

IBM SPSS Modeler Social Network  
Analysis 15 Guide de l'utilisateur



*Remarque* : Avant d'utiliser ces informations et le produit qu'elles concernent, lisez les informations générales sous Avis sur p. 24.

Cette version s'applique à IBM SPSS Modeler Social Network Analysis 15 et à toutes les publications et modifications ultérieures jusqu'à mention contraire dans les nouvelles versions.

Les captures d'écran des produits Adobe sont reproduites avec l'autorisation de Adobe Systems Incorporated.

Les captures d'écran des produits Microsoft sont reproduites avec l'autorisation de Microsoft Corporation.

Matériel sous licence - Propriété d'IBM

**© Copyright IBM Corporation 2012.**

Droits limités pour les utilisateurs au sein d'administrations américaines : utilisation, copie ou divulgation soumise au GSA ADP Schedule Contract avec IBM Corp.

---

# Préface

IBM® SPSS® Modeler Social Network Analysis traite des informations sur les relations entre les individus dans des champs décrivant le rôle d'un individu dans un réseau social, permettant ainsi d'intégrer les informations sociales aux modèles prédictifs. Ce manuel décrit l'utilisation des noeuds IBM SPSS Modeler Social Network Analysis dans l'environnement IBM® SPSS® Modeler vous permettant ainsi d'intégrer les noeuds dans vos flux. La combinaison du résultat des noeuds et des champs qui représentent les mesures sur les individus permet d'obtenir un profil plus complet des individus.

## ***A propos d'IBM Business Analytics***

Le logiciel IBM Business Analytics propose des informations complètes, cohérentes et précises auxquelles les preneurs de décisions peuvent se fier pour améliorer les performances de leur entreprise. Un porte-feuilles étendu de [veille économique](#), d'[analyses prédictives](#), de [gestion des performances et de stratégie financières](#) et d'[applications analytiques](#) vous offre des informations claires, immédiates et décisionnelles sur les performances actuelles et vous permet de prévoir les résultats futurs. Ce logiciel intègre des solutions dédiées à l'industrie, des pratiques éprouvées et des services professionnels qui permettent aux organisations de toute taille de maximiser leur productivité, d'automatiser leurs décisions sans risque et de proposer de meilleurs résultats.

Ce porte-feuilles intègre le logiciel IBM SPSS Predictive Analytics qui aide les organisations à prévoir les événements à venir et à réagir en fonction des informations afin d'améliorer leurs résultats. Les clients de l'industrie du commerce, de l'éducation et des administrations du monde entier font confiance à la technologie IBM SPSS qui offre un avantage concurrentiel en attirant et fidélisant les clients et en améliorant la base de données de la clientèle tout en diminuant la fraude et en réduisant les risques. En utilisant le logiciel IBM SPSS dans leurs opérations quotidiennes, les organisations deviennent des entreprises prédictives, capables de diriger et d'automatiser les décisions pour répondre aux objectifs commerciaux et obtenir un avantage concurrentiel mesurable. Pour des informations supplémentaires ou pour joindre un représentant, consultez <http://www.ibm.com/spss>.

## ***Assistance technique***

L'assistance technique est disponible pour les clients du service de maintenance. Les clients peuvent contacter l'assistance technique pour obtenir de l'aide concernant l'utilisation des produits IBM Corp. ou l'installation dans l'un des environnements matériels pris en charge. Pour contacter l'assistance technique, rendez-vous sur le site Web IBM Corp. à l'adresse <http://www.ibm.com/support>. Lorsque vous contactez l'assistance technique, soyez prêt à indiquer votre identité, le nom de votre société et votre contrat d'assistance.

---

# Contenu

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

À propos de IBM SPSS Modeler Social Network Analysis . . . . .	1
A propos de l'analyse de réseaux sociaux . . . . .	2
Affichage des réseaux. . . . .	3
Description des réseaux . . . . .	3
noeuds IBM SPSS Modeler Social Network Analysis . . . . .	6
Onglets Noeud. . . . .	6
Prévisualisation des résultats . . . . .	7
Analyse des données. . . . .	7
Applications . . . . .	8
Structure des données . . . . .	8
Propriétés de génération de scripts . . . . .	9

## **2 Analyse de groupe 11**

Présentation de l'analyse de groupe. . . . .	11
Déterminer la similarité sociale . . . . .	11
Scission en groupes . . . . .	12
Description des groupes et des membres de groupe . . . . .	13
Noeud Analyse de groupe . . . . .	14
Spécification des données pour l'analyse de groupe . . . . .	14
Configuration des options de création pour l'analyse de groupe. . . . .	15
Affichage des statistiques d'analyse de groupe . . . . .	15
Résultats de l'analyse de groupe . . . . .	16

## **3 Analyse de diffusion 19**

Présentation de l'analyse de diffusion. . . . .	19
Exemple d'analyse de diffusion . . . . .	20
Noeud Analyse de diffusion . . . . .	20
Spécification de données pour l'analyse de diffusion. . . . .	21
Configuration des options de création pour l'analyse de diffusion . . . . .	22
Affichage des statistiques d'analyses de diffusion . . . . .	22
Résultats de l'analyse de diffusion . . . . .	23

***Annexe***

<b><i>A Avis</i></b>	<b><i>24</i></b>
<b><i>Glossaire</i></b>	<b><i>26</i></b>
<b><i>Bibliographie</i></b>	<b><i>27</i></b>
<b><i>Index</i></b>	<b><i>28</i></b>



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

## ***À propos de IBM SPSS Modeler Social Network Analysis***

De nombreuses approches de modélisation du comportement sont centrées sur l'individu. Elles utilisent de nombreuses données différentes sur les individus pour générer un modèle qui utilise les indicateurs clés du comportement pour le prévoir. Si un individu a des valeurs pour les indicateurs clés qui sont associées à l'occurrence du comportement, cet individu peut être la cible d'une attention particulière permettant d'éviter ce comportement.

Considérer les approches de modélisation de l'attrition dans lesquelles un client met fin à sa relation avec une entreprise. Le coût de rétention des clients est beaucoup plus bas que le coût associé à leur remplacement. Par conséquent, la capacité à identifier les clients susceptibles d'attrition est essentielle. Un analyste utilise souvent un certain nombre d'indicateurs clés de performances pour décrire les clients, notamment des informations démographiques et des schémas d'appel récents pour chaque client individuel. Les modèles prédictifs basés sur ces champs utilisent les modifications des schémas d'appel des clients qui correspondent aux schémas d'appel des clients qui ont arrêté leur relation avec l'entreprise afin d'identifier les personnes ayant un risque d'attrition plus élevé. Les clients identifiés comme susceptibles d'attrition reçoivent un service client ou des options de service supplémentaires afin de les retenir.

Ces méthodes ne tiennent pas compte des informations sociales qui peuvent avoir une influence significative sur le comportement d'un client. Les informations sur une entreprise et sur ce que font les autres personnes imprègnent les relations et ont une influence sur les individus. Par conséquent, les relations avec d'autres personnes permettent à ces personnes d'avoir une influence sur les décisions et les actes d'un individu. Les analyses qui tiennent uniquement compte des mesures individuelles passent à côté de facteurs importants qui permettent d'effectuer des prévisions.

IBM® SPSS® Modeler Social Network Analysis traite ce problème en ajoutant des informations relationnelles dans des champs supplémentaires qui peuvent être intégrés aux modèles. Ces indicateurs clés de performance dérivés mesurent les caractéristiques sociales des individus. La combinaison de ces propriétés sociales et des mesures individuelles permet d'obtenir un panorama plus précis des individus et peut, par conséquent, améliorer la précision prédictive de vos modèles.

IBM SPSS Modeler Social Network Analysis est composé de deux composants principaux :

- les noeuds IBM SPSS Modeler Social Network Analysis ajoutés à l'environnement IBM® SPSS® Modeler qui permettent d'inclure les techniques d'analyses sociales dans les flux.
- IBM® SPSS® Modeler Server Social Network Analysis qui ajoute le traitement des spécifications de noeud à IBM® SPSS® Modeler Server. IBM SPSS Modeler Server Social Network Analysis traite efficacement d'énormes quantités de données réseau qui peuvent

contenir des millions d'individus et de relations en un nombre relativement petit de champs pour des analyses ultérieures.

Par exemple, IBM SPSS Modeler Social Network Analysis identifie les individus d'un réseau qui sont le plus influencés par l'attrition de personnes spécifiques. De plus, il est possible de découvrir des groupes d'individus dans un réseau qui ont un risque d'attrition accru. En intégrant à vos modèles les indicateurs clés de performance correspondant à ces effets, il est possible d'améliorer les performances générales de ces modèles.

*Remarque* : le traitement des noeuds IBM SPSS Modeler Social Network Analysis doit être effectué par IBM SPSS Modeler Server Social Network Analysis. Lors de l'exécution d'un flux contenant des noeuds IBM SPSS Modeler Social Network Analysis dans SPSS Modeler, vous devez vous connecter à une instance SPSS Modeler Server qui contient également IBM SPSS Modeler Server Social Network Analysis.

## ***A propos de l'analyse de réseaux sociaux***

Un réseau social est composé d'un ensemble d'individus et des relations qui les lient. L'analyse de réseaux sociaux examine ces relations pour décrire les individus et les groupes comme faisant partie d'une structure sociale. Les individus interagissent les uns avec les autres et ces schémas d'interactions offrent des informations sur les individus concernés. Les relations permettent aux informations de se propager dans un réseau ce qui permet à un individu d'en influencer un autre. L'importance des informations relationnelles différencie l'analyse des réseaux sociaux des autres approches. Au lieu de se focaliser sur chaque individu de manière séparée, l'unité d'étude est une dyade composée de deux individus et des relations qui les unissent.

Les relations à l'intérieur d'un réseau peuvent être classées comme dirigées ou non dirigées. Dans une **relation dirigée**, un individu est identifié comme l'initiateur, ou la source, de la relation et l'autre individu est identifié comme le récepteur, ou la destination. Par exemple, passer un appel téléphonique est une relation dirigée dans laquelle une personne en appelle une autre. En revanche, les rôles de la source et de la destination ne peuvent pas être définis dans des **relations non dirigées**. Dans ce cas, les deux parties participent à la relation de la même manière. Parler ensemble est un exemple d'une relation non dirigée.

Une autre propriété qui différencie les relations est de savoir si celles-ci sont dichotomiques ou de valeur. Les seules informations disponibles dans une **relation dichotomique** sont celles qui permettent de savoir si une relation existe ou non entre deux individus. Pour chaque dyade du réseau, la relation est soit présente, soit absente. De l'autre côté, une **relation de valeur** intègre une pondération indiquant la pondération de la relation. Les pondérations permettent de comparer les relations.

Le tableau « [Types de relation](#) » montre un exemple de classification croisée des relations par direction et par échelle. Dans les relations dirigées, *Joe* est la source de la relation et *Marie* est la destination. Dans les relations non dirigées, il n'existe pas d'indication quant à l'initiateur de la relation. Les relations de valeur utilisent la durée de la conversation comme pondération de relation alors que les relations dichotomiques ont simplement eu lieu ou non.

Table 1-1  
Types de relation

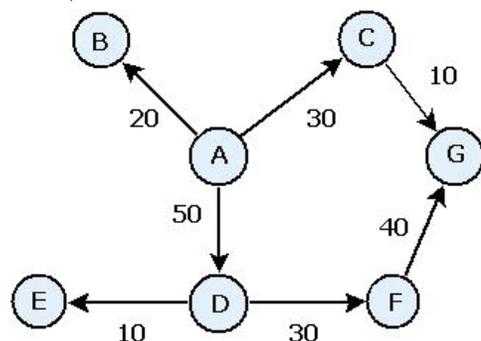
Direction	Echelle	Exemple
Non dirigée	Dichotomique	Joe et Marie parlent ensemble
Non dirigée	De valeur	Joe et Marie ont parlé ensemble pendant 20 minutes
Dirigée	Dichotomique	Joe a appelé Marie
Dirigée	De valeur	Joe a appelé Marie pendant 20 minutes

Pour plus d'informations sur le champ d'analyse de réseaux sociaux, consultez un des livres sur ce sujet (Wasserman et Faust, 1994).

## Affichage des réseaux

Un réseau social est généralement représenté par un **sociogramme**(Moreno, 1934). Dans ce genre de représentation visuelle, les individus correspondent à des points, ou à des noeuds, dans un espace. Les lignes, ou arêtes, reliant les points représentent les relations entre les individus. Si les relations sont dirigées, les arêtes contiennent une flèche qui indique la direction. Si les relations ont des pondérations, les étiquettes des arêtes indiquent les valeurs. Le graphique suivant affiche un réseau de sept individus.

Figure 1-1  
Exemple de réseau social



Imaginons que le réseau représente les appels téléphoniques effectués par des individus avec les pondérations relationnelles indiquant la durée des appels. Dans ce cas, la personne A a appelé trois personnes et a passé la majeure partie de son temps à parler à la personne D.

Ce réseau est bien plus petit que ceux rencontrés en réalité. Cependant, les concepts présentés dans des sociogrammes simples peuvent être appliqués à des réseaux de toute taille et de différente complexité.

## Description des réseaux

Les informations sur les réseaux, les groupes et les individus doivent être extraites sous la forme de caractéristiques descriptives qui permettent des comparaisons croisées et leur intégration dans des modèles prédictifs. Les réseaux doivent être décomposés en un ensemble fini d'indicateurs clés de performance pouvant être analysés. Par exemple, imaginons que vous souhaitiez comparer

des réseaux ou des groupes de noeuds à l'intérieur d'un réseau. Vous pourriez également vouloir comparer des individus dans le réseau ou identifier les individus les plus importants.

Les deux mesures les plus fréquemment utilisées pour décrire les réseaux sociaux sont la **densité** et le **degré**. Ces deux statistiques reflètent la connectivité mais la première cible le réseau en son entier ou des sous-groupes du réseau alors que la seconde définit les individus du réseau.

### **Densité du réseau**

Pour tout ensemble de noeuds d'un réseau, il existe un nombre fini de relations possibles. Chaque noeud peut être la source ou la cible d'une relation avec tous les autres noeuds. Imaginons un réseau composé de trois noeuds A, B et C. Le tableau suivant répertorie toutes les relations dirigées possibles entre les noeuds.

**Table 1-2**  
*Relations dirigées possibles pour trois noeuds*

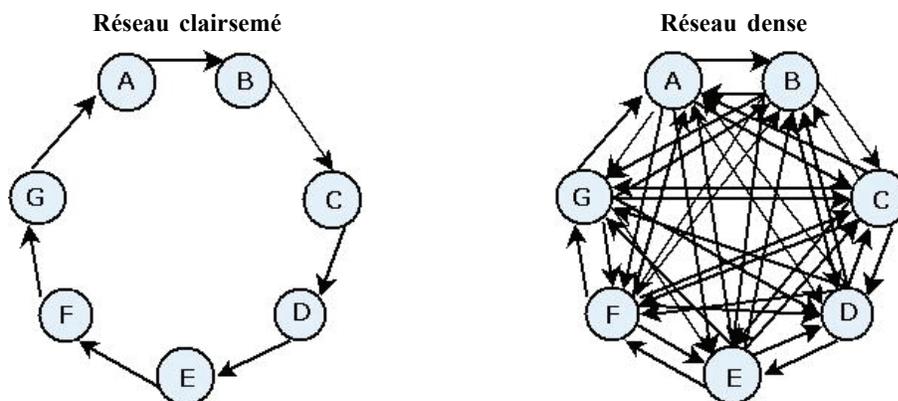
Source	Cible
A	B
A	C
B	A
B	C
C	A
C	B

Chaque noeud est la source d'une relation avec les deux autres noeuds. Cependant, dans la pratique, toutes les relations possibles peuvent ne pas être présentes. Il est possible que certains noeuds n'aient pas de relations directes avec d'autres noeuds. De plus, certaines relations dirigées peuvent ne pas être réciproques.

La statistique de **densité** représente la proportion de relations possibles dans le réseau qui sont réellement présentes. La valeur s'étend de 0 à 1, avec la limite inférieure correspondant aux réseaux sans relations et la limite supérieure représentant les réseaux avec toutes les relations possibles. Plus la valeur est proche de 1, plus le réseau est dense et plus la cohésion des noeuds est importante dans le réseau.

Les informations des réseaux denses sont transmises plus facilement que les informations des réseaux clairsemés. Le tableau « Réseaux clairsemés et denses » affiche deux réseaux constitués de sept noeuds. Le réseau clairsemé contient uniquement sept des 42 relations possibles entre les noeuds et génère une densité de 0,17. Le réseau dense, en revanche, contient toutes les relations possibles et affiche une densité de 1,0.

Table 1-3  
Réseaux clairsemés et denses



Dans le réseau clairsemé, pour que les informations soient transmises du noeud A au noeud G, elles doivent traverser cinq autres noeuds. Au contraire, dans le réseau dense, les informations peuvent passer directement du noeud A au noeud G.

### Degré des noeuds

Les individus importants dans le réseau sont souvent ceux qui sont impliqués dans le plus grand nombre de relations. Ces individus reçoivent des informations d'un grand nombre de sources et diffusent ces informations vers un grand nombre d'individus. Au contraire, les individus qui participent à un petit nombre de relations ne peuvent pas influencer directement un grand nombre d'individus du réseau.

Le **degré** d'un noeud, défini comme le nombre total de relations impliquant ce noeud, permet de comparer les participants du réseau. Les individus avec des valeurs de degré élevées sont plus actifs que ceux avec des valeurs moins élevées. Le degré ignore la direction des relations et propose une mesure générale de l'activité du noeud.

Pour les relations dirigées, vous pouvez vous concentrer sur le fait qu'un noeud est une source ou une cible pour compter le nombre de relations. Le **degré intérieur** d'un noeud est le nombre de relations dans lesquelles un noeud spécifique est la cible. Le **degré extérieur** est le nombre de relations dans lesquelles un noeud est la source. Le tableau suivant répertorie les valeurs de degré, degré intérieur et degré extérieur de chaque noeud dans le schéma « Exemple de réseau social ».

Table 1-4  
Exemples de valeurs de degré, degré intérieur et degré extérieur

Noeud	Degré	Degré intérieur	Degré extérieur
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

Le degré intérieur est souvent traité comme une mesure de prestige. Des valeurs de degré intérieur élevées correspondent à un grand nombre de relations se terminant au niveau de ce noeud. Pour le dire autrement, ces individus sont contactés par un grand nombre d'autres individus. De nombreux autres noeuds sont à l'origine des relations avec ce noeud. A l'inverse, le degré extérieur est traité comme une mesure de centralité. Des valeurs élevées correspondent à un grand nombre de relations provenant de ce noeud. Ces individus contactent un grand nombre d'autres individus.

Pour les noeuds de l'exemple de réseau, les valeurs de degré indiquent que les noeuds A et D sont les plus actifs alors que les noeuds B et E sont les moins actifs. Les valeurs de degré intérieur révèlent que le noeud G a le prestige le plus important. En fonction des valeurs de degré extérieur, le noeud A est le plus central.

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

En plus des nombreux noeuds standard fournis avec IBM® SPSS® Modeler, vous pouvez également utiliser les noeuds IBM® SPSS® Modeler Social Network Analysis pour qu'ils intègrent les résultats de l'analyse des réseaux sociaux dans vos flux. Le tableau « noeuds IBM SPSS Modeler Social Network Analysis » décrit ces noeuds qui sont stockés dans la palette Sources.

Table 1-5  
noeuds IBM SPSS Modeler Social Network Analysis

<b>Noeud</b>	<b>Icône</b>	<b>Description</b>
Analyse de groupe		Le noeud Analyse de groupe importe les données d'enregistrement des informations sur les appels depuis un fichier texte à champ fixe, identifie les groupes de noeuds dans le réseau défini par ces enregistrements et génère des indicateurs clés de performance pour les groupes et les individus du réseau. Pour plus d'informations, reportez-vous à la section <a href="#">Présentation de l'analyse de groupe</a> dans le chapitre 2 sur p. 11.
Analyse de diffusion		Le noeud Analyse de diffusion importe les données d'enregistrement des informations sur les appels depuis un fichier texte à champ fixe, propage un effet dans le réseau défini par ces enregistrements et génère des indicateurs clés de performance qui résument les résultats de cet effet sur les noeuds individuels. Pour plus d'informations, reportez-vous à la section <a href="#">Présentation de l'analyse de diffusion</a> dans le chapitre 3 sur p. 19.

Le traitement des noeuds IBM SPSS Modeler Social Network Analysis doit être effectué par IBM® SPSS® Modeler Server Social Network Analysis. Lors de l'exécution d'un flux contenant ces noeuds dans SPSS Modeler, vous devez vous connecter à une instance IBM® SPSS® Modeler Server qui contient également IBM SPSS Modeler Server Social Network Analysis.

### **Onglets Noeud**

Les noeuds IBM® SPSS® Modeler Social Network Analysis proposent les onglets suivants permettant de définir et d'obtenir un aperçu de l'analyse :

- **Onglet Données.** Permet d'identifier le fichier contenant les informations sur le réseau social.

- **Onglet Options de création.** Utilisé pour définir les paramètres des analyses.
- **Onglet Analyse.** Permet de consulter le récapitulatif des résultats préliminaires qui fournit une aide à la modification des paramètres d'entrée sur l'onglet Données afin de générer les meilleurs résultats possibles.

De plus, les noeuds proposent les onglets suivants dans tous les noeuds sources de IBM® SPSS® Modeler :

- **Onglet Filtrer.** Permet d'éliminer ou de renommer les champs de résultats produits par les noeuds. Cet onglet offre les mêmes fonctions que le noeud Filtrer.
- **Onglet Types.** Permet de définir les niveaux de mesure des champs de résultats produits par les noeuds. Cet onglet offre les mêmes fonctions que le noeud Typier.
- **Onglet Annotations.** Permet de renommer les noeuds, fournit des info-bulles personnalisées et stocke une annotation plus importante.

Pour des informations supplémentaires sur les onglets communs, consultez la documentation de SPSS Modeler.

## ***Prévisualisation des résultats***

Généralement, les résultats des noeuds sont utilisés comme entrée d'un modèle prédictif. Si vous souhaitez consulter les résultats, vous pouvez ajouter un noeud Tableau au flux. Cependant, la quantité de données est souvent extrêmement élevée et la génération de tableaux demande alors beaucoup de temps. De plus, le nombre de lignes du tableau limite son utilité.

Il est possible de générer un tableau de prévisualisation qui affiche un échantillon des résultats qui seront créés. La prévisualisation affiche les champs générés pour un nombre limité de lignes. Le nombre de lignes est défini dans les propriétés du flux. Pour prévisualiser les résultats d'un noeud, effectuez les opérations suivantes :

1. Ouvrez le noeud.
2. Spécifiez les paramètres de données dans l'onglet Données.
3. Définissez les paramètres analytiques dans l'onglet Options de création.
4. Cliquez sur Prévisualiser

Une fenêtre de prévisualisation s'ouvre et affiche les résultats. De plus, prévisualiser les résultats remplit l'onglet Analyse du noeud avec un récapitulatif des résultats.

## ***Analyse des données***

Déterminer les paramètres analytiques qui génèrent les meilleurs résultats est généralement un processus répétitif. Vous définissez les paramètres, effectuez l'analyse et observez les résultats. Si les résultats ne sont pas aussi utiles qu'ils pourraient l'être, vous modifiez les paramètres et réexécutez l'analyse.

Pour analyser les données d'entrée du noeud, effectuez les opérations suivantes :

1. Ouvrez le noeud.
2. Spécifiez les paramètres de données dans l'onglet Données.
3. Définissez les paramètres analytiques dans l'onglet Options de création.
4. Cliquez sur Analyser les données.

Si les options de création indiquent que les statistiques récapitulatives doivent être affichées, l'onglet Analyse affiche les résultats.

Si vous avez besoin d'effectuer à nouveau une analyse, cliquez sur Supprimer l'analyse pour nettoyer les résultats actuels avant de cliquer sur Analyser les données.

## ***Applications***

Les applications spécifiques dans lesquelles IBM® SPSS® Modeler Social Network Analysis peut être particulièrement utile sont notamment :

- **la prévision de l'attrition.** Les caractéristiques des groupes peuvent avoir une influence sur les taux d'attrition. En s'intéressant aux individus des groupes qui ont un risque d'attrition plus élevé, il peut être possible de l'éviter. De plus, vous pouvez identifier les individus susceptibles d'attrition en raison des informations qu'ils reçoivent de ceux qui ont déjà quitté l'entreprise.
- **Utilisation optimale des chefs de groupe.** Les chefs de groupe ont une forte influence sur les autres membres du groupe. S'il est possible d'éviter l'attrition d'un chef de groupe, le taux d'attrition des membres du groupe peut être réduit. Essayer de provoquer l'attrition d'un chef de groupe chez un concurrent peut également augmenter le taux d'attrition des membres du groupe associés à ce concurrent tout en réduisant le taux d'attrition des membres du groupe associés à votre entreprise.
- **Marketing.** Les chefs de groupe peuvent permettre de lancer de nouveaux produits ou offres de service. L'influence du chef peut pousser les autres membres du groupes à acheter cette offre. Vous pouvez utiliser l'analyse de diffusion pour identifier les individus les plus influencés par les chefs de groupe et cibler votre campagne marketing.

## ***Structure des données***

Les informations intéressantes sur les individus du réseau peuvent être propagées au sein d'un grand nombre de fichiers, bases de données et systèmes de votre entreprise. Pour analyser le réseau utilisant IBM® SPSS® Modeler Social Network Analysis, vous devez extraire les enregistrements et les champs appropriés de vos sources de données et les formater comme entrées des noeuds.

Les noeuds analytiques nécessitent que les enregistrements des informations sur les appels soient stockés dans un fichier texte à largeur fixe. Chaque ligne du fichier correspond à une relation, avec les données organisées dans les colonnes suivantes :

- l'identifiant de l'individu qui a initié la relation.

- l'identifiant de l'individu qui est la cible de la relation.
- une pondération facultative pour la relation.

Toutes les données doivent être numériques et les identifiants des individus doivent être des entiers. L'ajout des noms des champs comme première ligne des données du fichier est facultatif. Le tableau « [Exemple d'enregistrements des informations sur les appels](#) » illustre cette structure de données.

Table 1-6  
Exemple d'enregistrements des informations sur les appels

Source	Destination	Pondération
1000	5642	243
2190	8444	831
0299	9419	559

Les valeurs de pondération peuvent correspondre à toutes les mesures utilisées pour représenter l'importance de la relation par rapport aux autres relations du réseau. Pour les données d'appel, les pondérations habituelles incluent la durée ou la fréquence des appels.

Si vous souhaitez que l'analyse se concentre sur un sous-ensemble de l'historique des appels, vous devez utiliser ce sous-ensemble lors de la création du fichier d'entrée. Par exemple, vous pouvez limiter l'analyse aux mois passés ou aux appels plus récents d'un individu en incluant uniquement ces données dans le fichier texte d'entrée.

## Propriétés de génération de scripts

La génération de scripts IBM® SPSS® Modeler Social Network Analysis est un outil performant pour automatiser les processus dans l'interface utilisateur. Les scripts permettent d'effectuer les mêmes opérations qu'avec la souris ou le clavier. Vous pouvez les utiliser pour automatiser les tâches dont l'exécution manuelle s'avère très répétitive et très longue. Pour une explication sur l'utilisation des scripts, consultez le guide *ScriptingAutomation.pdf* disponible avec IBM® SPSS® Modeler.

### Propriétés du noeud Diffusion

Le tableau suivant répertorie les propriétés des scripts pour le noeud Diffusion.

Nom de la propriété	Le type de données	Description de la propriété
input_data_file_name	chaîne	
calling_field	champ	
called_field	champ	
frequency_weight_field	champ	
read_field_names	booléen	
diffusion_list_file_name	chaîne	
spreading_factor	double	Valeur par défaut = 50. Min = 1 Max = 99.

Nom de la propriété	Le type de données	Description de la propriété
max_number_iterations	<i>entier</i>	Valeur par défaut = 100. Min = 1.
accuracy_threshold	<i>double</i>	Valeur par défaut = 0.01. Min = 0.001.
calculate_statistics	<i>booléen</i>	

### **Propriétés du noeud Analyse de groupe**

Le tableau suivant répertorie les propriétés de script du noeud Analyse de groupe.

Nom de la propriété	Le type de données	Description de la propriété
input_data_file_name	<i>chaîne</i>	
calling_field	<i>champ</i>	
called_field	<i>champ</i>	
frequency_weight_field	<i>champ</i>	
read_field_names	<i>booléen</i>	
coverage_threshold	<i>double</i>	Valeur par défaut = 10. Min = 1 Max = 99.
min_group_size	<i>entier</i>	Valeur par défaut = 2. Min = 2.
max_group_size	<i>entier</i>	Valeur par défaut = 100. Min = 2.
calculate_statistics	<i>booléen</i>	

# ***Analyse de groupe***

## ***Présentation de l'analyse de groupe***

L'analyse de groupe utilise les modèles d'interaction des individus dans un réseau pour identifier les groupes d'individus communs. Les caractéristiques de ces groupes influencent le comportement des membres de groupe individuels. Par exemple, de petits groupes avec de nombreuses relations entre les membres et des chefs puissants, ont des risques accrus d'attrition même si aucun membre du groupe n'est encore parti. (Richter, Yom-Tov, et Slonim, 2010) Les modèles prédictifs qui intègrent à la fois les mesures sur les groupes et sur les individus seront plus efficaces que les modèles qui contiennent uniquement des mesures sur les individus.

L'analyse de groupe est composée des étapes de base suivantes :

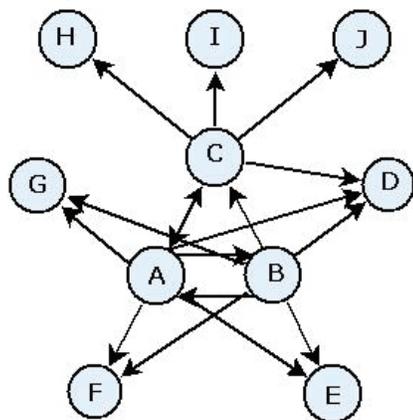
1. déterminer les forces relationnelles qui reflètent la proximité sociale. Pour plus d'informations, reportez-vous à la section [Déterminer la similarité sociale](#) sur p. 11.
2. Scinder le réseau en groupes en fonction de la pondération relationnelle tout en obéissant aux restrictions sur la taille. Pour plus d'informations, reportez-vous à la section [Scission en groupes](#) sur p. 12.
3. Effectuer un profil des groupes et des individus, en identifiant notamment les chefs des groupes. Pour plus d'informations, reportez-vous à la section [Description des groupes et des membres de groupe](#) sur p. 13.

## ***Déterminer la similarité sociale***

Les membres d'un groupe devraient plus ressembler aux autres membres du groupe qu'à des individus qui n'appartiennent pas à ce groupe. Dans l'analyse de réseau, la similarité de deux noeuds dépend de leur relation. Pour tout noeud d'un réseau, il existe un ensemble de noeuds qui sont les cibles des relations dirigées avec ce noeud. Pour les données de télécommunications, cet ensemble correspond à toutes les personnes contactées par un individu spécifique. Si deux individus contactent le même ensemble de personnes, ces individus sont considérés comme semblables. Plus les ensembles de cibles relationnelles de deux individus se recoupent, plus ces individus sont semblables.

Examinez le réseau présenté dans le schéma « [Exemple de réseau à dix noeuds](#) ».

Figure 2-1  
Exemple de réseau à dix noeuds



Le tableau « **Noeuds cibles** » montre les noeuds cibles des relations commençant aux noeuds A, B et C.

Table 2-1  
Noeuds cibles

Noeud source	Noeuds cibles
A	B, C, D, E, F, G
B	A, C, D, E, F, G
C	A, D, H, I, J

Les noeuds A et B ont cinq noeuds cibles en commun. Les noeuds A et C, en revanche, n'en ont qu'un. Par conséquent, les noeuds A et B sont socialement plus semblables que les noeuds A et C.

Les pondérations relationnelles, telles que la durée ou la fréquence des appels, ne capturent pas la similarité des noeuds dans un réseau. Pour inclure la notion de similarité, l'analyse de groupe utilise des informations mutuelles (Couverture et Thomas, 2006) telles que la pondération relationnelle. Cette statistique reflète la probabilité que deux noeuds soient connectés au même noeud. Si la relation dans le réseau contient des pondérations prédéfinies, telles que la fréquence des appels pour les données de télécommunications, les informations mutuelles intègrent ces pondérations.

### **Scission en groupes**

Les groupes doivent être composés d'individus qui ont des pondérations relationnelles élevées avec d'autres membres du groupe, où les pondérations mesurent la similarité des noeuds dans la relation. Ainsi, l'identification de groupe commence par ignorer les relations les plus faibles dans un réseau. Le **seuil de couverture** contrôle ce processus en définissant la fraction de la relation la plus forte à conserver. Par exemple, une valeur de seuil de couverture de 0,4 utilise les 40 % de relations les plus fortes pour l'identification du groupe alors que les 60% restants des relations sont ignorés.

Les relations restantes peuvent produire de très petits groupes ou des groupes très importants qui ont une utilité de prédiction limitée. Pour éviter que ce genre de groupes soient inclus dans l'analyse, vous pouvez spécifier des tailles de groupe minimales et maximales. Les groupes avec des tailles inférieures au minimum seront entièrement ignorés. Les groupes avec des tailles supérieures au maximum, en revanche, seront divisés en groupes plus petits dans la plage de tailles autorisées. Les groupes restants, après avoir ignoré les relations faibles et avoir mis en place des limites de taille, sont appelés des **groupes principaux**.

La suppression des relations du réseau d'origine peut avoir pour conséquence que certains individus n'appartiennent à aucun groupe principal. Cependant, ces individus peuvent avoir des connexions avec des membres d'un groupe ce qui leur garantit d'être intégrés à ce groupe. Un individu est ajouté à un groupe s'il existe de nombreuses relations relativement fortes avec les membres principaux de ce groupe, tant que les limites de taille des groupes sont respectées. Par conséquent, les groupes finaux sont composés d'un ensemble de membres principaux en plus des membres ajoutés en raison de leur connexion aux membres principaux.

### ***Description des groupes et des membres de groupe***

En plus de la densité, le degré intérieur et extérieur et d'autres statistiques décrivent les dynamiques de groupe. En particulier, les scores d'autorité et de diffusion proposent des mesures du statut social des individus au sein des groupes. Le rôle de chaque individu dans un groupe est extrêmement important pour essayer de prévoir le comportement du groupe et de ses membres.

Le **score d'autorité** d'un noeud mesure la tendance des autres noeuds du groupe à s'y connecter. Si de nombreux individus contactent un individu spécifique, pour lui demander des informations ou son opinion par exemple, cet individu a le rôle d'une autorité. Les scores d'autorité des noeuds dans le groupe, qui correspondent aux probabilités stationnaires d'un trajet aléatoire avec redémarrages dans le réseau du groupe, vont de 0 à 1. Plus le score d'autorité est proche de 1, plus ce noeud a d'autorité au sein du groupe. Le noeud dans le groupe ayant le score d'autorité le plus élevé s'appelle le **chef d'autorité** du groupe. Diviser le score le plus élevé d'un groupe par le plus petit génère une mesure de la puissance générale du chef d'autorité.

Le **score de diffusion** d'un noeud mesure la tendance du noeud à se connecter aux autres noeuds du groupe. Si un individu spécifique contacte de nombreuses personnes du groupe, cet individu peut avoir un impact important sur les opinions du groupe entier. Les scores de diffusion des noeuds du groupe, qui correspondent aux probabilités stationnaires d'un trajet aléatoire avec redémarrages dans le réseau du groupe dans l'ordre inverse, vont de 0 à 1. Plus le score de diffusion est proche de 1, plus ce noeud se connecte aux autres membres du groupe. Le noeud dans le groupe ayant le score de diffusion le plus élevé s'appelle le **chef de diffusion** du groupe. Diviser le score le plus élevé d'un groupe par le plus petit génère une mesure de la pondération globale du chef de diffusion.

## **Noeud Analyse de groupe**

Le noeud Analyse de groupe, qui est disponible dans la palette Sources, identifie les ensembles d'individus dans un réseau qui sont socialement semblables et détermine le statut social relatif des individus dans les groupes. Les groupes relativement petits qui ont de nombreuses connexions entre leurs membres et un chef important, ont un risque d'attrition plus élevé en raison de leur dynamique de groupe. Vous pouvez utiliser les résultats pour déterminer les scores d'attrition du groupe qui, lorsqu'ils sont combinés aux scores d'attrition individuels, améliorent la capacité de prévision de l'attrition en fonction de modèles basés sur les scores individuels seuls.

En raison de la quantité de données qui comprend généralement le réseau et la nature des interactions sociales qui sous-tendent les définitions des groupes, l'analyse de groupe est un processus très long. Généralement, vous enregistrez les résultats de l'analyse de groupe sur une base de données ou un fichier afin de les utiliser comme entrées d'un modèle prédictif. Les résultats du groupe sont actualisés peu fréquemment, une fois par mois par exemple, alors que le modèle peut être actualisé plus souvent.

**Conditions requises.** Le noeud nécessite un fichier texte à largeur fixe qui définit le réseau social à l'aide de trois champs. Un champ identifie la source de chaque relation dirigée, un champ définit la destination de chaque relation dirigée et le troisième champ spécifie une pondération facultative pour chaque relation. Toutes les relations du réseau doivent être dirigées.

### **Spécification des données pour l'analyse de groupe**

L'onglet Données de la fenêtre du noeud source Analyse de groupe vous permet de spécifier le fichier d'entrée contenant les relations des noeuds du réseau.

**Fichier.** Spécifiez le nom d'un fichier ou d'un dossier contenant les enregistrements des informations sur les appels. Vous pouvez saisir un nom ou cliquer sur le bouton des points de suspension (...) pour sélectionner un nom dans le système de fichiers. Le chemin d'accès apparaît lorsque le nom est sélectionné et son contenu est affiché avec des séparateurs. Si vous spécifiez un dossier, les enregistrements des informations sur les appels contenus dans ce dossier sont enchaînés pour l'analyse ; tous les fichiers du dossier doivent avoir la même structure.

**Lire les noms des champs à partir du fichier.** Sélectionnée par défaut, cette option traite la première ligne du fichier de données en tant que noms pour les colonnes. Si votre première ligne n'est pas un en-tête, désélectionnez cette option pour fournir automatiquement un nom générique à chaque champ, tel que *Champ1* et *Champ2*

#### **Paramètres de définition de réseau**

Les paramètres de définition de réseau définissent les rôles des champs.

**Champs.** Utilisez les boutons fléchés pour attribuer manuellement les éléments de cette liste aux différents champs de rôle. Les icônes indiquent les niveaux de mesure valides pour chaque champ de rôle. Cliquez sur le bouton Tous pour sélectionner tous les champs de la liste ou cliquez sur

un bouton de niveau de mesure individuelle pour sélectionner tous les champs avec un niveau de mesure.

**Source.** Sélectionnez un champ comme origine de la relation dirigée.

**Destination.** Sélectionnez un champ comme cible de la relation dirigée.

**Pondération.** En option, sélectionnez un champ correspondant au nombre de fois où la relation se produit dans le réseau. L'analyse pondère les enregistrements en fonction, au moment où les pondérations des relations sont déterminées.

## ***Configuration des options de création pour l'analyse de groupe***

L'onglet Options de création de la fenêtre du noeud source Analyse de groupe vous permet de définir les options d'identification des groupes dans le réseau.

### ***Paramètres du noeud Analyse de groupe***

Les paramètres des analyses de groupe influencent la taille et la pondération relative des groupes formés.

**Seuil de couverture.** Définit la fraction des relations réseau les plus fortes à utiliser dans l'analyse. Par exemple, un seuil de couverture de 0,2 indique que seules les relations dont les pondérations se trouvent parmi les 20 % des pondérations les plus élevées seront utilisées. Ce paramètre, qui va de 0 à 1, permet à l'analyse de cibler les relations les plus fortes qui ont lieu au sein des groupes. Des valeurs de seuil plus élevées produisent des groupes avec des relations moins fortes entre leurs membres. Pour plus d'informations, reportez-vous à la section [Scission en groupes](#) sur p. 12.

**Taille de groupe minimale.** Spécifie une limite inférieure pour la taille du groupe. Les groupes inférieurs à cette valeur ne seront pas renvoyés.

**Taille de groupe maximale.** Spécifie une limite supérieure pour la taille du groupe. Les groupes supérieurs à cette valeur seront divisés en groupes plus petits.

**Calculer et afficher les statistiques récapitulatives.** Si cette option est sélectionnée, le noeud calcule et affiche les statistiques récapitulatives de l'analyse en plus de calculer les résultats des indicateurs de performances clés. Le calcul de ces statistiques peut avoir un impact négatif sur les performances du noeud pour les réseaux de très grande taille.

## ***Affichage des statistiques d'analyse de groupe***

L'onglet Analyse de la fenêtre du noeud source Analyse de groupe offre un récapitulatif des groupes identifiés dans le réseau. Le tableau « Statistiques récapitulatives des analyses de groupe » répertorie les statistiques récapitulatives disponibles pour les groupes.

Table 2-2  
Statistiques récapitulatives des analyses de groupe

Statistiques	Description
Total des noeuds dans les groupes	Nombre de noeuds inclus dans les groupes identifiés
Total des liens dans les groupes	Nombre de liens inclus dans les groupes identifiés
Total des groupes	Nombre de groupes identifiés dans le réseau
Taille moyenne des groupes	Nombre de noeuds moyen dans un groupe
Densité moyenne des groupes	Fraction de connexions directes moyenne entre les noeuds d'un groupe. Pour plus d'informations, reportez-vous à la section <a href="#">Densité du réseau</a> dans le chapitre 1 sur p. 4.
Fraction moyenne des membres principaux	Fraction de noeuds moyenne dans un groupe qui sont des noeuds principaux du groupe. Pour plus d'informations, reportez-vous à la section <a href="#">Scission en groupes</a> sur p. 12.
Densité moyenne des groupes principaux	Fraction de connexions directes moyenne entre les noeuds principaux d'un groupe.
Degré intérieur moyen	Nombre de liens entrants moyen. Pour plus d'informations, reportez-vous à la section <a href="#">Degré des noeuds</a> dans le chapitre 1 sur p. 5.
Degré extérieur moyen	Nombre de liens sortants moyen. Pour plus d'informations, reportez-vous à la section <a href="#">Degré des noeuds</a> dans le chapitre 1 sur p. 5.

Sélectionnez une statistique spécifique dans le tableau Statistiques récapitulatives pour visualiser la distribution des valeurs, l'écart-type et l'asymétrie de la statistique.

La consultation de ces statistiques permet de mieux définir les paramètres d'analyse de groupe. Idéalement, les résultats doivent présenter des groupes relativement petits avec des valeurs de densité élevées. Par exemple, s'il existe des groupes de très grande taille, pensez à réduire la taille de groupe maximale pour les diviser en groupes plus petits. Autrement, si les valeurs de densité des groupes sont plutôt basses, pensez à réduire le seuil de couverture pour cibler les relations les plus fortes.

## Résultats de l'analyse de groupe

Le noeud Analyse de groupe génère de nombreux champs qui décrivent les groupes et les individus dans les groupes. Vous pouvez augmenter les modèles et données existants à l'aide de ces indicateurs de performances clés pour améliorer les prévisions générées par vos modèles. Par exemple, vous pouvez mettre à jour les valeurs de probabilité d'attrition individuelles pour qu'elles intègrent les influences des groupes.

Le tableau « Indicateurs de performances clés pour les groupes » répertorie les champs des indicateurs de performances clés des groupes indentifiés dans l'analyse.

Table 2-3  
Indicateurs de performances clés pour les groupes

Champ	Description
GAG_GroupNumber	Identifiant unique d'un groupe
GAG_Size	Nombre d'individus dans un groupe

Champ	Description
GAG_Density	Fraction de connexions directes entre les individus d'un groupe. Pour plus d'informations, reportez-vous à la section <a href="#">Densité du réseau</a> dans le chapitre 1 sur p. 4.
GAG_KernelDensity	Fraction de connexions directes entre les individus principaux d'un groupe.
GAG_CoreNodesFraction	Fraction des individus d'un groupe qui sont des individus principaux du groupe. Pour plus d'informations, reportez-vous à la section <a href="#">Scission en groupes</a> sur p. 12.
GAG_MaxRankType1	Score d'autorité maximum d'un des membres du groupe. Pour plus d'informations, reportez-vous à la section <a href="#">Description des groupes et des membres de groupe</a> sur p. 13.
GAG_MinRankType1	Score d'autorité minimum d'un des membres du groupe.
GAG_MaxMinRankRatioType1	Rapport entre le score d'autorité le plus élevé et le plus bas. Cette valeur reflète la pondération d'autorité du chef du groupe.
GAG_MaxRankType2	Score de diffusion maximum d'un des membres du groupe. Pour plus d'informations, reportez-vous à la section <a href="#">Description des groupes et des membres de groupe</a> sur p. 13.
GAG_MinRankType2	Score de diffusion minimum d'un des membres du groupe.
GAG_MaxMinRankRatioType2	Rapport entre le score de diffusion le plus élevé et le plus bas. Cette valeur reflète la pondération de diffusion du chef du groupe.

Le tableau « Indicateurs de performances clés pour les individus » répertorie les champs des indicateurs de performances clés des individus du réseau.

Table 2-4

*Indicateurs de performances clés pour les individus*

Champ	Description
GAI_NodeNumber	Identifiant unique d'un individu
GAI_CoreNode	Indicateur signalant si l'individu est un individu principal du groupe ou non. Pour plus d'informations, reportez-vous à la section <a href="#">Scission en groupes</a> sur p. 12.
GAI_RankType1	Score d'autorité de l'individu. Pour plus d'informations, reportez-vous à la section <a href="#">Description des groupes et des membres de groupe</a> sur p. 13.
GAI_RankOrderType1	Ordre des rangs dans le groupe en fonction des scores d'autorité
GAI_RankType2	Score de diffusion de l'individu. Pour plus d'informations, reportez-vous à la section <a href="#">Description des groupes et des membres de groupe</a> sur p. 13.
GAI_RankOrderType2	Ordre des rangs dans le groupe en fonction des scores de diffusion
GAI_InDegree	Nombre de relations dans lesquelles l'individu est la cible de la relation. Pour plus d'informations, reportez-vous à la section <a href="#">Degré des noeuds</a> dans le chapitre 1 sur p. 5.
GAI_OutDegree	Nombre de relations dans lesquelles l'individu est la source de la relation. Pour plus d'informations, reportez-vous à la section <a href="#">Degré des noeuds</a> dans le chapitre 1 sur p. 5.
GAI_GroupLeaderType1	Si le noeud est ou non un chef d'autorité, dont le score de chef est calculé à partir des liens entrants. Pour plus d'informations, reportez-vous à la section <a href="#">Description des groupes et des membres de groupe</a> sur p. 13.
GAI_GroupLeaderConfidenceType1	Le niveau de confiance au sujet du noeud comme chef d'autorité.

<b>Champ</b>	<b>Description</b>
GAI_GroupLeaderType2	Si le noeud est ou non un chef de diffusion, dont le score de chef est calculé à partir des liens sortants. Pour plus d'informations, reportez-vous à la section <a href="#">Description des groupes et des membres de groupe</a> sur p. 13.
GAI_GroupLeaderConfidenceType2	Le niveau de confiance au sujet du noeud comme chef de diffusion.

# Analyse de diffusion

## Présentation de l'analyse de diffusion.

L'analyse de diffusion identifie les individus les plus affectés par d'autres individus dans un réseau social et quantifie l'effet sous la forme d'**énergie diffusée**. Ce processus utilise une approche d'activation propagée au cours de laquelle un effet se propage à plusieurs reprises depuis des noeuds du réseau à leurs voisins immédiats et dont la taille diminue d'un noeud à l'autre (Dasgupta, Singh, Viswanathan, Chakraborty, Mukherjea, Nanavati, et Joshi, 2008). Lorsqu'un noeud reçoit l'énergie, il est activé et transmet une partie de cette énergie à ses voisins qui sont les cibles des relations dirigées avec ce noeud.

Le **facteur de propagation** définit la partie de l'énergie transmise par un noeud activé et la quantité restante conservée par le noeud. Tous les noeuds recevant cette énergie transmettront eux-mêmes la même proportion à leurs voisins générant ainsi un processus décroissant de transmission d'énergie. Des facteurs de propagation élevés correspondent à un envoi plus important d'énergie ce qui permet à cette énergie d'atteindre des noeuds plus éloignés des noeuds activés d'origine avant que le processus ne décroisse totalement. Des facteurs de propagation peu élevés génèrent des processus de diffusion qui décroissent rapidement et une énergie transmise restant relativement proche des noeuds d'origine.

La quantité totale d'énergie diffusée par un noeud activé est distribuée à tous les noeuds qui sont des cibles des relations dirigées avec ce noeud. La quantité que chaque noeud reçoit dépend de la force de la relation avec le noeud activé. La fraction envoyée à un noeud spécifique équivaut à la pondération de la relation divisée par le total des pondérations de toutes les relations dans lesquelles le noeud activé est la source. Par conséquent, les voisins ayant des relations avec des pondérations relatives plus élevées reçoivent plus d'énergie que leurs voisins avec des pondérations relatives inférieures.

Le processus de diffusion s'arrête lorsqu'une des conditions suivantes se produit :

- les noeuds activés ne sont la source d'aucune relation dirigée
- la quantité d'énergie transmise est inférieure au **seuil de précision**, une limite de la quantité d'énergie transférée pour la poursuite du processus
- le nombre d'itérations atteint une limite spécifiée

Lorsque la diffusion est terminée, les noeuds ayant le plus d'énergie diffusée sont les plus sensibles à l'effet ayant initié le processus. Par exemple, si le processus commence avec des noeuds concernés par l'attrition, les noeuds avec l'énergie la plus élevée sont les plus susceptibles d'être eux-mêmes concernés par l'attrition. Vous pouvez porter une attention particulière à ces noeuds pour éviter l'attrition.

### Exemple d'analyse de diffusion

Examinez le réseau présenté dans le chapitre 1 sur p. 3. Le réseau est composé de sept noeuds ayant des relations dirigées avec des pondérations différentes.

Le tableau « Exemple de processus de diffusion » montre un processus de diffusion dans le réseau qui utilise un facteur de propagation de 0,80. A l'origine, le noeud A contient toute l'énergie, qui reçoit une valeur arbitraire de 1,00. Dans l'étape 1, ce noeud est activé et propage 80 % de son énergie aux trois voisins qui sont les cibles des relations tout en conservant 20 % de l'énergie pour lui-même. Les relations ont une pondération totale de 100. La relation avec le noeud D totalise la moitié de cette pondération totale et par conséquent, ce noeud reçoit la moitié de l'énergie diffusée, soit 0,40. La relation avec le noeud B représente 20 % de la pondération totale, ce qui équivaut à une valeur d'énergie diffusée de 0,16. Le noeud C reçoit l'énergie restante, 0,24, soit 30 % de la quantité provenant du noeud A.

Table 3-1  
Exemple de processus de diffusion

Etape	Noeud A	Noeud B	Noeud C	Noeud D	Noeud E	Noeud F	Noeud 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

Ayant reçu de l'énergie à l'étape 1, les noeuds B, C et D sont activés à l'étap 2 et diffusent de l'énergie à tous les voisins cibles. Le noeud B n'a pas de noeuds cibles et ne diffuse aucune énergie. Le noeud C, en revanche, a une cible et diffuse 80 % de son énergie à ce noeud G. Finalement, le noeud D propage 80 % de son énergie aux noeuds E et F avec F qui reçoit trois fois la quantité de E en raison de sa relation plus forte avec D.

Dans l'étape 3, les noeuds E, F et G sont activés. Les noeuds E et G n'ont pas de noeuds cibles donc aucun d'eux ne diffuse d'énergie. Cependant, le noeud F diffuse 80 % de son énergie vers son seul noeud cible, le noeud G. Ce noeud a désormais une valeur d'énergie totale de 0,38 qui est composée d'énergie diffusée depuis les noeuds C et F.

A ce moment, le noeud G est activé mais n'a pas de noeuds cibles pour la diffusion et le processus s'arrête. L'énergie provenant du noeud A s'est propagée à travers tout le réseau avec le noeud G qui reçoit la quantité la plus importante. Si l'énergie introduite dans le réseau représente l'attrition, le noeud G sera probablement le plus affecté par l'attrition du noeud A.

### Noeud Analyse de diffusion

Le noeud Analyse de diffusion, qui est disponible depuis la palette Sources, propage un effet à partir d'un ensemble d'individus spécifié au travers d'un réseau, et utilise les relations du réseau pour identifier les individus les plus affectés par cet effet. Si l'effet est l'attrition, par exemple, le noeud identifie les individus les plus susceptibles d'attrition consécutive à l'attrition d'autres individus dans le réseau. Vous pouvez augmenter les modèles et données existants à l'aide des résultats des noeuds pour améliorer les prévisions générées par ces modèles. Par exemple, vous

pouvez mettre à jour les valeurs de probabilité d'attrition individuelles pour qu'elles intègrent les influences de la diffusion.

**Conditions requises.** Le noeud nécessite deux fichiers texte à largeur fixe contenant les données à analyser. Le premier fichier définit le réseau social utilisant les trois champs. Un champ identifie la source de chaque relation dirigée, un champ définit la destination de chaque relation dirigée et le troisième champ spécifie une pondération facultative pour chaque relation. Toutes les relations du réseau doivent être dirigées. Le deuxième fichier contient une liste des identifiants à partir desquels l'effet doit commencer.

### ***Spécification de données pour l'analyse de diffusion***

L'onglet Données de la fenêtre du noeud source Analyse de diffusion vous permet de spécifier le fichier d'entrée contenant les relations des noeuds du réseau.

**Fichier.** Spécifiez le nom d'un fichier ou d'un dossier contenant les enregistrements des informations sur les appels. Vous pouvez saisir un nom ou cliquer sur le bouton des points de suspension (...) pour sélectionner un nom dans le système de fichiers. Le chemin d'accès apparaît lorsque le nom est sélectionné et son contenu est affiché avec des séparateurs. Si vous spécifiez un dossier, les enregistrements des informations sur les appels contenus dans ce dossier sont enchaînés pour l'analyse ; tous les fichiers du dossier doivent avoir la même structure.

**Lire les noms des champs à partir du fichier.** Sélectionnée par défaut, cette option traite la première ligne du fichier de données en tant que noms pour les colonnes. Si votre première ligne n'est pas un en-tête, désélectionnez cette option pour fournir automatiquement un nom générique à chaque champ, tel que *Champ1* et *Champ2*

#### ***Paramètres de définition de réseau***

Les paramètres de définition de réseau définissent les rôles des champs.

**Champs.** Utilisez les boutons fléchés pour attribuer manuellement les éléments de cette liste aux différents champs de rôle. Les icônes indiquent les niveaux de mesure valides pour chaque champ de rôle. Cliquez sur le bouton Tous pour sélectionner tous les champs de la liste ou cliquez sur un bouton de niveau de mesure individuelle pour sélectionner tous les champs avec un niveau de mesure.

**Source.** Sélectionnez un champ comme origine de la relation dirigée.

**Destination.** Sélectionnez un champ comme cible de la relation dirigée.

**Pondération.** Sélectionnez un champ représentant la pondération de la relation. Pour les enregistrements d'informations sur les appels, la pondération peut correspondre à n'importe quel nombre de caractéristiques d'appel reflétant la pondération de la relation, telles que la durée ou la fréquence de l'appel.

## **Configuration des options de création pour l'analyse de diffusion**

L'onglet Options de création de la fenêtre du noeud source Analyse de diffusion vous permet de définir les options du processus de diffusion.

### **Liste de graines de diffusion**

La liste de graines de diffusion identifie les noeuds de réseau à partir desquels l'énergie doit commencer à se propager. Par exemple, lors de l'analyse de l'attrition, cette liste identifie les noeuds de réseau déjà concernés par l'attrition. Cette liste peut également contenir des noeuds ayant le risque d'attrition le plus élevé afin d'explorer les effets sur le réseau en cas d'attrition avérée de ces noeuds.

**Fichier.** Spécifiez le nom du fichier texte contenant les identifiants de noeud avec chaque identifiant correspondant à une ligne dans le fichier. Vous pouvez entrer un nom de fichier ou cliquer sur le bouton représentant des points de suspension (...) pour sélectionner un fichier.

### **Paramètres d'analyse de diffusion**

Les paramètres d'analyse de diffusion déterminent l'agressivité de la diffusion et le moment de l'arrêt du processus.

**Facteur de propagation.** Définit la fraction d'énergie qu'un noeud transfère à chaque étape du processus de diffusion. Les valeurs les plus petites limitent l'énergie aux noeuds les plus proches du noeud qui diffuse l'énergie. Inversement, les valeurs les plus élevées permettent à l'énergie d'atteindre des noeuds plus éloignés.

**Nombre maximum d'itérations :** Spécifiez une limite supérieure pour le nombre d'itérations de la diffusion. Le processus de diffusion se termine automatiquement si cette limite est atteinte.

**Seuil de précision.** Spécifiez la plus petite modification de l'énergie du noeud qui permet de poursuivre le processus de diffusion. Si la modification de l'énergie dans chaque noeud est inférieure à cette valeur, le processus de diffusion est arrêté.

**Calculer et afficher les statistiques récapitulatives.** Si cette option est sélectionnée, le noeud calcule et affiche les statistiques récapitulatives de l'analyse en plus de calculer les résultats des indicateurs de performances clés. Le calcul de ces statistiques peut avoir un impact négatif sur les performances du noeud pour les réseaux de très grande taille.

## **Affichage des statistiques d'analyses de diffusion**

L'onglet Analyse de la fenêtre du noeud source Analyse de diffusion propose un récapitulatif des résultats de diffusion. Le tableau « Statistiques récapitulatives de l'analyse de diffusion » répertorie les statistiques récapitulatives disponibles.

**Table 3-2**  
Statistiques récapitulatives de l'analyse de diffusion

Statistiques	Description
Total des noeuds dans le réseau	Nombre de noeuds dans le réseau
Total des liens dans le réseau	Nombre de liens dans le réseau
Total des graines de diffusion dans le réseau	Nombre de noeuds utilisés comme graines pour le processus de diffusion
Influence moyenne	Quantité moyenne d'énergie diffusée associée aux individus.
Degré intérieur moyen	Nombre moyen de relations dans lesquelles un individu est la cible de la relation. Pour plus d'informations, reportez-vous à la section <a href="#">Degré des noeuds</a> dans le chapitre 1 sur p. 5.
Degré extérieur moyen	Nombre moyen des relations dans lesquelles un individu est la source de la relation. Pour plus d'informations, reportez-vous à la section <a href="#">Degré des noeuds</a> dans le chapitre 1 sur p. 5.

Sélectionnez une statistique spécifique dans le tableau Statistiques récapitulatives pour visualiser la distribution des valeurs, l'écart-type et l'asymétrie de la statistique.

## Résultats de l'analyse de diffusion

Le noeud Analyse de diffusion génère de nombreux champs qui décrivent les individus dans le réseau. Vous pouvez augmenter les modèles et données existants à l'aide de ces indicateurs de performances clés pour améliorer les prévisions générées par vos modèles. Par exemple, vous pouvez mettre à jour les valeurs de probabilité d'attrition individuelles pour qu'elles intègrent les influences de la diffusion.

Le tableau « Indicateurs de performances clés de l'analyse de diffusion » répertorie les champs des indicateurs de performances clés des individus dans l'analyse.

**Table 3-3**  
Indicateurs de performances clés de l'analyse de diffusion

Champ	Description
DA_NodeNumber	Identifiant unique d'un individu
DA_DiffusedEnergy	Quantité d'énergie diffusée associée à l'individu. Pour l'analyse de l'attrition, les valeurs les plus élevées indiquent une probabilité plus importante d'attrition que les valeurs les plus basses.
DA_InDegree	Nombre de relations dans lesquelles l'individu est la cible de la relation. Pour plus d'informations, reportez-vous à la section <a href="#">Degré des noeuds</a> dans le chapitre 1 sur p. 5.
DA_OutDegree	Nombre de relations dans lesquelles l'individu est la source de la relation. Pour plus d'informations, reportez-vous à la section <a href="#">Degré des noeuds</a> dans le chapitre 1 sur p. 5.

## ***Avis***

Ces informations ont été développées pour les produits et services offerts dans le monde.

Il est possible qu'IBM n'offre pas dans les autres pays les produits, services et fonctionnalités décrits dans ce document. Contactez votre représentant local IBM pour obtenir des informations sur les produits et services actuellement disponibles dans votre région. Toute référence à un produit, programme ou service IBM n'implique pas que les seuls les produits, programmes ou services IBM peuvent être utilisés. Tout produit, programme ou service de fonctionnalité équivalente qui ne viole pas la propriété intellectuelle IBM peut être utilisé à la place. Cependant l'utilisateur doit évaluer et vérifier l'utilisation d'un produit, programme ou service non IBM.

IBM peut posséder des brevets ou des applications de brevet en attente qui couvrent les sujets décrits dans ce document. L'octroi de ce document n'équivaut aucunement à celui d'une licence pour ces brevets. Vous pouvez envoyer par écrit des questions concernant la licence à :

*IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785, États-Unis*

**Le paragraphe suivant ne s'applique pas au Royaume-Uni ni à aucun pays dans lequel ces dispositions sont contraires au droit local :** INTERNATIONAL BUSINESS MACHINES FOURNIT CETTE PUBLICATION « EN L'ETAT » SANS GARANTIE D'AUCUNE SORTE, IMPLICITE OU EXPLICITE, Y COMPRIS, MAIS SANS ETRE LIMITE AUX GARANTIES IMPLICITES DE NON VIOLATION, DE QUALITE MARCHANDE OU D'ADAPTATION POUR UN USAGE PARTICULIER. Certains états n'autorisent pas l'exclusion de garanties explicites ou implicites lors de certaines transactions, par conséquent, il est possible que cet énoncé ne vous concerne pas.

Ces informations peuvent contenir des erreurs techniques ou des erreurs typographiques. Ces informations sont modifiées de temps en temps ; ces modifications seront intégrées aux nouvelles versions de la publication. IBM peut apporter des améliorations et/ou modifications des produits et/ou des programmes décrits dans cette publications à tout moment sans avertissement préalable.

Toute référence dans ces informations à des sites Web autres qu'IBM est fournie dans un but pratique uniquement et ne sert en aucun cas de recommandation pour ces sites Web. Le matériel contenu sur ces sites Web ne fait pas partie du matériel de ce produit IBM et l'utilisation de ces sites Web se fait à vos propres risques.

IBM peut utiliser ou distribuer les informations que vous lui fournissez, de la façon dont il le souhaite, sans encourir aucune obligation envers vous.

Les personnes disposant d'une licence pour ce programme et qui souhaitent obtenir des informations sur celui-ci pour activer : (i) l'échange d'informations entre des programmes créés de manière indépendante et d'autres programmes (notamment celui-ci) et (ii) l'utilisation mutuelle des informations qui ont été échangées, doivent contacter :

*IBM Software Group, Attention: Licensing, 233 S. Wacker Dr., Chicago, IL 60606, États-Unis.*

Ces informations peuvent être disponibles, soumises à des conditions générales, et dans certains cas payantes.

Le programme sous licence décrit dans ce document et toute la documentation sous licence disponible pour ce programme sont fournis par IBM en conformité avec les conditions de l'accord du client IBM, avec l'accord de licence du programme international IBM et avec tout accord équivalent entre nous.

Ces informations contiennent des exemples de données et de rapports utilisés au cours d'opérations quotidiennes standard. Pour les illustrer le mieux possible, ces exemples contiennent des noms d'individus, d'entreprises, de marques et de produits. Tous ces noms sont fictifs et toute ressemblance avec des noms et des adresses utilisés par une entreprise réelle ne serait que pure coïncidence.

### ***Marques commerciales***

IBM, le logo IBM, ibm.com et SPSS sont des marques commerciales d'IBM Corporation, déposées dans de nombreuses juridictions du monde entier. Une liste à jour des marques IBM est disponible sur Internet à l'adresse <http://www.ibm.com/legal/copytrade.shtml>.

Linux est une marque déposée de Linus Torvalds aux Etats-Unis et/ou dans d'autres pays.

Microsoft, Windows, Windows NT et le logo Windows sont des marques commerciales de Microsoft Corporation aux Etats-Unis et/ou dans d'autres pays.

UNIX est une marque déposée de The Open Group aux Etats-Unis et dans d'autres pays.

Les autres noms de produits et de services peuvent être des marques d'IBM ou d'autres sociétés.

---

# Glossaire

**chef d'autorité.** Individu dans un groupe qui possède le plus haut score d'autorité.

**chef de diffusion.** Individu dans un groupe qui possède le plus haut score de diffusion.

**degré.** Mesure d'une activité de nœud dans un réseau, définie comme le nombre total de relations du nœud.

**degré extérieur.** Mesure de centralité d'un nœud dans un réseau consistant en relations dirigées, définie comme le nombre total de relations pour lesquelles le nœud est la source.

**degré intérieur.** Mesure de prestige d'un nœud dans un réseau consistant en relations dirigées, définie comme le nombre total de relations pour lesquelles le nœud est la cible.

**densité.** Mesure de la cohésion de réseau définie comme le nombre de relations observées divisé par le nombre de relations possibles.

**énergie diffusée.** Effet commençant avec des nœuds spécifiques dans un réseau qui se propage dans l'ensemble du réseau, et dont la taille diminue à mesure que la distance du nœud initial s'agrandit.

**facteur de propagation.** Fraction d'énergie qu'un nœud transfère à ses voisins dans une étape de processus de diffusion.

**groupe central.** Groupe de nœuds restant après l'omission de relations plus faibles d'un réseau et l'application de limites de taille de groupe.

**relation de valeur.** Connexion entre deux nœuds qui comporte une pondération associée indiquant la pondération de la relation.

**relation dichotomique.** Connexion entre deux nœuds qui ne peut prendre qu'une seule des deux valeurs, indiquant la présence ou l'absence de connexion.

**relation dirigée.** Connexion entre deux nœuds qui provient d'un nœud (la source) et se termine à un autre nœud (la destination).

**relation non dirigée.** Association entre deux nœuds.

**score d'autorité.** Mesure indiquant la capacité des autres membres d'un groupe à se lier à un individu. Les scores d'autorité associent l'importance d'un individu et le nombre de relations qui se terminent avec lui.

**score de diffusion.** Mesure indiquant la capacité d'un individu à se lier aux autres membres d'un groupe. Les scores de diffusion associent l'importance d'un individu et le nombre de relations qui commencent avec lui.

**seuil de couverture.** Fraction des relations les plus fortes utilisées pour l'identification de groupe.

**seuil de précision.** Critère d'arrêt pour un processus de diffusion correspondant à la plus petite modification dans l'énergie du nœud qui garantit une diffusion continue.

**sociogramme.** Représentation visuelle d'un réseau social dans laquelle les lignes reliant des points représentent des individus en relation.

---

# ***Bibliographie***

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

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

---

# Index

- analyse de diffusion, 6, 20
  - configuration requise, 21
  - facteur de propagation, 19, 22
  - graines, 22
  - options, 21
  - seuil de précision, 19, 22
- analyse de groupe, 6
  - configuration requise, 14
  - seuil de couverture, 12, 15
  - tailles des groupes, 13, 15
- analyse des données, 7
- avis juridiques, 24
  
- champs de destination, 9, 15, 21
- champs de pondération, 9, 15, 21
- champs source, 8, 15, 21
- chefs d'autorité, 13, 17
- chefs de diffusion, 13, 18
  
- degré extérieur, 5, 16–17, 23
- degré extérieur moyen, 16, 23
- degré intérieur, 5, 16–17, 23
- degré intérieur moyen, 16, 23
- degree, 5
- densité, 4, 16–17
- densité kernel, 16–17
- densité moyenne des groupes, 16
- densité moyenne des groupes principaux, 16
  
- énergie diffusée, 23
- enregistrements des informations sur les appels, 8
- exécution de flux
  - utilisation de IBM SPSS Modeler Server Social Network Analysis, 6
  
- facteur de propagation, 19, 22
- filtrage des résultats, 7
- fraction moyenne des membres principaux, 16
  
- GAG\_GroupNumber, 16
- génération de scripts
  - propriétés, 9
- graines
  - pour l'analyse de diffusion, 22
- groupes principaux, 13
  
- indicateurs de performances clés, 16, 23
- influence moyenne, 23
  
- liste de graines de diffusion, 22
  
- marques commerciales, 25
  
- niveaux de mesure, 7
- noeuds principaux, 16–17
  
- onglet analyse, 7
  - analyse de diffusion, 22
  - analyse de groupe, 15
- onglet annotations, 7
- onglet données, 6
  - analyse de diffusion, 21
  - analyse de groupe, 14
- onglet filtrer, 7
- onglet options de création, 7
  - analyse de diffusion, 22
  - analyse de groupe, 15
- onglet types, 7
  
- prévisualisation des résultats, 7
- propriétés
  - génération de scripts, 9
  
- relations
  - de valeur, 2
  - direction, 2
- relations de valeur, 2
- relations dichotomiques, 2
- relations dirigées, 2
- relations non dirigées, 2
  
- scores d'autorité, 13, 17
- scores de diffusion, 13, 17
- seuil de couverture, 12, 15
- seuil de précision, 19, 22
- suppression des résultats, 8
  
- taille moyenne des groupes, 16
- tailles des groupes, 16
  - limites, 13, 15