

*Руководство пользователя по
IBM SPSS Modeler Social
Network Analysis 17.1*

IBM

Примечание

Прежде чем использовать эту информацию и продукт, описанный в ней, прочтите сведения в разделе “Уведомления” на стр. 21.

Информация о продукте

Это издание применимо к версии 17, выпуску 1, модификации 0 IBM(r) SPSS(r) и ко всем последующим выпускам и модификациям до тех пор, пока в новых изданиях не будет указано иное.

Содержание

Предисловие v

О бизнес аналитике IBM. v

Техническая поддержка v

Глава 1. IBM SPSS Modeler Social Network Analysis. 1

О программе IBM SPSS Modeler Social Network Analysis 1

Об анализе социальных сетей 1

 Вывод сетей. 2

 Описание сетей. 3

Узлы IBM SPSS Modeler Social Network Analysis 5

 Вкладки узлов 5

 Предварительный просмотр вывода. 6

 Анализ данных. 6

Программы. 6

Структура данных 7

Свойства сценариев 8

Глава 2. Групповой анализ 11

Обзор анализа групп 11

 Определение социального сходства. 11

 Разделение на группы 12

 Описание групп и их членов 12

Узел анализа групп 13

 Задание данных для анализа групп 13

 Задание опций построения для анализа групп 14

 Просмотр статистик анализа групп. 14

 Вывод для анализа групп 15

Глава 3. Анализ распространения . . . 17

Обзор анализа распространения. 17

 Пример анализа распространения 17

Узел анализа распространения 18

 Задание данных для анализа распространения . . . 18

 Задание опций построения для анализа

 распространения. 19

 Просмотр статистик анализа распространения . . . 20

 Вывод для анализа распространения 20

Уведомления. 21

Товарные знаки 22

Индекс 27

Предисловие

IBM® SPSS Modeler Social Network Analysis обрабатывает информацию о взаимосвязях между людьми в полях, описывающих роли отдельных лиц в социальной сети, что позволяет включать социальную информацию в предсказательные модели. В этом руководстве даны описания узлов IBM SPSS Modeler Social Network Analysis в среде IBM SPSS Modeler, помогающие включать узлы в потоки. При объединении вывода узлов с полями, представляющими измерения для отдельных лиц, получается более полный профиль отдельных лиц.

О бизнес аналитике IBM

Программное обеспечение IBM для бизнес аналитики предоставляет полную, последовательную и точную информацию, которая повышает эффективность ведения бизнеса. Полный набор программного обеспечения для business intelligence, прогностической аналитики, управления финансовой эффективностью и стратегией и аналитических приложений позволяет ясно видеть текущую ситуацию, а также делать прогнозы, позволяющие предпринимать практические действия. В сочетании с решениями для конкретных отраслей, проверенной практикой и услугами бизнес аналитика IBM позволяет организациям любых размеров достигать наивысшей производительности, уверенно автоматизировать процессы принятия решений и добиться лучших результатов.

Как составная часть этого набора, программное обеспечение IBM SPSS Predictive Analytics помогает организациям предсказывать будущие события и предпринимать практические действия непосредственно на основе этих предсказаний. Коммерческие, правительственные и академические организации всего мира, полагаются на технологию IBM SPSS, обеспечивающую конкурентное преимущество в привлечении, удержании и повышении отдачи от клиентов. Включая программное обеспечение IBM SPSS в свои ежедневные операции, организации могут прогнозировать будущие события, направлять и автоматизировать решения для соответствия бизнес-целям и достигать ощутимых конкурентных преимуществ. Чтобы получить дальнейшую информацию или связаться с представителем, зайдите на <http://www.ibm.com/spss>.

Техническая поддержка

Техническая поддержка предоставляется клиентам, оплачивающим обновительные взносы. Пользователи могут обращаться в службу технической поддержки, если у них возникают какие-либо проблемы с использованием или установкой программного обеспечения IBM Corp.. За технической поддержкой обращайтесь на сайт IBM Corp.: <http://www.ibm.com/support>. При обращении за поддержкой будьте готовы назвать себя и организацию, в которой вы работаете.

Глава 1. IBM SPSS Modeler Social Network Analysis

О программе IBM SPSS Modeler Social Network Analysis

Множество подходов к моделированию поведения направлено на изучение отдельных лиц. В них при помощи разнообразных данных об отдельных лицах генерируется модель, в которой для предсказания поведения используются ключевые показатели поведения. Если значения ключевых показателей какого-либо отдельного лица связаны с проявлением определенного поведения, на это лицо можно обратить особое внимание для предотвращения такого поведения.

Рассмотрим подходы к моделированию оттока, в котором заказчик прекращает свою взаимосвязь с компанией. Стоимость удержания заказчиков значительно ниже стоимости их замены, что делает крайне важной возможность определения заказчиков с повышенным риском оттока. Аналитик часто использует для описания заказчиков ряд ключевых показателей эффективности, включая демографическую информацию и шаблоны последних вызовов для каждого отдельного заказчика. В предсказательных моделях на основе этих полей используются изменения в шаблонах вызовов заказчиков, согласующихся с шаблонами вызовов заказчиков, отток которых уже произошел, для идентификации людей с повышенным риском оттока. Заказчики, идентифицируемые как лица с повышенным риском оттока, получают дополнительное обслуживание заказчиков и опции услуг с целью их удержания.

Эти методы не учитывают социальную информацию, которая может существенно влиять на поведение различных заказчиков. Информация о компании и действиях других лиц проходит по взаимосвязям для влияния на поведение людей. В результате взаимосвязь с другими людьми позволяет этим людям влиять на решения и действия сотрудника. Методы анализа, включающие только отдельные меры, опускают важные факторы с возможностями предсказания.

IBM SPSS Modeler Social Network Analysis разрешает эту проблему посредством переработки информации о взаимосвязях в дополнительные поля, которые могут быть включены в модели. Полученные ключевые показатели эффективности измеряют социальные характеристики для отдельных лиц. Объединение этих социальных свойств с мерами, основанными на отдельных лицах, обеспечивает улучшенный обзор отдельных лиц и поэтому может улучшить точность предсказания для моделей.

IBM SPSS Modeler Social Network Analysis состоит из двух первичных компонентов:

- Узлы IBM SPSS Modeler Social Network Analysis, добавляемые в среду IBM SPSS Modeler, что разрешает включение социальных аналитических методов в потоки.
- IBM SPSS Modeler Server Social Network Analysis, добавляющий обработку спецификаций узлов в сервер IBM SPSS Modeler. IBM SPSS Modeler Server Social Network Analysis эффективно перерабатывает объемы сетевых данных, куда могут входить миллионы отдельных лиц и взаимосвязей, в относительно небольшое число полей для дальнейшего анализа.

Например, IBM SPSS Modeler Social Network Analysis идентифицирует в сети отдельных лиц, больше всего затрагиваемых оттоком конкретных людей. Кроме того, возможно обнаружение в сети групп отдельных лиц с повышенным риском оттока. Введя ключевые показатели эффективности для этих эффектов в состав моделей, можно повысить их общую эффективность.

Об анализе социальных сетей

Социальная сеть состоит из круга лиц и взаимосвязей между ними. Анализ социальных сетей исследует эти взаимосвязи, чтобы описать отдельных лиц и группы как сегменты социальной структуры. Отдельные лица взаимодействуют друг с другом, и шаблоны этих взаимодействий обеспечивают более глубокое понимание поведения участвующих в них лиц. Взаимосвязи обеспечивают перенос информации по сети, позволяя отдельным лицам влиять друг на друга. Важность информации о взаимосвязях определяет анализ

социальных сетей независимо от других подходов. Вместо сосредоточенного изучения каждого лица по отдельности единицей изучения становится диада, состоящая из двух отдельных лиц и их взаимосвязи.

Взаимосвязи в сети могут быть классифицированы либо как направленные, либо как ненаправленные. В *направленной взаимосвязи* одно отдельное лицо определяется как инициатор (или источник) взаимосвязи, а другое - как приемник взаимосвязи (или назначение). Например, телефонный звонок - это направленная взаимосвязь, в которой один человек вызывает другого. В отличие от этого, роли источника и назначения для *ненаправленных взаимосвязей* определить нельзя. В этом случае обе стороны участвуют во взаимосвязи одинаково. Пример ненаправленной взаимосвязи - разговор людей.

Другое свойство, проводящее различие между взаимосвязями - это то, является ли взаимосвязь оцениваемой или дихотомической. В случае *дихотомической взаимосвязи* доступна единственная информация: наличие или отсутствие данной взаимосвязи между двумя отдельными лицами. Для диады в сети взаимосвязь либо представлена, либо отсутствует. С другой стороны, *оцениваемая взаимосвязь* включает вес, указывающий силу взаимосвязи. Веса делают возможным сравнение взаимосвязей между собой.

В таблице "Типы взаимосвязей" приведены примеры перекрестной классификации взаимосвязей по направлениям и шкалам. В направленных взаимосвязях *Иван* - источник взаимосвязи, а *Мария* - ее назначение. В ненаправленных взаимосвязях нет указания, кто инициировал взаимосвязь. В оцениваемых взаимосвязях продолжительность разговора используется как вес взаимосвязи, тогда как дихотомические взаимосвязи либо имели место, либо нет.

Таблица 1. Типы взаимосвязей.

Направление	Масштаб	Пример
Ненаправленная	Дихотомическая	Иван и Мария разговаривают друг с другом
Ненаправленная	Оцениваемая	Иван и Мария разговаривают друг с другом в течение 20 минут
Направленная	Дихотомическая	Иван позвонил Марии
Направленная	Оцениваемая	Иван позвонил Марии и разговаривал с ней 20 минут

Дополнительную информацию по анализу социальных сетей можно найти в одной из общеобразовательных книг по этой тематике ¹.

Вывод сетей

Социальную сеть обычно иллюстрируют при помощи *социограммы* ². В визуальном выводе этого типа отдельные лица соответствуют точкам (или узлам) в пространстве. Линии (или ребра), соединяющие точки, представляют взаимосвязи между отдельными лицами. В случае направленных взаимосвязей ребра снабжены стрелкой, указывающей направление. Если у взаимосвязей есть веса, метки для ребер указывают значения. На следующей диаграмме показана сеть для семи отдельных лиц.

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..

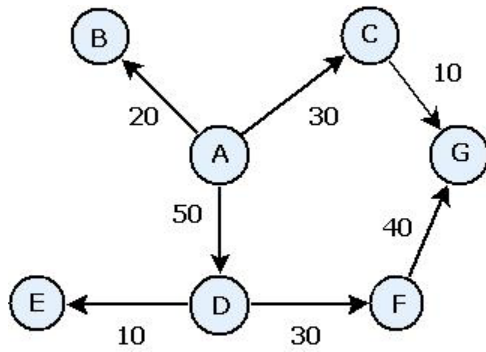


Рисунок 1. Пример социальной сети

Предположим, сеть представляет телефонные вызовы, сделанные отдельными лицами с весами взаимосвязей, указывающими продолжительность этих звонков. В данном случае А позвонил трем лицам, потратив основную часть времени на разговор с D.

Эта сеть значительно меньше реальных. Однако понятия, иллюстрируемые простыми социограммами, обобщаются на сети любого размера и сложности.

Описание сетей

Информацию о сетях, группах и отдельных лицах необходимо экстрагировать в описательные характеристики, допускающие перекрестные сравнения и включение в предсказательные модели. Сети требуется очистить, получив конечное множество ключевых показателей эффективности, поддающихся анализу. Например, вы хотите сравнить сети или группы узлов в сети друг с другом. Другой вариант - вы можете захотеть сравнить отдельных лиц в сети друг с другом или выявить наиболее важных лиц.

Обычно для описания социальных сетей используются две меры: **плотность** и **степень**. Обе эти статистики отражают соединения, но первая сосредоточена на всей сети или ее подгруппах, тогда как вторая характеризует отдельных лиц в сети.

Плотность сети

Для любого множества узлов в сети существует конечное число возможных взаимосвязей. Любой узел может выступать в качестве источника или назначения взаимосвязи с каждым другим узлом. Рассмотрим сеть, состоящую из трех узлов A, B и C. В следующей таблице приведены все возможные направленные взаимосвязи между этими узлами.

Таблица 2. Возможные направленные взаимосвязи для трех узлов.

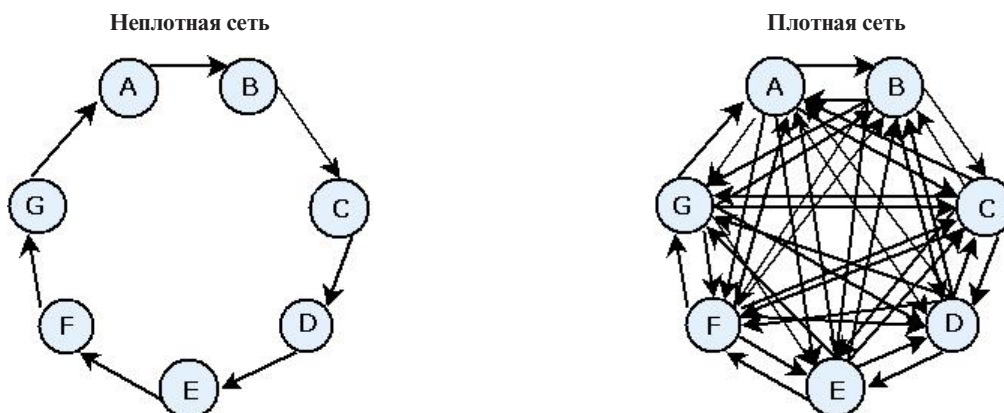
Источник	Назначение
A	D
A	C
D	A
D	C
C	A
C	D

Каждый узел является источником взаимосвязи с остальными двумя узлами. Однако на практике могут быть представлены фактически не все возможные взаимосвязи. У некоторых узлов может отсутствовать какая-нибудь взаимосвязь с другими узлами. Кроме того, для некоторых направленных взаимосвязей могут отсутствовать ответные взаимосвязи.

Статистика *плотности* представляет соотношение возможных взаимосвязей в сети, представленных фактически. Значение изменяется в диапазоне от 0 до 1 с нижним пределом, соответствующим сетям без взаимосвязей, и верхним пределом, представляющим сети со всеми возможными взаимосвязями. Чем ближе значение к 1, тем плотнее сеть и тем более связаны узлы в этой сети.

В плотных сетях перенос информации легче, чем в неплотных. В таблице "Неплотная и плотная сети" показаны две сети, состоящие из семи узлов. Неплотная сеть содержит только семь из 42 возможных взаимосвязей между узлами, что дает плотность 0,17. С другой стороны, плотная сеть содержит все возможные взаимосвязи, и ее плотность составляет 1,0.

Таблица 3. Неплотная и плотная сети.



В неплотной сети для переноса информации с узла А на узел G ее надо передать через пять остальных узлов. И наоборот, в плотной сети информация может пройти от узла А к узлу G напрямую.

Степень узлов

Чаще всего важные лица в сети - это те, кто участвует в большинстве взаимосвязей. Эти лица получают информацию из разнообразных источников и распространяют ее среди большого числа других лиц. И наоборот, лица, участвующие в небольшом числе взаимосвязей, не могут непосредственно влиять на большое число других лиц в сети.

Степень для узла, определяемая как общее число взаимосвязей, где участвует этот узел, разрешает операции сравнения между участниками сети. Лица с более высокими значениями степени более активны, чем лица с более низкими значениями. Степень игнорирует направление взаимосвязей, обеспечивая меру активности для узла.

Для направленных взаимосвязей можно сосредоточиться на том, является ли узел их источником или назначением, при подсчете числа взаимосвязей. *Входящая степень* для узла представляет собой число взаимосвязей, в которых отдельный узел является назначением. И наоборот, *исходящая степень* представляет собой число взаимосвязей, в которых узел является источником. В следующей таблице приведены значения степени, входящей степени и исходящей степени для каждого узла на рисунке "Пример социальной сети".

Таблица 4. Значения степени, входящей степени и исходящей степени примера.

Узел	Градус	Входящая степень	Исходящая степень
A	3	0	3
D	1	1	0
C	2	1	1
D	3	1	2
E	1	1	0

Таблица 4. Значения степени, входящей степени и исходящей степени примера (продолжение).

Узел	Градус	Входящая степень	Исходящая степень
Ж	2	1	1
Г	2	2	0



Входящая степень часто рассматривается как мера престижа. Более высокие значения входящей степени соответствуют большему числу взаимосвязей, заканчивающихся на этом узле. Другими словами, к таким лицам обращается большое число других лиц. Многие другие узлы инициируют взаимосвязи с этим узлом. И наоборот, исходящая степень рассматривается как мера средоточия. Более высокие значения соответствуют большему числу взаимосвязей, исходящих от этого узла. Такие лица обращаются к большому числу других лиц.

Для узлов в сети примера значения степени указывают, что узлы А и D наиболее активны, а узлы В и Е наименее активны. Значения входящей степени показывают, что у узла G наибольший престиж. На основе значений исходящей степени узел А - самый центральный.

Узлы IBM SPSS Modeler Social Network Analysis

Наряду с множеством стандартных узлов, поставляемых с IBM SPSS Modeler, вы можете также, работая с узлами IBM SPSS Modeler Social Network Analysis, включить результаты анализа социальных сетей в потоки. В таблице "Узлы IBM SPSS Modeler Social Network Analysis" описаны эти узлы, хранящиеся на палитре Источники.

Таблица 5. Узлы IBM SPSS Modeler Social Network Analysis.

Узел	Значок	Описание
Групповой анализ		Узел анализа групп импортирует данные записей с подробностями о вызовах из текстового файла в формате с полями фиксированной ширины, идентифицирует группы узлов в сети, определяемой записями, и генерирует ключевые показатели эффективности для групп и отдельных лиц в сети. Дополнительную информацию смотрите в разделе "Обзор анализа групп" на стр. 11.
Анализ распространения		Узел анализа распространения импортирует данные записей с подробностями о вызовах из текстового файла в формате с полями фиксированной ширины, распространяет эффект по сети, определяемой записями, и генерирует ключевые показатели эффективности, суммируя результаты эффекта на отдельных узлах. Дополнительную информацию смотрите в разделе "Обзор анализа распространения" на стр. 17.

Вкладки узлов

На узлах IBM SPSS Modeler Social Network Analysis для определения и предварительного просмотра анализа предлагаются следующие вкладки:

- **Вкладка Данные.** Используется для идентификации файла, содержащего информацию о социальной сети.
- **Вкладка Опции построения.** Используется для задания значений параметров для анализа.
- **Вкладка Анализ.** Используется для просмотра сводки предварительного вывода, обеспечивающей руководство по изменению значений параметров вывода на вкладке Данные с целью генерирования оптимальных результатов.

Кроме того, на узлах предлагается следующие вкладки, общие для узлов в IBM SPSS Modeler:

- **Вкладка Фильтр.** Используется для устранения или переименования выходных полей, сгенерированных узлами. Эта вкладка предлагает те же функции, что и узел фильтра.

- **Вкладка Типы.** Используется для задания шкалы измерений для выходных полей, сгенерированных узлами. Эта вкладка предлагает те же функции, что и узел типа.
- **Вкладка Аннотации.** Используется для переименования узлов, задания пользовательских подсказок и хранения очень длинных аннотаций.

Дополнительную информацию об общих вкладках смотрите в документации IBM SPSS Modeler.

Предварительный просмотр вывода

Обычно вывод узлов используется в качестве входных данных для предсказательной модели. Если вы хотите просмотреть вывод, в поток можно добавить узел таблицы. Однако объем этих данных обычно крайне велик, что делает генерирование таблицы процессом, занимающим много времени. Кроме того, число строк в таблице ограничивает ее полезность.

Как вариант - можно сгенерировать таблицу предварительного просмотра, показывающую пример вывода, который будет создан. Предварительный просмотр показывает сгенерированные поля для ограниченного числа строк. Число строк определяется в свойствах потока. Для предварительного просмотра вывода узла выполните следующие действия:

1. Откройте узел.
2. Задайте значения параметров данных на вкладке Данные.
3. Определите параметры анализа на вкладке Опции построения.
4. Нажмите кнопку **Предварительный просмотр**.

Откроется окно предварительного просмотра с выводом результатов. Кроме того, при предварительном просмотре вывода вкладка Анализ для узла заполняется информацией сводного предварительного просмотра результатов.

Анализ данных

Определение значений параметров анализа, генерирующих оптимальные результаты, - как правило, итерационный процесс. Вы определяете значения параметров, запускаете анализ и просматриваете результаты. Если результаты не настолько полезны, насколько они ими могут быть, вы можете изменить значения параметров и перезапустить анализ.

Чтобы проанализировать входные данные для узла, выполните следующие действия:

1. Откройте узел.
2. Задайте значения параметров данных на вкладке Данные.
3. Определите параметры анализа на вкладке Опции построения.
4. Нажмите кнопку **Анализировать данные**.

Если в опциях построения указано, что должны выводиться сводные статистики, результаты будут показаны на вкладке Анализ.

Если потребуется перезапуск анализа, нажмите кнопку **Очистить анализ** для очистки текущих результатов перед нажатием кнопки **Анализировать данные**.

Программы

К специальным областям применения, где IBM SPSS Modeler Social Network Analysis особенно выгоден, можно отнести следующие:

- **Предсказание оттока.** Характеристики групп могут влиять на коэффициенты оттока. Сосредоточив внимание на лицах в группах, риск оттока которых повышен, его можно предотвратить. Кроме того, можно идентифицировать отдельных лиц с риском оттока из-за информации, поступающей от лиц, отток которых уже произошел.

- **Выгодное использование лидеров групп.** Лидеры групп весьма влиятельны по отношению к другим членам групп. Если удастся предотвратить отток лидера группы, можно будет уменьшить коэффициент оттока для членов группы. Другой вариант - попытка получить лидера группы от конкурента с оттоком может увеличить коэффициент оттока членов группы, связанных с этим конкурентом при одновременном уменьшении коэффициента оттока для членов группы, связанных с вашей компанией.
- **Маркетинг.** Лидеров групп можно использовать для ввода новых товаров или предложений по обслуживанию. Влияние лидера может сделать других членов группы более склонными к приобретению предложения. При помощи анализа распространения можно идентифицировать отдельных лиц, сильнее всего затрагиваемых лидерами групп, и нацелить маркетинг на них.

С IBM SPSS Modeler Social Network Analysis поставляются два демонстрационных потока, которые дают примеры включения результатов анализа социальных сетей в ваши потоки. Файлы данных и потоки примеров устанавливаются в папке *Demos* в каталоге установки продукта.

- *DA demo streams.str* дает пример анализа данных для идентификации первых 300 клиентов, которые вероятнее всего могут уйти к конкурентам.
- *GA demo streams.str* дает пример использования ключевых показателей эффективности (КПЭ) для предсказания оттока как групп, так и отдельных клиентов, а также использования КПЭ для таргетирования на определенных клиентов при маркетинговых компаниях.

Структура данных

Информация об отдельных лицах в исследуемой сети может присутствовать в самых разнообразных файлах, базах данных и системах вашего предприятия. Для анализа сети при помощи IBM SPSS Modeler Social Network Analysis требуется извлечь из данных нужные записи и поля и сформатировать их для ввода в узлы.

Узлам анализа требуются записи с подробностями о вызовах, хранящиеся в одном текстовом файле с полями фиксированной ширины. Каждая строка этого файла соответствует взаимосвязи с данными, организованными в следующих столбцах:

- идентификатор иницилирующего взаимосвязь.
- идентификатор того, кому предназначена взаимосвязь.
- необязательный вес взаимосвязи.

Все данные должны быть числовыми, с идентификаторами лиц, ограниченными целочисленными значениями. Включение имен полей в качестве первой строки данных в файле необязательно. Эта структура данных показана в "Пример записей с данными о вызовах".

Заметим, что записи заголовков также должны использовать либо номера, либо пробельные значения. Например, заголовок со значением *EF BB BF* вызвал бы ошибку.

Таблица 6. Пример записей с данными о вызовах.

Источник	Назначение	Вес
1000	5642	243
2190	8444	831
0299	9419	559

Значения веса могут соответствовать любой мере, которую вы хотите использовать для представления важности взаимосвязи относительно других взаимосвязей в сети. Для данных вызовов общепринятые веса - это продолжительность или частота вызовов. Заметим, что это так для диффузионного анализа, но групповой анализ поддерживает только частоту вызовов.

Если вы хотите сосредоточить анализ на поднаборе хронологии вызовов, при создании входного файла нужно использовать именно этот поднабор. Например, анализ можно ограничить несколькими минувшими месяцами или самыми свежими вызовами для отдельного лица, включив во входной файл только эти данные.

Свойства сценариев

Сценарии в IBM SPSS Modeler Social Network Analysis - это мощный инструмент для автоматизации процессов в пользовательском интерфейсе. Сценарии могут выполнять действия того же типа, которые выполняются с помощью мыши или клавиатуры, и их можно использовать для автоматизации задач, которые при ручном выполнении могли бы быть многократно повторяемыми или требующими большого времени. Объяснение использования сценариев смотрите в руководстве *ScriptingAutomation.pdf*, доступном с IBM SPSS Modeler.

Свойства Diffusion node

В следующей таблице приведены свойства сценариев для узла распространения.

Имя свойства	Тип переменной	Описание свойства
input_data_file_name	строка	
calling_field	поле	
called_field	поле	
frequency_weight_field	поле	
read_field_names	логический	
diffusion_list_file_name	строка	
spreading_factor	двойной точности	Значение по умолчанию = 50. Минимум = 1 Максимум = 99.
max_number_iterations	целое	Значение по умолчанию = 100. Минимум = 1.
accuracy_threshold	двойной точности	Значение по умолчанию = 0,01. Минимум = 0,001.
calculate_statistics	логический	

Свойства узла анализа групп

В следующей таблице приведены свойства сценариев для узла анализа групп.

Имя свойства	Тип переменной	Описание свойства
input_data_file_name	строка	
calling_field	поле	
called_field	поле	
frequency_weight_field	поле	
read_field_names	логический	

Имя свойства	Тип переменной	Описание свойства
coverage_threshold	<i>двойной точности</i>	Значение по умолчанию = 10. Минимум = 1. Максимум = 99.
min_group_size	<i>целое</i>	Значение по умолчанию = 2. Минимум = 2.
max_group_size	<i>целое</i>	Значение по умолчанию = 100. Минимум = 2.
calculate_statistics	<i>логический</i>	

Глава 2. Групповой анализ

Обзор анализа групп

Анализ групп при помощи шаблонов взаимодействий для отдельных лиц в сети идентифицирует группы похожих людей. Характеристики этих групп влияют на поведение отдельных членов групп. Например, у небольших групп, где множество взаимосвязей между членами и сильные лидеры, повышен риск оттока, даже если отток ни одного из членов группы фактически не произошел.³ Предсказательные модели, объединяющие в себе меры и для групп, и для отдельных лиц, будут выполняться лучше, чем модели, включающие в себя только последнее.

Анализ групп состоит из следующих общих шагов:

1. Определение силы взаимосвязей, отражающей социальную близость. Дополнительную информацию смотрите в разделе “Определение социального сходства”.
2. Разделение сети на группы на основе силы взаимосвязей при соблюдении ограничений на размеры. Дополнительную информацию смотрите в разделе “Разделение на группы” на стр. 12.
3. Профилирование и групп, и отдельных лиц, включая идентификацию лидеров групп. Дополнительную информацию смотрите в разделе “Описание групп и их членов” на стр. 12.

Определение социального сходства

Степень сходства членов группы между собой должна быть выше, чем с лицами, не входящими в группу. В анализе сети сходство двух узлов зависит от их взаимосвязей. Для любого узла в сети существует набор узлов, являющихся назначениями направленных взаимосвязей с этим узлом. Для данных о нагрузке линий связи это соответствует всем людям, с которыми общается отдельное лицо. Если с одним и тем же кругом лиц общаются два отдельных лица, эти отдельные лица считаются схожими друг с другом. Чем больше назначений взаимосвязей перекрываются для двух лиц, тем больше эти лица схожи друг с другом.

Рассмотрим сеть, показанную на рисунке “Пример сети с 10 узлами”.

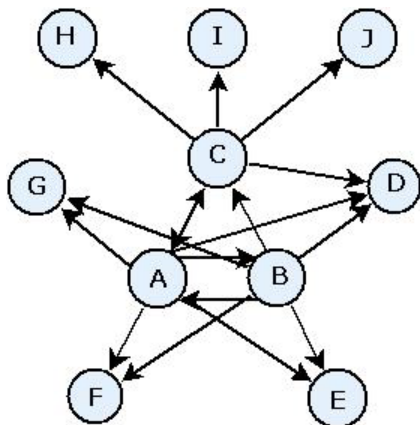


Рисунок 2. Пример сети с 10 узлами

В таблице “Узлы назначения” показаны узлы назначения для взаимосвязей, исходящих от узлов А, В и С.

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

Таблица 7. Узлы назначения.

Исходный узел	Узлы назначения
A	B, C, D, E, F, G
D	A, C, D, E, F, G
C	A, D, H, I, J

У узлов A и B - пять общих узлов назначения. Однако у узлов A и C - только один. Поэтому узлы A и B более схожи социально, чем узлы A и C.

Весы взаимосвязей, например, для продолжительности или частоты вызовов, не захватывают степень сходства узлов в сети. Для включения категории сходства в анализе групп в качестве веса взаимосвязи используется взаимная информация⁴. Эта статистика отражает вероятность соединения двух узлов с одним и тем же узлом. Если у взаимосвязей в сети есть заранее определенные веса (например, для продолжительности вызовов в данных о нагрузке линий связи), взаимная информация объединяет эти веса соответствующим образом.

Разделение на группы

Группы должны состоять из отдельных лиц с высокими весами взаимосвязей с другими членами групп, где веса являются мерой сходства узлов во взаимосвязи. Поэтому идентификация групп начинается опусканием более слабых взаимосвязей в сети. Этим процессом управляет *порог покрытия*, определяющий долю самых сильных взаимосвязей, которые следует сохранить. Например, значение порога покрытия 0,4 приводит к использованию для идентификации групп 40% самых сильных взаимосвязей и опусканию остальных 60%.

Остающиеся взаимосвязи могут образовать очень маленькие или очень большие группы, ограничивающие пригодность предсказания. Для предотвращения включения таких групп в анализ можно задать минимальный и максимальный размеры групп. Группы размером меньше указанного минимума будут полностью опущены. Однако группы, размер которых больше указанного максимума, будут разделены на более мелкие группы в диапазоне приемлемых размеров. Группы, остающиеся после отбрасывания слабых взаимосвязей и применения пределов к размерам, называют *основными группами*.

Удаление взаимосвязей из исходной сети может привести к тому, что некоторых лиц не окажется ни в одной из основных групп. Однако у этих отдельных лиц могут быть соединения с членами группы, гарантирующие их включение в эту группу. Отдельные лица добавляются в группу при наличии относительно сильных взаимосвязей с основными ее членами при условии ненарушения предельных размеров группы. Таким образом, окончательные группы будут состоять из набора основных их членов и членов, добавленных из-за наличия соединений с основными членами группы.

Описание групп и их членов

Помимо статистик плотности, входящей степени и исходящей степени динамику группового поведения описывают и другие статистики. В частности, оценки авторитета и рассеяния предлагают меры социального статуса отдельных лиц в группах. Роль каждого лица в группе крайне важна при попытке предсказать поведение группы и ее членов.

Оценка авторитета для узла измеряет стремление других узлов в группе соединиться с этим узлом. Если множество отдельных лиц обращаются к конкретному сотруднику (надеясь получить информацию или узнать его мнение), этот сотрудник обладает ролью авторитета. Авторитет оценивается для узлов в группе (соответствующих стационарным вероятностям для метода случайного блуждания с перезапусками через групповую сеть) в диапазоне от 0 до 1. Чем ближе оценка авторитета к 1, тем больше авторитет у этого узла в группе. Узел в группе с самой высокой оценкой авторитета называют *лидером авторитета* для этой группы. Деление наивысшей оценки в группе на наименьшую дает меру общей силы лидера авторитета.

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

В отличие от этого, *оценка рассеяния* для узла измеряет стремление узла соединиться с другими узлами в группе. Если конкретный человек обращается ко множеству других людей в группе, он может существенно влиять на мнения всей группы. Рассеяние оценивается для узлов в группе (соответствующих стационарным вероятностям для метода случайного блуждания с перезапусками через групповую сеть с обратным направлением) в диапазоне от 0 до 1. Чем ближе оценка рассеяния к 1, тем больше узел соединяется с другими членами группы. Узел в группе с самой высокой оценкой рассеяния называют *лидером рассеяния* для этой группы. Деление наивысшей оценки в группе на наименьшую дает меру общей силы лидера рассеяния.

Узел анализа групп

Узел анализа групп (доступный на палитре Источники) идентифицирует отдельных лиц в сети, социально похожих друг на друга, и определяет относительный социальный статус отдельных лиц в группе. Для групп относительно небольшого размера со множеством соединений между членами и сильным лидером повышен риск оттока из-за динамики группового поведения. Результаты анализа можно использовать для определения оценок оттока групп, которые при использовании в сочетании с оценками оттока для отдельных лиц повышают возможность предсказания оттока по отношению к моделям на основе оценок отдельных лиц, используемых автономно.

С учетом объема данных, из которых обычно состоит сеть, и характера социальных взаимодействий, лежащих в основе определений групп, анализ групп представляет собой процесс, занимающий много времени. Обычно результаты анализа групп сохраняются в базе данных или файле с целью использования в качестве входных данных для предсказательной модели. Групповые результаты будут обновляться относительно нечасто (например, раз в месяц), тогда как модель может обновляться намного чаще.

Требования. Узлу требуется файл с текстом фиксированной ширины, определяющий социальную сеть при помощи трех полей. Одно поле идентифицирует источник для каждой направленной взаимосвязи, другое определяет для каждой направленной взаимосвязи назначение и третье задает для каждой взаимосвязи необязательную силу. Все взаимосвязи в этой сети должны быть направленными.

Задание данных для анализа групп

На вкладке Данные окна узла источника Анализ групп можно задать входной файл, содержащий взаимосвязи узлов сети.

Файл. Задайте имя файла или папки, содержащих записи с подробностями о вызовах. Можно ввести имя или нажать кнопку с многоточием (...), чтобы выбрать имя из файловой системы. Путь появится, как только вы выберете имя; будет также выведено содержимое файла с разделителями. Если задать папку, для анализа будут сконкатенированы записи с подробностями о вызовах во всех файлах содержащихся в этой папке.

Читать имена полей из файла. Выбираемая по умолчанию эта опция обрабатывает первую строку в файле данных как строку имен для столбцов. Если первая строка не является строкой заголовка, отключите эту опцию, чтобы автоматически дать каждому полю типовое имя, такое как *Поле1* и *Поле2*.

Параметры определения сети

Параметры определения сети определяют роли для полей.

Поля. При помощи кнопок со стрелками назначьте вручную элементы из этого списка для различных полей ролей. Значки обозначают допустимые уровни измерения для каждого поля роли. Нажмите кнопку **Все**, чтобы выбрать все поля в списке, или нажмите кнопку отдельного уровня измерений, чтобы выбрать все поля с этим уровнем.

Источник. Выберите одно поле в качестве источника направленной взаимосвязи.

Назначение. Выберите одно поле в качестве назначения направленной взаимосвязи.

Вес. Необязательно: выберите одно поле, соответствующее числу вхождений той или иной взаимосвязи в сети. При анализе записи будут взвешиваться соответствующим образом при определении сил взаимосвязей.

Задание опций построения для анализа групп

На вкладке Опции построения окна узла источника Анализ групп можно определить опции для идентификации групп в сети.

Параметры анализа групп

Параметры анализа групп влияют на размер и относительную силу образуемых групп.

Порог покрытия. Определите долю самых сильных взаимосвязей для использования при анализе. Например, порог покрытия 0,2 указывает, что будут использоваться только взаимосвязи, веса которых попадают в верхние 20% всех весов взаимосвязей. Изменяющийся от 0 до 1, этот параметр позволяет сосредоточить анализ на самых сильных взаимосвязях, которые должны встречаться в группах. Более высокие значения порога ведут к образованию групп с более слабыми взаимосвязями между их членами. Дополнительную информацию смотрите в разделе “Разделение на группы” на стр. 12.

Минимальный размер группы. Задайте нижнюю границу для размера групп. Группы размером меньше этого значения возвращены не будут.

Максимальный размер группы. Задайте верхнюю границу для размера групп. Группы размером больше этого значения будут разделены на меньшие группы.

Вычислять и выводить сводные статистики. Если выбрана эта опция, узел, помимо получения вывода ключевых показателей эффективности, вычисляет и выводит сводные статистики для анализа. Вычисление этих статистик может отрицательно сказаться на производительности узла для очень больших сетей.

Просмотр статистик анализа групп

На вкладке Анализ окна узла источника Анализ групп предоставляется сводный обзор групп, идентифицированных в сети. В таблице "Сводные статистики анализа групп" указаны доступные сводные статистики для групп.

Таблица 8. Сводные статистики анализа групп.

Статистика	Описание
Общее количество узлов в группах	Количество узлов, включенных в идентифицированные группы
Общее количество связей в группах	Число связей, включенных в идентифицированные группы
Общее количество групп	Число групп, идентифицированных в сети
Средний размер группы	Среднее число узлов в группе
Средняя плотность группы	Средняя доля непосредственных соединений между узлами в группе. Дополнительную информацию смотрите в разделе “Плотность сети” на стр. 3.
Средняя доля основных членов	Средняя доля отдельных узлов в группе, являющихся основными узлами для этой группы. Дополнительную информацию смотрите в разделе “Разделение на группы” на стр. 12.
Средняя плотность основной группы	Средняя доля непосредственных соединений между основными узлами в группе.
Средняя входящая степень	Среднее число входящих связей. Дополнительную информацию смотрите в разделе “Степень узлов” на стр. 4.
Средняя исходящая степень	Среднее число исходящих связей. Дополнительную информацию смотрите в разделе “Степень узлов” на стр. 4.

Выберите в таблице сводных статистик конкретную статистику, чтобы просмотреть для нее распределение значений, среднее квадратичное отклонение и асимметрию.

Просмотр этих статистик помогает определить значения параметров анализа групп. В идеальном случае в результатах должны быть показаны относительно небольшие группы с высокими значениями плотности. Например, при наличии нескольких очень крупных групп попробуйте уменьшить максимальный размер группы, чтобы разделить эти группы на меньшие. Другой вариант - в случае в основном небольших значений плотности групп попробуйте уменьшить порог покрытия, чтобы сосредоточиться на более сильных взаимосвязях.

Вывод для анализа групп

Узел анализа групп генерирует разнообразные поля, описывающие группы и отдельных лиц в группах. Существующие модели и данные можно усилить указанными ключевыми показателями эффективности для улучшения предсказаний, генерируемых моделями. Например, можно изменить значения склонности отдельных лиц к оттоку, чтобы включить влияние групп.

В таблице "Ключевые показатели эффективности для групп" приведен список полей ключевых показателей эффективности для групп, идентифицируемых при анализе.

Таблица 9. Ключевые показатели эффективности для групп.

Поле	Описание
GAG_GroupNumber	Уникальный идентификатор для группы
GAG_Size	Число лиц в группе
GAG_Density	Доля непосредственных соединений между отдельными лицами в группе. Дополнительную информацию смотрите в разделе "Плотность сети" на стр. 3.
GAG_KernelDensity	Доля непосредственных соединений между отдельными лицами в группе.
GAG_CoreNodesFraction	Доля отдельных лиц в группе, являющихся основными лицами для этой группы. Дополнительную информацию смотрите в разделе "Разделение на группы" на стр. 12.
GAG_MaxRankType1	Максимальная оценка авторитета любого члена группы. Дополнительную информацию смотрите в разделе "Описание групп и их членов" на стр. 12.
GAG_MinRankType1	Минимальная оценка авторитета любого члена группы.
GAG_MaxMinRankRatioType1	Отношение наибольшей оценки авторитета к наименьшей. Это значение отражает силу авторитета лидера группы.
GAG_MaxRankType2	Максимальная оценка рассеяния любого члена группы. Дополнительную информацию смотрите в разделе "Описание групп и их членов" на стр. 12.
GAG_MinRankType2	Минимальная оценка рассеяния любого члена группы.
GAG_MaxMinRankRatioType2	Отношение наибольшей оценки рассеяния к наименьшей. Это значение отражает силу рассеяния лидера группы.

В таблице "Ключевые показатели эффективности для отдельных лиц" приведен список полей ключевых показателей эффективности для отдельных лиц в сети.

Таблица 10. Ключевые показатели эффективности для отдельных лиц.

Поле	Описание
GAI_NodeNumber	Уникальный идентификатор для отдельного лица
GAI_CoreNode	Указывает, является ли отдельное лицо основным лицом для группы. Дополнительную информацию смотрите в разделе "Разделение на группы" на стр. 12.

Таблица 10. Ключевые показатели эффективности для отдельных лиц (продолжение).

Поле	Описание
GAI_RankType1	Оценка авторитета для отдельного лица. Дополнительную информацию смотрите в разделе “Описание групп и их членов” на стр. 12.
GAI_RankOrderType1	Порядок ранга в группе на основе оценок авторитета
GAI_RankType2	Оценка рассеяния для отдельного лица. Дополнительную информацию смотрите в разделе “Описание групп и их членов” на стр. 12.
GAI_RankOrderType2	Порядок ранга в группе на основе оценок рассеяния
GAI_InDegree	Число взаимосвязей, в которых данное лицо является назначением взаимосвязи. Дополнительную информацию смотрите в разделе “Степень узлов” на стр. 4.
GAI_OutDegree	Число взаимосвязей, в которых данное лицо является источником взаимосвязи. Дополнительную информацию смотрите в разделе “Степень узлов” на стр. 4.
GAI_GroupLeaderType1	Является ли узел доверенным лидером, лидерство которого основано на входящих связях. Дополнительную информацию смотрите в разделе “Описание групп и их членов” на стр. 12.
GAI_GroupLeaderConfidenceType1	Показатель доверия того, что узел является доверенным лидером группы.
GAI_GroupLeaderType2	Является ли узел лидером рассеяния, оценка лидерства которого основана на исходящих связях. Дополнительную информацию смотрите в разделе “Описание групп и их членов” на стр. 12.
GAI_GroupLeaderConfidenceType2	Доверительная вероятность того, что узел является лидером рассеяния.

Глава 3. Анализ распространения

Обзор анализа распространения

При анализе распространения выявляются отдельные лица, наиболее затрагиваемые другими лицами в социальной сети с количественным выражением эффекта в виде *распространяемой энергии*. В этом процессе используется подход активации с распределением, при котором эффект итерационно распределяется с сетевых узлов на смежные с ними узлы, и уменьшением размеров, благодаря перемещению с одних узлов на другие⁵. При получении энергии узел становится активированным и передает часть этой энергии всем смежным узлам, являющимся назначениями в направленных взаимосвязях данного узла.

Показатель распределения определяет соотношение переданной активированным узлом энергии с оставшимся количеством, сохраняемым на узле. Любые узлы, получающие эту энергию, сами передают такую же ее долю своим непосредственным соседям, в результате чего получается затухающий процесс передачи энергии. Более высокие показатели распределения соответствуют более высокому передаваемому количеству энергии, что позволяет энергии достигать узлов на большем расстоянии от исходно активированных узлов перед полным затуханием процесса. Низкие показатели распределения приводят к быстро затухающим процессам распространения, с переданной энергией, остающейся относительно близко к исходным узлам.

Общее количество энергии, распространяемой активированным узлом, распределяется по всем узлам, являющимся назначениями направленных взаимосвязей с данным узлом. Количество, которое получает каждый узел, зависит от силы взаимосвязи с активированным узлом. Доля, отправляемая конкретному узлу, равна весу взаимосвязи, деленному на общую сумму весов для всех взаимосвязей, в которых активированный узел является источником. Следовательно, соседи, у взаимосвязей которых более высокие относительные веса, получают энергии больше соседей с более низкими относительными весами.

Процесс распространения останавливается при наличии одного из следующих условий:

- активированные узлы не являются источниками никаких направленных взаимосвязей;
- количество передаваемой энергии находится ниже *порога точности* (предельного количества передаваемой энергии, необходимой для продолжения процесса);
- число итераций достигло заданного предела.

По завершении процесса распространения узлы, у которых количество распространенной энергии самое высокое, являются наиболее чувствительными к эффекту, инициировавшему этот процесс. Например, если процесс начинается на узлах, где происходит отток, узлы с наибольшим количеством энергии будут сами представлять собой самый высокий риск оттока. Этим узлам можно уделить особое внимание, чтобы предотвратить их отток.

Пример анализа распространения

Рассмотрим сеть, показанную на в теме “Вывод сетей” на стр. 2. Сеть состоит из нескольких узлов, соединенных друг с другом направленными взаимосвязями различной сил.

В таблице "Пример процесса распространения" показан процесс распространения по сети с применением показателя распределения 0,8. Исходно узел A содержит всю энергию с произвольно назначенным значением 1,0. На шаге 1 этот узел активируется, распределяя 80% своей энергии на соседние узлы, являющиеся назначениями взаимосвязей, а 20% оставляя себе. Общий вес взаимосвязей составляет 100. На взаимосвязь с узлом D приходится половина от этого суммарного количества, из-за чего этот узел принимает половину

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. В сборнике: *Proceedings of the 11th international conference on extending database technology*. New York, NY: ADM.

распространяемой энергии (или 0,4). На взаимосвязь с узлом В приходится 20% от общего веса, что дает значение распространяемой энергии 0,16. Узел С получает остальную энергию (0,24), составляющую 30% объема, исходящего от узла А.

Таблица 11. Пример процесса распространения.

Шаг	Узел А	Узел В	Узел С	Узел D	Узел Е	Узел F	Узел 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

С энергией, полученной на шаге 1, на шаге 2 активируются узлы В, С и D и распространяют энергию на все соседние узлы назначения. У узла В нет узлов назначения, поэтому никакой энергии он не распространяет. У узла С, напротив, есть узел назначения, и 80% своей энергии он распространяет на этот узел G. И наконец, узел D 80% своей энергии распределяет по узлам Е и F, причем F получает энергии в 3 раза больше узла Е из-за своей более сильной взаимосвязи с D.

На шаге 3 активируются узлы Е, F и G. У узлов Е и G нет узлов назначения, поэтому ни один из них никакой энергии не распространяет. Однако узел F распространяет 80% своей энергии на свой единственный узел назначения, узел G. Теперь общее значение энергии этого узла составляет 0,38, куда входит энергия, распространяемая обоими узлами: С, и F.

В этот момент активируется узел G, но у него нет узлов назначения для распространения; поэтому процесс завершается. Энергия, образующаяся на узле А, распределяется по сети, причем узел G получает наибольшее ее количество. Если энергия, вносимая в сеть, представляет отток, узел G больше всего будет затронут оттоком через узел А.

Узел анализа распространения

Узел анализа распространения (доступный на палитре Источники) распространяет эффект от заданного набора лиц через социальную сеть, где используются сетевые взаимосвязи для идентификации пользователей, наиболее затрагиваемых эффектом. Например, если эффект представляет собой отток, узел идентифицирует таких лиц, отток которых наиболее вероятен из-за оттока других конкретных лиц в сети. Существующие модели и данные можно усилить выводом узлов для улучшения генерируемых этими моделями предсказаний. Например, можно изменить значения склонности отдельных лиц к оттоку, чтобы включить влияние операций распространения.

Требования. Узлу требуются два файла с текстом фиксированной ширины, содержащие данные для анализа. Первый файл определяет социальную сеть при помощи трех полей. Одно поле идентифицирует источник для каждой направленной взаимосвязи, другое определяет назначение для каждой направленной взаимосвязи, а третье задает необязательную силу для каждой взаимосвязи. Все взаимосвязи в этой сети должны быть направленными. Второе поле содержит список идентификаторов, откуда должен начинаться эффект.

Задание данных для анализа распространения

На вкладке Данные окна узла источника Анализ распространения можно задать входной файл, содержащий взаимосвязи узлов сети.

Файл. Задайте имя файла или папки, содержащих записи с подробностями о вызовах. Можно ввести имя или нажать кнопку с многоточием (...), чтобы выбрать имя из файловой системы. Путь появится, как только вы выберете имя; будет также выведено содержимое файла с разделителями. Если задать папку, для анализа будут сконкатенированы записи с подробностями о вызовах во всех файлах содержащихся в этой папке.

Читать имена полей из файла. Выбираемая по умолчанию эта опция обрабатывает первую строку в файле данных как строку имен для столбцов. Если первая строка не является строкой заголовка, отключите эту опцию, чтобы автоматически дать каждому полю типовое имя, такое как *Поле1* и *Поле2*.

Параметры определения сети

Параметры определения сети определяют роли для полей.

Поля. При помощи кнопок со стрелками назначьте вручную элементы из этого списка для различных полей ролей. Значки обозначают допустимые уровни измерения для каждого поля роли. Нажмите кнопку **Все**, чтобы выбрать все поля в списке, или нажмите кнопку отдельного уровня измерений, чтобы выбрать все поля с этим уровнем.

Источник. Выберите одно поле в качестве источника направленной взаимосвязи.

Назначение. Выберите одно поле в качестве назначения направленной взаимосвязи.

Вес. Выберите одно поле, представляющее вес взаимосвязи. Для записей с подробностями о вызовах вес может соответствовать любой из множества характеристик вызовов, отражая силу взаимосвязи, например, для продолжительности или частоты вызовов.

Задание опций построения для анализа распространения

На вкладке Опции построения окна узла источника Анализ распространения можно определить опции для процесса распространения.

Начальный список распространения

Начальный список процесса распространения идентифицирует узлы в сети, откуда должна распространяться энергия. Например, для анализа оттока этот список идентифицирует узлы сети, отток которых уже произошел. В других случаях этот список содержит узлы с наивысшим уровнем риска оттока для исследования эффектов в сети, приводящих к фактическому оттоку этих узлов.

Файл. Задайте имя текстового файла, содержащего идентификаторы узлов, где каждый идентификатор соответствует строке в файле. Можно ввести имя файла или нажать кнопку с многоточием (...), чтобы выбрать файл.

Параметры анализа распространения

Параметры анализа распространения определяют степень агрессивности распространения и момент остановки процесса.

Показатель распределения. Определите долю энергии, передаваемой узлом на каждом шаге процесса распространения. Меньшие значения ограничивают распространение энергии только узлами, ближайшими к распространяющему ее узлу. И наоборот, большие значения допускают влияние энергии на более удаленные узлы.

Максимум итераций. Задайте верхнюю границу для числа итераций процесса распространения. Процесс распространения завершается автоматически по достижении этого предела.

Порог точности. Укажите наименьшее изменение энергии узла, гарантирующее продолжение процесса распространения. Когда изменение энергии станет меньше этого значения для каждого узла, процесс распространения завершится.

Вычислять и выводить сводные статистики. Если выбрана эта опция, узел, помимо получения вывода ключевых показателей эффективности, вычисляет и выводит сводные статистики для анализа. Вычисление этих статистик может отрицательно сказаться на производительности узла для очень больших сетей.

Просмотр статистик анализа распространения

На вкладке Анализ в окне узла источника Анализ распространения предоставляется сводный обзор результатов процесса распространения. В таблице "Сводные статистики анализа распространения" указаны доступные сводные статистики.

Таблица 12. Сводные статистики анализа распространения.

Статистика	Описание
Общее количество узлов в сети	Число узлов в сети
Общее количество связей в сети	Число связей в сети
Общее количество начальных значений в сети	Число узлов, используемых в качестве начальных значений для процесса распространения.
Среднее влияние	Среднее количество распространяемой энергии, связанной с отдельными лицами.
Средняя входящая степень	Среднее число взаимосвязей, отдельное лицо в которых является назначением взаимосвязи. Дополнительную информацию смотрите в разделе "Степень узлов" на стр. 4.
Средняя исходящая степень	Среднее число взаимосвязей, отдельное лицо в которых является источником взаимосвязи. Дополнительную информацию смотрите в разделе "Степень узлов" на стр. 4.

Выберите в таблице сводных статистик конкретную статистику, чтобы просмотреть для нее распределение значений, среднее квадратичное отклонение и асимметрию.

Вывод для анализа распространения

Узел анализа распространения генерирует разнообразные поля, описывающие отдельных лиц в сети. Существующие модели и данные можно усилить указанными ключевыми показателями эффективности для улучшения предсказаний, генерируемых моделями. Например, можно изменить значения склонности отдельных лиц к оттоку, чтобы включить влияние операций распространения.

В таблице "Ключевые показатели эффективности анализа распространения" приведен список полей ключевых показателей эффективности для отдельных лиц в анализе.

Таблица 13. Ключевые показатели эффективности анализа распространения.

Поле	Описание
DA_NodeNumber	Уникальный идентификатор для отдельного лица
DA_DiffusedEnergy	Количество распространяемой энергии, связанной с отдельным лицом. Для анализа оттока более высокие значения указывают на более высокую склонность к оттоку, чем более низкие значения.
DA_InDegree	Число взаимосвязей, в которых данное лицо является назначением взаимосвязи. Дополнительную информацию смотрите в разделе "Степень узлов" на стр. 4.
DA_OutDegree	Число взаимосвязей, в которых данное лицо является источником взаимосвязи. Дополнительную информацию смотрите в разделе "Степень узлов" на стр. 4.

Уведомления

Эта информация относится к продуктам и сервису, предлагаемым по всему миру.

IBM может не предоставлять в других странах продукты, услуги и аппаратные средства, описанные в данном документе. За информацией о продуктах и услугах, предоставляемых в вашей стране, обращайтесь к местному представителю IBM. Ссылки на продукты, программы или услуги IBM не означают и не предполагают, что можно использовать только указанные продукты, программы или услуги IBM. Разрешается использовать любые функционально эквивалентные продукты, программы или услуги, если при этом не нарушаются права IBM на интеллектуальную собственность. Однако ответственность за оценку и проверку работы любого продукта, программы или сервиса, не произведенного корпорацией IBM, лежит на пользователе.

IBM может располагать патентами или рассматриваемыми заявками на патенты, относящимися к предмету данного документа. Предъявление данного документа не предоставляет какую-либо лицензию на эти патенты. Вы можете послать письменный запрос о лицензии по адресу:

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

По поводу лицензий, связанных с использованием наборов двухбайтных символов (DBCS), обращайтесь в отдел интеллектуальной собственности IBM в вашей стране или направьте запрос в письменной форме по адресу:

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

Следующий абзац не применяется в Великобритании или в любой другой стране, где подобные заявления противоречат местным законам: INTERNATIONAL BUSINESS MACHINES CORPORATION ПРЕДСТАВЛЯЕТ ДАННУЮ ПУБЛИКАЦИЮ "КАК ЕСТЬ", БЕЗ КАКИХ-ЛИБО ГАРАНТИЙ, КАК ЯВНЫХ, ТАК И ПОДРАЗУМЕВАЕМЫХ, ВКЛЮЧАЯ, НО НЕ ОГРАНИЧИВАЯСЬ ТАКОВЫМИ, ПРЕДПОЛАГАЕМЫЕ ГАРАНТИИ СОБЛЮДЕНИЯ ЧЬИХ-ЛИБО АВТОРСКИХ ПРАВ, ВОЗМОЖНОСТИ КОММЕРЧЕСКОГО ИСПОЛЬЗОВАНИЯ ИЛИ ПРИГОДНОСТИ ДЛЯ КАКИХ-ЛИБО ЦЕЛЕЙ И СООТВЕТСТВИЯ ОПРЕДЕЛЕННОЙ ЦЕЛИ. В некоторых штатах при определенных соглашениях не допускается отказ от выраженных или подразумеваемых гарантий, поэтому данное заявление может к вам не относиться.

Эта информация может содержать технические неточности и типографские ошибки. В представленную здесь информацию периодически вносятся изменения; эти изменения будут включаться в новые издания данной публикации. Фирма IBM может в любое время без уведомления вносить изменения и усовершенствования в продукты и программы, описанные в этой публикации.

Любые приводимые здесь ссылки на web-сайты, не относящиеся к компании IBM, даются исключительно для удобства и ни в коей мере не служат целям поддержки или рекламы этих web-сайтов. Материалы этих Web-сайтов не являются частью данного продукта IBM, и вы можете использовать их только на собственную ответственность.

Любую предоставленную вами информацию IBM может использовать или распространять любым способом, какой сочтет нужным, не беря на себя никаких обязательств по отношению к вам.

Если обладателю лицензии на данную программу понадобятся сведения о возможности: (i) обмена данными между независимо разработанными программами и другими программами (включая данную) и (ii) совместного использования таких данных, он может обратиться по адресу:

IBM Software Group
ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
U.S.A.

Такая информация может быть доступна при соответствующих условиях и соглашениях, включая в некоторых случаях взимание платы.

Описанную в данном документе лицензионную программу и все прилагаемые к ней лицензированные материалы IBM предоставляет на основе положений Соглашения между IBM и Заказчиком, Международного Соглашения о Лицензиях на Программы IBM или любого эквивалентного соглашения между IBM и заказчиком.

Любые данные о выполнении, содержащиеся здесь, были определены в контролируемой среде. Поэтому результаты, полученные в других операционных средах, могут существенно отличаться. Некоторые измерения могли быть сделаны на системах в стадии разработки, и поэтому нет гарантии, что соответствующие показатели останутся теми же на общедоступных системах. Более того, некоторые показатели могли быть оценены путем экстраполяции. Реальные результаты могут отличаться. Пользователи этого документа должны проверить приводимые данные в их конкретной среде.

Информация о продуктах, не принадлежащих компании IBM, была получена от поставщиков этих продуктов, из их опубликованных сообщений или других общедоступных источников. Компания IBM не тестировала эти продукты и не может подтвердить правильность их работы, совместимость и другие утверждения, касающиеся продуктов, не принадлежащих компании IBM. Вопросы о возможностях этих продуктов следует направлять их поставщикам.

Все заявления, касающиеся будущих направлений деятельности или намерений корпорации IBM, подвержены изменению или отмене без предупреждения и являются не более чем выражением целей или намерений.

Эти сведения содержат примеры данных и отчетов, используемых в повседневных деловых операциях. Чтобы проиллюстрировать их настолько полно, насколько это возможно, данные примеры включают имена индивидуумов, названия компаний, брендов и продуктов. Все эти имена и названия являются вымышленными, и любое совпадение с названиями и адресами, используемыми реально действующими компаниями, является чисто случайными.

При просмотре данного электронного информационного документа фотографии и цветные иллюстрации могут не показываться.

Товарные знаки

IBM, логотип IBM, и ibm.com являются товарными знаками или зарегистрированными товарными знаками компании International Business Machines Corp., зарегистрированными во многих странах мира. Прочие наименования продуктов и услуг могут быть товарными знаками, принадлежащими IBM или другим компаниям. Текущий список товарных знаков IBM можно найти в Интернете на странице "Copyright and trademark information" по адресу www.ibm.com/legal/copytrade.shtml.

Intel, логотип Intel, Intel Inside, логотип Intel Inside, Intel Centrino, логотип Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium и Pentium являются товарными знаками или зарегистрированными товарными знаками компании Intel или ее дочерних компаний в Соединенных Штатах и других странах.

Linux является зарегистрированным товарным знаком Linus Torvalds в Соединенных Штатах и других странах.

Microsoft, Windows, Windows NT и логотип Windows являются товарными знаками корпорации Microsoft в Соединенных Штатах и других странах.

UNIX является зарегистрированным товарным знаком The Open Group в Соединенных Штатах и других странах.

Java и все основанные на Java товарные знаки и логотипы - товарные знаки или зарегистрированные товарные знаки Oracle и/или его филиалов.

Другие названия продуктов и услуг могут являться товарными знаками IBM или других компаний.

Глоссарий

В

входящая степень (indegree) . Мера престижа для узла в сети, состоящая из направленных взаимосвязей, определяемая как общее число взаимосвязей, в которых узел является назначением.

Д

дихотомическая взаимосвязь (dichotomous relationship) . Соединение между двумя узлами, которое может принимать всего два значения, указывающие наличие или отсутствие данного соединения.

И

исходящая степень (outdegree) . Мера средоточия для узла в сети, состоящая из направленных взаимосвязей, определяемая как общее число взаимосвязей, в которых узел является источником.

Л

лидер авторитета (authority leader) . Отдельное лицо в группе с наивысшей оценкой авторитета.

лидер рассеяния (dissemination leader) . Отдельное лицо в группе с наивысшей оценкой рассеяния.

Н

направленная взаимосвязь (directional relationship) . Соединение между двумя узлами, исходящее от одного узла (источника) и заканчивающееся на другом (узле назначения).

ненаправленная взаимосвязь (nondirectional relationship) . Связь между двумя узлами.

О

основная группа (core group) . Группа узлов, остающаяся после опускания более слабых взаимосвязей в сети и применения пределов к размерам.

оцениваемая взаимосвязь (valued relationship) . Соединение между двумя узлами, с которым связан вес, указывающий силу взаимосвязи.

оценка авторитета (authority score) . Мера, указывающая степень соединения с отдельным лицом других членов группы. Оценки авторитета связывают важность отдельного лица с числом заканчивающихся на нем взаимосвязей.

оценка рассеяния (dissemination score) . Мера, указывающая степень соединения отдельного лица с другими членами в группе. Оценки рассеяния связывают важность отдельного лица с числом исходящих от него взаимосвязей.

П

плотность (density) . Мера связности сети, определяемая как число наблюдаемых взаимосвязей, деленное на число возможных взаимосвязей.

показатель распределения (spreading factor) . Доля энергии, передаваемая узлом соседям на шаге процесса распространения.

порог покрытия (coverage threshold) . Доля самых сильных взаимосвязей, используемых для идентификации групп.

порог точности (accuracy threshold) . Критерий остановки для процесса распространения соответствующий наименьшему изменению энергии узла, гарантирующему продолжение процесса распространения.

Р

распространяемая энергия (diffused energy) . Эффект, исходящий от конкретных узлов в сети, распространяемый по всей сети и уменьшающийся по размеру по мере увеличения расстояния от исходных узлов.

С

социограмма (sociogram) . Наглядное представление социальной сети, где соединяющиеся точки обозначают отдельных лиц, у которых есть взаимосвязи друг с другом.

степень (degree) . Мера активности узла в сети, определяемая как общее число взаимосвязей, где участвует этот узел.

Индекс

G

GAG_GroupNumber 15

O

outdegree 4, 14, 15, 20

A

анализ данных 6
анализ распространения 5, 17
 начальные значения 19
 параметры 18
показатель распределения 17, 19
порог точности 17, 19
требования 18

B

взаимосвязи
 направление 1
 оцениваемые 1
вкладка Анализ 5
 анализ распространения 20
 групповой анализ 14
вкладка аннотаций 5
вкладка данных 5
 анализ распространения 18
 групповой анализ 13
вкладка опций построения 5
 анализ распространения 19
 групповой анализ 14
вкладка типов 5
вкладка фильтра 5
входящая степень 4, 14, 15, 20
выполнение потоков
 использование IBM SPSS Modeler
 Server Social Network Analysis 5

Г

групповой анализ 5
 порог покрытия 12, 14
 размеры групп 12, 14
 требования 13

Д

дихотомические взаимосвязи 1

З

записи с подробностями о вызовах 7

И

исходящая степень 20

К

ключевые показатели эффективности 15, 20

Л

лидеры авторитета 12, 15
лидеры рассеяния 12, 15

Н

направленные взаимосвязи 1
начальные значения
 для анализа распространения 19
начальный список распространения 19
ненаправленные взаимосвязи 1

O

основная плотность 14, 15
основные группы 12
основные узлы 14, 15
оцениваемые взаимосвязи 1
оценки авторитета 12, 15
оценки рассеяния 12, 15
очистка результатов 6

П

плотность 3, 14, 15
показатель распределения 17, 19
поля веса 7, 13, 18
поля источника 7
поля назначения 7
порог покрытия 12, 14
порог точности 17, 19
предварительный просмотр вывода 6

Р

размеры групп 14, 15
 пределы 12, 14
распространяемая энергия 20

С

свойства
 сценарий 8
среднее влияние 20
средний размер группы 14
средняя входящая степень 14, 20
средняя доля основных членов 14
средняя исходящая степень 14, 20
средняя плотность группы 14
средняя плотность основных групп 14
степень 4
сценарии
 свойства 8

Ф

фильтрация вывода 5

Ш

шкалы измерений 5



Напечатано в Дании