

# Big data is being reshaped thanks to 100-year-old ideas about geometry

February 20 2019, by Ittay Weiss

---



Credit: Hassan Ali from Pexels

Your brain is made up of billions of neurons connected by trillions of synapses. And how they're arranged gives rise to the brain's functionality and to your personality. That's why [scientists in Switzerland](#) recently

produced the first-ever digital [3-D brain cell atlas](#), a complete mapping of the brain of a mouse. While this is a colossal achievement, the great challenge now lies in learning to decipher the atlas. And it's a huge one.

Science is full of this kind of problem: how to turn large amounts of information into useful insight. For many years, researchers relied on mathematics and statistics to explore data. The explosion of large datasets created by digital storage, the internet, and cheap sensors has led to the development of new techniques designed specifically to deal with this "big data".

And now there is an emerging new approach based on century-old ideas that's producing superior tools for understanding certain types of big data. Using the mouse's brain as an example, its physical [shape](#) determines its functionality. But a precise description of this shape, which we now have, doesn't automatically reveal everything about how the brain works.

Behind the physical shape lies a more abstract shape formed by the interconnections within the brain. Capturing aspects of this shape by applying techniques from the study of what's known as "topology" can help reveal a deeper understanding of the brain's functioning. This same guiding principle of using topological techniques on [big data](#) also has applications in [drug development](#) and other cutting-edge endeavours.

## Topology

Topology is a branch of modern geometry [with roots](#) going back to a foundational observation by the Swiss mathematician Leonhard Euler (1707-1783) about polyhedra, 3-D shapes with flat faces, straight edges and sharp corners or "vertices". In 1750, [Euler discovered](#) that for any convex (with all its faces pointing outwards) polyhedron, the number of vertices minus the number of edges plus the number of faces [always](#)

[equals two.](#)

You can apply the same formula to other shapes to get what is known as their Euler characteristic. This number doesn't change no matter how the shape is bent or deformed. And [topology](#) is the study of these kind of constant properties of shapes.

Topology went through rapid development during the 20th century as a prominent subject in pure mathematics. The researchers who created the subject didn't have real-world applications on their minds, they were just interested in what was mathematically true about shapes under certain conditions.

Yet some of these ideas from topology that have been around for over 100 years are now finding significant applications in [data science](#). Because topology focuses on constant properties, its techniques make it insensitive to various data inaccuracies or "noise". This makes it ideal for deciphering the true meaning behind the collected data.



Credit: AI-generated image ([disclaimer](#))

You are probably familiar with a common topological phenomenon. Wires placed neatly in your bag in the morning (your earphones or an adapter) have a tendency to produce a horrible mess by midday. A wire is a very simple shape. Whether or not it is knotted is a topological question, and the tendency to arrive at a topological nightmare in your bag is now [quite well understood](#).

Millions of years ago, evolution was confronted with a similar problem. [DNA](#) in cells is a molecule composed of two coiled up chains. Each chain is a very long wire, built up from a sequence of small molecules called nucleobases. When a cell divides, these wires unwind, replicate and then coil up again. But just like wires in a bag, the strands of DNA can become tangled, which prevents the cell from dividing and causes it to die.

Special enzymes in the cell called [topoisomerases](#) have the task of preventing such a catastrophe. And deliberately disrupting the topoisomerases of bacteria prevents them from spreading and so stops an infection. This means that a better understanding of how topoisomerases prevent the entanglement of DNA could help us design new antibiotics. And since entanglement is a purely topological feature, [topological techniques](#) can [help us do that](#).

## Drug development

Topology can also be used to improve the creation of new drugs. Pharmaceutical drugs are chemicals designed to interact with certain

cells in the body in a particular way. Specifically, cells have receptors on them that allow molecules of a certain shape to lock onto them, altering the behaviour of the cells. So producing drugs with these shaped molecules enables them to target and affect the right cells.

As it turns out, manufacturing a molecule to have a particular shape is a rather simple process. But the easiest way to get the drug to the target [cells](#) is to send them via the bloodstream, and for that, the drug must be water soluble. After a drug with a correct shape is produced, the million pound question is: does it dissolve in water? Unfortunately, this is a very difficult question to answer just from knowing the chemical structure of the molecule. Many [drug](#) discovery projects fail because of solubility issues.

This is where topology comes in. "Molecule space" refers to a way of thinking about an entire collection of molecules as a kind of mathematical entity that can be studied geometrically. Having a map of this space would be a tremendous tool for producing new drugs, particularly if the map included landmarks indicating higher chances of solubility.

In [recent work](#), researchers used topological data analysis tools as a first step to producing such a map. Analysing vast amounts of data linking molecule properties to water solubility, the new approach led to the discovery of new, previously unsuspected, indicators of solubility. This improved ability to produce water-soluble drugs has the potential to significantly shorten the time it takes to create a new treatment, and to make the whole process cheaper.

In more and more realms of science, researchers are finding themselves with more data than they can effectively make sense of. The response of modern mathematicians to meet the [mathematical challenges of big data](#) is still unfolding – and [topology](#), a theory bound only by the imagination

of its practitioners, is bound to help shape the future.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation

Citation: Big data is being reshaped thanks to 100-year-old ideas about geometry (2019, February 20) retrieved 19 April 2024 from <https://phys.org/news/2019-02-big-reshaped-year-old-ideas-geometry.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.