

IBM SPSS Modeler Social  
Network Analysis 15 用户指南



注意：使用本信息及其支持的产品之前，请阅读声明第 19 页码下的一般信息。

此版本适用于 IBM SPSS Modeler Social Network Analysis 15 及所有后续发布和修订，除非在新版本中另有说明。

Adobe 产品屏幕截图重印已获得 Adobe Systems Incorporated 的许可。

Microsoft 产品屏幕截图重印已获得 Microsoft Corporation 的许可。

受许可保护材料 – IBM 所有

**Copyright IBM Corporation 2012.**

美国政府用户受限权利 – 使用、复制或披露受与 IBM Corp. 签订的 GSA ADP Schedule Contract 的限制。

---

# 前言

IBM® SPSS® Modeler Social Network Analysis 将有关人们之间关系的信息处理成字段，这些字段介绍社交网络中的个人角色，允许在预测模型中包括社交信息。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。如果将节点输出与代表个人衡量的字段相结合，会产生更完整的个人简档。

## 关于 IBM Business Analytics

IBM Business Analytics 软件提供决策者可信任的完整、一致和准确的信息，以提高企业业绩。[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。配合丰富的行业解决方案、久经考验的实践和专业服务，各种规模的组织均可驱动最高生产力、安心地实现决策自动化并提供更出色的业绩。

作为此产品的组合部分，IBM SPSS Predictive Analytics 软件帮助组织预测未来的事件，并在该洞察力的基础上提前行动，驱动更好的企业成果。全世界的商业、政府和学术客户依靠 IBM SPSS 技术作为一项竞争优势，吸引、挽留和增长客户，同时减少欺诈和降低风险。通过在日常活动中融入 IBM SPSS 软件，成为预测企业的组织 - 可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

技术支持可用于维护客户。客户可就 IBM Corp. 产品使用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>。在请求帮助时，请做好准备，以便识别您自己、您的组织以及您的支持协议。

---

# 内容

<b>1</b>	<b>IBM SPSS Modeler Social Network Analysis</b>	<b>1</b>
	关于 IBM SPSS Modeler Social Network Analysis . . . . .	1
	关于社交网络分析. . . . .	2
	显示网络. . . . .	2
	介绍网络. . . . .	3
	IBM SPSS Modeler Social Network Analysis 节点. . . . .	5
	节点选项卡. . . . .	5
	预览输出. . . . .	6
	分析数据. . . . .	6
	应用程序 . . . . .	6
	数据结构 . . . . .	7
	脚本属性 . . . . .	7
<b>2</b>	<b>组分析</b>	<b>9</b>
	组分析概述 . . . . .	9
	确定社交相似性 . . . . .	9
	划分为组. . . . .	10
	介绍组和组成员 . . . . .	10
	组分析节点 . . . . .	11
	为组分析指定数据 . . . . .	11
	为组分析设置建构选项 . . . . .	12
	查看组分析统计量 . . . . .	12
	组分析的输出 . . . . .	13
<b>3</b>	<b>传播分析</b>	<b>15</b>
	传播分析概述 . . . . .	15
	传播分析示例 . . . . .	15
	传播分析节点 . . . . .	16
	为传播分析指定数据 . . . . .	16
	为传播分析设置建构选项. . . . .	17
	查看传播分析统计量 . . . . .	17
	传播分析的输出 . . . . .	18

## 附录

A 声明	19
词汇表	21
参考书目	22
索引	23



# IBM SPSS Modeler Social Network Analysis

## 关于 IBM SPSS Modeler Social Network Analysis

许多对行为建模的方法都侧重个人。它们使用有关个人的各种数据生成一个模型，并使用行为的关键指标进行预测。如果任何个人拥有的关键指标值与行为的发生相关联，则可以特殊关注该个人以防止该行为。

考虑为流失建模的方法，流失表示客户终止与公司的关系。挽留客户的成本比取代他们的成本低得多，因此要实现识别存在流失风险的客户的能力。分析师通常使用一系列关键绩效指标描述客户，包括人口统计信息和每位个人客户的最近通话模式。基于这些字段的预测模型使用符合过去已流失的客户通话模式的客户通话模式中的变化，以识别增加流失风险的人员。识别为存在风险的客户则获得额外客户服务或服务选项以挽留他们。

这些方法会忽略可能严重影响客户行为的社交信息。有关公司和其他人当前行动的信息通过关系流动影响人们。因此，和他人的关系允许那些人影响一个人的决定和行动。只包括个人衡量的分析则忽略了具有预测能力的重要因素。

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。这些导出的关键绩效指标衡量个人的社交特征。将这些社交属性与基于个人的衡量结合起来，提供对个人的更好概览，因此可提高您模型的预测精度。

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，可使流中包含社交分析技巧。
- 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，您可发现网络中流失风险提高的个人组。通过在您的模型中结合这些影响的关键绩效指标，可提高它们的总体绩效。

注意：IBM SPSS Modeler Social Network Analysis 节点的处理必须通过 IBM SPSS Modeler Server Social Network Analysis 完成。当在 SPSS Modeler 中执行包含 IBM SPSS Modeler Social Network Analysis 节点的流时，您必须连接到也包含 IBM SPSS Modeler Server Social Network Analysis 的 SPSS Modeler Server 实例。

## 关于社交网络分析

社交网络由一组个人和他们之间的关系组成。社交网络分析检查这些关系，以说明作为社交结构组成部分的个人和组。个人彼此交互，而这些交互模式提供对参与的个人深入了解。关系允许信息在网络中流动，让一个人影响另一个。关系信息的重要性将社交网络分析和其他方法区别开来。和分别关注每个人不同，研究单位是包括两个人和他们关系的成对单位。

网络中的关系可分为方向或非方向。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，或目的地。例如，打电话是一种方向关系，一个人呼叫另一个人。相反，源和目的地的角色不能定义为**非方向关系**。在这种情况下，双方都均等参与关系。彼此交谈是非方向关系的一种示例。

区别关系的另一个属性是识别关系为二分或重视。**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。对于网络中的每个成对单位，关系为存在或缺乏。另一方面，**重视关系**包括指示关系强度的一个权重。该权重允许关系互相比较。

“**关系类型**”表格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 是目的地。在非方向关系中，不会指示关系的发起人。重视关系使用会话长度作为关系权重，同时可能发生或不发生二分关系。

表 1-1  
关系类型

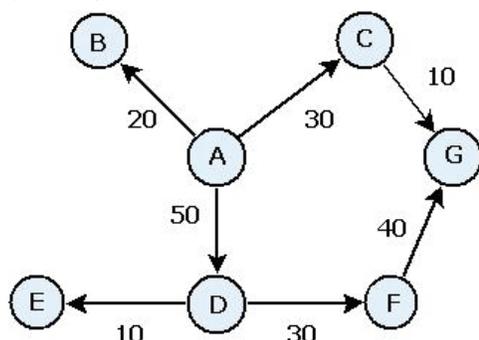
方向	尺度	示例
非方向	二分	Joe 和 Mary 彼此交谈
非方向	重视	Joe 和 Mary 彼此交谈 20 分钟
方向	二分	Joe 给 Mary 打电话
方向	重视	Joe 打电话给 Mary 交谈 20 分钟

有关社交网络分析的详细信息，请查阅此区域中的综合书籍之一(Wasserman 和 Faust, 1994)。

## 显示网络

社交网络通常使用**社会关系图**介绍(Moreno, 1934)。在这种视觉显示类型中，个人会对应至一个空间中的点，或节点。连接点的线，或边缘，代表个人之间的关系。如果是方向关系，边缘包括指示方向的一个箭头。如果关系有权重，边缘的标签则会标注值。下图显示七个人组成的一个网络。

图片 1-1  
社交网络示例



假设网络代表个人拨打的电话，关系权重表示通话长度。在这种情况下，人员 A 给三个人打电话，和人员 D 的通话时间最长。

此网络比实际遇到的那些小得多。不过，简单社会关系图介绍的概念可推广到任何规模和复杂程度的网络。

## 介绍网络

有关网络、组和个人的信息需要提取到描述特征中，以允许交叉比较并包括在预测模型中。需要将网络提取为可供分析的一组有限关键绩效指标。例如，可能希望彼此对比网络或网络中的节点组。也可能希望彼此对比网络中的个人，或识别最重要的人。

通常用于说明社交网络的两种衡量方法是**密度**和**程度**。两种统计数据都反应连通性，但前者关注整个网络或网络子组，而后者描述网络中个人的特征。

## 网络密度

对于网络中的任何节点组，都可能存在有限的关系数。每个节点都可作为与其他节点关系的源或目标。想象三个节点 A、B 和 C 组成的一个网络。下表列出节点之间所有可能的定向关系。

表 1-2  
三个节点的可能定向关系

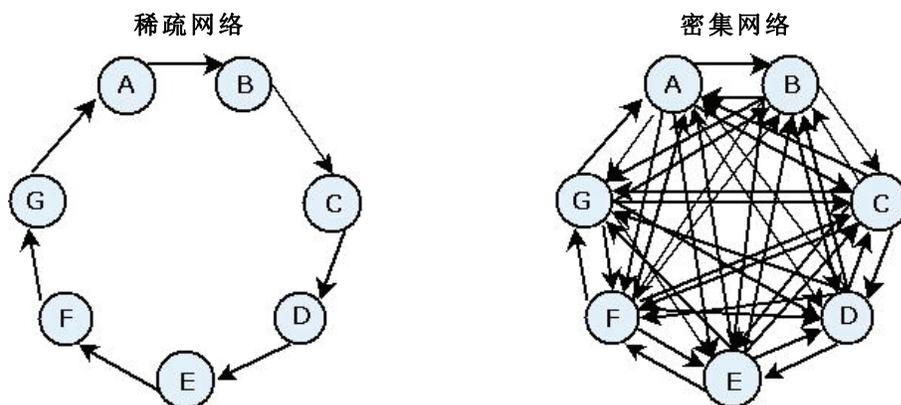
源	目标
A	B
A	C
B	A
B	C
C	A
C	B

每个节点都是和其他两个节点之间关系的源。但在实践中，所有可能的关系或许并不实际存在。有些节点可能和其他节点没有任何方向关系。此外，有些定向关系可能不往复。

**密度**统计量代表网络中实际存在的可能关系比例。值的范围从 0 到 1，下限对应于没有关系的网络，上限代表所有可能关系的网络。值越接近 1，网络的密度就越大，网络中的节点就越紧密。

密集网络中的信息比稀疏网络中的信息更容易流动。“稀疏和密集网络”表格显示包括七个节点的两个网络。稀疏网络只包括节点之间可能的 42 种关系中的七个，得出的密度为 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 1-3  
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，信息可直接从节点 A 流动到节点 G。

## 节点程度

网络中的重要个人通常是参与最多关系的人。这些个人从各种源取得信息，并向大量其他人传播该信息。相反，参与较少关系的个人无法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。程度值更高的人比程度更低的人活跃。程度忽略关系的方向，并提供对该节点活动的总体衡量。

对于定向关系，统计关系数时可侧重于节点是源还是目标。节点的**入度**是特定节点在其中作为目标的关系数。相反，**出度**是节点在其中作为源的关系数。下表为“[社交网络示例](#)”图中的每个节点列出程度、入度和出度值。

表 1-4  
程度、入度和出度值示例

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

入度通常被当做对威望的衡量。更高的入度值对应于在该节点结束的更多关系。换言之，那些个人得到许多其他人的联系。许多其他节点发起和节点的关系。相反，出度被当做对向心性的衡量。更高的值对应于从该节点发起的更多关系。那些个人联系更多其他个人。

对于示例网络中的节点，程度值表示节点 A 和 D 最活跃，而节点 B 和 E 最不活跃。入度值显示节点 G 有最多威望。根据出度值，节点 A 最向心。

## 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” 表格介绍的这些节点存储在源调色板中。

表 1-5  
IBM SPSS Modeler Social Network Analysis 节点

节点	图标	描述
组分析		组分析节点从一个固定字段文本文件导入通话详细记录数据、识别记录定义的网络中的节点组，并且为网络中的组和个人生成关键绩效指标。有关详细信息，请参阅第 9 页码第 2 章中的 <a href="#">组分析概述</a> 。
传播分析		传播分析节点从一个固定字段文本文件导入通话详细记录数据、在记录定义的网络中传播影响，并且生成关键绩效指标以汇总对个人节点影响的结果。有关详细信息，请参阅第 15 页码第 3 章中的 <a href="#">传播分析概述</a> 。

IBM SPSS Modeler Social Network Analysis 节点的处理必须通过 IBM® SPSS® Modeler Server Social Network Analysis 完成。当在 SPSS Modeler 中执行包含这些节点的流时，您必须连接到也包含 IBM SPSS Modeler Server Social Network Analysis 的 IBM® SPSS® Modeler Server 实例。

### 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。
- **建构选项选项卡。**用于为分析定义设置。
- **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。
- **注解选项卡。**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 SPSS Modeler 文档。

## 预览输出

通常，您使用节点输出作为一个预测模型的输入。如果希望查看输出，可将一个表格节点添加到流。但数据量通常非常大，让表格的生成非常耗时。此外，表格中的行数限制其有效性。

作为替代，您可生成一个预览表，显示将创建的输出示例。预览显示为有限行数生成的字段。行数在流属性中定义。要预览节点输出，请执行以下步骤：

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

一个预览窗口打开，显示结果。此外，预览使用结果的摘要填充节点“分析”选项卡的输出。

## 分析数据

确定产品优化结果通常是一个迭代过程的分析设置。您定义设置、运行分析并审核结果。如果结果不像预期那样有用，则修改设置并重新运行分析。

要为节点分析输入数据，请执行以下步骤：

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

如果建构选项指示应显示汇总统计量，则“分析”选项卡显示结果。

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。**可影响流失率的组特征。通过关注流失风险提高的组中个人，您可加以预防。此外，您可识别存在流失风险的个人，原因是来自那些个人的信息流已经流失。

- **充分利用组领导。**组领导显著影响其他组成员。如果可预防组领导流失，组成员的流失率可能降低。相反，尝试从竞争对手获得要流失的组领导会提高与该竞争对手关联的组成员流失率，同时降低与您的公司关联的组成员流失率。
- **营销。**可通过组领导发起新的商品和服务供应。领导的影响可能让其他组成员更可能购买产品。您可使用传播分析识别受组领导影响最大的个人，并针对他们开展营销。

## 数据结构

关于关注的网络中个人的信息，可能在您的整个企业的各种文件、数据库和系统中传播。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，并格式化它们以输入节点。

分析节点需要存储在一个单独的固定宽度文本文件中的通话详细记录。文件的每行对应一种关系，数据组织在以下列中：

- 发起关系的个人标识符。
- 作为关系目标的个人标识符。
- 关系的一个可选权重。

所有数据必须是数字，并且个人标识符限制为整数。可选择包括字段名，作为文件中数据的第一行。“[通话数据记录示例](#)”表格介绍这种数据结构。

表 1-6  
通话数据记录示例

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

相对于网络中的其他关系，权重值对应于您要用于代表关系重要性的任何衡量。对于通话数据，常用权重包括通话持续时间或通话频率。

如果希望分析侧重通话历史记录的子集，您必须在创建输入文件时使用该子集。例如，通过在输入文本文件中仅包括该数据，可将分析限制为过去几个月或最近对个人的通话。

## 脚本属性

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化那些手动执行将造成大量重复操作且高耗时的任务。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

下表列出传播节点的脚本属性。

属性名称	数据类型	属性说明
input_data_file_name	string	
calling_field	field	
called_field	field	
frequency_weight_field	field	
read_field_names	boolean	
diffusion_list_file_name	string	
spreading_factor	double	默认值 = 50。 最小值 = 1 最大值 = 99。
max_number_iterations	integer	默认值 = 100。 最小值 = 1。
accuracy_threshold	double	默认值 = 0.01。 最小值 = 0.001。
calculate_statistics	boolean	

### 组分析节点属性

下表列出组分析节点的脚本属性。

属性名称	数据类型	属性说明
input_data_file_name	string	
calling_field	field	
called_field	field	
frequency_weight_field	field	
read_field_names	boolean	
coverage_threshold	double	默认值 = 10。 最小值 = 1 最大值 = 99。
min_group_size	integer	默认值 = 2。 最小值 = 2。
max_group_size	integer	默认值 = 100。 最小值 = 2。
calculate_statistics	boolean	

# 组分析

## 组分析概述

组分析使用网络中的个人交互模式，识别类似个人的组。这些组的特征影响个人组成员的行为。例如，即便实际上没有组成员流失，拥有许多成员间关系和强大领导的小组仍增加了流失风险。(Richter, Yom-Tov, 和 Slonim, 2010)同时结合组和个人衡量的预测模型的执行效果将比仅包括后者的模型更好。

组分析包括以下常规步骤：

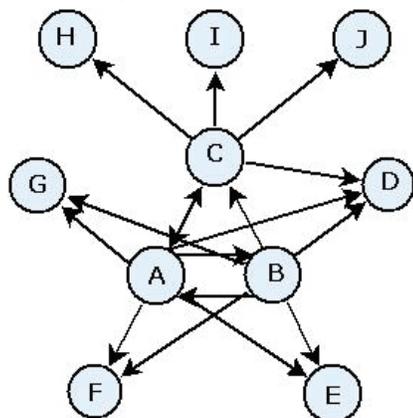
1. 确定反映社交邻近情况的关系强度。有关详细信息，请参阅第 9 页码中的[确定社交相似性](#)。
2. 根据关系强度将网络划分成组，同时遵守大小限制。有关详细信息，请参阅第 10 页码中的[划分为组](#)。
3. 同时简介组和个人，包括识别组领导。有关详细信息，请参阅第 10 页码中的[介绍组和组成员](#)。

## 确定社交相似性

组成员彼此之间应该比不在组中的个人更加相似。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，都有一组节点是该节点定向关系的目標。对于电信数据，这个组对应于特定个人所联系的所有人员。如果两位个人联系同一组人员，则将那些个人视为彼此相似。两位个人所重叠的关系目标组越多，他们就越相似。

考虑“十节点网络示例”图中显示的网络。

图片 2-1  
十节点网络示例



“目标节点”表格显示在节点 A、B 和 C 发起的关系目标节点。

表 2-1  
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
C	A, D, H, I, J

节点 A 和 B 有五个共用目标节点。另一方面，节点 A 和 C 只有一个。因此节点 A 和 B 比节点 A 和 C 的社交相似性更大。

诸如通话持续时间或频率等关系权重不会采集网络中节点的相似性。为了包括相似性的概念，组分析使用共同信息(覆盖和 Thomas, 2006)作为关系权重。此统计量反映两个节点连接到相同节点的似然度。如果网络中的关系有预定义的权重，例如电信数据的通话频率，则共同信息会相应结合那些权重。

## 划分为组

组应包括和其他组成员有高关系权重的个人，其中的权重衡量关系中节点的相似性。因此，通过忽略网络中较弱的关系开始组标识。**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，范围阈值 0.4 则会将最强关系的 40% 用于组标识，而忽略剩余 60% 的关系。

剩余关系可能产生一些非常小或非常大的组，预测实用性有限。为防止分析中包括类似组，您可指定最小和最大组大小。分析过程会完全忽略小于最小限制的组。但大于最大限制的组则会分为可接受大小范围内更小的组。丢弃弱关系和实施大小限制后剩余的组则为**核心组**。

从原始网络删除关系，可能导致某些个人不存在于任何核心组中。但那些个人可能与肯定包括在组中的组成员存在关联。假设没有违反组大小限制，如果与组的核心成员有许多相对较强的关系，则会将个人添加到组。因此，最终的组包括一组核心成员，外加由于与核心关联而添加的成员。

## 介绍组和组成员

除了密度、入度和出度，其他统计量也介绍组动力。特别是权威和传播分数，提供组中个人的社交状态的衡量。尝试预测组及其成员的行为时，组中每位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。如果许多个人联系一位特定人员，可能是咨询信息或意见，则该人员拥有权威角色。组中节点的权威分数，对应于通过组网络随机重启的稳态概率，范围从 0 到 1。权威分数越接近 1，该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**。将组中的最大分数除以最小分数，以提供权威领导总体强度的衡量。

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。如果一位特定人员联系组中许多其他人员，该人员可显著影响整个组的意见。组中节点的传播分数，对应于反向通过组网络随机重启的稳态概率，范围从 0 到 1。传播分数越接近 1，节点则连

接越多其他组成员。组中拥有最高传播分数的节点则为该组的**传播领导**。组中的最大分数除以最小分数，以提供传播领导总体强度的衡量。

## 组分析节点

“源”调色板提供的组分析节点可识别网络中彼此社交相似的个人组，并确定组中个人的相关社交状态。相对较小、在成员之间有许多联系并包括强大领导的组，会由于组动力而提高流失风险。当您结合个人流失分数时，可使用结果确定组流失分数，以提高单单一基于个人分数来预测模型流失的能力。

假设数据量通常包括网络和组定义下社交交互的性质，则组分析是耗时的过程。您通常会将组分析结果保存到数据库或文件，作为预测模型的输入。组结果的刷新频率相对不太频繁，例如每月刷新，而模型可能更频繁的刷新。

**要求。**节点需要一个使用三个字段定义社交网络的固定宽度文本文件。一个字段识别每个定向关系的源，一个字段定义每个定向关系的目的地，第三个字段定义每种关系的可选强度。网络中的全部关系必须都是方向关系。

## 为组分析指定数据

组分析源节点窗口的“数据”选项卡允许您指定包含网络节点关系的输入文件。

**文件。**指定包含调用明细记录的文件或文件夹名称。可输入名称或单击省略号按钮 (...) 从文件系统中选择名称。一旦您选择了名称，将显示路径；同时显示其内容，以分隔符间隔。如果您指定了一个文件夹，将会连接文件夹中所包含所有文件的调用明细记录以供分析；文件夹中的所有文件应具有相同结构。

**读取文件中的字段名。**在默认情况下选择此选项，它将数据文件中的第一行作为列的名称。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

网络定义设置可定义字段的角色。

**字段。**使用箭头按钮，将此列表的项目手动分配到各种角色字段。图标表示每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。

**目的地。**选择一个字段作为定向关系的目标。

**权重。**也可选择一个符合网络中关系发生次数的字段。确定关系强度时，分析相应衡量记录的权重。

## 为组分析设置建构选项

组分析源节点窗口的“建构选项”选项卡允许您定义选项，用于识别网络中的组。

### 组分析设置

组分析设置影响所形成组的大小和相对强度。

**范围阈值。**定义在分析中要使用的最强网络关系分数。例如，范围阈值 0.2 表示只使用其权重在所有权重中排名前 20% 的关系。这个参数的范围从 0 到 1，允许分析侧重于组中应发生的最强关系。较高阈值产生的组在成员间拥有较弱的关系。有关详细信息，请参阅第 10 页码中的[划分为组](#)。

**最小组大小。**指定组大小的下限。分析过程不会返回小于此值的组。

**最大组大小。**指定组大小的上限。大于此值的组将分成更小的组。

**计算并显示汇总统计量。**如选中，除了导出关键绩效指标输出，节点还为分析计算并显示汇总统计量。计算这些统计量会对极大网络的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摘要概述。“组分析汇总统计量”表格列出组的可用汇总统计量。

表 2-2  
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	组中节点之间的直接联系平均分数。有关详细信息，请参阅第 3 页码第 1 章中的 <a href="#">网络密度</a> 。
核心成员的平均分数	组中作为该组核心节点的平均分数。有关详细信息，请参阅第 10 页码中的 <a href="#">划分为组</a> 。
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	导入链接的平均数。有关详细信息，请参阅第 4 页码第 1 章中的 <a href="#">节点程度</a> 。
平均出度	导出链接的平均数。有关详细信息，请参阅第 4 页码第 1 章中的 <a href="#">节点程度</a> 。

从“汇总统计量”表选择特定统计量，以查看值的分配、标准差和统计量的偏度。

审核这些统计量有助于定义组分析设置。理想情况是，结果应显示具有高密度值的相对小的组。例如，如果有一些非常大的组，考虑减小最大组的大小，将它们分为较小的组。如果组密度值趋向于较小，请考虑减小范围阈值以关注更强的关系。

## 组分析的输出

组分析节点生成介绍组和组中个人的各种字段。可使用这些关键绩效指标增大现有模型和数据，改善您的模型生成的预测。例如，可更新个人流失倾向值以包括组的影响力。

“组的关键绩效指标”表格为分析中识别的组列出关键绩效指标字段。

表 2-3  
组的关键绩效指标

字段	描述
GAG_GroupNumber	组的唯一标识符
GAG_Size	组中的个人数
GAG_Density	组中个人之间的直接联系分数。有关详细信息，请参阅第 3 页码第 1 章中的 <a href="#">网络密度</a> 。
GAG_KernelDensity	组中核心个人之间的直接联系分数
GAG_CoreNodesFraction	组中作为该组核心个人的个人分数。有关详细信息，请参阅第 10 页码中的 <a href="#">划分为组</a> 。
GAG_MaxRankType1	任何组成员的最高权威分数。有关详细信息，请参阅第 10 页码中的 <a href="#">介绍组和组成员</a> 。
GAG_MinRankType1	任何组成员的最小权威分数
GAG_MaxMinRankRatioType1	最大和最小权威分数的比率。此值反映组领导的权威强度。
GAG_MaxRankType2	任何组成员的最高传播分数。有关详细信息，请参阅第 10 页码中的 <a href="#">介绍组和组成员</a> 。
GAG_MinRankType2	任何组成员的最小传播分数
GAG_MaxMinRankRatioType2	最大和最小传播分数的比率。此值反映组领导的传播强度。

“个人的关键绩效指标”表格为网络中的个人列出关键绩效指标字段。

表 2-4  
个人的关键绩效指标

字段	描述
GAI_NodeNumber	个人的唯一标识符
GAI_CoreNode	个人是否作为一个组的核心个人的指标。有关详细信息，请参阅第 10 页码中的 <a href="#">划分为组</a> 。
GAI_RankType1	个人的权威分数。有关详细信息，请参阅第 10 页码中的 <a href="#">介绍组和组成员</a> 。
GAI_RankOrderType1	基于权威分数的组中排名顺序
GAI_RankType2	个人的传播分数。有关详细信息，请参阅第 10 页码中的 <a href="#">介绍组和组成员</a> 。
GAI_RankOrderType2	基于传播分数的组中排名顺序
GAI_InDegree	个人在其中作为关系目标的关系数。有关详细信息，请参阅第 4 页码第 1 章中的 <a href="#">节点程度</a> 。
GAI_OutDegree	个人在其中作为关系源的关系数。有关详细信息，请参阅第 4 页码第 1 章中的 <a href="#">节点程度</a> 。
GAI_GroupLeaderType1	不管节点是否为权威领导，其领导分数均来自导入链接。有关详细信息，请参阅第 10 页码中的 <a href="#">介绍组和组成员</a> 。
GAI_GroupLeaderConfidenceType1	节点作为权威领导的置信度。

字段	描述
GAI_GroupLeaderType2	不管节点是否为传播领导，其领导分数均来自导出链接。有关详细信息，请参阅第 10 页码中的 <a href="#">介绍组和组成员</a> 。
GAI_GroupLeaderConfidenceType2	节点作为传播领导的置信度。

# 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展到它们的直接近邻，随着它在节点之间的移动而减少大小(Dasgupta, Singh, Viswanathan, Chakraborty, Mukherjea, Nanavati, 和 Joshi, 2008)。接收到能量时，则会激活一个节点，并将部分能量传送到作为该节点定向关系目标的任何近邻。

**传播系数**定义一个激活的节点可传送的能量比例，以及该节点保持的剩余量。接收此能量的任何节点会自己将相同比例传送给它们的近邻，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。传播系数小产生快速衰退的传播过程，剩余的已传送能量则相对较接近初始节点。

一个激活节点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。每个节点接收的量取决于和激活节点的关系强度。发送到特定节点的分数，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，拥有更高相关权重关系的近邻可接收的能量多于拥有较低相关权重的近邻。

如果发生下面的一种情况，扩散过程则停止：

- 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**，其为使过程能继续所传送的能量限制
- 迭代数达到指定限制

传播完成时，拥有最大传播能量的节点则表示对发起该过程的影响最敏感。例如，如果该过程开始于流失的节点，则拥有最高能量的节点其流失的风险就越大。可以为那些节点提供特殊关照，以防止它们流失。

## 传播分析示例

考虑第 1 章第 2 页码中显示的网络。该网络包括七个节点，拥有彼此强度各不相同的定向关系。

“传播过程示例”表格使用传播系数 0.80 介绍网络中的传播过程。最初，节点 A 包含所有能量，被任意分配一个值 1.00。第 1 步中，此节点激活，将其能量的 80% 传播到作为关系目标的三个近邻，而自己保留 20%。关系的总权重为 100。带有节点 D 的关系占此总数的一半，因此这个节点接收一半传播能量，或 0.40。带节点 B 的关系占总权重的 20%，产生 0.16 的传播能量值。节点 C 接收剩余能量，0.24，是从节点 A 发起量的 30%。

表 3-1  
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

在第 1 步中接收能量后，节点 B、C 和 D 在第 2 步中激活，向任何目标近邻传播能量。节点 B 没有目标节点，因此不传播能量。另一方面，节点 C 有一个目标并将其能量的 80% 传播到该节点 G。最终，节点 D 将其能量的 80% 传播到节点 E 和 F，而 F 由于和 D 更强的关系，因此接受相当于 E 的三倍的能量。

第 3 步中，节点 E、F 和 G 激活。节点 E 和 G 没有目标节点，因此不传播任何能量。但节点 F 将其能量的 80% 传播到其单独的目标节点 G。此节点现在拥有的总能量值为 0.38，包括从节点 C 和 F 传播的能量。

此时，节点 G 激活但没有用于传播的目标节点，因此过程终止。在节点 A 发起的能量通过网络传播，节点 G 接收最高的量。如果网络中介绍的能量代表流失，则节点 G 受节点 A 流失的影响最大。

## 传播分析节点

“源”调色板提供的传播分析节点使用网络关系识别受该影响最大的个人，并从社交网络中指定的个人组传播影响。例如，如果影响是流失，节点则会透过因网络中其他特定个人流失，来识别最可能流失的个人。可使用节点输出增大现有模型和数据，改善那些模型生成的预测。例如，可更新个人流失倾向值以包括传播影响力。

**要求。**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。一个字段识别每个定向关系的源，一个字段定义每个定向关系的目的地，第三个字段定义每种关系的可选强度。网络中的全部关系必须都是方向关系。第二个文件包含标识符的列表，影响应从其开始。

## 为传播分析指定数据

传播分析源节点窗口的“数据”选项卡允许您指定包含网络节点关系的输入文件。

**文件。**指定包含调用明细记录的文件或文件夹名称。可输入名称或单击省略号按钮 (...) 从文件系统中选择名称。一旦您选择了名称，将显示路径；同时显示其内容，以分隔符间隔。如果您指定了一个文件夹，将会连接文件夹中所包含所有文件的调用明细记录以供分析；文件夹中的所有文件应具有相同结构。

**读取文件中的字段名。**在默认情况下选择此选项，它将数据文件中的第一行作为列的名称。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

网络定义设置可定义字段的角色。

**字段。** 使用箭头按钮，将此列表的项目手动分配到各种角色字段。图标表示每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。

**目的地。** 选择一个字段作为定向关系的目标。

**权重。** 选择一个代表关系权重的字段。对于通话详细记录，权重可对应于反映关系强度的任何通话特征，例如通话持续时间或通话频率。

## 为传播分析设置建构选项

传播分析源节点窗口的“建构选项”选项卡允许您为传播过程定义选项。

### 传播种子列表

传播种子列表能识别应开始传播能量的网络节点。例如在流失分析中，此列表可识别已流失的网络节点。此列表也可包含拥有最高流失风险的节点，以在那些节点实际流失时探索网络上的影响。

**文件。** 指定包含节点标识符的文本文件的名称，每个标识符都对应于文件中的一行。可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性，以及过程何时停止。

**传播系数。** 定义一个节点在传播过程的每一步所传送的能量分数。更小的值将能量限制为距离传播能量的节点最近的节点。相反，更大的值允许能量影响更远的节点。

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

**精度阈值。** 指定节点能量中的最小变化，该变化能保证继续传播过程。如果每个节点的能量变化小于此值，则传播过程终止。

**计算并显示汇总统计量。** 如选中，除了导出关键绩效指标输出，节点还为分析计算并显示汇总统计量。计算这些统计量会对极大网络的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述。“传播分析汇总统计量”表格列出可用汇总统计量。

表 3-2  
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	与个人相关的平均传播能量。
平均入度	个人在其中作为关系目标的平均关系数。有关详细信息，请参阅第 4 页码第 1 章中的 <a href="#">节点程度</a> 。
平均出度	个人在其中作为关系源的平均关系数。有关详细信息，请参阅第 4 页码第 1 章中的 <a href="#">节点程度</a> 。

从“汇总统计量”表选择特定统计量，以查看值的分配、标准差和统计量的偏度。

## 传播分析的输出

传播分析节点生成介绍网络中个人的各种字段。可使用这些关键绩效指标增大现有模型和数据，改善您的模型生成的预测。例如，可更新个人流失倾向值以包括传播影响力。

“传播分析关键绩效指标”表格为分析中的个人列出关键绩效指标字段。

表 3-3  
传播分析关键绩效指标

字段	描述
DA_NodeNumber	个人的唯一标识符
DA_DiffusedEnergy	与个人相关的传播能量。对于流失分析，更高的值表示比更低的值拥有更大的倾向。
DA_InDegree	个人在其中作为关系目标的关系数。有关详细信息，请参阅第 4 页码第 1 章中的 <a href="#">节点程度</a> 。
DA_OutDegree	个人在其中作为关系源的关系数。有关详细信息，请参阅第 4 页码第 1 章中的 <a href="#">节点程度</a> 。

# 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/地区中不提供在本文档中讨论的产品、服务或功能。请咨询您当地的 IBM 代表以了解有关您所在地区当前可用产品和服务的信息。任何对 IBM 产品、程序或服务的引用，并不意味着仅可使用这些 IBM 产品、程序或服务。作为替代，可以使用任何功能相当的产品、程序或服务，前提是不侵犯任何 IBM 知识产权。不过，用户应自行负责评估和验证任何非 IBM 产品、程序或服务的工作情况。

在本文档中介绍的主题可能涉及 IBM 的专利或申请中的专利。您获得该文档并不意味着授予您任何这些专利许可。您可以将书面许可查询函件发送至：

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

**以下段落不适用于英国或此类条款与当地法律不符的其他国家/地区：** INTERNATIONAL BUSINESS MACHINES 公司“按原样”提供本出版物，不保证任何明示或暗示，包括但不限于对非侵权性、适销性或对特定用途适用性的暗示担保。有些州/省不允许针对某些交易的明示或暗示免责条款，因此本声明可能不适用于您。

本信息可能含有技术误差或排版错误。此处的信息会定期进行更改；这些更改会体现在本出版物的新版本中。IBM 可能随时对本出版物中所述的产品和/或程序进行改进和/或更改，恕不另行通知。

在本信息材料中对任何非 IBM 网站的引用仅为了方便用户，并不以任何方式表明对这些网站的认可。这些网站上的材料并非本 IBM 产品材料的一部分，您对这些网站的使用需自担风险。

IBM 可以自认为适当并且不会对您构成任何约束的任何方式使用或分发您提供的任何信息。

如果本程序的受许可方试图了解有关程序的信息以启用：(i) 在独立创建的程序和其他程序（包括本程序）之间交换信息；(ii) 相互使用交换的信息，则应联系：

IBM Software Group, Attention:Licensing, 233 S. Wacker Dr., Chicago, IL 60606,  
USA.

此类信息的提供应遵照相关条款和条件，其中包括在某些情况下支付适当费用。

在本文档中介绍的受许可保护程序，及其所有受许可保护材料由 IBM 在双方签署的“IBM 客户协议”、“IBM 国际程序许可证协议”或任何其他等同协议下提供。

本信息包含日常业务运营中使用的数据和报告的示例。为了尽可能详尽地对其进行说明，示例中包含了人员的姓名、公司、品牌和产品的名称。所有这些名称均为虚构，与真实商业企业使用的名称和地址的任何雷同纯属巧合。

## 商标

IBM、IBM 徽标、ibm.com 和 SPSS 是 IBM Corporation 的商标，在全球许多司法辖区注册。有关最新的 IBM 商标列表，请访问网页 <http://www.ibm.com/legal/copytrade.shtml>。

Linux 是 Linus Torvalds 在美国、其他国家或地区或两者的注册商标。

Microsoft、Windows、Windows NT 和 Windows 徽标是 Microsoft Corporation 在美国、其他国家或地区或两者的商标。

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

其他产品和服务名称可能是 IBM 或其他公司的商标。

---

# 词汇表

**二分关系.** 两个节点之间的一种联系，可只采用两个值中的一个，表示存在或缺乏联系。

**传播分数.** 用于表示个人连接到组的其他成员程度的衡量方式。传播分数将个人的重要性和个人发起的关系系数关联起来。

**传播系数.** 一个节点在传播处理步骤中传送到其邻居的能量分数。

**传播能量.** 从网络中特定节点发起的影响，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。

**传播领导.** 组中拥有最高传播分数的个人。

**入度.** 衡量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。

**出度.** 衡量定向关系组成的网络中节点的向心性，定义为节点作为源的关系总数。

**密度.** 网络凝聚力的一种衡量方式，定义为观测关系系数除以可能的关系系数。

**方向关系.** 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（目的地）结束。

**无方向关系.** 两个节点之间的一种联系。

**权威分数.** 用于表示组的其他成员连接到个人程度的衡量方式。权威分数将个人的重要性和结束于个人的关系系数关联起来。

**权威领导.** 组中拥有最高权威分数的个人。

**核心组.** 忽略网络中较弱关系并应用组大小限制后所剩余的一组节点。

**社会关系图.** 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。

**程度.** 衡量网络中的节点活动，定义为涉及节点的关系总数。

**精度阈值.** 传播过程的一个停止标准，符合保证继续传播的节点能量中的最微小变化。

**范围阈值.** 用于组标识的最强关系的分数。

**重视关系.** 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。

---

# 参考书目

Dasgupta, K., R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjee, A. A. Nanavati, 和 A. Joshi. 2008. Social ties and their relevance to churn in mobile telecom networks. 输入: Proceedings of the 11th international conference on extending database technology, New York, NY: ADM, 668 - 677.

Moreno, J. L. 1934. Who Shall Survive?: Foundations of Sociometry, Group Psychotherapy, and Sociodrama. Washington, D.C.: Nervous and Mental Disease Publishing Co..

Richter, Y., E. Yom-Tov, 和 N. Slonim. 2010. Predicting customer churn in mobile networks through analysis of social groups. 输入: Proceedings of the 2010 SIAM international conference on data mining, Columbus, OH: SDM 2010, 732 - 741.

Wasserman, S., 和 K. Faust. 1994. Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press.

覆盖, T. M., 和 J. A. Thomas. 2006. Elements of Information Theory, 2nd edition. New York: John Wiley and Sons, Inc.

GAG\_GroupNumber, 13

二分关系, 2

传播分数, 10, 13  
传播分析, 5, 15  
    传播系数, 15, 17  
    种子, 17  
    精度阈值, 15, 17  
    要求, 16  
    选项, 16  
传播种子列表, 17  
传播系数, 15, 17  
传播能量, 18  
传播领导, 10, 14

入度, 4, 12 - 13, 18

关系  
    方向, 2  
    重视, 2  
关键绩效指标, 13, 18

出度, 4, 12 - 13, 18

分析数据, 6  
分析选项卡, 5  
    传播分析, 17  
    组分析, 12

商标, 20

密度, 4, 12 - 13

属性  
    脚本编写, 7

平均入度, 12, 18  
平均出度, 12, 18  
平均影响, 18  
平均核心成员分数, 12  
平均核心组密度, 12  
平均组大小, 12  
平均组密度, 12

建构选项选项卡, 5  
    传播分析, 17  
    组分析, 12

执行流

    使用 IBM SPSS Modeler Server Social Network  
    Analysis, 5

数据选项卡, 5  
    传播分析, 16  
    组分析, 11

方向关系, 2

权威分数, 10, 13  
权威领导, 10, 13  
权重字段, 7, 11, 17  
核心密度, 12 - 13  
核心组, 10  
核心节点, 12 - 13

法律声明, 19  
注解选项卡, 5  
测量级别, 5  
清除结果, 6  
源字段, 7, 11, 17

目的地字段, 7, 11, 17

种子  
    用于传播分析, 17  
程度, 4

类型选项卡, 5  
精度阈值, 15, 17

组分析, 5  
    组大小, 10, 12  
    范围阈值, 10, 12  
    要求, 11  
组大小, 12 - 13  
    限制, 10, 12

脚本编写  
    属性, 7

范围阈值, 10, 12

过滤输出, 5  
过滤选项卡, 5  
通话详细记录, 7

重视关系, 2

非方向关系, 2

索引

预览输出, 6