

LA PRISE EN COMPTE DE LA DIMENSION TEMPORELLE DANS LA VISUALISATION DE DONNEES PAR MORPHING DE GRAPHE

Eloïse LOUBIER(*), Wahiba BAHOUN (*), Bernard DOUSSET(*)
loubier@irit.fr, wahsoun@irit.fr, dousset@irit.fr

(*)*Institut de Recherche en Informatique de Toulouse, IRIT-SIG*
Université Paul Sabatier,
118 route de Narbonne,
31062 Toulouse cedex 9
(France),

Mots clefs :

Réseau d'acteurs, réseau sémantique, graphe évolutif, alliance, morphing, position stratégique, analogie espace/temps, analyse relationnelle.

Keywords:

Actors network, semantic network, evolutionary graph, alliance, morphing, strategic position, spaces/time analogy, relational analysis.

Palabras clave :

Red de los agentes, red semántica, gráfico evolutivo, alianza, morphing, posición estratégica, analogía de los espacios/tiempo, análisis emparentado.

Résumé

L'étude de l'évolution des données relationnelles issues de la synthèse de grands corpus d'information est l'un des enjeux actuels de l'intelligence économique. Son principal objet est l'identification, l'analyse et la restitution des structures stratégiques sous jacentes qui relient entre eux et à des moments donnés les acteurs d'un domaine et les concepts qu'ils utilisent. Dans ce contexte, le recours à la visualisation par des graphes apporte un réel confort aux utilisateurs, qui, de façon intuitive, peuvent s'approprier une forme de connaissance difficile à décrire autrement. De plus, la prise en compte de l'aspect temporel au sein de l'analyse permet de situer les événements, les stratégies, les actions aussi bien dans le passé (reconstruction de la chronologie), le présent (orientation temporelle), que le futur (anticipation) pour tout ce qui concerne les organisations successives d'un réseau (collaborations, alliances, fusions, acquisitions, co-citations, co-signatures, co-occurrences de tous ordres).

Dans ce contexte, le prototype VisuGraph, module de la plate-forme de veille stratégique Tétralogie, permet déjà la visualisation et la classification interactive de données relationnelles. Afin de mieux analyser les données évolutives des différents réseaux d'acteurs ou sémantiques, nous proposons d'ajouter à VisuGraph une fonctionnalité portant sur le morphing de graphe. L'objectif est de faire ressortir les tendances significatives en réalisant une animation entre les visualisations successives des graphes attachés à chaque période et en se basant sur la représentation d'un graphe global toutes périodes confondues. Notre approche initiale se base sur le principe des horloges, pour lesquelles les repères temporels sont fixes et seules les aiguilles sont en mouvement. Nous assimilons les différentes périodes à des repères temporels fixés dans un ordre chronologique sur le pourtour de la fenêtre de visualisation et plaçons chaque sommet en fonction de ses appartenances aux différentes périodes. Ainsi, par l'analogie espace=temps, il est possible de distinguer les éléments persistants, des éléments temporaires et d'appréhender la chronologie des changements de structure dans les relations.

Dans cet article, nous présentons l'algorithme de morphing de graphe ajouté à VisuGraph, en expliquant son principe, ainsi que sa mise en place et sa validation sur un exemple à l'échelle.

Introduction

La plate-forme de veille stratégique Tétralogie [1], développée à l'IRIT, est dédiée à l'analyse de données textuelles par le croisement des informations contenues dans de gros corpus de données, pour en extraire une synthèse non accessible directement par la lecture. Elle comporte, entre autres, un module de graphe nommé VisuGraph [2], développé en Java, permettant 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.

Dans ce contexte d'analyse, la prise en compte de la dimension temporelle permet d'étudier l'évolution des structures en évaluant les organisations passées, actuelles et surtout en déduisant les futures. Or, cette analyse est souvent faussée, lorsqu'elle se base sur un graphe statique. En effet, si nous étudions une classe supposée connexe lors de la représentation des n périodes cumulées, on s'aperçoit que la connexité n'est pas forcément confirmée pour chaque phase et que les conclusions issues du graphe global sont souvent erronées. Ainsi, pour l'étude de l'évolution des données, il convient de visualiser successivement les n périodes, afin de pouvoir évaluer correctement les changements de structures et observer l'apparition/disparition de signaux forts/faibles, 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 cela, nous dotons VisuGraph d'une fonctionnalité de morphing de graphe, permettant de réaliser une cinématique, par la visualisation chronologique et successive des différentes périodes. 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. Dans notre cas, des repères temporels non visualisés, représentant les différentes périodes, sont fixés dans un ordre chronologique sur le pourtour de la fenêtre de visualisation. 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.

Dans cet article, nous effectuons dans un premier temps un état de l'art sur la visualisation de données relationnelles et évolutives via les graphes. Dans un second temps, nous étudions le prétraitement des données, permettant de transformer l'information issue des gros corpus en matrices de relations. Nous présentons ensuite le codage de l'information, par un certain nombre d'artifices visuels tels que les choix d'icônes et de couleurs afin d'améliorer l'ergonomie de la représentation. Nous définissons enfin le morphing, son principe et son fonctionnement que nous validons par une expérimentation.

1 Etat de l'art

« Détecter et visualiser les changements d'un réseau évolutif représente un défi significatif dans le domaine de la représentation graphique. La visualisation successive de graphes par période est une méthode essentielle pour comparer différents réseaux dans le temps », selon North [3] (1995).

Les représentations externes, ou conventionnelles, du temps sont multiples [4] et peuvent être considérées comme des cadres emboîtés à différentes échelles. Les travaux sur la récupération et la datation d'événements passés montrent l'utilisation de ce type de représentation [5] [6].

Un bon exemple de visualisation des changements de thématique, en fonction du temps, dans une collection de documents textes est ThemeRiver [7]. Assimilant le flot d'informations à un « fleuve » traversant le temps, la largeur de ce dernier varie pour dépeindre les changements des thématique des documents au cours du temps. Les thèmes des documents sont représentés en tant que « courants » colorés coulant dans le fleuve et qui se rétrécissent ou s'élargissent pour indiquer des diminutions ou des augmentations de la présence d'une thématique des documents associée à un point spécifique par rapport au temps. Le fleuve se base sur une chronologie et une visualisation de repères textuelles assimilés aux événements externes, comme le montre la figure 1.

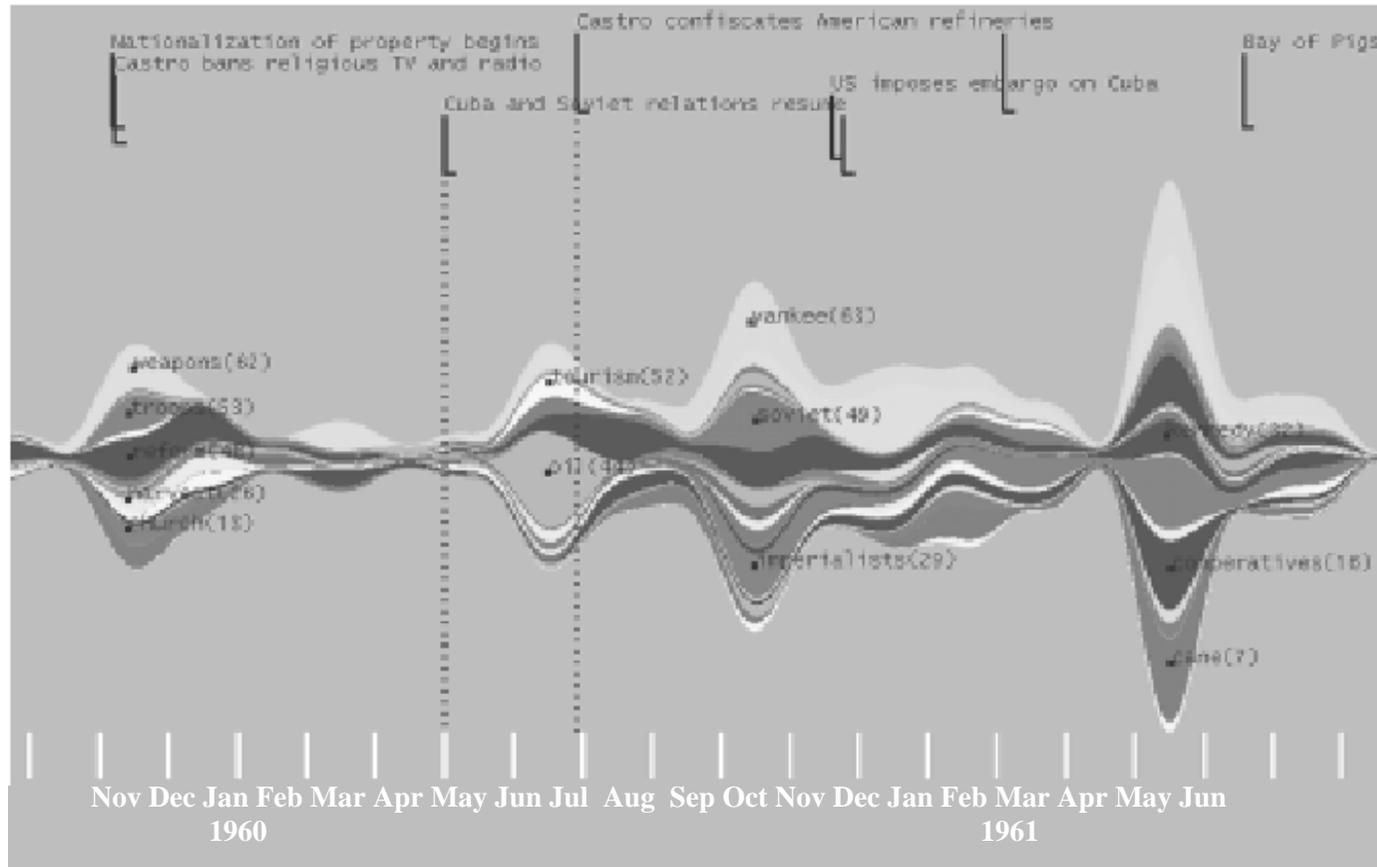


Figure 1: Représentation graphique des données (discours, interviews, articles, textes) sur Fidel Castro, de Novembre 1959 jusqu'à Juin 1961 effectuée avec ThemeRiver.

Parmi les outils permettant l'analyse de données évolutives, nous pouvons aussi citer TimeLine [8], qui permet la construction, l'exploration, et l'interprétation des lignes de temps afin d'identifier et de visualiser les citations d'articles de recherches sur des thématiques précises sur plusieurs périodes. Le graphe se base sur un repère dont les abscisses représentent le temps et les ordonnées sont les domaines de recherche, classés selon leurs similarités (cf. figure 2). Ainsi chaque sommet du graphe, dont la taille est relative à la valeur de sa métrique, correspond au nombre de citations, à une date et à une activité de recherche particulière.

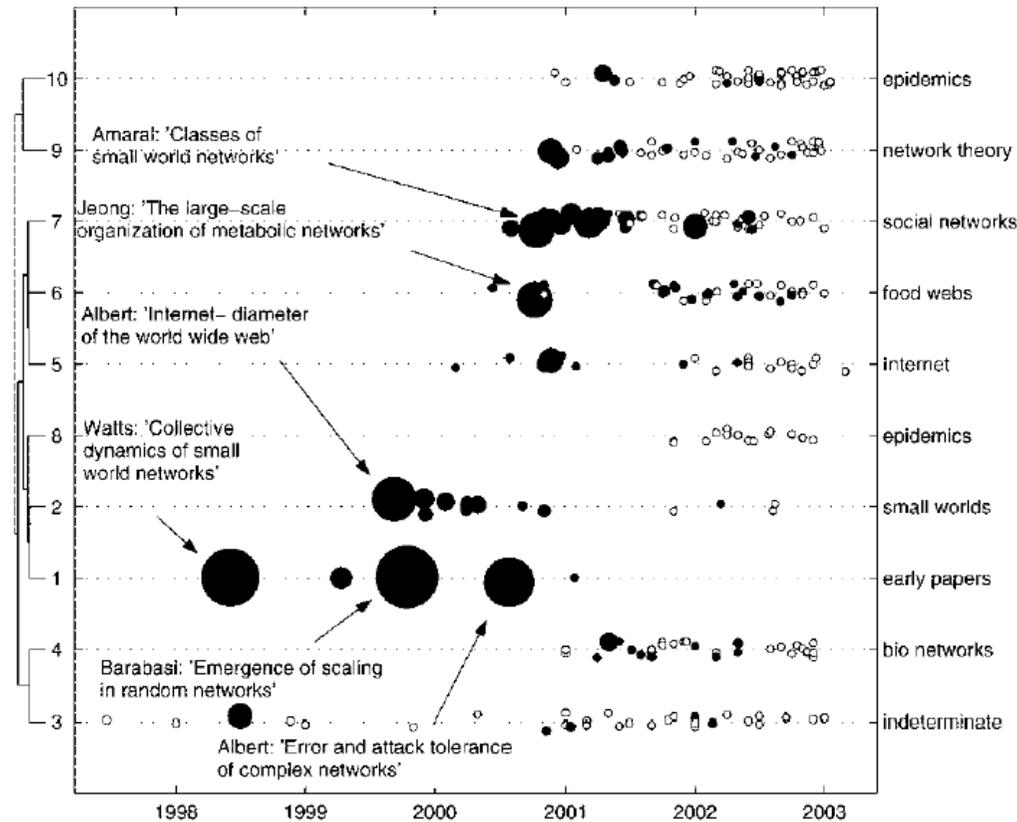


Figure 2 : TimeLine sur des articles de recherche. Les papiers sont représentés par des cercles dont la taille est proportionnelle au nombre total de fois où ils ont été cités. Les cercles remplis sont des papiers qui ont été cités huit fois ou plus dans 12 derniers mois.

Erten et al. [9] 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, 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. Comme le montre la figure 3, 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.

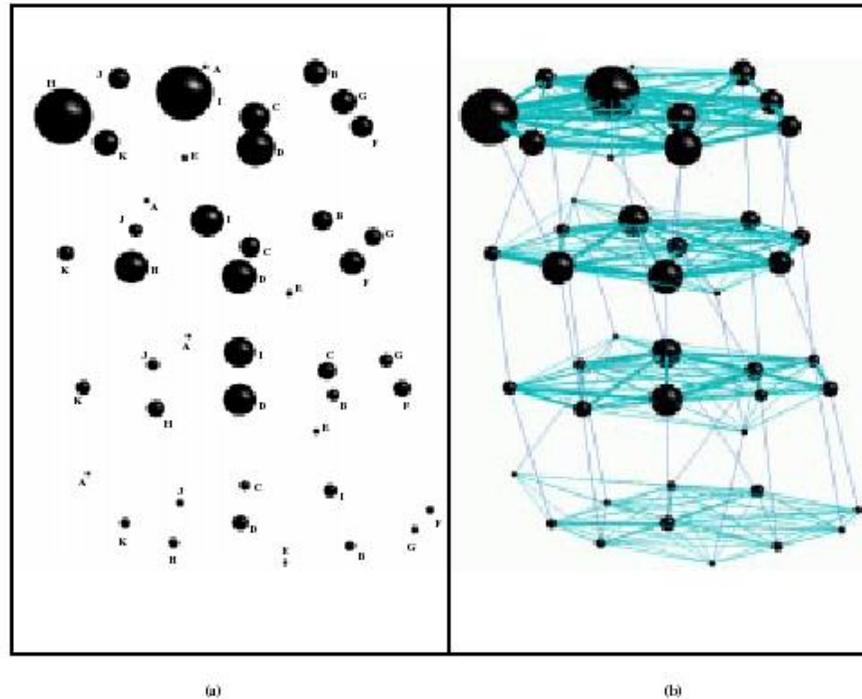


Figure 3: Représentation de l'évolution de collaborations par TGRIP (Erten et al, 2004). T1 (1981-1985), T2 (1986-1990), T3 (1991-1995), T4 (1996-2000); (a) Vue du graphe combiné, sans arête (b) vue du graphe combiné avec toutes les arêtes

Chen et al. [10] ont introduit les techniques de visualisation animées afin de resituer des citations ou co-citations dans un ordre chronologique. La figure 4, générée par l'outil CiteSpace, visualise les réseaux de co-citation les plus importants, à partir d'articles publiés dans le domaine considéré. Cette représentation met l'accent sur les acteurs persistants, dont la présence est effective sur plusieurs périodes et ceux qui ne le sont pas (turning point). Plus un noeud est relié à d'autres, plus son rôle de pivot est important.

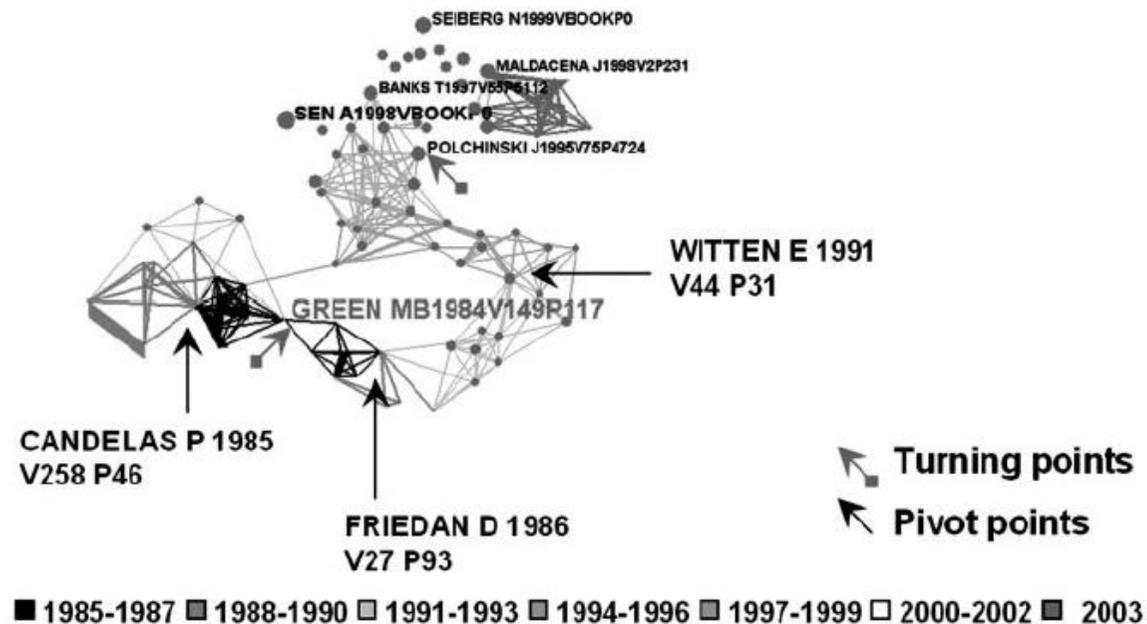


Figure 4 : Graphe généré par l'outil Citespace, visible en couleur sous www.pages.drexel.edu/~cc345/citespace/Figure2.png

2 La représentation graphique des données relationnelles : l'outil VisuGraph existant

Tétralogie, ainsi que son module graphique VisuGraph proposent initialement les fonctionnalités suivantes [11].

2.1 Le prétraitement des données évolutives

Les données analysées sont générées sous forme matricielle, dont le nombre de dimensions prises en compte est variable, selon les spécificités de l'analyse. Elles sont issues de gros corpus de données, dont l'origine provient de bases de données, de revues, journaux, périodiques, de thèses et de brevets, ainsi que de CD-ROMS. Elles sont traitées par une fonctionnalité de synonymie permettant d'obtenir des données homogènes, en regroupant sous un seul terme tous les synonymes correspondants, en gérant aussi les flexions morphologiques, sémantiques, ainsi que les inclusions et changements de granularité [12].

Les matrices analysées se basent sur des croisements entre deux champs, sous champs ou groupes de champs et peuvent être de type présence/absence (existence d'un lien dans au moins un document), de contingence (entre deux variables à item unique : journal, date, langue), ou encore de co-occurrence (une des variables peut contenir plusieurs items : auteurs, mots-clés....).

Pour introduire la notion d'évolution, nous croisons nos données avec la variable temps correctement discrétisée en « périodes » homogènes, nommées aussi « instances ». Ainsi, pour quatre périodes distinctes, nous aurons quatre matrices, chacune indiquant les valeurs des métriques de chaque item, ainsi que leurs liens, pour l'instance considérée.

2.2 Le codage de l'information évolutive

Les données des matrices sont représentées sous forme de graphe, $G = (V, E)$, comme un ensemble de sommets pondérés V et d'arcs valués E , orientés ou non selon le contexte d'analyse qui joignent chaque paire de sommets. Ces derniers sont généralement représentés sous forme de barres, reliés par des arcs sous forme de courbes ou segments.

Pour chaque matrice de croisements considérée, les entités sont assimilées à des sommets, dont le poids correspond au croisement de l'item avec lui-même; les autres valeurs indiquées correspondent aux croisement entre les différentes données, indiquant les valeurs des liens représentés sous forme d'arcs.

La représentation des différents fragments temporels s'effectue en deux étapes : une visualisation globale toutes périodes confondues, cas statique, puis individuelle pour chaque période.

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.

Ainsi chaque sommet est représenté sous forme de barre, dont la taille est relative à son poids. Dans un cadre d'analyse temporelle sur plusieurs périodes, nous avons une valeur de métrique par instance considérée, ce qui engendre de représenter les sommets sous forme d'un petit histogramme.

La taille des barres des sommets est basée sur la définition d'indices. Ces derniers sont définis à partir de la matrice $A = (a_{ij})$ associée au graphe sur une période donnée k . Chaque indice est calculé en effectuant le rapport de la valeur de la métrique du sommet v_i par la plus grande valeur de métrique de sommet de la matrice, toutes périodes considérées.

Soit $\mathbf{m}_i^k = \sum \mathbf{a}_{ij}^k$, la fonction qui associe à chaque sommet v_i la quantité $\mathbf{M}_i^k = \frac{\mathbf{m}_i^k}{\max(\mathbf{m}_j^k)}$ et définit un indice de v_i relatif à k .

Cet indice représente l'importance du sommet v_i dans la structure du graphe en terme de connexions et de taux d'information.

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 1 : Tableau des valeurs m_i des sommets

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 2 : Tableau des valeurs M_i des sommets

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 . Ainsi, la dernière barre de l'histogramme représentant le sommet « a » sera la plus grande de toutes. Chaque barre de l'histogramme est codée proportionnellement aux indices indiqués dans le tableau 7 final.

Pour la représentation du graphe global, une couleur, correspondant à une période spécifique, est affectée à chaque barre des histogrammes, permettant de distinguer les différentes valeurs du sommet entre les différentes instances, mais aussi de comparer les données sur l'ensemble des périodes ou de façon unitaire. Pour la représentation des liens, elle s'effectue sur la somme de leur valeur pour chaque période.

Pour la représentation par instance, toutes les barres des histogrammes sont de la même couleur, hormis celles correspondant à la période étudiées, afin de la distinguer. D'un point de vue unitaire, cela permet d'évaluer, pour chaque sommet, l'importance de son poids pour cette période, par rapport aux autres instances ; d'un point de vue général, cela permet de comparer toutes les données. Les arcs sont valués selon les valeurs indiquées dans la matrice temporelle correspondante par le croisement des deux items assimilés aux sommets qu'ils relient.

Ce type de codage permet de suivre facilement l'évolution d'un réseau. En effet, la taille de chaque barre étant proportionnelle à l'indice associé au sommet et chaque période étant indiquée par une couleur particulière, il est facile de détecter visuellement les sommets remarquables, dont la taille est maximale pour une période donnée et ceux en croissance (ou décroissance) progressive, ...

En ce qui concerne les arcs, l'intensité de la couleur est utilisée pour révéler l'importance de la valeur de leur métrique. Plus un lien a une valeur de métrique élevée, plus sa couleur sera de forte intensité et inversement, comme nous pouvons le voir dans la figure 5.

La couleur des arcs est établie selon le codage RVB (Rouge, Vert, Bleu) consistant au mélange des trois couleurs primaires. Le blanc est obtenu lorsque le rouge, le vert, le bleu sont additionnés en proportions égales (255, 255, 255). Le noir résulte de leur absence totale (0,0,0), correspondant à l'indice le plus élevé. Ainsi, plus la valeur de l'indice sera faible, plus les proportions RVB seront élevées et inversement.

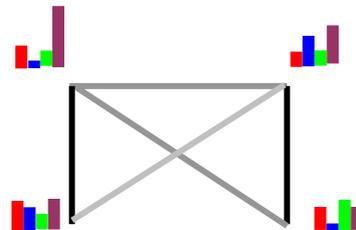


Figure 5 : Codage de l'évolution

Si la dimension temporelle est prise en compte, le calcul d'indices des liens et des sommets pour le graphe global s'effectue sur la somme des valeurs de la métrique du lien pour toutes les périodes considérées.

2.3 Application des forces d'attraction et de répulsion

Au-delà des procédés visuels utilisés pour faciliter la lisibilité du graphe, il est important de noter que plus le nombre d'entrecouplements d'arcs dans un graphe est limité, plus l'analyse de ce dernier est aisée. Afin d'obtenir un graphe aussi planaire que possible, nous proposons de mettre en place une méthode dynamique fondée sur le concept de « placement d'objets dans un champ de force » (« Force Directed Placement » ou FDP) proposés initialement par [13] Tutte et poursuivis, entre autres, par Eades [14], Kamada et al [15] et Frick et al [16]. Nous nous baserons sur Fruchterman et al. [17], qui offre une méthode générique de visualisation.

Deux forces antagonistes sont alors prises en compte : l'attraction appliquée à deux sommets reliés et la répulsion qui s'exerce sur tout couple de sommets liés ou non. L'intensité de ces deux forces dépend de la distance entre les deux sommets considérés. L'évolution du dessin 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. Deux sommets fortement liés se rapprocheront, via la force d'attraction qui leur est appliquée, contrairement à deux sommets non reliés, pour lesquels ne s'applique que la seule force de répulsion, provoquant leur éloignement si ces derniers sont proches.

A partir d'un état initial où les sommets sont placés le plus judicieusement possible dans l'espace cible, le système converge vers un état d'énergie minimum pour lequel les contraintes d'interaction entre les données sont satisfaites au mieux.

Notre travail se basant sur des graphes valués, nous proposons une extension des formules proposées par Fruchterman et al. [17] pour prendre en compte la valuation des arcs et la pondération des sommets.

La force d'attraction f_a (formule [1]) entre deux sommets peut être proportionnelle à la force de leur lien (valuation de l'arc). Pour les sommets v_i et v_j , elle est donnée par :

$$f_a(v_i, v_j) = \beta_{ij} \times d_{ij}^{\alpha_a} \times K^{-1} \quad [1]$$

β_{ij} est fonction du poids de l'arc (v_i, v_j) et du poids des sommets v_i et v_j . La constante K est calculée en fonction de l'aire du dessin et du nombre de sommets du graphe et d_{ij} est la distance euclidienne entre v_i, v_j . Si les sommets v_i, v_j ne sont pas reliés par un arc alors $f_a(v_i, v_j) = 0$.

La force de répulsion f_r (formule [2]) entre deux sommets v_i et v_j est définie par :

$$f_r(v_i, v_j) = -\beta_{ij} \times d_{ij}^{-\alpha_r} \times K^2 \quad [2]$$

La variable α_r , respectivement α_a , est une constante qui sert à définir le degré d'attraction (respectivement répulsion) entre deux sommets. Les valeurs des paramètres $\alpha_r, \alpha_a, \beta_{ij}$ permettant la mise en place d'un graphe lisible, ont été définies expérimentalement, minimisant les croisements d'arcs et maximisant la visibilité des clusters.

3 La prise en compte de la dimension temporelle

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 si celle-ci n'est pas prise en compte. Nous proposons d'enrichir VisuGraph de fonctionnalités qui facilitent l'analyse temporelle, en accordant une grande importance à l'aspect ergonomique mais aussi à l'interactivité avec l'utilisateur.

3.1 Le placement temporel des sommets

Comme nous l'avons vu précédemment au paragraphe 3.2, la prise en compte de la dimension temporelle s'effectue par une visualisation globale toutes périodes confondues, puis individuelle pour chaque période.

Afin de maintenir une bonne interactivité avec l'utilisateur, il faut préserver au mieux la stabilité des tracés. Pour cela, l'utilisateur doit garder en mémoire et pouvoir se référer au graphe global, c'est à dire sa carte mentale, en limitant les perturbations apportées sur le nouveau tracé de graphe par période, par rapport aux précédents.

Lorsque l'on introduit des fonctionnalités permettant l'ajout ou la suppression dynamique de sommets et d'arcs, la lecture des dessins dans un contexte évolutif conduit à des contraintes additionnelles. « *La stabilité est une notion complexe qui dépend des caractéristiques géométriques et combinatoires du tracé, mais aussi des facultés de perception et de mémorisation de l'utilisateur* » [18].

Il est donc indispensable que le graphe global de référence soit intelligible et clair, permettant, sans grand effort cognitif, de situer chaque graphe période dans le graphe global. Il convient que ce dernier repose sur un placement fixe par défaut des sommets, selon des repères spatiaux précis, permettant une mémorisation simple de la représentation.

Lors de la prise en compte de la dimension temporelle, nous proposons d'utiliser des repères temporels virtuels (non visibles dans le dessin mais dont la présence est prise en compte dans le graphe), assimilés à des sommets spécifiques, intégrés au dessin de graphe. Ainsi, il existe autant de repères temporels que de périodes analysées, représentés par des sommets situés à équidistance les uns des autres et dans un ordre chronologique, près des bords de la fenêtre de représentation (comme les heures sur un cadran d'horloge). L'intérêt est de placer les sommets représentant nos données, à une distance, des repères temporels, relative à leur appartenance (ou non) à chaque période. Plus la valeur de la métrique d'un sommet pour une date sera importante, plus ce dernier sera proche du repère temporel représentant cette période.

Ainsi les coordonnées temporelles de chaque sommet du graphe seront les suivantes :

$$\left\{ \frac{\sum \text{repx}^k * m_i^k}{\sum m_i^k} ; \frac{\sum \text{repy}^k * m_i^k}{\sum m_i^k} \right\} [3]$$

repx^k et repy^k correspondent aux coordonnées du repère temporel, représentant la période k .
 m_i^k correspond à la valeur de la métrique du sommet v_i pour la période k .

Une fonction est mise en place afin d'éviter la superposition des sommets. En effet, il est possible que deux sommets aient des appartenances identiques aux différentes périodes considérées. Selon nos formules, ces deux sommets ont les mêmes coordonnées, ce qui signifie qu'ils se superposent, empêchant ainsi leur distinction. Il convient donc, lors de chaque calcul de coordonnées, de vérifier si les abscisses et ordonnées résultantes ne sont pas déjà attribuées à un autre sommet. Si tel est le cas, nous modifions, au pixel près, la coordonnée afin d'éviter toute superposition.

Au niveau de l'interprétation, cette fonctionnalité apporte un réel confort à l'utilisateur, lors de l'analyse des données temporelles ; le positionnement temporel des données permet à l'utilisateur d'observer les regroupements de données par période, mais aussi de détecter les éléments ponctuels de ceux persistants sur la durée ; ou encore ceux en pleine évolution ou perte de vitesse. Dans la figure 6, nous pouvons observer plusieurs phénomènes temporels différents.

Dans le cas des sommets situés près du 1, les acteurs sont spécifiques à la première période. Ils apparaissent sous la forme d'une seule barre rouge (barre assimilée à la valeur de la métrique pour cette première période). Il s'agit donc d'éléments ponctuels, disparaissant pour les autres tranches de temps considérées.

Dans le cas 2, les sommets représentés appartiennent à deux périodes distinguées : la première et la seconde. Ces acteurs semblent être émergents puisqu'ils ne se limitent pas à une seule période. De par leur position, il est facile de voir que ces sommets ne persistent pas pour les troisièmes et quatrièmes périodes et que leur présence est prédominante au cours de la deuxième tranche de temps, puisqu'ils sont plus proches du repère temporel « 2 » que du « 1 ».

A l'inverse, dans le cas 3 les sommets appartiennent aux trois premières périodes, confirmant leur aspect persistant.

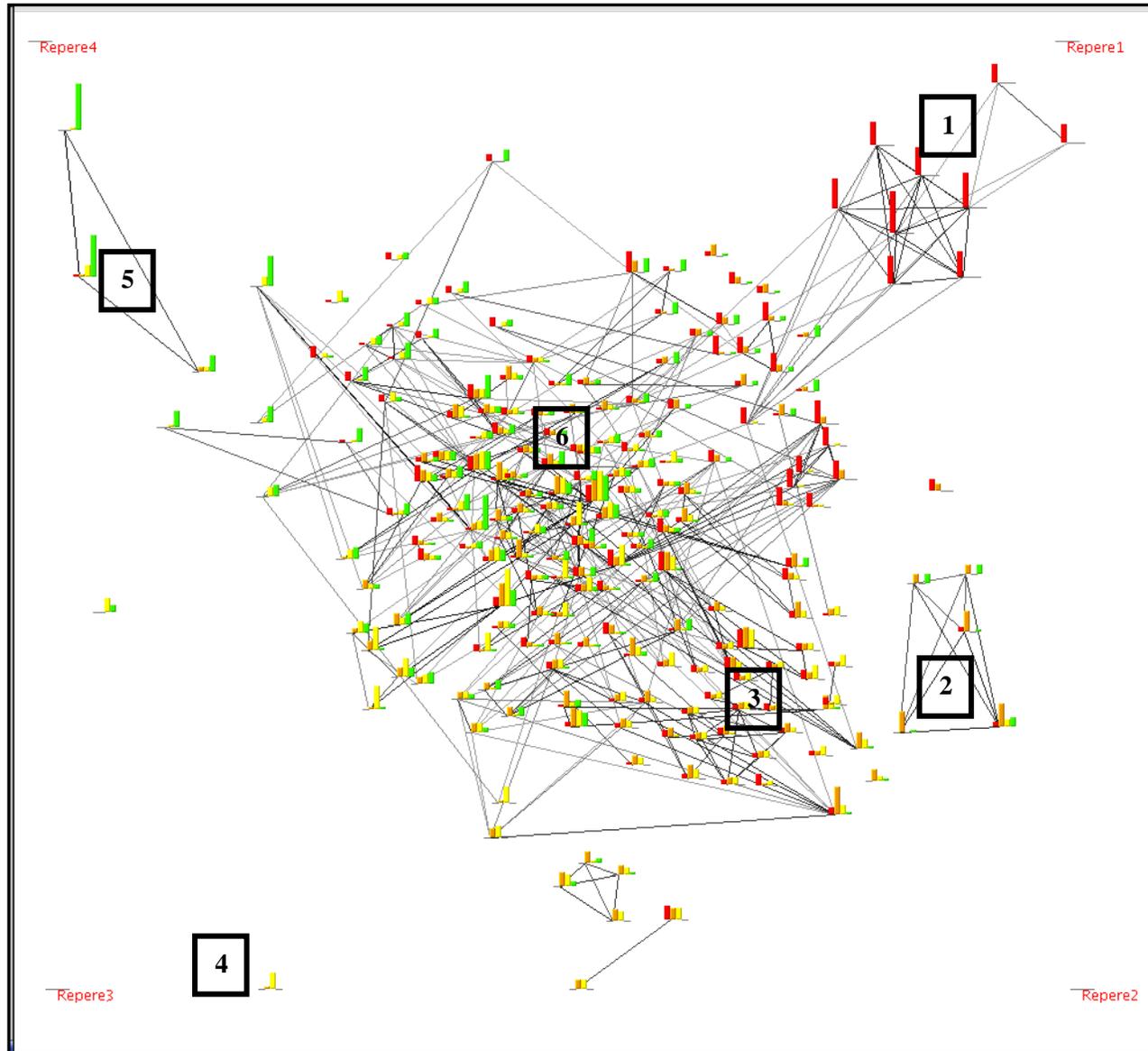


Figure 6 : Graphe global dont les sommets sont positionnés selon leurs caractéristiques temporelles.

Dans le cas 4, nous voyons que seul un sommet est spécifique à la troisième période, les autres sommets de cette tranche de temps appartenant aussi aux périodes précédentes/suivantes.

Dans le cas 5, les sommets appartiennent à la quatrième période. De par la grande taille des barres des histogrammes, nous pouvons penser que ces acteurs sont émergents et qu'ils seront présents dans les périodes futures.

Le cas 6 se situe en position centrale puisqu'il regroupe les données appartenant à toutes les périodes. Les positions se distinguent par les différences des valeurs des métriques pour chaque période. Ce cas là regroupe tous les sommets persistants. Le noyau central révèle les éléments omniprésents alors que ceux en périphérie sont plus caractéristiques des périodes vers lesquels ils sont proches.

De manière générale, nous pouvons donc dire que plus la distance entre le sommet et le centre de la figure est importante, moins le sommet est persistant (le nombre de périodes auxquelles appartient le sommet est faible), et inversement. Ainsi, plus le sommet est situé

- en périphérie de la fenêtre et à proximité d'un sommet, plus il est spécifique à cette période.
- Au centre de la fenêtre, plus il est omniprésent pour l'ensemble des périodes.

Le placement stratégique des sommets permet, non seulement, de les situer dans le temps mais aussi d'en évaluer la persistance et d'en déduire la tendance.

3.2 Le morphing de graphe

Le morphing consiste à réaliser une animation qui transforme de la façon la plus naturelle et la plus fluide possible un graphe initial vers un graphe final. L'objectif est de réaliser une lecture intuitive de l'évolution des données en répartissant à équidistance les périodes sur un cadran, et en positionnant les sommets selon leurs coordonnées temporelles, comme vu précédemment dans le paragraphe 3.1. À partir de la représentation du graphe global, il est possible d'observer dans un premier temps les différentes tendances des données, positionnées stratégiquement selon leur appartenance aux différentes.

Dans un second temps, le morphing de graphe permet de visualiser successivement chaque graphe de période, de façon animée et fluide. En se basant sur l'analogie espace/temps, il permet ainsi de détecter, comprendre et même prévoir les tendances significatives, au travers de la visualisation de l'évolution des données.

En se basant sur le graphe global, chaque période est visualisée en masquant les sommets et les arcs du graphe global n'appartenant pas à la période considérée, cf. figure 6. Il apparaît un graphe ne contenant que les sommets et les liens de la période étudiée. À partir de ce graphe, initialement placé vers le repère temporel de la période considérée, le graphe est centré par rapport à la totalité de la fenêtre de représentation, puis les forces d'attraction et de répulsion sont appliquées, comme vues dans le paragraphe 2.3, afin d'obtenir un graphe dont le nombre d'entrecouplements d'arcs est limité. Le graphe obtenu est organisé de façon à révéler les différentes organisations, alliances, collaborations.

Afin de conserver la carte mentale de l'utilisateur, la visualisation de la période $k+1$ passe par la reconstitution du graphe de la période k avant le centrage et l'application des forces d'attraction et de répulsion, tel qu'il est dans le graphe global. Le graphe global réapparaît avant de passer à la visualisation du graphe de la période suivante. Les mêmes opérations pour le passage à la période suivante sont effectuées pour n'afficher que les arcs et sommets du graphe global qui appartiennent à la période $k+1$, puis la représentation graphique est centrée et soumise aux forces d'attraction et de répulsion.

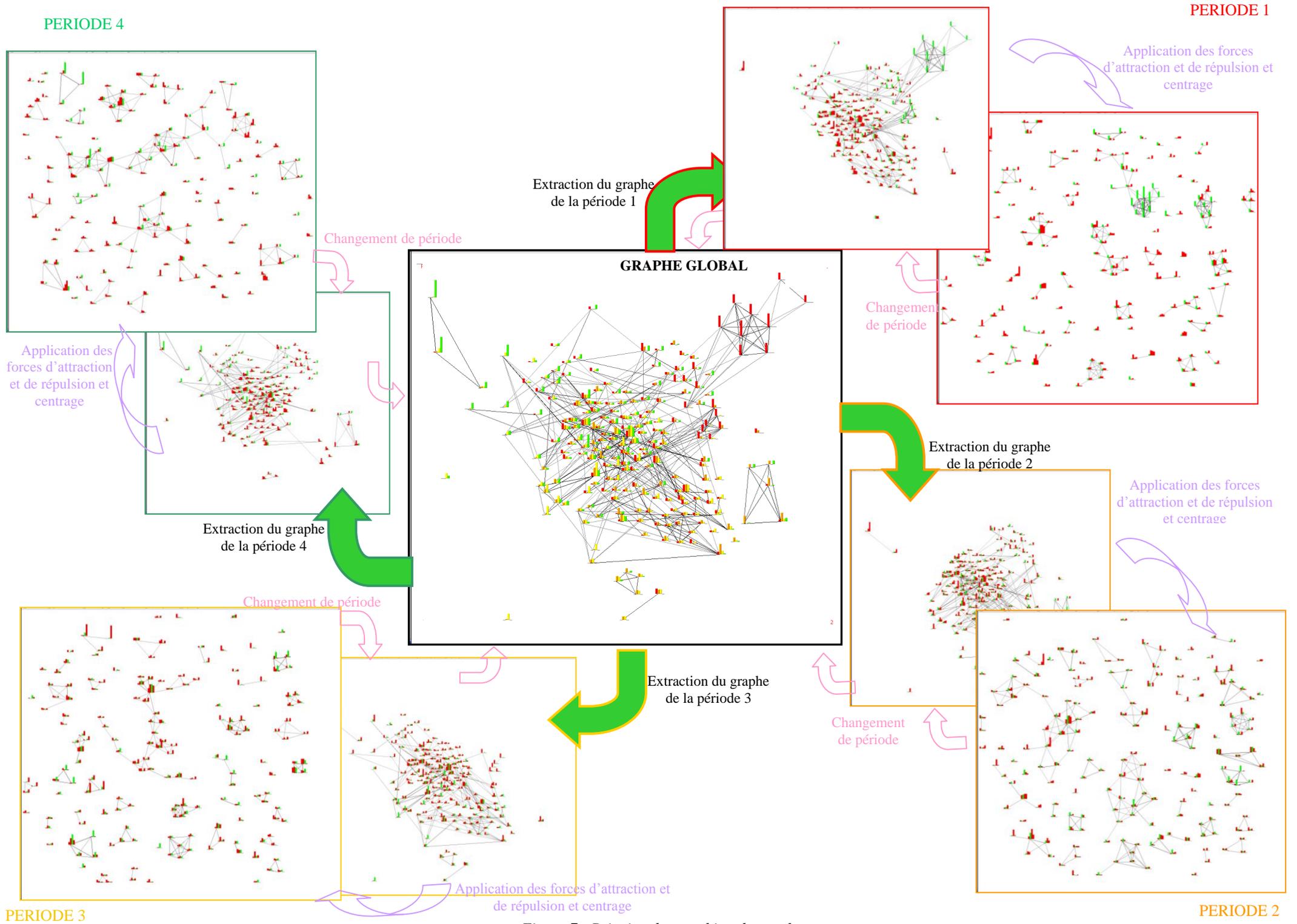


Figure 7 : Principe du morphing de graphe.

Le passage d'une instance à l'autre se fait par le biais d'un slider gradué, représenté sur la figure 8. Lorsque l'utilisateur le fait varier, la visualisation change afin d'afficher la représentation de la période suivante.

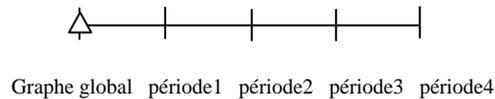


Figure 8 : slider permettant de choisir la période à visualiser

L'animation des visualisations successives des différentes périodes, dans le sens chronologique similaire à l'analogie espace/temps d'une horloge, permet de créer une certaine dynamique, révélant l'évolution des données au cours du temps, trouvant ainsi un bon compromis entre la préservation de la carte mentale de l'utilisateur et la lisibilité du tracé.

La figure 7 illustre le principe de morphing de graphe. Les données étudiées sur quatre périodes distinctes sont issues d'importants corpus de données, analysés sous Tétralogie, permettant d'obtenir quatre matrices temporelles de co-occurrences.

Les données sont représentées dans un premier temps sous forme de graphe global (au centre sur la figure 7), où les quatre périodes sont visualisées simultanément. Les quatre repères temporels sont placés aux angles de la fenêtre de représentation et les sommets sont disposés selon leur appartenance à chacune des tranches de temps. Ce graphe permet d'appuyer l'utilisateur sur son analyse temporelle. Par simple observation de ce graphe global, les sommets ponctuels, qui n'appartiennent qu'à une période spécifique se distinguent par leur positionnement (proche des bordures de la fenêtre). Les sommets positionnés au centre sont persistants (cf 3.1).

L'extraction du graphe de première période par masquage des sommets et des liens n'appartenant pas à cette dernière permet de mettre en avant le nombre d'acteurs présents pour cette tranche de temps. Un groupe de sommets spécifiques à cette période se distingue. Situés dans le cadran « Nord-Est » de la figure, ils sont visualisés sous forme d'une seule grande barre verte pour chaque sommet (l'absence d'autre barre de couleur pour ces sommets reflète leur absence au cours des autres périodes). Afin de rendre le graphe plus lisible, l'application des forces d'attraction et de répulsion permet d'obtenir un graphe le plus planaire possible et d'étudier plus précisément les différentes associations. Il apparaît que le nombre d'alliances importantes entre de nombreux acteurs (plus de trois acteurs) est faible. Nombreux sont les acteurs isolés ainsi que les alliances entre deux ou trois acteurs.

Ce graphe de première période peut être resitué dans son contexte de graphe global. Pour cela, avant de passer à la période suivante (lors du changement de position au niveau du slider), les sommets reprennent leur position initiale, c'est-à-dire celle qu'ils avaient dans le graphe global et les sommets des autres périodes réapparaissent.

A nouveau par masquage, le graphe de seconde instance est obtenu. La visualisation obtenue suite à l'application des forces d'attraction et de répulsion met en avant la complexité des alliances, qui sont plus nombreuses et de taille plus importante. Le nombre de sommets « libres » est faible.

Le passage à la période suivante s'effectue aussi par repositionnement des sommets dans leur contexte global, ainsi que par réaffichage des sommets toutes périodes confondues.

L'extraction de la période trois par modification du slider permet d'étudier les acteurs de cette tranche de temps. Leur organisation révèle une diminution légère du nombre d'alliances mais ces dernières sont composées de nombreux acteurs. Rares sont les sommets isolés.

Enfin, suite à l'affichage à nouveau du graphe global afin de permettre à l'utilisateur de conserver en mémoire la position initiale des sommets, la quatrième période est extraite. Cette dernière met en avant des acteurs persistants, situés dans le noyau (visible dans l'intitulé « extraction du graphe de la période 4 »), mais aussi les éléments naissants, apparaissant dans le secteur Nord-ouest du cadran. L'application de forces d'attraction et de répulsion permet de visualiser principalement une grande alliance qui relie une grande partie des acteurs. Les autres alliances sont de taille relativement importante.

4 Conclusion

Dans cet article, nous avons présenté le prototype VisuGraph, permettant de prendre en compte la dimension temporelle dans le cadre d'une analyse de données relationnelles évolutives. La fonctionnalité de morphing de graphe se base sur une animation réalisée à partir d'une représentation globale des données, constituant la carte mentale de l'utilisateur, suivie d'une visualisation animée et successive des différents graphes d'instances. Chaque période est assimilée à un sommet virtuel, agissant comme un repère invisible. La position de chaque sommet est calculée stratégiquement, de façon à révéler l'appartenance aux différentes périodes. La visualisation de chaque fragment temporel s'effectue par masquage des sommet et arcs n'appartenant pas à l'instance considérée. A partir de cette position initial du graphe de période, ce dernier est centré, les forces d'attraction et de répulsion sont appliquées afin de limiter l'entrecouplement des arcs et de rendre le graphe plus lisible. Avant de passer à la visualisation de la période suivante, le graphe reprend sa position initiale, avant déformation, au sein du graphe globale.

Les faiblesses de VisuGraph reposent sur l'interface graphique, qui devrait davantage être intuitive, afin d'en faciliter l'usage. De nombreux outils plus performants existent pour des analyses statiques (Aisee[19], Tulip[20], InfoVis[21], Guess[22]...). Cependant ces derniers ne traitent pas l'aspect évolutif des données tel que le fait VisuGraph. Ce module de Tétralogie est limité aussi au niveau du volume des données. En effet, plus le nombre de données est important, plus le système est ralenti à cause du nombre important de calculs. De même, l'algorithme de force d'attraction et de répulsion reste très lent et l'affichage des données reste massif face aux grands volumes de données.

Malgré ces faiblesses, VisuGraph est un outil permettant à l'utilisateur d'interagir avec le système et ainsi de le contrôler. Il reste maître de son analyse et des traitements appliqués aux données qu'il étudie. La visualisation dynamique des données permet de les organiser au mieux et d'étudier l'aspect évolutif des informations. Ce support visuel permet de conforter l'utilisateur dans son analyse et de l'aider dans un but décisionnel, en étudiant les différentes tendances, en observant les éléments majeurs, les acteurs persistants mais aussi les émergeant, pouvant être à l'origine de changements. Cet outil facilite la tâche du décideur, en lui offrant une visualisation simple et efficace d'importants volumes de données, difficilement analysable sous leur forme brute.

Ce travail ouvre de nombreuses perspectives, quant à l'évolution du morphing de graphe. Le graphe global, carte mentale de l'utilisateur, est basé sur des coordonnées temporelles, sans application des forces d'attraction et de répulsion. Pour des graphes portant sur un nombre de données importantes, le graphe peut facilement devenir illisible si le nombre d'entrecouplements d'arcs est trop important. Il conviendrait donc de trouver un juste milieu permettant de placer les sommets selon leur appartenance aux différentes périodes, tout en rendant le graphe le plus planaire possible par l'application des forces d'attraction et de répulsion.

De plus, ce morphing de graphe repose sur un retour au graphe global entre chaque période. Il serait intéressant de le compléter par une solution proposant le passage d'une période à la suivante sans retour aux positions initiales, rendant le procédé plus rapide, plus fluide et limitant les étapes intermédiaires. Actuellement, le morphing s'effectue entre plusieurs graphes de périodes sur lesquels les fonctionnalités de centrage et les forces ont été appliquées. Une perspective d'évolution de l'outil serait la suivante : les éléments persistants conserveraient leur position centrale (sans application d'algorithme de force et donc de changement de position) et les signaux apparaissant ou disparaissant seraient mis en valeur par un fondu, mettant ainsi l'accent sur les changements de structure au niveau des barres des histogrammes de chaque sommet (valeur de métrique pour chaque période). Cependant cette solution aurait pour difficulté de gérer la complexité de la représentation, à savoir le risque d'obtenir un nombre d'entrecouplement de liens important suite au positionnement temporel initial des sommets. La non application des forces d'attraction et de répulsion engendre la perte de l'atout majeur de cet algorithme : l'apport de lisibilité au graphe. Il conviendrait alors de trouver une solution pour faciliter le fondu entre deux visualisation, tout en traçant facilement tous les sommets et en analysant de façon la plus claire possible les graphes d'instance.

Ainsi, le morphing de graphe peut être approfondi, tant d'un point de vue méthodologique, qu'algorithmique, afin d'offrir une meilleure perception de l'évolution temporelle des données relationnelles, que ce soit pour la persistance des éléments forts ou encore par la naissance ou la disparitions de signaux, qu'ils soient forts ou faibles.

5 Bibliographie

- [1] DOUSSET B., BENJAMAA T., *Trilogie logiciel d'analyse de données*, Conférence sur les systèmes d'informations élaborées : Bibliométrie – Information Stratégique – Veille technologique, 1988.
- [2] KAROUACH S., DOUSSET B. Les graphes comme représentation synthétique et naturelle de l'information relationnelle de grande taille. Workshop sur la recherche d'information : un nouveau passage à l'échelle, associé à INFORSID'2003, (Nancy France), 3-6 juin 2003.
- [3] NORTH SC., *Incremental layout in dynadag*. Proceedings of Graph Drawing (GD'95), pp. 409-418, 1995.
- [4] FRIEDMAN W. J. *About time. Inventing the fourth dimension*. Cambridge, Massachusetts: TheMit Press,1990.
- [5] LIEURY A., AIELLO B., LEPREUX D., MELLET M.. *Le rôle des repères dans la récupération et la datation des souvenirs*. *L'Année Psychologique*, pp.149-167, 1980.
- [6] ROBINSON J.A., *Temporal reference systems and autobiographical memory*. In D. C. Rubin (Ed.), *Autobiographical Memory*,pp. 159-188, Cambridge: Cambridge University Press, 1986.
- [7] HAVRE, S., HETZLER E., WHITNEY P., NOWELL L. IEEE T. Vis. Comput. Graphics 8, pp.9–20, 2002.
- [8] MORRIS S.A., YEN G., WU Z., ASNAKE B., *Timeline : visualization of research fronts*. *J. Am. Soc. Inf. Sci.Technol.*, 55, 5. pp. 413-422, 2003.
- [9] ERTEN C., HARDING P.J., KOBOUROV S.G., WAMPLER K, YEE G. *Exploring the computing literature using temporal graph visualization*. In *Conference on Visualization and Data Analysis (VDA)* , 2004.
- [10] CHEN C., *Searching for intellectual turning points: Progressive knowledge domain visualization*. Proceedings of the National Academy of Sciences of the United States of America PNAS, 2004.
- [11] KAROUACH S., DOUSSET B., Manipulation de graphes de grande taille pour l'étude des réseaux d'acteurs et des réseaux sémantiques". 10ièmes journées d'études sur les systèmes d'information élaborée: Bibliométrie - Informatique stratégique - Veille technologique, (Ile Rousse Corse France), CD-ROM, 13-17 juin 2005.
- [12] LOUBIER E., CARBONNEL S. *VisuGraph : Un outil d'exploration de données relationnelles évolutives*. INFORSID 2007, à paraître, 2007.
- [13] TUTTE , W.T. *How to draw a graph*. Proceedings London Mathematical Society, Volume 13, p. 743-768, 1963.
- [14] EADES P. *A heuristic for Graph Drawing*. Congressus Numerantium, vol. 42, p. 149-160, 1984.
- [15] KAMADA T., Kawai S. *An algorithm for general undirected graphs*. In Information Processing Letters, Volume 31, p. 7-15, 1989.
- [16] FRICK A., Ludwig A., Mehldau. *A fast adaptive layout algorithm for undirected graphs*. In DIMACS International Work. Graph Drawing, GD'94, 1994.
- [17] FRUCHTERMAN TMJ., Reingold EM. *Graph drawing by force_directed placement*. Software – Practice and experience, 21, p. 1129-1164, 1991.
- [18] PINAUD B., KUNTZ P., *Un guide sur la Toile pour sélectionner un logiciel de tracé de graphes*. Congrès VSST'2004 : veille stratégique scientifique & technologique : Systèmes d'information élaborée, bibliométrie, linguistique, intelligence économique , pp. 546 ; 540 , 2004.
- [19] ABSINT ANGEWANDTE INFORMATIK GMBH. *Aisee Software*. <http://www.aisee.com/>
- [20] AUBER D. *Tulip : A huge graph visualisation framework*. In P. Mutzel and M. Jünger, editors, *Graph Drawing Softwares*, Mathematics and Visualization, pages 105-126. Springer-Verlag, 2003.
- [21] FEKETE J.D. *The InfoVis Toolkit*. Proceedings of the 10th IEEE Symposium on Information Visualization (InfoVis'04), Austin, TX, Oct 2004. IEEE Press. pp. 167-174.
- [22] ADAR E. AND MIRYUNG K. *SoftGUESS: Visualization and Exploration of Code Clones in Context*. ICSE 2007.