



HAL
open science

Introduction à la visualisation de données : l'analyse de réseau en histoire

Martin Grandjean

► **To cite this version:**

Martin Grandjean. Introduction à la visualisation de données : l'analyse de réseau en histoire. *Geschichte und Informatik*, 2015, 18/19, pp.109-128. halshs-01525543

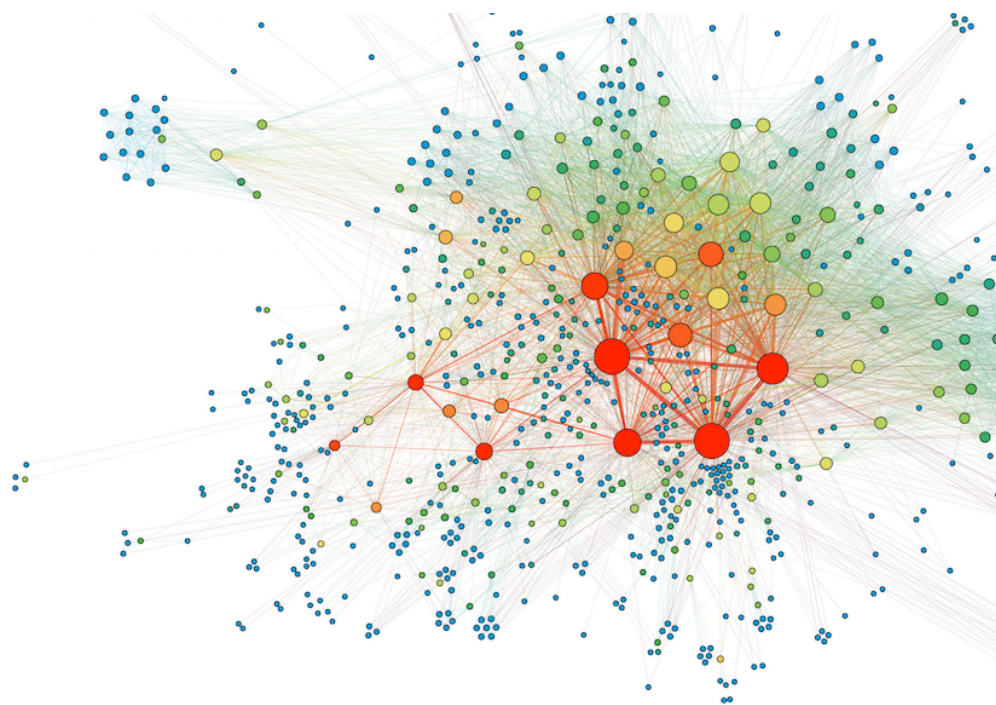
HAL Id: halshs-01525543

<https://shs.hal.science/halshs-01525543>

Submitted on 21 May 2017

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.



Introduction à la visualisation de données : l'analyse de réseau en histoire

MARTIN GRANDJEAN, UNIVERSITÉ DE LAUSANNE

**GRANDJEAN Martin (2015), « Introduction à la visualisation de données :
l'analyse de réseau en histoire », *Geschichte und Informatik*, 18/19, 109-128**

CHRONOS Verlag, Zürich

ISSN 1420-5955 / ISBN 978-3-0340-1289-8

L'utilisation de la visualisation de données en histoire engendre des réactions contradictoires : alors que certains sont fascinés par son potentiel heuristique à en oublier leur sens critique, d'autres rejettent par principe ces pratiques, les suspectant de cacher un vide explicatif.

Cet article introduit la distinction entre *visualisation de démonstration* et *visualisation de recherche*, rappelant que le chercheur ne doit pas uniquement utiliser la visualisation comme moyen de communication, mais également pour nourrir la recherche elle-même. C'est particulièrement dans sa forme la plus complexe, l'analyse de réseau, que cette catégorie de visualisations de recherche va être discutée.

The use of data visualization in history leads to contradictory reactions: some are fascinated by its heuristic potential and forget their critical faculties while others reject this practice, suspecting it of hiding explanatory weaknesses.

This paper proposes a distinction between *demonstration visualization* and *research visualization*, reminding that scholars should not only use data visualization for communication purposes, but also for the research itself. It is particularly in its more complex form that this research visualization category will be approached here: network analysis.

Avant-propos: fascination et préjugés

Inclure une visualisation de données dans une présentation académique fait courir au conférencier deux risques antagonistes liés à l'hétérogénéité de son assistance. D'une part, le risque de perdre la moitié du public qui, fasciné par les images, quitte la linéarité des propos du discourant et évacue tout esprit critique sur la source et l'apport épistémologique du matériel présenté. D'autre part, le risque que l'autre moitié du public rejette par principe cette manière de présenter un processus ou un résultat sur le simple soupçon que la démarche visuelle attrayante ne cache une faille méthodologique ou un vide explicatif.

Il va s'agir ici de répondre à la fascination comme aux craintes en pesant les potentialités comme les risques de la visualisation de données en histoire, en particulier dans sa forme la plus complexe et plus aboutie, l'analyse de réseaux. Le réseau, entre tentation de la beauté pour elle-même et difficile lisibilité, est un outil de recherche plus qu'une fin et c'est à ce titre que l'historien doit apprendre à l'appivoiser pour en tirer profit en évitant les principaux écueils. En ce qu'elle pose également les bases d'une typologie et qu'elle liste des erreurs courantes, cette démarche a un caractère introductif. Par respect pour les auteurs et pour conserver cette dimension didactique, il ne s'agit donc pas d'y construire une bibliographie exhaustive d'exemples et contre-exemples.

La visualisation de données en histoire aujourd'hui

Si la réputation de la visualisation de données est si teintée de cette ambiguïté entre fascination et préjugés, c'est probablement parce qu'elle est desservie par l'incompréhension profonde d'un public traditionnellement peu habitué à ce type de vecteur d'information. En outre, de nombreux exemples si ce n'est désastreux du moins fautifs parsèment la littérature scientifique et lui portent préjudice. Avant de s'attacher à combler le manque de compréhension qui entoure la visualisation (et son utilité pour les sciences historiques) et surtout de détailler certains des abus les plus fréquents, un retour aux fondamentaux s'impose.

Bien que le terme de visualisation renvoie aujourd'hui à une mise en image du traitement d'informations, la tendance à illustrer une argumentation par un objet visuel est très probablement aussi ancienne que l'écriture elle-même. Il n'est pas ici question de refaire une énième chronologie des traditions narratives qui ont abouti aux objets multiformes et multimédias qui nous servent aujourd'hui à dérouler nos propos, mais plutôt de rappeler que les enluminures, gravures, peintures et autres photographies, bien qu'étant comme on va le voir plus loin d'un autre ordre que les « visualisations de données », n'en sont pas moins des moyens de

soutenir l'expression d'un raisonnement. Même si elle se développe en particulier dès l'avènement des mathématiques modernes, la visualisation de données n'est en effet pas une pratique récente : des cartographies célestes aux traités d'anatomie en passant par les arbres généalogiques de monarques, les exemples ne manquent pas.

Emblématique depuis sa popularisation par Edward Tufte¹, la « Carte figurative des pertes successives en hommes de l'armée française dans la campagne de Russie 1812–1813 » de Charles Joseph Minard, datée de 1869, est fréquemment citée² comme la première visualisation de données en raison de la très grande lisibilité dont elle fait preuve malgré son importante densité d'informations. Sur un fond de carte géographique représentant la distance entre Kaunas et Moscou se déroule le trajet aller-retour des armées napoléoniennes. L'épaisseur de la ligne varie en fonction de l'effectif diminuant des soldats français alors que le tragique trajet du retour est augmenté de repères indiquant les températures endurées par la Grande Armée. Cette image tire sa puissante valeur narrative du fait qu'on y lit en un seul coup d'œil plusieurs mois de campagne et qu'on en tire rapidement un enseignement fondamental pour la compréhension de cet épisode historique : les effectifs napoléoniens ont subi une diminution relativement constante due aux maladies, au froid, aux escarmouches avec les troupes russes, alors qu'aucune bataille majeure n'a été perdue. Sans volonté de déconstruire le modèle de Minard et encore moins de porter atteinte au succès de cette réalisation, on notera que cette image n'est pas une « visualisation de données » au sens strict, un graphique sur lequel des points sont reportés en respectant fidèlement les valeurs qui leur sont associées mais plutôt un dessin sur la base de données, d'ailleurs évidemment incomplètes (puisqu'il s'agit de relevés d'effectifs ponctuels et irréguliers).

Sur les traces de réalisations réussies et visuellement efficaces comme celles de Minard et portée par l'essor des études statistiques et quantitatives, la pratique de la visualisation de données connaît un développement tout à fait remarquable dans le courant du XX^e siècle. Cette généralisation du graphique comme moyen d'explicitier un corpus de données s'accompagne, en particulier dans les disciplines peu familières au traitement d'informations chiffrées, de nombreux contre-exemples, souvent involontaires, qu'il s'agit ici de mettre à l'index pour mieux en éviter la répétition³.

- × **LE TABLEAU DE CHIFFRES ABSOLUS.** Par négligence ou par manque de données à comparer, il n'est pas rare d'exprimer un phénomène en chiffres absolus sans discours critique. La FIGURE 2 montre une augmentation nette du nombre de malades : il est toutefois extrêmement réducteur d'en conclure que la maladie est en progression

sur vingt ans (le nombre de malades a doublé) sans superposer à ces données la population totale, qui elle a été multipliée par quatre pendant le même intervalle.

- × **LA CARTE DE POPULATION NON PONDÉRÉE.** Cartographier les possesseurs de téléphones portables sur un territoire donné n'a aucun intérêt si les données ne sont pas pondérées en fonction de la population des subdivisions concernées. Sans pondération, et à moins que la répartition de la population-cible ne soit très irrégulière, le résultat ne sera qu'une carte de population globale (à ce titre, la FIGURE 3 peut aussi bien représenter les possesseurs de téléphones portables que les cas de grippe ou la consommation totale d'hydrocarbures). Cette erreur est extrêmement fréquente et témoigne de la fascination visuelle exercée par la cartographie de manière générale. Elle montre aussi à quel point ce genre de visualisation n'a pas fait l'objet d'un regard critique de la part de son concepteur.
- × **LE NUAGE DE MOTS-CLÉS INCOMPRIS.** Les conférences et couvertures d'ouvrages agrémentées de ces fameux *wordclouds* sont légion. Mais cette visualisation, pour autant que le corpus textuel ait été défini rigoureusement et les choix de dictionnaires (langue, mots à exclure, détection des expressions) effectués avec soin, apporte-t-elle une réelle valeur ajoutée vis-à-vis de la liste sèche du nombre d'occurrences de chaque notion (FIGURE 4)? Sans compter les visualisations obtenues à l'aide d'outils en ligne à de simples fins illustratives.
- × **LE RECOURS AVEUGLE À GOOGLE NGRAM⁴.** S'il est une mode dont les historiens ne semblent pas prêts de s'affranchir, c'est bien celle de céder à la griserie du prétendu *big data* de la bibliothèque numérisée du géant Google et de son outil de compilation d'occurrences. Non seulement l'opacité de la base de données (il est impossible de savoir quels sont les ouvrages pris en compte) de *Google Books* devrait décourager le chercheur de l'utiliser dans le cadre scientifique, mais celui-ci ne prend souvent même pas la peine d'en comprendre les fonctionnalités. Ainsi, il arrive fréquemment que soient comparés des termes incomparables: la FIGURE 5, directement inspirée d'une communication scientifique, compare « bible » à « Shakespeare » en oubliant la majuscule du premier terme, alors que le moteur de recherche est sensible à la casse. Une erreur aux conséquences lourdes puisque le terme « Bible » (avec sa majuscule) est massivement plus utilisé que son équivalent en bas de casse⁵.

On retiendra, à propos de la visualisation de données en histoire, qu'elle n'est efficace que si elle est accompagnée d'un discours critique qui en explicite les sources (et la méthode dans le cas de visualisations complexes).

Carte Figurative des pertes successives en hommes de l'Armée Française

Dressée par M. Minard, Inspecteur Général des Ponts et C

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millier des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. _____ Les n

dans les ouvrages de M.M. Chiers, de Ségur, de Fezensac, de Chambray et le journal inédit de

Pour mieux faire juger à l'oeil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et et Mohilow et ont rejoint vers Orscha et Witebsk, avaient toujours

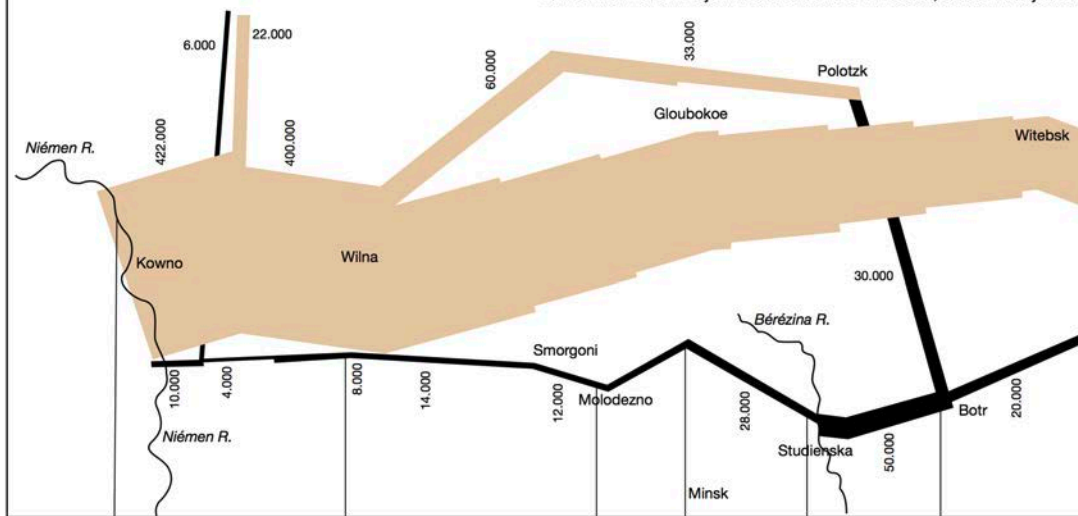
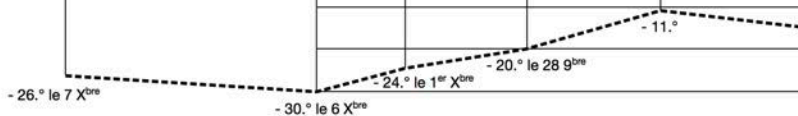


TABLEAU GRAPHIQUE de la température en degrés du thermomètre

Les Cosaques passent au galop le Niémen gelé.



Autog. par Regnier, 8. Pas. S^{re} Marie S^{re} G^{re} à Paris.

[Vectorization CC-BY-SA m

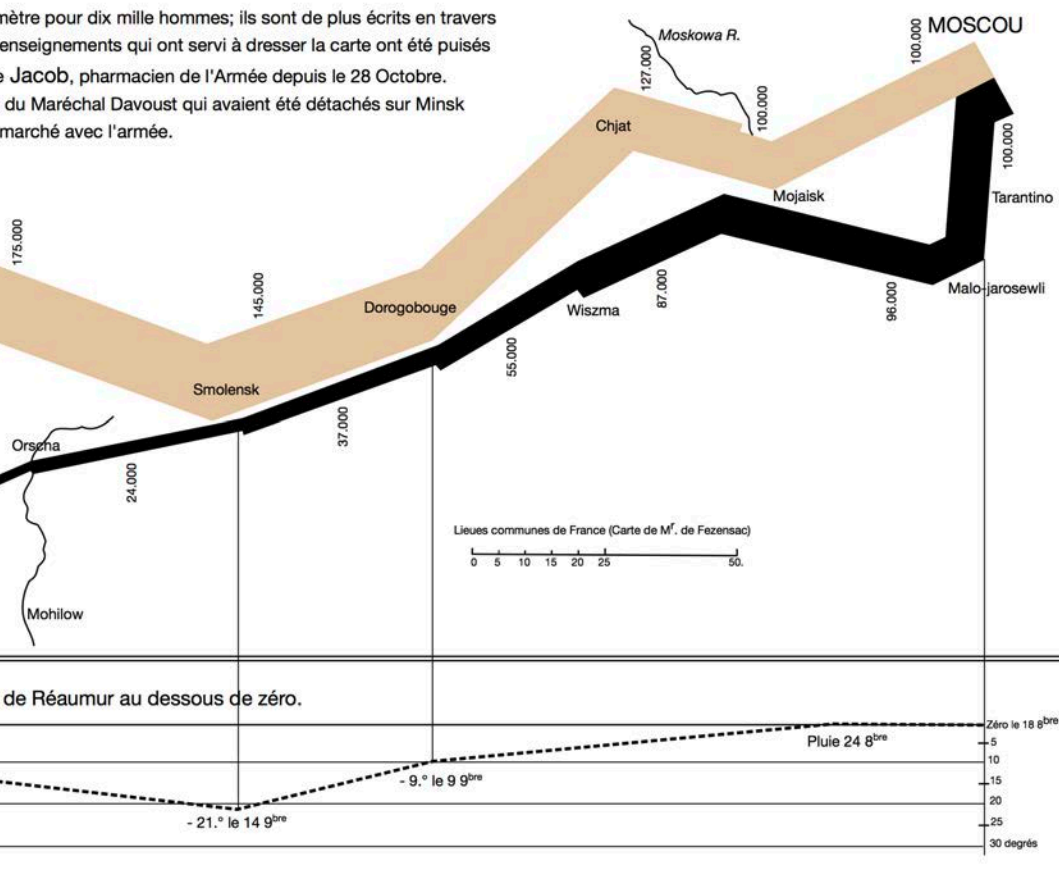
... dans la campagne de Russie 1812-1813.

Chaussées en retraite. Paris, le 20 Novembre 1869.

... mètre pour dix mille hommes; ils sont de plus écrits en travers
enseignements qui ont servi à dresser la carte ont été puisés

... Jacob, pharmacien de l'Armée depuis le 28 Octobre.

... du Maréchal Davoust qui avaient été détachés sur Minsk
marché avec l'armée.



de Réaumur au dessous de zéro.

[Martingrandjean.ch 2014]

Imp. Lith. Regnier et Dourdet.

FIGURE 1
« Carte figurative des pertes successives en hommes de
l'armée française dans la campagne de Russie 1812-1813 »,
Charles Joseph Minard, 1869.⁶

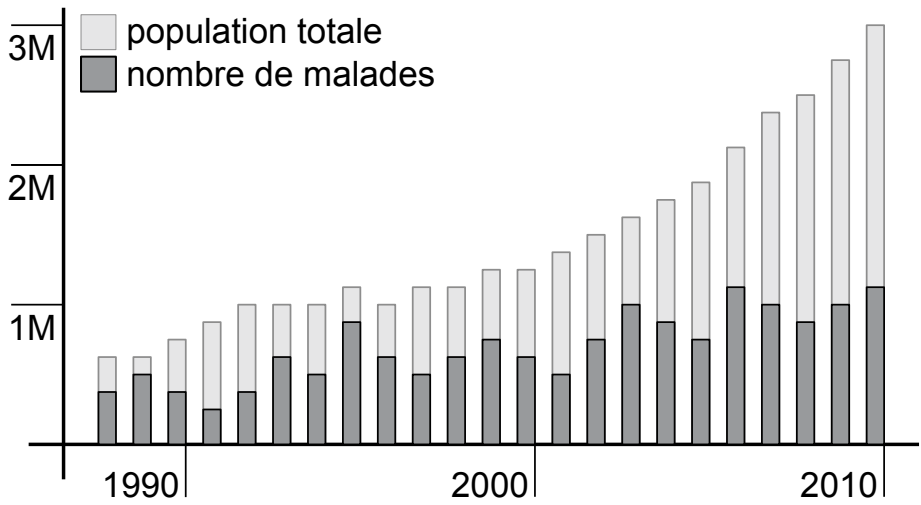


FIGURE 2 LE TABLEAU DE CHIFFRES ABSOLUS

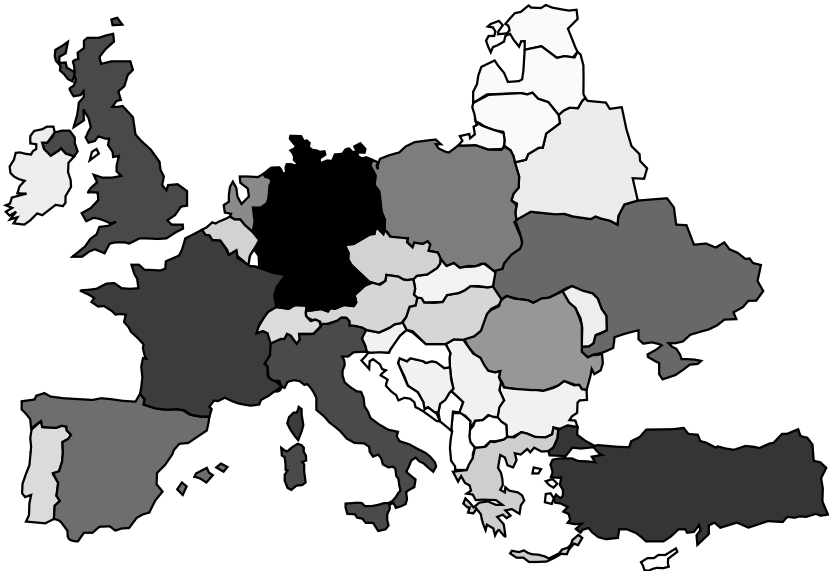


FIGURE 3 LA CARTE DE POPULATION NON PONDÉRÉE

Ce de mots
Nuage
 est attrayant mais est-il
 très vraiment **utile?**

FIGURE 4 LE NUAGE DE MOTS-CLÉS INCOMPRIS

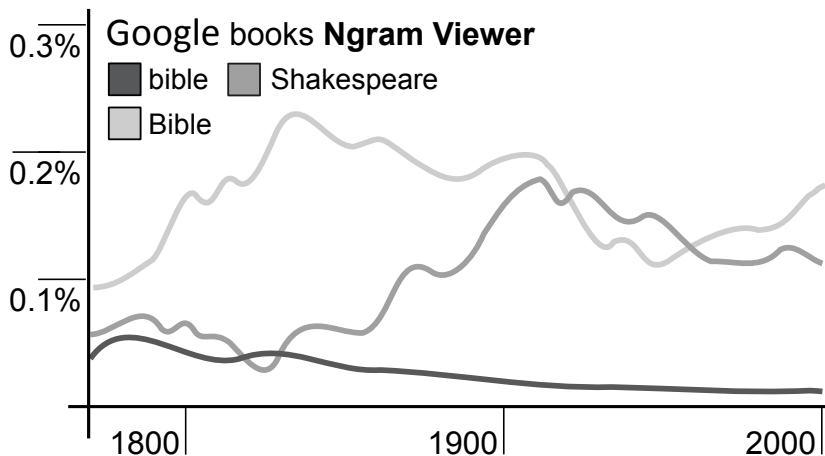


FIGURE 5 LE RECOURS AVEUGLE À GOOGLE NGRAM

Sans ce discours qui révèle ce qu'il y a derrière le produit fini visuel, elle perd son intérêt, voire son sens.

Quelle visualisation? Esquisse d'une typologie

Toutes les visualisations ne se valent pas, toutes ne se prêtent pas à la mise en image de tous les types de données. Au-delà d'aspects esthétiques subjectifs (qui ne sont évidemment pas à négliger puisqu'un visuel bâclé peut être rédhibitoire), le chercheur en histoire doit disposer d'une typologie qui guide et facilite ses expérimentations visuelles :

- × **L'USAGE, ENTRE DÉMONSTRATION ET RECHERCHE.** Il faut distinguer les visualisations qui découlent d'un savoir scientifique et les visualisations qui créent un savoir scientifique. Cette bipartition entre visualisation de « démonstration » et visualisation de « recherche » est tout à fait fondamentale et permet de se mettre au clair avec le but explicite de la démarche visuelle. Bien qu'il n'y ait pas de gradation de valeur entre ces deux types, il apparaît évident que, bien qu'on puisse avoir recours avec succès à la visualisation de démonstration à des fins pédagogiques (rendre un savoir compréhensible à une audience), c'est son pendant de recherche qui va attirer notre attention ici. Contrairement à l'idée reçue et comme on va avoir l'occasion de s'en rendre compte plus bas dans le cadre de l'analyse de réseau, la visualisation est parfois un moyen de faire surgir de nouvelles questions de recherche d'un jeu de données très complexes et à première vue inintelligibles. Ces visualisations ne sont donc pas toujours destinées à être publiées ou présentées mais simplement à soutenir le chercheur dans sa réflexion, lui ouvrant des perspectives inédites.
- × **LA SOURCE, ENTRE INFORMATION ET DONNÉES.** La nature des sources qui composent la visualisation est un critère qui détermine la nature du produit fini lui-même. Un organigramme de comité de direction renseigne son lecteur en lui fournissant des informations alors qu'un histogramme renseigne son lecteur en lui mettant à disposition l'expression de données. Triviale dans l'exemple organigramme/histogramme, cette typologie s'avère importante dans le contexte du partage de connaissances puisque fleurissent sur Internet et dans les médias les « infographies », ces images qui tiennent souvent plus de l'art graphique que de l'expression de données. L'infographie comparant trois modèles de voitures, quand bien même elle inclut des données chiffrées permettant de comparer les véhicules, n'est pas une visualisation de données mais un simple vecteur informatif.
- × **LA COMPLEXITÉ, ENTRE SIMPLICITÉ ET ILLISIBILITÉ.** Bien qu'elle différencie les objets visuels selon des critères différents, cet élément

de typologie rejoint parfois la question de l'usage en ce que les visualisations les plus compliquées sont généralement impropres à la « démonstration ». C'est le cas d'une partie des graphes de réseaux qui, bien qu'ils dégagent une apparence globale de cohérence, demandent une analyse minutieuse dont la complexité dépasse le lecteur non averti. Toutefois, cette typologie n'est pas dénuée d'un certain sens pour catégoriser les visualisations de données puisqu'il arrive que des corpus de données très conséquents puissent se résumer en un graphique en courbes tout à fait lisible et démonstrativement efficace.

La typologie esquissée ici n'est pas exhaustive, elle se veut une base de réflexion pour le chercheur face à ses données.

Analyse de réseau: le cas des archives de la coopération intellectuelle de la SDN

Dès 1919, et pendant l'immédiat après-guerre, des scientifiques européens cherchent à réactiver les nécessaires réseaux de communication académique. Alors que le continent se remet petit à petit des séquelles du premier conflit mondial, la Société des Nations (SDN) offre à ces chercheurs un espace de dialogue: la Commission Internationale de Coopération Intellectuelle (CICI). Présidée par le philosophe Henri Bergson et réunissant des éminences grises aux carrières internationales et aux attaches disciplinaires variées⁷, la CICI tente de se poser en rassembleuse dans un milieu très éprouvé par cinq années d'absence de relations. Entre projets d'enquêtes, soutiens à des institutions scientifiques en difficultés, rassemblements académiques et publications, la CICI a tout pour devenir un acteur incontournable de la restructuration des réseaux intellectuels.

Envisager l'étude de cette commission, cette nouvelle venue dans la scène traditionnellement très structurée des échanges académiques (entre revues, lieux de congrès et universités), fait poindre la nécessité de comprendre l'enchevêtrement des relations qu'elle tisse avec la toile préexistante. Puisque c'est vers l'analyse de réseau que l'historien se tourne pour cette étude, il est nécessaire de distinguer deux approches:

- × **LE RÉSEAU RECONSTRUIT.** Sur la base des nombreuses monographies et études documentant l'entre-deux-guerres scientifique, ainsi qu'avec les informations contenues dans les publications internes de la SDN, il est possible de reconstituer l'organigramme des institutions actrices de la coopération intellectuelle: sociétés savantes avec leurs comités et leurs membres, syndicats des travailleurs intellectuels, universités et leurs commissions de recherche, faïtières internationales et leurs sections, instituts indépendants et initiatives privées. Cette

approche plutôt traditionnelle permet de reconstruire un tableau global des relations entre acteurs (institutions et personnes) sur la base d'informations⁸.

- × **LE RÉSEAU DES DONNÉES.** Plutôt que de tenter un panorama global et interprété de la situation européenne, il s'agit ici de laisser parler un corpus d'archives. Le fonds de la CICI contient des milliers de correspondances internes et externes qui offrent une perspective nouvelle sur les relations entretenues par ses membres avec les autres acteurs du champ académique de l'entre-deux-guerres. Ces données (auteur-destinataire) s'offrent à une exploitation systématique en « réseau ».

On notera que ces deux approches reprennent à dessein la bipartition informations/données évoquée plus haut. Plutôt que de les opposer, l'intérêt est évidemment de comparer le réseau global reconstitué avec le réseau de correspondances de la CICI. On s'attardera ici sur la seconde, à savoir l'approche archivistique⁹.

Concrètement, c'est en listant tous les documents contenus dans les sections du fonds SDN qui traitent de la CICI¹⁰ avec leurs métadonnées standardisées que l'on obtient une base de données que l'on peut mettre en relation avec un index des acteurs des documents en question (leurs auteurs et destinataires). Avant même la visualisation du réseau formé par ces documents et leurs acteurs, la base de données relationnelle est déjà un réseau utilisable: il est aisé d'y lister toutes les occurrences communes de deux personnes ou de voir par quel document se concrétise le lien entre deux entités. C'est toutefois dans son exploitation visuelle que cette démarche trouve sa plus-value.

Cette base de données relationnelle peut être visualisée sous la forme d'un graphe bipartite (FIGURE 6) où chaque document est relié à ses acteurs. On procède ensuite à une « projection » qui va simplifier le graphe en résumant les connexions d'acteurs à un document commun en une relation simple entre acteurs. Dans la FIGURE 6, le document 31596 implique Marie Curie, Hendrik Lorentz et Henri Bergson. Cette situation est résumée dans la FIGURE 7 par une arête entre chacune de ces trois personnes (l'arête entre Marie Curie et Henri Bergson est plus épaisse car ils apparaissent simultanément dans trois documents dans cet exemple).

Vient l'heure de la visualisation du réseau complet (FIGURE 8), où tous les acteurs sont spatialisés sous la forme d'une galaxie à première vue inintelligible. La répartition des sommets n'en est pas pour autant aléatoire: ce *force-directed layout*¹¹ calcule les positions de façon à éloigner

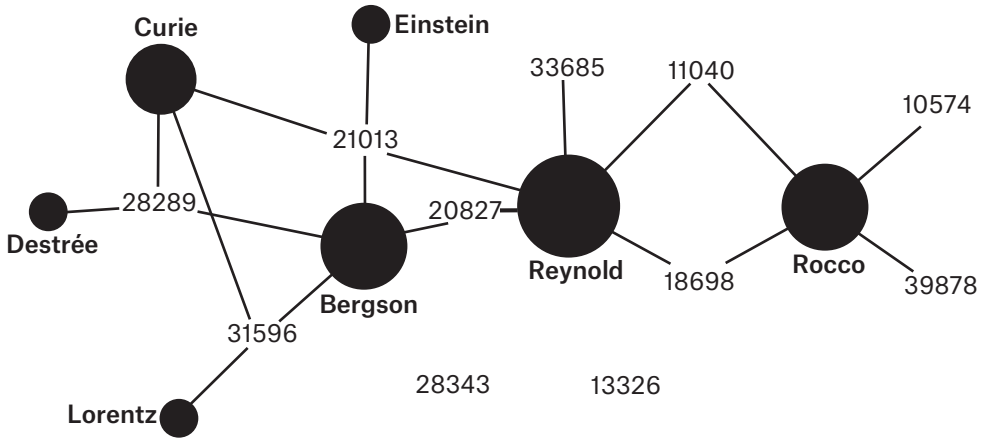


FIGURE 6

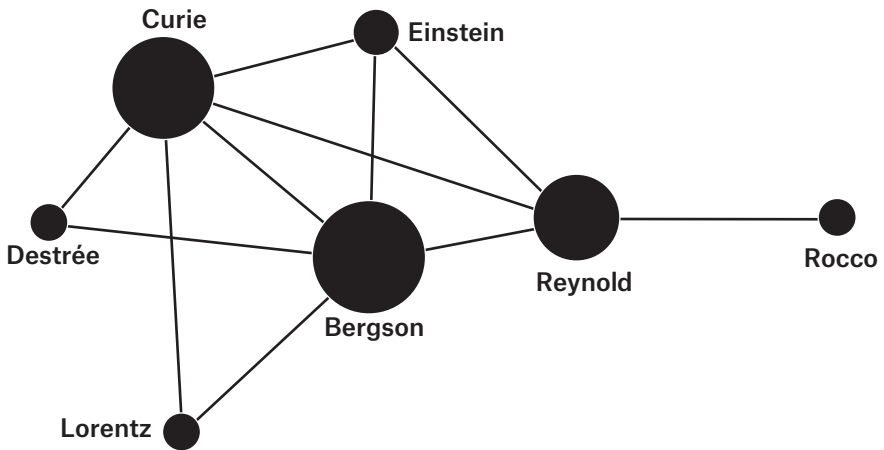
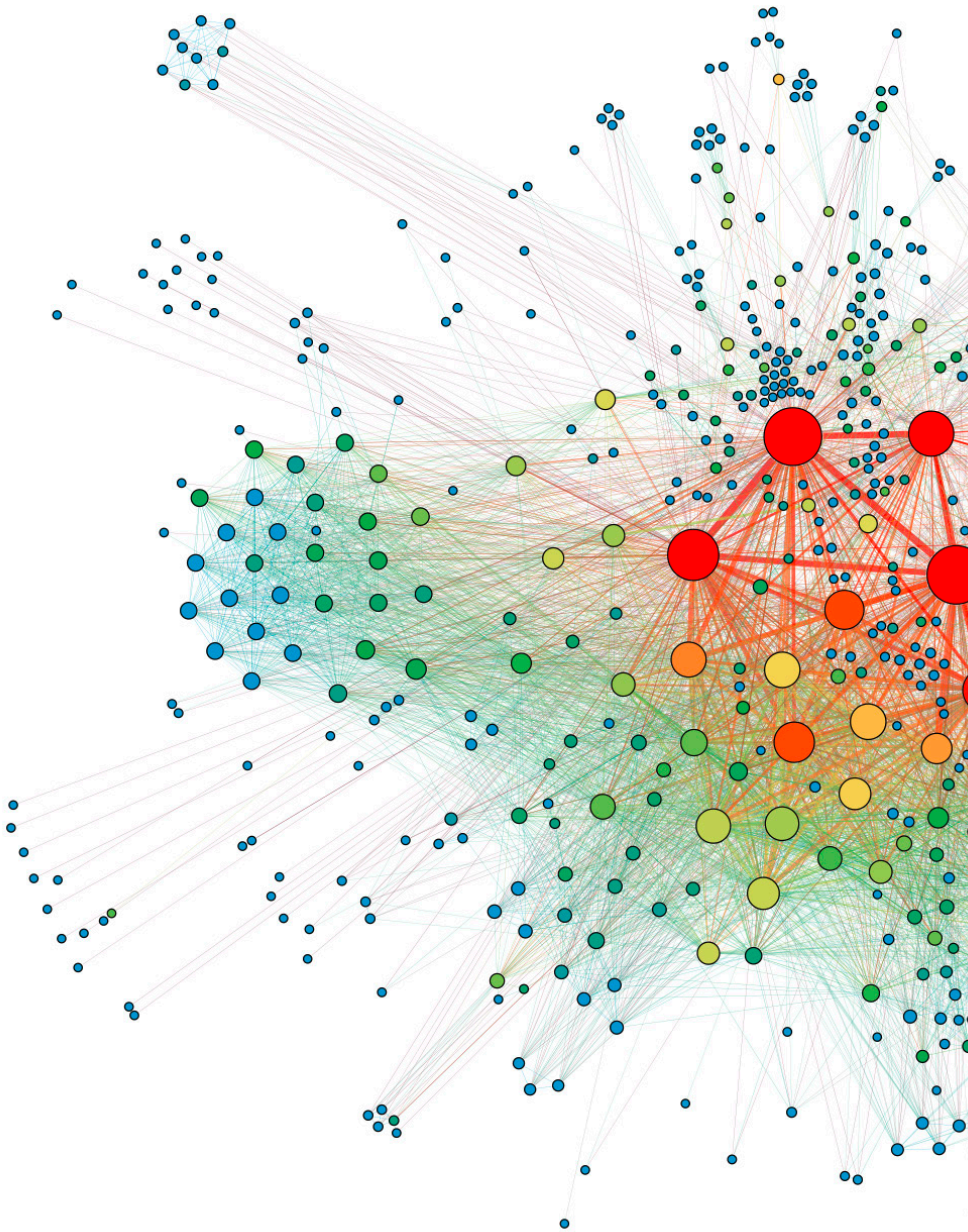


FIGURE 7

FIGURES 6 - 7
 Un exemple de réseau de documents¹², mis en lien avec les personnes qui y figurent (FIG. 6, en haut), puis « projeté » vers un réseau (FIG. 7, en bas) où seules les personnes apparaissent, connectées aux autres acteurs des documents communs.



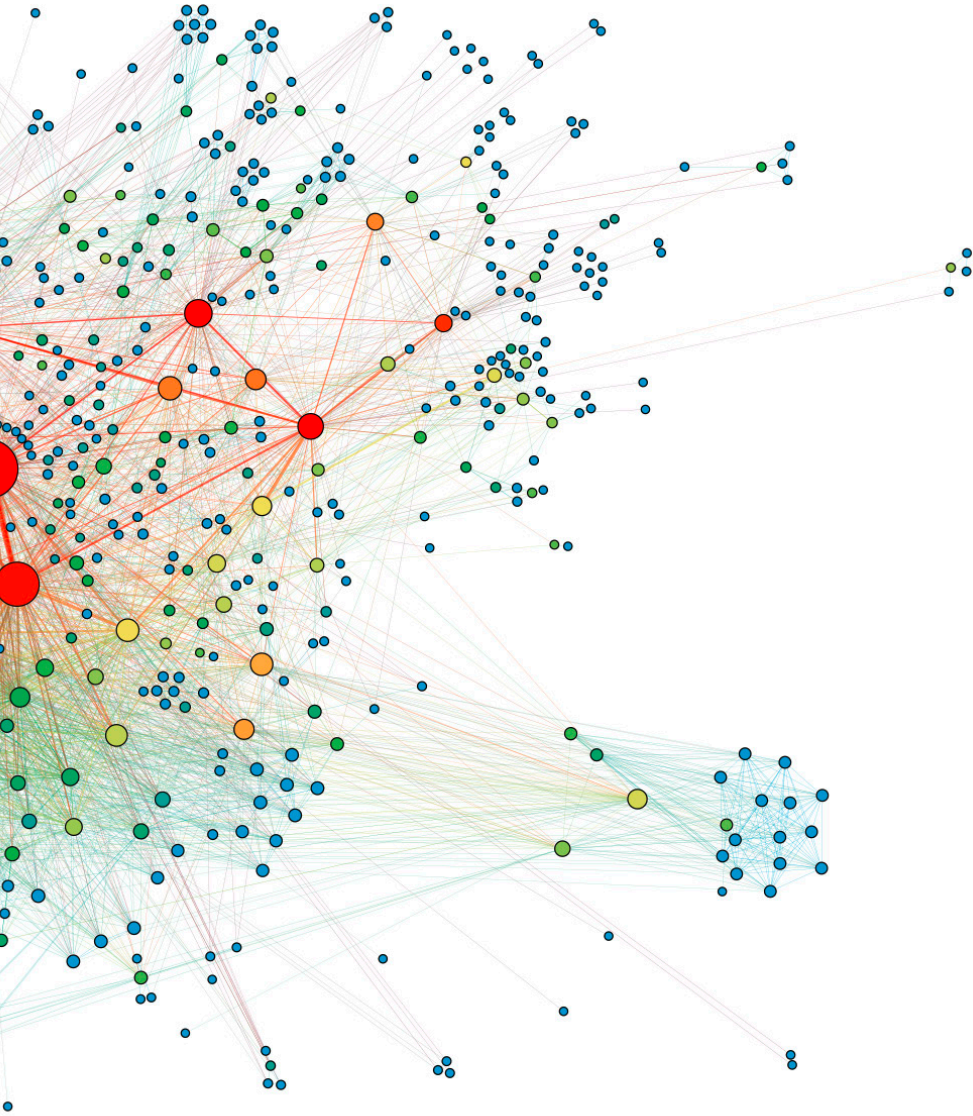


FIGURE 8

Le graphe du réseau obtenu après dépouillement (encore partiel) de plusieurs milliers de documents: plus de 800 personnes reliées par près de 6000 arêtes (représentant plus de 10000 relations, les arêtes s'épaississent proportionnellement au nombre d'apparitions simultanées de personnes comme acteurs d'un même document). La taille des cercles est fonction du degré de centralité des

personnes (le nombre de connexions qu'elles entretiennent) alors que la couleur indique leur centralité d'intermédiation (mesure, sur tous les chemins possibles dans le réseau, la proportion de chemins qui passent par cet acteur et renseigne ainsi sur sa potentielle fonction de « pont » dans le réseau en question).

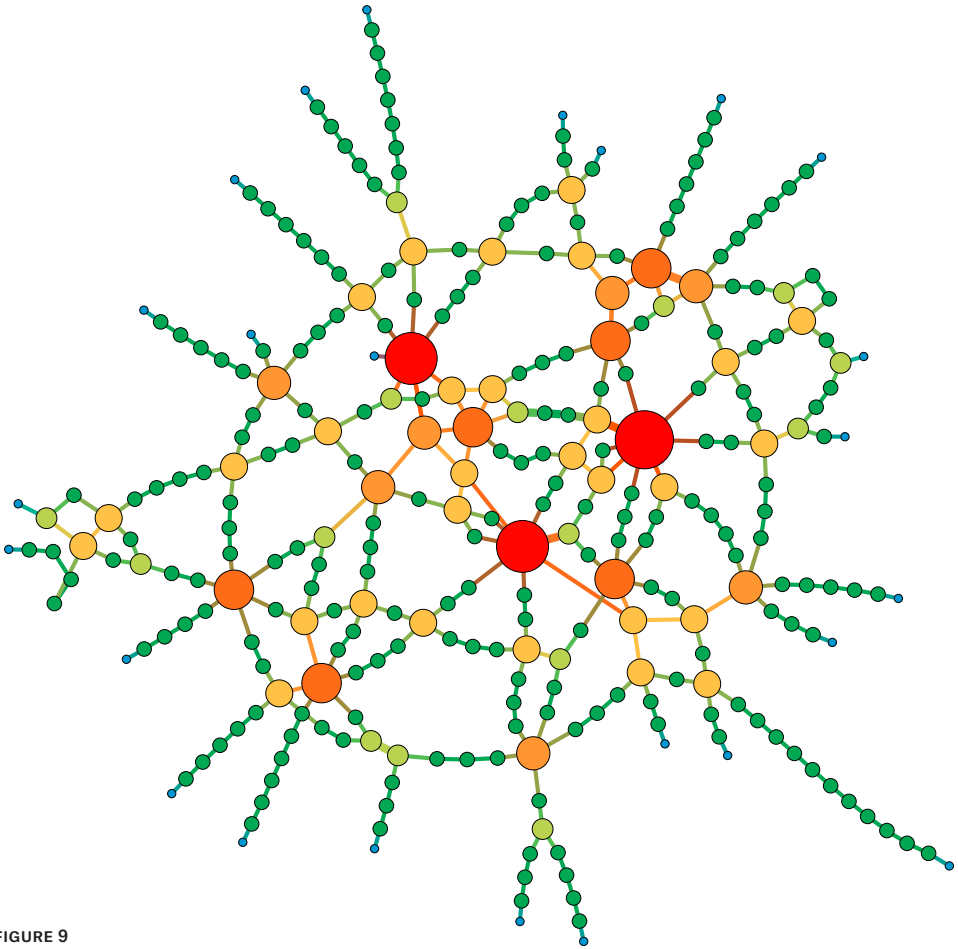


FIGURE 9
Un autre regard sur le réseau de métro parisien.

le plus possible les sommets les uns des autres, à l'exception des sommets reliés par une arête (auquel cas il tient compte du poids de celle-ci). Les positions sont donc calculées et recalculées afin de trouver, comme lorsqu'on déposerait un grand nombre d'aimants de même charge sur une table, une position d'équilibre.

Cette visualisation est tout sauf un produit fini. On imagine d'ailleurs mal comment s'en contenter puisque même si des communautés semblent se dégager du centre complexe de cette «galaxie», aucune conclusion ne peut en être tirée dès le premier coup d'œil. C'est l'exemple parfait de la visualisation de recherche évoquée plus haut: une visualisation destinée à attirer l'attention du chercheur sur une irrégularité, un phénomène global ou une relation étonnante. D'un point de vue documentaire, ce

type de visualisation peut évidemment être utilisé pour mettre en image le réseau de correspondance d'un individu¹³ (en le mettant en valeur par une couleur particulière, lui et les individus avec lesquels il est connecté, par exemple) ou d'un groupe d'individus, mais son réel intérêt réside dans les outils mathématiques qui s'offrent désormais à l'historien pour faire parler ces données dans un nouveau registre épistémologique.

Cette visualisation et les perspectives d'analyse de réseau qu'elle porte en puissance doit toutefois être elle aussi soumise à la critique du chercheur. Le principal malentendu véhiculé par ce type de visualisation est qu'elle représenterait une forme de « vision complète » d'un objet historique, alors qu'il ne s'agit que d'un artefact graphique visualisant une base de données qui a ses limites implicites. Le réseau donne en effet l'impression d'un objet fini, total. Il n'en est rien puisqu'on ne visualise que les données qui sont à disposition, nous heurtant à l'incomplétude des archives et au fait que les relations entre les scientifiques de l'entre-deux-guerres ne se résumaient pas à des échanges épistolaires. L'occasion une fois de plus de rappeler que l'analyse de réseau n'est qu'un outil. Un outil d'un très grand intérêt puisqu'il offre à l'historien une vision distante de son objet, pour mieux s'y plonger en détail ensuite.

Le réseau, au-delà de l'aspect visuel

L'analyse de réseau n'a pas besoin de son illustration visuelle pour être effective. Sa visualisation ne fait que contribuer à réduire son niveau d'abstraction, en particulier lors de son utilisation dans le cadre de recherches en histoire. Ses particularités en font un objet mathématique très bien balisé dans le domaine de la théorie des graphes¹⁴.

Lors de la création de nouvelles lignes de transports publics, c'est entre autres à cette théorie que l'on fait appel pour comprendre les flux de passagers et les tensions d'horaires. La FIGURE 9 représente le réseau de métro de Paris dans une configuration relativement différente du plan que l'utilisateur quotidien à l'habitude de trouver sur le panneau d'affichage de l'entrée la plus proche : les arrêts ne sont pas répartis en fonction de leurs coordonnées géographiques mais simplement ordonnés spatialement selon un algorithme *force-directed*. La taille des marqueurs des stations n'est pas fonction de l'importance des monuments qu'elles desservent, ni du nombre d'utilisateurs recensés lors du dernier relevé de la RATP, mais du nombre de lignes qui les traversent. C'est dans ce type d'application que les indices de centralité¹⁵ vont permettre de vérifier ou d'infirmer le calcul intuitif de l'utilisateur régulier en recherche du plus court trajet entre son domicile et son lieu de travail : ici, la distance géographique compte moins que le nombre d'arrêts.

On dénombre plusieurs mesures de centralité dont il s'agit ici d'évoquer quelques-unes des plus probantes dans le cadre d'études de réseau de transport public comme dans le cadre de l'analyse archivistique menée plus haut. La plus intuitive est la *centralité de degré*, déjà évoquée, qui classe les nœuds du réseau en fonction de leur nombre de connexions¹⁶. Mais avoir un nombre élevé de relations ne fait pas pour autant de l'élément en question un élément central du réseau, ce que la *centralité de proximité* mesure en évaluant la distance qui sépare chaque nœuds du réseau un à un, pour mettre en valeur ceux qui ont la distance moyenne la plus courte. Dans le cadre d'une étude sur un réseau épistolaire (le réseau des archives de la CICI s'y prête très bien), c'est une mesure de *centralité d'intermédiation* qui va indiquer au chercheur quelles sont les personnes qui font office de « ponts » entre les communautés en présence. Ces individus, collectionnant les « liens faibles »¹⁷, ne sont pas nécessairement les personnalités les plus connectées du réseau. C'est leur position à l'intérieur de celui-ci qui leur confère cette importance. Cette mesure est d'autant plus intéressante que ces individus ne sont pas toujours des personnalités de premier plan ; il est donc difficile de les détecter par un autre biais. On y découvre par exemple, dans les marges d'un réseau dense d'acteurs très médiatisés, des scientifiques dont le nom n'aura pas été retenu pour leurs travaux mais qui jouent efficacement ce rôle d'intermédiaires en occupant plusieurs positions administratives dans des sociétés savantes.

S'il est un réseau inextricable et impossible à visualiser dans toute sa complexité, c'est la bien nommée « toile » de l'Internet. Alors que les milliards de pages qui la composent contiennent elles-mêmes des liens vers d'autres pages, le moteur de recherche Google fait lui aussi usage de la théorie des graphes pour détecter les pages les plus dignes d'intérêt. C'est en effet à l'aide d'une mesure de *centralité de prestige*¹⁸, que Google attribue une note de 1 à 10 (échelle logarithmique), son fameux *pagerank*, aux sites web, qui impacte ensuite sur leur référencement. Cet algorithme attribue un score de « prestige » à chaque nœud du réseau (chaque page web dans cet exemple) qui va varier en fonction des scores des nœuds qui pointent sur lui. Une page web citée depuis des pages très bien référencées va donc voir son score augmenter, tout en distribuant ce « prestige » nouvellement acquis aux pages qu'elle cite elle-même (le score est donc constamment recalculé¹⁹).

Dans cette nouvelle relation qu'instaurent les outils mathématiques (et les outils statistiques déjà largement adoptés par la communauté des sciences historiques), il convient toutefois de faire preuve d'une certaine prudence vis-à-vis de la transformation de la source en une donnée chiffrée et interrogeable : ces aspects mathématiques, ces mesures de centralités

et leurs potentialités n'ont pas toujours à voir avec les comportements humains que décrivent les sources originales. Toutes ces démarches ont donc leur place dans la dimension exploratoire de la recherche et doivent passer le filtre critique du chercheur avant d'être érigées en conclusions automatisées. Ce n'est pas parce qu'un des acteurs du réseau de la CICI obtient un score d'intermédialité important qu'il est pour autant un personnage-clé des relations entre des sous-parties du réseau. Cette conclusion ne doit être tirée que par l'historien qui, alerté par ce résultat statistique, a replongé dans les correspondances de cet individu pour constater qualitativement sa position particulière.

Conclusion: organiser et visualiser l'information en réseau

Alors que le proverbe « une image vaut mieux que mille mots », attribué à Confucius, semble appuyé par les récents développements de la visualisation de données, on rappellera que sans discours critique cette visualisation n'a pas toute sa légitimité narrative ou explicative. Mais si l'image est effectivement parlante, pour illustrer un article scientifique comme pour guider un chercheur dans un questionnement, c'est probablement parce que la visualisation est un langage qui s'impose à son lecteur de manière quasi organique, viscérale et instinctive, en instaurant une nouvelle relation avec les données. D'ailleurs, l'organisation de la connaissance en réseau, que ce soit dans de grandes bases de données de recherche, dans des encyclopédies participatives en ligne comme Wikipédia ou concrétisée par les réseaux sociaux de chercheurs sur des nouveaux médias comme Twitter, est une réalité du monde académique d'aujourd'hui qu'on ne saurait amputer de sa dimension visuelle.

Prêchons donc pour que l'historien ne sous-traite pas les composantes techniques de sa recherche, mais qu'il se saisisse lui-même de la visualisation de données dans une démarche exploratoire et méthodologique pour réaliser que la mise en réseau de ses données est déjà un moyen de créer de nouveaux savoirs.

- 1 Tufte, Edward R., *The visual display of quantitative information*, Graphics Press, Michigan, 1983.
- 2 Robinson, Arthur H., *The thematic maps of Charles Joseph Minard*, in: *Imagio Mundi*, 21, 1, 1967, pp. 95-108 ou Palsky, G., *Des chiffres et des cartes: naissance et développement de la cartographie quantitative française au XIX^e siècle*, CTHS, Paris, 1996, p. 331.
- 3 Les exemples qui suivent sont tout quatre inspirés de communications scientifiques récentes.
- 4 books.google.com/ngrams.
- 5 Le paradoxe du *case-sensitive* et de la possibilité de rechercher des expressions de plusieurs mots est d'ailleurs lui-même entretenu par Google puisque l'exemple donné par son outil Ngrams (note précédente) montre que «Frankenstein» est plus utilisé que «Albert Einstein» et «Sherlock Holmes»... alors que «Einstein» et «Holmes» (sans leurs prénoms, puisque Viktor Frankenstein est cité sans son prénom) font des scores en réalité dix fois plus élevés que «Frankenstein»!
- 6 Illustration vectorisée par l'auteur.
- 7 Pour illustrer cette diversité, on y trouve des représentants d'une large palette d'intérêts académiques, comme Marie Curie, Albert Einstein, Hendrik Antoon Lorentz, Robert Andrew Millikan, Jules Destrée ou encore Gonzague de Reynold. À propos de la CICI, voir Renoliet, Jean-Jacques, *L'UNESCO oubliée, la Société des Nations et la coopération intellectuelle (1919-1946)*, Publications de la Sorbonne, Paris, 1999 et les articles de Laqua, Daniel, *Transnational intellectual cooperation, the League of Nations and the problem of order*, in: *Journal of Global History*, 6, 2, 2011, pp. 223-247 et Pernet, Corinne A., *Les échanges d'informations entre intellectuels: la conférence comme outil de coopération intellectuelle à la Société des Nations*, in: Clavien, Alain et Vallotton, François, *Devant le verre d'eau. Regards croisés sur la conférence comme vecteur de la vie intellectuelle 1880-1950*, Antipodes, Lausanne, 2007, pp. 91-106.
- 8 À l'exemple de l'excellente base de données LONSEA. Sibille, Christiane, *LONSEA - Der Völkerbund in neuer Sicht. Eine Netzwerkanalyse zur Geschichte internationaler Organisationen*, in: *Zeithistorische Forschungen*, 8, 2011, Online: <http://www.zeithistorische-forschungen.de/site/40209190/default.aspx> (consultation septembre 2013).
- 9 Cette approche a déjà été abordée dans Grandjean, Martin, *La connaissance est un réseau, perspective sur l'organisation archivistique et encyclopédique*, in: *Les Cahiers du Numérique*, 10, 3, 2014, pp. 37-54.
- 10 On parle ici des archives de la SDN conservées à Genève (United Nations Library, Palais des Nations), en particulier des sections 5 et 13 du secrétariat. Ne sont pas concernées les archives de l'Institut International de Coopération Intellectuelle (IICI), conservées à l'UNESCO (Paris).
- 11 En l'occurrence l'algorithme «Force Atlas 2» (Jacomy, Mathieu, Venturini, Tommaso, Heymann, Sebastien et Bastian, Mathieu, *ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software*, in: *PLoS ONE*, 9, 6, 2014) du logiciel open source Gephi (Bastian, Mathieu, Heymann, Sebastien et Jacomy, Mathieu, *Gephi: an open source software for exploring and manipulate networks*, in: *International AAAI Conference on Weblogs and Social Media*, 2009, pp. 361-362).
- 12 Archives SDN, cotes 13C/14297/20827; 13C/14297/21013; 13C/20085/28289; 13C/20085/31596; 13C/20085/33685; 5B/2975/11040; 5B/2975/18698; 5B/5353/10574 et 5B/5353/39878.
- 13 Un exemple de réflexion à ce sujet chez De Nooy, Wouter, *Fields and networks: correspondence analysis and social network analysis in the framework of field theory*, in: *Poetics*, 31, 2003, pp. 305-327.
- 14 U. Brandes, M. Hennig, I. Mergel, J. Pfeffer, *Studying Social Networks: A guide to empirical research*, Frankfurt am Main, 2012, pp. 149-182.
- 15 Linton C. Freeman, *Centrality in social networks: Conceptual clarification*, in: *Social Networks*, 1, 1979, pp. 215-239.
- 16 À propos des mesures de centralité, voir Freeman, Linton C., *Centrality in social networks: conceptual clarification*, in: *Social Networks*, 1, 3, 1978, pp. 215-239 et en particulier Koschützi, Dirk, Lehmann, Katharina Anna, Peeters, Leon, Richter, Setfan, Tenfelde-Podehl, Dagmar et Zlotowski Oliver, *Centrality indices*, in: Brandes, Ulrik et Erlebach, Thomas, *Network Analysis*, Springer, Berlin, 2005, pp. 16-61.
- 17 Mark S. Granovetter, *The strength of weak ties*, in: *American Journal of Sociology*, 78, 6, 1973, pp. 1360-1380.
- 18 Ou *Eigenvector centrality*.
- 19 Ce qui n'est pas tout à fait le cas dans l'exemple de Google qui procède à des réajustements ponctuels.