

Visualisation de l'évolution des informations relationnelles par morphing de graphe

Eloïse Loubier, Wahiba Bahsoun et Bernard Dousset

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

Résumé. La visualisation par dessin de graphes est un outil puissant pour analyser l'évolution des informations relationnelles. Elle permet d'identifier et d'évaluer les structures passées, actuelles et d'en déduire les futures entre acteurs et concepts extraits de la synthèse de grands corpus de données. Dans ce contexte, notre laboratoire 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 prototype 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'un algorithme de morphing. L'objectif est d'animer avec fluidité la représentation d'un graphe entre 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 en illustrant notre propos d'exemples à l'échelle.

1 Introduction

Selon Fayyad (1996), “ 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”. 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 et/ou les concepts d'un domaine. De plus, l'étude de la structure d'un réseau dans le temps (collaborations, co-citations, co-signatures, co-occurrences, 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.

Visualisation de l'évolution des informations relationnelles par morphing de graphe

Dans ce contexte d'information stratégique, notre laboratoire 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 prototype VisuGraph y ajoute la visualisation et la classification interactive des données relationnelles.

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 dynamiques 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 de morphing de graphe, qui par animation du dessin entre différentes périodes, révèle les structures successives, les changements significatifs et les acteurs et/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, sa mise oeuvre, ainsi que les tests réalisés afin de la valider.

Nous mettons l'accent sur la structure des données, l'ergonomie de l'interface (interaction homme/machine et visualisation) et les notions de points de vue utilisateurs. Puis, nous proposons d'optimiser la représentation graphique, ainsi que son animation et nous illustrons notre propos par des exemples à l'échelle.

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, Tufte (1983, 1990, 1997). Par exemple, on peut poser la question : Existe t'il 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 de la donnée est importante.

L'analyse de l'évolution d'informations relationnelles est basée sur la visualisation de graphes dynamiques. 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 liens. Nos travaux reprennent ce principe en y ajoutant une troisième dimension temporelle.

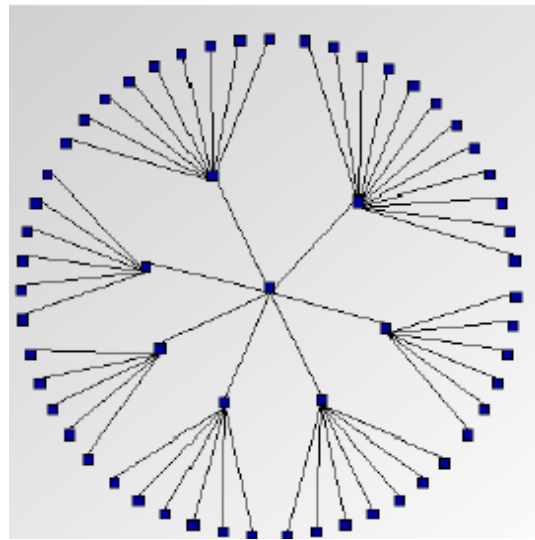


FIG. 1 : *Visualisation circulaire sous forme d'arbre : Fedeke (2004).*

Brandes (2003) présente un système pour la visualisation de l'évolution de réseaux en 3D. La présentation s'effectue sous forme de couches dont chacune représente le réseau pour une tranche de temps donnée. Les sommets, correspondant 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 liens et non pas sur la structure.

Visualisation de l'évolution des informations relationnelles par morphing de graphe

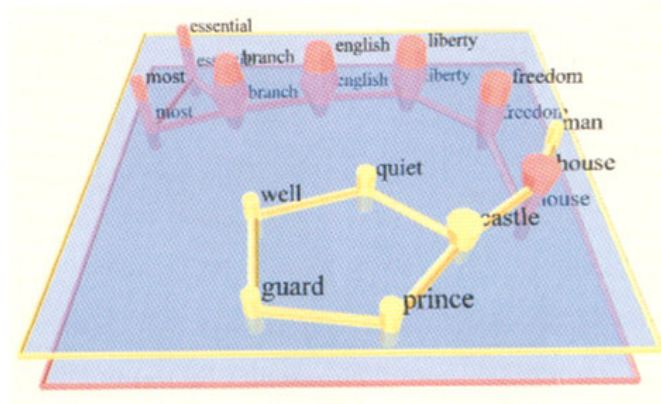


FIG. 2 : Représentation 3D d'un réseau évolutif : Brandes (2003).

L'approche proposée par Chen (1999) (cf. FIG. 2) 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.

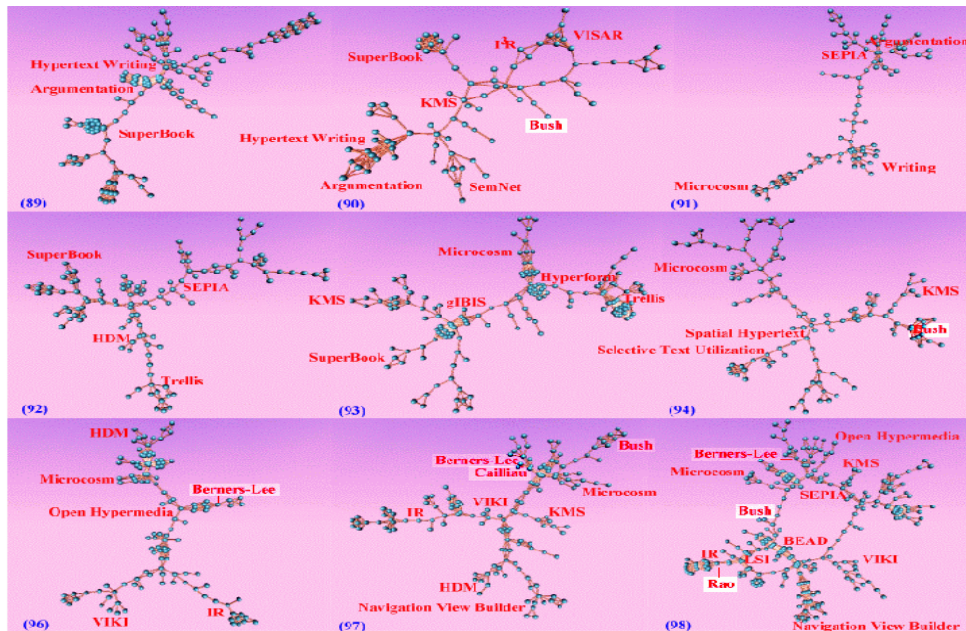


FIG. 3 : Représentation séparée des réseaux évolutifs : Chen (1999).

Erten et al (2004) présentent le système TGRIP d'analyse visuelle de l'évolution de collaborations entre chercheurs d'un domaine donné, comme extension de GRIP (2000). TGRIP produit une série de représentations 2D (cf. FIG. 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.

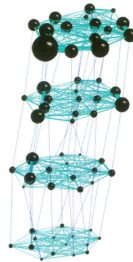


FIG. 4 : Représentation 2D de l'évolution de collaborations par TGRIP : Erten et al (2004).

Chen et Kuljis (2003) ont introduit les techniques de visualisation animées afin de resituer des citations ou co-citations dans un ordre chronologique. La FIG. 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 Reingold (1991), puis de Tamassia et al (1988). Dans notre exemple, cela se traduit par la distinction de branches encerclées de noir foncé pour les périodes 1993-1995 et une mise en évidence de la période 1999-2000, encerclée de noir clair.



FIG. 5 : Représentation graphique A co-citation map of graph drawing articles (1990 – 2003) by CiteSpace : Chen (2004).

3 Le prototype VisuGraph

Selon Tufte (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 ”. Toute représentation visuelle de l'information, possède un certain degré d'interactivité, ainsi qu'une capacité à transmettre de l'information complexe à haute densité. Un graphe $G = (V, E)$ est un ensemble de sommets V et d'arêtes E , joignant chaque paire de sommets. Ces derniers sont généralement représentés sous forme de cercles, reliés par des arcs sous forme de courbes ou segments. Il existe plusieurs types de représentations graphiques, suivant les objectifs de la visualisation. Principalement, le dessin d'un graphe s'effectue en suivant les cinq règles suivantes (Fruchterman et al., 1991) :

- Disposition des sommets dans la fenêtre de représentation.
- Minimisation des croisements des arêtes.
- Uniformisation de la taille des arêtes.
- Respect d'une certaine symétrie dans la disposition du graphe (répartition équitable).
- Les sommets ne doivent pas être placés hors de la fenêtre.

Basé sur ce principe, VisuGraph a été développé par Karouach (2004) et nous proposons d'en étendre certaines fonctionnalités. Les relations sont représentées à l'aide d'un graphe dont les sommets sont les objets et les arêtes les liens assimilés à des ressorts, suivant la théorie de Eades (1984), qui s'inspire des lois de la physique. A partir d'un état initial de forte énergie, nous laissons se relaxer le système afin que les sommets se positionnent harmonieusement les uns par rapport aux autres sans se superposer.

Au niveau de chaque sommet, la valeur de la métrique qui leur est associée est représentée sous forme d'une ou plusieurs barres d'histogramme, selon le cas statique ou dynamique, comme dans les FIG. 6. et FIG. 7.

Pour l'aspect évolutif, ce prototype se base sur la fragmentation temporelle, en plusieurs périodes homogènes, d'une matrice d'adjacence. La représentation des différents fragments temporels s'effectue en deux étapes.

- Une visualisation globale toutes périodes confondues, cas statique, puis individuelle pour chaque période. Le graphe global est optimisé tout au long de sa représentation : non chevauchement des sommets et croisement minimal des arêtes; il sert de base au dessin des graphes partiels restituant chaque période. Nous illustrons notre propos, FIG. 6, par la représentation globale d'une classe supposée connexe du graphe des co-signatures lors de quatre éditions successives d'un même colloque.
- 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 graphes de grande taille, VisuGraph propose un partitionnement de l'ensemble des sommets, ce qui conduit à travailler sur un graphe réduit plus facilement manipulable.

Chaque classe peut être visualisée soit séparément, soit avec ses connecteurs la liant aux autres classes, soit dans le contexte général reconstruit, à partir du graphe partiel, en figeant un représentant par classe.

4 La contribution

4.1 Fonctionnalités et paramétrisation

Dans la seconde étape, 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...).
- La prise en compte de la transitivité sert à extraire des sous-graphes qui permettent d'étudier l'environnement spécifique d'un sommet. Une fois celui-ci sélectionné, nous pouvons obtenir le nombre d'éléments connectés pour chaque degré de transitivité, la distance moyenne, la centralité, et afficher ou effacer, pas à pas, les voisins.
- 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, l'échelle des nuances des liens, la granularité du partitionnement.

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

4.2 Le morphing de graphe

4.2.1 De la représentation statique à la représentation dynamique

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 cf. FIG. 6 (toutes les périodes y sont cumulées au lieu de les représenter individuellement comme dans la FIG. 8).

En effet, la visualisation globale révèle 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 acteur à un instant donné.

Par contre, dans une étude sur les co-citations (phénomène plutôt cumulatif), la représentation statique permet de situer l'ensemble des influences (école, courant, communauté) qui fait l'originalité d'une démarche de recherche. Dans ce cas là, il est intéressant de placer nos sommets de façon à ce qu'ils apportent le plus d'information possible. Dans le cas de la FIG. 6, nous avons une visualisation globale des données, placées selon leur attirances les uns envers les autres, sans prendre en compte la dimension temporelle.

Dans le cas de la FIG. 7, les données sont orientées et regroupées selon leur appartenance à chacune des périodes. Nous avons recours à des repères temporels caractérisant ces dernières, comme nous l'expliquons dans le paragraphe suivant.

Visualisation de l'évolution des informations relationnelles par morphing de graphe

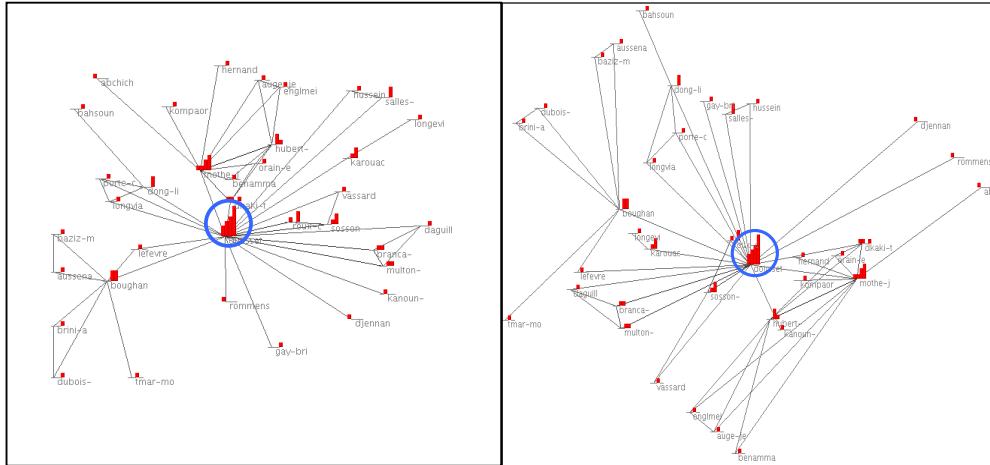


FIG. 6 : Graphe global statique de quatre éditions successives d'un colloque.

FIG. 7: Graphe global avec prise en compte de la dimension temporelle.

4.2.2 Le principe

Le morphing 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. Dans notre cas, des repères temporels non visibles, qui représentent les différentes périodes, sont fixés dans un ordre chronologique et de façon équidistante sur le pourtour de la fenêtre de visualisation (comme les heures sur un cadran). Dans notre exemple, chaque repère est un sommet invisible symbolisant une édition de la conférence auquel nous nous référerons. Le dessin de graphe est influencé par l'attribution de nouveaux liens reliant chacun des sommets aux repères temporels, qui le concernent. Ceci engendre un déplacement, vers certains repères, en fonction de la plus ou moins forte présence d'un sommet dans chaque période.

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).

Les graphes des différentes périodes peuvent être représentés individuellement, en masquant simplement sommets et liens non concernés par la période choisie cf. chaque rectangle de la FIG. 8. La visualisation successive des différentes périodes (représentées par chaque rectangle), dans le sens indiqué par la flèche jaune, 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 notre exemple, nous en avons quatre pour chaque sommet, dont la taille de chacun 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.

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. Comme le mode de représentation des sommets sous forme d'histogramme d'évolution permet de les situer dans le temps, nous remarquons que le cadran nord-ouest (repère temporel 4) contient une majorité de sommets récents. C'est ici que nous devons rechercher les fameux signaux faibles et tenter de prévoir leur évolution.

Dans l'exemple proposé dans les FIG. 6 à 8, le sommet entouré, est situé, dans l'approche statique, au centre de sa classe connexe (FIG. 6), la répartition des autres sommets s'effectuant au hasard. Dans l'approche dynamique (FIG. 7), le dessin du même graphe global permet de décanter les sommets en fonction de leurs périodes spécifiques. Dans la FIG. 7, nous appliquons les forces d'attraction et de répulsion, permettant le déplacement des sommets vers les repères des périodes les caractérisant le mieux. Cette disposition est reprise dans le dessin des graphes partiels (FIG. 8). Nous pouvons remarquer, dans le cadran nord-ouest, que deux équipes (e1, e2) se détachent du sommet central mais que celui-ci gagne, par ailleurs, trois nouvelles collaborations (c1, c2, c3).

Visualisation de l'évolution des informations relationnelles par morphing de graphe

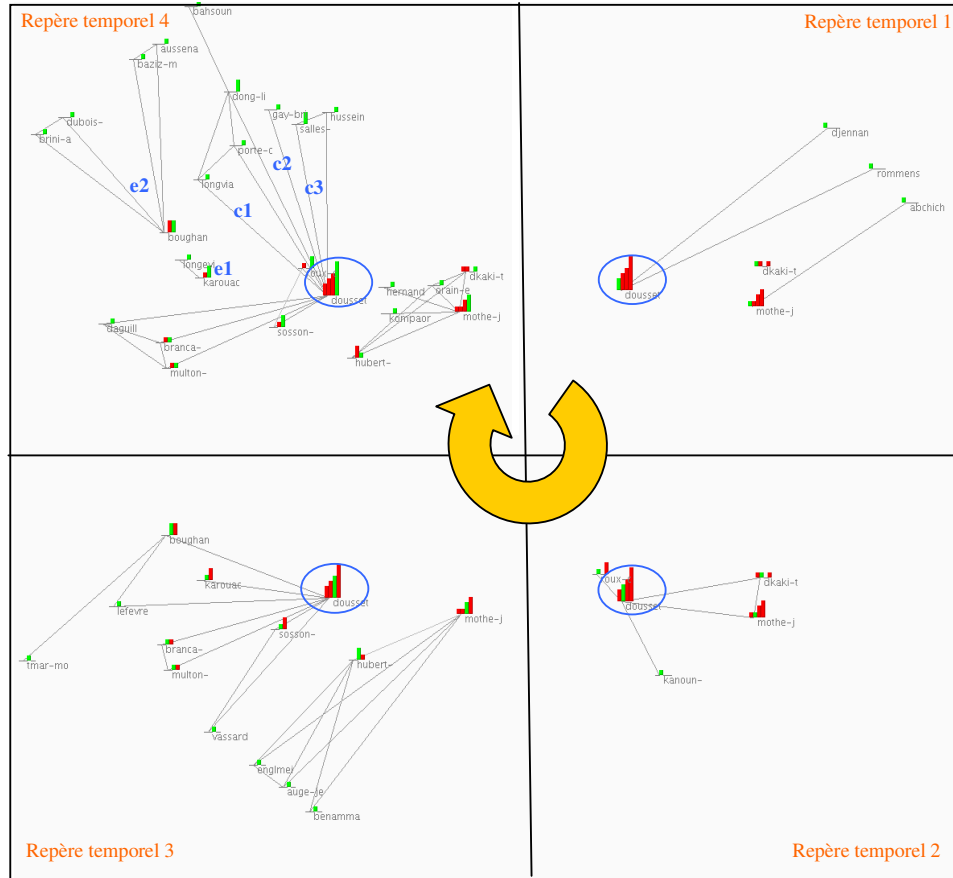


FIG. 8 : Mise en évidence du mouvement circulaire par la visualisation individuelle des périodes par morphing de graphe.

5 Exemples, à l'échelle, d'utilisation de VisuGraph

Ce prototype est utilisé dans l'étude des réseaux d'alliances contractées dans le domaine des biotechnologies, permettant ainsi d'évaluer la position stratégique des leaders et d'identifier les nouveaux entrants à forte potentialité.

De même, il est possible d'étudier l'évolution des équipes de recherche par le biais des co-signatures et des collaborations dans les contrats, notamment européens.

VisuGraph est aussi bien adapté à l'analyse des co-dépôts de brevets entre sociétés, afin de révéler des réseaux de collaboration ainsi que les liens privilégiés avec la sous-traitance. Dans le contexte d'Internet, le co-référencement de sites Web est décrypté, afin d'évaluer, entre autres, la pertinence des pages visitées.

Dans le domaine de la recherche d'informations, VisuGraph permet la détection de documents pertinents. Si nous utilisons, par exemple, une matrice de co-occurrence croisant des termes et des documents, les liens représentés permettront de révéler la pertinence d'un mot-clé dans un contenu textuel. En effet, pour chaque paire « mot-clé/document », nous obtenons le nombre d'occurrences, c'est-à-dire le nombre d'apparitions du terme dans le document. En affichant la valeur des liens, nous pouvons visualiser directement les documents les plus pertinents, c'est-à-dire ceux dont les liens avec le terme choisi sont les plus importants. Dans un contexte évolutif, cela permet de visualiser les documents les plus riches en informations sur plusieurs périodes, en révélant les compléments d'informations au cours du temps.

6 Conclusion

VisuGraph apparaît comme un outil d'analyse des données dynamiques, qui est puissant et ergonomique et qui permet de révéler, comprendre et anticiper les structures sous-jacentes afin d'identifier leurs implications stratégiques. Tout d'abord, ce prototype est limité au niveau du nombre de données pouvant être représentées. En effet, plus le nombre de données est important (plus de mille nœuds), plus les animations apparaissent saccadées. Il conviendrait donc d'améliorer les performances de cet outil. De plus, ce prototype nécessite cependant d'être amélioré, l'algorithme de morphing présenté n'étant qu'une première tentative. Il est, en effet, très directif car les sommets sont fortement attirés par les repères temporels, au détriment de leurs relations initiales. Il nous faudrait donc trouver un compromis pour une animation plus souple du mouvement entre deux périodes, puis une cinématique appropriée, qui pourrait en enchaîner plusieurs. De plus, ce morphing est conditionné 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 privilégie.

Références

- Brandes U., Corman S., (2003). *Visual unrolling of network evolution and the analysis of dynamic discourse*. InfoVis'02 Vol. 2, N°1, 40-50.
- Chen C., (2004). *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), 5303-5316.
<http://www.pnas.org/cgi/reprint/0307513100v1.pdf>
- Chen C., Kuljis J., (2003). *The rising landscape: A visual exploration of superstring revolutions in physics*. Journal of the American Society for Information Science and Technology, 54(5), 435-446.
- DiBattista G., Eades P., Tamassia R., Tollis IG, (1999). *Graph drawing: Algorithms for the visualisation of graphs*. Upper Saddle River, NJ: Prentice Hall, 1999.
- Eades P., (1984). *A heuristic for Graph Drawing*. Congressus Numerantium, vol. 42, pp. 149-160.

Visualisation de l'évolution des informations relationnelles par morphing de graphe

- Erten C., Harding P., Kobourov S., Wampler K., Yee G., (2004). *Exploring the computing literature using temporal graph visualization*. Conference on Visualization and Data Analysis.
- Fekete JD., (2004). Proceedings of the 10th IEEE Symposium on Information Visualization (InfoVis'04), Austin, TX, Oct 2004. IEEE Press. pp. 167-174.
- Fruchterman TMJ., Reingold EM., (1991). *Graph drawing by force_directed placement*. Software – Practice and experience, 21, 1129-1164.
- Gager P., Kobourov S., (2000). *GRIP : Graph Drawing with Intelligent Placement*, 8th Symposium on Graph Drawing (GD), p. 222-228.
- Karouach S., Dousset B., (2004). *Analyse d'information relationnelle par des graphes interactifs de grandes tailles*. EGC'04, Clermont Ferrand.
- Tamassia R., DiBattista G., Batini C., (1988). *Automatic graph drawing and readability of diagrams*. IEEE Transactions on Systems, Man and Cybernetics, 18(1), 61-79.
- Tufte E., (1983). *The visual display of quantitative information*. Graphic Press. Cheshire, p. 198, Connecticut.
- Tufte E., (1990). *Envisioning Information*. Graphics Press.
- Tufte E., (1997). *Visual Explanations*. Graphics Press.

Summary

Visualization based on graph drawing makes it possible to identify and to evaluate the passed and current structures and to deduce the future ones between actors and concepts. VisuGraph is developed in order to offer to the users the visualization and the interactive classification of relational data. We propose to complete this prototype with a morphing algorithm which animates with fluidity the representation between different periods, emphasizing major elements and significant tendencies.