Extended Abstract: Constructing Area-Proportional Venn and Euler Diagrams with Three Circles

Stirling Chow* Peter Rodgers†

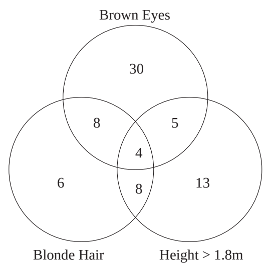

Figure 1: Typical 3-Venn diagram showing distribution of characteristics.

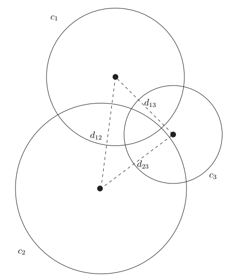

Figure 2: The triangle formed by the circles' centers.

1 Introduction

A 3-Venn diagram is used to represent all possible combinations of three characteristics and is most commonly drawn with three overlapping congruent circles as shown in Fig. 1. The diagram's zones are often labelled to indicate the size of the population with that respective combination of characteristics. If some of the combinations have zero population, they may be omitted and the resulting diagram is referred to as a 3-Euler diagram.

Suppose, in addition to having the circles overlap as required, the circles and zones are scaled according to their respective populations; such diagrams are referred to as *area-proportional* [1] and are intended to enhance readability. In [1], the authors describe an algorithm for constructing area-proportional 3-Venn diagrams using rectangles after briefly indicating that these diagrams don't generally exist for three circles.

In this paper, we more thoroughly investigate the conditions under which area-proportional circular 3-Venn diagrams exist, and for those untenable cases, we present an optimization strategy for approximating a solution. In the conclusion, we describe how these results can be extended to the case of 3-Euler diagrams.

2 Existence

Let c_1, c_2, and c_3 be the circles in a 3-Venn diagram, and let $\omega(s)$ be the population size of zone s where $s \subseteq \{1,2,3\}$ indicates the circles that exclusively enclose the zone. WLOG we assume a 1:1 scale so the actual area of c_i is

$$area(c_i) = \sum_{s \subseteq \{1,2,3\} \text{ and } i \in s} \omega(s).$$

Each pair of circles, c_i and c_j, is a 2-Venn diagram whose zone of overlap has area

$$area(c_i \cap c_j) = \sum_{s \subseteq \{1,2,3\} \text{ and } i,j \in s} \omega(s).$$

As discussed in [1], the distance d_{ij} between c_i and c_j is uniquely determined by $area(c_i \cap c_j)$; therefore, the centers of the circles of an area-proportional 3-Venn diagram form a triangle with sides d_{12}, d_{13}, and d_{23} as shown in Fig. 2. Modulo translations, rotations, and reflections, there is no freedom in positioning the circles to achieve a specific area for $c_1 \cap c_2 \cap c_3$; hence, area-proportional circular 3-Venn diagrams generally don't exist.

Based on the previous discussion, the question of whether or not there exists an area-proportional circular 3-Venn diagram for a specific ω can be answered using the following steps:

1. Compute d_{12}, d_{13}, and d_{23} using the bisection method described in [1].
2. If d_{12}, d_{13}, and d_{23} satisfy the *triangle inequality* [2], continue; otherwise, no diagram exists.
3. Compute the radii of c_1, c_2, and c_3 and place their centers at the vertices of a triangle with sides d_{12}, d_{13}, and d_{23}.
4. If $area(c_1 \cap c_2 \cap c_3) = \omega(\{1,2,3\})$ then the diagram exists; otherwise, no diagram exists.

*Department of Computer Science, University of Victoria, Canada, `schow@cs.uvic.ca`. Research supported by an NSERC CGS.

†Computing Laboratory, University of Kent, UK, `P.J.Rodgers@kent.ac.uk`.

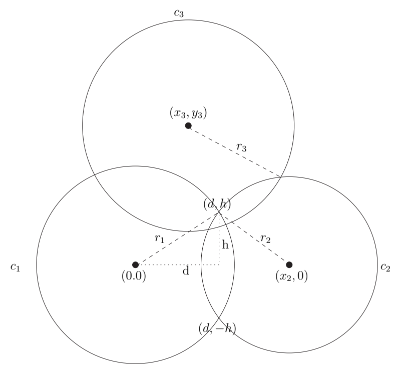

Figure 3: Layout parameters.

3 Optimization

Given the limited circumstances for existence, we now consider the question of approximating area-proportional circular 3-Venn diagrams. The approximation maintains the circles, but allows both their area and the zones' areas to deviate from ω. We begin by defining the approximation as a constraint satisfaction problem.

Let each circle c_i, $1 \le i \le 3$, be defined by center (x_i, y_i) and radius r_i. WLOG we set $(x_1, y_1) = (0, 0)$, align c_2's center along the positive x-axis, and c_3's center in the positive y quadrants (see Fig. 3).

Circles c_1 and c_2 intersect at (d, h) and $(d, -h)$ where

$$h^2 = {r_1}^2 - d^2 = {r_2}^2 - (x_2 - d)^2 \tag{1}$$

$$\Rightarrow d = \frac{{r_1}^2 - {r_2}^2 + {x_2}^2}{2x_2} \tag{2}$$

$$\Rightarrow h = \sqrt{{r_1}^2 - ({r_1}^2 - {r_2}^2 + {x_2}^2)^2/(2x_2)^2} \tag{3}$$

In order for the three circles to form a valid Venn diagram, c_1 and c_2 must intersect non-tangentially and c_3 must bisect zones $\{1\}$, $\{2\}$, and $\{1, 2\}$. The following conditions are necessary and sufficient for the proper intersection of the three circles according to this layout:

- $x_1 = y_1 = 0$
- $x_2 > 0$ and $y_2 = 0$
- $y_3 > 0$
- $|r_1 - r_2| < x_2 < r_1 + r_2$
- $(x_3 - d)^2 + (y_3 - h)^2 < {r_3}^2 < (x_3 - d)^2 + (y_3 + h)^2$

Let $\omega(s)$ be the desired area of zone s and ω_{tot} be the desired total area of the diagram. Similarly define $\alpha(s)$ and α_{tot} for the actual areas. In addition, order the zones $s_1, s_2, \ldots, s_7$ so that $i < j \Rightarrow \omega(s_i) \le \omega(s_j)$.

The goal of the optimization is to determine values for x_2, x_3, y_3, r_1, r_2, and r_3 that minimize the "difference" between the actual areas of the diagram's zones and the desired areas. For our purposes, the difference metric is a combination of variance and "out-of-orderness". As a result, the goal is to minimize the following function:

$$\Delta = \sum_{s \subseteq \{c_1, c_2, c_3\}} (\omega(s)/\omega_{tot} - \alpha(s)/\alpha_{tot})^2 + \sum_{i<j} [\alpha(s_i) > \alpha(s_j)]$$

The above fitness function is idealistic and equally weights the two components; in practice, there are many interpretations of what makes a 'good' diagram. In the following sections we describe a hill climbing approach for optimizing a more complex, but ad hoc fitness function and present examples of its effectiveness.

4 Hill Climbing

The triangle layout described in Sec. 2 is used as the starting point for the search; this produces a layout where the areas of combinations of some zones are exact, but the areas of the single zones themselves may be far from an acceptable solution. In order to improve the layout further, we have developed a metric that attempts to quantify our perception of what is a good relationship between zone areas. The optimization process changes the layout and so improves the diagram as measured by the metric.

The metric is based on our perception that it is not necessary to ensure that zone areas are exactly proportional to their required size. In fact, users will be more concerned with relative sizes. In particular, when comparing the appearance of two zones, the one with the higher population should have a larger area, and it is not particularly important how much bigger it is. Also, two zones that are more or less equal in population should appear roughly equal in size in the diagram.

In order to measure this by a metric, we perform a pairwise comparison of all zones. We calculate a value for each pair of zones. The values are summed to produce the metric, which forms the fitness for the diagram. A lower number means a 'better' diagram. For each pair, we calculate an allowable range for the relative areas. If the areas fall in this range, the value is zero, and when they fall outside the range, the value increases with the square of the distance from the allowable range. This squaring factor penalizes zone pairs falling far from the allowable range harshly, and so these pairs are more likely to be corrected than pairs with areas lying just outside the allowable range.

A pair of zones is considered to be equal if their populations are within 10% of each other. In this case, the allowable range for the area ratios is 10%. Hence zones with population 50 and 53 would have a value of 0 if their areas were 93 and 100, but if their areas were 90 and 110 then the value would be 100.

Where the zones are not considered to be equal, the allowable range for the area ratios is between $1 + 0.3 * rp$ and $1 + 2 * rp$, where rp is how much greater the large zone population is than the small zone population; that is, $rp = (largepopulation/smallpopulation) - 1$. For example, in the case where zone populations were 10 and 20, ratios for the areas between 1.3 and 3 would result in a value of 0.

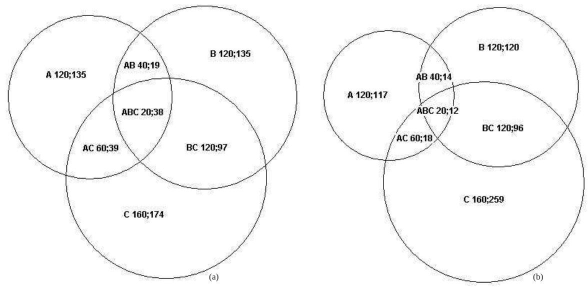

Figure 4: An example of an initial layout (a) that is improved by the hill climber (b).

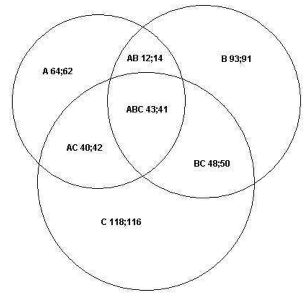

Figure 5: An example of a good initial layout that does not need improving.

In addition, we consider the two cases where zones are not equal in area when their populations are equal and where two unequal zones have an area ratio less than the allowable range (i.e. the zone with the large population has too small an area) to be of greater importance than the case where the two unequal zones have an area ratio greater than the allowable range (i.e. the zone with the large population has a too large an area). Hence we multiply the values for the first two cases by 100.

Clearly, the above metric is ad hoc. Both the computational method and the constants used in the calculation were developed after examining numerous example cases and comparing various alternative metrics.

In order to change the diagram to improve the metric, we take a hill climbing approach. This consists of a number of iterations, which move the circle centres, and change the circle radii. The movement is controlled by a cooling schedule, where the amount of movement reduces on each iteration, allowing large changes at the start of the search process and refinement of the diagram layout towards the end of the process. On each iteration each circle is modified. First the circle is moved successively in 8 possible directions: horizontally, vertically and diagonally. Then the radius is expanded and contracted. When a move improves the diagram layout as measured by the metric, the move is kept and the next circle is tried.

Moves are restricted so that they do not modify the structure of the diagram; that is, moves which add or remove zones are not made.

5 Examples

This section illustrates the method described in this paper.

Figure 4(a) shows a 3-Venn diagram after the initial layout. Each zone is labelled with the zone description followed by its population and actual area (separated by ';'). Exact diagrams would have all zones with their populations equal to their actual areas.

Figure 4(a) illustrates a layout where some of the zone areas are bad for their desired populations. AB and AC have populations much larger than ABC, but AB has a smaller area and AC has a nearly equal area. A, B, and BC have the same population, but whilst A and B have equal areas, BC's area is smaller. C has a larger population and area than the other zones, so this is a reasonable layout for C.

Figure 4(b) shows the diagram after running the hill climber. The areas of AB and AC are now bigger than ABC, and although the relative size of their populations is much larger than would be indicated by the their areas, we consider that having area relationships in correct order to be a reasonable result. A and B now have slightly different areas, but the percentage error is small and unlikely to be noticeable. BC still has a smaller area than A and B, but the difference has been reduced significantly. C's area is still, correctly, larger than the rest, although it has increased in size. In terms of fitness values, Fig. 4(a) scores 275 and Fig. 4(b) scores 18.

Figure 5 shows a diagram that has a good initial layout since most of the areas are close to their populations. Such diagrams often occur is cases where the outer zones (i.e., A, B and C) have nearly equal populations that are large compared to the inner zones' populations. The inner zones are similarly close in population, except for AB which is smaller; the layout method can usually adapt to single variations of this sort. The fitness value of the initial layout is approx. 0, and running the hill climber has very little impact on the diagram.

Figure 6(a) shows a diagram with a bad initial layout whose fitness value is 3277. When comparing zone pairs, often the zone with a larger population has a smaller area. After running the hill climber, the diagram in Fig. 6(b) still has a bad layout whose fitness value is 2009. As with this example, when some of the 2-set zones' populations are large and the 3-set zone's population is small, often no good layout exists; this demonstrates a limitation of circular 3-Venn diagrams.

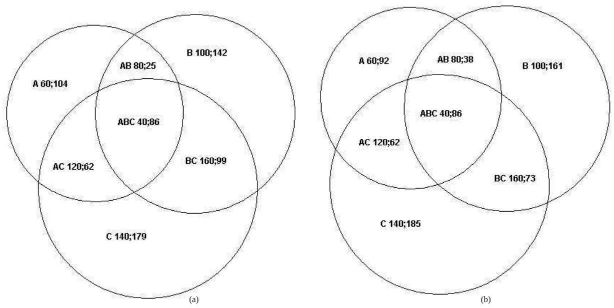

Figure 6: An example of a bad initial layout (a) that is not improved by the hill climber (b).

6 Conclusion

We have introduced a method for generating an area-proportional visualization of all intersections of three sets using circles. We have implemented the method as an applet, accessible at `http://theory.cs.uvic.ca/venn/EulerianCircles`. The method operates almost instantaneously from a user perspective and works effectively for a reasonable subset of possible populations, assuming some error in zone size is permitted.

There are several areas of improvement planned. Firstly, there are cases where no approximate solution is generated because the distances between the circles do not satisfy the triangle inequality. We could solve this by modifying the circle distances to satisfy this inequality; whilst the initial embedding would not be as desirable, in many cases, it is likely the hill climber will improve the layout adequately. Another difficulty with the initial layout is that on occasion, even when the triangle inequality is satisfied, a zone may be split into two disconnected areas; this typically occurs when the triangle formed from the centers of the circles has an angle close to 180 degrees. A solution is to again modify the circle distances so that both the triangle inequality is satisfied and one of the split areas disappears.

At the moment, we do not draw diagrams with zero size zones (i.e. we do not draw Euler diagrams, apart from those which are also Venn diagrams). A solution, drawn from the original visualization of Venn diagrams [3], is to shade these zones; however, in many cases there are simple and effective visualizations of circular Euler diagrams. For instance, the diagram A, B, C, AB, BC often has an exact solution for an area-proportional layout. Other diagrams, such as A, B, C, AB, AC, ABC, have a visualization with circles, but many population instances cannot be visualized well because the areas of the zones are highly constrained. Still others, such as the diagram AB, AC, have no visualization with three circles. Fortunately, since there are a limited number of combinations of Euler diagrams with three sets, we will tackle them on a case by case basis.

The hill climber could be improved. At the moment, a move either shifts the center of a circle in an absolute direction or changes its radius. Relative movement, such as rotation of a circle about a point, may prove useful, particularly with some Euler diagrams that have triple points (i.e., a point where all three circles meet, which occurs when visualizing diagrams such as A, B, C, AB, AC). Rotation of a circle around the triple point is likely to be the only way of moving a circle, as any other movement would break the triple point. Furthermore, the current approach of using a hill climber with a small number of iterations means we can generate diagrams in real time; however, to find a better solution, a wider search might be desirable using simulated annealing or genetic algorithms. Since drawing circular Venn diagrams is a well-defined constraint satisfaction problem, another possibility is to encode the problem in a symbolic math package such as Maple or Matlab, and use some of their optimization solvers. We also acknowledge that the metric applied in the hill climber is entirely ad hoc; whilst some alternatives have already been explored, various users may have different perceptions of what is most desirable for their diagrams, which is likely to mean the development of alternative fitness functions.

It should also be noted that circles cannot adequately draw all population instances, even when a level of error is permissible. We have concentrated on this simple shape because it is desirable for many users, and it reduces the difficulty of finding a visualization. Using other shapes, such as ovals, would increase the closeness of many diagrams to their desired area-proportionality; however, for every convex shape there is some set of desired areas that can't be represented exactly proportionally, so other methods are required for a general solution. One future goal is to develop an embedding method that produces exact area-proportional diagrams with shapes that are as smooth and regular as possible. Similarly, as the number of sets increase past three, circles are not a reasonable shape to use exclusively unless the Euler diagrams are fairly simple, hence alternative approaches are required.

References

[1] S. Chow and F. Ruskey. Drawing area-proportional Venn and Euler diagrams. In *Proc. of Graph Drawing 2003*, Spring-Verlag LNCS 2912, pages 466–477, 2004.

[2] E. Weisstein. Triangle Inequality. In Mathworld, `http://mathworld.wolfram.com/TriangleInequality.html/`.

[3] J. Venn, On the diagrammatic and mechanical representation of propositions and reasonsings. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 9, pages 1–18, 1880.