

## ***MANIPULATION DE GRAPHES DE GRANDE TAILLE POUR L'ETUDE DES RESEAUX D'ACTEURS ET DES RESEAUX SEMANTIQUES.***

---

**Bernard DOUSSET, Saïd KAROUACH**  
[dousset@irit.fr](mailto:dousset@irit.fr), [karouach@irit.fr](mailto:karouach@irit.fr)

Institut de Recherche en Informatique de Toulouse, Equipe Systèmes d'Information Généralisés,  
Université Paul Sabatier- 31077 Toulouse cedex 9 (France).

---

**Mots-clés :** veille stratégique, Hyper espace, dessin de graphes, morphing, analyse relationnelle, ergonomie, Tétralogie

### **Résumé :**

La veille stratégique consiste à surveiller les sources d'information les plus diverses (bases de données bibliométriques, brevets, presse, news, agences, internet, bases internes, ...), à croiser les informations qu'elles contiennent pour en extraire une synthèse non accessible directement par la lecture parce-que non explicitée par les auteurs eux-mêmes. Cette information souvent stratégique, pour ne pas dire sensible, revêt plusieurs formes, mais la plus parlante est celle qui peut être matérialisée par la visualisation des réseaux : entre acteurs du domaine (co-publications, co-citations, collaborations, prises de participation, accords, alliances, co-dépôts de brevets, inventions communes, ...), ou entre éléments terminologiques ou concepts (co-occurrences entre termes, réseaux sémantiques). Une méthode naturelle de représentation de ces relations est le dessin de graphes dans lequel est optimisé le tracé (minimisation des croisements d'arêtes) et qui a recours à un certain nombre d'artifices visuels afin d'améliorer l'ergonomie de la représentation : couleurs, icônes, visualisation des valeurs, recherche assistée, classification, simplification, seuillages, interactivité.

Mais un graphe statique est le plus souvent trompeur car il montre une situation qui n'a jamais existé car elle est le résultat de la superposition de plusieurs périodes clés des relations pertinentes qui se sont succédées dans l'environnement du domaine analysé. Pour s'en convaincre, il suffit d'étudier une classe connexe (par exemple la plus grande) en la décomposant suivant plusieurs périodes, on s'aperçoit alors que la connexité n'est plus vérifiée pour chaque phase et donc qu'en fait la cohésion montrée par le graphe global n'a jamais existé.

D'où l'idée d'étudier l'évolution de ces graphes (d'acteurs ou sémantiques) par des techniques permettant d'en apprécier la cinématique et ce afin de comprendre les stratégies et les enjeux que cachent les restructurations successives du domaine cible. Pour cela, nous avons mis au point un outil de morphing qui simule une évolution continue passant par des étapes centrées sur des périodes prédéfinies. Il est alors possible d'apprécier les convergences et les divergences, les apparitions et les disparitions ainsi que les mobilités ou les essais de structuration. Les éléments clés apparaissent alors ainsi que l'histoire récente du domaine et la genèse de sa situation actuelle. Il est parfois même possible d'en déduire l'évolution future et donc la stratégie générale sous-jacente.

## 1. Introduction

L'outil de dessin de graphe que nous avons développé pour la plate-forme Tétralogie, était jusqu'à présent dédié aux graphes statiques représentés par des matrices carrées ou rectangulaires (graphes bipartis) correspondant à des métriques de proximité (présence-absence, contingence, co-occurrence, adjacence, ...). Plusieurs fonctionnalités étaient proposées dans cet outil :

- Dessin initial aléatoire ou circulaire (après tri par blocs diagonaux)
- Choix des icônes représentant les sommets (nuances, cercles ou histogrammes)
- Choix de l'intensité des liens (nuances) et affichage de leurs valeurs
- Affichage des noms des sommets (long, court, par survol)
- Dessin optimal par un algorithme de force paramétrable (attraction, répulsion)
- Possibilité d'intervenir interactivement dans le processus de dessin (déplacement ou verrouillage de sommets), choix de la taille de fenêtre, paramétrage des forces
- Tracé d'arbre partiel (maximum ou minimum)
- Classification stochastique des sommets avec choix de la densité
- Extraction des graphes des classes et optimisation de leurs dessins
- Dans chaque classe, visualisation des connecteurs directs dans les autres classes
- Graphe initial coloré par les classes avec un représentant fixé par classe
- Recherche sur les noms par visualisation fish eye de la liste alphabétique
- Graphe transitif à partir d'un sommet quelconque
- Pour les graphes bipartis (2 ensembles distincts de sommets)
  - 2 colorations
  - tracé initial suivant deux cercles concentriques
  - héritage d'une classification conjointe des lignes et des colonnes
- Pour les graphes sémantiques
  - Choix initial de la métrique (relative, absolue, équivalence, ...)
  - Méthodes de seuillage et de masquage

Les applications étaient diverses : représentation de réseaux d'acteurs de même type (auteurs, inventeurs, déposants, équipes de recherche, organismes, pays, sites web, ...) ou croisant deux types différents d'acteurs (auteurs - pays, inventeurs - déposants, équipes de recherche - organismes, sites web citants – sites web cités, ...), réseaux sémantiques (mots clés, index, multi-termes, concepts, ...) ou graphes mixtes croisant acteurs et sémantique. Mais la nécessité d'une étude dynamique est vite apparue, car les interprétations en veille stratégique ne peuvent absolument pas s'affranchir de la variable temps, elles peuvent même être totalement erronées.

## 2. L'outil VisuGraph existant

### 2.1. Liens entre graphes et matrices

Les tableaux de contingence ou de cooccurrence issus de croisements d'informations peuvent être considérés comme des matrices valuées attachées à des graphes non orientés. Si la matrice est symétrique, les lignes (ou les colonnes) représentent les sommets du graphe et les éléments de la matrices les arêtes (simultanément présence d'un lien et valeur de ce lien). Si la matrice est asymétrique, les lignes et les colonnes représentent deux type différents de sommets, le graphe est alors dit biparti (les arêtes du graphe ne relient, dans ce cas, que des sommets de types différents).

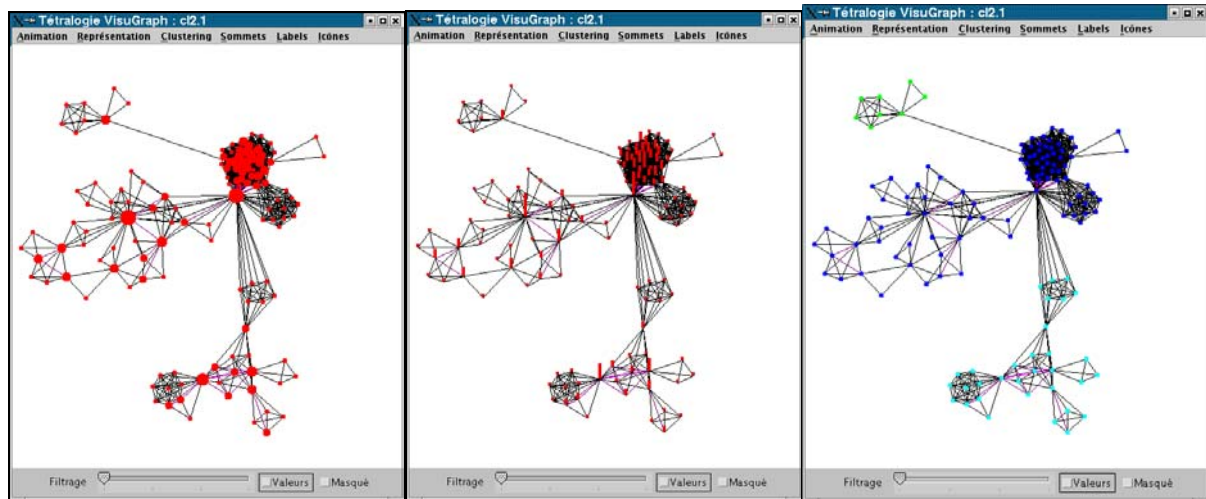
Actuellement, nous développons des méthodes de visualisation interactive de graphes de grande taille basées sur la notion d'attraction et de répulsion entre les sommets et qui aboutissent à un dessin optimisé aussi lisible que possible. Cette méthode de restitution est très appréciée par les utilisateurs de notre plate-forme de veille Tétralogie, car elle ne nécessite pas de connaissance mathématique particulière [KARO]. Le tout est de proposer un graphe à la fois fidèle à la réalité, suffisamment lisible et autorisant des manipulations interactives très nombreuses axées sur la découverte de relations pertinentes et la mise en relief de stratégies. Nous avons testé plusieurs modes de représentations graphiques qui permettent d'apprécier à la fois l'importance de chaque sommet et de chaque connexion (arête) dans la structure visualisée. Pour les sommets, il est possible de traduire leur

importance (valeur diagonale ou fréquence, norme de leur ligne ou de leur colonne, ...), pour les arêtes leur module obtenu le plus souvent après une transformation adaptée au cas considéré avec, éventuellement, l'application d'un seuil.

Voici quelques exemples de ces représentations extraits d'analyses stratégiques récentes.

## 2.2. Codage des sommets

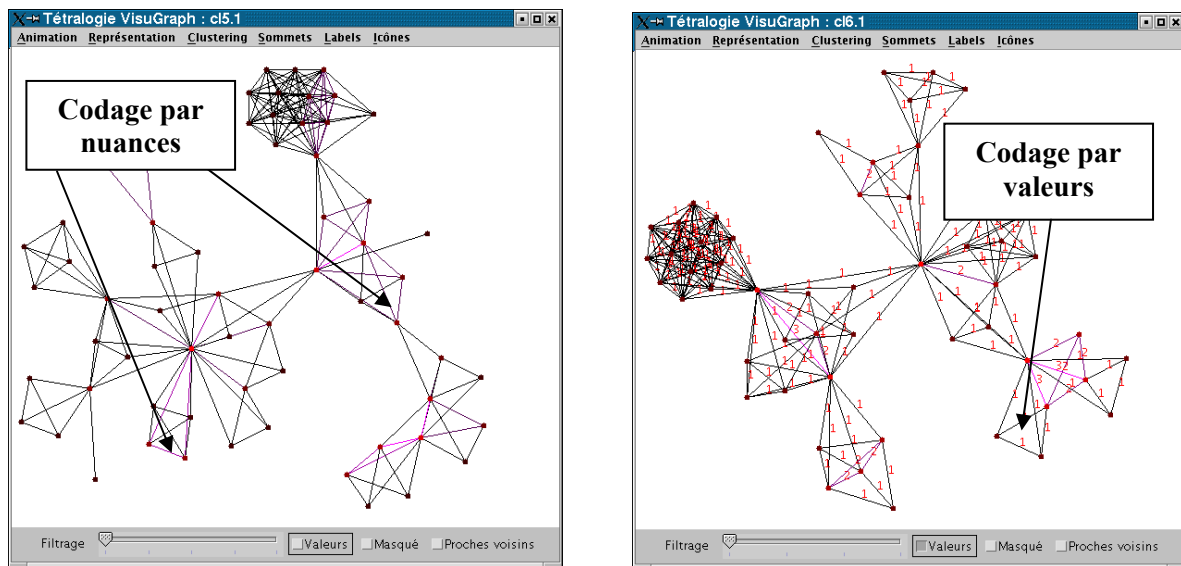
Pour traduire les échelles de valeurs, nous utilisons des points colorés suivant un nuancier, des barres verticales ou de cercles de taille variable. Pour visualiser l'appartenance à des classes prédéfinies nous utilisons une palette de couleurs.



*Figure 1. : Tailles des sommets ( icônes circulaires ou barres), classes (couleurs)*

## 2.3. Codage des arêtes

Les échelles de valeurs sont ici traduites, soit par des nuances d'une couleur, soit par l'épaisseur. Il est toujours possible d'indiquer la valeur d'une arête en son milieu.



*Figure 2. : Codage de la taille des arêtes par des nuances (et valeurs).*

## 2.4. Extension du graphe depuis un sommet

Afin de réaliser une navigation locale dans le graphe initial souvent trop complexe, il est possible de travailler sur un sous-graphe. Pour cela, nous partons d'un sommet particulier choisi dans une liste alphabétique et nous étendons progressivement le graphe, par transitivité, depuis ce sommet. Cette technique permet, par un changement de focus, de nous concentrer sur un extrait pertinent issu d'une information ciblée (acteur, mot-clé, concept).

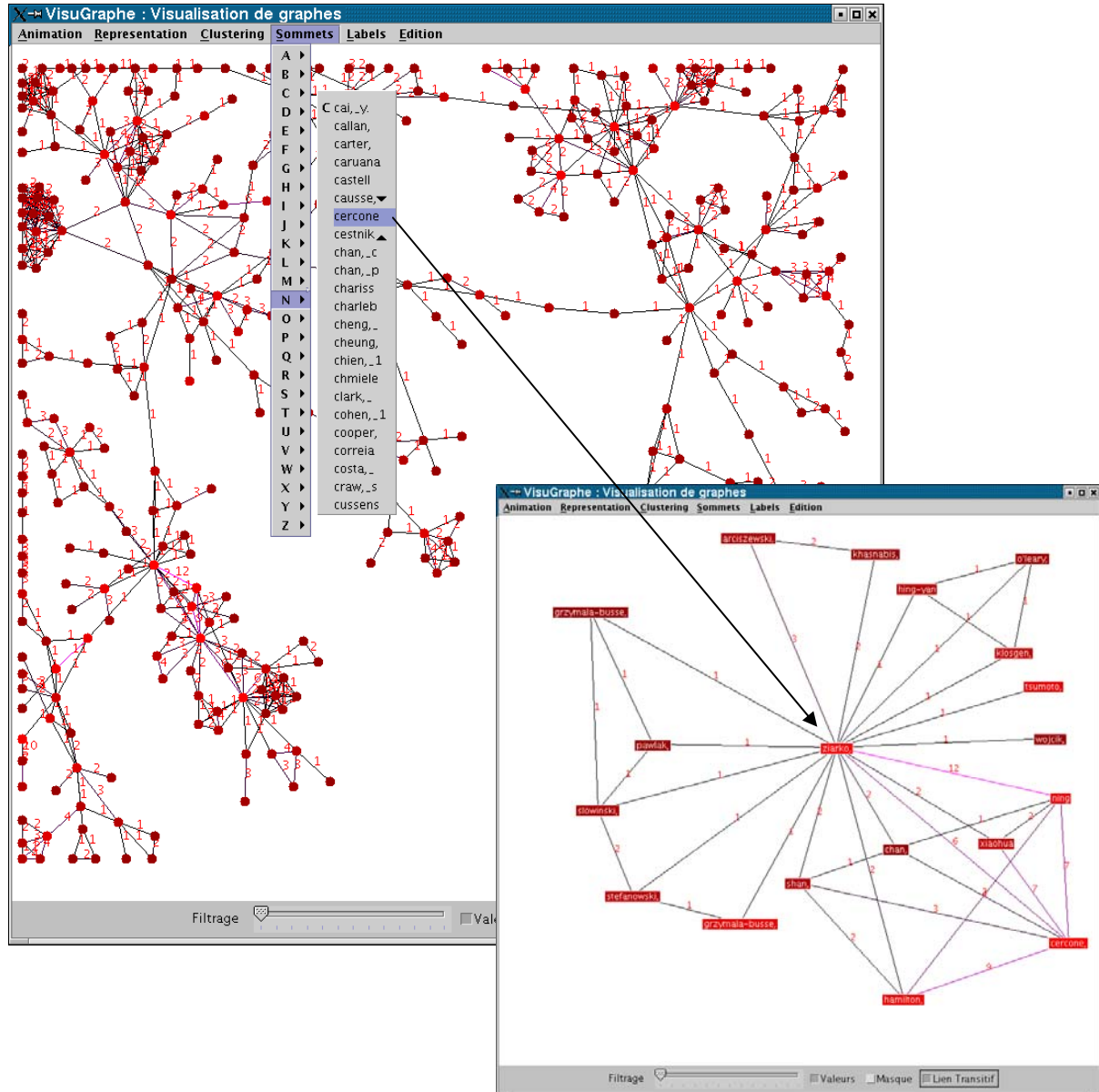


Figure 3. : Extension d'un sous graphe par transitivité depuis un sommet sélectionné.

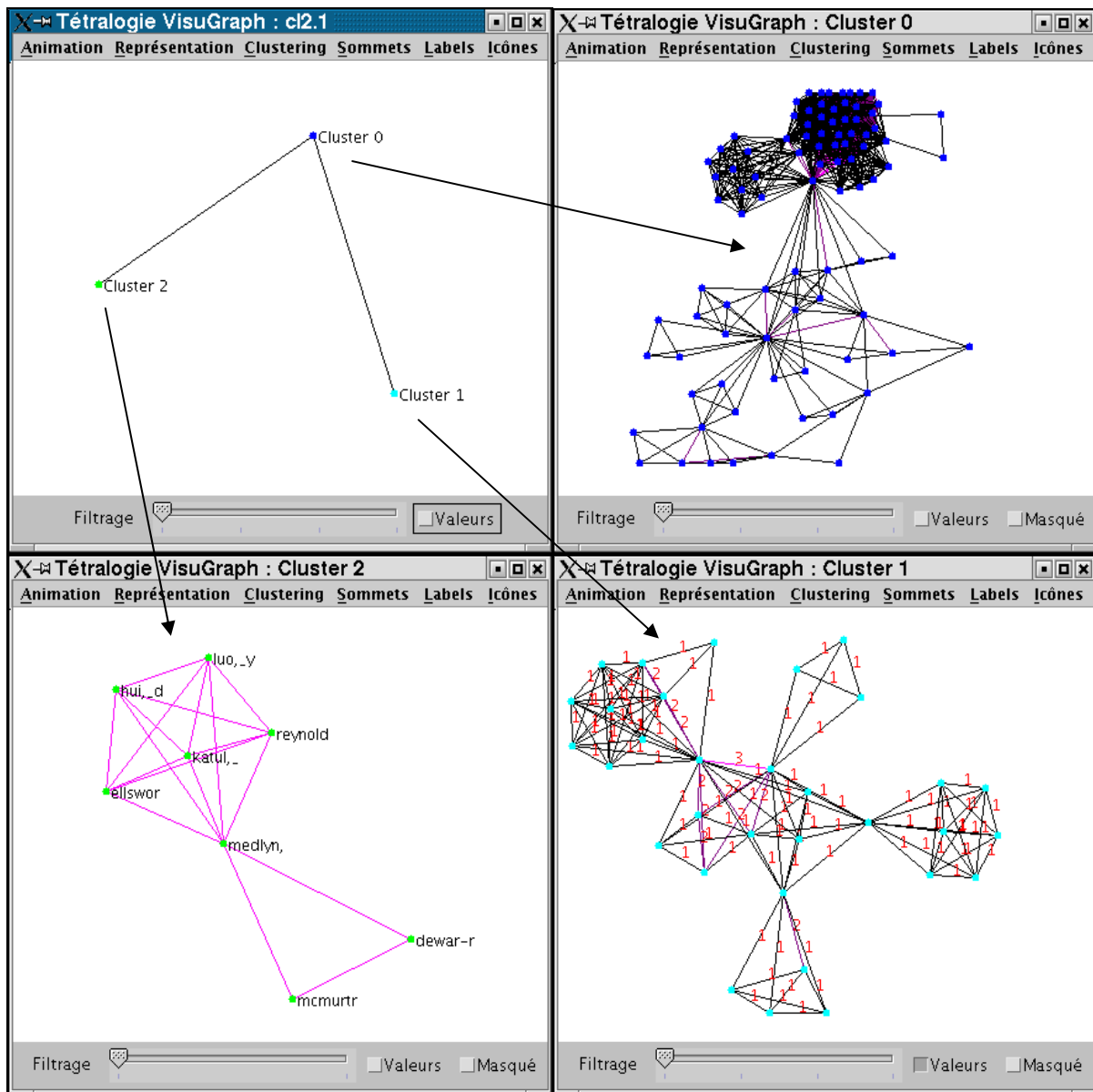
## 2.5. Classification, graphe réduit et extraction des classes

Une autre méthode pour contourner la complexité de certaines structures est de travailler sur le graphe des classes. Il s'agit alors de décomposer le graphe initial en classes homogènes, afin de regrouper les sommets les plus fortement connectés, nous utilisons pour cela la méthode de Markov Clustering. Un graphe réduit est alors obtenu : ses sommets sont définis par les classes elles-mêmes et ses arêtes par les connexions entre classes. Une fois ce nouveau graphe généré et étudié, il est possible de revenir au graphe initial en gardant visible l'appartenance de chaque sommet à sa classe grâce à une coloration

identique. Cette technique permet donc une approche macroscopique de la structure avec la possibilité de zoomer si nécessaire sur certains détails utiles.

A partir du graphe réduit ainsi obtenu, nous avons la possibilité de visualiser séparément chacune de ces classes afin de réaliser une analyse détaillée de leur structure et d'en comprendre leurs mécanismes locaux. Chaque classe est alors considérée comme un nouveau graphe et bénéficie de l'ensemble des fonctions interactives vues plus haut. Comme le nombre de sommets est maintenant réduit, il est beaucoup plus facile de détecter des stratégies et d'identifier des éléments importants tels que les connecteurs, les leaders, les sous groupes.

Dans l'exemple ci-dessous, nous avons décomposé le graphe connexe en 3 classes et nous avons, à partir du graphe réduit, demandé l'affichage séparé de chaque classe. Les sommets gardent la couleur qui leur a été attribuée et en leur appliquant une nouvelle migration il est possible de simplifier encore le dessin de ce graphe. Ici, le graphe réduit permet de garder en mémoire la structure initiale : les clusters 1 et 2 ne sont pas connectés, le cluster 0 est donc le cluster central.



*Figure 4. : Graphe réduit et graphes de ses classes.*

### 3. Le passage aux graphes dynamiques : matrices 3D

Les tableaux de présence absence, de contingence ou de cooccurrence issus de croisements d'informations peuvent être considérés comme des matrices valuées attachées à des graphes non orientés. Si la matrice est symétrique, les lignes (ou les colonnes) représentent les sommets du graphe et les éléments de la matrice les arêtes (simultanément présence d'un lien et valeur de ce lien). Si la matrice est asymétrique, les lignes et les colonnes représentent deux type différents de sommets, le graphe est alors dit biparti (les arêtes du graphe ne relient, dans ce cas, que des sommets de types différents). Pour introduire la notion d'évolution, nous croisons le tout avec la variable temps correctement discrétisée en périodes homogènes (plus en quantité qu'en découpage régulier : les volumes relationnels doivent rester assez comparables). Nous présentons ci-dessous l'outil de génération de ces matrices (issu de Tétralogie) avec le paramétrage nécessaire à l'obtention de matrices 3D compatibles avec notre outil de dessin de graphes évolutifs.

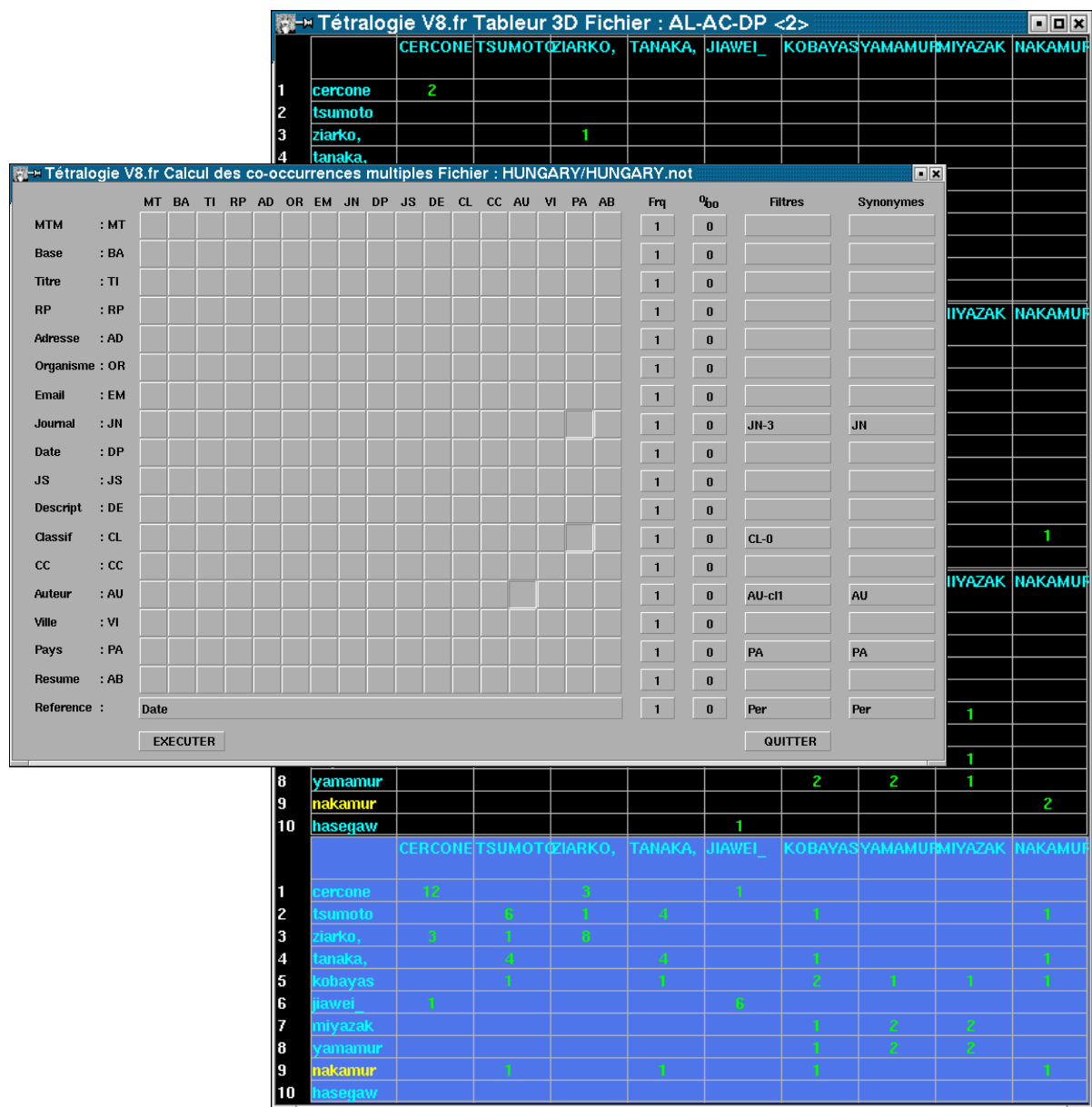


Figure 5. : Génération des matrices 3D (graphes évolutifs).



## 5. Les dessins à chaque étape de l'évolution

Afin de rendre compte de l'évolution du graphe d'une période à l'autre, nous ne visualisons au départ que le graphe correspondant à un plan de la matrice 3D. Les sommets non présents dans la période correspondante peuvent ou non être visualisés dans la position qu'ils occupent dans le graphe global. Le passage contrôlé d'une période à l'autre permet d'identifier les évolutions remarquables (apparitions, disparitions, maintiens, passage par un maximum ou un minimum, liens sporadiques, ...). La juxtaposition des dessins, comme ci-dessous, permet d'appréhender l'évolution globale et conduit ainsi à une interprétation plus fine des stratégies sous-jacentes. Des rapprochements entre équipes sont mis en évidence, des fins de collaborations, des mouvement de chercheurs. Dans l'exemple que nous proposons, il est évident que le graphe global ne permet pas l'extraction d'une information aussi fine, pire il induit une fausse collaboration générale qui n'a en fait jamais existé et qui peut amener à conclure que l'on est en présence d'un groupe assez homogène, connexe et persistant alors qu'il n'en est rien. La dernière période montre un éclatement du groupe en 10 classes connexes différentes, l'équipe principale est en légère chute, par contre une équipe assez bien structurée semble se dessiner à droite à mi-hauteur et, en bas à gauche, l'équipe très liée bien qu'en léger déclin sur le nombre de publications a l'air de renforcer ses relations avec deux équipes voisines.

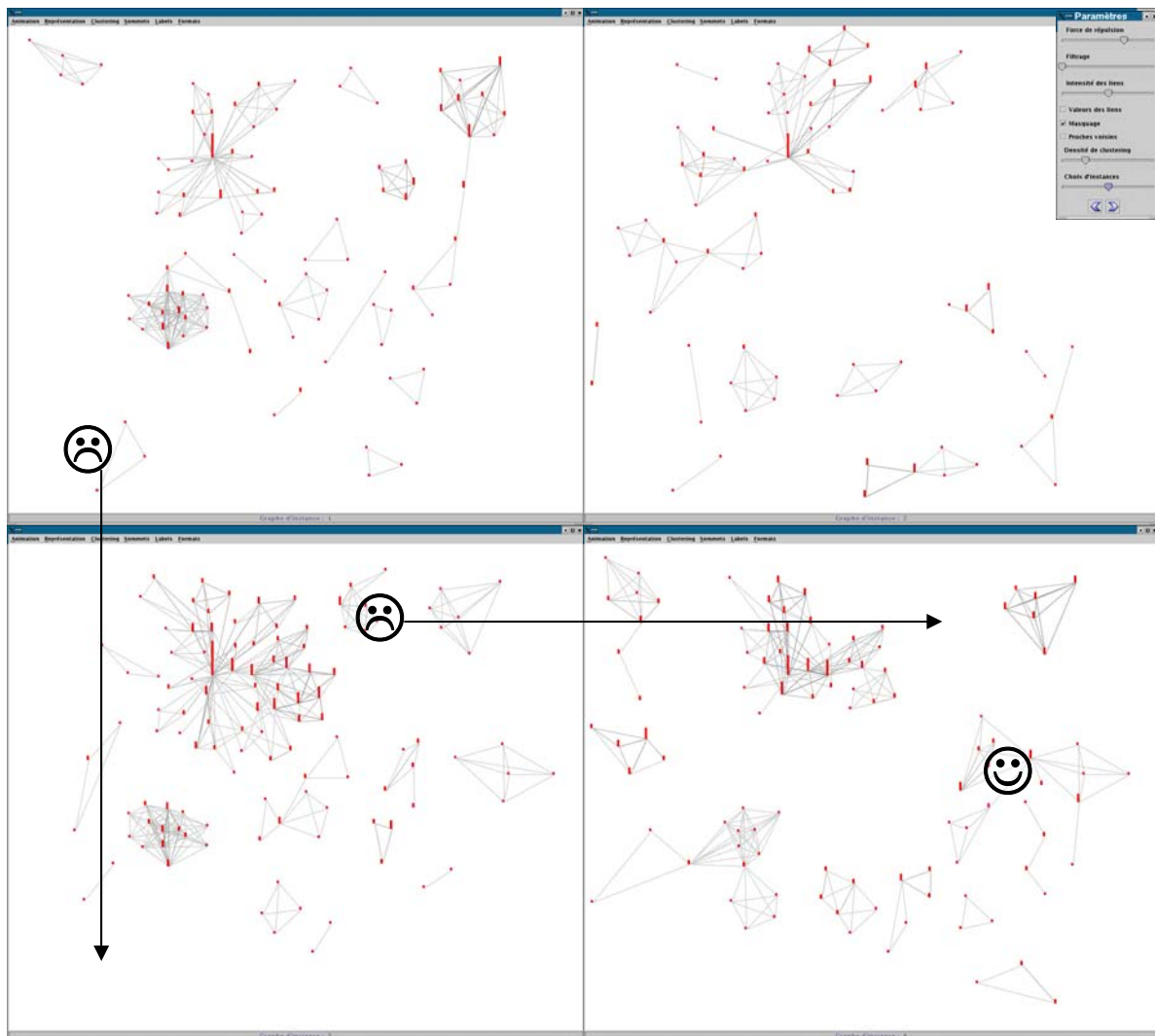


Figure 7. : Placement des items en fonction d'un tri par blocs diagonaux.

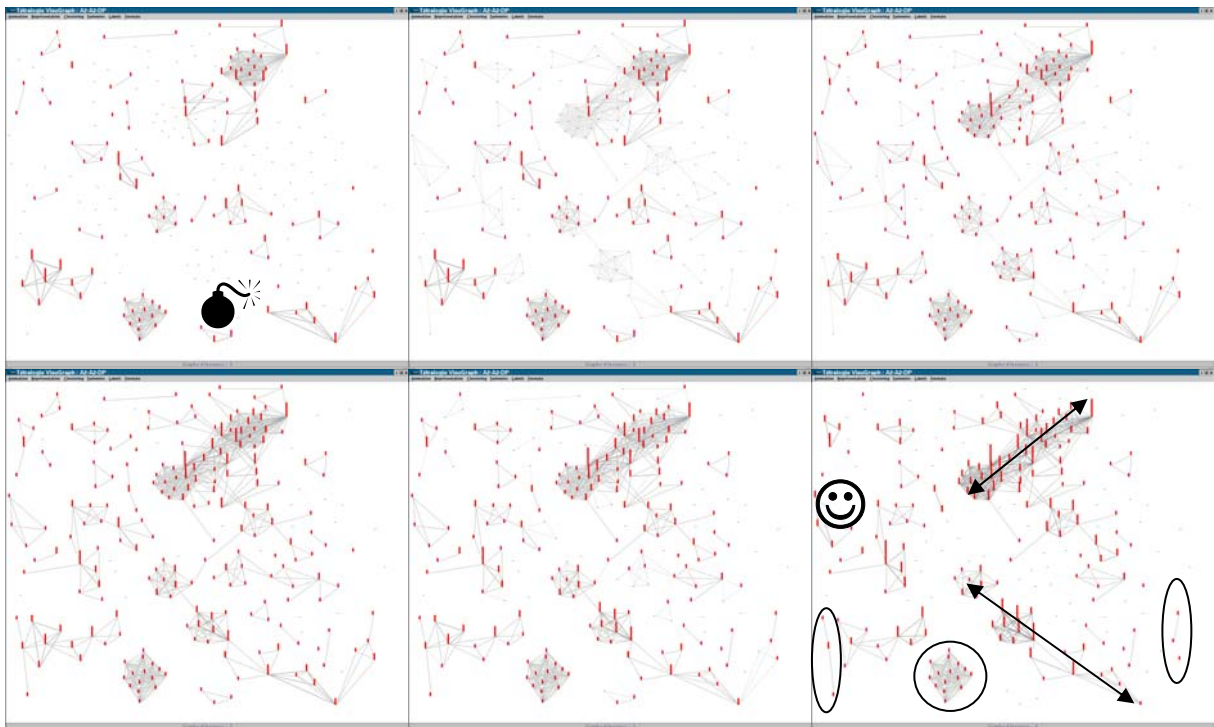


## 6. Morphing entre deux périodes

Afin de mieux comprendre l'évolution des relations entre les sommets, nous avons développé une technique de morphing qui dans un premier temps lisse le passage entre deux périodes adjacentes. Des liens apparaissent progressivement, d'autres disparaissent, certains se renforcent ou s'atténuent, les histogrammes représentant l'activité de chaque sommet subissent eux aussi des modifications (augmentation, diminution, apparition, disparition). En faisant plusieurs fois l'aller-retour entre deux périodes, il est possible de mieux comprendre l'évolution globale et de remarquer certains détails utiles pour l'étude stratégique. Un balayage de l'ensemble des périodes est aussi recommandé pour avoir une vue d'ensemble de l'évolution et ce afin de reconstruire l'historique du domaine et identifier les faits marquants qui ont conduit à la situation actuelle.

Dans le morphing ci-dessous, plusieurs choses sont à relever :

- En haut au centre, nous voyons fusionner plusieurs équipes qui captent simultanément de nouveaux acteurs, par contre l'équipe tout en haut est perdue.
- En bas au centre, une équipe (de 15 auteurs) est totalement stable alors qu'une autre (de 3 auteurs) disparaît complètement.
- Dans l'équipe en bas à droite, nous constatons deux événements diamétralement opposés : l'auteur placé à gauche de cette équipe se détache et forme sa propre équipe avec 2 auteurs nouveaux, l'auteur placé à droite forme une autre équipe tout en restant connecté et c'est cette équipe qui devient leader du groupe ainsi formé.
- A gauche, apparition d'une nouvelle équipe (de 4 personnes) et restructuration de plusieurs équipes autour d'un nouveau leader.
- En bas à droite, une nouvelle équipe apparaît, elle fusionne avec des équipes existantes et devient leader, mais 3 auteurs se détachent et forment, tout à droite, une nouvelle équipe.



*Figure 8. : Passage par morphing d'une période à l'autre.*

Ces spéculations ne sont bien entendu que des hypothèses de travail, elles demandent à être confirmées, mais le rôle de la veille est avant tout de surveiller l'évolution d'un domaine, nous avons simplement identifié des phénomènes à surveiller. Le futur viendra confirmer ou infirmer nos hypothèses, mais notre expérience montre que ce sont très souvent les bonnes.

## 7. Conclusion

Nous avons appliqué cette nouvelle technique à de nombreux exemples : évolution des co-signatures de publications scientifiques, de co-inventions ou de co-dépôts de brevets, de collaborations internationales, de graphes sémantiques (mots clés, multi-termes indexant le texte libre, concepts) . A chaque fois, de nouvelles interprétations étaient possibles qui tennaient compte non seulement de la structure générale mais aussi de son évolution récente. Dans certains cas même, les interprétations initiales étaient infirmées par cette nouvelle approche : collaborations arrêtées, sujets perdus ou transformés, divergences récentes, associations temporaires, signaux émergents et leur contexte.

Il reste malgré tout des améliorations possibles pour cet outil. En premier lieu, trouver un algorithme de placement des sommets qui tienne compte des non connexités des graphes partiels (par période) afin de faire ressortir les liens passagers qui relie des classes bien distinctes au cours du temps. Essayer de placer les sommets en fonction de la distribution de leur présence dans les différentes périodes (effet d'horloge). Ensuite, améliorer le morphing de graphe et le rendre paramétrable (choix des fonctions de transition, lissage de l'évolution), générer un déplacement des sommets afin de simuler rapprochements et déconnexions, extrapoler des positions ou des structures futures.

## 8. Bibliographie

- [CARM] Liran Carmel, David Harel, Yehuda Koren: Combining Hierarchy and Energy Drawing Directed Graphs. IEEE Transactions on Visualisation and Computer Graphics 10 (1), 46-57 (2004)
- [DOUS] B. Dousset. *Visualisation interactive d'informations endogènes extraites de grands corpus hétérogènes*. Journée visualisation de l'information INIST-INRIA-LORIA (Nancy France), 19 février 2003.
- [GIME0] R. Gimeno, P. Mitrano. *Théorie cartographique et sémiologie graphique*. [http://www.sciences-po.fr/cartographie/cartographie\\_html](http://www.sciences-po.fr/cartographie/cartographie_html), 2001.
- [KARO] S. Karouach. *Système de visualisations interactives pour la découverte de connaissances*. Thèse en informatique de l'Université Paul Sabatier, (Toulouse III France), juillet 2003.
- [KORE] Yehuda Koren: Graph Drawing by Subspace Optimization. IEEE TVGC Symposium on Visualization, 65-74, (2004)
- [MARS] M. Marshall. *Methods and tools for the visualization and navigation of graphs*. Thèse, Département de mathématiques et d'informatique, Université de Bordeaux, (France), juin 2001.
- [MEGH] G. Meghabghab. *Discovering authorities and hubs in different topological Web graph structures*. Information Processing & Management, vol. 38 (1), pp 111-140, 2002.
- [PATR] M. Patrignani. *Visualization of large graphs*. Thèse de doctorat, Università di Roma « La Sapienza », 2001.
- [VAN D] S. van Dongen. *Graph Clustering by Flow Simulation*. Thèse de l'université d'Utrecht, Allemagne, May 2000.