Single valued neutrosophic minimum spanning tree and its clustering method

Jun Ye*

Department of Electrical and Information Engineering, Shaoxing University, 508 Huancheng West Road, Shaoxing, Zhejiang Province 312000, P.R. China

Abstract

Clustering plays an important role in data mining, pattern recognition, and machine learning. Then, single valued neutrosophic sets (SVNSs) are useful means to describe and handle indeterminate and inconsistent information, which fuzzy sets and intuitionistic fuzzy sets cannot describe and deal with. To cluster the data represented by single value neutrosophic information, the paper proposes a single valued neutrosophic minimum spanning tree (SVNMST) clustering algorithm. Firstly, we defined a generalized distance measure between SVNSs. Then, we present a SVNMST clustering algorithm for clustering single value neutrosophic data based on the generalized distance measure of SVNSs. Finally, two illustrative examples are given to demonstrate the application and effectiveness of the developed approach.

Keywords: Neutrosophic set; Single valued neutrosophic set; Minimum spanning tree; Clustering algorithm; Generalized distance measure

* Tel.: +86-575-88327323
E-mail address: yehjun@aliyun.com (Jun Ye)

1. Introduction

Clustering plays an important role in data mining, pattern recognition, and machine learning. Clustering data sets into disjoint groups is a problem arising in many domains. Generally, the goal of clustering is to find groups that are both homogeneous and well separated, that is, entities within the same group should be similar and entities in different groups dissimilar. Then, data sets can be represented by weighted graphs, where nodes correspond to the entities to be clustered and edges correspond to a dissimilarity or similarity measure between those entities. Graph theory [1] provided us with a very convenient tool to describe clustering problems. However, a minimum spanning tree (MST) is a very useful graph structure and can capture perceptual grouping [2]. Zahn [3] defined several criteria of edge inconsistency for detecting clusters of different shapes and proposed the clustering algorithm using MST. Xu et al. [4] introduced three MST algorithms and applied them to clustering gene expression data. Due to the fuzziness and uncertainty of many practical problems in the real world, Ruspini [5] first presented the concept of fuzzy division and a fuzzy clustering approach. Dong et al. [6] introduced a hierarchical clustering algorithm based on fuzzy graph connectedness. Then, Chen et al. [7] put forward a fuzzy graph maximal tree clustering method of the fuzzy graph constructing the fuzzy similarity relation matrix and used the threshold of fuzzy similarity relation matrix to cut maximum spanning tree, and then obtained the classification on level. Zhao et al. [8] proposed two intuitionistic fuzzy MST (IFMST) clustering algorithms to deal with intuitionistic fuzzy information and extended them to clustering interval-valued intuitionistic fuzzy information. Furthermore, Zhang and Xu [9] introduced a MST algorithm-based clustering method under hesitant fuzzy environment.

To represent uncertainty, imprecise, incomplete, and inconsistent information which exist in real world, Smarandache [10] gave the concept of a neutrosophic set from philosophical point of view. The neutrosophic set [10] is a powerful general formal framework which generalizes the concept of the classic set, fuzzy set, interval valued fuzzy set, intuitionistic fuzzy set, interval valued intuitionistic fuzzy set, paraconsistent set, dialetheist set, paradoxist set, and tautological set. In the neutrosophic set, truth-membership, indeterminacy-membership, and falsity-membership are represented independently. However, the neutrosophic set generalizes the above mentioned sets from philosophical point of view and its functions $T_A(x)$, $I_A(x)$ and $F_A(x)$ are real standard or nonstandard subsets of $]^-0, 1^+[$, i.e., $T_A(x): X \to]^-0, 1^+[$, $I_A(x): X \to]^-0, 1^+[$, and $F_A(x): X \to]^-0, 1^+[$. So it will be difficult to apply in real scientific and engineering areas. Thus, Wang et al. [11] introduced a single valued neutrosophic set (SVNS), which is an instance of a neutrosophic set. It can describe and handle indeterminate information and inconsistent information, which fuzzy sets and intuitionistic fuzzy sets (IFSs) cannot describe and deal with. Recently, Ye [12, 13] presented the correlation coefficient of SVNSs and the cross-entropy measure of SVNSs and applied them to single valued neutrosophic decision-making problems. Yet, until now there have been no study on clustering the data represented by single valued neutrosophic information. However, the existing MST clustering algorithms cannot cluster the single valued neutrosophic data. Therefore, this paper proposes a single valued neutrosophic MST (SVNMST) clustering algorithm to deal with the data represented by SVNSs. To do so, the rest of the article is organized as follows. Section 2 introduces some basic concepts of SVNSs, the graph and its MST. Section 3 defines a generalized distance measure between SVNSs. In Section 4, a single valued neutrosophic clustering algorithm is proposed based on the MST. In Section 5, two illustrative examples are given to demonstrate the applications and the effectiveness of the proposed approach. Conclusions and further research are contained in Section 6.

2. Preliminaries

In this section, some basic concepts of SVNSs, the graph and its MST are introduced to be utilized in the next sections.

2.1. Some concepts of SVNSs

Smarandache [10] introduced the concept of a neutrosophic set from philosophical point of view.

Definition 1 [10]. Let X be a space of points (objects), with a generic element in X denoted by x. A neutrosophic set A in X is characterized by a truth-membership function $T_A(x)$, an indeterminacy-membership function $I_A(x)$, and a falsity-membership function $F_A(x)$. The functions $T_A(x)$, $I_A(x)$ and $F_A(x)$ are real standard or nonstandard subsets of $]^-0, 1^+[$. That is $T_A(x): X \rightarrow\ ]^-0, 1^+[$, $I_A(x): X \rightarrow\ ]^-0, 1^+[$, and $F_A(x): X \rightarrow\ ]^-0, 1^+[$, with the condition $^-0 \leq \sup T_A(x) + \sup I_A(x) + \sup F_A(x) \leq 3^+$

Obviously, it is difficult to apply in real scientific and engineering fields. Thus, Wang et al. [11] introduced the concept of a SVNS, which is an instance of a neutrosophic set.

Definition 2 [11]. Let X be a space of points (objects) with generic elements in X denoted by x. A SVNS A in X is characterized by truth-membership function $T_A(x)$, indeterminacy-membership function $I_A(x)$, and falsity-membership function $F_A(x)$. For each point x in X, there are $T_A(x)$, $I_A(x)$, $F_A(x) \in [0, 1]$ and $0 \leq T_A(x) + I_A(x) + F_A(x) \leq 3$. Therefore, a SVNS A can be represented by

$$A = \left\{ \langle x, T_A(x), I_A(x), F_A(x) \rangle \mid x \in X \right\}.$$

The following expressions are defined in [11] for SVNSs A, B:

(1) $A \subseteq B$ if and only if $T_A(x) \leq T_B(x)$, $I_A(x) \geq I_B(x)$, $F_A(x) \geq F_B(x)$ for any x in X,

(2) $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$,

(3) $A^c = \{\langle x, F_A(x), 1 - I_A(x), T_A(x) \rangle \mid x \in X\}$.

2.2. Graph and minimum spanning trees

A graph G is a pair of sets $G = (V, E)$, where V is the set of vertices (or nodes) and E is the set of edges. There are two kinds of graphs: an undirected graph and a directed graph. Each edge in the undirected graph is an unordered pair $\{v_i, v_j\}$, while each edge in the directed graph is an ordered pair $\{v_i, v_j\}$, where the vertices v_i and v_j are called the endpoints of an edge. A sequence of edges and vertices that can be traveled between two different vertices is called a path. Suppose 5 nodes are given, then a graph with 5 nodes and 7 edges is shown in Fig. 1.

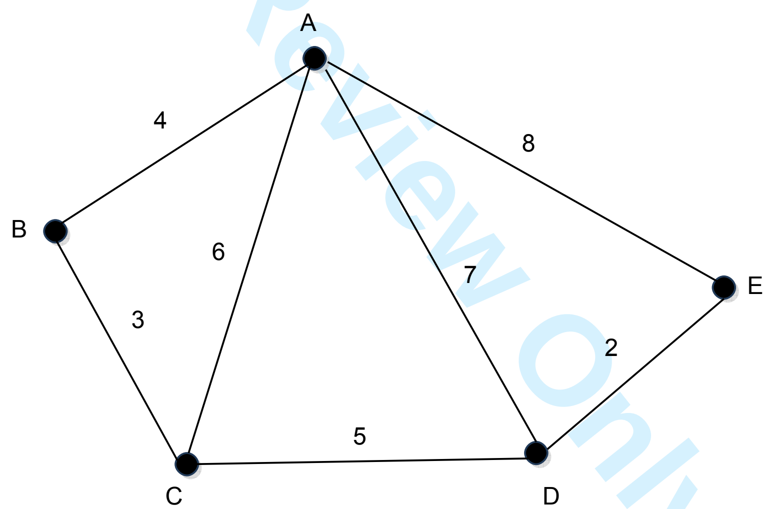

Fig. 1 Graph

In Fig. 1, there are different paths from the node B to the node E, such as the path (BAE) and the other path (BCDE). For a path, the start node and destination node is the same, which is called a circuit, like (ABCA) or (ADEA).

A connected graph has paths between any pair of nodes. A connected acyclic graph that contains all nodes of G is called a spanning tree of the graph, which is any set of straight line segments connecting pairs of nodes such that:

(1) no closed loops occur;

(2) each node is visited by at least one line;

(3) a tree is connected.

For example, a spanning tree with integer segment lengths is shown in Fig. 2, which is obtained from Fig. 1. When the nodes *B* and *C* or *D* and *E* were connected, a closed loop would be formed and the resulting figure would not be a tree. Then, we define the length of a tree to be the sum of the lengths of its constituent edges, the length of the tree in Fig. 2 is 4 + 6 + 5 + 8 = 23 units.

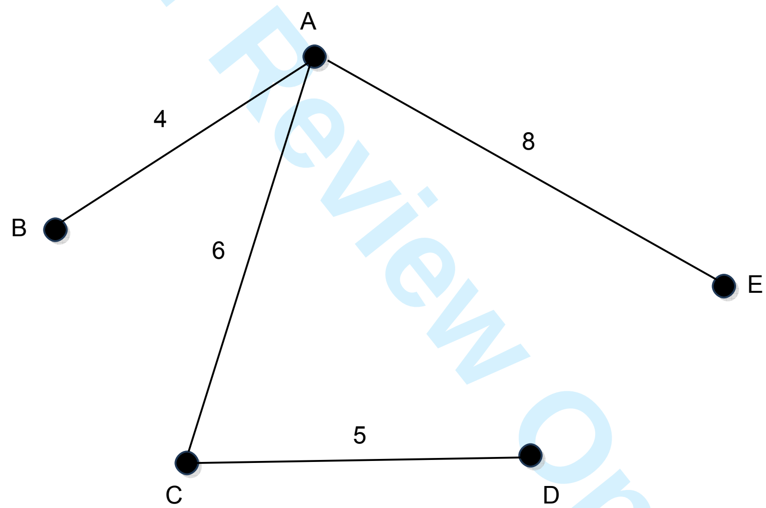

Fig. 2 Spanning tree

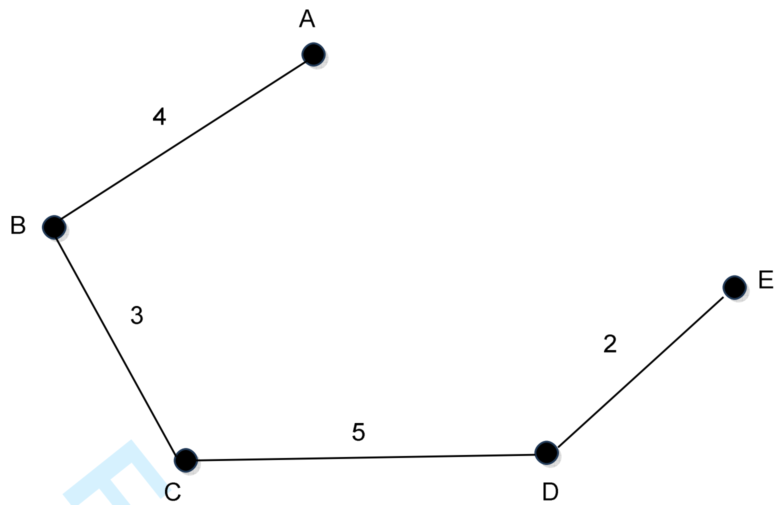

Fig. 3 Minimum spanning tree

A MST is the spanning tree of the minimum length which is often required. For instance, Fig. 3 shows the MST obtained from Fig. 1, and then its minimum length is 4 + 3 + 2 + 5 = 14 units.

Various algorithms to find the MST have been proposed in [4, 14, 15]. However, there are the two most popular algorithms to find the MST of a graph *G* by operating iteratively [14, 15], which are needed later. At any stage the segment belong to one of two sets, i.e., the set *A* containing those segments assigned to the MST and the set *B*, those not assigned.

The algorithm given by Kruskal [14] is to assign interactively to the set *A* the shortest segment in the set *B* which does not form a closed loop with any of the segments already in *A*. Initially *A* is empty and the iteration stops when *A* contains (n-1) segments.

The algorithm given by Prim [15] starts with any one of the given nodes and initially assign to *A* the shortest segment starting from this node. The procedure continues to add the shortest segment from *B* which connects to at least one segment from *A* without forming a closed loop amongst the segments already in *A*. The iteration stops when *A* contains (n-1) segments.

Usually, clustering data sets can be represented as weighted graphs, where nodes correspond to the entities to be clustered and edges correspond to distance measure (or called dissimilarity measure) between those entities. If a fuzzy relation R over $V\times V$ is defined, then the membership function $\mu_R(v_1, v_2)$, where $(v_1, v_2) \in V\times V$, takes various values from 0 to 1, and then such a graph is called a fuzzy graph. When R is an intuitionistic fuzzy relation over $V\times V$, then such a graph is called an intuitionistic fuzzy graph [8].

3. Distance measures of SVNSs

For two SVNSs A and B in a universe of discourse $X = \{x_1, x_2, \ldots, x_n\}$, which are denoted by $A = \{\langle x_i, T_A(x_i), I_A(x_i), F_A(x_i)\rangle \mid x_i \in X\}$ and $B = \{\langle x_i, T_B(x_i), I_B(x_i), F_B(x_i)\rangle \mid x_i \in X\}$, where $T_A(x_i)$, $I_A(x_i)$, $F_A(x_i)$, $T_B(x_i)$, $I_B(x_i)$, $F_B(x_i) \in [0, 1]$ for every $x_i \in X$. Let us to consider the weight w_i ($i = 1, 2, \ldots, n$) of an element x_i ($i = 1, 2, \ldots, n$), with $w_i \geq 0$ ($i = 1, 2, \ldots, n$) and $\sum_{i=1}^{n} w_i = 1$. Then, we define the generalized single valued neutrosophic weighted distance:

$$d_\lambda(A,B) = \left\{\frac{1}{3}\sum_{i=1}^{n} w_i\left[\left|T_A(x_i) - T_B(x_i)\right|^\lambda + \left|I_A(x_i) - I_B(x_i)\right|^\lambda + \left|F_A(x_i) - F_B(x_i)\right|^\lambda\right]\right\}^{1/\lambda}, \quad (1)$$

where $\lambda > 0$.

Especially, if $\lambda = 1, 2$, Eq. (1) reduces to the single valued neutrosophic weighted Hamming distance and the single valued neutrosophic weighted Euclidean distance, respectively, as follows:

$$d_1(A,B) = \frac{1}{3}\sum_{i=1}^{n} w_i\left[\left|T_A(x_i) - T_B(x_i)\right| + \left|I_A(x_i) - I_B(x_i)\right| + \left|F_A(x_i) - F_B(x_i)\right|\right], \quad (2)$$

$$d_2(A,B) = \left\{\frac{1}{3}\sum_{i=1}^{n} w_i\left[\left|T_A(x_i) - T_B(x_i)\right|^2 + \left|I_A(x_i) - I_B(x_i)\right|^2 + \left|F_A(x_i) - F_B(x_i)\right|^2\right]\right\}^{1/2}. \quad (3)$$

Then, for the distance measure we have the following proposition.

Proposition 1. The above defined distance $d_\lambda(A, B)$ for $\lambda > 0$ satisfies the following properties:

(P1) $d_\lambda(A, B) \geq 0$;

(P2) $d_\lambda(A, B) = 0$ if and only if $A = B$;

(P3) $d_\lambda(A, B) = d_\lambda(B, A)$;

(P4) If $A \subseteq B \subseteq C$, C is a SVNS in X, then $d_\lambda(A, C) \geq d_\lambda(A, B)$ and $d_\lambda(A, C) \geq d_\lambda(B, C)$.

Proof. It is easy to see that $d_\lambda(A, B)$ satisfies the properties (P1)–(P3). Therefore, we only prove (P4). Let $A \subseteq B \subseteq C$, then, $T_A(x_i) \leq T_B(x_i) \leq T_C(x_i)$, $I_A(x_i) \geq I_B(x_i) \geq I_C(x_i)$, and $F_A(x_i) \geq F_B(x_i) \geq F_C(x_i)$ for every $x_i \in X$. Then, we obtain the following relations:

$$\left|T_A(x_i) - T_B(x_i)\right|^\lambda \leq \left|T_A(x_i) - T_C(x_i)\right|^\lambda, \quad \left|T_B(x_i) - T_C(x_i)\right|^\lambda \leq \left|T_A(x_i) - T_C(x_i)\right|^\lambda,$$

$$\left|I_A(x_i) - I_B(x_i)\right|^\lambda \leq \left|I_A(x_i) - I_C(x_i)\right|^\lambda, \quad \left|I_B(x_i) - I_C(x_i)\right|^\lambda \leq \left|I_A(x_i) - I_C(x_i)\right|^\lambda,$$

$$\left|F_A(x_i) - F_B(x_i)\right|^\lambda \leq \left|F_A(x_i) - F_C(x_i)\right|^\lambda, \quad \left|F_B(x_i) - F_C(x_i)\right|^\lambda \leq \left|F_A(x_i) - F_C(x_i)\right|^\lambda.$$

Hence

$$\left|T_A(x_i)-T_B(x_i)\right|^\lambda + \left|I_A(x_i)-I_B(x_i)\right|^\lambda + \left|F_A(x_i)-F_B(x_i)\right|^\lambda \leq \left|T_A(x_i)-T_C(x_i)\right|^\lambda + \left|I_A(x_i)-I_C(x_i)\right|^\lambda + \left|F_A(x_i)-F_C(x_i)\right|^\lambda,$$

$$\left|T_B(x_i)-T_C(x_i)\right|^\lambda + \left|I_B(x_i)-I_C(x_i)\right|^\lambda + \left|F_B(x_i)-F_C(x_i)\right|^\lambda \leq \left|T_A(x_i)-T_C(x_i)\right|^\lambda + \left|I_A(x_i)-I_C(x_i)\right|^\lambda + \left|F_A(x_i)-F_C(x_i)\right|^\lambda.$$

Combining the above inequalities with the above defined distance formula (1), we can obtain that

$d_\lambda(A, B) \leq d_\lambda(A, C)$ and $d_\lambda(B, C) \leq d_\lambda(A, C)$ for $\lambda > 0$.

Thus the property (P4) is obtained. □

In the following, we give the definition of single valued neutrosophic distance matrix.

Definition 3. Let A_j ($j = 1, 2, \ldots, m$) be a collection of m SVNSs, then $D = (d_{ij})_{m\times m}$ is called a single valued neutrosophic distance matrix, where $d_{ij} = d_\lambda(A_i, A_j)$ is the distance between A_i and A_j and its properties are as follows:

(1) $0 \leq d_{ij} \leq 1$ for all $i, j = 1, 2, \ldots, m$;

(2) $d_{ij} = 0$ if and only if $A_i = A_j$;

(3) $d_{ij} = d_{ji}$ for all $i, j = 1, 2, \ldots, m$.

4. SVNMST clustering algorithm

In this section, a SVNMST clustering algorithm is proposed as a generalization of an IFMST clustering algorithm [8].

Let $X = \{x_1, x_2, \ldots, x_n\}$ be an attribution space and the weight vector of an element x_i ($i = 1, 2, \ldots, n$) be $w = \{w_1, w_2, \ldots, w_n\}$ with $w_i \geq 0$ ($i = 1, 2, \ldots, n$) and $\sum_{i=1}^{n} w_i = 1$. Assume that A_j ($j = 1, 2, \ldots, m$) is a collection of m SVNSs, which express m samples to be clustered. Then there is the following form:

$$A_j = \left\{ \left\langle x_i, T_{A_j}(x_i), I_{A_j}(x_i), F_{A_j}(x_i) \right\rangle \mid x_i \in X \right\}.$$

Thus, we propose a SVNMST clustering algorithm, which is described by the following steps:

Step 1. Calculate the distance of $d_{ij} = d_\lambda(A_i, A_j)$ ($i, j = 1, 2, \ldots, m$) by Eq. (1) to establish the single valued neutrosophic distance matrix $D = (d_{ij})_{m \times m}$.

Step 2. Draw the single valued neutrosophic graph $G(V, E)$ where every edge between A_i and A_j is assigned the weight (single valued neutrosophic distance) d_{ij} coming from an element of the single valued neutrosophic distance matrix $D = (d_{ij})_{m \times m}$, which represents the dissimilarity degree between the samples A_i and A_j.

Step 3. Build the MST of the single valued neutrosophic graph $G(V, E)$ by Kruskal's method [14] or Prim's method [15]:

(1) Sort the edges of G in increasing order by weights.

(2) Keep a sub-graph S of G, which is initially empty, and at each step choose the edge e with the smallest weight to be added to the sub-graph S, where the endpoint of e are disconnected.

(3) Repeat the process (2) until the sub-graph S spans all vertices. Hence, the MST of the single valued neutrosophic graph $G(V, E)$ is obtained.

Step 4. Perform clustering by use of the SVNMST. We can get a certain number of sub-trees (clustering) by disconnecting all the edges of the MST with weights greater than a threshold r. The clustering results induced by the sub-trees do not depend on some particular MST [8, 9].

5. Illustrative examples

In this section, two illustrative examples are presented to demonstrate the real applications and the effectiveness of the proposed approach.

Example 1. A car market is going to classify eight different cars of A_j (j = 1, 2, …, 8). Every car has six evaluation factors (attributes): (1) x_1: fuel consumption; (2) x_2: coefficient of friction; (3) x_3: price; (4) x_4: comfortable degree; (5) x_5: design; (6) x_6: security coefficient. The characteristics of each car under the six attributes are represented by the form of SVNSs, and then the single valued neutrosophic data are as follows:

A_1 = {< x_1, 0.3, 0.2, 0.5>, < x_2, 0.6, 0.3, 0.1>, < x_3, 0.4, 0.3, 0.3>, < x_4, 0.8, 0.1, 0.1>, < x_5, 0.1, 0.3, 0.6>, < x_6, 0.5, 0.2, 0.4>},

A_2 = {< x_1, 0.6, 0.3, 0.3>, < x_2, 0.5, 0.4, 0.2>, < x_3, 0.6, 0.2, 0.1>, < x_4, 0.7, 0.2, 0.1>, < x_5, 0.3, 0.1, 0.6>, < x_6, 0.4, 0.3, 0.3>},

A_3 = {< x_1, 0.4, 0.2, 0.4>, < x_2, 0.8, 0.2, 0.1>, < x_3, 0.5, 0.3, 0.1>, < x_4, 0.6, 0.1, 0.2>, < x_5, 0.4, 0.1, 0.5>, < x_6, 0.3, 0.2, 0.2>},

A_4 = {< x_1, 0.2, 0.4, 0.4>, < x_2, 0.4, 0.5, 0.1>, < x_3, 0.9, 0.2, 0.0>, < x_4, 0.8, 0.2, 0.1>, < x_5, 0.2, 0.3, 0.5>, < x_6, 0.7, 0.3, 0.1>},

A_5 = {< x_1, 0.2, 0.3, 0.3>, < x_2, 0.3, 0.2, 0.6>, < x_3, 0.5, 0.1, 0.4>, < x_4, 0.7, 0.1, 0.1>, < x_5, 0.4, 0.2, 0.4>, < x_6, 0.3, 0.2, 0.6>},

A_6 = {< x_1, 0.3, 0.2, 0.4>, < x_2, 0.2, 0.1, 0.7>, < x_3, 0.4, 0.2, 0.5>, < x_4, 0.8, 0.0, 0.1>, < x_5, 0.4, 0.3, 0.5>, < x_6, 0.2, 0.1, 0.7>},

A_7 = {< x_1, 0.4, 0.4, 0.3>, < x_2, 0.5, 0.3, 0.1>, < x_3, 0.6, 0.1, 0.2>, < x_4, 0.2, 0.3, 0.7>, < x_5, 0.3, 0.1, 0.5>, < x_6, 0.7, 0.2, 0.1>},

A_8 = {< x_1, 0.4, 0.1, 0.2>, < x_2, 0.6, 0.1, 0.1>, < x_3, 0.8, 0.2, 0.1>, < x_4, 0.7, 0.2, 0.1>, < x_5, 0.1, 0.1, 0.8>, < x_6, 0.2, 0.1, 0.8>}.

If the weight vector of the attribute x_i (i = 1, 2, …, 6) is w = (0.16, 0.12, 0.25, 0.2, 0.15, 0.12)^T. Then we utilize the SVNMST clustering algorithm to group the eight different cars of A_j (j =1, 2, …, 8):

Step 1. Calculate the distance $d_{ij} = d_\lambda(A_i, A_j)$ by Eq. (1) (take $\lambda = 2$). Then we can establish the single valued neutrosophic distance matrix $D = (d_{ij})_{m\times m}$ as follows:

$$D = \begin{bmatrix} 0.0000 & 0.2737 & 0.2551 & 0.3583 & 0.4022 & 0.4559 & 0.5052 & 0.4100 \\ 0.2737 & 0.0000 & 0.2269 & 0.2605 & 0.4064 & 0.5334 & 0.4271 & 0.4117 \\ 0.2551 & 0.2269 & 0.0000 & 0.3311 & 0.4577 & 0.5527 & 0.3983 & 0.4450 \\ 0.3583 & 0.2605 & 0.3311 & 0.0000 & 0.5442 & 0.6720 & 0.4559 & 0.5794 \\ 0.4022 & 0.4064 & 0.4577 & 0.5442 & 0.0000 & 0.1987 & 0.5957 & 0.4811 \\ 0.4559 & 0.5334 & 0.5527 & 0.6720 & 0.1987 & 0.0000 & 0.7155 & 0.5248 \\ 0.5052 & 0.4271 & 0.3983 & 0.4559 & 0.5957 & 0.7155 & 0.0000 & 0.6370 \\ 0.4100 & 0.4117 & 0.4450 & 0.5794 & 0.4811 & 0.5248 & 0.6370 & 0.0000 \end{bmatrix}.$$

Step 2. Draw the single valued neutrosophic graph $G(V, E)$ where every edge between A_i and A_j (i, j = 1, 2, …, 8) is assigned the weight (single valued neutrosophic distance) d_{ij} coming from an element of the single valued neutrosophic distance matrix $D = (d_{ij})_{m\times m}$, which represents the dissimilarity degree between the samples A_i and A_j. Then, the single valued neutrosophic graph $G(V, E)$ is shown in Fig. 4.

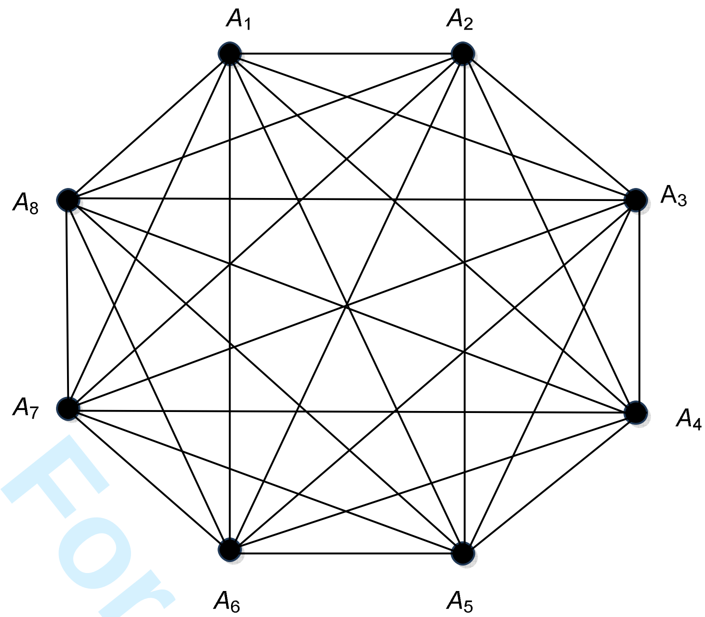

Fig. 4 Single valued neutrosophic graph $G(V, E)$ with the eight nodes

Step 3. Establish the MST of the single valued neutrosophic graph $G(V, E)$ by Kruskal's method [14] or Prim's method [15]:

(1) Sort the edges of G in increasing order by weights:

$d_{56} < d_{23} < d_{13} < d_{24} < d_{12} < d_{34} < d_{14} < d_{37} < d_{15} < d_{25} < d_{18} < d_{28} < d_{27} < d_{38} < d_{16} = d_{47} < d_{35} < d_{58} < d_{17} < d_{68} < d_{26} < d_{45} < d_{36} < d_{48} < d_{57} < d_{78} < d_{46} < d_{67}$

(2) Keep an empty sub-graph S of G and add the edge e with the smallest weight to S, where the endpoints of e are disconnected, thus we choose the edge e_{56} between A_5 and A_6.

(3) Repeat the process (2) until the sub-graph S spans eight nodes. Thus, the MST of the single valued neutrosophic graph $G(V, E)$ is obtained, as shown in Fig. 5.

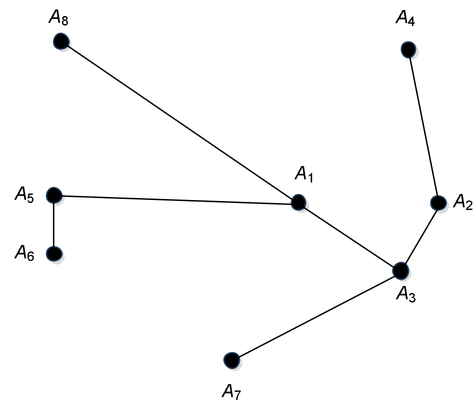

Fig. 5 The MST of the single valued neutrosophic graph $G(V, E)$ with the eight nodes

Step 4. Select a threshold r and disconnect all the edges of the MST with weights greater than r to obtain a certain number of sub-trees (clusters), as listed in Table 1.

Table 1. SVNMST clustering results of the eight different cars

r	Corresponding to clustering result
$r = d_{18} = 0.41$	$\{A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8\}$
$r = d_{15} = 0.4022$	$\{A_1, A_2, A_3, A_4, A_5, A_6, A_7\}, \{A_8\}$
$r = d_{37} = 0.3983$	$\{A_1, A_2, A_3, A_4, A_7\}, \{A_5, A_6\}, \{A_8\}$
$r = d_{24} = 0.2605$	$\{A_1, A_2, A_3, A_4\}, \{A_5, A_6\},\{A_7\}, \{A_8\}$
$r = d_{13} = 0.2551$	$\{A_1, A_2, A_3\}, \{A_4\}, \{A_5, A_6\},\{A_7\}, \{A_8\}$
$r =d_{23} = 0.2269$	$\{A_1\}, \{A_2, A_3\}, \{A_4\}, \{A_5, A_6\}, \{A_7\}, \{A_8\}$
$r =d_{56} = 0.1987$	$\{A_1\}, \{A_2\}, \{A_3\}, \{A_4\}, \{A_5, A_6\}, \{A_7\}, \{A_8\}$
$r = 0$	$\{A_1\}, \{A_2\}, \{A_3\}, \{A_4\}, \{A_5\}, \{A_6\}, \{A_7\}, \{A_8\}$

To compare the SVNMST clustering algorithm with the intuitionistic fuzzy MST clustering algorithm and the fuzzy MST clustering algorithm, we introduce the following example discussed in [8, 9] for comparative convenience.

Example 2. To complete an operational mission, the six sets of operational plans are made (adapted from [8, 9]). To group these operational plans with respect to their comprehensive function, a military committee has been set up to provide assessment information on them. The attributes which are considered here in assessment of the six operational plans A_j ($j = 1, 2, \ldots, 6$) are: (1) x_1 is the effectiveness of operational organization; and (2) x_2 is the effectiveness of operational command.

The weight vector of the attributes x_i (i =1, 2) is w = (0.45, 0.55)^T. The military committee evaluates the performance of the six operational plans A_j (j =1, 2, …, 6) with respect to the attributes x_i (i = 1, 2) and gives the SVNSs as follows:

$A_1 = \{< x_1, 0.7, 0.2, 0.15>, < x_2, 0.6, 0.3, 0.2>\}$,

$A_2 = \{< x_1, 0.4, 0.3, 0.35>, < x_2, 0.8, 0.1, 0.1>\}$,

$A_3 = \{< x_1, 0.55, 0.2, 0.25>, < x_2, 0.7, 0.1, 0.15>\}$,

$A_4 = \{< x_1, 0.44, 0.2, 0.35>, < x_2, 0.6, 0.2, 0.2>\}$,

$A_5 = \{< x_1, 0.5, 0.15, 0.35>, < x_2, 0.75, 0.1, 0.2>\}$,

$A_6 = \{< x_1, 0.55, 0.2, 0.25>, < x_2, 0.57, 0.2, 0.15>\}$.

Then we employ SVNMST clustering algorithm to group these operational plans A_j (j =1, 2, …, 6):

Step 1. Calculate the distance $d_{ij} = d_\lambda(A_i, A_j)$ by Eq. (1) (take $\lambda = 2$). Then we can establish the single valued neutrosophic distance matrix $D = (d_{ij})_{m\times m}$ as follows:

$$D = \begin{bmatrix} 0.0000 & 0.2327 & 0.1507 & 0.1637 & 0.1940 & 0.1051 \\ 0.2327 & 0.0000 & 0.1127 & 0.1326 & 0.1131 & 0.1546 \\ 0.1507 & 0.1127 & 0.0000 & 0.1056 & 0.0764 & 0.0802 \\ 0.1637 & 0.1326 & 0.1056 & 0.0000 & 0.0940 & 0.0784 \\ 0.1940 & 0.1131 & 0.0764 & 0.0940 & 0.0000 & 0.1210 \\ 0.1051 & 0.1546 & 0.0802 & 0.0784 & 0.1210 & 0.0000 \end{bmatrix}.$$

Step 2. Draw the single valued neutrosophic graph $G(V, E)$ where every edge between A_i and A_j (i, j = 1, 2, …, 6) is assigned the weight (single valued neutrosophic distance) d_{ij} coming from an element of the single valued neutrosophic distance matrix $D = (d_{ij})_{m\times m}$, which represents the dissimilarity degree between the samples A_i and A_j. The single valued neutrosophic graph $G(V, E)$ is shown in Fig. 6.

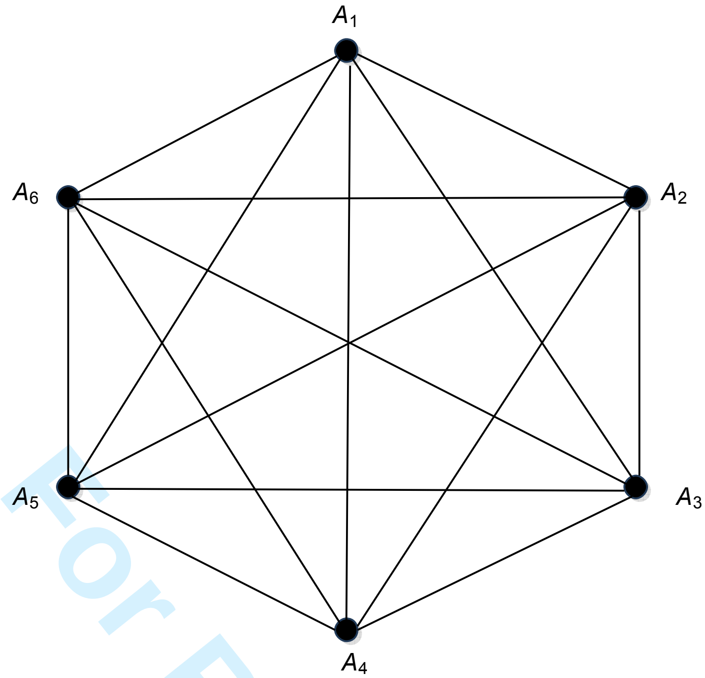

Fig. 6 Single valued neutrosophic graph $G(V, E)$ with the six nodes

Step 3. Build the MST of the single valued neutrosophic graph $G(V, E)$ by Kruskal's method [14] or Prim's method [15]:

(1) Sort the edges of G in increasing order by weights:

$d_{35} < d_{46} < \mathrm{d}_{36} < d_{45} < d_{16} < d_{34} < d_{23} < d_{25} < d_{56} < d_{24} < d_{13} < d_{26} < d_{14} < d_{15} < d_{12}$.

(2) Keep an empty sub-graph S of G and add the edge e with the smallest weight to S, where the endpoints of e are disconnected, thus we choose the edge e_{35} between A_3 and A_5.

(3) Repeat the process (2) until the sub-graph S spans six nodes. Thus, the MST of the single valued neutrosophic graph $G(V, E)$ is obtained, as shown in Fig. 7.

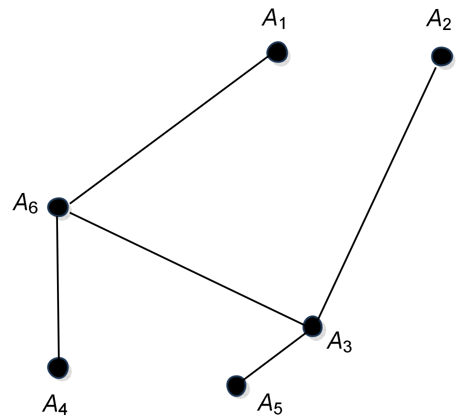

Fig. 7 The MST of the single valued neutrosophic graph $G(V, E)$ with the six nodes

Step 4. Select a threshold r and disconnect all the edges of the MST with weights greater than r to obtain a certain number of sub-trees (clusters), as listed in Table 2.

Table 2. SVNMST clustering results of the six operational plans

r	Corresponding to clustering result
$r = d_{23} = 0.1127$	$\{A_1, A_2, A_3, A_4, A_5, A_6\}$
$r = d_{16} = 0.1051$	$\{A_2\}, \{A_1, A_3, A_4, A_5, A_6\}$
$r = d_{36} = 0.0802$	$\{A_1\}, \{A_2\}, \{A_3, A_4, A_5, A_6\}$
$r = d_{46} = 0.0784$	$\{A_1\}, \{A_2\}, \{A_3, A_5\}, \{A_4, A_6\}$
$r = d_{35} = 0.0764$	$\{A_1\}, \{A_2\}, \{A_3, A_5\}, \{A_4\}, \{A_6\}$
$r = 0$	$\{A_1\}, \{A_2\}, \{A_3\}, \{A_4\}, \{A_5\}, \{A_6\}$

To compare the SVNMST clustering algorithm with the IFMST clustering algorithm in [8], assume that the indeterminacy-membership function $I_A(x_i)$ is not considered independently in a SVNS A_j, then the samples' information given by the military committee will be the intuitionistic fuzzy data (adopted in [8]):

$A_1 = \{< x_1, 0.7, 0.15>, < x_2, 0.6, 0.2>\}$, $A_2 = \{< x_1, 0.4, 0.35>, < x_2, 0.8, 0.1>\}$,

$A_3 = \{< x_1, 0.55, 0.25>, < x_2, 0.7, 0.15>\}$, $A_4 = \{< x_1, 0.44, 0.35>, < x_2, 0.6, 0.2>\}$,

$A_5 = \{< x_1, 0.5, 0.35>, < x_2, 0.75, 0.2>\}$, $A_6 = \{< x_1, 0.55, 0.25>, < x_2, 0.57, 0.15>\}$.

Then, the operational plans A_j (j = 1,2, …, 6) can be clustered by the following IFMST clustering algorithm [8]:

Step 1. Calculate $d_{ij} = d(A_i, A_j)$ by the intuitionistic fuzzy distance measure:

$$d(A,B) = \left\{ \frac{1}{2} \sum_{i=1}^{n} w_i \left[\left| \mu_A(x_i) - \mu_B(x_i) \right|^2 + \left| \nu_A(x_i) - \nu_B(x_i) \right|^2 + \left| \pi_A(x_i) - \pi_B(x_i) \right|^2 \right] \right\}^{1/2}. \quad (4)$$

Thus, we get the intuitionistic fuzzy distance matrix:

$$D = \begin{bmatrix} 0.0000 & 0.2450 & 0.1225 & 0.1170 & 0.1725 & 0.1115 \\ 0.2450 & 0.0000 & 0.1225 & 0.1280 & 0.1000 & 0.1940 \\ 0.1225 & 0.1225 & 0.0000 & 0.1045 & 0.1000 & 0.0715 \\ 0.1170 & 0.1280 & 0.1045 & 0.0000 & 0.1095 & 0.0935 \\ 0.1725 & 0.1000 & 0.1000 & 0.1095 & 0.0000 & 0.1715 \\ 0.1115 & 0.1940 & 0.0715 & 0.0935 & 0.1715 & 0.0000 \end{bmatrix}.$$

Step 2. Structuring the intuitionistic fuzzy graph $G = (V, E)$, see Step 2 in the SVNMST clustering algorithm and Fig. 6.

Step 3. Get the MST of the intuitionistic fuzzy graph $G = (V, E)$ by Kruskal's method [14] or Prim's method [15]:

(1) Sort the edges of G in increasing order by weights:

$d_{36} < d_{46} < d_{35} = d_{25} < d_{34} < d_{45} < d_{16} < d_{14} < d_{13} = d_{23} < d_{24} < d_{56} < d_{15} < d_{26} < d_{52}$.

(2) Keep an empty sub-graph S of G and add the edge e with the smallest weight to S, where the endpoints of e are disconnected, thus we choose the edge e_{36} between A_3 and A_6.

(3) Repeat the process (2) until the sub-graph S spans six nodes. Thus, the MST of the intuitionistic fuzzy graph $G(V, E)$ is obtained, as shown in Fig. 8.

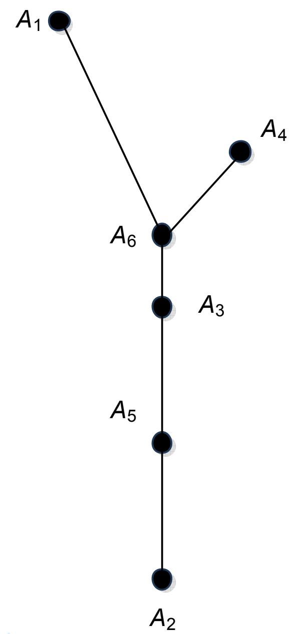

Fig. 8 The MST of the intuitionistic fuzzy graph $G = (V, E)$

Step 4. Select a threshold r and disconnect all the edges of the MST with weights greater than r to obtain a certain number of sub-trees (clusters), as listed in Table 3.

Table 3. IFMST clustering results of the six operational plans

r	Corresponding to clustering result
$r = d_{16} = 0.1115$	$\{A_1, A_2, A_3, A_4, A_5, A_6\}$
$r = d_{25} = d_{35} = 0.1$	$\{A_1\}, \{A_2, A_3, A_4, A_5, A_6\}$
$r = d_{46} = 0.088$	$\{A_1\}, \{A_2\}, \{A_5\}, \{A_3, A_4, A_6\}$
$r = d_{36} = 0.0715$	$\{A_1\}, \{A_2\}, \{A_4\}, \{A_5\}, \{A_3, A_6\}$
$r = 0$	$\{A_1\}, \{A_2\}, \{A_3\}, \{A_4\}, \{A_5\}, \{A_6\}$

It is well know that the fuzzy set is only composed of the membership degree. Therefore, we only consider the membership degrees of intuitionistic fuzzy data, and then the operational plan information given by the following fuzzy data [8]:

$A_1 = \{< x_1, 0.7>, < x_2, 0.6>\}$, $A_2 = \{< x_1, 0.4>, < x_2, 0.8>\}$, $A_3 = \{< x_1, 0.55>, < x_2, 0.7>\}$,

$A_4 = \{< x_1, 0.44>, < x_2, 0.6>\}$, $A_5 = \{< x_1, 0.5>, < x_2, 0.75>\}$, $A_6 = \{< x_1, 0.55>, < x_2, 0.57>\}$.

Then, the operational plans A_j (j =1, 2, …, 6) can be clustered by the fuzzy MST clustering algorithm:

Step 1. Calculate $d_{ij} = d(A_i, A_j)$ by the fuzzy distance measure:

$$d(A,B) = \left\{ \sum_{i=1}^{2} w_i \left[\left| \mu_A(x_i) - \mu_B(x_i) \right|^2 \right] \right\}^{1/2}. \tag{5}$$

Thus, we obtain the fuzzy distance matrix:

$$D = \begin{bmatrix} 0.0000 & 0.2500 & 0.1250 & 0.1744 & 0.1743 & 0.1031 \\ 0.2500 & 0.0000 & 0.1250 & 0.1507 & 0.0766 & 0.1980 \\ 0.1250 & 0.1250 & 0.0000 & 0.1046 & 0.0500 & 0.0964 \\ 0.1744 & 0.1507 & 0.1046 & 0.0000 & 0.1183 & 0.0771 \\ 0.1743 & 0.0766 & 0.0500 & 0.1183 & 0.0000 & 0.1376 \\ 0.1031 & 0.1980 & 0.0964 & 0.0771 & 0.1376 & 0.0000 \end{bmatrix}.$$

Step 2. Draw the fuzzy graph $G = (V, E)$. See also Step 2 in the SVNMST clustering algorithm and Fig. 6.

Step 3. Build the MST of the fuzzy graph $G = (V, E)$ by Kruskal's method [14] or Prim's method [15]:

(1) Sort the edges of G in increasing order by weights:

$d_{35} < d_{25} < d_{46} < d_{36} < d_{16} < d_{34} < d_{45} < d_{13} = d_{23} < d_{56} < d_{24} < d_{15} < d_{14} < d_{26} < d_{12}$.

(2) Keep an empty sub-graph S of G and add the edge e with the smallest weight to S, where the endpoints of e are disconnected, thus we choose the edge e_{35} between A_3 and A_5.

(3) Repeat the process (2) until the sub-graph S spans six nodes. Thus, the MST of the fuzzy graph $G(V, E)$ is obtained, as shown in Fig. 9.

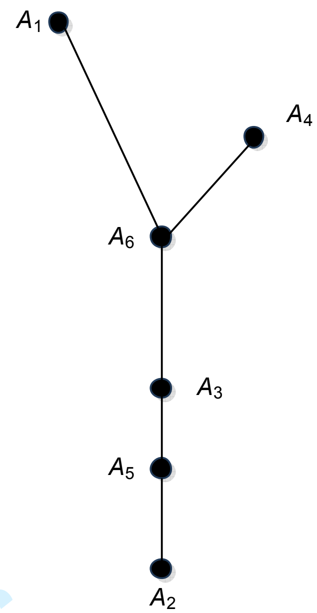

Fig. 9 The MST of the fuzzy graph $G = (V, E)$

Step 4. Select a threshold r and disconnect all the edges of the MST with weights greater than r to obtain a certain number of sub-trees (clusters), as listed in Table 4.

Table 4. Fuzzy MST clustering results

r	Corresponding to clustering result
$r = d_{16} = 0.1031$	$\{A_1, A_2, A_3, A_4, A_5, A_6\}$
$r = d_{36} = 0.0964$	$\{A_1\}, \{A_2, A_3, A_4, A_5, A_6\}$
$r = d_{46} = 0.0774$	$\{A_1\}, \{A_2, A_3, A_5\}, \{A_4, A_6\}$
$r = d_{25} = 0.0766$	$\{A_1\}, \{ A_2, A_3, A_5\}, \{A_4\}, \{A_6\}$
$r = d_{35} = 0.0500$	$\{A_1\}, \{A_2\},\{A_3, A_5\}, \{A_4\}, \{A_6\}$
$r = 0$	$\{A_1\}, \{A_2\}, \{A_3\}, \{A_4\}, \{A_5\}, \{A_6\}$

For convenient comparisons, we put the clustering results of three kinds of clustering algorithms into Table 5.

Table 5. Clustering results of three kinds of clustering algorithms

Class	SVNMST clustering algorithm	IFMST clustering algorithm	Fuzzy MST clustering algorithm
1	$\{A_1, A_2, A_3, A_4, A_5, A_6\}$	$\{A_1, A_2, A_3, A_4, A_5, A_6\}$	$\{A_1, A_2, A_3, A_4, A_5, A_6\}$
2	$\{A_2\}, \{A_1, A_3, A_4, A_5, A_6\}$	$\{A_1\}, \{A_2, A_3, A_4, A_5, A_6\}$	$\{A_1\}, \{A_2, A_3, A_4, A_5, A_6\}$
3	$\{A_1\}, \{A_2\}, \{A_3, A_4, A_5, A_6\}$		$\{A_1\}, \{A_2, A_3, A_5\}, \{A_4, A_6\}$
4	$\{A_1\}, \{A_2\}, \{A_3, A_5\}, \{A_4, A_6\}$	$\{A_1\}, \{A_2\}, \{A_5\}, \{A_3, A_4, A_6\}$	$\{A_1\}, \{A_2, A_3, A_5\}, \{A_4\}, \{A_6\}$

5	$\{A_1\}, \{A_2\}, \{A_3, A_5\}, \{A_4\}, \{A_6\}$	$\{A_1\}, \{A_2\}, \{A_4\}, \{A_5\}, \{A_3, A_6\}$	$\{A_1\}, \{A_2\}, \{A_3, A_5\}, \{A_4\}, \{A_6\}$
6	$\{A_1\}, \{A_2\}, \{A_3\}, \{A_4\}, \{A_5\}, \{A_6\}$	$\{A_1\}, \{A_2\}, \{A_3\}, \{A_4\}, \{A_5\}, \{A_6\}$	$\{A_1\}, \{A_2\}, \{A_3\}, \{A_4\}, \{A_5\}, \{A_6\}$

From Table 5, we can see that the clustering results of the three clustering algorithms are quite different. The main reason can be given by the following comparative analysis.

As is known to all, the single valued neutrosophic information is a generalization of intuitionistic fuzzy information, and intuitionistic fuzzy information is a further generalization of fuzzy information. On the one hand, a SVNS is an instance of a neutrosophic set, which give us an additional possibility to represent uncertainty, imprecise, incomplete, and inconsistent information which exist in real world. It can describe and handle indeterminate information and inconsistent information. However, the connector in the fuzzy set is defined with respect to *T*, i.e. membership only, hence the information of indeterminacy and nonmembership is lost. The connectors in the intuitionistic fuzzy set are defined with respect to *T* and *F*, i.e. membership and nonmembership only, hence the indeterminacy is what is left from 1, and then the IFS can only handle incomplete information but not the indeterminate information and inconsistent information. While in the SVNS, its truth-membership, indeterminacy-membership, and falsity-membership are represented independently, and then they can be defined with respect to any of them (no restriction). So the notion of SVNSs is more general. On the other hand, the clustering analysis under single valued neutrosophic environment is suitable for capturing imprecise, uncertain, and inconsistent information in clustering the data. Thus, the SVNMST clustering algorithm clusters the single valued neutrosophic information, while the IFMST clustering algorithm clusters the intuitionistic fuzzy information and the fuzzy MST clustering algorithm clusters the fuzzy information. Obviously, the SVNMST clustering algorithm is the extension of both the IFMST clustering algorithm and the fuzzy MST clustering algorithm. Therefore, compared with the IFMST clustering algorithm and the fuzzy MST clustering algorithm, the SVNMST clustering algorithm is more general. Furthermore, when we encounter some situations, which are represented by indeterminate information and inconsistent information, the SVNMST clustering algorithm demonstrates its great superiority in clustering those single valued neutrosophic data.

6. Conclusion

This paper defined a generalized single valued neutrosophic weighted distance and proposed the SVNMST clustering algorithm as a generalization of the IFMST clustering algorithm. Through the computational tests on the SVNMST clustering algorithm, the IFMST clustering algorithm and the fuzzy MST clustering algorithm, the clustering results have shown that the SVNMST clustering algorithm is more general and more reasonable than the IFMST clustering algorithm and the fuzzy MST clustering algorithm. Furthermore, in the situations which are represented by indeterminate information and inconsistent information, the SVNMST clustering algorithm demonstrates its great superiority in clustering those single valued neutrosophic data, since the SVNSs are a powerful tool to deal with uncertainty, imprecise, incomplete, and inconsistent information. In the future, the developed algorithm can be applied to many areas such as information retrieval, investment decision making, and data mining.

References

[1] J. Harary, Graph theory, Reading, Addison –Wesley, 1969.

[2] A. K. Jain, R. C. Dubes, Algorithms for Clustering Data, Prentice-Hall, Englewood Cliffs, NJ, 1988.

[3] C. T. Zahn, Graph-theoretical methods for detecting and describing gestalt clusters, IEEE Trans. Comput. C-20 (1971) 68–86.

[4] Y. Xu, V. Olman, D. Xu, Clustering gene expression data using a graph-theoretic approach: an application of minimum spanning tree, Bioinformatics 18 (2002) 536–545.

[5] E. H. Ruspini, A new approach to clustering. Information and Control 15 (1969) 22–32.

[6] Y. H. Dong, Y. T. Zhuang, K. Chen, X. Y. Tu, A hierarchical clustering algorithm based on fuzzy graph connectedness, Fuzzy Sets and Systems 157(11) (2006) 1760-1774.

[7] D. S. Chen, K. X. Li, L. B. Zhao, Fuzzy graph maximal tree clustering method and its application. Operations Research and Management Science 16 (2007) 69–73.

[8] H. Zhao, Z. Xu, S. Liu, Z. Wang, Intuitionistic fuzzy MST clustering algorithms, Computers & Industrial Engineering 62 (2012) 1130–1140.

[9] X. Zhang, Z. Xu, An MST cluster analysis method under hesitant fuzzy environment, Control and Cybernetics 41(3) (2012) 645-666.

[10] F. Smarandache, A unifying field in logics. neutrosophy: Neutrosophic probability, set and logic, American Research Press, Rehoboth, 1998.

[11] H. Wang, F. Smarandache, Y. Q. Zhang, et al., Single valued neutrosophic sets, Multispace and Multistructure 4 (2010) 410-413.

[12] J. Ye, Multicriteria decision-making method using the correlation coefficient under single-valued neutrosophic environment, International Journal of General Systems 42(4) (2013) 386-394.

[13] J. Ye, Single valued neutrosophic cross-entropy for multicriteria decision making problems, Applied Mathematical Modelling (2013) doi: 10.1016/j.apm.2013.07.020.

[14] J. B. Kruskal, On the shortest spanning subtree of a graph and the traveling salesman problem. Proceedings of the American Mathematical Society 7 (1956) 48–50.

[15] R. C. Prim, Shortest connection networks and some generalizations. Bell System Technology Journal, 36 (1957) 1389–1401.