Topological Data Analysis in the Biomedical Research: an Epistemological Study

Andrei Rodin

philomatica.org

8 december 2025

Voevodsky's Unfinished Project (*Bridging the Gap*, Palma de Mallorca, 21-23 May, 2024)

Could recently developed mathematics serve as the all-purpose language for science and engineering like PDEs of the 19th c.?

BioSystems 204 (2021) 104391

Contents lists available at ScienceDirect

BioSystems

journal homepage: http://www.elsevier.com/locate/biosystems

Voevodsky's unfinished project: Filling the gap between pure and applied mathematics

Andrei Rodin

Institute of Philosophy, Russian Academy of Sciences and Saint-Petersburg State University, Russia

Voevodsky's Unfinished Project (*Bridging the Gap*, Palma de Mallorca, 21-23 May, 2024)

Could recently developed mathematics serve as the all-purpose language for science and engineering like PDEs of the 19th c.? Yes, perhaps via Information Technologies (link to the Univalent Foundations).

BioSystems 204 (2021) 104391

Contents lists available at ScienceDirect

BioSystems

journal homepage: http://www.elsevier.com/locate/biosystems

ELSEVIER

Voevodsky's unfinished project: Filling the gap between pure and applied mathematics

Andrei Rodin

Institute of Philosophy, Russian Academy of Sciences and Saint-Petersburg State University, Russia

Topological Data Analysis

Perhaps TDA is such an example? (After 2017).

Topological Data Analysis

Perhaps TDA is such an example? (After 2017).

... even if it does not apply quite in the same way as do PDEs.

Topological Data Analysis

Perhaps TDA is such an example? (After 2017).

... even if it does not apply quite in the same way as do PDEs.

Does TDA help to build *models* of natural phenomena?

End of Theory?

WIRED

<< Back to Article

WIRED MAGAZINE: 16.07

The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

By Chris Anderson 06.23.08

Chris Anderson 2008

The new availablility of huge amounts of data, along with statistical tools to crunch these numbers, offers a whole new way of understanding the world. Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanical explanation at all. There is no reason to cling to our old ways.

Chris Anderson 2008

The new availablility of huge amounts of data, along with statistical tools to crunch these numbers, offers a whole new way of understanding the world. Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanical explanation at all. There is no reason to cling to our old ways.

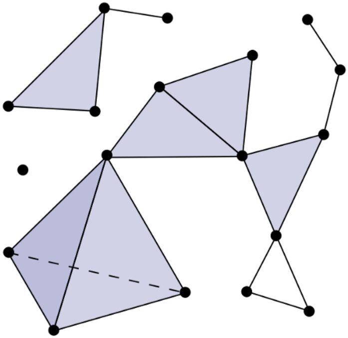
It's time to ask: What can science learn from Google?

Yu.I Manin 2013

Think! Otherwise no Google will help you.

Homological Algebra

Simplicial complexes are generalized polyhedra:

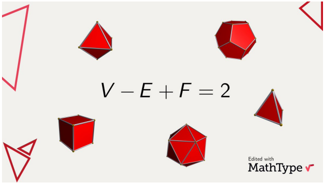


Betti Numbers

- b_0 is the number of connected components;
- b_1 is the number of 1-dimensional “circular” holes;
- b_2 is the number of 2-dimensional holes aka cavities;
- $\ldots$
- b_k is the number of k-dimensional holes.

Euler Characteristic

$$\chi = b_0 - b_1 + b_2 - b_3 + \dots$$

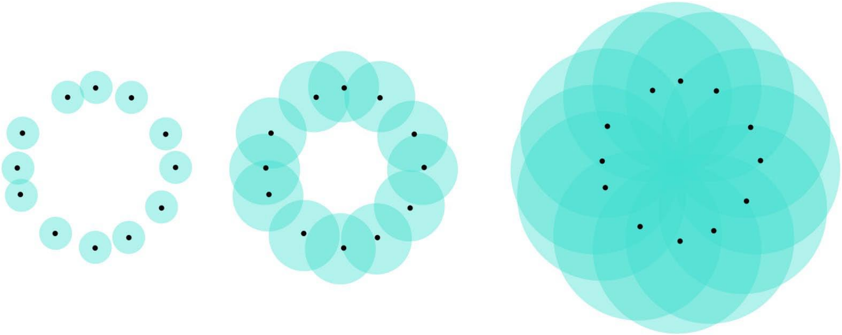


$$\chi(\text{tetrahedron}) = V - E + F = 4 - 6 + 4 =$$

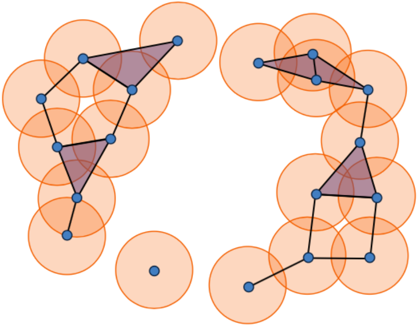
$$= b_0 - b_1 + b_2 = 1 - 0 + 1 = 2$$

Persistent homology : Computational Topology + Morse theory + CS Applications

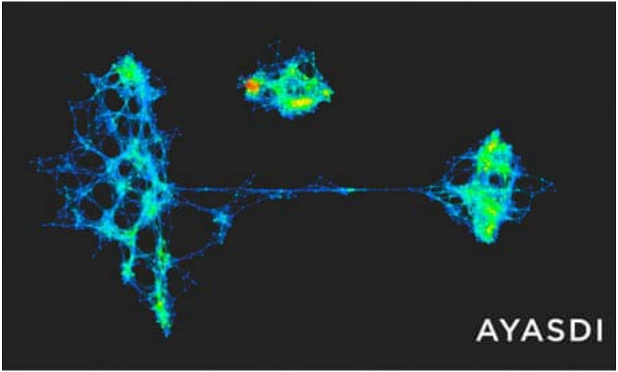
Edelsbrunner, Letscher & Zomorodian 2002



Persistent homology



Topological profile of a dataset



What a topological profile says about the data?

Nobody, to the best of my knowledge, has an exact answer, however it is clear that this is quite a lot!

End of Theory?

Revival of the old Empiricism vs. Rationalism debate?

Euler 1750-1758

How it happened that nobody (with minimal historical reservations) discovered Euler Characteristic of polyhedra before Euler simply by counting vertexes, edges and faces of well-known polyhedra and then using a "numerological induction"?

Euler 1750-1758

How it happened that nobody (with minimal historical reservations) discovered Euler Characteristic of polyhedra before Euler simply by counting vertexes, edges and faces of well-known polyhedra and then using a "numerological induction"?

Was it a mere lack of attention to available data and an excessive focus on higher abstractions that postponed this discovery until the 18th century?

Euler 1750-1758

How it happened that nobody (with minimal historical reservations) discovered Euler Characteristic of polyhedra before Euler simply by counting vertexes, edges and faces of well-known polyhedra and then using a "numerological induction"?

Was it a mere lack of attention to available data and an excessive focus on higher abstractions that postponed this discovery until the 18th century?

Could this discovery (at least in a form of observation without supporting theorems) be possibly made by using a brute computational force (Machine Learning)?

Probably not: cf. "acies vocabo"

A part of the answer is that the concept of *edge* (of a polyhedron) did not exist before Euler introduced it. So the lack of appropriate conceptual optics did not allow this to happen before Euler.

§. 6. Ad ambitum autem cuiusque ſolidi figuris planis incluſi pertinent 1^{mo} ipſae figurae planae eius ambitum conſtituentes, quae hedrae vocantur; 2^{do} binarum hedrarum ſecundum latera concurſus, quibus termini lineares ſolidi oriuntur: hos terminos, quoniam apud ſcriptores Stereometriae nullum nomen proprium reperio, acies vocabo; 3^{tio} puncta, in quibus tres pluresue hedrae concurrunt, quae puncta anguli ſolidi appellantur.

Moral:

There is no literally “raw data”. There is no “pure and innocent” Data Analysis. Google has its algorithms. Data Analysis is fully corrupted by human conceptual thinking at many different levels.

.

Moral:

There is no literally “raw data”. There is no “pure and innocent” Data Analysis. Google has its algorithms. Data Analysis is fully corrupted by human conceptual thinking at many different levels. But Anderson can be nevertheless quite right that the success of Google algorithms can teach some interesting lessons to scientists and to philosophers of science..

TDA and the “brute computational force”

Concept-based TDA combined with Machine Learning is more effective than the “brute” ML (whatever it may mean) alone!

Robert May 2004 (contra Anderson)

“A paradigmatic account of the uses of mathematics in the natural sciences comes, in deliberately oversimplified fashion, from the classic sequence of Brahe, Kepler, Newton: observed facts, patterns that give coherence to the observations, fundamental laws that explain the patterns [. . .].
"“Consider the role played by applications of mathematics in

sequencing the human and other genomes [. . .]. The sequence information, however, represents only the Tycho Brahe stage. ”

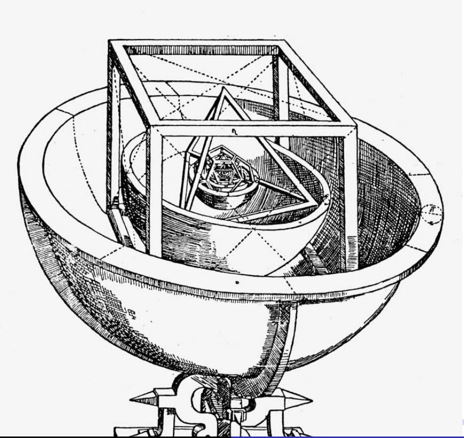
“Current work on various genomes uses pattern-seeking programs to sort out coding sequences corresponding to individual genes [. . .]. Again, elegant and sometimes novel mathematics is involved in this Keplerian stage of the work in progress. ”

Robert May 2004

"We are only just beginning, if that, the Newtonian stage of addressing the deeper evolutionary questions posed by these patterns."

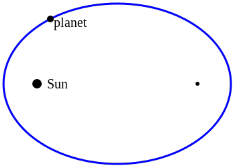
Kepler 1596

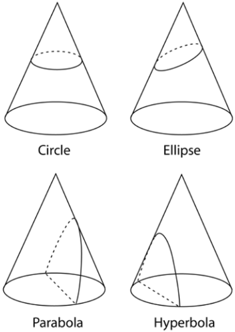
Mysterium Cosmographicum:



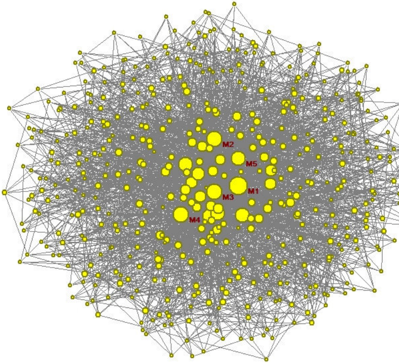
Kepler 1607

Astronomia nova Aitiologitis 1607:





the old and the new image of the stock market



Bare patter recognition / visualisation: (after the review by Skaf and Laubenbacher 2022)

(after the review by Skaf and Laubenbacher 2022)

1. the identification of a new type breast cancer tumour earlier undetected by other methods (Nicolau:2011);
2. the identification of two new subtypes of asthma (Kuo:2017);
3. the visualisation of high-dimensional immunological data pertaining to co-infection of mice with influenza and pneumococcus, allowing for characterisation of distinct immune responses in various infection scenarios and stages within immune responses to co-infection (Sasaki:2020);

Basis of classification (after the review by Skaf and Laubenbacher 2022):

4. a highly accurate classification of subtypes of stomach lesions (Dunaeva:2016);
5. a classification of type 2 diabetes mellitus (T2DM) disease trajectories (Dagliati:2020);

Functional features (after the review by Skaf and Laubenbacher 2022):

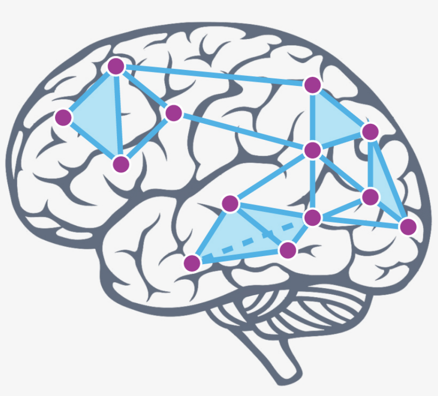
6. using molecular topological fingerprints (MTFs) for protein identification, classification, quantitative analysis of rigidity, and evolution of structural features during folding (Xia and Wei:2014);
7. the identification of distinct molecular states in the RNA hairpin folding process, providing insight into its mechanism (Bowman:2008);

Full-fledged mathematical modelling (after the review by Skaf and Laubenbacher 2022):

8. a study of the spread of contagions across a network from geometric and topological perspectives (Taylor:2015).

Brain Research (Simplicial Connectome)

A paradigm for new science?



Conclusion 1

TDA is a rear example of successful application of recently developed abstract mathematical concepts in a large variety of empirical studies;

Conclusion 1

TDA is a rear example of successful application of recently developed abstract mathematical concepts in a large variety of empirical studies;

In the majority of cases TDA at the "Keplerian stage" (May) but there are also case of genuine theorising and modelling. (Contra May): there is no reason to devaluate that stage and hope that it will be left wholly behind at some point of scientific development. Science not only continually progresses but also continually revises its foundations.

Conclusion 1 (continued)

Scientific breakthroughs may occur (and do occur) both in highly selective theory-based experimentation (the case of LIGO 1915-2015) and in a quasi-blind empirical research and basic pattern recognition.

Conclusion 1 (continued)

Scientific breakthroughs may occur (and do occur) both in highly selective theory-based experimentation (the case of LIGO 1915-2015) and in a quasi-blind empirical research and basic pattern recognition.

In order to emphasise the importance of data-driven science one does not to claim that theories and models are obsolete. This is just the other side of the same erroneous thinking about science and its progress.

Conclusions 2 (Paraphrasing Immanuel Kant)

Berechnungen (und Datenanalyse) ohne Begriffe sind blind (und langweilig ☹), Begriffe ohne Berechnungen sind steril.

Conclusions 2 (Paraphrasing Immanuel Kant)

Berechnungen (und Datenanalyse) ohne Begriffe sind blind (und langweilig ☹), Begriffe ohne Berechnungen sind steril.

But there is no universal receipt how to combine the two. TDA is one promising attempt among many others ...

Thank you for you attention!