



**HAL**  
open science

## Comparing Fuzzy Partitions: A Generalization of the Rand Index and Related Measures

Eyke Hüllermeier, Maria Rifqi, Sascha Henzgen, Robin Senge

► **To cite this version:**

Eyke Hüllermeier, Maria Rifqi, Sascha Henzgen, Robin Senge. Comparing Fuzzy Partitions: A Generalization of the Rand Index and Related Measures. *IEEE Transactions on Fuzzy Systems*, 2012, 20 (3), pp.546 - 556. 10.1109/TFUZZ.2011.2179303 . hal-00734389

**HAL Id: hal-00734389**

**<https://hal.science/hal-00734389>**

Submitted on 21 Sep 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Comparing Fuzzy Partitions: A Generalization of the Rand Index and Related Measures

Eyke Hüllermeier, Maria Rifqi, Sascha Henzgen, and Robin Senge

**Abstract**—In this paper, we introduce a fuzzy extension of a class of measures to compare clustering structures, namely, measures that are based on the number of concordant and the number of discordant pairs of data points. This class includes the well-known Rand index but also commonly used alternatives, such as the Jaccard measure. In contrast with previous proposals, our extension exhibits desirable metrical properties. Apart from elaborating on formal properties of this kind, we present an experimental study in which we compare different fuzzy extensions of the Rand index and the Jaccard measure.

**Index Terms**—Clustering, distance, fuzzy partition, Jaccard index, Rand index, similarity.

## I. INTRODUCTION

THE problem of comparing two partitions of a set of objects occurs quite naturally in various domains, notably in data analysis and clustering. For example, one way to evaluate the result of a clustering algorithm is to compare the clustering structure that is produced by the algorithm with a correct partition of the data (which of course presumes that this information is available). In cluster analysis, the so-called *external evaluation measures* have been developed for this purpose [1], [2]. However, measures of that kind are not only of interest as evaluation criteria, i.e., to compare a hypothetical partition with a true one. Instead, distance measures for partitions are interesting in their own right and can be used for different purposes.

Just to give a motivating example, consider the problem to compare two different representations of the same set of objects. More concretely, in [3], the problem of clustering data in a very high dimensional space is considered. To increase efficiency, they propose to map the data into a low-dimensional space first and to cluster the transformed data thus obtained afterward. In this context, a distance measure to cluster structures (partitions) is useful to measure the loss of information that is incurred by the data transformation: If the transformation is (almost) lossless, the clustering structures in the two spaces should be highly similar, i.e., their distance should be small. On the other hand, a significant difference between the two partitions would

indicate that the transformation does have a strong effect in the sense of distorting the structure of the dataset.

Even though a large number of evaluation criteria and similarity indexes to cluster structures have been proposed in the literature, their extension to the case of fuzzy partitions has received much less attention so far. This is especially true for external evaluation criteria and measures comparing two clustering structures, whereas *internal criteria* to evaluate a single partition<sup>1</sup> have been studied more thoroughly (see, e.g., [4] and [5] for early proposals).

Nevertheless, a few measures to compare fuzzy partitions, notably extensions of the well-known Rand index, have recently been proposed in the literature. In this paper, which is an extended version of a previous conference version [6], we make another proposal for a measure of that kind, namely a fuzzy variant of the Rand index and related measures. In contrast with previous proposals, our measure satisfies desirable properties of a metric (when being used as a distance function).

The remainder of this paper is organized as follows. In the next section, we briefly recall the definition of the well-known Rand index to compare clustering structures. In Section III, we review existing approaches to compare fuzzy partitions. In Section IV, we introduce our new measure and elaborate on its formal properties. In Section V, we address the question of how to generalize our approach to other types of similarity measures. In Section VI, we compare our measure experimentally with previous proposals. This paper concludes with a short summary and an outlook on future work in Section VII.

## II. RAND INDEX

Let  $\mathbf{P} = \{P_1, \dots, P_k\} \subset 2^X$  and  $\mathbf{Q} = \{Q_1, \dots, Q_\ell\} \subset 2^X$  be two (crisp) partitions of a finite set  $X = \{x_1, x_2, \dots, x_n\}$  with  $n$  elements, which means that  $P_i \neq \emptyset$ ,  $P_i \cap P_j = \emptyset$  for all  $1 \leq i \neq j \leq k$ , and  $P_1 \cup P_2 \cup \dots \cup P_k = X$  (and analogously for  $\mathbf{Q}$ ). Let

$$C = \{(x_i, x_j) \in X \times X \mid 1 \leq i < j \leq n\}$$

denote the set of all tuples of elements in  $X$ .<sup>2</sup> We say that two elements  $(x, x') \in C$  are *paired* in  $\mathbf{P}$  if they belong to the same cluster, i.e., if there is a cluster  $P_i \in \mathbf{P}$  such that  $x \in P_i$  and  $x' \in P_i$ . Moreover, we distinguish the following subsets of  $C$ :

- $C_1 \equiv$  the set of tuples  $(x, x') \in C$  that are paired in  $\mathbf{P}$  and paired in  $\mathbf{Q}$ ;

<sup>1</sup>Typically, such criteria compare the intracluster variability, i.e., the variability among objects within the same cluster (which should be small) with the intercluster variability, i.e., the variability among objects from different clusters (which should be high).

<sup>2</sup>Since we consider unordered tuples, we should more correctly write  $\{x_i, x_j\}$  instead of  $(x_i, x_j)$ .

Manuscript received March 29, 2011; revised June 20, 2011 and August 29, 2011; accepted October 27, 2011. Date of publication December 13, 2011; date of current version.

E. Hüllermeier, S. Henzgen, and R. Senge are with the Department of Mathematics and Computer Science, University of Marburg, 35037 Marburg, Germany (e-mail: eyke@mathematik.uni-marburg.de; henzgen@mathematik.uni-marburg.de; senge@mathematik.uni-marburg.de).

M. Rifqi is with the University Pierre et Marie Curie, CNRS UMR 7606, LIP6, Paris 75005, France (e-mail: maria.rifqi@lip6.fr).

Digital Object Identifier 10.1109/TFUZZ.2011.2179303

- $C_2 \equiv$  the set of tuples  $(x, x') \in C$  that are paired in  $\mathbf{P}$  but not paired in  $\mathbf{Q}$ ;
- $C_3 \equiv$  the set of tuples  $(x, x') \in C$  that are not paired in  $\mathbf{P}$  but paired in  $\mathbf{Q}$ ;
- $C_4 \equiv$  the set of tuples  $(x, x') \in C$  that are neither paired in  $\mathbf{P}$  nor in  $\mathbf{Q}$ .

Obviously,  $\{C_1, C_2, C_3, C_4\}$  is a partition of  $C$ , and  $a + b + c + d = |C| = n(n-1)/2$ , where

$$a = |C_1|, \quad b = |C_2|, \quad c = |C_3|, \quad d = |C_4|. \quad (1)$$

The tuples  $(x, x') \in C_1 \cup C_4$  are the *concordant* pairs, i.e., the pairs for which there is agreement between  $\mathbf{P}$  and  $\mathbf{Q}$ , while the tuples  $(x, x') \in C_2 \cup C_3$  are the *discordant* pairs for which the two partitions disagree. The Rand index is defined by the number of concordant pairs that are divided by the total number of pairs:

$$R(\mathbf{P}, \mathbf{Q}) = \frac{a + d}{a + b + c + d}. \quad (2)$$

Thus defined, the Rand index is a similarity measure which assumes values between 0 and 1. It can easily be turned into a distance function by defining

$$D_R(\mathbf{P}, \mathbf{Q}) = 1 - R(\mathbf{P}, \mathbf{Q}) = \frac{b + c}{a + b + c + d}.$$

It is worth mentioning that  $D_R$  satisfies the classical properties of a distance (reflexivity, separation, symmetry, and triangular inequality).

### III. GENERALIZATIONS OF THE RAND INDEX

In this section, we briefly review existing measures that have been proposed in the literature to compare fuzzy partitions. We start with the proposals of Campello [7] and Frigui *et al.* [8], which are intimately connected in the sense that the latter can be seen as a special case of the former. Subsequently, an alternative extension of the Rand index is discussed, namely the one proposed by Brouwer [9], which is based on the idea of a “measure of bonding” between pairs of objects. While this approach is still quite similar to Campello and Frigui *et al.*, the measure that is put forward by Anderson *et al.* [10] proceeds from a different idea and takes the so-called contingency matrix of two fuzzy partitions as a point of departure. Finally, we also discuss other proposals that are not direct extensions of the Rand index.

#### A. Campello and Frigui *et al.*

In order to extend the Rand index to the case of fuzzy partitions, Campello [7] first reformulates this measure within a set-theoretic framework. An extension to the fuzzy case can then be accomplished in a straightforward way by using generalized set-theoretical operators. Recall that  $k = |\mathbf{P}|$  and  $\ell = |\mathbf{Q}|$ , and consider the following sets:

- $V \equiv$  the set of pairs  $(x, x') \in C$  that belong to the same cluster in  $\mathbf{P}$ ; it can be expressed as  $V = \bigcup_{1 \leq i \leq k} V_i$ , where  $V_i$  is the set of pairs that both belong to the  $i$ th cluster  $P_i \in \mathbf{P}$ .

- $W \equiv$  the set of pairs  $(x, x') \in C$  that belong to different clusters in  $\mathbf{P}$ ; it can be expressed as  $W = \bigcup_{1 \leq i \neq j \leq k} W_{ij}$ , where  $W_{ij}$  is the set of pairs such that  $x \in P_i$  and  $x' \in P_j$ .
- $Y \equiv$  the set of pairs  $(x, x') \in C$  that belong to the same cluster in  $\mathbf{Q}$ ; it can be expressed as  $Y = \bigcup_{1 \leq i \leq \ell} Y_i$ , where  $Y_i$  is the set of pairs that both belong to the  $i$ th cluster  $Q_i \in \mathbf{Q}$ .
- $Z \equiv$  the set of pairs  $(x, x') \in C$  that belong to different clusters in  $\mathbf{Q}$ ; it can be expressed as  $Z = \bigcup_{1 \leq i \neq j \leq \ell} Z_{ij}$ , where  $Z_{ij}$  is the set of pairs such that  $x \in Q_i$  and  $x' \in Q_j$ .

The Rand index can directly be written in terms of the cardinalities of these sets, since the four quantities (1) are obviously given by

$$\begin{aligned} a &= |V \cap Y|, & b &= |V \cap Z| \\ c &= |W \cap Y|, & d &= |W \cap Z|. \end{aligned} \quad (3)$$

In the fuzzy case, the above sets become fuzzy sets. Let  $P_i(x) \in [0, 1]$  denote the degree of membership of element  $x \in X$  in the cluster  $P_i \in \mathbf{P}$ . The sets  $V$ ,  $W$ ,  $Y$ , and  $Z$  can then be defined through fuzzy-logical expressions that involve a t-norm  $\top$  and t-conorm  $\perp$  [11]:

$$\begin{aligned} V(x, x') &= \perp_{i=1}^k \top(P_i(x), P_i(x')) \\ W(x, x') &= \perp_{1 \leq i \neq j \leq k} \top(P_i(x), P_j(x')) \\ Y(x, x') &= \perp_{i=1}^{\ell} \top(Q_i(x), Q_i(x')) \\ Z(x, x') &= \perp_{1 \leq i \neq j \leq \ell} \top(Q_i(x), Q_j(x')). \end{aligned} \quad (4)$$

Moreover, defining the intersection of sets by the t-norm combination of membership degrees and resorting to the commonly used sigma-count principle [12] to define set cardinality, one obtains

$$\begin{aligned} a &= |V \cap Y| = \sum_{(x, x') \in C} \top(V(x, x'), Y(x, x')) \\ b &= |V \cap Z| = \sum_{(x, x') \in C} \top(V(x, x'), Z(x, x')) \\ c &= |W \cap Y| = \sum_{(x, x') \in C} \top(W(x, x'), Y(x, x')) \\ d &= |W \cap Z| = \sum_{(x, x') \in C} \top(W(x, x'), Z(x, x')). \end{aligned} \quad (5)$$

As before, the Rand index can then be defined as in (2), namely as the fraction

$$\frac{a + d}{a + b + c + d}.$$

In passing, we note that Campello is actually only interested in comparing a fuzzy partition  $\mathbf{P}$  with a non-fuzzy partition  $\mathbf{Q}$ . On the other hand, he notes himself that, formally, the measure can also be applied to compare two fuzzy partitions.

A very similar proposal was made by Frigui *et al.* [8]. Essentially, their measure can be seen as a special case of Campello's, using the product as a t-norm in (4) and summation (bounded

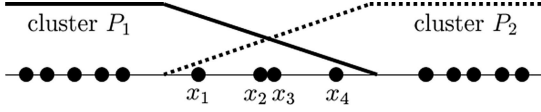


Fig. 1. Simple fuzzy partition of a subset of the reals (indicated by circles). The partition consists of two clusters, i.e.,  $P_1$  (left) and  $P_2$  (right). While some elements definitely belong to only one of the clusters, some “critical” points in the middle have partial membership in both clusters.

sum) as a t-conorm:

$$\begin{aligned}
 a &= \sum_{(x,x') \in C} \psi^{(P)}(x,x') \psi^{(Q)}(x,x') \\
 b &= \sum_{(x,x') \in C} \psi^{(P)}(x,x') (1 - \psi^{(Q)}(x,x')) \\
 c &= \sum_{(x,x') \in C} (1 - \psi^{(P)}(x,x')) \psi^{(Q)}(x,x') \\
 d &= \sum_{(x,x') \in C} (1 - \psi^{(P)}(x,x')) (1 - \psi^{(Q)}(x,x'))
 \end{aligned} \tag{6}$$

where

$$\psi^{(P)}(x,x') = \sum_{i=1}^k P_i(x)P_i(x') = \mathbf{P}(x) \cdot \mathbf{P}(x')^T \tag{7}$$

with  $\mathbf{P}(x) = (P_1(x), \dots, P_k(x))$  being the membership vector of  $x$  in the partition  $\mathbf{P}$ .

Having defined a similarity or, equivalently, a distance function, it is natural to ask for desirable metrical properties of that function. When doing so, it turns out quickly that the previous measures fail to be a proper metric. In fact, they not even satisfy reflexivity, the perhaps most basic axiom: Even for two identical partitions  $\mathbf{P}$  and  $\mathbf{Q}$ , the quantities  $b$  and  $c$  in (5) will generally not vanish, a necessary condition to have  $R(\mathbf{P}, \mathbf{Q}) = 1$ .

Consider, for example, the simple fuzzy partition  $\mathbf{P}$  that is illustrated in Fig. 1, which consists of two clusters  $P_1$  and  $P_2$ . Instead of a hard boundary, there is a “soft” transition between  $P_1$  and  $P_2$ ; the elements  $x_1$ ,  $x_2$ ,  $x_3$ , and  $x_4$  partially belong to both clusters and have membership degrees, respectively, of  $3/4$ ,  $1/2$ ,  $1/2$ ,  $1/4$  in  $P_1$  and  $1/4$ ,  $1/2$ ,  $1/2$ ,  $3/4$  in  $P_2$ . Comparing  $\mathbf{P}$  with itself in terms of either Campello’s or Frigui’s fuzzy Rand index, we obtain  $R(\mathbf{P}, \mathbf{P}) < 1$ .

Upon closer examination, it seems that the core principle of the aforementioned extensions is not suitable to compare partitions in a fuzzy sense. This becomes especially obvious when using the product as a t-norm and the (bounded) sum as t-conorm, that is, for the special case of Frigui *et al.* These operators suggest a kind of “probabilistic” interpretation. Indeed, if  $P_i(x)$  is interpreted as the probability that  $x$  belongs to the  $i$ th cluster, then  $V(x,x') = \psi^{(P)}(x,x')$  is nothing else but the probability that  $x$  and  $x'$  are put in the same cluster, given that the two corresponding clusters are chosen *independently of each other* according to the distributions  $(P_1(x), P_2(x), \dots, P_k(x))$  and  $(P_1(x'), P_2(x'), \dots, P_k(x'))$ , respectively. Likewise,  $W(x,x') = 1 - \psi^{(P)}(x,x')$  is the probability that  $x$  and  $x'$  are put into different clusters.

Even if one accepts the probabilistic interpretation of a single membership degree, the additional assumption of independence is clearly not tenable. In fact, this property is obviously violated

when comparing a partition with itself, since for each element  $x \in X$ , a cluster can then only be chosen once and not two times independently of each other. However, even if  $\mathbf{P}$  and  $\mathbf{Q}$  are not identical, independence of cluster membership is in conflict with the topological relationships between the elements and clusters. In the example in Fig. 1, for instance, it is not reasonable to put  $x_1$  and  $x_4$  into cluster  $P_2$  and  $x_2$  and  $x_3$  into cluster  $P_1$ . When putting elements independently of each other into clusters, however, this is a possible scenario. And indeed, this scenario contributes to the fuzzy Rand index according to (4) and (6).

What the fuzzy partition in our example truly suggests is that we are uncertain about the boundary between the two clusters. More concretely, the fuzzy partition suggests four possible non-fuzzy partitions:

- $\mathbf{P}_1$ , which puts the boundary left to  $x_1$ ;
- $\mathbf{P}_2$  with boundary between  $x_1$  and  $x_2$ ;
- $\mathbf{P}_3$  with boundary between  $x_3$  and  $x_4$ ;
- $\mathbf{P}_4$ , which puts the boundary right to  $x_4$ .

Thus, it seems reasonable to define an extension of the Rand index as an aggregation (e.g., weighted average) of the results of the non-fuzzy comparisons, namely

$$R(\mathbf{P}_1, \mathbf{Q}), R(\mathbf{P}_2, \mathbf{Q}), R(\mathbf{P}_3, \mathbf{Q}), R(\mathbf{P}_4, \mathbf{Q}) .$$

In Campello’s and Frigui’s approach, there are not 4 but 16 scenarios which have an influence on the result, since each of the four cluster memberships is determined independently of each other. In general, the result will, therefore, be different. In fact, differences already occur for single pairs of elements. For example, since  $x_2$  and  $x_3$  are always in the same cluster in  $\mathbf{P}_1, \dots, \mathbf{P}_4$ , it is natural to say that they are paired with degree 1. According to Campello’s approach, however, the degree to which  $x_2$  and  $x_3$  are in the same cluster in  $\mathbf{P}$  is given by

$$V(x_2, x_3) = \perp(\top(1/2, 1/2), \top(1/2, 1/2)),$$

which corresponds to the truth degree of the proposition that “ $x_1$  is put into  $P_1$  AND  $x_2$  is put into  $P_1$  OR  $x_1$  is put into  $P_2$  AND  $x_2$  is put into  $P_2$ .” In general, this degree will be  $< 1$  (except for special  $(\top, \perp)$  combinations such as  $\top = \min$  and  $\perp = \text{bounded sum}$ ).

## B. Brouwer

An alternative extension of the Rand index has been recently proposed by Brouwer [9]. His “measure of bonding” between pairs of objects arguably improves upon the comparison of objects as proposed by Frigui *et al.* (and Campello). Besides, this measure shares some similarities with our proposal to be introduced later on: Brouwer’s bonding between two objects is closely related to what we shall call their degree of equivalence. Apart from that, however, the approach is still quite similar to Campello and Frigui *et al.*

More specifically, Brouwer notes that the dot product (7) is a questionable measure to compare two membership vectors. Therefore, he proposes to replace this measure by the cosine

similarity between the membership vectors  $\mathbf{P}(x)$  and  $\mathbf{P}(x')$ :

$$b^{(P)}(x, x') = \frac{|\mathbf{P}(x)|}{|\mathbf{P}(x')|} \cdot \frac{|\mathbf{P}(x')^T|}{|\mathbf{P}(x)|}. \quad (8)$$

Thus, the main difference to Frigui is the normalization of the membership vectors. Brower considers (8) as a degree of “bonding” of the objects  $x$  and  $x'$ . The four values (5) are derived in the same way, namely according to (6) with  $\psi^{(P)}(x, x')$  and  $\psi^{(Q)}(x, x')$  replaced by  $b^{(P)}(x, x')$  and  $b^{(Q)}(x, x')$ , respectively.

As an illustration, consider again our previous example. Restricting to the objects  $x_1, x_2, x_3, x_4$ , the partition  $\mathbf{P}$  is given by the membership matrix

$$\mathbf{P} = \begin{pmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \\ 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix}.$$

Brouwer’s “bonding matrix” is then given by the pairwise cosine similarities between the rows of  $\mathbf{P}$ :

$$\mathbf{B} = \begin{pmatrix} 1 & 0.8944 & 0.8944 & 0.6 \\ 0.8944 & 1 & 1 & 0.8944 \\ 0.8944 & 1 & 1 & 0.8944 \\ 0.6 & 0.8944 & 0.8944 & 1 \end{pmatrix}.$$

Comparing  $\mathbf{P}$  with itself, one derives  $a = 4.56, b = c = 0.6177$ , and  $d = 0.2046$  from this matrix, and hence a similarity degree of 0.7941. In the case of Frigui, the corresponding “bonding matrix” is given by

$$\mathbf{B} = \mathbf{P} \cdot \mathbf{P}^T = \begin{pmatrix} 0.625 & 0.5 & 0.5 & 0.375 \\ 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 \\ 0.375 & 0.5 & 0.5 & 0.625 \end{pmatrix}$$

yielding a smaller similarity degree of 0.5052. Although the result of Brouwer looks more reasonable than the one of Frigui, the example also reveals that his approach is not reflexive either.

### C. Anderson *et al.*

Anderson *et al.* [10] proceed from the *contingency matrix* that is associated with two partitions  $\mathbf{P}$  and  $\mathbf{Q}$ , defined as  $\mathbf{C} = \mathbf{P}^T \mathbf{Q}$ . If  $\mathbf{P}$  consists of  $k$  clusters and  $\mathbf{Q}$  consists of  $\ell$  clusters, then  $\mathbf{C} = (c_{i,j})$  is a  $(k \times \ell)$ -matrix. In the non-fuzzy case, the entry  $c_{i,j}$  corresponds to the number of objects  $x$  that are put into the  $i$ th cluster  $P_i$  in  $\mathbf{P}$  and into the  $j$ th cluster  $Q_j$  in  $\mathbf{Q}$ .

It was already observed by Brower in [9] that the quantities  $a, b, c, d$  can directly be derived from  $\mathbf{C}$ . For example, what is the number  $a = |C_1|$  of tuples  $(x, x')$  of objects that are both paired in  $\mathbf{P}$  and in  $\mathbf{Q}$ ? If  $(x, x') \in C_1$ , then there are clusters  $P_i$  and  $Q_j$  such that  $x$  and  $x'$  are both in  $P_i$  and both in  $Q_j$ , and hence both in  $P_i \cap Q_j$ . The other way around, there are

$\binom{c_{i,j}}{2} = c_{i,j}(c_{i,j} - 1)/2$  possibilities to choose a pair  $(x, x')$  of that kind. In total, since we can choose  $(x, x')$  from the intersection  $P_i \cap Q_j$  of any pair of clusters  $P_i$  and  $Q_j$ , this

yields

$$a = |C_1| = \sum_{i=1}^k \sum_{j=1}^{\ell} c_{i,j}(c_{i,j} - 1)/2.$$

The other quantities can be derived analogously:

$$\begin{aligned} a &= \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^{\ell} c_{i,j}(c_{i,j} - 1) \\ b &= \frac{1}{2} \left( \sum_{j=1}^{\ell} c_{\bullet,j}^2 - \sum_{i=1}^k \sum_{j=1}^{\ell} c_{i,j}^2 \right) \\ c &= \frac{1}{2} \left( \sum_{i=1}^k c_{i,\bullet}^2 - \sum_{i=1}^k \sum_{j=1}^{\ell} c_{i,j}^2 \right) \\ d &= \frac{1}{2} \left( n^2 + \sum_{i=1}^k \sum_{j=1}^{\ell} c_{i,j}^2 - (\sum_{i=1}^k c_{i,\bullet}^2 + \sum_{j=1}^{\ell} c_{\bullet,j}^2) \right) \end{aligned} \quad (9)$$

where  $c_{i,\bullet} = \sum_{r=1}^n P_i(x_r)$  denotes the size of the cluster  $P_i$ ,  $c_{\bullet,j} = \sum_{r=1}^n Q_j(x_r)$  denotes the size of the cluster  $Q_j$ , and  $n = |X|$  is the number of objects.

Mathematically, the expressions (9) can of course also be used in the case of fuzzy partitions  $\mathbf{P}$  and  $\mathbf{Q}$ . Then, the entries of the matrix  $\mathbf{C} = \mathbf{P}^T \mathbf{Q}$  are no longer integers, but  $a, b, c, d$  can still be computed. This is precisely the idea of Anderson *et al.* [10].

While formally correct, at least at first sight, this idea can be called into question from a semantical point of view. In fact, one should note that (9) is *one* particular possibility to express  $a, b, c, d$  (in the non-fuzzy case). However, mathematically, these quantities can be expressed in many other ways, too, and in each of these cases, a straightforward fuzzification will probably yield a different result. In particular, note that the binomial coefficient  $\binom{c_{i,j}}{2}$  equals  $c_{i,j}(c_{i,j} - 1)/2$  only if  $c_{i,j}$  is an integer. In the fuzzy case, however, the meaning of the number  $c_{i,j}(c_{i,j} - 1)/2$  is not at all obvious. Besides, the latter expression is *not* the standard way to extend the binomial coefficient to real arguments. Instead, this is normally done using the well-known Gamma function. Thus, one may argue that a more proper generalization would have been

$$a = \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^{\ell} \frac{\Gamma(c_{i,j} + 1)}{\Gamma(c_{i,j})}.$$

Indeed, it is worth mentioning that  $c_{i,j}(c_{i,j} - 1) < 0$  if  $0 < c_{i,j} < 1$ , and this situation may well occur in the fuzzy case. Consequently,  $a$  can even become negative, and examples of this kind are easy to construct. This immediately implies that measures like the Jaccard coefficient (which is defined as  $a/(a + b + c)$ ; see Section V) can become negative as well.

Unsurprisingly, this approach does indeed fail to guarantee desirable metrical properties apart from symmetry. Again, for example, it is not even reflexive. As one advantage, however, also emphasized by the authors, we note its computational efficiency. This efficiency is mainly due to the fact that, in contrast with all other methods, the measure does not need to consider all pairs of object (making the complexity inherently quadratic in  $n = |X|$ ).

### D. Other Measures

Apart from the measures discussed above, other proposals can be found in the literature. These proposals, however, are not

extensions of the Rand index and related measures, insofar, as they are not based on a generalization of the four quantities (1).

Beringer and Hüllermeier [3] proceed from the following intuitive idea: A partition  $\mathbf{P} = \{P_1, \dots, P_k\}$  is similar to a partition  $\mathbf{Q} = \{Q_1, \dots, Q_\ell\}$  if, for each cluster in  $P_i \in \mathbf{P}$ , there is a similar cluster  $Q_j \in \mathbf{Q}$ , and *vice versa* if, for each cluster  $Q_j \in \mathbf{Q}$ , there is a similar cluster  $P_i \in \mathbf{P}$ . Formally, this can be expressed as follows:

$$S(\mathbf{P}, \mathbf{Q}) = \top (s(\mathbf{P}, \mathbf{Q}), s(\mathbf{Q}, \mathbf{P})) \quad (10)$$

where  $s(\mathbf{P}, \mathbf{Q})$  denotes the similarity of  $\mathbf{P}$  to  $\mathbf{Q}$  (in the aforementioned sense) and *vice versa*  $s(\mathbf{Q}, \mathbf{P})$  the similarity of  $\mathbf{Q}$  to  $\mathbf{P}$ :

$$s(\mathbf{P}, \mathbf{Q}) = \top_{1 \leq i \leq k} \perp_{1 \leq j \leq \ell} s(P_i, Q_j), \quad (11)$$

where  $\top$  is a t-norm (modeling a logical conjunction),  $\perp$  is a t-conorm (modeling a logical disjunction), and  $s(P_i, Q_j)$  denotes the similarity between clusters  $P_i$  and  $Q_j$ . Regarding the latter, note that one can refer to standard measures for the similarity of fuzzy sets, such as

$$s(P_i, Q_j) = \frac{|P_i \cap Q_j|}{|P_i \cup Q_j|} = \frac{\sum_{x \in X} \min(P_i(x), Q_j(x))}{\sum_{x \in X} \max(P_i(x), Q_j(x))}. \quad (12)$$

In order to take the different size of clusters into account, Beringer and Hüllermeier propose to generalize this approach by using a *weighted* t-norm aggregation [13]:

$$s(\mathbf{P}, \mathbf{Q}) = \top_{1 \leq i \leq k} m \left( w_i, \perp_{1 \leq j \leq \ell} s(P_i, Q_j) \right) \quad (13)$$

where  $w_i = |P_i|/|X|$  is the relative size of cluster  $P_i$ .

A similar approach was recently put forward by Runkler [14]. The measure he proposes is almost the same as the unweighted version (10), except that the two similarity degrees are combined *disjunctively* instead of *conjunctively*:

$$S(\mathbf{P}, \mathbf{Q}) = \perp (s(\mathbf{P}, \mathbf{Q}), s(\mathbf{Q}, \mathbf{P})). \quad (14)$$

Thus, in a sense, it is more an inclusion than a similarity measure, and consequently loses reflexivity. As fuzzy-logical operators, Runkler suggests  $\top = \min$  and  $\perp = \max$ .

#### IV. NEW FUZZY RAND INDEX

In this section, we propose a new fuzzy variant of the Rand index that exhibits desirable metric properties. In the following, we focus on the view of the Rand index as a distance function. Thanks to the affine transformation  $D_R = 1 - R$ , all results can directly be transferred to the original conception as a measure of similarity.

##### A. Definition

Given a fuzzy partition  $\mathbf{P} = \{P_1, P_2, \dots, P_k\}$  of  $X$ , each element  $x \in X$  can be characterized by its membership vector

$$\mathbf{P}(x) = (P_1(x), P_2(x), \dots, P_k(x)) \in [0, 1]^k, \quad (15)$$

where  $P_i(x)$  is the degree of membership of  $x$  in the  $i$ th cluster  $P_i$ . We define a fuzzy equivalence relation on  $X$  in terms of a

similarity measure on the associated membership vectors (15). Generally, this relation is of the form

$$E_{\mathbf{P}}(x, x') = 1 - \|\mathbf{P}(x) - \mathbf{P}(x')\|, \quad (16)$$

where  $\|\cdot\|$  is a proper metric on  $[0, 1]^k$ . The basic requirement on this metric is that it yields values in  $[0, 1]$ . Relation (16) generalizes the equivalence relation that is induced by a conventional partition (where each cluster forms an equivalence class). Indeed, it is easy to verify that relation (16) is not only reflexive and symmetric but  $T_L$ -transitive as well, where  $T_L$  is the Lukasiewicz t-norm  $(u, v) \mapsto \max(u + v - 1, 0)$  [15]. In passing, we also note that definition (16) is invariant toward a permutation (renumbering) of the clusters in  $\mathbf{P}$ , which is clearly a desirable property.

Now, given two fuzzy partitions  $\mathbf{P}$  and  $\mathbf{Q}$ , the idea is to generalize the concept of concordance as follows. We consider a pair  $(x, x')$  as being concordant insofar as  $\mathbf{P}$  and  $\mathbf{Q}$  agree on their degree of equivalence. This suggests to define the *degree of concordance* as

$$\text{conc}(x, x') = 1 - |E_{\mathbf{P}}(x, x') - E_{\mathbf{Q}}(x, x')| \in [0, 1]. \quad (17)$$

Analogously, the *degree of discordance* is

$$\text{disc}(x, x') = |E_{\mathbf{P}}(x, x') - E_{\mathbf{Q}}(x, x')|.$$

Our distance measure on fuzzy partitions is then defined by the normalized sum of degrees of discordance:

$$d(\mathbf{P}, \mathbf{Q}) = \frac{\sum_{(x, x') \in C} |E_{\mathbf{P}}(x, x') - E_{\mathbf{Q}}(x, x')|}{n(n-1)/2}. \quad (18)$$

Likewise,

$$R_E(\mathbf{P}, \mathbf{Q}) = 1 - d(\mathbf{P}, \mathbf{Q}) \quad (19)$$

corresponds to the normalized degree of concordance and, therefore, is a direct generalization of the original Rand index.

##### B. Formal Properties

In this section, we first show that our proposal is indeed a proper generalization of the Rand index. Afterward, we study the metrical properties of the measure.

*Proposition 1:* In the case where  $\mathbf{P}$  and  $\mathbf{Q}$  are non-fuzzy partitions, measure (19) reduces to the original Rand index.

*Proof:* In the non-fuzzy case, the membership vectors (15) are 0/1-vectors. More specifically, each vector has a single entry  $P_i(x) = 1$ , while all other entries are 0. Consequently, the fuzzy equivalence (16) reduces to the conventional equivalence, that is,  $E_{\mathbf{P}}(x, x') = 1$  if  $x$  and  $x'$  are in the same cluster and  $E_{\mathbf{P}}(x, x') = 0$  otherwise. Likewise, (17) yields 1 if  $(x, x')$  is a concordant pair and 0 otherwise. Consequently, measure (19) is the (normalized) sum of concordant pairs and, therefore, equals the original Rand index.  $\square$

Recall that a nonnegative  $Z^2 \rightarrow \mathbb{R}$  mapping  $d(\cdot)$  is called a metric on  $Z$  if it satisfies the following properties for all  $z, z', z'' \in Z$ :

- Reflexivity:  $d(z, z) = 0$ .
- Separation:  $d(z, z') = 0$  implies  $z = z'$ .
- Symmetry:  $d(z, z') = d(z', z)$ .

- Triangle inequality:  $d(z, z'') \leq d(z, z') + d(z', z'')$ .

The properties of reflexivity and symmetry are quite obviously valid for our measure (18). To show the triangle inequality, consider three fuzzy partitions  $\mathbf{P}$ ,  $\mathbf{Q}$ ,  $\mathbf{R}$  and fix a single tuple  $(x, x') \in C$ . Let

$$u = E_{\mathbf{P}}(x, x'), \quad v = E_{\mathbf{Q}}(x, x'), \quad w = E_{\mathbf{R}}(x, x').$$

Since  $u, v$ , and  $w$  are the real numbers (from the unit interval), and the simple difference on the reals satisfies the triangle inequality, we have  $|u - w| \leq |u - v| + |v - w|$ . Now, since this inequality holds for each pair  $(x, x') \in C$ , it remains valid when summing over all these pairs. In other words, it is also satisfied by (18), which means that

$$d(\mathbf{P}, \mathbf{R}) \leq d(\mathbf{P}, \mathbf{Q}) + d(\mathbf{Q}, \mathbf{R}).$$

The separation property is not immediately valid for (18). Roughly speaking, this is due to the fact that, by mapping elements to their membership vectors (15), some information about the partition itself is lost. In particular, it is possible that two partitions, even though they are not identical, cannot be distinguished in terms of the distances between these vectors.

Nevertheless, we can guarantee the separation property by restricting to a reasonable subclass of fuzzy partitions. We call a fuzzy partition  $\mathbf{P} = \{P_1, P_2 \dots P_k\}$  *normal*, if it satisfies the following:

N1 For each  $x \in X$ :  $P_1(x) + \dots + P_k(x) = 1$ .

N2 For each  $P_i \in \mathbf{P}$ , there exists  $x \in X$  such that  $P_i(x) = 1$ .

In other words, we consider Ruspini partitions [16] and assume that each cluster has a prototypical element. Moreover, we assume the following equivalence relation on  $X$ :

$$E_{\mathbf{P}}(x, x') = 1 - \frac{1}{2} \sum_{i=1}^k |P_i(x) - P_i(x')| \quad (20)$$

$$= 1 - \|\mathbf{P}(x) - \mathbf{P}(x')\|$$

with  $\|\cdot\|$  being the  $L_1$ -norm divided by 2. Note that  $0 \leq E_{\mathbf{P}}(x, x') \leq 1$  for all  $(x, x') \in X^2$  under assumption N1.

Now, consider two normal fuzzy partitions  $\mathbf{P}$  and  $\mathbf{Q}$ , and suppose that  $d(\mathbf{P}, \mathbf{Q}) = 0$ . According to our definition of  $d(\cdot)$ , this obviously means that

$$E_{\mathbf{P}}(x, x') = E_{\mathbf{Q}}(x, x') \quad (21)$$

for all  $(x, x') \in C$ . We call a set  $\{p_1, p_2, \dots, p_k\} \subset X$  a prototype set for  $\mathbf{P}$ , if  $P_i(p_i) = 1$  for all  $i = 1, \dots, k$  (note that a prototype set is not necessarily unique). We distinguish two cases.

a) There are no identical prototype sets for  $\mathbf{P}$  and  $\mathbf{Q}$ ; note that this is necessarily the case if  $\mathbf{P}$  and  $\mathbf{Q}$  have a different number of clusters. Let  $k$  and  $\ell$  denote the number of clusters in  $\mathbf{P}$  and  $\mathbf{Q}$ , respectively, and let  $\ell \leq k$  without loss of generality. (Remark that  $k > 1$ , since otherwise  $k = \ell = 1$ , which means that both  $\mathbf{P}$  and  $\mathbf{Q}$  consist of a single cluster and are, therefore, identical.) Moreover, let  $\{p_1, \dots, p_k\}$  be a prototype set of  $\mathbf{P}$ . Note that N1 and N2 jointly imply that a prototype is represented by a 0/1 membership vector, and that  $\|\mathbf{P}(p_i) - \mathbf{P}(p_j)\| = 1$  for two different prototypes  $p_i$  and  $p_j$ . Moreover, these properties imply that the extreme distance of 1

can only be assumed for membership vectors  $(m_1, \dots, m_k)$  and  $(m'_1, \dots, m'_k)$  if  $\min(m_i, m'_i) = 0$  for all  $i \in \{1, \dots, k\}$ ; that is,  $m_i > 0$  implies  $m'_i = 0$ , and  $m'_i > 0$  implies  $m_i = 0$ . Now, consider the membership vectors  $\mathbf{Q}(p_1), \dots, \mathbf{Q}(p_k)$ , which can be combined into a  $(k \times \ell)$ -matrix:

$$\begin{array}{l} \mathbf{Q}(p_1) : \quad m_{11} \quad m_{12} \quad \dots \quad m_{1\ell} \\ \mathbf{Q}(p_2) : \quad m_{21} \quad m_{22} \quad \dots \quad m_{2\ell} \\ \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\ \mathbf{Q}(p_k) : \quad m_{k1} \quad m_{k2} \quad \dots \quad m_{k\ell} \end{array}$$

Since  $\ell \leq k$ , and since not all  $p_i$  are prototypes in  $\mathbf{Q}$ , there is necessarily a column  $c$  and rows  $i$  and  $j$  such that  $m_{ic} > 0$  and  $m_{jc} > 0$  (in other words, it is not possible that there is only one positive entry in each column). Consequently, there exist at least two prototypes  $p_i$  and  $p_j$  of clusters  $P_i$  and  $P_j$ , respectively, for which  $\|\mathbf{Q}(p_i) - \mathbf{Q}(p_j)\| < 1$ , and therefore,  $E_{\mathbf{Q}}(p_i, p_j) > 0$ . Since  $E_{\mathbf{P}}(p_i, p_j) = 0$ , condition (21) is, hence, violated.

b) There are identical prototype sets  $\{p_1, \dots, p_k\} = \{q_1, \dots, q_\ell\}$ , respectively, for  $\mathbf{P}$  and  $\mathbf{Q}$  (which means that  $k = \ell$ , i.e.,  $\mathbf{P}$  and  $\mathbf{Q}$  do have the same number of clusters). We can then establish a one-to-one correspondence between prototypes such that, without loss of generality,  $p_i = q_i$  for  $i = 1, \dots, k$ . From properties N1 and N2, it follows that the membership degree of any element  $x$  in the cluster  $P_i$  is a function of  $E_{\mathbf{P}}(x, p_i)$ . In fact, noting that  $\mathbf{P}(p_i)$  is a 0/1 vector with a single 1 on position  $i$ , we get

$$\begin{aligned} E_{\mathbf{P}}(x, p_i) &= 1 - \frac{1}{2} \sum_{j=1}^k |P_j(x) - P_j(p_i)| \\ &= 1 - \frac{1}{2} \left( (1 - P_i(x)) + \sum_{j \neq i} P_j(x) \right) \\ &= 1 - \frac{1}{2} ((1 - P_i(x)) + (1 - P_i(x))) \\ &= P_i(x). \end{aligned} \quad (22)$$

From (21), it thus follows that  $P_i(x) = Q_i(x)$  for all  $x \in X$ , i.e., the  $i$ th cluster in  $\mathbf{P}$  and the  $i$ th cluster in  $\mathbf{Q}$  are identical. Since this holds for all  $i \in \{1, 2, \dots, k\}$ , we have shown that  $\mathbf{P} = \mathbf{Q}$ .

The aforementioned results can be summarized as follows.

*Theorem 1:* The distance function (18) on fuzzy partitions is a pseudo-metric, i.e., it is reflexive, symmetric, and subadditive. Moreover, on the restricted class of normal fuzzy partitions or, more specifically, under the assumptions N1, N2, and (20), it also satisfies the separation property, and therefore, it is a metric.

*Remark 1:* Since our comparison of two fuzzy partitions  $\mathbf{P}$  and  $\mathbf{Q}$  is eventually reduced to the comparison of the corresponding equivalence relations  $E_{\mathbf{P}}$  and  $E_{\mathbf{Q}}$ , it appears legitimate to ask for the formal relationship between the partitions and the equivalence relations (for more general studies of this type of question, see, e.g., [17]–[19]). In this regard, it is worth mentioning that the following result follows immediately from (22): If  $\mathbf{P} = \{P_1, \dots, P_k\}$  is a normal fuzzy partition with prototypes  $\{p_1, \dots, p_k\}$ , then  $\mathbf{P} = \{E_1, \dots, E_k\}$ , where the fuzzy equivalence class  $E_i$  is defined by  $E_i(x) = E_{\mathbf{P}}(x, p_i)$  for all  $x \in X$ . In other words, just like in the non-fuzzy case, the original partition  $\mathbf{P}$  corresponds to the collection of (fuzzy) equivalence classes associated with its prototypical elements.

Finally, regarding the computational complexity of our approach, note that our measure is essentially derived by comparing the equivalence degrees (16) for each pair of elements  $x$  and  $x'$ , and that the number of such pairs is  $n(n-1)/2$ . The computation of the equivalence degrees in turn comes down to comparing vectors of dimension  $k$  in the case of the first and of dimension  $\ell$  in the case of the second partition. Thus, as a result, the overall complexity is  $\mathcal{O}(\max(k, \ell) \cdot n^2)$ .

## V. EXTENSIONS

Apart from the Rand index itself, a number of related comparison measures have been proposed in the literature, many of which can be expressed in terms of the four cardinalities  $a$ ,  $b$ ,  $c$ , and  $d$  in (1). Important examples include the Jaccard measure [20] (also known as Tanimoto coefficient)

$$\frac{a}{a+b+c} \quad (23)$$

and the related Dice index

$$\frac{a}{a + \frac{1}{2}(b+c)}. \quad (24)$$

An obvious idea is to extend our approach to measures of this kind. In this regard, it is important to note that, for many measures, the two types of concordance are not treated in a symmetric way, as done by the Rand index. In the two aforesaid measures, for example, only  $a$  appears, while  $d$  is omitted from both the nominator and the denominator. Now, since our generalization of concordance (17) is an expression of the sum  $a+d$ , an important prerequisite to apply our approach to other measures is to split (17) into two parts, say, a-concordance and d-concordance.

Essentially, (17) expresses that  $x$  and  $x'$  are concordant insofar as their degree of equivalence in structure  $\mathbf{P}$  is the same as their degree of equivalence in structure  $\mathbf{Q}$ , that is

$$u = E_{\mathbf{P}}(x, x') = E_{\mathbf{Q}}(x, x') = v.$$

In the non-fuzzy case, where  $u$  and  $v$  are either 0 or 1, we have a-concordance if  $u = v = 1$  and d-concordance if  $u = v = 0$ . Specifically, a-concordance can be considered as a strict version of concordance, which not only assumes that  $u$  is equal to  $v$ , but also that both values are large, i.e.,  $x$  and  $x'$  are regarded as equivalent in both structures. In fact, this is the reason why a-concordance is in a sense more relevant than d-concordance.

An obvious formalization of a-concordance, in the fuzzy case, is therefore

$$a = \top(1 - |u - v|, \top(u, v)),$$

where  $\top$  is a t-norm operator [11]. Thus,  $x$  and  $x'$  are a-concordant insofar as their degree of equivalence in  $\mathbf{P}$  and  $\mathbf{Q}$  is similar *and* their degree of equivalence in  $\mathbf{P}$  is high *and* their degree of equivalence in  $\mathbf{Q}$  is high. In other words, the additional restriction, which is distinguishing a-concordance from concordance, is the condition  $\top(u, v)$ , which is conjunctively combined with the original degree of concordance. Consequently, d-concordance corresponds to that part of the concordance for which this condition is *not* satisfied or, stated differently, for

which the negation of this condition holds, which means that either  $u$  (the degree of equivalence in  $\mathbf{P}$ ) is not high or  $v$  (the degree of equivalence in  $\mathbf{Q}$ ) is not high:

$$d = \top(1 - |u - v|, \perp(1 - u, 1 - v))$$

where  $\perp$  is a t-conorm. To make this definition of a-concordance and d-concordance consistent with our previous proposal, we should require that the sum of these two types of concordance equals the original concordance, which leads to

$$w = \top(w, \top(u, v)) + \top(w, \perp(1 - u, 1 - v)), \quad (25)$$

where  $w = 1 - |u - v|$ . An interesting question, then, concerns the choice of the t-norm  $\top$  and t-conorm  $\perp$ : Which operators  $(\top, \perp)$  can guarantee that (25) holds for all  $0 \leq u, v, w \leq 1$ ? Interestingly, this question can be answered in a unique way, thanks to a theorem that is proved by Alsina in [21]: The only admissible choice is the product t-norm and its associated t-conorm, namely the algebraic sum. Thus, we end up with the following definitions of a-concordance and d-concordance:

$$a = (1 - |u - v|) \cdot u \cdot v,$$

$$d = (1 - |u - v|) \cdot (1 - u \cdot v)$$

where  $u = E_{\mathbf{P}}(x, x')$  and  $v = E_{\mathbf{Q}}(x, x')$ . These quantities can be directly plugged into (23) or (24), thus allowing us to generalize measures of this kind.

In a similar way, the degree of discordance could be split into, say, b- and c-discordance (although a distinction of this kind is used by less measures). The case of b-discordance occurs when the degree of equivalence of  $x$  and  $x'$  in  $\mathbf{P}$  is larger than that in  $\mathbf{Q}$  (which necessarily means  $E_{\mathbf{P}}(x, x') = 1$  and  $E_{\mathbf{Q}}(x, x') = 0$  in the non-fuzzy case), and *vice versa* for c-discordance. A generalization of this distinction calls for a fuzzy extension of the “larger than” relation, which, in the simplest case, is given by the standard order relation on  $[0, 1]$ . This yields

$$b = \max(E_{\mathbf{P}}(x, x') - E_{\mathbf{Q}}(x, x'), 0)$$

$$c = \max(E_{\mathbf{Q}}(x, x') - E_{\mathbf{P}}(x, x'), 0)$$

Thus, we end up with a consistent generalization of all four quantities that are used by measures based on the concordance and discordance of pairs of data points, suggesting a direct fuzzy extension for each measure of that kind.

## VI. EXPERIMENTAL VALIDATION

An experimental comparison of similarity measures for fuzzy partitions, or similarity measures in general, is far from trivial, mainly because a clear reference is normally missing: If two measures produce different similarity degrees for a pair of partitions  $(\mathbf{P}, \mathbf{Q})$ , it is often difficult to say which is the more correct one.

Our experiments are, therefore, based on two settings for which there is at least a reasonable expectation. This is accomplished by producing a *sequence* of fuzzy partitions  $(\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_m)$  with a natural linear order, where  $\mathbf{P}_i$  is the solution to a clustering task  $T_i$ . As will be seen, due to the specific construction of the sequences of tasks  $T_i$  and partitions  $\mathbf{P}_i$ ,



the following assumption appears legitimate: The closer  $i$  and  $j$ , the more similar the tasks  $T_i$  and  $T_j$ , and hence the more similar  $\mathbf{P}_i$  and  $\mathbf{P}_j$  should be.

More specifically, since larger effects can be expected for smaller indexes, we measure the similarity between tasks  $T_i$  and  $T_j$  by

$$S(i, j) = S(T_i, T_j) = \frac{\min(i, j)}{\max(i, j)}. \quad (26)$$

The performance of a similarity measures  $R$  for fuzzy partitions is then defined in terms of the correlation between  $S$  and  $R$ , that is, by comparing the set of similarity degrees  $\{s_{i,j} = S(i, j)\}_{1 \leq i < j \leq m}$  as defined in (26) with the set of similarity degrees  $\{r_{i,j} = R(\mathbf{P}_i, \mathbf{P}_j)\}_{1 \leq i < j \leq m}$ . Since the numbers  $S(i, j)$  themselves might be disputable, whereas their comparison is definitely meaningful (i.e.,  $S(i, j) < S(k, l)$  means that tasks  $T_i$  and  $T_j$  are less similar to tasks  $T_k$  and  $T_l$ ), we compute a rank correlation measure, namely the Kendall tau coefficient [22], instead of a numerical correlation measure. This coefficient ranges between  $-1$  and  $+1$ , with  $+1$  meaning perfect correlation,  $0$  no correlation, and  $-1$  perfect anti-correlation.

#### A. First Experiment: Comparing Partitions with Different Numbers of Clusters

Sequences of fuzzy partitions are generated in two different ways. In our first experiment, we applied fuzzy  $C$ -means (FCM) clustering [23], [24] with different values of  $C$ .<sup>3</sup> In principle, any other clustering method could of course also be used. Our decision in favor of FCM is simply driven by its popularity. In order to avoid local optima, FCM was started ten times, and the best result was adopted.

More concretely, the task  $T_i$  was defined as partitioning a given dataset into  $C = i + 1$  groups, with  $i \in \{1, 2, \dots, 7\}$ . Our assumption of task similarity as explained earlier does clearly make sense for this problem. For example, the task to partition a dataset into three clusters is more similar to finding four clusters than to finding, say, six clusters; correspondingly, the three-cluster structure  $\mathbf{P}_2$  is expected to be more similar to the four-cluster structure  $\mathbf{P}_3$  than to the six-cluster structure  $\mathbf{P}_5$ .

As an illustration, consider the dataset shown in Fig. 2, which has been generated synthetically using four Gaussian distributions. The optimal number of clusters is, thus,  $C = 4$ , though in general, this number is of course unknown. Applying FCM with  $C$  ranging from 2 to 8 yields seven different fuzzy partitions  $\mathbf{P}_1, \dots, \mathbf{P}_7$ . These partitions can be compared with each other using our extension of the Rand index (or any of the other extensions that are discussed in Section III). Table I provides a summary of the corresponding similarity degrees  $R(\mathbf{P}_i, \mathbf{P}_j)$ .

#### B. Second Experiment: Comparing Partitions of Related Datasets

Our second experiment is motivated by the fact that, prior to applying a clustering method, the original data are often prepro-

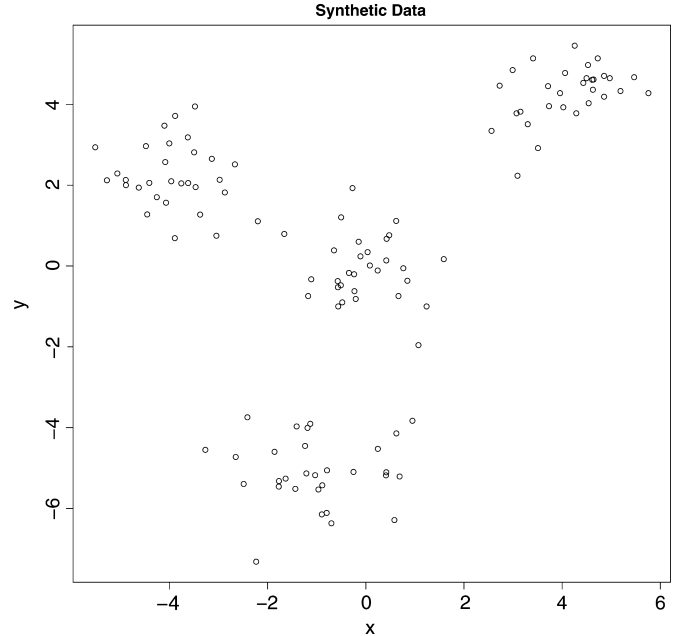


Fig. 2. Synthetic dataset generated by sampling from four Gaussian distributions.

TABLE I  
SIMILARITY  $R(\mathbf{P}_i, \mathbf{P}_j)$  BETWEEN FUZZY PARTITIONS WITH DIFFERENT NUMBERS OF CLUSTERS

|                | $\mathbf{P}_1$ | $\mathbf{P}_2$ | $\mathbf{P}_3$ | $\mathbf{P}_4$ | $\mathbf{P}_5$ | $\mathbf{P}_6$ | $\mathbf{P}_7$ |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| $\mathbf{P}_1$ | 1              | 0,732          | 0,631          | 0,610          | 0,593          | 0,577          | 0,555          |
| $\mathbf{P}_2$ | 0,732          | 1              | 0,868          | 0,850          | 0,833          | 0,811          | 0,792          |
| $\mathbf{P}_3$ | 0,631          | 0,868          | 1              | 0,965          | 0,939          | 0,918          | 0,898          |
| $\mathbf{P}_4$ | 0,610          | 0,850          | 0,965          | 1              | 0,966          | 0,938          | 0,917          |
| $\mathbf{P}_5$ | 0,593          | 0,833          | 0,939          | 0,966          | 1              | 0,940          | 0,933          |
| $\mathbf{P}_6$ | 0,577          | 0,811          | 0,918          | 0,938          | 0,940          | 1              | 0,938          |
| $\mathbf{P}_7$ | 0,555          | 0,792          | 0,898          | 0,917          | 0,933          | 0,938          | 1              |

cessed, for example, using dimensionality reduction techniques (an example of this kind was given in Section I). More specifically, as task  $T_i$ , we considered the problem to partition a low-dimensional projection of the original dataset, namely the projection given by the first  $i$  principal components (as determined by the principal component analysis), into a pre-defined (and data-dependent) number of clusters. Obviously, the assumption of task similarity does again appear reasonable, simply because the closer the  $i$  and  $j$ , the more similar the datasets to be partitioned, namely the  $i$ -dimensional and  $j$ -dimensional projections of the original data.

The dimensionality  $i$  ranges between  $d_{\min} = 1$  and the maximum value  $d_{\max}$  which depends on the dimensionality of the original dataset. Like in the first experiment, FCM was used to cluster, and the pre-defined number  $C$  of clusters was given by the number of classes of the respective dataset (all datasets have a specific class attribute, which is typically used as a target in classification). The datasets used in both experiments are taken from the UCI repository for machine learning.<sup>4</sup> We removed all non-numerical attributes; see Table II for a summary.

<sup>3</sup>Note that the FCM algorithm produces fuzzy partitions that are not necessarily normal in the sense as defined in Section IV.

<sup>4</sup><http://archive.ics.uci.edu/ml/>

TABLE II

DATASETS USED IN THE EXPERIMENTS: SIZE (NUMBER OF INSTANCES), NUMBER OF CLASSES, AND NUMBER OF ATTRIBUTES (DIMENSIONS) IN THE FIRST (E1) AND SECOND EXPERIMENT (E2)

| data set   | size | # classes | # attr (E1) | # attr (E2) |
|------------|------|-----------|-------------|-------------|
| wine       | 178  | 3         | 13          | 1–10        |
| vehicle    | 846  | 4         | 18          | 1–7         |
| sonar      | 208  | 2         | 60          | 1–30        |
| schizo     | 340  | 2         | 12          | 1–10        |
| pima       | 768  | 2         | 8           | 1–9         |
| ionosphere | 351  | 2         | 34          | 1–23        |
| flag       | 194  | 8         | 17          | 1–27        |
| cancer     | 683  | 2         | 9           | 1–7         |
| autos      | 205  | 6         | 15          | 1–24        |
| homerun    | 163  | 2         | 13          | 1–26        |

TABLE III

AVERAGE RANK OF HRHS IS COMPARED WITH THE AVERAGE RANK OF OTHER METHODS USING THE HOLM TEST

|          | 1st exp.<br>Rand | 1st exp.<br>Jaccard | 2nd exp.<br>Rand | 2nd exp.<br>Jaccard |
|----------|------------------|---------------------|------------------|---------------------|
| HRHS     | 1.40             | 1.95                | 1.25             | 1.20                |
| Brouwer  | 3.25             | 3.25                | <b>3.20</b>      | <b>3.20</b>         |
| Campello | 2.75             | 2.70                | <b>3.70</b>      | <b>4.40</b>         |
| Anderson | <b>5.75</b>      | <b>4.85</b>         | <b>5.85</b>      | <b>5.35</b>         |
| Frigui   | <b>5.05</b>      | <b>4.85</b>         | <b>5.05</b>      | <b>5.05</b>         |
| Runkler  | 2.80             | 3.40                | 1.95             | 1.80                |

Average ranks significantly worse than the one of HRHS (at the 5% level) are highlighted in bold.

TABLE IV

FIRST EXPERIMENT: CORRELATION FOR RAND IN TERMS OF KENDALL TAU

| data set   | HRHS           | Brouwer        | Campello       | Anderson | Frigui  | Runkler        |
|------------|----------------|----------------|----------------|----------|---------|----------------|
| wine       | <b>0.82792</b> | 0.78406        | 0.75116        | 0.67440  | 0.67440 | 0.76213        |
| vehicle    | 0.82792        | <b>0.85534</b> | 0.82792        | 0.73471  | 0.73471 | 0.50991        |
| sonar      | 0.61409        | -0.09869       | 0.87179        | 0.44412  | 0.44412 | <b>0.96500</b> |
| schizo     | <b>0.87179</b> | 0.80599        | 0.76761        | 0.70730  | 0.70730 | 0.84437        |
| pima       | 0.81147        | 0.78954        | <b>0.86082</b> | 0.69085  | 0.69085 | 0.75664        |
| ionosphere | 0.87727        | 0.57022        | 0.91565        | 0.44960  | 0.44960 | <b>0.94855</b> |
| homerun    | <b>0.82792</b> | 0.80599        | 0.80051        | 0.60861  | 0.60861 | 0.69633        |
| flag       | <b>0.86630</b> | 0.82244        | 0.82792        | 0.82244  | 0.82244 | 0.42767        |
| cancer     | 0.80599        | 0.47153        | 0.70730        | 0.71826  | 0.71826 | <b>0.89920</b> |
| autos      | 0.84437        | <b>0.84985</b> | <b>0.84985</b> | 0.79502  | 0.79502 | 0.67988        |
| mean       | 0.81750        | 0.66563        | <b>0.81805</b> | 0.66453  | 0.66453 | 0.74897        |

### C. Results

The results of the first experiment are summarized in Table IV for the Rand index as a similarity measure and in Table V for the Jaccard measure. Both tables show the aforementioned correlation between the similarity of tasks and the similarity of fuzzy partitions for our proposal (which is denoted HRHS) as well as the approaches that are discussed in Section III: Brouwer, Campello,<sup>5</sup> Anderson *et al.*, Frigui *et al.*, and Runkler. The same results are summarized for the second experiment in Table VI (Rand) and Table VII (Jaccard). Obviously, our approach performs quite well in comparison with the other methods. This is confirmed by a statistical test that is based on the average rank<sup>6</sup> of each method, which follows the two-step procedure proposed in [25] and is summarized in Table III.

<sup>5</sup>with t-norm min and t-conorm max.

<sup>6</sup>On each dataset, the best method (i.e., with the highest correlation value) receives rank 1, the second-best method receives rank 2, etc., and these ranks are averaged over the datasets, giving rise to one average rank per experiment and method.

TABLE V

FIRST EXPERIMENT: CORRELATION FOR JACCARD IN TERMS OF KENDALL TAU

| data set   | HRHS           | Brouwer  | Campello | Anderson | Frigui   | Runkler        |
|------------|----------------|----------|----------|----------|----------|----------------|
| wine       | <b>0.83341</b> | 0.73471  | 0.74020  | 0.63054  | 0.63602  | 0.76213        |
| vehicle    | <b>0.87179</b> | 0.52636  | 0.52636  | 0.17545  | 0.18094  | 0.50991        |
| sonar      | 0.59216        | -0.10966 | 0.70182  | -0.20835 | -0.14256 | <b>0.96500</b> |
| schizo     | <b>0.93210</b> | 0.88823  | 0.88275  | 0.79502  | 0.80051  | 0.84437        |
| pima       | <b>0.87727</b> | 0.23577  | 0.54829  | 0.02741  | 0.02741  | 0.75664        |
| ionosphere | 0.87727        | 0        | 0.80051  | -0.08773 | -0.07676 | <b>0.94855</b> |
| homerun    | <b>0.92113</b> | 0.64699  | 0.78406  | 0.35639  | 0.37284  | 0.69633        |
| flag       | <b>0.81081</b> | 0.68984  | 0.66696  | 0.66042  | 0.66042  | 0.42829        |
| cancer     | 0.77309        | 0.04386  | 0.33446  | 0.02741  | 0.02741  | <b>0.89920</b> |
| autos      | <b>0.88275</b> | 0.75664  | 0.74020  | 0.54829  | 0.56474  | 0.67988        |
| mean       | <b>0.83718</b> | 0.44128  | 0.67256  | 0.29249  | 0.30510  | 0.74903        |

TABLE VI

SECOND EXPERIMENT: CORRELATION FOR RAND IN TERMS OF KENDALL TAU

| data set   | HRHS           | Brouwer  | Campello | Anderson | Frigui   | Runkler        |
|------------|----------------|----------|----------|----------|----------|----------------|
| wine       | <b>0.96202</b> | 0.44380  | -0.03446 | -0.27014 | -0.27014 | 0.94824        |
| vehicle    | <b>0.89372</b> | 0.60861  | 0.18094  | -0.04935 | -0.05483 | 0.85534        |
| sonar      | <b>0.79291</b> | 0.57766  | 0.10631  | -0.37083 | -0.35336 | 0.76866        |
| schizo     | 0.95926        | 0.70291  | -0.11302 | -0.29632 | -0.29632 | <b>0.96064</b> |
| pima       | <b>0.76502</b> | 0.51416  | 0.21354  | -0.08708 | -0.08500 | 0.70697        |
| homerun    | <b>0.77074</b> | 0.57028  | -0.12143 | -0.39774 | -0.39957 | 0.75811        |
| flag       | <b>0.77872</b> | 0.59855  | 0.05592  | 0.63464  | 0.62717  | 0.64877        |
| ionosphere | 0.89233        | 0.66136  | -0.27518 | -0.43294 | -0.43315 | <b>0.90229</b> |
| cancer     | <b>0.92662</b> | -0.19190 | -0.19739 | -0.25221 | -0.25221 | 0.92113        |
| autos      | <b>0.88984</b> | 0.68562  | 0.11653  | 0.62413  | 0.60951  | 0.73976        |
| mean       | <b>0.86312</b> | 0.51710  | -0.00682 | -0.08978 | -0.09079 | 0.82099        |

TABLE VII

SECOND EXPERIMENT: CORRELATION FOR JACCARD IN TERMS OF KENDALL TAU

| data set   | HRHS           | Brouwer  | Campello | Anderson | Frigui   | Runkler        |
|------------|----------------|----------|----------|----------|----------|----------------|
| wine       | 0.92343        | 0.79663  | 0.23017  | -0.27289 | -0.27289 | <b>0.94824</b> |
| vehicle    | <b>0.89920</b> | 0.75664  | 0.67988  | -0.18642 | -0.18642 | 0.85534        |
| sonar      | <b>0.77143</b> | 0.57217  | 0.02727  | -0.38668 | -0.24234 | 0.76866        |
| schizo     | <b>0.96615</b> | 0.59816  | -0.11715 | -0.30597 | -0.30459 | 0.96064        |
| pima       | <b>0.75258</b> | 0.45611  | 0.18659  | -0.07464 | -0.07256 | 0.70697        |
| homerun    | <b>0.76570</b> | 0.56226  | 0.06214  | -0.44561 | 0.28465  | 0.75811        |
| flag       | <b>0.76453</b> | 0.57289  | 0.70582  | -0.52207 | -0.51299 | 0.64877        |
| ionosphere | 0.88178        | 0.60861  | -0.26966 | -0.42324 | -0.42297 | <b>0.90229</b> |
| cancer     | <b>0.92113</b> | -0.13707 | -0.20287 | -0.24673 | -0.24673 | <b>0.92113</b> |
| autos      | <b>0.86544</b> | 0.60430  | 0.76618  | -0.28674 | -0.27315 | 0.73976        |
| mean       | <b>0.85114</b> | 0.53907  | 0.20684  | -0.31510 | -0.22500 | 0.82099        |

In addition, Fig. 3 provides a visual impression of the performance of the different methods (for the sake of clearness, we omit Brouwer and Frigui, which behave quite similarly to Campello and Anderson, respectively), albeit only for a single example. What is shown here, for the wine dataset, is a graphical illustration of the similarity matrix ( $r_{i,j}$ ) with entries  $r_{i,j} = R(\mathbf{P}_i, \mathbf{P}_j)$  for the Jaccard measure. Each of these entries is shown as a black bar whose length is proportional to the value. What should be expected, therefore, is to find the longest bars on the diagonal, whereas the length decreases toward the lower left and upper right corner; besides, the bars on the main and the secondary diagonals should have a similar length. As can be seen, this expectation is met quite well by our approach, for which this tendency is rather pronounced. What can also be seen from these pictures is that, in general, the different methods show different qualitative behaviors.

## VII. SUMMARY AND OUTLOOK

The main contribution of this paper is a proposal of a generalized Rand index to compare fuzzy clustering structures. Elaborating on the formal properties of our measure, we have shown that it is a pseudo-metric and, on a subclass of fuzzy partitions obeying certain normality assumptions, even a metric.

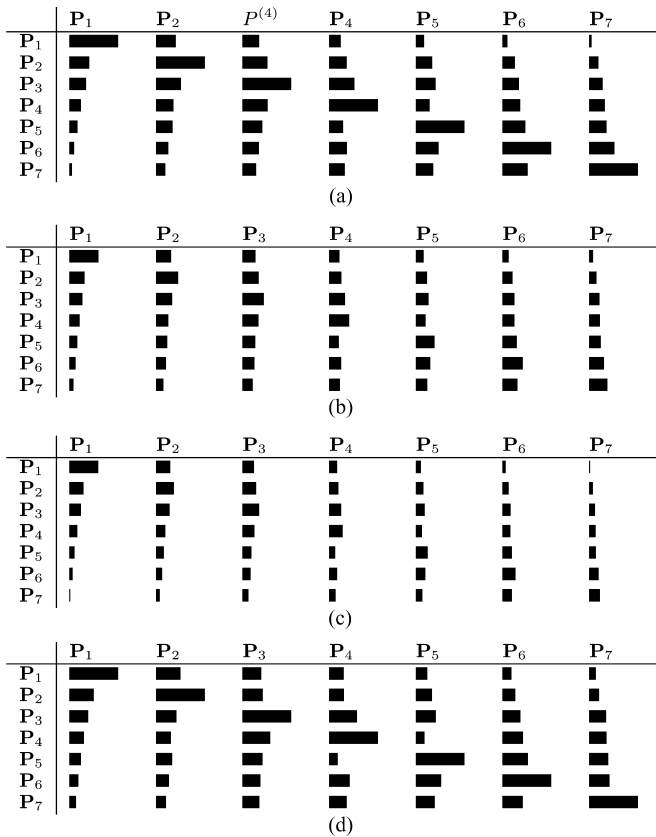


Fig. 3. Visual representation of similarity between fuzzy partitions (first experiment, Wine data, Jaccard). (a) HRHS. (b) Campello. (c) Anderson. (d) Runkler.

Thus, in contrast with previous proposals, our extension exhibits desirable metrical properties. Indeed, our review of existing approaches has revealed a number of potential shortcomings, which are, in a sense, also confirmed by our experimental study.

Apart from generalizing the Rand index, we also provided the basis to extend our approach to other similarity measures which are defined in terms of the same basic quantities, namely the numbers  $a$ ,  $b$ ,  $c$ , and  $d$  of concordant and discordant object pairs. Even though our results allow such measures to be extended to the case of fuzzy partitions, it is of course not clear which of the metrical properties will be preserved by this extension. Just like in the case of the Rand index, we are, therefore, interested in studying the formal properties of fuzzy extensions of specific measures such as, for example, the Jaccard index.

## REFERENCES

- [1] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "Clustering validity checking methods: Part I," *ACM SIGMOD Rec.*, vol. 31, no. 2, 2002.
- [2] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "Clustering validity checking methods: Part II," *ACM SIGMOD Rec.*, vol. 31, no. 3, pp. 19–27, 2002.
- [3] J. Beringer and E. Hüllermeier, "Fuzzy clustering of parallel data streams," in *Advances in Fuzzy Clustering and Its Application*, J. V. de Oliveira and W. Pedrycz, Eds. New York: Wiley, 2007, pp. 333–352.
- [4] M. Windham, "Cluster validity for the fuzzy c-means algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-4, no. 4, pp. 357–363, Jul. 1982.
- [5] X. Xie and G. Beni, "Validity measure for fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 3, no. 8, pp. 841–846, Aug. 1991.
- [6] E. Hüllermeier and M. Rifqi, "A fuzzy variant of the Rand index for comparing clustering structures," in *Proc. Int. Joint 2009 Fuzzy System*

*Association World Congress and 2009 Euro. Soc. Fuzzy Logic and Technol. Conf.*, Lisbon, Portugal, 2009, pp. 1294–1298.

- [7] R. Campello, "A fuzzy extension of the Rand index and other related indexes for clustering and classification assessment," *Pattern Recognit. Lett.*, vol. 28, no. 7, pp. 833–841, 2007.
- [8] H. Frigui, C. Hwang, and F. C.-H. Rhee, "Clustering and aggregation of relational data with applications to image database categorization," *Pattern Recognit.*, vol. 40, pp. 3053–3068, 2007.
- [9] R. Brouwer, "Extending the Rand, adjusted Rand and Jaccard indices to fuzzy partitions," *J. Intell. Inf. Syst.*, vol. 32, pp. 213–235, 2009.
- [10] D. Anderson, J. Bezdek, M. Popescu, and J. Keller, "Comparing fuzzy, probabilistic and possibilistic partitions," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 5, pp. 906–918, 2010.
- [11] E. Klement, R. Mesiar, and E. Pap, *Triangular Norms*. Norwell, MA: Kluwer, 2002.
- [12] A. D. Luca and S. Termini, "A definition of non-probabilistic entropy in the setting of fuzzy sets theory," *Inf. Control*, vol. 24, pp. 301–312, 1972.
- [13] U. Kaymak and H. van Nauta Lemke, "A sensitivity analysis approach to introducing weight factors into decision functions in fuzzy multicriteria decision making," *Fuzzy Sets Syst.*, vol. 97, no. 2, pp. 169–182, 1998.
- [14] T. Runkler, "Comparing partitions by subset similarities," in *Proc. Information Processing and Management of Uncertainty*, Dortmund, Germany, 2010, pp. 29–38.
- [15] B. D. Baets, S. Janssens, and H. D. Meyer, "Meta-theorems on inequalities for scalar fuzzy set cardinalities," *Fuzzy Sets Syst.*, vol. 157, no. 11, pp. 1463–1476, 2006.
- [16] E. Ruspini, "A new approach to clustering," *Inf. Control*, vol. 15, pp. 22–32, 1969.
- [17] N. Schmechel, "On the isomorphic lattices of fuzzy equivalence relations and fuzzy partitions," *Multi-Valued Logic*, vol. 2, pp. 1–46, 1996.
- [18] H. Thiele and N. Schmechel, "On the mutual definability of fuzzy equivalence relations and fuzzy partitions," in *Proc. FUZZ-IEEE-95*, Yokohama, Japan, 1995, pp. 1383–1390.
- [19] R. Mesiar, B. Reusch, and H. Thiele, "Fuzzy equivalence relations and fuzzy partitions," *J. Multi-Valued Logic Soft Comput.*, vol. 12, pp. 167–181, 2006.
- [20] P. Jaccard, "étude comparative de la distribution florale dans une portion des Alpes et des Jura," *Bulletin de la Société Vaudoise des Sciences Naturelles*, vol. 37, pp. 547–579, 1901.
- [21] C. Alsina, "On a family of connectives for fuzzy sets," *Fuzzy Sets Syst.*, vol. 16, pp. 231–235, 1985.
- [22] M. Kendall, *Rank Correlation Methods*. London, U.K.: Griffin, 1955.
- [23] J. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithm*. New York: Plenum, 1981.
- [24] F. Höppner, F. Klawonn, F. Kruse, and T. Runkler, *Fuzzy Cluster Analysis*. Chichester, U.K.: Wiley, 1999.
- [25] S. Garcia, A. Fernandez, J. Luengo, and F. Herrera, "Advanced non-parametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power," *Inf. Sci.*, vol. 180, pp. 2044–2064, 2010.



**Eyke Hüllermeier** received the M.Sc. degrees in mathematics and business computing, the Ph.D. degree in computer science, and the Habilitation degree, all from the University of Paderborn, Paderborn, Germany.

He is currently a Full Professor with the Department of Mathematics and Computer Science, University of Marburg, Marburg, Germany. His research interests include machine learning and data mining, fuzzy set theory, uncertainty and approximate reasoning, and applications in bioinformatics. He has

published numerous research papers on these topics in journals and major international conferences.

Dr. Hüllermeier is a member of the IEEE Computational Intelligence Society and a board member of the European Society for Fuzzy Logic and Technology (EUSFLAT). He is the Coeditor-in-Chief of *Fuzzy Sets and Systems* and is on the editorial board of several other journals. Moreover, he is the Coordinator of the EUSFLAT working group on learning and data mining, and the Head of the IEEE Computational Intelligence Society task force on machine learning.



**Maria Rifqi** received the M.Sc. degree in computer science and operational research in 1993 and the Ph.D. degree, as well as the Habilitation degree in computer science, from the University Pierre et Marie Curie, Paris, France, in 1996 and 2010, respectively.

She has been an Associate Professor with the University Panthéon-Assas, Paris, and with LIP6, which is the computer science laboratory of the University Pierre et Marie Curie, Paris, since 1998. Her research interests include fuzzy logic, approximate reasoning, and prototype learning. She specializes in similarity

measures and machine learning algorithms.

Dr. Rifqi was the Scientific Chair of the 2010, the French conference on fuzzy logic and its applications, as well as the Special Sessions Chair of FUZZ'IEEE 2010 and the local arrangement chair of IEEE Symposium Series on Computational Intelligence 2011. She is the Machine Learning and Data Mining Area Editor of the *International Journal of Uncertainty and Fuzziness in Knowledge Based Systems*.



**Sascha Henzgen** received the Bachelor of Science degree in computer science from the University of Siegen, Siegen, Germany, in 2010. He is currently working toward the Master's degree with the Phillips University of Marburg, Marburg, Germany, and is a graduate student with the group of Prof. Hüllermeier in the area of machine learning.



**Robin Senge** received the M.Sc. degree in computer science from the University of Marburg, Marburg, Germany, in 2006.

After graduation, he worked as a Software Engineer in industry, where he was consigned to consulting and developed software for financial applications like trading and portfolio management systems. In 2009, he joined the Knowledge Engineering and Bioinformatics Laboratory, University of Marburg, as a doctoral student. His research interests include machine learning and fuzzy systems.