Charting the Worlds[a]

by Étienne Ghys

I. Charting the Lands

Cartography has accompanied mathematics since its very beginnings[b] and continually renews its set of problems. Originally, it was a matter of drawing maps as accurately as possible of continents that were more or less well known. Above is a map dating from 1570, in which to the south we see a "Terra Australis Incognita" — which does not exist. Today, this mission of the cartographer remains relevant, but there are other new worlds we would like to represent in atlases. Cyberspace and the brain are only two examples of these modern-day terra incognita.

In this series of articles, I would like to present a few selected pieces of this interaction between cartography and mathematics. The subject is broad, and I will only touch upon it briefly, but I hope to show through these examples how the two disciplines mutually enrich one another.

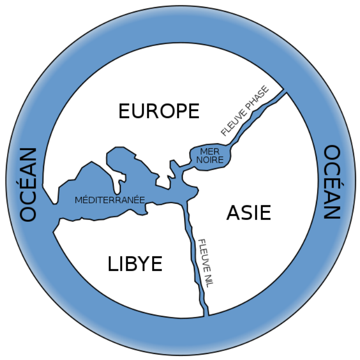
What is a map?

Below are two maps. The one on the left is a reconstruction of Anaximander's map (around 550 BCE). Described by Herodotus, it is one of the oldest maps representing the world as a whole. The map on the right is the famous "Carte du

[a]English translation by Rostislav Matveev of two articles *Représenter des Mondes*, and *Représenter des Mondes II* by Étienne Ghys, (2010), https://images-des-maths.pages.math.cnrs.fr/freeze/Representer-les-mondes.html and https://images-archive.math.cnrs.fr/Representer-les-mondes-II.html.

[b]This is also evident from the name of the vast and possibly most ancient area of mathematics — "geometry", that is "measuring the Earth." (Ed.)

Tendre," dating from the 17th century, which depicts the most mysterious land: that of romantic emotions.



A map is an image. It is a representation of a continent, a country, a region, and so on, on a medium — most often a sheet of paper — which one can, for example, slip into one's pocket when traveling through the country in question. If we denote by X the country and by Y the sheet of paper, the map is a mapping

$$f : X \to Y$$

What do we expect from a map?

That it be faithful, of course! We hope to gather enough information in Y to find our way around in X. We will see later — though the reader can already imagine — that if it is a matter of making a map of the New York subway, a sheet of paper will do perfectly well, but if we want to represent the network of connections among the billions of neurons that make up our brain, we will probably have to look for another medium!

Faithfulness in our language means injectivity of a mapping f.

Distances

We also expect a map to be accurate. In general, there is so much information in X that we would like to represent in Y, but Y is only a small sheet of paper and choices have to be made. Here, I will focus on a single aspect: that the map preserves relative distances, though there would be many other possibilities.

In ordinary cartography, X is a country on Earth, so two points x_1 and x_2 of X are separated by a distance that I will denote $\mathsf{distance}_X(x_1, x_2)$. When Y is a sheet of paper, two of its points y_1 and y_2 are likewise separated by a distance $\mathsf{distance}_Y(y_1, y_2)$.

An ideal map would be such that the distance between two points at the source, in X, is exactly the same as the one we measure on the map, in the target. That is

$$\mathsf{distance}_Y\big(f(x_1), f(x_2)\big) = \mathsf{distance}_X(x_1, x_2)$$

That said, such a map of France would have to be a thousand kilometers across, which is not very practical. Of course, we use maps at a certain scale. For example, if we want one centimeter on the map to represent one kilometer "in real life," that means that distances are multiplied in the target (on paper) by $k = 0.00001 =$

1/100000. Thus, for an ideal map at scale k we would instead require:

$$\mathsf{distance}_Y\left(f(x_1), f(x_2)\right) = k \cdot \mathsf{distance}_X(x_1, x_2)$$

The ideal does not exist!

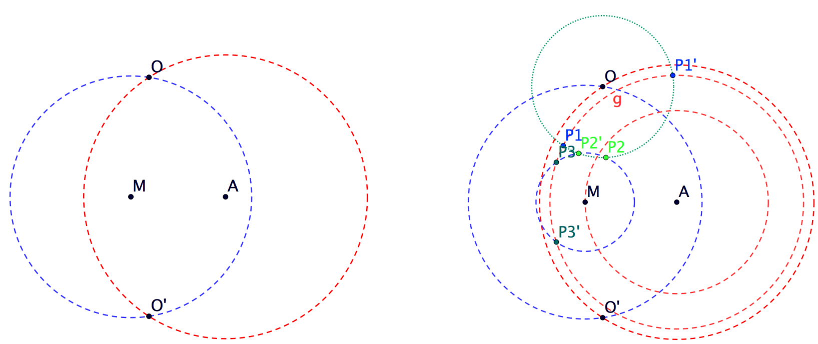
I claim that there is no map of Europe, for instance, that preserves distances exactly.

To prove this, I consider four cities: Athens, Madrid, Paris, and Oslo. Here is a table giving the distances between them.[1]

	Athens	Madrid	Paris	Oslo
Athens	0 km	1743 km	2414 km	2612 km
Madrid	1743 km	0 km	936.6 km	2224 km
Paris	2414 km	936.6 km	0 km	1351 km
Oslo	2612 km	2224 km	1351 km	0 km

Let us try to construct an exact map of Europe, say at scale $k = 1/100000$ (one centimeter equals one kilometer).[2] To do this, we must first mark the (image of a) first city, say Athens. This is the point A. Next, we place a second city, say Madrid: the point M must be 17.43 centimeters from A. No problem so far... We can place M at any point on a circle, but in any case all these positions are equivalent, up to rotating the sheet of paper around A.

Next, we place Oslo. We must place a point O at 26.12 cm from A and 22.24 cm from M. So we must construct two circles, centered at A and M, with radii 26.12 cm and 22.24 cm, and place O at their intersection. There are two intersection points, and we may choose either one, since the other is obtained from the first by symmetry. Let us therefore choose the one that respects the orientation we are used to, that is "to the North". So far so good.



It remains to place Paris on the map.

- We know the distance from Paris to Oslo and to Athens: this gives us only two possibilities for placing Paris, at P_1 or P_1'.
- We know the distance from Paris to Oslo and to Madrid: this also gives us only two possibilities for placing Paris, at P_2 or P_2'.

[1] I had intended to compute these distances myself from the latitudes and longitudes of the cities, but I discovered a website that does it for me in just a few clicks... The calculations assume that the Earth is perfectly spherical, which it is not quite..., but this does not change anything in the argument.

[2] Images are created using the excellent software GeoGebra. https://www.geogebra.org/

- We know the distance from Paris to Athens and to Madrid: this gives us only two possibilities for placing Paris, at P_3 or P_3'.

The problem is that these possibilities are not consistent: the points P_1 and P_1' do not coincide with P_2 and P_2'... They are not very far from one another, but they do not coincide...

It is impossible to draw an exact map of Europe!

Exercise 1. *Find four cities that, unlike the ones that we considered above, can be represented in the plane in such a way as to respect the distances exactly (up to a scale). Can you find all the situations in which this is possible?*

Exercise 2 (Difficult)**.** *Given six positive numbers $d_{1,2}$, $d_{1,3}$, $d_{1,4}$, $d_{2,3}$, $d_{2,4}$ and $d_{3,4}$, under what condition is it possible to find four points in the plane P_1, P_2, P_3, P_4 such that the pairwise distances $\mathsf{distance}(P_i, P_j)$ between them are exactly $d_{i,j}$?*

What can we do nevertheless?

Since there is no perfect map, we have to resign ourselves to approximations. For example, we can say that a map is precise "within 10%" if the distances measured in X and in Y do not differ by more than 10%, taking the scale k into account, of course. Formally speaking, we will say that a map f has precision ε if:

$$\mathsf{distance}\left(f(x_1), f(x_2)\right) \leq k \cdot (1+\varepsilon) \cdot \mathsf{distance}(x_1, x_2)$$

and

$$\mathsf{distance}(x_1, x_2) \leq \frac{1}{k} \cdot (1+\varepsilon) \cdot \mathsf{distance}\left(f(x_1), f(x_2)\right)$$

for all points x_1 and x_2 of X. If $\varepsilon = 0$, this would mean that a map is exact..., which does not exist. If $\varepsilon = 0.1$, this corresponds to a map accurate to within 10%.

These two formulas express the fact that the map works in both directions:

- if I know the distance $\mathsf{distance}_X(x_1, x_2)$ between two points "in reality", then I also know the distance $\mathsf{distance}_Y\left(f(x_1), f(x_2)\right)$ between the points that represent x_1 and x_2 on the map, up to the precision ε. I just need to multiply by the scale, and the error will be at most the percentage given by ε.
- if I know the distance $\mathsf{distance}_Y\left(f(x_1), f(x_2)\right)$ between two points on the map, then to find the distance $\mathsf{distance}_X(x_1, x_2)$ between the points represented by $f(x_1)$ and $f(x_2)$ on the map I have to divide by the scale the result will be at most ε percentage off.

Of course, we strive for maps for which ε is as small as possible.

The search for accurate maps began as soon as precise measurements of latitudes, and especially of longitudes, became possible. One anecdote illustrates the imperfection of maps in the 17th century.

> In a map published in the early *Mémoires de l'Académie des Sciences*, vol. VII, p. 430 — one of the first in which longitudes were counted from the Paris Observatory — the shaded boundaries represent those established from astronomical observations carried out, under the orders of Louis XIV, by various members of the Académie des Sciences. The boundaries shown in simple outline, with names in italics, reproduce the map of the famous Sanson, drawn up in 1679 and far superior to those that had preceded it. One sees that in Sanson's map the errors

were generally considerable, and that Brittany was displaced more than 100 kilometers to the west. Thus, when Louis XIV saw this map, he complained to the academicians that they had considerably reduced the extent of his dominions.[3]

A theorem and a question of J. Milnor

John Milnor is an American mathematician who has profoundly influenced twentieth-century mathematics, particularly in the fields of topology and dynamics. In 1969, he wrote a nice article[4] in which he proves a theorem about cartography and poses a problem... which still, up until today, has no solution.

Theorem. *Among all maps $f : X \to Y$, there exists at least one with a maximal possible precision.*

This result is an example of what we call a "compactness theorem".[5] But it is also a typical example of an existence theorem in mathematics, which is extremely frustrating for at least two reasons.

The first is that the theorem states that optimal maps exist, but it does not say how to find them! It is good to know that something exists, but without any concrete indication of where it can be found, its appeal is diminished... For example, if X is France, how can one construct such a map?

The second is that this is a theorem of existence and not of uniqueness. There may be many optimal maps. Of course, starting from a map $f : X \to Y$, one can translate or rotate the sheet of paper Y, or even change the scale, and thereby produce another map that clearly has the same precision. But this "new map" is really the same one: we have merely rotated the sheet of paper. Only mathematicians would consider them different!

Question. *Up to these operations of rotating the sheet of paper, shifting and rescaling, is there a unique optimal map?*

If that were the case, how could one construct it, what would it look like, what would its regularity be? All these are open questions, even for countries much simpler than France.

For example, consider a "rectangular" country X, bounded by two parallels and two meridians. Can one solve the above problems at least in this particular case? Uniqueness? Practical construction? Nice challenges for the readers!

A case where everything is understood...

The situation is not completely hopeless. In his article, Milnor completely solves the question for a circular country. Admittedly, there are few regions of the world that are circular and that one would want to map (apart from the polar regions). But at least it is a simple case in which everything is understood. This seems to be one of the cases of so called "Spherical cow in a vacuum".[c] Nevertheless...

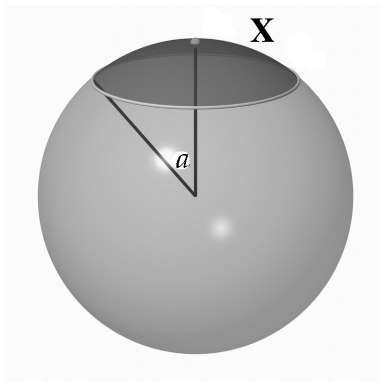
---

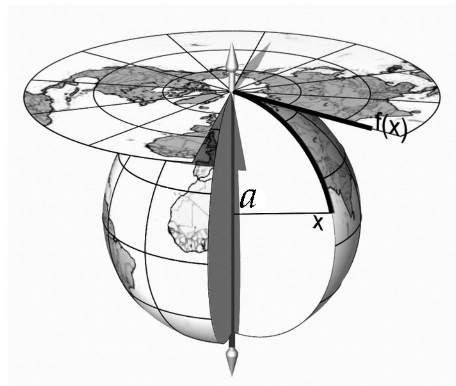
[3]Bigourdan, *La carte de France,* Annales de Géographie (1899) Volume 8 Numéro 42, pp. 427-437, Download: https://www.persee.fr/doc/geo_0003-4010_1899_num_8_42_6155(In French)

[4]J. Milnor, *A problem in cartography.* Amer. Math. Monthly 76 (1969) pp. 1101–1112.

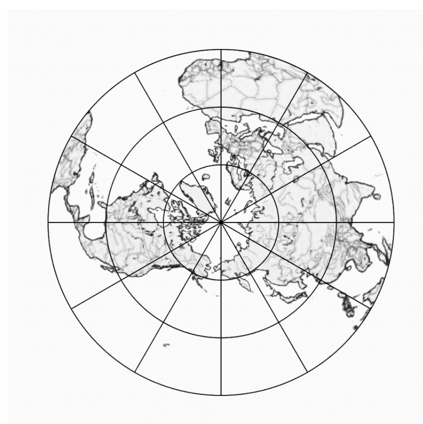
[5]Indeed, the result follows quite easily from Arzelà–Ascoli Theorem

[c]This alludes to a joke highlighting the oversimplifications common in theoretical science: A dairy farm hires a mathematician to improve the efficiency of milk production. After collecting massive amounts of data at the farm and spending a few sleepless nights analyzing it, the mathematician presents his report, which begins: *"Consider the case of a spherical cow in a vacuum..."*





The left figure above is a spherical cap X, centered for instance at the North Pole. On the right figure the tangent plane Y at the North Pole is shown, on which we are going to draw the map of X. Each point x of X lies on a meridian starting from the North Pole. We then consider the ray in Y that is tangent to this meridian at the North Pole, and we place $f(x)$ on this ray, at the same distance from the origin as x is from the North Pole.



I stress an important point: in X, the distances are *intrinsic distances* the sphere — that is, the length of shortest path drawn on the sphere from one point to another, without cutting through the interior of the Earth to go faster... The distance in X between x and the North Pole is therefore the length of the meridian arc from x to the North Pole. The resulting map is shown on the figure.

In cartography, this map is called the *azimuthal equidistant projection*[d] (attributed to Guillaume Postel), while in mathematical terms, it is known as the (inverse) *exponential map.*

Theorem. *The azimuthal equidistant projection is the only map of a spherical cap that has the best possible precision.*

Milnor Conjecture

It is not difficult to compute the distortion ε of the azimuthal equidistant map. The result is a formula:

$$\varepsilon = \sqrt{\frac{a}{\sin(a)}} - 1$$

where a denotes the aperture of the cap — the angle shown in the figure. Here is a small table giving the precision ε for various values of the angle.

[d] See, for example, the flag of the UN. (Ed.)

Angle a	Distortion ε	
$23°26'$	1.41%	Arctic polar region
$45°$	5.39%	
$66°34'$	12.5%	North of Tropic of Cancer
$90°$	25.3%	Northen Hemisphere
$113°26'$	46.9%	North of Tropic of Capricorn
$135°$	82.5%	
$156°34'$	162%	North of Antarctic Polar Circle
$179°$	1238%	

Instead of computing the precision as a function of the angle a, one can compute it as a function of the number

$$u = \frac{\mathsf{area}(\text{Cap})}{\mathsf{area}(\text{Earth})}$$

which lies between 0 and 1 and represents the proportion of the Earth covered by the cap.

One obtains a complicated and uninteresting formula, but Milnor shows that this precision is less (that is, better) than

$$\frac{1}{3}u + \frac{1}{2}u^2$$

whenever $u \leq 1/2$, that is, when the cap is smaller than a hemisphere.[6]

Milnor proposes a problem: show that the same estimate for the precision holds for noncircular, convex countries. A country X is convex if whenever two points x_1 and x_2 lie in X, the shortest path on the sphere joining them is entirely contained in X. In plain language, this rules out "fjords"...

Problem. *Let X be a convex country lying in a hemisphere, and let u be the proportion of the Earth's surface occupied by X. Is it possible to draw a map of X with precision $\frac{1}{3}u + \frac{1}{2}u^2$?*

More than forty years later, this problem still has not been solved.

Here is an example: let us surround the United States of America by a rectangle. This rectangle covers about 1.5% of the Earth.[7] According to Milnor, the best known map of the United States is accurate to within 2.2%, whereas according to his conjecture, there ought to exist one accurate to within 1%...

The literature on cartography is immense, and many types of maps exist, each trying to represent one feature or another as well as possible.[8,e]

Other trade-offs, other distances...

There are not only distances in kilometers that one might want to represent on a map. If x_1 and x_2 are two cities in France, we can compute the time $\mathsf{time}(x_1, x_2)$ needed to go from x_1 to x_2. One would need to be more precise, specifying whether one is traveling by car or by high-speed train, the time of day and the day of

[6] In fact, Milnor does not give this estimate, but it follows easily from his calculations...

[7] Such a powerful country that covers such a small percentage of the world...

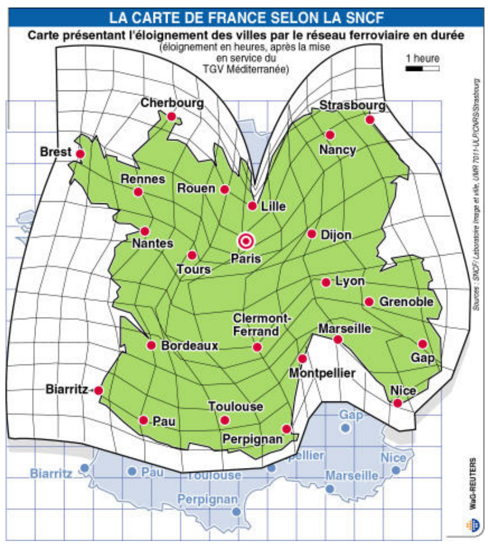
[8] For a mathematical approach see, for example, http://www.yann-ollivier.org/carto/carto.php (In French)

[e] There are numerous websites devoted to the mathematics of cartographic projections. An internet search will produce a list of resources to suit any taste. (Ed.)

travel, etc. We encounter a new problem: the space X that we want to represent is not known with complete precision, and one could even say it is not completely defined. Let us ignore this difficulty for the moment, promising to return to it in other articles. A precise temporal map would be very useful: one would read off in centimeters on the map the distance between $f(x_1)$ and $f(x_2)$ the time $\mathsf{time}(x_1, x_2)$, expressed for example in hours.

Of course, for the same reasons we saw above, it is in general impossible to construct an exact temporal map, and one has to do "as well as possible." A good exercise, following the reading of Milnor's article, is to check that, given a "temporal" country X, it is possible to find a map that represents it "as well as possible," with the same issues as before: no idea how to actually construct such a map!

Here is a temporal map of France for the SNCF.[f]



It was constructed to be exact "from Paris," in the same way that the azimuthal equidistant map is exact "from the North Pole." It is therefore an SNCF azimuthal equidistant map.

Along transverse routes, one shouldn't expect very high precision. How can one draw a more accurate temporal map of the SNCF network? There's work here for mathematicians.

There is another example, so called Travel Time Tube Map[9], using a cool modern technology. It is an interactive map of the London Underground network. Each time user chooses a station on the map, the map deforms to represent "the azimuthal equidistant projection from that station." In plain language, you can read directly on the map the travel time from that station.

Yet another method

Suppose we have a large number of points $x_1, x_2, \ldots, x_N$ and we have measured all the distances $d_{i,j}$ between these points. These distances may be in kilometers

[9] https://www.tom-carden.co.uk/p5/tube_map_travel_times/applet/

[f] SNCF (**S**ociété **n**ationale des **c**hemins de **f**er français) is the French railway operator. (Ed.)

or in hours, for that matter. How can we "best" represent this situation by points $y_1 = f(x_1), y_2 = f(x_2), \ldots, y_N = f(x_N)$ in the plane, given that we have just seen that no method is known for finding the "best map," the one whose existence is guaranteed by Milnor?

Once we choose the points y_i, the defect of the map will be smaller the closer $\mathsf{distance}_Y(y_i, y_j)$ is to $k \cdot d_{i,j}$, where k is our chosen scale. One idea is to compute the sum of the squares of the differences $\mathsf{distance}_Y(y_i, y_j) - k \cdot d_{i,j}$:

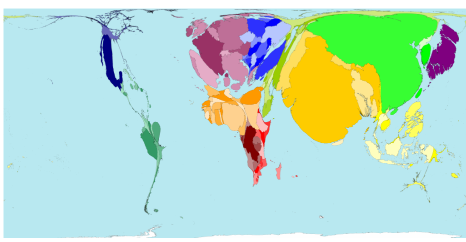
$$\mathsf{defect} := \sum \big(\mathsf{distance}_Y(y_i, y_j) - k \cdot d_{i,j}\big)^2$$

and try to find points y_i for which the defect is as small as possible.

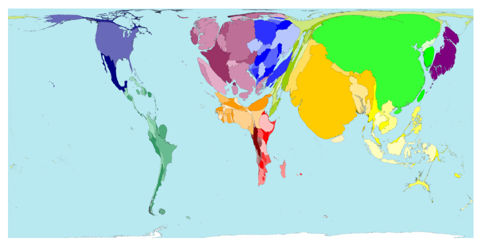
Why the sum of squares? There are several good reasons for this, particularly coming from physics, but the best reason is that this new minimization problem is much simpler than Milnor's. We know good methods for finding the y_i's that minimize this sum of squares, and computers can compute it without difficulty. We have changed our ambitions; our new problem may be less natural, but at least we know how to solve it! Cartography is also a matter of "cookbook recipes."

Other criteria

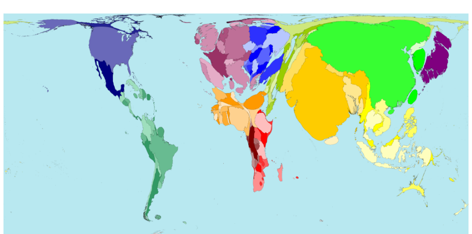
There are many other things one would like to represent on a map. Distances, even temporal ones, are only one example. Here I would like to speak about *cartograms*.[g] Here are some cartograms representing the world population.[10]



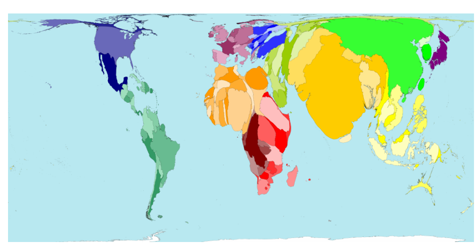
World population in 1500



World population in 1900



World population in 1960



Projected world population in 2050

[10]Images are taken from a very interesting site (https://worldmapper.org/) that offers many other kinds of cartograms.

[g]A cartogram is a map distorted so that the areas of geographic regions on the map represent a specific quantity, such as population or GDP, instead of their actual land area. Of course, such maps are not suitable for navigation, but serve as visualization of features other than distances. (Ed.)

As you have understood, the map $f : X \to Y$ tries to represent the population of a country P (in millions of inhabitants, for example) by the area of its image $f(P)$ (in cm^2, for example). In formulas, for every country P in X, we would like to have the relation

$$\mathsf{area}\,\big(f(P)\big) = \mathsf{population}(P)$$

Here again, the existence of such a map is anything but obvious; it follows from the theorem, due to Oxtoby and Ulam.

We will not state the theorem in full generality, but only a corollary sufficient for our purposes. We are given a *measure* m on X, which may represent a density of population, of cars, of wealth, etc. Each subset P of X thus has a certain measure[h] $m(P)$, which may represent the number of inhabitants in P, the number of cars, the total wealth of the inhabitants of P, yearly rainfall at P, and so on. We choose the units of measurement in such a way, that measure of the whole X is equal to 1. Not every measure is suitable for our purposes. We assume that measure m of some region is zero if and only if its area is zero. In particular, only two-dimensional regions (not points or curves) can have positive measure and conversely, the measure of any "country" must be positive. Theorem of Oxtoby–Ulam[i] implies the following statement.

Theorem. *There is a homeomorphism*[j] *from X to the square $f : X \to [0, 1]^2$ such that*

$$\mathsf{area}\,\big(f(P)\big) = m(P)$$

for all subsets P of X. In other words, one can deform the image of X to a square in such a way, that area of each distorted region will be equal to the measure assigned to it.

The good news is that the theorem is *constructive*, meaning that one can program a computer to draw the map, as in the examples above.

A couple of remarks are in order here:

- The cartogram is not at all unique. One can exploit this flexibility to try to impose additional constraints on the map, such as best preserving "something else." *Optimal transport theory*,[k] which has been growing rapidly in recent years, offers an approach that has already been discussed in the archives of *Images des maths* and to which we will probably return.
- Moser's theorem[l] and the practical method it provides — easy to program on a computer — went completely unnoticed among cartographers. Admittedly, Moser himself was likely not particularly interested in cartography.

[h] Not quite every subset is measurable, there are "exotic" subsets, that can not be assigned a measure. However, this phenomenon is essentially "invisible" in applications and can safely be ignored. (Ed.)

[i] An additional assumption about region X is necessary, namely that X is connected (consists of one piece) and simply-connected (has no holes inside). (Ed.)

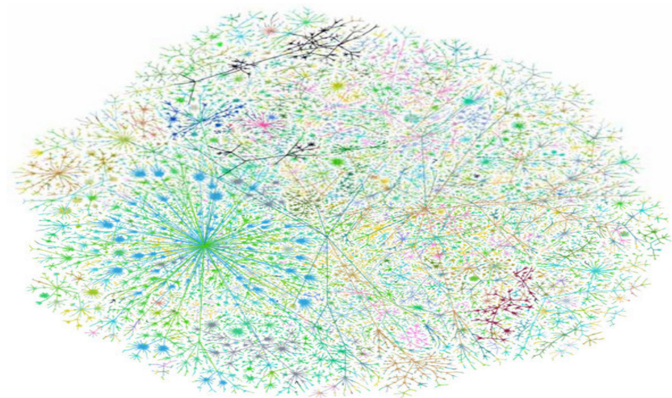
[j] A continous one-to-one map. (Ed.)

[k] *Optimal transport theory* (in the Monge formulation) concerns, roughly speaking, finding *optimal way* to move one pile of sand of a certain shape to another pile of sand of a different shape but the same volume. Different optimality criteria lead to different *optimal transportation plans*, that is prescriptions for where each grain from the first pile should be transported in order to fill the second pile. (Ed.)

[l] The difference between theorems of Moser and of Oxtoby–Ulam is that Moser assumes that the density of measure m is infinitely differentiable. In return, the map guaranteed by Moser is infinitely differentiable as well and also easily constructible. (Ed.)

It was only recently, in 2004, that the theorem was rediscovered by cartographers.[11] Good ideas are often born more than once.

II. Charting the Networks



"We live in a networked world." — this is a comment we hear more and more often.

Facebook, Twitter, and the Internet in general are obvious examples, but there are many others:

- The *human brain* contains billions of neurons connected to each other by axons.
- A *modern microchip* can contain roughly a billion of transistors interconnected in a highly complex pattern.
- A major part of physics is devoted to studying systems composed of vast numbers of interacting particles — for example, the molecules that make up the atmosphere.
- One can also think of *ecological, social, human, and commercial networks* — not to mention telephone networks, etc.

A major challenge in contemporary mathematics is to "map" these new virtual worlds — to produce visual representations that help us understand how they function. This is not easy, since these are truly enormous structures, sometimes containing billions of elements, which certainly cannot be drawn on a sheet of paper like a subway map.

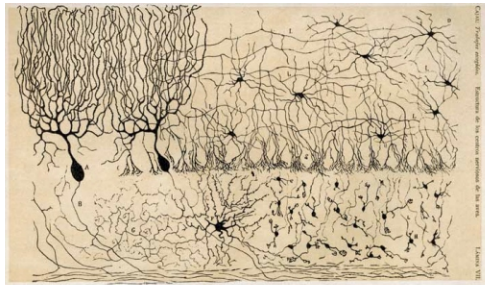
The discussion below raises more questions than it answers. I would first like to show a few examples, and then discuss a theorem that gives reason for optimism: *it may be possible to represent these immense networks in a reasonable way.*

The Brain

The brain is a mystery. It was toward the end of the nineteenth century that the neuronal dogma was accepted. The brain is not a continuous tissue but a network containing roughly ten billion neurons connected to each other by axons. Each neuron is connected to many other neurons — typically about ten thousand. How

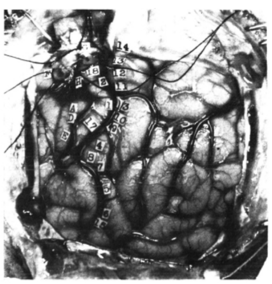
[11]Michael T. Gastner, M. E. J. Newman, *Diffusion-based method for producing density equalizing maps*, https://arxiv.org/pdf/physics/0401102v1

can we picture such a structure? Below is one of the earliest sketches, dating from 1888:

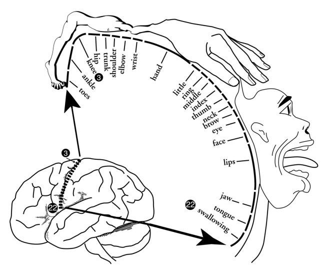


Cajal 1888
http://scienceblogs.com/neurophilosophy/2007/07/the_discovery_of_the_neuron.php

One of the first "cartographers" of the brain was Wilder Penfield. In the 1930s, during brain surgeries, he stimulated certain areas of the brain with electrodes and recorded the patient's reactions. The photograph below, taken in 1937, shows a series of numbered tags marking the locations of the electrodes.



The photograph of tags marking the location of electrodes.



The homunculus.

Each number corresponded to a reaction. For example, number 14 produced a tickling sensation from the knee down to the left foot. Number 5 produced numbness on the right side of the tongue, and so on.

This was the beginning of the *homunculus theory*, which sought to establish a map of the brain corresponding to regions of the body. The picture on the right is of the famous *homunculus*: the "map of the brain", where particular parts of the brain correspond to parts of the body.

It is a kind of map $f : X \to Y$, where the domain is the brain and the target is the human body. But how much credence should we give to such a "map"? Penfield was exploring completely uncharted territory! At the very least, one might be surprised not to see any sexual organs represented![12]

Today, this theory has lost any scientific pretension. Penfield himself readily admitted this toward the end of his life, writing: "The figurines [...] have the defects, and the virtues, of cartoons in that they are inaccurate anatomically. [...]

[12] See, however, the article by Damien Mascret, *L'homunculus retrouve son pénis. (The Homunculus Regains His Penis.)* Le Généraliste № 2340, 09.09.2005.

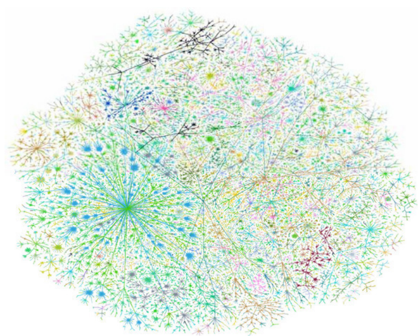
They are aids to memory, no more."[13] Wilder Penfield is regarded as one of the pioneers of modern neurosurgery.

Thanks to advances in medical imaging, it is now possible to map the brain in great detail — see, for instance, a modern brain atlas.[14] Yet it goes without saying that the fundamental problem remains, since these images, however precise, do not capture the connectivity or the functioning of the neural network.

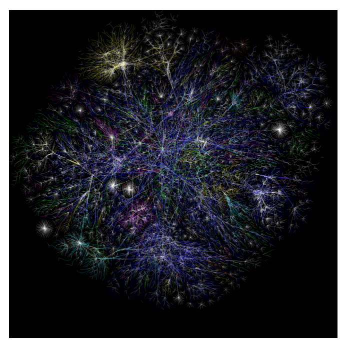
Internet

Nowadays, the number of web pages in the world is around ten billion — roughly the same order of magnitude as the number of neurons in a human brain. However, an average web page is linked to only about twenty others, which means that the Internet is less "connected" than a brain. On the other hand, the electrical or chemical impulses that travel through axons are much slower than Internet communications.

How can we get an idea of the structure of such a gigantic network? Below are the images produced in 1998 and 2005, reminiscent of the sketch of the neural network from 1888 that we saw earlier. In this drawings, the points represent web pages, and the lines connect pages that reference each other through hyperlinks.



The image of internet from 1998.



The image computed in 2005, from the Opte Project

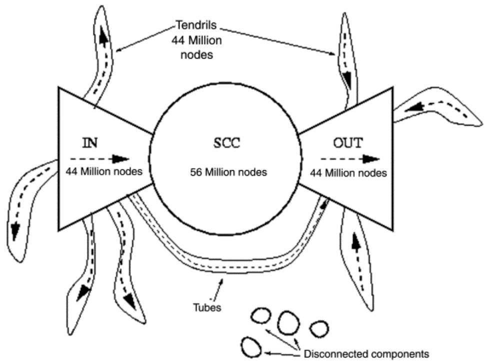
Just as with a detailed image of the brain, these pictures don't help much when it comes to understanding the global functioning of the system. In essence, they are analogous to how one sees the brain under a microscope.

The first "web explorers" proposed around the year 2000 an approximate map called the "Bow Tie":[15]

[13] Penfield, Wilder, and Herbert Jasper. *Epilepsy and the functional anatomy of the human brain.* (1954), pp. 105-106.

[14] http://www.med.harvard.edu/AANLIB/home.html

[15] Broder, Andrei, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener. *Graph structure in the web.* Computer networks 33, no. 1-6 (2000): 309-320. https://www.sciencedirect.com/science/article/abs/pii/S1389128600000839



This map remains relevant even though the numbers have changed. While one should not expect great precision, it does tell us something about the structure of the Web, it can be divided into four main regions, roughly of equal size:

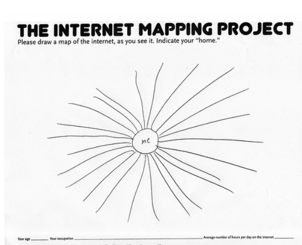
- The **S**trongly **C**onnected **C**ore, consisting of a block of pages that are highly interconnected. This is the part where connections are fully active — in every direction.
- The **OUT** zone, made up of pages that are heavily cited by the core but that, in turn, cite very few others. These are, in a sense, the pages that everyone is interested in, but that are interested in no one.
- The **IN** zone, consisting of pages that cite the core frequently but are rarely cited themselves. These are the pages that are interested in everything but that interest no one.
- The **Disconnected zone**, made up of pages that are not, or are only slightly, connected to the main core — for example, small communities with very specific interests that seek no external contact.

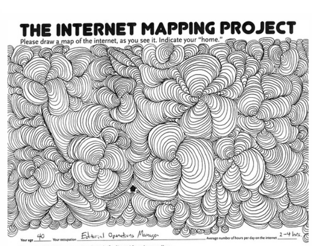
Even though it is simplistic, still this "map" is meaningful. How can we go further?

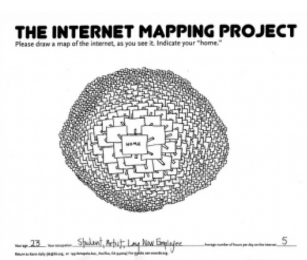
There is a vast and rapidly growing body of literature on this subject, attracting the interest of computer scientists, physicists, mathematicians, as well as psychologists and sociologists.

Faced with a structure we do not yet fully understand — and which nevertheless plays an increasingly important role in our society — it is fascinating to ask how we imagine the Web. The Internet Mapping Project[16] invited many people to draw their personal vision of the Web and to indicate their own position within it. The result is about a hundred highly instructive sketches.

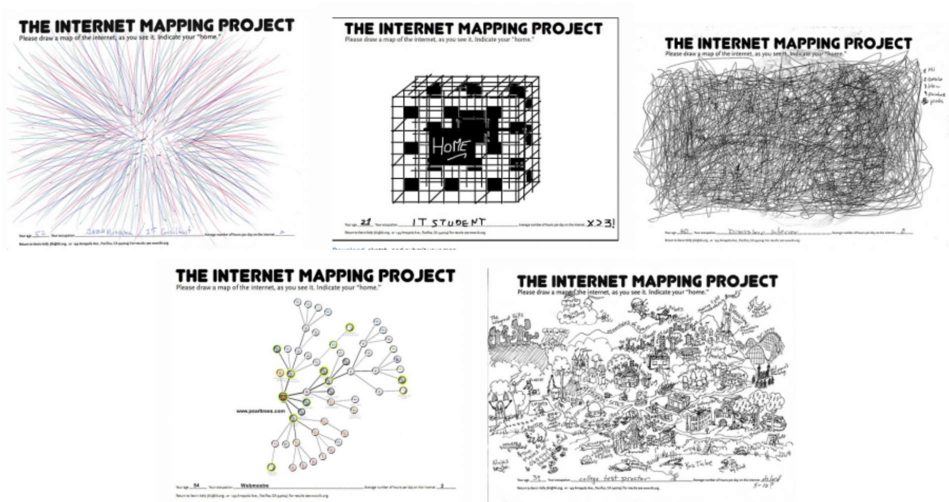
Here are a few of them:







[16] https://cheswick.com/ches/map/



The website Internet Geography: A Collection of Ways to Visually Organize and Explore the Internet[17] is well worth a visit.

The Sciences

The various sciences, of course, interact. But how is the network of sciences structured?

As in our previous examples, the first "explorers" had to make do with basic and simple facts. In his *Cours de philosophie positive*[m] (published between 1830 and 1842), Auguste Comte arranged the sciences from the most general to the most particular, from the most abstract to the most concrete: mathematics, astronomy, physics, chemistry, biology, sociology. Naturally, at the very top of Comte's pyramid stood... mathematics!

Today, the interactions between disciplines are more complex than ever. Many scientists are interested in mapping the relations between different branches of the sciences. Below is a map of the scientific world, taken from Ranking and Mapping Scientific Knowledge.[18]

The mathematician may perhaps regret no longer occupying the top of the pyramid.

What is the meaning of such a map?

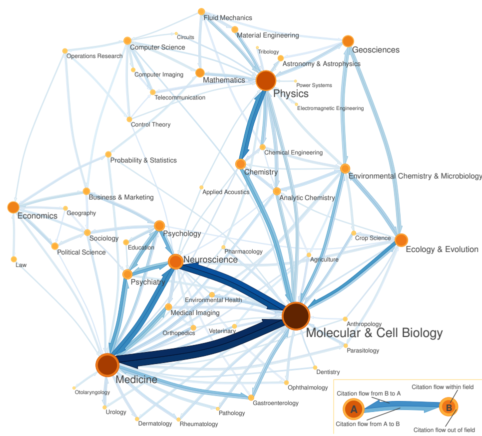
Here is the method used — highly debatable, of course — but can we really do better?

The starting point is a database that lists a large number of scientific journals (around 7 000), and that makes it possible, in particular, to count how many articles, published in one journal, cite, in their bibliographies, articles from another. The database contains about 60 000 000 citations from the past decade.

We can view this data as a large table with 7 000 rows and 7 000 columns. In the i^{th} row and j^{th} column, we record the number $N(i, j)$ of articles published by

[17] https://internetgeography.blogspot.com/
[18] http://www.eigenfactor.org/projects/mappingScience/

[m] English translation by Harriet Martineau is titled *The Positive Philosophy of Auguste Comte*. (Ed.)



journal i that are cited in articles published in journal j. Alternatively, we can think of each journal i as a vector in a 7 000-dimensional space, whose coordinates are $\big(N(i,1), N(i,2), \ldots, N(i,7000)\big)$.

The first idea is to take two journals, i and j. They are considered to be thematically close if the angle between their corresponding vectors is small. Indeed, if two vectors point in almost the same direction, it means they assign similar relative importance to all other journals — in other words, they are interested in roughly the same topics.

Next comes a clustering operation — that is, grouping together journals that are very close to one another. There are many ways to do this, and they do not all yield the same result.

Once we have formed the clusters — the groups of journals — we interpret each cluster as representing a “science.” To determine the name of each science automatically, the computer selects a keyword that appears frequently in the titles of the journals within the cluster. In this particular study, the computer identified 88 “sciences.”

Once these “sciences” are identified, we proceed as if all journals within the same group formed a single entity. We can then count the citations between sciences, yielding a new table with only 88 rows and 88 columns. By computing the angles between the corresponding vectors, we can determine a “distance” between two sciences.

Finally, the goal is to represent these 88 sciences as 88 points in the plane, in such a way that the distances between them correspond as closely as possible to the computed angles. This can be done, for example, by a minimization method such as that described in the first part of the article — though, again, many other approaches could be used. The result is the map shown above.[n]

I leave it to the reader to judge the interest and relevance of such constructions. But let us not forget the lessons of Penfield’s homunculus or the Web’s bow tie: sometimes, simple ideas can produce good images that are far better than having none at all. However we should not forget certain scientific errors of the past — such

[n]The thickness of the graph edges indicates citation traffic between the “sciences”, with only significant links shown. See the original publication for more detailed explanation. (Ed.)

as phrenology, which claimed that the bumps on a human skull revealed character traits.

Many articles are currently being published on this type of question, each proposing its own methods — often debatable ones.

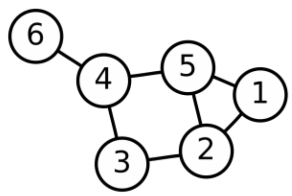
Mapping the scientific world, the Internet, or the brain is certainly not an easy task. Nevertheless, in at least two significant cases — the Internet and the world of science — we are able to draw on a sheet of paper a sketch that shows, in broad strokes, how the parts of such a network communicate with one another.

I would now like to state a theorem that highlights a regularity property common to all networks, regardless of their size — one that allows us to hope for a "reasonable" representation.

Networks and Graphs

Mathematically speaking, the networks that concern us here are called graphs. A graph consists of a set of objects called vertices (or nodes), which, depending on the context, may represent neurons, web pages, scientific journals, or any other entities. Most importantly, two vertices can be connected by an edge, which symbolizes — depending on the situation — a neural connection, a hyperlink between two web pages, a citation from one journal to another, or a "friendship" link on Facebook, and so on.

When the number of vertices and edges is small, we can easily draw such a graph on a sheet of paper.[o] For example, here is a graph with 6 vertices and 7 edges:



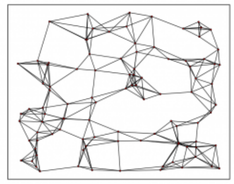
Graph theory is a fairly old branch of mathematics, but in recent years it has faced new challenges. The focus now lies on graphs whose number of vertices is so big that it defies imagination. Earlier, we saw some attempts to draw representations of the Web graph, with its billions of vertices.

The Web graph can be treated as a physical object of study, one that can be measured experimentally — for example, assessing the speed at which information circulates, somewhat like an electrical engineer might study the flow of electricity through a network. But the Web graph can also be viewed as a mathematical object.

One of the first questions raised about it concerned its randomness. Suppose we take a very large number of points — say, ten billion — and, for each pair of points, connect them with an edge at random, with a probability of 2 out of a billion. We repeat this independently for every pair of points. In this way, we create a graph with roughly the same number of vertices and edges as the Web graph.

Here is what the result of such a random construction looks like with 100 points and a probability of connecting two points equal to 5/100:

[o]Even relatively small graphs, e.g. with five vertices, may not fit on a paper in an embedded way, that is, some edges may have to intersect in their interiors. (Ed.)



This is what is known as an Erdős–Rényi random graph, named after the mathematicians who developed the theory of such structures in the late 1950s — long before the Internet existed, and with entirely different motivations.

But will such random graph with appropriate number of vertices and edges resemble the Web graph? Can the connections that arise in the Web be attributed purely to chance? The answer is no. The Web graph is constructed according to another kind of randomness, that is its own, and one that we are only beginning to model accurately.

Let us now set these developments aside and turn to Szemerédi's theorem. For further information about the Web graph (at an undergraduate level), I recommend Bonato's book.[19]

Szemerédi's Theorem

This theorem was proven in the 1970s — again, with motivations quite different from those we have in mind here.[20]

We wish to say something meaningful about a graph that has a very large number of vertices — too large for us to consider them all individually. In fact, the graphs we have in mind are so vast that it's not even clear they are well-defined objects. And even if they were, they are not static: for example, new connections appear and disappear constantly on the Web, and our neurons, unfortunately, slowly degrade and deactivate over time...

So we find ourselves in a situation where the object we wish to describe is only approximately defined and largely inaccessible. We can therefore hope for no more than an approximate understanding, say, within 1%.

Let G denote an arbitrary graph, and let me try to explain what it could mean to have a "reasonably sized understanding to within 1%."

Consider two subsets A and B of the vertex set of G, which we may imagine to be disjoint. We want to study how A and B are connected.

Let $X \subset A$ and $Y \subset B$. We count the number of edges connecting a vertex in X to a vertex in Y, and denote this number by $N(X, Y)$. If $N(X)$ and $N(Y)$ are the numbers of vertices in X and Y, respectively, we can think of the ratio

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

as representing the connection probability between X and Y.

---

[19]Bonato, Anthony. *A course on the web graph.* Vol. 89. American Mathematical Soc., 2008.

[20]A recurring theme in the development of mathematics: the same idea can be useful multiple times, often in unexpected ways.

A priori, this probability depends on the specific choice of X and Y. We shall say that A and B are *well connected* (to within 1%) if this probability does not depend too much on the choice of subsets $X \subset A$ and $Y \subset B$.[p]

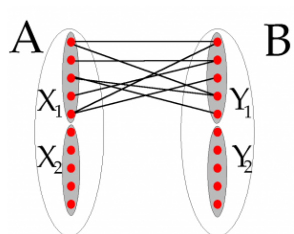
More precisely, we require that there exists a number $p_{A,B} \in [0,1]$ such that, for all $X \subset A$ and $Y \subset B$ one has

$$0.99 p_{A,B} \le \frac{N(X,Y)}{N(X) \cdot N(Y)} \le 1.01 p_{A,B}$$

However, this definition doesn't quite work. If X and Y each contain only one vertex, then the ratio $\frac{N(X,Y)}{N(X) \cdot N(Y)}$ is either 0 or 1, which can hardly be close to $p_{A,B}$.

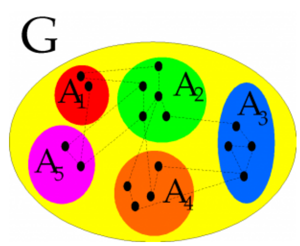
So, we refine the definition: the inequality above should hold only when X and Y are large enough — say, when each contains at least 1% of the vertices of A and B, respectively.

That gives us a definition of what it means for A and B to be "well connected (within 1%)."



Here is an example where A and B are not well connected. You can see that X_1, X_2 and Y_1, Y_2 each contain half of the vertices of A and B, respectively. But there are no connections between X_2 and Y_2, while there are many between X_1 and Y_1. For A and B to be well connected, there should be roughly the same number of connections in both cases.

We will need yet another definition. We say that we have a partition of a graph when we have divided its vertex set into subsets $A_1, A_2, \ldots, A_n$, such that every vertex of the graph belongs to exactly one of them.



We encountered such a partition earlier when grouping scientific journals into "clusters" that we called "sciences." An example of a partition of a graph into 5 parts containing 2, 5, 4, 4, and 2 vertices each is pictured above.

A partition is called an equipartition if each A_i contains the same number of vertices. Of course, if we want an equipartition of a graph G into n parts, then n must

[p]Well-connectedness essentially says, that the bipartite subgraph spanned by the sets of vertices A and B has no easily detectable structure and it looks like a typical random graph at a first glance.

divide the total number $N(G)$ vertices. In our enormous graphs, it's completely unrealistic to expect that $N(G)$ will be divisible by 3 or 157!

So we relax the condition slightly. We say that the n parts A_i form an almost-equipartition if the number of vertices in each A_i differs from $N(G)/n$ by at most 1:

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

Naturally, the smaller the number n of parts, the more useful the partition is — at least compared to the astronomical number of vertices in the original graph (which we can barely imagine!). So let's take a reasonable number, say, $n = 1000$.

We are looking for almost-equipartitions that are well connected — meaning that the parts A_i and A_j well connected. All of them? No, that would be too much to ask. "Almost all" will suffice. We require that among the $n^2 = 1\,000\,000$ possible pairs (i, j), at least $0.99n^2$ (that is, $990\,000$) of them are well connected.[q]

Szemerédi's Theorem (informal statement)

Now we can state the theorem — but first, a warning: the following statement may not be exactly correct! We'll need to refine it.

Theorem (Szemerédi)**.** *For any graph, no matter how large, it is always possible to find an almost-equipartition into* 1000 *parts that are well connected.*

Szemerédi's Theorem[r]

There are several versions of Szemerédi's Theorem, which is also known in the literature as Szemerédi's Regularity Lemma. One of the statements is given below. To make it simpler we first introduce another definition. For a fixed precision level $\varepsilon > 0$ the *ε-regular partition* of a graph G is a partition of the vertex set of G into k parts $(A_1, \ldots, A_k)$, such that among the pairs of distinct A_i's there are at most εk^2 "bad" pairs, about which we can not say anything. The rest of the pairs of distinct A_i's are ε-well-connected. One may or may not require the partition to be an almost equipartition. The theorem holds in both setups, however the known bounds on the size of the partition will be different.

Now we are ready to state the theorem. For any precision level ε (1% in the text above) the theorem provides the upper bound M on the size of the ε-regular partition, for any sufficiently large graph.

Theorem (Szemerédi's Regularity Lemma)**.** *For any $\varepsilon > 0$ (precision level) there exist $M > 0$ (maximal number of parts) such that for every graph G with at least M vertices there exist an ε-regular partition of G with the number of parts not exceeding M.*

The known bounds on M are just astounding,[s] they are in the form of exponential towers (repeated exponentiation) of the height proportional to ε^{-2}. There is no saying, how large this number for $\varepsilon = 0.01$ is, there is nothing in the world to compare it to. It has also been shown that there are families of graphs for which the required number of parts is indeed about that big, so that the bound can not

[q]Stricktly speaking, well-connectedness was only defined for pairs of disjoint subsets of vertices. We should imagine, that the pairs (A_i, A_i) are among those that we allow to be "bad". (Ed.)

[r]This section was added by the translator. (Ed.)

[s]W.T. Gowers, *Quasirandomness, counting and regularity for 3-uniform hypergraphs*, Combin. Probab. Comput. 15 (2006), 143–184.

be improved by much. We may still hope that such synthetic examples are rare and there are large classes of graphs for which Szemerédi Regularity Lemma with some reasonable M holds.

The Significance of the Theorem and a Few Comments

Once a well-connected, almost-equipartition is known, we can draw a "reasonable" representation of the graph we are studying.

We imagine drawing 1 000 points on a plane, each representing one of the A_i, which we can think of as a "cluster," a "science," or perhaps an "organ" of the network — each labeled with its corresponding number i. Then, along the segment joining the points i and j, we write the number $P_{A_i A_j}$.

This simple diagram serves as a genuine map of the graph to within 1% accuracy! And this remains true even if the original graph contains billions upon billions of vertices. Of course, such a map does not allow us to reconstruct the entire graph — but it does tell us, to within 1%, the number of connections between any two sufficiently large parts X and Y. It suffices to know how X and Y intersect the "clusters" A_i

In short, anyone interested in the global functioning of the graph can be quite satisfied with such a map. If we wish, we can even position the 1 000 points on the plane in such a way that the distances between them reflect, as closely as possible, the degree of connectivity: the larger $p_{A_i A_j}$ is, the closer the corresponding points will appear.

Two caveats (that somewhat spoil the theorem's magic)

I have saved two comments for the end because they somewhat diminish the theorem's appeal — though they certainly do not destroy the desire to improve it, and to overcome the two shortcomings I will now highlight.

First, the theorem as stated above is not quite correct. Unfortunately, I cannot guarantee a partition with exactly 1 000 clusters, as I somewhat boldly claimed earlier. The truth is that for each desired precision (say, 1% in our example), there exists some integer M that makes the theorem valid.

What is the actual value of M for a 1% precision? Is it 1 000, as I suggested? The important point is that this M depends only on the chosen precision — not on the size of the graph itself. If it depended on the size, the theorem would lose all appeal.

I would not be surprised if the reader felt frustrated by such a statement — "an integer M exists, but we don't know it." That's how mathematics often is: we do what we can!

Now, the hunt is on to determine the best possible M for a given precision. Researchers are striving to make the theorem effective — to obtain actual, reasonable values of M for realistic levels of precision at least for some large classes of graphs or a slightly modified conditions imposed on the partition. The theorem is flexible, and the goal is to make it usable.

Finding the partition itself

Along similar lines, it is not enough to assert that there exists a partition with, say, 1 000 clusters. We must also be able to find it if we wish to draw a map!

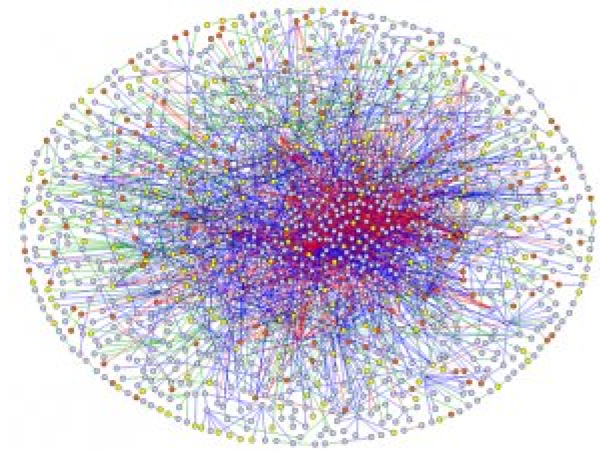
This, too, is an area of active research: developing efficient (and not overly time-consuming) computational methods to actually find Szemerédi's clusters.

Readers with a solid mathematical background may enjoy reading an engaging preprint on "Very Large Graphs",[21] which delves deeper into these questions.

Other Networks...

To conclude, here is another network we would like to understand better. It is a graph describing 3 200 interactions among 1 700 human proteins — a work that earned its authors a scientific award in 2005.

This graph is rather complicated! One would certainly like to apply Szemerédi's theorem to it...



[21] Lovász, László. *Very large graphs.* In Current developments in mathematics, 2008, vol. 2008, pp. 67-129. International Press of Boston, 2009. https://arxiv.org/pdf/0902.0132