The Evolution Of Mathematics Over The Last Centuries

SD201 - Final Project

Erwan Fagnou

SD201
Télécom Paris
France
December 12, 2021

Github: https://github.com/ErwanFagnou/history_of_maths

Contents

Introduction

> *Mathematics is a vast adventure; its history reflects some of the noblest thoughts of countless generations.*
>
> Dirk Jan Struik

In this project, I tried a data-driven analysis of mathematics, in order to understand its structure and history. What have been the trends in mathematics? Is there a way to determine the most important discoveries, or the most active fields during the last centuries? To answer these questions, I collected data to create my own dataset, which contains information about thousands of mathematical concepts and mathematicians related to them. Even if this dataset was actually quite incomplete, it still allowed me to produce very satisfying and interesting results.

Chapter 1

The dataset

1.1 Wolfram MathWorld

The Wolfram MathWorld website is a huge database of mathematical discoveries (concepts, theorems, conjectures...). It is "the web's most extensive mathematical resource", and it "has been assembled over more than a decade by Eric W. Weisstein with assistance from thousands of contributors".

I chose this website to make my own dataset for several reasons:

- I did not find any dataset that contained what I was looking for.
- Wolfram MathWorld is a specialized and extremely complete encyclopedia.
- Other websites, like Wikipedia, may contain the same information, but it would be quite hard to determine which article is related to maths and which isn't.

1.2 Collecting the data

As the website does not provide the dataset nor tools to obtain its data, I used scraping. As written in the terms of use, I have been careful not to cause any bandwidth degradation. The features I extracted are:

- The name of the mathematical statement.
- The fields it belongs to (the fields have a tree-like structure, for example the Pythagorean theorem is in "Geometry / Plane Geometry / Triangles / Triangle Properties").
- The links to other similar MathWorld articles (from the "see also" section).
- The mathematicians who published about it, along with the year of the publication (from the "references" section).

Note that we do not have the name of the mathematician(s) who made the discovery. However we can find them for recent works, by taking the author(s) of the first publication in the references.

The data obtained is very clean, the major source of errors being the extraction of the author(s) and year of the references, as the formats and spellings are not always consistent.

1.3 The resulting dataset

The dataset I obtained contains:

- 13925 mathematical concepts
- 49069 directed edges between math concepts (the "see also" edges)
- 11254 mathematicians (with some unfortunate redundancy when the spelling of a name isn't consistent)
- 46027 undirected edges between math concepts and mathematicians (the "reference" edges), labeled with the year of the first publication

Chapter 2

Analyzing the data

2.1 Finding the most important mathematical discoveries

I first focused only on the graph of mathematical concepts, and ignored the mathematicians. I had to make a choice for the edges: when a concept refers to another in the "see also" section, is the edge directed? And if so, in which direction? With the PageRank algorithm in mind, I chose to orient the edges towards the referred concept.

It is then very straightforward to apply the PageRank algorithm[1] in order to get the most important mathematical concepts. Here is the top 15:

Name	**PageRank score**
Sphere	1.25e-03
Circle	1.17e-03
Prime Number	1.15e-03
Group	1.05e-03
Fourier Transform	9.61e-04
Tree	9.34e-04
Archimedean Solid	8.64e-04
Normal Distribution	8.45e-04
Integer Sequence Primes	8.30e-04
Polygon	8.25e-04
Finite Group	8.19e-04
Large Number	7.93e-04
Ring	7.73e-04
Vector	7.69e-04
Riemann Zeta Function	7.62e-04

Figure 2.1: Top 15 of the mathematical discoveries with the highest PageRank score

As expected, this list contains some of the most fundamental mathematical objects and concepts, from many different fields.

2.2 Visualizing the graph of mathematics

In order to have a better understanding of the structure of the graph, I projected it onto a 2D plan[2]. Using the previously computed PageRank scores, I made the most important nodes bigger and redder.

[1] I used the NetworkX Python library, which implements the PageRank algorithm, with the damping factor $\alpha = 0.85$.

[2] The iGraph Python library provides a lot of customization for drawing graphs. I used the default layout, which is "Dr. L", a force-directed algorithm adapted for large graphs.

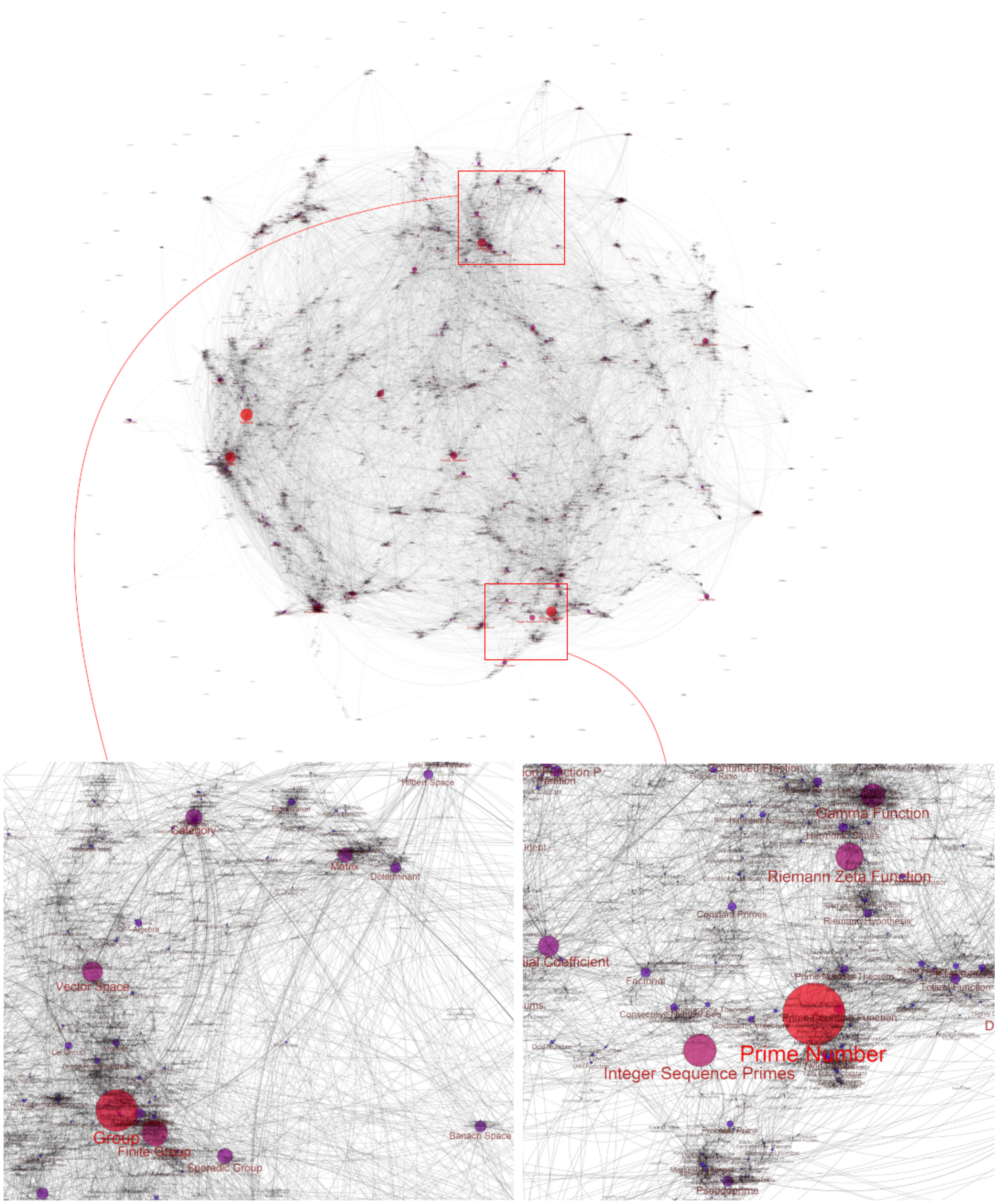

Figure 2.2: 2D projection of the mathematical discoveries graph[3], zooming on two interesting areas

This image clearly shows the presence of clusters, both at small and big scale. We can notice some interesting structures, like how the discrete and continuous mathematics interact, or the links between maths and computer science... The use of hierarchical clustering here would give some very interesting results, but this is not my goal for now, and the dataset already contains a lot of information about the categories of each mathematical concept.

[3]I uploaded this image online so that anyone can explore it and zoom anywhere, as I could not include such a large file in this document: https://www.easyzoom.com/imageaccess/4042cfe20d1a4e188ca34729b1168888

2.3 Completing the graph with mathematicians

So far, only the mathematical concepts have been used. The previous graph does not take into account the links between mathematicians and math concepts. One way of including them is to simply add mathematicians and their edges in the previous graph. The resulting graph then contains two types of nodes, and two types of edges. The issue we have to face is the weight we give to the different types of edges. Is a “see also” edge (between math discoveries) as important as a “reference” edge (between discoveries and mathematicians)? For now, I chose to give the same weight to each edge.

Before analyzing this new graph, some cleaning must be done. When looking into the mathematicians in the database, I realized that some have a huge number of irrelevant references. Neil Sloane, the creator of the OEIS (On-Line Encyclopedia of Integer Sequences), has by far the most references in the database (1552, the second being Richard Guy with “only” 511 references), because of the many references to the OEIS. I fixed this by only keeping the reference which have a year of publication, reducing Sloane’s references to 115.

Name	PageRank score
Riemann Zeta Function	7.82e-04
Global Optimization	7.45e-04
Prime Number	6.90e-04
Logistic Map	6.56e-04
Fibonacci Number	6.19e-04
Ramsey Number	6.17e-04
Pi	5.68e-04
Topology	5.50e-04
Fourier Transform	5.38e-04
Sphere	5.29e-04
Chaos	5.25e-04
Graceful Graph	5.17e-04
Knot	5.17e-04
Finite Group	5.14e-04
Circle	4.90e-04

Name	PageRank score
Guy, R	3.22e-03
Wells, D	2.54e-03
Coxeter, H	2.36e-03
Zwillinger, D	1.97e-03
Gardner, M	1.96e-03
Abramowitz, M	1.80e-03
Stegun, I	1.80e-03
Skiena, S	1.75e-03
Finch, S	1.66e-03
Arfken, G	1.62e-03
Conway, J	1.54e-03
Kimberling, C	1.51e-03
Borwein, J	1.46e-03
Wenninger, M	1.36e-03
Gray, A	1.30e-03

Figure 2.3: Top 15 of the mathematical concepts (left) and mathematicians (right) with the highest PageRank score in the resulting graph

We can already notice some changes in Figure 2.3 compared to Figure 2.1. Now, the top mathematical fields and objects are more recent, where as basic things like “Sphere” or “Group” have gone down in the rankings. This difference indicates that adding the mathematicians in the graph gave more importance to fields that interest mathematicians.

The rank of mathematicians is however more difficult to interpret. It seems greatly influenced by the number of references, but the presence of John Conway, who has only 18 references in the database (see Figure 2.4), shows that this rank measures in some way how important a mathematician has been.

Name	Number of references
Guy, R	509
Sloane, N	115
Ramanujan, S	39
Euler, L	36
Gauss, C	20
Conway, J	18
Newton, I	7
Turing, A	4

Figure 2.4: Number of references in the database (after filtering) of some noticeable mathematicians and scientists

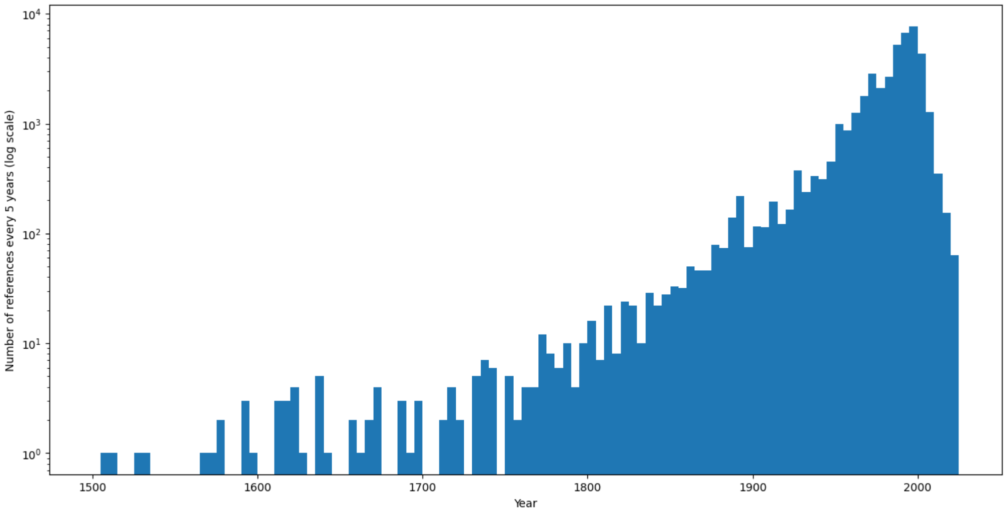

Figure 2.5: Distribution of the date of the references in the database

2.4 Finding mathematical trends over the last two centuries

Given a specific time interval, we can keep only the references that are in this interval and do as in section 2.3. This way, we have an idea of the most active mathematical fields of this period. As we can see in Figure 2.5, we can only focus on the last two centuries because of the lack of references before. I chose to use intervals of 25 years. Even if the intervals are quite large, there still is a big difference in the number of references to mathematicians between the early 19$^{\text{th}}$ century and today. In order to compensate for that, and to have coherent results, I changed the weight of the "reference" edges (the "see also" edges always have a weight equal to 1):

$$w_{\text{references}} = \frac{\text{number of "see also" edges}}{\text{number of "reference" edges}} \text{ and } w_{\text{see also}} = 1$$

The final results are shown in Figure 2.6.

Some mathematical objects or concepts have always been in the top, like "Tree", "Fourier Transform", "Group", "Prime Number"... which is not very interesting as they already appeared in Figure 2.1. However, there are some which got a high rank during a shorter and more precise period of time. Unfortunately, checking if every result is coherent is impossible to automate, so I only did it a few times for some results:

- **Knot**: The table shows that knot theory has been at its peak between 1825 and 1900. And this is indeed the case:

  > "Mathematical studies of knots began in the 19th century with Carl Friedrich Gauss, who defined the linking integral (Silver 2006). In the 1860s, Lord Kelvin's theory that atoms were knots in the aether led to Peter Guthrie Tait's creation of the first knot tables for complete classification. Tait, in 1885, published a table of knots with up to ten crossings, and what came to be known as the Tait conjectures." *Wikipedia*

- **Hamiltonian Graph**: The period during which it was in the top 10 is 1850-1875, which corresponds to the first publications about finding a Hamiltonian cycle:

  > "The Icosian game [...] is the problem of finding a Hamiltonian cycle along the edges of an dodecahedron [and] was invented in 1857 by William Rowan Hamilton." *MathWorld Wolfram*

- **Global Optimization**: This field doesn't really have a creation year, but all of the algorithms and references in MathWorld Wolfram and Wikipedia date from between the end of the 20$^{\text{th}}$ century and today, which matches perfectly with the results in Figure 2.6.

But there still are some cases which do not easily correlate with the historical facts, like the Riemann zeta function. This function was studied by Riemann who published about it in 1859. This date does not match with Figure 2.6, which places this function in the early 19$^{\text{th}}$ century. However Euler had studied the real case of this function in 1737, which suggests that there already were some works about topics close to the Riemann zeta function even before it existed.

Name	1800-25	1825-50	1850-75	1875-1900	1900-25	1925-50	1950-75	1975-2000	2000-today
Triangle	9			16	3				
Tetrahedron	8								
Tree	7	9	9	12	20	17			
Fourier Transform	6	8	18	10		18	14		9
Group	5	6	16	8	11	20			
Prime Number	4	7	13	4	1		15	20	
Circle	3	4	12	9	19				
Fibonacci Number	20				2	3	6		19
Sphere	2	2	5	5	13	13			
Archimedean Solid	18	11	14	14	9	15	16		
Vector	17	16					3		
Ring	16	18			4	1			2
Gamma Function	12			1		10			
Normal Distribution	11	10		15					
Finite Group	10	12	8	3		2			1
Riemann Zeta Function	1							4	4
Divisor Function		5			18				
Bessel Function of the First Kind		3	3						
Knot		1	1	2		5	18		12
Hamiltonian Graph			7						
Conic Section			6	7					
Matrix			4						8
Logistic Map			20					10	6
Ellipse			2	6		14			
Polynomial			11				4		
Quintic Equation			10						
Pi				13			8	11	
Cube					8				
Quaternion					7				
Graph					6				
Partition Function P					5				
Dual Polyhedron					10				
Hilbert Space						9			
Perfect Square Dissection						8			
Trigonometry						7			
Determinant						6			
Tensor						4		19	
Polyomino							9		
Limit							7		3
Problem							5	13	
Antichain							2		
Amicable Pair							10		
Topology							1	9	
Fractal								8	
Graceful Graph								7	
Ramsey Number								6	
Radon Transform								5	
Wavelet								3	
Chaos								2	
Global Optimization								1	17
Interval Arithmetic									7
Superellipse									5
Cyclic Quadrilateral									10

Figure 2.6: Evolution of the rank of mathematical concepts from 1800 to today. Only the ones that have been at least once in the top 10 are listed. An empty tile means the rank was above 20 during that period.

Conclusion

The results are really promising, as they generally match quite well with the history of mathematics. The dataset is however far from being perfect, as we do not have much information about what happened before the 18^{th} century. Having more references would also allow us to use smaller time intervals, in order to obtain more precise results. Furthermore, it would have been very useful to be able to filter the more important references, as the current graph gives the same importance to, for exemple, a major breakthrough and a synthesis of the state of the art. Still, the PageRank algorithm was able to determine the most important mathematicians, giving more importance to their references.

The lack of an efficient evaluation method is however a shame, as there is no simple way of measuring the reliability of those results. Nevertheless, they still allow us to better understand the evolution of contemporary mathematics.

Github repository: https://github.com/ErwanFagnou/history_of_maths