Neutrosophic Sets and Systems, Vol. 97, 2026
University of New Mexico

Agglomerative Hierarchical Clustering Method under Neutrosophic Trapezoidal Fuzzy Numbers

Madineh Farnam [1], Gholam Hassan Shirdel [2], Majid Darehmiraki [3]

[1] Department of Electrical Engineering, Shohadaye Hoveizeh Campus of Technology, Shahid Chamran University of Ahvaz, Dasht-e Azadegan, Khuzestan, Iran; m.farnam@scu.ac.ir
[2] Department of Mathematic, Faculty of Basic Sciences, University of Qom, Qom, Iran; g.h.shirdel@qom.ac.ir
[3] Department of Mathematics, Behbahan Khatam Alanbia University of Technology, Behbahan, Khouzestan, Iran; darehmiraki@bkatu.ac.ir

* Correspondence: g.h.shirdel@qom.ac.ir; Tel.: (optional; include country code)

Abstract: Clustering, as one of the most basic data mining strategies, has a prominent role in various fields of application, especially decision making in management systems. Experts can make and implement decisions according to the features of each cluster by examining the characteristics, nature, and essence of the data that are in the same cluster. Since neutrosophic sets in various application fields have inspired tremendous research endeavors to model problems from varius of aspects, the main goal of this research is to combine hierarchical clustering with neutrosophic trapezoidal fuzzy data. For this purpose, a new distance measure is first introduced to calculate the difference between neutrosophic trapezoidal fuzzy data. In the following, while proving some characteristics of the measure, the hierarchical clustering algorithm based on the new distance measure with neutrosophic trapezoidal fuzzy data is explained. By extending continuous fuzzy data from an explanatory example in fuzzy literature, the effectiveness and efficiency of the proposed algorithm are tested in MATLAB software. Although the resulting dendrogram provides appropriate clustering to the decision maker, two criteria gap, and silhouette, are also used to determine the optimal number of clusters. The hybrid process developed in this research can not only be used in the study areas of clustering but also makes it possible to propose the optimal number of clusters for neutrosophic trapezoidal fuzzy data.

Keywords: clustering; agglomerative hierarchical clustering; neutrosophic set; neutrosophic trapezoidal fuzzy number; distance measure

1. Introduction

Clustering methods are recognized as efficient tools for extracting, analyzing, and identifying hidden patterns from existing data. The concept of clusters in clustering involves grouping together data or objects with the highest similarity (or least distance) within the group and the greatest dissimilarity (or least similarity) with members of other clusters. Therefore, measures of distance and similarity in the data space are crucial for the final clustering outcome. Various methods, including partitional and hierarchical clustering algorithms, are available for deterministic data, many of which are detailed in [1].

While numerous techniques for deterministic data clustering have been proposed theoretically, there has been relatively less research on non-deterministic data in this domain. With the introduction of fuzzy numbers by Zadeh [2] and their extensions, particularly Neutrosophic fuzzy numbers [3], researchers have become interested in their applications in scientific and engineering problems such as decision science, pattern recognition, medicine, networking, and cluster analysis. For example, Ye and Smarandache (2016) [4] developed measures such as Cosine, Jaccard, and Dice for multi-criteria decision-making scenarios using single-valued neutrosophic sets. Khalifa and Kumar (2020) [5] proposed a framework for the neutrosophic assignment problem under interval-valued trapezoidal neutrosophic numbers, transforming it into an interval-valued assignment problem using the score function and considering ordinal relationships based on decision makers' preferences to optimize the interval objective function. Additionally, Zulqarnain et al. [6] advocated for the use of fuzzy TOPSIS as a practical methodology for neutrosophic fuzzy TOPSIS in the context of supplier selection in the manufacturing sector. Haq et al. [7] developed a fuzzy multi-objective framework to solve complex multi-site hybrid supply chain problems under neutrosophic data, focusing on optimizing transportation cost and delivery time simultaneously. Poonia and Bajaj [8] devised an architecture of exponential-function-based similarity measures and weighted types for neutrosophic collections, applying these measures to pattern recognition and decision-making problems to demonstrate their effectiveness and computational manageability. The Kruskal approach for finding the minimum spanning tree in an undirected network with trapezoidal neutrosophic edge weights was introduced by Patro et al. in 2017 [9]. Subsequently, Broumi et al. [10] proposed hybrid structures and methods for addressing medical diagnosis issues under neutrosophic information to expedite disease identification and treatment. Recently, Jdid and Smarandache (2023) [11] developed a comprehensive model for optimal agricultural land distribution using neutrosophic science, asserting that neutrosophic values can more accurately represent environmental fluctuations, natural factors affecting production, and price fluctuations, all of which impact final profitability. Broumi et al. (2023) [12] formulated the shortest route problem based on the interval set of a Fermatean neutrosophic setting and proposed a new score function for its solution. Rosli et al. (2023) [13] visualized the neutrosophic Bézier curve (NBC) of the quartic version, demonstrating the approximation of the neutrosophic control point to the NBC and introducing its mathematical construction and logical properties. Research in the field of clustering methods with neutrosophic numbers, particularly continuous numbers, is much scarcer compared to other neutrosophic problems. Ye (2014) [14] proposed an extended distance measure for single-valued neutrosophic sets (SVNSs) and introduced two similarity measures based on this distance measure, leading to a practical clustering algorithm for SVN data. Long et al. [15] utilized the neutrosophic association matrix to propose a clustering approach, constructing a neutrosophic equivalent matrix and then deriving the lambda-cutting matrix. Zhang et al. [16] are among the authors who formulated a clustering model using neutrosophic datasets, considering both original clusters and noise clusters representing real and false memberships, respectively. They defined uncertainty, factoring in data density, and employed Lagrangian multipliers to further optimize the cost function under the neutrosophic dataset.

Hierarchical clustering stands out among clustering methods for providing an optimal display of clusters through dendrograms, from the highest to the lowest possible number of clusters [17]. Despite being considered computationally intensive for extensive data, hierarchical clustering offers diverse cluster numbers, catering to expert preferences in data segmentation. Its advantages, including simplicity in utilizing distance and similarity criteria, accessibility, and flexibility, have sparked interest among researchers in expanding such algorithms. Consequently, this study aims to develop an agglomerative hierarchical clustering algorithm using trapezoidal fuzzy neutrosophic numbers. Geva (1999) [18] pioneered a fuzzy hierarchical algorithm, while Ghasemigol et al. (2010) [19] introduced a hierarchical clustering approach for fuzzy data clustering, incorporating dendrogram drawing and evaluating its performance against noisy data. Vandhana and Anuradha [20] employed fuzzy hierarchical clustering and neutrosophic fuzzy hierarchical clustering to identify dengue-affected foci in Sri Lanka, revealing clusters with varying risk levels of dengue attack. Elhassouny [21] outlined the steps of the DIANA hierarchical clustering algorithm using single-valued neutrosophic sets (SVNs). In light of the indisputable role played by distance measures in data clustering methodologies, it becomes imperative to spearhead further research aimed at introducing innovative distance metrics tailored for neutrosophic trapezoidal fuzzy numbers. Although some clustering methods have been applied to different types of neutrosophic numbers, research specifically investigating hierarchical clustering on neutrosophic trapezoidal fuzzy numbers remains scarce. Thus, the primary motivation of this research is to introduce a distance measure between neutrosophic trapezoidal fuzzy numbers based on a novel structure capable of effectively facilitating hierarchical clustering. To illustrate this, we implemented the developed hierarchical clustering method on a set of neutrosophic trapezoidal fuzzy numbers. Furthermore, in the comparative analysis, while evaluating the performance of the proposed algorithm against existing measures, we also assess its performance based on a linear combination of these measures. To proceed, the remainder of this paper is structured as follows: In Sect. 2, we gather some preliminaries and notations regarding neutrosophic sets and neutrosophic trapezoidal fuzzy numbers (NTraFNs). In Sect. 3, we introduce the novel structure and its main properties. In Sect. 4, we present the agglomerative hierarchical clustering method based on the proposed distance within the NTraFNs dataset. We provide illustrative examples, discuss the determination of the optimal number of clusters, and conduct sensitivity analyses. Furthermore, in Sect. 5, we conduct a comparative study and discuss the advantages of the suggested approach. Finally, Sect. 6 includes some concluding remarks and outlines prospects for further research.

2. Preliminaries

In this section, we will recall the preliminary concepts about Neutrosophic sets and neutrosophic trapezoidal fuzzy numbers (NTraFNs) and notations that will be used often in the rest of the paper.

Definition 1 (Samarandache 1999). Let a crisp set M be fixed. A neutrosophic set $\widetilde{N}$ in M is an object of the following representation [3]:

$$\widetilde{N} = \{\langle x, \mu_{\widetilde{N}}(x), \pi_{\widetilde{N}}(x), \rho_{\widetilde{N}}(x)\rangle \mid 0 \leq \mu_{\widetilde{N}}(x), \pi_{\widetilde{N}}(x), \rho_{\widetilde{N}}(x) \leq 1, x \in M\}, \quad (1)$$

Where functions $\mu_{\widetilde{N}}, \pi_{\widetilde{N}}, \rho_{\widetilde{N}}: \mathrm{M} \to [0,1]$ are the degree of truth-membership, indeterminacy-membership, and falsity-membership of the $x \in M$ to $\widetilde{N}$. Furthermore, for every $x \in M$, $0 \leq \mu_{\widetilde{N}}(x) + \pi_{\widetilde{N}}(x) + \rho_{\widetilde{N}}(x) \leq 3$.

Definition 2 [22]. Let a crisp set M be fixed, $\tilde{n} = \langle \mu_{\tilde{N}}(x), \pi_{\tilde{N}}(x), \rho_{\tilde{N}}(x) \rangle$ is a neutrosophic fuzzy number in the set of M. Then its truth membership function is defined as

$$\mu_{\tilde{n}}(x) = \begin{cases} \mu_{\tilde{n}}^{l}(x), & a_1 \le x \le a_2 \\ 1, & a_2 \le x \le a_3 \\ \mu_{\tilde{n}}^{r}(x), & a_3 \le x \le a_4 \\ 0, & o.w. \end{cases} \tag{2}$$

The membership function of indeterminacy is defined as

$$\pi_{\tilde{n}}(x) = \begin{cases} \pi_{\tilde{n}}^{l}(x), & b_1 \le x \le b_2 \\ 0, & b_2 \le x \le b_3 \\ \pi_{\tilde{n}}^{r}(x), & b_3 \le x \le b_4 \\ 1, & o.w. \end{cases} \tag{3}$$

Furthermore, the falsehood membership function is defined as

$$\rho_{\tilde{n}}(x) = \begin{cases} \rho_{\tilde{n}}^{l}(x), & c_1 \le x \le c_2 \\ 0, & c_2 \le x \le c_3 \\ \rho_{\tilde{n}}^{r}(x), & c_3 \le x \le c_4 \\ 1, & o.w, \end{cases} \tag{4}$$

where $0 \le \mu_{\tilde{n}}(x), \pi_{\tilde{n}}(x), \rho_{\tilde{n}}(x) \le 1$ and $0 \le \mu_{\tilde{n}}(x) + \pi_{\tilde{n}}(x) + \rho_{\tilde{n}}(x) \le 3$.

Definition 3 [3]. Let $\tilde{n}_1$ and $\tilde{n}_2$ be two neutrosophic fuzzy numbers given as $\tilde{n}_1 = \langle \mu_{\tilde{n}_1}(x), \pi_{\tilde{n}_1}(x), \rho_{\tilde{n}_1}(x) \rangle$ and $\tilde{n}_2 = \langle \mu_{\tilde{n}_2}(x), \pi_{\tilde{n}_2}(x), \rho_{\tilde{n}_2}(x) \rangle$. Then, $\tilde{n}_1 \subseteq \tilde{n}_2$ if and only if

$$\mu_{\tilde{n}_1}(x) \le \mu_{\tilde{n}_2}(x), \pi_{\tilde{n}_1}(x) \ge \pi_{\tilde{n}_2}(x), \rho_{\tilde{n}_1}(x) \ge \rho_{\tilde{n}_2}(x) \quad , for \text{ every } x \in \mathrm{M}$$

Definition 4 [22]. Let a crisp set M be fixed, $\tilde{n} = \langle (a_1, a_2, a_3, a_4), (b_1, b_2, b_3, b_4), (c_1, c_2, c_3, c_4) \rangle$ is a neutrosophic trapezoidal fuzzy number of M. Then its truth-membership function is defined as

$$\mu_{\tilde{n}}(u) = \begin{cases} \dfrac{(x - a_1)}{a_2 - a_1}, & a_1 \le x < a_2 \\ 1, & a_2 \le x \le a_3 \\ \dfrac{(a_4 - x)}{a_4 - a_3}, & a_3 \le x < a_4 \\ 0, & o.w. \end{cases} \tag{5}$$

The membership function of indeterminacy is defined as

$$\pi_{\tilde{n}}(x) = \begin{cases} \dfrac{(b_2 - x)}{b_2 - b_1}, & b_1 \le x < b_2 \\ 0, & b_2 \le x < b_3 \\ \dfrac{(x - b_3)}{b_4 - b_3}, & b_3 \le x < b_4 \\ 0, & o.w, \end{cases} \tag{6}$$

Furthermore, the falsehood membership function is defined as

$$\rho_{\tilde{n}}(x) = \begin{cases} \dfrac{(c_2 - x)}{c_2 - c_1}, & c_1 \le x < c_2 \\ 0, & c_2 \le x < c_3 \\ \dfrac{(x - c_3)}{c_4 - c_3}, & c_3 \le x < c_4 \\ 0, & o.w, \end{cases} \tag{7}$$

Where $0 \leq \mu_{\tilde{n}}(x), \pi_{\tilde{n}}(x), \rho_{\tilde{n}}(x) \leq 1$ and $0 \leq \mu_{\tilde{n}}(x) + \pi_{\tilde{n}}(x) + \rho_{\tilde{n}}(x) \leq 3$.

3. Proposed distance measure for neutrosophic trapezoidal fuzzy numbers

In this section, we will conduct a comprehensive survey of the novel distance measure, its structure, and its properties.

3.1. Structure

Here, we will to design a meaningful scheme to model the distance measure between two neutrosophic trapezoidal fuzzy numbers. The primary idea of this subsection is based on the ranking method provided by Abbasi and Darehmiraki [23]. Assume

$$\tilde{n}_i = \langle (a_{1\tilde{n}_i}, a_{2\tilde{n}_i}, a_{3\tilde{n}_i}, a_{4\tilde{n}_i}), (b_{1\tilde{n}_i}, b_{2\tilde{n}_i}, b_{3\tilde{n}_i}, b_{4\tilde{n}_i}), (c_{1\tilde{n}_i}, c_{2\tilde{n}_i}, c_{3\tilde{n}_i}, c_{4\tilde{n}_i}) \rangle,$$

be a neutrosophic trapezoidal fuzzy number. The left and right line formulas of the truth-membership function, the complement indeterminacy-membership function, and the complement falsity-membership function can be considered for $\tilde{n}_i$ to the vertical line ω =0, respectively, as follows:

$$[\sigma^{l_{\tilde{n}_i}}(\omega), \sigma^{r_{\tilde{n}_i}}(\omega)] = [(a_{1\tilde{n}_i}) + \omega(a_{2\tilde{n}_i} - a_{1\tilde{n}_i}), (a_{4\tilde{n}_i}) + \omega(a_{3\tilde{n}_i} - a_{4\tilde{n}_i})], \tag{8}$$

$$[\tau^{l_{\tilde{n}_i}}(\omega), \tau^{r_{\tilde{n}_i}}(\omega)] = [(b_{1\tilde{n}_i}) + \omega(b_{2\tilde{n}_i} - b_{1\tilde{n}_i}), (b_{4\tilde{n}_i}) + \omega(b_{3\tilde{n}_i} - b_{4\tilde{n}_i})], \tag{9}$$

$$[\varphi^{l_{\tilde{n}_i}}(\omega), \varphi^{r_{\tilde{n}_i}}(\omega)] = [(c_{1\tilde{n}_i}) + \omega(c_{2\tilde{n}_i} - c_{1\tilde{n}_i}), (c_{4\tilde{n}_i}) + \omega(c_{3\tilde{n}_i} - c_{4\tilde{n}_i})]. \tag{10}$$

In this sense, the area of the left and right half of the truth-membership interval respective to the ω =0 can state:

$$\mathcal{S}^{l_\mu}(\tilde{n}_i) = \int_0^1 \left((a_{1\tilde{n}_i}) + \omega(a_{2\tilde{n}_i} - a_{1\tilde{n}_i})\right) d\omega = (a_{1\tilde{n}_i}) + \frac{(a_{2\tilde{n}_i} - a_{1\tilde{n}_i})}{2}, \tag{11}$$

$$\mathcal{S}^{r_\mu}(\tilde{n}_i) = \int_0^1 \left((a_{4\tilde{n}_i}) + \omega(a_{3\tilde{n}_i} - a_{4\tilde{n}_i})\right) d\omega = (a_{4\tilde{n}_i}) + \frac{(a_{3\tilde{n}_i} - a_{4\tilde{n}_i})}{2}. \tag{12}$$

The area of the left and right half of the indeterminacy-membership interval respective to the ω =0 can state:

$$\mathcal{S}^{l_\pi}(\tilde{n}_i) = \int_0^1 \left(b_{1\tilde{n}_i}) + \omega(b_{2\tilde{n}_i} - b_{1\tilde{n}_i})\right) d\omega = (b_{1\tilde{n}_i}) + \frac{(b_{2\tilde{n}_i} - b_{1\tilde{n}_i})}{2}, \tag{13}$$

$$\mathcal{S}^{r_\pi}(\tilde{n}_i) = \int_0^1 \left((b_{4\tilde{n}_i}) + \omega(b_{3\tilde{n}_i} - b_{4\tilde{n}_i})\right) d\omega = (b_{4\tilde{n}_i}) + \frac{(b_{3\tilde{n}_i} - b_{4\tilde{n}_i})}{2}. \tag{14}$$

Furthermore, the area of the left and right half of the falsity-membership interval respective to the ω =0 can state:

$$\mathcal{S}^{l_\rho}(\tilde{n}_i) = \int_0^1 \left((c_{1\tilde{n}_i}) + \omega(c_{2\tilde{n}_i} - c_{1\tilde{n}_i})\right) d\omega = (c_{1\tilde{n}_i}) + \frac{(c_{2\tilde{n}_i} - c_{1\tilde{n}_i})}{2}, \tag{15}$$

$$\mathcal{S}^{r_\rho}(\tilde{n}_i) = \int_0^1 \left((c_{4\tilde{n}_i}) + \omega(c_{3\tilde{n}_i} - c_{4\tilde{n}_i})\right) d\omega = (c_{4\tilde{n}_i}) + \frac{(c_{3\tilde{n}_i} - c_{4\tilde{n}_i})}{2}. \tag{16}$$

Based on the relations 11-16, the score functions corresponding to the truth-membership function, indeterminacy-membership function, and falsity-membership function are:

$$\mathcal{R}^{\mu}(\tilde{n}_i) = \frac{1}{2}\{\mathcal{S}^{l_\mu}(\tilde{n}_i) + \mathcal{S}^{r_\mu}(\tilde{n}_i)\} = \frac{1}{2}\left\{(b_{1\tilde{n}_i}) + \frac{(b_{2\tilde{n}_i} - b_{1\tilde{n}_i})}{2} + (a_{4\tilde{n}_i}) + \frac{(a_{3\tilde{n}_i} - a_{4\tilde{n}_i})}{2}\right\}, \tag{17}$$

$$\mathcal{R}^{\pi}(\tilde{n}_i) = \frac{1}{2}\{\mathcal{S}^{l_\pi}(\tilde{n}_i) + \mathcal{S}^{r_\pi}(\tilde{n}_i)\} = \frac{1}{2}\left\{(a_{1\tilde{n}_i}) + \frac{(a_{2\tilde{n}_i} - a_{1\tilde{n}_i})}{2} + (b_{4\tilde{n}_i}) + \frac{(b_{3\tilde{n}_i} - b_{4\tilde{n}_i})}{2}\right\}, \tag{18}$$

$$\mathcal{R}^{\rho}(\tilde{n}_i) = \frac{1}{2}\{\mathcal{S}^{l_\rho}(\tilde{n}_i) + \mathcal{S}^{r_\rho}(\tilde{n}_i)\} = \frac{1}{2}\left\{(c_{1\tilde{n}_i}) + \frac{(c_{2\tilde{n}_i} - c_{1\tilde{n}_i})}{2} + (c_{4\tilde{n}_i}) + \frac{(c_{3\tilde{n}_i} - c_{4\tilde{n}_i})}{2}\right\}. \tag{19}$$

Let

$$\tilde{n}_i = \langle (a_{1\tilde{n}_i}, a_{2\tilde{n}_i}, a_{3\tilde{n}_i}, a_{4\tilde{n}_i}), (b_{1\tilde{n}_i}, b_{2\tilde{n}_i}, b_{3\tilde{n}_i}, b_{4\tilde{n}_i}), (c_{1\tilde{n}_i}, c_{2\tilde{n}_i}, c_{3\tilde{n}_i}, c_{4\tilde{n}_i}) \rangle$$

and

$$\tilde{n}_j = \left\langle \left(a_{1\tilde{n}_j}, a_{2\tilde{n}_j}, a_{3\tilde{n}_j}, a_{4\tilde{n}_j}\right), \left(b_{1\tilde{n}_j}, b_{2\tilde{n}_j}, b_{3\tilde{n}_j}, b_{4\tilde{n}_j}\right), \left(c_{1\tilde{n}_j}, c_{2\tilde{n}_j}, c_{3\tilde{n}_j}, c_{4\tilde{n}_j}\right) \right\rangle$$

are two neutrosophic trapezoidal fuzzy numbers. Then our proposed distance measure defines as

$$Dis(\tilde{n}_i, \tilde{n}_j) = \frac{1}{\sqrt{3}} \left\{ \left(\mathcal{R}^{\mu}(\tilde{n}_i) - \mathcal{R}^{\mu}(\tilde{n}_j)\right)^2 + \left(\mathcal{R}^{\pi}(\tilde{n}_i) - \mathcal{R}^{\pi}(\tilde{n}_j)\right)^2 + \left(\mathcal{R}^{\rho}(\tilde{n}_i) - \mathcal{R}^{\rho}(\tilde{n}_j)\right)^2 \right\}^{1/2}. \quad (20)$$

3.2. Theorems and properties

Now, the validity and reasonable structure of the suggested distance measure are proven according to the following theorem.

Theorem1: Show the proposed distance measure in equation (20) for all $\tilde{n}_1$, $\tilde{n}_2$ and $\tilde{n}_3$ of a dataset, including the NTraFNs are satisfied following principles:

1) $Dis(\tilde{n}_1, \tilde{n}_2) \in [0,1]$
2) $\tilde{n}_1 = \tilde{n}_2 \Longrightarrow Dis(\tilde{n}_1, \tilde{n}_2) = 0$
3) $Dis(\tilde{n}_1, \tilde{n}_2) = Dis(\tilde{n}_2, \tilde{n}_1)$
4) If $\tilde{n}_1 \subseteq \tilde{n}_2 \subseteq \tilde{n}_3 \Longrightarrow Dis(\tilde{n}_1, \tilde{n}_2) \leq Dis(\tilde{n}_1, \tilde{n}_3)$ and $Dis(\tilde{n}_1, \tilde{n}_2) \leq Dis(\tilde{n}_1, \tilde{n}_3)$.

Proof:

1. Since

$$0 \leq \left(\mathcal{R}^{\mu}(\tilde{n}_1) - \mathcal{R}^{\mu}(\tilde{n}_2)\right)^2 \leq 1,\ 0 \leq \left(\mathcal{R}^{\pi}(\tilde{n}_1) - \mathcal{R}^{\pi}(\tilde{n}_2)\right)^2 \leq 1,\ 0 \leq \left(\mathcal{R}^{\rho}(\tilde{n}_1) - \mathcal{R}^{\rho}(\tilde{n}_2)\right)^2 \leq 1,$$

Certainly, we have:

$$0 \leq Dis(\tilde{n}_1, \tilde{n}_2) \leq 1$$

2. If $\tilde{n}_1 = \tilde{n}_2$, then

$$\mathcal{R}^{\mu}(\tilde{n}_1) = \mathcal{R}^{\mu}(\tilde{n}_2), \qquad \mathcal{R}^{\pi}(\tilde{n}_1) = \mathcal{R}^{\pi}(\tilde{n}_2), \qquad \mathcal{R}^{\rho}(\tilde{n}_1) = \mathcal{R}^{\rho}(\tilde{n}_2)$$

So, $Dis(\tilde{n}_1, \tilde{n}_2) = 0$.

3. For every $\tilde{n}_1$ and $\tilde{n}_2$, Since the each of the expressions in (20) is positive value. We have:

$$Dis(\tilde{n}_1, \tilde{n}_2) = \frac{1}{\sqrt{3}} \left\{ \left(\mathcal{R}^{\mu}(\tilde{n}_1) - \mathcal{R}^{\mu}(\tilde{n}_2)\right)^2 + \left(\mathcal{R}^{\pi}(\tilde{n}_1) - \mathcal{R}^{\pi}(\tilde{n}_2)\right)^2 + \left(\mathcal{R}^{\rho}(\tilde{n}_1) - \mathcal{R}^{\rho}(\tilde{n}_2)\right)^2 \right\}^{1/2}$$

$$= \frac{1}{\sqrt{3}} \left\{ \left(\mathcal{R}^{\mu}(\tilde{n}_2) - \mathcal{R}^{\mu}(\tilde{n}_1)\right)^2 + \left(\mathcal{R}^{\pi}(\tilde{n}_2) - \mathcal{R}^{\pi}(\tilde{n}_1)\right)^2 + \left(\mathcal{R}^{\rho}(\tilde{n}_2) - \mathcal{R}^{\rho}(\tilde{n}_1)\right)^2 \right\}^{1/2}$$

$$= Dis(\tilde{n}_2, \tilde{n}_1).$$

4. If $\tilde{n}_1 \subseteq \tilde{n}_2 \subseteq \tilde{n}_3$, then:

$$\left(\mathcal{R}^{\mu}(\tilde{n}_1) - \mathcal{R}^{\mu}(\tilde{n}_2)\right) \leq \left(\mathcal{R}^{\mu}(\tilde{n}_1) - \mathcal{R}^{\mu}(\tilde{n}_3)\right),$$
$$\left(\mathcal{R}^{\pi}(\tilde{n}_1) - \mathcal{R}^{\pi}(\tilde{n}_2)\right) \leq \left(\mathcal{R}^{\pi}(\tilde{n}_1) - \mathcal{R}^{\pi}(\tilde{n}_3)\right),$$
$$\left(\mathcal{R}^{\rho}(\tilde{n}_1) - \mathcal{R}^{\rho}(\tilde{n}_2)\right) \leq \left(\mathcal{R}^{\rho}(\tilde{n}_1) - \mathcal{R}^{\rho}(\tilde{n}_3)\right).$$

Hence

$$\frac{1}{\sqrt{3}} \left\{ \left(\mathcal{R}^{\mu}(\tilde{n}_1) - \mathcal{R}^{\mu}(\tilde{n}_2)\right)^2 + \left(\mathcal{R}^{\pi}(\tilde{n}_1) - \mathcal{R}^{\pi}(\tilde{n}_2)\right)^2 + \left(\mathcal{R}^{\rho}(\tilde{n}_1) - \mathcal{R}^{\rho}(\tilde{n}_2)\right)^2 \right\}^{1/2} \leq$$

$$\frac{1}{\sqrt{3}} \left\{ \left(\mathcal{R}^{\mu}(\tilde{n}_1) - \mathcal{R}^{\mu}(\tilde{n}_3)\right)^2 + \left(\mathcal{R}^{\pi}(\tilde{n}_1) - \mathcal{R}^{\pi}(\tilde{n}_3)\right)^2 + \left(\mathcal{R}^{\rho}(\tilde{n}_1) - \mathcal{R}^{\rho}(\tilde{n}_3)\right)^2 \right\}^{1/2},$$

Therefore $Dis(\tilde{n}_1, \tilde{n}_2) \leq Dis(\tilde{n}_1, \tilde{n}_3)$.

Similarly, $Dis(\tilde{n}_2, \tilde{n}_3) \leq Dis(\tilde{n}_1, \tilde{n}_3)$ also holds.

Property 1: If $\tilde{n}_1 = \langle (a,a,a,a), (a,a,a,a), (a,a,a,a) \rangle$ and $\tilde{n}_2 = \langle (b,b,b,b), (b,b,b,b), (b,b,b,b) \rangle$ then, $Dis(\tilde{n}_1, \tilde{n}_2) = |a - b|$.

Property 2: If $\tilde{n}_1 = \langle (a,a,a,a), (a,a,a,a), (a,a,a,a) \rangle$ and $\tilde{n}_2 = \langle (0,0,0,0), (0,0,0,0), (0,0,0,0) \rangle$ then, $Dis(\tilde{n}_1, \tilde{n}_2) = a$.

Property 3: If $\tilde{n}_1 = \langle (a,a,a,a),(a,a,a,a),(a,a,a,a)\rangle$ and $\tilde{n}_2 = \langle (1,1,1,1),(1,1,1,1),(1,1,1,1)\rangle$ then, $Dis(\tilde{n}_1, \tilde{n}_2) = |1-a|$.

Example 1: Consider the following TrNFNs. $\tilde{n}_1 = \langle (0,0,0,0),(0,0,0,0),(0,0,0,0)\rangle, \tilde{n}_2 = \langle (1,1,1,1),(1,1,1,1),(1,1,1,1)\rangle$, $\tilde{n}_3 = \langle (0.5,0.5,0.5,0.5),(0.5,0.5,0.5,0.5),(0.5,0.5,0.5,0.5)\rangle$, $\tilde{n}_4 = \langle (0.8,0.8,0.8,0.8),(0.8,0.8,0.8,0.8),(0.8,0.8,0.8,0.8)\rangle$.

Distance measures; $Dis(\tilde{n}_i, \tilde{n}_j)$ for every $i,j = 1,2,3,4$ calculated in Table 1.

Table 1. Distance measures of Ex 1.

$Dis(\tilde{n}_i, \tilde{n}_j)$	$\tilde{n}_1$	$\tilde{n}_2$	$\tilde{n}_3$	$\tilde{n}_4$
$\tilde{n}_1$	0	1.0000	0.5000	0.8000
$\tilde{n}_2$	1.0000	0	0.5000	0.2000
$\tilde{n}_3$	0.5000	0.5000	0	0.3000
$\tilde{n}_4$	0.8000	0.2000	0.3000	0

The results of Table 1 show the logical performance related to the properties of the proposed distance measure.

4. Agglomerative hierarchical clustering (AHC) under neutrosophic trapezoidal fuzzy numbers

Hierarchical clustering methods that use a dendrogram tree to depict the status of data relative to each other at different levels are known as the most widely used methods in clustering. Therefore, while providing a proper picture of the data placement in the clusters, it reveals a proper picture of the sub-clusters at each stage. Two general views, bottom-up and top-down, are known as the main approaches in hierarchical clustering. In the implementation mechanism of this method, based on the first approach, each data point is assigned to a cluster. Then the pairs of clusters that are closest to each other are combined. How to combine the clusters with spare parts will be influential in determining the final shape of the cluster. Some of these criteria to determine the distance between two clusters are:

Single-link method: The similarity or closeness between two clusters is measured based on the smallest distance between the members of the two clusters. In this case, for two clusters, we have:

$$Dis(C_i, C_j) = \min_{a\in C_i,\ b\in C_j} Dis(a,b) \tag{21}$$

Complete-link method: The similarity or closeness between two clusters is measured based on the highest distance between the members of the two clusters. In this case, for two clusters, we have:

$$Dis(\mathrm{C_i}, \mathrm{C}_j) = \max_{a\in \mathrm{C_i},\ b\in \mathrm{C}_j} Dis(\mathrm{a,b}) \tag{22}$$

Average-link method: The similarity or closeness between two clusters is measured based on the unweighted average distance between the members of the two clusters. In this case, for two clusters, we have:

$$Dis(\mathrm{C_i}, \mathrm{C}_j) = \frac{1}{|\mathrm{C_i}||\mathrm{C_j}|} \sum_{a\in \mathrm{C_i},\ b\in \mathrm{C}_j} Dis(\mathrm{a,b}) \tag{23}$$

Our goal in this part is to present a hybrid algorithm based on the first approach. The details of the second approach can be found in [24].

4.1. AHC algorithem under neutrosophic trapezoidal fuzzy numbers

Clustering is the separation of a heterogeneous group of data into several clusters, with the goal of maximizing the inter-cluster difference and minimizing the intra-cluster difference. In the AHC algorithm, each data point is placed in a separate cluster at first. Then, according to the dispersion of the data, the two closest clusters (according to the selected distance comparison) are merged. This process continues until all the data are in one cluster.

The steps of the AHC procedure under the NtraFN dataset are described as follows:

Step 1: Record and identify the elements of NtraFN dataset

$$D_N = \{\tilde{n}_1, \tilde{n}_2, \dots, \tilde{n}_m\}$$

Where $\tilde{n}_i = \langle (a_{1\tilde{n}_i}, a_{2\tilde{n}_i}, a_{3\tilde{n}_i}, a_{4\tilde{n}_i}), (b_{1\tilde{n}_i}, b_{2\tilde{n}_i}, b_{3\tilde{n}_i}, b_{4\tilde{n}_i}), (c_{1\tilde{n}_i}, c_{2\tilde{n}_i}, c_{3\tilde{n}_i}, c_{4\tilde{n}_i}) \rangle$ and m denote the number of data.

Step 2: Set $\delta = \max_{\substack{l=1,2,3,4 \\ i=1,2,\dots,m}} \{a_{l\tilde{n}_i}, b_{l\tilde{n}_i}, c_{l\tilde{n}_i}\}$, the normalized NTraF-dataset denote by $\overline{D}_N$ which each element obtain as

$$\tilde{n}_i^{nor} = \tilde{n}_i/\delta = \langle (a_{1\tilde{n}_i}, a_{2\tilde{n}_i}, a_{3\tilde{n}_i}, a_{4\tilde{n}_i}), (b_{1\tilde{n}_i}, b_{2\tilde{n}_i}, b_{3\tilde{n}_i}, b_{4\tilde{n}_i}), (c_{1\tilde{n}_i}, c_{2\tilde{n}_i}, c_{3\tilde{n}_i}, c_{4\tilde{n}_i}) \rangle/\delta.$$

Step 3: Transform $\overline{D}_N$ to $\overline{D}_N^R$ such that:

$$\overline{D}_N^R = \{\mathcal{R}(\tilde{n}_1^{nor}), \mathcal{R}(\tilde{n}_2^{nor}), \dots, \mathcal{R}(\tilde{n}_m^{nor})\}.$$

And $\mathcal{R}(\tilde{n}_i^{nor}) = \langle \mathcal{R}^{\mu}(\tilde{n}_i^{nor}), \mathcal{R}^{\pi}(\tilde{n}_i^{nor}), \mathcal{R}^{\rho}(\tilde{n}_i^{nor}) \rangle$ for every $i \epsilon \{1,2,\dots,m\}$.

Step 4: Place each element of $\overline{D}_N^R$ in a separate cluster. So, the number of clusters is $k = m$.

Step 5: Form the Neutrosophic distance matrix, between clusters according to equation (20) as follows

$$[Dis(\tilde{n}_i, \tilde{n}_j)]_{m*m}$$

Step 6: To choose the shortest distance between two clusters for merging, select (fix), and use one of the criteria presented in equations 21, 22 and 23.

Step 7: Put $k = m - 1$. Then, repeat steps 5 to 7 until we reach k=1.

Note: In big data, for the termination condition, k=t can be considered as the optimal number of clusters (1≤t≤m).

4.2. Illustrative example

In this part, the experimental study is conducted using the previously suggested method. A typical example includes collecting trapezoidal fuzzy numbers, which is used for clustering methods in [19]. We, have

$$D_F = \{\tilde{f}_1, \tilde{f}_2, \tilde{f}_3, \tilde{f}_4, \tilde{f}_5, \tilde{f}_6\} = \left\{ \begin{matrix} \langle (1.0, 1.5, 3.5, 4.5) \rangle, \langle (1.5, 2.5, 2.5, 3.5) \rangle, \langle (6.0, 6.5, 7.5, 8.5) \rangle, \\ \langle (7.0, 8.0, 8.0, 9.0) \rangle, \langle (8.0, 8.5, 8.5, 9.0) \rangle, \langle (3.5, 4.5, 4.5, 5.5) \rangle \end{matrix} \right\}.$$

Due to the intrinsic limitations of NTraF data, Without making a noticeable change in the distribution of fuzzy data from D_F, by applying the following equation $\tilde{f}_i = \langle (a_{1\tilde{n}_i}, a_{2\tilde{n}_i}, a_{3\tilde{n}_i}, a_{4\tilde{n}_i}) \rangle$ so

$$\tilde{n}_i = \left\langle \begin{matrix} (a_{1\tilde{n}_i}, a_{2\tilde{n}_i}, a_{3\tilde{n}_i}, a_{4\tilde{n}_i}), (a_{1\tilde{n}_i} - 0.1, a_{2\tilde{n}_i} - 0.1, a_{3\tilde{n}_i} - 0.1, a_{4\tilde{n}_i} - 0.1), \\ (a_{1\tilde{n}_i} + 0.1, a_{2\tilde{n}_i} + 0.1, a_{3\tilde{n}_i} + 0.1, a_{4\tilde{n}_i} + 0.1) \end{matrix} \right\rangle \quad (24)$$

We can construct the following NTraF-dataset as follows

$D_N = \{\tilde{n}_1, \tilde{n}_2, \tilde{n}_3, \tilde{n}_4, \tilde{n}_5, \tilde{n}_6\}$
$= \{\langle (1.0, 1.5, 3.5, 4.5), (0.9, 1.4, 3.4, 4.4), (1.1, 1.6, 3.6, 4.6) \rangle, \langle (1.5, 2.5, 2.5, 3.5), (1.4, 2.4, 2.4, 3.4), (1.6, 2.6, 2.6, 3.6) \rangle,$
$\langle (6.0, 6.5, 7.5, 8.5), (5.9, 6.4, 7.4, 8.4), (6.1, 6.6, 7.6, 8.6) \rangle, \langle (7.0, 8.0, 8.0, 9.0), (6.9, 7.9, 7.9, 8.9), (7.1, 8.1, 8.1, 9.1) \rangle,$

$\langle(8.0, 8.5, 8.5, 9.0), (7.9, 8.4, 8.4, 8.9), (8.1, 8.6, 8.6, 9.1)\rangle, \langle(3.5, 4.5, 4.5, 5.5), (3.4, 4.4, 4.4, 5.4), (3.6, 4.6, 4.6, 5.6)\rangle\}$

In Figure 1 a, b, the elements of D_F and D_N is represented.

From step 2, since $\max_{\substack{l=1,2,3,4\\ i=1,2,\dots,6}} \{a_{l\tilde{n}_i}, b_{l\tilde{n}_i}, c_{l\tilde{n}_i}\} = 9.1$, the normalized NTraF-dataset is obtained as

$$\overline{D}_N = \{\langle(0.1099, 0.1648, 0.3846, 0.4945), (0.0989, 0.1538, 0.3736, 0.4835), (0.1209, 0.1758, 0.3956, 0.5055)\rangle,$$
$$\langle(0.1648, 0.2747, 0.2747, 0.3846), (0.1538, 0.2637, 0.2637, 0.3736), (0.1758, 0.2857, 0.2857, 0.3956)\rangle,$$
$$\langle(0.6593, 0.7143, 0.8242, 0.9341), (0.6484, 0.7033, 0.8132, 0.9231), (0.6703, 0.7253, 0.8352, 0.9451)\rangle,$$
$$\langle(0.7692, 0.8791, 0.8791, 0.9890), (0.7582, 0.8681, 0.8681, 0.9780), (0.7802, 0.8901, 0.8901, 1.0000)\rangle,$$
$$\langle(0.8791, 0.9341, 0.9341, 0.9890), (0.8681, 0.9231, 0.9231, 0.9780), (0.8901, 0.9451, 0.9451, 1.0000)\rangle,$$
$$\langle(0.3846, 0.4945, 0.4945, 0.6044), (0.3736, 0.4835, 0.4835, 0.5934), (0.3956, 0.5055, 0.5055, 0.6154)\rangle\}.$$

For each element of the $\overline{D}_N$, we calculate the triple ranks corresponding to the truth-membership, indeterminacy-membership, and falsity-membership functions. Triple-ranks replace the previous elements of $\overline{D}_N$. The new set is named $\overline{D}_N^R$ with the following representation.

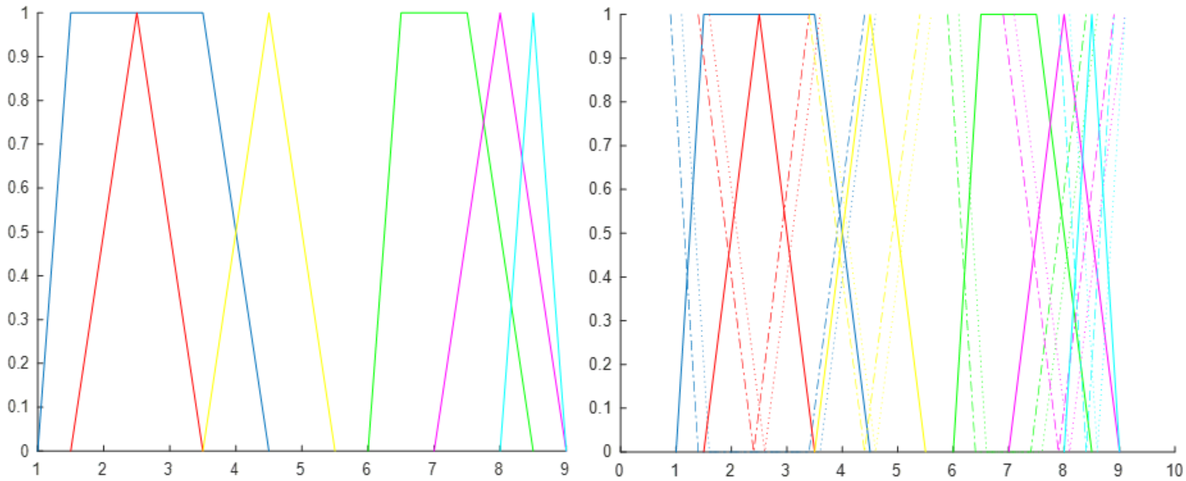

Figure 1. a) The elements of D_F (TraF-dataset), b) The elements of D_N(NTraF-dataset).

$$\overline{D}_N^R = \{\langle 0.2885, 0.3187, 0.2995\rangle, \langle 0.2747, 0.2363, 0.2857\rangle, \langle 0.7830, 0.7857, 0.7940\rangle,$$
$$\langle 0.8791, 0.8407, 0.8901\rangle, \langle 0.9341, 0.9093, 0.9451\rangle, \langle 0.4945, 0.4560, 0.5055\rangle\}.$$

From step 4, and equation 20 the distance measure matrix is obtained, which is shown in Table 2.

Table 2. Distance measure matrix of illustrative example (k=6).

$Dis(\tilde{n}_i, \tilde{n}_j)$	$\tilde{n}_1$	$\tilde{n}_2$	$\tilde{n}_3$	$\tilde{n}_4$	$\tilde{n}_5$	$\tilde{n}_6$
$\tilde{n}_1$	0	0.0489	0.4855	0.5687	0.6287	0.1860
$\tilde{n}_2$	0.0489	0	0.5223	0.6044	0.6640	0.2198
$\tilde{n}_3$	0.4855	0.5223	0	0.0847	0.1425	0.3028
$\tilde{n}_4$	0.5687	0.6044	0.0847	0	0.0599	0.3864
$\tilde{n}_5$	0.6287	0.6640	0.1425	0.0599	0	0.4442
$\tilde{n}_6$	0.1860	0.2198	0.3028	0.3846	0.4442	0

The distance results of Table 2 show $Dis(\tilde{n}_1, \tilde{n}_2) = 0.0489$ has the smallest value among all distances. So, $\tilde{n}_1$ and $\tilde{n}_2$ are placed together in one cluster. The information of the new distance matrix is updated by merging the rows and columns corresponding to these two elements and based on the selected distance criterion (single linkage is selected) in Table 3.

Table 3. Distance measure matrix of illustrative example (k=5).

	$\{\tilde{n}_1, \tilde{n}_2\}$	$\tilde{n}_3$	$\tilde{n}_4$	$\tilde{n}_5$	$\tilde{n}_6$
$\{\tilde{n}_1, \tilde{n}_2\}$	0	0.4855	0.5687	0.6278	0.1860
$\tilde{n}_3$	0.4855	0	0.0847	0.1425	0.3028
$\tilde{n}_4$	0.5687	0.0847	0	0.0599	0.3846
$\tilde{n}_5$	0.6278	0.1425	0.0599	0	0.4442
$\tilde{n}_6$	0.1860	0.3028	0.3846	0.4442	0

With the continuation of this iterative process, the results of the distance measure matrices are summarized in Tables 4 to 6 for k=4, 3, 2.

Table 4. Distance measure matrix of illustrative example (k=4).

	$\{\tilde{n}_1, \tilde{n}_2\}$	$\tilde{n}_3$	$\{\tilde{n}_4, \tilde{n}_5\}$	$\tilde{n}_6$
$\{\tilde{n}_1, \tilde{n}_2\}$	0	0.4855	0.5687	0.1860
$\tilde{n}_3$	0.4855	0	0.0847	0.3028
$\{\tilde{n}_4, \tilde{n}_5\}$	0.5687	0.0847	0	0.3846
$\tilde{n}_6$	0.1860	0.3028	0.3846	0

Table 5. Distance measure matrix of illustrative example (k=3).

	$\{\tilde{n}_1, \tilde{n}_2\}$	$\{\tilde{n}_3, \tilde{n}_4, \tilde{n}_5\}$	$\tilde{n}_6$
$\{\tilde{n}_1, \tilde{n}_2\}$	0	0.4855	0.1860
$\{\tilde{n}_3, \tilde{n}_4, \tilde{n}_5\}$	0.4855	0	0.3028
$\tilde{n}_6$	0.1860	0.3028	0

Table 6. Distance measure matrix of illustrative example (k=2).

	$\{\tilde{n}_1, \tilde{n}_2, \tilde{n}_6\}$	$\{\tilde{n}_3, \tilde{n}_4, \tilde{n}_5\}$
$\{\tilde{n}_1, \tilde{n}_2, \tilde{n}_6\}$	0	0.4855
$\{\tilde{n}_3, \tilde{n}_4, \tilde{n}_5\}$	0.4855	0

Finally, all NTraF data are in the same cluster. Figure 2 depicts the elements of D_N . The first one shows the elements of $\overline{D}_N^R$, before clustering, while the second shows the elements of $\overline{D}_N^R$, after clustering; two groups of 6 subjects are classified as cluster 1 (red color) and cluster 2 (cyan color) have been simulated.

A visual display in the form of a dendrogram diagram is given in Figure 3 for a better analysis of the findings related to tables 2 to 6.

The comparison between AHC-fuzzy numbers and AHC-neutrosophic numbers about suggested distance is summarized in Table 7.

Table 7. Comparison between AHC-fuzzy numbers and AHC-neutrosophic numbers.

Cluster	AHC-fuzzy numbers	AHC-neutrosophic numbers
6	$\{\tilde{f}_1\},\{\tilde{f}_2\},\{\tilde{f}_3\},\{\tilde{f}_4\},\{\tilde{f}_5\},\{\tilde{f}_6\}$	$\{\tilde{n}_1\},\{\tilde{n}_2\},\{\tilde{n}_3\},\{\tilde{n}_4\},\{\tilde{n}_5\},\{\tilde{n}_6\}$
5	$\{\tilde{f}_1,\tilde{f}_2\},,\{\tilde{f}_3\},\{\tilde{f}_4\},\{\tilde{f}_5\},\{\tilde{f}_6\}$	$\{\tilde{n}_1,\tilde{n}_2\},\{\tilde{n}_3\},\{\tilde{n}_4\},\{\tilde{n}_5\},\{\tilde{n}_6\}$
4	$\{\tilde{f}_1,\tilde{f}_2\},\{\tilde{f}_3\},\{\tilde{f}_4,\tilde{f}_5\},\{\tilde{f}_6\}$	$\{\tilde{n}_1,\tilde{n}_2\},\{\tilde{n}_3\},\{\tilde{n}_4,\tilde{n}_5\},\{\tilde{n}_6\}$
3	$\{\tilde{f}_1,\tilde{f}_2\},\{\tilde{f}_3,\tilde{f}_4,\tilde{f}_5\},\{\tilde{f}_6\}$	$\{\tilde{n}_1,\tilde{n}_2\},\{\tilde{n}_3,\tilde{n}_4,\tilde{n}_5\},\{\tilde{n}_6\}$

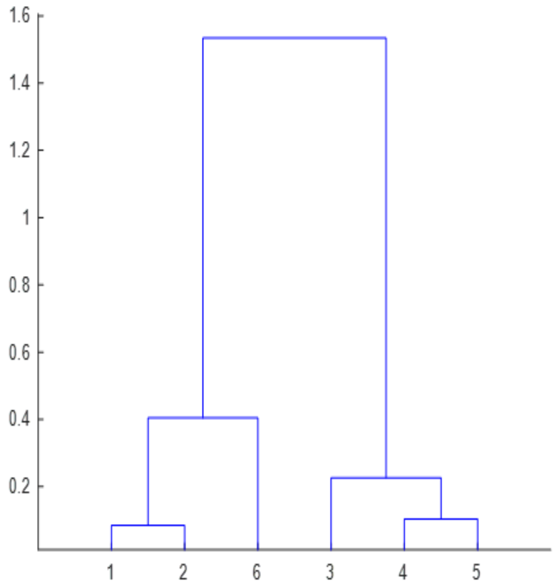

Figure 2. Dendrogram of D_N (NTraF-dataset).

2	$\{\tilde{f}_1,\tilde{f}_2,\tilde{f}_6\},\{\tilde{f}_3,\tilde{f}_4,\tilde{f}_5\}$	$\{\tilde{n}_1,\tilde{n}_2,\tilde{n}_6\},\{\tilde{n}_3,\tilde{n}_4,\tilde{n}_5\}$
1	$\{\tilde{f}_1,\tilde{f}_2,\tilde{f}_3,\tilde{f}_4,\tilde{f}_5,\tilde{f}_6\}$	$\{\tilde{n}_1,\tilde{n}_2,\tilde{n}_3,\tilde{n}_4,\tilde{n}_5,\tilde{n}_6\}$
run time	0.063351	0.066300

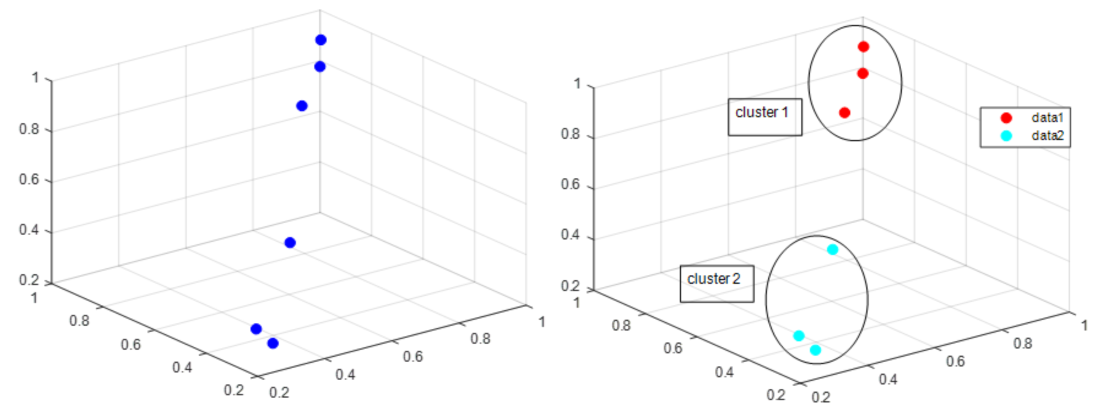

Figure 3. The elements of $\bar{D}_N^R$, a) before clustering and b), after clustering (k=2).

4.3. Optimal number of clusters

Although in hierarchical clustering methods, it is not necessary to determine the optimal number of clusters, instead, the focus is on the relationships between clusters, however, from the dendrogram presented in Figure 3. It is possible to guess the optimal number of clusters for the appropriate separation of the data. Of course, in this part, for more certainty, two criteria, silhouette method and gap statistic, are used to determine the optimal number of clusters after removing the apparent states of one cluster and six clusters. In the first one, the average silhouette value for each member is calculated and depicted for any arbitrary number of clusters. The maximum point in the silhouette diagram represents the optimal number of clusters. The second one compares the sum of the intra-cluster differences of the data with their expected values (data distribution assuming the null hypothesis is authentic) for any arbitrary number of clusters.

Here, the point for which the gap statistic is maximum shows the optimal number of clusters. More methods of evaluation criteria found in [25-26]. As can be seen in Figure 4, the results of two

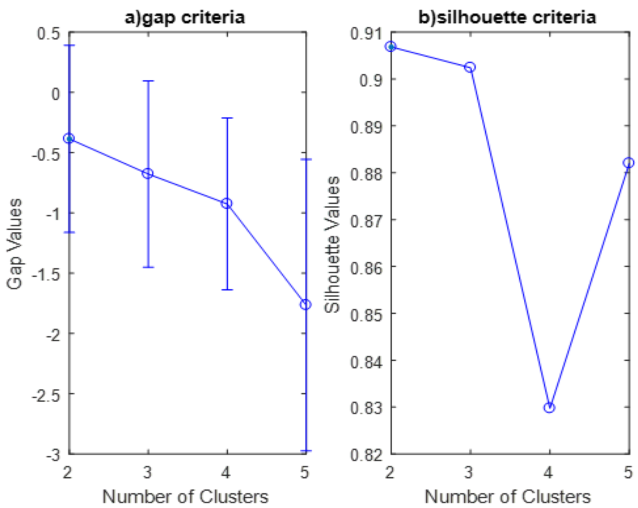

Figure 4. Optimal number of clusters based on two criteria.

indexes also recommend the number of two for the optimal number of clusters.

4.4. Sensitivity analysis

Consider the data set in the D_N. It is clear that the same and simultaneous perturbation on all numbers in D_N cannot provide a different evaluation than that obtained in Subsection 4.2. Hence, it makes more sense to examine the changes in each of the numbers separately. It is also assumed that when applying each of the changes, no changes occur in other samples. We want to know how much the increase or decrease in the values related to the NTraF will not change the hierarchical clustering structure obtained in Table 7. It is clear that any decrease or increase in the α will not necessarily change the current results. For this purpose, we want to check the turbulence performance based on the following four modes for $\tilde{n}_4 = \langle(7.0, 8.0, 8.0, 9.0), (6.9, 7.9, 7.9, 8.9), (7.1, 8.1, 8.1, 9.1)\rangle$.

Case 1. Consider the following changes in all values of $\tilde{n}_4$

$$\tilde{n}_4^{new} = \langle(7.0, 8.0, 8.0, 9.0), (6.9, 7.9, 7.9, 8.9), (7.1, 8.1, 8.1, 9.1)\rangle + \langle(\alpha, \alpha, \alpha, \alpha), (\alpha, \alpha, \alpha, \alpha), (\alpha, \alpha, \alpha, \alpha)\rangle.$$

Changes for $\alpha\epsilon(-0.12, 0.11)$ shows the result in steps of AHC-neutrosophic numbers remain stable. Meanwhile, the clustering results change for $\alpha = -0.12$ and $\alpha = 0.11$ according to what is given in Table 8.

Table 8. AHC-neutrosophic results based on sensitive analysis for Case 1.

Cluster	AHC-neutrosophic numbers ($\alpha = 0.11$)	AHC-neutrosophic numbers ($\alpha = -0.12$)
6	$\{\tilde{n}_1\}, \{\tilde{n}_2\}, \{\tilde{n}_3\}, \{\tilde{n}_4\}, \{\tilde{n}_5\}, \{\tilde{n}_6\}$	$\{\tilde{n}_1\}, \{\tilde{n}_2\}, \{\tilde{n}_3\}, \{\tilde{n}_4\}, \{\tilde{n}_5\}, \{\tilde{n}_6\}$
5	$\{\tilde{n}_1\}, \{\tilde{n}_2\}, \{\tilde{n}_3\}, \{\tilde{n}_4, \tilde{n}_5\}, \{\tilde{n}_6\}$	$\{\tilde{n}_1, \tilde{n}_2\}, \{\tilde{n}_3\}, \{\tilde{n}_4\}, \{\tilde{n}_5\}, \{\tilde{n}_6\}$
4	$\{\tilde{n}_1, \tilde{n}_2\}, \{\tilde{n}_3\}, \{\tilde{n}_4, \tilde{n}_5\}, \{\tilde{n}_6\}$	$\{\tilde{n}_1, \tilde{n}_2\}, \{\tilde{n}_3, \tilde{n}_4\}, \{\tilde{n}_5\}, \{\tilde{n}_6\}$
3	$\{\tilde{n}_1, \tilde{n}_2\}, \{\tilde{n}_3, \tilde{n}_4, \tilde{n}_5\}, \{\tilde{n}_6\}$	$\{\tilde{n}_1, \tilde{n}_2\}, \{\tilde{n}_3, \tilde{n}_4, \tilde{n}_5\}, \{\tilde{n}_6\}$
2	$\{\tilde{n}_1, \tilde{n}_2, \tilde{n}_6\}, \{\tilde{n}_3, \tilde{n}_4, \tilde{n}_5\}$	$\{\tilde{n}_1, \tilde{n}_2, \tilde{n}_6\}, \{\tilde{n}_3, \tilde{n}_4, \tilde{n}_5\}$
1	$\{\tilde{n}_1, \tilde{n}_2, \tilde{n}_3, \tilde{n}_4, \tilde{n}_5, \tilde{n}_6\}$	$\{\tilde{n}_1, \tilde{n}_2, \tilde{n}_3, \tilde{n}_4, \tilde{n}_5, \tilde{n}_6\}$

Case 2. Consider the following changes in $\tilde{n}_4$

$\tilde{n}_4^{new} = \langle(7.0, 8.0, 8.0, 9.0), (6.9, 7.9, 7.9, 8.9), (7.1, 8.1, 8.1, 9.1)\rangle + \langle(\alpha, \alpha, \alpha, \alpha), (0,0,0,0), (0,0,0,0)\rangle.$

Changes for $\alpha\epsilon(-0.33, 1.88)$ shows the result in steps of AHC-neutrosophic numbers remain stable. Meanwhile, the clustering results change for $\alpha = -0.33$ and $\alpha = 1.88$ according to what is given in Table 9.

Table 9. AHC-neutrosophic results based on sensitive analysis for Case 2.

Cluster	AHC-neutrosophic numbers ($\alpha = 1.88$)	AHC-neutrosophic numbers ($\alpha = -0.33$)
6	$\{\tilde{n}_1\}, \{\tilde{n}_2\}, \{\tilde{n}_3\}, \{\tilde{n}_4\}, \{\tilde{n}_5\}, \{\tilde{n}_6\}$	$\{\tilde{n}_1\}, \{\tilde{n}_2\}, \{\tilde{n}_3\}, \{\tilde{n}_4\}, \{\tilde{n}_5\}, \{\tilde{n}_6\}$
5	$\{\tilde{n}_1, \tilde{n}_2\}, \{\tilde{n}_3\}, \{\tilde{n}_4\}, \{\tilde{n}_5\}, \{\tilde{n}_6\}$	$\{\tilde{n}_1, \tilde{n}_2\}, \{\tilde{n}_3\}, \{\tilde{n}_4\}, \{\tilde{n}_5\}, \{\tilde{n}_6\}$
4	$\{\tilde{n}_1, \tilde{n}_2\}, \{\tilde{n}_3\}, \{\tilde{n}_4, \tilde{n}_5\}, \{\tilde{n}_6\}$	$\{\tilde{n}_1, \tilde{n}_2\}, \{\tilde{n}_3, \tilde{n}_4\}, \{\tilde{n}_5\}, \{\tilde{n}_6\}$
3	$\{\tilde{n}_1, \tilde{n}_2, \tilde{n}_6\}, \{\tilde{n}_3\}, \{\tilde{n}_4, \tilde{n}_5\}$	$\{\tilde{n}_1, \tilde{n}_2\}, \{\tilde{n}_3, \tilde{n}_4, \tilde{n}_5\}, \{\tilde{n}_6\}$
2	$\{\tilde{n}_1, \tilde{n}_2, \tilde{n}_6\}, \{\tilde{n}_3, \tilde{n}_4, \tilde{n}_5\}$	$\{\tilde{n}_1, \tilde{n}_2, \tilde{n}_6\}, \{\tilde{n}_3, \tilde{n}_4, \tilde{n}_5\}$
1	$\{\tilde{n}_1, \tilde{n}_2, \tilde{n}_3, \tilde{n}_4, \tilde{n}_5, \tilde{n}_6\}$	$\{\tilde{n}_1, \tilde{n}_2, \tilde{n}_3, \tilde{n}_4, \tilde{n}_5, \tilde{n}_6\}$

Case 3. Consider the following changes in $\tilde{n}_4$

$\tilde{n}_4^{new} = \langle(7.0, 8.0, 8.0, 9.0), (6.9, 7.9, 7.9, 8.9), (7.1, 8.1, 8.1, 9.1)\rangle + \langle(0,0,0,0), (\alpha, \alpha, \alpha, \alpha), (0,0,0,0)\rangle.$

Changes for $\alpha\epsilon(-0.40, 2.16)$ shows the result in steps of AHC-neutrosophic numbers remain stable. Meanwhile, the clustering results change for $\alpha = -0.40$ and $\alpha = 2.16$ according to what is given in Table 10.

Table 10. AHC-neutrosophic results based on sensitive analysis for Case 3.

Cluster	AHC-neutrosophic numbers ($\alpha = 2.16$)	AHC-neutrosophic numbers ($\alpha = -0.40$)
6	$\{\tilde{n}_1\},\{\tilde{n}_2\},\{\tilde{n}_3\},\{\tilde{n}_4\},\{\tilde{n}_5\},\{\tilde{n}_6\}$	$\{\tilde{n}_1\},\{\tilde{n}_2\},\{\tilde{n}_3\},\{\tilde{n}_4\},\{\tilde{n}_5\},\{\tilde{n}_6\}$
5	$\{\tilde{n}_1,\tilde{n}_2\},\{\tilde{n}_3\},\{\tilde{n}_4\},\{\tilde{n}_5\},\{\tilde{n}_6\}$	$\{\tilde{n}_1,\tilde{n}_2\},\{\tilde{n}_3\},\{\tilde{n}_4\},\{\tilde{n}_5\},\{\tilde{n}_6\}$
4	$\{\tilde{n}_1,\tilde{n}_2\},\{\tilde{n}_3\},\{\tilde{n}_4,\tilde{n}_5\},\{\tilde{n}_6\}$	$\{\tilde{n}_1,\tilde{n}_2\},\{\tilde{n}_3,\tilde{n}_4\},\{\tilde{n}_5\},\{\tilde{n}_6\}$
3	$\{\tilde{n}_1,\tilde{n}_2,\tilde{n}_6\},\{\tilde{n}_3\},\{\tilde{n}_4,\tilde{n}_5\}$	$\{\tilde{n}_1,\tilde{n}_2\},\{\tilde{n}_3,\tilde{n}_4,\tilde{n}_5\},\{\tilde{n}_6\}$
2	$\{\tilde{n}_1,\tilde{n}_2,\tilde{n}_6\},\{\tilde{n}_3,\tilde{n}_4,\tilde{n}_5\}$	$\{\tilde{n}_1,\tilde{n}_2,\tilde{n}_6\},\{\tilde{n}_3,\tilde{n}_4,\tilde{n}_5\}$
1	$\{\tilde{n}_1,\tilde{n}_2,\tilde{n}_3,\tilde{n}_4,\tilde{n}_5,\tilde{n}_6\}$	$\{\tilde{n}_1,\tilde{n}_2,\tilde{n}_3,\tilde{n}_4,\tilde{n}_5,\tilde{n}_6\}$

Case 4. Consider the following changes in $\tilde{n}_4$

$\tilde{n}_4^{new} = \langle(7.0, 8.0, 8.0, 9.0), (6.9, 7.9, 7.9, 8.9), (7.1, 8.1, 8.1, 9.1)\rangle + \langle\,(0,0,0,0), (0,0,0,0), (\alpha,\alpha,\alpha,\alpha)\rangle.$

Changes for $\alpha\epsilon(-0.33,1.88)$ shows the result in steps of AHC-neutrosophic numbers remain stable. Meanwhile, the clustering results change for $\alpha = -0.33$ and $\alpha = 1.88$ according to what is given in Table 11.

Table 11. AHC-neutrosophic results based on sensitive analysis for Case 4.

Cluster	AHC-neutrosophic numbers ($\alpha = 1.88$)	AHC-neutrosophic numbers ($\alpha = -0.33$)
6	$\{\tilde{n}_1\},\{\tilde{n}_2\},\{\tilde{n}_3\},\{\tilde{n}_4\},\{\tilde{n}_5\},\{\tilde{n}_6\}$	$\{\tilde{n}_1\},\{\tilde{n}_2\},\{\tilde{n}_3\},\{\tilde{n}_4\},\{\tilde{n}_5\},\{\tilde{n}_6\}$
5	$\{\tilde{n}_1,\tilde{n}_2\},\{\tilde{n}_3\},\{\tilde{n}_4\},\{\tilde{n}_5\},\{\tilde{n}_6\}$	$\{\tilde{n}_1,\tilde{n}_2\},\{\tilde{n}_3\},\{\tilde{n}_4\},\{\tilde{n}_5\},\{\tilde{n}_6\}$
4	$\{\tilde{n}_1,\tilde{n}_2\},\{\tilde{n}_3\},\{\tilde{n}_4,\tilde{n}_5\},\{\tilde{n}_6\}$	$\{\tilde{n}_1,\tilde{n}_2\},\{\tilde{n}_3,\tilde{n}_4\},\{\tilde{n}_5\},\{\tilde{n}_6\}$
3	$\{\tilde{n}_1,\tilde{n}_2,\tilde{n}_6\},\{\tilde{n}_3\},\{\tilde{n}_4,\tilde{n}_5\}$	$\{\tilde{n}_1,\tilde{n}_2\},\{\tilde{n}_3,\tilde{n}_4,\tilde{n}_5\},\{\tilde{n}_6\}$
2	$\{\tilde{n}_1,\tilde{n}_2,\tilde{n}_6\},\{\tilde{n}_3,\tilde{n}_4,\tilde{n}_5\}$	$\{\tilde{n}_1,\tilde{n}_2,\tilde{n}_6\},\{\tilde{n}_3,\tilde{n}_4,\tilde{n}_5\}$
1	$\{\tilde{n}_1,\tilde{n}_2,\tilde{n}_3,\tilde{n}_4,\tilde{n}_5,\tilde{n}_6\}$	$\{\tilde{n}_1,\tilde{n}_2,\tilde{n}_3,\tilde{n}_4,\tilde{n}_5,\tilde{n}_6\}$

It is clear that for other numbers in the D_N in a similar way, the effect of disturbance in creating different clusters can be obtained.

5. Comparative analyses and advantages

5.1. Comparison

Distance measure as one of the most basic concepts in information theory can play a prominent role in determining the degree of difference and distinction between NTraFNs. Most of the researchers explain and define the distance measure from their theoretical and practical point of view. In the following, we refer to researches that have specifically introduced a new distance measure between NTraFNs. For this purpose, consider two NTraFNs as follows:

$$\tilde{n}_1 = \langle\left(a_{1\tilde{n}_1}, a_{2\tilde{n}_1}, a_{3\tilde{n}_1}, a_{4\tilde{n}_1}\right), \left(b_{1\tilde{n}_1}, b_{2\tilde{n}_1}, b_{3\tilde{n}_1}, b_{4\tilde{n}_1}\right), \left(c_{1\tilde{n}_1}, c_{2\tilde{n}_1}, c_{3\tilde{n}_1}, c_{4\tilde{n}_1}\right)\rangle$$

and

$$\tilde{n}_2 = \langle\left(a_{1\tilde{n}_2}, a_{2\tilde{n}_2}, a_{3\tilde{n}_2}, a_{4\tilde{n}_2}\right), \left(b_{1\tilde{n}_2}, b_{2\tilde{n}_2}, b_{3\tilde{n}_2}, b_{4\tilde{n}_2}\right), \left(c_{1\tilde{n}_2}, c_{2\tilde{n}_2}, c_{3\tilde{n}_2}, c_{4\tilde{n}_2}\right)\rangle$$

Tan et al. (2019) [27] defined the following distance measure to determine the distance between two NTraFNs.

$$Dis_t(\tilde{n}_1,\tilde{n}_2) = \left\{\frac{1}{12}\left(\sum_{i=1}^{4}\left|a_{i\tilde{n}_1} - a_{i\tilde{n}_2}\right|^{\lambda} + \sum_{i=1}^{4}\left|b_{i\tilde{n}_1} - b_{i\tilde{n}_2}\right|^{\lambda} + \sum_{i=1}^{4}\left|c_{i\tilde{n}_1} - c_{i\tilde{n}_2}\right|^{\lambda}\right)\right\}^{1/\lambda}, \lambda \geq 0. \tag{25}$$

If $\boldsymbol{\lambda = 1}$, then equation (25) is reduced to the following Hamming distance:

$$Dis_{tH}(\tilde{n}_1,\tilde{n}_2)=\left\{\frac{1}{12}(\sum_{i=1}^{4}\left|a_{i\tilde{n}_1}-a_{i\tilde{n}_2}\right|+\sum_{i=1}^{4}\left|b_{i\tilde{n}_1}-b_{i\tilde{n}_2}\right|+\sum_{i=1}^{4}\left|c_{i\tilde{n}_1}-c_{i\tilde{n}_2}\right|)\right\}, \tag{26}$$

If $\boldsymbol{\lambda}=\mathbf{2}$, then equation (25) is reduced to the following Euclidean distance:

$$Dis_{tE}(\tilde{n}_1,\tilde{n}_2)=\left\{\frac{1}{12}(\sum_{i=1}^{4}\left|a_{i\tilde{n}_1}-a_{i\tilde{n}_2}\right|^2+\sum_{i=1}^{4}\left|b_{i\tilde{n}_1}-b_{i\tilde{n}_2}\right|^2+\sum_{i=1}^{4}\left|c_{i\tilde{n}_1}-c_{i\tilde{n}_2}\right|^2)\right\}^{1/2}, \tag{27}$$

The performance of the distance measure defined by them in combination with the hierarchical algorithm in some cases shows the equality between the differences, which prevents the effective performance of the method.

Broumi et al. (2020) [28] defined the following distance measure to determine the distance between two NTraFNs.

$$Dis_B(\tilde{n}_1,\tilde{n}_2)$$
$$=\frac{1}{2}\left\{\sqrt{\begin{array}{c}(\sum_{i=1}^{4}a_{i\tilde{n}_1}+\sum_{i=1}^{4}b_{i\tilde{n}_1}+\sum_{i=1}^{4}c_{i\tilde{n}_1}-\sum_{i=1}^{4}a_{i\tilde{n}_2}-\sum_{i=1}^{4}b_{i\tilde{n}_2}-\sum_{i=1}^{4}c_{i\tilde{n}_2}-(\frac{a_{4\tilde{n}_1}a_{3\tilde{n}_1}-a_{1\tilde{n}_1}a_{2\tilde{n}_1}}{(a_{4\tilde{n}_1}+a_{3\tilde{n}_1})-(a_{1\tilde{n}_1}+a_{2\tilde{n}_1})}-\frac{a_{4\tilde{n}_2}a_{3\tilde{n}_2}-a_{1\tilde{n}_2}a_{2\tilde{n}_2}}{(a_{4\tilde{n}_2}+a_{3\tilde{n}_2})-(a_{1\tilde{n}_2}+a_{2\tilde{n}_2})})\\ -(\frac{b_{4\tilde{n}_1}b_{3\tilde{n}_1}-b_{1\tilde{n}_1}b_{2\tilde{n}_1}}{(b_{4\tilde{n}_1}+b_{3\tilde{n}_1})-(b_{1\tilde{n}_1}+b_{2\tilde{n}_1})}-\frac{b_{4\tilde{n}_2}b_{3\tilde{n}_2}-b_{1\tilde{n}_2}b_{2\tilde{n}_2}}{(b_{4\tilde{n}_2}+b_{3\tilde{n}_2})-(b_{1\tilde{n}_2}+b_{2\tilde{n}_2})})-(\frac{c_{4\tilde{n}_1}c_{3\tilde{n}_1}-c_{1\tilde{n}_1}c_{2\tilde{n}_1}}{(c_{4\tilde{n}_1}+c_{3\tilde{n}_1})-(c_{1\tilde{n}_1}+c_{2\tilde{n}_1})}-\frac{c_{4\tilde{n}_2}c_{3\tilde{n}_2}-c_{1\tilde{n}_2}c_{2\tilde{n}_2}}{(c_{4\tilde{n}_2}+c_{3\tilde{n}_2})-(c_{1\tilde{n}_2}+c_{2\tilde{n}_2})}))^2\\ +\left(\left(\frac{a_{4\tilde{n}_1}a_{3\tilde{n}_1}-a_{1\tilde{n}_1}a_{2\tilde{n}_1}}{(a_{4\tilde{n}_1}+a_{3\tilde{n}_1})-(a_{1\tilde{n}_1}+a_{2\tilde{n}_1})}-\frac{a_{4\tilde{n}_2}a_{3\tilde{n}_2}-a_{1\tilde{n}_2}a_{2\tilde{n}_2}}{(a_{4\tilde{n}_2}+a_{3\tilde{n}_2})-(a_{1\tilde{n}_2}+a_{2\tilde{n}_2})}\right)+(\frac{b_{4\tilde{n}_1}b_{3\tilde{n}_1}-b_{1\tilde{n}_1}b_{2\tilde{n}_1}}{(b_{4\tilde{n}_1}+b_{3\tilde{n}_1})-(b_{1\tilde{n}_1}+b_{2\tilde{n}_1})}-\frac{b_{4\tilde{n}_2}b_{3\tilde{n}_2}-b_{1\tilde{n}_2}b_{2\tilde{n}_2}}{(b_{4\tilde{n}_2}+b_{3\tilde{n}_2})-(b_{1\tilde{n}_2}+b_{2\tilde{n}_2})})\right.\\ +(\frac{c_{4\tilde{n}_1}c_{3\tilde{n}_1}-c_{1\tilde{n}_1}c_{2\tilde{n}_1}}{(c_{4\tilde{n}_1}+c_{3\tilde{n}_1})-(c_{1\tilde{n}_1}+c_{2\tilde{n}_1})}-\frac{c_{4\tilde{n}_2}c_{3\tilde{n}_2}-c_{1\tilde{n}_2}c_{2\tilde{n}_2}}{(c_{4\tilde{n}_2}+c_{3\tilde{n}_2})-(c_{1\tilde{n}_2}+c_{2\tilde{n}_2})}))^2\end{array}}\right\}$$

Despite some intuitive advantages, their suggested distance measure does not work properly in some cases. For example, consider the following two NTraFNs:

$$\tilde{n}_1=\langle(0,0,0,0),(0.1,0.2,0.3,0.4),(0.3,0.4,0.5,0.6)\rangle$$

and

$$\tilde{n}_2=\langle(0.1,0.3,0.5,0.7),(0.2,0.4,0.6,0.8),(0.1,0.2,0.3,0.4)\rangle$$

In this case, the distance measure between $\tilde{\boldsymbol{n}}_1$ and $\tilde{\boldsymbol{n}}_2$ cannot be defined based on the $\boldsymbol{Dis_B}$.

Now we want to evaluate the performance of the hierarchical clustering method based on the available measures. The results of clustering based on different measure are given in Table 12.

Table 12. Comparison between AHC-neutrosophic numbers based on proposed distance and existing distance measures.

Cluster	$Dis_{tH}(\tilde{n}_1,\tilde{n}_2)$ and $Dis_{tE}(\tilde{n}_1,\tilde{n}_2)$	Proposed distance measure, $Dis_B(\tilde{n}_1,\tilde{n}_2)$ and combination of distance measures
6	$\{\tilde{n}_1\},\{\tilde{n}_2\},\{\tilde{n}_3\},\{\tilde{n}_4\},\{\tilde{n}_5\},\{\tilde{n}_6\}$	$\{\tilde{n}_1\},\{\tilde{n}_2\},\{\tilde{n}_3\},\{\tilde{n}_4\},\{\tilde{n}_5\},\{\tilde{n}_6\}$
5	$\{\tilde{n}_1\},\{\tilde{n}_2\},\{\tilde{n}_3\},\{\tilde{n}_4,\tilde{n}_5\},,\{\tilde{n}_6\}$	$\{\tilde{n}_1,\tilde{n}_2\},\{\tilde{n}_3\},\{\tilde{n}_4\},\{\tilde{n}_5\},\{\tilde{n}_6\}$
4	$\{\tilde{n}_1,\tilde{n}_2\},\{\tilde{n}_3\},\{\tilde{n}_4,\tilde{n}_5\},\{\tilde{n}_6\}$	$\{\tilde{n}_1,\tilde{n}_2\},\{\tilde{n}_3\},\{\tilde{n}_4,\tilde{n}_5\},\{\tilde{n}_6\}$
3	$\{\tilde{n}_1,\tilde{n}_2\},\{\tilde{n}_3,\tilde{n}_4,\tilde{n}_5\},\{\tilde{n}_6\}$	$\{\tilde{n}_1,\tilde{n}_2\},\{\tilde{n}_3,\tilde{n}_4,\tilde{n}_5\},\{\tilde{n}_6\}$
2	$\{\tilde{n}_1,\tilde{n}_2,\tilde{n}_6\},\{\tilde{n}_3,\tilde{n}_4,\tilde{n}_5\}$	$\{\tilde{n}_1,\tilde{n}_2,\tilde{n}_6\},\{\tilde{n}_3,\tilde{n}_4,\tilde{n}_5\}$
1	$\{\tilde{n}_1,\tilde{n}_2,\tilde{n}_3,\tilde{n}_4,\tilde{n}_5,\tilde{n}_6\}$	$\{\tilde{n}_1,\tilde{n}_2,\tilde{n}_3,\tilde{n}_4,\tilde{n}_5,\tilde{n}_6\}$

From the information in Table 12, it can be concluded that the hierarchical clustering method based on the $Dis_B(\tilde{n}_1, \tilde{n}_2)$ and combination of distance measures has the same output as the proposed method in subsection 4.2. In addition, the results of clustering based on the $Dis_{tH}(\tilde{n}_1, \tilde{n}_2)$ and $Dis_{tE}(\tilde{n}_1, \tilde{n}_2)$ are different from other approaches only in cluster 5.

5.2. Managerial implications and advantages

Clustering serves as a conduit for encapsulating latent knowledge within the final categories. Each cluster's output embodies characteristics closely intertwined within the cluster's data, offering invaluable insights that can inform organizational actions such as marketing strategies, customer relationship enhancement, resource allocation, and investment decisions. This act of data labeling facilitates precise managerial decisions by elucidating internal data patterns. By delving into the intricate relationships within the data, managers gain access to more reasoned options, thereby enhancing organizational performance. Additionally, decision-makers can harness the strengths of various distance metrics simultaneously by amalgamating the output of the distance matrix, thus yielding more accurate results. For instance, in the context of customer data, appropriate clustering enables organizations to devise effective promotional and incentive strategies tailored to each cluster, fostering customer retention, satisfaction, and acquisition.

Among the myriad management advantages offered by hierarchical clustering methods, one standout feature is the decision-maker's autonomy in determining the final number of clusters. This flexibility allows for adjustments in resource allocation based on fluctuating resource availability (be it human, financial, or temporal), enabling organizations to tailor their goals accordingly.

As previously highlighted, existing distance metrics may not fully capture all the nuances inherent in NTraFNs. Hence, in Section 3, we introduce a novel distance measure based on vertical axis levels. Some key advantages of this distance measure include:

- Intuitive and logical interpretation of NTraFN dimensions.
- Theorems and properties outlined in subsection 3.2 affirm the accuracy of the proposed distance measure.
- Simplicity in calculations facilitates its applicability in diverse distance- or similarity-based problems.
- Combined with existing distance measures, the proposed measure yields consistent clustering outputs, offering decision-makers confidence in their analytical endeavors.

6. Conclusions

In this study, we introduced a novel and efficacious conceptual distance measure for comparing NTraFNs. The validity of the proposed distance measure structure was substantiated through the verification of its underlying principles and logical properties. Furthermore, the performance of the clustering algorithm, predicated on this distance measure, underscores the practical utility of our research. Among the limitations encountered in this study is the paucity of a comprehensive database of NTraFNs, which hampers our ability to tackle problems with larger dimensions. Nonetheless, the dearth of research pertaining to data mining under NTraFNs, coupled with the significance of information theory, underscores the importance and continued exploration of this field. Several avenues for future research warrant consideration:

- Developing similarity measures and entropy metrics based on the proposed distance measure between NTraFNs.
- Generalizing the conceptual framework to encompass other distance, similarity, and entropy measures.
- Integrating the suggested distance measure with alternative clustering methodologies.
- Employing the proposed distance measure to address challenges associated with larger-dimensional problems and other distance-based predicaments.
- These avenues not only hold promise for advancing our understanding of NTraFNs but also offer opportunities to enhance the efficacy and versatility of clustering methodologies in diverse problem domains.

References

1. Gan, G.; Ma, C.; Wu, J. *Data clustering: theory, algorithms, and applications*, 2nd ed.; Society for Industrial and Applied Mathematics: 3600 Market Street, 6th Floor Philadelphia, USA, 2020.
2. Zadeh, L.A. Fuzzy sets. *Information and control* **1965**, *8*, 338-353.
3. Smarandache, RF. A unifying field in logics. In: Neutrosophy. Neutrosophic probability, set and logic. *American Research Press, Rehoboth* **1999**, 7-8.
4. Ye, J.; Smarandache, F. Similarity measure of refined single-valued neutrosophic sets and its multicriteria decision making method. *Neutrosophic Sets and Systems* **2016**, *12*, 41–44.
5. Khalifa, H.A.E.W. and Kumar, P., A novel method for neutrosophic assignment problem by using interval-valued trapezoidal neutrosophic number (Vol. 36). *Infinite Study* **2020**.
6. Zulqarnain, R. M.; Xin, X. L.; Saeed, M.; Smarandache, F.; Ahmad, N. Generalized neutrosophic TOPSIS to solve multi-criteria decision-making problems. *Neutrosophic Sets and Systems* **2020**, *38*, 276-292.
7. Haq, A.; Gupta, S.; Ahmed, A. A multi-criteria fuzzy neutrosophic decision-making model for solving the supply chain network problem. *Neutrosophic Sets and Systems* **2021**, *46*, 50-66.
8. Poonia, M.; Bajaj, R.K. On Measures of Similarity for Neutrosophic Sets with Applications in Classification and Evaluation Processes. *Neutrosophic Sets and* Syst*ems* **2021**, *39*, 86-100.
9. Patro, Santanu & Said, Broumi & Smarandache, Florentin & Talea, Mohamed & Bakali, Assia. (2017). Minimum Spanning tree problem with single valued Trapezoidal neutrosophic number. IEEE Technically Sponsored Computing Conference 2018. At: London, UK.
10. Broumi, S.; Dhar, M.; Bakhouyi, A.; Bakali, A.; Talea, M. Medical diagnosis problems based on neutrosophic sets and their hybrid structures: A survey. *Neutrosophic Sets and Systems* **2022**, *49*, 1-18.
11. Jdid, Maissam & Smarandache, Florentin. (2023). Optimal Agricultural Land Use: An Efficient Neutrosophic Linear Programming Method. *Neutrosophic Systems with Applications.* 10. 53-59. 10.61356/j.nswa.2023.76.
12. Broumi, S.; Prabha, S.; Krishna.; Uluçay, Vakkas. (2023). Interval-Valued Fermatean Neutrosophic Shortest Path Problem via Score Function. *Neutrosophic Systems with Applications*. 11. 1-10. 10.61356/j.nswa.2023.83.
13. Rosli, Siti Nur & Zulkifly, Mohammad Izat Emir. (2023). 3-Dimensional Quartic Bézier Curve Approximation Model by Using Neutrosophic Approach. *Neutrosophic Systems with Applications.* 11. 11-21. 10.61356/j.nswa.2023.78.
14. Ye, J. Clustering methods using distance-based similarity measures of single-valued neutrosophic sets. *Journal of Intelligent Systems* 2014, 23, 379-389.
15. Long, H.V.; Ali, M.; Khan, M.; Tu, D.N. A novel approach for fuzzy clustering based on neutrosophic association matrix. *Computers and Industrial Engineering* **2019**, *127*, 687-697.
16. Zhang, D.; Ma, Y.; Dai, X.; Qiao, Y. Clustering algorithm based on data indeterminacy in neutrosophic set. *Neutrosophic Sets and Systems* **2022**, *51*, 556-569.
17. Johnson, S.C.; Hierarchical clustering schemes, *Psychometrika* **1967**, *32*, 241–254.
18. Geva, A.B. Hierarchical unsupervised fuzzy clustering, *IEEE Transactions on Fuzzy Systems* **1999**, *7*, 723-733.
19. Ghasemigol, M.; Sadoghi Yazdi, H.; Monsefi, R. A new hierarchical clustering algorithm on fuzzy data (FHCA). *International Journal of Computer and Electrical Engineering* **2010**, *2*, 134-140.

20. Vandhana, S.; Anuradha, J. Neutrosophic fuzzy hierarchical clustering for dengue analysis in Sri Lanka. *Neutrosophic Sets and Systems* **2020,** *31*, 179-199.
21. Elhassouny, A. Neutrosophic Logic-based DIANA Clustering algorithm. *Neutrosophic Sets and Systems* 2023, 55, 498-509.
22. Biswas, P.; Pramanik, S.; Giri, B.C. A new methodology for neutrosophic multi-attribute decision making with unknown weight information. *Neutrosophic Sets and Systems* **2014**, *3*, 42–52.
23. Abbasi Shureshjani. R.; Darehmiraki, M. A new parametric method for ranking fuzzy numbers. *Indagationes Mathematicae* **2013**, *24*, 518–529.
24. Laros, D. *Discovering knowledge in data: An introduction to data mining*, 2nd ed.; Wiley, New York, America, 2005.
25. Charrad, M.; Ghazzali, N.; Boiteau, V.; Niknafs, A. NbClust: An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software* **2014**, 61, 1-36.
26. Kaufman, L.; Rousseeuw, P. *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, New York, America, 1990.
27. Tan, Ruipu & Zhang, Wende & Chen, Shengqun. (2019). Decision-Making Method Based on Grey Relation Analysis and Trapezoidal Fuzzy Neutrosophic Numbers under Double Incomplete Information and Its Application in Typhoon Disaster Assessment. *IEEE Access.* PP. 1-1. 10.1109/ACCESS.2019.2962330.
28. Broumi, Said; Malayalan Lathamaheswari; Ruipu Tan; Deivanayagampillai Nagarajan; Talea Mohamed; Florentin Smarandache; and Assia Bakali. "A new distance measure for trapezoidal fuzzy neutrosophic numbers based on the centroids." *Neutrosophic Sets and Systems* 35, 1 (2020). https://digitalrepository.unm.edu/nss_journal/vol35/iss1/27

Received: April 22, 2025. Accepted: Sep 30, 2025