SAS/GRAPH® 9.4: Network Visualization Workshop User's Guide

SAS® Documentation

October 13, 2017

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2013. *SAS/GRAPH® 9.4: Network Visualization Workshop User's Guide*. Cary, NC: SAS Institute Inc.

SAS/GRAPH® 9.4: Network Visualization Workshop User's Guide

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

Electronic book 2, December 2013

SAS® Publishing provides a complete selection of books and electronic products to help customers use SAS software to its fullest potential. For more information about our e-books, e-learning products, CDs, and hard-copy books, visit the SAS Publishing Web site at **support.sas.com/publishing** or call 1-800-727-3228.

SAS/GRAPH Network Visualization Workshop has changed its release number to align with the current SAS/GRAPH release number. Apart from this new release number, the documentation remains unchanged from the previous release.

Contents

1

Overview of SAS/GRAPH Network Visualization Workshop

Notice of Discontinuation

Starting with SAS 9.4M5, SAS/GRAPH Network Visualization Workshop is discontinued and is no longer supported.

Introduction to SAS/GRAPH Network Visualization Workshop

SAS/GRAPH Network Visualization Workshop is an interactive, graphics-oriented application for visualizing and investigating data. The application uses visualization techniques that enable you to detect patterns and extract information that might be concealed, often in very large quantities of data. SAS/GRAPH Network Visualization Workshop is particularly useful for examining network data (that is, data that is structured into nodes and links that connect the nodes). SAS/GRAPH Network Visualization Workshop is one of the SAS/GRAPH suite of products and requires SAS/GRAPH to be installed on the system.

Here are the major tasks that you can perform:

- explore both network and non-network relational data
- render data in any of five different types of statistical graph: bar charts, box plots, histograms, pie charts, and scatter plots
- render network data in any of five layout patterns: circular, hierarchical, hexagonal, multi-level force, and fixed position
- explore and manipulate data in data tables (for example, by sorting the data)
- visualize and manipulate data in graphs using tools such as the zoom and lens tool
- use a combination of data tables, statistical graphics, and network graphs to selectively view and filter your data
- use different selection modes on your data tables, statistical graphics, and network graphs to graphically subset data at multiple levels
- save your data tables and associated graphs as projects that can be reopened with a single menu command

See Also

“Getting Started with Network Visualization” on page 7

Accessibility Features of SAS/GRAPH 9.4 Network Visualization Workshop

About the Accessibility Features

SAS/GRAPH Network Visualization Workshop includes accessibility and compatibility features that improve usability of the product for users with disabilities, with exceptions noted below. These features are related to accessibility standards for electronic information technology that were adopted by the U.S. Government under Section 508 of the U.S. Rehabilitation Act of 1973, as amended.

If you have questions or concerns about the accessibility of SAS products, send e-mail to accessibility@sas.com.

Standard Keyboard Navigation

SAS/GRAPH Network Visualization Workshop can be navigated by using the keyboard. The following table includes some guidelines:

Table 1.1 *Standard Keyboard Navigation*

Task	Keyboard Control
Move forward through graph and table windows	CTRL + TAB
Move backward through graph and table windows	CTRL + SHIFT + TAB
Move forward through controls on a window or cells in a table	TAB
Move backward through controls or cells in a table	SHIFT + TAB
Display menus when they have focus	DOWN ARROW

Task	Keyboard Control
Display the right-click menu in a graph window that has focus Display the right-click menu in a table window when a variable or observation is selected	SHIFT + F10
Select a variable in a table that has focus	CTRL + SPACE BAR
Select an observation in a table that has focus	SHIFT + SPACE BAR
Display the contents of drop-down lists	ALT + DOWN ARROW
Scroll through the contents of drop-down lists or right-click menus	DOWN ARROW and UP ARROW
Select check boxes when they have focus	SPACEBAR
Select a different radio button when a radio button has focus	DOWN ARROW, UP ARROW, RIGHT ARROW, and LEFT ARROW

Accessibility Exceptions

The following table describes accessibility compliance with Section 508. All known exceptions to accessibility standards are documented in the table.

Section 508 Accessibility Criteria	Support Status	Explanation
(a) When software is designed to run on a system that has a keyboard, product functions shall be executable from a keyboard where the function itself or the result of performing a function can be discerned textually.	Supported with exceptions	Exceptions include the following: ■ There are no keyboard equivalent keys to zoom, pan, lens and highlight the diagram. ■ There are no keyboard keys available to de-select the observation or variable once it is selected. ■ There are no keyboard equivalent keys to minimize, maximize, and close the child windows.
(b) Applications shall not disrupt or disable activated features of other products that are identified as accessibility features, where those features are developed and documented according to industry standards. Applications also shall not disrupt or disable activated features of any operating system that are identified as accessibility features where the application programming interface for those accessibility features has been documented by the manufacturer of the operating system and is available to the product developer.	Supported	The software does not disrupt or disable any of the keyboard accessibility features incorporated within the operating system.
(c) A well-defined on-screen indication of the current focus shall be provided that moves among interactive interface elements as the input focus changes. The focus shall be programmatically exposed so that Assistive Technology can track focus and focus changes.	Supported with minor exceptions	On-screen indication of the current focus is provided except in the following cases: ■ In the diagram window, the visual focus indicator is not prominent. ■ Tab navigation skips over the icon toolbar. However, the toolbar commands are available from the application menu.

Section 508 Accessibility Criteria	Support Status	Explanation
(d) Sufficient information about a user interface element including the identity, operation and state of the element shall be available to Assistive Technology. When an image represents a program element, the information conveyed by the image must also be available in text.	Supported with exception	Most interface elements, including menu bar options and shortcut keys, are read by a screen reader. However, JAWS is not able to read the data tables
(e) When bitmap images are used to identify controls, status indicators, or other programmatic elements, the meaning assigned to those images shall be consistent throughout an application's performance.	Supported	Images are used consistently throughout the interface.
(f) Textual information shall be provided through operating system functions for displaying text. The minimum information that shall be made available is text content, text input caret location, and text attributes.	Supported	The software uses standard operating system functions for displaying text.
(g) Applications shall not override user selected contrast and color selections and other individual display attributes.	Supported	The software inherits user-selected contrast and color selections for display attributes to the extent supported by the platform.
(h) When animation is displayed, the information shall be displayable in at least one non-animated presentation mode at the option of the user.	Not applicable	The software contains no animation.
(i) Color coding shall not be used as the only means of conveying information, indicating an action, prompting a response, or distinguishing a visual element.	Supported	Color alone is not used to convey meaning.

Section 508 Accessibility Criteria	Support Status	Explanation
(j) When a product permits a user to adjust color and contrast settings, a variety of color selections capable of producing a range of contrast levels shall be provided.	Supported with minor exceptions	The software offers user options to enable users to choose alternating row colors for grids to ensure ample color contrast for a range of vision abilities.
(k) Software shall not use flashing or blinking text, objects, or other elements having a flash or blink frequency greater than 2 Hz and lower than 55 Hz.	Supported	The software uses no flashing or blinking elements beyond the system caret.
(l) When electronic forms are used, the form shall allow people using Assistive Technology to access the information, field elements, and functionality required for completion and submission of the form, including all directions and cues.	Supported	Compatibility with assistive technology was assessed using JAWS screen reader version 9.

Start SAS/GRAPH Network Visualization Workshop

You can start SAS/GRAPH Network Visualization Workshop from the **Start** ▶ **Programs** ▶ **SAS** menu path.

Getting Started with Network Visualization

SAS/GRAPH Network Visualization Workshop assumes that the network to be investigated is described by two data sets:

- One data set contains information about the nodes in the network.
- A second data set contains information about the links that connect the nodes.

  Note: If only link data is available, SAS/GRAPH Network Visualization Workshop can generate a simple node data set from the link data.

For information about these data sets, see "About Data Sources" on page 11.

For information about the main tasks associated with network visualization, see "Main Steps for Visualizing Network Data" on page 8.

SAS/GRAPH Network Visualization Workshop is installed with several sample data sets. The following topics describe the samples and provide example use cases that can help you learn to use the product:

- "Using the Sample Data" on page 120
- "Example Use Case: Credit Card Fraud" on page 128
- "Example Use Case: Fortune 100 Boards of Directors" on page 137
- "Example Use Case: Web Path Data" on page 148

Main Steps for Visualizing Network Data

There are multiple ways to investigate network data, and your approach can vary from what is described here. In general though, a typical session might consist of the following steps:

1. Load your node and link data into SAS/GRAPH Network Visualization Workshop. For instructions, see "Load Data into SAS/GRAPH Network Visualization Workshop" on page 14.

2. Use the Edit Data Attributes dialog box to identify the FROM and TO variables in the link data set and the NODE variable in the node data set. When you set data

attributes, you can also specify which variables to use for link colors as well as for node colors, shapes, and labels. For more information, see "Specify Data Attributes" on page 17.

3 Create one or more statistical graphs to see standard relational views of node and link data separately. For instructions, see "Create a Statistical Graph" on page 47.

4 Create one or more network graphs to see the network of nodes and links. For instructions, see "Create a Network Graph" on page 83.

   You choose from several layout patterns when you create a network graph. You might not initially be aware of any underlying structure in your data. It can be helpful to try multiple layout patterns to determine which are the most useful for the network being investigated. After creating the graph, you can later change its layout.

5 Explore and investigate your data. You can investigate data in the following ways:

   - Explore and manipulate data in the data tables, for example, by sorting the data.
   - Use a combination of statistical graphics, network graphs, and data tables to selectively view and filter your data. To filter the data, select individual observations or groups of observations in a data table or in a graph. The observations are automatically selected in all graphs and data tables. You can also use local selection mode to graphically subset data at multiple levels.
   - Use the tools available to help visualize the graphs. For example, you can zoom in on a portion of a graph or apply a magnifying lens to a network graph. You can also show the labels for particular nodes in a network graph.

6 You might want to change the appearance and behaviors associated with a graph by changing the graph's properties. For example, the properties of a statistical graph determine the graph's colors, whether and where the graph displays titles or a legend, and more. For more information, see "Changing the Properties of a Statistical Graph" on page 49 and "Changing the Properties of a Network Graph" on page 86.

7 Save your data as a project so that you can easily reopen the data tables and associated graphs. For instructions, see "Save a Project" on page 114.

2

Working with Data Sources

About Data Sources

This section describes the format and structure of network data and provides some examples of real-world network data. This section also explains how you can use SAS/GRAPH Network Visualization Workshop to explore non-network relational data.

Network Data

SAS/GRAPH Network Visualization Workshop uses SAS data sets to define a network. In simple terms, a network is a system of interconnected items, and network data is the information that describes such a system.

Network data consists of two types of data:

Table 2.1 *Network Data*

Data Type	What the Data Contains
Node	Information about the items that are connected
Link	Information about the connections between nodes

Many real-world problems can be represented by using a collection of nodes and links. Common examples include supply chains, Web sites, database schema, communication networks, and software module dependencies. For a supply chain, the nodes might represent manufacturing plants, warehouses, and customer locations. The links might represent the flow of goods or products between the locations. For a communications network, the nodes might be switches, routers, and other hardware devices with attributes such as capacity, device type, traffic volume, and number of dropped packets. The links might represent transmission facilities or media connecting the nodes with attributes such as failure rates, error rates, and traffic volume.

To create a network graph, SAS/GRAPH Network Visualization Workshop requires two SAS data sets: a link data set and a node data set. Together these data sets constitute the network data. Each link connects two nodes, though a node can have multiple connecting links. There can be thousands of links connecting thousands of nodes.

SAS/GRAPH Network Visualization Workshop also enables you to create statistical graphs that are based on either node data or link data. For more information, see “Non-Network Data ” on page 14.

Node Data

A node data set defines the nodes in a network. In this data set, each row represents one node in the network. At a minimum, the node data set must contain a node identifier variable. You specify the name of this variable when you set data attributes. For details, see “Specify Data Attributes” on page 17.

SAS/GRAPH Network Visualization Workshop provides the capability to automatically create a minimal node data set from the link data set that you provide. However, you might want to provide your own node data set. To be truly useful, the node data set also includes one or more variables for attributes that describe the nodes. These variables can be useful in several ways:

- When you set data attributes, you can specify one of these variables to use for node colors and the same or a different variable to be used for node shapes. You can also specify a variable to be used for text that appears when you apply labels to a network graph.
- You can use these variables in conjunction with statistical, non-network graphs to filter data.
- You can use these variables in data tables to sort the data.

Link Data

A link data set defines the links in the network. In this data set, each row represents one link between two nodes in the network.

The data set must contain at least two variables to identify the link. The values of these variables must be node identifiers; that is, the values for these variables must exist in the node identifier column in the corresponding node data set. The variables perform a From and a To role. The From variable lists the node at which the link originates, and the To variable lists the node at which the link terminates. You specify the names for these variables when you set data attributes. For details, see “Specify Data Attributes” on page 17.

Other variables in the link data set can be used to store attributes related to the links. These variables can be useful in several ways:

- When you set data attributes, you can specify one of these variables to use for link colors.
- You can use these variables in conjunction with statistical, non-network graphs to filter data.
- You can use these variables in data tables to sort the data.

Non-Network Data

In addition to using SAS/GRAPH Network Visualization Workshop to explore network data, you can also explore standard relational data that does not constitute a network. The data is based on a node or a link data set.

You can investigate relational data in the following ways:

- explore and manipulate data in the data table
- create one or more statistical graphs based on node or link data
- use a combination of statistical graphics with the data table to selectively view and filter your data
- use different selection modes on your data table and statistical graphics to graphically subset data at multiple levels

Load Data into SAS/GRAPH Network Visualization Workshop

You must load source data into SAS/GRAPH Network Visualization Workshop before you can work with the data. You load data by opening a project, a node data set, or a link data set. You can load data in one of three ways:

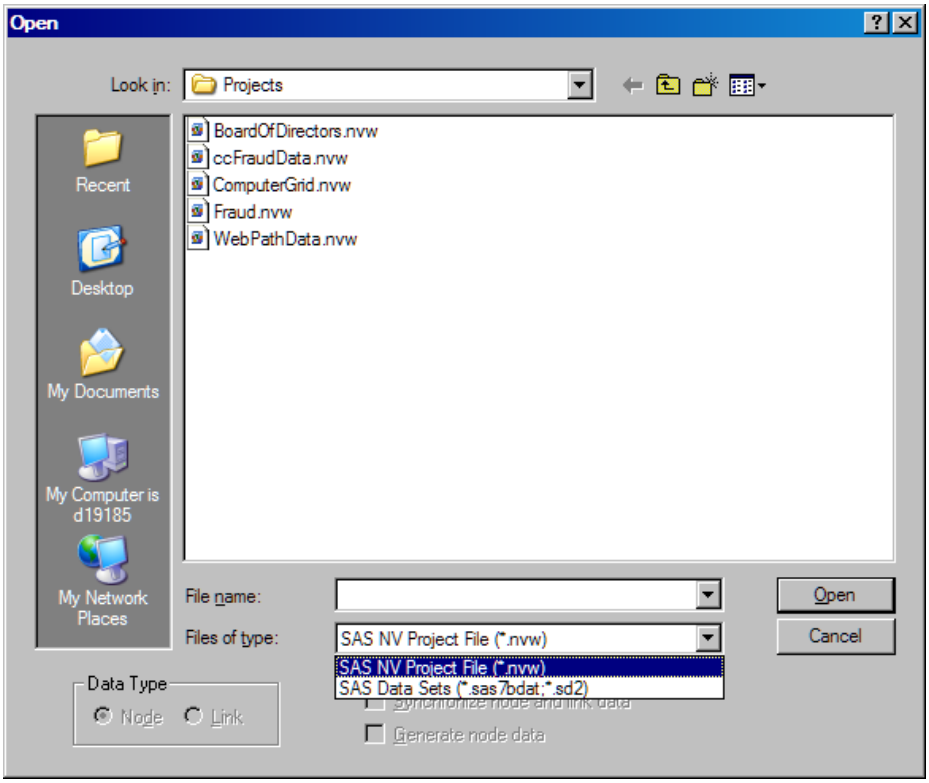
- select a project from the Welcome dialog box

- select one of the last eight opened data sets or projects from the **File** menu
- use the Open dialog box

You can have at most one link and one node data set open at any time.

To use the Open dialog box:

1 Select **File** ▶ **Open**. The Open dialog box opens.



2 In the **Files of type** list box, select either project files or SAS data sets, depending on the type of file that you want to open.

3 If you are opening a data set, then select either the **Node** or **Link** button, depending on the type of data set that you want to open.

4 If you are opening a data set and want SAS/GRAPH Network Visualization Workshop to synchronize node and link data, select **Synchronize node and link**

data. This option is available only when you are opening the second data set of your network data, and that data set has not been synchronized with your first data set.

For more information about synchronization, see "Synchronizing Data" on page 19.

5 If you are opening a link data set and want SAS/GRAPH Network Visualization Workshop to generate the node data set, select **Generate node data**.

6 Select the file you want to open.

7 Click **Open**.

SAS/GRAPH Network Visualization Workshop displays the data sets in data tables. If you opened a project file, then SAS/GRAPH Network Visualization Workshop also displays any graphs that are associated with the project.

Generate Node Data from Link Data

SAS/GRAPH Network Visualization Workshop provides the capability to create a minimal node data set from the link data set that you provide. This minimal node data set can be useful when you plan to create statistical graphs based on link data and don't require node attributes for your analysis.

To generate node data from link data, do any one of the following:

- When you open a link data set, select **Generate node data** in the Open dialog box.
- When you create a new project, in the New Project dialog box, select the link data set and then select the **Create nodes** check box.
- After you have opened a link data set, select **Data** ▶ **Create Node Data**.

When you generate node data from link data, you are prompted to specify the FROM and TO variables in your link data set. For details, see "Specify Data Attributes" on page 17.

SAS/GRAPH Network Visualization Workshop creates a table that contains the nodes along with the following other variables:

- The Label variable determines the text that appears when you apply a label to nodes in a network graph.
- The In Arcs, Out Arcs, and Total Arcs variables are used to help keep track of the links that are associated with each node. The value for Total Arcs equals the value for In Arcs plus the value for Out Arcs.

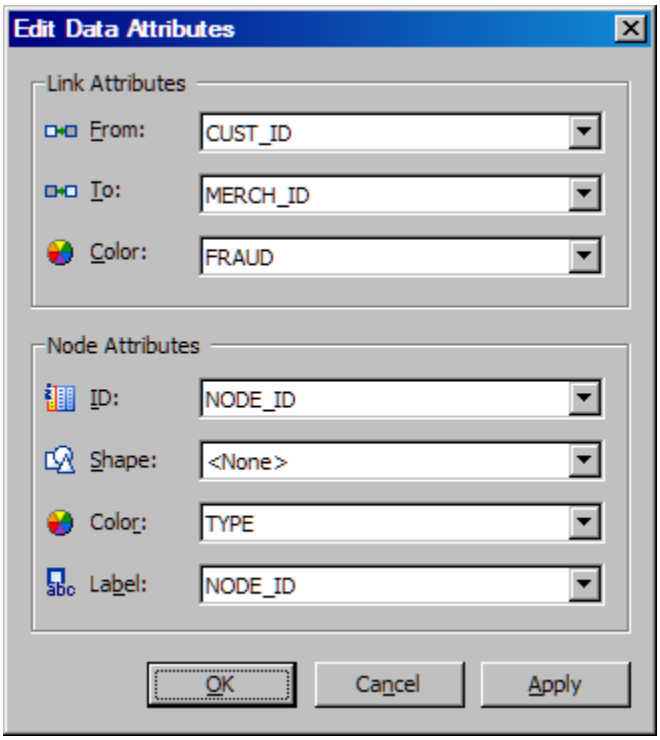
You cannot save the node data set by itself. However, if you save the data as part of a project, then the node data is preserved with the project.

Specify Data Attributes

Before you can analyze network data, you must specify which variables to use for node identification, link origin, and termination. For more information about these variables, see "About Data Sources" on page 11.

To specify attributes:

1 Select **Data** ▶ **Edit Attributes**. The Edit Data Attributes dialog box opens.



Note: This dialog box opens automatically in response to particular commands. For example, the dialog box opens if you attempt to create a network graph but haven't yet specified the required data attributes.

2 For link attributes, specify the following:

From — Select the variable that corresponds to the link's origin node.

To — Select the variable that corresponds to the node at which the link terminates.

Color — (Optional) Select the variable to use for link colors. The link colors vary based on the value of the variable that you select.

3 For node attributes, specify the following:

ID — Select the variable that identifies the nodes in the data set.

Shape — (Optional) Select the variable to use for the node marker shapes. The shapes vary based on the value of the variable that you select.

Color (Optional) Select the variable to use for the node marker colors. The colors vary based on the value of the variable that you select.

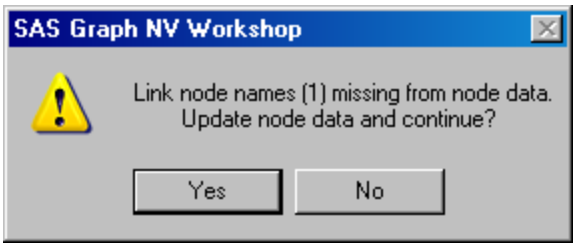
Label (Optional) Select the variable to use for the node labels. If you apply labels to nodes in a network graph, the labels are determined by the value of the variable that you select.

4 Click **OK**.

Synchronizing Data

About Data Synchronization

When you load network data or create a network graph, SAS/GRAPH Network Visualization Workshop synchronizes your node and link data. Synchronization ensures that the node data set contains all the nodes that are represented by the FROM and TO values in the link data set. If any nodes are missing from the node data set, then SAS/GRAPH Network Visualization Workshop displays the following prompt:



- If you choose **No**, then you are unable to create network graphs.
- If you choose **Yes**, then SAS/GRAPH Network Visualization Workshop adds the missing nodes to the data table. You can then create a network graph.

Regardless of your choice, the underlying data set is unchanged. The next time you update your data, it is recommended that you add the missing nodes to your node data set in order to avoid receiving this message.

The data must be validated before it can be synchronized. If you attempt to synchronize data that has not been validated, SAS/GRAPH Network Visualization Workshop validates the data first. For more information, see “Validating Data” on page 20.

Synchronize the Data

SAS/GRAPH Network Visualization Workshop synchronizes data automatically when you load network data or create a network graph. You can also synchronize your data manually.

There are two ways to synchronize data manually:

- Select the **Synchronize node and link data** option when you load the data. For instructions about loading data, see “Load Data into SAS/GRAPH Network Visualization Workshop” on page 14.
- Once the data has been loaded, you can select **Data** ▶ **Synchronize**.

If your data has already been synchronized, then these options are dimmed, indicating that they are not available.

Validating Data

About Data Validation

SAS/GRAPH Network Visualization Workshop validates data in order to ensure that the data contains the required variables. Validation also checks whether the data sets contain observations and whether there are duplicate nodes in the node data. If the node data contains duplicate nodes, SAS/GRAPH Network Visualization Workshop prompts you to either keep or delete the duplicates. Regardless of your choice, the underlying data set is unchanged. The next time you update your data, it is recommended that you remove the duplicate nodes from your node data set.

The data’s NODE ID, FROM, and TO variables must be assigned before the data can be validated. If you attempt to validate data whose variables have not been assigned, then SAS/GRAPH Network Visualization Workshop opens the Edit Data Attributes

dialog box. For details about using this dialog box, see “Specify Data Attributes” on page 17.

Validate Data

SAS/GRAPH Network Visualization Workshop validates data automatically when it synchronizes the data. You can also validate your data manually.

To validate data manually:

1. Load either or both of the node and the link data sets. For instructions, see “Load Data into SAS/GRAPH Network Visualization Workshop” on page 14.
2. Select **Data** ▶ **Validation**.

If your data has already been validated, then the menu item is dimmed, indicating that it is not available.

3

Working with Tables

About Data Tables

When you open a data set or a project, SAS/GRAPH Network Visualization Workshop displays the data in one or two tables. A table operates somewhat like a typical spreadsheet. You can edit data, sort data, and perform other tasks.

All views of the data in SAS/GRAPH Network Visualization Workshop (all tables and graphs) are linked through an underlying data model. Therefore, any changes or selections made in a table are also reflected in any other views of that same data.

Note: Changes made to the table remain in effect only while the current project is open. For example, you might change the value of an observation in a table. Then, when you close and later reopen the data sets or the project, your change is not saved.

You can display a context menu by right-clicking on a column heading or a row heading. The context menu varies depending on the cursor's position when you right-click.

Here is a sample table that is based on link data:

Figure 3.1 *Example Link Data Table*

1	2	3	4	5	
	5	Customer ID	Merchant ID	Merchant Category	From Index	To Index
1	■	Customer01	Merchant01	Retail	0	3
2	■	Customer01	Merchant02	Services	0	4
3	■	Customer02	Merchant01	Retail	1	3
4	■	Customer02	Merchant03	Medical	1	5
5	■	Customer02	Merchant04	Construction	1	6
6	■	Customer02	Merchant06	Banking	1	8
7	■	Customer02	Merchant05	Retail	1	7
8	■	Customer02	Merchant02	Services	1	4
9	■	Customer03	Merchant05	Retail	2	7
10	■	Customer03	Merchant02	Services	2	4
11	■	Customer03	Merchant06	Banking	2	8
12	■	Customer03	Merchant01	Retail	2	3
13	■	Customer03	Merchant04	Construction	2	6

1 The first column in the table contains the observation label, which is used to identify observation markers in graphs. You can change the labels for variables and observations. For details, see "Changing Header Labels" on page 28.

2 The second column contains marker icons that indicate whether an observation is included in graphs. If you exclude an observation from graphs, then its marker icon disappears. The shape and color of the marker icons indicate the following:

- In a node data table, a marker icon indicates the marker color and shape associated with the corresponding observation in graphs. The default color is blue, and the default shape is a square.
- In a link data table, a marker icon indicates the marker color associated with the corresponding observation in graphs. The default color is plum.

You can configure which variable is used to determine marker shapes and colors by setting data attributes. For details, see "Specify Data Attributes" on page 17.

The numeric value in the column heading indicates how many variables the data set contains. If you select one or more variables, then the value changes to indicate the number of selected variables, and the value is displayed within brackets.

3	The third and fourth columns function as the FROM and TO nodes for your links. Though the FROM and TO variables are the first variables in this example data set, they can correspond to any two columns of your data set. You specify the names for the FROM and TO variables when you set the data attributes.
4	Though not required, you can have additional variables that might be useful for sorting and exploring data.
5	When you load both link and node data, SAS/GRAPH Network Visualization Workshop adds two variables to the links table: From Index and To Index. These variables are derived from the FROM and TO variables in the data set. The derived variables are zero-based indexes into the node data set. Each variable corresponds to a node observation in the node table (the observation is the index value plus one). The variables enable you to look up a node quickly in the node data table rather than scanning through the node data table row by row.

See Also

"Selecting Data in a Table" on page 26

Selecting Data in a Table

Select Cells in a Table

Use any of the following methods to select one or more cells in a table:

- To select a single cell, click the cell.
- To select multiple cells, click and drag the cursor across the cells that you want to select.

Note: When you double-click a cell, the cell becomes editable.

Select Variables in a Table

Use any of the following methods to select one or more variables in a table:

- To select a single variable, click on the header for the variable that you want to select. The header and entire column are highlighted.
- To select multiple variables, press and hold the CTRL key and select the variables that you want. To select a block of contiguous variables, select the first variable in the block, press the Shift key, and then select the last variable in the block.
- To select a large number of variables, it might be easier to select a few variables that you don't want and then invert your selection. To invert a selection:

  1. In the data table, select one or more variables that you don't want selected.
  2. Right-click and choose **Invert Selection** from the pop-up menu. All currently selected variables become deselected, and all deselected variables become selected.

Select Observations in a Table

Use any of the following methods to select one or more observations in a table:

- To select a single observation, click on the header for the row that you want to select. The header and entire row are highlighted.
- To select multiple observations, press and hold the CTRL key, and select the rows that you want. To select a block of contiguous observations, select the first row in the block, press the Shift key, and then select the last row in the block.

  To select all observations, click the table and then press CTRL-A.
- To select a large number of observations, it might be easier to select a few observations that you don't want and then invert your selection. To invert a selection:

  1. In the data table, select one or more observations that you don't want selected.
  2. Right-click and choose **Invert Selection** from the pop-up menu. All currently selected observations become deselected, and all deselected observations become selected.

Note: When selecting observations, you might want to sort the data table first so the observations appear in a particular order. For instructions, see "Sorting a Table" on page 29.

Copy and Paste Table Data

You can copy and paste data from a table into another application such as Microsoft Excel or Microsoft Word.

To copy table data:

1. In a data table, select the data that you want to copy. You can select rows, columns, or cells. For instructions, see "Selecting Data in a Table" on page 26.
2. Select **Edit** ▶ **Copy**. The data is copied to your system clipboard.

You can paste the data into the target application using that application's paste command.

Changing Header Labels

Change the Label Used for Variable Headers

Variable headers appear in the first row of the table.

To change the label used for variable headers:

1. Select the data table that you want to change.
2. Select **Table View** ▶ **Variables**, and then select either of the following:
   - **Display Name** displays the name of each variable in the header row.

- **Display Label** replaces variable names with any descriptive labels that have been defined for the variables. This is the default setting. If no label exists in the data set for a variable, then the application displays the variable name.

Any change you make remains in effect only while the current project is open. If you close and then reopen the project, the change is not saved.

Change the Label Used for Observation Headers

Observation headers appear in the first column of the table.

To change the label used for observation headers, complete these steps:

1. Select the data table that you want to change.
2. Select **Table View** ▶ **Observations**, and then select either of the following:
   - **Label by Observation Number** displays observation numbers in the header column. This is the default setting.
   - **Label by Variable** opens a dialog box where you can select the variable to use for the header column.

     Additionally, you can hide the selected variable from the table in order to prevent it from being displayed twice. In the dialog box, select **Hide Label Variable**.

Any change you make remains in effect only while the current project is open. If you close and then reopen the project, then the change is not saved.

Sorting a Table

About the Sort Function

By default, a table displays data in the order that it appears in your data set. You can change the order by sorting the data based on the values of one or more variables.

Here are some characteristics of the sort function:

- The default behavior is to sort alphanumerically in ascending order (A to Z, zero to 9, or earliest to latest date). You can change the order when you perform a sort.
- Your sort can be based on more than one variable. For example, suppose that the data contains a variable for last name and another variable for first name. You can use last name as the primary sort and first name as the secondary sort. The last name variable has the higher priority in the sort.
- You can use a specialized order type to use in a sort. For example, if your data contains a category variable, you might want to sort by frequency of the variable's occurrences rather than sort alphabetically. For more information, see "Using a Specialized Sort Order" on page 31.

The sort order remains in effect only while the current project is open. If you close and then reopen the project, the sort order reverts to the default order.

Sort a Table

To sort a table:

1. (Optional) Select one or more variables to use in the sort.
2. Select **Table View** ▶ **Variables** ▶ **Sort**. The Sort dialog box opens. Note the following details about this dialog box:
   - If any variables were selected when you opened the dialog box, then those variables appear in the **Sort by** list.
   - Any variables that appear in the **Sort by** list are used in the sort.
   - The order in which variables are listed determines their sort priority.
   - In the **Sort by** list, an A or D precedes each variable name. These letters indicate whether the sort order is ascending or descending, respectively.
   - If an asterisk appears next to the A or D, then a specialized sort order has been specified for the variable. For details, see "Using a Specialized Sort Order" on page 31.

3 You can perform any of the following actions in the Sort dialog box:

- To sort by a variable, select the variable in the **Variables** list and click >>>. The variable is added to the **Sort by** list.
- To remove a variable from the **Sort by** list, select the variable and click <<<.
- To change the variable's ascending or descending sort status, select the variable in the **Sort by** list and click **Ascending** or **Descending**. You can also double-click the variable's name in the list.
- To change the sort priority of a variable, select the variable in the **Sort by** list and click **Up** or **Down**.

4 Click **OK**.

Using a Specialized Sort Order

Specialized Sort Order Overview

By default, the sort order is based on the alphanumeric value of the sort variable. You can specify other factors to use for a sort. For example, you can sort by frequency of a variable occurrence.

The sort function behaves differently depending on whether you have associated a format with the variable used for the sort. For example, suppose that an MMDDYY10 format has been applied to a date variable, and you choose to sort the formatted values rather than the numerical date values. In your results, sorted in ascending order, a date of 02/05/1982 comes before 07/07/1969 because 02 comes before 07.

The following figure shows a formatted sort on birthdate values.

Figure 3.2 *Table Sorted by Formatted Dates*

	3	Name	Age	Birth Date
1	■	Doe_J	25	02/05/1982
2	■	Davids_J	25	02/05/1982
3	■	Jones_S	25	02/05/1982
4	■	Scott_S	37	07/07/1969
5	■	Jones_D	15	07/07/1992
6	■	May_V	24	08/10/1982
7	■	Stark_T	24	08/10/1982
8	■	Thompson_T	22	10/04/1984
9	■	Clark_S	33	10/10/1973
✳				

For some sort orders, application of a format produces the same result as no format. In the previous example, if you perform a frequency sort on the date variable, then the formatted results are the same as the non-formatted results. A date that appears three times in the data set also appears three times in the sorted table regardless of whether the date is formatted.

The following figure shows the frequency of birthdate values formatted as MMDDYY10. The sort order is the same for non-formatted values.

Figure 3.3 *Table Sorted by Frequency of Formatted Dates*

	3	Name	Age	Birth Date
1	■	Doe_J	25	02/05/1982
2	■	Davids_J	25	02/05/1982
3	■	Jones_S	25	02/05/1982
4	■	May_V	24	08/10/1982
5	■	Stark_T	24	08/10/1982
6	■	Thompson_T	22	10/04/1984
7	■	Clark_S	33	10/10/1973
8	■	Jones_D	15	07/07/1992
9	■	Scott_S	37	07/07/1969
✳				

However, if you apply a format such as DOWNAME to the variable, then the results differ. In this case, you might see three observations grouped together with a Friday date, three observations for Tuesday, and so on.

The following figure shows a frequency sort on the same data, but with birthdate values formatted as DOWNAME. In this case, the format affects the sort order.

Figure 3.4 *Table Sorted by Frequency, Dates Formatted as DOWNAME*

	3	Name	Age	Birth Date
1	■	Doe_J	25	Friday
2	■	Davids_J	25	Friday
3	■	Jones_S	25	Friday
4	■	May_V	24	Tuesday
5	■	Stark_T	24	Tuesday
6	■	Jones_D	15	Tuesday
7	■	Thompson_T	22	Thursday
8	■	Clark_S	33	Wednesday
9	■	Scott_S	37	Monday
*				

Use a Specialized Sort Order

To use a specialized sort order:

1 Select the variable that you want to use for your sort.

2 Right-click on the variable, select **Ordering** from the pop-up menu, and then select one of the ordering options shown in the following table:

Option	Type of Sort	Applies Format*
Standard	This option is the typical alphanumeric sort. This option is the only ordering available for interval variables. If necessary, change the variable's measurement level to nominal. See "Change a Variable's Measurement Level" on page 36.	no
by Format	Sort order is based on the variable's format. For example, if the data contains a date variable formatted as MMDDYY10., the formatted values are sorted rather than the numerical date values.	yes
by Frequency	Sort order is based on the frequency of variable values.	yes

Option	Type of Sort	Applies Format*
by Data	Data items are collated and sorted by the order in which data appears in the selected variable of the data set.	yes
by Frequency (unformatted)	Sort order is based on the frequency of variable values.	no
by Data (unformatted)	Data items are collated and sorted by the order in which data appears in the selected variable of the data set.	no
Custom	This option is not used and is therefore dimmed.	not applicable

* The Applies Format column is relevant only if you have associated a format with the variable used for the sort. This column indicates whether the sort operation is performed on the formatted value (Yes) or on the raw numeric values (No).

3 Sort the table. For instructions, see "Sort a Table" on page 30. When you sort the table, your specialized sort order is used.

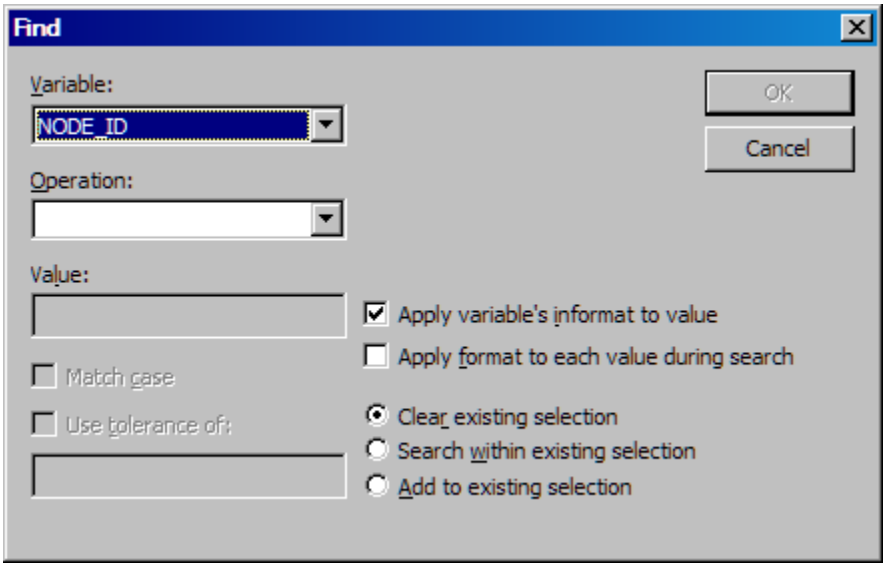
Find Specific Data in a Table

Use the search feature to locate specific data in a table.

To search for data:

1 Select the table that you want to search.

2 Select **Edit** ▶ **Find**. The Find dialog box opens.



3 From the **Variable** list box, select the variable that contains the data you want to find.

4 From the **Operation** list box, select the operator for your search. For example, if you want to locate all observations that have a value less than 100 for a particular variable, then you would select **Less than**.

5 In the **Value** field, depending on which operator you selected, enter the value that completes your search.

6 Select any of the other options that are applicable to your search. For example, if you are searching for a character string, then you can select **Match case**. If you are searching for a numeric value, then you can set a tolerance that indicates the amount of deviation from the number you want to allow.

7 Click **OK** to perform the search.

Enable Automatic Scrolling in Tables

When you select observations in network or statistical graphs, the corresponding observations are also selected in one or more of the data tables. However, you might need to scroll down a table in order to see the selections. You can enable automatic scrolling to an observation that is selected in a table.

To enable automatic scrolling:

1 Select the table that you want to scroll.

2 Select **Table View** ▶ **Auto Scroll**.

Now, if you select observations in a graph, the table automatically scrolls down to display the first selected observation. The **Auto Scroll** option stays selected until you deselect it.

Change a Variable's Measurement Level

The measurement level of a variable is a classification that describes the type of data that a variable contains. Variables can be numeric or character depending on their measurement levels. In SAS/GRAPH Network Visualization Workshop, the measurement level can be one of the following measurement levels:

Nominal	Variables can be either numeric or character.
Interval	Variables must be numeric.

To change the measurement level of a variable:

1 Select the variable in the data table.

2 Right-click and select either **Interval** or **Nominal**.

Your change remains in effect only while the current project is open. If you close and then reopen the project, then your change is not saved.

Hide a Variable

To hide a variable:

1. In the data table, select the variable that you want to hide.
2. Select **Table View** ▶ **Variables** ▶ **Hide**.

Your change remains in effect only while the current project is open. If you close and then reopen the project, the variable reappears in the table.

To restore a variable that you have hidden, select **Table View** ▶ **Variables** ▶ **Show All**.

Exclude an Observation from Graphs

To exclude an observation from graphs:

1. Select one or more observations that you want to exclude.
2. Select **Table View** ▶ **Observations** ▶ **Exclude from Plots**.

   The observation's marker disappears from the second column of the table. The observation is excluded from network graphs and from scatter plots.

Your change remains in effect only while the current project is open. If you close and then reopen the project, the observation is included in the graphs.

To restore an observation that you have excluded, select the observation and then select **Table View** ▶ **Observations** ▶ **Include in Plots**.

Change the Marker's Appearance for One or More Observations

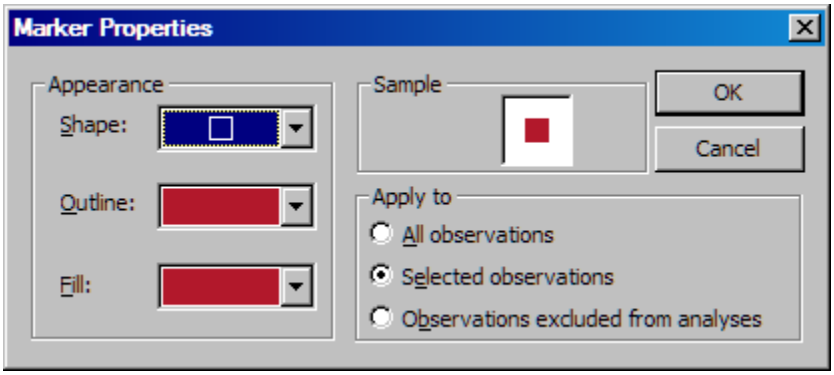
A node marker is used to identify the node in a data table. You can change the shape, outline, and fill used for node markers. When you change the node marker's appearance, you can apply your changes to one of the following observations:

- all observations
- selected observations

Note: Any change you make here remains in effect only while the current project is open. If you close and then reopen the project, your change is not saved.

To change a marker's appearance:

1. If you want to apply changes to selected observations, then select one or more observations that you want to change.
2. Select **Table View** ▶ **Observations** ▶ **Marker Properties**. The Marker Properties dialog box opens.



3. In the **Appearance** box, select the shape, outline color, and fill color that you want for the marker.

   To select a color, select **Set color** from the **Outline** or **Fill** list box. This action opens the Color dialog box, which contains colors to select from and the option to define

your own color. For help using this dialog box, see “Using the Color Dialog Box” on page 62.

4 In the **Apply to** box, select whether you want the changes applied to all observations or to just the selected observations. (The **Observations excluded from analyses** option is not supported.)

5 Click **OK**.

The changes described here override settings that have been made in the Edit Data Attributes dialog box. For information about the Edit Data Attributes dialog box, see “Specify Data Attributes” on page 17.

Examine the Variable Attributes for Selected Observations

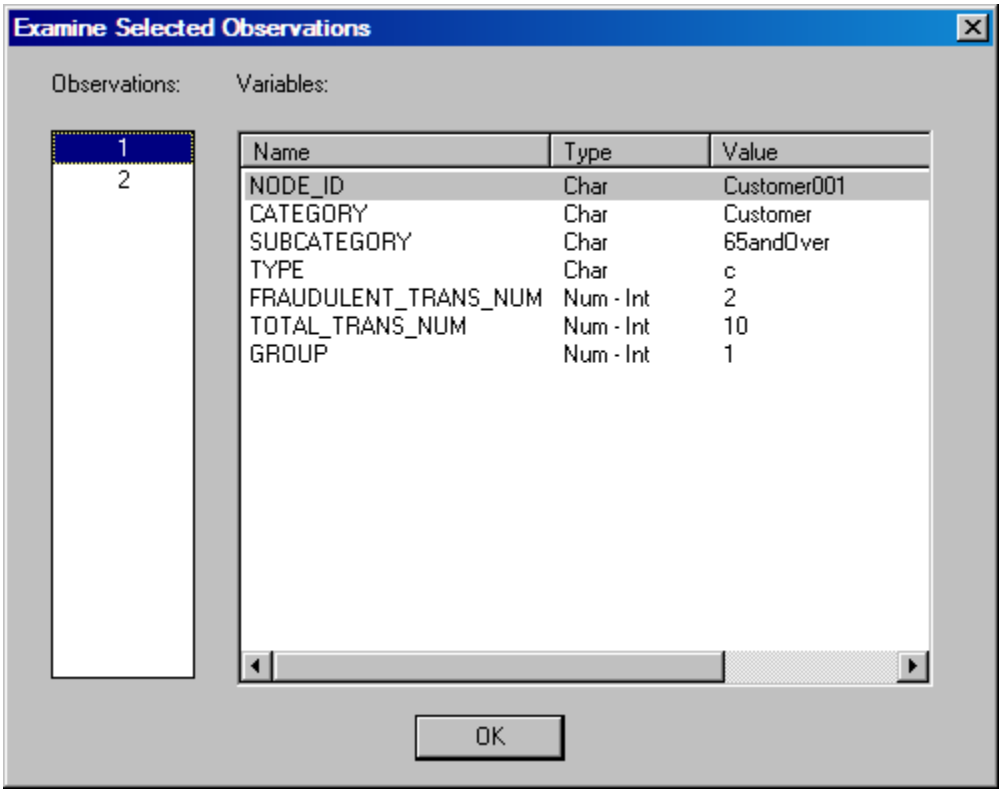
Variable attributes for observations include the following:

- the name of the variable
- the variable type (character or numeric)
- the current variable value

To examine variable attributes:

1 Select one or more observations in a data table.

2 Select **Table View** ▶ **Observations** ▶ **Examine Selected Observations**. The Examine Selected Observations dialog box opens.



3 Select an observation number on the left side of the dialog box. Its variable attributes are displayed on the right side.

4 Repeat the previous step to examine variable attributes for other observations, if applicable.

5 When you are finished examining variable attributes, click **OK**.

Reopen a Table

If you close a table that has an associated statistical graph, then you can easily reopen the table.

To reopen a table:

1. Select any statistical graph that is based on the table that you closed.
2. Select **View** ▶ **Show Table**. The table opens.

4

Creating and Modifying Statistical Graphs

About Statistical Graphs

After you have loaded your data into SAS/GRAPH Network Visualization Workshop, you can investigate the data with a variety of statistical graphs. A statistical graph is based on either link data or node data.

Statistical graphs serve two distinct roles in SAS/GRAPH Network Visualization Workshop:

- The graphs enable exploratory data analysis on either the link or the node data. Distributional patterns or correlations in the data can often be detected using these graphs. For a summary of tasks that you can perform, see “Non-Network Data ” on page 14.
- The graphs can provide data filters for network data. Since all views (tables or graphs) of a data set are linked, observation selections in one graph are reflected in all graphs using that data set. Using the statistical graphs, you can selectively filter the observations displayed in the network graphs to uncover important relationships between nodes.

You can create the following types of statistical graphs:

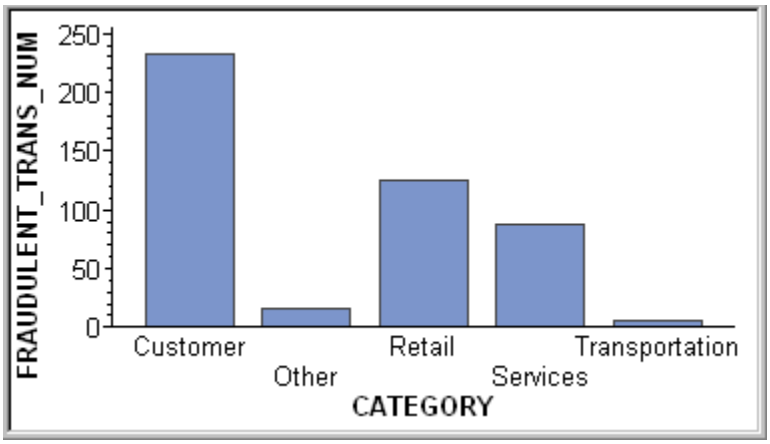
See Also

- “Create a Statistical Graph” on page 47
- “Changing the Properties of a Statistical Graph” on page 49

Types of Statistical Graphs

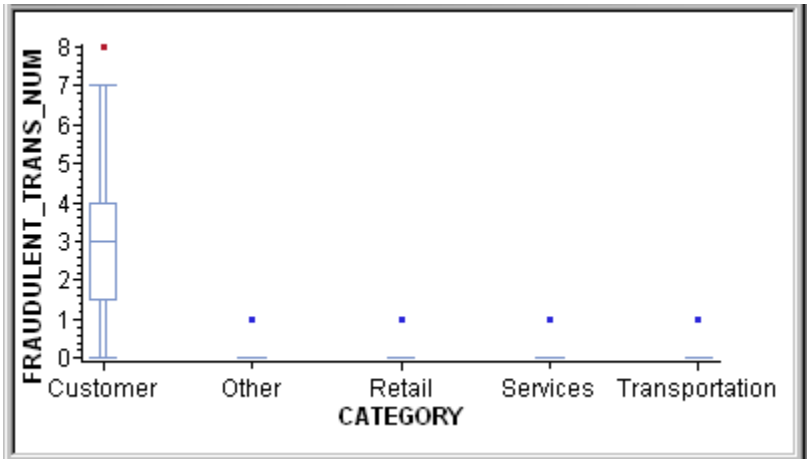
Bar Charts

A bar chart consists of a grid and some vertical or horizontal columns (bars). Each column represents quantitative data.



Box Plots

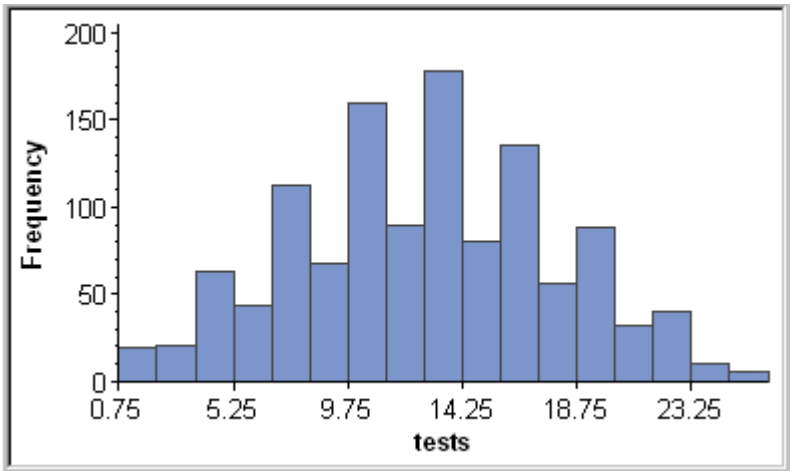
A box plot displays summary statistics for the distribution of values for a variable. The outer bounds of the box represent the first and third quartiles. The line inside the box represents the median. The markers outside the box, referred to as outliers, represent data points that are outside of the 25th and 75th percentiles.



Histograms

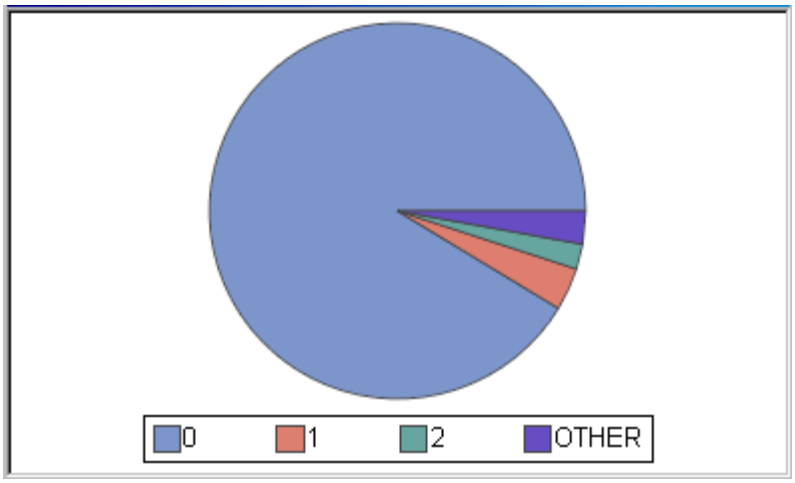
A histogram is a bar chart that displays the observed frequencies of data that have been binned (divided into contiguous, equally spaced intervals). The heights of the bars

indicate the relative frequency of observations in each bin. Histograms can also show binned response data if you choose a response variable other than Frequency.



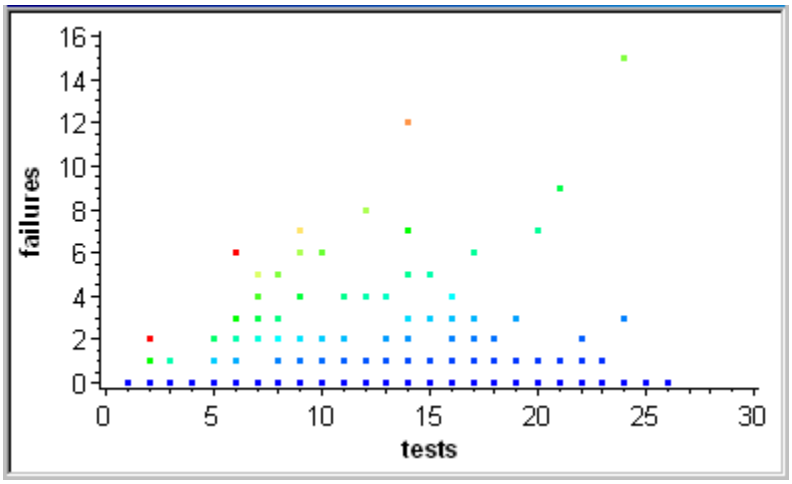
Pie Charts

A pie chart is a circular chart that is divided into slices by radial lines. Each slice represents the relative contribution of each part to the whole.



Scatter Plots

A scatter plot is a two- or three-dimensional plot that shows the joint variation of two (or three) variables from a group of observations. The coordinates of each point in the plot correspond to the data values for a single observation.



Create a Statistical Graph

To create a statistical graph:

1. Select the data table on which you want to base the graph. A graph can be based on either link or node data. Alternatively, if you already have a graph open, you can select that graph and base the new graph on the selected graph.

2. From the **Graphs** menu, select the type of graph that you want to create. For an overview of the types, see "About Statistical Graphs" on page 43.

3. Select the data variables to be used to create the graph. The information that you provide varies with the type of graph, as described in the following table:

Bar chart	Select the category variable (X axis) from the **Category** list box. Only character data type variables are available. The default selection is the first variable in the data table. Select the measure variable (Y axis) from the **Response** list box. Only numeric data type variables are available. If you choose the default selection <Frequency>, then the application displays a frequency chart for the Category variable. (Optional) If you want the bar chart to have an additional subgroup category, then select the variable from the **Subgroup** list box. Numeric and character data type variables are available. The default selection is <None>.
Box plot	Select the category variable (X axis) from the **Category** list box. Only character data type variables are available. The default selection is the first variable in the data table. Select the measure variable (Y axis) from the **Response** list box. Only numeric data type variables are available.
Histogram	Select the X variable for the histogram from the **X Variable** list box. Only numeric data type variables are available. The default selection is the first numeric variable in the data table. Select the Y variable for the histogram from the **Y Variable** list box. Only numeric data type variables are available. If you choose the default selection <Frequency>, then the application displays a frequency histogram for the X variable.
Pie chart	Select the category variable from the **Category** list box. Numeric and character data type variables are available. The default selection is the first variable in the data table.
Scatter plot	Select the X variable for the plot from the **X Variable** list box. Only numeric data type variables are available. The default selection is the first numeric variable in the data table. Select the Y variable for the plot from the **Y Variable** list box. Only numeric data type variables are available. (Optional) Select the Z variable for the plot from the **Z Variable** list box. Only numeric data type variables are available. The default selection is <None>.

Changing the Properties of a Statistical Graph

Change the Properties of a Statistical Graph

The properties of a statistical graph determine features that affect the overall appearance of the graph. These properties include (but are not limited to) the following items:

- the graph's style and colors
- the orientation and appearance of the graph's axes
- whether and where the graph displays a legend
- whether and where a graph displays titles

Note: Any changes that you make to a graph's properties remain in effect only while the current project is open. If you close and then reopen the project, the properties revert to their default values.

To change a statistical graph’s properties:

1. Right-click on the graph you want to change.
2. From the pop-up menu, choose **Graph Properties**. The Properties dialog box opens.
3. Change the settings for the graph.

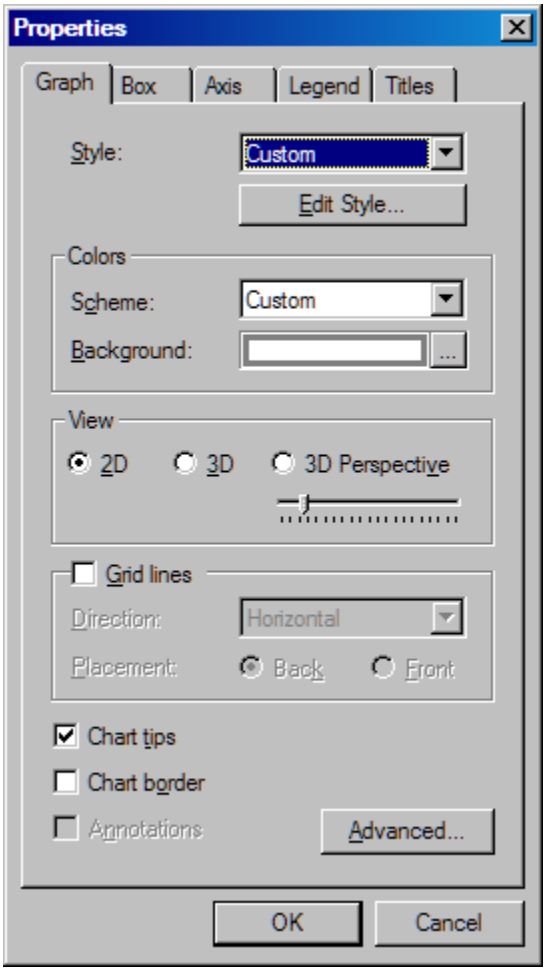
   The Properties dialog box contains several tabs that organize the properties you can change. In general, there are two types of properties:

   - The **Graph**, **Axis**, **Legend**, and **Titles** tabs contain properties that all or most statistical graphs have in common. For a description of these properties, see “General Properties” on page 50.

- The dialog box also contains a tab that has graph-specific properties. The name and contents of this tab vary depending on the type of graph you are modifying. For a description of these properties, see "Graph-Specific Properties" on page 64.

General Properties

Graph Tab



The **Graph** tab in the Properties dialog box contains the following properties:

Style

Select a predefined graph style name from the list box. The style that you select overrides the style that has been selected for the project. For details about the project style, see "Apply a Style to a Project" on page 114.

You can also select **Edit Style** and then customize the currently selected style. For instructions, see "Customize the Style Applied to a Graph" on page 59.

Colors

Select a color scheme from the **Scheme** list box to change the colors of the graph.

You can also select ... to open the Background dialog box, where you can choose the background color, gradient, and texture. For information about this dialog box, see "Changing Background Color, Gradient, and Texture" on page 62.

View

Select whether you want to view the graph in two-dimensional, three-dimensional, or three-dimensional perspective format. If you select **3D Perspective**, then you can use the **Perspective** slider to increase or decrease the depth of the graph. This format is not available for pie charts.

Grid lines

Select **Grid lines** to display grid lines for the graph. If you display grid lines, then you can make the following changes:

Direction

Select **Horizontal**, **Vertical**, or **Both** from the **Direction** list box.

Placement

Change the placement of the grid lines from background to foreground by selecting the appropriate button.

The grid lines feature is not available for pie charts.

Chart tips

Select or clear the **Chart tips** check box to display or hide chart tips. Chart tips contain data details and appear when you move the cursor over particular areas on a graph. For example, moving the cursor over the bars in a bar chart displays the data values associated with the bars.

Chart border

Select **Chart border** if you want a border around the graph. This option is not available for pie charts.

Annotations

This option is not used and is therefore dimmed.

Advanced settings

Click the **Advanced** button to open the Advanced Settings dialog box, which contains the following items:

Mouse Preferences

Use the radio buttons to select either graph-style rotation around the X and Y axes or map-style, three-dimensional arcball rotation around all three axes (X, Y, and Z). To use arcball rotation, imagine a virtual sphere surrounding the object, select a point on the sphere, and rotate that point to a new location. If you select a position outside the sphere, then the object is rotated around the Z axis only. Arcball rotation is most useful (and the default setting) for maps.

PowerPoint Sessions

Use the check boxes to determine the behavior when you copy and paste a graph into PowerPoint. The check boxes enable you to display interactive graphs and to remove the graph background.

Scrollbars

Select or clear the **Scrollbars** check box to control whether scroll bars appear for a graph. Though scroll bars are enabled, the threshold for displaying scroll bars varies with the type of graph. You can control the threshold for displaying scroll bars by changing the values in the **Horizontal** and **Vertical** list boxes. (If you select the zoom tool and zoom in on any of the graphs, the scroll bars appear.)

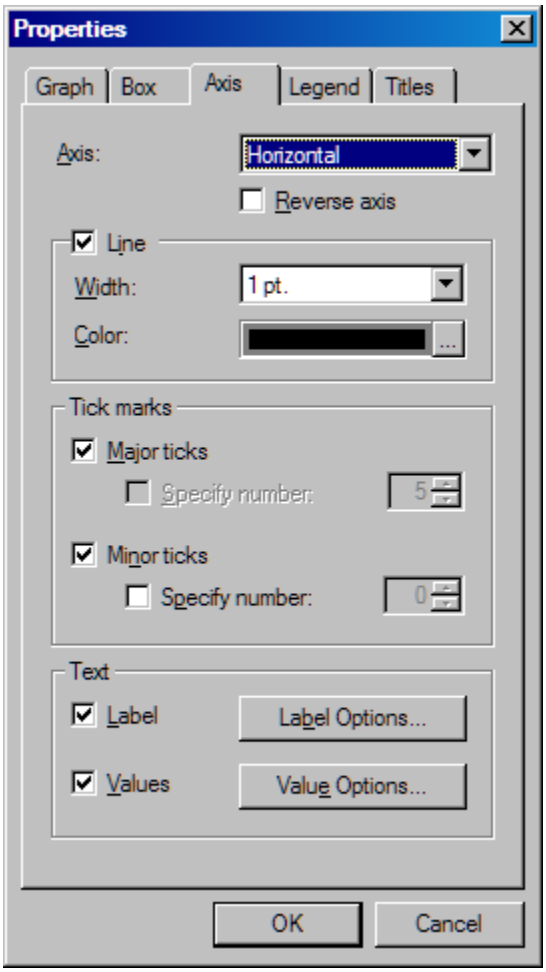
Graph Rendering

Use the slider to control graph quality. The rendering quality slider affects texture mapping, line anti-aliasing, and scene anti-aliasing.

Lighting

Use the sliders to control the level of direct and ambient light. Ambient light is light coming from all directions that casts no shadows. Direct light is light emitting from a point light source that provides light direction and shadow effects.

Axis Tab



The **Axis** tab in the Properties dialog box enables you to control the display and formatting of axes. These properties are not available for pie charts. The **Axis** tab contains the following items:

Axis

From the list box, select the axis whose properties you want to change.

Select **Reverse Axis** to move the data origin to the opposite side of the selected axis.

Line

Select or clear the **Line** check box to control whether the specified axis line is displayed or hidden.

Use the **Width** list box to set the line width from one point to six points. The default width depends on the graph type.

To change the color of the line, click ... to open the Color dialog box, which contains colors to choose from and the option to define your own color. See "Using the Color Dialog Box" on page 62.

Tick marks

Select the **Major ticks** and **Minor ticks** check boxes to display major and minor tick marks.

To control the number of tick marks that are displayed for either axis, select the **Specify number** check box and enter the number of tick marks that you want.

Note: You cannot change the number of major tick marks for bar charts that show the category axis.

Text

Control axis labels and values using the following items:

Labels

Select or clear the **Labels** check box to control whether axis labels are displayed. For graphs that use a group variable, you can also display or hide the **Group** values.

Label Options

Change the label text or font by selecting the **Label Options** button. In the Label Options dialog box, you can specify the text of axis labels by selecting **Custom label** and then entering the desired label text into the text box.

To change the font, select the **Font** button, and then use the Font dialog box to specify a font type and font characteristics such as size and color.

Values

Use the **Values** check box to control whether axis values are displayed. For graphs that use a group variable, you can also display or hide the **Group** values.

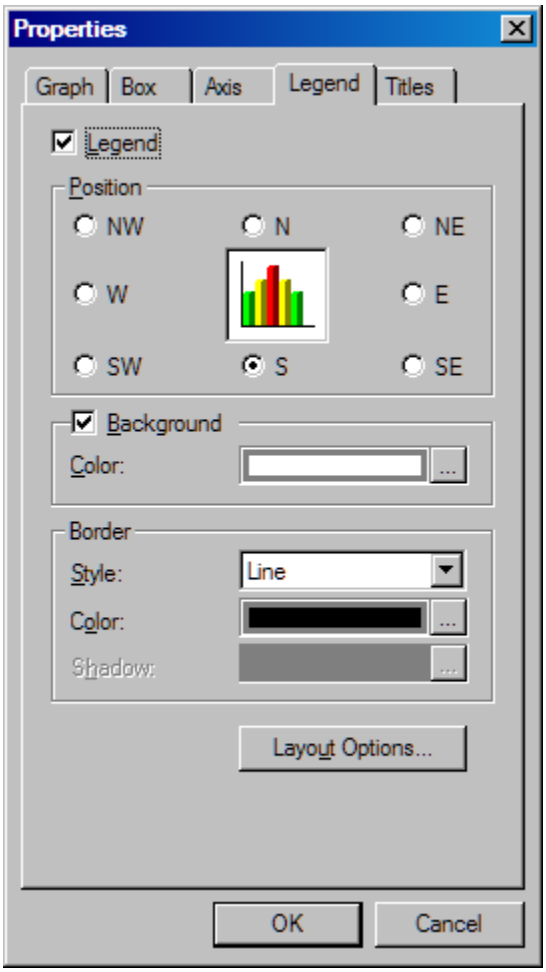
Value Options

Change the layout or the font of axis values by selecting the **Value Options** button. In the Value Options dialog box, select one of the following layouts from the **Arrange values** list box:

- **Automatically** specifies an optimized fit, as determined by the graph.
- **Horizontally** specifies that the text layout is horizontal.
- **Vertically** specifies that the text layout is vertical.
- **Angle 45** specifies that the values are rotated at a 45 degree angle.
- **By offsetting (stagger)** specifies that the text is staggered to fit.
- **By abridging (thin)** specifies that the values are thinned out to fit (a subset of values is displayed).

To change the font of the axis values, click **Font** in the Value Options dialog box. Then use the Font dialog box to select a font type and font characteristics such as size and color.

Legend Tab



You can change the legend position, background, border, or text in the **Legend** tab of the Properties dialog box. The **Legend** tab contains the following items:

Legend

Select or clear the **Legend** check box to control whether the legend is displayed for your graph.

Position

Set the legend's position relative to the graph by selecting one of the position choices. For example, choose **NW** to position the legend in the top left corner of the graph or **S** to position the legend in bottom center of the graph.

Background

The legend's background is the area within its border. Select or clear the **Background** check box to control the legend background's display. Click ... to open a Legend dialog box that enables you to change the background's color, gradient, and texture. For details, see "Changing Background Color, Gradient, and Texture" on page 62.

Border

Select a border style from the **Style** list box. Choose **None** to remove the border. After you choose a style, you can select the border color. If the style that you choose uses a shadow, then you can also choose a color for the shadow. Click ... to open the Color dialog box. The Color dialog box contains colors to choose from and the option to define your own color for the border and shadow. For details, see "Using the Color Dialog Box" on page 62.

Layout Options

When you select the **Layout Options** button, the Layout Options dialog box opens. This dialog box contains the following items:

Title

Enter the title that you want for the legend. To hide the legend title temporarily, clear the **Title** check box.

In the **Position** list box, select a directional position or select **Automatic** to specify that the position be automatically assigned.

To change the font that is used for the legend's title text, click the **Font** button. Then use the Font dialog box to choose a font type and font characteristics such as size and color.

Labels

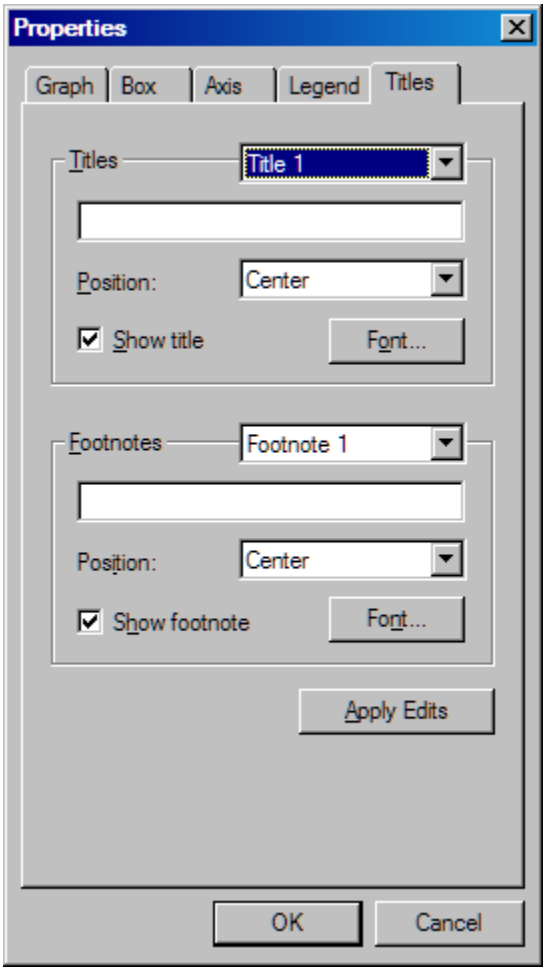
To set the label's position relative to the graph, select one of the position choices from the **Position** list box. Select **Automatic** to specify that the position be automatically assigned.

The data in the legend can be placed in rows or columns. From the **Layout** list box, select rows or columns, or select **Automatic** to specify that the layout be automatically assigned. If you specified rows or columns as the layout of the

legend, then you can use the **Numbers** list box to specify the number of rows or columns in the legend.

To change the font that is used for the label text, click the **Font** button. Then use the Font dialog box to choose a font type and font characteristics such as size and color.

Titles Tab



The **Titles** tab in the Properties dialog box enables you to control the display of titles and footnotes. The **Titles** tab contains the following items:

Titles

You can specify the text, position, and font for up to four titles. Select which title you want to edit from the list box, and enter the text for that title in the text box.

Select the location for the title from the Position list box. You can locate the title in the center, left, or right portion of the display.

Click the **Font** button to open the Font dialog box. The Font dialog box enables you to specify the font type and font characteristics such as size and color for your title.

If you want to temporarily hide a title that you have defined, then clear the **Show title** check box. You can select the check box again when you are ready to display the title.

Footnotes

You can specify the text, position, and font for up to two footnotes. Select which footnote you want to edit from the list box, and enter the text for that footnote in the text box.

Select the location for the footnote from the **Position** list box. You can locate the footnote in the center, left, or right portion of the display.

Click the **Font** button to open the Font dialog box. The Font dialog box enables you to specify the font type and font characteristics such as size and color for your footnote.

If you want to temporarily hide a footnote that you have defined, then clear the **Show footnote** check box. You can select the check box again when you are ready to display the footnote.

Apply Edits

Select the **Apply Edits** button to apply any changes that you make to a title or footnote.

Customize the Style Applied to a Graph

You can customize the style that is used for a graph by entering changes in the Edit Style dialog box. To open this dialog box:

1. Select the graph whose style you want to change.

2 Right-click and choose **Graph Properties**. The Properties dialog box opens.

3 Select the style you want to customize from the **Style** list box.

4 Click **Edit Style**.

You can change the following components of your graph using the Style Editor dialog box:

Style Name

Select a different style from the list box, if appropriate. The style that you select overrides the style that has been selected for the project. (For details about the project style, see "Apply a Style to a Project" on page 114.)

Colors

Select a scheme from the **Scheme** list box to change the graph color scheme.

Click ... to open a dialog box where you can change the background color, gradient, and texture. For information about using this dialog box, see "Changing Background Color, Gradient, and Texture" on page 62.

Chart tab

Use the selection buttons ... to open separate dialog boxes to change the legend, floor, and wall fill color, gradient, and texture. For help using these dialog boxes, see "Changing Background Color, Gradient, and Texture" on page 62. For the chart's walls and data element fill, you can use the **Transparency** slider to increase (move to the right) or decrease (move to the left) the level of transparency.

Click a data element fill color from the color palette and then select **Edit** to open the Data Element dialog box. This dialog box enables you to change the color, gradient, and texture for the data element fill values. Clicking the **Edit All** button enables you to select and apply either or both of the following:

- texture file and a fit (for the selected texture file) to the entire chart. Select the **Texture** check box to specify a texture image file and file fitting. Use the **File** list box to select a predefined image. To select an image from your file system, click the file selection button ..., and then choose the image from the Open window. Use the **Fit** list box to specify a fitting setting for your texture file.

- color ramp. You can select a predefined color ramp or define your own color ramp by selecting starting and ending colors. The predefined color ramps available in the **Scheme** list box are lighter to darker colors that are used to indicate lower and high data values respectively. For example, if **White to Orange** is the selected scheme, then the lowest data values are white. The middle data values are light to darker orange. The highest data values are the darkest shade of orange. Select the **Color ramp** check box to specify a color ramp. Use the **Scheme** list box to select a predefined color ramp scheme. Click on the **Start** and **End** selection buttons ... to open the Color dialog box, where you can select start and end colors for the color ramp. For help using the Color dialog box, see "Using the Color Dialog Box" on page 62.

Note: For the lowest data values, you cannot specify a color for three-dimensional surfaces that are transparent or textured.

Lines tab

Click the **Color** selection buttons ... for **Outlines**, **Grid**, and **Axis and borders** to open the Color dialog boxes for each. The Color dialog boxes enable you to select the color for the specified line type. Click the **Width** list box to select line widths for the different types of lines. For both the **Grid** and **Axis and borders**, you can also select a defined line style from the **Style** list box.

Text tab

Click the **Color** selection buttons ... for Title, Labels, and Values to open a Color dialog box for each. The Color dialog boxes enable you to select the color for specified text type. Click on the **Font** buttons to specify a font for titles, labels, and text data values. Use the **Shadow** and **Border** check boxes to select or clear text shadows and borders.

Effects tab

Use the Style list boxes to select a defined style for the **Shadows**, **Text borders**, and **Selection**. Click the **Color** selection buttons ... to open the Color dialog boxes for each effect. The Color dialog boxes enable you to select the color for the specified effect. For shadows, use the arrows to the right of **Offset** to assign a shadow offset value. For text borders, you can select a border width from the **Width** list box.

Using the Color Dialog Box

Several tabs in the Properties dialog box provide a button that opens the Color dialog box. The Color dialog box contains colors to choose from and the option to define your own color.

To select a predefined color, select a color in the Basic colors section and then click **OK**.

Clicking the **Define Custom Colors** button extends the Color window and adds a color palette, a vertical slider thumb, a preview box, and several text boxes. To define a color, you can use the methods listed below singly or in combination:

- Position the target on a color you would like to define. Use the vertical slider thumb to increase (move up) or decrease (move down) the red, green, blue, and lumination values. As the slider thumb is moved, the values in the **Red**, **Green**, **Blue**, and **Lum** text boxes increase or decrease to reflect the slider thumb's position. When the slider thumb is moved or values are entered into any of the text boxes, the color that you are defining is displayed in the **Color Solid** preview box. Once the color has been displayed in the preview box, click the **Add to Custom Colors** button. The color is then available for selection on the left side of the Color dialog box in the **Custom colors** section.
- Use the **Hue, Saturation, Lum, Red, Green**, and **Blue** text boxes to define your color. As you enter values into these text boxes, the target and slider thumb move to positions reflecting the color being defined. When the color you want to define is displayed in the **Color Solid** preview box, click the **Add to Custom Colors** button. The color is then available for selection on the left side of the Color dialog box in the **Custom colors** section.

You can select a color that you have defined by clicking on the color in the **Custom colors** section of the Color dialog box.

Changing Background Color, Gradient, and Texture

Several tabs in the Properties dialog box provide a button that opens a dialog box containing three tabs: **Solid Color**, **Gradient**, and **Texture**. Open this dialog box to change the background appearance for a graph. You can change the following items:

Solid Color tab

Select a solid color from the **Color** palette. Click the **More** button to open the Color dialog box. The Color dialog box contains additional colors to select from and the option to define your own color. For details, see "Using the Color Dialog Box" on page 62.

Gradient tab

A gradient blends from one color to another on your graph. The **Preview** box shows you how the blend looks.

Use the **Type** list box to set the type of gradient. The gradient types available are dependent on the graph generated.

To set the direction for the color blend, choose **Left to right**, **Right to left**, **Top to bottom**, or **Bottom to top** from the **Directions** list box.

To select the starting and ending colors, use the **Start color** and **End color** buttons to open the Color dialog box for each. For details, see "Using the Color Dialog Box" on page 62.

Texture tab

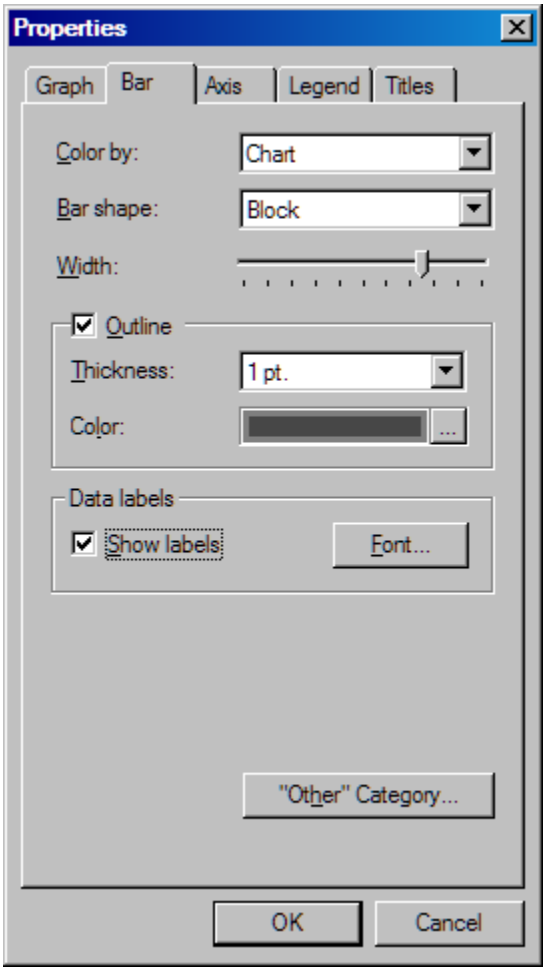
Select a background image. The **Preview** box shows you the image that is currently selected in the **File** list box. To select an image from your file system, click ..., and then choose the image from the Open dialog box.

Use the **Fit** list box to select **Tile** or a relative alignment for your image such as **Center** or **Right**.

The **Color fill** buttons enable you to specify that the image and the color fill are blended or that the image replaces the color fill. Select **Blend** to blend the color selected in the **Solid Color** tab with the image. Select the **Replace** radio button to specify that the image replaces the color selected in the **Solid Color** tab.

Graph-Specific Properties

Bar Tab



The **Bar** tab in the Properties dialog box contains the following items for customizing bar charts:

Color by

Specify how colors are used in the graph. Select one of the following items from the **Color by** list box:

- **Chart** assigns the same color to all axis variables in the chart.

- **Category** assigns a different color to each category axis variable.
- **Subgroup** assigns the same color to each subgroup (for example, each segment within a bar).

Except for **Chart**, the colors are determined by the role assigned to each variable. For example, if you select **Category**, then the bar colors are determined by the variable that is selected as the category variable.

Bar shape

Specify the shape of the bars. Select **Block, Cylinder, Hexagon, Prism**, or **Star** from the **Bar shape** list box.

Width

Set the width of the bars. To increase the width of the bars, drag the slider thumb to the right or click to the right of the thumb. To decrease the width of the bars, drag the slider thumb to the left or click to the left of the thumb.

Outline

Select or clear the **Outline** check box to control whether bars are outlined. Use the **Thickness** list box to change the thickness of the bar outline. Click the **Color** selection button to open the Color dialog box. The Color dialog box contains additional colors to select from and the option to define your own color for the bar outline. See "Using the Color Dialog Box" on page 62.

Data labels

Select the **Show labels** check box to display data labels. Data labels indicate the value of the variable represented by each bar.

To change the font of the data labels, click **Font** and then use the Font dialog box to select a font type and font characteristics such as size and color.

Show subgroups as

Select one of the buttons to specify whether the subgroups are shown as bars stacked horizontally or as bars grouped vertically. These buttons are available only when the graph contains subgroups.

"Other" category

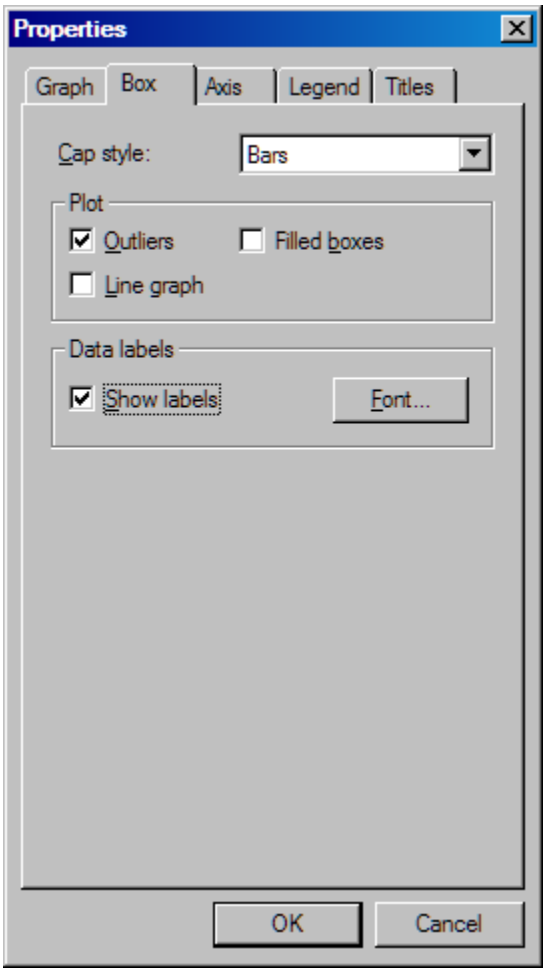
Select the **"Other" category** button to specify display parameters for categories that are four percent or less of the total bar chart. The "Other" Category dialog box opens. You can make the following changes in this dialog box:

Select or clear the **Show smaller values as "Other"** check box to control whether to show smaller values with the label of "Other."

Specify how you want the application to determine which values to include in "Other."

- Values can be determined based on a maximum number of categories that you select from the **Number** list box.
- Values can be determined based on the value's percentage of the total. You select the maximum percentage from the **Value** list box.

Box Tab



The **Box** tab in the Properties dialog box contains the following items for customizing box plots:

Cap style

Change the style of the box caps. You can select **None**, **Whiskers**, **Brackets**, or **Bars** from the **Cap Style** list box.

Plot

Use the check boxes to add or remove the following options:

Outliers

Toggle the display of outliers on and off.

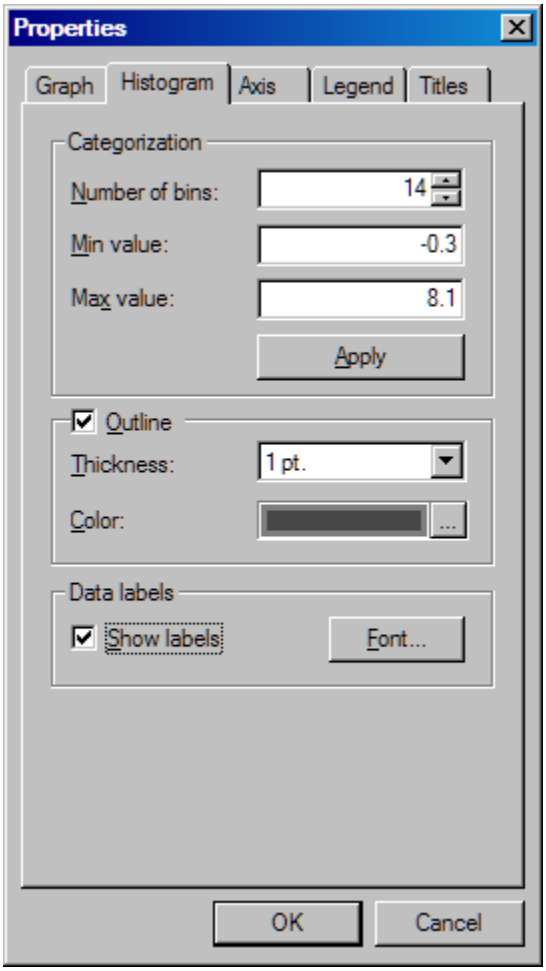
Line graph
: Toggle the display of a line joining the boxes.

Filled boxes
: Toggle between solid-fill boxes and outline-only boxes.

Data labels
: Select the **Show labels** check box to display data labels. Data labels indicate the value of variables represented by sections of the boxes displayed in the plot.

To change the font of the data labels, click **Font** and then use the Font dialog box to select a font type and font characteristics such as size and color.

Histogram Tab



The **Histogram** tab in the Properties dialog box contains the following items:

Categorization

Use the Categorization section to set the following properties:

Number of bins

Select the number of bins that you want to use to divide the category data.

Min value

Specify the minimum data value to be included in the histogram binning. Values smaller than this are ignored.

Max value

Specify the maximum data value to be included in the histogram binning. Values larger than this are ignored.

Outline

Select or clear the **Outline** check box to control whether bins are outlined. If outlines are displayed, then you can change the thickness of the outline by selecting a different value from the **Thickness** list box.

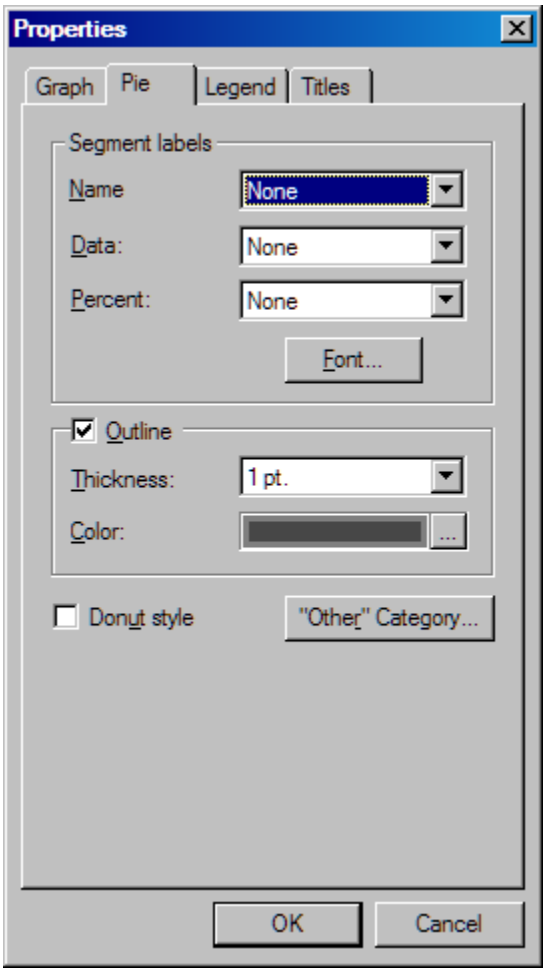
To change the color, click ... to open the Color dialog box. The Color dialog box contains colors to choose from and the option to define your own color for the outline. See "Using the Color Dialog Box" on page 62.

Data labels

Select the **Show labels** check box to display data labels. Data labels indicate the value of variables represented by the bins displayed in the plot.

To change the font of the data labels, click **Font** and then use the Font dialog box to select a font type and font characteristics such as size and color.

Pie Tab



The **Pie** tab in the Properties dialog box contains the following items for customizing pie charts:

Segment labels

Use the Segment Labels section to set the placement of the following segment labels:

Name

Specify the label option for the names of your pie segments (the category variable values).

Data

Specify the label option for the data values for the pie segments (the response variable values).

Percent

Specify the label option for the text that indicates the percentage of the full pie that is represented by each slice.

You can choose one of the following for each of the labels:

- **None** turns off the display of labels.
- **Inside** displays each label inside the pie slice that it represents.
- **Outside** displays each label next to the pie slice that it represents.
- **Arrow** displays the labels outside the pie and connects each label to its pie slice with a line.
- **Best** attempts to display each label inside the slice it represents. If the label does not fit, then it is displayed outside the slice.

To change the font for the labels, click **Font** and then use the Font dialog box to select a font type and font characteristics such as size and color. All labels are displayed in the same font.

Outline

Select or clear the **Outline** check box to control whether pie slices are outlined. If outlines are displayed, then you can change the thickness of the outline by selecting a different value from the **Thickness** list box. Use the Color selection button to open the Color dialog box. The Color dialog box enables you to select a color to be used for your pie outline.

To change the color of the outlines, click ... to open the Color dialog box. The Color dialog box contains colors to choose from and the option to define your own color. See "Using the Color Dialog Box" on page 62.

Donut style

Select the **Donut style** check box to change a pie chart to a donut chart.

"Other" category

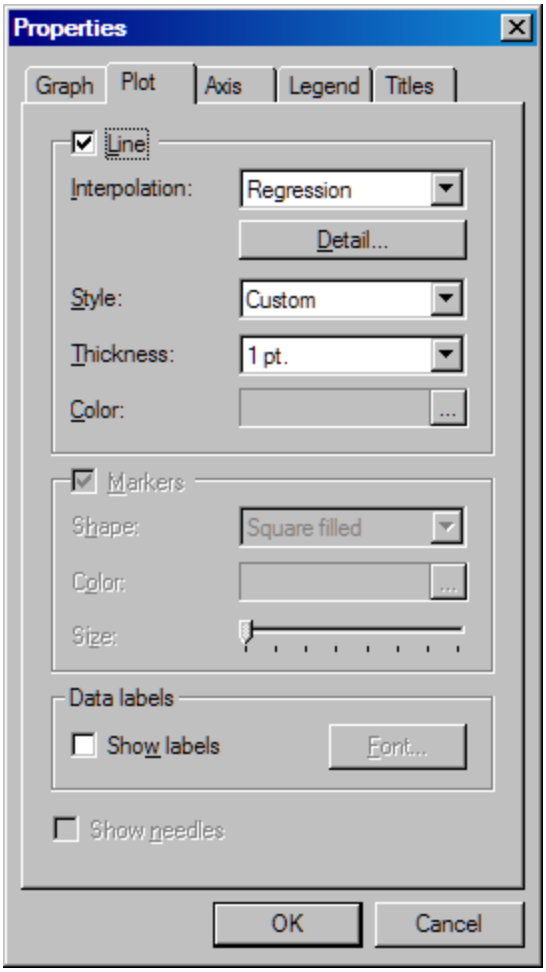
Select the **"Other" category** button to specify display parameters for categories that are four percent or less of the total pie chart. The "Other" Category dialog box opens. You can make the following changes in this dialog box:

Select or clear the **Show smaller values as "Other"** check box to control whether to show smaller values with the label of "Other."

Specify how you want the application to determine which values to include in "Other."

- Values can be determined based on a maximum number of categories that you select from the **Number** list box.
- Values can be determined based on the value's percentage of the total. You select the maximum percentage from the **Value** list box.

Plot Tab



The **Plot** tab in the Properties dialog box contains the following items for customizing scatter plots:

Line

You can display lines between markers in a scatter plot. Select the **Line** check box to enable the display of lines.

Interpolation

Use the Interpolation list box to change the interpolation method for the graph. Depending on the interpolation type that you select, the **Detail** button might be enabled. Click the **Detail** button to specify more details for the lines. The

following list summarizes the interpolation methods and describes any details that apply to the interpolation:

- **Join** connects data points with straight lines.
- **Step** plots data with a step function.

  For details, you can select the step placement and indicate whether to show join lines.
- **Spline** uses a spline routine to produce a smooth line or surface connecting the data points. Spline produces the smoothest line.

  For details, you can enable the sort and parametric options.
- **Smooth** fits a smooth line to the data using a spline routine. This option smooths noisy data. The points on the plot do not necessarily fall on the line.

  For details, you can specify a percentage for the plot smoothness. You can also enable the sort and parametric options.
- **Lagrange** specifies a Lagrange interpolation to smooth the plot line.

  For details, you can specify the degree of interpolation. You can also enable the sort and parametric options.
- **Map** specifies a polygon that has been defined and connected by the data points.
- **Regression** specifies that the plot is a regression analysis.

  For details, you can specify a regression type, confidence type, and a confidence level. You can also enable an intercept.

Style

Choose one of the available line styles from the **Style** list box.

Thickness

Choose a line width from the **Thickness** list box.

Color

To change the color of the lines, click ... to open the Color dialog box. The Color dialog box contains additional colors to select from and the option to define your own color. See "Using the Color Dialog Box" on page 62.

Markers

You can customize the markers in a scatter plot. Select or clear the **Markers** check box to enable or hide markers. If markers are enabled, then you can make the following changes:

Shape

Change the shape of the markers by selecting a new shape from the **Shape** list box.

Color

To change the color of the markers, click ... to open the Color dialog box. See "Using the Color Dialog Box" on page 62.

Size

Move the slider's thumb to the right or click on the slider right of the thumb to increase the marker size. Move the slider's thumb to the left or click on the slider left of the thumb to decrease the marker's size.

Note: You cannot change the marker shape and color if the chart is subgrouped.

Data labels

Select or clear the **Show labels** check box to enable or hide labels. Data labels indicate the value of variables represented by objects on the graph.

Click the **Font** button to specify font information for your data labels.

Show needles

Select **Show needles** to add vertical lines that connect each data point to the horizontal axis or plane. You can select **Show needles** only if the lines are not displayed.

5

Creating and Modifying Network Graphs

About Network Graphs

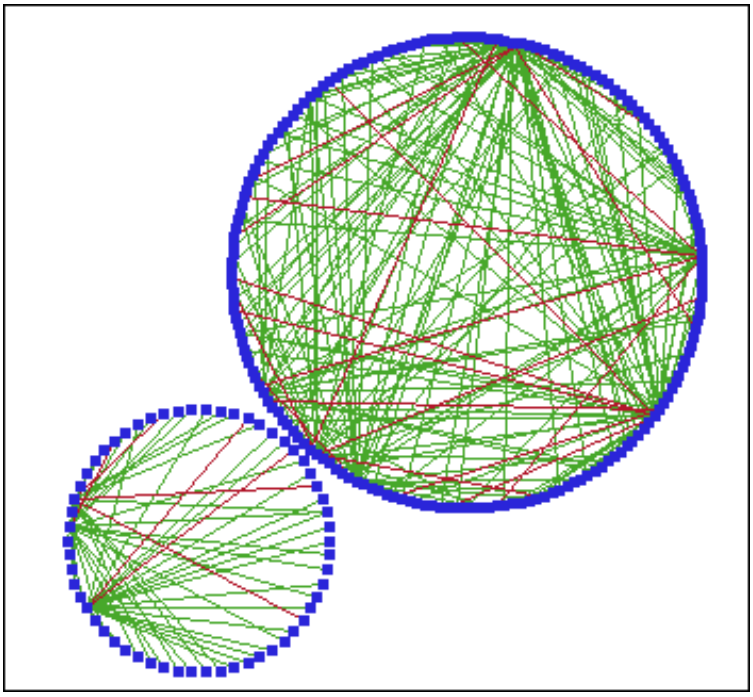
A network graph uses information from both the link and node data sets to generate a graphical depiction of the network.

The nodes and links in a network graph can be arranged in a variety of layout patterns. You can choose from the following patterns:

- Circular: the application sequentially assigns all the connected nodes in the network within a circle.

  Here is an example of a circular layout:

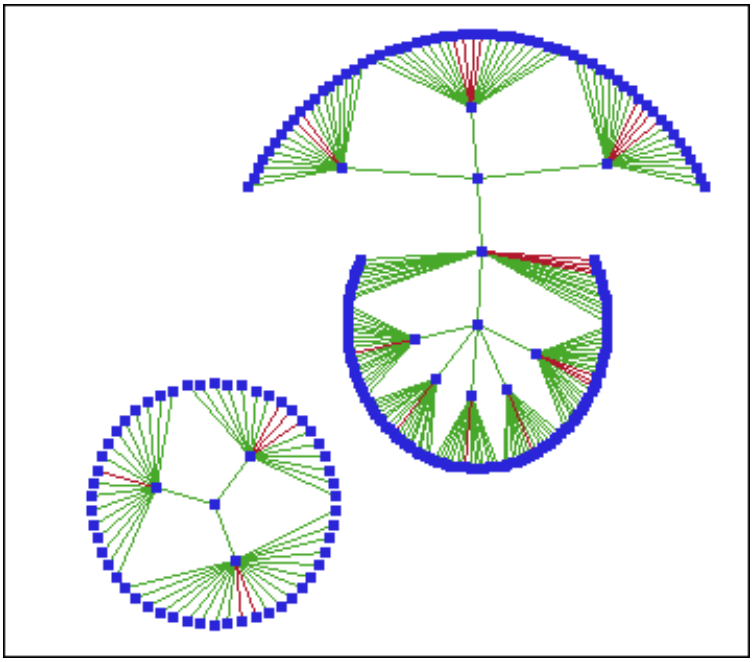
  Figure 5.1 *Circular Layout*



- Hierarchical: The application constructs a tree-like depiction of the network, and places nodes with more connecting links closer to the center.

  Here is an example of a hierarchical layout:

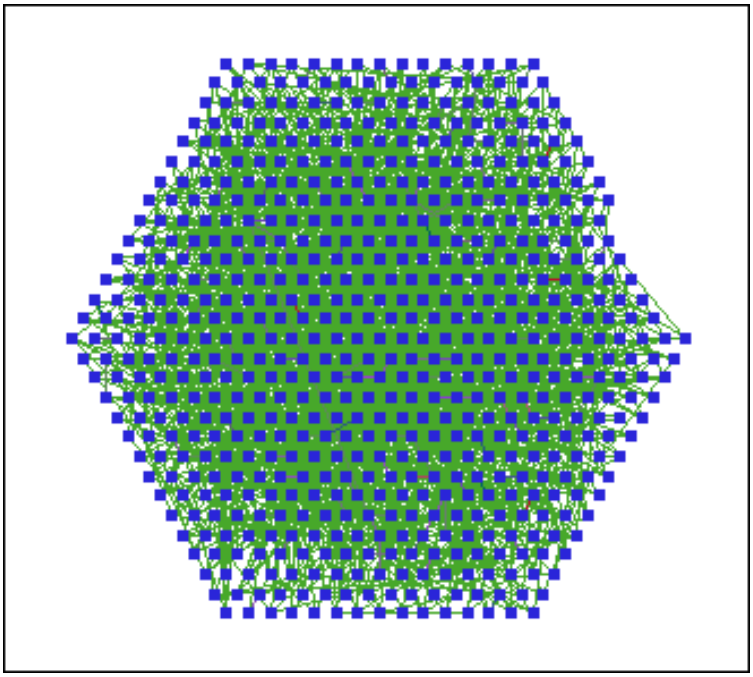
Figure 5.2 *Hierarchical Layout*



- Hexagonal: The application sequentially chooses nodes and attempts to distribute them evenly across one or more hexagons.

  Here is an example of a hexagonal layout:

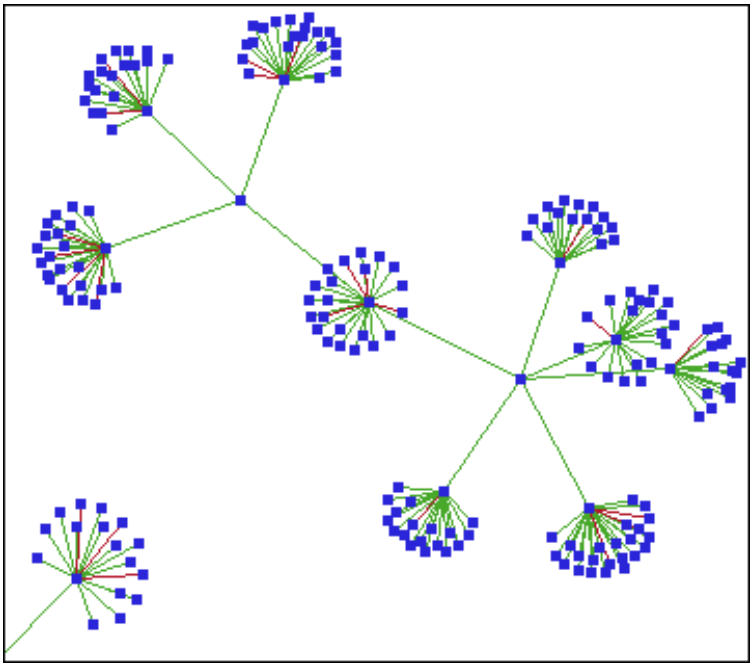
Figure 5.3 *Hexagonal Layout*



- Multi-level force-directed: The application combines graph partitioning and spring or force-directed layout heuristics to determine node positions.

  Here is an example of a multi-level force layout:

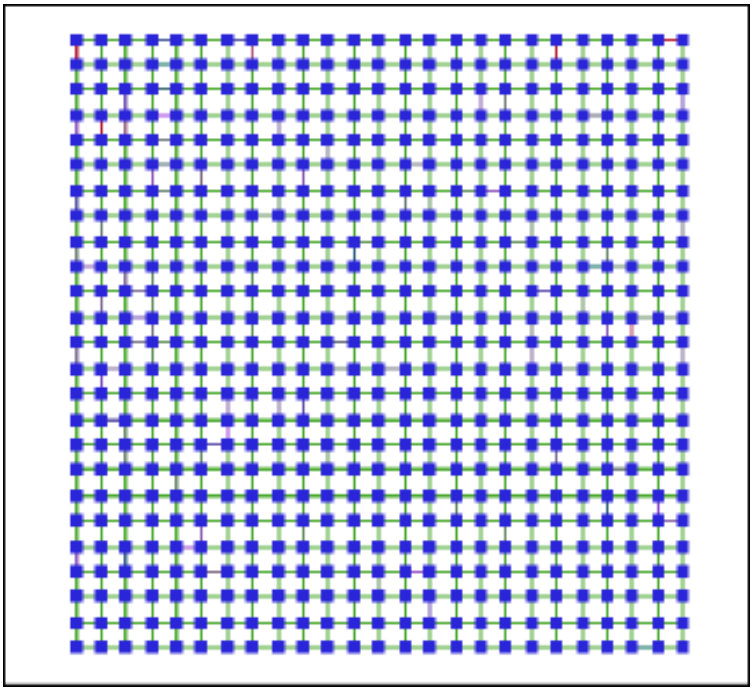
Figure 5.4 *Multi-Level Force-Directed Layout*



- Fixed position: The application arranges nodes according to (x; y) coordinates that you supply with a SAS data set.

  Here is an example of a fixed position layout:

Figure 5.5 *Fixed Position Layout*



For information about the position data set, see “Creating a Positions File for a Fixed Position Network Graph” on page 85.

You choose the initial layout pattern when you create the graph. Depending on the nature of the data, your choice of layout pattern can greatly affect the insight you can glean from the visualization of your network. Because you might not initially be aware of any underlying structure in your data, it can be helpful to try multiple layout patterns to determine which are the most useful for the network being investigated. After creating the graph, you can later change its layout.

Once you have created a network graph, you can analyze graph data in the following ways:

- Select individual observations or groups of observations in order to focus on a subset of the data. You can then analyze selected links, nodes, or both. For details about selecting data, see “Selecting the Data That Appears in Graphs” on page 95.
- Use statistical graphs to subset the data that is displayed. For example, you can create a scatter plot based on node data and then select some of the markers in the

scatter plot. When you select markers in the scatter plot, the network graph is updated to show only the nodes that correspond to the markers you selected.

- Use a local selection mode to graphically subset data at multiple levels. For more information, see “Understanding Selection Mode” on page 103.
- Regardless of the layout pattern you are using for a graph, you can fine tune the layout by adjusting its layout tuning algorithm. You can make this adjustment when you set the graph's properties.
- Use the tools available to manipulate the graphs. For example, you can zoom in on a portion of a graph or apply a magnifying lens to a network graph. You can also show the labels for particular nodes in a network graph.

You might want to change the appearance and behaviors associated with a graph by changing the graph's properties. The properties of a network graph determine features such as colors, line widths for links, the layout algorithm used for the graph, and the behavior of the lens tool.

SAS/GRAPH Network Visualization Workshop assigns colors to nodes and links based on user-designated variables in the respective node and link data sets. In addition, you can assign different shapes to the nodes based on a user-designated variable in the node data set. You can assign these variables using the Edit Data Attributes dialog box. For details, see “Specify Data Attributes” on page 17.

See Also

- “Create a Network Graph” on page 83
- “Change the Layout of a Network Graph” on page 84
- “Changing the Properties of a Network Graph” on page 86

Create a Network Graph

To create a network graph:

1 Make sure that both the node and link data sets have been loaded. For instructions, see “Load Data into SAS/GRAPH Network Visualization Workshop” on page 14.

2 Select **Graphs** ▶ **Visualize Network**, and then select a layout pattern from the cascading menu. For descriptions of the layout pattern, see “About Network Graphs” on page 77. Note the following:

- If you choose the **Fixed Position** layout style, then you are prompted to open a positions file and to specify which variables in the file correspond to the X and Y axes. For information about the positions file, see “Creating a Positions File for a Fixed Position Network Graph” on page 85.
- If you haven’t already specified the link and node attributes (the FROM, TO, and NODE ID variables), then the Edit Data Attributes dialog box is displayed. For help with this dialog box, see “Specify Data Attributes” on page 17.
- Networks with high depth can cause your system to run out of memory when you attempt to create a network graph. The depth of a network is the number of nodes that are linearly connected (A is connected to B, B is connected to C, C is connected to D, and so on). SAS/GRAPH Network Visualization Workshop displays a message when system memory has been exceeded. If you receive this type of message, you should close other applications before you try to create the network graph again. The actual maximum number of nodes that can cause this condition varies depending on system resources. Typically, the threshold is a depth of approximately 11,000 nodes.

You can fine-tune the layout pattern by adjusting the layout algorithm used for the graph. For details, see “Layout Tab” on page 89.

Change the Layout of a Network Graph

To change the layout pattern of an existing network graph:

1 Right-click inside the graph that you want to change.

2 From the pop-up menu, select **Layout Style**, and then select the layout that you want. For a description of the layout patterns, see "About Network Graphs" on page 77.

Creating a Positions File for a Fixed Position Network Graph

When you create a network graph with a fixed position layout pattern, you must specify the positions file for the graph. A positions file is a SAS data set that meets the following requirements:

- The data set contains the following variables:
  - a variable containing the X coordinates for the nodes
  - a variable containing the Y coordinates for the nodes

  The data set can contain additional variables, but must have at least these two.
- The data must be in the same order as the data in the node data set. The reason is that the positions data set doesn't require a "node" variable for its orientation; it relies on the position of the nodes in the node data set.

TIP One easy way to create a positions data set is simply to add two variables (for the X and Y coordinates) to the node data set.

Here is an example of a small data set:

	Name	xaxis	yaxis
1	Customer01	5	10
2	Customer02	5	15
3	Customer03	5	20
4	Customer04	5	25
5	Merchant01	10	10
6	Merchant02	10	15
7	Merchant03	10	20
8	Merchant04	10	25
9	Merchant05	10	30
10	Merchant06	10	35

When you create a network graph using the positions file, you are prompted to specify which variables correspond to the X and Y axes. SAS/GRAPH Network Visualization Workshop then graphs the nodes and their links based on your network data and the values provided for the X and Y axes.

Changing the Properties of a Network Graph

Change the Properties of a Network Graph

The properties of a network graph determine features such as colors, line widths for links, the layout algorithm used for the graph, and behavior of the lens tool.

Note: Any changes that you make to a graph's properties remain in effect only while the current project is open. If you close and then reopen the project, the properties revert to their default values.

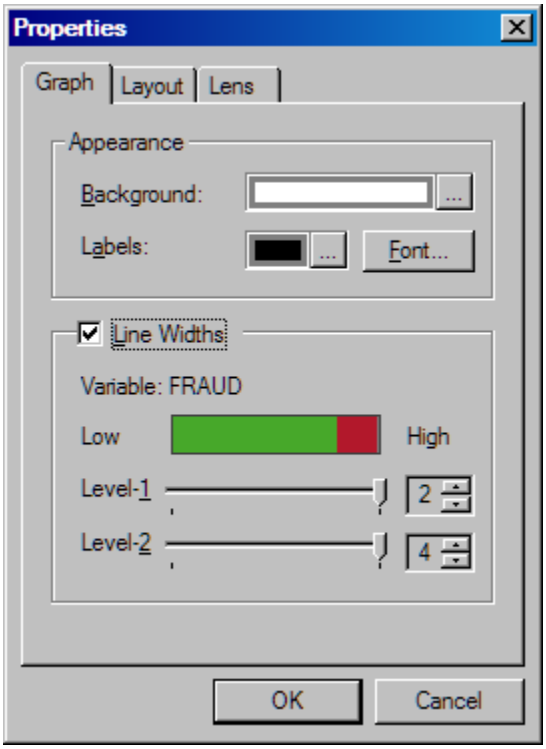
To change a network graph’s properties:

1. Right-click on the graph you want to change.
2. From the pop-up menu, choose **Graph Properties**. The Properties dialog box opens.
3. Change the settings for the graph.

The Properties dialog box contains several tabs that organize the properties you can change:

- "Graph Tab" on page 87
- "Layout Tab" on page 89
- "Lens Tab" on page 91

Graph Tab



The **Graph** tab in the Properties dialog box contains the following items:

Appearance

You can select the background color for the graph and the color that is used for graph labels. Click the color selection buttons ... for **Background** and **Labels** to open the Color dialog box for each. The Color dialog box contains colors to choose

from and the option to define your own color. For help using the Color dialog box, see “Using the Color Dialog Box” on page 62.

To change the font used for the labels, click **Font** and then use the Font dialog box to select a font type and font characteristics such as size and color.

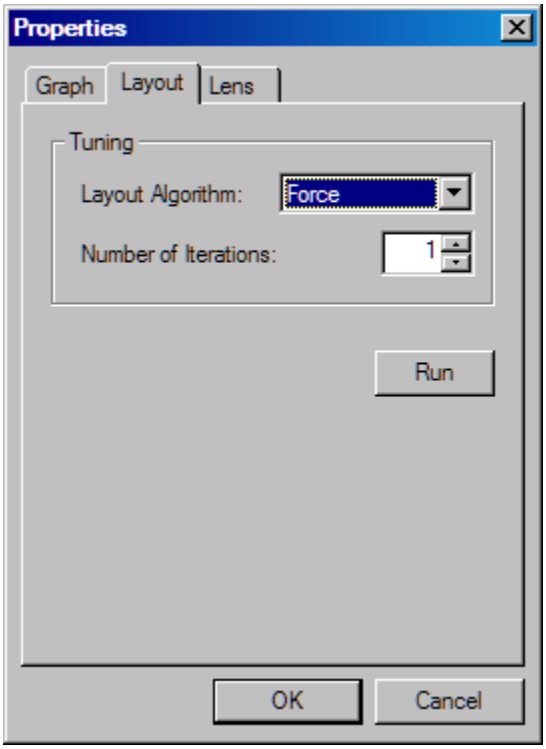
Line Widths

By default, all links are one pixel in width. Select the **Line Width** check box to set up to three different link widths.

This section uses a color spectrum to display the link colors, and lists the variable that is used to determine those colors. (You set this variable in the Edit Data Attributes dialog box. See “Specify Data Attributes” on page 17.)

Use the sliders to set the boundaries between the different line width levels. Use the arrows to the right of the list boxes to increase or decrease the width of the lines. The **Level-1** and **Level-2** sliders enable you to set two line-width levels for particular link colors. As you drag a slider across the color spectrum, the respective links change the width of their lines.

Layout Tab



Regardless of the layout pattern you are using for a graph, you can fine tune the layout by adjusting its layout tuning algorithm. A layout tuning algorithm adjusts the current node layout in a prescribed manner in an attempt to create a more visually useful layout. When you first create a network graph, you might not know which layout algorithm is best for your graph, so it's often helpful to explore the available algorithms.

From the **Layout Algorithm** list box, select one of the following algorithms:

- **Swap** chooses two nodes at random and determines whether it would be beneficial to swap or exchange the location of those two nodes. You specify the number of iterations that the algorithm should run by selecting the number from the **Number of Iterations** list box.

- **Force** is more complex than Swap, and applies a multi-level force algorithm to attempt to improve the layout. The algorithm is computationally intensive, so you should be conservative when entering the number of iterations for this option.

Click **Run** to apply the algorithm to your graph. You can run the algorithm multiple times to see different layouts. The Force algorithm typically changes the layout each time you click **Run**.

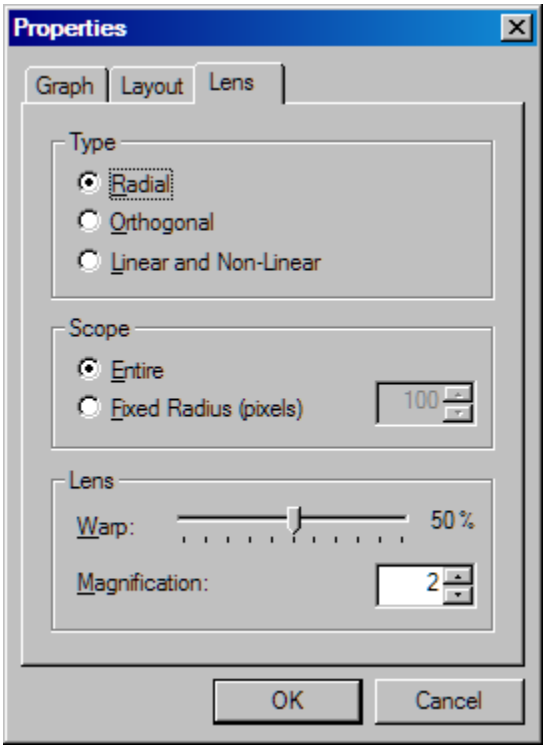
The Swap algorithm might or might not change the layout when you click **Run**, depending on the following factors:

- the number of iterations you specified
- the number of nodes and links in your graph
- the fundamental structure of the graph

Note: After you click **Run**, any changes made to the layout cannot be undone by clicking the **Cancel** button. However, the **Cancel** button cancels all other changes made in the properties dialog box.

For information about the layout patterns available for network graphs, see “About Network Graphs” on page 77. To change a graph's layout pattern, see “Change the Layout of a Network Graph” on page 84.

Lens Tab



The **Lens** tab enables you to set properties for the lens tool, which transforms (magnifies and warps) the region of the graph near the cursor. The **Lens** tab contains the following items:

Type

Select the type of transformation that the lens tool uses:

- **Radial** applies the lens radially around the cursor.
- **Orthogonal** applies the lens along an X and a Y axis.
- **Linear and Non-Linear** combines two types of transformations into one lens. The points closest to the cursor are zoomed using a linear transformation (all are moved the same amount). The points farther away from the cursor are scaled using a nonlinear technique. The effect is that the farther away the point is from the cursor, the less zoom it gets. Changing the **Fixed Radius** value controls

where the demarcation line is between the linear and the nonlinear transformation.

Scope

Select a button to set the range of the transformation from the center of focus. This setting controls the size of the lens. Select **Entire** if you want the warped portion to cover the entire graph. Select **Fixed Radius** to specify in pixels the portion of the graph that is to be warped. The default value is 100 pixels. If you select a **Linear or Non-Linear** lens type, then **Fixed Radius** is the only option available.

Lens

The **Warp** slider enables you to set the relative weight given to the transformed node location with respect to the original location. Move the slider to the left for less warping, or to the right for more warping.

Use the **Magnification** list box to set the relative power of the transformation. You can increase magnification of the warped portion of the graph by selecting a higher value. The default value is 2.

Change the Node Markers in a Network or Statistical Graph

SAS/GRAPH Network Visualization Workshop uses markers to indicate nodes in a network graph. You can change the appearance of the markers in a graph by setting their color and shape.

Note: You can also change the markers that appear in some statistical graphs, such as scatter plots and box plots.

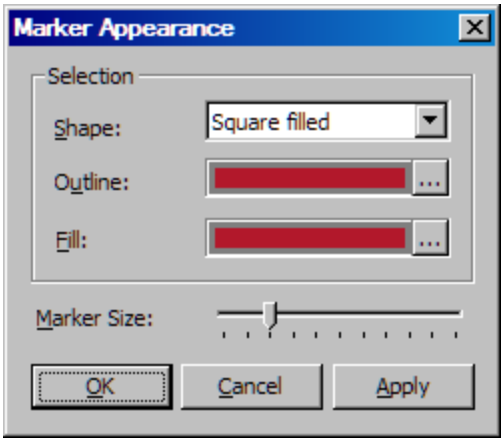
The changes described here override settings that have been made in the Edit Data Attributes dialog box. For information about the Edit Data Attributes dialog box, see “Specify Data Attributes” on page 17. Any changes that you make here remain in effect only while the current project is open. If you close and then reopen the project, the marker reverts to its default properties.

To change the marker:

1 Right-click on the graph whose marker you want to change.

2 If you want to change the marker color and shape for particular nodes only, then select those nodes. Do not select any nodes if you want to change the size of the marker.

3 From the pop-up menu, choose **Marker Properties**. The Marker Appearance dialog box opens.



4 Make the desired changes. The dialog box contains the following items:

Selection

From the **Shape** list box, select the shape that you want for the markers. There are a number of shapes to choose from, and each can be either solid-filled or outline-only.

Click the color selection buttons ... for **Outline** and **Fill** to open the Color dialog boxes for each, where you can change the marker's outline color and the fill color respectively. The **Fill** area is available only if you selected a filled shape for the marker.

The Color dialog box contains colors to choose from and the option to define your own color. For details, see "Using the Color Dialog Box" on page 62.

Marker Size

Use the **Marker Size** slider to decrease (move to the left) or increase (move to the right) the size of the marker.

Note: Color and shape changes apply to any nodes that have been selected. If no node is selected, then the color and shape changes apply to all the nodes. Changes to the size always apply to all the nodes. The size slider is dimmed if any nodes have been selected.

5 Click **Apply** to apply your changes without closing the Marker Appearance dialog box.

6 When you are finished making changes, click **OK**.

6

Working with Graphs

Selecting the Data That Appears in Graphs

About Selecting Data

You select data in order to filter the view and focus on the selected data. You can select data in a data table, in a network graph, or in a statistical graph.

Selecting data in one view affects all views (unless you are using local selection mode). For example, if you select an observation in a data table, then that observation is automatically selected in all graphs associated with the project. Similarly, you can use statistical graphs to subset the data in a network graph. For example, you might select all the nodes for a particular category in a bar chart. Then, you can view only those nodes in your network graph, along with their links.

Note: When you select a portion of a statistical graph that is based on nodes, the selected nodes display in your network graphs along with all their links. Likewise, when you select a portion of a graph that is based on links, the selected links display in your network graphs along with all their nodes.

To select data, you can use the selection tool or the highlight tool.

See Also

- "Understanding Selection Mode" on page 103
- "Use the Selection Tool to Select Data" on page 96
- "Use the Highlight Tool to Select Data " on page 97

Use the Selection Tool to Select Data

The selection tool is enabled by default when you open a project. If you use a different tool, then you might need to re-enable the selection tool afterward.

To enable the selection tool, select **Tools** ▶ **Selection**.

To select data:

- In a statistical graph, click the portion of the graph that you want to select. For example, click on a histogram bin to select all observations in the same bin.

  Hold down the CTRL key to make multiple selections.

- In a statistical or a network graph, click and drag the cursor across the portion of the graph that you want to select. The cursor forms a rectangle as it sweeps across the nodes. Dragging a rectangle selects all observations within that rectangle. Once a rectangle is created, you can move the rectangle by placing the cursor inside the

rectangle and dragging it to a new location. As the rectangle passes over observations, those observations are selected.

- In a data table, select cells, variables, or observations. See "Selecting Data in a Table" on page 26.

Use the Highlight Tool to Select Data

The highlight tool is similar to the selection tool, except that the deselected nodes and links appear as gray in a network graph. (When you use the selection tool, the deselected nodes and links don't appear at all.)

To use the highlight tool:

1. Select **Tools** ▶ **Highlight**.
2. Select the data of interest. For instructions, see "Use the Selection Tool to Select Data" on page 96.

Label the Nodes in a Network Graph

The label tool enables you to attach labels to nodes in a network graph. You can label individual nodes or groups of nodes.

To label the nodes:

1. Right-click in the network graph whose nodes you want to label.
2. Select **Tools** ▶ **Label Tool**.
3. Select the nodes to be labeled using any of the following methods:
   - Click on a node. The label appears next to the node.
   - Press and hold the CTRL key to click multiple nodes.

- Click and drag the cursor across a group of nodes. The cursor forms a rectangle as it sweeps across the nodes. This action labels all the nodes located within the rectangle.
- Press and hold the CTRL key to drag and form multiple rectangles.
- To clear all labels, click on the white space outside the graph.

4 When you are finished labeling nodes, change the cursor back to the selection state. Select **Tools** ▶ **Selection**.

You can customize the labels in the following ways:

- In the Edit Data Attributes dialog box, you can change the variable that is used for the label text. For details, see "Specify Data Attributes" on page 17.

  If you have not selected a variable to be used for labeling, then the default label is the observation number in the node data set.
- In the **Graph** tab of the graph's Properties dialog box, you can change the label's color. For details, see "Changing the Properties of a Network Graph" on page 86.

Zoom In and Out of a Graph

You can zoom in to get a close-up view of a graph or zoom out to see more of the graph at a reduced size. The zoom tool works on all graphs except pie charts. For statistical graphs, the tool zooms on the axes.

To zoom in and out:

1 Select **Tools** ▶ **Interactive Zoom**.

2 To zoom in, do one of the following:

- For network graphs, click the graph that you want to zoom.
- For statistical graphs, select a portion of the graph that you want to zoom. Scroll bars appear for the graph. You can use the scroll bars to navigate the axes.

3 To zoom out, do one of the following:

- For network graphs, press the Shift key and click the graph.
- For statistical graphs, click the graph.

4 When you are finished zooming graphs, change the cursor back to the selection state. Select **Tools** ▶ **Selection**.

Move a Graph in Its Window

Use the pan tool to move a graph up, down, or to the side within its graph window. The pan tool is often used in conjunction with the zoom tool. You can zoom in on a graph and then move the graph to reposition the portion that's of interest.

To move a graph:

1 Select **Tools** ▶ **Pan**.

2 Click the graph and drag the cursor in the direction that you want the graph to move.

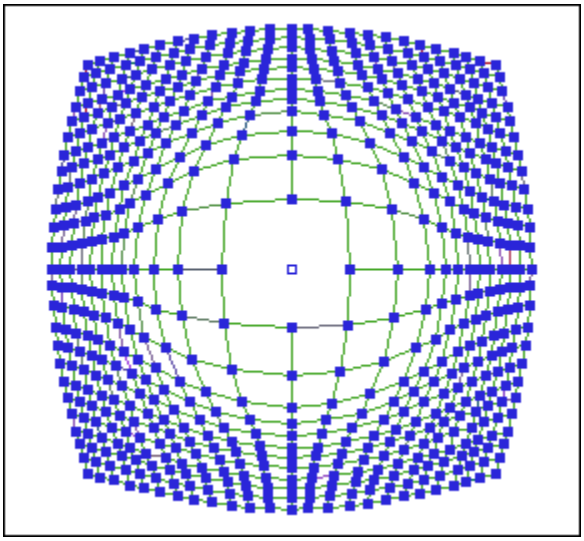
3 When you are finished moving graphs, change the cursor back to the selection state. Select **Tools** ▶ **Selection**.

Apply a Lens to a Network Graph

The lens tool magnifies and warps the graph in a way that is similar to a magnifying lens. The tool affects the region of the graph near the cursor. As you move the cursor, the lens moves. Network graphs tend to be very dense, and the use of a lens enables you to focus on the detail in a particular region of the graph.

The following figure shows the use of a lens applied to a network graph.

Figure 6.1 Lens Applied to a Network Graph



There are multiple types of lenses available for a network graph, and each has various options. The default lens is the radial lens. You can change the lens type and customize some of the properties of the lens tool using the graph's Properties dialog box. See "Lens Tab" on page 91.

To use the lens tool:

1. Select **Tools** ▶ **Lens**.
2. Click the network graph and drag the lens across the graph.
3. When you are finished with the lens, change the cursor back to the selection state. Select **Tools** ▶ **Selection**.

Save a Graph as a Graphic File

You can save a statistical or network graph as a graphic file in one of the following formats: JPG, PNG, GIF, and BMP (24-bit).

To save a graph as a graphic file:

1 Right-click the graph that you want to save.

2 From the pop-up menu, select **Save As**. The Save As dialog box opens.

3 In the Save As dialog box, select the location where you want the file to be stored. You can also change the filename, the file extension, and the file type. By default, the file is saved with the name **NVGraph.jpg**.

Print a Graph

You can print a statistical or network graph.

To print a graph:

1 Right-click the graph that you want to print.

2 From the pop-up menu, select **Print**. The Print dialog box opens.

3 In the Print dialog box, make any changes that you want to the printer selection or print options, and click **OK**. The graph prints to the selected printer.

Copy and Paste a Graph

You can copy and paste a graph into another application such as Microsoft Excel or Microsoft Word.

To copy a graph:

1 Right-click the graph that you want to copy.

2 From the pop-up menu, select **Copy**. The graph is copied to your system clipboard.

You can paste the graph into the target application using that application's paste command.

7

Using Selection Mode to Subset Data

Understanding Selection Mode

By default, when you select an observation in a graph or data table, that observation is automatically selected in all graphs and data tables associated with the project. You can change the default behavior so that a selection is limited to the graph or data table in which the selection is made.

SAS/GRAPH Network Visualization Workshop supports two selection modes:

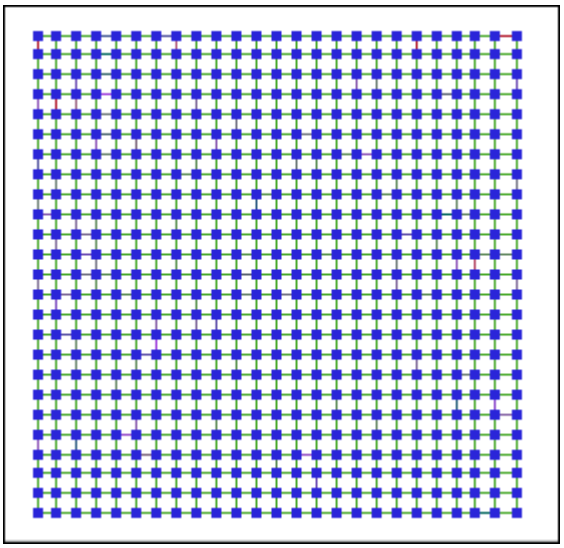
Global selection mode	The default mode. All data views (graphs and data tables) share a common observation selection state. Global selection mode enables you to graphically subset data at a single level. Suppose you have three graphs: A, B, and C. Selecting a subset of the observations in graph A causes graphs B and C to treat that same subset as selected.

Local selection mode	Each data view has a private observation selection state. When you select an observation in a graph or data table, the observation is marked as selected only for that view. Local selection mode enables you to graphically subset data at multiple levels. Suppose you have three graphs: A, B, and C. You can select some observations in graph A, and then select different observations in graph B. Finally, you can configure graph C to display either the union or the intersection of the selections made in graphs A and B.

While in local selection mode, a graph operates in one of two roles:

- Selector role: A view that has a selector role enables you to manually select observations. When you select an observation, the observation is marked as selected only for that view.
- Observer role: A view that has an observer role does not allow you to select observations. An observer view treats an observation as selected based on the observation's selection state in the selector views and on the observer view's scheme. There are two observer view schemes:
  - The union scheme treats an observation as selected if the observation is selected in any of the selector views.
  - The intersection scheme treats an observation as selected if the observation is selected in all of the selector views.

The following display shows an example of a fixed position network visualization using the Computer Grid sample data.

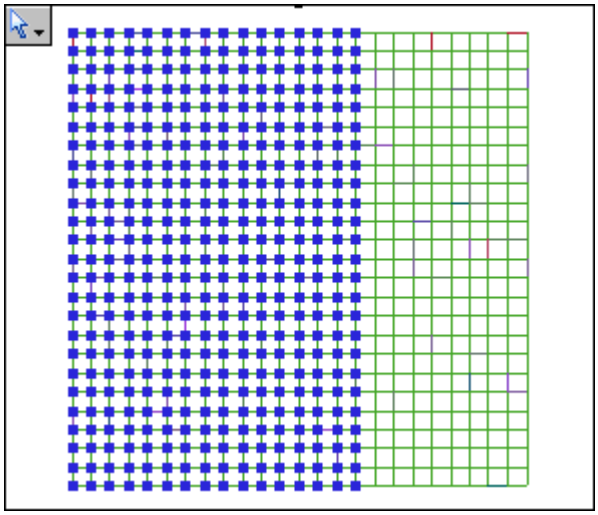


The following display shows two additional network graphs of the same data. Here are the characteristics of these network graphs:

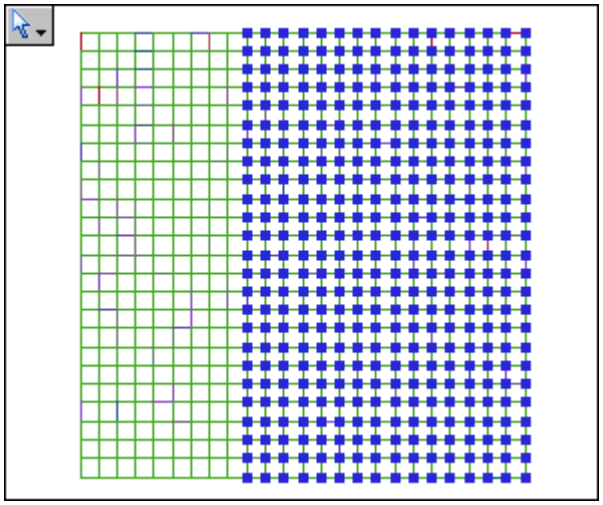
- Local selection mode has been applied.
- Both of these graphs have been changed to a selector role, as indicated by the icon in the upper left corner.
- Different observations are selected for each network graph, though there is some overlap (intersection).

Note: All tables and any statistical graphs that are open have been changed to observer role. They do not participate in the selection of data.

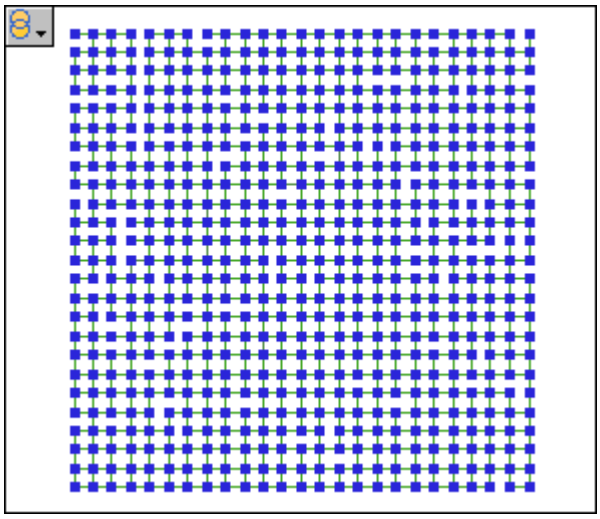
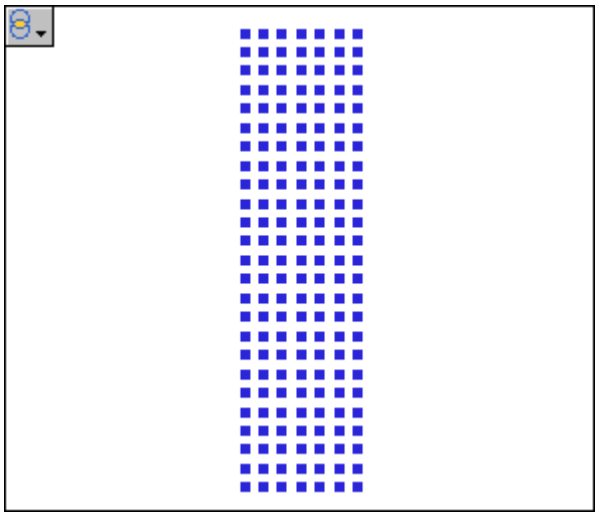
Some data selected in a network graph



Different data selected in another network graph



The following display shows the union and the intersection of the selected data.

Union	
Intersection	

See Also

“Configure Local Selection Mode” on page 107

Configure Local Selection Mode

To configure local selection mode:

1 Select **Data** ▶ **Selection Mode** ▶ **Local**. All data views switch to local selection mode. An icon is displayed in the upper left corner of each graph and table to indicate its respective role.

Graphs and data tables can have one of the following icons:

Table 7.1 *Local Selection Icons Used in Graphs and in Tables*

Graph or Table	Description
	Indicates that the view has a selector role. Initially, all statistical graphs and data tables are configured as selector roles.
	Indicates that the view has an observer role with a union scheme. Initially, all network graphs are configured as observer views with a union scheme.
	Indicates that the view has an observer role with an intersection scheme.

For descriptions of these roles, see "Understanding Selection Mode" on page 103.

2 To change the role for a view, click its icon and choose a different role from the drop-down menu. You can change a selector view to an observer view, and vice versa. You can also change the scheme of an observer view from union to intersection, and vice versa.

Note: If any statistical graphs or data tables do not participate in the selection of data, be sure to change them to observer roles.

If you change an observer view to a selector view, the view initially displays the union of the selections made in the other selector views. To clear this selection and make a new one, click in the graph area to deselect all observations in the view, and then select the desired observations.

3 When you are finished using local selection mode, restore global selection mode. See "Restore Global Selection Mode" on page 109.

Restore Global Selection Mode

When you are finished using local selection mode, to restore global selection mode, select **Data** ▶ **Selection Mode** ▶ **Global**. All data views switch to global selection mode.

When you return to global selection mode, all views initially display the union of the selections that were made in local selection mode.

8

Working with Projects

About Projects

The most efficient way to manage related data tables and graphs is to work with projects. A project is a collection of one or more data tables, graphs, and a style. The project saves all of your settings so that SAS/GRAPH Network Visualization Workshop can retrieve the settings for use in later sessions. All of this information is stored in a project file that has an extension of NVW.

You can perform the following tasks with projects:

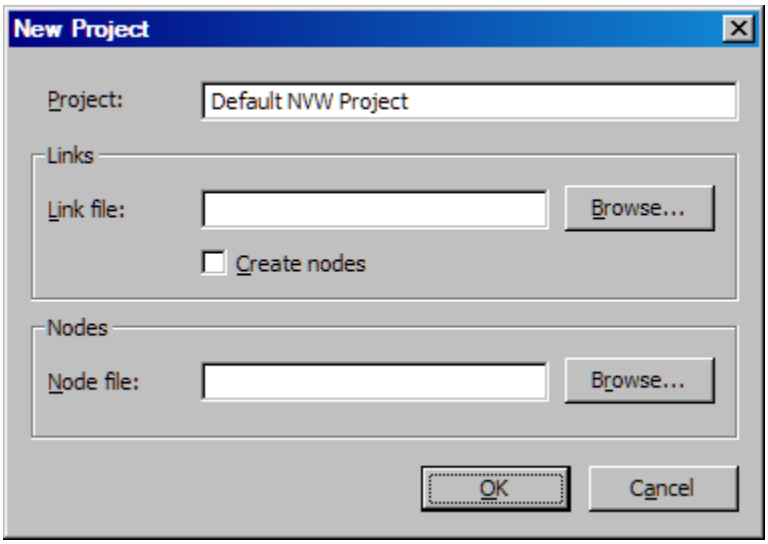
- “Create a New Project” on page 112
- “Open an Existing Project” on page 113
- “Save a Project” on page 114

■ “Apply a Style to a Project” on page 114

Create a New Project

To create a new project:

1 Select **File** ▶ **New**. The New Project dialog box opens.



Note: You can also open this dialog box by selecting **New Network Visualization Project** on the Welcome screen that appears when you first open SAS/GRAPH Network Visualization Workshop.

2 In the New Project dialog box, specify the following information:

Project
: Enter a name for the project. Use any combination of spaces and alphanumeric characters except the following: < > : " / \ | ? *

Links
: If the project uses a link data set, click **Browse** to locate the link data set that you want to use.

Nodes

If the project uses a node data set, click **Browse** to locate the node data set that you want to use.

If you want SAS/GRAPH Network Visualization Workshop to generate the node data set, select the **Create nodes** check box.

Note: For your project, you can specify either a link data set, a node data set, or both.

3 Click **OK**.

4 Create graphs for your project as appropriate.

- "Create a Statistical Graph" on page 47
- "Create a Network Graph" on page 83

5 Save the project. For instructions, see "Save a Project" on page 114.

If you already have one or more data tables opened, then you can create a project simply by selecting **File** ▶ **Save Project As**, and giving the project a filename.

Open an Existing Project

You can open a project in one of three ways:

- select a project from the Welcome dialog box
- select one of the last eight opened projects from the **File** menu
- use the Open dialog box

To use the Open dialog box:

1 Select **File** ▶ **Open**. The Open dialog box opens.

2 Locate the project that you want to open, and then click **Open**. Project filenames have an extension of NVW.

Save a Project

You can save a project as follows:

- If you have modified an existing project and want to save your changes, select **File** ▶ **Save Project**.
- If you have created a new project that has not yet been saved, select **File** ▶ **Save Project As** and then choose a location for the project.

  SAS/GRAPH Network Visualization Workshop saves the project using the name that you supplied when you created the project, along with the extension NVW. If you change the filename, SAS/GRAPH Network Visualization Workshop changes the project name accordingly. For the filename, use any combination of spaces and alphanumeric characters except the following: < > *: " / \ | ? **

Apply a Style to a Project

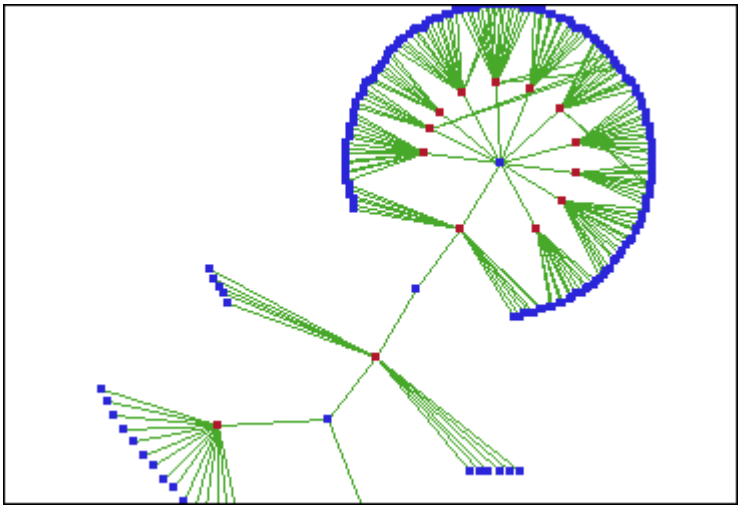
SAS/GRAPH Network Visualization Workshop contains pre-defined styles that specify different colors, fonts, and graphics for the user interface. You can apply a style to a project or to data tables and graphs that you have open. Most often, you work with projects.

To apply a style:

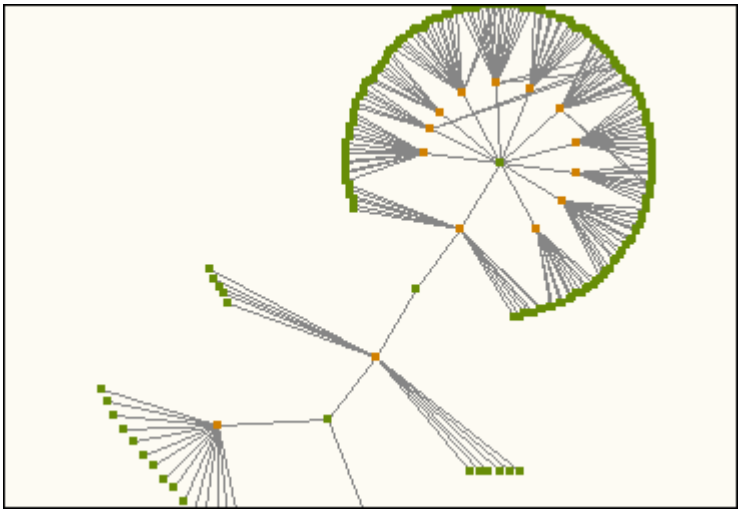
1. Select **View** ▶ **Style**.
2. From the cascading menu, select the style that you want.

The following table shows a portion of a hierarchical network graph with the different styles applied:

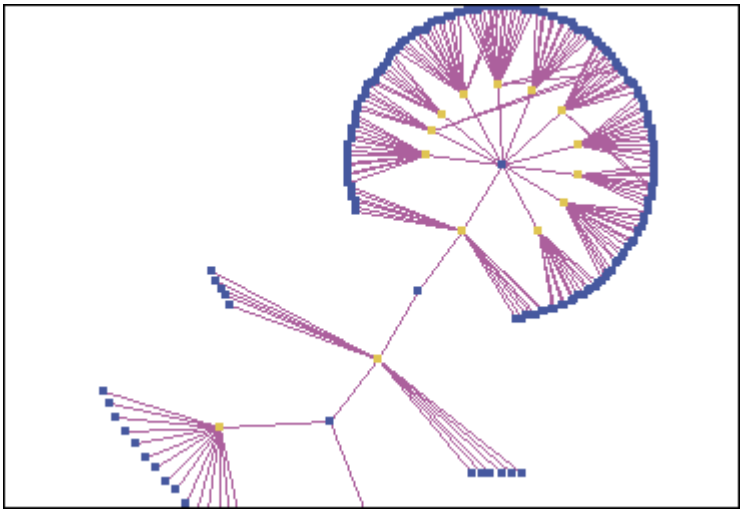
Default



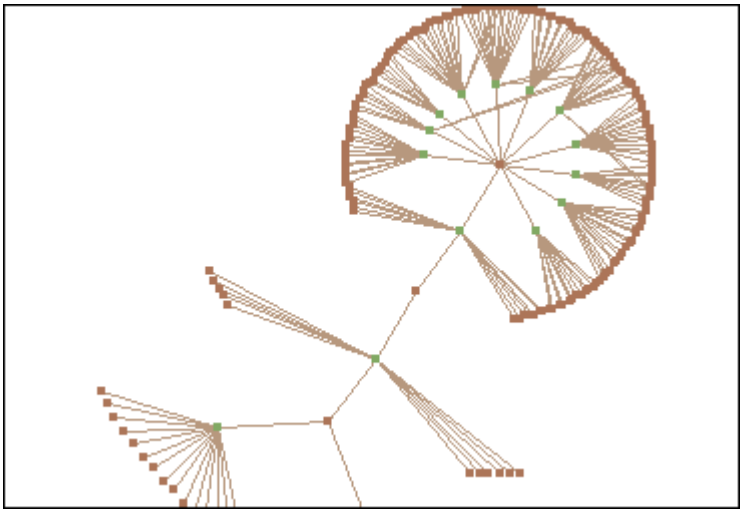
Analysis

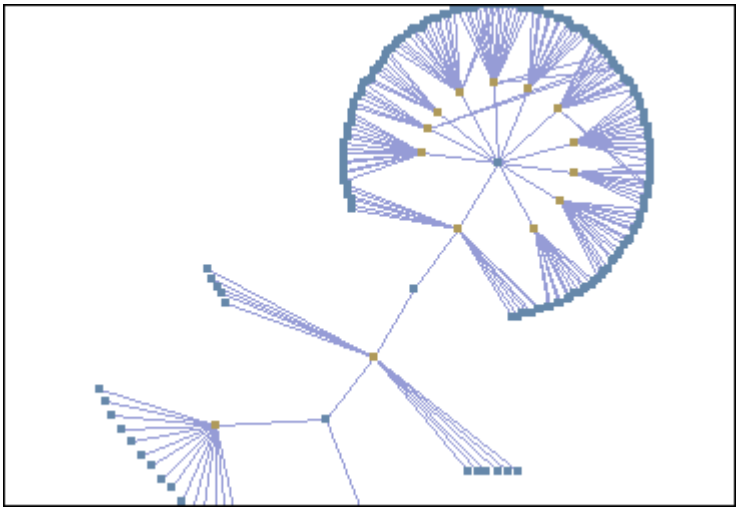
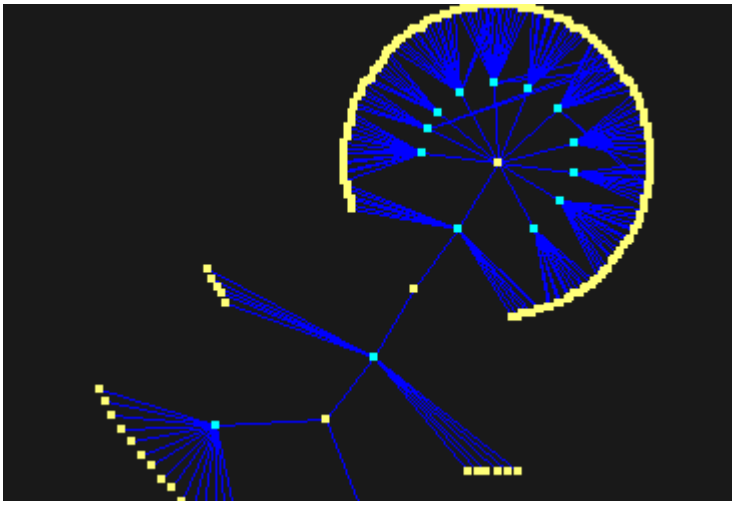


Festival



Meadow



Seaside	
NV Classic	

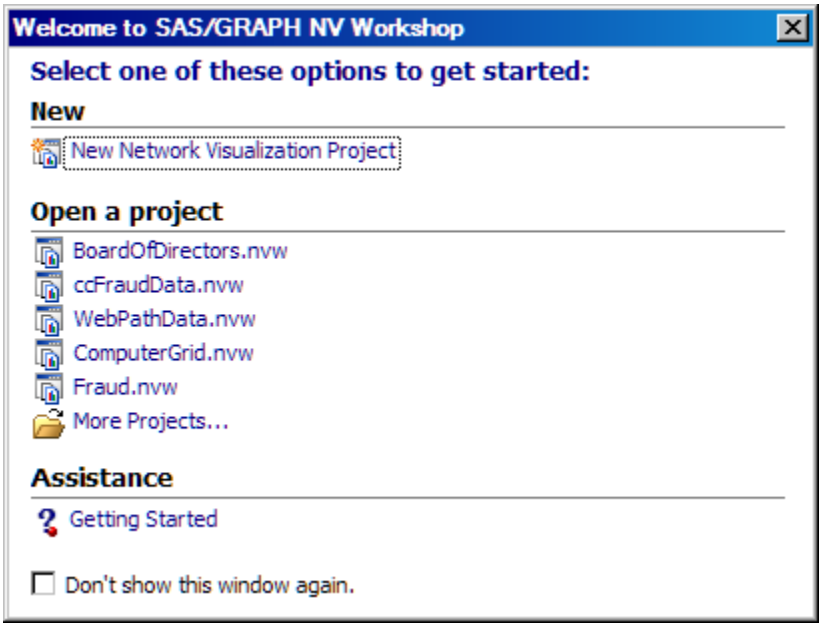
Show the Welcome Dialog Box

When you start SAS/GRAPH Network Visualization Workshop, the Welcome dialog box opens. From this dialog box, you can do the following:

- create a new project

- open an existing project
- display a Getting Started topic in the documentation (this topic is also available from the Help menu)
- select the **Don't show this window again** check box

If you select the **Don't show this window again** check box, the Welcome dialog box no longer appears when you open SAS/GRAPH Network Visualization Workshop. If you later decide that you want to display the dialog box again, then select **View** ▶ **Show Welcome Dialog**. This menu option toggles the Welcome dialog box on and off.



9

Sample Data and Example Use Cases

Using the Sample Data

About the Sample Data

Several sample projects and data sets are included with SAS/GRAPH Network Visualization Workshop.

The samples are located in the **Samples** directory of your installation. This directory contains the following subdirectories:

Projects	Contains the sample project files. When you open a project file, SAS/GRAPH Network Visualization Workshop automatically opens the data sets and all graphs that have been associated with the project.
Data	Contains the sample data sets. You can load individual data sets without opening a project.

You can explore the following samples:

- "Credit Card Fraud" on page 120
- "Credit Card Fraud (Simplified)" on page 121
- "Computer Grid" on page 122
- "Board of Directors" on page 125
- "Web Path Data" on page 127

Credit Card Fraud

This sample enables you to investigate patterns of credit card fraud. The data contains transactions, both valid and fraudulent, related to customers with at least one fraudulent credit card transaction.

This sample uses the following files:

Project File:	ccFraudData.nvw
Link Data Set:	cclinks.sas7bdat
Node Data Set:	ccnodes.sas7bdat

For descriptions of the variables and an example use case, see "Example Use Case: Credit Card Fraud" on page 128.

Credit Card Fraud (Simplified)

This sample is similar to the previous sample but contains less data and is easier to examine. For this sample, the node data is generated from the link data set.

This sample uses the following files:

Project File:	Fraud.nvw
Link Data Set:	fraud.sas7bdat
Node Data Set:	auto-generated

The following table summarizes the variables in the link data set:

Table 9.1 *Link Variables for the Credit Card Fraud Sample*

Variable	Description
Customer ID	Serves as the From variable that originates the link.
Transaction ID	A value that identifies every transaction.
Merchant ID	Serves as the To variable that terminates the link.
Fraud	Indicates whether the transaction is fraudulent (1) or not fraudulent (0). This variable also determines link colors.

SAS/GRAPH Network Visualization Workshop generates the node data from the link data. The following table summarizes the variables in the node data set.

Table 9.2 Node Variables for the Credit Card Fraud Sample

Variable	Description
Node	Serves as the Node ID variable that identifies all customers and merchants.
Label	Determines the text that appears when you apply a label to nodes in a network graph.
In Arcs; Out Arcs; Total Arcs	Created by SAS/GRAPH Network Visualization Workshop to keep track of the links that are associated with each node.

Here are suggestions for exploring this data:

1. Open the project. The project loads the two data sets and opens a hierarchical network graph.
2. Use the label tool to identify customers and merchants in the graph. Observe that the fraudulent transactions correspond to different customers that have transactions with a common merchant.

   Although this approach is not proof of fraud, it identifies specific merchants that had access to customer credit card numbers that have fraudulent transactions associated with them. These merchants warrant further investigation.

Computer Grid

This sample represents a mock physical layout of computer circuitry that is undergoing operational testing. This sample is useful for exploring fixed position and hexagonal network graphs. The sample can also be used to demonstrate the power of local selection.

This sample uses the following files:

Project File:	ComputerGrid.nvw
Link Data Set:	gridlink4.sas7bdat
Node Data Set:	gridnodes.sas7bdat
Position File:	gridnodes.sas7bdat

The following table summarizes the variables in the link data set:

Table 9.3 *Link Variables for the Computer Grid Sample*

Variable	Description
From	Serves as the From variable that originates the link.
To	Serves as the To variable that terminates the link.
Failures	Provides the number of times that the connection failed.
Tests	Provides the number of times that the connection was tested.
Weight3	Indicates an arbitrary variable that is available for the purpose of exploring the data.
Pcnt_fail	Indicates failure as a percentage of the number of failures divided by the number of tests.

The following table summarizes the variables in the node data set:

Table 9.4 *Node Variables for the Computer Grid Sample*

Variable	Description
Node	Serves as the Node ID variable that identifies all computers in the data. This variable also determines the text that appears when you apply a label to nodes in a network graph.

Variable	Description
Weight;Weight2;Weight3	Indicates arbitrary variables that are available for the purpose of exploring data that has multiple variables.
X	Provides the X coordinate for a fixed position network graph.
Y	Provides the Y coordinate for a fixed position network graph.

Here are suggestions for exploring this data:

1. Open the project. The project loads the two data sets and the graphs that are associated with the project, including a histogram and a fixed position network graph.
2. In the histogram of failure percentages, select the bars with an X coordinate greater than six. The network graph changes to display only those links that correspond to a significant failure rate (more than six percent).
3. Explore multiple points of potential failure by using local selection mode:
   a. Change to local selection mode. See “Configure Local Selection Mode” on page 107.
   b. Change both tables to the observer-union role. The network graph already has the observer-union role.
   c. In the histogram of tests, select the first two bars.
   d. In the histogram of failure percentages, select the bars with an X coordinate greater than six percent.

   The network graph now shows which links warrant closer scrutiny either because they have high failure percentages or because they have not received sufficient testing.

Board of Directors

This sample enables you to investigate relationships among boards of directors for Fortune 100 companies. In particular, you can detect directors who belong to several boards.

This sample uses the following files:

Project File:	BoardOfDirectors.nvw
Link Data Set:	bdlinks.sas7bdat
Node Data Set:	bdnodes.sas7bdat

The following table summarizes the variables in the link data set:

Table 9.5 *Link Variables for the Board of Directors Sample*

Variable	Description
Individual	Serves as the From variable that originates the link.
Corporation	Serves as the To variable that terminates the link.
Board Chairman	Indicates whether the individual is (1) or is not (0) the board chairman of the corresponding corporation.
CEO	Indicates whether the individual is (1) or is not (0) the chief executive officer of the corresponding corporation.
Number of Boards	Provides the number of boards of which the individual is a member. This variable also determines link colors.
Board Size	Provides the size of the board.

The following table summarizes the variables in the node data set:

Table 9.6 Node Variables for the Board of Directors Sample

Variable	Description
Category	Classifies the corporations into categories such as retail, energy, manufacturing, and others. For individual people, this variable identifies the node as an individual and also indicates which individuals are chief executive officers.
Name	Serves as the Node ID variable that identifies all individuals and corporations.
Full Name	Provides the first and last name for individuals and the full company name for corporations.
Type	Indicates whether the node is an individual (I) or a corporation (C). This variable also determines the shapes of node markers.
Number of Boards	Provides the number of boards of which the individual is a member. For a corporation, this variable gives the size of the corporation's board.
Revenue	Provides the revenues for corporations. For individuals, this value is ignored.

Here are suggestions for exploring this data:

1. Open the project. The project loads the two data sets and opens associated graphs, including a hierarchical network graph.
2. Create a histogram based on link data, and select **NUMBER OF BOARDS** as the X variable.
3. In the histogram, select the bars with an X coordinate greater than or equal to four. The network graph changes to display only those links that correspond to multiple boards (four or more).

For an example use case, see “Example Use Case: Fortune 100 Boards of Directors” on page 137.

Web Path Data

This sample enables you to investigate Web path data. You can explore frequently visited sites and frequently followed links.

The sample uses the following files:

Project File:	WebPathData.nvw
Link Data Set:	webpath_links.sas7bdat
Node Data Set:	webpath_nodes.sas7bdat

The following table summarizes the variables in the link data set:

Table 9.7 *Link Variables for the Web Path Sample*

Variable	Description
ID1	Serves as the From variable that originates the link.
ID2	Serves as the To variable that terminates the link.
LinkID	Represents the ID for a link.
Count	Indicates the number of times the particular path (ID1 to ID2) is followed. This variable acts as a weight for the links.

The following table summarizes the variables in the node data set:

Table 9.8 *Node Variables for the Web Path Sample*

Variable	Description
ID	Provides a unique ID for each Web page. This variable serves as the Node ID variable in the Edit Data Attributes dialog box.
Count	Represents the number of hits for the particular Web page.

Variable	Description
Value	Represents the actual Web page name.

For an example use case, see “Example Use Case: Web Path Data” on page 148.

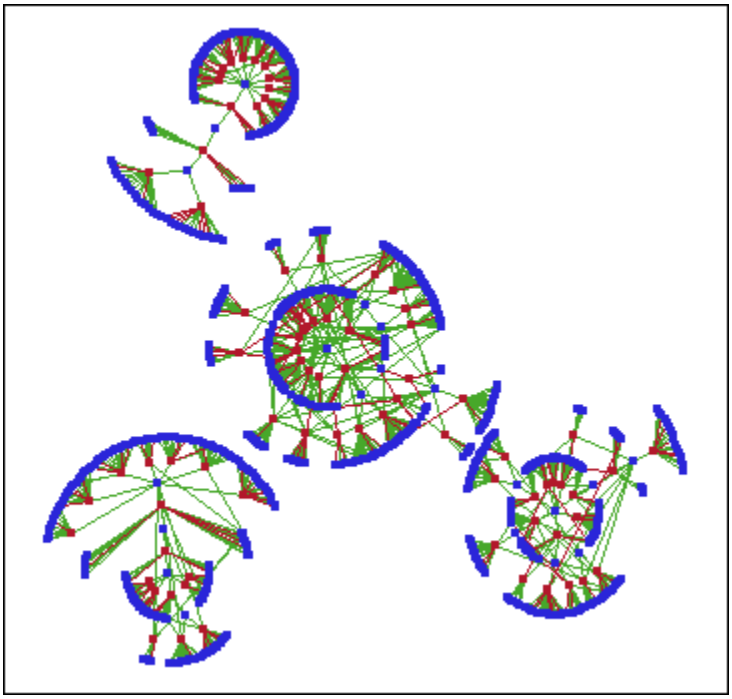
Example Use Case: Credit Card Fraud

Introduction to the Credit Card Fraud Example

This example describes how to use SAS/GRAPH Network Visualization Workshop to investigate patterns of credit card fraud. The data used here represent a sampling from a large credit card transaction database. All transactions included here, both valid and fraudulent, relate to customers with at least one fraudulent credit card transaction.

A visualization of this data is shown in the following figure. The blue nodes in the network correspond to merchants, and the red nodes correspond to customers. Each link represents a transaction between a customer and a merchant. The green links correspond to valid transactions, whereas the red links correspond to fraudulent transactions. (Remember that all links pertain to customers with at least one fraudulent credit card transaction in the database.)

Figure 9.1 *Credit Card Transaction Network*



As you can see from the figure, this example uses a lot of data. The goal of this example is to examine this large amount of data in order to detect patterns of credit card fraud.

Open the Credit Card Fraud Project

To open the project:

1. Open the ccFraudData.nvw project, which is located in the `Samples\Projects` subdirectory of your installation.

   This project automatically loads the node and link data sets. The project also loads the hierarchical network graph that you examine in this example.

2. Select **Data** ▶ **Edit Attributes**, and set or make sure that the attributes have been set as indicated here:

Link Attributes

Select **CUST_ID** from the **From** list box.

Select **MERCH_ID** from the **To** list box.

Select **FRAUD** from the **Color** list box.

Node Attributes

Select **NODE_ID** from the **ID** list box.

Leave the **Shape** list box set to **<None>**.

Select **TYPE** from the **Color** list box.

Select **NODE_ID** from the **Label** list box.

Note: This example describes the default settings for the ccFraudData project. If you have made changes earlier to this project, then your settings might differ from those described here. For example, if you have applied a different style (other than Default), then your nodes and links will have different colors.

About the Data Used in the Example

The data for this example is contained in two data sets: a node data set (ccnodes.sas7bdat) and a link data set (cclinks.sas7bdat).

The following figure shows a portion of the node data set:

Figure 9.2 *Node Data Set (ccnodes.sas7bdat)*

	7	NODE_ID	CATEGORY	SUBCATEGORY	TYPE	FRAUDULENT_TRANS_NUM	TOTAL_TRANS_NUM	GROUP
75	■	Customer075	Customer	25to34	c	3	15	4
76	■	Customer076	Customer	25to34	c	1	16	4
77	■	Customer077	Customer	Under25	c	2	11	4
78	■	Customer078	Customer	45to54	c	4	17	4
79	■	Customer079	Customer	65andOver	c	2	9	4
80	■	Customer080	Customer	25to34	c	4	13	4
81	■	Merchant0001	Retail	DrugStores	m	0	1	1
82	■	Merchant0002	Retail	FoodStore	m	0	1	1
83	■	Merchant0003	Services	Restaurants	m	1	1	1
84	■	Merchant0004	Services	Restaurants	m	0	1	1
85	■	Merchant0005	Services	OtherServices	m	0	1	1
86	■	Merchant0006	Services	OtherServices	m	0	1	1
87	■	Merchant0007	Retail	General	m	1	1	1
88	■	Merchant0008	Services	OtherServices	m	0	1	1
89	■	Merchant0009	Retail	GasStation	m	1	1	1

The following table summarizes the variables in this data set:

Variable	Description
NODE_ID	Lists the names of the customers and merchants. This variable serves as the Node ID variable in the Edit Data Attributes dialog box.
CATEGORY	Classifies the merchants (along with SUBCATEGORY).
SUBCATEGORY	Classifies the customers. Provides a second-level classification for merchants.
TYPE	Indicates whether the node is a customer (c) or merchant (m). This variable also determines the colors of the node markers.
FRAUDULENT_TRANS_NUM	Provides the number of fraudulent transactions.
TOTAL_TRANS_NUM	Provides the total number of transactions.
GROUP	Provides an additional grouping in order to facilitate examination of the data.

The following figure shows a portion of the link data set:

Figure 9.3 *Link Data Set (cclinks.sas7bdat)*

	9	CUST_ID	CUST_CAT	MERCH_ID	MERCH_CAT	MERCH_SUBCAT	FRAUD	GROUP
1	■	Customer001	65andOver	Merchant0001	Retail	DrugStores	0	1
2	■	Customer001	65andOver	Merchant0002	Retail	FoodStore	0	1
3	■	Customer001	65andOver	Merchant0003	Services	Restaurants	1	1
4	■	Customer001	65andOver	Merchant0004	Services	Restaurants	0	1
5	■	Customer001	65andOver	Merchant0005	Services	OtherServices	0	1
6	■	Customer001	65andOver	Merchant0006	Services	OtherServices	0	1
7	■	Customer001	65andOver	Merchant0007	Retail	General	1	1
8	■	Customer002	35to44	Merchant0008	Services	OtherServices	0	1
9	■	Customer002	35to44	Merchant0009	Retail	GasStation	1	1
10	■	Customer002	35to44	Merchant0010	Retail	OtherRetail	0	1
11	■	Customer002	35to44	Merchant0011	Services	OtherServices	0	1
12	■	Customer002	35to44	Merchant0012	Retail	NonStore	1	1
13	■	Customer002	35to44	Merchant0013	Services	OtherServices	0	1
14	■	Customer002	35to44	Merchant0014	Retail	FoodStore	0	1

The following table summarizes the variables in this data set:

Variable	Description
CUST_ID	Serves as the FROM variable in the Edit Data Attributes dialog box. Data values can be found in the NODE_ID variable of the node data set.
CUST_CAT	Indicates the customer classification data values. These values are taken from the SUBCATEGORY variable of the node data set.
MERCH_ID	Serves as the TO variable in the Edit Data Attributes dialog box. Data values can be found in the NODE_ID variable of the node data set.
MERCH_CAT; MERCH_SUBCAT	Indicates the merchant classification data values. These values are taken from the CATEGORY and the SUBCATEGORY variables of the node data set.
FRAUD	Indicates whether the transaction is fraudulent (1) or not fraudulent (0). This variable also determines link colors.
GROUP	Provides an additional grouping in order to facilitate examination of the data.

Note: The From Index and To Index variables are not shown in the figure or included in the previous table. These variables are automatically generated, zero-based indexes into the node data set.

Determining Your Strategy

One scenario for credit card fraud involves employees who work for the merchants in our data sets. In this scenario, the employee steals customer credit card numbers and either uses them or sells the numbers. The fraudulent transaction is typically not directly associated with the merchant where the employee works. Instead, the fraudulent transaction is connected to another merchant through a common customer with the original merchant. In a network graph, the fraudulent links (depicted in red here) are typically one link removed from the problem merchant. Therefore, rather than focus on fraudulent transactions, you will look at merchants with a significant number of customers. Keep in mind that all customers in the data set experienced at least one

fraudulent transaction. A merchant that has connections with many customers in the transaction network raises suspicions of fraud. Although this approach is not proof of fraud, it identifies specific merchants that had access to customer credit card numbers that have fraudulent transactions associated with them. These merchants warrant further investigation.

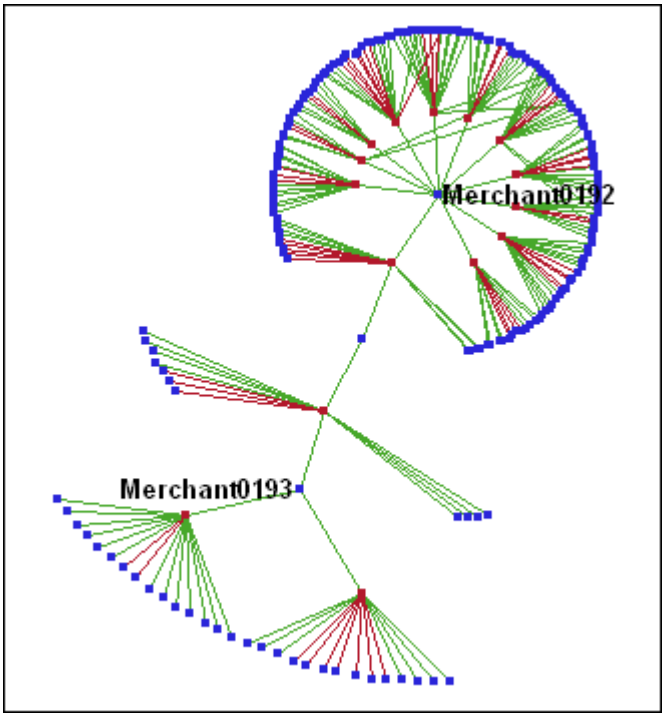
The degree of a node is the number of links having that node as an endpoint. In this example, the degree of a merchant is the number of customers that had a transaction with the merchant. The goal, then, is to identify merchants with high degree, which is defined here as greater than or equal to three.

Identifying Suspicious Merchants

Examine Merchants in Group 1

The subnetwork in the upper left portion of the network graph contains transactions between merchants and customers that are in Group 1. The following figure magnifies and shows this subnetwork.

Figure 9.4 Subnetwork of Transactions in Group 1



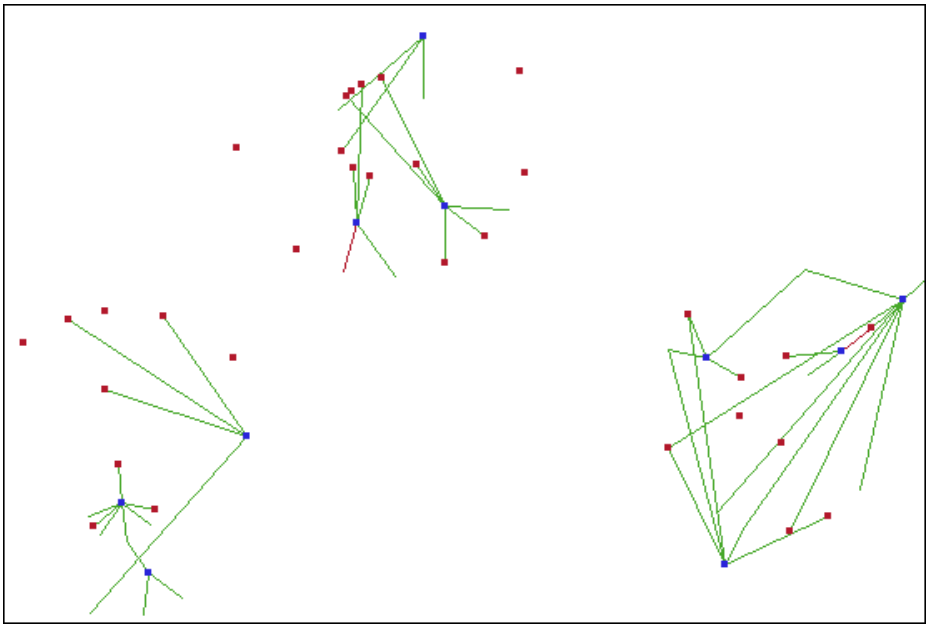
The simplicity of this subnetwork enables you to identify the merchants with high degree by direct observation. (Keep in mind that blue markers represent merchants, and red markers represent customers.) When you apply labels to the merchants with high degree, you can identify these merchants as Merchant0192 and Merchant0193 with a degree of 11 and 3, respectively.

Examine the Remaining Merchants

For the remaining three clusters in the network graph, the density of the subnetworks makes it difficult to detect all the merchants with high degree using direct observation. You can use a statistical graph and the local selection feature to filter the data in the network graph.

The following figure shows the result of using a scatter plot with local selection mode to parse the visualized data. In this graph, the blue nodes that are visible represent all the merchants with high degree in groups 2, 3, and 4.

Figure 9.5 *Merchants with High Degree in Groups 2, 3, and 4*



The following steps describe how to create this graph in the ccFraudData.nvw project:

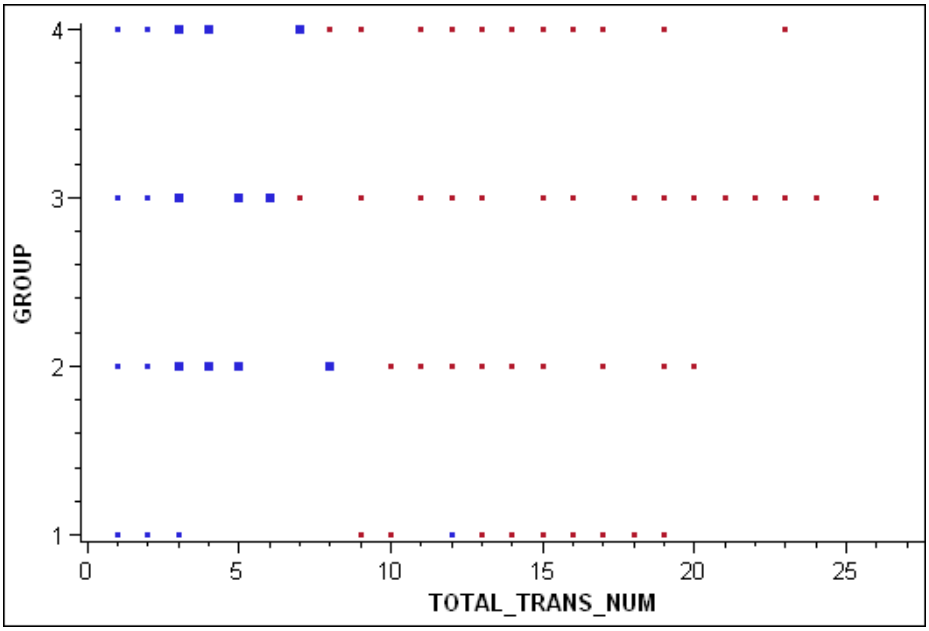
1 Click the node data table to activate it.

2 Create a scatter plot using the **Graphs** ▶ **Scatter Plot** menu option. Select **TOTAL_TRANS_NUM** as the X variable and **GROUP** as the Y variable.

3 In the scatter plot, select the merchants with high degree in groups 2, 3, and 4. To do this, select all blue nodes with an X coordinate greater than or equal to 3 and a Y coordinate greater than or equal to 2. Hold the CTRL key to select multiple nodes.

   The following figure shows the scatter plot with the merchants selected:

Figure 9.6 *Scatter Plot with Some Merchants Selected*



4 Select **Data** ▶ **Selection Mode** ▶ **Local**. An icon is displayed in the upper left corner of each graph. The icon for the network graph ( ) indicates that the graph has an observer-union role.

5 In the scatter plot, select all customers (all red nodes) with a Y coordinate greater than or equal to 2. The customer nodes appear in the network graph.

   In the network graph, the blue nodes that are visible represent all the merchants in groups 2, 3, and 4 that have high degree.

6 Select **Tools** ▶ **Interactive Zoom**. Then click the network graph to zoom in on the graph.

7 If you want to see the names of the merchants, select **Tools** ▶ **Label**, and then click the blue nodes in the network graph.

In summary, this example shows how to use SAS/GRAPH Network Visualization Workshop to investigate credit card fraud. You used the visualization features and observation filtering capabilities to identify merchants that warrant additional scrutiny with regard to fraudulent credit card transactions.

Example Use Case: Fortune 100 Boards of Directors

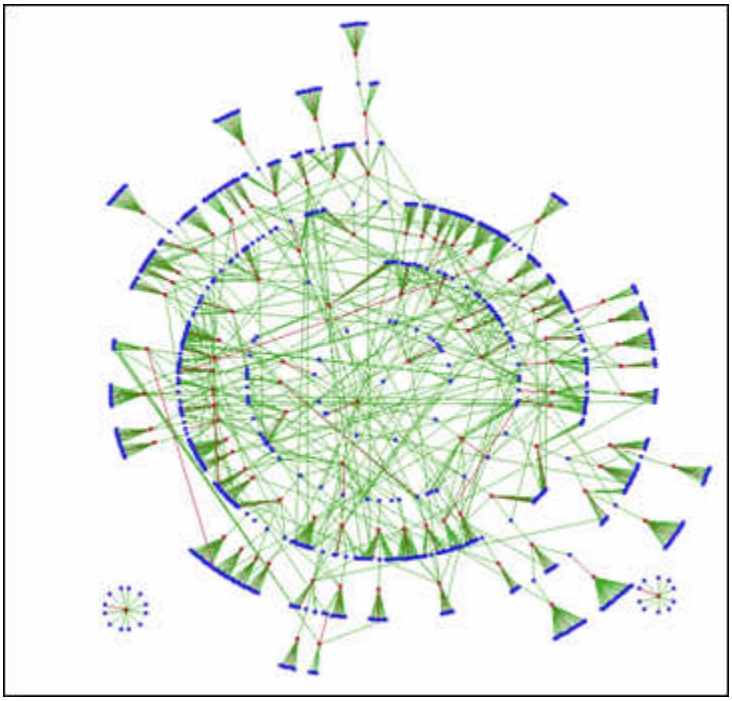
Introduction to the Boards of Directors Example

This example describes how to use SAS/GRAPH Network Visualization Workshop to investigate relationships among boards of directors of Fortune 100 companies circa the year 2001. The goal is to identify influential board members and uncover relationships among companies that share common directors.

The data for this example was obtained directly from various corporate Web sites.

A visualization of this data is shown in the following figure. The blue nodes in the network represent directors or chief executive officers (CEOs), and the red nodes represent corporations. Each arc links a director to a corporation. The red arcs link CEOs and their corporations, while green arcs link general directors and their corporations.

Figure 9.7 Boards of Directors Network



Open the Boards of Directors Project

To open the project:

1 Open the BoardOfDirectors.nvw project, which is located in the `Samples\Projects` subdirectory of your installation.

This project automatically loads the node and link data sets. The project also loads the hierarchical network graph that you examine in this example.

2 Select **Data** ▶ **Edit Attributes**. The Edit Data Attributes dialog box opens. Set or make sure that the attributes have been set as indicated here:

Link Attributes

Select **INDIVIDUAL** from the **From** list box.

Select **CORPORATION** from the **To** list box.

Select **CEO** from the **Color** list box.

Node Attributes

Select **NAME** from the **ID** list box.

Leave the **Shape** list box set to **<None>**.

Select **TYPE** from the **Color** list box.

Select **FULL_NAME** from the **Label** list box.

3 Select **View** ▶ **Style** ▶ **Default** to use the Default style that is seen in this example.

About the Data Used in the Example

The data for this example is contained in two data sets: a node data set (bdnodes.sas7bdat) and a link data set (bdlinks.sas7bdat).

The following figure shows a portion of the node data set:

Figure 9.8 *Node Data Set (bdnodes.sas7bdat)*

	6	CATEGORY	NAME	FULL NAME	TYPE	NUMBER OF BOARDS	REVENUES
1071	■	Individual	Coan	Gaylord_Coan	I	1	15000
1072	■	Individual	De Boon	Herman_De Boon	I	1	15000
1073	■	Individual	Mimran	David_Mimran	I	1	15000
1074	■	Individual	Mulroney_M	M_Mulroney	I	1	15000
1075	■	Individual	Strauss	Robert_Strauss	I	1	15000
1076	■	Individual	Vanier	John_Vanier	I	1	15000
1077	■	Individual	Webb_O	O_Webb	I	1	15000
1078	■	Individual	Young_A	Andrew_Young	I	1	15000
1079	■	Retail Trade	Walmart	Wal-Mart_Stores	C	16	219812
1080	■	Energy	ExxonMobil	Exxon_Mobil	C	13	191581
1081	■	Manufacturing	GM	General_Motors	C	14	177260
1082	■	Manufacturing	Ford	Ford_Motor	C	14	162412
1083	■	Energy	Enron	Enron	C	17	138718
1084	■	Manufacturing	GE	General_Electric	C	16	125913

The following table summarizes the variables in this data set:

Table 9.9 Node Variables for the Boards of Directors Sample

Variable	Description
NAME	Lists the names of individuals and corporations. This variable serves as the Node ID variable in the Edit Data Attributes dialog box.
FULL NAME	Lists the full names of individuals and corporations.
CATEGORY	Classifies the corporations and individuals. This variable also indicates which individuals are CEOs.
TYPE	Indicates whether the node is an individual (I) or a corporation (C). This variable also determines the colors of the node markers. This variable serves as the Color variable in the Edit Data Attributes dialog box and, therefore, determines node colors.
NUMBER_OF_BOARDS	Represents the number of boards of which an individual is member. For a corporation, this variable gives the size of the corporation's board.
REVENUES	Provides the revenues of corporations. For individuals, this value is ignored.

The following figure shows a portion of the link data set:

Figure 9.9 Link Data Set (bdlinks.sas7bdat)

	8	INDIVIDUAL	CORPORATION	BOARD CHAIRMAN	CEO	NUMBER OF BOARDS	BOARD SIZE
1		Breyer	Walmart	0	0	1	16
2		Chambers	Walmart	0	0	2	16
3		Coughlin	Walmart	0	0	1	16
4		Friedman_S	Walmart	0	0	2	16
5		Gault	Walmart	0	0	1	16
6		Glass	Walmart	0	0	1	16
7		Hernandez	Walmart	0	0	1	16
8		Lepore	Walmart	0	0	1	16
9		Reason	Walmart	0	0	1	16
10		Sanders	Walmart	0	0	1	16
11		Scott_H	Walmart	0	1	1	16
12		Shewmaker	Walmart	0	0	1	16
13		Soderquist	Walmart	0	0	1	16
14		Villarreal	Walmart	0	0	1	16

The following table summarizes the variables in this data set:

Table 9.10 Link Variables for the Boards of Directors Sample

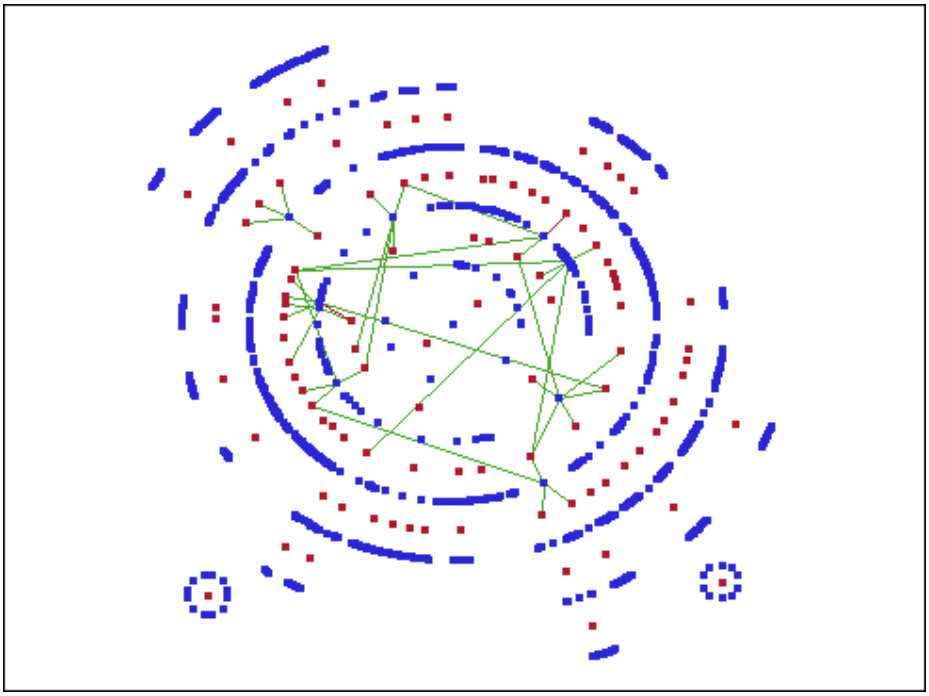
Variable	Description
INDIVIDUAL	Serves as the From variable that originates the link.
CORPORATION	Serves as the To variable that terminates the link.
BOARD CHAIRMAN	Indicates whether the individual is (1) or is not (0) the board chairman of the corresponding corporation.
CEO	Indicates whether the individual is (1) or is not (0) the chief executive officer of the corresponding corporation. This variable serves as the Color variable in the Edit Data Attributes dialog box and, therefore, determines link colors.
NUMBER OF BOARDS	Indicates how many boards the individual is a member of.
BOARD SIZE	Indicates the size of the board.

Note: The From Index and To Index variables are not shown in the figure or included in the previous table. These variables are automatically generated, zero-based indexes into the node data set.

Identifying Corporate Bonds

The density of the network makes it difficult to detect patterns involving individuals and corporations of interest, such as the individuals who sit on multiple boards. You can use a statistical graph in combination with the network graph in order to produce a subnetwork of the data. The result is a network graph that shows individuals who sit on four or more different boards.

Figure 9.10 *Directors on Four or More Boards*



In this graph, you can see that nine individuals sit on four or more different boards. There are nine blue nodes that connect with four or more red nodes. Two of these directors sit on five boards, and one director sits on six different boards. These patterns suggest possible relationships between companies that might not be readily apparent.

The following steps describe how to create this graph:

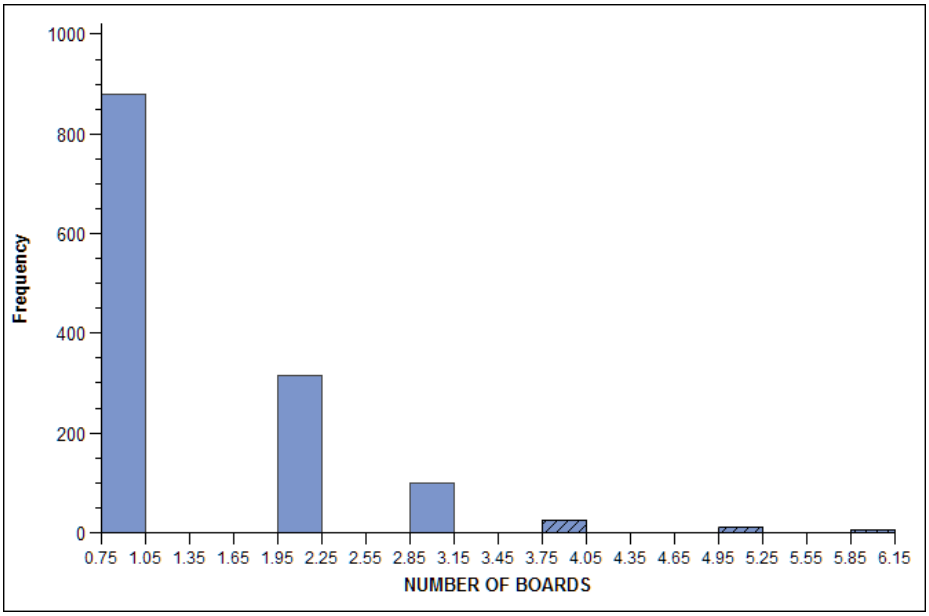
1 Click the link data table to activate it.

2 Create a histogram using the **Graphs** ▶ **Histogram** menu option. Select **NUMBER_OF_BOARDS** as the X variable. Leave the default **<Frequency>** for the Y variable.

3 In the histogram, select the bars with an X value greater than or equal to four. Hold the CTRL key to make multiple selections. Alternatively, drag a rectangle around the bars that you want to select.

   The following display shows the histogram with the bars selected:

Figure 9.11 *Histogram with Some Bars Selected*



4 Select **Tools** ▶ **Interactive Zoom**. Then click the network graph to zoom in on the graph.

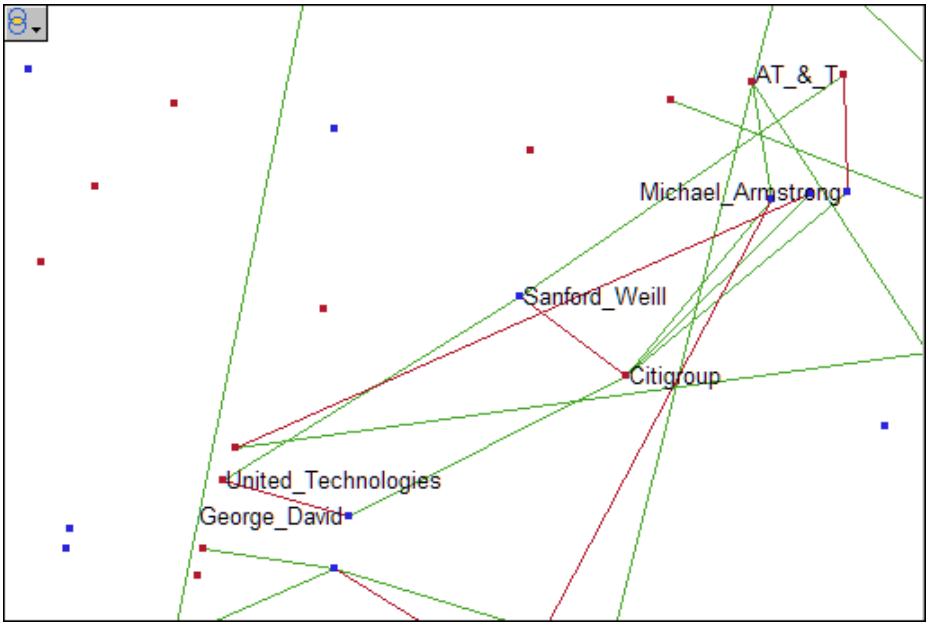
5 If you want to see the names of the CEOs and corporations, select **Tools** ▶ **Label**, and then click the blue nodes in the network graph.

Detecting Interlocks

According to Investor Responsibility Research Center's (IRRC) definition, there is an interlock when two CEOs sit on the boards of each other's companies. Interlocks are often viewed with a degree of suspicion; therefore, it's important to identify interlocks in the network. The subnetwork of interlocks is shown in the following figure. Here you can see there are two interlocks:

- Sanford Weill, the CEO of Citigroup, sits on the boards of AT&T and United Technologies.
- Michael Armstrong and George David, the CEOs of AT&T and United Technologies, respectively, sit on the board of Citigroup.

Figure 9.12 Subnetwork of Interlocks

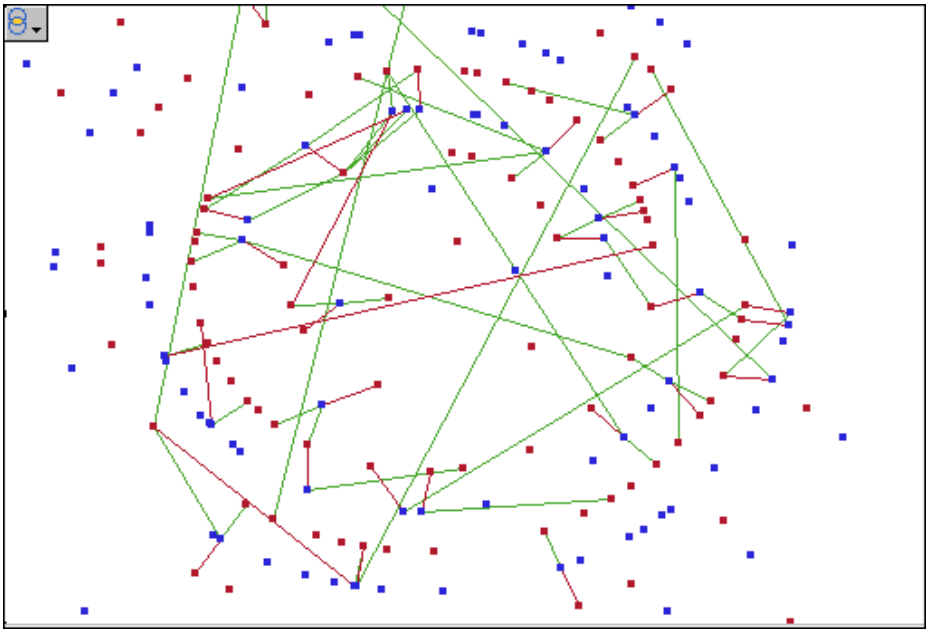


A quadrilateral consisting of two CEOs and two corporations represents an interlock. The interlocks can be detected by finding the quadrilaterals in this much sparser subnetwork.

You can use a statistical graph and the local selection feature to filter the data in the network graph.

The following figure shows the result of using a bar chart with local selection mode to parse the visualized data. In this graph, links are visible only for CEOs who sit on two or more boards.

Figure 9.13 *CEOs Who Sit on Two or More Boards*



The following steps describe how to create this graph in the BoardOfDirectors.nvw project:

1 Close any statistical graphs that you have open. Keep the single network graph open.

2 Click the node data table to activate it.

3 Create a bar chart using the **Graphs** ▶ **Bar Chart** menu option. Select **CATEGORY** as the category variable. Leave the default values in the other fields.

4 In the bar chart, select the bar for **CEO**.

5 Select **Data** ▶ **Selection Mode** ▶ **Local**. An icon is displayed in the upper left corner of each graph. The icon for the network graph ( ) indicates that the graph has an observer-union role.

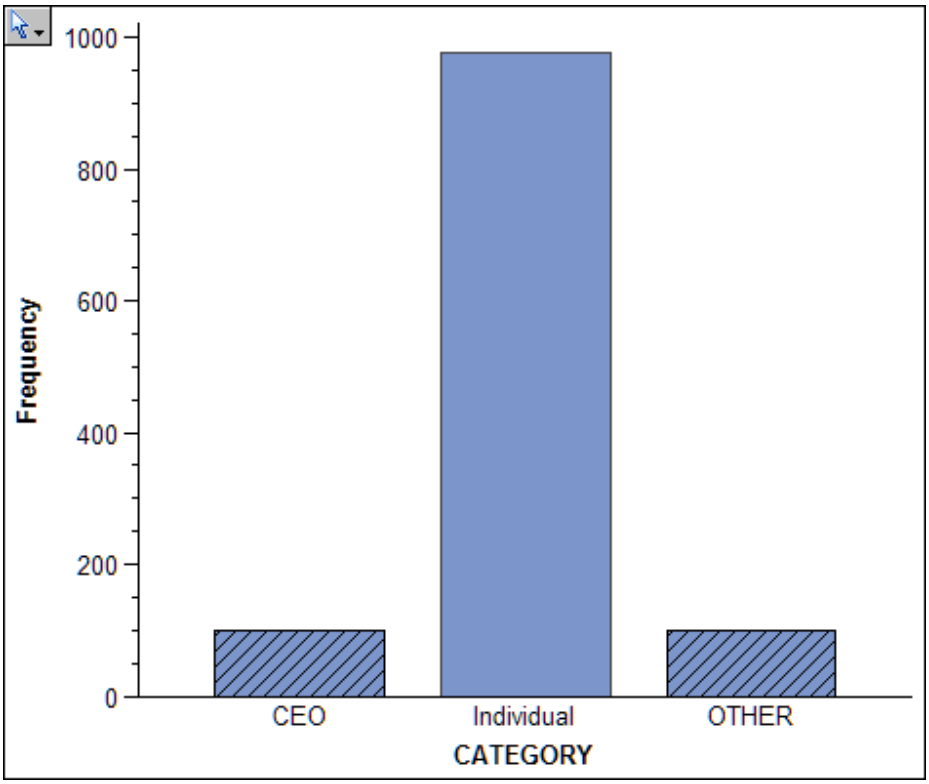
6 Select the network plot. Then click the arrow in the icon and select the observer-intersection icon ( ). The graph now has an observer-intersection role.

7 Select all observations in the node data table.

> **TIP** You can click the table and then press CTRL-A to select all observations.

8 In the bar chart, press CTRL and select the bar for **Other**.

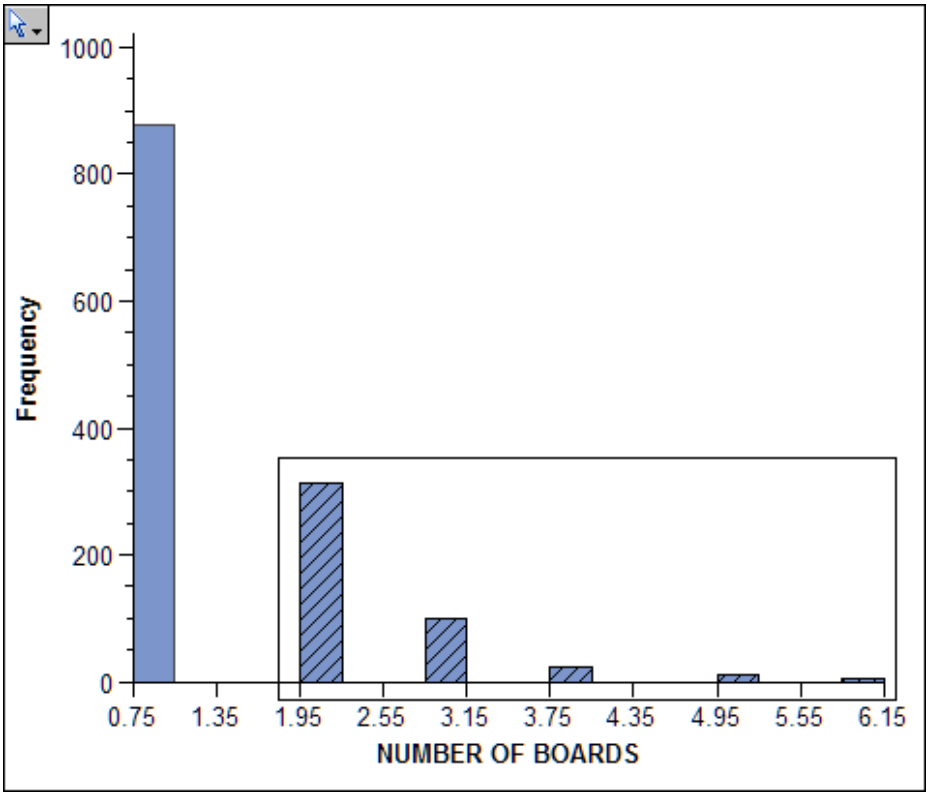
Figure 9.14 Bar Chart with CEO and OTHER Selected



9 Click the link data table to activate it. Then, create a histogram using the **Graphs** ▶ **Histogram** menu option. Select **NUMBER_OF_BOARDS** as the X variable. Leave the default **<Frequency>** for the Y variable.

10 In the histogram, select the bars for the numbers greater than or equal to two.

Figure 9.15 *Histogram with Number of Boards Greater than Two*



11 Select **Tools** ▶ **Interactive Zoom**. Then click the network graph to zoom in on the graph.

12 Select **Tools** ▶ **Pan**. Then click and drag the network graph so you can focus on the portion that contains the quadrilateral.

13 If you want to see the names of the CEOs and corporations, select **Tools** ▶ **Label**, and then click the blue nodes in the network graph.

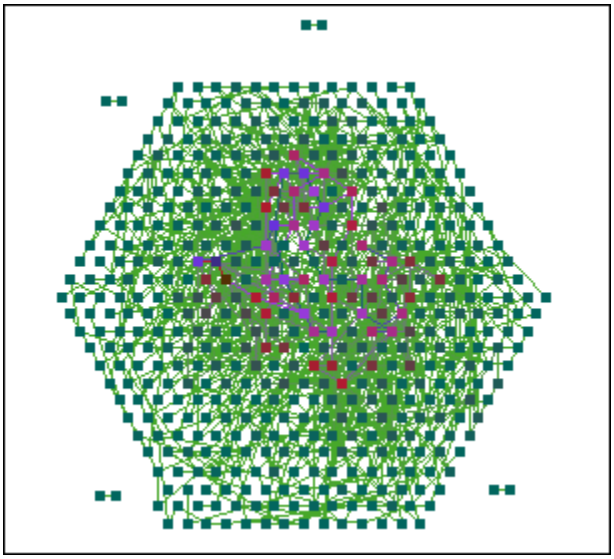
In summary, this example shows how to use SAS/GRAPH Network Visualization Workshop to identify relationships between directors and the boards on which they sit. You first identified individuals who sit on four or more different boards. Then you identified CEOs who sit on the boards of each other's companies.

Example Use Case: Web Path Data

Introduction to the Web Path Data Example

This example describes how to use SAS/GRAPH Network Visualization Workshop to investigate Web path data. The data used here represents Web paths that users follow when they browse Web pages on an intranet. The data has information about different Web pages (nodes) that users visited and the paths (links) that users followed after visiting a page.

A visualization of this data is shown in the following figure. The nodes other than green are the nodes that are visited the most. The nodes connecting frequently followed links are kept close to each other, and these nodes form the cluster of links at the center of the hexagon.



Open the Web Path Data Project

To open the project:

1 Open the WebPathData.nvw project, which is located in the **`Samples\Projects`** subdirectory of your installation.

   This project automatically loads the node and link data sets. The project also loads the hexagonal network graph that you examine in this example.

2 Select **Data** ▶ **Edit Attributes**. The Edit Data Attributes dialog box opens. Set or make sure that the attributes have been set as indicated here:

   Link Attributes

   Select **ID1** from the **From** list box.

   Select **ID2** from the **To** list box.

   Select **COUNT** from the **Color** list box.

   Node Attributes

   Select **ID** from the **ID** list box.

   Leave the **Shape** list box set to **<None>**.

   Select **COUNT** from the **Color** list box.

   Select **VALUE** from the **Label** list box.

3 Select **View** ▶ **Style** ▶ **Default** to use the Default style that is seen in this example.

About the Data Used in the Example

The data for this example is contained in two data sets: a node data set (webpath_nodes.sas7bdat) and a link data set (webpath_links.sas7bdat).

The following figure shows a portion of the node data set:

Figure 9.16 *Node Data Set (webpath_nodes.sas7bdat)*

	3	VALUE	ID	COUNT
1		/Home.jsp	N10	160
2		/Subcategory.jsp	N29	73
3		/Department.jsp	N6	32
4		/Product.jsp	N17	31
5		/Cookie_Check.jsp	N5	18
6		/Product_Detail.jsp	N18	7
7		/Cart.jsp	N2	7
8		/Summary.jsp	N30	7
9		/Order_Status.jsp	N15	5
10		/Billing.jsp	N1	4
11		/Product_Info.jsp	N19	4
12		/MiniDisc/Error.htm	N14	3
13		/Shipping.jsp	N25	3
14		/Site_Map.htm	N26	3
15		/Store/Store.htm	N28	3

The following table summarizes the variables in this data set that are relevant to the example:

Table 9.11 *Node Variables for the Web Path Sample*

Variable	Description
ID	Provides a unique ID for each Web page. This variable serves as the Node ID variable in the Edit Data Attributes dialog box.
COUNT	Represents the number of hits for the particular Web page. This variable is set as the color attribute in the Edit Data Attributes dialog box and, therefore, determines node colors.
VALUE	Represents the actual Web page name.

The following figure shows a portion of the link data set:

Figure 9.17 *Link Data Set (webpath_links.sas7bdat)*

	6	ID1	ID2	LINKID	COUNT
1141	■	N331	N352	L268	43
1142	■	N88	N118	L846	44
1143	■	N29	N59	L652	45
1144	■	N352	N375	L322	46
1145	■	N118	N147	L917	49
1146	■	N286	N311	L152	50
1147	■	N311	N331	L212	51
1148	■	N234	N261	L1184	52
1149	■	N205	N234	L1119	54
1150	■	N147	N179	L994	55
1151	■	N261	N286	L93	57
1152	■	N179	N205	L1060	60
1153	■	N37	N70	L722	96
1154	■	N10	N37	L593	100

The following table summarizes the variables in this data set:

Table 9.12 *Link Variables for the Web Path Sample*

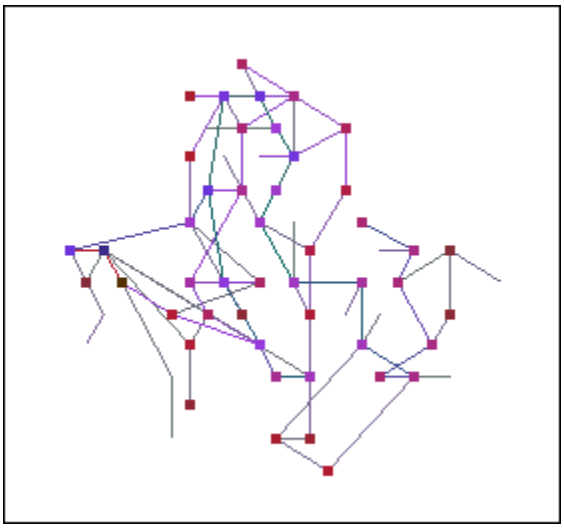
Variable	Description
ID1	Serves as the FROM variable that originates the link. Data values can be found in the ID variable of the node data set.
ID2	Serves as the TO variable that terminates the link. Data values can be found in the ID variable of the node data set.
LINKID	Represents the ID for a link.
COUNT	Indicates the number of times the particular path (ID1 to ID2) is followed. This variable acts as a weight for the links. This variable is set as the color attribute in the Edit Data Attributes dialog box and, therefore, determines link colors.

Note: The From and To variables are not shown in the figure or included in the previous table. These variables are automatically generated, zero-based indexes into the node data set.

Identifying the Most Popular Web Sites

You can use statistical graphs in combination with the network graph in order to produce a subnetwork of the data. The result is a network graph that shows frequently followed links and nodes with the maximum number of visits.

Figure 9.18 *Most Popular Web Sites*

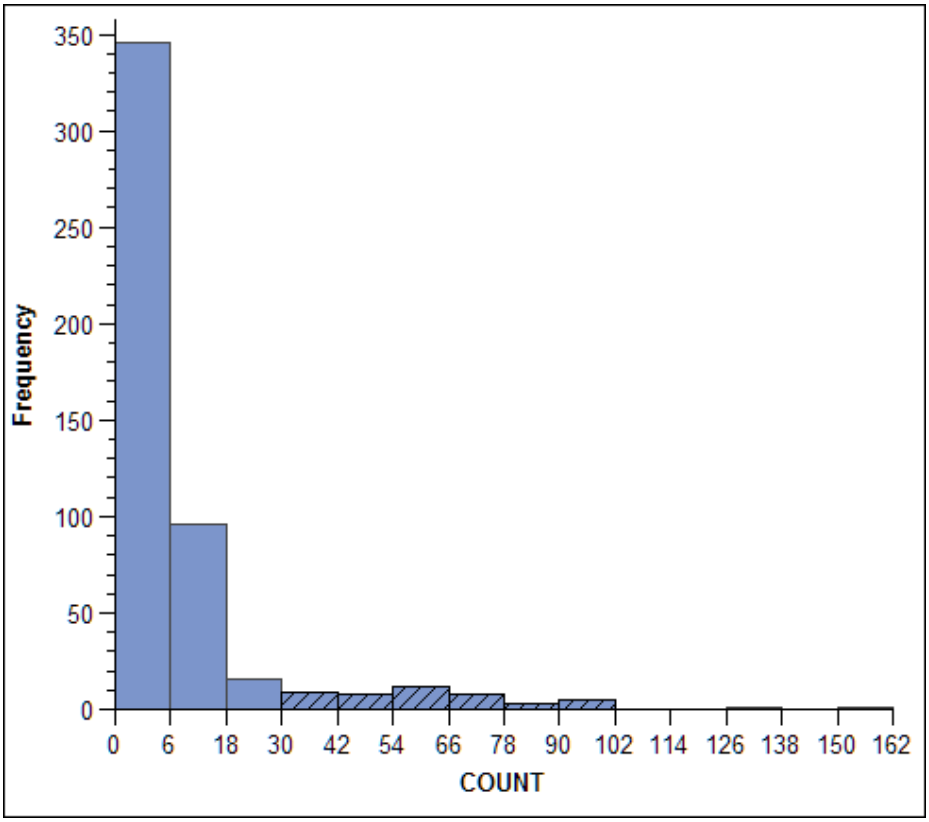


The following steps describe how to create this graph:

1. Click the node data table to activate it.
2. Create a histogram using the **Graphs** ▶ **Histogram** menu option. Select **COUNT** as the X variable. Leave the default **<Frequency>** for the Y variable.
3. In the histogram, select the bars with an X value greater than or equal to 30. Hold the CTRL key to make multiple selections. Alternatively, drag a rectangle around the bars that you want to select.

   The following figure shows the histogram with the bars selected:

Figure 9.19 *Nodes with a Count of 30 or More*



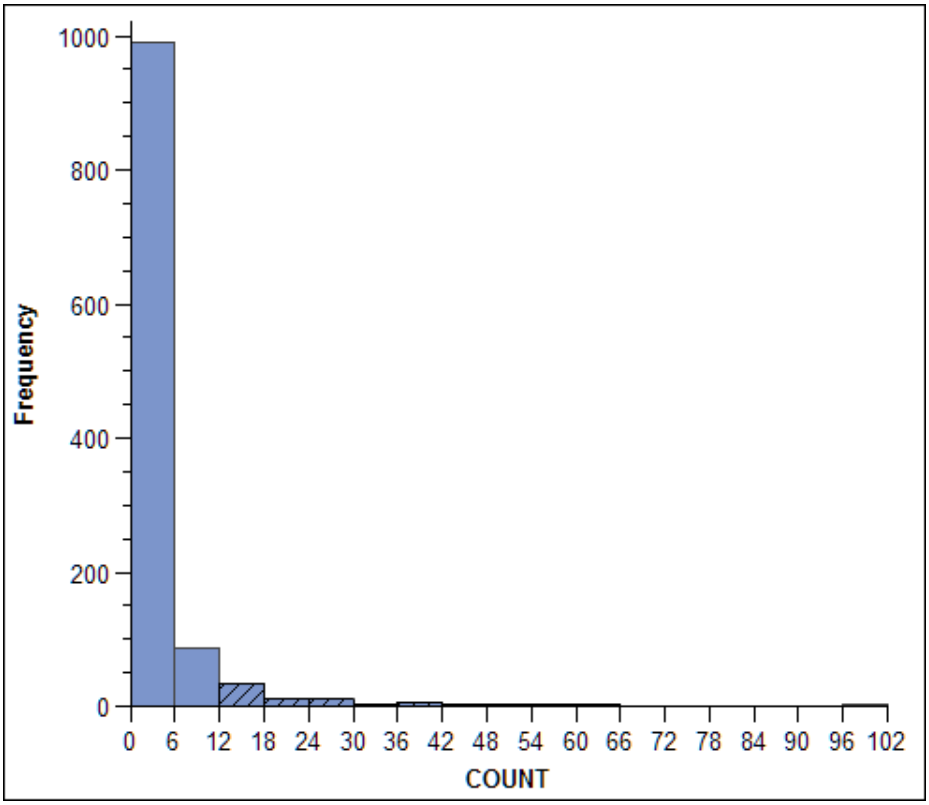
4 Click the link data table to activate it.

5 Create a histogram using the **Graphs** ▶ **Histogram** menu option. Select **COUNT** as the X variable. Leave the default **<Frequency>** for the Y variable.

6 In the histogram, select the bars with an X value greater than or equal to 12. Hold the CTRL key to make multiple selections. Alternatively, drag a rectangle around the bars that you want to select.

The following figure shows the histogram with the bars selected:

Figure 9.20 *Links with a Count of 12 or More*



7 Select **Tools** ▶ **Interactive Zoom**. Then click the network graph to zoom in on the graph.

8 If you want to see the names of the Web pages, select **Tools** ▶ **Label**, and then click the nodes in the network graph.

Index

F

G

H

L

M

N

O

P

R

S

T

V

W

Z