

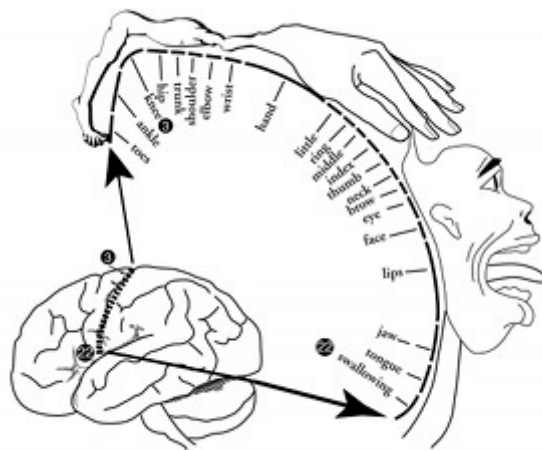
Échos de la recherche ([-Echos-de-la-recherche-.html](#))

## REPRÉSENTER LES MONDES II

Les réseaux

Piste rouge ([spip.php?page=mot&id\\_mot=22](#)) Le 24 mars 2010 - Ecrit par Ghys, Étienne

([\\_Ghys-Etienne\\_.html](#))



« Nous vivons dans un monde en réseaux. »

Voilà un commentaire qu'on entend de plus en plus souvent.

Facebook, Twitter, et Internet sont des exemples évidents, mais il y en a beaucoup d'autres :

- Le *cerveau humain* contient des milliards de neurones connectés entre eux par des axones.

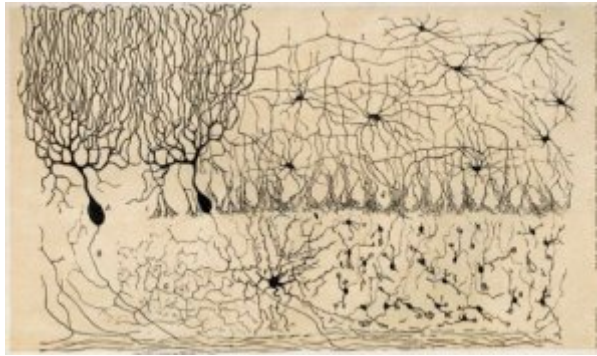
- Une *puce électronique* d'aujourd'hui peut contenir un milliard de transistors !
- On peut dire qu'une bonne partie de la physique consiste à étudier un grand nombre de particules (comme les molécules qui constituent l'atmosphère) qui sont en interaction suivant un réseau complexe.
- Mais on peut aussi penser aux *réseaux écologiques, sociaux, humains, commerciaux*, sans parler de réseaux téléphoniques etc.

Un enjeu majeur des mathématiques d'aujourd'hui est de réussir à « *cartographier* » ces nouveaux mondes virtuels. Il s'agit d'en donner des représentations visuelles qui nous aident à comprendre leur fonctionnement. Ce n'est pas facile puisqu'il s'agit de structures véritablement énormes, contenant parfois des milliards d'éléments, qu'on ne peut certainement pas dessiner sur une feuille de papier comme un plan du métro.

Cet article fait suite à [celui-ci](#) (*Représenter-les-mondes.html*) ; il pose plus de questions qu'il n'en résout. Je voudrais d'abord montrer quelques exemples. Ensuite, je discuterai d'un théorème qui permet d'être optimiste : *il est peut-être possible de représenter ces réseaux immenses de manière raisonnable.*

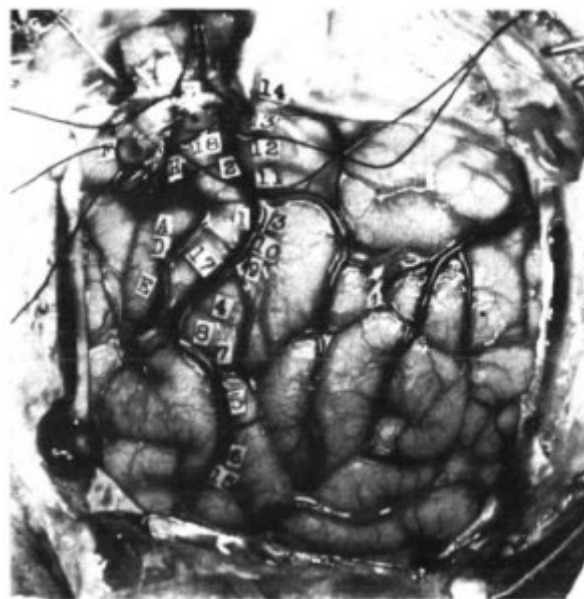
## Le cerveau

Le cerveau est un mystère. C'est vers la fin du dix-neuvième siècle que le « *dogme neuronal* » a été accepté [1 (#nb1)]. Le cerveau n'est pas un tissu continu mais bien un réseau contenant environ dix milliards de *neurones* connectés entre eux par des *axones*. Un neurone est connecté à beaucoup d'autres neurones : typiquement une dizaine de milliers. Comment se *représenter* cette structure ? Voici l'un des premiers croquis, datant de 1888 :



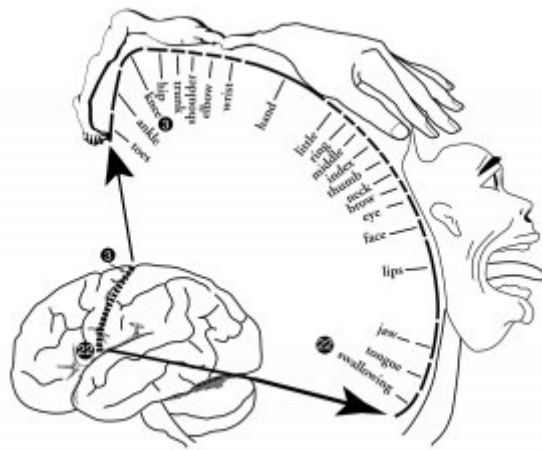
Cajal 1888 ([http://scienceblogs.com/neurophilosophy/2007/07/the\\_discovery\\_of\\_the\\_neuron.php](http://scienceblogs.com/neurophilosophy/2007/07/the_discovery_of_the_neuron.php))

L'un des premiers « *cartographes* » du cerveau est Penfield [2 (#nb2)]. Dans les années 1930, lors d'opérations du cerveau, il excitait certaines zones par des électrodes et notait les réactions du patient. Voici une photographie datant de 1937 sur laquelle on voit un certain nombre de tickets numérotés qui marquent l'emplacement des électrodes.



A chaque numéro correspond une réaction. Par exemple, le numéro 14 entraîne un chatouillement du genou jusqu'au pied gauche. Le numéro 5 entraîne une insensibilité de la partie droite de la langue etc. C'est le début de la théorie de l'**homonculus** qui consiste à établir une carte du cerveau en rapport avec les zones du corps qui lui correspondent.

Voici ce fameux homonculus : la « **carte du cerveau** » :



Il s'agit donc d'une carte  $f : X \rightarrow Y$  dont la source est le cerveau et le but le corps humain. Quel crédit apporter à ce genre de « carte » ? Penfield explorait un territoire totalement inconnu !

Au minimum, on peut s'étonner de ne pas voir de sexe [3 (#nb3)] !

Aujourd'hui, cette théorie a perdu toute prétention scientifique. Penfield le reconnaissait bien volontiers sur la fin de sa vie en écrivant « *Ces figures... ont les défauts, et les qualités, des bandes dessinées dans le sens qu'elles ne sont pas précises anatomiquement. [...] Elles sont des aides pour la mémoire, rien de plus* » [4 (#nb4)]. Penfield est **considéré** ([http://fr.wikipedia.org/wiki/Wilder\\_Penfield](http://fr.wikipedia.org/wiki/Wilder_Penfield)) comme l'un des pionniers de la neurochirurgie moderne [5 (#nb5)].

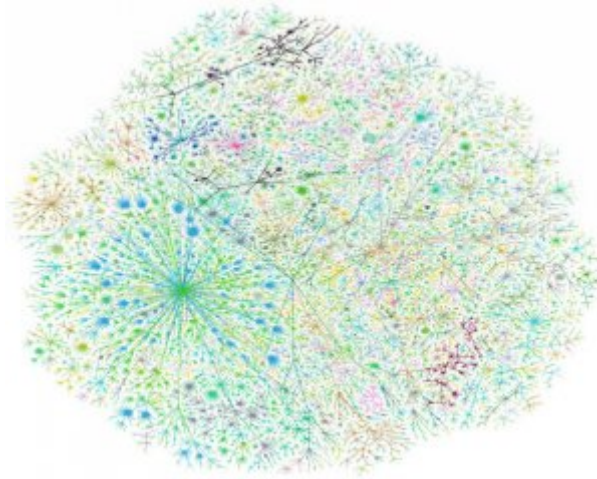
Grâce aux développements de l'imagerie médicale, on peut maintenant *cartographier* le cerveau en détail. Voyez par exemple un **atlas du cerveau** (<http://www.med.harvard.edu/AANLIB/home.html>). Mais il va de soi que le problème reste entier puisque ces images, si précises soient-elles, ne rendent pas compte de la *connectivité* et du *fonctionnement* du réseau de neurones.

## Internet

De nos jours, le nombre de pages internet dans le monde est environ de **dix milliards** : du même ordre de grandeur que le nombre de neurones d'un cerveau humain. Par contre, une page internet n'est liée en moyenne qu'à une vingtaine d'autres, ce qui fait qu'internet est moins « connecté »

qu'un cerveau. D'un autre côté, les influx électriques ou chimiques qui circulent dans les axones sont bien plus lents que les communications internet.

*Comment se faire une idée de la structure d'un tel réseau gigantesque ?*  
Voici une image réalisée en 1998, évoquant le croquis du réseau de neurones de 1888 que nous avons vu plus haut.



Les points dans ce dessin représentent des pages internet et les segments connectent des pages qui se font référence via un hyperlien.

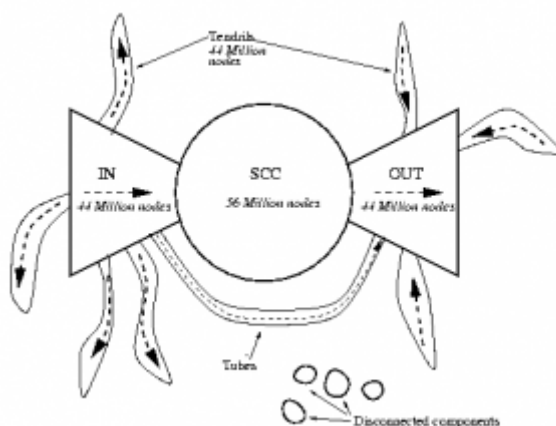
Voici une autre image, calculée en 2005, provenant du **Projet Opte** (<http://www.opte.org/>):



De même qu'une image précise du cerveau, ces images ne vont pas nous aider beaucoup lorsqu'il s'agira de comprendre un fonctionnement global.

Au fond, elles sont analogues à ce qu'on voit au microscope.

Les premiers « *web-explorateurs* » ont proposé vers 2000 une carte approximative, appelée « **Nœud papillon** » [6 (#nb6)] :



Cette carte reste d'actualité même si les chiffres ont changé. S'il ne faut pas en espérer une grande précision, elle dit « *quelque chose* » sur la structure du Web :

On peut décomposer le Web en *quatre grands domaines*, à peu près de même tailles.

- le **cœur**, qui est constitué d'un bloc de pages qui se connectent très largement entre elles (Strongly Connected Core) . C'est la partie où les connections fonctionnent à plein, dans tous les sens...
- la zone **OUT**. Il s'agit d'un bloc de pages qui sont beaucoup citées par le cœur mais qui par contre ne citent que peu de pages. En quelque sorte les pages qui nous intéressent tous mais qui ne sont intéressées par personne.
- la zone **IN**. Il s'agit des pages qui citent beaucoup le cœur mais qui ne sont presque pas citées. Les pages qui n'intéressent personne mais qui s'intéressent à tout !
- La zone **déconnectée**. Les pages qui ne sont pas, ou peu, connectées au cœur principal. Il s'agit par exemple de petites communautés qui ont des intérêts très précis et qui ne cherchent aucun contact extérieur.

Même si elle est simpliste, cette « carte » a du *sens*...

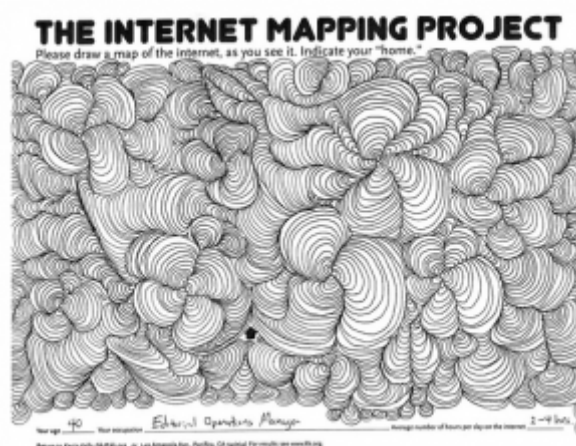
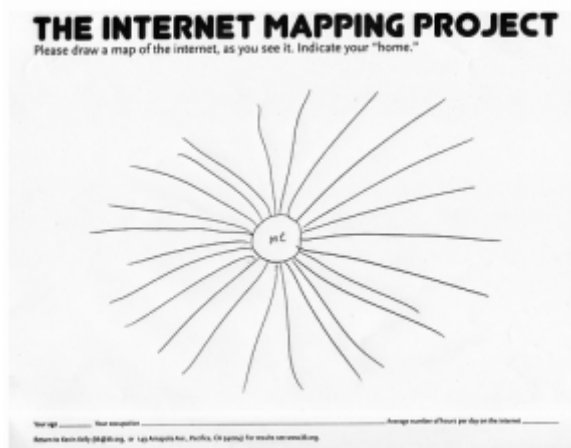


Comment peut-on aller plus loin ?

Il y a une vaste littérature sur ce sujet, en plein développement, qui intéresse les informaticiens, les physiciens, les mathématiciens, mais aussi les psychologues et les sociologues.

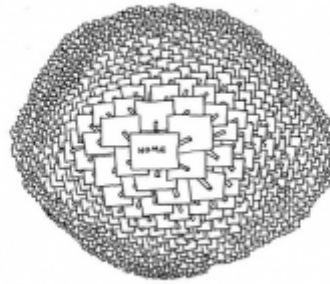
Face à une structure que nous ne comprenons pas suffisamment et qui joue pourtant un rôle de plus en plus important dans notre société, il est intéressant de se demander comment nous *imaginons* le Web. L'*Internet Mapping Project* a demandé à un grand nombre de personnes de *dessiner* leur vision du Web et d'indiquer leur présence sur leur dessin. On trouve [ici](http://www.flickr.com/photos/kevinkelly/sets/72157613562011932/) (<http://www.flickr.com/photos/kevinkelly/sets/72157613562011932/>) une petite centaine de dessins très instructifs.

En voici quelques-uns :



## THE INTERNET MAPPING PROJECT

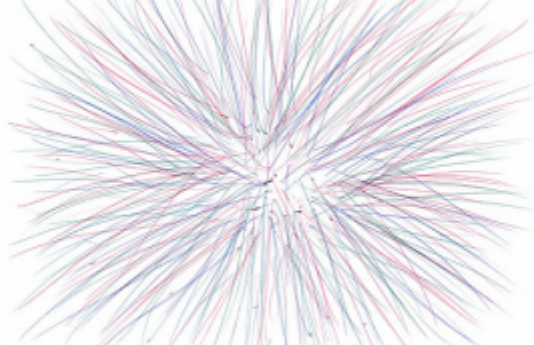
Please draw a map of the internet, as you see it. Indicate your "home."



Your age: 23 Your occupation: Student/Artist/Low New Employee Average number of hours per day on the internet: 5  
Return to Keri Kelly (k@kelly.org, or k@kelly.net, Fairfax, VA) for results and more info.

## THE INTERNET MAPPING PROJECT

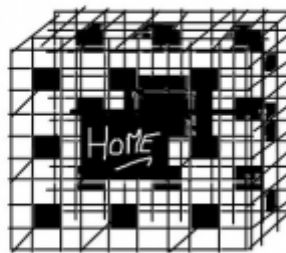
Please draw a map of the internet, as you see it. Indicate your "home."



Your age: 52 Your occupation: Senior Manager IT Consultant Average number of hours per day on the internet: 2  
Return to Keri Kelly (k@kelly.org, or k@kelly.net, Fairfax, VA) for results and more info.

## THE INTERNET MAPPING PROJECT

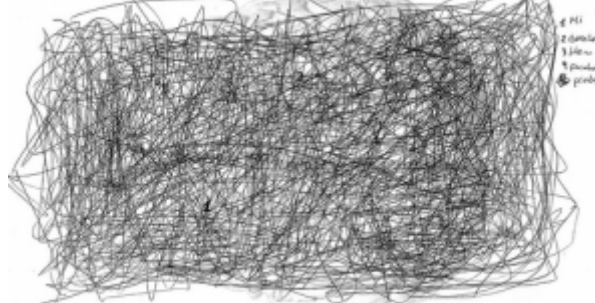
Please draw a map of the internet, as you see it. Indicate your "home."



Your age: 21 Your occupation: IT STUDENT Average number of hours per day on the internet: >3!  
Return to Keri Kelly (k@kelly.org, or k@kelly.net, Fairfax, VA) for results and more info.

## THE INTERNET MAPPING PROJECT

Please draw a map of the internet, as you see it. Indicate your "home."



Your age: 20 Your occupation: Disability Advocate Average number of hours per day on the internet: 2  
Return to Keri Kelly (k@kelly.org, or k@kelly.net, Fairfax, VA) for results and more info.



Please draw a map of the internet, as you see it. Indicate your "home."



Le site internet « **Internet geography** (<http://internetgeography.blogspot.com/>), a collection of ways to vizuallly organize and explore the internet » mérite une visite.

# Les Sciences

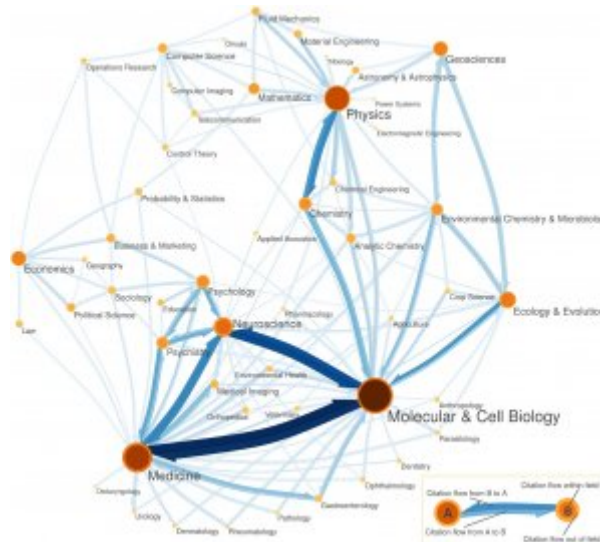
Les diverses sciences interagissent bien sûr. Comment est *structuré* le *réseau des sciences*?

De même que dans nos autres exemples, les premiers « *explorateurs* » ont dû se contenter de faits basiques et simples. Dans son cours de philosophie positive (publié entre 1830 et 1842), Auguste Comte range les sciences du plus général au plus particulier, du plus abstrait au plus concret : mathématiques, astronomie, physique, chimie, biologie, sociologie.

Bien sûr, tout en haut de la pyramide de Comte, il y a ... les mathématiques !

Aujourd'hui, les interactions entre les disciplines sont plus complexes que jamais. De nombreux scientifiques s'intéressent à la *cartographie des sciences*.

Voici une carte du monde scientifique, extraite de **Ranking and mapping Scientific Knowledge** (<http://www.eigenfactor.org/map/>).



Le mathématicien regrettera peut-être de ne plus occuper le sommet de la pyramide ?

*Quelle est la signification d'une telle carte ?*

Voici la méthode suivie, très discutable bien sûr, mais peut-on faire mieux ?

Le point de départ est une base de données qui recense un grand nombre de journaux scientifiques (7 000 environ) et qui permet en particulier de dénombrer le nombre d'articles publiés dans un journal qui citent dans leur bibliographie un article d'un autre journal. La base contient 60 000 000 de citations dans la dernière décennie.

On peut envisager cette donnée comme un grand tableau à 7 000 lignes et 7 000 colonnes. A la  $i$ -ème ligne et la  $j$ -ème colonne on indique le nombre  $N(i, j)$  d'articles publiés par le journal numéro  $i$  qui sont cités dans un article publié dans le journal numéro  $j$ . On peut aussi penser que chaque journal  $i$  donne un vecteur dans un espace de dimension 7 000 : celui dont les coordonnées sont  $(N(i, 1); N(i, 2), \dots, N(i, 7000))$ .

La première idée est de considérer que deux journaux  $i$  et  $j$  sont

*thématiquement proches* si l'angle entre les vecteurs correspondants est petit. On peut en effet penser que si les deux vecteurs « pointent dans la même direction », cela veut dire qu'ils accordent la même importance relative à tous les journaux, ce qui signifie plus ou moins qu'ils s'intéressent aux mêmes problèmes. Si vous vous souvenez du « produit scalaire de deux vecteurs », vous comprendrez qu'on peut facilement calculer l'angle entre deux vecteurs du plan à partir des coordonnées des vecteurs. C'est la même chose dans un plan ou dans l'espace, qu'il soit de dimension 3 ou ... 7 000 ! La géométrie en dimension 7 000 ne doit pas effrayer nos lecteurs ! Quoi qu'il en soit, vous pourrez admettre que d'une manière ou d'une autre, on est capable d'estimer des « distances » entre les journaux.

Il s'agit ensuite de faire une opération de « *clustering* », c'est-à-dire qu'on va regrouper par paquets les journaux qui sont très proches entre eux. Il y a beaucoup de manières de procéder, qui ne donnent d'ailleurs pas toutes le même résultat.

On considère ensuite les « clusters », c'est-à-dire les groupes de journaux qui ont été regroupés, et on les considère comme représentant « *une science* ». Pour déterminer le nom de cette science de manière automatique, l'ordinateur choisit un mot clé qu'il trouve en commun dans beaucoup de titres des journaux du « cluster ». Dans le cas présent, l'ordinateur a sélectionné 88 « sciences ».

Une fois les « sciences » isolées et définies, on fait « comme si » tous les journaux du même groupe n'en formaient qu'un seul, si bien qu'on peut maintenant compter les citations entre sciences et on a un tableau qui n'a *que* 88 lignes et 88 colonnes. En calculant à nouveau un angle, on peut ainsi déterminer une « distance » entre deux sciences.

Reste à représenter tout cela par 88 points dans le plan dont les distances s'accordent au mieux avec ces angles. On y arrive par exemple avec une méthode de minimisation, comme décrit dans l'article précédent, mais là-encore il faut bien dire qu'on pourrait procéder de beaucoup de manières différentes.

Le résultat est la carte présentée ci-dessus.

Je laisse le lecteur juger de l'intérêt et de la pertinence de ce genre de constructions. Mais qu'il n'oublie pas les leçons de l'homoncule de Penfield ou du nœud papillon du web : parfois des idées simples peuvent donner de bonnes images qui valent bien mieux qu'une absence d'images. Qu'il n'oublie pas non plus certaines erreurs commises par la science dans le passé : la « **phrénologie** (<http://fr.wikipedia.org/wiki/Phrénologie>) » par exemple qui prétendait que les bosses du crâne d'un être humain reflètent son caractère.

De nombreux articles paraissent actuellement autour de ce genre de questions, chacun apportant ses méthodes... souvent discutables.

Cartographier le monde scientifique, internet ou le cerveau, n'est certainement pas facile.

Il n'empêche que dans deux cas significatifs au moins, internet et le monde des sciences, nous sommes en mesure de faire un dessin sur une feuille de papier qui indique « en gros » comment les parties de ce réseau communiquent entre elles.

Je voudrais énoncer un théorème qui met en évidence une propriété de régularité qui vaut pour **TOUS** les réseaux, quelles que soient leurs tailles, et qui permet d'en espérer une représentation « raisonnable ».

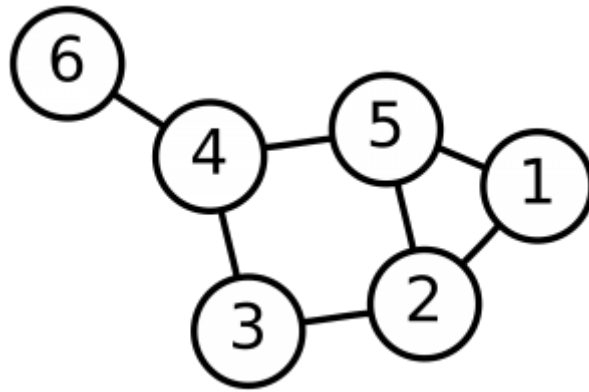
## Réseaux et graphes

Mathématiquement, les réseaux qui nous intéressent ici sont ce qu'on appelle des **graphes**. Il s'agit de considérer un ensemble d'objets qu'on appelle des **sommets** et qui, suivant le cas, représenteront les neurones, ou les pages internet, des journaux scientifiques, ou tout autre chose.

Mais surtout, il s'agit de considérer que deux **sommets** peuvent être reliés par une **arête** qui symbolisera une connexion neuronale, un hyperlien entre deux pages, une citation d'un journal vers un autre, ou le fait d'être « amis » sur Facebook etc.

Lorsque le nombre de sommets et d'arêtes est petit, on peut dessiner cela sur une feuille de papier. Voici par exemple un graphe qui a 6 sommets et

7 arêtes.



La *théorie des graphes* est une partie assez ancienne des mathématiques, mais elle est confrontée depuis quelques années à de nouveaux enjeux. Il s'agit de se concentrer sur des graphes dont le nombre de sommets dépasse l'imagination. Nous avons vu plus haut quelques tentatives de dessins représentant le graphe du web et ses milliards de sommets.

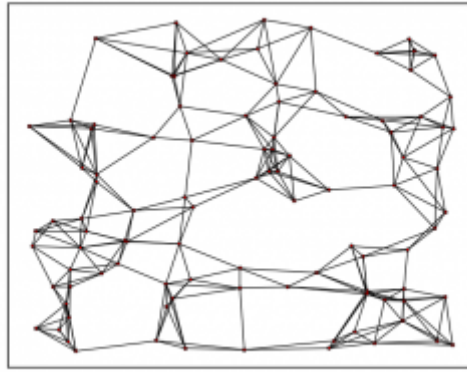
Le *graphe du web* peut être considéré comme un objet d'étude physique, sur lequel on peut faire des mesures expérimentales ; évaluer par exemple la rapidité avec laquelle les informations circulent, un peu à la manière du physicien qui étudie la circulation de l'électricité dans un réseau électrique...

Mais on peut aussi considérer que le graphe web est un objet qui concerne les mathématiciens.

L'une des premières questions qui se sont posées à son sujet concernait son caractère *aléatoire*. Prenez un grand nombre de points, par exemple dix milliards et pour chaque paire de points, connectez-les par une arête aléatoirement, disons avec une probabilité de 2 sur un milliard. Faites cela de manière indépendante pour chaque paire de points. De cette manière, vous créerez un graphe qui possède à peu près le même nombre de sommets et d'arêtes que le graphe web.

Voici ce qu'on obtient avec 100 points et une probabilité de 5/100 de connecter deux points :





C'est ce qu'on appelle le **graphe aléatoire de Erdős-Rényi** (qui ont développé toute une théorie à ce sujet à la fin des années 1950, mais avec des motivations bien éloignées d'internet). Ce graphe aléatoire ressemblera-t-il au graphe web ? Les connexions qui se créent dans le web peuvent-elles être attribuées au hasard ? La réponse est **NON**. Le graphe web fonctionne selon un aléa qui lui est propre et qu'on commence à savoir modéliser...

Mais laissons-là ces développements et dirigeons-nous vers le théorème de Szemerédi. Je recommande un livre [7 (#nb7)] de Bonato pour plus d'informations sur le graphe du web (à un niveau de premier cycle universitaire).

## Le théorème de Szemerédi

Ce théorème a été démontré dans les années 1970, également avec des motivations bien différentes [8 (#nb8)].

Nous voulons dire quelque chose de significatif sur un graphe qui a *un très grand nombre de sommets*, trop grand pour qu'on puisse les considérer tous. D'ailleurs les graphes auxquels nous pensons sont si grands qu'il n'est même pas clair qu'ils soient bien définis. Même s'ils l'étaient, *ils ne sont pas statiques* de toutes façons. De nouvelles connexions apparaissent et disparaissent à tout instant dans le web et nos neurones se désactivent malheureusement peu à peu... Nous sommes donc dans une situation où l'objet que nous voulons décrire n'est qu'à peu près défini et inaccessible. Nous ne pouvons donc pas rêver mieux que de chercher une compréhension approximative, disons à 1 % près.

Je noterai donc  $G$  un **graphe quelconque** et je vais essayer d'expliquer ce que pourrait signifier « *une compréhension de taille raisonnable à 1 % près* ».

Considérons deux parties  $A$  et  $B$  de l'ensemble des sommets du graphe  $G$ . Vous pouvez imaginer que  $A$  et  $B$  ne se rencontrent pas. Nous voulons étudier comment  $A$  et  $B$  sont connectés.

Soit  $X$  une partie de  $A$  et  $Y$  une partie de  $B$ . Comptons le nombre d'arêtes qui relient un point de  $X$  à un point de  $Y$  et notons-le  $N(X, Y)$ . Si nous notons  $N(X)$  et  $N(Y)$  le nombre d'éléments de  $X$  et de  $Y$ , on peut penser au quotient

$$\frac{N(X, Y)}{N(X) \cdot N(Y)}$$

comme la *probabilité de connexion* entre un point de  $X$  et un point de  $Y$ . A priori, cette probabilité dépend des parties  $X$  et  $Y$  bien sûr. On dira que  $A$  et  $B$  sont *bien connectés* (sous-entendu à 1 % près) si cette probabilité n'en dépend pas trop. Précisément, on demande qu'il existe un nombre  $p_{A,B}$  compris entre 0 et 1 tel que, pour tous  $X$  et  $Y$  contenus dans  $A$  et  $B$ , on a

$$0,99p_{A,B} \leq \frac{N(X, Y)}{N(X) \cdot N(Y)} \leq 1,01p_{A,B}.$$

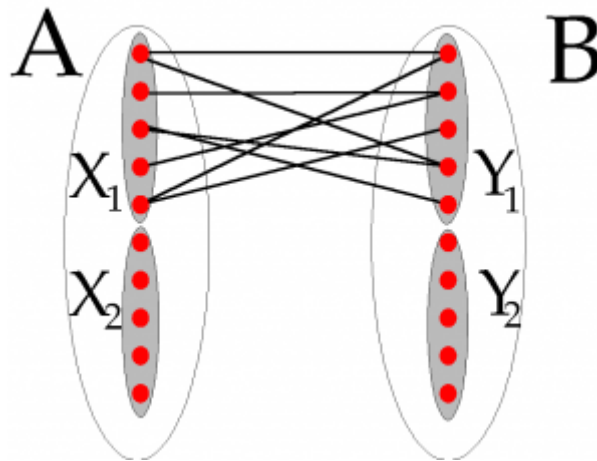
Que les lecteurs ne se laissent pas impressionner par cette formule : elle ne signifie rien d'autre que  $p_{A,B}$  et  $\frac{N(X, Y)}{N(X) \cdot N(Y)}$  sont égaux à 1% près.

Malheureusement, cette définition ne convient pas ! En effet si  $X$  et  $Y$  n'ont par exemple qu'un seul élément,  $\frac{N(X, Y)}{N(X) \cdot N(Y)}$  est égal à 0 ou à 1 et peut donc difficilement être proche de  $p_{A,B}$ . Alors, il faut modifier la définition pour tenir compte de cela. On demande plutôt que l'inégalité ci-dessus soit valide si  $X$  et  $Y$  sont « assez gros », c'est-à-dire s'ils contiennent au moins 1 % des sommets de  $A$  et  $B$ .

Voilà, nous avons une définition de «  $A$  et  $B$  sont bien connectés » (sous-

entendu à 1 % près).

Voici un exemple où  $A$  et  $B$  ne sont **pas** bien connectés. Vous voyez que  $X_1$ ,  $X_2$ ,  $Y_1$ ,  $Y_2$  contiennent la moitié des sommets de  $A$  et  $B$  respectivement. Mais il n'y a aucune connexion entre  $X_2$  et  $Y_2$  alors qu'il y en a beaucoup entre  $X_1$  et  $Y_2$ . Pour que  $A$  et  $B$  soient bien connectés, il aurait fallu qu'il y ait à peu près le même nombre de connexions dans les deux cas.

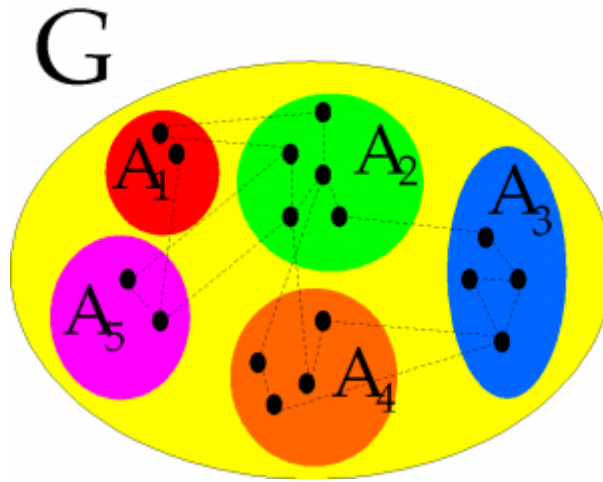


Passons maintenant à une deuxième définition.

Je dirai que j'ai une *partition* du graphe lorsque j'ai déterminé des parties  $A_1, A_2, \dots, A_n$  de l'ensemble des sommets du graphe de telle sorte que :

- les parties  $A_i$  ne se rencontrent pas.
- tous les sommets du graphe appartiennent à l'un des  $A_i$  (et donc à un seul).

Nous avons déjà rencontré une partition lorsque nous avons regroupé les journaux scientifiques en « clusters » que nous avons baptisé « sciences ». Voici une partition en 5 parties, contenant respectivement 2,5,4, 4 et 2 sommets.



Il s'agira d'une *équipartition* si tous les  $A_i$  contiennent le *même* nombre de sommets. Bien entendu, si l'on souhaite une équirépartition d'un graphe  $G$  en un certain nombre  $n$  de parties, il faut que  $n$  soit un diviseur du nombre  $N(G)$  de sommets, c'est-à-dire que  $N(G)/n$  soit un nombre entier : ceci n'a aucune raison d'être le cas. D'ailleurs, comme nous l'avons vu, le nombre de sommets des graphes auxquels nous pensons sont si grands qu'il est tout à fait illusoire de se demander s'il est divisible par 3 ou par 157 par exemple !

Alors, nous modifions légèrement la définition pour tenir compte de ce détail.

On dirons que les  $n$  parties  $A_i$  définissent une *presque équipartition* si le nombre de sommets dans chaque  $A_i$  est égal à  $N(G)/n$  à une unité près :

$$|N(A_i) - \frac{N(G)}{n}| \leq 1.$$

Bien entendu, les partitions seront d'autant plus intéressantes pour nous que le nombre  $n$  de parties est petit, tout au moins par rapport au nombre de sommets du graphe initial, que nous n'osons même pas imaginer ! Alors, prenons un nombre  $n$  raisonnable, par exemple  $n = 1000$ .

Nous cherchons des presque équipartitions bien connectées. Cela signifie que les parties  $A_i$  et  $A_j$  sont bien connectées. Toutes ? Non, ce serait trop demander. « Presque toutes » fera l'affaire. Nous demandons donc que parmi les  $n \times n$  ( $= 1\,000\,000$ ) choix possibles de  $i, j$ , au moins  $0,99n^2$

(soit 990 000) d'entre eux sont bien connectés.

Voilà, je peux énoncer le théorème mais avant cela je voudrais mettre en garde le lecteur que l'énoncé qui vient n'est (peut-être) pas correct ! Nous devons le rectifier ensuite.

**Théorème (Szemerédi)** : *Pour tout graphe, aussi grand soit-il, on peut toujours trouver une presque équipartition bien connectée en 1000 parties.*

## L'intérêt du théorème et quelques commentaires...

Une fois connue une presque équipartition bien connectée, on peut faire un dessin « *raisonnable* » du graphe qui nous intéresse.

Il s'agit de dessiner 1000 points dans le plan, chacun symbolisant un  $A_i$  qu'on peut donc penser comme un « cluster », une « science », ou peut-être un « organe » du réseau, et portant le numéro  $i$ .

Ensuite, sur le segment joignant les points correspondant à  $i$  et  $j$ , on écrit le nombre  $p_{A_i, A_j}$ .

Ce petit dessin est une véritable carte du graphe à 1 % près ! Et ceci, même si le graphe possède des milliards de milliards de sommets. En effet, il ne permet évidemment pas de reconstruire tout le graphe mais on connaît à 1% près les nombres de connexions entre deux parties  $X, Y$  (suffisamment grosses) du graphe : il suffit de savoir comment  $X, Y$  rencontrent les « clusters »  $A_i$ . En bref, celui qui est intéressé par le « *fonctionnement global* » du graphe se contentera tout à fait d'une telle carte. Si on le souhaite, on peut représenter les 1000 points dans le plan de manière à faire en sorte que les distances entre ces points représentent « au mieux » la connectivité. Deux  $i$  et  $j$  seront représentés par des points d'autant plus proches que le nombre  $p_{A_i, A_j}$  est grand.

Restent **deux commentaires** que j'ai gardés pour la fin puisqu'ils détruisent quelque peu l'intérêt du théorème ! Mais ils ne détruisent certainement pas l'envie de l'améliorer pour lui ôter les deux défauts que



je vais mettre en évidence.

Tout d'abord, le théorème tel qu'il est énoncé ci-dessus n'est *pas correct* car malheureusement, je ne peux garantir une partition avec 1000 clusters, comme je l'ai annoncé ci-dessus de manière quelque peu péremptoire. **La vérité est que pour chaque précision souhaitée (ici j'avais pris 1%), il existe un entier  $n$  qui rend le théorème correct.** Quel est la valeur de ce  $n$  pour la précision 1 % ? Est-ce 1000 comme je l'avais avancé ? Je n'en sais rien. Tout ce que je sais, *c'est qu'il existe un entier  $n$ .* Ce qui est important c'est que cet entier  $n$  ne dépend que de la précision (que je choisis comme je veux) mais **il ne dépend pas de la taille du graphe** que je cherche à décrire. S'il en dépendait, le théorème n'aurait aucun intérêt... Je ne serais pas étonné que le lecteur soit frustré par cet énoncé : il existe un entier  $n$  mais on ne le connaît pas. C'est comme ça : on ne fait souvent que ce qu'on peut en mathématiques...

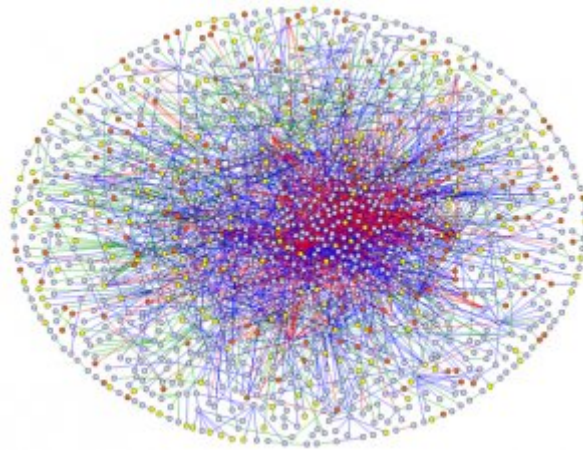
Alors, la chasse est ouverte pour trouver le meilleur entier  $n$  qui fait l'affaire à une précision donnée. Des travaux récents de T. Gowers semblent indiquer que  $n$  est de l'ordre de l'inverse du carré de la précision, ce qui n'est pas satisfaisant : pour une précision de 1 %, soit 0,01, cela donnerait  $n = 10000$ . Les chercheurs s'efforcent de rendre le **théorème effectif**, avec de vraies valeurs raisonnables de  $n$  pour des précisions raisonnables. Pour cela, ils pourront peut-être adapter quelque peu les conditions qu'on impose à la partition : l'énoncé est assez flexible, et on cherche un théorème utilisable avant tout.

Dans le même ordre d'idées, il ne suffit pas de dire qu'il existe une partition avec 1000 clusters. Encore faut-il les trouver si on souhaite produire une carte. Là encore, il s'agit d'un thème de recherche active. Comment développer des méthodes de calculs efficaces (et pas trop longs) qui permettent de trouver les clusters de Szemerédi ?

Les lecteurs qui possèdent un bagage mathématique solide pourront lire une prépublication passionnante sur les « Très grands graphes » [9 (#nb9)].

**D'autres réseaux...**

Pour terminer, voici un autre réseau qu'on aimerait mieux comprendre. Il s'agit d'un graphe décrivant 3 200 interactions entre 1 700 protéines humaines. Ce graphe a valu à ses auteurs un **prix scientifique** ([http://www.mdc-berlin.de/en/news/2008/20080910-erwin\\_schr\\_dinger\\_prize\\_2008\\_goes\\_to\\_resea/index.html](http://www.mdc-berlin.de/en/news/2008/20080910-erwin_schr_dinger_prize_2008_goes_to_resea/index.html)) en 2005.



Un peu compliqué ce graphe ! On aimerait lui appliquer le théorème de Szemerédi...

Article édité par Ghys, Étienne ([\\_Ghys-Etienne\\_.html](#))

## NOTES

[1 (#nh1)] **The discovery of the neuron** ([http://scienceblogs.com/neurophilosophy/2007/07/the\\_discovery\\_of\\_the\\_neuron.php](http://scienceblogs.com/neurophilosophy/2007/07/the_discovery_of_the_neuron.php)).

[2 (#nh2)] **Wilder Penfield, Neural cartographer** ([http://scienceblogs.com/neurophilosophy/2008/08/wilder\\_penfield\\_neural\\_cartographer.php](http://scienceblogs.com/neurophilosophy/2008/08/wilder_penfield_neural_cartographer.php)).

[3 (#nh3)] Voir cependant l'article de Damien Mascaret, *L'homunculus retrouve son pénis*, *Le Généraliste* n° 2340, 09.09.2005

[4 (#nh4)] « The figurines [...] have the defects, and the virtues, of cartoons in that they are inaccurate anatomically. [...] They are aids to memory no more » (Penfield et Jasper, 1954, pp. 105-106).

[5 (#nh5)] Mathieu Arminjon : **L'homuncule de Penfield** (<http://www.espacestems.net/document7825.html>); une icône neuropsychologique.

[6 (#nh6)] Broder et al. [Graph structure of the web](http://www9.org/w9cdrom/160/160.html) (<http://www9.org/w9cdrom/160/160.html>).

[7 (#nh7)] Bonato, A course on the web graph, AMS 2008.

[8 (#nh8)] Une grande constante dans le développement des mathématiques : une même idée peut servir plusieurs fois, souvent là où on ne l'attend pas.

[9 (#nh9)] Laszlo Lovasz : Very Large graphs, prépublication 2009.  
[Téléchargeable ici](http://arxiv.org/abs/0902.0132). (<http://arxiv.org/abs/0902.0132>)