

IBM SPSS Modeler Social Network  
Analysis 15 Manual del usuario



*Nota:* Antes de utilizar esta información y el producto que admite, lea la información general en Avisos el p. 24.

Esta edición se aplica a IBM SPSS Modeler Social Network Analysis 15 y a todas las versiones y modificaciones posteriores hasta que se indique lo contrario en nuevas ediciones.

Capturas de pantalla de productos de Adobe reimpresas con permiso de Adobe Systems Incorporated.

Capturas de pantalla de productos de Microsoft reimpresas con permiso de Microsoft Corporation.

Materiales bajo licencia: Propiedad de IBM

**© Copyright IBM Corporation 2012.**

Derechos restringidos para los usuarios del gobierno de Estados Unidos: Uso, duplicación o revelación restringidos por GSA ADP Schedule Contract con IBM Corp.

---

# Prefacio

IBM® SPSS® Modeler Social Network Analysis procesa información sobre las relaciones entre personas en campos que describen el papel de un individuo en una red social, permitiendo la inclusión de información social en modelos predictivos. Este manual describe el uso de los nodos de IBM SPSS Modeler Social Network Analysis en el entorno de IBM® SPSS® Modeler, lo que le permite incluir los nodos en sus flujos. Cuando el resultado de los nodos se combina con campos que representan medidas de individuos, se obtiene un perfil más completo de los individuos.

## ***Acerca de IBM Business Analytics***

El software IBM Business Analytics ofrece información completa, coherente y precisa en la que confían los encargados de la toma de decisiones para mejorar el rendimiento comercial. Un conjunto integral de [inteligencia empresarial](#), [análisis predictivo](#), [rendimiento financiero y gestión de estrategias](#) y [aplicaciones de análisis](#) que ofrece una perspectiva clara, inmediata e interactiva del rendimiento actual y la capacidad de predecir resultados futuros. En combinación con extensas soluciones sectoriales, prácticas probadas y servicios profesionales, las organizaciones de cualquier tamaño pueden conseguir el máximo de productividad, automatizar las decisiones de forma fiable y alcanzar mejores resultados.

Como parte de esta familia, el software de análisis predictivo de IBM SPSS ayuda a las organizaciones a predecir eventos futuros y actuar proactivamente según esa información para lograr mejores resultados comerciales. Los clientes comerciales, gubernamentales y académicos de todo el mundo confían en la tecnología de IBM SPSS como ventaja ante la competencia para atraer, retener y hacer crecer a los clientes, reduciendo al mismo tiempo el fraude y el riesgo. Al incorporar el software de IBM SPSS en sus operaciones diarias, las organizaciones se convierten en empresas predictivas, capaces de dirigir y automatizar decisiones para alcanzar los objetivos comerciales y lograr una ventaja considerable sobre la competencia. Para obtener más información o contactar con un representante, visite <http://www.ibm.com/spss>.

## ***Asistencia técnica***

Hay asistencia técnica disponible para los clientes de mantenimiento. Los clientes podrán ponerse en contacto con el servicio de asistencia técnica si desean recibir ayuda sobre la utilización de los productos de IBM Corp. o sobre la instalación en los entornos de hardware admitidos. Para ponerse en contacto con el servicio de asistencia técnica, consulte el sitio web de IBM Corp. en <http://www.ibm.com/support>. Tenga a mano su acuerdo de asistencia y esté preparado para identificarse a sí mismo y a su organización al solicitar ayuda.

---

# Contenido

## **1 IBM SPSS Modeler Social Network Analysis 1**

Acerca de IBM SPSS Modeler Social Network Analysis . . . . .	1
Acerca del análisis de redes sociales . . . . .	2
Visualización de redes . . . . .	3
Descripción de redes . . . . .	3
Nodos de IBM SPSS Modeler Social Network Analysis . . . . .	6
Pestañas de nodos . . . . .	7
Presentación preliminar del resultado . . . . .	7
Análisis de datos . . . . .	8
Aplicaciones . . . . .	8
Estructura de datos . . . . .	9
Propiedades de los procesos . . . . .	9

## **2 Análisis de grupos 11**

Visión general del análisis de grupos . . . . .	11
Determinación de la similitud social . . . . .	11
División en grupos . . . . .	12
Descripción de grupos y miembros de grupos . . . . .	13
Nodo Análisis de grupos . . . . .	14
Especificación de datos para el análisis de grupos . . . . .	14
Establecimiento de opciones de creación para el análisis de grupos . . . . .	15
Visualización de estadísticos del análisis de grupos . . . . .	15
Resultado del análisis de grupos . . . . .	16

## **3 Análisis de difusión 19**

Visión general del análisis de difusión . . . . .	19
Ejemplo de análisis de difusión . . . . .	20
Nodo Análisis de difusión . . . . .	20
Especificación de datos para el análisis de difusión . . . . .	21
Configuración de opciones de creación para el análisis de difusión . . . . .	22
Visualización de estadísticos del análisis de difusión . . . . .	22
Resultado del análisis de difusión . . . . .	23

## ***Apéndice***

<b><i>A Avisos</i></b>	<b><i>24</i></b>
<b><i>Glosario</i></b>	<b><i>26</i></b>
<b><i>Bibliografía</i></b>	<b><i>27</i></b>
<b><i>Índice</i></b>	<b><i>28</i></b>



# ***IBM SPSS Modeler Social Network Analysis***

## ***Acerca de IBM SPSS Modeler Social Network Analysis***

Muchos enfoques para la creación de modelos de comportamiento se centran en el individuo. Utilizan una variedad de datos sobre los individuos para generar un modelo que utilice los indicadores clave del comportamiento para predecirlo. Si algún individuo presenta determinados valores de los indicadores clave asociados a la aparición del comportamiento, se puede prestar especial atención a ese individuo con el fin de evitar dicho comportamiento.

Consideremos los enfoques para crear modelos de abandono, en el que un cliente finaliza su relación con una empresa. El coste de mantenimiento de clientes es significativamente inferior al coste de sustitución de clientes, lo que hace que la capacidad de identificar clientes con riesgo de abandono sea esencial. Un analista a menudo utiliza cierto número de indicadores clave de rendimiento para describir a los clientes, entre los que se incluyen información demográfica y patrones de llamadas recientes de cada cliente individual. Los modelos predictivos basados en estos campos utilizan cambios en los patrones de llamadas de los clientes que concuerden con patrones de llamadas de clientes que abandonaron en el pasado para identificar a las personas con un mayor riesgo de abandono. Los clientes a los que se identifique en una situación de riesgo reciben una atención al cliente u opciones de servicio adicionales para intentar mantenerlos.

Estos métodos pasan por alto información social que podría afectar significativamente al comportamiento de un cliente. Con las relaciones se transmite información sobre una empresa y sobre lo que hacen otras personas, la cual influye en los individuos. Como resultado, las relaciones con otros permiten que esas personas influyan en las decisiones y acciones de dicho individuo. Los análisis que sólo incluyen medidas individuales están omitiendo factores importantes que cuentan con capacidades predictivas.

IBM® SPSS® Modeler Social Network Analysis se enfrenta a este problema procesando información de relaciones en campos adicionales que pueden incluirse en los modelos. Estos indicadores clave de rendimiento derivados miden características sociales de los individuos. Al combinar estas propiedades sociales con medidas basadas en el individuo se obtiene una mejor visión general de los individuos y, por consiguiente, se puede mejorar la precisión predictiva de sus modelos.

IBM SPSS Modeler Social Network Analysis consta de dos componentes principales:

- Nodos de IBM SPSS Modeler Social Network Analysis añadidos al entorno de IBM® SPSS® Modeler que permiten incluir técnicas analíticas sociales en los flujos.
- IBM® SPSS® Modeler Server Social Network Analysis, que añade el procesamiento de las especificaciones de nodo a IBM® SPSS® Modeler Server. IBM SPSS Modeler Server Social Network Analysis procesa de manera eficaz enormes cantidades de datos de red, que

pueden incluir millones de individuos y relaciones, en un número relativamente bajo de campos para realizar análisis posteriores.

Por ejemplo, IBM SPSS Modeler Social Network Analysis identifica a los individuos de una red que se ven más afectados por el abandono de personas específicas. Además, puede detectar a grupos de individuos de una red con un mayor riesgo de abandono. Al incorporar indicadores clave de rendimiento a estos efectos en sus modelos, puede mejorar su rendimiento global.

*Nota:* El procesamiento de los nodos de IBM SPSS Modeler Social Network Analysis debe realizarse mediante IBM SPSS Modeler Server Social Network Analysis. Al ejecutar un flujo que contenga nodos de IBM SPSS Modeler Social Network Analysis en SPSS Modeler, debe conectarse a una instancia de SPSS Modeler Server que también incluya IBM SPSS Modeler Server Social Network Analysis.

## ***Acerca del análisis de redes sociales***

Una red social se compone de un conjunto de individuos y las relaciones entre ellos. El análisis de redes sociales examina estas relaciones para describir a los individuos y a los grupos como parte de una estructura social. Los individuos interactúan entre sí y estos patrones de interacción ofrecen información sobre los individuos implicados. Las relaciones permiten el flujo de información a través de una red, lo que hace que un individuo pueda influir en otro. La importancia de la información de las relaciones distingue al análisis de redes sociales de otros enfoques. En lugar de centrarse en cada individuo por separado, la unidad de estudio es una díada compuesta por dos individuos y sus relaciones.

Las relaciones de una red pueden clasificarse como direccionales o no direccionales. En una **relación direccional**, un individuo se identifica como el iniciador, u origen, de la relación y el otro se identifica como el receptor, o destinatario. Por ejemplo, realizar una llamada de teléfono es una relación direccional en la que una persona llama a otra. Por el contrario, en las **relaciones no direccionales** no pueden definirse claramente los papeles de iniciador y destinatario. En este caso, ambas partes participan en la relación en la misma medida. Mantener una conversación es un ejemplo de relación no direccional.

Otra propiedad que distingue una relación de otra es si la relación es dicotómica o valorada. La única información disponible en una **relación dicotómica** es si ésta existe o no entre dos individuos. Por cada díada de la red, la relación está presente o ausente. Por otro lado, una **relación valorada** incluye una ponderación que indica la fuerza de la relación. Las ponderaciones permiten comparar las relaciones entre sí.

La tabla “Tipos de relaciones” muestra ejemplos de la clasificación cruzada de relaciones según dirección y escala. En las relaciones direccionales, *Juan* es el iniciador de la relación y *María* es la destinataria. En las relaciones no direccionales, no hay ninguna indicación de quién inició la relación. Las relaciones valoradas utilizan la duración de la conversación como ponderación de la relación, mientras que las relaciones dicotómicas simplemente si sucedieron o no.

Tabla 1-1  
Tipos de relaciones

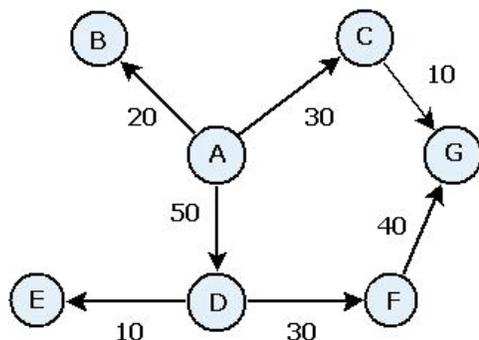
Dirección	Escala	Ejemplo
No direccional	Dicotómica	Juan y María mantuvieron una conversación.
No direccional	Valorada	Juan y María mantuvieron una conversación de 20 minutos.
Direccional	Dicotómica	Juan llamó a María.
Direccional	Valorada	Juan llamó a María y mantuvieron una conversación de 20 minutos.

Para obtener más información sobre el campo del análisis de redes sociales, consulte uno de los completos libros de esta área (Wasserman y Faust, 1994).

## Visualización de redes

Una red social suele ilustrarse mediante un **sociograma** (Moreno, 1934). En este tipo de visualización, los individuos se corresponden con puntos, o nodos, en un espacio. Las líneas, o bordes, que conectan los puntos representan las relaciones entre los individuos. Si las relaciones son direccionales, los bordes incluyen una flecha para indicar la dirección. Si las relaciones tienen ponderaciones, las etiquetas de los bordes indican los valores. El siguiente gráfico muestra una red de siete individuos.

Figura 1-1  
Red social de ejemplo



Supongamos que la red representa las llamadas de teléfono realizadas por individuos y que las ponderaciones de las relaciones indican la duración de las llamadas. En este caso, la persona A llamó a tres personas y pasó la mayor parte del tiempo hablando con la persona D.

Esta red es mucho más pequeña que las que se dan en la práctica. Sin embargo, los conceptos ilustrados en sociogramas sencillos son generales para redes de cualquier tamaño y complejidad.

## Descripción de redes

Es necesario extraer la información sobre redes, grupos e individuos en características descriptivas que permitan comparaciones cruzadas y la inclusión en modelos predictivos. Las redes deben reducirse a un conjunto finito de indicadores clave de rendimiento que puedan analizarse. Por

ejemplo, puede que quiera comparar redes o grupos de nodos de una red entre sí. Asimismo, puede que quiera comparar individuos de la red entre sí o identificar a los individuos más importantes.

Dos medidas que suelen utilizarse para describir las redes sociales son la **densidad** y el **grado**. Ambos estadísticos reflejan la conectividad, pero el primero se centra en la totalidad de la red o en subgrupos de la red mientras que el segundo caracteriza a los individuos de la red.

### **Densidad de red**

Para cualquier conjunto de nodos de una red, hay un número finito de relaciones posible. Cada nodo puede servir como origen o destino de una relación con respecto a cualquier otro nodo. Consideremos una red compuesta por tres nodos: A, B y C. La siguiente tabla muestra todas las posibles relaciones dirigidas entre los nodos.

Tabla 1-2  
*Posibles relaciones dirigidas de los tres nodos*

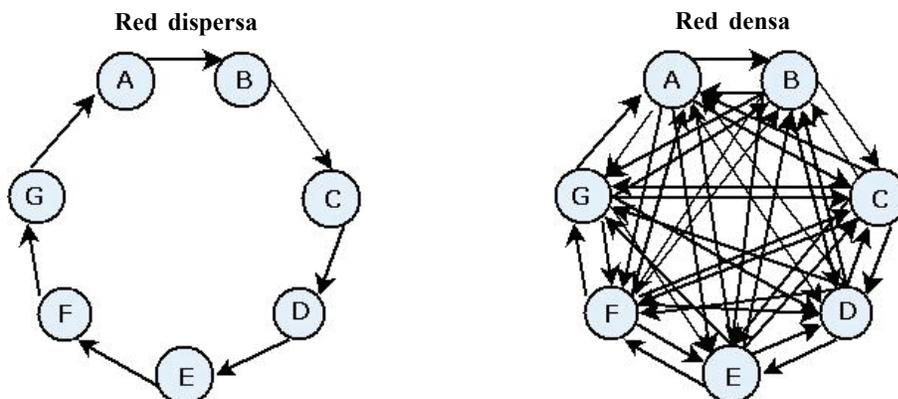
<b>Origen</b>	<b>Destino</b>
A	B
A	C
B	A
B	C
C	A
C	B

Cada nodo es el origen de una relación con los otros dos nodos. Sin embargo, en la práctica puede que todas las posibles relaciones no estén presentes. Puede que algunos nodos no tengan una relación dirigida con otros nodos. Además, puede que algunas relaciones dirigidas no sean recíprocas.

El estadístico **densidad** representa la proporción de posibles relaciones en la red que de hecho están presentes. El valor va de 0 a 1, donde el límite inferior se corresponde con redes sin relaciones y el límite superior representa redes con todas las posibles relaciones. Cuanto más cerca esté el valor a 1, mayor será la densidad de la red y sus nodos estarán más unidos.

La información de redes densas puede fluir más fácilmente que la información de redes dispersas. La tabla “Redes dispersas y densas” muestra dos redes compuestas por siete nodos. La red dispersa sólo incluye siete de las 42 posibles relaciones entre los nodos, con lo que su densidad es de 0,17. Por el contrario, la red densa contiene todas las posibles relaciones y tiene una densidad de 1,0.

Tabla 1-3  
Redes dispersas y densas



En la red dispersa, para que la información fluya del nodo A al nodo G, debe pasar a través de otros cinco nodos. Por el contrario, en la red densa la información puede ir directamente del nodo A al nodo G.

### Grado nodal

Los individuos importantes de una red suelen ser los implicados en el mayor número de relaciones. Estos individuos adquieren información de una variedad de fuentes y distribuyen dicha información entre un gran número de individuos diferentes. Por el contrario, los individuos que participan en pocas relaciones no pueden influir directamente en un gran número de personas de la red.

El **grado** de un nodo, definido como el número total de relaciones en las que participa dicho nodo, permite realizar comparaciones entre los participantes de la red. Los individuos con valores superiores de grado son más activos que los que tienen valores inferiores. El grado ignora la dirección de las relaciones, lo que proporciona una medida global de la actividad del nodo.

En el caso de relaciones dirigidas, puede centrarse en si un nodo es origen o destino al determinar el número de relaciones. El **grado de entrada** de un nodo es el número de relaciones en las que un nodo específico es el destino. A la inversa, el **grado de salida** es el número de relaciones en las que un nodo es el origen. La siguiente tabla muestra los valores de grado, grado de entrada y grado de salida de cada nodo en la ilustración “Red social de ejemplo”.

Tabla 1-4  
Valores de grado, grado de entrada y grado de salida de ejemplo

Nodo	Grado	Grado de entrada	Grado de salida
A	3	0	3
B	1	1	0
C	2	1	1
D	3	1	2
E	1	1	0

Nodo	Grado	Grado de entrada	Grado de salida
F	2	1	1
G	2	2	0

A menudo el grado de entrada se considera una medida de prestigio. Los valores más elevados del grado de entrada se corresponden con un mayor número de relaciones que finalizan en ese nodo. Dicho de otro modo, un gran número de individuos se pone en contacto con estas personas. Otros muchos nodos inician relaciones con este nodo. A la inversa, el grado de salida se considera una medida de centralidad. Los valores más elevados se corresponden con un mayor número de relaciones cuyo origen es ese nodo. Esos individuos se ponen en contacto con un gran número de personas diferentes.

En el caso de los nodos de la red de ejemplo, los valores de grado indican que los nodos A y D son los más activos mientras que los nodos B y E son los menos activos. Los valores de grado de entrada revelan que el nodo G tiene el mayor prestigio. Según los valores de grado de salida, el nodo A es el más central.

## Nodos de IBM SPSS Modeler Social Network Analysis

Junto con los diversos nodos estándar proporcionados por IBM® SPSS® Modeler, también puede trabajar con nodos de IBM® SPSS® Modeler Social Network Analysis para incluir los resultados del análisis de redes sociales en sus flujos. La tabla “Nodos de IBM SPSS Modeler Social Network Analysis” describe estos nodos, que se almacenan en la paleta Fuentes.

Tabla 1-5  
Nodos de IBM SPSS Modeler Social Network Analysis

Nodo	Icono	Descripción
Análisis de grupos		El nodo Análisis de grupos importa datos de registros de detalles de llamadas de un archivo de texto de campos fijos, identifica grupos de nodos en la red definidos por los registros y genera indicadores clave de rendimiento para los grupos e individuos de la red. Si desea obtener más información, consulte el tema <a href="#">Visión general del análisis de grupos</a> en el capítulo 2 el p. 11.
Análisis de difusión		El nodo Análisis de difusión importa datos de registros de detalles de llamadas de un archivo de texto de campos fijos, propaga un efecto por la red definido por los registros y genera indicadores clave de rendimiento que resumen los resultados del efecto sobre nodos individuales. Si desea obtener más información, consulte el tema <a href="#">Visión general del análisis de difusión</a> en el capítulo 3 el p. 19.

El procesamiento de los nodos de IBM SPSS Modeler Social Network Analysis debe realizarse mediante IBM® SPSS® Modeler Server Social Network Analysis. Al ejecutar un flujo que contenga estos nodos en SPSS Modeler, debe conectarse a una instancia de IBM® SPSS® Modeler Server que también incluya IBM SPSS Modeler Server Social Network Analysis.

## ***Pestañas de nodos***

Los nodos de IBM® SPSS® Modeler Social Network Analysis ofrecen las siguientes pestañas para definir y obtener una presentación preliminar del análisis:

- **Pestaña Datos.** Utilizada para identificar el archivo que contiene la información de la red social.
- **Pestaña Opciones de creación.** Utilizada para definir los ajustes de los análisis.
- **Pestaña Análisis.** Utilizada para ver un resumen del resultado preliminar que sirve de guía para modificar los ajustes de entrada en la pestaña Datos para producir resultados óptimos.

Además, los nodos ofrecen las siguientes pestañas comunes entre los nodos de origen de IBM® SPSS® Modeler:

- **Pestaña Filtro.** Utilizada para eliminar o cambiar el nombre de los campos de resultados producidos por los nodos. Esta pestaña ofrece la misma funcionalidad que el nodo Filtro.
- **Pestaña Tipos.** Utilizada para establecer los niveles de medida de los campos de resultados producidos por los nodos. Esta pestaña ofrece la misma funcionalidad que el nodo Tipo.
- **Pestaña Anotaciones.** Utilizada para cambiar el nombre de los nodos, proporcionar información sobre herramientas personalizada y almacenar una anotación de gran tamaño.

Para obtener más información sobre las pestañas comunes, consulte la documentación de SPSS Modeler.

## ***Presentación preliminar del resultado***

Normalmente, el resultado de los nodos se utiliza como entrada para un modelo predictivo. Si quiere ver el resultado, puede añadir un nodo Tabla al flujo. Sin embargo, la cantidad de datos suele ser extremadamente grande, por lo que se tarda mucho tiempo en generar la tabla. Además, el número de filas de la tabla limita su utilidad.

Como alternativa, puede generar una tabla de presentación preliminar en la que se visualice una muestra del resultado que se creará. La presentación preliminar muestra los campos generados para un número limitado de filas. El número de filas se define en las propiedades de flujo. Para obtener una presentación preliminar del resultado de los nodos, realice los pasos siguientes:

1. Abra el nodo.
2. Especifique la configuración de los datos en la pestaña Datos.
3. Defina los parámetros analíticos en la pestaña Opciones de creación.
4. Pulse en Presentación preliminar.

Aparecerá una ventana de presentación preliminar con los resultados. Además, al obtener una presentación preliminar del resultado, se cumplimenta la pestaña Análisis del nodo con una visión general de resumen de los resultados.

## **Análisis de datos**

Normalmente determinar la configuración analítica que produce resultados óptimos es un proceso repetitivo. El usuario define la configuración, ejecuta el análisis y revisa los resultados. Si los resultados no son todo lo útiles que debieran ser, el usuario modifica la configuración y vuelve a ejecutar el análisis.

Para analizar los datos de entrada del nodo, realice los pasos siguientes:

1. Abra el nodo.
2. Especifique la configuración de los datos en la pestaña Datos.
3. Defina los parámetros analíticos en la pestaña Opciones de creación.
4. Pulse en Analizar datos.

Si las opciones de creación indican que deben mostrarse los estadísticos de resumen, la pestaña Análisis mostrará los resultados.

Si necesita volver a ejecutar un análisis, pulse en Borrar análisis para eliminar los resultados actuales antes de pulsar en Analizar datos.

## **Aplicaciones**

Entre las aplicaciones específicas para las que IBM® SPSS® Modeler Social Network Analysis puede ser especialmente útil se incluyen las siguientes:

- **Predicción de abandono.** Las características de grupo pueden influir en las tasas de abandono. Al centrarse en individuos de grupos con un mayor riesgo de abandono, quizás pueda evitarlo. Además, puede identificar a individuos con riesgo de abandono gracias al flujo de información de los que ya han abandonado.
- **Aprovechamiento de líderes de grupo.** Los líderes de grupo influyen en gran medida en los otros miembros del grupo. Si se puede evitar el abandono de un líder de grupo, puede que se reduzca la tasa de abandono de los miembros del grupo. Si no, intentar que un líder de grupo de la competencia abandone puede aumentar la tasa de abandono de los miembros del grupo asociado a dicho competidor a la vez que se reduce la tasa de abandono de los miembros del grupo asociado a su empresa.
- **Marketing.** Se puede utilizar a los líderes de grupo para presentar nuevos bienes u ofertas de servicio. La influencia del líder puede aumentar la probabilidad de que otros miembros del grupo adquieran la oferta. Puede utilizar el análisis de difusión para identificar a los individuos sobre los que los líderes de grupo tengan una mayor influencia y centrar su estrategia de marketing en ellos.

## Estructura de datos

La información sobre los individuos de la red de interés puede distribuirse a través de una variedad de archivos, bases de datos y sistemas en toda su empresa. Para analizar la red mediante IBM® SPSS® Modeler Social Network Analysis, debe extraer los registros y campos relevantes de sus fuentes de datos y aplicarles el formato necesario para introducirlos en los nodos.

Los nodos analíticos requieren que los registros de detalles de llamadas se almacenen en un único archivo de texto de ancho fijo. Cada fila del archivo se corresponde con una relación, con los datos organizados en las siguientes columnas:

- El identificador del individuo que inicia la relación.
- El identificador del individuo que es el destino de la relación.
- Una ponderación opcional de la relación.

Todos los datos deben ser numéricos y los identificadores de los individuos deben ser únicamente enteros. La inclusión de nombres de campos en la primera fila de datos del archivo es opcional. La tabla “Registros de datos de llamadas de ejemplo” ilustra esta estructura de datos.

Tabla 1-6  
Registros de datos de llamadas de ejemplo

Origen	Destino	Ponderación
1000	5642	243
2190	8444	831
0299	9419	559

Los valores de ponderación pueden corresponderse con cualquier medida que quiera utilizar para representar la importancia de la relación con respecto a las otras relaciones de la red. En el caso de los datos de llamadas, las ponderaciones comunes incluyen la duración o la frecuencia de las llamadas.

Si desea que el análisis se centre en un subconjunto del historial de llamadas, debe utilizar dicho subconjunto al crear el archivo de entrada. Por ejemplo, puede limitar el análisis a los últimos meses o a las llamadas más recientes de un individuo incluyendo únicamente esos datos en el archivo de texto de entrada.

## Propiedades de los procesos

El procesamiento en IBM® SPSS® Modeler Social Network Analysis es una herramienta potente para automatizar procesos en la interfaz de usuario. Los procesos pueden realizar los mismos tipos de acciones que se realizan con el ratón o el teclado y se utilizan para automatizar tareas que resultarían extremadamente repetitivas o llevarían mucho tiempo si se realizaran manualmente. Para obtener una explicación del uso del procesamiento, consulte la guía *ScriptingAutomation.pdf* que se incluye con IBM® SPSS® Modeler.

**Propiedades del nodo de difusión**

La tabla siguiente muestra las propiedades de procesamiento del nodo de difusión.

Nombre de la propiedad	Tipo de datos	Descripción de la propiedad
input_data_file_name	<i>cadena</i>	
calling_field	<i>campo</i>	
called_field	<i>campo</i>	
frequency_weight_field	<i>campo</i>	
read_field_names	<i>booleano</i>	
diffusion_list_file_name	<i>cadena</i>	
spreading_factor	<i>double (doble)</i>	Valor por defecto = 50. Mín. = 1 Máx. = 99.
max_number_iterations	<i>entero</i>	Valor por defecto = 100. Mín. = 1.
accuracy_threshold	<i>double (doble)</i>	Valor por defecto = 0.01. Mín. = 0.001.
calculate_statistics	<i>booleano</i>	

**Propiedades del nodo Análisis de grupos**

La tabla siguiente muestra las propiedades de procesamiento del nodo Análisis de grupos.

Nombre de la propiedad	Tipo de datos	Descripción de la propiedad
input_data_file_name	<i>cadena</i>	
calling_field	<i>campo</i>	
called_field	<i>campo</i>	
frequency_weight_field	<i>campo</i>	
read_field_names	<i>booleano</i>	
coverage_threshold	<i>double (doble)</i>	Valor por defecto = 10. Mín. = 1 Máx. = 99.
min_group_size	<i>entero</i>	Valor por defecto = 2. Mín. = 2.
max_group_size	<i>entero</i>	Valor por defecto = 100. Mín. = 2.
calculate_statistics	<i>booleano</i>	

# ***Análisis de grupos***

## ***Visión general del análisis de grupos***

El análisis de grupos utiliza los patrones de interacción de los individuos de una red para identificar grupos de individuos similares. Las características de estos grupos influyen en el comportamiento de los miembros individuales de los grupos. Por ejemplo, los grupos pequeños con muchas relaciones entre los miembros y líderes influyentes tienen un mayor riesgo de abandono, aunque de hecho ningún miembro del grupo haya abandonado. (Richter, Yom-Tov, y Slonim, 2010) Los modelos predictivos que incorporen medidas individuales y de grupo tendrán un mejor rendimiento que los modelos que sólo incluyan las primeras.

El análisis de grupos se compone de los siguientes pasos generales:

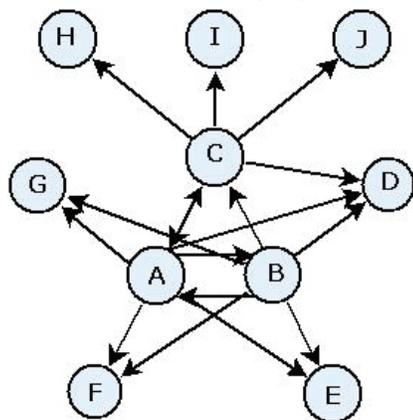
1. Determine las fuerzas de las relaciones que reflejan la proximidad social. Si desea obtener más información, consulte el tema [Determinación de la similitud social](#) el p. 11.
2. Divida la red en grupos según la fuerza de la relación a la vez que respeta las restricciones de tamaño. Si desea obtener más información, consulte el tema [División en grupos](#) el p. 12.
3. Cree perfiles tanto de los grupos como de los individuos e identifique a los líderes de grupo. Si desea obtener más información, consulte el tema [Descripción de grupos y miembros de grupos](#) el p. 13.

## ***Determinación de la similitud social***

Los miembros de un grupo deberían parecerse más entre ellos que a los individuos que no pertenezcan al grupo. En el análisis de redes, la similitud de dos nodos depende de sus relaciones. Para cualquier nodo de una red, existe un conjunto de nodos que son el destino de relaciones dirigidas desde el nodo. En el caso de los datos de telecomunicaciones, este conjunto se corresponde con todas las personas con las que se ha puesto en contacto un determinado individuo. Si dos individuos tienen contacto con el mismo conjunto de personas, esos individuos se consideran similares entre sí. Cuanto más se solapan los conjuntos de destinos de las relaciones de dos individuos, mayor será su similitud.

Consideremos la red mostrada en la ilustración “[Red de diez nodos de ejemplo](#)”.

Figura 2-1  
Red de diez nodos de ejemplo



La tabla “**Nodos de destino**” muestra los nodos de destino de las relaciones cuyo origen está en los nodos A, B y C.

Tabla 2-1  
Nodos de destino

Nodo de origen	Nodos de destino
A	B, C, D, E, F, G
B	A, C, D, E, F, G
C	A, D, H, I, J

Los nodos A y B tienen cinco nodos de destino comunes. Por otra parte, los nodos A y C sólo tienen uno. Por lo tanto, los nodos A y B son más parecidos socialmente que los nodos A y C.

Las ponderaciones de las relaciones como la duración o la frecuencia de las llamadas no captan la similitud de los nodos de una red. Para incluir la noción de similitud, el análisis de grupos utiliza información mutua (Cover y Thomas, 2006) como ponderación de la relación. Este estadístico refleja la probabilidad de que dos nodos estén conectados al mismo nodo. Si las relaciones de la red tienen ponderaciones predefinidas, como la frecuencia de las llamadas en los datos de telecomunicaciones, la información mutua incorpora esas ponderaciones en consecuencia.

## **División en grupos**

Los grupos deberían estar compuestos por individuos que tengan altas ponderaciones de las relaciones con otros miembros del grupo, donde las ponderaciones midan la similitud de los nodos en la relación. Como resultado, la identificación de un grupo comienza mediante la omisión de las relaciones más débiles de una red. El **umbral de cobertura** controla este proceso definiendo la fracción de las relaciones más sólidas que debe mantenerse. Por ejemplo, un umbral de cobertura con un valor de 0,4 da como resultado que se utilice el 40% de las relaciones más sólidas para la identificación de un grupo, mientras que el 60% restante de las relaciones se omite.

Las relaciones restantes pueden constituir grupos muy pequeños o muy grandes, que presentan una utilidad predictiva limitada. Para evitar la inclusión de dichos grupos en el análisis, puede especificar tamaños mínimos y máximos para los grupos. Los grupos que sean más pequeños que el tamaño mínimo se omitirán por completo. Sin embargo, los grupos que sean más grandes que el tamaño máximo se dividirán en grupos más pequeños dentro del rango de tamaños aceptable. Los grupos restantes tras eliminar las relaciones más débiles y aplicar los límites de tamaño se denominan **grupos principales**.

La eliminación de relaciones de la red original puede provocar que algunos individuos no se incluyan en ningún grupo principal. Sin embargo, puede que esos individuos tengan conexiones con miembros de un grupo que garanticen su inclusión en el grupo. Un individuo se añade a un grupo si hay muchas relaciones relativamente sólidas con los miembros principales del grupo, siempre que se respeten los límites de tamaño del grupo. De este modo, los grupos definitivos se componen de un conjunto de miembros principales y otros miembros añadidos debido a su conexión con los miembros principales.

### ***Descripción de grupos y miembros de grupos***

Además de la densidad, el grado de entrada y el grado de salida, otros estadísticos describen la dinámica de grupos. En particular, las puntuaciones de autoridad y diseminación ofrecen medidas del estado social de los individuos dentro de los grupos. El papel de cada individuo dentro de un grupo es de vital importancia cuando se trata de predecir el comportamiento del grupo y de sus miembros.

La **puntuación de autoridad** de un nodo mide la tendencia de otros nodos del grupo a conectarse a él. Si muchos individuos se ponen en contacto con una persona en particular, posiblemente pidiendo información o su opinión, esa persona tiene un papel de autoridad. Las puntuaciones de autoridad de los nodos de un grupo, que se corresponden con las probabilidades estacionarias de un paseo aleatorio con reinicios a través de la red del grupo, van de 0 a 1. Cuanto más cerca esté la puntuación de autoridad a 1, mayor será la autoridad que el nodo tiene dentro del grupo. El nodo del grupo con la mayor puntuación de autoridad se denomina **líder de autoridad** del grupo. La división de la mayor puntuación de un grupo por la más pequeña proporciona una medida de la fuerza general del líder de autoridad.

Por el contrario, la **puntuación de diseminación** de un nodo mide la tendencia del nodo a conectarse a otros nodos del grupo. Si una persona en particular se pone en contacto con otras muchas personas del grupo, esa persona puede afectar significativamente a las opiniones de todo el grupo. Las puntuaciones de diseminación de los nodos de un grupo, que se corresponden con las probabilidades estacionarias de un paseo aleatorio con reinicios a través de la red del grupo a la inversa, van de 0 a 1. Cuanto más cerca esté la puntuación de diseminación a 1, más se conectará el nodo a los otros miembros del grupo. El nodo del grupo con la mayor puntuación de diseminación se denomina **líder de diseminación** del grupo. La división de la mayor puntuación de un grupo entre la más pequeña proporciona una medida de la fuerza general del líder de diseminación.

## **Nodo Análisis de grupos**

El nodo Análisis de grupos, que está disponible en la paleta Fuentes, identifica conjuntos de individuos de una red que son parecidos socialmente entre sí y determina el estado social relativo de los individuos de los grupos. Los grupos que son relativamente pequeños, tienen muchas conexiones entre sus miembros e incluyen a un líder influyente tienen un mayor riesgo de abandono debido a la dinámica de grupos. Puede utilizar los resultados para determinar las puntuaciones de abandono de grupos que, cuando se combinan con puntuaciones de abandono individuales, mejoran la capacidad de predecir el abandono en modelos basados sólo en puntuaciones individuales.

Dada la cantidad de datos de la que suele estar compuesta la red y la naturaleza de las interacciones sociales que subyacen en las definiciones de grupos, el análisis de grupos es un proceso que requiere mucho tiempo. Por lo general, los resultados del análisis de grupos suelen guardarse en una base de datos o archivo para su uso como entrada para un modelo predictivo. Los resultados de grupos se actualizarían con relativamente poca frecuencia, por ejemplo, mensualmente, mientras que el modelo podría actualizarse con mucha más frecuencia.

**Requisitos.** El nodo requiere un archivo de texto de ancho fijo que defina la red social mediante tres campos. Un campo identifica el origen de cada relación dirigida, otro campo define el destino de cada relación dirigida y el tercero especifica una fuerza opcional para cada relación. Todas las relaciones de la red deben ser direccionales.

### **Especificación de datos para el análisis de grupos**

La pestaña Datos de la ventana del nodo de origen Análisis de grupos le permite especificar el archivo de entrada que contiene las relaciones de los nodos de la red.

**Archivo.** Especifique el nombre del archivo o la carpeta que contiene los registros de detalles de llamadas. Puede introducir un nombre o pulsar en el botón de puntos suspensivos (...) para seleccionar un nombre en el sistema de archivos. La ruta se mostrará cuando haya seleccionado un nombre y su contenido se mostrará con los delimitadores. Si especifica una carpeta, los registros de detalles de llamadas de todos los archivos de la carpeta estarán concatenados para el análisis; todos los archivos de la carpeta deben tener la misma estructura.

**Lea los nombres de campo del archivo.** Seleccionada de manera predeterminada, esta opción trata la primera fila del archivo de datos como nombres para las columnas. Si la primera fila no es un encabezado, cancele la selección de esta opción para proporcionar un nombre genérico a cada campo automáticamente, como *Campo1* y *Campo2*.

### **Configuración de la definición de red**

La configuración de la definición de red define los papeles de los campos.

**Campos.** Utilice los botones de flecha para asignar elementos de esta lista manualmente a los diversos campos de papeles. Los iconos indican los niveles de medidas válidos para cada campo de papel. Pulse en el botón Todos para seleccionar todos los campos de la lista o pulse en un botón de nivel de medida individual para seleccionar todos los campos con ese nivel de medida.

**Origen.** Seleccione un campo como origen de la relación dirigida.

**Destino.** Seleccione un campo como destino de la relación dirigida.

**Ponderación.** De forma opcional, seleccione un campo correspondiente al número de veces que se produzca la relación en la red. El análisis pondera los registros en consecuencia al determinar las fuerzas de las relaciones.

## ***Establecimiento de opciones de creación para el análisis de grupos***

La pestaña Opciones de creación de la ventana del nodo de origen Análisis de grupos le permite definir opciones para identificar grupos en la red.

### ***Configuración del análisis de grupos***

La configuración del análisis de grupos influye en el tamaño y la fuerza relativa de los grupos formados.

**Umbral de cobertura.** Defina la fracción de las relaciones de la red más sólidas que debe utilizarse en el análisis. Por ejemplo, un umbral de cobertura de 0,2 indica que sólo se utilizarán las relaciones cuyas ponderaciones estén dentro del 20% que incluye las mejores de todas las ponderaciones. Este parámetro, que va de 0 a 1, permite que el análisis se centre en las relaciones más sólidas, que deberían producirse dentro de los grupos. Los valores de umbral superiores producen grupos con relaciones más débiles entre sus miembros. Si desea obtener más información, consulte el tema [División en grupos](#) el p. 12.

**Tamaño mínimo del grupo.** Especifique un límite inferior para el tamaño del grupo. Los grupos con un valor inferior a éste no se incluirán.

**Tamaño máximo del grupo.** Especifique un límite superior para el tamaño del grupo. Los grupos con un valor superior a éste se dividirán en grupos más pequeños.

**Calcula y muestra estadísticas de resumen.** Si se selecciona, el nodo calcula y muestra estadísticas de resumen para el análisis además de derivar el resultado de los indicadores clave de rendimiento. El cálculo de estos estadísticos puede afectar negativamente al rendimiento del nodo en redes extremadamente grandes.

## ***Visualización de estadísticos del análisis de grupos***

La pestaña Análisis de la ventana del nodo de origen Análisis de grupos ofrece una visión general de resumen de los grupos identificados en la red. La tabla “Estadísticos de resumen del análisis de grupos” muestra los estadísticos de resumen disponibles para los grupos.

Tabla 2-2  
Estadísticos de resumen del análisis de grupos

Estadístico	Descripción
Número total de nodos en los grupos	Número de nodos incluidos en los grupos identificados.
Número total de enlaces en los grupos	Número de enlaces incluidos en los grupos identificados.
Número total de grupos	Número de grupos identificados en la red.
Tamaño medio del grupo	Número medio de nodos de un grupo.
Densidad media del grupo	Fracción media de conexiones directas entre nodos de un grupo. Si desea obtener más información, consulte el tema <a href="#">Densidad de red</a> en el capítulo 1 el p. 4.
Fracción media de miembros principales	Fracción media de nodos de un grupo que son nodos principales del grupo. Si desea obtener más información, consulte el tema <a href="#">División en grupos</a> el p. 12.
Densidad media del grupo principal	Fracción media de conexiones directas entre nodos principales de un grupo.
Grado de entrada medio	Número medio de enlaces entrantes. Si desea obtener más información, consulte el tema <a href="#">Grado nodal</a> en el capítulo 1 el p. 5.
Grado de salida medio	Número medio de enlaces salientes. Si desea obtener más información, consulte el tema <a href="#">Grado nodal</a> en el capítulo 1 el p. 5.

Seleccione un estadístico específico de la tabla de estadísticos de resumen para ver la distribución de los valores, la desviación típica y la asimetría del estadístico.

La revisión de estos estadísticos ayuda a definir la configuración del análisis de grupos. Teóricamente, los resultados deberían mostrar un número relativamente pequeño de grupos con valores de densidad altos. Por ejemplo, si hay varios grupos muy grandes, considere reducir el tamaño máximo de grupo para dividirlos en grupos más pequeños. Asimismo, si los valores de densidad de los grupos tienden a ser bajos, considere reducir el umbral de cobertura para centrarse en relaciones más sólidas.

## Resultado del análisis de grupos

El nodo Análisis de grupos genera una variedad de campos que describen los grupos y los individuos de los grupos. Puede aumentar los modelos y los datos existentes con estos indicadores clave de rendimiento para mejorar las predicciones generadas por sus modelos. Por ejemplo, puede actualizar los valores de propensión de abandono individuales para incluir las influencias de grupos.

La tabla “Indicadores clave de rendimiento para grupos” muestra los campos de indicadores clave de rendimiento para los grupos identificados en el análisis.

Tabla 2-3  
Indicadores clave de rendimiento para grupos

Campo	Descripción
GAG_GroupNumber	Identificador exclusivo de un grupo.
GAG_Size	Número de individuos de un grupo.
GAG_Density	Fracción de conexiones directas entre individuos de un grupo. Si desea obtener más información, consulte el tema <a href="#">Densidad de red</a> en el capítulo 1 el p. 4.

<b>Campo</b>	<b>Descripción</b>
GAG_KernelDensity	Fracción de conexiones directas entre individuos principales de un grupo.
GAG_CoreNodesFraction	Fracción de individuos de un grupo que son individuos principales del grupo. Si desea obtener más información, consulte el tema <a href="#">División en grupos</a> el p. 12.
GAG_MaxRankType1	Puntuación de autoridad máxima de cualquier miembro de grupo. Si desea obtener más información, consulte el tema <a href="#">Descripción de grupos y miembros de grupos</a> el p. 13.
GAG_MinRankType1	Puntuación de autoridad mínima de cualquier miembro de grupo.
GAG_MaxMinRankRatioType1	Proporción de la puntuación de autoridad mayor a la menor. Este valor refleja la fuerza de autoridad del líder de grupo.
GAG_MaxRankType2	Puntuación de diseminación máxima de cualquier miembro de grupo. Si desea obtener más información, consulte el tema <a href="#">Descripción de grupos y miembros de grupos</a> el p. 13.
GAG_MinRankType2	Puntuación de diseminación mínima de cualquier miembro de grupo.
GAG_MaxMinRankRatioType2	Proporción de la puntuación de diseminación mayor a la menor. Este valor refleja la fuerza de diseminación del líder de grupo.

La tabla “Indicadores clave de rendimiento para individuos” muestra los campos de indicadores clave de rendimiento para los individuos de la red.

Tabla 2-4  
Indicadores clave de rendimiento para individuos

<b>Campo</b>	<b>Descripción</b>
GAI_NodeNumber	Identificador exclusivo de un individuo.
GAI_CoreNode	Indicador de si el individuo es principal para un grupo o no. Si desea obtener más información, consulte el tema <a href="#">División en grupos</a> el p. 12.
GAI_RankType1	Puntuación de autoridad del individuo. Si desea obtener más información, consulte el tema <a href="#">Descripción de grupos y miembros de grupos</a> el p. 13.
GAI_RankOrderType1	Orden de rango del grupo según las puntuaciones de autoridad.
GAI_RankType2	Puntuación de diseminación del individuo. Si desea obtener más información, consulte el tema <a href="#">Descripción de grupos y miembros de grupos</a> el p. 13.
GAI_RankOrderType2	Orden de rango del grupo según las puntuaciones de diseminación.
GAI_InDegree	Número de relaciones en las que el individuo es el destino de la relación. Si desea obtener más información, consulte el tema <a href="#">Grado nodal</a> en el capítulo 1 el p. 5.
GAI_OutDegree	Número de relaciones en las que el individuo es el origen de la relación. Si desea obtener más información, consulte el tema <a href="#">Grado nodal</a> en el capítulo 1 el p. 5.
GAI_GroupLeaderType1	Si el nodo es un líder de autoridad, cuya puntuación de liderazgo se deriva de los enlaces entrantes. Si desea obtener más información, consulte el tema <a href="#">Descripción de grupos y miembros de grupos</a> el p. 13.
GAI_GroupLeaderConfidenceType1	La confianza de que el nodo es un líder de autoridad.

<b>Campo</b>	<b>Descripción</b>
GAI_GroupLeaderType2	Si el nodo es un líder de diseminación, cuya puntuación de liderazgo se deriva de los enlaces salientes. Si desea obtener más información, consulte el tema <a href="#">Descripción de grupos y miembros de grupos</a> el p. 13.
GAI_GroupLeaderConfidenceType2	La confianza de que el nodo es un líder de diseminación.

# **Análisis de difusión**

## **Visión general del análisis de difusión**

El análisis de difusión identifica a los individuos que se ven más afectados por otros individuos en una red social, cuantificando el efecto como **energía difundida**. El proceso utiliza un enfoque de activación de dispersión, en el que un efecto se dispersa de forma iterativa desde los nodos de una red por sus elementos afines inmediatos, disminuyendo su tamaño a medida que pasa de un nodo a otro (Dasgupta, Singh, Viswanathan, Chakraborty, Mukherjee, Nanavati, y Joshi, 2008). Al recibir energía, un nodo se activa y transmite una parte de dicha energía a los elementos afines que sean el destino de relaciones dirigidas con el nodo.

El **factor de dispersión** define la proporción de energía transmitida por un nodo activado, quedando la cantidad restante en el nodo. Los nodos que reciban esta energía transmitirán ellos mismos la misma proporción a sus elementos afines, lo que da como resultado un proceso de debilitación de la transmisión de energía. Los factores de dispersión de gran tamaño se corresponden con un mayor envío de energía, lo que permite que la energía llegue a nodos más distantes de los nodos activados inicialmente antes de que el proceso se debilite por completo. Los factores de dispersión de tamaño pequeño dan como resultado procesos de difusión que se debilitan rápidamente y en los que la energía transmitida permanece relativamente cerca de los nodos iniciales.

La cantidad total de energía difundida por un nodo activado se distribuye entre todos los nodos que sean el destino de relaciones dirigidas con el nodo. La cantidad que recibe cada nodo depende de la fuerza de la relación con el nodo activado. La fracción enviada a un nodo específico equivale a la ponderación de la relación dividida entre el total de ponderaciones de todas las relaciones de las que el nodo activado es el origen. Por consiguiente, los elementos afines que tengan relaciones con ponderaciones relativas superiores recibirán más energía que los elementos afines con ponderaciones relativas inferiores.

El proceso de difusión se detiene cuando se da una de las siguientes condiciones:

- Los nodos activados no son el origen de ninguna relación dirigida.
- La cantidad de energía transmitida está por debajo del **umbral de precisión**, el límite sobre la cantidad de energía transmitida que debe cumplirse para que el proceso continúe.
- El número de iteraciones alcanza un límite especificado.

Cuanto termine la difusión, los nodos con mayor cantidad de energía difundida serán los más sensibles al efecto que inició el proceso. Por ejemplo, si el proceso comienza en los nodos que abandonan, los nodos con mayor energía serán los que presenten un mayor riesgo de abandono. Puede prestar especial atención a esos nodos para evitar su abandono.

### **Ejemplo de análisis de difusión**

Consideremos la red mostrada en el capítulo 1 el p. 3. La red se compone de siete nodos con relaciones dirigidas de fuerzas variables entre sí.

La tabla “Proceso de difusión de ejemplo” ilustra un proceso de difusión a través de la red con un factor de dispersión de 0,80. Inicialmente, el nodo A contiene toda la energía, a la que se ha asignado un valor arbitrario de 1,00. En el paso 1, este nodo se activa, dispersando el 80% de su energía a los tres elementos afines que son el destino de las relaciones a la vez que conserva el 20% para sí mismo. Las relaciones tienen una ponderación total de 100. La relación con el nodo D representa la mitad de este total, de modo que este nodo recibe la mitad de la energía difundida: 0,40. La relación con el nodo B representa el 20% de la ponderación total, lo que constituye una energía difundida con un valor de 0,16. El nodo C recibe la energía restante, 0,24, que es el 30% de la cantidad cuyo origen está en el nodo A.

Tabla 3-1  
Proceso de difusión de ejemplo

Paso	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

Al haber recibido energía en el paso 1, los nodos B, C y D se activan en el paso 2 y difunden energía a los elementos afines de destino. El nodo B no tiene ningún nodo de destino, de modo que no difunde energía. Por otra parte, el nodo C tiene un destino y difunde el 80% de su energía a dicho nodo, G. Por último, el nodo D dispersa el 80% de su energía a los nodos E y F. El nodo F recibe tres veces la cantidad recibida por el nodo E debido a que su relación con D es más sólida.

En el paso 3, los nodos E, F y G se activan. Los nodos E y G no tienen nodos de destino, así que ninguno de ellos difunde energía. Sin embargo, el nodo F difunde el 80% de su energía a su único nodo de destino, G. Este nodo tiene un valor total de energía de 0,38 compuesto por energía difundida por los nodos C y F.

En este punto, el nodo G se activa, pero como no tiene ningún nodo de destino al que difundir energía el proceso termina. La energía cuyo origen está en el nodo A se ha distribuido por la red, siendo el nodo G el que ha recibido la mayor cantidad. Si la energía introducida en la red representa el abandono, el nodo G sería el más afectado por el abandono del nodo A.

### **Nodo Análisis de difusión**

El nodo Análisis de difusión, que está disponible en la paleta Fuentes, propaga un efecto desde un conjunto específico de individuos por una red social, utilizando las relaciones de la red para identificar a los individuos más afectados por el efecto. Si, por ejemplo, el efecto es el abandono, el nodo identifica a los individuos con mayor probabilidad de abandonar debido al abandono de otros individuos específicos de la red. Puede aumentar los modelos y los datos existentes con el resultado de los nodos para mejorar las predicciones generadas por esos modelos. Por ejemplo,

puede actualizar los valores de propensión de abandono individuales para incluir las influencias de difusión.

**Requisitos.** El nodo requiere dos archivos de texto de ancho fijo con los datos que van a analizarse. El primer archivo define la red social mediante tres campos. Un campo identifica el origen de cada relación dirigida, otro campo define el destino de cada relación dirigida y el tercero especifica una fuerza opcional para cada relación. Todas las relaciones de la red deben ser direccionales. El segundo archivo contiene una lista de identificadores desde los que debería iniciarse el efecto.

### ***Especificación de datos para el análisis de difusión***

La pestaña Datos de la ventana del nodo de origen Análisis de difusión le permite especificar el archivo de entrada que contiene las relaciones de los nodos de la red.

**Archivo.** Especifique el nombre del archivo o la carpeta que contiene los registros de detalles de llamadas. Puede introducir un nombre o pulsar en el botón de puntos suspensivos (...) para seleccionar un nombre en el sistema de archivos. La ruta se mostrará cuando haya seleccionado un nombre y su contenido se mostrará con los delimitadores. Si especifica una carpeta, los registros de detalles de llamadas de todos los archivos de la carpeta estarán concatenados para el análisis; todos los archivos de la carpeta deben tener la misma estructura.

**Lea los nombres de campo del archivo.** Seleccionada de manera predeterminada, esta opción trata la primera fila del archivo de datos como nombres para las columnas. Si la primera fila no es un encabezado, cancele la selección de esta opción para proporcionar un nombre genérico a cada campo automáticamente, como *Campo1* y *Campo2*.

#### ***Configuración de la definición de red***

La configuración de la definición de red define los papeles de los campos.

**Campos.** Utilice los botones de flecha para asignar elementos de esta lista manualmente a los diversos campos de papeles. Los iconos indican los niveles de medidas válidos para cada campo de papel. Pulse en el botón Todos para seleccionar todos los campos de la lista o pulse en un botón de nivel de medida individual para seleccionar todos los campos con ese nivel de medida.

**Origen.** Seleccione un campo como origen de la relación dirigida.

**Destino.** Seleccione un campo como destino de la relación dirigida.

**Ponderación.** Seleccione un campo que represente la ponderación de la relación. En el caso de registros de detalles de llamadas, la ponderación puede corresponder a cualquier número de características de llamadas que reflejen la fuerza de la relación, como la duración o la frecuencia de las llamadas.

## **Configuración de opciones de creación para el análisis de difusión**

La pestaña Opciones de creación de la ventana del nodo de origen Análisis de difusión le permite definir opciones para el proceso de difusión.

### **Lista de semillas de difusión**

La lista de semillas de difusión identifica los nodos de la red desde los que la energía debería empezar a difundirse. Por ejemplo, en el análisis del abandono, esta lista identifica los nodos de la red que ya han abandonado. Asimismo, esta lista puede contener los nodos con el mayor riesgo de abandono para examinar los efectos que tendría en la red su abandono real.

**Archivo.** Especifique el nombre del archivo de texto que contiene los identificadores de nodos, donde cada identificador se corresponde con una línea del archivo. Puede introducir un nombre de archivo o pulsar en el botón de puntos suspensivos (...) para seleccionar un archivo.

### **Configuración del análisis de difusión**

La configuración del análisis de difusión determina el grado de agresividad de la difusión y cuándo se detiene el proceso.

**Factor de dispersión.** Defina la fracción de energía que un nodo transfiere en cada paso del proceso de difusión. Los valores más pequeños limitan la energía a los nodos más cercanos al nodo que está difundiendo la energía. A la inversa, los valores más elevados permiten que la energía afecte a nodos más distantes.

**Número máximo de iteraciones.** Especifique un límite superior para el número de iteraciones de difusión. El proceso de difusión termina automáticamente si se alcanza este límite.

**Umbral de precisión.** Especifique el cambio más pequeño en la energía del nodo que garantice la continuación del proceso de difusión. Si el cambio en la energía de cada nodo es más pequeño que este valor, el proceso de difusión termina.

**Calcula y muestra estadísticas de resumen.** Si se selecciona, el nodo calcula y muestra estadísticos de resumen para el análisis además de derivar el resultado de los indicadores clave de rendimiento. El cálculo de estos estadísticos puede afectar negativamente al rendimiento del nodo en redes extremadamente grandes.

## **Visualización de estadísticos del análisis de difusión**

La pestaña Análisis de la ventana del nodo de origen Análisis de difusión ofrece una visión general de resumen de los resultados de difusión. La tabla “Estadísticos de resumen del análisis de difusión” muestra los estadísticos de resumen disponibles.

Tabla 3-2  
*Estadísticos de resumen del análisis de difusión*

<b>Estadístico</b>	<b>Descripción</b>
Número total de nodos en la red	Número de nodos en la red.
Número total de enlaces en la red	Número de enlaces en la red.

Estadístico	Descripción
Número total de semillas de difusión en la red	Número de nodos utilizados como semillas para el proceso de difusión.
Influencia media	Cantidad media de energía difundida asociada a individuos.
Grado de entrada medio	Número medio de relaciones en las que un individuo es el destino de la relación. Si desea obtener más información, consulte el tema <a href="#">Grado nodal</a> en el capítulo 1 el p. 5.
Grado de salida medio	Número medio de relaciones en las que un individuo es el origen de la relación. Si desea obtener más información, consulte el tema <a href="#">Grado nodal</a> en el capítulo 1 el p. 5.

Seleccione un estadístico específico de la tabla de estadísticos de resumen para ver la distribución de los valores, la desviación típica y la asimetría del estadístico.

### **Resultado del análisis de difusión**

El nodo Análisis de difusión genera una variedad de campos que describen a los individuos de la red. Puede aumentar los modelos y los datos existentes con estos indicadores clave de rendimiento para mejorar las predicciones generadas por sus modelos. Por ejemplo, puede actualizar los valores de propensión de abandono individuales para incluir las influencias de difusión.

La tabla “Indicadores clave de rendimiento del análisis de difusión” muestra los campos de indicadores clave de rendimiento para los individuos del análisis.

**Tabla 3-3**  
*Indicadores clave de rendimiento del análisis de difusión*

Campo	Descripción
DA_NodeNumber	Identificador exclusivo de un individuo.
DA_DiffusedEnergy	Cantidad de energía difundida asociada al individuo. En el caso del análisis del abandono, los valores superiores indican una mayor propensión de abandono que los valores inferiores.
DA_InDegree	Número de relaciones en las que el individuo es el destino de la relación. Si desea obtener más información, consulte el tema <a href="#">Grado nodal</a> en el capítulo 1 el p. 5.
DA_OutDegree	Número de relaciones en las que el individuo es el origen de la relación. Si desea obtener más información, consulte el tema <a href="#">Grado nodal</a> en el capítulo 1 el p. 5.

# Avisos

Esta información se ha desarrollado para los productos y servicios ofrecidos en todo el mundo.

Puede que IBM no ofrezca los productos, los servicios o las características de los que se habla en este documento en otros países. Consulte a su representante local de IBM para obtener información acerca de los productos y servicios que está disponibles actualmente en su zona. Toda referencia que se haga de un producto, programa o servicio de IBM no implica que sólo se deba utilizar ese producto, programa o servicio de IBM. En su lugar, puede utilizarse todo producto, programa o servicio con funcionalidades equivalentes que no infrinjan los derechos de propiedad intelectual de IBM. Sin embargo, es responsabilidad del usuario evaluar y comprobar el funcionamiento de todo producto, programa o servicio que no sea de IBM.

IBM puede tener patentes o aplicaciones de patentes pendientes que cubren el asunto descrito en este documento. Este documento no le otorga ninguna licencia para estas patentes. Puede enviar preguntas acerca de las licencias, por escrito, a:

*IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785, Estados Unidos*

**El párrafo siguiente no se aplica a los Reino Unido o cualquier otro país donde tales disposiciones son incompatibles con la legislación local:** INTERNATIONAL BUSINESS MACHINES PROPORCIONA ESTA PUBLICACIÓN “TAL CUAL” SIN GARANTÍA DE NINGÚN TIPO, YA SEA EXPRESA O IMPLÍCITA, INCLUYENDO, PERO NO LIMITADA A, LAS GARANTÍAS IMPLÍCITAS DE NO INFRACCIÓN, COMERCIALIZACIÓN O IDONEIDAD PARA UN PROPÓSITO PARTICULAR. Algunos estados no permiten la renuncia a expresar o a garantías implícitas en determinadas transacciones, por lo tanto, esta declaración no se aplique a usted.

Esta información puede incluir imprecisiones técnicas o errores tipográficos. Periódicamente, se efectúan cambios en la información aquí y estos cambios se incorporarán en nuevas ediciones de la publicación. IBM puede realizar mejoras y/o cambios en los productos y/o los programas descritos en esta publicación en cualquier momento sin previo aviso.

Cualquier referencia a sitios Web que no sean de IBM en esta información sólo es ofrecida por comodidad y de ningún modo sirve como aprobación de esos sitios Web. Los materiales en esos sitios Web no forman parte del material de este producto de IBM y el uso de estos sitios Web es bajo su propio riesgo.

IBM puede utilizar cualquier información que le suministre en cualquier forma que considere adecuada, sin incurrir en ninguna obligación para usted.

Los licenciarios de este programa que deseen tener información sobre el mismo con el objetivo de habilitar: (i) el intercambio de información entre programas creados independientemente y otros programas (incluyendo este) y (ii) el uso común de la información que se ha intercambiado, deben ponerse en contacto con:

*IBM Software Group, a la atención de: Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA.*

---

Esta información estará disponible, bajo las condiciones adecuadas, incluyendo en algunos casos el pago de una cuota.

IBM proporciona el programa bajo licencia que se describe en este documento y todo el material bajo licencia disponible para el mismo bajo los términos de IBM Customer Agreement (Acuerdo de cliente de IBM), IBM International Program License Agreement (Acuerdo de licencia de programa internacional de IBM) o cualquier acuerdo equivalente entre las partes.

Esta información contiene ejemplos de datos e informes utilizados en operaciones comerciales diarias. Para ilustrarlos lo máximo posible, los ejemplos incluyen los nombres de las personas, empresas, marcas y productos. Todos esos nombres son ficticios y cualquier parecido con los nombres y direcciones utilizados por una empresa real es pura coincidencia.

### ***Marcas comerciales***

IBM, el logotipo de IBM, ibm.com y SPSS son marcas comerciales de IBM Corporation, registradas en muchas jurisdicciones de todo el mundo. Existe una lista actualizada de marcas comerciales de IBM en Internet en <http://www.ibm.com/legal/copytrade.shtml>.

Linux es una marca registrada de Linus Torvalds en Estados Unidos, otros países o ambos.

Microsoft, Windows, Windows NT, y el logotipo de Windows son marcas comerciales de Microsoft Corporation en Estados Unidos, otros países o ambos.

UNIX es una marca registrada de The Open Group en Estados Unidos y otros países.

Otros productos y nombres de servicio pueden ser marcas comerciales de IBM u otras empresas.

---

# Glosario

**densidad.** Medida de cohesión de red definida como el número de relaciones observadas dividido entre el número de posibles relaciones.

**energía difundida.** Efecto cuyo origen está en nodos específicos de una red que se distribuye por toda la red, disminuyendo su tamaño a medida que la distancia desde los nodos iniciales se hace mayor.

**factor de dispersión.** Fracción de energía que un nodo transfiere a sus elementos afines en un paso del proceso de difusión.

**grado.** Medida de la actividad de un nodo de una red, definida como el número total de relaciones en las que participa dicho nodo.

**grado de entrada.** Medida de prestigio de un nodo de una red compuesta por relaciones dirigidas, definida como el número total de relaciones en las que el nodo es el destino.

**grado de salida.** Medida de centralidad de un nodo de una red compuesta por relaciones dirigidas, definida como el número total de relaciones en las que el nodo es el origen.

**grupo principal.** Grupo de nodos que permanecen tras omitir las relaciones más débiles de una red y aplicar los límites de tamaño de grupo.

**líder de autoridad.** Individuo de un grupo con la mayor puntuación de autoridad.

**líder de diseminación.** Individuo de un grupo con la mayor puntuación de diseminación.

**puntuación de autoridad.** Medida que indica en qué grado se conectan los otros miembros de un grupo a un individuo. Las puntuaciones de autoridad asocian la importancia de un individuo al número de relaciones que finalizan en él.

**puntuación de diseminación.** Medida que indica en qué grado se conecta un individuo a los otros miembros de un grupo. Las puntuaciones de diseminación asocian la importancia de un individuo al número de relaciones cuyo origen es él mismo.

**relación dicotómica.** Conexión entre dos nodos que sólo puede tomar un valor de entre dos, indicando la presencia o la ausencia de la conexión.

**relación direccional.** Conexión entre dos nodos cuyo origen es uno de ellos y el destino es el otro.

**relación no direccional.** Asociación entre dos nodos.

**relación valorada.** Conexión entre dos nodos con una ponderación asociada que indica la fuerza de la relación.

**sociograma.** Representación visual de una red social en la que las líneas que conectan los puntos se corresponden con individuos que tienen relaciones entre sí.

**umbral de cobertura.** Fracción de las relaciones más sólidas utilizada para la identificación de grupo.

**umbral de precisión.** Criterio de parada para un proceso de difusión correspondiente al cambio más pequeño en la energía del nodo que garantiza la continuación de la difusión.

---

# ***Bibliografía***

Cover, T. M., y 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, y A. Joshi. 2008. Social ties and their relevance to churn in mobile telecom networks. En: *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, y N. Slonim. 2010. Predicting customer churn in mobile networks through analysis of social groups. En: *Proceedings of the 2010 SIAM international conference on data mining*, Columbus, OH: SDM 2010, 732–741.

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

---

# Índice

- análisis de datos, 8
- análisis de difusión, 6, 20
  - factor de dispersión, 19, 22
  - opciones, 21
  - requisitos, 21
  - semillas, 22
  - umbral de precisión, 19, 22
- análisis de grupos, 6
  - requisitos, 14
  - tamaños de grupos, 13, 15
  - umbral de cobertura, 12, 15
- avisos legales, 24
  
- borrado de resultados, 8
  
- campos de destino, 9, 15, 21
- campos de origen, 9, 15, 21
- campos de ponderación, 9, 15, 21
  
- densidad, 4, 16
- densidad de kernel, 16–17
- densidad media del grupo, 16
- densidad media del grupo principal, 16
  
- ejecución de flujos
  - uso de IBM SPSS Modeler Server Social Network Analysis, 6
- energía difundida, 23
  
- factor de dispersión, 19, 22
- filtrado de resultados, 7
- fracción media de miembros principales, 16
  
- GAG\_GroupNumber, 16
- grado, 5
- grado de entrada, 5, 16–17, 23
- grado de entrada medio, 16, 23
- grado de salida, 5, 16–17, 23
- grado de salida medio, 16
- Grado de salida medio, 23
- grupos principales, 13
  
- indicadores clave de rendimiento, 16, 23
- influencia media, 23
  
- líderes de autoridad, 13, 17
- líderes de diseminación, 13, 18
- lista de semillas de difusión, 22
  
- marcas comerciales, 25
  
- niveles de medida, 7
- nodos principales, 16–17
  
- pestaña análisis, 7
  - análisis de difusión, 22
  - análisis de grupos, 15
- pestaña anotaciones, 7
- pestaña datos, 7
  - análisis de difusión, 21
  - análisis de grupos, 14
- pestaña filtro, 7
- pestaña opciones de creación, 7
  - análisis de difusión, 22
  - análisis de grupos, 15
- pestaña tipos, 7
- presentación preliminar del resultado, 7
- procesos
  - propiedades, 9
- propiedades
  - procesos, 9
- puntuaciones de autoridad, 13, 17
- puntuaciones de diseminación, 13, 17
  
- registros de detalles de llamadas, 9
- relaciones
  - dirección, 2
  - valoradas, 2
- relaciones dicotómicas, 2
- relaciones direccionales, 2
- relaciones no direccionales, 2
- relaciones valoradas, 2
  
- semillas
  - para el análisis de difusión, 22
  
- tamaño medio del grupo, 16
- tamaños de grupos, 16
  - límites, 13, 15
  
- umbral de cobertura, 12, 15
- umbral de precisión, 19, 22