

Mémoire présenté devant
l'UFR de Mathématique et Informatique
pour l'obtention du du Diplôme Universitaire d'Actuaire de Strasbourg
et l'admission à l'Institut des Actuaire

le 22/11/2018

Par : Thomas Bourdin

Titre: Application de la théorie des graphes à l'étude des modèles actuariels

Confidentialité : NON OUI Durée : 1 an 2 ans 3 ans 4 ans 5 ans

Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

Signature : _____

Membres du jury de l'Unistra :

Entreprise : **ACTUARIS**

Directeur de mémoire en entreprise :

Nom : Arnaud Becquet

Signature : _____

Invité : _____

P. ARTZNER
J. BERARD
F. BERTRAND
A. COUSIN
K.-T. EISELE
J. FRANCHI
M. MAUMY-BERTRAND
V. VIGON

Jury de l'Institut des
Actuaire :

Nom : _____

Signature : _____

**Autorisation de publication et de
mise en ligne sur un site de
diffusion de documents
actuariels (après expiration de
l'éventuel délai de confidentialité)**

Signature du responsable entreprise

Signature du candidat

Bourdin

Secrétariat : *Mme Stéphanie Richard*

Bibliothèque : *Mme Christine Disdier*

Résumé

Les modèles sont de plus en plus présents en actuariat depuis les exigences mises en place par Solvabilité 2. La complexité réglementaire se retrouve dans les travaux de modélisation. Étant donné que les modèles actuariels jouent un rôle prépondérant dans l'évaluation des risques, leur maîtrise représente un véritable enjeu dans la gestion des risques. L'objectif de ce mémoire est de proposer des méthodes s'appuyant sur la théorie des graphes pour représenter les interactions entre les différents modules des modèles actuariels afin de les contrôler et de faciliter leur compréhension. Une approche statistique appliquée à des simulations sera ensuite proposée pour pondérer les liens de nos graphes.

Mots-clés : Modélisation actuarielle, contrôle de l'implémentation, théorie des graphes, DAG, dépendance informatique, réseaux bayésiens.

Abstract

Models are more and more used in actuarial science since Solvency II came into effect. This complex regulation is part of every actuarial model. As actuarial models play a key role in risk management, their mastery represent an important issue. The main objective of this report is to develop methods which use graph theory in order to analyze the relations between the different parts of actuarial models. The tools have to control and summarize these relations. A statistical method will be applied to simulations in order to improve the first method.

Keywords : Actuarial modeling, test implementation, graph theory, *DAG*, coupling, bayesian network.

Table des matières

| | |
|--|----|
| Résumé | 2 |
| Abstract..... | 3 |
| Table des matières | 4 |
| Liste des tableaux | 7 |
| Liste des figures..... | 8 |
| Liste des acronymes..... | 10 |
| Remerciements | 12 |
| Introduction | 13 |
| 1. Rationaliser la modélisation actuarielle..... | 15 |
| 1.1. Contexte réglementaire..... | 15 |
| 1.2. Modèle Prévoyance Santé | 16 |
| 1.2.1. Présentation du modèle Prévoyance Santé | 17 |
| 1.2.2. Résumé de l'architecture | 17 |
| 1.2.2.1. Passif..... | 18 |
| 1.2.2.2. Actif | 20 |
| 1.2.2.3. Solvabilité 2..... | 21 |
| 1.3. Contrôles des modèles actuariels..... | 21 |
| Synthèse..... | 23 |
| 2. Cartographie informatique des modèles actuariels..... | 24 |
| 2.1. Applications de la théorie des graphes | 25 |
| 2.1.1. Présentation des données à disposition..... | 25 |
| 2.1.2. Principes de la théorie des graphes..... | 28 |
| 2.1.2.1. Définitions et applications de la théorie des graphes | 29 |
| 2.1.2.2. Tracés des graphes..... | 36 |
| 2.1.2.2.1. Algorithmes de placement par des forces | 36 |
| 2.1.2.2.2. Algorithmes hiérarchiques | 38 |
| Synthèse de la partie 2.1 | 39 |
| 2.2. Réalisation des cartes..... | 42 |
| 2.2.1. Méthode pour visualiser les interactions | 42 |

| | | |
|----------|--|----|
| 2.2.1.1. | Première approche | 42 |
| 2.2.1.2. | Méthode de regroupement pour la restitution de l'architecture du modèle... | 44 |
| 2.2.2. | Méthode pour visualiser les calculs..... | 49 |
| | Synthèse de la partie 2.2..... | 52 |
| 2.3. | Applications à la modélisation actuarielle..... | 53 |
| 2.3.1. | Présentation des hypothèses de modélisation..... | 53 |
| 2.3.2. | Calcul de la provision pour risque d'exigibilité | 57 |
| 2.3.2.1. | Réorganisation du calcul de la provision pour risque d'exigibilité..... | 57 |
| 2.3.2.2. | Impacts de la provision pour risque d'exigibilité | 60 |
| 2.3.3. | Calcul de la cession des <i>Best Estimates</i> | 64 |
| 2.3.3.1. | Réorganisation du calcul des <i>Best Estimates NSLT</i> | 64 |
| 2.3.3.2. | Impacts sur la cession des <i>Best Estimates NSLT</i> | 68 |
| | Synthèse de la partie 2.3..... | 71 |
| | Synthèse de la cartographie par les dépendances informatiques..... | 72 |
| 3. | Approche statistique de la cartographie..... | 73 |
| 3.1. | Présentation des réseaux bayésiens | 73 |
| 3.1.1. | Définition d'un réseau bayésien | 73 |
| 3.1.2. | Dépendances probabilistes | 74 |
| 3.1.3. | Loi jointe d'un réseau bayésien | 76 |
| 3.2. | Apprentissage d'un réseau bayésien..... | 77 |
| 3.2.1. | Présentation des réseaux bayésiens gaussiens | 77 |
| 3.2.2. | Apprentissage de la structure..... | 78 |
| 3.2.2.1. | Les algorithmes par contraintes | 78 |
| 3.2.2.2. | Les algorithmes par maximisation du score | 80 |
| 3.3. | Application pour le modèle standard..... | 80 |
| 3.3.1. | Mise en place de la démarche..... | 81 |
| 3.3.1.1. | Présentation des variables étudiées | 81 |
| 3.3.1.2. | Vérification des conditions de normalité..... | 82 |
| 3.3.2. | Alternative non paramétrique | 86 |
| 3.3.2.1. | Coefficient de corrélation de Spearman | 86 |
| 3.3.2.2. | Pondération des liens de la cartographie informatique..... | 87 |

| | |
|---|----|
| Synthèse de l'approche statistique..... | 92 |
| Conclusion..... | 93 |
| Bibliographie | 95 |
| Annexe A..... | i |
| Annexe B | ii |

Liste des tableaux

| | | |
|---------------|---|----|
| Tableau I. | Tableau récapitulatif des concepts de la théorie des graphes | 40 |
| Tableau II. | Tableau comparatif des deux familles algorithmes de visualisation | 41 |
| Tableau III. | Tableau de synthèses des cartes des modèles actuariels..... | 52 |
| Tableau IV. | Correspondance entre les produits modélisées et les lignes d'activité | 54 |
| Tableau V. | Principales caractéristiques des produits de l'organisme tableau étudié..... | 55 |
| Tableau VI. | Passif du bilan prudentiel de l'organisme..... | 55 |
| Tableau VII. | Actif du bilan prudentiel de l'organisme | 56 |
| Tableau VIII. | Capitaux de solvabilité réglementaire de l'organisme | 56 |
| Tableau IX. | Décomposition des fonds propres S2 pour les deux situations étudiées | 62 |
| Tableau X. | Hypothèses retenues pour le taux d'imposition..... | 63 |
| Tableau XI. | Evolution des postes de passif du bilan prudentiel entre les deux modèles | 69 |
| Tableau XII. | Evolution des postes d'actif du bilan prudentiel entre les deux modèles..... | 69 |
| Tableau XIII. | Evolution des montants de <i>MCR</i> entre les deux modèles..... | 70 |
| Tableau XIV. | Tableau de synthèse des développements effectués à l'aide des cartes | 71 |
| Tableau XV. | Tableau des sorties des tests de Shapiro Wilk..... | 85 |

Liste des figures

| | | |
|------------|--|----|
| Figure 1. | Schéma des trois piliers de Solvabilité 2..... | 16 |
| Figure 2. | Schéma de l'architecture d'un modèle type sur addactis® Modeling..... | 17 |
| Figure 3. | Architecture commune des modèles de la partie Passif | 19 |
| Figure 4. | Fichier des dépendances du modèle Comptes | 26 |
| Figure 5. | Schéma de la transcription des formules | 27 |
| Figure 6. | Onglet Architecture du modèle Comptes | 28 |
| Figure 7. | Exemple de visualisation du graphe du modèle Comptes | 29 |
| Figure 8. | Visualisation du plus court chemin des cotisations moyennes aux prestations payées 31 | |
| Figure 9. | Visualisation des antécédents des chargements dans le modèle Comptes | 32 |
| Figure 10. | Visualisation d'un cycle dans le modèle Comptes | 33 |
| Figure 11. | Visualisation du modèle Comptes après l'ajout du calcul des frais | 35 |
| Figure 12. | Visualisation à l'aide des degrés des nœuds..... | 36 |
| Figure 13. | Visualisation du modèle Comptes à l'aide de l'algorithme de Fruchterman Reingold | 38 |
| Figure 14. | Architecture du modèle Prévoyance Santé..... | 42 |
| Figure 15. | Première version de la carte du modèle Prévoyance Santé | 43 |
| Figure 16. | Visualisation de la structure modifiée du graphe du modèle Comptes | 45 |
| Figure 17. | Visualisation du modèle Prévoyance Santé à l'aide de la méthode de regroupement | 46 |
| Figure 18. | Visualisation des segments de passif du modèle Prévoyance Santé | 47 |
| Figure 19. | Visualisation du module Cotisations | 50 |
| Figure 20. | Visualisation des descendants de l'indicatrice de <i>run-off</i> | 51 |
| Figure 21. | Carte des variables en sortie du module Santé | 57 |
| Figure 22. | Carte des variables de l'Actif dépendant de la Santé..... | 58 |
| Figure 23. | Visualisation des descendants de la PRE | 59 |
| Figure 24. | Visualisation des descendants de la PRE après correction..... | 60 |
| Figure 25. | Evolution des capitaux de solvabilité réglementaire dans un scénario de rendement défavorable | 61 |

| | | |
|------------|---|----|
| Figure 26. | Visualisation de la première version du module de <i>Best Estimates NSLT</i> | 65 |
| Figure 27. | Visualisation des ancêtres de la table <i>BEL_C</i> | 66 |
| Figure 28. | Visualisation du module de calculs de <i>Best Estimates</i> pour la santé après le développement..... | 67 |
| Figure 29. | Visualisation des ancêtres de la table <i>BEL_C</i> après le développement..... | 68 |
| Figure 30. | Exemple de DAG | 75 |
| Figure 31. | Tracé des diagrammes quantile quantile pour les principales variables..... | 83 |
| Figure 32. | Histogrammes des variables <i>SCR</i> et <i>BE Life</i> | 84 |
| Figure 33. | Graphe du modèle de l'étude obtenues par la cartographie informatique | 88 |
| Figure 34. | Lien de dépendance informatique entre le <i>BE SLT</i> et le <i>SCR</i> immobilier..... | 88 |
| Figure 35. | Graphe obtenu avec la coloration des liens par l'estimation du coefficient de corrélation | 89 |
| Figure 36. | Lien entre l'hypothèse de ratio S/P S2 en santé et le <i>Best Estimate NSLT</i> | 90 |
| Figure 37. | Pseudo code de l'algorithme de Fruchterman Reingold | i |

Liste des acronymes

ALM : *Asset liability management* : gestion actif passif

BE : *Best Estimate* : Meilleure Estimation

DAG : *Directed acyclic graph* : graphe acyclique orienté

ESG : *Economic scenario generator* : générateur de scénarios économiques

IBNR : *Incured but not reported* : provision pour sinistre survenu non déclaré

IDA : Impôt différé actif

IDP : Impôt différé passif

LGD : *Loss given default* : perte en cas de défaut

LoB : *Line of business* : ligne d'activité

MCR : *Minimum capital requirement* : capital minimal requis

NSLT : *Non similar to life techniques* : santé similaire à la non-vie

OPCVM : Organisme de placements collectif en valeurs mobilières

PANE : Primes acquises non émises

PEg : Provisions pour égalisation

PPE : Provisions pour participations aux excédents

PPNA : Provision pour primes non acquises

PRE : Provision pour risque d'exigibilité

PSAP : Provision pour sinistre à payer

QRT : *Quantitative reporting template* : état de reporting quantitatif

QQ Plot : *Quantile quantile plot* : diagramme quantile-quantile

Risk margin : marge de risque

RM : *Risk mitigation* : atténuation du risque

RSR : *Regular supervisory report* : rapport régulier au contrôleur

S1 : Solvabilité 1

S2 : Solvabilité 2

S/P : Ratio des sinistres à primes

SCR : *Solvency capital requirement* : capital de solvabilité requis

SFCR : *Solvency and Financial Condition Report* : rapport sur la solvabilité et la situation financière

Remerciements

Je tiens à remercier l'ensemble de mes collègues de l'équipe modélisation qui ont veillé à ma bonne intégration au sein du cabinet ACTUARIS. Les nombreux échanges que nous avons eu m'ont permis de progresser sur des sujets variés de l'actuariat et se sont toujours tenus dans un cadre convivial.

Je pense aussi à mes collègues d'addactis ® Software qui ont su me transmettre leurs connaissances techniques sur leurs logiciels.

Introduction

Les organismes d'assurance ont besoin de disposer de visibilité sur leurs activités futures afin de prendre les décisions stratégiques adéquates. Pour ce faire ils ont recours à la modélisation pour appréhender les différents risques auxquels ils sont confrontés. Solvabilité 2 a accentué cette nécessité avec de nouvelles exigences réglementaires. Ainsi la directive les oblige à évaluer leur solvabilité mais également à mesurer et contrôler leurs risques dans le temps à l'aide de modèles actuariels.

La modélisation se traduit par des développements informatiques qui doivent permettre de prendre en compte précisément les différents aspects métier, techniques, réglementaires tout en étant réalisée conjointement par plusieurs personnes. Elle doit par conséquent être maîtrisée et contrôlée pour que le modèle fournisse des résultats justes et cohérents. Divers tests sont ainsi réalisés à chaque étape des développements pour assurer la pertinence du modèle. Cependant la validation de programmes informatiques faisant intervenir des milliers de variables reste complexe. En effet il est nécessaire pour apporter d'éventuelles corrections ou améliorations d'avoir une vision globale de l'organisation des calculs ainsi qu'une connaissance approfondie dans les montants modélisés. C'est pour répondre à ces problèmes que nous proposons un outil utilisant la théorie des graphes. Cette discipline est utilisée dans de nombreux domaines pour représenter et résoudre les problèmes engendrés par des réseaux. Étant donné les nombreuses interactions entre les calculs effectués, il est donc justifié de les utiliser pour nos problèmes de modélisation actuarielle. Notre outil utilisera donc les différents algorithmes de la théorie des graphes pour répondre à ses deux objectifs :

- résumer les calculs effectués ;
- améliorer les contrôles.

Pour ce faire nous utiliserons dans un premier temps les données de dépendances informatiques entre les variables générées par le logiciel de modélisation. Une variable A dépend informatiquement d'une autre variable B si B intervient dans le calcul de A. Nous verrons comment cette dépendance informatique peut nous permettre de répondre aux objectifs énoncés précédemment pour le modèle Prévoyance Santé développé par le cabinet de conseil

ACTUARIS. Pour cela nous appliquerons différents concepts de la théorie des graphes et nous utiliserons des algorithmes de tracés.

Dans un deuxième temps, nous chercherons à déterminer des dépendances probabilistes entre les variables de nos modèles actuariels. Pour ce faire nous lancerons à plusieurs reprises le modèle Prévoyance Santé pour obtenir les valeurs calculées des variables. Pour les utiliser nous ferons appel aux réseaux bayésiens et plus précisément aux algorithmes d'apprentissage de la structure. Cependant l'examen des hypothèses nous conduira à privilégier une autre méthode non-paramétrique.

1. Rationaliser la modélisation actuarielle

L'un des objectifs du modèle actuariel que nous considérons est de répondre aux exigences réglementaires imposées par la directive Solvabilité 2. Cette finalité oriente l'ensemble des développements effectués, il est donc essentiel de commencer par la présenter. Nous expliquerons par la suite le fonctionnement général du modèle Prévoyance Santé qui en découle. Enfin nous nous intéresserons aux contrôles effectués permettant de s'assurer de la justesse des calculs effectués.

1.1. Contexte réglementaire

L'objectif principal de la directive est de mesurer la solvabilité des organismes d'assurance en considérant l'ensemble de leurs risques. Pour ce faire, la réforme s'organise en 3 piliers distincts :

- **Pilier 1** (Exigences quantitatives) : expose les méthodologies de calcul des provisions techniques (vision *Best Estimate*), du *MCR* et du *SCR* ainsi que la classification des fonds propres ;
- **Pilier 2** (Exigences qualitatives et de supervision) : définit le système de gouvernance et la mise en place d'un système de gestion des risque interne à la société ;
- **Pilier 3** (Discipline de marché) : définit les différents reporting narratifs (*SFCR* pour le public et *RSR* pour le régulateur) et quantitatifs (*QRT*).



Figure 1. Schéma des trois piliers de Solvabilité 2

Le premier et le deuxième pilier de Solvabilité 2 renforcent considérablement la modélisation des différents risques. En effet les *Best Estimates* et les *SCR* sont obtenus par des calculs sophistiqués qui font intervenir un nombre important d'hypothèses techniques par rapport à ceux effectués précédemment sous Solvabilité 1. Nous avons par exemple : les notions de valeur de marché, la courbe des taux, les probabilités de défaut des réassureurs, les matrices de corrélation des *SCR*, les différents chocs de la formule standard, etc...

Même si la complexité des exigences réglementaires a augmenté, il est essentiel de garder le contrôle sur la modélisation des risques pour y répondre de manière juste et efficace. Une modélisation organisée facilitera ainsi l'analyse des résultats et la réalisation de divers tests de sensibilité qui permettront aux organismes d'assurance de pleinement maîtriser leurs risques.

1.2. Modèle Prévoyance Santé

Après avoir présenté les enjeux de la modélisation réglementaire, nous allons voir comment l'organisation du modèle Prévoyance Santé permet d'y répondre. Nous présenterons succinctement dans cette partie le modèle de notre étude. Cette présentation n'a pas pour objectif d'être exhaustive étant donné le nombre conséquent de calculs effectués.

1.2.1. Présentation du modèle Prévoyance Santé

Nous considérons pour notre étude le modèle Prévoyance Santé implémenté sous le logiciel de modélisation actuarielle addactis® Modeling. Dans son fonctionnement standard, le modèle Prévoyance Santé effectue les différents calculs réglementaires de manière déterministe. La modélisation déterministe permet notamment d'effectuer de nombreux scénarios étant donné que les temps de calcul se retrouvent réduits.

Par ailleurs, le modèle fonctionne avec une architecture modulaire qui permet de :

- Répliquer une partie identique de modèle à travers différents modèles ou au sein d'un même modèle ;
- Maintenir et mettre à jour automatiquement les différentes parties du modèle.

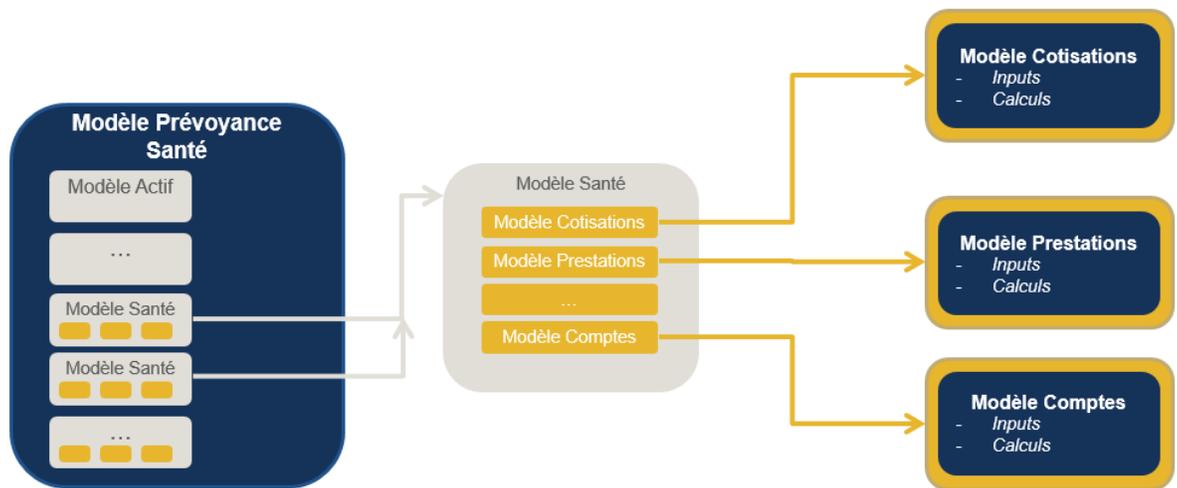


Figure 2. Schéma de l'architecture d'un modèle type sur addactis® Modeling

1.2.2. Résumé de l'architecture

Nous passerons rapidement en revue les modules du modèle Prévoyance Santé. Le modèle comporte trois parties principales :

- une partie « **actif** » qui fait évoluer les indicateurs économiques, projette le portefeuille d’actifs, calcule les chocs de la formule standard et réalise l’*ALM* ;
- une partie « **passif** » organisée en « segment ». Les « segments » sont des modules effectuant la modélisation d’une garantie précise d’assurance. Ils disposent tous d’une même architecture modulaire. Les « segments » disponibles sont les suivants : santé, décès, arrêt de travail, rente de conjoint, rente d’éducation et flux ;
- une partie « **solvabilité 2** » calculant les exigences en capital (*MCR*, *SCR*) et la marge pour risque puis valorisant les comptes en normes prudentielle.

1.2.2.1. Passif

La partie concernant le passif est composée de plusieurs « segments » qui correspondent aux garanties évoquées précédemment. Ces modèles disposent tous de la même architecture.

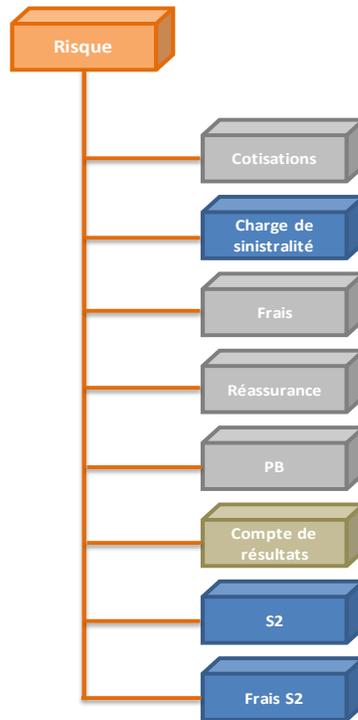


Figure 3. Architecture commune des modèles de la partie Passif

Le module principal associé à la garantie est représenté en orange. Les sous modules communs sont représentés en gris alors que les autres spécifiques à la garantie sont représentés en bleus. Enfin le sous module comptes de résultats est différent selon que la garantie soit classée en vie ou en non-vie.

Ces différents modules ont les fonctions suivantes :

- **Cotisations** : projeter l'évolution des cotisations et des provisions associées (provisions pour primes non acquises PPNA¹ et provisions pour primes acquises non émises PANE²) ;

¹ PPNA : Provision pour primes non acquises, concerne les garanties portant sur plusieurs exercices comptables. Elle correspond à un *prorata temporis* de la période de garantie dépassant l'exercice comptable.

² PANE : Primes acquises non émises, concerne les primes afférentes à l'exercice mais n'ayant pu être émises. C'est le cas notamment des garanties dont la prime porte sur la masse salariale, l'assureur ne dispose de l'information que l'année suivante.

- **Charge de Sinistralité** : projeter l'évolution des prestations et des provisions associées (provisions mathématiques PM, provisions pour sinistres à payer PSAP³, IBNR⁴) ;
- **Frais** : projeter l'évolution des frais d'acquisition, d'administration et de gestion ;
- **Réassurance** : projeter l'évolution des montants cédés (provisions, prestations et cotisations) dans le cas de la réassurance en quote-part ;
- **Participations aux bénéfices (PB)** : projeter l'évolution des provisions d'égalisation (PEg⁵) et des provisions de participation aux excédents (PPE⁶) ;
- **Compte de résultat** : projeter l'évolution des comptes techniques Vie et Non Vie, et non technique ;
- **Solvabilité 2 (S2)** : calculer les provisions en vision *Best Estimate* (brutes, cédées et nettes) et les différents chocs associés ;
- **Frais S2** : calculer les frais associés aux provisions *Best Estimates* ainsi que les chocs associés.

1.2.2.2. Actif

Cette partie comporte quatre modules :

- **ESG** (*Economic scenario generator*) : ce module projette plusieurs indicateurs économiques comme les indices d'évolution des actions ainsi que la courbe des taux. Le sous modèle Actif permet la projection du portefeuille d'actifs de l'organisme assureur modélisé ;
- **Actif** : projection du portefeuille d'actifs en ligne à ligne à l'aide des éléments de l'*ESG* ;
- **Actif S2** : calcul des chocs de la formule standard ;

³ PSAP : estimation des dépenses pour les sinistres survenus et non payés.

⁴ IBNR (*Incurring but not reported*) : provisions pour sinistres survenus non déclarés.

⁵ PEg : provision pour faire face aux fluctuations de sinistralité et permettant de mutualiser les risques sur plusieurs années. Elle est également utilisée pour lisser le résultat technique.

⁶ PPE : provision dotée par le montant de participation aux bénéfices non distribuée par l'assureur.

- **ALM** (*Asset and liability management*) : le module récupère les flux de trésorerie générés par l'actif et le passif afin d'acheter et de vendre des titres selon une allocation déterminée.

1.2.2.3. Solvabilité 2

Cette partie effectue l'ensemble des calculs de la formule standard du Pilier 1 de Solvabilité 2 à partir des montants modélisés dans les deux autres parties du modèle. Elle est composée des modules *MCR*, *SCR*, *Risk Margin*, *Comptes S2* et *Risk margin*.

1.3. Contrôles des modèles actuariels

Nous avons vu précédemment comment le modèle Prévoyance Santé s'organisait pour répondre aux exigences imposées par Solvabilité 2. Ces contraintes réglementaires évoluent régulièrement et de nouvelles fonctionnalités sont ajoutées au modèle. Pour assurer leur conformité, un processus détaillé est appliqué pour contrôler les développements sur le modèle. Nous en donnons les principales étapes.

1. Tout d'abord un développement est signalé pour un module sur une plateforme de gestion de projet.
2. Il est ensuite confirmé et précisé par un expert du module concerné.
3. Puis le développement est réalisé et divers tests sont effectués pour vérifier que l'implémentation de la fonctionnalité a bien été effectuée. Nous distinguons les tests suivants par ordre de déroulé :
 - **Test unitaire** : vérifier le bon fonctionnement d'un module de manière indépendante.

Exemple : Pour un développement sur un module de calcul de *Best Estimates*, nous pouvons vérifier que la somme actualisée des flux projetés probabilisés correspond bien à la sortie du module.

- **Test d'intégration** : vérifier le bon fonctionnement du module lorsqu'il communique avec d'autres.

Exemple : Nous pouvons vérifier que les montants des *Best Estimates* choqués en sortie du module de calcul sont bien ceux présent dans les assiettes choquées du module de calcul du *SCR*.

- **Test de non régression** : vérifier que l'implémentation d'une nouvelle fonctionnalité n'entrave pas le fonctionnement d'une autre.

Exemple : Les développements que nous avons fait sur le module de calculs de provisions en vision *Best Estimate* ne doivent pas changer les résultats des modules de calculs de provisions en vision Solvabilité 1.

- **Test de validation** : vérifier que les résultats du modèle évoluent de manière cohérente avec la nouvelle fonctionnalité apportée.

Exemple : Pour un développement consistant à retrancher les PANE du *Best Estimate* de prime, les *SCR* de souscription ne sont pas censés être modifiés. En effet les PANE ne sont pas choqués et sont donc annulés par différence entre l'assiette de passif choquée et l'assiette de passif central.

4. Si le modèle passe l'ensemble des tests de contrôles avec la nouvelle fonctionnalité, il devient la dernière version en cours.
5. Enfin la fonctionnalité est contrôlée et validée ultérieurement par une personne experte du module concerné.

Synthèse

Nous avons vu comment le contexte réglementaire a complexifié la modélisation actuarielle avec l'ajout de calculs plus sophistiqués que la norme précédente. Pour ce faire les organismes d'assurance se sont dotés de modèles avec une architecture composée des différents modules nécessaires : actif, passif et solvabilité 2. Cependant les évolutions constantes de la réglementation nécessitent de retravailler les modules de cette architecture régulièrement. Afin d'assurer la conformité à chaque étape des développements il est nécessaire de mettre en place un processus précis de contrôle. Le processus actuellement mis en place nous permet de valider indépendamment chaque nouvelle fonctionnalité implémentée. Nous souhaiterions également disposer d'un outil permettant de valider l'architecture des modèles actuariels. Le développement de cet outil sera exposé dans la partie suivante.

2. Cartographie informatique des modèles actuariels

Le modèle Prévoyance et Santé prend en entrée un nombre important d'hypothèses et réalise de multiples calculs. Par exemple pour une projection sur 5 ans de l'activité d'un organisme assureur, le modèle réalise plus de 33 000 000 de calculs et prend en compte près de 200 000 dépendances entre des variables. Cette quantité impressionnante d'informations reflète la complexité de la modélisation sous Solvabilité 2. Cependant il est essentiel de connaître les interactions entre ses différentes variables afin de maîtriser pleinement le modèle et les résultats qu'il fournit. Les principaux obstacles rencontrés par les équipes de modélisation actuarielle sont les suivants :

- Les conséquences de la modification d'une variable s'avèrent compliquées à anticiper étant donné les nombreuses dépendances entre elles.

Exemple : l'impact des frais de gestion ou du taux de quote-part d'un traité de réassurance sur le SCR taux.

- Il est difficile de décrire un modèle actuariel de manière synthétique. Or il est essentiel de connaître l'ordre et l'organisation de ses calculs afin d'intervenir efficacement sur celui-ci.

Exemple : les étapes de calculs d'un *Best Estimate*.

- Après l'implémentation, certaines dépendances entre des parties différentes du modèle ne sont pas cohérentes au sens de la modélisation actuarielle.

Exemple : les variables de la partie concernant l'actif ne doivent pas dépendre directement de celles de la partie portant sur le passif, mais plutôt de la partie ALM.

Ces différents points mettent en évidence le besoin de visualiser simplement les interactions entre les variables du modèle Prévoyance Santé. Il est donc logique de se tourner

vers les graphes qui sont utilisés dans de nombreux autres domaines pour faciliter la compréhension des systèmes.

Suite à la mise en évidence du besoin de contrôler et comprendre les modèles, nous allons aborder dans cette partie une première méthode basée sur la théorie des graphes. Nous verrons tout d'abord comment elle peut être mise en pratique à partir des données à disposition. Puis nous détaillerons sa mise en place pour le modèle Prévoyance Santé. Enfin nous expliquerons différents cas pratiques de modélisation actuarielle où les cartographies peuvent être utilisées.

2.1. Applications de la théorie des graphes

Nous considérons dans cette partie le modèle Comptes pour illustrer les concepts abordés.

Description du modèle Comptes :

Le modèle construit et projette sur plusieurs années le résultat d'un produit santé. Tout d'abord, il prend en entrée un nombre de cotisants et une cotisation moyenne afin de déterminer un montant de cotisations versées. Puis des frais sont calculés sur ce montant à partir des taux entrés. Ces frais sont ensuite retranchés au montant de cotisations pour déterminer la prime pure. Cette prime pure multipliée par le ratio sinistre à primes (S/P) permet d'obtenir une charge de prestations. Cette charge est ensuite liquidée sur plusieurs années à l'aide d'une chronique : les montants des PSAP et des prestations payées sont ainsi déduits. Enfin le résultat est calculé en retranchant des cotisations versées la variation des provisions, les commissions et les charges.

2.1.1. Présentation des données à disposition

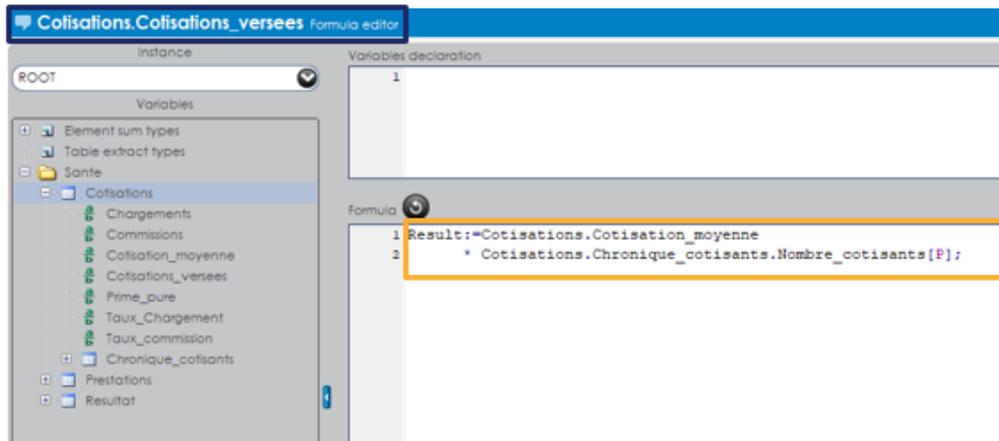
Les données utilisées pour effectuer la cartographie sont issues d'addactis® Modeling. En effet le logiciel dispose d'une fonctionnalité de débogage permettant de générer un fichier texte contenant l'ensemble des dépendances informatiques à partir des formules des variables. Les données que nous utilisons sont produites à la compilation du logiciel. Il n'y a par conséquent aucun calcul effectué à partir du modèle.

Le fichier généré pour le modèle Comptes est le suivant :

| OBJECT | DEP | | | |
|---|--|--|--|--|
| MODEL.Cotisations.Cotisations.Cotisations_versees | MODEL.Cotisations.Cotisations.Chronique_cotisants.Nombre_cotisants | | | |
| MODEL.Cotisations.Cotisations.Cotisations_versees | MODEL.Cotisations.Cotisations.Cotisation_moyenne | | | |
| MODEL.Cotisations.Cotisations.Commissions | MODEL.Cotisations.Cotisations.Cotisations_versees | | | |
| MODEL.Cotisations.Cotisations.Commissions | MODEL.Cotisations.Cotisations.Taux_commission | | | |
| MODEL.Cotisations.Cotisations.Chargements | MODEL.Cotisations.Cotisations.Cotisations_versees | | | |
| MODEL.Cotisations.Cotisations.Chargements | MODEL.Cotisations.Cotisations.Taux_Chargement | | | |
| MODEL.Cotisations.Cotisations.Prime_pure | MODEL.Cotisations.Cotisations.Chargements | | | |
| MODEL.Cotisations.Cotisations.Prime_pure | MODEL.Cotisations.Cotisations.Commissions | | | |
| MODEL.Cotisations.Cotisations.Prime_pure | MODEL.Cotisations.Cotisations.Cotisations_versees | | | |
| MODEL.Prestations.Prestations.Charge_de_prestation | MODEL.Cotisations.Cotisations.Prime_pure | | | |
| MODEL.Prestations.Prestations.Charge_de_prestation | MODEL.Prestations.Prestations.S_P | | | |
| MODEL.Prestations.Prestations.Prestations_payees | MODEL.Prestations.Prestations.Charge_de_prestation | | | |
| MODEL.Prestations.Prestations.Prestations_payees | MODEL.Prestations.Prestations.Chronique_des_PSAP.Cadence_liquidation | | | |
| MODEL.Prestations.Prestations.Provisions_de_sinistres | MODEL.Prestations.Prestations.Charge_de_prestation | | | |
| MODEL.Prestations.Prestations.Provisions_de_sinistres | MODEL.Prestations.Prestations.Chronique_des_PSAP.Cadence_liquidation | | | |
| MODEL.Resultats.Resultat.Resultat | MODEL.Cotisations.Cotisations.Chargements | | | |
| MODEL.Resultats.Resultat.Resultat | MODEL.Cotisations.Cotisations.Commissions | | | |
| MODEL.Resultats.Resultat.Resultat | MODEL.Cotisations.Cotisations.Cotisations_versees | | | |
| MODEL.Resultats.Resultat.Resultat | MODEL.Prestations.Prestations.Prestations_payees | | | |
| MODEL.Resultats.Resultat.Resultat | MODEL.Prestations.Prestations.Provisions_de_sinistres | | | |

Figure 4. Fichier des dépendances du modèle Comptes

Chaque ligne du fichier correspond à une **dépendance informatique** entre deux variables. La notion de dépendance informatique s'interprète ici de la façon suivante : **les variables de la colonne DEP interviennent dans la formule permettant de calculer des variables de la colonne OBJECT**. Il est important de la dissocier de la dépendance probabiliste qu'il est possible de mesurer par la corrélation. Par la suite nous la nommerons **dépendance informatique**. Par exemple, pour le modèle Comptes les variables Nombre_cotisants et Cotisation_moyenne servent à déterminer Cotisations_versees comme nous pouvons le voir sur le schéma suivant:



| OBJECT | DEP |
|---|--|
| MODEL.Cotisations.Cotisations.Cotisations_versees | MODEL.Cotisations.Cotisations.Chronique_cotisants.Nombre_cotisants |
| MODEL.Cotisations.Cotisations.Cotisations_versees | MODEL.Cotisations.Cotisations.Cotisation_moyenne |
| MODEL.Cotisations.Cotisations.Commissions | MODEL.Cotisations.Cotisations.Cotisations_versees |
| MODEL.Cotisations.Cotisations.Commissions | MODEL.Cotisations.Cotisations.Taux_commission |
| MODEL.Cotisations.Cotisations.Chargements | MODEL.Cotisations.Cotisations.Cotisations_versees |
| MODEL.Cotisations.Cotisations.Chargements | MODEL.Cotisations.Cotisations.Taux_Chargement |
| MODEL.Cotisations.Cotisations.Prime_pure | MODEL.Cotisations.Cotisations.Chargements |
| MODEL.Cotisations.Cotisations.Prime_pure | MODEL.Cotisations.Cotisations.Commissions |
| MODEL.Cotisations.Cotisations.Prime_pure | MODEL.Cotisations.Cotisations.Cotisations_versees |
| MODEL.Prestations.Prestations.Charge_de_prestation | MODEL.Cotisations.Cotisations.Prime_pure |
| MODEL.Prestations.Prestations.Charge_de_prestation | MODEL.Prestations.Prestations.5_P |
| MODEL.Prestations.Prestations.Prestations_payees | MODEL.Prestations.Prestations.Charge_de_prestation |
| MODEL.Prestations.Prestations.Prestations_payees | MODEL.Prestations.Prestations.Chronique_des_PSAP.Cadence_liquidation |
| MODEL.Prestations.Prestations.Provisions_de_sinistres | MODEL.Prestations.Prestations.Charge_de_prestation |
| MODEL.Prestations.Prestations.Provisions_de_sinistres | MODEL.Prestations.Prestations.Chronique_des_PSAP.Cadence_liquidation |
| MODEL.Resultats.Resultat.Resultat | MODEL.Cotisations.Cotisations.Chargements |
| MODEL.Resultats.Resultat.Resultat | MODEL.Cotisations.Cotisations.Commissions |
| MODEL.Resultats.Resultat.Resultat | MODEL.Cotisations.Cotisations.Cotisations_versees |
| MODEL.Resultats.Resultat.Resultat | MODEL.Prestations.Prestations.Prestations_payees |
| MODEL.Resultats.Resultat.Resultat | MODEL.Prestations.Prestations.Provisions_de_sinistres |

Figure 5. Schéma de la transcription des formules

Par ailleurs les noms des variables de ce fichier permettent de savoir à quelle partie du modèle elles appartiennent. Nous pouvons voir que le modèle Comptes est composé de trois sous parties : Cotisations, Prestations et Resultats. Ainsi, le nom de la variable *Cotisations.Cotisations_versees* nous permet de savoir que la variable appartient à la partie Cotisations du modèle. Cette organisation est visible depuis l'onglet architecture du logiciel :



Figure 6. Onglet Architecture du modèle Comptes

Afin d'exploiter efficacement les données du fichier des dépendances informatiques, un traitement des données est effectuée. Cette étape consiste à simplifier les noms des variables, puis à isoler la partie du modèle à laquelle elles appartiennent et enfin à formater le fichier pour l'import dans le logiciel d'analyse de graphe.

Le logiciel choisi pour notre étude est Gephi. Il permet notamment d'effectuer des visualisations de graphes interactives et propose différents outils d'analyse (degrés, ancêtres, descendants etc...). Il a été retenu pour sa prise en main facile, la diversité de ses algorithmes de visualisation ainsi que pour sa capacité à manipuler des graphes avec un nombre de nœuds et de liens importants.

2.1.2. Principes de la théorie des graphes

Nous présentons dans cette sous-partie les concepts de la théorie des graphes appliqués à la cartographie des modèles actuariels. Les graphes sont utilisés dans de nombreux domaines pour décrire les interactions entre des objets en fournissant une représentation intuitive. Par conséquent les utiliser pour traduire la complexité des dépendances informatiques entre les différentes variables du modèle est justifié. Nous présentons tout d'abord les principales définitions de la théorie des graphes. Les définitions sont tirées de la section *Basic Graph Concepts* du livre *Bayesian Reasoning and Machine Learning* (Barber, 2012). Puis nous nous intéresserons aux manières de représenter les graphes dans le plan.

2.1.2.1. Définitions et applications de la théorie des graphes

Définition 2.1 (Graphe) Un graphe G est un couple composé d'un ensemble de **nœuds** V et d'un ensemble de **liens** E entre les nœuds. Les liens peuvent être **orientés** (i.e. ils comportent une flèche vers une direction) ou non. Ils peuvent aussi comportés des **poids**. Un graphe dont tous les liens sont orientés est un **graphe orienté**.

Application 2.1 : Les **nœuds** sont les **variables** et les **liens** sont les **dépendances informatiques** entre les variables. Notre graphe est un **graphe orienté** et le sens des liens s'interprète de la façon suivante :

Le lien de A vers B signifie que la variable A intervient dans la formule de la variable B

Exemple 2.1 : Nous reprenons l'exemple du modèle Comptes. La représentation du graphe est obtenu à partir de la méthode de Sugiyama décrite dans la section suivante. Les variables du modèle appartenant à la même catégorie sont représentées avec une couleur identique. La direction des liens est indiquée par une flèche. Les différents niveaux s'interprètent comme l'ordre des calculs effectués par le modèle.

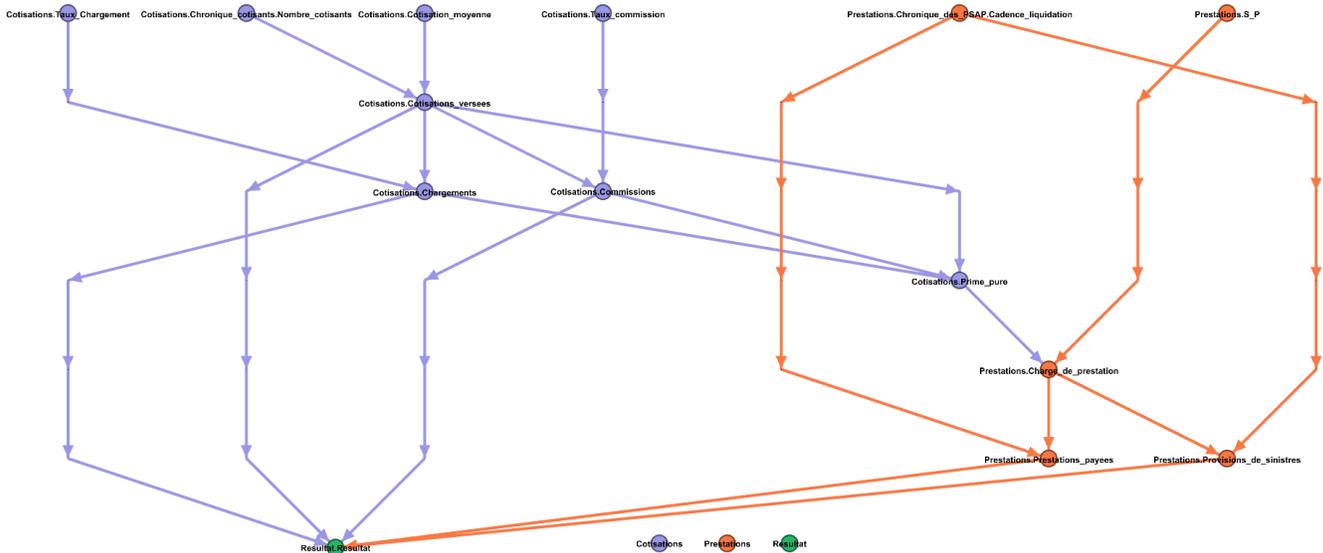


Figure 7. Exemple de visualisation du graphe du modèle Comptes

Définition 2.2 (Chemins) Un **chemin** d'un nœud A à un nœud B est une suite de nœuds reliant A à B. Un chemin est donc de la forme $A_0, A_1, \dots, A_{n-1}, A_n$ avec $A_0 = A$ et $A_n = B$, où chaque lien (A_{k-1}, A_k) , pour $k = 1, \dots, n$, fait partie du graphe.

Un **chemin orienté** d'un nœud A à un nœud B est une suite de nœuds reliant A à B, obtenu en suivant la direction des liens.

A est **connecté** à B si il existe un chemin reliant A à B. On notera $A \rightarrow B$.

Application 2.2 : Nous considérons des **chemins orientés**. Si la variable A est connecté à la variable B alors cela signifie que la variable A intervient dans la formule de la variable B par le biais des variables composant le chemin. Nous pouvons donc savoir si la variable A peut avoir un impact sur la variable B et par l'intermédiaire de quels autres variables.

Exemple 2.2 : Nous voulons savoir si la moyenne des cotisations a un impact sur les prestations payées. Cela revient à savoir si la variable *Cotisations.Cotisation_moyenne* est connectée à la variable *Prestations.Prestations_payees*. Nous effectuons la requête du plus court chemin entre les deux variables. Les nœuds des variables intervenant dans le chemin sont représentés en rouge. Nous obtenons ainsi dans l'ordre le chemin : *Cotisations.Cotisations_versees*, *Cotisations.Prime_pure* et *Prestations.Charge_de_prestation*. Ce chemin est bien conforme avec la description du modèle effectué précédemment. Ainsi les cotisations versées sont obtenues en multipliant la cotisation moyenne par le nombre de cotisants. La prime pure est ensuite obtenue en retranchant les commissions et les charges. Puis cette prime est multipliée par le ratio S/P pour déterminer la charge de prestation. Enfin une partie de cette charge de prestation est payée la première année, ce qui constitue les prestations payées.

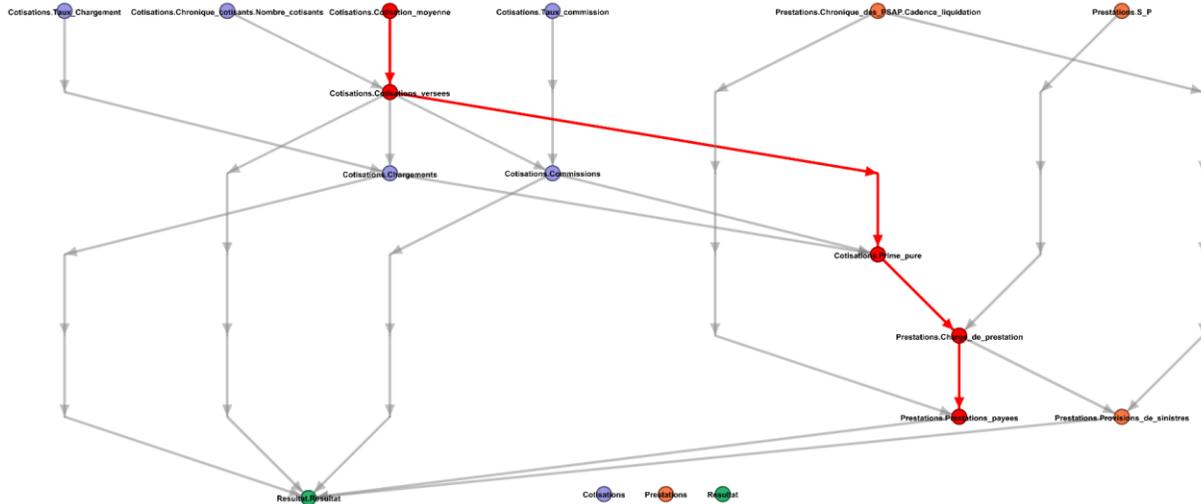


Figure 8. Visualisation du plus court chemin des cotisations moyennes aux prestations payées

Définition 2.3 (Ancêtres et descendants) : Les **ancêtres** d'un nœud B sont les nœuds A tels que $A \rightarrow B$ et $\text{non}(B \rightarrow A)$.

Les **descendants** d'un nœud B sont les nœuds A tels que $B \rightarrow A$ et $\text{non}(A \rightarrow B)$.

Application 2.3 : Les ancêtres d'une variable sont toutes les variables intervenant dans sa formule. Cette notion permet notamment de retracer les différentes étapes d'un calcul.

Les descendants d'une variable sont toutes les variables l'utilisant dans leur formule. Cela permet donc de connaître les possibles conséquences de la modification d'une variable.

Exemple 2.3 : Nous voulons savoir quelles variables peuvent avoir un impact sur les Chargements. Pour cela nous effectuons une requête pour déterminer ses ancêtres et nous affectons une couleur au nœud en fonction du résultat. Nous pouvons voir que les antécédents sont le taux de chargement, le nombre de cotisants, la cotisation moyenne et les cotisations versées.

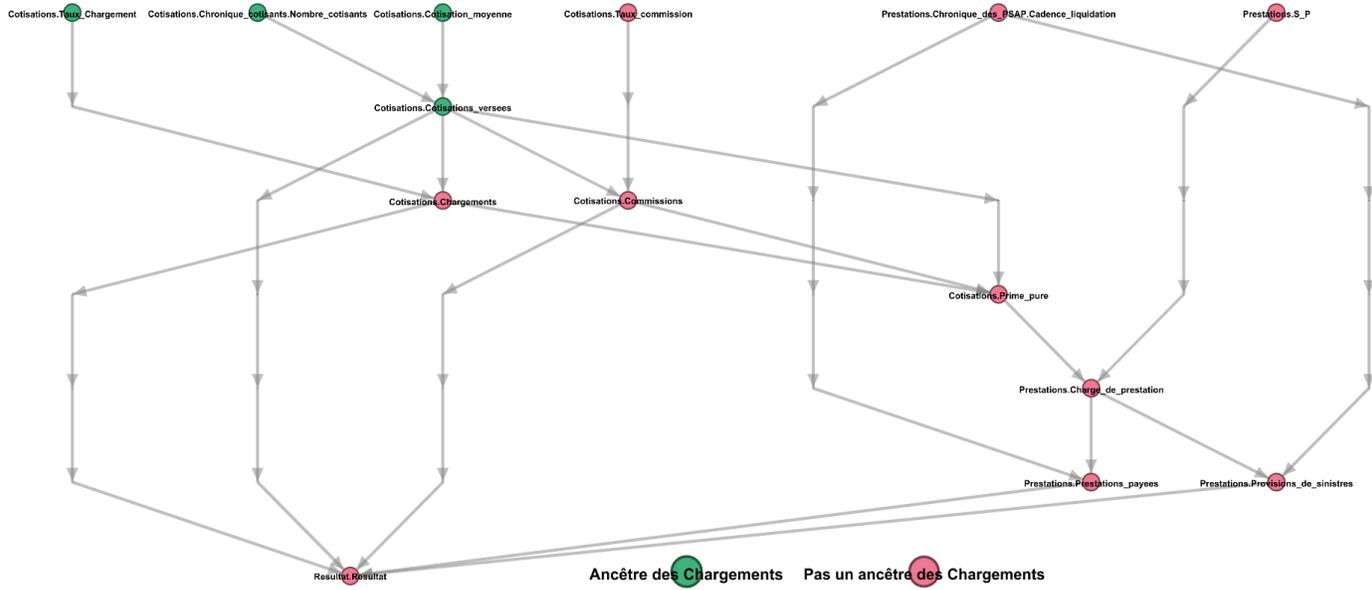


Figure 9. Visualisation des antécédents des chargements dans le modèle Comptes

Définition 2.4 (Cycle) Un **cycle** est un chemin orienté qui commence par un nœud et se termine par le même nœud : $A \rightarrow B \rightarrow \dots \rightarrow Z \rightarrow A$.

Un **graphe acyclique** ne contient pas de cycle.

Un **graphe acyclique orienté** est noté **DAG** (*Directed Acyclic Graph*).

Application 2.4 : Nous avons vu précédemment que les graphes considérés étaient orientés. Ils ne comportent pas de cycle car cela correspondrait à une référence circulaire. En effet cela signifierait qu'une variable s'appelle elle-même dans sa propre formule. Ce genre d'anomalie est détectée à la compilation du code du logiciel.

Exemple 2.4 : Nous pouvons voir que le graphe du modèle Comptes ne comporte pas de cycle. Cependant si nous ajoutons un lien de la prime pure à la cotisation moyenne, alors il y a formation d'un cycle représenté en rouge.

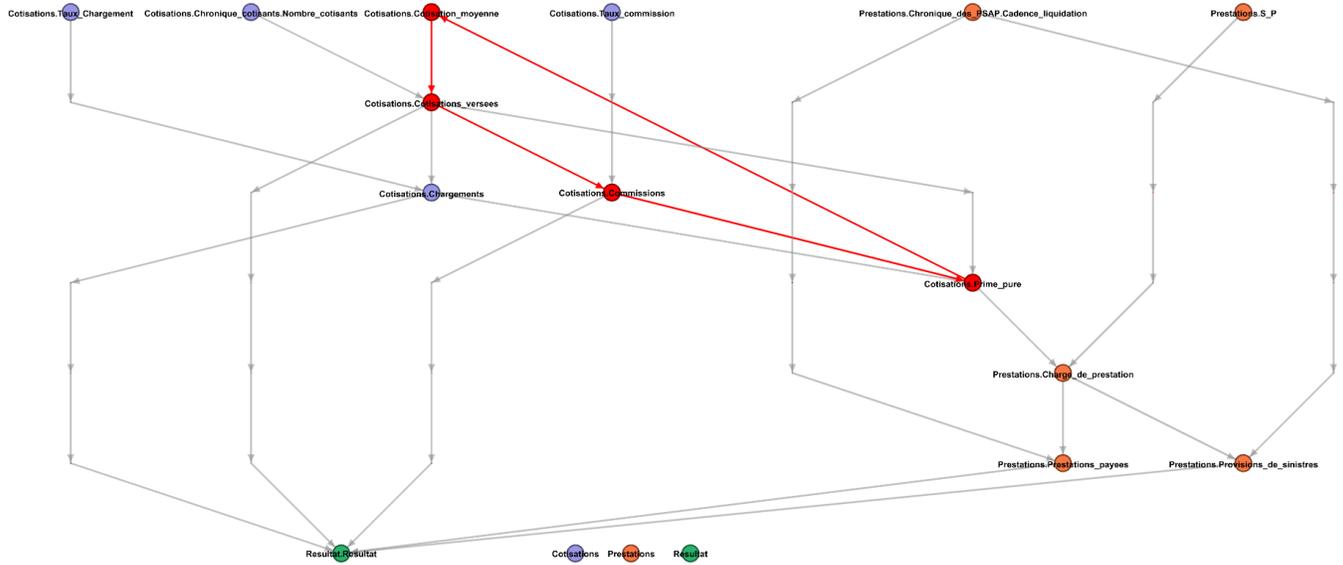


Figure 10. Visualisation d'un cycle dans le modèle Comptes

Définition 2.5 (Hiérarchie⁷) Une hiérarchie de niveau n est un graphe orienté (V, E) vérifiant les deux conditions suivantes :

- 1) L'ensemble des nœuds V est partitionné en n sous-ensembles $V_1, \dots, V_i, \dots, V_n$ où V_i représente le i ème niveau de la hiérarchie de longueur n .
- 2) Chaque lien part d'un niveau inférieur pour aller vers un niveau supérieur.

Application 2.5 : Les graphes des modèles actuariels peuvent être représentés comme des hiérarchies. Les niveaux correspondent à l'ordre des calculs effectués. Ainsi les variables des niveaux inférieurs interviennent dans les calculs de celles situés aux niveaux supérieurs.

Exemple 2.5 : Nous pouvons interpréter les niveaux obtenus à l'aide de la méthode de Sugiyama. Au premier niveau se trouvent les hypothèses des calculs comme le ratio S/P où le taux de commission. Aux niveaux intermédiaires, nous retrouvons les différentes étapes de calcul comme les cotisations versées ou les prestations payées. Enfin le dernier niveau est réservé à la variable finale de notre modèle : le résultat technique.

⁷ La définition est tirée de l'article (Sugiyama, et al., 1981)

modèle Comptes qui n'utilise pas les autres variables, alors le graphe n'est plus simplement connexe. Il peut se décomposer en deux composantes connexes : celle associée au modèle précédent (à gauche sur la figure) et celle associée au calcul des frais (à droite sur la figure).

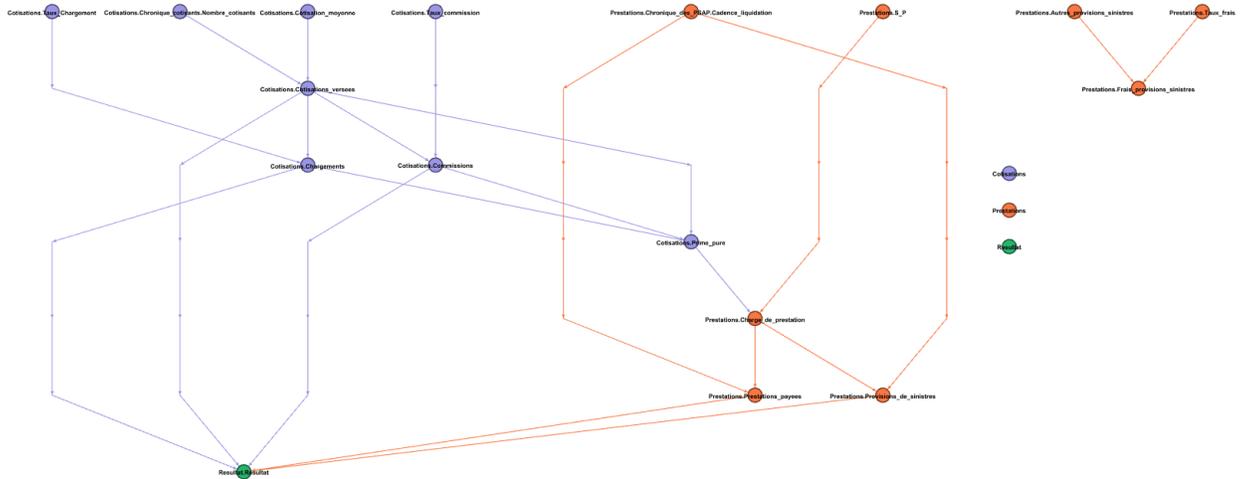


Figure 11. Visualisation du modèle Comptes après l'ajout du calcul des frais

Définition 2.7 (Degrés) Le **degré** d'un nœud est le nombre de liens incident à ce nœud.

Le **degré sortant** d'un nœud est son nombre de liens sortant incidents.

Le **degré entrant** d'un nœud est son nombre de liens entrant incidents.

Application 2.7 : La notion de degré sortant pour une de nos variables correspond au nombre de fois où elle est appelée, alors que le degré entrant est son nombre d'hypothèses nécessaires à son calcul.

Exemple 2.7 : Sur la visualisation, les nœuds sont d'autant plus grands que leur degré est élevé. Cette indicateur permet notamment de repérer les variables jouant un rôle central dans la modélisation de notre modèle, ce sont ici les cotisations versées et le résultat.

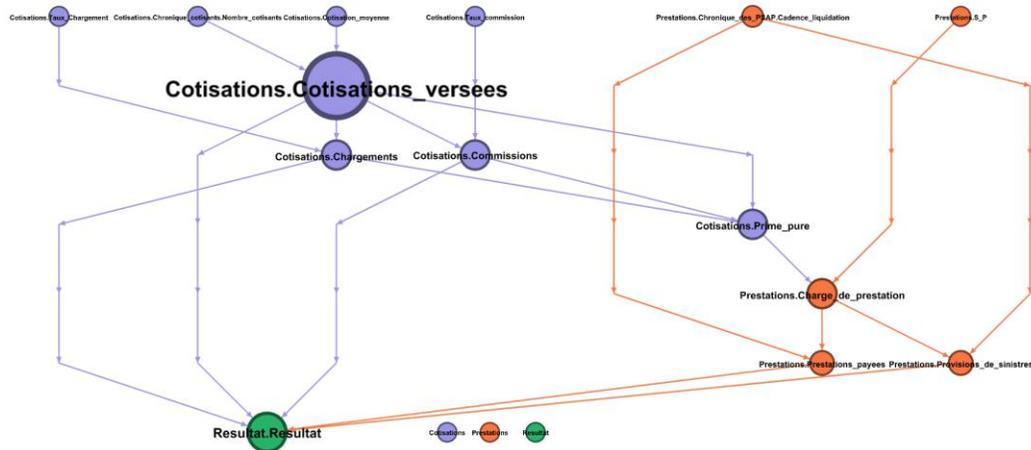


Figure 12. Visualisation à l'aide des degrés des nœuds

2.1.2.2. Tracés des graphes

Nous avons vu précédemment comment les concepts de la théorie des graphes pouvaient nous aider à analyser les dépendances informatiques entre les variables de nos modèles. Nous allons maintenant nous intéresser à la manière de les représenter.

La visualisation de graphes est un domaine faisant appel aux mathématiques et à l'algorithmique. L'objectif de cette discipline est de positionner de manière esthétique les nœuds et les liens d'un graphe afin d'en résumer sa structure. Étant donné le caractère subjectif de l'esthétique et la diversité des structures de graphe possibles, il existe une très grande variété d'algorithmes permettant d'atteindre ce but. Nous présentons deux familles d'algorithmes qui nous seront utiles pour effectuer la cartographie des modèles actuariels.

2.1.2.2.1. Algorithmes de placement par des forces

Ces algorithmes se basent sur des analogies avec la physique et notamment la mécanique classique. Ils essaient de placer les nœuds selon les deux critères suivants :

- a) Deux nœuds connectés par un lien doivent être proches l'un de l'autre ;
- b) Les nœuds ne doivent pas être trop rapprochés les uns des autres.

En considérant que les nœuds sont des particules de même charge et que les liens sont des ressorts, ces critères peuvent être matérialisés par deux forces :

- a) **Force attractive** (c.-à-d. la force de rappel) : les ressorts rapprochent les particules ;
- b) **Force répulsive** (c.-à-d. la force de Coulomb) : toutes les particules se repoussent.

À chaque itération, ces algorithmes calculent les forces s'appliquant sur chaque nœud, puis les déplacent en conséquence en se calquant sur le système physique.

L'un des premiers algorithmes utilisant le principe des forces attractives et répulsives est celui de Eades (Eades, 1984). Un autre très utilisé est celui de Fruchterman Reingold (Fruchterman, et al., 1991) qui ajoute la notion de « température ». Celle-ci permet de limiter les déplacements de chaque nœud à chaque itération. La « température » commence à une valeur initiale et décroît à chaque itération de l'algorithme de sorte que les ajustements apportés à la modélisation sont de plus en plus fins.

Ces algorithmes s'appliquent aux graphes non-orientés. Étant donné leur complexité élevée⁸, ils ne s'appliquent qu'aux graphes de taille modestes (environ 1 000 nœuds). Pour pallier à ce problème⁹, des versions plus élaborées, comme OpenOrd (Martin, et al., 2011), utilisent des approximations pour représenter des graphes de très grandes tailles (jusqu'à 600 000 nœuds). Ces algorithmes sont par ailleurs très flexibles et peuvent représenter les symétries de la plupart des structures de graphe. Nous reprenons le graphe du modèle Comptes et nous lui appliquons l'algorithme de Fruchterman Reingold implémenté dans le logiciel Gephi. Par rapport à la version de l'algorithme décrite précédemment, celle-ci ajoute la notion « gravité » qui agit comme une force attractive dirigé vers le centre sur l'ensemble des nœuds.

⁸ Pour l'algorithme de Fruchterman Reingold : la complexité est de l'ordre de $O(|V|^2+|E|)$, où $|V|$ représente le nombre de nœuds et $|E|$ le nombre de liens.

⁹ Pour l'algorithme OpenOrd : la complexité est de l'ordre de $O(|V| \log |V|)$.

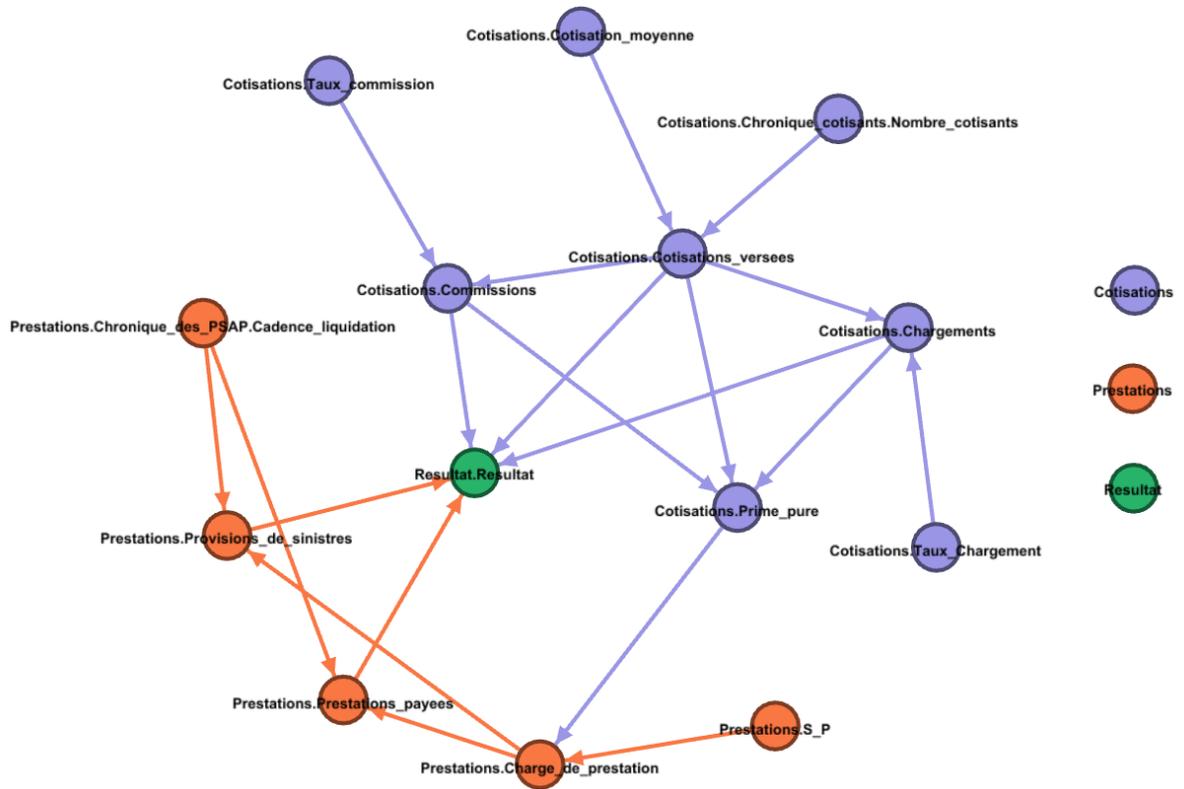


Figure 13. Visualisation du modèle Comptes à l'aide de l'algorithme de Fruchterman Reingold

Nous pouvons voir que les hypothèses de notre modèle sont situées en périphérie sur la représentation. Étant donné qu'elles n'ont qu'un lien sortant vers une ou deux variables, elles sont soumises à une force d'attraction plus faible que les autres. Par ailleurs, la variable Résultat en sortie du modèle se retrouve au centre de la visualisation car elle est reliée à de nombreuses variables. Enfin, nous ne retrouvons pas l'ordre des calculs effectués comme dans les visualisations effectuées dans la section 2.1.2.1.

2.1.2.2.2. Algorithmes hiérarchiques

Cette famille d'algorithmes permet de représenter la hiérarchie des graphes orientés. La méthode la plus utilisée est celle de Sugiyama (Sugiyama, et al., 1981). Elle considère les critères esthétiques suivants :

- Les liens doivent pointer dans la même direction ;

- Les liens doivent être des lignes droites ;
- Les liens courts sont plus lisibles ;
- Les croisements de liens rendent la compréhension difficile.

Pour les réaliser le tracé, la méthode effectue les étapes suivantes :

- a) Déterminer le niveau auquel appartient chaque nœud. Placer tous les nœuds appartenant au même niveau à la même ordonnée.
- b) Pour chaque niveau déterminer l'ordre des nœuds qui minimise les croisements de liens.
- c) Pour chaque niveau déterminer les abscisses des nœuds qui permettent d'obtenir les liens les plus courts possibles, sans utiliser de ligne brisée.

Cette algorithmme est appliqué à l'aide du package *igraph* (Csardi, 2015) du logiciel R. Un script permet d'importer les données, effectuer la visualisation et d'exporter le graphe au format du logiciel Gephi.

La méthode de Sugiyama s'applique uniquement à des *DAG*. Les algorithmes employés pour les trois étapes de la méthode sont de complexité élevée¹⁰, ce qui rend la méthode applicable uniquement aux graphes de taille modeste (environ 1 000 nœuds).

Synthèse de la partie 2.1

Nous avons vu dans cette partie les données de dépendances informatiques mise à disposition par addactis® Modeling. Ce fichier est ensuite traité afin d'être exploité par le logiciel de visualisation de graphes Gephi. Les différents principes de la théorie des graphes (chemins, connexité, ancêtres...) nous permettent d'analyser l'organisation des calculs effectués dans les modèles actuariels.

¹⁰ Pour l'algorithme de Sugiyama : la complexité est de l'ordre de $O(|V| |E| \log |E|)$

| Concepts de la théorie des graphes | Applications à l'analyse des modèles actuariels |
|---|--|
| Nœuds | Variables du modèle |
| Liens | Relations de dépendance informatique |
| Chemin entre deux nœuds | Étapes de calculs entre deux variables |
| Ancêtres d'un nœud A | Variables pouvant avoir un impact sur A |
| Descendants d'un nœud A | Variables pouvant être impactées par A |
| Cycle | Référence circulaire dans les calculs |
| Hiérarchie | Ordre des calculs effectués |
| Graphe connexe | Absence de calculs isolés |
| Degrés | Mesure de l'importance d'une variable |

Tableau I. Tableau récapitulatif des concepts de la théorie des graphes

Enfin la visualisation des graphes est effectuée à l'aide de deux familles d'algorithmes : placement par des forces et hiérarchique.

| Caractéristiques | Placement par des forces | Hierarchique |
|--------------------------------|--|--|
| <u>Type de graphe</u> | Non-orienté | <i>DAG</i> |
| <u>Fonctionnement</u> | Simulation de forces physiques | Détermination des niveaux et minimisation des croisements de liens |
| <u>Objectif</u> | Représenter les symétries et effectuer des regroupements | Représenter la hiérarchie d'un <i>DAG</i> |
| <u>Complexité¹¹</u> | Élevée – (Fruchterman, et al., 1991) Moyenne – (Martin, et al., 2011) | Très élevée – (Sugiyama, et al., 1981) |

Tableau II. Tableau comparatif des deux familles algorithmes de visualisation

¹¹ Complexité : quantité de ressources nécessaire à l'exécution d'un algorithme. $|V|$ désigne le nombre de nœuds et $|E|$ désigne le nombre de liens.

2.2. Réalisation des cartes

Nous avons vu dans la sous-partie précédente l'ensemble des éléments nécessaires à l'exploitation, l'analyse et la visualisation des données de dépendances informatiques de nos modèles actuariels. Nous allons maintenant expliquer les méthodologies pour réaliser une carte générale du modèle Prévoyance Santé et une carte d'un module composant le modèle.

2.2.1. Méthode pour visualiser les interactions

2.2.1.1. Première approche

Le fichier de données du modèle Prévoyance Santé comporte 200 000 liens de dépendances associés à 30 000 variables. Nous avons déjà une idée de l'organisation générale du modèle Prévoyance Santé. Les différentes parties du modèle sont disponibles dans l'onglet architecture du logiciel.

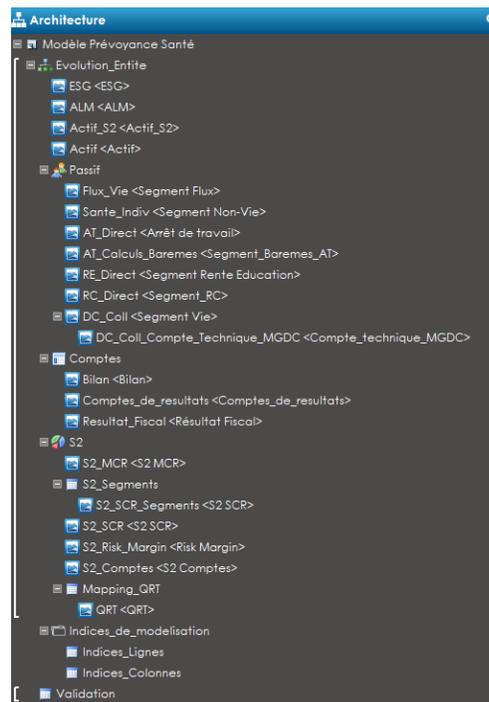


Figure 14. Architecture du modèle Prévoyance Santé

Avant d'être importé sous Gephi, le fichier passe par une étape de traitement de données. Ainsi nous décidons de ne pas retenir les parties du dossier Indices de modélisation. Ces indices permettent d'accéder aux lignes des tables avec des formats prédéfinis comme le bilan ou les comptes de résultats. Par conséquent leur intérêt est relativement limité dans la modélisation.

Nous voulons que la visualisation du graphe fasse ressortir cette architecture. Les différentes parties du modèle doivent être facilement visibles et identifiables sur notre graphe. Par ailleurs le nombre de nœuds et de liens est très élevée. Il nous faut donc utiliser un algorithme avec une complexité plus faible afin de pouvoir obtenir une carte dans un temps raisonnable. Nous essayons de produire une carte à l'aide de l'algorithme de visualisation OpenOrd (Martin, et al., 2011).

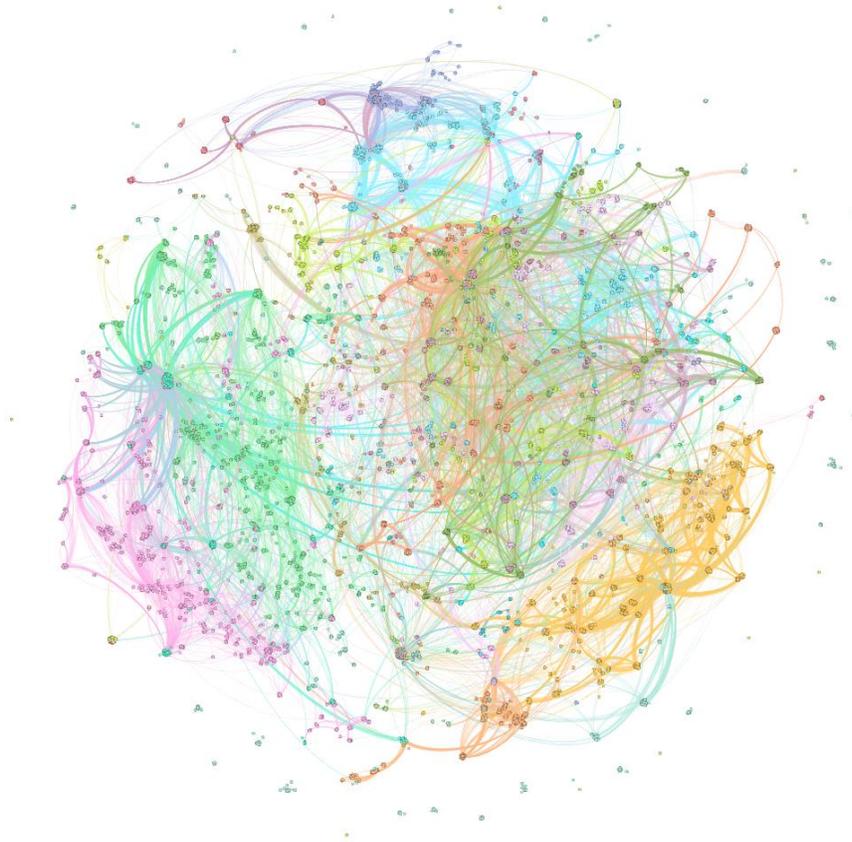


Figure 15. Première version de la carte du modèle Prévoyance Santé

Nous représentons avec une couleur identique les variables appartenant à la même partie du modèle comme précédemment. Pour rappel la carte représente l'ensemble des dépendances informatiques entre les variables du modèle. La carte obtenue est difficilement interprétable car

la répartition des nœuds dans l'espace ne respecte pas l'organisation du modèle. En effet nous pouvons voir que des variables appartenant à des parties différentes du modèle sont représentées côte à côte, ce qui est contre-intuitif du fait de l'architecture retenue sous addactis® Modeling.

Pour faciliter la compréhension de notre carte, nous allons donc répartir les points de manière à retranscrire l'architecture du modèle.

2.2.1.2. Méthode de regroupement pour la restitution de l'architecture du modèle

Nous avons vu précédemment que les algorithmes de placement par des forces plaçaient les nœuds selon les deux critères suivants :

- deux nœuds connectés par un lien doivent être proches l'un de l'autre ;
- les nœuds ne doivent pas être trop rapprochés les uns des autres.

Nous utilisons le premier point pour regrouper les variables appartenant à la même partie du modèle. La méthode consiste à utiliser une structure de graphe modifiée pour appliquer l'algorithme de forces de placements. Pour cela, nous ajoutons sur notre graphe un nœud de groupe pour chaque partie du modèle que nous relient aux nœuds appartenant à la partie correspondante. Nous appellerons par la suite ses liens ajoutés : liens de groupe. Nous faisons donc apparaître directement dans la structure modifiée du graphe l'architecture que nous voulons faire ressortir. Étant donné le premier point de l'algorithme, les nœuds appartenant à la même partie du modèle se retrouvent donc autour du nœud de groupe correspondant.

Cette modification de la structure du graphe est effectuée dans une étape de traitement de données en amont de la visualisation. Les divers retraitements que nous effectuons sur le graphe sont notifiées durant cette étape afin de pouvoir être filtrée lors de la visualisation finale. Nous illustrons cette méthode de modification de la structure à l'aide du modèle Comptes

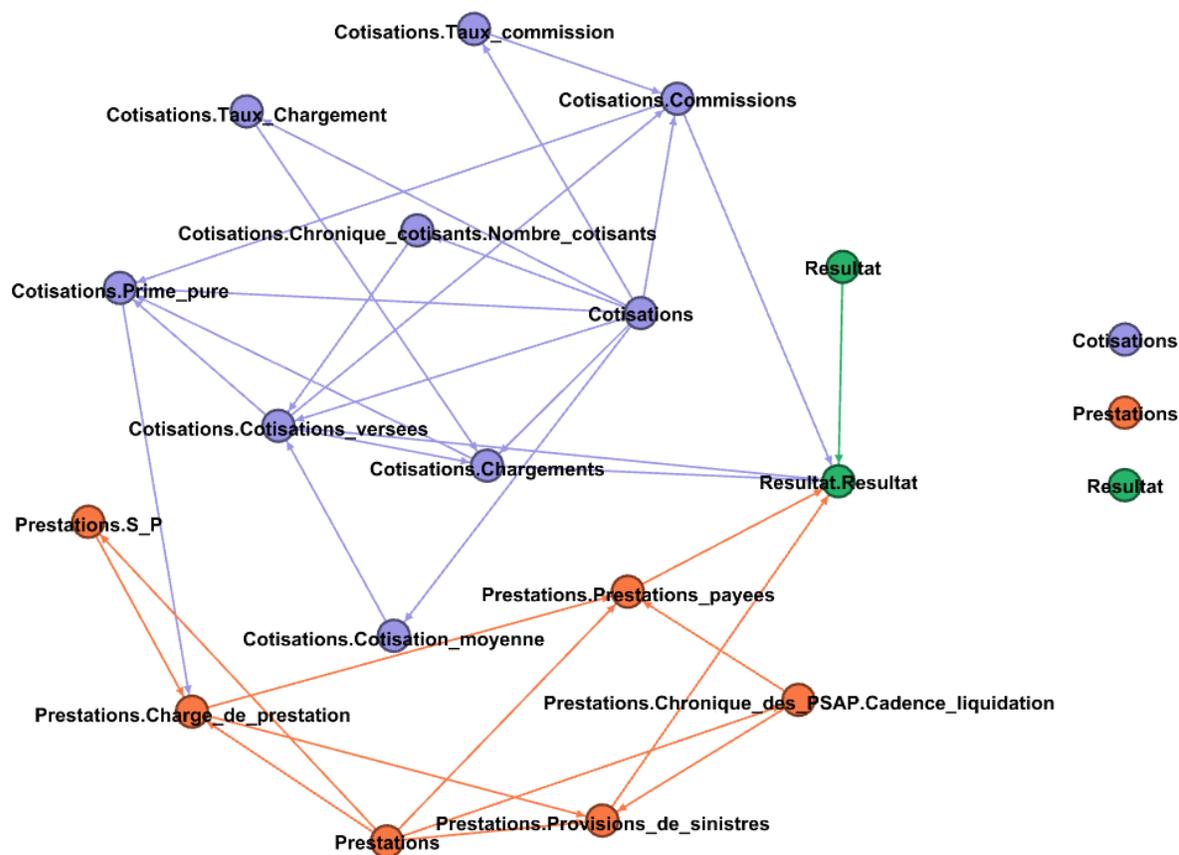


Figure 16. Visualisation de la structure modifiée du graphe du modèle Comptes

Les nœuds de groupe ajoutés sont *Cotisations*, *Prestations* et *Resultat*. Nous pouvons voir que pour chaque partie du modèle, les différents nœuds sont reliés au nœud de groupe correspondant. Par exemple, le lien allant du nœud *Prestation* au nœud *Prestations.Prestations_payees* fait partie des liens qui ont été ajoutés.

Par ailleurs la fonction cible que l'algorithme OpenOrd (Martin, et al., 2011) essaie de minimiser fait intervenir les poids des liens. Pour améliorer le regroupement, nous attribuons un poids plus élevé aux liens de groupe afin de favoriser leur effet sur le placement des nœuds. En appliquant la méthodologie au graphe du modèle Prévoyance Santé, nous obtenons le graphe suivant :

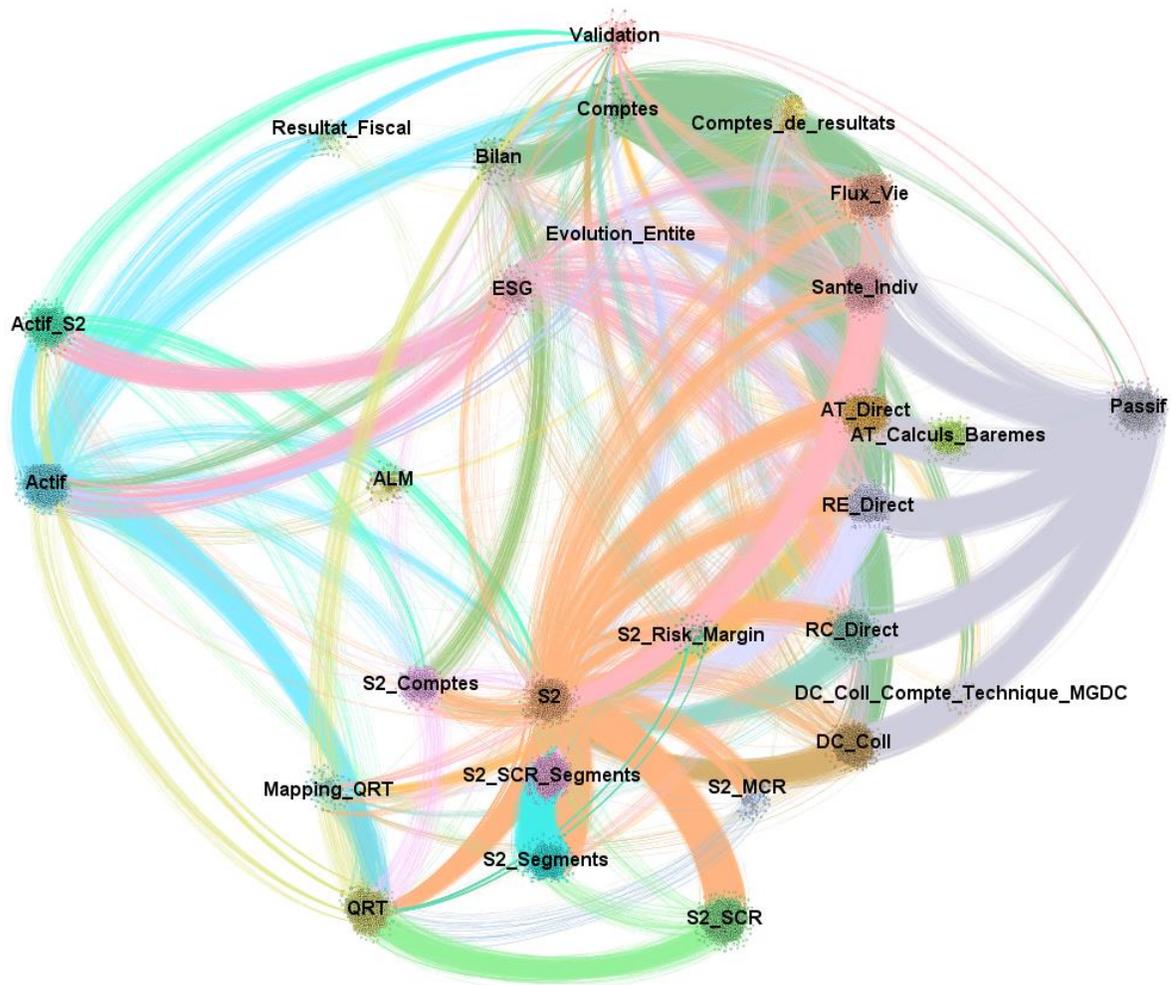


Figure 17. Visualisation du modèle Prévoyance Santé à l'aide de la méthode de regroupement

Nous distinguons maintenant clairement les parties de l'architecture du modèle Prévoyance Santé. Pour rappel, les différentes parties du modèle sont représentées avec la même couleur et leur nom est affiché. La couleur des liens est déterminée par leur source. Si un lien va de la partie S2 vers la partie Santé_Indiv, alors le lien aura la couleur de la partie S2.

L'orientation des courbes permet de déterminer le sens des liens sur la visualisation. Si un lien va d'une variable portant sur le passif vers une variable portant sur le segment santé, alors la courbe sera dans le sens horaire.

De manière générale nous pouvons voir sur notre carte les grandes parties du modèle Prévoyance et Santé évoquées dans sa présentation :

- le passif à droite avec l'organisation par segment ;

- les comptes sur la partie haute ;
- les éléments de Solvabilité 2 sur la partie basse ;
- l'actif à gauche.

Par ailleurs il est possible à l'aide de la carte de résumer rapidement les interactions entre les différentes parties du modèle.

Prenons comme exemple l'organisation des segments de passif du modèle.

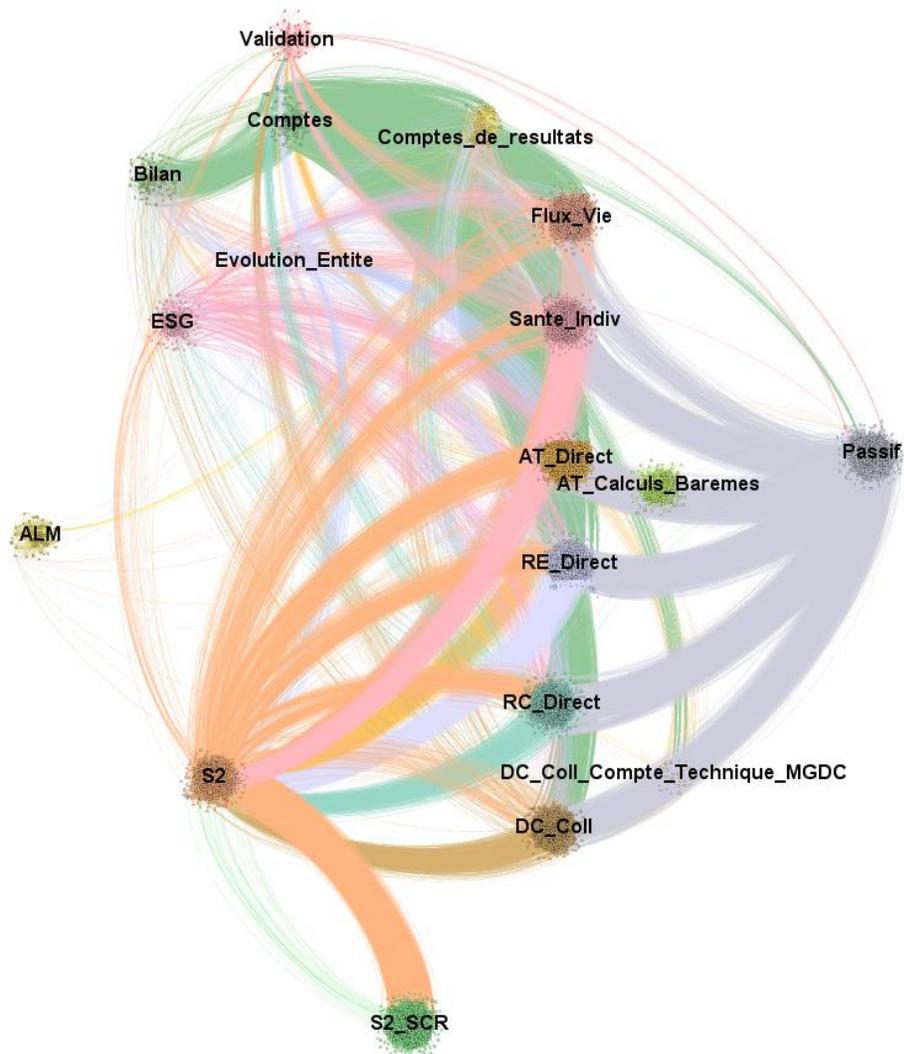


Figure 18. Visualisation des segments de passif du modèle Prévoyance Santé

Analyse des entrées :

D'une part nous pouvons voir que les segments du modèle (Flux_Vie, Sante_Indiv, AT_Direct...) sont alimentés par les parties Passif, Comptes, ESG et S2. En effet ce sont dans ces parties que sont chargées les différentes hypothèses utilisées pour leur modélisation.

- La partie Passif contient des informations comme les taux de frais, les caractéristiques des portefeuilles de rente ou les ratios S/P.
- La partie Comptes contient les tables des bilans comptables et des comptes techniques. Les montants de provisions comptables du bilan ainsi que les montants de cotisations sont notamment utiles pour la modélisation des segments de passif.
- La partie ESG contient la courbe des taux et les coefficients d'actualisation utilisés pour calculer les *Best Estimates*.
- La partie S2 contient l'ensemble des chocs de la formule standard permettant le calcul des *Best Estimates* choqués.

Analyse des sorties :

Les segments alimentent les parties suivantes Comptes de résultat, Bilan, S2 et Validation.

- La partie Comptes de résultat récupère l'ensemble des postes du compte de résultat projetés dans les segments de passif comme les prestations ou les cotisations.
- La partie Bilan récupère l'ensemble des postes du bilan projetés dans les segments de passif comme les provisions comptables.
- La partie ALM récupère l'ensemble des flux de trésorerie générés par les comptes techniques des segments.
- La partie S2 récupère l'ensemble des *Best Estimates* choqués et centraux calculés dans les segments de passif. Les différentes tables de *Best Estimates* alimentent par la suite la partie SCR où les calculs des capitaux de solvabilité requis sont effectués.
- La partie Validation récupère l'ensemble des contrôles effectués dans les différents segments de passif. Des tests sont effectués par exemple sur la cohérence des notations des réassureurs renseignées par les utilisateurs et sur les âges des portefeuilles de rentes.

2.2.2. Méthode pour visualiser les calculs

Dans la section précédente, nous avons détaillé la mise en place de la méthodologie sur la globalité des calculs effectués par le modèle Prévoyance Santé. Nous avons ainsi obtenu une représentation des interactions entre les différentes parties du modèle. Cependant elle ne nous permet pas de visualiser l'organisation des calculs effectués au sein de ces parties. Nous nous intéresserons donc dans cette section à la représentation des calculs à l'intérieur des modules composant le modèle.

Comme nous l'avons vu dans la première partie, le modèle dispose d'une structure modulaire. Pour obtenir les données de dépendances informatiques il nous suffit d'utiliser la fonction de débogage générant les données de dépendance au niveau d'un module. Après la génération du fichier, nous appliquons le traitement des données de la section 2.1.1. Pour la visualisation nous utilisons la méthode de Sugiyama qui est la plus adaptée pour représenter la hiérarchie des *DAG*.

Pour illustrer notre méthode nous considérons le module de Cotisations qui est commun à l'ensemble des segments du passif. Nous obtenons le graphe suivant :

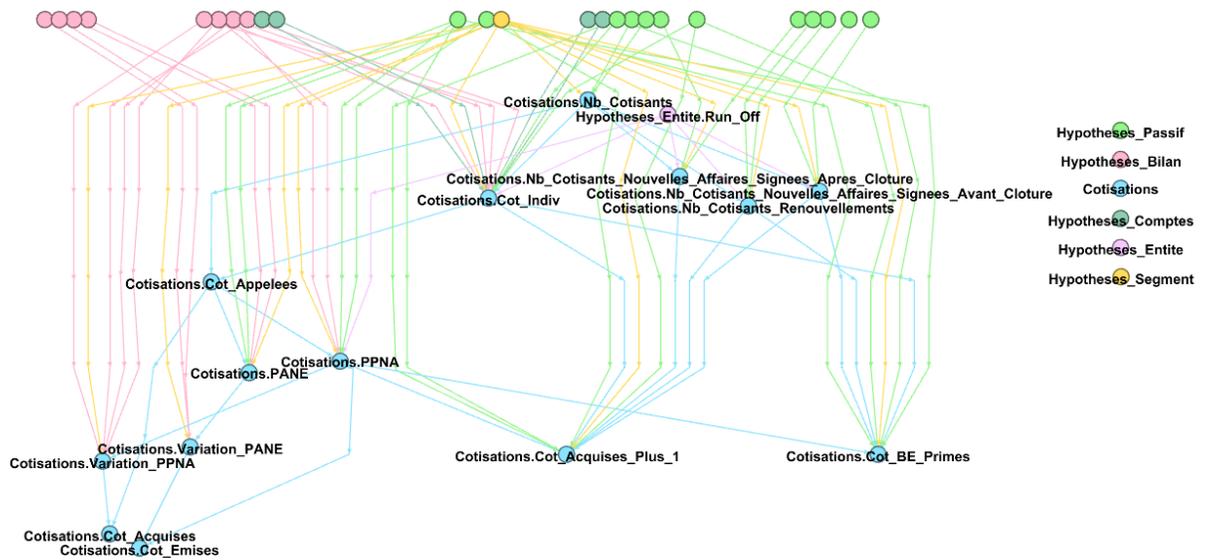


Figure 19. Visualisation du module Cotisations

Pour faciliter la visualisation nous n'affichons que les noms des nœuds situés en dessous du premier niveau. Un algorithme permettant de rendre les noms visibles a été appliqué, ce qui explique pourquoi les nœuds ne sont pas complètement alignés sur le même niveau. Comme nous l'avons vu précédemment, les hypothèses pour le calcul des cotisations se situent au premier niveau du graphe :

- Les hypothèses sur le bilan comportent les postes de PPNA et de PANE de l'année N ainsi que ceux de l'année N-1. Les renseigner sur l'année N-1 permet notamment de calculer la variation des PPNA et des PANE lors de l'année N.
- Les hypothèses sur le passif comportent le nombre initial de cotisants, les taux d'évolution des cotisations, du nombre de cotisants (avant et après la clôture) ainsi que les taux de PPNA et de PANE appliqués aux cotisations appelées.
- Les hypothèses sur les comptes comportent les postes de cotisations vie et non-vie de l'année N. Ces postes sont divisés par le nombre initial de cotisants afin d'initialiser la cotisation en 0.
- L'hypothèse sur le segment est un numéro permettant d'accéder à son emplacement dans les tables regroupant des hypothèses pour l'ensemble des segments.

- L’hypothèse sur l’entité est une indicatrice de *run-off*¹². Si elle est activée, la projection des cotisations ne dépassent pas l’année N+1. Nous pouvons voir son effet sur les différents éléments de cotisations projetés en sélectionnant ses descendants :

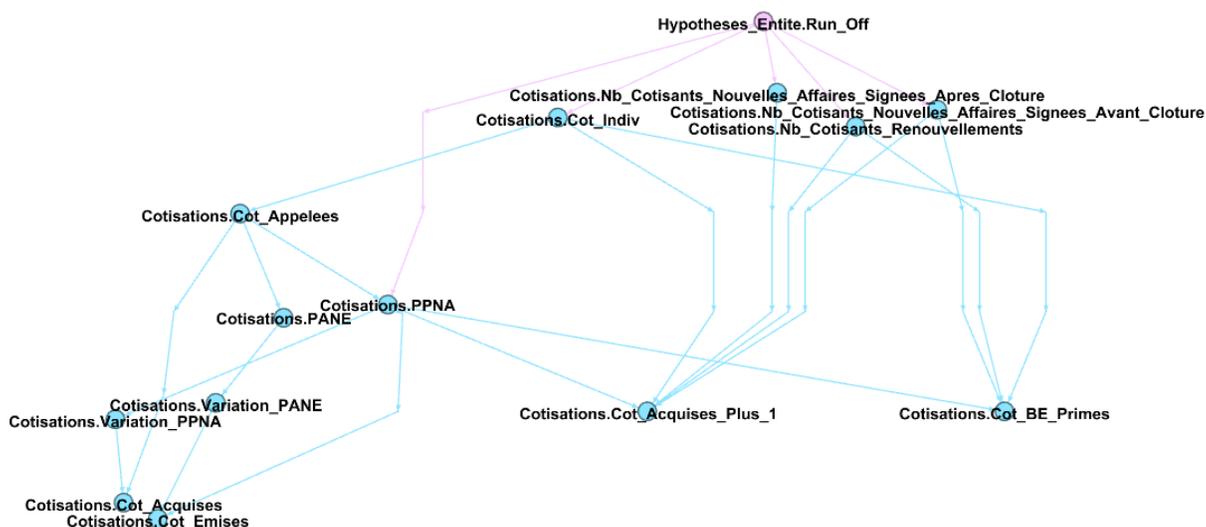


Figure 20. Visualisation des descendants de l’indicatrice de *run-off*

Nous pouvons voir que l’indicatrice intervient au niveau de la cotisation individuelle, de la PPNA, des nombres de cotisants issus des renouvellements et des affaires nouvelles signées avant et après la clôture. Ainsi lorsqu’elle est activée, le montant d’affaire nouvelle après la clôture est annulé en N alors que les autres montants sont annulés en N+1.

- Enfin les sorties du module, situées au dernier niveau, sont les cotisations acquises et les cotisations émises. Elles sont obtenues en retranchant respectivement la variation des PPNA et des PANE.

¹² *run-off* (liquidation) : une entité est en *run-off* si elle ne réalise plus de souscription.

Synthèse de la partie 2.2

Nous avons vu deux méthodes distinctes permettant de créer des cartes à partir des données de dépendances informatiques des modèles actuarielles. La synthèse est effectuée dans le tableau ci-dessous :

| Caractéristiques | Carte des interactions (2.2.1) | Carte de la hiérarchie des calculs (2.2.2) |
|--------------------------------|--|---|
| <u>Modèles concernés</u> | Modèles composés de plusieurs modules | Modèles composé d'un unique module |
| <u>Algorithme de placement</u> | OpenOrd | Méthode de Sugiyama |
| <u>Méthode</u> | S'appuyer sur l'architecture présente dans le modèle et appliquer l'algorithme | Appliquer l'algorithme de placement directement |
| <u>Objectif</u> | Visualiser les interactions entre les différents modules | Représenter l'ordre des calculs dans un module |
| <u>Utilisations</u> | Contrôler les interactions entre les différents modules | Comprendre l'organistaion des calculs |

Tableau III. Tableau de synthèses des cartes des modèles actuariels

2.3. Applications à la modélisation actuarielle

Nous avons vu dans la section précédente la mise en œuvre de la cartographie des modèles développés par le cabinet. Nous avons ainsi distingué deux méthodes en fonction des modèles étudiés. Dans cette section elles seront mises en pratique pour résoudre des problèmes de modélisation actuarielle. Pour chaque problème nous verrons des impacts chiffrés fournis par le modèle prévoyance santé standard à partir des hypothèses d'un organisme d'assurance que nous présenterons au préalable.

2.3.1. Présentation des hypothèses de modélisation

Les hypothèses de l'organisme modélisé ont été obtenues en agrégeant les données de plusieurs institutions de prévoyance. Cet organisme propose cinq produits d'assurance qui correspondent aux « segments » de passif présents dans le modèle standard :

- le produit **frais de soin de santé** : la garantie est annuelle et consiste à rembourser les actes médicaux de l'assuré ;
- le produit **décès** : la garantie est annuelle et consiste à verser un capital aux ayants-droit lors du décès de l'assuré ;
- le produit **rente de conjoint** : la garantie consiste au versement d'une rente viagère et/ou temporaire destinée au conjoint de l'assuré après le décès de dernier ;
- le produit **rente d'éducation** : la garantie consiste au versement d'une rente temporaire jusqu'à la fin des études destinée aux enfants de l'assuré après le décès de dernier ;
- le produit **arrêt de travail** : il se décompose en trois garanties distinctes : l'incapacité, l'invalidité et le maintien de la garantie décès. Les garanties incapacité et invalidité consistent en une rente périodique destinée à l'assuré en fonction de la durée et de son degré d'inaptitude au travail. Le maintien de la garantie décès consiste à verser un capital aux ayants-droit lors du décès de l'assuré après que celui-ci soit entré en arrêt de travail.

Ces différents produits d'assurance sont classés de la façon suivante sous Solvabilité 2 :

| Produits modélisés | Ligne d'activité (<i>LoB</i>¹³) |
|-------------------------------|---|
| Santé | <i>NSLT</i> ¹⁴ – Assurance des frais médicaux (1) |
| Décès | <i>Life</i> – Autres assurances vie (32) |
| Incapacité | <i>NSLT</i> – Assurance de protection de revenu (2) |
| Invalidité | <i>SLT</i> ¹⁵ – Rentes découlant des contrats d'assurance non-vie liées aux engagements d'assurance santé (33) |
| Maintien de la garantie décès | <i>Life</i> – Autres assurances vie (32) |
| Rente d'éducation | <i>Life</i> – Autres assurances vie (32) |
| Rente de conjoint | <i>Life</i> – Autres assurances vie (32) |

Tableau IV. Correspondance entre les produits modélisés et les lignes d'activité

Nous donnons ensuite les principales caractéristiques de ces différents produits dans l'activité de l'organisme au 31/12/2017:

¹³ *LoB (Line of business)* : ligne d'activité en français, classification sous Solvabilité 2 des produits d'assurances.

¹⁴ *NSLT (Non similar to life techniques)* : classification sous Solvabilité 2 regroupant les garanties d'assurance santé ne reposant pas sur une base technique similaire à la vie : frais de soins de santé, accident de la vie, indemnités journalières et d'incapacité.

¹⁵ *SLT (Similar to life techniques)* : classification sous Solvabilité 2 regroupant les garanties d'assurance santé reposant sur une base technique similaire à la vie : dépendance, invalidité.

| Produits | Cotisations acquises (en millions €) | Best Estimates bruts (en millions €) | Duration (en années) |
|----------------------------------|---|---|---------------------------------|
| Santé | 143,6m€ | 52,0m€ | 1,14 |
| Décès | 4,4m€ | 7,3m€ | 1,03 |
| Incapacité | 32,1m€ | 41,6m€ | 6,61 |
| Invalidité | - | 84,8m€ | 6,32 |
| Maintien de la garantie décès | - | 13,7m€ | 6,20 |
| Rente d'éducation | 1,3m€ | 15,4m€ | 6,20 |
| Rente de conjoint | 6,1m€ | 54,2m€ | 14,74 |

Tableau V. Principales caractéristiques des produits de l'organisme tableau étudié

Nous voyons que deux types de produits se distinguent par leur durée : la santé et le décès dont la durée est proche de 1 et les produits de rente dont la durée est plus élevée. En effet le premier type de produits concerne des engagements de courtes durées.

Le résumé du bilan prudentiel de l'organisme est donné dans les tableaux suivants :

| Postes du bilan | Solvabilité 2 (en millions €) | Solvabilité 1 (en millions €) |
|-------------------------------|--|--|
| Fonds Propres | 327,0m€ | 304,8 m€ |
| Provisions techniques brutes | 344,2 m€ | 308,6 m€ |
| <i>Best Estimates bruts</i> | 332,7 m€ | - |
| Marge de risque | 11,5 m€ | - |
| Autres passifs | 224,9 m€ | 218,1 m€ |
| <u>Total du passif</u> | 896,1 m€ | 831,5 m€ |

Tableau VI. Passif du bilan prudentiel de l'organisme

| Postes | Solvabilité 2 (en millions €) | Solvabilité 1 (en millions €) |
|--------------------------------|--|--|
| Placements | 654,9 m€ | 614,0 m€ |
| Provisions techniques cédées | 63,9 m€ | 72,1 m€ |
| Autres actifs | 177,3 m€ | 145,4 m€ |
| <u>Total de l'actif</u> | 896,1 m€ | 831,5 m€ |

Tableau VII. Actif du bilan prudentiel de l'organisme

Les montants des capitaux de solvabilité réglementaires de l'organisme de prévoyance santé sont les suivants :

| <i>SCR</i> | <i>MCR</i> | Ratio de couverture du <i>SCR</i> | Ratio de couverture du <i>MCR</i> |
|-------------------|-------------------|--|--|
| 127,9m€ | 33,7m€ | 256% | 970% |

Tableau VIII. Capitaux de solvabilité réglementaire de l'organisme

Ainsi nous modélisons l'activité d'un organisme de prévoyance santé dont les engagements se répartissent principalement entre des produits arrêt de travail, rente de conjoint et santé. Par ailleurs la couverture des exigences de solvabilité est largement respectée car les deux ratios sont supérieurs à 100%.

Après la présentation de l'organisme, nous allons nous intéresser à différents développements effectués sur le modèle prévoyance santé. Les hypothèses de l'organisme nous permettront d'étudier des impacts chiffrés liés à ces améliorations.

2.3.2. Calcul de la provision pour risque d'exigibilité

Nous nous intéresserons dans cette partie au calcul de la provision pour risque d'exigibilité effectué au sein du modèle prévoyance santé. Nous visualiserons tout d'abord à l'aide de nos cartographies les interactions entre les modules permettant ce calcul et nous proposerons une amélioration de cette organisation. Enfin un exemple chiffré de l'impact de cette provision sera donné à l'aide des hypothèses de notre organisme.

2.3.2.1. Réorganisation du calcul de la provision pour risque d'exigibilité

Nous détaillons l'analyse ayant conduit à un développement sur le modèle Prévoyance Santé. Par conséquent la première cartographie est réalisée à partir d'une version antérieure du modèle, ce qui explique pourquoi la carte est différente de celle présente dans la section 2.2.

Nous considérons l'étude des variables en sortie du segment Santé. Pour cela nous affichons les variables qui dépendent informatiquement de ce module.

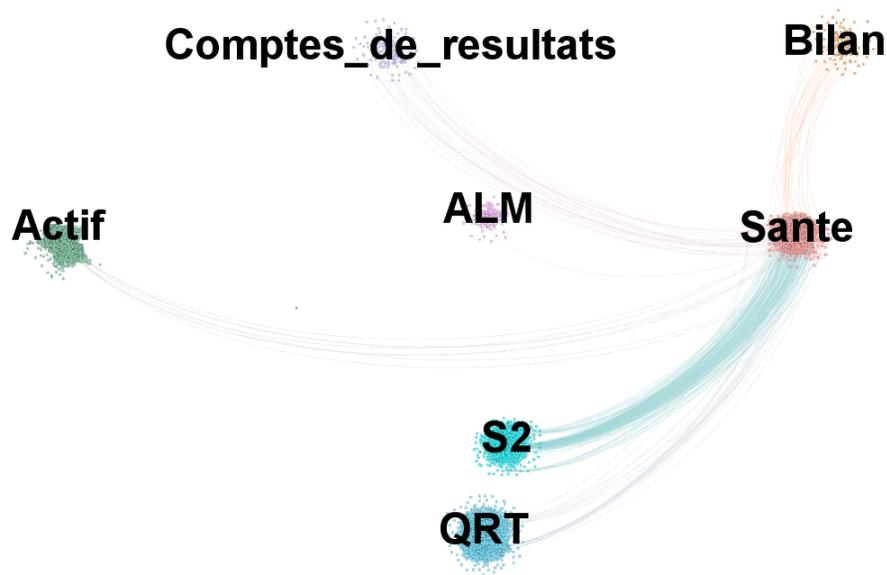


Figure 21. Carte des variables en sortie du module Santé

Nous remarquons que des variables du module Santé interviennent directement dans le module de l'Actif. Or la modélisation du passif au sein du modèle Prévoyance Santé n'est censée communiquer avec l'Actif que par l'intermédiaire du module ALM. Les variables concernées sont les suivantes :

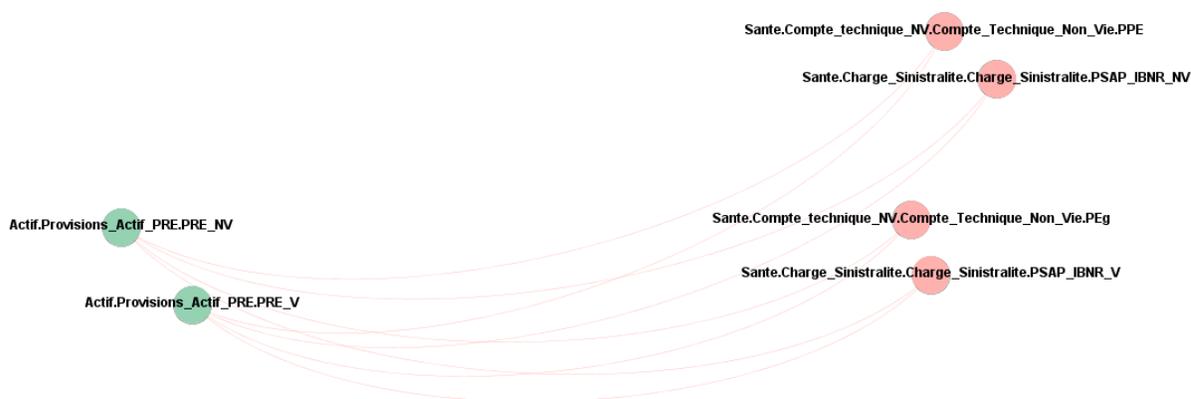


Figure 22. Carte des variables de l'Actif dépendant de la Santé

Ainsi nous pouvons voir que les provisions pour sinistres à payer (PSAP), les IBNR, les provisions pour participations aux excédents (PPE) et les provisions pour égalisation (PEg) du module Santé interviennent dans le calcul de la provision pour risque d'exigibilité (PRE) vie et non-vie.

La PRE est, d'après l'article R 331 - 6 du code des assurances, une provision destinée à faire face aux engagements dans le cas de moins-value de l'ensemble des actifs mentionnés à l'article R 332 - 20 (actions, OPCVM, prêts et dépôts et immobiliers principalement) qui sont comptabilisés en prix de revient. Un actif est dit en moins-value latente lorsque sa valeur de réalisation est inférieure à sa valeur nette comptable. Son but est de compenser une éventuelle moins-value latente globale sur ces actifs après dotation de la provision pour dépréciation durable (PDD¹⁶). La charge de la dotation à la PRE peut être étalée sur une période allant de 3 à 8 ans. Étant donné que cette provision concerne l'actif, son calcul est effectué dans le module

¹⁶ PDD : provision pour compenser les moins-values durables relatives aux placements mentionnés à l'article R 332 – 20 du code des assurances (actions, OPCVM, prêts et dépôts et immobiliers principalement)

correspondant. Afin de visualiser son utilisation dans le reste du modèle, nous considérons ces descendants au troisième niveau.

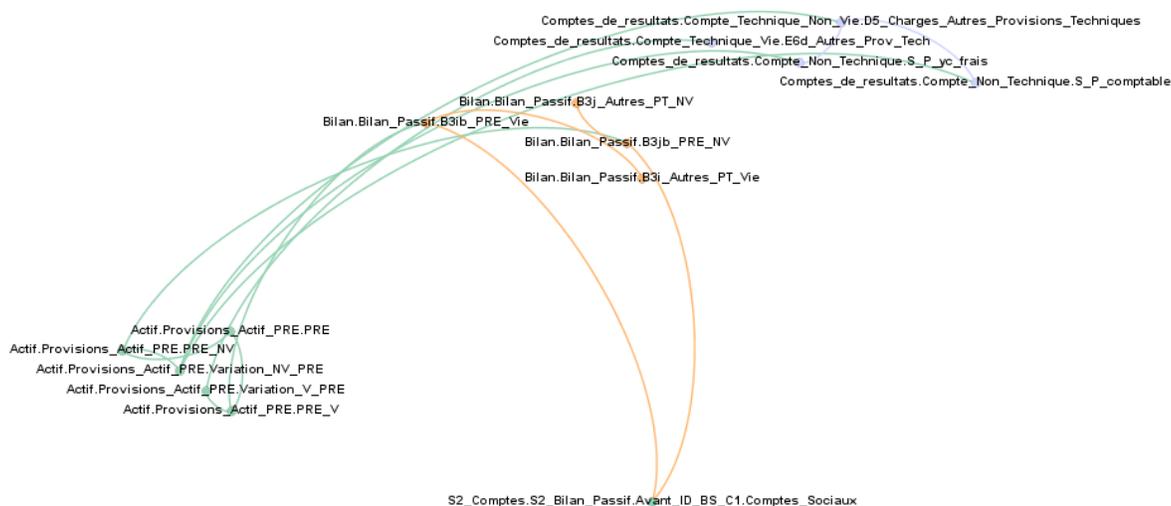


Figure 23. Visualisation des descendants de la PRE

Cependant pour alimenter les postes du bilan, il est nécessaire de ventiler cette provision entre les risques vie et non-vie. Cette allocation est faite au prorata des provisions techniques correspondantes et fait donc intervenir les montants de provisions calculés dans les segments de passif. Nous avons ainsi :

$$PRE_{Vie} = PRE \cdot \frac{\text{Provisions techniques Vie}}{\text{Provisions techniques Vie} + \text{Provisions techniques non Vie}}$$

$$PRE_{Non Vie} = PRE \cdot \frac{\text{Provisions techniques Non Vie}}{\text{Provisions techniques Vie} + \text{Provisions techniques non Vie}}$$

Étant donné que la ventilation intervient pour alimenter le bilan, nous déplaçons ce calcul au niveau du module adéquat.

Par ailleurs nous remarquons aussi que la variation entre les exercices comptables des PRE vie et non-vie est effectuée dans le module Actif (variables sur le graphe : *Actif.Provisions_Actif_PRE.Variation_NV_PRE* et de même pour la vie). Cependant elle n'est

utilisée qu’au niveau du compte de résultats. Nous suivons donc la même logique que précédemment et nous déplaçons le calcul de la variation au niveau du module du Compte de Résultats.

Après avoir apporté la correction au modèle, nous la validons en effectuant la nouvelle cartographie du modèle :

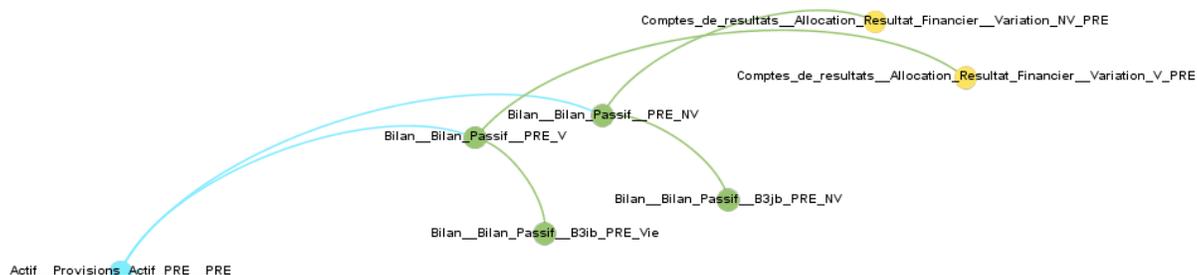


Figure 24. Visualisation des descendants de la PRE après correction

Après le développement, nous voyons que seul le calcul de la PRE est effectué au sein du module Actif puis que la ventilation et la variation sont effectués dans les modules correspondants.

Ainsi la carte nous a permis de rendre les interactions entre les parties du modèle Prévoyance Santé plus cohérentes. L’intérêt principal de ce développement est de rendre autonome les différents modules. Cela peut par exemple faciliter la création d’un modèle de projection de l’actif car les entrées de la brique seront clairement identifiées.

2.3.2.2. Impacts de la provision pour risque d’exigibilité

Pour comprendre l’intérêt de ces développements, nous considérons le cas où la PRE n’est pas ventilée car les provisions sont manquantes dans le module actif. Cette situation était tout à fait envisageable car l’interaction directe entre les modules de l’Actif et du Passif n’était normalement pas prévue dans le modèle prévoyance santé et nous aurions donc pu nous retrouver avec une PRE nulle. Ainsi nous allons comparer les résultats fournis par deux modèles : le premier prenant en compte la PRE et la deuxième ne la prenant pas en compte.

Pour constater l'impact d'une PRE mise à zéro par erreur, nous devons projeter l'activité de notre organisme selon un scénario de rendement défavorable des actifs car la PRE n'est dotée qu'en cas de moins-value latente globale. En effet dans la situation centrale de l'organisme considéré les rendements sont supposés positifs et le montant de PRE inscrit au bilan est nul. Nous considérons par exemple la situation du choc de marché de la formule standard, soit une baisse de 39% de la valeur de marché des actions et de 25% de la valeur de marché de l'immobilier. La couverture des capitaux de solvabilité évolue de la façon suivante :

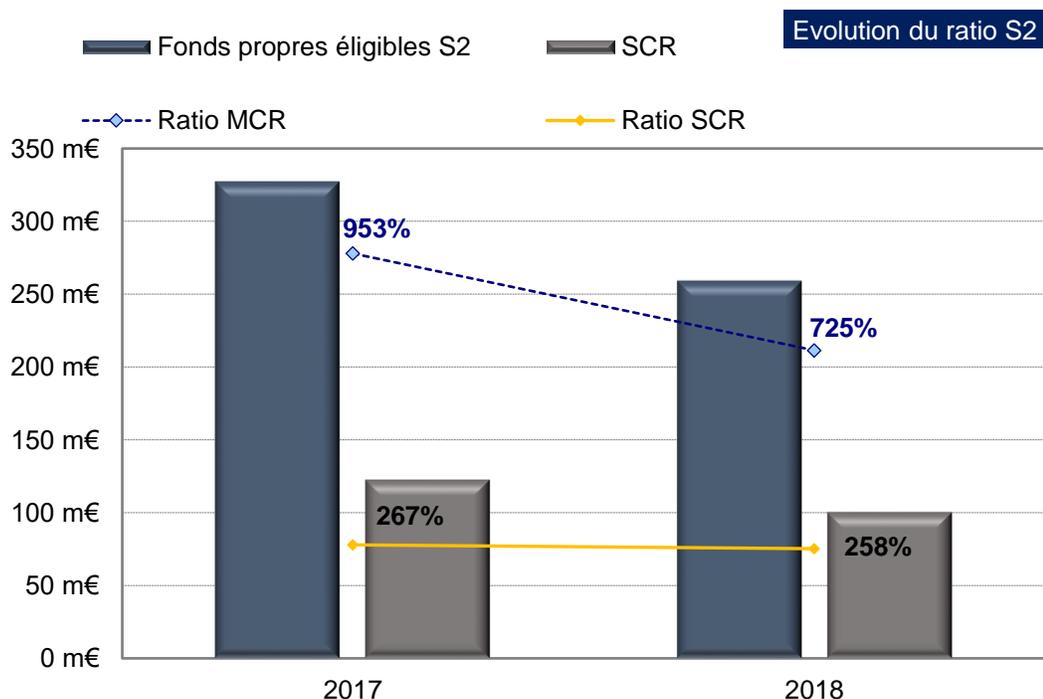


Figure 25. Evolution des capitaux de solvabilité réglementaire dans un scénario de rendement défavorable

Ainsi les chocs sur les valeurs de marché des actifs entraînent une baisse des fonds propres économiques d'environ 70 millions d'euros. Cette baisse des valeurs de marché provoque d'autre part une baisse des SCR pour les risques immobilier et action notamment. La conséquence de ces deux effets qui se compensent partiellement est que le ratio de solvabilité ne baisse que de 9 points.

Nous allons maintenant comparer les résultats de ce scénario de choc avec et sans prise en compte de la PRE. Tout d'abord nous observons un montant de PRE de **10,3 millions d'euros** alors que ce montant est nul dans l'autre cas. Nous constatons ensuite que le montant de fonds

propres économiques a légèrement augmenté dans le modèle sans PRE avec un écart de **155 000 euros**. Cet écart provient du montant net d'impôts différés comme nous pouvons le voir dans le tableau suivant :

| <u>Décomposition des fonds propres S2</u> | <u>2018 avec PRE</u> | <u>2018 sans PRE</u> |
|---|-----------------------------|-----------------------------|
| Fonds Propres S1 | 285,6 m€ | 295,9 m€ |
| Passifs subordonnés acceptés en S2 | 0,7 m€ | 0,7 m€ |
| Autres fonds propres ne rentrant pas dans S2 | 1,6 m€ | 1,6 m€ |
| Fonds de dotation | 1,6 m€ | 1,6 m€ |
| Passage de S1 à S2 (Actif) | 12,5 m€ | 9,0 m€ |
| Total actif S1 | 819,9 m€ | 819,9 m€ |
| Total actif S2 | 832,4 m€ | 828,9 m€ |
| Passage de S1 à S2 (Passif) | 9,3 m€ | 15,9 m€ |
| Total passif S1 | 534,3 m€ | 524,0 m€ |
| Total passif S2 | 543,6 m€ | 539,9 m€ |
| Marge pour risque | 11,0 m€ | 11,0 m€ |
| Marge de risque NL | 0,0 m€ | 0,0 m€ |
| Marge de risque Health NSLT | 9,8 m€ | 9,8 m€ |
| Marge de risque Health SLT | 0,2 m€ | 0,2 m€ |
| Marge de risque Life | 1,0 m€ | 1,0 m€ |
| Position nette des ID | 33,3 m€ | 33,4 m€ |
| Impôts différés actifs | 50,5 m€ | 46,9 m€ |
| Impôts différés passifs | 17,2 m€ | 13,5 m€ |
| Fonds Propres Non Eligibles | 18,2 m€ | 18,4 m€ |
| Fonds propres S2 | 292,0 m€ | 292,2 m€ |

Tableau IX. Décomposition des fonds propres S2 pour les deux situations étudiées

En effet les impôts différés sont des montants d'impôts qui seront dus ou qui seront remboursables dans le futur du fait des écarts de valorisation entre le bilan prudentiel et le bilan fiscal. Ces écarts constituent les différences temporelles imposables qui donneront lieu à des impôts différés passifs et les différences temporelles déductibles qui donneront lieu à des impôts différés actifs.

Le montant net d'impôts différé peut ainsi augmenter les fonds propres économiques sous réserve que l'organisme justifie qu'il est en mesure de réaliser des bénéfices dans le futur. Pour l'organisme que nous modélisons, nous avons supposé qu'il recouvrait la totalité de ses impôts différés nets.

Pour notre exemple, nous constatons une compensation des impôts différés :

- Dans le modèle prenant en compte la PRE, les IDA augmentent du montant du déficit, provoqué par la dotation de la provision, multiplié par le **taux d'imposition sur les sociétés** ;
- Dans le modèle ne prenant pas en compte la PRE, ce sont les IDP qui baissent car la différence de valorisation entre le bilan comptable et le bilan prudentiel est désormais nulle. La baisse est donc constituée du montant de PRE multipliée par le **taux imposition utilisé pour le calcul des impôts différés**.

L'écart que nous constatons sur les ID nets est dû à la différence entre le taux d'imposition utilisé sur le résultat comptable et le taux d'imposition utilisé pour la valorisation des impôts différés qui s'élève à **1,51%**.

Nous avons en effet utilisé les hypothèses suivantes pour les taux d'imposition :

| Année | Taux d'impôt sur les sociétés utilisés pour le résultat | Taux d'impôt sur les sociétés utilisés pour la valorisation des impôts différés |
|-------|---|---|
| 2 017 | 34,43% | 35,94% |
| 2 018 | 34,43% | 35,94% |

Tableau X. Hypothèses retenues pour le taux d'imposition

En effet le différentiel de taux d'imposition appliqué à la provision pour risque d'exigibilité nous donne l'écart constaté sur les fonds propres économiques.

Finalement l'impact de la PRE est relativement mineur sur les fonds propres économiques avec les hypothèses retenues pour la projection de l'activité de l'organisme que nous modélisons et seul la répartition des montants du bilan prudentiel est modifiée.

2.3.3. Calcul de la cession des *Best Estimates*

Nous nous intéresserons dans cette partie au calcul de la cession des *Best Estimates* effectué au sein du modèle prévoyance santé. Nous visualiserons tout d'abord à l'aide de nos cartographies l'ordre des calculs effectués au sein du module permettant ce calcul et nous proposerons une amélioration de cette organisation. Enfin un exemple chiffré de l'impact de cette cession sera donné à l'aide des hypothèses de notre organisme.

2.3.3.1. Réorganisation du calcul des *Best Estimates NSLT*

Dans cette section, nous nous intéresserons aux développements ayant eu lieu sur le module produisant les *Best Estimates* pour les segments appartenant à la classification *NSLT* excepté l'arrêt de travail. Nous visualiserons donc le module avant et après le développement.

Par rapport aux cartes précédentes, nous souhaitons diminuer le nombre de nœuds et de liens à afficher afin de pouvoir discerner clairement la hiérarchie des calculs. Pour cela, nous ajoutons une étape dans notre traitement des données pour supprimer les nœuds et les liens associés aux colonnes des tables. Pour la visualisation nous appliquons la méthode de Sugiyama qui est la plus adaptée pour représenter l'ordre des calculs des modèles composés d'un unique module. Nous obtenons le graphe suivant :

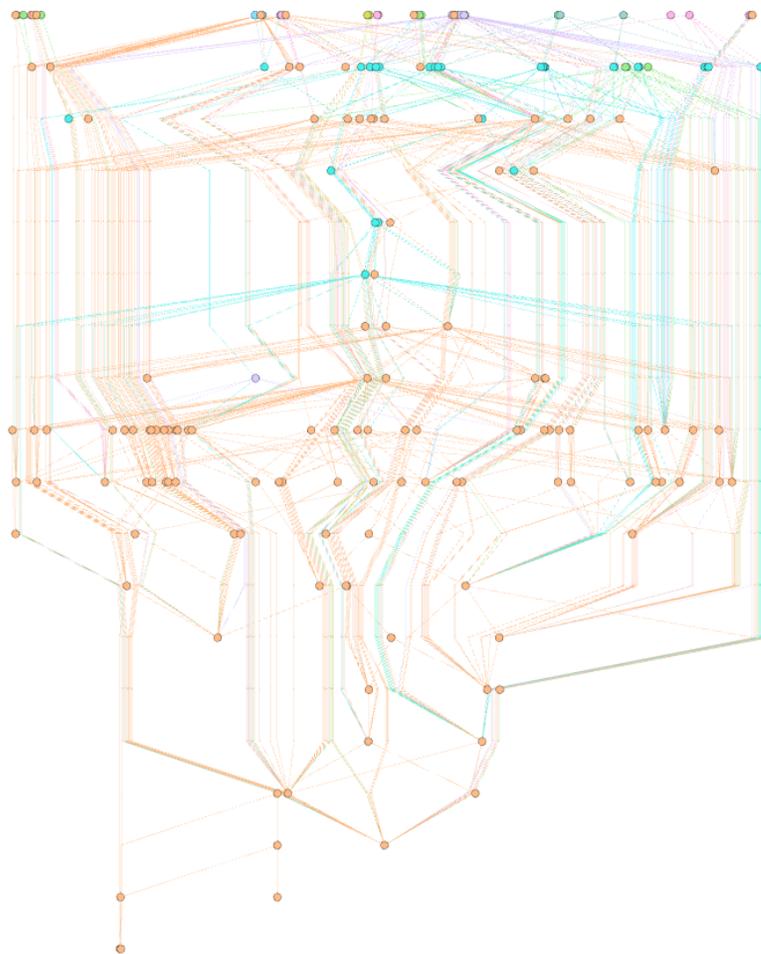


Figure 26. Visualisation de la première version du module de *Best Estimates NSLT*

Nous allons rappeler la prise en compte de la réassurance sous Solvabilité 2 qui doit prendre en compte le risque de défaut du réassureur.

Pour chaque année de survenance et chacun des traités, le *Best Estimate* cédé avant ajustement est obtenu par application du taux de quote-part sur l'assiette de *Best Estimate* brut correspondante.

Puis afin de prendre en compte le défaut moyen du réassureur, un ajustement est calculé en fonction de la probabilité de défaut de réassureur, de son taux de recouvrement et de la duration du *Best Estimate* brut sur lequel porte le traité.

Le calcul de l'ajustement est effectué de la façon suivante :

$$Ajustement = (1 - Taux \text{ Recouvrement}) \frac{Probabilité \text{ Défaut}}{(1 - Probabilité \text{ Défaut})} * Duration_{BE \text{ brut}}$$

Puis nous obtenons le *Best Estimate* cédé après ajustement :

$$BE \text{ Cédé}_{\text{après ajustement}} = BE \text{ Cédé}_{\text{avant ajustement}} * (1 - \text{Ajustement})$$

Les travaux à effectuer sur ce module concernent la réorganisation des hypothèses et le calcul de la cession. Pour les hypothèses, nous voulons supprimer une étape dans l'import des montants calculés et projetés dans les autres modules. Cette étape de transition permettait d'effectuer la recette unitaire lors des premières phases d'implémentation du module. Elle est visible au deuxième niveau du graphe précédent, les hypothèses se situant au premier niveau. Pour la réassurance nous cherchons à réduire le nombre de variables intermédiaires intervenant dans le calcul des *Best Estimates* cédés. Cela rendra notamment le module plus lisible et améliorera la traçabilité des calculs.

Les montants de *Best Estimates* cédés sont évalués par année de survenance, par application du traité de réassurance correspondant, au sein des tables *BE_C_surv*.

Pour comprendre l'organisation du calcul de la cession nous considérons les ancêtres de la table *BEL_C* en sortie du module :

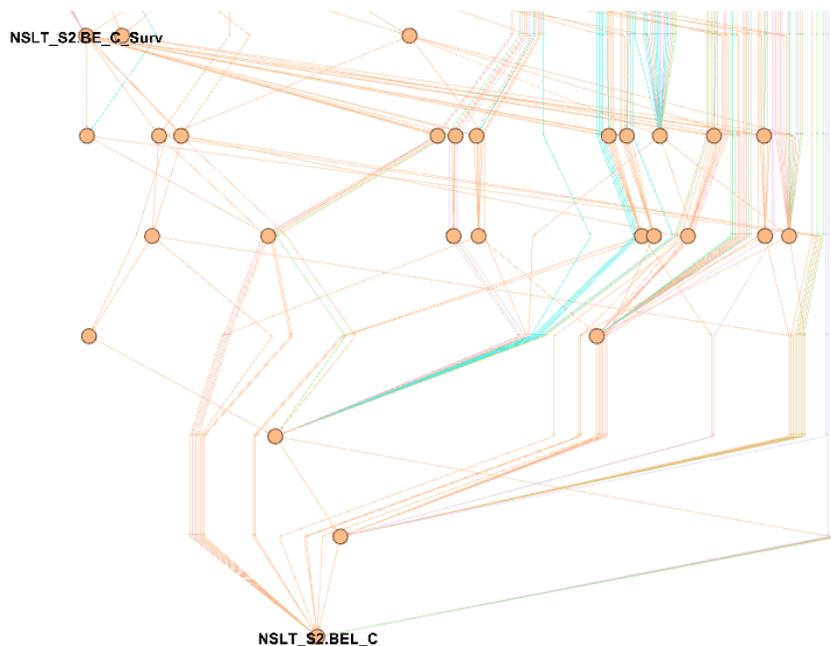


Figure 27. Visualisation des ancêtres de la table *BEL_C*

Le calcul de la cession est effectué par année de survenance des sinistres pour chaque réassureur dans la table *BE_C_Surv* dont le nom apparaît sur le graphe. Nous pouvons visualiser ces étapes intermédiaires. Ce sont notamment au cours de celles-ci que sont retraités les montants de PANE, de PEG et de PPE cédées ainsi que les commissions de réassurance. Pour simplifier cette organisation nous créons deux tables effectuant ces différents retraitements.

La première table est *Calcul_Reass*. Elle permet de récupérer les montants de provisions comptables cédées provenant d'un autre module et de calculer les *Best Estimates* cédés par réassureur.

La deuxième table est *Detail_Reass*. Elle agrège la table précédente par réassureur et effectue le retraitement des PANE des *Best Estimates* de primes.

Nous obtenons le graphe suivant à la fin de notre développement :

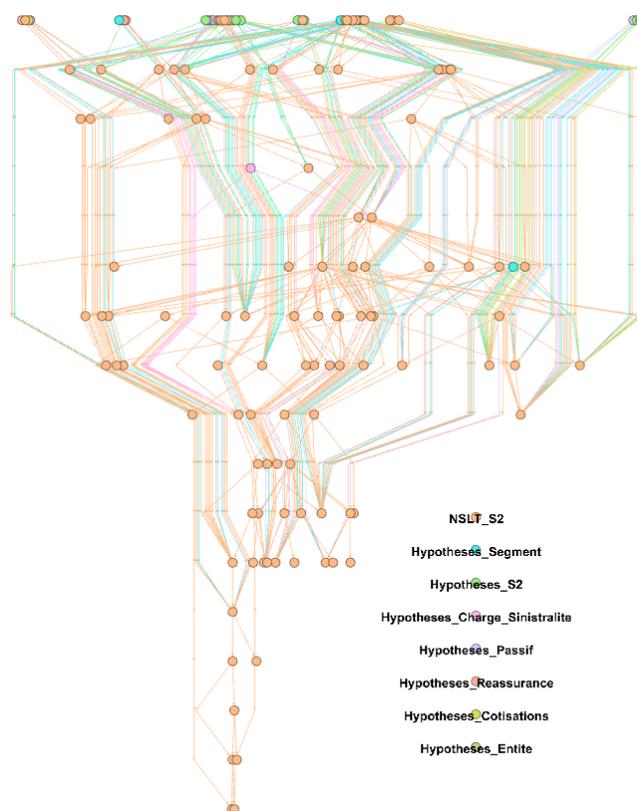


Figure 28. Visualisation du module de calculs de *Best Estimates* pour la santé après le développement

Tout d'abord nous remarquons que le deuxième niveau n'est plus occupé par les variables de *Calculs_Temporaires*. Le nombre de variables utilisés dans le module est ainsi réduit. Puis

pour confirmer notre deuxième développement, nous considérons les ancêtres de la table *BEL_C* :

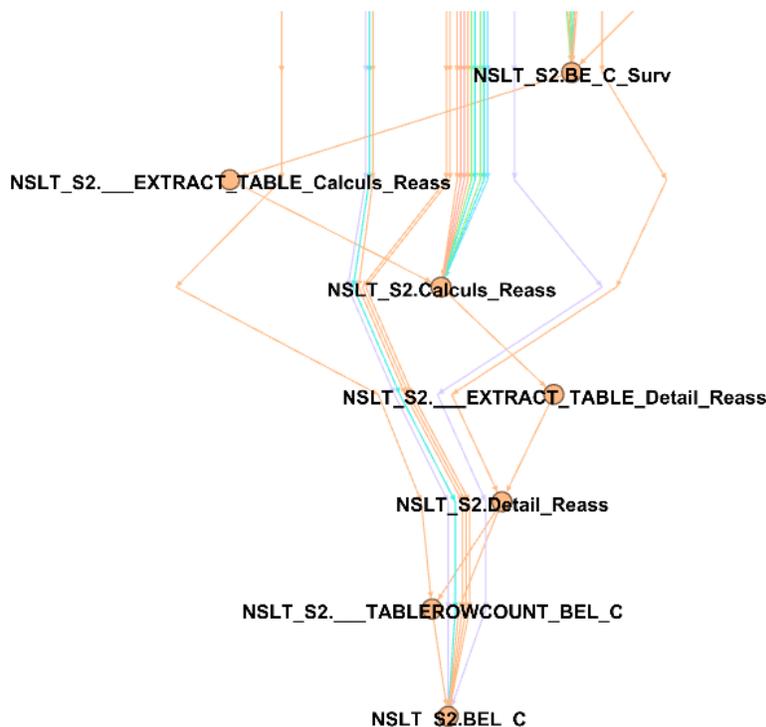


Figure 29. Visualisation des ancêtres de la table *BEL_C* après le développement

Nous retrouvons sur le graphe les tables évoquées précédemment. Ainsi nous pouvons voir que le nombre d'étapes entre la table *BE_C_Surv* et la table *BEL_C* est grandement diminué.

2.3.3.2. Impacts sur la cession des *Best Estimates NSLT*

Pour comprendre l'intérêt de ces développements, nous considérons le cas où les PANE cédées ne sont pas retranchées du *Best Estimate* de prime dans le module *NSLT*. Cette erreur pouvait survenir du fait de manque de lisibilité dans les calculs de cession effectués au sein du module. Ainsi nous allons comparer les résultats fournis par deux modèles : le premier prenant en compte le retraitement des PANE cédées et le deuxième ne la prenant pas en compte. Nous rappelons que sous Solvabilité 2, les PANE (Primes Acquisées Non Émises) sont à retrancher du

Best Estimate (conformément aux directives de l'autorité de contrôle (Conférence annuelle IA – ACPR du 9 mars 2017, 2017)).

Nous comparons l'évolution des postes à l'actif et au passif du bilan prudentiel :

| Postes | Avec retraitement des PANE | Sans retraitement des PANE | Avec retraitement des PANE | Sans retraitement des PANE | Ecart Solvabilité 2 | |
|---|----------------------------|----------------------------|----------------------------|----------------------------|---------------------|-------------|
| | Comptes sociaux | Comptes sociaux | Solvabilité 2 | Solvabilité 2 | | |
| Impôts différés passifs | 0,0 m€ | 0,0 m€ | 19,3 m€ | 17,6 m€ | -1,7 m€ | ↓ -9% |
| Dettes nées d'opérations de réassurance | 23,1 m€ | 23,1 m€ | 11,8 m€ | 16,6 m€ | 4,8 m€ | ↑ 40% |
| Total du passif | 525,4 m€ | 525,4 m€ | 569,1 m€ | 572,1 m€ | 3,0 m€ | ↑ 1% |

Tableau XI. Evolution des postes de passif du bilan prudentiel entre les deux modèles

| Postes | Avec retraitement des PANE | Sans retraitement des PANE | Avec retraitement des PANE | Sans retraitement des PANE | Ecart Solvabilité 2 | |
|---|----------------------------|----------------------------|----------------------------|----------------------------|---------------------|-------------|
| | Comptes sociaux | Comptes sociaux | Solvabilité 2 | Solvabilité 2 | | |
| Impôts différés actifs | 0,0 m€ | 0,0 m€ | 24,9 m€ | 23,1 m€ | -1,7 m€ | ↓ -7% |
| Provisions techniques cédées | 72,1 m€ | 72,1 m€ | 63,9 m€ | 68,6 m€ | 4,8 m€ | ↑ 7% |
| Non vie et santé similaire à la non-vie | 18,1 m€ | 18,1 m€ | 11,1 m€ | 15,9 m€ | 4,8 m€ | ↑ 43% |
| Non vie hors santé | 0,0 m€ | 0,0 m€ | 0,0 m€ | 0,0 m€ | | → 0% |
| Santé similaire à la non vie | 18,1 m€ | 18,1 m€ | 11,1 m€ | 15,9 m€ | 4,8 m€ | ↑ 43% |
| Vie et santé similaire à la vie, hors UC ou indexés | 53,9 m€ | 53,9 m€ | 52,8 m€ | 52,8 m€ | | → 0% |
| Santé similaire à la vie | 35,4 m€ | 35,4 m€ | 34,4 m€ | 34,4 m€ | | → 0% |
| Vie hors santé, UC ou indexés | 18,5 m€ | 18,5 m€ | 18,4 m€ | 18,4 m€ | | → 0% |
| UC ou indexés | 0,0 m€ | 0,0 m€ | 0,0 m€ | 0,0 m€ | | → 0% |
| Total de l'actif | 830,2 m€ | 830,2 m€ | 897,0 m€ | 900,0 m€ | 3,0 m€ | ↑ 0% |

Tableau XII. Evolution des postes d'actif du bilan prudentiel entre les deux modèles

Nous avons en hypothèse un montant de **4,8 millions** d'euros de PANE cédées pour la garantie santé.

D'un côté, les écarts que nous constatons sur les postes du passif sont dus à l'alimentation des *dettes nées d'opérations de réassurance*. En effet elles sont alimentées par le montant hors PANE du poste *dettes nées d'opérations de réassurance et de cessions en substitution* du bilan comptable renseigné en entrée du modèle. Cet écart génère par conséquent une baisse des impôts différés passifs.

De l'autre côté, les écarts que nous constatons à l'actif sont dus au poste des provisions techniques cédées pour la classification *NSLT* dont le montant de PANE cédée n'a pas été retranché. Cet écart génère par conséquent une baisse des impôts différés actifs qui vient compenser celle des impôts différés passifs.

Ainsi pouvons voir que le retraitement des PANE cédées n'a pas d'impact sur les fonds propres disponibles mais modifie la répartition des montants du bilan prudentiel.

Par ailleurs le retraitement des PANE a des impacts sur le calcul du *MCR*. Effectivement le calcul du *MCR* linéaire non-vie fait intervenir les montants **nets** de *Best Estimate* ainsi que les primes **émises** nettes qui sont obtenues en retranchant la variation des PANE aux cotisations acquises. Nous pouvons le voir dans la formule suivante :

$$MCR\ NL = \sum_{i=1}^I \alpha_i \times BE\ Nets_i + \beta_i \times Primes\ émises\ nettes_i$$

- i : numéro de la *LoB* considérée
- α_i et β_i : facteurs pour la *LoB* i de la formule standard

Les écarts constatés sur le *MCR* sont présentés dans le tableau suivant :

| Décomposition complète du MCR | Avec retraitement des PANE | Sans retraitement des PANE | Ecarts | |
|-------------------------------|----------------------------|----------------------------|---------------|-----|
| MCR | 33,7m€ | 33,5m€ | -0,2m€ | -1% |
| MCR Non Life | 16,5m€ | 16,2m€ | -0,2m€ | -1% |
| MCR Life | 17,3m€ | 17,3m€ | | 0% |
| AMCR | 3,7m€ | 3,7m€ | | 0% |

Tableau XIII. Evolution des montants de *MCR* entre les deux modèles

Synthèse de la partie 2.3

Nous avons abordé dans cette partie deux développements effectués à l'aide des méthodes de cartographie évoquée précédemment. La synthèse des travaux menés est effectuée dans le tableau suivant :

| Caractéristiques | Calcul de la PRE | Réorganisation des calculs de <i>Best Estimates NSLT</i> |
|--------------------------------------|--|---|
| <u>Carte utilisée</u> | Carte des interactions (2.2.1) | Carte de la hiérarchie des calculs (2.2.2) |
| <u>Problèmes détectés</u> | Interactions incohérentes entre les parties du modèle Prévoyance Santé | Manque de lisibilité dans les calculs du module |
| <u>Concepts des graphes utilisés</u> | Chemins, Descendants | Chemins, Ancêtres, Hiérarchie |
| <u>Amélioration apportée</u> | Interactions respectant la logique de modélisation actuarielle | Simplification des calculs et réduction des étapes intermédiaires |

Tableau XIV. Tableau de synthèse des développements effectués à l'aide des cartes

Synthèse de la cartographie par les dépendances informatiques

La théorie des graphes appliquée aux données de dépendances informatiques nous a permis de définir deux méthodes pour analyser les interactions entre les variables du modèle Prévoyance Santé. Ces méthodes ont ainsi permis des développements qui ont amélioré l'organisation des calculs effectués. Ils contribuent donc à faciliter l'analyse des résultats fournis par le modèle et à la maîtrise des risques modélisés. Nous avons ainsi pu mettre en place un outil complétant le processus de validation des modèles abordé dans la première partie.

Cependant cette méthode d'analyse des modèles actuariels comporte des limites.

Pour la carte des interactions (2.2.1), seuls les modules qui se situent au premier niveau de l'architecture du modèle Prévoyance Santé sont utilisés pour effectuer les regroupements des nœuds. Par exemple les modules communs de la partie passif n'apparaissent pas sur le graphe.

La relation de dépendance informatique entre deux variables du modèle peut s'avérer contre-intuitive. Il est tout à fait possible qu'une variable intervienne dans la formule d'une autre sans pour autant intervenir dans son calcul. Par conséquent nos cartographies ont tendance à être surchargée en liens. C'est le jugement d'un expert de la modélisation actuarielle qui permet de déterminer si le lien entre deux variables est pertinent ou non.

Pour combler ce dernier problème, nous allons essayer de mesurer les relations entre les variables apparaissant sur nos graphes.

3. Approche statistique de la cartographie

Jusqu'à maintenant, nous avons considéré les « dépendances » entre les variables sur le plan informatique. En effet une variable dépendait d'une autre si elle intervenait dans le code permettant de la calculer. A présent, nous considérons que les variables calculées par le modèle Prévoyance Santé sont aléatoires et nous nous intéressons aux relations de dépendances conditionnelles. Nous allons ainsi évoquer une nouvelle approche basée sur les réseaux bayésiens pour cartographier le modèle.

Cette partie a été rédigée à l'aide du livre *Réseaux bayésiens avec R* (Denis, et al., 2015).

3.1. Présentation des réseaux bayésiens

L'intérêt des réseaux bayésiens pour notre cartographie est la représentation graphique simple des dépendances entre les variables aléatoires à l'aide du *DAG*. Cette représentation devrait permettre une compréhension intuitive et qualitative du modèle Prévoyance Santé.

3.1.1. Définition d'un réseau bayésien

Un réseau bayésien définit la loi conjointe d'un couple de variables aléatoires à l'aide des lois univariées (marginales ou conditionnelles) de chacune des variables aléatoires du couple. En effet, il s'agit d'un un modèle graphique probabiliste composé de deux aspects :

- Un couple de n variables aléatoires $X = (X_1, \dots, X_n)$. Pour rappel, la loi du couple X est appelé la loi conjointe. La loi d'une variable aléatoire X_i est appelée la loi marginale. La loi de X_i sachant X_j est appelée la loi conditionnelle de X_i sachant X_j . Ces trois lois sont distinctes ;

- Un *DAG* (*Directed acyclic graph*) qui signifie en français graphe acyclique orienté. Chaque nœud de ce graphe est associé à une variable aléatoire du couple X . Par la suite nous confondrons la variable aléatoire X_i et son nœud sur le *DAG*. Les liens entre les nœuds représentent les dépendances probabilistes entre les variables aléatoires du modèle.

Nous allons maintenant détailler et préciser les propriétés engendrées par ces deux aspects.

3.1.2. Dépendances probabilistes

Tout d'abord nous présentons la notion de d-séparation qui permet de retranscrire les relations de dépendances entre les variables aléatoires d'un réseau bayésien à partir du *DAG*.

Définition de la d-séparation : Soient A , B et C trois nœuds d'un *DAG* G . On dit que A et B sont d-séparés par C lorsque pour tous les chemins reliant A à B , une des conditions suivantes est vérifiée :

- le chemin entre A et B passe par C et forme soit une connexion en série ($A \rightarrow C \rightarrow B$) soit une connexion divergente ($A \leftarrow C \rightarrow B$) ;
- le chemin entre A et B converge en un nœud D ($A \rightarrow D \leftarrow B$), avec D différent de C et qui n'est pas lié directement à C .

Pour le *DAG* d'un réseau bayésien, si A et B sont d-séparés par rapport à C alors les variables aléatoires A et B sont indépendantes conditionnellement à C .

Exemple : Considérons un couple de variables aléatoires $X = (X_1, \dots, X_6)$ dont le *DAG* est le suivant :

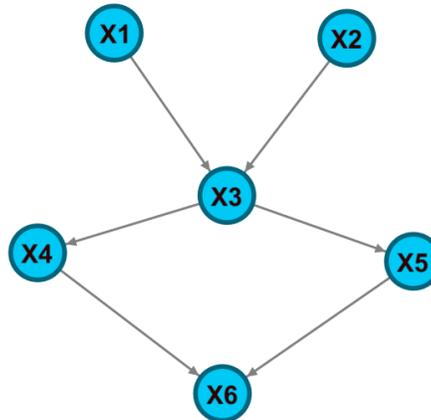


Figure 30. Exemple de DAG

Nous voulons savoir à partir de ce *DAG* si X_5 est indépendante de X_2 conditionnellement à X_3 . Le premier chemin reliant X_2 à X_5 est le suivant ($X_2 \rightarrow X_3 \rightarrow X_5$). Ce chemin passe par X_3 et forme une connexion en série. De même pour le chemin ($X_2 \rightarrow X_3 \rightarrow X_4 \rightarrow X_6 \leftarrow X_5$). Par conséquent la première condition de la d-séparation est vérifiée, ce qui signifie que X_2 et X_5 sont d-séparées par X_3 et cela implique donc que X_5 est indépendante de X_2 sachant X_3 .

Nous voulons maintenant savoir si X_6 est indépendante de X_2 sachant X_4 . Le chemin ($X_2 \rightarrow X_3 \rightarrow X_4 \rightarrow X_6$) vérifie la première condition de la d-séparation. Cependant le deuxième chemin ($X_2 \rightarrow X_3 \rightarrow X_5 \rightarrow X_6$) ne vérifie aucune des deux conditions. Par conséquent la variable X_6 dépend de X_2 conditionnellement à X_4 . Cela n'implique pas forcément que X_2 est la cause d'un effet sur X_6 lorsque la valeur de X_4 est connue. C'est uniquement l'avis de la personne experte sur les quantités modélisées qui permet de déterminer les relations causales.

Un autre résultat similaire est celui de la propriété de Markov locale.

Propriété de Markov locale : Chaque nœud X_i est indépendant de ses non-descendants¹⁷ conditionnellement à ses parents¹⁸.

Cette propriété permet aussi d'associer les dépendances conditionnelles entre les variables aléatoires du couple et le *DAG*.

3.1.3. Loi jointe d'un réseau bayésien

Nous avons évoqué précédemment qu'un réseau bayésien utilisait les lois univariées (marginales et conditionnelles) de chaque variable pour obtenir la loi jointe du couple X . En effet la loi jointe de X est obtenue par le théorème de factorisation qui repose sur la propriété de Markov locale.

Théorème de factorisation : La distribution de la loi jointe du couple X peut s'écrire :

$$P(X) = \prod_{i=1}^n P(X_i | Parents(X_i))$$

Avec $Parents(X_i)$: l'ensemble des parents de X_i .

Ainsi les variables aléatoires qui sont associées à des nœuds racines¹⁹ se voient attribuer une loi marginale alors que les variables aléatoires possédant des parents se voient attribuer une loi conditionnelle à leurs parents.

Cette factorisation propre aux réseaux bayésiens permet de réduire grandement le nombre de paramètres à estimer pour déterminer la loi jointe.

¹⁷ Non-descendant d'un nœud X_i : tous les nœuds que l'on ne peut pas atteindre depuis X_i en suivant les chemins orientés du *DAG*.

¹⁸ Parents d'un nœud X_i : tous les nœuds ayant un lien dirigé vers X_i .

¹⁹ Nœud racine : nœud sans parent.

3.2. Apprentissage d'un réseau bayésien

L'apprentissage d'un réseau bayésien consiste en deux étapes distinctes :

- l'apprentissage de la structure qui consiste à déterminer le *DAG* associé au couple de variables aléatoires ;
- l'estimation des paramètres des lois intervenant dans la loi conjointe sous la forme factorisée qui a été déduite du *DAG*.

Nous ne présenterons uniquement que la première étape dans le cas des réseaux bayésiens gaussiens car nous nous intéressons en priorité au *DAG* pour notre cartographie.

3.2.1. Présentation des réseaux bayésiens gaussiens

Nous considérons que les variables aléatoires du couple X sont continues. Dans le cadre des réseaux bayésiens gaussiens linéaires, nous supposons que la loi jointe de X est une loi normale multivariée $N(\mu, \Sigma)$ dont la densité est :

$$f(x) = (2\pi)^{-\frac{n}{2}} \cdot |\Sigma|^{-\frac{1}{2}} \cdot \exp\left[-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu)\right]$$

Avec :

T : la transposée

μ : le vecteur moyenne de dimension n ;

Σ : la matrice de variance covariance de X ;

$|\Sigma|$: le déterminant de Σ .

Par ailleurs nous supposons que l'effet du conditionnement des parents sur les variables aléatoires se traduit par une fonction affine sur l'espérance et que la variance reste constante. Notons Y une variable aléatoire du couple X et X_1, \dots, X_k les parents de Y .

Par conséquent la loi de Y sachant X_1, \dots, X_k est donnée par :

$$Loi(Y | X_1 = x_1, \dots, X_k = x_k) = N(\beta^T x, \sigma^2)$$

Avec :

- $\beta = (\beta_0, \beta_1, \dots, \beta_k)^T$: vecteur des coefficients associés à chaque parent X_1, \dots, X_k ;
- $x = (1, x_1, \dots, x_k)^T$: vecteur des valeurs prises par chaque parent X_1, \dots, X_k ;
- σ^2 : variance de la loi conditionnelle de Y .

Ainsi nous supposons que les variables aléatoires du réseau bayésien sont des régressions linéaires de leurs parents sur le *DAG*, ce qui peut s'écrire :

$$Y = \beta^T x + \varepsilon$$

avec ε : terme d'erreur de loi $N(0, \sigma^2)$.

3.2.2. Apprentissage de la structure

Maintenant que le modèle est spécifié nous allons présenter les différentes approches possibles pour déterminer le *DAG*.

Une première approche consiste à construire le *DAG* à partir des connaissances d'un expert sur le domaine. Cependant cette méthode ne tient pas compte des données disponibles et devient difficile à mettre en œuvre lorsque le nombre de variables aléatoires devient conséquent.

Une deuxième approche consiste à utiliser sur nos données des algorithmes spécifiques pour la recherche du *DAG*. Il en existe deux sortes :

- algorithmes par contraintes : utilisation de tests d'indépendance conditionnelles ;
- algorithmes par maximisation de score : optimisation d'un score de réseau.

Nous allons présenter les principes de ces deux familles d'algorithmes.

3.2.2.1. Les algorithmes par contraintes

Cette famille se base sur l'algorithme IC de causalité inductive introduit par Verma et Pearl en 1991. Il consiste à établir les liens entre les nœuds à l'aide de tests statistiques sur les corrélations conditionnelles.

Le test de corrélation conditionnelle le plus utilisé est le test du coefficient de corrélation linéaire partiel dont nous détaillons le principe.

Test du coefficient de corrélation linéaire partiel

Soient X , Y et Z trois variables aléatoires, nous voulons savoir s'il existe une relation linéaire entre X et Y sachant Z . Cela revient donc à tester :

$H_0 : \rho_{XY|Z} = 0$ contre l'hypothèse alternative **$H_1 : \rho_{XY|Z} \neq 0$**

où $\rho_{XY|Z}$ est le coefficient de corrélation partielle de X et Y sachant Z .

Pour appliquer le test, il est nécessaire que l'échantillon $(x_i, y_i, z_i)_{i=1, \dots, n}$ soit des réalisations indépendantes du couple de variables aléatoires (X, Y, Z) et que (X, Y, Z) suive une loi normale multivariée. Pour vérifier la deuxième condition d'application, il est possible d'utiliser un test de normalité multivarié comme celui de Royston. Il s'agit d'une généralisation du test de Shapiro Wilk pour des échantillons multivariés. Ce test est disponible dans le package *R mvnortest*.

Nous avons alors sous l'hypothèse nulle H_0 , la statistique de test :

$$t = \widehat{\rho}_{X,Y|Z} \sqrt{\frac{n-3}{1 - \widehat{\rho}_{X,Y|Z}^2}}$$

qui suit une loi de Student avec $n-3$ degrés de liberté, où n est la taille de l'échantillon.

Nous comparons donc la statistique de test observée notée t_{obs} à la valeur critique du test, notée c_α pour une loi de Student à $n-3$ degrés de liberté.

Si $|t_{\text{obs}}| > c_\alpha$, alors le test est significatif. Nous rejetons l'hypothèse H_0 et nous décidons que H_1 est vraie en commettant une erreur de première espèce α .

Dans le cas contraire, le test n'est pas significatif. Nous décidons de conserver l'hypothèse H_0 et nous commettons une erreur de deuxième espèce β .

Le nombre de tests de corrélation linéaire à effectuer est a priori très élevé étant donné le nombre important de relations d'indépendance conditionnelle possibles dans un *DAG*. Pour diminuer ce nombre de tests, les algorithmes par contraintes effectuent un apprentissage préalable de la structure pour diminuer le nombre de tests à effectuer.

Exemples d'algorithmes par contraintes : *Grow-Shrink* (2003), *Incremental Association Markov Blanket* (2003), *Semi-Interleaved Hiton-PC* (2010)

3.2.2.2. Les algorithmes par maximisation du score

Ces algorithmes associent à chaque *DAG* un score de réseau qui traduit sa bonne adéquation aux données disponibles. Le score le plus souvent employé pour les réseaux bayésiens gaussiens est le *BIC*²⁰. Ils cherchent donc à trouver le *DAG* pour lequel le score de réseau est le plus élevé.

La méthode employée pour l'optimisation est une recherche dite « gloutonne » qui procède de la façon suivante :

- i. Poser un graphe initial G à partir des nœuds V et calculer son score. Bien souvent le graphe initial est le graphe vide sans lien.
- ii. Répéter les étapes suivantes tant que le score du graphe continue de croître :
 - a. Ajouter, retirer ou inverser un lien du graphe G . Nous obtenons un graphe G^* .
 - b. Calculer le score de G^* .
 - c. Si le score de G^* est supérieur à celui de G , alors G devient G^* .

Exemple d'algorithmes par maximisation de score : *Hill-Climbing* et *Tabu search* (1995).

3.3. Application pour le modèle standard

Pour notre étude nous allons considérer à nouveau l'organisme de prévoyance santé présenté dans la partie 2.3. En plus des hypothèses au 31/12/2017, nous disposons de celles concernant la projection de son activité dans le temps. Pour obtenir un jeu de données afin d'appliquer les algorithmes d'apprentissage de la structure, nous allons ajouter un bruit blanc gaussien aux principales hypothèses de projection. Nous vérifierons par la suite que les hypothèses nécessaires à l'application de ces algorithmes sont bien vérifiées. Enfin les

²⁰ *BIC* : *Bayesian Information Criterion*

conclusions que nous tirerons de ces vérifications nous amènerons à proposer une autre méthode pour considérer les dépendances entre nos variables.

3.3.1. Mise en place de la démarche

Nous présentons dans cette section la démarche devant nous permettre d'établir un *DAG* du modèle prévoyance santé. Pour ce faire nous considérerons les principales hypothèses de projection de notre organisme et les sorties majeures de notre modèle. Des simulations seront effectuées pour obtenir différents jeux d'hypothèses. Nous vérifierons ensuite que les variables calculées par le modèle vérifient les conditions de normalité.

3.3.1.1. Présentation des variables étudiées

Nous allons considérer les hypothèses suivantes pour notre étude :

- ratio S/P utilisé pour la projection du business plan (pour chaque segment) ;
- ratio S/P utilisé pour la projection du *Best Estimate* de prime (pour chaque segment) ;
- rendement moyen des actions ;
- rendement moyen de l'immobilier ;
- taux de cession en réassurance ;
- taux de rémunération des dépôts ;
- courbe des taux ;
- taux de frais d'acquisition, d'administration et de gestion (pour chaque segment) ;
- taux de renouvellement des cotisants (pour chaque segment);
- taux évolution des cotisations individuelles (pour chaque segment) ;
- taux d'affaires nouvelles signées avant clôture (pour chaque segment) ;
- taux d'affaires nouvelles signées après la clôture (pour chaque segment).

Nous considérons ensuite les variables de sortie suivantes :

- SCR sous modules et le SCR ;
- fonds propres éligibles pour la couverture du SCR ;
- provisions techniques cédées en vision *Best Estimate* ;
- placements en valeur de marché ;
- résultat de l'exercice comptable ;
- cotisations des comptes de résultat projetés.

Les hypothèses et les sorties retenues se veulent représentatives de la projection de l'entité. Elles ont été choisies en amont par le jugement d'un expert.

Les paramètres des lois normales utilisées pour les simulations sont déterminés de la façon suivante :

- la moyenne : valeur de l'hypothèse central de l'organisme ;
- l'écart-type : déterminé pour chaque hypothèse par un avis d'expert.

Nous lançons ainsi le modèle prévoyance santé 1000 fois sur une horizon d'un an pour chacun des jeux d'hypothèses. Nous obtenons ainsi un jeu de données composée de 1000 lignes et 82 colonnes. Nous allons maintenant analyser les résultats de l'organisme sur une horizon d'un an.

3.3.1.2. Vérification des conditions de normalité

Nous avons vu précédemment dans la section 3.2.1 que les variables aléatoires intervenant dans le réseau bayésien gaussien devaient être normales, notamment pour l'application des algorithmes d'apprentissage de la structure.

Pour le vérifier nous effectuons dans un premier temps une analyse graphique par un *QQ Plot*. Ce graphe permet de comparer la loi empirique de l'échantillon avec la loi normale. Cela consiste à tracer le nuage de points $(x_{(i)}, F^{[-1]}(p_i))$, $i \in \llbracket 1, T \rrbracket$ avec

- $x_{(i)}$: les valeurs ordonnées de l'échantillon ;
- $F^{[-1]}$: la fonction quantile de la loi normale ajustée ;

- $p_i = i/(T+1)$.

Si la loi normale est similaire à la loi empirique de l'échantillon, alors les points sont alignés sur la droite d'équation $y = \hat{\sigma} x + \hat{\mu}$, avec

- $\hat{\sigma}$: l'estimation empirique de l'écart-type de l'échantillon ;
- $\hat{\mu}$: l'estimation empirique de la moyenne de l'échantillon.

Les tracés sont effectués à l'aide du package R *ggplot2* (Wickham, 2018).

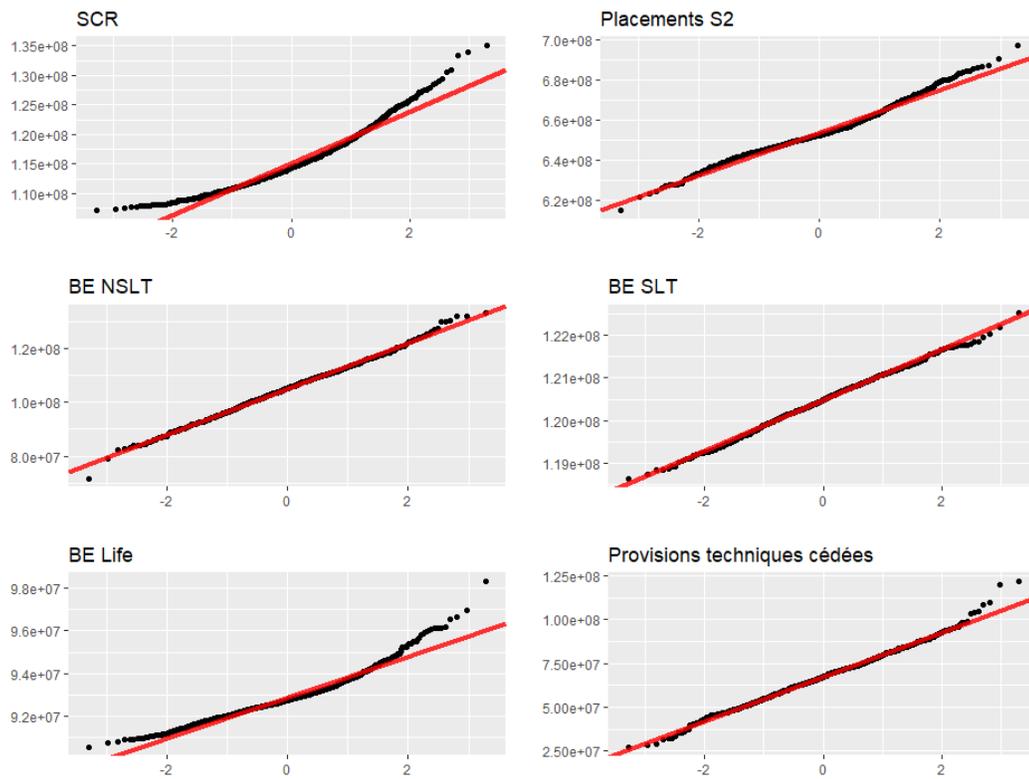


Figure 31. Tracé des diagrammes quantile quantile pour les principales variables

Sur les graphiques nous pouvons voir que l'hypothèse de normalité ne semble pas être satisfaite par l'ensemble des variables. D'un côté, l'adéquation à la loi normale est plutôt satisfaisante pour les *Best Estimates NSLT*, *SLT*, les *Best Estimates* cédés et le montant des placements en vision Solvabilité 2. D'un autre côté le *SCR* et le *Best Estimate Life* ne semblent pas satisfaire l'hypothèse. Nous confirmons cette observation avec le tracé des histogrammes :

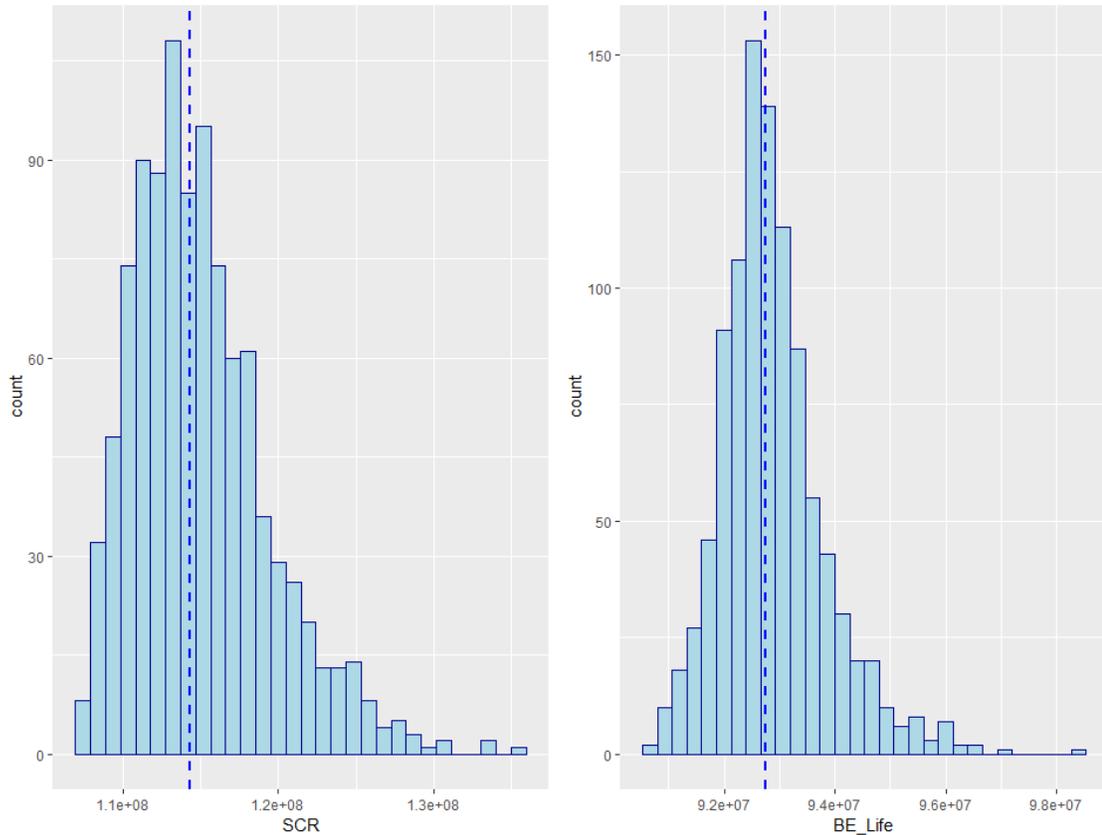


Figure 32. Histogrammes des variables *SCR* et *BE Life*

Les estimations des médianes des échantillons de nos deux variables sont représentées en pointillé bleu. Ainsi nous pouvons voir que les distributions ne sont pas symétriques et plutôt décalées vers la gauche de la médiane. Ces observations vont à l'encontre de la symétrie de la loi normale.

Dans un deuxième temps, nous allons vérifier l'hypothèse de normalité à l'aide du test statistique de Shapiro Wilk (Ferrigno, et al., 2013).

Condition : L'échantillon x_1, \dots, x_n est formé d'au moins 30 réalisations indépendantes de la variable aléatoire X .

L'hypothèse nulle est

$$H_0 : \text{La variable aléatoire } X \text{ suit une loi normale}$$

Contre l'hypothèse alternative

$$H_1 : \text{La variable aléatoire } X \text{ ne suit pas une loi normale}$$

La statistique du test de Shapiro Wilk est la suivante :

$$W = \frac{(\sum_{i=1}^{\lfloor n/2 \rfloor} a_{n-i+1,n} (X_{n-i+1,n} - X_{i,n}))^2}{\sum_{i=1}^T (X_i - \bar{X})^2}$$

- $X_{i,n}$: la ième variable de l'échantillon aléatoire trié ;
- a_i : les coefficients calculés à partir des quantiles d'une loi normale centrée réduite
- \bar{X} : l'estimateur de la moyenne de X

Sous l'hypothèse H_0 la statistique W suit une loi de Shapiro Wilk.

Lorsque $n > 50$, si la p valeur est inférieure ou égale au seuil α alors le test est significatif et nous rejetons l'hypothèse nulle H_0 . Nous commettons alors une erreur de première espèce α .

Au contraire, si la p valeur est strictement supérieure au seuil α alors nous décidons de conserver l'hypothèse H_0 . Nous commettons alors une erreur de deuxième espèce β .

Nous effectuons les tests de Shapiro Wilk à l'aide du package R *MVN* (Selcuk, 2018). Nous obtenons les sorties suivantes :

| Test | Variable | Statistic | p value | Normality |
|--------------|---------------------------------|-----------|---------|-----------|
| Shapiro-Wilk | SCR | 0.9461 | <0.001 | NO |
| Shapiro-Wilk | Fonds_Propres_Eligibles | 0.1787 | <0.001 | NO |
| Shapiro-Wilk | Placements_S2 | 0.9792 | <0.001 | NO |
| Shapiro-Wilk | Provisions_techniques_cedees_S2 | 0.9952 | 0.0032 | NO |
| Shapiro-Wilk | Total_Actif_S2 | 0.9941 | 0.0005 | NO |
| Shapiro-Wilk | Total_Passif_S2 | 0.2100 | <0.001 | NO |
| Shapiro-Wilk | BE_NSLT | 0.9981 | 0.3304 | YES |
| Shapiro-Wilk | BE_SLT | 0.9980 | 0.2721 | YES |
| Shapiro-Wilk | BE_Life | 0.9505 | <0.001 | NO |
| Shapiro-Wilk | SCR_Mkt_Int | 0.8916 | <0.001 | NO |
| Shapiro-Wilk | SCR_Mkt_Eq | 0.9990 | 0.8501 | YES |
| Shapiro-Wilk | SCR_Mkt_Prop | 0.9990 | 0.8893 | YES |
| Shapiro-Wilk | SCR_Mkt_Currency | 0.9481 | <0.001 | NO |
| Shapiro-Wilk | SCR_Mkt_Spread | 0.9093 | <0.001 | NO |
| Shapiro-Wilk | SCR_Mkt_Conc | 0.9879 | <0.001 | NO |
| Shapiro-Wilk | SCR_Default_Type_1 | 0.2617 | <0.001 | NO |
| Shapiro-Wilk | SCR_Default_Type_2 | 0.8324 | <0.001 | NO |
| Shapiro-Wilk | SCR_Life_Mortalite | 0.9988 | 0.7379 | YES |
| Shapiro-Wilk | SCR_Life_Longevite | 0.9987 | 0.6839 | YES |
| Shapiro-Wilk | SCR_Life_Morbidity | 0.9982 | 0.3762 | YES |
| Shapiro-Wilk | SCR_Life_Frais | 0.9979 | 0.2402 | YES |
| Shapiro-Wilk | SCR_Life_Revision | 0.9985 | 0.5863 | YES |
| Shapiro-Wilk | SCR_Life_CAT | 0.9125 | <0.001 | NO |
| Shapiro-Wilk | SCR_HSLT_Longevite | 0.9986 | 0.6125 | YES |
| Shapiro-Wilk | SCR_HSLT_Frais | 0.9981 | 0.3292 | YES |
| Shapiro-Wilk | SCR_HSLT_Revision | 0.9983 | 0.4467 | YES |
| Shapiro-Wilk | SCR_H_Primes_Reserve | 0.9105 | <0.001 | NO |
| Shapiro-Wilk | SCR_Health_CAT | 0.9833 | <0.001 | NO |
| Shapiro-Wilk | SCR_Operational_Risk | 0.9934 | 0.0002 | NO |
| Shapiro-Wilk | Ajustement_FDB | 0.9424 | <0.001 | NO |
| Shapiro-Wilk | Resultat_Exercice | 0.9815 | <0.001 | NO |
| Shapiro-Wilk | D1_Cotisations_Acquises_Brutes | 0.9931 | 0.0001 | NO |
| Shapiro-Wilk | E1_Cotisations | 0.8904 | <0.001 | NO |

Tableau XV. Tableau des sorties des tests de Shapiro Wilk

La colonne *Normality* permet de déterminer si le test est significatif au risque d'erreur $\alpha=5\%$.

L'application des tests de Shapiro Wilk nous conduit à rejeter l'hypothèse H_0 de normalité pour 21 des 33 variables aléatoires en sortie du modèle. Au vue des analyses graphiques et statistiques que nous avons effectuées, nous ne pouvons pas considérer que l'hypothèse de normalité est satisfaite.

Nous proposons dans la suite de l'étude une alternative non-paramétrique.

3.3.2. Alternative non paramétrique

L'hypothèse de normalité des réseaux bayésiens n'a pas pu être vérifiée à partir de nos données. Nous ne mesurerons donc pas la dépendance probabiliste mais simplement la corrélation. Nous allons dorénavant mener des analyses qui ne supposent pas la normalité. Nous présenterons dans un premier temps le coefficient de corrélation de Spearman. Puis dans un deuxième temps nous verrons comment les calculs de ces coefficients pourront compléter notre cartographie informatique.

3.3.2.1. Coefficient de corrélation de Spearman

Le coefficient de corrélation de Spearman consiste à appliquer le coefficient de corrélation de Pearson aux rangs des variables aléatoires. (Rakotomalala, 2017) La formule de son estimateur est la suivante :

$$\widehat{\rho}_{X,Y} = \frac{\sum_{i=1}^n (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2} \sqrt{\sum_{i=1}^n (S_i - \bar{S})^2}}$$

Avec :

- R_i : rang de la variable aléatoire X_i ;
- S_i : rang de la variable aléatoire Y_i ;

À l'aide de propriétés sur le rang nous avons l'expression suivante de l'estimateur en posant $D_i = R_i - S_i$:

$$\widehat{\rho}_{X,Y} = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n(n^2 - 1)}$$

Le coefficient de corrélation de Spearman est compris entre -1 et 1. Une valeur proche de 1 signifiera que la relation entre les variables est croissante alors qu'une valeur proche de -1 signifiera que la relation est décroissante.

L'avantage principal du coefficient de corrélation de Spearman pour notre étude est qu'il est non-paramétrique, à l'inverse de celui de Pearson. Ainsi pour l'appliquer nous n'avons plus besoin de l'hypothèse de normalité qui n'est pas vérifiée avec nos données.

Par ailleurs le coefficient permet de mesurer la relation monotone entre deux variables alors que celui de Pearson n'est limité qu'à la relation linéaire. Or nous savons à l'avance que certains liens entre nos variables ne sont pas linéaires. C'est notamment le cas pour la relation entre la courbe des taux et les *Best Estimates* : une hausse des taux n'entraîne pas une diminution linéaire de la provision.

Par la suite les estimations des coefficients de corrélation de Spearman sont données à titre indicatif, de manière analogue à des statistiques descriptives. En effet les valeurs des coefficients dépendent grandement des caractéristiques de l'organisme modélisé lors de l'année de projection initiale. Il n'est donc pas possible de tirer à l'aide des valeurs de conclusion générale sur les relations entre les différentes variables modélisées.

3.3.2.2. Pondération des liens de la cartographie informatique

L'objectif de cette section est de compléter la cartographie informatique du modèle prévoyance santé en pondérant les liens par les coefficients de corrélation de Spearman calculés à partir de nos simulations.

À l'aide de la carte des interactions informatiques du modèle prévoyance santé que nous avons établie, nous construisons le graphe associé aux variables de notre étude. Cet étape est réalisée à l'aide du package R *igraph* (Csardi, 2015).

Nous calculons ensuite pour les couples de variables présents dans l'ensemble des liens les estimations des coefficients de corrélation de Spearman. Ce calcul est effectué à l'aide du package *stats* (R Core Team, 2018).

Enfin les coefficients de corrélation sont importés dans le logiciel Gephi. Nous obtenons le graphe suivant dans un premier temps sans appliquer de filtre de couleur :

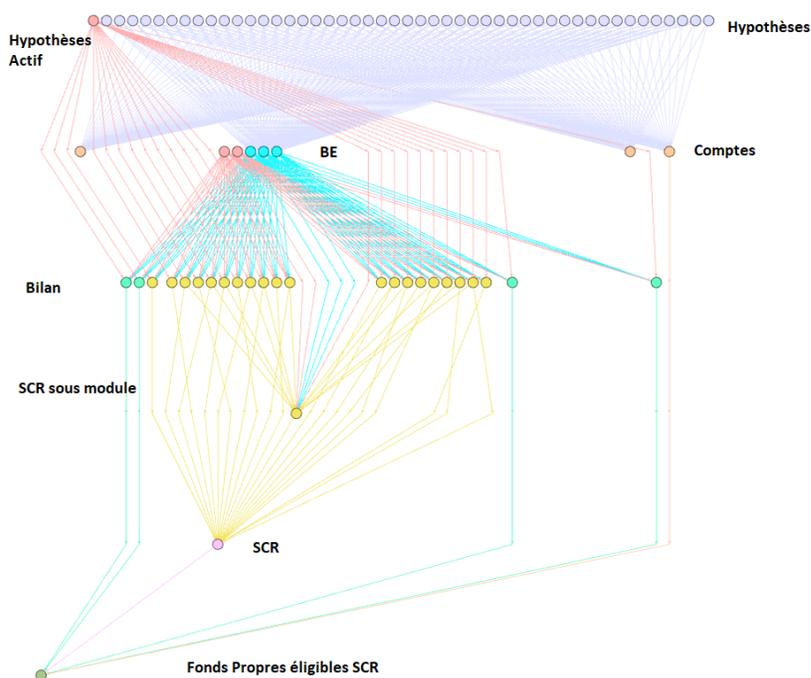


Figure 33. Graphe du modèle de l'étude obtenues par la cartographie informatique

Nous pouvons voir que les liens présents sur le graphe sont très nombreux. En effet certaines relations de dépendance informatique sont trompeuses. Par exemple nous observons un lien entre le *Best Estimate SLT* et le *SCR* immobilier. Or dans le modèle prévoyance santé nous savons que les *Best Estimates* n'interviennent pas dans le calcul du *SCR* immobilier qui ne dépend que du portefeuille d'immobilier détenus par l'organisme.



Figure 34. Lien de dépendance informatique entre le *BE SLT* et le *SCR* immobilier

Pour améliorer la pertinence des liens représentés, nous colorons les liens en fonction de la valeur de l'estimation du coefficient de corrélation de Spearman.

- Plus la couleur du lien est rouge, plus l'estimation du coefficient est proche de 1 ;
 - Plus la couleur du lien est blanche, plus l'estimation du coefficient est proche de 0 ;
 - Plus la couleur du lien est noir, plus l'estimation du coefficient est proche de -1.
- Nous obtenons alors le graphe suivant :

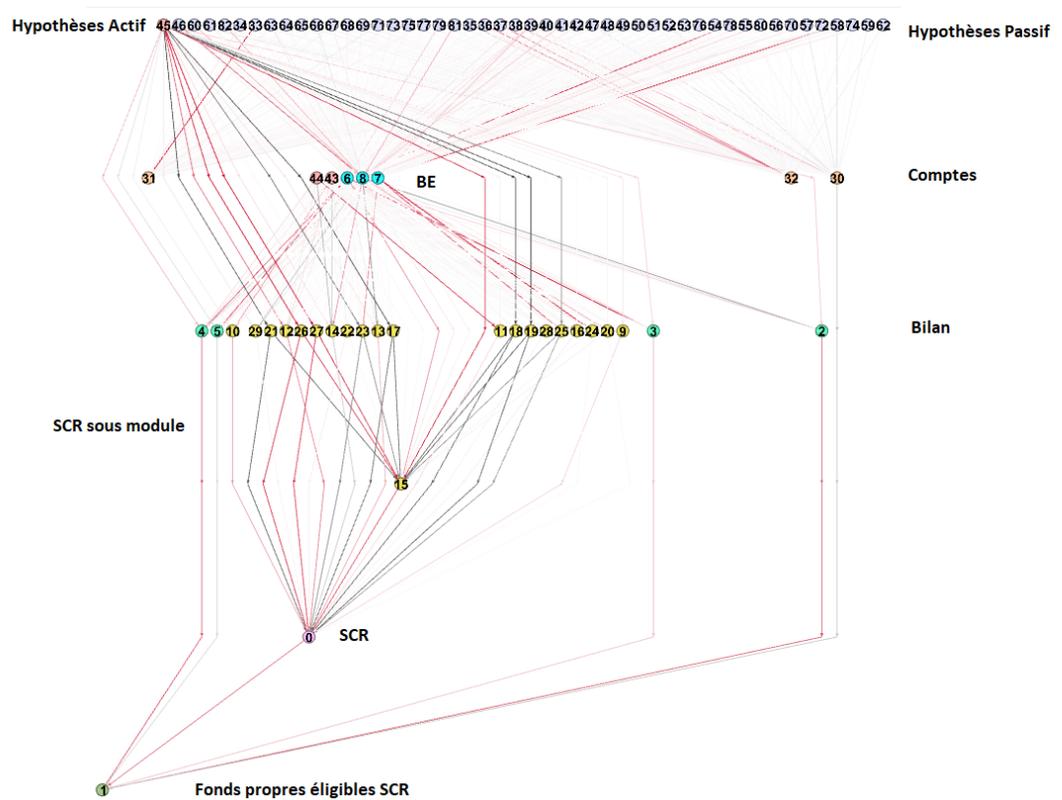


Figure 35. Graphe obtenu avec la coloration des liens par l'estimation du coefficient de corrélation

La correspondance entre les numéros du graphe et les variables est donnée dans l'annexe B. Nous observons une grande diminution du nombre de liens visibles grâce à la coloration. Notre graphe devient ainsi beaucoup plus lisible et facile à appréhender.

Pour notre exemple précédent, le lien entre le *Best Estimate SLT* et le *SCR* immobilier n'est plus visible. En effet nous obtenons une valeur de 0,0092 pour $\hat{\rho}$, très proche de 0.

Nous poursuivons notre analyse en considérant le lien entre le *Best Estimate Life* et le *SCR* de spread. À nouveau, le *Best Estimate Life* n'intervient pas dans le calcul de ce *SCR* sous-module. Cependant nous obtenons une valeur de -0,6186 pour $\hat{\rho}$. Nous disposons ainsi d'un exemple où le coefficient de corrélation ne traduit pas l'absence de relation entre les deux variables.

Nous considérons ensuite l'hypothèse du ratio S/P de la santé utilisé pour la projection du *Best Estimate* de prime. Nous obtenons une valeur de 0,9172 pour $\hat{\rho}$. En effet dans le cas de la santé, ce ratio est multiplié par les cotisations engagées pour déterminer le montant de sinistre futur entrant dans le périmètre du *Best Estimate* de primes. La corrélation obtenue est donc cohérente avec la modélisation effectuée.

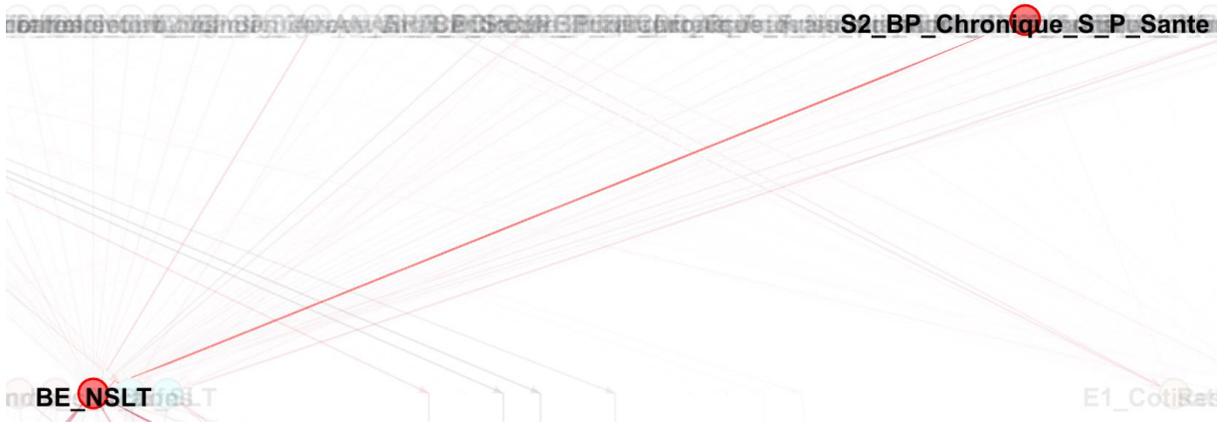


Figure 36. Lien entre l'hypothèse de ratio S/P S2 en santé et le *Best Estimate NSLT*

Les estimations des coefficients obtenues pour les liaisons entre les hypothèses et les *Best Estimates* sont cohérents. En effet les hypothèses semblent disposer d'une corrélation plus élevée pour les montants calculés correspondant à leur segment. Par exemple l'estimation de la corrélation entre le ratio S/P pour le segment rente de conjoint et le *Best Estimate Life* est de 0,2101 alors qu'elle est de -0,0151 avec le *Best Estimate SLT*.

De même pour le taux de cession qui semble être corrélé avec les montants de *SCR*. En effet toute chose égale par ailleurs, nous savons que la cession entraîne une baisse des *SCR* sous module de souscription. Nous avons par exemple une estimation de la corrélation entre le taux de cession et le *SCR Life* mortalité de 0,9971. Cependant nous observons une estimation de

0,5809 entre le taux de cession et le *SCR* marché pour le risque de change alors que l'hypothèse n'intervient à aucun moment dans le calcul du *SCR*.

Pour les autres hypothèses de projection de l'actif comme le rendement des actions ou de l'immobilier, les estimations des corrélations sont élevées avec les *SCR* sous module de marché ce qui est bien cohérent. Nous avons par exemple une estimation de la corrélation de 0,3697 entre le *SCR* pour le risque action et le rendement des actions.

Enfin les estimations des corrélations entre les postes du bilan et les fonds propres éligibles pour la couverture du *SCR* sont cohérents avec les interactions entre les comptes.

Synthèse de l'approche statistique

Afin de compléter l'approche développée dans la partie précédente, nous avons essayé d'appliquer les algorithmes des réseaux bayésiens à la construction d'un DAG du modèle prévoyance santé. Etant donné que les hypothèses de normalité n'étaient pas vérifiées pour l'ensemble de nos variables, nous avons décidé de nous tourner vers une approche non paramétrique utilisant le coefficient de corrélation de Spearman. Les estimations des coefficients viennent ainsi pondérer les liens du graphe obtenus par l'exploitation des dépendances informatiques. Ces estimations des coefficients ont l'avantage d'être faciles à mettre en œuvre et permettent de facilement détecter les liaisons entre les variables de notre modèle. Dans l'ensemble les estimations sont cohérentes avec les calculs effectuées dans le modèle prévoyance santé. Cependant, certaines estimations des corrélations restent élevées alors que les variables sont calculées séparément. Il est donc nécessaire de garder un sens critique à la lecture des graphes.

Conclusion

La réforme Solvabilité 2 a modifié en profondeur la modélisation actuarielle de la solvabilité des organismes d'assurance. La complexité des calculs demandés nécessite de disposer de processus de validation permettant d'assurer à chaque instant la conformité des modèles actuariels.

Afin de compléter les contrôles déjà effectués, nous avons mis en place un outil d'analyse des interactions entre les variables des modèles actuariels en exploitant les données de dépendances informatiques. À l'aide de la théorie des graphes nous avons élaboré deux méthodologies pour analyser les modèles actuariels : la carte des interactions et la carte de la hiérarchie des calculs. Ces méthodes nous ont apporté un nouveau regard sur l'organisation des calculs. En effet l'aspect visuel nous permet de facilement détecter les anomalies et les améliorations à effectuer dans la modélisation. Ces cartes ont ainsi permis d'effectuer divers développements sur le modèle prévoyance santé.

Pour enrichir cette analyse, nous souhaitons utiliser les réseaux bayésiens et leurs algorithmes d'apprentissage de la structure pour représenter les dépendances probabilistes entre les variables de nos modèles. Cependant l'examen de notre jeu de données nous a conduit à rejeter l'hypothèse de normalité. Par conséquent nous nous sommes tournés vers des méthodes non-paramétriques comme le coefficient de corrélation de Spearman. Les estimations de ces coefficients sont ainsi venues compléter les cartes de dépendance informatique précédemment établies.

Pour s'affranchir de cette hypothèse de normalité, il aurait été envisageable de considérer des tests de corrélation par permutation dans les algorithmes d'apprentissage de la structure. Ces tests utilisant le ré-échantillonnage permettent en effet d'obtenir directement la distribution de la statistique de test envisagée. Cependant les temps de calcul auraient été bien plus élevés par rapport à la méthode retenue.

Une autre alternative aurait consisté à appliquer des transformations comme celles de Box Cox aux données ne satisfaisant pas l'hypothèse. Nous aurions d'un côté gardé des temps de calcul corrects mais d'un autre côté perdu en interprétation de nos résultats.

L'outil ainsi développé dans ce mémoire suit une démarche de rationalisation de la modélisation actuarielle et participe pleinement à la gestion pérenne des risques.

Bibliographie

Barber, D. 2012. *Bayesian Reasoning and Machine Learning*. s.l. : Cambridge University Press, 2012.

Conférence annuelle IA – ACPR du 9 mars 2017. ACPR. 2017. 2017.

Csardi, Gabor. 2015. *Package igraph version 1.2.2*. 2015.

Denis, Jean-Baptiste and Scutari, Marco. 2015. *Réseaux bayésiens avec R*. s.l. : EDP Sciences, 2015.

Eades. 1984. A heuristic for graph drawing. *Congressus Numerantiunt*. 1984, pp. 149-160.

Ferrigno, Sandie, et al. 2013. *Mathématiques pour les sciences de l'ingénieur : tout le cours en fiches : licence, prépas, IUT*. s.l. : Dunod, 2013.

Fruchterman, Thomas M. J. and Reingold, Edward M. 1991. Graph Drawing by Force-directed Placement. *SOFTWARE—PRACTICE AND EXPERIENCE*. 1991, pp. 1129-1164.

Goeman, Jelle J and Solari, Aldo. 2012. *Tutorial in biostatistics: multiple hypothesis testing in genomics*. s.l. : John Wiley & Sons, 2012.

Harrell Jr, Frank. 2018. *Package Hmisc, version 4.1-1*. 2018.

Martin, Shawn, et al. 2011. OpenOrd: An Open-Source Toolbox for Large Graph Layout. *Proceedings of SPIE - The International Society for Optical Engineering*. Janvier 2011.

R Core Team. 2018. *Package stats, version 3.6.0*. 2018.

Rakotomalala, Ricco. 2017. *Analyse de corrélation : Étude des dépendances - Variables quantitatives.* 2017.

Selcuk, Korkmaz. 2018. *package MVN version 5.5.* 2018.

Sugiyama, K, Tagawa, S and Toda, M. 1981. Methods for Visual Understanding of Hierarchical System Structures. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS.* 1981, pp. 109-125.

Tamassia, Roberto. 2013. *Handbook of Graph Drawing and Visualization.* s.l. : CRC Press, 2013. ISBN : 9781138034242.

Wickham, Hadley. 2018. *Package ggplot 2 version 3.1.0.* 2018.

Annexe A

Nous donnons ici le détail du pseudo code de l'algorithme de Fruchterman et Reingold fournit par l'ouvrage (Tamassia, 2013)

```
area :=  $W * L$ ; { $W$  and  $L$  are the width and length of the frame}
 $G := (V, E)$ ; {the vertices are assigned random initial positions}
 $k := \sqrt{\text{area}/|V|}$ ;
function  $f_a(x) := \text{begin return } x^2/k \text{ end}$ ;
function  $f_r(x) := \text{begin return } k^2/x \text{ end}$ ;
for  $i := 1$  to  $\text{iterations}$  do begin
  {calculate repulsive forces}
  for  $v$  in  $V$  do begin
    {each vertex has two vectors:  $.pos$  and  $.disp$ }
     $v.disp := 0$ ;
    for  $u$  in  $V$  do
      if ( $u \neq v$ ) then begin
        { $\delta$  is the difference vector between the positions of the two vertices}
         $\delta := v.pos - u.pos$ ;
         $v.disp := v.disp + (\delta/|\delta|) * f_r(|\delta|)$ 
      end
    end
  end
  {calculate attractive forces}
  for  $e$  in  $E$  do begin
    {each edges is an ordered pair of vertices  $.v$  and  $.u$ }
     $\delta := e.v.pos - e.u.pos$ ;
     $e.v.disp := e.v.disp - (\delta/|\delta|) * f_a(|\delta|)$ ;
     $e.u.disp := e.u.disp + (\delta/|\delta|) * f_a(|\delta|)$ 
  end
  {limit max displacement to temperature  $t$  and prevent from displacement
  outside frame}
  for  $v$  in  $V$  do begin
     $v.pos := v.pos + (v.disp/|v.disp|) * \min(v.disp, t)$ ;
     $v.pos.x := \min(W/2, \max(-W/2, v.pos.x))$ ;
     $v.pos.y := \min(L/2, \max(-L/2, v.pos.y))$ 
  end
  {reduce the temperature as the layout approaches a better configuration}
   $t := \text{cool}(t)$ 
end
```

Figure 37. Pseudo code de l'algorithme de Fruchterman Reingold

Annexe B

Nous donnons la table de correspondance entre les numéros présents sur le graphe de la partie 3.3.2.2 et les noms des variables associées.

| <i>Numéro sur le graphe</i> | <i>Nom de la variable</i> |
|-----------------------------|---|
| 0 | SCR |
| 1 | Fonds_Propres_Eligibles |
| 2 | Placements_S2 |
| 3 | Provisions_techniques_cedees_S2 |
| 4 | Total_Actif_S2 |
| 5 | Total_Passif_S2 |
| 6 | BE_NSLT |
| 7 | BE_SLT |
| 8 | BE_Life |
| 9 | SCR_Mkt_Int |
| 10 | SCR_Mkt_Eq |
| 11 | SCR_Mkt_Prop |
| 12 | SCR_Mkt_Currency |
| 13 | SCR_Mkt_Spread |
| 14 | SCR_Mkt_Conc |
| 15 | SCR_Default_Type_1 |
| 16 | SCR_Default_Type_2 |
| 17 | SCR_Life_Mortalite |
| 18 | SCR_Life_Longevite |
| 19 | SCR_Life_Morbidite |
| 20 | SCR_Life_Frais |
| 21 | SCR_Life_Revision |
| 22 | SCR_Life_CAT |
| 23 | SCR_HSLT_Longevite |
| 24 | SCR_HSLT_Frais |
| 25 | SCR_HSLT_Revision |
| 26 | SCR_H_Primes_Reserve |
| 27 | SCR_Health_CAT |
| 28 | SCR_Operational_Risk |
| 29 | Ajustement_FDB |
| 30 | Resultat_Exercice |
| 31 | D1_Cotisations_Acquises_Brutes |
| 32 | E1_Cotisations |
| 33 | Taux_affaires_nouvelles_apres_cloture_Sante_Indiv |
| 34 | Taux_affaires_nouvelles_apres_cloture_AT_Direct |
| 35 | Taux_affaires_nouvelles_apres_cloture_RE_Direct |
| 36 | Taux_affaires_nouvelles_apres_cloture_RC_Direct |
| 37 | Taux_affaires_nouvelles_apres_cloture_DC_Coll |
| 38 | BP_SP_Sante_Indiv |
| 39 | BP_SP_AT_Direct |
| 40 | BP_SP_RE_Direct |
| 41 | BP_SP_RC_Direct |
| 42 | BP_SP_DC_Coll |
| 43 | Rdmt_Indices_Actions |
| 44 | Indices_Immo_capital |
| 45 | Taux_cession |
| 46 | Decalage_taux |
| 47 | Chronique_Frais_Acquisition_Sante_indiv |
| 48 | Chronique_Frais_Acquisition_AT |
| 49 | Chronique_Frais_Acquisition_RE |
| 50 | Chronique_Frais_Acquisition_RC |
| 51 | Chronique_Frais_Acquisition_DC |
| 52 | Chronique_Frais_Administration_Sante |
| 53 | Chronique_Frais_Administration_AT |
| 54 | Chronique_Frais_Administration_RE |

| | |
|----|---|
| 55 | Chronique_Frais_Administration_RC |
| 56 | Chronique_Frais_Administration_DC |
| 57 | Chronique_Frais_Gestion_Sante |
| 58 | Chronique_Frais_Gestion_AT |
| 59 | Chronique_Frais_Gestion_RE |
| 60 | Chronique_Frais_Gestion_RC |
| 61 | Chronique_Frais_Gestion_DC |
| 62 | Taux_Remu_Depot |
| 63 | Taux_Renouvellement_Sante |
| 64 | Taux_Renouvellement_AT |
| 65 | Taux_Renouvellement_RE |
| 66 | Taux_Renouvellement_RC |
| 67 | Taux_Renouvellement_DC |
| 68 | Taux_Evol_Cot_Indiv_Sante |
| 69 | Taux_Evol_Cot_Indiv_AT |
| 70 | Taux_Evol_Cot_Indiv_RE |
| 71 | Taux_Evol_Cot_Indiv_RC |
| 72 | Taux_Evol_Cot_Indiv_DC |
| 73 | Taux_Affaires_Nouvelles_Signees_Avant_Cloture_Sante |
| 74 | Taux_Affaires_Nouvelles_Signees_Avant_Cloture_AT |
| 75 | Taux_Affaires_Nouvelles_Signees_Avant_Cloture_RE |
| 76 | Taux_Affaires_Nouvelles_Signees_Avant_Cloture_RC |
| 77 | Taux_Affaires_Nouvelles_Signees_Avant_Cloture_DC |
| 78 | S2_BP_Chronique_S_P_Sante |
| 79 | S2_BP_Chronique_S_P_AT |
| 80 | S2_BP_Chronique_S_P_RE |
| 81 | S2_BP_Chronique_S_P_RC |
| 82 | S2_BP_Chronique_S_P_DC |