DATA

Square

DONNER DU FUTUR AU TALENT

GRAPH THINKING

QUELLE VALEUR AJOUTÉE EN DATA SCIENCE ?

TEVA ATANI,
EMILIEN JUSSIAUME-MILLET,
JULIETTE PILET.

ÉTUDE I JUIN 2020

RISK & FINANCE

Square

DONNER DU FUTUR AU TALENT

GRAPH THINKING

QUELLE VALEUR AJOUTÉE EN DATA SCIENCE ?

Auteurs : Teva Atani,
Emilien Jussiaume-Millet
et Juliette Pilet


Welcome to Gentallela Alela Admin Theme.
Avarage Time
1.51 Sec
3% From last Week
Total Males
2,500
34% From last Week
Total Females
4,567
12% From last Week
Total Collections
2,315
34% From last Week
Total Connections
7,325
User Signup
Converted Sales
Profit Made
Some shitty stuff 1
Some shitty stuff 2
Some shitty stuff 3
Some shitty stuff 4
Link One
Link Eight
January
February
March
April
May
June
July
Daily active users
Sessions
All users vs users affected crash
App Usage across versions
123k
53k
23k
3k
Top 5
Disbursement
Progress
Marketing 30%
Media 10%
Engagement 20%
Awareness 15%
Settings
Subscription
Auto Renewal
Achievements
Account Balance:
€0.00EUR
€0.00EUR per month
Basic Subscription

ÉDITO

La consécration de la “Donnée”, comme une nouvelle forme de patrimoine au sein de l’entreprise, a bouleversé les organisations, les systèmes d’information et parfois même le rapport au client final.

Son exploitation est désormais incontournable et repose sur de nouvelles compétences, encore rares, regroupées sous le vocable de la science de la donnée, un vaste champ de possibles au sein duquel évolue le Data Scientist, le spécialiste en charge de transformer les données en informations exploitables pour la prise de décision.

Parmi les différentes spécialités, la théorie des graphes propose une approche mathématique différente, combinant une analyse en réseaux avec des visuels plus facilement décryptables pour les non spécialistes.

Cette méthode se veut donc complémentaire aux autres spécialités (Machine Learning, Deep Learning, etc.), pour ainsi apporter un regard différent notamment sur des sujets déjà pourtant bien défrichés : le marketing et la finance. Après une brève revue méthodologique, c’est ce que ce Book Square propose d’analyser en détail.

Au nom de toute l’équipe Data, je vous en souhaite une excellente lecture.

Adrien Aubert,
Associate Partner en charge du domaine d’excellence Data

Square

DONNER DU FUTUR AU TALENT

SOMMAIRE

Square

DONNER DU FUTUR AU TALENT

1. PRÉSENTATION GÉNÉRALE DE LA THÉORIE DES GRAPHES

1.1 DÉFINITION CONTEMPORAINE

L'une des complexités majeures pour les Data Scientist est de réussir à expliquer leur métier et les intérêts concrets de celui-ci. En effet, si l'utilité de l'analyse de données n'est plus à prouver, il est parfois difficile de visualiser ses champs d'application. La théorie est complexe pour les non-experts et la vulgarisation est parfois trop simpliste. Pour pallier cela, il est possible d'expliquer et de comprendre un business en mêlant pédagogie et représentation visuelle grâce à **LA THÉORIE DES GRAPHES (ou Graph Thinking)**[1].

Depuis les premiers travaux, jusqu'à l'essor de son utilisation dans les années 1990, la théorie des graphes n'a cessé de s'enrichir et de s'améliorer. Elle permet à tout un chacun d'appréhender des problématiques toujours plus variées sous une forme plus visuelle et graphique.

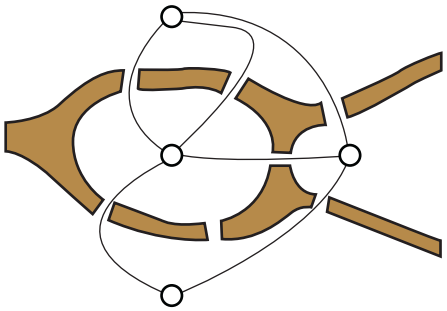
1.2 PREMIÈRES APPARITIONS HISTORIQUES

Il existe un célèbre problème dans la théorie des graphes : celui des **7 ponts de Königsberg.**

La ville de Königsberg est constituée de 2 îles, qui sont reliées entre elles et à la terre par 7 ponts. Le problème consiste à traverser ces 7 ponts une seule et unique fois, et revenir à son point de départ.

Euler a mathématiquement prouvé que c'était impossible, même en ne considérant les ponts que d'une seule de ces îles, on ne peut pas traverser un pont une fois sans devoir le traverser une 2e.

Figure 1. Représentation graphique de la ville de Königsberg et de ses 7 ponts



Cet exemple est considéré comme étant l'un des pionniers concernant la représentation graphique afin de résoudre une problématique. Sans forcément en avoir conscience, les graphes sont présents partout autour de nous.

1. *Par la suite, les deux termes seront utilisés.*

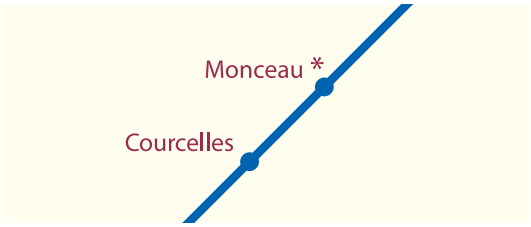
1.3 USAGES (MÉ)CONNUS

Les applications de la théorie des graphes sont en effet nombreuses et le plus souvent méconnues. Préparer un déplacement avec un plan de métro, consulter son arbre généalogique, ou encore, dans un cadre plus professionnel, planifier un projet avec les notions de jalons et de marges (représentations de Gantt par exemple), sont autant d'usages concrets de la théorie des graphes. En matière de gestion de projet, une bonne organisation des tâches optimise le travail collectif et permet à chaque collaborateur de travailler de manière efficace. De plus, la vision globale de cette représentation des activités permet au chef de projet de connaître les attentes en termes de délais des différentes tâches et leurs interdépendances.

1.4 SÉMANTIQUE & TERMINOLOGIES SPÉCIFIQUES

Prenons chacune des stations du métro parisien et définissons-les **en tant que sommets** (ou nœuds). Les différentes stations peuvent être reliées entre elles, ces jonctions sont nommées des **arêtes** (ou arcs).

Figure 2. Extrait du plan de métro parisien



Sa position dans l'espace n'a pas d'importance. En effet, un plan de métro vous donne la même information, peu importe le sens dans lequel vous le tenez.

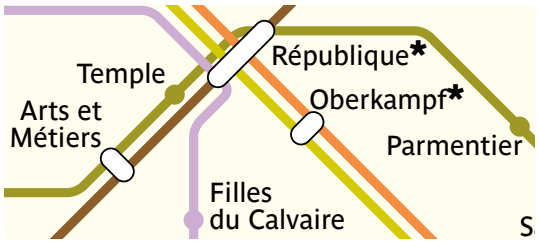
L'ordre d'un graphe est l'un des premiers éléments qui permet de décrire ce dernier. Il s'agit du **nombre de sommets** dont est composé un graphe. Dans notre exemple, nous sommes en présence de 302 nœuds/stations.

Si deux stations sont liées entre elles (par une arête), comme ci-avant les stations Courcelles et Monceau, nous pouvons dire qu'elles sont **adjacentes**.

Si toutes les stations étaient reliées entre elles, nous pourrions parler d'un graphe **complet**.

Mais, les lignes de métro ne sont pas distinctes, il existe des correspondances entre celles-ci. Plus une station donne la possibilité de correspondre avec d'autres lignes, plus son **degré** sera élevé. Par exemple, la station République, avec ses 5 correspondances, est la station ayant le plus haut degré sur le réseau, égal à 5.

Figure 3. Extrait du plan de métro parisien



1.5 CONCEPT DE CENTRALITÉ

L'influence d'un sommet peut se mesurer de diverses façons. L'une des plus utilisées est la **centralité**. Elle permet de situer le sommet dans son environnement (le graphe) et ainsi d'en définir l'importance. Plusieurs types de centralité existent. Selon la problématique, certaines centralités seront plus utiles que d'autres.

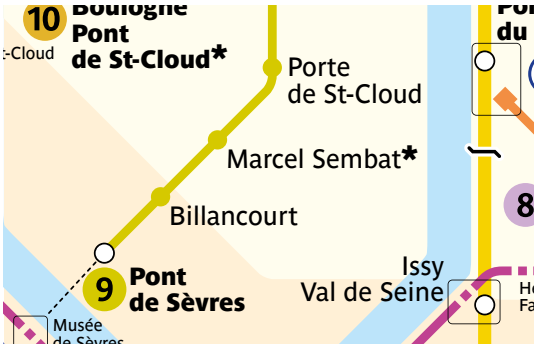
Dans notre cas, seules les centralités de degré et de proximité ont un sens :

- **La centralité de degré** permet d'identifier les stations avec le plus de correspondances ;
- **La centralité de proximité** permet d'identifier la station à partir de laquelle nous sommes le plus proche de chaque station du réseau.

1.5 CONCEPT DE DÉPENDANCE

Lorsqu'il est impossible de rejoindre un nœud sans passer par un autre, on parlera de **dépendance**. En effet, pour rejoindre Porte de Saint-Cloud à partir de Pont de Sèvres, on sera obligé de passer par les stations Billancourt et Marcel Sembat.

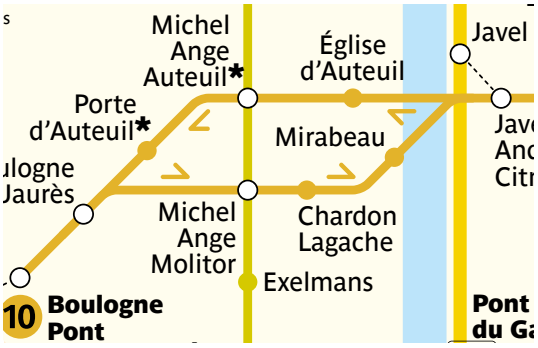
Figure 4. Extrait du plan de métro parisien



1.6 CONCEPT D'ORIENTATION

Un graphe, en plus d'être une succession de nœuds et d'arêtes, peut avoir un sens. On dira de ce graphe qu'il est **orienté**. Toujours sur notre réseau, la ligne 10 du métro comporte des parties orientées.

Figure 5. Extrait du plan de métro parisien



1.7 SYNTHÈSE

La théorie des graphes est une des spécialités de la science de la donnée. Elle repose sur une culture prédominante de la représentation visuelle fondée sur une analyse de la relation entre des observations statistiques. Elle introduit de nouveaux concepts, qui se veulent complémentaires aux analyses traditionnelles de corrélations à plusieurs dimensions. Pour en explorer le potentiel, nous vous proposons, sur cette base, de détailler deux des principaux usages opérationnels, le marketing d'influence et la finance de marché.

2.

RÉSEAUX SOCIAUX ET MARKETING D'INFLUENCE

2.1 QU'EST-CE QUE LE MARKETING D'INFLUENCE ?

Avant toute chose, il faut définir ce qu'est un **influenceur**. Il s'agit d'un individu qui, au travers de divers médias (écrits/photos/vidéos/etc.), a la capacité d'impacter les opinions et les habitudes de consommation des personnes suivant ses posts sur Internet : ces personnes seront par la suite nommés **followers** ou **abonnés**.

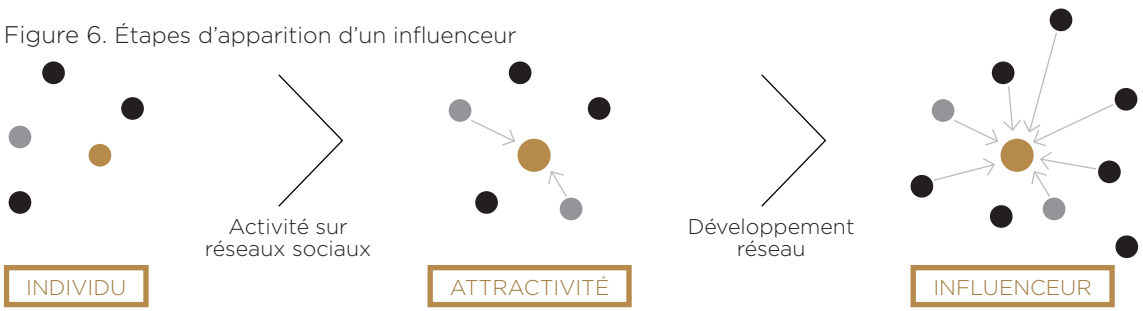
Le panel de sites sur lesquels nous pouvons trouver des influenceurs est immense : parmi les plus connus, nous pouvons citer par exemple Youtube, Twitter ou Snapchat. Ainsi, les influenceurs peuvent poster sur tous types de médias, et sur tous types de sujets : mode, gaming, sport, musique, beauté... Leur champ d'action est illimité.

A la genèse de chaque influenceur, nous pouvons trouver un individu qui a su, pas à pas, attirer l'intérêt d'autres internautes, partageant ses valeurs et ses centres d'intérêts, jusqu'à devenir une figure emblématique dans son domaine. Ainsi, cet individu est devenu un **élément central**. Par le système d'abonnement, de suivi ou "d'acceptation d'amis", l'influenceur tisse une toile d'araignée, un véritable réseau, entre lui et ses followers, représentant eux-mêmes des **nœuds**.

Ainsi, de par leur attractivité et leur popularité, ces influenceurs peuvent être sollicités par des marques dans le but d'en faire la promotion via leurs réseaux sociaux[2]. Ces partenariats peuvent avoir de multiples avantages, à commencer par accroître la notoriété de la marque elle-même, en faisant passer ses messages de manière moins intrusive et en bénéficiant de la viralité de certains contenus. Il existe de nombreuses stratégies de partenariat, parmi lesquelles le buzzkit et le placement de produit.

2. Un réseau social est un site internet mettant en relation des utilisateurs.

Figure 6. Étapes d'apparition d'un influenceur



Le **buzzkit** consiste à envoyer un produit à l'influenceur, qui va ensuite le tester et partager son avis avec sa communauté. Généralement, ce partage se fait en direct, permettant de cerner les différentes émotions de l'influenceur instantanément. Nous pouvons trouver sur de nombreux sites de partage des vidéos d'influenceurs se filmant lors de l'ouverture d'un produit et jugeant de ses caractéristiques.
Les prospects ont alors la possibilité de formuler un avis a priori avant achat, en constatant les caractéristiques du produit et une mise en situation réelle.

Le **placement de produit** est une technique tout aussi connue mais parfois plus discrète. Elle consiste à faire apparaître un produit porté ou utilisé par un influenceur lors de l'une de ses interventions, sans pour autant qu'il en fasse explicitement la promotion. Cette méthode est commune dans les clips vidéo, mais pas uniquement : on peut en trouver dans les films, les séries, ou même dans la vie quotidienne de personnes influentes.

> ***L'usage de cette force de recommandation ou de prescription émanant des influenceurs par les entreprises est ce qui est communément appelé le marketing d'influence. Néanmoins, derrière ce phénomène, désormais entré dans les mœurs des entreprises, s'articulent des réflexions complexes dans le but d'impacter le plus de prospects, tout en minimisant le nombre de partenariats. Ainsi, aux questions habituellement rencontrées en marketing : "quel produit ? par quel canal ? et à qui ?", le marketing d'influence ajoute une nouvelle question : "qui ?".***

2.2 L'INFLUENCEUR AU CENTRE DES LIENS

Les stratégies d'influence et de marketing viral doivent prendre en compte ces nouveaux acteurs, qui apparaissent comme de potentiels amplificateurs de message. Quels sont les facteurs nécessaires pour que ce potentiel soit révélé ? Comment la marque peut identifier l'importance d'un influenceur sur les réseaux ?

La théorie des graphes permet de répondre à ces questions. En effet, chacun de ces individus, y compris l'influenceur, représente un nœud. Toutefois, ce qui fait la différence entre le nœud "internaute lambda" et le nœud "influenceur" est la capacité de ce dernier, grâce à son aura médiatique, à attirer, voire à fidéliser des individus. Autrement dit, un influenceur développe de nombreux liens, lui conférant un certain poids sur les réseaux sociaux jusqu'à devenir un élément **central** de sa communauté. Et c'est grâce à ce concept de **centralité** qu'il est possible d'identifier l'importance relative d'un influenceur par rapport aux autres internautes dans l'univers des réseaux sociaux !

La centralité d'un sommet dans un graphe renvoie à la position de ce sommet par rapport aux autres sommets. De manière intuitive, nous pouvons déjà définir un individu comme central lorsque le nombre de contacts directs qu'il possède est important. Le nombre de ses contacts (ou **liens**) directs correspond au **degré du nœud**. Ces liens peuvent être directionnels (lien entrant/lien sortant) ou pondérés : plus le nœud auquel est rattaché notre sommet est influent, plus le lien avec celui-ci aura un poids fort (par exemple, le poids d'une connexion/abonnement entre deux influenceurs). Cette mesure est appelée "**centralité de vecteur propre**".

Un autre indicateur de la centralité est la proximité (closeness) de l'influenceur par rapport à ses abonnés. Cette proximité peut être mesurée en prenant en compte ses liens directs (lien entre

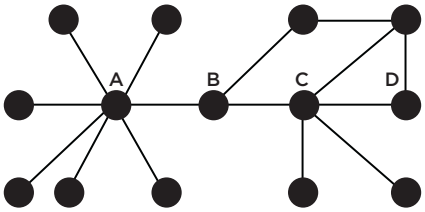
deux sommets) et indirects (deux sommets tous deux en relation avec un troisième).

Enfin, nous appelons **centralité d'intermédiarité** le plus court chemin du graphe passant par chaque sommet.

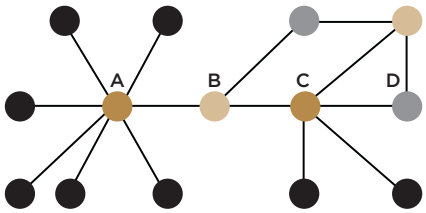
Toutes ces mesures de centralité peuvent conduire à des conclusions différentes, et même contradictoires. Nous allons illustrer toute la difficulté d'identifier un influenceur à travers l'exemple ci-contre.

Dans ce réseau, nous considérons que chaque internaute, représenté par des points, peut potentiellement être un influenceur pour notre marque. Pour simplifier, nous nous intéresserons uniquement aux comportements des individus A, B, C et D. Notons qu'il ne s'agit pas d'un graphe complet, c'est à dire que les individus ne possèdent pas tous des liens entre eux.

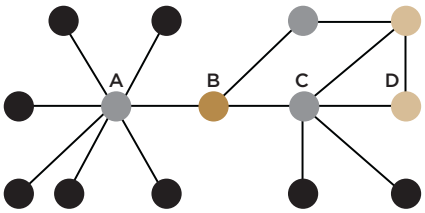
Figure 7. Exemple de réseau



ENCADRÉ 1. LES DIFFÉRENTS TYPES DE CENTRALITÉ

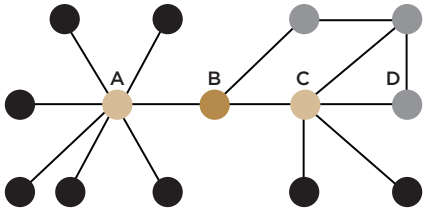


La CENTRALITÉ DE DEGRÉ hiérarchise les individus en fonction de la quantité de leurs liens directs.
Dans notre exemple, les individus A et C possèdent a minima cinq voisins directs et ont par conséquent le degré le plus élevé. Les individus en gris et en beige ayant moins de connexions, leur degré sera moins élevé.

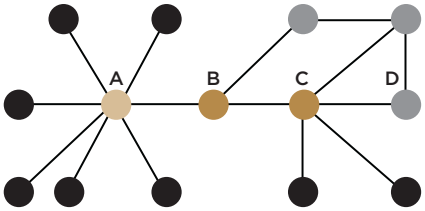


La CENTRALITÉ DE VECTEUR PROPRE hiérarchise les individus en fonction de la quantité de liens considérés comme influents.
Sur le graphique ci-dessus, nous mesurons l'importance d'un individu, non pas en fonction de son nombre de liens directs, mais en fonction de l'importance de ses relations. Ainsi, nous considérons les sommets A et C comme influents. On constate que l'individus B est directement relié à ces influenceurs : il est donc considéré comme plus important au sens du vecteur propre.

ENCADRÉ 1. LES DIFFÉRENTS TYPES DE CENTRALITÉ (SUITE)



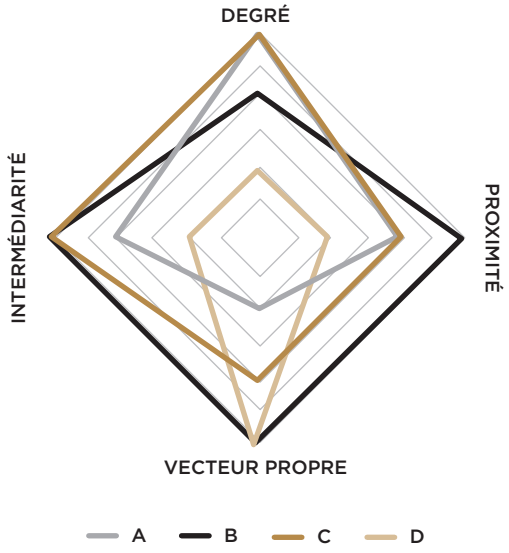
La CENTRALITÉ DE PROXIMITÉ hiérarchise les individus en fonction de leur position relative aux autres individus au sein du réseau.
L'individu B est le plus central car il ne faut parcourir que deux liens pour relier n'importe quel individu du réseau. À partir des individus A et C, le nombre maximal de liens à parcourir est de 3 pour atteindre n'importe quel individu.



La CENTRALITÉ D'INTERMÉDIARITÉ hiérarchise les individus en fonction du nombre de chemins qui passent par eux.
Pour atteindre tous les individus du réseau, il faut a minima passer par les individus A, B et C. Les individus B et C sont centraux au sens du parcours entre la partie gauche et la partie droite du graphe : un nombre important de « plus courts chemins » passe par ces derniers.

Nous constatons que les individus dits centraux (beige et marron) diffèrent selon la mesure considérée. Pour résumer, nous pouvons synthétiser toutes les informations comme suit :

Figure 8. Représentation des différents degrés de centralité de nos 4 individus



Sur ce cadran, il apparaît que les individus A et C se ressemblent mais leurs profils diffèrent en fonction de la nature des connexions : l'individu A dispose d'un grand nombre de connexions, tandis que l'individu C a développé des liens avec des individus plus influents, et apparaît comme un intermédiaire entre les différents individus du graphe. Les individus B et D ont des profils moins comparables : le profil B correspond à un individu par lequel transite l'information d'un groupe à l'autre, alors que l'individu D a, quant à lui, peu de relations mais elles sont toutes influentes.

> ***La qualification d'un influenceur en fonction des indicateurs de centralité permet d'expliciter son impact sur le réseau. Par conséquent, la marque a la possibilité de préférer un influenceur à un autre en fonction de ses stratégies de communication. Après avoir identifié et établi un partenariat avec un influenceur, l'objectif pour l'entreprise est de mesurer l'efficacité de sa campagne.***

2.3 COMMENT MESURER L'EFFICACITÉ DU MARKETING D'INFLUENCE ?

Ces indicateurs peuvent se définir en trois catégories de KPI[3] :

- "**Reach**" ou "**portée**" : Quelle est la popularité de l'influenceur ? Le reach est un élément intervenant à toutes les étapes de la campagne, et correspond à l'importance du réseau d'un influenceur. Il est important de connaître le nombre de followers, le trafic généré par l'influenceur, mais aussi d'analyser le taux d'engagement.
- **Engagement** : Quel est l'impact de l'influenceur sur sa communauté : son audience est-elle active ou s'agit-il d'un faux influenceur ? L'objectif de ces KPIs est d'identifier le nombre de personnes réellement intéressées par la marque et donc susceptibles de devenir des acheteurs. Pour cela, le nombre de clics, de likes, de partages ou de retweets sont des indicateurs quantitatifs de l'attention et de l'intérêt portés au message, ainsi que de sa viralité. Les analyses des commentaires ou des mentions (grâce notamment au Natural Language Processing) sont des indicateurs qualitatifs. A noter que ces indicateurs interviennent également dans le calcul de l'EMV - "**Earned Media Value[4]**".
- **Vente** : L'influenceur a-t-il réussi à convaincre son auditoire ? Il s'agit de mesurer le taux de conversion (ou de transformation) de la campagne marketing. Différentes stratégies peuvent être mises en place en fonction des réseaux. Par exemple, sur Facebook ou Twitter, il est possible de placer des liens trackés dans les posts pour suivre l'origine du trafic. Une autre technique consiste à associer à chacun des influenceurs un code promotionnel différent, permettant d'observer le nombre de leads que ces derniers ont générés.

Ces informations sont récupérables grâce aux APIs proposées par les réseaux sociaux eux-mêmes.

2.4 CONCLUSION

Le marketing d'influence n'est plus une option pour les marques. Nous le constatons chaque jour, cette méthode est présente partout et via de nombreux cas d'application. Une récente étude, réalisée par Tomoson[5] démontre que cette méthode surpasse les autres canaux marketing et qu'il constitue la méthode d'acquisition de clients la plus rapide et la plus rentable. Néanmoins, tout comme les habitudes de consommation des clients, les besoins évoluent, c'est pourquoi il est nécessaire de toujours rester à l'affût des dernières tendances pour ne pas se laisser dépasser. Et ce suivi peut s'avérer complexe dans la majeure partie des cas.

3. Acronyme pour "Key Performance Indicator", traduit par "Indicateur clé de performance" en francais. Il s'agit d'indicateurs mesurables permettant le pilotage et le suivi de l'activité, et de mesurer l'efficacité des actions marketings lors d'une campagne.

4. Earned Media Value : Chaque like, commentaire, mention... a une valeur. Si leur somme est supérieure au coût d'activation de la campagne, alors celle-ci est réussie.

5. https://blog.tomoson.com/influencer-marketing-study/


fn
control
option
command
Q
W
A
S
D
X
C
V

3.

LES APPLICATIONS FINANCE & RÉGLEMENTAIRE

Le marché financier regorge d'une grande diversité de produits qui le composent et le caractérisent : les titres immobiliers, les obligations, les actions... Ces titres voient leur valeur évoluer au cours du temps, influencée par différents facteurs pouvant être regroupés dans deux grandes catégories :

> Les **événements exogènes**, tels que la situation économique, politique ou sociétale d'un pays, l'état de la concurrence.

> Les **événements endogènes**, tels que les facteurs micro-économiques comme la condition financière d'une société ou sa stratégie à moyen ou long terme, peuvent faire évoluer à la hausse ou à la baisse le cours du produit financier, un gain ou une perte pour un investisseur. Ainsi, des stratégies de management du risque visent à optimiser le gain ou maîtriser la perte, notamment grâce à la diversification du portefeuille.

> ***Comment choisir un portefeuille d'actions, simplement en minimisant les risques de perte ?***
> ***La théorie des graphes offre la possibilité de schématiser les interactions entre les différentes actions, afin d'optimiser l'arbitrage par rapport à des matrices de chiffres plus complexes.***

3.1 LA DIVERSIFICATION DE PORTEFEUILLE AVEC LA THÉORIE DES GRAPHES

3.1.1 La théorie moderne du portefeuille

La théorie moderne du portefeuille expose la manière dont les investisseurs, supposés rationnels, utilisent la diversification afin d'optimiser la rentabilité de leurs actifs financiers. Tout l'enjeu de cet exercice consiste à maximiser le rendement du portefeuille d'actions pour un niveau de risque donné, ou a contrario, à minimiser le risque pour un niveau de rendement donné. Cette maîtrise du risque passe notamment par la constitution d'un portefeuille composé de titres faiblement ou négativement liés entre eux. Plus communément, l'investisseur a pour objectif de "ne pas mettre tous ses œufs dans le même panier".

Ainsi, l'enjeu va être d'identifier l'existence de lien ou de dépendance entre les cours des actions, et la manière la plus simple pour mesurer ces liens est le calcul d'un indicateur appelé **corrélation**.

L'illustration de cette méthodologie peut difficilement s'appuyer sur des exemples concrets, de par la confidentialité des compositions de portefeuilles d'actifs au sein

des établissements financiers. Par conséquent, il s'agira dans la suite de cette application de s'appuyer sur une théorie illustrative.

3.1.2 De la corrélation dans les graphes !

Le coefficient de corrélation se calcule en considérant les cours historiques d'une paire d'actifs. Lorsque plus de deux actifs sont considérés, afin d'exploiter le résultat global, le coefficient de corrélation doit être calculé pour chaque paire d'actifs.

Les résultats de ces calculs sont souvent synthétisés dans une matrice, représentée ci-dessous, appelée matrice de corrélation.

Figure 9. Matrice de corrélation

	A	B	C	D	E	F	G	H	I
A	1.0000000	0.8724332	0.92730670	-0.419750528	-0.4346084	0.143437958	0.46766268	0.3175530	0.8254469
B	0.8724332	1.0000000	0.74783386	-0.497227095	-0.2992901	0.173186053	0.42716227	0.3089994	0.9836630
C	0.9273067	0.7478339	1.00000000	-0.058566349	-0.4086673	0.162486359	0.45027991	0.3174286	0.7007417
D	-0.4197505	-0.4972271	-0.05856635	1.000000000	0.2060044	-0.000543089	-0.15738819	-0.0660683	-0.4783142
E	-0.4346084	-0.2992901	-0.40866727	0.206004429	1.0000000	-0.441041897	0.11108096	0.2853231	-0.1226329
F	0.1434380	0.1731861	0.16248636	-0.000543089	-0.4410419	1.000000000	-0.01512747	-0.1356710	0.0969256
G	0.4676627	0.4271623	0.45027991	-0.157388185	0.1110810	-0.015127472	1.00000000	0.9482029	0.4652614
H	0.3175530	0.3089994	0.31742856	-0.066068298	0.2853231	-0.135670987	0.94820287	1.0000000	0.3752306
I	0.8254469	0.9836630	0.70074171	-0.478314158	-0.1226329	0.096925595	0.46526140	0.3752306	1.0000000

Cette matrice met en évidence le lien entre chaque paire des neufs actifs considérés. Pour mieux comprendre, quelques règles sont à prendre en compte. Ces valeurs sont comprises entre -1 et 1 avec l'interprétation suivante :

> Si le coefficient de corrélation est proche de 1, alors les deux actions évoluent dans le même sens ;
> Si le coefficient de corrélation est proche de -1, alors les deux actions évoluent dans le sens contraire ;
> Enfin, si le coefficient de corrélation est proche de 0, alors les variables ne sont pas corrélées linéairement.

D'autres remarques sont également importantes pour comprendre l'indicateur dans sa globalité :

> La corrélation entre A et lui-même vaut 1 ;
> La corrélation entre AB et BA est identique (il s'agit d'une mesure symétrique).

Plus le nombre d'actifs considérés devient important, plus la lecture de la matrice de corrélation devient complexe, c'est pourquoi la représentation graphique de ces informations est une alternative intéressante et appropriée à ce genre de problématique. En effet, elle permet de visualiser plus rapidement les liens les plus importants entre les actifs.

Afin de rendre l'information la plus allégée possible, il est intéressant de ne conserver que les corrélations qui sont supérieures à un certain seuil, déterminé par l'intuition de l'expert après son analyse de la problématique (par exemple ici 0,7) en les annotant 1 et 0 dans le cas contraire. La matrice devient donc :

Figure 10. Matrice d'adjacence

	A	B	C	D	E	F	G	H	I
A	0	1	1	0	0	0	0	0	1
B	1	0	1	0	0	0	0	0	1
C	1	1	0	0	0	0	0	0	1
D	0	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0	1	0
H	0	0	0	0	0	0	1	0	0
I	1	1	1	0	0	0	0	0	0

En considérant que chaque actif représente un sommet, et que chaque arête entre paires de sommets est représentée par un "1" depuis la matrice. Ce premier exemple représente les

interactions considérées comme fortes entre l'ensemble des actifs, autrement dit les plus corrélés. A noter que ce graphique est spécifique au cas où le seuil est de 0,7. En effet, en imaginant d'autres seuils, d'autres représentations auraient émergé :

Figure 11. Représentation graphique de matrice de corrélation (seuil > 0.7)

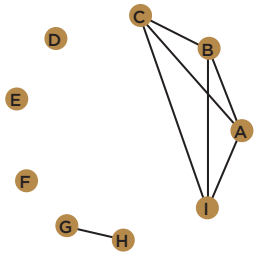


Figure 12. Représentation graphique de matrice de corrélation (seuil > 0.4)

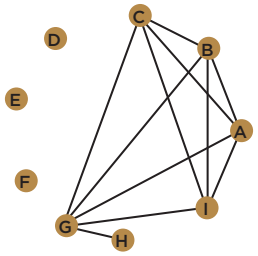
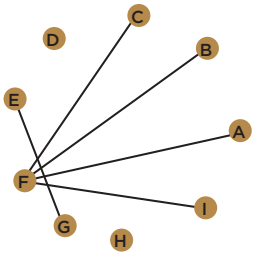


Figure 13. Représentation graphique de matrice de corrélation (seuil entre 0 et 0.2)



Ainsi, le choix du seuil permet de retranscrire des informations différentes : la première illustration met en évidence les liens de « forte corrélation » entre les actifs, tandis que la seconde met en évidence les actifs faiblement ou non corrélés. Bien que cette dernière représentation ne soit pas courante, elle est un moyen d'illustrer les actifs potentiellement intéressants dans la constitution d'un portefeuille diversifié.

Cette illustration expose la possibilité de retranscrire des informations matricielles via un outil appelé graphe. Dans la théorie, ces dernières sont nommées **matrices d'adjacence** (cf. glossaire).

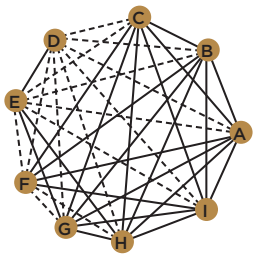
> ***La transposition des matrices de corrélation en graphe permet d'affiner et de visualiser rapidement les liens ainsi mis en évidence. Toutefois, l'exemple ci-dessus ne révèle pas l'intensité de la dépendance entre les deux actifs : le graphe ne permet pas d'identifier si la corrélation est positive ou négative, information importante lorsque l'on souhaite diversifier notre portefeuille. Comment représenter explicitement ces différentes dépendances ?***

3.1.3 Complexification des représentations

La représentation matricielle précédente identifie les arêtes entre deux sommets en attribuant la valeur 1 lorsque ceux-ci sont reliés, 0 dans le cas inverse. Or, cette représentation ne permet pas de prendre en compte l'intensité de la relation.

L'iconographie des corrélations est une méthode qui consiste à remplacer une matrice de corrélation par un graphe. Dans ce type de représentation, les corrélations positives sont généralement représentées avec un trait plein et les corrélations négatives avec un trait en pointillé. Les pondérations attribuées aux arêtes peuvent également figurer sur le graphe.

Figure 14. Exemple d'iconographie des corrélations

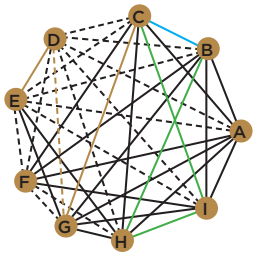


Grâce à cette représentation, nous constatons rapidement que l'action D est corrélée négativement avec toutes les autres actions, sauf l'action E.

A noter : Les représentations graphiques peuvent se trouver bien plus complexes, en faisant intervenir davantage de paramètres.

La matrice d'adjacence est un outil essentiel en théorie des graphes afin de synthétiser l'ensemble des liens entre les actifs, avec une structure ordonnée de l'information. Dans les exemples précédents, celle-ci découle du calcul du coefficient de corrélation entre chaque paire d'actifs. D'autres méthodes plus complexes, associées au Deep Learning, existent et aident à nous projeter sur le problème de manière différente. Malgré tout, les représentations évoquées jusqu'alors ne permettent pas de déterminer le sens des liens existants entre les différents actifs, en raison de la nature symétrique du coefficient de corrélation. Quels sont les outils qui permettent d'identifier le sens des liens ? Comment les représenter ?

Figure 15. Iconographie des corrélations avec paramètres supplémentaires



3.2 CAUSALITÉ ET GRAPHE ORIENTÉ

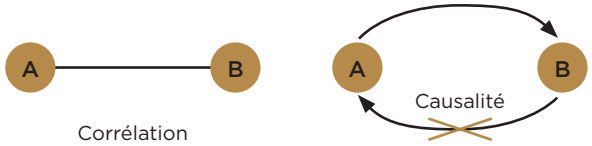
Comme vu précédemment, la corrélation entre les titres A/B est la même qu'entre les titres B/A. Pour autant, le comportement sur les marchés financiers est plus complexe : l'évolution du cours de l'action A peut influencer le cours du titre B sans que la réciproque ne soit vérifiée. Il s'agit du phénomène **d'interdépendance des marchés**, qui n'est pas identifiable ou mesurable avec le coefficient de corrélation.

Une illustration de la causalité peut être considérée à travers l'évolution du prix du pétrole et le cours du dollar. De nombreuses études[6] mettent en évidence une relation de dépendance positive entre les deux actifs financiers, de sorte que lorsque le prix du pétrole augmente, le cours du dollar augmente également. Afin de compléter cette analyse, Throop (1993) étudie le lien de causalité qui pourrait exister et identifie que le sens de la relation va du prix du pétrole vers le dollar. Amano et Van Norden (1993, 1995) justifient ce lien de causalité comme la conséquence d'une préférence pour les pays exportateurs de pétrole à réaliser des investissements en dollar. Autrement dit, la flambée des prix du pétrole stimule la richesse des pays producteurs et la demande d'actifs en dollar[7].

6. *Throop, 1993 ; Zhou, 1995 ; Dibooglu, 1996 ; Amano et Van Norden, 1998 ; Bénassy-Quéré, Mignon et Penot, 2007 ; Coudert, Mignon et Penot, 2007*
7. *A noter toutefois que les conclusions sur cette question ne sont pas tranchées dans la littérature scientifique. Les résultats ne sont pas robustes et dépendent de la période d'observation considérée, du modèle choisi, et le recours à d'autres variables explicatives dans le modèle.*

ENCADRÉ 2 : DIFFÉRENCE ENTRE CORRÉLATION ET CAUSALITÉ

Une **corrélation** est un lien statistique, sans pouvoir identifier quelle variable agit sur l'autre. Une **causalité** est un lien qui affirme qu'une variable agit sur une autre.



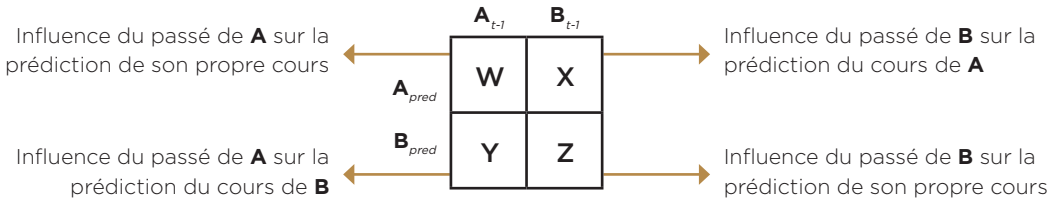
A noter que la version la plus standard de la causalité est dite linéaire (A, donc B, avec éventuellement des étapes causales intermédiaires) ; elle suppose une antériorité, même infime, de la cause sur l'effet.

L'étude de ces co-mouvements sur les marchés financiers a fait couler beaucoup d'encre dans le monde de la recherche et a conduit aux développements de modèles statistiques permettant de prendre en compte ces phénomènes : c'est notamment le cas des **modèles VAR** (modèles vectoriels auto-régressifs). Ce modèle de séries temporelles a l'avantage de considérer les relations dynamiques entres les variables et de permettre l'étude de la propagation des chocs au sein d'un système économique.

Les historiques de prix de deux actions (A et B) sont considérés pour illustrer un système VAR. La modélisation/prédiction du cours de l'action A via un modèle VAR repose sur une régression linéaire prenant en compte les valeurs passées de l'action A et les valeurs passées de l'action B. De cette modélisation, il en résulte des estimations de coefficients liant les valeurs passées des cours de A et B à la valeur présente de A. Autrement dit, les estimations des valeurs de A sont issues d'une combinaison d'effets des valeurs passées de A et de B simultanément.

Parallèlement, l'estimation de B repose également sur une combinaison de ses propres valeurs passées et de celles de l'action A.

Selon le modèle utilisé, l'historique pris en compte dans le calcul des valeurs actuelles remonte plus ou moins loin dans le temps. Ces coefficients sont présentés sous format matriciel. Avec cet exemple, la matrice finale obtenue est de la forme suivante :



Ainsi, l'estimation d'un tel modèle permet de repérer les interactions existantes entre les différentes composantes de ce portefeuille. L'analyse des coefficients de régression indique **le sens de la causalité** (au sens de Granger[8]) entre deux variables lorsqu'elle existe et **l'ampleur de la dynamique temporelle**.

8. *Causalité au sens de Granger : Granger a développé un test de causalité simple pour tenter d'identifier l'origine de la causalité entre deux séries temporelles. Le principe est le suivant : une série en cause une autre si elle permet d'en expliquer les variations futures.*

ENCADRÉ 3 : LES MODÈLES VAR

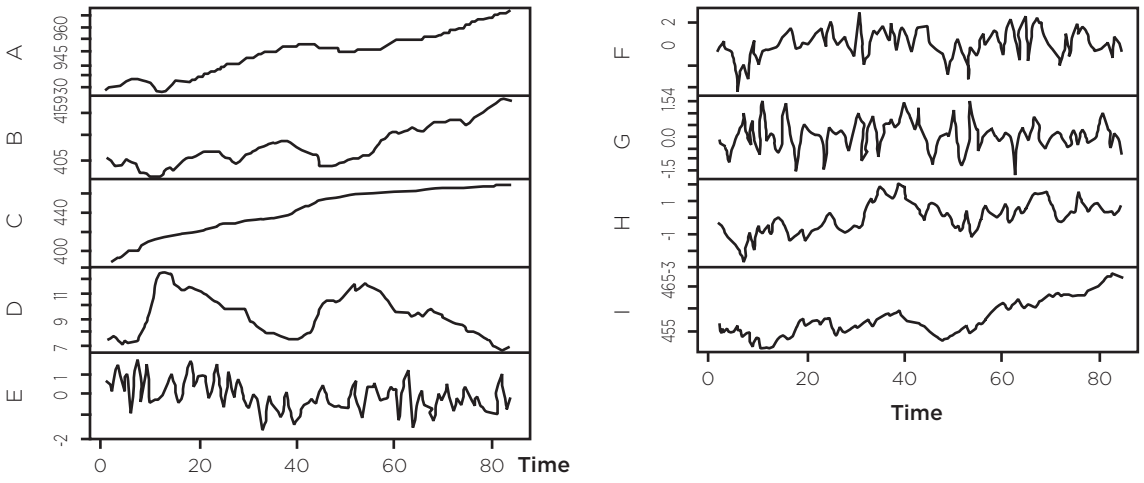
Ils sont utilisés lorsqu'une causalité est constatée. Un modèle VAR(1) s'écrit matriciellement comme suit :

$$r_t = a + B \times r_{t-1} + u_t$$

Avec :

- a la constante du modèle ;
- u_t les erreurs d'estimation ;
- r_t un vecteur de dimension n contenant les rendements des actifs considérés dans le modèle à la date t.
- La matrice des coefficients estimés, communément notée B, est une matrice de dimension n x n, tel que n soit le nombre d'actifs considéré dans le modèle. L'élément b_{ij} représente l'impact de l'actif j en période t-1 sur l'actif i en période t.

Figure 16. Évolution des cours des actions



En estimant un modèle VAR(1) sur ces données, l'ensemble des coefficients obtenus via l'algorithme est restitué dans cette matrice, représentant les interactions des cours passés des actions du portefeuille sur le cours présent pour chacune de ces actions :

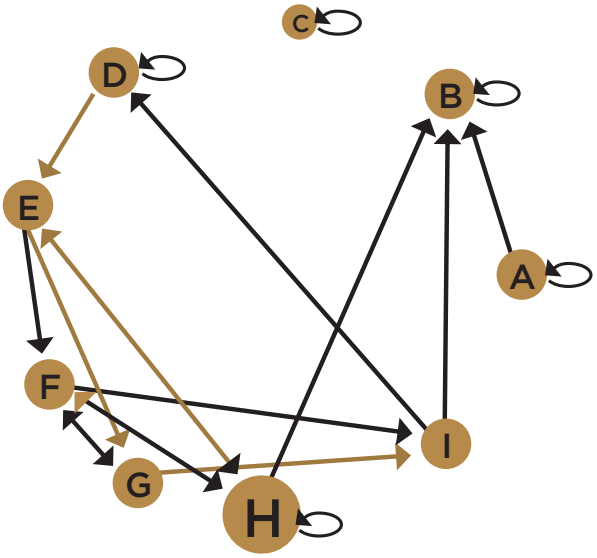
Figure 17. Matrice récapitulative des coefficients du modèle

	A	B	C	D	E	F	G	H	I
A	0.9410216	0.1375167	0.0000000	0.0000000	0.00000000	0.0000000	0.000000	0.0000000	0.0000000
B	0.0000000	1.0003531	0.0000000	0.0000000	0.00000000	0.0000000	0.000000	0.0000000	0.0000000
C	0.0000000	0.0000000	0.9743427	0.0000000	0.00000000	0.0000000	0.000000	0.0000000	0.0000000
D	0.0000000	0.0000000	0.0000000	0.9091642	-0.05353131	0.0000000	0.000000	0.0000000	0.0000000
E	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.2448813	-3.037608	1.9101045	0.0000000
F	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	5.119234	-3.2741151	1.0829799
G	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.2049523	0.000000	0.2466938	-0.4341991
H	0.0000000	0.5751568	0.0000000	0.0000000	-0.51232811	0.3642815	-1.828850	1.3058895	0.0000000
I	0.0000000	1.1142389	0.0000000	0.3719189	0.00000000	0.0000000	0.000000	0.0000000	0.0000000

Interprétation : le cours de l'action B dépend de sa valeur à la période précédente (notée généralement t-1) et des valeurs t-1 de l'action A, H et I, les coefficients affectés étant positifs.

De la même façon que pour les corrélations, les interactions entre les différentes actions ne sont pas facilement visibles à travers cette représentation matricielle. L'avantage de la théorie des graphes est de fournir une représentation graphique permettant de repérer visuellement et plus rapidement des interactions entre les différents actifs financiers.

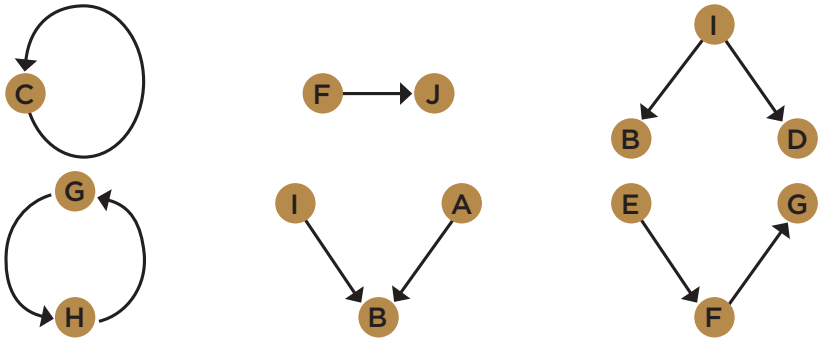
Figure 18. Représentation graphique des interactions



Les sommets correspondent aux actions sélectionnées pour l'analyse et la modélisation. Un arc noir (resp. marron) représente un lien de causalité positif (resp. négatif).

A partir de ce graphe, différentes structures relationnelles particulières entre les actions peuvent être identifiées, et notamment six motifs d'interactions distincts :

Figure 19. Différents types d'interactions



Chacun de ces motifs influence la variance du portefeuille de manière différente, permettant de mettre en pratique des stratégies de portefeuille différentes.

La combinaison de l'ensemble des méthodes décrites permet d'affiner l'identification des liens et de leur sens d'existence entre les actifs, afin d'aider à la prise de décision dans l'élaboration, ou dans la gestion du portefeuille au cours du temps.

3.3 AUTRES CHAMPS APPLICATIFS

3.3.1 Identifier les opportunités d'arbitrages grâce aux graphes

La théorie des graphes est un outil mathématique intéressant dans le cadre d'une analyse des dépendances entre différentes actions, en vue de la constitution d'un portefeuille diversifié. D'autres applications sont également envisageables, telle que **l'identification des opportunités d'arbitrage**. Il s'agit d'opérations financières ayant pour objectif de réaliser un gain en profitant d'écarts temporaires de prix constatés entre différents titres. Ces écarts sont le reflet d'incohérences résultant d'un marché imparfait, comme par exemple lorsque deux actifs sont cotés sur différents marchés ou encore lorsque deux actifs financiers rapportent le même montant de flux de trésorerie dans le futur, mais dont les prix d'achat sont différents aujourd'hui.

Afin de déterminer s'il existe une opportunité d'arbitrage, voici un exemple qui s'inspire des taux de change réels entre quatre devises principales[9] : USD, EUR, CHF et GBP. Les cours en temps réels relevés sont récapitulés dans le tableau ci-dessous :

Tableau 1. Taux de change

	EUR	USD	GBP	CHF
EUR		1,1442	0,895	1,1363
USD	0,8734		0,7836	0,9949
GBP	1,117	1,2758		1,2682
CHF	0,8798	1,0049	0,7885	

$R_{devise\,i devise\,j}$ correspond au nombre d'unités de la devise j qu'il est possible d'acheter avec une unité de la devise i. Lorsque plusieurs devises sont disponibles sur le marché, il est possible d'enchaîner les échanges. Par exemple, en échangeant une unité de la devise EUR en devise USD, il est possible ensuite d'échanger une unité de devise USD avec une unité de devise CHF. Le résultat est alors R_{EURUSD} R_{USDCHF} unités de la devise CHF.

Une opportunité d'arbitrage apparaîtra lorsque le nombre d'unités de la devise finale obtenu est supérieur au nombre d'unités de la devise initiale.

Cela se traduit, formellement, par une suite d'échanges dont le produit est supérieur à 1 :

$$R_{EURUSD}\ R_{USDGBP}.....R_{devise\,n\text{-}1\ EUR}>1$$

Avec $R_{devise\,n\text{-}1\ EUR}$ le dernier arbitrage réalisé sur l'ensemble de l'opération.

Via certaines propriétés liées à des fonctions mathématiques (logarithme népérien), il est possible de transformer cette opération multiplicative en une opération d'addition.

Par la suite, résoudre ce problème grâce à la théorie des graphes revient à identifier un **circuit absorbant** (s'il existe).

9. Les valeurs sont inspirées du site : https://www.oanda.com/lang/fr/currency/live-exchange-rates/ en date du 20 Aout 2018 à 14h45

ENCADRÉ 4 : CYCLE

Dans un graphe non orienté, un cycle est une suite d'arêtes consécutives dont les deux sommets aux extrémités sont identiques. Autrement dit, le sommet de départ correspond au sommet d'arrivée. Dans un graphe orienté, la notion équivalente est celle de circuit, même si on parle parfois également de cycle.

Cycle absorbant

Dans un graphe orienté, on appelle circuit ou cycle un chemin fermé tel que le sommet de départ et le sommet terminal coïncident et que tous les arcs parcourus par le chemin sont différents. Un circuit absorbant est un circuit ayant un coût négatif, c'est-à-dire que la somme des poids des arcs est négative.

L'algorithme de Floyd et Warshall permet d'identifier un tel circuit s'il existe. Il permet de parcourir l'ensemble des cycles dans un graphe orienté pondéré, et plus particulièrement de détecter par la suite le plus court chemin entre 2 sommets. Le calcul du coût du circuit est opéré en réalisant la somme des poids affectés à chacune des arêtes traversées par le circuit.

Figure 20. Cycle d'échange de devises

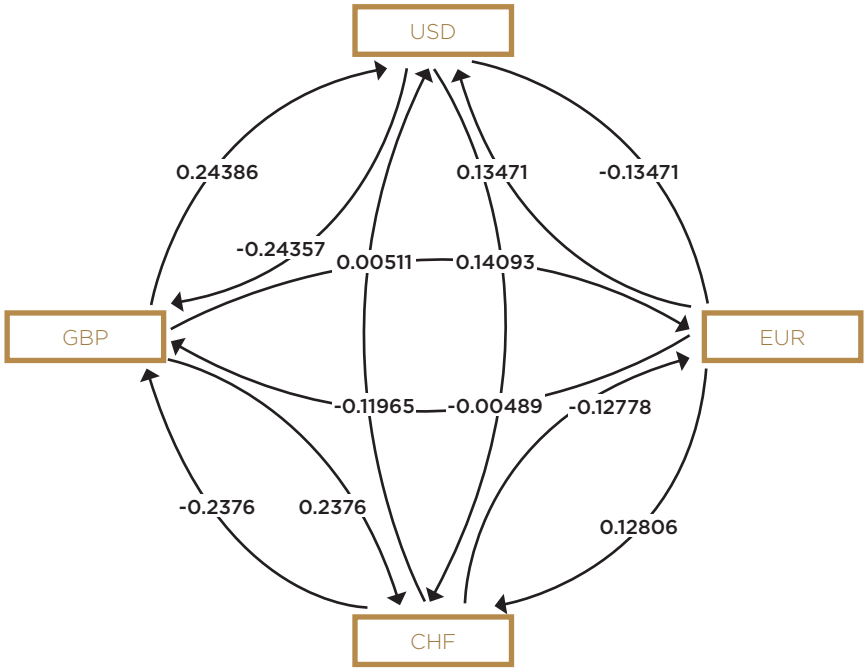


Tableau 2. Coûts des différents cycles d'échange de devises

Type de Circuit	Coût du circuit
[EUR, USD, CHF, GBP, EUR]	-0,002640601
[EUR, USD, CHF, EUR]	-0,001531983
[EUR, USD, GBP, EUR]	-0,00149563
[USD, CHF, GBP, USD]	-0,000861811
[EUR, CHF, GBP, EUR]	-0,000825318
[EUR, USD, GBP, CHF, EUR]	-0,000387012
[EUR, USD, EUR]	0
[GBP, CHF, GBP]	0
[USD, CHF, USD]	0,000225015
[USD, GBP, USD]	0,00028316
[EUR, CHF, EUR]	0,0002833
[EUR, GBP, EUR]	0,000285041
[EUR, GBP, USD, CHF, EUR]	0,000531847
[EUR, CHF, USD, GBP, EUR]	0,000544669
[EUR, CHF, GBP, USD, EUR]	0,000953472
[USD, GBP, CHF, USD]	0,001369987
[EUR, GBP, CHF, EUR]	0,001393658
[EUR, CHF, USD, EUR]	0,002040299
[EUR, GBP, USD, EUR]	0,002063831
[EUR, GBP, CHF, USD, EUR]	0,003150657

Sur notre exemple, seuls les six premiers correspondent à un circuit absorbant. Ainsi, si on réalise l'opération EUR - USD - CHF - GBP - EUR, le gain de l'opération est de 0.0026€ par euro investi.

Cette opération n'est pas triviale surtout si le volume des données disponibles est important. De plus, et compte tenu du fait que l'information circule plus facilement aujourd'hui, les opportunités d'arbitrages se raréfient et/ou la fenêtre temporelle de cette opportunité est faible.

Ce type de problématique s'apparente à celle du contexte plus connu de High Frequency Trading dans le milieu de la finance, où les opérations s'exécutent en quelques fractions de secondes par des algorithmes gérant un volume important de données.

Dans ce cadre, la théorie des graphes permet de simplifier la représentation et l'actualisation des informations matricielles, **les calculs algorithmiques tels que celui-ci sont facilement implémentables grâce à des packages sur R et Python**, notamment **IGraph** ou encore **NetworkX**, et conservent une part interprétable pour les analystes via des représentations intuitives.

3.3.2 Risques systémiques et théorie des graphes

Les deux opérations évoquées jusqu'à présent (la spéculation et l'arbitrage), sont deux manières de gérer le profit. Ces deux fonctions s'appuient sur l'interdépendance des actifs et l'interconnexion des marchés pour la mise en place d'une stratégie. Bien que ces inter-relations puissent être profitables dans certains contextes, elles sont également vecteurs de risque via la contagion des différentes parties du système. Typiquement, la crise des Subprimes en 2008 est une crise caractérisée par la propagation d'événements négatifs aux différentes institutions et places financières mondiales. C'est cette réaction en chaîne qui est appelée **risque systémique**.

La mesure du risque systémique est fondée soit sur la mesure du degré de contagion, soit sur des indicateurs descriptifs de l'architecture du réseau bancaire, qualifiée de « topologie » en mathématiques. De nombreux travaux utilisent la théorie des graphes pour en déduire ces mesures.

Toutefois, et avant de pouvoir calculer ces indicateurs, la construction du graphe associée aux relations inter-institutionnelles est une difficulté à laquelle il faut faire face. En effet, pour pouvoir dessiner le système financier, les informations les plus pertinentes seraient les obligations contractuelles entre établissements (comme par exemple les prêts interbancaires et les échanges de crédits). Malheureusement, ces données sont généralement peu disponibles et/ou opaques et le recours à des mesures de dépendances peut procurer des résultats instables et non robustes en fonction des données utilisées.

La représentation du système financier peut donc s'avérer complexe. Toutefois, le lien entre banques reste identifiable à travers des informations plus indirectes, tels que les rendements des actions (equity return) et les *portfolio holding*. A partir de ces historiques de prix, les méthodes statistiques permettent de retrouver les liens. Typiquement, la causalité au sens de Granger, présentée dans les parties précédentes, permet d'établir des liens à partir de relations linéaires sur la base des données historiques des *equity return*.

A partir des graphes, qu'ils aient été construits directement ou indirectement, les mesures de centralités (dont les notions ont été introduites et définies précédemment) permettent d'appréhender les acteurs clés dans le risque systémique. Par exemple, la **centralité d'intermédiarité**, qui fournit une indication de l'exclusivité du nœud dans le réseau global, est importante pour identifier les nœuds dont la suppression exercerait la plus forte incidence sur la résilience du réseau. Autre mesure intéressante, la **centralité de vecteurs propres**, qui donne une indication sur l'importance des nœuds dans la propagation d'un choc. Autrement dit, cette mesure révèle la capacité d'une institution financière à engendrer de lourdes pertes pour d'autres institutions par contagion. Les centralités d'intermédiarité et de vecteurs propres permettent donc d'identifier les institutions dites *"too big to fail"*, c'est-à-dire celles dont la faillite pourrait déclencher un effondrement systémique.

Outre l'identification des *"too big to fail"*, un autre aspect important à mettre en évidence est le chemin de cette contagion. La recherche d'un **arbre couvrant de poids minimal** (ACM, ou Minimum Spanning Tree en anglais) est une solution à ce problème.

Cet algorithme dessine un chemin, en parcourant l'ensemble des sommets du graphe sans boucle, assimilable au chemin de propagation de choc le plus court qui soit, et donc le plus probable, au sein du système de prix.

La représentation des ACM permet d'extraire la topologie du système, et d'identifier, selon le schéma, les institutions jouant un rôle capital dans le risque systémique.

En effet, un tel arbre parcourt l'ensemble des sommets du graphe, sans boucle, et en minimisant la somme des poids des arêtes. L'arbre dessine un chemin, assimilable au chemin de propagation de choc le plus court qui soit, et donc le plus probable, au sein du système de prix. La représentation des ACM permet d'en extraire sa topologie. Par exemple, dans le cas où l'arbre aurait une forme linéaire, autrement dit formerait une "ligne", un choc provenant à une extrémité n'aura qu'un chemin possible et devra passer successivement par tous les nœuds avant d'atteindre l'autre extrémité.

Dans un arbre structuré en étoile, la propagation des chocs est plus complexe à prévoir : en provenance du centre de l'étoile, le choc peut se propager sur un ou plusieurs autres nœuds, rendant complexe la prédiction de cette direction. De plus, dans cette configuration, ce nœud central revêt un caractère capital dans le risque systémique, confirmant le caractère *"too big to fail"* des institutions financières concernées.

L'appréhension des risques systémiques par la théorie des graphes permet de passer d'une surveillance micro-prudentielle à une surveillance macro-prudentielle, où l'intérêt est porté sur la stabilité d'un système dans son ensemble et non pas sur la stabilité individuelle de chaque institution financière.

3.4 CONCLUSION

Les actifs, les marchés, les facteurs économiques et les sociétés regorgent de relations, d'influences, de corrélations et de causes, qui peuvent évoluer rapidement avec le temps et difficilement identifiables. Par ailleurs, avec la prolifération des données, l'analyse du marché s'avère plus complexe pour un opérateur humain : difficile en effet de dégager une stratégie opportuniste ou de gestion de risque dans un contexte "Big Data".

La théorie des graphes apparaît alors comme un moyen de renouveler les méthodes quantitatives "traditionnelles", en offrant la possibilité d'intégrer et d'analyser ces dépendances et de les appréhender en les représentant.

La théorie des graphes peut par ailleurs être vue comme une méthode de « vulgarisation » des concepts théoriques, avec son aspect visuel beaucoup plus intuitif et accessible à des profils de compétences diversifiés au sein de l'entreprise.


fn
control
option
command
Q
W
E
A
S
D
Y
X
C

4.

CONCLUSION

La consécration de la donnée comme étant une nouvelle forme de patrimoine au sein de l'entreprise a profondément bouleversé le fonctionnement des processus, la structure des systèmes d'information et l'organisation des compétences. Son exploitation s'industrialise de plus en plus, une tendance observée sur de nombreux secteurs.

Si des freins culturels peuvent parfois persister sur son usage, la théorie des graphes peut constituer un accélérateur intéressant dans la diffusion d'une compétence et d'une meilleure compréhension des possibles. En tant qu'approche plus visuelle et plus pédagogique que la Data Science, la théorie des graphes peut être utilisée dans des domaines variés, et pour un nombre important d'applications différentes : en gestion de projet pour présenter un planning et la gestion des dépendances entre les activités, en marketing pour une amélioration de l'expérience client, dans le domaine de la finance pour la gestion des risques ou la quête de rentabilité.

Applicable de façon totalement transverse, cette pratique a fait ses preuves pour présenter simplement et visuellement des processus complexes, et devient indissociable des grands piliers de la Data Science. Les avancées dans le domaine sont permanentes, et nous permettront de répondre au mieux à des problèmes que nous commençons juste à imaginer.


Code File Edit Selection View Go Debug Tasks Window Help
EXPLORER
OPEN EDITORS
package.json edgeweb-workspace/reseller
typings.json edgeweb-workspace/reseller
tsconfig.json edgeweb-workspace/reseller
api.ts edgeweb-workspace/reseller/src/app/s...
EDGESTACK
edge-deployment
edge-workspace
edgeweb-workspace
admin
org
reseller
node_modules
src
app
datatypes
directives
pages
services
api
api.ts
index.ts
app
auth
brokers
data
dates
jwt
login
orgs
readings
user
widgets
app.component.html
import { Injectable, Inject, Injector } from '@angular
import { Router } from '@angular/router';
import { AuthHttp, JwtHelper } from '../jwt/jwt';
import { Observable } from 'rxjs/Rx';
import { AuthHelper } from '../auth/auth';
import { RequestOptions, Headers, Response } from '@a
import { Config } from '../../config';
import { appInjector } from '../app/app-injector';
@Injectable()
export class ApiService {
jwt: string;
private static parseResponse(res: Response) {
var info = res.headers.get('X-Message-Info');
var error = res.headers.get('X-Message-Error'
var success = res.headers.get('X-Message-Succ
return {
message: {
info: info,
error: error,
success: success
},
json: function() {
let json;
try {
json (local var) Exception: any
} catch (Except]on) {
json = res.json();
}
return json;
}(),
res: res
};
}
constructor(
public auth: AuthHelper,
public authHttp: AuthHttp,

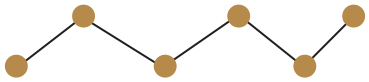
5.

GLOSSAIRE & ANNEXE

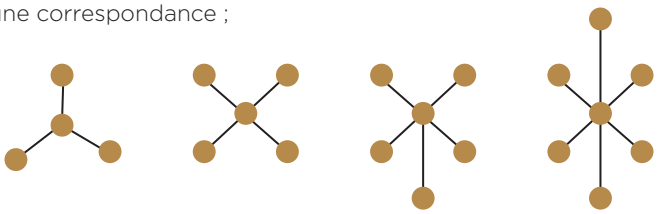
5.1 GLOSSAIRE

Plusieurs types de graphes sont possibles. Parmi les plus fréquents nous pouvons trouver :

- La **chaîne**, une suite d'arêtes adjacentes, qui dans notre exemple pourrait s'apparenter à une ligne complète de métro sans correspondance ;

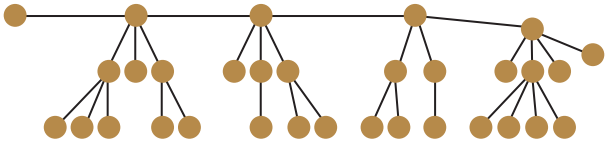


- L'**étoile**, un graphe dont seul un sommet est de degré supérieur à 1. Par exemple dans ce cas, la représentation d'une correspondance ;

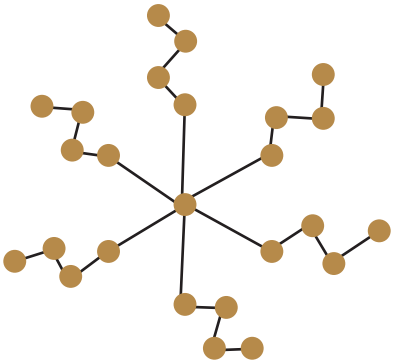


- La **chenille**, une suite d'arêtes adjacentes dont chaque noeud s'apparente à une étoile. On peut trouver ce type de graphe pour schématiser les différentes correspondances possibles si on emprunte cette ligne ;

- Le **homard** s'obtient en partant d'une chenille et en rajoutant quelques brindilles supplémentaires à ses sommets ;



- L'**araignée** est un graphe ayant un sommet central duquel partent plusieurs chemins (ses pattes). Si on évolue au niveau national, on peut imaginer le réseau SNCF comme une araignée centrée sur Paris.



- **Matrice d'adjacence :** en considérant N titres financiers représentés sous forme de graphe, chacun des liens entre paires de sommets peut être répertorié/synthétisé dans une matrice, appelée matrice d'adjacence. Cette matrice comporte N lignes et colonnes, et ses éléments représentent numériquement les liens entre les sommets.

	A	B	C	D
A	0	1	1	0
B	1	0	0	1
C	1	0	0	1
D	0	1	1	0

Cette matrice d'adjacence nous indique qu'il existe un lien entre l'actif A et l'actif B. À contrario, il n'y a pas de lien entre l'actif B et l'actif C.

Sur l'exemple ci-dessus, les liens sont représentés de manière binaire : "1" représentant un lien existant, "0" symbolisant l'absence de lien. De plus, et dans cet exemple, la matrice d'adjacence est symétrique : le sommet A est lié au sommet B et le sommet B est lié au sommet A. Il n'y a pas de sens dans le lien entre deux actifs. Cette représentation peut donc être enrichie en considérant :

> des valeurs quelconques : les liens seront plus ou moins importants en fonction de la valeur prise par l'élément de la matrice ;

> une matrice d'adjacence non symétrique : les liens pourront être unidirectionnels ou bidirectionnels avec des intensités plus ou moins fortes.

L'avantage de cette représentation matricielle des graphes est la facilité de stockage des informations et la réalisation d'opérations matricielles sur celle-ci : union, produit, intersection, etc.

5.2 ANNEXE

Théorie moderne du portefeuille de Markowitz et la corrélation

La théorie moderne du portefeuille, développée et formalisée par Harry Markowitz, économiste et prix Nobel en sciences économiques, vise à construire le portefeuille le plus efficient possible, c'est-à-dire dont la rentabilité est maximale pour un niveau de risque donné. Autrement dit, il s'agit d'un problème d'optimisation entre le rendement espéré ou attendu, représenté par le rendement moyen des actifs, pondérés de leur poids dans le portefeuille, et le risque inhérent aux titres représenté par la volatilité du portefeuille.

$$\sigma_{\rho}^{2} = \sum_{i=1}^{n} w_i^2 \sigma_i^2 + 2 \sum_{i=1}^{n} \sum_{i=1}^{n} w_i w_j \sigma_{ij}$$

Avec : $\sigma_{ij} = \sigma_i \sigma_j \rho_{ij}$ Où :
- ρ_i correspond à la corrélation entre le titre i et j
- σ_i (resp. σ_i) l'écart type du titre i (resp. j) et σ_i^2 (resp. σ_j^2) sa variance
- w_i (resp. w_j) la proportion du titre i (resp. j) dans le portefeuille p.

La mutualisation des risques par le biais de la diversification est à la base de cette construction, et consiste à établir un portefeuille de telle sorte que le risque attaché à un titre compense celui d'un autre. Le risque d'un actif financier est calculé en fonction de sa volatilité, et plus précisément d'un indicateur appelé variance. Un actif a une forte volatilité, et par conséquent un risque élevé, lorsque ses rentabilités journalières sont très dispersées autour de la moyenne de ses rentabilités.

Formellement, le risque du portefeuille s'écrit comme une fonction somme dépendant de la corrélation, ainsi que de la dispersion de chaque titre autour de sa moyenne pondérée par sa proportion dans le portefeuille.

En imaginant désormais trois scénarii afin de comprendre l'impact de la corrélation dans l'évaluation du risque d'un portefeuille donné : deux scénarii correspondent aux cas où les actifs sont corrélés, soit négativement soit positivement, et le troisième où il n'y a aucune corrélation. Dans chacun des scénarii, la proportion d'actifs A et B est identique.

D'après la formule de calcul du risque, il est facile d'imaginer que plus les actions sont corrélées, donc avec un coefficient de corrélation proche de 1, plus le terme additif dépendant de la corrélation viendra accroître le calcul du risque. A l'inverse, plus la corrélation est proche de -1, plus le terme dépendant de la corrélation débouchera sur une diminution du résultat global du calcul du risque.


(! function_exists('incode_starter_s
Sets up theme defaults and registers s
Note that this function is hooked into
runs before the init hook. The init hoo
as indicating support for post thumbnai
function incode_starter_setup() {

6.

DOMAINES D'EXCELLENCE & CONTACTS

DONNER DU FUTUR AU TALENT

Fondé en 2008, Square est un cabinet de conseil en stratégie et organisation. 1er cabinet de conseil indépendant en France, en Belgique et au Luxembourg, Square est, avec ses 700 consultants, l'un des rares acteurs du marché à proposer une gamme d'expertises aussi étendue.

Square guide ses clients en mettant à leur disposition ses compétences et son expérience sur 8 domaines d'excellence :

INNOVATION
Square accompagne ses clients dans la transformation de leur dynamique d'innovation. Nos consultants, par leur approche sur-mesure, aident à concevoir, industrialiser et gouverner l'innovation pour assurer la croissance durable des entreprises et leur transformation en entité socialement et écologiquement responsable.

PEOPLE & CHANGE
Square aide ses clients à acquérir, fédérer et développer le capital humain de leur organisation. Afin de créer davantage d'engagement au sein des équipes, nos interventions portent principalement sur l'adaptation des méthodes de travail aux changements opérationnels et culturels, l'efficacité des directions des ressources humaines et le développement des compétences.

MARKETING
Square accompagne ses clients sur l'ensemble du spectre marketing : marketing stratégique, marketing relationnel, marketing de l'offre, communication, tarification, satisfaction clients. Nos expertises, initialement centrées sur les secteurs de la banque et de l'assurance, s'adressent désormais à l'ensemble des industries ou services B2C.

DATA
Square élabore des stratégies Data et assure leurs déclinaisons opérationnelles à travers la conduite de projets de Data Management, Data Analyse et Data Science. Notre approche experte et pragmatique vise à valoriser et sécuriser le patrimoine de données des entreprises.

DIGITAL
Square accompagne ses clients dans l'élaboration de leur stratégie digitale, la conception et la mise en œuvre de nouveaux parcours digitaux pour leurs clients ou leurs collaborateurs, ainsi que dans l'ensemble des chantiers d'acculturation interne et d'accompagnement aux nouvelles méthodes de conception.

RISK & FINANCE
Square prend en charge le pilotage des programmes de maîtrise des risques financiers et non financiers, ainsi que la transformation des fonctions Risque et Finance face à l'évolution des dispositifs prudentiels et à l'irruption des problématiques liées à la maîtrise de la donnée.

REGULATORY & COMPLIANCE
Square conseille ses clients dans le déploiement des nouvelles réglementations, ainsi que dans l'optimisation et le renforcement des dispositifs de contrôle. Ce domaine d'excellence s'appuie sur une communauté d'experts de 130 consultants qui, outre les missions auprès des clients, conduit d'importants travaux d'investigation et de publication.

SUPPLY-CHAIN
Square assure l'excellence opérationnelle de la logistique, des achats aux derniers kilomètres, avec des parcours clients différenciants. Nos experts conçoivent des solutions omnicanales mettant en œuvre les meilleures pratiques des systèmes d'informations, de la mécanisation à la robotisation.

Rédigé par les consultants Square du domaine d'excellence Data, ce book propose de revenir sur une spécialisation méconnue au sein des sciences de la donnée : la théorie des graphes. Loin d'être une nouveauté, le Graph Thinking présente un intérêt particulier au moment où Intelligence Artificielle et Machine Learning prennent une place grandissante dans nos quotidiens et inquiètent autant qu'elles fascinent. En effet, plutôt que de choisir une approche ultra technique réservée à une poignée d'experts, la théorie des graphes propose de vulgariser la science de la donnée avec des représentations visuelles et des usages à fort ancrage opérationnel. Ce focus en propose donc un aperçu avec une présentation théorique accessible et deux exemples de mise en application métier.

Crédits photos - Unsplash :Shunya Koide (pp. 1, 2), Bantersnaps (p.6), Vincent Nicolas (p.18), Toa Heftiba (p.20), Manuel Nobauer (p.25).

CONTACTS

JULIEN GUIBERT
PARTNER
+33 6 67 56 90 02
julien.guibert@square-management.com

MARC CAMPI
PARTNER SQUARE
+33 6 84 02 68 59
marc.campi@square-management.com

ADRIEN AUBERT
ASSOCIATE PARTNER
+33 6 69 63 06 01
Adrien.aubert@square-management.com

Square

Square

DONNER DU FUTUR AU TALENT

square-management.com