**Knowledge Maps, méthodologie (Shou)**

**Survol**

Les *fuzzy cognitive maps* (FCM) ont pour but de faciliter la compréhension des flux d’information d’un système sur la base d’une métaphore spatiale, en caractérisant les relations entre les entités/objets formant ce système par des règles simples (*fuzzy rules*). Mais alors que les FCM reposent souvent sur une base experte, étant par là-même sujettes aux biais subjectifs de l’analyste et se révélant rapidement illisibles au fur et à mesure de l’ajout de nouvelles variables issues du système, les *knowledge maps* (KM) exploitent directement et automatiquement les données brutes descriptives du système, de manière exclusivement statistique, sans passer par la modélisation. Dans ce sens, les KM représentent une formalisation et un avancement des FCM.

Une KM peut être exprimée sous la forme , où l’on trouve :

* , soit un ensemble de **nœuds** représentant des variables ou des concepts décrivant le système analysé.
* , soit les **liens** joignant des paires de nœuds dans ce système, tels que et . **Ces liens incluent les *fuzzy rules* décrivant les relations causales entre les variables**.
* , soit un ensemble de **connecteurs** mappant à . est le **coefficient généralisé de** [**corrélation croisée**](https://en.wikipedia.org/wiki/Cross-correlation) (*generalized cross correlation coefficient*) entre et . est **l’entropie** de l’image du graphique de dispersion correspondant.

L’utilisation des connecteurs est la différence principale entre les FCM et les KM ( ?). Le coefficient de corrélation croisée décrit le degré de *fuzzyness* de la règle : plus la mesure de corrélation est forte, plus la relation entre les variables indique une influence mutuelle et une causalité probable. L’entropie décrit le degré de désorganisation dans la relation entre les variables, respectivement dans le système.



Exemple de KM, représentée comme une variante de [matrice structurelle](https://en.wikipedia.org/wiki/Design_structure_matrix). Les nœuds sont alignés sur la diagonale et les liens sont représentés par deux segments vertical et horizontal, avec un point noir au coude. Différents ensembles de nœuds (variables provenant de différentes sources, ou bien représentant par exemple des inputs et outputs du système) peuvent être représentés par différentes couleurs. Les nœuds importants (*hubs*) sont représentés en cercles rouges, et les nœuds non corrélés au reste du système apparaissent comme des carrés blancs (ici, invisibles).

La lecture de la carte doit permettre :

* L’identification de relations de cause à effet entre les variables
* La visualisation du flux d’information dans le système
* Le classement des variables d’importance grâce aux hubs et aux nœuds inactifs

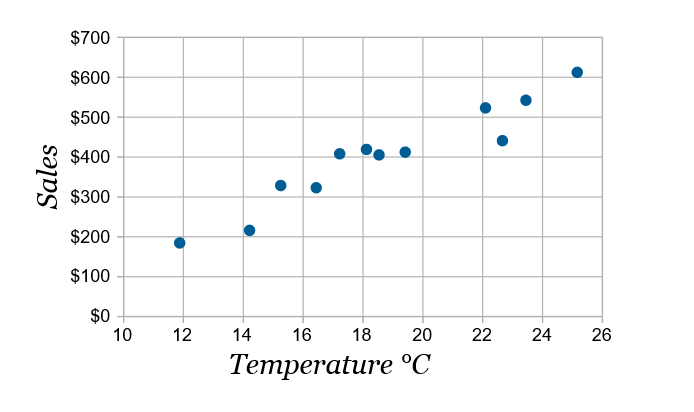
**Méthode de construction de la KM**

Les données récoltées sont entrées dans une matrice :

Ici, représente la k-ème observation de la j-ème variable . L’objectif est le minage de cette matrice afin d’obtenir une KM. Ce minage s’effectue en quatre étapes.

1. Construction des graphiques de dispersion

Toute les paires de variables et telles que sont utilisées afin de construire une série de graphiques de dispersion (nombre total de graphiques : ).



Pour chaque graphique de dispersion, une première mesure est effectuée : **le coefficient généralisé de corrélation**, qui peut prendre une valeur de 0 à 1 et se calcule de la manière suivante :

où représente l’[information mutuelle](https://fr.wikipedia.org/wiki/Information_mutuelle) entre et et permet de mesurer la dépendance linéaire et colinéaire entre les deux variables. L’information mutuelle représente le degré de dépendance statistique entre deux variables, qui ne doit pas être confondu avec la causalité, bien qu’en pratique, l’un implique souvent l’autre. A son tour, se calcule ainsi :

[[python](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mutual_info_score.html) : [sklearn.metrics](https://scikit-learn.org/stable/modules/classes.html#module-sklearn.metrics).mutual\_info\_score]

Ici, représente la fonction de probabilité de densité (*probability density function*) des paires de variables. Chaque variable est ici discrétisée et mappée sur intervalles. Si , alors ne contient aucune information sur . A l’inverse, signifie que et sont si parfaitement corrélées que l’une peut entièrement déterminer l’autre.

Une seconde mesure est effectuée : **l’entropie**. L’entropie de l’image d’un graphique de dispersion se mesure de la façon suivante :

[[python](https://kite.com/python/answers/how-to-calculate-shannon-entropy-in-python) : [pandas.core.algorithms.value\_counts()](https://kite.com/python/docs/pandas.core.algorithms.value_counts) [scipy.stats.entropy()](https://kite.com/python/docs/scipy.stats.entropy)]

L’entropie est une mesure de la désorganisation d’un graphique. Les graphiques contenant une grande quantitié d’incertitude contiennent également beaucoup d’entropie.

De manière générale, si l’entropie d’un graphique est élevée tandis que le coefficient généralisé de corrélation est bas, cela indique l’absence d’une relation significative entre les deux variables. Dans ce cas, le graphe ne contient pas de fuzzy rule non plus et aucun lien n’est créé entre les variables dans la KM.

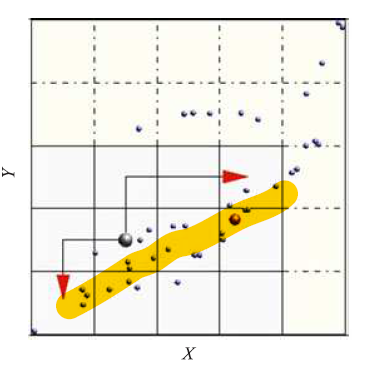
1. Génération des fuzzy rules

Les fuzzy rules sont des règles représentant le comportement d’une variable par rapport à l’autre, ou encore l’influence que le comportement de l’une a sur l’autre. Ces règles peuvent prendre quatre formes :

* IF + DELTA X THEN DELTA Y
* IF – DELTA X THEN DELTA Y
* IF + DELTA Y THEN DELTA X
* IF – DELTA Y THEN DELTA X

Les fuzzy rules effectives sont représentées par des quadruples, tels que . Ce quadruple signifie que si X augmente ou décroit d’une unité, Y change simultanément d’une unité, et inversément. Un quadruple peut être établi sur la base de chaque cellule originale (ou bien de chaque point de chaque cellule ? Il est de toute manière claire que plusieurs quadruples peuvent être extraits de chaque graphe, puisque le critère de création des liens à l’étape suivante implique qu’au moins un quadruple existe par graphe pour établir un lien entre les deux variables concernées).

L’espace du graphe est segmenté en **fuzzy levels** (généralement 3, 5 ou 7). Chaque variable est ensuite classée dans l’une de trois catégories (si l’on a choisi 3) : low, medium, high. PARAGRAPHE PAS CLAIR.



Dans cet espace, la fuzzy rule est .

On analyse ensuite les graphes spécifiques (segmentés en fuzzy levels) : chaque cellule est successivement définie comme origine, puis on bouge vers la droite (+ DELTA X), la gauche (- DELTA X), le haut (+ DELTA Y) ou le bas (- DELTA Y) afin de déterminer vers quelle nouvelle cellule on va le plus probablement se diriger (= dans quelle mesure l’autre variable va se modifier).

La tendance est ensuite enregistrée, puis transformée en fuzzy rule et en quadruple. S’il n’y a pas de tendance claire, ou s’il est possible d’arriver également sur plusieurs autres cellules, alors aucune fuzzy rule n’est établie.

1. Construction de la carte

La carte est graphiquement construite selon le modèle présenté dans le survol. Les liens entre les nœuds sont établis sur la base des fuzzy rules dégagées à l’étape 2. S’il existe au moins un quadruple dans le graphique de dispersion, alors les deux variables en questions se verront attribuer un lien dans la KM. Sinon, les variables ne sont pas connectées. Les liens entre deux variables sont assimilés à des connecteurs, lesquels représentent le coefficient généralisé de corrélation croisée et l’entropie d’image existant entre les variables, indiquant dans quelle mesure celle-ci s’influencent, ainsi que le niveau d’incertitude dans leur relation.

Les entropies des chaque graphique de dispersion sont sommées afin d’obtenir l’entropie totale de la carte :

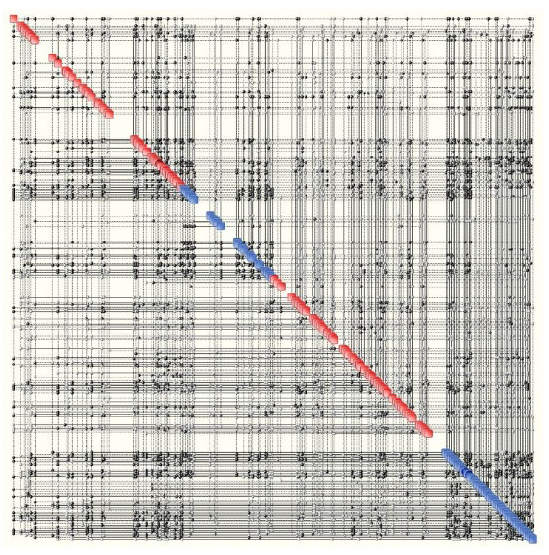
1. Identification des hubs et des nœuds inactifs

La KM permet ensuite d’identifier les variables qui ont le plus d’influence dans le système, et celles qui sont relativement indépendantes. Des mesures de théorie des graphes sont appliquées aux nœuds. Premièrement, [l’intermédiarité](https://fr.wikipedia.org/wiki/Centralité_intermédiaire) de chaque nœud est mesurée (*betweenness centrality*), à savoir le nombre de fois que ce nœud se trouve sur le chemin le plus court entre deux autres nœuds du graphe. Deuxièmement, la [centralité](https://fr.wikipedia.org/wiki/Centralité) de chaque noeud (*degree centrality*) est mesuré par le nombre de liens qui y touchent. Le nœud doté de la plus forte centralité est qualifié de hub.

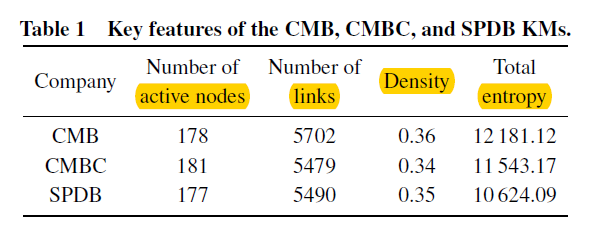
Enfin, la densité de la carte de calculée sur la base du nombre total de liens et du nombre total de nœuds actifs :

Analyse statique (exemple)

L’analyse statique utilise la totalité des données entre 2002 et 2010, avec un fuzzy level fixé à 5. Un exemple de KM est donné ici :



Les données du bilan et de la trésorerie sont en rouge, tandis que les données fiscales et boursières sont en bleu. Les quatre mesures principales de la carte (ainsi que ces mêmes mesures pour deux autres compagnies), à savoir le nombre de nœuds actifs, le nombre de liens, la densité et l’entropie totale, sont rassemblées dans le tableau suivant :



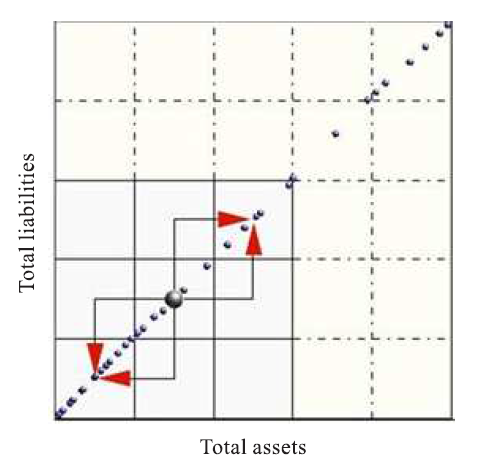
Le nombre de nœuds signifie qu’approximativement 39% des variables n’ont pas de relation spécifique particulière avec le reste de la carte. Une partie de ces cas résulte du fait que les variables n’ont pas assez d’observations.

Le nombre de liens et la densité indiquent que les relations entre les variables sont plus rapprochées pour CMB, ce qui signifie que cette compagnie est relativement plus difficile à contrôler, du fait de la plus forte propagation de la moindre perturbation à l’ensemble des nœuds.

L’entropie également plus élevée de CMB signifie que la structure de l’entreprise a plus de chance de collapser au contact d’un environnement turbulent.

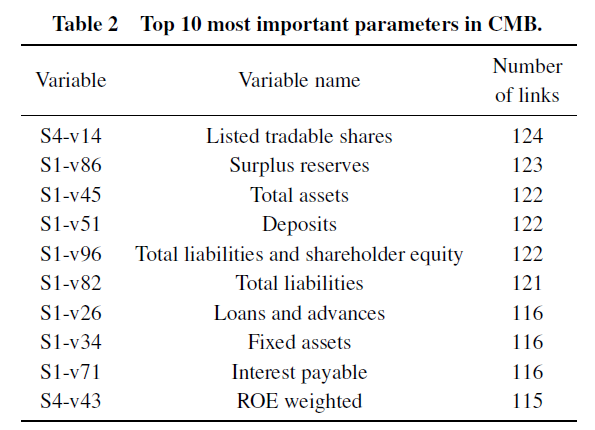
CMB présente la structure la plus compliquée et la plus porteuse d’incertitude. Elle est capable de réaliser plus de tâches, mais également la plus susceptible de rencontrer un échec. Un système très complexe présente plusieurs modes de fonctionnement et les [transitions de phase](https://en.wikipedia.org/wiki/Phase_transition) entre ces modes sont aussi soudaines qu’imprévisibles à mesure que croissent complexité et entropie. De plus, dans un environnement très perturbant, la compagnie peut éventuellement devenir incontrôlable.

Les paramètres de bilan et boursiers interagissent le plus avec les autres paramètres afin de créer le plus de liens. De tous les liens de la KM de CMB, ‘Total Assets’ et ‘Total Liabilities’ ont une corrélation de 0.91. Leur graphe de dispersion se présente sous la forme suivante :



L’entropie d’image de ce graphe est de 1.63 (relativement basse, presque nulle, c’est-à-dire que la relation entre les variables est structurée et informante. Le comportement de l’une des deux variables permet de prédire fidèlement celui de l’autre). Les variables sont coordonnées, ce qui est exprimé par la fuzzy rule du graphe : (+1, -1, +1, -1). Cette analyse permet de saisir les interdépendances entre les données : ici, l’interdépendance est forte. Un graphe de dispersion flou révèle au contraire une interdépendance forte.

Les dix paramètres les plus importants sont données dans ce tableau :

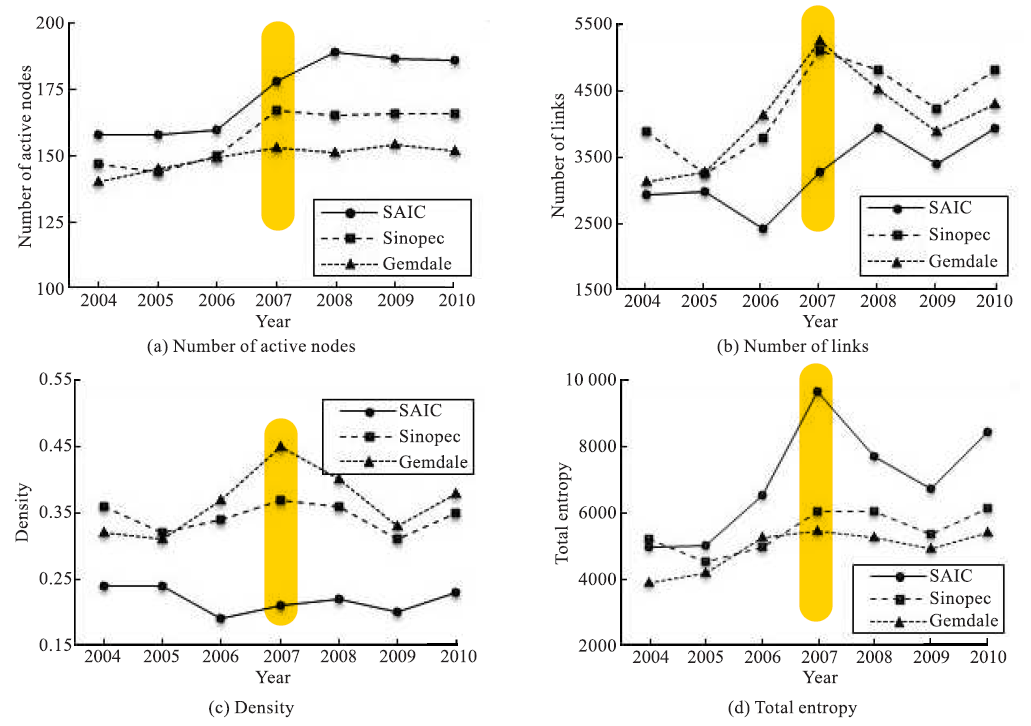


C’est le nombre de liens originaires du paramètre qui détermine le classement. Ici, c’est la variable S4-v14 (‘Listed Tradable Shares’) qui est la plus importante dans la structure. Notons que de ces dix variables, huit émanent du bilan. Ce sont ces variables, respectivement le traitement de ces postes de bilan qui représentent le meilleur levier pour une action interne ou externe efficace sur la compagnie.

Analyse temporelle

Dans l’analyse temporelle, les échantillons pour trois compagnies non financières sont splittés en différentes fenêtres continues. Chaque fenêtre est analysée séparément et toutes les fenêtres se recoupent afin de ne pas introduire de différence significative entre les étapes. Ici, la largeur de la fenêtre (*window width*) est fixée à 12 et le pas de recoupement (*overlap step*) à 11. On a donc des fenêtres annuelles qui se recoupent d’un mois.

Voici les données dynamiques regroupées dans quatre tableaux (nombre de nœuds actifs, nombre de liens, densité, entropie totale) :



En 2007, le **nombre de nœuds actifs et de liens (structure)** a cru dans les trois compagnies. La **densité (complexité)** et **l’entropie (incertitude)** ont également crû cette année-là. Ces signaux doivent être considérés comme des précurseurs d’une crise. Une analyse plus approfondie des compagnies composants l’indice Shanghai 50 a montré que 75% de celles-ci ont expérimenté une hausse des liens en 2007. Il semble que les mesures critiques des KM aient généralement augmenté en 2007.