

RAPPORT DE STAGE DE FIN D'ETUDES – PROMOTION MACS 2013

---

# VISUALISATION DES CORRELATIONS STATISTIQUES ENTRE ACTIFS FINANCIERS

---

De *Suzanne Ketsia NZE ONDO* – 19 Septembre 2013.



Responsable de stage : Bernard Main.

Tuteur universitaire : El Nouty.



# REMERCIEMENTS

---

Je ne saurais commencer ma rédaction sans d'abord et de tout cœur, remercier toutes les personnes qui ont contribué au bon déroulement de mon stage et plus généralement, de mes études.

Un grand merci à mon responsable Bernard Main, pour m'avoir permise d'intégrer l'équipe R&D du Groupe BBSP pour ce stage de fin d'études, et d'avoir veillé à la bonne évolution de mes résultats. Ses conseils, critiques et orientations m'ont guidée tout au long de ce stage.

Je remercie également Hugues Leroy qui a été présent sur toute la partie purement informatique du projet. Il a été d'un grand et indispensable apport pour mettre sur pied l'outil de visualisation. Aussi, je remercie toute les équipes CREO (informatique) et BBSP (analystes techniques) qui ont contribué à mon travail avec une disponibilité, une serviabilité et un regard critique qui m'ont été nécessaires pour avancer. Merci pour les connaissances qu'ils m'ont fait partager durant cette expérience.

Je pense ici évidemment à mes professeurs de Sup Galilée qui m'ont donné une très bonne formation d'ingénieur en mathématiques appliquées, tant sur le fond que sur la forme.

Pour finir, je tiens à remercier mon Grand Amour, mes parents, ma famille et mes amis qui ont été d'un soutien indéfectible et à toute épreuve durant toute ma scolarité. Votre confiance m'a poussée à me dépasser même dans les difficultés et j'espère aujourd'hui vous avoir rendu fiers de votre investissement dans ma vie.

A tous, pour tout, Merci.

# RESUME

---

BBSP est le leader européen du conseil en analyse graphique sur les marchés financiers de taux, d'indices et de devises. La société CREO est le département informatique filialisé du Groupe BBSP. C'est dans cette dernière entité que j'ai eu à effectuer mon stage de fin d'études, en vue de l'obtention de mon diplôme d'ingénieur.

Le sujet de ce stage vise à répondre au besoin des analystes techniques de BBSP de pouvoir visualiser de façon systématique les corrélations entre les actifs financiers, corrélations pas forcément évidentes à penser, afin d'appréhender/de suivre des pistes décisionnaires pas forcément instinctives.

Pour ce faire, il s'agit de mettre en place un démonstrateur fonctionnel d'un outil qui prendra en entrée les flux de données boursières de BBSP et rendra en sortie une représentation simple en forme de réseau, des corrélations statistiques détectées parmi les historiques de cours des actifs. Ce travail part d'un algorithme interne combinant le calcul des corrélations statistiques des rendements d'actifs financiers et la méthode du *Minimal Spanning Tree* (Arbre Couvrant de Poids Minimal) pour la répartition dans l'espace des actifs corrélés.

L'on se sert des données historiques des cours financiers afin de conclure sur d'éventuels liens plus ou moins durables et exploitables entre les différents actifs.

# Table des matières

<b>Introduction .....</b>	<b>8</b>
 <b>Partie I : Présentation des entreprises BBSP et CREO &amp; Mission du stage .....</b>	<b>9</b>
I. La société BBSP et sa société partenaire, CREO .....	10
I.1 Présentation .....	10
I.2 Historique de développement .....	10
I.3 Le métier BBSP et son approche des marchés .....	11
I.4 La méthodologie BBSP .....	13
I.4.1 Le support écrit .....	13
I.4.2 La Hotline .....	13
I.4.3 Les devises .....	14
I.4.4 Deux produits : TAO et ACE .....	14
I.5 Chiffre d'affaire et stratégie .....	15
I.6 Perspectives d'avenir .....	16
II. Mission du stage .....	16
 <b>Partie II : Développement du corps du sujet .....</b>	<b>19</b>
I. Notions essentielles à la compréhension du sujet .....	20
I.1 Les corrélations en finance .....	20
I.1.1 Méthode de calcul .....	20
I.1.2 Les utilisations de la corrélation .....	22
I.1.3 Corrélations entre les différents marchés .....	22
I.1.4 Les limites du coefficient de corrélation .....	23
I.2 Le Minimal Spanning Tree (Arbre Couvrant Minimal) .....	24
I.2.1 Contexte d'utilisation .....	24
I.2.2 Algorithmes liés à l'arbre couvrant minimal .....	25
I.3 Traitement des données .....	28
I.3.1 Définitions .....	28

I.3.2	Les différents traitements des données manquantes .....	30
I.3.3	Description des méthodes courantes dans le traitement par imputation des données manquantes .....	32
II.	Analyse du sujet .....	33
II.1	Présentation simplifiée de l'algorithme interne à BBSP .....	34
II.1.1	Présentation de l'étude de Rosario N. Mantegna.....	34
II.1.2	Exemple d'usage du Minimal Spanning Tree .....	37
II.2	Récapitulatif de l'algorithme .....	38
III.	Application et résultats .....	39
III.1	Les données .....	40
III.1.1	Le CAC 40.....	41
III.1.2	Le Stoxx 600 ou le Dow Jones Stoxx 600.....	41
III.1.3	Les Commodities (Matières Premières).....	41
III.2	Démarche adoptée lors du projet .....	42
III.2.1	Traitement des données .....	42
III.2.2	Formule de calcul de la corrélation empirique .....	44
III.2.3	Représentation graphique .....	44
III.3	Evolution du travail.....	46
III.3.1	Point de lancement.....	46
III.3.2	Evolution des moyens techniques .....	49
III.3.3	Une évolution en fonction des besoins des analystes.....	50
III.3.4	Evolution du démonstrateur.....	52
IV.	Perspective d'évolution de l'outil.....	67
IV.1	Autres idées à exploiter.....	67
IV.2	Une évolution de l'affichage vers la 3D .....	68
<b>Partie III : Les apports du stage.....</b>		<b>69</b>
I.	Environnement de travail.....	70

I.1	Un travail en équipe .....	70
I.2	Organisation du travail .....	70
II.	Bilan final .....	71
<b>Conclusion .....</b>		<b>72</b>
<b>Bibliographie .....</b>		<b>73</b>

# INTRODUCTION

---

Aujourd'hui, les stratégies d'investissement s'inscrivent de plus en plus dans les logiques d'optimisation de portefeuille. Le challenge quotidien des professionnels de la finance tels les analystes financiers, traders et autres décisionnaires du monde de l'investissement financier se résume à prendre les meilleures décisions, faire les meilleurs arrangements, afin d'obtenir un gain optimal. A cela se rajoute l'interpénétration croissante et de plus en plus complexe des marchés financiers qui favorise le développement d'approches inter-marchés.

C'est dans ce contexte que la recherche de BBSP, dans le souci de s'adapter et d'optimiser ses outils d'aide à la décision, s'intéresse aux moyens d'identifier et d'analyser, de façon systématique, des corrélations statistiques pouvant exister entre les cours de plusieurs actifs financiers. Pour cela, BBSP souhaite se doter d'un outil de visualisation de ces dites corrélations statistiques.

Mon stage vise à poser les bases de cet outil. Il s'est déroulé au sein de la SSII interne CREO, département informatique de BBSP. L'approche retenue par BBSP pour développer cet outil est celle du *Minimal Spanning Tree (Arbre Couvrant de Poids Minimal)* - noté MST tout au long de ce rapport - pour lequel BBSP dispose d'un algorithme général. Il s'agira avec l'aide mise à ma disposition, tant en ressources techniques qu'humaines, de contribuer à la mise en place d'un démonstrateur fonctionnel pour cet outil.

Je vais m'atteler durant tout le rapport, à décrire la tâche qui m'a été confiée, à développer les différents moyens dont j'ai usé pour remédier à certaines problématiques rencontrées et enfin, à exposer de façon progressive mes résultats.

Pour ce faire et dans un souci de clarté, j'ai organisé mon rapport de stage en trois grandes parties, la première étant consacrée à la présentation des entreprises partenaires BBSP et CREO ainsi qu'à la présentation de mon sujet de stage, la deuxième analysant les différents aspects du sujet. Nous aurons aussi là le détail de mon travail et les résultats obtenus. La troisième enfin aborde mon environnement de travail, les évolutions éventuelles du sujet et les apports de ce stage dans ma connaissance du métier d'ingénieurs MACS.



**PARTIE I**  
**PRESENTATION DES ENTREPRISES BBSP ET CREO**  
**&**  
**MISSION DU STAGE.**

---

## **I. La société BBSP et sa société partenaire, CREO**

### **I.1 Présentation**

Créée en 1990, **BBSP**<sup>1</sup> est aujourd'hui le leader européen du conseil en analyse graphique<sup>2</sup> sur les marchés financiers de taux, d'indices et de devises. Il est l'un des principaux cabinets d'études techniques indépendants avec des bureaux en Europe, en Amérique du Nord et en Asie, au service de clients dans plus de 20 pays. Sa position de leader a été acquise par :

- Une approche originale d'analyse de l'évolution des prix de marchés et la compétence de l'équipe d'analystes consultants de BBSP ;
- Une forte dynamique commerciale et la qualité du service client ;
- Un outil informatique de communication sophistiqué en constante évolution.

**CREO**, filiale de BBSP créée en 1995, assure quant à elle le développement et la maintenance des moyens informatiques et internet nécessaires à la consultation des conseils de BBSP

Avec la création en septembre 2001 de BBSP Inc. Montréal (Québec) et BBSP Ltd. en Singapour en 2005, le Groupe couvre désormais tous les fuseaux horaires. Cette implantation est soutenue par la mise en place du conseil BBSP sur internet à partir d'Octobre 1999 via **son site TAO**.

Le groupe compte aujourd'hui à peu près 70 employés dont la majorité travaille au siège situé à Paris et il compte 8/10 des plus grosses banques mondiales parmi leurs clients.

### **I.2 Historique de développement**

#### **Avril 1990**

La société BBSP est une S.A. au capital de 500.000 Francs français, dont le siège social est situé à Paris. Initialement créée sous forme de S.A.R.L., BBSP s'est transformée en S.A. le 18 mai 1995.

---

<sup>1</sup> Le sigle **BBSP** est tiré des noms des quatre associés qui ont fondé cette entreprise. Ce sont **Boury** Stéphane, **Blanco** Jean-Michel, **Sadock** Philippe et **Pauwels** Antoine.

<sup>2</sup> C'est de l'analyse technique consistant à « **Interroger l'ancien pour connaître le nouveau** ».

BBSP a pour principale activité l'analyse des marchés financiers et la production, à partir d'études graphiques, de conseils de gestion dont la consultation est offerte à la clientèle essentiellement par le biais du téléphone et d'Internet.

### **Juin 1995**

La société CREO (Computer Research and Evolutive Object) est une S.A.R.L. au capital de 50000 Francs français, domiciliée à Paris.

Partenaire de BBSP, CREO assure la maintenance du parc informatique et le fonctionnement du réseau de BBSP. Elle développe, met en place et assure la maintenance de progiciels conçus par BBSP pour son besoin propre et pour celui de sa clientèle. La société compte à peu près une dizaine d'employés.

De plus, CREO développe le logiciel **ACE** (Analysis Comes Easy). Il s'agit d'un logiciel qui fonctionne en temps réel comportant les principaux supports nécessaires à l'élaboration d'analyses techniques. CREO est ainsi le département informatique de BBSP et gère toutes les plateformes (Asie – Amérique – Europe) de son centre à Paris.

### **Aout 1998**

La société BBSP Partners est une société par actions simplifiée au capital de 100.000 euros domiciliée à Paris.

Cette filiale règle les affaires courantes. Elle a pour vocation de vendre à une clientèle institutionnelle la production de BBSP avec une valeur ajoutée fournie par des vendeurs seniors qui ont une expertise complémentaire à celle des analystes de BBSP. Les clients sont aussi bien des traders compte propre que des hedge funds ou des sociétés de gestion.

Aujourd'hui, BBSP, CREO et BBSP Partners sont domiciliés au même endroit au 23 rue de Balzac à Paris 8<sup>ème</sup>.

## **I.3 Le métier BBSP et son approche des marchés**

Ces vingt dernières années ont été marquées par la mondialisation des marchés financiers et le développement des transactions sur produits dérivés, en particulier des Futures. Parallèlement, les déséquilibres monétaires ont rendu l'interprétation des données fondamentales plus difficiles. La multiplication des nouveaux produits comme les Fonds à capital garanti ou les Hedge Funds a également accru la volatilité des marchés.

Dans ce contexte de forte fluctuation et de grande technicité des marchés, le métier de BBSP est d'élaborer des prévisions sur l'évolution de l'ensemble des actifs (futures sur indices boursiers, marché des taux, marché des devises...), en se basant sur les techniques anglo-saxonnes d'analyse chartiste (analyse technique). **Les concepts macro-économiques et conjoncturels de l'analyse fondamentale<sup>3</sup> ne sont pas pris en compte dans ces prévisions.**

Pour réaliser leurs conseils, les analystes se reposent sur l'analyse technique basée sur trois concepts :

- L'analyse graphique : elle consiste essentiellement en la reconnaissance de formes caractéristiques. Il en existe 3 groupes :
  - la détermination d'une tendance ;
  - les figures de retournement de tendance ;
  - les figures de prolongement de tendance.
- Les indicateurs mathématiques : les analystes disposent aussi de modèles mathématiques (indicateurs) de modèles dits simples (type moyennes mobiles) et d'indicateurs de contre tendance (type RSI, momentum, stochastique, etc...).
- « Les vagues d'Elliot » : Ce sont des indicateurs empiriques de « rythme » de marché populaires chez les traders. Les vagues d'Elliot consistent en une théorie sur le développement des marchés financiers proposé par Ralph Nelson Elliott dans les années 1940.

L'analyse technique permet d'optimiser les décisions de gestion des opérateurs de marchés financiers. Elle consiste en la recherche de configurations graphiques caractéristiques du comportement des investisseurs qui permettent d'anticiper les prises de position massives à venir. Cette analyse, couplée avec l'étude d'indicateurs mathématiques (fonctions mathématiques appliquées au prix) est le cœur de métier de BBSP. En suivant les prix de marché, il est possible **d'appréhender l'état psychologique des opérateurs** et donc, l'évolution la plus probable de ces groupes en environnement de risque.

L'analyse technique répond aux trois points cruciaux suivants :

- Les mouvements de tendance : Où va le marché ?

---

<sup>3</sup> Le but de l'analyse fondamentale est d'évaluer au plus juste la valeur d'une société ou d'un secteur d'activité particulier. Il ne faut pas la voir comme une science mais comme un croisement de nombreuses disciplines. Parfois confondue à l'analyse financière.

- L'évolution des sous-tendances : Comment se dessine le mouvement ?
- La détermination d'un degré de conviction sur les niveaux charnières de tendance de prix<sup>4</sup> : Où est le risque ?

## **I.4 La méthodologie BBSP**

BBSP fournit à ses clients un service personnalisé à leur horizon d'investissement, l'objectif étant l'amélioration de la performance par l'apport d'une valeur ajoutée clairement mesurable. Le conseil s'articule autour de plusieurs moyens de diffusion :

### **I.4.1 Le support écrit**

BBSP propose des analyses sur les principaux marchés de futures sur taux et sur indices, ainsi que les principaux marchés de devises. Le support écrit matérialise la prestation de service.

De plus, BBSP fournit au client un commentaire analytique de l'évolution antérieure des marchés, une description de l'état présent et une projection des évolutions possibles en scénarios alternatifs. Le conseil définit des niveaux d'intervention spécifiques, des « points d'entrée », « pivot » et « cible ».

Aussi, BBSP édite des analyses à long-terme (*bi-annual*), mensuelles (*monthly*), des analyses hebdomadaires (*weekly*), des analyses quotidiennes (*daily*), ainsi que des actualisations de court terme (*intraday*).

### **I.4.2 La Hotline**

La personnalisation du service BBSP passe par la hotline. Les équipes des différents services taux, indices et devises assurent un commentaire téléphonique quotidien à chaque client. De plus, le client accède à tout moment en séance à une actualisation immédiate et personnalisée de l'analyse en fonction de l'évolution des marchés.

La Hotline ne remplace jamais le conseil écrit, mais vient illustrer les points forts, les convictions, les zones de risques, et permet de faire des parallèles entre les marchés.

Si la hotline personnalise le service BBSP, elle lui confère aussi originalité et réactivité aux besoins spécifiques des clients et en fait un produit haut de gamme sans équivalent.

---

<sup>4</sup> Niveaux sur lesquels le prix va s'appuyer lors d'un mouvement.

### **I.4.3 Les services**

Aujourd'hui, le conseil de BBSP s'articule autour de trois services :

#### **Service Taux**

Le service Taux est consacré au marché obligataire. C'est le service « phare » de BBSP et utilisé par 60% de la clientèle. BBSP analyse quotidiennement les EuroBund, Tbond, Gilt et CGB. Des analyses hebdomadaires (EuroSchatz, EuroBobl et TNote 10) et apériodiques (JGB, Euribor, EuroDollar et BAX) complètent le panel des marchés étudiés. La Hotline assure le suivi quotidien de l'ensemble des marchés précités.

#### **Service devise**

Utilisé par 30% de la clientèle, ce service s'est fortement développé depuis 1995. BBSP analyse quotidiennement l'euro/dollar, le dollar/yen et l'euro/sterling. Des analyses hebdomadaires sont éditées sur le dollar/suisse et l'euro/suisse. Le Cable (sterling/dollar), le sterling/yen, l'euro/yen, le canadien (USD/CAD) ainsi que l'Aussie (AUD/USD) sont analysés de manière apériodique. La Hotline assure le suivi quotidien des marchés précités.

#### **Service indice**

Le potentiel du service indices est à la mesure de l'engouement croissant pour le marché des actions. BBSP analyse quotidiennement les Dow Jones, DAX et CAC40. La hotline assure le suivi quotidien des marchés précités ainsi que des marchés analysés une fois par semaine (Footsie et Nasdaq) et de manière apériodique (Nikkei, MIB 30, smic).

### **I.4.4 Deux produits : TAO et ACE**

#### **TAO : logiciel de consultation et de transmission du conseil**

BBSP s'est doté en 1996 d'un outil de communication de pointe : **TAO**.

TAO, propriété de BBSP, est développé et maintenu par CREO. Le client accède pour tous les marchés auquel il « s'abonne » à un conseil à divers horizons temporels, soutenu par une représentation graphique de l'étude BBSP.

D'abord accessible par modem, le service BBSP est diffusé depuis Octobre 1999 via internet avec TAO web. La diffusion par internet renforce la dimension temps réel du conseil et la réactivité de son apport écrit. De plus, internet élargit la gamme des utilisateurs potentiels.

## **ACE**

Le logiciel graphique ACE (*Analysis Comes Easy*) a également été développé par CREO en coopération avec les spécialistes de BBSP (10 ans d'expérience en analyse technique). Ce logiciel est interfacé avec le flux de données COMSTOCK.

ACE est un outil d'aide à la décision destinée aux opérateurs des salles de marché. ACE permet de visualiser graphiquement l'évolution des marchés et de les analyser. De nombreuses fonctions graphiques sont disponibles.

## **1.5 Chiffre d'affaire et stratégie**

BBSP s'adresse à une clientèle de professionnels de marchés financiers, au sein d'établissements tels que des banques, des compagnies d'assurance, des caisses de retraite, des groupes industriels et des « Hedge funds ». Les utilisateurs des services de BBSP sont principalement des responsables de salles de marché, des responsables de desk, des teneurs de marché, des traders propriétaires, des arbitragistes et bien entendu des gestionnaires de portefeuille. De par la technicité de son conseil et la forte volatilité des marchés analysés, BBSP privilégie une clientèle de professionnels.

BBSP réalise plus de 60% de son chiffre d'affaires hors de France, principalement au Royaume-Uni, en Allemagne, en Suisse, au Benelux, en Espagne, au Canada et aux Etats-Unis. BBSP compte environ 200 clients dont les plus grosses banques d'investissement mondiales.

Le chiffre d'affaires de BBSP est récurrent, les clients souscrivent un abonnement annuel renouvelable par tacite reconduction.

La relation directe de BBSP avec les bénéficiaires de ses services donne une dimension particulière au turnover de la clientèle, compris entre 10 et 15% annuellement. Les clients de BBSP sont globalement fidèles : un client qui a changé d'employeur introduira le plus souvent le service BBSP auprès de sa nouvelle équipe.

Cette stratégie de fidélisation a permis à la société de ne pas être trop affectée par les restructurations de la profession : concentration des activités sur un nombre réduit de

places financières, fusions et absorptions de grandes banques, difficultés des banques asiatiques.

## I.6 Perspectives d'avenir

Pour la deuxième année consécutive, BBSP s'est vu attribuer la récompense de « Meilleur centre de recherche de l'année » et « Meilleure recherche indépendante sur le marché des actions » en 2013 par un groupe d'experts de l'analyse technique.



BBSP s'attelle ainsi à consolider sa position sur le marché face à la concurrence.

## II. Mission du stage

Nous l'aurons ainsi compris avec la partie précédente : BBSP fait de l'analyse technique et cela se résume à « **Interroger l'ancien pour connaître le nouveau** ».

Comme évoqué en Introduction, BBSP veut faire face à l'interpénétration des marchés dans ses stratégies d'investissement et pour ce faire, souhaite développer un moyen de visualiser et d'analyser les corrélations statistiques pouvant exister entre plusieurs actifs financiers.

Il s'agit d'un outil appelé à servir de **paramètre d'étude** aux analystes de l'entreprise BBSP mais aussi **d'information mise en ligne** à destination de la clientèle : la qualité et la clarté de la représentation graphique sont donc des critères importants.

Cette mission est ensuite chargée d'être mise sur pied par CREO mais avant cela, le stage portera sur la mise en place d'un **démonstrateur fonctionnel** pour cet outil.

### La demande



Après des recherches faites en collaboration avec Centrale Paris, un algorithme interne, celui du *Minimal Spanning Tree* appliqué au calcul des corrélations statistiques, a été adopté.

Travaillant sur un univers restreints d'actifs dont les historiques seront quotidiennement mis à jour, cet outil doit être capable :

- De rendre visibles les corrélations les plus significantes au sein d'un ensemble d'actifs dont le périmètre aura été défini par l'utilisateur ;
- Au sein d'un ensemble plus étendu, de rendre visibles les actifs les plus fortement corrélés à un actif donné, choisi par l'utilisateur.

## Spécifications

Le démonstrateur doit fonctionner de façon autonome et sur une base quotidienne. Il comportera des fonctionnalités de paramétrage simples et accessibles pour ce qui concerne la création et l'édition d'ensembles d'études.

### En entrée :

Le module acceptera en entrée les flux de données fournis par BBSP.

### En sortie :

Le démonstrateur devra fournir en sortie une **représentation graphique** simple des corrélations détectées (il s'agit d'un réseau).

Il doit également fournir une **caractérisation** de cette représentation dans un format ouvert qui puisse être ensuite traité par des modules d'affichage plus sophistiqués. Par exemple, les formats XML, JSON ou langage DOT se prêtent à ce type de traitement.

Idéalement, le démonstrateur pourra prendre la forme d'une **application Web** autonome, exploitant les ressources d'un framework, et tirer parti d'un module open source de représentation graphique.

## Difficultés

Les difficultés envisagées sont de plusieurs ordres :

- Prétraitement et conformation des données qui peuvent parfois être aberrantes ou hétérogènes ;
- Calcul sur des séries massives. Il est possible que ces derniers ne puissent avoir lieu en temps réel, et devront être mémorisés. On craint donc ici un grand temps de calcul qui nécessiterait de laisser tourner le programme toute la nuit ;

- Optimisation de l'algorithme et des ressources de calcul, en vue d'un changement d'échelle pour aboutir à un prototype industriel.

Rajoutons quand même pour clarifier tout ceci, que des outils simples de visualisation de corrélation existent déjà sous Bloomberg (que les analystes de BBSP utilisent également) par exemple, pour connaître la corrélation entre 2 actifs. Mais la problématique de BBSP concerne l'analyse massive de corrélations au sein d'un panier d'actifs. Il s'agit de rendre lisibles toutes les corrélations existant à l'intérieur d'un panier d'actifs pris dans un ou plusieurs secteurs différents. On pourra, par exemple, chercher s'il existe une corrélation entre l'indice d'une devise (par exemple l'EUR/US) et les différents actifs constituant le CAC 40.

D'autres tentatives pour visualiser les corrélations ont également déjà été testées par BBSP mais celle-ci s'avèrera être la plus aboutie.

### **Répartition des tâches**

Je précise ici que compte tenu de la forte composante informatique exigée pour la représentation graphique, je me suis essentiellement occupée d'analyser le modèle mathématique et de réaliser les tests nécessaires... Je détaillerai cela dans la partie II.

C'est M. Hugues Leroy, qui s'est chargé d'afficher les résultats des matrices de corrélations et de rendre l'outil le plus fonctionnel possible.

Ceci étant dit, passons à la partie pratique du stage.

## **PARTIE II**

### **DEVELOPPEMENT DU CORPS DU SUJET.**

---

Dans cette partie, nous allons, après avoir abordé les notions essentielles au projet, passer à l'analyse du modèle (algorithme interne de l'entreprise BBSP) et ensuite, nous en viendrons à l'application que nous en avons faite. Précisons que nous travaillions avec les analystes qui définissent leurs besoins par rapport à cet outil. Les différents changements et évolutions tiennent compte de cela.

## I. Notions essentielles à la compréhension du sujet

### I.1 Les corrélations en finance

La corrélation mesure la force du lien existant entre deux variables sur une période donnée. Elle peut être positive ou négative. Dans le premier cas, cela signifie que les deux variables ont tendance à évoluer dans le même sens. A l'inverse, une corrélation négative intervient lorsque deux variables évoluent dans un sens opposé.

La corrélation entre deux variables est mesurée par le **coefficient de corrélation linéaire**. Celui-ci n'est pas sensible aux unités de chacune des variables. En revanche, les valeurs extrêmes impactent fortement le coefficient.

#### I.1.1 Méthode de calcul

La corrélation est un concept issu de la biologie. C'est par le biais des travaux de Francis Galton<sup>5</sup> que la corrélation devient un concept statistique. Toutefois pour Galton, la notion de corrélation n'est pas définie précisément et il l'assimile dans un premier temps à la droite de régression d'un modèle de régression linéaire.

C'est ensuite Karl Pearson qui propose en 1896 une formule mathématique pour la notion de corrélation et un estimateur de cette grandeur.

Soient  $X$  et  $Y$  deux variables. Le coefficient de corrélation sera :

$$\rho(X, Y) = \frac{\sum_{i=1}^m (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^m (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^m (Y_i - \bar{Y})^2}} = \frac{Cov(X, Y)}{\sqrt{V(X)} \cdot \sqrt{V(Y)}}$$

---

<sup>5</sup> Homme de science britannique (né en 16 Février 1822 à Sparkbrook – mort le 17 Janvier 1911 à Haslemere). Anthropologue, géographe, inventeur, météorologue, proto-généticien, psychométricien et statisticien.

Le coefficient de corrélation, noté  $\rho$  est compris entre -1 et 1.

- -1 signifie que les deux variables sont corrélées négativement de façon parfaite. Elles évoluent donc dans le sens contraire à chaque mouvement de marché.
- 1 signifie qu'il y a corrélation positive parfaite. Les deux variables évoluent dans le même sens et avec la même intensité.
- 0 signifie qu'il n'existe aucun lien entre les mouvements des deux variables. Elles sont totalement décorélées. Toutefois, cela ne veut pas dire que les variables sont indépendantes. Deux variables indépendantes sont forcément dé-corrélées mais l'inverse n'est pas forcément vrai. En effet, deux variables dé-corrélées peuvent être corrélées de façon non linéaire.

Toutefois, les corrélations parfaites ou la non corrélation interviennent très rarement. On parle davantage d'une corrélation positive (ou négative) forte ou faible.

**Remarque** : Notons que le fait que deux variables soit « fortement corrélées » ne démontre pas qu'il y ait une relation de causalité entre l'une et l'autre. Le contre-exemple le plus typique est celui où elles sont en fait liées par une causalité commune.

Généralement, nous avons ces seuils définis :

Corrélation	Négative	Positive
Faible	[-0.5, 0.0]	[0.0, 0.5]
Forte	[-1.0, 0.5]	[0.5, 1.0]

Il est aussi important de préciser que le coefficient de corrélation empirique est un estimateur biaisé. Fort heureusement, le biais devient négligeable lorsque l'effectif (ou disons, la taille du vecteur) augmente. L'espérance de l'estimateur s'écrit<sup>6</sup> :

$$E[\hat{\rho}] = \rho - \frac{\rho(1 - \rho^2)}{2n}$$

Avec  $n$  la taille des vecteurs  $X$  et  $Y$ .

Pour cette raison, certains logiciels proposent un coefficient ajusté<sup>7</sup>

$$\hat{\rho}_{aj} = \sqrt{1 - \frac{n-1}{n-2}(1 - \hat{\rho}^2)}$$

<sup>6</sup> Page 107 du livre référencé : Aivazian, S., *Etude statistique des dépendances*, Mir, Moscou, 1978

<sup>7</sup> Page 274 du livre référencé

Bien entendu, l'ajustement est d'autant plus sensible que l'effectif est faible. Lorsque  $n$  est élevé,  $\hat{\rho}$  et  $\hat{\rho}_{aj}$  se confondent. Dans notre cas, nous nous contenterons d'implémenter le coefficient de corrélation empirique classique défini.

### **I.1.2 Les utilisations de la corrélation**

#### **Diversification du portefeuille**

L'étude des corrélations permet de réaliser une bonne diversification au sein de son portefeuille. En sélectionnant des actifs peu corrélés entre eux, le risque de votre portefeuille sera réduit.

#### **Détecter des valeurs non corrélés au marché de référence**

Certaines valeurs n'évoluent pas en fonction des variations de marché mais simplement des nouvelles économiques sur l'entreprise. Pour cela, il faut par exemple regarder la corrélation entre l'indice CAC40 et les valeurs qui le composent.

#### **Détecter les valeurs qui sont corrélées négativement au marché**

En période de baisse des marchés, si une valeur est en hausse, cela montre la force de la valeur et peut donc être un bon investissement.

#### **Détecter les corrélations décalées**

Certains actifs évoluent dans le même sens mais de manière décalés. Par exemple, l'actif A va suivre les variations de B mais avec 1 mois de décalage. L'investisseur peut alors anticiper la hausse de l'actif B.

### **I.1.3 Corrélations entre les différents marchés**

Les relations inter-marchés sont très importantes dans le trading. Un investisseur qui comprend ces relations va pouvoir changer de marché en fonction de l'évolution de la

conjoncture et des conditions de marchés. Voici les principales corrélations entre les différents marchés :

- Prix des matières premières / Prix des obligations : Il y a une forte corrélation négative. En effet, une hausse des matières premières signifie souvent que l'économie est en bonne santé, et donc la demande de matières premières est en hausse. Les taux d'intérêts vont alors monter pour lutter contre l'inflation et une hausse des taux fera baisser le cours des obligations.
- Actions / Prix des obligations : Il y a une corrélation négative. En cas de baisse des taux (et donc de hausse du prix des obligations), les actions deviennent plus attractives. En revanche, si l'inflation est faible ou que nous sommes en déflation (une baisse des taux), c'est l'inverse qui va se produire, les obligations deviennent plus attractives. C'est ce que l'on appelle le "flight to quality", se traduisant par un transfert des liquidités des marchés actions vers les marchés d'obligations.
- Prix des matières premières / Dollar US : Il y a une forte corrélation négative. En effet, les matières premières sont cotées en dollar US. Ainsi, si le dollar s'apprécie, le prix des matières premières va baisser du fait d'une baisse de l'inflation.
- Actions / Dollar US : Il y a une corrélation négative. Si le dollar US s'apprécie, c'est que les investisseurs ne sont pas optimistes. Le dollar US est considéré comme une valeur refuge. Ainsi, les marchés actions vont baisser. En revanche, si la confiance revient, les investisseurs vont se tourner vers des investissements plus risqués, les actions ou encore l'euro.

#### **I.1.4 Les limites du coefficient de corrélation**

Bien que la notion de corrélation soit largement adoptée dans le milieu de la finance de marché et notamment dans le trading à haute fréquence, elle est critiquée par beaucoup comme relevant plus du charlatanisme.

En effet, la corrélation étudie la relation entre deux variables uniquement de manière linéaire mais il peut exister une relation non linéaire entre deux actifs. C'est pour cette raison que deux actifs indépendants ont une corrélation de 0, mais l'inverse n'est pas forcément vrai. Deux actifs avec une corrélation de 0 ne sont forcément indépendants.

De plus, la corrélation étudie la moyenne des variations. Or si les variations d'un actif sont très hétérogènes, la dispersion autour de la moyenne est importante. L'actif aura toutefois une corrélation importante avec un actif dont la moyenne des variations est sensiblement la même mais dont la dispersion autour de la moyenne est beaucoup moins importante.

Cette crainte de l'utilisation des corrélations comme système de mesure n'empêche néanmoins pas le coefficient de corrélation soit encore et pour un bon moment encore, un facteur plus qu'essentiel et simple dans la gestion des actifs financiers.

## I.2 Le Minimal Spanning Tree (Arbre couvrant de poids minimal)

En théorie des graphes<sup>8</sup>, étant donné un graphe non orienté connexe dont les arêtes sont pondérées, un arbre couvrant de poids minimal de ce graphe est un arbre couvrant (sous-ensemble qui est un arbre et qui connecte tous les sommets ensemble) dont la somme des poids des arêtes est minimale. L'arbre couvrant de poids minimal est aussi connu sous certains autres noms, tel qu'arbre couvrant minimum ou encore *arbre sous-tendant minimum*.

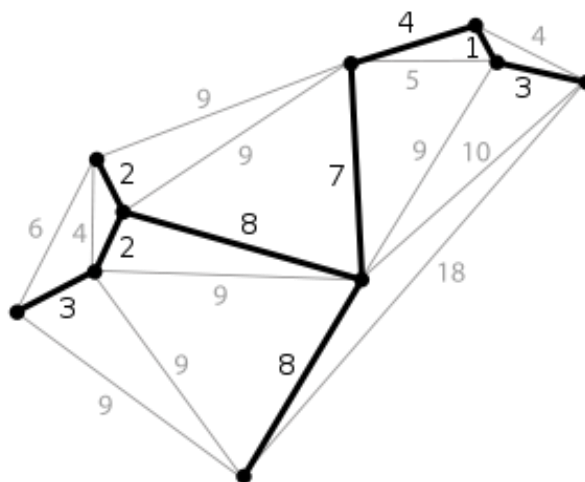


Figure 1. Arbre couvrant de poids minimal d'un graphe planaire. Chaque arête étant identifiée par son poids, qui ici, est approximativement sa longueur.

### I.2.1 Contexte d'utilisation

<sup>8</sup> La théorie des graphes est une théorie informatique et mathématique.



Les graphes sont un outil de représentation puissant et l'arbre couvrant minimum peut s'interpréter de différentes manières selon ce que représente le graphe. De manière générale, si on considère un réseau où un ensemble d'objets doivent être reliés entre eux (par exemple un réseau électrique et des habitations), l'arbre couvrant minimum est la façon de construire un tel réseau en minimisant un coût représenté par le poids des arêtes (par exemple la longueur totale de câble utilisé pour construire un réseau électrique).

C'est donc une technique de représentation présente dans plusieurs domaines dont les Telecom & Réseaux ou les réseaux routiers. Plusieurs algorithmes ont été mis en place afin de résoudre ce problème.

### I.2.2 Algorithmes liés à l'arbre couvrant de poids minimal

Il existe de nombreux algorithmes de recherche d'un arbre couvrant de poids minimal. On citera, entre autres, l'algorithme de Borůvka (le premier algorithme inventé pour ce problème), l'algorithme de Prim et l'algorithme de Kruskal.

#### Algorithme de Borůvka

C'est le premier algorithme de recherche d'arbre couvrant minimal découvert publié par Otakar Borůvka<sup>9</sup> en 1926 dans l'article « *O jistém problému minimálním* » (Sur un certain problème minimal). Au départ, il était destiné à rendre le réseau électrique de Moravie efficace. Il fut redécouvert à de nombreuses reprises par la suite.

Le principe est de réduire G en contractant des arêtes. On choisit peu à peu les arêtes qui seront dans l'arbre, et à chaque fois que l'on en choisit une, on fusionne les nœuds que cette arête relie. Ainsi, il ne reste plus qu'un sommet à la fin.

#### Algorithme

On prend G le graphe et F l'ensemble des arêtes choisies (on le remplit peu à peu).

- 1  $F \leftarrow \emptyset$
- 2 Tant que G n'est pas réduit à un sommet faire
- 3     Détruire les boucles de G (\*)

---

<sup>9</sup> **Otakar Borůvka** (10 mai 1899, Uherský Ostroh – 22 juillet 1995, Brno) est un mathématicien tchèque connu pour son travail en théorie des graphes.

```

4   Remplacer les arêtes multiples entre deux sommets par une seule dont
le poids est le minimum
5   Pour tout  $x$ , sommet de  $G$  faire
6       Trouver l'arête  $e_x$  de poids minimum adjacente à  $x$ 
7        $F \leftarrow F \cup \{e_x\}$ 
8       Contracter  $e_x$  (**)
9   fin pour
10 fin tant que
11 renvoyer  $F$ ;

```

(\*) Les arêtes qui partent et arrivent au même sommet ; ce genre d'arête peut apparaître après une itération.

(\*\*) Fusionner les sommets aux extrémités de  $e_x$ .

L'algorithme a une complexité en  $\Theta(A \ln(S))$ , où  $A$  est le nombre d'arêtes et  $S$  le nombre de sommets du graphe considéré.

## Algorithme de Prim

L'algorithme de Prim est un algorithme glouton<sup>10</sup> qui permet de trouver un arbre couvrant minimal dans un graphe connexe valué et non-orienté. En d'autres termes, cet algorithme trouve un sous-ensemble d'arêtes formant un arbre sur l'ensemble des sommets du graphe initial, et tel que la somme des poids de ces arêtes soit minimale. Si le graphe n'est pas connexe, alors l'algorithme ne déterminera l'arbre couvrant minimal que d'une composante connexe du graphe. Il a été conçu en 1957 par Robert C. Prim.

L'algorithme consiste à choisir arbitrairement un sommet et à faire croître un arbre à partir de ce sommet. Chaque augmentation se fait de la manière la plus économique possible.

### Algorithme

```

Procédure PRIM
Paramètres locaux : entier  $s$ , graphe  $G$ 
Paramètres globaux : graphe  $T$ 
Variables :
entier  $i, m, y$ 
réel :  $v$ 
ensemble :  $M$ 
TvectNent : pp
TvectNReel : d
Début
1   $T \leftarrow \text{graphe\_vide}$ 
2   $M \leftarrow \text{ensemble\_vide}$ 
3  Pour  $i \leftarrow 0$  jusqu'à  $N$  Faire
4       $d[i] \leftarrow \text{coût}(s, i, G)$ 

```

---

<sup>10</sup> Un algorithme glouton (greedy algorithm) est un algorithme qui suit le principe de faire, étape par étape, un choix optimum local, dans l'espoir d'obtenir un résultat optimum global.

```

5   pp[i] ← s
6   M ← Ajouter (i,M)
7   Fin pour
8   M ← Supprimer (s,M)
9   Tant que M <> Ensemble_vide Faire
10  m ← Choisir (M,d)
11  M ← Supprimer (m,M)
12  z ← pp[m]
13  v ← coût (m,z,G)
14  T ← Ajout arête <m,z> de coût v à T
15  Pour i ← 1 jusqu'à d° m dans G Faire
16    y ← i ième_succ_de m dans G
17    Si y ∈ M et (coût(m,y,G) < d[y]) alors
18      d[y] ← coût(m,y,G)
19      pp[y] ← m
20    Fin Si
21  Fin Pour
22 Fin Tant que
Fin algo

```

L'initialisation consiste à choisir un sommet quelconque du graphe initial.

**Remarque :** La complexité de l'algorithme dépend fortement de la manière dont est implémenté le choix de l'arête/sommet à ajouter dans l'ensemble à chaque étape. Avec une représentation naïve, on obtient une complexité en  $\theta(A * S)$  avec  $A$  le nombre d'arêtes et  $S$  le nombre de sommets. Si l'on utilise un *tas min binaire*, la complexité devient alors  $\theta(A * \ln(S))$ . En utilisant un *tas de Fibonacci*<sup>11</sup>, on peut encore descendre à  $\theta(A + S * \ln(S))$ .

## Algorithme de Kruskal

L'algorithme de Kruskal est un algorithme de recherche d'arbre recouvrant de poids minimum (ARPM) ou arbre couvrant minimum (ACM) dans un graphe connexe valué et non-orienté. Il a été conçu en 1956 par Joseph Kruskal.

L'algorithme consiste à d'abord ranger par ordre de poids croissant les arêtes d'un graphe, puis à retirer une à une les arêtes selon cet ordre et à les ajouter à l'ACM cherché tant que cet ajout ne fait pas apparaître un cycle dans l'ACM.

### Algorithme

```

KRUSKAL (G,w)
1 E := ∅
2 pour chaque sommet v de G
3   faire CRÉER-ENSEMBLE (v)
4 trier les arêtes de G par ordre croissant de poids w
5 pour chaque arête (u,v) de G prise par ordre de poids croissant
6   faire si ENSEMBLE-REPRÉSENTATIF (u) ≠ ENSEMBLE-REPRÉSENTATIF (v)

```

<sup>11</sup> En informatique, un **tas de Fibonacci** est une structure de données similaire au tas binomial, mais avec un meilleur temps d'exécution amorti.

```

7         alors ajouter l'arête (u,v) à l'ensemble E
8         UNION (u,v)
9 retourner E

```

- $W$  est une fonction qui associe à chaque arête du graphe  $G$  une valeur qui est son poids.
- Les fonctions ENSEMBLE-REPRÉSENTATIF et UNION sont les deux opérations d'une structure Union-Find (qui, respectivement, renvoie un élément représentatif d'un ensemble et fusionne deux ensembles).

La complexité de l'algorithme, dominée par l'étape de tri des arêtes, est  $\theta(A \log A)$  avec  $A$  le nombre d'arêtes du graphe  $G$ . On remarquera que lors du déroulement de l'algorithme, l'ACM n'est pas nécessairement connexe, il ne le sera qu'à la fin.

Ainsi, le but du Minimal Spanning Tree est de relier les sommets en attribuant un poids minimal à l'arête. Ce poids peut être la longueur de l'arête.

Pour finir avec cette partie, le fait de travailler avec des données massives extraites des historiques des cours d'actifs financiers nous forcera à nous adapter au problème du traitement des données.

### I.3 Traitement des données

Lorsqu'on a à traiter des données, on se confronte régulièrement au problème des valeurs manquantes et/ou aberrantes. Cette partie nous renvoie sur des méthodes de traitement statistiques.

#### I.3.1 Définitions








##### Les données manquantes







Pour prendre le cas des données manquantes sur le marché financier, cela s'explique notamment par le type de l'actif et par son activité. Si on prend l'exemple des devises, elles cotent tous les jours de l'année sans interruption, et on entend là les week-ends et les jours fériés également. Ce n'est pas le cas d'un actif du CAC 40 qui généralement, ne cotera pas le week-end et les jours fériés en France. Le marché est également international. Cela implique des horaires différents d'entrée et de fermeture du marché. Ce sont autant de paramètres qui font que sur une minute, une heure ou un jour, on n'ait pas forcément la cotation pour tous les actifs du marché et en l'occurrence, pour tous les actifs constituant un panier.

Ci-après une illustration.

Classeur1 - Mic

Fichier Accueil Insertion Mise en page Formules Données Révision Affichage Développeur

Calibri 11 A A     Renvoyer à la ligne automatiquement Standard  % 000  000  000

G I S   Fusionner et centrer    

Presse-papiers Police Alignement Nombre

	A	B	C	D	E	F	G	H	I	J
1	CYRE3 BZ Ex	Close Price		SPX Index	Close Price					
2		(USD)			(USD)					
3	21/10/2008	5,5281		21/10/2008	955,0500					
4	22/10/2008	4,3155		22/10/2008	896,7800					
5	23/10/2008	4,0993		23/10/2008	908,1100					
6	24/10/2008	3,5991		24/10/2008	876,7700					
7	27/10/2008	3,4611		27/10/2008	848,9200					
8	28/10/2008	4,7164		28/10/2008	940,5100					
9	29/10/2008	5,3789		29/10/2008	930,0900					
10	30/10/2008	5,7466		30/10/2008	954,0900					
11	31/10/2008	4,9120		31/10/2008	968,7500					
12	03/11/2008	5,4544		03/11/2008	966,3000					
13	04/11/2008	6,0047		04/11/2008	1005,7500					
14	05/11/2008	5,5766		05/11/2008	952,7700					
15	06/11/2008	4,4136		06/11/2008	904,8800					
16	07/11/2008	4,2657		07/11/2008	930,9900					
17	10/11/2008	3,6023		10/11/2008	919,2100					
18	11/11/2008	3,6474		11/11/2008	898,9500					
19	12/11/2008	3,1965		12/11/2008	852,3000					
20	13/11/2008	3,2798		13/11/2008	911,2900					
21	14/11/2008	3,2722		14/11/2008	873,2900					
22	17/11/2008	3,0363		17/11/2008	850,7500					
23	18/11/2008	2,8139		18/11/2008	859,1200					
24	19/11/2008	2,6471		19/11/2008	806,5800					
25	20/11/2008			20/11/2008	752,4400					
26	21/11/2008	2,3594		21/11/2008	800,0300					
27	24/11/2008	2,7122		24/11/2008	851,8100					
28	25/11/2008	3,1018		25/11/2008	857,3900					
29	26/11/2008	3,4612		26/11/2008	887,6800					
30	27/11/2008	3,3794		27/11/2008						
31	28/11/2008	3,3892		28/11/2008	896,2400					
32	01/12/2008	3,1670		01/12/2008	816,2100					
33	02/12/2008	3,1203		02/12/2008	848,8100					
34	03/12/2008	2,8325		03/12/2008	870,7400					
35	04/12/2008	2,9730		04/12/2008	845,2200					
36	05/12/2008	3,2383		05/12/2008	876,0700					
37	08/12/2008	3,5590		08/12/2008	909,7000					
38	09/12/2008	3,6629		09/12/2008	888,6700					
39	10/12/2008	3,9967		10/12/2008	899,2400					
40	11/12/2008	4,1349		11/12/2008	873,5900					
41	12/12/2008	4,0059		12/12/2008	879,7300					
42	15/12/2008	3,7639		15/12/2008	868,5700					
43	16/12/2008	3,9003		16/12/2008	913,1800					
44	17/12/2008	3,8022		17/12/2008	904,4200					
45	18/12/2008	3,6618		18/12/2008	885,2800					
46	19/12/2008	3,8128		19/12/2008	887,8800					
47	22/12/2008	3,7433		22/12/2008	871,6300					

Feuil1 Feuil2 Feuil3

Prêt

*Tableau. Données extraites de Bloomberg de CYRE3 BZ et SPX Index – Prix de clôture*

**Remarque :** Le but est d'avoir la matrice de corrélations estimée à partir d'un panier défini par l'utilisateur. Cela implique que dans ce panier, nous pourrions avoir des actifs provenant de différents pays. Ces différents actifs pourront ainsi présenter des jours de non-cotation différents en fonction des jours fériés de leurs pays respectifs ou d'autres raisons propres à leur activité. Souvent également dans le traitement des données manquantes en finance, les week-ends sont considérés comme des sauts pour pouvoir anticiper la reprise. Cependant, nous ne considérerons pas les week-ends comme des sauts à part si cela est nécessaire (s'il y a un ou plusieurs actifs du panier qui ont coté le week-end, les autres actifs considéreront alors que les week-ends sont des sauts).

### **Les données aberrantes**

Une donnée aberrante est une observation qui se trouve « loin » des autres observations. La présence d'une donnée aberrante peut signifier toutes sortes de choses. Cela peut être par exemple un cas qui ne fait pas partie de la population que l'on étudie (un adulte parmi un jeu de données concernant des enfants), ou bien une erreur de saisie ou de mesure. Dans le cas des flux de cotation, les données aberrantes sont bien des erreurs de saisie/mesure. Elles sont en général immédiatement apparentes à l'œil nu. Cependant :

- On traite des séries massives, donc elles se perdent dans la masse.
- Les cours étant par nature volatils, il faut disposer d'un algorithme robuste pour les détecter.

Certaines données aberrantes peuvent être aisément identifiées. Toutefois, l'appellation « donnée aberrante » reste très subjective. Il n'existe pas de définition mathématique ou statistique des données aberrantes. Leur traitement aussi est un vrai challenge car il faut pouvoir non seulement la détecter, mais aussi la traiter. C'est un cas qui fait de plus en plus l'objet de recherche afin de pouvoir le traiter efficacement. Vous verrez dans les applications de notre travail, la résolution prise pour régler ce problème.

### **I.3.2 Les différents traitements des données manquantes**

Statistiquement, les méthodes les plus utilisées afin de pallier à ce problème sont :

1. Ne rien faire ;
2. Utiliser uniquement les enregistrements pour lesquels les données sont complètes ;
3. Utiliser une méthode de repondération ;
4. Imputer une valeur.

### **Traitement 1 : Ne rien faire**

Cela oblige à travailler sur un fichier de données incomplet qui ressemble à un morceau de fromage gruyère. Si les valeurs manquantes sont peu nombreuses, on peut les oublier sans que cela ne pose problème.

### **Traitement 2 : Utiliser uniquement les enregistrements pour lesquels les données sont complètes**

Si les données sont présentées sous forme de tableau, cela revient à oublier une ligne dès qu'il manque une valeur dans cette ligne : on oublie donc aussi les autres valeurs de cette ligne, qui sont effectivement présentes.

Bien que cette option soit simple et permette d'utiliser un fichier complet, elle présente certains risques. En effet :

- L'échantillon de ceux qui contiennent toutes les données sur la période de temps exploitée peut être soit trop réduit pour être significatif, soit non représentatif de l'ensemble de la période.
- Cette méthode peut mener à des estimateurs fortement biaisés, à moins que la non-réponse ne dépende d'aucune des variables d'intérêts (cas MCAR<sup>12</sup>).

### **Traitement 3 : Repondération**

On remplace là les valeurs non-existantes par une valeur existante.

Les résultats de diverses analyses peuvent ne pas concorder, c'est pourquoi on utilise peu les méthodes de repondération pour tenir compte de la non-réponse.

### **Traitement 4 : L'imputation.**

Cette méthode consiste à produire une valeur artificielle pour remplacer la valeur manquante, avec pour objectif de produire des estimations approximativement sans biais. Les méthodes courantes d'imputation sont :

- La moyenne
- Le ratio

---

<sup>12</sup> MCAR pour *Missing Completely at random*. Réponse manquant entièrement au hasard.

- La régression
- Le hot-deck aléatoire
- Le plus proche voisin

Les trois premiers étant utilisés pour les méthodes déterministes et les deux dernières pour les méthodes stochastiques ou aléatoires.

Nous nous sommes nous, penchés sur la méthode de l'imputation pour décider du meilleur moyen de remplacement des données manquantes.

### **I.3.3 Description des méthodes courantes dans le traitement des données manquantes par imputation.**

#### **Imputation par la moyenne**

On remplace chacune des valeurs manquantes par la valeur moyenne de l'ensemble des réponses obtenues.

#### **Imputation par le ratio**

Chaque valeur manquante  $y_i$  est remplacée par la valeur prévue  $y_i^*$  obtenue par régression de  $y$  sur  $x$ .

#### **Imputation par régression**

C'est une extension naturelle de l'imputation par la méthode du ratio où l'on se sert de  $q$  variables auxiliaires  $x_1 \dots x_q$ .

#### **Imputation par la méthode de hot-deck aléatoire**

Cela consiste à attribuer la valeur de  $y$  fournie par un répondant (donneur), sélectionné au hasard avec remise parmi l'ensemble des répondants, pour remplacer la valeur manquante pour l'unité non-répondante (receveur).

#### **Imputation par la méthode du plus proche voisin**



On attribue à la valeur manquante, la valeur qui suit immédiatement après.

Méthode d'imputation	Moyenne	Ratio	Régression	Hot-deck aléatoire	Plus proche voisin
Valeur imputée	$y_i^* = \frac{1}{r} \sum_{i \in S_r} y_i = \overline{y_r}$	$y_i^* = \frac{\overline{y_r}}{\overline{x_r}} x_i$	$y_i^* = a + b x_{1i} + \dots + p x_{qi}$	$y_i^* = y_j$ pour certains $j \in S_r$ tels que $P(y_i^* = y_j) = 1/r$	$y_i^* = y_j$ pour certains $j \in S_r$ tels que $dist(x_i, x_j)$ soit minimal.

Il existe néanmoins des inconvénients à l'imputation.

1. Même si l'imputation produit un fichier complet de données, l'inférence, en particulier l'estimation ponctuelle, n'est valide que si les hypothèses sous-jacentes sont satisfaites.
2. L'imputation modifie les relations entre les variables.
3. Si les valeurs imputées sont traitées comme des valeurs observées, la variance de l'estimateur risque d'être considérablement sous-estimée, surtout si la proportion de non-réponses est appréciable.

L'imputation des données manquantes n'est pas une affaire banale. On constate qu'il existe de nombreuses méthodes et qu'il n'existe donc pas, de recette définitive. Il faudrait ainsi s'adapter à la situation qui se présente.

Dans notre cas, nous avons testé deux des méthodes de l'imputation (l'imputation par la moyenne et l'imputation par le ratio – cf *tableau ci-haut*) et l'utilisation des données pour lesquelles les données sont complètes (traitement 2), pour ne retenir que cette dernière. C'est la méthode qui nous rapproche le plus des résultats exacts.

## II. Analyse du projet

Cette étude part de l'intention de BBSP d'améliorer ou disons, d'optimiser la qualité d'analyse des différents analystes en poste sur les différents marchés financiers (Equities, Global Macro etc.). Dans cette optique, l'entreprise a mené ces dernières années un chantier de recherche tournant autour du data mining et des approches quantitatives, en partenariat avec plusieurs laboratoires universitaires.

Etant partis d'une problématique posée à l'Ecole Centrale Paris qui n'a pu trouver réponse, cette dernière les a orientés vers cette étude menée par Rosario N. Mantegna et qu'ils utilisent dans le cadre du trading haute fréquence.

## **II.1 Présentation simplifiée de l'algorithme interne à BBSP.**

Le modèle à reproduire et à tester est celui développé par Rosario N. Mantegna dans « *Hierarchical Structure in Financial Market* ».

### **II.1.1 Présentation de l'étude de Rosario N. Mantegna**

Il est nécessaire de présenter cette étude étant donné que c'est elle qui nous donnera de conclure sur une version simplifiée de l'algorithme interne à BBSP.

Rosario N. Mantegna propose un arrangement topologique des actions (stocks) échangées sur un marché financier associé à une taxonomie<sup>13</sup> économique significative. L'espace topologique est un graphe reliant les actions du portefeuille analysé. Le graphe est obtenu partant de la matrice de corrélation calculée avec toutes les paires d'actions du portefeuille, compte tenu de l'évolution synchrone de la différence du logarithme du cours de l'action quotidienne. L'arborescence de l'espace ultramétrique<sup>14</sup> sous-dominante associée avec le graphique fournit des informations utiles pour enquêter sur le nombre et la nature des facteurs économiques communs affectant l'évolution temporelle du logarithme du prix de groupes d'actions bien définis.

La motivation de cette étude est double. La première motivation concerne la recherche d'un type d'arrangement topologique pour les actions d'un portefeuille coté sur un marché financier. La seconde motivation est la recherche d'une preuve empirique de l'existence et de la nature de facteurs économiques communs qui conduisent l'évolution temporelle des prix des actifs.

Pour cette étude, M. Mantegna a enquêté sur la structure hiérarchique d'un portefeuille de  $n$  actions cotées sur un marché financier. Le facteur permettant de détecter

---

<sup>13</sup> Science des lois de la classification. Terme rencontré aussi en bactériologie, en botanique et en zoologie et qui donne de regrouper les êtres vivants en fonction de leurs caractéristiques.

<sup>14</sup> En mathématiques et plus précisément en topologie, se dit d'un espace métrique défini par une distance  $d$  sur un ensemble  $X$  vérifiant l'inégalité ultratriangulaire :

$$d(x, y) = \max(d(x, z), d(y, z))$$

l'arrangement topologique des actions d'un portefeuille est le **coefficient de corrélation synchrone de la différence journalière du logarithme du prix de clôture**. Le calcul de la corrélation s'effectue pour chaque paire d'actions selon une période de temps donnée.

### Démarche

Il s'agit donc d'abord de quantifier le degré de similitude entre l'évolution temporelle synchrone d'une paire de prix de l'action par le coefficient de corrélation.

$$\rho_{i,j} = \frac{\langle Y_i Y_j \rangle - \langle Y_i \rangle \langle Y_j \rangle}{\sqrt{(\langle Y_i^2 \rangle - \langle Y_i \rangle^2)(\langle Y_j^2 \rangle - \langle Y_j \rangle^2)}} \quad (1)$$

Avec  $i$  et  $j$  représentant les indices numériques des actions,  $Y_i = \ln P_i(t) - \ln P_i(t-1)$  et  $P_i(t)$  est le prix de clôture de l'action  $i$  au jour  $t$ .

La moyenne statistique est une moyenne temporelle effectuée sur tous les jours de cotations boursières de la période étudiée.

La formule (1) permet de définir une matrice de corrélation de taille  $n \times n$ , telle que:

- $(\rho)_{1 \leq i, j \leq n}$  est une matrice symétrique, donc  $\frac{n(n-1)}{2}$  coefficients définissent complètement la matrice ;
- $-1$  (*actifs inversement corrélés*)  $\leq \rho_{i,j} \leq 1$  (*actifs complètement corrélés*) ;
- $\rho_{i,j} = 0 \rightarrow$  les actifs  $i$  et  $j$  sont décorrélés ;
- $\rho_{i,i} = 1$ .

Ensuite, l'auteur Rosario N. Mantegna a recherché une métrique adaptée à la tâche, c'est-à-dire, définir la distance spatiale de deux actifs dont on a le coefficient de corrélation. Ce dernier ne peut à lui seul, définir une distance car il ne répond pas aux trois axiomes qui définissent une métrique euclidienne.

Une fonction de la corrélation peut tout de même être définie comme métrique générale.

Celle choisie est :

$$d(i, j) = 1 - \rho_{i,j}^2$$

Cette fonction remplit les trois axiomes de la métrique euclidienne :

$$(i) \quad d(i, j) = 0 \text{ si et seulement si } i = j;$$

$$(ii) \quad d(i, j) = d(j, i);$$

$$(iii) \ d(i, j) \leq d(i, k) + d(k, j)$$

Pour tout  $(i, j)$ .

D'où la matrice  $\mathbf{D} = (d)_{1 \leq i, j \leq n}$ , symétrique, qui est la matrice de distance qui permettra de construire le *Minimal Spanning Tree (MST)*.

Cette approche par le MST est intéressante car elle permet un arrangement des actions qui sélectionne les liens les plus forts, pour chaque point de chaque ensemble. De plus, le MST de façon directe, donne l'organisation hiérarchique ultramétrique sous-dominante des points (actions) du portefeuille étudié.

Ci-après un exemple de *Minimal Spanning Tree* pour les constituants du Dow Jones Industrial Average (DJIA). Cf *Hierarchical Structure in Financial Markets* de Rosario N. Mantegna.

Respectivement par ordre de grandeur croissant des corrélations, on a les couleurs **jaune – vert – turquoise – cyan – bleu – violet**. Voir les différents seuils ainsi que les noms correspondants aux abréviations des nœuds dans la partie FIGURES du rapport.

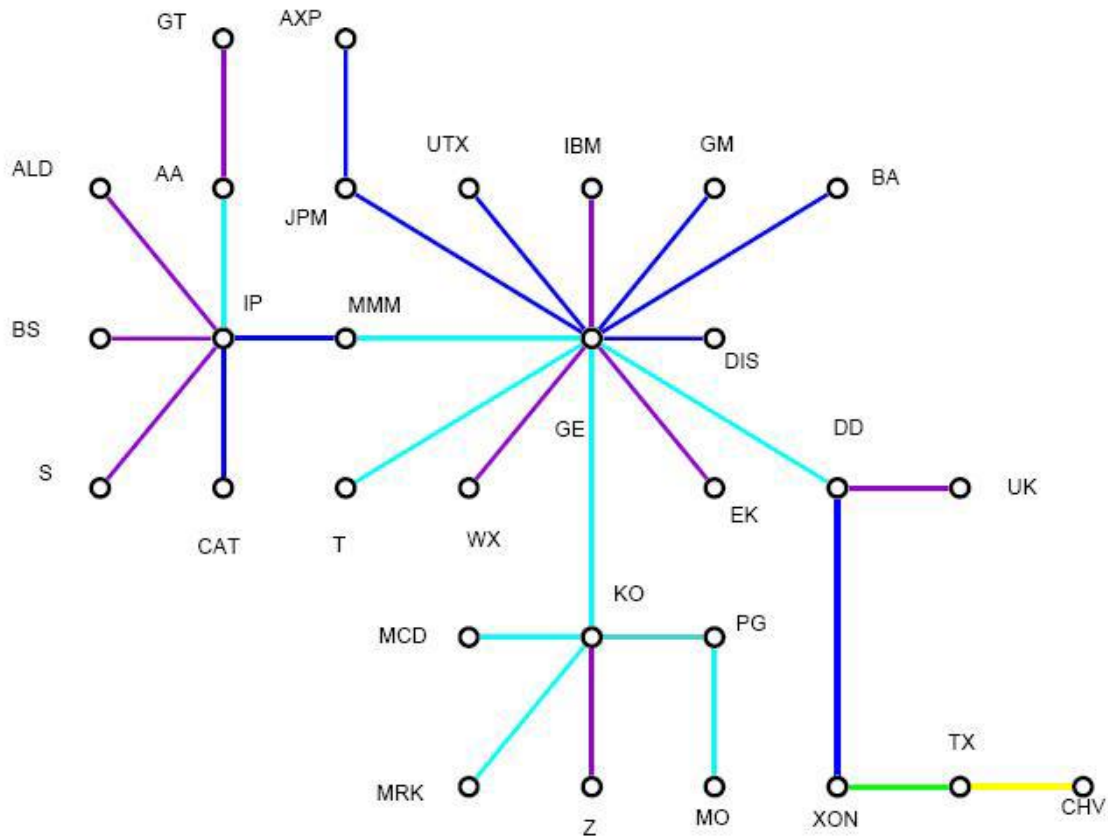


Figure 2. Minimal Spanning Tree pour les constituants du DJIA (période de Juillet 89 à Octobre 95).

**PS** : Voir la légende complète des abréviations et les seuils des couleurs dans la partie FIGURES du rapport.

### II.1.2 Exemple d'usage du Minimal Spanning Tree

Supposons par exemple qu'un analyste veuille connaître les actions les plus corrélées à International Paper (IP). On peut isoler en jouant sur les seuils afin de ne retenir que les liens les plus forts. Ci-après la figure représentative.

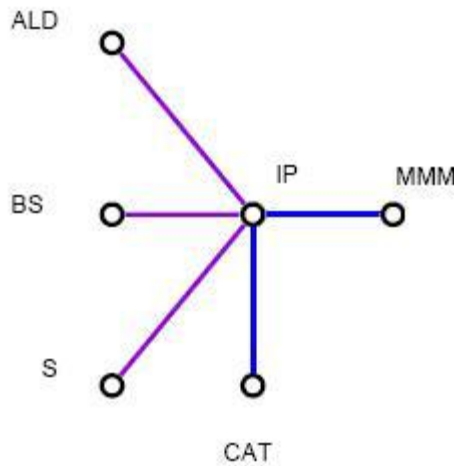


Figure 3. Groupe représentant les stocks les plus corrélés à IP

L'analyste en tirera ainsi les conclusions dont il a besoin dans son travail. Il a là une représentation claire de l'état du marché. Elle sera plus simple à exploiter qu'un tableau de nombres, qui peut s'avérer très gros selon le nombre d'actifs sur lequel on fait des recherches (Exemple du Stoxx 600 plus loin). Cet arrangement spacial s'avèrera d'autant plus utile et plus simple à manipuler.

## II.2 Récapitulatif de l'algorithme.

Dans son étude, l'auteur a essentiellement travaillé sur les actions, mais cette démarche peut être adaptée à tout type d'actif.

Tout simplement, le modèle à suivre peut être défini comme suit :

**Soit  $n$  actifs financiers  $S_i, i = 1, \dots, n$ . Pour chaque  $S_i$ , nous avons un temps de séries d'observations  $S_i(t_k), t_k = 0, \dots, m$  et une série de log-returns**

$$X_i(t_k) = \ln \frac{S_i(t_k)}{S_i(t_{k-1})}.$$

**Pour chaque paire  $i, j$  nous pouvons calculer la corrélation empirique entre les séries  $X_i$  et  $X_j$  que nous notons  $C_{i,j}$ . Cela définit une matrice de dimensions  $n \times n$  qui est la matrice**

de corrélation. Notons que sachant que cette matrice est symétrique, et qu'elle est identiquement égale à 1 sur la diagonale, elle va posséder  $\frac{n(n-1)}{2}$  nombres.

Ainsi, la procédure pour avoir le Minimal Spanning Tree est la suivante :

1. Ranger les nombres  $|C_{i,j}|$  par ordre décroissant ;
2. Soit  $i, j$  la première paire de ce rang. Tracer une ligne reliant  $S_i$  et  $S_j$ .
3. Tracer ensuite la ligne reliant la seconde paire de ce rang.
4. Répéter l'opération pour toutes les paires corrélées, sans tracer de boucle dans l'arbre.

Ainsi, nous obtenons un arbre simple contenant tous les actifs, de façon à ce que chaque paire soit liée uniquement avec ses actifs les plus corrélées.

Le fait que l'on travaille sur des données boursières a en statistique une importance. Il se trouve en effet que sur un ensemble d'actifs, il peut en avoir qui ne cotent pas d'un jour à l'autre, ou sur plusieurs jours consécutifs. Nous ne sommes donc pas, quand on considère la réalité du marché, sur un **système de temps continu**. Pour cette raison, les actifs sont considérés comme des modèles à saut. Il a donc fallu trouver des méthodes statistiques pour traiter ces données manquantes. Pour avoir un résultat juste de la corrélation, il faudra alors rajouter ce point d'étude.

### III. Application et résultats

Cet outil de calcul des corrélations a pour objectif premier de servir aux analystes qui travaillent sur les Global Macro (Stoxx 600 etc.). Il pourra aussi servir aux autres analystes évidemment. Ce travail a ainsi été fait en collaboration avec les analystes de BBSP qui seront les premiers à s'en servir une fois le projet opérationnel. Prendre en compte leur demande et leurs besoins était plus qu'essentiel.

Pour rappeler les difficultés qui avaient été notées devant ce projet, nous avons :

- Prétraitement et conformation des données qui peuvent parfois être aberrantes ou hétérogènes ;
- Calcul sur des séries massives. Il est possible que ces derniers ne puissent avoir lieu en temps réel, et devront être mémorisés. On craint donc ici un grand temps de calcul qui nécessiterait de laisser tourner le programme toute la nuit ;

- Optimisation de l'algorithme et des ressources de calcul, en vue d'un changement d'échelle pour aboutir à un prototype industriel.

Déjà et avant toute chose, je rappelle comme je l'ai dit plus haut dans le rapport, que je me suis occupée de la partie mathématique du problème. Il s'agissait d'implémenter le modèle avec les contraintes qui se présentaient, telles les données manquantes. Toute la partie affichage de ces résultats s'est faite par M. Hugues Leroy, l'un des informaticiens de l'équipe CREO.

Pour appliquer notre algorithme interne, il fallut faire des adaptations aux problématiques rencontrées. De plus, dans le travail de M. Rosario N. Mantegna, il est essentiellement question des rendements :  $X_i(t) = \ln(S_i(t)) - \ln(S_i(t-1))$  ;

avec  $S_i(t)$  le prix de clôture quotidien de l'actif financier  $i$  du jour  $t$  ;  $t = 1, \dots, m$ .

Seulement nous avons parallèlement, et pour la raison que je cite dans la partie **Evolution du travail en fonction des besoins de l'analyste** par la suite, utilisé la même démarche pour une évolution simple des prix de clôture de l'actif  $S_i$ . Dans ce cadre donc,  $X_i(t) = S_i(t)$ ;

avec  $S_i(t)$  le prix de clôture quotidien de l'actif financier  $i$  du jour  $t$  ;  $t = 1, \dots, m$ .

Le reste de la procédure reste le même.

### III.1 Les données

Nous avons réalisé les tests sur trois paniers différents : le CAC 40, le Stoxx 600 et un panier comportant des Commodities (Matières premières). Dans tous les cas, ce sont des données pris dans les bases de données de BBSP et qu'ils suivent au quotidien.



### **III.1.1 Le CAC 40<sup>15</sup>**

Il comprend les 40 actions cotées en continu sur le premier marché<sup>16</sup> parmi les 100 dont les échanges sont les plus abondants sur Euronext Paris qui fait partie d'Euronext, la première bourse européenne.

De la base de données de CREO nous extrayions les valeurs du CAC 40 cotées, à savoir à peu près 38 parmi les 40 (le CAC 40 regroupant les 40 premières capitalisations, l'ordre ou la composition de ce panier peuvent changer à long terme. Tous les actifs ne sont donc pas comparables sur la période d'étude.).

### **III.1.2 Le Stoxx 600 ou Dow Jones Stoxx 600<sup>17</sup>**

Il est constitué des 600 premières capitalisations boursières du marché européen (Allemagne, Autriche, Belgique, Danemark, Espagne, Finlande, France, Grèce, Irlande, Italie, Luxembourg, Norvège, Pays-Bas, Portugal, Royaume-Uni, etc.).

De la base de données de CREO nous extrayions les valeurs du Stoxx 600 cotées, à savoir à peu près 438 parmi les 600.

### **III.1.3 Les Commodities (Matières premières)**

Ce sont les cours des matières premières cotées. Elles sont réparties en différentes catégories : Matières premières d'origine végétale (blé, maïs, sucre, cacao, soja, caoutchouc) – métaux non ferreux (cuivre, aluminium, etc.) – métaux précieux (or, argent, platine) – énergie (pétrole, gaz) etc.

---

<sup>15</sup> CAC pour Cotation Assistée en Continu. C'est le principal indice boursier de la place de Paris.

<sup>16</sup> Le premier marché appartient à la bourse de Paris et regroupe les entreprises les plus importantes (c'est-à-dire ayant une capitalisation boursière d'au moins 1 milliard d'euros et ayant placé 25% de son capital dans le public ; Elle doit avoir publié ses comptes depuis trois ans).

<sup>17</sup> Un indice du même nom, l'indice Dow Jones Stoxx 600, est construit à partir des 600 plus grandes capitalisations boursières en terme de flottant, opérant dans 18 secteurs économiques définis par Dow Jones.

Ce panier est intéressant pour les analystes de BBSP car ils travaillent sur deux hypothèses différentes quant au comportement des Commodities :

- Lorsque la valeur des Commodities baisse, cela traduit une croissance mondiale en baisse, ce qui implique que les indices et actions auront une valeur plus basse.
- De l'autre côté, l'on pense que lorsque la valeur baisse, cela permet aux industriels de se fournir en matières premières moins chères et donc, cela favorise leur croissance, d'où le prix des actions qui montent.

En surveillant ces corrélations, les analystes pourront être aidés dans leurs travaux.

## III.2 Démarche adoptée lors du projet

### III.2.1 Traitements des données

Nous distinguons là les données manquantes et les valeurs aberrantes.

#### Traitement des valeurs manquantes (sauts)

Nous traitons les sauts différemment selon qu'il s'agit du calcul des corrélations avec les rendements, ou du calcul des corrélations sur les cours des prix de clôture seuls.

- Dans le premier cas, nous avons adopté la démarche qui est faite sous Bloomberg.

Soient  $S_i$  et  $S_j$  deux actifs présentant des sauts et dont on cherche à calculer le coefficient de corrélation de leurs rendements respectivement  $X_i$  et  $X_j$ .

**Supposons que  $S_i$  présente  $p$  jours de non cotation successifs, à partir de la date  $t_a$  jusqu'à la date  $t_b$  (reprise des cotations). Notons  $T = t_b - t'_a$ , avec  $t'_a = t_a - 1$ , jour précédant à  $t_a$ .**

***Tant que le temps est continu,***

***Faire*** 
$$X_i(t) = \ln(S_i(t)) - \ln(S_i(t-1));$$

$$X_j(t) = \ln(S_j(t)) - \ln(S_j(t-1));$$

***Si  $t = t_a$  (jour de non cotation),***

$$X_i(t_a) = \frac{\ln(S_i(t_b)) - \ln(S_i(t'_a))}{\sqrt{T}}$$

$$X_j(t_a) = \frac{\ln(S_j(t_b)) - \ln(S_j(t'_a))}{\sqrt{T}}$$

**Ensuite, on reprend à la date  $t_b : t = t_b$ .**

- Si nous sommes sur des corrélations simples et j'entends par là, le calcul des coefficients de corrélation simplement avec les cours de clôture des actifs, nous retiendrons la méthode de la **suppression des sauts**. Nous avons choisi celle-là car elle se rapproche le plus des résultats exacts lorsqu'on a peu de jours qui ne cotent pas, ce qui est souvent le cas.

**Soient  $X$  et  $Y$  deux actifs.**

**S'il existe un saut à la date  $i$  pour l'actif  $X$ ,**

**Supprimer  $X_i$  et  $Y_i$ .**

On procède donc là par paire d'actifs, le but étant d'avoir une base de comparaison similaire.

**PS :** Précisons que les sauts sont marqués en référence à l'ensemble des dates qui cotent dans le panier. C'est ensuite qu'on procède paire par paire au traitement des sauts, pour avoir une base de comparaison similaire pour chaque paire d'actifs.

- Néanmoins, une limite a été établie au niveau de mon programme. Elle prévoit de donner un coefficient de corrélation nul pour si l'un des actifs ne cote pas suffisamment sur la période donnée.

### **Traitement des valeurs aberrantes**

Déjà, il existe un prétraitement des données par les outils de collecte des données internes de BBSP, sur ACE par exemple. Dans le cadre de mon étude, je n'ai pas trouvé d'aberration.

Ce point est assez problématique à traiter. Une solution à laquelle j'ai pensé pour identifier d'éventuels points aberrants était :

**Soit  $(S_i(t))_{1 \leq t \leq m}$  un vecteur contenant les prix de clôture quotidiens sur la période  $1 \leq t \leq m$ .**

***Pour chaque prix de clôture, le comparer avec la moyenne des deux points l'entourant. Si ce point est  $k$  fois supérieur à cette moyenne, le remplacer par cette moyenne.***

Seulement, des pics existent sur les cours des marchés financiers et il arrive que ces pics soient très importants d'un jour à l'autre en fonction de la volatilité du marché. Cette méthode risque donc de supprimer une information importante qui faussera le calcul de la corrélation.

Au final, on va travailler en considérant qu'il n'existe pas de valeurs aberrantes. Ce sont des cas très rares et si nous les traitons, cela représentera un coût important en termes de calcul.

### III.2.2 Calcul empirique de la corrélation statistique

La formule qui sera utilisée pour le calcul du coefficient de corrélation entre deux vecteurs  $X$  et  $Y$ , sur une période  $t = 1, \dots, m$  sera :

$$\frac{\sum_{i=1}^m (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^m (X_i - \bar{X})^2 (Y_i - \bar{Y})^2}}$$

Avec  $\bar{p} = \frac{\sum_{i=1}^m p_i}{m}$ , la moyenne empirique de  $p$ .

C'est la même formule que celle de l'auteur mais réécrite différemment. C'est également la traditionnelle formule de calcul de ce coefficient en mathématiques.

### III.2.3 Représentation graphique

Cette partie a été effectuée par M. Hugues Leroy. Il avait en sa possession un script en langage Ruby qui soit capable d'intégrer un outil de visualisation : Graphviz.

Son script en Ruby va prendre en entrée la matrice de corrélation d'un ensemble d'actifs. Il va ensuite adapter cette matrice à la visualisation à savoir, attribuer pour chaque coefficient de corrélation, la paire d'actifs qui lui correspond. Un fichier .dot est ainsi construit et sera transmis à Graphviz qui sait le lire.

On remarquera ici que **nous n'utilisons pas la matrice des distances  $D$**  telle que décrite dans l'algorithme de l'auteur. La raison est simple.

Graphviz va se charger de répartir dans l'espace, les actifs répondant au critère de sortie (seuil) qui lui est donné. Plus la corrélation entre deux actifs est forte et plus les actifs sont proches. On se rapproche là alors de la pensée de départ de l'auteur et du principe du Minimal Spanning Tree, même si l'application n'en est pas strictement rigoureuse.

Cette méthode du Minimal Spanning Tree est également intégrée sous Matlab mais le rendu visuel bien que pouvant servir pour un usage de tests personnels, n'est pas assez convaincant pour être utilisé dans l'entreprise.

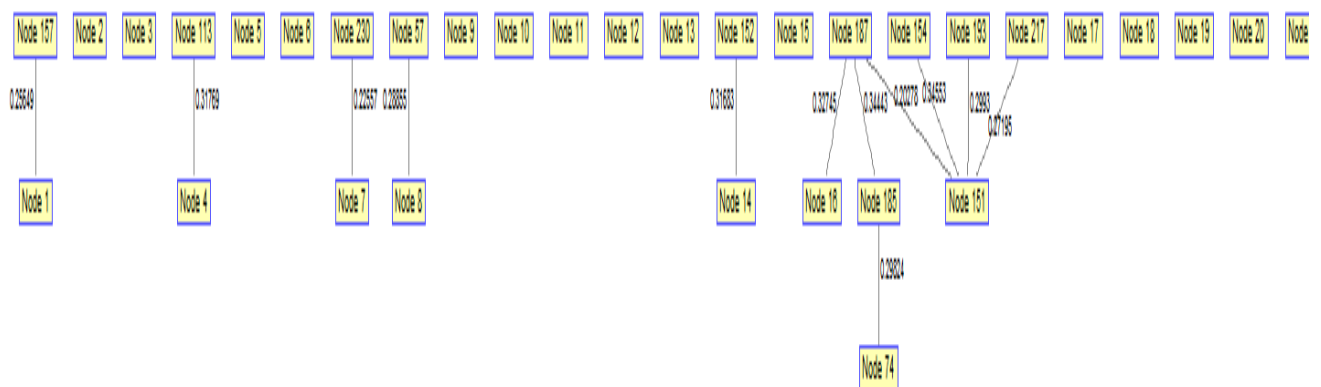


Figure 4. Modèle d'un minimal spanning tree sous Matlab

Par contre Graphviz offre d'énormes avantages et pourra aisément être incorporé à une application Web par exemple lorsque tout sera le processus sera automatisé.

Si nous pouvons relever une limite (encore non résolue) à cela, ce serait que du fait que Graphviz optimise l'espace pour lier au mieux toutes les paires d'actifs corrélées selon le seuil défini, un actif peut changer d'emplacement si on glisse dans le temps en étudiant la corrélation d'un jour à un autre (sur la même période de temps). *Cf exemples dans la suite.*

Cela peut donner visuellement l'impression que le graphe change beaucoup d'un jour à l'autre, ce qui signifierait qu'il y a peu de corrélations stables. Il faudra donc faire attention à cela à bien relever les actifs qui s'affichent sur le graphe.

### III.3 Evolution du travail

#### III.3.1 Point de lancement

Après le premier mois consacré essentiellement à l'étude du sujet et à la prise de connaissance des documents fournis au départ pour effectuer le stage, ce fut le point de départ d'une longue phase expérimentale.

N'ayant pas de contrainte quant au langage informatique sous lequel implémenter le modèle pour obtenir la matrice de corrélation, j'ai décidé de la faire sous **Matlab** pour sa rapidité de calcul matriciel. Comme énoncé lors de la présentation du projet, les tests se sont faits sur des données massives pour pouvoir notamment estimer le temps de calcul.

J'ai notamment réalisé des tests sur le CAC 40 (37 à 40 des actifs du CAC 40) pour commencer, puis sur le Stoxx 600 (231 à 440 des actifs du Stoxx) et plus tard, sur les Commodities (Matières premières suivies par BBSP). Ces données étaient au départ extraites sous des fichiers Excel via la base de données Bloomberg. Je marquais les sauts en leur attribuant le nombre -1 que mon programme Matlab reconnaissait mais ce système d'extraction a été provisoire car trop fastidieux pour des données plus nombreuses.

De l'autre côté, Hugues Leroy avait un script en langage Ruby qui lui, génère un fichier .dot pouvant être visualisé par Graphviz. C'est cette alternative qui a été trouvée pour construire le graphe. La première figure que je tenais à montrer qui a été générée est la suivante.

## Résultat initial

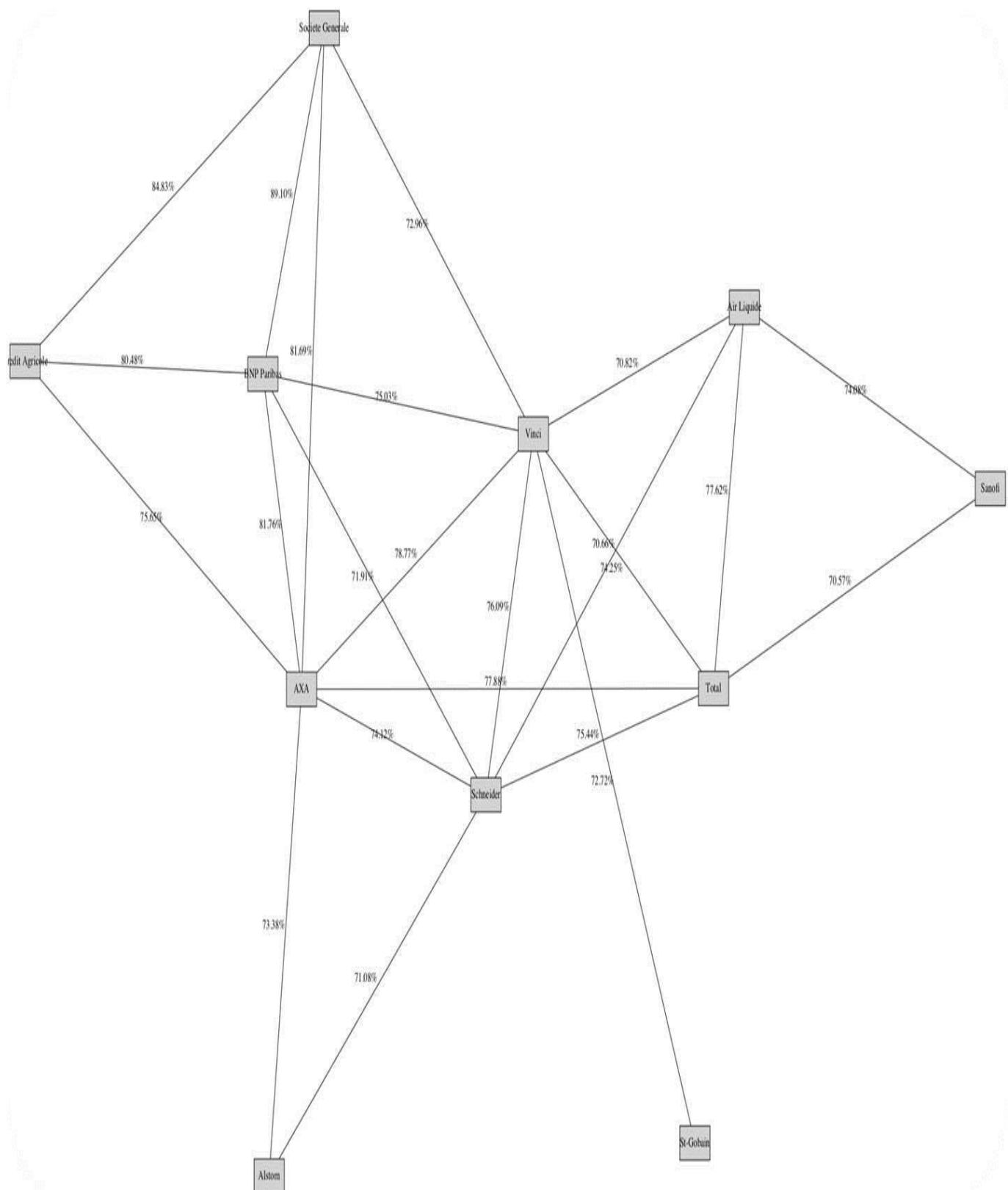


Figure 5. Corrélations sur les rendements des actifs du CAC 40 du 02/05/2012 au 03/05/2013 ; Seuil défini à 0.7

C'est le premier test d'affichage que nous avons fait. Nous avons traité les sauts en les supprimant. Dans cette première génération, j'ai relevé quelques erreurs au niveau de mon programme qui faussent quelque peu le résultat des corrélations. Mais le but ici est de montrer à quoi ressemblait le premier résultat.

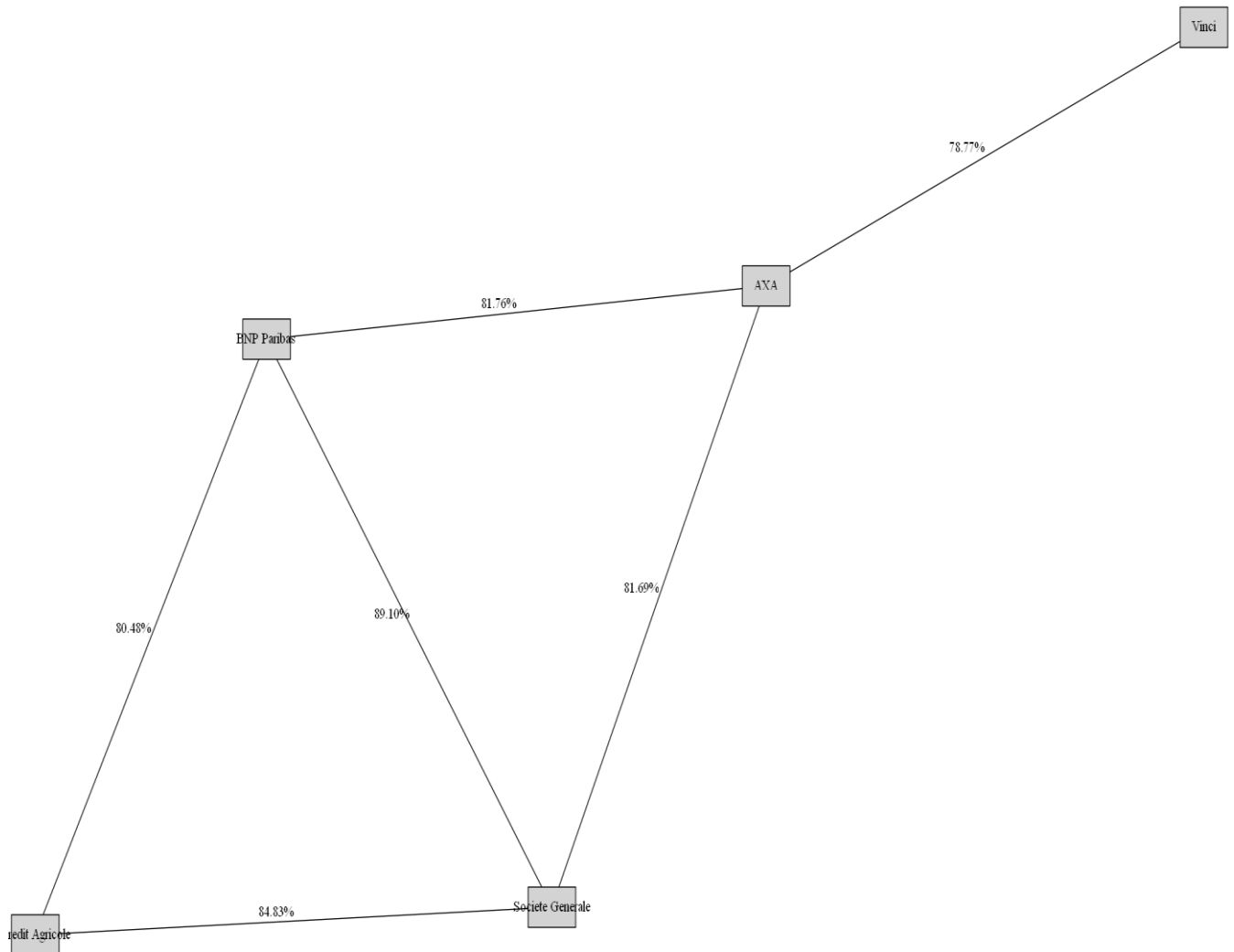


Figure 6. *Corrélations sur les rendements des actifs du CAC 40 du 02/05/2012 au 03/05/2013 ; Seuil défini à 0.78*

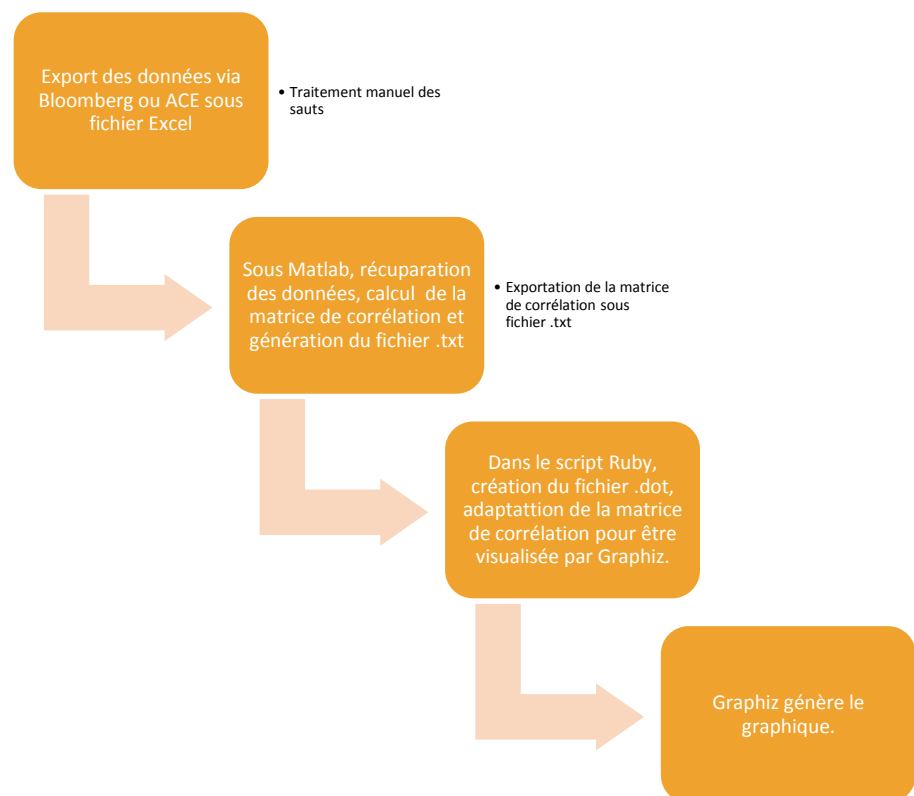
C'est à partir de là que nous avons compris que nous avions les outils nécessaires pour réaliser le démonstrateur et donc, mener le projet à terme. L'affichage était satisfaisant pour toute l'équipe car montrant bien les liens entre les actifs avec éventuellement, le résultat chiffré des coefficients de corrélation significatifs. Nous avons donc décidé de continuer avec ce système d'affichage.

Nous sommes là au 17 Mai 2013. Nous étions encore sur un outil assez « démantelé ». La seule chose que l'on pouvait modifier directement en ligne de commande était le seuil.



Les procédures séparées de traitement étaient au départ:

1. Récupération des sauts et marquage manuel des sauts (que j'ai effectué sur le CAC 40 uniquement au départ car trop fastidieux pour les plus gros paniers).
2. Récupération du fichier de données ainsi traité et calcul par mon programme de la matrice de corrélation.
3. Export de la matrice de corrélation ainsi générée sous fichier .txt pour être adaptée par le script Ruby qui va non attribuer à chaque corrélation, la paire d'actif qui lui correspond, puis générer un fichier .dot, description sémantique du graphe qui sera lue et affichée par Graphviz.



Il existait ainsi encore à ce niveau beaucoup de travail manuel. Etant donné que je devais travailler sur des paniers plus gros comme le Stoxx 600, je devais changer ce système.

### III.3.2 Evolution des moyens techniques

A mi-parcours, un outil d'extraction automatique des données a été mis à ma disposition. Ce programme avait été développé par CREO pour les besoins d'un des analystes de BBSP. Hugues l'a ensuite intégré au script Ruby afin que je puisse avoir un fichier complet

des données extraites où il y avait les sauts marqués automatiquement. Cela a été un gain de temps considérable surtout pour les tests sur le Stoxx 600. Je pus ainsi avoir un fichier .csv avec les sauts marqués que j'utilisai dans Matlab pour le calcul de la matrice des corrélations.

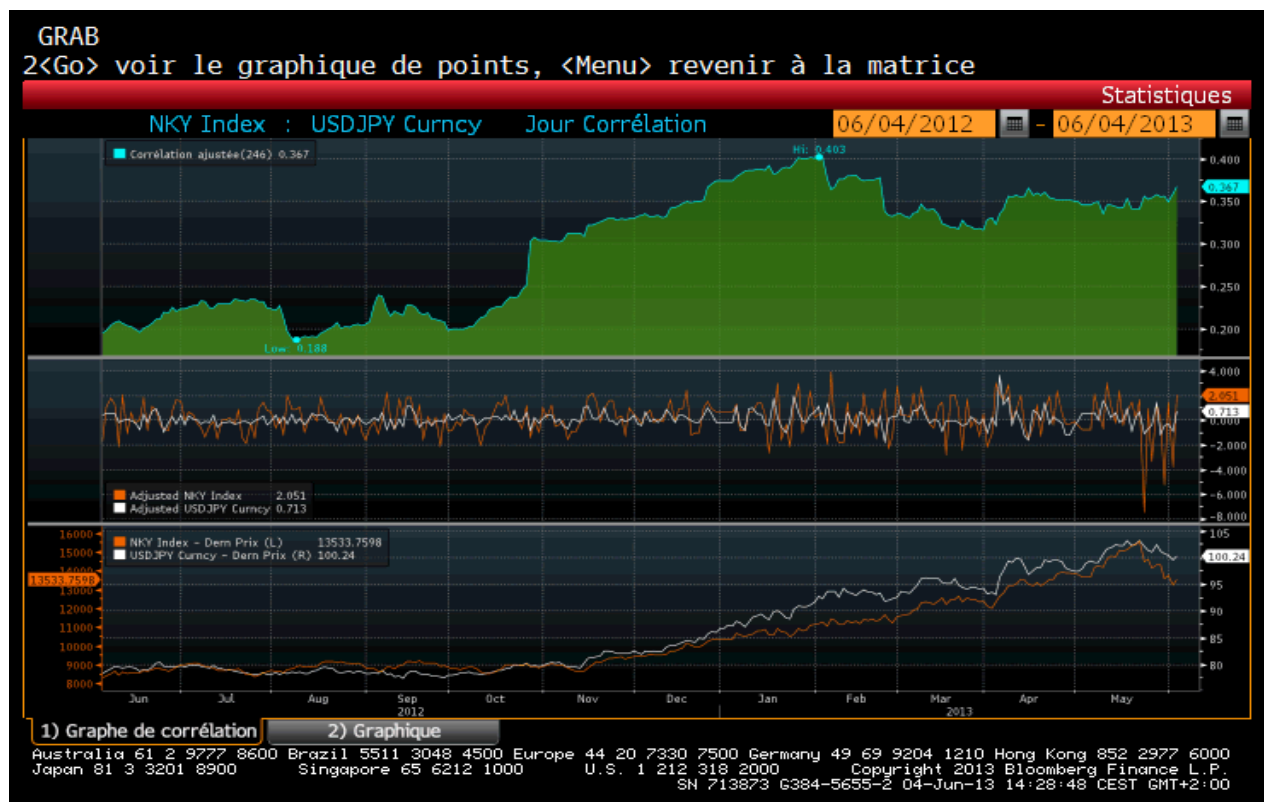
### III.3.3 Une évolution en fonction des besoins des analystes

Il a paru dès le début, important de garder en vue la possibilité de modifier des détails directement en ligne de commande : *créer son panier et définir la période d'analyse des données*, ainsi que *définir le seuil d'affichage des corrélations par exemple*.

Ce projet devant servir aux analystes de BBSP, il était important qu'ils définissent leurs attentes par rapport à l'outil.

#### Réévaluation du modèle.

L'idée de départ était de travailler sur les rendements des actifs financiers. Mais un des analystes nous a soumis le problème ci-après. Il se demandait comment malgré les cours qui évoluent à peu près de façon similaire sur le graphique qui va suivre, la corrélation ne pouvait être que de 0.367. Voir l'affichage qui suit.



Il s'agit ici de la comparaison du **NKY Index** et de **l'USDJPY Curncy**. Ce sont des données Bloomberg.

Sur le graphe du bas, on voit bien que ces deux actifs sont très corrélés entre eux, mais sur celui du haut, la corrélation affichée est de 0.367.

Après avoir récupéré ces données et les avoir testées avec mes codes, cette corrélation de 0.367 est celle calculée sur les rendements des deux actifs. Si on calcule simplement la corrélation des deux cours boursiers sur la même période (valeurs de clôture journalières), on trouve 0.98, ce qui est plus crédible vue le graphe du bas.

**Remarque** : instinctivement, on pourrait penser qu'un coefficient de corrélation calculé sur les rendements des actifs financiers donnera à peu près le même résultat que si on calculait ce même coefficient avec l'évolution simple des cours de clôture de ces mêmes actifs, mais cet exemple est un bon contre-exemple.

En effet, le rendement consiste à mesurer la performance d'un actif ou d'un produit. Il est un critère de sélection pour un agent économique au même titre que la sécurité ou la liquidité de son investissement.

Le rendement financier correspond à la rentabilité d'une somme placée. Le rendement d'un placement financier se calcule différemment selon le type de placement. Certains rendements financiers (livret épargne) sont connus à l'avance alors que d'autres rendements financiers ne seront connus qu'à posteriori (actions en bourse). *Voir la formule du calcul du rendement exploitée dans l'étude de Rosario N. Mantegna.*

Nous ne pouvons donc pas faire un rapprochement évident entre un coefficient de corrélation déduit avec le rendement de l'actif et un coefficient de corrélation déduit par l'évolution de ce même cours. Il n'a pas de proportionnalité entre les deux.

Dans notre cas, le problème qui se pose aux analystes avec ce calcul des corrélations sur le rendement de l'actif est que cela n'est pas représentatif visuellement de l'évolution des cours (voir exemple ci-haut), et eux, sont très visuels dans leur travail.

**Ainsi, nous avons abandonné le calcul des corrélations au profit du calcul simple des corrélations des cours journaliers.**

### **Autres demandes des analystes.**

- Selon le secteur d'activité, les analystes préfèrent soit les courtes et moyennes périodes (Equities), soit les plus longues notamment pour les Global Macro.
- Les seuils aussi ont été définis par les analystes. Entre  $[-0.3 ; 0.3]$  on considère les actifs comme étant non corrélés. A partir d'un seuil supérieur à 0.7, on commence à considérer les corrélations comme étant importantes. Inversement, un seuil inférieur à - 0.7 montre des actifs inversement corrélés qui peuvent présenter un intérêt.
- Il est aussi important pour les analystes, qu'une alerte soit lancée lorsqu'un lien de corrélation se brise ou se crée (selon le seuil).

Autant de facteurs donc à prendre en compte dans la conception de l'outil.

### **III.3.3 Evolution au niveau du démonstrateur.**

Nous sommes passés avec le script Ruby, rapidement d'une version en noir et blanc du graphique, à un affichage en couleurs. Celui-ci permet de rendre visible l'affiliation sectorielle (une couleur pour chaque secteur : banques, assurances, énergétique etc.) de chaque actif et le type de lien détecté (rouge pour les corrélations inverses, vert pour les corrélations positives).

Ci-dessous la représentation des corrélations (en rendement) entre les actifs du CAC 40 au seuil 0.7.

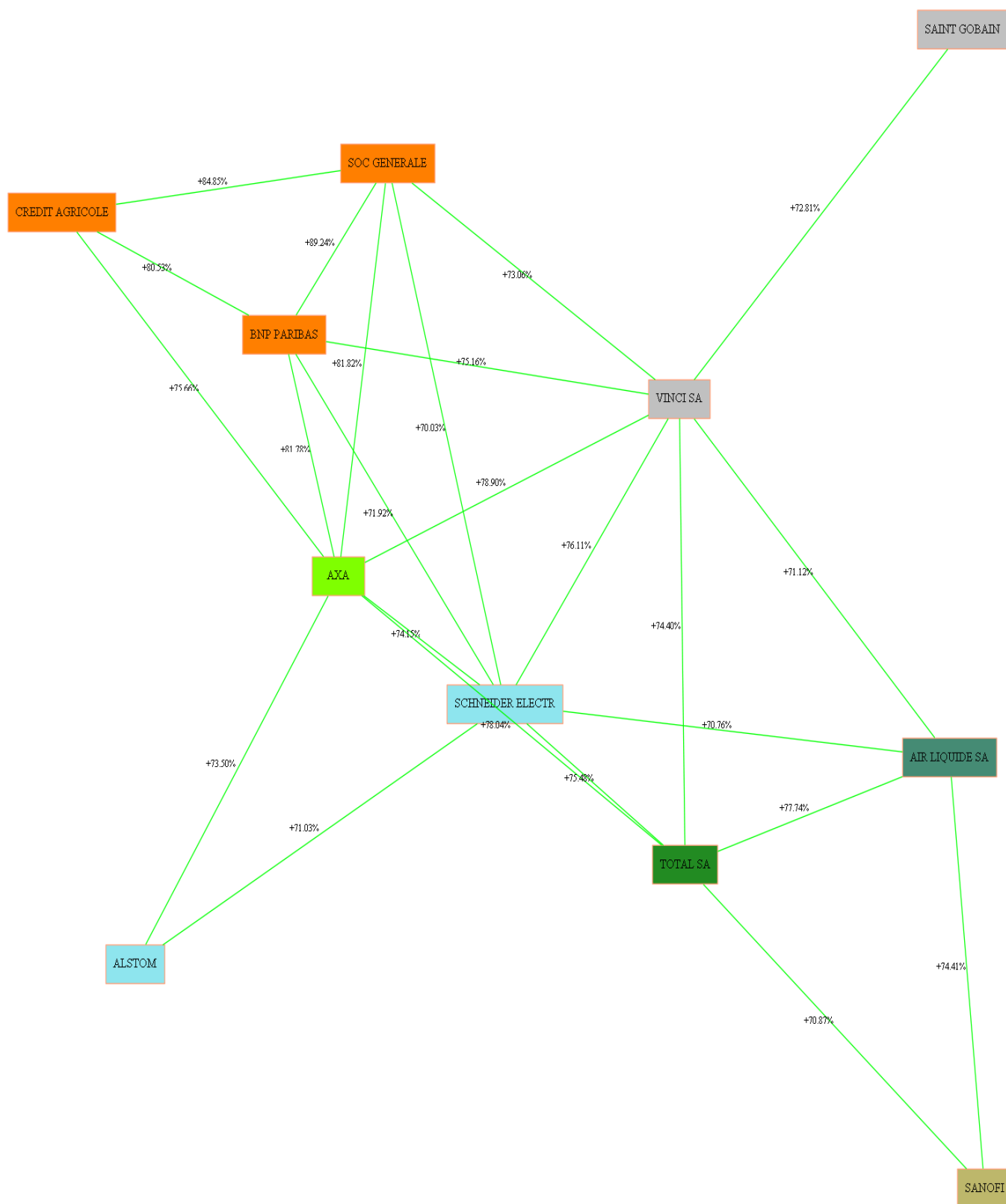


Figure 7. Calcul (en rendements) des corrélations sur le CAC 40, période du 02/05/2012 au 03/05/2013, seuil 0.7.

En augmentant le seuil à 0.8, on a la figure qui suit :

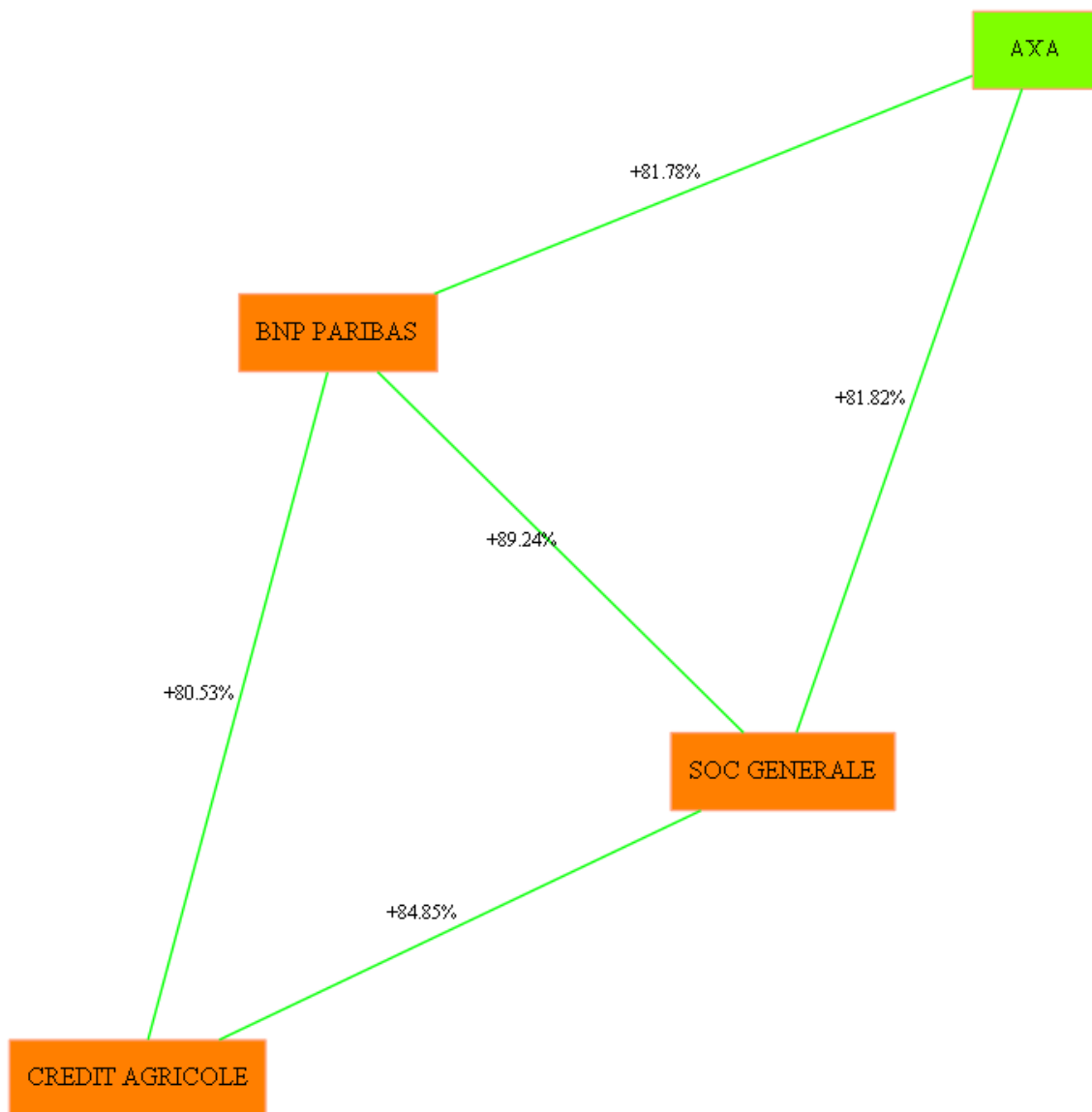


Figure 8. Calcul (en rendements) des corrélations sur le CAC 40, période du 02/05/2012 au 03/05/2013, seuil 0.8

### Interprétation

Sur tous les graphes du CAC 40, nous observons le triangle **Crédit agricole – Société Générale – BNP Paribas** (trois valeurs bancaires). Concrètement en investissement, constituer un panier avec ces trois éléments donnera soit de gagner trois fois, soit de tout perdre. D'où le choix de diversifier le panier pour ne pas prendre trop de risques.

La visualisation des graphiques des cours confirme le résultat des corrélations obtenues.

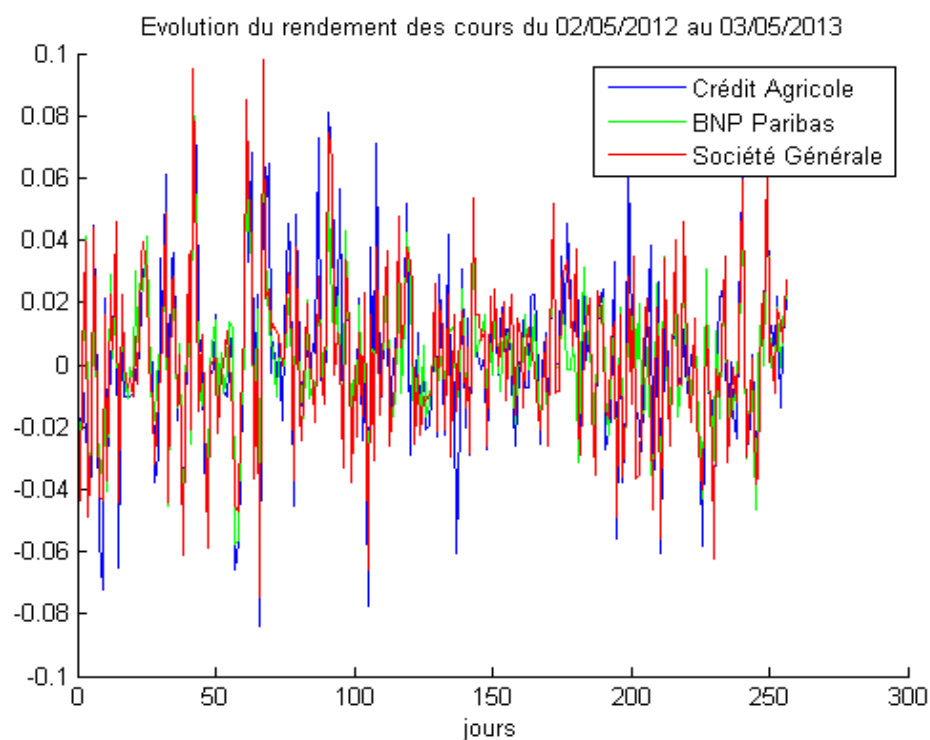


Figure 9. Evolution du rendement des cours des actifs CA – BNP – SG du 02/05/2012 au 03/05/2013, soit 257 jours.

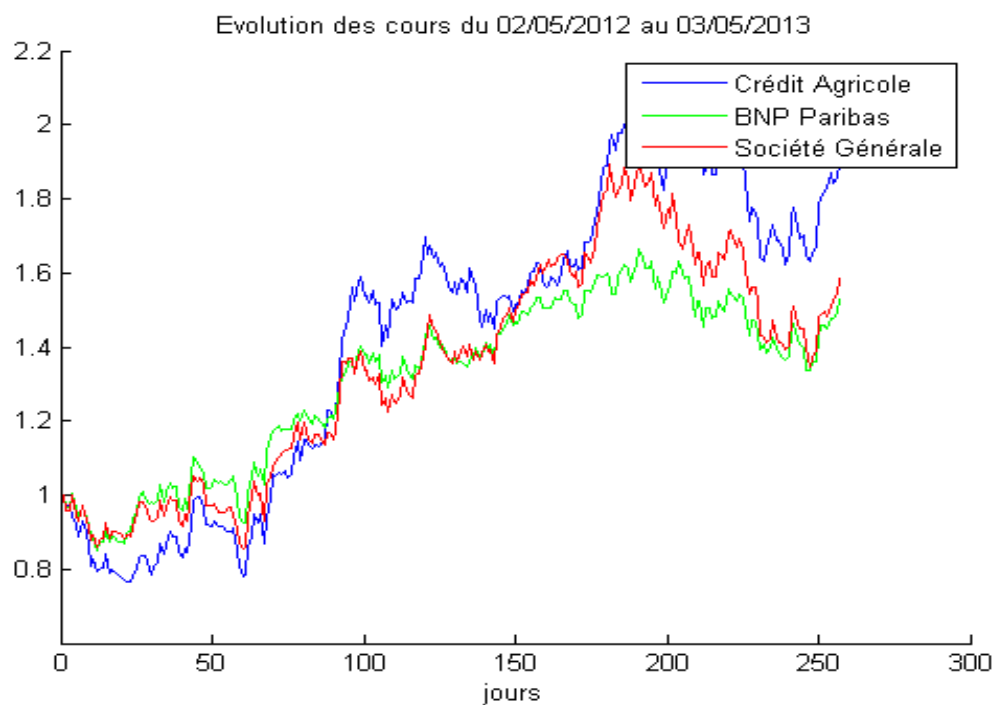


Figure 10. Evolution des cours des actifs CA – BNP – SG du 02/05/2012 au 03/05/2013, soit 257 jours.

**PS :** On observe naturellement des corrélations plus élevées lorsque nous évaluons les cours de clôture simples au lieu des rendements de ces mêmes cours. Rappelons que les analystes techniques de BBSP préfèrent travailler avec les corrélations des cours simplement car cela est plus parlant visuellement pour eux. Les résultats concordent en effet avec l'évolution graphique des cours avec lesquels ils travaillent.

Ci-dessous la visualisation des corrélations des cours des actifs du CAC 40 sur la même période que celle du haut, du 02/05/2012 au 03/05/2013.

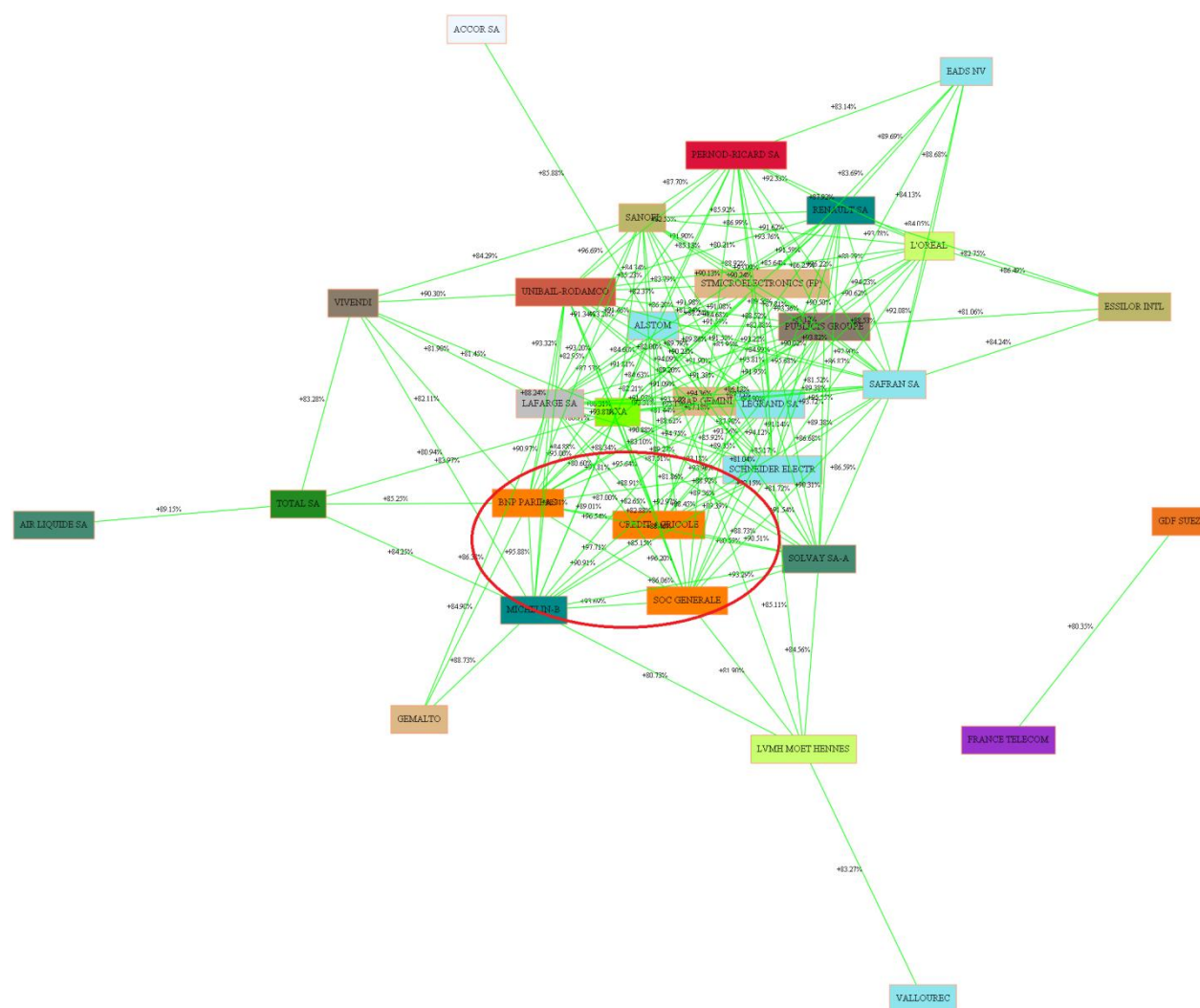


Figure 11. Calcul des corrélations des cours (prix de clôture) sur le CAC 40, période du 02/05/2012 au 03/05/2013, seuil 0.7.

Il y a déjà plus d'actifs représentés. Filtrons avec un seuil à 0.95 pour mieux identifier le triangle Crédit agricole – BNP Paribas – Société Générale.



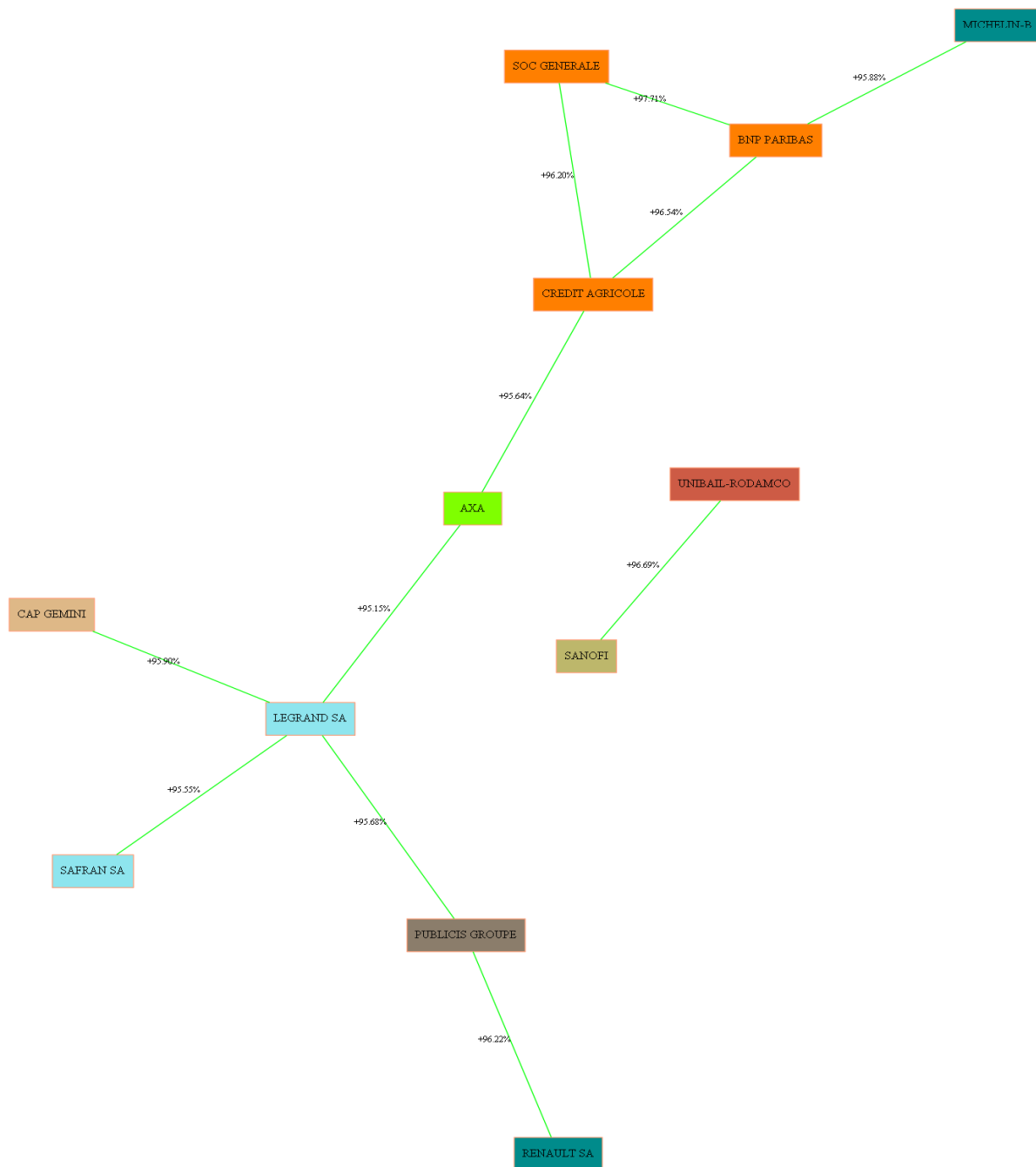


Figure 12. Calcul des corrélations des cours (prix de clôture) sur le CAC 40, période du 02/05/2012 au 03/05/2013, seuil 0.95.

**PS :** Les deux graphiques précédents portent sur 37 des 40 actifs du CAC 40. Les figures du CAC 40 qui suivent, portent sur les 40 actifs. C'est pourquoi on voit apparaître d'autres actifs tels que Carrefour.

A mi-parcours, des options ont été ajoutées au script en plus du seuil. On peut donc les appeler via la boîte de commande. On a notamment :

- - t : pour définir le seuil ;
- -- help : pour voir les options disponibles ;
- -- focus : pour visualiser les corrélations en étant focalisé sur un actif défini par l'utilisateur. Exemple ci-après.

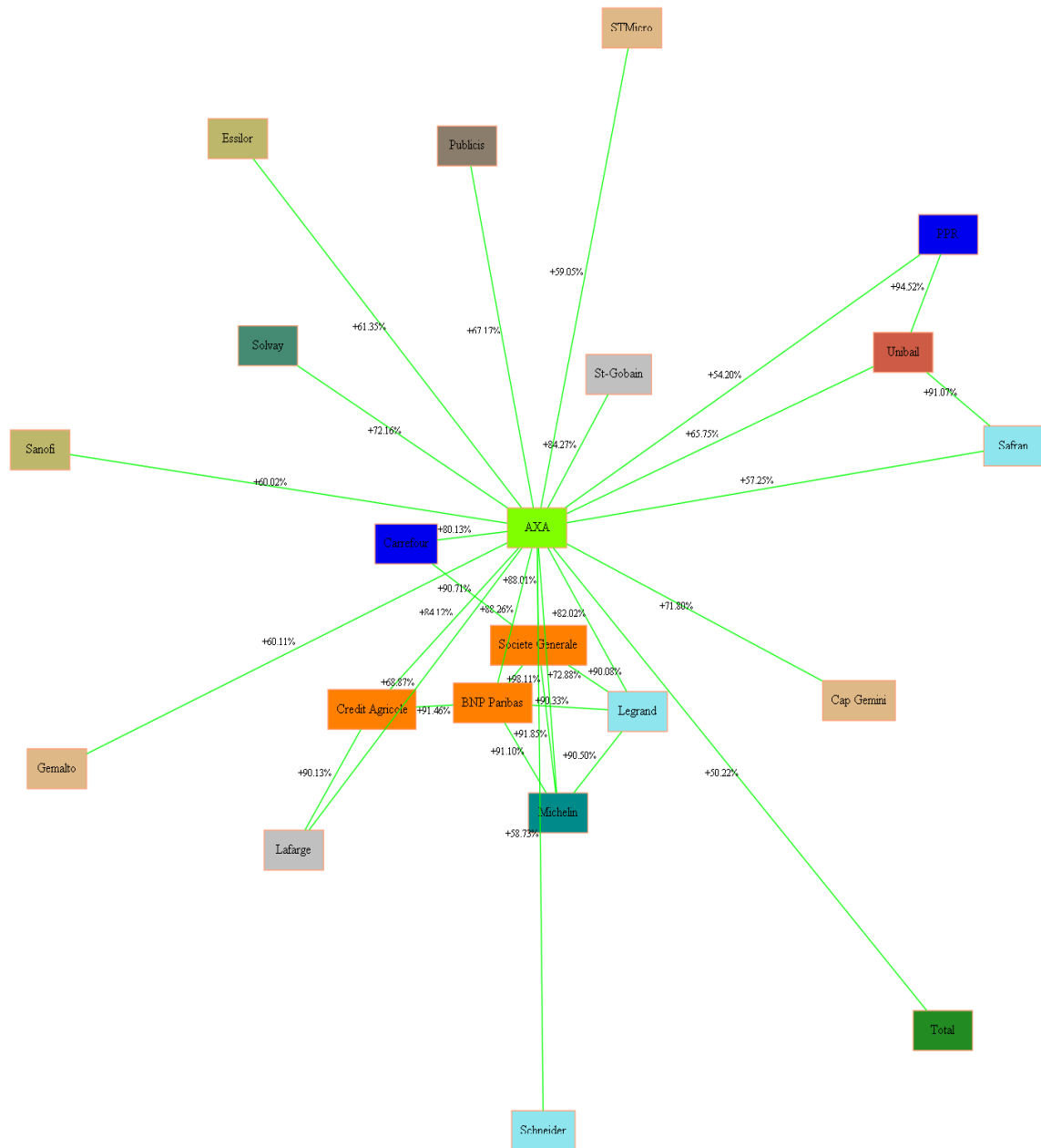
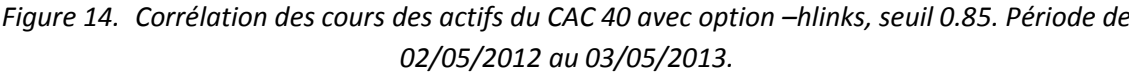


Figure 13. Corrélation des cours. Focus sur AXA dans le panier du CAC 40, période de 02/05/2012 au 03/05/2013, seuil 0.9.

**PS :** La figure qui précède a été faite avec les 40 actifs du CAC 40. C'est pourquoi on y voit d'autres actifs comme Carrefour (secteur en bleu foncé). Les autres figures sont faites avec 37 des 40 actifs du CAC 40.

- -- hlinks : qui donne pour chaque nœud présent au-delà du seuil demandé, les trois plus fortes liaisons.



- Ces différentes options sont très utiles notamment pour les gros paniers tels que le Stoxx 600.

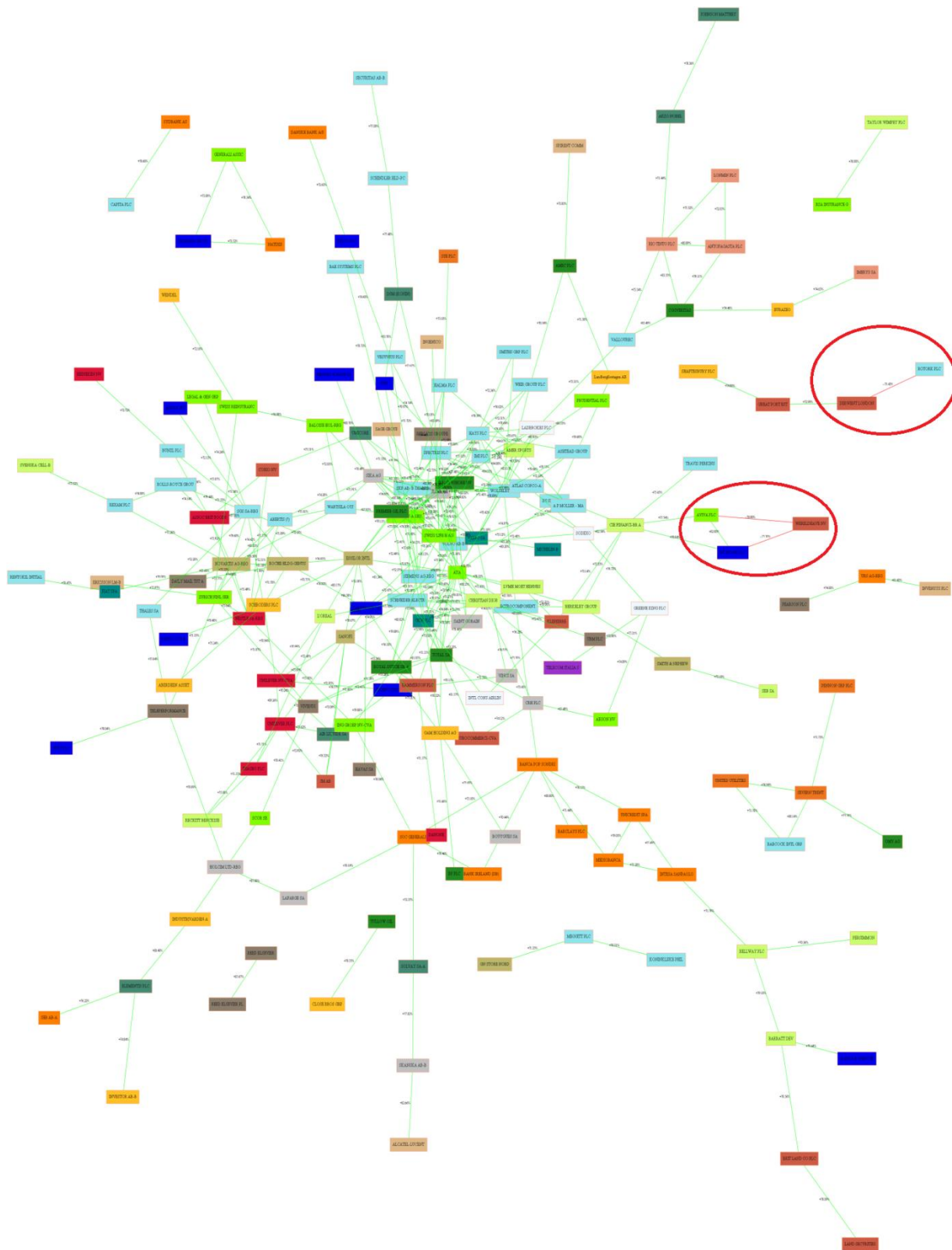
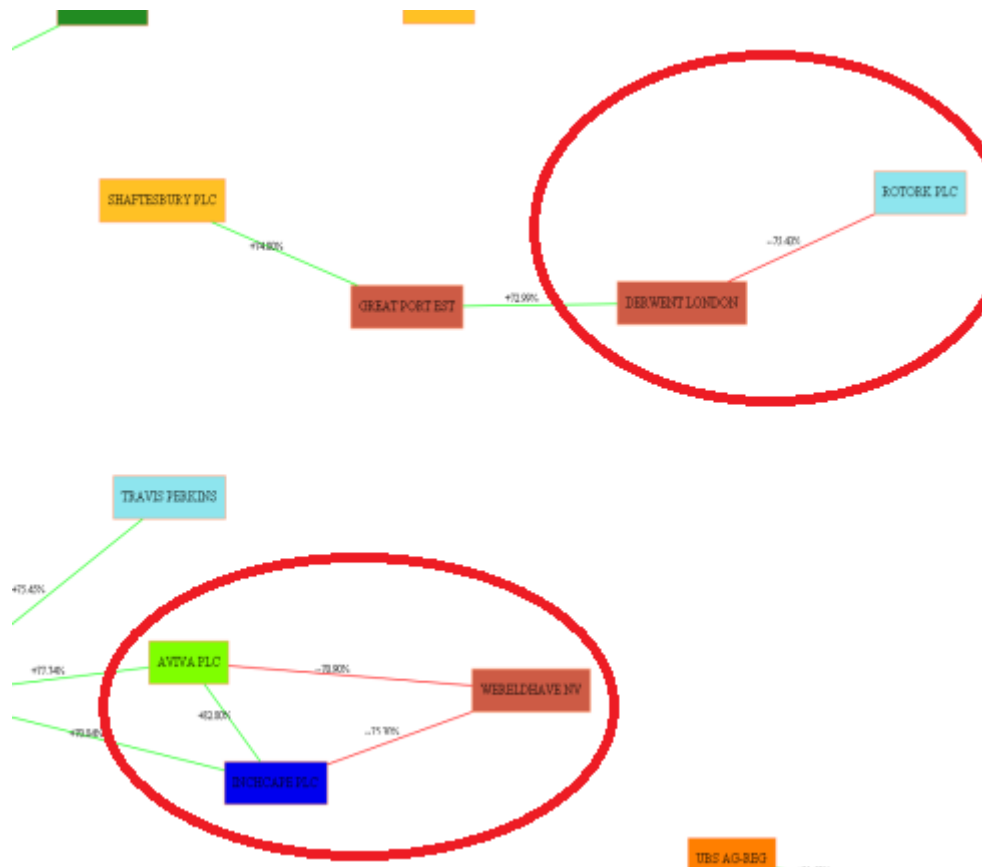


Figure 15. Corrélation en rendement des cours de 231 actifs du Stoxx 600, seuil 0.7, période du 19/04/2013 au 17/05/2013, soit 20 jours.



On voit ici apparaître une corrélation négative de -0.71 entre Rotork PLC<sup>18</sup> et Derwent London<sup>19</sup>. Cette corrélation illustre de fait deux stratégies de placement opposées : spéculative (le boom des pays émergents) ou plus défensive (l'immobilier londonien).

En filtrant les données avec un seuil de détection plus élevé pour la même période c'est déjà un peu plus clair avec l'apparition de clusters (groupes).

<sup>18</sup> Rotork est une société industrielle qui réalise une grande partie de son activité à l'international, notamment dans le secteur parapétrolier.

<sup>19</sup> Derwent London est un leader britannique de l'investissement immobilier.



Figure 16. Corrélations en rendement des cours de 231 actifs du Stoxx 600, seuil 0.8, période du 19/04/2013 au 17/05/2013, soit 20 jours.

Avec les options, je fais un focus sur AVIVA. Cela donne :

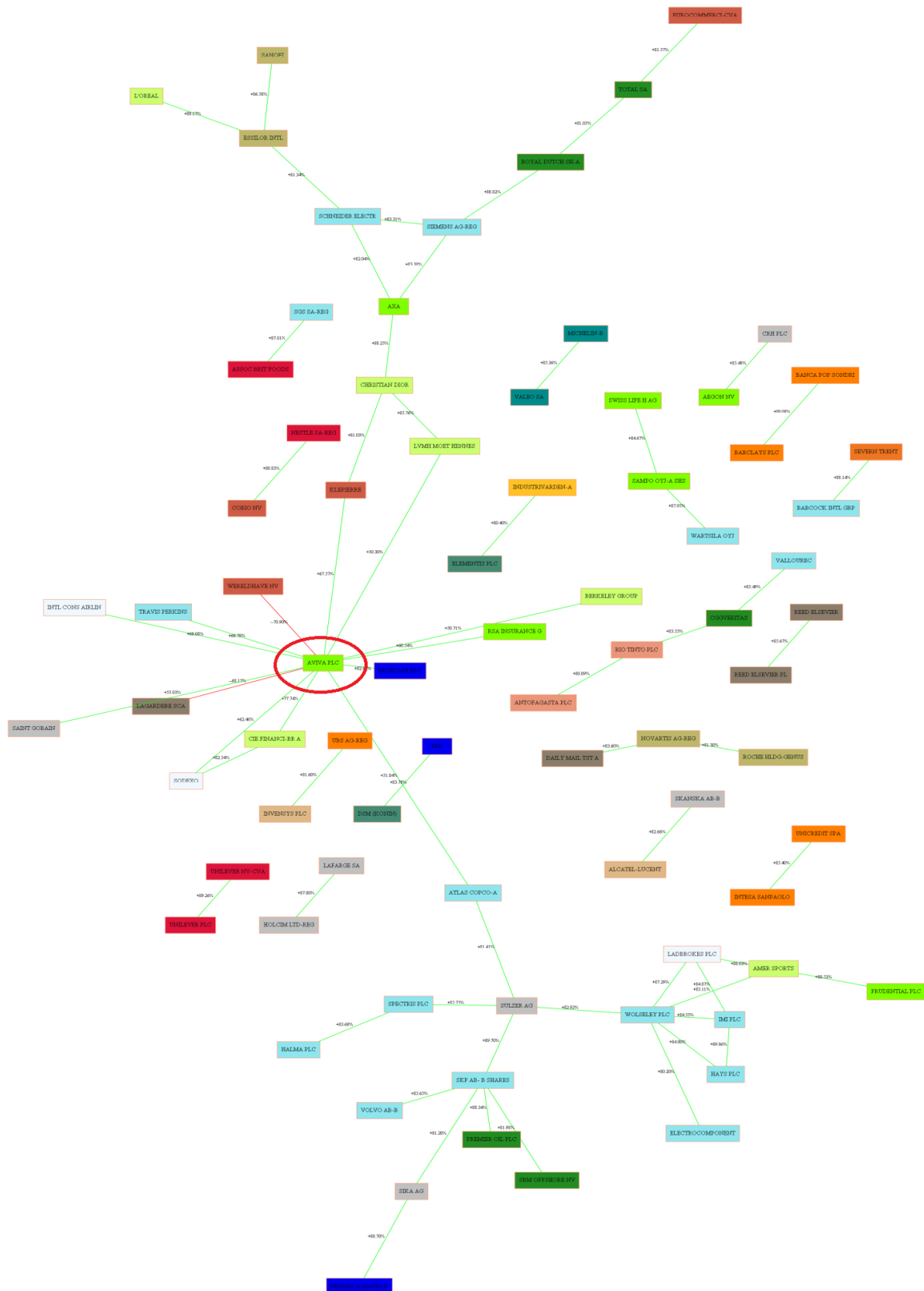


Figure 17. Corrélations en rendement des cours de 231 actifs du Stoxx 600, focus sur AVIVA, période du 19/04/2013 au 17/05/2013, soit 20 jours.

Un autre choix au lieu des visualisations ponctuelles est de regarder l'évolution des corrélations d'une période à une autre. Ci-après une illustration de cela avec l'évolution du CAC 40 sur 3 ans à peu près, année après année au seuil 0.85.

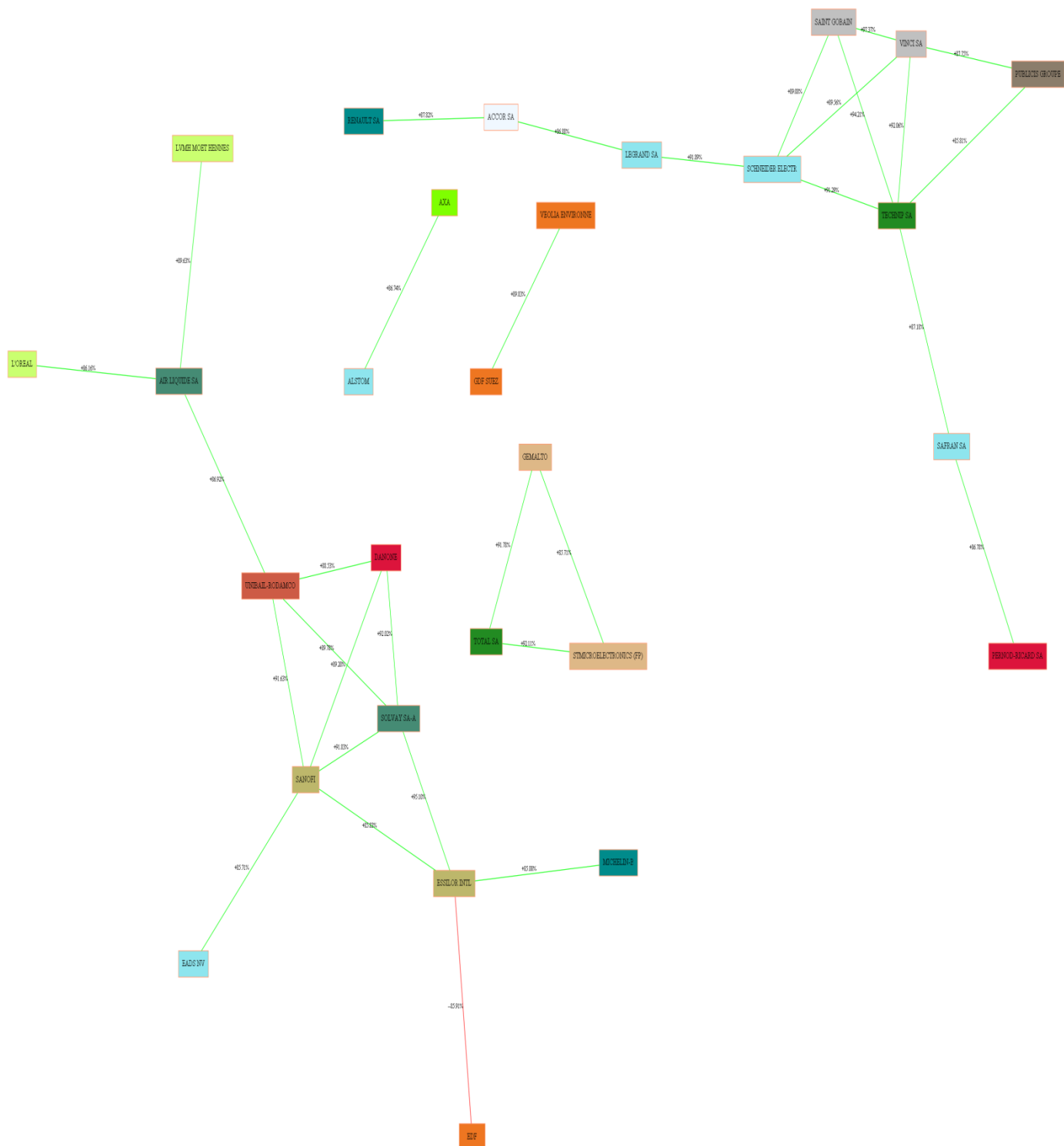
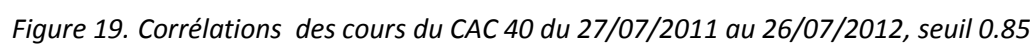


Figure 18. Corrélations des cours du CAC 40 du 27/07/2010 au 26/07/2011, seuil 0.85





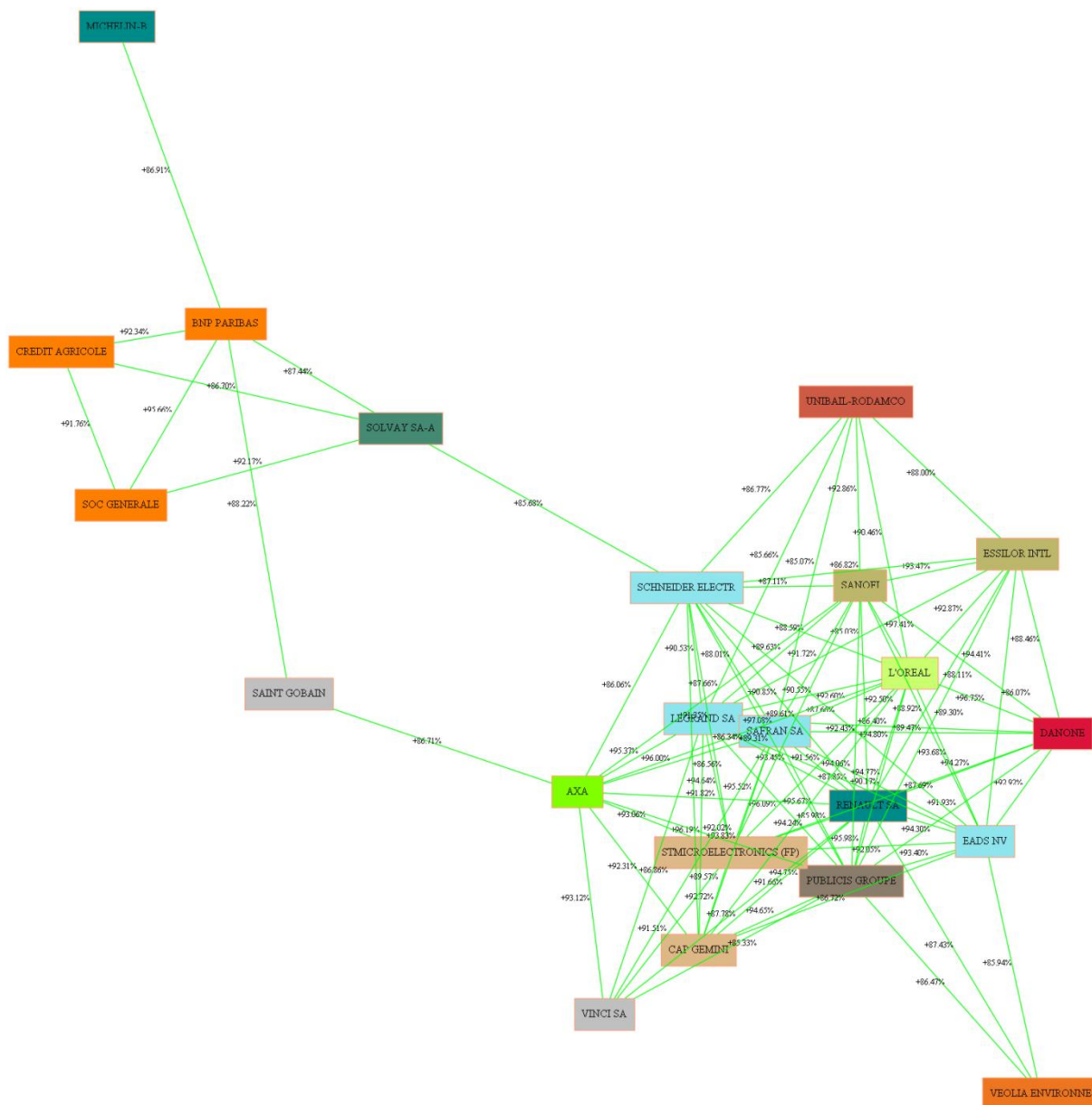


Figure 20. Corrélation des cours du CAC 40 du 27/07/2011 au 24/07/2012, seuil 0.85

Le triangle Crédit Agricole – Société Générale – BNP Paribas apparait sur ce dernier graphique. On peut y voir aussi que des corrélations se sont faites et défaits.

A ce niveau du rapport, le démonstrateur donne encore ce que je peux appeler, des affichages séparés, c'est-à-dire qu'il n'y a pas de ressemblance spatiale entre l'affichage d'une période à une autre.

## **IV. Perspectives d'évolution de l'outil**

### **IV.1 Autres idées à exploiter**

Dans l'optique de rendre notre outil plus fonctionnel, nous avons pensé explorer de nombreuses pistes parmi lesquelles les corrélations décalées et les cycles de corrélations. Il y a aussi les évolutions d'un point de vue informatique de l'outil.

#### **Les corrélations décalées.**

Il est en effet possible qu'une corrélation se manifeste avec un décalage de  $n$  jours, semaines ou mois. Il faudra ainsi  $n$  jours pour qu'un actif  $X$  puisse réagir en réponse à/ en concordance avec un actif  $Y$ . Par exemple, l'actif  $X$  va suivre les variations de  $Y$  mais avec 1 mois de décalage. L'investisseur peut alors anticiper la hausse de l'actif  $Y$ .

La recherche de ces corrélations décalées pourra donc être incorporée dans l'outil lors de sa mise en place. J'ai prévu à cette fin un algorithme que j'ai implémenté en Matlab.

#### **Les cycles de corrélation.**

Quant aux cycles de corrélations, l'idée m'a aussi parue intéressante. L'un des aspects sur lesquels BBSP travaille est de proposer à ses clients des éclairages inédits sur les variations de cours d'actifs. Repérer des corrélations cycliques entre ces derniers serait ainsi un excellent moyen de comprendre leur fonctionnement.

Cette idée n'a cependant pas été développée car déjà, je n'ai pas trouvé les outils mathématiques nécessaires pour (en dehors du simple calcul de corrélation). Ensuite chercher les corrélations entre 2 actifs et en dégager un cycle est faisable certes, mais très coûteux en temps de calcul. Cette idée a donc été mise de côté.

Les recherches sur les cycles de corrélations m'ont conduite sur des études sur les cycles économiques qui sont, naturellement, à l'origine d'éventuels liens de corrélations entre actifs. Ces cycles économiques sont eux, suivis au quotidien par les analystes de BBSP qui ont pour le moins qu'on puisse dire, une bonne maîtrise du marché sur lequel ils travaillent. Cela couplé à une alerte qui signalera une corrélation qui s'est faite ou défaite (selon le seuil défini par l'utilisateur) pourrait offrir une alternative à cette idée.

En effet, la corrélation pourra être calculée automatiquement de façon journalière, hebdomadaire, mensuelle etc. selon les paramètres définis. Une alerte sera mise dans l'outil pour signaler, sur les actifs qui sont suivis, une corrélation qui se détache ou qui se crée. L'analyste pourra ainsi, derrière, en chercher la cause.

## **IV.2 Une évolution de l’affichage vers la 3D.**

Une évolution en 3D de l’outil pourra également être implémentée plus tard. Il s’agira par exemple de visualiser les corrélations qui se font ou se défont entre les grands marchés mondiaux et ce, représenté sur la carte mondiale ou même sur la carte d’une région déterminée...

Cet outil offre donc de belles perspectives d’exploitation.

## **PARTIE III**

### **LES APPORTS DU STAGE**

---

## **I. Environnement de travail**

Ce stage de fin d'études a été l'occasion, pour moi, d'intégrer une entreprise au cœur de métier spécialisé, BBSP, au sein de son département informatique CREO. Cela a donné une dimension réelle à mon travail d'ingénieure car il s'agissait de répondre à un besoin, de le penser et de réfléchir à son développement. C'est dans un bureau avec une vue parfaite sur la Tour Eiffel en plein centre de Paris que j'ai vécu cette agréable expérience.

### **I.1 Un travail en équipe**

J'ai rapidement pu constater que ce milieu était constamment en mouvement. BBSP et CREO ont leurs locaux séparés à peine par une porte et un mur vitrés. Il faut dire en effet que ce sont deux services qui travaillent en collaboration quotidienne. Les uns sont au front avec l'analyse technique et les autres s'attellent à gérer leurs outils et, de façon réactive, à corriger un quelconque défaut sur les outils et postes de travail.

C'est une grande équipe interdépendante et indissociable dans son travail quotidien et mon travail n'a pas échappé à cette règle. Il a fallu un apport des deux partis BBSP et CREO pour diriger ce projet dans un sens utile.

A mon arrivée, l'équipe CREO m'a chaleureusement et naturellement accueillie et m'a aidée à connaître les outils informatiques de l'entreprise. J'ai eu également besoin d'eux pour mes extractions de données sans lesquelles je n'aurais pu faire mes tests.

### **I.2 Organisation du travail**

Nous faisons régulièrement, au moins une fois par semaine des points avec notre responsable Bernard Main, Hugues Leroy ainsi que Naoufal Stitou, l'autre stagiaire CREO qui travaillait sur un autre projet mais avec qui on échangeait des idées. C'était un point pour statuer sur l'évolution des travaux. Un autre point également était fait avec toute l'équipe CREO à qui nous présentions nos travaux. Durant tous ces échanges actifs, il y a eu beaucoup d'évolutions en réponse aux commentaires ou critiques formulés et qui ont permis d'avancer de façon constructive.

On apprend ainsi vite qu'en entreprise, l'humilité et le partage constructif sont indispensables à une bonne formation.

En plus, comme nous l'avons dit précédemment, l'outil doit servir aux analystes BBSP. Nous nous concertons donc avec eux également au fur et à mesure que nous avançons dans le projet afin qu'ils puissent nous définir leurs besoins et que nous, nous puissions les appréhender au mieux.

C'est une interaction de tout cela qui a donné le produit avec lequel nous pouvons présenter les résultats présents.

## **II. Bilan final**

Pour reprendre les attentes du stage, il a été demandé avant la fin du stage idéalement, un démonstrateur fonctionnel qui visualise les corrélations statistiques sur un panier d'actifs. Ce démonstrateur doit prendre en entrée un flux de données définies par l'utilisateur, et renvoyer en sortie un graphe lucide représentant les corrélations entre les différents actifs financiers.

Toute la partie mathématique du projet a été achevée et il ne reste à ce jour, qu'à automatiser le démonstrateur, à savoir, réunir l'extraction des données, le calcul des corrélations et l'affichage graphique dans un seul et même programme. On pourra ainsi lancer tout le processus via une interface paramétrable. Ce travail est en train d'être fait par M. Hugues Leroy actuellement qui va compléter son script Ruby en conséquence. Les programmes Matlab que j'ai fait sont ainsi, réimplémentés en Ruby.

Les initiateurs du projet craignaient que le temps de calcul soit important, mais nous pouvons d'ores et déjà affirmer que le programme tourne relativement bien et rapidement même pour un nombre de données important. Pour le moment, nous ne pouvons donner le temps total entre l'extraction des données et l'affichage car l'outil n'est pas encore achevé, mais nous serons certainement sur un temps raisonnable. Sous Matlab, le calcul des matrices de corrélations n'a jamais excédé les 3 minutes (pour les plus gros paniers), et l'affichage des données également est rapide. Nous pensons a priori ne pas dépasser les 4-5 minutes pour les affichages statiques (sur une seule période) avec les plus gros paniers.

Ce projet a été validé comme étant réalisable et sera mis sur pied par l'équipe de R&D de BBSP (l'équipe CREO).

# Conclusion

---

Le but de ce stage était de répondre au besoin d'une entreprise, BBSP, qui dans le contexte d'une interpénétration des marchés financiers, veut visualiser les corrélations liant différents actifs financiers. Partant de l'algorithme interne que BBSP voulait tester, les travaux menés permettent de conclure que le projet est tout à fait réalisable et il en sera ainsi : l'équipe de R&D est chargée par la suite, d'intégrer l'outil de visualisation des corrélations, aux autres outils de travail des analystes techniques de BBSP.

Pour ma première expérience significative en entreprise, j'attendais de travailler sur un projet mathématique qui soit en accord avec ma formation. Dans le département informatique CREO du groupe BBSP, j'ai eu l'occasion de réaliser cela par ce projet. J'en ressors avec une bonne conscience des attentes de mon métier et plus, ce fut une expérience pratique, technique et humaine.

Enfin, arrivée au terme de ce rapport et à la porte de l'obtention de mon diplôme d'ingénieur en mathématiques appliquées, je terminerai par cette phrase d'Albert Einstein qui conclut en beauté, toute cette expérience :

*« La connaissance s'acquiert par l'expérience, tout le reste n'est que de l'information. »*



# BIBLIOGRAPHIE

---

Corps du sujet :

« ***Hierarchical Structure in Financial Markets*** » de Rosario N. Mantegna parue le 24 Février 1998.

« ***Corrélations*** » d'Olaf Torné qui travaillait à BBSP et qui résuma cette étude de Rosario N. Mantegna.

Documentation interne à la boîte :

**CorrelationEmergentV3.xls** : une ancienne étude des corrélations menée dans la boîte.

**How corr is calculated on Bloomberg.xls** : qui décrit la méthode de calcul des corrélations pour le rendement des actifs financiers présentant des sauts.

« Le guide d'Egisthe l'Analyste »

Une liste non exhaustive de sites internet parmi lesquels :

[www.trader-finance.fr](http://www.trader-finance.fr)

[http://eric.univ-lyon2.fr/~ricco/cours/cours/Analyse de Correlation.pdf](http://eric.univ-lyon2.fr/~ricco/cours/cours/Analyse_de_Correlation.pdf)

[http://www.lesechos.fr/finance-marches/vernimmen/definition\\_correlation.html](http://www.lesechos.fr/finance-marches/vernimmen/definition_correlation.html)

<http://www.immobilier-finance-gestion.com/article-correlation-entre-indice-cac-40-et-s-p-500-57904652.html>

<http://www.management-supplychain.fr/wp-content/uploads/2012/02/Performances-financieres-et-Supply-chain-1.pdf>

[http://www.abcbourse.com/apprendre/19\\_correlation.html](http://www.abcbourse.com/apprendre/19_correlation.html)

## FIGURES

---

FIGURE 2 & 3. Arbre couvrant minimal reliant les 30 valeurs utilisées pour calculer l'indice du Dow Jones Industrial Average. Ce sont :

AA – Alcoa ; ALD – Allied Signal ; AXP – American Express Co ; BA – Boeing Co ; BS – Bethlehem Steel ; CAT – Caterpillar Inc., CHV – Chevron Corp., DD – Du Pont, DIS – Walt Disney Co., EK – Eastman Kodak Co., GE – General Electric, GM – General Motors, GT – Goodyear Tire, IBM – IBM Corp., IP – International Paper, JPM – Morgan JP, KO – CocaCola Co., MCD – McDonalds Corp., MMM – Minnesota Mining, MO – Philips Morris, MRK – Merck & Co Inc., PG – Procter & Gamble, S – Sears Roebuck, T – AT & T, TX – Texaco Inc., UK – Union Carbide, UTX – United Tech, WX – Westinghouse, XON – Exxon Corp., et Z – Woolworth.

La matrice de distance ***D*** permet de définir des seuils de corrélation qui seront interprétés par des couleurs, ces dernières allant de vives à foncées. Soient *i* et *j* deux actifs. Nous avons :

Jaune :  $0.65 < d(i, j) \leq 0.70$  ; Vert :  $0.70 < d(i, j) \leq 0.75$  ; Turquoise :  $0.75 < d(i, j) \leq 0.80$  ; Cyan :  $0.80 < d(i, j) \leq 0.85$  ; Bleu :  $0.85 < d(i, j) \leq 0.90$  ; Violet :  $0.90 < d(i, j) \leq 0.95$ .

# ANNEXES

---

## Les actifs du CAC 40

CREDIT AGRICOLE – SAFRAN – AIR LIQUIDE – CARREFOUR – TOTAL – L'OREAL – VALLOUREC – ACCOR – BOUYGUES – LAFARGE – SANOFI – AXA – DANONE – PERNOD RICARD – LVMH – MICHELIN – KERING – ESSILOR INTL. – SCHNEIDER ELECTRIC – VEOLIA ENVIRON. – UNIBAIL RODAMCO - SAINT GOBAIN – CAP GEMINI – VINCI – VIVENDI – PUBLICIS GROUPE SA – SOCIETE GENERALE – BNP PARIBAS ACT.A – TECHNIP – RENAULT – ORANGE – GDF SUEZ – ALSTROM – EDF – LEGRAND – STMICROELECTRONICS – EADS – ARCELORMITTAL – SOLVAY – GEMALTO.

Les Commodities (matières premières) ont également été visualisées. Voici un affichage de corrélation sur une période de 20 jours. Notons ici que les différents secteurs ne sont pas reconnus dans le programme car non rentrés encore dans les références du fichier qu'utilise Graphiz pour les distinguer par couleur. D'où la même couleur pour tous les actifs ici représentés.

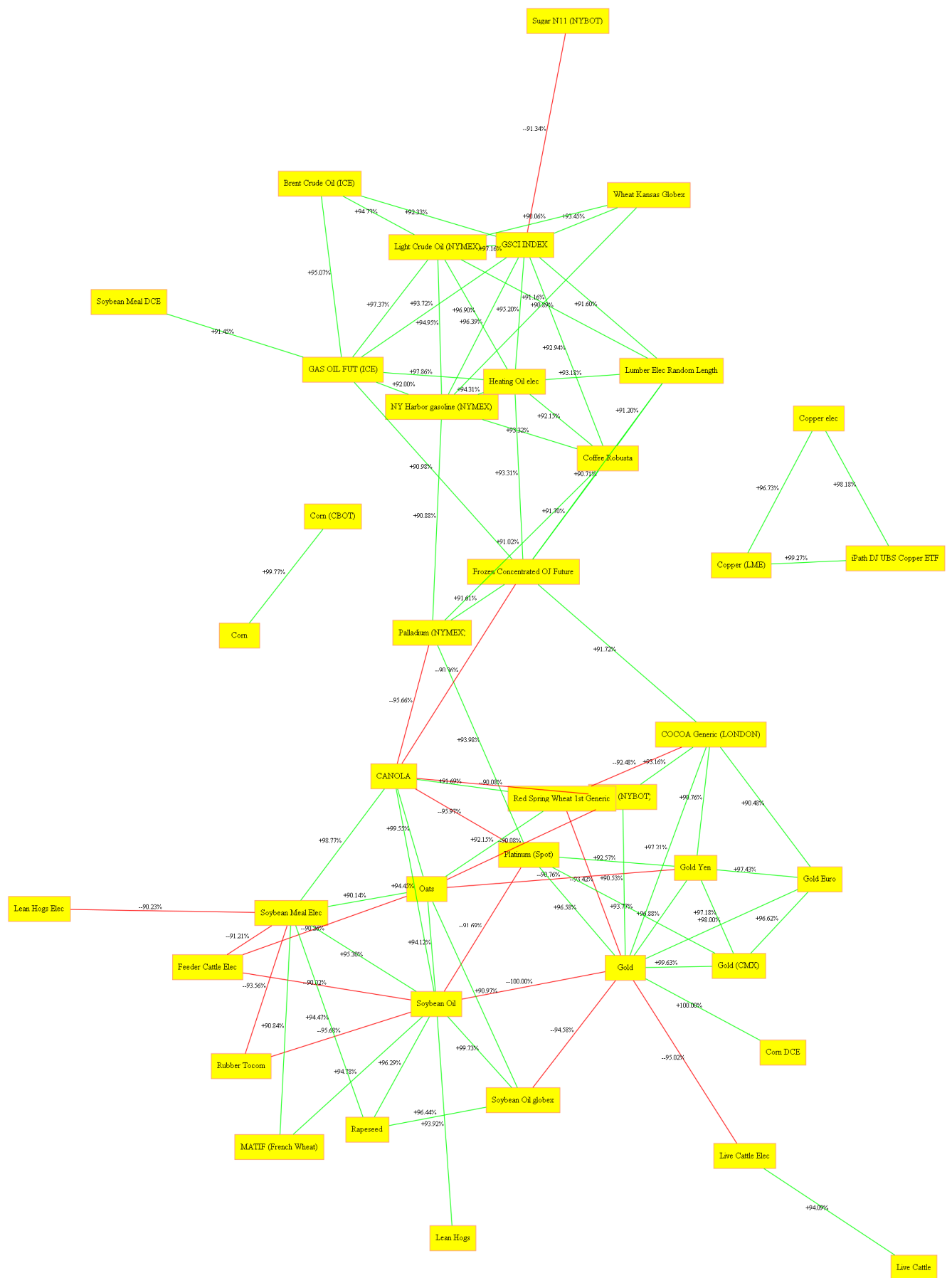


Figure 21. Corrélation de quelques matières premières, période de mi-Juin à mi-Juillet, seuil 0.90