

*IBM SPSS Modeler Social Network
Analysis 17 - Guide d'utilisation*

IBM

Important

Avant d'utiliser le présent document et le produit associé, prenez connaissance des informations générales figurant à la section «Remarques», à la page 21.

LE PRESENT DOCUMENT EST LIVRE EN L'ETAT SANS AUCUNE GARANTIE EXPLICITE OU IMPLICITE. IBM DECLINE NOTAMMENT TOUTE RESPONSABILITE RELATIVE A CES INFORMATIONS EN CAS DE CONTREFACON AINSI QU'EN CAS DE DEFAUT D'APTITUDE A L'EXECUTION D'UN TRAVAIL DONNE.

Ce document est mis à jour périodiquement. Chaque nouvelle édition inclut les mises à jour. Les informations qui y sont fournies sont susceptibles d'être modifiées avant que les produits décrits ne deviennent eux-mêmes disponibles. En outre, il peut contenir des informations ou des références concernant certains produits, logiciels ou services non annoncés dans ce pays. Cela ne signifie cependant pas qu'ils y seront annoncés.

Pour plus de détails, pour toute demande d'ordre technique, ou pour obtenir des exemplaires de documents IBM, référez-vous aux documents d'annonce disponibles dans votre pays, ou adressez-vous à votre partenaire commercial.

Vous pouvez également consulter les serveurs Internet suivants :

- <http://www.fr.ibm.com> (serveur IBM en France)
- <http://www.ibm.com/ca/fr> (serveur IBM au Canada)
- <http://www.ibm.com> (serveur IBM aux Etats-Unis)

*Compagnie IBM France
Direction Qualité
17, avenue de l'Europe
92275 Bois-Colombes Cedex*

Cette édition s'applique à la version 17.0.0 d'IBM(r) SPSS(r) Modeler et à toutes les éditions et modifications ultérieures sauf mention contraire dans les nouvelles éditions.

Table des matières

Avis aux lecteurs canadiens	v
Préface	vii
A propos d'IBM Business Analytics	vii
Assistance technique	vii
Chapitre 1. IBM SPSS Modeler Social Network Analysis.	1
A propos de IBM SPSS Modeler Social Network Analysis	1
A propos de l'analyse de réseaux sociaux	2
Affichage des réseaux	2
Description des réseaux.	3
Noeuds IBM SPSS Modeler Social Network Analysis	5
Onglets Noeud	5
Prévisualisation des résultats	6
Analyse des données	6
Applications	6
Structure des données	7
Propriétés de génération de scripts	8
Chapitre 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	12
Noeud Analyse de groupe	13
Spécification des données pour l'analyse de groupe	13
Configuration des options de création pour l'analyse de groupe.	14
Affichage des statistiques d'analyse de groupe.	14
Résultats de l'analyse de groupe	15
Chapitre 3. Analyse de diffusion.	17
Présentation de l'analyse de diffusion.	17
Exemple d'analyse de diffusion.	17
Noeud Analyse de diffusion.	18
Spécification de données pour l'analyse de diffusion	18
Configuration des options de création pour l'analyse de diffusion	19
Affichage des statistiques d'analyses de diffusion	20
Résultats de l'analyse de diffusion.	20
Remarques	21
Marques	22
Index	27

Avis aux lecteurs canadiens

Le présent document a été traduit en France. Voici les principales différences et particularités dont vous devez tenir compte.

Illustrations

Les illustrations sont fournies à titre d'exemple. Certaines peuvent contenir des données propres à la France.

Terminologie

La terminologie des titres IBM peut différer d'un pays à l'autre. Reportez-vous au tableau ci-dessous, au besoin.

IBM France	IBM Canada
ingénieur commercial	représentant
agence commerciale	succursale
ingénieur technico-commercial	informaticien
inspecteur	technicien du matériel

Claviers

Les lettres sont disposées différemment : le clavier français est de type AZERTY, et le clavier français-canadien de type QWERTY.








OS/2 et Windows - Paramètres canadiens

Au Canada, on utilise :

- les pages de codes 850 (multilingue) et 863 (français-canadien),
- le code pays 002,
- le code clavier CF.

Nomenclature

Les touches présentées dans le tableau d'équivalence suivant sont libellées différemment selon qu'il s'agit du clavier de la France, du clavier du Canada ou du clavier des États-Unis. Reportez-vous à ce tableau pour faire correspondre les touches françaises figurant dans le présent document aux touches de votre clavier.

France	Canada	Etats-Unis
 (Pos1)		Home
Fin	Fin	End
 (PgAr)		PgUp
 (PgAv)		PgDn
Inser	Inser	Ins
Suppr	Suppr	Del
Echap	Echap	Esc
Attn	Intrp	Break
Impr écran	ImpEc	PrtSc
Verr num	Num	Num Lock
Arrêt défil	Défil	Scroll Lock
 (Verr maj)	FixMaj	Caps Lock
AltGr	AltCar	Alt (à droite)

Brevets

Il est possible qu'IBM détienne des brevets ou qu'elle ait déposé des demandes de brevets portant sur certains sujets abordés dans ce document. Le fait qu'IBM vous fournisse le présent document ne signifie pas qu'elle vous accorde un permis d'utilisation de ces brevets. Vous pouvez envoyer, par écrit, vos demandes de renseignements relatives aux permis d'utilisation au directeur général des relations commerciales d'IBM, 3600 Steeles Avenue East, Markham, Ontario, L3R 9Z7.

Assistance téléphonique

Si vous avez besoin d'assistance ou si vous voulez commander du matériel, des logiciels et des publications IBM, contactez IBM direct au 1 800 465-1234.

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

Les logiciels IBM Business Analytics aident les entreprises à mesurer, comprendre et anticiper leur performance financière et opérationnelle en fournissant des informations exactes, cohérentes et complètes. 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.

Intégrée dans ce portefeuille, la solution logicielle IBM SPSS Predictive Analytics permet aux entreprises de prévoir les événements et d'agir proactivement en fonction de ces informations, afin d'obtenir de meilleurs résultats. Les clients des secteurs privé, public et universitaire du monde entier font appel à la technologie IBM SPSS, qui les dote d'un atout concurrentiel pour attirer, fidéliser et développer leur clientèle, tout en réduisant les fraudes et en atténuant les risques. L'intégration du logiciel IBM SPSS aux opérations quotidiennes transforme les organisations en entreprises prédictives, capables de guider et d'automatiser leurs décisions de manière à répondre aux objectifs métier et à obtenir un avantage concurrentiel mesurable. Pour plus d'informations ou pour contacter un représentant, visitez le site <http://www.ibm.com/spss>.

Assistance technique

L'assistance technique est réservée aux clients ayant signé un contrat 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>. Votre nom, celui de votre société, ainsi que votre contrat d'assistance vous seront demandés.

Chapitre 1. IBM SPSS Modeler Social Network Analysis

A 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 l'exactitude 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.

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.

Tableau 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 sujet¹.

Affichage des réseaux

Un réseau social est généralement représenté par un *sociogramme*². 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 libellés des arêtes indiquent les valeurs. Le graphique suivant affiche un réseau de sept individus.

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

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

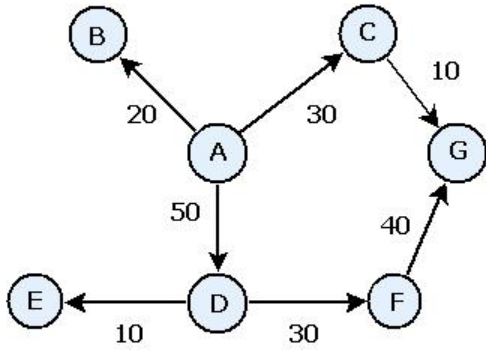


Figure 1. Exemple de réseau social

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.

Tableau 2. Relations dirigées possibles pour trois noeuds.

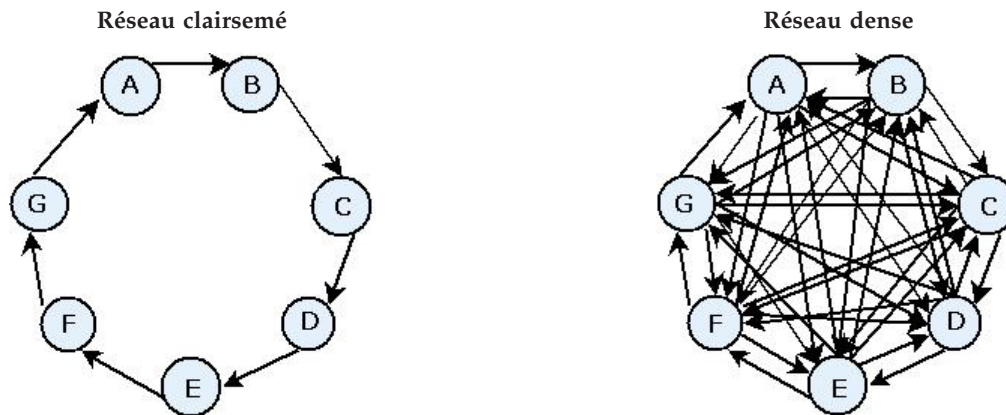
Source	Cible
A	B
A	C
B	A
B	C
C	A
C	B

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.

Tableau 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".

Tableau 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

Tableau 4. Exemples de valeurs de degré, degré intérieur et degré extérieur (suite).

Noeud	Degré	Degré intérieur	Degré extérieur
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.

Tableau 5. noeuds IBM SPSS Modeler Social Network Analysis.

Noeud	Icône	Description
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, voir la rubrique «Présentation de l'analyse de groupe», à la page 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 résumant les résultats de cet effet sur les noeuds individuels. Pour plus d'informations, voir la rubrique «Présentation de l'analyse de diffusion.», à la page 17.

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 génération.** 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 Filtre.** Permet d'éliminer ou de renommer les champs de résultats produits par les noeuds. Cet onglet offre la même fonctionnalités que le nœud Filtrer.
- **Onglet Types.** Permet de définir les niveaux de mesure des champs de résultats produits par les noeuds. Cet onglet offre la même fonctionnalité que le noeud type.
- **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 IBM 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 :

- **Prévision d'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.

Deux flux de démonstration sont fournis avec IBM SPSS Modeler Social Network Analysis pour vous proposer des exemples montrant comment vous pouvez inclure les résultats de l'analyse des réseaux sociaux dans vos flux. Les fichiers de données et les flux d'échantillons sont installés dans le dossier *Demos*, sous le répertoire d'installation du produit.

- *DA demo streams.str* fournit un exemple d'analyse de données permettant d'identifier les 300 personnes ayant la probabilité d'attrition la plus élevée.
- *GA demo streams.str* fournit un exemple d'utilisation d'indicateurs clés de performance permettant de prédire l'attrition de groupes et d'individus, et d'utilisation d'indicateurs clés de performance visant à cibler des individus spécifiques pour une 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.

Veillez noter que les enregistrements en-tête doivent également utiliser soit des nombres, soit des valeurs non renseignées. Par exemple, un en-tête dont la valeur est *EF BB BF* provoquerait une erreur.

Tableau 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. Notez que cela est vrai pour l'analyse de diffusion, mais que l'analyse de groupe prend uniquement en charge pour les fréquences 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
nom_fichier_données_entrée	chaîne	
champ_appel	champ	
champ_appelé	champ	
champ_pondération_ fréquence	champ	
noms_champ_lecture	booléen	
nom_fichier_liste_ diffusion	chaîne	
facteur_propagation	double	Valeur par défaut = 50. Min = 1 Max = 99.
itérations_nombre_max	entier	Valeur par défaut = 100. Min = 1.
seuil_exactitude	double	Valeur par défaut = 0.01. Min = 0.001.
calcul_statistiques	booléen	

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
nom_fichier_données_entrée	chaîne	
champ_appel	champ	
champ_appelé	champ	

Nom de la propriété	Le type de données	Description de la propriété
champ_pondération_ fréquence	<i>champ</i>	
noms_champ_lecture	<i>booléen</i>	
seuil_couverture	<i>double</i>	Valeur par défaut = 10. Min = 1 Max = 99.
taille_groupe_min	<i>entier</i>	Valeur par défaut = 2. Min = 2.
taille_groupe_max	<i>entier</i>	Valeur par défaut = 100. Min = 2.
calcul_statistiques	<i>booléen</i>	

Chapitre 2. 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.³ 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, voir la rubrique «Déterminer la similarité sociale».
2. Scinder le réseau en groupes en fonction de la pondération relationnelle tout en obéissant aux restrictions sur la taille. Pour plus d'informations, voir la rubrique «Scission en groupes», à la page 12.
3. Effectuer un profil des groupes et des individus, en identifiant notamment les chefs des groupes. Pour plus d'informations, voir la rubrique «Description des groupes et des membres de groupe», à la page 12.

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".

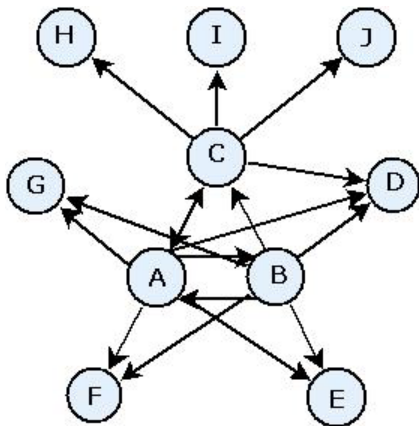


Figure 2. Exemple de réseau à dix noeuds

3. 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.

Le tableau "Noeuds cible" présente les noeuds cibles des relations commençant aux noeuds A, B et C.

Tableau 7. 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 mutuelles⁴ 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

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

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 délimiteurs. 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, voir la rubrique «Scission en groupes», à la page 12.

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

Taille maximale du groupe. 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.

Tableau 8. 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, voir la rubrique «Densité du réseau», à la page 3.

Tableau 8. Statistiques récapitulatives des analyses de groupe (suite).

Statistiques	Description
Fraction moyenne des membres principaux	Fraction de noeuds moyenne dans un groupe qui sont des noeuds principaux du groupe. Pour plus d'informations, voir la rubrique «Scission en groupes», à la page 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, voir la rubrique «Degré des noeuds», à la page 4.
Degré extérieur moyen	Nombre de liens sortants moyen. Pour plus d'informations, voir la rubrique «Degré des noeuds», à la page 4.

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.

Tableau 9. Indicateurs de performances clés pour les groupes.

Champ	Description
GAG_GroupNumber	Identifiant unique d'un groupe
GAG_Size	Nombre d'individus dans un groupe
GAG_Density	Fraction de connexions directes entre les individus d'un groupe. Pour plus d'informations, voir la rubrique «Densité du réseau», à la page 3.
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, voir la rubrique «Scission en groupes», à la page 12.
GAG_MaxRankType1	Score d'autorité maximum d'un des membres du groupe. Pour plus d'informations, voir la rubrique «Description des groupes et des membres de groupe», à la page 12.
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, voir la rubrique «Description des groupes et des membres de groupe», à la page 12.

Tableau 9. Indicateurs de performances clés pour les groupes (suite).

Champ	Description
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.

Tableau 10. 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, voir la rubrique «Scission en groupes», à la page 12.
GAI_RankType1	Score d'autorité de l'individu. Pour plus d'informations, voir la rubrique «Description des groupes et des membres de groupe», à la page 12.
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, voir la rubrique «Description des groupes et des membres de groupe», à la page 12.
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, voir la rubrique «Degré des noeuds», à la page 4.
GAI_OutDegree	Nombre de relations dans lesquelles l'individu est la source de la relation. Pour plus d'informations, voir la rubrique «Degré des noeuds», à la page 4.
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, voir la rubrique «Description des groupes et des membres de groupe», à la page 12.
GAI_GroupLeaderConfidenceType1	Le niveau de confiance au sujet du noeud comme chef d'autorité.
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, voir la rubrique «Description des groupes et des membres de groupe», à la page 12.
GAI_GroupLeaderConfidenceType2	Le niveau de confiance au sujet du noeud comme chef de diffusion.

Chapitre 3. 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'autre⁵. 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 d'exactitude*, 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 «Affichage des réseaux», à la page 2. 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.

5. Dasgupta, K., R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjee, 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.

Tableau 11. 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 conséquente 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 délimiteurs. 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 valeurs de départ de diffusion

La liste de valeurs de départ 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.

Itérations maximales. 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 d'exactitude. 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 des analyses de groupe" répertorie les statistiques récapitulatives disponibles.

Tableau 12. Statistiques récapitulatives de l'analyse de diffusion.

Statistiques	Description
Total des noeuds dans le réseau	Nombre de noeuds dans le réseau
Nombre total de liens sur le réseau	Nombre de liens dans le réseau
Total des valeurs de départ de diffusion dans le réseau	Nombre de noeuds utilisés comme valeurs de départ 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, voir la rubrique «Degré des noeuds», à la page 4.
Degré extérieur moyen	Nombre moyen des relations dans lesquelles un individu est la source de la relation. Pour plus d'informations, voir la rubrique «Degré des noeuds», à la page 4.

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 de l'analyse.

Tableau 13. 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, voir la rubrique «Degré des noeuds», à la page 4.
DA_OutDegree	Nombre de relations dans lesquelles l'individu est la source de la relation. Pour plus d'informations, voir la rubrique «Degré des noeuds», à la page 4.

Remarques

Ces informations ont été développées pour des produits et des services proposés dans le monde entier.

Le présent document peut contenir des informations ou des références concernant certains produits, logiciels ou services IBM non annoncés dans ce pays. Pour plus de détails, référez-vous aux documents d'annonce disponibles dans votre pays, ou adressez-vous à votre partenaire commercial IBM. Toute référence à un produit, programme ou service IBM n'implique pas que seul ce produit, programme ou service IBM puisse être utilisé. Tout produit, programme ou service fonctionnellement équivalent peut être utilisé s'il n'enfreint aucun droit de propriété intellectuelle d'IBM. Cependant l'utilisateur doit évaluer et vérifier l'utilisation d'un produit, programme ou service non IBM.

IBM peut détenir des brevets ou des demandes de brevet couvrant les produits mentionnés dans le présent 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
U.S.A.

Pour le Canada, veuillez adresser votre courrier à :

IBM Director of Commercial Relations
IBM Canada Ltd
3600 Steeles Avenue East
Markham, Ontario
L3R 9Z7
Canada

Pour toute demande au sujet des licences concernant les jeux de caractères codés sur deux octets (DBCS), contactez le service Propriété intellectuelle IBM de votre pays ou adressez vos questions par écrit à :

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 Japon

Le paragraphe suivant ne s'applique ni au Royaume-Uni, ni dans aucun pays dans lequel il serait contraire aux lois locales. LE PRESENT DOCUMENT EST LIVRE EN L'ETAT SANS AUCUNE GARANTIE EXPLICITE OU IMPLICITE. IBM DECLINE NOTAMMENT TOUTE RESPONSABILITE RELATIVE A CES INFORMATIONS EN CAS DE CONTREFACON AINSI QU'EN CAS DE DEFAUT D'APTITUDE A L'EXECUTION D'UN TRAVAIL DONNE. 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. Ce document est mis à jour périodiquement. Chaque nouvelle édition inclut les mises à jour. IBM peut, à tout moment et sans préavis, modifier les produits et logiciels décrits dans ce document.

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. Les éléments figurant sur ces sites Web ne font pas partie des éléments du présent produit IBM et l'utilisation de ces sites relève de votre seule responsabilité.

IBM pourra utiliser ou diffuser, de toute manière qu'elle jugera appropriée et sans aucune obligation à votre égard, tout ou partie des informations qui lui seront fournies.

Les licenciés souhaitant obtenir des informations permettant : (i) l'échange des données entre des logiciels créés de façon indépendante et d'autres logiciels (dont celui-ci), et (ii) l'utilisation mutuelle des données ainsi échangées, doivent adresser leur demande à :

IBM Software Group
ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
U.S.A.

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

Le programme sous licence décrit dans le présent document et tous les éléments sous licence disponibles s'y rapportant sont fournis par IBM conformément aux dispositions du Livret Contractuel IBM, des Conditions internationales d'utilisation des Logiciels IBM ou de tout autre contrat équivalent.

Toutes les données sur les performances contenues dans le présent document ont été obtenues dans un environnement contrôlé. Par conséquent, les résultats obtenus dans d'autres environnements d'exploitation peuvent varier de manière significative. Certaines mesures peuvent avoir été effectuées sur des systèmes en cours de développement et il est impossible de garantir que ces mesures seront les mêmes sur les systèmes commercialisés. De plus, certaines mesures peuvent avoir été estimées par extrapolation. Les résultats réels peuvent être différents. Les utilisateurs de ce document doivent vérifier les données applicables à leur environnement spécifique.

les informations concernant les produits autres qu'IBM ont été obtenues auprès des fabricants de ces produits, leurs annonces publiques ou d'autres sources publiques disponibles. IBM n'a pas testé ces produits et ne peut confirmer l'exactitude de leurs performances ni leur compatibilité. Aucune réclamation relative à des produits non IBM ne pourra être reçue par IBM. Les questions sur les capacités de produits autres qu'IBM doivent être adressées aux fabricants de ces produits.

Toutes les déclarations concernant la direction ou les intentions futures d'IBM peuvent être modifiées ou retirées sans avertissement préalable et représentent uniquement des buts et des objectifs.

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.

Si vous consultez la version papier de ces informations, il est possible que certaines photographies et illustrations en couleurs n'apparaissent pas.

Marques

IBM, le logo IBM et ibm.com sont des marques d'International Business Machines dans de nombreux pays. Les autres noms de produits et de services peuvent être des marques d'IBM ou d'autres sociétés. La liste actualisée de toutes les marques d'IBM est disponible sur la page Web "Copyright and trademark information" à l'adresse www.ibm.com/legal/copytrade.shtml.

Intel, le logo Intel, Intel Inside, le logo Intel Inside, Intel Centrino, le logo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium, et Pentium sont des marques commerciales ou des marques déposées de Intel Corporation ou de ses filiales aux Etats-Unis et dans d'autres pays.

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 marques commerciales Java et basées sur Java ainsi que les logos sont des marques commerciales ou déposées de Oracle et/ou de ses filiales.

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

Glossaire

C

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.

D

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.

E

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.

F

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

G

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.

R

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.

S

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 d'exactitude : Critère d'arrêt pour un processus de diffusion correspondant à la plus petite modification dans l'énergie du noeud qui garantit une diffusion continue.

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

Index

A

analyse de diffusion 5, 17
 configuration requise 18
 facteur de propagation 17, 19
 options 18
 seuil d'exactitude 17, 19
 valeurs de départ 19
analyse de groupe 5
 configuration requise 13
 seuil de couverture 12, 14
 tailles des groupes 12, 14
analyse des données 6

C

champs de destination 7
champs de pondération 7, 13, 18
champs source 7
chefs d'autorité 12, 15
chefs de diffusion 12, 15

D

degré 4
degré extérieur 4, 14, 15, 20
degré extérieur moyen 14, 20
degré intérieur 4, 14, 15, 20
degré intérieur moyen 14, 20
densité 3, 14, 15
densité kernel 14, 15
densité moyenne des groupes 14
densité moyenne des groupes
 principaux 14

E

énergie diffusée 20
enregistrements des informations sur les
 appels 7
exécution de flux
 utilisation d'IBM SPSS Modeler Server
 Social Network Analysis 5

F

facteur de propagation 17, 19
filtrage des résultats 5
fraction moyenne des membres
 principaux 14

G

GAG_GroupNumber 15
génération de scripts
 propriétés 8
groupes principaux 12

I

indicateurs de performances clés 15, 20
influence moyenne 20

L

liste de valeurs de départ de
 diffusion 19

N

niveaux de mesure 5
noeuds principaux 14, 15

O

onglet analyse 5
 analyse de diffusion 20
 analyse de groupe 14
onglet annotations 5
onglet données 5
 analyse de diffusion 18
 analyse de groupe 13
onglet filtrer 5
onglet options de création 5
 analyse de diffusion 19
 analyse de groupe 14
onglet types 5

P

prévisualisation des résultats 6
propriétés
 génération de scripts 8

R

relations
 de valeur 2
 direction 2
relations de valeur 2
relations dichotomiques 2
relations dirigées 2
relations non dirigées 2

S

scores d'autorité 12, 15
scores de diffusion 12, 15
seuil d'exactitude 17, 19
seuil de couverture 12, 14
suppression des résultats 6

T

taille moyenne des groupes 14
tailles des groupes 14, 15

tailles des groupes (*suite*)
 limites 12, 14

V

valeurs de départ
 pour l'analyse de diffusion 19

