

**IBM SPSS Modeler Social
Network Analysis 17 ユーザ
ー・ガイド**

IBM

お願い

本書および本書で紹介する製品をご使用になる前に、21 ページの『特記事項』に記載されている情報をお読みください。

本書は、IBM SPSS Modeler バージョン 17 リリース 0 モディフィケーション 0、および新しい版で明記されていない限り、以降のすべてのリリースおよびモディフィケーションに適用されます。

お客様の環境によっては、資料中の円記号がバックスラッシュと表示されたり、バックスラッシュが円記号と表示されたりする場合があります。

原典： IBM SPSS Modeler Social Network Analysis 17
User Guide

発行： 日本アイ・ビー・エム株式会社

担当： トランスレーション・サービス・センター

目次

前書き	v	グループへの区分	12
IBM Business Analytics について	v	グループおよびグループ・メンバーの記述	13
技術サポート	v	グループ分析ノード	13
第 1 章 IBM SPSS Modeler Social Network Analysis	1	グループ分析のデータの指定	13
IBM SPSS Modeler Social Network Analysis について	1	グループ分析のビルド・オプションの設定	14
ソーシャル・ネットワーク分析について	2	グループ分析統計の表示	15
ネットワークの表示	2	グループ分析の出力	15
ネットワークの記述	3	第 3 章 拡散分析	17
IBM SPSS Modeler Social Network Analysis のノード	5	拡散分析の概要	17
ノードのタブ	5	拡散分析の例	17
出力のプレビュー	6	拡散分析ノード	18
データの分析	6	拡散分析のデータの指定	18
適用例	6	拡散分析のビルド・オプションの設定	19
データ構造	7	拡散分析統計の表示	20
スクリプト・プロパティ	8	拡散分析の出力	20
第 2 章 グループ分析	11	特記事項	21
グループ分析の概要	11	商標	22
社会的類似性の決定	11	索引	25

前書き

IBM® SPSS® Modeler Social Network Analysis は、人間関係に関する情報を処理して、ソーシャル・ネットワークでの個人の役割を記述するフィールドを作成します。これにより、社会的な情報を予測モデルに組み込むことが可能になります。このマニュアルでは、IBM SPSS Modeler 環境で IBM SPSS Modeler Social Network Analysis ノードを使用して、ストリームにこれらのノードを含めることができるようにする方法を説明します。個人に関する指標を表すフィールドにノード出力を組み合わせると、個人のプロファイルの完成度が高まります。

IBM Business Analytics について

IBM Business Analytics ソフトウェアは、意思決定者が業績向上を図る際に頼りになる、完全で一貫性のある正確な情報を提供します。ビジネス・インテリジェンス、予測分析、財務業績/戦略管理、および分析アプリケーションからなる包括的なポートフォリオにより、現在の業績に関する明確で直接的かつ実用的な洞察を得て、将来の成果を予測できるようになります。豊富な業界ソリューション、実証済みの手法、および専門のサービスを組み合わせることにより、あらゆる規模の企業が、高い生産性を実現するとともに、自信を持って意思決定を自動化し、優れた結果を出すことができます。

このポートフォリオの一部である IBM SPSS Predictive Analytics ソフトウェアにより、企業は将来の事象を予測し、その洞察に基づいて事前対処的に行動して、ビジネス成果を高めることができます。全世界の民間、行政、学術の各分野のお客様が、IBM SPSS テクノロジーを活用して、不正の削減とリスクの軽減を図りながら、顧客の獲得、維持、拡大での競争優位性を高めています。IBM SPSS ソフトウェアを日々の業務に取り入れることで、企業は予測的に対応できるようになり、ビジネス目標を達成し、高い競争優位性を実現するように意思決定を方向付け、自動化することが可能になります。詳細情報、および担当員の連絡先については、<http://www.ibm.com/spss> をご覧ください。

技術サポート

保守契約があるお客様は、テクニカル・サポートをご利用いただけます。IBM Corp. 製品の使用方法、または対応するハードウェア環境へのインストールについてサポートが必要な場合は、テクニカル・サポートにご連絡ください。テクニカル・サポートをご利用になる場合は、IBM Corp. Web サイト (<http://www.ibm.com/support>) をご覧ください。ご利用の際には、ご本人、所属する組織、およびサポート契約を確認できるものをご用意ください。

第 1 章 IBM SPSS Modeler Social Network Analysis

IBM SPSS Modeler Social Network Analysis について

行動をモデリングするための手法の多くは、個人に重点を置いています。これらの手法では、個人に関するさまざまなデータを使用して、行動の主要指標から行動を予測するモデルを生成します。行動の発生に関連する主要指標の値がある個人には、その行動を防止することを目的とした特別な注意を払うことができます。

顧客が企業との関係を終了させる解約をモデリングするための手法について考えます。顧客を維持するコストは、顧客を入れ替えるコストよりもはるかに低いため、解約のリスクがある顧客を特定できることはきわめて重要です。アナリストは多くの場合、人口統計情報や個々の顧客の最近の通話パターンといった、複数の主要業績評価指標を使用して顧客を記述します。このようなフィールドに基づいた予測モデルでは、過去に解約した顧客の通話パターンと一致する顧客の通話パターンの変化を使用して、解約のリスクが高まっている顧客を特定します。リスクがあると特定された顧客には、つなぎ止めるための追加のカスタマー・サービスやサービス・オプションが提供されます。

このような方法では、顧客の行動に大きく影響する可能性がある社会的情報が見落とされます。企業に関する情報、および他人の行動に関する情報は、関係を介して広がり、人々に影響を与えます。その結果、他の人々が自身の関係を通じて、個人の意思決定やアクションに影響を与えることが可能となります。個人の指標のみを対象とした分析では、予測能力を持つ重要な要素が欠落してしまいます。

IBM SPSS Modeler Social Network Analysis では、関係の情報を処理して、モデルに組み込み可能な追加フィールドにすることで、この問題に対処します。こうして導出された主要業績評価指標では、個人の社会的特性が測定されます。このような社会的プロパティを個人ベースの指標と組み合わせることで、個人の概要をよりの確に把握できるようになり、結果的にモデルの予測精度が向上します。

IBM SPSS Modeler Social Network Analysis は、以下の 2 つの主要なコンポーネントで構成されています。

- IBM SPSS Modeler 環境に追加された IBM SPSS Modeler Social Network Analysis ノード。社会的分析技術をストリームに組み込むことを可能にします。
- IBM SPSS Modeler Server Social Network Analysis。IBM SPSS Modeler Server にノードの仕様の処理を追加します。IBM SPSS Modeler Server Social Network Analysis は、膨大な数の個人と関係を含む大量のネットワーク・データを、比較的少数のフィールドへと効率的に処理して、その後の分析に備えます。

例えば、IBM SPSS Modeler Social Network Analysis では、解約しようとする特定の人々から最も影響を受けるネットワーク内の個人を特定します。さらに、解約のリスクが高いネットワーク内の個人のグループを見つけ出すことができます。このような効果を示す主要業績評価指標をモデルに取り込むことによって、モデルの全体的な性能を高めることができます。

ソーシャル・ネットワーク分析について

ソーシャル・ネットワークは、個人の集合および個人間の関係で構成されます。ソーシャル・ネットワーク分析では、これらの関係を調査して、個人およびグループを社会構造の一部として記述します。個人は相互に作用し、その相互作用のパターンから、関係する個人についての洞察を得ることができます。関係によって情報をネットワーク全体に広めることができ、一個人が他の個人に影響を与えることが可能となります。関係の情報が重視される点で、ソーシャル・ネットワーク分析は他の手法と異なります。各個人に個別に重点を置くのではなく、2人の個人とその関係で構成されるダイアド（二者関係）が調査の単位となります。

ネットワークでの関係は、有向関係と無向関係に分類できます。有向関係では、一方の個人が関係の開始者（ソース）、もう一方の個人が受動者（宛先）と識別されます。例えば、電話をかける行為は、ある個人が別の個人を呼び出すという有向関係です。これに対し、無向関係では、ソースと宛先の役割を定義できません。この場合は、両方の当事者が等しく関係に関与します。無向関係の一例として会話が挙げられます。

関係を区別するもう1つのプロパティが、関係が二分であるか評価付きであるかです。二分関係で得られる情報は、2人の個人間に関係が存在するかどうかのみです。ネットワークのすべてのダイアドには、関係が存在するか、または存在しません。一方、評価付き関係には、関係の強度を示す重みが含まれています。重みによって、関係を互いに比較することができます。

『関係の種類』の表に、方向およびスケールによる関係の交差分類の例を示します。有向関係では、Joeが関係のソースであり、Maryが関係の宛先です。無向関係では、関係の開始者が示されません。評価付き関係では、会話の長さが関係の重みとして使用されます。一方、二分関係は発生したかしないかのどちらかです。

表 1. 関係の種類：

方向	尺度	例
無向	二分	Joe と Mary が会話した
無向	評価付き	Joe と Mary が 20 分間会話した
有向	二分	Joe が Mary に電話した
有向	評価付き	Joe が Mary に電話して 20 分間会話した

ソーシャル・ネットワーク分析について詳しくは、この分野に関する総合的な書籍¹を参照してください。

ネットワークの表示

ソーシャル・ネットワークは、一般にソシオグラム²を使用して図示されます。このタイプの視覚的表示では、個人が空間における点、つまりノードに対応します。点を結ぶ線、つまり辺は、個人間の関係を表します。関係に方向がある場合は、方向を示す矢印が辺に付加されます。関係に重みがある場合は、辺のラベルに値が示されます。次のグラフは、7人の個人のネットワークを示しています。

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..

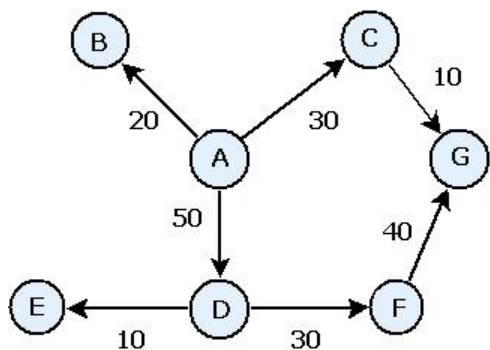


図1. ソーシャル・ネットワークの例

このネットワークは個人がかけた電話を表し、関係の重みが通話の長さを示しているとしします。この場合、個人 A が 3 人に電話し、個人 D との通話に大部分の時間を費やしています。

このネットワークは、実際に発生するネットワークに比べるとはるかに小規模です。ただし、単純なソシオグラムで示される概念によって、あらゆるサイズと複雑さのネットワークが一般化されます。

ネットワークの記述

ネットワーク、グループ、および個人に関する情報を抽出して、相互比較と予測モデルへの組み込みが可能な記述的特性を作成する必要があります。ネットワークは、分析可能な主要業績評価指標の有限集合へと抽出する必要があります。例えば、ネットワーク間またはネットワーク内のノード・グループ間を相互に比較することができます。あるいは、ネットワーク内の個人間を相互に比較したり、最も重要な個人を特定したりすることができます。

ソーシャル・ネットワークを記述する際によく使用される 2 つの指標が、**密度**と**次数**です。どちらの統計も接続性を反映しますが、前者はネットワーク全体またはネットワーク・サブグループに重点を置き、後者はネットワーク内の個人の特徴を記述します。

ネットワークの密度

ネットワーク内のどのノード・セットについても、有限の数の関係が発生する可能性があります。各ノードが、他のすべてのノードとの関係のソースまたはターゲットになることができます。A、B、C の 3 つのノードで構成されたネットワークがあるとしします。次の表に、これらのノード間で可能なすべての有向関係を示します。

表2. 3 つのノード間で可能な有向関係：

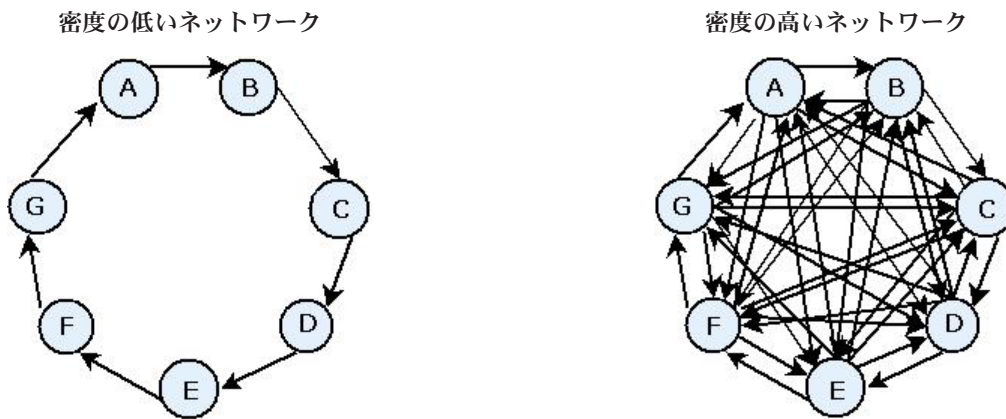
ソース	対象
A	B
A	C
B	A
B	C
C	A
C	B

各ノードが、他の 2 つのノードとの関係のソースになっています。ただし、実際のところ、可能な関係がすべて存在するわけではありません。ノードによっては、他のノードとの直接的な関係がない場合もあります。また、有向関係が相互に発生しない場合もあります。

密度の統計値は、ネットワーク内の可能な関係が実際に存在する割合を示します。値の範囲は 0 から 1 までです。下限値の 0 は関係が存在しないネットワーク、上限値の 1 は可能な関係がすべて存在するネットワークを表します。値が 1 に近いほど、ネットワークの密度が高くなり、ネットワーク内のノードの結合度も高くなります。

密度の高いネットワークでは、密度の低いネットワークよりも情報が拡散しやすくなります。『密度の低いネットワークと高いネットワーク』の図は、7 つのノードで構成される 2 つのネットワークを示しています。密度の低いネットワークには、ノード間で可能な 42 個の関係のうち 7 つの関係しか存在しないため、その密度は 0.17 です。一方、密度の高いネットワークには可能な関係がすべて存在するため、その密度は 1.0 です。

表 3. 密度の低いネットワークと高いネットワーク：



密度の低いネットワークでは、情報をノード A からノード G に流すために、他の 5 つのノードを通過する必要があります。これに対し、密度の高いネットワークでは、情報をノード A からノード G に直接送ることができます。

ノードの次数

ネットワーク内の重要な個人は、多くの場合ほとんどの関係に関与しています。このような個人は、さまざまなソースから情報を取得し、その情報を多数の他の個人に拡散します。これに対し、関与している関係が少ない個人は、ネットワーク内の多数の他の個人に直接影響を与えることはできません。

ノードの次数は、特定のノードが関与している関係の総数と定義され、ネットワーク参加者間の比較を可能にします。次数値が高い個人は、次数値が低い個人よりもアクティブです。次数は、関係の方向を無視して、ノードの活動の全体的な指標を示します。

有向関係では、関係の数を計算する際に、ノードがソースとターゲットのどちらであるかに重点を置くことができます。ノードの入次数は、ある特定のノードがターゲットとなっている関係の数です。これに対して、ノードの出次数は、あるノードがソースとなっている関係の数です。次の表では『ソーシャル・ネットワークの例』の図における各ノードの次数、入次数、および出次数の値を示します。

表 4. 度数、入次数、および出次数の値の例：

ノード	次数	入次数	出次数
A	3	0	3
B	1	1	0
C	2	1	1
D	3	1	2

表 4. 度数、入次数、および出次数の値の例 (続き) :

ノード	次数	入次数	出次数
E	1	1	0
F	2	1	1
G	2	2	0



入次数は、多くの場合、威信の指標として扱われます。入次数の値が高いほど、そのノードを終点とする関係が多くなります。つまり、このような個人は、多数の他の個人から接触されています。その他多数のノードは、そのノードとの関係の起点となっています。これに対して、出次数は中心性の指標として扱われます。出次数の値が高いほど、そのノードを起点とする関係が多くなります。このような個人は、多数の他の個人に接触しています。

例に示したネットワーク内のノードの場合は、次数の値から、ノード A と D が最もアクティブで、ノード B と E が最もアクティブでないことがわかります。入次数の値は、ノード G の威信が最も高いことを示しています。出次数の値に基づくと、ノード A が最も中心的です。

IBM SPSS Modeler Social Network Analysis のノード

IBM SPSS Modeler で提供される多数の標準ノードとともに、IBM SPSS Modeler Social Network Analysis のノードを使用して、ソーシャル・ネットワーク分析の結果をストリームに加えることができます。「ソース」パレットに格納されているこれらのノードについて、『IBM SPSS Modeler Social Network Analysis ノード』の表で説明します。

表 5. IBM SPSS Modeler Social Network Analysis のノード :

ノード	アイコン	説明
グループ分析		グループ分析ノードは、固定フィールド・テキスト・ファイルから通話詳細レコード・データをインポートして、レコードによって定義されるネットワーク内のノード・グループを特定し、ネットワーク内のグループおよび個人に関する主要業績評価指標を生成します。詳細については、11 ページの『グループ分析の概要』のトピックを参照してください。
拡散分析		拡散分析ノードは、固定フィールド・テキスト・ファイルから通話詳細レコード・データをインポートして、レコードによって定義されるネットワーク全体に効果を伝搬し、個人のノードへの効果の結果を要約する主要業績評価指標を生成します。詳細については、17 ページの『拡散分析の概要』のトピックを参照してください。

ノードのタブ

IBM SPSS Modeler Social Network Analysis ノードには、分析の定義とプレビューを行うための以下のタブがあります。

- 「データ」タブ: ソーシャル・ネットワーク情報を含んでいるファイルを特定するために使用します。
- 「ビルド・オプション」タブ: 分析の設定を定義するために使用します。
- 「分析」タブ: 「データ」タブの入力設定を変更して最適な結果を得るためのガイダンスとなる事前出力の要約を表示するために使用します。

また、ノードには IBM SPSS Modeler の入力ノード間で共通の以下のタブがあります。

- 「フィルター」タブ: ノードで生成される出力フィールドを除外したり名前変更したりするために使用します。このタブは、フィルター・ノードと同じ機能を提供します。
- 「データ型」タブ: ノードで生成される出力フィールドの尺度を設定するために使用します。このタブは、データ型ノードと同じ機能を提供します。
- 「注釈」タブ: ノード名の変更、カスタム・ツールチップの指定、長い注釈の保管を行うために使用します。

共通のタブについて詳しくは、IBM SPSS Modeler の資料を参照してください。

出力のプレビュー

一般に、ノードの出力は予測モデルへの入力として使用します。出力を表示したい場合は、ストリームにテーブル・ノードを追加できます。ただし、通常はデータの量が非常に多いため、テーブルの生成に時間がかかります。また、テーブルの行数によって、その有用性が制限されます。

代わりに、作成される出力のサンプルを表示するプレビュー・テーブルを生成することができます。プレビューには、限られた数の行の生成フィールドが表示されます。この行数は、ストリーム・プロパティーで定義されます。ノード出力をプレビューするには、次の手順を実行します。

1. ノードを開きます。
2. 「データ」タブでデータ設定を指定します。
3. 「ビルド・オプション」タブで分析パラメーターを定義します。
4. 「プレビュー」をクリックします。

プレビュー・ウィンドウが開いて結果が表示されます。また、出力をプレビューすると、ノードの「分析」タブに結果の概要が表示されます。

データの分析

最適な結果を出す分析設定を決定する作業は、通常は反復的なプロセスです。設定の定義、分析の実行、および結果のレビューを行います。結果が思っていたほど有益でない場合は、設定を変更して分析を再実行します。

ノードの入力データを分析するには、次の手順を実行します。

1. ノードを開きます。
2. 「データ」タブでデータ設定を指定します。
3. 「ビルド・オプション」タブで分析パラメーターを定義します。
4. 「データの分析」をクリックします。

ビルド・オプションで要約統計を表示するように指定した場合は、「分析」タブに結果が表示されます。

分析を再実行する必要がある場合は、「分析をクリア」をクリックして現在の結果を消去してから「データの分析」をクリックします。

適用例

IBM SPSS Modeler Social Network Analysis が特に役に立つ特定の適用例としては、以下のものではありません。

- **解約予測。**グループの特性が解約率に影響する場合があります。解約のリスクが高いグループに属する個人に重点を置くことで、解約を回避できる可能性があります。また、既に解約した人からの情報の流れが原因で解約するリスクがある個人を特定できます。
- **グループ・リーダーの活用。**グループ・リーダーは、他のグループ・メンバーに大きな影響を与えます。グループ・リーダーの解約を回避できれば、グループ・メンバーの解約率が下がる可能性があります。あるいは、競合他社のグループ・リーダーを解約させることにより、その会社に関連するグループ・メンバーの解約率が上がり、自社に関連するグループ・メンバーの解約率が下がる可能性があります。
- **マーケティング。**グループ・リーダーを使用して、新しい商品やサービスの提供を開始することができます。リーダーの影響で、他のグループ・メンバーが商品を購入する可能性が高くなる場合があります。拡散分析を使用して、グループ・リーダーの影響を最も受ける個人を特定し、マーケティングの対象にすることができます。

IBM SPSS Modeler Social Network Analysis には 2 つのデモ ストリームが用意されており、ソーシャル・ネットワーク分析の結果をストリームに含める方法の例として使用できます。データ ファイルとサンプル ストリームは、製品のインストール ディレクトリーの *Demos* フォルダにインストールされています。

- *DA demo streams.str* では、データを分析して解約する可能性が最も高い上位 300 人の個人を識別する例が示されます。
- *GA demo streams.str* では、主要業績評価指標 (KPI) を使用してグループと個人の両方の解約を予測すると共に、KPI を使用してマーケティング キャンペーンのために特定の個人を対象に設定する例も示されます。

データ構造

対象となるネットワーク内の個人に関する情報が、社内のさまざまなファイル、データベース、およびシステムに分散している可能性があります。IBM SPSS Modeler Social Network Analysis を使用してネットワークを分析するには、関連するレコードおよびフィールドをデータ・ソースから抽出し、ノードに入力できるようにフォーマットする必要があります。

分析ノードでは、1 つの固定幅テキスト・ファイルに格納された通話詳細レコードが必要となります。このファイルの各行は関係に対応しており、データは次の列に編成されています。

- 関係の起点である個人の識別子
- 関係のターゲットである個人の識別子
- 関係の重み (オプション)

データはいずれも数値でなければなりません。個人の識別子は整数に限定されます。オプションで、ファイルの 1 行目のデータとしてフィールド名を入れることができます。『通話データ・レコードの例』の表は、このデータ構造を示しています。

ヘッダー・レコードには数値か空白値のいずれかを使用する必要があることにも注意してください。例えば、*EF BB BF* の値が指定されたヘッダーはエラーの原因となります。

表 6. 通話データ・レコードの例:

ソース	宛先	重み
1000	5642	243
2190	8444	831
0299	9419	559

重みの値は、関係の重要度をネットワーク内の他の関係と比較して表すために使用する任意の指標に対応させることができます。通話データの場合は、通話時間や通話頻度などが重みとしてよく使用されます。これは拡散分析に当てはまりますが、グループ分析は通話頻度のみをサポートする点に注意してください。

通話履歴のサブセットに分析を集中させる場合は、入力ファイル作成時にそのサブセットを使用する必要があります。例えば、ある個人の過去数カ月間または最近の通話データのみを入力テキスト・ファイルに取り込んで、分析の対象をこれらのデータに限定することができます。

スクリプト・プロパティ

IBM SPSS Modeler Social Network Analysis のスクリプトは、ユーザー・インターフェースでのプロセスを自動化するための強力なツールです。スクリプトでは、マウスやキーボードで実行するアクションと同じ種類のアクションを実行できます。また、何度も繰り返し実行されるタスクや、手動で実行すると時間がかかるタスクを自動化する場合にもスクリプトを使用できます。スクリプトの使用については、IBM SPSS Modeler 付属の *ScriptingAutomation.pdf* ガイドを参照してください。

拡散ノードのプロパティ

次の表に、拡散ノードでのスクリプト・プロパティを示します。

プロパティ名	データ型	プロパティの説明
input_data_file_name	文字列	
calling_field	フィールド	
called_field	フィールド	
frequency_weight_field	フィールド	
read_field_names	ブール	
diffusion_list_file_name	文字列	
spreading_factor	倍精度浮動小数点数	デフォルト値 = 50 最小値 = 1 最大値 = 99
max_number_iterations	整数	デフォルト値 = 100 最小値 = 1
accuracy_threshold	倍精度浮動小数点数	デフォルト値 = 0.01. 最小値 = 0.001
calculate_statistics	ブール	

グループ分析ノードのプロパティ

次の表に、グループ分析ノードでのスクリプト・プロパティを示します。

プロパティ名	データ型	プロパティの説明
input_data_file_name	文字列	
calling_field	フィールド	
called_field	フィールド	
frequency_weight_field	フィールド	
read_field_names	ブール	

プロパティ名	データ型	プロパティの説明
coverage_threshold	倍精度浮動小数点数	デフォルト値 = 10 最小値 = 1 最大値 = 99
min_group_size	整数	デフォルト値 = 2 最小値 = 2
max_group_size	整数	デフォルト値 = 100 最小値 = 2
calculate_statistics	ブール	

第 2 章 グループ分析

グループ分析の概要

グループ分析では、ネットワークにおける個人の相互作用パターンを使用して、類似する個人のグループを特定します。これらのグループの特性は、個々のグループ・メンバーの行動に影響します。例えば、メンバー間の関係が多く、強いリーダーがいる小規模なグループは、実際に解約したメンバーがグループにいない場合でも解約のリスクが高くなります。³グループと個人の両方の指標を取り込む予測モデルは、個人の指標のみを使用するモデルに比べて性能が高くなります。

グループ分析は、一般に次の手順で行われます。

1. 社会的近接性を反映する関係の強度を特定します。詳細については、『社会的類似性の決定』のトピックを参照してください。
2. サイズの制限に従いながら、関係の強度に基づいてネットワークをグループに区分します。詳細については、12 ページの『グループへの区分』のトピックを参照してください。
3. グループと個人の両方のプロフィールを作成し、その際にグループ・リーダーを特定します。詳細については、13 ページの『グループおよびグループ・メンバーの記述』のトピックを参照してください。

社会的類似性の決定

グループのメンバーは、グループに入っていない個人と比べて、互いの類似性が高いことが必要です。ネットワーク分析では、2 つのノードの類似性はその関係によって決まります。ネットワーク内の各ノードには、そのノードとの有向関係のターゲットである一連のノードが存在します。通信データの場合は、この一連のノードが、特定の個人が連絡を取るすべての人物に相当します。2 人の個人が同じ一連の人物に連絡を取る場合、この 2 人は互いに類似していると見なされます。2 人の個人の間で重複している関係のターゲットのセットが多いほど、類似性は高くなります。

『10 ノードのネットワークの例』の図に示すネットワークについて考えてみます。

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.

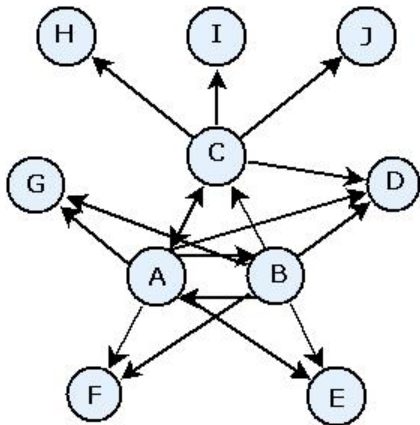


図 2. 10 ノードのネットワークの例

『ターゲット・ノード』の表に、ノード A、B、C を起点とする関係のターゲット・ノードを示します。

表 7. ターゲット・ノード：

入力ノード	ターゲット・ノード
A	B、C、D、E、F、G
B	A、C、D、E、F、G
C	A、D、H、I、J

ノード A と B には、共通のターゲット・ノードが 5 つあります。一方、ノード A と C の共通ターゲット・ノードは 1 つだけです。したがって、ノード A と B は、ノード A と C に比べて社会的類似性が高いこととなります。

通話時間や通話頻度などの関係の重みからは、ネットワークにおけるノードの類似性を確認できません。類似性の概念を加えるために、グループ分析では、相互情報量⁴を関係の重みとして使用します。この統計値は、2 つのノードが同じノードにつながっている可能性を表します。ネットワーク内の関係に、通信データの通話頻度などの重みが事前に定義されている場合は、相互情報量にこの重みが適宜取り込まれます。

グループへの区分

グループは、他のグループ・メンバーとの関係の重みが大きい個人で構成する必要があります。この場合、重みは関係におけるノードの類似性を測定します。このため、グループを特定するには、まずネットワーク内の弱い関係を除外します。カバレッジしきい値は、保持の対象となる最も強い関係の割合を定義することで、このプロセスを制御します。例えば、カバレッジしきい値を 0.4 にすると、最も強い 40% の関係がグループの特定に使用され、残りの 60% の関係が除外されます。

残りの関係から、予測の有用性が限られている非常に小さいグループまたは非常に大きいグループが生成される場合があります。このようなグループが分析に含まれないようにするために、グループの最小サイズと最大サイズを指定することができます。最小サイズより小さいグループは完全に除外されます。最大サイズより大きいグループは、妥当なサイズの範囲内でより小さいグループに分割されます。弱い関係を削除し、サイズの制限を適用した後に残ったグループを、コア・グループと呼びます。

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

元のネットワークとの関係を削除すると、一部の個人がどのコア・グループにも属さなくなる場合があります。ただし、このような個人は、グループに入ることが保証されているグループのメンバーとつながりがある場合があります。グループのコア・メンバーとの比較的強い関係が多数ある場合は、グループ・サイズの制限に違反しない限り、グループに個人が追加されます。したがって、最終的なグループは、コア・メンバーおよびコアとつながっているために追加されたメンバーで構成されます。

グループおよびグループ・メンバーの記述

密度、入次数、および出次数のほかにも、グループ力学を記述する統計があります。特に、権限スコアと拡散スコアは、グループ内での個人のソーシャル・ステータスの指標となります。グループとそのメンバーの両方の行動を予測する場合は、グループでの各個人の役割が非常に重要となります。

ノードの権限スコアは、対象のノードにグループ内の他のノードがつながる傾向を測定します。特定の個人に多数の個人が（場合によっては、情報や意見を求めて）接触している場合、その人物には権限の役割があります。グループ内のノードの権限スコアは、グループ・ネットワークで再開するランダム・ウォークの定常確率に対応し、その範囲は 0 から 1 です。権限スコアが 1 に近いほど、グループ内でのそのノードの権限が高くなります。グループ内で権限スコアが最も高いノードを、グループの権限リーダーと呼びます。グループ内の最大スコアを最小スコアで除算すると、権限リーダーの全体的な強度が測定されます。

これに対し、ノードの拡散スコアは、対象のノードがグループ内の他のノードにつながる傾向を測定します。特定の人物がグループ内の他の多数の人々に接触している場合、その人物はグループ全体の意見に大きな影響を与える可能性があります。グループ内のノードの拡散スコアは、グループ・ネットワークで逆順に再開するランダム・ウォークの定常確率に対応し、その範囲は 0 から 1 です。拡散スコアが 1 に近いほど、そのノードがつながっている他のグループ・メンバーの数が多くなります。グループ内で拡散スコアが最も高いノードを、グループの拡散リーダーと呼びます。グループ内の最大スコアを最小スコアで除算すると、拡散リーダーの全体的な強度が測定されます。

グループ分析ノード

「ソース」パレットから使用できるグループ分析ノードは、ネットワーク内の社会的に互いに類似する個人のセットを特定し、そのグループでの個人の相対的なソーシャル・ステータスを判別します。比較的規模が小さく、メンバー間のつながりが多く、かつ強力なリーダーがいるグループは、グループ力学により解約のリスクが高くなります。この結果を使用してグループの解約スコアを判別し、これを個人の解約スコアと組み合わせることで、個人のスコアのみに基づいたモデルに比べて、解約の予測能力を高めることができます。

データ量が、通常はネットワークを構成し、グループ定義の基礎となる社会的相互作用の性質を包含している場合は、グループ分析のプロセスに時間がかかります。通常、グループ分析の結果は、予測モデルの入力として使用されるデータベースまたはファイルに保存します。グループの結果はそれほど頻繁に更新されませんが（月 1 回程度）、モデルはかなり頻繁に更新される可能性があります。

要件: このノードには、3 つのフィールドでソーシャル・ネットワークを定義する固定幅のテキスト・ファイルが必要です。1 つ目のフィールドでは各有向関係のソースが識別され、2 つ目のフィールドでは各有向関係の宛先が定義され、3 つ目のフィールドでは各関係の強度（オプション）が指定されます。ネットワーク内の関係はすべて有向である必要があります。

グループ分析のデータの指定

「グループ分析」入力ノード・ウィンドウの「データ」タブでは、ネットワーク・ノードの関係が含まれている入力ファイルを指定することができます。

ファイル: 通話詳細レコードを含んだファイルまたはフォルダーの名前を指定します。名前を入力するか、省略符号ボタン (「...」) をクリックして、ファイル・システムから名前を選択できます。名前を選択するとパスが表示され、その内容が区切り文字とともに表示されます。フォルダーを指定した場合は、そのフォルダーに含まれるすべてのファイル内の通話詳細レコードが分析用に連結されます。そのため、フォルダー内のすべてのファイルが同じ構造であることが必要です。

ファイルからフィールド名を取得: このオプションを選択すると (デフォルトで選択されます)、データ・ファイルの 1 行目が列の名前として扱われます。1 行目がヘッダーでない場合は、このオプションの選択を解除します。これにより、各フィールドに *Field1* や *Field2* などの総称が自動的に割り当てられます。

ネットワーク定義設定

ネットワーク定義設定では、フィールドの役割を定義します。

フィールド: 矢印ボタンを使用して、このリストの項目をさまざまな役割フィールドに手動で割り当てます。アイコンは、各役割フィールドの有効な尺度を示します。「すべて」ボタンをクリックしてリスト内のすべてのフィールドを選択するか、個々の尺度ボタンをクリックしてその尺度のフィールドをすべて選択します。

ソース: 有向関係の起点となるフィールドを 1 つ選択します。

宛先: 有向関係のターゲットとなるフィールドを 1 つ選択します。

重み: オプションとして、ネットワーク内で関係が発生する回数に対応するフィールドを 1 つ選択します。これにより、分析で関係の強度を決定する際に、レコードに重みが付けられます。

グループ分析のビルド・オプションの設定

「グループ分析」入力ノード・ウィンドウの「ビルド・オプション」タブでは、ネットワーク内でグループを特定するためのオプションを定義することができます。

グループ分析設定

グループ分析設定は、形成されるグループのサイズおよび相対的な強度に影響します

カバレッジしきい値: 分析で使用する最も強力なネットワーク関係の割合を定義します。例えば、カバレッジしきい値を 0.2 にすると、重みが全体の上位 20% に入る関係のみが使用されることを指定します。0 から 1 までを範囲とするこのパラメーターを使用して、グループ内で発生する最も強力な関係に分析の焦点を置くことができます。しきい値が大きいほど、生成されるグループのメンバー間の関係は弱くなります。詳細については、12 ページの『グループへの区分』のトピックを参照してください。

最小グループ・サイズ: グループ・サイズの下限を指定します。この値より小さいグループは返されません。

最大グループ・サイズ: グループ・サイズの上限を指定します。この値より大きいグループは、より小さいグループに分割されます。

要約統計量の計算および表示: このオプションを選択すると、ノードは主要業績評価指標の出力を導出するほかに、分析の要約統計量を計算して表示します。非常に大きなネットワークでは、この統計量の計算によって、ノードのパフォーマンスに悪影響が生じることがあります。

グループ分析統計の表示

「グループ分析」入力ノード・ウィンドウの「分析」タブには、ネットワークで特定されたグループの概要が表示されます。『グループ分析の要約統計』の表に、グループについて表示される要約統計を示します。

表 8. グループ分析の要約統計：

統計値	説明
グループ内の総ノード数	特定されたグループに含まれているノードの数
グループ内の総リンク数	特定されたグループに含まれているリンクの数
グループの総数	ネットワーク内で特定されたグループの数
平均グループ・サイズ	グループ内の平均ノード数
平均グループ密度	グループ内のノード間が直接つながっている割合の平均。詳細については、3 ページの『ネットワークの密度』のトピックを参照してください。
コア・メンバーの割合の平均	グループ内のノードがグループのコア・ノードである割合の平均。詳細については、12 ページの『グループへの区分』のトピックを参照してください。
平均コア・グループ密度	グループ内のコア・ノード間が直接つながっている割合の平均
平均入次数	受信リンクの平均数。詳細については、4 ページの『ノードの次数』のトピックを参照してください。
平均出次数	送信リンクの平均数。詳細については、4 ページの『ノードの次数』のトピックを参照してください。

要約統計表から特定の統計を選択すると、値の分布、標準偏差、および統計の歪度が表示されます。

これらの統計をレビューすると、グループ分析設定を定義する際に役立ちます。統計の結果が、密度の値が高い比較的小さなグループを示していることが理想的です。例えば、非常に大きなグループがある場合は、最大グループ・サイズを減少させて小さなグループに分割することを検討してください。また、グループ密度の値が低い場合は、カバレッジしきい値を小さくして、より強力な関係に重点を置くことを検討してください。

グループ分析の出力

グループ分析ノードは、グループおよびグループ内の個人について記述するさまざまなフィールドを生成します。既存のモデルとデータをこれらの主要業績評価指標で増補することで、モデルによって生成される予測を改善できます。例えば、グループの影響が反映されるように、個人の解約傾向の値を更新することができます。

『グループの主要業績評価指標』の表に、分析で特定されたグループの主要業績評価指標フィールドをリストします。

表 9. グループの主要業績評価指標：

フィールド	説明
GAG_GroupNumber	グループの一意の識別子
GAG_Size	グループ内の個人の数
GAG_Density	グループ内の個人同士が直接つながっている割合。詳細については、3 ページの『ネットワークの密度』のトピックを参照してください。
GAG_KernelDensity	グループ内のコア個人同士が直接つながっている割合。

表 9. グループの主要業績評価指標 (続き) :

フィールド	説明
GAG_CoreNodesFraction	グループ内の個人がグループのコア個人である割合。詳細については、12 ページの『グループへの区分』のトピックを参照してください。
GAG_MaxRankType1	グループ・メンバーの最大権限スコア。詳細については、13 ページの『グループおよびグループ・メンバーの記述』のトピックを参照してください。
GAG_MinRankType1	グループ・メンバーの最小権限スコア。
GAG_MaxMinRankRatioType1	最小権限スコアに対する最大権限スコアの比率。この値はグループ・リーダーの権限の強度を反映します。
GAG_MaxRankType2	グループ・メンバーの最大拡散スコア。詳細については、13 ページの『グループおよびグループ・メンバーの記述』のトピックを参照してください。
GAG_MinRankType2	グループ・メンバーの最小拡散スコア。
GAG_MaxMinRankRatioType2	最小拡散スコアに対する最大拡散スコアの比率。この値はグループ・リーダーの拡散の強度を反映します。

『個人の主要業績評価指標』の表に、ネットワーク内の個人の主要業績評価指標フィールドをリストします。

表 10. 個人の主要業績評価指標 :

フィールド	説明
GAI_NodeNumber	個人の一意の識別子
GAI_CoreNode	個人がグループのコア個人かどうかを示す指標。詳細については、12 ページの『グループへの区分』のトピックを参照してください。
GAI_RankType1	個人の権限スコア。詳細については、13 ページの『グループおよびグループ・メンバーの記述』のトピックを参照してください。
GAI_RankOrderType1	権限スコアに基づいたグループ内の順位。
GAI_RankType2	個人の拡散スコア。詳細については、13 ページの『グループおよびグループ・メンバーの記述』のトピックを参照してください。
GAI_RankOrderType2	拡散スコアに基づいたグループ内の順位。
GAI_InDegree	対象の個人が関係のターゲットとなっている関係の数。詳細については、4 ページの『ノードの次数』のトピックを参照してください。
GAI_OutDegree	対象の個人が関係のソースとなっている関係の数。詳細については、4 ページの『ノードの次数』のトピックを参照してください。
GAI_GroupLeaderType1	ノードが権限リーダーかどうか。ノードのリーダーシップ・スコアは受信リンクから導出されます。詳細については、13 ページの『グループおよびグループ・メンバーの記述』のトピックを参照してください。
GAI_GroupLeaderConfidenceType1	ノードが権限リーダーであること確信度。
GAI_GroupLeaderType2	ノードが拡散リーダーかどうか。ノードのリーダーシップ・スコアは送信リンクから導出されます。詳細については、13 ページの『グループおよびグループ・メンバーの記述』のトピックを参照してください。
GAI_GroupLeaderConfidenceType2	ノードが拡散リーダーであること確信度。

第 3 章 拡散分析

拡散分析の概要

拡散分析では、ソーシャル・ネットワーク内で他の個人に最も影響を受ける個人を特定して、効果を拡散エネルギーとして定量化します。このプロセスでは、効果がネットワーク・ノードから直接の近隣に繰り返し拡散する活性化拡散手法を使用し、効果がノード間を移動するにつれてそのサイズを縮小します⁵。ノードがエネルギーを受信すると、そのノードはアクティブ化され、エネルギーの一部をそのノードとの有向関係のターゲットである近隣に送信します。

拡散係数は、アクティブ化されたノードから送信されるエネルギーの割合を、ノードで保持される残りの量とともに定義します。このエネルギーを受信したノードは、同じ割合を近隣に送信し、結果的にエネルギー送信のプロセスが崩壊していきます。拡散係数を大きくすると、送信されるエネルギーが増えるため、プロセスが完全に崩壊するまでに、最初にアクティブ化されたノードからより離れたノードにエネルギーが送信されるようになります。拡散係数を小さくすると、拡散プロセスが速く崩壊することになり、送信されたエネルギーが最初のノードから比較的近い場所に留まります。

アクティブ化されたノードから拡散される総エネルギー量は、そのノードとの有向関係のターゲットであるすべてのノード間で分散されます。各ノードが受信するエネルギー量は、アクティブ化されたノードとの関係の強度によって異なります。特定のノードに送信される割合は、その関係の重みを、アクティブ化されたノードがソースとなっている全関係の重みの合計で除算した値となります。したがって、相対的な重みが大きい関係を持つ近隣は、相対的な重みが小さい近隣よりも多くのエネルギーを受信します。

次のいずれかの状況が発生すると、拡散プロセスは停止します。

- アクティブ化ノードが有向関係のソースになっていない
- 送信されるエネルギーの量が、プロセスを続行するための送信エネルギー量の限度である精度しきい値を下回っている
- 反復数が指定の限度に達している

拡散が完了した時点で、拡散エネルギーを最も多く受信したノードが、プロセスを開始した効果に対して最も感度が高くなります。例えば、解約を行うノードでプロセスが開始された場合は、最もエネルギー量の多いノードが最も解約のリスクが高くなります。これらのノードに特別な注意を払って、解約を防止することができます。

拡散分析の例

2 ページの『ネットワークの表示』に示すネットワークについて考えてみます。このネットワークは 7 個のノードで構成されており、各ノードが強度の異なる有向関係で互いに結ばれています。

『拡散プロセスの例』の表は、拡散係数 0.80 を使用するネットワークでの拡散プロセスを示しています。最初は、ノード A にすべてのエネルギーが含まれ、1.00 の値が任意に割り当てられます。ステップ 1 で、このノードがアクティブ化されて、エネルギーの 80% が関係のターゲットである 3 つの近隣に拡散され、20% がノード A に保持されます。関係の重みは合計 100 になります。ノード D との関係がこの

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.

合計の半分を占めるため、ノード D は拡散エネルギーの半分 (0.40) を受信します。ノード B との関係は、全重みの 20% を占めるため、拡散エネルギー値は 0.16 になります。ノード C は残りのエネルギー (0.24) を受信します。これは、ノード A で発生したエネルギー量の 30% です。

表 11. 拡散プロセスの例:

ステップ	ノード A	ノード B	ノード C	ノード D	ノード E	ノード F	ノード 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

ステップ 1 でエネルギーを受信したノード B、C、D は、ステップ 2 でアクティブ化され、ターゲットの近隣にエネルギーを拡散します。ノード B はターゲット・ノードを持たないため、エネルギーを拡散しません。一方、ノード C はターゲットを持ち、エネルギーの 80% をそのノード (G) に拡散します。最後に、ノード D は、エネルギーの 80% をノード E と F に拡散しますが、ノード F はノード D との関係がより強いため、ノード E の 3 倍の量を受信します。

ステップ 3 で、ノード E、F、G がアクティブ化されます。ノード E と G はターゲット・ノードを持たないため、どちらもエネルギーを拡散しません。ただし、ノード F は、80% のエネルギーを 1 つのターゲット・ノード (G) に拡散します。これでノード G の総エネルギー値は、0.38 となります (C と F の両ノードから拡散されたエネルギーで構成されたもの)。

この時点で ノード G がアクティブ化されますが、拡散対象のターゲット・ノードがないため、プロセスが終了します。ノード A で発生したエネルギーは、ネットワーク全体に拡散され、ノード G が最も多くの量を受信します。ネットワークに導入されたエネルギーが解約を表す場合は、ノード G がノード A の解約による影響を最も強く受けることになります。

拡散分析ノード

「ソース」パレットから使用できる拡散分析ノードでは、指定した 1 組の個人からネットワークの関係をを使用してソーシャル・ネットワーク全体に効果を伝搬し、その効果の影響を最も受ける個人を特定します。例えば効果が解約である場合は、ネットワークの解約において、他の特定の個人が原因で解約する可能性が最も高い個人が、このノードによって特定されます。既存のモデルとデータをこのノード出力で増補することで、これらのモデルによって生成される予測を改善できます。例えば、拡散の影響が反映されるように、個人の解約傾向の値を更新することができます。

要件: このノードには、分析するデータが格納された 2 つの固定幅テキスト・ファイルが必要です。一方のファイルでは、3 つのフィールドを使用してソーシャル・ネットワークが定義されます。1 つ目のフィールドでは各有向関係のソースが識別され、2 つ目のフィールドでは各有向関係の宛先が定義され、3 つ目のフィールドでは各関係の強度 (オプション) が指定されます。ネットワーク内の関係はすべて有向である必要があります。もう一方のファイルには、効果の始点となる識別子のリストが含まれています。

拡散分析のデータの指定

「拡散分析」入力ノード・ウィンドウの「データ」タブでは、ネットワーク・ノードの関係が含まれている入力ファイルを指定することができます。

ファイル: 通話詳細レコードを含んだファイルまたはフォルダーの名前を指定します。名前を入力するか、省略符号ボタン (「...」) をクリックして、ファイル・システムから名前を選択できます。名前を選択する

とパスが表示され、その内容が区切り文字とともに表示されます。フォルダーを指定した場合は、そのフォルダーに含まれるすべてのファイル内の通話詳細レコードが分析用に連結されます。そのため、フォルダー内のすべてのファイルが同じ構造であることが必要です。

ファイルからフィールド名を取得: このオプションを選択すると (デフォルトで選択されます)、データ・ファイルの 1 行目が列の名前として扱われます。1 行目がヘッダーでない場合は、このオプションの選択を解除します。これにより、各フィールドに *Field1* や *Field2* などの総称が自動的に割り当てられます。

ネットワーク定義設定

ネットワーク定義設定では、フィールドの役割を定義します。

フィールド: 矢印ボタンを使用して、このリストの項目をさまざまな役割フィールドに手動で割り当てます。アイコンは、各役割フィールドの有効な尺度を示します。「すべて」ボタンをクリックしてリスト内のすべてのフィールドを選択するか、個々の尺度ボタンをクリックしてその尺度のフィールドをすべて選択します。

ソース: 有向関係の起点となるフィールドを 1 つ選択します。

宛先: 有向関係のターゲットとなるフィールドを 1 つ選択します。

重み: 関係の重みを表すフィールドを 1 つ選択します。通話詳細レコードの場合は、関係の強度を反映する複数の通話特性 (通話時間や通話頻度など) のいずれかが重みに対応します。

拡散分析のビルド・オプションの設定

「拡散分析」入力ノード・ウィンドウの「ビルド・オプション」タブでは、拡散プロセスのオプションを定義することができます。

拡散シード・リスト

拡散シード・リストは、エネルギーが拡散する始点となるネットワーク・ノードを特定します。例えば解約分析では、既に解約したネットワーク・ノードをこのリストで特定します。あるいは、解約のリスクが最も高いノードをこのリストに含めて、これらのノードが実際に解約した場合のネットワークへの影響を調べることができます。

ファイル: ノード識別子を含んだテキスト・ファイルの名前を指定します。各識別子がファイルの各行に対応します。ファイル名を入力するか、省略符号ボタン (「...」) をクリックしてファイルを選択できます。

拡散分析設定

拡散分析設定では、拡散の強さと、プロセスを停止するタイミングを決定します。

拡散係数: 拡散プロセスの各ステップでノードから送信されるエネルギーの割合を定義します。値が小さいほど、エネルギーを拡散しているノードに最も近いノードにエネルギーの影響が制限されます。逆に、値が大きいほど、離れたノードにエネルギーの影響を与えることができます。

最大反復回数: 拡散の反復回数の上限を指定します。この上限に達すると、拡散プロセスは自動的に終了します。

精度しきい値: 拡散プロセスの続行を保証する、ノードのエネルギーの変化の最小値を指定します。すべてのノードのエネルギーの変化がこの値より小さくなると、拡散プロセスが終了します。

要約統計量の計算および表示: このオプションを選択すると、ノードは主要業績評価指標の出力を導出するほかに、分析の要約統計量を計算して表示します。非常に大きなネットワークでは、この統計量の計算によって、ノードのパフォーマンスに悪影響が生じることがあります。

拡散分析統計の表示

「拡散分析」入力ノード・ウィンドウの「分析」タブには、拡散の結果の概要が表示されます。『拡散分析の要約統計』の表に、表示される要約統計を示します。

表 12. 拡散分析の要約統計：

統計値	説明
ネットワーク内の総ノード数	ネットワーク内のノードの数
ネットワーク内の総リンク数	ネットワーク内のリンクの数
ネットワーク内の拡散シードの総数	拡散プロセスのシードとして使用されるノードの数
平均の影響	個人に関連する拡散エネルギーの平均量
平均入次数	個人が関係のターゲットとなっている関係の平均数。詳細については、4 ページの『ノードの次数』のトピックを参照してください。
平均出次数	個人が関係のソースとなっている関係の平均数。詳細については、4 ページの『ノードの次数』のトピックを参照してください。

要約統計表から特定の統計を選択すると、値の分布、標準偏差、および統計の歪度が表示されます。

拡散分析の出力

拡散分析ノードは、ネットワーク内の個人を記述するさまざまなフィールドを生成します。既存のモデルとデータをこれらの主要業績評価指標で増補することで、モデルによって生成される予測を改善できます。例えば、拡散の影響が反映されるように、個人の解約傾向の値を更新することができます。

『拡散分析の主要業績評価指標』の表に、分析における個人の主要業績評価指標フィールドをリストします。

表 13. 拡散分析の主要業績評価指標：

フィールド	説明
DA_NodeNumber	個人の一意の識別子
DA_DiffusedEnergy	個人に関連する拡散エネルギーの量。解約分析の場合は、値が高いほど、低い値に比べて解約傾向が大きいことを示します。
DA_InDegree	対象の個人が関係のターゲットとなっている関係の数。詳細については、4 ページの『ノードの次数』のトピックを参照してください。
DA_OutDegree	対象の個人が関係のソースとなっている関係の数。詳細については、4 ページの『ノードの次数』のトピックを参照してください。

特記事項

本書は、全世界で提供する製品およびサービスについて作成したものです。

本書に記載の製品、サービス、または機能が日本においては提供されていない場合があります。日本で利用可能な製品、サービス、および機能については、日本 IBM の営業担当員にお尋ねください。本書で IBM 製品、プログラム、またはサービスに言及していても、その IBM 製品、プログラム、またはサービスのみが使用可能であることを意味するものではありません。これらに代えて、IBM の知的所有権を侵害することのない、機能的に同等の製品、プログラム、またはサービスを使用することができます。ただし、IBM 以外の製品とプログラムの操作またはサービスの評価および検証は、お客様の責任で行っていただきます。

IBM は、本書に記載されている内容に関して特許権 (特許出願中のものを含む) を保有している場合があります。本書の提供は、お客様にこれらの特許権について実施権を許諾することを意味するものではありません。実施権についてのお問い合わせは、書面にて下記宛先にお送りください。

〒103-8510

東京都中央区日本橋箱崎町19番21号

日本アイ・ビー・エム株式会社

法務・知的財産

知的財産権ライセンス渉外

以下の保証は、国または地域の法律に沿わない場合は、適用されません。IBM およびその直接または間接の子会社は、本書を特定物として現存するままの状態を提供し、商品性の保証、特定目的適合性の保証および法律上の瑕疵担保責任を含むすべての明示もしくは黙示の保証責任を負わないものとします。国または地域によっては、法律の強行規定により、保証責任の制限が禁じられる場合、強行規定の制限を受けるものとします。

この情報には、技術的に不適切な記述や誤植を含む場合があります。本書は定期的に見直され、必要な変更は本書の次版に組み込まれます。IBM は予告なしに、随時、この文書に記載されている製品またはプログラムに対して、改良または変更を行うことがあります。

本書において IBM 以外の Web サイトに言及している場合がありますが、便宜のため記載しただけであり、決してそれらの Web サイトを推奨するものではありません。それらの Web サイトにある資料は、この IBM 製品の資料の一部ではありません。それらの Web サイトは、お客様の責任でご使用ください。

IBM は、お客様が提供するいかなる情報も、お客様に対してなんら義務も負うことのない、自ら適切と信ずる方法で、使用もしくは配布することができるものとします。

本プログラムのライセンス保持者で、(i) 独自に作成したプログラムとその他のプログラム (本プログラムを含む) との間での情報交換、および (ii) 交換された情報の相互利用を可能にすることを目的として、本プログラムに関する情報を必要とする方は、下記に連絡してください。

IBM Software Group

ATTN: Licensing

200 W. Madison St.

Chicago, IL; 60606

U.S.A.

本プログラムに関する上記の情報は、適切な使用条件の下で使用することができますが、有償の場合もあります。

本書で説明されているライセンス・プログラムまたはその他のライセンス資料は、IBM 所定のプログラム契約の契約条項、IBM プログラムのご使用条件、またはそれと同等の条項に基づいて、IBM より提供されます。

この文書に含まれるいかなるパフォーマンス・データも、管理環境下で決定されたものです。そのため、他の操作環境で得られた結果は、異なる可能性があります。一部の測定が、開発レベルのシステムで行われた可能性があります。その測定値が、一般に利用可能なシステムのものと同じである保証はありません。さらに、一部の測定値が、推定値である可能性があります。実際の結果は、異なる可能性があります。お客様は、お客様の特定の環境に適したデータを確かめる必要があります。

IBM 以外の製品に関する情報は、その製品の供給者、出版物、もしくはその他の公に利用可能なソースから入手したものです。IBM は、それらの製品のテストは行っておりません。したがって、他社製品に関する実行性、互換性、またはその他の要求については確認できません。IBM 以外の製品の性能に関する質問は、それらの製品の供給者をお願いします。

IBM の将来の方向または意向に関する記述については、予告なしに変更または撤回される場合があります、単に目標を示しているものです。

本書には、日常の業務処理で用いられるデータや報告書の例が含まれています。より具体性を与えるために、それらの例には、個人、企業、ブランド、あるいは製品などの名前が含まれている場合があります。これらの名称はすべて架空のものであり、名称や住所が類似する企業が実在しているとしても、それは偶然にすぎません。

この情報をソフトコピーでご覧になっている場合は、写真やカラーの図表は表示されない場合があります。

商標

IBM、IBM ロゴおよび ibm.com は、世界の多くの国で登録された International Business Machines Corporation の商標です。他の製品名およびサービス名等は、それぞれ IBM または各社の商標である場合があります。現時点での IBM の商標リストについては <http://www.ibm.com/legal/copytrade.shtml> をご覧ください。

インテル、Intel、Intel ロゴ、Intel Inside、Intel Inside ロゴ、Centrino、Intel Centrino ロゴ、Celeron、Xeon、Intel SpeedStep、Itanium、および Pentium は、Intel Corporation または子会社の米国およびその他の国における商標または登録商標です。

Linux は、Linus Torvalds の米国およびその他の国における登録商標です。

Microsoft、Windows、Windows NT および Windows ロゴは、Microsoft Corporation の米国およびその他の国における商標です。

UNIX は The Open Group の米国およびその他の国における登録商標です。

Java およびすべての Java 関連の商標およびロゴは Oracle やその関連会社の米国およびその他の国における商標または登録商標です。

用語集

A

精度しきい値 (*accuracy threshold*) . 拡散プロセスを停止する基準。拡散の継続を保証するノード・エネルギーの変化の最小値に対応します。

権限リーダー (*authority leader*) . グループ内で権限スコアが最も高い個人。

権限スコア (*authority score*) . ある個人にグループの他のメンバーがどの程度つながっているかを示す指標。権限スコアによって、個人の重要度が、その個人を終点とする関係の数と関連付けられます。

C

コア・グループ (*core group*) . 弱い関係をネットワークから除外し、グループ・サイズの制限を適用した後に残ったノードのグループ。

カバレッジしきい値 (*coverage threshold*) . グループの特定に使用される最も強力な関係の割合。

D

次数 (*degree*) . ネットワークにおけるノードの活動の指標。ノードが関与している関係の総数と定義されます。

密度 (*density*) . ネットワーク結合度の指標。観測された関係の数を可能な関係の数で除算した値と定義されます。

二分関係 (*dichotomous relationship*) . 2つのノード間のつながり。つながりの有無を示す2つの値のいずれか1つのみを取ります。

拡散エネルギー (*diffused energy*) . ネットワーク内の特定のノードを起点としてネットワーク全体に拡散する効果。起点のノードからの距離が遠くなるほどサイズは小さくなります。

有向関係 (*directional relationship*) . 一方のノードを起点 (ソース)、もう一方のノードを終点 (宛先) とする2つのノード間のつながり。

拡散リーダー (*dissemination leader*) . グループ内で拡散スコアが最も高い個人。

拡散スコア (*dissemination score*) . 個人がグループの他のメンバーにどの程度つながっているかを示す指標。拡散スコアによって、個人の重要度が、その個人を起点とする関係の数と関連付けられます。

I

入次数 (*indegree*) . 有向関係で構成されるネットワーク内の、あるノードの威信を示す指標。そのノードがターゲットとなっている関係の総数と定義されます。

N

無向関係 (*nondirectional relationship*) . 2つのノード間の関連性。

O

出次数 (outdegree) . 有向関係で構成されるネットワーク内の、あるノードの中心性を示す指標。そのノードがソースとなっている関係の総数と定義されます。

S

ソシオグラム (sociogram) . ソーシャル・ネットワークの視覚的な表現。点を結ぶ線によって、複数の個人間の相互関係が表されます。

拡散係数 (spreading factor) . 拡散プロセスのステップでノードが近隣に送信するエネルギーの割合。

V

評価付き関係 (valued relationship) . 関係の強度を示す重みに関連付けられた 2 つのノード間のつながり。

索引

日本語, 数字, 英字, 特殊文字の順に配列されています。なお, 濁音と半濁音は清音と同等に扱われています。

[ア行]

宛先フィールド 7
重みフィールド 7, 13, 18

[カ行]

カーネル密度 15
拡散エネルギー 20
拡散係数 17, 19
拡散シード・リスト 19
拡散スコア 13, 15
拡散分析 5, 17
 オプション 18
 拡散係数 17, 19
 シード 19
 精度しきい値 17, 19
 要件 18
拡散リーダー 13, 15
カバレッジしきい値 12, 14
関係
 評価付き 2
 方向 2
グループ分析 5
 カバレッジしきい値 12, 14
 グループ・サイズ 12, 14
 要件 13
グループ・サイズ 15
 制限 12, 14
結果の消去 6
権限スコア 13, 15
権限リーダー 13, 15
コア・グループ 12
コア・ノード 15
コア・メンバーの割合の平均 15

[サ行]

シード
 拡散分析 19
次数 4
尺度 5
出力のフィルタリング 5
出力のプレビュー 6
主要業績評価指標 15, 20

スクリプト
 プロパティ 8
ストリームの実行
 IBM SPSS Modeler Server Social
 Network Analysis を使用 5
精度しきい値 17, 19
ソース・フィールド 7

[タ行]

「注釈」タブ 5
通話詳細レコード 7
「データ」タブ 5
 拡散分析 18
 グループ分析 13
「データ型」タブ 5
データの分析 6
出次数 4, 15, 20

[ナ行]

二分関係 2
入次数 4, 15, 20

[ハ行]

評価付き関係 2
「ビルド・オプション」タブ 5
 拡散分析 19
 グループ分析 14
「フィルター」タブ 5
プロパティ
 スクリプト 8
「分析」タブ 5
 拡散分析 20
 グループ分析 15
平均グループ密度 15
平均グループ・サイズ 15
平均コア・グループ密度 15
平均出次数 15, 20
平均入次数 15, 20
平均の影響 20

[マ行]

密度 3, 15
無向関係 2

[ヤ行]

有向関係 2

G

GAG_GroupNumber 15



Printed in Japan