
VisuGraph : Un outil d'exploration de données relationnelles évolutives

Eloïse Loubier, Sabine Carbonnel

Institut de Recherche en Informatique de Toulouse, IRIT-SIG
Université Paul Sabatier,
118 route de Narbonne,
31062 Toulouse cedex 9
[loubier, carbonnel](mailto:{loubier, carbonnel}@irit.fr)@irit.fr

RÉSUMÉ. La visualisation par dessin de graphe est un outil puissant pour analyser l'évolution des informations relationnelles. Elle permet, par la synthèse de grands corpus de données, d'identifier et d'évaluer les structures, entre acteurs et concepts, passées ou actuelles et d'en déduire les futures. Dans ce contexte, l'équipe SIG de l'IRIT propose la plate-forme de veille stratégique Tétralogie orientée analyse de données et le portail Web Xplor dédié à la navigation en ligne au cœur de l'information relationnelle. Le module VisuGraph ajoute à ces deux outils la visualisation et la classification interactive des données relationnelles. Afin de mieux décrypter l'aspect stratégique déduit de l'évolution des différents réseaux, nous avons doté VisuGraph d'une animation fluide des représentations successives des graphes des différentes périodes, en faisant ressortir visuellement éléments majeurs et tendances significatives. Nous développons cette approche en insistant sur la structure des données, l'ergonomie logicielle, l'optimisation du dessin de graphe et de son animation.

ABSTRACT. Visualization based on graph drawing is a powerful tool to analyze relational information evolution. It is based on big data corpora synthesis. It makes it possible to identify and to evaluate passed and current structures, and to deduce the future ones between actors and concepts. In this context, IRIT SIG team propose the strategic data analyzes platform Tétralogie and the gate Web Xplor dedicated to on line relational information navigation. VisuGraph tool adds to these two tools visualization and interactive relational data classification. We add to VisuGraph a fluid animation of graph representations of various periods, while emphasizing visually major elements and significant tendencies. We develop this approach while insisting on data structure, software ergonomics, graph drawing optimization and its animation.

MOTS-CLÉS : Dessin de graphe, analyse relationnelle, réseau d'acteur, réseau sémantique, animation, alliance, graphe évolutif, analogie espace/temps.

KEYWORDS: Graph drawing, relational analysis, actor network, semantic network, animation, alliance, evolutionary graph, space/time analogy.

1. Introduction

« L'extraction de connaissances, à partir des données est un processus non trivial d'identification de structures inconnues, valides et potentiellement exploitables dans les bases de données », (Fayyad, 1996). La représentation graphique permet de compléter l'extraction de connaissance en visualisant une grande quantité d'informations de façon compréhensible et en fournissant au lecteur un maximum de renseignements synthétiques, qui ne sont que très rarement explicités dans les données brutes. Les topologies remarquables sont identifiées, révélant les relations entre les différents acteurs (auteurs, laboratoires, entreprises, pays) et les termes ou les concepts d'un domaine. De plus, l'étude de la structure d'un réseau dans le temps (collaborations, co-citations, co-signatures, cooccurrences, alliances, fusions, acquisitions, licences, ...) permet d'évaluer ses organisations passées et actuelles et surtout d'en déduire ses organisations futures et leurs implications en terme de décision.

Dans ce contexte d'analyse de l'information stratégique, l'équipe SIG de l'IRIT propose deux outils puissants de visualisation des données relationnelles : la plate-forme Tétralogie et le portail Web Xplor.

Tétralogie permet (à distance et à plusieurs utilisateurs) d'effectuer des analyses stratégiques à partir de données textuelles hétérogènes, par le biais de méthodes d'analyse de données classiques ou innovantes. Xplor est un système de navigation en ligne au cœur de l'information relationnelle obtenue par le croisement statistique des différentes variables associées aux données thématiques étudiées. Ces deux systèmes se complètent, puisque Tétralogie est dédiée aux macros analyses qui synthétisent au mieux l'ensemble des informations déduites, alors qu'Xplor permet aux experts du domaine de réaliser une navigation ciblée sur leurs objectifs spécifiques : identification, connaissance et suivi des acteurs et des thèmes porteurs constituant leur environnement. Le module VisuGraph y ajoute la visualisation et la classification interactive des données relationnelles par le biais de graphes.

Comme les analyses statiques, sur une longue période, et leur représentation graphique peuvent être à l'origine de graves erreurs d'interprétation surtout dans le cas où elles visualisent des phénomènes non ou peu cumulatifs, il nous paraît nécessaire de proposer des analyses beaucoup plus fidèles et rigoureuses.

Pour cela, nous développons de nouvelles fonctionnalités pour VisuGraph afin de visualiser l'évolution des différents réseaux, d'analyser, d'un point de vue stratégique, la dynamique de leurs relations et de révéler les tendances significatives. Pour ce faire, nous avons conçu un algorithme, qui par animation du dessin entre différentes périodes, révèle les structures successives, les changements significatifs et les acteurs ou concepts déterminants. En se basant sur l'analogie espace/temps utilisée pour une horloge, l'objectif est d'obtenir une lecture intuitive de l'évolution en répartissant séquentiellement les périodes sur un cadran. Le placement stratégique des sommets permet alors, non seulement, de les situer dans le temps mais aussi d'en évaluer la persistance et d'en déduire la tendance. Nous proposons donc, dans cet article, de développer cette fonctionnalité en expliquant son principe et sa mise œuvre.

Dans un premier temps, nous synthétisons un état de l'art de la représentation graphique de données (section 2), puis nous présentons la plateforme Tétralogie et le module VisuGraph (dans la section 3). Nous proposons ensuite un algorithme d'animation de graphe (section 4), que nous illustrons par un exemple (section 5). Dans la section 6, nous faisons un bilan des travaux proposés dans l'article et nous présentons nos perspectives de recherche.

2. Etat de l'art

La représentation graphique permet de fournir au lecteur un maximum de renseignements synthétiques, qui ne sont que très rarement explicités dans les données brutes. La représentation graphique est un excellent vecteur d'analyse des données complexes, (Tufté, 1983), (Tufté, 1990), (Tufté, 1997). Par exemple, on peut se demander s'il existe des regroupements dans ce réseau ? La visualisation graphique peut nous donner une vue sur l'organisation des données ou en faire apparaître les propriétés structurelles pour la question tel élément est-il important dans le réseau ? Ces tâches d'analyse seraient très difficiles, voire impossibles, en basant l'analyse sur du texte brut, en particulier quand la taille des données est importante.

L'analyse de l'évolution d'informations relationnelles est basée sur la visualisation de graphes de périodes et en particulier de leurs changements de structure. De nombreux chercheurs ont développé des systèmes de visualisation de réseaux, (DiBattista et al., 1999), en prenant en compte une cartographie des connectivités liées à Internet, les réseaux d'appels téléphoniques, les réseaux de citation ainsi que la visualisation progressive des domaines évolutifs de connaissances.

L'outil InfoViz développé par (Fekete, 2004), permet la création, l'extension et la visualisation en 2D d'informations pour une période spécifique. Cet outil permet la représentation graphique de gros volumes d'informations. Les données peuvent être représentées sous forme d'arbre, par le biais de graphe circulaire, permettant la visualisation non croisée des arcs. Nos travaux se basent sur le même procédé de représentation des données, en y ajoutant la dimension temporelle.

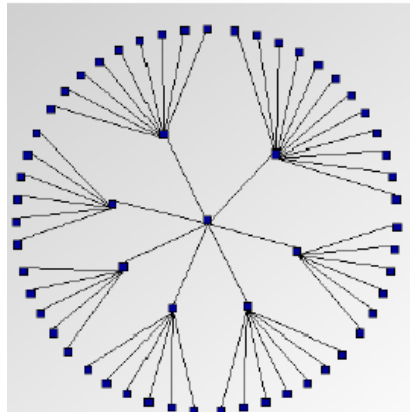


Figure 1. *Visualisation circulaire sous forme d'arbre (Fedeke, 2004).*

(Brandes, 2003) présente un système pour la visualisation de l'évolution de réseaux en 3D. La représentation s'effectue sous forme de couches dont chacune représente le réseau pour une tranche de temps donnée. Les sommets, correspondants à une entité, restent dans des positions semblables d'une couche à une autre. La visualisation de l'évolution ne porte

donc que sur les arcs entre sommets et non pas sur la structure du graphe pour chaque période considérée.

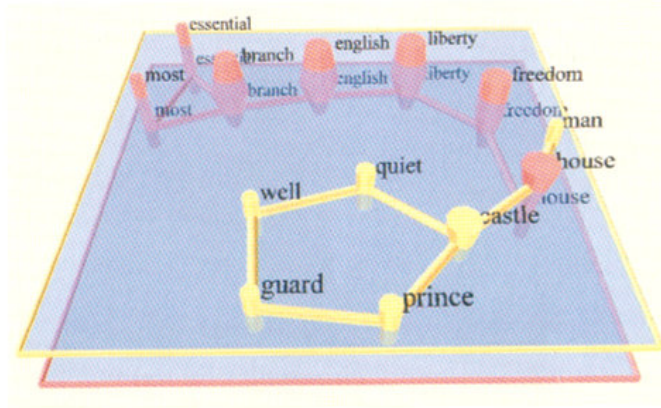


Figure 2. Représentation 3D d'un réseau évolutif (Brandes, 2003).

L'approche proposée par (Chen, 1999) (cf. Figure 3.) consiste à visualiser séparément les réseaux pour chaque période. L'inconvénient majeur de cette approche est que l'utilisateur ne dispose pas de points de repères sur les différentes représentations. Les positions d'un sommet changent d'une représentation à une autre, ce qui perturbe la carte mentale de l'utilisateur. D'autre part, l'utilisateur ne dispose pas d'une vue globale sur le réseau.

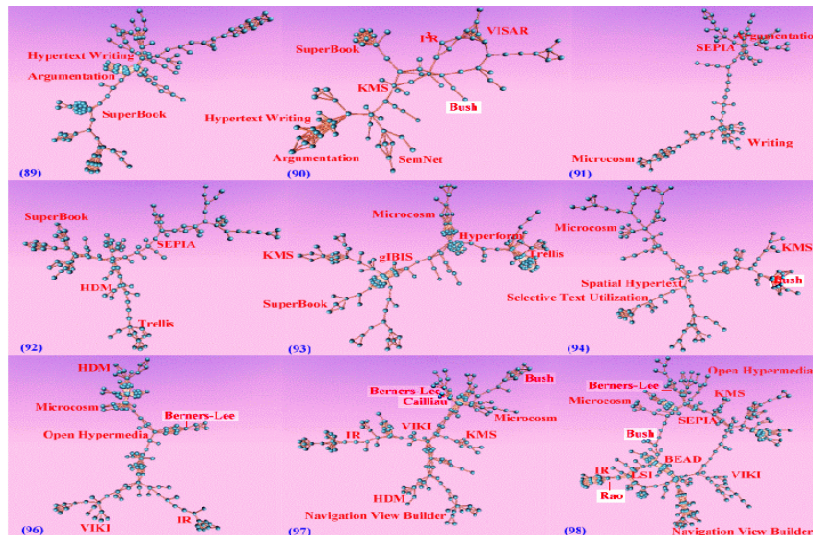


Figure 3. Représentation séparée des réseaux évolutifs (Chen, 1999).

(Erten et al., 2004) présentent le système TGRIP d'analyse visuelle de l'évolution de collaborations entre chercheurs d'un domaine donné. TGRIP produit une série de représentations 2D (cf. Figure 4), une pour chaque période, en fixant tous les sommets communs à chaque période. Les sommets et les arêtes du graphe étudié possèdent un poids calculé en fonction de la structure du graphe. Ainsi, chaque sommet a une taille relative à son poids. Le poids d'une arête est utilisé pour calculer la force d'attraction entre les sommets lors du dessin de graphe.

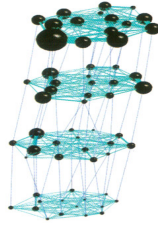


Figure 4. Représentation 2D de l'évolution de collaborations par TGRIP (Erten et al, 2004).

(Chen et al., 2003) ont introduit les techniques de visualisation animées afin de resituer des citations ou co-citations dans un ordre chronologique. La Figure 5, générée par l'outil CiteSpace (Chen, 2004), visualise les réseaux de co-citation les plus importants, à partir d'articles publiés dans le domaine considéré.

Dans cette figure, le domaine étudié est l'évolution des travaux effectués sur le dessin de graphe. Un raccord des différentes périodes considérées, par tranche de temps, des réseaux de co-citation permet une visualisation panoramique. Le dessin de graphe inclut la prise en compte du placement stratégique des sommets du graphe, tels qu'ils ont été étudiés par (Fruchterman et al., 1991), puis de (Tamassia et al., 1988). Dans l'exemple de la Figure 5, cela se traduit par la distinction de branches entourées d'un trait épais pour les périodes 1993-1995 et une mise en évidence de la période 1999-2000, entouré d'un trait fin.



Figure 5. Représentation graphique des co-citations d'auteurs d'articles publiés entre 1990 et 2003.

3. Le module VisuGraph

Selon (Tufté, 1983), « Un excellent graphique est celui qui fournit au lecteur un nombre maximum d'idées dans le plus court laps de temps en utilisant le moins d'encre et le plus petit espace possible ». Basé sur ce principe, Tétralogie permet en plus à l'utilisateur d'interagir avec le système, tout en possédant la capacité de transmettre de l'information complexe à haute densité.

VisuGraph, module de Tétralogie ¹, a été développé en Java par (Karouach, 2004) et nous proposons d'en étendre les fonctionnalités interactives de représentation.

3.2. Matrice de cooccurrences

Les données relationnelles que nous traitons sont issues d'un processus de traitement d'information effectué sous la plateforme Tétralogie, mise en place par (Dousset *et al.*, 1988). Ces informations, issues de corpus textuels semi structurés, sont représentées sous forme matricielle par croisements d'entités, dont le contenu correspond aux cooccurrences ². Ces matrices révèlent les groupes de mots apparaissant fréquemment ensemble. Un des intérêts de Tétralogie est de pouvoir prendre en compte trois dimensions, ce qui permet, entre autre, d'ajouter la notion de temporalité à une analyse classique. Dans ce cas là, le module VisuGraph permet une analyse évolutive des données relationnelles.

Considérons l'exemple d'une matrice, qui par croisement d'auteurs, indique le nombre d'articles co-écrits par chaque paire. Nous avons alors une matrice de cooccurrence entre quatre auteurs {A, B, C, D}. Ainsi le nombre d'occurrences entre les auteurs A et D pour la période 1 vaut 1.

	A	B	C	D
A	3	1	0	1
B	1	2	0	1
C	0	0	1	0
D	1	1	0	3

Tableau 1. *Matrice de cooccurrences*

3.3. Représentation graphique des matrices de cooccurrences

Notre démarche consiste alors à transformer les données de la matrice étudiée dans l'exemple du tableau 1, en une représentation sous forme de graphe dont les sommets représentent les auteurs {A, B, C, D} et les liens définissent leurs relations, comme le montre la figure 6.

¹ Tétralogie (ainsi que son module VisuGraph) sont disponibles sur station SUN pour les systèmes d'exploitation SUN/OS et Solaris, est accessible aussi bien à partir de terminaux X que de PC ou de Macintosh

² Dans le contexte de Tétralogie la cooccurrence correspondent au croisement de deux termes dans un même document.

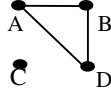


Figure 6. Représentation sous forme de graphe de la matrice de croisement des auteurs.

Dans VisuGraph, nous définissons un graphe $G = (V, E)$, comme un ensemble de sommets pondérés V et d'un ensemble d'arcs valués E , orientés ou non selon le contexte d'analyse et joignant chaque paire de sommets. Ces derniers sont généralement représentés sous forme de cercles, reliés par des arcs sous forme de courbes ou segments. Dans cet article, nous ne présentons que des exemples de graphes dont les arcs ne sont pas orientés.

Dans un graphe, tous les éléments (sommets et arcs) n'ont pas la même importance ou le même rôle dans la structure locale ou globale du graphe. Pour pouvoir identifier visuellement les caractéristiques de chacun de ces éléments, il est indispensable d'introduire des variables visuelles afin de rendre la représentation du graphe plus riche en information. Pour cela, nous utilisons la couleur (ou intensité de couleur). Plus le lien entre deux sommets sera important (forte valeur associée à l'arc), plus il sera représenté avec une forte intensité de couleur et inversement. La valeur initiale est relative au maximum de la matrice de cooccurrence, permettant ainsi un dessin des arcs non agressif pour l'utilisateur. Cette fonctionnalité est réglable à travers un slider gradué, permettant de faire varier le niveau de couleur associé à l'intensité de chaque arc.

Ainsi la Figure 7 illustre, à partir d'un extrait de matrice de cooccurrence, ce principe d'intensité de la couleur des arcs. Il en est de même pour les sommets, représentés par défaut sous forme de barres ou d'histogrammes colorés, dont la taille est proportionnelle aux valeurs des métriques. Dans l'exemple du Tableau 2, la valeur de la métrique du sommet A vaut 12, celle de B vaut 11, celle de C vaut 9, celle de D vaut 9.

	A	B	C	D
A	12	2	2	7
B	2	11	7	1
C	2	7	9	0
D	7	1	0	9

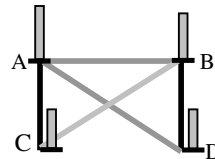


Tableau 2. Matrice de cooccurrences

Figure 7. Codage de la métrique des arcs

3.4. Visualisation de graphes par forces et ressorts

La représentation graphique des données relationnelles par VisuGraph, comme défini dans le paragraphe précédent, a des limites. En effet, cette méthode ne permet pas d'éviter les croisements d'arêtes, rendant le graphe complexe. L'objectif est de minimiser le nombre de croisement d'arc et, quand c'est possible, d'obtenir un graphe planaire. Dans cette perspective, nous proposons d'assimiler les arcs des graphes à des ressorts, en nous basant sur les travaux de recherche sur les algorithmes de ressort existants. Le principe repose sur un processus itératif, qui permet de recalculer les positions des sommets en fonction de leur liaison et des forces appliquées, jusqu'à stabilisation.

4. La contribution

Notre contribution concerne l'amélioration des fonctionnalités de représentation de données, ainsi que leur paramétrisation et l'animation des différents graphes de période.

4.1. Fonctionnalités et paramétrisation

L'amélioration de l'ergonomie du prototype passe par la facilité d'accès aux différentes fonctionnalités, aussi bien dans leur sélection que dans leur application. Pour ce faire, nous avons réalisé un menu en deux parties distinctes.

- les **fonctions** permettent à l'utilisateur de spécifier, entre autres, les méthodes à appliquer à la représentation (graphe circulaire, optimisé ou réduit, icônes représentant les sommets, coloration des sommets et des arêtes, affichage des noms des sommets, recherche d'un sommet spécifique, choix de la couleur de l'écran...). Une des fonctionnalités proposée est l'extraction de sous graphe permettant d'étudier l'environnement spécifique d'un sommet. Une fois celui-ci sélectionné, le sous graphe est obtenu par transitivité, offrant les informations suivantes : le nombre de sommets connectés pour chaque degré de transitivité, la distance moyenne, la centralité. L'augmentation (diminution) du degré de transitivité engendre l'affichage (la suppression) des voisins pas à pas.
- le **paramétrage** permet de contrôler le dessin avec la possibilité de modifier la profondeur de la transitivité, les coefficients des forces d'attraction et de répulsion, le seuil d'affichage des sommets en fonction de leur pondération, le slider de l'intensité des arcs, la granularité du partitionnement.

La place occupée par le menu reste volontairement limitée, permettant un gain de place pour l'affichage du graphe.

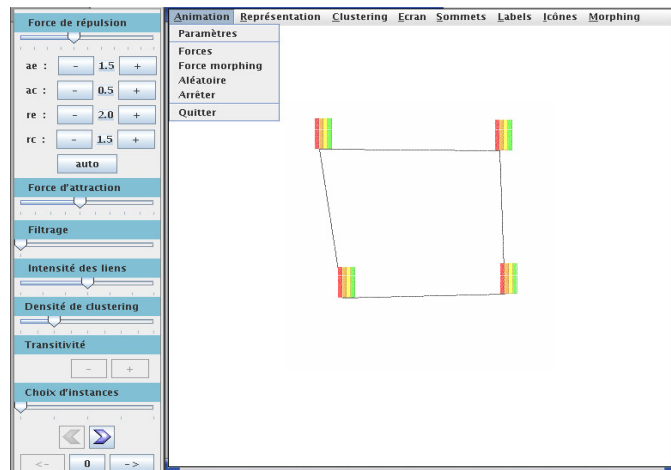


Figure 9: Menu permettant le choix des fonctions (barre horizontale en haut de la fenêtre de représentation) et la paramétrisation (colonne à gauche)

4.2. L'animation de graphes pour révéler l'évolution des données

4.2.1. De la représentation globale à la représentation par période

Concernant l'aspect évolutif, ce prototype se base sur la fragmentation en plusieurs périodes homogènes du point de vue de l'unité temporelle, d'une matrice de cooccurrence. La représentation des différentes périodes temporelles s'effectue en deux étapes.

1^{ère} étape : Toutes les périodes confondues sont représentées sur un même graphe global.

Puis, nous proposons une visualisation individuelle de chaque période. La lisibilité du graphe global est améliorée par l'application des forces, tout au long de sa représentation : non chevauchement des sommets et croisement minimal des arêtes. Ce graphe sert de base aux graphes partiels restituant chaque période.

2^{ème} étape : Le choix, par l'utilisateur, des fonctionnalités à appliquer sur le graphe (visualisation circulaire, affichage par seuil,...), pour obtenir davantage d'informations sur la structure des données, ciblées sur son axe de recherche (détection d'acteurs importants, recherche de signaux faibles, ...). Il peut aussi choisir le mode de représentation des sommets traduisant leur importance pour chaque période, sous forme de nuance, de cercle ou d'histogramme proportionnels à la métrique du sommet.

Pour des phénomènes peu ou non cumulatifs, l'interprétation des données relationnelles est souvent faussée, si l'extraction et la visualisation des connaissances se limite au cas statique (toutes les périodes y sont cumulées au lieu de les représenter individuellement). En effet, la visualisation globale peut révéler un graphe connexe, alors que les représentations partielles ne confirment pas cette propriété : le graphe n'a à aucun moment été connexe.

De plus, le graphe global ne nous fournit que des renseignements d'ordre général, qui ne permettent pas de préciser la fonction de chaque entité à un instant donné. Par contre, dans une étude sur les co-citations (phénomène qui peut être intéressant sous un aspect cumulatif), la représentation statique permet de situer l'ensemble des influences (école, courant, communauté) qui fait l'originalité d'une démarche de recherche. Il est intéressant de placer les sommets de façon à ce qu'ils apportent le plus d'information possible, en leur attribuant une position stratégique, plutôt qu'aléatoire.

4.2.2. Le principe

L'animation de graphe permet, en se basant sur l'analogie espace/temps, de détecter, comprendre et même prévoir les tendances significatives, au travers de la visualisation de l'évolution des données. L'objectif est de réaliser une animation se basant sur le graphe global et visualisant de manière fluide et successive les différentes représentations par période. Nous attribuons un sommet virtuel (non visible dans le dessin mais dont la présence est prise en compte dans le graphe) qui servira de repère pour chaque période considérée. Ces sommets virtuels sont fixés dans un ordre chronologique et de façon équidistante sur le contour de la fenêtre de visualisation (comme les heures sur un cadran). Le dessin de graphe est influencé par l'attribution de nouveaux arcs reliant chacun des sommets aux repères temporels, qui le concernent, en leur attribuant un poids plus important que la valeur de la métrique d'arête la plus grande. Ceci engendre un déplacement, vers certains repères, en fonction de la plus ou moins forte présence d'un sommet dans chaque période. Dans cette approche, le dessin du graphe global permet de décanter les sommets en fonction de leurs périodes spécifiques. Les forces d'attraction et de répulsion permettent le déplacement des sommets vers les repères des périodes les caractérisant le mieux.

Après stabilisation du dessin, chaque secteur périphérique de la fenêtre correspond à une typologie d'évolution particulière, seul le centre peut éventuellement contenir plusieurs types de persistance (présence continue ou sur quelques périodes espacées).

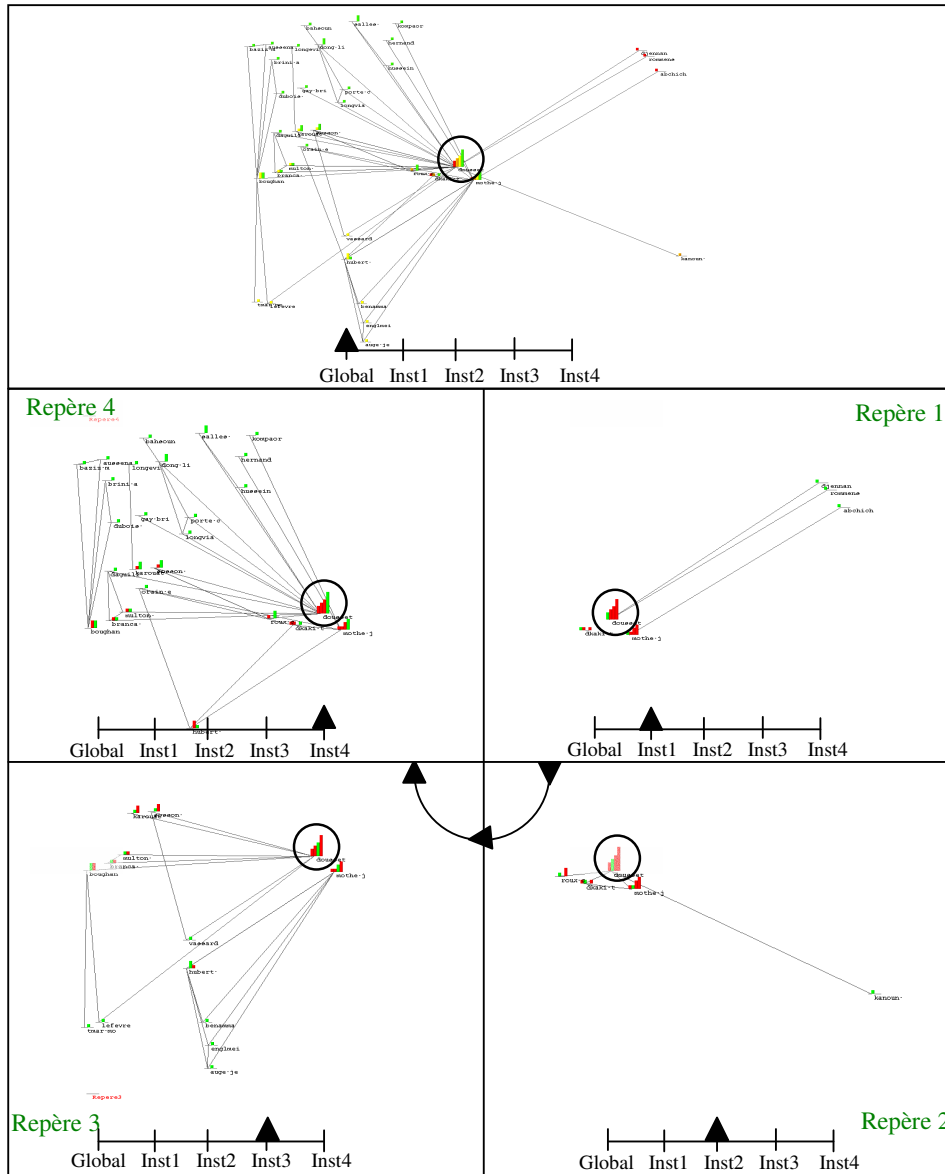


Figure 10. Cinématique réalisée à partir de graphe de plusieurs sessions d'un colloque. Les sommets représentent des auteurs et les arcs le nombre de publications communes.

Les graphes des différentes périodes peuvent être représentés individuellement, en masquant simplement sommets et arcs non concernés par la période choisie (cf. chaque rectangle de la figure 10). La visualisation successive des différentes périodes (représentées par chaque rectangle), dans le sens indiqué par les flèches, permet de créer une certaine cinématique, révélant l'évolution des données au cours du temps. Chacun des sommets est caractérisé par un histogramme, comportant autant de barres que de périodes étudiées. Dans cet exemple, chaque histogramme de sommet contient quatre barres, dont la taille est proportionnelle à la valeur de la métrique du sommet pour l'instance temporelle considérée.

Afin de distinguer la période que nous visualisons, nous affichons de couleur plus claire la barre de l'histogramme correspondant à la période visualisée.

Ainsi l'animation, présentée en exemple, révèle un auteur important, encerclé sur la figure, dont le nombre de collaborations augmente d'une période à l'autre. Nous pouvons également observer, en troisième période, la naissance de nouvelles alliances, qui se confirment dans le dernier graphe (repère 4). C'est ici que nous devons rechercher les fameux signaux faibles et tenter de prévoir leur évolution. Il est alors possible de détecter, par exemple, une structure naissante ou un changement organisationnel et d'en vérifier la pertinence au niveau de la période suivante.

5. Un exemple d'utilisation de VisuGraph

Les travaux de visualisation de (Marks *et al.*, 2005) s'appuient sur le constat suivant : un graphe peut représenter, de façon claire, plus de deux cents sommets alors qu'un écran d'ordinateur ne peut pas afficher plus de vingt lignes consécutives, résultantes d'un moteur de recherche classique.

Ainsi, il devient plus facile d'analyser les arcs entre les sommets, ainsi que les différents regroupements de sommets. La visualisation globale des documents croisés aux mots-clés permet de révéler des informations non détectables dans les données brutes.

Appliquée à la recherche d'information, la visualisation des données par croisement d'informations peut être une aide précieuse, quant à l'évaluation de la pertinence des documents.

5.1. Les apports de VisuGraph pour la Recherche d'Information

La recherche d'information se base sur des requêtes, traduisant le besoin de l'utilisateur à travers des mots-clés. En effectuant un croisement entre les documents d'une base et le type des mots-clés (concepts, termes, noms d'auteurs...), les documents les plus proches et les plus liés aux différents composants de la requête, sont détectés, via un graphe, retrouvant avec aisance, les documents les plus pertinents et cela, pour des périodes particulières.

Prenons l'exemple d'une requête : « Quels sont les articles écrits par LIU ? ». Nous générons la matrice de présence/absence, croisant le champ « auteur » avec le champ « publication », nous obtenons des arcs unitaires liant auteurs et documents. VisuGraph permet de visualiser le contenu de cette matrice. L'application d'une forte force d'attraction mène à distinguer différentes classes, constituées généralement d'un ou plusieurs articles et de leurs co-auteurs, bien représentées visuellement dans le graphe de la figure 11. Il est possible de sélectionner l'auteur « LIU » et de visualiser, par transitivité, ses publications. D'autres informations sont fournies par cette représentation graphique. Ainsi sont révélés pour chaque article ses différents co-auteurs mais aussi les différentes collaborations de chaque auteur.

Plusieurs classes de documents peuvent être reliées par un auteur. En effet, si un individu a co-écrit deux articles appartenant à deux classes distinctes, il peut se trouver à mi chemin de ces deux classes.

Dans la Figure 11, le résultat de la requête est encadré. Nous pouvons voir que LIU a écrit quatre articles. Si nous étudions ses collaborations, nous constatons que Liu a collaboré avec plusieurs équipes, composées de nombreux auteurs. Il est relié à quatre classes, bien représenté par les quatre articles qui l'entourent.

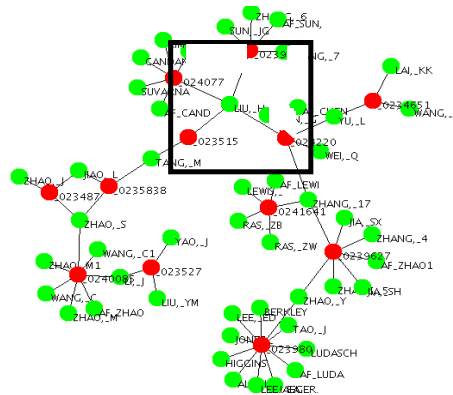


Figure 11. Graphe des auteurs (sommets clairs) et de leurs publications (sommets foncés).

5.2. L'apport de la cinématique effectuée à partir des graphes de période

L'animation entre les différents graphes de période, présenté précédemment dans la section 4, peut être également utilisé dans le cadre de la recherche d'information. Soit la requête suivante «Quels sont les articles écrits sur thème 1, thème 2 et thème 3, pour les périodes 1, 2, 3 et 4 ? ». Pour répondre à cette requête, en utilisant Tétralogie, il faut générer les matrices de cooccurrence de chaque période, croisant des auteurs et les trois thèmes. Dans une matrice, chaque valeur correspond au nombre d'articles écrits sur le thème par l'auteur. L'animation de graphes de période présenté en Figure 12, nous permet de visualiser le graphe global, qui indique, par les liaisons entre « auteur » et « thème », tous les articles parus sur chacun des thèmes, puis le graphe de chaque période. Pour cet exemple, il nous semble plus judicieux d'utiliser une représentation globale, pour laquelle les sommets sont représentés sous forme d'histogramme. Nous pouvons ainsi visualiser et comparer, d'une part l'importance (ou non) du nombre de publications pour chaque période, mais aussi observer les tendances des différents thèmes. Il apparaît ainsi clairement que le premier thème est fréquemment abordé pour la première période, mais il l'est faiblement pour toutes les autres. A l'inverse, les thèmes 2 et 3 ne sont fréquemment abordés qu'après la première période. L'utilisateur a, de plus, la possibilité de visualiser le contenu de chaque article, par simple clic sur le sommet correspondant, engendrant automatiquement l'ouverture du document dans un éditeur de texte.

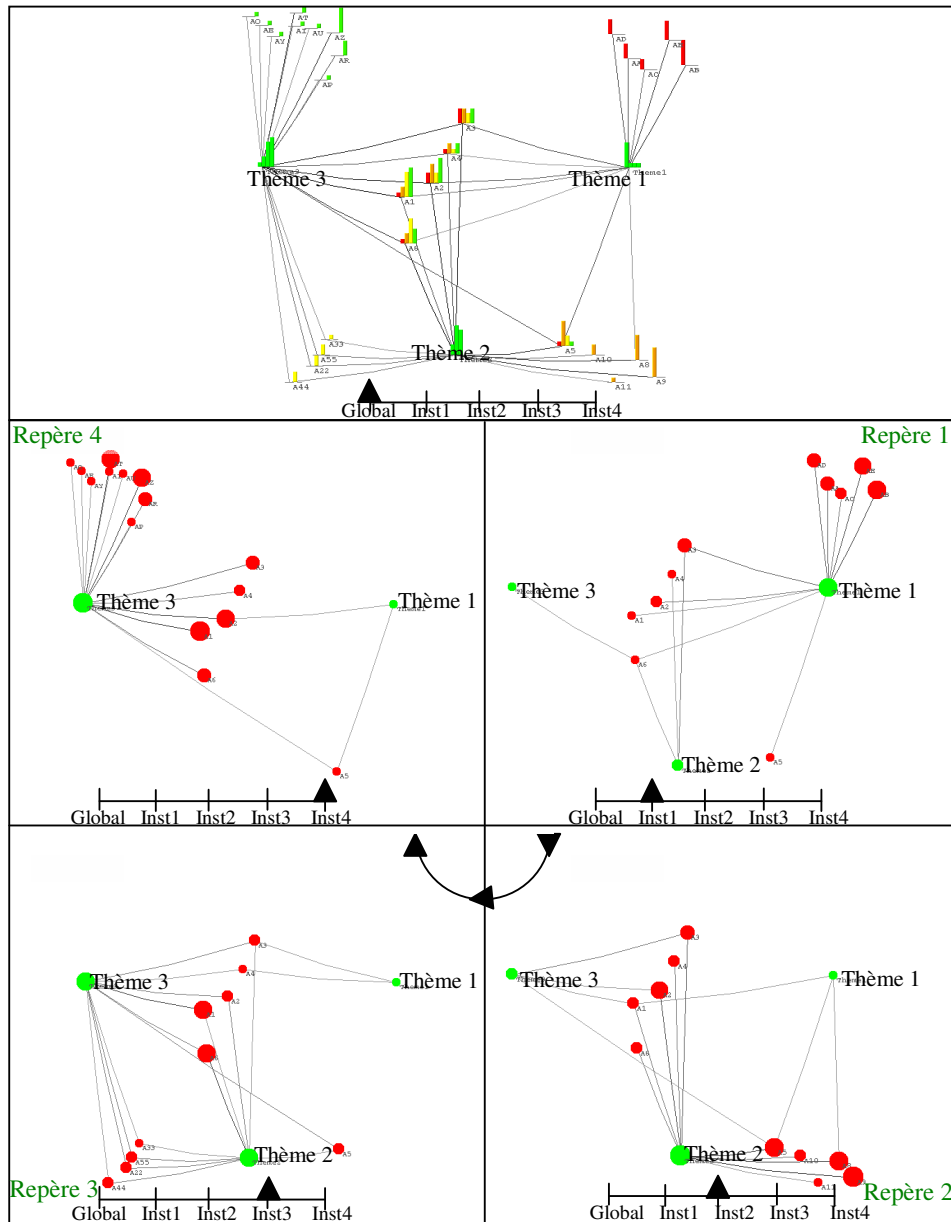


Figure 12 : Animation entre les différents graphes de période pour l'étude des publications écrites par des auteurs sur 3 thèmes spécifiques.

Pour la visualisation successive de chaque période, il semble intéressant de représenter les sommets sous forme de cercles proportionnels à la valeur de la métrique, les auteurs étant représentés par une couleur foncée et les thèmes par une couleur claire. Ainsi, nous pouvons distinguer facilement les auteurs ayant le plus publié pour la période considérée. La visuali-

sation successive et animée des graphes de période permet de montrer, l'importance du thème 1 pour la première période, puis sa perte de vitesse et l'émergence du thème 2.

La visualisation de la troisième période montre l'évolution du thème 3, tout en conservant un grand nombre de publications sur le thème 2. Enfin, lors de la dernière période, le thème 2 disparaît, alors que le thème 3 est toujours sujet à de nombreuses publications.

Ainsi pour une requête basée sur un thème spécifique, le graphe permet de retrouver avec aisances les articles écrits, mais il permet aussi, dans le cas où l'intérêt (via la requête) est porté sur un auteur spécifique, de découvrir ses publications et ses domaines d'activités.

6. Conclusion

VisuGraph apparaît comme un outil d'analyse des données relationnelles puissant et ergonomique et qui permet de révéler, comprendre et anticiper les structures sous-jacentes des données afin d'identifier leurs implications stratégiques. Ce prototype propose une visualisation globale de données relationnelles, toutes périodes temporelles confondues, mais aussi une visualisation individuelle et successive de ces dernières. Ainsi, l'analyse des données relationnelles au cours du temps est facilitée par la représentation graphique proposée, et par la visualisation de l'aspect évolutif, révélant les structures émergentes, celles disparaissant, ainsi que les nouvelles alliances.

Cependant, ce prototype est limité au niveau du nombre de données. En effet, plus le volume de données est important, plus le temps de traitement est conséquent, ce qui peut être gênant pour l'utilisateur. Il conviendrait donc d'améliorer les performances d'affichage de cet outil.

De plus, l'animation de graphes présentée est une première implémentation. Il est très directif car les sommets sont fortement attirés par les repères temporels, au détriment de leurs relations initiales. Il serait intéressant de trouver un compromis pour une animation plus souple du mouvement entre deux périodes, qui pourrait en enchaîner plusieurs.

Enfin, cette animation de graphe est conditionnée par le point de vue utilisateur qui peut être, par exemple, orienté vers la détection de signaux forts (importants ou persistants) ou de signaux faibles (apparitions, disparitions, réorganisations d'acteurs qui peuvent être potentiellement intéressantes). Ainsi, nous devons situer précisément la problématique de chacun, afin de mettre en relief ses centres d'intérêt par une visualisation accentuée des structures qu'il a envie de privilégier.

7. Références

- Brandes U., Corman S. *Visual unrolling of network evolution and the analysis of dynamic discourse*. InfoVis'02, IEEE Computer Society Press. Vol. 2, N°1, p. 40-50, 2003.
- Chen C., Carr L. *Visualizing the evolution of subject domain: A case study*. InfoVis'99, IEEE Computer Society Press. p. 449-452, 1999.
- Chen C. *Searching for intellectual turning points: Progressive Knowledge Domain Visualization*. Proceedings of the National Academy of Sciences of the United States of America, 101(suppl. 1), p. 5303-5316, 2004.

- Chen C., Kuljis J. *The rising landscape: A visual exploration of superstring revolutions in physics*. Journal of the American Society for Information Science and Technology, 54(5), p. 435-446, 2003.
- DiBattista G., Eades P., Tamassia R., Tollis IG, (1999). *Graph drawing: Algorithms for the visualisation of graphs*. Upper Saddle River, NJ:Prentice Hall, 1999.
- Dousset B., Benjamaa T., *Trilogie logiciel d'analyse de données*, Conférence sur les systèmes d'informations élaborées : Bibliométrie – Information Stratégique – Veille technologique, 1988.
- Eades P. *A heuristic for Graph Drawing*. Congressus Numerantium, vol. 42, p. 149-160, 1984.
- Erten C., Harding P., Kobourov S., Wampler K., Yee G. *Exploring the computing literature using temporal graph visualization*. Conference on Visualization and Data Analysis, 2004.
- Fayyad U. , Piatetsky-Shapiro G., Smyth P. *The KDD process for extracting useful knowledge from volumes of data*, Communications of the ACM, v.39 n.11, p.27-34, Nov. 1996.
- Fekete JD. *The InfoVis Toolkit*, Proceedings of the 10th IEEE Symposium on Information Visualization (InfoVis'04), Austin, TX, Oct 2004. IEEE Press. p. 167-174, 2004.
- Frick A., Ludwig A., Mehldau. *A fast adaptive layout algorithm for undirected graphs*. In DIMACS International Work. Graph Drawing, GD'94, 1994.
- Fruchterman TMJ., Reingold EM. *Graph drawing by force_directed placement*. Software – Practice and experience, 21, p. 1129-1164, 1991.
- Gager P., Kobourov S. *GRIP : Graph Drawing with Intelligent Placement*, 8th Symposium on Graph Drawing (GD), p. 222-228, 2000.
- Kamada T., Kawai S. *An algorithm for general undirected graphs*. In Information Processing Letters, Volume 31, p. 7-15, 1989.
- Karouach S., Dousset B. *Visualisation de relations par des graphes interactifs de grande taille*. Journal of ISDM (Information Sciences for Decision Making), V. 6 N. 57, p. 12, mars 2003.
- Karouach S., Dousset B. *Analyse d'information relationnelle par des graphes interactifs de grandes tailles*. EGC'04, 2004.
- Marks L., Hussell J., McMahon T., Luce E., *ActiveGraph : A digital library visualization tool*, Research Library, Los Alamos National Laboratory, USA, Springer-Verlag, 2005.
- Tamassia R., DiBattista G., Batini C. *Automatic graph drawing and readability of diagrams*. IEEE Transactions on Systems, Man and Cybernetics, 18(1), p. 61-79, 1988.
- Tutte, W.T. *How to draw a graph*. Proceedings London Mathematical Society, Volume 13, p. 743-768, 1963.
- Tufte E. *The visual display of quantitative information*. Graphics Press. Cheshire, p. 198, Connecticut, 1983.
- Tufte E. *Envisioning Information*. Graphics Press, 1990.
- Tufte E. *Visual Explanations*. Graphics Press, 1997.
- Van Dongen S. *A cluster algorithm for graphs*. Technical Report INS-R0010, National Research Institute for Mathematics and Computer Science in the Netherlands, Amsterdam, Mai 2000.