SPSS Modeler Social Network Analysis processes information about relationships between people into fields describing an individual's role in a social network, allowing social information to be included in predictive models.

About SPSS Modeler Social Network Analysis

Many approaches to modeling behavior focus on the individual. They use a variety of data about individuals to generate a model that uses the key indicators of the behavior to predict it. If any individual has values for the key indicators that are associated with the occurrence of the behavior, that individual can be targeted for special attention designed to prevent the behavior.

Consider approaches to modeling churn, in which a customer terminates his or her relationship with a company. The cost of retaining customers is significantly lower than the cost of replacing them, making the ability to identify customers at risk of churning vital. An analyst often uses a number of Key Performance Indicators to describe customers, including demographic information and recent call patterns for each individual customer. Predictive models based on these fields use changes in customer call patterns that are consistent with call patterns of customers who have churned in the past to identify people having an increased churn risk. Customers identified as being at risk receive additional customer service or service options in an effort to retain them.

These methods overlook social information that may significantly affect the behavior of a customer. Information about a company and about what other people are doing flows across the relationships to impact people. As a result, relationships with other people allow those people to influence a person's decisions and actions. Analyses that include only individual measures are omitting important factors having predictive capabilities.

SPSS Modeler Social Network Analysis addresses this problem by processing relationship information into additional fields that can be included in models. These derived key performance indicators measure social characteristics for individuals. Combining these social properties with individual-based measures provides a better overview of individuals and consequently can improve the predictive accuracy of your models.

SPSS Modeler Social Network Analysis consists of two primary components:

- SPSS Modeler Social Network Analysis nodes added to the SPSS Modeler environment that enable the inclusion of social analytic techniques in streams.
- SPSS Modeler Server Social Network Analysis, which adds processing of the node specifications to SPSS Modeler Server. SPSS Modeler Server Social Network Analysis efficiently processes massive amounts of network data, which may include millions of individuals and relationships, into a relatively small number of fields for further analysis.

For example, SPSS Modeler Social Network Analysis identifies the individuals in a network that are most affected by specific people churning. Furthermore, you can discover groups of individuals in a network that are at an increased risk of churn. By incorporating Key Performance Indicators for these effects in your models, you can improve their overall performance.

About social network analysis

A social network consists of a set of individuals and the relationships between them. Social network analysis examines these relationships to describe individuals and groups as parts of a social structure. Individuals interact with each other and these interaction patterns provide insight into the individuals involved. Relationships enable information to flow across a network, enabling one individual to influence another. The importance of the relationship information sets social network analysis apart from other approaches. Instead of focusing on each individual separately, the unit of study is a dyad consisting of two individuals and their relationships.

Relationships in a network can be classified as either directional or nondirectional. In a *directional relationship*, one individual is identified as the initiator, or source, of the relationship and the other is identified as the receiver, or destination. For example, making a phone call is a directional relationship in which one person calls another. In contrast, the roles of source and destination cannot be defined for *nondirectional relationships*. In this case, both parties participate in the relationship equally. Speaking to each other is an example of a nondirectional relationship.

Another property that distinguishes between relationships is whether the relationship is dichotomous or valued. The only information available in a *dichotomous relationship* is whether or not the relationship exists between two individuals. For every dyad in the network, the relationship is either present or absent. A *valued relationship*, on the other hand, includes a weight indicating the strength of the relationship. The weights allow the relationships to be compared to each other.

The "Relationship types" table lists examples for the cross-classification of relationships by direction and scale. In the directional relationships, *Joe* is the source of the relationship and *Mary* is the destination. In the nondirectional relationships, there is no indication of who initiated the relationship. The valued relationships use the length of the conversation as the relationship weight, while the dichotomous relationships either occurred or they did not.

Table 01. Relationship types

Direction	Scale	Example
Nondirectional	Dichotomous	Joe and Mary spoke to each other
Nondirectional	Valued	Joe and Mary spoke to each other for 20 minutes
Directional	Dichotomous	Joe called Mary
Directional	Valued	Joe called Mary for a 20 minute conversation

Displaying networks

A social network is typically illustrated using a *sociogram* [2]. In this type of visual display, individuals correspond to points, or nodes, in a space. Lines, or edges, connecting the points represent relationships between the individuals. If the relationships are directional, the edges include an arrow to indicate the direction. If the relationships have weights, the labels for the edges denote the values. The following graph displays a network for seven individuals.

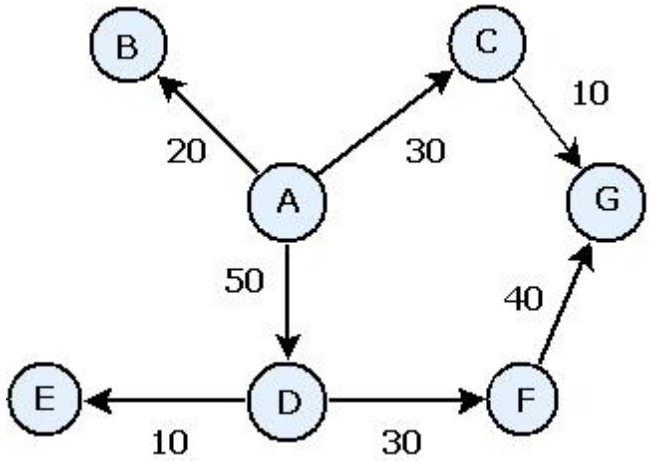


Figure 1. Example social network

Suppose the network represents the phone calls made by individuals with the relationship weights indicating the length of the calls. In this case, Person A called three people, spending the majority of time talking to person D.

This network is much smaller than those encountered in practice. However, the concepts illustrated by simple sociograms generalize to networks of any size and complexity.

Describing networks

Information about networks, groups, and individuals needs to be extracted into descriptive characteristics that allow cross-comparisons and inclusion in predictive models. Networks need to be distilled into a finite set of key performance indicators that can be analyzed. For example, you may want to compare networks or groups of nodes within a network to each other. Alternatively, you may want to compare individuals in the network to each other or identify the most important individuals.

Two measures commonly used to describe social networks are **density** and **degree**. Both statistics reflect connectivity, but the former focuses on the entire network or on network subgroups while the latter characterizes the individuals in the network.

Network density

For any set of nodes in a network, there is a finite number of relationships possible. Each node can serve as the source or the target of a relationship with every other node. Consider a network consisting of the three nodes A, B, and C. The following table lists all possible directed relationships between the nodes.

Table 2. Possible directed relationships for three nodes.

Source	Target
A	B
A	C
B	A
B	C
B	A
C	B

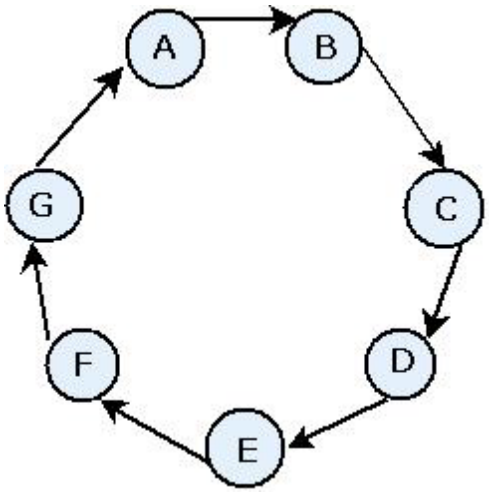
Each node is the source of a relationship with the other two nodes. However, in practice, all possible relationships may not actually be present. Some nodes may not have any direct relationship with other nodes. In addition, some directed relationships may not be reciprocated.

The *density* statistic represents the proportion of possible relationships in the network that are actually present. The value ranges from 0 to 1, with the lower limit corresponding to networks with no relationships and the upper limit representing networks with all possible relationships. The closer the value is to 1, the more dense is the network and the more cohesive are the nodes in the network.

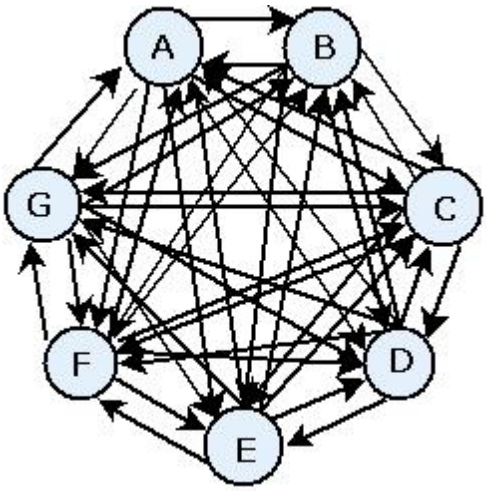
Information in dense networks can flow more easily than information in sparse networks. The "Sparse and dense networks" table displays two networks consisting of seven nodes. The sparse network includes only seven of the possible 42 relationships between the nodes, yielding a density of 0.17. The dense network, on the other hand, contains all possible relationships and has a density of 1.0.

Table 3. Sparse and dense networks.

Sparse network



Sparse network



In the sparse network, for information to flow from node A to node G, it must pass through five other nodes. In contrast, in the dense network, the information can go directly from node A to node G
.

Nodal degree

The important individuals in a network are often the ones involved in the most relationships. These individuals acquire information from a variety of sources and spread that information to a large number of other individuals. In contrast, individuals who participate in a few relationships cannot directly influence a large number of others in the network.

The *degree* for a node, defined as the total number of relationships involving that node, permits comparisons between network participants. Individuals with higher degree values are more active than those with lower values. Degree ignores the direction of the relationships, providing an overall measure of activity for the node.

For directed relationships, you can focus on whether a node is a source or target when tallying the number of relationships. The *indegree* for a node is the number of relationships in which a particular node is the target. Conversely, the *outdegree* is the number of relationships in which a node is the source. The following table lists the degree, indegree, and outdegree values for each node in the "Example social network" figure.

Table 5. SPSS Modeler Social Network Analysis nodes.

Node	Degree	Indegree	Outdegree
A	3	0	3
B	1	1	0
C	2	1	1
D	3	1	2
E	1	1	0
F	2	1	1
G	3	2	0

Indegree is often treated as a measure of prestige. Higher indegree values correspond to more relationships ending at that node. In other words, those individuals are contacted by a high number of other individuals. Many other nodes are initiating relationships with the node. Conversely, outdegree is

treated as a measure of centrality. Higher values correspond to more relationships originating from that node. Those individuals contact a high number of other individuals.

For the nodes in the example network, the degree values indicate that nodes A and D are the most active while nodes B and E are the least active. The indegree values reveal that node G has the most prestige. Based on the outdegree values, node A is the most central.

SPSS Modeler Social Network Analysis nodes

Along with the many standard nodes delivered with SPSS Modeler, you can also work with SPSS Modeler Social Network Analysis nodes to include the results of social network analysis in your streams. The “SPSS Modeler Social Network Analysis nodes" table describes these nodes, which are stored in the Sources palette.

Table 5. SPSS Modeler Social Network Analysis nodes.

Node	Description
Group Analysis	The Group Analysis node imports call detail record data from a fixed-field text file, identifies groups of nodes within the network defined by the records, and generates key performance indicators for the groups and individuals in the network.
Diffusion Analysis	The Diffusion Analysis node imports call detail record data from a fixed-field text file, propagates an effect across the network defined by the records, and generates key performance indicators summarizing the results of the effect on individual nodes

Node tabs

The SPSS Modeler Social Network Analysis nodes offer the following tabs for defining and previewing the analysis:

- **Data tab.** Used to identify the file containing the social network information.
- **Build Options tab.** Used to define settings for the analyses.
- **Analysis tab.** Used to view a summary of preliminary output that provides guidance for modifying the input settings on the Data tab to produce optimal results.

In addition, the nodes offer the following tabs common across source nodes in SPSS Modeler:

- **Filter tab.** Used to eliminate or rename the output fields produced by the nodes. This tab offers the same functionality as the Filter node.
- **Types tab.** Used to set measurement levels for the output fields produced by the nodes. This tab offers the same functionality as the Type node.
- **Annotations tab.** Used to rename nodes, supply a custom ToolTip, and store a lengthy annotation.

For more information about common tabs, consult the SPSS Modeler documentation.