SAS® Customer Link Analytics 6.5: User's Guide

SAS® Documentation

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2016. *SAS® Customer Link Analytics 6.5: User's Guide*. Cary, NC: SAS Institute Inc.

SAS® Customer Link Analytics 6.5: User's Guide

SAS Institute Inc., SAS Campus Drive, Cary, NC 27513-2414

June 2016

Contents

Using This Book

Audience

This document explains 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. In addition, they can enrich the project's output data, load data to the SAS Customer Link Analytics LASR Analytic Server, and create and view the Community report.
- business users who can view all the projects, load data to the SAS Customer Link Analytics LASR Analytic Server, and create and view the Community report.

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.

Document Conventions

Document Conventions

Convention	Descritpion
<SAS configuration directory>	Represents the path to the folder where SAS configuration data is stored. For example, on a Windows computer, this path can be C:/SAS/Config. For more information, see **SAS Customer Link Analytics Administrator's Guide**.

Convention	Descritpion
<Project path>	Represents the path to the folder where the project's data is stored. This path is configured as a software component property. For more information, see **SAS Customer Link Analytics Administrator's Guide**.

What's New

What's New in SAS Customer Link Analytics 6.5

Overview

SAS Customer Link Analytics has the following enhancements:

- community segmentation
- support for production adoption events in viral effect analysis
- reporting data for viral effect analysis
- pulling a scenario into design mode

Community Segmentation

When you configure the Centrality Measures Computation workflow step, SAS Customer Link Analytics enables you to perform an additional sub-step, community segmentation. *Community segmentation* groups the communities into various segments so that communities within each segment have similar patterns for the transactional measures. SAS Customer Link Analytics uses the statistical clustering procedure to divide communities into segments. Each segment represents a group of communities that have homogenous behavior for centrality measures, transactional measures, and link weights. Conversely, there is heterogeneity across segments.

In addition, SAS Customer Link Analytics generates a Community Segmentation report that shows the distribution of communities across various segments. You can view this report in SAS Visual Analytics Viewer in a seamless manner by using the Flex application switcher.

Support for Production Adoption Events in Viral Effect Analysis

In addition to customer churn, SAS Customer Link Analytics enables you to perform step-by-step viral effect analysis for the production adoption event. SAS Customer Link Analytics predicts the probability of a node to adopt a particular product based on the connections that it has with nodes that have already adopted this product. As a result, SAS Customer Link Analytics helps business analysts define appropriate business strategies to increase product adoption.

Reporting Data for Viral Effect Analysis

SAS Customer Link Analytics enables you to generate reporting data for viral effect analysis and load it to the SAS Customer Link Analytics LASR Analytic Server. You can further explore this data in SAS Visual Analytics tools.

Moving a Scenario into Design Mode

If the scoring run of the scenario does not produce results according to your business requirements, you might want to rebuild the analytical model for viral effect analysis. Also, if the configuration of the project that is associated with the scenario changes, you might want to rework the tasks of the modeling run. For all these scenarios, SAS Customer Link Analytics enables you to pull the scenario back into design mode.

Accessibility

Accessibility Notice

For more information about the accessibility of this product, see Accessibility Features of SAS Customer Link Analytics 6.5 at support.sas.com.

Part 1

Introduction

1

Introduction to 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 customer acquisitions. Also, with the increasing market penetration, 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 customer base, such as how much they interact, with whom they interact, and so on. The strength of relationships within their network and outside their network can reveal more information than static attributes such as their demographic information. 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 network analysts to analyze network data, identify communities within the network, and quantify the relative importance of nodes within a community or network from various aspects. It also enables them to identify the role that each node plays within its community. In addition, network analysts can further enrich these results and produce reports with information about the network's nodes, links, and communities. Using these

reports, business analysts can define appropriate strategies to increase revenue.

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, aggregated transactional data (such as banking transactions; call detail records in a communications network; or likes, comments, and shares between two nodes in a social network) is configured into the SAS Customer Link Analytics application. Therefore, 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.

Data Enrichment

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 data enrichment feature of SAS Customer Link Analytics. SAS Customer Link Analytics produces certain predefined reports based on the enriched data and enables you to view them in the SAS Visual Analytics Viewer. You can further explore the data in suitable tools such as SAS Customer Intelligence to enhance your campaigning strategy.

Viral Effect Analysis

Unlike the traditional predictive modeling that is used in the customer intelligence area of various industries, viral effect analysis emphasizes the interactions between customers 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 the *distributed mode* of operation. Running on a single computer system with multiple CPUs is referred to as the *non-distributed mode* of operation. For more information, 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 network data. It operates in distributed and non-distributed mode. For more information, see *SAS Customer Link Analytics: Administrator's Guide*.

Figure 1.1 Solution Flow Diagram: Non-Distributed Mode

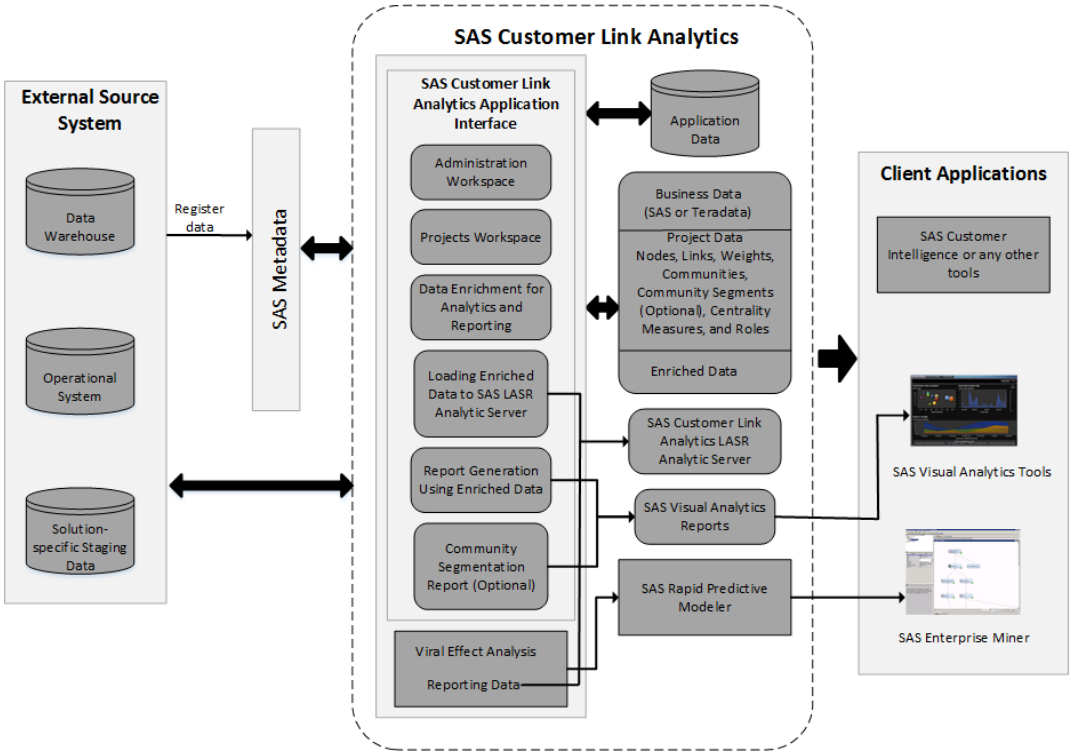
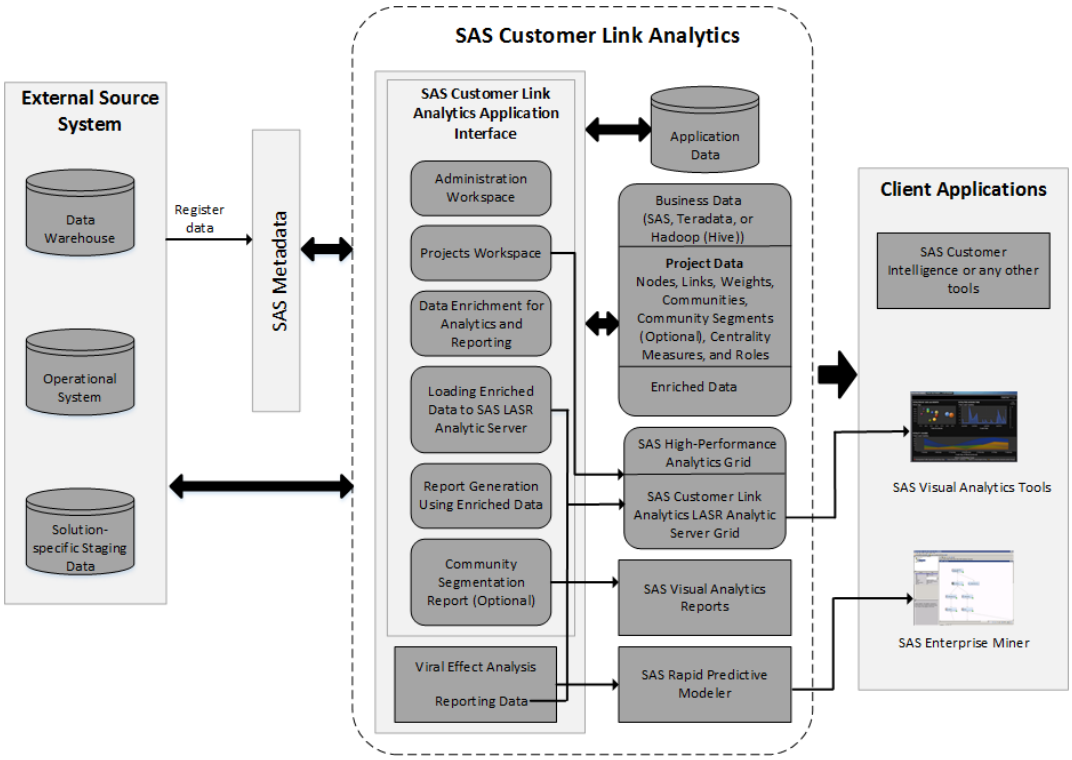


Figure 1.2 Solution Flow Diagram: Distributed Mode



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. In addition, SAS Customer Link Analytics, registers the project's output data, enriched data, viral effect analysis results, and the Community report in the metadata.

SAS Customer Link Analytics interface
: a workflow-based application that enables you to perform the important tasks listed here:

  Administration workspace
  : enables you to configure the metadata that is required for the SAS Customer Link Analytics workflow.

  Projects workspace
  : enables you to define projects, configure and run workflow steps, enrich the project's output data, load the enriched data to the SAS Customer Link Analytics LASR Analytic Server, and view the Community report.

  Data enrichment for analytics and reporting
  : enables you to enrich the project's output data based on predefined categories. SAS Customer Link Analytics uses this data for reporting. Moreover, you can use the enriched data in tools such as SAS Customer Intelligence for further analysis.

  Loading enriched data to the SAS Customer Link Analytics LASR Analytic Server
  : SAS Customer Link Analytics enables you to load the enriched data to the SAS Customer Link Analytics LASR Analytic Server. SAS Customer Link Analytics uses this data to produce the Community report.

  Report generation using enriched data
  : SAS Customer Link Analytics produces the Community report based on the enriched data that it loads into the SAS Customer Link Analytics LASR Analytic Server. It also enables you to view this report in the SAS Visual Analytics Viewer seamlessly by using the Flex application switcher.

  Community Segmentation report (Optional)
  : SAS Customer Link Analytics produces the Community Segmentation report if you choose to create community segments when you run the Centrality Measures Computation workflow step. It also enables you to view this report in the SAS Visual Analytics Viewer seamlessly by using the Flex application switcher.

  Viral effect analysis
  : enables analysis of SAS Customer Link Analytics output, construction of an analytical model using SAS Rapid Predictive Modeler, and generation of analytical scores and reporting data 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.

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.

  Application data also contains the configuration details that are required for running the data enrichment process.

Business data
: stores 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 data that SAS Customer Link Analytics produces when you run the data enrichment process.

  Business data also contains the intermediate tables that SAS Customer Link Analytics creates when the workflow steps are run.

SAS High-Performance Analytics Server Grid
: in distributed mode, provides tools for performing analytic tasks of community building and centrality measure computation in a high-performance environment. This environment is characterized by massively parallel processing (MPP) on a distributed system. For more information, see *SAS Customer Link Analytics: Administrator's Guide*.

SAS Customer Link Analytics LASR Analytic Server Grid
: an analytic platform that provides a secure, multi-user grid environment for concurrent access to enriched data that SAS Customer Link Analytics loads into memory to produce the Community report. In the distributed mode, the server distributes data and the workload among multiple machines and performs massively parallel processing. However, in the non-distributed mode, server is deployed on a single machine as the workload and data volumes do not require a distributed computing environment. In addition, if the business data is stored in Hadoop, SAS Customer Link Analytics runs the analytical processes of Community Building and Centrality Measures Computation workflow steps along-side LASR.

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.

SAS Customer Intelligence or any other similar tools
: a suite of marketing automation tools that enable organizations to manage interactions along the customer journey in a personalized and profitable way. SAS Customer Intelligence provides analytically driven capabilities in the four areas that the modern marketing organization needs in today's digital world: strategy and operations, marketing analytics, multichannel engagement, and digital experience.

  You can use the project's output data and the enriched data that SAS Customer Link Analytics produces for defining target lists or campaigns in SAS Customer Intelligence and taking marketing actions. You can thereby enhance your campaigning strategies.

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 more information, see Chapter 13, “Working with Tables,” on page 101.
   b. Define source data profiles. For more information, see Chapter 14, “Working with Source Data Profiles,” on page 113.
3. Log on to SAS Customer Link Analytics as a network analyst and define a project. For more information, see “Create a Project” on page 23.
4. Complete the workflow steps that are listed here:
   a. Select the nodes and links whose data you want to analyze and then extract summarized transaction data. For more information, see “Configure the Data Extraction Workflow Step” on page 27.
   b. Filter the links based on specific parameters and assign weights to the links. For more information, see “Configure the Link and Node Processing Workflow Step” on page 31.
   c. Build communities by selecting the appropriate analytical approach. For more information, see “Configure the Community Building Workflow Step” on page 39.
   d. Select centrality measures that you want to compute and provide input parameters to compute these measures. You can also generate community segments and view the Community Segmentation report. For more information, see “Configure the Centrality Measures Computation Workflow Step” on page 45.
   e. Assign a role to each node of the communities. For more information, see “Define a Role” on page 57.
5. Promote a project to batch mode. For more information, see Chapter 10, “Batch Processing,” on page 67.
6. Log on to SAS Customer Link Analytics as a network analyst, and complete the following tasks:
   a. Enrich the output data of a project. For more information, see Chapter 11, “Working with Data Enrichment and Data Loading,” on page 71.
   b. Load the node-level enriched data to the SAS Customer Link Analytics LASR Analytic Server. For more information, see “Load Enriched Data to the SAS Customer Link Analytics LASR Analytic Server” on page 81.

      Note: You can also perform this step if you log on as a business user.
   c. Create and view the Community report. For more information, see Chapter 12, “Generating the Community Report,” on page 85.

      Note: You can also view the report if you log on as a business user.
7. (Optional) Perform viral effect analysis. For more information, see Chapter 15, “Introduction to Viral Effect Analysis,” on page 119.

SAS Customer Link Analytics Documentation

In addition to the **Help**, the SAS Customer Link Analytics documentation contains the following documents:

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

These documents are available at the following site: http://support.sas.com/documentation/solutions/clac/index.html. You have to provide the following user name and password to access this site:

User Name: *sas*

Password: *UserDocCLA*

2

Getting Started with 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 the following predefined roles—network analysts, administrators, and business users. 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, enrich the project's output data, load enriched data to the SAS Customer Link Analytics LASR Analytic Server, and create and view the Community report.
Cust Link Analytics: Administration	Provides capabilities to view, create, and delete source data profiles and tables.
Cust Link Analytics: Business User	Provides capabilities to view all projects, load enriched data to the SAS Customer Link Analytics LASR Analytic Server, and create and view the Community report.

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.
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.
Execute Data Enrichment	Enables you to enrich the project's output data.

Capability Name	Description
Load Data on LASR	Enables you to load the enriched data into the SAS Customer Link Analytics LASR Analytic Server.
Create Reports	Enables you to create the Community report for a project.
Open Reports	Enables you to view the Community report of a project.
Administrative capabilities	
Create New Source Profile	Enables you to create a new source data profile.
Delete Source Profile	Enables you to delete a 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.

Logging 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 ID. As a result, this user ID cannot perform the following tasks:

- create a project
- run a workflow step
- change the mode of a project
- refresh data in a transactional table
- enrich a project's output data
- load enriched data to the SAS Customer Link Analytics LASR Analytic Server

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

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

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

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 the *distributed mode* of operation. Running on a single computer system with multiple CPUs is referred to as the *non-distributed mode* of operation. For more information, 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.

Working in 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.

Table 2.3 *Workspaces*

Button	Button Name	Purpose
	Projects	Enables you to define new projects and work on their workflow steps. In addition, it enables you to complete the data enrichment and data loading tasks. You can also create and view the Community report in this workspace.

Button	Button Name	Purpose
	Administration	Enables you to import tables and define source data profiles.

Introduction to Projects

Overview of 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 node. 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*. For more information, see “Create a Project” on page 23.

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. For more information, see “Change the Mode of a Project” on page 26.

Introduction to Workflow Steps

Overview of Workflow Steps

When you define a project, a default workflow is assigned to it. The workflow contains a set of predefined workflow steps.

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 more information, see “Configure the Data Extraction Workflow Step” on page 27.

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 more information, see “Configure the Link and Node Processing Workflow Step” on page 31.

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 more information, see “Configure the Community Building Workflow Step” on page 39.

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.

(Optional) Select the **Create community segments** check box if you want to group communities into various segments. Also, specify appropriate values for the parameters. The clustering procedure produces community segments depending on the values that you specify here.

For more information, see “Configure the Centrality Measures Computation Workflow Step” on page 45.

5 **Role Assignment**

Assign roles based on the centrality measures statistics. You can define roles according to your business requirements. For more information, see “Define a Role” on page 57.

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 2.4 Processing Status of Workflow Steps

Icon	Status	Description
	Successfully executed	The workflow step ran successfully without any errors.

Icon	Status	Description
	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 Options

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 2.5 Toolbar Options

Button	Purpose
	Customizes the default setup of the workflow step. **Note:** When you edit the default setup of 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.
	Reverts the changes that were made to the setup of the workflow step and resets them to the default. For more information, see “Reset a Workflow Step” on page 18.
	Runs the workflow step.

Reset a Workflow Step

When you reset a workflow step, the business data tables that contain the workflow step data for the current run are deleted along with their registered metadata. Similarly, the tables of the current run of the subsequent workflow steps are also deleted. For more information, see "Storing the Workflow Step Data" on page 149. As a result, if you changed the default values for these workflow steps, you will lose them. Therefore, you have to configure these workflow steps and then run them again.

If you have selected the **Create community segments** check box in the Centrality Measures Computation workflow step, certain additional tasks are performed.

- The values that you have specified for **Maximum number of segments** and **Convergence criterion** are not reset to default values.
- The modeling ABT that is created for the project is deleted.
- The clustering model that is currently marked as the active model for the project is deactivated.
- The information that is displayed on the **Community Segmentation Results** tab is also deleted.

Workflow Step Pane

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

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

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 2.6 *Sample Data Details*

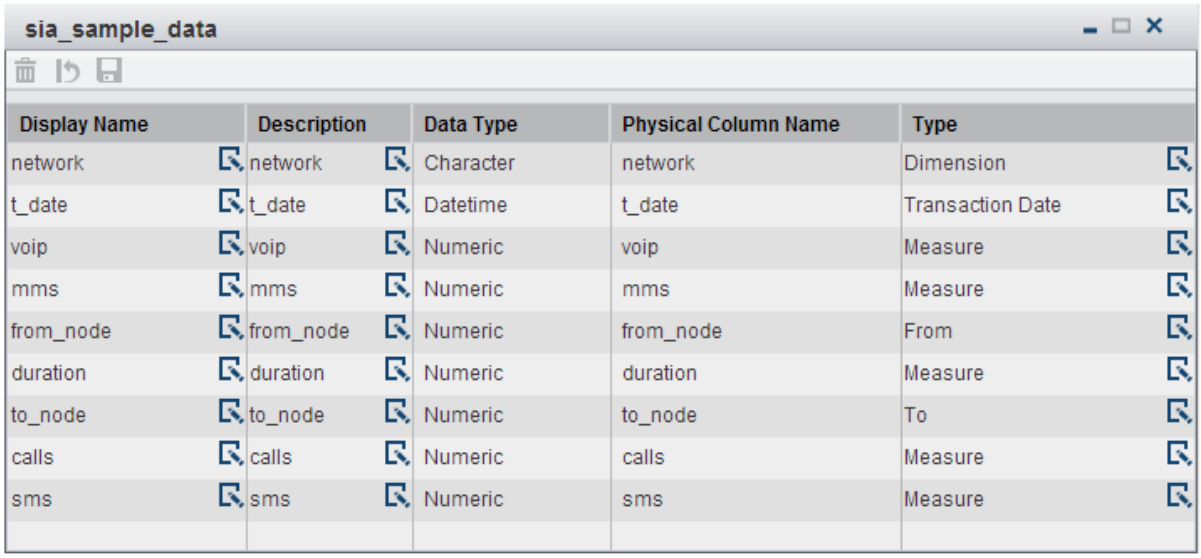
Table name	sia_sample_data

Library name	sia_sample
Metadata location	`/Products/SAS Customer Link Analytics/Cust Link Analytics 6.5/Data Sources/Sample Data`
Physical location	`<SAS configuration directory>/Lev1/AppData/SASCustomerLinkAnalytics/6.5/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 more information, see Chapter 13, “Working with Tables,” on page 101.

Figure 2.1 *Columns of Sample Data Table*



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 more information, see Chapter 14, “Working with Source Data Profiles,” on page 113.

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 more information, see Chapter 3, “Managing Projects,” on page 23.

After you define a project, you can begin working on its workflow steps. For more information, see “Introduction to Workflow Steps” on page 15.

Part 2

Projects Workspace

3

Managing Projects

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.

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

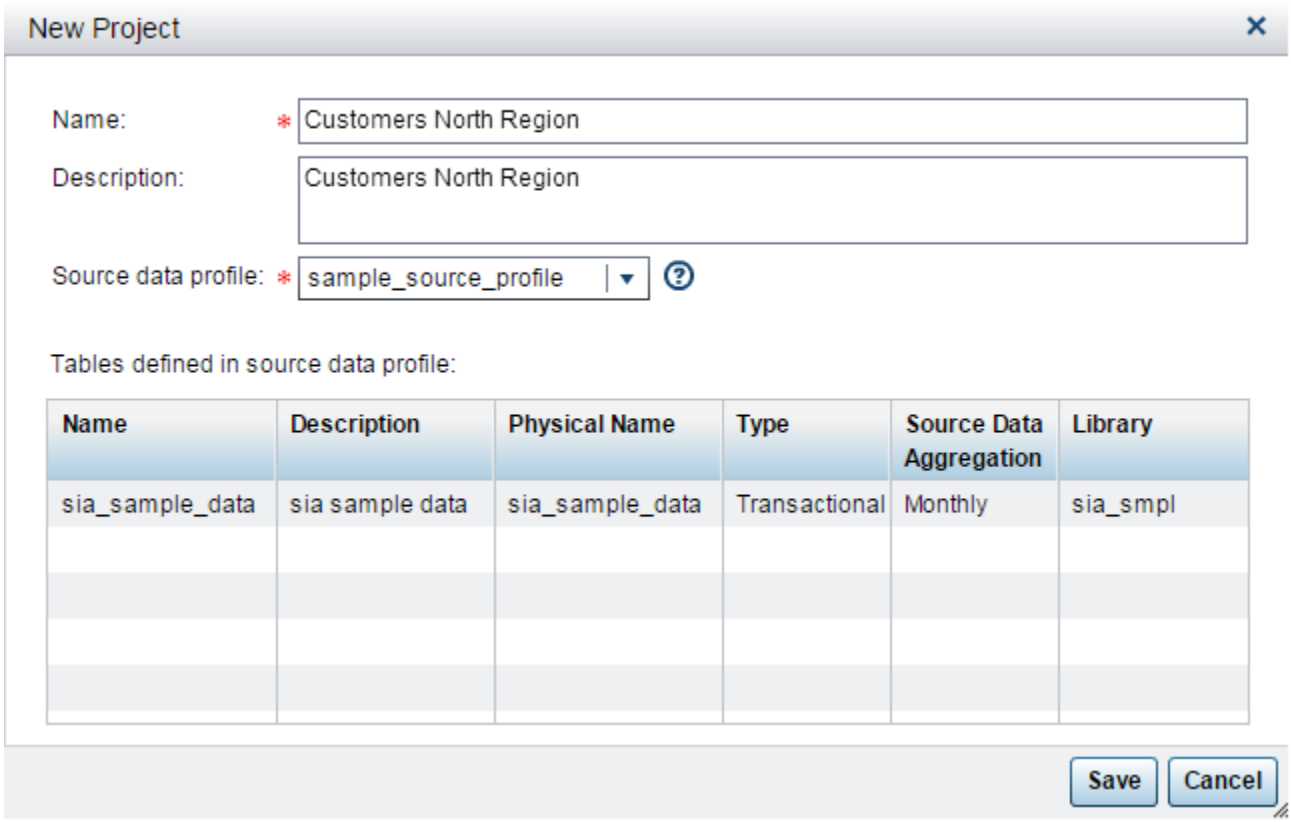
4 (Optional) Enter a description of the project. In this field, you can include the purpose for which you are defining the project.

5 Select the source data profile. The list displays the source data profiles that your administrator has defined. For more information, see “Create a Source Data Profile” on page 113.

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.

Figure 3.1 *Source Data Profile Selection*



6 Click **Save**. The project that you have created is added to the All Projects list. 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.

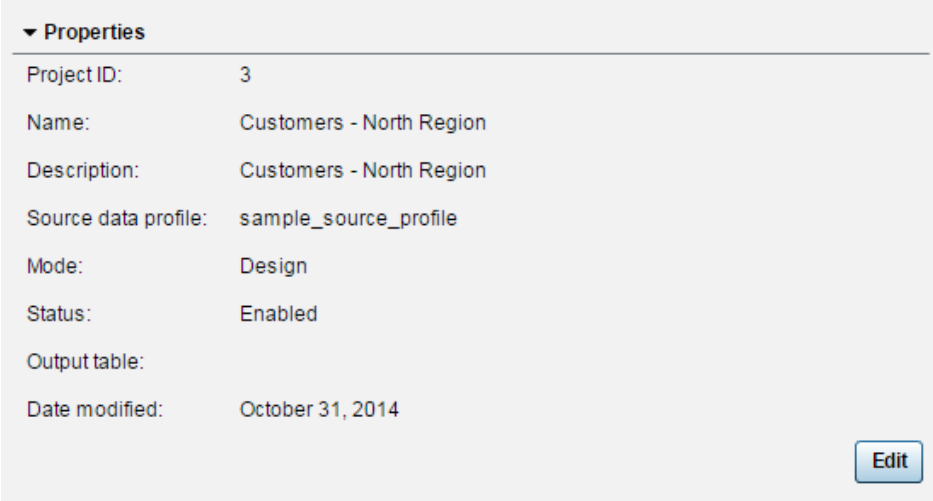
Note: If your administrator has granted you the capability to view all projects, along with your own projects, you can view the projects that other users might have created. However, you can edit the properties of only those projects that you have created.

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:

Figure 3.2 *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 more information, see “Modes of a Project” on page 15.

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 more information, see “Overview of Source Data Profiles” on page 113. 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 more information, see “Modes of a Project” on page 15.

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

Table 3.1 *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 more information, see “Final Output Data of SAS Customer Link Analytics” on page 65.

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.

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 more information, see Chapter 10, “Batch Processing,” on page 67.

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**.

4

Data Extraction

Configure 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 and select an inclusion list. For more information, see "About Inclusion Lists" on page 29.

When you run the Data Extraction workflow step, the transactional data associated with these links and nodes is extracted from the preconfigured external source system for the specified time period.

To configure the Data Extraction workflow step:

1. In the Workflow Diagram pane, select **Data Extraction**.
2. On the toolbar, select  .
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 more information, see "Create a Source Data Profile" on page 113. 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 more information, see "Refresh Data in a Transactional Table" on page 108. 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**.

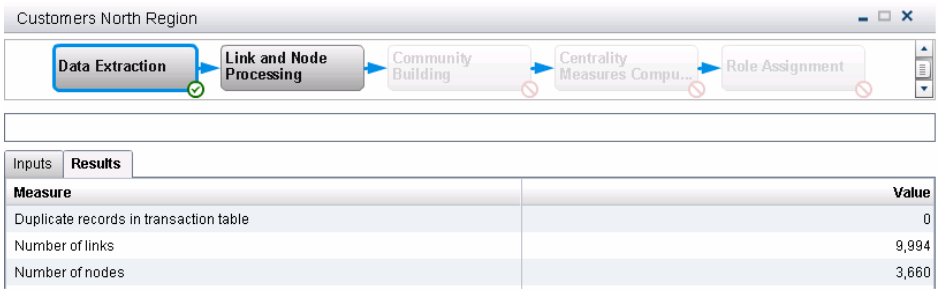
7 On the toolbar, click ▶. The required transactional data is extracted and the results are produced.

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.

In addition, this tab informs you if there are any duplicate records. It displays the number of duplicate records that are retrieved in the node inclusion list, the link inclusion list, and the transaction table that are associated with the project. It is recommended that you remove these duplicate records before you proceed further.

Figure 4.1 *Results of Data Extraction Workflow Step*



About Inclusion Lists

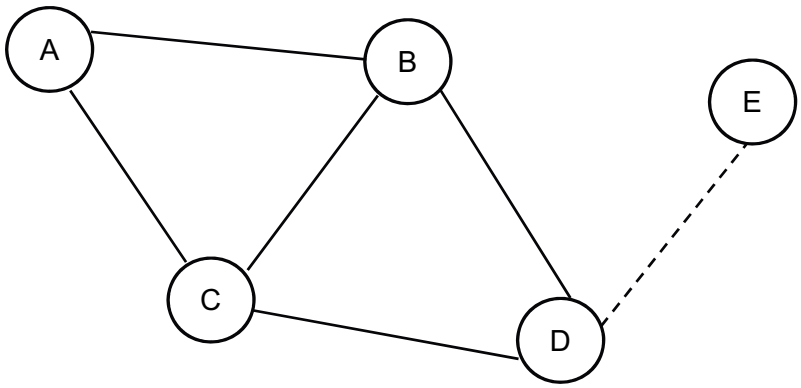
Overview of Inclusion Lists

Your administrator defines one or more lists of links or nodes that you can consider when building communities. For more information, see "Attribute Tables" on page 103. 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 4.2 *Selected Nodes in a Network*

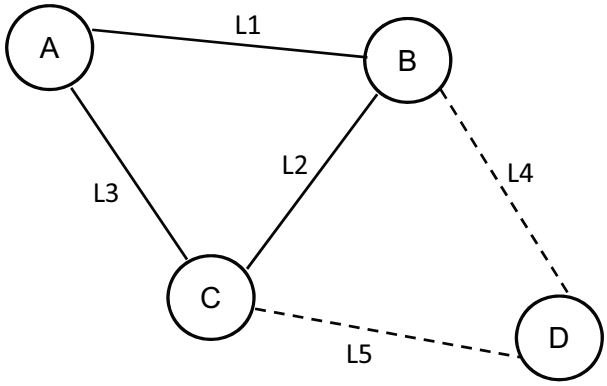


As depicted in the network diagram, transactional data of links that involve nodes A, B, and C is extracted. The inclusion list restricts the link that exists between D and E from being considered for extracting data.

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 4.3 *Selected Links in a Network*



5

Link and Node Processing

Configure the Link and Node Processing Workflow Step

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 transactional measures. For more information, see “Transactional Measures for Filtering Links and Nodes” on page 36.

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.

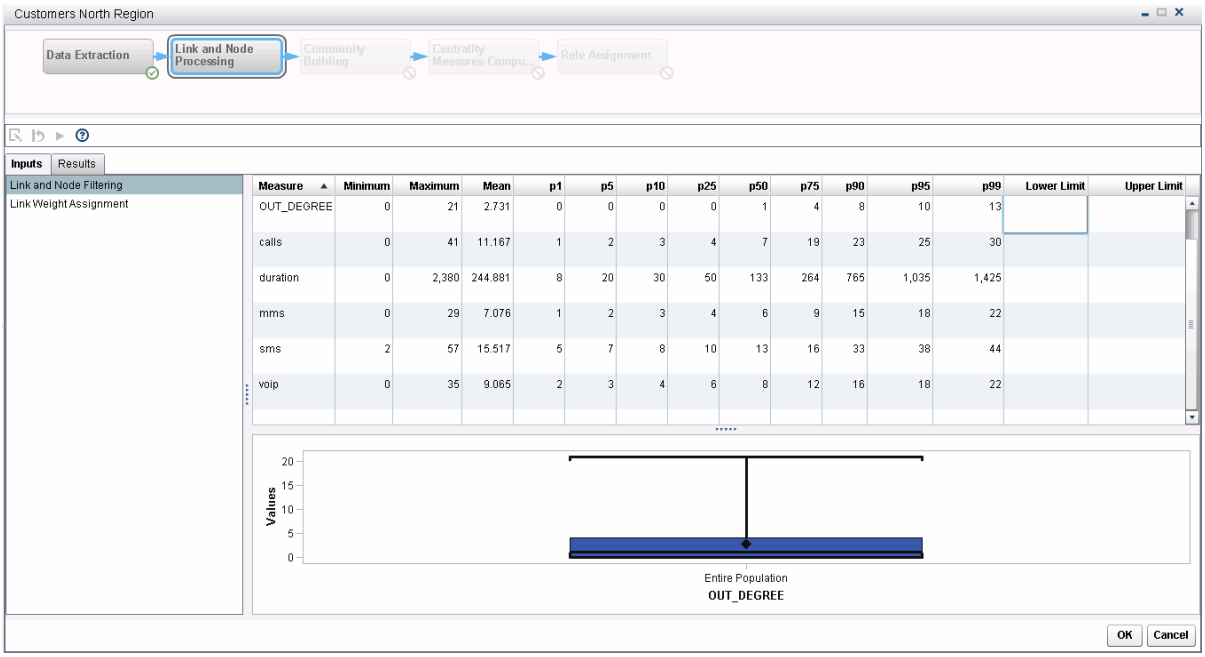
When you run the Link and Node Processing workflow step, links and nodes of the entire network that satisfy the limit values that you have specified for all transactional measures are selected. Also, a weight is assigned for each link that is selected.

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 transactional measures that are predefined for filtering links and nodes are displayed. For more

information, see “Transactional Measures for Filtering Links and Nodes” on page 36.

Figure 5.1 *Link and Node Filtering Page*



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

5 View the statistics that are produced for the transactional measure. You can also view this information graphically. For more information, see “Interpreting the Graphical Statistics for Transactional Measures” on page 35.

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

Here are a few examples that indicate how you can filter nodes of a communications network.

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

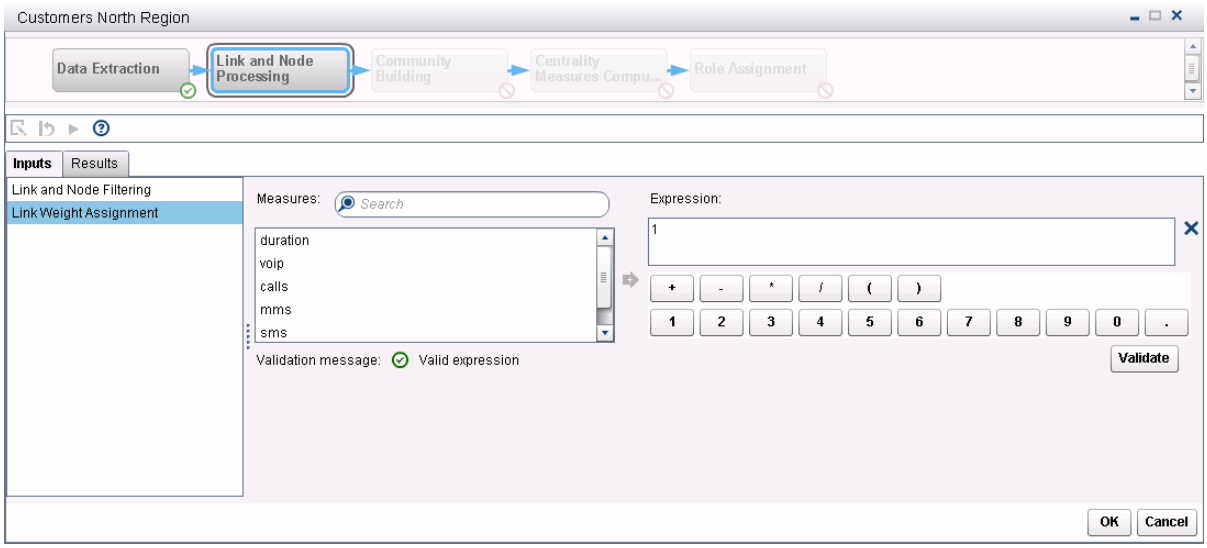
Extremely high values for call duration can be outliers in the data. Assume that 90% (percentile 90 column) of the nodes have call duration values less than or equal to 400 minutes. However, the maximum value is 7,500. You specify the **Upper Limit** as 2,000. All the subscriptions (nodes) that have call duration values greater than 2,000 minutes will not be considered for the analysis.

7 Repeat steps 4 through 6 for the other transactional 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 transactional measures are filtered and made available in subsequent steps.

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

Figure 5.2 *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 transactional measures, arithmetic operators, and numeric values.

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

Here are a few examples that indicate how you can define link weights in a communications network.

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 transactional 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 transactional 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 more information, see "Configure the Community Building Workflow Step" on page 39.

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 transactional measure in the denominator, then make sure that the lower limit that you specify for this transactional 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.

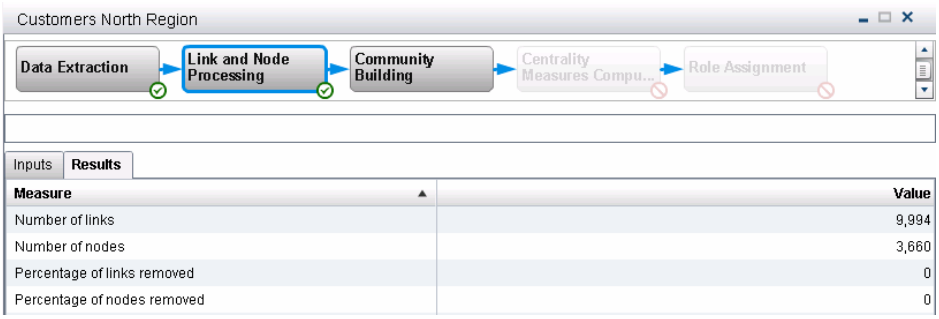
View Results of the Link and Node Processing Workflow Step

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 more information, see Step 6 on page 32.

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.

Figure 5.3 *Results of Link and Node Processing Workflow Step*



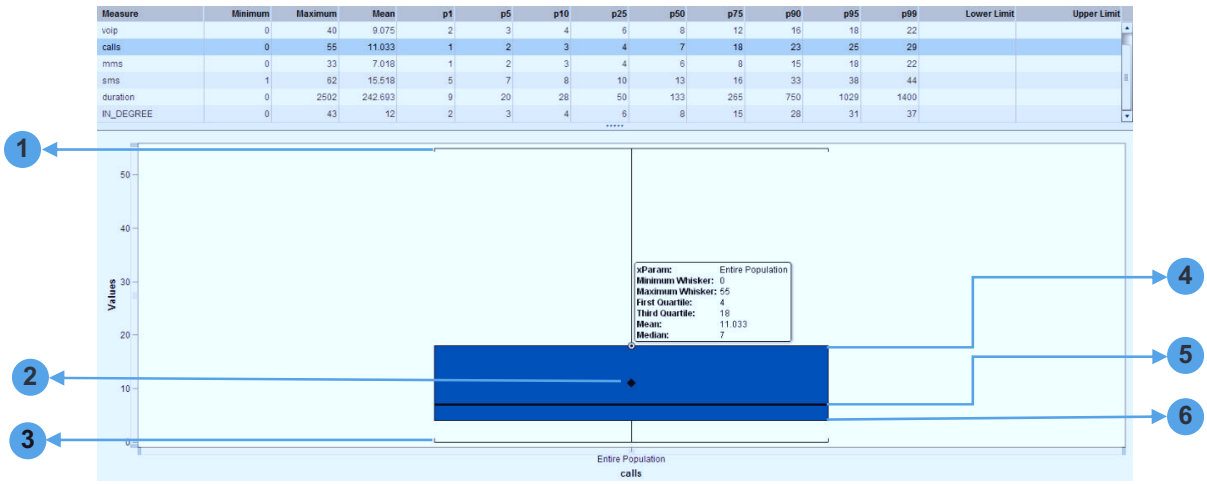
If the results that are produced do not meet your business requirements, you can reconfigure the limit values that you have specified. For example, if almost 80% of your links or nodes are removed, then you can change the limit values accordingly.

Interpreting the Graphical Statistics for Transactional Measures

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

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

Figure 5.4 *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 transactional 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 transactional measure. That is, the minimum call count of a node is 0.

4 The third quartile indicates that 75% of the observations of the transactional 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 transactional 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 transactional 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.

Transactional Measures for Filtering Links and Nodes

You can filter links and nodes based on the limit values that you enter for certain transactional measures. SAS Customer Link Analytics provides the in-degree and out-degree measures. However, the rest of the transactional measures are extracted from the transactional data. You configure these transactional measures in the Administration workspace. For more information, see "Configure the Column Values of a Table" on page 106.

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

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

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

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

6

Community Building

Configure 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 this method, the workflow step recursively splits larger communities into smaller units until certain conditions are satisfied.

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

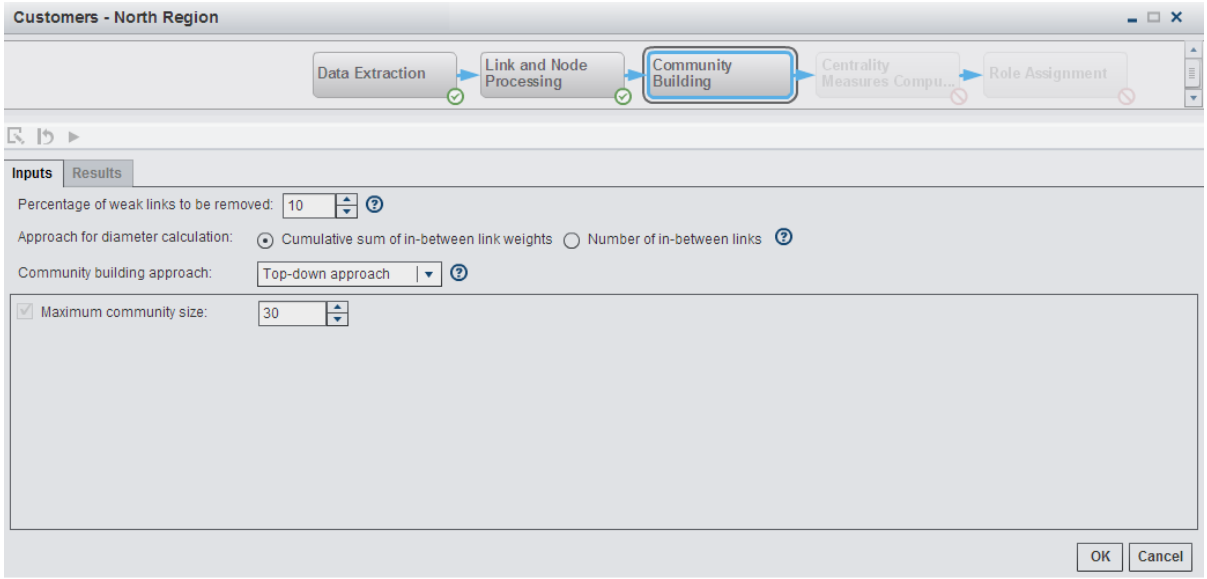
You can choose either of these methods, depending on your business requirements.

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.

To configure the Community Building workflow step:

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

Figure 6.1 *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 more information, see Step 11 on page 33.

   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 that you want to use for calculating the diameter. For more information, see “Diameter of a Community” on page 142.

   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 more information, see “Interpreting the Results Summary” on page 43.

   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.

   For more information, see “Diameter of a Community” on page 142.

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. However, 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. However, 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:

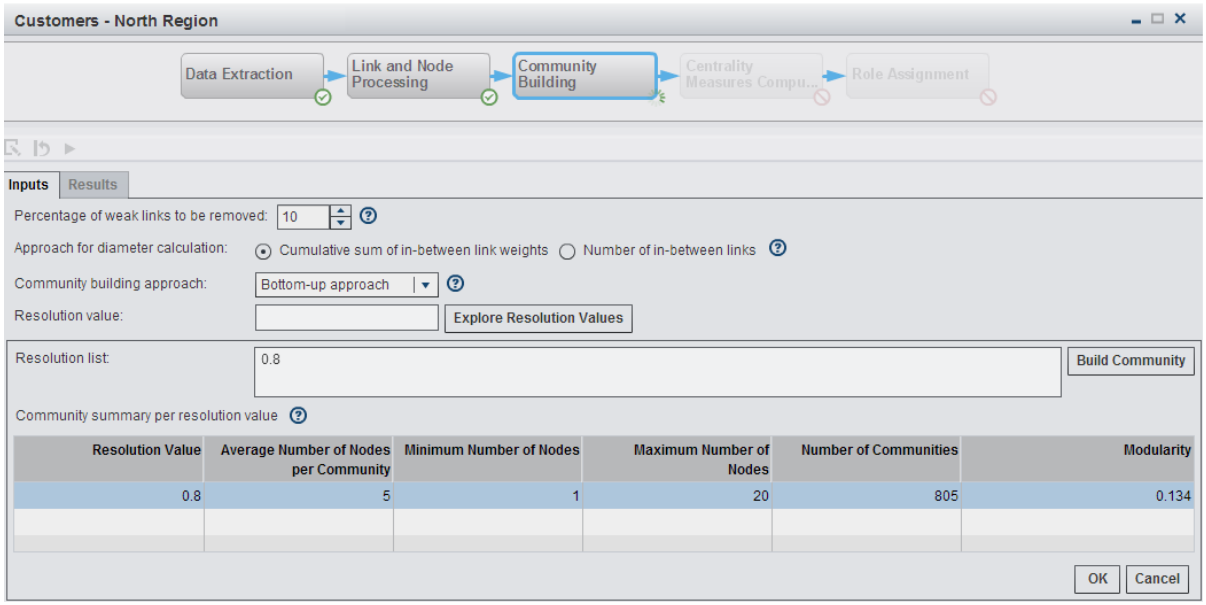
a Click **Explore Resolution Values**.

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 one or a maximum of 20 values, each value enclosed in parentheses.

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.

Figure 6.2 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 6.1 *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 more information, see Step 4 on page 40.
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 more information, see Step 4 on page 40.
Maximum number of nodes	Indicates the maximum community size.

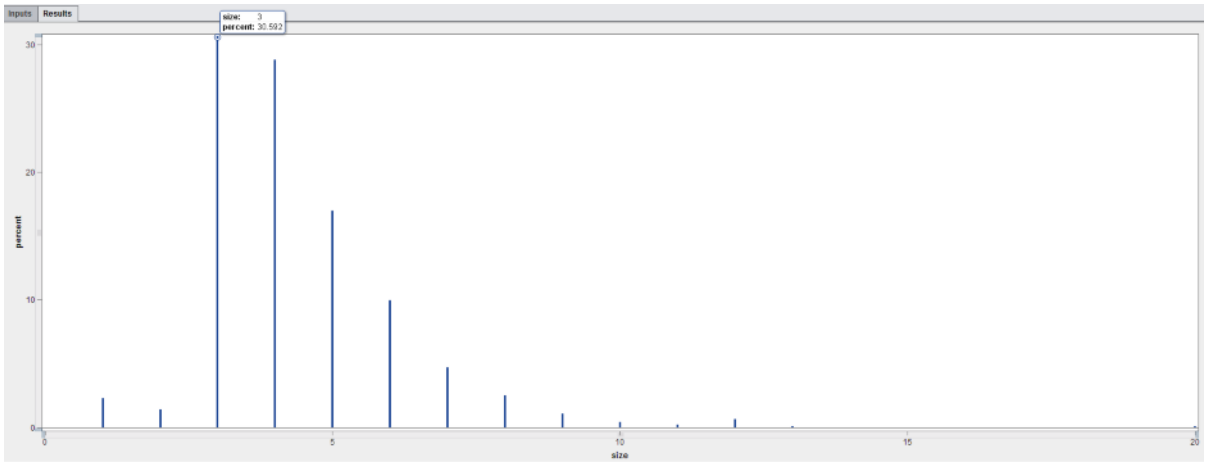
Measure Name	Description
Modularity of the network	Indicates the quality of a division of a network into communities.
Number of communities in the network	Indicates the number of communities that are built in the network.

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.

Figure 6.3 *Community Building: Graphical Results*



7

Centrality Measures Computation

Configure the Centrality Measures Computation Workflow Step

The Centrality Measures Computation workflow step enables you to provide input for the parameters that SAS Customer Link Analytics requires for 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. For details about these centrality measures, see Appendix 2, “Overview of Centrality Measures and Clustering Parameters,” on page 145.

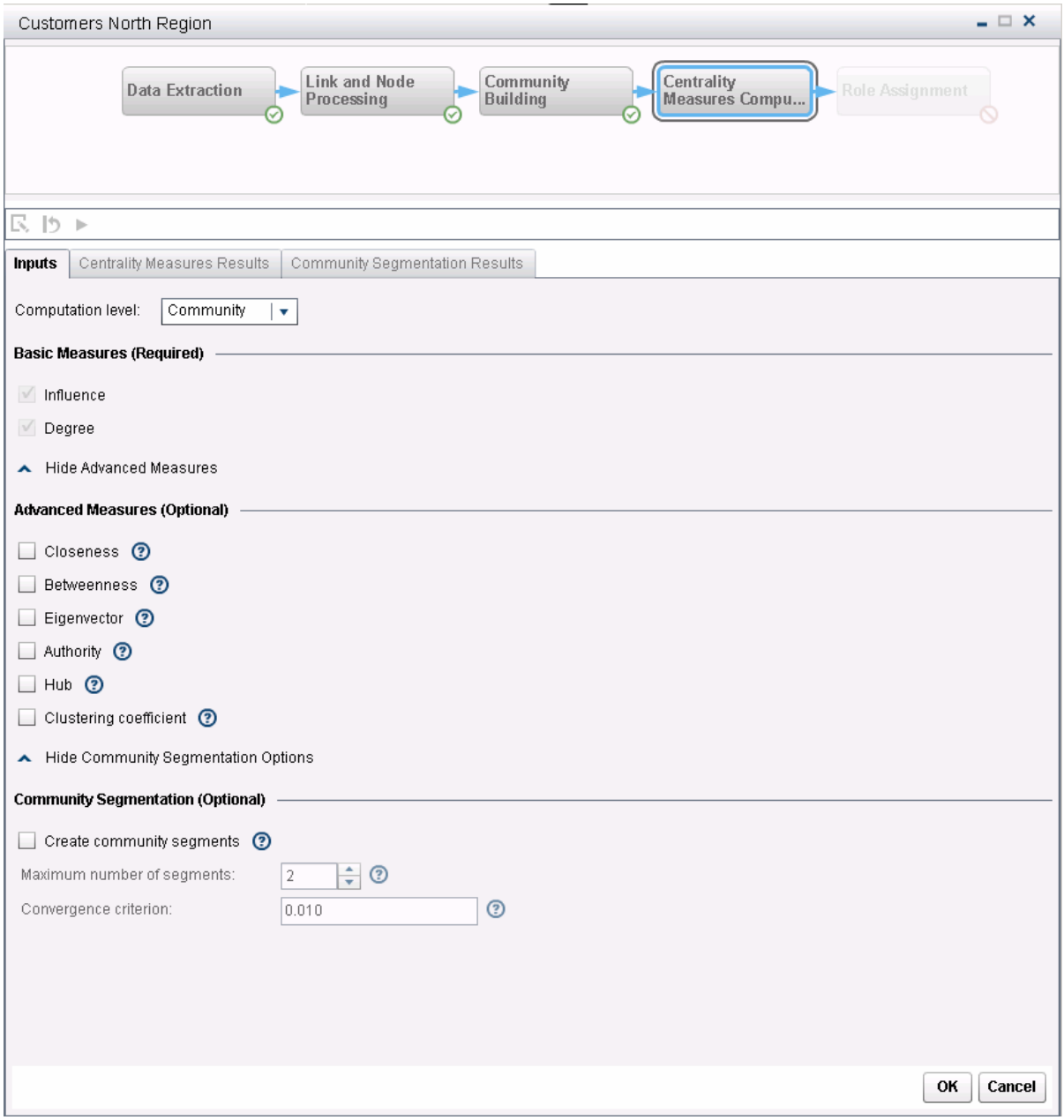
When you run this workflow step, the centrality measures are computed and their graphical results are displayed. These results can help you understand how significant the nodes are within the network or community. In addition, this workflow step enables you to group communities into various segments. This sub-step is called *community segmentation*. Depending on your business requirements, you can decide whether you want to segment communities. For more information, see “How Community Segmentation Works” on page 52.

Note: If you want to segment communities, make sure that you do not run the Centrality Measures Computation workflow step with its default setup. On the toolbar, click  to configure community segmentation. For more information, see step 6 Step 6 on page 48.

To configure parameters for computing centrality measures:

1. In the Workflow Diagram pane, click **Centrality Measures Computation**.
2. On the toolbar, select  .

Figure 7.1 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 check box is selected by default. This centrality measure indicates the number of immediate neighbors of a node.

The following variables are computed for this centrality measure:

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

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

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

5 Click ⌄ to display the advanced centrality measures.

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 ⌄ to display the options that are defined for community segmentation.

7 Select **Create community segments** if you want to group communities into various segments. For more information, see “How Community Segmentation Works” on page 52.

8 Specify appropriate values for the parameters. The clustering procedure produces community segments depending on the values that you specify here.

Maximum number of segments
specifies the maximum number of segments that the clustering procedure should create.

Convergence criterion
specifies the threshold value for the relative change in the cluster centroids in each iteration of the clustering procedure.

For more information, see “Convergence Criterion” on page 148.

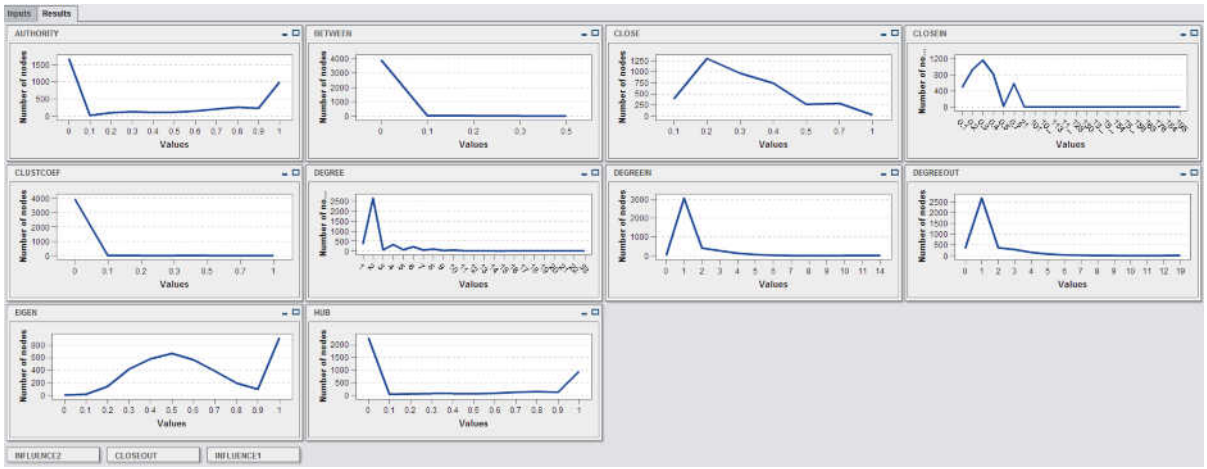
9 Click **OK**.

10 On the toolbar, select ▶. The centrality measures that you have selected are computed for each community or network. In addition, if you have chosen community segmentation, one or more communities are grouped into a segment. Also, a segment ID is assigned to each community to indicate the segment to which it belongs.

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.

Figure 7.2 *Centrality Measures Computation: Graphical Results*



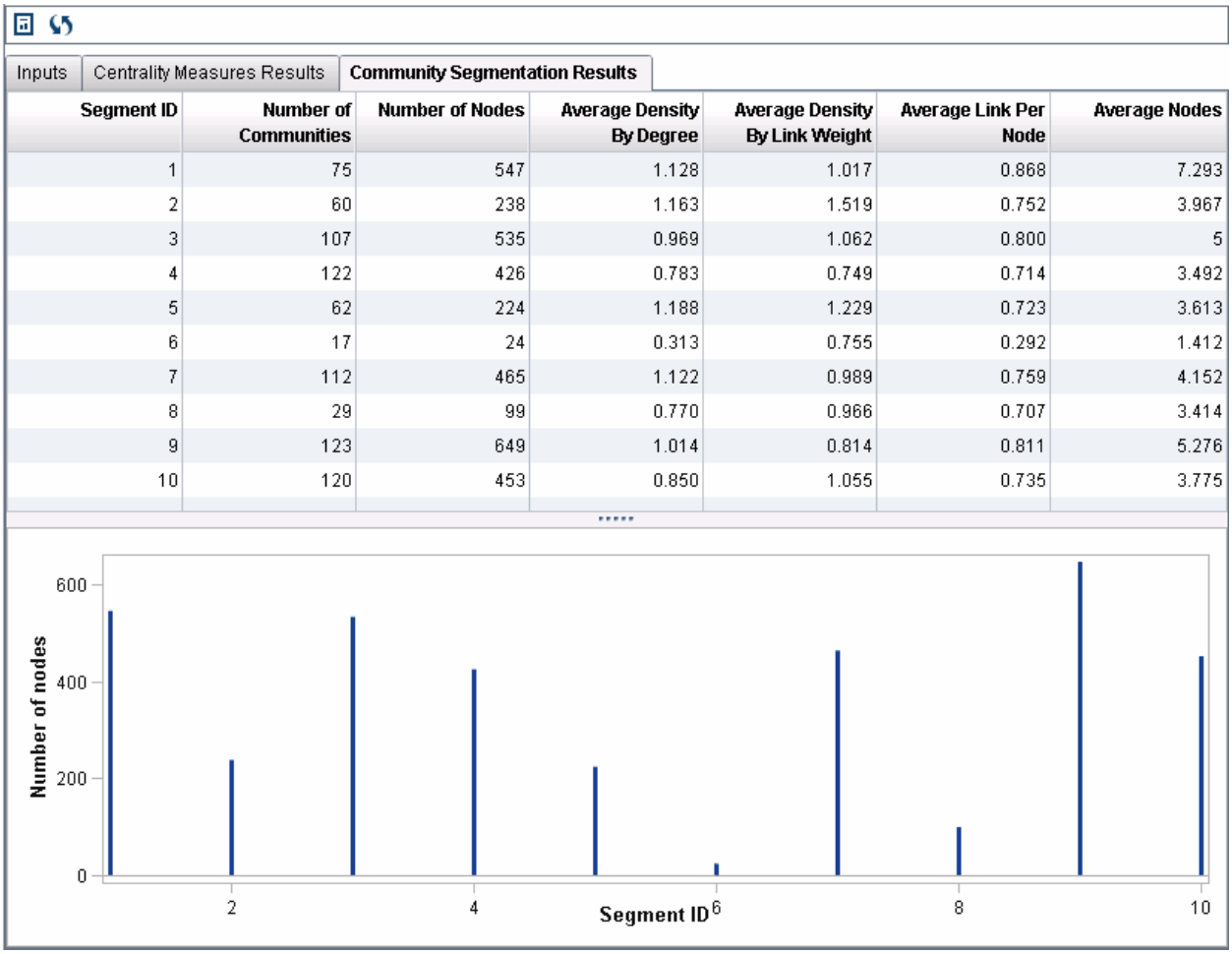
Based on the distribution of centrality measures within a community or a network, you can understand the relative significance of the nodes. Also, you can use this information for defining roles within your community.

View Results of Community Segmentation

If you select the **Create community segments** check box, the results of the community segmentation sub-step are displayed on the **Community Segmentation Results** tab. This tab displays information about the segments that SAS Customer Link Analytics produces when you run the Centrality Measures Computation workflow step.

The following information is displayed for each segment that is created for the project. Each segment represents the communities that are grouped within it.

Figure 7.3 *Community Segmentation: Results Tab*



Segment ID
indicates the unique ID that is assigned to the segment.

Number of Communities
indicates the number of communities that are grouped in the segment.

Number of Nodes
indicates the total number of nodes that belong to the segment.

Average Density By Degree
is the ratio of the average degree for a community over the average degree for the entire segment. A higher value indicates that the nodes of the segment have more connections.

Average Density By Link Weight
is the ratio of the average link weight for a community over the average link weight for the entire segment. A higher value indicates that the nodes of the segment have strong connections.

Average Link Per Node
indicates the average number of links that a node of the segment is connected with.

Average Nodes
indicates the average number of nodes in a community that belongs to the segment.

In addition, a needle plot is displayed on this tab. This plot indicates the distribution of nodes across various segments. The X-axis represents the segment ID and the Y-axis represents the number of nodes in a segment.

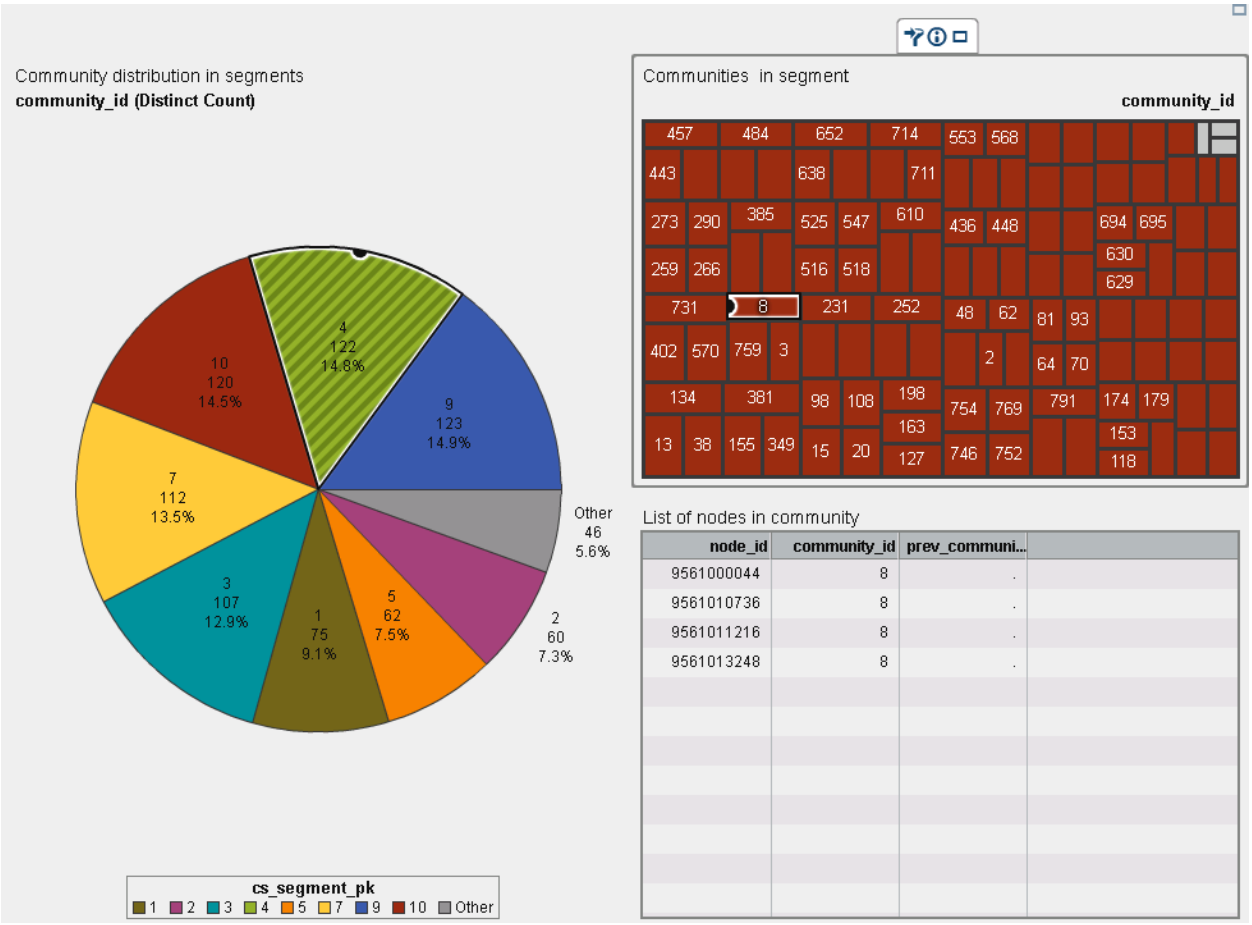
View the Community Segmentation Report

SAS Customer Link Analytics generates the Community Segmentation report if you create community segments when you run the Centrality Measures Computation workflow step. This report indicates the distribution of communities across various segments.

To view the Community Segmentation report:

1 On the toolbar of the **Community Segmentation Results** tab, click . The Community Segmentation report opens in SAS Visual Analytics Viewer.

Figure 7.4 *Community Segmentation Report*



2 The report displays the following information:

Community distribution in segments

The pie chart indicates the percentage distribution of communities into various segments. It also indicates the total number of communities that are grouped in each segment.

Communities in segment
The treemap displays the communities of the segment that you selected in the pie chart as a set of rectangles (called *tiles*). Each tile represents a community, and the size of each tile represents the number of nodes in that community.

List of nodes in community
The table displays the nodes of the community that you selected in the treemap. For each node, the current and previous community to which it belongs is displayed.

3 To return to the Projects workspace, on the banner, click **SAS Customer Link Analytics**.

Reload Reporting Data to SAS Customer Link Analytics LASR Analytic Server

You can view the Community Segmentation report based on the reporting data that SAS Customer Link Analytics loads on the SAS Customer Link Analytics LASR Analytic Server. The reporting data is loaded on the SAS Customer Link Analytics LASR Analytic Server when you run the Centrality Measures Computation workflow step. If the SAS Customer Link Analytics LASR Analytic Server goes down, you cannot view the report because the LASR data will be unavailable. In this case, to view the report, reload the data on the SAS Customer Link Analytics LASR Analytic Server.

To reload the reporting data to SAS Customer Link Analytics LASR Analytic Server:

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

2 Select the **Community Segmentation Results** tab.

3 On the toolbar, click . SAS Customer Link Analytics reloads the reporting data to SAS Customer Link Analytics LASR Analytic Server.

How Community Segmentation Works

Overview

When you configure the Centrality Measures Computation workflow step, SAS Customer Link Analytics enables you to perform an additional sub-step, community segmentation. *Community segmentation* groups the communities into various segments that have similar patterns for the transactional measures. SAS Customer Link Analytics uses the statistical clustering procedure to divide communities into segments. Each segment represents a group of communities that have homogenous behavior for centrality measures, transactional measures, and link weights. Conversely, there is heterogeneity across segments.

Analytical Flow for Community Segmentation

When you segment communities, SAS Customer Link Analytics performs the following analytical tasks:

1 ABT building

SAS Customer Link Analytics builds the modeling ABT with the name cla_cs_abt_<*Project ID*> and registers it in the following SAS metadata location: **`/Shared Data/SAS Customer Link Analytics/Cust Link Analytics 6.5/Projects/<Project ID>/Data Sources/Business Data`**.

This ABT contains the following types of variables. For more information about these variables, see *SAS Customer Link Analytics: Data Reference Guide.*

Variables based on centrality measures
: These are the community-level variables that are created based on the centrality measures that are computed for the project. For each community and centrality measure, the following variables are created:

- Average value of the centrality measure
- Minimum value of the centrality measure
- Maximum value of the centrality measure
- Total value of the centrality measure

Variables based on transactional measures
: These are the community-level variables that are created based on the transactional measures that are defined for the project. For each community and transactional measure, the following variables are created:

- Minimum incoming value of a transactional measure
- Minimum outgoing value of a transactional measure
- Maximum incoming value of a transactional measure
- Maximum outgoing value of a transactional measure
- Average incoming value of a transactional measure
- Average outgoing value of a transactional measure
- Total incoming value of a transactional measure
- Total outgoing value of a transactional measure

Variables based on link weights
: These are the community-level variables that are created based on the link weights that are assigned during the Link and Node Processing workflow step. For each community, variables are created for the incoming link weight, outgoing link weight, and total link weight. For example, for each community, the following variables are created for the incoming link weight:

- Maximum incoming link weight
- Minimum incoming link weight
- Average incoming link weight

- Total incoming link weight

Similar variables are created for each community for the outgoing link weight and the total link weight.

Variables based on diameter and density of community
These are the community-level variables that indicate the homogeneity of the community. For distributed mode, the following variables are computed for each community:

Density of the community by degree (Density_1)
is the ratio of the average degree for a community over the average degree for the entire network.

Density of the community by link weight (Density_2)
is the ratio of the average link weight for a community over the average link weight for the entire network.

Average number of links per node
is the ratio of the total number of links in a community over the total number of nodes in the community.

In addition to these variables, for the non-distributed mode, the following variables are created:

Density of the community
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.

Diameter of the community
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 more information, see Step 4 on page 40.

In addition, the ABT contains the variable that computes the total number of nodes in each community.

2 Model building

The following tasks are completed in this step:

a Using the appropriate statistical procedure, variables that explain the maximum variance in data are extracted from the modeling ABT. These variables are stored in the application data.

b Based on the variables that are stored in the application data, a reduced ABT is built.

c A clustering model is built using the reduced ABT, and the score code is generated with the name sia_segm_score_code_<*Project ID*>. This code is available in the following location: **<*SAS configuration directory*>/Lev1/AppData/SASCustomerLinkAnalytics/6.5/projects/batchcode**.

In addition, the scored ABT (cla_segm_score_<*Project ID*>) is generated and stored in the business data, and the scores are stored in the cla_clust_scr_<*Project ID*>_<*Active Model PK*> business data table.

d The following information is stored in the application data:

Clustering model information
: contains information about the clustering models that are built for the project. Only a single model can be active at a time.

Community segments
: contains information about the segments that are created for the active model.

Variable details
: contains information about the reduced variables for the active model.

Variable behavior
: contains the behavior of the reduced variables across segments.

For more information about these tables, see *SAS Customer Link Analytics: Data Reference Guide.*

3 Publishing results

In this step, SAS Customer Link Analytics produces segment-level statistics and publishes this information on the **Community Segmentation Results** tab. For more information, see "View Results of Community Segmentation" on page 49.

4 Preparing report data and viewing the Community Segmentation report

In this step, SAS Customer Link Analytics populates the cla_comseg_dtl_data_<*Project ID*> table. This table contains information about community-level statistics, node-level centrality measures, and community segments for two successive runs of the community segmentation flow. For more information about the column details, see *SAS Customer Link Analytics: Data Reference Guide*. SAS Customer Link Analytics uses this table for generating the Community Segmentation report. SAS Customer Link Analytics also loads this table into the SAS Customer Link Analytics LASR Analytic Server with the name cla_comseg_dtl_data_lasr_<*Project ID*>. Both these tables are registered in the following metadata location: **`<Project Path>/<Project ID>/Data Sources/Report Data`**. For example, the project path can be **`/Shared Data/SAS Customer Link Analytics/Cust Link Analytics 6.5/Projects`**. For more information about the project path, see *SAS Customer Link Analytics: Administrator's Guide*.

8

Role Assignment

Define a Role

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.

SAS Customer Link Analytics provides you with a default predefined role. In addition, you can define your own roles. For more information, see “Sample Role Definitions” on page 63.

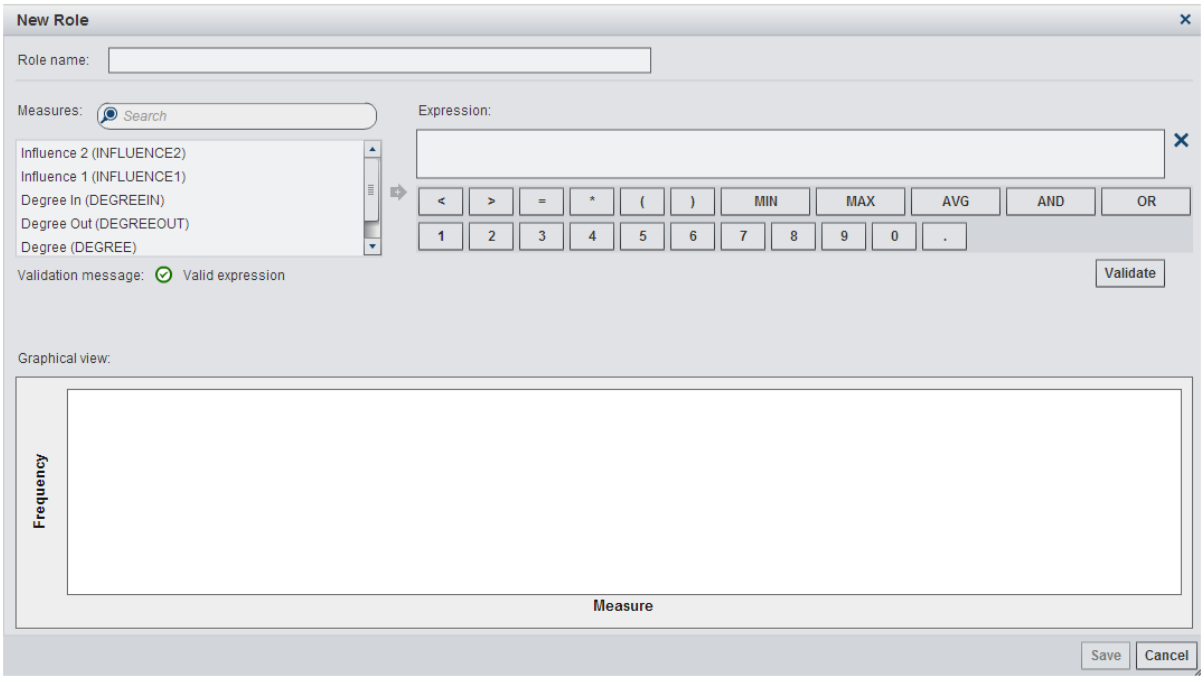
To define a role:

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

2 On the toolbar, select  .

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

Figure 8.1 *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:

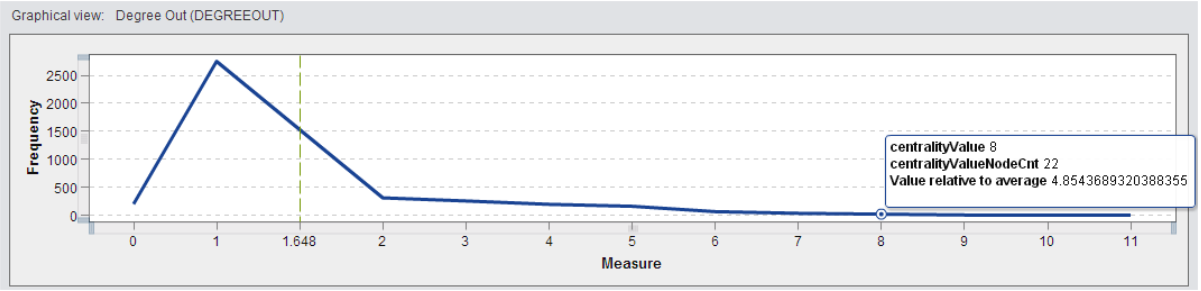
- A node that is interacting with at least 8 other nodes. In other words, nodes that have an in-degree or out-degree value greater than 8.
- A node that is close to the other nodes in the community. In other words, nodes that have a higher closeness value (for example, greater than 0.5).

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 **>**.

Figure 8.2 Graph of Degree-out Centrality



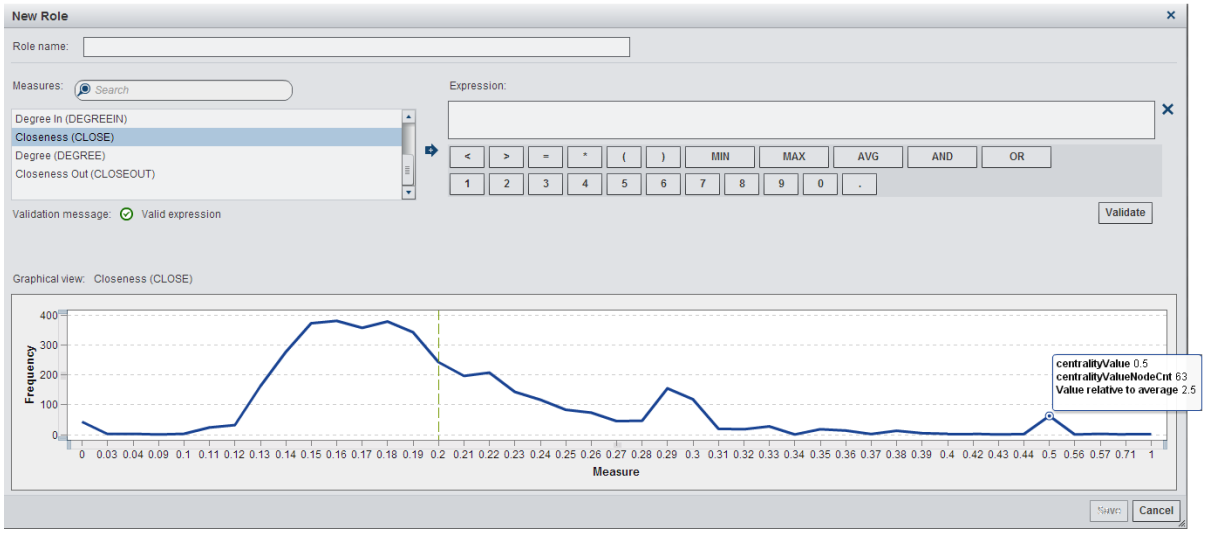
c In the graph, click the data point that corresponds to a value of degree-out equal to 8 on the X-axis. In the expression, this value is displayed as **4.85 * AVG(DEGREEOUT)**.

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)**.

Figure 8.3 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. A unique ID is assigned to the role that you define. This ID is used in the naming convention of variables that are created when you run the data enrichment process. For more information about these variables, see *SAS Customer Link Analytics: Data Reference Guide*.

Note: Before you run the workflow step, make sure that you define the precedence for the roles. For more information, see "Define the Precedence of a Role" on page 61.

Edit a Role

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

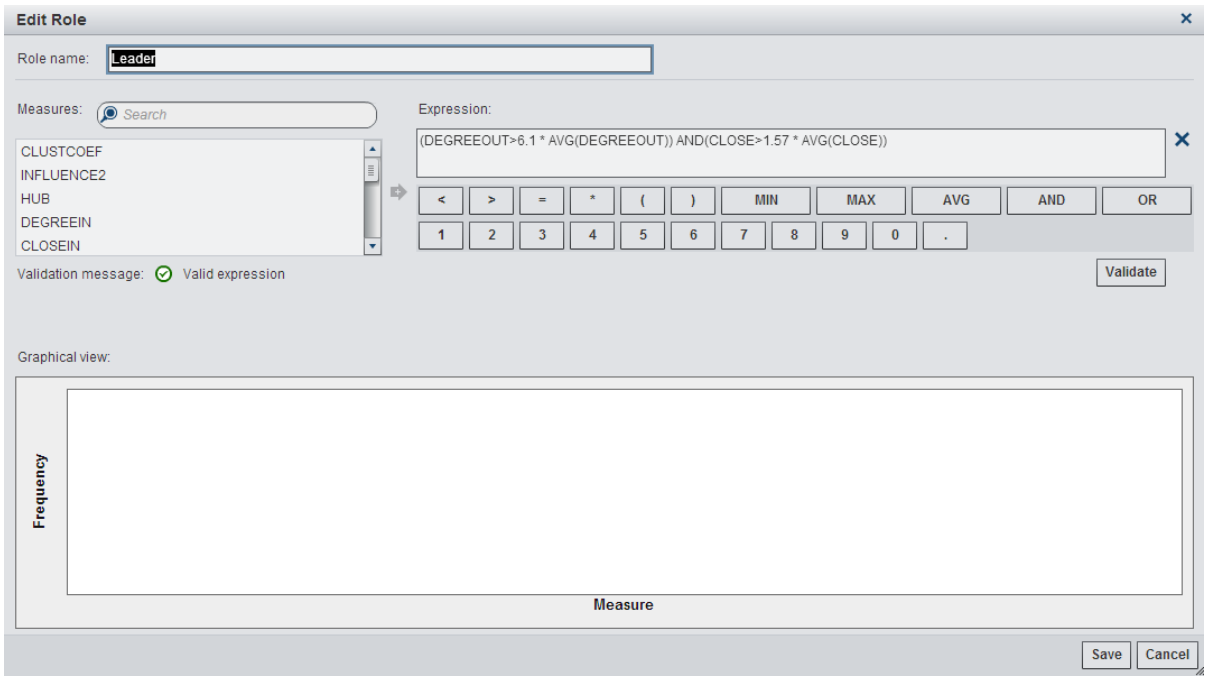
To edit a role:

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.

Figure 8.4 *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

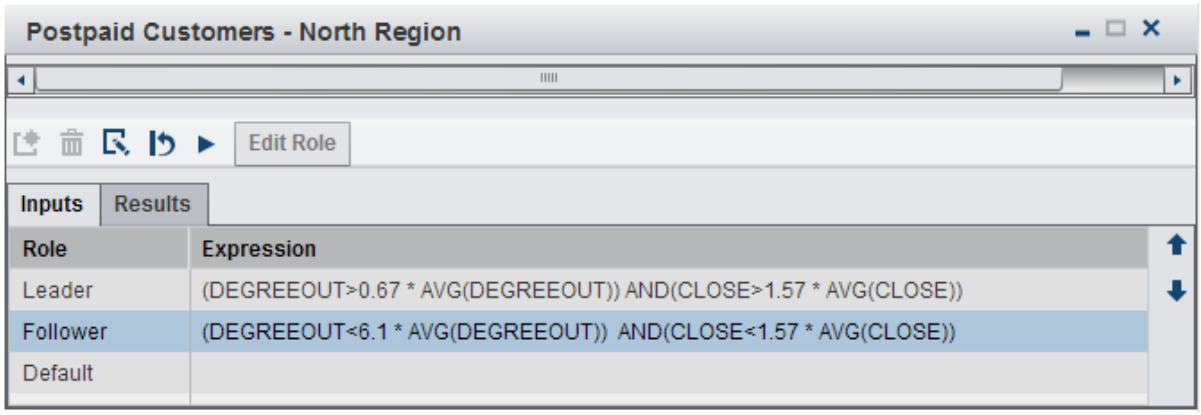
When you define all the roles that you want to consider for the nodes in the community, you can define the precedence for each role. When you run the Role Assignment workflow step, a role is assigned to a node if it satisfies the expression that you have defined for that role. A single node can satisfy the expressions that are defined for one or more nodes. In this case, the role with the highest precedence is assigned to the node.

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.

Figure 8.5 *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:

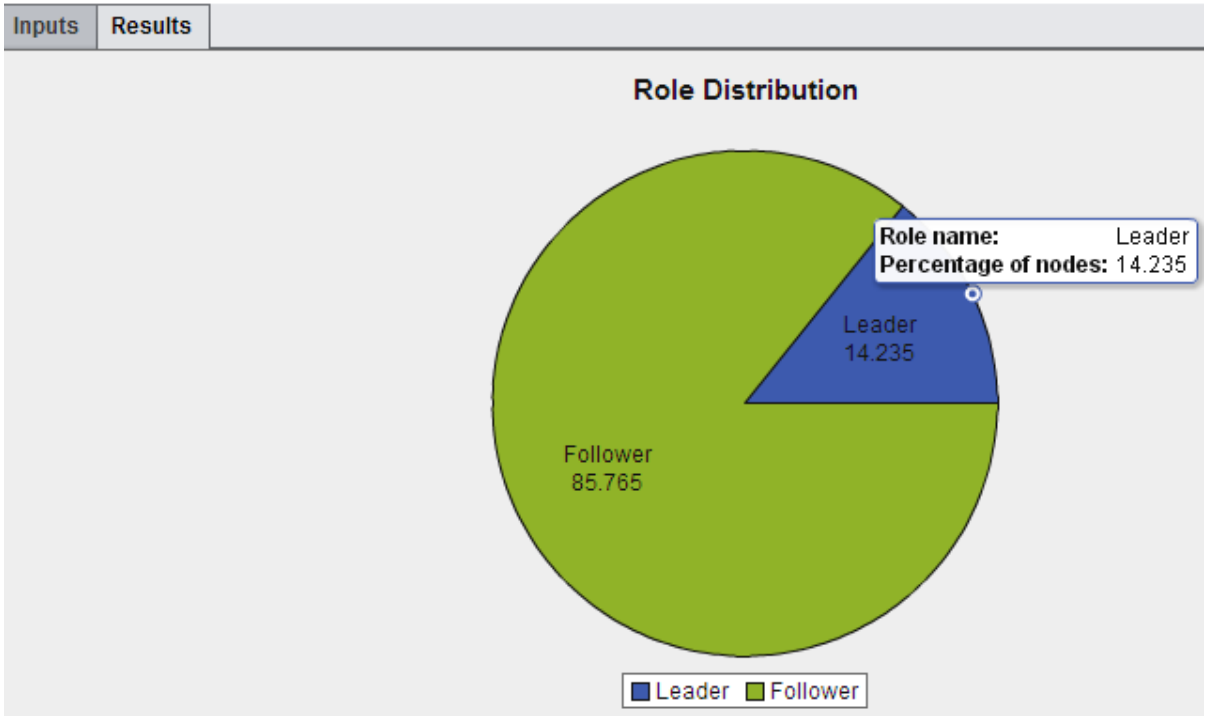
- The node uniquely satisfies the conditions that are defined in the expression of a single role. The role is assigned to the node.
- The node satisfies the conditions that are defined in the expression of two or more roles. The role with the highest precedence order is assigned to the node.
- The node 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.

Figure 8.6 *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.

Sample Role Definitions

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 8.1 *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.
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 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.

9

Output Data of Project Workflow

Final Output Data of SAS Customer Link Analytics

When a project completes one run in design mode, SAS Customer Link Analytics produces certain results (also called the final output of SAS Customer Link Analytics) for each node and stores them in a business table. This table, cla_ra_p1_<*project ID*>, is created in the following location: `<SAS configuration directory>/AppData/SASCustomerLinkAnalytics/6.5/projects/data/business_output`. For more information 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 belongs. This ID is assigned to the node when you run the Community Building workflow step.

Segment ID
: indicates the ID of the segment to which a node belongs. This ID is assigned to the node if you have opted for community segmentation when you run the Centrality Measures Computation 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 more information, see "Configure the Centrality Measures Computation Workflow Step" on page 45.

10

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 change only 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 view only 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 directory>/Lev1/AppData/SASCustomerLinkAnalytics/6.5/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 more information, 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 the 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

11

Working with Data Enrichment and Data Loading

Enriching Project Output Data

Overview

When a project completes at least one run in design or batch mode, you can decide whether you want to process its output data. This data is collectively called the *project's output data*. It comprises the workflow step data, the final output of SAS Customer Link Analytics, the node attribute tables that you configure in the Administrative workspace, and certain derived information. The process of transforming and enriching the project's output data so that you can use it for further analysis is called *data enrichment*.

SAS Customer Link Analytics offers you a set of predefined enrichment categories at node and link levels. Each enrichment category enables you to enrich the project's output data, so that you can use it to perform specific tasks. The enriched data provides insights for customer behavior and communities and enables you to make well-informed business decisions. With comprehensive information about customers and their communities, you can focus on creating effective marketing campaigns.

Here are a few examples that explain how you can use the enriched data:

Community report
: SAS Customer Link Analytics enables you to create a Community report based on the enriched data. For more information, see "About the Community Report" on page 85.

Community visualization
: If you have installed SAS Visual Analytics Explorer, you can explore and visualize how communities are formed.

Campaign management
: With data enrichment, you have more detailed information about your customers. You can use this data in a suitable campaign management tool and target potential customers.

Data visualization
: If you have installed SAS Visual Analytics Explorer, you can visualize the enriched data according to your business requirements and gain insights about customer behavior.

Proactive indicators
: You can track changes in the roles of nodes over time. In addition, you can identify if a threshold value is reached for a particular node so that you can take corrective action. You can also track and monitor changes in the centrality measures of a node over time.

Node-Level Enrichment Categories

SAS Customer Link Analytics provides the following node-level enrichment categories:

- Roles and communities
- Node attributes
- Associations with neighboring roles
- Community-level statistics
- Aggregated transactional data
- Roles and communities over time
- Churn and acquisition indicators
- Relation with churned and acquired nodes

For more information, see Node-Level Enrichment Categories on page 74.

SAS Customer Link Analytics stores the node-level enriched data in a project-specific business data table, CLA_DP_NODE_LVL_<*Project ID*>. It also registers this table in the following SAS metadata location: **<*Project Path*>/<*Project ID*>/Data Sources/Business Data**. For example, the <*Project Path*> can be **/Shared Data/SAS Customer Link Analytics/Cust Link Analytics 6.5/Projects**. For more information about the project path, see the *SAS Customer Link Analytics: Administrator's Guide*.

Link-Level Enrichment Categories

SAS Customer Link Analytics provides the following predefined link-level enrichment categories:

- Roles and communities
- Node attributes
- Churn and acquisition indicators

For more information, see Link-Level Enrichment Categories on page 78.

SAS Customer Link Analytics stores the link-level enriched data in a project-specific business data table, CLA_DP_LINK_LVL_<*Project ID*>. It also registers this table in the following SAS metadata location: `<Project Path>/<Project ID>/Data Sources/Business Data`. For example, the <*Project Path*> can be `/Shared Data/SAS Customer Link Analytics/Cust Link Analytics 6.5/Projects`.

Recommendations for Node-Level Enrichment Categories Selection

If you want to enrich data using the node-level enrichment categories, make sure that the source data profile contains a node attribute table. For more information, see "Create a Source Data Profile" on page 113. Otherwise, you will not be able to save your enrichment category selection. For more information, see "Select an Enrichment Category for a Project" on page 73.

Select the node-level enrichment categories, **Roles and communities over time** and **Relation with churned and acquired nodes**, for projects that are in batch mode. During the data enrichment process, these categories process workflow step data of two successive runs of a project. In the design run, when you re-run a project, the role of a node is bound to change. As a result, the comparative data in the two successive runs always indicates a role change for all the nodes. Therefore, you cannot use this information to filter or target nodes that have a role change.

Select an Enrichment Category for a Project

You can select enrichment categories for a project that has completed at least one successful run in batch or design mode (that is, if the status of the project is **Successfully executed**). If the project has any other status, you can view the current selection. However, you cannot modify it.

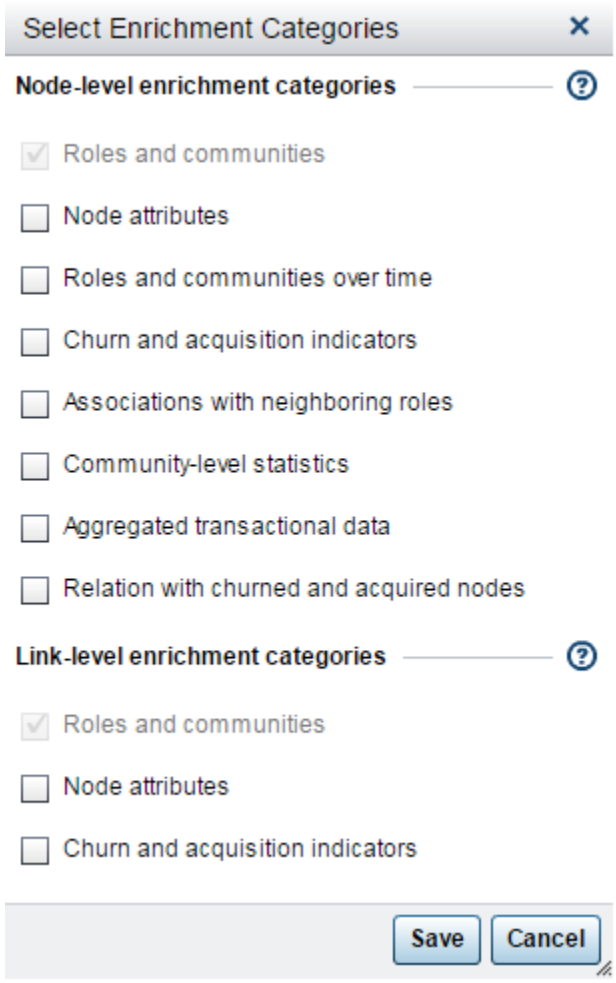
To select an enrichment category for a project:

1 In the Projects workspace, select the project whose output data you want to enrich.

Note: You can enrich data using the default enrichment category, **Roles and communities** that is defined at node and link level. To do so, you can directly proceed to enrich the project's output data. For more information, see "Enrich the Output Data of a Project" on page 79. However, if you want to enrich the project's output data using other enrichment categories also, perform the following steps.

2 In the Data Enrichment for Exploration and Analysis pane, click **Modify**. The Select Enrichment Categories window appears.

Figure 11.1 Select Enrichment Categories Window



3 Select the enrichment categories.

a Select the node-level enrichment categories. The following categories are available for your selection:

Note: If you have not included a node attribute table in the source data profile of the project, you will not be able to save the node-level categories that you select. For more information, see "Create a Source Data Profile" on page 113.

Roles and communities

Identifies the relative significance of a node in a community. The enriched data contains the final output that SAS Customer Link Analytics produces when a project completes one run in design or batch mode. For more information, see "Final Output Data of SAS Customer Link Analytics" on page 65.

You can use the enriched data to identify various communities of the network and the nodes in each community. Moreover, you can use this data to define campaigns that target specific communities or nodes with specific roles such as leaders or followers.

Node attributes

Provides non-transactional information about the nodes. For example, this information can include the geographical location or special dates,

such as the birth date or anniversary date of a node. You can use this information to target communities such as friends and family and gift them special offers on the special dates. Based on the geographical locations, you can target communities that indicate a low-market penetration.

The non-transactional information is derived from the node attribute tables that your administrator has selected for the source data profile of the project. For more information, see "Import a Table" on page 104.

Associations with neighboring roles

Indicates how a node is associated with different roles, such as leaders and followers of its neighboring nodes. It provides information about various characteristics of a node with its neighboring roles. This information helps you identify the usage, transactions, and aggregated statistics of a node across various roles. The association of a node with its neighboring roles is computed based on the incoming and outgoing values of its transactional measures. For example, in a communications network, total incoming and outgoing call duration can be computed for each role that a node is connected to. Similarly, in a social network, the number of likes that a node receives and the number of likes that a node hits can be computed for each role that a node is connected to.

The enriched data contains the total number of roles to which a node is connected. In addition, the following list indicates the enriched information that this category produces for a node for each combination of a transactional measure and a role:

- Total incoming value of a transactional measure for a node from its neighboring nodes that have a specific role.
- Total outgoing value of a transactional measure for a node to its neighboring nodes that have a specific role.
- Maximum incoming value of a transactional measure for a node from its neighboring nodes that have a specific role.
- Maximum outgoing value of a transactional measure for a node to its neighboring nodes that have a specific role.
- Minimum incoming value of a transactional measure for a node from its neighboring nodes that have a specific role.
- Minimum outgoing value of a transactional measure for a node to its neighboring nodes that have a specific role.

Community-level statistics

Identifies the significance of a community within the network. It provides information about various characteristics of the community. This information helps you identify the usage, transactions, and aggregated statistics of a community as a whole. The community-level statistics are based on the incoming and outgoing values of transactional measures that are computed for a community.

The following list indicates the enriched information that this category computes for a community for each transactional measure:

- Total incoming value of a transactional measure for a community
- Total outgoing value of a transactional measure for a community

- Minimum incoming value of a transactional measure for a community
- Minimum outgoing value of a transactional measure for a community
- Maximum incoming value of a transactional measure for a community
- Maximum outgoing value of a transactional measure for a community
- Average incoming value of a transactional measure for a community
- Average outgoing value of a transactional measure for a community

Similarly, the following list indicates the enriched information that this category produces for each transactional measure and for nodes of a particular community that have a specific role:

- Total incoming value of a transactional measure for nodes of a community that have a specific role
- Total outgoing value of a transactional measure for nodes of a community that have a specific role
- Average incoming value of a transactional measure for nodes of a community that have a specific role
- Average outgoing value of a transactional measure for nodes of a community that have a specific role

Aggregated transactional data

Identifies the significance of a node in the network by considering its aggregated transactional summary. This information helps you identify the usage, transactions, and aggregated statistics of a node in the community. The aggregated transactional summary is based on the incoming and outgoing values of transactional measures that are computed for a node.

The following list indicates the enriched information that this category produces for a node for each transactional measure:

- Total incoming value of a transactional measure for a node
- Total outgoing value of a transactional measure for a node
- Maximum incoming value of a transactional measure for a node
- Maximum outgoing value of a transactional measure for a node
- Minimum incoming value of a transactional measure for a node
- Minimum outgoing value of a transactional measure for a node

Roles and communities over time

Provides comparative information about the roles, communities, and centrality measures of nodes that are currently available in the network. The comparative information is available for the current and previous runs of a project. For the first run, the information of the previous run will not be available. Therefore, you need to run the project at least twice to produce comparative results.

You can use the comparative information to identify the change in roles of nodes and to measure the shift in their centrality measures. For example, you can target a customer whose role has changed from leader to follower and then define a suitable campaign for that customer.

Note: Select this enrichment category for a project that is in batch mode. For more information, see "Recommendations for Node-Level Enrichment Categories Selection" on page 73.

Churn and acquisition indicators

Identifies whether a node is newly acquired, churned, or retained. This information can give you insights about the impact of the newly acquired nodes in the community and help you measure their cascading impact on business volume. You can gain similar insights for nodes that churn from a community.

Relation with churned and acquired nodes

Provides comparative information about various characteristics of a node that identify its usage, transactions, and aggregated statistics in association with churned and acquired nodes. In addition, for each node, the enriched data measures the shift in values of transactional measures and link weights. You can use this comparative information to target customers who have reached a threshold value and then define appropriate campaign strategies for them.

The comparative information is available for the current and previous runs of a project. For the first run, the information of the previous run will not be available. Therefore, you need to run the project at least twice to produce comparative results.

The enriched data contains the following information for each node:

- Percentage change in the incoming value of a transactional measure
- Percentage change in the outgoing value of a transactional measure
- Number of churned nodes with a specific role that were connected to a node
- Number of newly acquired nodes with a specific role that are connected to a node
- Number of churned nodes that were connected to a node
- Number of newly acquired nodes that are connected to a node
- Percentage change in the link weight of a node for incoming links
- Percentage change in the link weight of a node for outgoing links
- Total link weight of a node for incoming links
- Total link weight of a node for outgoing links
- An indicator that identifies whether the role of a node has changed

Note: Select this enrichment category for a project that is in batch mode. For more information, see "Recommendations for Node-Level Enrichment Categories Selection" on page 73.

For each enrichment category, the enriched data is stored as a set of variables in the node-level output table. For more information about these

variables and how to understand their naming convention, see *SAS Customer Link Analytics: Data Reference Guide*. For more information about how many variables the enriched data contains for each enrichment category, see "Formulas for Computing the Number of Node-Level Data Enrichment Variables" on page 153.

b Select the link-level enrichment categories. The following link-level enrichment categories are available for your selection:

Roles and communities
Identifies the relative significance of each pair of nodes in a link of a community. You can use the enriched data enriched to identify various communities and the nodes and links in each community. Moreover, you can use this data to define campaigns that target specific communities or nodes of a link with specific roles such as leaders or followers.

This category processes the final output that SAS Customer Link Analytics produces when a project completes one run in design or batch mode. The enriched data contains information for each link that is defined as a pair of nodes (from and to nodes). For the from and to nodes of each link, it indicates the role that is assigned to the node, the community, and the segment to which the node belongs. It also contains information about the centrality measures that are computed for the from and to nodes of each link.

Note: The segmentation information is available only if you chose community segmentation when you ran the Community Building workflow step.

Node attributes
Provides non-transactional information about the nodes of a link. For example, if you are visualizing a network diagram, this information can be displayed as tooltips for each of the nodes.

The non-transactional information is derived from the node attribute tables that your administrator has selected for the source data profile of the project. For more information, see "Create a Source Data Profile" on page 113. The enriched data typically contains non-transactional information for the from node and the to node of each link.

Churn and acquisition indicators
Identifies whether a link is newly acquired, churned, or retained. This information can give you insights about the impact of the newly acquired links on the community and help you measure their cascading impact on business volume. You can gain similar insights for links that churn from a community. Also, you can get information about the existing nodes that are connected to the newly acquired nodes and those that were connected to the nodes that churned out. You can use this information to enhance campaign definitions and target appropriate customers.

For each enrichment category, the enriched data is stored as a set of variables in the link-level output table. For more information about these variables and how to understand their naming convention, see *SAS Customer Link Analytics: Data Reference Guide*.

4 Click **Save**. The window closes. If you modify the default selection, the status changes to **Default selection modified**. You can now proceed to enrich the project's output data.

Note: At any point of time, you can modify the enrichment categories that you have selected. However, you can do so only when the data enrichment process is not being run and the project's status is **Successfully executed**. If the project has any other status, you can view the enrichment categories that you have selected. However, you cannot modify it.

Enrich the Output Data of a Project

When you enrich the project's output data, SAS Customer Link Analytics stores the node-level and link-level data in separate business data tables. If you enrich data by using the default enrichment categories, these tables contain the final output of SAS Customer Link Analytics at node level and link level. The data that is enriched by other categories is added to this base information.

To enrich the project's output data:

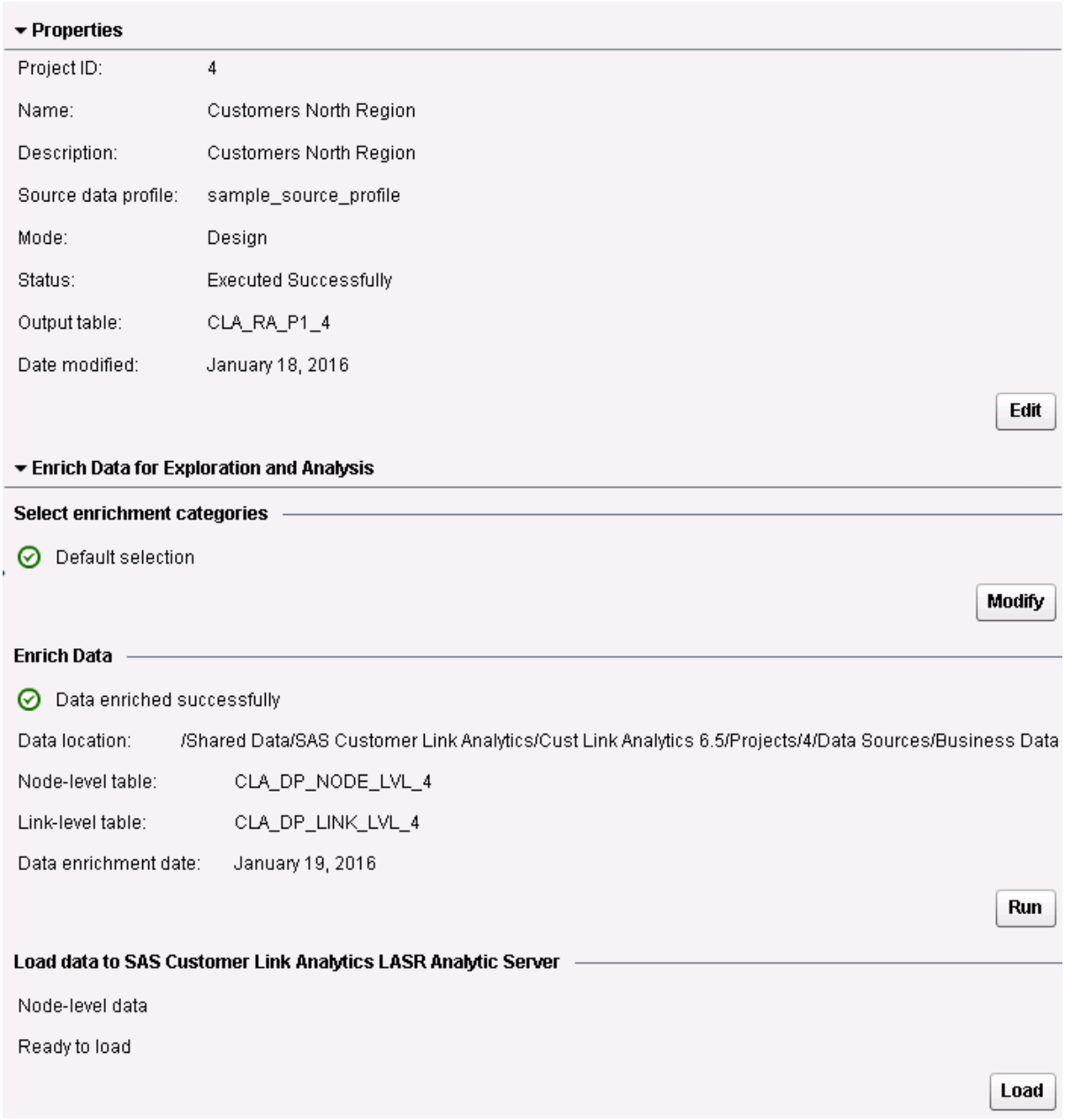
1 In the Projects workspace, select the project whose data you want to enrich.

2 In the Data Enrichment for Exploration and Analysis pane, click **Run**.

3 In the confirmation dialog box that appears, click **Yes**. The status changes from **Ready to enrich** to **Data enrichment in progress**.

4 On the toolbar, select to view the updated status of the data enrichment process.

Figure 11.2 Successful Data Enrichment



Note:

- You must click at regular intervals to see the updated status of the process.
- If the data enrichment process runs for a long time without a status update, contact your administrator for assistance. For more information, see *SAS Customer Link Analytics: Administrator's Guide*.

When the data enrichment is successful, the following information is displayed:

- The status changes to **Data enriched successfully**. This status indicates that the node-level and link-level data is enriched and stored in separate business data tables that are created for the project. Moreover, these tables are registered in SAS metadata.

- The metadata location in which the business data tables are registered is displayed.
- The node-level and link-level business tables in which the enriched data is stored are displayed.
- The date on which you last enriched the project's output data is displayed.

For more information about the enriched data that is produced for each data enrichment category, see "Using Node-Level Enriched Data" on page 159 and "Using Link-Level Enriched Data" on page 169.

You can now proceed to load the enriched data to the SAS Customer Link Analytics LASR Analytic Server.

TIP If the data enrichment fails, click **View Log**. An appropriate error message is displayed. Correct the error and click **Run** again. For more information about the log file, see *SAS Customer Link Analytics: Administrator's Guide*.

Load Enriched Data to the SAS Customer Link Analytics LASR Analytic Server

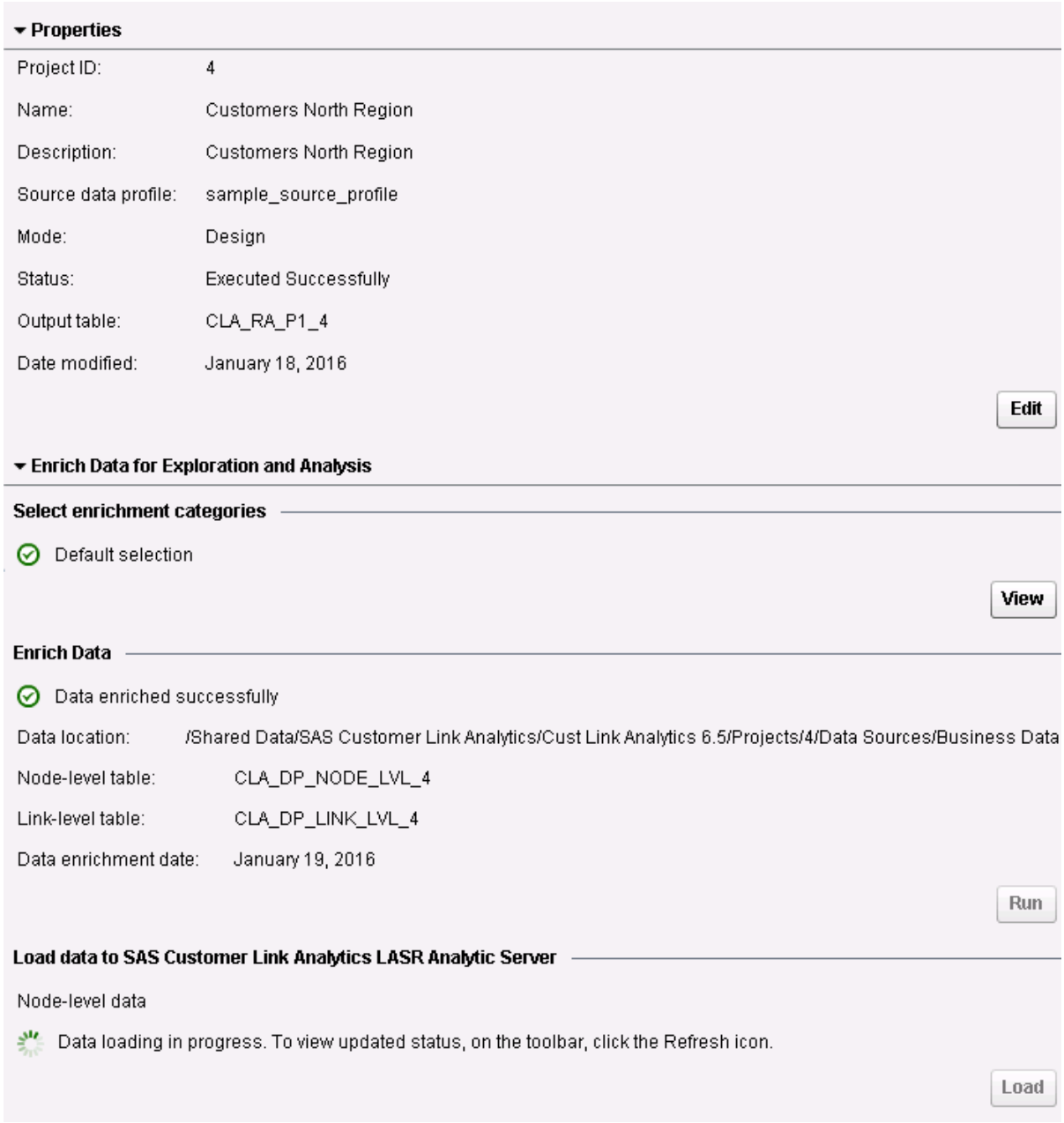
After you complete the data enrichment process, SAS Customer Link Analytics enables you to copy the enriched data to the SAS Customer Link Analytics LASR Analytic Server. SAS Customer Link Analytics uses this data to create the Community report. For more information, see "About the Community Report" on page 85.

When the enriched data is loaded into the SAS Customer Link Analytics LASR Analytic Server, it is copied into the CLA_ND_LVL_LASR_<*Project ID*> table. This table is registered in the following SAS metadata location: `<Project path>/<Project ID>/Data Sources/Report Data`. For example, the project path can be: `/Shared Data/SAS Customer Link Analytics/Cust Link Analytics 6.5/Projects`. For more information about the project path, see *SAS Customer Link Analytics 6.5: Administrator's Guide*.

To load the enriched data into the SAS Customer Link Analytics LASR Analytic Server:

1. In the Projects workspace, select the project whose enriched data you want to load into the SAS Customer Link Analytics LASR Analytic Server.
2. Click **Load**.
3. In the confirmation dialog box that appears, click **Yes**. The status changes from **Ready to load** to **Data loading in progress**.

Figure 11.3 Loading Data into the SAS Customer Link Analytics LASR Analytic Server



4 On the toolbar, select to view the updated status of the data loading process.

Note: You must click at regular intervals to see the updated status of the process.

When the data loading process is successful, the following tasks are performed:

- The status changes from **Data loading in progress** to **Data loading successful**.
- The SAS metadata location in which the table that stores the enriched data is registered is displayed.
- The name of the table in which the enriched data is stored is displayed.

- The date on which the enriched data was last loaded into the SAS Customer Link Analytics LASR Analytic Server is displayed.

TIP If the data loading process fails, click **View Log**. An appropriate error message is displayed. Correct the error and click **Load** again. For more information about the log file, see *SAS Customer Link Analytics: Administrator's Guide*.

12

Generating the Community Report

About the Community Report

When SAS Customer Link Analytics loads the enriched data into the SAS Customer Link Analytics LASR Analytic Server successfully, you can create the Community report. This report gives summarized information about the communities that are created for the project.

The Community report is registered in the following SAS metadata location: **`<Project Path>/<Project ID>/Reports`**. For example, the *<Project Path>* can be **`/Shared Data/SAS Customer Link Analytics/Cust Link Analytics 6.5/Projects`**. For more information about the project path, see *SAS Customer Link Analytics: Administrator's Guide*.

The information that is populated in the Community report is organized on various tabs. The report information is populated depending on the node-level enrichment categories that you choose. For more information, see "Select an Enrichment Category for a Project" on page 73. The basic Community report contains information based on the default selection for the node-level enrichment category. If you select one or more other node-level enrichment categories, the report contains more detailed information. The number of tabs that are displayed categorizes the Community report into the following types.

Table 12.1 Types of Community Report

Selection of Node-Level Enrichment Categories	Report Filename	Report Information
■ Roles and communities	CLA_sample_default_node_level_rpt	Displays the following information: ■ Distribution of various roles in the network ■ Top ten percentage of communities with maximum size ■ Bottom ten percentage of communities with minimum size ■ Distribution of centrality measures across various roles For more information, see "The Default Community Report" on page 89.
■ Roles and communities ■ Churn and acquisition indicators	CLA_sample_derived_indicator_node_level_rpt	Displays the following information: ■ Distribution of various roles in the network ■ Top ten percentage of communities with maximum size ■ Bottom ten percentage of communities with minimum size ■ Distribution of centrality measures across various roles ■ Top ten percentage of communities by newly acquired nodes ■ Bottom ten percentage of communities by newly acquired nodes For more information, see "Community Report for Churned and Acquired Indicators" on page 90.

Selection of Node-Level Enrichment Categories	Report Filename	Report Information
■ Roles and communities ■ Relation with churned and acquired nodes	CLA_sample_acquisition_churn_node_level_rpt	Displays the following information: ■ Distribution of various roles in the network ■ Top ten percentage of communities with maximum size ■ Bottom ten percentage of communities with minimum size ■ Distribution of centrality measures across various roles ■ Top ten percentage of communities with maximum role changes for its nodes ■ Top ten percentage of communities with nodes that had maximum connections with churned nodes For more information, see "Community Report for Relation with Churned and Acquired Nodes" on page 91.
■ Roles and communities ■ Churn and acquisition indicators ■ Relation with churned and acquired nodes	CLA_sample_acquisition_churn_and_derived_node_level_rpt	Displays the following information: ■ Distribution of various roles in the network ■ Top ten percentage of communities with maximum size ■ Bottom ten percentage of communities with minimum size ■ Distribution of centrality measures across various roles ■ Top ten percentage of communities by newly acquired nodes ■ Bottom ten percentage of communities by newly acquired nodes ■ Top ten percentage of communities with maximum role changes for its nodes ■ Top ten percentage of communities with nodes that had maximum connections with churned nodes For more information, see "Community Report for Relation with Churned and Acquired Nodes and Their Indicators" on page 92.

TIP SAS Customer Link Analytics enables you to view the Community report in SAS Visual Analytics Viewer. You can explore the report information at various levels and export the required information in a Microsoft Excel workbook or print it as a PDF file.

Create the Community Report

To create the Community report:

1. In the Projects workspace, select the project for which you want to create the report.
2. On the toolbar, click , and then select **Create Report**.

   Note: This option is activated only if the following conditions are fulfilled:

   - The data enrichment process has run successfully.
   - The node-level data is loaded into the SAS Customer Link Analytics LASR Analytic Server successfully.
   - The data enrichment process was completed before the data loading process.

   The icon indicates that the report is being created. The Community Report dialog box appears when the report has been created and registered in the SAS metadata successfully.
3. Click **OK**.

View the Community Report

After you create the Community report for a project, you can open and view it in SAS Visual Analytics Viewer.

To view a Community report:

1. In the Projects workspace, select the project whose Community report you want to view.
2. On the toolbar, click , and then select **View Report**. The report is displayed in SAS Visual Analytics Viewer depending on the node-level enrichment categories that you choose. For more information, see “Types of Community Reports” on page 89.

   Note: The **Report** menu is activated for projects for which node-level data has been successfully loaded into the SAS Customer Link Analytics LASR Analytic Server.
3. To return to the Projects workspace, on the banner, click **SAS Customer Link Analytics**.

Types of Community Reports

Overview

The information in the Community report is populated on various tabs depending on the node-level enrichment categories that you choose. For more information see, Step 3a on page 74.

The following topics explain the types of Community reports depending on the information that it contains.

The Default Community Report

The default Community report helps you analyze the role distribution in the network. In addition, you can analyze the communities that have the maximum and minimum community sizes. Also, you can explore the distribution of the default centrality measures across various roles of the community.

This Community report is created when you run the data enrichment process for the default node-level enrichment category, **Roles and communities**.

The report information is displayed on the following tabs:

Role Distribution In Network
displays a pie chart that indicates the percentage distribution of various roles in the network. Each slice of the pie chart indicates a role. You can select a role and explore that role further.

Top Communities by Size
displays the list of the top ten percentage of communities in descending order of their community size. You can select each community and further view its role distribution as a pie chart. You can select a role and view the list of nodes that have this role. For each node, you can view the values that are computed for the default centrality measures.

Bottom Communities by Size
displays the list of the bottom ten percentage of communities in ascending order of their community size. You can select each community and further view its role distribution as a pie chart. You can select a role and view the list of nodes that have this role. For each node, you can view the values that are computed for the default centrality measures.

Centrality Measure by Role
displays the distribution of the default centrality measures across various roles of the network. The distribution of each centrality measure is displayed as a separate bar chart. For the influence centrality, you can view the comparative values for the incoming and outgoing influence of the nodes with their neighbors. Moreover, if you select a role in the bar chart, you can view the distribution of the centrality measure across the nodes.

Community Report for Churned and Acquired Indicators

In addition to the information that the default report produces, the Community report for churned and acquired indicators enables you to analyze communities based on the interactions of their existing nodes with the newly acquired nodes. You can identify communities that have maximum and minimum interactions with the newly acquired nodes.

This Community report is created when you run the data enrichment process for the following node-level enrichment categories: **Roles and communities** and **Churned and acquired indicators**.

The report information is displayed on the following tabs:

Role Distribution in Network
displays a pie chart that indicates the percentage distribution of various roles in the network. You can select a role and view the distribution of newly acquired nodes and the churned nodes in separate pie charts.

Note: In the pie chart, the missing role indicates the nodes that have churned out of the network.

Top Communities by Size
displays the list of the top ten percentage of communities in descending order of their community size. You can select each community and further view its role distribution as a pie chart. You can select a role and view its distribution of newly acquired nodes in another pie chart. Moreover, for each role, you can view the list of nodes that have this role. From this list, you can identify the nodes that are newly acquired. These nodes have the acquisition indicator value of 1.

Bottom Communities by Size
displays the list of the bottom ten percentage of communities in ascending order of their community size. You can select each community and further view its role distribution as a pie chart. You can select a role and view its distribution of newly acquired nodes in another pie chart. Moreover, for each role, you can view the list of nodes that have this role. From this list, you can also identify the nodes that are newly acquired. These nodes have the acquisition indicator value of 1.

Centrality Measure by Role
displays the distribution of the default centrality measures across various roles of the network. The distribution of each centrality measure is displayed as a separate bar chart. For the influence centrality, you can view the comparative values for the incoming and outgoing influence of the nodes with their neighbors. Moreover, if you select a role in the bar chart, you can view the distribution of the centrality measure across the nodes.

Top Communities by Acquired Nodes
displays the list of the top ten percentage of communities whose nodes have maximum connections with the newly acquired nodes. For each community in the list, you can see its size and the number of newly acquired nodes. You can select each community and further view its role distribution as a pie chart. You can select a role and view its distribution of newly acquired nodes in another pier chart. Moreover, for each role, you can view the list of nodes that have this role. From this list, you can also identify the nodes that are newly acquired. These nodes have the acquisition indicator value of 1.

Bottom Communities by Acquired Nodes
displays the list of the bottom ten percentage of communities whose nodes are least connected with the newly acquired nodes. For each community in the list, you can see its size and the number of newly acquired nodes. You can select each community and further view its role distribution as a pie chart. You can select a role and view its distribution of newly acquired nodes in a separate pie chart. Moreover, for each role, you can view the list of nodes that have this role. From this list, you can also identify the nodes that are newly acquired. These nodes have the acquisition indicator value of 1.

Community Report for Relation with Churned and Acquired Nodes

In addition to the information that the default report produces, the Community report for relation with churned and acquired nodes enables you to analyze communities whose nodes indicate a maximum role change. It also provides information about the topmost communities whose nodes had a strong affinity with nodes that churned out.

This Community report is created when you run the data enrichment process for the following node-level enrichment categories: **Roles and communities** and **Relation with churned and acquired nodes**.

The report information is displayed on the following tabs:

Role Distribution in Network
displays a pie chart that indicates the percentage distribution of various roles in the network.

Note: In the pie chart, the missing role indicates the nodes that have churned out of the network.

Top Communities by Size
displays the list of the top ten percentage of communities in descending order of their community size. You can select each community and further view its role distribution as a pie chart. You can select a role and view the list of nodes that have this role.

Bottom Communities by Size
displays the list of the bottom ten percentage of communities in ascending order of their community size. You can select each community and further view its role distribution as a pie chart. You can select a role and view the list of nodes that have this role.

Centrality Measure by Role
displays the distribution of various roles in the network as a bar chart. Also, it displays the distribution of the default centrality measures across various roles of the network. The distribution of each centrality measure is displayed as a separate bar chart. For the influence centrality, you can view the comparative values for the incoming and outgoing influence of the nodes with their neighbors. Moreover, if you select a role in the bar chart, you can view the distribution of the centrality measure across the nodes.

Top Communities with Role Change Nodes
displays the list of the top ten percentage of communities whose nodes have maximum role changes. You can select each community and further view its role distribution as a pie chart. You can select a role and view its distribution of nodes that have a role change in a separate pie chart. Moreover, for each role, you can view the list of nodes that have this role. From this list, you can

identify the nodes whose role has changed. These nodes have the role change indicator value of 1.

Top Communities with Connection to Churned Nodes
provides the list of the top ten percentage of communities whose nodes had maximum connections with the churned nodes. You can select each community and view its role distribution as a treemap. The role that has the maximum size indicates that nodes with this role had maximum connections with the churned nodes. For each role, you can further view the list of nodes that have this role. For each node, you can view the number of churned nodes that it was connected to.

Community Report for Relation with Churned and Acquired Nodes and Their Indicators

The Community report for relation with churned and acquired nodes and their indicators gives you comprehensive information that is provided by all three other Community reports.

This Community report is created when you run the data enrichment process for the following node-level enrichment categories: **Roles and communities**, **Churned and acquired indicators**, and **Relation with churned and acquired nodes**.

The report information is displayed on the following tabs:

Role Distribution in Network
displays a pie chart that indicates the percentage distribution of various roles in the network. You can select a slice of the pie chart and explore that role further. For each role, this tab further gives the distribution of newly acquired nodes and the churned nodes in separate pie charts.

Note: In the pie chart, the missing role indicates the nodes that have churned out of the network.

Top Communities by Size
displays the list of the top ten percentage of communities in descending order of their community size. You can select each community and further view its role distribution as a pie chart. You can select a role and view its distribution of newly acquired nodes in another pie chart. Moreover, for each role, you can view the list of nodes that have this role. From this list, you can identify the nodes that are newly acquired. These nodes have the acquisition indicator value of 1.

Bottom Communities by Size
displays the list of the bottom ten percentage of communities whose nodes are least connected with the newly acquired nodes. You can select each community and further view its role distribution as a pie chart. You can select a role and view its distribution of newly acquired nodes in a separate pie chart. Moreover, for each role, you can view the list of nodes that have this role. From this list, you can also identify the nodes that are newly acquired. These nodes have the acquisition indicator value of 1.

Centrality Measure by Role
displays the distribution of the default centrality measures across various roles of the network. The distribution of each centrality measure is displayed as a separate bar chart. For the influence centrality, you can view the comparative values for the incoming and outgoing influence of the nodes with

their neighbors. Moreover, if you select a role in the bar chart, you can view the distribution of the centrality measure across the nodes.

Top Communities by Acquired Nodes
displays information about the top ten percentage of communities whose nodes have maximum connections with the newly acquired nodes. You can select each community and further view its role distribution. For each role, you can view the distribution of newly acquired nodes. Moreover, for each role, you can view the node details. You can also identify the nodes that are newly acquired. These nodes have the acquisition indicator value of 1.

Bottom Communities by Acquired Nodes
displays information about the bottom ten percentage of communities whose nodes are the least connected with the newly acquired nodes. You can select each community and further view its role distribution. For each role, you can view the distribution of newly acquired nodes. Moreover, for each role, you can view the node details. You can also identify the nodes that are newly acquired. These nodes have the acquisition indicator value of 1.

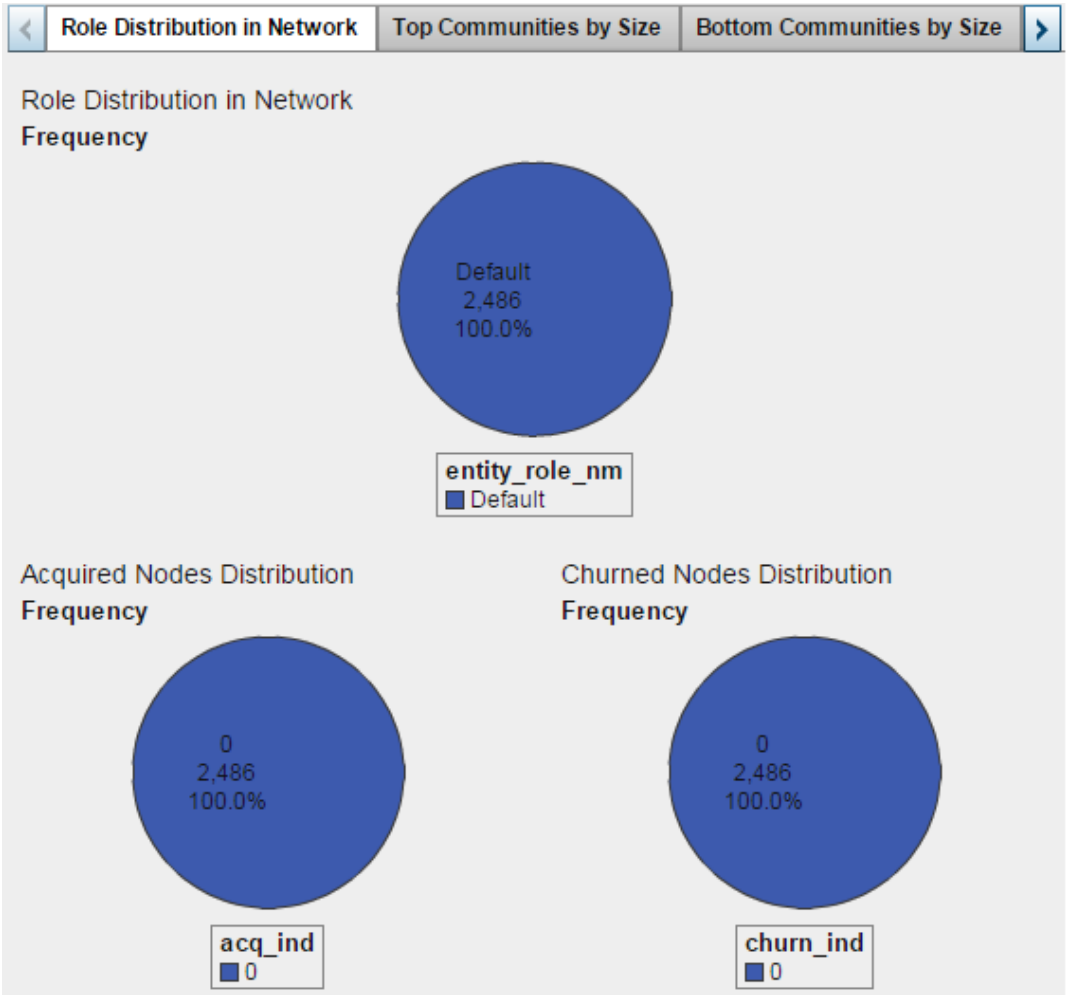
Top Communities with Role Change Nodes
displays the list of the top ten percentage of communities whose nodes have maximum role changes. You can select each community and further view its role distribution as a pie chart. You can select a role and view its distribution of nodes that have a role change in a separate pie chart. Moreover, for each role, you can view the list of nodes that have this role. From this list, you can identify the nodes whose role has changed. These nodes have the role change indicator value of 1.

Top Communities with Connection to Churned Nodes
provides the list of the top ten percentage of communities whose nodes had maximum connections with the churned nodes. You can select each community and view its role distribution as a treemap. The role that has the maximum size indicates that nodes with this role had maximum connections with the churned nodes. For each role, you can further view the list of nodes that have this role. For each node, you can view the number of churned nodes that it was connected to.

Sample Tabs of the Community Report

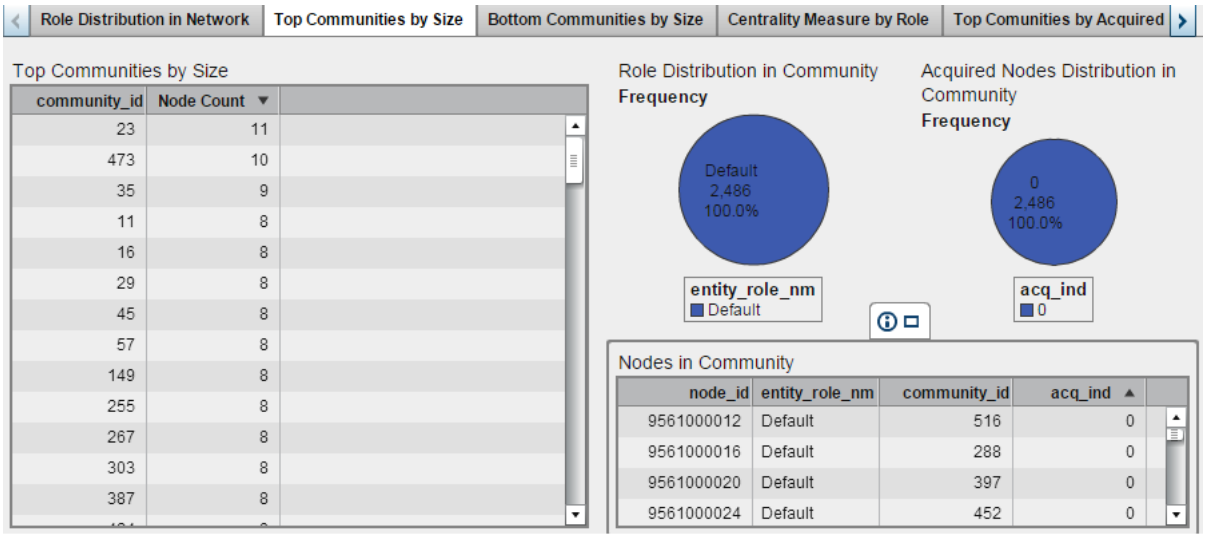
Role Distribution in Network

Figure 12.1 *Role Distribution in Network Tab*



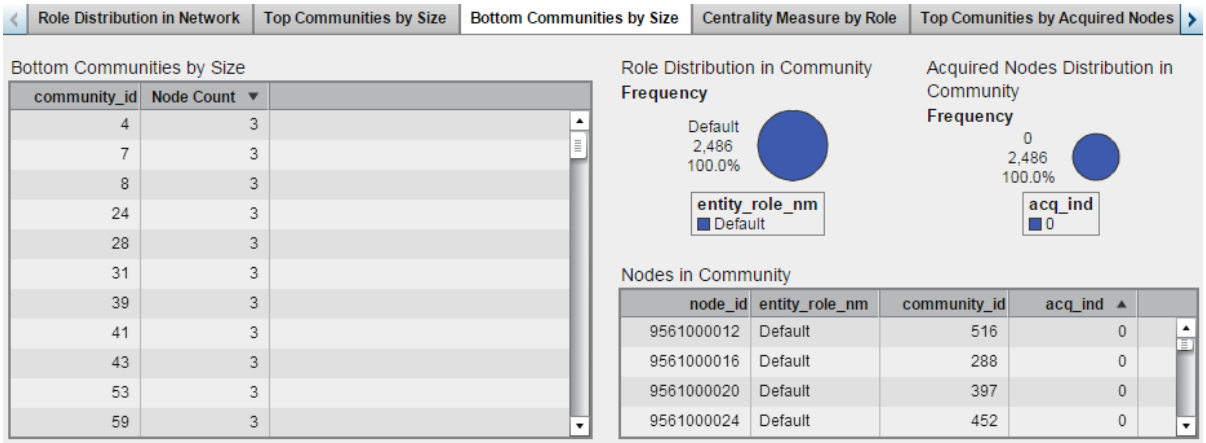
Top Communities by Size

Figure 12.2 Top Communities by Size Tab



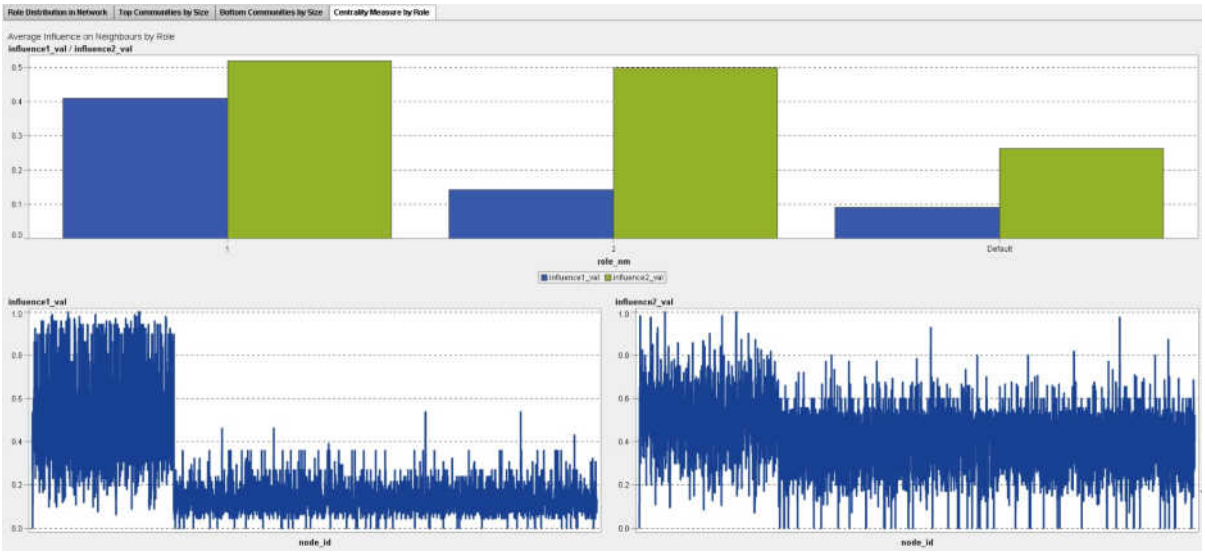
Bottom Communities by Size

Figure 12.3 Bottom Communities by Size Tab



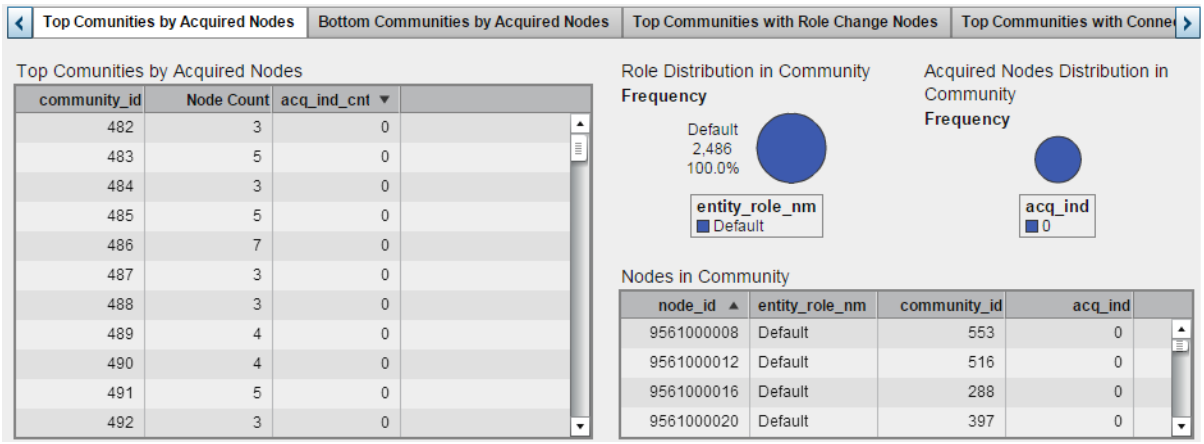
Centrality Measure by Role

Figure 12.4 *Centrality Measure by Role*



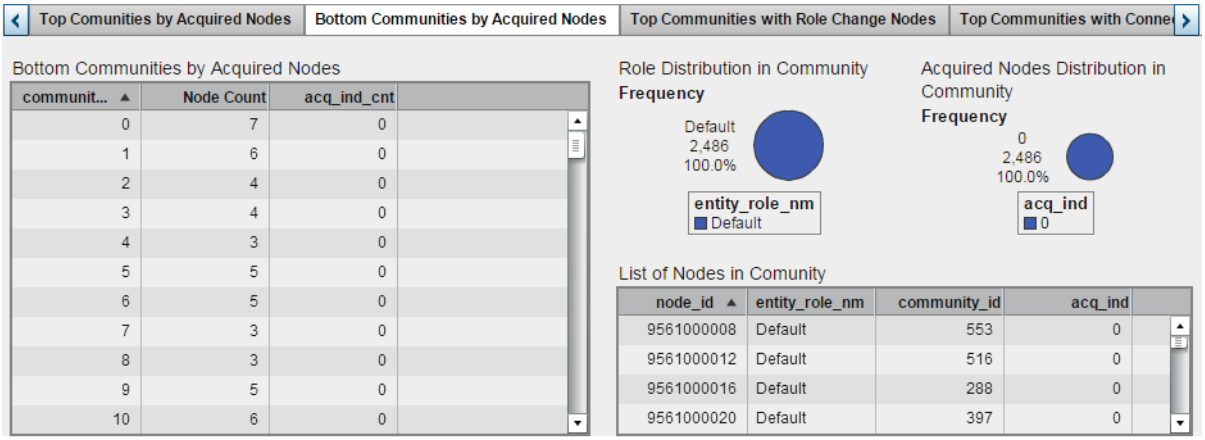
Top Communities by Acquired Nodes

Figure 12.5 *Top Communities by Acquired Nodes Tab*



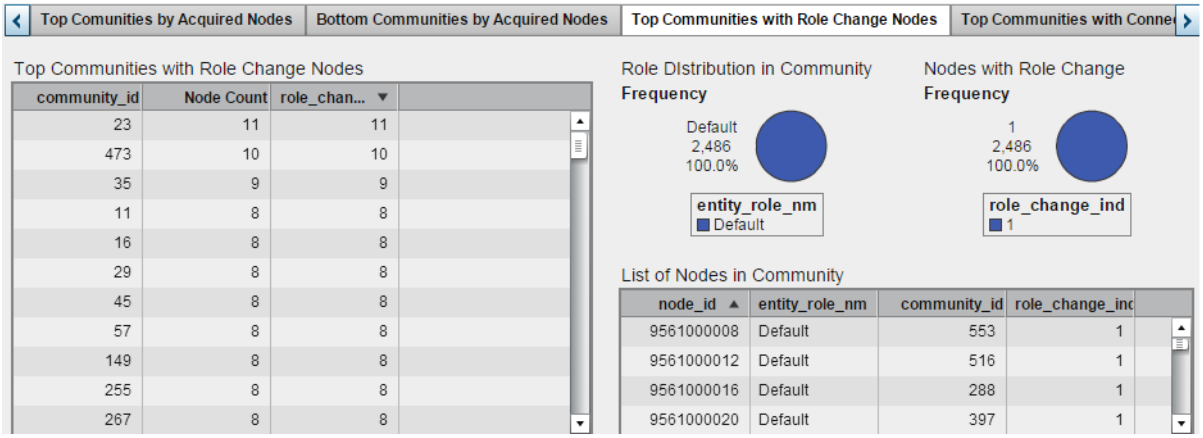
Bottom Communities by Acquired Nodes Tab

Figure 12.6 Bottom Communities by Acquired Nodes Tab



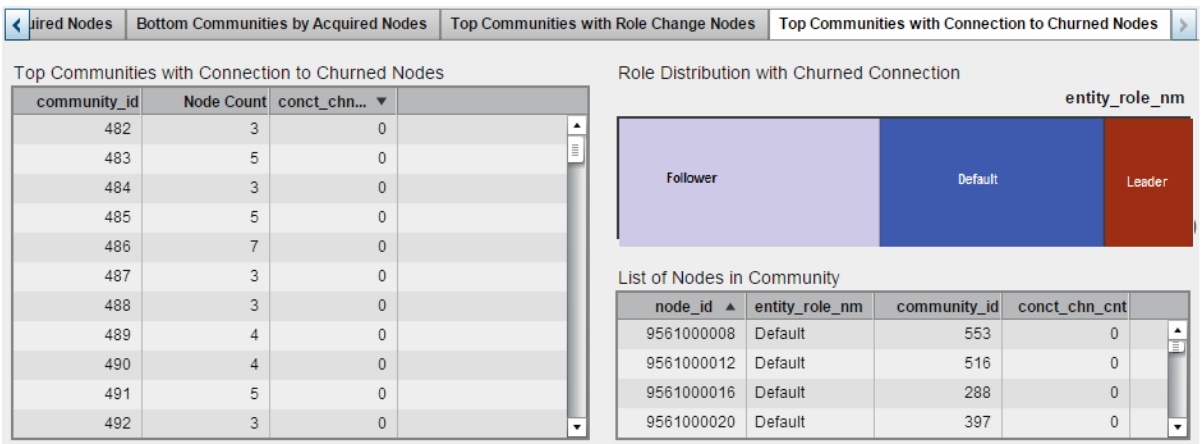
Top Communities with Role Change Nodes

Figure 12.7 Top Communities with Role Change Nodes Tab



Top Communities with Connection to Churned Nodes

Figure 12.8 Top Communities with Connection to Churned Nodes



Recommendations for Re-Running Processes

Re-running the Data Enrichment Process

Each time you modify the selection of enrichment categories, you have to re-run the data enrichment process. Otherwise, the enriched data will not contain the information according to the latest selection of enrichment categories. SAS Customer Link Analytics displays the ⚠ icon to indicate that the enriched data is out of date and you have to re-run the data enrichment process.

Note: Assume that you have created a Community report based on your earlier enrichment categories selection. If you modify the enrichment categories, you will still see this report. You will not be able to see the new Community report unless you re-run the data enrichment process, reload the data to the SAS Customer Link Analytics LASR Analytic Server, and re-create the report.

Reloading Node-Level Data

Each time you run the data enrichment process, you must load the node-level data to the SAS Customer Link Analytics LASR Analytic Server. To ensure that you have the loaded the latest enriched data to the server, you can compare the **Data enrichment date** and the **Data loading date**. If the **Data loading date** is earlier than the **Data enrichment date**, you have to reload the data to the SAS Customer Link Analytics LASR Analytic Server. In addition, SAS Customer Link Analytics displays the ⚠ icon to indicate that the data that you have loaded to the SAS Customer Link Analytics LASR Analytic Server is not in synchronization with the enriched data. In this case, you have to reload the latest enriched data to SAS LASR Analytic Server.

Re-creating the Community Report

Each time you reload the node-level data to the SAS Customer Link Analytics LASR Analytic Server, you have to re-create the Community report. Otherwise, the report will not be generated based on the latest data. In this scenario, to ensure that you have the correct Community report, SAS Customer Link Analytics forces you to re-create the report and deactivates the **Open report** option. Also, assume that you modify the enrichment categories and do not re-run the data enrichment process or reload the data to the SAS Customer Link Analytics LASR Analytic Server. In this scenario, the report that you view, will be based on your previous selection of enrichment categories. Therefore, it might show incorrect or insufficient information.

Part 4

Administration Workspace

13

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 more information, see “Import a Table” on page 104.

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 transactional measures. For more information, see “Configure the Column Values of a Table” on page 106.

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 more information, see “Create a Source Data Profile” on page 113.

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 more information, see “Refresh Data in a Transactional Table” on page 108.

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

Also, make sure that the table name and the column names of that table do not contain any special characters or DBCS characters. Otherwise, workflow steps of projects that use these tables will not run successfully.

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 13.1 *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, from node ID (from), or to node ID (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 13.2 *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. For more information about the structure of transactional tables, see *SAS Customer Link Analytics: Administrator's Guide*.

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 either be a list of nodes (also called a node inclusion list) or a list of links (also called a link inclusion list). 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 more information, see "Overview of Inclusion Lists" on page 29.

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, from node ID (from), or to node ID (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 13.3 Mandatory Column Mapping for Node Inclusion List and Node Attribute Table

Column Name	Type
Node ID	From

Table 13.4 Mandatory Column Mappings for Link Inclusion List and Link Attribute Table

Column Name	Type
From Node ID	From
To Node ID	To

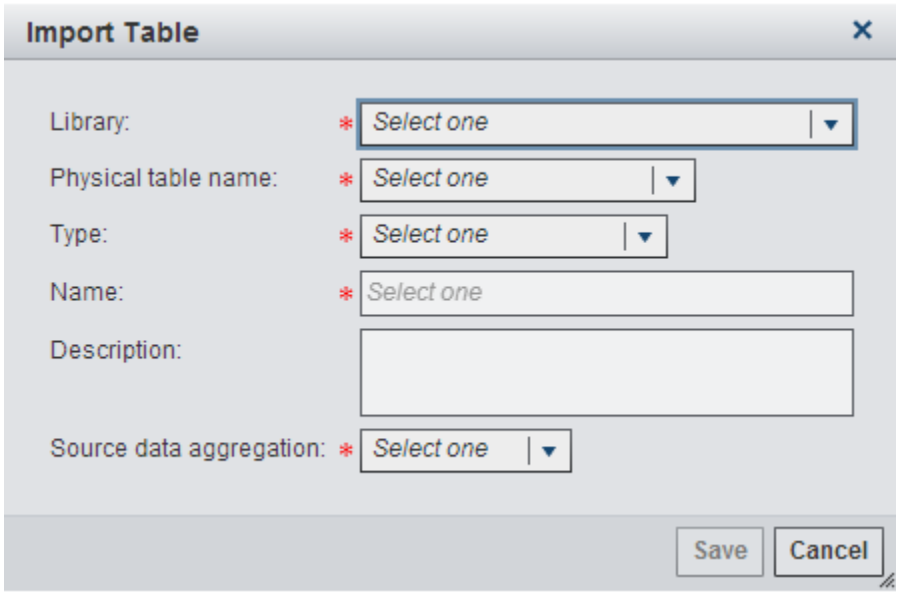
For more information about the structure of attribute tables, see *SAS Customer Link Analytics: Administrator's Guide.*

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.

Figure 13.1 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. For more information about the structure of tables that you can import, see *SAS Customer Link Analytics: Administrator's Guide*.

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. SAS Customer Link Analytics does not use this table for generating reports. However, you can use this table in external systems by combining the link-level information that it stores with some specific data that is stored in the external system.

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 more information, see "Example 2: Link Inclusion List" on page 30.

Node attribute
indicates that the table stores information about nodes. SAS Customer Link Analytics uses this table for generating reports that contain node-level information.

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 more information, see "Example 1: Node Inclusion List" on page 29.

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.

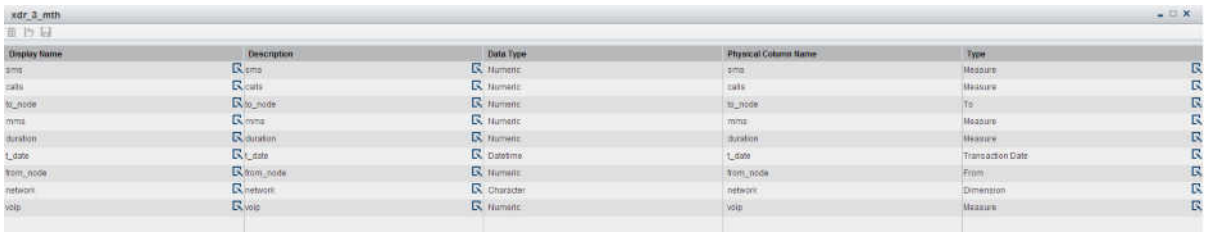
Configure the Column Values of a Table

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.

Figure 13.2 *Columns of a Table*



The following details are displayed for each column:

Column ID
indicates the unique ID that is assigned to the column. This ID is used in the naming convention of variables that are created when you run the data enrichment process. For more information about these variables, see *SAS Customer Link Analytics: Data Reference Guide*.

You cannot change the value of this field.

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. For example, you can use this data type for columns that store the birth date and the anniversary date.

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. This column is called a transactional measure. For example, a communications network, call duration or number of calls be transactional measures. Similarly, for a social network, number of likes can be a transactional measure.

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 an 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.

Process date
indicates that the column contains a date value that is used for auditing purposes. These columns are not used for any further analysis. For example, you can use this data type for columns that indicate when a record was created or last processed in the table.

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 13.5 *Mandatory Column Types*

Table Type	Mandatory Column Type
Transactional	■ Transaction date ■ From ■ To ■ Measure

Table Type	Mandatory Column Type
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 13.5 on page 107.

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 more information, see “Configure the Data Extraction Workflow Step” on page 27.

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 more information, see “Configure the Column Values of a Table” on page 106. 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.

Refresh a Transactional Table and a Project for Incremental Data

Assume that you load incremental data in the transactional table. In this case you have to refresh the transactional table. Otherwise, the start and end dates that are displayed for the transactional table will not be synchronized with the actual data. To do so, refresh the data in the transactional table and confirm that the start and end dates are updated. For more information, see "Refresh Data in a Transactional Table" on page 108.

After you refresh the transactional table with incremental data, you have to also refresh the projects that are in design mode and are associated with this table. Otherwise, the default end data that is populated in the Data Extraction workflow step will not be updated. As a result, this workflow step will extract data for the incorrect historical period. For more information, see "Configure the Data Extraction Workflow Step" on page 27.

To refresh a project for incremental data:

1. In the Projects workspace, double-click the project that is associated with the transactional table that you have refreshed for incremental data. .
2. In the workflow diagram pane, select **Data Extraction**.
3. Verify that you have not changed the end date that is populated by default.
4. On the toolbar, click  .
5. Make sure that end date is updated according to the end date that is displayed for the transactional table in the Administration workspace.

   Note: If you have changed the default end date, then you will not notice any change in the date.
6. Repeat steps from 1 through 5 for the other projects that are associated with the transactional table.

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.

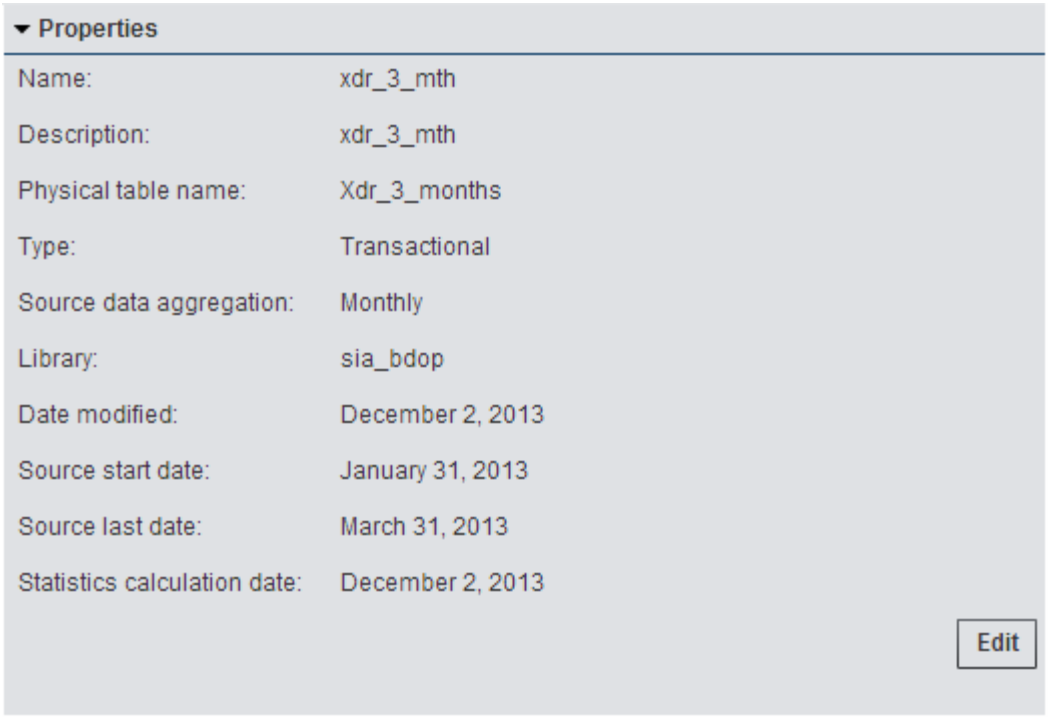
To edit the properties of a table:

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**.

Figure 13.3 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.

Figure 13.4 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

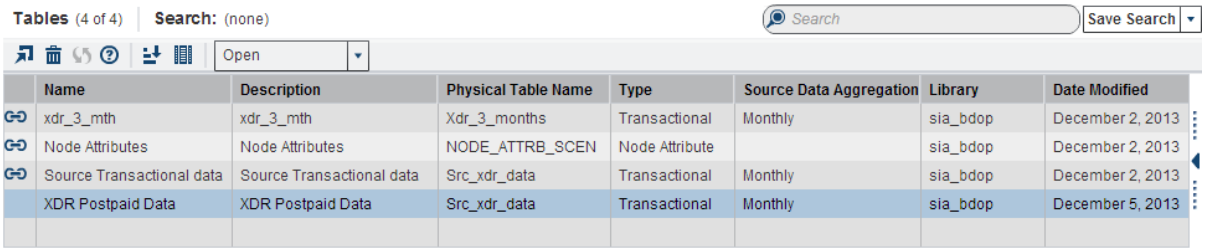
You can delete a table if it is not used in the definition of a source data profile. For more information, see “Create a Source Data Profile” on page 113.

To delete a table:

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

2 Select the table that you want to delete.

Figure 13.5 Delete Table



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.

14

Working with Source Data Profiles

Overview of Source Data Profiles

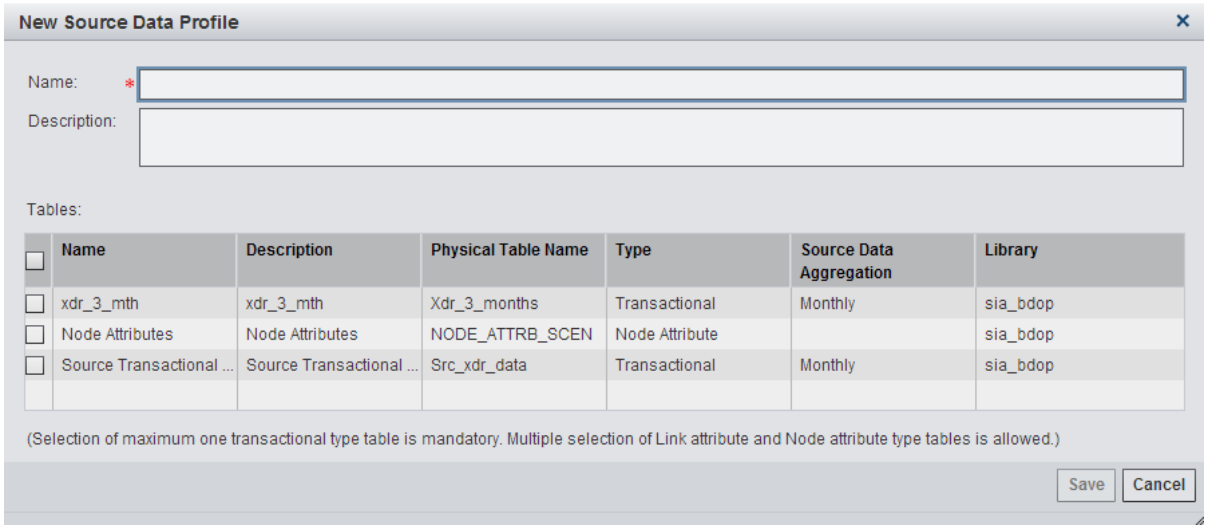
After you import tables and configure their columns, you can group a set of tables to define a source data profile. A source data profile comprises a transactional table and one or more attribute tables. These tables are the ones that you have imported. For more information, see “Import a Table” on page 104. You have to associate each project with a source data profile. For more information, see “Overview of Projects” on page 15. 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.

Figure 14.1 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 more information, see “Import a Table” on page 104.

You must select only one transactional table. Alternatively, you can decide whether you want to select any attribute tables. You can select only one attribute table of the **Node Attribute** type. However, there is no restriction on the other attribute tables that you can include. These types include the following: **Link Attribute**, **Link Inclusion List**, and **Node Inclusion List**. For more information, see “Table Types” on page 102.

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

Edit a Source Data Profile

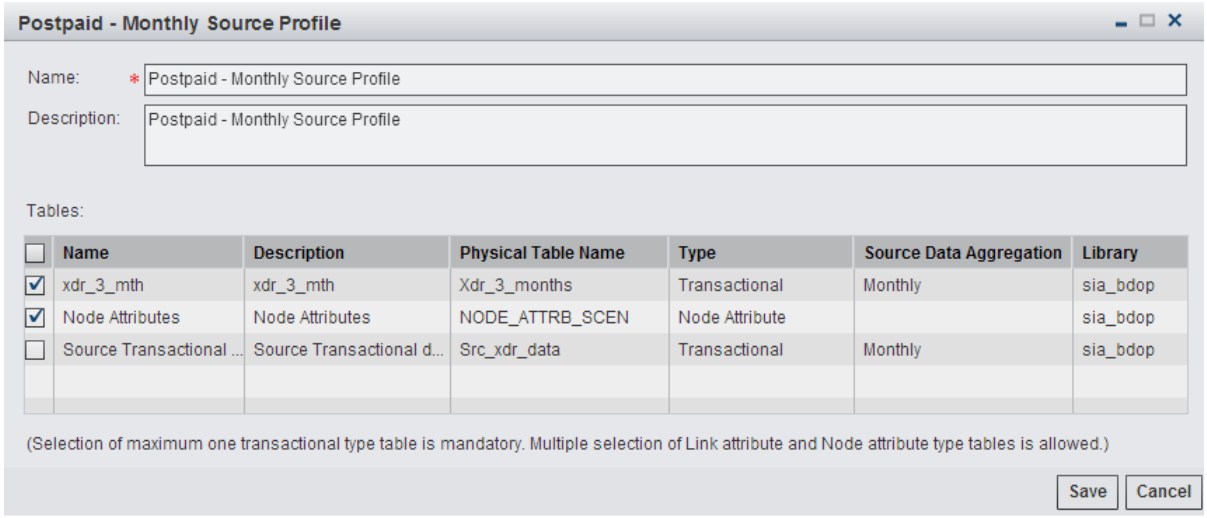
When you create a source data profile, you can view its details in the Properties pane. You can change the name and description of the source data profile. To do so, in the Properties pane, click **Edit**. In addition, you can change the tables that you have included in the source data profile if it 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.

Figure 14.2 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.

  Note:

  - You cannot deselect an attribute table of the **Inclusion list** type if you have selected this table in the Data Extraction workflow step of a project that is associated with this source data profile. For more information, see "Configure the Data Extraction Workflow Step" on page 27.
  - You cannot deselect an attribute table of the **Node attribute** type if you have selected the **Node attributes** enrichment category in the data enrichment process. For more information, see "Select an Enrichment Category for a Project" on page 73.
- 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 more information, see "Create a Project" on page 23.

To delete a source data profile:

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

Figure 14.3 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

15

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 adoption or reduced customer churn. The ultimate goal of business analysts is to increase revenue by defining business strategies based on the results of viral effect analysis. Viral effect analysis uses the transactional data of links that are formed by direct interactions between customers.

SAS Customer Link Analytics provides network analysts a step-by-step analytical process to perform viral effect analysis for the following events:

Churn
: SAS Customer Link Analytics predicts the probability of a node to churn from the network based on the connections that it has with nodes that have already churned from the network. As a result, SAS Customer Link Analytics helps business analysts define appropriate business strategies to reduce customer churn.

Product adoption
: SAS Customer Link Analytics predicts the probability of a node to adopt a particular product based on the connections that it has with nodes that have already adopted this product. As a result, SAS Customer Link Analytics helps business analysts define appropriate business strategies to increase product adoption.

Prerequisites for Viral Effect Analysis

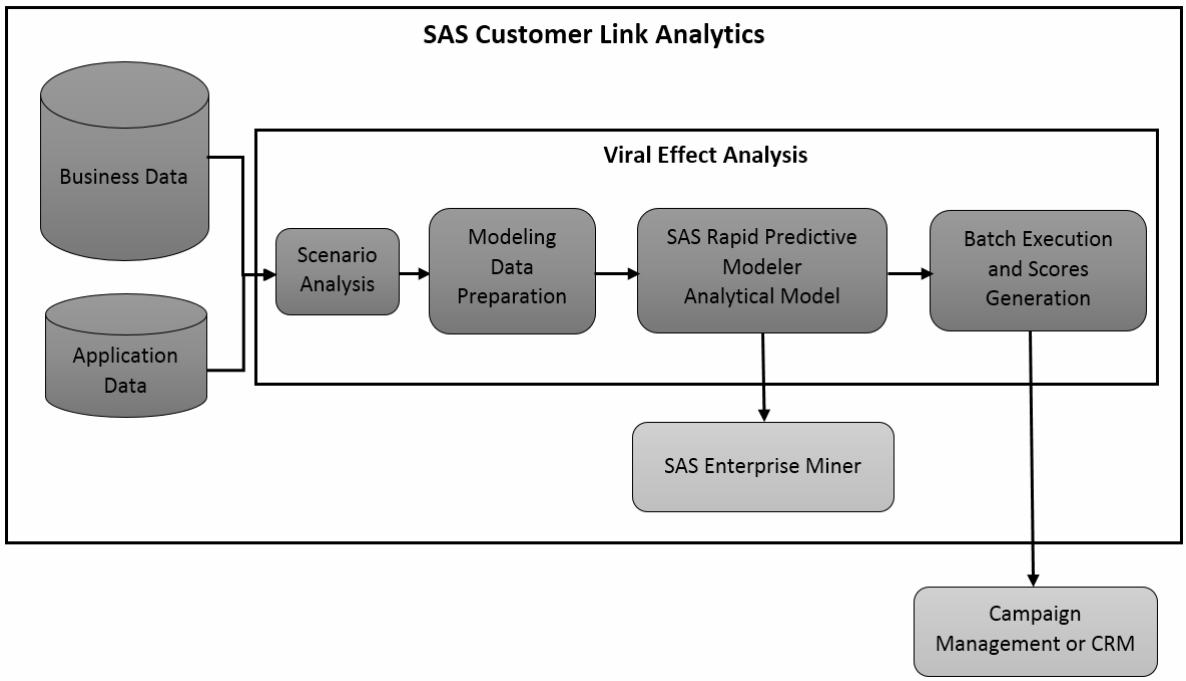
Before you start working with the analytical flow of viral effect analysis, make sure that you complete the following tasks:

- 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.

Solution Flow of Viral Effect Analysis

Figure 15.1 Viral Effect Analysis Solution Flow



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

1 Using the project information, define an appropriate scenario for that project. For each scenario, indicate the purpose of your viral effect analysis. The purpose also indicates whether the event of analysis is churn or product adoption.

2 Prepare modeling data based on the project information. This data contains information about the roles and communities of the project that is associated with the scenario. It also contains transactional data and details of the event of analysis.

   Note: SAS Customer Link Analytics does not support viral effect analysis if the transactional data is fully aggregated.

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.

Analytical Flow for Viral Effect Analysis

SAS Customer Link Analytics provides you a step-by-step analytical flow to analyze the viral effect. When you publish a scenario, the scenario completes one run in design mode and then it is pushed to batch mode. The modeling tasks are completed in design mode. Similarly, the scoring tasks are performed in batch mode.

1 **Scenario Definition**

   Define a scenario to indicate the business objective of your analysis. For example, you can define a scenario to reduce churn or to increase product adoption. For more information, see “Create a Scenario” on page 124.

2 **Parameter Update**

   Specify appropriate values for the parameters that are defined for the scenario. For more information, see “Update Parameter Values” on page 125.

3 **ABT Building**

   Define the modeling analytical base table (ABT). The modeling ABT contains all the variables that are required for the viral effect analysis. For more information, see “Build the Modeling ABT” on page 128.

4 **Model Building**

   Build an analytical model by using the SAS Rapid Predictive Modeler macro. For more information, see “Build the Analytical Model” on page 128.

5 **Model Registration**

   Register the model in the metadata to extract the metadata attributes of the model. For more information, see “Register the Model” on page 129.

6 **Model Capture**

   Capture and store the model’s information in application data. For more information, see “Capture Model Information” on page 129.

7 **Scenario Publishing**

   Publish the scenario for the scoring run (also called the *batch run*). For more information, see “Publish a Scenario” on page 130.

8 **Batch Processing**

   Complete the scoring tasks for a scenario. For more information, see *SAS Customer Link Analytics: Administrator’s Guide.*

9 **Analysis of Reporting Data**

   Analyze the reporting data that is uploaded to the SAS Customer Link Analytics LASR Analytic Server. For more information see, “Load Reporting Data for Viral Effect Analysis into the SAS Customer Link Analytics LASR Analytic Server” on page 135.

10 (Optional) **Pulling Scenario to Design Mode**

Pull the scenario from batch mode to design mode. For more information, see “Pull a Scenario into Design Mode” on page 135.

16

Working with Viral Effect Analysis

Log On to SAS Stored Process Web Application

To perform the tasks of the analytical flow, log on to the 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.

Create a Scenario

When you define a scenario, you indicate the business objective of your viral effect analysis. For example, you might want to define a scenario to analyze the impact of viral churn or viral product adoption. In addition, a scenario defines the event of analysis for which you need to build the analytical model. Each scenario is associated with a project that has completed at least one successful run either in design or batch mode.

To create a scenario:

1 In the SAS Stored Process Web Application window, select **Stored Processes** ▸ **Products** ▸ **SAS Customer Link Analytics** ▸ **Cust Link Analytics 6.5** ▸ **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 primary key
Enter the unique ID of the project for which you want to perform the viral effect analysis. To ensure that you enter the correct ID, refer to the **Project ID** that is displayed in the Properties pane of the Projects workspace. Also, make sure that the project has completed at least one successful run either in design or batch mode.

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.

Scenario type
Select the purpose for which you are defining the scenario. The scenario type indicates the event for which the node behavior is observed. The following options are available:

Churn
indicates that the nodes are observed for their churn behavior. For this scenario type, SAS Customer Link Analytics predicts the probability of a node to churn from the network based on the connections that it has with nodes that have already churned from the network.

Product Adoption
indicates that the nodes are observed for adopting a specific product. SAS Customer Link Analytics predicts the probability of a node to adopt a particular product based on the connections that it has with nodes that have already adopted this product.

4 Click **Run**. A unique sequence number is generated for the scenario. This sequence number is used as the *Scenario ID* in the subsequent analytical tasks.

In addition, SAS Customer Link Analytics creates the following folder structure in SAS metadata.

For each scenario that you create, a separate folder is created with the sequence number of the scenario as its name. Within this folder, the following subfolders are created:

Data Sources
contains data that is required for viral effect analysis. This data is stored in the following subfolders:

Analytics Data
stores the source table that is required for building the analytical model for viral effect analysis.

Report Data
stores the data that is required for generating SAS Visual Analytics reports.

Score Data
stores the output that the scoring process produces.

Models
stores the model that the SAS Rapid Predictive Modeler macro creates. SAS Customer Link Analytics registers this model in SAS metadata.

Reports
stores the SAS Visual Analytics reports that are created to analyze the viral effect analysis results.

Update Parameter Values

For each scenario that you create, you have to specify values for certain predefined parameters. These parameters are required for processing the additional tasks of the analytical workflow. For more information about the parameters that are defined for viral effect analysis, see *SAS Customer Link Analytics: Administrator's Guide*. SAS Customer Link Analytics sets up default values for some of these parameters. However, you can change the default values according to your requirements.

To update the parameter values:

1 Retrieve the sequence number of the scenario whose parameter values you want to update. For more information, see "Retrieve the Scenario ID" on page 130.

2 In the SAS Stored Process Web Application window, select **Stored Processes** ▸ **Products** ▸ **SAS Customer Link Analytics** ▸ **Cust Link Analytics 6.5** ▸ **Viral Effect Analysis**.

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

4 Specify values for the following parameters:

Scenario primary key
Enter the scenario ID.

Library name of event details table
Specify the library in which the event details table resides. Your administrator registers this table in SAS metadata. For more information, see *SAS Customer Link Analytics: Administrator's Guide*.

The event details table contains certain key columns that store the event information of all nodes. The key columns differ depending on whether the scenario type is **Churn** or **Product Adoption**. For more information, see "About the Event Details Table" on page 131.

Name of event details table
Specify the name of the table that stores the event details.

Column name of unique entity ID in event details table
Specify the column name from the event details table that uniquely identifies each node. For example, the column name can be subscription_id, customer_id, node_id, and so on.

Column name of event date in event details table
Specify the column name that stores the event date. For example, for the **Churn** scenario type, this column stores the churn date. Similarly, for the **Product Adoption** scenario type, this column can store the product activation date or the product usage start date.

Note: Make sure that the length of the column name that you specify does not exceed 22 characters and that it is a valid SAS column name.

Column name of product ID values
This parameter applies to the **Product Adoption** scenario type. Specify the column name of the event details table that stores the unique IDs of the products that the nodes have adopted.

Product ID value
This parameter applies to the **Product Adoption** scenario type. Enter the value of the product ID for which you want to perform viral impact analysis. Make sure that the value that you enter is exactly the same as the value that is available in the event details table.

Note: You can specify only one product for each scenario. If you want to perform viral effect analysis for another product, then you have to define another scenario.

SAS Rapid Predictive Modeler model type
Select the type of analytical model that you want to build by using SAS Rapid Predictive Modeler. Make sure that you choose any one of the following options:

Basic
enables you to perform a simple regression analysis. This methodology involves three data mining operations: variable selection, transformation, and modeling.

Intermediate
enables you to perform a more sophisticated analysis in addition to the basic analysis. The intermediate methodology involves the following additional data mining operations: imputation, union of variable selection techniques, and champion model selection. This methodology compares the basic model with the intermediate model and chooses the best one.

Advanced
enables you to perform a more sophisticated analysis in addition to the basic and intermediate analyses. The advanced method performs all the data mining operations that the basic and intermediate methodologies perform. In addition, it compares the advanced model to the intermediate and basic champion models and chooses the best performing champion model as the result.

Note: If you select the blank value, the stored process will not run successfully.

For more information, see "Build the Analytical Model" on page 128.

SAS Enterprise Miner project location
Specify the path where you want to create the workspace for the SAS Enterprise Miner project. For more information, see "Build the Analytical Model" on page 128.

Note:

- The only special character that you can use for the project location 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 not exceed 100 bytes.

Note: Make sure that you have the Write permission to the folder location that you specify for this parameter.

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 is created with the name that you provide here. For more information, see "Build the Modeling ABT" on page 128.

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, Hadoop, or Teradata.

5 Click **Run**. The default values of the parameters are updated with the values that you enter here.

Build the Modeling ABT

A modeling ABT is the input table for the analytical model. Before you build the modeling ABT, make sure that your administrator has deployed the siascenloopjob job. For more information, see *SAS Customer Link Analytics: Administrator's Guide*.

To build the modeling ABT:

1. Retrieve the sequence number of the scenario for which you want to build the modeling ABT. For more information, see "Retrieve the Scenario ID" on page 130.
2. In the SAS Stored Process Web Application window, select **Stored Processes** ▸ **Products** ▸ **SAS Customer Link Analytics** ▸ **Cust Link Analytics 6.5** ▸ **Viral Effect Analysis**.
3. Select the **sia_stp_build_abt** stored process. The parameters that are defined for this stored process are displayed.
4. In the right pane, enter the scenario ID that you retrieved in step 1.
5. Click **Run**. The modeling ABT is built with the name that you specified for the **Modeling ABT name** parameter. For more information, see "Update Parameter Values" on page 125.

   The variables that are populated in the modeling ABT are called *ABT variables*. Each ABT variable is associated with an ABT variable category. In the application data, these variables are populated in the SCENARIO_STEP_OUTPUT_COLUMN table. For more information about this table, see *SAS Customer Link Analytics: Data Reference Guide*.

   Moreover, the modeling ABT is registered in the following SAS metadata location: **`/Shared Data/SAS Customer Link Analytics/Cust Link Analytics 6.5/Viral Effect Analysis/<Scenario ID>/Data Sources/Analytics Data`**.

   For more information about the ABT variable categories, see "About the ABT Variable Categories" on page 132.

Build the Analytical Model

SAS Customer Link Analytics builds a predictive model using the modeling ABT and the SAS Rapid Predictive Modeler macro. The model is built depending on the value that you selected for the **SAS Rapid Predictive Modeler model type** parameter. For more information, see "Update Parameter Values" on page 125.

To build the analytical model:

1. Retrieve the sequence number of the scenario for which you want to build the model. For more information, see "Retrieve the Scenario ID" on page 130.

2 In the SAS Stored Process Web Application window, select **Stored Processes** ▸ **Products** ▸ **SAS Customer Link Analytics** ▸ **Cust Link Analytics 6.5** ▸ **Viral Effect Analysis**.

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

4 Enter the scenario ID that you retrieved in step 1.

5 Click **Run**. A SAS Enterprise Miner project workspace is created in the location that you specified for the **SAS Enterprise Miner project location** parameter. For more information, see "Update Parameter Values" on page 125.

Register the Model

When you register the model, it is registered in SAS metadata and uniquely identified with a sequence number.

To register the model:

1 Retrieve the sequence number of the scenario whose model you want to register. For more information, see "Retrieve the Scenario ID" on page 130.

2 In the SAS Stored Process Web Application window, select **Stored Processes** ▸ **Products** ▸ **SAS Customer Link Analytics** ▸ **Cust Link Analytics 6.5** ▸ **Viral Effect Analysis**.

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

4 Enter the scenario ID that you retrieved in step 1.

5 Click **Run**. In the metadata tree, the model is registered at the following location: `/Shared Data/SAS Customer Link Analytics/Cust Link Analytics 6.5/Viral Effect Analysis/<Scenario ID>/Models`. Also, a unique sequence number is generated for the model. This sequence number is referred to as *Model ID* in the subsequent tasks of the analytical flow.

Capture Model Information

In this task, the information about the model that you have registered is extracted from the metadata, and a metadata ID is generated. This ID is stored in the application data. After you capture the model information, the scenario completes one run in design mode.

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

To capture the model information:

1 Retrieve the sequence number of the scenario whose model information you want to capture. For more information, see "Retrieve the Scenario ID" on page 130.

2 In the SAS Stored Process Web Application window, select **Stored Processes** ▸ **Products** ▸ **SAS Customer Link Analytics** ▸ **Cust Link Analytics 6.5** ▸ **Viral Effect Analysis**.

3 In the left pane, select the **sia_stp_capture_store_model_info** stored process.

4 Enter the scenario ID that you retrieved in step 1.

5 Click **Run**. The application data is updated with the metadata ID of the model. In addition, the score code of the model is created with the filename sia_model_score_code_<*Scenario ID*>_<*Model ID*>. This file is stored in the following location: `<SAS configation path>/Lev1/AppData/SASCustomerLinkAnalytics/6.5/scenario/modelscorecode`.

   The score code identifies the significant variables and the corresponding category of each. This information is stored in the application data. In the scoring run, the variables that belong to these categories are generated and populated in the final scoring ABT. For more information, see *SAS Customer Link Analytics: Administrator's Guide*.

Publish a Scenario

When you publish a scenario, it is pushed in batch mode. The scenario is now ready for the scoring run. For more information about the processes involved in the scoring run, see *SAS Customer Link Analytics: Administrator's Guide*.

To publish a scenario:

1 Retrieve the sequence number of the scenario that you want to publish. For more information, see "Retrieve the Scenario ID" on page 130.

2 In the SAS Stored Process Web Application window, select **Stored Processes** ▸ **Products** ▸ **SAS Customer Link Analytics** ▸ **Cust Link Analytics 6.5** ▸ **Viral Effect Analysis**.

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

4 Enter the scenario ID that you retrieved in step 1.

5 Click **Run**. A batch code file is generated at the following location: `<SAS configuration directory>/Lev1/AppData/SASCustomerLinkAnalytics/6.5/scenario/batchcode`. Your administrator can schedule this file for scoring. For more information, see *SAS Customer Link Analytics: Administrator's Guide*.

Retrieve the Scenario ID

To perform the tasks of the analytical flow of viral effect analysis, specify the unique sequence number that is generated for the scenario.

To retrieve the sequence number of a scenario:

1 In the SAS Stored Process Web Application, select **Stored Processes** ▸ **Products** ▸ **SAS Customer Link Analytics** ▸ **Cust Link Analytics 6.5** ▸ **Viral Effect Analysis**.

2 Click **sia_stp_get_project_scenario_dtl**.

3 In the right pane, enter the following details:

Entity
Select **Scenario** from the list to retrieve the sequence number of a scenario. Make sure that you do not select the blank value from the list. Otherwise, the stored process will not run successfully.

Name
Enter the name of the scenario for which you want to perform viral effect analysis.

4 Click **OK**. The sequence number that is assigned to the scenario is displayed, along with other details. Use this sequence number as the scenario primary key or scenario ID when you perform the analytical tasks.

About the Event Details Table

Overview of the Event Details Table

The event details table contains key columns depending on the scenario type under consideration. These key columns store the event information of the nodes. For the **Churn** scenario type, these key columns indicate whether a churn event has occurred for a node. Similarly, for the **Product Adoption** scenario type, these key columns indicate whether a node has bought a particular product. Your administrator registers this table in SAS metadata. For more information about the structure of the event details table, see *SAS Customer Link Analytics: Administrator's Guide*.

Event Details Table for Viral Churn

For the **Churn** scenario type, the event details table contains two types of key columns: the unique ID for a node and the date on which the node churns out from the network. For example, these columns can be node_id and churn_date

Table 16.1 *Sample Data in Event Details Table for Churn*

node_id	churn_date
A	15Jan2015:13:16:47
B	10Dec2014:05:47:58
C	23Dec2014:09:57:35

node_id	churn_date
D	25Feb2015:11:50:31

Event Details Table for Viral Product Adoption

For the **Product Adoption** scenario type, the event details table stores three types of key columns: the unique ID for a node, the product adoption date, and the unique ID for the product. The product adoption can be the date on which the node starts using the product or the date on which the product is activated for the node. For example, the key columns in the event details table can be node_id, product_id, and product_start_dt.

Table 16.2 *Sample Records of Event Details Table for Product Adoption*

node_id	product_id	product_start_id
A	Product A	01Jan2015:07:41:24
A	Product C	20Mar2015:02:53:44
B	Product B	01Dec2014:11:50:32
C	Product C	20Mar2015:02:05:13

If you want to analyze the viral product adoption for Product A and Product C, then you have to define two separate scenarios and specify appropriate parameters for each of them.

About the ABT Variable Categories

Input Variables

Enrichment Variables

You have built the modeling ABT for a scenario that is associated with a SAS Customer Link Analytics project. In this case, some of the ABT variables are the same as the variables that are generated when you run the data enrichment process for certain enrichment categories. For more information, see “Enrich the Output Data of a Project” on page 79.

SAS Rapid Predictive Modeler uses these ABT variables as input variables to build the analytical model. These input variables are the variables that are generated for the following node-level data enrichment categories:

- **Roles and communities**
- **Associations with neighboring roles**
- **Community-level statistics**

- **Aggregated transactional data**
- **Node attributes**

Assume that you have not run the data enrichment process for one or more of these data enrichment categories. In this case, SAS Customer Link Analytics first runs the data enrichment process when you built the ABT, and then it generates the ABT variables.

For more information about the variables that are generated for the data enrichment categories, see *SAS Customer Link Analytics: Data Reference Guide*.

Count-Based Variables

The modeling ABT contains variables that give various counts for a node. These count variables are generated based on the scenario type and the number of roles that are defined for the project. The *count variables* indicate the number of connections that a node has with various roles. The counts are computed based on the occurrence of the event under consideration (churn or product adoption) in the observation period.

For example, consider the churn scenario type. Assume that two roles, followers and leaders, are defined for the project. In this case, the ABT contains four count-based variables. These variables give the following information:

- the node under consideration was connected to how many follower nodes that churned in the observation period
- the node under consideration is connected to how many follower nodes that did not churn in the observation period
- the node under consideration was connected to how many leader nodes that churned in the observation period
- the node under consideration is connected to how many leader nodes that did not churn in the observation period

For more information about the count-based variables, see *SAS Customer Link Analytics: Data Reference Guide*.

Variables Based on Roles and Transactional Measures

The modeling ABT variables contain variables that give information about the aggregated transactional summary of a node. These variables are generated based on the scenario type and transactional measures that are configured for a project. In addition, these variables consider the roles that a node is connected to. The values of these variables are computed based on the occurrence of the event under consideration (churn or product adoption) in the observation period.

For example, consider the product adoption scenario type. In this case, the modeling ABT contains the following types of variables for each transactional measure:

- Total incoming value of a transactional measure for a node based on its connections with nodes that adopted the product in the observation period
- Total outgoing value of a transactional measure for a node based on its connections with nodes that adopted the product in the observation period

- Total incoming value of a transactional measure for a node based on its connections with nodes that did not adopt the product in the observation period
- Total outgoing value of a transactional measure for a node based on its connections with nodes that did not adopt the product in the observation period

Considering the role that the node is connected to and the transactional measure, the modeling ABT contains additional variables. These variables are generated for three types of aggregations: minimum, maximum, and total. Here is list of variables that are generated for the total aggregation type. Similar variables are generated for the minimum and maximum aggregation type.

- Total incoming value of a transactional measure for a node based on its connections with nodes of a specific role that adopted the product in the observation period
- Total outgoing value of a transactional measure for a node based on its connections with nodes of a specific role that adopted the product in the observation period
- Total incoming value of a transactional measure for a node based on its connections with nodes of a specific role that did not adopt the product in the observation period
- Total outgoing value of a transactional measure for a node based on its connections with nodes of a specific role that did not adopt the product in the observation period

For more information about the variables based on roles and transactional measures, see *SAS Customer Link Analytics: Data Reference Guide*.

Target Variables

The modeling ABT contains one target variable depending on the scenario type for which you are building the modeling ABT. The target variables indicate whether the event under consideration has occurred for the nodes in the performance window. For more information, see “Historical, Performance, and Observation Periods” on page 151.

For the churn scenario type, the target variable is churn_flg. This variable indicates whether the node has churned from the network in the performance period.

For the product adoption scenario type, the target variable is prod_adopt_flg. This variable indicates whether the node adopted the specified product in the performance period. In addition to the target variable, the ABT contains the prod_id variable that indicates the unique ID of the product for which you have built the modeling ABT.

For more information about these variables, see *SAS Customer Link Analytics: Data Reference Guide*.

Load Reporting Data for Viral Effect Analysis into the SAS Customer Link Analytics LASR Analytic Server

If you want to explore the reporting data for viral effect analysis, you can load it into the SAS Customer Link Analytics LASR Analytic Server. SAS Customer Link Analytics generates the reporting data during the scoring run of a scenario. For more information about the scoring run of a scenario, see *SAS Customer Link Analytics: Administrator's Guide*.

The reporting data helps you compare the variable values of the two successive scoring runs. You can use this information to define various campaign strategies. For example, the comparative churn scores can help you identify if there is a significant increase in the scores. You can target these nodes and define appropriate campaign strategies to prevent them from churning from the network. Similarly, if the product adoption scores indicate a significant increase, you can define campaign strategies to target these nodes for product adoption.

To load the reporting data into the SAS Customer Link Analytics LASR Analytic Server:

1. Retrieve the sequence number of the scenario that you want to publish. For more information, see "Retrieve the Scenario ID" on page 130.
2. In the SAS Stored Process Web Application window, select **Stored Processes** ▸ **Products** ▸ **SAS Customer Link Analytics** ▸ **Cust Link Analytics 6.5** ▸ **Viral Effect Analysis**.
3. In the left pane, select the **sia_stp_load_rptdata_lasr** stored process.
4. Enter the scenario ID that you retrieved in step 1.
5. Click **Run**. SAS Customer Link Analytics completes the following tasks:
   - Stores the reporting data in the scen_rpt_LASR_<*Scenario ID*> LASR table.
   - Registers the LASR table in the following metadata location: **`/Shared Data/SAS Customer Link Analytics/Cust Link Analytics 6.5/Viral Effect Analysis/<Scenario ID>/Data Sources/Report Data`**.
   - Loads the LASR table into the SAS Customer Link Analytics LASR Analytic Server.

   For more information, see "Sample Reporting Data for Viral Effect Analysis".

Pull a Scenario into Design Mode

If the scoring run does not produce results according to your business requirements, you might want to rebuild the analytical model. Also, if the configuration of the project that is associated with the scenario changes, you

might want to rework the tasks of the modeling run. In all these scenarios, you have to pull the scenario back into design mode.

To pull the scenario into design mode:

1. Retrieve the sequence number of the scenario for which you want to build the model. For more information, see “Retrieve the Scenario ID” on page 130.
2. In the SAS Stored Process Web Application window, select **Stored Processes** ▸ **Products** ▸ **SAS Customer Link Analytics** ▸ **Cust Link Analytics 6.5** ▸ **Viral Effect Analysis**.
3. In the left pane, select the **sia_stp_push_scen_to_design** stored process.
4. Enter the scenario ID that you retrieved in step 1.
5. Click **Run**. The batch code for the scenario is deleted and the scenario is returned to design mode. In addition, all the ABT variable categories and the ABT variables are marked as significant. For more information, see “About the ABT Variable Categories” on page 132.

   You can now start working on the tasks of the modeling run. For more information, see “Analytical Flow for Viral Effect Analysis” on page 121.

Part 6

Appendixes

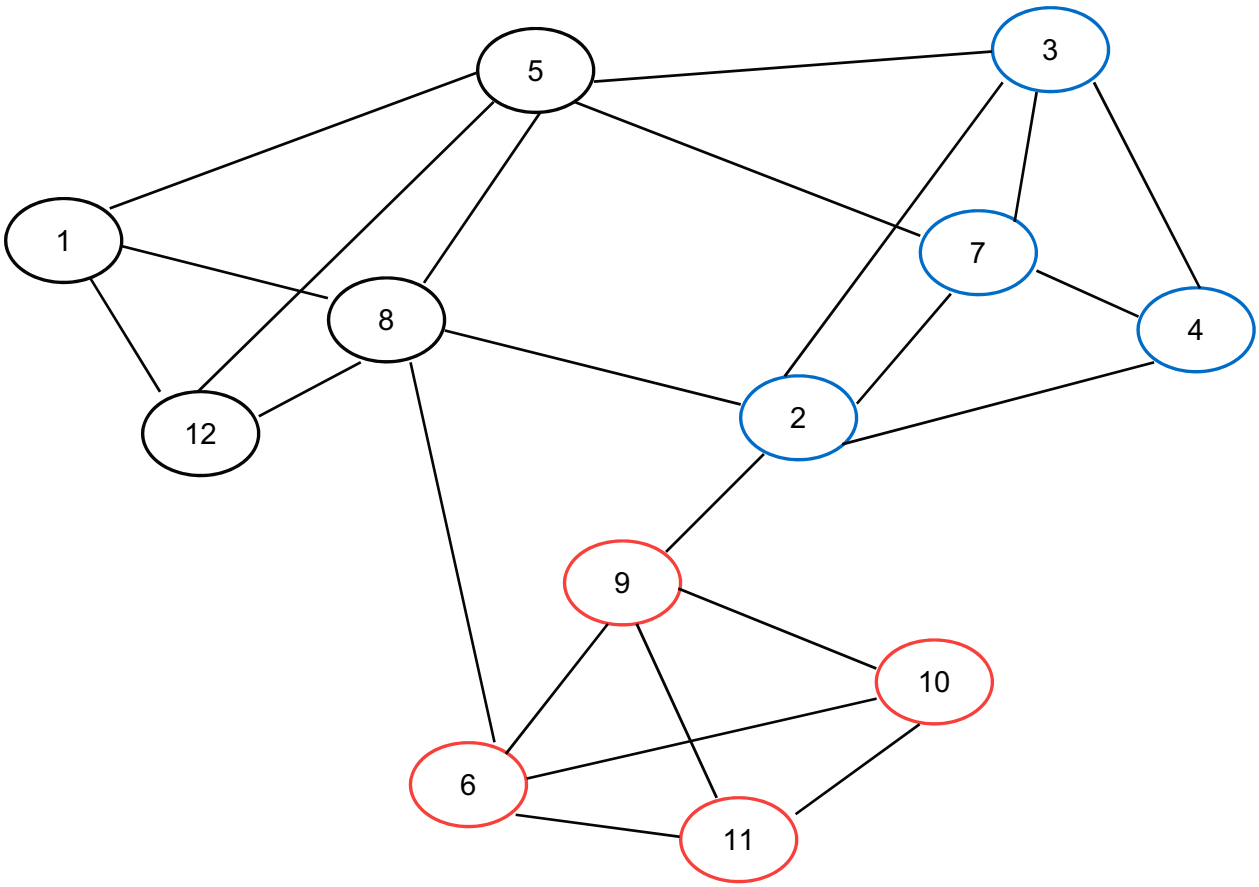
Appendix 1

Common Network Analysis Concepts

About Network Analysis

Network analysis examines the structure of relationships between entities. These entities can be persons, groups, organizations, websites, 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*

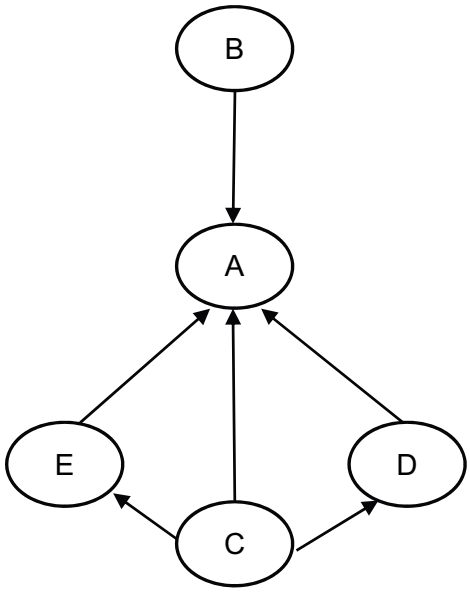


In a network diagram, nodes represent the entities that are connected and links represent the connections between nodes. Each link connects two nodes. However, a node can have multiple connecting links. The nodes and the links together constitute a community. One or more such communities forms a network. 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

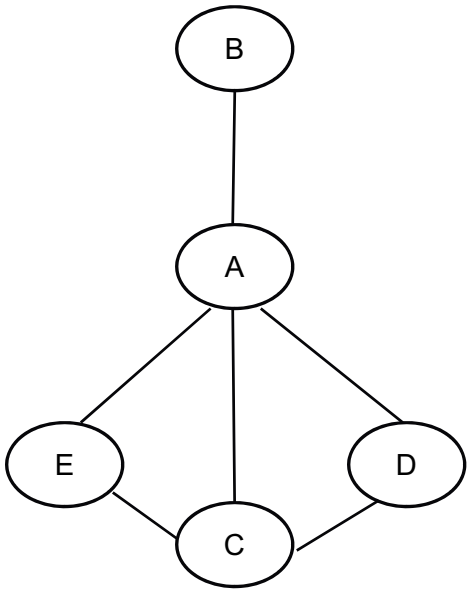
A network diagram in which each link has a sense of direction from one node to another is called a *directed graph*. Each link indicates whether it is an incoming or outgoing link for a node. However, in an undirected graph, a link has no sense of direction.

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

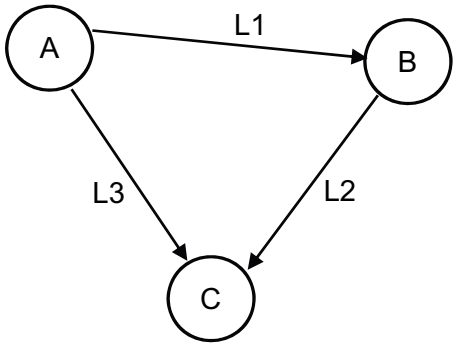


For example, in the network that is represented in Figure A1.3, assume that the nodes represent people at an alumni party. There is a link between two people if they shake hands. In this case, a link between person A and person B indicates that if person A shook hands with person B, then person B also 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

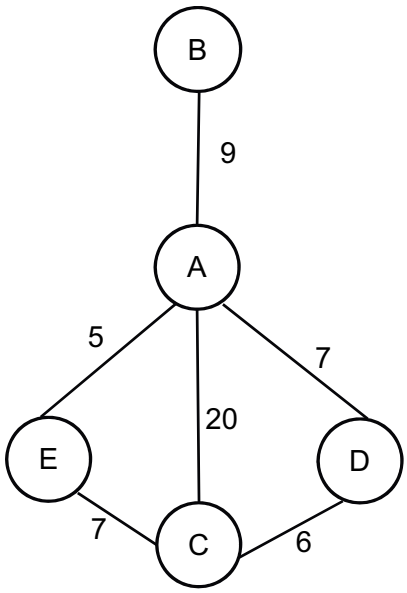
Node Name	In-Degree Measure	Out-Degree Measure
A	0	2

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*

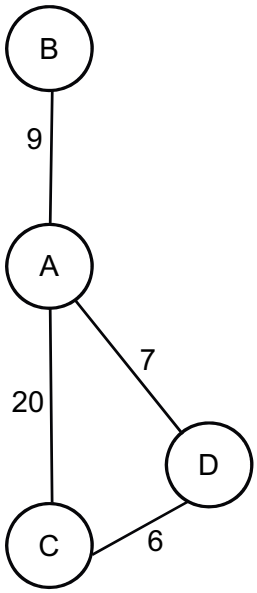


For example, in this figure, consider the shortest path between nodes A and C. Considering the link weights, the shortest path between nodes A and C is A->E->C, with the cumulative sum of weights as 12. Considering the number of links, the shortest path is A->C because the number of links that separate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

About 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 as follows:

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

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

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:

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

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:

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

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.

Convergence Criterion

The clustering procedure runs in iterations. After each iteration of the clustering procedure, the cluster centroids are updated. Iterations stop when the relative change in the cluster centroids is less than or equal to the convergence criterion. For complete convergence, it is recommended that you enter the minimum value for this parameter. However, complete convergence also depends on the number of iterations of the clustering procedure. To achieve complete convergence, set the number of iterations of the clustering procedure to a large value.

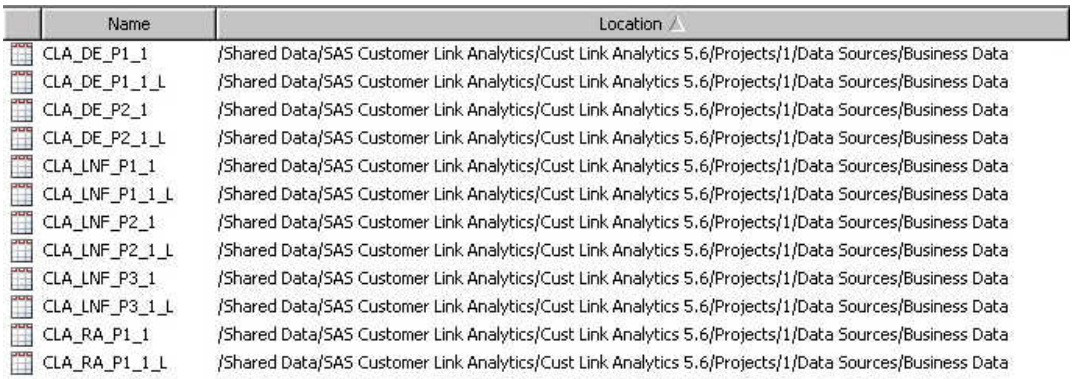
Appendix 3

Workflow Step Data

The information that SAS Customer Link Analytics produces when you run a workflow step is stored in the project-specific business data tables. This information is called the *workflow step data*. For more information about the business data tables, see *SAS Customer Link Analytics: Data Reference Guide*.

When you run a workflow step for the first time, SAS Customer Link Analytics stores the workflow step data of the current run in the business data tables. When you run the workflow step subsequently, the workflow step data of the previous run is retained, and the data of the current run is added. As a result, comparative data of two successive runs of a workflow step is available in the business data tables. Moreover, SAS Customer Link Analytics registers these tables in SAS metadata.

Figure A3.1 *Workflow Step Data*



Name	Location
CLA_DE_P1_1	/Shared Data/SAS Customer Link Analytics/Cust Link Analytics 5.6/Projects/1/Data Sources/Business Data
CLA_DE_P1_1_L	/Shared Data/SAS Customer Link Analytics/Cust Link Analytics 5.6/Projects/1/Data Sources/Business Data
CLA_DE_P2_1	/Shared Data/SAS Customer Link Analytics/Cust Link Analytics 5.6/Projects/1/Data Sources/Business Data
CLA_DE_P2_1_L	/Shared Data/SAS Customer Link Analytics/Cust Link Analytics 5.6/Projects/1/Data Sources/Business Data
CLA_LNF_P1_1	/Shared Data/SAS Customer Link Analytics/Cust Link Analytics 5.6/Projects/1/Data Sources/Business Data
CLA_LNF_P1_1_L	/Shared Data/SAS Customer Link Analytics/Cust Link Analytics 5.6/Projects/1/Data Sources/Business Data
CLA_LNF_P2_1	/Shared Data/SAS Customer Link Analytics/Cust Link Analytics 5.6/Projects/1/Data Sources/Business Data
CLA_LNF_P2_1_L	/Shared Data/SAS Customer Link Analytics/Cust Link Analytics 5.6/Projects/1/Data Sources/Business Data
CLA_LNF_P3_1	/Shared Data/SAS Customer Link Analytics/Cust Link Analytics 5.6/Projects/1/Data Sources/Business Data
CLA_LNF_P3_1_L	/Shared Data/SAS Customer Link Analytics/Cust Link Analytics 5.6/Projects/1/Data Sources/Business Data
CLA_RA_P1_1	/Shared Data/SAS Customer Link Analytics/Cust Link Analytics 5.6/Projects/1/Data Sources/Business Data
CLA_RA_P1_1_L	/Shared Data/SAS Customer Link Analytics/Cust Link Analytics 5.6/Projects/1/Data Sources/Business Data

As indicated in this diagram, a pair of business tables is created. The table that contains the suffix _L stores the workflow step data of the previous run. For example, for the Data Extraction workflow step, the following tables store the workflow step data for the two successive runs:

- cla_de_p1_<*Project ID*>_L (previous run)
- cla_de_p1_<*Project ID*> (current run)

The workflow step data of two successive runs of a project provides comparative information over time. SAS Customer Link Analytics enables you to enrich this data and use it to perform reporting and analytical tasks. For more information, see “Enriching Project Output Data ” on page 71.

Appendix 4

Performance and Observation Periods

Modeling

The modeling phase considers three types of time periods:

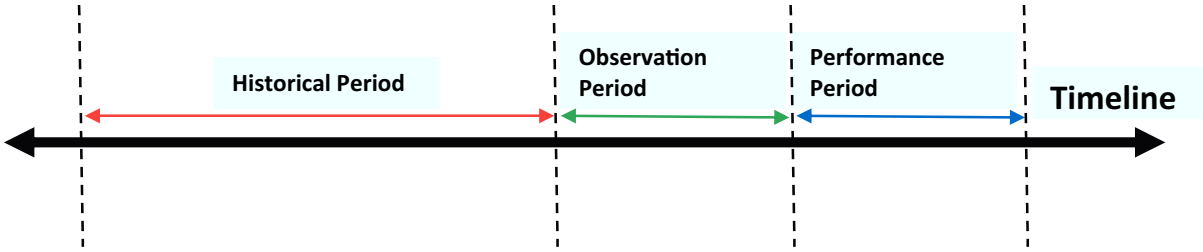
Historical period
: In the modeling phase, ABT variables are built by considering the entire historical period. For a scenario that is linked to aSAS Customer Link Analytics project, you select the historical period when you configure the Data Extraction workflow step.

Observation period
: In the observation period, the event under consideration is observed and input variables are computed based on these observations. That is, the node behavior is observed depending on the scenario type. For example, for the **Churn** scenario type, nodes are observed for their churn behavior. The event of whether any of the nodes churn from the network in the specified period is observed. Similarly, for the **Product Adoption** scenario type, the event of whether any of the nodes adopt a particular product in the specified period is observed.

Performance period
: In the performance period, the target variable is computed as an impact of the occurrence of the event under consideration in the observation window. That is, the node behavior is observed considering the occurrence of the event in the observation window. For example, for the **Churn** scenario type, the nodes that churn in the performance window because of their connections with nodes that churn in the observation window are observed. Similarly, for **Product Adoption**, the nodes that adopt a product in the performance window because of their connections with nodes that adopt the same product in the observation window are observed.

  Note: SAS Customer Link Analytics computes the duration for the observation period and the performance period for a scenario while computing the target variable.

Figure A4.1 *Sample Time Windows: Modeling*

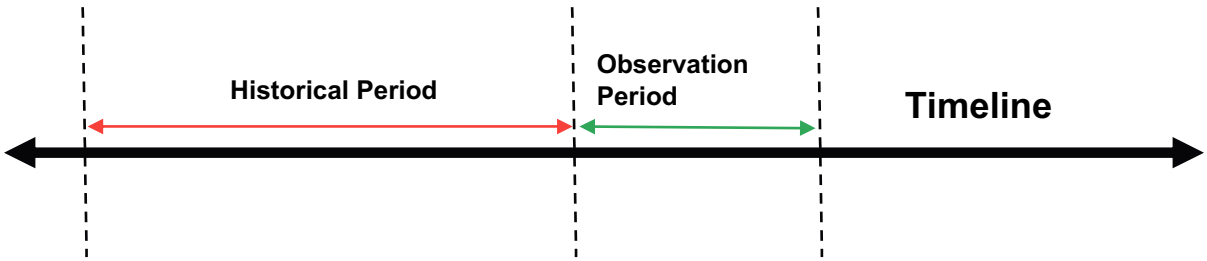


Note: The observation period can overlap the historical period. However, the performance period can start only after the observation period ends.

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 A4.2 *Time Windows: Scoring*



Appendix 5

Node-Level Data Enrichments

The following table indicates the number of variables that the enriched data contains for each node-level enrichment category when you run the data enrichment process.

The information in the table is based on the following assumptions that are considered for a project:

- The transactional table that is associated with the project contains five columns that are configured as transactional measures.
- The Role Assignment workflow step has assigned any one of the three roles (Leader, Follower, or Boundary Spanner) to a node.
- Two types of directions, incoming and outgoing, are considered for the interaction between two nodes.

The number of roles and transactional measures that you need to include in the formula is project-specific. Therefore, to compute the number of variables for your project, replace the number of roles with the number of roles that the Role Assignment workflow step has produced for your project. Similarly, replace the number of transactional measures with the number of columns that are configured as measures in the transactional table selected in the source data profile of your project.

Table A5.1 *Formula for Number of Variables Created*

Category Name	Number of TransactionalMeasures	Number of Possible Aggregations	Number of Direction Types	Number of Roles	Number of Derived Measures	Formula	Total Number of Variables
Roles and communities	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Same as the number of rows in the CLA_RA_P1_<*Project ID*> table.
Node attributes	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Total number of columns, except for the Process date columns that are available in the Node Attribute tables configured in the source data profile of the project.
Associations with neighboring roles	5	3 (Minimum, Maximum, and Total)	2	3	Not applicable	Number of Transactional Measures * Number of Possible Aggregations * Number of Direction Types * Number of Roles	5*3*2*3 = 90
	Not applicable	1 (Total number of connected nodes with a specific role)	Not applicable	3	Not applicable	Number of Possible Aggregations * Number of Roles	1*3 = 3
	Total variables						90+3=93

Category Name	Number of TransactionalMeasures	Number of Possible Aggregations	Number of Direction Types	Number of Roles	Number of Derived Measures	Formula	Total Number of Variables
Community-level statistics	5	1 (Total)	2	3 (Leader, Follower, and Boundary Spanner)	Not applicable	Number of Transactional Measures * Number of Possible Aggregations * Number of Direction Types * Number of Roles	5*1*2*3 = 30
	5	1 (Average)	2	3 (Leader, Follower, and Boundary Spanner)	Not applicable	Number of Transactional Measures * Number of Possible Aggregations * Number of Direction Types * Number of Roles	5*1*2*3 = 30
	5	1(Average)	2	Not applicable	Not applicable	Number of Transactional Measures * Number of Possible Aggregations * Number of Direction Types	5*1*2 = 10
	5	3 (Minimum, Maximum, and Total)	2	Not applicable	Not applicable	Number of Transactional Measures * Number of Possible Aggregations * Number of Direction Types	5*3*2 = 30
	Total variables						30+30+10+30 = 100

Category Name	Number of TransactionalMeasures	Number of Possible Aggregations	Number of Direction Types	Number of Roles	Number of Derived Measures	Formula	Total Number of Variables
Aggregated transactional data	5	3 (Minimum, Maximum, and Total)	2	Not applicable	Not applicable	Number of Transactional Measures * Number of Possible Aggregations * Number of Direction Types	5*3*2 = 30
Roles and communities over time	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Same as the number of rows in the CLA_RA_P1_<*Project ID*> table.
Churn and acquisition indicators	Not applicable	Not applicable	Not applicable	Not applicable	2 (Churn indicator and Acquisition indicator)	Number of Derived Measures	2
Relation with churned and acquired nodes	5	1 (Percentage)	2	Not applicable	Not applicable	Number of Transactional Measures * Number of Possible Aggregations * Number of Direction Types	5*1*2=10
	Not applicable	1 (Percentage)	2	Not applicable	1 (Link weight)	Number of Possible Aggregations * Number of Direction Types * Number of Derived Measures	1*2*1=2
	Not applicable	1 (Total)	2	Not applicable	1 (Link weight)	Number of Possible Aggregations * Number of Direction Types * Number of Derived Measures	1*2*1=2

Category Name	Number of TransactionalMeasures	Number of Possible Aggregations	Number of Direction Types	Number of Roles	Number of Derived Measures	Formula	Total Number of Variables
	Not applicable	1 (Total)	Not applicable	3	1 (Connected churned count)	Number of Possible Aggregations * Number of Roles * Number of Derived Measures	1*3*1=3
	Not applicable	1 (Total)	Not applicable	3	1 (Connected acquired count)	Number of Possible Aggregations * Number of Roles * Number of Derived Measures	1*3*1=3
	Not applicable	1 (Total)	Not applicable	Not applicable	1 (Total connected churned nodes)	Number of Possible Aggregations * Number of Derived Measures	1*1=1
	Not applicable	1 (Total)	Not applicable	Not applicable	1 (Total connected acquired nodes)	Number of Possible Aggregations * Number of Derived Measures	1*1=1
	Not applicable	Not applicable	Not applicable	Not applicable	1 (Role change indicator)	Number of Derived Measures	1

Appendix 6

More About Enriched Data

Using Node-Level Enriched Data

Overview

The node-level business table that SAS Customer Link Analytics populates when you run the data enrichment process contains columns depending on the node-level enrichment categories that you choose. The following topics explain the columns that are populated as enriched data for each node-level enrichment category. These topics also explain how you can leverage the enriched data and use it further to enhance your campaign definitions. For more information about the data enrichment columns, see *SAS Customer Link Analytics: Data Reference Guide*.

Roles and Communities

When you enrich the project's output data using the default node-level enrichment category, **Roles and communities**, the enriched data contains the following columns.

Note: The cs_segment_pk column is applicable only if you have opted for community segmentation when you run the Centrality Measures Computation workflow step.

Table A6.1 *Default Enriched Data*

node_id	community_id	cs_segment_pk	entity_role_nm	between_val	closein_val	closeout_val	close_val	degreein_val	degreeout_val	degree_val	influence1_val	influence2_val
A	11	3	Follower	0.16667	0.30211	0.23804	0.32242	2	1	1	0.43846	0.12537
B	9	5	Follower	0	0.16898	0.07796	0.24989	2	1	1	0.28846	0.31343
C	11	3	Leader	0.33333	0.40438	0.27155	0.49091	4	1	3	0.85385	0.25075
D	3	8	Follower	0.01786	0.13489	0.025	0.09465	1	1	2	0.03205	0

The **Roles and communities** enrichment category provides statistics such as community ID, segment ID, role, and value of centrality measures. The data that the other node-level enrichment categories produce is appended as additional columns in the node-level table.

For example, in Table A6.1 on page 160, nodes A and C belong to the same community. However, they have different roles, follower and leader respectively. You can use the enriched data to define campaigns that are based on roles, community sizes, and centrality measure values.

Node Attributes

Assume that your administrator has configured three columns (customer type, DOB, and city) in the node attribute table that is associated with the project. When you enrich the project's output data by selecting the **Node attributes** enrichment category, these three columns are added to the default enriched data, Table A6.1 on page 160.

Table A6.2 *Data Enriched by Node Attributes Enrichment Category*

Few Columns of Default Enriched Data			Columns of Data Enriched by Node Attributes Category		
node_id	**community_id**	**entity_role_nm**	**cust_type_cd**	**dob**	**city**
A	11	Follower	Individual	16-May-1980	Florida
B	9	Follower	Corporate	25-Jul-1975	Chicago
C	11	Leader	Individual	5-May-1983	New York
D	3	Follower	Individual	25-Dec-1991	California

You can use the enriched data that contains different node attributes to design targeted and automatic campaigns. For example, you can use the data displayed in Table A6.2 on page 162 to design a campaign that is based on the date of birth (DOB). Both nodes A and C can be offered discounts in the month of May for up-selling products. Similarly, based on the customer type, you can define different campaigns for Corporate type nodes to increase the daily usage of services.

Roles and Communities over Time

Table A6.3 *Data Enriched by the Roles and Communities over Time Enrichment Category*

Columns of Default Enriched Data													Columns of Data Enriched by Roles and Communities over Time Enrichment Category									
node_id	between_val	closein_val	closeout_val	close_val	degreein_val	degreeout_val	degree_val	influence1_val	influenece2_val	community_id	cs_segment_pk	entity_role_nm	prev_closein_val	prev_closeout_val	prev_close_val	prev_degreein_val	prev_degreeout_val	prev_degree_val	prev_influence1_val	prev_influence2_val	prev_community_id	prev_entity_role_nm
A	0	0.73	0	0.44654	1	0	1	0	0	7	3	Leader	0.48	0	0.33865	1	0	1	0	0	7	Leader
B	0.25	0.47881	0.33377	0.45489	2	1	3	0.32308	0.66269	10	7	Follower	0.25187	0.18142	0.28887	2	1	3	0.16667	0.54348	8	Leader
C	0.05357	0.10315	0.18173	0.15131	1	3	4	0.26282	0	1	10	Follower	0.13565	0.12078	0.16972	1	3	4	0.24603	0	1	Leader

Note: The cs_segment_pk column is available only if you have opted for community segmentation when you run the Centrality Measures Computation workflow step.

Behavior and positioning of each node changes over time. In each run of a project, the nodes might shift from one community to the other and their centrality scores can also change. The **Roles and communities over time** enrichment category enables users to compare all the values between two consecutive runs and analyze the shift in the behavior of each node. For example, in Table A6.3 on page 164, node B has shifted from community 8 to community 10 and its role has also changed from Follower to Leader. Similarly, for node C, the role has changed from Follower to Leader. However, node A continues to be the Leader of the community.

Churn and Acquisition Indicators

Churn and acquisition indicators can be used to design campaigns for communities that have a high acquired node count and for communities for which the churn indicator is 1. These communities are more likely to disintegrate due to viral churn.

Table A6.4 *Data Enriched by Churn and Acquisition Indicators Enrichment Category*

Few Columns of Default Enriched Data			**Columns of Data Enriched by Churn and Acquisition Indicators Data Category**	
node_id	**community_id**	**entiry_role_nm**	**acq_ind**	**churn_ind**
A	9	Follower	1	0
B	11	Follower	0	0
C	1	Follower	0	1

In the table here, node A is the newly acquired node and belongs to community ID 9. Conversely, node C has churned out of community 1. Therefore, community ID 1 must be targeted for campaigns that are designed for retention.

Relation with Churned and Acquired Nodes

This data enrichment category helps derive relations of a node with a churned or acquired node. Marketing teams can identify the number of newly acquired nodes that are connected to a node and the roles that they have after the project is run. Similarly, statistics are available for churned nodes. The enriched data such as the percentage of change in incoming or outgoing link weights (pct_chng_inc_lnk_wt and pct_chng_out_lnk_wt) helps in identifying whether the communities are weakening or strengthening over time.

Table A6.5 *Data Enriched by Relation with Churned and Acquired Nodes Enrichment Category*

Few Columns of Default Enriched Data			Columns of Data Enriched by Relation with Churned and Acquired Nodes Category										
node_id	community_id	entity_role_nm	conct_acq_cnt	conct_acq_rl25	conct_acq_rl26	conct_chn_cnt	conct_chn_rl25	conct_chn_rl26	pct_chng_inc_lnk_wt	pct_chng_inc_msr20	pct_chng_out_lnk_wt	pct_chng_out_msr20	role_change_ind
A	11	Follower	1	0	1	0	0	0	100	100	64.444	62.5	0
B	9	Follower	0	0	0	2		1	0	0	.	.	0
C	11	Leader	3	1	2	0		0	88.889	41.176	14.286	37.838	0
D	9	Follower	0	0	0	2		1	0	0	0	0	1

Consider the table Table A6.5 on page 166. A value of 1 in the role change indicator column (role_change_ind) indicates that the role of a node has changed compared to the previous run of the project. Node C is connected to three new acquired nodes (conct_acq_cnt). Out of these three nodes, one node has role ID 25 (conct_acq_rl25) and the other two have role ID 26 (conct_acq_rl26). Assume that role ID 25 is a leader and role ID 26 is a follower. In this case, community ID 11 is an important community and node C is a stronger leader.

Aggregated Transactional Data

Assume that two transactional measures with IDs 20 and 22 are configured in the transactional table. For each of these transactional measures, the enriched data contains maximum, minimum, and total values for incoming and outgoing transactions. The columns of Table A6.6 on page 168 table indicate these values for node A, B, and C.

Table A6.6 *Data Enriched by Aggregated Transactional Data Enrichment Category*

Few Columns of Default Enriched Data		Columns of Data Enriched by Aggregated Transactional Data Category										
node_id	community_id	max_inc_msr20	max_inc_msr22	max_out_msr20	max_out_msr22	min_inc_msr20	min_inc_msr22	min_out_msr22	tot_inc_msr20	tot_inc_msr22	tot_out_msr20	tot_out_msr22
A	3	32	20	72	18	32	20	18	32	20	72	18
B	6	15	8	32	18	15	8	3	15	8	59	31
C	5	30	14	28	14	30	14	14	30	14	28	14

The transactional data can be used to identify the behavior of nodes in a community based on different transactional measures. The minimum, maximum, and total value of the transactional measures helps identify the transactional measure on which a node is performing, whether high or low. For example, for certain nodes, these transactional measures might indicate very low values. These nodes should potentially be targeted and the transactional measures of other nodes in the same community tracked and analyzed. The results can then be used to gain any insights into any of these nodes churning out of the network.

Community-Level Statistics and Associations with Neighboring Roles

The **Community-level statistics** and **Associations with neighboring roles** categories provide the statistics for each node and the relation that each node has with other nodes in the network. Statistics such as average incoming, maximum incoming, and minimum incoming indicate the pattern in which a node transacts with other nodes in the network. Similarly, these categories compile the data and enable network analysts to understand what role a node plays in the network. In addition, by analyzing the transactional usage pattern, you can further use the data enriched by these categories to identify nodes that have stronger links with other nodes.

Using Link-Level Enriched Data

Overview

The link-level business table that SAS Customer Link Analytics populates when you run the data enrichment process contains columns depending on the link-level enrichment categories that you choose. The following topics explain the columns that are populated as enriched data for each link-level enrichment category. These topics also explain how you can leverage the enriched data and use it further to enhance your campaign definitions. For more information about the data enrichment columns, see *SAS Customer Link Analytics: Data Reference Guide*.

Roles and Communities

When you enrich the project's output data using the default node-level enrichment category, **Roles and communities**, the enriched data contains the following columns.

Table A6.7 *Default Enriched Data*

from_node	to_node	from_betweenness_val	from_closeness_val	from_closeout_val	from_close_val	from_degreeinn_val	from_degreeout_val	from_degree_val	from_influence1_val	from_influence2_val	from_community_id	from_cs_segment_pk	from_entity_role_nm	link_weight_val	to_betweenness_val	to_closeness_val	to_closeout_val	to_close_val	to_degreeinn_val	to_degreeout_val	to_degree_val	to_influence1_val	to_influence2_val	to_community_id	to_cs_segment_pk	to_entity_role_nm
A	B	0.13333	0.03703	0.19104	0.11538	1	3	4	0.25549	0.05117	6	2	Follower	15	0	0.07661	0	0.04686	1	0	1	0	0	6	2	Follower
C	D	0.25	0.25391	0.11846	0.21282	1	1	2	0.16154	0.39403	5	1	Leader	14	0.33333	0.18223	0.07333	0.14711	1	1	2	0.50769	0.01791	5	1	Leader
E	A	0.01786	0.07141	0.09	0.08737	1	1	2	0.11538	0	1	4	Follower	18	0	0.10649	0	0.06514	2	0	2	0	0	1	4	Follower
F	G	0.06667	0.06171	0.05333	0.06364	1	1	2	0.06593	0	6	2	Follower	8	0	0.06632	0	0.04057	1	0	1	0	0	6	2	Follower
H	I	0	0	0.58577	0.28432	0	2	2	0.47308	0.54627	4	3	Follower	19	0.16667	0.15076	0.57311	0.37039	1	2	3	0.70385	0	4	3	Leader
J	C	0.25	0.47881	0.33377	0.45489	2	1	3	0.32308	0.66269	10	6	Leader	28	0.33333	0.27155	0.49091	0.40438	1	3	4	0.85385	0.25075	10	6	Leader

Note: The from_cs_segment_pk and to_cs_segment_pk columns are available only if you have opted for community segmentation when you run the Centrality Measures Computation workflow step.

The link-level default enriched data contains the characteristics of a node (from node) to another node (to node). The enriched data helps identify the values for each node in a link and also provides the link weight value. The link weight value signifies the strength of the two nodes: the higher the value, the more closely the nodes are linked to each other. You can use the enriched data to visualize the communities and the network in tools such as SAS Visual Analytics Explorer.

Node Attributes

Table A6.8 *Data Enriched by the Node Attributes Category*

from_node	to_node	from_city	from_cust_type_cd	from_payment_method_cd	from_subscrp_type_cd	to_city	to_cust_type_cd	to_payment_method_cd	to_subscrp_type_cd
A	B	Miami	Individual	Cash	Postpaid	Miami	Organization	Cash	Prepaid
C	D	Florida	Individual	Cash	Postpaid	LA	Individual	Credit card	Prepaid
E	F	NY	Individual	Credit Card	Postpaid	Cary	Individual	Cash	Prepaid

The node attributes at link level provide data that helps the marketing team establish connections between the nodes based on the attributes that they have. Nodes that are geographically separated from each other can be targeted for campaigns such as international subscriber dialing (ISD) campaigns for more usage.

Churn and Acquisition Indicators

Table A6.9 *Data Enriched by Churn and Acquisition Category*

from_node	to_node	from_acq_ind	from_churn_ind	link_acq_ind	link_churn_ind	to_acq_ind	to_churn_ind
A	B	1	0	0	1	0	0
C	D	0	0	1	0	0	0
A	E	1	0	0	0	0	0
G	H	0	1	0	1	0	1
J	K	0	0	0	0	0	0

New nodes can be acquired in the network and existing nodes can churn from the network. The data enriched by this category enables the campaign management teams to identify links that exist with the newly acquired nodes or links for which the nodes have churned out of the network. A value of 1 for the column link_acq_ind shows that C->D is a newly acquired link. Similarly, node G has churned out of the network and the link that existed between nodes G and H has also churned out. Marketing teams would be interested in designing campaigns for all such nodes that are linked with churned or acquired nodes.

Appendix 7

Sample Reporting Data for Viral Effect Analysis

The reporting data is produced during the first or subsequent scoring run of a scenario. In the first scoring run, the reporting data contains information about the nodes for the current run. However, for subsequent runs, the reporting data contains comparative information about nodes for two successive scoring runs of a scenario. The variables of the previous run are identified with the prefix PREV.

Note: Data for two consecutive scoring runs is available in the table only if the analytical model of the scenario is the same in both runs. However, if you rebuild the model, the reporting table contains data only for the current scoring run.

The reporting data can be categorized based on the type of variables that are populated in the table:

- Significant variables that the analytical model identifies.
- Variables that are produced for the node-level ABT variable category, **Roles and communities**. For more information about these variables, see *SAS Customer Link Analytics: Data Reference Guide*.
- Scenario-specific scores. For example, for the churn scenario type, the scores indicate the probability of a node to churn from the network based on whether its neighboring nodes have churned from the network. Similar scores are populated for the product adoption scenario. In addition, the product ID for which the scenario is defined is also captured.

The following table indicates the sample columns and records that are populated in the LASR reporting table that is created for viral effect analysis. Assume that the scenario ID is 7. For this scenario, the scen_LASR_rpt_7 database table is created and registered in the following metadata location: `/Shared Data/SAS Customer Link Analytics/Cust Link Analytics 6.5/Viral Effect Analysis/7/Data Sources/Report Data`. The following table indicates the sample records that are populated for the first scoring run of the scenario.

Table A7.1 Sample Records for First Scoring Run

node_id	mnog_msr4_torl2	toog_msr11_torl2	tot_out_msr11	scenario_score	between_val	influence2_val	influence1_val	role_nm	degreeout_val	degreein_val	degree_val	community_id	closeout_val	closein_val	close_val	product_id
1	20	40	46	0.141	0.41667	0.18621	0.61875	Leader	3	2	5	1	0.56386	0.25938	0.65709	Prod_1
4	33	98	124	0.126	0.5	0	0.3125	Default	1	1	2	207	0.35625	0.1441	0.39949	Prod_1
34	39	63	76	0.369	0	0.26724	0	Follower	0	1	1	205	0.68705	0.12529	0.09924	Prod_1

Here is the categorization of variables that are populated in this table:

- Significant variables
  - mnog_msr4_torl2
  - toog_msr11_torl2
  - tot_out_msr11
- Scenario-specific score and variables
  - scenario_score
  - product_id
- Node-level variables for the **Roles and communities** ABT variables category
  - node_id
  - between_val
  - influence2_val
  - influence1_val
  - role_nm
  - degreeout_val
  - degreein_val
  - degree_val
  - community_id
  - closeout_val
  - closein_val
  - close_val

In the next scoring run, these variables, except for node_id and product_id, are renamed with a prefix PREV. A new set of columns with the names mentioned here is added to the reporting table.

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*
- *SAS Customer Link Analytics: Migration 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 publications, go to sas.com/store/books. If you have questions about which titles you need, please contact a SAS Representative:

SAS Books
SAS Campus Drive
Cary, NC 27513-2414
Phone: 1-800-727-0025
Fax: 1-919-677-4444
Email: sasbook@sas.com
Web address: sas.com/store/books

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.

data store
a table, view, or file that is registered in a data warehouse environment. Data stores can contain either individual data items or summary data that is derived from the data in a database.

geodesic
the shortest distance between a pair of nodes.

grid
a collection of networked computers that are coordinated to provide load balancing of multiple SAS jobs, accelerated processing of parallel jobs, and scheduling of SAS workflows.

grid host
the machine to which the SAS client makes an initial connection in a SAS High-Performance Analytics application.

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

locked-down server
a SAS server that is configured with the LOCKDOWN system option, so that the server can access only designated host resources.

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.

SMP
See symmetric multiprocessing.

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

symmetric multiprocessing
a type of hardware and software architecture that can improve the speed of I/O and processing. An SMP machine has multiple CPUs and a thread-enabled operating system. An SMP machine is usually configured with multiple controllers and with multiple disk drives per controller.

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.

transactional measure
in a transactional table, a type of column that contains an aggregated value. For example, "call duration" is a transactional measure for a communications network, and "number of likes" is a transactional measure for a social network.

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.

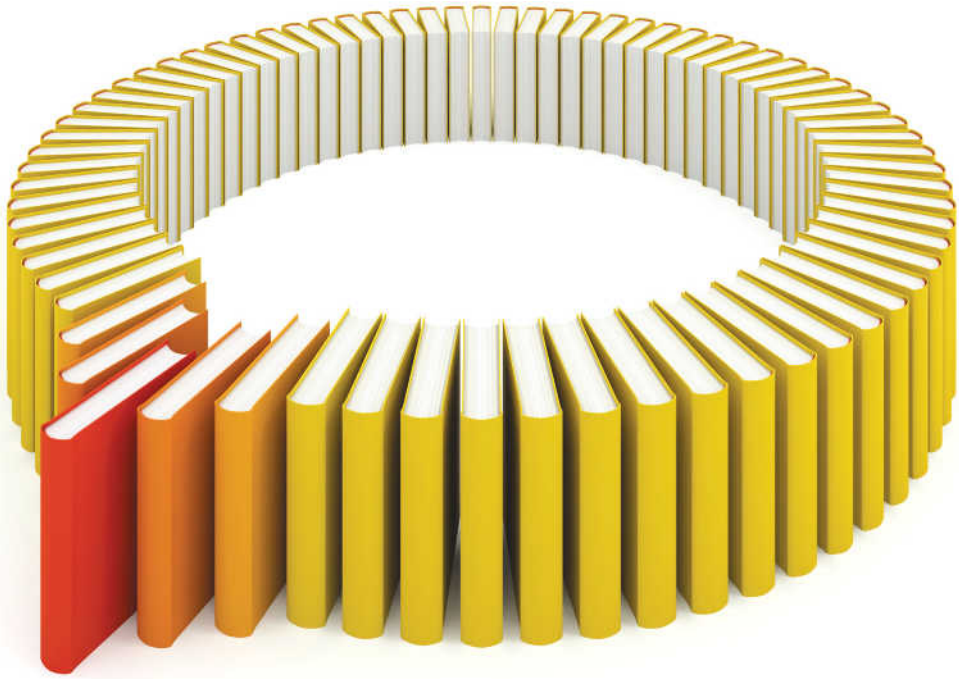
Index

S

T

V

W



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.