

IBM SPSS Modeler Social Network
Analysis 15 - Manuale dell'utente



Nota: Prima di utilizzare queste informazioni e il relativo prodotto, leggere le informazioni generali disponibili in Note a pag. 22.

Questa versione si applica a IBM SPSS Modeler Social Network Analysis 15 e a tutte le successive versioni e modifiche fino a eventuali disposizioni contrarie indicate in nuove versioni.

Le schermate dei prodotti Adobe sono state ristampate su autorizzazione di Adobe Systems Incorporated.

Le schermate dei prodotti Microsoft sono state ristampate su autorizzazione di Microsoft Corporation.

Materiali concessi in licenza - Proprietà di IBM

© **Copyright IBM Corporation 2012.**

Tutti i diritti riservati.

Prefazione

IBM® SPSS® Modeler Social Network Analysis elabora informazioni sui rapporti tra le persone compilando dei campi che descrivono il ruolo di un individuo all'interno di una rete sociale, permettendo così di includere informazioni sociali nei modelli predittivi. Il presente manuale descrive come usare i nodi di IBM SPSS Modeler Social Network Analysis nell'ambiente IBM® SPSS® Modeler al fine di includere i nodi nei propri stream di dati. Combinando l'output dei nodi a campi che rappresentano misurazioni effettuate su individui si ottengono profili più completi delle persone.

Informazioni su IBM Business Analytics

Il software IBM Business Analytics fornisce informazioni complete, coerenti e accurate a cui i responsabili delle decisioni possono affidarsi per ottimizzare le prestazioni dell'azienda. Un ampio portafoglio di applicazioni di [business intelligence](#), [analisi predittiva](#), [gestione delle prestazioni e delle strategie finanziarie](#) e [analisi](#) offre una panoramica chiara, istantanea e interattiva delle prestazioni attuali e la possibilità di prevedere i risultati futuri. Utilizzato in combinazione con potenti soluzioni di settore, prassi consolidate e servizi professionali, questo software consente alle aziende di tutte le dimensioni di ottimizzare la produttività, automatizzare le decisioni senza problemi e fornire risultati migliori.

Come parte di questo portafoglio, il software IBM SPSS Predictive Analytics consente alle aziende di prevedere gli eventi futuri e di agire tempestivamente in modo da migliorare i risultati delle attività aziendali. Le aziende, gli enti governativi e le università di tutto il mondo si affidano alla tecnologia IBM SPSS perché rappresenta un vantaggio concorrenziale in termini di attrazione, retention e aumento dei clienti, riducendo al tempo stesso le frodi e limitando i rischi. Incorporando il software IBM SPSS nelle attività quotidiane, le aziende diventano imprese in grado di effettuare previsioni e di gestire e automatizzare le decisioni, per raggiungere gli obiettivi aziendali e vantaggi tangibili sulla concorrenza. Per ulteriori informazioni o per contattare un rappresentante, visitare il sito <http://www.ibm.com/spss>.

Supporto tecnico

Il supporto tecnico è a disposizione dei clienti che dispongono di un contratto di manutenzione. I clienti possono contattare il supporto tecnico per richiedere assistenza per l'utilizzo di IBM Corp. o per l'installazione di uno degli ambienti hardware supportati. Per contattare il supporto tecnico, visitare il sito Web IBM Corp. all'indirizzo <http://www.ibm.com/support>. Ricordare che durante la richiesta di assistenza sarà necessario fornire i dati di identificazione personali, i dati relativi alla propria società e il numero del contratto di manutenzione.

Contenuto

1 IBM SPSS Modeler Social Network Analysis 1

Informazioni su IBM SPSS Modeler Social Network Analysis	1
Informazioni sull'analisi delle reti sociali	2
Visualizzazione delle reti	3
Descrizione delle reti	3
nodi IBM SPSS Modeler Social Network Analysis	6
Schede dei nodi	6
Anteprima dell'output	6
Analisi dei dati	7
Applicazioni	7
Struttura dei dati	8
Proprietà script	9

2 Analisi di gruppo 10

Panoramica dell'analisi di gruppo	10
Determinazione della similarità sociale	10
Suddivisione in gruppi	11
Descrizione dei gruppi e dei membri	12
Nodo Analisi di gruppo	12
Definizione dei dati per l'analisi di gruppo	13
Impostazione delle opzioni di creazione per l'analisi di gruppo	14
Visualizzazione delle statistiche dell'analisi di gruppo	14
Output per analisi di gruppo	15

3 Analisi di diffusione 17

Panoramica dell'analisi di diffusione	17
Esempio di analisi di diffusione	17
Nodo Analisi di diffusione	18
Definizione dei dati per l'analisi di diffusione	19
Impostazione delle opzioni di creazione per l'analisi di diffusione	19
Visualizzazione delle statistiche dell'analisi di diffusione	20
Output dell'analisi di diffusione	21

Appendice

<i>A Note</i>	<i>22</i>
<i>Glossario</i>	<i>24</i>
<i>Bibliografia</i>	<i>25</i>
<i>Indice</i>	<i>26</i>

IBM SPSS Modeler Social Network Analysis

Informazioni su IBM SPSS Modeler Social Network Analysis

Molte metodologie di modellazione dei comportamenti si concentrano sugli individui. Usano una grande quantità di dati sugli individui per generare un modello che usi gli indicatori chiave del comportamento per la sua previsione. Se i valori degli indicatori chiave di un individuo indicano una propensione al comportamento, quell'individuo verrà ritenuto meritevole di un'attenzione speciale allo scopo di prevenire il comportamento.

Si considerino le metodologie di modellazione dell'abbandono in cui un cliente interrompe il suo rapporto con un'azienda. Poiché il costo da sostenere per trattenere i clienti è decisamente inferiore al costo per acquisirne di nuovi, la capacità di individuare i clienti con alta propensione all'abbandono diventa fondamentale. Per descrivere i clienti, gli analisti usano spesso vari indicatori delle prestazioni (KPI), come i dati demografici e gli schemi di chiamata recenti di ogni singolo cliente. Per individuare le persone che presentano la maggiore propensione all'abbandono, i modelli predittivi che fanno leva su questi campi confrontano i mutamenti negli schemi di chiamata dei clienti con gli schemi di chiamata dei clienti che hanno interrotto il loro rapporto con l'azienda. I clienti ritenuti a rischio ricevono un'attenzione speciale o servizi aggiuntivi volti a trattenerli.

Questi metodi, tuttavia, trascurano l'influenza che i rapporti sociali hanno sul comportamento dei clienti. Le informazioni circa un'azienda e il comportamento adottato da altre persone della propria cerchia sono dati che si diffondono e influenzano le persone. Di conseguenza, i rapporti che intercorrono tra le persone influenzano le decisioni e le azioni. Le analisi che esaminano esclusivamente dati individuali trascurano importanti fattori di previsione.

IBM® SPSS® Modeler Social Network Analysis affronta questo problema elaborando informazioni sulle relazioni e compilando campi aggiuntivi che possono essere inclusi nei modelli. Questi indicatori di prestazione chiave derivati misurano le caratteristiche sociali degli individui. La combinazione tra queste caratteristiche sociali e le metriche individuali delinea un profilo più completo degli individui e di conseguenza può migliorare l'accuratezza predittiva dei modelli.

IBM SPSS Modeler Social Network Analysis è costituito da due componenti principali:

- I nodi IBM SPSS Modeler Social Network Analysis aggiunti all'ambiente IBM® SPSS® Modeler tramite i quali si possono inserire tecniche di analisi sociale negli stream di dati
- IBM® SPSS® Modeler Server Social Network Analysis che abilita l'elaborazione delle specifiche dei nodi in IBM® SPSS® Modeler Server. IBM SPSS Modeler Server Social Network Analysis elabora in modo efficiente enormi quantità di dati (una rete può contenere milioni di individui e relazioni) ed estrapola un numero di campi relativamente contenuto da sottoporre a successive analisi.

Per esempio, IBM SPSS Modeler Social Network Analysis riconosce gli individui di una rete che sono maggiormente influenzati dall'abbandono da parte di altri. Inoltre, permette di scoprire gruppi di individui che all'interno di una rete presentano una maggiore propensione all'abbandono. Incorporando nei modelli i valori KPI che registrano tali effetti, è possibile migliorare le prestazioni globali dell'analisi.

Nota: L'elaborazione dei nodi IBM SPSS Modeler Social Network Analysis deve essere eseguita da IBM SPSS Modeler Server Social Network Analysis. Quando si esegue uno stream contenente nodi IBM SPSS Modeler Social Network Analysis in SPSS Modeler, è necessario connettersi a un'istanza SPSS Modeler Server che includa anche IBM SPSS Modeler Server Social Network Analysis.

Informazioni sull'analisi delle reti sociali

Una rete sociale è caratterizzata da una serie di individui tra cui intercorrono delle relazioni. L'analisi delle reti sociali esamina tali relazioni per descrivere gli individui e i gruppi come costituenti di una struttura sociale. Gli schemi che descrivono le interazioni tra gli individui di una rete forniscono preziose indicazioni sugli individui. Le relazioni facilitano il passaggio di informazioni all'interno della rete e favoriscono l'emulazione. L'importanza che le informazioni sulle relazioni rivestono per l'analisi delle reti sociali rende questo tipo di approccio diverso da tutti gli altri. Aniché concentrarsi su ogni individuo singolarmente, questa analisi utilizza come unità di studio il binomio composto da due individui e dalle loro relazioni.

Le relazioni presenti in una rete possono essere classificate come direzionali o non direzionali. In una **relazione direzionale**, un individuo funge da origine e l'altro da destinazione. Per esempio, una chiamata telefonica è un tipo di relazione direzionale con un'origine e una destinazione. Per contrasto, i ruoli di origine e destinazione non possono essere definiti per le **relazioni non direzionali**. In questo caso, entrambi gli elementi partecipano equamente alla relazione. Una conversazione tra due persone è un esempio di relazione non direzionale.

Un'altra proprietà che caratterizza le relazioni è la dicotomia o il valore. L'unica informazione disponibile in una **relazione dicotomica** è se la relazione tra due individui esiste o meno. Per ogni coppia di individui della rete, la relazione è semplicemente presente oppure assente. Una **relazione di valore**, invece, comprende anche un peso che descrive la forza della relazione. I pesi permettono di mettere a confronto le relazioni.

La tabella "Tipi di relazioni" contiene alcuni esempi di classificazione incrociata delle relazioni in base alla direzione e alla scala. Nelle relazioni direzionali *Joe* è l'origine e *Mary* la destinazione. Nelle relazioni non direzionali, non sono presenti indicazioni circa chi ha dato avvio alla relazione. Nelle relazioni di valore si usa la durata della conversazione per valutare il peso della relazione; mentre le relazioni dicotomiche semplicemente avvengono o non avvengono.

Tabella 1-1
Tipi di relazioni

Direzione	Scala	Esempio
Non direzionale	Dicotomica	Joe e Mary hanno una conversazione
Non direzionale	Di valore	Joe e Mary hanno avuto una conversazione di 20 minuti

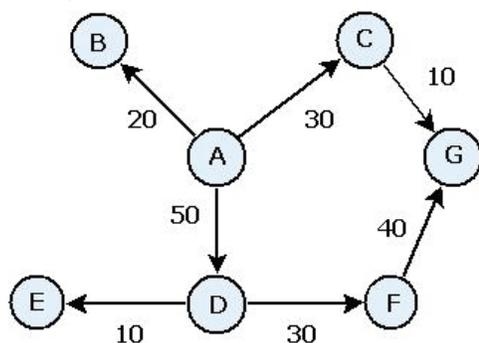
Direzione	Scala	Esempio
Direzionale	Dicotomica	Joe ha chiamato Mary
Direzionale	Di valore	Joe ha chiamato Mary e la conversazione è durata 20 minuti

Per ulteriori informazioni sulle analisi delle reti sociali, consultare uno dei testi dedicati presenti in quest'area (Wasserman e Faust, 1994).

Visualizzazione delle reti

Per rappresentare graficamente una rete sociale si usa generalmente un **sociogramma** (Moreno, 1934). In questo tipo di rappresentazione gli individui corrispondono a punti, o nodi, in uno spazio. Le linee che collegano i punti sono le relazioni interpersonali. Se le relazioni sono di tipo direzionale, la linea è sostituita da una freccia. Se le relazioni hanno un peso, le etichette delle linee lo indicano. Il grafico seguente rappresenta una rete di sette individui.

Figura 1-1
Esempio di rete sociale



Supponiamo che la rete rappresenti telefonate intercorse tra alcuni individui dove il peso della relazione è indicato dalla durata della chiamata. In questo caso, la Persona A ha chiamato tre persone, parlando più a lungo con la Persona D.

Nonostante questa rete abbia una dimensione nettamente inferiore rispetto a quelle reali, i concetti illustrati tramite semplici sociogrammi sono applicabili a reti di qualsiasi portata e complessità.

Descrizione delle reti

Le informazioni su reti, gruppi e individui devono essere estratte sotto forma di caratteristiche descrittive che possano essere sottoposte a confronti incrociati e incluse in modelli predittivi. Dalle reti si deve estrapolare una serie finita di indicatori di prestazioni chiave che possano essere analizzati. Per esempio, deve essere possibile confrontare tra loro reti o gruppi di nodi di una stessa rete. In alternativa, si potrebbe voler confrontare tra loro gli individui di una rete o riconoscere gli individui più importanti.

Le due metriche comunemente usate per descrivere le reti sociali sono la **densità** e il **grado**. Entrambe queste statistiche riflettono la connettività ma mentre la prima si riferisce all'intera rete o a sottogruppi di essa, la seconda caratterizza gli individui della rete.

Densità della rete

Per ogni serie di nodi di una rete esiste un numero finito di relazioni possibili. Ogni nodo può fungere da origine o da destinazione nella relazione con tutti gli altri nodi. Si consideri una rete contenente i nodi A, B e C. La tabella seguente riporta tutte le relazioni dirette possibili tra i tre nodi.

Tabella 1-2
Relazioni dirette possibili tra i tre nodi

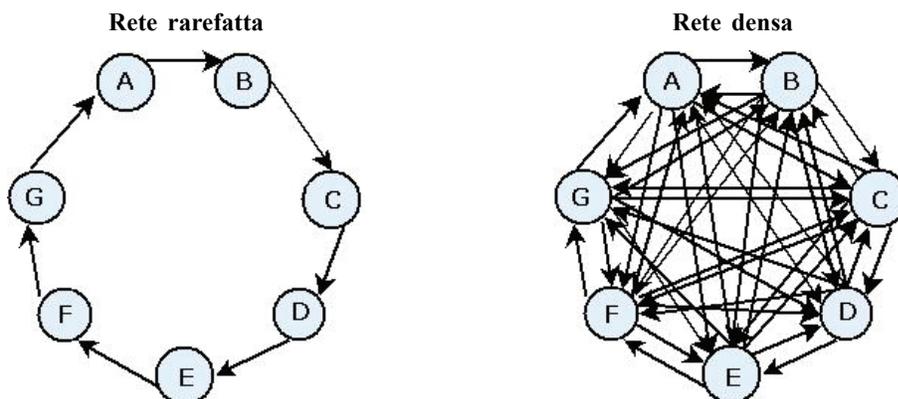
Origine	Obiettivo
A	B
A	C
B	A
B	C
C	A
C	B

Ogni nodo è l'origine della relazione con gli altri due nodi. In pratica, però, non è detto che tutte le relazioni possibili siano presenti. Alcuni nodi potrebbero non presentare alcuna relazione diretta con gli altri nodi. Inoltre, alcune relazioni dirette potrebbero non essere reciproche.

La statistica **densità** rappresenta la proporzione delle relazioni possibili nella rete che sono effettivamente presenti. Il valore è compreso tra 0 e 1, dove il limite inferiore corrisponde alle reti senza relazioni e il limite superiore rappresenta le reti con tutte le relazioni possibili. Più il valore si avvicina a 1, maggiore è la densità della rete e più sono coesi i nodi al suo interno.

All'interno delle reti dense, le informazioni viaggiano più facilmente rispetto alle reti più rarefatte. La tabella "Reti rarefatte e dense" illustra due reti contenenti sette nodi. La rete meno densa presenta solo sette delle 42 relazioni possibili, il che corrisponde a una densità di 0,17. La rete densa, al contrario, contiene tutte le relazioni possibili, il che corrisponde a una densità di 1,0.

Tabella 1-3
Reti rarefatte e dense



Nella rete rarefatta, per passare dal nodo A al nodo G un'informazione deve attraversare altri cinque nodi. Nella rete densa, invece, l'informazione può passare direttamente dal nodo A al nodo G.

Grado nodale

Gli individui importanti di una rete sono spesso quelli coinvolti in più relazioni interpersonali. Questi individui acquisiscono informazioni da varie fonti e le diffondono a una grande quantità di altri individui. Al contrario, gli individui che partecipano a poche relazioni hanno meno possibilità di influenzare direttamente un numero elevato di altri membri della rete.

Il **grado** di un nodo, cioè il numero totale di relazioni in cui il nodo è coinvolto, permette di eseguire confronti tra i membri della rete. Gli individui con grado più alto sono più attivi di quelli con grado inferiore. Il grado ignora la direzione delle relazioni e fornisce una metrica complessiva del livello di attività di un determinato nodo.

Il conteggio delle relazioni dirette può tener conto del fatto che un nodo sia l'origine o la destinazione della relazione. Il valore **legami in entrata** di un nodo rappresenta il numero di relazioni per cui il nodo funge da destinazione. Il valore **legami in uscita**, invece, rappresenta il numero di relazioni in cui il nodo funge da origine. La tabella seguente riporta i valori di grado, legami in entrata e legami in uscita di ogni nodo della figura "Esempio di rete sociale".

Tabella 1-4

Esempio di valori di grado, legami in entrata e legami in uscita

Nodo	Grado	Legami in entrata	Legami in uscita
A	3	0	3
B	1	1	0
C	2	1	1
D	3	1	2
E	1	1	0
F	2	1	1
G	2	2	0

Il valore Legami in entrata viene spesso considerato un'indicazione del prestigio. Alti valori di legami in entrata corrispondono a grandi quantità di relazioni che hanno come destinazione quel nodo. In altre parole, questi individui vengono contattati da tanti individui. Molti nodi avviano relazioni con questo nodo. Il valore Legami in uscita, invece, viene considerato un'indicazione della centralità. Alti valori corrispondono a grandi quantità di relazioni che hanno origine da quel nodo. Questi individui contattano molti altri individui.

Nella nostra rete di esempio, i gradi indicano che i nodi A e D sono i più attivi mentre i nodi B e E i meno attivi. L'analisi dei valori dei Legami in entrata rivela che il nodo G è quello che gode di maggiore prestigio. L'analisi dei valori dei Legami in uscita indica che il nodo A è il più centrale.

nodì IBM SPSS Modeler Social Network Analysis

Oltre ai nodi standard proposti con IBM® SPSS® Modeler, è possibile utilizzare anche i nodi IBM® SPSS® Modeler Social Network Analysis per includere negli stream i risultati derivanti dall'analisi della rete sociale. La tabella "nodì IBM SPSS Modeler Social Network Analysis" descrive tali nodi che sono disponibili nella palette Input.

Tabella 1-5
nodi IBM SPSS Modeler Social Network Analysis

Nodo	Icona	Descrizione
Analisi di gruppo		Il nodo Analisi di gruppo importa i record dei dati di chiamata da un file di testo a campi fissi, individua gruppi di nodi all'interno della rete definita dai record e genera indicatori di prestazioni chiave per i gruppi e gli individui della rete. Per ulteriori informazioni, vedere l'argomento Panoramica dell'analisi di gruppo in il capitolo 2 a pag. 10.
Analisi di diffusione		Il nodo Analisi di diffusione importa i record dei dati di chiamata da un file di testo a campi fissi, propaga un effetto all'interno della rete definita dai record e genera indicatori di prestazioni chiave che riassumono i risultati provocati sui singoli nodi dall'effetto. Per ulteriori informazioni, vedere l'argomento Panoramica dell'analisi di diffusione in il capitolo 3 a pag. 17.

L'elaborazione dei nodi IBM SPSS Modeler Social Network Analysis deve essere eseguita da IBM® SPSS® Modeler Server Social Network Analysis. Quando si esegue uno stream contenente questi nodi in SPSS Modeler, è necessario connettersi a un'istanza IBM® SPSS® Modeler Server che includa anche IBM SPSS Modeler Server Social Network Analysis.

Schede dei nodi

I nodi IBM® SPSS® Modeler Social Network Analysis contengono le schede seguenti tramite le quali l'utente definisce l'analisi e ne richiede un'anteprima:

- **Scheda Dati.** Identifica il file che contiene le informazioni sulla rete sociale.
- **Scheda Opzioni di creazione.** Utilizzata per definire le impostazioni per le analisi.
- **Scheda Analisi.** Utilizzata per vedere un riassunto dell'output preliminare da usare come guida per perfezionare le impostazioni di input della scheda Dati e ottenere risultati ottimali.

Inoltre, i nodi di input contengono le seguenti schede comuni in IBM® SPSS® Modeler:

- **Scheda Filtro.** Utilizzata per eliminare o ridenominare i campi di output generati dai nodi. Questa scheda include le stesse opzioni disponibili nel nodo Filtro.
- **Scheda Tipi.** Utilizzata per impostare i livelli di misurazione per i campi di output generati dai nodi. Questa scheda include le stesse opzioni disponibili nel nodo Tipo.
- **Scheda Annotazioni.** Utilizzata per ridenominare i nodi, definire un testo descrittivo personalizzato e registrare un commento particolarmente lungo.

Per ulteriori informazioni sulle schede comuni, vedere la documentazione relativa a SPSS Modeler.

Anteprima dell'output

Generalmente, i dati di output dei nodi vengono utilizzati come dati di input del modello predittivo. Per vedere l'output di un nodo, è possibile aggiungere un nodo Tabella allo stream. Tuttavia, poiché la quantità di dati generata è spesso estremamente elevata, la compilazione delle tabelle è laboriosa. Inoltre, il numero di righe della tabella ne limita l'utilità.

Come alternativa si consiglia di generare una tabella di anteprima che visualizzi un campione dell'output che verrà creato. L'anteprima mostra i campi generati per un limitato numero di righe. Il numero di righe è definito nelle proprietà dello stream. Per creare un'anteprima dell'output del nodo, attenersi alla procedura descritta di seguito:

1. Aprire il nodo.
2. Specificare le impostazioni dei dati nella scheda Dati.
3. Definire i parametri analitici nella scheda Opzioni di creazione.
4. Fare clic su Anteprima.

Si apre una finestra di anteprima con i risultati richiesti. La generazione dell'anteprima, inoltre, compila la scheda Analisi del nodo con il riassunto dei risultati.

Analisi dei dati

La definizione di impostazioni di analisi in grado di generare risultati ottimali è generalmente un processo iterativo, vale a dire che si definiscono le impostazioni, si esegue l'analisi e si valutano i risultati. Se i risultati non sembrano abbastanza utili, si modificano le impostazioni e si esegue nuovamente l'analisi.

Per analizzare i dati di input per il nodo, attenersi alla procedura descritta di seguito:

1. Aprire il nodo.
2. Specificare le impostazioni dei dati nella scheda Dati.
3. Definire i parametri analitici nella scheda Opzioni di creazione.
4. Fare clic su Analizza dati.

Se le opzioni di creazione prevedono la visualizzazione di statistiche riassuntive, i risultati vengono riportati nella scheda Analisi.

Se si rende necessario un nuovo ciclo di analisi, fare clic su Cancella analisi per eliminare i risultati attuali prima di fare clic su Analizza dati.

Applicazioni

Esempi di applicazioni specifiche in cui l'impiego di IBM® SPSS® Modeler Social Network Analysis può risultare particolarmente vantaggioso:

- **Previsione del tasso di abbandono.** Le caratteristiche di un gruppo possono influenzare i tassi di abbandono. Concentrando l'attenzione sugli individui del gruppo che presentano una maggiore propensione all'abbandono è probabile che si riesca a prevenire tale comportamento. Inoltre, è possibile riconoscere gli individui "a rischio" analizzando il flusso di informazioni proveniente dalle persone che già si sono ritirate.

- **Sfruttare i leader dei gruppi.** All'interno dei gruppi i leader esercitano una grande influenza sugli altri membri. Prevenendo l'abbandono da parte del leader di un gruppo è possibile ridurre il tasso di abbandono degli altri membri del gruppo. In alternativa, è possibile tentare di facilitare l'abbandono del leader di un gruppo di un concorrente per aumentare il tasso di abbandono degli altri membri e contestualmente ridurre il tasso di abbandono dei membri associati alla propria azienda.
- **Marketing.** Si possono usare i leader dei gruppi per introdurre nuove offerte di prodotti e servizi. Esercitando la sua influenza, il leader può convincere altri membri del gruppo ad acquistare l'offerta. Usare l'analisi di diffusione per riconoscere gli individui maggiormente influenzati dai leader e creare una campagna di marketing mirata.

Struttura dei dati

Le informazioni sugli individui della rete di interesse potrebbero essere segmentate in vari file, database e sistemi dell'azienda. Per poter analizzare la rete tramite IBM® SPSS® Modeler Social Network Analysis è necessario estrarre tutti i record e i campi di interesse dalle origini e strutturarli in modo che possano essere inseriti nei nodi.

Per i nodi da analizzare è necessario disporre di record dei dati di chiamata organizzati all'interno di un unico file di testo a larghezza fissa. Ogni riga corrisponderà a una relazione e i dati saranno organizzati nelle colonne seguenti:

- l'identificatore dell'individuo che ha dato avvio alla relazione.
- l'identificatore dell'individuo che funge da destinazione della relazione.
- peso della relazione (facoltativo).

Tutti i dati devono essere in formato numerico e gli identificatori degli individui devono essere numeri interi. È facoltativo aggiungere i nomi dei campi come prima riga dei dati. La tabella "Esempio di record dei dati di chiamata" illustra questa struttura di dati.

Tabella 1-6
Esempio di record dei dati di chiamata

Origine	Destinazione	Peso
1000	5642	243
2190	8444	831
0299	9419	559

I valori del peso possono corrispondere a qualsiasi metrica si voglia utilizzare per rappresentare l'importanza della relazione rispetto alle altre relazioni della rete. Per i dati di chiamata, i parametri di peso comuni sono la durata della chiamata e la frequenza.

Se si desidera che l'analisi si concentri su una porzione di chiamate, sarà necessario usare tale sottogruppo per creare il file di input. Per esempio, si può limitare l'analisi ai mesi più recenti o alle chiamate più recenti creando un file di input contenente solo questi dati.

Proprietà script

Gli script di IBM® SPSS® Modeler Social Network Analysis sono un potente strumento per automatizzare i processi nell'interfaccia utente. Tramite gli script è possibile eseguire gli stessi tipi di azioni eseguite con il mouse o la tastiera, nonché automatizzare le attività ripetitive o la cui esecuzione manuale richiederebbe un tempo molto maggiore. Per una spiegazione sull'uso degli script, consultare la guida *ScriptingAutomation.pdf* disponibile con IBM® SPSS® Modeler.

proprietà nodo Diffusione

Nella tabella seguente vengono elencate le proprietà degli script per il nodo Diffusione.

Nome proprietà	Tipo di dati	Descrizione proprietà
input_data_file_name	stringa	
calling_field	campo	
called_field	campo	
frequency_weight_field	campo	
read_field_names	boolean	
diffusion_list_file_name	stringa	
spreading_factor	double	Valore di default = 50. Min = 1 Max = 99.
max_number_iterations	intero	Valore di default = 100. Min = 1.
accuracy_threshold	double	Valore di default = 0.01. Min = 0.001.
calculate_statistics	boolean	

proprietà nodo Analisi di gruppo

Nella tabella seguente vengono elencate le proprietà degli script per il nodo Analisi di gruppo.

Nome proprietà	Tipo di dati	Descrizione proprietà
input_data_file_name	stringa	
calling_field	campo	
called_field	campo	
frequency_weight_field	campo	
read_field_names	boolean	
coverage_threshold	double	Valore di default = 10. Min = 1 Max = 99.
min_group_size	intero	Valore di default = 2. Min = 2.
max_group_size	intero	Valore di default = 100. Min = 2.
calculate_statistics	boolean	

Analisi di gruppo

Panoramica dell'analisi di gruppo

L'analisi di gruppo analizza gli schemi di interazione tra gli individui di una rete per rilevare gruppi di individui simili. Le caratteristiche di questi gruppi influenzano il comportamento dei singoli membri dei gruppi. Per esempio, piccoli gruppi caratterizzati da molteplici relazioni interpersonali e leader influenti presentano una propensione all'abbandono superiore anche se effettivamente nessun membro del gruppo si è mai ritirato. (Richter, Yom-Tov, e Slonim, 2010) I modelli predittivi che esaminano sia metriche di gruppo che individuali generano risultati più accurati rispetto ai modelli che analizzano solo metriche individuali.

L'analisi di gruppo prevede i seguenti passaggi generici:

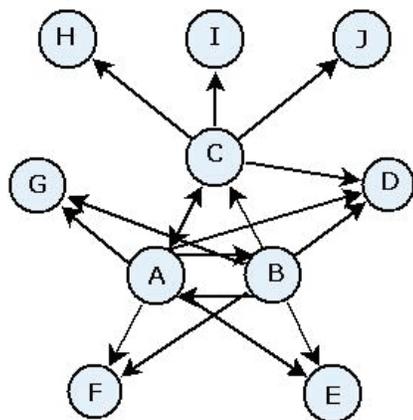
1. Determinare la forza delle relazioni che riflettono la prossimità sociale. Per ulteriori informazioni, vedere l'argomento [Determinazione della similarità sociale](#) a pag. 10.
2. Suddividere la rete in gruppi basandosi sulla forza delle relazioni pur rispettando le dimensioni limite. Per ulteriori informazioni, vedere l'argomento [Suddivisione in gruppi](#) a pag. 11.
3. Tracciare un profilo dei gruppi e degli individui e indicare i leader dei gruppi. Per ulteriori informazioni, vedere l'argomento [Descrizione dei gruppi e dei membri](#) a pag. 12.

Determinazione della similarità sociale

I membri di un gruppo devono essere caratterizzati da un livello di “somiglianza” maggiore rispetto agli individui che non appartengono al gruppo. Nell'analisi di rete, la similarità tra due nodi dipende dalle relazioni di ognuno. Per ogni nodo della rete è presente un gruppo di nodi che rappresentano la destinazione delle relazioni dirette con quel nodo. Nel settore delle telecomunicazioni, questo gruppo corrisponde a tutte le persone contattate da un particolare individuo. Due individui che contattano lo stesso gruppo di persone vengono considerati simili. Più è alto il numero di sovrapposizioni tra le destinazioni delle relazioni di due individui, più questi due individui saranno simili.

Consultare la rete illustrata nella figura [“Esempio di rete con dieci nodi”](#)..

Figura 2-1
Esempio di rete con dieci nodi



La tabella “Nodi di destinazione” mostra i nodi di destinazione per le relazioni che hanno come origine i nodi A, B e C.

Tabella 2-1
Nodi di destinazione

Nodo di origine	Nodi di destinazione
A	B, C, D, E, F, G
B	A, C, D, E, F, G
C	A, D, H, I, J

I nodi A e B hanno cinque nodi di destinazione in comune. I nodi A e C, invece, ne hanno solo uno. Dal punto di vista sociale, i nodi A e B sono più simili tra loro rispetto ai nodi A e C.

Pesi della relazione come durata o frequenza di chiamata non esprimono il grado di similarità dei nodi di una rete. Per includere l’aspetto della similarità, l’analisi di gruppo deve usare informazioni reciproche (Cover e Thomas, 2006) come peso della relazione. Questa statistica riflette il tasso di probabilità che due nodi siano collegati allo stesso nodo. Se le relazioni all’interno di una rete hanno pesi predefiniti, come la frequenza di chiamata nel settore delle telecomunicazioni, le informazioni reciproche integreranno tali dati sul peso.

Suddivisione in gruppi

I gruppi dovrebbero essere formati da individui tra cui intercorrono relazioni di peso elevato, dove i pesi misurano la similarità dei nodi nella relazione. Di conseguenza, il primo passo per identificare un gruppo consiste nell’omettere le relazioni più deboli della rete. La **soglia di copertura** regola questo processo definendo la percentuale delle relazioni più forti da considerare. Per esempio, con una soglia di copertura pari a 0,4, il 40% delle relazioni più forti verrà usato per l’identificazione di gruppi mentre il restante 60% delle relazioni verrà omesso.

Le relazioni restanti possono generare alcuni gruppi molto piccoli o molto grandi caratterizzati da un’utilità predittiva limitata. Per evitare che tali gruppi vengano inclusi nell’analisi, è possibile specificare la dimensione minima e massima dei gruppi. I gruppi con dimensioni inferiori

al minimo vengono completamente omessi. I gruppi con dimensioni superiori al massimo, tuttavia, vengono suddivisi in gruppi più piccoli di dimensioni accettabili. I gruppi restanti dopo l'omissione delle relazioni deboli e l'applicazione dei limiti delle dimensioni vengono chiamati **gruppi principali**.

In conseguenza all'eliminazione di alcune relazioni dalla rete originale, alcuni individui potrebbero non risultare inseriti in alcun gruppo principale. Tuttavia, questi individui potrebbero avere delle relazioni con membri di un gruppo tali da motivarne l'inclusione nel gruppo. Un individuo può essere aggiunto a un gruppo se presenta molte relazioni relativamente forti con i membri principali del gruppo e a patto che non vengano violati i limiti relativi alle dimensioni del gruppo. I gruppi finali saranno quindi composti da un insieme di membri principali e da membri aggiunti in virtù dei loro rapporti con i membri principali.

Descrizione dei gruppi e dei membri

Oltre alla densità, ai legami in entrata e ai legami in uscita esistono altre statistiche che descrivono le dinamiche dei gruppi. In particolare, i punteggi di autorità e di disseminazione descrivono lo stato sociale degli individui all'interno dei gruppi. Il ruolo di ogni individuo all'interno di un gruppo ha un'importanza vitale per prevedere il comportamento del gruppo nel suo complesso e dei singoli membri.

Il **punteggio di autorità** di un nodo misura la tendenza degli altri nodi del gruppo a consultarlo. Se molti individui contattano una persona particolare, potenzialmente per richiedere informazioni o punti di vista, quella persona gode di un ruolo di autorità. I punteggi di autorità per i nodi del gruppo, che corrispondono alle probabilità stazionarie di una passeggiata aleatoria con riprese nella rete di gruppi, variano da 0 a 1. Più il punteggio di autorità si avvicina a 1, maggiore è l'autorità di cui gode il nodo all'interno del gruppo. Il nodo del gruppo con maggiore autorità è definito **leader con autorità** del gruppo. Suddividendo il punteggio più alto di un gruppo per il punteggio più basso si ottiene la misura della forza globale del leader con autorità.

Per contrasto, il **punteggio di disseminazione** di un nodo misura la tendenza del nodo a consultare altri nodi del gruppo. Una persona che contatta molte altre persone di un gruppo è in grado di influenzare in modo significativo le opinioni dell'intero gruppo. I punteggi di disseminazione per i nodi del gruppo, che corrispondono alle probabilità stazionarie di una passeggiata aleatoria con riprese nella rete di gruppi al contrario, variano da 0 a 1. Più il punteggio di disseminazione si avvicina a 1, maggiore è la frequenza di collegamento del nodo agli altri membri del gruppo. Il nodo del gruppo con il punteggio di disseminazione più alto è definito **leader disseminante** del gruppo. Suddividendo il punteggio più alto di un gruppo per il punteggio più basso si ottiene la misura della forza globale del leader disseminante.

Nodo Analisi di gruppo

Il nodo Analisi di gruppo è disponibile nella palette Input e serve a individuare gruppi di individui socialmente simili tra loro all'interno di una rete. Inoltre, determina lo stato sociale relativo dei membri dei gruppi. Nei gruppi relativamente piccoli caratterizzati da membri con molti collegamenti tra di loro e presieduti da un leader influente, si delinea una propensione

all'abbandono maggiore determinata dalle dinamiche di gruppo. I punteggi di abbandono del gruppo calcolati combinati ai punteggi di abbandono individuali migliorano la capacità di prevedere la propensione all'abbandono rispetto ai modelli che analizzano solo i punteggi individuali.

Se consideriamo la quantità di dati che generalmente una rete contiene e la natura delle interazioni sociali che sono alla base delle definizioni dei gruppi, risulta chiaro quanto l'analisi di gruppo sia un processo laborioso. È pertanto consigliabile salvare i risultati dell'analisi di gruppo in un database o in un file da usare come input per il modello predittivo. I risultati dei gruppi verrebbero aggiornati abbastanza raramente, per esempio una volta al mese, mentre il modello potrebbe essere aggiornato più frequentemente.

Requisiti. Il nodo deve disporre di un file di testo a larghezza fissa che definisca la rete sociale tramite tre campi. Un campo indica l'origine di ogni relazione diretta, un altro campo definisce la destinazione di ogni relazione diretta e il terzo campo, opzionale, specifica la forza di ogni relazione. Tutte le relazioni presenti nella rete devono avere una direzione (origine-destinazione).

Definizione dei dati per l'analisi di gruppo

La scheda Dati nella finestra del nodo di origine Analisi di gruppo permette di specificare il file di input contenente le relazioni esistenti nel nodo di rete.

File. Specificare il nome di un file o di una cartella contenente i record dei dati di chiamata. È possibile immettere un nome o fare clic sul pulsante con i puntini di sospensione (...) per selezionare un nome dal file system. Il percorso viene mostrato quando il nome viene selezionato; i relativi contenuti vengono visualizzati separati da delimitatori. Se si specifica una cartella, i record dei dati di chiamata di tutti i file della cartella vengono concatenati per l'analisi; tutti i file della cartella devono avere la medesima struttura.

Leggi i nomi dei campi dal file. Questa opzione è selezionata per default e consente di utilizzare i nomi dei campi inclusi nella prima riga nel file di dati come nomi per le colonne. Se la prima riga non è un'intestazione, deselezionare l'opzione per assegnare automaticamente un nome generico a ogni campo, come *Campo1* e *Campo2*.

Impostazioni definizione di rete

Le impostazioni definizione di rete stabiliscono i ruoli dei campi.

Campi. Usare i pulsanti freccia per assegnare manualmente gli elementi dell'elenco ai vari campi di ruolo. Le icone indicano i livelli di misurazione validi per ogni campo di ruolo. Fare clic sul pulsante Tutto per selezionare tutti i campi dell'elenco o fare clic su un singolo pulsante di livello di misurazione per selezionare tutti i campi con tale livello.

Origine. Selezionare un campo come origine della relazione diretta.

Destinazione. Selezionare un campo come destinazione della relazione diretta.

Peso. Scegliere il campo corrispondente al numero di iterazioni della relazione all'interno della rete (facoltativo). Per determinare la forza delle relazioni, l'analisi calcola il peso di ciascun record.

Impostazione delle opzioni di creazione per l'analisi di gruppo

La scheda Opzioni di creazione nella finestra del nodo di origine Analisi di gruppo permette di definire le opzioni per l'identificazione dei gruppi nella rete.

Impostazioni analisi di gruppo

Le impostazioni per l'analisi di gruppo influiscono sulla dimensione e sulla forza relativa dei gruppi formati.

Soglia di copertura. Definisce la percentuale delle relazioni più forti della rete da includere nell'analisi. Per esempio, una soglia di copertura di 0,2 indica che verranno usate solo le relazioni il cui peso rientra nella fascia del 20% dei pesi maggiori. Questo parametro, il cui valore è compreso tra 0 e 1, permette all'analisi di concentrarsi sulle relazioni più forti presenti nei gruppi. Valori di sogli più alti generano gruppi in cui tra i membri intercorrono relazioni più deboli. Per ulteriori informazioni, vedere l'argomento [Suddivisione in gruppi](#) a pag. 11.

Dimensione gruppo min. Specifica la dimensione minima del gruppo. I gruppi la cui dimensione è inferiore a questo valore non vengono restituiti.

Dimensione gruppo max. Specifica la dimensione massima del gruppo. I gruppi la cui dimensione è superiore a questo valore vengono suddivisi.

Calcola e visualizza statistiche riassuntive. Quando questa opzione è selezionata, oltre a calcolare i KPI, il nodo calcola e visualizza statistiche riassuntive dell'analisi. Nelle reti estremamente grandi, il calcolo delle statistiche riassuntive ha un impatto negativo sulla velocità di elaborazione dei nodi.

Visualizzazione delle statistiche dell'analisi di gruppo

La scheda Analisi nella finestra del nodo di origine Analisi di gruppo contiene un riassunto dei gruppi individuati nella rete. La tabella "Statistiche riassuntive dell'analisi di gruppo" elenca tutte le statistiche riassuntive disponibili per i gruppi.

Tabella 2-2

Statistiche riassuntive dell'analisi di gruppo

Statistica	Descrizione
Nodi totali nei gruppi	Numero di nodi presenti nei gruppi individuati
Legami totali nei gruppi	Numero di legami presenti nei gruppi individuati
Numero di gruppi totale	Numero di gruppi individuati nella rete
Dimensione del gruppo media	Numero medio di nodi in un gruppo
Densità del gruppo media	Frazione media di collegamenti diretti tra i nodi di un gruppo. Per ulteriori informazioni, vedere l'argomento Densità della rete in il capitolo 1 a pag. 4.

Statistica	Descrizione
Frazione media dei membri principali	Frazione media di nodi di un gruppo che sono nodi principali. Per ulteriori informazioni, vedere l'argomento Suddivisione in gruppi a pag. 11.
Densità media del gruppo principale	Frazione media di collegamenti diretti tra i nodi principali di un gruppo
Legami in entrata medi	Numero medio di legami in entrata. Per ulteriori informazioni, vedere l'argomento Grado nodale in il capitolo 1 a pag. 5.
Legami in uscita medi	Numero medio di legami in uscita. Per ulteriori informazioni, vedere l'argomento Grado nodale in il capitolo 1 a pag. 5.

Selezionare una singola statistica nella tabella Statistiche riassuntive per consultare la distribuzione dei valori, la deviazione standard e l'asimmetria della statistica.

La revisione di queste statistiche aiuta a definire le impostazioni per l'analisi di gruppo. Idealmente i risultati dovrebbero delineare gruppi relativamente piccoli con alti valori di densità. Per esempio, se sono presenti gruppi molto grandi, prendere in considerazione la possibilità di suddividerli in gruppi più piccoli. In alternativa, se i valori di densità tendono a essere contenuti, provare a ridurre la soglia di copertura per concentrarsi su relazioni più forti.

Output per analisi di gruppo

Il nodo Analisi di gruppo genera una serie di campi che descrivono i gruppi e gli individui al loro interno. I modelli e i dati esistenti possono essere ampliati tramite i KPI per migliorare le previsioni generate dai modelli. Per esempio, è possibile perfezionare i valori di propensione all'abbandono di un individuo aggiungendo le influenze dei gruppi.

La tabella "KPI per i gruppi" elenca i campi KPI per i gruppi individuati dall'analisi.

Tabella 2-3
KPI per i gruppi

Campo	Descrizione
GAG_GroupNumber	Identificatore univoco di un gruppo
GAG_Size	Numero di individui di un gruppo
GAG_Density	Frazione di collegamenti diretti tra gli individui di un gruppo. Per ulteriori informazioni, vedere l'argomento Densità della rete in il capitolo 1 a pag. 4.
GAG_KernelDensity	Frazione di collegamenti diretti tra gli individui principali di un gruppo
GAG_CoreNodesFraction	Frazione di individui di un gruppo che sono individui principali del gruppo. Per ulteriori informazioni, vedere l'argomento Suddivisione in gruppi a pag. 11.
GAG_MaxRankType1	Punteggio di autorità massimo per ogni membro del gruppo. Per ulteriori informazioni, vedere l'argomento Descrizione dei gruppi e dei membri a pag. 12.
GAG_MinRankType1	Punteggio di autorità minimo per ogni membro del gruppo.
GAG_MaxMinRankRatioType1	Rapporto tra il punteggio di autorità massimo e minimo. Il valore riflette la forza dell'autorità del leader del gruppo.

Campo	Descrizione
GAG_MaxRankType2	Punteggio di disseminazione massimo per ogni membro del gruppo. Per ulteriori informazioni, vedere l'argomento Descrizione dei gruppi e dei membri a pag. 12.
GAG_MinRankType2	Punteggio di disseminazione minimo per ogni membro del gruppo.
GAG_MaxMinRankRatioType2	Rapporto tra il punteggio di disseminazione massimo e minimo. Il valore riflette la forza di disseminazione del leader del gruppo.

La tabella “KPI per gli individui” elenca i campi KPI per gli individui della rete.

Tabella 2-4
KPI per gli individui

Campo	Descrizione
GAI_NodeNumber	Identificatore univoco di un individuo
GAI_CoreNode	Indica se l'individuo è un individuo principale all'interno del gruppo oppure no. Per ulteriori informazioni, vedere l'argomento Suddivisione in gruppi a pag. 11.
GAI_RankType1	Punteggio di autorità dell'individuo. Per ulteriori informazioni, vedere l'argomento Descrizione dei gruppi e dei membri a pag. 12.
GAI_RankOrderType1	Ordine di classificazione nel gruppo in base al punteggio di autorità
GAI_RankType2	Punteggio di disseminazione dell'individuo. Per ulteriori informazioni, vedere l'argomento Descrizione dei gruppi e dei membri a pag. 12.
GAI_RankOrderType2	Ordine di classificazione nel gruppo in base al punteggio di disseminazione
GAI_InDegree	Numero di relazioni in cui un individuo è la destinazione della relazione. Per ulteriori informazioni, vedere l'argomento Grado nodale in il capitolo 1 a pag. 5.
GAI_OutDegree	Numero di relazioni in cui un individuo è l'origine della relazione. Per ulteriori informazioni, vedere l'argomento Grado nodale in il capitolo 1 a pag. 5.
GAI_GroupLeaderType1	Indica se il nodo è un leader con autorità, il cui punteggio di leadership è derivato dai legami in entrata. Per ulteriori informazioni, vedere l'argomento Descrizione dei gruppi e dei membri a pag. 12.
GAI_GroupLeaderConfidenceType1	La confidenza che il nodo sia un leader con autorità.
GAI_GroupLeaderType2	Indica se il nodo è un leader disseminante, il cui punteggio di leadership è derivato dai legami in uscita. Per ulteriori informazioni, vedere l'argomento Descrizione dei gruppi e dei membri a pag. 12.
GAI_GroupLeaderConfidenceType2	La confidenza che il nodo sia un leader disseminante.

Analisi di diffusione

Panoramica dell'analisi di diffusione

All'interno di una rete sociale, l'analisi di diffusione rileva gli individui che sono maggiormente influenzati dagli altri individui della rete, quantificando l'effetto come **energia diffusa**. Il processo impiega un metodo di attivazione dell'allargamento nel quale un effetto si diffonde a ripetizione da alcuni nodi della rete ai nodi immediatamente circostanti diminuendo di intensità man mano che si sposta (Dasgupta, Singh, Viswanathan, Chakraborty, Mukherjea, Nanavati, e Joshi, 2008). Quando riceve l'energia, il nodo si attiva e trasmette una parte dell'energia ricevuta a quei vicini a cui è legato da una relazione diretta.

Il **fattore di allargamento** definisce la percentuale di energia trasmessa da un nodo attivato rispetto alla quantità conservata dal nodo. I nodi che riceveranno l'energia trasmetteranno a loro volta la stessa proporzione di energia ai loro vicini, determinando il progressivo esaurimento del processo di trasmissione. Grandi fattori di allargamento trasmettono maggiori quantità di energia, che arriverà a toccare nodi più distanti dai nodi originari prima che il processo si esaurisca completamente. Fattori di allargamento ridotti attivano processi di diffusione che si esauriscono velocemente e in cui l'energia trasmessa rimane relativamente vicina ai nodi iniziali.

La quantità totale di energia diffusa da un nodo attivato è distribuita tra tutti i nodi che fungono da destinazione di relazioni dirette con il nodo di partenza. La quantità ricevuta da ciascun nodo dipende dalla forza della relazione esistente con il nodo attivato. La porzione di energia trasmessa a un nodo specifico corrisponde al peso della relazione suddiviso per il totale dei pesi di tutte le relazioni in cui il nodo attivato funge da origine. Di conseguenza, i nodi vicini che godono di relazioni con pesi relativi maggiori riceveranno più energia dei vicini con pesi relativi inferiori.

Il processo di diffusione si interrompe quando si verifica una delle condizioni seguenti:

- i nodi attivati non sono l'origine di alcuna relazione diretta
- la quantità di energia trasmessa è inferiore alla **soglia di accuratezza**, cioè alla quantità minima di energia trasmessa che permette al processo di continuare
- il numero di iterazioni raggiunge il limite specificato

Quando la diffusione si interrompe, i nodi con la maggiore quantità di energia diffusa sono quelli più sensibili all'effetto che ha dato avvio al processo. Per esempio, se il processo parte da nodi caratterizzati dall'abbandono, i nodi che presentano la maggiore quantità di energia saranno quelli con la più alta propensione all'abbandono. Questi saranno i nodi su cui concentrarsi per scongiurare l'abbandono.

Esempio di analisi di diffusione

Consultare la rete illustrata il capitolo 1 a pag. 3. La rete comprende sette nodi con relazioni dirette di forza variabile.

La tabella “Esempio di processo di diffusione” illustra un processo di diffusione nella rete con fattore di allargamento di 0,80. All’inizio, il nodo A contiene tutta l’energia, a cui è assegnato arbitrariamente il valore di 1,00. Nella prima fase, il nodo A si attiva diffondendo a tre vicini a cui è legato da relazione diretta l’80% della sua energia e trattenendone il 20%. Il peso totale delle relazioni è di 100. La relazione con il nodo D corrisponde alla metà del valore totale pertanto questo nodo riceve metà dell’energia diffusa: 0,40. La relazione con il nodo B corrisponde al 20% del peso totale, generando un valore di energia diffusa pari a 0,16. Il nodo C riceve l’energia rimanente, 0,24, che rappresenta il 30% della quantità originata dal nodo A.

Tabella 3-1
Esempio di processo di diffusione

Fase	Nodo A	Nodo B	Nodo C	Nodo D	Nodo E	Nodo F	Nodo 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

Avendo ricevuto l’energia alla fase 1, i nodi B, C e D si attivano nella fase 2 e diffondono a loro volta energia ai loro vicini. Il nodo B non dispone di nodi di destinazione e non diffonde energia. Il nodo C, invece, ha un nodo di destinazione, G, a cui trasmette l’80% della sua energia. Infine, il nodo D trasmette l’80% della sua energia ai nodi E e F, dove F riceve il triplo di energia di E in virtù del maggiore peso della sua relazione con D.

Nella fase 3, si attivano i nodi E, F e G. I nodi E e G non dispongono di nodi di destinazione quindi nessuno di loro trasmette energia. Il nodo F trasmette l’80% della sua energia all’unico suo nodo di destinazione, G. Questo nodo ora conta una quantità di energia pari a 0,38 proveniente dai nodi C e F.

A questo punto il nodo G si attiva ma, non disponendo di nodi di destinazione, determina l’esaurimento del processo. L’energia che ha avuto origine nel nodo A si è diffusa nell’intera rete e G è il nodo che ne ha ricevuto la quantità maggiore. Se l’energia introdotta nella rete rappresenta il rischio di abbandono, il nodo G sarà quello maggiormente influenzato da un eventuale abbandono del nodo A.

Nodo Analisi di diffusione

Il nodo Analisi di diffusione, presente nella palette Input, propaga un effetto da un determinato gruppo di individui all’interno di una rete sociale. Facendo leva sulle relazioni esistenti nella rete, sarà possibile determinare gli individui che saranno maggiormente influenzati dall’effetto. Se l’effetto è l’abbandono, per esempio, il nodo individua le persone per cui la probabilità di abbandono è maggiore come reazione all’abbandono di altri individui della rete. I modelli e i dati esistenti possono essere ampliati tramite l’output del nodo per migliorare le previsioni generate dai modelli. Per esempio, è possibile perfezionare i valori di propensione all’abbandono di un individuo aggiungendo le influenze della diffusione.

Requisiti. Il nodo deve disporre di due file di testo a larghezza fissa contenenti i dati da analizzare. Il primo file definisce la rete sociale mediante tre campi. Un campo indica l’origine di ogni relazione diretta, un altro campo definisce la destinazione di ogni relazione diretta e il terzo

campo, opzionale, specifica la forza di ogni relazione. Tutte le relazioni presenti nella rete devono avere una direzione (origine-destinazione). Il secondo file contiene l'elenco degli identificatori da cui deve partire l'effetto.

Definizione dei dati per l'analisi di diffusione

La scheda Dati nella finestra del nodo di origine Analisi di diffusione permette di specificare il file di input contenente le relazioni esistenti nel nodo di rete.

File. Specificare il nome di un file o di una cartella contenente i record dei dati di chiamata. È possibile immettere un nome o fare clic sul pulsante con i puntini di sospensione (...) per selezionare un nome dal file system. Il percorso viene mostrato quando il nome viene selezionato; i relativi contenuti vengono visualizzati separati da delimitatori. Se si specifica una cartella, i record dei dati di chiamata di tutti i file della cartella vengono concatenati per l'analisi; tutti i file della cartella devono avere la medesima struttura.

Leggi i nomi dei campi dal file. Questa opzione è selezionata per default e consente di utilizzare i nomi dei campi inclusi nella prima riga nel file di dati come nomi per le colonne. Se la prima riga non è un'intestazione, deselezionare l'opzione per assegnare automaticamente un nome generico a ogni campo, come *Campo1* e *Campo2*.

Impostazioni definizione di rete

Le impostazioni definizione di rete stabiliscono i ruoli dei campi.

Campi. Usare i pulsanti freccia per assegnare manualmente gli elementi dell'elenco ai vari campi di ruolo. Le icone indicano i livelli di misurazione validi per ogni campo di ruolo. Fare clic sul pulsante Tutto per selezionare tutti i campi dell'elenco o fare clic su un singolo pulsante di livello di misurazione per selezionare tutti i campi con tale livello.

Origine. Selezionare un campo come origine della relazione diretta.

Destinazione. Selezionare un campo come destinazione della relazione diretta.

Peso. Selezionare il campo che rappresenta il peso della relazione. Per i record dei dati di chiamata, il peso può corrispondere a qualsiasi numero di caratteristiche della chiamata che riflettano la forza della relazione, come la durata o la frequenza della chiamata.

Impostazione delle opzioni di creazione per l'analisi di diffusione

La scheda Opzioni di creazione nella finestra del nodo di origine Analisi di diffusione permette di definire le opzioni del processo di diffusione.

Elenco semi di diffusione

L'elenco semi di diffusione indica i nodi della rete da cui deve partire la diffusione dell'energia. Per esempio, nel caso di un'analisi dell'abbandono, l'elenco definisce i nodi della rete che hanno già abbandonato. In alternativa, l'elenco potrebbe contenere i nodi per cui la propensione all'abbandono è maggiore così da esplorare gli effetti sulla rete nell'eventualità in cui si verificano gli abbandoni.

File. Specificare il nome del file di testo che contiene gli identificatori dei nodi (ogni identificatore corrisponde a una riga del file). È possibile immettere un nome di file o fare clic sul pulsante con i puntini di sospensione (...) per selezionare un file.

Impostazioni analisi di diffusione

Le impostazioni dell'analisi di diffusione determinano la pervasività della diffusione e il momento in cui il processo si interrompe.

Fattore di allargamento. Definisce la porzione di energia che un nodo trasferisce a ogni fase del processo di diffusione. Valori minori limitano la diffusione dell'energia ai nodi più vicini. Al contrario, valori maggiori permettono all'energia di toccare nodi più distanti.

Massimo numero di iterazioni. Specificare un limite massimo per il numero di iterazioni della diffusione. Il processo di diffusione termina automaticamente quando questo limite viene raggiunto.

Soglia di accuratezza. Specifica l'incremento minimo di energia del nodo che giustifica la ripetizione del processo di diffusione. Se l'incremento della quantità di energia di ciascun nodo è inferiore a questo valore, il processo di diffusione si interrompe.

Calcola e visualizza statistiche riassuntive. Quando questa opzione è selezionata, oltre a calcolare i KPI, il nodo calcola e visualizza statistiche riassuntive dell'analisi. Nelle reti estremamente grandi, il calcolo delle statistiche riassuntive ha un impatto negativo sulla velocità di elaborazione dei nodi.

Visualizzazione delle statistiche dell'analisi di diffusione

La scheda Analisi nella finestra del nodo di origine Analisi di diffusione contiene un riassunto dei risultati del processo di diffusione. La tabella "Statistiche riassuntive dell'analisi di diffusione" elenca tutte le statistiche riassuntive disponibili.

Tabella 3-2
Statistiche riassuntive dell'analisi di diffusione

Statistica	Descrizione
Nodi totali nella rete	Numero di nodi presenti nella rete
Legami totali nella rete	Numero di legami presenti nella rete
Semi di diffusione totali nella rete	Numero di nodi usati come semi dal processo di diffusione
Influenza media	Quantità media di energia diffusa associata agli individui.

Statistica	Descrizione
Legami in entrata medi	Numero medio di relazioni in cui un individuo è la destinazione della relazione. Per ulteriori informazioni, vedere l'argomento Grado nodale in il capitolo 1 a pag. 5.
Legami in uscita medi	Numero medio di relazioni in cui un individuo è l'origine della relazione. Per ulteriori informazioni, vedere l'argomento Grado nodale in il capitolo 1 a pag. 5.

Selezionare una singola statistica nella tabella Statistiche riassuntive per consultare la distribuzione dei valori, la deviazione standard e l'asimmetria della statistica.

Output dell'analisi di diffusione

Il nodo Analisi di diffusione genera una serie di campi che descrivono gli individui della rete. I modelli e i dati esistenti possono essere ampliati tramite i KPI per migliorare le previsioni generate dai modelli. Per esempio, è possibile perfezionare i valori di propensione all'abbandono di un individuo aggiungendo le influenze della diffusione.

La tabella "KPI per l'analisi di diffusione" elenca i campi KPI per gli individui sottoposti ad analisi.

Tabella 3-3
KPI per l'analisi di diffusione

Campo	Descrizione
DA_NodeNumber	Identificatore univoco di un individuo
DA_DiffusedEnergy	Quantità di energia diffusa associata all'individuo. Per l'analisi dell'abbandono, valori più alti indicano una maggiore propensione all'abbandono rispetto a valori bassi.
DA_InDegree	Numero di relazioni in cui un individuo è la destinazione della relazione. Per ulteriori informazioni, vedere l'argomento Grado nodale in il capitolo 1 a pag. 5.
DA_OutDegree	Numero di relazioni in cui un individuo è l'origine della relazione. Per ulteriori informazioni, vedere l'argomento Grado nodale in il capitolo 1 a pag. 5.

Note

Queste informazioni sono state preparate per prodotti e servizi offerti in tutto il mondo.

IBM potrebbe non offrire i prodotti, i servizi o le funzionalità di cui si tratta nel presente documento in altri paesi. Contattare il rappresentante IBM locale per informazioni sui prodotti e i servizi attualmente disponibili nella propria zona. Qualsiasi riferimento a un prodotto, programma o servizio IBM non intende dichiarare o implicare che sia possibile utilizzare esclusivamente tale prodotto, programma o servizio IBM. Potrà invece essere utilizzato qualsiasi prodotto, programma o servizio con funzionalità equivalente e che non violi i diritti di proprietà intellettuale di IBM. Tuttavia, è responsabilità dell'utente valutare e verificare il funzionamento di qualsiasi prodotto, programma o servizio non IBM.

IBM può essere titolare di brevetti o domande di brevetto relativi alla materia oggetto del presente documento. La consegna del presente documento non conferisce alcuna licenza rispetto a questi brevetti. Rivolgere per iscritto i quesiti sulle licenze a:

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

Il seguente paragrafo non si applica per il Regno Unito o altri paesi in cui le presenti disposizioni non sono conformi alle leggi locali: INTERNATIONAL BUSINESS MACHINES FORNISCE QUESTA PUBBLICAZIONE “COSÌ COM'È” SENZA GARANZIA DI ALCUN TIPO, SIA ESSA ESPRESSA O IMPLICITA, INCLUSE, MA NON LIMITATE A, LE GARANZIE IMPLICITE DI NON VIOLAZIONE, COMMERCIALIZZABILITÀ O IDONEITÀ A UNO SCOPO SPECIFICO. Alcuni stati non consentono limitazioni di garanzie espresse o implicite in determinate transazioni, pertanto quanto sopra potrebbe non essere applicabile.

Le presenti informazioni possono includere imprecisioni tecniche o errori tipografici. Le modifiche periodiche apportate alle informazioni contenute in questa pubblicazione verranno inserite nelle nuove edizioni della pubblicazione. IBM può apportare miglioramenti e/o modifiche al/ai prodotto/i e/o al/ai programma/i descritti nella presente pubblicazione in qualsiasi momento senza preavviso.

Qualsiasi riferimento nelle presenti informazioni a siti Web non IBM viene fornito esclusivamente per facilitare la consultazione e non rappresenta in alcun modo un'approvazione o sostegno da parte nostra di tali siti Web. I materiali contenuti in tali siti Web non fanno parte dei materiali di questo prodotto IBM e il loro utilizzo è esclusivamente a rischio dell'utente.

IBM può utilizzare o distribuire eventuali informazioni fornite dall'utente nei modi che ritiene appropriati senza incorrere in alcun obbligo nei confronti dell'utente.

I licenziatari del programma che desiderassero informazioni su di esso allo scopo di abilitare: (i) lo scambio di informazioni tra programmi creati indipendentemente e altri programmi (questo compreso) e (ii) l'utilizzo in comune delle informazioni scambiate, dovranno rivolgersi a:

IBM Software Group, All'attenzione di: Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA.

Tali informazioni saranno fornite in conformità ai termini e alle condizioni in vigore e, in alcuni casi, dietro pagamento.

Il programma concesso in licenza descritto nel presente documento e tutto il materiale correlato disponibile sono forniti da IBM in base ai termini del contratto di licenza cliente IBM, del contratto di licenza internazionale IBM o del contratto equivalente esistente tra le parti.

Le presenti informazioni includono esempi di dati e report utilizzati in operazioni aziendali quotidiane. Per fornire una descrizione il più possibile esaustiva, gli esempi includono nomi di persone, società, marchi e prodotti. Tutti questi nomi sono fittizi e ogni somiglianza a nomi e indirizzi utilizzati da aziende reali è puramente casuale.

Marchi

IBM, il logo IBM, ibm.com e SPSS sono marchi di IBM Corporation, registrati in numerose giurisdizioni nel mondo. Un elenco aggiornato dei marchi IBM è disponibile sul Web all'indirizzo <http://www.ibm.com/legal/copytrade.shtml>.

Linux è un marchio registrato di Linus Torvalds negli Stati Uniti e/o negli altri paesi.

Microsoft, Windows, Windows NT e il logo Windows sono marchi di Microsoft Corporation negli Stati Uniti e/o negli altri paesi.

UNIX è un marchio registrato di The Open Group negli Stati Uniti e in altri paesi.

Altri nomi di prodotti e servizi possono essere marchi commerciali di IBM o di altre aziende.

Glossario

densità. Misura della coesione di una rete definita dal numero di relazioni osservate suddiviso per il numero di relazioni possibili.

energia diffusa. Effetto propagato all'intera rete da nodi specifici la cui entità diminuisce all'aumentare della distanza dal nodo di partenza.

fattore di allargamento. Frazione di energia che un nodo trasferisce ai nodi vicini a ogni fase del processo di diffusione.

grado. Misura del livello di attività di un nodo in una rete definito come il numero totale di relazioni in cui un determinato nodo è coinvolto.

gruppo principale. Gruppo di nodi restante dopo l'esclusione delle relazioni più deboli della rete e l'applicazione dei limiti delle dimensioni del gruppo.

leader con autorità. Individuo del gruppo che presenta il punteggio di autorità più alto.

leader disseminante. Individuo del gruppo che presenta il punteggio di disseminazione più alto.

legami in entrata. Misura del prestigio di un nodo in una rete contenente relazioni dirette; rappresenta il numero di relazioni in cui il nodo funge da destinazione.

legami in uscita. Misura della centralità di un nodo in una rete contenente relazioni dirette; rappresenta il numero di relazioni in cui il nodo funge da origine.

punteggio di autorità. Misura di quanto gli altri membri del gruppo si collegano a un determinato individuo. I punteggi di autorità associano l'importanza di un individuo al numero di relazioni in cui tale individuo funge da destinazione.

punteggio di disseminazione. Misura di quanto un determinato individuo si collega agli altri membri del gruppo. I punteggi di disseminazione associano l'importanza di un individuo al numero di relazioni in cui tale individuo funge da origine.

relazione dicotomica. Connessione tra due nodi descritta da soli due valori possibili: presenza della connessione o assenza della connessione.

relazione direzionale. Connessione tra due nodi che parte da un nodo (l'origine) e arriva a un altro nodo (destinazione).

relazione di valore. Connessione tra due nodi che comprende anche un peso che descrive la forza della relazione.

relazione non direzionale. Associazione tra due nodi.

sociogramma. Rappresentazione di una rete sociale in cui le linee che collegano i punti rappresentano individui tra cui intercorre una relazione.

soglia di accuratezza. Criterio di interruzione del processo di diffusione corrispondente all'incremento minimo di energia del nodo che giustifica la ripetizione del processo di diffusione.

soglia di copertura. Frazione delle relazioni più forti usata per l'identificazione del gruppo.

Bibliografia

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

Dasgupta, K., R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjea, A. A. Nanavati, e 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, 668–677.

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

Richter, Y., E. Yom-Tov, e 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, 732–741.

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

Indice

- analisi dei dati, 7
- analisi di diffusione, 6, 17
 - fattore di allargamento, 17, 20
 - opzioni, 19
 - requisiti, 18
 - semi, 20
 - soglia di accuratezza, 17, 20
- analisi di gruppo, 6
 - dimensioni dei gruppi, 11, 14
 - requisiti, 13
 - soglia di copertura, 11, 14
- anteprima dell'output, 6

- campi di destinazione, 8, 13, 19
- campi di origine, 8, 13, 19
- campi peso, 8, 14, 19
- cancellazione dei risultati, 7

- densità, 4, 14–15
- densità del gruppo media, 14
- densità kernel, 15
- densità media del gruppo principale, 15
- dimensione del gruppo media, 14
- dimensioni dei gruppi, 14–15
 - limiti, 11, 14

- elenco semi di diffusione, 20
- energia diffusa, 20–21
- esecuzione di stream
 - utilizzo di IBM SPSS Modeler Server Social Network Analysis, 6

- fattore di allargamento, 17, 20
- filtro dell'output, 6
- frazione media membri principali, 15

- GAG_GroupNumber, 15
- grado, 5
- gruppi centrali, 11

- influenza media, 20

- KPI, indicatori di prestazioni chiave, 15, 21

- leader con autorità, 12, 16
- leader disseminanti, 12, 16
- legami in entrata, 5, 15–16, 21
- legami in entrata medi, 15, 21
- legami in uscita, 5, 15–16, 21
- legami in uscita medi, 15, 21
- livelli di misurazione, 6

- marchi, 23

- nodi principali, 15–16
- note legali, 22

- proprietà
 - script, 9
- punteggi di autorità, 12, 15–16
- punteggi di disseminazione, 12, 16

- record dei dati di chiamata, 8
- relazioni
 - di valore, 2
 - direzione, 2
- relazioni di valore, 2
- relazioni dicotomiche, 2
- relazioni direzionali, 2
- relazioni non direzionali, 2

- scheda Analisi, 6
 - analisi di diffusione, 20
 - analisi di gruppo, 14
- scheda Annotazioni, 6
- scheda Dati, 6
 - analisi di diffusione, 19
 - analisi di gruppo, 13
- scheda Filtro, 6
- scheda Opzioni di creazione, 6
 - analisi di diffusione, 19
 - analisi di gruppo, 14
- scheda Tipi, 6
- script
 - proprietà, 9
- semi
 - per l'analisi di diffusione, 20
- soglia di accuratezza, 17, 20
- soglia di copertura, 11, 14