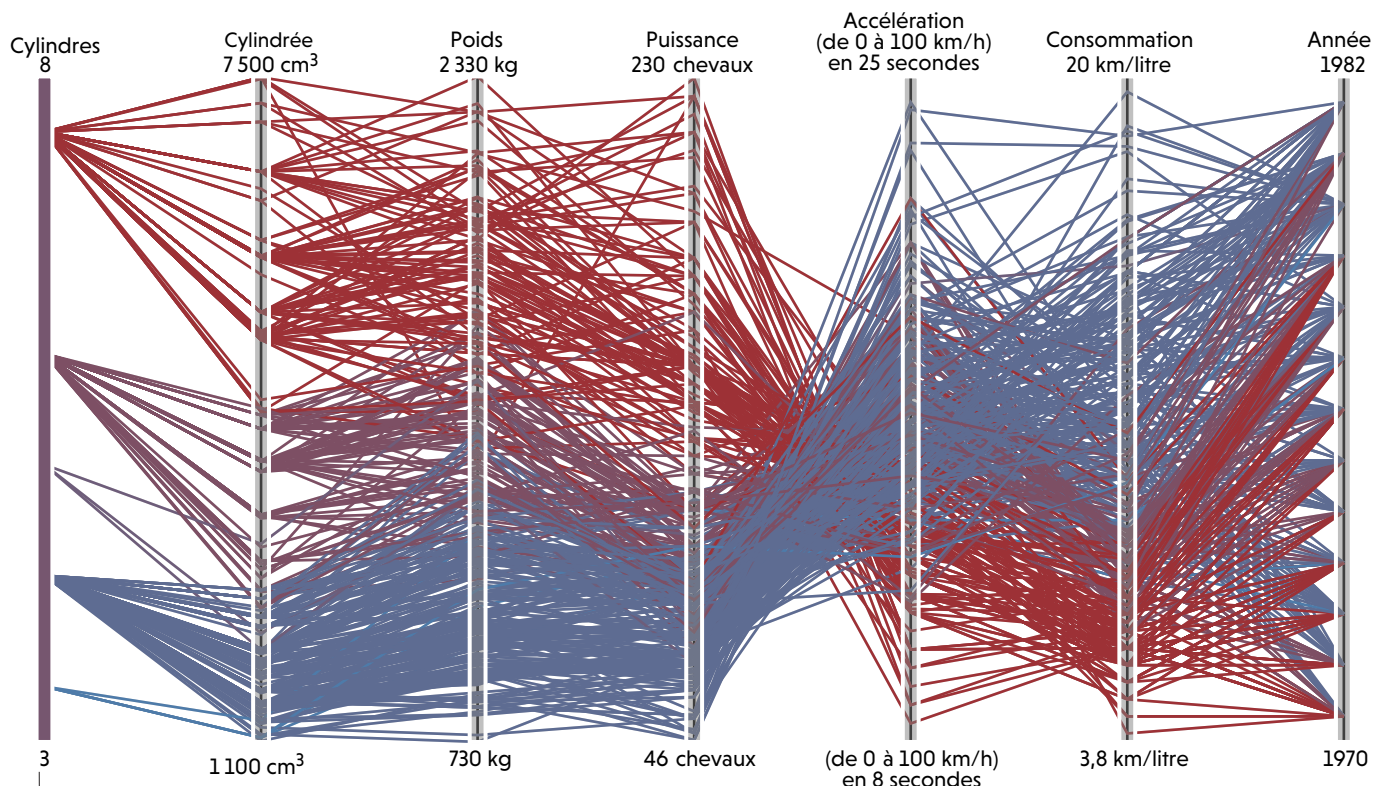
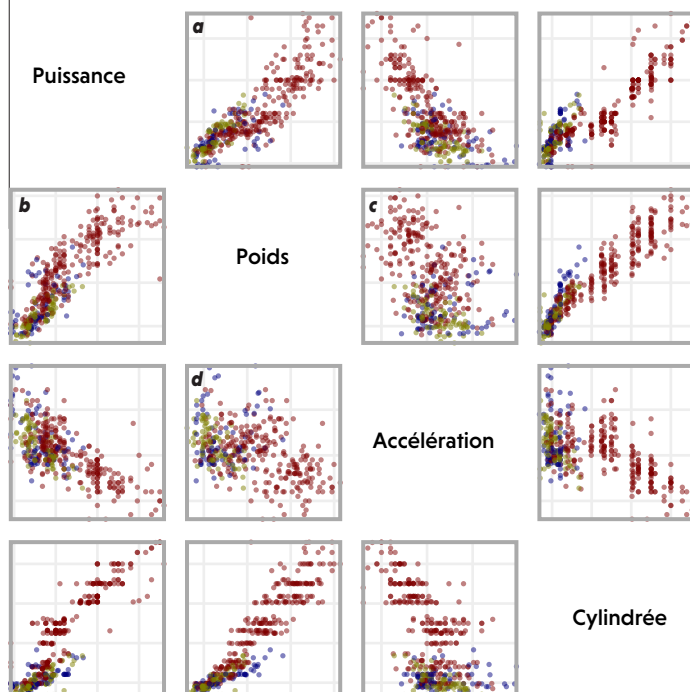




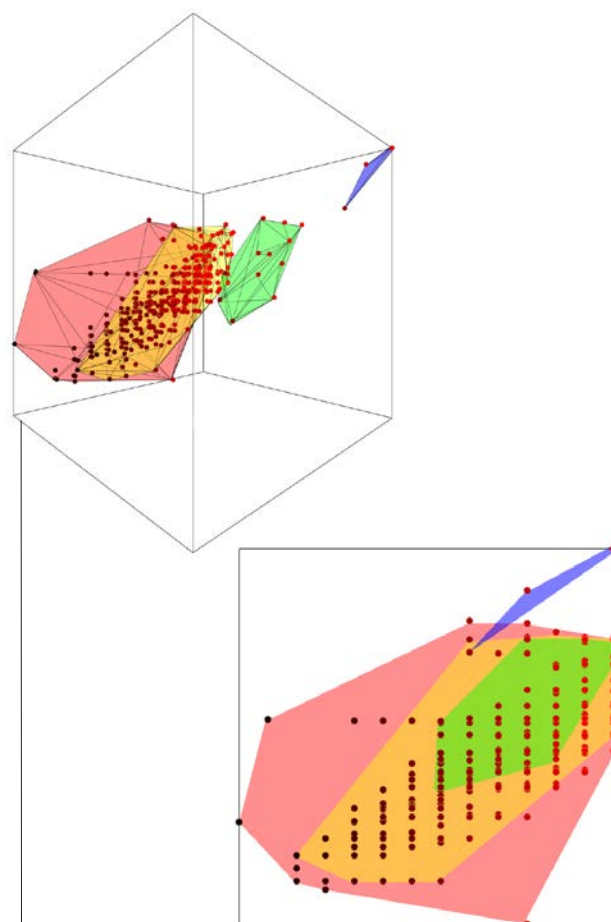
© J. M. Heer / U. Washington

DONNÉES MULTIDIMENSIONNELLES

Quand les données concernent plusieurs caractéristiques d'un ensemble d'éléments (ici, des voitures), on peut les afficher avec des coordonnées parallèles (*ci-dessus*) : chaque voiture est une ligne brisée qui relie ses caractéristiques selon chacune des dimensions (*les barres verticales*). Ce graphe révèle des motifs, telle la corrélation inverse entre puissance et accélération. Pour une analyse plus fine, on préfère les matrices de corrélation (*ci-dessous*), qui montrent pour chaque paire de dimensions le nuage de points correspondant. On repère une corrélation forte entre le poids et la puissance (a et b), mais pas entre le poids et l'accélération (c et d).



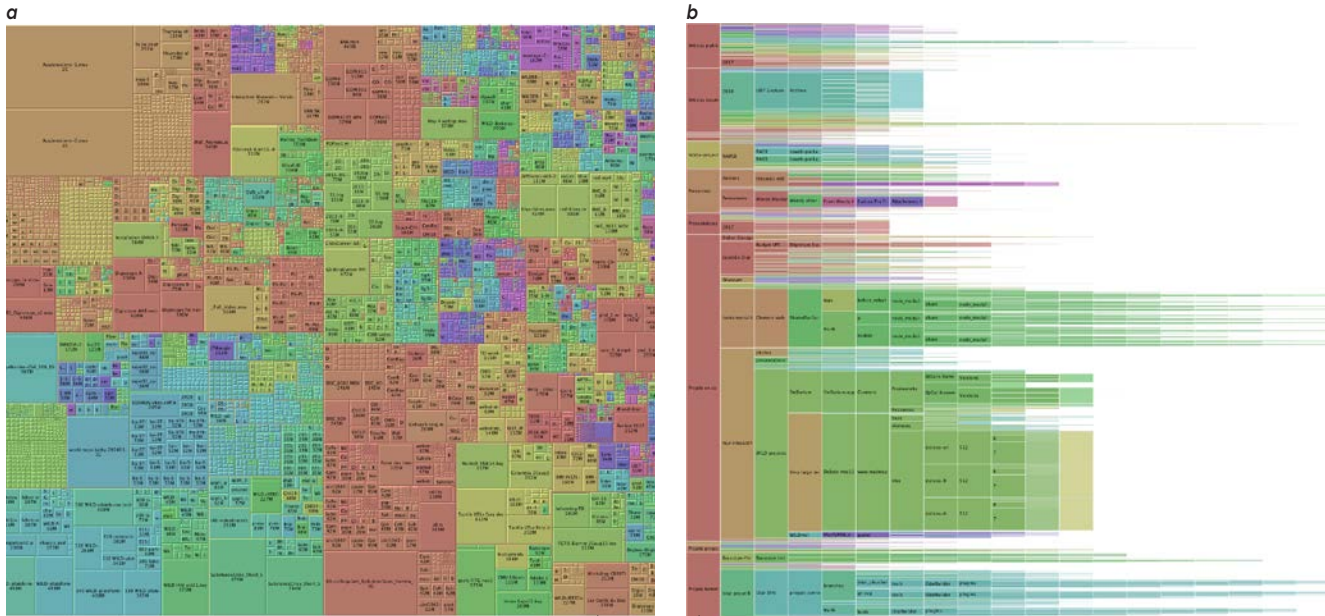
● Etats-Unis ● Union Européenne ● Japon



«ROLLING THE DICE»

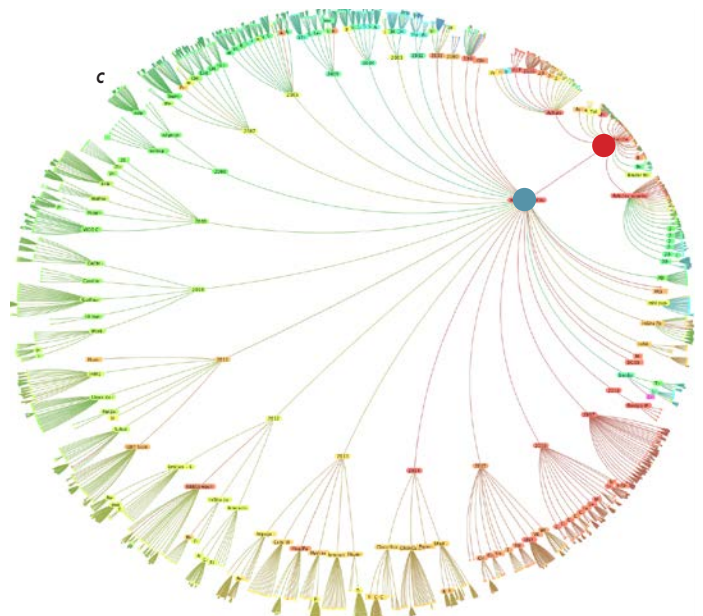
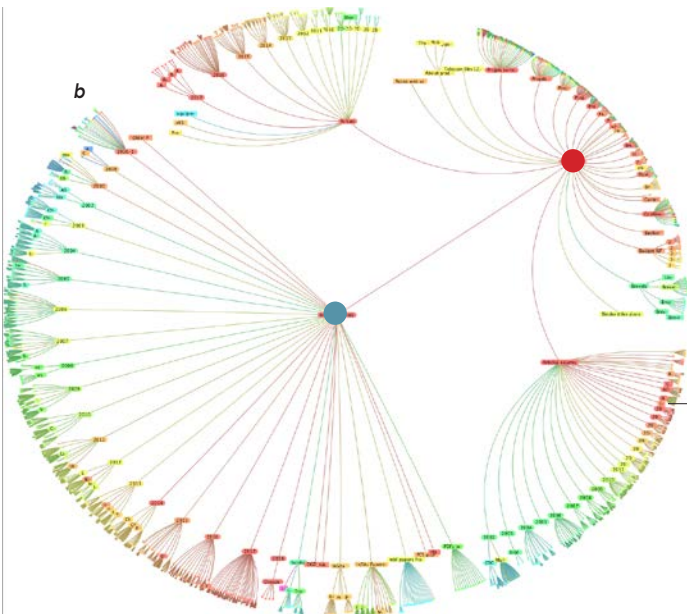
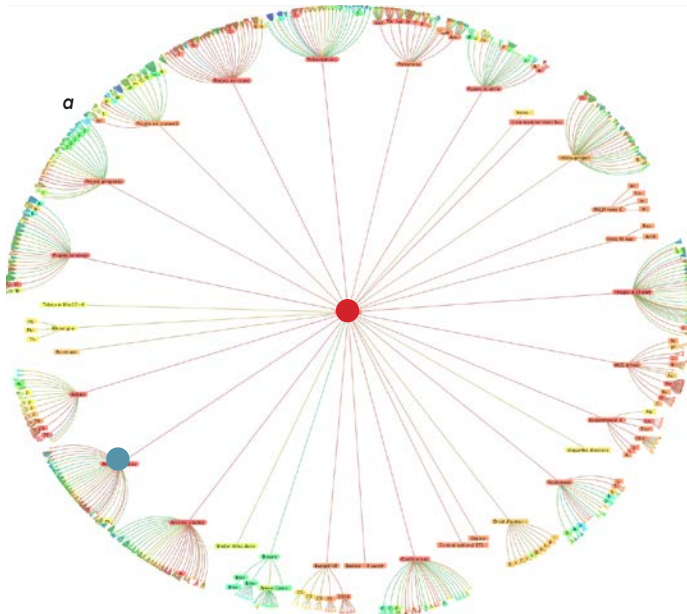
Avec cette technique d'animation 3D (littéralement «Faire rouler les dés»), on peut projeter dans l'espace les matrices de corrélations (*ci-contre*) et passer de l'une à l'autre afin de mieux comprendre leurs relations. On peut voir cette animation ici : <http://bit.ly/RolDice>

© N. Elmqvist, P. Dragicevic et J.-D. Fekete



UNE FORÊT D'ARBRES

La représentation traditionnelle des arbres (des graphes sans cycle où chaque nœud a un ou plusieurs descendants) sous forme... d'arbre n'est pas très lisible lorsqu'ils sont gros. On préfère des représentations qui occupent mieux l'espace, comme les arbres planaires, ou *treemap* (a), où la taille des rectangles est proportionnelle au volume de son contenu en données. On peut par exemple visualiser l'espace qu'occupe chacun des éléments du disque dur d'un ordinateur et l'arborescence selon laquelle ils sont organisés. La représentation en stalactites (b) est plus proche de la version classique.



L'HYPERBOLE

En utilisant le disque de Poincaré, un modèle de la géométrie hyperbolique, on peut représenter un arbre de taille arbitraire (les bords du disque sont mathématiquement situés à l'infini). La navigation dans un tel arbre hyperbolique (de a à c, on pousse le nœud rouge vers le Nord-Est) offre une vue détaillée d'une partie quelconque de l'arbre, sans perdre la vue d'ensemble.

L'ARBRE DE LA VIE

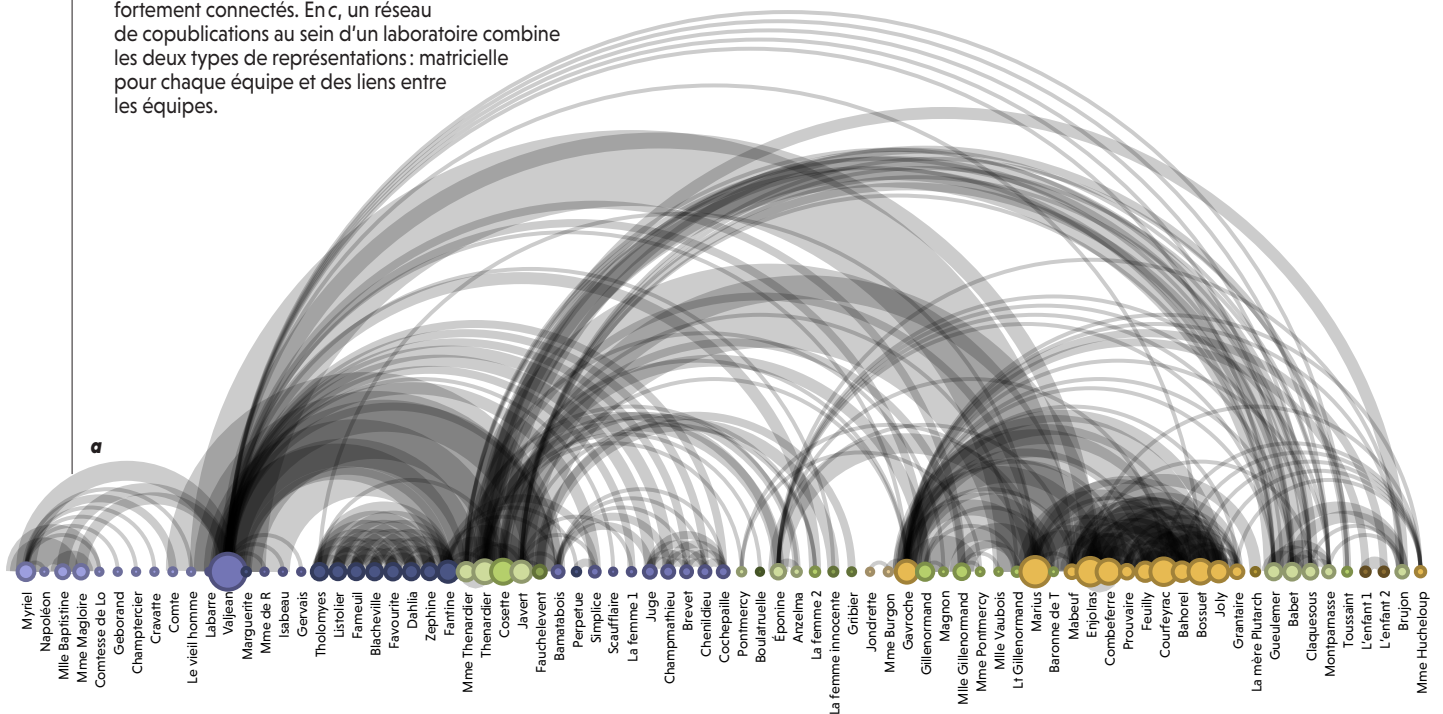
Sur le mur d'écrans Wilder du projet Digiscope (<http://digiscope.fr>), de six mètres de largeur et de 70 millions de pixels, s'affiche l'arbre planaire rassemblant près de 200 000 espèces d'organismes vivants. La technique des images hybrides permet de lire de loin les étiquettes des grandes catégories, mais aussi les détails lorsque l'on s'approche (dans le cartouche).



© Michel Beaudouin-Lafon

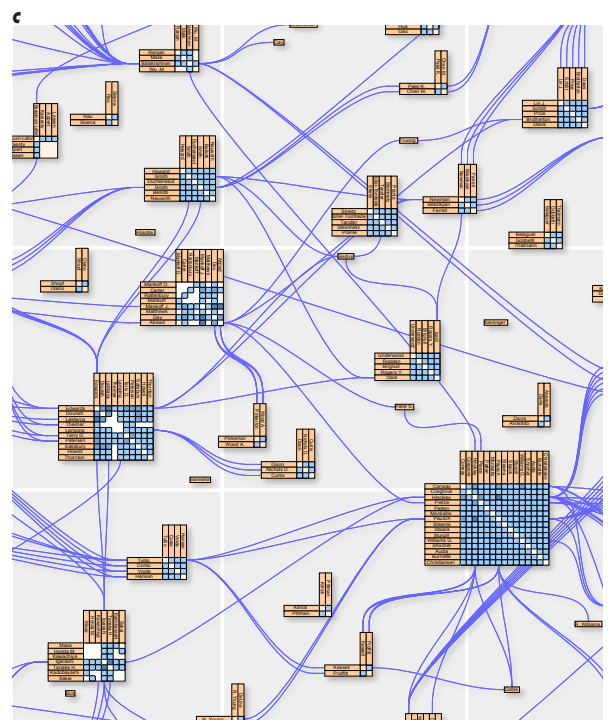
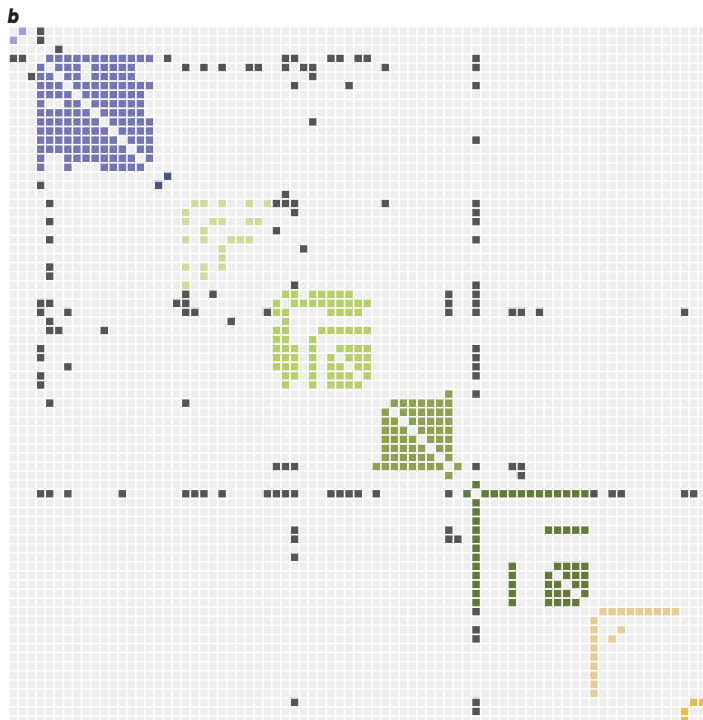
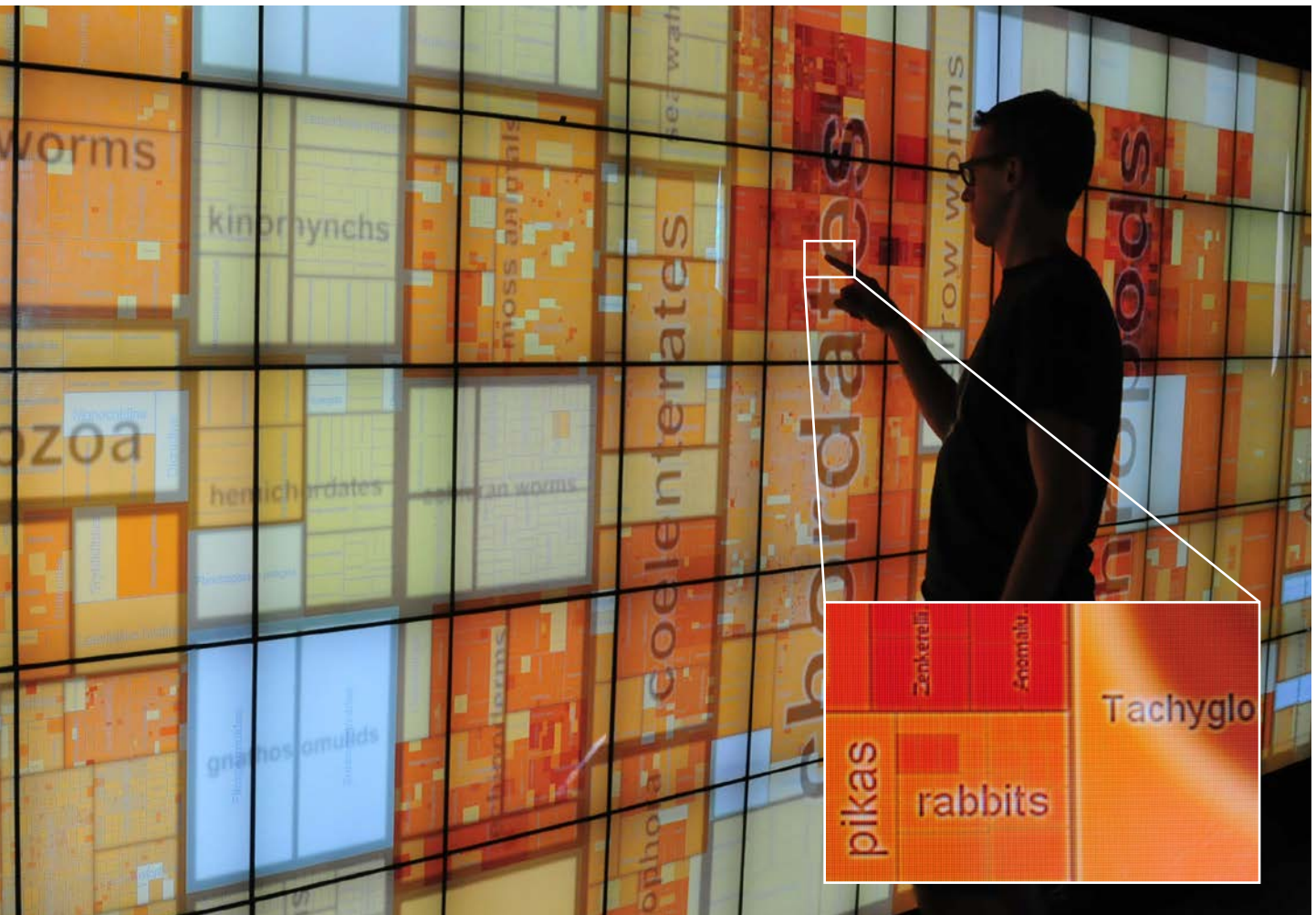
NON AUX SPAGHETTIS!

Les données en réseau sont difficiles à représenter car on obtient en général un « plat de spaghettis » illisible. Dans la représentation *a*, les arcs relient les personnages des *Misérables* qui apparaissent en même temps dans le livre. Ils ont été ordonnés et rassemblés par couleur de façon à rendre plus lisibles les regroupements. Plus un personnage intervient, plus le cercle correspondant est grand. En *b*, ce même réseau est représenté sous la forme de matrice où l'agencement des lignes et des colonnes fait apparaître des groupes fortement connectés. En *c*, un réseau de copublications au sein d'un laboratoire combine les deux types de représentations : matricielle pour chaque équipe et des liens entre les équipes.



Images 1 et 2 : © J. M. Heer/ U. Washington
Image 3 : © 2007, N. Henry, P. Dragicevic, J.-D. Fekete et Inria

Je vois, donc je comprends



L'ESSENTIEL

● La notion de cause a connu des fortunes diverses au cours des siècles et a cédé la place, en science, au principe de causalité.

● La science a continué de se bâtir sur des théories physiques et des lois universelles.

● Aujourd'hui, l'essor du *big data* s'accompagne d'une possible nouvelle science fondée sur les corrélations. Est-ce la fin de la théorie ?

● Si elle peut remporter quelques succès, cette voie ne conduira jamais à des théories du niveau de la relativité générale.

L'AUTEUR



ÉTIENNE KLEIN
est philosophe des sciences et
directeur de recherche au CEA.

Vers une nouvelle SCIENCE ?

Aux échecs, au jeu de go, le silicium écrase parfois le neurone. Est-ce une raison pour confier à des machines, gavées de données, l'activité scientifique de demain ?



Cet article est extrait en partie du livre *Matière à contredire*, essai de philo-physique, à paraître le 7 février 2018 aux Éditions de l'Observatoire.

E

n juin 2008, Chris Anderson, rédacteur en chef du magazine *Wired*, publiait un article au titre provocateur : «*The end of theory: The data deluge makes the scientific method obsolete*» («La fin de la théorie : le *big data* rend obsolète la méthode scientifique»). Il y défendait l'idée que lorsque nous disposons de suffisamment de données, les nombres

parleront d'eux-mêmes et les corrélations qu'ils dévoileront remplaceront les relations de causalité que manifestent les lois théoriques. La science changerait alors de visage puisqu'elle pourrait se développer sans plus énoncer d'hypothèses, et sans plus s'appuyer sur des théories explicites. Prophétie ou délire techno-utopiste ?

DE LA MONARCHIE ANGLAISE

Pour trancher, ou au moins disposer d'éléments de réponse, nous devons d'abord revenir sur l'idée de causalité. On dit souvent des sciences dures, notamment de la physique, qu'elles visent à identifier les causes des phénomènes naturels. En réalité, la notion de cause, au sens fort du terme, a posé tant de problèmes aux physiciens qu'ils ont fini par quasiment l'abandonner. Ils ne l'invoquent en tout cas presque plus de façon explicite, même si le concept de cause demeure présent dans certains de leurs discours.