THE POWER TO KNOW.

SAS® Customer Link Analytics 5.5

User's Guide

SAS® Documentation

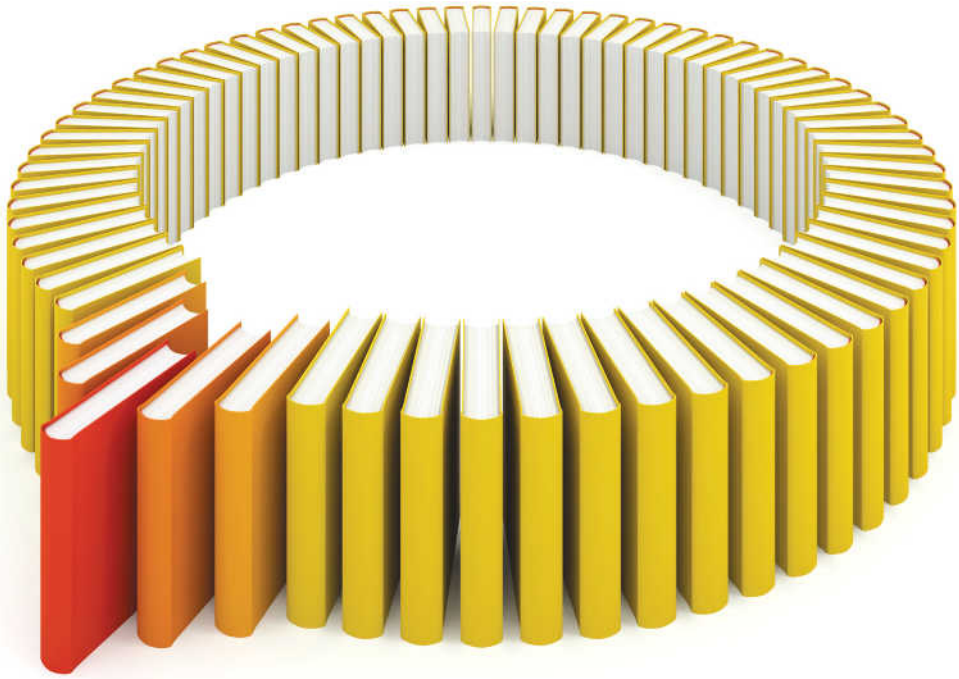
The correct bibliographic citation for this manual is as follows: SAS Insitute Inc 2014. *SAS® Customer Link Analytics 5.5: User's Guide*. Cary, NC : SAS Institute Inc.

SAS® Customer Link Analytics 5.5: User's Guide

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

Electronic book 1, March 2014

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.



Gain Greater Insight into Your SAS® Software with SAS Books.

Discover all that you need on your journey to knowledge and empowerment.

support.sas.com/bookstore
for additional books and resources.

Contents

Using This Book

Audience

This document focuses on explaining the tasks that you can perform by using the SAS Customer Link Analytics interface. You might be assigned to a specific role, which determines the tasks that you can perform. SAS Customer Link Analytics is designed for the following roles:

- administrators who are responsible for defining the transactional data and other data that is required for the network analysis.
- network analysts who perform the end-to-end tasks from extracting transactional data to building communities and assigning roles.

Prerequisites

Users of SAS Customer Link Analytics should have high-level analytical capabilities, strong reporting skills, and high-level knowledge of data management. Familiarity with concepts of network analysis and graph theory can be an added advantage.

x

Accessibility

Accessibility Notice

For more information about the accessibility of this product, see **Accessibility Features of SAS Customer Link Analytics 5.5 at support.sas.com**.

Recommended Reading

SAS Customer Link Analytics is supported by the following documents:

- *SAS Customer Link Analytics: Administrator's Guide*
- *SAS Customer Link Analytics: Data Reference Guide*

Other relevant documents include the following:

- *SAS OPTGRAPH Procedure: Graph Algorithms and Network Analysis*
- *SAS OPTGRAPH Procedure: High-Performance Features*
- *SAS Visual Analytics: User's Guide*
- *SAS Visual Analytics: Installation and Configuration Guide*
- *SAS Visual Analytics: Administration Guide*

For a complete list of SAS books, go to support.sas.com/bookstore. If you have questions about which titles you need, please contact a SAS Book Sales Representative:

SAS Books
SAS Campus Drive
Cary, NC 27513-2414
Phone: 1-800-727-3228
Fax: 1-919-677-8166
E-mail: sasbook@sas.com
Web address: support.sas.com/bookstore

Part 1

Introduction

1

About SAS Customer Link Analytics

Overview of SAS Customer Link Analytics

In recent years, customers have become more sophisticated and well-informed in their buying decisions than ever before. They rely on and seek advice from their network of friends, family, and acquaintances. As a result, there is explosive growth in the number of subscriptions on the communications service provider's network. Also, with the increasing penetration of people subscribing to communications services, the traditional methods of campaigning, such as telemarketing and advertising, are no longer necessarily applicable. Therefore, it is imperative for marketers to develop marketing strategies based on meaningful insights that are gained from network or transactional data analysis. This data captures interactions in the subscriber base, such as how much they talk, with whom they talk, and so on. The strength of relationships within their network subscriptions and outside their network can reveal more information than such static attributes as the demographics of their subscriptions. Marketers can then use these insights to target their customers more accurately and effectively.

SAS Customer Link Analytics is a comprehensive solution for analyzing and controlling network data processing. It enables marketers to analyze network data, identify communities within the network, and quantify the relative importance of nodes within a community or network from various aspects. In addition, it enables marketing analysts to identify the role that each node plays within its community.

Business Benefits

More Effective Marketing Strategies

The roles that are identified within the network or the community-based variables enable creation of enhanced marketing actions and campaigns. Also, targeting campaigns for subscriptions (nodes) that have certain roles such as leaders has a higher chance of reaching multiple subscriptions within the neighborhood and the community of leaders. This strategy can result in reducing campaign costs and increasing return on investment (ROI). In addition, by using SAS Customer Link Analytics, marketers can understand the relative effectiveness of communications within different networks of their subscriptions and create more relevant marketing messages. They can also analyze the behavior that is embedded in networks of their subscriptions. This analysis enables them to exploit different patterns to broadcast processes that are related to products and services acquisition. Ultimately, this strategy prevents churn.

Better Identification of Customer Segments

Instead of merely profiling customers based on their demographics and usage patterns, SAS Customer Link Analytics empowers customer profiling based on network or transactional data intelligence. The solution extracts customers' transactional information from external data sources and uses it to unveil communities that exist within their subscriptions. Marketers can understand how the networks of their subscriptions operate and who they should target. In addition, the solution enables marketers to assign roles such as leaders and followers within networks and communities. These role-based variables can further be used in traditional segmentation of customers as value-added information. This enables marketers to provide more relevant and better targeted campaigns. Some key customers, the leaders, can play a crucial role in targeting the right subset of customers. Also, dense communities can be targeted for faster revenue realization.

Insights into Viral Effect

Traditionally, demographic and behavioral variables are used in predictive models in the customer intelligence area of the communications industry. SAS Customer Link Analytics opens up avenues for including variables that can be defined on the basis of type as well as on the strength of relationships and the interactions between your subscriptions. These additional variables can enrich predictive models that are used for customer intelligence. Also, as information flows within the networks through links among customers, a campaign is initiated from leaders and spreads using their connections.

Solution Features

Embedded Analytics

SAS Customer Link Analytics leverages the power of analytics to analyze network data or transactional data. It uses various network data processing algorithms of advanced analytics to analyze networks. In addition, the solution provides the capability to incorporate static data (demographics) and transactional data.

Quick Implementation

SAS Customer Link Analytics is a stand-alone solution. It does not require a separate data layer that needs to be configured and populated for its use. For example, for the communications industry, after the aggregated call detail record (CDR) data is configured into the SAS Customer Link Analytics application, users can analyze the network data instantly.

Workflow-Based Approach

SAS Customer Link Analytics is built around the concept of a project that is associated with a predefined workflow. The workflow provides a guided approach to perform all the tasks that are involved in building and analyzing communities. In addition, SAS Customer Link Analytics provides an intuitive interface that is designed for marketers. At each workflow step, it provides reports that provide node-level and community-level metrics. These results provide insight into the significance of each node or link both within and across communities.

Reporting Data Preparation

A SAS Customer Link Analytics project produces outputs such as communities, centrality measures (that indicate the relative importance of nodes), and roles. This output information and the node-level or link-level attributes can be automatically processed by using the reporting data preparation feature of SAS Customer Link Analytics. If SAS Visual Analytics is available, the resulting data can be further exploited within SAS Visual Analytics Explorer to build more meaningful reports.

Viral Effect Analysis

Unlike the traditional predictive modeling that is used in the customer intelligence area of the communications industry, viral effect analysis emphasizes the interactions between subscriptions that are captured through network or transactional data. The viral effect modeling capability assesses the behavior of a leader within a community to understand its impact within both the community and the neighborhood of that leader. Viral effect analysis uses the output of SAS Customer Link Analytics, such as communities, centrality measures, and roles along with transactional and demographic measures that are configured in the application. It uses SAS Rapid Predictive Modeler to build

and register analytical models automatically. Its automated data processing also creates an enhanced set of meaningful variables that serve as an input to predictive models. As an output, viral effect analysis predicts the score of a node for the analysis under consideration.

Distributed and Non-Distributed Modes

SAS Customer Link Analytics can run on a computer grid or on a single computer system with multiple CPUs. Running on a computer grid is referred to as *distributed mode* of operation. Running on a single computer system with multiple CPUs is referred to as *non-distributed mode* of operation. For details, see *SAS Customer Link Analytics: Administrator's Guide*.

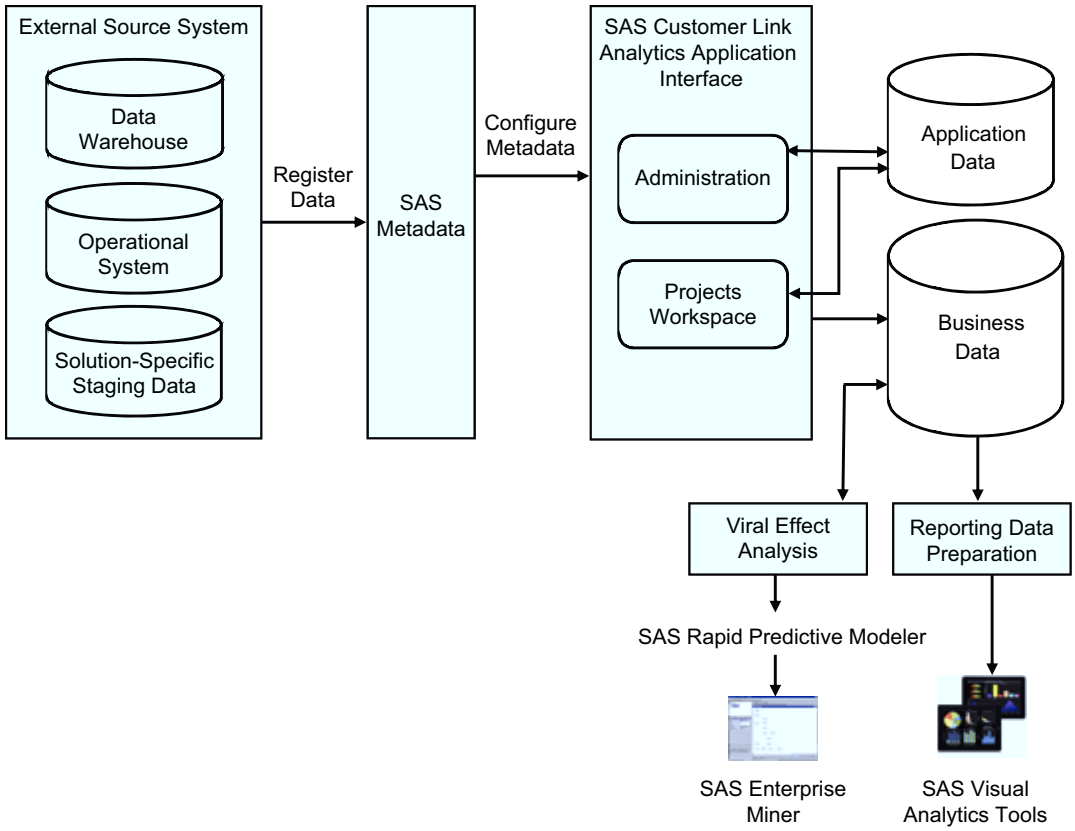
Depending on whether SAS Customer Link Analytics is operating in distributed or non-distributed mode, the fields and the information that is displayed for certain workflow steps differs. The relevant chapters of this guide provide information for both these modes wherever applicable.

How SAS Customer Link Analytics Works

Overview

SAS Customer Link Analytics is a comprehensive solution that interacts with an external source system to extract and process subscription network data.

Figure 1.1 Solution Flow Diagram



The solution comprises the following components:

External source system
: the system with which SAS Customer Link Analytics interfaces to extract transactional data and other information such as node and link attributes. SAS Customer Link Analytics uses this data to build communities. The external source system can be a data warehouse, an operational system, or staging data that is specific to SAS Customer Link Analytics.

SAS metadata
: the data from the external source system has to be registered in SAS metadata. This registered data can then be imported into SAS Customer Link Analytics and used by the application.

SAS Customer Link Analytics interface
: a workflow-based application to configure metadata, define projects, configure and run workflow steps, and view reports.

Application data
: stores project-specific data and configuration details of the source data. Also, stores the summary of results that the SAS Customer Link Analytics solution produces when each workflow step is run. These results include information about communities, roles, and centrality measures.

Business data
: stores the intermediate tables that SAS Customer Link Analytics creates when the workflow steps are run. The business data also contains the final output that SAS Customer Link Analytics produces when all the workflow steps of a project are run. This output contains node-level information such as the role ID, community ID, and centrality values. In addition, business data

contains reporting data that SAS Customer Link Analytics can produce in a format that can be further exploited by reporting tools such as SAS Visual Analytics to build meaningful reports.

SAS Visual Analytics tools
: reporting tools for business analysts to explore, view, and analyze data and create and view reports that help them make business decisions.

Viral effect analysis
: analysis of SAS Customer Link Analytics output, construction of an analytical model using SAS Rapid Predictive Modeler, and generation of analytical scores for viral effect analysis. The analytical model can be further explored in SAS Enterprise Miner and the scores can be used by marketing automation tools.

Solution Flow

The SAS Customer Link Analytics solution flow includes the following steps:

1. Register data from the external source system in SAS metadata. This step is not within the scope of this document.
2. Log on to SAS Customer Link Analytics as an administrator and perform the following tasks:
   a. Import tables that are registered in SAS metadata, configure them, and then refresh the transactional tables. For details, see Chapter 15, “Working with Tables,” on page 85.
   b. Define source data profiles. For details, see Chapter 16, “Working with Source Data Profiles,” on page 97.
3. Log on to SAS Customer Link Analytics as a network analyst and define a project. For details, see “Create a Project” on page 26.
4. Complete the workflow steps here:
   a. Select the nodes and links whose data you want to analyze and then extract summarized transaction data. For details, see “Configure the Data Extraction Workflow Step” on page 40.
   b. Filter the links based on specific parameters and assign weights to the links. For details, see “Configure the Link and Node Processing Workflow Step” on page 44.
   c. Build communities by selecting the appropriate analytical approach. For details, see “Configure the Community Building Workflow Step” on page 51.
   d. Select centrality measures that you want to compute and provide input parameters to compute these measures. For details, see “Configure the Centrality Measures Computation Workflow Step” on page 58.
   e. Assign a role to each node of the communities. For details, see “Define a Role” on page 64.
5. (Optional) Promote a project to batch mode. For details, see Chapter 12, “Batch Processing,” on page 71.

6 (Optional) View and explore the output produced by SAS Customer Link Analytics. For details, see Chapter 14, “Output of SAS Customer Link Analytics ,” on page 77.

7 (Optional) Perform viral effect analysis. For details, see Chapter 17, “Introduction to Viral Effect Analysis,” on page 103.

2

Managing Access to SAS Customer Link Analytics

Roles and Capabilities

Different users might have access to different functionality depending on their assigned roles. Each role is mapped to a set of predefined capabilities. A *capability*, also known as an *application action*, defines the operations that a user can perform.

SAS Customer Link Analytics has two predefined roles—network analysts and administrators. A predefined set of capabilities is available for each role. Capabilities are further categorized into two levels—project-related and administrative.

Table 2.1 *Predefined Roles*

Role Name	Description
Cust Link Analytics: Network Analysis	Provides capabilities to view and create projects.
Cust Link Analytics: Administration	Provides capabilities to view, create, and delete source data profiles and tables.

Table 2.2 *Predefined Capabilities*

Capability Name	Description
Project-related capabilities	
Create New Project	Enables you to create a new project.
View All Projects	Enables you to view any project.

Capability Name	Description
View Owned Projects	Enables you to view only those projects that you have created.
Edit All Projects	Enables you to edit any project.
Edit Owned Projects	Enables you to edit only those projects that you have created.
Delete All Projects	Enables you to delete any project.
Delete Owned Projects	Enables you to delete only those projects that you have created.
Change All Projects Owner	Enables you to change the owner of any project.
Change Owned Projects Owner	Enables you to change the owner of only those projects that you have created.
Change All Projects Mode	Enables you to change the mode of any project.
Change Owned Projects Mode	Enables you to change the mode of only those projects that you have created.
Administrative capabilities	
Create New Source Profile	Enables you to create a new source data profile.
Delete Source Profile	Enables you to delete an existing source data profile.
Import New Table	Enables you to import a table.
Remove Imported Table	Enables a user to remove a table that you have imported.

Log On to SAS Customer Link Analytics as a Registered User

When your administrator installs SAS Customer Link Analytics, an internal user ID (sasadm@saspw) is created by default. You can log on to SAS Customer Link Analytics by using this user ID. However, all the capabilities of the software are not granted to this user. As a result, this user cannot perform the following tasks:

- create a project
- run a workflow step

- change the mode of a project
- refresh data in a transactional table

Therefore, it is recommended that you define your own users and assign appropriate roles and capabilities to them. For details, see *SAS Customer Link Analytics: Administrator's Guide*.

To log on to SAS Customer Link Analytics:

1. Click on the URL that is supplied to you by your system administrator or paste it into the address field of your browser. For example, you can enter `http://server01.abc.com/SASCustLinkAnlytics` as the URL. The logon window appears.
2. In the **User ID** field, enter your user ID.
3. In the **Password** field, enter the password for the user ID that you have specified.
4. Click **Log On**.

   The main application window appears. For details, see Chapter 3, "The SAS Customer Link Analytics Interface," on page 15.

   Note: Your password is case-sensitive. Also, your user ID might be case-sensitive depending on the operating system that is used to host the web application server. If you need assistance, contact your system administrator.

Log Off from SAS Customer Link Analytics

To log off from SAS Customer Link Analytics, on the Application bar, click **Log Off** .

If your connection to SAS Customer Link Analytics remains inactive for a certain time, then your session might time out. By default, the session time-out due to inactivity is 30 minutes. Your administrator can change this duration. The application prompts you about your inactive session, and you can log on and continue with your session. However, if you lose your connection, then you must begin at the same point where you last saved your work. As a best practice, save your work frequently.

Log On to SAS Customer Link Analytics from SAS Visual Analytics Hub

If SAS Visual Analytics is available, you can log on to SAS Customer Link Analytics from SAS Visual Analytics Hub.

To log on to SAS Customer Link Analytics from SAS Visual Analytics Hub:

1. Log on to SAS Visual Analytics Hub. The main page appears.
2. In the Common Actions pane, select **Analyze Customer Link Data**. The main application window of SAS Customer Link Analytics appears. For

details, see Chapter 3, “The SAS Customer Link Analytics Interface,” on page 15.

3 To return to SAS Visual Analytics Hub, on the banner area, click **Home**. To log off from SAS Customer Link Analytics, on the Application bar, click **Log Off** .

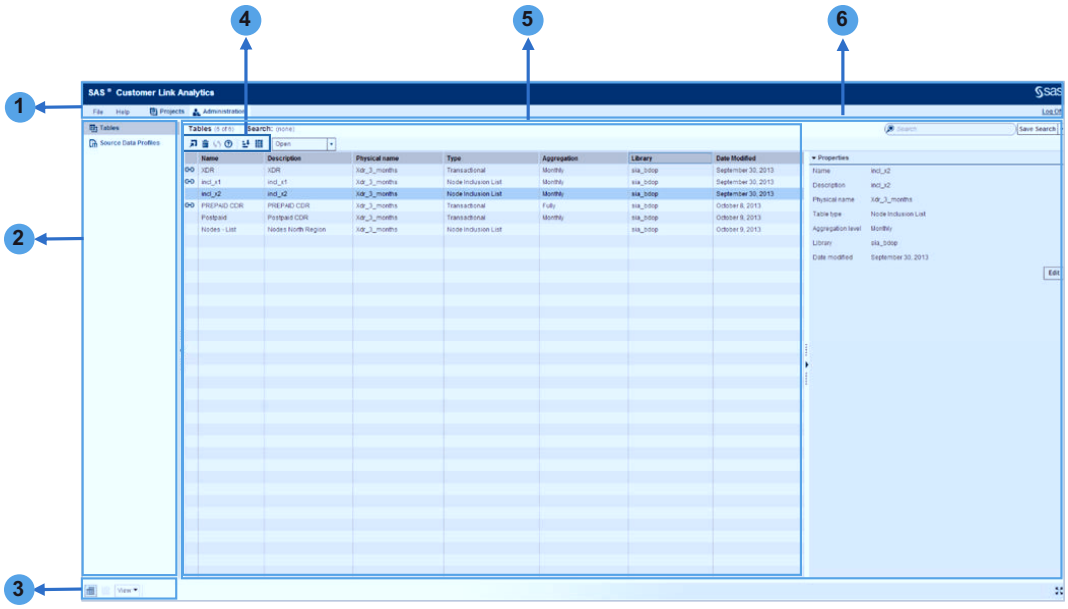
3

The SAS Customer Link Analytics Interface

Overview of the SAS Customer Link Analytics Interface

When you log on to SAS Customer Link Analytics, the main application window appears. The main application window comprises the application bar and the workspace. The application bar and workspace are further divided into various sections, as depicted in the following diagram.

Figure 3.1 The SAS Customer Link Analytics Interface



1 Application bar
2 Category pane
3 Tile pane
4 Toolbar
5 Object navigator
6 Details pane

Using the Application Bar

Overview

The application bar is located at the top of the application window and is part of the application banner. The application bar contains the menu bar, the workspace buttons, and the **Log Off** option.

The Menu Bar

The File Menu

The following options are available on the **File** menu:

Preferences
enables you to define your preferences for displaying information in the user interface (UI).

Log Off
enables you to log off from the application.

The Help Menu

The following options are available on the **Help** menu:

How-To Topics
displays a list of Help topics. These topics provide information about how to perform certain tasks.

User's Guide (PDF)
opens *SAS Customer Link Analytics: User's Guide* in a separate browser window.

SAS on the Web
includes links to the product support page, SAS customer support, and the SAS home page. The product support page gives information about the documentation that is available for SAS Customer Link Analytics.

About SAS Customer Link Analytics
displays copyright and other information about SAS Customer Link Analytics.

The Workspace Buttons

The workspace buttons enable you to navigate across workspaces.

Table 3.1 *Workspace Buttons*

Button	Button Name	Purpose
	Projects	Enables you to define new projects and work on their workflow steps.
	Administration	Enables you to import tables and define source data profiles.

The Log Off Option

The **Log Off** option enables you to log off from the application.

Working in the Workspaces

Overview of the Workspaces

The SAS Customer Link Analytics interface contains workspaces for performing a group of related tasks within the application. You can navigate across the workspaces by using the workspace buttons that are available on the application bar.

The SAS Customer Link Analytics interface contains the following workspaces:

- Projects
- Administration

The Category Pane

The Category pane enables you to select an object such as a table or a source data profile that you want to work on.

The Toolbar

The toolbar contains options that enable you to quickly complete certain commonly performed tasks.

Note: The toolbar options might not be the same for each workspace.

Table 3.2 Toolbar Options

Icon	Description	Action
	New	Defines a new object such as a project, a role, or a source data profile.
	Delete	Deletes an object.
	Refresh	Restores the details of an object.
	Import	Enables you to import a table.
Not available	Menu	Provides options to open an object or make it available in the Tile pane.

Object Navigator

The object navigator displays a list of objects that are defined for the type of object that you have selected in the Category pane.

Details Pane

The Details pane displays information about the object that you have selected in the object navigator.

Tile Pane

The Tile pane displays the objects that you have opened from the object navigator.

Help Overview

Help is embedded in the SAS Customer Link Analytics interface as various components. For example, Help pop-ups and tooltips give required information to users whenever needed. In addition, the *SAS Customer Link Analytics: User's Guide* is available from the **Help** menu.

The following types of online Help are available with SAS Customer Link Analytics:

Tooltip Help
: displays short, descriptive information about fields, columns, and icons in a pane. Tooltip Help appears automatically, when you move your pointer over an item on your screen. For example, tooltips display the purpose of each toolbar option.

Pop-up Help
: displays detailed information about fields and columns in a pane. Pop-up Help is represented using the Help icon. Click  to view the information, which is displayed in a pop-up dialog box. For example, while working in a window, users might need complete details of certain fields in order to provide appropriate information.

How-To Topics
: displays a list of Help topics that can be accessed from the **Help** menu. These topics describe how to perform certain tasks.

4

Performing Common Tasks

Overview of Commonly Performed Tasks

Here are the common components of the SAS Customer Link Analytics interface:

- panes
- sections
- tables
- lists

This chapter explains the common tasks that you can perform when you work on these components.

Resize a Pane

You can resize a pane if the default size of the pane does not display complete information. Also, you can resize a certain pane in order to provide more area for other panes.

To resize a pane:

1 Move your pointer to the horizontal or vertical divider.

2 When your pointer changes to a ⇹, drag to resize the pane.

Working with Tables

Change the Width of a Column

If the default width of a column does not display complete information, you can change the column width. To change the width of a column, drag the boundary on the right side of the column heading until the column is the width that you want.

Move a Column

If the order in which the columns are displayed does not meet your needs, you can change the sequence of the columns. To move a column, click the column heading and drag the column to the desired location.

Change the Sort Order of a Column

You can sort data by only a single column at a time. To change the order in which a column is sorted, click the up or down arrow that is displayed in the column heading.

Working in Multiple Windows at a Time

SAS Customer Link Analytics enables you to work in multiple windows at a time. This feature enables you to compare definitions of various objects such as projects, tables, and source data profiles. For example, you can compare the results of workflow steps of two or more projects in separate windows. You can also use this feature to work on multiple objects of similar types together.

To open multiple windows:

1. From the list that is displayed in the respective workspace, select the object that you want to open.
2. On the toolbar, click the menu, and then select **Send to Tile Pane**.
3. Repeat steps 1 and 2 for the other objects.
4. In the Tile pane, hold down the Ctrl key and select the objects that you want to view. Each object opens in a separate window.

Part 2

Projects Workspace

5

Managing Projects

About Projects

SAS Customer Link Analytics is built around the concept of projects. A project focuses on a specific network within which you want to build communities and assign a role to each customer. For example, in the communications industry, you might want to analyze the subscriptions in the North region. In this case, you can create your project so that the network includes subscriptions that belong to the North region. Moreover, a project enables you to group, organize, and track all your tasks that you need to perform in order to build communities and assign roles. These tasks are called *workflow steps*. When you define a project, a set of predefined workflow steps is assigned to it. You complete all the tasks that are relevant to a project in the Projects workspace.

About the Projects Workspace

Overview of the Projects Workspace

The Projects workspace is divided into three sections:

Category pane
: collapse this pane in order to provide more space for the other sections of the workspace.

Details pane
: displays a list of projects that you have defined so far.

Properties pane
: displays the properties of the project that you have selected in the Details pane. Also, you can modify the project's properties in this pane.

View and Configure the Workflow Steps of a Project

Each project is associated with a set of predefined workflow steps. The default values that are required for each workflow step are pre-populated. However, you can configure these values according to your requirements. To view or configure the workflow steps of a project, double-click the project in the Details pane. Alternatively, you can select the project and then select **Open** from the menu that is displayed on the toolbar.

Modes of a Project

A project can be in any one of the following modes:

Design mode
: When you successfully run all workflow steps, a project completes one run in design mode. You can then push a project to batch mode.

Batch mode
: After you push a project to batch mode, you have to schedule the batch run of the project. In the batch run, all workflow steps that you performed in design mode are automatically run. You can view the progress of the project run in the Workflow Diagram pane. However, you cannot make any changes to a project's properties or its workflow steps. Neither can you delete a project that is in batch mode. If you are not satisfied with the results in batch mode, you can pull a project back into design mode. You can configure and run the workflow steps again and then push it to batch mode.

The current mode of the project is displayed in the Properties pane and the Details pane.

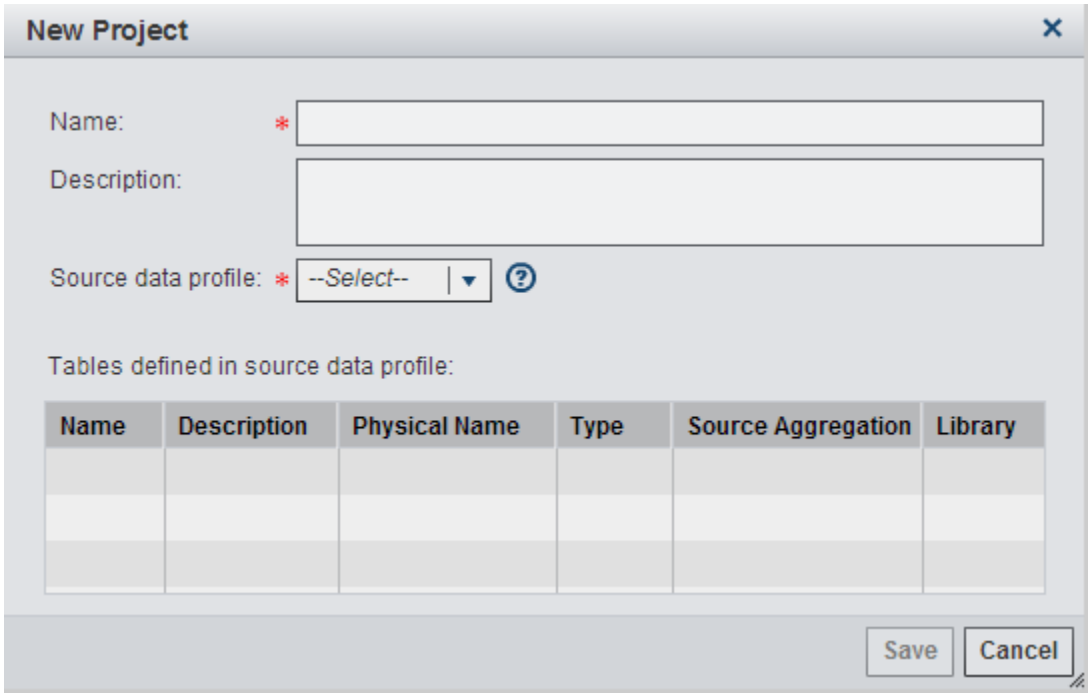
Create a Project

After you create a project, a default workflow is attached to it. You can view this workflow when you open the project.

To create a project:

1. Select the Projects workspace.
2. On the toolbar, click  . The New Project window appears.

Display 5.1 *New Project Window*



3 Enter a suitable name for the project. The project is identified with this name in the SAS Customer Link Analytics interface.

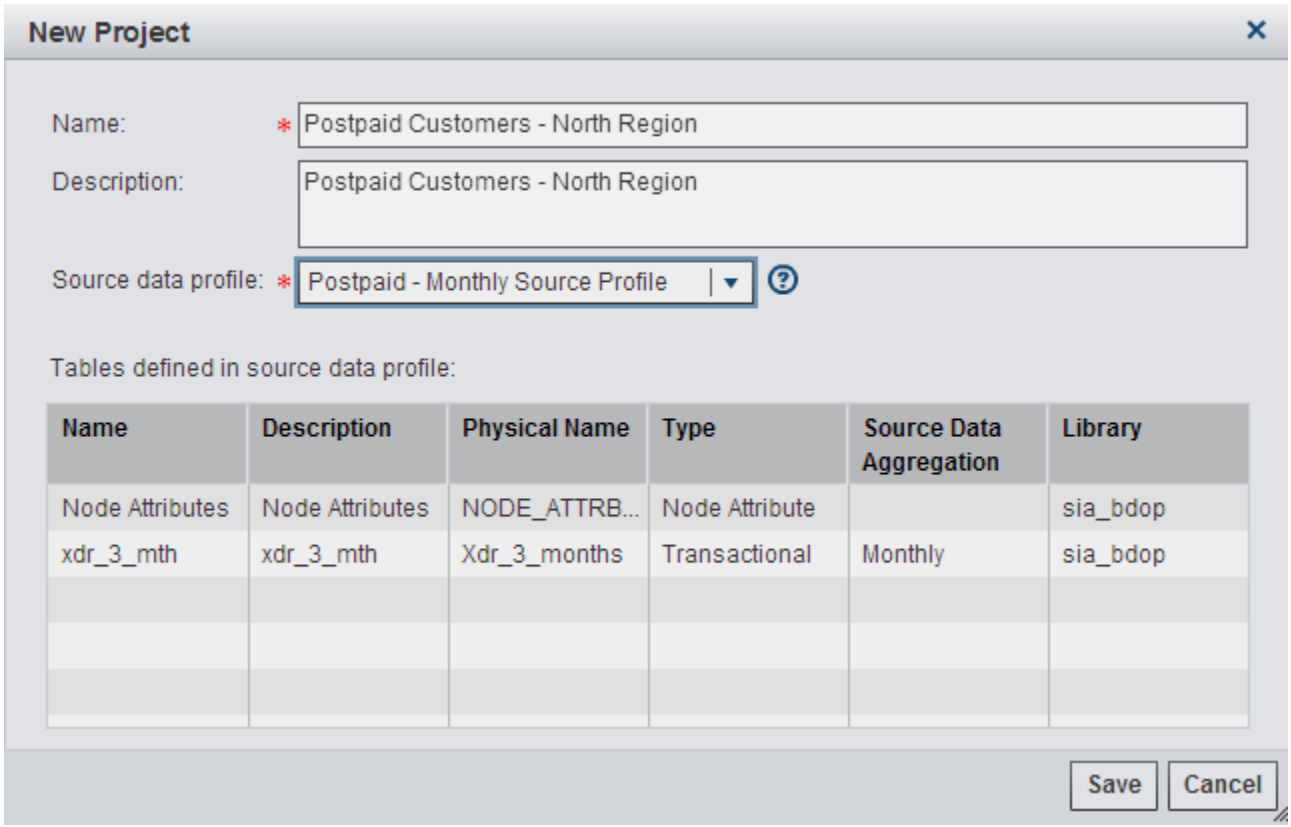
4 (Optional) Enter a description of the project. In this field, you can include the purpose for which you have defined the project.

5 Select the source data profile. The list displays the source data profiles that your administrator has defined. For details, see "Create a Source Data Profile" on page 97.

Note: Make sure that you select the correct source data profile. You cannot change the source data profile after you save the project. If you need to associate another source data profile, you have to delete the project and define a new one.

After you select the source data profile, the tables that are defined in this data profile are displayed. These tables store the transactional data and the information about the link and node attributes. The workflow steps use this information to analyze data that is required to build communities and define roles.

Display 5.2 *Source Data Profile Selection*



6 Click **Save**. The project that you have created is added to the All Projects. Also, a unique sequence number is generated for the project. This ID is displayed in the Properties pane. You can use this ID to uniquely identify the project.

View or Edit Properties of a Project

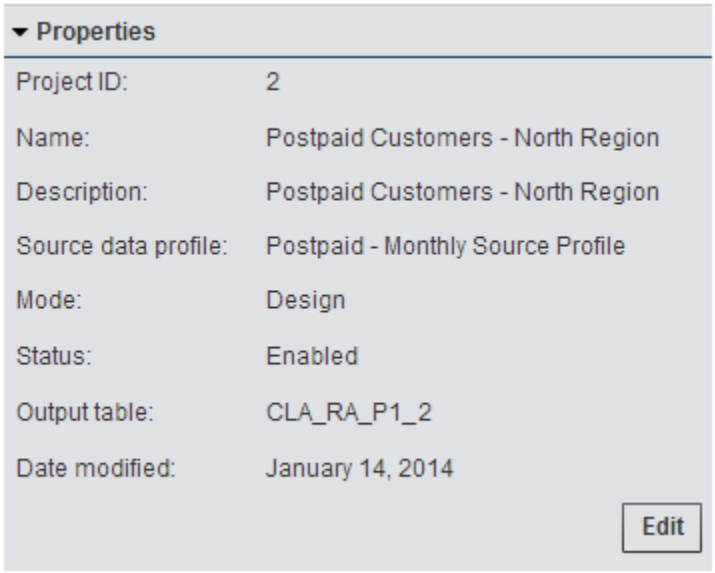
You can view the details about a project in the Properties pane. In addition, you can change the name, the description, and the processing mode of a project.

To view or edit the properties of a project:

1 In the Projects workspace, select the project whose properties you want to change.

2 In the Properties pane, you can view the following details:

Display 5.3 Properties Pane



Project ID
displays the sequence number that is generated for the project. This ID helps you uniquely identify a project.

Name
displays the name that you have specified for the project. You can change the name of the project if the project is in design mode. For details, see "Modes of a Project" on page 26.

Description
displays a short description of the project. You can change the description of the project.

Source data profile
displays the source data profile that you have selected for the project. For details, see "Overview of Source Data Profiles" on page 97. You cannot change the source data profile.

Mode
indicates the current mode of the project. A project can be in either design or batch mode. For details, see "Modes of a Project" on page 26.

Status
indicates the current status of the project. A project can have any one of the following statuses:

Table 5.1 Project Status

Status	Description
Enabled	Indicates that a project has been configured to have one or more active workflow steps.
Modified	Indicates that a workflow step of a project is being configured.
In progress	Indicates that one of the workflow steps is running.

Status	Description
Executed successfully	Indicates that all of the workflow steps of a project workflow have run successfully.
Executed with errors	Indicates that a workflow step of a project has not run successfully.

Output table
indicates the name of the output table that SAS Customer Link Analytics produces when the project completes one run in design mode. For details, see “Output of Workflow Steps” on page 77.

Date modified
indicates the last date on which the project’s properties were changed.

3 (Optional) Click **Edit**.

4 Click **OK** to save the changes that you have made.

Delete a Project

You can delete a project if you no longer need it in your application. You can delete a project only if it is in design mode.

To delete a project:

1 In the Projects workspace, select the project that you want to delete.

2 On the toolbar, select  .

3 In the confirmation dialog box that appears, click **Yes**.

Change the Mode of a Project

You can either push a project from design mode to batch mode or pull it into design mode from batch mode. In either case, you have to change the current mode of the project. For details, see Chapter 12, “Batch Processing,” on page 71.

To change the mode of a project:

1 In the Projects workspace, select the project whose mode you want to change.

2 In the Properties pane, click **Edit**.

3 Select the mode of the project. If the project is in design mode, select **Batch**. However, you can do so only if all the workflow steps of the project are complete. That is, the status of the project is **Executed Successfully**. After

you push the project to batch mode, all its workflow steps are deactivated. If you want to configure any workflow step, you have to pull the project back into design mode. To do so, select **Design**.

4 Click **OK**.

6

Managing Workflow Steps

About Project Workflows

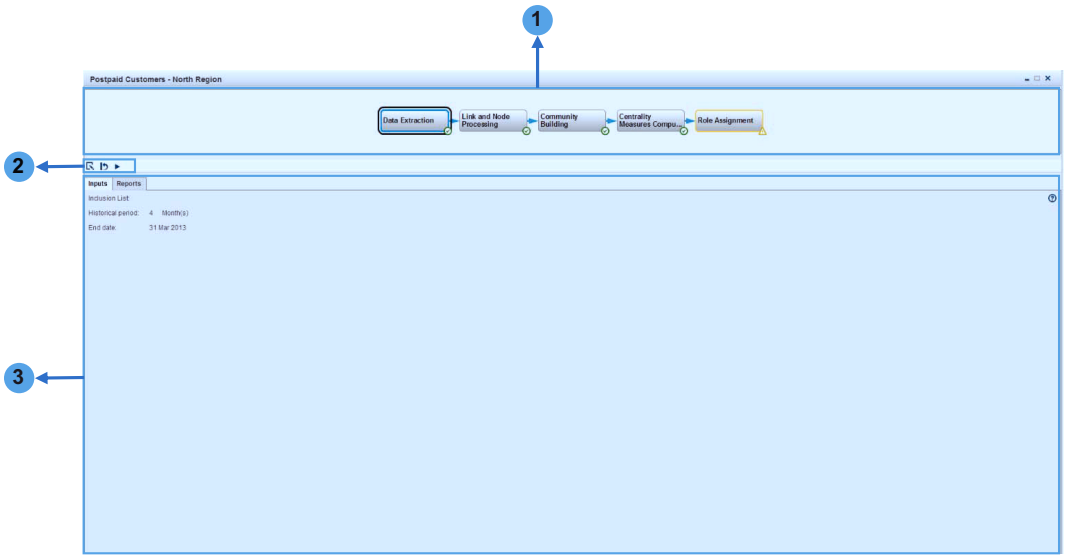
In SAS Customer Link Analytics, each project has a workflow that enables you to complete your tasks in a structured manner. The project workflow provides you with a guided approach for performing tasks that are associated with your project. When you open a project, its default workflow is displayed in the Workflow Diagram pane. The workflow diagram indicates the workflow steps of the project. The details of the current workflow step are displayed in the Workflow Step pane.

Working with Workflow Steps

Overview

When you open a project, you can view the following components:

Figure 6.1 Components of a Project



1 Workflow Diagram pane

2 Toolbar

3 Workflow Step pane

Workflow Diagram pane
: displays the diagrammatic representation of the workflow steps. It also provides the sequence of workflow steps and displays the processing status of each workflow step.

toolbar
: enables you to perform tasks for each workflow step.

Workflow Step pane
: displays information on the selected workflow step on two tabs, **Inputs** and **Results**. On the **Inputs** tab, you can specify the configuration details for the workflow step. On the **Results** tab, you can view the output that the workflow step produces.

Workflow Diagram Pane

The Workflow Diagram pane shows the workflow steps of a project. Each workflow step indicates an individual stage of the workflow. Here is the list of project workflow steps and the objective of each step:

1 **Data Extraction**

Extract transactional data from the pre-configured external source system and aggregate this data to produce links and nodes for the project based on the parameter values that are specified. In this workflow step, data pertaining to specific links and nodes can be made available. For details, see “Configure the Data Extraction Workflow Step” on page 40.

2 **Link and Node Processing**

Filter the links and nodes further based on the specified conditions and assign link weights. In this workflow step, the outlier links and nodes are

removed. For details, see “Configure the Link and Node Processing Workflow Step” on page 44.

3 **Community Building**

Specify parameter values as an input to build communities. You can choose the approach by which you want to build the communities. For details, see “Configure the Community Building Workflow Step” on page 51.

4 **Centrality Measures Computation**

Specify parameter values as an input to compute centrality measures. The centrality measures help in identifying the relative importance of a link or node in a network or community. For details, see “Configure the Centrality Measures Computation Workflow Step” on page 58.

5 **Role Assignment**

Assign roles based on the centrality measures statistics. You can define roles according to your business requirements. For details, see “Define a Role” on page 64.

Processing Status

Each workflow step is associated with a processing status. A workflow step is automatically enabled when the previous workflow step runs successfully. An enabled workflow step can have any one of the following processing statuses:

TIP The current workflow step is highlighted with a bounding box.

Table 6.1 Processing Status of Workflow Steps

Icon	Status	Description
	Successfully executed	The workflow step was successfully run without any errors.
	Executed with errors	One or more errors occurred when this workflow step was run. You have to resolve the errors and run the workflow step again.
	In progress	The workflow step is currently running.
	Modified	The configuration setup of the workflow step is modified.
Not available	Enabled	The workflow step is configured, and you can run the workflow step.

Toolbar

You can use the toolbar to perform tasks on a workflow step of a project that is in design mode. If the project is in batch mode, the options on this toolbar are deactivated.

Table 6.2 Toolbar Options

Button	Purpose
	Customizes the default setup of the workflow step.
	Reverts the changes that were made to the setup of the workflow step and resets them to the default.
	Runs the workflow step.

Workflow Step Pane

The Workflow Step pane displays information about the selected workflow step on two tabs.

Inputs tab
: This tab displays the default configuration that is set up for the workflow step. You can specify the configuration details for the workflow step.

Results tab
: On this tab, you can view the results that the workflow step produces after you run it.

About Common Tasks for Workflow Steps

After you define a project, the default workflow with predefined workflow steps is displayed in the Workflow Diagram pane. You must perform workflow steps in the sequence that is shown in the diagram. For each workflow step, you can perform certain common tasks. The tasks that you can perform differ depending on the workflow step that you are currently working on. Using the toolbar, you can perform the following common tasks:

- Configure a workflow step.
- Run a workflow step.
- View the results of a workflow step.
- Reset a workflow step.

Note:

- You can perform any of these tasks on a workflow step only if the project is in design mode. In addition, you have to be the owner of the project or have the

capability to work on any project. For details, see “Roles and Capabilities” on page 11.

- The toolbar is deactivated if any of the workflow steps is running in design mode or if the project is in batch mode.

Configure or Edit a Workflow Step

SAS Customer Link Analytics provides a default configuration setup for each workflow step. If you do not want to customize this setup, you can run the workflow step directly. You can also change the setup according to your business requirements. You can configure and run a workflow step until you are satisfied with the results of the workflow step.

At any time, you can also change the current configuration of a workflow step. In this case, you are editing a workflow step. When you are editing a workflow step, you cannot access any other workflow step. After you edit a workflow step and run it, all the subsequent workflow steps that have successfully run are deactivated. However, their configuration values are retained. You have to run these workflow steps again.

Note: You can configure a workflow step only if the project is in design mode.

To configure a workflow step:

1. In the Workflow Diagram pane, select the workflow step that you want to configure. The current configuration details are displayed.
2. Select the **Inputs** tab.
3. On the toolbar, select  .
4. Change the details as needed.
5. Click **OK**.
6. On the toolbar, select ▶.

Run a Workflow Step

After you configure a workflow step, you can run the workflow step using the toolbar.

Note: It is not mandatory that you run the workflow step as soon as you configure it. However, you will not be able to view the results that are generated for the workflow step until you run it. Also, you cannot proceed with the next workflow step until you have successfully run the current the workflow step.

To run a workflow step:

1. In the Workflow Diagram pane, select the workflow step that you want to run.
2. On the toolbar, select ▶.

View the Results of a Workflow Step

When you run a workflow step, certain results are generated. These results are displayed on the **Results** tab.

To view the results of a workflow step:

1. In the Workflow Diagram pane, select the workflow step.
2. Select the **Results** tab. This tab displays the results that the workflow step produces after you run it.

Reset a Workflow Step

In order to discard the changes that you have made to the default configuration for a workflow step, you can reset the workflow step. However, when you reset a workflow step, the subsequent workflow steps that you have configured and run are deactivated. If you have changed the default values for these workflow steps, you will lose them. Therefore, you have to configure these workflow steps again and then run them.

To reset a workflow step:

1. In the Workflow Diagram pane, select the workflow step.
2. On the toolbar, select  . The default configuration is displayed for the workflow step and the subsequent steps are deactivated.

7

Data Extraction

About the Data Extraction Workflow Step

Overview of the Data Extraction Workflow Step

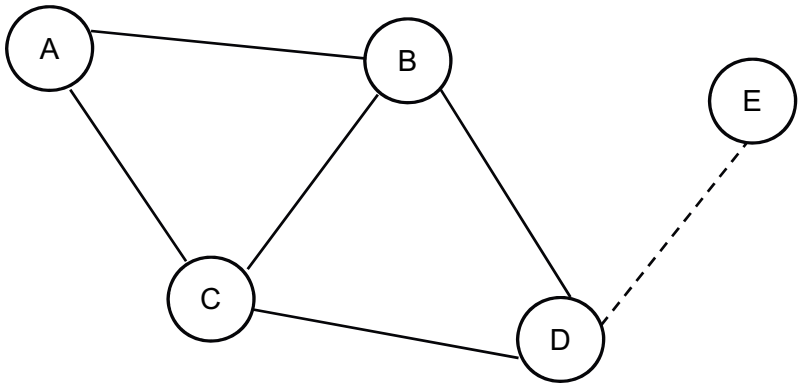
The Data Extraction workflow step enables you to define the links and nodes that you want to consider for building communities within your network. In addition, you can specify the historical period for which you want to consider the transactional data. When you run this workflow step, the transactional data associated with these links and nodes is extracted from the preconfigured external source system for the specified time period.

Your administrator defines one or more lists of links or nodes that you can consider when building communities. For details, see “Attribute Tables” on page 87. In the Data Extraction workflow step, you can choose any one of these lists. These lists are called *inclusion lists*. An inclusion list can either be a list of nodes (also called a *node inclusion list*) or a list of links (also called a *link inclusion list*). Based on the inclusion list that you choose, the associated links and nodes of the network are considered in the subsequent workflow steps.

Example 1: Node Inclusion List

You choose an inclusion list that contains three nodes: A, B, and C. In this case, the Data Extraction workflow step extracts the transactional data of links that involve nodes A, B, and C. That is, transactional data of interactions between the pairs of nodes A and B, B and C, A and C, C and D, and B and D is extracted.

Figure 7.1 *Selected Nodes in a Network*

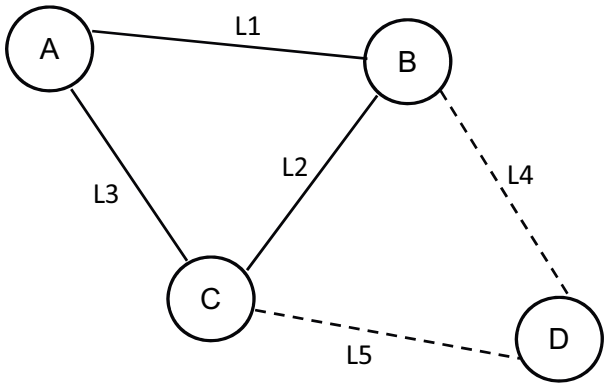


As depicted in the network diagram, transactional data of links that involve nodes A, B, and C is extracted. The inclusion list restricts the link that exists between D and E from being considered for extracting data.

Example 2: Link Inclusion List

You choose an inclusion list that contains links L1, L2, and L3. In this case, the workflow step extracts the transactional data of interactions between A and B, B and C, and A and C. The inclusion list restricts the transactional data of interactions between B and D and C and D from being considered for extracting data.

Figure 7.2 *Selected Links in a Network*



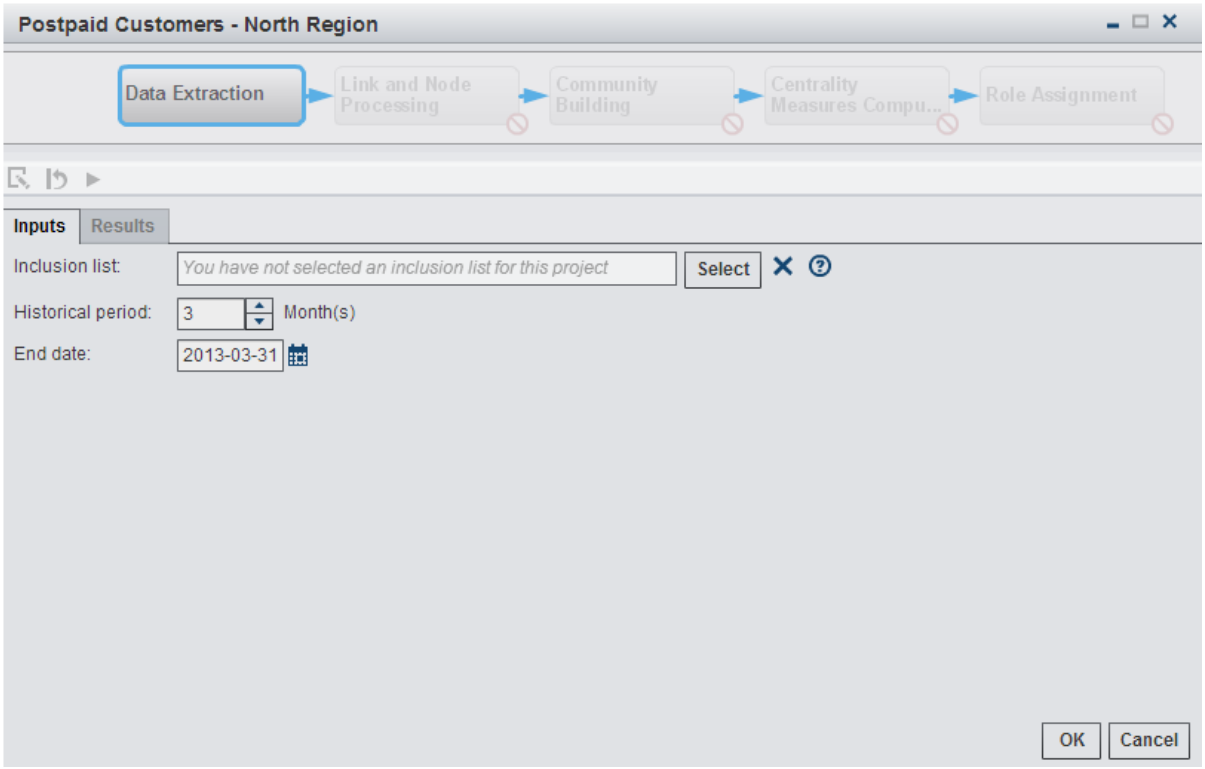
Configure the Data Extraction Workflow Step

To configure the Data Extraction workflow step:

1 In the Workflow Diagram pane, select **Data Extraction**.

2 On the toolbar, select .

Display 7.1 Data Extraction Workflow Step



3 (Optional) Click **Select** to add an inclusion list. The Inclusion List window appears.

   a Select the inclusion list that you want to attach to the project. The list displays the attribute tables that your administrator has included while defining the source data profile that is associated with the project. For details, see "Create a Source Data Profile" on page 97. Make sure that you select the correct type of inclusion list. An inclusion list that contains links or nodes is of the **Link Inclusion List** or **Node Inclusion List** type, respectively. If you do not select an inclusion list, all the links and nodes that are available in the network are considered.

      Note:

      - You can attach only one inclusion list for each project.
      - You can attach or remove an inclusion list anytime later. However, if you have configured and run this workflow step or subsequent workflow steps, you have to run these workflow steps again.

   b Click **OK**.

4 Specify the historical period if you want to use a different historical period than the default period that is displayed. The unit of the historical period is displayed depending on the data aggregation level of the transactional table of the source data profile that you have selected for the project. Therefore, this field is deactivated if the transactional table is fully aggregated. However, if the data aggregation level of the transactional table is Monthly or Daily, you have to specify the historical period in months or days, respectively.

   When you run this workflow step, the transactional data of the relevant links and nodes is extracted for the specified historical period. A default historical

period is preconfigured. If you run this workflow step without configuring it, then the transactional data is extracted for the default historical period.

5 A default date is displayed in the **End date** field. This date is populated when your administrator refreshes the transactional table that is associated with this project. For details, see "Refresh Data in a Transactional Table" on page 92. Specify the end date of the historical period if you want to use a different date than the default date that is displayed. Only certain dates are available for selection depending on the data aggregation level of the transactional table. For example, for a monthly aggregated transactional table, you specify the historical period as 3 months and the end date as 31-Mar-2013. In this case, data is extracted for the period from 1-Jan-2013 to 31-Mar-2013. Because it is a monthly aggregated table, the transactional table would contain 3 records for each unique combination of the From ID and To ID columns.

   Note: This field is deactivated if the transactional table is fully aggregated.

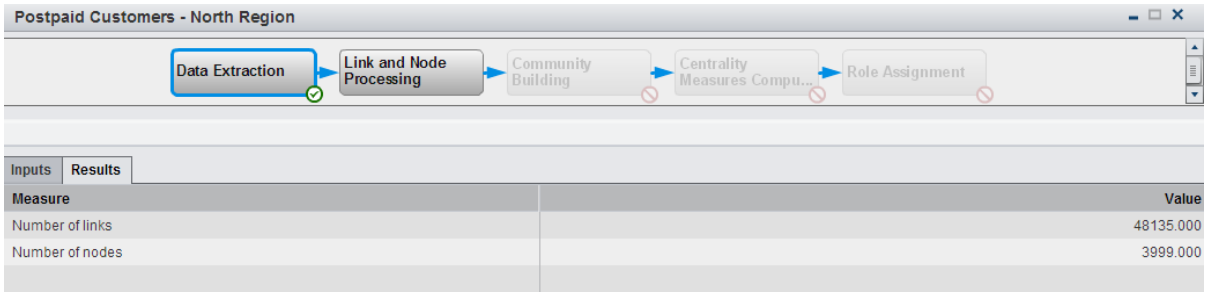
6 Click **OK**.

7 On the toolbar, click ▶. The required transactional data is extracted and the results are produced.

View Results of the Data Extraction Workflow Step

When you run the Data Extraction workflow step, the results are published on the **Results** tab. On this tab, the number of links and nodes for which transactional data is extracted is displayed. This data is used in the subsequent workflow steps.

Display 7.2 *Results of Data Extraction Workflow Step*



8

Link and Node Processing

About the Link and Node Processing Workflow Step

Overview

The Link and Node Processing workflow step contains two sub-steps.

Link node filtering
: In this sub-step, you filter the links and nodes that you have selected in the Data Extraction workflow step. This task helps you remove the links and nodes that are insignificant for network analysis and community building. For example, while analyzing the communications network, you would want to remove links that are associated with call centers or any other promotional offers. As a result, link filtering helps you focus your marketing activities on target customers.

  In order to filter the links and nodes, you have to specify limit values for certain measures. For details, see “Measures for Filtering Links and Nodes” on page 44.

Link weight assignment
: In this sub-step, you assign weights to the links that you have filtered. Subsequently, this task provides you with a basis to remove weaker links in the network.

When you run the Link and Node Processing workflow step, links and nodes of the entire network that satisfy the limit values that you have specified for all measures are selected. Also, a weight is assigned for each link that is selected.

Measures for Filtering Links and Nodes

You can filter links and nodes based on the limit values that you enter for certain measures. SAS Customer Link Analytics provides the in-degree and out-degree measures. However, the rest of the measures are extracted from the transactional data. You configure these measures in the Administration workspace. For details, see “Configure the Column Values of a Table” on page 90.

For example, if you are analyzing call detail records of subscriptions of the communications industry, you can use the following measures to filter links and nodes. Except for the in-degree and out-degree measures, the rest of the measures are defined according to the transactional data.

In-degree
: indicates the number of incoming links that are incident to a certain node. For example, it can be the number of inbound calls of a certain customer. This measure helps you filter nodes.

Out-degree
: indicates the number of outgoing links for a certain node. For example, it can be the number of outbound calls of a certain customer. This measure helps you filter nodes.

Number of calls
: indicates the total number of incoming and outgoing links for a node. This measure helps you filter nodes.

Number of short message services (SMS)
: indicates the total number of SMS that are received by or sent from this node. This measure helps you filter nodes.

Call duration
: indicates the duration of a call between two nodes. This measure helps you filter links.

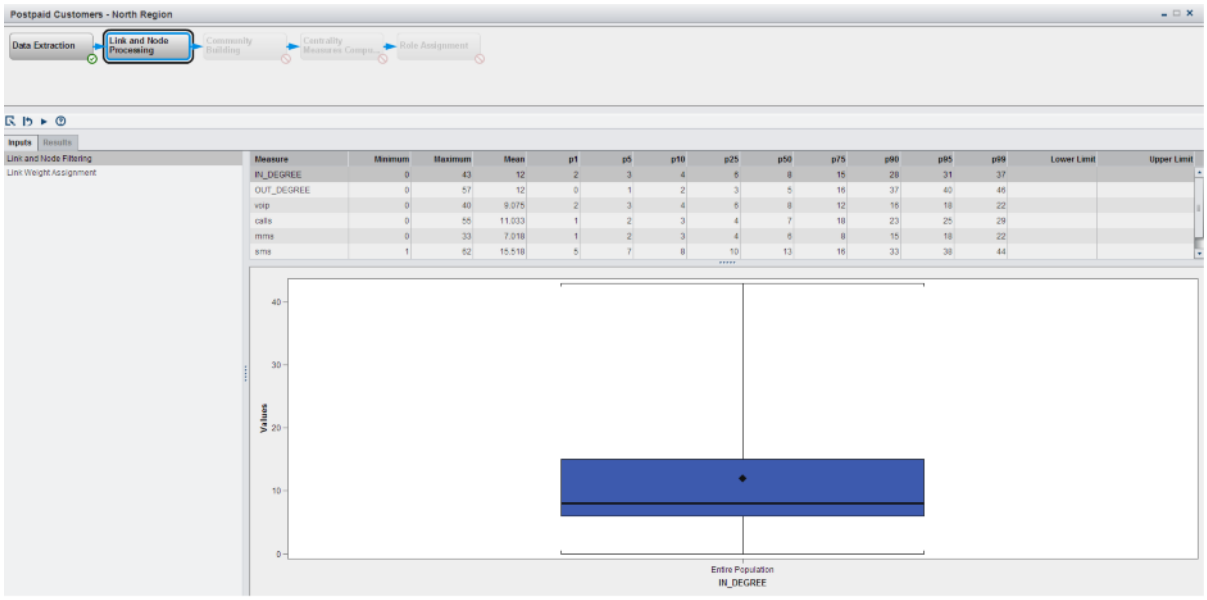
For each of these measures, you can view the minimum, maximum, average, and percentile values. Based on these values, you can specify the limit values for the respective measures.

Configure the Link and Node Processing Workflow Step

To configure the Link and Node Processing workflow step:

1. In the Workflow Diagram pane, select **Link and Node Processing**.
2. On the toolbar, select .
3. Select the **Link and Node Filtering** page. The measures that are predefined for filtering links and nodes are displayed. For details, see “Measures for Filtering Links and Nodes” on page 44.

Display 8.1 *Link and Node Filtering Page*



4 Select the measure that you want to consider for filtering the links and nodes.

5 View the statistics that are produced for the measure. You can also view this information graphically. For details, see “Interpreting the Graphical Statistics for Measures” on page 48.

6 Specify either one or both of the limit values for the measure. These values indicate the filter conditions that you want to specify for the measure. The links and nodes are filtered based on the limit values that you enter.

Example 1: Filter nodes with extreme in-degree values

A node that has extremely high in-degree values might be a service center or a restaurant. The presence of these nodes can impact the processing time and the results that are produced in the Community Building workflow step. Also, such nodes are never part of campaigns that target individual subscriptions (nodes). Therefore, these nodes should be filtered from the analysis. Assume that for the in-degree measure, the minimum and maximum values are 1 and 700, respectively. In this case, you can specify the **Upper Limit** as 50 in order to filter the insignificant nodes.

Example 2: Filter nodes with extreme out-degree values (also called star nodes)

A node that has extremely high out-degree values might be a call center. The presence of these nodes can impact the processing time and the results that are produced in the Community Building workflow step. Also, such nodes are never part of campaigns that target individual subscriptions (nodes). Therefore, these nodes should be filtered from the analysis. Assume that for the out-degree measure, the minimum and maximum values are 0 and 1,200, respectively. In this case, you can specify the **Upper Limit** as 100 in order to filter the insignificant nodes.

Example 3: Filter extreme values for call duration

Extremely high values for call duration can be outliers in the data. Assume that 90% (percentile 90 column) of the nodes have call duration values less than or equal to 400 minutes. However, the maximum value is 7,500. You

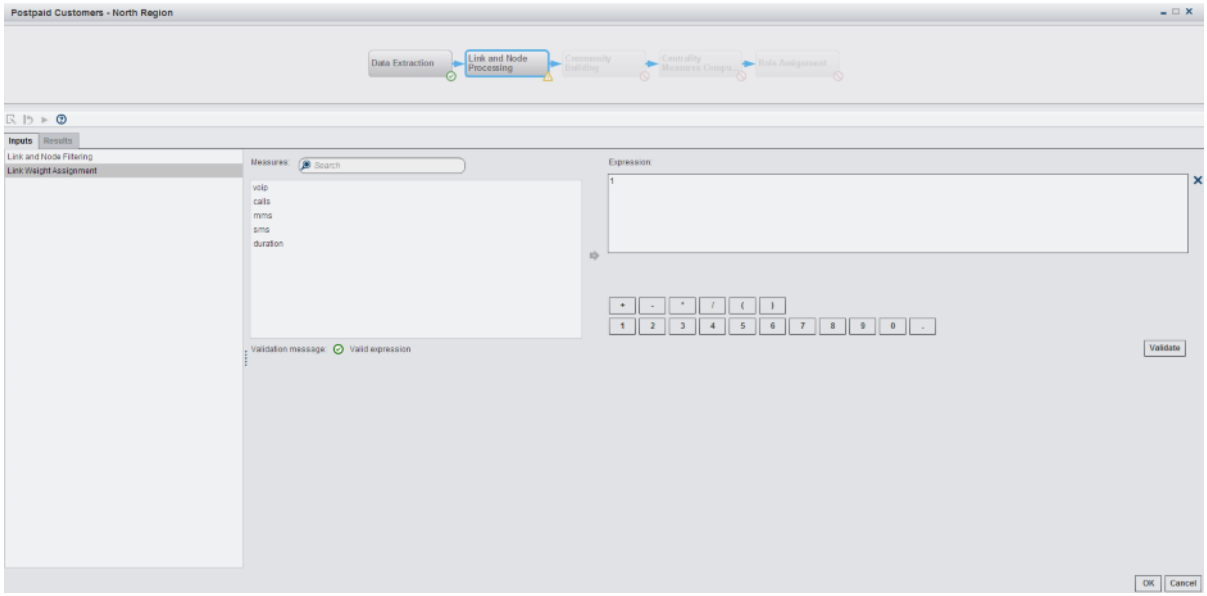
specify the **Upper Limit** as 2,000. All the subscriptions (nodes) that have call duration values greater than 2,000 minutes will not be considered for the analysis.

7 Repeat steps 4 through 6 for the other measures based on how you want to filter the nodes and links.

8 Click **OK**. Links and nodes that satisfy the limit values that you have specified for the selected measures are filtered and made available in subsequent steps.

9 Select the **Link Weight Assignment** page.

Display 8.2 *Link Weight Assignment Page*



10 On the toolbar, select . A default weight of value 1 is displayed in the **Expression** field.

Note: If you do not change the default value, 1 is assigned as the link weight. This indicates that all links are equally important. As a result, no links are removed.

TIP Click to clear the default expression.

11 Build an arithmetic expression or enter a static value that you want to assign as a link weight. You can build the expression by using appropriate combinations of the available measures, arithmetic operators, and numeric values.

You can either enter the expression or build it by selecting the appropriate measures and options of the expression builder.

Example 1: Expression for analyzing voice calls

If you want to analyze the transactional data of voice calls, you might want to use the call duration measure in your expression.

Example 2: Using weighted expression for analyzing SMS and call counts

If you want to analyze the counts of calls and messages, you might want to use the following expression for the link weights. Also, assume that call counts are more important for your analysis than message counts. In this case, you can assign a weight to the call count measure, as described in the following expression:

*sms + (4*calls)*

When you run this workflow step, the link weight is computed for each link of the network. You can remove weaker links from the network in the Community Building workflow step. For details, see "Configure the Community Building Workflow Step" on page 51.

Note:

- If the link weight expression for a link returns a zero, a negative, or a missing value, then the weight of that link is considered to be 1.
- If your expression contains a measure in the denominator, then make sure that the lower limit that you specify for this measure is a nonzero value.

12 Click **Validate** to confirm that the expression that you have entered is valid.

13 Click **OK**. The expression that you have specified is displayed.

14 On the toolbar, select ▶. The link weight is assigned to each link according to the expression that you have specified.

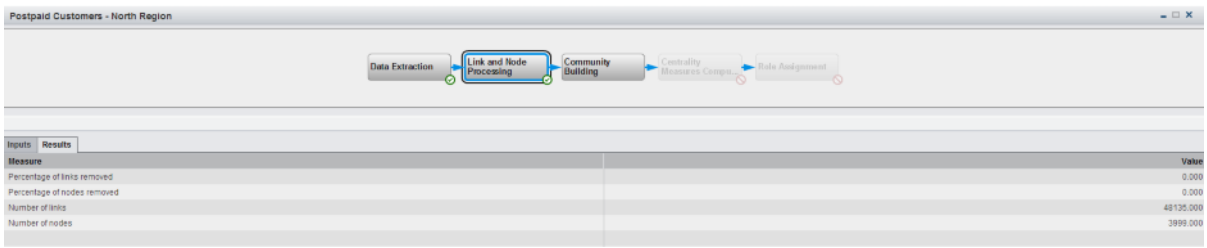
View Results of the Link and Node Processing Workflow Step

When you run the workflow step, the results appear on the **Results** tab. These results are produced based on the limit values that you specified in the Link and node filtering page. For details, see Step 6 on page 45.

The results include the following details:

- percentage of links that is removed from the network.
- percentage of nodes that is removed from the network.
- number of links of the network that are selected based on the limit values that you have entered.
- number of nodes of the network that are selected based on the limit values that you have entered.

Display 8.3 *Results of Link and Node Processing Workflow Step*



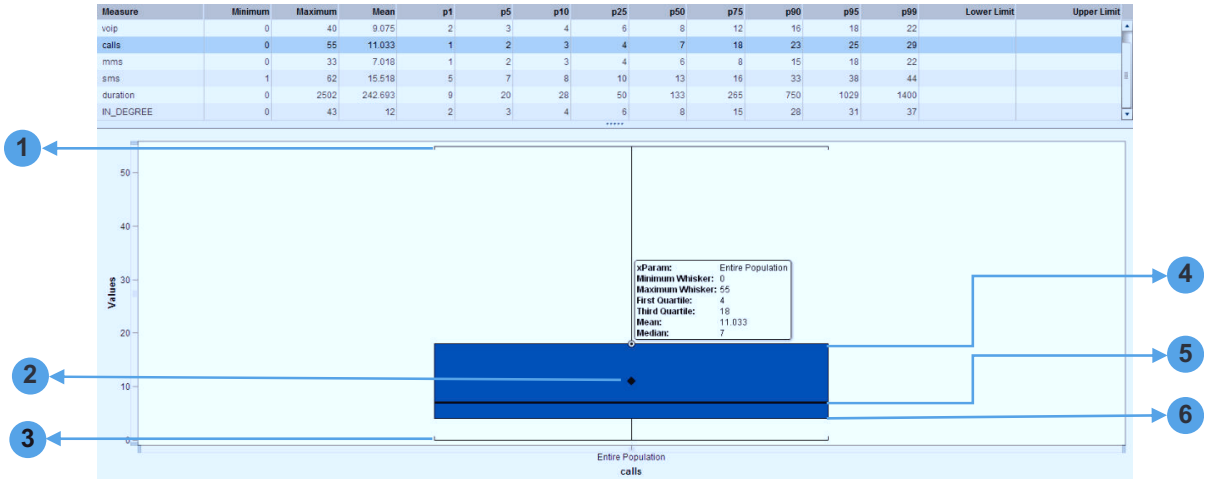
If the results that are produced do not meet your business requirements, you can reconfigure the limit values that you have specified. For example, if almost 80% of your links or nodes are removed, then you can change the limit values accordingly.

Interpreting the Graphical Statistics for Measures

The box plot graph helps you understand the distribution of measures within the network. Based on the information that you get from the graph, you can specify the limit values for the measures.

For example, consider that the following graph is displayed for the Calls measure.

Figure 8.1 *Box Plot Graph*



The graph gives the following statistical values:

1. The maximum whisker indicates the largest non-outlier value for the **calls** measure. That is, the maximum call count of a node is 55.
2. The mean value indicates the average number for the measure. That is, the average call count of a node is 11.033.
3. The minimum whisker indicates the smallest non-outlier value for the measure. That is, the minimum call count of a node is 0.
4. The third quartile indicates that 75% of the observations of the measure have a value less than this value. That is, the call count of 75% of nodes in the network is less than 18.
5. The median value indicates that 50% of the observations of the measure have a value less than this value. That is, the call count of 50% of nodes in the network is less than 7.
6. The first quartile value indicates that 25% of the observations of the measure have a value less than this value. That is, the call count of 25% of nodes in the network is less than 4.

In addition, the graph indicates that the length between the third quartile and the maximum whisker is significantly greater than the length between the first quartile and the minimum whisker. Therefore, the graph is more skewed at the top than at the bottom. This skewness indicates that observations (the call count of a node) are concentrated on the higher end of the scale.

9

Community Building

Overview of the Community Building Workflow Step

The Community Building workflow step enables you to build communities within your network. This workflow step partitions the network into communities so that the links within the communities are more densely connected than the links between communities. The Community Building workflow step implements two methods for finding communities:

Top-down approach
: In the method, the workflow step recursively splits larger communities into smaller units until certain conditions are satisfied.

Bottom-up approach
: In the method, the workflow step recursively merges smaller communities until certain conditions are satisfied.

You can choose either of these methods, depending on your business requirements.

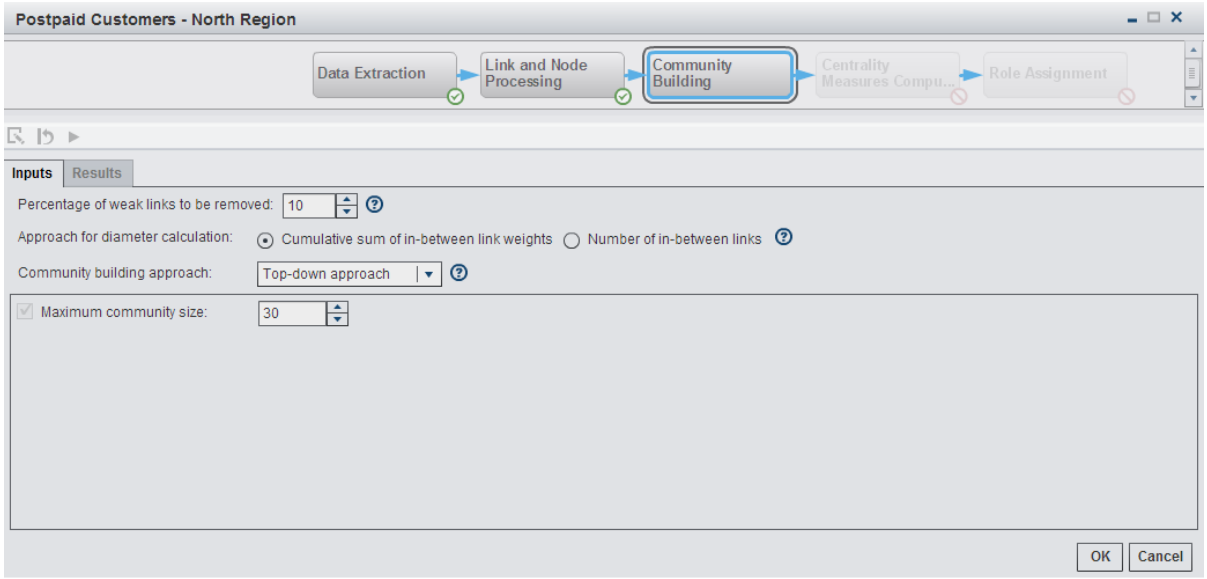
When you run the Community Building workflow step, communities are built within your network based on the approach that you chose and the parameter values that you specified.

Configure the Community Building Workflow Step

To configure the Community Building workflow step:

1. In the Workflow Diagram pane, select **Community Building**.
2. On the toolbar, select  .

Display 9.1 Community Building Workflow Step



3 Specify the percentage of weak links that you want to remove. The valid range is between 0 and 100. The default value is 10. This option enables you to considerably improve the run time for building communities. You can remove weak links so that removing them does not remove the underlying node. The strength of a link is determined on the basis of the link weight that you assign in the Link and Node Processing workflow step. For details, see Step 11 on page 46.

The Link and Node Processing workflow step filters links and nodes by applying the same filtering condition to the entire network. However, the weak link removal option removes relatively weak links in the neighborhood of the node.

4 Select the approach you want to use for calculating the diameter. For details, see “Diameter of a Community” on page 120.

The option that you choose here does not have any impact on how the communities are built when you run this workflow step. However, the diameter values that are displayed on the **Results** tab when you run this workflow step are computed based on your selection. For details, see “Interpreting the Results Summary” on page 55.

The following options are available:

Cumulative sum of in-between link weights
Select this option if you want to compute the diameter based on the cumulative sum of the weight of links (also called *edges*).

Number of in-between links
Select this option if you want to compute the diameter based on the total number of links.

5 Choose the community-building approach. The following options are available:

Top-down approach
Select the top-down approach if you have a fairly good idea about the maximum community size that your business requires. By using this approach, you can restrict the community size according to your needs.

In this approach, the workflow step recursively breaks down larger communities into smaller ones until the conditions that you have specified are satisfied.

Bottom-up approach

Select the bottom-up approach if you want the community-building algorithm to decide the optimal community size.

In this approach, the workflow step recursively merges communities at various resolution values. The resolution value can be interpreted as the minimal density of communities in an undirected and unweighted graph. The *density* of a community is defined as the number of links inside the community divided by the total number of possible links. A larger resolution value produces more communities, each of which contains a smaller number of nodes.

Note: Perform step 6 or step 7 depending on whether you chose the **Top-down approach** or the **Bottom-up approach**.

6 Specify the following parameters for the top-down approach:

Note: Parameters are displayed depending on whether you are working in distributed or non-distributed mode. For distributed mode, only the **Maximum community size** parameter is displayed. However, for non-distributed mode, both **Maximum community size** and **Maximum diameter** parameters are displayed.

a The **Maximum community size** check box is selected by default. Specify the maximum number of nodes that can be contained in any community. The community–building algorithm runs iteratively until each community contains on average the maximum number of nodes that you have specified.

Note: In certain cases, a community cannot be split further even if it does not meet the threshold value that you specified for this parameter. For example, a star-shaped community can contain 200 nodes, while the maximum community size that you might have specified is 100.

b Select **Maximum diameter** and specify the maximum number of links on the shortest paths between any pair of nodes in any community. The community–building algorithm runs iteratively until the number of links on the shortest path in each community is less than or equal to the value that you have specified.

Note: In certain cases, a community cannot be split further even if it does not meet the threshold value that you specified for this parameter. For example, a symmetric community can have a diameter of 2, while the maximum diameter that you might have specified is 1.

c Decide whether you want the algorithm to recursively split the communities until both the conditions are satisfied. Select the appropriate option accordingly.

7 Specify parameters for the bottom-up approach:

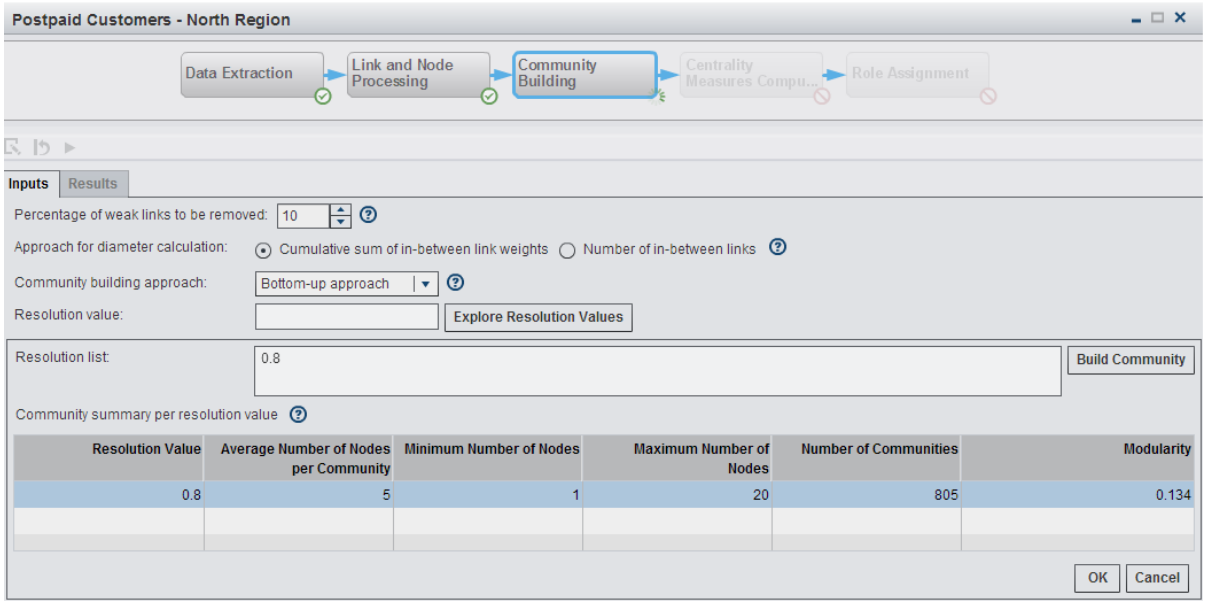
a Click **Explore Resolution Values**.

b Enter one or more resolution values depending on whether you are working in distributed or non-distributed mode. For distributed mode, you can enter only one value. However, for non-distributed mode, you can enter at most 20 values, each value separated by a space.

The higher the resolution value, the greater the number of communities that are built. However, two different runs might not produce the same result even if the two runs share a common resolution level. For example, the algorithm can produce different results at resolution value 0.5 in two runs: one with resolution list as 1 0.8 0.5 and the other with resolution list as 1 0.5.

c Click **Build Community**. The community building algorithm runs iteratively for each resolution value, builds the communities, and publishes the results for each resolution value.

Display 9.2 *Community Building: Bottom-up Approach*



Average number of nodes per community
Specifies the average number of nodes in a community.

Maximum number of nodes
Specifies the maximum number of nodes in a community.

Minimum number of nodes
Specifies the minimum number of nodes in a community.

Number of communities
Specifies the number of communities that are built.

Modularity
Specifies the quality of a division of a network into communities. It also indicates the stopping criteria for the iterative process of the community-building approach and identifies the optimal community size. The higher the modularity, the better the results of the community-building process are.

d Based on the results, choose the resolution value at which you want to build the communities.

e Click **OK**. The resolution value that you select from the list is displayed in the **Resolution value** field.

8 Click **OK**.

9 On the toolbar, select ▶ and view the results that are produced.

View Results of the Community Building Workflow Step

Interpreting the Results Summary

After you run the Community Building workflow step, communities are built based on the approach that you chose and the parameters that you specified. On the **Results** tab, you can view the following information for all the communities that are built in the network:

Note: This information is displayed only if you are working in non-distributed mode.

Table 9.1 *Summary of Results for Non-distributed Mode*

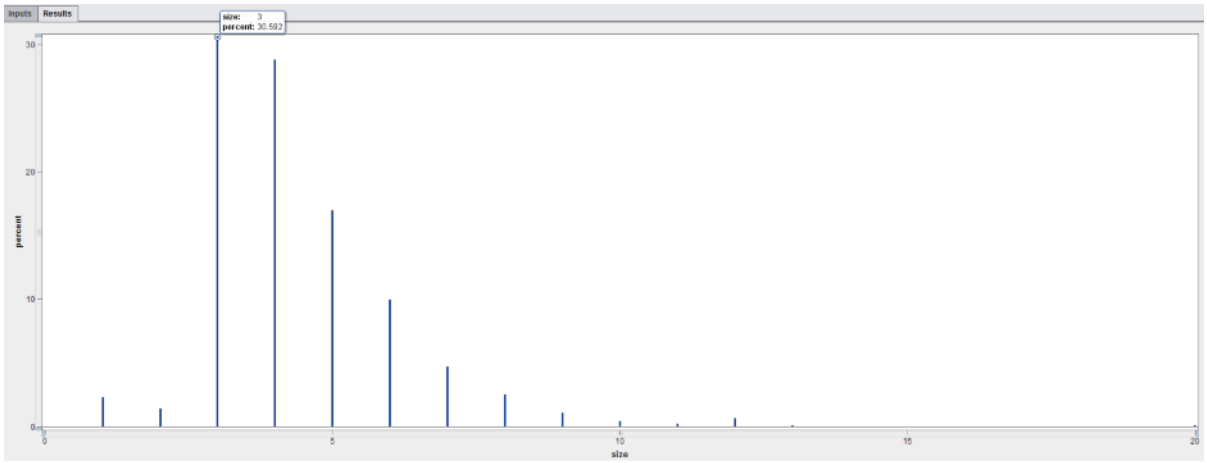
Measure Name	Description
Average density	Indicates the density of a community, which is computed as the number of links in a community divided by the maximum possible number of links in the community. Therefore, the average density is the average of all community densities.
Average diameter	Indicates the average diameter value. The diameter is computed according to the approach that you select when you configure the Community Building workflow step. For details, see Step 4 on page 52.
Average number of nodes	Indicates the average number of nodes in a community.
Maximum density	Indicates the maximum density value.
Maximum diameter	Indicates the maximum diameter value. The diameter is computed according to the approach that you select when you configure the Community Building workflow step. For details, see Step 4 on page 52.
Maximum number of nodes	Indicates the maximum community size.
Modularity of the network	Indicates the quality of a division of a network into communities.
Number of communities in the network	Indicates the number of communities that are built in the network.

Interpreting the Graph Results

The graph indicates the percentage of communities that have a particular size. The X-axis indicates the community size and the Y-axis indicates the percentage value. This graph helps you verify whether the communities that are built serve your business objective.

For example, in the graph here, 30.59% of communities have a size of 3 nodes. Similarly, 28.83% of communities have a size of 4 nodes.

Display 9.3 *Community Building: Graphical Results*



10

Centrality Measures Computation

Overview of Centrality Measures Computation Workflow Step

The Centrality Measures Computation workflow step enables you to provide input for the parameters thatSAS Customer Link Analytics require computing centrality measures. The centrality measures determine the relative significance of a node in a network or a community. In addition, you can use some of these centrality measures when you define roles for a community.

SAS Customer Link Analytics uses the following centrality measures:

- Influence
- Degree
- Closeness
- Betweenness
- Eigenvector
- Authority
- Hub
- Clustering coefficient

For details about these centrality measures, see Appendix 2, “Overview of Centrality Measures,” on page 123.

When you run this workflow step, the centrality measures are computed and their graphical results are displayed. These results can help you understand how significant the nodes are within the network or community.

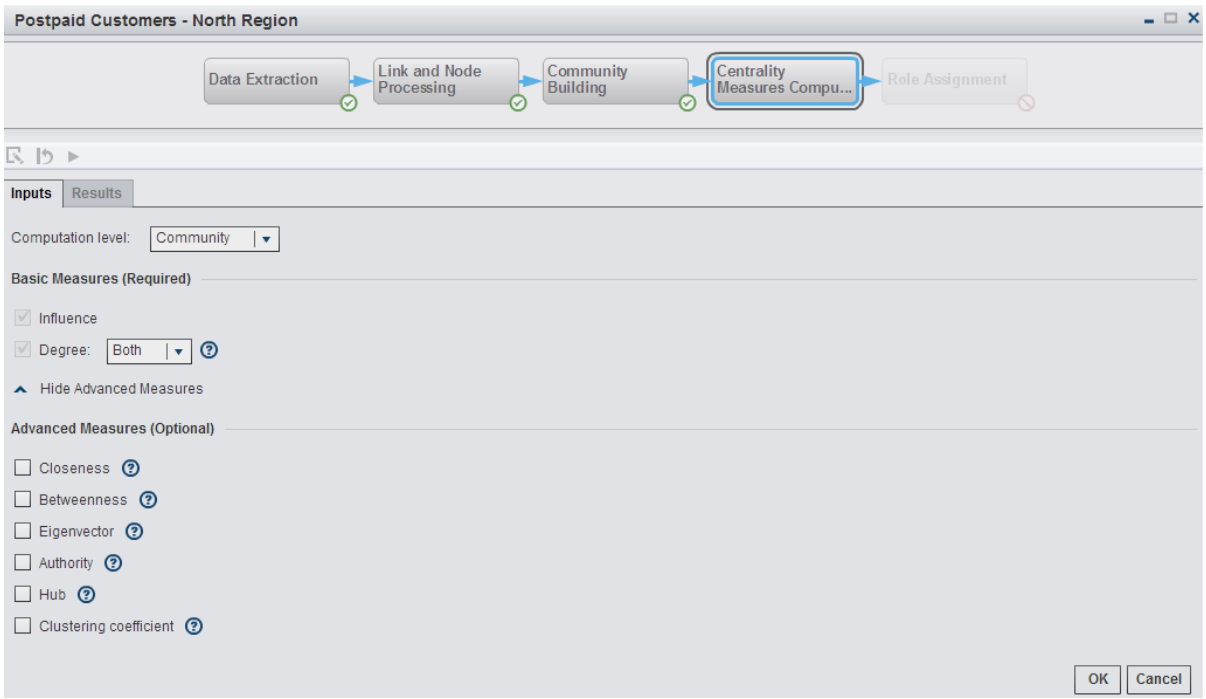
Configure the Centrality Measures Computation Workflow Step

To configure parameters for computing centrality measures:

1 In the Workflow Diagram pane, click **Centrality Measures Computation**.

2 On the toolbar, select .

Display 10.1 *Centrality Measures Workflow Step*



3 Select the level at which you want to compute centrality measures.

The following options are available:

Note: For distributed mode, only the **Community** option is available. However, for non-distributed mode, you can select the level as either **Community** or **Network**.

Network

Select this option if you want to compute centrality measures for a node by considering its relative significance with all other nodes of the entire network.

Note: This option does not have any impact on the Closeness and Betweenness centralities. Therefore, even if you choose this option, the Closeness and Betweenness centralities are computed at the community level.

Community

Select this option if you want to compute centrality measures for a node by restricting its relative significance with other nodes within its

community. In addition, this option helps you produce results faster compared to the **Network** option.

4 Specify parameter values for the basic centrality measures:

Influence
This check box is selected by default. The Influence centrality computes the relative influence of a node on its immediate neighbors. A higher influence value of a node indicates that the node has a capability to influence its immediate neighbors.

Note: SAS Customer Link Analytics normalizes the value that is computed for the Influence centrality. Therefore, the influence values are always greater than or equal to 0 and less than or equal to 1.

Degree
The Degree centrality indicates the number of immediate neighbors of a node.

Select any one of the following options to indicate how you want to compute the degree centrality:

In
calculates the degree based on incoming links that are incident to a node.

Out
calculates the degree based on outgoing links that are emerging from a node.

Both
calculates the degree based on both incoming and outgoing links.

5 Click ⌄ to display the advanced options.

Closeness
This centrality indicates the proximity of the node to the entire network. A higher closeness value indicates that the node is closer to the rest of the nodes. A lower value indicates that the node is away from the rest of the nodes.

Select the **Closeness** check box if you want to compute this measure.

Note: SAS Customer Link Analytics normalizes the value that is computed for the Closeness centrality. Therefore, the closeness values are always greater than or equal to 0 and less than or equal to 1.

Betweenness
This centrality indicates the connecting potential of the node. A higher Betweenness value indicates that the node is pivotal in connecting multiple other nodes in the network.

Select the **Betweenness** check box if you want to compute this centrality.

Eigenvector
This centrality is an extension of the degree centrality, in which centrality points are awarded for each neighbor. However, not all neighbors are equally important. Intuitively, a connection to an important node should contribute more to the centrality score than a connection to a less important node. This is the basic idea behind Eigenvector centrality. Eigenvector centrality of a node is defined to be proportional to the sum of the scores of all nodes that are connected to it.

Select the **Eigenvector** check box if you want to compute this centrality.

Authority and **Hub**

Certain nodes tend to distribute and disseminate the information that they receive. These nodes are called hubs. Similarly, certain nodes receive more information. These nodes are called authorities. Both hubs and authorities influence the flow of information within the network. A good hub node is one that points to many good authorities, and a good authority node is one that is pointed to by many good hub nodes.

Select the respective check boxes depending on the centrality required.

Clustering coefficient

The Clustering coefficient of a node indicates the compactness of its links to neighbors. A higher clustering coefficient value indicates a well-connected neighborhood.

Select the **Clustering coefficient** check box if you want to compute this centrality.

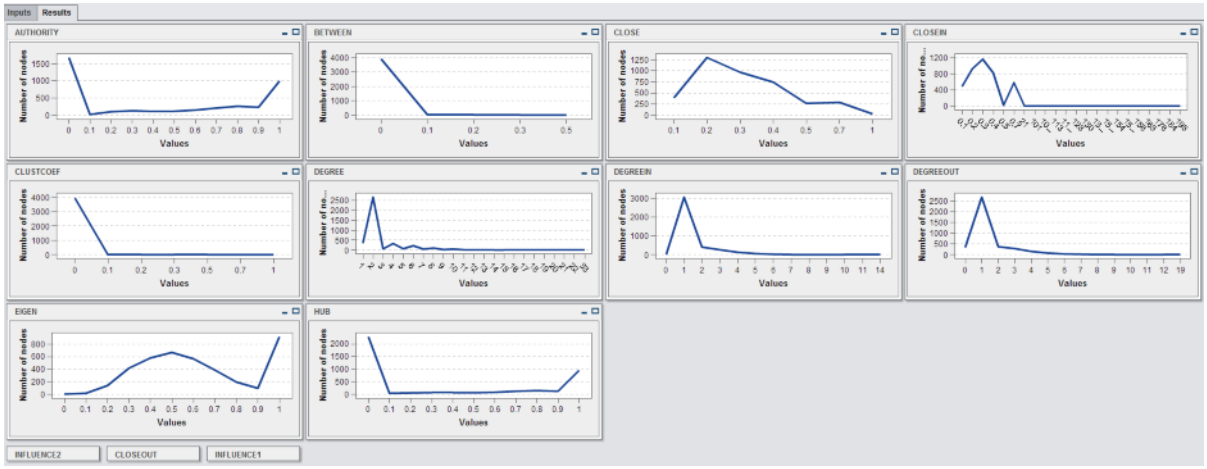
6 Click **OK**.

7 On the toolbar, select ▶. The centrality measures that you have selected are computed for each community or network.

View Results of the Centrality Measures Computation Workflow Step

When you run the Centrality Measures Computation workflow step, graphical results are displayed for the basic centrality measures and the advanced centrality measures that you select. A separate graph is plotted for each centrality measure. The X-axis of the graph indicates the values that are computed for the centrality measure. The Y-axis indicates the number of nodes in the community or the network that have this centrality value.

Display 10.2 *Centrality Measures Computation: Graphical Results*



Based on the distribution of centrality measures within a community or a network, you can understand the relative significance of the nodes. Also, you can use this information for defining roles within your community.

11

Role Assignment

Overview of the Role Assignment Workflow Step

The Role Assignment workflow step enables you to define roles for the nodes in a community. Role definitions enable you to summarize the output that SAS Customer Link Analytics produces. For example, by using centrality measures, you can categorize your nodes within a community into various roles. As a business user, you can easily interpret these roles and thus focus on the important nodes. The role definitions can be used downstream for marketing and campaign activities. In addition, you can define the order of precedence for each role. When you run this workflow step, a role is assigned to a node if it satisfies the expression that you have defined for that role. A single node can satisfy the expressions that are defined for one or more nodes. In this case, the role with the highest precedence is assigned to the node.

SAS Customer Link Analytics provides you with a default predefined role. In addition, you can define your own roles. For example, you can define the following set of roles:

Table 11.1 *Sample Role Definitions*

Role Name	Description
Leader	Node that typically has the highest values for the Betweenness, In-degree, Out-degree, and Closeness centrality measures.
Follower	Node that has centrality measures similar to that of a leader but to a lesser extent.

Role Name	Description
Bridge (also called *boundary spanner*)	Node that links to multiple communities. Also, has values of centrality measures lower than that of leaders and followers.
Neutral (also called *sub-follower*)	Node that is similar to a follower but to a lesser extent. In terms of weighted data, it can be an outlier, with its weight being in the higher range.
Outlier (also called *peripheral player*)	Node that has the lowest values for the Betweenness, In-degree, Out-degree, and Closeness centrality measures. It can be linked to only one other node.

Define a Role

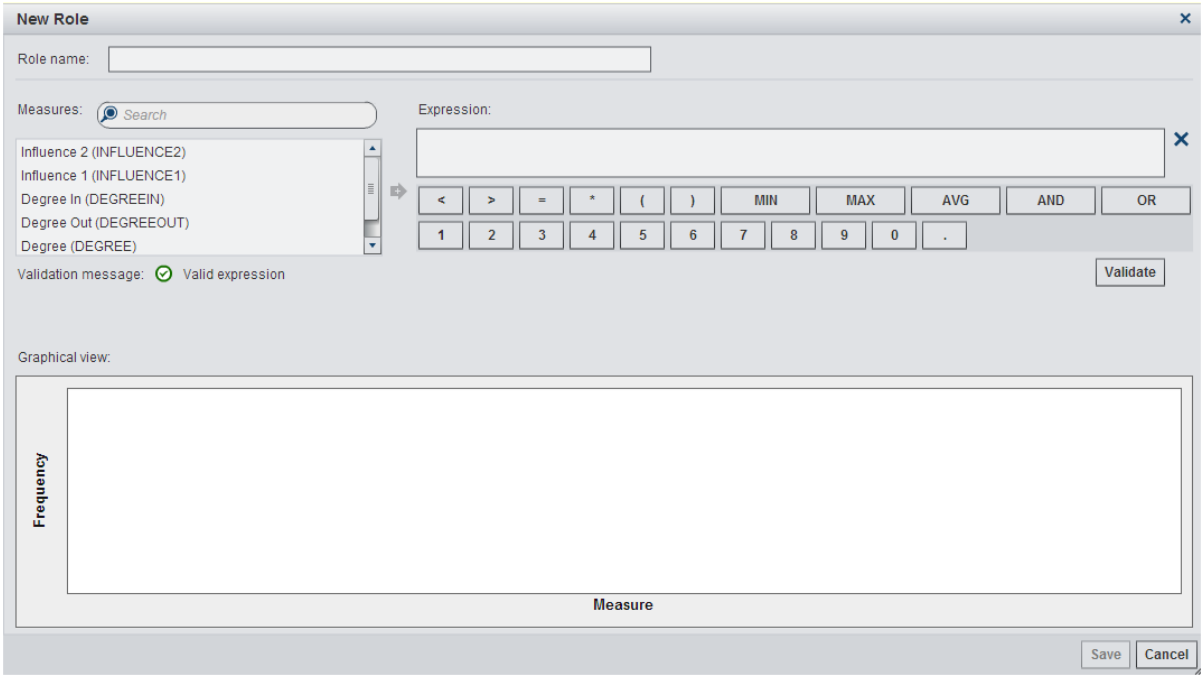
To define a role:

1 In the Workflow Diagram pane, click **Role Assignment**.

2 On the toolbar, select .

3 On the toolbar, select . The New Role window appears.

Display 11.1 *New Role Window*



4 Enter the role name. For example, you can define a role such as leader, follower, and so on.

5 Define an expression for the role. You define an expression by using a valid combination of centrality measures, constant values, relative values, arithmetic operators, and logical operators.

If you are including a centrality measure in your expression, you can view its graph. The graph can help you understand how the centrality measure is distributed within the network.

Example: Define a role of a leader based on the following business requirements:

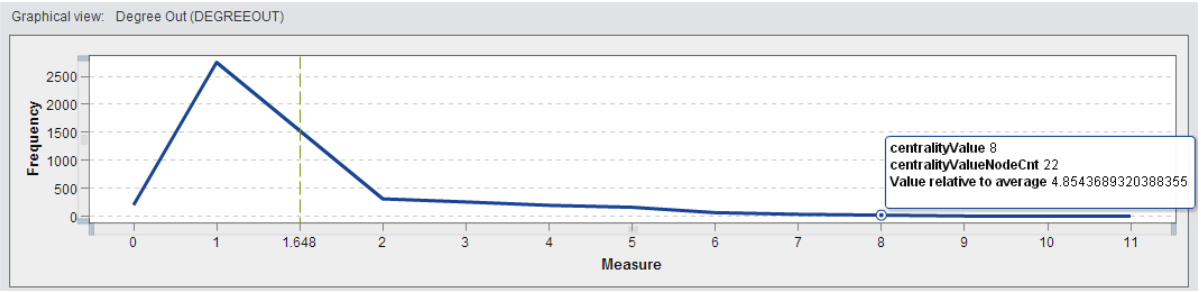
- A subscriber (node) who is making calls to at least 8 subscriptions. In other words, nodes that have an in-degree or out-degree value greater than 8.
- A subscriber (node) who is close to the other nodes in the community. In other words, nodes that have a higher closeness value (for example, greater than 0.5).

For this role, complete the following steps to define the expression:

a Click **(**.

b Double-click **Degree Out (DEGREEOUT)** from the **Measures** list, and click **>**.

Display 11.2 *Graph of Degree-out Centrality*



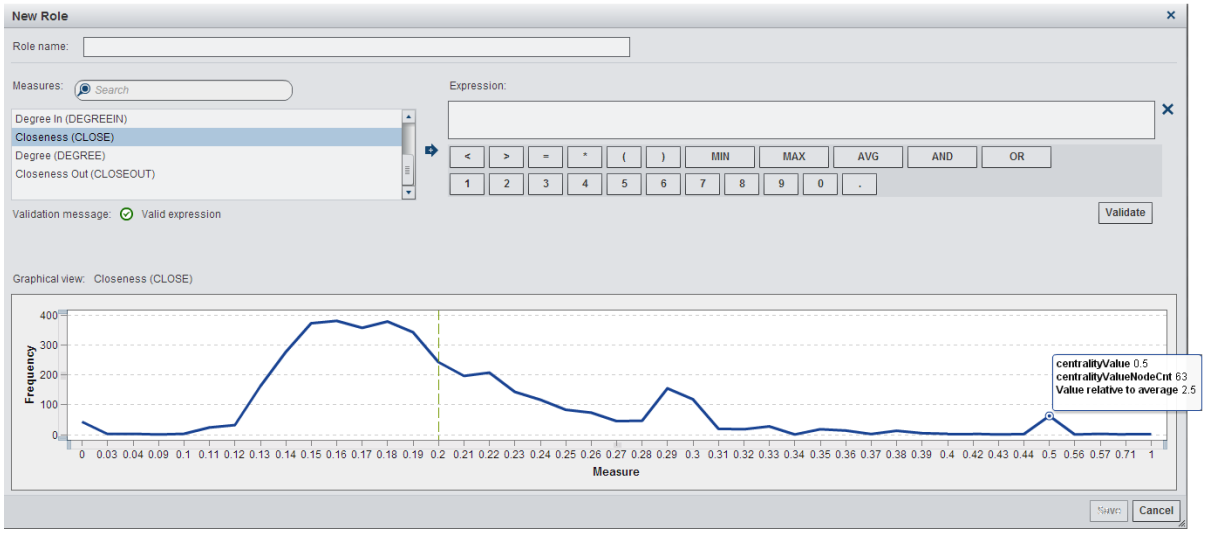
c In the graph, click the data point that corresponds to a value of degree-out equal to 8 on the X-axis. In the expression, this value is displayed as **4.85 * AVG(DEGREEOUT)**.

d Click **)**, and then click **AND**.

e Double-click **Closeness (CLOSE)** from the **Measures** list, and click **>**.

f In the graph, select a data point that corresponds to a closeness value of 0.5 on the X-axis. In the expression, this value is displayed as **2.5 * AVG(CLOSE)**.

Display 11.3 Graph of Closeness Centrality



g Click **)**. The final expression is displayed as follows:

*(DEGREEOUT > 4.85 * AVG(DEGREEOUT)) AND (CLOSE>2.5 * AVG(CLOSE))*

TIP You can also enter the expression using the keyboard. However, you will not be able to use the relative values of the centrality measures effectively.

6 Click **Validate** to ensure that you have specified the correct expression. Alternatively, you can click **Save**.

Note: If there are any errors, a message is displayed. Correct the errors.

7 Click **Save**. The role that you defined is displayed in the Details pane.

8 Click **OK** to confirm the role definition.

Note: Before you run the workflow step, make sure that you define the precedence for the roles. For details, see "Define the Precedence of a Role" on page 67.

Edit a Role

You can edit the roles that you have defined. However, you cannot edit the default role.

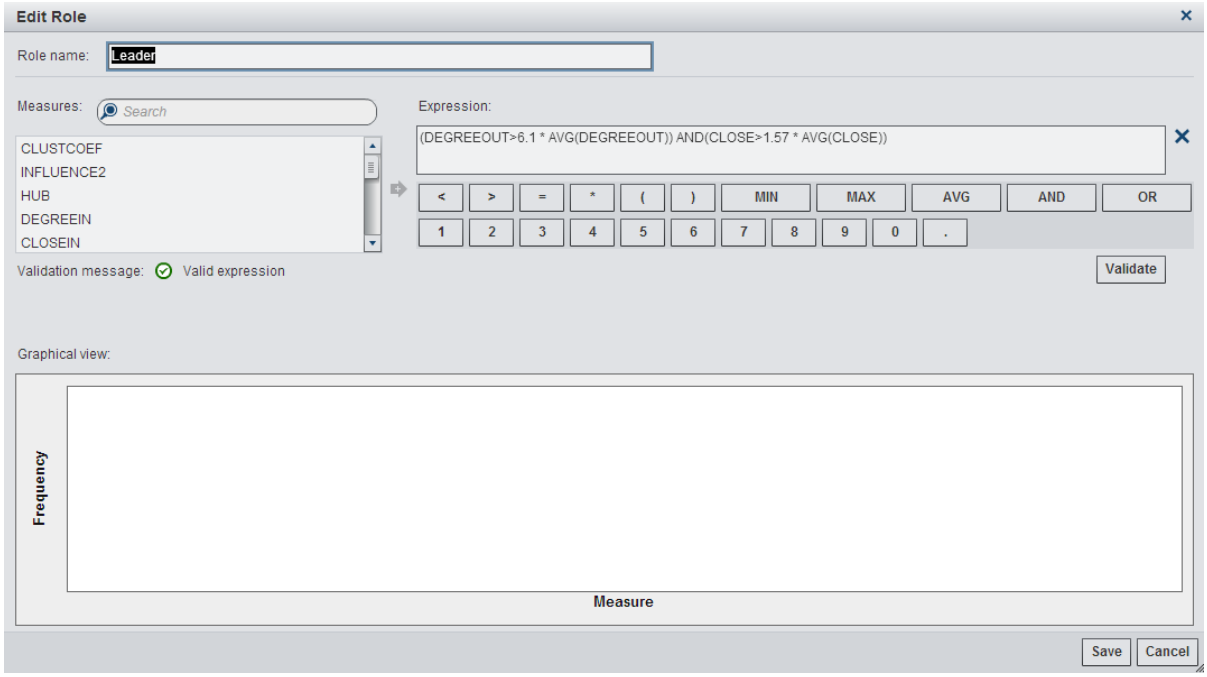
To edit a role:

1 In the Workflow Diagram pane, select **Role Assignment**. The default role and the roles that you have defined are displayed.

2 On the toolbar, select .

3 Double-click the role that you want to edit. Alternatively, on the toolbar, click **Edit Role**. The Edit Role window appears.

Display 11.4 *Edit Role Window*



4 Make the required changes.

5 Click **Save**. The window closes.

6 Click **OK** to confirm the changes that you have made.

Define the Precedence of a Role

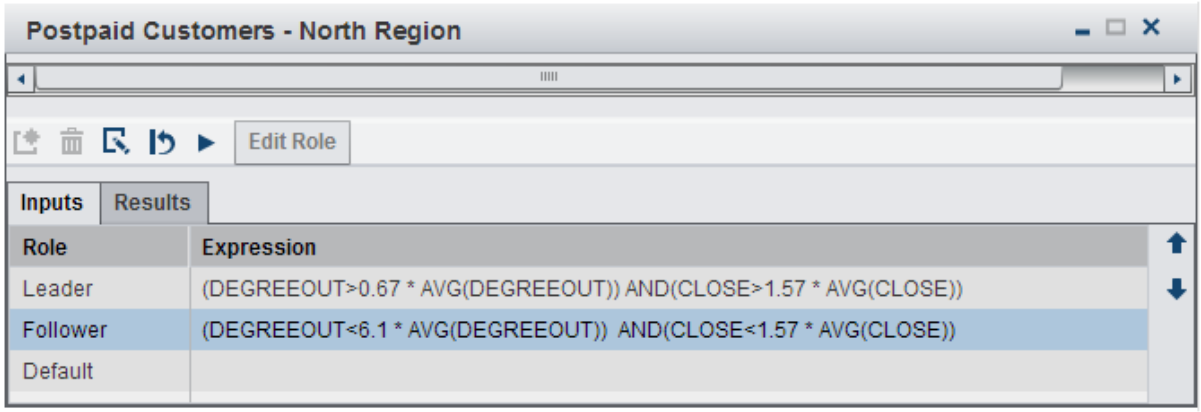
When you define all the roles that you want to consider for the nodes in the community, you can define the precedence for each role. The role with the highest precedence is assigned to a role if a node satisfies the expression that is defined for two or more nodes.

To change the precedence of a role:

1 In the Projects workspace, open the project on which you want to work.

2 In the Workflow Diagram pane, select **Role Assignment**. The roles that you have defined are displayed.

Display 11.5 Role Assignment: Change Precedence



3 Select the role whose precedence you want to change.

4 In the right pane, click ⬆ to move the role up in the list. This increases the precedence of the role. Alternatively, click ⬇ to reduce the precedence of the role.

Assign a Role to a Node

In order to assign a role to each node in a community, you have to run the Role Assignment workflow step. Before you run the workflow step, make sure that you have defined all the roles that you need. Also, make sure that you have defined the precedence order for these roles.

To assign a role to a node:

1 In the Workflow Diagram pane, select **Role Assignment**. The roles that you have defined are displayed, along with the default role.

2 Make sure that you have defined all the required roles. Also, make sure that the correct precedence order is set up for these roles.

3 On the toolbar, select ▶. A role is assigned to a node based on the following conditions:

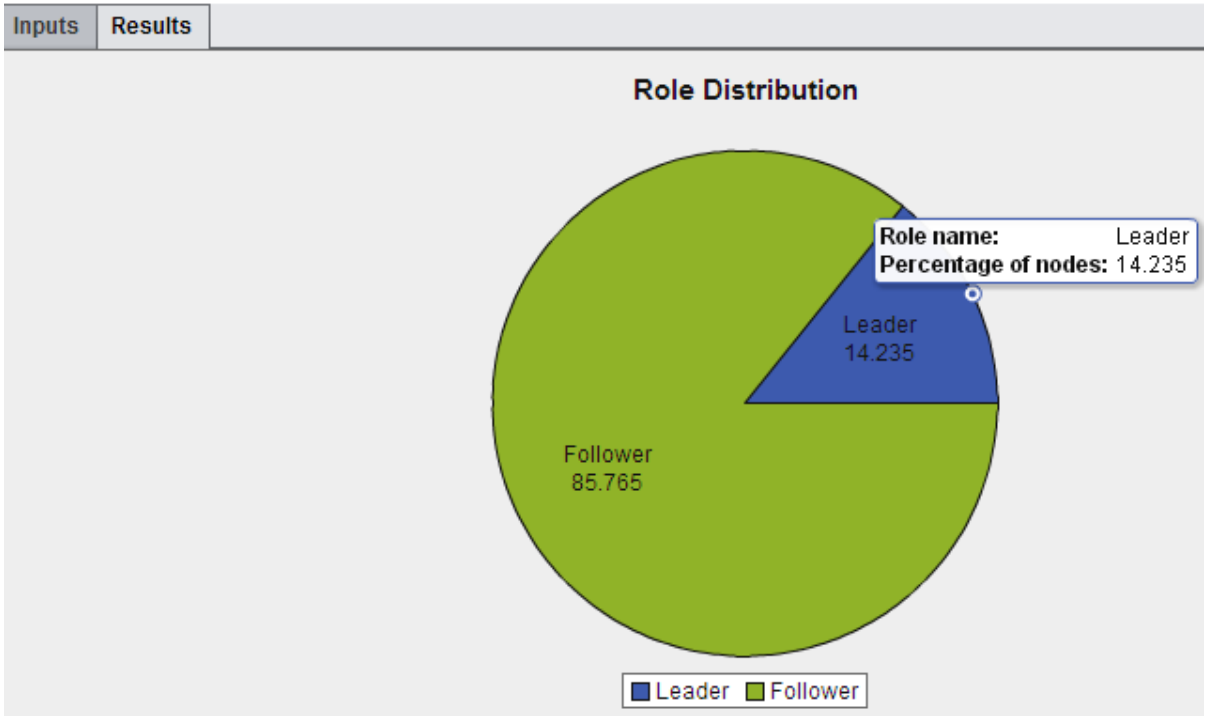
- The node uniquely satisfies the conditions that are defined in the expression of a single role. The role is assigned to the node.
- The node satisfies the conditions that are defined in the expression of two or more roles. The role with the highest precedence order is assigned to the node.
- The node does not satisfy the conditions that are defined for any of the roles. The default role is assigned to the node.

After you run the workflow step, you can also view how the roles are distributed within the network. If a default role is assigned to any of the nodes, you might want to reconsider the expressions that you have defined for the roles.

View the Distribution of Roles

When you run the Role Assignment workflow step, an appropriate role is assigned to each node of the community. SAS Customer Link Analytics enables you to view the distribution of these roles within your network. To do so, select the Role Assignment workflow step, and then select the **Results** tab. A pie chart that indicates the distribution of the roles across your network is displayed.

Display 11.6 *Distribution of Roles*



This pie chart shows the distribution of the leader and follower roles in the network. The pie chart indicates that 85.765% of the nodes are assigned to the follower role and 14.235% of the nodes are assigned to the leader role.

12

Batch Processing

Overview of Batch Processing

After you create a project and successfully run all its workflow steps, a project completes one run in design mode. If you are satisfied with the results that are produced in design mode and you want the same project parameters to run on the source data at regular time intervals, you can push the project to batch mode. Once the project is in batch mode, you cannot configure any of its workflow steps or change its properties. You can only change the description of the project. If you want to make any other changes, you have to pull the project back into design mode.

Push a Project to Batch Mode

You can push a project to batch mode only if its status in design mode is **Executed Successfully**. This status indicates that all the workflow steps of that project have run successfully.

To push a project to batch mode:

1. In the Projects workspace, select the project that you want to push to batch mode. Make sure that the project's status is **Executed Successfully**.
2. In the Properties pane, click **Edit**.
3. Select **Batch** as the mode of the project.
4. Click **OK**. The project is now in a view-only state. You can only view the details of the workflow steps.

   After you push the project to batch mode, a batch code file is created in the following location: `<SAS configuration path>/Lev1/AppData/SASCustomerLinkAnalytics/CustLinkAnalytcs5.5/projects/batchcode`. The filename is as follows: sia_batch_exe_<*Project name*>_<*Project ID*>.sas. The project name is the name that you entered for the project, and the Project ID is the sequence number that is generated for

the project. For example, assume that the project's name is Project1 and the sequence number that is generated for this project is 25. In this case, the batch code file, sia_batch_exe_project1_25.sas is created in the specified location. Your administrator can use this file to schedule the batch run of the project. For details, see *SAS Customer Link Analytics: Administrator's Guide*.

When a project is run in batch mode, its execution status is reflected in the Properties pane and results are displayed on the **Results** tab.

Pull a Project into Design Mode

If you are not satisfied with the results that are produced in batch mode, you might need to configure the workflow steps again according to your requirements. However, in order to configure a workflow step, you have to pull the project back into design mode. You can pull the project into design mode only if none of its workflow steps are running.

To pull a project into design mode:

1. In the Projects workspace, select the project that you want to pull into design mode. Make sure that the project's status is not **In Progress**.
2. In the Properties pane, click **Edit**.
3. Select **Design** as the mode of the project.
4. Click **OK**. Now you can configure the workflow steps of this project.

Part 3

Extended Capabilities

13

Using Sample Data

Working with Sample Data

Sample Data Details

SAS Customer Link Analytics provides a sample data set to help you familiarize yourself with the tasks that you can perform in the Projects and the Administration workspaces. This sample data is readily available to you when your administrator installs SAS Customer Link Analytics. The following table provides details about the sample data.

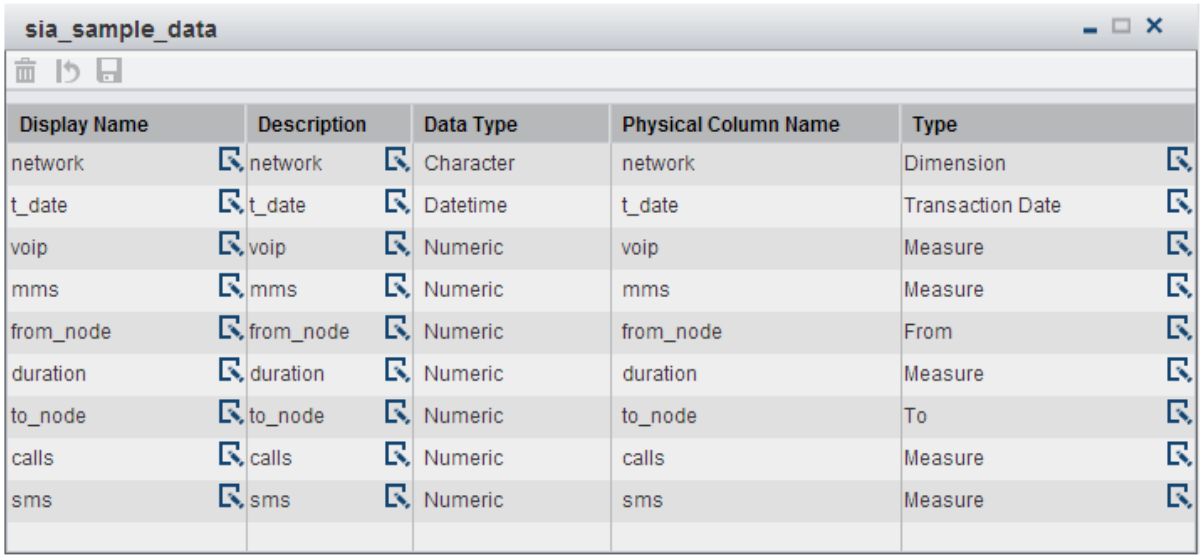
Table 13.1 *Sample Data Details*

Table name	sia_sample_data
Library name	sia_sample
Metadata location	`/Products/SAS Customer Link Analytics/Cust Link Analytics 5.5/Data Sources/Sample Data`
Physical location	`<SAS configuration path>/Lev1/AppData/SASCustomerLinkAnalytics/CustLinkAnalytics5.5/projects/data/sample`

Accessing the Sample Data in the Administration Workspace

You can access the sample data table, sia_sample_data, in the Administration workspace. It is configured as a transactional table and contains the following columns. You can perform all the relevant tasks for this table. For details, see Chapter 15, “Working with Tables,” on page 85.

Display 13.1 *Columns of Sample Data Table*



sia_sample_data

Display Name	Description	Data Type	Physical Column Name	Type
network	network	Character	network	Dimension
t_date	t_date	Datetime	t_date	Transaction Date
voip	voip	Numeric	voip	Measure
mms	mms	Numeric	mms	Measure
from_node	from_node	Numeric	from_node	From
duration	duration	Numeric	duration	Measure
to_node	to_node	Numeric	to_node	To
calls	calls	Numeric	calls	Measure
sms	sms	Numeric	sms	Measure

A source data profile, sample_source_profile, is defined for this table. You can perform all the relevant tasks for this source data profile. For details, see Chapter 16, “Working with Source Data Profiles,” on page 97.

Working in the Projects Workspace Using Sample Data

Because the sample data is preconfigured in the Administration workspace, you can immediately start working on a new project by selecting the sample source file. In the Projects workspace, define a new project and select **sample_source_profile** from the **Source data profile** list. For details, see Chapter 5, “Managing Projects,” on page 25.

After you define a project, you can begin working on its workflow steps. For details, see Chapter 6, “Managing Workflow Steps,” on page 33.

14

Output of SAS Customer Link Analytics

Output Data of SAS Customer Link Analytics

Overview

The output that SAS Customer Link Analytics produces is stored in business data tables. These data tables are classified as follows:

Final and intermediate tables
: These tables contain the output data that is produced as a result of running workflow steps. For details, see "Output of Workflow Steps" on page 77.

Reporting data tables
: These output tables contain data that can be further visualized in SAS Visual Analytics tools. For details, see For details, see "Preparing Reporting Data for Visual Analytics Tools" on page 78.

Output of Workflow Steps

When you complete one run in design mode, SAS Customer Link Analytics produces certain results for each node and stores them in an output table. This table, cla_ra_p1_<*project ID*>, is created in the following location: `<SAS configuration path>/AppData/SASCustomerLinkAnalytics/CustLinkAnalytics5.5/projects/data/business_output`. You can use this table as input for various marketing automation or campaign management tools that are used for further analysis. For details about the structure of this table, see *SAS Customer Link Analytics: Data Reference Guide*.

Note: If the project is in batch mode, the table is refreshed with batch results.

The cla_ra_p1_<*project ID*> table stores the following details:

Node ID
: indicates the unique identifier that is assigned to a node.

Role ID
: indicates the ID of the role that is assigned to a node when you run the Role Assignment workflow step.

Role name
: indicates the role that is assigned to a node when you run the Role Assignment workflow step. For example, the role name can be leader, follower, and so on.

Community ID
: indicates the ID of the community to which a node belongswhen you run the Community Building workflow step.

Centrality measures
: displays the values of the centrality measures that SAS Customer Link Analytics computes for the centrality measures that you select in the Centrality Measures workflow step. For details, see “Configure the Centrality Measures Computation Workflow Step” on page 58.

Preparing Reporting Data for Visual Analytics Tools

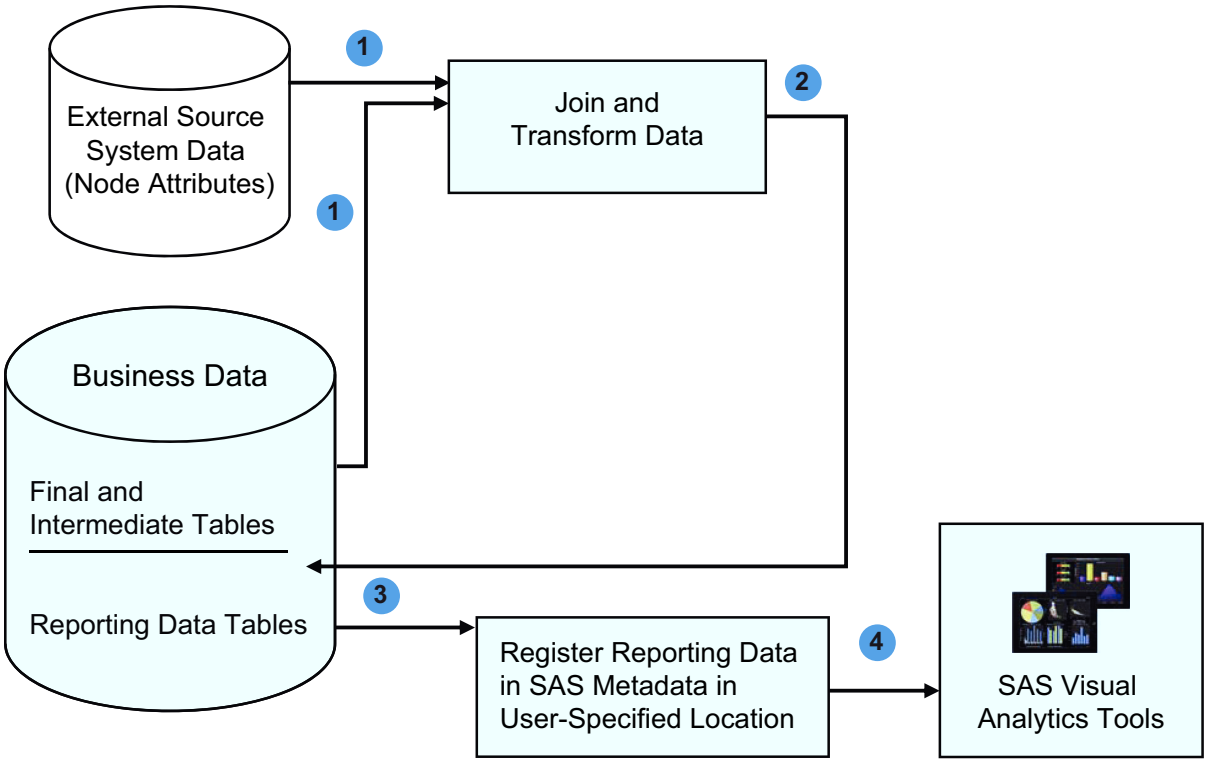
Overview

SAS Customer Link Analytics enables you to prepare data that can be used to produce business reports in visual analytics tools such as SAS Visual Analytics Explorer. In order to prepare the data, you have to run the appropriate stored process depending on whether your project is in design mode or batch mode. This data is stored in an output table. For details about the sample output data that is created, see *SAS Customer Link Analytics: Data Reference Guide*.

Note: Before you run the stored process, make sure that you have **Create**, **Read**, **Write**, **Delete**, **ReadMetadata**, and **WriteMetadata** access to the library metadata object and the metadata folder in which you want to register the output table. Contact your administrator if you need any assistance.

The following diagram indicates the flow of output data.

Figure 14.1 Flow of Reporting Data



1 Data extraction
2 Data transformation and storage
3 Data registration
4 Data exploration

Prepare Data for a Project in Design Mode

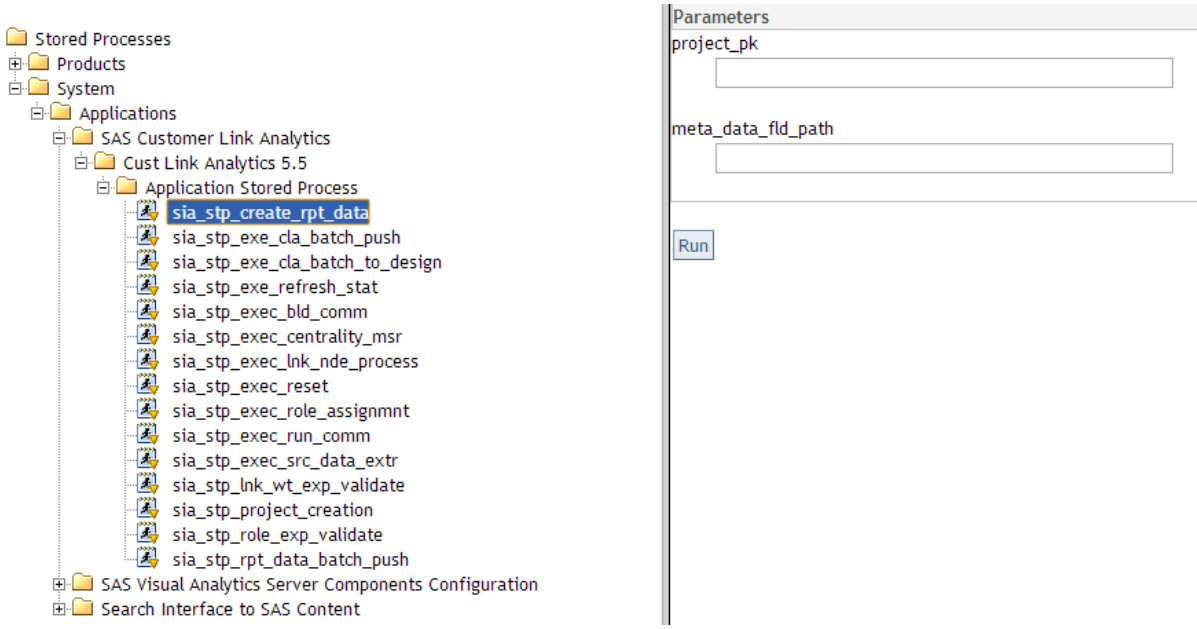
If the project is in design mode, complete the following steps:

1 Log on to SAS Stored Process Web Application by using the appropriate credentials. For details, see "Log On to SAS Stored Process Web Application" on page 107.

2 In the application window, click **List Available Stored Processes and Reports**. The list of stored processes is displayed in the left pane.

3 Expand **Stored Processes** ▸ **System** ▸ **Applications** ▸ **SAS Customer Link Analytics** ▸ **Cust Link Analytics 5.5** ▸ **Application Stored Processes** to view various stored processes.

Display 14.1 *Stored*



4 Select the **sia_stp_create_rpt_data** stored process and specify values for the following parameters:

project_pk
Specify the unique ID of the project for which you want to prepare data.

meta_data_fld_path
Specify the metadata path in which you want to register the output table. If you do not specify any path, the output table is registered in the **Shared Data** metadata folder.

5 Click **Run**.

6 When you run the stored process, the following tasks are performed:

- An output table, RPT_DATA_<*Project ID*>, is created and registered in the metadata path that you specify.
- A log file, sia_create_rpt_data_<*Project ID*>, is created in the following location: `<SAS configuration path>/AppData/SASCustomerLinkAnalytics/CustLinkAnalytics5.5/projects/logs`.

7 Log off from SAS Stored Process Web Application.

Prepare Data for a Project in Batch Mode

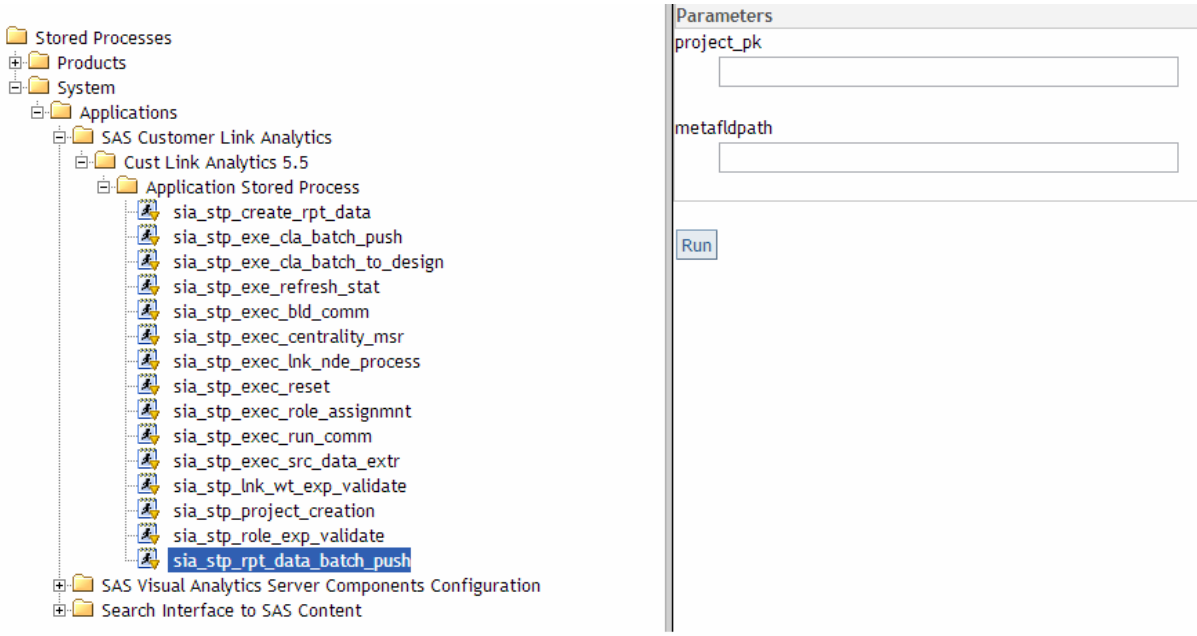
If the project is in batch mode, complete the following steps:

1 Log on to SAS Stored Process Web Application by using the appropriate credentials. For details, see “Log On to SAS Stored Process Web Application” on page 107.

2 In the application window, click **List Available Stored Processes and Reports**. The list of stored processes is displayed in the left pane.

3 Expand **Stored Processes** ▸ **System** ▸ **Applications** ▸ **SAS Customer Link Analytics** ▸ **Cust Link Analytics 5.5** ▸ **Application Stored Processes** to view various stored processes.

Display 14.2 *Stored Processes*



4 Select the **sia_stp_rpt_data_batch_push** stored process and specify values for the following parameters:

project_pk
Specify the unique ID of the project for which you want to prepare data.

metafldpath
Specify the metadata path in which you want to register the output table. If you do not specify any path, the output table is registered in the **Shared Data** metadata folder.

5 Click **Run**.

6 When you run the stored process, the following tasks are performed:

- A batch code file, sia_rpt_data_exe_<*Project name*>_<*Project ID*>.sas, is created in the following location: `<SAS configuration path>/AppData/SASCustomerLinkAnalytics/CustLinkAnalytics5.5/projects/batchcode`.
- A log file, sia_rpt_data_batch_push_<*Project ID*>.log, is created in the following location: `<SAS configuration path AppData/SASCustomerLinkAnalytics/CustLinkAnalytics5.5/projects/logs`.

7 Log off from SAS Stored Process Web Application.

8 Run the batch code file in the SAS environment. When you run the batch code, the following tasks are performed:

- An output table, RPT_DATA_<*Project ID*>, is created and registered in the metadata path that you specify.

- A log file, create_rpt_data_batch_<*Project ID*>, is created in the following location: **<SAS configuration path>/Lev1/AppData/SASCustomerLinkAnalytics/CustLinkAnalytics5.5/projects/logs**.

Alternatively, you can request your administrator to schedule a batch run for the code file. For details, see *SAS Customer Link Analytics: Administrator's Guide*.

Part 4

Administration Workspace

15

Working with Tables

Overview of the Administration Workspace

The Administration workspace enables you to complete the prerequisite tasks that are required for defining a project and completing its workflow. You can perform the following tasks in the Administration workspace:

Importing tables
: You can import tables that are registered in SAS metadata. When you import tables, SAS Customer Link Analytics tags these tables according to the type of data that they contain. This tagging ensures that SAS Customer Link Analytics uses the correct table types when you define source data profiles. For details, see “Import a Table” on page 88.

Mapping table columns
: You have to map the columns of tables that you have imported. When you map columns, SAS Customer Link Analytics tags each column according to the type of data that it can contain. This tagging ensures that SAS Customer Link Analytics uses the correct columns when you work with workflow steps. For example, when you define an expression for a link weight, SAS Customer Link Analytics provides a list columns that are mapped as measures. For details, see “Configure the Column Values of a Table” on page 90.

Defining source data profiles
: When you define a source data profile, SAS Customer Link Analytics groups the relevant tables that are required to define a project and run its workflow steps. This task restricts you from using any incorrect data for your analysis. For details, see “Create a Source Data Profile” on page 97.

Refreshing data
: When you refresh tables that are tagged as transactional tables, SAS Customer Link Analytics provides information about the latest state of the data in the transactional data. This task helps you understand the start date and the end date of the period for which data is available in the transactional table. For details, see “Refresh Data in a Transactional Table” on page 92.

About Tables

SAS Customer Link Analytics requires certain information such as transactional data of links and nodes and their relevant attributes in order to build communities. Therefore, to use this information, you have to import the tables that store these details. You can import only those tables that your administrator has registered in SAS Management Console.

Before you import a table, you must be aware of the following information about the table:

- the library to which the table belongs
- the physical name of the table that you want to import
- the type of information that the table stores
- the level at which data is aggregated in the table

Table Types

Transactional Tables

A transactional table typically stores graph-type data. A transactional table in SAS Customer Link Analytics must contain the following mandatory columns:

- from
- to
- transaction date
- measure

For example, in the communications industry, a transactional table typically stores the call detail records. Each record can provide the following information:

Table 15.1 Sample Columns of a Transactional Table

Column Name	Description
From number	Contains the calling number.
To number	Contains the called number.
Usage date	Contains the date on which the communication event occurred.
Aggregated value	Contains the aggregated value for the event, such as call duration, volume of data downloaded, and so on.

After you import a transactional table, you have to configure the type of each column. That is, you have to identify whether the column is a dimension, measure, date, transaction date, calling number (from), or called number (to). For a transactional table, you must define only one column that maps to each of the Transaction date, From , or To columns.. Also, you must define at least one column as a Measure type. For example, for the previous table, you must define the following mappings between the column and type:

Table 15.2 Mandatory Column Mappings for Transactional Tables

Column Name	Type
From number	From
To number	To
Usage date	Transaction date
Aggregated information	Measure

The rest of the columns can be of date or dimension type.

Attribute Tables

An attribute table is further classified depending on whether it contains an inclusion list of nodes or links or attributes of links or nodes.

An inclusion list helps to filter nodes or links that are relevant for network analysis. An inclusion list can be a node inclusion list or a link inclusion list. For details, see “Example 1: Node Inclusion List” on page 39 and “Example 2: Link Inclusion List” on page 40. The transactional data of nodes or links that are available in the inclusion list is extracted from the source tables in the Data Extraction workflow step. For details, see “Overview of the Data Extraction Workflow Step” on page 39.

The link attribute or node attribute tables contain information about the links or nodes, respectively. The columns of these tables are usually of the Dimension type.

After you import an attribute table, you have to configure the type of each column. That is, you have to identify whether the column is a dimension, date, calling number (from), or called number (to). In an attribute table that contains a node inclusion list or node attributes, there must be only one column that maps to the From type. Similarly for an attribute table that contains a link inclusion list or link attributes, there must be only one column that maps to the From and To type.

Table 15.3 Mandatory Column Mapping for Node Inclusion List and Node Attribute Table

Column Name	Type
Node ID	From

Table 15.4 Mandatory Column Mappings for Link Inclusion List and Link Attribute Table

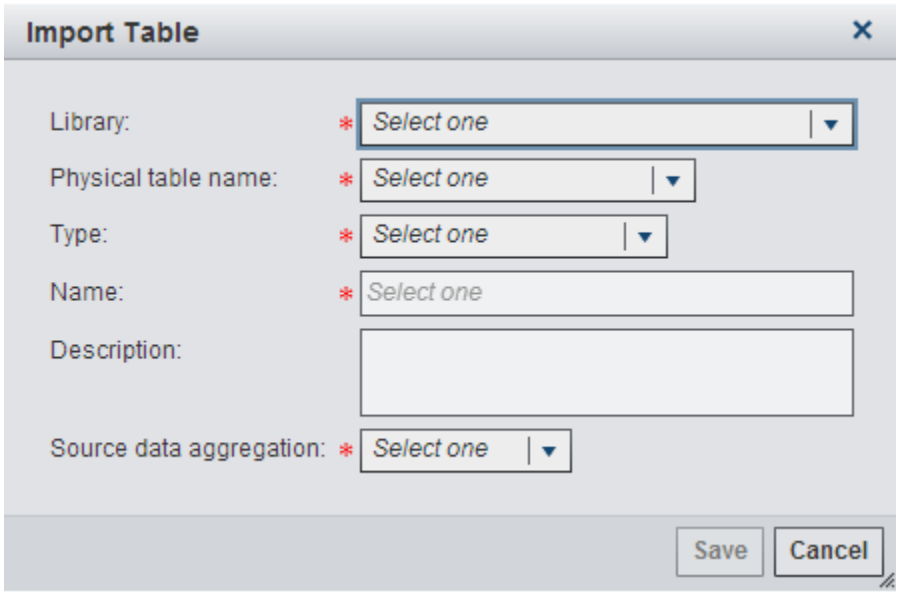
Column Name	Type
From Node ID	From
To Node ID	To

Import a Table

To import a table:

1. In the Administration workspace, select **Tables** in the Category pane.
2. On the toolbar, select  . The Import Table window appears.

Display 15.1 *Import Table Window*



3 Select the library to which the table that you want to import belongs.

4 Select the table that you want to import. The list displays the tables that belong to the library that you have selected and those that your administrator has registered in SAS Management Console.

5 Select the type of table depending on the information that it stores. The following options are available:

Link attribute
indicates that the table stores information about links.

Link inclusion list
indicates that the table stores an inclusion list for links. This inclusion list is used in the Data Extraction workflow step to filter the relevant links. For details, see “Example 2: Link Inclusion List” on page 40.

Node attribute
indicates that the table stores information about nodes.

Node inclusion list
indicates that the table stores an inclusion list for nodes. This inclusion list is used in the Data Extraction workflow step to filter the relevant nodes. For details, see “Example 1: Node Inclusion List” on page 39.

Transactional
indicates that the table stores transactional information for links and nodes.

6 Enter the name of the table. The table will be uniquely identified in the SAS Customer Link Analytics interface with this name.

7 Enter the description of the table. You can enter brief information about the table.

8 Select the level at which data is aggregated in the table. The following options are available: **Monthly**, **Daily**, and **Fully**. This field is applicable only if you are importing a transactional table.

9 Click **Save**. The table is added to the list of tables.

10 (Optional) Double-click the table to view its columns.

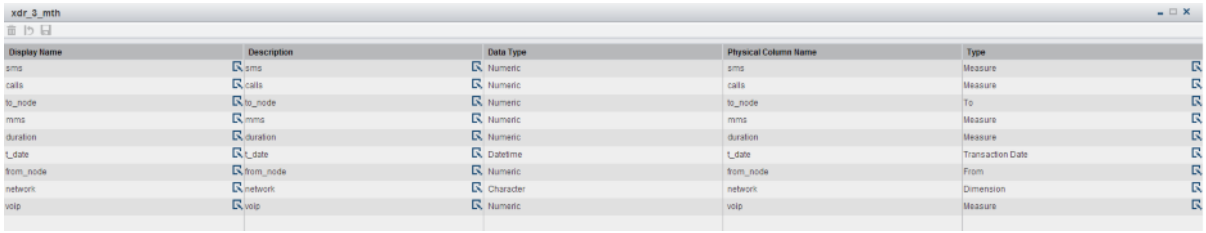
Configure the Column Values of a Table

When you import a table, its column metadata information is automatically imported. This information includes the display name, description, data type, physical column name, and type of each column. SAS Customer Link Analytics assigns a default value for the metadata details of each column. You can change the display name, description, and type of a column. However, you cannot change the physical column name or data type of the column. In addition, you can change the type of a column only if the table is not used in the definition of a source data profile.

To change the default value of a column's metadata:

1 In the Administration workspace, double-click the table whose column values you want to change.

Display 15.2 *Columns of a Table*



The following details are displayed for each column:

Display Name
displays the name of the column. You can change the value of this field.

Description
displays a short description of the column. You can change the value of this field.

Data Type
displays the data type of the column. You cannot change this value.

Physical Column Name
displays the column name as defined in the source table.

Type
displays type of the column. You can choose any one of the following values:

Note: Make sure that you select the correct type for each column. Otherwise, you will produce incorrect results.

Date
indicates that the column contains a date value.

From
indicates that the column contains a calling number.

To
indicates that the column contains a called number.

Measure
indicates that the column contains an aggregated value such as call duration or number of calls.

Dimension
indicates that the column contains a character type of information such as node ID, service ID, or service provider's ID.

Transaction date
indicates that the column contains the date on which the communication event occurred. For example, this column can contain the usage date. The historical period that you want to consider to extract data for your analysis is based on the value that you specify for the transaction date type column.

Note: Depending on the type of table that you have imported, certain types of columns are mandatory. You can define only one column for the **Transaction date**, **From**, and **To** data types. However, you can define one or more columns for the **Measure** data type.

Table 15.5 *Mandatory Column Types*

Table Type	Mandatory Column Type
Transactional	■ Transaction date ■ From ■ To ■ Measure
Link attribute	■ From ■ To
Node attribute	■ From
Link inclusion list	■ From ■ To
Node inclusion list	■ From

2 (Optional) Change the default value. To do so, click the that is available next to the value.

3 Make sure that you choose the mandatory column types as explained in Table 15.5 on page 91.

4 On the toolbar select .

Refresh Data in a Transactional Table

After you import a transactional table and configure it, you have to refresh it before you define a project that is based on this transactional table. When you refresh a table, the start date and the end date of the period for which the transactional data is available in the table are populated in the respective columns. In the Data Extraction workflow step, the default value for the end date and the duration of the historical period is populated based on these dates. For details, see “Configure the Data Extraction Workflow Step” on page 40.

Note: You must refresh a transactional table that you want to associate with a project. Otherwise, you will not be able to run the Data Extraction workflow step for that project.

To refresh a transactional table:

1. In the Administration workspace, select the transactional table that you want to refresh.
2. On the toolbar, click  . This icon is activated only if you have configured the columns of the transactional table. For details, see “Configure the Column Values of a Table” on page 90. In the Properties pane, the following fields are populated:

   Source start date
   indicates the start date of the period for which transactional data is available in the table.

   Source end date
   indicates the end date of the period for which transactional data is available in the table.

   Statistics calculation date
   indicates the date on which you refreshed the table the last time.

   This table can now be used in the Data Extraction workflow step.

Edit the Properties and Columns of a Table

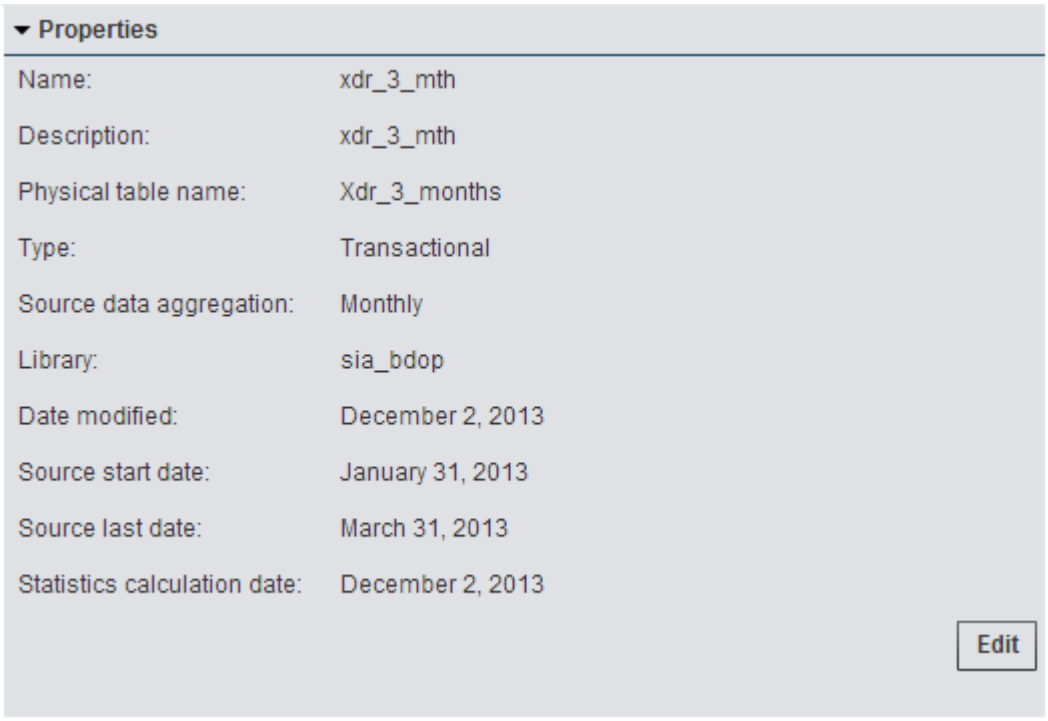
When you import a table, you can view its details in the Properties pane of the Administration workspace. If you want, you can edit the name and description of the table.

To edit the properties of a table:

1. In the Administration workspace, select the **Tables** category.
2. Select the table whose properties or columns you want to edit.
3. To edit the properties:

   a. In the Properties pane, click **Edit**.

Display 15.3 *Properties Pane*



b Change the name or description as required.

c Click **OK**.

4 To edit the columns in the table:

a Double-click the table. Its columns are displayed.

Note: You can edit the columns in a table only if the table is not used in the definition of a source data profile. The toolbar is deactivated, as shown in the diagram here, if the table is used in the definition of a source data profile.

Display 15.4 *Columns in a Table*

b Click the icon that is displayed adjacent to the **Display Name**, **Description**, and **Type** columns and make the required changes.

c On the toolbar, select .

Delete a Table

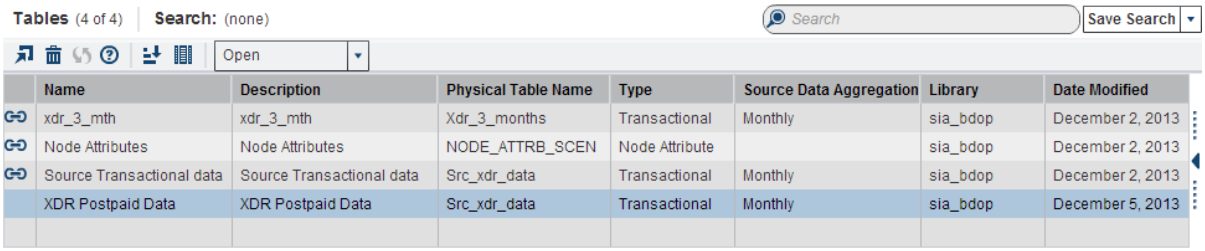
You can delete a table if it is not used in the definition of a source data profile. For details, see “Create a Source Data Profile” on page 97.

To delete a table:

1 In the Administration workspace, select **Tables** in the Category pane.

2 Select the table that you want to delete.

Display 15.5 *Delete Table*



	Name	Description	Physical Table Name	Type	Source Data Aggregation	Library	Date Modified
🔗	xdr_3_mth	xdr_3_mth	Xdr_3_months	Transactional	Monthly	sia_bdop	December 2, 2013
🔗	Node Attributes	Node Attributes	NODE_ATTRB_SCEN	Node Attribute		sia_bdop	December 2, 2013
🔗	Source Transactional data	Source Transactional data	Src_xdr_data	Transactional	Monthly	sia_bdop	December 2, 2013
	XDR Postpaid Data	XDR Postpaid Data	Src_xdr_data	Transactional	Monthly	sia_bdop	December 5, 2013

3 On the toolbar, select 🗑.

Note: You can delete the table if it is not used in the definition of a source data profile. The 🔗 icon indicates that a table is used in the definition of a source data profile.

Delete a Column in a Table

You can delete a column in a table if it is not included in a source data profile.

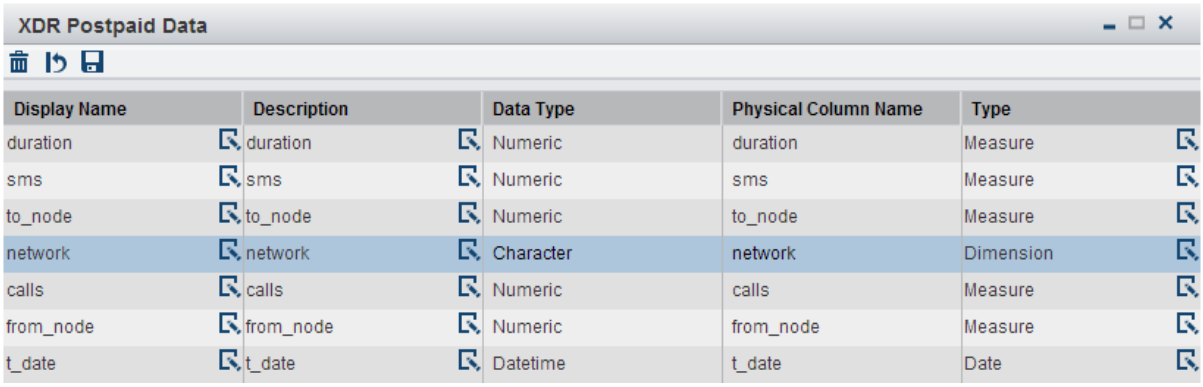
To delete a column in a table:

1 In the Administration workspace, double-click the table whose column you want to delete. Make sure that the table is not included in any of the source data profiles.

TIP A 🔗 indicates that the table is used in the definition of one or more source data profiles.

2 In the column list, select the column that you want to delete.

Display 15.6 *Delete Column in a Table*



Display Name	Description	Data Type	Physical Column Name	Type
duration	duration	Numeric	duration	Measure
sms	sms	Numeric	sms	Measure
to_node	to_node	Numeric	to_node	Measure
network	network	Character	network	Dimension
calls	calls	Numeric	calls	Measure
from_node	from_node	Numeric	from_node	Measure
t_date	t_date	Datetime	t_date	Date

3 On the toolbar, select .

4 In the confirmation window that appears, select **Yes**.

Reset the Value of a Column

You might change the current value of a column. However, later you might want to discard the changed value and retain the value that you had saved the last time. For example, the value that is last saved for the display name of a column is To_node. You change it to Called number. You might want to retain the previous value. In this case, you can use the reset feature.

Note: You can reset a value of a column only if the table is not used in the definition of a source data profile.

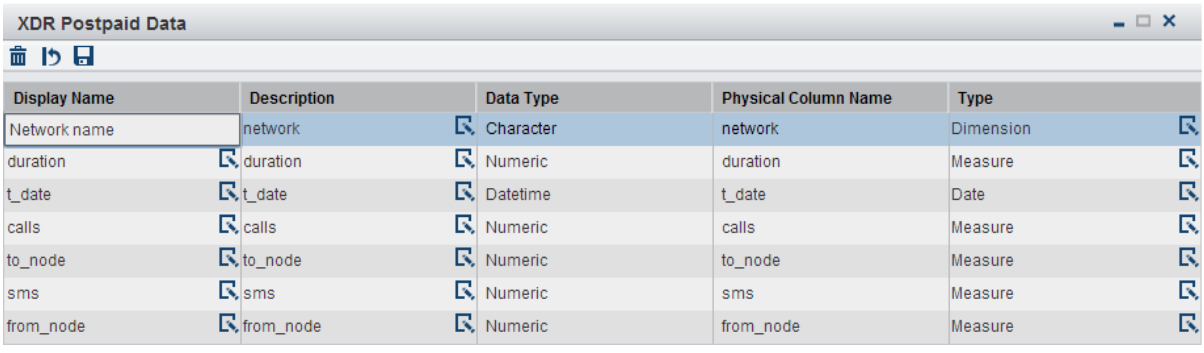
To reset the value of a column:

1 In the Administration workspace, double-click the table whose column value you want to reset. The list of columns in the table is displayed.

2 Click for the column value that you want to reset.

Note: You can reset the value of **Display Name**, **Description**, and **Type** columns.

Display 15.7 *Reset Column*



Display Name	Description	Data Type	Physical Column Name	Type
Network name	network	Character	network	Dimension
duration	duration	Numeric	duration	Measure
t_date	t_date	Datetime	t_date	Date
calls	calls	Numeric	calls	Measure
to_node	to_node	Numeric	to_node	Measure
sms	sms	Numeric	sms	Measure
from_node	from_node	Numeric	from_node	Measure

3 Click .

4 In the confirmation window that appears, click **Yes**. The value changes to the value that was last saved for that column.

16

Working with Source Data Profiles

Overview of Source Data Profiles

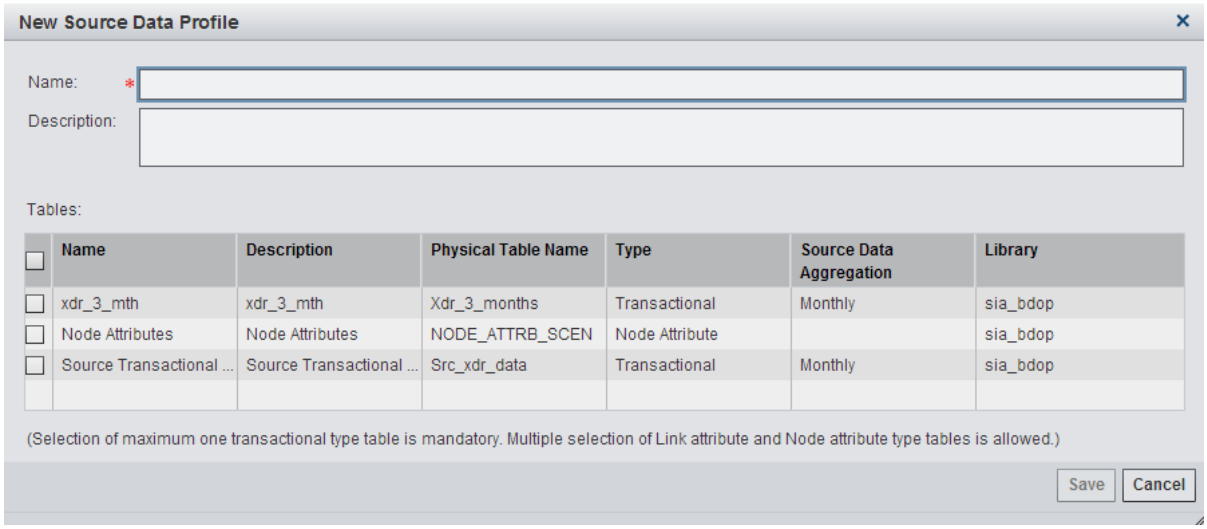
After you import tables and configure their columns, you can group a set of tables to define a source data profile. A source data profile comprises a transactional table and one or more attribute tables. These tables are the ones that you have imported. For details, see “Import a Table” on page 88. You have to associate each project with a source data profile. For details, see “About Projects” on page 25. Communities are built based on the transactional data and other information about the attributes that is available in the source data profile.

Create a Source Data Profile

To create a source data profile:

1 In the Administration workspace, select on the toolbar. The New Source Data Profile window appears.

Display 16.1 *New Source Data Profile Window*



2 Enter the name of the source data profile.

3 Enter the description of the source data profile.

4 Select the tables that you want to include in the source data profile. The list contains the tables that your administrator has imported and configured. For details, see “Import a Table” on page 88.

You must select only one transactional table. However, there is no restriction on the attribute tables that you can include. You can also choose not to add any attribute tables. Attribute tables can be of the following types: **Link Attribute**, **Link Inclusion List**, **Node Attribute**, and **Node Inclusion List**. For details, see “Table Types” on page 86.

5 Click **Save**. The source data profile is now available for selection when you create a new project.

Edit a Source Data Profile

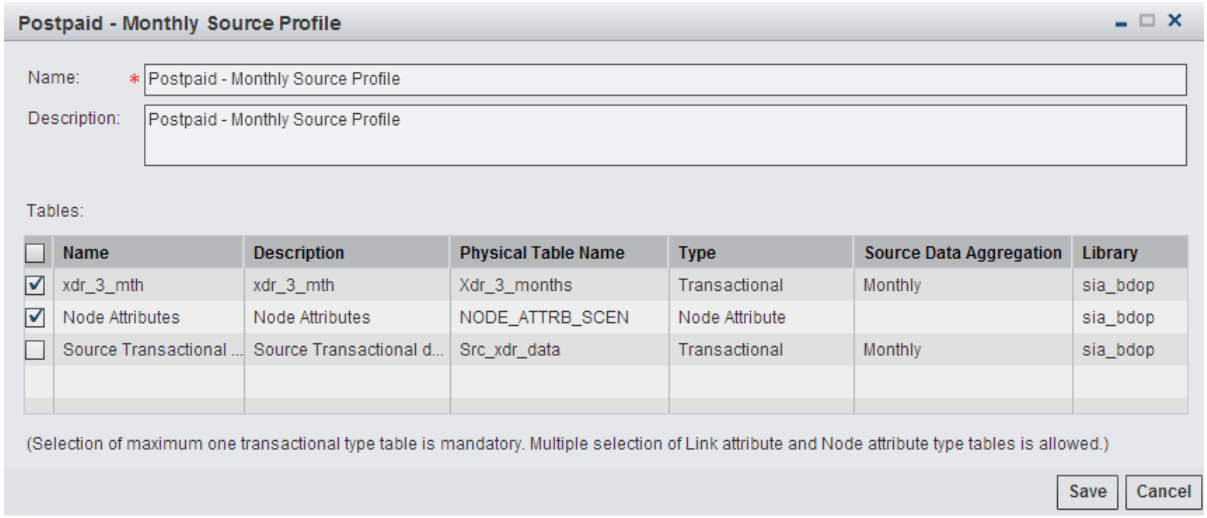
When you create a source data profile, you can view its details in the Properties pane. You can change the name and description of the source data profile. To do so, in the Properties pane, click **Edit**. In addition, you can change the tables that you have included in the source data profile if it is not associated with any project. However, there is no such restriction if you want to include any additional attribute tables in the source data profile.

To change the tables in a source data profile:

1 In the Category pane, select **Source Data Profiles**.

2 Double-click the source data profile whose definition you want to change. The tables that are selected in the definition are displayed along with the other tables that you have imported.

Display 16.2 *Source Data Profile Definition*



3 Change the name and description of the source data profile, if required.

4 You can make the following changes to the table selection based on whether a source data profile is associated with a project:

TIP On the toolbar, if the is deactivated, it indicates that the source data profile is associated with a project.

- Deselect the existing transactional table and select a new one. You can do so only if the data source profile is not associated with any project.
- Deselect an existing attribute table and select a new one. You can do so only if the data source profile is not associated with any project.
- Select one or more attribute tables.

5 Click **Save**.

Delete a Source Data Profile

You can delete a source data profile if it is not associated with a project. For details, see “Create a Project” on page 26.

To delete a source data profile:

1 In the Administration workspace, select **Source Data Profiles** in the Category pane.

Display 16.3 Source Data Profiles

2 Select the source data profile that you want to delete.

3 On the toolbar, select .

Note: If the icon is deactivated, it indicates that the source data profile is associated with a project. Therefore, you cannot delete this source data profile.

Part 5

Viral Effect Analysis

17

Introduction to Viral Effect Analysis

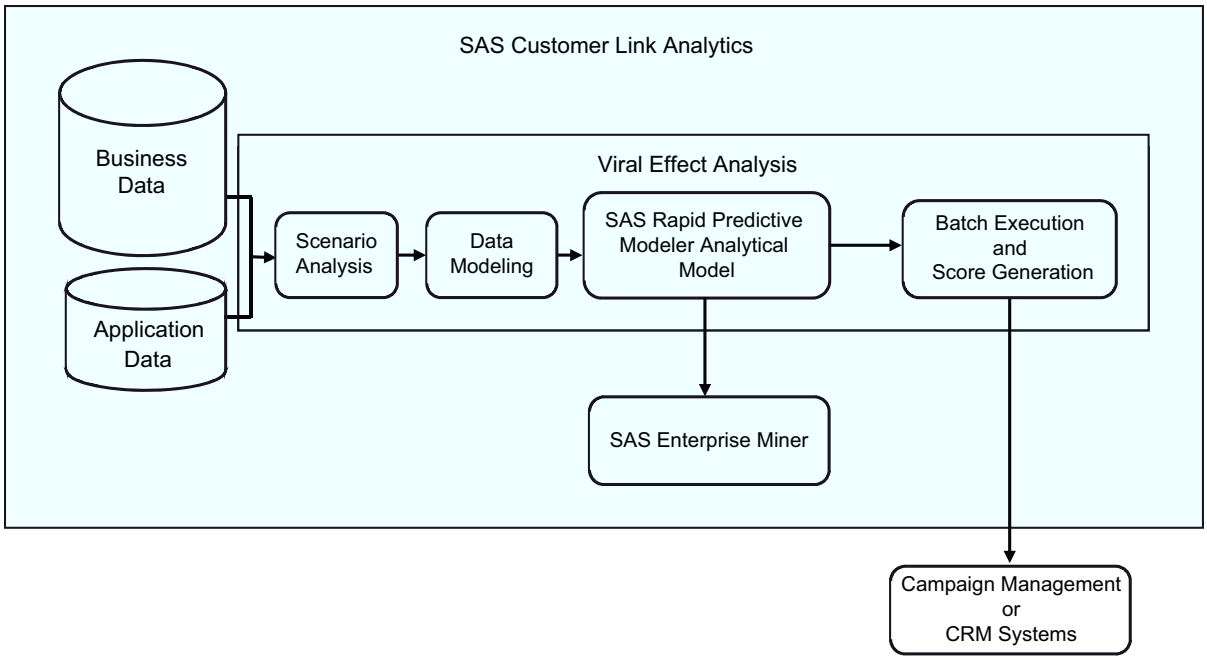
Overview of Viral Effect Analysis

Viral effect analysis is a marketing technique that uses existing networking services and other technologies to achieve marketing objectives such as increased product sales and reduced customer churn. The ultimate goal of marketers is to increase sales by using information about links formed by direct interactions between consumers.

SAS Customer Link Analytics provides you a step-by-step process to perform viral churn analysis. It helps you identify at-risk or recently churned customers and to calculate their impact on the existing customer base. It also enables you to calculate potential revenue impacts and determine appropriate business strategies to reduce customer churn.

Solution Flow of Viral Effect Analysis

Figure 17.1 Viral Effect Analysis Solution Flow



The solution flow of viral effect analysis in SAS Customer Link Analytics includes the following steps:

1. Use the output that a SAS Customer Link Analytics project produces and define an appropriate scenario for that project. The scenario indicates the purpose of your viral effect analysis. For example, the objective of your scenario might be to study the impact of churn of leader nodes on the rest of the community.
2. Prepare modeling data based on project data that is available in both application data and business data. This data contains information about roles and communities of the project that is associated with the scenario. It also contains transaction data and event of analysis data.
3. Build an analytical model by using the SAS Rapid Predictive Modeler macro. You can import this model in SAS Enterprise Miner.
4. Schedule a batch run and generate analytical scores. These scores can be further explored in marketing automation tools such as campaign management or customer relationship management (CRM) systems.

   Note: The subsequent topics explain the analytical flow for viral churn analysis. However, you can build a similar viral analysis model according to your business needs.

Analytical Flow for Viral Churn Analysis

SAS Customer Link Analytics provides you a step-by-step analytical flow to analyze viral churn. The analytical flow for viral churn analysis includes the following steps:

1 **Project Selection**

Select the project that you want to use for your analysis. The project that you select must have successfully completed at least one run in batch mode. For details, see Chapter 12, “Batch Processing,” on page 71 and “Select a Project” on page 108.

2 **Scenario Definition**

Define a scenario to indicate the business objective of your analysis. For example, you can define a scenario to reduce viral churn or to increase product uptakes. The scenario that you define is uniquely identified with a sequence number. This sequence number is used in the subsequent steps of the analytical flow. For details, see “Create a Scenario” on page 109.

3 **Parameter Update**

Specify appropriate values for the parameters that are required for the analysis. When you define a scenario, default values are configured for the parameters that are significant for the analysis. You need to specify your own values to meet your business requirements. For details, see “Update Parameter Values” on page 109.

4 **ABT Building**

Define the analytical base table (ABT). The ABT contains all the variables that are required for the analysis. These variables are categorized as node-level and community-level variables, role-level link variables, churn-level link variables, and node attributes. For details, see “Build ABT” on page 111.

5 **Model Building**

Build the analytical model. As the target event of churn analysis is binary (churned or not churned), a predictive model is built for viral analysis. The model is built by using the SAS Rapid Predictive Modeler macro. As a result, a SAS Enterprise Miner project workspace is created for the model. For details, see “Build the Analytical Model” on page 111.

6 **Model Registration**

Register the model in the metadata. After you register the model in the metadata, you can extract the metadata attributes of the model. For details, see “Register the Model” on page 111.

7 **Model Capture**

Capture and store the model’s information in application data. For details, see “Capture Model Information” on page 112.

8 **Modeling-Time Scoring**

With the latest information that is available at the time of model building, predict the event that will occur in the next time period. For details, see “Perform Modeling-Time Scoring” on page 112.

9 **Scenario Publishing**

Publish the scenario for the scoring run (also called the *batch run*). For details, see “Publish a Scenario” on page 113.

Prerequisites for Viral Churn Analysis

Before you start working with the analytical flow of viral churn analysis, make sure that you complete the following tasks:

- Your administrator has provided you with the user ID and password to log on to SAS Stored Process Web Application.
- You are familiar with the concept of building an analytical model.
- Your user ID should have metadata WRITE permission.

18

Working With Viral Effect Analysis

Log On to SAS Stored Process Web Application

To log on to SAS Stored Process Web Application:

1. In your browser window, type or paste the URL that your administrator has provided you. For example, the URL can be `http://server1.com/SASStoredProcess/do`. The logon window appears.
2. In the **User ID** field, enter your user ID.
3. In the **Password** field, enter the password for the user ID that you have specified.
4. Click **Log On**. The application window appears.
5. Click **List Available Stored Processes and Reports**. The list of stored processes is displayed in the left pane.
6. Expand the appropriate folders depending on the stored process that you want to run.

Select a Project

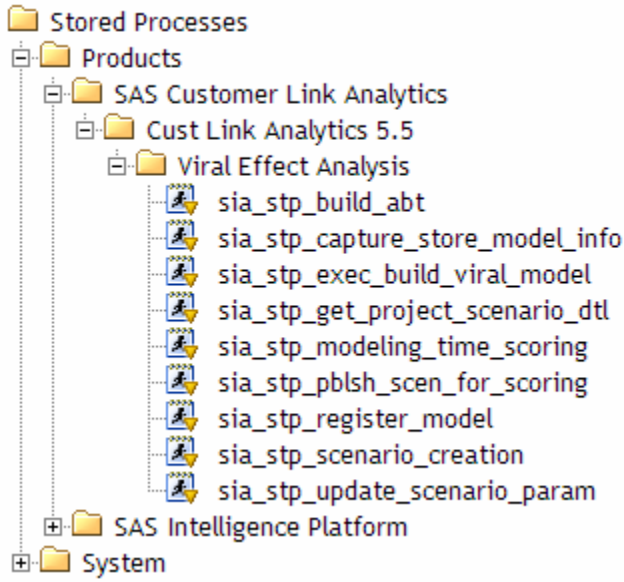
In order to analyze viral churn, you have to define a scenario. A scenario is associated with a project that you have defined in the Projects workspace. Moreover, at least one batch run must be complete for that project.

To associate a project with a scenario, you must know the unique sequence number that is assigned to the project. The **Project ID** that is displayed in the Properties pane of the Projects workspace is the unique sequence number of the project. Thus, you can retrieve the project ID from the Projects workspace. Alternatively, you can run the following steps to retrieve the project ID.

To retrieve the sequence number of a project:

1 In SAS Stored Process Web Application, expand **Stored Processes** ▸ **Products** ▸ **SAS Customer Link Analytics** ▸ **Cust Link Analytics 5.5** ▸ **Viral Effect Analysis**.

Display 18.1 *Stored Processes for Viral Effect Analysis*



2 Click **sia_stp_get_project_scenario_dtl**.

3 In the right pane, enter the following details:

Project
Select the project from the list to indicate that you want to retrieve the sequence number of a project.

Name
Enter the name of the project for which you want to perform viral churn analysis. Make sure that the project has completed at least one run in batch mode.

4 Click **OK**. The unique ID that is assigned to the project is displayed, along with other project details. You need to use this unique ID as the project key in the subsequent analytical tasks.

Create a Scenario

A scenario indicates the purpose of your analysis. For example, you might want to define a scenario to analyze the impact of viral churn.

To create a scenario:

1 In the SAS Stored Process Web Application window, expand **Stored Processes** ▸ **Products** ▸ **SAS Customer Link Analytics** ▸ **Cust Link Analytics 5.5** ▸ **Viral Effect Analysis**.

2 Click the **sia_stp_scenario_creation** stored process. The parameters that are defined for this stored process are displayed in the right pane.

3 Specify the values for the following parameters:

Project PK
Enter the unique ID of the project for which you want to perform the viral churn analysis. The scenario that you are creating is associated with this project.

Scenario name
Enter the name of the scenario that you are creating.

Note: The only special character that you can use for the scenario name is an underscore (_).

Scenario description
Enter a description of the scenario.

4 Click **Run**. A unique sequence number is generated for the scenario. This sequence number uniquely identifies the scenario. It will be used in the subsequent analytical tasks. Also, this process creates a copy of the parameters with their default values. These parameters are required for subsequent analysis.

Update Parameter Values

After you define a scenario, you have to specify parameter values.

To update the parameter values:

1 In the SAS Stored Process Web Application window, expand **Stored Processes** ▸ **Products** ▸ **SAS Customer Link Analytics** ▸ **Cust Link Analytics 5.5** ▸ **Viral Effect Analysis**.

2 Click the **sia_stp_update_scenario_param** stored process. The parameters that are defined for this stored process are displayed in the right pane.

3 Specify values for the following parameters:

Scenario primary key
Enter the unique number of the scenario.

Node attribute library name
Specify the library reference under which the node attribute table is registered in metadata.

Node attribute table name
Specify the name of the node attribute table. Your administrator must have imported this table and configured its columns. For details, see Chapter 15, “Working with Tables,” on page 85.

Note: Your administrator might have configured multiple node attribute tables as part of a data source profile for a given project. However, you need to specify the table that stores information about the event under consideration, such as different nodes and their event dates.

Event date column name of node attribute table
Specify the column name that stores the event date. For example, if the event under consideration is churn analysis, this column stores the churn date.

Note: Make sure that the length of the column name that you specify does not exceed 22 characters. Also, it should be a valid SAS column name.

Unique node identifier column of node attribute table
Specify the column name from the node attribute table that you have selected for the **Node attribute table name**. This column identifies each unique node. For example, the column name can be subscription_id, customer_id, node_id, and so on.

RPM Model type
Select the type of analytical model that you want to build by using SAS Rapid Predictive Modeler.

EMiner project location
Specify the path where you want to create the workspace for the SAS Enterprise Miner project.

Note:

- The only special character that you can use in the project location that you specify is an underscore (_). If you use any other special character, then the model-building process will result in an error.
- Make sure that the length of the path that you specify does exceed 100 characters.

Observation window length
Enter the length of the observation window in days. In this time period, nodes are observed for their churn behavior.

Performance window length
Enter the length of the performance window in days. In this time period, the impact of nodes (that have churned in the observation window) on the remaining nodes is observed. The impact that is considered in terms of whether the remaining nodes churned in the performance window is analyzed in this window.

Modeling ABT name
Enter the name of the modeling ABT that you want to create. The name that you enter must not exceed 25 characters. When the ABT-building stored process is executed, a physical table with the name that you provide is created.

Note: The modeling ABT name that you specify must be a valid table name depending on whether the business data table is stored in SAS or Teradata.

4 Click **Run**. The default values of the parameters are updated with the values that you have entered here.

Build ABT

To build an ABT:

1 In the SAS Stored Process Web Application window, expand **Stored Processes** ▸ **Products** ▸ **SAS Customer Link Analytics** ▸ **Cust Link Analytics 5.5** ▸ **Viral Effect Analysis**.

2 Select the **sia_stp_build_abt** stored process. The parameters that are defined for this stored process are displayed.

3 Enter the sequence number of the scenario for which you want to build the ABT. To retrieve the sequence number of the scenario, run the **sia_stp_get_project_scenario_dtl** stored process and provide the scenario name for which you want to build the ABT.

4 Click **Run**. The ABT is built in the specified location.

Build the Analytical Model

To build the analytical model:

1 In the SAS Stored Process Web Application window, expand **Stored Processes** ▸ **Products** ▸ **SAS Customer Link Analytics** ▸ **Cust Link Analytics 5.5** ▸ **Viral Effect Analysis**.

2 In the left pane, select the **sia_stp_exec_build_viral_model** stored process.

3 Specify the sequence number of the scenario that you want to execute. If you are not sure of the scenario ID, you can confirm it by running the **sia_stp_get_project_scenario_dtl** stored process.

4 Click **Run**. A SAS Enterprise Miner project workspace is created.

Register the Model

To register the model:

1 In the SAS Stored Process Web Application window, expand **Stored Processes** ▸ **Products** ▸ **SAS Customer Link Analytics** ▸ **Cust Link Analytics 5.5** ▸ **Viral Effect Analysis**.

2 In the left pane, select the **sia_stp_register_model** stored process.

3 Specify the sequence number of the scenario that you want to execute. If you are not sure of the scenario ID, you can confirm it by running the **sia_stp_get_project_scenario_dtl** stored process.

4 Click **Run**. In the metadata tree, the model is registered at the following location: `User Folders/<User name>/My Folder`. In this path, *user name* indicates the user name that your administrator has provided while configuring your user ID in SAS Management Console.

Capture Model Information

To capture model information:

1 In the SAS Stored Process Web Application window, expand **Stored Processes** ▸ **Products** ▸ **SAS Customer Link Analytics** ▸ **Cust Link Analytics 5.5** ▸ **Viral Effect Analysis**.

2 In the left pane, select the **sia_stp_register_model** stored process.

3 Specify the unique ID of the scenario that you want to execute. If you are not sure of the scenario ID, you can confirm it by running the **sia_stp_get_project_scenario_dtl** stored process.

4 Click **Run**. The application data is updated with the metadata ID of the model, and the score code of the model is stored in the following location: `<SAS configation path>/Lev1/AppData/SASCustomerLinkAnalytics/CustLinkAnalytics5.5/scenario/modelscorecode`.

Note: At any point of time, only a single model is valid for a given scenario.

Perform Modeling-Time Scoring

In the modeling ABT phase, the impact of nodes that churn in the observation window is analyzed on the remaining nodes that are available in the performance window. During modeling-time scoring, you can use this information to analyze how many nodes churn in the subsequent time window, assuming that some nodes churn in the performance window.

To perform modeling-time scoring:

1 In the SAS Stored Process Web Application window, expand **Stored Processes** ▸ **Products** ▸ **SAS Customer Link Analytics** ▸ **Cust Link Analytics 5.5** ▸ **Viral Effect Analysis**.

2 In the left pane, select the **sia_stp_modeling_time_scoring** stored process.

3 Specify the unique ID of the scenario that you want to execute. If you are not sure of the scenario ID, you can confirm it by running the **sia_stp_get_project_scenario_dtl** stored process.

4 Click **Run**.

Publish a Scenario

To publish a scenario:

1 In the SAS Stored Process Web Application window, expand **Stored Processes** ▸ **Products** ▸ **SAS Customer Link Analytics** ▸ **Cust Link Analytics 5.5** ▸ **Viral Effect Analysis**.

2 In the left pane, select the **sia_stp_pblsh_scen_for_scoring** stored process.

3 Specify the unique ID of the scenario that you want to execute. If you are not sure of the scenario ID, you can confirm it by running the **sia_stp_get_project_scenario_dtl** stored process.

4 Click **Run**. A batch code file is generated at the following location: `<SAS configuration path>/Lev1/AppData/SASCustomerLinkAnalytics/CustLinkAnalytics5.5/scenario/batchcode`. Your administrator can schedule this file for scoring. For details, see *SAS Customer Link Analytics: Administrator's Guide*.

Part 6

Appendixes

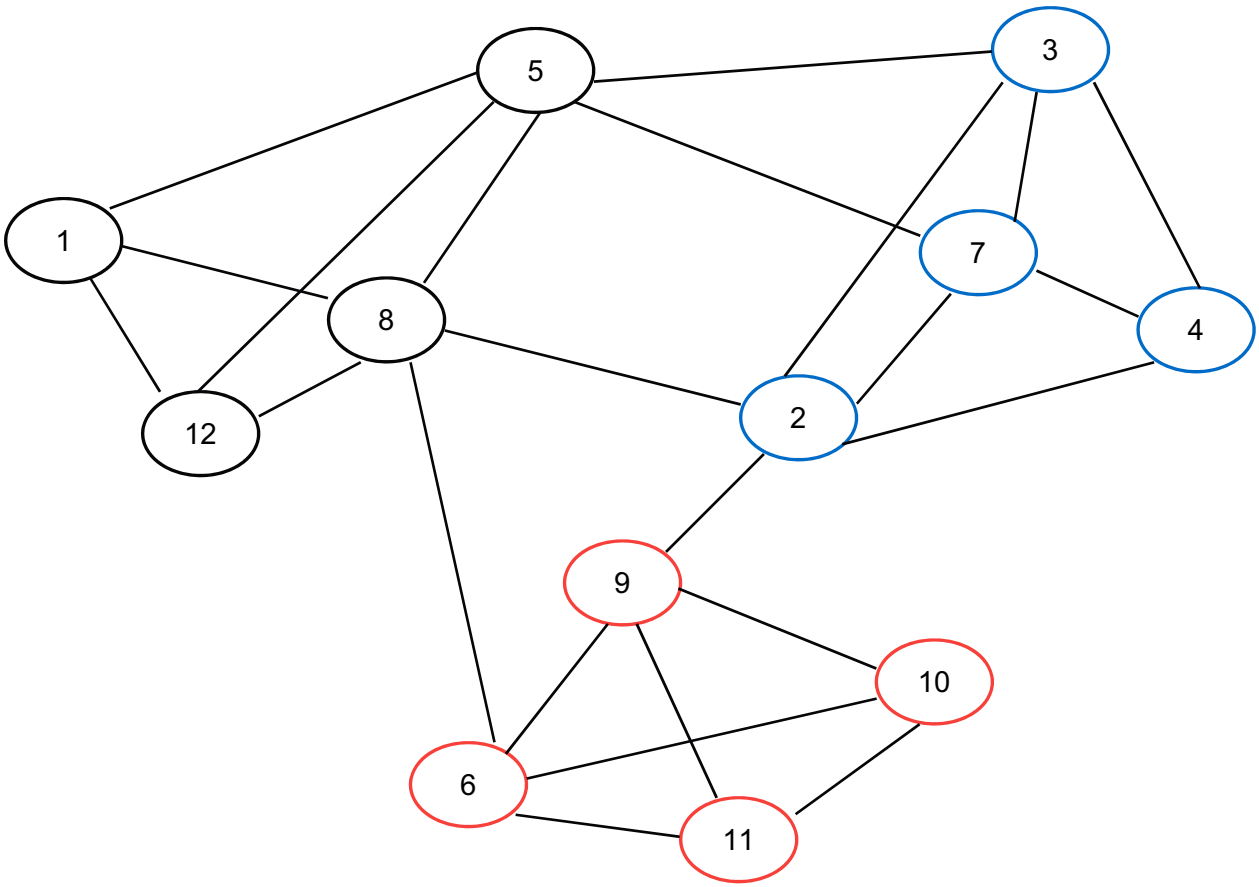
Appendix 1

Common Network Analysis Concepts

About Network Analysis

Network analysis examines the structure of relationships between entities. These entities can be persons, groups, organizations, web sites, and so on. One of the most important objectives of network analysis is the detection of cohesive and self-contained structures that are called *communities*. These communities are defined intuitively as groups of nodes that are more tightly connected to each other than they are to the rest of the network. The diagrammatic representation of communities within a network is called a *network diagram*.

Figure A1.1 *Network Diagram*

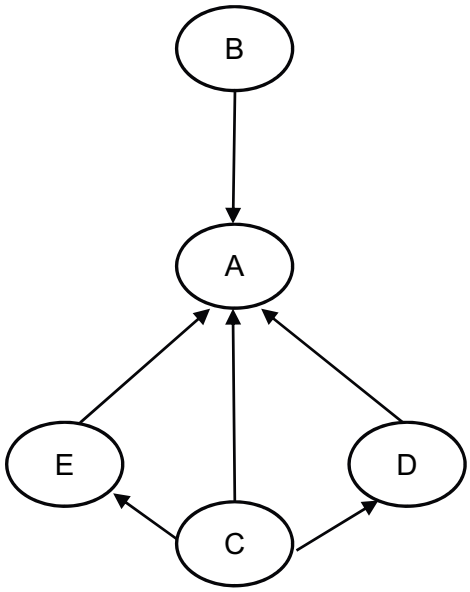


In a network diagram, nodes represent the entities that are connected and links represent the connections between nodes. Each link connects two nodes. However, a node can have multiple connecting links. The nodes and the links together constitute a community. One or more such communities forms a network. In a communications network, a node typically indicates a calling or a called number and a link represents the communication that is initiated between two nodes. The communication can be through different type of services, such as voice calls, messages, data upload, or data download.

Directed and Undirected Graphs

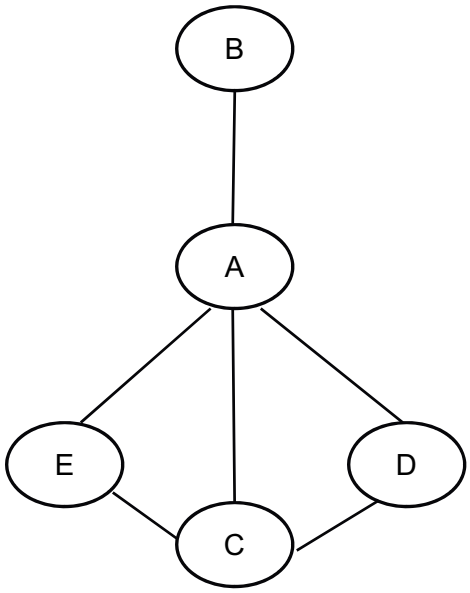
A network diagram in which each link has a sense of direction from one node to another is called a *directed graph*. Each link indicates whether it is an incoming or outgoing link for a node. However, in an undirected graph, a link has no sense of direction.

Figure A1.2 *Directed Graph*



For example, in the network that is represented in Figure A1.2, assume that the nodes represent people at a fan club party. There is a link from one node to another when one person knows another. A has the maximum number of incoming links. These links indicate that a lot of people know person A, but the reverse is not true. Similarly, node C has the maximum number of outgoing links. These links indicate that person C knows a lot of people, but the reverse is not true. These types of links indicate that one person knowing another person does not necessarily imply that the reverse is also true. These types of relationships can exist between celebrities and fans. Many fans might know of a celebrity. However, it is unlikely that celebrities know all their fans.

Figure A1.3 Undirected Graph

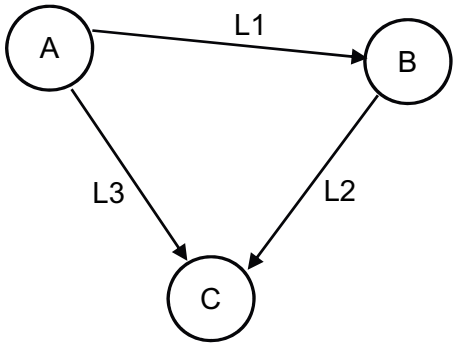


For example, in the network that is represented in Figure A1.3, assume that the nodes represent people at an alumni party. There is a link between two people if they shake hands. In this case, a link between person A and person B indicates that if person A shook hands with person B, then person B also shook hands with person A. These types of links indicate that one person knowing another person necessarily implies that the reverse is also true.

In-Degree and Out-Degree Measures

Consider the following network diagram containing three nodes, A, B, and C, with three links, L1, L2, and L3.

Figure A1.4 In-Degree and Out-Degree Measures



The following table indicates the in-degree and out-degree values for each node.

Table A1.1 In-Degree and Out-Degree Measures

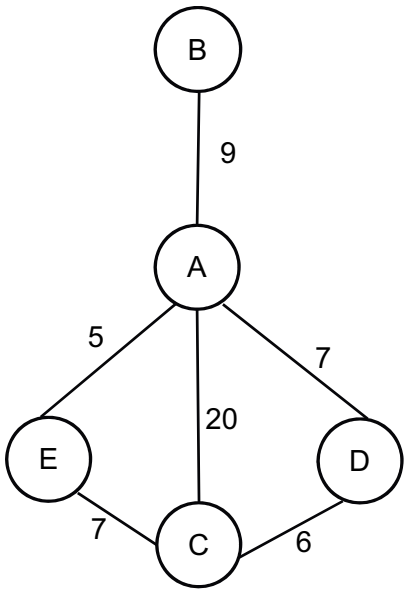
Node Name	In-Degree Measure	Out-Degree Measure
A	0	2

Node Name	In-Degree Measure	Out-Degree Measure
B	1	1
C	2	0

Shortest Path

The shortest path between a pair of nodes of a network is defined as the minimum distance between them. The distance between a pair of nodes can be computed based on the weights of links that connect them or the number of links between them.

Figure A1.5 *Shortest Path*

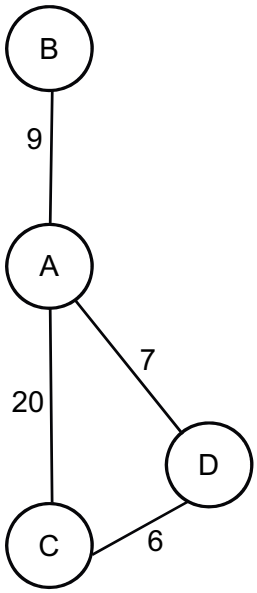


For example, in this figure, consider the shortest path between nodes A and C. Considering the link weights, the shortest path between nodes A and C is A->E->C, with the cumulative sum of weights as 12. Considering the number of links, the shortest path is A->C because the number of links that separate nodes A and C is 1.

Diameter of a Community

The diameter of a community is computed as the longest of all the shortest paths between all the pairs of a community.

Figure A1.6 *Diameter of a Community*



For example, consider that this figure represents a community. The following table lists the shortest paths for each pair of nodes.

Table A1.2 *Shortest Paths*

Node Pair	Shortest Path Considering Cumulative Weights	Shortest Path Considering Number of Links
A and B	9	1
A and C	20	1
A and D	7	1
B and C	22	2
B and D	16	2

At the community level, considering the cumulative weights, the diameter is 22, which the is the maximum of all the shortest paths. However, considering the number of links, the diameter is 2.

Appendix 2

Overview of Centrality Measures

Influence Centrality

Influence centrality is a generalization of degree centrality that considers the link and node weights of adjacent nodes (C_1) in addition to the link weights of nodes that are adjacent to adjacent nodes (C_2). The metric C_1 is referred to as the *first-order influence centrality*, and the metric C_2 is referred to as the *second-order influence centrality*.

Let ω_{uv} define the link weight for link (u, v), and let ω_u define the node weight for node u. Let δ_u represent the list of nodes connected to node u (that is, its neighbors); this list is called the *adjacency list*. For directed graphs, the neighbors are the out-links. The general formula for influence centrality is asfollows:

$$C_1(u) = \frac{\sum_{v \in \delta_u} \omega_{uv}}{\sum_{v \in N} \omega_v}$$

$$C_2(u) = \sum_{v \epsilon \delta_u} C_1(v)$$

As the name suggests, this metric gives some indication of the potential influence, performance, or ability to transfer knowledge.

Degree Centrality

The degree of a node v in an undirected graph is the number of links that are node v. The out-degree of a node in a directed graph is the number of out-links incident to that node; the in-degree is the number of in-links incident to that node. The terms degree and out-degree are interchangeable for an undirected graph. Degree centrality is simply the (in- or out-) degree of a node and can be interpreted as some form of relative importance to a network. For example, consider a network where nodes are people and you are tracking the flow of a virus. In this case, the degree centrality gives some idea of the magnitude of the risk of spreading the virus. People with a higher out-degree can lead to a quicker and more widespread transmission. In a friendship network, in-degree often indicates popularity.

Closeness Centrality

Closeness centrality is the reciprocal of the average of the shortest paths (geodesic distances) to all other nodes. Closeness can be thought of as a

measure of how long it would take information to spread from a given node to other nodes in the network.

The general formula for the closeness centrality is as follows:

$$C_C(u) = \frac{|C| - 1}{\sum_{v \in N \setminus u} d_{uv}}$$

In this equation, C is the component that contains u and d_{uv} is the shortest path from node u to node v.

Betweenness Centrality

Betweenness centrality counts the number of times a particular node (or link) occurs on the shortest paths between other nodes. Betweenness can be thought of as a measure of the control a node (or link) has over the communication flow among the rest of the network. In this sense, the nodes (or links) with high betweenness are the gatekeepers of information because of their relative location in the network.

The formula for node betweenness centrality is as follows:

$$C_b(u) = \sum_{\substack{s \neq u \neq t \in N \\ s \neq t}} \frac{\sigma_{st}(u)}{\sigma st}$$

In this equation, σ_{st} is the number of shortest paths from s to t and $\sigma_{st}(u)$ is the number of shortest paths from s to t that pass through node u.

The formula for link betweenness centrality is as follows:

$$C_b(u, v) = \sum_{\substack{s,t \in N \\ s \neq t}} \frac{\sigma_{st}(u, v)}{\sigma st}$$

In this equation, $\sigma_{st}(u, v)$ is the number of shortest paths from s to t that pass through link (u, v).

Eigenvector Centrality

Eigenvector centrality is an extension of degree centrality, in which centrality points are awarded for each neighbor. However, not all neighbors are equally important. Intuitively, a connection to an important node should contribute more to the centrality score than a connection to a less important node. This is the basic idea behind eigenvector centrality. Eigenvector centrality of a node is defined to be proportional to the sum of the scores of all nodes that are connected to it. Mathematically, it is represented as follows:

$$x_i = \frac{1}{\lambda} \sum_{j \in \delta_i} x_j = \frac{1}{\lambda} \sum_{j \in N} A_{ij} x_j$$

In this equation, x_i is the eigenvector centrality of node i, λ is constant, δ_i is the set of nodes that connects to node i, and A_{ij} is the weight of the link from node i to node j.

Eigenvector centrality can be written as an eigenvector equation in matrix form as

$$Ax = \lambda x$$

As can be seen from the preceding equation, *x* is the eigenvector and λ is the eigenvalue. Because *x* should be positive, only the principal eigenvector that corresponds to the largest eigenvalue is of interest.

Hub and Authority Centralities

Hub and authority centralities were originally developed by Kleinberg (1998) to rank the importance of web pages. Certain web pages are important because they point to many important pages (these are called *hubs*). On the other hand, some web pages are important because they are linked by many important pages (called *authorities*). In other words, a good hub node is one that points to many good authorities. Similarly, a good authority node is one that is pointed to by many good hub nodes. This idea can be applied to many other types of graphs besides web pages. For example, it can be applied to a citation network for journal articles. A review article that cites many good authority papers has a high hub score, whereas a paper that is referenced by many other papers has a high authority score.

The authority centrality of a node is proportional to the sum of the hub centrality of nodes that point to it. Similarly, the hub centrality of a node is proportional to the sum of the authorities of nodes that it points to.

That is,

$$x_i = \alpha \sum_{j \in N} A_{ij} y_j$$
$$y_i = \beta \sum_{j \in N} A_{ji} y_{xj}$$

In this equation, x_i is the authority centrality of node i , y_i is the hub centrality of node i , Aij is the weight of the link from node i to node j , and α and β are constants.

The definition can be written in matrix form as follows:

$AA^T x = \lambda x$

$A^T y = \lambda y$

Thus, the authority and hub centralities are the principal eigenvectors of $A^T A$ and AA^T, respectively.

Clustering Coefficient

The clustering coefficient for a node is the number of links between the nodes within its neighborhood divided by the number of links that could possibly exist between them.

Let δ^u represent the list of nodes that are connected to node *u*. The formula for the clustering coefficient is as follows:

$$C(i) = \frac{\left| \{(u, v) \in A : u, v \in \delta_i\} \right|}{|\delta_i| (|\delta_i| - 1)}$$

For a particular node *i* , the clustering coefficient determines how close to being a clique (complete subgraph) the subgraph induced by itself and its neighbor set δ_i are. In social networks, a high clustering coefficient can help predict relationships that might not be known, confirmed, or realized yet. The fact that person A knows person B and person B knows person C does not guarantee

that person A knows person C. However, it is much more likely that person A knows person C than that person A knows some random person.

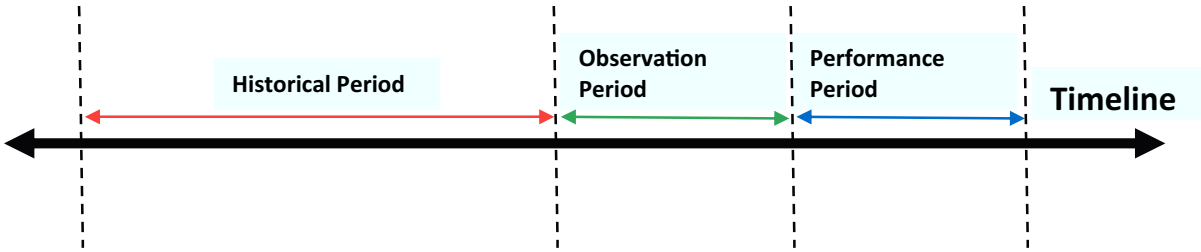
Appendix 3

Performance and Observation Periods

Modeling

In the modeling phase, ABT variables are built by considering the entire historical period. In the observation period, the event under consideration is observed and input variables are computed based on these observations. In the performance period, the target variable is computed as an impact of the occurrence of the event under consideration in the observation window.

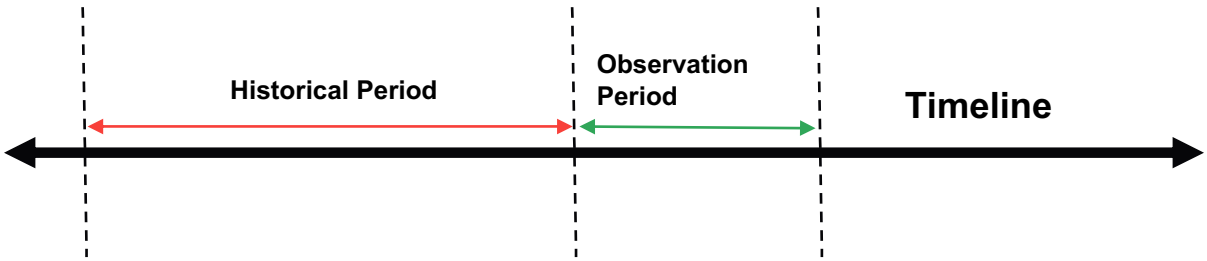
Figure A3.1 *Time Windows: Modeling*



Scoring

In the scoring phase, in the observation period the event under consideration is observed and input variables are computed based on these observations. The impact of the occurrence of the event in the observation period is predicted for the subsequent time period. In this phase, no target variable is defined or computed. Therefore, there is no performance period.

Figure A3.2 *Time Windows: Scoring*



Glossary

ABT variable
See analytical base table variable

analytical base table
a highly denormalized data structure that is designed to build an analytical model or to generate scores based on an analytical model.

analytical base table variable
a column in an analytical base table that is used to build a statistical model to predict defaults.

analytical model
a statistical model that is designed to perform a specific task or to predict the probability of a specific event.

box plot
a graphical display of five statistics (the minimum, lower quartile, median, upper quartile, and maximum) that summarize the distribution of a set of data. The lower quartile (25th percentile) is represented by the lower edge of the box, and the upper quartile (75th percentile) is represented by the upper edge of the box. The median (50th percentile) is represented by a central line that divides the box into sections. The extreme values are represented by whiskers that extend out from the edges of the box.

centrality measure
in graph theory and network analysis, a factor that indicates the relative importance of a vertex within a graph. A few examples of centrality measures are Degree, Closeness, Betweenness, and Eigenvector.

community
a group of nodes in a network that are more densely connected internally than with the rest of the network. A network can contain one or more communities.

geodesic
the shortest distance between a pair of nodes.

link
in a network diagram, a line that represents a relationship between two nodes.

model scoring
the process of applying a model to new data in order to compute outputs.

network
a collection of one or more communities.

node
in a network diagram, a dot or point that represents an individual actor within the network.

outlier
a data point that differs from the general trend of the data by more than is expected by chance alone. An outlier might be an erroneous data point or one that is not from the same sampling model as the rest of the data.

project
the named collection of activities and reports to implement a business strategy for addressing a business pain. For example, a project can be created for reducing churn of highly profitable customers in the North region.

quartile
any of the three points that divide the values of a variable into four groups of equal frequency, or any of those groups. The quartiles correspond to the 25th percentile, the 50th percentile (or median), and the 75th percentile.

scoring
See model scoring

symmetric community
a community in which each central node is symmetrically linked to other nodes of the community.

transactional data
timestamped data collected over time at no particular frequency. Some examples of transactional data are point-of-sale data, inventory data, call center data, and trading data.

whisker
a vertical line on a box plot that represents values larger than the third quartile or smaller than the first quartile but within 1.5 interquartile ranges of the box.

workflow
a series of tasks, together with the participants and the logic that is required to execute the tasks. A workflow includes policies, status values, and data objects.

workflow diagram
a diagram that indicates the order in which activities of a project are to be performed.

workflow step
each individual activity of a project that is depicted in a workflow diagram.

Index

S

T

V

W