



HAL
open science

LinkWave: une Liste d'Adjacence Visuelle Interactive pour explorer les Réseaux Pondérés Dynamiques

Nathalie Henry Riche, Yann Riche, Nicolas Roussel, Sheelagh Carpendale,
Tara Madhyastha, Thomas Grabowski

► To cite this version:

Nathalie Henry Riche, Yann Riche, Nicolas Roussel, Sheelagh Carpendale, Tara Madhyastha, et al.. LinkWave: une Liste d'Adjacence Visuelle Interactive pour explorer les Réseaux Pondérés Dynamiques. IHM'14, 26e conférence francophone sur l'Interaction Homme-Machine, Oct 2014, Lille, France. pp.113-122, 10.1145/2670444.2670461 . hal-01090425

HAL Id: hal-01090425

<https://hal.science/hal-01090425v1>

Submitted on 3 Dec 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LinkWave : une Liste d'Adjacence Visuelle Interactive pour explorer les Réseaux Pondérés Dynamiques

Nathalie Henry Riche
Microsoft Research
Redmond, WA, USA
nath@microsoft.com

Sheelagh Carpendale
University of Calgary
Calgary, Canada
sheelagh@ucalgary.ca

Yann Riche
Microsoft Surface
Redmond, WA, USA
yannr@microsoft.com

Tara Madhyastha
University of Washington
Seattle, WA, USA
madhyt@u.washington.edu

Nicolas Roussel
INRIA
Villeneuve d'Ascq, France
nicolas.roussel@inria.fr

Thomas J. Grabowski
University of Washington
Seattle, WA, USA
tgrabow@u.washington.edu



Figure 1: Représentation de la connectivité de 23 régions du cerveau regroupées en 4 groupes fonctionnels lors d'un scan de 5 minutes (300 points). Le diagramme noeuds-liens (à gauche) représente la connectivité à un seul instant donné. La liste d'adjacence visuelle (au milieu et à droite) que nous proposons représente l'évolution temporelle des poids de chacun des liens du réseau. Ces liens peuvent être agrégés pour permettre la découverte de motifs de connexion entre groupes de noeuds. Cette représentation est au coeur du système LinkWave, permettant aux utilisateurs d'explorer interactivement des ensembles de liens. Les étoiles indiquent une période où le poids d'un grand nombre de liens a diminué. La vue agrégée permet de discerner que le poids des liens colorés en turquoise (liens du groupe DMN) n'ont pas suivi ce motif.

RÉSUMÉ

Nous présentons LinkWave, un système pour visualiser interactivement des réseaux pondérés dynamiques. LinkWave est basé sur le concept simple d'une liste visuelle de liens qui peuvent être triés, filtrés et agrégés. LinkWave a été développé en collaborations avec des chercheurs en neurosciences pour analyser les réseaux de connectivité du cerveau issus de données IRMf. Nous expliquons la démarche de conception du système et rendons compte des commentaires des neuroscientifiques impliqués et des premières découvertes réalisées avec le système.

Mots Clés

IRMf ; connectivité fonctionnelle cérébrale ; réseaux pondérés dynamiques ; visualisation d'information

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous.

INTRODUCTION

La recherche en neurosciences s'intéresse de près aux mécanismes cognitifs et neuraux qui sous-tendent les processus complexes de la pensée et du raisonnement^{1, 2}. Une avancée majeure dans ce domaine est l'*imagerie par résonance magnétique fonctionnelle* (IRMf) qui permet d'étudier *in vivo* de manière indirecte l'activité cérébrale. De nombreux chercheurs utilisent les données d'IRMf pour étudier les interactions entre différentes régions du cerveau [29]. Ils modélisent pour cela l'activité cérébrale sous forme de réseaux complexes, les noeuds de ces réseaux correspondant aux régions du cerveau et des liens pondérés dynamiques indiquant la force d'interaction entre deux régions à un instant donné. Les données d'IRMf sont cependant à la fois complexes et volumineuses, ce qui rend leur analyse difficile.

© ACM, 2014. This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in Actes de la 26^{ième} conférence francophone sur l'Interaction Homme-Machine, 2014.

<http://dx.doi.org/10.1145/2670444.2670461>

1. <http://www.nih.gov/science/brain/>
2. <https://www.humanbrainproject.eu>

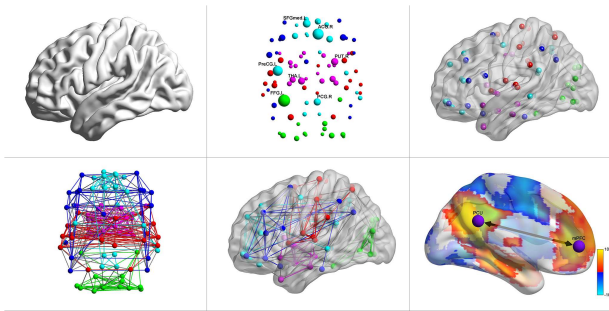


Figure 2: Le fonctionnement du cerveau représenté par des diagrammes nœuds-liens entre régions. Six représentations 3D générées par *BrainNet Viewer* /citebrainnetviewer

L'identification de motifs dans ces réseaux dynamiques est un enjeu important pour la compréhension générale du cerveau et pour l'identification et la compréhension de phénomènes particuliers tels que la maladie d'Alzheimer. Or, cette tâche s'avère extrêmement difficile de par sa nature exploratoire lorsqu'on se limite aux outils d'analyse statistique classiques tels que l'environnement R. Quelques outils de visualisation tel que *BrainNet Viewer* [33] proposent de représenter les interactions déterminées à partir des IRMf en utilisant la représentation la plus commune des réseaux : les diagrammes nœuds-liens. Le placement des nœuds respecte généralement la morphologie du cerveau, chaque nœud étant virtuellement positionné sur la région qu'il représente. Cependant, les représentations nœuds-liens sont généralement difficile à lire pour ces réseaux denses [18], et ce placement diminue plus encore la lisibilité puisqu'il ne permet pas de minimiser le croisement des liens (Figure 2).

De plus, l'aspect dynamique des réseaux est difficile à représenter, les approches les plus courantes consistant à calculer une moyenne des poids de connexion dans le temps, ou à générer une animation omettant éventuellement les liens de poids inférieur à un seuil pour diminuer l'encombrement visuel. S'ils offrent une vue d'ensemble des connexions fonctionnelles, les outils de visualisation existants ne permettent alors pas aux scientifiques d'analyser l'évolution des poids de connexion dans le temps et d'en extraire des motifs. Or, ces tâches sont critiques pour les chercheurs qui tentent de comprendre l'impact des maladies dégénératives telles que Alzheimer ou Parkinson. Une des hypothèses concernant ces maladies est en effet qu'elles dégradent l'interaction entre certaines régions du cerveau, ce qui pourrait se traduire par une modification de certains poids de connexion ou de leur séquençage dans le temps.

Dans cet article, nous décrivons un système d'exploration visuelle destiné à faciliter l'identification de motifs dans les réseaux dynamiques pondérés, appliqué ici au réseaux formés par les zones d'activité du cerveau humain et leurs connexions. Ce système, *LinkWave*, repose sur une représentation qui met en avant l'évolution des poids des liens du réseau dans le temps et les représentent comme des séries temporelles (Figure 1). *LinkWave* offre un nouveau paradigme d'exploration des réseaux dynamiques dans lequel l'utilisateur interagit principalement avec les liens et

non les nœuds du réseau. Nous offrons des interactions simples qui permettent de trier, agréger et comparer ces liens les uns aux autres pour extraire des motifs temporels. Nous expliquons la démarche de conception du système de plus de deux ans avec des chercheurs en neurosciences de l'Université de Washington. Nous rendons également compte des premières découvertes réalisées par ces chercheurs avec *LinkWave* concernant des motifs liés à l'âge du sujet ou à des maladies dégénératives telles que les maladies de Parkinson ou d'Alzheimer.

ANALYSE DES BESOINS

Pour informer la conception de *LinkWave*, nous avons formé une collaboration avec l'équipe de neuroscience Integrated Brain Imaging Center (IBIC) de l'Université de Washington. Ces chercheurs étudient les dégradations des connexions dans le cerveau humain dues à certaines maladies dégénératives comme la maladie de Parkinson ou d'Alzheimer, issues de traumatismes ou de chirurgies, ou simplement dues au vieillissement naturel. Ils analysent des données IRMf comparant les différents motifs de connexion extraits de leurs sujets à un groupe contrôle. Après une série d'entretiens avec six neuroscientifiques (rencontres bi-mensuelles, de deux heures environ, pendant six mois) et l'observation de leurs pratiques existantes d'analyse avec les outils statistiques et visuels traditionnels, nous avons formalisé une liste de tâches typiquement réalisées sur les graphes pondérés dynamiques produits à partir d'IRMf.

Caractéristiques des Données

Les données IRMf mesurent l'activité des régions cérébrales de manière indirecte, via leur consommation d'oxygène. Les neurologues nettoient ces données pour éliminer les bruits provenant des éventuels mouvements du sujet pendant l'acquisition et d'autres facteurs physiologiques tels que la respiration. Dû à la présence de fluide cérébrospinal et à des parties de matière blanche présents dans les images IRMf mais qui ne représentent pas l'activité neuronale, les valeurs des signaux brutes à un moment donné ne sont pas comparables aux valeurs brutes à un autre moment dans le temps. L'information que les scientifiques utilisent est donc toujours une information sur le niveau d'activation d'une région par rapport aux autres à un moment donné. A cet effet, nous calculons des corrélations de Pearson entre régions d'intérêt à différents points dans le temps (à noter que d'autres méthodes existent incluant l'utilisation de modèles statistiques). Le résultat de ces calculs est un réseau pondéré dynamique dont il est généralement supposé que la dynamique au cours du temps reflète le fonctionnement du cerveau.

Les réseaux produits possèdent très peu de nœuds, d'une vingtaine à quelques centaines, chaque nœud pouvant appartenir à un groupe fonctionnel. Par exemple, les réseaux que nos collaborateurs analysent pour identifier des motifs associés à la maladie de Parkinson sont constitués de 23 nœuds faisant partie de 4 groupes fonctionnels différents. Ces groupes représentent les régions qui interagissent ensemble pour accomplir une certaine fonction. Par exemple le *Default Mode Network* (abrégé DMN et coloré en turquoise sur les figures) est un groupe de régions qui interagit lorsque le cerveau est au repos. Ce groupe est un

des premiers affecté chez les sujets atteints de la maladie de Parkinson. Il est utile de noter que le terme “réseau” (“network” en anglais) utilisé en neurosciences ne représente pas le réseau complet mais un groupe de nœuds. Les groupes fonctionnels résultent de nombreux travaux de recherche et sont identifiés par les neurologues a priori.

Le réseau présente une forte densité de liens à chaque instant donné. Le réseau agrégé sur l'intervalle de temps total est un graphe complet (chaque nœud est connecté à chacun des autres nœuds à un moment donné). Cette forte densité est source d'occultation dans les diagrammes nœuds-liens, ce qui nécessite l'utilisation de représentations alternatives telles que les matrices d'adjacence. Le nombre d'instantanés considérés est relativement important, de l'ordre de 300. Ce nombre rend l'utilisation de “*small multiples*” en pratique très difficile.

Formalisation des Tâches

A partir de ces entretiens, nous formalisons une liste de 7 tâches typiquement réalisées sur les graphes pondérés dynamiques extraits des données IRMF de nos collaborateurs.

T1. Caractériser la connectivité entre deux nœuds

Cette tâche consiste à caractériser l'évolution dans le temps de la force de connectivité entre deux nœuds. Il est par exemple important d'identifier les augmentations de poids d'un lien et de considérer leur intensité et leur fréquence.

T2. Caractériser la connectivité d'un nœud

Cette tâche consiste à caractériser l'évolution dans le temps de la connectivité d'un nœud. Il est par exemple important d'identifier un groupe de nœuds auxquels celui-ci est fortement connecté en moyenne, et d'identifier si ces nœuds varient à certains moments.

T3. Identifier des groupes de nœuds similaires

Cette tâche consiste à identifier des nœuds qui présentent des motifs de connexions similaires. Ces nœuds peuvent être considérés similaires de par leur connectivité à des mêmes nœuds et/ou en rapport à l'évolution des poids de leurs liens.

T4. Identifier des cas particuliers

Cette tâche consiste à identifier des nœuds (ou des liens) dont les motifs de connectivité diffèrent des autres. Il est important d'identifier si ces exceptions ont trait à la topologie (différents voisins), à l'évolution des poids de connexion (poids différents du groupe), ou sont dus au séquençement (délai de connexion par rapport au groupe).

T5. Caractériser la connectivité entre plusieurs groupes

Cette tâche consiste à caractériser la force de connexion entre des groupes de nœuds. Identifier les nœuds qui jouent un rôle particulier pour ces groupes, un nœud pivot par lequel passeraient la majorité des connexions entre deux groupes par exemple. Évaluer l'évolution temporelle de ces rôles est également important.

T6. Identifier des états du graphe

Cette tâche consiste à identifier des motifs de haut niveau qui permettent de caractériser un état du réseau. Par

exemple, un état extrême (et peu courant) est un réseau complet, où tous les nœuds sont connectés, ou bien un réseau déconnecté, où aucune connexion n'apparaît. En réalité, ces états sont bien plus nuancés. Ils peuvent être définis par un intervalle de temps dans lequel les nœuds du réseau semblent suivre un motif de connexion particulier. Idéalement, ces états peuvent se répéter dans le temps et il est important de pouvoir les dissocier les uns des autres.

T7. Comparer plusieurs graphes pondérés dynamiques

Étant donné que la structure anatomique du cerveau varie d'un individu à un autre, on peut estimer que la structure fonctionnelle du cerveau va également présenter des variations entre individus. Il est important de pouvoir comparer les motifs entre individus et d'identifier les motifs communs ainsi que leurs variations.

LINKWAVE

Pour répondre aux besoins de nos collaborateurs en neurosciences, nous avons conçu une visualisation interactive qui permet d'accomplir les tâches décrites ci-dessus. Les travaux de recherche sur la visualisation de graphes s'intéressent essentiellement à des tâches centrées sur les nœuds et groupes de nœuds [23]. À l'inverse, les tâches identifiées pour les neuroscientifiques mettent en avant les liens du réseau, et non ses nœuds. Si peu de travaux en visualisation se sont concentrés sur ces questions, elles sont pourtant importantes pour l'analyse de réseaux en général et applicables à d'autres domaines. Par exemple, les chercheurs en sciences humaines et sociales accomplissent des tâches similaires lorsqu'ils analysent les communications par courriers électroniques ou des réseaux sociaux. Dans ces réseaux la force de connexion peut être liée au nombre de messages échangés et identifier les motifs temporels (la fréquence et le rythme général de communication) est crucial à l'analyse des comportements sociaux.

Aider les utilisateurs à accomplir les tâches ci-dessus nécessite un changement de perspective dans lequel *les liens du réseau deviennent les éléments centraux de la représentation visuelle*. Par exemple, analyser l'évolution de la connexion entre deux groupes de nœuds revient à analyser l'évolution des poids des liens connectant les deux

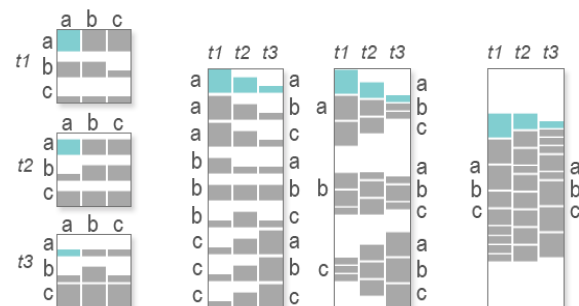


Figure 3: De gauche à droite : matrices d'adjacence pour représenter l'évolution d'un réseau pondéré à trois instants donnés, liste d'adjacence visuelle dans laquelle chaque ligne représente l'évolution des poids d'un lien, agrégation de cette liste illustrant la connectivité de chaque nœud, agrégation de cette liste illustrant la connectivité totale du réseau.



Figure 4: Capture d’écran de la fenêtre principale de LinkWave lors de l’exploration de la connectivité d’un sujet atteint de la maladie de Parkinson (à gauche) et d’un sujet du groupe contrôle (à droite). Les liens vers la région latérale temporale du groupe DMN (lattemp) sont agrégés par leurs sources (régions des groupes DAN, FPT et SAL). Chaque paire de lien d’une couleur représente la connexion à la région lattemp de l’hémisphère droit (DMNRlattemp) et gauche (DMNLlattemp). Alors que ces liens ont des motifs similaires chez le sujet contrôle, la symétrie est altérée chez le sujet avec Parkinson. L’étoile indique une de ces différences.

groupes. LinkWave est une visualisation interactive permettant d’explorer ce nouveau paradigme d’exploration des réseaux dynamiques. Le principe de cette visualisation est illustré par la Figure 3. Il consiste à représenter le réseau dynamique complet par une liste d’adjacence : une table dans laquelle chaque ligne représente un lien du réseau et chaque colonne un point dans le temps. Une cellule dans cette table représente le poids porté par un lien (ligne) à un instant donné (colonne). De fait, une ligne de la table encode l’évolution temporelle de la force de connexion entre deux nœuds. La représentation de l’évolution temporelle des poids des liens dans une représentation unique rend possible la comparaison de plusieurs liens et permet de comparer les rythmes de leur augmentations ou réductions de poids. Bien que les listes d’adjacence soient couramment utilisées pour décrire les données brutes des réseaux, à notre connaissance, ces structures n’ont jamais été utilisées comme outils de visualisation et d’exploration.

Représentation visuelle

LinkWave présente l’évolution des liens pondérés du réseau sous forme de liste. Il s’agit donc de représenter visuellement plusieurs séries de données temporelles. La littérature sur ce sujet est abondante et plusieurs représentations sont possibles. Les représentations les plus communes sont celles à base de lignes (reliant les points), de barres (une par point) ou d’aires (en partie délimitées par les points). Considérant que les poids des liens sont le résultat d’un calcul de corrélation sur une fenêtre glissante, une représentation lissant les changements pour indiquer une tendance d’évolution plutôt que des valeurs précises

semble préférable. Nous nous sommes donc intéressés aux représentations à base de lignes et d’aires. Considérant le grand nombre d’éléments à représenter et la prédisposition humaine à percevoir des motifs symétriques [14], nous nous sommes intéressés aux graphiques d’aires symétriques. Nous avons également exploré des représentations plus compactes tels que les heatmap et graphes d’horizons, mais celles-ci ont été critiquées par nos utilisateurs en raison de la difficulté à discerner les variations de couleur et d’opacité et, plus généralement, à identifier des motifs saillants.

Etant donné que chaque série temporelle représente l’évolution du poids d’un lien du réseau, un noeud source (placé à gauche de l’aire) et un noeud destination (placé à droite) y sont donc associés. Ces informations topologiques peuvent être utilisées pour agréger les liens et révéler la connectivité à un plus haut niveau d’abstraction. Par exemple, agréger les liens qui ont le même noeud source permet de visualiser la connectivité générale de chacun des noeuds du réseau.

Dans les données visualisées dans la Figure 4, nous employons seulement trois niveaux d’agrégation possibles pour la source et/ou la destination des liens : un noeud, un groupe de noeud, ou tous les noeuds du réseau. Cela revient à attacher une hiérarchie de trois niveaux à droite et cette même hiérarchie à gauche, résultant en 9 combinaisons (noeud-noeud ; noeud-groupe ; noeud-tous ; groupe-noeud ; groupe-groupe ; groupe-tous ; tous-noeud ; tous-groupe ; tous-tous). Ces agrégations peuvent être généralisées à une hiérarchie de profondeur arbitraire. Pour représenter ces agrégations, nous utilisons des *stream-*

graphs. Nous n'avons pas choisi cette représentation pour permettre une analyse détaillée des diverses couches, mais pour leur effet visuel qui permet de rappeler à l'utilisateur qu'il s'agit d'une agrégation de liens et non d'un objet unique.

Interactions

Les liens du réseau sont exponentiellement plus nombreux que les noeuds. En effet, un réseau complet de n noeuds comporte $n \times (n - 1) \div 2$ liens non dirigés. Le plus petit réseau que nous avons étudié comporte 23 noeuds et 253 liens évoluant sur 280 points dans le temps. Le principal défi est donc de proposer des techniques d'interaction permettant le passage à l'échelle de la visualisation et son utilisation sur un (ou deux) écran(s) d'ordinateur standard(s). LinkWave offre aux utilisateurs des interactions standards sur les données temporelles tels que zoomer, trier, filtrer et agréger. Ces interactions ont été également adaptées pour explorer la topologie du réseau et permettre aux utilisateurs de spécifier un ensemble particulier de liens.

LinkWave permet le tri des séries temporelles selon des valeurs à des instants spécifiques ou à partir de mesures statistiques simples (minimum, maximum, médiane). Le tri peut être également spécifique à la topologie du réseau, en réordonnant la liste alphabétiquement par les noms des noeuds ou des groupes source ou destination. Le filtrage est également adapté à l'exploration de réseaux. Il est par exemple possible d'explorer la connectivité intra-groupe (en filtrant les autres groupes en sources et destinations des liens), ainsi que la connectivité inter-groupes (en spécifiant un groupe donné en source et un groupe différent en destination). La visualisation interactive du réseau s'effectue donc en interagissant avec différents ensembles de liens au travers du panneau de contrôle situé en haut de la fenêtre LinkWave (Figure 4).

Scenario d'utilisation

Pour illustrer le type de découvertes que permet LinkWave, nous présentons un court scenario d'utilisation. Emma souhaite explorer les données issues de son étude avec des sujets atteints de la maladie de Parkinson. Elle charge un premier sujet et isole les régions du groupe fonctionnel DMN (régions actives à l'état de repos du sujet), que la littérature scientifique désigne comme particulièrement affecté par la maladie. Elle utilise les boutons de filtrage de LinkWave pour isoler les 117 liens (sur un total de 506) dont la destination est une région du groupe DMN. Elle effectue ensuite un tri par le poids moyen de chaque lien pour avoir un aperçu des connexions les plus fortes chez ce sujet. Elle note que les 9 liens dont la moyenne des poids est la plus élevée sont des liens internes de DMN (la source et la destination sont des régions de DMN). Intriguée, elle se demande si le tri serait similaire chez un sujet du groupe contrôle. Elle copie la configuration de la vue d'un clic et charge les données d'un sujet contrôle. Cette fois, seuls 3 liens internes du groupe DMN apparaissent dans le top 10 (Figure 5).

Encouragée par la confirmation d'une différence entre les motifs de connexion intra-groupe DMN, Emma décide d'explorer s'il existe une différence entre les liens

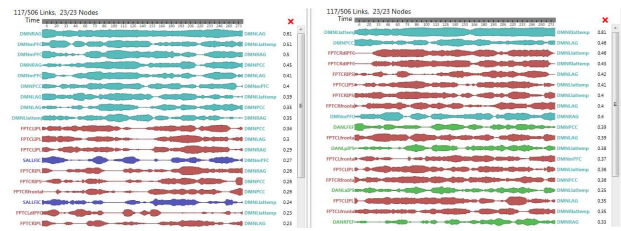


Figure 5: Listes de liens pondérés ordonnée par le poids moyen pour un sujet atteint de Parkinson (à gauche) et un sujet contrôle (à droite). Les liens internes au groupe DMN sont colorés en bleu.

inter-groupes. Elle filtre les sources des liens qui appartiennent au groupe DMN pour éliminer les liens intra-groupe. Elle agrège ensuite les noeuds sources des 102 liens restant par leurs 3 groupes respectifs (FPT, DAN et SAL). En étudiant l'évolution de ces 18 liens agrégés, elle identifie des points de synchronisation particulièrement salients dans le sujet avec Parkinson. Emma note également que certains liens semblent avoir un motif différent des autres. Il s'agit des liens avec la région latérale temporale (lattemp) au sein du groupe DMN. Elle effectue une recherche sur ce libellé pour ne garder que les 6 liens correspondants chez les deux sujets. Chaque paire de liens représente les connexions agrégés des régions d'un groupe fonctionnel à la région lattemp dans l'hémisphère droit (DMNRIlattemp) et gauche (DMNLIlattemp) du cerveau. Le sujet contrôle exhibe des motifs similaires pour chacune de ces paires alors que cette symétrie est altérée chez le sujet avec Parkinson (Figure 4). Cette découverte semble confirmer l'hypothèse que certaines régions du cerveau se désynchronisent due à la maladie de Parkinson. Cependant, bien d'autres explorations et analyses statistiques seront nécessaires pour accepter ou rejeter définitivement cette hypothèse.

RETOUR D'EXPÉRIENCE

Méthodologie

Afin d'itérer sur l'utilisabilité de LinkWave et de comprendre si ce système peut permettre de faire des découvertes lors de l'exploration d'IRMF, nous avons suivi la méthode d'évaluation longitudinale MILCS proposée par Shneiderman et Plaisant en 2006 [28]. Cette méthode ethnographique consiste à collaborer étroitement avec une poignée d'utilisateurs potentiels dans les conditions les plus réalistes possibles (utilisant leurs données et leur cadre de travail). LinkWave a été utilisé pendant trois mois par deux chercheurs de l'Université de Washington (occasionnellement par leurs étudiants). Les scientifiques ont installé une copie du logiciel et ont conduit leurs explorations seuls. Nous leurs avons demandé de documenter leurs découvertes en faisant des captures d'écran et en utilisant les points de sauvegarde du logiciel. Enfin, deux fois par mois, pendant deux heures environ, nous avons discuté avec eux de leurs explorations faites avec LinkWave, des éventuels problèmes rencontrés et des évolutions possibles de notre système.

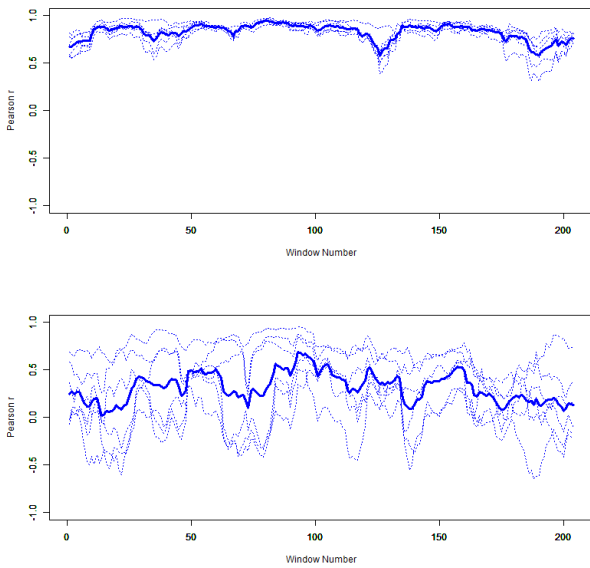


Figure 6: Graphiques des corrélations entre plusieurs régions du cerveau générés avec R pour deux sujets d'âges différents. Le sujet de 64 ans est en haut, celui de 88 ans est en bas et démontre une plus forte variation des corrélations.

Pratique usuelle

Nos collaborateurs neuroscientifiques cherchent à identifier des motifs dans les IRMf afin de mieux comprendre les déficits associés à la maladie de Parkinson. Les réseaux pondérés dynamiques qu'ils ont analysés avec LinkWave étaient ceux sur lesquels ils travaillent habituellement, extraits de données IRMf et "nettoyés" selon les procédures généralement recommandées par leur communauté de recherche. Les liens pondérés représentent dans ces réseaux les corrélations entre certaines régions d'intérêt (ROIs) qu'ils ont eux-même spécifié. Les outils de visualisation existants sont peu nombreux et présentent généralement le réseau dans le contexte de la structure anatomique du cerveau, représentée en 3D (Figure 2). Les croisements de liens rendent alors difficile l'identification de motifs.

La pratique d'analyse actuelle de nos collaborateurs neuroscientifiques est essentiellement fondée sur la formation d'hypothèses et leur acceptation ou réfutation sur la base de tests et modèles statistiques plus ou moins avancés. Durant leurs analyses, nos collaborateurs ont généré un certain nombre de graphiques simples utilisant les environnements tels que R (voir Figure 6 par exemple). Ils représentent dans une même vue toutes les corrélations entre ROIs (*i.e.* tous les liens pondérés) pour deux participants différents. Ces simples graphiques (non interactifs) leur ont déjà permis de découvrir des motifs intéressants telle qu'une plus forte variabilité des corrélations chez les sujets plus âgés. D'autres découvertes de ce genre clairement liées à la visualisation des données ont amené nos collaborateurs à former de nouvelles hypothèses et à effectuer des analyses qu'ils n'avaient pas envisagées avant. Cette prise de conscience du potentiel lié à la visualisation a par la suite joué un rôle moteur dans notre collaboration.

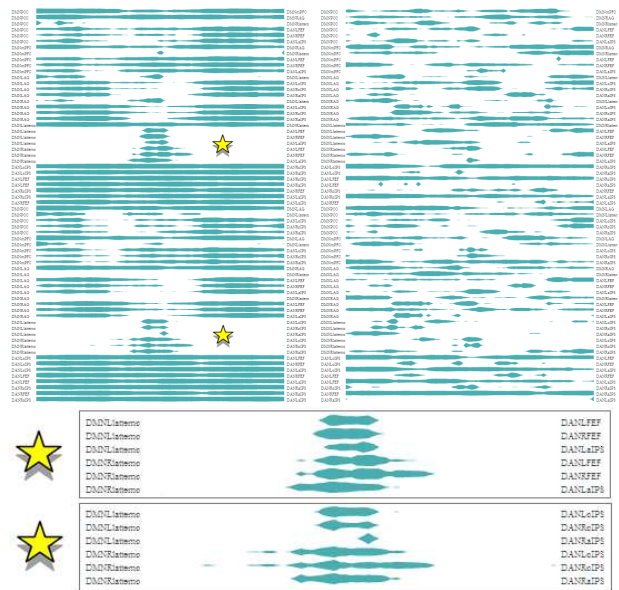


Figure 7: Représentations générées avec LinkWave pour les mêmes données que les graphiques de la Figure 6. Le sujet de 64 ans est à gauche, celui de 88 ans est à droite. La forte variation des poids des liens est identifiable par les variations des diagrammes en aires sur l'image de droite. L'étoile indique un nouveau motif découvert par nos collaborateurs. Il s'agit des liens issus de la région latérale temporelle (latterp) dont l'évolution des poids diffèrent des autres.

Utilisation de LinkWave

Nous avons conçu et implémenté un prototype de LinkWave (Figure 4) puis raffiné les interactions proposées de manière itérative avec les neuroscientifiques sur un intervalle de trois mois. Durant ces trois mois, les 3 itérations majeures ont concerné la paramétrisation des calculs des fenêtres de corrélation (afin qu'ils puissent contrôler la taille de ces fenêtres), l'implémentation de plusieurs algorithmes de tri de la liste visuelle des liens, et l'ajout de vues synchronisées avec LinkWave (vue matricielle et vue noeuds-liens). Nous leur avons ensuite livré le logiciel qu'ils ont pu utiliser pendant les trois mois suivant pour analyser leurs données. Dans ce qui suit, nous décrivons les principaux commentaires et découvertes faits avec le système par les neuroscientifiques.

Confirmation de découvertes précédentes

Durant les premières utilisations de LinkWave, nos collaborateurs ont tout d'abord cherché à retrouver un certain nombre de découvertes réalisées auparavant. La première analyse fut faite pour mettre en avant la différence de synchronisation due à l'âge des sujets. Nos collaborateurs avaient en particulier généré avec R des graphiques mettant en évidence la variation des corrélations chez les sujets les plus âgés (Figure 6). Utilisant ces mêmes données dans LinkWave, nos collaborateurs ont créé les représentations présentées dans la Figure 7. La différence de variabilité des poids est directement reflété dans LinkWave mais ils ont également fait une seconde découverte. Les liens d'une des régions chez le sujet moins âgé présentent un motif différent des autres, ce qui n'était pas visible dans les graphiques précédents. Intrigués par cette région, ils explorèrent alors des données de participants moins âgés. A ce jour, ils continuent de poursuivre cette

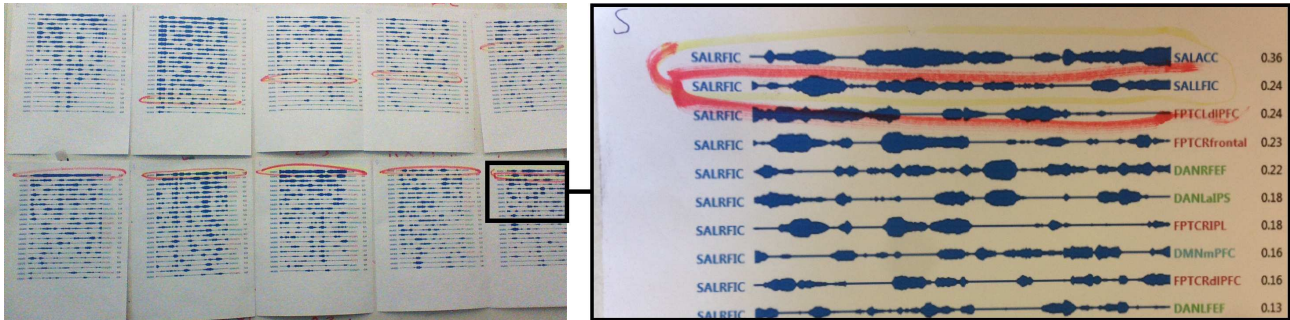


Figure 8: Impréssions annotées de la fenêtre principale de LinkWave. La rangée du haut présente un sous-ensemble de liens pour cinq participants du groupe contrôle ; la rangée du bas ce même sous-ensemble pour cinq participants du groupe de Parkinson. Les liens sont ordonnés par leur intensité moyennes, le lien entouré semble avoir un poids plus important dans le groupe de Parkinson.

direction de recherche. Cette nouvelle découverte leur a donc permis de faire de nouvelles hypothèses sur des données analysées auparavant.

Nos collaborateurs ont aussi cherché à reproduire la découverte d'une différence entre les sujets d'un groupe de contrôle et les sujets atteints de la maladie Parkinson. En réordonnant les liens par leur poids moyen et en comparant les sujets des deux groupes à l'aide de versions papier de visualisations LinkWave (Figure 8), ils ont identifié un lien spécifique entre deux ROIs qui s'avère beaucoup plus fort dans les sujets avec Parkinson que dans les sujets du groupe de contrôle. Ce lien est entouré sur les documents imprimés et apparaît en haut de la liste. Ce résultat n'était pas surprenant au vu de leur analyse initiale, mais il furent surpris qu'une interaction de quelques minutes mette en évidence ce motif découvert après plusieurs jours d'analyses statistiques. LinkWave leur permit en outre de continuer leur exploration et les mécanismes de filtrage par noeud source et destination leur permit d'inspecter systématiquement les liens de ces deux ROIs au reste du réseau, pour essayer d'identifier des motifs distinctifs entre le groupe contrôle et le groupe Parkinson.

Remise en question du seuillage des liens pondérés

Comme mentionné précédemment, il est commun de réduire le nombre de liens du réseau en filtrant les liens dont le poids est inférieur à une valeur donnée. Ce principe de seuillage est particulièrement utile pour les techniques de visualisation basées sur les diagrammes noeuds-liens car il peut améliorer leur lisibilité. Comme LinkWave ne bénéficie pas directement du filtrage des liens à un instant donné, nous n'avons pas implémenté ce seuillage. Cette décision a permis à nos collaborateurs de découvrir que, dans certains cas, des motifs intéressants émergent lorsque tous les poids sont conservés. La Figure 9 illustre ce type de découverte. Ces explorations avec LinkWave ont donc permis à nos collaborateurs de remettre en question cette pratique de seuillage et de considérer le recours à l'exploration visuelle en amont de l'analyse pour mieux comprendre l'impact du nettoyage des données.

Découverte de points de synchronisation

La découverte la plus importante faite par nos collaborateurs avec le système est l'identification de points temporels pour lesquels les poids de nombreux liens augmentent ou diminuent. Ces "points de synchronisation" se traduisent par des lignes verticales visibles à tous les niveaux

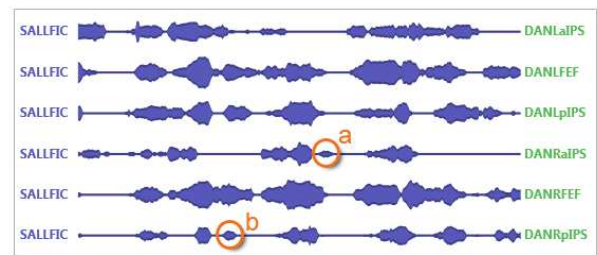


Figure 9: Filtrer les poids faibles peut permettre de réduire le bruit mais peut aussi cacher certains motifs. Dans cet exemple le motif annoté (a) est certainement du bruit. Par contre ; le motif annoté (b) peut se révéler être un motif car il fait écho à un motif dans les deuxième, troisième et cinquième lignes.

d'agrégation et illustrés par des flèches sur la Figure 10. Bien que ce type de synchronisation soit une hypothèse discutée par les neuroscientifiques depuis plusieurs années, ils étaient extrêmement surpris et enthousiastes lors de cette découverte, demandant aux autres chercheurs et étudiants présents dans le laboratoire de venir voir les visualisations sur leurs écrans.

Au cours des discussions qui suivirent leur réaction fut de vérifier que ces motifs ne soient pas dus à un bogue du prototype ou à un artefact durant l'acquisition des données. Après plusieurs analyses complémentaires, ils ne furent pas en mesure de corrélérer ces motifs avec les informations physiologiques collectées lors de l'acquisition (mouvements du sujet, respirations, etc.). A ce jour, nos collaborateurs continuent leurs analyses de ce phénomène de manière plus systématique pour comprendre ce que représentent ces points de synchronisation et leur importance dans le fonctionnement du cerveau.

Découverte de séquençement

Un autre type de motif qui intéresse nos collaborateurs est illustré dans la Figure 10. Il s'agit de séquençements possibles dans les forces de connexions entre deux groupes de liens. Ces séquençements peuvent indiquer un rôle particulier ou une synchronisation entre des régions d'intérêt. Ces motifs, difficiles à identifier dans la représentation individuelle des liens, deviennent apparents après agrégation noeud-groupe.

En résumé, l'utilisation du prototype LinkWave a permis à

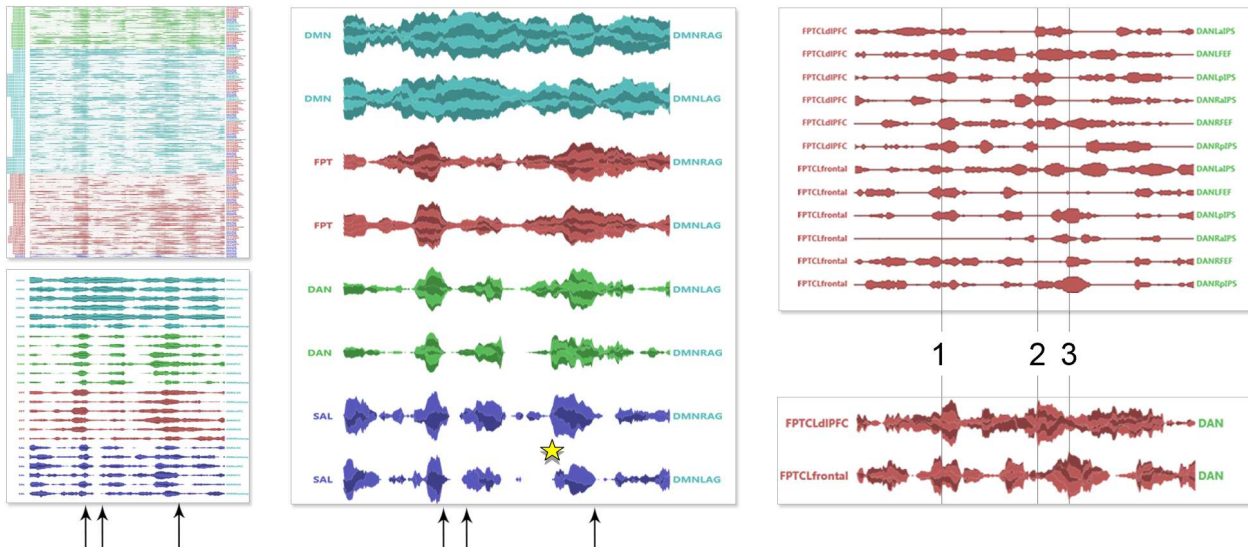


Figure 10: Les points de synchronisation sont annotés par des flèches dans ces figures. Il s'agit d'un instant donné auquel les poids des liens suivent une forte variation. L'étoile sur l'image du milieu indique une synchronisation plus subtile où le poids du lien placé à l'avant-dernière ligne semble augmenter légèrement avant les autres. Ce type de séquençement est difficile à capturer dans une vue non agrégée des liens. L'image de droite est un autre exemple qui indique ces motifs de séquençement et possible synchronisation entre certaines régions du cerveau. Les liens concernent la connectivité de deux régions FPTCLdLPFS et FPTCLfrontal au groupe DAN. Au temps 1, la force de connectivité des deux régions au groupe DAN s'inverse soudainement : FPTCLfrontal est connecté moins fortement alors que FPTCLdLPFS devient plus fortement connecté. Un autre de ces changements est plus progressif entre les temps 2 et 3.

des neuroscientifiques de découvrir des motifs temporels dans leurs données et surtout de former des hypothèses et d'effectuer des analyses qu'ils n'avaient pas envisagées auparavant. Il est intéressant de noter que les vues synchronisées, sollicitées par nos collaborateurs ne semblent pas avoir été utilisées lors des découvertes. D'après leurs commentaires, elles leurs ont cependant permis de se familiariser avec la représentation visuelle de LinkWave pendant les premiers jours d'utilisations du système et de les aider à interpréter les nouveaux motifs lors de l'exploration. Bien que de nombreuses recherches et analyses systématiques soient encore nécessaires pour conclure sur la validité et l'importance des motifs découverts avec notre outil, ce premier retour d'expérience a permis de démontrer le potentiel de ce nouveau paradigme d'exploration visuelle des réseaux pondérés dynamiques.

TRAVAUX ANTÉRIEURS

Nous décrivons les travaux antérieurs dans le domaine de la visualisation d'information et non spécifiques au domaine des neurosciences.

Représentations de réseaux dynamiques

Les réseaux ont deux représentations visuelles communes : les diagrammes nœuds-liens [5] et les matrices d'adjacence [21]. L'animation est une technique standard pour adapter ces représentations aux graphes dynamiques [6]. Mais bien que l'utilisation de la dimension temporelle paraîsse naturelle pour représenter des données dynamiques, l'efficacité de ces animations a été remise en cause [4, 16, 17]. En particulier, Tversky *et al.* ont démontré que l'identification des changements entre plusieurs images successives est laborieuse étant donné que le cerveau humain ne peut suivre qu'un petit nombre (cinq) d'objets en mouvement à la fois [31]. Les "small

multiples" [30], qui consistent à offrir une série de représentations pour chaque point dans le temps, sont une alternative à l'animation. Cette technique permet d'accomplir certaines tâches plus efficacement, comme la comparaison de topologies à des instants éloignés [4]. Mais elle ne passe pas à l'échelle lorsque le nombre d'instants considérés augmente. Quelques travaux ont proposé de représenter le temps comme une troisième dimension [2, 8], mais les techniques proposées souffrent des habituels problèmes d'occultation inhérents aux représentations 3D.

Plusieurs chercheurs se sont intéressés au cas particulier des réseaux pondérés dynamiques. Burch *et al.* [12] proposent une technique basée sur les coordonnées parallèles [22]. Leur technique représente les nœuds du réseaux par des lignes verticales répliquées pour chaque point considéré dans le temps. Les liens pondérés connectent deux lignes verticales, indiquant le poids par l'intensité du lien. Malheureusement, la lisibilité de la représentation dépend fortement de la densité du réseau et la technique proposée ne permet pas aisément de caractériser l'évolution du poids des liens. Les autres techniques proposées sont basées sur les matrices. Perer and Sun proposent une approche basée sur les "small multiples" [25]. Yi *et al.* [34] et Brandes & Nick [9] définissent des glyphes pour représenter l'évolution des poids des liens. Les glyphes sont une représentation visuelle (telle qu'un histogramme) de petite taille qui peut donc être insérée dans la cellule correspondante de la matrice. L'utilisation de glyphe permet d'ajouter une dimension de données aux matrices, mais rend difficile l'extraction de leurs motifs temporels car les glyphes ne sont pas alignés.

Le système le plus proche de nos travaux est Cubix [24]. Ce système est basé sur une matrice d'adjacence dont la

troisième dimension est utilisée pour le temps. Ce cube espace-temps permet aux utilisateurs de réaliser des opérations très avancées. Il est par exemple possible de créer une matrice d'adjacence standard mais de remplacer une colonne par son évolution dans le temps. Par contraste, LinkWave opte pour une des structures les plus simples : une liste réordonnable. Un avantage majeur de cette liste est la possibilité d'obtenir une vue d'ensemble d'un (très) grand nombre de liens en utilisant des interactions familières aux utilisateurs tels que pan et zoom, filtrage et tri. Notre système a été exclusivement conçu pour accomplir les tâches basées sur les liens du réseaux. Intégrer des fonctionnalités de tri plus avancées basées sur les différentes configurations de cubix, tout en préservant la simplicité de LinkWave est une direction intéressante.

Représentations de données temporelles

La représentation visuelle de données temporelles a donné lieu à une quantité considérable de travaux [3, 1]. Les techniques les plus communes pour représenter ce type de données sont les graphiques en aires ou d'horizons [20], ou les animations. Les représentations utilisées dans notre travail sont inspirées des graphiques empilés [32] et *streamgraphs* tels que *ThemeRiver* [19] ou *TouchWave* [7]. Plusieurs travaux antérieurs se sont concentrés sur l'identification de séquences et de motifs dans les données temporelles tels que *PatternFinder* [15] et *LifeLines* [26]. Ces systèmes permettent de créer et raffiner interactivement des requêtes afin de trouver des séquences particulières dans un grand nombre de données temporelles. Aucun de ces systèmes n'a été appliqué aux réseaux pondérés dynamiques.

Concernant l'exploration de structures plus complexes, plusieurs techniques se sont focalisées sur la représentation de données temporelles associées à des hiérarchies. *Timeline trees* [10], son équivalent circulaire *TimeRadar-Trees* [11] et *TimeEdgeTrees* [13] sont de bons exemples. Dans ces représentations, les relations entre les données temporelles sont visualisées par des diagrammes nœuds-liens. Les analystes peuvent ouvrir et fermer les branches de l'arbre interactivement, ce qui permet d'agréger ou de désagréger les visualisations temporelles. Ces techniques sont efficaces pour les structures hiérarchiques dans lesquelles les liens ne se croisent pas, mais ne sont pas appropriées aux réseaux plus complexes. Quelques travaux se sont enfin intéressés à représenter des données temporelles associées aux nœuds des réseaux [27], mais ces techniques ne permettent pas de représenter la temporalité des liens, de fait beaucoup plus nombreux.

CONCLUSION

Cet article présente les premiers résultats de notre collaboration avec des neuroscientifiques qui cherchent à identifier des motifs dans les réseaux pondérés dynamiques extraits de données d'imagerie par résonance magnétique fonctionnelle (IRMf). À l'issue d'entretiens et de séances d'observation, nous avons dressé une liste de tâches à accomplir pour permettre l'exploration des réseaux considérés. Ces tâches, centrées sur l'évolution des poids des liens, se retrouvent également dans d'autres domaines d'application. Comme les techniques de visualisation de réseaux actuelles ne sont pas centrées sur l'évolution

des poids mais plutôt l'évolution des nœuds, nous avons conçu un nouveau système, LinkWave, permettant d'accomplir ce type de tâches. LinkWave met en œuvre un nouveau paradigme de représentation des réseaux pondérés dynamiques dans lequel les liens et l'évolution de leur poids sont mis en avant. Les utilisateurs interagissent directement avec les liens et peuvent les agréger pour explorer la connectivité du réseau au cours du temps, et ce à plusieurs niveaux d'abstraction. Nous avons développé LinkWave en collaboration avec des neuroscientifiques et présenté dans cet article leurs retours d'expérience ainsi que quelques découvertes préliminaires.

Dans le cadre de notre collaboration avec ces neuroscientifiques, nous développons une deuxième version de LinkWave incluant de nouveaux mécanismes pour former des requêtes sur la topologie du réseau et sur les séquences temporelles. Nous explorons également des moyens de mieux intégrer les outils statistiques à notre système de visualisation. En effet, même si les motifs découverts par exploration visuelle permettent aux scientifiques de former de nouvelles hypothèses, ceux-ci ont toujours recours à leurs outils statistiques traditionnels pour les rejeter ou les valider. Une intégration de ces outils pourrait permettre un processus d'exploration plus efficace. Enfin, la direction de recherche la plus prometteuse est la conception d'interactions destinées à faciliter la comparaison de plusieurs liens ; groupes de liens ou réseaux. Cette tâche est particulièrement importante pour nos collaborateurs car ils explorent les données issues de plusieurs sujets et cherchent à identifier des motifs représentatifs.

REMERCIEMENTS

Nous remercions Sohaib Ghani qui a contribué au développement du système durant son stage à Microsoft Research. Nous remercions également Bongshin Lee qui a participé aux discussions sur la conception du système. This research was supported by grants from the National Institutes of Health 1RC4NS073008-01.

BIBLIOGRAPHIE

1. Adjanor, K., Lecolinet, E., Guiard, Y., and Ribière, M. Visualisation interactive de données temporelles: Un aperçu de l'état de l'art. In *Actes d'IHM '10*, ACM (2010), 97–104.
2. Ahmed, A., Dwyer, T., Murray, C., Song, L., and Wu, Y. X. WilmaScope graph visualisation. In *Proceedings of InfoVis '04*, IEEE (Oct. 2004), r4.
3. Aigner, W., Miksch, S., Müller, W., Schumann, H., and Tominski, C. Visualizing time-oriented data – a systematic view. *Computers and Graphics* 31, 3 (June 2007), 401–409.
4. Archambault, D., Purchase, H., and Pinaud, B. Animation, small multiples, and the effect of mental map preservation in dynamic graphs. *IEEE TVCG* 17, 4 (April 2011), 539–552.
5. Auber, D., Chiricota, Y., Delest, M., Melançon, G., Domenger, J.-P., and Mary, P. Visualisation de graphes avec Tulip : exploration interactive de grandes masses de données en appui à la fouille de données et à l'extraction de connaissances. In *Actes d'EGC '07*, Cépaduès (Jan. 2007), 147–156.

6. Bach, B., Pietriga, E., and Fekete, J.-D. GraphDiaries: animated transitions and temporal navigation for dynamic networks. *IEEE TVCG 20*, 5 (May 2014), 740–754.
7. Baur, D., Lee, B., and Carpendale, S. TouchWave: kinetic multi-touch manipulation for hierarchical stacked graphs. In *Proceedings of ITS '12*, ACM (2012), 255–264.
8. Brandes, U., and Corman, S. R. Visual unrolling of network evolution and the analysis of dynamic discourse. *Information Visualization 2*, 1 (Mar. 2003), 40–50.
9. Brandes, U., and Nick, B. Asymmetric relations in longitudinal social networks. *IEEE TVCG 17*, 12 (Dec. 2011), 2283–2290.
10. Burch, M., Beck, F., and Diehl, S. Timeline trees: Visualizing sequences of transactions in information hierarchies. In *Proceedings of AVI '08*, ACM (2008), 75–82.
11. Burch, M., and Diehl, S. TimeRadarTrees: visualizing dynamic compound digraphs. In *Proceedings of EuroVis '08*, Eurographics Association (2008), 823–830.
12. Burch, M., Vehlow, C., Beck, F., Diehl, S., and Weiskopf, D. Parallel edge splatting for scalable dynamic graph visualization. *IEEE TVCG 17*, 12 (Dec 2011), 2344–2353.
13. Burch, M., and Weiskopf, D. Visualizing dynamic quantitative data in hierarchies - TimeEdgeTrees: attaching dynamic weights to tree edges. In *IMAGAPP/IVAPP*, SciTePress (2011), 177–186.
14. Ellis, W. D. *A source book of Gestalt psychology*. Routledge & Kegan Paul, 1938.
15. Fails, J., Karlson, A., Shahamat, L., and Shneiderman, B. A visual interface for multivariate temporal data: finding patterns of events across multiple histories. In *Proceedings of VAST '06*, IEEE (2006), 167–174.
16. Farrugia, M., and Quigley, A. Effective temporal graph layout: A comparative study of animation versus static display methods. *Information Visualization 10*, 1 (Jan. 2011), 47–64.
17. Ghani, S., Elmqvist, N., and Yi, J. S. Perception of animated node-link diagrams for dynamic graphs. In *Proceedings of EuroVis '12*, John Wiley & Sons (June 2012), 1205–1214.
18. Ghoniem, M., Fekete, J.-D., and Castagliola, P. Comparaison de la lisibilité des graphes en représentation noeuds-liens et matricielle. In *Actes d'IHM 04*, ACM (2004), 77–84.
19. Havre, S., Hetzler, B., and Nowell, L. ThemeRiver: visualizing theme changes over time. In *Proceedings of InfoVis '00*, IEEE (Oct. 2000), 115–123.
20. Heer, J., Kong, N., and Agrawala, M. Sizing the horizon: The effects of chart size and layering on the graphical perception of time series visualizations. In *Proceedings of CHI '09*, ACM (2009), 1303–1312.
21. Henry, N., and Fekete, J.-D. MatrixExplorer: un système pour l'analyse exploratoire de réseaux sociaux. In *Actes d'IHM '06*, ACM (2006), 67–74.
22. Inselberg, A. The plane with parallel coordinates. *The Visual Computer 1*, 2 (1985), 69–91.
23. Lee, B., Plaisant, C., Parr, C. S., Fekete, J.-D., and Henry, N. Task taxonomy for graph visualization. In *Proceedings of BELIV '06*, ACM (2006), 1–5.
24. Melo, C., Mikheev, A., Le-Grand, B., and Aufaure, M.-A. Cubix: A visual analytics tool for conceptual and semantic data. In *Proceedings of ICDMW '12*, IEEE (Dec. 2012), 894–897.
25. Perer, A., and Sun, J. MatrixFlow: temporal network visual analytics to track symptom evolution during disease progression. In *Proceedings of AMIA '12*, American Medical Informatics Association (Nov. 2012), 716–725.
26. Plaisant, C., Milash, B., Rose, A., Widoff, S., and Shneiderman, B. Lifelines: Visualizing personal histories. In *Proceedings of CHI '96*, ACM (1996), 221–227.
27. Saraiya, P., Lee, P., and North, C. Visualization of graphs with associated timeseries data. In *Proceedings of InfoVis '05*, IEEE (Oct. 2005), 225–232.
28. Shneiderman, B., and Plaisant, C. Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies. In *Proceedings of the 2006 AVI Workshop on Beyond Time and Errors: Novel Evaluation Methods for Information Visualization*, BELIV '06, ACM (New York, NY, USA, 2006), 1–7.
29. Sporns, O. *Networks of the Brain*. MIT Press, Oct. 2010.
30. Tufte, E. *Envisioning Information*. Graphics Press, May 1990.
31. Tversky, B., Morrison, J. B., and Betrancourt, M. Animation: can it facilitate? *IJHCS 57*, 4 (Oct. 2002), 247–262.
32. Wattenberg, M. Baby names, visualization, and social data analysis. In *Proceedings of InfoVis '05*, IEEE (Oct. 2005), 1–7.
33. Xia, M., Wang, J., and He, Y. BrainNet Viewer: a network visualization tool for human brain connectomics. *PLoS ONE 8*, 7 (07 2013), e68910.
34. Yi, J. S., Elmqvist, N., and Lee, S. TimeMatrix: Analyzing Temporal Social Networks Using Interactive Matrix-Based Visualizations. *IJHCI 26*, 11 (Nov. 2010), 1031–1051.