Visualizing Sets and Set Relations

Oliver Deussen
University of Konstanz, Germany
Background

Computer Graphics
Abstract Computer Graphics
Information Visualization
e-David: A painting robot

- Visual feedback
- Target function (image)
- Constraints (styles)
- Drawing commands
All of Inflation’s Little Parts

Each month, the Bureau of Labor Statistics gathers data on about 80,000 consumer goods and services, including prices for food, fuel, shelter, and medical care—among others—and follows a similar list of 800 goods and services to track the Consumer Price Index, or CPI, measure of inflation. It is among the statistics that the Federal Reserve uses when it sets interest rates on Wall Street.

This category is weighted according to an estimate of what the average adult- aged consumer spends, as shown below:

Larger shapes mean a larger part of spending. Colors correspond to change in prices over the last year.

---

Uncertainty
Graph Layouts
Content

History of Set Visualizations
SpEuler
RectEuler
History of Set Visualizations
Visualizing logical propositions - Gottfried Wilhelm Leibniz

All $B$ is $C$

No $B$ is $C$

Some $B$ is $C$

Some $B$ is not $C$

Baron, Margaret E. (May 1969). "A Note on The Historical Development of Logic Diagrams". The Mathematical Gazette
"These four species of propositions may, likewise, be represented by figures, so as to exhibit their nature to the eye. This must be a great assistance to comprehending, more distinctly, wherein the accuracy of a chain of reasoning consists”

Leonhard Euler, *Lettres à une princesse d'Allemagne sur divers sujets de physique et de philosophie*, Letter CII, 1768
John Venn (1880)

“All that is necessary for this purpose is to describe a series of closed figures, of any kind, so that each successive one shall intersect all the compartments already produced, and thus double the number.”

Baron, Margaret E. (May 1969). “A Note on The Historical Development of Logic Diagrams”. *The Mathematical Gazette*
John Venn (1880)

All X is Y

Venn Diagram

Euler Diagram
John Venn (1880)

Visualizations of the propositions:

All X is either Y and Z Or not–Y
If any XY is Z, then it is W
No WX is YZ

Venn, John (1880). “On the diagrammatic and mechanical representation of propositions and reasonings”. Philosophical Magazine Series 5
For all practical purposes, however, any outline which is not very simple and easy to follow with the eye fails entirely in its main purpose of affording intuitive and sensible illustration.

*John Venn in “On the diagrammatic and mechanical representation of propositions and reasonings” (1880)*
Abstract Description

-> does not determine the visual representation

for sets A, B and C show all intersections: \(\emptyset, a, b, c, ab, ac, bc, abc\)
Related Techniques

EulerView (2009)

Related Techniques

Untangling Euler Diagrams (2010)

Related Techniques

SetNet (2016)

Related Techniques

Edeap (2021)

Related Techniques

MosaicSets (2023)

Related Techniques

Metroset (2020)

SPEuler, a first approach
Motivation

A manually created Venn diagram:

Our automatic solution:

XKCD Size Diagram: https://xkcd.com/2122/
The dual graph

Euler diagram ↔ dual graph ↔ rank-based dual graph

The dual graph

non-simple diagram ↔ irregular dual

Non-pairwise intersection

Irregular (non-monotonic) face

The dual graph

simple diagram ↔ monotonic dual graph

pairwise intersection AB ↔ monotonic face AB
(a) well-matched

(b) well-formed and well-matched
Creating the set enclosures

(1) Link ordering for each set

(2) Create Catmull-Rom splines

(3) Final diagram
Final Venn Diagram with Labels
RectEuler – making things easier...
Our Inspiration
Idea: use Mixed Integer Linear Programming
Idea: use Mixed Integer Linear Programming

Set names must not overlap (H5)

Fully-enclosed sets must be contained in parent set (H7)

Must remain in bounding rect (H4)

Must keep minimal distance of disconnected sets (H6)

Must be in participating sets (H1)

Must not be in other set (H2)

Must keep minimum distance between lines (H3)

Empty intersections should be avoided if possible (H6)
Idea: use Mixed Integer Linear Programming

Mixed Integer Linear Programming as optimization procedure:

Minimize Rectangles:

\[ o_{\text{rects}} = \sum_{r \in U} (r_{x_2} - r_{x_1}) + (r_{y_2} - r_{y_1}) \]
Splitting

- Complex datasets: cannot be shown with a single rectangular Euler diagram
- Our approach: if no feasible MILP for the whole dataset, we split
- individual sub-diagrams are generated for each part of the dataset
Cluster Split
Cluster Split
Splitting
Back to our motivation:

Human made EU dataset

Ours
Limitations

Not Well-Matched:
- Extra zones are present

Not Well-Formed:
- Disconnected Zones
- Duplications in split diagrams
Summary

- Set visualizations are old, no perfect solution so far
- SpEuler – works with curves
- RectEuler – a solution with rectangles