
Mieux comprendre les enjeux stratégiques via l'analyse relationnelle : le morphing de graphes.

Eloïse Loubier* — Bernard Dousset*

loubier@irit.fr — dousset@irit.fr

* Institut de Recherche en Informatique de Toulouse, IRIT-SIG

Université Paul Sabatier, 118, Route de Narbonne, F-31062 Toulouse cedex 4

Résumé : Dans l'utilisation de notre plateforme de veille stratégique Tétralogie, nous sommes souvent conduits à générer des structures de données relationnelles (matrices ou graphes) afin d'identifier et d'analyser les réseaux d'acteurs et les réseaux sémantiques associés aux domaines spécifiques que nous étudions. Une propriété récurrente de ces réseaux est leur complexité si on les étudie de façon statique (cumul de plusieurs périodes), mais lorsqu'on distingue les différentes phases de leur évolution, leurs instances prises séparément sont beaucoup plus lisibles et interprétables. Il est même possible de pousser plus loin l'analyse en traduisant de façon dynamique leur évolution dans le temps par le biais de l'analogie espace/temps. Il devient alors beaucoup plus facile d'appréhender non seulement la fonctionnalité des structures implicites découvertes mais aussi de comprendre leur évolution et donc de détecter les événements clés et les stratégies mises en œuvre. Dans cet article, nous allons successivement aborder les notions de dessin de graphe, de graphe évolutif, de visualisation de la dynamique des relations, d'extraction de connaissance par la cinématique simulée. Nous illustrerons notre propos de nombreux exemples touchant notamment les co-publications entre auteurs ou équipes de recherche, les jeux d'alliances entre entreprises, les relations internationales, les co-citations, les réseaux sémantiques, l'aide à la création ou la mise à jour d'ontologies. Un premier algorithme de morphing de graphe sera proposé, il sera basé sur l'analogie de la représentation du temps dans l'espace empruntée aux horloges dites analogiques.

Mots-clés : Réseau, graphes, évolution, alliances, centralité, cartographie, position, structure, stratégie, ontologie.

Keywords : Network, graphs, evolution, alliances, centrality, mapping, position, structure, strategy, ontology.

1. Introduction

Il est établi par les praticiens de l'Analyse de Données (AD) que des supports visuels appropriés permettent d'expliquer aux décideurs les structures sous-jacentes des données cachées dans de longues tables. Les outils de visualisation contribuent à l'efficacité des processus mis en oeuvre en extraction de connaissances, en offrant aux utilisateurs des représentations intelligibles et facilitant l'interaction. La visualisation intervient à différentes étapes de la chaîne de traitement : dans les phases amont pour appréhender les données et effectuer les premières sélections, lors du processus de fouille, et dans la phase aval pour évaluer les résultats obtenus et les communiquer plus facilement.

Nos travaux proposent, dans un premier temps, de visualiser les données sous forme de graphe, dont les sommets représentent les principaux acteurs et les arêtes, les liens entre ces derniers. Dans un second temps, il est intéressant de s'intéresser à l'évolution temporelle des données. En effet, si nous analysons des données du réseau sur des périodes différentes, nous pourrions en déduire davantage d'informations sur leurs tendances leurs structures. Afin de faciliter le travail de l'expert en analyse de données, nous proposons une approche, visant à placer nos sommets de façon stratégique, c'est-à-dire les situer selon des repères temporels, afin de les regrouper par période et de mieux cibler leurs changements au cours du temps. Cette approche a pour but de comprendre les mécanismes de formation et d'évolution de réseaux, permettant d'expliquer les relations entre la structure d'un réseau d'alliances interentreprises et l'innovation. L'observation de la structure du réseau d'alliances sur plusieurs périodes sera alors plus simple

Dans un premier lieu, nous présenterons un état de l'art sur l'analyse de l'évolution d'information relationnelle. Nous caractériserons ce type de données et la technique de visualisation que nous proposons. Ensuite, nous présenterons notre contribution, quant à l'analyse et la visualisation des informations relationnelles par morphing de graphe, en prenant en compte la dimension temporelle.

2. Etat de l'art

L'analyse de l'évolution d'information relationnelle est typiquement basée sur des techniques de visualisation dynamique de graphes. L'affichage d'un graphe dans sa totalité donne une indication sur la structure sous-jacente, mais aussi sur sa complexité. Il permet ainsi de faciliter l'accès à des structures mathématiques complexes et, peut être considéré comme un guide efficace pour la fouille de règles.

Brandes [1] 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 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 liens et non pas sur la structure.

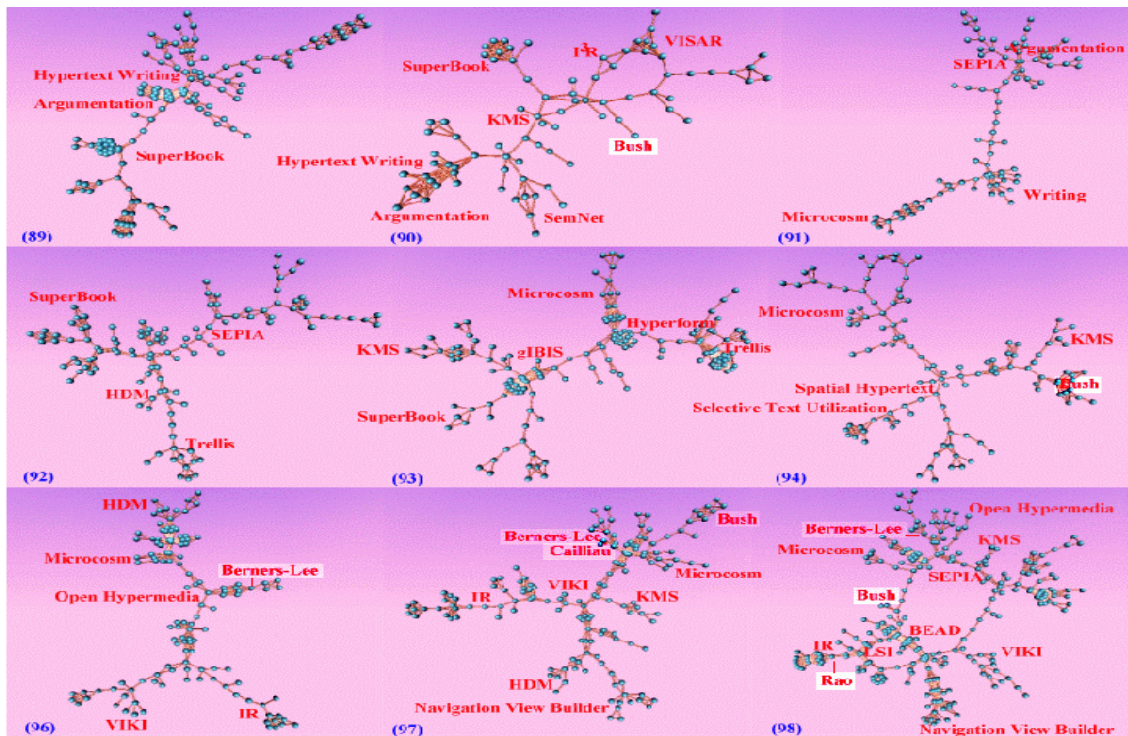


Figure 1 : Représentation séparée des réseaux évolutifs : Chen [2]

L'approche proposée par Chen [2] (cf. figure 1) 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. Le point commun entre toutes ces techniques est l'utilisation de la même technique de dessin basée sur la notion de forces.

Une première approche, sur notre prototype VisuGraph, a été proposée par Karouach [3], visant à partir d'une matrice d'adjacence (binaire ou valuée) en 3D, « la troisième dimension représentant la variable temps », la décomposition en périodes suffisamment homogènes. Le graphe global (toutes périodes confondues) est ensuite dessiné et son dessin est optimisé. Dans une seconde phase sont dessinés les graphes partiels correspondants aux différentes périodes en prenant le graphe global comme point de départ. Enfin, par morphing l'utilisateur passe d'une période à l'autre visualisant au mieux les évolutions de la structure, les positions relatives des sommets et la force des liens qui les unissent.

3. Pré requis

La plupart des spécialistes de l'intelligence technologique reconnaissent au moins trois étapes dans le processus : la collecte, le traitement et la diffusion de l'information selon Jakobiak [4].

Le logiciel Tétralogie, développé à l'IRIT par Dousset [5], Dkaki [6], Chrisment[7] permet le traitement des données et la visualisation graphique des résultants, dans un contexte de veille stratégique.

3.1. Les données relationnelles

Les graphes que nous concevons se basent sur des données relationnelles, issues de matrices tridimensionnelles, dont la troisième dimension représente le temps. Les travaux fondamentaux de Bertin [8] constituent une référence incontournable à ce niveau. Dans sa représentation la plus simple, un réseau est constitué de nœuds/sommets/vertices ou agents/acteurs/personnes et de relations qui les unissent, les liens/arcs.

Notre approche se cible sur une première représentation globale des sommets et de leurs liens, c'est-à-dire, leur représentation pour l'ensemble des instances considérées. Nous proposons ensuite une représentation spécifique à chaque instance.

3.2. Dessin de graphe

Afin de placer au mieux nos sommets, nous avons décidé d'avoir recours aux fonctions d'attraction et de répulsion des sommets. Selon Eades [9], un graphe est comparable à un modèle de ressort en s'inspirant des lois physiques pour dessiner le graphe. Il associe les sommets à des objets et les arêtes à des ressorts reliant ceux-ci. Un tel système engendre des forces entre les sommets ce qui entraîne leur déplacement. En simulant des frottements qui diminuent progressivement l'énergie initiale, le graphe se stabilise sous une forme plus lisible.

L'objectif est de rendre notre graphe plus visible permettant de distinguer plus facilement les sommets qui ont de fort liens (suite à l'application de la force d'attraction ces derniers se rapprocheront), de ceux qui n'ont pas de fortes liaisons (sommets se repoussant d'avantage).

La force d'attraction entre les sommets peut être proportionnelle à la force du lien entre eux. La force d'attraction entre deux sommets v_i et v_j est donnée par :

$$fa(v_i, v_j) = \beta_{ij} \times d_{ij}^{aa} / K$$

β_{ij} est fonction du poids de l'arête (v_i, v_j) et du poids des sommets v_i et v_j . Le facteur K est calculé en fonction de l'aire du dessin et du nombre de sommets du graphe et d_{ij} est la distance entre v_i et v_j dans le dessin.

3.3. Valuation de l'information

Chaque sommet n' a pas la même importance dans un graphe. Pour rendre compte de la valeur des sommets et de leurs liens, il est important de leur attribuer une valeur calculée, nommée « valeur de métrique ».

La valeur de la métrique de chaque sommet correspond à sa valeur indiquée dans sa matrice, en sa diagonale. Il s'agit de la somme de ses occurrences pour la période considérée.

La valeur des liens correspond à la valeur de la matrice ayant pour abscisse le premier sommet et pour ordonnée le second (et vice-versa). L'exemple suivant illustre cette application.

	DOUSSET	MOTHE	DKAKI	HUBERT	KAROUAC	ROUX
DOUSSET	3	1	1	0	0	1
MOTHE	1	1	1	0	0	0
DKAKI	1	1	1	0	0	0
HUBERT	0	0	0	0	0	0
KAROUAC	0	0	0	0	0	0
ROUX	1	0	0	0	0	1

Tableau 1 : Extrait de la matrice de données de VSST

Pour chaque sommet, nous établissons un histogramme, dont chaque barre correspond à la valeur de la métrique pour chaque instance, au prorata de la métrique maximale.

4. Mise en place d'un morphing évolutif

Notre travail repose sur un graphe à plusieurs instances. L'objectif est d'orienter la disposition de nos sommets de façon temporelle. Ainsi chaque sommet devra être placé selon son appartenance à une période. L'orientation des données simplifiera la visibilité puisque les données seront situées selon un repère temporel. L'expert verra plus facilement les instances durant lesquelles les données ont été présentes. Nous disposons nos repères telles les heures sur une horloge. Ainsi nous aurons une vision circulaire de l'évolution des données au cours du temps. Les sommets seront fixés dès notre première visualisation globale du réseau. Le passage d'une instance n'engendre aucun déplacement de sommet, permettant ainsi à l'utilisateur d'avoir pour repère la première visualisation globale, fixant la position des sommets.

Les repères sont des sommets invisibles à l'œil nu. Chaque sommet se verra relié (de façon invisible pour l'utilisateur) à chaque repère d'instance auquel il appartient. Nous avons décidé d'attribuer comme métrique de lien entre chaque sommet et son (ou ses) repère(s) temporel(s), la valeur de la métrique du sommet pour l'instance considérée. Ainsi le lien entre le sommet et son

repère est forcément supérieur à tous les autres liens inter sommets. Le principe de dessin de graphe, que nous utilisons, permet de stabiliser le graphe. En effet, les liens entre les sommets et leurs repères étant forts, les sommets seront naturellement attirés par leur repère.

Nous proposons une première visualisation globale des données, orientées selon nos repères temporels, par le biais des forces d'attraction et de répulsion.

Puis, nous proposons une représentation individuelle, pour chaque instance.

4.1. Principe

L'objectif du morphing est d'animer graphiquement les données entre différentes instances tout en faisant ressortir visuellement les éléments majeurs et les tendances significatives. Pour ce faire, il nous faut placer les sommets de notre graphe de façon stratégique. Nous attribuons pour chaque instance considérée un repère temporel.

Ainsi, nous aurons autant de repères que nous avons d'instances. Ces repères sont placés de façons équidistantes sur le périmètre du rectangle de la fenêtre de représentation. Chaque sommet est relié aux repères des instances auxquelles il appartient. Ainsi, si un sommet apparaît aux instances 3 et 4, il sera relié aux repères 3 et 4. Les liens entre les sommets et les repères sont invisibles. La valeur de ces liens correspond à la valeur de la métrique du sommet considéré.

En effet, comme nous l'avons vu dans le chapitre précédent, la valeur de la métrique des sommets est supérieure ou égale à la somme des liens avec ce sommet. Ainsi, en attribuant la valeur de la métrique du sommet au lien entre le sommet et le repère temporel, nous avons un lien entre le sommet et le repère plus fort que n'importe quel autre lien entre le sommet considéré et tout autre sommet. Ainsi, nous pouvons appliquer notre dessin de graphe.

Par l'application des forces d'attraction et de répulsion, le sommet sera davantage attiré par les repères auxquels il est relié, que par les autres sommets. L'illustration suivante montre ce principe de placement des sommets vis-à-vis des repères temporels fixés.

Chaque sommet va donc se placer de façon stratégique vers les repères temporels des instances auxquelles il appartient. Ainsi, un sommet appartenant à toutes les instances se placera vers le centre de la fenêtre de représentation. Un sommet n'appartenant qu'à deux instances se placera entre les deux repères correspondants.

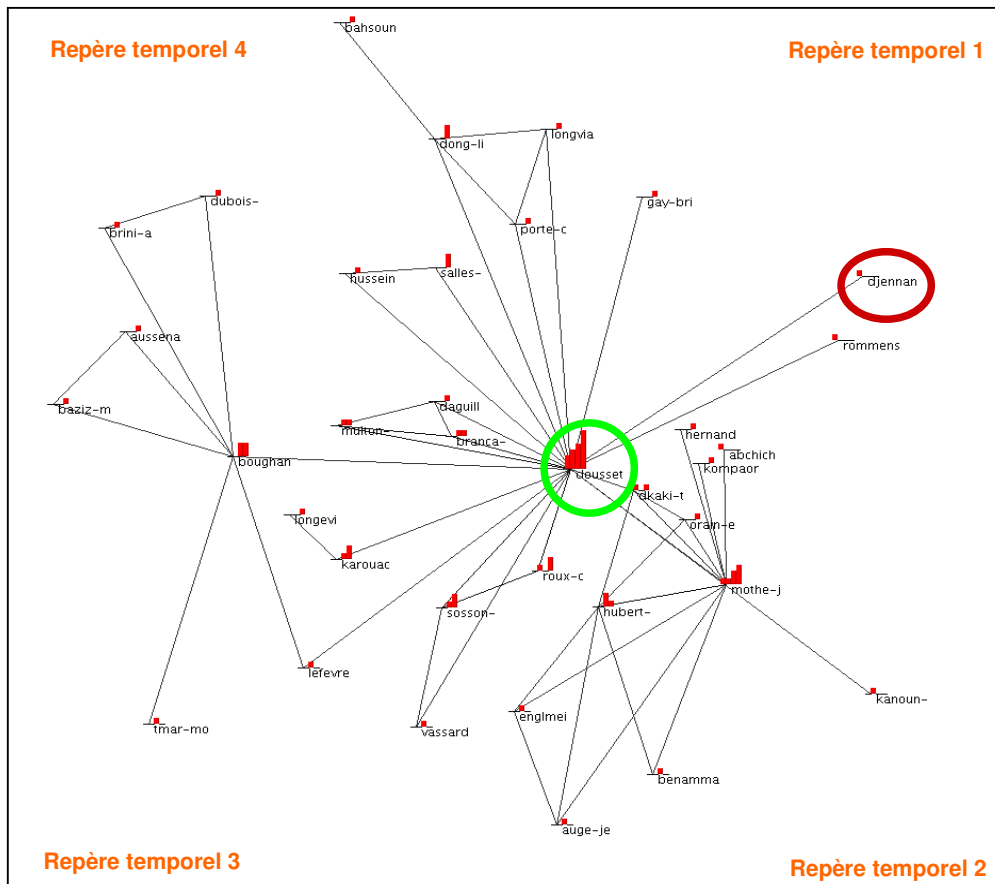


Figure 3 : Placement des sommets et des repères temporels

Dans la figure 3, nous avons placé dans un premier temps nos sommets. Puis, nous avons placé quatre repères temporels, puisque nous avons quatre instances.

Nous appliquons nos forces d'attraction et de répulsion. Nos sommets sont alors placés de manière stratégique et nous pouvons voir que le sommet encerclé de vert est placé au centre de la fenêtre. Il appartient donc aux quatre instances. Le sommet encerclé de rouge est placé dans la partie supérieure et droite de la figure, très proche du repère temporel 1. Nous pouvons donc affirmer que ce sommet appartient uniquement à l'instance 1.

L'objectif de cette représentation est de permettre de déduire l'évolution des données au cours du temps. Cette première ébauche permet cette approche.

En effet, suite à l'application décrite ci-avant, si un sommet est situé, par exemple entre les repères temporels 1 et 2, et si nous visualisons la représentation pour l'instance 1, nous pouvons déduire, de par le positionnement, que le sommet sera présent pour la visualisation de l'instance 2.

En activant la visualisation successive, instance par instance, nous pouvons voir l'évolution des données par le mouvement perçu. Les sommets sont fixés dès la visualisation globale. Ainsi l'utilisateur a des repères visuels fixes.

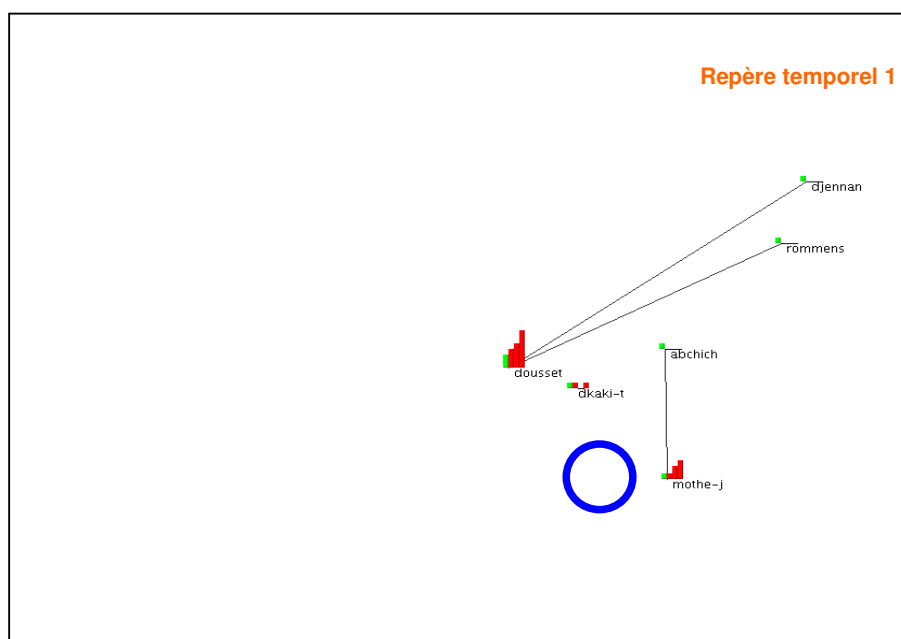
Les repères temporels étant placés de façon chronologique et sur le périmètre du rectangle de représentation, la visualisation des repères est telle la visualisation des heures sur une horloge.

4.2. Exemple

Si nous reprenons l'exemple vu précédemment, nous pouvons, suite à la visualisation globale, voir nos données par instance.

Ainsi, seuls les sommets présents pour l'instance considérée sont affichés. L'exemple 4 illustre la représentation de l'instance 1.

Nous voyons que seuls les points situés vers le repère temporel 1 apparaissent. L'exemple du sommet Mothe, encadré de bleu, montre que cet élément appartient à l'instance 1, mais que de par sa position, il appartient aussi à l'instance 2.



Exemple 4 : Représentation du morphing de graphe pour l'instance 1.

De même, le codage par histogramme permet de confirmer cette déduction. En effet, comme nous l'avons vu dans le chapitre précédent, chaque sommet est doté d'un histogramme dont chaque barre correspond au codage de la valeur de la métrique du sommet, pour l'instance considérée.

Ainsi, comme nous avons quatre instances, notre histogramme sera composé de quatre barres, invisibles si le sommet est absent pour l'instance considérée.

Pour tout sommet à une seule barre, et apparaissant pour l'instance 1, nous pouvons affirmer que ce dernier n'apparaîtra pas pour les autres instances.

Inversement, pour l'exemple du sommet Mothe, de notre exemple précédent, nous pouvons voir les trois premières barres de l'histogramme apparaître. Nous pouvons donc en déduire que ce sommet sera présent pour les instances 1, 2 et 3. Cela confirme notre première prédiction, par visualisation du graphe.

Il en est de même, pour les instances suivantes, comme le montrent les exemples 5,6 et 7.

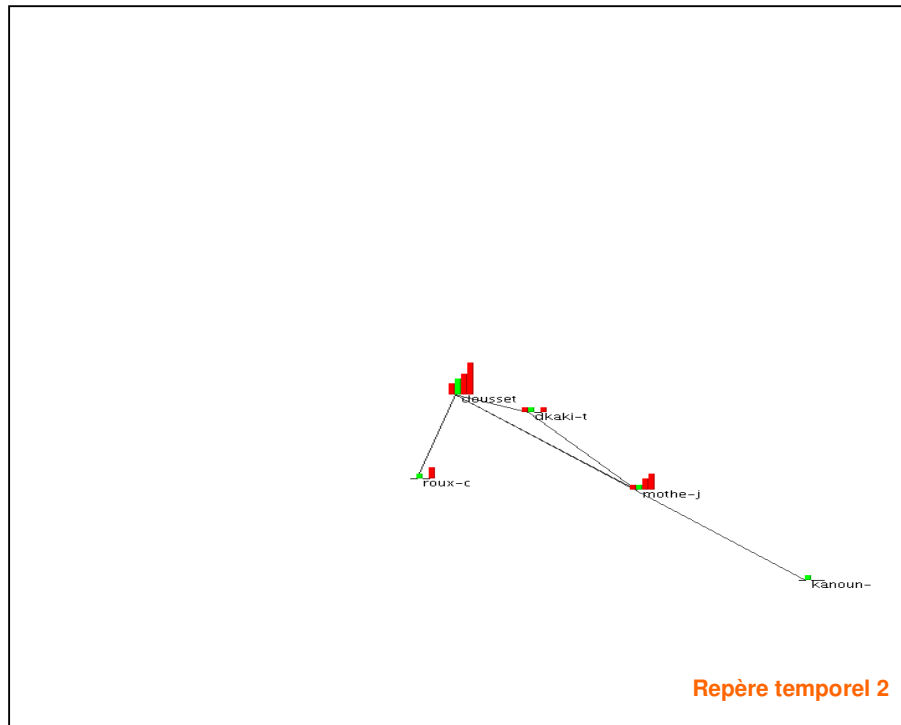


Figure 5 : Représentation de l'instance 2.

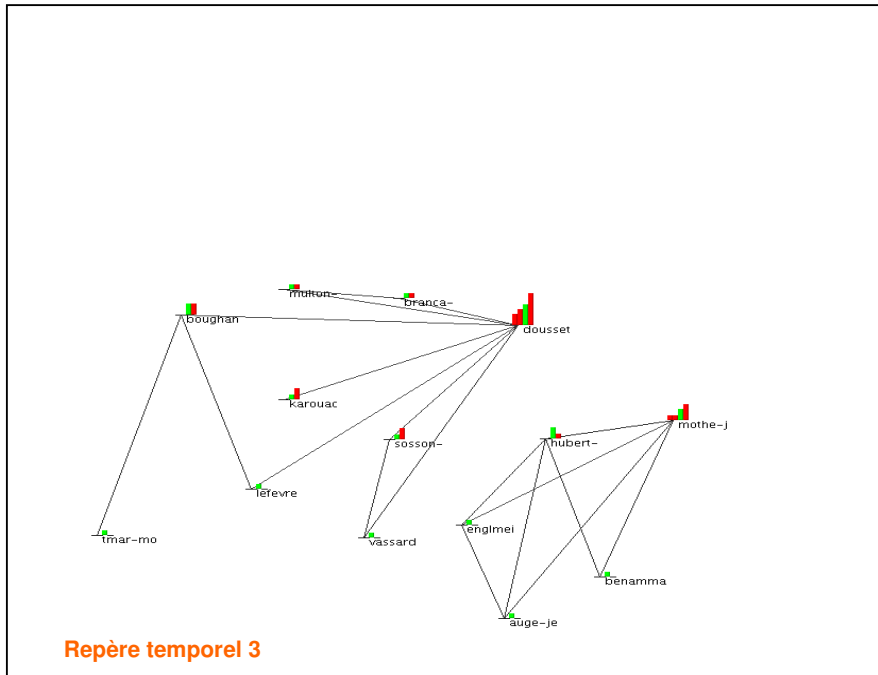


Figure 6 : Représentation de l'instance 3.

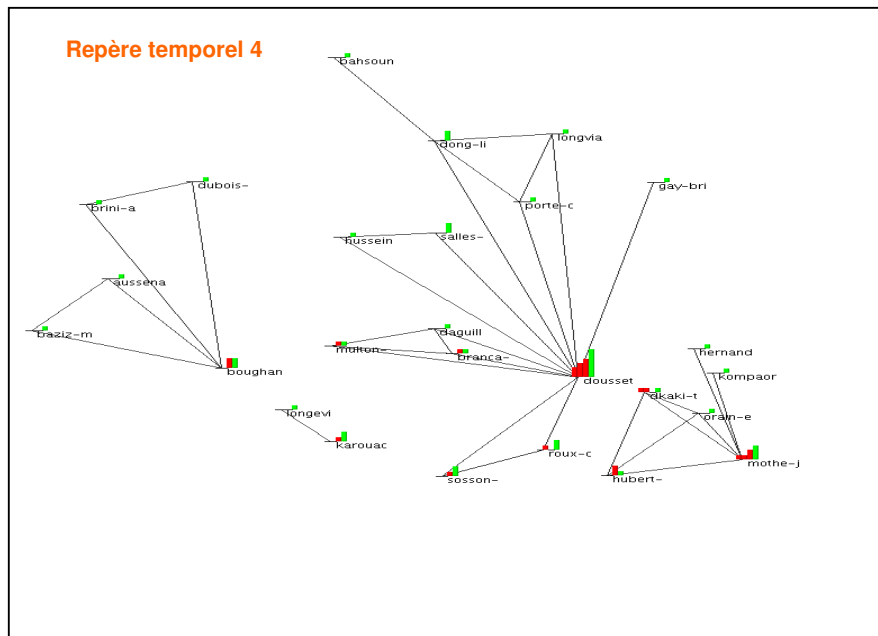


Figure 7: Représentation de l'instance 4.

Nous obtenons une vision circulaire de nos données au cours du temps. La figure 8 reprend les exemples vus précédemment pour mettre en évidence cette notion de mouvement et d'apport d'information au cours du temps.

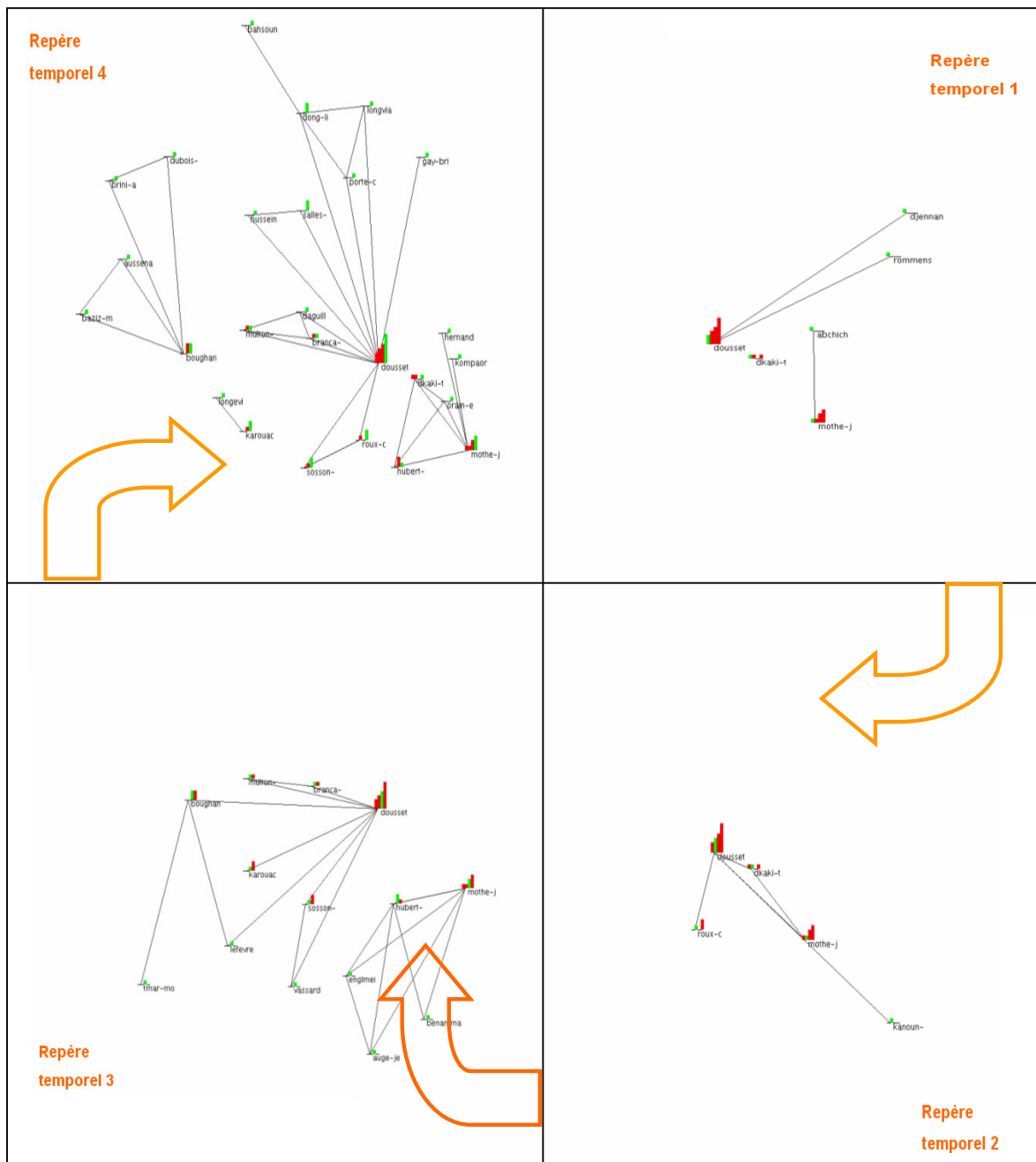


Figure 8: Mise en évidence du mouvement circulaire par la visualisation individuelle des instances par morphing de graphe.

5. Conclusion et perspectives

Les travaux présentés dans cet article, présentent une approche de la visualisation interactive des données, modélisée par des graphes. A partir de matrices de données, nous sommes capable de représenter nos données, afin de faciliter l'interprétation et l'exploration des données.

Les solutions proposées doivent cependant faire face à des problèmes, tels que le volume de données (taille des matrices), le choix du mode de représentation et les moyens d'exploration et de navigation, mis à la disposition de l'utilisateur. Pour résoudre ces problèmes, notre approche s'appuie sur plusieurs techniques issues de la théorie des graphes, partitionnement de graphes, dessin de graphes, analyse relationnelle et visualisation d'informations.

L'objectif de l'analyse des données relationnelles et de leur visualisation est la mise en évidence d'éléments stratégiques difficilement détectables à la main. Par le biais d'outils spécifiques, l'utilisateur doit pouvoir identifier les informations statiques les plus importantes, mais il doit surtout appréhender l'évolution dans le temps des structures existantes et anticiper l'émergence de nouvelles organisations à forte potentialité.

L'approche proposée est le morphing de graphe, principe qui consiste à animer graphiquement les données entre différentes instances tout en faisant ressortir visuellement les éléments majeurs et les tendances significatives. Notre première ébauche permet la mise en évidence des structures sous jacentes et de l'évolution des données au cours du temps. Cependant, elle nécessiterait d'être davantage développée.

L'objectif serait de réaliser un morphing complet assimilable à une séquence vidéo, mettant réellement les éléments importants en évidence, ainsi que les changements encourus par nos données au cours du temps. Pour cela, il nous faut améliorer notre algorithme en améliorant le dessin de graphe pour chaque instance et en fluidifiant le passage d'une instance à une autre, tout en mettant davantage en évidence les évolutions des données mais aussi les éléments importants.

6. Bibliographie

- [1] BRANDES U., CORMAN S. *Visual unrolling of network evolution and the analysis of dynamic discourse*. InfoVis'02 Vol. 2, N°1, 40-50, 2003.
- [2] CHEN C., CARR L. *Visualizing the evolution of subject domain: A case study*. InfoVis'99, IEEE Computer Society Press. pp. 449-452, 1999.
- [3] KAROUACH S., DOUSSET B. *Analyse d'information relationnelle par des graphes interactifs de grandes tailles*. EGC'04, Clermont Ferrand, janvier 2004.
- [4] JAKOBIAK F, "Maîtriser l'information critique", Editions d'Organisation, Paris, 1988.
- [5] DOUSSET B., BENJAMAA T., « Trilogie logiciel d'analyse de données », Conférence sur les systèmes d'informations élaborées : Bibliométrie – Information Stratégique – Veille technologique, 1988.
- [6] DKAKI T., « Outils mathématiques et méthodes automatiques pour la veille technologique », thèse de Doctorat, Université Paul Sabatier – IRIT, France, Août 1993.
- [7] CHRISMENT C., DKAKI T., MOTHE J., DOUSSET B, "Extraction et synthèse de connaissances à partir de bases de données hétérogènes", Ingénierie des Systèmes d'Information, vol. 5, n°3, 1997, pp. 367-400.
- [8] BERTIN J., « La sémiologie graphique: les diagrammes, les réseaux et les cartes », Editions Gauthier Villars, Paris, 1967, p.8.
- [9] EADES P., *A heuristic for Graph Drawing*. Congressus Numerantium, vol. 42, pp. 149-160, 1984.