the concept of concordance. For a pair \((x, x')\), Hüllermeier et al. (2012) defined the degree of concordance as:

\[
\text{conc}(x, x') = 1 - \|E_P(x, x') - E_Q(x, x')\| \in [0, 1].
\]

A distance measure is then defined by the normalized sum of concordant pairs:

\[
d(P, Q) = \frac{\sum_{(x, x') \in X} \|E_P(x, x') - E_Q(x, x')\|}{n(n-1)/2}.
\]

(4)

A direct generalization of the Rand index corresponds to the normalized degree of concordance (NDC) and is equal to:

\[
\text{NDC}(P, Q) = 1 - d(P, Q).
\]

(5)

The index in Eq. 5 reduces to the original Rand index when partitions \( P \) and \( Q \) are nonfuzzy.

Note that the distance defined in Eq. 4 is a pseudo-metric, which is a proper metric when we consider particular assumptions, which can be summarized as considering Ruspini’s partitions and the existence of a prototypical element for each cluster (see Hüllermeier et al., 2012, for a formal proof). Hüllermeier et al. (2012) proposed extensions of a large number of comparison measures that can be expressed in the cardinalities \( a, b, c, \) and \( d \) by extending the quantities \( a, b, c, \) and \( d \) to fuzzy logic. The fuzzy extensions are given by:

\[
\begin{align*}
a &= \top(1 - |E_P(x, x') - E_Q(x, x')|), \top(E_P(x, x'), E_Q(x, x')); \\
b &= \max(E_P(x, x') - E_Q(x, x'), 0); \\
c &= \max(E_Q(x, x') - E_P(x, x'), 0); \\
d &= \top(1 - |E_P(x, x') - E_Q(x, x')|), \bot(1 - E_P(x, x'), 1 - E_Q(x, x')).
\end{align*}
\]

(6)

where \( a \) and \( d \) indicate concordance, while \( b \) and \( c \) indicate discordance. \( \top \) denotes a triangular product norm, and \( \bot \) is the associated triangular conorm (algebraic sum) (Klement et al. 2010).

The normalized degree of concordance varies between 0 and 1, and it always equals 1 when comparing a fuzzy partition with itself. The properties of the NDC index play a key role in the definition of the adjusted concordance index, which is defined in the following section.

### 5 Adjusted Concordance Index

Our proposal is to use any fuzzy Rand-like index and normalize the difference between itself and the point estimate of its expected value. Specifically, we use “the general form of an index corrected for chance” of Hubert and Arabie (1985, p. 198):

\[
\text{Adjusted } I = \frac{I - E(I)}{\max(I) - E(I)},
\]

(7)

where \( I \) is a generic index and \( E(I) \) and \( \max(I) \) denote its expectation under statistical independence and its maximum, respectively. Then, we define the adjusted concordance index as follows:

\[
\text{ACI} = \frac{NDC - \overline{NDC}}{1 - \overline{NDC}},
\]

(8)
where $\bar{NDC}$ is the mean value of the NDC over all permutations. Here, we are interested in the reflexivity property of an index, as it must return an unequivocal attainable maximum value. For the purposes of this paper, we focus on those indexes having $\max(I) = 1$.

For a better understanding of our point of view, let us consider the following toy example in which there are two random fuzzy (probabilistic) partitions of a set of $n = 7$ objects:

$$
P = \begin{pmatrix}
0.626 & 0.374 \\
0.536 & 0.464 \\
0.271 & 0.729 \\
0.715 & 0.285 \\
0.074 & 0.926 \\
0.357 & 0.643 \\
0.317 & 0.683
\end{pmatrix};
Q = \begin{pmatrix}
0.601 & 0.399 \\
0.410 & 0.590 \\
0.571 & 0.429 \\
0.562 & 0.438 \\
0.365 & 0.635 \\
0.347 & 0.653 \\
0.499 & 0.501
\end{pmatrix}.
$$

The fuzzy Rand indexes were first computed between the two partitions and then, in turn, between partition $P$ and itself and between $Q$ and itself. The results are reported in Table 1.

Note that except for the NDC (Hüllermeier et al., 2012 in Table 1), no index achieves a unique maximum value in comparing a partition with another identical partition. Moreover, the fuzzy Rand index of Brouwer (2009) between the two partitions $P$ and $Q$ is larger than the same index computed between partition $P$ and itself. The same occurs for the fuzzy Rand index of Anderson et al. (2010), for which the results tend to be larger when comparing partitions $P$ and $Q$ than when comparing $Q$ with itself. The results of this toy example show that indexes that do not satisfy the reflexivity property may produce inconsistent values when the purpose is to build a corrected index by means of Eq. 7, which turns out to be a desirable goal. Thus, the indexes that satisfy such a property appear to be more consistent with the class of indexes corrected for chance, as defined by Hubert and Arabie (1985), that explicitly refer to Eq. 7. This consideration leads us to not consider the proposals of Campello (2007), Frigui et al. (2007), Brouwer (2009), and Anderson et al. (2010) as raw fuzzy Rand indexes. Hence, the NDC index turns out to be a suitable choice for our purposes.

However, reflexivity is only one of the desirable properties according to our viewpoint. Indeed, two partitions can be similar, but their similarity could be due to chance.

In a recent work (Suleman 2017), the quantities $a$, $b$, $c$, and $d$ as defined in Eq. 6 have been used in a comparative study to adjust the NDC index by using the formulation expressed in Eq. 2, which corresponds to adjusting the NDC by means of Eq. 1. However, the use of this formulation is not straightforward.

| Fuzzy Rand index (FR) version | $FR_{P,Q}$ | $FR_{P,P}$ | $FR_{Q,Q}$ |
|-----------------------------|------------|------------|------------|
| Campello (2007)             | 0.505      | 0.572      | 0.531      |
| Brouwer (2009)              | 0.820      | 0.787      | 0.923      |
| Anderson et al. (2010)      | 0.418      | 0.440      | 0.417      |
| Hüllermeier et al. (2012)   | 0.819      | 1.000      | 1.000      |
Considering the two previously introduced random fuzzy partitions $\mathbf{P}$ and $\mathbf{Q}$, $\mathbf{E}_\mathbf{P}$ and $\mathbf{E}_\mathbf{Q}$ correspond to:

$$\mathbf{E}_\mathbf{P} = \begin{pmatrix}
1.00 & 0.91 & 0.64 & 0.91 & 0.44 & 0.73 & 0.69 \\
0.91 & 1.00 & 0.73 & 0.82 & 0.54 & 0.82 & 0.78 \\
0.64 & 0.73 & 1.00 & 0.55 & 0.80 & 0.91 & 0.95 \\
0.91 & 0.82 & 0.55 & 1.00 & 0.35 & 0.64 & 0.60 \\
0.44 & 0.54 & 0.80 & 0.35 & 1.00 & 0.71 & 0.75 \\
0.73 & 0.82 & 0.91 & 0.64 & 0.71 & 1.00 & 0.96 \\
0.69 & 0.78 & 0.95 & 0.60 & 0.75 & 0.96 & 1.00
\end{pmatrix};$$

$$\mathbf{E}_\mathbf{Q} = \begin{pmatrix}
1.00 & 0.80 & 0.97 & 0.96 & 0.76 & 0.74 & 0.89 \\
0.80 & 1.00 & 0.83 & 0.84 & 0.95 & 0.93 & 0.91 \\
0.97 & 0.83 & 1.00 & 0.99 & 0.79 & 0.77 & 0.92 \\
0.96 & 0.84 & 0.99 & 1.00 & 0.80 & 0.78 & 0.93 \\
0.76 & 0.95 & 0.79 & 0.80 & 1.00 & 0.98 & 0.86 \\
0.74 & 0.93 & 0.77 & 0.78 & 0.98 & 1.00 & 0.84 \\
0.89 & 0.91 & 0.92 & 0.93 & 0.86 & 0.84 & 1.00
\end{pmatrix},$$

resulting in $NDC = 0.8179$.

Computing the adjustment of the NDC by using Eq. 2 provides a value of 0.6201, which is in turn substituted into Eq. 1 that, when solved for the expected value of the NDC, returns a value of 0.5208. In contrast, when considering all possible permutations of the $n$ elements of the membership vectors of one of the two partitions and keeping the other fixed, the following summary measures for the distribution of the NDC are obtained: $\min=0.7949$, $Q_1=0.8107$, $Q_2=0.8162$, mean=0.8189, $Q_3=0.8240$, and $\max=0.8549$, where $Q_i$ represents the $i$-th quartile. Note that the measure that is supposed to be the expected value of the index is not included in the support of the distribution of the permuted NDC since 0.5208 is much smaller than the minimum value of 0.7949.

It is worth stressing that all possible permutations (i.e., $7!=5040$) were used in this toy example. This example led us to consider the meaning of using Eq. 2 in correcting the NDC index. Specifically, this index does not appear to be adjusted with respect to its expectation.

An intensive Monte Carlo study conducted by Suleman (2017) showed that the version of the NDC adjusted by using Eq. 2, henceforth denoted $ARI_{HA}$, exhibits good behavior as an external fuzzy clustering validation criterion and, in some circumstances, functions better than a selection of other indexes built starting from the fuzzy cardinalities as reported in Eq. 6 (i.e., the fuzzy extension of the Jaccard index, the Dice index, and the Fowlkes and Mallows index). The use of Eq. 2 would mean correcting the index according to the original idea of Hubert and Arabie. However, it remains unclear with respect to what such normalization is applied.

In this regard, to ensure a proper interpretation of the correction for agreement due to chance (i.e., we must know with respect to what the index is corrected), we estimate the expected value of the NDC by considering the average value of the index after permuting the $n$ membership vectors of one of the two partitions. This procedure allows one to preserve the fuzzy (probabilistic) cluster composition of the perturbed partition and thus to proceed as in the case of crispy clustering when it is assumed that both the row and column marginals of the contingency table are kept fixed. In other words, the exchangeability condition is preserved (see, for example, Pesarin and Salmaso, 2010b). For the toy example presented...
above, the mean value of the NDC, considering all possible permutations, corresponds to 0.8189. Hence:

$$ACI = \frac{0.8179 - 0.8189}{1 - 0.8189} = -0.0054,$$

Similarly to Hubert’s ARI, negative values of the ACI are possible but not interesting since they indicate less agreement than expected by chance, and the index can be set equal to zero. On the other hand, another desirable property that a generic association index should have, namely, a minimum value of minus one independent of the marginal distributions (Warrens 2008a), is not required for this class of indexes. It is worth stressing that the NDC can be corrected because, to the best of our knowledge, it is the only fuzzy extension of the Rand index that possesses the reflexivity property. Note that it is easily verifiable that the resulting ACI is itself a reflexive index: indeed, it equals 1 if and only if the raw index achieves its maximum, meaning that the two partitions are identical. The same procedure can also be applied to estimate the expected values of the other considered indexes. The problem is that since they are not reflexive, the resulting adjusted indexes are not reflexive either.

6 Experimental Evaluation

This section is dedicated to evaluating the proposed index through simulated data.

6.1 Determination of the Number of Random Permutations

Theoretically, the identification of the mean value in Eq. 8 is based on considering all possible permutations ($n!$). When determining all of the permutations is not practical, for example, when $n > 20$, the expected value is estimated by taking into account a very large number $h$ of randomly selected permutations of the total $n!$ permutations. Our approach is similar to the permutation testing procedure (Hoeffding 1952; Pesarin and Salmaso 2010a), but no hypothesis testing is performed. Our goal is to obtain a point estimate of the expected value of the NDC when, given two partitions, the membership probabilities are by chance assigned to preserve the fuzzy clustering structure. This can be accomplished with “without replacement resampling” (Pesarin and Salmaso 2010b) under the invariance of the fuzzy cluster structure of the partition that is perturbed. An experiment to evaluate whether the choice of the number of permutations can influence the results was conducted. A total number of 900 data sets were generated: 300 data sets with a sample size equal to 100, 300 data sets with a sample size equal to 500, and 300 data sets with a sample size equal to 1000. The data were generated by sampling from a multivariate normal distribution with a number of dimensions randomly chosen between 2 and 10. The mean vector was randomly sampled from a uniform distribution between 10 and 30. The covariance matrix was set as a diagonal matrix with uniformly random numbers between 0.1 and 3. A random number of clusters between 2 and 10 were randomly chosen for each data set, storing the crisp clustering composition. The K-means algorithm was run, and for each data set, both the ARI of Hubert and Arabie and the ACI were computed. The latter was computed five times with the following different numbers of random permutations: 100, 500, 1000, 5000, and 10,000. The difference between the ACI and ARI was then computed. The results are summarized in Table 2.
Table 2  Summary of the differences between the ACI and ARI partitioned by sample size and number of random permutations

| Sample size | Number of random permutations | \( Q_1 \) | \( Q_2 \) | Mean | \( Q_3 \) |
|-------------|-------------------------------|---------|---------|------|---------|
| 100         | 100                           | -0.0004 | 0.0000  | -0.0002 | 0.0000 |
| 500         | 1000                          | -0.0001 | 0.0000  | -0.0001 | 0.0000 |
| 100         | 5000                          | -0.0001 | 0.0000  | 0.0000  | 0.0000 |
| 1000        | 10000                         | 0.0000  | 0.0000  | 0.0000  | 0.0000 |
| 100         | 500                           | -0.0005 | 0.0000  | -0.0002 | 0.0000 |
| 500         | 1000                          | -0.0001 | 0.0000  | -0.0001 | 0.0000 |
| 500         | 5000                          | -0.0001 | 0.0000  | 0.0000  | 0.0000 |
| 1000        | 10000                         | -0.0001 | 0.0000  | -0.0001 | 0.0000 |
| 100         | 100                           | -0.0006 | 0.0000  | -0.0006 | 0.0000 |
| 500         | 500                           | -0.0003 | 0.0000  | -0.0001 | 0.0000 |
| 1000        | 1000                          | -0.0001 | 0.0000  | 0.0000  | 0.0000 |
| 100         | 500                           | -0.0001 | 0.0000  | 0.0000  | 0.0000 |
| 1000        | 10000                         | -0.0001 | 0.0000  | 0.0000  | 0.0000 |

The results of this Monte Carlo experiment indicate an almost insignificant negative bias that tends to become smaller as the number of random permutations increases. It can also be noted that the sample size does not affect the final results. It seems that in real analysis, 5000 random permutations are sufficient to achieve good results.

6.2 Comparing Two Fuzzy Partitions

The following simulation studies aim to investigate the behavior of the proposed adjusted concordance index (ACI) and compare it with the adjusted NDC using Eq. 2, as used in (Suleman 2017) (ARI_{HA}). For the sake of completeness, the comparison is also made with respect to the indexes proposed by Brouwer (2009), Campello (2007), and Anderson et al. (2010), henceforth denoted Brouwer_{HA}, Campello_{HA}, and Anderson_{HA}, indicating that the adjustment is made according to Eq. 2, as stated by the authors. It is worth noting that we cannot adjust these indexes according to our procedure because none of them is reflexive. We would like to note that we do not evaluate clustering algorithms.

6.2.1 Comparing Two Fuzzy Partitions: Clustered Data

For the first simulation study, six data sets with \( K = 2, 3, \ldots, 7 \) cluster centers were generated by incrementally merging seven different bivariate normal distributions with the following mean vectors: \( \mu_1 = (-2, -2) \), \( \mu_2 = (2, 2) \), \( \mu_3 = (0, 0) \), \( \mu_4 = (-2, 2) \), \( \mu_5 = (2, -2) \), \( \mu_6 = (-4, 4) \), and \( \mu_7 = (4, -4) \). The covariance matrix is equal to \( \alpha' \times I \), where \( \alpha \) is a vector composed of two draws from a uniform distribution in \([0.1, 1]\). The structure of the data sets is presented in Table 3. The sample size was set to 120.

For each data set, a kind of “true” fuzzy probabilistic partition was computed through the probabilistic-distance (PD) clustering of Ben-Israel and Iyigun (2008) and compared with
Table 3  Second simulation study: data set structure

| Data set number | Centers     | Size within clusters |
|-----------------|-------------|----------------------|
| 1               | $\mu_1, \mu_2$ | 60, 60               |
| 2               | Add $\mu_3$ to data set 1 | 40, 40, 40          |
| 3               | Add $\mu_4$ to data set 2 | 30, 30, 30, 30      |
| 4               | Add $\mu_5$ to data set 3 | 20, 15, 15, 30, 40  |
| 5               | Add $\mu_6$ to data set 4 | 15, 15, 15, 20, 30, 25 |
| 6               | Add $\mu_7$ to data set 5 | 17, 16, 16, 14, 17, 18, 18 |

the estimated one. It is worth noting that in this experiment, the PD framework was used to generate a kind of true probabilistic partition, but these partitions cannot be considered true by any means. We do not assert that the PD clustering aims to discover true probabilistic partitions. The ground truth should be deterministic and known a priori. In practice, one can hardly find true fuzzy partitions in real-life data sets. With this premise and the generated data sets, pseudo-real partitions have been generated by using the Euclidean distance and applying the probabilistic-distance method.

For each data set, the PD clustering algorithm was run by setting $K = 2, 3, \ldots, 7$. Then, all indexes under evaluation between the estimated probabilistic partitions and the true probabilistic partitions were computed for the six data sets. The results are displayed in Table 4.

The highest value of the ACI is sometimes not in the cell we would expect it to be in (i.e., data sets 4 and 6). It should also be noted that the values of this index are generally high and similar to each other. The same holds for $\text{ARI}_{HA}$ and, with significantly lower values, the other considered indexes. However, it is worth stressing that the values of the ACI are more consistent with the nature of the experiment for two reasons: First, the range of variation of the index is larger with respect to the others. Second, the index is noticeably lower in situations in which the true probabilistic partition is expected to differ from the estimated one. Note also that the index of Anderson ($\text{Anderson}_{HA}$) provides the smallest values among the considered indexes, which, in many cases, are almost equal to zero. In addition, the value of Anderson’s index is not the highest when the two partitions perfectly agree, as in the cases of data sets 1 and 6.

6.2.2 Comparing Fuzzy Partitions: Homogeneous Data

Drawing from an experiment proposed in Albatineh and Niewiadomska-Bugaj (2011), for the second simulation study, a data set was generated by randomly sampling from a bivariate normal distribution with parameters $\mu = \begin{pmatrix} 10 \\ 10 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 1.00 & 0.25 \\ 0.25 & 1.00 \end{pmatrix}$, with a sample size of 500. The statistical units were then assigned to the clusters according to a multinomial sampling scheme with probability vector equal to $\begin{pmatrix} 1/2 \\ 1/2 \\ 1/3 \\ 1/3 \\ 1/3 \\ 1/8 \\ 1/8 \\ 1/8 \\ 1/8 \end{pmatrix}$ for $C = 2, 3, \ldots, 7$ clusters, respectively. For each data set, the true probabilistic partitions were generated as described in the previous section, and the PD clustering algorithm was then applied by setting $C = 2, \ldots, 7$. Table 5 shows the behavior of the five indexes under evaluation.

The first column highlights that when comparing a partition with itself, the ACI and $\text{ARI}_{HA}$ are always equal to the theoretical maximum (i.e., 1). This situation derives from the
Table 4 Comparing two fuzzy partitions: clustered data

| Data set | Adjusted index | raw | True vs | C=2 | C=3 | C=4 | C=5 | C=6 | C=7 |
|----------|----------------|-----|---------|-----|-----|-----|-----|-----|-----|
| 1        | ACI            | 1.000 | 0.9501 | 0.7479 | 0.7579 | 0.7093 | 0.6352 | 0.5847 |
|          | ARI<sub>HA</sub> | 1.000 | 0.9641 | 0.8251 | 0.8251 | 0.7874 | 0.7298 | 0.6943 |
|          | Brouwer<sub>HA</sub> | 0.5064 | 0.5058 | 0.4508 | 0.4444 | 0.4243 | 0.3968 | 0.3811 |
|          | Campello<sub>HA</sub> | 0.4620 | 0.4552 | 0.2596 | 0.1562 | 0.1332 | 0.1183 | 0.1097 |
|          | Anderson<sub>HA</sub> | 0.2672 | 0.2680 | 0.1529 | 0.1263 | 0.1077 | 0.0942 | 0.0818 |
| 2        | ACI            | 1.000 | 0.5369 | 0.7901 | 0.6056 | 0.5638 | 0.4271 | 0.6081 |
|          | ARI<sub>HA</sub> | 1.000 | 0.7360 | 0.8797 | 0.7744 | 0.7594 | 0.6823 | 0.7635 |
|          | Brouwer<sub>HA</sub> | 0.3204 | 0.2441 | 0.3201 | 0.2896 | 0.2668 | 0.2413 | 0.2771 |
|          | Campello<sub>HA</sub> | 0.2847 | 0.1330 | 0.2706 | 0.2068 | 0.1476 | 0.1224 | 0.0959 |
|          | Anderson<sub>HA</sub> | 0.2847 | 0.1330 | 0.2706 | 0.2068 | 0.1476 | 0.1224 | 0.0959 |
| 3        | ACI            | 1.000 | 0.5005 | 0.7101 | 0.6056 | 0.5638 | 0.4271 | 0.6081 |
|          | ARI<sub>HA</sub> | 1.000 | 0.7360 | 0.8797 | 0.7744 | 0.7594 | 0.6823 | 0.7635 |
|          | Brouwer<sub>HA</sub> | 0.2867 | 0.1405 | 0.2188 | 0.2730 | 0.2573 | 0.2035 | 0.2382 |
|          | Campello<sub>HA</sub> | 0.2843 | 0.0682 | 0.1543 | 0.2597 | 0.2170 | 0.1464 | 0.1589 |
|          | Anderson<sub>HA</sub> | 0.0522 | 0.0261 | 0.0402 | 0.0481 | 0.0334 | 0.0164 | 0.0182 |
| 4        | ACI            | 1.000 | 0.5363 | 0.7183 | 0.5183 | 0.4588 | 0.4623 | 0.5478 |
|          | ARI<sub>HA</sub> | 1.000 | 0.6351 | 0.7078 | 0.7623 | 0.7550 | 0.7100 | 0.7185 |
|          | Brouwer<sub>HA</sub> | 0.2679 | 0.1309 | 0.1806 | 0.2448 | 0.2149 | 0.2346 | 0.2631 |
|          | Campello<sub>HA</sub> | 0.2868 | 0.0813 | 0.1473 | 0.2300 | 0.2405 | 0.1771 | 0.2174 |
|          | Anderson<sub>HA</sub> | 0.0238 | 0.0238 | 0.0255 | 0.0341 | 0.0243 | 0.0138 | 0.0112 |
| 5        | ACI            | 1.000 | 0.4537 | 0.5063 | 0.5909 | 0.6181 | 0.3792 | 0.6689 |
|          | ARI<sub>HA</sub> | 1.000 | 0.5934 | 0.7078 | 0.7365 | 0.7799 | 0.7933 | 0.6689 |
|          | Brouwer<sub>HA</sub> | 0.2704 | 0.1237 | 0.1774 | 0.1899 | 0.2339 | 0.2666 | 0.1808 |
|          | Campello<sub>HA</sub> | 0.3535 | 0.0661 | 0.1406 | 0.1641 | 0.2375 | 0.3254 | 0.1875 |
|          | Anderson<sub>HA</sub> | 0.0250 | 0.0237 | 0.0298 | 0.0132 | 0.0255 | 0.0310 | 0.0047 |
| 6        | ACI            | 1.000 | 0.3802 | 0.5312 | 0.3624 | 0.4286 | 0.2921 | 0.6537 |
|          | ARI<sub>HA</sub> | 1.000 | 0.5105 | 0.6342 | 0.7261 | 0.6334 | 0.6710 | 0.5637 |
|          | Brouwer<sub>HA</sub> | 0.2751 | 0.0752 | 0.1641 | 0.2437 | 0.1805 | 0.1999 | 0.1507 |
|          | Campello<sub>HA</sub> | 0.3349 | 0.0360 | 0.0954 | 0.1836 | 0.1760 | 0.1922 | 0.1450 |
|          | Anderson<sub>HA</sub> | 0.0183 | 0.0048 | 0.0192 | 0.0390 | 0.0107 | 0.0098 | −0.0158 |

properties of the NDC index. In all other cases, the considered indexes attain values very close to zero (indicating that even if the partitions are similar, this closeness can be due to chance). This is true except for ARI<sub>HA</sub>, with a value of approximately 0.5 in each case, and there are no cases in which it can suggest significant discordance between 2 partitions. This result must be combined with the results shown in Table 4: when data are structured, there is no great difference between the ACI and ARI<sub>HA</sub>, a consideration that could lead the reader to conclude that the two indexes behave in essentially the same manner. On the other hand, from Table 5 it is clear that only the ACI suggests that the two partitions are far from similar. In this case, considering all data sets, ARI<sub>HA</sub> returns values within the range of 0.4564 to 0.6446, suggesting good similarity between the partitions.
### Table 5 Comparing two fuzzy partitions: homogeneous data

| Center number | Adjusted raw index | True vs true | C=2   | C=3   | C=4   | C=5   | C=6   | C=7   |
|---------------|--------------------|--------------|-------|-------|-------|-------|-------|-------|
| 2             | ACI                | 1.0000       | 0.0484| 0.0676| 0.0127| 0.0242| 0.0117| 0.0100|
|               | ARI\textsubscript{HA} | 1.0000       | 0.6147| 0.6019| 0.5197| 0.5224| 0.4929| 0.4912|
|               | Brouwer\textsubscript{HA} | 0.0710       | 0.0138| 0.0190| 0.0044| 0.0118| 0.0068| 0.0042|
|               | Campello\textsubscript{HA} | 0.0252       | 0.0178| 0.0035| 0.0017| 0.0011| 0.0006| 0.0002|
|               | Anderson\textsubscript{HA} | −0.0017      | −0.0011| −0.0022| −0.0025| −0.0027| −0.0030| −0.0031|
| 3             | ACI                | 1.0000       | 0.0327| 0.0515| 0.0035| 0.0054| 0.0026| 0.0003|
|               | ARI\textsubscript{HA} | 1.0000       | 0.5813| 0.5663| 0.4853| 0.4834| 0.4588| 0.4564|
|               | Brouwer\textsubscript{HA} | 0.1166       | 0.0127| 0.0206| 0.0060| 0.0034| 0.0021| 0.0017|
|               | Campello\textsubscript{HA} | 0.0141       | 0.0051| 0.0091| 0.0077| 0.0038| 0.0021| 0.0010|
|               | Anderson\textsubscript{HA} | −0.0040      | −0.0021| −0.0037| −0.0045| −0.0051| −0.0055| −0.0058|
| 4             | ACI                | 1.0000       | 0.0536| 0.0892| 0.0152| 0.0311| 0.0279| 0.0113|
|               | ARI\textsubscript{HA} | 1.0000       | 0.6419| 0.6373| 0.5504| 0.5555| 0.5304| 0.5218|
|               | Brouwer\textsubscript{HA} | 0.0993       | 0.0154| 0.0247| 0.0061| 0.0145| 0.0127| 0.0060|
|               | Campello\textsubscript{HA} | 0.0211       | 0.0018| 0.0036| 0.0126| 0.0153| 0.0119| 0.0087|
|               | Anderson\textsubscript{HA} | −0.0058      | −0.0024| −0.0044| −0.0055| −0.0064| −0.0071| −0.0077|
| 5             | ACI                | 1.0000       | 0.0121| 0.0345| 0.0074| 0.0097| 0.0055| 0.0052|
|               | ARI\textsubscript{HA} | 1.0000       | 0.5891| 0.5730| 0.4989| 0.4969| 0.4707| 0.4697|
|               | Brouwer\textsubscript{HA} | 0.0415       | 0.0042| 0.0128| 0.0029| 0.0053| 0.0033| 0.0028|
|               | Campello\textsubscript{HA} | 0.0122       | 0.0002| 0.0004| 0.0045| 0.0080| 0.0095| 0.0084|
|               | Anderson\textsubscript{HA} | −0.0081      | −0.0031| −0.0052| −0.0067| −0.0079| −0.0088| −0.0096|
| Center number | Adjusted raw index | True vs true | C=2  | C=3  | C=4  | C=5  | C=6  | C=7  |
|---------------|--------------------|--------------|------|------|------|------|------|------|
| 6             | ACI                | 1.0000       | 0.0509| 0.1149| 0.0213| 0.0424| 0.0206| 0.0125|
|               | ARI<sub>HA</sub>   | 1.0000       | 0.6372| 0.6446| 0.5513| 0.5587| 0.5257| 0.5210|
|               | Brouwer<sub>HA</sub> | 0.1442       | 0.0227| 0.0382| 0.0099| 0.0275| 0.0129| 0.0088|
|               | Campello<sub>HA</sub> | 0.0263       | 0.0005| 0.0004| 0.0042| 0.0112| 0.0163| 0.0173|
|               | Anderson<sub>HA</sub> | −0.0099    | −0.0027| −0.0054| −0.0070| −0.0084| −0.0096| −0.0106|
| 7             | ACI                | 1.0000       | 0.0238| 0.0739| 0.0038| 0.0102| 0.0025| 0.0021|
|               | ARI<sub>HA</sub>   | 1.0000       | 0.5900| 0.5876| 0.4969| 0.4969| 0.4695| 0.4686|
|               | Brouwer<sub>HA</sub> | 0.1645       | 0.0082| 0.0349| 0.0027| 0.0071| 0.0015| 0.0024|
|               | Campello<sub>HA</sub> | 0.0143       | 0.0001| 0.0001| 0.0012| 0.0026| 0.0069| 0.0093|
|               | Anderson<sub>HA</sub> | −0.0121    | −0.0032| −0.0058| −0.0079| −0.0095| −0.0109| −0.0120|
7 Concluding Remarks

In this paper, we propose an adjusted version of the normalized degree of concordance index (NDC) defined by Hüllermeier et al. (2012) for comparing fuzzy partitions, called the adjusted concordance index (ACI). The proposed measure is constructed based on a reasoning similar to that for the well-known adjusted Rand index of Hubert and Arabie (1985), and it is derived by normalizing the difference between the NDC and the point estimate of its expected value obtained by considering a large number of permutations of the membership vector of one of the two fuzzy partitions. The rationale of our approach derives from the consideration that adjusting the NDC by using Eq. 2 leads to unclear results in the sense that this formulation does not adjust the index as one would expect, that is, correcting it by providing a baseline using the expected similarity of all pairwise comparisons between partitions specified by a random model. The condition under which our approach functions is the invariance of the fuzzy cluster composition of a partition under random permutations of the membership vectors.

Our proposal theoretically works with any raw fuzzy index, provided that two “desiderata” are satisfied: the reflexivity and a proper interpretation of correction of the index for agreement due to chance. The latter property is obtainable either analytically or (as in our case) through appropriate resampling methods.

A simulation study was conducted in order to compare our proposal with the fuzzy version of the NDC index, as already used by Suleman (2017), for the goal of comparing two fuzzy partitions of the same data sets. Moreover, for the sake of completeness, the comparison was also made with other existing indexes proposed in the literature, even though these latter indexes cannot be adjusted according to our procedure because they are not reflexive. The results highlight that when data sets are well structured, there is no great difference in the behavior of the ACI and $\text{ARI}_{HA}$. When the data are instead unstructured, $\text{ARI}_{HA}$ returns values that suggest a high degree of similarity between the partitions, whereas the ACI always returns values near zero. For real problems for which the structure of data sets is not known a priori, the use of the ACI is preferable. In other words, $\text{ARI}_{HA}$ always suggests that two fuzzy partitions are similar.

As adjustment of the NDC by means of Eq. 2 is extensively used (Suleman 2017), the main issues concern both the measure with respect to which the NDC is normalized and the meaning of Eq. 2 in producing the adjusted NDC. On the other hand, it can be noted that Suleman has always compared fuzzy and crisp partitions in his work. In any case, it is clear that the correction of the NDC by using Eq. 2 is not a “correction-for-chance” as originally thought.

Recently, Warrens and van der Hoef (2019) noted that various external validation indexes for hard clustering based on the pair counting approach (for example, Rand, ARI, and Jaccard) are sensitive to cluster size imbalance. Hence, further research may consider how the ideas in their work can also be applied to the fuzzy extensions of the Rand index and adjusted Rand index.

Acknowledgments The authors would like to thank both the Editor and an anonymous reviewer, whose comments and remarks highly contributed to improve the quality of this manuscript.

References

Albatineh, A.N., & Niewiadomska-Bugaj, M. (2011). Correcting Jaccard and other similarity indices for chance agreement in cluster analysis. Advances in Data Analysis and Classification, 5(3), 179–200.
Albatineh, A.N., Niewiadomska-Bugaj, M., Mihalko, D. (2006). On similarity indices and correction for chance agreement. *Journal of Classification*, 23(2), 301–313.
Anderberg, M.R. (1973). *Cluster analysis for applications: probability and mathematical statistics: a series of monographs and textbooks*, 1st edn. New York: Academic press.
Anderson, D.T., Bezdek, J.C., Popescu, M., Keller, J.M. (2010). Comparing fuzzy, probabilistic, and possibilistic partitions. *IEEE Transactions on Fuzzy Systems*, 18(5), 906–918.
Ben-Israel, A., & Iyigun, C. (2008). Probabilistic d-clustering. *Journal of Classification*, 25(1), 5–26.
Berkhin, P. (2006). A survey of clustering data mining techniques, in Grouping multidimensional data. In Kogan, J., Nicholas, C., Teboulle, M. (Eds.) (pp. 25–71). Berlin: Springer.
Bezdek, J.C., Ehrlich, R., Full, W. (1984). FCM: the Fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2), 191–203.
Böck, H.H. (1974). *Automatische Klassifikation*, 1st edn. Göttingen: Vandenhoeck & Ruprecht.
Brouwer, R.K. (2009). Extending the Rand, adjusted Rand and Jaccard indices to fuzzy partitions. *Journal of Intelligent Information Systems*, 32(3), 213–235.
Campello, R.J. (2007). A fuzzy extension of the Rand index and other related indices for clustering and classification assessment. *Pattern Recognition Letters*, 28(7), 833–841.
Dice, L.R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3), 297–302.
Downton, M., & Brennan, T. (1980). Comparing classifications: an evaluation of several coefficients of partition agreement. *Classification Society Bulletin*, 4(4), 53–54.
Duran, B.S., & Odell, P.L. (2013). *Cluster analysis: a survey*, 2nd edn. Heidelberg: Springer Science & Business Media.
D’Urso, P. (2015). Fuzzy clustering, in Handbook of cluster analysis. In Hennig, C., Meila, M., Murtagh, F., Rocci, R. (Eds.) (pp. 545–574). Boca Raton: CRC Press, chap. 24.
Everitt, B.S., Landau, S., Leese, M., Stahl, D. (2011). *Cluster analysis*, 5th edn. Chichester: Wiley.
Fasulo, D. (1999). *An analysis of recent work on clustering algorithms*. Department of Computer Science & Engineering, University of Washington. Available at https://citeseer.ist.psu.edu/viewdoc/download?doi=10.1.1.44.2946&rep=rep1&type=pdf.
Fowlkes, E.B., & Mallows, C.L. (1983). A method for comparing two hierarchical clusterings. *Journal of the American Statistical Association*, 78(383), 553–569.
Frigui, H., Hwang, C., Rhee, F.C.-H. (2007). Clustering and aggregation of relational data with applications to image database categorization. *Pattern Recognition*, 40(11), 3053–3068.
Gower, J.C., & Legendre, P. (1986). Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification*, 3(1), 5–48.
Halkidi, M., Vazigiannis, M., Hennig, C. (2015). Method-independent indices for cluster validation and estimating the number of clusters. In Hennig, C., Meila, M., Murtagh, F., Rocci, R. (Eds.) *Handbook of cluster analysis*, chap. 26 (pp. 595–618). Boca Raton: CRC Press.
Hamann, U. (1961). *Merkmalenbestand und Verwandtschaftsbeziehungen der Farinosae: ein beitrag zum system der Monokotyledonen (639–768). Willdenowia*.
Han, J., Pei, J., Kamber, M. (2012). *Data mining: concepts and techniques*, 3rd edn. Amsterdam: Elsevier.
Hartigan, J.A. (1975). *Clustering algorithms*, 1st edn. New York: Wiley.
Hennig, C., & Meila, M. (2015). Cluster analysis: an overview, in Handbook of cluster analysis. In Hennig, C., Meila, M., Murtagh, F., Rocci, R. (Eds.) (pp. 1–20). Boca Raton: CRC Press, chap. 1.
Hoeffding, W. (1952). The large-sample power of tests based on permutations of observations. *The Annals of Mathematical Statistics*, 169–192.
Höppner, F., Klawonn, F., Kruse, R., Runkler, T. (1999). *Fuzzy cluster analysis: methods for classification, data analysis and image recognition*, 1st edn. Chichester: Wiley.
Hubert, L. (1977). Nominal scale response agreement as a generalized correlation. *British Journal of Mathematical and Statistical Psychology*, 30(1), 98–103.
Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2(1), 193–218.
Hullermeier, E., Rifqi, M., Henzgen, S., Senge, R. (2012). Comparing fuzzy partitions: a generalization of the Rand index and related measures. *IEEE Transactions on Fuzzy Systems*, 20(3), 546–556.
Jaccard, P. (1901). Distribution de la Flore Alpine: dans le Bassin des dranses et dans quelques régions voisines. *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37(140), 241–272.
Jain, A.K., & Dubes, R.C. (1988). *Algorithms for clustering data*, 1st edn. Englewood Cliffs: Prentice-Hall, Inc.
Jain, A.K., Murty, M.N., Flynn, P.J. (1999). Data clustering: a review. *ACM Computing Surveys (CSUR)*, 31(3), 264–323.
Kaufman, L., & Rousseeuw, P.J. (2005). *Finding groups in data: an introduction to cluster analysis*, 2nd edn. Hoboken: Wiley.
Klement, E.P., Mesiar, R., Pap, E. (2010). *Triangular norms*, 1st edn. Dordercht: Springer Science & Business Media.

Kulczynski, S. (1927). Die pflanzenassocationen der pienenen. *Bulletin International de l’académie Polonaise des Sciences et des lettres, classe des sciences mathématiques et naturelles, Serie B Supplement, II*, 2, 57–203.

Meilă, M. (2007). Comparing clusterings - an information based distance. *Journal of Multivariate Analysis*, 98(5), 873–895.

Mirkin, B. (1998). Mathematical classification and clustering: from how to what and why. In Balderjahn, I., Mathar, R., Schader, M. (Eds.) *Classification, data analysis, and data highways* (pp. 172–181). Heidelberg: Springer.

Morey, L.C., & Agresti, A. (1984). The measurement of classification agreement: an adjustment to the Rand statistic for chance agreement. *Educational and Psychological Measurement*, 44(1), 33–37.

Pesarin, F., & Salmaso, L. (2010a). The permutation testing approach: a review. *Statistica*, 70(4), 481–509.

Pesarin, F., & Salmaso, L. (2010b). *Permutation tests for complex data: theory, applications and software*, 1st edn. Chippenham: Wiley.

Rand, W.M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336), 846–850.

Ruspini, E.H. (1970). Numerical methods for fuzzy clustering. *Information Sciences*, 2(3), 319–350.

Spath, H. (1980). *Cluster analysis algorithms for data reduction and classification of objects*, 1st edn. Chichester: Ellis Horwood, Ltd.

Stahl, D., & Sallis, H. (2012). Model-based cluster analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(4), 341–358.

Suleman, A. (2017). Assessing a fuzzy extension of Rand index and related measures. *IEEE Transactions on Fuzzy Systems*, 25(1), 237–244.

Warrens, M.J. (2008a). On association coefficients for $2 \times 2$ tables and properties that do not depend on the marginal distributions. *Psychometrika*, 73(4), 777–789.

Warrens, M.J. (2008b). On the equivalence of Cohen’s kappa and the Hubert-Arabie adjusted Rand index. *Journal of Classification*, 25(2), 177–183.

Warrens, M.J., & van der Hoef, H. (2019). Understanding partition comparison indices based on counting object pairs. Available at arXiv:1901.01777.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.