

REPRÉSENTATIONS VISUELLES ALTERNATIVES POUR LES RÉSEAUX SOCIAUX

Nathalie HENRY

Jean-Daniel FEKETE

Grâce à la démocratisation des technologies internet et l'avènement du web 2.0, il devient plus facile de collecter de grandes quantités de données et d'extraire de grands réseaux sociaux afin de les étudier. Par exemple, au travers de sites tels que Facebook ou Friendster, les réseaux sociaux disponibles sont très vastes (des millions d'utilisateurs), riches (contenant de multiples relations ou informations sur les acteurs du réseau) et évoluent dans le temps. Face à ces phénomènes sociaux récents et dont l'effet boule-de-neige génère des réseaux de taille démesurée – tels que Facebook, qui, lancé en 2004, annonce 12 millions d'utilisateurs actifs en 2006, 50 millions fin 2007 et plus de 120 millions en 2008, avec un taux de croissance hebdomadaire de 3 % et une population doublant tous les 6 mois¹ – les sociologues n'ont pas seulement besoin de puissants outils d'analyse mais de véritables outils d'*exploration*.

Effectivement, si la collecte des données est grandement simplifiée, le problème de l'analyse se pose de manière aigüe car ces nouveaux réseaux ont une structure et une évolution très méconnues. Le processus d'exploration est donc essentiel car il permet de susciter de nouvelles questions sur les données, de mettre en évidence d'éventuelles informations surprenantes et ainsi établir des hypothèses qui n'auraient pas été imaginées *a priori*. Or, si les outils actuels² proposent un large panel de fonctionnalités statistiques et analytiques, ils n'ont pas été conçus pour l'exploration de données. UCINET³, par exemple, permet de répondre à un grand nombre de questions, e.g. calculer de multiples mesures de centralité des acteurs ; ou de valider plusieurs hypothèses sur le réseau, e.g. vérifier si le jeu de données correspond à un modèle statistique donné ; mais ce système n'offre que très peu de support (limité à un lot de statistiques descriptives) pour générer de nouvelles questions et aider les sociologues à découvrir de nouvelles pistes d'analyse.

1. <http://www.facebook.com/press/info.phstatistics> et <http://adscriptum.blogspot.com/2007/10/facebook-statistiques-fin-octobre-2007.html>

2. HUISMAN et DUIJN, section du livre sur l'analyse des réseaux sociaux avec Pajek.

3. Voir <http://www.analytictech.com/>

Aider les analystes à formuler de nouvelles questions sur leurs données est le but de l'Analyse Exploratoire des Données⁴, introduite en 1970 par le statisticien John Tukey. Dans son livre, il démontre que la représentation graphique des données brutes est un puissant outil d'analyse et que représenter les données sous de multiples perspectives permet de susciter des questions et d'observer des phénomènes saillants dont la découverte aurait requis beaucoup d'efforts ou d'intuition en utilisant les statistiques classiques. Le Quartet du statisticien Anscombe⁵ constitue un bon exemple. Pour démontrer la puissance de l'exploration visuelle, Anscombe a généré quatre séries de points (coordonnées x,y) ayant les mêmes statistiques générales (moyenne, écart type, etc.), indissociables par régression linéaire, mais dont la représentation graphique révèle en un coup d'œil la différence. L'exploration visuelle est donc une étape cruciale pour l'analyse des réseaux sociaux.

Aujourd'hui, les sociologues disposent d'un panel d'outils de représentation graphique (tel que Pajek) mais là encore, ceux-ci ne constituent pas des outils d'exploration performants. Effectivement, la majorité de ces systèmes sont basés sur des représentations visuelles en nœud-lien (figure 1). Ces représentations souffrent d'importants problèmes de lisibilité pour les réseaux grands (beaucoup de nœuds qui se superposent) ou denses (beaucoup de liens qui se croisent), rendant l'exploration des réseaux actuels impossible car la représentation générée n'est bien souvent qu'un amas indistinct de nœuds et de liens.

Depuis les cinq dernières années, le domaine de la visualisation d'information⁶ a connu une effervescence autour de l'exploration des réseaux et des graphes⁷, très certainement liée à la popularité des réseaux sociaux et aux défis lancés par leur très grande taille. Ces nouveaux systèmes, basés sur l'exploration interactive visuelle, constituent une solution pour aborder l'analyse des vastes réseaux sociaux actuels. Dans cet article, nous faisons un état de l'art de ces nouvelles visualisations des réseaux, constituant des alternatives aux diagrammes nœud-lien et favorisant la manipulation interactive des données.

4. TUKEY, 1977.

5. ANSCOMBE, 1973.

6. Les deux auteurs de cet article sont des chercheurs en visualisation d'information.

7. Dans cet article, nous utiliserons indifféremment les termes graphe et réseau.

REPRESENTATIONS NŒUD-LIEN

L'analyse des réseaux sociaux a débuté il y a plus de 70 ans, avec les travaux empiriques de Jacob Moreno⁸. Il a été un des pionniers à utiliser la représentation nœud-lien pour communiquer sur ses travaux (figure 1). Dans cette représentation, un nœud représente un acteur et un lien représente une relation du réseau. Il s'agit de la représentation la plus courante des réseaux⁹. Wasserman et Faust¹⁰ présentent les diverses catégories de méthodes : analyses statistiques, structurelles et exploratoires. Freeman¹¹ retrace l'histoire des visualisations de réseaux sociaux et montre que les représentations visuelles peuvent être un outil efficace pour illustrer des concepts tels que les acteurs centraux ou les groupes sociaux. Les figure 1 et figure 2 présentent deux exemples utilisant les diagrammes nœud-lien.

Le très grand avantage des diagrammes nœud-lien est leur intuitivité : la grande majorité des lecteurs peut les comprendre. En revanche, qu'ils soient dessinés manuellement (figure 1) ou générés automatiquement (figure 2), leur lisibilité dépend totalement du placement des nœuds dans le plan. Ce problème épineux a d'ailleurs donné naissance à un domaine de recherche à part entière nommé le dessin de graphes (*graph drawing*).

Dessin de graphes et visualisation d'information

L'ouvrage de Di Battista *et al.*¹² constitue une bonne introduction au domaine du dessin de graphes, présentant plus de trois cents algorithmes. Un état de l'art de la visualisation de graphes¹³, plus récent, recense également un grand nombre de techniques, parfois interactives, pour représenter les graphes et les réseaux. En général, ces techniques tentent d'optimiser un certain nombre de critères esthétiques, comme, par exemple, limiter le nombre de croisements des liens ou bien placer les nœuds pour avoir une longueur uniforme de tous les liens. Plusieurs études ont été réalisées pour

8. MORENO, 1934.

9. En parcourant le répertoire de l'International Network for Social Network Analysis (<http://www.insna.org>), il est possible de compter plus de 50 systèmes basés sur les représentations nœud-lien.

10. WASSERMAN et FAUST, 1994.

11. FREEMAN, 2000.

12. Di BATTISTA, EADES, TAMASSIA et TOLLIS, 1998.

13. HERMAN, MELANÇON et MARSHALL, 2000.

déterminer quels critères sont les plus importants et ainsi tenter de découvrir le meilleur algorithme de dessin de graphe. Cependant, il existe un tel nombre de critères qu'il est difficile de former des conclusions générales.

Figure 1. Réseau d'amitiés entre garçons (triangles) et filles (cercles) par J. Moreno. Un acteur central lie les deux groupes (triangle du milieu gauche)

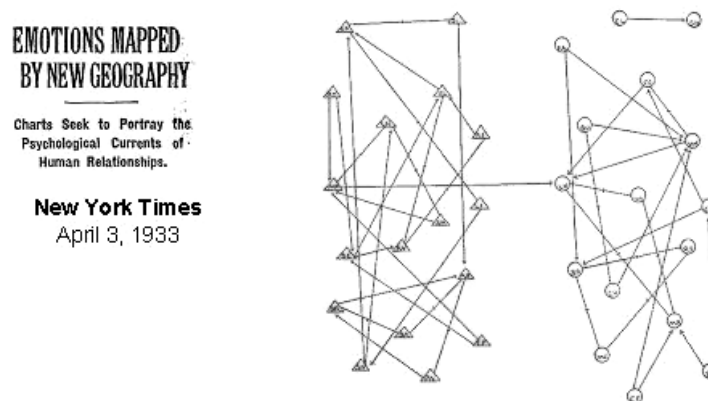
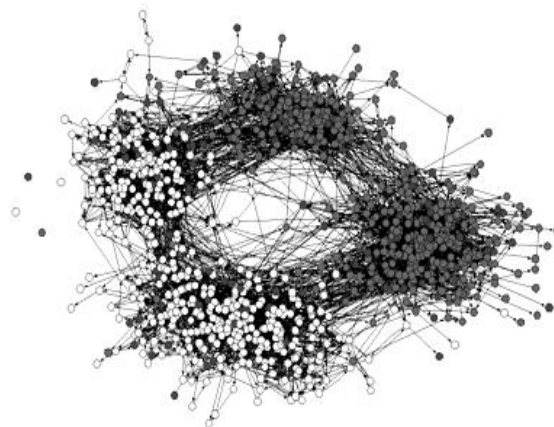


Figure 2. Réseau d'amitiés entre lycéens par J. Moody, la couleur marque l'origine ethnique. Quatre groupes sociaux émergent après application d'un algorithme de clustering. Ces quatre groupes montrent que les relations d'amitiés sont fortement corrélées avec l'âge et l'origine ethnique des lycéens



Le domaine de la visualisation d'information a une perspective différente sur le sujet. Ce domaine de recherche se focalise sur l'exploration visuelle et la

découverte ou la communication d'informations. Ainsi, les réseaux présentés en figure 1 et figure 2 n'optimisent certainement pas des critères esthétiques, comme la minimisation du nombre de croisements de liens, mais ils permettent de mettre en lumière certaines informations importantes sur la structure du réseau. Dans cet article, nous présentons une série de représentations alternatives aux diagrammes nœud-lien, non pas pour les remplacer mais plutôt pour les compléter. Dans le contexte de l'analyse exploratoire, nous pensons qu'il est nécessaire d'avoir plusieurs perspectives sur les données pour en découvrir toutes les arcanes.

Exploration de réseau et passage à l'échelle

Un problème récurrent se pose quand le réseau à représenter contient beaucoup de nœuds ou de liens : le diagramme nœud-lien se transforme en un amas de traits et de points difficile voire impossible à transformer en représentation lisible, ni manuellement ni automatiquement. Dans la littérature, il existe trois techniques distinctes pour tenter de résoudre ce problème :

1. *réduire la quantité d'information représentée par filtrage ou agrégation*. Il existe plusieurs méthodes pour échantillonner les réseaux¹⁴ ou bien calculer des clusters¹⁵ (groupe d'éléments similaires), qui peuvent être ensuite agrégés en un élément représentatif du groupe ;

2. *proposer une représentation interactive pour explorer la totalité du réseau*. Plusieurs outils de visualisation ne représentent qu'une partie du réseau et proposent de naviguer interactivement pour explorer le reste du réseau. Les stratégies varient : certaines se concentrent sur un nœud particulier et permettent la navigation en suivant les connexions¹⁶, d'autres sont basées principalement sur du filtrage interactif¹⁷ ou bien permettent la navigation dans un réseau agrégé¹⁸ ;

3. *utiliser des représentations alternatives aux diagrammes nœud-lien*. Cette stratégie consiste à utiliser différentes métaphores visuelles pour la représentation de réseaux. L'idée est de découvrir des représentations permettant d'« augmenter l'espace visuel », permettant d'afficher un plus grand nombre d'information de façon plus lisible.

14. FRANK, 1978.

15. JAIN, MURTY et FLYNN, 1999.

16. HEER et BOYD, 2005 et LEE *et al.*, 2006.

17. PERER et SHNEIDERMAN, 2006 et SHNEIDERMAN et ARIS, 2006.

18. ABELLO *et al.*, 2006.

La psalmodie de la visualisation d'information¹⁹ est : « vue d'ensemble d'abord, zoomer et filtrer, puis détails à la demande » (*overview first, zoom and filter, then details on demand*). Cette phrase résume le processus d'exploration visuel.

Ainsi, il est primordial de présenter une vue d'ensemble du réseau afin d'initier le processus d'exploration. Les deux premières stratégies, visant à supprimer l'information ou à n'en présenter qu'une partie, ne sont donc pas satisfaisantes. Nous nous sommes donc intéressés à la troisième solution : utiliser des représentations alternatives pour explorer les réseaux.

REPRESENTATIONS ALTERNATIVES

3D et espaces géométriques non euclidiens

Afin d'augmenter l'espace visuel, un certain nombre de chercheurs ont tenté d'utiliser la 3^e dimension. L'idée est qu'une troisième dimension permet d'avoir plus d'espace pour dessiner de plus grands réseaux mais aussi permet d'optimiser certains critères esthétiques tels que la minimisation du nombre de croisements. Plusieurs programmes utilisent la 3D, la figure 3a montre un exemple créé avec la boîte à outils de dessin de graphes Tulip²⁰. Le principal problème lié à l'utilisation de la 3D est l'occlusion : les utilisateurs ont l'impression que les représentations sont « encombrées » et trouvent difficile d'avoir une représentation mentale de la totalité du graphe²¹. Pour palier à ces problèmes, plusieurs stratégies ont été adoptées²², comme l'utilisation de multiples vues ou des techniques de navigation pour voir la représentation sous de multiples angles. Malheureusement, dans la majorité des cas, ces techniques désorientent l'utilisateur et rendent les visualisations peu efficaces²³. Plusieurs études ont montré que, si les visualisations 3D sont attractives, elles n'améliorent pas les performances et les détériorent même pour certaines tâches²⁴.

19. SHNEIDERMAN, 1996.

20. AUBER, 2003.

21. COCKBURN et MCKENZIE, 2002.

22. ELMQVIST et TSIGAS, 2007.

23. SUTCLIFFE et PATER, 1996.

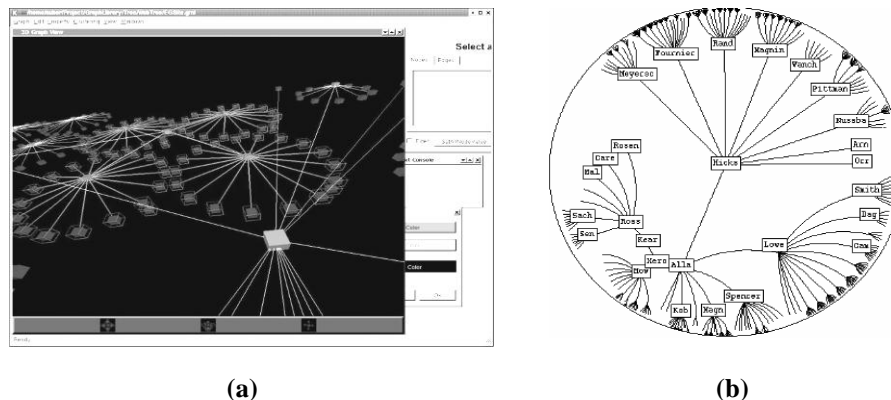
24. COCKBURN et MCKENZIE, 2000.

Une autre façon d'augmenter l'espace visuel est l'utilisation de la géométrie hyperbolique au lieu de la géométrie Euclidienne. Dans l'espace hyperbolique, le postulat du parallélisme est rejeté : deux lignes parallèles dans l'espace euclidien divergent l'une de l'autre. Donc, si l'on considère un disque dans l'espace hyperbolique, l'espace augmente exponentiellement lorsque l'on s'éloigne de son centre. Ainsi, un réseau dessiné sur ce disque dispose d'un espace infini sur ses bords, ce qui permet d'afficher plus d'information, comme le montre figure 3b. La métaphore s'applique aussi à la 3D²⁵. Malheureusement, similairement à la 3D, naviguer dans un espace géométrique non euclidien désoriente l'utilisateur. Il est de plus difficile de se constituer une carte mentale du réseau sans recours à la géométrie 2D euclidienne²⁶.

Représenter un graphe par un arbre

Une autre stratégie pour représenter des réseaux peu denses et augmenter leur lisibilité est de les représenter sous forme d'arbres et de le « réparer » avec des connexions supplémentaires. Les algorithmes de dessin d'arbres permettent effectivement d'ôter tous croisements voire d'exploiter au maximum l'espace utilisé si l'arbre est dessiné en utilisant des inclusions de formes.

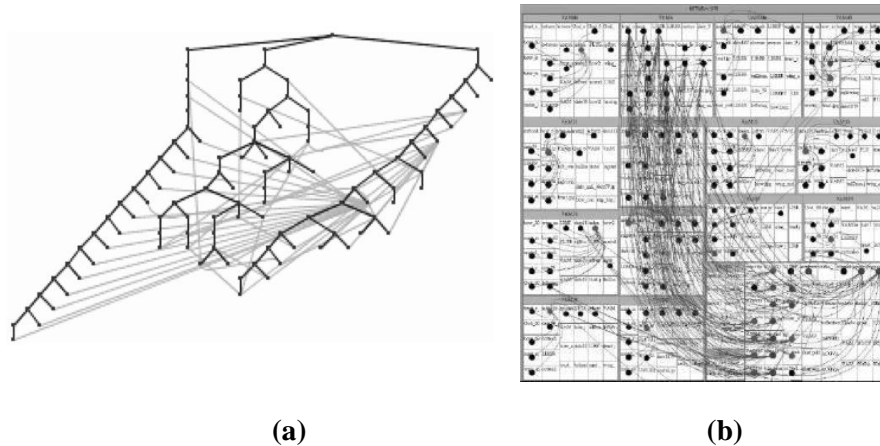
Figure 3. Représentation d'un réseau en 3D (a) et dans l'espace hyperbolique (b).



25. MUNZNER, 1997.

26. RISDEN *et al.*, 2000.

Figure 4. Représentation d'un réseau sous forme d'arbre avec des liens supplémentaires (a) et sous forme de Treemap avec des liens supplémentaires (b)



Pour dessiner un réseau sous forme d'arbre, il est nécessaire de calculer un arbre recouvrant (généralement on prend le plus grand), de dessiner cet arbre puis d'ajouter les liens supplémentaires²⁷. Un exemple est présenté dans la figure 4a. Cette technique fonctionne bien pour des réseaux ayant une structure différant peu de celle d'un arbre tels que les arbres généalogiques par exemple, mais la représentation devient illisible pour des réseaux plus denses.

Une alternative est la représentation sous forme de Treemap²⁸. Le principe est identique au précédent, mais la représentation utilisée pour l'arbre recouvrant diffère : il s'agit de découper l'espace successivement pour représenter la hiérarchie (imbrications de boîtes). Par exemple, un arbre très simple composé d'une racine, de deux nœuds et de quatre feuilles sur chacun d'eux, est alors représenté par un rectangle occupant tout l'espace visuel, contenant deux rectangles de taille identique (représentant les nœuds), eux-mêmes sub-divisés en quatre rectangles de taille identique. Les liens supplémentaires du réseau sont alors dessinés par dessus le Treemap²⁹

27. HERMAN *et al.*, 1999.

28. SHNEIDERMAN, 1992.

29. FEKETE *et al.*, 2003.

comme montré sur la figure 4b. Cette représentation devient également difficile à lire si le nombre de liens supplémentaires est important.

L'ajout d'interaction est une piste intéressante pour améliorer la lisibilité de ces représentations. Elle est particulièrement bien exploitée dans TreePlus³⁰, un système visualisant une portion du réseau sous forme d'arbre et permettant de naviguer interactivement dans le reste. Afin de résoudre le problème des cycles, dans cette représentation sous forme d'arbre, des nœuds sont dupliqués. Si ce système permet de mieux réaliser un certain nombre de tâches, il rend difficile la constitution d'une carte mentale de la totalité du réseau.

Représentation matricielle

Les graphes ont deux représentations canoniques : les diagrammes nœud-lien et les matrices d'adjacence. Une matrice d'adjacence représente chaque sommet d'un réseau à la fois comme une ligne et comme une colonne. Si deux sommets sont connectés, la case correspondant à l'intersection de la ligne et de la colonne est marquée. Traditionnellement, on utilise une valeur numérique (0 marquant l'absence de connexion, 1 marquant la présence), soit par une marque graphique comme dans la figure 5. Le tableau 1 liste les principaux avantages et inconvénients des représentations matricielles par rapport aux diagrammes nœud-lien.

Tableau 1. Avantages et inconvénients des représentations matricielles

<i>Pour</i>	<i>Contre</i>
Absence de superpositions des nœuds (ce qui permet de pouvoir toujours lire les étiquettes portées par les nœuds)	Taille de l'espace visuel requis à un niveau de détail équivalent plus important que le diagramme nœud-lien
Absence de croisements des liens (ce qui permet de toujours identifier la source et destination des connexions)	Difficulté à effectuer des tâches de suivi de chemin (suivre les liens de Pierre à Paul passant par Marie)
Facilité avec laquelle il est possible d'identifier les absences de connexions	Manque de familiarité, les matrices paraissent moins intuitives que les diagrammes nœud-lien

30. LEE *et al.*, 2006.

Plus de détails sur les performances des deux représentations de réseaux sont présentés dans les travaux de Ghoniem *et al*³¹. Il s'agit d'une comparaison plus quantitative de la lisibilité des représentations matricielles et des représentations nœud-lien pour des graphes de différentes tailles et densités. Dans cette étude, une liste de tâches de base a été utilisée pour évaluer la lisibilité des représentations : estimer le nombre de nœuds, estimer le nombre de connexions, trouver le nœud le plus connecté, trouver le nœud portant une étiquette donnée, trouver la connexion entre deux nœuds donnés, trouver un nœud commun à deux nœuds donnés et trouver un chemin entre deux nœuds donnés.

De manière générale, les résultats de leur expérimentation montrent que les nœud-lien sont plus efficaces pour les graphes de petite taille et peu denses alors que les matrices ont de meilleures performances dès que les graphes sont plus grands (>20 nœuds !) et plus denses pour toutes les tâches excepté le suivi de chemin. Ces résultats ainsi que les pros et cons listés précédemment révèlent le vaste potentiel de la représentation matricielle pour explorer les réseaux sociaux.

UTILISER LA REPRESENTATION MATRICIELLE

Deux facteurs principaux entrent en jeu lors qu'il s'agit d'utiliser les représentations matricielles pour explorer de grands réseaux :

1. être capable de réordonner leurs lignes et colonnes ;
2. pouvoir naviguer dans des matrices de grande taille.

Réordonner les matrices

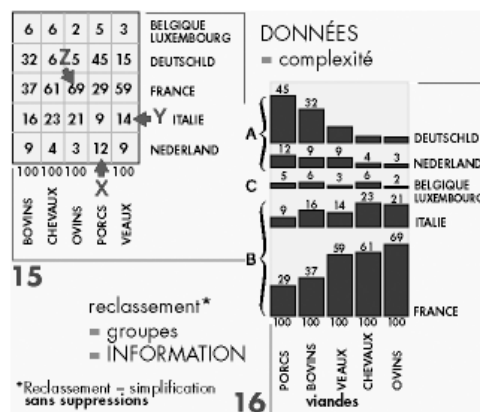
Dans son livre faisant référence en visualisation d'information³², Bertin présente la matrice réordonnable. Au travers de plusieurs exemples, il démontre que réordonner les lignes et colonnes d'une matrice permet d'améliorer significativement sa compréhension.

31. GHONIEM, FEKETE et CASTAGLIOLA, 2005.

32. La Sémiologie Graphique, BERTIN, 1967.

Prenons pour exemple la **figure 5** extraite de ce livre ; la matrice présentée contient seulement 5 lignes et 5 colonnes représentant la production de 5 types de viande dans 5 pays. En transformant la matrice originale, contenant des valeurs numériques, en matrice visuelle, codant les valeurs par des formes géométriques de différentes tailles et en réordonnant « intelligemment » ses lignes et ses colonnes, il devient beaucoup plus simple d'interpréter ce jeu de données. En effet, il est pratiquement immédiat d'observer que la France est le pays ayant la plus grande production, tout type de viande confondu ; que la Belgique a la plus petite et qu'il existe trois différents profils de production (marqué par A, B et C dans la figure). Pour aller une étape plus loin dans l'interprétation de ce simple exemple, imaginons qu'une loi doive être votée pour limiter la production d'un type de viande produit en grande quantité par les pays.

Figure 5. Matrice réordonnable de J. Bertin, extraite de la Sémiologie graphique. Cette image montre la production de 5 pays pour 5 types de viande. Bertin démontre que la représentation visuelle et la permutation de lignes et colonnes permet d'identifier rapidement des tendances dans les données (ici, les groupes A, B et C)

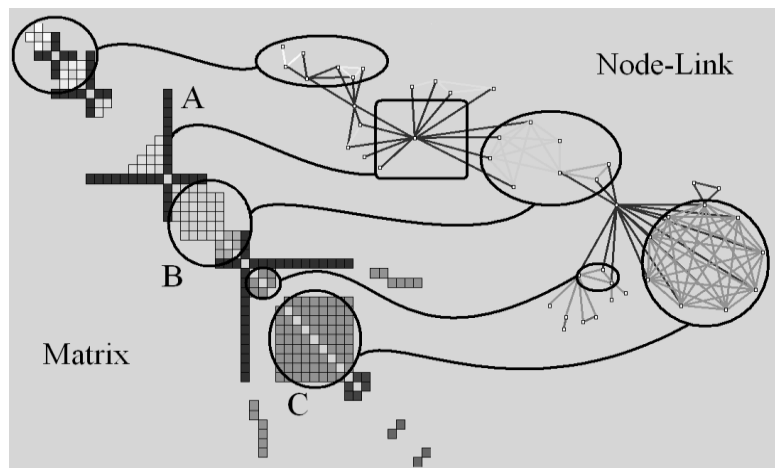


Disposant de ces données et de la représentation visuelle associée, un analyste pourra rapidement observer que la Belgique est le pays à convaincre de ne pas accepter cette loi, car son profil de production est plutôt neutre alors que les autres pays voteront dans leur propre intérêt, contraire à celui de l'autre groupe.

Cet exemple démontre l'importance du réordonnement des matrices : si les représentations matricielles dont les lignes et les colonnes sont ordonnées

aléatoirement sont lisibles, une fois réordonnées elles deviennent réellement exploitables.

Figure 6. Correspondance entre les motifs visuels des matrices et ceux des diagrammes nœud-lien : A montre un acteur connectant plusieurs communautés, B deux communautés et C une clique



Algorithmes de réordonnancement

Effectuer un état de l'art des techniques de réordonnancement automatique est un véritable défi car un grand nombre de problèmes sont liés à la permutation des lignes et colonnes. Par exemple, un algorithme de linéarisation de graphe (alignement de tous les sommets d'un graphe de façon à minimiser un critère esthétique tel que le nombre de croisements) ou bien un algorithme d'optimisation de tables de calculs peut être utilisé pour réordonner une matrice d'adjacence. Les algorithmes sont donc disséminés dans divers domaines tels que la biologie ou le dessin de graphe ; et leurs critères de qualité varient en fonction du contexte d'utilisation.

Plusieurs outils tels que PermutMatrix³³ ou VisuLab³⁴ offrent une représentation visuelle des matrices (ou de tables) et permettent d'expérimenter avec diverses techniques de réordonnancement et leurs paramètres. Nous ne détaillerons pas les diverses catégories de méthodes dans cet article, mais il est important de comprendre que la méthode de

33. CARAUX *et al.*, 2005.

34. Voir <http://www.inf.ethz.ch/personal/hinterbe/Visulab>

réordonnement sélectionnée a un fort impact sur l'interprétation de la représentation, similairement aux méthodes de placement de nœuds des diagrammes nœud-lien. La **figure 6** montre un exemple de motifs visuels émergeant lorsque l'ordonnement de matrice et le placement du nœud-lien sont optimisés.

Naviguer dans de grandes matrices

A niveau de détail équivalent, la représentation matricielle utilise plus d'espace que les diagrammes nœud-lien. Par exemple, sur un écran de 17 pouces, il est possible d'afficher seulement une centaine de nœuds si l'on souhaite pouvoir lire confortablement leurs noms. Ainsi, afin de tirer parti du potentiel des matrices, il est nécessaire d'avoir des outils efficaces de navigation.

De nombreuses techniques existent dans le domaine de l'Interaction Homme-Machine pour permettre de se déplacer dans de grands espaces visuels et/ou à divers niveaux de détails. Par exemple, les techniques de Focus+Context³⁵ telles que la vue à vol d'oiseau (*bird's eye view*) permettent d'avoir une vue d'ensemble miniature et une vue zoomée d'une partie de la représentation. Ces techniques ne sont pas directement utilisables dans notre contexte, elles requièrent quelques aménagements. Il est notamment indispensable de toujours voir les étiquettes des lignes et colonnes des matrices et de faciliter les transitions entre les divers niveaux de détails pour éviter à l'analyste de se perdre lors de l'exploration. Plusieurs systèmes intègrent ces outils de navigation tels que MatrixZoom³⁶ et la matrice d'appels multi-niveaux présentée par van Ham³⁷.

COMBINER MATRICES ET NŒUD-LIEN

Les représentations matricielle et nœud-lien ont chacune des avantages et inconvénients à la fois pour l'exploration et la communication de résultats.

35. CARPENDALE, 1999.

36. ABELLO et VAN HAM, 2004.

37. VAN HAM, 2003.

Dans cette section, nous présentons notre système MatrixExplorer³⁸, basé sur cette représentation duale et tentant de combiner le meilleur de deux mondes.

Initier l'exploration

L'avantage majeur des matrices par rapport aux diagrammes nœud-lien est de permettre d'avoir une vue d'ensemble du réseau toujours exploitable et ce, même pour de très grands graphes, alors que les diagrammes nœud-lien ne montrent bien souvent qu'un amas de liens et de nœuds superposés.

Prenons l'exemple d'un réseau social contenant les données d'échanges de courrier électronique entre plus de 450 personnes pendant un an (les sommets sont les personnes et les liens sont les courriers). La **figure 7** montre deux représentations de ce réseau. Alors que la représentation nœud-lien (utilisant un algorithme de placement traditionnel par forces) permet éventuellement de former l'hypothèse que le réseau est très dense ; la représentation matricielle, elle, donne beaucoup plus d'information. Pour le démontrer, analysons brièvement la matrice présentée à la **figure 7**.

Dans cette représentation, chacun des points noirs représente une connexion entre une ligne et une colonne (un lien dans le réseau), le gris marquant donc l'absence de connexion. Tout d'abord, on peut clairement rejeter l'hypothèse précédente : le réseau social n'est pas très dense car on peut observer une majorité de gris dans cette matrice. Deuxièmement, si l'on regarde de plus près la représentation, on peut voir plusieurs blocs de points noirs. Il s'agit de clusters : des groupes de recherche dont les membres échangent des courriers électroniques les uns avec les autres. Une autre observation permet de voir une sorte de croix : une ligne constituée de points noirs verticale et une horizontale (à peu près de la longueur d'une moitié de la matrice). Il s'agit d'un membre du service des missions de cet institut, celui-ci échange effectivement des courriers électroniques avec beaucoup des membres d'équipes, il a donc un grand nombre de connexions.

Ce simple exemple démontre que les matrices procurent une vue d'ensemble permettant d'initier l'exploration d'un jeu de données en mettant en évidence plusieurs éléments saillants. Le point important est d'apprendre à les décoder, car ces représentations sont beaucoup moins familières que les

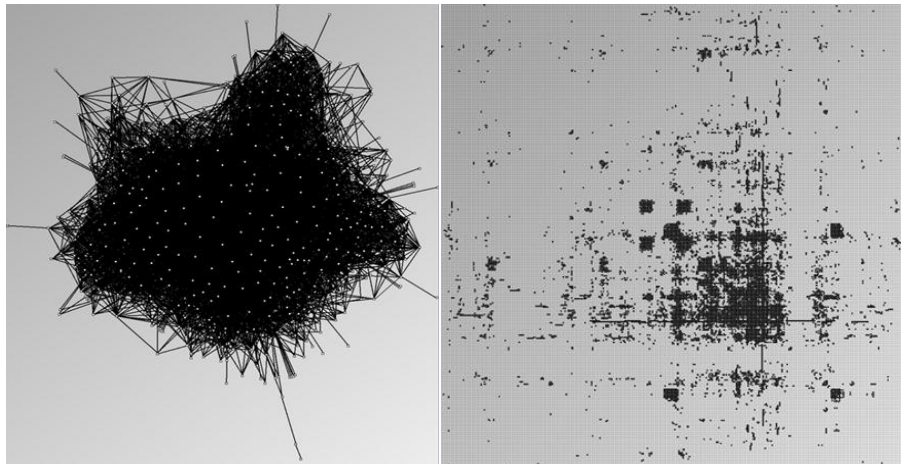
38. HENRY *et al.*, 2005 et 2006.

diagrammes nœud-lien. La **figure 6** donne quelques exemples de correspondance entre les motifs visuels courant dans les réseaux sociaux à la fois pour les représentations nœud-lien et matricielles.

Explorer interactivement

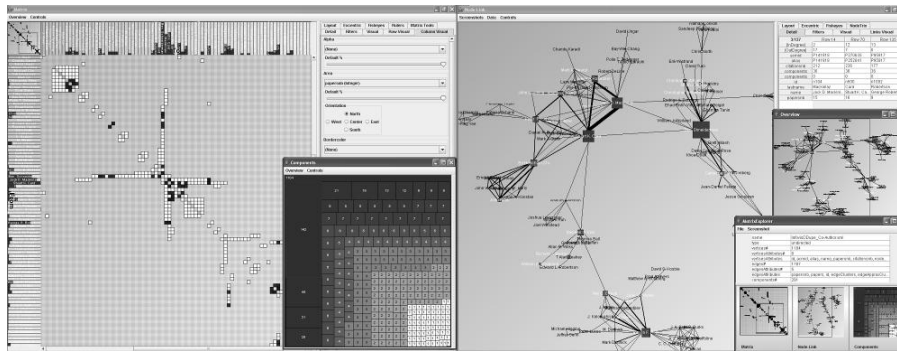
A l'issue d'entretiens avec des analystes et de séances d'observation, nous nous sommes rendu compte que le processus d'exploration visuel est constitué d'opérations successives pour configurer les représentations (ajuster la position des éléments graphiques et les attributs visuels), filtrer certains éléments, les grouper et éventuellement les agréger. Il est important de créer de vues multiples d'un jeu de données, l'analysant sous divers aspects et plusieurs niveaux de détail afin d'identifier sa structure. Par exemple, une représentation matricielle peut permettre d'identifier plusieurs communautés alors qu'un diagramme nœud-lien égo-centré peut permettre l'analyse détaillée des connexions d'un acteur important. Dans MatrixExplorer, nous avons donc décidé de proposer plusieurs vues sur les données (**figure 8**) et d'offrir un grand nombre d'outils pour manipuler directement³⁹ les représentations.

Figure 7. Réseau d'échange de courriers électroniques entre plus de 450 chercheurs, pendant un an. Nœud-lien (a), matrice (b)



39. Manipulation directe, SHNEIDERMAN, 1983 et requêtes dynamiques, AHLBERG *et al.*, 1992.

Figure 8. Interface de MatrixExplorer, combinant matrices (grande fenêtre de gauche) et nœud lien (grande fenêtre de droite). Une représentation Treemap (petite fenêtre de gauche) montre la macrostructure du réseau, *i.e.* ses composantes connexes. Des vues miniatures des représentations permettent la navigation (petites fenêtres de droite)



Tout d'abord, nous avons synchronisé les deux visualisations (matricielle et nœud-lien) afin de combiner leurs avantages et de faciliter l'identification des motifs visuels. Il est ainsi possible d'utiliser les matrices pour certaines tâches et les diagrammes nœud-lien pour d'autres. Si un analyste souhaite avoir une vue d'ensemble sur un réseau et, par exemple, identifier ses principales communautés, il est préférable qu'il utilise une représentation matricielle. Par contre, lors d'une analyse plus détaillée, s'il est nécessaire d'identifier un plus court chemin entre deux acteurs du réseau, le diagramme nœud-lien constitue une meilleure alternative.

De plus, synchroniser les deux représentations par la sélection permet de comparer la représentation d'un motif visuel dans l'une et l'autre représentation, facilitant ainsi leur compréhension (et l'apprentissage des matrices).

Pour manipuler interactivement les deux représentations nous avons mis en place un certain nombre de fonctionnalités :

1. spécification interactive des attributs visuels comme la couleur des nœuds, des en-têtes des lignes et de colonnes dans les matrices, ou bien encore l'épaisseur des liens ou la couleur des cellules de la matrice ;
2. méthodes automatiques et interactives pour le placement des nœuds des diagrammes nœud-lien et le réordonnancement des lignes et colonnes des matrices. En particulier, nous avons développé plusieurs outils d'aide au

réordonnancement en proposant d'appliquer des algorithmes automatiques sur des portions de matrices ou bien de bloquer l'ordre de plusieurs lignes et colonnes lorsque l'analyste souhaite les conserver ;

3. filtrage interactif de nœuds ou de connexions en fonction d'une sélection ou d'un attribut particulier présent dans les données (l'âge ou le sexe des personnes contenues dans le réseau social par exemple) ;

4. clustering interactif permettant d'identifier les différents groupes ou communautés détectées pour les marquer par un attribut visuel tel que la couleur ;

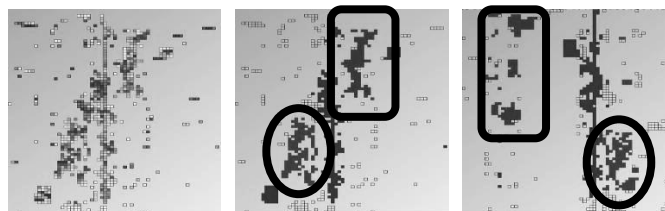
5. finalement, plusieurs vues d'ensemble sont proposées : miniatures des visualisations dans lesquelles l'analyste peut se déplacer ou représentations sous forme de Treemap de la macrostructure du réseau (voir **figure 8**).

En combinant les deux représentations et permettant de les manipuler interactivement, les utilisateurs de MatrixExplorer peuvent créer de multiples vues sur leurs données et les explorer à divers niveaux de détail.

Trouver un consensus

Comme chaque visualisation présente une perspective différente, nous avons tenté de favoriser l'établissement de consensus à l'aide d'interactions simples. La **figure 9** donne un exemple. Pour faciliter cette recherche de consensus, nous avons également introduit la sélection approximative : possibilité de marquer le degré d'appartenance d'un élément à un cluster (représenté par une couleur plus claire par exemple) ; et le clustering multiple : possibilité de marquer l'appartenance d'un élément à plusieurs clusters.

Figure 9. L'image du milieu montre l'identification de plusieurs clusters de la matrice initiale. Après réordonnancement, nous constatons que l'un des clusters (entouré par un ovale) est conservé alors que l'autre (entouré par un rectangle) se retrouve divisé en trois morceaux. Il y donc seulement consensus sur une partie des clusters identifiés ; il faut continuer à explorer pour en identifier éventuellement d'autres



Présenter l'information

Alors que la matrice permet principalement d'explorer les données et d'en manipuler une grande quantité, le diagramme nœud-lien est indispensable pour présenter des résultats et communiquer des découvertes. Ainsi, une fois le réseau filtré et éventuellement agrégé, MatrixExplorer permet d'extraire directement des images nœud-lien du résultat.

REPRESENTATIONS HYBRIDES

Combiner deux représentations présente beaucoup d'avantages, mais aussi plusieurs inconvénients. Tout d'abord, la taille de l'espace visuel : une utilisation confortable nécessite au moins deux écrans (un par représentation), un troisième offrant un confort supplémentaire en permettant de consulter les données détaillées éventuellement textuelles ainsi que plusieurs vues d'ensemble. De plus, nous avons observé que le passage d'une représentation à l'autre entraînait un coût cognitif non négligeable : un nœud unique se transforme en deux éléments dans la matrice (ligne et colonne) et un lien (visuellement représenté par un trait) devient une cellule (un rectangle) dans la matrice. Pour limiter ce coût cognitif, nous avons donc tenté d'améliorer les représentations matricielles pour limiter le recours au diagramme nœud-lien lors de l'exploration.

Augmenter les matrices

Comme l'ont montré les travaux sur la lisibilité de la matrice⁴⁰, la tâche de base pour laquelle cette représentation est moins performante que les diagrammes nœud-lien est le suivi de chemin. Pour remédier à ce problème, nous avons créé MatLink⁴¹, une représentation matricielle augmentée (figure 10).

Le principe de MatLink est de dessiner des liens sur les bords de la matrice afin de mieux représenter la connectivité et faciliter le suivi de chemin. Deux catégories de liens sont ajoutées sur la matrice : des liens statiques (en blanc sur l'image) et des liens interactifs (en plus foncé sur l'image) qui

40. GHONIEM, FEKETE et CASTAGLIOLA, 2005.

41. HENRY *et al.*, 2007.

apparaissent uniquement lors d'un survol ou d'un clic de la souris sur les en-têtes des lignes et colonnes. Lorsque l'utilisateur sélectionne deux éléments, les liens apparaissent montrant un plus court chemin.

Figure 10. MatLink augmente les matrices en ajoutant des liens statiques (en blanc) et des liens interactifs (en plus foncés sur les en-têtes) pour aider au suivi de chemin

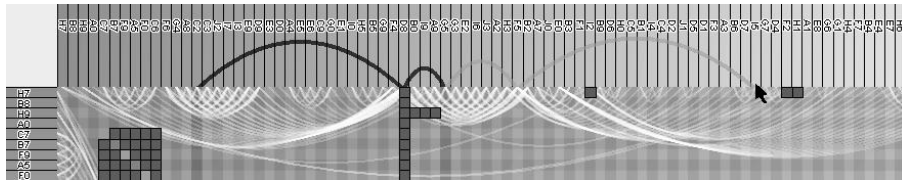
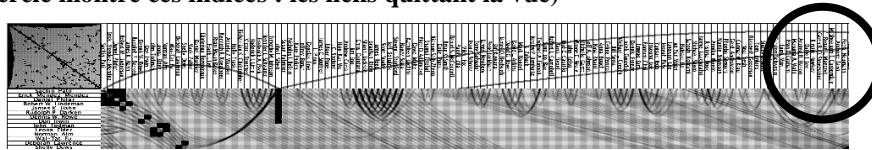


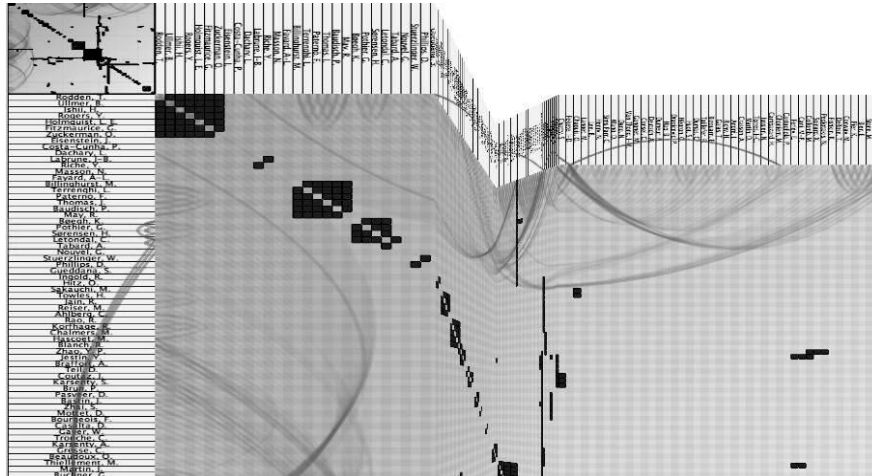
Figure 11. Les liens permettent à l'analyste d'être conscient que des éléments connectés aux nœuds qu'il étudie existent en dehors de la vue courante (le cercle montre ces indices : les liens quittant la vue)



Nous avons réalisé une expérimentation contrôlée similaire à celle de Ghoniem *et al.* afin d'étudier si cette représentation améliorerait significativement la représentation matricielle standard. Nous avons testé un lot de tâche de lisibilité de base ainsi que plusieurs tâches spécifiques à l'analyse de réseaux sociaux : trouver un point d'articulation, trouver une clique et identifier plusieurs communautés (groupe fortement connecté).

D'une manière générale, nous avons démontré que MatLink améliore significativement la représentation matricielle standard, en particulier pour la tâche de suivi de chemin, auparavant très laborieuse à réaliser. Pour ce cas précis, MatLink montre même des performances supérieures au diagramme nœud-lien pour les graphes denses (et ce, même augmenté de surlignage interactif) à cause du grand nombre de superpositions et de croisements. La seule tâche pour laquelle les diagrammes nœud-lien restent plus performants est l'identification de points d'articulation. Cette tâche requiert assez grande familiarité avec la représentation matricielle, ce qui n'était pas le cas de nos participants (qui, en moyenne, disposaient de 30 minutes d'entraînement avec les matrices).

Figure 12. Mélange permet de « plier » l'espace pour voir deux régions éloignées de la matrice tout en préservant le contexte intermédiaire



Naviguer dans de grandes matrices

Le second avantage de MatLink est l'amélioration que la technique offre pour la navigation dans de grandes matrices. Un des inconvénients des matrices est la grande taille requise pour les représenter. Ainsi, il peut (souvent) arriver qu'un acteur soit connecté à plusieurs autres acteurs placés en dehors de la vue. Il est nécessaire de naviguer sur toute la largeur (ou hauteur) de la matrice pour identifier tous les voisins d'un acteur particulier, ce qui est laborieux. Grâce aux liens affichés dans MatLink, et par un simple clic sur l'acteur étudié, l'analyste a un retour visuel immédiat sur tous les éléments connectés à cet acteur et possède une indication pour ceux placés hors de l'écran. La **figure 11** présente un exemple.

En outre, afin de faciliter l'exploration de ces grandes matrices (et le déplacement jusqu'aux éléments placés hors de la vue), nous avons mis au point une technique de navigation « pliant l'espace » : Mélangé⁴². Cette technique permet de visualiser deux parties de la matrice dans une même vue, en offrant un aperçu de l'espace intermédiaire, comme illustré dans La **figure 12**. Mélange offre même la possibilité de zoomer sur l'une des régions d'intérêt indépendamment de l'autre.

42. ELMQVIST *et al.*, 2008.

Figure 13. Représentation nœud-lien (à gauche) dans laquelle il est difficile d'identifier que les nœuds en noir ne sont pas connectés. La représentation NodeTrix (au milieu) visualise les sous-parties denses en matrices, il devient alors possible de voir les absences de connexions (matrice du haut gauche avec des cellules manquantes). Enfin, dupliquer des nœuds (à droite) permet d'améliorer la lisibilité en supprimant des croisements de liens inter-communautés

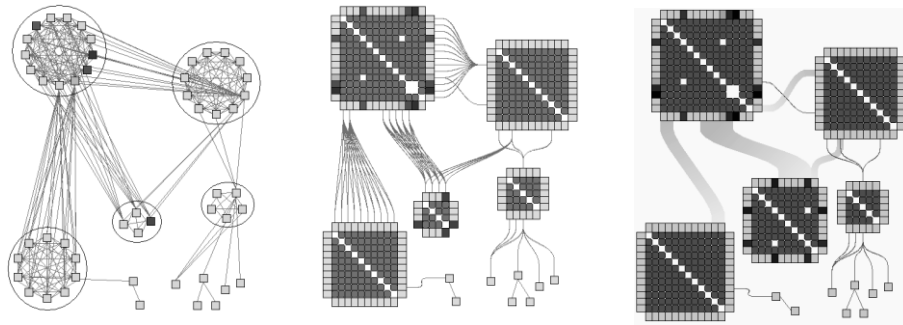
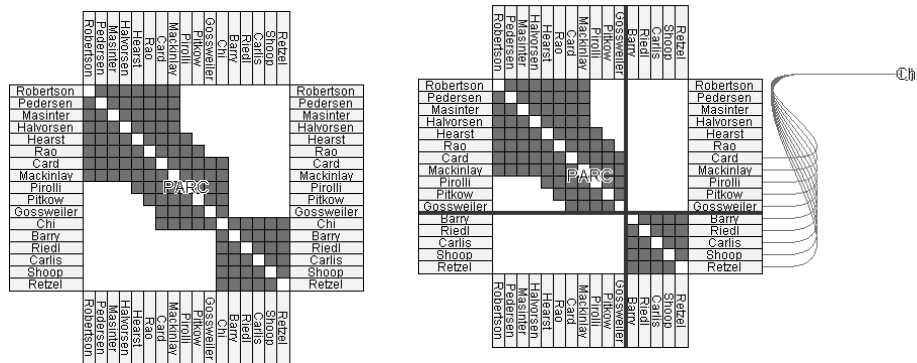


Figure 14. Interagir avec la représentation matricielle pour mieux comprendre la structure de la communauté. Ici, en ôtant une personne de la communauté, il est possible de voir qu'il s'agit d'un point d'articulation

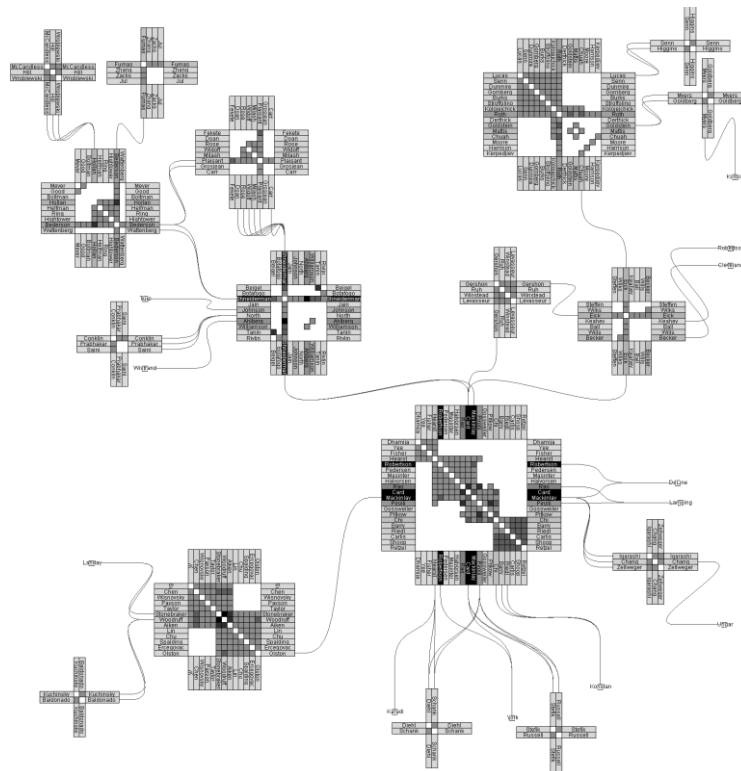


Fusionner matrices et nœud-lien

La structure des réseaux sociaux varie beaucoup : certains exhibent une structure d'arbre (ou presque) très peu dense tels les arbres généalogiques, d'autres ont des structures quasi complètes (réseau bipartite très dense) tels

que les réseaux d'import/export entre pays du monde et enfin, certains ont une structure intermédiaire, localement dense mais globalement creuse, il s'agit des réseaux petit-monde⁴³ tels que les réseaux d'amitiés ou encore de communication.

Figure 15. Représentation NodeTrix pour l'exploration. Les étiquettes sont toutes lisibles et permettent l'interaction pour éditer la représentation. (Réseau de copublication d'infovis contenant plus d'une centaine d'acteurs)



Les représentations visuelles que nous avons présentées précédemment sont plus ou moins performantes selon le type de réseau à visualiser. Par exemple, les diagrammes nœud-lien ainsi que les représentations basées sur des arbres avec liens supplémentaires (Treemap+liens par exemple) sont particulièrement efficaces pour les réseaux peu denses présentant une

43. WATTS et STROGATZ, 1998.

structure d'arbres. Au contraire, si le réseau à représenter est plutôt dense ou bipartite et présentant une structure quasi complète, la représentation matricielle est alors tout indiquée. Par contre, le choix n'est pas si simple pour les réseaux petit-monde.

Le problème que posent les réseaux petit-monde réside dans la variation de la densité locale. Si l'on choisit un diagramme nœud-lien, alors les parties denses (communautés) ne sont pas lisibles. Inversement, si l'on choisit une matrice, alors la majorité de l'espace est vide et l'exploration requiert beaucoup de navigation. Une possibilité est alors d'utiliser MatrixExplorer, combinant les deux représentations, mais le coût cognitif est plus élevé.

Pour ne plus avoir à choisir, nous avons mis au point une seconde représentation hybride, fusionnant matrice et diagramme nœud-lien : NodeTrix⁴⁴. Le principe de NodeTrix est de représenter le réseau global comme un nœud-lien mais d'utiliser une représentation matricielle pour ses sous parties les plus denses. Deux exemples sont présentés [figure 13](#) et [figure 15](#).

Explorer

Afin de faciliter la création, l'exploration et l'édition des matrices représentant les parties denses du réseau, nous avons développé un ensemble d'outils interactifs basés sur la technique du glisser-déposer. Il suffit de sélectionner au lasso un ensemble de nœuds fortement connecté pour transformer cette partie du réseau en matrice. Des informations sur la connectivité intra-communauté apparaissent immédiatement comme, par exemple, l'absence de connexion entre deux membres de la communauté, auparavant difficile à percevoir à cause de la multitude de croisements de liens à cet endroit.

En outre, la représentation matricielle offre l'avantage de placer linéairement en ligne et colonnes les membres de la communauté ; il devient plus aisé d'identifier les membres connecté à des acteurs extérieurs. Pour permettre l'exploration de cette représentation, il est possible d'ajouter ou d'ôter interactivement un acteur à la communauté. L'interaction est simple : il suffit de cliquer sur l'élément (nœud ou bien en-tête d'une ligne ou colonne dans la matrice) et ensuite le glisser-déposer à l'intérieur ou à l'extérieur de

44. HENRY *et al.*, 2007.

la matrice (**figure 14**). D'autres interactions permettent de fusionner deux matrices ou bien de les casser pour revenir à une représentation normale. Enfin, pour aider l'utilisateur à comprendre le changement de représentation, une courte animation permet de suivre la transformation⁴⁵.

Les matrices des parties denses peuvent être créées manuellement ou automatiquement en utilisant un algorithme de clustering. NodeTrix étant intégré à MatrixExplorer, il est également possible de glisser-déposer des parties d'une matrice dans une fenêtre NodeTrix. Cela permet de mieux comprendre la connectivité inter-communauté, ce qui est surtout important si les deux sous-matrices ne sont pas côte à côte dans la représentation matricielle standard.

Un inconvénient de cette représentation est la représentation concrète des communautés. Si un élément est membre de deux communautés, il faut faire un choix et le placer soit dans l'une des matrices, soit le laisser entre les deux (c'est un problème classique posé par l'agrégation hiérarchique de réseaux). Placer un acteur dans une des communautés peut biaiser l'interprétation de la représentation. Par contre, le laisser entre les deux communautés dégrade la lisibilité en introduisant beaucoup plus de liens et de croisements. Pour résoudre ce problème, nous avons proposé de dupliquer les acteurs pour les placer dans deux ou plusieurs communautés à la fois !

Notre étude utilisateur⁴⁶ a montré que l'impact potentiellement négatif des duplications sur l'interprétation du réseau peut être minimisé en représentant explicitement le lien de duplication. Après avoir exploré l'espace de conception possible, nous avons décidé de représenter ce lien par une bande (un lien plus épais) de couleur claire permettant de minimiser les interférences avec la représentation des connexions « réelles » du réseau. Un exemple est présenté dans la **figure 13**. Notre évaluation portait sur six tâches de lisibilité du réseau (dont certaines spécifiques à l'analyse de réseaux sociaux) pour deux réseaux de densités différentes. Les résultats ont démontré que les duplications permettaient de significativement améliorer la lisibilité de représentation et d'offrir une vue non biaisée de chaque communauté.

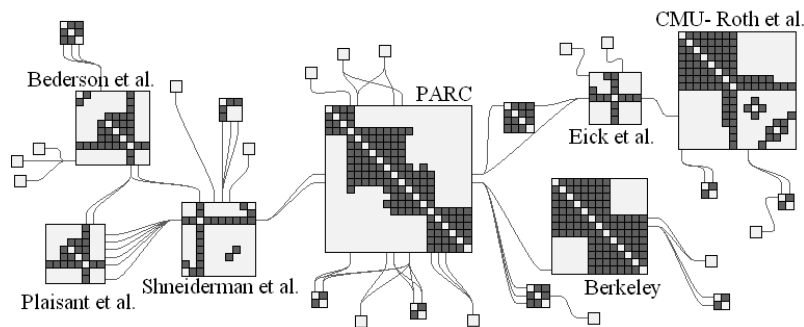
45. Une vidéo est disponible sur <http://www.aviz.fr/~nhenry>

46. HENRY *et al.*, 2008.

Présenter l'information

Le grand avantage de NodeTrix est sa versatilité : il est possible de l'utiliser pour explorer un réseau et découvrir sa structure mais aussi pour en faire une représentation très compacte et l'utiliser pour présenter des résultats. Un exemple est présenté [figure 16](#).

Figure 16. Représentation NodeTrix compacte pour la communication (réseau de co-publication d'InfoVis contenant plus d'une centaine d'acteurs). Des types de collaboration sont clairement visibles. Croix : un chercheur senior et ses étudiants (Bederson, Plaisant, Shneiderman, Eick). Blocs : des chercheurs collaborant beaucoup les uns avec les autres (Parc, Berkeley). Motif visuel intermédiaire (CMU-Roth)



Cette souplesse d'utilisation est liée à l'intégration des représentations matricielles et à leur capacité à produire à la fois une vue d'ensemble et une vue de détail. En mode présentation des résultats, il est alors possible de réduire au maximum la taille des communautés représentée par des matrices, faisant disparaître les étiquettes individuelles de chacun des membres (mais en conservant une étiquette pour la communauté par exemple) mais permettant de conserver un aperçu de la connectivité à l'intérieur de la communauté. [La figure 15 et la figure 16](#) sont un bon exemple, toutes deux représentent le même réseau, à des niveaux de détail différents.

CONCLUSION

Face aux défis lancés par l'analyse des réseaux sociaux en ligne, les sociologues ont besoin d'outils d'exploration efficaces. Aujourd'hui, les sociologues utilisent des outils d'analyse statistiques et analytiques ainsi que des outils de représentations graphiques (statiques) pour analyser leurs

données. Si ces outils permettent de répondre à de nombreuses questions et de valider nombre d'hypothèses, ils n'ont pas été conçus pour supporter le processus d'exploration.

Faciliter l'exploration des réseaux sociaux actuels est indispensable car leur taille est très souvent démesurée et leur structure méconnue. Les récentes avancées en visualisation d'information ouvrent une nouvelle voie : elles constituent à la fois une solution pour faciliter l'exploration de très grands réseaux et pour communiquer les résultats de cette analyse. Dans cet article, nous avons présenté un état de l'art de ces nouveaux outils d'exploration visuelle interactive.

Nous avons tout d'abord mis en lumière les problèmes de passage à l'échelle des diagrammes nœud-lien. Ces représentations sont familières et intuitives mais leurs problèmes de lisibilité pour les réseaux vastes et denses les rendent rédhibitoires pour explorer les réseaux sociaux actuels. Nous avons donc présenté plusieurs représentations alternatives permettant de résoudre ces problèmes de lisibilités. En particulier, nous nous sommes concentrés sur les représentations basées sur la matrice d'adjacence de graphe, qui permet de proposer une vue d'ensemble exploitable (et ce même pour de très vastes et denses réseaux).

Cette représentation présente plusieurs avantages par rapport aux nœud-lien, le plus important étant de supprimer toute superposition de nœuds et croisements de liens ce qui est en particulier utile pour les réseaux de forte densité. Ces représentations ont aussi plusieurs inconvénients comparées aux diagrammes nœud-lien : il est notamment très laborieux d'utiliser les matrices pour identifier des chemins entre plusieurs nœuds.

Nous avons cherché à combiner les avantages des deux représentations. Nous avons présenté MatrixExplorer, un système interactif permettant d'explorer les réseaux sociaux en manipulant matrices et diagrammes nœud-lien pour avoir de multiples perspectives sur les données. Nous avons également présenté deux représentations hybrides, fusionnant les deux représentations : MatLink, permettant d'améliorer significativement les tâches de suivi de chemin dans les matrices ; et NodeTrix, particulièrement efficace pour visualiser les réseaux petit-monde et constituant une solution compacte pour présenter des résultats.

Les travaux présentés dans cet article sont centrés sur l'exploration visuelle des réseaux et tentent d'aider les analystes à générer de nouvelles questions sur les données ou d'ouvrir de nouvelles pistes d'analyse. Nous avons constaté que l'exploration basée sur la création de multiples perspectives et la manipulation interactive des données permet d'identifier des éléments saillants des données et de générer de nouvelles questions, facilitant et stimulant ainsi l'analyse des réseaux sociaux.

Par exemple, l'exploration visuelle des réseaux de collaboration de chercheurs en interface homme-machine⁴⁷ a permis de révéler rapidement différents modes de production d'articles de recherche : « compétitif » et « collaboratif » (figure 16). Dans notre domaine de recherche, les professeurs d'université ont en général un comportement de mentor plutôt compétitif : ils encadrent de nombreux étudiants, ceux-ci collaborant rarement entre eux. Ils collaborent très peu entre eux et exhibent donc un mode de production d'articles plutôt compétitif, nécessaire aux Etats-Unis pour démontrer leur talent individuel et l'indépendance requise pour obtenir une tenure (équivalent de l'habilitation à diriger les recherches). Si l'on se réfère à l'apparence visuelle de la matrice des relations, nous observons une sorte de « croix » ; dans la figure 16, nous pouvons identifier les professeurs saillants en visualisation d'information : Plaisant, Bederson, Shneiderman et Eick.

Au contraire, les organismes de recherche privés (ou groupes universitaires ayant une majorité de chercheurs avec tenure) exhibent un motif visuel de collaboration « en bloc ». Ces chercheurs collaborent beaucoup les uns avec les autres, leur but étant de produire un grand nombre d'articles (et non pas d'obtenir une tenure), ils constituent généralement des cliques par thématique de recherche. Dans la figure 16, nous pouvons identifier Parc (4 sous-groupes) et Berkeley (2 cliques connectées). Certains motifs de collaboration sont intermédiaires tels que celui de CMU-Roth, où CMU représente un groupe de chercheurs seniors collaborant beaucoup entre eux et le groupe constitué du mentor Roth et de ces étudiants.

Ce rapide exemple démontre que l'exploration visuelle peut mener à d'intéressantes découvertes et rapidement attirer l'attention de l'analyste sur des éléments saillants de la structure d'un réseau social (tels que ces différents types de collaboration). Tout au long de cet article, nous avons

47. Exploration visuelle présentée dans HENRY *et al.* : *20 years of 4 HCI : a visual exploration*.

montré que les représentations matricielles nécessitent un certain apprentissage (pour pouvoir les décoder et ainsi les exploiter) mais constituent une piste prometteuse pour explorer des réseaux de dizaines de milliers de nœuds⁴⁸ (voire de quelques millions de nœuds⁴⁹). Par contre, elles ne permettent pas aujourd'hui d'explorer des réseaux de taille démesurée tels que Facebook. La recherche en visualisation d'information a donc encore de beaux jours, notamment en proposant des outils spécifiques pour l'exploration de réseaux orientés ou multi-variés.

Dans le futur, il nous semble indispensable de proposer des outils intégrant plus intimement l'exploration visuelle et analyse statistique. Effectivement, si l'exploration génère un lot de questions nouvelles ou orientent l'analyse, une analyse statistique est très souvent nécessaire pour répondre à ces questions et éliminer les biais potentiellement introduit par la représentation graphique et l'interprétation de cette représentation.

48. Dans *20 years of 4 HCI : a visual exploration*, HENRY *et al.* visualisent un réseau de plus de 26 000 acteurs et 118 000 relations.

49. ELMQVIST *et al.* présentent ZAME, utilisant les matrices d'adjacence zoomable pour analyser 6 millions de pages dans le wikipedia français.

RÉFÉRENCES

- ABELLO J., VAN HAM F. (2004), "Matrix zoom: A visual interface to semi-external graphs", *Proceedings of the 2004 IEEE Symposium on Information Visualization (INFOVIS'04)*, Austin, Texas, p. 183-190.
- ABELLO J., VAN HAM F., KRISHNAN N. (2006), "Ask-graphview: A large scale graph visualization system", *IEEE TVCG journal*, vol. 12, n° 5, p. 669-676.
- AHLBERG C., WILLIAMSON C., SHNEIDERMAN B. (1992), "Dynamic queries for information exploration: An implementation and evaluation", *Proceedings of the ACM CHI'92 Conference on Human Factors in Computing Systems*, New York, NY, USA, p. 619-626.
- ANSCOMBE F.J. (1973), "Graphs in Statistical Analysis", *American Statistician*, 27, p 17-21.
- AUBER D. (2003), "Tulip: A huge graph visualisation framework", *Graph Drawing Software*, Springer-Verlag, p. 105-126.
- BERTIN J. (1967), *Sémiologie graphique : les diagrammes - Les réseaux - Les cartes* (Les réimpressions ed.), Paris, France, Editions de l'Ecole des Hautes Etudes en Sciences.
- CARAUX G., PINLOCHE S. (2005), "Permutmatrix: A graphical environment to arrange gene expression profiles in optimal linear order", *Bioinformatics*, 21, p. 1280-1281.
- CARPENDALE M. S. T. (1999), RA Framework for Elastic Presentation Space, Ph. D. thesis, Simon Fraser University.
- COCKBURN A., MCKENZIE B. (2000), "An evaluation of cone trees. People and Computers XV", *Proceedings of the 2000 British Computer Society Conference on HumanComputer Interaction*.
- COCKBURN A., MCKENZIE B. (2002), Evaluating the effectiveness of spatial memory in 2d and 3d physical and virtual environments.
- COCKBURN A., MCKENZIE B. J. (2001), "3D or not 3D ? Evaluating the effect of the third dimension in a document management system", *Proceedings of SIG CHI 2001*, p. 434-441.
- DI BATTISTA G., EADES P., TAMASSIA R., TOLLIS I. G. (1998), *Graph Drawing: Algorithms for the Visualization of Graphs*, Prentice Hall PTR. 14.
- ELMQVIST N., DO T.-N., GOODELL H., HENRY N., FEKETE J.-D. (2008), "ZAME: Interactive Large-Scale Graph Visualization", *Proceedings of the IEEE*

Pacific Visualization Symposium 2008, p. 215-222.

ELMQVIST N., HENRY N., RICHE Y., FEKETE J.-D. (2008), "Mélange: space-folding for multi-focus interaction", *CHI'08: Proceedings of the SIGCHI conference on Human Factors in computing systems*, New York, NY, USA.

ELMQVIST N., TSIGAS P. (2007), "A taxonomy of 3d occlusion management techniques", *Proceedings of the IEEE Conference on Virtual Reality*, 2007, p. 51-58.

FEKETE J.-D., WANG D., DANG N., PLAISANT C. (2003), "Overlaying graph links on treemaps", *IEEE Symposium on Information Visualization Conference Compendium* (demonstration).

FRANK O. (1978), "Sampling and estimation in large social networks", *Social networks*, 1, p. 91-101.

FREEMAN L. C. (2000), "Visualizing social networks", *Journal of social structure* vol. 1, n° 1.

GHONIEM M., FEKETE J.-D., CASTAGLIOLA P. (2005), "On the readability of graphs using node-link and matrix-based representations: a controlled experiment and statistical analysis", *Information Visualization*, vol. 4, n° 2, p. 114-135.

HEER J., BOYD D. (2005), "Vizster: Visualizing Online Social Networks", *Proceedings of the IEEE Symposium on Information Visualization*, p. 5.

HENRY N., BEZERIANOS A., FEKETE J.-D. (2008), "Improving the readability of clustered social networks by node duplication", *IEEE Transactions on Visualization and Computer Graphics (Proceedings of Visualization/Information Visualization 2008)*, vol. 14.

HENRY N., FEKETE J.-D. (2006a), "MatrixExplorer: a Dual-Representation System to Explore Social Networks", *IEEE TVCG (Infovis '06 proceedings)*, vol. 12, n° 5, p. 677-684.

HENRY N., FEKETE J.-D. (2007), "Matlink: Enhanced matrix visualization for analyzing social networks", *Lecture Notes in Computer Science (Proceedings of the 13th IFIP TC13 International Conference on Human-Computer Interaction, INTERACT'07)*, vol. 4663, p. 288-302.

HENRY N., FEKETE J.-D. (2006b), « Matrixexplorer : un système pour l'analyse exploratoire de réseaux sociaux », *Proceedings of IHM2006*, International Conference Proceedings Series, p. 67-74.

HENRY N., FEKETE J.-D., MCGUFFIN M. (2007), "Nodetrix: Hybrid representation for analyzing social networks", *TVCG (Proceedings of IEEE Information Visualization conference)*, vol. 13, n° 6, p. 1302-1309.

HENRY N., GOODELL H., ELMQVIST N., FEKETE J.-D. (2007), "20 Years of 4 HCI Conferences: a Visual Exploration", *International Journal of Human Computer Interaction — Reflections on Human-Computer Interaction, A Special*

Issue in Honor of Ben Shneiderman's 60th Birthday, vol. 23, n° 3, p 239-285.

HERMAN I., MELANÇON G., DE RUITER M. M., DELEST M. (1999), "Latour, a tree visualisation system", *Proceedings of the Graph Drawing symposium (GD'99)*, LNCS 1731, p. 392-399.

HERMAN I., MELANÇON G., MARSHALL S. (2000), "Graph visualization and navigation in information visualization: A survey", *IEEE TVCG*, vol. 6, n° 1, p. 24-43.

HUISMAN M., DUIJN M. A. J. (2005), "Software for social network analysis" (chapter 13), Nooy V. B. W. de, Mrvar A. (Ed.), *Exploratory Social Network Analysis with Pajek*.

JAIN A. K., MURTY M. N., FLYNN P. J. (1999), "Data clustering: a review", *ACM Computing Surveys*, vol. 31, n° 3, p. 264-323.

LAMPING J., RAO R. (1996), "The Hyperbolic Browser: A focus + context technique for visualizing large hierarchies", *Journal of Visual Languages and Computing*, vol. 7, n° 1, p. 33-35.

Lee B., Parr C. S., Plaisant C., Bederson B. B., Veksler V. D., Gray W. D., Kotfila C. (2006), "Treeplus: Interactive exploration of networks with enhanced tree layouts", *IEEE TVCG (Infovis'06 proceedings)*, vol. 12, n° 6, p. 1414-1426.

MORENO J. L. (1934), *Who shall survive ? Foundations of sociometry, group psychotherapy and sociodrama*. Beacon House, New York.

MUNZNER T. (1997), "H3: Laying out large directed graphs in 3d hyperbolic space", *Symposium on Information Visualization (InfoVis'97)*, p. 2-10.

PERER A., SHNEIDERMAN B. (2006), "Balancing Systematic and Flexible Exploration of Social Networks", *IEEE TVCG (Infovis'06 proceedings)*, vol. 12, n° 5, p. 693-700.

RISDEN K., CZERWINSKI M., MUNZNER T., COOK D. (2000), "An initial examination of ease of use for 2D and 3D information visualizations of web content", *International Journal of Human Computer Studies*, vol. 53, n° 5, p. 695-714.

SHNEIDERMAN B. (1983), "Direct manipulation: A step beyond programming languages", *IEEE Computer*, vol. 16, n° 8, p. 57-69.

SHNEIDERMAN B. (1992), "Tree visualization with tree-maps: A 2-D space-filling approach", *ACM Transactions on Graphics*, vol. 11, n° 1, p. 92-99.

SHNEIDERMAN B. (1996), "The Eyes Have It: A Task by Data Taxonomy for Information Visualization", *Visual Languages*, p. 336-343.

SHNEIDERMAN B., ARIS A. (2006), "Network visualization by semantic substrates", *IEEE TVCG (Infovis'06 proceedings)*, vol. 12, n° 5.

SUTCLIFFE A. G., PATER U. (1996), "3D or not 3D : is it nobler to the mind?", *British Human Computer Interaction Conference*, Cambridge University Press,

p. 79-94.

TUKEY J. (1977), *Exploratory Data Analysis*, Addison-Wesley.

VAN HAM F. (2003), "Using multilevel call matrices in large software projects", *Proceedings of the 2003 IEEE Symposium on Information Visualization*, Seattle, WA, USA, IEEE Press, p. 227-232.

WASSERMAN S., FAUST K. (1994), *Social Network Analysis*, Cambridge Univ. Press.

WATTS D. J., STROGATZ S. H. (1998), "Collective dynamics of 'small-world' networks", *Nature* 393, p. 440-442.

ZIEGLER J., KUNZ C., BOTSCH V. (2002), "Matrix browser visualizing and exploring large networked information spaces", *Extended Abstracts of Human Factors in Computing Systems (CHI'02)*, p. 602-603.