

**IBM SPSS Modeler Social  
Network Analysis 17.1 用户  
指南**

**IBM**

**注释**

在使用本资料及其支持的产品之前，请阅读第 19 页的『声明』中的信息。

**产品信息**

此版本适用于 IBM(r) SPSS(r) Modeler V17.1.0 以及所有后续发行版和修订版，直到在新版本中另有声明为止。

---

# 目录

前言 . . . . .	v	划分为组 . . . . .	10
关于 IBM Business Analytics . . . . .	v	介绍组和组成员 . . . . .	10
技术支持 . . . . .	v	组分析节点 . . . . .	10
<b>第 1 章 IBM SPSS Modeler Social Network Analysis . . . . .</b>	<b>1</b>	为组分析指定数据 . . . . .	11
关于 IBM SPSS Modeler Social Network Analysis . . . . .	1	为组分析设置构建选项 . . . . .	11
关于社交网络分析 . . . . .	1	查看组分析统计量 . . . . .	12
显示网络 . . . . .	2	组分析的输出 . . . . .	12
介绍网络 . . . . .	2	<b>第 3 章 传播分析 . . . . .</b>	<b>15</b>
IBM SPSS Modeler Social Network Analysis 节点 . . . . .	4	传播分析概述 . . . . .	15
节点选项卡 . . . . .	4	传播分析示例 . . . . .	15
预览输出 . . . . .	5	传播分析节点 . . . . .	16
分析数据 . . . . .	5	为传播分析指定数据 . . . . .	16
应用程序 . . . . .	5	为传播分析设置构建选项 . . . . .	16
数据结构 . . . . .	6	查看传播分析统计量 . . . . .	17
脚本编制属性 . . . . .	6	传播分析的输出 . . . . .	17
<b>第 2 章 组分析 . . . . .</b>	<b>9</b>	声明 . . . . .	<b>19</b>
组分析概述 . . . . .	9	商标 . . . . .	20
确定社交相似性 . . . . .	9	索引 . . . . .	<b>23</b>



---

## 前言

IBM® SPSS® Modeler Social Network Analysis 将有关人与人之间的关系的信息处理成描述社交网络中的个人角色的字段，从而允许将社交信息包括在预测模型中。本手册介绍 IBM SPSS Modeler Social Network Analysis 节点在 IBM SPSS Modeler 环境中的使用方式，允许您在流中包括节点。当节点输出与表示个人度量的字段相结合时，会产生更完整的个人档案。

---

## 关于 IBM Business Analytics

IBM Business Analytics 软件提供完整、一致和正确的信息，决策者依据此信息来提高业务性能。企业智能、预测分析、财务业绩和战略管理的完整产品组合，和分析应用程序一起提供对当前业绩的清晰、直接和实用的洞察力，以及预测未来结果的能力。结合丰富的行业解决方案，久经证明的实践和专业服务，各种规模的组织都能够实现最高生产力、确信地自动作出决策以及获得更好的结果。

作为此产品服务组合的组成部分，IBM SPSS Predictive Analytics 软件可帮助组织预测未来事件，并在深入了解的基础上提前行动以实现更好的业务结果。减少欺诈和降低风险时，全球的商业、政府和学术客户都依赖 IBM SPSS 技术作为吸引、保留和增加客户的竞争优势。通过在日常运营中融入 IBM SPSS 软件，组织将成为预测型企业 - 即可以指引并实现决策的自动化，以满足企业目标并实现可衡量的竞争优势。有关详细信息或要联系一位代表，请访问 <http://www.ibm.com/spss>。

---

## 技术支持

技术支持可供维护客户使用。客户可就 IBM Corp. 产品使用问题或某一受支持硬件环境的安装帮助寻求技术支持。要寻求技术支持，请访问 IBM Corp. 网站 <http://www.ibm.com/support>。请求帮助时，请准备好您自己、您的组织和支持协议的相关信息。



---

# 第 1 章 IBM SPSS Modeler Social Network Analysis

---

## 关于 IBM SPSS Modeler Social Network Analysis

许多行为建模方法关注个人。这些方法使用各种有关个人的数据来生成使用行为的关键指标来预测行为的模型。如果任何个人具有与行为发生相关联的关键指标的值，那么可以特别关注该个人来防止相应行为。

请考虑流失建模方法，在这些方法中，客户终止其与公司的关系。挽留客户的成本远远低于更换客户的成员，这使得确定具有流失风险的客户的能力至关重要。分析员通常使用大量关键业绩指标来描述客户，包括人口统计信息以及每个单独的客户的最近通话模式。基于这些字段建立的预测模型使用与以往流失客户相同的客户通话模式的变化，确定流失风险不断上升的人员。确定为有流失风险的客户将获得额外的客户服务或服务选项，这样做是为了对他/她们进行挽留。

这些方法忽略了可能会极大地影响客户行为的社会信息。关于公司的信息以及关于其他人的动向的信息通过关系流动，从而对人们产生影响。因此，与其他人的关系允许这些人影响一个人的决策和行为。只包含个人度量的分析将省略具有预测功能的重要因素。

IBM SPSS Modeler Social Network Analysis 通过将关系信息处理为可包括在模型中的附加字段解决此问题。这些派生的关键业绩指标测量个人的社会特征。将这些社交属性与基于个人的度量相结合可以更好地概述个人，并因此能够提高模型预测准确性。

IBM SPSS Modeler Social Network Analysis 包括两个主要组件：

- IBM SPSS Modeler Social Network Analysis 节点添加到 IBM SPSS Modeler 环境，可使流中包括社交分析技巧。
- IBM SPSS Modeler Server Social Network Analysis，将节点规范的处理添加到 IBM SPSS Modeler Server。IBM SPSS Modeler Server Social Network Analysis 将包括数以百万计的个人和关系的极大量网络数据高效处理为相对少量的字段，以进行深入分析。

例如，IBM SPSS Modeler Social Network Analysis 识别网络中最受特定人员流失影响的个人。此外，您还可以发现网络中流失风险不断上升的个人群体。通过这些影响的关键业绩指标合并到模型中，您可以提高模型的整体性能。

---

## 关于社交网络分析

社交网络由一组个人以及他们之间的关系组成。社交网络分析将检查这些关系，以便将个人和组作为社交结构组成部分进行描述。个人之间进行交互，并且这些交互模式使得对所涉及个人有更深入的了解。关系使信息能够在网络中流动，从而使一个人能够影响另一个人。关系信息的重要性将社交网络分析与其他方法区分开来。学习单元是一个包含两个人及其关系的二分体，它不单独关注每个人。

网络中的关系可以分为方向关系或非方向关系。在方向关系中，会将一个人识别为关系的发起者，或源，另一个人则识别为接收者，或目标。例如，打电话是一个方向关系，在此关系中，一个人打电话给另一个人。相反，源和角色的角色不能定义为非方向关系。在这种情况下，双方都同等地参与到关系中。相互说话是一个非方向关系示例。

另一个用于区分关系的属性是关系是二分关系还是重视关系。二分关系中可用的唯一信息是该关系是否存在于两位个人之间。对于网络中的每个二分体，关系要么存在，要么不存在。另一方面，重视关系包括指示关系强度的一个权重。权重允许将关系进行比较。

“关系类型”表按方向和尺度列出关系的交叉分类示例。在方向关系中，*Joe* 是关系的源，*Mary* 是目标。在非方向关系中，未指示发起该关系的人员。重视关系使用对话长度作为关系权重，而二分关系要么发生，要么不发生。

表 1. 关系类型.

方向	尺度	示例
非方向	二分	Joe 和 Mary 相互通话
非方向	重视	Joe 和 Mary 相互通话 20 分钟
方向	二分	Joe 打电话给 Mary
方向	重视	Joe 打电话给 Mary 进行 20 分钟的对话

有关社交网络分析字段的更多信息，请参阅此领域的任一综合书籍<sup>1</sup>。

## 显示网络

通常使用社会关系图<sup>2</sup>来显示社交网络。在这种视觉显示中，个人对应于空间中的点或节点。用于连接点的线或边缘表示个人之间的关系。如果关系具有方向性，那么边缘将包括箭头来指示方向。如果关系具有权重，那么边缘标签将指示值。下图显示包含七个个人的网络。

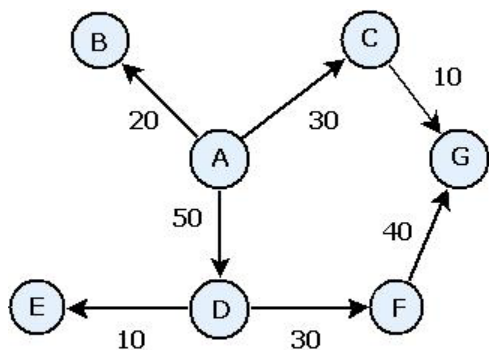


图 1. 社交网络示例

假定该网络表示个人所拨打的电话，并且关系权重指示通话长度。在本例中，人员 A 给三个人打电话，并且大部分时间用于与人员 D 通话。

该网络比实际上遇到的网络要小很多。然而，用简单的社会关系图说明的概念可推广到任意规模和复杂性的网络。

## 介绍网络

必须将有关网络、组和个人的信息抽取成允许进行交叉比较并包括在预测模型中的描述性特征。需要将网络提取为可以分析的有限关键业绩指标集合。例如，您可能想要将网络或网络中的节点组相互比较。另外，您可能想要将网络中的个人相互比较，或者确定最重要的个人。

通常用于说明社交网络的两种衡量方法是**密度**和**程度**。这两个统计量都反映连通性，但前者关注整个网络或网络子组，而后者描述网络中的个人的特征。

1. Wasserman, S., and K. Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.  
 2. Moreno, J. L. 1934. *Who Shall Survive?: Foundations of Sociometry, Group Psychotherapy, and Sociodrama*. Washington, D.C.: Nervous and Mental Disease Publishing Co..



## 网络密度

对于网络中的任何一组节点，都可能存在有限数目的关系。在与其他各个节点的关系中，每个节点都可以是源或目标。请考虑包含三个节点 A、B 和 C 的网络。下表列出了这些节点之间所有可能的定向关系。

表 2. 三个节点的可能定向关系。

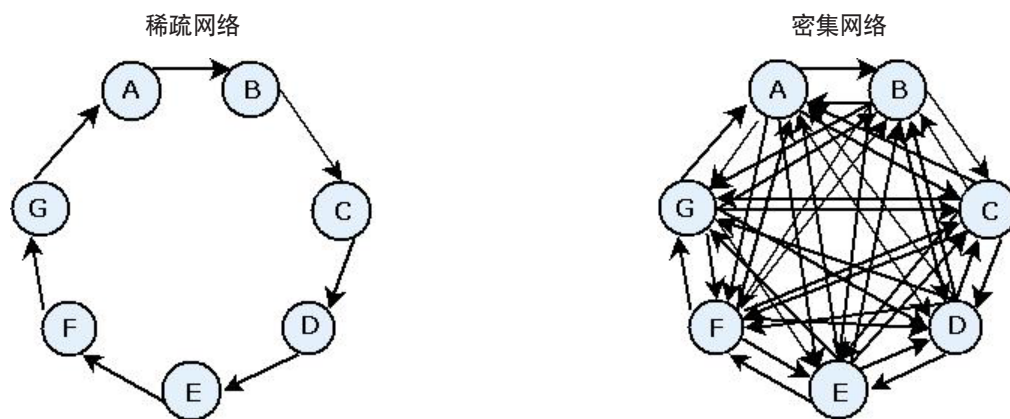
源	目标
A	B
A	C
B	A
B	C
C	A
C	B

每个节点都是与其他两个节点的关系中的源。然而，事实上，所有可能的关系实际上可能都不存在。某些节点可能与其他节点没有任何直接关系。另外，某些定向关系可能未给出。

密度统计量代表网络中实际存在的可能关系比例。值的范围为 0 到 1，其中较小的限制对应于没有任何关系的网络，而较大的限制表示具有所有可能关系的网络。值越接近 1，网络就越密集，并且网络中的节点更具凝聚力。

密集网络中的信息可能比稀疏网络中的信息更容易流动。“稀疏和密集网络”表显示两个包含七个节点的网络。稀疏网络只包括节点之间可能存在的 42 种关系中的 7 种关系，因此生成的密度为 0.17。另一方面，密集网络包含全部可能的关系，密度为 1.0。

表 3. 稀疏和密集网络。



在稀疏网络中，信息要从节点 A 流动到节点 G 必须通过五个其他节点。相反，在密集网络中，信息可以直接从节点 A 到达节点 G。

## 节点程度

网络中的重要个人通常是大多数关系中涉及的那些个人。这些个人从各种源获取信息，然后将信息分布给大量其他个人。相反，参与较少关系的个人不能直接影响网络中的大量其他人。

节点的程度，定义为涉及该节点的关系总数，允许在网络参与者之间对比。具有较高程度值的个人比具有较低程度值的个人更活跃。程度忽略关系的方向，对节点活动进行总体衡量。

对于定向关系，您可以在计算关系系数时关注节点是源还是目标。节点的人度是特定节点在其中作为目标的关系系数。相反，出度是节点在其中作为源的关系系数。下表列出了“社交网络示例”图中每个节点的程度值、入度值和出度值。

表 4. 程度值、入度值和出度值示例.

节点	程度	入度	出度
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	2	2	0



入度通常被视为对威望的测量。较高的人度值对应于更多终止于该节点的关系。换言之，那些个人得到许多其他人的联系。许多其他节点正启动与该节点的关系。相反，出度被视为对中心性的衡量。较高的值对应于更多源自该节点的关系。这些个人与大量其他个人联系。

对于网络示例中的节点，程度值指示节点 A 和节点 D 最活跃，而节点 B 和节点 E 最不活跃。入度值显示节点 G 具有最高威望。根据出度值，节点 A 处于最中央。

## IBM SPSS Modeler Social Network Analysis 节点

和 IBM SPSS Modeler 提供的许多标准节点一起，您还可使用 IBM SPSS Modeler Social Network Analysis 节点，将社交网络分析的结果包括在您的流中。“IBM SPSS Modeler Social Network Analysis 节点”表描述这些存储在“源”选用板中的节点。

表 5. IBM SPSS Modeler Social Network Analysis 节点.

节点	图标	描述
组分析		“组分析”节点从固定字段的文本文件中导入通话详细记录数据、标识记录所定义的网络中的节点组，并生成网络中的组和个人的关键业绩指标。请参阅主题第 9 页的『组分析概述』以获取更多信息。
传播分析		“传播分析”节点从固定字段的文本文件中导入通话详细记录数据、在记录所定义的网络中传播影响，并生成对各个节点的影响结果进行汇总的关键业绩指标。请参阅主题第 15 页的『传播分析概述』以获取更多信息。

## 节点选项卡

IBM SPSS Modeler Social Network Analysis 节点为定义和预览分析提供以下选项卡：

- “数据”选项卡。用于确定包含社交网络信息的文件。
- “构建选项”选项卡。用于定义分析的设置。
- “分析”选项卡。用于查看初步输出的摘要，它提供修改“数据”选项卡上的输入设置以产生优化结果的指导。

此外，节点提供 IBM SPSS Modeler 中源节点常用的以下选项卡：

- “过滤器”选项卡。用于消除或重命名节点产生的输出字段。此选项卡所提供的功能与过滤节点相同。

- “类型”选项卡。用于设置节点产生的输出字段的测量级别。此选项卡所提供的功能与类型节点相同。
- “注释”选项卡。用于重命名节点、提供自定义工具提示和存储较长的注解。

有关常用选项卡的详细信息，请参阅 IBM SPSS Modeler 文档。

## 预览输出

通常，您将节点的输出用作预测模型的输入。如果要查看输出，可以向流中添加表节点。然而，数据量通常相当大，使得生成表要耗用很长时间。此外，表中的行数还限制表的有用性。

替代方法是，可以生成将显示要创建的输出样本的预览表。预览将显示为有限数目的行生成的字段。行数在流属性中定义。要预览节点输出，请执行下列步骤：

1. 打开节点。
2. 在“数据”选项卡上指定数据设置。
3. 在“构建选项”选项卡上定义分析参数。
4. 单击**预览**。

此时预览窗口将打开，此窗口中显示结果。此外，预览输出还会在节点的“分析”选项卡上填充结果摘要概括。

## 分析数据

确定生成最佳结果的分析设置通常是一个迭代式过程。您需要定义设置，运行分析，然后查看结果。如果结果并非那么有用，您可以修改设置并重新运行分析。

要分析节点的输入数据，请执行下列步骤：

1. 打开节点。
2. 在“数据”选项卡上指定数据设置。
3. 在“构建选项”选项卡上定义分析参数。
4. 单击**分析数据**。

如果构建选项指示应显示汇总统计量，那么“分析”选项卡将显示结果。

如果需要重新运行分析，请单击**清除分析**以清空当前结果，然后单击**分析数据**。

---

## 应用程序

IBM SPSS Modeler Social Network Analysis 中特别有益的具体应用程序包括：

- **流失预测。** 组特征可以影响流失率。通过关注组中流失风险不断上升的个人，您可能能够防止流失。另外，您还可以确定由于来自已流失个人的信息流而有流失风险的个人。
- **利用组领导。** 组领导显著影响其他组成员。如果可以防止组领导流失，那么可以降低组成员的流失率。另外，尝试使竞争对手的组领导流失可以提高与该竞争对手相关联的组成员的流失率，并同时降低与您的公司相关联的组成员的流失率。
- **市场营销。** 可通过组领导发起新的商品和服务供应。领导的影响力可以使其他组成员更可能购买产品。您可以使用传播分析来确定受组领导影响最大的个人，并且将这些个人作为您的市场营销目标。

IBM SPSS Modeler Social Network Analysis 提供了两种演示流，为您提供了有关可如何在流中包含社交网络分析结果的示例。数据文件和样本流安装在产品安装目录下的 *Demos* 文件夹中。

- **DA 演示** *streams.str* 提供了有关分析数据以识别最有可能流失的排名前 30 位个人的示例。

- GA 演示 *streams.str* 提供的示例使用关键绩效指标 (KPI) 同时预测组和个人的流失，并且使用 KPI 定位市场营销活动的特定个人。

## 数据结构

有关网络中感兴趣的个人的信息可以通过各种文件、数据库和系统在整个企业内传播。要使用 IBM SPSS Modeler Social Network Analysis 分析网络，您需要从数据源提取相关记录和字段，并对它们进行格式化以输入节点中。

分析节点需要存储在单个固定宽度的文本文件中的通话详细记录。该文件的每一行对应于一种关系，而数据则按以下几列进行组织：

- 发起关系的个人的标识。
- 作为关系目标的个人的标识。
- 关系的权重（可选）。

所有数据都必须是数值，并且个人的标识限制为整数。可选择将字段名称作为第一个数据行包括在文件中。“通话数据记录示例”表显示了这种数据结构。

请注意，头记录必须使用数字或者空白值。例如，值为 *EF BB BF* 的头将会导致错误。

表 6. 通话数据记录示例。

源	目标	权重
1000	5642	243
2190	8444	831
0299	9419	559

权重值可以对应于任何要用来表示关系相对于网络中的其他关系的重要性的度量。对于通话数据，常用权重包括通话持续时间或通话频率。请注意，这适用于“传播分析”，但“组分析”仅支持呼叫频率。

如果您希望分析专注于通话历史记录子集，那么必须在创建输入文件时使用该子集。例如，通过仅将过去数月或某人的最近通话包括在输入文本文件中，可以限制为只分析这些数据。

## 脚本编制属性

IBM SPSS Modeler Social Network Analysis 中的脚本编写是用于在用户界面上实现过程自动化的强大工具。您使用鼠标或键盘进行的操作，借助脚本同样可以完成，而且使用脚本可以自动执行那些手动执行将造成大量重复操作或高耗时的任务。有关使用脚本的说明，请参阅 IBM SPSS Modeler 提供的 *ScriptingAutomation.pdf* 指南。

### 传播节点属性

下表列出了“传播”节点的脚本编制属性。

属性名称	数据类型	属性说明
<code>input_data_file_name</code>	字符串	
<code>calling_field</code>	字段	
<code>called_field</code>	字段	
<code>frequency_weight_field</code>	字段	

属性名称	数据类型	属性说明
read_field_names	布尔型	
diffusion_list_file_name	字符串	
spreading_factor	双精度型	缺省值: 50。 最小值: 1 最大值: 99。
max_number_iterations	整型	缺省值: 100。 最小值: 1。
accuracy_threshold	双精度型	缺省值: 0.01。 最小值: 0.001。
calculate_statistics	布尔型	

### 组分析节点属性

下表列出了“组分析”节点的脚本编制属性。

属性名称	数据类型	属性说明
input_data_file_name	字符串	
calling_field	字段	
called_field	字段	
frequency_weight_field	字段	
read_field_names	布尔型	
coverage_threshold	双精度型	缺省值: 10。 最小值: 1 最大值: 99。
min_group_size	整型	缺省值: 2。 最小值: 2。
max_group_size	整型	缺省值: 100。 最小值: 2。
calculate_statistics	布尔型	



## 第 2 章 组分析

### 组分析概述

组分析使用网络中个人的交互模式来识别包含相似个人的组。这些组的特征将影响个人组成员的行为。例如，拥有许多成员间关系和强大领导的小型组面临不断上升的流失风险，即使实际上并没有组成员流失。<sup>3</sup>与只包含个人度量的模型相比，同时包含组度量和个人度量的预测模型表现更好。

组分析包含下列一般步骤：

1. 确定反映社会邻近性的关系强度。请参阅主题『确定社交相似性』以获取更多信息。
2. 根据关系强度并遵循大小限制将网络划分为多个组。请参阅主题第 10 页的『划分为组』以获取更多信息。
3. 描述组和个人，包括确定组领导。请参阅主题第 10 页的『介绍组和组成员』以获取更多信息。

### 确定社交相似性

组成员之间应该比组成员与非组成员之间的相似性更高。在网络分析中，两个节点的相似性取决于它们的关系。对于网络中的任何节点，都有一组节点是与该节点的定向关系的目标。对于远程通信数据，这组节点对应于特定个人联系的所有人员。如果两个人与同一组人员联系，那么这两个人被视为彼此相似。两个人的关系目标集重叠越多，那么这两个人越相似。

请考虑“十节点网络示例”图中显示的网络。

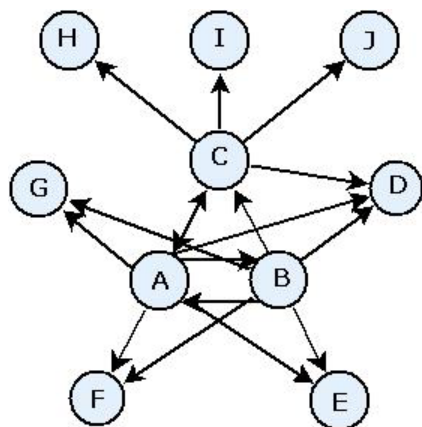


图 2. 十节点网络示例

“目标节点”表显示源自节点 A、B 和 C 的关系的目标节点。

表 7. 目标节点.

源节点	目标节点
A	B, C, D, E, F, G
B	A, C, D, E, F, G

3. Richter, Y., E. Yom-Tov, and N. Slonim. 2010. Predicting customer churn in mobile networks through analysis of social groups. In: *Proceedings of the 2010 SIAM international conference on data mining*. Columbus, OH: SDM 2010.

表 7. 目标节点 (续).

源节点	目标节点
C	A, D, H, I, J

节点 A 和 B 有五个共同的目标节点。另一方面，节点 A 和 C 只有一个共同的目标节点。因此，与节点 A 和 C 相比，节点 A 和 B 在社交方面更相似。

关系权重（例如通话持续时间或频率）不捕获网络中节点的相似性。为了包括相似性的概念，组分析使用交互信息<sup>4</sup>作为关系权重。此统计量反映两个节点与同一个节点连接的可能性。如果网络中的关系具有预定义权重（例如远程通信数据的通话频率），那么交互信息会相应地包括这些权重。

## 划分为组

组应该包含与其他组成员之间的关系权重较高的个人，其中权重用于度量关系中节点的相似性。因此，组标识从省略网络中的较弱关系开始。范围阈值通过定义要保持的最强关系分数控制此过程。例如，如果范围阈值为 0.4，那么将会导致 40% 的最强关系用于组标识，并且剩余的 60% 关系被忽略。

其余关系可能会生成一些非常小或非常大的组，这些组的预测能力有限。为了防止将这种组包括在分析中，您可以指定最小组大小和最大组大小。大小小于最小值的组将被完全省略。然而，大小大于最大值的组将被划分成位于可接受大小范围内的较小组。丢弃弱关系和实施大小限制后剩余的组则为核心组。

从原始网络中除去关系可能会导致某些个人不属于任何核心组。然而，这些个人可能与保证包括在组中的组成员具有联系。如果存在许多与组核心成员的相对较强关系，那么只要不违反组大小限制，就会向该组中添加个人。因此，最终组包含一组核心成员以及由于与核心成员联系而添加的成员。

## 介绍组和组成员

除了密度、入度和出度外，还有其他统计量描述组动态。特别是，权威分数和传播分数提供对组内个人的社会地位的测量。尝试同时预测组及其成员的行为时，组中每个人的角色至关重要。

节点的权威分数衡量组中其他节点连接到自身的趋势。如果许多人都与特定人员联系（可能是咨询信息或征求意见），那么该人员具有权威角色。组中节点的权威分数（对应于在整个组网络中随机游走的平稳概率）的范围是 0 到 1。权威分数越接近 1，节点在组中的权威就越大。组中拥有最高权威分数的节点则为该组的权威领导。用组中的最高分数除以最低分数提供了一种测量权威领导的整体强度的方法。

相反，节点的传播分数衡量节点连接到组中其他节点的趋势。如果特定人员与组中的许多其他人员联系，那么该人员可能会极大地影响整个组的意见。组中节点的传播分数（对应于按相反方向在整个组网络中随机游走的平稳概率）的范围是 0 到 1。传播分数越接近 1，节点与其他组成员的联系就越紧密。组中拥有最高传播分数的节点则为该组的传播领导。用组中的最高分数除以最低分数提供了一种测量传播领导的整体强度的方法。

## 组分析节点

“组分析”节点（可以从“源”选用板访问）识别网络中那些在社交方面彼此相似的个人集合，并确定组中个人的相对社会地位。拥有许多成员间联系且包含强大领导的相对较小组由于组动态而面临不断上升的流失风险。您可以使用结果来确定组流失分数，此分数与个人流失分数相结合能够提高只基于个人分数预测模型流失的能力。

4. Cover, T. M., and J. A. Thomas. 2007. *Elements of Information Theory, 2nd edition*. New York: John Wiley and Sons, Inc.



考虑到通常构成网络的数据量以及作为组定义基础的社交互动的性质，组分析是一个耗时的过程。通常，您可以将组分析结果保存至要用作预测模型的输入的数据库或文件。组结果的刷新频率相对较低（例如每月刷新一次），而模型的刷新频率要高很多。

**需求。** 该节点需要一个使用三个字段来定义社交网络的固定宽度文本文件。一个字段标识每个定向关系的源，一个字段定义每个定向关系的目标，第三个字段为每个关系指定可选的强度。网络中的所有关系都必须具有方向性。

## 为组分析指定数据

通过“组分析”源节点窗口的“数据”选项卡，您可以指定包含网络节点关系的输入文件。

**文件。** 指定包含通话详细记录的文件或文件夹的名称。您可以输入名称，也可以单击省略号按钮 (...) 从文件系统中选择名称。在您选择名称后，路径将会显示，并且其内容使用定界符显示。如果您指定文件夹，那么该文件夹中包含的所有文件内的通话详细记录将会合并以供分析；该文件夹中的所有文件应具有相同的结构。

**读取文件中的字段名称。** 缺省情况下，此选项处于选中状态，它将数据文件中的第一行视为列名。如果第一行不是标题，请取消选择此选项，以自动为每个字段提供普通名称，例如 *Field1* 和 *Field2*。

### 网络定义设置

网络定义设置定义字段的角色。

**字段。** 使用方向按钮可以将此列表中的项手动分配给各种角色字段。图标指示各个角色字段的有效测量级别。单击**全部**按钮可以选择列表中的所有字段，或单击单独的测量级别按钮以选择具有此测量级别的所有字段。

**源。** 选择一个字段作为定向关系的起点。

**目标。** 选择一个字段作为定向关系的目标。

**权重。**（可选）选择一个对应于网络中关系发生次数的字段。在确定关系强度时，分析相应地确定记录的权重。

## 为组分析设置构建选项

通过“组分析”源节点窗口的“构建选项”选项卡，您可以定义用于标识网络中的组的选项。

### 组分析设置

组分析设置将影响所构成的组的大小和相对强度。

**范围阈值。** 定义要在分析中使用的最强网络关系分数。例如，范围阈值为 0.2 表示仅使用其权重在所有权重中位列前 20% 的关系。范围在 0 到 1 之间的这个参数允许分析关注组中应发生的最强关系。阈值越大，所生成的组的成员之间的关系就越弱。请参阅主题第 10 页的『划分为组』以获取更多信息。

**最小组大小。** 指定组大小的下限。将不返回小于此值的组。

**最大组大小。** 指定组大小的上限。大于此值的组将被划分为较小的组。

**计算并显示汇总统计量。** 选中此项后，除了导出关键业绩指标输出外，节点还计算并显示分析的汇总统计量。对于非常大的网络，计算这些统计量可能会对节点性能产生负面影响。

## 查看组分析统计量

“组分析”源节点窗口的“分析”选项卡提供网络中识别的组的汇总概括。“组分析汇总统计量”表列出组的可用汇总统计量。

表 8. 组分析汇总统计量.

统计量	描述
组中的节点总数	识别的组中包括的节点数
组中的链接总数	识别的组中包括的链接数
组的总数	网络中识别的组总数
平均组大小	组中的平均节点数
平均组密度	组中节点之间的平均直接连接分数。请参阅主题第 3 页的『网络密度』以获取更多信息。
核心成员的平均分数	组中核心节点的平均分数。请参阅主题第 10 页的『划分为组』以获取更多信息。
核心组的平均密度	组中核心节点之间的直接联系平均分数
平均入度	平均传入链接数。请参阅主题第 3 页的『节点程度』以获取更多信息。
平均出度	平均传出链接数。请参阅主题第 3 页的『节点程度』以获取更多信息。

从“汇总统计量”表中选择特定统计量可以查看值的分布、标准偏差和统计量的偏度。

查看这些统计量有助于定义组分析设置。理想情况下，结果应该显示具有较高密度值的相对较小组。例如，如果有一些非常大的组，请考虑减小最大组大小，以将它们分为较小的组。另外，如果组密度值往往较小，请考虑减小范围阈值以关注更强的关系。

## 组分析的输出

“组分析”节点生成各种描述组以及组中的个人的字段。您可以使用这些关键业绩指标来扩充现有模型和数据，以改进模型所生成的预测。例如，您可以更新个人流失倾向值以包括组影响。

“组的关键业绩指标”表列出分析中标识的组的关键业绩指标字段。

表 9. 组的关键业绩指标.

字段	描述
GAG_GroupNumber	组的唯一标识
GAG_Size	组中的个人数
GAG_Density	组中个人之间的直接联系分数。请参阅主题第 3 页的『网络密度』以获取更多信息。
GAG_KernelDensity	组中核心个人之间的直接联系分数
GAG_CoreNodesFraction	组中核心个人的分数。请参阅主题第 10 页的『划分为组』以获取更多信息。
GAG_MaxRankType1	任何组成员的最大权威分数。请参阅主题第 10 页的『介绍组和组成员』以获取更多信息。
GAG_MinRankType1	任何组成员的最小权威分数
GAG_MaxMinRankRatioType1	最大权威分数与最小权威分数的比率。此值反映组领导的权威强度。
GAG_MaxRankType2	任何组成员的最大传播分数。请参阅主题第 10 页的『介绍组和组成员』以获取更多信息。

表 9. 组的关键业绩指标 (续).

字段	描述
GAG_MinRankType2	任何组成员的最小传播分数
GAG_MaxMinRankRatioType2	最大传播分数与最小传播分数的比率。此值反映组领导的传播强度。

“个人的关键业绩指标”表列出网络中的个人的关键业绩指标字段。

表 10. 个人的关键业绩指标.

字段	描述
GAI_NodeNumber	个人的唯一标识
GAI_CoreNode	这是指示个人是否是组的核心个人的指示符。请参阅主题第 10 页的『划分为组』以获取更多信息。
GAI_RankType1	个人的权威分数。请参阅主题第 10 页的『介绍组和组成员』以获取更多信息。
GAI_RankOrderType1	基于权威分数的组中排名顺序
GAI_RankType2	个人的传播分数。请参阅主题第 10 页的『介绍组和组成员』以获取更多信息。
GAI_RankOrderType2	基于传播分数的组中排名顺序
GAI_InDegree	个人作为关系目标的关系数。请参阅主题第 3 页的『节点程度』以获取更多信息。
GAI_OutDegree	个人作为关系源的关系数。请参阅主题第 3 页的『节点程度』以获取更多信息。
GAI_GroupLeaderType1	无论节点是否是权威领导，其领导能力评分都派生自传入链接。请参阅主题第 10 页的『介绍组和组成员』以获取更多信息。
GAI_GroupLeaderConfidenceType1	表明节点是权威领导的置信度。
GAI_GroupLeaderType2	无论节点是否是传播领导，其领导能力评分都派生自传出链接。请参阅主题第 10 页的『介绍组和组成员』以获取更多信息。
GAI_GroupLeaderConfidenceType2	表明节点是传播领导的置信度。



## 第 3 章 传播分析

### 传播分析概述

传播分析可识别在社交网络中受其他人影响最大的个人，并将影响量化为传播能量。该过程使用一种扩散激活方法，使用此方法时，影响会从网络节点反复扩散到它们的直接近邻，随着它在节点之间的移动而减少大小<sup>5</sup>。接收到能量时，节点将被激活，并且将该能量的一部分传送到作为该节点定向关系目标的所有近邻节点。

传播系数定义一个激活的节点传送的能量比例，剩余量将由该节点保留。任何接收到此能量的节点本身都会将相同比例传送给其近邻，从而使能量传送成为一个衰退过程。传播系数大表示发送更多能量，允许能量在过程完全衰退前，到达距离初始激活的节点更远的节点。较小的传播系数会导致传播过程快速衰退，剩余的已传送能量则相对较接近初始节点。

激活的节点所传播的能量总量会分布在作为该节点定向关系目标的所有节点之间。每个节点接收到的数量取决于和激活的节点的关系强度。向特定节点发送的部分等于关系权重除以激活的节点作为源的所有关系的总权重。因此，拥有较高相对权重关系的近邻比拥有较低相对权重的近邻接收的能量要多。

如果发生下列其中一种情况，那么传播过程将停止：

- 激活的节点不是任何定向关系的源
- 传送的能量低于精度阈值（对为使过程能继续而传送的能量的限制）
- 迭代数达到指定限制

传播完成后，拥有最多传播能量的节点对发起该过程的影响最敏感。例如，如果该过程从流失的节点开始，那么拥有最高能量的节点就面临最大的流失风险。您可以特别关注这些节点，以防止它们流失。

### 传播分析示例

请考虑第 2 页的『显示网络』中显示的网络。该网络包含七个节点，每个节点之间都具有强度不同的定向关系。

“传播过程示例”表显示网络中使用传播系数 0.80 的传播过程。最初，节点 A 包含所有能量，被任意分配一个值 1.00。在步骤 1 中，此节点将激活，并将其 80% 的能量传播给三个作为关系目标的近邻，同时为自己保留 20%。关系的总权重为 100。与节点 D 的关系占此总量的一半，因此这个节点接收一半传播能量，或 0.40。与节点 B 的关系占总权重的 20%，因此生成的传播能量值为 0.16。节点 C 接收剩余能量 0.24，这占源自节点 A 的能量数的 30%。

表 11. 传播过程示例.

步骤	节点 A	节点 B	节点 C	节点 D	节点 E	节点 F	节点 G
0	1.00	0	0	0	0	0	0
1	0.20	0.16	0.24	0.40	0	0	0
2	0.20	0.16	0.05	0.08	0.08	0.24	0.19
3	0.20	0.16	0.05	0.08	0.08	0.08	0.38

5. Dasgupta, K., R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjee, A. A. Nanavati, and A. Joshi. 2008. Social ties and their relevance to churn in mobile telecom networks. In: *Proceedings of the 11th international conference on extending database technology*. New York, NY: ADM.

接收到步骤 1 中的能量后，节点 B、C 和 D 将在步骤 2 中激活，并将能量传播给任何目标近邻。节点 B 没有目标节点，因此它不传播任何能量。另一方面，节点 C 具有目标，并且将其 80% 的能量传播到节点 G。最后，节点 D 将其 80% 的能量传播到节点 E 和 F，其中 F 接收的能量数是 E 的三倍，这是因为 F 与 D 的关系更强。

在步骤 3 中，节点 E、F 和 G 将激活。节点 E 和 G 没有目标节点，因此这两个节点都不传播任何能量。然而，节点 F 将其 80% 的能量传播到它的单一目标节点 G。此节点现在拥有的总能量值为 0.38，这既包括从节点 C 传播的能量，也包括从节点 F 传播的能量。

此时，节点 G 将激活，但没有传播的目标节点，因此该过程将终止。源自节点 A 的能量已在整个网络中传播，其中节点 G 接收到最多能量。如果网络中引入的能量表示流失，那么节点 G 受节点 A 流失的影响最大。

---

## 传播分析节点

“传播分析”节点（可以从“源”选用板访问）将影响从一组指定的个人通过社交网络进行传播，并使用网络关系来识别受该影响最大的个人。例如，如果影响是流失，那么该节点将识别那些由于网络中的其他特定个人流失而最可能流失的个人。您可以使用节点输出来扩充现有模型和数据，以改进这些模型所生成的预测。例如，您可以更新个人流失倾向值以包括传播影响。

**需求。** 节点需要包括要分析数据的两个固定宽度文本文件。第一个文件使用三个字段定义社交网络。一个字段标识每个定向关系的源，一个字段定义每个定向关系的目标，第三个字段为每个关系指定可选的强度。网络中的所有关系都必须具有方向性。第二个文件包含一个标识列表，影响应该从这些标识开始。

## 为传播分析指定数据

通过“传播分析”源节点窗口的“数据”选项卡，您可以指定包含网络节点关系的输入文件。

**文件。** 指定包含通话详细记录的文件或文件夹的名称。您可以输入名称，也可以单击省略号按钮 (...) 从文件系统中选择名称。在您选择名称后，路径将会显示，并且其内容使用定界符显示。如果您指定文件夹，那么该文件夹中包含的所有文件内的通话详细记录将会合并以供分析；该文件夹中的所有文件应具有相同的结构。

**读取文件中的字段名称。** 缺省情况下，此选项处于选中状态，它将数据文件中的第一行视为列名。如果第一行不是标题，请取消选择此选项，以自动为每个字段提供普通名称，例如 *Field1* 和 *Field2*。

网络定义设置

网络定义设置定义字段的角色。

**字段。** 使用方向按钮可以将此列表中的项手动分配给各种角色字段。图标指示各个角色字段的有效测量级别。单击**全部**按钮可以选择列表中的所有字段，或单击单独的测量级别按钮以选择具有此测量级别的所有字段。

**源。** 选择一个字段作为定向关系的起点。

**目标。** 选择一个字段作为定向关系的目标。

**权重。** 选择一个表示关系权重的字段。对于通话详细记录，权重可以对应于反应关系强度的许多通话特征中的任意特征，例如通话持续时间或通话频率。

## 为传播分析设置构建选项

通过“传播分析”源节点窗口的“构建选项”选项卡，您可以定义传播过程的选项。

传播种子列表

传播种子列表标识应该从其开始传播能量的网络节点。例如，在流失分析中，此列表标识已流失的网络节点。另外，此列表还包含流失风险最大的节点，以便在这些节点实际流失时探查网络上的影响。

**文件。** 指定包含节点标识的文本文件的名称，每个标识都对应于该文件中的一行。您可以输入文件名或单击省略号按钮 (...) 来选择文件。

### 传播分析设置

传播分析设置确定传播的积极程度以及该过程的停止时间。

**传播系数。** 定义节点在传播过程中的每一步所传输的能量分数。较小的值限制能量只能传播到距离正在传播能量的节点最近的节点。相反，值越大，能量能够影响的节点就越远。

**最大迭代次数。** 指定传播迭代次数的上限。在达到此限制时，传播过程自动终止。

**准确性阈值。** 指定节点能量中的最小变化，该变化能保证传播过程继续。如果每个节点的能量变化均小于此值，那么传播过程将终止。

**计算并显示汇总统计量。** 选中此项后，除了导出关键业绩指标输出外，节点还计算并显示分析的汇总统计量。对于非常大的网络，计算这些统计量可能会对节点性能产生负面影响。

## 查看传播分析统计量

“传播分析”源节点窗口的“分析”选项卡提供传播结果的摘要概括。“传播分析汇总统计量”表列出可用的汇总统计量。

表 12. 传播分析汇总统计量.

统计量	描述
网络中的节点总数	网络中的节点数
网络中的链接总数	网络中的链接数
网络中的传播种子总数	用作为传播过程种子的节点数
平均影响力	与个人相关联的传播能量的平均数量。
平均入度	个人作为关系目标的平均关系数。请参阅主题第 3 页的『节点程度』以获取更多信息。
平均出度	个人作为关系源的平均关系数。请参阅主题第 3 页的『节点程度』以获取更多信息。

从“汇总统计量”表中选择特定统计量可以查看值的分布、标准偏差和统计量的偏度。

## 传播分析的输出

“传播分析”节点生成各种描述网络中的个人的字段。您可以使用这些关键业绩指标来扩充现有模型和数据，以改进模型所生成的预测。例如，您可以更新个人流失倾向值以包括传播影响。

“传播分析关键业绩指标”表列出分析中个人的关键业绩指标字段。

表 13. 传播分析关键业绩指标.

字段	描述
DA_NodeNumber	个人的唯一标识
DA_DiffusedEnergy	与个人相关联的传播能量。对于流失分析，较大的值比较小的值所指示的流失倾向更大。

表 13. 传播分析关键业绩指标 (续).

字段	描述
DA_InDegree	个人作为关系目标的关系数。请参阅主题第 3 页的『节点程度』以获取更多信息。
DA_OutDegree	个人作为关系源的关系数。请参阅主题第 3 页的『节点程度』以获取更多信息。



---

## 声明

这些信息开发用于在全球提供的产品和服务。

IBM 可能在其他国家或地区不提供本文中讨论的产品、服务或功能特性。有关您当前所在区域的产品和服务的信息，请向您当地的 IBM 代表咨询。任何对 IBM 产品、程序或服务的引用并非意在明示或暗示只能使用 IBM 的产品、程序或服务。只要不侵犯 IBM 的知识产权，任何同等功能的产品、程序或服务，都可以代替 IBM 产品、程序或服务。但是，评估和验证任何非 IBM 产品、程序或服务，则由用户自行负责。

IBM 公司可能已拥有或正在申请与本文档内容有关的各项专利。提供本文档并未授予用户使用这些专利的任何许可。您可以用书面方式将许可查询寄往：

IBM Director of Licensing  
IBM Corporation  
North Castle Drive  
Armonk, NY 10504-1785  
U.S.A.

有关双字节 (DBCS) 信息的许可查询，请与您所在国家或地区的 IBM 知识产权部门联系，或用书面方式将查询寄往：

Intellectual Property Licensing  
Legal and Intellectual Property Law  
IBM Japan Ltd.  
1623-14, Shimotsuruma, Yamato-shi  
Kanagawa 242-8502 Japan

本条款不适用英国或任何这样的条款与当地法律不一致的国家或地区：International Business Machines Corporation“按现状”提供本出版物，不附有任何种类的（无论是明示的还是暗含的）保证，包括但不限于暗含的有关非侵权、适销和适用于某种特定用途的保证。某些国家或地区在某些交易中不允许免除明示或暗含的保证。因此本条款可能不适用于您。

本信息中可能包含技术方面不够准确的地方或印刷错误。此处的信息将定期更改；这些更改将编入本资料的新版本中。IBM 可以随时对本出版物中描述的产品进行改进和/或更改，而不另行通知。

本信息中对非 IBM Web 站点的任何引用都只是为了方便起见才提供的，不以任何方式充当对那些 Web 站点的保证。那些 Web 站点中的资料不是 IBM 产品资料的一部分，使用那些 Web 站点带来的风险将由您自行承担。

IBM 可以按它认为适当的任何方式使用或分发您所提供的任何信息而无须对您承担任何责任。

本程序的许可证持有者如果要了解有关程序的信息以达到如下目的：(i) 允许在独立创建的程序和其他程序（包括本程序）之间进行信息交换，以及 (ii) 允许对已经交换的信息进行相互使用，请与下列地址联系：

IBM Software Group  
ATTN: Licensing

200 W. Madison St.  
Chicago, IL; 60606  
U.S.A.

只要遵守适当的条件和条款，包括某些情形下的一定数量的付费，都可获得这方面的信息。

本资料中描述的许可程序及其所有可用的许可资料均由 IBM 依据 IBM 客户协议、IBM 国际软件许可协议或任何同等协议中的条款提供。

此处包含的任何性能数据都是在受控环境中测得的。因此，在其他操作环境中获得的数据可能会有明显的不同。有些测量可能是在开发级的系统上进行的，因此不保证与一般可用系统上进行的测量结果相同。此外，有些测量是通过推算而估计的，实际结果可能会有差异。本文档的用户应当验证其特定环境的适用数据。

涉及非 IBM 产品的信息可从这些产品的供应商、其出版说明或其他可公开获得的资料中获取。IBM 没有对这些产品进行测试，也无法确认其性能的精确性、兼容性或任何其他关于非 IBM 产品的声明。有关非 IBM 产品性能的问题应当向这些产品的供应商提出。

所有关于 IBM 未来方向或意向的声明都可随时更改或收回，而不另行通知，它们仅仅表示了目标和意愿而已。

本信息包含在日常业务操作中使用的数据和报告的示例。为了尽可能完整地说明这些示例，示例中可能会包括个人、公司、品牌和产品的名称。所有这些名字都是虚构的，若现实生活中实际业务企业使用的名字和地址与此相似，纯属巧合。

如果您正在查看本信息的软拷贝，图片和彩色图例可能无法显示。

---

## 商标

IBM、IBM 徽标和 `ibm.com` 是 International Business Machines Corp. 在全球许多行政管辖地区的商标或注册商标。其他产品和服务名称可能是 IBM 或其他公司的商标。Web 页面“Copyright and trademark information” ([www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml)) 提供了 IBM 商标的最新列表。

Intel、Intel 徽标、Intel Inside、Intel Inside 徽标、Intel Centrino、Intel Centrino 徽标、Celeron、Intel Xeon、Intel SpeedStep、Itanium 和 Pentium 是 Intel Corporation 或其子公司在美国和其他国家或地区的商标或注册商标。

Linux 是 Linus Torvalds 在美国和/或其他国家或地区的注册商标。

Microsoft、Windows、Windows NT 和 Windows 徽标是 Microsoft Corporation 在美国和/或其他国家或地区的商标。

UNIX 是 The Open Group 在美国和其他国家或地区的注册商标。

Java 和所有基于 Java 的商标和徽标是 Oracle 和/或其子公司的商标或注册商标。

其他产品和服务名称可能是 IBM 或其他公司的商标。

---

## 词汇表

---

### A

精度阈值：这是传播过程的一个中止条件，对应于节点能量中能保证传播继续的最小变化。

权威领导 (*authority leader*)：这是组中拥有最高权威分数的个人。

权威分数 (*authority score*)：此度量表示其他组成员与个人的联系程度。权威分数将个人重要性与其终止于该个人的关系数相关联。

---

### C

核心组 (*core group*)：这是在忽略网络中的较弱关系并应用组大小限制后剩余的一组节点。

范围阈值 (*coverage threshold*)：这是用于组标识的最强关系的分数。

---

### D

*degree*：用于衡量网络中的节点活动，定义为涉及节点的关系总数。

密度 (*density*)：用于衡量网络凝聚力，定义为观察到的关系数除以可能的关系数。

二分关系 (*dichotomous relationship*)：这是两个节点之间的一种联系，只能采用两个值中的一个，表示存在或缺少这种联系。

传播能量：这是源自网络中的特定节点的影响，它会扩散到整个网络，并随着和初始节点距离的增大而减少大小。

方向关系 (*directional relationship*)：这是两个节点之间的一种联系，从一个节点（源）开始，在另一个节点（目标）结束。

传播领导 (*dissemination leader*)：这是组中拥有最高传播分数的个人。

传播分数 (*dissemination score*)：此度量表示个人与其他组成员的联系程度。传播分数将个人重要性与其源自该个人的关系数相关联。

---

### R

入度：用于衡量由定向关系组成的网络中节点的威望，定义为节点作为目标的关系总数。

---

### N

非方向关系 (*nondirectional relationship*)：这是两个节点之间的一种关联。

---

### O

出度：用于衡量由定向关系组成的网络中节点的中心性，定义为节点作为源的关系总数。

---

## S

社会关系图 (*sociogram*) : 这是社交网络的一种视觉表示, 其中连接点的线表示彼此之间有关系的个人。

传播系数 : 这是节点在传播过程步骤中传送到其近邻的能量分数。

---

## V

重视关系 (*valued relationship*) : 这是两个节点之间的一种联系, 带有一个关联的权重, 表示关系的强度。

# 索引

## [ C ]

测量级别 4  
程度 3  
重视关系 1  
出度 3, 12, 17  
传播分数 10, 12  
传播分析 4, 15  
    传播系数 15, 16  
    精度阈值 15, 16  
    选项 16  
    要求 16  
    种子 16  
传播领导 10, 12  
传播能量 17  
传播系数 15, 16  
传播种子列表 16

## [ E ]

二分关系 1

## [ F ]

范围阈值 10, 11  
方向关系 1  
非方向关系 1  
分析数据 5  
分析选项卡 4  
    传播分析 17  
    组分析 12

## [ G ]

构建选项选项卡 4  
    传播分析 16  
    组分析 11  
关键业绩指标 12, 17  
关系  
    重视 1  
    方向 1  
过滤输出 4  
过滤选项卡 4

## [ H ]

核心节点 12  
核心密度 12  
核心组 10

## [ J ]

脚本编写  
    属性 6  
精度阈值 15, 16

## [ L ]

类型选项卡 4

## [ M ]

密度 3, 12  
目标字段 6

## [ P ]

平均出度 12, 17  
平均核心成员分数 12  
平均核心组密度 12  
平均入度 12, 17  
平均影响力 17  
平均组大小 12  
平均组密度 12

## [ Q ]

清除结果 5  
权重字段 6, 11, 16  
权威分数 10, 12  
权威领导 10, 12

## [ R ]

入度 3, 12, 17

## [ S ]

数据选项卡 4  
    传播分析 16  
    组分析 11  
属性  
    脚本编写 6

## [ T ]

通话详细记录 6

## [ Y ]

预览输出 5  
源字段 6

## [ Z ]

执行流  
    使用 IBM SPSS Modeler Server Social  
        Network Analysis 4  
种子  
    用于传播分析 16  
注解选项卡 4  
组大小 12  
    限制 10, 11  
组分析 4  
    范围阈值 10, 11  
    要求 10  
    组大小 10, 11

## G

GAG\_GroupNumber 12







Printed in China