Chapter 478

Minimum Spanning Tree

Introduction

A minimum spanning tree links all nodes (points or vertices) of a network with the minimum length of all arcs. This procedure finds the minimum spanning tree of a network using a greedy algorithm. If the network is not connected, the solution, called a *minimum spanning forest*, is a combination of minimum spanning trees formed on the connected subsets.

The algorithm is used in applications such as transportation networks, computer networks, and water networks where a tree connecting all nodes must have minimum length.

Greedy Algorithm

The algorithm proceeds as follows:

1. Start with any node.
2. Connect this node to its nearest neighbor using any of the available branches.
3. Find the unconnected node that is closest any of the connected nodes. Connect these nodes.
4. Repeat steps 2 and 3 until all nodes have been connected.

Data Structure

This procedure requires a special data format in which the rows represent the network branches. Each branch is defined by two nodes and the distance between its nodes.

Consider the following example. Each row represents an electric cable that connects two orchards on a large fruit farm that produces many types of fruit. The nodes are identified by the type of fruit grown in that section. The length of the cable is given in the last column. These data are stored in the dataset *Wiring Network*.

Wiring Network dataset

Site1	Site2	Length
A1	A2	6.8
A1	A3	3.2
A2	B1	9.1
A2	A3	8.5
A2	B2	7.2
B1	B2	10.3
A3	B2	14.8
A3	C1	5.6
B2	C1	8.2
B2	C2	9.1
C1	C2	10.9
Z1	Z2	0.5
Z2	Z3	3.4
Z1	Z3	2.2

Procedure Options

This section describes the options available in this procedure.

Specifications Tab

Set the specifications for the analysis.

Columns Specifying Arcs

Node 1 Column

Specify the column containing the first node (also called vertex or point) of an arc (also called edge, line, link, or branch). Each row gives the specification of a single arc by giving its two ends (nodes) as well as the length between them. Here, we define an arc as two connected nodes along with the distance between them.

A node may be a number or a text value. It does not matter which node is specified first.

Node 2 Column

Specify the column containing the second node (also called vertex or point) of an arc (also called edge, line, link, or branch). Each row gives the specification of a single arc by giving its two ends (nodes) as well as the length between them. Here, we define an arc as two connected nodes along with the distance between them.

A node may be a number or a text value. It does not matter which node is specified second.

Length Column

Specify the column containing the length (distance or cost) between the two nodes of an arc. The values should be positive numbers.

Reports Tab

Select Reports

Minimum Spanning Tree – Possible Network Arcs

Indicate which reports you want to view.

Report Options

Column Names

This option lets you select whether to display only variable names, variable labels, or both.

Example 1 – Wiring Network

This section presents an example of how to create a minimum spanning tree from the data presented in the example given above. The data are contained in the *Wiring Network* database. Each row represents a possible arc (connection or branch) of the network by identifying the end points along with the length between them.

Wiring Network dataset

Site1	Site2	Length
A1	A2	6.8
A1	A3	3.2
A2	B1	9.1
A2	A3	8.5
A2	B2	7.2
B1	B2	10.3
A3	B2	14.8
A3	C1	5.6
B2	C1	8.2
B2	C2	9.1
C1	C2	10.9
Z1	Z2	0.5
Z2	Z3	3.4
Z1	Z3	2.2

You may follow along here by making the appropriate entries or load the completed template **Example 1** by clicking on Open Example Template from the File menu of the Minimum Spanning Tree window.

1 Open the Wiring Network dataset.

- From the **File** menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **Wiring Network.NCSS**.
- Click **Open**.

2 Open the Minimum Spanning Tree window.

- Using the Analysis menu or the Procedure Navigator, find and select the **Minimum Spanning Tree** procedure.
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

3 Specify the problem.

- On the **Minimum Spanning Tree** window, select the **Specifications tab**.
- Double-click in the **Node 1 Column** text box. This will bring up the column selection window.
- Select columns **Site1** from the list of columns and then click **Ok**. "Site1" will appear in this box.
- Double-click in the **Node 2 Column** text box. This will bring up the column selection window.
- Select columns **Site2** from the list of columns and then click **Ok**. "Site2" will appear in this box.
- Double-click in the **Length Column** text box. This will bring up the column selection window.
- Select **Length** from the list of columns and then click **Ok**. "Length" will appear in this box.

4 Run the procedure.

- From the **Run** menu, select **Run Procedure**. Alternatively, just click the green Run button.

Minimum Spanning Tree

Minimum Spanning Forest

Tree	Arc (Row)	Length (Length)	Site1 (Node 1)	Site2 (Node 2)
1	1	6.8	A1	A2
1	2	3.2	A1	A3
1	3	9.1	A2	B1
1	5	7.2	A2	B2
1	8	5.6	A3	C1
1	10	9.1	B2	C2
2	12	0.5	Z1	Z2
2	14	2.2	Z1	Z3
Total		43.7		

This report lists the solution by giving the arcs that form the minimum spanning tree(s). Note that in this example, there were two trees so the solution is called a forest. You can see that there are no defined arcs between the first and second trees.

Node Connections

Node Connections

Node	Connected Nodes
A1	A2*, A3*
A2	A1*, B1*, A3, B2*
A3	A1*, A2, B2, C1*
B1	A2*, B2
B2	A2*, B1, A3, C1, C2*
C1	A3*, B2, C2
C2	B2*, C1
Z1	Z2*, Z3*
Z2	Z1*, Z3
Z3	Z2, Z1*

This report presents a list of the nodes followed by the nodes they are connected with. If the pair of nodes is in the minimum spanning tree, the connected note is starred.

Possible Network Arcs

Possible Network Arcs

Tree	Arc (Row)	Length (Length)	Site1 (Node 1)	Site2 (Node 2)
1	1*	6.8	A1	A2
1	2*	3.2	A1	A3
1	3*	9.1	A2	B1
	4	8.5	A2	A3
1	5*	7.2	A2	B2
	6	10.3	B1	B2
	7	14.8	A3	B2
1	8*	5.6	A3	C1
	9	8.2	B2	C1
1	10*	9.1	B2	C2
	11	10.9	C1	C2
2	12*	0.5	Z1	Z2
	13	3.4	Z2	Z3
2	14*	2.2	Z1	Z3

* The starred arcs are in the minimum spanning tree.

This report lists the arcs that were input from the dataset. The started arcs are in the minimum spanning tree.