A Case Study in Fitting Area-Proportional Euler Diagrams with Ellipses Using **eulerr**

Johan Larsson[1]   Peter Gustafsson

Department of Statistics, School of Economics and Management, Lund University, Lund, Sweden

June 18, 2018

[1] johanlarsson@outlook.com

Motivation

We prefer circular Euler diagrams, yet they often fail to produce acceptable diagrams for set relationships with three or more intersecting sets. Consider

$$A = B = C = 4,$$
$$A \cap B = A \cap C = B \cap C = 1, \text{ and}$$
$$A \cap B \cap C = \emptyset.$$

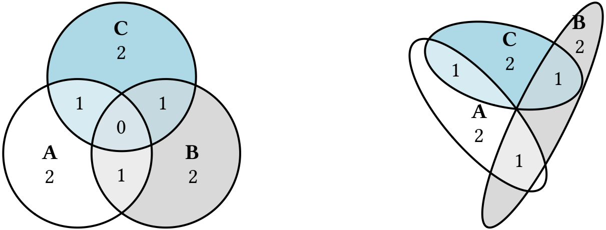


Figure: The merit of elliptical diagrams.

eulerAPE

eulerAPE [Micallef and Rodgers, 2014] introduced elliptical Euler Diagrams, yet only for three-set diagrams with three, fully-intersecting ellipses.

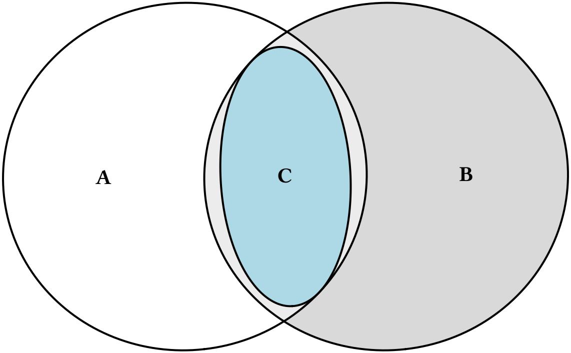


Figure: Impossible with **eulerAPE**.

eulerr's Algorithm

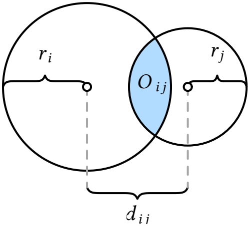
The Euler diagram is fit in two steps:

- first, an initial layout is formed with *circles* using only the sets' pairwise relationships;
- secondly, this layout is fine-tuned, optionally using *ellipses*, asnd taking all $2^N - 1$ intersections into consideration.

The Initial Layout

First, we need the circles' pairwise overlaps, which we find numerically.

$$O_{ij} = r_i^2 \arccos\left(\frac{d_{ij}^2 + r_i^2 - r_j^2}{2d_{ij}r_i}\right) + r_j^2 \arccos\left(\frac{d_{ij}^2 + r_j^2 - r_i^2}{2d_{ij}r_j}\right) - \frac{1}{2}\sqrt{(-d_{ij} + r_i + r_j)(d_{ij} + r_i - r_j)(d_{ij} - r_i + r_j)(d_{ij} + r_i + r_j)}.$$



Constrained Multi-Dimensional Scaling

$$\mathcal{L}(h,k) = \sum_{1 \le i < j \le N} \begin{cases} 0 & (1) \\ 0 & (2) \\ \left(\left(h_i - h_j\right)^2 + \left(k_i - k_j\right)^2 - d_{ij}^2 \right)^2 & \text{otherwise} \end{cases} .$$

$$\vec{\nabla} f(h_i) = \sum_{j=1}^{N} \begin{cases} \vec{0} & (1) \\ \vec{0} & (2) \\ 4\left(h_i - h_j\right)\left(\left(h_i - h_j\right)^2 + \left(k_i - k_j\right)^2 - d_{ij}^2 \right) & \text{otherwise,} \end{cases}$$

with

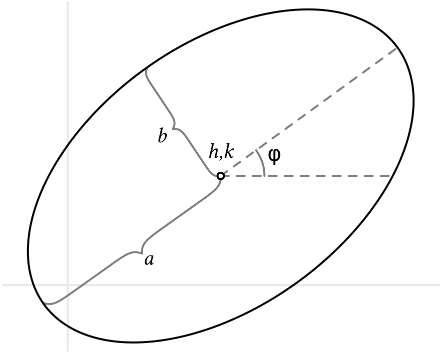
$$(1) \coloneqq F_i \cap F_j = \emptyset \text{ and } O_{ij} = 0$$

$$(2) \coloneqq (F_i \subseteq F_j \text{ or } F_i \supseteq F_j) \text{ and } O_{ij} = \min\left(F_i, F_j\right)$$

Final Layout

For our final layout, we extend ourselves to ellipses.

$$\frac{[(x-h)\cos\phi + (y-k)\sin\phi]^2}{a^2} + \frac{[(x-h)\sin\phi - (y-k)\cos\phi]^2}{b^2} = 1,$$



However, because an ellipse is a conic it can be represented in quadric form,

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$$

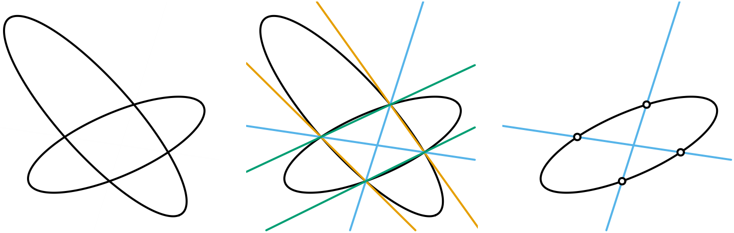
that in turn can be represented as a matrix,

$$\begin{bmatrix} A & B/2 & D/2 \\ B/2 & C & E/2 \\ D/2 & E/2 & F \end{bmatrix},$$

which is the form we need to intersect our ellipses.

Intersecting Ellipses

1. Form three degenerate conics from a linear combination of the two ellipses we wish to intersect,
2. split one of these degenerate conics into two lines, and
3. intersect one of the ellipses with these lines, yielding 0 to 4 intersection points points.



Overlap Area

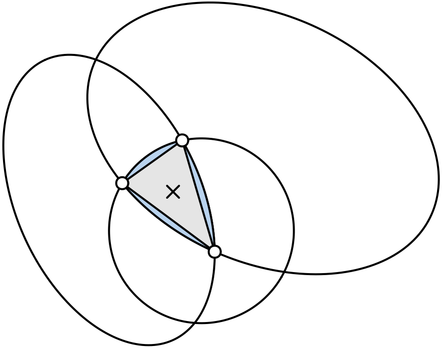


Figure: The area we are interested in is the convex polygon made up of all the intersecting points that lie inside all the sets.

Elliptical Segment

We first obtain elliptical sectors $F(\theta_0)$ and $F(\theta_1)$ and then subtract the smaller sector from the larger. The elliptical segment (in blue) is then found by subtracting the triangle part (in grey) from $F(\theta_1) - F(\theta_0)$.

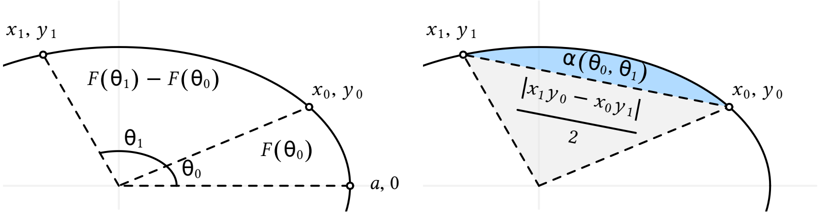


Figure: Obtaining the elliptical segment between two points x_0, y_0 and x_1, y_1.

Optimization Procedure

The optimization target is *stress* [Wilkinson, 2012],

$$\frac{\sum_{i=1}^{n}(A_i - \beta\omega_i)^2}{\sum_{i=1}^{n} A_i^2},$$

where

$$\beta = \frac{\sum_{i=1}^{n} A_i\omega_i}{\sum_{i=1}^{n} \omega_i^2}.$$

We use the quasi-Newton optimizer `nlm()` [Schnabel et al., 1985]; optionally, a last-ditch optimizer [Xiang et al., 2013], **GenSA**, may be employed for relationships that are particularly difficult to fit.

A Six-Set Combination

We begin our examination of **eulerr** by studying a difficult set relationship from Wilkinson [2012]:

$$A = 4, \quad B = 6, \quad C = 3, \quad D = 2, \quad E = 7, \quad F = 3,$$
$$A\&B = 2, \quad A\&F = 2, \quad B\&C = 2, \quad B\&D = 1,$$
$$B\&F = 2, \quad C\&D = 1, \quad D\&E = 1, \quad E\&F = 1$$
$$A\&B\&F = 1, \quad \text{and} \quad B\&C\&D = 1,$$

Diagrams from **venneuler**, **eulerr** (circles), and **eulerr** (ellipses). Stress values are 0.0066562, 0.0042077, and $3.1733942 \times 10^{-13}$ respectively."

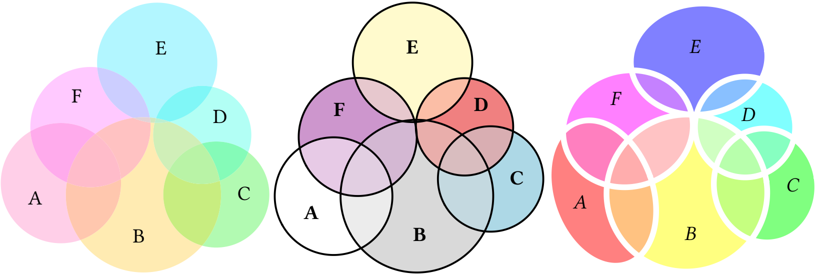


Figure: Diagrams based on the six-set relationship from Wilkinson.

Consistency

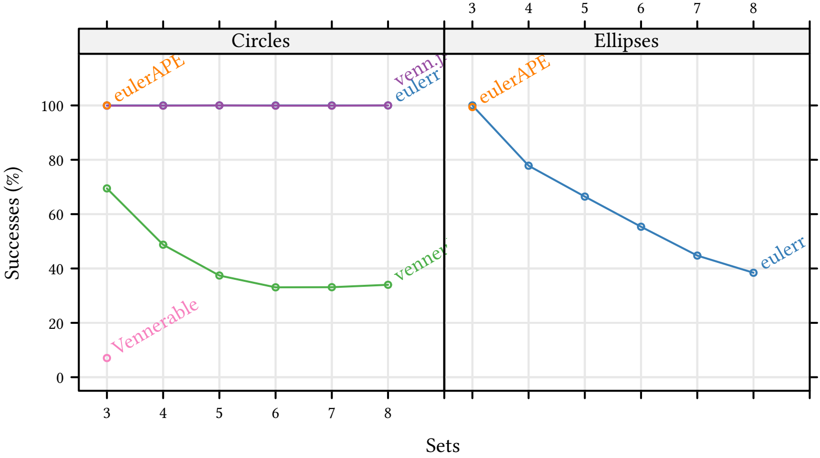


Figure: Consistency in reproducing randomly sampled diagrams.

Accuracy (Three-Set Relationships)

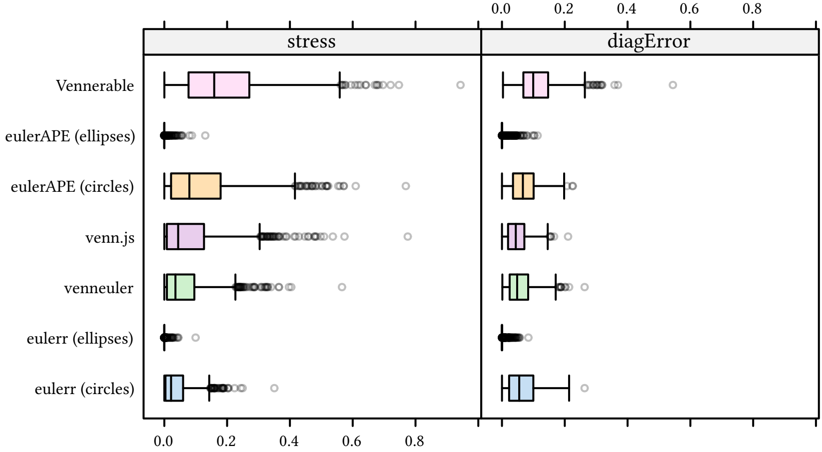


Figure: Error in reproducing random three-set relationships.

Accuracy

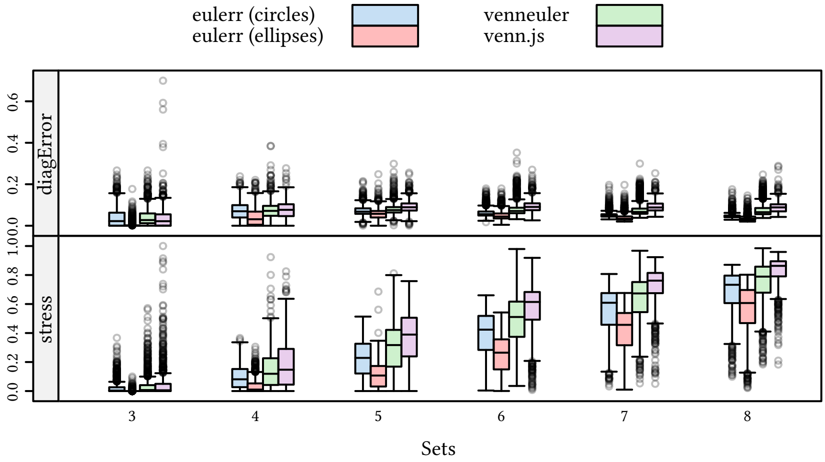


Figure: Error in reproducing randomly samples set relationships.

Performance

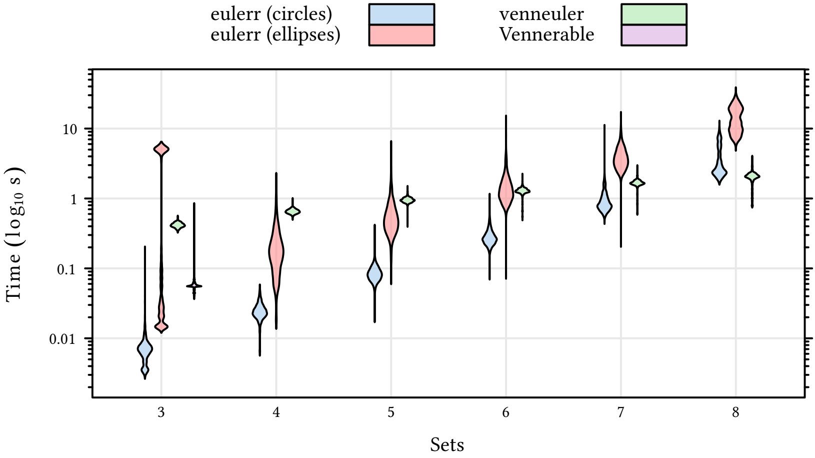


Figure: Wall clock performance for various R-based solutions.

Availability and Sample Code

eulerr is available on the Comprehensive R Archive Network. In R, it can be installed, loaded, and used to fit a simple diagrams easily:

```
install.packages("eulerr")
library(eulerr)

# A simple diagram
fit <- euler(c("A" = 1,
               "B" = 3,
               "A&B" = 1.3))
plot(fit)
```

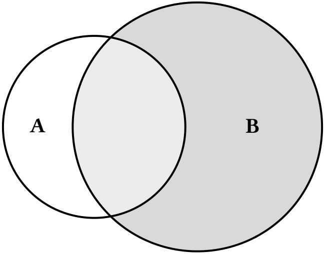
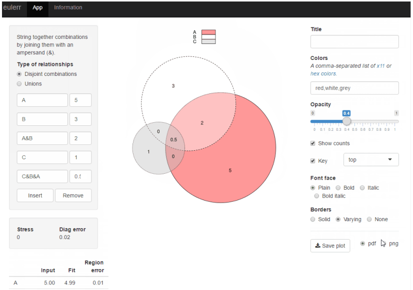


Figure: A simple diagram.

Shiny Application

eulerr is also available as a Shiny application at `http://eulerr.co`.



Thank you for your attention!

References

Luana Micallef and Peter Rodgers. eulerAPE: drawing area-proportional 3-Venn diagrams using ellipses. *PLOS ONE*, 9(7):e101717, July 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0101717.

L. Wilkinson. Exact and approximate area-proportional circular Venn and Euler diagrams. *IEEE Transactions on Visualization and Computer Graphics*, 18(2): 321–331, February 2012. ISSN 1077-2626. doi: 10.1109/TVCG.2011.56.

Robert B. Schnabel, John E. Koonatz, and Barry E. Weiss. A modular system of algorithms for unconstrained minimization. *ACM Trans Math Softw*, 11(4): 419–440, December 1985. ISSN 0098-3500. doi: 10.1145/6187.6192.

Yang Xiang, Sylvain Gubian, Brian Suomela, and Julia Hoeng. Generalized simulated annealing for global optimization: the GenSA package. *The R Journal*, 5(1):13–28, June 2013. URL `https://journal.r-project.org/archive/2013/RJ-2013-002/index.html`.