Visualization of Mathematicians' Genealogy

Weicheng Zhu & Fangyuan Liu

Abstract As mathematics developed, lots of mathematicians are involved. They are connected with each other in a lot of dimensions. This paper introduces a way to visualize those connections, thus providing the users with a sense of the mathematicians' genealogy. Force-directed layout algorithm was used to demonstrate the advisor-student relationship between mathematicians. Zoomable tree is implemented to show the categorization and change of interest of mathematical research fields. Spatio-temporal designs are developed to show the relationship between mathematicians from different countries.

1. Introduction

1.1 Background and Motivation

Mathematics is a subject of a long history which can be dated back to ancient Greece, but there were not many Mathematicians along the history, unlike the researches in other subjects. The concentration of mathematical genius formed a close relationship with each other as a "family". For instance, Lagrange was the student of Euler; Fourier and Poisson were students of Lagrange; Dirichlet was the student of Poisson. Some of professors in our university have "ancestors" of these famous mathematicians, which is a surprising finding. Hence, we would like to represent this relationship with visualization and further explore the interconnection of their research fields, eras, distributions among the history and nationalities through the visualization analysis. Students or researchers in mathematics field might be interested in our visualization project since it can provide them with a direct insight into the history, focus and development of mathematics.

1.2 Challenges in Existing Techniques

We have a couple of tasks to achieve, including 1) tell how influential a mathematician is to his or her peers and younger generations, 2) find the ancestors of a given mathematician, 3) give users of the visualization a general understanding of how the mathematical world is divided into different fields, 4) how did the research interest of a branch of mathematicians changed over time and 5) whether geographic location is significant in influencing the spread of mathematical knowledge. We decided to implement three visualizations, since we did not find any visualization algorithm which achieves all our tasks in one visualization after doing research.

We decided to implement one of our visualizations using force-directed layout, because we have a lot of data and force-directed layout is able to cluster similar data. However, existing force-directed algorithms are inefficient in showing the links between nodes. Therefore, we need to implement an algorithm to satisfy our need.

1.3 Our Methods

We adopted the force-directed layout to visualize how mathematicians are related with each other. We used branches to denote the advisor-student relationship between professors. Implementing the tree structure with a force-directed graph would allow the tree to spread over the screen in a more compact way. If the user of our visualization would like to see the detailed information of one branch of mathematicians, he or she can move the mouse to make it hover on the target. Then the specific mathematician, his ancestors and first generation students, and the links between them would be highlighted. This visualization can also tell a story to users and potential researchers that which fields are closely related to each other and how a new subject evolved from the previous one.

In order to make the users see how mathematical research fields are categorized and the researchers involved more clearly, we implemented a zoomable treemap layout. The detailed introduction of the mathematical field is provided on the nodes, while clicking on a research field would bring the user to the next layer with subfields. All fields would be colored with the heat of the topic throughout the development of mathematics. This visualization can also provide the user (students or researchers in mathematics) an impression of the history of mathematics and predict, to some extent, the trend of future development in mathematics.

For the connections between countries, we used a spatio-temporal approach. Mathematicians are categorized by their countries and times, and we count up all pairs of advisor-student relationship. Then we visualize the relationship with curves connecting the countries where the two mathematicians are from. The thicker the curve is, the more connections mathematicians from the two countries have. We would like to divide the time into several periods so the trend of changes in connections between countries can be observed.

1.4 Our Contributions

We aim at providing a way to visualize data with strong chronological and spatial relations. Since the data have connections in multiple dimensions, we wish to amplify those connections by visualization. Force-directed graph is an effective approach, but it is not very effective in visualizing the links between nodes. We want to develop our own version of force-directed graph which has good visualization of links, which can also be used by other researchers interested in other fields. We believe this would become an effective approach when the data being visualized has multiple hierarchies, when we care about the connections between all nodes, and when the dataset is large.

1.5 The Organization of the Paper

To begin with, we would introduce the related visualization methods which other researchers had used before to implement visualization for a similar aim. In the next section we will talk about the overview of the system. After that, we will conclude the methods we used for data preprocessing. We will discuss the design and evaluations of our visualization in the next sections. Finally we are going to give a conclusion of our visualization, while providing some of the potential future works.

2. Related Work

In this section, we discuss the related work on three aspects: (1) effectively presenting connection information in a graph layout (2) visualization a large number of nodes in treemap.

Effective graph layout:

Graph layout algorithms were widely researched in literatures [10,11,12,13,15], including force simulation algorithms, clustering methods and multi-level arrangement. Typically, the output of these layout can show the feature of graphs with a large number of nodes and huge dataset through either clusters or multilevel hierarchical in information. In graph theory, the path is another important property other than node and links, especially for visualization the family of mathematicians.

There are several utilitarian methods to highlight groups in a graph. For instance, Feng [10] researched on algorithms for cluster graph drawing; Hu made attempts to visualize groups with multi level layout. With these methods, designs may occupy one addition channel and fail to provide detailed information. Developed from several force directed graph layout mentioned above, we put more emphasis on visualizing information of individual nodes and family information with BFS algorithm. Our visualization can be applied to highlight a group of nodes in a graph with complex relationship better than simply use clusters and multilevel visualizations.

Large Treemap:

Different types of treemap layout were extensively studied in literatures.[6,7,8,9], optimizing the arrangement of the grids in treemaps in 2D space. With the squarified layout, the treemap will not be as narrow as that in slice-and-dice layout. There were also a lot of special forms of treemap layout, such as Voronoi Treemap by Balzer and Deussen and Ordered Treemap by Shneiderman and Wattenberg. These algorithms were enlightening on presenting nodes with better sequences and shapes.

There was another important problem for treemap was to visualize as many nodes as possible in limited space. Because the treemap was data-ink consuming, presenting 100 nodes in one view clearly may require 10000px*20000px which was not feasible on the web page. We developed a hierarchical treemap based on the researches on single-view treemaps and stratified nodes in to different levels. Our visualization can be applied to visualize a large number of nodes within a treemap with a main view and several subviews, presenting detailed descriptions in squares clearly.

3. System Overview

We have altogether three visualizations focusing on different aims. The first is a force-directed graph visualizing the selected mathematicians and the links between them. The directed links represent that there is an advisor-student relationship between the two mathematicians connected by it, and it points from the advisor to the student. Colors of the node are used to represent the research fields a mathematician was interested in. Clicking on a node will redirect the user to the Wikipedia page of the mathematician, and hovering on it would highlight all links that are related to him.

The second visualization is a zoomable treemap. The grids represent different mathematical research fields, and clicking on a major field would bring the user to a visualization of its subfields. The color of the grids demonstrates whether the field is under heated discussion or not. The basic introduction of major fields is also provided in the grids, together with the number of mathematicians working on it and the subfields it included.

Finally, the last visualization is a world map with spatio-temporal design. Countries are represented by nodes, and links between nodes represent the connections between countries during a certain time period, which can be modified by the user. the thicker the curve is, the more connections mathematicians from those two countries have.

4. Data Preprocessing

Because all the data on the website are in the text form, we would like to read these raw data, transform them and store them as abstract data (a tree or a network with each mathematician as an item combined with their attributes stored in a node).

We first used Scrapy as a tool of ‘spider’ to capture information from HTML according to the ID of each mathematician on URL of the website. After this step, we have a file in json format with more than two hundred thousand pieces of data, each as a dictionary.

It is very hard to visualize such a large number of data points in a force-directed layout graph, so we first picked out around two hundred most influential mathematicians by applying breadth first search to the dataset, starting from the roots of the tree we found. Those mathematicians all have total descendants of more than 5000, i.e. more than five thousand mathematicians have them as advisor, or the advisor of their advisor, etc. Besides, they all have more than 5 first-generation students, implying that they are influential among their peers at their times.

We kept those data as nodes, and established links from them. For each mathematician, we found his or her students, and determine whether the student is in the two hundred people subset. However, the name of a mathematician is expressed differently as a student and as an advisor, which brought us some extra work in data cleaning.

After that, we categorized all fields of interest provided by the dataset as subfields, and put them into several general research fields.

Finally, we constructed the links between different nations from the entire dataset. We divided all links into five classes according to the year of it: 1800-1850, 1850-1900, 1900-1950, 1950-2000 and 2000-2017. The reason why we excluded the links before 1800 is that the nationality of lots of mathematicians at that time were not recorded in the system, and there are very few available links.

5. Visualization Design

5.1 Tasks

The tasks of our project includes:

1) Present the basic information (name, field of study, etc.) of mathematicians
2) Present a mathematician in the scope of all mathematicians in the history
    i. Derive and present the importance of a mathematician
    ii. Discover the ancestors of a mathematician
    iii. Present the links between a mathematicians and all his or her ancestors
3) Browse connections and families of mathematicians
4) Discover the trend of research field variations within a family
5) Analyze the associativity of mathematical theorems by name
6) Compare heat of different math fields
7) Present categories of subfields in mathematical research interests
8) Browse the cross national connection of mathematicians in different ages along history
9) Analyze the trend of globalization in mathematics study

5.2 Visualization Design

5.2.1 Force-Directed Graph for Mathematician's Genealogy

5.2.1.1 Task Analysis & Design Rationale

In this phase, connections of mathematicians which assembled the genealogy of the community were presented. The task of this graph was to trace back to all the ancestors of mathematicians and their direct students, to analyze the trend of research field variations in a family and to help mathematics learners to understand the development of some mathematics theorems named after mathematicians by associating those theorems with other theorems established by advisors or students of those mathematicians.

The graph was implemented as a tool to visualize the connections and research field information. Tree was another important tool for genealogy visualization, but it cannot be applied here since some of mathematicians had more than one advisors. For instance, Klein has two advisors who can date back to Gauss and Euler, which are two important branches in mathematicians. Hence, the graph here was a better design than a tree to present a more comprehensive information. On the node, different colors were used to present different research fields from which the trend of researches in families can be easily interpreted.

5.2.1.2 Visual Encoding Schema

- Marks:
  - Points: a node represented a mathematician;
  - Connections: links among mathematicians;
  - Areas: a node occupied areas according to its importance;
- Channels:
  - Color(Hue): different research fields;
  - Size: the importance of mathematicians;
  - Position: locating a mathematician with his node;

5.2.1.3 Layout Design

A force-directed graph was implemented here to avoid too much crossings of edges and overlapping of nodes. Because the number of mathematicians were great, a strong repulsive force was applied in this layout. To make the graph more flexible, the layout enables users to drag nodes and zoom-in/out with mouse so that they can see the graph more clearly. Also, the size and color of nodes was used to represent the importance and research fields of mathematicians, and the width of links was used to represent the strength of connection between advisors and students.

Breadth-First-Search algorithm was implemented in the layout of animation. If one put the mouse on a node of mathematician, all of his ancestors and direct students would be highlighted. Since the number of advisors to a mathematician was uncertain, simply calling the parent of the node was useless, BFS algorithm was applied to visit all the ancestors.

If the user clicks on the node, a Wikipedia page of corresponding mathematicians would pop out. This interactive function in layout can provider more information for user to learn related mathematics theorems and papers of this mathematician. The visualization itself is also zoomable via scrolling, making it easier for the users to see a specific part of the visualization.

5.2.2 Hierarchical Treemap for Volumes of Fields in Mathematics

5.2.2.1 Task Analysis & Design Rationale

In this view, different mathematics fields were visualized with a two-level hierarchical treemap. The task was to analyze the relationship among different research subjects inherited from the first view to browse basic information of specific fields and to discover the hottest fields in recent years so that the user can learn the latest trend of mathematical researches from the map.

A hierarchical treemap was implemented for these tasks with size of squares for number of mathematicians in certain field and color for the heat of field in 10 years. A treemap was applied because in squares there were enough area to present brief introduction on this field so that users can immediately know what the field was about. Otherwise, some user may not know what the meaning of "Analysis" in mathematics was by simply looking at the word. Also, the reason why a hierarchical treemap was applied as an improvement from traditional treemap was that there are about 100 sub-fields categories in the dataset. It was messy to present all of them in one flat treemap, and the hierarchical treemap can greatly reduce the number of squares in a single view making the visualization more organized. Based on these reason, the hierarchical treemap was a rational choice of this view.

5.2.2.2 Visual Encoding Schema

- Marks:
  - Containment: relationship between a field and its subfield;
  - Areas: squares for different research fields;
- Channels:
  - Color(Hue): heat of research fields: the redder, the hotter; the greener, less hot;
  - Size: the number of mathematicians in this field;
  - Position: locating a field with its square;

5.2.2.3 Layout Design

The dataset was transformed to a two-level tree including general fields and their sub-fields. In the first stage, each square represented a general field, and on-click either of them, it would show the treemap of the subfields. With this animation, treemap showed its properties of enclosure and helped user to learn the relationship among different fields. Also, the size of the squares showed the number of mathematicians in that field; the color showed the heat of fields, according to the ratio between the number of mathematicians in recent ten years in that field and the total number. Together with the texts on squares about the total number and descriptive information, this graph provided a comprehensive visualization on mathematics fields.

5.2.3 *Temporal World Map for Cross-national Connections of Mathematicians*

5.2.3.1 Task Analysis & Design Rationale

As a subtask of browsing the connection among mathematicians and analyzing the development of mathematics, this graph would like to compare the scale of cross-national communication in different ages (1800s, 1850s, 1900s. 1950s and 2000s) and analyze the trend of globalization of mathematics researches and trends of development in specific countries.

In this view, a spatial-temporal design was implemented because the task was to compare information of different time. The nationalities can be abstracted as GPS locations on the map. Hence, it was reasonable to design follow a spatial-temporal problem on a world map which a time axis to flip between different ages.

5.2.3.2 Visual Encoding Schema

- Marks:
    - Lines: time axis
    - Points: nationalities of mathematicians;
    - Areas: location of different countries;
    - Connection: the number of connections in maths among different countries
- Channels:
    - Size: the stroke width represents the strength of connection between two counties;
    - Position: the location of countries;

5.2.3.3 Layout Design

The spatial-temporal design was implemented in an interactive layout. The view showed the connections of mathematicians between two countries. The scale of communication between two countries was quantified by the sum of the number of students in country A being supervised by advisors in country B and the number of students in country B being supervised by advisors in country A. The large scale communications were, the wider the connection lines would be. Then this layout projected these connections onto a world map and represented them in a spatial visualization, providing users a more direct observation on these countries then simply using an abstract graph.

The layout to present the temporal information was a drag slider on which user can choose the year range to visualize. On this time axis, five equally split year-range was presented to show the development of globalization in mathematics study, and highlighted several centers of mathematics education and researches in different ages.

5.3 Interactions

We propose following interactions to effectively make use of our visualizations:
Highlight. When hovering the mouse over a user's node in a view, the same user shown in different views is highlighted at the same time, showing the innate connection of the data

- **Highlight.** When hovering the mouse over a mathematician's node in the first visualization, his first-generation students and all his ancestors would be highlighted, together with the links between them.
- **Inspection.** The users of the visualization would be directed to the Wikipedia page of the mathematician by clicking on the node of the mathematician in the first visualization. He can browse through the personal information and get a more thorough understanding of the mathematician via the Wikipedia page.
- **Zoom and Pan.** The first visualization supports zooming and panning, making it more convenient for the users to explore the data.
- **Explore.** The users are allowed to click on any of the grids in the second visualization, which would zoom in the visualization and bring them to the subfields of the major research field they are interested in. After that, they would be able to see which subfields are under heated discussion in the history of mathematics.
-

Drag the slider. In the third visualization, users can drag the slider to see the relationship between mathematicians from different countries in different times.

6. Evaluation

We captured a couple of interesting patterns when implementing this visualization.

6.1 Force-Directed Graph for Mathematician's Genealogy

- Less than 10% of mathematicians is neither a descendant of Johann Bernoulli nor Christian August Hausen. Johann Bernoulli, who is a member of the famous Bernoulli family in mathematics, did a lot of research in integration and catenaries. He is also the advisor of Leonhard Euler, one of the most famous mathematicians of all times. Hausen is not very famous for common people, but his third-generation student, Carl Friedrich Gauss, who was sometimes referred to as "the foremost of mathematicians", is also one of the most influential mathematicians.
- Most mathematicians after 1850 is a descendant of both of them. This is because the connections between mathematicians greatly increased after 1850. The phenomenon can be explained by the third visualization, in which the connection between mathematicians according to nationality after 1800 is visualized.
- The families of mathematicians turned to applied mathematics fields were not likely to switch back to pure mathematics. We are not really sure about what is happening here, but we have a hypothesis. If a mathematician turned to applied mathematics because he

thought pure mathematics is too far-fetched with real world, he might refuse to advise PhD students working in pure mathematics field since he thought it was useless. But if a mathematician turned to, for example, functional analysis from number theory, he is still in the pure mathematics domain, therefore would not refuse to advise the same student with the reason mentioned above. Therefore, we concluded that the gap between pure mathematics and applied mathematics is very deep.

- The mathematics world is really small, as some mathematicians have families that form a loop, even multiple loops sometimes. Such a situation happens because some mathematicians have two advisors, and both of them are a descendent of some ancient mathematician. Another phenomenon is that those gaps are often small, and restricted in few countries. While mathematicians communicate with others increasingly often, they seem to be more and more willing to promote young mathematicians from their own countries, making use of their occupation and resources as a professor.
- The mathematicians who expertise in probability theory may be closely connected to mathematicians of measure theory, which implies that probability theory largely developed from measure theory.

6.2 Hierarchical Treemap for Volumes of Fields in Mathematics

- More than half of all mathematicians devote themselves in applied mathematics research. This is not a surprising phenomenon to us, because almost all mathematical knowledge has some kind of practical application and abstract theories are often more complicated than their applications. Therefore, the connection between mathematics and real life is much tighter than common people believe.
- The most studied pure mathematics subfield is differential equations.

6.3 Temporal World Map for Cross-national Connections of Mathematicians

- The connections between mathematicians increased after 1800, and as time passed by, the connections were increasing faster and faster. Transportation in the 19th century is much faster and more convenient than the past, which is the most important reason that the connections between mathematicians increased. Another reason might be that the problems mathematicians worked on are becoming more and more difficult, thus other people's help becomes necessary.
- United States is really the core of mathematics in the scope of the spread of mathematical knowledge. Around 40% of all advisor-student relationships in the recent two decades are related to American mathematicians, and the ratio was even higher in the old times.

7. Conclusion and Future Work

In this paper, we proposed three visualizations to visualize the multi-dimensional data of all mathematicians in the history. Those visualizations allow users to have a better idea about the connections between mathematicians from different generations, and the development of mathematical research over the years. Future work includes finding additional data of mathematical theorems and visualize them with reference to the mathematicians who discovered them, making the visualization more powerful. People would be able to look up the mathematical formulas they are interested in, and browse through the story of the discoverer of the theorem using our visualization. A formal user study could also be used to improve our visualization.

Bibliography

1. D. Myers and J. McGuffee "Choosing Scrapy", *Journal of Computing Sciences in Colleges, Volume 31 Issue 1, October 2015, Pages 83-89*.
2. N. Cao, D. Gotz, J. Sun and H. Qu, "DICON: Interactive Visual Analysis of Multidimensional Clusters," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2581-2590, Dec. 2011.
3. Lei Shi *et al*., "HiMap: Adaptive visualization of large-scale online social networks," *2009 IEEE Pacific Visualization Symposium*, Beijing, 2009, pp. 41-48.
4. R. Kosara and J. Mackinlay, "Storytelling: The Next Step for Visualization," in *Computer*, vol. 46, no. 5, pp. 44-50, May 2013.
5. Aigner, Wolfgang, *et al. "*Visualization of time-Oriented data". *London, Springer, 2011.*
6. M. Balzer and O. Deussen. "Voronoi treemap" *IEEE Symposium on Information Visualization*, pages 49 - 56, 2005.
7. M. Bruls, K. Huizing, and J. Van Wijk. "Squarified treemaps". *Proceedings of the joint Eurographics and IEEE Symposium on Visualization,* pages 33 - 42, 2000.
8. B. Shneiderman and M. Wattenberg. "Ordered treemap layouts". *IEEE Symposium on Information Visualization*, pages 73 - 80, 2001.
9. B. Shneiderman. "Tree visualization with treemaps: 2d space-filling approach". *ACM Transactions on Graphics: 92 - 99*, 1992
10. Q. Feng. "Algorithms for Drawing Clustered Graphs" PhD thesis, University of Newcastle, 1997.
11. N. Henry and J. Fekete. "NodeTrix: a Hybrid Visualization of Social Networks" *IEEE Symposium on Information Visualization*, 2007.

12. T. Kamada and S. Kawai. “An algorithm for drawing general undirected graphs”. *Inf. Process. Lett.,*31(1):7-15,1989.
13. M. L. Huang and P. Eades. “A Fully Animated Interactive System for Clustering and Navigating Huge Graphs” *Proceedings of International Symposium on Graph Drawing*, 1998.
14. K. Misue, P. Eades, W. Lai, and K. Sugiyama. “Layout Adjustment and the Mental Map”. Journal of Visual Languages and Computing, 6(2): 183-201, June 1995.
15. P. Eades and Q.- W. Feng. “Multilevel Visualization of Clustered Graphs” Proceedings of the International Symposium on Graph Drawing. London, UK, 1997. Springer-Verlag.
16. Lau DC, Murnighan JK. Demographic diversity and faultlines: the compositional dynamics of organizational groups. Acad Manage Rev 1998;23 (2):325–40.
17. Bezrukova K, Jehn KA, Zanutto EL, Thatcher SM. Do workgroup faultlines help or hurt? A moderated model of faultlines, team identification, and group performance Organ Sci 2009;20(1):35–50.
18. Frank AU. Different types of ‘‘Times’’ in GIS. In: Egenhofer MJ, Golledge RG, editors. Spatial and temporal reasoning in geographic information systems. New York, USA: Oxford University Press; 1998. p. 40–62 [Chapter 3].
19. Van Wijk JJ. Image based flow visualization. In: Proceedings of SIGGRAPH, San Antonio, USA; 2002.
20. Reinders F, Post FH, Spoelder HJW. Visualization of time- dependent data with feature tracking and event detection. The Visual Computer 2001;17(1):55–71.
21. Havre S, Hetzler E, Whitney P, Nowell L. ThemeRiver: visualizing thematic changes in large document collections. IEEE Transactions on Visualization and Computer Graphics 2002;8(1):9–20.
22. Tominski C, Abello J, Schumann H. Axes-based visualizations with radial layouts. In: Proceedings of ACM symposium on applied computing, Nicosia, Cyprus; 2004.
23. Tominski C, Schulze-Wollgast P, Schumann H. 3D information visualization for time dependent data on maps. In: Proceedings of international conference on information visualisation, London, UK; 2005.
24. Moere AV. Time-varying data visualization using information flocking boids. In: Proceedings of IEEE symposium on information visualization, Austin, USA; 2004.
25. van Wijk JJ, van Selow ER. Cluster and calendar based visualization of time series data. In: Proceedings of IEEE symposium on information visualization, San Francisco, USA; 1999.
26. Aigner W, Miksch S, Thurnher B, Biffl S. PlanningLines: novel glyphs for representing temporal uncertainties and their evaluation. In: Proceedings of international conference on information visualisation, London, UK; 2005.
27. N. Gershon and W. Page. What storytelling can do for information visualization. Commun. ACM, 44(8):31–37, 2001.

28. W. Wojtkowski and W. G. Wojtkowski. Storytelling: its role in information visualization. In European Systems Science Congress, 2002.
29. K. Cukier. Show Me: New ways of visualising data. http://www.economist.com/node/15557455, 2010.
30. D. Bordwell and K. Thompson. Film Art: An Introduction. McGraw-Hill, 2003.