

*Guia do Usuário do IBM SPSS Modeler
Social Network Analysis 17.1*

IBM

Nota

Antes de utilizar estas informações e o produto suportado por elas, leia as informações em “Avisos” na página 21.

Informações sobre o Produto

Esta edição se aplica à versão 17, liberação 1, modificação 0 do IBM(r) SPSS(r) Modeler e a todas as liberações e modificações subsequentes até que seja indicado de outra forma em novas edições.

Índice

| | |
|--|----------|
| Prefácio | v |
| Sobre o IBM Business Analytics | v |
| Suporte técnico | v |

Capítulo 1. IBM SPSS Modeler Social Network Analysis **1**

| | |
|--|---|
| Sobre o IBM SPSS Modeler Social Network Analysis | 1 |
| Sobre a análise de rede social | 2 |
| Exibindo redes | 2 |
| Descrevendo redes | 3 |
| IBM SPSS Modeler Social Network Analysisnodes | 5 |
| Guias Nó | 5 |
| Visualizando saída | 6 |
| Analisando dados | 6 |
| Aplicativos | 6 |
| Estrutura de dados | 7 |
| Propriedades de script | 8 |

Capítulo 2. Análise de grupo **11**

| | |
|--|----|
| Visão geral de análise de grupo | 11 |
| Determinando a similaridade social | 11 |
| Particionamento em grupos | 12 |

| | |
|--|----|
| Descrevendo grupos e membros do grupo | 12 |
| Nó Análise de grupo | 13 |
| Especificando dados para análise de grupo | 13 |
| Configurando opções de criação para análise de grupo | 14 |
| Visualizando estatísticas de análise de grupo | 14 |
| Saída para análise de grupo | 15 |

Capítulo 3. Análise de difusão **17**

| | |
|--|----|
| Visão geral da análise de difusão | 17 |
| Exemplo de Análise de difusão | 17 |
| Nó Análise de Difusão | 18 |
| Especificando dados para análise de difusão | 18 |
| Configurando opções de criação para análise de difusão | 19 |
| Visualizando estatísticas de análise de difusão | 20 |
| Saída para análise de difusão | 20 |

Avisos **21**

| | |
|-----------------------------|----|
| Marcas comerciais | 22 |
|-----------------------------|----|

Índice Remissivo **27**

Prefácio

O IBM® SPSS Modeler Social Network Analysis processa as informações sobre os relacionamentos entre pessoas em campos que descrevem o papel de um indivíduo em uma rede social, permitindo que as informações sociais sejam incluídas em modelos preditivos. Este manual descreve o uso dos nós do IBM SPSS Modeler Social Network Analysis no ambiente do IBM SPSS Modeler, permitindo incluir os nós em seus fluxos. Quando a saída do nó é combinada com campos que representam as medidas de indivíduos, resulta em um perfil mais completo dos indivíduos.

Sobre o IBM Business Analytics

O software IBM Business Analytics fornece informações completas, consistentes e exatas nas quais os tomadores de decisão confiam para melhorar o desempenho de negócios. Um portfólio abrangente de inteligência de negócios, análise preditiva, gerenciamento de desempenho e estratégia financeira e aplicativos analíticos fornece insights claros, imediatos e práticos sobre o desempenho atual e a capacidade de prever resultados futuros. Combinado com soluções para segmentos do mercado, práticas comprovadas e serviços profissionais completos, organizações de qualquer tamanho poderão conduzir maior produtividade, automatizar as decisões de modo confiável e entregar melhores resultados.

Como parte deste portfólio, o software IBM SPSS Predictive Analytics ajuda as organizações a preverem eventos futuros e agirem proativamente nesse insight para conduzir os melhores resultados de negócios. Clientes comerciais, governamentais e acadêmicos do mundo todo confiam na tecnologia IBM SPSS como uma vantagem competitiva para atrair, reter e aumentar clientes, enquanto reduz a fraude e minimiza riscos. Ao incorporar o software IBM SPSS em suas operações diárias, as organizações se tornam empresas preditivas, ou seja, capazes de direcionar e de automatizar as decisões para atender às metas de negócios e obter vantagem competitiva mensuráveis. Para obter mais informações ou entrar em contato com um representante, visite <http://www.ibm.com/spss>.

Suporte técnico

O suporte técnico está disponível para clientes de manutenção. Os clientes podem entrar em contato com o Suporte Técnico para obterem assistência com o uso de produtos IBM Corp. ou para obterem ajuda com a instalação de um dos ambientes de hardware suportados. Para entrar em contato com o Suporte Técnico, consulte o website do IBM Corp. em <http://www.ibm.com/support>. Esteja preparado para se identificar, identificar sua organização e sua concordância de suporte ao solicitar assistência.

Capítulo 1. IBM SPSS Modeler Social Network Analysis

Sobre o IBM SPSS Modeler Social Network Analysis

Muitas abordagens para o comportamento de modelagem concentram-se no indivíduo. Elas usam uma variedade de dados sobre os indivíduos para gerar um modelo que usa os principais indicadores do comportamento para prevê-lo. Se qualquer indivíduo possui valores para os principais indicadores que estão associados com a ocorrência do comportamento, esse indivíduo pode ser direcionado para atenção especial projetada para evitar o comportamento.

Considere as abordagens para a modelagem de migração para o concorrente, nas quais um cliente encerra seu relacionamento com uma empresa. O custo da retenção de clientes é significativamente menor do que o custo da substituição, tornando vital a capacidade de identificar clientes em risco de migrar para o concorrente. Um analista geralmente usa uma série de principais indicadores de desempenho para descrever os clientes, incluindo informações demográficas e padrões de chamada recente para cada cliente individual. Os modelos preditivos com base nesses campos usam as mudanças nos padrões de chamada do cliente, que são consistentes com os padrões de chamadas dos clientes que migraram para o concorrente no passado, para identificar as pessoas com um maior risco de migrar para o concorrente. Os clientes identificados como estando em risco recebem atendimento ao cliente adicional ou opções de serviço como um esforço para mantê-los.

Esses métodos negligenciam as informações sociais que podem afetar significativamente o comportamento de um cliente. As informações sobre uma empresa e sobre o que as outras pessoas estão fazendo fluem entre os relacionamentos para afetar as pessoas. Como resultado, os relacionamentos com outras pessoas permitem que essas pessoas influenciem as decisões e ações de uma pessoa. As análises que incluem apenas as medidas individuais estão omitindo fatores importantes que possuem recursos preditivos.

O IBM SPSS Modeler Social Network Analysis aborda esse problema ao processar as informações sobre relacionamentos em campos adicionais que podem ser incluídas em modelos. Estes principais indicadores de desempenho derivados medem as características sociais dos indivíduos. A combinação dessas propriedades sociais com medidas baseadas no indivíduo fornece uma visão geral melhor dos indivíduos e, conseqüentemente, pode melhorar a precisão preditiva de seus modelos.

O IBM SPSS Modeler Social Network Analysis consiste em dois componentes principais:

- Os nós do IBM SPSS Modeler Social Network Analysis incluídos no ambiente do IBM SPSS Modeler que permitem a inclusão de técnicas analíticas sociais em fluxos.
- O IBM SPSS Modeler Server Social Network Analysis, que inclui o processamento das especificações de nó para o IBM SPSS Modeler Server. O IBM SPSS Modeler Server Social Network Analysis processa eficientemente grandes quantias de dados de rede, que podem incluir milhões de indivíduos e relacionamentos, em um número relativamente pequeno de campos para análise adicional.

Por exemplo, o IBM SPSS Modeler Social Network Analysis identifica os indivíduos em uma rede que são mais afetados pela migração de pessoas específicas para o concorrente. Além disso, é possível descobrir grupos de indivíduos em uma rede que estão em um risco maior de migração. Ao incorporar os principais indicadores de desempenho para esses efeitos em seus modelos, é possível melhorar seu desempenho geral.

Sobre a análise de rede social

Uma rede social consiste de um conjunto de indivíduos e os relacionamentos entre eles. A análise de rede social examina esses relacionamentos para descrever os indivíduos e grupos como partes de uma estrutura social. Os indivíduos interagem uns com os outros e esses padrões de interação fornecem insight para as pessoas envolvidas. Os relacionamentos permitem que informações fluam através de uma rede, permitindo que um indivíduo influencie outro. A importância das informações de relacionamento configura análise de rede social além de outras abordagens. Em vez de concentrar-se em cada indivíduo separadamente, a unidade de estudo é uma díade composta de duas pessoas e seus relacionamentos.

Os relacionamentos em uma rede podem ser classificados como direcionais ou não direcionais. Em um *relacionamento direcional*, um indivíduo é identificado como o inicializador ou a origem do relacionamento, e a outra é identificado como o receptor ou destino. Por exemplo, fazer uma chamada telefônica é uma relação direcional no qual uma pessoa chama a outra. Em contraste, as funções de origem e destino não podem ser definidas para os *relacionamentos não direcionais*. Neste caso, ambas as partes participam no relacionamento igualmente. Falar um com o outro é um exemplo de um relacionamento não direcional.

Outra propriedade que se distingue entre relacionamentos é se o relacionamento é valorizado ou dicotômico. A única informação disponível em um *relacionamento dicotômico* é se o relacionamento existe entre dois indivíduos ou não. Para cada díade na rede, o relacionamento está presente ou ausente. Um *relacionamento valorizado*, por outro lado, inclui um peso indicando a intensidade do relacionamento. Os pesos permitem que os relacionamentos sejam comparados uns com os outros.

A tabela "Tipos de Relacionamento" lista exemplos para a classificação cruzada de relacionamentos por direção e escala. Nos relacionamentos direcionais, *Joe* é a origem do relacionamento e *Mary* é o destino. Nos relacionamentos não direcionais, não há indicação de quem iniciou o relacionamento. Os relacionamentos valorizados usam o comprimento da conversa como o peso do relacionamento, enquanto os relacionamentos dicotômicos, se elas ocorreram ou não.

Tabela 1. Tipos de relacionamento.

| Direção | Escala | exemplo |
|----------------|------------|---|
| Não direcional | Dicotômico | Joe e Mary falam um com o outro |
| Não direcional | Valorizado | Joe e Mary falam um com o outro por 20 minutos |
| Direcional | Dicotômico | Joe fez uma chamada para Mary |
| Direcional | Valorizado | Joe fez uma chamada de 20 minutos de conversa para Mary |

Para obter mais informações sobre o campo de análise de rede social, consulte um dos livros abrangente nesta área ¹.

Exibindo redes

Uma rede social geralmente é ilustrada usando um *sociograma* ². Neste tipo de exibição visual, indivíduos correspondem aos pontos, ou nós, em um espaço. Linhas, ou bordas, que conectam os pontos representam relacionamentos entre os indivíduos. Se os relacionamentos são direcionais, as bordas incluem uma seta para indicar a direção. Se os relacionamentos possuem pesos, os rótulos para as bordas denotam os valores. O gráfico a seguir exibe uma rede para sete pessoas.

1. Wasserman, S., e 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..

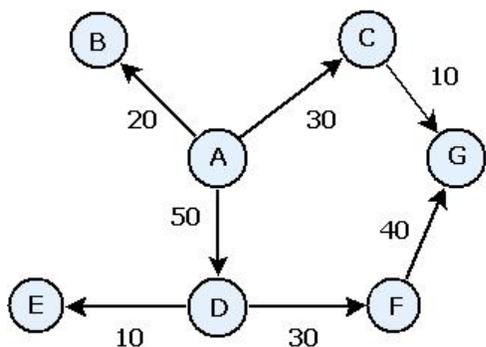


Figura 1. Rede social de exemplo

Suponha-se que a rede represente as chamadas telefônicas feitas por indivíduos, com os pesos de relacionamento indicando o comprimento das chamadas. Neste caso, a Pessoa A fez chamadas para três pessoas, gastando a maior parte do tempo falando com a pessoa D.

Esta rede é muito menor do que aquelas encontradas na prática. No entanto, os conceitos ilustrados por sociogramas simples são generalizados para redes de qualquer tamanho e complexidade.

Descrevendo redes

Informações sobre redes, grupos e indivíduos precisam ser extraídas em características descritivas que permitam comparações cruzadas e inclusão em modelos preditivos. As redes precisam ser extraídas para um conjunto finito dos principais indicadores de desempenho que podem ser analisados. Por exemplo, você pode desejar comparar redes ou grupos de nós em uma rede uns com os outros. Como alternativa, pode ser desejado comparar indivíduos na rede, uns com os outros, ou identificar os indivíduos mais importantes.

Duas medidas comumente usadas para descrever as redes sociais são **densidade** e **grau**. Ambas as estatísticas refletem a conectividade, mas a primeira se concentra na rede inteira ou em subgrupos de rede, enquanto a última caracteriza os indivíduos na rede.

Densidade de rede

Para qualquer conjunto de nós em uma rede, há um número finito de relacionamentos possíveis. Cada nó pode servir como a origem ou o destino de um relacionamento com todos os outros nós. Considere uma rede composta de três nós, A, B e C. A tabela a seguir lista todos os possíveis relacionamentos direcionados entre os nós.

Tabela 2. Possíveis relacionamentos direcionados para três nós.

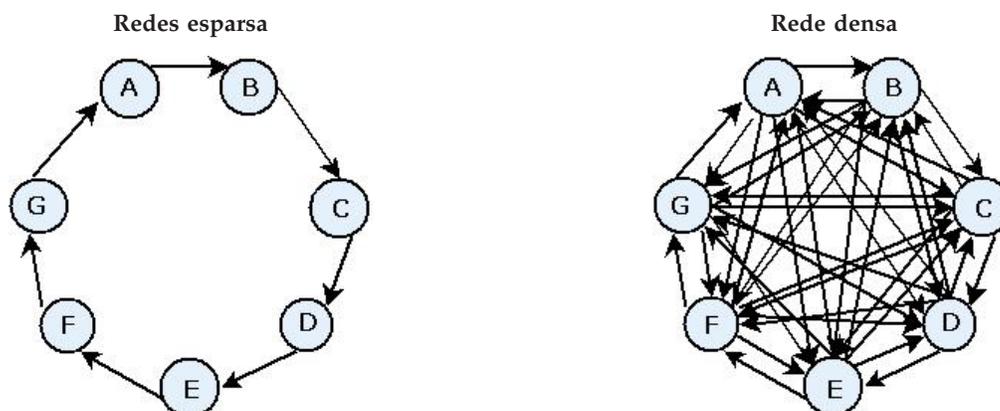
| Fonte | Resposta |
|-------|----------|
| A | B |
| A | C |
| B | A |
| B | C |
| C | A |
| C | B |

Cada nó é a origem de um relacionamento com os outros dois nós. No entanto, na prática, todos os relacionamentos possíveis não podem, na verdade, estar presentes. Alguns nós podem não ter nenhum relacionamento direto com os outros nós. Além disso, alguns relacionamentos direcionados podem não ser recíprocos.

A estatística *densidade* representa a proporção de possíveis relacionamentos na rede que estão realmente presentes. O valor varia de 0 a 1, com o limite inferior correspondendo às redes sem relacionamentos, e o limite superior representando as redes com todos os relacionamentos possíveis. Quanto mais próximo o valor for de 1, mais densa é a rede e mais coesos são os nós na rede.

As informações nas redes densas podem fluir mais facilmente do que as informações em redes esparsas. A tabela "Redes esparsas e densas" exibe duas redes compostas de sete nós. A rede esparsa inclui apenas sete dos 42 relacionamentos possíveis entre os nós, resultando em uma densidade de 0,17. A rede densa, por outro lado, contém todos os relacionamentos possíveis e possui uma densidade de 1,0.

Tabela 3. Redes esparsas e densas.



Na rede esparsa, para que as informações fluam do nó A para o G, ele deve passar pelos cinco outros nós. Em contraste, na rede densa, as informações podem ir diretamente do nó A para o nó G.

Grau do nó

Os indivíduos importante em uma rede são frequentemente aqueles envolvidos na maioria dos relacionamentos. Esses indivíduos adquirem informações de diversas origens e espalham essas informações para um grande número de outros indivíduos. Em contraste, os indivíduos que participam de poucos relacionamentos não podem influenciar diretamente um grande número de outros indivíduos na rede.

O *grau* para um nó, definido como o número total de relacionamentos que envolvem esse nó, permite comparações entre os participantes da rede. Indivíduos com valores superiores de grau mais altos são mais ativos do que aqueles com valores mais baixos. O grau ignora a direção dos relacionamentos, fornecendo uma medida geral da atividade para o nó.

Para relacionamentos direcionados, é possível se concentrar em se um nó é uma origem ou destino ao calcular o número de relacionamentos. O *indegree* para um nó é o número de relacionamentos nos quais um nó específico é o destino. De forma contrária, o *outdegree* é o número de relacionamentos nos quais um nó é a origem. A tabela a seguir lista os valores grau, indegree e outdegree para cada nó na figura "Exemplo de rede social".

Tabela 4. Exemplo de valores de grau, indegree e outdegree.

| Nó | Grau | Indegree | Outdegree |
|----|------|----------|-----------|
| A | 3 | 0 | 3 |
| B | 1 | 1 | 0 |
| C | 2 | 1 | 1 |
| D | 3 | 1 | 2 |
| E | 1 | 1 | 0 |

Tabela 4. Exemplo de valores de grau, indegree e outdegree (continuação).

| Nó | Grau | Indegree | Outdegree |
|----|------|----------|-----------|
| F | 2 | 1 | 1 |
| G | 2 | 2 | 0 |

Indegree é frequentemente tratado como uma medida de prestígio. Os valores indegree mais altos correspondem a mais relacionamentos terminando nesse nó. Em outras palavras, esses indivíduos são contatados por um número alto de outros indivíduos. Muitos outros nós estão iniciando relacionamentos com o nó. Por outro lado, outdegree é tratado como uma medida de centralidade. Os valores mais altos correspondem a mais relacionamentos que se originam desse nó. Esses indivíduos entram em contato com um número alto de outros indivíduos.

Para os nós na rede de exemplo, os valores de grau indicam que os nós A e D são os mais ativos, enquanto os nós B e E são os menos ativos. Os valores indegree revelam que o nó G possui o maior prestígio. Com base nos valores outdegree, o nó A é o mais central.

IBM SPSS Modeler Social Network Analysisnodes

Juntamente com a maioria dos nós padrão fornecidos com o IBM SPSS Modeler, também é possível trabalhar com os nós do IBM SPSS Modeler Social Network Analysis para incluir os resultados de análise de rede social em seus fluxos. A tabela Nós do "IBM SPSS Modeler Social Network Analysis" descreve estes nós, que estão armazenados na paleta Fontes.

Tabela 5. IBM SPSS Modeler Social Network Analysisnodes.

| Nó | Ícone | Descrição |
|--------------------|---|--|
| Análise de grupo |  | O nó Análise de Grupo importa os dados de registro de detalhe de chamada a partir de um arquivo de texto de campo fixo, identifica os grupos de nós na rede definida pelos registros e gera os principais indicadores de desempenho para os grupos e indivíduos na rede. Consulte o tópico "Visão geral de análise de grupo" na página 11 para obter mais informações. |
| Análise de difusão |  | O nó Análise de Difusão importa os dados de registro de detalhe de chamada a partir de um arquivo de texto de campo fixo, propaga um efeito em toda a rede definida pelos registros e gera os principais indicadores de desempenho que resumem os resultados do efeito em nós individuais. Consulte o tópico "Visão geral da análise de difusão" na página 17 para obter mais informações. |

Guias Nó

Os nós do IBM SPSS Modeler Social Network Analysis oferecem as seguintes guias para definir e visualizar a análise:

- **Guia Dados.** Usada para identificar o arquivo que contém as informações de rede social.
- **guia Opções de Criação.** Usada para definir as configurações para as análises.
- **guia Análise.** Usada para visualizar um resumo de saída preliminar, que fornece orientação para modificar as configurações de entrada na guia Dados para produzir melhores resultados.

Além disso, os nós oferecem as seguintes guias comuns entre os nós de origem em IBM SPSS Modeler:

- **guia Filtro.** Usada para eliminar ou renomear os campos de saída produzidos pelos nós. Este guia oferece a mesma funcionalidade que o nó Filtro.
- **guia Tipos.** Usada para definir níveis de medição para os campos de saída produzidos pelos nós. Esta guia oferece a mesma funcionalidade que o nó Tipo.

- **guia Anotações.** Usada para renomear nós, forneça um ToolTip customizado e armazene uma anotação longa.

Para obter mais informações sobre as guias comuns, consulte a documentação do IBM SPSS Modeler.

Visualizando saída

Geralmente, você usa a saída dos nós como entrada para um modelo preditivo. Se você deseja visualizar a saída, poderá incluir um nó de Tabela no fluxo. No entanto, a quantidade de dados geralmente é extremamente grande, tornando a geração de tabela demorada. Além disso, o número de linhas na tabela limita a sua utilidade.

Como alternativa, é possível gerar uma tabela de visualização que exibe uma amostra da saída que será criada. A visualização mostra os campos gerados para um número limitado de linhas. O número de linhas é definido nas propriedades do fluxo. Para visualizar a saída do nó, execute as seguintes etapas:

1. Abra o nó.
2. Especifique as configurações de dados na guia Dados.
3. Defina os parâmetros analíticos na guia Opções de Criação.
4. Clique em **Visualizar**.

Uma janela de visualização é aberta, exibindo os resultados. Além disso, visualizar a saída preenche a guia Análise do nó com uma visão geral resumida dos resultados.

Analisando dados

Determinar as configurações analíticas que produzem resultados ideais é geralmente um processo iterativo. Defina as configurações, execute a análise e revise os resultados. Se os resultados não forem tão úteis quanto poderiam ser, modifique as configurações e execute novamente a análise.

Para analisar os dados de entrada para o nó, execute as seguintes etapas:

1. Abra o nó.
2. Especifique as configurações de dados na guia Dados.
3. Defina os parâmetros analíticos na guia Opções de Criação.
4. Clique em **Analisar de Dados**.

Se as opções de criação indicam que estatísticas básicas devem ser exibidas, a guia Análise mostrará os resultados.

Se for necessário executar novamente uma análise, clique em **Limpar análise** para limpar os resultados atuais antes de clicar em **Analisar dados**.

Aplicativos

Os aplicativos específicos nos quais o IBM SPSS Modeler Social Network Analysis pode ser particularmente útil, incluem o seguinte:

- **Predição de migração para o concorrente.** Características de grupo podem influenciar as taxas de migração. Ao concentrar em indivíduos em grupos que estão em um risco de migração maior, será possível evitar isso. Além disso, é possível identificar os indivíduos que estão em risco de migração devido ao fluxo de informações a partir daqueles que já migraram.
- **Alavancando líderes de grupo.** Os líderes de grupo são altamente influentes sobre outros membros do grupo. Se um líder de grupo pode ser impedido de migrar, a taxa de migração para os membros do grupo pode ser reduzida. Como alternativa, tentar fazer com que um líder de grupo migre do concorrente pode aumentar a taxa de migração de membros do grupo associado a esse concorrente, reduzindo a taxa de migração de membros do grupo associado à sua empresa.

- **Marketing.** Os líderes de grupo podem ser usados para iniciar novas ofertas de mercadorias e serviços. A influência do líder pode tornar mais provável que outros membros do grupo adquiram a oferta. É possível usar a análise de difusão para identificar os indivíduos mais afetados pelos líderes de grupo e direcionar seu marketing para eles.

Dois fluxos de demonstração são fornecidos com o IBM SPSS Modeler Social Network Analysis para fornecer exemplos de como é possível incluir os resultados da análise de rede social em seus fluxos. Os arquivos de dados e os fluxos de amostra são instalados na pasta *Demos* no diretório de instalação do produto.

- *DA demo streams.str* fornece um exemplo de análise de dados para identificar os 300 principais indivíduos que têm maior probabilidade de migrar para o concorrente.
- *GA demo streams.str* fornece um exemplo do uso dos principais indicadores de desempenho (KPIs) para prever a rotatividade de ambos os grupos e indivíduos, e também como usar os KPIs para atingir indivíduos específicos para uma campanha de marketing.

Estrutura de dados

Informações sobre os indivíduos na rede de interesse podem ser distribuídas entre uma variedade de arquivos, bancos de dados e sistemas em toda a sua empresa. Para analisar a rede usando o IBM SPSS Modeler Social Network Analysis, é necessário extrair os registros e campos relevantes de suas origens de dados e formatá-los para a entrada para os nós.

Os nós analíticos requerem registros de detalhes de chamada armazenados em um único arquivo de texto de largura fixa. Cada linha do arquivo corresponde a um relacionamento, com os dados organizados nas seguintes colunas:

- o identificador para o indivíduo que inicia o relacionamento.
- o identificador para o indivíduo que é o destino do relacionamento.
- um peso opcional para o relacionamento.

Todos os dados devem ser numéricos, com os identificadores para indivíduos limitados para números inteiros. A inclusão de nomes de campos como a primeira linha de dados no arquivo é opcional. A tabela "Exemplo de registros de dados da chamada" ilustra esta estrutura de dados.

Observe que os registros de cabeçalho também devem usar números ou valores em branco. Por exemplo, um cabeçalho com o valor de *EF BB BF* pode causar um erro.

Tabela 6. Exemplo de registros de dados de chamadas.

| Fonte | Destino | Peso |
|-------|---------|------|
| 1000 | 5642 | 243 |
| 2190 | 8444 | 831 |
| 0299 | 9419 | 559 |

Os valores de peso podem corresponder a qualquer medida que você deseja usar para representar a importância do relacionamento relativo para os outros relacionamentos na rede. Para dados de chamada, os pesos comuns incluem a duração de chamada ou a frequência de chamada. Observe que isso é verdadeiro para a Análise de Difusão, mas a Análise de Grupo suporta apenas a frequência de chamada.

Se você deseja que a análise mantenha o foco em um subconjunto do histórico de chamada, deverá usar esse subconjunto ao criar o arquivo de entrada. Por exemplo, é possível limitar a análise para vários meses no passado ou para as chamadas mais recentes para um indivíduo, incluindo apenas estes dados no arquivo de texto de entrada.

Propriedades de script

A criação de script no IBM SPSS Modeler Social Network Analysis é uma ferramenta poderosa para automatizar processos na interface com o usuário. Os scripts podem executar os mesmos tipos de ações que podem ser executadas com o mouse ou teclado e podem ser utilizados para automatizar tarefas que seriam altamente repetitivas ou que demorariam muito tempo para serem executadas manualmente. Para uma explicação sobre como usar o script, consulte o guia *ScriptingAutomation.pdf*, disponível com o IBM SPSS Modeler.

Propriedades do nó de difusão

A tabela a seguir lista as propriedades de script para o nó Difusão.

| Nome da propriedade | Tipo de dados | Descrição da propriedade |
|--------------------------|----------------|---|
| input_data_file_name | <i>string</i> | |
| calling_field | <i>campo</i> | |
| called_field | <i>campo</i> | |
| frequency_weight_field | <i>campo</i> | |
| read_field_names | <i>boolean</i> | |
| diffusion_list_file_name | <i>string</i> | |
| spreading_factor | <i>double</i> | Valor padrão= 50. Mín = 1 Máx = 99. |
| max_number_iterations | <i>integer</i> | Valor padrão = 100. Mín = 1. |
| accuracy_threshold | <i>double</i> | Valor padrão = 0.01. Mín = 0.001. |
| calculate_statistics | <i>boolean</i> | |

Propriedades do nó Análise de Grupo

A tabela a seguir lista as propriedades de script para o nó Análise de Grupo.

| Nome da propriedade | Tipo de dados | Descrição da propriedade |
|------------------------|----------------|--|
| input_data_file_name | <i>string</i> | |
| calling_field | <i>campo</i> | |
| called_field | <i>campo</i> | |
| frequency_weight_field | <i>campo</i> | |
| read_field_names | <i>boolean</i> | |
| coverage_threshold | <i>double</i> | Valor padrão = 10. Mín = 1 Máx = 99. |

| Nome da propriedade | Tipo de dados | Descrição da propriedade |
|----------------------------|----------------------|---------------------------------|
| min_group_size | <i>integer</i> | Valor padrão = 2. Mín = 2. |
| max_group_size | <i>integer</i> | Valor padrão = 100. Mín = 2. |
| calculate_statistics | <i>boolean</i> | |

Capítulo 2. Análise de grupo

Visão geral de análise de grupo

A Análise de Grupo usa os padrões de interação de indivíduos em uma rede para identificar grupos de indivíduos semelhantes. As características desses grupos influenciam o comportamento dos membros individuais do grupo. Por exemplo, grupos pequenos que têm vários relacionamentos entre membros e fortes líderes têm um risco maior de migração para a concorrência, mesmo que nenhum membro do grupo tenha migrado realmente.³ Modelos preditivos que incorporam as medidas de grupo e indivíduo serão melhor executados do que os modelos que incluem apenas o último.

A Análise de Grupo consiste nas etapas gerais a seguir:

1. Determine a intensidade dos relacionamentos que refletem proximidade social. Consulte o tópico “Determinando a similaridade social” para obter mais informações.
2. Particione a rede em grupos com base na intensidade do relacionamento enquanto obedece às restrições de tamanho. Consulte o tópico “Particionamento em grupos” na página 12 para obter mais informações.
3. Trace o perfil de grupos e indivíduos, incluindo a identificação de líderes de grupo. Consulte o tópico “Descrevendo grupos e membros do grupo” na página 12 para obter mais informações.

Determinando a similaridade social

Os membros de um grupo devem ser mais semelhantes entre si do que são para pessoas que não estão no grupo. Na análise de rede, a semelhança de dois nós depende dos seus relacionamentos. Para qualquer nó em uma rede, há um conjunto de nós que são os destinos de relacionamentos direcionados com o nó. Para dados de telecomunicações, este conjunto corresponde a todas as pessoas contatadas por um indivíduo específico. Se dois indivíduos entram em contato com o mesmo conjunto de pessoas, esses indivíduos são considerados semelhantes entre si. Quanto mais os conjuntos de destinos de relacionamento se sobrepõem para dois indivíduos, mais semelhantes eles serão.

Considere a rede mostrada na figura "Rede de dez nós de exemplo".

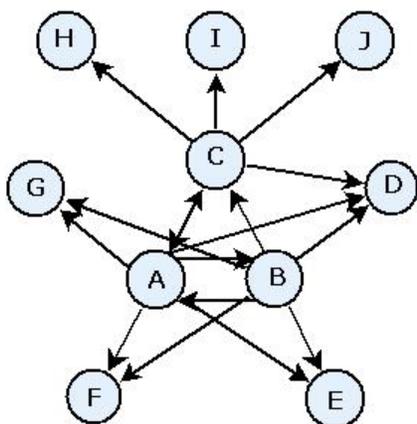


Figura 2. Rede de dez nós de exemplo

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

A tabela "Nós de destino" mostra os nós de destino para relacionamentos que se originam nos nós A, B e C.

Tabela 7. Nós de destino.

| Nó de origem | Nós de destino |
|--------------|------------------|
| A | B, C, D, E, F, G |
| B | A, C, D, E, F, G |
| C | A, D, H, I, J |

Os Nós A e B possuem cinco nós de destino em comum. Os Nós A e C, por outro lado, possuem apenas um. Assim, os nós A e B são socialmente mais semelhantes que os nós A e C.

Os pesos dos relacionamentos, como duração de chamada ou frequência, não capturam a similaridade de nós em uma rede. Para incluir a noção de semelhança, a análise de grupo usa informações mútuas⁴ como o peso do relacionamento. Essa estatística reflete a probabilidade de que dois nós estejam conectados ao mesmo nó. Se os relacionamentos na rede possuem pesos predefinidos, como frequência de chamada para os dados de telecomunicações, as informações mútuas incorporam estes pesos em conformidade.

Particionamento em grupos

Os grupos devem consistir em indivíduos que possuem pesos de relacionamento altos com outros membros do grupo, em que os pesos medem a similaridade dos nós no relacionamento. Como resultado, a identificação do grupo começa omitindo os relacionamentos mais fracos em uma rede. O *limite de cobertura* controla esse processo definindo a fração dos relacionamentos mais fortes para reter. Por exemplo, um valor do limite de cobertura de 0,4 resulta em 40% de relacionamentos mais fortes sendo usados para a identificação do grupo, com os restantes 60% dos relacionamentos sendo omitidos.

Os relacionamentos restantes podem gerar alguns grupos muito pequenos ou muito grandes, que têm utilitário de previsão limitado. Para impedir que esses grupos sejam incluídos na análise, é possível especificar tamanhos mínimo e máximo do grupo. Grupos que possuem o tamanho menor que o mínimo são omitidos completamente. Grupos com tamanhos maiores do que o máximo, no entanto, são divididos em grupos menores dentro do intervalo de tamanho aceitável. Os grupos restantes, após eliminar os relacionamentos fracos e impingir limites de tamanho, são referidos como *grupos principais*.

Remover relacionamentos da rede original pode fazer com que alguns indivíduos não estejam em nenhum grupo principal. No entanto, esses indivíduos podem ter conexões com membros de um grupo que garantem ser incluídos no grupo. Um indivíduo será incluído em um grupo se houver muitos relacionamentos relativamente forte com os membros principais do grupo, desde que os limites de tamanho do grupo não sejam violados. Assim, os grupos finais consistem de um conjunto de membros principais mais membros incluídos devido à conectividade com o núcleo.

Descrevendo grupos e membros do grupo

Além da densidade, indegree e outdegree, outras estatísticas descrevem a dinâmica de grupo. Em particular, os escores de autoridade e disseminação oferecem medidas do status social dos indivíduos dentro dos grupos. O papel de cada indivíduo em um grupo é de importância vital quando se tenta prever o comportamento do grupo e seus membros.

O *escore de autoridade* para um nó mede a tendência de outros nós no grupo se conectarem a ele. Se muitos indivíduos estão em contato uma pessoa em particular, potencialmente pedindo informações ou opiniões, essa pessoa tem o papel de uma autoridade. Os escores de autoridade para os nós no grupo, que correspondem às probabilidades estacionárias para um walk aleatório com reinicializações por meio da rede do grupo, variam de 0 a 1. Quanto mais próximo o escore de autoridade estiver de 1, mais

4. Cover, T. M., e J. A. Thomas. 2006. *Elements of Information Theory*, 2ª edição. Nova York: John Wiley and Sons, Inc.

autoridade este tem dentro do grupo. O nó no grupo com escore de autoridade mais alto é referido como o *líder de autoridade* para o grupo. Dividir maior escore dentro de um grupo pelo menor fornece uma medida da intensidade geral do líder de autoridade.

Em contraste, o *escore de disseminação* para um nó mede a tendência do nó de se conectar a outros nós no grupo. Se uma pessoa específica entra em contato com muitas outras pessoas no grupo, essa pessoa poderá afetar significativamente os pareceres do grupo inteiro. Os scores de disseminação dos nós no grupo, que correspondem às probabilidades estacionárias para um walk aleatório com reinicializações por meio da rede do grupo em inverso, variam de 0 a 1. Quanto mais próximo o escore de disseminação estiver de 1, mais o nó se conecta aos outros membros do grupo. O nó no grupo com o maior escore de disseminação é referido como o *líder de disseminação* para o grupo. Dividir o maior escore dentro de um grupo pelo menor fornece uma medida da intensidade geral do líder de disseminação.

Nó Análise de grupo

O nó Análise de grupo, que está disponível na paleta Fontes, identifica conjuntos de indivíduos em uma rede que são socialmente semelhantes uns com os outros e determina o estatuto social relativo dos indivíduos nos grupos. Grupos que são relativamente pequenos têm muitas conexões entre os membros, e ao incluir um líder forte, estão em um risco maior de migração devido à dinâmica de grupo. É possível usar os resultados para determinar escores de migração de grupo que, quando combinado com escores de migração individuais, melhoram a capacidade de prever a migração em relação aos modelos com base em escores individuais sozinhos.

Dada a quantidade de dados que normalmente inclui a rede e a natureza das interações sociais subjacentes às definições de grupo, a análise de grupo é um processo muito demorado. Geralmente, será possível salvar os resultados de análise de grupo em um banco de dados ou arquivo a ser usado como entrada para um modelo preditivo. Os resultados do grupo serão atualizados relativamente com pouca frequência, como mensalmente, enquanto o modelo pode ser atualizado com muito mais frequência.

Requisitos. O nó requer um arquivo de texto de largura fixa, que define a rede social usando três campos. Um campo identifica a origem para cada relacionamento direcionado, outro campo define o destino para cada relacionamento direcionado, e o terceiro campo especifica uma intensidade opcional para cada relacionamento. Todos os relacionamentos na rede devem ser direcionais.

Especificando dados para análise de grupo

A guia Dados da janela do nó de origem Análise de Grupo permite especificar o arquivo de entrada que contém os relacionamentos de nó de rede.

Arquivo. Especifique o nome de um arquivo ou pasta que contém os registros de detalhes de chamada. É possível inserir um nome ou clicar no botão de reticências (...) para selecionar um nome no sistema de arquivos. O caminho é mostrado depois que um nome tiver sido selecionado, e seu conteúdo é exibido com delimitadores. Se uma pasta for especificada, os registros de detalhes de chamada em todos os arquivos contidos na pasta serão concatenados para a análise; todos os arquivos na pasta devem ter a mesma estrutura.

Ler os nomes do campo no arquivo. Selecionada por padrão, esta opção trata a primeira linha no arquivo de dados como os nomes para as colunas. Se a sua primeira linha não for um cabeçalho, desmarque essa opção para fornecer automaticamente a cada campo um nome genérico, como *Campo1* e *Campo2*.

Configurações de definição de rede

As configurações de definição de rede definem os papéis para os campos.

Campos. Use os botões de seta para designar manualmente itens dessa lista para os vários campos de papel. Os ícones indicam os níveis de medição válidos para cada campo de papel. Clique no botão **Todos** para selecionar todos os campos da lista, ou clique em um botão de nível de medição individual para selecionar todos os campos com esse nível de medição.

Origem. Selecione um campo como a origem do relacionamento direcionado.

Destino. Selecione um campo como o destino do relacionamento direcionado.

Ponderação. Opcionalmente, selecione um campo correspondente ao número de vezes que o relacionamento ocorre na rede. A análise pondera os registros conforme a determinação da intensidade do relacionamento.

Configurando opções de criação para análise de grupo

A guia Opções de Criação da janela Nó de origem do Grupo de Análise permite definir opções para identificar grupos na rede.

Configurações de Análise de Grupo

As configurações de análise de grupo influenciam o tamanho e a intensidade relativa dos grupos formados.

Limite de cobertura. Defina a fração dos relacionamentos de rede mais fortes para usar na análise. Por exemplo, um limite de cobertura de 0,2 indica que apenas os relacionamentos cujos pesos estiverem entre os principais 20% de todos os pesos serão usados. Este parâmetro, que varia de 0 a 1, permite que a análise se concentre nos relacionamentos mais fortes, que devem ocorrer dentro de grupos. Lucro com valores de limite mais altos agrupa relacionamentos mais fracos entre seus membros. Consulte o tópico "Particionamento em grupos" na página 12 para obter mais informações.

Tamanho mínimo do grupo. Especifique um limite inferior para o tamanho do grupo. Grupos menores que este valor não serão retornados.

Tamanho máximo do grupo. Especifique um limite superior para o tamanho do grupo. Grupos maiores que esse valor serão divididos em grupos menores.

Calcular e exibir estatísticas básicas. Se selecionado, o nó calcula e exibe estatísticas básicas para a análise, além da derivação da saída do principal indicador de desempenho. O cálculo dessas estatísticas pode afetar negativamente o desempenho do nó para redes muito grandes.

Visualizando estatísticas de análise de grupo

A guia Análise da janela de nó de origem Análise de Grupo fornece uma visão geral resumida dos grupos identificados na rede. A tabela "estatísticas básicas da Análise de Grupo" lista as estatísticas básicas disponíveis para os grupos.

Tabela 8. Estatísticas básicas da Análise de Grupo.

| Estatística | Descrição |
|--------------------------|--|
| Total de nós em grupos | Número de nós incluídos nos grupos identificados |
| Total de links em grupos | Número de links incluídos nos grupos identificados |
| Número total de grupos | Número de grupos identificados na rede |
| Tamanho médio do grupo | Número médio de nós em um grupo |
| Densidade média do grupo | Fração média de conexões diretas entre nós em um grupo. Consulte o tópico "Densidade de rede" na página 3 para obter mais informações. |

Tabela 8. Estatísticas básicas da Análise de Grupo (continuação).

| Estatística | Descrição |
|------------------------------------|--|
| Fração média de membros principais | Fração média de nós em um grupo que são nós principais para o grupo. Consulte o tópico "Particionamento em grupos" na página 12 para obter mais informações. |
| Densidade média do grupo principal | Fração média de conexões diretas entre nós principais em um grupo |
| Média InDegree | Número médio de links de entrada. Consulte o tópico "Grau do nó" na página 4 para obter mais informações. |
| Média OutDegree | Número médio de links de saída. Consulte o tópico "Grau do nó" na página 4 para obter mais informações. |

Selecione uma estatística específica da tabela Estatísticas Básicas para visualizar a distribuição de valores, o desvio padrão e a assimetria para a estatística.

Revisar estas estatísticas auxilia na definição das configurações de análise de grupo. Idealmente, os resultados deverão mostrar grupos relativamente pequenos que tenham valores de densidade altos. Por exemplo, se houver alguns grupos muito grandes, considere reduzir o tamanho máximo do grupo para dividi-los em grupos menores. Como alternativa, se os valores de densidade do grupo tendem a ser pequenos, considere a redução do limite de cobertura para se concentrar em relacionamentos mais fortes.

Saída para análise de grupo

O nó Análise de Grupo gera uma variedade de campos que descrevem os grupos e os indivíduos nos grupos. É possível aumentar os modelos existentes e dados com esses principais indicadores de desempenho para melhorar as previsões geradas por seus modelos. Por exemplo, é possível atualizar os valores de propensão de migração individuais para o concorrente para incluir as influências do grupo.

A tabela "Principais indicadores de desempenho para grupos" lista os campos dos principais indicadores de desempenho para os grupos identificados na análise.

Tabela 9. Principais indicadores de desempenho para grupos.

| Campo | Descrição |
|--------------------------|--|
| GAG_GroupNumber | Identificador exclusivo para um grupo |
| GAG_Size | Número de indivíduos em um grupo |
| GAG_Density | Fração de conexões diretas entre os indivíduos em um grupo. Consulte o tópico "Densidade de rede" na página 3 para obter mais informações. |
| GAG_KernelDensity | Fração de conexões diretas entre indivíduos principais em um grupo |
| GAG_CoreNodesFraction | Fração de indivíduos em um grupo que são indivíduos principais para o grupo. Consulte o tópico "Particionamento em grupos" na página 12 para obter mais informações. |
| GAG_MaxRankType1 | Escore máximo de autoridade de qualquer membro do grupo. Consulte o tópico "Descrevendo grupos e membros do grupo" na página 12 para obter mais informações. |
| GAG_MinRankType1 | Escore mínimo de autoridade de qualquer membro do grupo. |
| GAG_MaxMinRankRatioType1 | Proporção do maior escore de autoridade para o menor. Esse valor reflete a intensidade de autoridade do líder do grupo. |
| GAG_MaxRankType2 | Escore máximo de disseminação de qualquer membro do grupo. Consulte o tópico "Descrevendo grupos e membros do grupo" na página 12 para obter mais informações. |
| GAG_MinRankType2 | Escore mínimo de disseminação de qualquer membro do grupo. |

Tabela 9. Principais indicadores de desempenho para grupos (continuação).

| Campo | Descrição |
|--------------------------|---|
| GAG_MaxMinRankRatioType2 | Proporção do maior escore de disseminação para o menor. Esse valor reflete a intensidade de disseminação do líder do grupo. |

A tabela "Principais indicadores de desempenho para indivíduos" lista os campos dos principais indicadores de desempenho para os indivíduos na rede.

Tabela 10. Principais indicadores de desempenho para indivíduos.

| Campo | Descrição |
|--------------------------------|--|
| GAI_NodeNumber | Identificador exclusivo para um indivíduo |
| GAI_CoreNode | Indicador que determina se o indivíduo é um indivíduo principal para um grupo ou não. Consulte o tópico "Particionamento em grupos" na página 12 para obter mais informações. |
| GAI_RankType1 | Escore da autoridade para o indivíduo. Consulte o tópico "Descrevendo grupos e membros do grupo" na página 12 para obter mais informações. |
| GAI_RankOrderType1 | Ordem de ranqueamento no grupo com base nos escores de autoridade |
| GAI_RankType2 | Escore de disseminação para o indivíduo. Consulte o tópico "Descrevendo grupos e membros do grupo" na página 12 para obter mais informações. |
| GAI_RankOrderType2 | Ordem de ranqueamento no grupo com base nos escores de disseminação |
| GAI_InDegree | Número de relacionamentos nos quais o indivíduo é o destino do relacionamento. Consulte o tópico "Grau do nó" na página 4 para obter mais informações. |
| GAI_OutDegree | Número de relacionamentos nos quais o indivíduo é a origem do relacionamento. Consulte o tópico "Grau do nó" na página 4 para obter mais informações. |
| GAI_GroupLeaderType1 | Se o nó é um líder de autoridade, cujo escore de liderança é derivado de links de entrada. Consulte o tópico "Descrevendo grupos e membros do grupo" na página 12 para obter mais informações. |
| GAI_GroupLeaderConfidenceType1 | A confiança de que o nó é um líder de autoridade. |
| GAI_GroupLeaderType2 | Se o nó é um líder de disseminação, cujo escore de liderança é derivado de links de saída. Consulte o tópico "Descrevendo grupos e membros do grupo" na página 12 para obter mais informações. |
| GAI_GroupLeaderConfidenceType2 | A confiança de que o nó é um líder de disseminação. |

Capítulo 3. Análise de difusão

Visão geral da análise de difusão

A Análise de Difusão identifica indivíduos que são mais afetados por outros indivíduos em uma rede social, quantificando o efeito como *energia difundida*. O processo usa uma abordagem de ativação de difusão, no qual um efeito espalha-se de forma iterativa a partir de nós de rede para seus vizinhos imediatos, diminuindo em tamanho à medida em que é movido de nó para nó⁵. Após o recebimento de energia, um nó é ativado e transmite uma parte dessa energia para qualquer vizinho que for o destino de relacionamentos direcionado com o nó.

O *fator de difusão* define a porcentagem de energia transmitida por um nó ativado, com a quantia restante retida pelo nó. Todos os nós que recebem esta energia, irão transmitir-se com a mesma proporção a seus vizinhos, resultando em um processo de decomposição para transmissão de energia. Grande fatores de difusão correspondem a mais energia sendo enviada, permitindo que a energia alcance os nós mais distantes dos nós iniciais ativados antes de o processo se decompor completamente. Pequenos fatores de difusão resultam em processos de difusão que se decompõem rapidamente, com a energia transmitida permanecendo relativamente perto dos nós iniciais.

A quantia total de energia que está sendo difundida por um nó ativado é distribuída entre todos os nós que são respostas de relacionamentos direcionados com o nó. A quantia que cada nó recebe depende da intensidade do relacionamento com o nó ativado. A fração enviada para um nó específico é igual ao peso do relacionamento, dividido pelo total de pesos de todos os relacionamentos em que o nó ativado é a origem. Consequentemente, os vizinhos que têm relacionamentos com pesos relativos maiores recebem mais energia do que os vizinhos com pesos relativos menores.

O processo de difusão para quando uma das seguintes condições ocorre:

- os nós ativados não são a origem de qualquer um dos relacionamentos direcionados
- a quantidade de energia que está sendo transmitida é abaixo do *limite de precisão*, um limite sobre a quantidade de energia que está sendo transferida para o processo continuar
- o número de iterações atinge um limite especificado

Quando a difusão é concluída, os nós que possuem a energia mais difundida são os mais sensíveis ao efeito que iniciou o processo. Por exemplo, se o processo começa nos nós que migram para o concorrente, os nós com a energia mais alta são os que têm maior risco de migrarem para o concorrente. É possível fornecer uma atenção especial a esses nós para impedir que eles migrem para o concorrente.

Exemplo de Análise de difusão

Considere a rede mostrada "Exibindo redes" na página 2. A rede é composta de sete nós que têm relacionamentos direcionados de intensidades que variam uns com os outros.

A tabela "Exemplo do processo de difusão" ilustra um processo de difusão na rede usando um fator de propagação de 0,80. Inicialmente, o nó A contém toda a energia, arbitrariamente atribuída a um valor de 1,00. Na etapa 1, esse nó ativa, espalhando 80% da sua energia para os três vizinhos que são destinos de relacionamentos, enquanto mantém 20% para si mesmo. Os relacionamentos possuem um peso total de 100. O relacionamento com o nó D responde por metade deste total, então este nó recebe metade da

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

energia difundida, ou 0,40. O relacionamento com o nó B é responsável por 20% do peso total, gerando um valor de energia difundida de 0,16. O nó C recebe a energia restante, 0,24, que é 30% da quantia que origina do nó A.

Tabela 11. Exemplo do processo de difusão.

| Step | Nó A | Nó B | Nó C | Nó D | Nó E | Nó F | Nó 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 |

Tendo recebido energia na etapa 1, os nós B, C e D são ativados na etapa 2 e difundem energia para qualquer vizinho de destino. O nó B não possui nós de destino, por isso ele não difunde energia. O nó C, por outro lado, possui um destino e difunde 80% da sua energia para esse nó, o G. Finalmente, o nó D difunde 80% da sua energia para os nós E e F, com o F recebendo três vezes a quantia do E devido ao seu relacionamento mais forte com o D.

Na etapa 3, os nós E, F e G são ativados. Os nós E e G não possuem nós de destino, por isso não difundem energia. Entretanto, o nó F difunde 80% da sua energia para seu único nó de destino, o G. Este nós agora tem um valor de energia total de 0,38, que consiste em energia difundida a partir dos nós C e F.

Neste ponto, o nó G não é ativado, mas não possui nós de destino para difusão, por isso o processo termina. A energia que se origina no nó A espalhou-se por toda a rede, com o nó G recebendo a maior quantidade. Se a energia introduzida na rede representar migração para o concorrente, o nó G seria o mais afetado pela migração do nó A.

Nó Análise de Difusão

O nó Análise de Difusão, que está disponível a partir da paleta Origens, propagará um efeito a partir de um conjunto especificado de indivíduos em uma rede social, usando os relacionamentos de rede para identificar os indivíduos mais afetados pelo efeito. Se o efeito for migrar para o concorrente, por exemplo, o nó identifica os indivíduos com maior probabilidade de migração devido a outros indivíduos específicos na migração de rede. É possível aumentar os modelos existentes e os dados com a saída do nó para melhorar as previsões geradas por esses modelos. Por exemplo, é possível atualizar os valores de propensão de migração individuais para o concorrente para incluir influências de difusão.

Requisitos. O nó requer dois arquivos de texto de largura fixa que contenham os dados a serem analisados. O primeiro arquivo define a rede social usando três campos. Um campo identifica a origem para cada relacionamento direcionado, outro campo define o destino para cada relacionamento direcionado, e o terceiro campo especifica uma intensidade opcional para cada relacionamento. Todos os relacionamentos na rede devem ser direcionais. O segundo arquivo contém uma lista de identificadores a partir dos quais o efeito deve começar.

Especificando dados para análise de difusão

A guia Dados da janela de nó de origem Análise de Difusão permite especificar o arquivo de entrada que contém os relacionamentos de nó de rede.

Arquivo. Especifique o nome de um arquivo ou pasta que contém os registros de detalhes de chamada. É possível inserir um nome ou clicar no botão de reticências (...) para selecionar um nome no sistema de arquivos. O caminho é mostrado depois que um nome tiver sido selecionado, e seu conteúdo é exibido

com delimitadores. Se uma pasta for especificada, os registros de detalhes de chamada em todos os arquivos contidos na pasta serão concatenados para a análise; todos os arquivos na pasta devem ter a mesma estrutura.

Ler os nomes do campo no arquivo. Seleccionada por padrão, esta opção trata a primeira linha no arquivo de dados como os nomes para as colunas. Se a sua primeira linha não for um cabeçalho, desmarque essa opção para fornecer automaticamente a cada campo um nome genérico, como *Campo1* e *Campo2*.

Configurações de definição de rede

As configurações de definição de rede definem os papéis para os campos.

Campos. Use os botões de seta para designar manualmente itens dessa lista para os vários campos de papel. Os ícones indicam os níveis de medição válidos para cada campo de papel. Clique no botão **Todos** para seleccionar todos os campos da lista, ou clique em um botão de nível de medição individual para seleccionar todos os campos com esse nível de medição.

Origem. Selecione um campo como a origem do relacionamento direcionado.

Destino. Selecione um campo como o destino do relacionamento direcionado.

Ponderação. Selecione um campo que representa o peso do relacionamento. Para registros de detalhe de chamada, o peso pode corresponder a qualquer uma da série de características de chamada que refletem a intensidade do relacionamento, como duração de chamada ou frequência de chamada.

Configurando opções de criação para análise de difusão

A guia Opções de Criação da janela de nó de origem Análise de Difusão permite definir as opções para o processo de difusão.

Lista de valores iniciais de difusão

A lista de valores iniciais de difusão identifica os nós de rede a partir dos quais a energia deve iniciar a difusão. Por exemplo, em análise de migração para o concorrente, esta lista identifica os nós de rede que já migraram. Como alternativa, essa lista pode conter nós que apresentam maior risco de migração para explorar os efeitos sobre a rede que esses nós devem realmente migrar.

Arquivo. Especifique o nome do arquivo de texto que contém os identificadores de nós, com cada identificador correspondendo a uma linha no arquivo. É possível inserir um nome de arquivo ou clicar no botão de reticências (...) para seleccionar um arquivo.

Configurações de Análise de difusão

As configurações de análise de difusão determinam quão agressiva a difusão está e quando o processo parará.

Fator de difusão. Defina a fração de energia que um nó transfere em cada etapa do processo de difusão. Valores menores limitam a energia aos nós mais próximos do nó que está difundindo a energia. Por outro lado, valores maiores permitem que a energia afete os nós mais distantes.

Máximo de Iterações. Especifique um limite superior para o número de iterações de difusão. O processo de difusão é encerrado automaticamente se esse limite for atingido.

Limite de precisão. Especifique a menor mudança na energia do nó que garante a continuação do processo de difusão. Se a mudança na energia para cada nó for menor do que este valor, o processo de difusão terminará.

Calcular e exibir estatísticas básicas. Se selecionado, o nó calcula e exibe estatísticas básicas para a análise, além da derivação da saída do principal indicador de desempenho. O cálculo dessas estatísticas pode afetar negativamente o desempenho do nó para redes muito grandes.

Visualizando estatísticas de análise de difusão

A guia Análise da janela de nó de origem de Análise de Difusão fornece uma visão geral resumida dos resultados de difusão. A tabela "Estatísticas básicas da Análise de Difusão" lista as estatísticas básicas disponíveis.

Tabela 12. Estatísticas básicas da Análise de Difusão.

| Estatística | Descrição |
|--|---|
| Total de nós na rede | Número de nós na rede |
| Total de links na rede | Número de links na rede |
| Total de valores iniciais de difusão na rede | Número de nós usados como valores iniciais para o processo de difusão |
| Média de influência | Quantia média de energia difundida associada a indivíduos. |
| Média InDegree | Número médio de relacionamentos nos quais um indivíduo é o destino do relacionamento. Consulte o tópico "Grau do nó" na página 4 para obter mais informações. |
| Média OutDegree | Número médio de relacionamentos nos quais um indivíduo é a origem do relacionamento. Consulte o tópico "Grau do nó" na página 4 para obter mais informações. |

Selecione uma estatística específica da tabela Estatísticas Básicas para visualizar a distribuição de valores, o desvio padrão e a assimetria para a estatística.

Saída para análise de difusão

O nó Análise de Difusão gera uma variedade de campos que descrevem os indivíduos na rede. É possível aumentar os modelos existentes e dados com esses principais indicadores de desempenho para melhorar as previsões geradas por seus modelos. Por exemplo, é possível atualizar os valores de propensão de migração individuais para o concorrente para incluir influências de difusão.

A tabela "Principais indicadores de desempenho da Análise de Difusão" lista os campos dos principais indicadores de desempenho para os indivíduos na análise.

Tabela 13. Principais indicadores de desempenho da Análise de Difusão.

| Campo | Descrição |
|-------------------|--|
| DA_NodeNumber | Identificador exclusivo para um indivíduo |
| DA_DiffusedEnergy | Quantidade de energia difundida associada ao indivíduo. Para a análise de migração, valores mais altos indicam uma maior propensão de migração do que valores mais baixos. |
| DA_InDegree | Número de relacionamentos nos quais o indivíduo é o destino do relacionamento. Consulte o tópico "Grau do nó" na página 4 para obter mais informações. |
| DA_OutDegree | Número de relacionamentos nos quais o indivíduo é a origem do relacionamento. Consulte o tópico "Grau do nó" na página 4 para obter mais informações. |

Avisos

Estas informações foram desenvolvidas para produtos e serviços oferecidos no mundo todo.

É possível que a IBM não ofereça os produtos, serviços ou recursos discutidos nesta publicação em outros países. Consulte um representante IBM local para obter informações sobre produtos e serviços disponíveis atualmente em sua área. Qualquer referência a produtos, programas ou serviços IBM não significa que apenas produtos, programas ou serviços IBM possam ser usados. Qualquer produto, programa ou serviço funcionalmente equivalente, que não infrinja nenhum direito de propriedade intelectual da IBM poderá ser usado em substituição. Entretanto, a avaliação e verificação da operação de qualquer produto, programa ou serviço não IBM são de responsabilidade do Cliente.

A IBM pode ter patentes ou solicitações de patentes relativas a assuntos tratados nesta publicação. O fornecimento desta publicação não lhe garante direito algum sobre tais patentes. Pedidos de licença devem ser enviados, por escrito, para:

Gerência de Relações Comerciais e Industriais da IBM Brasil
IBM Corporation
Av. Pasteur, 138-146, Botafogo
Rio de Janeiro. RJ
CEP 22290-240
Brasil

Para pedidos de licença relacionados a informações de byte duplo (DBCS), entre em contato com o Departamento de Propriedade Intelectual da IBM em seu país ou envie pedidos de licença, por escrito, para:

Licença de Propriedade Intelectual
Lei de Propriedade Intelectual
IBM Japan Ltd.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 Japan

O parágrafo a seguir não se aplica a nenhum país em que tais disposições não estejam de acordo com a legislação local: A INTERNATIONAL BUSINESS MACHINES CORPORATION FORNECE ESTA PUBLICAÇÃO "NO ESTADO EM QUE SE ENCONTRA", SEM GARANTIA DE NENHUM TIPO, SEJA EXPRESSA OU IMPLÍCITA, INCLUINDO, MAS A ELAS NÃO SE LIMITANDO, AS GARANTIAS IMPLÍCITAS DE NÃO INFRAÇÃO, COMERCIALIZAÇÃO OU ADEQUAÇÃO A UM DETERMINADO PROPÓSITO. Alguns países não permitem a exclusão de garantias expressas ou implícitas em determinadas transações, portanto, essa disposição pode não se aplicar ao Cliente.

Essas informações podem conter imprecisões técnicas ou erros tipográficos. São feitas alterações periódicas nas informações aqui contidas; tais alterações serão incorporadas em futuras edições desta publicação. A IBM pode, a qualquer momento, aperfeiçoar e/ou alterar os produtos e/ou programas descritos nesta publicação, sem aviso prévio.

As referências nestas informações a websites não IBM são fornecidas apenas por conveniência e não representam de forma alguma um endosso a estes websites. Os materiais contidos nesses Web sites não fazem parte dos materiais deste produto IBM e a utilização desses Web sites é de inteira responsabilidade do Cliente.

A IBM pode usar ou distribuir as informações fornecidas da forma que julgar apropriada sem incorrer em qualquer obrigação para com o Cliente.

Licenciados deste programa que desejam obter informações sobre este assunto com objetivo de permitir: (i) a troca de informações entre programas criados independentemente e outros programas (incluindo este) e (ii) a utilização mútua das informações trocadas, devem entrar em contato com:

IBM Software Group
ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
CEP 22290-240

Tais informações podem estar disponíveis, sujeitas a termos e condições apropriadas, incluindo em alguns casos o pagamento de uma taxa.

O programa licenciado descrito nesta publicação e todo o material licenciado disponível são fornecidos pela IBM sob os termos do IBM Customer Agreement, Contrato de Licença do Programa Internacional IBM ou qualquer outro contrato equivalente.

Quaisquer dados de desempenho aqui contidos foram determinados em um ambiente controlado. Portanto, os resultados obtidos em outros ambientes operacionais podem variar significativamente. Algumas medidas podem ter sido tomadas em sistemas em nível de desenvolvimento e não há garantia de que estas medidas serão as mesmas em sistemas geralmente disponíveis. Além disso, algumas medidas podem ter sido estimadas através de extrapolação. Os resultados reais podem variar. Os usuários desta publicação devem verificar os dados aplicáveis para seu ambiente específico.

As informações relacionadas a produtos não IBM foram obtidas junto aos fornecedores destes produtos, de seus anúncios publicados ou de outras fontes disponíveis publicamente. A IBM não testou estes produtos e não pode confirmar a precisão de desempenho, compatibilidade nem qualquer outra reivindicação relacionada a produtos não IBM. Perguntas sobre os recursos de produtos não IBM devem ser encaminhadas aos fornecedores desses produtos.

Todas as instruções relativas aos objetivos ou intenção futura da IBM estão sujeitas a alterações ou cancelamento sem aviso prévio e representam apenas metas e objetivos.

Essas informações contêm exemplos de dados e relatórios usados em operações diárias de negócios. Para ilustrá-los da forma mais completa possível, os exemplos incluem nomes de indivíduos, empresas, marcas e produtos. Todos estes nomes são fictícios e qualquer semelhança com nomes e endereços usados por uma empresa real é mera coincidência.

Se essas informações estiverem sendo exibidas em formato eletrônico, as fotografias e ilustrações coloridas poderão não aparecer.

Marcas comerciais

IBM, o logotipo IBM e ibm.com são marcas comerciais ou marcas registradas da International Business Machines Corp., registradas em muitas jurisdições em todo o mundo. Outros nomes de produtos e serviços podem ser marcas comerciais da IBM ou de outras empresas. Uma lista atual de marcas comerciais da IBM trademarks está disponível na web em "Informações de Copyright e marcas comerciais" em www.ibm.com/legal/copytrade.shtml.

Intel, logotipo Intel, Intel Inside, logotipo Intel Inside, Intel Centrino, logotipo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium e Pentium são marcas registradas ou marcas comerciais da Intel Corporation ou suas subsidiárias nos Estados Unidos e outros países.

Linux é uma marca registrada da Linus Torvalds nos Estados Unidos e/ou em outros países.

Microsoft, Windows, Windows NT e o logotipo Windows são marcas comerciais da Microsoft Corporation nos Estados Unidos e/ou em outros países.

UNIX é uma marca registrada do The Open Group nos Estados Unidos e em outros países.

Java e todas as marcas e logotipos baseados em Java são marcas comerciais ou marcas registradas da Oracle e/ou suas afiliadas.

Outros nomes de produtos e serviços podem ser marcas comerciais da IBM ou de outras empresa.

Glossário

A

limite de precisão . Um critério de parada para um processo de difusão correspondente à menor mudança na energia do nó que garante a continuação da difusão.

líder de autoridade . O indivíduo em um grupo que tem o escore de autoridade mais alto.

escore de autoridade . Medida que indica o quanto os outros membros de um grupo se conectam a um indivíduo. Os escores de autoridade associam a importância de um indivíduo com o número de relacionamentos que terminam nele.

C

grupo principal . Um grupo de nós que restam após omitir os relacionamentos fracos de uma rede e aplicar os limites de tamanho do grupo.

limite de cobertura . A fração dos relacionamentos mais fortes usados para identificação do grupo.

D

degree . Medida de atividade do nó em uma rede, definida como o número total de relacionamentos que envolvem o nó.

density . Uma medida de coesão rede definida como o número de relacionamentos observados, dividido pelo número de relacionamentos possíveis.

relacionamento dicotômicos . Uma conexão entre dois nós que podem ter apenas um dos dois valores, indicando a presença ou ausência da conexão.

energia difundida . Efeito que se origina de nós específicos em uma rede que difunde em toda a rede, diminuindo de tamanho à medida que a distância dos nós iniciais aumenta.

relacionamento direcional . Uma conexão entre dois nós que se origina de um (a origem) e termina em outro (o destino).

líder de disseminação . O indivíduo em um grupo que tem o escore de disseminação mais alto.

escore de disseminação . Medida que indica o quanto um indivíduo se conecta aos outros membros de um grupo. Os escores de disseminação associam a importância de um indivíduo com o número de relacionamentos que se originam dele.

I

indegree . Medida de prestígio para um nó em uma rede composta de relacionamentos direcionados, definida como o número total de relacionamentos nos quais o nó é o destino.

N

relacionamento não direcional . uma associação entre dois nós.

O

outdegree . Medida de centralidade para um nó em uma rede composta de relacionamentos direcionados, definida como o número total de relacionamentos nos quais o nó é a origem.

S

sociograma . Uma representação visual de uma rede social na qual as linhas que conectam pontos representam indivíduos que têm um relacionamento entre si.

fator de difusão . A fração de energia que um nó transfere para seus vizinhos em uma etapa do processo de difusão.

V

relacionamento valorizado . Ocorreu uma conexão entre dois nós que tem um peso associado indicando a intensidade do relacionamento.

Índice Remissivo

A

analisando dados 6
análise de difusão 5, 17
 fator de difusão 17, 19
 limite de precisão 17, 19
 opções 18
 requisitos 18
 valores iniciais 19
análise de grupo 5
 limite de cobertura 12, 14
 requisitos 13
 tamanhos de grupo 12, 14

C

campos de destino 7
campos de origem 7
campos de ponderação 7, 13, 18

D

densidade 3, 14, 15
densidade do kernel 14, 15
densidade média do grupo 14
densidade média do grupo principal 14

E

energia difundida 20
escores de autoridade 12, 15
escores de disseminação 12, 15
executando fluxos
 usando o IBM SPSS Modeler Server
 Social Network Analysis 5

F

fator de difusão 17, 19
filtrando saída 5
fração média do membro principal 14

G

GAG_GroupNumber 15
grau 4
grupos principais 12
guia Análise 5
 análise de difusão 20
 análise de grupo 14
guia Anotações 5
guia Dados 5
 análise de difusão 18
 análise de grupo 13
guia Filtro 5
guia Opções de criação 5
 análise de difusão 19
 análise de grupo 14
guia Tipos 5

I

indegree 4, 14, 15, 20

L

líderes de autoridade 12, 15
líderes de disseminação 12, 15
limite de cobertura 12, 14
limite de precisão 17, 19
limpando resultados 6
lista de valores iniciais de difusão 19

M

média de influência 20
média indegree 14, 20
média outdegree 14, 20

N

níveis de medição 5
nós principais 14, 15

O

outdegree 4, 14, 15, 20

P

principais indicadores de
 desempenho 15, 20
propriedades
 script 8

R

registros de detalhes de chamada 7
relacionamentos
 direção 2
 valorizados 2
relacionamentos dicotômicos 2
relacionamentos direcionais 2
relacionamentos não direcionais 2
relacionamentos valorizados 2

S

script
 propriedades 8

T

tamanho médio do grupo 14
tamanhos de grupo 14, 15
 limites 12, 14

V

valores iniciais
 para análise de difusão 19
visualizando saída 6



Impresso no Brasil