



UNIVERSITÉ PARIS SUD

ECOLE DOCTORALE D'INFORMATIQUE PARIS-SUD

LABORATOIRE: INRIA, SACLAY, ILE DE FRANCE

DISCIPLINE: INFORMATIQUE

THÈSE DE DOCTORAT

Soutenue le 9 Mai 2014 par

BENJAMIN BACH

CONNECTIONS, CHANGES, AND CUBES:
UNFOLDING DYNAMIC NETWORKS FOR
VISUAL EXPLORATION

— SYNTHÈSE EN FRANÇAIS —

Directeur de thèse :	M. Jean-Daniel Fekete	Directeur de Recherche (INRIA Saclay, France)
Co-directeur de thèse :	M. Emmanuel Pietriga	Chargé de Recherche (INRIA Saclay, France, and INRIA Chile, Chile)
Composition du jury :		
Rapporteurs :	M. Jarke J. van Wijk M. Tim Dwyer	Professeur (Eindhoven University of Technology, Pays-Bas) Maitre de Conférence avec habilitation (<i>Senior Lecturer</i>) (Monash University, Australie)
Examineurs :	Mme. Silvia Miksch M. Guy Melançon Mme. Chantal Reynaud	Professeur (Vienna University of Technology, Autriche) Professeur (Université Bordeaux, France) Professeur (Laboratoire de Recherche en Informatique, Université Paris Sud, France)

Resumé de These

LES réseaux sont des modèles qui nous permettent de comprendre les relations entre éléments du monde réel. Une grande quantité de réseaux sont dynamiques, c'est-à-dire que leur connexité change au cours du temps. Comprendre les changements de connexité signifie comprendre les interactions entre les éléments de systèmes complexes: comment se forment les relations sociales et commerciales, comment sont transmis les signaux entre les régions du cerveau, comment s'organisent les réseaux trophiques après des catastrophes environnementales.

Une manière d'analyser ces réseaux est l'analyse quantitative qui fournit des mesures des caractéristiques spécifiques comme la taille, la densité (indiquant le nombre de liens par rapport au nombre de nœuds), ou des mesures de clustering. Au-delà de ce que nous permet la technologie et les algorithmes d'analyses, l'humain dispose de capacités uniques pour comprendre et interpréter les données : la vision et la cognition.

La vision permet de percevoir des structures complexes et des relations entre des objets. Par rapport à la visualisation des données et des réseaux, ce sens est une opportunité unique pour explorer des données et pour obtenir une vue globale. Un tel processus d'exploration est naturellement interactif. L'interaction homme-machine est considérée essentielle pour chaque système de visualisation.

Cette thèse développe et examine des moyens pour explorer les réseaux dynamiques d'une manière interactive et visuelle. Le concept central que je décris est la métaphore de *dépliage* des réseaux, c'est-à-dire, de réduire leur complexité, dans le but de les rendre compréhensibles, de les regarder à partir de perspectives différentes, d'examiner leurs composantes. Déplier des réseaux est une métaphore, comme la création des cartes bidimensionnelles d'objets tridimensionnels comme la Terre : chaque méthode de projection a comme résultat une carte différente qui permet de voir des relations différentes dans la taille des continents et des océans, des distances, etc.

Definitions et Etat-de-l'Art (Chapitre 2)

Le terme "*réseau dynamique*" décrit un réseau, composé de nœuds, de liens et potentiellement d'attributs associés avec ces nœuds et ces liens, qui évolue au cours du temps; des nœuds apparaissent et disparaissent, des liens s'établissent, se brisent, se consolident ou se fragilisent (en termes de poids). Ces changements simples sont responsables de changements complexes, comme la création de groupes et leur évolution au cours du temps. Dans cette thèse, le terme réseau dynamique est utilisé comme synonyme pour d'autres termes dans la littérature comme "réseau temporel" ou "graphe temporel". Le terme est surtout utilisé pour contraster avec les "réseaux statiques", qui ne modélisent aucun changement temporel.

Les visualisations permettent mieux comprendre les changements dans des réseaux dynamiques sont nombreuses, telles que les techniques utilisées pour l'encodage visuel

de l'information et la navigation interactive. Je classe les visualisations existantes en 7 groupes majeurs :

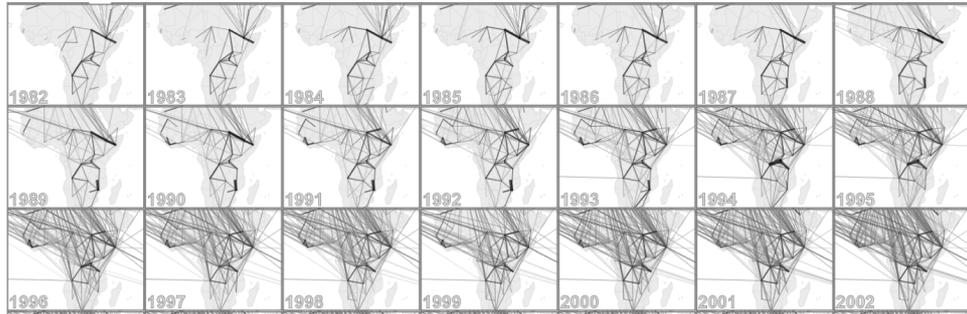
1. *Temporal Multiples* (des vignettes de chaque instant temporel sont présentées individuellement et simultanément sur l'écran juxtaposées, Figure 1(a)),
2. *Animated Sequences*, représentant le passage du temps par des animations entre les états du réseau à chaque instant,
3. *Aggregation* sont techniques pour l'agrégation de temps en utilisant des fonctions d'agrégation algorithmiques ou visuelles (Figure 1(b)),
4. des techniques de *composition* d'une représentation temporelle dans une représentation topologique du réseau (Figure 1(c)),
5. *Timelines*, des vues historiques qui représentent bien le cours du temps mais qui présentent justement une vue abstraite de la topologie (Figure 1(d)),
6. des *ego networks*, représentant un noeud focus ("ego") et les changements de sa connectivité avec ses noeuds voisins (Figure 1(e)),
7. des *Space-time cubes*. Celle-ci utilisent la métaphore d'un espace 3D composé de a) deux dimensions pour représenter la topologie du réseau et b) la troisième pour indiquer le temps (Figure 1(f)).

Toutes ces techniques ont des avantages et désavantages par rapport aux tâches que l'utilisateur accomplit au cours de son exploration. Les différences respectives peuvent être discutées à un niveau conceptuel, comme discuté dans le chapitre 8 de cette thèse, ou à partir des études quantitatives. Quant aux études quantitatives, les résultats et conclusions sont difficiles à comparer et synthétiser.

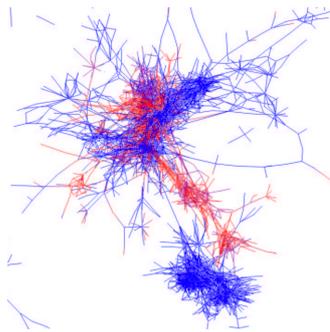
Vers une taxonomie de tâches pour les réseaux dynamiques (Chapitre 3)

Pour créer des visualisations et interactions efficaces, il est indispensable d'obtenir une vue générale des tâches que chaque visualisation doit supporter. Ce chapitre décrit une taxonomie de tâches, constituée de trois dimensions: 1) *composants de réseau* (noeuds, liens, groupes, etc.), 2) *temps* (instants temporels, périodes, durations, etc.) et 3) *changement* des composant durant le temps (apparition de noeuds, disparition de liens, etc). Ces trois dimensions nous permet de décrire des tâches simples en fonction de la dimension qui contient la réponse. La tâche "A quel moment disparaît le noeud *a*?" cherche à identifier un instant de temps, par rapport à un élément de réseau (noeud *a*) et un changement spécifique (disparition).

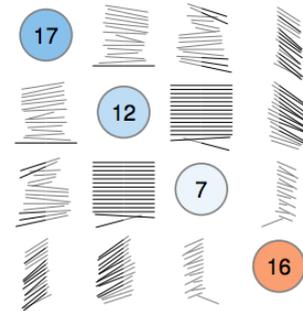
En se fondant sur cette méthode de classification et de description des tâches simples, nous pouvons décrire des tâches composées comme séquence de tâches simples; la réponse à une tâche initiale permet d'aborder la tâche suivante. Autrement dit, la réponse est nécessaire pour effectuer la tâche suivante. Au-delà de cette classification systématique, certaines tâches qui sont difficile à classifier et à décrire. Ces sont les tâches



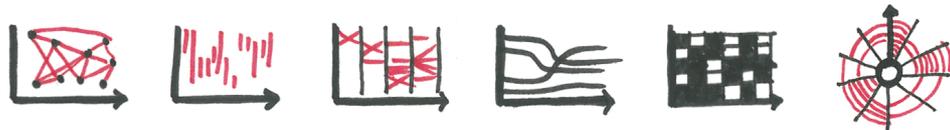
(a)



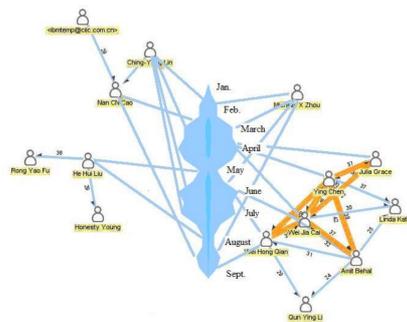
(b)



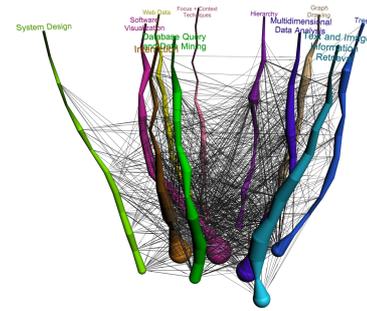
(c)



(d)



(e)



(f)

Figure 1: Exemples et catégories de visualisations de réseaux dynamiques. (a) Temporal multiples (Boyandin, Bertini, & Lalanne, 2012), (b) agrégation utilisant la couleur pour encoder le temps (Collberg, Kobourov, Nagra, Pitts, & Wampler, 2003), (c) composition (Brandes & Nick, 2011), (d) types pour utiliser une timeline pour visualiser des réseaux, (e) ego-network (Shi, Wang, & Wen, 2011), (f) représentation 3D (Ahmed et al., 2005).

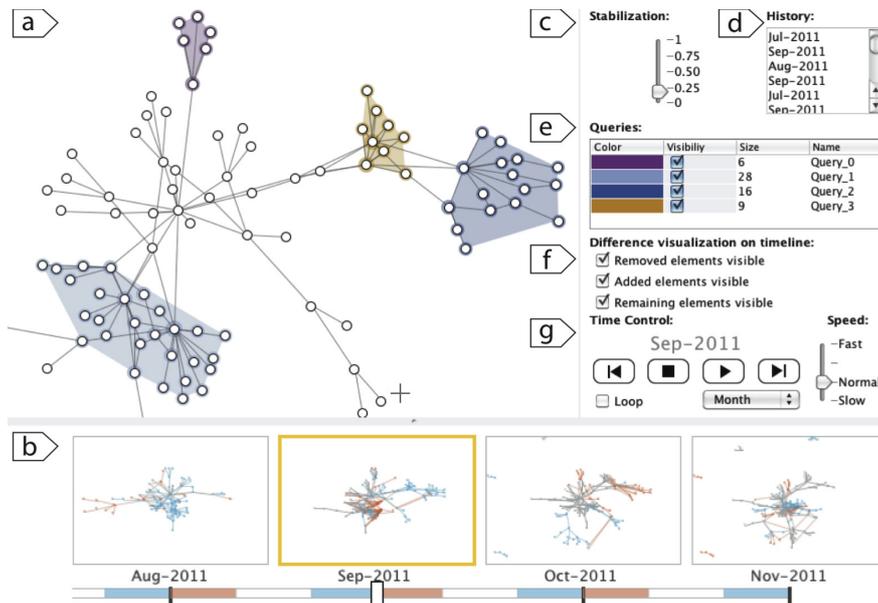


Figure 2: *GraphDiaries*: a) Vue topologique, b) Timeline, c) Stabilisation du layout, d) histoire de navigation, e) Node queries, f) Visibilité des éléments bleus, rouges, ou blancs dans la Timeline, g) panneau de configuration d'animations.

de haut-niveau comme “l’exploration des groupes” ou la “descriptions des tendances de changement au cours du temps.”

Le chapitre conclue avec des critères pour informer le design des interfaces de visualisation. Ces critères incluent la flexibilité de la navigation dans le temps, la présentation de la topologie du réseau, et de la possibilité de rendre explicites certains changements, par exemple, en utilisant des variables visuelles.

Transitions animées et navigation dans le temps (Chapitre 4)

La taxonomie de tâches, et les critères de design, nous permettent de concevoir une interface pour l’accomplissement de tâches. Cette interface, que nous appelons *GraphDiaries* (Figure 2), combine différentes techniques, telles que décrites dans l’état de l’art, et nous sert comme base pour leur extension. *GraphDiaries* utilise une vue topologique qui montre le réseau à un instant précis en utilisant la représentation noeud-lien (Figure 2-a)). Les autres instants temporel sont accessibles grâce à une représentation Small Multiples (Figure 2-b)). En cliquant sur une de ces représentations, une transition visuelle montre les changements dans le réseau dans la vue topologique. Pour montrer les changements temporels, la transition utilise de la couleur pour montre quels noeuds et liens apparaissent ou disparaissent. L’utilisateur a plusieurs options pour faire apparaître une transition: a) un simple clic rapide ne montre aucune animation entre des instants mais permet une navigation rapide, b) un click long pour voire la transition entière, et c) l’exploration interactive des transitions en utilisant un slider (Figure 2-c)).

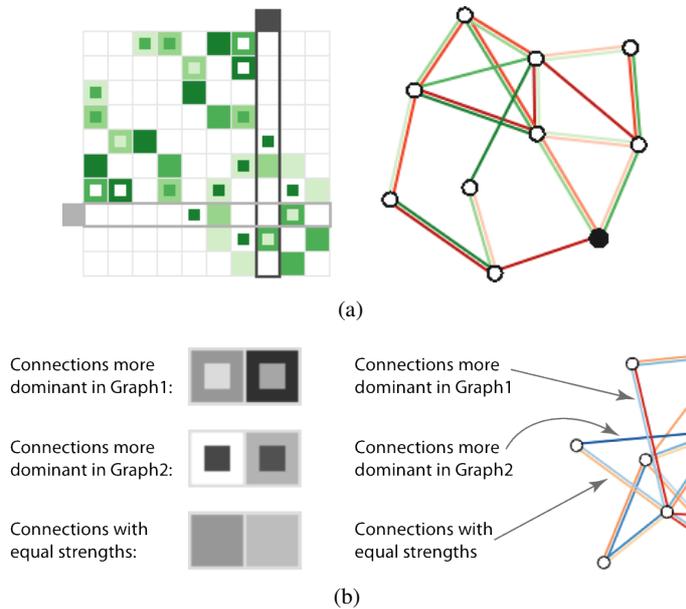


Figure 3: Comparaison des graphes (a) Encodage final pour matrices (gauche) et représentations noeud-lien. (b) Détail et explication de la représentation des liens dans la matrice (gauche) et dans la représentation noeud-liens, tels qu'ils apparaissent dans les explications pour l'étude qualitative.

Dans une étude quantitative on a mesuré la performance (temps pour compléter une tâche, taux d'erreur) de plusieurs techniques implémentées dans GraphDiaries. Dans cette étude, GraphDiaries a été comparé à des interfaces existantes par rapport à trois tâches, représentant les trois dimensions de notre taxonomie de tâches (Chapitre 3). Les résultats montrent que les techniques de GraphDiaries fournissent un meilleur support à la réalisation des tâches testées. GraphDiaries est facilement extensible par rapport aux nouvelles techniques de layout, aux nouvelles transitions et aux représentations de graphes ainsi que l'encodage visuel des données.

Comparaison de réseaux denses et pondérés (Chapitre 4)

Les diagrammes noeuds-liens, utilisés dans la plupart des visualisations actuelles, ne suffisent pas pour une représentation de réseaux denses avec des changements de poids de liens. Premièrement, ce chapitre décrit un cas particulier de ce type de réseau; les réseaux anatomiques et fonctionnels du cerveau où le poids des liens indique l'intensité des signaux, de la corrélation ou de la covariance entre les régions dans le cerveau. Ainsi, nous proposons un design pour faciliter la comparaison de deux de ces réseaux, soit la comparaison de la connexité de deux patients, soit la comparaison dans l'évolution d'un seul cerveau au cours du temps.

La littérature propose peu de techniques pour la comparaison des graphes. Les techniques incluent les représentations simultanées (juxtaposées) ainsi que des représentations intégrées, utilisant une représentation qui encode les différences explicitement.

Dans ce chapitre, nous proposons plusieurs designs, basés sur des représentations noeuds-liens et matricielles, respectivement ; les représentations noeuds-liens sont plus facile à comprendre, les matrices ont l'avantage de montrer tous les connexions dans le réseau sans les cacher, et peuvent mieux encoder l'information sur des liens dans leurs cellules. La Figure 3 montre les designs les plus prometteurs pour chaque représentation (noeud-lien ou matrices). Dans la matrices, le carré intérieur dans chaque cellule figure le poids dans le graphe 1, la forme extérieure représente le poids dans le graphe 2. Les couleurs foncées indiquent un poids supérieur.

Les deux représentations de la Figure 3(a) ont été choisies suite à des discussions et des expériences pilotes. Ce chapitre discute en détail les différents essais, les raisons pour chaque design ainsi que la discussion pour faire une sélection finale.

Dans la troisième partie de ce chapitre, nous comparons les deux design choisis dans une étude quantitative qui montre que notre représentation matricielle permet aux utilisateurs d'accomplir les tâches liées à l'exploration des graphes des manières plus efficace.

Déplier les réseaux avec le *Matrix Cube* (Chapitre 6)

GraphDiaries montre comment des techniques différentes sont utilisées et intégrées dans une interface commune. Pour des réseaux complexes et denses, nous avons montré dans le chapitre 6 que les représentations matricielles sont supérieures aux représentations noeuds-liens. Les représentations matricielles peuvent donc facilement être intégrées dans des outils comme GraphDiaries. Une telle intégration pose le problème de la *complexité* de l'interface. Cette complexité augmente avec le nombre de représentations et visualisations, de fenêtres simultanées, ainsi qu'avec les possibilités d'interaction et le besoin de changer de techniques. La complexité de l'interface est diamétral à l'objectif d'une exploration générale des réseaux et leur complexité basé des vues complémentaires décrit comme *dépliage* des réseaux dans l'introduction de cette thèse.

Ce chapitre propose un modèle simple afin de créer des interfaces compréhensibles et efficaces. Nous proposons la structure d'une matrice tridimensionnelle, composée de matrices individuelles pour chaque instant temporel dans le réseau (Figure 4). Chaque cellule dans ce *Matrix Cube* correspond à une connexion entre deux noeuds à un instant spécifique.

Cependant, une visualisation tridimensionnelle est moins efficace que de la visualisations bidimensionnelles, à cause de la superposition des éléments graphiques, de la déformation du modèle de la projection perspective et difficulté de manipulation dans un espace 3D. Le cube peut néanmoins servir pour obtenir une idée générale du réseau et pour communiquer une *métaphore* de visualisation; explorer le cube comme on explorerait un cube physique permet d'explorer les données. L'espace des *opérations* pour

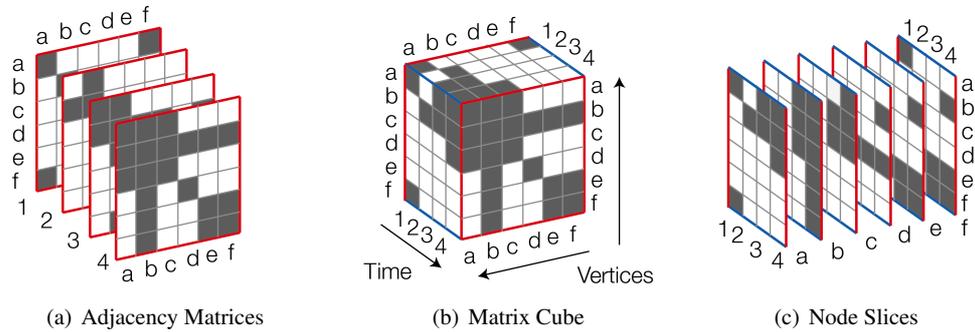


Figure 4: Le Matrix Cube. (a) Chaque instant dans le réseau (1,2,3,4) est représenté par une matrice individuelle. (b) Le cube est créé par l'empilement des matrices. Les dimensions rouges représentent les noeuds dans le réseaux; la dimension bleue correspond au temps. (c) découpage du cube le long d'une dimension rouge engendre des node slices.

	Time	Space		Time	Space
Extraction	Cell		Shift		
	Vectors	Time Vector Neighborhood Vector		Rotation	
Slices	Time Slice Vertex Slice	Filtering			
Projection	Time Projection		Vertex Projection	Time Coloring	

Figure 5: Opérations supportées par Cubix pour visualiser et décomposer les Matrix Cubes.

manipuler, décomposer et agréger le cube et ses parties est grand et décrit en détail dans le chapitre 8 (Figure 5). Le focus de ce chapitre est de créer une interface pour une telle exploration interactive et d'évaluer la compréhension ainsi que l'utilité de notre approche.

La Figure 6 montre des images obtenues par notre interface *Cubix*, montrant des signaux IRMf entre huit régions du cerveau. Les cellules sont colorées en fonction d'orientation des liens et leur taille indique la puissance du signal dans cet instant. Pour faciliter l'exploration du réseau, Cubix propose des *vues* pré-définies, telles que des small multiples et agrégations. D'autres vues sont créées par manipulation directe. Chaque vue dans Cubix est définie par les opérations mentionnées et communiquées par des transitions animées. Les figures (a)-(d) dans la Figure 6 montrent les mêmes données dans des vues

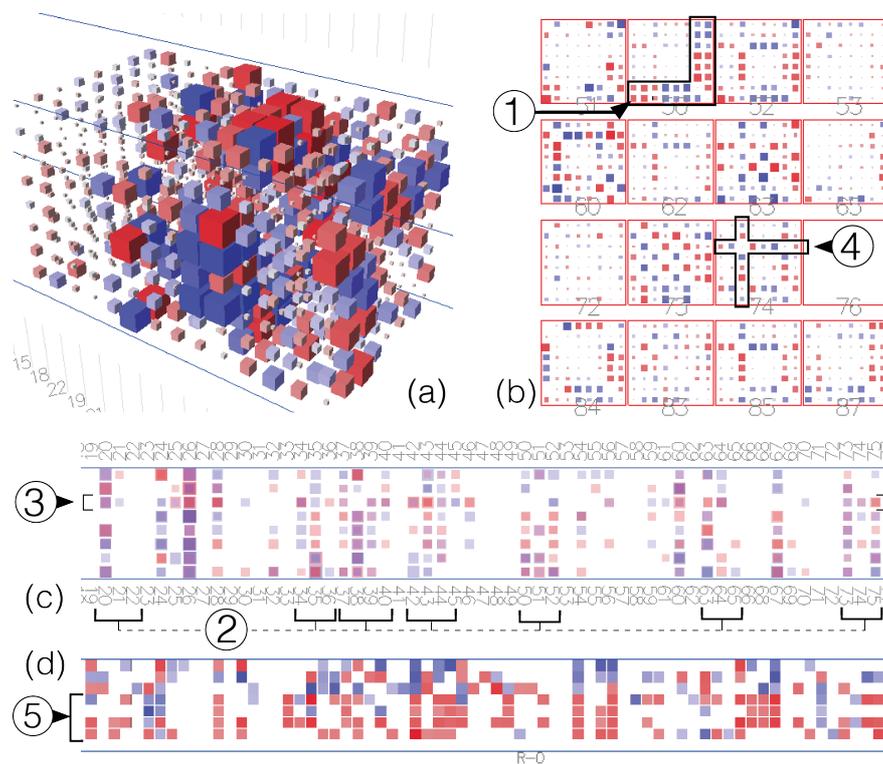


Figure 6: Connexité cérébrale pour 8 régions et 15 instants. (a) La vue globale montre que le réseau est plus actif à certains instants. (b) Vue détaillée de 16 instants, juxtaposées, (c) activité de tous les noeuds (superposés) au cours du temps (horizontalement), (d) connexions d'un seul noeud au cours du temps (également horizontalement).

différentes. Des annotations ①-⑤ montrent des motifs particuliers découverts dans les données, et difficilement visibles dans la représentation 3D.

Au-delà du cas illustré dans la Figure 6, nous avons utilisé Cubix pour visualiser et explorer des signaux entre 34 antennes de radio-télescopes ALMA, au Chili. Les changements dans les signaux indiquent des problèmes de calibration ainsi que des problèmes entre des paires d'antennes. Pour ces deux cas d'utilisation, télescopes et IRMf, nous avons travaillé avec des spécialistes en neurologie et astronomie. Nos observations ainsi que leur feedback montrent que le Matrix Cube est compréhensible et permet de visualiser des données complexes de manière nouvelle et puissante.

Des opérations pour déplier les cubes espace-temps (Chapitre 7)

Cubix implémente un certain nombre d'opérations, telles que la rotation, la coupe et la translation des pièces du cube, pour créer des vues (Figure 5). Ces opérations transforment le cube tridimensionnel dans des structures bidimensionnelles (plans, lignes,

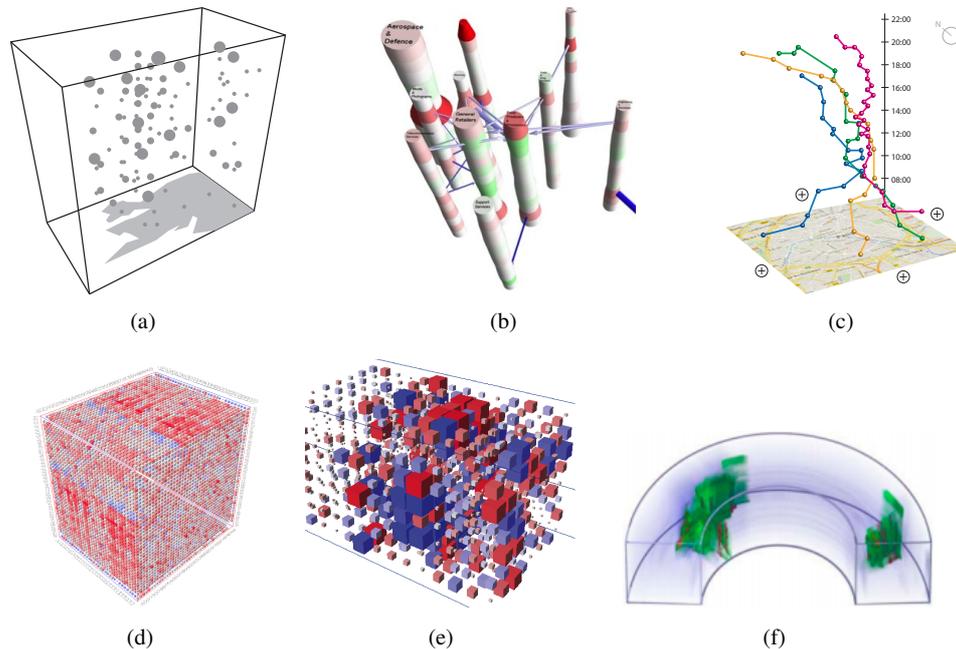


Figure 7: Exemples de cubes espace-temps montrant des données diverses: (a) Tremblements de terre (Andrienko & Andrienko, 2006), (b) Fund manager holdings (Dwyer & Eades, 2002), (c) Trajectoires des personnes dans une ville (Kapler & Wright, 2004), (d) signaux entre des antennes d'un radio-télescope (Cubix, Chapter 6), (e) connectivité cérébrale (Cubix, Chapter 6), (f) vidéo surveillance (Daniel & Chen, 2003).

points). Néanmoins, Cubix ne couvre pas l'ensemble des opérations. L'espace design pour des opérations est beaucoup plus large.

Cet chapitre propose une taxonomie structurée pour décrire les opérations sur des cubes espace-temps en général. Le cube peut être utilisé pour décrire des données diverses qui changent au cours du temps, tel que les données géographiques, les vidéos, ou les réseaux représentés par des représentations noeud-liens. Même si certaines des visualisations dans ces domaines utilisent une représentation explicite du cube (Figure 7), la plupart des visualisations sont bidimensionnelles et peuvent être décrites par des *opérations sur un cube espace-temps conceptuel*; le cube sert comme modèle de représentation pour les données d'un certain type et les opérations dérivent des visualisations finales.

Notre taxonomie des opérations considère des exemples parmi tous les domaines dont les données peuvent être représentées par des cubes espace-temps conceptuels. Figure 5 montre que en détail cette taxonomie qui pour l'instant décrit 41 opérations générales. Les opérations peuvent être composées de manière séquentielle pour décrire et créer des vues plus élaborées.

Notre taxonomie nous permet donc non seulement de décrire des visualisations existantes, mais aussi de créer des visualisations nouvelles. Un troisième but de cette taxonomie est d'informer les évaluations, c'est-à-dire de comparer les effets des opérations

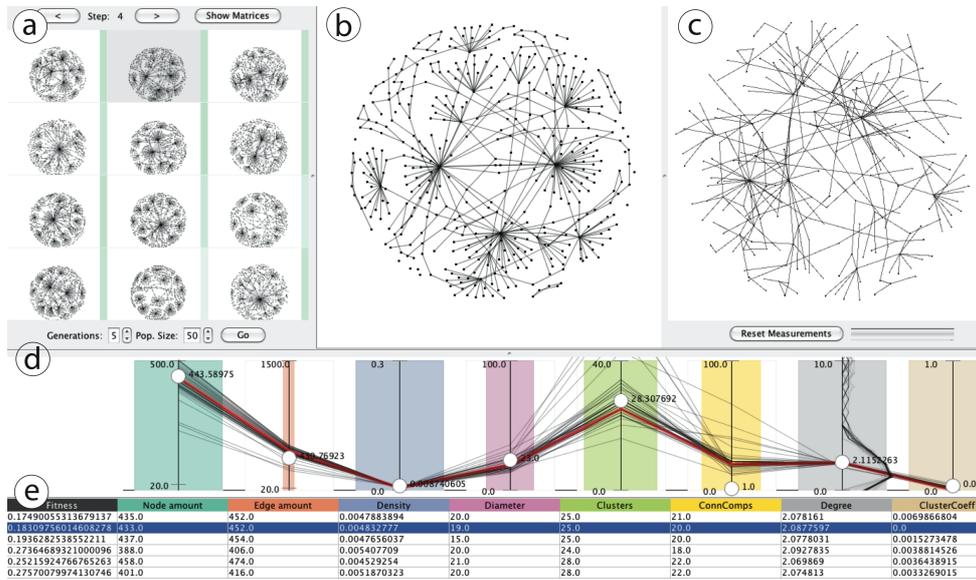


Figure 8: Interface de GraphCuisine: (a) différents réseaux générés (population view), (b) détail d'un réseaux choisi (detail view), (c) réseaux importé (pour des raisons de comparaison) (data set view), (d) parallel coordinates plot montrant des mesures des réseaux générés (mesures view) and (e) table montrant des mesures précises pour les réseaux générés.

dans les visualisations finales. Les opérations peuvent également servir à mieux décrire les évaluations existantes. Pour faciliter le choix et discussion des opérations, nous proposons aussi une description structurelle des cube espace-temps. Les exemples montrés dans la Figure 7 peuvent être décrites par a) leur densité, b) leurs objets géométriques, c) le degré de changement (variation) et d) le changement des attributs visuels tel comme la couleur de ses composantes. Ces caractéristiques des cubes sont essentielles pour choisir les opérations et créer des visualisations pertinentes.

En proposant cet espace de design pour les visualisations temporelles, nous pouvons mieux comprendre et discuter les différentes manières de visualiser des données complexes. Il nous sert également comme pont entre les diverses disciplines concernées avec ces mêmes problèmes et par conséquent de relier des idées, telles que l'expérience de l'utilisation des visualisations et opérations entre les domaines. Certaines des opérations pourront être aussi appliquées à des Matrix Cubes, mais aussi à des cubes résultants des représentations noeud-liens (Figure 7(b)).

Génération de graphes pour aider l'évaluation des visualisations (Appendix)

Evaluer des visualisations nécessite des données contrôlées, c'est-à-dire des réseaux qui correspondent à des critères spécifiques; nombre de noeuds, densité de connections, nombre de clusters. Comme les réseaux existants ne correspondent pas facilement à de tel critères, où ils ne sont pas disponibles en quantité suffisante, nous avons créé un générateur des réseaux interactif, appelé *GraphCuisine*.

GraphCuisine commence par proposer des réseaux arbitraires parmi l'utilisateur peut choisir des réseaux correspondant le mieux à ses critères. Un tel choix est fait par des valeurs précises de mesures de graphe, ou par des visualisations de ces mêmes graphs (Figure 8). A partir des choix faits par l'utilisateur, GraphCuisine calcule de nouveaux réseaux qu'il propose ensuite à l'utilisateur de manière itérative.

Conclusion

La question générale adressée dans cette thèse est "*Comment mieux comprendre des changements dans des réseaux dynamiques, à travers leur exploration visuelle et interactive?*". La thèse propose une catégorisation des techniques (Chapitre 2), puis définit une taxonomie de tâches pour mieux comprendre les besoins des utilisateurs (Chapitre 3). Chapitre 4 présente le système GraphDiaries qui combine plusieurs techniques diverses et qui peut être augmenté par des matrices telles que décrite dans le chapitre 5. Pour explorer l'espace design des matrices et pour créer une interface cohérente, le chapitre 6 décrit le système Cubix et le modèle des Matrix Cubes. Cette idée est généralisé aux cubes espace-temps, utilisés dans d'autres domaines, afin de structurer l'espace de design pour les visualisations temporelles en général.

La question initiale est loin d'être complètement résolue, mais cette thèse propose des techniques et des méthodes pour mieux comprendre le processus de visualisation des réseaux dynamiques.

References

- Ahmed, A., Dywer, T., Hong, S.-H., Murray, C., Song, L., & Wu, Y. X. (2005). Visualisation and analysis of large and complex scale-free networks. In *Proceedings of the seventh joint eurographics / IEEE VGTC conference on visualization* (pp. 239–246). Aire-la-Ville, Switzerland, Switzerland: Eurographics Association. doi: 10.2312/VisSym/EuroVis05/239-246
- Andrienko, N., & Andrienko, G. (2006). *Exploratory analysis of spatial and temporal data*. Springer.
- Boyandin, I., Bertini, E., & Lalanne, D. (2012). A qualitative study on the exploration of temporal changes in flow maps with animation and small-multiples. *Computer Graphics Forum*, 31(3pt2), 1005–1014. doi: 10.1111/j.1467-8659.2012.03093.x

- Brandes, U., & Nick, B. (2011, December). Asymmetric Relations in Longitudinal Social Networks. *IEEE Transactions on Visualization and Computer Graphics*, 17(12), 2283–2290. doi: 10.1109/TVCG.2011.169
- Collberg, C., Kobourov, S., Nagra, J., Pitts, J., & Wampler, K. (2003). A system for graph-based visualization of the evolution of software. In *Proceedings of the 2003 acm symposium on software visualization* (pp. 77–ff). New York, NY, USA: ACM. doi: 10.1145/774833.774844
- Daniel, G., & Chen, M. (2003). Video visualization. In *Proceedings of the 14th IEEE conference on visualization* (pp. 54–). Washington, DC, USA: IEEE Computer Society. doi: 10.1109/VISUAL.2003.1250401
- Dwyer, T., & Eades, P. (2002). Visualising a fund manager flow graph with columns and worms. In *Proceedings of IV* (p. 147-152).
- Kapler, T., & Wright, W. (2004). Geotime information visualization. In *Proceedings of IEEE information visualization* (p. 25-32). Washington, DC, USA: IEEE Computer Society. doi: 10.1109/INFOVIS.2004.27
- Shi, L., Wang, C., & Wen, Z. (2011). Dynamic network visualization in 1.5D. In *Proceedings of the 2011 IEEE pacific visualization symposium* (pp. 179–186). Washington, DC, USA: IEEE Computer Society.