



HAL
open science

Visualisation des réseaux: apports, défis et enjeux du travail sur les données historiques

Anne Baillot

► To cite this version:

Anne Baillot. Visualisation des réseaux: apports, défis et enjeux du travail sur les données historiques. Numérisation de masse et traitement des grands corpus de textes utilisant des méthodes des humanités numériques, Mar 2015, Stuttgart, Allemagne. <halshs-01130425>

HAL Id: halshs-01130425

<https://shs.hal.science/halshs-01130425v1>

Submitted on 11 Mar 2015

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.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License



Visualisation des réseaux : apports, défis et enjeux du travail sur les données historiques

Anne Baillot, mars 2015¹

« Un petit dessin vaut mieux qu'un long discours » : l'assertion semble avoir envahi non seulement la vie courante, mais aussi, sans doute plus souvent qu'il ne serait bon, la communication scientifique. L'infographie est omniprésente, comme si elle pouvait transmettre plus efficacement les données essentielles d'un problème qu'un tableau de chiffres de type matriciel ou un texte descriptif/analytique. Partir du principe qu'une représentation visuelle est plus intuitive qu'une représentation chiffrée ou textuelle reste cependant une prémisse contestable.

D'une part, le design visuel repose sur des codes de lectures précis ; l'ouvrage de référence de Jacques Bertin sur ce sujet, *Sémiologie graphique*, met bien en évidence qu'il s'agit d'un système de signes à part entière. Dans les premières pages de son ouvrage, il met en relation le fonctionnement de ce système de signes et sa possible contribution à la recherche :

« La représentation graphique fait partie des systèmes de signes que l'homme a construits pour retenir, comprendre et communiquer les observations qui lui sont nécessaires. [...] La graphique tient ses lettres de noblesse de sa double fonction de mémoire artificielle et d'instrument de recherche. [...] La grande différence que l'on perçoit maintenant entre la représentation graphique d'hier, mal dissociée de l'image figurative, et la graphique de demain, c'est la disparition de la fixité congénitale de l'image. Devenue manipulable par superposition, juxtapositions, transformations, permutations, autorisant groupements et classements, l'image graphique est passée de l'image morte, de l'illustration, à l'image vivante, à l'instrument de recherche accessible à tous. La graphique n'est plus seulement la re-présentation de la simplification finale, c'est aussi, c'est surtout, le point de départ exhaustif et l'instrument qui permet de découvrir et de défendre cette simplification. »²

Or, rares sont les disciplines des sciences humaines dans lesquelles le décryptage des codes inhérents aux systèmes graphiques fait partie de la formation universitaire. Comment et où, donc, apprendre à les lire ?

D'autre part, pour revenir à ma formulation introductive, il est rare qu'un petit dessin soit effectivement totalement en mesure de remplacer un long discours. La plupart des visualisations sont illisibles sans légende, et requièrent même souvent un texte analytique élucidant les différents éléments « visibles » pour développer la totalité de leur potentiel d'information.

¹ L'ensemble des liens cités dans cet article a été consulté le 10 mars 2015.

² Jacques Bertin, *Sémiologie graphique. Les diagrammes – Les réseaux – Les cartes*, Paris, 1967, p. 6-8.



La visualisation de réseaux fait partie des outils devenus quasi incontournables dans les humanités numériques : réseaux de Tweets (Fig. 1), réseaux de correspondants (Fig 2.), réseaux de co-auteurs (Fig. 3).

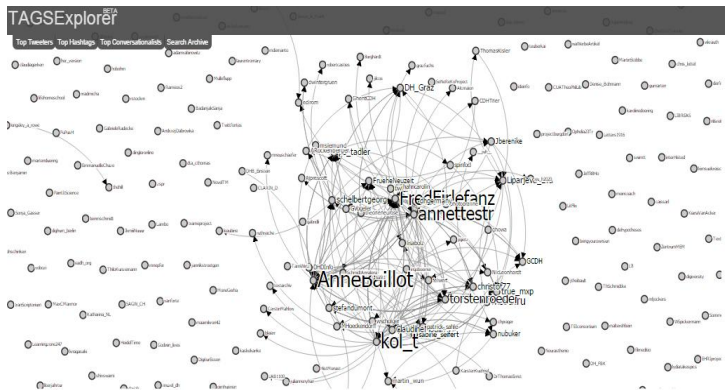


Fig. 1 : Réseau de tweets autour du mot-dièse #dhd2015

Source :

<http://hawksey.info/tagsexplorer/?key=1HMujSXE9OrBO2VEnIW2xrLK1BWNrSGurhbdfk3WxD4&gid=400689247>



Fig. 2 : Réseau généré à partir de métadonnées archivistiques.

Source : <http://kalliope-verbund.info/de/ueber-kalliope/tourdhorizon/kapitel-4.html>

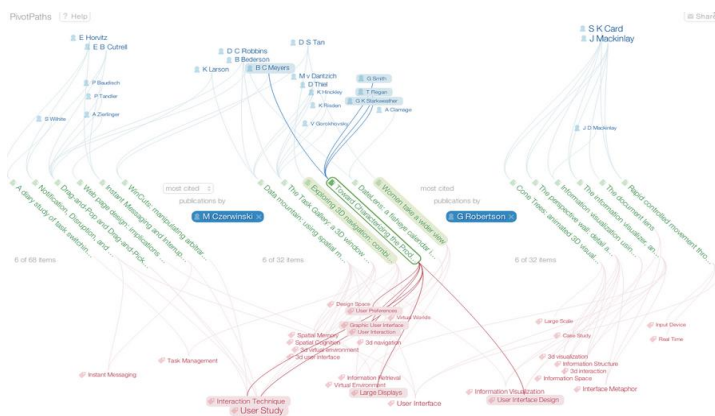


Fig. 3 : Réseau de co-auteurs sous la forme de PivotPaths

Source : <http://mariandoerk.de/pivotpaths/>



Dans tous ces cas, il s'agit de mettre en relation une production (artistique, scientifique) et des acteurs (auteurs, lecteurs). Cet usage de la visualisation de réseaux met en évidence les racines sociologiques de la méthode qui y préside, la Social Network Analysis. Les origines sociologiques de la visualisation de réseaux à partir de données historiques sont essentielles pour comprendre leur fonctionnement. Shneiderman, Hansen et Smith mettent bien en évidence ces enjeux lorsqu'ils écrivent :

« The focus of social network analysis is between, not within people. Whereas traditional social-science research methods such as surveys focus on individuals and their attributes (e.g. gender, age, income), network scientists focus on the connections that binds individuals together, not exclusively on their internal qualities or abilities. This change in focus from attribute data to relational data dramatically affects how data are collected, represented, and analyzed. Social network analysis complements methods that focus more narrowly on individuals, adding a critical dimension that captures the connective tissue of societies and other complex interdependencies. »³

Mais il n'y a pas que les relations sociales qui puissent être visualisées sous la forme de réseaux ; ceux-ci s'appliquent potentiellement à une palette beaucoup plus large de données de recherche. Isabelle Meirelles ouvre son chapitre sur les réseaux sur le constat que nous sommes entourés de réseaux et indique les caractéristiques générales de la recherche sur les réseaux :

« The study of networks is not new, as shown by early research in fields as varied as biology, sociology, and mathematics [...]. The scientific study of networks – network science – is, however, more recent and focuses on the study of patterns of connections in real-world systems. According to Barabási, four key characteristics distinguish network science as a discipline from early studies of networks: it is interdisciplinary; it examines empirical data; it is quantitative and mathematical in nature; and it relies on computational tools. »⁴

La visualisation de réseaux est une tentation d'autant plus grande qu'elle peut sembler au premier abord offrir un doux remède à un certain nombre de maux propres aux recherches en sciences humaines. Elle permet ainsi d'intégrer, avec les changements d'échelle, le passage du très grand au plus petit (et inversement). Au sein d'un même grand réseau, on peut zoomer sur une zone particulière et visualiser le détail d'un réseau dans lequel entités et relations sont singularisées. Martin Grandjean parle de réseaux multi-couches (multi-layered)⁵, on trouve également l'expression « échelonnable » (scalable). De manière générale, les réseaux offrent la promesse de pouvoir représenter la totalité des relations pertinentes dans un domaine donné, de la vision générale de la grande masse au détail local.

³ Ben Shneidermann et al., *Analyzing Social Media Networks with NodeXL: Insights form a Connected World* (2010), cité par Isabel Meirelles, *Design for Information*, Beverly, 2013, p. 47.

⁴ Ibid, p. 48.

⁵ Cf. <http://www.martingrandjean.ch/intellectual-cooperation-multi-level-network-analysis/>.



Un premier caveat est ici nécessaire avant d'en venir au bon usage de la visualisation de réseaux dans le contexte de données historiques. Il règne en effet une certaine confusion terminologique. Le terme « analyse de réseaux » est souvent employé dans des situations de recherche où il ne s'agit pas à proprement parler de visualisation de réseaux, sujet dont je parle plus spécifiquement aujourd'hui. On peut en effet avoir recours à des réseaux soit pour en faire l'analyse (éventuellement en s'appuyant sur une visualisation, mais pas forcément), soit la visualisation a en elle-même une fonction principalement d'illustration. En pratique, c'est plus souvent cette dernière constellation qui est mise en œuvre dans la recherche dite historique sur les réseaux, même lorsqu'est annoncé un travail d'« analyse » de réseaux. De manière intéressante, il y a souvent une certaine réticence à admettre cet usage du réseau visualisé à titre illustratif, comme si l'illustration n'avait pas la noblesse argumentative de l'analyse. Or, ce mépris est infondé si l'on intègre la relation entre le texte (analytique) et l'image (illustrative) à une épistémologie bien définie de la démarche de recherche.

Deuxième caveat propre lui aussi aux humanités numériques : l'internet étant lui-même un réseau, la métaphore du réseau a aussi structurellement un charme narcissique. Le réseau se fait volontiers reflet du réseau, et réfléchit, ou plutôt auto-réfléchit son propre fonctionnement dans ce modèle métaphorique. La notion de réseau sur laquelle repose ce fonctionnement est cependant assez floue, il s'agit le plus souvent de connecter des entités, sans définition particulière ni des entités ni de la nature ou de la force des connections. Sebastian Gießmann résume ainsi :

« Wir sind alle nur 'Netzeffekte', die nie über gebaute Netze in ihrer Gesamtheit verfügen können, sondern immer auf den Anschluss an diese angewiesen sind. Dahingegen verspricht der Begriff des Netzwerks die Möglichkeit aktiver Teilhabe. »⁶

Ce qui m'amène à mon troisième et dernier caveat : le français ne dispose que d'un seul terme, réseau, recouvrant les deux termes allemands « Netz » et « Netzwerk ». Sebastian Gießmann revient sur cette différence linguistique (également présente en anglais avec net et network) et en conclut que le réseau-Netzwerk (par exemple dans le cas de la toile d'araignée) renvoie à une performativité qui n'est pas présente dans le réseau-Netz (filet). Ce sur quoi porte mon propos, ce sont donc davantage les réseaux au sens de Netzwerke, en tant que leur visualisation permet de générer de l'information scientifique.

Toutes ces précautions liminaires étant prises – et sans doute faudrait-il y en ajouter d'autres, ce à quoi je renonce pour pouvoir entrer enfin dans le vif de mon propos –, il s'agira donc ici de tâcher de s'affranchir des effets de miroir, en partie déformants, et des déterminations linguistiques, et de s'interroger plus particulièrement sur ce que permet la visualisation de réseaux lorsqu'elle est appliquée à des données historiques.

⁶ Sebastian Gießmann, *Netze und Netzwerke. Archäologie einer Kulturtechnik 1740-1840*, Bielefeld, 2006, p. 16: „Nous sommes donc tous des „effets de réseau“ qui ne peuvent jamais disposer dans leur totalité des réseaux (filets) construits. Au contraire, nous sommes à la merci des connections qu'ils peuvent offrir. Inversement, le concept de réseau (toile) offre la promesse de pouvoir participer activement. »



Qu'est-ce que les réseaux permettent de visualiser ? Principalement, des relations.⁷ Les points figurent les acteurs et les traits qui les relient les types de relations dont il est question. Isabel Meirelles, Claire Lemerrier⁸, Marten Düring, bref, à peu près toutes les approches insistent sur le fait que l'essentiel, dans un réseau, repose sur la définition de la ou des relations représentées. Scott Weingart ouvre sa série de posts de blog sur le sujet « Demystifying Networks » par la remarque suivante :

“Networks can be used on any project. Networks should be used on far fewer. Networks in the humanities are experiencing quite the awakening, and this is due in part to until-recently untapped resources. There is a lot of low-hanging fruit out there on the networks+humanities tree, and they ought to be plucked by those brave and willing enough to do so. However, that does not give us an excuse to apply networks to everything.”⁹

Les traits reliant les points entre eux peuvent indiquer ou non une réciprocité ; lorsque la direction de la relation doit également être représentée, c'est le plus souvent par l'ajout au trait d'une pointe de flèche pour désigner la ou les directions pertinentes. On parle alors de réseau dirigé (directed). Les relations ainsi représentées peuvent être de tout genre : amitié/inimitié, collaboration/concurrence, égalitaire/hiérarchique. Dans son schéma générique, une visualisation de réseau ne comporte donc que peu de codes imposés allant au-delà du trait entre deux points. Elle offre en apparence une certaine marge de manœuvre – et il est certain que, sous l'angle du design, la recherche en visualisation a fait preuve ces dernières années d'une grande créativité.¹⁰

Ce principe générique est déclinable en différents types de visualisation, qui permettent de mettre l'accent sur différents aspects, certains tout à fait à même d'être mis en corrélation avec l'analyse historique dans son sens traditionnel. On peut ainsi mettre en valeur dans la représentation qu'on élabore le rôle joué par certains acteurs ou le type de connexion qui les relie à d'autres (liens forts ou faibles, nombre d'intermédiaires reliant 2 acteurs) ; on met alors l'accent sur la connectivité, ce qui permet notamment de faire émerger des figures centrales. On peut également visualiser l'intensité des relations.

Pour les données historiques, quatre types de visualisation de réseaux sont particulièrement pertinents : la visualisation chronologique (évolution au fil du temps), la visualisation géographique (projection des relations entre acteurs sur une carte, sachant que se pose dans ce contexte le problème de l'historicité des cartes, puisque la plupart du temps, c'est une carte tirée de googlemaps ou openstreetview et les coordonnées géographiques tirées

⁷ „Relationships matter“; cf. „Networks in Archaeology: Phenomena, Abstraction, Representation“, introduction co-signé par Anna Collar, Fiona Coward, Tom Brughmans et Barbara J. Mills: *Journal of Archaeological Method and Theory*, March 2015, Volume 22, Issue 1, pp 1-32 (DOI 10.1007/s10816-014-9235-6), ici p. 5.

⁸ Cf. Claire Lemerrier, „Formal network methods in history: why and how“; halshs-00521527.

⁹ Cf. <http://www.scottbot.net/> pour la série de posts ; ici <http://www.scottbot.net/HIAL/?p=6279>. Voir aussi C. Lemerrier, Formal network methods (cf. note ci-dessus), p. 4-5.

¹⁰ Je renvoie sur ce point au magnifique ouvrage d'Isabel Meirelles (cf. note 3).



de Wikipedia qui sont utilisés), la visualisation sociale (englobant un grand nombre d'acteurs) et égo-réseau (centré autour d'un acteur). Il n'existe aucun code graphique pré-défini, mais l'utilisation de couleurs connotées (contraste rouge/bleu par exemple) ou le jeu sur la largeur des traits pour signifier l'importance relative d'une relation font partie des présupposés visuels qui ont à peine besoin d'être spécifiés en légende. En revanche, il est rare que la totalité des conventions visuelles présupposables épuise la complexité et la subtilité souhaitées, a fortiori pour des données historiques.

Mais si l'on peut ainsi voir dans des conventions issues du design ou des sciences de l'information des obstacles à la représentation adéquate de données historiques sous la forme de réseau, c'est, pour être tout à fait honnête, dans la structure de ces données même qu'il faut en réalité aller chercher la pierre d'achoppement de tout l'édifice. La question fondamentale est la suivante : Quelles données historiques sont à même de permettre de représenter des relations sous forme de réseau ?

Partons de la notion de relation. Il existe un certain nombre de relations objectivables, comme les relations familiales par exemple. Ces relations (mariages, naissances, baptêmes, divorces, décès) sont documentées dans des sources historiques tels que les registres des paroisses ou des mairies. Si des erreurs s'y glissent, il y a tout lieu de croire qu'il ne s'agit là que de phénomènes marginaux qui n'ont aucune raison d'entamer la crédibilité de la source. Mais une relation parent-enfant ou une fratrie peuvent tout aussi bien donner lieu à des phénomènes de coopération qu'à des phénomènes de concurrence – et ce n'est pas dans un registre de mariage ou de baptême que l'on peut trouver cette information, qui est celle qui spécifie la qualité de la relation, potentiellement plus pertinente que la seule affiliation familiale. Qui plus est, plus l'on s'écarte de la documentation d'une identité quasi notariale, plus se pose la question de la pertinence des données historiques en termes d'objectivité. Par exemple, les relations de maître à élèves sont à première vue, elles aussi, repérables objectivement dans les listes de présence en cours (l'Université de Berlin a fait en tenir dans les années qui suivirent sa fondation, par exemple). Mais il faut réunir bien plus d'indices pour pouvoir en conclure la qualité réelle de l'impact exercé par (le plus souvent) le maître sur l'élève : par Hegel sur Hotho et Rosenkranz ou par Boeckh sur Lepsius et Trendelenburg. On peut se référer à ce qu'ils en disent dans leur correspondance, aux dédicaces de leurs ouvrages, aux citations qu'on trouve dans ceux-ci pour s'en faire une idée plus précise. Mais tous ces éléments mis ensemble aboutissent davantage à une interprétation d'une qualité de relation qu'à son objectivation complète. On observera une tendance, mais on ne pourra pas entrer avec certitude dans le détail de la relation.

Je pense ne pas me tromper en disant qu'actuellement, ce sont en très large partie des égo-documents qui sont utilisés dans la recherche historique sur les réseaux, ou du moins des documents mettant en scènes des acteurs le plus souvent réels, de manière à mettre en évidence leurs stratégies relationnelles dans un contexte donné. Mais peut-on prendre des égo-documents au pied de la lettre ? L'analyse *littéraire* d'égo-documents met en valeur leur dimension construite, la mise en scène et en œuvre qui préside à la transmission des



informations, tandis que l'analyse *historique* du même corpus va considérer les mêmes éléments comme des faits. Mais peut-être Humboldt n'a-t-il pas, en réalité, rendu visite à Rahel Levin, peut-être Schlegel ne trouvait-il pas Tieck aussi brillant que cela... pour tirer de documents de ce type des informations historiquement solides, il faudrait pouvoir les recouper au moins trois fois, ce qui est la plupart du temps impossible soit par manque de temps, soit par manque de sources.

La question de l'accès aux sources est cruciale pour le travail sur des données historiques. Elle contribue à déplacer les méthodes pertinentes pour les corpus du XX^{ème} siècle, conservés de manière plus complète et systématique, par comparaison avec les méthodes utilisées pour des corpus des siècles antérieurs qui sont, eux, conservés de manière beaucoup plus fragmentaire. Pour tirer une analyse statistiquement significative d'égo-documents, il faut en effet non seulement pouvoir les recouper entre eux afin d'obtenir des informations fiables, mais aussi s'assurer de la représentativité intrinsèque de l'ensemble du corpus. S'il est parfois possible de travailler avec des substituts (surrogates) permettant de compenser l'absence d'une partie des informations nécessaires pour garantir la cohérence du corpus, c'est loin d'être toujours possible.

Lorsque l'on se demande s'il peut y avoir des big data en sciences humaines, et plus particulièrement en histoire, il s'agit donc en réalité moins d'une question de quantité des données que d'une question de structure de celles-ci. Sans revenir sur la question du workflow de la numérisation et de l'OCR-isation, se pose la question fondamentale de l'inhomogénéité des données historiques, pointée notamment par Andreas Haka dans sa thèse sur les réseaux sociaux dans le domaine du génie mécanique dans les institutions de recherche en Allemagne entre 1920 et 1970¹¹. Pour visualiser un réseau de relations, il faut disposer d'informations structurées de manière similaire pour l'ensemble des acteurs et l'ensemble de relations qu'il s'agit de représenter. Il faut pouvoir, en d'autres termes, comparer ce qui est comparable. C'est à ce niveau-là que se pose la question du changement d'échelle de la manière la plus aiguë, car plus le corpus est grand, plus il est difficile de disposer de données homogènes. Or plus le corpus est grand, plus il est utile de pouvoir en visualiser la structure, car la visualisation permet de se repérer dans un ensemble de relations qui n'est plus simplement saisissable par l'esprit.

À mon sens, la question centrale qui se pose pour la visualisation de réseaux à partir de données historiques est celle des conditions de possibilité d'agrégation des données-sources. Les changements d'échelle, la lisibilité et la pertinence de la visualisation, la possibilité de mettre sur pied des comparaisons : tout en dépend. Pour donner une idée plus concrète de ce que cela signifie, je vais prendre un exemple tiré de mon travail de recherche sur les correspondances d'intellectuels autour de 1800.

¹¹ Andreas Haka, *Soziale Netzwerke im Maschinenbau an deutschen Hochschul- und außeruniversitären Forschungseinrichtungen 1920-1970*, Berlin, 2014.



Pour visualiser des réseaux à partir de correspondances, plusieurs approches sont possibles. On peut par exemple se contenter de relever les noms des acteurs et le nombre de lettres connus entre chaque acteur, ainsi que la date de la lettre. L'intérêt de la visualisation, dans ce cas, consiste à repérer l'importance respective des échanges comme marqueurs de transferts scientifiques potentiels. Cela suppose de considérer que le nombre de lettres encore conservés est statistiquement pertinent par rapport à la masse réelle de départ, qu'il est impossible de reconstruire complètement. Ce genre de réseaux est très général (voir Fig. 4), mais fait émerger des tendances. On peut aussi travailler avec un nombre plus réduit d'entités, mais en spécifiant davantage le type de relations, c'est-à-dire en entrant dans le contenu de la correspondance.

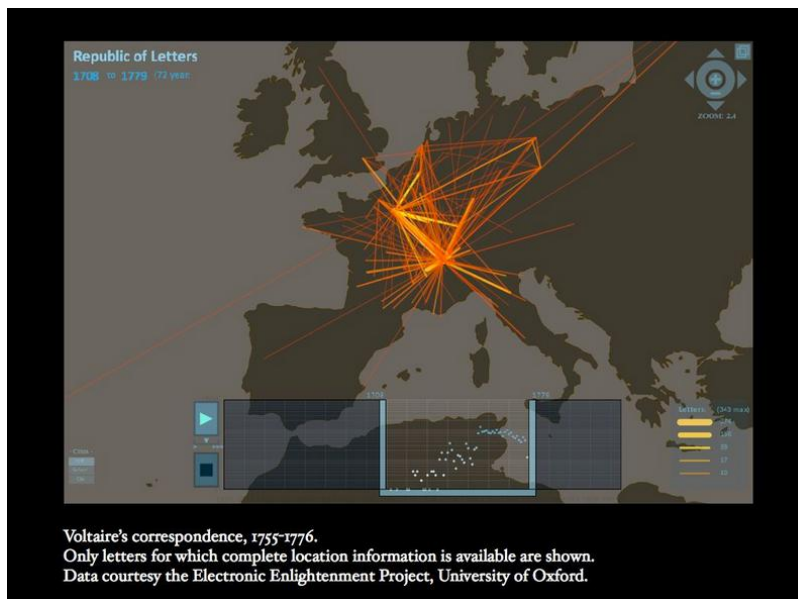


Fig. 4 : Les réseaux de correspondance de Voltaire combinant les dimensions géographique et temporelle.

Source : <http://republicofletters.stanford.edu/casestudies/voltaire.html>

Les lettres, et plus particulièrement les correspondances (donc les échanges de lettres), renseignent sur l'échange d'informations ; les correspondances savantes permettent ainsi d'approfondir les transferts scientifiques. Les savants s'écrivent les uns les autres au sujet de ce qu'ils ont lu, de ce qu'ils aimeraient lire ou vont lire ; ils détaillent les ouvrages qu'ils souhaitent acheter, mentionnent les revues qu'ils ont consultées. Ils s'écrivent ce qu'ils pensent, leur avis sur des livres, des mises en scène, des personnes, des idées qui sont dans l'air du temps. En théorie, on pourrait, en partant d'un savant, retracer le réseau de ses habitudes de lecture et d'écriture en les articulant autour d'idées et d'auteurs mentionnés dans sa correspondance avec d'autres savants.

En isolant les correspondances singulières, on pourrait également voir comment le réseau varie selon le destinataire de la lettre, c'est-à-dire si certains sujets ou certains auteurs sont davantage évoqués dans le cadre d'une correspondance plutôt que d'une autre, ce qui permettrait dans une certaine mesure d'opérer un correctif aux éventuelles distorsions qui



peuvent apparaître si l'on considère l'ensemble. De cette manière, on peut à la fois présenter la position d'un savant donné dans un contexte donné, voire même introduire une dimension temporelle dynamique sur une période donnée. On pourrait ainsi rassembler un grand nombre d'ego-réseaux d'une période donnée et utiliser les recoupements de personnes, d'œuvre ou de thèmes, pour extrapoler à partir de là un macro-réseau. C'est ce que se propose de faire le projet „Reassembling the Republic of Letters,„ financé par l'Union Européenne¹². Ce réseau de chercheurs, travaillant pour la plupart sur des correspondances, rassemble plus de 200 partenaires.

Sans pouvoir complètement extrapoler sur la nature des discussions qui animeront nos travaux à cette occasion et dans les années qui viennent, je peux d'ores et déjà pressentir au moins trois aspects incontournables.

Pour relier entre eux les réseaux savants issus des différentes éditions de correspondance, nous aurons besoin d'interfaces et d'interopérabilité. Dans le cas de correspondances savantes, les entités pertinentes sont les personnes, les œuvres et les idées.

- 1) Les personnes : Dans une édition numérique (ou un registre archivistique), les personnes sont relevées et, dans le meilleur des cas, mises en relation avec le fichier d'autorité qui leur attribue un numéro d'identification. On intègre, autant que possible, les dates et lieux de naissance et de mort et une activité. Si on a un peu plus temps, on peut également saisir les relations familiales entre différentes personnes. Jusque là, on n'a rien saisi que de très schématique, mais cela suffit pour prendre conscience du nombre considérable de personnes qui ne figurent pas dans les fichiers d'autorité. En Allemagne, la Gemeinsame Normdatei,¹³ qui attribue ces identifiants uniques, est un instrument bibliothécaire et informe donc en priorité des auteurs d'ouvrages, au détriment d'autres catégories de personnes. Même si la plupart des savants sont auteurs et ont donc un tel identifiant, il arrive bien souvent que leurs correspondants, leurs femmes, leurs éditeurs, n'en aient pas. En d'autres termes, les fichiers d'autorité perpétuent le canon. Inversement, il arrive que plusieurs identifiants soient attribués à une seule et même personne, dans le cas par exemple où deux bibliothèques différentes ont fait des saisies concurrentes sur la base de données éventuellement légèrement différentes (la précision dans les noms, les dates, les lieux, peut varier). Enfin, il existe des généalogies de savants dans lesquelles les fils portent le même nom que leurs pères, ce qui peut entraîner une certaine confusion. Absence d'informations d'un côté, doublons de l'autre : la question de l'enrichissement des fichiers d'autorité est une question essentielle. Il serait souhaitable qu'un pourcentage imposé de tout projet travaillant sur des données biographiques soit attribué à l'enrichissement des fichiers d'autorité.

¹² COST Action IS1310.

¹³ Cf. http://www.dnb.de/EN/Standardisierung/GND/gnd_node.html pour l'Allemagne. En France et au niveau mondial, le Virtual Authority File (<http://viaf.org/>).



- 2) Les œuvres : Là aussi, l'infrastructure bibliothécaire existe, mais elle n'est pas unifiée. Lorsqu'un savant évoque une œuvre, cela peut être à un niveau très variable de matérialité de ladite œuvre. Il peut s'agir de l'œuvre en tant que telle, d'une édition particulière, voire même de l'exemplaire posé sur son bureau au moment où il rédige sa lettre. Or, il n'est pas rare que le choix du niveau auquel on renseigne un ouvrage (l'œuvre en général, l'édition précise, l'exemplaire singulier) soit essentielle pour l'analyse du corpus qu'il sera possible de réaliser, mais l'existence de ces différents niveaux complexifie évidemment d'une part l'architecture des données de chaque édition, d'autre part l'interopérabilité entre ressources numériques.
- 3) Les idées : l'annotation sémantique est, dans un contexte comme celui-ci, condamnée à rester de l'ordre de la science-fiction pour quelques temps encore. Tout d'abord, c'est quelque chose qu'il est quasiment impossible d'automatiser sur des corpus datant de périodes où l'orthographe n'est pas normalisée, a fortiori dans des échanges comme les correspondances savantes, dans lesquelles les changements de langue se produisent fréquemment (renvoi au terme technique latin, à la citation française, etc., ne nécessitant pas de traduction pour être compris par le correspondant). De plus, la définition d'une structure d'annotation sémantique (en constituant un thesaurus¹⁴ ou en prédéfinissant des catégories éventuellement structurées en arbre) est également un processus qui est loin d'être rôdé et standardisé, et dans lequel intervient inévitablement la subjectivité des personnes qui réalisent l'annotation.

10

En d'autres termes : les visualisations de réseaux ne peuvent être rendues interopérables que si l'architecture des données l'est, ce qui suppose que les points et les arêtes puissent être annotés de manière interopérable. Pour y parvenir, il faut notamment pouvoir faire interagir les métadonnées de la recherche et celles des archives/bibliothèques, mais aussi les données de la recherche entre elles. Les avancées scientifiques que l'on peut attendre de telles visualisations vont au-delà de la simple illustration. Il s'agit, dans ce cas, de faire émerger des structures de transferts scientifique et culturel allant au-delà du cas d'école, et de matérialiser, par exemple, des phénomènes souvent difficiles à embrasser dans leur ensemble comme l'intertextualité ou la diffusion. Pour y parvenir, il est nécessaire d'avoir non seulement beaucoup de données, mais surtout et principalement des données bien structurées : not necessarily big data, but a lot of good data. Des données vérifiées, riches, interopérables et, de préférence, ouvertes.

L'expérience de collaboration avec nodegoat illustre bien l'intérêt de ce type de structure de données. Lors de la conférence « Von Daten zu Erkenntnissen »¹⁵, des chercheurs ont présenté le module *correspDesc*¹⁶, destiné à encoder de manière cohérente les données essentielles à la structure d'échanges épistolaires dans un header TEI, puis la plateforme

¹⁴ Comme le fait par exemple le projet « Vernetzte Korrespondenzen » (<http://exilnetz33.de/de/>).

¹⁵ Voir la page web de la conférence : <http://dhd2015.uni-graz.at/>.

¹⁶ Cf. <https://github.com/TEI-Correspondence-SIG/correspDesc>.



correspSearch¹⁷, qui permet de moissonner les métadonnées de corpus épistolaires. En quelques minutes, les développeurs de nodegoat ont pu générer on the fly la visualisation suivante (Fig. 5) à partir des données actuellement disponibles dans correspSearch. C'est à mon avis un excellent exemple de ce que, à partir du moment où les données sont bien structurées et accessibles, les premiers obstacles techniques sont aisés à surmonter, ce qui permet de passer très rapidement au raffinement des questions de recherche.

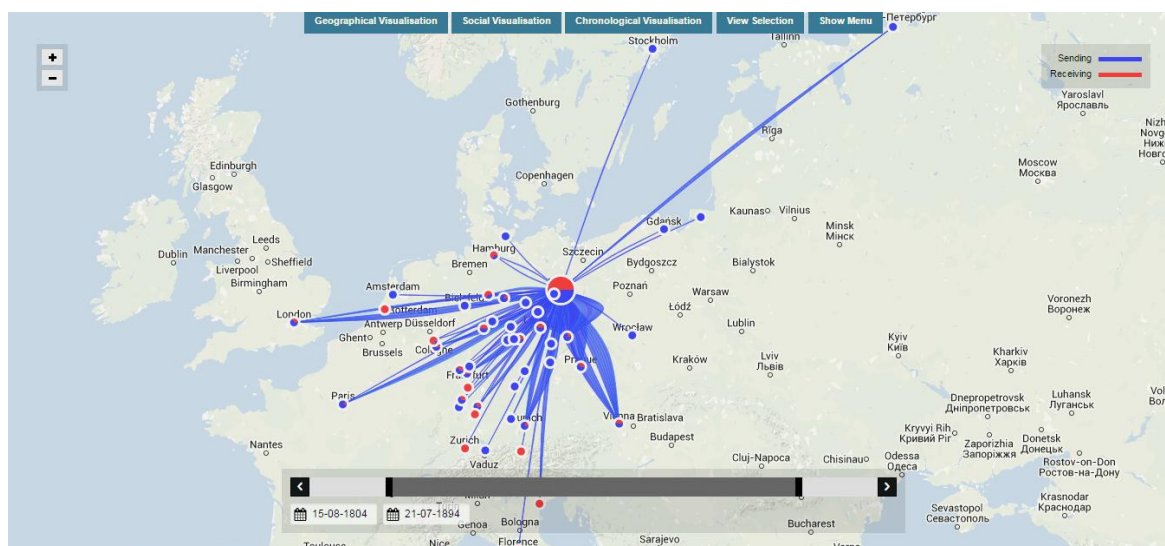


Fig. 5 : Réseau de correspondances généré à partir des données de correspSearch au 26 février 2015. Source : <http://correspsearch-test.nodegoat.net/viewer.p/4/136/scenario/1/geo/fullscreen>

Cela me ramène à ma question de départ, à savoir : quelle est la fonction réelle d'une visualisation de réseau ? Est-ce de rendre justice à une réalité issue des données et permettant de découvrir de nouvelles structures, ou d'illustrer de façon plus élégante qu'exacte une analyse de toute façon contenue dans un texte ou une matrice ? Dans les échanges avec Martin Grandjean et Marten Düring lors des twitterinterviews données dans le cadre du séminaire « réseaux historiques »¹⁸, deux aspects ont émergé que je n'ai pas encore évoqués jusqu'ici et qui peuvent peut-être tenir lieu de conclusion. Tout d'abord, la notion de jeu. Les chercheurs travaillant avec des visualisations de réseaux insistent sur la dimension ludique de la réalisation de la visualisation, même lorsqu'ils disent qu'ils mettent 99% de leurs visualisations à la poubelle. C'est que le 1% restant, celui qui va être utilisé pour mettre en valeur leurs résultats de recherche, a une qualité unique, irremplaçable, et d'une valeur inestimable : le 1% des visualisations de réseaux qui sert vraiment à quelque chose est *beau*. Cette qualité esthétique va au-delà du simple ornement, elle est la condition d'une autre forme d'appréhension par l'esprit, qui contribue à produire des savoirs nouveaux.

¹⁷ Cf. <http://correspsearch.bbaw.de/>.

¹⁸ Cf. <http://digigw.hypotheses.org/1153> et <http://digigw.hypotheses.org/1190>.