

RESEARCH

Open Access



Novel similarity measure between hesitant fuzzy set and their applications in pattern recognition and clustering analysis

Rakhi Gupta^{1*} and Satish Kumar¹

*Correspondence:
rakhigupta.k@gmail.com

¹ Department of Mathematics,
Maharishi Markandeshwar
(Deemed To Be University),
Mullana, Ambala 133207, India

Abstract

The extension of classical fuzzy sets are hesitant fuzzy sets (HFSs), in which each element has a possible value from $[0,1]$. Similarity and distance measures are useful implements for solving medical, clustering and pattern-recognition problems. Most of the researchers have suggested their ideas for HFSs using distance measures and extract the similarity measure from distance measure but most of them are getting inadequate results. Therefore, we proposed a new similarity measure to resolve these problems and also satisfied the properties of proposed measure for HFSs. Additionally, numerous examples are taken in consideration using HFS and compared the performance of existing measures with proposed measure for different cases. Furthermore, we have applied proposed measure for pattern recognition problems using three different examples and also calculate performance index (i.e., Degree of Confidence) to explore the behavior of different measures. Finally, we suggested MST based clustering algorithm using HF-environment and contrast the performance of proposed measure with existing ones. All these comparison illustrate that proposed measure is getting efficient and reasonable results and it also verified that proposed measure is not restricted to particular domain, it can be effectively applied for diverse field of application.

Keywords: Hesitant fuzzy sets (HFSs), Similarity measure, Pattern-recognition, Clustering, Maximum spanning tree (MST), Degree of Confidence (DOC)

Introduction

Initially, the only approach to estimate the ambiguity was probability. Nonetheless, all kind of unpredictability in daily life cannot computed through probability such as very smart, low price and fast speed etc. as these indistinct terms could not exhibited by exact terms. Thus, Zadeh [1] introduced fuzzy set theory to tackle these uncertainties and found suitable in many applications like approximate reasoning, decision-making, and fuzzy control. Yager [2] suggested his fuzzy information measures. However it is found difficult to solve some of the practical applications using fuzzy set. Consequently, some new strategies for-instance Intuitionistic fuzzy set (IFS) suggested by Atanassov [3], interval-valued fuzzy set by Zadeh [4], Type 2 fuzzy set introduced by Dubois [5] and

fuzzy multi-set by Yager [6] were suggested which are the expansion of FS. The above expansions are centered on the hypothesis that it is uncertain to allocate the belongingness degree of an element to a fixed set Torra and Narukawa [7, 8]. Gupta and Kumar [9, 10] suggested their approach using IFs in MCDM (Multi criteria decision making) problems. Membership degree has one specific value in all of these extensions including fuzzy set (FS). However, in practical it is not always true that same membership value will assign to an alternative. This situation can be better explain using following example.

Suppose a company wants to take a decision through its governing council. As we know governing council has lot of members having different backgrounds, knowledge, expertise and qualifications. Thus, for a particular decision it is not necessary that all the members will assign same membership value to an alternative. For instance, some decision makers assigned 0.4, some provided 0.5 and others have assigned 0.6 as the membership degree. Therefore, it is not possible for them to commemorate each other. In that case HFS is found more powerful to cope up with this problem than all the other extensions of FS. Thus, we can represent this problem with hesitant fuzzy element (HFE) $\{0.4,0.5,0.6\}$ and it will express the problem more impartially than interval-valued fuzzy numbers $[0.4,0.6]$ or crisp number 0.4 (or 0.5 or 0.6) or intuitionistic fuzzy numbers $(0.4,0.6)$. Subsequently, idea of HFSs were presented by [7, 8] using the membership function with possible set values. However, HFS can consider the human hesitance more equitably as compared to different expansion of fuzzy set, Thus HFS became an effective concept to tackle with unpredictability or unreliability. Thus, in the short span of time it has fascinated the curiosity of most of the researchers [11–14] dealing in evaluation and decision-making process. Chen et al. [15] and Singh and Lalotra [16] explored the HFS using correlation coefficient and applied in clustering analysis. Some researchers [17–19] also suggested their approach using HFSs in the application of decision-making. Suo et al. [20] suggested information measure using HFS whereas Singh [21] suggested dual HFS based similarity and distance measure and applied in decision making. Dual hesitant fuzzy set using the concept of correlation coefficient was suggested by Tyagi [22]. Further, hesitant fuzzy prioritized operator was explored by Wei [23] whereas, generalized HF Bonferroni mean operator was suggested by Yu et al. [24]. Some of them introduced their concept using HFS theory in the field of decision-making [25–27]. Using the idea of Archimedean t-conorm and t-norm dual HF power aggregation operator was suggested by Wang et al. [28] for the problems of MAGDM. Further, Frank aggregation operator suggested by Qin et al. [29] and Hamacher aggregation operator by Tan et al. [30] using the notion of HFSs and applied in the field of MCDM.

The two main indexes used in FS theory are similarity and distance measure, considerably used in different fields like pattern-recognition, appropriate reasoning, decision-making, machine learning and market prediction. First of all, concept of similarity measure was presented by Wang [31]. Geometric distance and Hausdorff metrics was explored by Zwick et al. [32] and further comparison of similarity measure of FS have also given by them. Thereafter, the fundamental definition of inclusion measure similarity measure (SM) was examined by Zeng and Li [33]. Gupta and Kumar [34, 35] suggested their similarity measures in different field like pattern recognition and clustering. From the last few years distinct researchers have suggested their work using hesitant fuzzy set such as similarity and distance measure using HFSs was suggested by Xu and

Xia [36], correlation and distance measure was suggested by [37, 38], entropy measure by Xu and Xia [39] whereas generalized HFSWDM (hesitant fuzzy synergetic weighted distance measure) was suggested by Peng et al. [40] and used in MCDM problems. Some new similarity and distance measures were suggested by Zhang and Xu [41] using HFSs in the application of clustering. Ahmad and Khan [42] recognized different themes for the study of mixed data clustering. Different authors suggested K-means algorithm and applied in different applications like distributed memory multiprocessor [43], edge computing environment [44] and multi core CPU [45, 46]. Lapegna and Stranieri [47] suggested direction based clustering algorithm. Lapegna et al. [48] suggested an adaptive approach with K-means clustering algorithm. Laccetti et al. [49, 50] suggested different clustering algorithm for edge computing environment. Distance, similarity and information measure of HFSs were studied by Farhadinia [51, 52]. He further extended his work for interval-valued and higher order HFSs. The idea of hesitance degree was suggested by Li et al. [53, 54] and gave new formulas for calculation of similarity measures. Zeng et al. [55] also suggested his similarity measures using hesitant fuzzy set in pattern-recognition. Cosine based similarity and distance measure was suggested by Liao et al. [56] and used for decision-making problems.

Methods

It should be noted that existing measures have calculated their values based on distance while some of the researchers have converted their distance measures into similarity measures. Some of the existing measures are not achieving reasonable results in some of the cases while some of them have introduced hesitant degree based measures for HFS and get inadequate results. This encouraged us to come up with new similarity measure for HFS. In the view of complication of practical problems, it is necessary for us to proposed similarity measures which will make their calculation simpler and they are more useful to solve the different problems like clustering, pattern-recognition and approximate reasoning. Therefore, *main highlights of this paper is:*

- A novel similarity measure using HFSs has been introduced and also proved the properties of proposed measure for HFSs.
- Thereafter, numerical and comparative analysis has been performed that includes the observation of different cases of HFSs, pattern-recognition problems.
- Moreover, we have also calculated DOC as performance index with the aid of numerical illustration.
- Finally, hesitant fuzzy clustering algorithm has been developed and with the help of proposed measure applied it in a numerical example to compare its potency than existing measure and demonstrate the usefulness and acceptance of proposed method.

The outline of paper is as follows: the “**Preliminaries**” section deals with basic FS, HFS and its associated properties. Existing distance and similarity measures suggested by distinct authors are covered in the “**Existing similarity measures**” section. The “**Proposed new HF-similarity measure**” section covers the proposed similarity measure with its validation. The “**Numerical and comparative analysis**” section covers the numerical

experiment for pattern-recognition and clustering problems. Last section forwards the conclusion and scope of improvement.

Preliminaries

Definition 1 Suppose $Y = \{y_1, y_2, \dots, y_v\}$, be a finite universe of discourse. A fuzzy set (FS) U for $y_n \in Y$ defined by Zadeh [1] as:

$$U = \{(y_n, f_U(y_n)) | y_n \in Y, n = 1, 2, \dots, v\},$$

where $f_U(y_n)$ denotes membership degree such that $0 \leq f_U(y_n) \leq 1$.

Definition 2 Torra and Narukawa [7, 8]. Consider hesitant fuzzy set (HFS) B on Y which is function whenever apply to Y return a subset between $[0,1]$ and described as:

$$B = \{(y_n, h_B(y_n)) | y_n \in Y, n = 1, 2, \dots, v\},$$

where $h_B(y_n)$ is a set having values between $[0,1]$ with degree of membership of element $y \in Y$ in set B . Xu and Xia [36] considers $h_B(y_n)$ as hesitant fuzzy element (HFE) for comfort.

The different operations like union, intersection and complement can be represented as defined in subsequent definition.

Definition 3 For hesitant fuzzy elements h_{B_1} , h_{B_2} and h_B , the different operations described by Torra and Narukawa [7, 8] as:

1. Lower bound: $h_B^-(y_n) = \min h_B(y_n)$;
2. Upper bound: $h_B^+(y_n) = \max h_B(y_n)$;
3. $h_B^c = \cup_{\alpha \in h_B} \{1 - \alpha\}$;
4. $h_{B_1} \cup h_{B_2} = \{h_B \in h_{B_1} \cup h_{B_2} | h_B \geq \max(h_{B_1}^-, h_{B_2}^-)\}$;
5. $h_{B_1} \cap h_{B_2} = \{h_B \in h_{B_1} \cup h_{B_2} | h_B \leq \min(h_{B_1}^+, h_{B_2}^+)\}$.

Xu and Xia [36] describes the above two operations in the following form:

6. $h_{B_1} \cup h_{B_2} = \cup_{\alpha_1 \in h_{B_1}, \alpha_2 \in h_{B_2}} \max\{\alpha_1, \alpha_2\}$;
7. $h_{B_1} \cap h_{B_2} = \cup_{\alpha_1 \in h_{B_1}, \alpha_2 \in h_{B_2}} \min\{\alpha_1, \alpha_2\}$,

and also described operational laws regarding HFEs h_{B_1} , h_{B_2} and h_B as:

Definition 4 Xu and Xia [36] Let us consider different HFEs h_{B_1} , h_{B_2} and h_B and β can be consider as positive real number, then

- (a) $h_B^\beta = \cup_{\alpha \in h_B} \{\alpha^\beta\}$;
- (b) $\beta h_B = \cup_{\alpha \in h_B} \{1 - (1 - \alpha)^\beta\}$;
- (c) $h_{B_1} \oplus h_{B_2} = \cup_{\alpha_1 \in h_{B_1}, \alpha_2 \in h_{B_2}} \{\alpha_1 + \alpha_2 - \alpha_1 \alpha_2\}$;
- (d) $h_{B_1} \otimes h_{B_2} = \cup_{\alpha_1 \in h_{B_1}, \alpha_2 \in h_{B_2}} \{\alpha_1 \alpha_2\}$.

Let h_{B_k} ($k = 1, 2, \dots, j$) consider as HFEs collection, (c) and (d) of Definition 4 was summarized by Liao et al. [57] as:

$$\begin{aligned} \text{(e)} \quad \oplus_{k=1}^j h_{B_k} &= \cup_{\alpha_k \in h_{B_k}} \left\{ 1 - \prod_{k=1}^j (1 - \alpha_k) \right\}; \\ \text{(f)} \quad \otimes_{k=1}^j h_{B_k} &= \cup_{\alpha_k \in h_{B_k}} \left\{ \prod_{k=1}^j \alpha_k \right\}. \end{aligned}$$

For different HFEs the number of values may be different. Let $\bar{l}_{h_B(y_n)}$ be the length of $h_B(y_n)$. consider $\bar{l} = \max\{\bar{l}_{h_{B_1}}, \bar{l}_{h_{B_2}}\}$ for two hesitant fuzzy elements h_{B_1} and h_{B_2} . To work properly, a rule has been provided by Xu and Xia [36] based on a hypothesis that decision makers are negative which is as follows: For $\bar{l}_{h_{B_1}} < \bar{l}_{h_{B_2}}$, we will add the minimum value in h_{B_1} until its length will be same as that of h_{B_2} ; if $\bar{l}_{h_{B_1}} > \bar{l}_{h_{B_2}}$, then we add the minimum value in h_{B_2} so that its length will same as that of h_{B_1} .

Liao et al. [57] introduced the following theorem according to above operational laws:

Theorem 1 Liao et al. [57]. Let us consider h_{B_1} and h_{B_2} as two HFEs, we can write

$$\bar{l}_{h_{B_1} \oplus h_{B_2}} = \bar{l}_{h_{B_1}} \times \bar{l}_{h_{B_2}}, \bar{l}_{h_{B_1} \otimes h_{B_2}} = \bar{l}_{h_{B_1}} \times \bar{l}_{h_{B_2}} \tag{1}$$

This operation also holds having j different HFEs, i.e.,

$$\bar{l}_{\oplus_{m=1}^j h_{B_m}} = \prod_{m=1}^j \bar{l}_{h_{B_m}}, \bar{l}_{\otimes_{m=1}^j h_{B_m}} = \prod_{m=1}^j \bar{l}_{h_{B_m}} \tag{2}$$

Example 1 Let $h_{B_1} = \langle 0.3, 0.4, 0.6 \rangle$ and $h_{B_2} = \langle 0.1, 0.3 \rangle$ are two HFEs, according to operational laws of HFSs defined in Definition 4, we can write

$$\begin{aligned} h_{B_1} \oplus h_{B_2} &= \cup_{\alpha_1 \in h_{B_1}, \alpha_2 \in h_{B_2}} \{ \alpha_1 + \alpha_2 - \alpha_1 \alpha_2 \} \\ &= \{ 0.3 + 0.1 - 0.3 \times 0.1, 0.3 + 0.3 - 0.3 \times 0.3, 0.4 + 0.1 - 0.4 \times 0.1, 0.4 + 0.3 - 0.4 \times 0.3, \\ &\quad 0.6 + 0.1 - 0.6 \times 0.1, 0.6 + 0.3 - 0.6 \times 0.3 \}; \\ &= \{ 0.37, 0.51, 0.46, 0.58, 0.64, 0.72 \} \end{aligned}$$

$$\begin{aligned} h_{B_1} \otimes h_{B_2} &= \cup_{\alpha_1 \in h_{B_1}, \alpha_2 \in h_{B_2}} \{ \alpha_1 \alpha_2 \} = \{ 0.3 \times 0.1, 0.3 \times 0.3, 0.4 \times 0.1, 0.4 \times 0.3, 0.6 \times 0.1, 0.6 \times 0.3 \} \\ &= \{ 0.03, 0.09, 0.04, 0.12, 0.06, 0.18 \} \end{aligned}$$

and then, $\bar{l}_{h_{B_1} \oplus h_{B_2}} = 6 = 3 \times 2 = \bar{l}_{h_{B_1}} \times \bar{l}_{h_{B_2}}, \bar{l}_{h_{B_1} \otimes h_{B_2}} = 6 = 3 \times 2 = \bar{l}_{h_{B_1}} \times \bar{l}_{h_{B_2}}$.

We observe that length of the derived HFE will increase after applying the above mentioned operations. Thus, calculations complexity will also increase. Therefore, they also suggested new methodology so that length of the derived HFE will decrease while handling HFEs. The modified operational laws of Definition 4 are as follows:

Definition 5 Liao et al. [57] Let us consider a collection of HFEs $B = \{ h_{B_1}, h_{B_2}, \dots, h_{B_j} \}$ and consider β as positive real number, then

$$1. h_B^\beta = \left\{ \left(h_B^{q(s)} \right)^\beta \mid s = 1, 2, \dots, t \right\};$$

$$\begin{aligned}
 2. \beta h_B &= \left\{ 1 - \left(1 - h_B^{q(s)} \right)^\beta \mid s = 1, 2, \dots, t \right\}; \\
 3. h_{B_1} \oplus h_{B_2} &= \left\{ h_{B_1}^{q(s)} + h_{B_2}^{q(s)} - h_{B_1}^{q(s)} h_{B_2}^{q(s)} \mid s = 1, 2, \dots, t \right\}; \\
 4. h_{B_1} \otimes h_{B_2} &= \left\{ h_{B_1}^{q(s)} h_{B_2}^{q(s)} \mid s = 1, 2, \dots, t \right\}; \\
 5. \bigoplus_{k=1}^j h_{B_k} &= \left\{ 1 - \prod_{k=1}^j \left(1 - h_{B_k}^{q(s)} \right) \mid s = 1, 2, \dots, t \right\}; \\
 6. \bigotimes_{k=1}^j h_{B_k} &= \left\{ \prod_{k=1}^j h_{B_k}^{q(s)} \mid s = 1, 2, \dots, t \right\},
 \end{aligned}$$

where $h_{B_k}^{q(s)}$ is the sth smallest value in h_{B_k} .

Example 2 Consider $h_{B_1} = \langle 0.1, 0.2, 0.4, 0.7 \rangle$ and $h_{B_2} = \langle 0.3, 0.5, 0.6 \rangle$ are two HFEs. After implementing summation and multiplication operation as defined in Definition 5, we have

$$\begin{aligned}
 h_{B_1} \oplus h_{B_2} &= \left\{ h_{B_1}^{q(s)} + h_{B_2}^{q(s)} - h_{B_1}^{q(s)} h_{B_2}^{q(s)} \mid s = 1, 2, 3, 4 \right\} \\
 &= \{0.1 + 0.3 - 0.1 \times 0.3, 0.2 + 0.4 - 0.2 \times 0.4, 0.4 + 0.5 - 0.4 \times 0.5, 0.7 + 0.6 - 0.7 \times 0.6\} \\
 &= \{0.37, 0.52, 0.7, 0.88\}
 \end{aligned}$$

$$\begin{aligned}
 h_{B_1} \otimes h_{B_2} &= \left\{ h_{B_1}^{q(s)} h_{B_2}^{q(s)} \mid s = 1, 2, 3, 4 \right\} = \{0.1 \times 0.3, 0.2 \times 0.4, 0.4 \times 0.5, 0.7 \times 0.6\} \\
 &= \{0.03, 0.08, 0.20, 0.42\}
 \end{aligned}$$

The score function of HFE was described by Xia and Xu [58] to obtain the MAX and MIN operators of two HFEs.

Definition 6 Xia and Xu [58] For a HFE h_B , $c(h_B) = \frac{1}{l_{h_B}} \sum_{\alpha \in h_B} \alpha$ is known as the score function of h_B , l_{h_B} consider as number of values of h_B . Let us consider HFEs as h_{B_1} and h_{B_2} , if $c(h_{B_1}) > c(h_{B_2})$, then $h_{B_1} > h_{B_2}$; if $c(h_{B_1}) = c(h_{B_2})$, then $h_{B_1} = h_{B_2}$.

Although, this comparison rule could not discriminate between two HFEs in some specific cases. Therefore, to cope up with this problem the variance function of HFE was suggested by Liao et al. [57] and also introduced a new method for ranking HFEs.

Definition 7 Liao et al. [57] consider a HFE h_B , $\bar{v}(h_B) = \frac{1}{l_{h_B}} \sqrt{\sum_{\alpha_k, \alpha_m \in h_B} (\alpha_k - \alpha_m)^2}$ is consider as variance function of h where l_{h_B} = number of values of h_B and $\bar{v}(h_B)$ = variance degree of h . Considering HFEs h_{B_1} and h_{B_2} , if $\bar{v}(h_{B_1}) > \bar{v}(h_{B_2})$ then $h_{B_1} < h_{B_2}$; if $\bar{v}(h_{B_1}) = \bar{v}(h_{B_2})$, then $h_{B_1} = h_{B_2}$.

The connection of score function and variance function is same as that of mean and variance in statistics. To contrast two HFEs, a strategy can be obtained simply by taking score function $c(h_B)$ and the variance function $\bar{v}(h_B)$ into consideration:

Strategy:

if $c(h_{B_1}) < c(h_{B_2})$, then $h_{B_1} < h_{B_2}$, **Max** $\{h_{B_1}, h_{B_2}\} = h_{B_2}$, and **Min** $\{h_{B_1}, h_{B_2}\} = h_{B_1}$; if $c(h_{B_1}) = c(h_{B_2})$, then

(a) if $\bar{v}(h_{B_1}) < \bar{v}(h_{B_2})$, then $h_{B_1} > h_{B_2}$, **Max** $\{h_{B_1}, h_{B_2}\} = h_{B_1}$, and **Min** $\{h_{B_1}, h_{B_2}\} = h_{B_2}$;

(b) if $\bar{v}(h_{B_1}) = \bar{v}(h_{B_2})$, then $h_{B_1} = h_{B_2}$, $\mathbf{Max}\{h_{B_1}, h_{B_2}\} = h_{B_1} = \mathbf{Min}\{h_{B_1}, h_{B_2}\} = h_{B_2}$.

Example 3 Let $h_{B_1} = \langle 0.2, 0.2, 0.5 \rangle$ and $h_{B_2} = \langle 0.3, 0.3 \rangle$ be two HFEs. Then by Definition 7, we have

$$c(h_{B_1}) = \frac{0.2 + 0.2 + 0.5}{3} = 0.3, \quad c(h_{B_2}) = \frac{0.3 + 0.3}{2} = 0.3$$

$$\bar{v}(h_{B_1}) = \frac{\sqrt{0 + 0.3^2 + 0.3^2}}{3} = 0.14, \quad \bar{v}(h_{B_2}) = \frac{\sqrt{0}}{2} = 0.0$$

Then, $\bar{v}(h_{B_1}) > \bar{v}(h_{B_2})$, that is the variance degree of h_{B_2} is smaller than h_{B_1} , thus, $h_{B_1} < h_{B_2}$.

Definition 8 Xu and Xia [36] For HFSs B, F and D , Distance measure $\bar{d}(B, F)$ should satisfied the following conditions:

1. $0 \leq \bar{d}(B, F) \leq 1$.
2. $\bar{d}(B, F) = \bar{d}(F, B)$.
3. $\bar{d}(B, F) = 1$ iff $B = F$.
4. If $B \subseteq F \subseteq D$, then $\bar{d}(B, D) \geq \bar{d}(B, F)$ and $\bar{d}(B, D) \geq \bar{d}(F, D)$.

Definition 9 Xu and Xia [36] For HFSs B, F , and D , similarity measure $\bar{S}(B, F)$ should satisfied the following conditions:

1. $0 \leq \bar{S}(B, F) \leq 1$.
2. $\bar{S}(B, F) = \bar{S}(F, B)$.
3. $\bar{S}(B, F) = 1$ iff $B = F$.
4. If $B \subseteq F \subseteq D$, then $\bar{S}(B, D) \leq \bar{S}(B, F)$ and $\bar{S}(B, D) \leq \bar{S}(F, D)$.

Remark 1 It noticed that $\bar{S}(B, F) = 1 - \bar{d}(B, F)$ is distance measure.

Existing similarity measures

Some distinct distance measures were suggested by Xu and Xia [36] for HFEs.

(1). For any two HFEs $h_B^{q(k)}$ and $h_F^{q(k)}$, the Manhattan distance $s(h_B, h_F)$ is given by Xu and Xia [36]

$$\bar{s}_{man}(h_B, h_F) = 1 - \bar{d}_{man}(h_B, h_F) = 1 - \frac{1}{\bar{l}_{y_n}} \sum_{k=1}^{\bar{l}_{y_n}} \left| h_B^{q(k)}(y_n) - h_F^{q(k)}(y_n) \right|, \quad (3)$$

where $h_B^{q(k)}(y_n)$ and $h_F^{q(k)}(y_n)$ are the k^{th} largest values in $h_B(y_n)$ and $h_F(y_n)$, respectively and $\bar{l}_{y_n} = \max\{\bar{l}(h_B(y_n)), \bar{l}(h_F(y_n))\}$.

(2). Consider two HFEs $h_B^{q(k)}$ and $h_F^{q(k)}$ for a universal set B , Xu and Xia [36] suggested hesitant normalized Hamming distance, the hesitant normalized Euclidean distance and the generalized hesitant normalized distance using the concept of Hamming and Euclidean distance and can be represented as:

$$\bar{s}_{hnh}(h_B, h_F) = 1 - \bar{d}_{hnh}(h_B, h_F) = 1 - \frac{1}{v} \sum_{n=1}^v \left[\frac{1}{\bar{l}_{y_n}} \sum_{k=1}^{\bar{l}_{y_n}} \left| h_B^{q(k)}(y_n) - h_F^{q(k)}(y_n) \right| \right], \quad (4)$$

$$\bar{s}_{hne}(h_B, h_F) = 1 - \bar{d}_{hne}(h_B, h_F) = 1 - \left[\frac{1}{v} \sum_{n=1}^v \left(\frac{1}{\bar{l}_{y_n}} \sum_{k=1}^{\bar{l}_{y_n}} \left| h_B^{q(k)}(y_n) - h_F^{q(k)}(y_n) \right|^2 \right) \right]^{\frac{1}{2}}, \tag{5}$$

$$\bar{s}_{ghn}(h_B, h_F) = 1 - \bar{d}_{ghn}(h_B, h_F) = 1 - \left[\frac{1}{v} \sum_{n=1}^v \left(\frac{1}{\bar{l}_{y_n}} \sum_{k=1}^{\bar{l}_{y_n}} \left| h_B^{q(k)}(y_n) - h_F^{q(k)}(y_n) \right|^\gamma \right) \right]^{\frac{1}{\gamma}}, \tag{6}$$

where $h_B^{q(k)}(y_n)$ and $h_F^{q(k)}(y_n)$ are the k^{th} largest values in $h_B(y_n)$ and $h_F(y_n)$, respectively and $\bar{l}_{y_n} = \max\{\bar{l}(h_B(y_n)), \bar{l}(h_F(y_n))\}$ and $\gamma > 0$.

(3). The hesitant normalized Hamming-Hausdorff distance, normalized Euclidean-Hausdorff distance, Hybrid hesitant normalized Hamming distance, and Hybrid hesitant normalized Euclidean distance were also suggested by Xu and Xia [36] which can be represented as:

$$\bar{s}_{hnhh}(h_B, h_F) = 1 - \bar{d}_{hnhh}(h_B, h_F) = 1 - \frac{1}{v} \sum_{n=1}^v \max_k \left| h_B^{q(k)}(y_n) - h_F^{q(k)}(y_n) \right|, \tag{7}$$

$$\bar{s}_{hneh}(h_B, h_F) = 1 - \bar{d}_{hneh}(h_B, h_F) = 1 - \left[\frac{1}{v} \sum_{n=1}^v \max_k \left| h_B^{q(k)}(y_n) - h_F^{q(k)}(y_n) \right|^2 \right]^{\frac{1}{2}}, \tag{8}$$

$$\bar{s}_{hhmh}(h_B, h_F) = 1 - \bar{d}_{hhmh}(h_B, h_F) = 1 - \frac{1}{2v} \sum_{n=1}^v \left[\frac{1}{\bar{l}_{y_n}} \sum_{k=1}^{\bar{l}_{y_n}} \left| h_B^{q(k)}(y_n) - h_F^{q(k)}(y_n) \right| + \max_k \left| h_B^{q(k)}(y_n) - h_F^{q(k)}(y_n) \right| \right], \tag{9}$$

$$\bar{s}_{hhne}(h_B, h_F) = 1 - \bar{d}_{hhne}(h_B, h_F) = 1 - \frac{1}{2v} \sum_{n=1}^v \left[\frac{1}{\bar{l}_{y_n}} \sum_{k=1}^{\bar{l}_{y_n}} \left| h_B^{q(k)}(y_n) - h_F^{q(k)}(y_n) \right|^2 + \max_k \left| h_B^{q(k)}(y_n) - h_F^{q(k)}(y_n) \right| \right]^{\frac{1}{2}} \tag{10}$$

Now we will describe the definition suggested by Zeng [55] which is as mentioned in next definition.

Definition 10 Zeng [55]. Consider a HFS B on the universal set $Y = \{y_1, y_2, \dots, y_v\}$ and for any $y_n \in Y$, $\bar{l}(h_B(y_n))$ be the length of $h_B(y_n)$, $\rho(h_B(y_n))$ will be considered as the hesitance degree of $h_B(y_n)$ and is defined as follows

$$\rho(h_B(y_n)) = 1 - \frac{1}{\bar{l}(h_B(y_n))} \tag{11}$$

Various similarity and distance measures were suggested by Zeng [55] using above definition.

(4). For a universal set Y let us consider B and F as HFSs. The normalized Hamming, Euclidean and generalized distance including hesitance degree between B and F were introduced by Zeng [55] as:

$$\bar{s}_{hhd}(h_B, h_F) = 1 - \bar{d}_{hhd}(h_B, h_F) = 1 - \frac{1}{v} \sum_{n=1}^v \left[\frac{1}{1 + \bar{l}_{y_n}} \left(\frac{|\rho(h_B(y_n)) - \rho(h_F(y_n))|}{+\sum_{k=1}^{\bar{l}_{y_n}} |h_B^{q(k)}(y_n) - h_F^{q(k)}(y_n)|} \right) \right] \tag{12}$$

$$\bar{s}_{ehd}(h_B, h_F) = 1 - \bar{d}_{ehd}(h_B, h_F) = 1 - \frac{1}{v} \sum_{n=1}^v \left[\frac{1}{1 + \bar{l}_{y_n}} \left(\frac{|\rho(h_B(y_n)) - \rho(h_F(y_n))|^2}{+\sum_{k=1}^{\bar{l}_{y_n}} |h_B^{q(k)}(y_n) - h_F^{q(k)}(y_n)|} \right) \right]^{\frac{1}{2}} \tag{13}$$

$$\bar{s}_{ghd}(h_B, h_F) = 1 - \bar{d}_{ghd}(h_B, h_F) = 1 - \frac{1}{v} \sum_{n=1}^v \left[\frac{1}{1 + \bar{l}_{y_n}} \left(\frac{|\rho(h_B(y_n)) - \rho(h_F(y_n))|^\gamma}{+\sum_{k=1}^{\bar{l}_{y_n}} |h_B^{q(k)}(y_n) - h_F^{q(k)}(y_n)|} \right) \right]^{\frac{1}{\gamma}} \tag{14}$$

Definition 11 K. Rezaei [59] For HFS B , on the universal set Y such that $y_n \in Y$, $\varpi(h_B(y_n))$ is called the range of $h_B(y_n)$ and defined as follows:

$$\varpi(h_B(y_n)) = h_B^+(y_n) - h_B^-(y_n) \tag{15}$$

where $h_B^+(y_n) = \max h_B(y_n)$ and $h_B^-(y_n) = \min h_B(y_n)$. The range of hesitant fuzzy set B can be described as:

$$\varpi(B) = \frac{1}{v} \sum_{n=1}^v \varpi(h_B(y_n)) \tag{16}$$

(5). For a universal set Y , consider B and F as HFS. K. Rezaei [59] suggested various distance measures between B and F which can be defined as:

$$\bar{s}_{nhmh}(h_B, h_F) = 1 - \bar{d}_{nhmh}(h_B, h_F) = 1 - \frac{1}{v} \sum_{n=1}^v \left[\frac{\frac{1}{2 + \bar{l}_{y_n}} \left(|\varpi(h_B(y_n)) - \varpi(h_F(y_n))| + |\rho(h_B(y_n))| \right)}{-\rho(h_F(y_n)) + \sum_{k=1}^{\bar{l}_{y_n}} |h_B^{q(k)}(y_n) - h_F^{q(k)}(y_n)|} \right] \tag{17}$$

$$\bar{s}_{nhme}(h_B, h_F) = 1 - \bar{d}_{nhme}(h_B, h_F) = 1 - \frac{1}{v} \sum_{n=1}^v \left[\frac{\frac{1}{2 + \bar{l}_{y_n}} \left(|\varpi(h_B(y_n)) - \varpi(h_F(y_n))|^2 + |\rho(h_B(y_n))| \right)}{-\rho(h_F(y_n)) + \sum_{k=1}^{\bar{l}_{y_n}} |h_B^{q(k)}(y_n) - h_F^{q(k)}(y_n)|} \right]^{\frac{1}{2}} \tag{18}$$

$$\bar{s}_{nghm}(h_B, h_F) = 1 - \bar{d}_{nghm}(h_B, h_F) = 1 - \frac{1}{v} \sum_{n=1}^v \left[\frac{\frac{1}{2 + \bar{l}_{y_n}} \left(|\varpi(h_B(y_n)) - \varpi(h_F(y_n))|^\gamma + |\rho(h_B(y_n))| \right)}{-\rho(h_F(y_n)) + \sum_{k=1}^{\bar{l}_{y_n}} |h_B^{q(k)}(y_n) - h_F^{q(k)}(y_n)|} \right]^{\frac{1}{\gamma}} \tag{19}$$

where $\gamma > 0$.

Proposed new HF-similarity measure

Similarity measure for HFSs were suggested by distinct researchers as described in previous section. However, number of them are incapable to sort out the problems of decision making effectively. Some of them getting identical values on different cases while few of them are getting contrary and unreasonable results. Thus, it is necessary to characterize a new similarity measure which can overcome the short comings of existing ones. Therefore, we proposed a new similarity measure which is more advantageous to explore different applications effectively.

Consider a universal set $Y = \{y_1, y_2, \dots, y_v\}$ for two HFSs, i.e., B and F in Y . Now, we proposed the following HF-Similarity measure

$$\hat{s}(B, F) = \frac{1}{\bar{l}_{y_n}} \sum_{k=1}^{\bar{l}_{y_n}} \left(\frac{1}{1 + |h_B^{q(k)}(y_n) - h_F^{q(k)}(y_n)|} \right), \tag{20}$$

where $h_B^{q(k)}(y_n)$ and $h_F^{q(k)}(y_n)$ are the Hesitant Fuzzy elements in hesitant fuzzy sets B and F on Y . \bar{l}_{y_n} is the length of HFSs.

Definition 12 For a universal set $Y = \{y_1, y_2, \dots, y_v\}$, we consider two HFSs as B and F in Y . After that, we describe Similarity Measure $\bar{S} : HFS(Y) \times HFS(Y) \rightarrow [0, 1]$ as below:

$$\begin{aligned} \bar{S}(B, F) &= \frac{1}{v} \sum_{n=1}^v \hat{s}(B, F) \\ \bar{S}(B, F) &= \frac{1}{v} \sum_{n=1}^v \left[\frac{1}{\bar{l}_{y_n}} \sum_{k=1}^{\bar{l}_{y_n}} \left(\frac{1}{1 + |h_B^{q(k)}(y_n) - h_F^{q(k)}(y_n)|} \right) \right]. \end{aligned} \tag{21}$$

The normalized and generalized similarity measures

$$\bar{S}(B, F) = \frac{1}{v} \sum_{n=1}^v \left[\frac{1}{\bar{l}_{y_n}} \sum_{k=1}^{\bar{l}_{y_n}} \left(\frac{1}{1 + |h_B^{q(k)}(y_n) - h_F^{q(k)}(y_n)|} \right)^2 \right]^{\frac{1}{2}}. \tag{22}$$

$$\bar{S}(B, F) = \frac{1}{v} \sum_{n=1}^v \left[\frac{1}{\bar{l}_{y_n}} \sum_{k=1}^{\bar{l}_{y_n}} \left(\frac{1}{1 + |h_B^{q(k)}(y_n) - h_F^{q(k)}(y_n)|} \right) \right]^{\frac{1}{\lambda}}. \tag{23}$$

Example 4 For $Y = \{y_1\}$, we define the two HFSs on Y as $B = \{(y_1, (0.3658, 0.4655, 0.5659))\}$ and $F = \{(y_1, (0.2758, 0.3955, 0.6559))\}$. Here, we calculate the value of proposed similarity measure $\bar{S}(B, F)$ which is as follows:

$$\bar{S}(B, F) = \frac{1}{3} \left(\frac{1}{1 + |0.3658 - 0.2758|} + \frac{1}{1 + |0.4655 - 0.3955|} + \frac{1}{1 + |0.5659 - 0.6559|} \right) = 0.9232$$

Example 5 For $Y = \{y_1, y_2\}$, we define the two HFSs on Y as

$$B = \{(y_1, (0.3658, 0.4655, 0.5659)), (y_2, (0.2598, 0.3347, 0.5214))\}$$

and

$$F = \{(y_1, (0.2758, 0.3955, 0.6559)), (y_2, (0.3987, 0.4001, 0.6754))\}.$$

Here, we calculate the value of proposed similarity measure $\bar{S}(B, F)$ which is as follows:

$$\begin{aligned} \bar{S}(B, F) &= \frac{1}{2} \left[\frac{1}{3} \left(\frac{1}{1+|0.3658-0.2758|} + \frac{1}{1+|0.4655-0.3955|} + \frac{1}{1+|0.5659-0.6559|} + \frac{1}{1+|0.2598-0.3987|} \right) \right. \\ &\quad \left. + \frac{1}{1+|0.3347-0.4001|} + \frac{1}{1+|0.5214-0.6754|} \right] \\ &= 0.9089 \end{aligned}$$

Theorem 2 Measure $\bar{S}(B, F)$ holds the subsequent properties.

1. $0 \leq \bar{S}(B, F) \leq 1$
2. $\bar{S}(B, F) = 1$ iff $B = F$
3. $\bar{S}(B, F) = \bar{S}(F, B)$
4. If $B \subseteq F \subseteq D$, then $\bar{S}(B, D) \leq \bar{S}(B, F)$ and $\bar{S}(B, D) \leq \bar{S}(F, D)$.

Proof

(1) Let $x^{**} = |h_B^{q(k)}(y_n) - h_F^{q(k)}(y_n)|$ then we have $0 \leq \frac{1}{1+x^{**}} \leq 1$, for all $x^{**} \in \langle \text{spanclass} = 'reftype' \rangle [0, 1] \langle /span \rangle$.

Therefore, $0 \leq \hat{S}(B, F) \leq 1$. Using Eq. (21), $0 \leq \bar{S}(B, F) \leq 1$.

(2) Let $B = F$, then $h_B^{q(k)}(y_n) = h_F^{q(k)}(y_n)$ and $\forall y_n$. So $|h_B^{q(k)}(y_n) - h_F^{q(k)}(y_n)| = 0$, which gives $\bar{S}(B, F) = 1$.

(3) It is clear from Eq. (21).

(4) Let $B \subseteq F \subseteq D$, then $h_B^{q(k)}(y_n) \geq h_F^{q(k)}(y_n) \geq h_D^{q(k)}(y_n)$. So, we have

$$\left| h_B^{q(k)}(y_n) - h_F^{q(k)}(y_n) \right| \leq \left| h_B^{q(k)}(y_n) - h_D^{q(k)}(y_n) \right|, \left| h_F^{q(k)}(y_n) - h_D^{q(k)}(y_n) \right| \leq \left| h_B^{q(k)}(y_n) - h_D^{q(k)}(y_n) \right|$$

So, $\bar{S}(B, D) \leq \bar{S}(B, F)$ and $\bar{S}(B, D) \leq \bar{S}(F, D)$.

This proves the Theorem.

We will present numerical analysis in the coming section which reveals the effectiveness of proposed measure.

Numerical and comparative analysis

This section comprises numerical experiment on different sets, Pattern-recognition and Clustering analysis.

Numerical experiment

Here firstly numerical experiment has been employed to differentiate the similarity degree among HFSs. To perceive the ability of distinct measures, we have considered four different cases. The calculated values of existing and proposed measure at different pairs are as mentioned in Table 1, where contradictory results are shown by bold values. It can be easily observed from the Table 1 that all the existing measures are getting same values for different HFSs whereas proposed measure is getting different values. Therefore, it can be easily observed that proposed measure is getting reasonable and better results.

Let us take another experiment of HFSs with length two, to further explore the performance of different measures. For this purpose we have taken another three cases and the compiled results are as given in Table 2. \bar{s}_{man} getting similar values 1 and 3 case whereas \bar{s}_{hnh} is getting same results for all the three cases. \bar{s}_{hne} is also getting same value on 1 and 2 cases. \bar{s}_{hnhh} and \bar{s}_{hneh} both are getting similar values. In the similar fashion \bar{s}_{hhd} is also getting same outputs for all the three cases. \bar{s}_{ehd} , \bar{s}_{ghd} ($\gamma = 10$), \bar{s}_{nhne} and \bar{s}_{nghn} ($\gamma = 6$) are also getting same values for 1 and 2 cases. Now consider another HFSs with length three. The computed values are as tabulated in Table 3. \bar{s}_{man} has obtained the similar value at I^{st} and 3^{rd} case whereas \bar{s}_{hnh} has obtained the similar result for all

Table 1 Calculated values at different pairs

Cases	1	2	3	4
	$B = \langle y, 0.39 \rangle$	$B = \langle y, 0.3, 0.5 \rangle$	$B = \langle y, 0.45, 0.5, 0.6 \rangle$	$B = \langle y, 0.08, 0.17, 0.25 \rangle$
	$F = \langle y, 0.45 \rangle$	$F = \langle y, 0.4 \rangle$	$F = \langle y, 0.65, 0.7, 0.8 \rangle$	$F = \langle y, 0.33, 0.33, 0.42, 0.50 \rangle$
\bar{s}_{man}	0.9400	0.9000	0.8000	0.7500
\bar{s}_{hnh}	0.9400	0.9000	0.8000	0.7500
\bar{s}_{hne}	0.9400	0.9000	0.8000	0.7500
$\bar{s}_{ghn} \gamma = 6$	0.9400	0.9000	0.8000	0.7500
$\bar{s}_{ghn} \gamma = 10$	0.9400	0.9000	0.8000	0.7500
\bar{s}_{hnhh}	0.9400	0.9000	0.8000	0.7500
\bar{s}_{hneh}	0.9400	0.9000	0.8000	0.7500
\bar{s}_{hnhh}	0.9400	0.9000	0.8000	0.7500
\bar{s}_{hnhne}	0.9400	0.9000	0.8000	0.7500
\bar{S} (proposed measure)	0.9434	0.9091	0.8333	0.8000

Table 2 Calculated values at different pairs

Cases	1	2	3
	$B = (y, 0.20, 0.30)$	$B = (y, 0.32, 0.35)$	$B = (y, 0.50, 0.52)$
	$F = (y, 0.40, 0.50)$	$F = (y, 0.53, 0.54)$	$F = (y, 0.67, 0.75)$
\bar{s}_{man}	0.2000	0.1900	0.2000
\bar{s}_{hnh}	0.8000	0.8000	0.8000
\bar{s}_{hne}	0.7998	0.7998	0.7978
$\bar{s}_{ghn} \gamma = 6$	0.8000	0.7988	0.7899
$\bar{s}_{ghn} \gamma = 10$	0.8000	0.7978	0.7844
\bar{s}_{hnhh}	0.8000	0.8100	0.7700
\bar{s}_{hneh}	0.8000	0.8100	0.7700
\bar{s}_{hhnh}	0.8000	0.8050	0.7850
\bar{s}_{hhne}	0.8000	0.8048	0.7834
\bar{s}_{hhd}	0.8667	0.8667	0.8667
\bar{s}_{ehd}	0.8367	0.8367	0.8349
$\bar{s}_{ghd} \gamma = 6$	0.8131	0.8119	0.8036
$\bar{s}_{ghd} \gamma = 10$	0.8079	0.8079	0.7929
\bar{s}_{nhnh}	0.9000	0.8950	0.8850
\bar{s}_{nhne}	0.8580	0.8580	0.8539
$\bar{s}_{ngnh} \gamma = 6$	0.8218	0.8218	0.8128
$\bar{s}_{ngnh} \gamma = 10$	0.8134	0.8114	0.7988
\bar{S} (proposed measure)	0.8333	0.8336	0.8339

Table 3 Calculated values at different pairs

Cases	1	2	3
	$B = (y, 0.41, 0.46, 0.51)$	$B = (y, 0.59, 0.67, 0.69)$	$B = (y, 0.06, 0.02, 0.03)$
	$F = (y, 0.45, 0.49, 0.55)$	$F = (y, 0.55, 0.69, 0.74)$	$F = (y, 0.29, 0.31, 0.59)$
\bar{s}_{man}	0.2000	0.1900	0.2000
\bar{s}_{hnh}	0.9633	0.9633	0.9633
\bar{s}_{hne}	0.9613	0.9613	0.9596
$\bar{s}_{ghn} \gamma = 6$	0.9621	0.9567	0.9499
\bar{s}_{hnhh}	0.9600	0.9500	0.9700
\bar{s}_{hneh}	0.9600	0.9500	0.9700
\bar{s}_{hhnh}	0.9617	0.9567	0.9667
\bar{s}_{hhne}	0.9615	0.9553	0.9644
\bar{s}_{hhd}	0.9725	0.9725	0.9725
\bar{s}_{ehd}	0.9680	0.9680	0.9650
$\bar{s}_{ghd} \gamma = 6$	0.9639	0.9587	0.9522
\bar{s}_{nhnh}	0.9780	0.9600	0.9780
\bar{s}_{nhne}	0.9714	0.9498	0.9659
$\bar{s}_{ngnh} \gamma = 6$	0.9652	0.9308	0.9539
$\bar{s}_{ngnh} \gamma = 10$	0.9634	0.9234	0.9489
\bar{S} (proposed measure)	0.9645	0.9649	0.9653

the three cases. \bar{s}_{hnhh} and \bar{s}_{hneh} both have obtained the similar value with respect to each other. \bar{s}_{hne} and \bar{s}_{ehd} both have obtained the similar results on 1st and 2nd case whereas \bar{s}_{nhnh} has similar value on 1st and 3rd case. In the similar fashion \bar{s}_{hhd} has obtained the similar value on all the three cases. Thus, we have observed that some of the existing

measures are not getting accurate and reasonable results but proposed measure is getting consistent and rational result.

Pattern recognition

To classify the unknown pattern from known pattern in pattern-recognition problems, we generally use two main indexes one is distance measure and another is similarity measure. Most of the existing measures have used distance measures but in our case similarity measure has been used which can easily tackle different kind of pattern-recognition and real life problems. The major requirement for any measure is to take proper decision. Therefore, it is necessary for us to formulate the pattern recognition problem in an efficient manner.

Consider B_j as the known patterns and F is the unknown pattern which we have to classify from one of the known pattern ($j = 1, 2, \dots, f$). Therefore, problem can be described as:

$$B_j = \left\{ \left(y_n, h_{B_j}(y_n) \right) \mid y_n \in Y, n = 1, 2, \dots, v \right\}, j = 1, 2, \dots, f.$$

$$F = \left\{ \left(y_n, h_{B_j}(y_n) \right) \mid y_n \in Y, n = 1, 2, \dots, v \right\}.$$

Thus, we will distinguish the unknown pattern from known pattern which has more resemblance with the assistance of identification principle. Thus, we will show the identification principle using subsequent examples and compare the performance of distinct measures.

Example 6 Let us consider $X^* = \{(\alpha, \beta, \gamma) \mid \alpha = \beta = \gamma = 60^\circ\}$ as set of equilateral triangle. B is denoted as fuzzy set in X^* for every triangle and its membership degree reflects the degree to which triangle (fuzzy set). B is associated to equilateral triangle. Let us take an example of triangle having three angles ($65^\circ, 70^\circ, 45^\circ$). Some of the analyst thinks it seems like equilateral triangle, some thinks it is close to equilateral triangle while other thinks, it may not seem like equilateral triangle. Thus, membership value allotted to fuzzy set can be hesitate between different values. Thus, with the help of hesitant fuzzy set, we can represent this triangle.

Let us consider two triangles which are denoted in the form of HFSs

$$B_1 = \{y_1, 0.95, 1.0\} \text{ and } B_2 = \{y_1, 0.45, 0.50, 1.0\}$$

and we have to recognize triangle $F = \{y_1, 0.70, 0.75, 0.80\}$ which is considered as unknown pattern. The different values obtained by existing and proposed measure are as mentioned in Table 4. As we know pattern recognition problems should allocate one of the known pattern according to the identification principle. The classified results which we have calculated using different measures are as compiled in Table 4.

Example 7 Consider a company who wants to hire HR Manager for its company. The assessment of candidates in the form of HFSs are $B_1 = \{y_1, 0.17, 0.20\}$, $B_2 = \{y_1, 0.19, 0.20\}$, and $B_3 = \{y_1, 0.18, 0.19\}$. The standard characteristics set by interviewer which is based on the requirement of job are $F = \{y_1, 0.20, 0.20\}$. Now the

Table 4 Computed results corresponding to Example 1

Similarity measures	(B_1, F)	(B_2, F)	Ordering	Comment
\bar{s}_{man}	0.7833	0.7667	$B_1 > B_2$	Reasonable
\bar{s}_{hnh}	0.7833	0.7667	$B_1 > B_2$	Reasonable
\bar{s}_{hne}	0.7821	0.7655	$B_1 > B_2$	Reasonable
$\bar{s}_{ghn} \gamma = 6$	0.7767	0.7615	$B_1 > B_2$	Reasonable
$\bar{s}_{ghn} \gamma = 10$	0.7716	0.7587	$B_1 > B_2$	Reasonable
\bar{s}_{hnhh}	0.8000	0.8000	$B_1 = B_2$	Unreasonable
\bar{s}_{hneh}	0.8000	0.8000	$B_1 = B_2$	Unreasonable
\bar{s}_{hhnh}	0.7917	0.7833	$B_1 > B_2$	Reasonable
\bar{s}_{hhne}	0.7908	0.7821	$B_1 > B_2$	Reasonable
\bar{s}_{hhd}	0.7958	0.8250	$B_1 < B_2$	Unreasonable
\bar{s}_{ehd}	0.8113	0.7969	$B_1 > B_2$	Reasonable
$\bar{s}_{ghd} \gamma = 6$	0.7851	0.7727	$B_1 > B_2$	Reasonable
$\bar{s}_{ghd} \gamma = 10$	0.7778	0.7655	$B_1 > B_2$	Reasonable
\bar{s}_{nhnh}	0.8267	0.7700	$B_1 > B_2$	Reasonable
\bar{s}_{nhne}	0.8141	0.7289	$B_1 > B_2$	Reasonable
$\bar{s}_{nghn} \gamma = 6$	0.7930	0.6522	$B_1 > B_2$	Reasonable
$\bar{s}_{nghn} \gamma = 10$	0.7827	0.6167	$B_1 > B_2$	Reasonable
\bar{S} (proposed measure)	0.8222	0.8111	$B_1 > B_2$	Reasonable

problem is to find out the best candidates for the vacant position. Thus, according to identification principle we will check the closeness of candidate characteristic with standard characteristic set by interviewer. The Candidate which will have higher similarity measure value, will be the better candidate. For this purpose, we will calculate the values of existing and proposed measures using this example and also discriminate their performances. The different values obtained by existing and proposed measure are as mentioned in Table 5.

Table 5 Computed results corresponding to Example 2

Similarity measures	(B_1, F)	(B_2, F)	(B_3, F)	Ordering	Comment
\bar{s}_{man}	0.9850	0.9950	0.9850	$B_1 = B_3 < B_2$	Unjustified
\bar{s}_{hnh}	0.9850	0.9950	0.9850	$B_1 = B_3 < B_2$	Unjustified
$\bar{s}_{ghn} \gamma = 6$	0.9733	0.9911	0.9821	$B_2 > B_3 > B_1$	Justified
$\bar{s}_{ghn} \gamma = 10$	0.9720	0.9907	0.9813	$B_2 > B_3 > B_1$	Justified
\bar{s}_{hnhh}	1.0000	1.0000	0.9900	$B_1 = B_2 > B_3$	Unjustified
\bar{s}_{hneh}	1.0000	1.0000	0.9900	$B_1 = B_2 > B_3$	Unjustified
\bar{s}_{hhnh}	0.9925	0.9975	0.9975	$B_2 = B_3 > B_1$	Unjustified
\bar{s}_{hhne}	0.9850	0.9850	0.9868	$B_1 = B_2 < B_3$	Unjustified
\bar{s}_{hhd}	0.8233	0.8300	0.8233	$B_1 = B_3 < B_2$	Unjustified
\bar{s}_{ehd}	0.7108	0.7113	0.7108	$B_1 = B_3 < B_2$	Unjustified
$\bar{s}_{ghd} \gamma = 6$	0.5837	0.5387	0.5837	$B_1 = B_2 = B_3$	Indistinguished
$\bar{s}_{ghd} \gamma = 10$	0.5520	0.5520	0.5520	$B_1 = B_2 = B_3$	Indistinguished
\bar{s}_{nhnh}	0.8600	0.8700	0.8650	$B_2 > B_3 > B_1$	Justified
\bar{s}_{nhne}	0.7497	0.7499	0.7497	$B_1 = B_3 < B_2$	Unjustified
$\bar{s}_{nghn} \gamma = 6$	0.6031	0.6031	0.6031	$B_1 = B_2 = B_3$	Indistinguished
$\bar{s}_{nghn} \gamma = 10$	0.5647	0.5647	0.5647	$B_1 = B_2 = B_3$	Indistinguished
\bar{S} (proposed measure)	0.9854	0.9950	0.9852	$B_2 > B_1 > B_3$	Justified

Example 8 Now we will consider an example of Asian rice (*Oryza Sativa*). This can be further divided into three types—Javonica, Japonica, and Indica. Total 21 sample has to be taken into consideration means 7 samples of each type. Different parameter/attributes for each are considered as milling degree (MD), foreign matter (FM), whiteness (WT), moisture content (MC), and grain shape (GS).

After that similarity between Javonica with Japonica and Javonica with Indica has to be discussed on different existing measures and proposed measure for contrast purposes. Further, we will calculate another term i.e Degree of Confidence (DOC) to explore the behavior of different measures. Initially, Hatzimichailidis et al. [60] had given the notion of DOC which can be described as:

$$DoC(X) = \left| X(\text{Javonica, Japonica}) \right| + \left| X(\text{Javonica, Indica}) - Y \right|, \quad (24)$$

where X is any HF-compatibility measure.

For HF similarity/correlation measure X ; $Y = \text{Max}\{X(\text{Javonica, Japonica}), X(\text{Javonica, Indica})\}$ and For HF dissimilarity/distance measure X ; $Y = \text{Min}\{X(\text{Javonica, Japonica}), X(\text{Javonica, Indica})\}$. We have taken the random data as the different parameters of Asian rice. But this data cannot exactly used on different measures for calculation purposes. Firstly, we have to modify this data into HF-domain.

Therefore, we established transformation formulas as:

$$\begin{aligned} \hat{mf}_1(y_{ij}) &= \frac{1}{(1 + A^*)^2}, \text{ where } A^* \text{ is the Attributes Value} \\ \hat{mf}_2(y_{ij}) &= \frac{1}{(1 + A^*)}, \\ \hat{mf}_3(y_{ij}) &= \frac{1}{(1 + A^*)^{\frac{1}{2}}} \end{aligned}$$

Where $\hat{mf}_1(y_{ij})$, $\hat{mf}_2(y_{ij})$ and $\hat{mf}_3(y_{ij})$ = membership functions of the element y_{ij} .

The computed values corresponding to Example 3 are as mentioned in Table 6 which outlined the similarity between Javonica with Japonica and Javonica with Indica along with their DOC.

Results and discussion

After observing Table 4 it can be easily reveals that \bar{s}_{hnhh} and \bar{s}_{hneh} are not classifying any of the known pattern because (B_1, F) and (B_2, F) both are getting same values, thus getting unclassified result. Although \bar{s}_{man} and \bar{s}_{hnh} are classifying B_1 as known pattern but both are attaining same value on both cases whereas all the other measures including proposed measure are classifying B_1 as the known pattern means F is classified as the class of B_1 , i.e. B_1 has particular shape. Thus, proposed measure is getting consistent result with existing measures.

Table 5 reveals that \bar{s}_{man} and \bar{s}_{hnh} are getting same value on (B_1, F) and (B_3, F) . Thus, they are not able to distinguish between candidates. \bar{s}_{hnhn} and \bar{s}_{hneh} are also attaining

Table 6 Different values corresponding to Example 3

Similarity measures	(Javonica, Japonica)	(Javonica, Indica)	DOC
\bar{s}_{man}	0.9986	0.9983	0.0003
\bar{s}_{hnh}	0.9986	0.9983	0.0003
\bar{s}_{hne}	0.9979	0.9973	0.0006
$\bar{s}_{ghn} \gamma = 6$	0.9964	0.9955	0.0009
$\bar{s}_{ghn} \gamma = 10$	0.9958	0.9948	0.0010
\bar{s}_{hnhh}	0.9987	0.9973	0.0014
\bar{s}_{hneh}	0.9992	0.9983	0.0009
\bar{s}_{hhnh}	0.9987	0.9978	0.0009
\bar{s}_{hhne}	0.9980	0.9967	0.0013
\bar{s}_{hhd}	0.9990	0.9987	0.0003
\bar{s}_{ehd}	0.9982	0.9977	0.0005
$\bar{s}_{ghd} \gamma = 6$	0.9966	0.9957	0.0009
$\bar{s}_{ghd} \gamma = 10$	0.9959	0.9949	0.0010
\bar{s}_{nhnh}	0.9991	0.9978	0.0013
\bar{s}_{nhne}	0.9983	0.9967	0.0016
$\bar{s}_{nghn} \gamma = 6$	0.9967	0.9950	0.0017
$\bar{s}_{nghn} \gamma = 10$	0.9960	0.9944	0.0016
\bar{S} (proposed measure)	0.9989	0.9970	0.0019

unclassified results. In the similar fashion most of the existing measures are getting unclassified results means they are not distinguished the assessment of candidates. But proposed measure (\bar{S}) is capable to distinguish among candidates and getting B_2 as one of the best candidate because the similarity measure value of B_2 is larger than B_1 and B_3 . This shows that proposed measure is getting consistent and better results as compared to others.

We observe that Javonica has more similar with Japonica than Javonica with Indica (please refer Table 6). (Javonica, Japonica) has attained more value than (Javonica, Indica), this can be outlined with Fig. 1. With the help of Table 6 it can also perceive that proposed measure has maximum DOC than existing measures (please see Fig. 2). From all these consideration we concluded that proposed measure is more reliable and efficient than others.

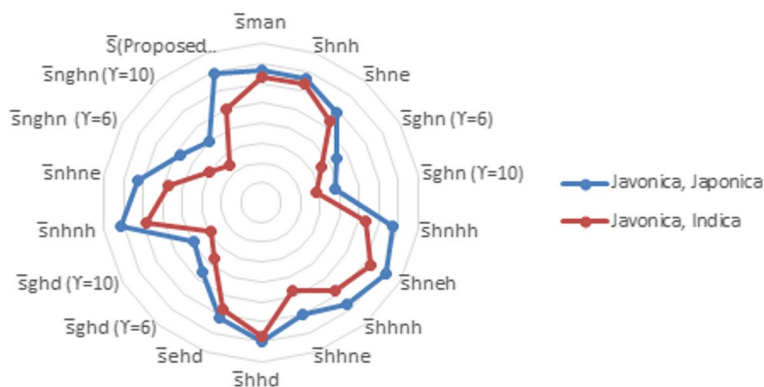


Fig. 1 Different measures

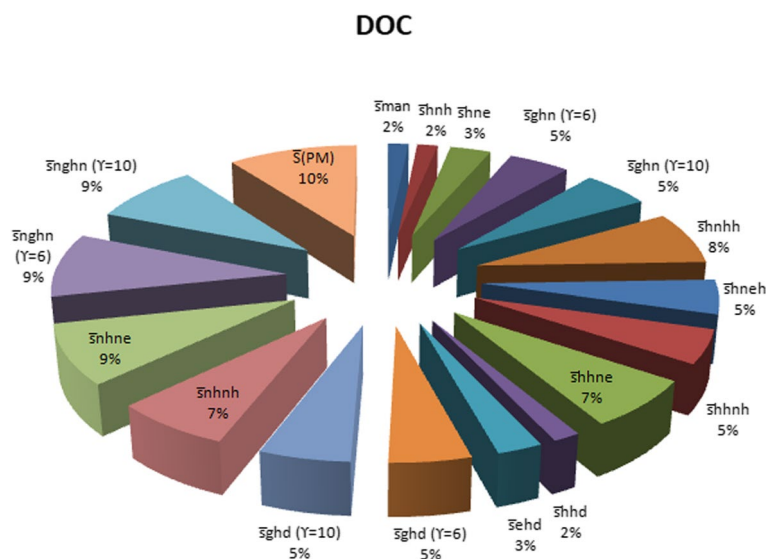


Fig. 2 Degree of Confidence (DOC)

Clustering analysis

Clustering is one of significant modeling technique. Some of the researchers explored the concept and elongated into IFSs and HFSs. Wang [31] and Yao et al. [61] explored the concept of clustering using FS and employed in decision-making, production predict and assessment process. Association coefficient based IFSs clustering algorithm was suggested by Xu et al. [62] and expanded it for Interval valued FS. The correlation coefficient of HFSs was examined by Chen et al. [15] and implemented for clustering. Farhadinia [51] suggested similarity measure using HFSs and applied in clustering process. A clustering algorithm was investigated by Wen et al. [63] using HF Lukasiewicz implication operator. Yang and Hussian [64] suggested their measures using HFSs based on Hausdorff metric and applied in the field of MCDM and clustering. Some new similarity and distance measures was suggested by Zhang and Xu [41, 65] using HFSs in the application of clustering. Zhang and Xu presented an Hierarchical clustering algorithm using hesitant fuzzy. Further, we will suggest clustering algorithm for HF-environment which is based on MST.

Algorithm: The steps of MST based clustering algorithm using HFSs are:

1. Firstly Consider HFSs (B_1, B_2, \dots, B_i) in B. Next, we have to calculate the values of proposed similarity measure using these HFSs and construct hesitant fuzzy similar matrix.
2. Draw HF-graph by interconnecting each node with the help of edge. Assign value to each edge from HF similar matrix.
3. Next, maximum spanning tree (MST) is to be formed using Kruskal [66] or Prim method [67].
 - (i) An edge is connected between two nodes and each edge has its weight. First of all we have to arrange the weight in decreasing order.
 - (ii) Then, Pick an edge having highest value.

Table 7 HF-Ship Data

	\bar{c}_1	\bar{c}_2	\bar{c}_3	\bar{c}_4
h_{B_1}	{0.6,0.7,0.75}	{0.25,0.35,0.45}	{0.5,0.8}	{0.6,0.75,1.0}
h_{B_2}	{0.7,0.85}	{0.45,0.55,0.7}	{0.3,0.7}	{0.9,1.0}
h_{B_3}	{0.55,0.65,0.8}	{0.8,1.0}	{0.3,0.5}	{0.6,0.75,1.0}
h_{B_4}	{0.4,0.9}	{0.2,0.25,0.7}	{0.2,0.7}	{0.5,0.6,0.9}
h_{B_5}	{0.7,0.8}	{0.15,0.4,0.55}	{0.8,1.0}	{0.1,0.3,0.5}
h_{B_6}	{0.45,0.55,0.7}	{0.55,0.6}	{0.2,0.8}	{0.2,0.3,0.8}
h_{B_7}	{0.35,0.5,0.6}	{0.1,0.25,0.6}	{0.9,1.0}	{0.35,0.4}
h_{B_8}	{0.6,0.9}	{0.4,0.5}	{0.55,0.65}	{0.6,0.7}
h_{B_9}	{0.2,0.25,0.7}	{0.5,0.55,0.7}	{0.35,0.45}	{0.65,0.7,0.85}
$h_{B_{10}}$	{0.15,0.35,0.6}	{0.8,0.9}	{0.7,0.8}	{0.7,1.0}

(iii) After that, pick the another edge having next larger value among the remaining one and it should not formed the closed circuit with already selected edges.

(iv) Repeat the previous steps until (n-1) edges have been picked. In this way MST graph is formed.

4. Make the distinct clusters using grouping of nodes by selecting threshold (Ψ).

Example 9 Consider the data of ten Ships with its four attributes in the form of HFSs. These attributes can be characterized as Cost value (\bar{c}_1), Speed (\bar{c}_2), Design (\bar{c}_3), and Fuel economy (\bar{c}_4) which is shown in Table 7.

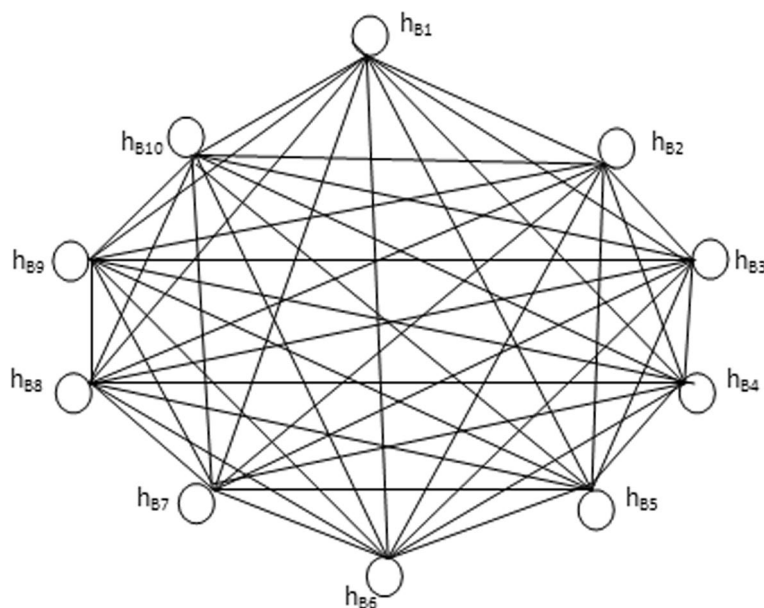


Fig. 3 Hesitant fuzzy graph

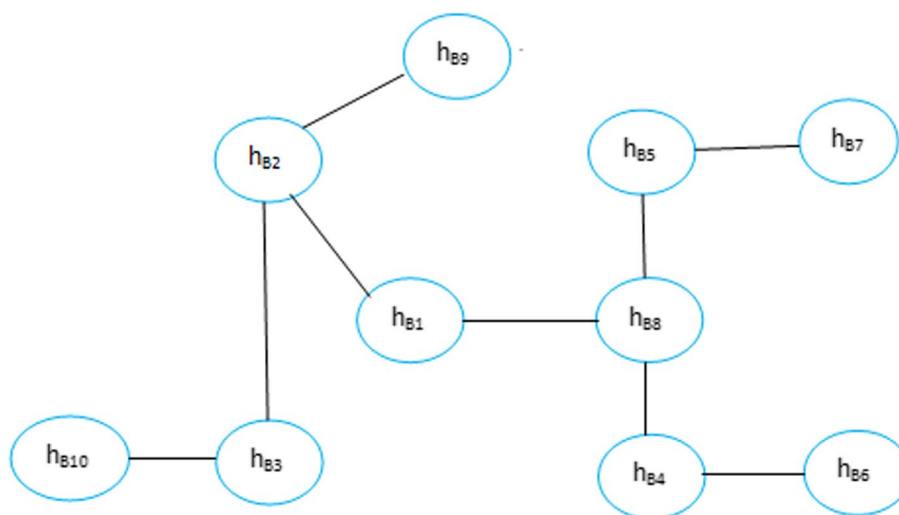


Fig. 4 MST Graph

1. HF-similar matrix using above data is as follows:

1	0.8750	0.8561	0.8557	0.8356	0.8283	0.7960	0.9138	0.8370	0.8210
0.8750	1	0.8797	0.8559	0.7842	0.8443	0.7273	0.8572	0.8655	0.7970
0.8561	0.8797	1	0.8298	0.7379	0.8222	0.7117	0.8348	0.8612	0.8571
0.8557	0.8559	0.8298	1	0.7797	0.8692	0.8109	0.8649	0.8621	0.7933
0.8356	0.7842	0.7379	0.7797	1	0.8230	0.8884	0.8408	0.7392	0.7448
0.8283	0.8443	0.8222	0.8692	0.8230	1	0.8146	0.8373	0.8556	0.7857
0.7960	0.7273	0.7117	0.8109	0.8884	0.8146	1	0.8015	0.7589	0.7760
0.9138	0.8572	0.8348	0.8649	0.8408	0.8373	0.8015	1	0.8443	0.7998
0.8370	0.8655	0.8612	0.8621	0.7392	0.8556	0.7589	0.8443	1	0.8514
0.8210	0.7970	0.8571	0.7933	0.7448	0.7857	0.7760	0.7998	0.8514	1

2. Hesitant fuzzy graph $\hat{Z} = (V, E)$ by interconnecting the nodes is as shown in Fig. 3.

3. Next, we have formed the MST of Hesitant Fuzzy graph using all the following steps.

(i) Arrange the edges according to their values.

$$\begin{aligned}
 e_{1,8} &> e_{5,7} > e_{2,3} > e_{1,2} > e_{4,6} > e_{2,9} > e_{4,8} > e_{4,9} > e_{3,9} > e_{2,8} > e_{3,10} > e_{1,3} > e_{2,4} > e_{1,4} > e_{6,9} > e_{9,10} \\
 &> e_{2,6} = e_{8,9} > e_{5,8} > e_{6,8} > e_{1,9} > e_{1,5} > e_{3,8} > e_{3,4} > e_{1,6} > e_{5,6} > e_{3,6} > e_{1,10} > e_{6,7} > e_{4,7} \\
 &> e_{7,8} > e_{8,10} > e_{2,10} > e_{1,7} > e_{4,10} > e_{6,10} > e_{2,5} > e_{4,5} > e_{7,10} > e_{7,9} > e_{5,10} > e_{5,9} > e_{3,5} \\
 &> e_{2,7} > e_{3,7}
 \end{aligned}$$

(ii) $e_{1,8}$ is the edge having max. weight as it is connected between node 1 and 8.

(iii) Next selected edge is $e_{5,7}$ which is having next higher value $e_{1,8}$ and it is also not forming the closed circuit with $e_{1,8}$.

(iv) In this way, we repeated the above steps until 9 edges has been selected and attain the MST graph (see Fig. 4).

4. Lastly, threshold (ψ) has been selected to form the clusters into groups which are as mentioned in Table 8.

Comparison of clustering results: According to above algorithm, the comparative results are shown in Tables 8, 9, and 10.

Table 8 Proposed measure (\bar{S})

Groups	Threshold (ψ)	Clusters
1	0.8408	$\{h_{B_1}, h_{B_2}, h_{B_3}, h_{B_4}, h_{B_5}, h_{B_6}, h_{B_7}, h_{B_8}, h_{B_9}, h_{B_{10}}\}$
2	0.8571	$\{h_{B_1}, h_{B_2}, h_{B_3}, h_{B_4}, h_{B_6}, h_{B_8}, h_{B_9}, h_{B_{10}}\} \{h_{B_5}, h_{B_7}\}$
3	0.8649	$\{h_{B_1}, h_{B_2}, h_{B_3}, h_{B_4}, h_{B_6}, h_{B_8}, h_{B_9}\} \{h_{B_5}\} \{h_{B_{10}}\}$
4	0.8655	$\{h_{B_1}, h_{B_2}, h_{B_3}, h_{B_8}, h_{B_9}\} \{h_{B_4}, h_{B_6}\} \{h_{B_5}\} \{h_{B_{10}}\}$
5	0.8692	$\{h_{B_1}, h_{B_2}, h_{B_3}, h_{B_8}\} \{h_{B_5}, h_{B_7}\} \{h_{B_4}, h_{B_6}\} \{h_{B_{10}}\} \{h_{B_9}\}$
6	0.8750	$\{h_{B_1}, h_{B_2}, h_{B_3}, h_{B_8}\} \{h_{B_5}, h_{B_7}\} \{h_{B_4}\} \{h_{B_6}\} \{h_{B_{10}}\} \{h_{B_9}\}$
7	0.8797	$\{h_{B_1}, h_{B_8}\} \{h_{B_2}, h_{B_3}\} \{h_{B_5}, h_{B_7}\} \{h_{B_4}\} \{h_{B_6}\} \{h_{B_{10}}\} \{h_{B_9}\}$
8	0.8884	$\{h_{B_1}, h_{B_8}\} \{h_{B_2}\} \{h_{B_3}\} \{h_{B_5}, h_{B_7}\} \{h_{B_4}\} \{h_{B_6}\} \{h_{B_{10}}\} \{h_{B_9}\}$
9	0.9138	$\{h_{B_1}, h_{B_8}\} \{h_{B_2}\} \{h_{B_3}\} \{h_{B_5}\} \{h_{B_7}\} \{h_{B_4}\} \{h_{B_6}\} \{h_{B_{10}}\} \{h_{B_9}\}$
10	1	$\{h_{B_1}\} \{h_{B_2}\} \{h_{B_3}\} \{h_{B_4}\} \{h_{B_5}\} \{h_{B_6}\} \{h_{B_7}\} \{h_{B_8}\} \{h_{B_9}\} \{h_{B_{10}}\}$

Table 9 Existing Measure (\bar{S}_{hnh}) Case 1:

Groups	Threshold (ψ)	Clusters
1	0.7958	$\{h_{B_1}, h_{B_2}, h_{B_3}, h_{B_4}, h_{B_5}, h_{B_6}, h_{B_7}, h_{B_8}, h_{B_9}, h_{B_{10}}\}$
2	0.8000	$\{h_{B_1}, h_{B_2}, h_{B_3}, h_{B_4}, h_{B_6}, h_{B_8}, h_{B_9}, h_{B_{10}}\} \{h_{B_5}, h_{B_7}\}$
3	0.8125	$\{h_{B_1}, h_{B_2}, h_{B_4}, h_{B_6}, h_{B_8}, h_{B_9}\} \{h_{B_5}, h_{B_7}\} \{h_{B_3}\} \{h_{B_{10}}\}$
4	0.8250	$\{h_{B_1}, h_{B_2}, h_{B_4}, h_{B_6}, h_{B_8}, h_{B_9}\} \{h_{B_5}, h_{B_7}\} \{h_{B_3}\} \{h_{B_{10}}\}$
5	0.8333	$\{h_{B_1}, h_{B_2}, h_{B_4}, h_{B_6}, h_{B_8}\} \{h_{B_5}, h_{B_7}\} \{h_{B_3}\} \{h_{B_{10}}\} \{h_{B_9}\}$
6	0.8375	$\{h_{B_1}, h_{B_2}, h_{B_8}\} \{h_{B_5}, h_{B_7}\} \{h_{B_4}, h_{B_6}\} \{h_{B_{10}}\} \{h_{B_3}\} \{h_{B_9}\}$
7	0.8500	$\{h_{B_1}, h_{B_2}, h_{B_8}\} \{h_{B_5}, h_{B_7}\} \{h_{B_4}\} \{h_{B_6}\} \{h_{B_{10}}\} \{h_{B_3}\} \{h_{B_9}\}$
8	0.8667	$\{h_{B_1}, h_{B_8}\} \{h_{B_2}, h_{B_3}\} \{h_{B_5}, h_{B_7}\} \{h_{B_4}\} \{h_{B_6}\} \{h_{B_{10}}\} \{h_{B_9}\} \{h_{B_9}\}$
9	0.9000	$\{h_{B_1}, h_{B_8}\} \{h_{B_2}\} \{h_{B_3}\} \{h_{B_5}, h_{B_7}\} \{h_{B_4}\} \{h_{B_6}\} \{h_{B_{10}}\} \{h_{B_9}\} \{h_{B_9}\}$
10	1	$\{h_{B_1}\} \{h_{B_2}\} \{h_{B_3}\} \{h_{B_4}\} \{h_{B_5}\} \{h_{B_6}\} \{h_{B_7}\} \{h_{B_8}\} \{h_{B_9}\} \{h_{B_{10}}\}$

Table 10 Existing Measure (\bar{S}_{hnh}) Case 2:

Groups	Threshold (ψ)	Clusters
1	0.7958	$\{h_{B_1}, h_{B_2}, h_{B_3}, h_{B_4}, h_{B_5}, h_{B_6}, h_{B_7}, h_{B_8}, h_{B_9}, h_{B_{10}}\}$
2	0.8125	$\{h_{B_1}, h_{B_2}, h_{B_3}, h_{B_4}, h_{B_6}, h_{B_8}, h_{B_9}, h_{B_{10}}\} \{h_{B_5}, h_{B_7}\}$
3	0.8167	$\{h_{B_1}, h_{B_2}, h_{B_3}, h_{B_4}, h_{B_6}, h_{B_8}, h_{B_9}\} \{h_{B_5}, h_{B_7}\} \{h_{B_{10}}\}$
4	0.8250	$\{h_{B_1}, h_{B_4}, h_{B_6}, h_{B_8}, h_{B_9}\} \{h_{B_5}, h_{B_7}\} \{h_{B_2}\} \{h_{B_3}\} \{h_{B_{10}}\}$
5	0.8333	$\{h_{B_4}, h_{B_6}, h_{B_8}\} \{h_{B_5}, h_{B_7}\} \{h_{B_1}, h_{B_9}\} \{h_{B_2}\} \{h_{B_3}\} \{h_{B_{10}}\}$
6	0.8375	$\{h_{B_4}, h_{B_8}\} \{h_{B_5}, h_{B_7}\} \{h_{B_1}, h_{B_9}\} \{h_{B_2}\} \{h_{B_3}\} \{h_{B_{10}}\}$
7	0.8500	$\{h_{B_4}\} \{h_{B_8}\} \{h_{B_5}, h_{B_7}\} \{h_{B_1}, h_{B_9}\} \{h_{B_2}\} \{h_{B_3}\} \{h_{B_{10}}\}$
8	0.8667	$\{h_{B_4}\} \{h_{B_8}\} \{h_{B_5}\} \{h_{B_7}\} \{h_{B_1}, h_{B_9}\} \{h_{B_2}\} \{h_{B_3}\} \{h_{B_{10}}\}$
9	0.9000	$\{h_{B_4}\} \{h_{B_8}\} \{h_{B_5}\} \{h_{B_7}\} \{h_{B_1}, h_{B_9}\} \{h_{B_2}\} \{h_{B_3}\} \{h_{B_{10}}\}$
10	1	$\{h_{B_1}\} \{h_{B_2}\} \{h_{B_3}\} \{h_{B_4}\} \{h_{B_5}\} \{h_{B_6}\} \{h_{B_7}\} \{h_{B_8}\} \{h_{B_9}\} \{h_{B_{10}}\}$

Comparative analysis

Now, we will compare the MST based Clustering with hierarchical hesitant fuzzy k-means clustering algorithm to show the ability of proposed method.

Example 10 Suppose a company want to hire a manager. The different attribute on which experts evaluate are communication, leadership, experience and confidence and respective HF information is as shown in Table 11:

Table 11 Hesitant Fuzzy Information

	\bar{c}_1	\bar{c}_2	\bar{c}_3	\bar{c}_4
h_{B_1}	{0.6,0.7,0.75}	{0.25,0.35,0.45}	{0.5,0.8}	{0.6,0.75,1.0}
h_{B_2}	{0.7,0.85}	{0.45,0.55,0.7}	{0.3,0.7}	{0.9,1.0}
h_{B_3}	{0.55,0.65,0.8}	{0.8,1.0}	{0.3,0.5}	{0.6,0.75,1.0}
h_{B_4}	{0.4,0.9}	{0.2,0.25,0.7}	{0.2,0.7}	{0.5,0.6,0.9}
h_{B_5}	{0.7,0.8}	{0.15,0.4,0.55}	{0.8,1.0}	{0.1,0.3,0.5}

Hierarchical K-means clustering method:-

First of all we will consider each HFS as an independent cluster $\{h_{B_1}\}, \{h_{B_2}\}, \{h_{B_3}\}, \{h_{B_4}\}, \{h_{B_5}\}$. After this, we will calculate the distance between different HFS with the help of Eq. (20).

$$\begin{aligned}
 d(h_{B_1}, h_{B_2}) &= 0.125 & d(h_{B_1}, h_{B_2}) &= 0.1439 \\
 d(h_{B_1}, h_{B_4}) &= 0.1443 & d(h_{B_1}, h_{B_5}) &= 0.1644 \\
 d(h_{B_2}, h_{B_3}) &= 0.1203 & d(h_{B_2}, h_{B_4}) &= 0.1441 \\
 d(h_{B_2}, h_{B_5}) &= 0.2158 & d(h_{B_3}, h_{B_4}) &= 0.1702 \\
 d(h_{B_3}, h_{B_5}) &= 0.2621 & d(h_{B_4}, h_{B_5}) &= 0.2203
 \end{aligned}$$

The distance between clusters $d(h_{B_2}, h_{B_3})$ is minimum so we have to merge these clusters. This results that hesitant fuzzy sets are divided into four clusters $\{h_{B_1}\}, \{h_{B_2}, h_{B_3}\}, \{h_{B_4}\}, \{h_{B_5}\}$. After taking the average of h_{B_2} and h_{B_3} , we will further calculate distance between each cluster

$$\begin{aligned}
 d(\{h_{B_2}, h_{B_3}\}, h_{B_1}) &= 0.135 & d(\{h_{B_2}, h_{B_3}\}, h_{B_4}) &= 0.1619 \\
 d(\{h_{B_2}, h_{B_3}\}, h_{B_5}) &= 0.2370 & d(h_{B_1}, h_{B_4}) &= 0.1443 \\
 d(h_{B_1}, h_{B_5}) &= 0.1644
 \end{aligned}$$

We get $d(\{h_{B_2}, h_{B_3}\}, h_{B_1})$ is the shortest distance. Thus, we get the clusters as $\{h_{B_1}, h_{B_2}, h_{B_3}\}, \{h_{B_4}\}, \{h_{B_5}\}$. Further we merge h_{B_1}, h_{B_2} and h_{B_3} and calculate the distance between each cluster. Then we have $d(\{h_{B_1}, h_{B_2}, h_{B_3}\}, h_{B_4})$ as the minimum distance. Thus, new clusters are converted into single cluster $\{h_{B_1}, h_{B_2}, h_{B_3}, h_{B_4}, h_{B_5}\}$.

MST Based method:-After applying the different steps of MST based clustering algorithm using Example 5, we get the different clusters which are tabulated in Table 12

Results and discussion

Although lot of researchers have used different technique to find the clusters like association coefficient based clustering, Hausdorff metric and hierarchical K-means clustering. But we have implemented MST based clustering algorithm and different clusters obtained

Table 12 Comparative Result

Groups	Hierarchical K-means clustering	Proposed MST Based Clustering
1	$\{h_{B_1}, h_{B_2}, h_{B_3}, h_{B_4}, h_{B_5}\}$	$\{h_{B_1}, h_{B_2}, h_{B_3}, h_{B_4}, h_{B_5}\}$
2	$\{h_{B_1}, h_{B_2}, h_{B_3}, h_{B_4}\}, \{h_{B_5}\}$	$\{h_{B_1}, h_{B_2}, h_{B_3}, h_{B_4}\}, \{h_{B_5}\}$
3	$\{h_{B_1}, h_{B_2}, h_{B_3}\}, \{h_{B_4}\}, \{h_{B_5}\}$	$\{h_{B_1}, h_{B_2}, h_{B_3}\}, \{h_{B_4}\}, \{h_{B_5}\}$
4	$\{h_{B_1}\}, \{h_{B_2}, h_{B_3}\}, \{h_{B_4}\}, \{h_{B_5}\}$	$\{h_{B_1}\}, \{h_{B_2}, h_{B_3}\}, \{h_{B_4}\}, \{h_{B_5}\}$
5	$\{h_{B_1}\}, \{h_{B_2}\}, \{h_{B_3}\}, \{h_{B_4}\}, \{h_{B_5}\}$	$\{h_{B_1}\}, \{h_{B_2}\}, \{h_{B_3}\}, \{h_{B_4}\}, \{h_{B_5}\}$

are tabulated in Table 8. We have also find the clusters for existing measure (\bar{s}_{hnh}) using the same example and clustering obtained are as mentioned in Tables 9 and 10. Tables 9 and 10 shows that clusters are not fixed for existing measure (\bar{s}_{hnh}) as two different clustering results are formed. Therefore, it is difficult for users to select the right clustering results. Thus, we concluded that existing measures are not reliable and efficient whereas proposed measure is getting unique, better and efficient results using HF-MST algorithm technique.

To show the ability of proposed measure further, we have compared hierarchical k-means clustering with MST based clustering using another example. The computed results are as mentioned in Table 12. Although, same results can be obtained using both methods but in hierarchical K-means clustering we have to calculate clustering center in each step and merge the clusters into single cluster. Thus, in the existing clustering technique lot of computation is required where as proposed clustering technique is simpler and effective as compared to other techniques.

Conclusion and future scope

Similarity and distance measures are two important implements for solving clustering, pattern-recognition, medical diagnosis problems and so on. Although, various authors have suggested their measures but they are mainly distance measures while some of them have extracted similarity measures from distance measures and can be applied for different applications. But it is observed that some of them are not getting appropriate results means they are getting unreasonable or counter intuitive results. This motivate us to develop new similarity measure which can tackle with all these problem and satisfying different properties for HFSs. Further, numerical experiment and pattern-recognition problems are taken into consideration. In numerical experiment we considered cases using different length of HFSs to explore the performance of different measures which exhibits that proposed measure is attaining consistent and rational results. To verify the proposed measure validity for pattern recognition problems we have considered the different examples. In first two examples classification of unknown pattern from one of the known patterns has been carried out and in third example we calculated another term i.e. DOC (Degree of Confidence) by taking an example of Asian Rice. Furthermore, clustering algorithm using maximum spanning tree (MST) has been suggested for HF-environment and comparison has been performed with existing measures. From all these comparative study we determined that proposed measure is reliable and getting superior outcomes than others. This work expanded to IV-HFS (interval valued HFS), HIFS (Hesitant intuitionistic Fuzzy set), Hesitant picture fuzzy set (HPFS), Hesitant spherical fuzzy set, Hesitant q-rung orthopair fuzzy set for future scope of improvement. In future, we can extend our work using Adaptive K-means and hybrid clustering algorithm.

Abbreviations

HFSs	Hesitant fuzzy sets
MST	Maximum spanning tree
DOC	Degree of Confidence
MCDM	Multi criteria decision making
HFE	Hesitant fuzzy element
FS	Fuzzy set

Acknowledgements

Not applicable.

Authors' contributions

Both the authors contributed to the design and implementation of the research, to the analysis of the results and to the writing of the manuscript. All authors read and approved the final manuscript.

Funding

No funds, grants or other support was received.

Availability of data and materials

All data generated or analysed during this study are included in this article.

Declarations**Competing interests**

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Received: 17 January 2023 Accepted: 2 December 2023

Published online: 03 January 2024

References

- Zadeh LA (1965) Fuzzy sets. *Inf Control* 8:338–356
- Yager RR (1979) On measures of fuzziness and negation, part I: membership in the unit interval. *Int J Gen Syst* 5:221–229
- Atanassov KT (1986) Intuitionistic fuzzy sets. *Fuzzy Sets Syst* 20:87–96
- Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning. (I), (II), (III). *Inf Sci* 8:199–249, 8:301–357, 9:43–80
- Dubois D, Prade H (1980) *Fuzzy Sets and Systems: Theory and Applications*. Academic Press, New York
- Yager RR (1986) On the theory of bags. *Int J Gen Syst* 13:23–37
- Torra V, Narukawa Y (2009) On hesitant fuzzy sets and decision. The 18th IEEE International Conference on Fuzzy Systems, Jeju Island, pp 1378–1382
- Torra V (2010) Hesitant fuzzy sets. *Int J Intell Syst* 25:529–539
- Gupta R, Kumar S (2021) Intuitionistic fuzzy scale-invariant entropy with correlation coefficients-based VIKOR approach for multi-criteria decision-making. *Granul Comput* 7:77–93. <https://doi.org/10.1007/s41066-020-00252-0>
- Gupta R, Kumar S (2022) Correlation Coefficient Based Extended VIKOR Approach Under Intuitionistic Fuzzy Environment. *Int J Inf Manag Sci* 33(1):35–54
- Wei C, Yan F, Rodriguez RM (2016) Entropy measures for hesitant fuzzy sets and their application relations and fuzzy in multi-criteria decision-making. *J Intell Fuzzy Syst* 31(1):673–685
- Wang YM, Que CP, Lan YX (2017) Hesitant fuzzy TOPSIS multi-attribute decision method based on prospect theory. *Control Decis (Chin)* 32(5):864–870
- Karaaslan F, Karamaz F (2021) Hesitant fuzzy parameterized hesitant fuzzy soft sets and their applications in decision making. *Int J Comput Math*. Taylor and Francis. <https://doi.org/10.1080/00207160.2021.2019715>
- Chen X, Suo C, Li Y, (2021) Distance measures on intuitionistic hesitant fuzzy set and its application in decision-making. *Comput Appl Math* 40(3). <https://doi.org/10.1007/s40314-021-01478-7>
- Chen N, Xu Z, Xia M (2013) Correlation coefficients of hesitant fuzzy sets and their applications to clustering analysis. *Appl Math Model* 37(4):2197–2211
- Singh S, Lalotra S (2019) On generalized correlation coefficients of the hesitant fuzzy sets with their application to clustering analysis. *Comput Appl Math*. <https://doi.org/10.1007/s40314-019-0765-0>
- Liao HC, Xu ZS, Zeng XJ (2015) Novel correlation coefficients between hesitant fuzzy sets and their application in decision making. *Knowl-Based Syst* 82:115–127
- Joshi R, Kumar S (2019) A new approach in multiple attribute decision making using exponential hesitant fuzzy entropy. *Int J Inf Manag Sci* 30:305–322
- Lalotra S, Singh S (2020) Knowledge measure of hesitant fuzzy set and its application in multi-attribute decision-making. *Comput Appl Math* 39(2). <https://doi.org/10.1007/s40314-020-1095-y>
- Suo CF, Li YM, Li ZH (2020) An (R, S)-norm information measure for hesitant fuzzy sets and its application in decision-making. *Comput Appl Math*. <https://doi.org/10.1007/s40314-020-01339-9>
- Singh P (2017) Distance and similarity measures for multiple-attribute decision-making with dual hesitant fuzzy sets. *Comput Appl Math* 36(1):111–126
- Tyagi SK (2015) Correlation coefficient of dual hesitant fuzzy sets and its applications. *Appl Math Model* 39:7082–7092
- Wei GW (2012) Hesitant fuzzy prioritized operators and their application to multiple attribute decision making. *Knowl-Based Syst* 31:176–182
- Yu DJ, Wu YY, Zhou W (2012) Generalized hesitant fuzzy Bonferroni mean and its application in multi-criteria group decision making. *J Inf Comput Sci* 9:267–274
- Rodriguez RM, Martinez L, Herrera F (2012) Hesitant fuzzy linguistic term sets for decision making. *IEEE Trans Fuzzy Syst* 20:109–119
- Xu ZS (2015) *Hesitant Fuzzy Sets Theory*. Springer-Verlag, Berlin
- Xia MM, Xu ZS, Chen N (2013) Some hesitant fuzzy aggregation operators with their application in group decision making. *Group Decis Negot*. 22:259–279

28. Wang L, Shen Q, Zhu L (2015) Dual hesitant fuzzy power aggregation operators based on Archimedean t-conorm and t-norm and their application to multiple attribute group decision making. *Appl Soft Comput* 38:23–50
29. Qin J, Liu X, Pedrycz W (2016) Frank aggregation operators and their application to hesitant fuzzy multiple attribute decision making. *Appl Soft Comput* 41:428–452
30. Tan CQ, Yi WT, Chen XH (2015) Hesitant fuzzy Hamacher aggregation operators for multicriteria decision making. *Appl Soft Comput* 26:325–349
31. Wang PZ (1983) *Fuzzy Sets and Its Applications*. Shanghai Science and Technology Press, Shanghai (in Chinese)
32. Zwick R, Carlstein E, Budescu DV (1987) Measures of similarity among fuzzy concepts: a comparative analysis. *Int J Approx Reason* 1:221–242
33. Zeng WY, Li HX (2006) Inclusion measure, similarity measure and the fuzziness of fuzzy sets and their relations. *Int J Intell Syst* 21:639–653
34. Gupta R, Kumar S (2021) A new similarity measure between picture fuzzy sets with applications to pattern recognition and clustering problems. *Granul Comput*. <https://doi.org/10.1007/s41066-021-00283-1>
35. Gupta R, Kumar S (2022) Intuitionistic Fuzzy Similarity-Based Information Measure in the Application of Pattern Recognition and Clustering. *Int J Fuzzy Syst*. <https://doi.org/10.1007/s40815-022-01272-5>
36. Xu ZS, Xia MM (2011) Distance and similarity measures for hesitant fuzzy sets. *Inf Sci* 181:2128–2138
37. Xu ZS, Xia MM (2011) On distance and correlation measures of hesitant fuzzy information. *Int J Intell Syst* 26:410–425
38. Peng D, Peng B, Wang T (2020) Reconfiguring IVHF-TOPSIS decision making method with parameterized reference solutions and a novel distance for corporate carbon performance evaluation. *J Ambient Intell Human Comput* 11:3811–3832. <https://doi.org/10.1007/s12652-019-01603-9>
39. Xu ZS, Xia MM (2012) Hesitant fuzzy entropy and cross-entropy and their use in multiattribute decision-making. *Int J Intell Syst* 27(9):799–822
40. Peng DH, Gao CY, Gao ZF (2013) Generalized hesitant fuzzy synergetic weighted distance measures and their application to multiple criteria decision-making. *Appl Math Model* 37:5837–5850
41. Zhang X, Xu Z (2015) Novel distance and similarity measures on hesitant fuzzy sets with applications to clustering analysis. *J Intell Fuzzy Syst* 28(5):2279–2296
42. Ahmad A, Khan SS (2019) Survey of state-of-the-art mixed data clustering algorithms. *IEEE Access* 7:31883–31902
43. Dhillon IS, Modha DS (2002) A Data-Clustering Algorithm on Distributed Memory Multiprocessors. In: Zaki MJ, Ho CT (eds) *Large-Scale Parallel Data Mining*, vol 1759. Springer, Berlin/Heidelberg, pp 245–260. *Lecture Notes in Computer Science*
44. Lapegna M, Balzano W, Meyer N, Romano D (2021) Clustering Algorithms on Low-Power and High-Performance Devices for Edge Computing Environments. 21:5395. <https://doi.org/10.3390/s21165395>
45. Laccetti G et al. (2020) A high performance modified K-means algorithm for dynamic data clustering in multi-core CPUs based environments. In: Montella R, et al. (eds) *Internet and Distributed Computing Systems, IDC19*, vol 11874. Springer, Cham. *Lecture Notes in Computer Science*. <https://doi.org/10.1007/978-3-030-34914-19>
46. Laccetti G et al (2020) Performance enhancement of a dynamic K-means algorithm through a parallel adaptive strategy on multi-core CPUs. *J Parallel Distrib Comput* 145:34–41
47. Lapegna M, Stranieri S (2022) A Direction-Based Clustering Algorithm for VANETs Management. In: Barolli L, Yim K, Chen H-C. (eds) *Innovative Mobile and Internet Services in Ubiquitous Computing*. *Lecture Notes in Networks and Systems*. <https://doi.org/10.1007/978-3-030-79728-7>
48. Lapegna M, Mele V, Romano D (2019,2020) An Adaptive Strategy for Dynamic Data Clustering with the K-Means Algorithm. In: Wyrzykowski R, Deelman E, Dongarra J, Karczewski K (eds) *Parallel Processing and Applied Mathematics PPAM*, vol 12044. Springer, Cham, pp 101–110. *Lecture Notes in Computer Science*
49. Laccetti G, Lapegna M, Romano D (2022) A hybrid clustering algorithm for high-performance edge computing devices [Short]. In: 21st International Symposium on Parallel and Distributed Computing (ISPDC), Basel, Switzerland, pp 78–82. <https://doi.org/10.1109/ISPDC5340.2022.00020>
50. Laccetti G, Lapegna M, Montella R (2022) Toward a high-performance clustering algorithm for securing edge computing environments. In: 22nd IEEE International Symposium on Cluster, Cloud and Internet Computing (CCGrid), Taormina, Italy, pp 820–825. <https://doi.org/10.1109/CCGrid54584.2022.00097>
51. Farhadinia B (2013) Information measures for hesitant fuzzy sets and interval valued hesitant fuzzy sets. *Inf Sci* 240:129–144
52. Farhadinia B (2014) Distance and similarity measures for higher order hesitant fuzzy sets. *Knowl-Based Syst* 55:43–48
53. Li DQ, Zeng WY, Zhao YB (2015) Note on distance measure of hesitant fuzzy sets. *Inf Sci* 321:103–115
54. Li DQ, Zeng WY, Li JH (2015) New distance and similarity measures on hesitant Fuzzy sets and their applications in multiple criteria decision making. *Eng Appl Artif Intell* 40:11–16
55. Zeng W, Li D, Yin Q (2016) Distance and similarity measures between hesitant fuzzy sets and their application in pattern recognition. *Pattern Recogn Lett* 84:267–271
56. Liao HC, Xu ZS (2015) Approaches to manage hesitant fuzzy linguistic information based on the cosine distance and similarity measures for HFLTSs and their application in qualitative decision making. *Expert Syst Appl* 42:5328–5336
57. Liao HC, Xu ZS, Xia MM (2013) Multiplicative consistency of hesitant fuzzy preference relation and its application in group decision making. Technical report
58. Xia MM, Xu ZS (2011) Hesitant fuzzy information aggregation in decision making. *Int J Approx Reason* 52:395–407
59. Rezaei K, Rezaei H (2020) New distance and similarity measures for hesitant fuzzy sets and their application. *J Intell Fuzzy Syst* 20. <https://doi.org/10.3233/JIFS-200364>
60. Hatzimichailidis AG, Papakosta GA, Kaburlasos VG (2012) A novel distance measure of intuitionistic fuzzy sets and its application to pattern recognition problems. *Int J Intell Syst*. 27:396–409
61. Yao J, Dash M (2000) Fuzzy clustering and fuzzy modeling. *Fuzzy Sets Syst* 113:381–388
62. Xu Z, Chen J, Wu J (2008) Clustering algorithm for intuitionistic fuzzy sets. *Inf Sci* 178:3775–3790
63. Wen M, Zhao H, Xu Z (2019) Hesitant fuzzy Lukasiewicz implication operation and its application to alternatives' sorting and clustering analysis. *Soft Comput* 23:393–405

64. Yang MS, Hussain Z (2019) Distance and similarity measures of hesitant fuzzy sets based on Hausdorff metric with applications to multi-criteria decision making and clustering. *Soft Comput* 23(14):5835–5848
65. Zhang X, Xu Z (2015) Hesitant fuzzy agglomerative hier-archical clustering algorithms. *Int J Syst Sci* 46(3):562–576
66. Kruskal JB (1956) On the shortest spanning subtree of a graph and the travelling salesman problem. *Proc Am Math Soc* 7:48–50
67. Prim RC (1957) Shortest connection networks and some generalizations. *Bell Syst Tech J* 36:1389–1401

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
