Some Results for Drawing Area Proportional Venn3 With Convex Curves

Peter Rodgers
University of Kent
p.j.rodgers@kent.ac.uk

Jean Flower
Autodesk

Gem Stapleton John Howse
University of Brighton
{g.e.stapleton,john.howse}@brighton.ac.uk

Abstract

Many data sets are visualized effectively with area proportional Venn diagrams, where the area of the regions is in proportion to a defined specification. In particular, Venn diagrams with three intersecting curves are considered useful for visualizing data in many applications, including bioscience, ecology and medicine. To ease the understanding of such diagrams, using restricted 'nice' shapes for the curves is considered beneficial. Many research questions on the use of such diagrams are still open. For instance, a general solution to the question of when given area specifications can be represented by Venn3 using convex curves is still unknown.

In this paper we study symmetric Venn3 drawn with convex curves and show that there is a symmetric area specification that cannot be represented with such a diagram. In addition, by using symmetric diagrams drawn with polygons, we show that, if area specifications are restricted so that the double intersection areas are no greater than the triple intersection area then the specification can be drawn with convex curves. We also propose a construction that allows the representation of some area specifications when the double intersection areas are greater than the triple intersection area. Finally, we present some open questions on the topic.

Keywords--- **Venn diagrams; diagram layout**

1. Introduction

Venn diagrams [10] are a widely used method for visualizing intersecting data sets. Often, the data comes with cardinality information and so might best be visualized using area proportional Venn diagrams. Here, the data set forms an area specification, as the areas of the regions of the diagram are proportional to the relevant numerical set cardinalities. In figures 1 and 2, area proportional Venn3 diagram are shown, with the data values written inside the regions. Applications where such diagrams are used include: bioscience, for example when visualizing lists of intersecting biological identifiers [6]; ecology, for example when visualizing species distribution [5] and medicine, for example when visualizing the results of online medical surveys [7]. Well known shapes, such as triangles [1] and rectilinear curves [4] have been used to draw Venn diagrams. In particular, circles are considered an effective way to visualize area proportional Venn and Venn-like diagrams [1][3][6][8]. However, it is known that circles cannot always be used to achieve exact proportions for the regions of a Venn3 diagram [4]. In general, it would be useful to know which area specifications can be represented by diagrams with particular shapes.

In this paper, we take some steps in exploring which shapes can used to draw area proportional Venn3 diagrams. In particular, we look at convex curves. This is motivated by, firstly, the observation that Venn diagrams are often more usable if they are drawn with convex curves. Secondly, we note that many (if not all) of the desirable shapes for curves are convex, so that if a non-convex curve is needed for a data set, then the diagram cannot be drawn with such a desirable shape.

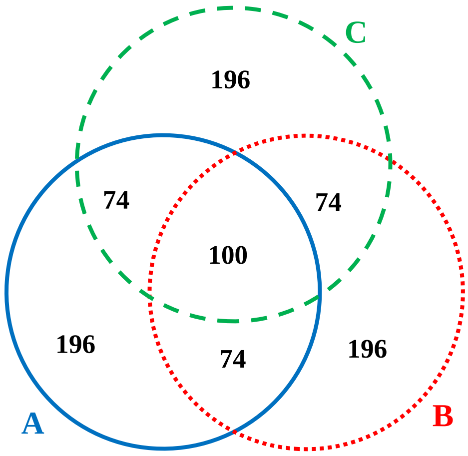

Figure 1: Convex, symmetric Venn3

Venn3 diagrams are surprisingly complex to reason about, so we restrict the problem to manageable size by considering symmetric Venn3 diagrams with rotational symmetry of order 3 and reflection symmetry of 3 axes, for example as seen in Figure 1. Conversely, a non-convex, non-symmetric diagram is shown in Figure 2.

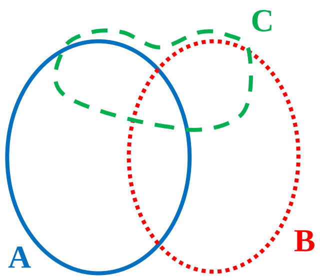

Figure 2: Non-convex, non-symmetric Venn3

After defining our terms in Section 2, we go on in Section 3 to outline a proof that some symmetric area specifications cannot be drawn with any convex symmetric Venn3 diagram. In sections 4, 5 and 6, we state some conditions relating diagram areas which can be realised with convex Venn 3, and we describe how to construct a suitable diagram from an area specification. In Section 6 we introduce a third diagram type that extends the conditions found in Section 4. Section 7 gives some open questions, and finally Section 8 gives our conclusions and further work.

2. Definitions

Here we define some of the concepts that will be used throughout this paper.

Venn diagrams are collections of simple closed curves where every possible intersection between curves is a non-empty, connected region. **Venn3** is a Venn diagram with exactly three curves. A Venn diagram is **simple** if, at most, only two curves meet at a particular point. We deal only with simple Venn diagrams in this paper. A Venn diagram is **convex** if all of its curves are convex.

For simple Venn3, there is only one topologically distinct embedding [9], as shown in Figure 1. In fact this figure shows a symmetric, convex Venn3 diagram. In this paper we will label the curves with a fixed set of labels: *A*, *B* and *C*. Regions of a Venn diagram are described by exactly the labels of the curves that contain them. For instance, the region inside all three curves is described by $\{A, B, C\}$. This region is called a triple intersection, since it is the intersection of three curves. Similarly, the three regions that are each contained by two curves, such as $\{A, B\}$, are called double intersections and the three regions that are each contained by one curve, such as $\{A\}$ are called single intersections.

$R = \mathbb{P}(\{A, B, C\}) - \{\emptyset\}$ is the set of all **region descriptions** for any Venn3 diagram and $w: R \rightarrow \mathbb{R}^+ \cup \{0\}$ is an **area specification**. If

$$w(A) = w(B) = w(C)$$
$$w(AB) = w(AC) = w(BC)$$

then **w is a symmetric area specification**. In this paper we use a_1, a_2 and a_3 to be the three data values for a symmetric area specification where:

$$a_1 = w(A) = w(B) = w(C)$$
$$a_2 = w(AB) = w(AC) = w(BC)$$
$$a_3 = w(ABC)$$

Given an area specification, w, we say that a Venn diagram, d, **represents** w if, for each region, *r*, the area of *r* is $w(des(r))$, where $des(r)$ is the description of *r*. Venn3 diagrams are dimensionless, which implies that there is relative size for each region of d.

If a Venn3 diagram, d, has rotational symmetry of order 3, and reflection symmetry with 3 axes, then d is **symmetric**. It is trivial to show that a symmetric Venn3 diagram has a symmetric area specification.

A convex symmetric area proportional Venn3 is shown in Figure 1 where the area specification is $a_1 = 100$, $a_2 = 74$ and $a_3 = 196$.

3. Some area proportions cannot be represented by symmetric convex diagrams

In this section, we are concerned with the non-representability of certain area proportions. In particular, we will show that some area proportions are not representable by a symmetric, convex Venn3 diagrams. We are not the first to investigate non-representability, and we first discuss a previous attempt at proving certain area proportions could not be represented by a convex Venn3 diagram.

3.1. Previous proof attempts

A significant study of area proportional Venn diagrams was conducted by Chow, and a large number of results can be found in [2]. Chow was interested in whether there existed an area specification that could not be represented by a convex Venn3 diagram. He attempted to prove that this was indeed the case but, unfortunately, his proof contains a flaw. In particular, he sets up a symmetric area specification with the double intersections having twice the area of the triple intersection, thus making a_2 'large' with respect to a_3.

Then, he supposes that there is a Venn3 diagram that represents this specification for any a_1, thus treating a_1 as a variable. He then considers the core of the diagram; the core is the region (including its boundary) consisting of the three double intersections and the triple intersection. The argument then proceeds to suppose that for each given value of a_1, the core is not convex and, therefore, the area of the convex hull of the core, less the area of the core, is positive. He then incorrectly deduces that the set of all such area differences (recall that a_1 is a variable, so there is one such area difference for each a_1) has an infimum that is strictly positive. This is a key step in his proof, since he uses the positivity of the infimum to derive a contradiction. Chow's insight, however, about the relationship between the area proportions (making a_2 large and a_3 small to force a_1 to have a positive lower bound on its area) was correct, and is exemplified by our result below. We note, though, that our proof that follows relies on symmetry whereas Chow's proof was aiming for a general non-representability result.

3.2. A non-representability result

Theorem There exists an area specification which cannot be represented by a convex symmetric Venn 3 diagram.

Proof (Sketch) The idea of the proof is to show that, by deriving some conditions on the triple intersection, we sometimes have a non-zero lower bound on the area a_2. First, we define notation for points, angles and lengths as illustrated in Figure 3:

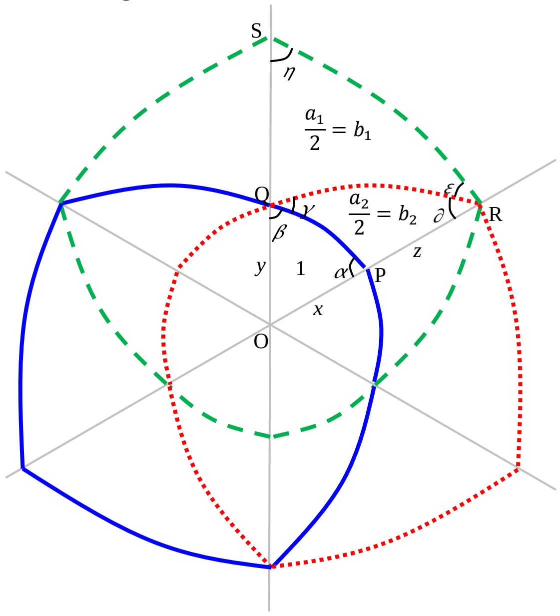

Figure 3

The idea of the proof is to show that if b_2 is large, then $x + z$ is large compared to y. Convexity then means that b_1 is bound below, so choosing a large enough b_2 and a small enough b_1 gives an unrepresentable area specification. Using convexity at the points P, Q and R, we can derive constraints on the angles, for example $\eta \leq \frac{\pi}{2}$. From these, we can derive constraints on the relationships between the lengths x, y and z in terms of the areas:

$$x + z \leq \frac{2\sqrt{2}}{\sqrt[4]{3}}\sqrt{1 + b_1 + b_2}$$

$$x + z \geq \frac{2}{\sqrt{3}}(1 + b_2) + \frac{1}{2\sqrt{2}}$$

Combining these two bounds:

$$\frac{2}{\sqrt{3}}(1 + b_2) + \frac{1}{2\sqrt{2}} \leq \frac{2\sqrt{2}}{\sqrt[4]{3}}\sqrt{1 + b_1 + b_2}$$

Taking $b_2 = 10$ shows that there is a lower bound on b_1 and we conclude that there are area specifications that cannot be represented by any convex symmetric Venn3 diagram.

Note that our proof gives some conditions that allow us to determine a area proportions that cannot be represented by any symmetric, convex Venn3 diagram (i.e. those for which the above inequality fails). Our attention now turns to identifying area proportions that can be represented.

4. A class of diagrams with $\frac{1}{3}a_3 \leq a_2 \leq a_3$

In the following three sections we discuss classes of diagrams formed from three regular polygons. Firstly we find some results for a limited set of values of a_2.

4.1. When $a_2 = a_3$

Figure 4 depicts a class of diagrams with $a_2 = a_3$. Here, the triple intersection and double intersection regions are equilateral triangles. We can choose x to be any positive real number, moving the point P along the axis of symmetry. As we change the value of x, we can increase the area a_1 from zero to any value, and the diagram remains convex throughout.

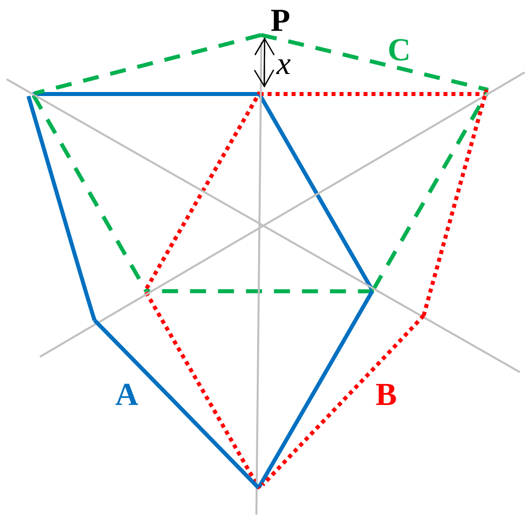

Figure 4: Diagrams with $a_2 = a_3$

We conclude that any area specifications with $a_2 = a_3$ can be drawn with a convex symmetric Venn diagram. We now generalise this class of diagrams to include some cases where $a_2 < a_3$.

4.2. When $\frac{1}{3}\boldsymbol{a_3} \leq \boldsymbol{a_2} \leq \boldsymbol{a_3}$.

Figure 5 depicts a class of diagrams, where the marked distances are constrained by the inequalities

$$x \geq 0 \text{ and } {}^{1}/_{3} \leq y \leq 1$$

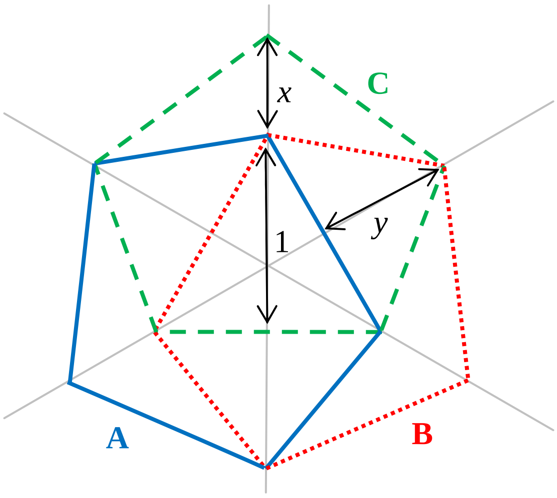

Figure 5: Diagrams with $\frac{1}{3}\boldsymbol{a_3} \leq \boldsymbol{a_2} \leq \boldsymbol{a_3}$

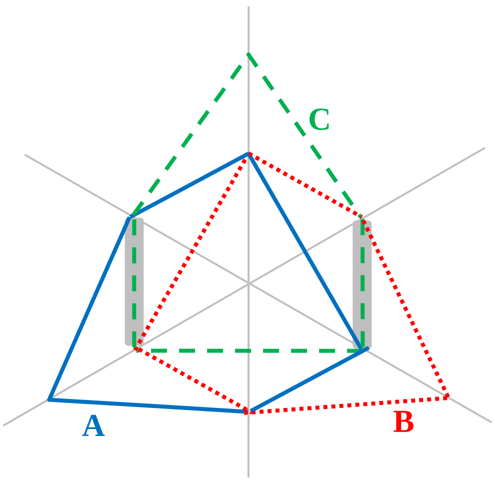

Figure 6: $\boldsymbol{a_2} = \frac{1}{3}\boldsymbol{a_3}$

The condition that $y \leq 1$ ensures that, for a fixed y, we can decrease x to zero without violating convexity of the contours.

The condition that $y \geq \frac{1}{3}$ ensures that, for a fixed value of y, x can be increased arbitrarily without violating convexity of the contours. When $y = \frac{1}{3}$, we obtain the family shown in Figure 6. Any smaller value of y would allow the lines shown with a shaded background to intersect above the diagram and convexity would impose an upper limit on x.

Hence, using the class of diagrams shown in Figure 5, we can provide a convex diagram for any area constraint that satisfies $\frac{1}{3}\boldsymbol{a_3} \leq \boldsymbol{a_2} \leq \boldsymbol{a_3}$.

5. When $\boldsymbol{a_2} > \boldsymbol{a_3}$

As shown in Section 2, there are some area specifications that cannot be drawn with a convex symmetric Venn3 diagram. The undrawable cases have large values of a_2 and small values of a_3. The classes of diagrams we have provided in Section 4 have $a_2 \leq a_3$. In this section, we describe a class for drawing some cases where $a_2 > a_3$. In these cases, a_1 has a non-zero lower bound.

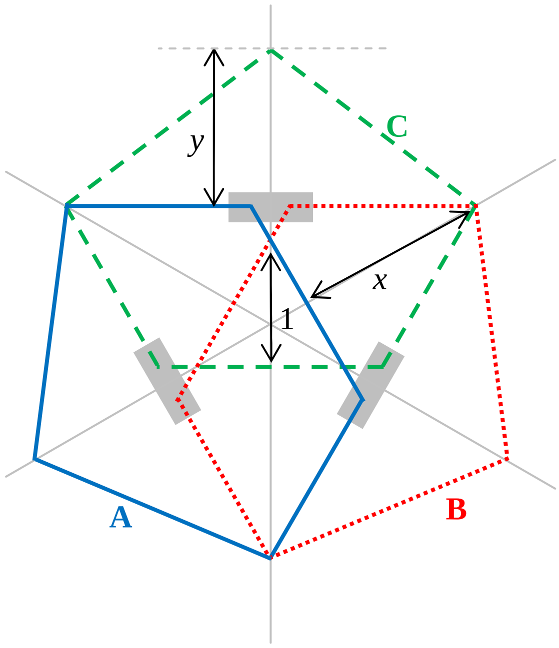

Figure 7: Diagrams with $\boldsymbol{a_2} > \boldsymbol{a_3}$

Figure 7 shows a class of diagrams similar to that of Figure 5, but the points of the polygons that border the triple intersection region are now free to be placed away from the intersection (shown with a grey background in Figure 7). We also impose the limits $x > 1$ and $y > 0$ on this diagram class.

It is clear that a_1 can be increased arbitrarily. For a fixed value of a_2 there is a lower bound on a_1 and the diagram which achieves this lower bound is shown in Figure 8. The ratios of areas in these lower bound diagrams, In fact, utilizing, geometry, we find that, when $a_2 > a_3$ the ratio $a_1 : a_2 : a_3$ is $t^2 : t^2 + 4t + 1 : 1$. For example, if we set $t = 1$ and then scale the diagram so that $a_3 = 1$ then $a_1 = 1$ and $a_2 = 7$.

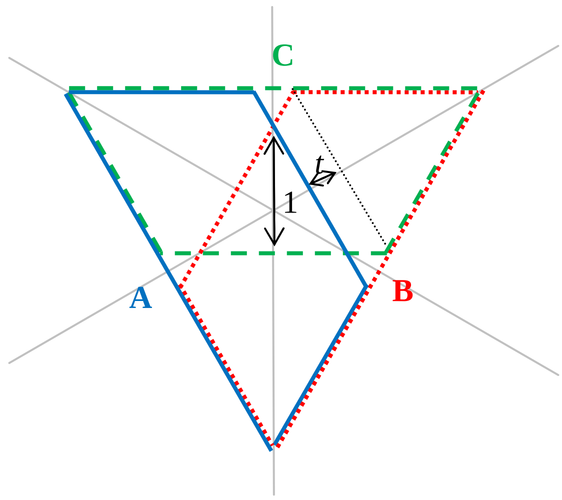

Figure 8: Minimal a_1 when $a_2 > a_3$

We note that if this diagram presents a minimal a_1 with fixed a_2 and a_3, then the formulae give a decision procedure for indicating whether a symmetric area specification can be drawn with convex curves when $a_2 > a_3$.

6. When $a_2 < \frac{1}{3}a_3$

In Sections 4 and 5 the diagrams were restricted to $a_2 \geq \frac{1}{3}a_3$. Now we address the remaining cases, where $a_2 < \frac{1}{3}a_3$. Figure 9 shows a diagram class, similar to that of Figure 3, however each curve has an extra point, so that the triple intersection region forms a regular hexagon.

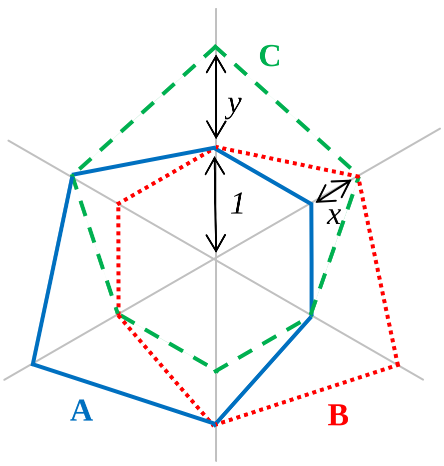

Figure 9: Diagram class for $a_2 < \frac{1}{3}a_3$

Here $x \leq 1$ and $y > 0$ and we have diagrams where $a_2 < \frac{1}{3}a_3$. Any area specification with $a_2 < \frac{1}{3}a_3$ can be drawn using a diagram from this class.

7. Open Questions

In this Section we pose some questions that arise from this work and examine the consequences of answering them.

1. **Does there exist a symmetric area specification that can be represented by a non-symmetric Venn3 that cannot be represented by a symmetric Venn3?**

If the answer to this is no, then the result in Section 3 implies that there are area specifications that cannot be represented with a convex Venn3 diagram.

2. **Does there exist a symmetric area specification that can be represented by a convex Venn3 diagram that cannot be represented by the classes shown in this paper.**

The diagram class in Figure 8 is the best we have formulated so far in terms of being able to find the lowest value for the ratio $a_1 : a_2$ when $a_2 > a_3$. However, it needs to be shown that there is no better diagram. If this can be shown, then the formulae that results from the version of this diagram with lowest a_1 (in Section 5) does not just indicate which diagrams can be represented with convex curves, but becomes a decision procedure for deciding whether an area specification can be drawn with convex curves.

8. Conclusions

In this paper we have shown that there are some symmetric area specifications that cannot be drawn with area proportional symmetric Venn3 diagrams. We have also found some results towards finding out which area specifications can be drawn, and have provided constructions for drawing the diagrams. We have also posed two open questions. The first, if answered in the negative means that, from the theorem in Section 3.2, we could deduce that there are area specifications that cannot be represented by convex Venn3 diagrams. If both questions are answered in the negative, would mean our techniques form a decision procedure for deciding whether an area specification has a convex drawing, and providing a construction for a such a area proportional convex Venn3 diagram.

Further work clearly includes addressing the questions we pose. However, this paper is limited to symmetric diagrams, and to be of practical significance we also need to investigate the non-symmetric implications. We believe that the techniques in this paper are easily applied to some non-symmetric are specifications, and we will explore this issue in detail. However, a longer term goal is to find an approach to the problem of: given any area specification, is there a convex Venn3 diagram with which it can be represented?

More generally, further work will look at Venn-like diagrams, as diagrams that have regions with zero area (commonly called ‘Euler diagrams’) can have radically different embeddings than the familiar Venn3 embedding. In addition, work on area proportional diagrams with more curves would be another important avenue to explore.

Acknowledgements

We would like to thank Marc Brodsky for communication that led to us reappraising previous assumptions about convex diagrams. This work is supported by the EPSRC under grants EP/E010393/1 and EP/E011160/1 for the Visualization with Euler Diagrams project, web site: www.eulerdiagrams.com.

References

[1] J.R. Benson, D. Crist, P. Lafleur & B.A. Watson. Agent-based visualization of streaming text. I7 Information Visualization Conference 2008 Poster.

[1] J. Carroll. Drawing Venn Triangles. Technical Report HPL-2000-73, HP Labs (2000).

[2] S. Chow. Generating and Drawing Area-Proportional Venn and Euler Diagrams. PhD Thesis. University of Victoria, Canada. 2008.

[3] S. Chow and P. Rodgers. Constructing Area-proportional Venn and Euler Diagrams with Three Circles. In Euler Diagrams Workshop 2005.

[4] S. Chow and F. Ruskey: Drawing Area-Proportional Venn and Euler Diagrams. Graph Drawing 2003. LNCS 2912: 466-477: 50.

[5] A. Coreau, J.-L. Martin. Multi-scale study of bird species distribution and of their response to vegetation change: a Mediterranean example. Landscape Ecol (2007) 22:747–764. DOI 10.1007/s10980-006-9074-2.

[6] T. Hulsen, J. de Vlieg1, W. Alkema. BioVenn – a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. BMC Genomics 2008, 9:488, doi:10.1186/1471-2164-9-488.

[7] O. Lenz A. Fornoni. Chronic kidney disease care delivered by US family medicine and internal medicine trainees: results from an online survey. BMC Medicine 2006, 4:30, doi:10.1186/1741-7015-4-30.

[8] H. A. Kestler, A. Müller, T. M. Gress, and M. Buchholz. Generalized Venn Diagrams: A new method of visualizing complex genetic set relations. Bioinformatics, 21(6):733-739, 2005.

[9] F. Ruskey. A Survey of Venn Diagrams. Electronic J. Combinatorics Dynamical Survey DS5, June 18, 2005.

[10] J. Venn. On the Diagrammatic and Mechanical Representation of Propositions and Reasonings. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 9 (1880) 1-18.