

Analyse interactive de l'évolution des réseaux par morphing de graphes

Saïd KAROUACH *, Christophe LONGEVIALLE **

karouach@irit.fr, longevialle.christophe@wanadoo.fr

(*) Université Paul Sabatier - IRIT - Equipe SIG
118, route de Narbonne - 31062 Toulouse cedex 4

(**) Université Marne La Vallée - I.F.I.S/ C.E.S.D
Cité Descartes - Champs sur Marne - 77454 Marne la Vallée Cedex 2

Mots-clés :

Analyse réseaux, réseaux sémantiques,, analyse de l'évolution, visualisation de graphes, veille stratégique.

Keywords :

Network analysis, semantic networks, evolution analysis, graph visualization, strategic watch.

Résumé :

La recherche d'informations stratégiques conduit très souvent à utiliser les graphes comme outil de représentation synthétique et intuitif des réseaux d'acteurs ou des réseaux sémantiques. De nombreuses méthodes de dessin de graphes sont aujourd'hui proposées, elles permettent d'obtenir un dessin optimisé qui minimise les croisements des arêtes tout en harmonisant le placement des sommets. Ces graphes permettent de facilement repérer des regroupements, des connecteurs entre les groupes ainsi que des structures révélatrices de stratégies globales : collaborations, ententes, prises de contrôle, encerclement, lobbying,... Mais le plus souvent, une étude en dynamique est nécessaire pour bien comprendre la mise en place de ces stratégies ou pour aider à leur détection précoce. Nous proposons, dans cet article, une méthode de dessin de graphes évolutifs basée sur la notion de morphing. Le point de départ est une matrice d'adjacence (binaire ou valuée) en 3D, « la troisième dimension représentant la variable temps » décomposée en périodes suffisamment homogènes. Le graphe global (toutes périodes confondues) est ensuite dessiné et son dessin est optimisé. Dans une seconde phase, nous dessinons les graphes partiels correspondants aux différentes périodes en prenant le graphe global comme point de départ. Enfin, par morphing nous proposons à l'utilisateur de passer d'une période à l'autre afin qu'il puisse visualiser au mieux les évolutions de la structure, les positions relatives des sommets et la force des liens qui les unissent. D'autres types de structures et de codages sont proposés, afin de s'adapter à tous les cas rencontrés (graphes binaires, graphes valués, graphes bipartis, graphes partiels non connexes,).

Erten et al [5] présente le système TGRIP, comme extension de GRIP [6], d'analyse visuelle de l'évolution de collaborations entre chercheurs d'un domaine donné. TGRIP produit une série de représentations 2D (cf. figure 2), une pour chaque période, tout en préservant la carte mentale 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 par la technique des forces.

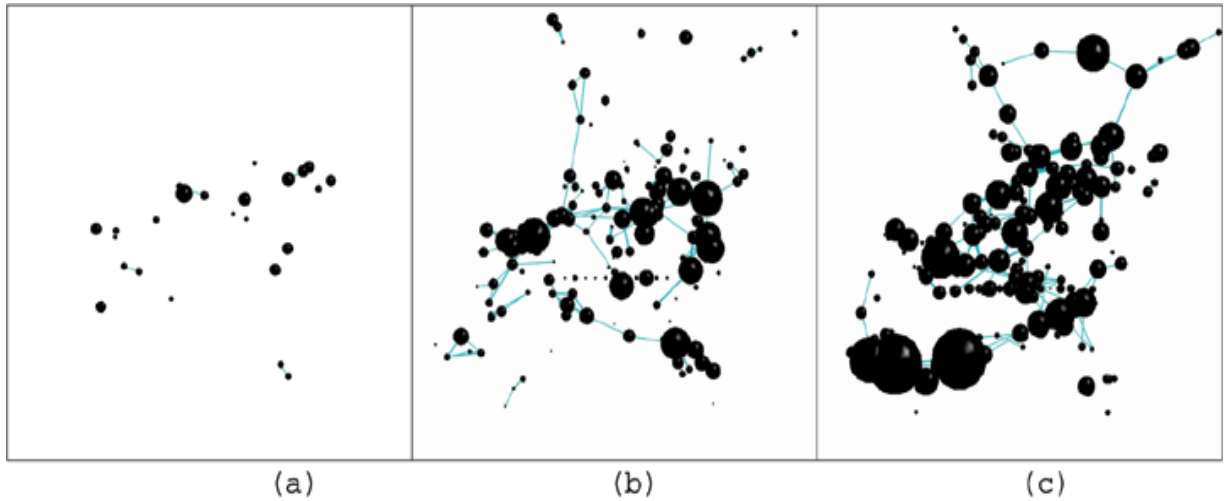


Figure 2 : Représentation 2D de l'évolution de collaborations par TGRIP : Erten et al [5]

Une représentation 3D du même graphe global peut également être visualisée où

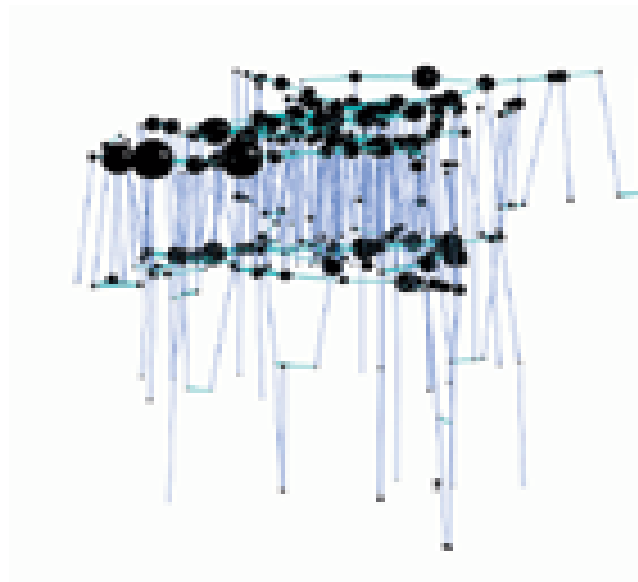


Figure 3 : Représentation 3D de l'évolution de collaborations par TGRIP : Erten et al [5]

Collberg et al [3] décrit le système GEVOL pour la visualisation de l'évolution de logiciel. Ce système utilise une modification de l'algorithme de dessin de GRIP.

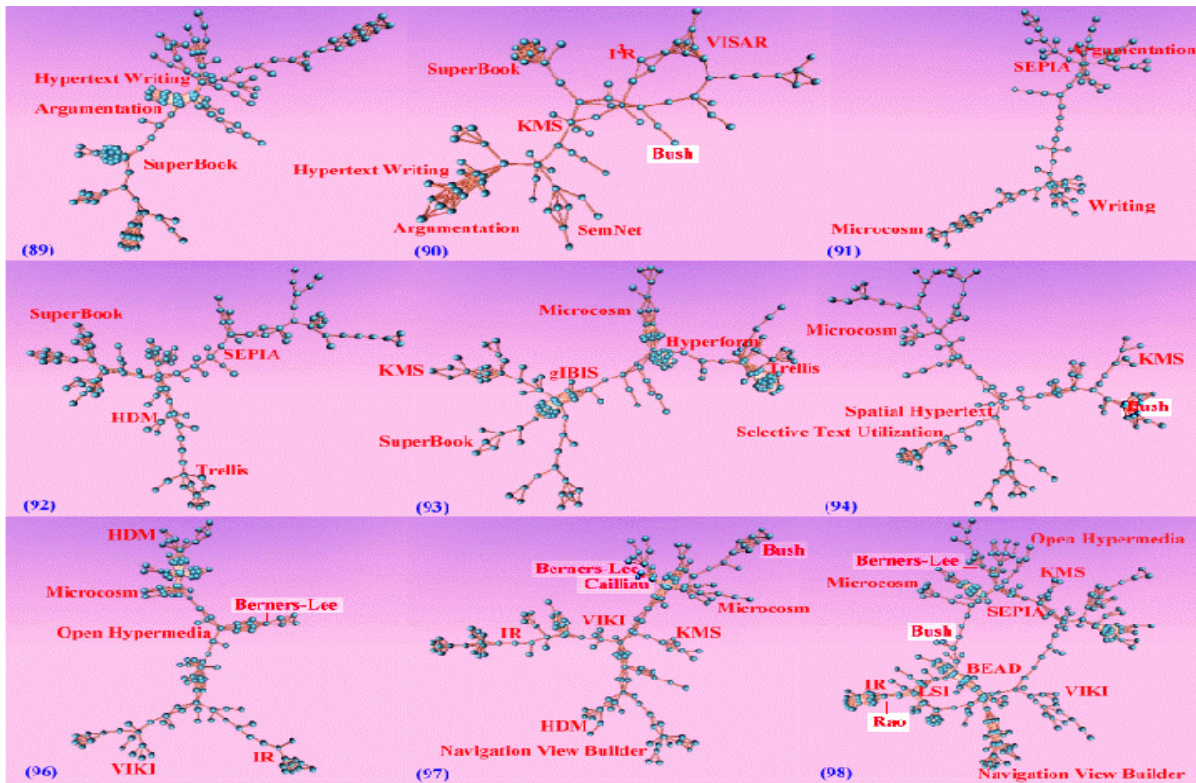


Figure 4 : Représentation séparée des réseaux évolutifs : Chen [2]

L'approche proposée par Chen [2] (cf. figure 4) consiste à visualiser séparément les réseaux pour chaque période. L'inconvénient majeure de cette approche est que l'utilisateur ne dispose pas de points de repères sur les différentes représentations. Autrement, 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.

3 Représentation de données évolutives

Les données relationnelles que nous traitons sont issues d'un processus de traitement d'information représentées sous formes matricielles tridimensionnelles dont la troisième dimension représente le temps. Il s'agit essentiellement de croisements entre des entités sur plusieurs segments temporels (ou périodes) homogènes. Afin de mieux illustrer notre démarche, nous considérons un exemple simple de croisement entre les entités {a, b, c, d} sur quatre périodes.

	a	b	c	d
a	2			1
b		1	1	
c		1	2	
d	1			3

Tableau 1 : période 1

	a	b	c	D
a	1			
b		3		1
c			1	
d		1		2

Tableau 2: période 2

	a	b	c	D
a	1		1	
b		1	1	
c	1	1	2	
d				2

Tableau 3: période 3

	a	b	c	d
a	5	1		2
b	1	3	1	
c		1	2	
d	2			2

Tableau 4: période 4

Notre démarche consiste alors à transformer ces données en une représentation sous forme de réseaux dont les sommets représentent $\{a, b, c, d\}$ et les liens définissent les relations entre elles. Ce type de données peut être représenté de deux manières en fonction des objectifs souhaités. Il est possible de définir un graphe pour chaque valeur de la troisième dimension [2].

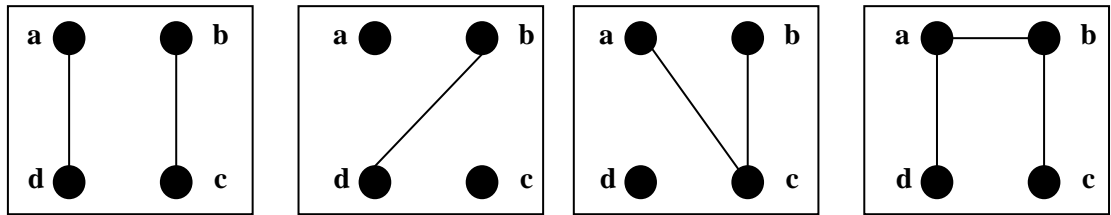


Figure 2 : Graphes associés aux différentes périodes

Dans la figure 2, chaque graphe représente l'ensemble des relations entre les entités $\{a, b, c, d\}$ pour la période correspondante. Cette solution ne permet d'analyser que les instances séparément les unes des autres et jamais combinées. Une autre approche consiste à construire le graphe « global » G_{1-n} , comme la combinaison des graphes G_1, G_2, \dots, G_n de chaque période. Le graphe global est associé à la matrice résultante des matrices de toutes les périodes (cf. tableau 5). Cette représentation a l'avantage d'avoir une vue globale de l'ensemble des données sur l'ensemble des périodes de l'analyse : positionnement global.

	a	b	c	d
a	9	1	1	3
b	1	8	3	1
c		3	7	
d	3	1		9

Tableau 5: Matrice globale

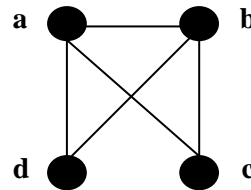


Figure 6: Matrice globale

Par contre, une telle représentation est pauvre en information et ne favorise pas l'analyse de l'évolution du réseau. D'où, la nécessité d'utiliser des artefacts visuels supplémentaires pour enrichir la représentation et pour caractériser l'évolution. En effet, l'utilisation d'artefacts comme la taille permet d'identifier visuellement des sommets potentiellement importants, par exemple ceux dont la taille est maximale.

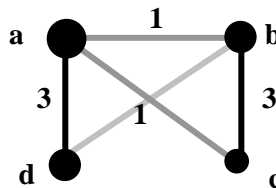


Figure 4 : Codage par la taille

4 Analyse évolutive d'information relationnelle

Une méthode de visualisation qui vise à analyser l'évolution d'un réseau sur différents segments temporels doit aborder les trois points suivants :

- Lisibilité des différents sous réseaux ;
- Identification des sommets potentiellement importants ;
- Mise en évidence de l'évolution.

4.1 Dessin de graphes

Afin que la représentation du réseau soit lisible pour être bien interprétée par l'utilisateur, nous utilisons, pour placer les sommets de façon optimale, la technique du dessin de graphes basée sur la notion des forces d'attraction et de répulsion entre sommets. Eades [4] comparait un graphe 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. Après quelques déplacements, le système finira par se stabiliser. L'algorithme d'Eades est basé sur ce principe de ressort, auquel il ajoute une notion de force de répulsion entre les sommets et une attraction matérialisée par les arêtes. Cette technique permet un dessin plus lisible en reflétant les structures macroscopiques du graphe analysé.

Dans le cas de données évolutives, notre soucis principal est d'obtenir des représentations qui soient à la fois lisibles et qui préservent au mieux la carte mentale de l'utilisateur, c'est à dire que les mêmes parties du graphe demeurent dans leurs positions initiales pour toutes les périodes. Après un dessin du graphe « global » à l'aide de la technique des forces qui met en évidence la morphologie des données sur toute les périodes, il est possible ensuite de visualiser les différents graphes de chaque période.

Le fait que les sommets de ces graphes ne sont pas repositionnés automatiquement permet à l'utilisateur de repérer facilement les liens qui apparaissent, qui disparaissent ou qui sont stables d'un graphe à l'autre. Bien entendu, nous mettons à la disposition de l'utilisateur des moyens interactifs pour repositionner automatiquement et/ou manuellement les sommets du graphe visualisé.

4.2 Codage de l'information

Dans un graphe, tous les éléments (sommets et liens) 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 indispensables 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) et la taille pour mettre en valeur ces caractéristiques. Ce codage d'information est basé sur la définition d'indices (métriques [8]). Ces indices sont définis à partir de la matrice $A = (a_{ij})$ associée au graphe sur une période donnée k . Soit $m_i^k = \sum_j a_{ij}^k$, la fonction qui associe à chaque sommet v_i

la quantité $M_i^k = \frac{m_i^k}{\max_j(m_j^k)}$ définit un indice de v_i relatif à k . Il représente l'importance du

sommet v_i dans la structure du graphe en terme de connexions et de taux d'information. Il est possible un indice absolu sur toutes les périodes de l'analyse par la série des indices relatifs.

m	t_1	t_2	t_3	t_4
a	3	1	2	8
b	2	4	2	5
c	3	1	4	3
d	4	3	2	4

Tableau 6 : Tableau des valeurs m_i

M	t_1	t_2	t_3	t_4
a	3/8	1/8	2/8	1
b	2/8	4/8	2/8	5/8
c	3/8	1/8	4/8	3/8
d	4/8	3/8	2/8	4/8

Tableau 7 : Tableau des valeurs M_i

Les tableaux ci-dessus donnent un récapitulatif des calculs des valeurs de m_i et M_i . Le maximum des m_i (en gras) est atteint par le sommet a à la période t_4 . Ces indices nous permettent plusieurs types de codages. Dans le cas relatif, l'indice d'un sommet sera codé soit par la taille sous forme de barre. Dans le cas absolu, nous représentons l'information sur toutes les périodes par un histogramme des indices relatifs. Ce type de codage permet facilement la comparaison visuelle de l'évolution de la dynamique d'un réseau. En effet, la taille de chaque barre est proportionnelle à l'indice associé au sommet et

chaque période est indiquée par une couleur particulière. Ainsi il est facile détecter visuellement les sommets particuliers :

- ceux dont la taille est maximale pour une période donnée
- ceux en croissance (ou décroissance) progressive, ...

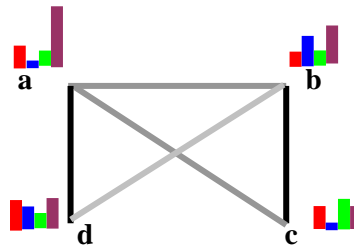


Figure 5 : Codage de l'évolution

5 Exemple :

Considérons la matrice tridimensionnelle représentant un croisement entre un ensemble d'entités pris sur 4 périodes.

Tetralogie TABLEUR											
	MEDAREX	ABGENIX	CAMBRID	GENPHAR	GENENTE	PROTEIN	GENZYME	XOMA	CELLTEC	ENZON	IMMUNOG IDEC
1	medarex	5									
2	abgenix										
3	cambrid		4								
4	genphar			32							
5	genente				7				2		
6	protein					7					
7	genzyme						10				
	MEDAREX	ABGENIX	CAMBRID	GENPHAR	GENENTE	PROTEIN	GENZYME	XOMA	CELLTEC	ENZON	IMMUNOG IDEC
1	medarex	9									
2	abgenix		4			2					
3	cambrid			2							
4	genphar				3						
5	genente					8		2			
6	protein						8				2
7	genzyme							5			
	MEDAREX	ABGENIX	CAMBRID	GENPHAR	GENENTE	PROTEIN	GENZYME	XOMA	CELLTEC	ENZON	IMMUNOG IDEC
1	medarex	43				2					
2	abgenix		34								
3	cambrid			27				2			
4	genphar										
5	genente				5						
6	protein	2				5					
7	genzyme		2				8				
	MEDAREX	ABGENIX	CAMBRID	GENPHAR	GENENTE	PROTEIN	GENZYME	XOMA	CELLTEC	ENZON	IMMUNOG IDEC
1	medarex	15									
2	abgenix		12								
3	cambrid			16				2			
4	genphar										
5	genente				8			2			
6	protein					6					
7	genzyme						2				

Figure 6: Tableau de cooccurrence à trois dimensions sur quatre périodes.

Une première représentation (cf. figure 7) de l'ensemble du réseau est donnée par le graphe global après convergence vers un dessin lisible. Chaque sommet est représenté par l'histogramme des 4 valeurs des indices relatifs pour chaque période schématisée par une couleur (rouge, bleu, vert et magenta).

L'analyse directe du graphe global permet de détecter les particularités de chaque sommet à chaque période et donne une bonne idée de l'évolution. Par exemple, la forte présence de certains sommets est rapidement détectée pour chaque période. Celle-ci est matérialisée par la taille de chacune des barres de l'histogramme. Globalement, le maximum est atteint sur la troisième période.

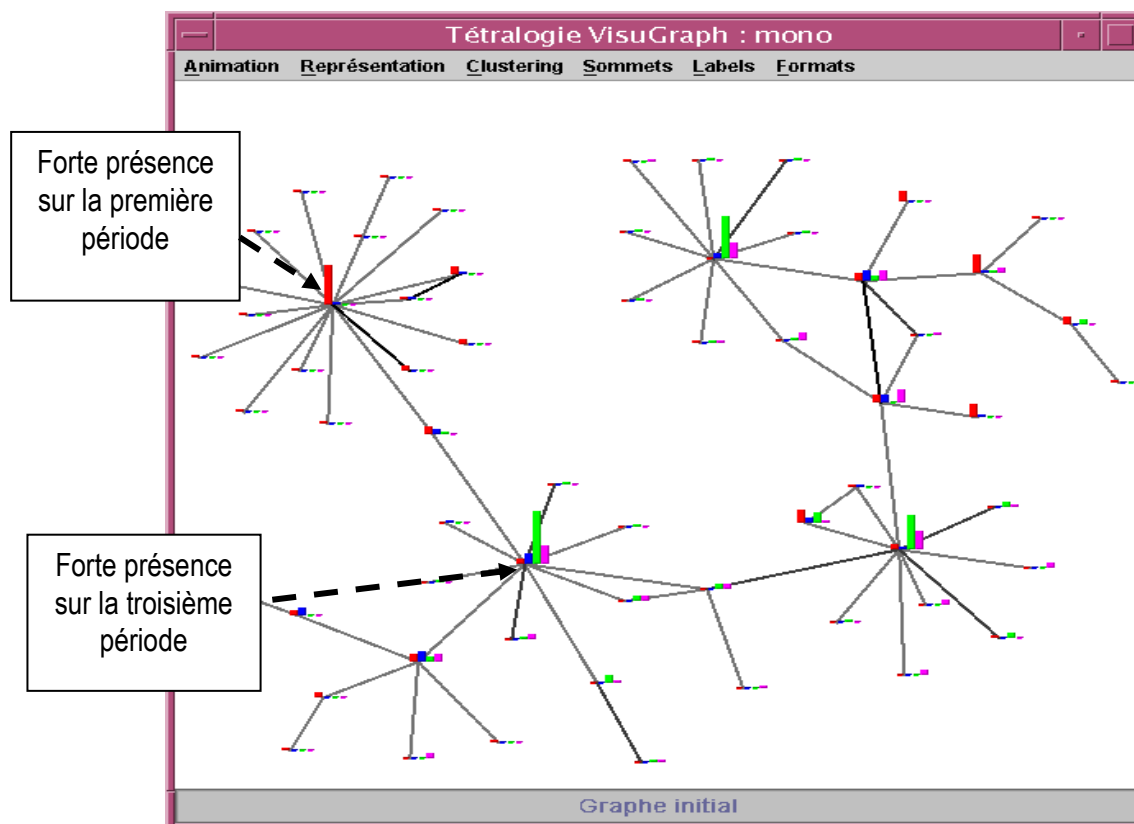


Figure 7 : Représentation du graphe global

La cinématique sur les quatre périodes rend bien mieux compte de l'évolution aussi bien générale que locale. Par contre, les apparitions et disparitions de certaines arêtes ne sont pas décelables sur cette représentation. Néanmoins, chaque arête du graphe global a un poids qui représente le cumul d'éventuelles arêtes de toutes les périodes. Ce poids permet d'attribuer un niveau de gris à chaque arête. Un tel codage permet la comparaison de la force des liens et ainsi il favorise la détection des liens forts.

Dans la figure 8, nous montrons l'état du graphe au niveau de chaque période ce qui nous permet de remarquer immédiatement les diverses phases d'évolution. Le morphing de graphe vient bien entendu renforcer cette impression de dynamique. Afin de mieux conserver la carte mentale de l'utilisateur, tous les sommets du graphe demeurent visibles et dans les mêmes positions pour tous les graphes.

Dans chaque graphe d'instance, la taille d'un sommet est recalculée en fonction des valeurs de la matrice associée à la période en cours. De la même manière, les niveaux de gris des arêtes sont recalculés en fonction des poids des arêtes.

Nous comptons, très prochainement, proposer une autre fonction de morphing de graphe qui positionne les sommets de chaque instance en tenant compte à la fois des liens absolus (somme de toutes les instances) et des liens instantanés (l'instance en cours). Nous pourrions ainsi visualiser, non seulement l'évolution des sommets et de liens, mais aussi l'évolution des positions relatives de chaque groupe. La migration des sommets sera alors calculée à partir d'un mixage paramétrable entre le

tableau absolu et chaque tableau relatif. Des mouvements de convergence ou de divergence seront alors perceptibles ce qui facilitera grandement l'interprétation.

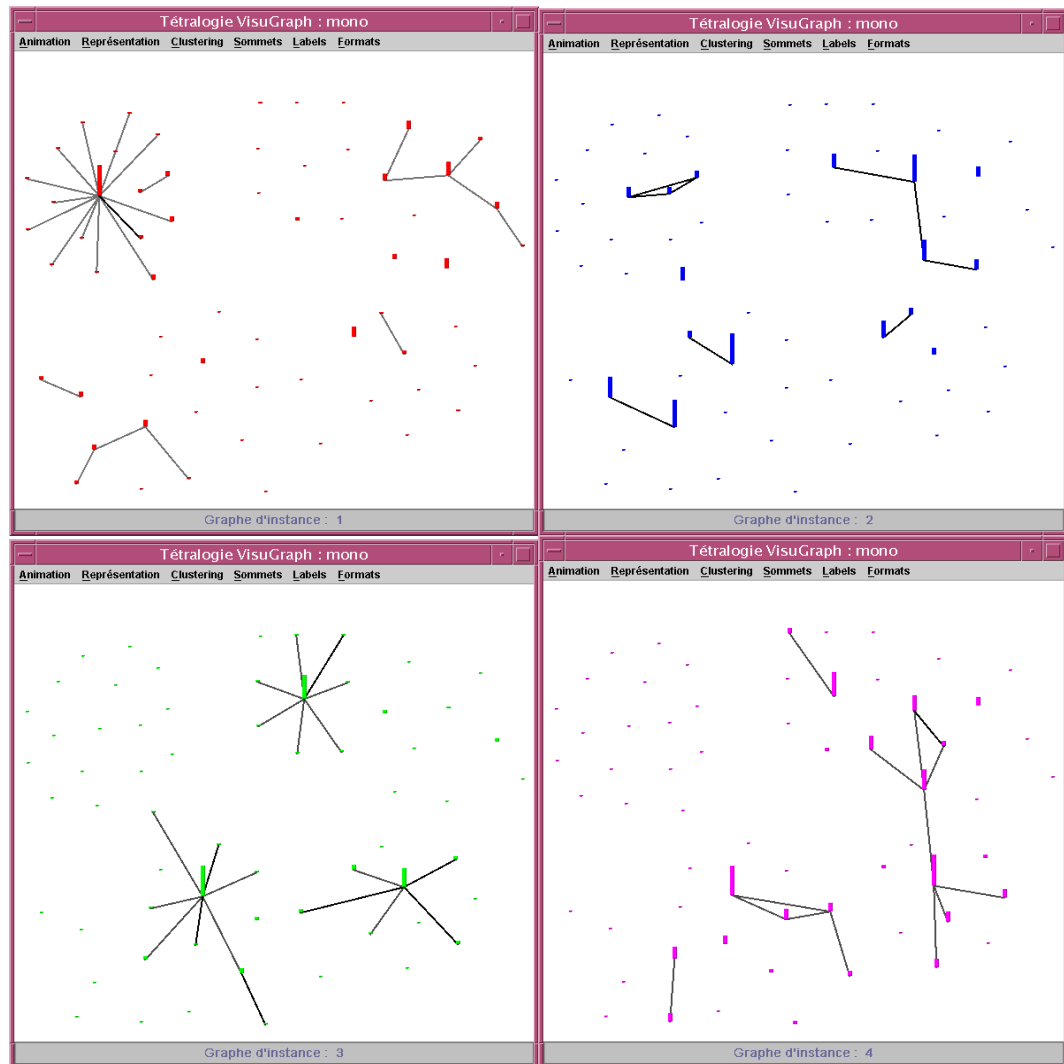


Figure 8: Comparaison des graphes d'instances

6 Conclusion et perspectives

Dans cet article, nous proposons d'intégrer à nos outils de veille stratégique essentiellement basés sur l'analyse des données (analyses factorielles, classifications, cartographies, ...) une approche graphique et plus intuitive de l'aspect relationnel inhérent à tout croisement d'information. L'étude des réseaux, qui représente l'essentiel de l'information endogène recherchée, est grandement facilitée par une représentation optimisée et interactive de leur structure. Ces fonctions, déjà intégrées dans notre prototype *VisuGraph*, permettent, aux utilisateurs non initiés, d'assimiler, sous une forme ergonomique et intuitive, l'ensemble des informations stratégiques. Cependant, dans un contexte de veille, la prise en compte du facteur temporel est essentiel pour une analyse complète. D'où la nécessité de travailler sur l'évolution du phénomène étudié. Nous proposons alors une approche de visualisation interactive pour l'analyse d'informations relationnelles et évolutives. L'aspect évolutif est mis en valeur par la définition d'indicateurs permettant des comparaisons visuelles des différentes instances de chaque éléments sur le réseau global. Quant aux graphes d'instances, ils consistent à fournir des informations complémentaires sur l'apparition et disparition de certains liens du réseau global. Le codage par la couleur et la taille permet à l'utilisateur de localiser les sommets

potentiellement importants et de caractériser la nature de l'évolution. Quant au dessin par la technique des forces, il permet d'identifier visuellement la morphologie structurelle du graphe analysé.

Dans le cas de données massives, le partitionnement est un moyen efficace pour réduire la taille des réseaux à visualiser. Après avoir découvert une partition du graphe étudié, nous pouvons réduire le nombre d'éléments à afficher en limitant notre vue aux groupes eux-mêmes. Cette représentation du graphe de classes (ou graphe réduit) diminue considérablement la complexité du dessin à base de forces. Dans *VisuGraph*, cette technique est déjà utilisée pour les graphes non évolutifs. A court terme, nous comptons l'étendre pour les graphes évolutifs. Dans ce cas, nous serons amenés à analyser l'évolution d'un groupe de sommets par rapport aux autres classes. Le codage d'information relative à chaque classe doit être judicieusement choisie afin de pouvoir identifier les changements le long du temps.

7 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.] COLLBERG C., KOBOUROV S. G., NAGRA J., PITTS J., and WAMPLER K. *A system for graph-based visualization of the evolution of software*. In ACM Symposium on Software Visualization, 2003.
- [4.] EADES P., *A heuristic for Graph Drawing*. Congressus Numerantium, vol. 42, pp. 149-160, 1984.
- [5.] ERTEN C., HARDING P. J., KOBOUROV S. G, WAMPLER K., and YEE G. *Exploring the computing literature using temporal graph visualization*, Conference on Visualization and Data Analysis, 2004.
- [6.] GAGER P.; KOBOUROV S. G., *GRIP: Graph Drawing with Intelligent Placement*, 8th Symposium on Graph Drawing (GD), p. 222-228, 2000.
- [7.] KAROUACH S., DOUSSET B. *Analyse d'information relationnelle par des graphes interactifs de grandes tailles*. EGC'04, Clermont Ferrand, janvier 2004.
- [8.] MELANCON G., HERMAN I., DELEST M. *Indices visuels et métriques combinatoires pour la visualisation de données hiérarchiques*. In Proceedings of the IHM'99 Workshop, Montpellier, pp. 166-173, (1999).