



Les indicateurs structurels d'un graphe: calculs, visualisation, interactivité

Brigitte Gay, Bernard Dousset, Radwen Wanassi

► To cite this version:

Brigitte Gay, Bernard Dousset, Radwen Wanassi. Les indicateurs structurels d'un graphe: calculs, visualisation, interactivité. 5ème séminaire de Veille Stratégique Scientifique & Technologique (VSST 2013), Oct 2013, Nancy, France. pp.0. hal-02639447

HAL Id: hal-02639447

<https://hal.science/hal-02639447>

Submitted on 28 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Open Archive Toulouse Archive Ouverte

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible

This is an author's version published in: <https://oatao.univ-toulouse.fr/22141>

To cite this version:

Gay, Brigitte and Dousset, Bernard and Wanassi, Radwen *Les indicateurs structurels d'un graphe: calculs, visualisation, interactivité*. (2013) In: 5ème séminaire de Veille Stratégique Scientifique & Technologique (VSST 2013), 23 October 2013 - 25 October 2013 (Nancy, France). (Unpublished)

Any correspondence concerning this service should be sent to the repository administrator: tech-oatao@listes-diff.inp-toulouse.fr

Les indicateurs structurels d'un graphe: calculs, visualisation, interactivité

Brigitte GAY (*), Bernard DOUSSET (**), Radwen WANASSI (**)

b.gay@esc-toulouse.fr, bernard.dousset@irit.fr, wanesradwen@gmail.com

(*) Ecole Supérieure de Commerce de Toulouse (France)

(**) IRT Université Paul Sabatier, Toulouse (France)

Mots clefs :

[Veille scientifique et technologique](#), systèmes complexes, graphes, réseaux sociaux, indicateurs, centralité, intermédiarité, visualisation, interactivité

Keywords:

Scientific and technological vision, complex systems, graphs, social networks, measures, centrality, betweenness, visualization, interactivity

Palabras clave :

Vision científica y tecnológica, los sistemas complejos, gráficos, redes sociales, indicadores , centralidad, intermediación, visualización, interactividad

Résumé

L'analyse des réseaux à grande échelle et notamment des réseaux sociaux est de nos jours un défi majeur pour chercher des informations utiles comme l'identification des acteurs les plus importants pour certains critères donnés. Dans cette communication, nous visons à présenter l'état actuel de la recherche sur les mesures de centralité et leurs applications dans l'analyse des réseaux de la recherche et ceux concernant les relations entre entreprises. Nous illustrons également l'extrême importance d'une utilisation réfléchie des mesures de centralité existante. A cet effet, nous présenterons quelques études expérimentales et théoriques de corrélations entre les mesures de centralités issues de l'analyse des réseaux sociaux.

Abstract

The analysis of large-scale networks, including social networks, is nowadays a major challenge specially when finding useful information such as the identification of the most important players within a certain criteria. In this paper, we aim to present the current state of research on the centrality measures and their applications in network analysis and research concerning the relationship between companies. We also illustrate the extreme importance of the wise use of existing centrality measures. To this end, we present some experimental and theoretical studies of correlations between measures of centralities from the analysis of social networks.

Introduction

Ces dix dernières années, la modélisation des systèmes d'interdépendances a pris une place importante dans le décisionnel avec notamment l'avènement des réseaux sociaux (Facebook, LinkedIn, Viadeo, Myspace, Flickr, Dailymotion, ...), des blogs, des tweets, (Internet) sans compter les multiples réseaux sociaux d'entreprises (Intranet) et les liens entre entreprises : fusions, acquisitions, alliances, licences, échanges de connaissances, alignement des systèmes d'information (Extranet). L'analyse du fonctionnement de ces réseaux permet d'en comprendre les mécanismes, d'observer les comportements, de prendre en compte le relationnel dans l'apprentissage, la transmission des savoirs et son impact sur la vie sociale et sur le développement de l'entreprise. Pour arriver à analyser le fonctionnement de telles structures souvent complexes et fortement connectées il est nécessaire d'avoir recours à des outils performants de visualisation qui, en mettant en avant certains éléments remarquables de la structure, permettent de situer les éléments clés souvent centraux par lesquels passe une part importante des relations et des flux d'information. Dans ce but, nous avons ajouté, à notre outils VisuGraph dédié à l'analyse des graphes en statique comme en évolutif, différentes mesures de centralités permettant de guider l'utilisateur vers les éléments remarquables en fonction de son point de vue et des propriétés qu'il veut mettre en avant. Après un panorama de diverses mesures proposées dans la littérature, nous proposerons une implémentation de ces mesures dans nos outils et nous présenterons quelques exemples d'utilisation sur des graphes issus de nos études stratégiques.

1 Les développements actuels de la théorie des graphes

La théorie des graphes est une branche un peu à part des mathématiques. Son efficacité dans la résolution des problèmes de recherche opérationnelle et son aspect synthétique pour la représentation des problèmes les plus complexes via les relations topologiques entre les entités d'un graphe sont les deux facteurs essentiels qui ont fait son succès. Ceci explique son développement croissant dans les sciences sociales, physique et biologique d'où l'apparition de modèles de graphes théoriques tel que les « réseaux petits mondes » et les « réseaux invariants d'échelle ».

1.1 Les modèles de graphes théoriques

1.1.1 Réseaux aléatoires

Un graphe aléatoire est un graphe qui est généré par un processus aléatoire. Un graphe aléatoire de taille N est un graphe à N sommets dont on a choisi aléatoirement les arêtes. Le modèle d'Erdős-Rényi suppose que l'existence de chaque arête est indépendante de celle des autres et que chaque arête a une probabilité p d'exister. La distribution des degrés des sommets du graphe suit une loi de Poisson. Ce modèle de graphe théorique repose donc sur un modèle de construction qui se caractérise par une loi statistique. En cela, il constitue le premier modèle de graphe théorique. Les graphes aléatoires s'opposent aux graphes réguliers peu présents dans la nature (un graphe est dit régulier quand tous les sommets ont le même nombre de voisins, c'est à dire le même degré).

1.1.2 Réseaux petits mondes (small world network)

Les réseaux de type « petit monde » (small-world) décrits par [Watts-Strogatz 98] sont des graphes où chaque nœud est connecté avec son voisin le plus proche avec des connexions réciproques aléatoires, et une distance topologique moyenne faible, ce qui traduit une forte connectivité. Ce modèle de réseau a connu un succès rapide, principalement dû à son lien direct avec les travaux de Stanley Milgram concernant les fameux « six degrés de séparation » : seules six personnes nous séparent de n'importe quelle autre personne dans le monde. Ce genre de graphe possède alors un coefficient d'agglomération supérieur (en fait très supérieur) et un diamètre inférieur à un graphe aléatoire de même ordre et de même taille.

1.1.3 Réseau invariant d'échelle (Scale-free network)

Introduit par Barabasi et Albert, ces graphes sont caractérisés par une distribution de degré qui suit une loi de puissance, c'est-à-dire à des graphes où il existe une large distribution de degrés différents, décroissant très lentement, ce qui traduit une hiérarchisation des nœuds selon leur degré. Pour expliquer cette distribution inégale des degrés on peut prendre l'exemple d'Internet. La création d'un site Internet sur un sujet précis implique la création de rubrique de liens vers les sites de référence de ce sujet qui envoie de façon préférentielle vers les sites de référence concernant ce sujet, et la centralité de ces sites de référence tend à augmenter au cours du temps. Les nœuds

« hub » de ce genre de réseau sont connectés directement à ceux de plus petites tailles et ceci permet une tolérance aux pannes même en cas de faillite d'un nœud central en terme de degré. La distribution du coefficient d'agglomération dans ce modèle de réseau suit une loi de puissance : ce coefficient est inversement proportionnelle à l'augmentation du degré d'un nœud, ce qui implique que les nœuds de faible degré appartiennent à des « small world networks ». Plusieurs réseaux du monde réel sont modélisés par ce modèle de graphe tel que les réseaux sociaux, les réseaux de paiement interbancaires, ...

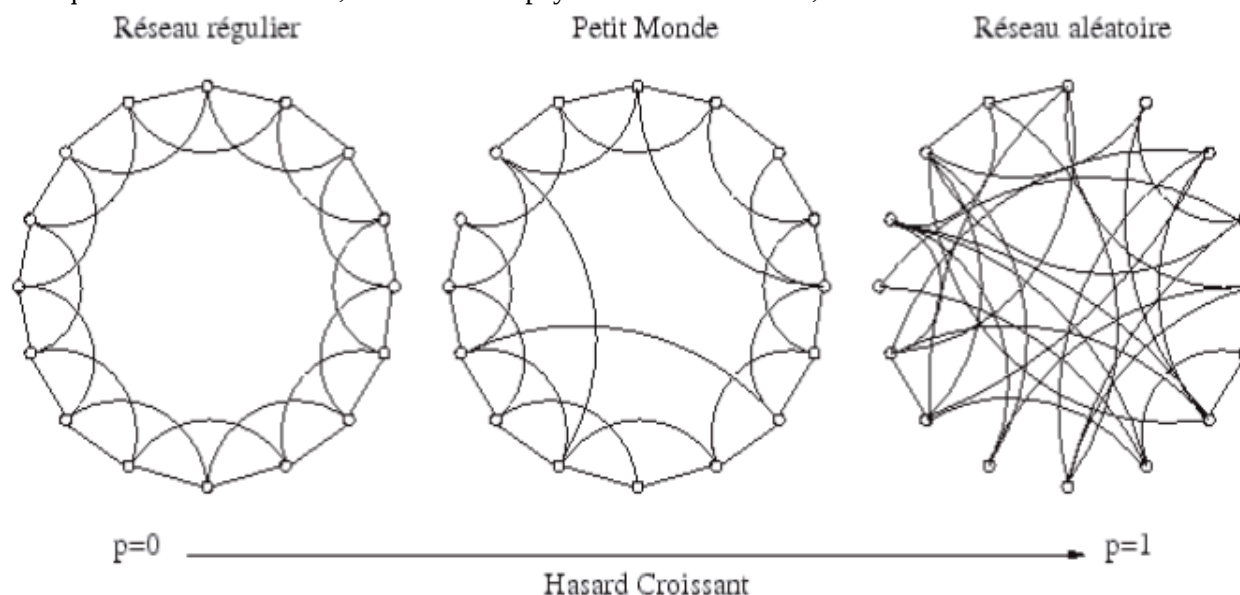


Figure 1: Les différents types de réseaux (Source: [Watts et Strogatz 98])

1.2 Théorie des graphes et réseaux sociaux

Suivant le type de graphe, on peut déterminer plusieurs critères relationnels du réseau. L'étude des graphes binaires dans lesquels les valeurs des liens sont soit 0, soit 1, l'apport d'information vient de l'existence d'un lien entre deux nœuds-acteurs du réseau. Par contre, dans les réseaux pondérés, où les liens sont pondérés par des nombres réels, on indique non seulement l'existence du lien mais aussi on donne une description du lien qui pourrait être, par exemple, la fréquence d'interaction entre les agents. [Granovetter 73], [Granovetter 82] classe les liens faibles et les liens forts en fonction de la durée de la relation, son intensité. Les liens forts sont représentatifs de l'intégration locale alors que les liens faibles assurent l'intégration globale du réseau. Les différents liens sont ainsi les indicateurs d'une position structurale des agents, montrant leurs pouvoir et leur confère une position privilégiée dans les échanges d'informations et de flux des données étudiées.

Les graphes symétriques (respectivement asymétriques) sont des réseaux ayant une matrice symétrique (respectivement asymétrique). Les matrices symétriques ont des valeurs symétriques par rapport à la diagonale. Les réseaux asymétriques sont aussi dits réseaux orientés. Dans les graphes orientés, les nœuds possèdent deux types de liens (arcs), à savoir des liens entrants et des liens sortants. Chaque nœud possède alors deux mesures relatives aux deux types de liens : une mesure relative à ses liens sortants, appelée d'influence ou de centralité sortante, et une autre mesure relative à ses liens entrants, dite mesure de prestige, de popularité, d'autorité ou de centralité entrante [Wasserman & Faust 94].

1.3 Application : Les réseaux des entreprises

Dans le domaine de l'économie de l'information et plus précisément l'économie de l'organisation centrée sur l'information, les économistes sont déjà basé sur les mesures utilisés dans les sciences sociales pour développer leurs propres outils analytiques leur permettant ainsi de détecter les possibilités d'alliances et l'existence de dépendance et la distribution de pouvoir dans un réseau donné. Les indicateurs de centralités dont les caractéristiques structurelles de porté, permettent à l'économiste de combiner les différents niveaux d'analyses globales, locales ou bilatérales et de comprendre les relations de dépendance ou d'interdépendance entre les acteurs d'un marché donné.

1.4 Influence du type de graphe sur les indicateurs de centralités

L'analyse d'un réseau complexe de données réelles doit commencer par s'interroger sur la nature de la centralité qu'il faut prendre en compte, à savoir la centralité locale, ou la centralité globale. Savoir quelle est la combinaison de nœuds et de liens qu'il faut prendre en compte pour analyser un réseau dans un contexte donné est primordial, mais avant de poser la question de choix de mesures de centralités, l'évaluation des corrélations qui peuvent exister entre les mesures à utiliser est nécessaire. En effet, l'existence d'une forte corrélation implique une redondance de l'information et donc l'analyse est peu informative quant aux particularités qui peuvent rester indétectable par cette famille liée de mesures. Cependant, si les mesures sont peu corrélées, l'analyse du réseau, via ces mesures, sera meilleure en quantité et en qualité. Plusieurs recherches faites dans ce sens ont prouvé qu'une étude de la topologie du réseau en lui-même peut favoriser le choix de certaines mesures. Les études expérimentales sur des données réelles ou expérimentales ont été menées par des chercheurs comme [Bolland 88] pour déterminer les corrélations qui peuvent exister entre les différentes mesures de centralités. Une corrélation forte implique que peu d'informations peuvent être utile et que les mesures utilisés formes une famille liée. En effet, plusieurs facteurs peuvent influencer le degré de corrélation entre les mesures tels que le type de la matrice : symétrique ou asymétrique, le contexte étudié et les propriétés du réseau comme sa densité.

2 Etude des indicateurs et métriques des graphes

Les modèles théoriques de graphe étudiés ci-dessus ont entraîné la constitution d'un courant de recherche très actif. De très nombreuses mesures ont été proposées et, si certaines ne font que reprendre sous un autre nom des mesures déjà connues, d'autres sont innovantes, même s'il n'est pas toujours facile de déterminer leur degré d'utilité. Dans ce qui suit, nous allons en présenter une sélection.

2.1 Distribution de degré

La représentation de la fréquence de la distribution des degrés donne une information utile sur le degré de hiérarchie présent dans le graphe. Si le graphe est orienté, représenter le degré des liens entrant IN et celle des liens sortants OUT peut s'avérer intéressant.

2.2 Assortativité

L'assortativité est utilisé par les sociologues : les sommets partageant une caractéristique commune sont plus souvent reliés que les sommets ne partageant pas cette caractéristique. Ainsi, un réseau peut être dit assortatif quand la corrélation entre les degrés des sommets et ceux de leurs voisins est élevée. Autrement dit, les sommets ayant un degré élevé sont connectés entre eux et les sommets ayant un degré faible sont connectés entre eux et il y a peu de liens entre les sommets de degré élevé et ceux de degré faible. L'assortativité, qui dans son sens strict, rend compte de la tendance des nœuds à être connectés préférentiellement à des nœuds de degré comparable. Dans le cas contraire, le réseau est dissortatif.

2.3 Mesures de Centralités

Dans ce qui suit, nous décrivons quelques mesures de centralité dans les graphes orientés et non-orientés en provenance de l'analyse des réseaux sociaux. Celles-ci incluent les centralités de degré, d'intermédierité, de proximité, combinée ainsi que la centralité du vecteur propre (appelée aussi centralité spectrale).

2.3.1 Centralité de degré

La centralité de degré est influencée par la nature du graphe : pondéré ou non. Dans le cas d'un graphe orienté, la centralité de degré peut être divisée en trois sous-mesures: la centralité de degré, celle en rapport aux liens entrants et celles des liens sortants.

- **Degré binaire**

La centralité de degré [Freeman 79] est la forme la plus intuitive de toutes les mesures de centralité. Elle définit le nombre d'acteurs avec lesquels un nœud est lié directement et donc elle mesure l'importance d'un acteur qui est ici directement proportionnelle à sa valeur de centralité de degré. Cette mesure, exprime le nombre d'alternatives disponibles pour un acteur de la structure. Augmenter le nombre d'alternatives augmente le pouvoir d'un acteur. La centralité de degré est aussi appelée mesure de centralité locale [Scott 00] car elle ne prend pas en compte la structure globale du graphe vu qu'elle est déterminée directement par le voisinage immédiat du nœud pris en considération. L'apport d'information de cette mesure peut être limité dans certains cas.

Soit $G = (V, E)$ un graphe d'ordre N et $[a_{ij}]$ sa matrice d'adjacence binaire. Dans le cas où le graphe G est non-orienté, la centralité de degré d'un nœud $v_i \in V$ est défini par :

$$C^{deg}(v_i) = \frac{1}{N-1} \sum_{j=1}^N a_{ij}$$

Dans le cas où le graphe G est orienté, chaque nœud $v_i \in V$ possède alors deux mesures de centralité de degré : une par rapport aux liens entrants et une par rapport aux liens sortants.

- **Degré de Pondération**

La notion de centralité de degré a ensuite été étendue à la somme des poids, si le graphe est pondéré [Barrat et al., 2004] [Newman, 2004] [Opsahl et al., 2008] . Cette mesure exprime un effet de « résistance » du nœud dans le graphe. Bien que préféré pour analyser les réseaux pondérés, le degré pondéré et le degré simple expriment deux notions différentes, car le degré pondéré mesure l'engagement d'un acteur dans un réseau indépendamment du nombre de voisins auxquels il est connecté. Cette mesure est formulée comme suit:

$$S_i = C_D^W(i) = \sum_j^N a_{ij}$$

Où A est maintenant la matrice d'adjacence dans laquelle a_{ij} a une valeur supérieure à zéro si le nœud i est connectée au nœud j .

- **Centralité Combinée**

D'après [Opsahl & al. 10] la combinaison des deux mesures précédentes (degré binaire et degré pondéré) sera plus utile en vue d'extraire une information pertinente : car elle combine la topologie et le poids. En effet, le degré binaire considère tous les liens de la même façon (0 absence du lien et 1 présence du lien), alors que le degré pondéré ne

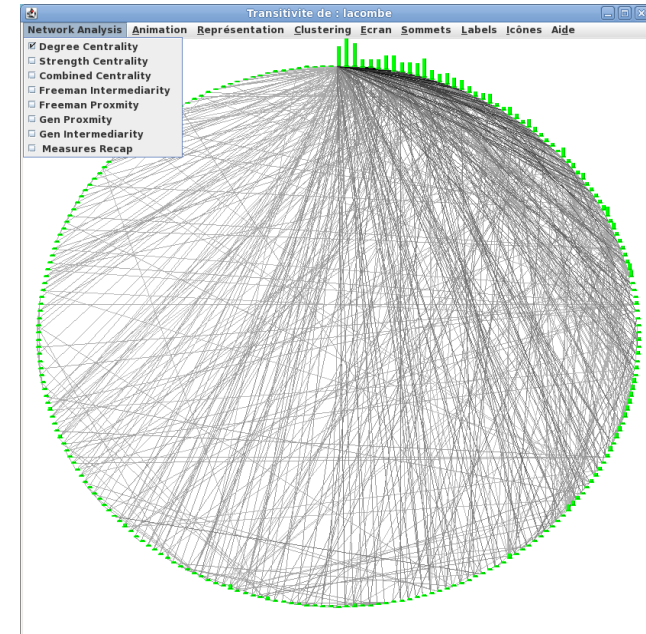


Figure 2: Centralité de degré (graphe non orienté circulaire)

fait pas de distinction entre un seul lien de valeur 2 et deux liens de valeur 1. Cette combinaison est faite en définissant un paramètre alpha, dont la valeur est positive et ne dépend que de la nature du réseau et de son contexte d'étude : un coefficient positif inférieur à 1 donnera plus d'importance au degré binaire, tandis qu'un coefficient supérieur à 1 favorisera le rôle du poids des liens. Cette mesure est définie comme suit :

$$C_D^{W\alpha}(i) = K_i \times \left(\frac{S_i}{K_i}\right)^\alpha \quad \text{où } K_i = C_D(i) = \sum_j^N a_{ij} : \text{degré du nœud } i.$$

Pour les graphes orientés, les auteurs ont défini les mesures suivantes pour qualifier l'activité d'un nœud :

$$C_{D-out}^{W\alpha}(i) = K_{i-out} \times \left(\frac{S_{i-out}}{K_{i-out}}\right)^\alpha \quad C_{D-in}^{W\alpha}(i) = K_{i-in} \times \left(\frac{S_{i-in}}{K_{i-in}}\right)^\alpha$$

La valeur d'alpha, dans ces deux équations, est similaire à celle pour un graphe non orienté.

2.3.2 Centralité de proximité

La centralité de proximité [Freeman 79] est une mesure de centralité globale définie via le calcul de distance séparant un nœud des autres nœuds du graphe. Dans un réseau social, cette mesure exprime la notion d'effort : un acteur ayant une grande valeur de centralité de proximité est capable de contacter la majorité des acteurs du réseau avec un minimum d'effort, c'est-à-dire de distance à parcourir dans le réseau (ou le plus court chemin) en prenant en compte les liens directs et indirects et ceci dans le but de minimiser le nombre d'intermédiaire ce qui permet de s'échapper à leur contrôle relatif à chaque passage. Cette mesure permet de déceler les cas particuliers d'agents qui, même peu connecté au réseau, sont centraux parce que relié à plusieurs agents centraux de la structure.

Soit $G = (V, E)$ un graphe d'ordre N représenté par sa matrice d'adjacence $[a_{ij}]$. Dans le cas où le graphe G est non-orienté, la centralité de proximité d'un nœud $v_i \in V$ est définie par :

$$C^{pro}(v_i) = \frac{N-1}{\sum_{j=1}^N dist(v_i v_j)}$$

Où $dist(v_i v_j)$ est la distance entre les deux sommets v_i et v_j .

Dans le cas où le graphe G est orienté, chaque nœud $v_i \in V$ possède alors deux mesures de centralité de proximité : une par rapport aux liens sortants et une par rapport aux liens entrants.

Pour le calcul des distances entre sommets, plusieurs métriques peuvent être utilisées. Freeman propose par exemple d'utiliser la distance géodésique (i.e. taille du chemin le plus court) entre les nœuds. D'autres mesures de distance telle que la distance euclidienne, quand elle s'y prête, peuvent également être utilisées pour le calcul de la centralité de proximité.

2.3.3 Centralité d'intermédiarité

La centralité d'intermédiarité [Freeman 79] mesure le nombre de fois où un agent est l'intermédiaire sur le plus court chemin entre deux autres, il s'agit donc du degré de médiation d'un agent sur les transferts d'information entre deux autres agents. Cette mesure paraît particulièrement appropriée dans l'étude du pouvoir dans les réseaux de communication parce que l'intermédiation permet la manipulation de l'information transmise. Finalement, la centralité de proximité mesure la manière dont on peut éviter le contrôle d'autres membres de l'organisation là où l'intermédiarité représente au contraire la manière dont on peut accroître la dépendance des autres. La centralité peut être définie comme la probabilité qu'une information transmise entre deux nœuds passe par un nœud donné. [Borgatti and Everett 06].

Soit $G = (V, E)$ un graphe (orienté ou non) d'ordre N . La centralité d'intermédiarité d'un nœud $v_i \in V$ est définie par :

$$C^{int}(v_i) = \sum_{j=1}^N \sum_{k=1}^N \frac{g_{jk}(v_i)}{g_{jk}}$$

Où $g_{jk}(v_i)$ est le nombre total de chemins géodésiques entre les nœuds v_j et v_k qui passent par le nœud v_i , et g_{jk} est le nombre total des plus courts chemins entre les nœuds v_j et v_k . La centralité d'intermédiarité est basée sur l'idée que les nœuds ne communiquent entre eux qu'à travers les chemins les plus courts mais suivant le contexte d'étude on peut avoir intérêt à prendre en compte le fait que les nœuds peuvent interagir en utilisant des chemins autres que les chemins géodésiques. La notion de centralité d'intermédiarité de nœud proposé par Freeman a été généralisée à la centralité d'intermédiarité de lien ou intermédiarité de lien par [Girvan & Newman 02], [Newman & Girvan 04]. Cette mesure est définie comme suit:

$$C^{int}(a_{k,l}) = \sum_{i=1}^N \sum_{\substack{j=1 \\ j < i}}^N \frac{g_{ij}(a_{k,l})}{g_{ij}}$$

Où $g_{ij}(a_{k,l})$ est le nombre de plus courts chemins entre i et j passant pas le lien en considération et g_{ij} est le nombre global de plus courts chemins entre i et j . Une des applications directe de cette mesure d'intermédiarité relative aux liens est la détection des communautés dans les réseaux complexes. Newman et Giravan ont proposé un algorithme qui fonctionne en deux phases. La première phase supprime itérativement des liens du réseau, par ordre d'intermédiarité de lien décroissante, elle est décrite dans [Girvan & Newman 02] puis une deuxième phase qui rétablit un à un les liens du réseau, dans l'ordre inverse de leur suppression. Le but de cette seconde phase est de déterminer la partition qui maximise la modularité du réseau.

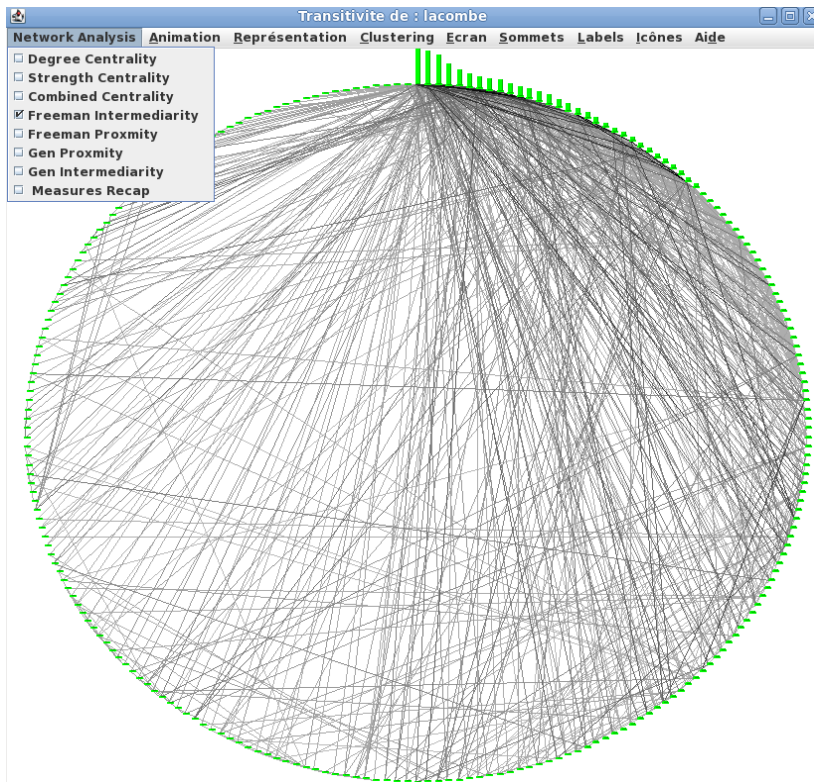


Figure 3: Intermédiarité de Freeman (graphe circulaire)

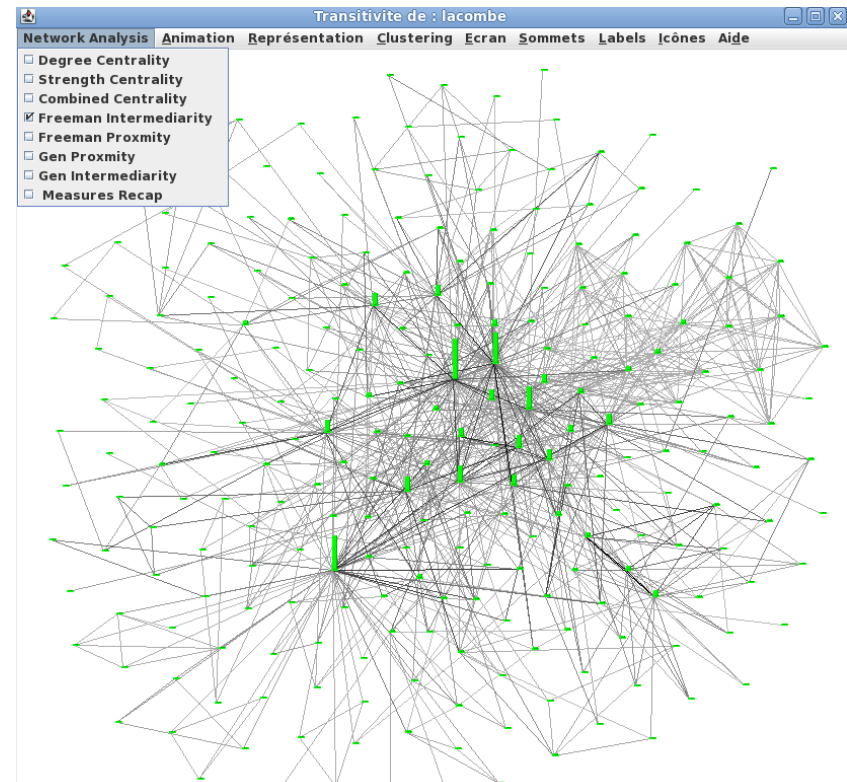


Figure 4: Intermédiarité de Freeman (graphe convergé)

2.3.4 Centralité de proximité et d'intermédiarité généralisée

Le calcul de centralité classique se base sur le plus court chemin en tant que distance calculé sans tenir compte du nombre de nœuds intermédiaires. T. Opsahl a essayé d'établir un compromis entre ces deux paramètres ce qui doit permettre aux chercheurs d'établir l'importance relative entre le nombre de liens et les poids des liaisons. Une extension de l'algorithme de Dijkstra a été prise en compte pour cela :

$$d^{w\alpha}(i,j) = \min \left(\frac{1}{(w_{ih})^\alpha} + \dots + \frac{1}{(w_{hj})^\alpha} \right) \quad C^{pro}(v_i) = \frac{N-1}{\sum_{j=1}^N dist^\alpha(v_i v_j)} \quad C^{int}(v_i) = \sum_{j=1}^N \sum_{k=1}^N \frac{g_{jk}^\alpha(v_i)}{g_{jk}^\alpha}$$

h représente le nombre de nœud intermédiaire entre i et j et en se basant sur cette formule on peut redéfinir la centralité de proximité et d'intermédiarité.

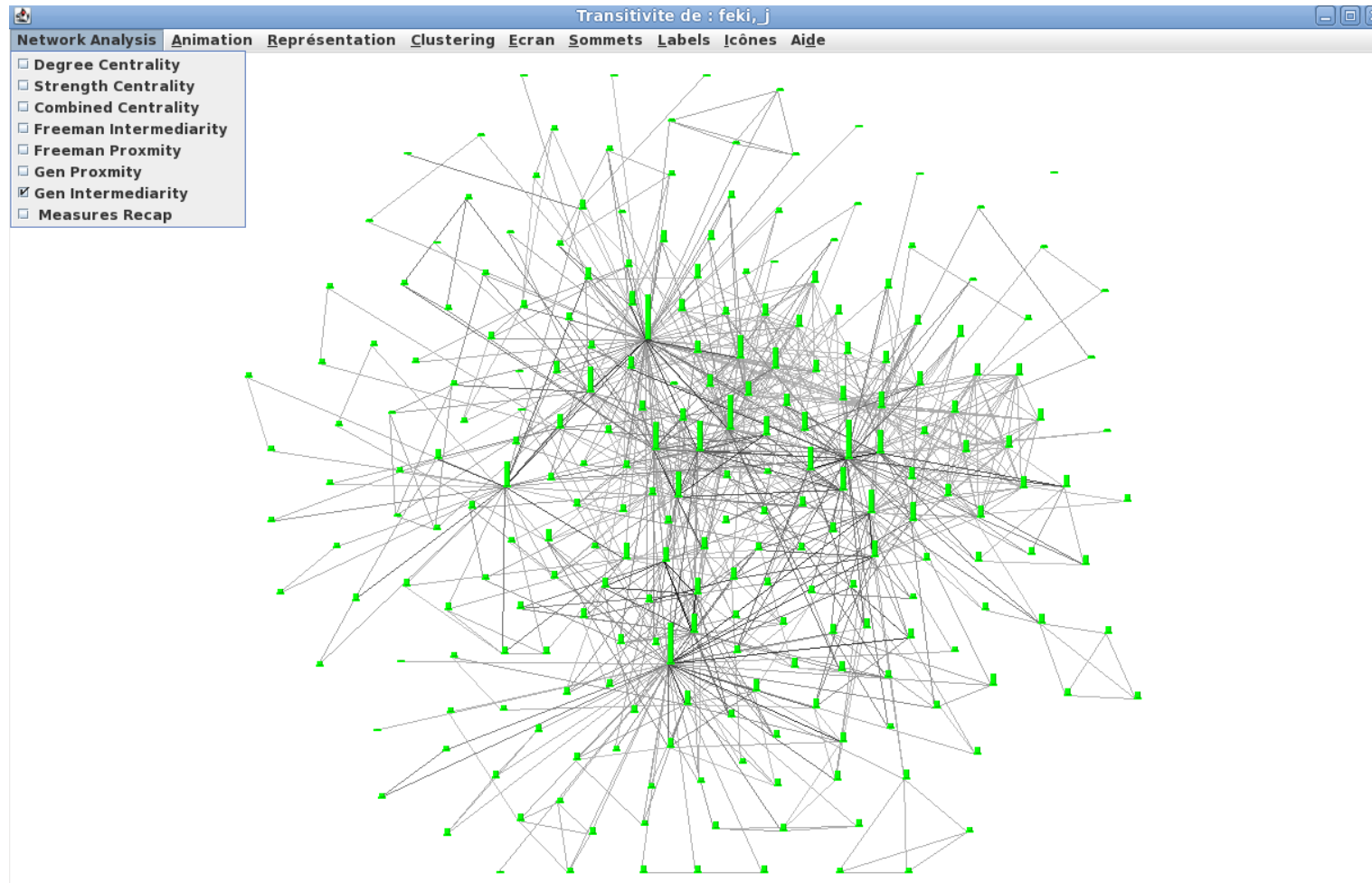


Figure 5: Implémentation de l'intermédiarité généralisée (graphe convergé)

2.3.5 Centralités de proximité et de d'intermédiarité paramétrées

Les auteurs cherchent dans leurs articles « *Node Centrality in weighted networks* » à inclure une notion de coût liée au nombre de nœuds intermédiaires dans le calcul du plus court chemin. On cherche alors à étendre l'algorithme de Dijkstra qui néglige ce coût pour déterminer le plus court chemin entre deux nœuds et ainsi la formulation de Dijkstra du calcul de distance ne se base que sur les pondérations :

$$d^w(i, j) = \min((\frac{1}{w_{ih}}) + \dots + (\frac{1}{w_{hj}}))$$

Il peut être reformulé comme suit pour tenir compte de ces deux coûts:

$$d^w(i, j) = \min((\frac{1}{(w_{ih})^\alpha}) + \dots + (\frac{1}{(w_{hj})^\alpha}))$$

Où le paramètre α est un paramètre de valeur positive.

En tenant compte de cet algorithme de calcul du plus court chemin, les mesures de centralité d'intermédiarité et de proximité peuvent être reformulées respectivement comme suit :

$$C_c^{w\alpha}(i) = \left[\sum_j^N d^{w\alpha}(i, j) \right]^{-1} \quad C^{w\alpha}_{int}(v_i) = \sum_{j=1}^N \sum_{k=1}^N \frac{g_{jk}^{w\alpha}(v_i)}{g_{jk}^{w\alpha}}$$

2.3.6 Centralité spectrale (Eigenvector Centrality)

La mesure proposée par [Bonacich 07] considère que la centralité d'un nœud est déterminée par la centralité des nœuds auxquels il est directement connecté. Cette mesure donne autant d'importance à l'acteur qu'il entretient des liens avec des autres acteurs du réseau qui sont eux même importants. La qualification d'importance d'un acteur est indiquée par son poids associé. Cette mesure est définie comme combinaison linéaire des centralités des nœuds voisins du nœud en considération.

Soit $G = (V, E)$ un graphe d'ordre N représenté par sa matrice d'adjacence A . Dans le cas où le graphe G est non-orienté, la centralité spectrale $C^{sp}(v_i)$ d'un nœud $v_i \in V$ est donnée par l'équation :

$$\mu C^{sp}(v_i) = a_{1i} C^{sp}(v_1) + a_{2i} C^{sp}(v_2) + \dots + a_{ni} C^{sp}(v_n)$$

Dans le cas où le graphe G est orienté, chaque nœud $v_i \in V$ possède alors deux déclinaisons de centralité spectrale : une par rapport aux liens sortants et une par rapport aux liens entrants. Elles sont données respectivement par les équations suivantes:

$$\mu C_{out}^{sp}(v_i) = a_{1i} C_{out}^{sp}(v_1) + a_{2i} C_{out}^{sp}(v_2) + \dots + a_{ni} C_{out}^{sp}(v_n) \quad \theta C_{in}^{sp}(v_i) = a_{1i} C_{in}^{sp}(v_1) + a_{2i} C_{in}^{sp}(v_2) + \dots + a_{ni} C_{in}^{sp}(v_n)$$

Où μ et θ sont des réels strictement positifs. En termes de calcul matriciel, la centralité spectrale s'écrit :

$$\mu C^{sp} = M^T C^{sp}$$

où μ est un réel strictement positif, C^{sp} est le vecteur de centralité spectrale. M correspond soit à la matrice d'adjacence dans le cas du calcul de la centralité entrante, soit à la transposée de cette matrice dans le cas du calcul de la centralité sortante (dans le cas où le graphe est non-orienté, nous avons $M = M^T$). Bonacich montre que le vecteur de

centralité spectrale C^{spe} correspond en fait au vecteur propre dominant (ou principal) de la matrice M . Pour le calcul du vecteur C^{spe} , il est possible d'utiliser, par exemple, la méthode des puissances itérées qui est particulièrement efficace lorsque la matrice M est creuse.

2.3.7 Centralisation :

[Freeman 79] définit également la centralisation qui est une mesure qui caractérise le réseau dans sa globalité. Cette valeur donne une idée sur la distribution des valeurs de centralité. Cette mesure est définie comme la somme des différences entre le nœud le plus central et tous les autres nœuds, divisé par le maximum possible de cette somme des différences comme l'indique sa formule:

$$C_X = \frac{\sum_{u=1}^{u=|N|} C_X(u^*) - C_X(u)}{\max(\sum_{u=1}^{u=|N|} C_X(u^*) - C_X(u))}$$

Dans l'équation ci-dessus, $C_X(u^*)$ est la valeur maximale de centralité.

2.4 Etude comparative des mesures de centralités

Les mesures locales au sein d'un réseau servent à définir la situation d'un élément via ses voisins directs, la communauté auquel il appartient et par rapport au réseau global. Paul Chapron distingue deux types d'indicateurs qui définissent deux manières d'analyser les réseaux, la première contextuelle et la deuxième structurelle :

- **Indicateurs situés** : Ce sont des indicateurs relatifs à une valeur particulière du vecteur d'état de l'organisation. Donc, un réseau étiqueté avec de telles mesures sera considéré comme réseau contextuel qui représentera un mode de fonctionnement particulier de l'organisation.
- **Indicateurs structurels** : Ce sont des mesures indépendants du vecteur d'état ce qui permettra d'avoir une vision plus général du réseau détaché de la valeur du vecteur d'état de l'organisation.

Les mesures structurelles peuvent être classées en mesures locales et mesures globales. Les mesures locales sont décomposées en deux sous catégories : mesures locales de voisinage et mesures locales d'ensembles.

- **Mesures locales de voisinages** : Le degré de voisinage peut être calculé en fonction du degré des voisins, de la pondération des liens ou bien d'une combinaison des deux : topologie et poids (Opsahl). La centralité spectrale se différencie du degré par le fait qu'elle tient compte du degré des sommets adjacents. Cette centralité tient compte à la fois du nombre et de la qualité du lien.
- **Mesures locales d'ensembles** : Ce genre de mesure se base essentiellement sur le calcul des distances géodésiques, ce qui donne une idée sur l'accessibilité prise sous différents angles tel que la centralité d'intermédiation dans ses deux versions topologiques ou pondérés et la centralité de proximité. Ces mesures permettent de détecter les ponts qui interconnectent les différents modules d'un graphe.

Les mesures globales sont les études de distribution statistique des mesures locales telles que la distribution de degré, croisement des centralités d'intermédiation et proximité. En effet, la distribution de degré des sommets permet de détecter les structures invariantes d'échelle. Une autre solution utilisée par les chercheurs est de croiser deux mesures de centralités pour mettre en œuvre les sommets qui sont mieux placés aussi au niveau global qu'au niveau local. Le croisement de la centralité de degré avec celle d'intermédiation permet de mettre en évidence les connecteurs, bridges et c'est sur cette idée que sont basées quelques mesures de centralités dans des recherches récentes telles que « Bridging centrality ». Un autre croisement pratique, et très utilisé dans l'analyse des réseaux complexe, est celui de degré simple avec degré pondéré.

2.5 Etude de la complexité dans les mesures de centralité

Les mesures de centralités ont été typiquement étudié dans les graphes de taille réduit, mais avec la montée en puissance des réseaux sociaux, du commerce électronique, de l'évolution des structures de données et des capacités de stockages informatiques nous devons faire face aujourd'hui à des graphes qui peuvent atteindre des millions de

nœuds et de liens et l'exemple le plus parlant est celui des réseaux sociaux : Facebook, Flickr, ... Les algorithmes de calcul de centralité dans ce genre de réseau posent plusieurs défis tels que le temps de calcul, l'espace mémoire, le parallélisme, sans compter la complexité de certaines de ces mesures.

Bien que le calcul de la centralité degré, qui a une complexité de $O(N)$, reste simple à exécuter, le calcul de la centralité d'intermédiation basé sur le plus court chemin ou sur la marche aléatoire (en anglais « Random Walk ») requiert au moins $O(N^3)$ temps de calcul et $O(N^2)$ comme espace mémoire où N est le nombre de nœud du graphe. La solution de Ulrik Brandes ne requiert que $O(N + M)$ d'espace mémoire et $O(NM)$ et $O(NM + n^2 \log N)$ de temps de calcul pour les graphes binaires et les graphes pondérés respectivement où M est le nombre de lien du graphe. Cet algorithme a été utilisé par plusieurs plateformes d'analyse de données des réseaux sociaux tels que Gephi et Ucinet. Selon [Kiss & Bichler 08], la complexité liée au calcul de la centralité spectrale est quant-à elle de l'ordre de $O(N^2)$.

Nous venons de présenter une synthèse des différents axes de recherches actuels dans le domaine de l'analyse structurelle des réseaux. Ce genre d'analyse est devenu primordial de nos jours pour faire émerger les éléments de compréhension concernant le fonctionnement, les forces et les faiblesses de ces réseaux, ce qui permettra d'établir le lien entre plusieurs phénomènes d'ordre global et local.

3 Visualisation de données dans un contexte statique et Temporel

3.1 Introduction

La plateforme de visualisation de graphe VisuGraph développée au sein de l'équipe EVI de l'IRIT est dédiée à l'exploration de grands graphes. Elle autorise un utilisateur expert à visualiser un graphe à partir d'algorithmes de dessin parmi les plus récents. Ces algorithmes facilitent le rendu visuel des graphes. Conçue pour la manipulation et le calcul du clustering de grands graphes, l'interface utilisateur permet de fouiller dans l'information ainsi décomposée en inspectant chaque cluster séparément.

De fait, VisuGraph se prête à la fouille interactive de données munies de relations pondérées et enrichies d'attributs, numériques ou textuels (graphes étiquetés). L'analyse d'un graphe à l'aide d'indices et de mesures structurels est un exemple typique où l'exploration interactive vient en appui du travail de fouille de données et d'extraction de connaissances.

Si nous prenons par exemple un jeu de données qui décrit les travaux des chercheurs, l'objectif de notre travail est de savoir comment se distribue le pouvoir entre les acteurs dans un tel réseau c'est-à-dire : qui contrôle une communauté donnée, un réseau? Qui sert ou détient une information pertinente et qui contrôle un flux de données? Comment évolue un acteur ? Comment et Quand s'est fait une alliance, une fusion ou une rupture ?, ... Nous allons chercher à résoudre ce genre de question dans un contexte statique et temporel et présenter le résultat via des techniques qui faciliteront la compréhension et l'interprétation des résultats obtenus.

3.2 Visualisation des données statiques :

La visualisation des données relationnelles se base sur la notion de graphe qui est une représentation mathématique définie par deux ensembles : les sommets et les arêtes. Dans ce qui suit, nous allons noter par V l'ensemble des nœuds du réseau et par A l'ensemble des arêtes. Nous définirons ainsi un graphe $G = (V, A)$ avec $|V| = N$ le nombre de sommets.

La représentation des graphes statiques repose sur la définition de l'espace de présentation. En effet, la définition des éléments qui forme la composante graphique de cet espace revient à définir les variables visuelles à utiliser qui décrivent la forme et l'aspect visuelles des données. L'espace de visualisation est un plan et chaque sommet est défini par ses coordonnées x et y qui correspondent à l'abscisse et l'ordonnée de ce sommet. Nous avons aussi fait des essais en 3D.

Pour satisfaire les besoins de l'utilisateur et faciliter la découverte, la navigation et l'exploration de la structure de données, nous allons définir un ensemble de paramètres et de métriques pour garantir une interface riche et compréhensible par l'utilisateur qui peut être expert ou novice.

3.2.1 Définition de l'espace de visualisation des données :

L'espace de représentation des données regroupe par définition les données du corpus et les éléments graphiques par lesquels on veut les représenter : la relation entre ces données et les éléments graphiques est une relation associative. On se base sur le concept du nœud métrique : c'est-à-dire une quantification de l'information qui passe par les arêtes et les nœuds.

Soit f une fonction qui associe à chaque sommet v_i une valeur m_i qui désigne l'information détenue par ce sommet.

$$f(v_i) = m_i$$

Et g la fonction qui associe à chaque arête a_i une valeur k_i qui désigne l'information qu'elle détient.

$$g(a_i) = k_i$$

La valeur m_i correspond, dans le cas des matrices symétriques, au croisement de v_i avec lui-même. Cette métrique adoptée dans des travaux précédents [Loubier E. 09] était dans VisuGraph la seule à pouvoir identifier l'importance d'un sommet dans le graphe. Dans notre approche, nous allons partir du modèle théorique de représentation de ces données relationnelles en nous basant sur les propriétés structurelles telles que les mesures de centralités sur les sommets mais aussi les relations qui détiennent la plus grande quantité d'information ou de flux de donnée. En prenant comme exemple le cas d'une matrice symétrique croisant les auteurs, on ne peut pas définir le statut d'un acteur et juger de son importance simplement par le nombre de ses publications, vu que le pouvoir et l'influence de cet auteur peuvent être liés au type de ses publications, à ses coopérations nationales et internationales, ... Ceci doit faire l'objet de mesures et d'indicateurs spécifiques et structurels provenant de l'analyse fine du graphe.

D'autre part, le modèle adopté par VisuGraph pour identifier les liens les plus influents du réseau se limite à la fonction g qui n'exploite pas la pondération des liens et ne donne pas ainsi une idée sur l'importance du lien au niveau local et global.

3.2.2 Représentation graphique des métriques

Pour visualiser d'une façon rigoureuse les données, sans encombrement ou perte qui peuvent affecter le jugement de l'utilisateur, on doit procéder à la normalisation des paramètres de visualisation. La normalisation des variables visuelles permet d'établir une comparaison entre des données hétérogènes. Ceci permet, par exemple, la comparaison de réseaux de tailles différentes en divisant ces mesures par un facteur choisi. Ce facteur est généralement la valeur maximale pour chaque paramètre utilisé.

$$\mu_i = \frac{m_i}{m_{max}}$$

Ainsi on peut décrire cela par une fonction

Avec μ_i la mesure normalisée pour un paramètre de visualisation donné et m_i la valeur de ce paramètre pour un sommet ou une arête et m_{max} sa valeur maximale.

Nous allons procéder à l'encodage visuel des valeurs normalisés en utilisant les variables de Bertin définies précédemment.

- La forme : dans le but de simplifier la représentation graphique nous allons utiliser une seule forme de représentation pour réduire l'effort cognitif mobilisé par l'utilisateur pour détecter les signes. L'utilisation d'une seule forme à chaque fois : barre, ou cercle facilite l'interprétation du graphe.
- La taille : La variation de taille relativement à une métrique que nous voulons mettre en évidence permet une comparaison simple et aisée entre les différents sommets du graphe. Soit $D(v_i)$ la fonction qui calcule le diamètre du cercle relativement à la valeur du paramètre en considération :

$$D(v_i) = \sqrt{\mu_i}$$

- La variation de valeur : l'utilisation de nuance de couleur permet de mettre en évidence les différents niveaux d'importance des sommets. Il existe une vingtaine de niveau de couleur et ce nombre correspond au nombre de dégradation de couleur détectable par l'œil humain. L'utilisateur peut sélectionner la distribution de couleur qui lui donne un meilleur accès aux informations recherchées. Cette distribution de couleur suit la fonction suivante :

$$f_n(v_i) = \frac{(n+1).x}{nx+1}$$

Avec x : valeur de la métrique et n représente un amplificateur qui sera utile si la valeur d'une métrique sélectionnée est faible. Une fonction similaire est utilisée pour la distribution des couleurs pour les arêtes dans le cas de métrique lié aux arêtes.

La distribution non linéaire de couleurs et le fait de proposer à l'utilisateur de choisir le mode visuel qui lui convient, ont permis une meilleure interactivité car celui-ci peut se concentrer sur les éléments les plus remarquables ou, au contraire, faire ressortir des détails s'ils sont trop écrasés par la présence de quelques point forts.

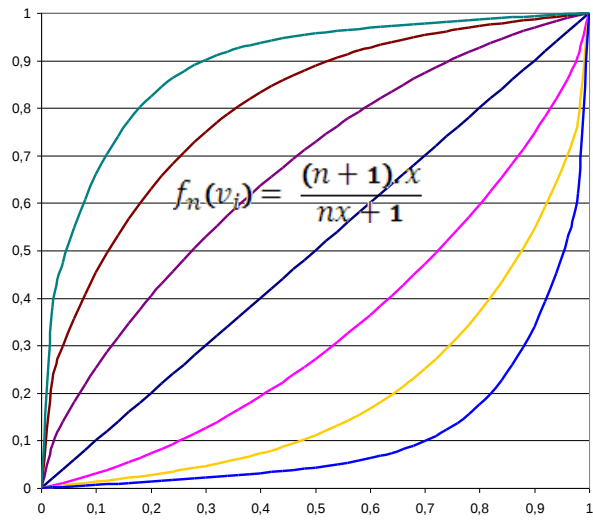


Figure 6: Codage non linéaire des couleurs

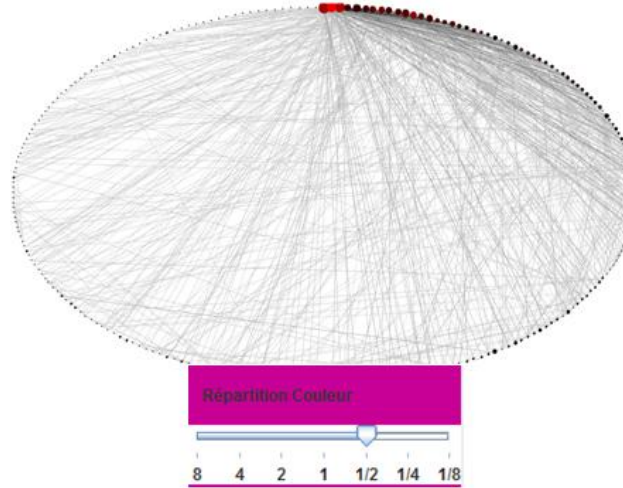


Figure 7: Scénario pour la distribution des couleurs

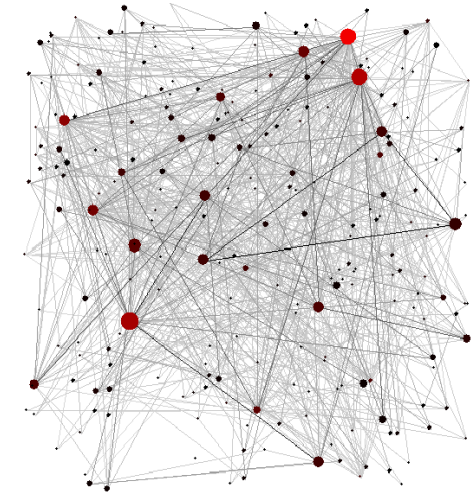


Figure 8: Sur un graphe aléatoire

3.2.3 Processus d'identification des Sommets et Liens les plus influents

L'interface graphique de l'outil VisuGraph comporte trois fenêtres dans lesquelles sont présentées les différentes fonctionnalités disponibles. En effet, pour garantir un confort visuel et réduire la charge cognitive, nous proposons un paramétrage de l'affichage : couleur de l'écran, modification de la forme et de la couleur des sommets, des liens et de la police utilisée pour la nomenclature des sommets. Pour analyser d'une façon rigoureuse les données du graphe, nous avons intégré plusieurs fonctionnalités qui seront utiles au partitionnement du graphe et à la recherche en profondeur et en largeur des sommets et liens les plus influents. Dans ce qui suit, nous allons décrire les fonctionnalités disponibles dans l'analyse et la visualisation de données dans un contexte statique.

Les mesures de centralités employées dans cet outil ont été choisies en se basant sur les études théoriques et expérimentales menés par plusieurs chercheurs sur des jeux de données pour mettre en évidence les corrélations qui peuvent exister entre ces mesures [Landher, 2010]. Un agent sera central dans plusieurs cas : s'il possède un nombre important de liens, alors il a une forte centralité de degré ; s'il peut atteindre facilement les autres agents du graphe, alors il a une grande centralité de proximité ; s'il est nécessaire à la communication au sein du graphe, alors il a une intermédiarité forte ; s'il est lié ou proche d'agents qui sont eux-mêmes centraux, alors sa centralité par les valeurs propres est importante. Pour chaque indicateur, il est possible de définir trois mesures complémentaires : l'indicateur brut qui donne la position d'un agent dans le graphe, l'indicateur relatif qui permet de prendre en compte la taille du graphe et l'indicateur général qui concerne tout le graphe et qui permet de comparer celui-ci à un graphe en forme d'étoile.

La centralité de degré qui compte le nombre de liens que possède un acteur. Cette mesure est facile à calculer, directement interprétable mais n'exprime que la présence ou l'absence d'un lien sans tenir compte de sa pondération et c'est pour cette raison qu'on a défini la pondération du lien comme mesure qui met en œuvre l'implication d'un agent dans le réseau et la définition du degré généralisé qui combine les deux approches.

La centralité de proximité prend en compte la position relative de l'agent par rapport à l'ensemble des autres agents du réseau. Son inconvénient majeur est qu'elle n'utilise qu'un seul chemin pour mesurer la distance entre deux agents : le chemin géodésique ce qui ne présente pas vraiment une grande utilité dans le monde réel et pour palier à cet inconvénient on a introduit une mesure généralisée de proximité.

La centralité d'intermédiarité utilise un concept de centralité radicalement différent qui accorde peu d'importance au nombre de liens. Cet indicateur permet de localiser les agents nécessaires à la connexion du réseau. Inconvénients : l'indicateur est calculé uniquement sur les chemins géodésiques.

La centralité par les valeurs propres prend en compte la centralité de tous les autres agents pour calculer la centralité d'un agent. C'est une mesure précise qui permet de discriminer les agents. Son inconvénient est qu'elle est difficilement calculable quand le nombre d'agent est trop important et lorsque le graphe n'est pas connexe, seule la centralité des agents appartenant à la plus grande composante est calculée.

Pour chercher des caractéristiques particulières et qui peuvent être liées au contexte de l'étude, nous avons mis en œuvre un mécanisme de classement ordonné pour chaque mesure. Ainsi, la sélection d'une mesure de centralité nous affichera sous un format choisit les nœuds du réseau. Puis nous cumulon sur ces résultats affichés au premier plan, les résultats d'un choix qui permet de n'afficher que les n premières sommets pour un autre paramètre. L'ordre des mesures est décroissant le premier classé est celui qui détient la valeur maximale pour chaque mesure. Ainsi un sommet affiché après une certaine configuration doit être certainement parmi les n les plus grands pour chaque mesure.

Pour chercher à exploiter au maximum ces mesures, nous pouvons combiner avec le filtrage visuel, qu'on a adopté pour mettre en évidence ces métriques visuelles, quelques paramètres analytiques qui serviront effectuer une analyse prédictive. Nous analysons les scénarios possibles en cas d'établissement d'un nouveau lien, l'élimination d'un lien existant ou la suppression d'un acteur du réseau et la détection du résultat peut se faire en regardant l'évolution de certaines propriétés structurelles du réseau.

Ces paramètres analytiques sont :

- Diamètre de graphe : mesure de la plus grande distance géodésique qui existe dans le graphe.
- Densité du graphe : rapport du nombre d'arêtes dans le graphe sur le nombre maximum qui peut exister.
- Nombre de nœud isolés
- Le nombre de composante connexe.

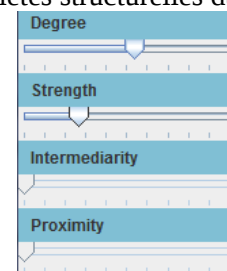


Figure 9: Filtre configurable sur les mesures de centralités

3.3 Visualisation des données temporelles

Dans cette partie, nous nous concentrons sur le thème de la visualisation des données temporelles et nous mettons l'accent sur notre solution pour la visualisation interactive pour des données relationnelles évolutives.

3.3.1 Définition de l'espace de visualisation des données temporelles

Contrairement au modèle statique qui se base sur un seul matrice enregistré dans un seul fichier, dans la visualisation des données temporelles les données en entrée sont sous forme de plusieurs matrices et chacune est enregistrée sur un fichier différent.

Nous allons adopter la même définition d'espace de présentation avec ajout de variables temporelles :

- R : nombre d'instances, périodes en étude.
- m_g est la métrique globale associée au sommet v_i , elle est définie par :

$$m_g = \sum_{t=0}^R m_t$$

Avec $t \in \{1, \dots, R\}$ et m_t la métrique associée à un sommet v_i pour une période donnée.

Si on considère les paramètres de Bertin, alors on va utiliser, pour représenter les nœuds, des histogrammes composés de barres de couleurs et tailles différentes ce qui permet aisément la comparaison des valeurs de chaque métrique en considération pour chaque période. Cette forme de représentation facilite la détection des tendances et l'évolution

des mesures. Pour la coloration des barres verticales, on va utiliser le rouge pour marquer la première période et le vert pour marquer la dernière, le reste des périodes sera représenté par des dégradés entre ces deux couleurs.

Contrairement à la représentation statique, l'emplacement des nœuds doit être justifié : Pour cela, nous allons définir des nœuds repères temporels placés sur un cercle près des bords de la fenêtre. L'emplacement des nœuds du graphe sera alors régi par les forces d'attraction de ces nœuds repères en fonction de la valeur de la métrique au cours d'une période t [Loubier E. 2009]. Pour ne pas augmenter la charge cognitive on ne va pas masquer ces nœuds temporels une fois qu'a été sélectionnée une des mesures de centralités.

Le calcul des mesures de centralités se fait par période et sur chaque fichier indépendamment et la représentation se fait sur une seule interface tout en appliquant l'algorithme de placement des nœuds.

3.3.2 Analyse de la structure temporelle :

L'analyse structurelle temporelle de ce genre de données se fait en fonction de l'emplacement des nœuds par apport aux repères temporels. Ainsi, les données qui ne concernent qu'une seule période sont placées à coté du repère qui lui correspond. Si les données concernent deux périodes alors elles sont placées de façon équidistante aux deux périodes considérés et si notre acteur se présente sur les trois périodes alors il est placé au milieu du cercle passant par tous les nœuds repères. Cette représentation évolutive des données permet de faire le bilan sur l'apparition, disparition de thèmes, les fusions et les alliances entre les acteurs du marché aussi bien que la détection des nouveaux acteurs.

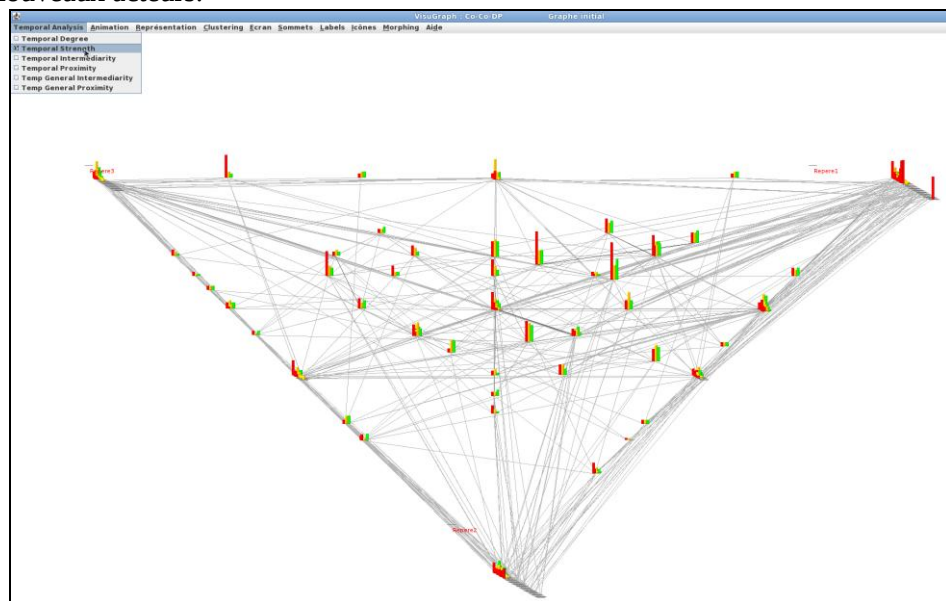


Figure 10: Evolution du degré pondéré sur 3 périodes

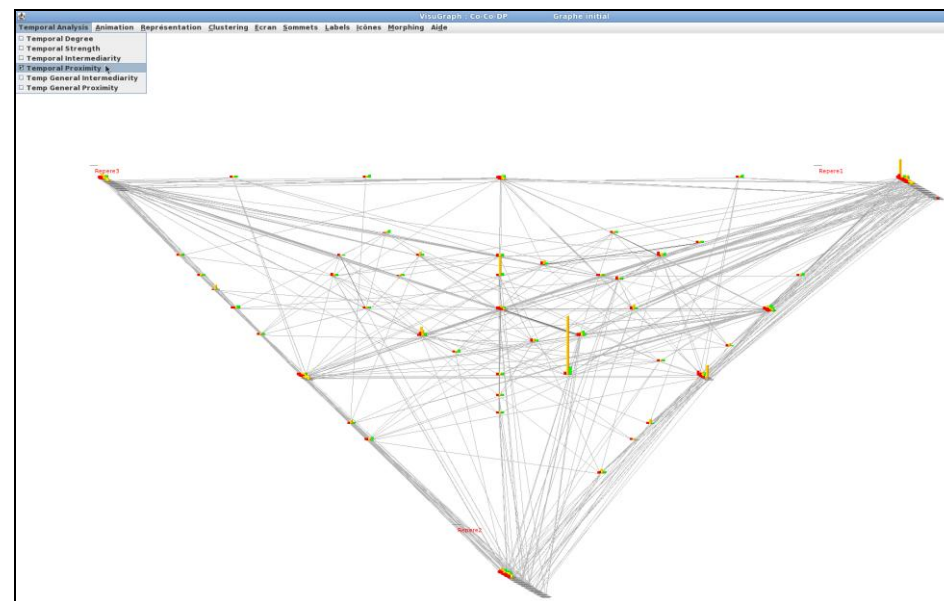


Figure 11: Evolution de la proximité sur 3 périodes

Cette représentation évolutive des mesures couplée avec le morphing de graphe permet la fouille de grands volumes de données par période ce qui permet de connaître la cause d'une modification de positionnement d'un acteur. Par exemples, la formation d'une nouvelle alliance, la perte de lien ou l'apparition d'un nouveau concurrent. Les valeurs des liens, qui quantifient les forces des relations, augmentent ou diminuent et par suite nous pouvons constater des changements dans l'emplacement des acteurs dans l'espace 2D de présentation.

La visualisation des données dans le contexte statiques et temporel est une solution pour l'analyse et la recherche sur des motifs, signes et indices pour déterminer des tendances. Le modèle de graphe montre bien son efficacité dans la recherche de sommets qui détiennent l'information. L'intégration des mesures de centralités au sein de VisuGraph nous a permis d'en améliorer la production interactive de connaissances.

4 Conclusion

La visualisation de connaissance est un domaine de recherche extrêmement récent. Les applications de la visualisation de l'information facilitent la découverte et le transfert de connaissance, l'apprentissage, l'enrichissement ou la vérification des connaissances visualisées. Notre objectif initial était de mettre au point des approches graphiques adaptées à la visualisation des connaissances sur les données relationnelles, permettant de découvrir plus rapidement les connaissances cachées dans de grands volumes de données.

Le panel d'outils intégré dans VisuGraph est riche par les techniques d'analyse et les algorithmes de visualisation de données qu'il propose tels que les algorithmes de placements paramétrables des sommets par des forces qui permettent d'obtenir des graphes planaires, clairs et exploitables. Bien qu'il existe beaucoup de logiciel qui effectuent du dessin de graphes et de l'analyse de données dans un contexte statique, VisuGraph restera parmi la minorité de ces plateformes d'analyses qui traitent aussi les données évolutives. L'objectif de l'intégration des mesures de centralités dans VisuGraph en statique et en dynamique était de faciliter d'extraction de signes structurels pertinents avec lesquelles il est possible de construire un plan d'action fiable.

La performance et le temps de calcul qui augmentent avec la charge des données en entrée dans VisuGraph restent un problème à régler en cherchant une solution matérielles ou par l'application de calculs distribués qui peuvent donner des résultats plus rapides. Nous proposons d'exploiter le progrès des technologies 3D pour planifier le passage en 3D de tout le module de visualisation pour améliorer la qualité de visualisation et faciliter encore l'interaction.

5 Bibliographie

- [1] **[Loubier E. 09]** *Analyse et visualisation de données relationnelles par morphing de graphe en prenant en compte la dimension temporelle.*
- [2] **[Freeman L. C. 79]** *Centrality in social networks conceptual clarification.* Social Networks, 1(3):215-239, 1979.
- [3] **[Wasserman & Faust 94]** *Social Network Analysis: Methods and Applications.* Cambridge University Press, 1994.
- [4] **[Scott J. P. 00]** *Social Network Analysis: A Handbook.* Sage Publications Ltd, 2000.
- [5] **[Kleinberg 99]** *Authoritative sources in a hyperlinked environment.* J. ACM, 46(5): 604-632, 1999.
- [6] **[Bertin 83]** *Semiology of Graphics,* University of Wisconsin Press.
- [7] **[Bonacich P. 87]** *Power and Centrality.* The American Journal of Sociology, Vol.92, No.5 1170-1182, 1987.
- [8] **[Bonacich P. 07]** *Some unique properties of eigenvector centrality.* Social Networks, 29(4):555-564, 2007.
- [9] **[Brandes U. 01]** *A Faster Algorithm for Betweenness Centrality,* Journal of Mathematical Sociology 25(2) pp.163-177, 2001.
- [10] **[Watts D. J. & Strogatz S. 98]** *Collective dynamics of 'small-world' networks.* Nature, 393:440-442.
- [11] **[Granovetter 73]** *The Strength of Weak Ties.* American Journal of Sociology 78(6): 1360- 1380.
- [12] **[Opsahl 08]** *Prominence and control: The weighted rich-club effect.* Physical Review Letters 101 (168702).
- [13] **[Bolland J. M. 88]** *Sorting out centrality: An analysis of the performance of four centrality models in real and simulated networks.* Social Networks 10(3):233-253

- [14][**Borgatti S. P. & Everett M. G. 06**] *A graph theoretic perspective on centrality*. Social Networks 28(4):466–484
- [15][**Newman M. E. J. 04**] *Fast algorithm for detecting community structure in networks*. Physical Review E, 69(6):066133, 2004.
- [16][**Chapron P. 10**] César Ducruet, CNRS, UMR Géographie-cités, *Les mesures locales d'un réseau*.
- [17][**Landher 10**] Business & Information Systems Engineering 2 (2010) 6, p. 371-385
- [18][**Kiss C. & Bichler M. 08**] *Identification of influencers – measuring influence in customer networks*. Decis Support Syst 46(1):233–253