Kent Academic Repository

On the drawability of 3D Venn and Euler diagrams ☆

Jean Flower [a], Gem Stapleton [b,*], Peter Rodgers [c]

[a] *Autodesk*
[b] *Visual Modelling Group, University of Brighton, UK*
[c] *University of Kent, UK*

ARTICLE INFO

ABSTRACT

3D Euler diagrams visually represent the set-theoretic notions of intersection, containment and disjointness by using closed, orientable surfaces. In previous work, we introduced 3D Venn and Euler diagrams and formally defined them. In this paper, we consider the drawability of data sets using 3D Venn and Euler diagrams. The specific contributions are as follows. First, we demonstrate that there is more choice of layout when drawing 3D Euler diagrams than when drawing 2D Euler diagrams. These choices impact the topological adjacency properties of the diagrams and having more choice is helpful for some Euler diagram drawing algorithms. To illustrate this, we consider the well-known class of Venn-3 diagrams in detail. We then proceed to consider drawability questions centered around which data sets can be visualized when the diagrams are required to possess certain properties. We show that any diagram description can be drawn with 3D Euler diagrams that have unique labels. We then go on to define a set of necessary and sufficient conditions for wellformed drawability in 3D.

1. Introduction

Euler diagrams are a widely used notation for visualizing set theoretic relationships such as containment and disjointness. Their ability to effectively convey these relationships has inspired a large body of research. Of particular interest has been identifying which collections of sets and relationships can be visualized using Euler diagrams drawn under certain constraints, such as using curves that do not run concurrently with each other. These conditions impact on the effectiveness of the diagrams drawn and, thus, are important for usability reasons [13].

Some of the major contributions on 2D Euler diagram drawing have identified necessary and sufficient conditions for diagram drawability under particular constraints. This includes the first work, by Flower and Howse, on 2D

Euler diagram drawing, which established necessary and sufficient conditions for so-called wellformed drawability [7]. Since that first work, a number of methods have been devised to automatically draw 2D Euler diagrams under varying sets of wellformedness properties, including research by Chow and Ruskey [3], Kestler et al. [8], Rodgers et al. [12], Simonetto et al. [17], Stapleton et al. [19] and Wilkinson [25]. In addition, the impact that various properties have on user understanding has been studied empirically [5,13]. To summarize key findings from this existing work: it is known that drawing wellformed 2D Euler diagrams is important for usability, we have necessary and sufficient conditions for wellformed drawability, some sets can only be visualized with non-wellformed diagrams, and a variety of methods for automatically drawing Euler diagrams in 2D exist.

Recent advances in technology include 3D televisions and 3D interfaces such as Microsoft Kinect, as well as 3D printing technology. This means that visualization in 3D has the potential to become more mainstream, leading to a requirement for a better understanding of what can be visualized in 3D and how to produce those visualizations.
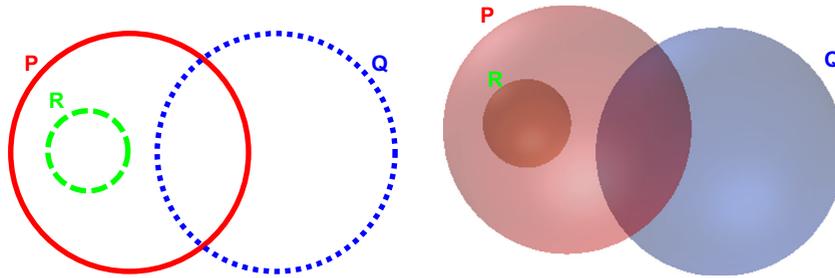
Fig. 1. A 2D Euler diagram with an equivalent 3D Euler diagram.

Thus, it seems timely to extend Euler diagrams to 3D. Fig. 1 shows a 2D and a 3D Euler diagram, both of which represent the same information.

In [11], we introduced 3D Euler diagrams. Rather than closed curves (used in 2D Euler diagrams), they use surfaces to represent sets. We generalized the 2D well-formedness properties to 3D and established that every wellformed 2D Euler diagram can also be drawn well-formed in 3D. In this paper, we extend the theoretical investigations of 3D Euler diagrams undertaken in [11]. In Section 2 we present some definitions that are required in the remainder of the paper. In Section 3, we discuss the importance of the topological properties of Euler diagrams (in both 2D and 3D). In particular, we demonstrate that there are more topologically different representations of sets in 3D than in 2D, which is important for some automated drawing methods. We provide a classification theorem for 3D Venn-3 drawn with surfaces equivalent to spheres. By contrast to the 2D case, where there is only one topologically distinct drawing of a wellformed Venn-3 diagram, we prove that there are four equivalence classes of wellformed 3D Venn-3 diagrams drawn with surfaces equivalent to spheres.

In Section 4, we provide a series of drawability results, establishing necessary and sufficient conditions for draw-ability under varying wellformedness properties. This includes demonstrating that any diagram description can be drawn with a 3D Euler diagram that has unique labels. Section 4 culminates in providing necessary and sufficient conditions for wellformed drawability of Euler diagrams in 3D. Further, we establish that our conditions are also necessary, but not sufficient in 2D. Consequently, the sets that can be visualized wellformed in 2D can all be visualized wellformed in 3D but not vice versa. This could be a major advantage for 3D Euler diagrams, although empirical studies are needed to determine how their effectiveness as visuali-zations of sets compares to 2D Euler diagrams.

Finally, in Section 5, we give our conclusions and discuss further work. We detail some open questions in the theory of 3D Euler diagrams and we discuss some directions for research that might demonstrate the usabil-ity of 3D Euler diagrams in practical situations.

2. Definitions of 3D Venn and Euler diagrams

Here we present the core definitions needed through-out the paper for our study of 3D Euler diagrams. They are extended from definitions given in [11]. We refer the
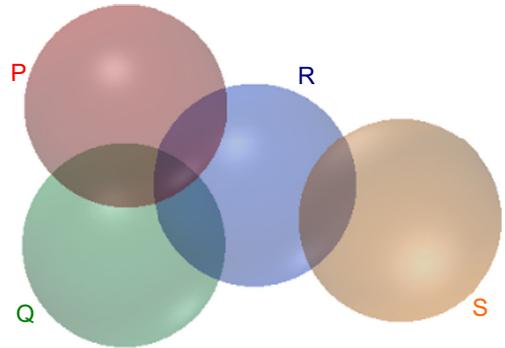
Fig. 2. A 3D Euler diagram.

reader to [20] for a formal definition of 2D Euler diagrams and associated wellformedness properties.

3D Euler diagrams are made up of a set of closed surfaces[1] embedded in $\mathbb{R}^3$. The surfaces are assigned labels from a set $\mathcal{L}$. An example can be seen in Fig. 2, which has four surfaces, each with a distinct label.

Definition 2.1. A **3D Euler diagram** is a pair, $d = (\mathcal{S}, l)$, where

1. $\mathcal{S}$ is a finite set of closed surfaces embedded in $\mathbb{R}^3$, and
2. $l : \mathcal{S} \to \mathcal{L}$ is a function that labels each surface.[2]

Requiring surfaces to be closed and embedded in $\mathbb{R}^3$ implies that they are orientable and gives us a well-understood notion of what constitutes the interior.

Given an Euler diagram, $d = (\mathcal{S}, l)$, and a label, $L \in \mathcal{L}$, we define the *contour* in d with label L to be the largest set of surfaces in d that have label L. The set of d's contours is denoted $\mathcal{C}$. We extend l, so that given a contour, c, $l(c) = L$. A point is *inside* a contour whenever it is inside an odd number of its surfaces, otherwise the point is *outside* the contour.

---

[1] Formally, we consider a surface, S, to be function that embeds a closed, orientable 2-manifold in $\mathbb{R}^3$, although we find it convenient to blur the distinction between S and the set of points to which S maps.

[2] In [11], we required l to be injective, for simplicity. Here, we remove this constraint, and promote the injective property to be a wellformed-ness condition (see Section 2.1), as has often been the case for 2D.

To illustrate, the diagram in Fig. 3 has two contours, P and Q. The contour P, unlike the contour Q, comprises two surfaces. The points that are inside either one of the surfaces labeled P are inside the contour labeled P.

The semantics of the diagram are captured precisely by its *zones*. A zone is a region in the diagram that is described as being inside some (or no) contours and outside the rest of the contours. In Fig. 2, the 3D diagram has ten zones. Between them, the ten zones represent all of the non-empty set intersections. So, for example, $R \cap S = \varnothing$.

Definition 2.2. A *zone*, z, in a 3D Euler diagram, $d = (\mathcal{S}, l)$, is a set of points in $\mathbb{R}^3$ for which there exists a set, $C \subseteq \mathcal{C}$, of d's contours such that

1. every point, p_{in}, in z is inside all of the contours in C and outside all of d's remaining contours, and
2. z is maximal with this property.

Such a zone, z, is *described* by $\{l(c) : c \in C\}$. The set of zones in d is denoted $\mathcal{Z}(d)$. A *minimal region* of a zone, z, is a maximal connected subset of z.

All zones in Fig. 2 are connected and, thus, comprise a single minimal region. In Fig. 3, however, the zones are not connected. The diagram has 4 zones, but 6 minimal regions, with the zones with descriptions $\{P\}$ and $\{P, Q\}$ each comprising two minimal regions.

Definition 2.3. A *diagram description*, D, is a subset of $\mathbb{P}\mathcal{L}$ that includes $\varnothing$. Given a 3D Euler diagram, d, its description is the set of the descriptions of d's zones.

For example, the diagram in Fig. 2 has the description $\{\varnothing, \{P\}, \{Q\}, \{R\}, \{S\}, \{P, Q\}, \{P, R\}, \{Q, R\}, \{Q, S\}, \{P, Q, R\}\}$.
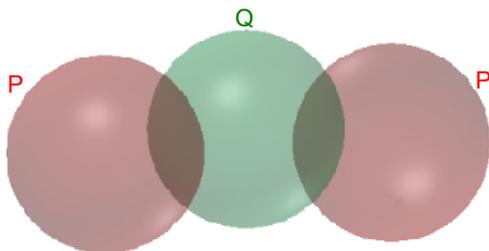
Fig. 3. Illustrating contours.

We will sometimes abuse notation by writing the zone description $\{P, Q\}$ as PQ, for example. Further, we will blur the distinction between a zone and its description, so referring to a zone with description PQ as the zone PQ.

The concept of two zones being *adjacent* will be important in later sections of the paper. There are two notions of adjacency, one relying solely on the zone descriptions and another relying on the zones as they appear in $\mathbb{R}^3$. The two notions of adjacency are related, but not equivalent.

Definition 2.4. If a pair of zones, z_1 and z_2, have descriptions whose symmetric difference contains exactly one label then they are *combinatorially adjacent*. If z_1 and z_2 have boundaries whose intersection includes a set of points that form a (possibly open or disconnected) surface then z_1 and z_2 are *topologically adjacent*.

For example, in Fig. 2 the two zones in the 3D diagram with descriptions P and PQR are neither topologically adjacent nor combinatorially adjacent. The zones P and PQ are both topologically adjacent and combinatorially adjacent. In this diagram, all pairs of combinatorially adjacent zones are also topologically adjacent. However in general, combinatorial adjacency does not imply topological adjacency or vice versa. In wellformed diagrams (defined below), though, topological adjacency implies combinatorial adjacency.

2.1. Wellformedness properties of 3D Euler diagrams

There are various wellformedness properties that can be possessed by 2D Euler diagrams [20]. In [11], we demonstrated how they generalized to 3D Euler diagrams. Here we simply summarize them in Table 1, and give examples in Table 2.

Definition 2.5. A 3D Euler diagram, d, is *wellformed* provided all of the wellformedness properties hold.

In 2D, studies have shown that diagrams with various wellformedness properties are more effective for data analysis [13] than diagrams without those wellformedness properties. We conjecture that wellformed 3D Euler diagrams will also be more effective visualizations of data than diagrams which fail to be wellformed.

Table 1
Wellformedness properties.

Property	2D case	3D case
Unique labels	No two curves have the same label (l is injective)	No two surfaces have the same label (l is injective)
Connected zones	Every zone is a connected component of $\mathbb{R}^2$	Every zone is a connected component of $\mathbb{R}^3$
n-point	Every point in $\mathbb{R}^2$ is passed through at most $n = 2$ times by the curves	Every point in $\mathbb{R}^3$ is passed through at most $n = 3$ times by the surfaces
Crossings	Whenever two curves intersect, they cross transversely	Whenever two surfaces intersect, they cross transversely
Line concurrency	No two curves share a common line segment	No three surfaces share a common line segment
Surface concurrency	N/A	No two surfaces share a set of points that form a disc

Table 2
Examples of non-wellformed diagrams.

Property	2D case	3D case
Connected zones	 The zone PQ is disconnected	 Here, P is a sphere with a 'sausage', Q, through it. The zone inside Q and outside P is disconnected
n-point	 The curves P, Q, and R form two 3-points	 The spheres P, Q, and R form a 4-point where they all intersect with S
Crossings	 The curves P and Q intersect at a point where they do not cross (as do R and S)	 The sphere R intersects with Q but does not cross Q; a cross-section is shown on the right
Line concurrency	 The two curves Q and R share a common line segment	 The three tori share a common line segment; a cross-section is shown on the right
Surface concurrency	N/A	 The two 'squashed' spheres share a disc-like surface
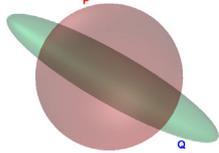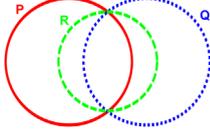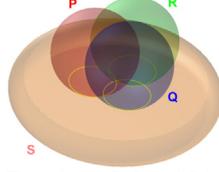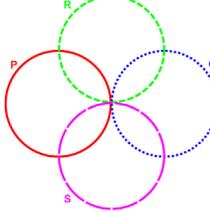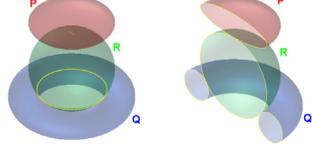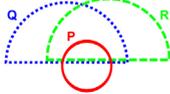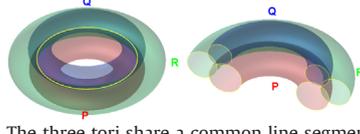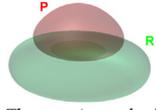
2.2. The drawability problem

A major theme of Euler diagrams research has been on establishing when a diagram exists as a visualization of information, often under some additional constraints including being wellformed [7], or being drawable with circles [22]. Moreover, algorithms for automatically producing Euler diagrams, given the information to be visualized, are sought and a number of them now exist for the 2D case, including [3,7,8,14,16,17,19,24,25]. These algorithms require a description of a required diagram, d, to be provided which typically comprises precisely the descriptions of the zones in d.

We can now state the classic Euler diagram drawability problem, generalized to 3D:

given a diagram description, D, draw a 3D Euler diagram, d, with description D such that d satisfies some specified conditions.

The conditions that are often enforced are that all, or a subset of, the wellformedness properties are possessed by d. For instance, we may wish to find a diagram that has no concurrency between surfaces. Whatever set of conditions has been specified, we will call this set the *drawability constraints*. The remainder of this paper is largely focused on providing drawability results in 3D, as well as establishing that more descriptions are drawable in 3D than are drawable in 2D.

3. Choices of representation

In this section, we provide results about zone topological adjacency. In particular, we demonstrate two key results:

1. For all diagram descriptions, D, if D can be drawn in 2D then there exists a 3D diagram with description D, with

the same zone topological adjacency properties and the same wellformedness properties as the 2D diagram. This is captured in Theorem 3.1.

2. There exists a diagram description, D, such that D can be drawn in 3D and there does not exist a 2D diagram with the same description, the same zone topological adjacency properties and wellformedness properties as the 3D diagram.

The significance of these results lies in the demonstration that there are fundamentally more choices of diagram in 3D than in 2D.

3.1. Diagram equivalences

Given a diagram description, D, there are infinitely many different Euler diagrams that are drawings of D (or none, if it cannot be drawn given the drawability constraints). This can be seen by taking a drawing of D and tweaking its layout slightly, altering its geometry. The top left and top middle diagrams in Fig. 4 are geometrically different, but have the same description and possess the same wellformedness properties. Changing the geometry to produce a different layout can impact on both the usability and the aesthetic quality of the diagram, but it does not impact on the drawability problem when the drawability constraints are, for example, that the diagram is wellformed.

The fundamental properties of an Euler diagram that *do* impact on drawability, wellformed or otherwise, arise from the topological structure of the diagram rather than its geometric properties. Informally, two Euler diagrams are topologically equivalent whenever we can convert one of the diagrams into the other diagram by a continuous distortion of the space in which the diagram is embedded. More formally, recalling that surfaces are functions:

Definition 3.1. Let $d_1 = (\mathcal{S}_1, l_1)$ and $d_2 = (\mathcal{S}_2, l_2)$ be 3D Euler diagrams. Then d_1 and d_2 are *topologically equivalent* provided that there exists a continuous function, $F : \mathbb{R}^3 \times [0,1] \to \mathbb{R}^3$, and a bijection, $\sigma : \mathcal{S}_1 \to \mathcal{S}_2$, such that

1. restricting the domain of F to $\mathbb{R}^3 \times \{0\}$ yields the identity map, that is $F(p, 0) = p$ for all points p,

2. for each $S \in \mathcal{S}_1$, F is an ambient isotopy[3] where, given the function $f : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $f(p) = F(p, 1)$, $f \circ S = \sigma(S)$ and

3. for each $S \in \mathcal{S}_1$, $l_1(S) = l_2(\sigma(S))$.

If all but condition 3 hold then d_1 and d_2 are *topologically equivalent up to labeling*.

The conditions of the above definition ensure that F continuously transforms each surface, S, into $\sigma(S)$ with condition 3 simply ensuring that the labels coincide.

To show that two diagrams are topologically equivalent, we can demonstrate (or imagine) a continuous transformation from one to the other. To show that two diagrams are topologically different, we use topological invariants. These are diagram properties that do not change under a topological equivalence. Some examples of topological invariants are

1. the number of surfaces in the diagram,
2. the diagram description,
3. the connectedness of zones (the number of minimal regions of each zone),
4. the zone topological adjacency relations,
5. whether a zone is simply connected (e.g. the inside of a torus is not simply connected), and
6. the Euler characteristic of the surfaces and the boundaries of the zones.

The top left and top right diagrams in Fig. 4 have the same description and zone topological adjacency relationships but the surfaces labeled P have different Euler characteristics, so these diagrams are different. The top left and bottom left diagrams have different zone adjacency (zones PR and R). In addition, the zone Q is connected in both of these diagrams but it is not simply connected in the bottom left diagram where the surface R is a tube forming a tunnel through the zone Q. Either of these differences is enough to show topological distinctness. The bottom left and bottom right diagrams are also topologically different, but none of the topological invariants listed above justify this. It is not possible to undo the knot with a continuous transformation of space whilst preserving the diagram's structure.

An understanding of the range of topologically different diagrams with a given description is important for researchers developing inductive drawing algorithms [4,19,21–23]. These inductive (or incremental) approaches add one contour at a time to a drawing, building up the diagram until all of the contours are present. A choice of diagram at each stage may impact upon whether we can add the next contour in the required manner. If the contour cannot be added and the diagram is topologically unique then we cannot produce the required diagram following that incremental path. However, if the diagram is not topologically unique then an alternative choice of diagram may allow the contour to be added. If we have multiple topologically different diagrams with a given
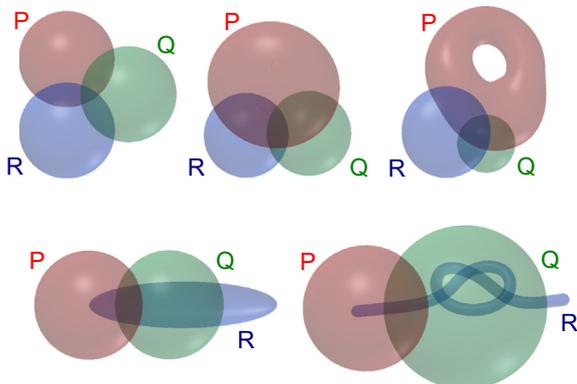
Fig. 4. Some 3D diagrams for Venn-3.

---

[3] See [2] for a definition of ambient isotopy.
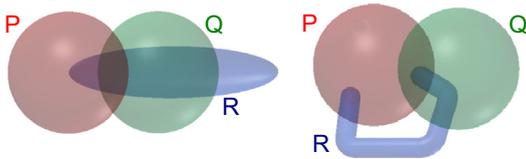
Fig. 5. Choices in diagram layout.

description then we have potentially enlarged (and certainly not shrunk) the set of diagrams that can be drawn incrementally, given the drawability constraints in question.

An example is given in Fig. 5. In the lefthand diagram, the two zones PR and PQR are not topologically adjacent whereas they are topologically adjacent in the righthand diagram; these diagrams have the same description. We can add a surface S to the righthand diagram so that it splits only PR and PQR into two zones whilst preserving wellformedness. However, we cannot add a similar surface to the lefthand diagram without either splitting other zones or breaking wellformedness. Using 3D Euler diagrams, we can represent more diagram descriptions than in 2D, and we can represent them in more different ways. We now capture when two diagrams have the same zone topological adjacency properties and wellformedness properties.

Definition 3.2. Let d_1 and d_2 be diagrams both with description D. For all zone descriptions Z and Z' in D, if the zones with those descriptions are topologically adjacent in both d_1 or d_2 or in neither of d_1 and d_2 then d_1 and d_2 are *topological adjacency equivalent*.

Definition 3.3. Let d_2 be a 2D diagram and let d_3 be a 3D diagram, both with description D. We say that d_2 and d_3 are *wellformed equivalent* provided they possess essentially the same wellformedness properties, ignoring any surface concurrency.

3.2. 2D to 3D: topological zone adjacency and wellformedness

We now prove our first main result of this section, that diagrams drawn in 2D can be drawn in 3D with the same topological zone adjacency properties and the same wellformedness properties.

Theorem 3.1. *Let d_2 be a 2D Euler diagram with description D. Then there exists a 3D Euler diagram, d_3, with description D such that d_2 and d_3 are topological adjacency equivalent and are wellformed equivalent.*

Proof. Consider d_2 and choose a straight line in $\mathbb{R}^2$ such that all of d_2 is on one side of the line. Rotate d_2 around the line by 2π, thus converting each simple closed curve into a torus. This yields d_3. It is trivial to verify that d_3 and d_2 have the same diagram descriptions and that they are topological adjacency equivalent and wellformed equivalent. □

3.3. 3D: more topological zone adjacency and wellformedness choices

We now proceed to demonstrate that the following proposition is *false*:

Proposition 1. *Let d_3 be a 3D Euler diagram with description D. Then there exists a 2D Euler diagram, d_2, with description D such that d_2 and d_3 are topological adjacency equivalent and are wellformed equivalent.*

The simplest way to prove this is to present a 3D Euler diagram which has no 2D counterpart: the description

$$\{\varnothing, P, Q, R, PQ, PR, QR\}$$

can be drawn wellformed in 3D but there is no corresponding 2D diagram. Using results in [7], it can be shown that any 2D diagram with this description is not wellformed. Conditions for 2D wellformed drawability that draw upon results in this paper are further discussed in Section 4.4.3.

A key example for contrasting 2D and 3D choices of representation is the Venn diagram on 3 contours, so we will use it to investigate differences between 2D and 3D. A Venn diagram is an Euler diagram in which all 2^n possible zones are present, where n is the number of contours.

Definition 3.4. A *3D Venn diagram*, $d = (\mathcal{S}, l)$, is a 3D Euler diagram where there are $2^{|C|}$ zones.

Every wellformed 2D Euler diagram on 3 contours is topologically unique. This was established for wellformed 2D Venn-3 in [15], where the notation Venn-n denotes a Venn diagram with n contours (Fig. 6).

Lemma 3.1. *There is only one topologically distinct wellformed 2D Venn-3 [15].*

By contrast, wellformed 3D Venn-3 diagrams are far from unique. It is helpful for us to define a Venn-3 diagram drawn with spheres (where we mean actual spheres and not surfaces topologically equivalent to spheres) as a *standard* Venn-3. An example of a standard Venn-3 can be seen in Fig. 7. There is only one standard Venn-3 up to topological equivalence, ignoring labels, so we will simply say *the* standard Venn-3.
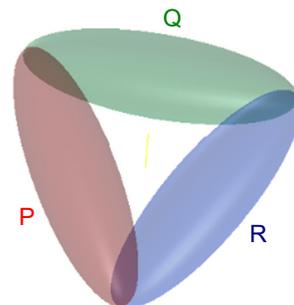
Fig. 6. A wellformed 3D diagram.

In fact, there are an infinite number of different representations of wellformed Venn-3 in 3D as we can trivially convert Venn-3 into a topologically different diagram by adding handles to any of the surfaces, whilst retaining the same diagram description. This is illustrated in Fig. 8 where three handles have been added to the diagram shown in Fig. 7. In fact, this argument applies in to wellformed Euler diagrams in general:

Lemma 3.2. *Let D be a diagram description such that there exists a wellformed diagram, d, with description D. Then there exist infinitely many wellformed diagrams with description D.*

Proof (*Sketch*). Add handles to some surface and argue about the Euler characteristic. □

Lemma 3.2 begins to give us an idea of the differences that exist between 2D and 3D representations of Euler diagrams. However, it is not terribly insightful to generate topologically distinct diagrams by simply adding handles to surfaces that are already drawn: adding handles does not alter the topological adjacency of zones, which is important for drawability as discussed above. Adding a handle is simply a theoretical construct which serves to create a different diagram, it does not make the diagram any more usable. Thus, it seems interesting to ask whether for a diagram description there are wellformed diagrams that are not topologically equivalent to each other but whose surfaces are all topologically equivalent to spheres. This prevents us from considering cases where we merely add handles to surfaces.

Are there 3D Venn-3 diagrams without handles but where the topological adjacency of zones differ? We answered this question in [11], showing that there are Venn-3 diagrams whose surfaces are all equivalent to spheres but where the diagrams are not topologically equivalent to the standard Venn-3. Moreover, we established that some topologically different representations have different zone adjacency properties.
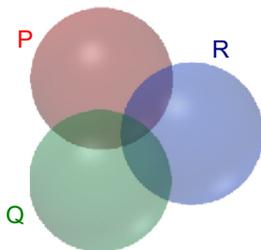
Fig. 7. A standard 3D Venn-3 diagram.
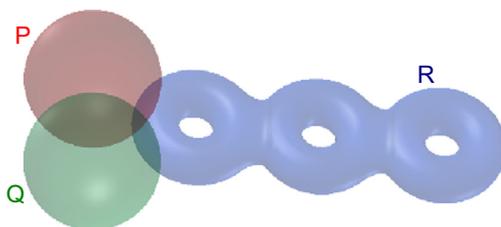
Fig. 8. Creating topologically distinct 3D Venn-3 diagrams.

Definition 3.5. A *simply connected*[4] surface in a 3D Euler diagram is topologically equivalent to a sphere.

We now present the first results which classify and provide a construction for *all* possible wellformed drawings of 3D Venn-3 using simply connected surfaces, up to topological equivalence. Given a wellformed Venn-3 with simply connected surfaces, its zone topological adjacency properties identify in which of the classes it sits. The remainder of this section is devoted to this classification result.

3.3.1. The classification theorem for wellformed 3D Venn-3 diagrams with simply connected surfaces

We will prove that every wellformed 3D Venn-3 diagram with simply connected surfaces is topologically equivalent, up to labeling, to either the standard Venn-3, which is constructed from three spheres, or a diagram constructed from a tube, T, with two knots,[5] K_1 and K_2 (see Fig. 9, which shows T containing two trefoil knots; in general the knots can be arbitrarily complex), and two simply connected surfaces, P and Q, combined together in one of three ways:

1. *Class* 1: Cap the tube T at its ends to create R; add a simply connected surface, P, containing both knots K_1 and K_2 and all of T that is between the knots, but not containing either end of T; add the second simply connected surface, Q, containing exactly one knot and the adjacent end of T. This construction is illustrated in Fig. 10 which also shows the simplest diagram in this class, where the two knots are the unknot.
2. *Class* 2: Cap the tube T at its ends to create R; add a simply connected surface, P, containing both ends of the tube but neither of the knots or any part of the tube between the knots; add the second simply connected surface, Q, containing exactly one knot and the adjacent end of T. This construction is illustrated in Fig. 11 which also shows the simplest diagram in this class, where the two knots are the unknot.
3. *Class* 3: Cap the tube at one end, but stretch the other end and extend it back over the tube and then cap it to create R; add a simply connected surface, P, containing both knots and all of the tube between the knots; add the second simply connected surface, Q, containing the end of T that was capped first and the adjacent knot. This construction is illustrated in Fig. 12 which also shows the simplest diagram in this class, where the two knots are the unknot.

We define *class* 0 to be the collection of diagrams topologically equivalent to the standard Venn-3.

---

[4] This is not, in general, equivalent to the definition of a simply connected topological space (a space is simply connected provided any embedding of a simple closed curve in to the space can be continuously transformed to a constant map into the space, see, for example, [18]). However, in the context of closed surfaces embedded in $\mathbb{R}^3$, which is the case for all surfaces in 3D Euler diagrams, the definitions are equivalent.

[5] We are using the word 'knot' for ease of understanding, but a knot theorist would require knots to have no end points. Knot theorists would call our 'knots' tangles [1].

Fig. 9. A tube with two knots.
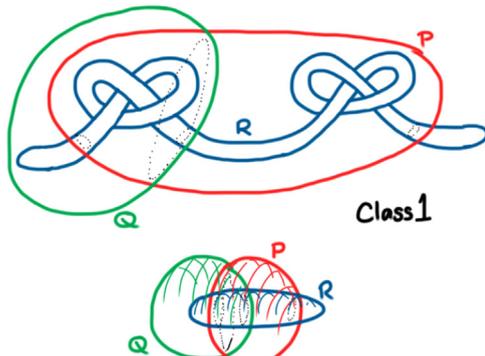
Fig. 10. Class 1 Venn-3 diagrams.
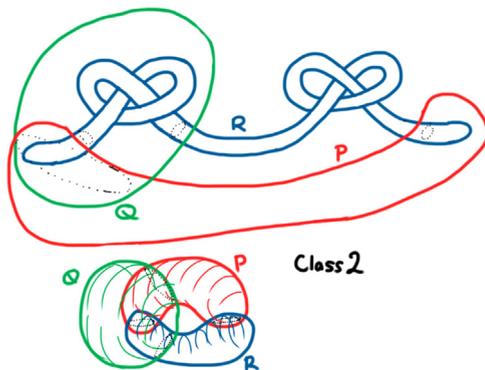
Fig. 11. Class 2 Venn-3 diagrams.
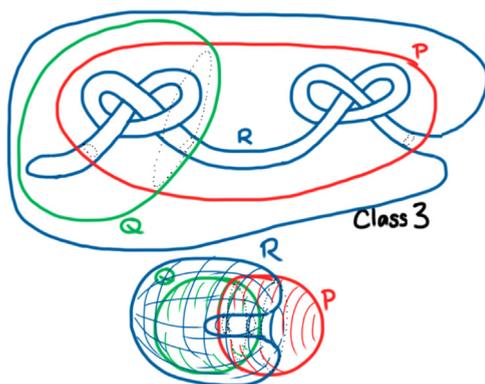
Fig. 12. Class 3 Venn-3 diagrams.

Theorem 3.2. *Let* $d = (\{P, Q, R\}, l)$ *be a wellformed Venn-3 diagram whose surfaces are simply connected. Then d is in either class* 0, *class* 1, *class* 2 *or class* 3.

The proof of Theorem 3.2 is contained in the appendix, and is preceded by a sequence of supporting lemmas. The

proof follows a rather lengthy case-by-case argument and it is not clear how readily it might generalize to proving similar results.

The two knots can be arbitrarily complex and although the resulting Venn-3 diagrams might seem quite unusable for, say, data visualization purposes, it is important to document the existence of such diagrams not least to prevent researchers from making unsupported claims about the topological layout of 3D Euler diagrams in general. Moreover, such documentation helps us to understand the topological structure of the zones in 3D Euler diagrams. We can see that the boundaries of the zones in Venn-3s from these classes are not always equivalent to spheres and those zones that are not simply connected can have boundaries that are more complex than a torus.

A classification becomes powerful when we are able to take an arbitrary Venn-3 diagram which probably looks, visually, nothing like any of these constructions and identify to which class it belongs. Theorem 5.2 in the appendix demonstrates how to identify to which class a Venn-3 belongs.

In this section we have provided a classification of all Venn-3 diagrams whose surfaces are simply connected. The steps involved allowed us to demonstrate that we can draw Venn-3 with different zone topological adjacency properties which is important for drawability. In addition, we have also demonstrated that the zone topological adjacency properties in 2D can always be achieved in 3D. In conclusion, therefore, in 3D we can draw a superset of the diagram descriptions that can be drawn in 2D under constraints such as being wellformed.

4. Drawability

We will now proceed to derive some general drawability results for 3D Euler diagrams. Firstly, we justify that any diagram description can be drawn with unique labels and connected zones. We do this by providing a construction mechanism, although this mechanism leads to diagrams that may be difficult to use in practice because of the large amount of concurrency between surfaces that results.

A further key contribution of this section is to provide necessary and sufficient conditions that encapsulate when diagram descriptions can be drawn wellformed in 3D. In the build-up to this result, we present a series of weaker results, where only a subset of the wellformedness properties are required.

4.1. Everything can be drawn

In this section, we demonstrate that all diagram descriptions can be drawn by providing a construction algorithm. The diagram construction process ensures that the diagram has unique labels and that the zones are connected.

Lemma 4.1. *Let D be a diagram description. Then there exists a* 3D *Euler diagram, d, with description D such that d has unique labels and connected zones.*

Proof. Given D, embed $n = |D - \{\varnothing\}|$ disjoint spheres in $\mathbb{R}^3$. Add a network of non-overlapping tubes to join each of these spheres to every other sphere. Each sphere gains $n-1$ tubes attached. Also, for each tube, T, choose a point mid-way along T, and add a disc which cuts T at that point:
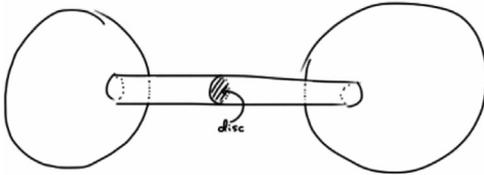
This set of punctured spheres, tubes and discs (which form caps for the two halves of the tubes) is the template for the 3D Euler diagram. For each contour label, L, build the associated surface by choosing the punctured spheres that correspond to zone descriptions in D that contain L. For any two of these chosen spheres that are joined by a tube, also choose the entire tube between them. For the remaining tubes attached to the chosen spheres, choose just the part of the tube attached to the sphere up to, and including, the cap. These chosen items combine to make the contour with label L. The resulting set of contours is a drawing of D with unique labels and connected zones. □

An example can be seen in Fig. 13, where the diagram description $\{\varnothing, PQ, PR, PS, QST\}$ is drawn using four spheres joined by tubes; the resulting diagram is shown on the left. The righthand side of the figure zooms in to show how two of the spheres (for PQ and for QST) give rise to concurrent contours, pulled apart a little for visual clarity. All but the most trivial diagrams constructed using this approach will fail to be wellformed because of concurrency between surfaces, and other wellformedness properties are likely to fail too.

4.2. Conditions for wellformed drawability except for non-unique labels

In our buildup to presenting necessary and sufficient conditions for wellformed drawability in 3D, this section focuses on the problem of drawing diagrams where we allow contours to comprise multiple surfaces, that is the unique labels property need not be possessed, but all other wellformedness properties are possessed. Our strategy again is to provide a construction of the required diagram. This result is somewhat interesting in its own right, but
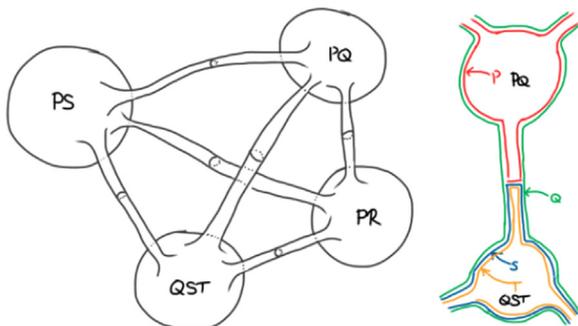
our primary motivation for including it is that afterwards, for our main drawability result (Theorem 4.2), we will extend the construction further to build a completely wellformed diagram, given any description which is capable of being drawn wellformed.

The constructive approach used for the proof of Lemma 4.2 below uses the superdual [7] which can be constructed from a diagram description.

Definition 4.1. Let D be a diagram description. The *superdual* of D is a graph, $G = (V, E)$ where $V = D$ and there is an edge between any two vertices whose symmetric difference contains exactly one label.

An example of the superdual for Venn-3's description can be seen in Fig. 14. An edge between two zone descriptions represents the fact that zones with those descriptions are combinatorially adjacent.

We show that the *only* condition required for drawability with, possibly, non-unique labels is that the superdual is connected. To illustrate the proof strategy, consider the diagram description

$$D = \{\varnothing, \{P\}, \{Q\}, \{P, Q\}, \{P, R\}, \{P, Q, R\}\}.$$

The (connected) superdual, G, of D is shown in Fig. 15. The first step in the construction is to choose a spanning tree of G, as shown in Fig. 16. We then turn the tree into a directed, rooted tree, with $\varnothing$ as the root node, see Fig. 17. Thus, each vertex is on a set of paths from $\varnothing$ to some of the leaf nodes. The next stage is to create spheres around the leaves of the directed spanning tree. These spheres are contours in the to-be-created 3D Euler diagram and the label of each sphere is the symmetric difference of the vertices incident with the edge of the spanning tree cut by the sphere; see the top left diagram in Fig. 18, where the lefthand sphere will be labeled Q and the righthand sphere will be labeled R. Next, we iteratively choose vertices,
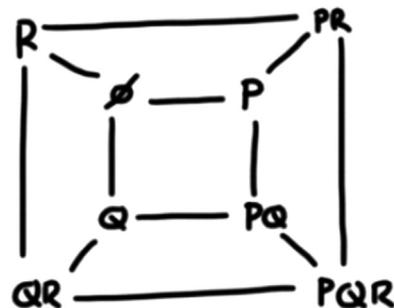
Fig. 14. The superdual of Venn-3.

Fig. 15. A connected superdual.

Fig. 13. Constructing 3D Euler diagrams from diagram descriptions.
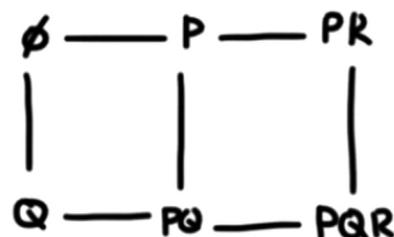
v say, other than $\varnothing$, such that all directed paths from v to leaves contain only vertices enclosed by surfaces. In our example, we choose the vertex PQR for this step and we start by enclosing it with a sphere, just as we did for the leaves. There is only one path from PQR to a leaf and we fatten the surface containing that leaf to create a new sphere. The sphere around PQR and the fattened sphere around PQ are then joined by a tube that encloses the edge between PQ and PQR; this is illustrated in the top right diagram in Fig. 18. This process has created another surface and, this time, we label it Q; the spanning tree edge cut by the surface is incident with two vertices whose symmetric difference is Q. This process is repeated until all vertices are enclosed by surfaces, except for $\varnothing$. The final diagram can be seen in Fig. 19 (the diagram with the spanning tree is on the left and the final diagram is on the right), where the contour Q comprises two surfaces. The unique labels condition is the only wellformedness condition broken because of the way in which the diagram was formed. In the general case, this construction method can be more complex when there is more than one directed path from a

Fig. 16. A spanning tree of a connected superdual.

Fig. 17. A directed spanning tree of a connected superdual.

Fig. 18. Fattening subtrees of the chosen spanning tree to create surfaces.
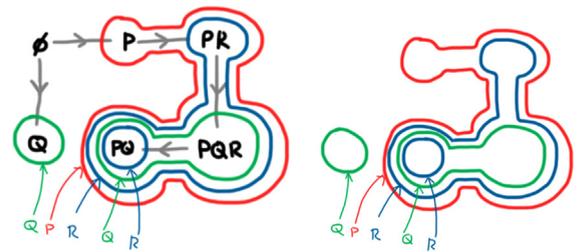
Fig. 19. A diagram created from a spanning tree.

vertex to more than one leaf. This eventuality is covered in the proof of Lemma 4.2 below.

For Lemma 4.2 we need to find a spanning tree of a connected superdual to show that the required diagram exists. For our results on wellformed drawability below we need to make use of subgraphs of the superdual. Since the proof of our wellformed drawability result (Theorem 4.2) directly extends the construction that we present in the proof of Lemma 4.2, we state that lemma using subgraphs. These are not arbitrary subgraphs, though they must contain all of the vertices

Definition 4.2. Let D be a diagram description with superdual $G = (V, E)$. A *spanning dual* of G, and of D, is a graph, $G' = (V, E')$ where $E' \subseteq E$.

To proceed with our exposition, we also consider properties of graphs formed from wellformed diagrams. We can form a spanning dual of a wellformed diagram's superdual by inspecting the topological adjacency of its zones:

Definition 4.3. Given a 3D Euler diagram, the *topological dual* is a graph with a vertex for each zone description and an edge between each pair of topologically adjacent zones.

Lemma 4.2. *Let D be a diagram description. There exists a 3D Euler diagram, d, with description D such that d possesses all of the wellformedness properties except that it might have non-unique labels iff D has a spanning dual that is connected.*

Proof. First, suppose that such a d exists. Create the topological dual of d. The topological dual is a connected graph. Topologically adjacent zones are separated by single contours, because there is no concurrency. Therefore, topologically adjacent zones are also combinatorially adjacent. Hence the topological dual is a spanning dual of D, thus completing the first direction of the proof.

For the converse, choose any connected spanning dual of D and choose any spanning tree, T, of the spanning dual. Embed T in $\mathbb{R}^3$; since we construct an embedding, no pair of edges intersect. Turn T into a directed tree with a unique root node $\varnothing$. This implies that the edges form directed paths from $\varnothing$ to the leaves of T. We use T to construct d.

The diagram construction process begins by forming a sphere around each leaf, except for $\varnothing$ should this be a leaf. If this step results in all vertices except for $\varnothing$ being enclosed by spheres then all that remains is to label the surfaces; this process is described later. Alternatively, there is at least one such vertex, except for $\varnothing$, that is not enclosed by a surface. Then, since T has no cycles, there
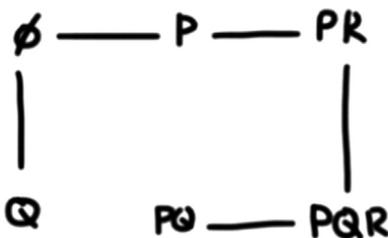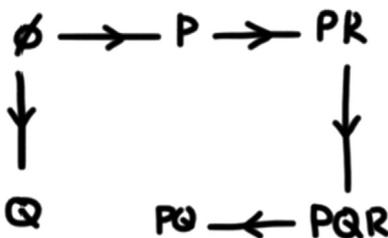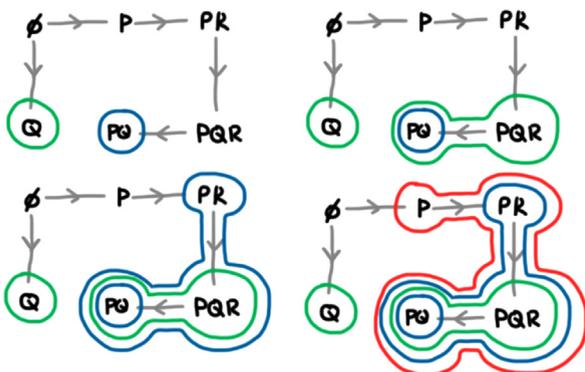
also exists at least one vertex, v, ($\neq \varnothing$) where all vertices on paths from v to leaves are enclosed by surfaces. Choose such a vertex v and draw a sphere, S, around v. Next, for each vertex, v', incident with v and on a path from v to a leaf, fatten the unique surface that encloses v' to create a new surface, $S_{v'}$, and grow $S_{v'}$ along the edge that is incident with v and v' until it joins S and $S_{v'}$. This process for adding a sphere, fattening surfaces and connecting them to the sphere is illustrated here:
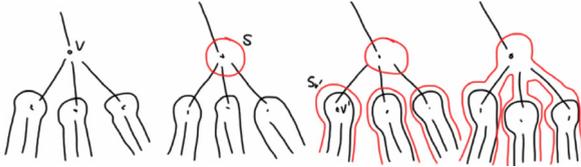
Thus, the fattened surfaces, the connecting tubes and the sphere around v combine to form a single surface that encloses v and all vertices on paths from v to leaves. Repeat this process until all vertices except for $\varnothing$ are enclosed by surfaces. This final set of surfaces is $\mathcal{S}$.

Now, each surface, S, cuts exactly one edge, e, of T. The label of S is the symmetric difference of the vertices incident with e, thus defining the labeling function l. The result is a 3D Euler diagram, $d = (\mathcal{S}, l)$, with description D.

We must now show that d possesses the required properties. Each vertex is in exactly one minimal region and, since there is no intersection between surfaces, the construction ensures that every minimal region contains a vertex. The vertex set is the diagram description, so the number of minimal regions is precisely the number of zones. Hence the zones are connected. Since no pair of surfaces intersect, it is trivial that the 3-point, crossings, line concurrency and surface concurrency properties are possessed. Hence, d is a 3D Euler diagram with description D and possesses all wellformedness properties except, perhaps, for unique labels. □

4.3. Conditions for wellformed drawability

The construction described above provides us with insight into how to build wellformed diagrams, under certain conditions, as we now illustrate by example. Continuing with the example used to illustrate the proof strategy for Lemma 4.2 (i.e. Figs. 15–19), we can join disconnected contours by creating tubes that run along edges in the superdual. Starting with the diagram in Fig. 19, we observe that the two surfaces labeled R are separated by just a single surface, that labeled Q. We can, therefore, connect the two surfaces labeled R by growing a tube between them:
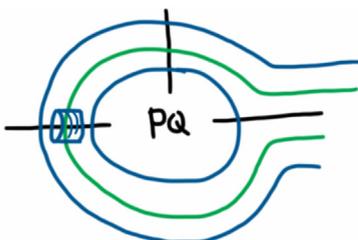
The result is shown in the top diagram in Fig. 20. Now, the only contour comprising more than one surface is Q. After the joining of R, the two surfaces labeled Q are separated by a single surface so we can, again, join them using a tube. The bottom diagram in Fig. 20 shows the final, wellformed diagram.

By contrast, given the diagram in Fig. 21, we cannot join the disconnected contour P or the contour Q using tubes without creating either additional zones or a non-wellformed coincidence of contours. These problems occur because of the ordering of the contours in the layers that surround the vertices Q and QR in the superdual.

The question arises: when can we follow this process of adding tubes to join together separated surfaces belonging to the same contour? Answering this question is the heart of our approach used to establish necessary and sufficient conditions that identify which diagram descriptions can be drawn wellformed. We identify conditions on the superdual, or more
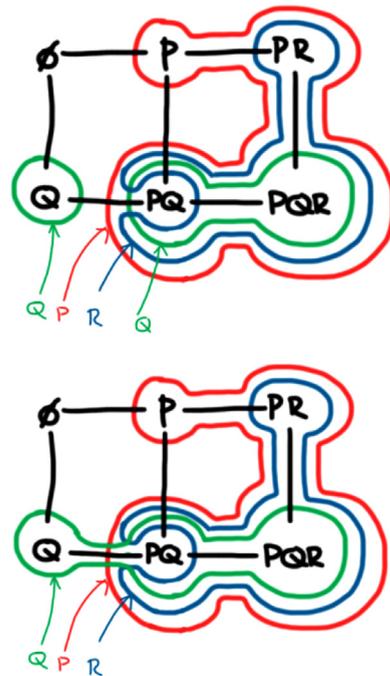
Fig. 20. Connecting the contour R and then Q.
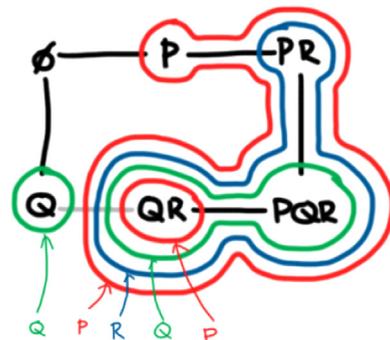
Fig. 21. Unable to connect the contour P or the contour Q.

specifically spanning duals, that correspond to wellformed drawability.

4.3.1. Connectivity conditions

In the topological dual, for any surface, the maximal sub-graph containing precisely the vertices arising from zones inside that surface is connected. Similarly, the maximal sub-graph formed from the vertices outside the surface is also connected. For wellformed diagrams, where surface labels are unique, this corresponds to the *connectivity conditions*:

Definition 4.4. Let $G = (V, E)$ be a graph such that $V \subseteq \mathbb{PL}$. The *connectivity conditions* for G are:

1. G is connected,
2. for each label, L, in $\mathcal{L}$, the maximal subgraph of G whose vertices include L is connected, and
3. for each label, L, in $\mathcal{L}$, the maximal subgraph of G whose vertices do not include L is connected [7].

We will show that the connectivity conditions are necessary for 3D wellformed drawability as part of Theorem 4.1. Unfortunately, the connectivity conditions alone are not sufficient for 3D wellformed drawability. The diagram description

$$D = \{\varnothing, P, PQ, R, QR, PQR, PQS, QRS, PQRS\}$$

passes the connectivity conditions but, as we will demonstrate later, cannot be drawn wellformed. The superdual of D can be seen in Fig. 22.

4.3.2. WF-moves on paths in graphs

Our next task is to give some additional conditions which can be used in conjunction with the connectivity conditions to give sufficient conditions for 3D wellformed drawability. Rather than deriving those conditions by working through examples we will, for now, present a series of definitions that enable us to state additional conditions. Justification and background explaining how those conditions came about will follow. We start by defining when two paths in a spanning dual are *WF-related*.

Definition 4.5. Let D be a diagram description and let $G' = (V, E')$ be a spanning dual of D. Let p_1 and p_2 be paths in G' such that p_1 has the same start vertex as p_2 and, also, the same end vertex as p_2. We say that p_1 and p_2 are

WF-related if p_1 can be transformed into p_2 via a sequence of operations, called *WF-moves*.

WF1 given a path, an edge e in G' can be inserted into the path, immediately followed by edge e again (so the path goes along e and then turns back and goes along e in the other direction) or, similarly, the pair of traversals along e can be removed:

WF2 a sequence of three edges, e_1, e_2, and e_3, in can be replaced by a single edge e_4 in G' where:
 (a) e_1 adds or removes some contour label L_1,
 (b) e_2 adds or removes some contour label L_2, and
 (c) e_3 adds or removes L_1, and
 (d) e_4 adds or removes L_2:
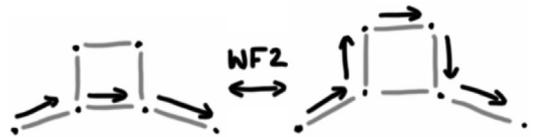
In Fig. 23, the paths p_1 and p_2 are not WF-related whereas p_1 and p_3 are WF-related. To establish that p_1 and p_3 are WF-related, we need to find a sequence of WF-moves that transforms p_1 into p_3. To aid our exposition, we will describe paths by their vertices, and identify the vertex that changes at each step by the label that changes. Using this description, p_1 can therefore be written as $PQ\ (+R)\ PQR\ (-P)\ QR$. The sequence of operations to obtain p_3 is as follows:

$$p_1 = PQ\ (+R)\ PQR\ (-P)\ QR$$

$$WF1 \rightarrow PQ\ (+S)\ PQS\ (-S)\ PQ\ (+R)\ PQR\ (-P)\ QR$$

$$WF1 \rightarrow PQ\ (+S)\ PQS\ (+R)\ PQRS\ (-R)\ PQS\ (-S)$$
$$PQ\ (+R)\ PQR\ (-P)\ QR$$

$$WF2 \rightarrow PQ\ (+S)\ PQS\ (+R)\ PQRS\ (-S)\ PQR\ (-P)\ QR$$

$$WF1 \rightarrow PQ\ (+S)\ PQS\ (+R)\ PQRS\ (-P)\ QRS\ (+P)$$
$$PQRS\ (-S)\ PQR\ (-P)\ QR$$

$$WF2 \rightarrow PQ\ (+S)\ PQS\ (+R)\ PQRS\ (-P)\ QRS\ (-S)\ QR = p_3.$$

Fig. 22. A superdual that passes the connectivity conditions.
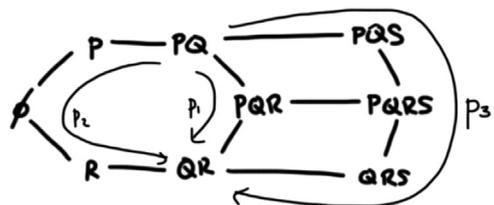
Fig. 23. WF-related paths.

It is useful for us to define a further WF-move that is a composite of the first two moves:

Definition 4.6. Given a path in a spanning dual, G', *WF3* allows us to remove two consecutive edges, e_1 and e_2, and replace them with two consecutive edges e_3 and e_4 in G' whenever:

1. e_1 and e_4 both add some label L or both remove some label L, and
2. e_2 and e_3 both add some label L' or both remove some label L':

It can readily be shown that WF3 is the composition of WF1 and WF2. For instance, if our path includes $v_1 (+L_1) v_2 (+L_2) v_3$ and we can apply WF3 to yield $v_1 (+L_2) v_4 (+L_1) v_3$ then we can firstly apply WF1 to obtain $v_1 (+L_2) v4 (-L_2) v_1 (+L_1) v_2 (+L_2) v_3$. Then we can apply WF2 to obtain $v_1 (+L_2) v_4 (+L_1) v_3$ as required. The previous example, which demonstrated that p_1 and p_3 were WF-related using WF1 and WF2, can be simplified using WF3:

$$p_1 = PQ (+R) PQR (-P) QR$$
$$WF1 \rightarrow PQ (+S) PQS (-S) PQ (+R) PQR (-P) QR$$
$$WF3 \rightarrow PQ (+S) PQS (+R) PQRS (-S) PQR (-P) QR$$

$$WF3 \rightarrow PQ (+S) PQS (+R) PQRS (-P) QRS (-S) QR$$
$$= p_3.$$

4.3.3. Wellformed spanning duals

We now define what it means for a spanning dual to be *wellformed*:

Definition 4.7. A spanning dual, G', is *wellformed* provided it passes the connectivity conditions and any two paths in G' that share their start vertex and end vertex are WF-related.

To illustrate, neither of the spanning duals in Fig. 24 are wellformed. The first fails the connectivity conditions and the second has paths which share end-vertices but are not WF-related. We will prove that, given a wellformed 3D Euler diagram, its topological dual is wellformed. Thus, for wellformed drawability in 3D, it is necessary for the superdual of description D to possess a spanning dual that is wellformed. Our main result of Section 4.3.4 is that this condition is both necessary and sufficient. To see that it is necessary, we will examine some properties of the topological duals of wellformed diagrams.
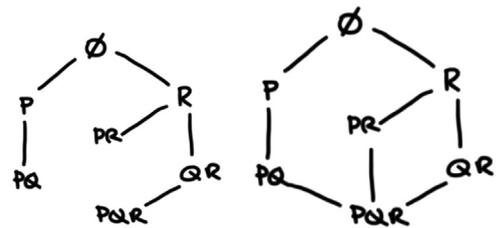
Fig. 24. Two not wellformed spanning duals.

4.3.4. Necessary and sufficient conditions for wellformed drawability

We start by linking paths in the topological dual with paths in space through the wellformed diagram. Recall that every edge in the topological dual is between zones whose descriptions are combinatorially adjacent. Moving along such an edge corresponds to passing through a single surface. Thus, traversing a path in the topological dual intuitively corresponds to a path in $\mathbb{R}^3$ that passes through a sequence of surfaces. Similarly, some paths in $\mathbb{R}^3$ correspond to paths in the topological dual.

In a wellformed diagram, if we choose two points in space that lie in zones then there is a path in $\mathbb{R}^3$ that joins them. Such a path can be chosen so that it meets the surfaces of d at a finite set of points, only passes through one surface at any time, and only intersects surfaces transversely. By choosing such a path, which we call a *canonical path*, we have essentially chosen a path in the topological dual: each time the path crosses a surface, moving from one zone to another, traverse the corresponding edge in the topological dual to create a path. If we have two canonical paths with the same start points and the same end points then it is possible to continuously change one path into the other whilst keeping the end points fixed. Using language from mathematical topology, we can construct a *homotopy* between the two original paths, where almost all intermediate paths are canonical. As the paths change in $\mathbb{R}^3$, the corresponding paths in the topological dual also undergo changes, which exactly correspond to the defined WF-moves.

Fig. 25 shows an Euler diagram (top left) along with two vertices in its topological dual, corresponding to points P_1 and P_2. The illustrated path, λ, is canonical and gives rise to the illustrated path in the topological dual (bottom left). One feature of canonical paths is that we can nudge them and the resulting path is also canonical. The smaller illustrations on the right show various non-canonical choices of paths, λ, between two points, P_1 and P_2: λ passes through a surface crossing point or touches a surface.

Definition 4.8. Let d be a 3D Euler diagram and let $\lambda : [0, 1] \rightarrow \mathbb{R}^3$ be a path in $\mathbb{R}^3$. We say that λ is *canonical* provided

1. λ meets the surfaces of d only a finite number of times,
2. at any point λ meets a surface, S, of d, it meets only S, and
3. whenever λ meets a surface, S, of d it crosses S transversely.

Fig. 26 shows Venn-3 (top left) with two canonical paths, λ_1 and λ_2, along with the paths to which each gives rise in the topological dual (bottom left) between vertices $P = v_1$ and $Q = v_2$. We demonstrate how a homotopy, transforming of λ_1 into λ_2, can be used to induce a sequence of WF-moves on path in the topological dual corresponding to λ_1 to yield the path corresponding to λ_2. The sequence of diagrams on the right shows how we can transform λ_1 into λ_2 and how that transformation corresponds to applying specific moves WF1, WF2 and WF3. These WF moves thus demonstrate a corresponding transformation of one path in topological dual into the other path, establishing that they are WF-related.

It is possible that some alternative choices of homotopy would not give rise to a sequence of WF-moves in the manner just illustrated. This is because some of the intermediate paths may not correspond to paths in the topological dual; we must take care to avoid this situation. Therefore, we now define a *canonical* homotopy. Recall
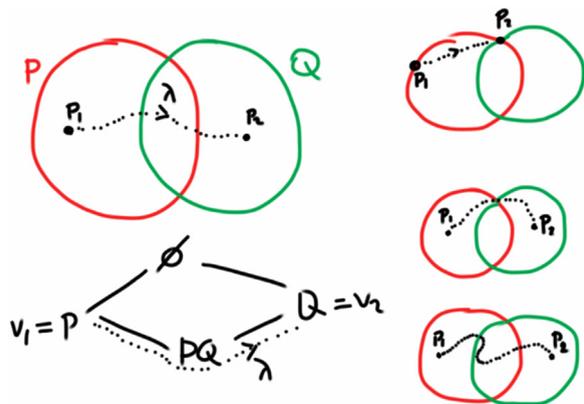
Fig. 25. Relating paths through a diagram with paths in its topological dual.

that a homotopy, in $\mathbb{R}^3$, is a function of the form $h : [0, 1] \times [0, 1] \to \mathbb{R}^3$ that continuously transforms one path into another path. For any fixed $s \in [0, 1]$, the function obtained by restricting the domain of h to $\{s\} \times [0, 1]$ is a path.

Definition 4.9. Let d be a 3D Euler diagram and let $\lambda_1 : [0, 1] \to \mathbb{R}^3$ and $\lambda_2 : [0, 1] \to \mathbb{R}^3$ be canonical paths such that $\lambda_1(0) = \lambda_2(0)$ and $\lambda_1(1) = \lambda_2(1)$. Let $h : [0, 1] \times [0, 1] \to \mathbb{R}^3$ be a homotopy such that, for all $t \in [0, 1]$, $h(0, t) = \lambda_1(t)$ and $h(1, t) = \lambda_2(t)$. Then h is a *canonical* homotopy provided there exists a finite subset, X, of $[0, 1]$ such that

1. for all $s \in [0, 1] - X$, restricting the domain of h to $\{s\} \times [0, 1]$ yields a canonical path, and
2. for all $s \in X$, restricting the domain of h to $\{s\} \times [0, 1]$ yields a path, λ, that is almost canonical, in that
   (a) λ passes through the surfaces of d a finite number of times,
   (b) there exists a unique point, t, such that
      i. at any point other than $\lambda(s, t)$ where λ passes through a surface, S, of d it passes through only S,
      ii. at any point other than $\lambda(s, t)$ where λ meets a surface, S, of d it crosses S transversely, and
      iii. either $\lambda(s, t)$ meets exactly one surface but does not cross it or $\lambda(s, t)$ meets exactly two surfaces of d.

The points in X are called *critical*.

Fig. 27 shows two paths, λ, that are not canonical paths, but could form part of a canonical homotopy. The lefthand example $\lambda(s, t)$ meets the surface S but does not cross it. In the righthand example $\lambda(s, t)$ meets exactly two surfaces.

Using canonical homotopies, we are now able to prove that the topological dual of a wellformed diagram is itself wellformed.
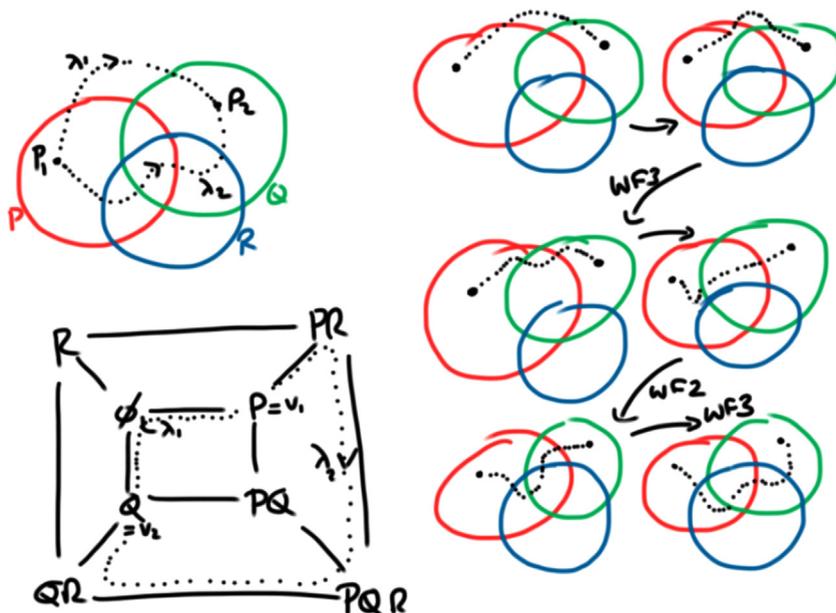
Fig. 26. Using a canonical homotopy to construct a sequence of WF-moves.
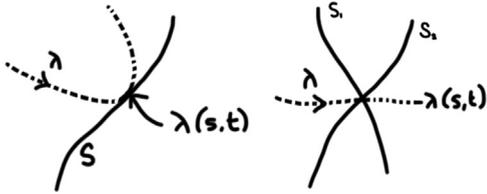
Fig. 27. A canonical homotopy.

Theorem 4.1. *Let d be a wellformed 3D Euler diagram. The topological dual of d is wellformed.*

Proof. The proof that the topological dual passes the connectivity conditions is a direct analogue of the proof of the same result in 2D [7], but we will give some details of the connectivity proof anyway, as it follows some of the same steps as the proof that the other WF conditions on the topological dual hold.

To prove that the topological dual of d is a connected graph, take any two of its vertices, v_1 and v_2. We need to show that there exists a path in the topological dual between v_1 and v_2. Each of the vertices corresponds to a zone of the diagram d. Choose points P_1 and P_2 in the interior of each of the two corresponding zones. We can construct a canonical path, λ, in $\mathbb{R}^3$ from P_1 to P_2. As already demonstrated, λ describes a path in the topological dual. This shows that the topological dual is connected, and the same argument can be used to show that the interior or exterior of any surface corresponds to a connected subgraph of the topological dual. Since d is wellformed, contours comprise single surfaces. Thus we deduce that for each label, L, the maximal subgraph of the topological dual whose vertices include L is connected, as is the maximal subgraph whose vertices do not include L. Hence the topological dual passes the connectivity conditions.

The new part of the proof addresses the other WF conditions on the topological dual. We need to begin with two paths in the topological dual which share start vertices and end vertices and show that there is a sequence of WF moves which transforms one path into the other. Choose points P_1 and P_2 in the zones which correspond to the shared start vertices and end vertices of the paths. Then choose canonical paths λ_1 and λ_2 in $\mathbb{R}^3$, from P_1 to P_2, which correspond to the two topological dual paths. The space $\mathbb{R}^3$ is simply connected so we can construct a homotopy between these paths. Moreover, because the diagram is wellformed we can construct a canonical homotopy. Consider a critical point, $s \in X$. Then there exists a unique $t \in [0, 1]$ such that either $\lambda(s, t)$ meets exactly one surface but does not cross it or or $\lambda(s, t)$ meets exactly two surfaces in d. The proof now considers each of these two cases.

1. *Case* 1: $\lambda(s, t)$ meets exactly one surface but does not cross it. Then the two canonical paths, $h_{s-\delta_s}$ and $h_{s+\delta_s}$, obtained by restricting the domain of h to $\{s-\delta_s\} \times [0, 1]$ and $\{s+\delta_s\} \times [0, 1]$ respectively either give rise to the same path in the topological dual or they give rise to different paths. In the former case, no WF move

arises. In the latter case, the two distinct paths in the topological dual, arising from $h_{s-\delta_s}$ and $h_{s+\delta_s}$ respectively, differ by a WF1 move in the topological dual. This is because the homotopy has pushed the path h through exactly one surface.

2. *Case* 2: $\lambda(s, t)$ meets exactly two surfaces in d. Here, $\lambda(s, t)$ meets a surface intersection curve and at this point we have two surfaces that cross transversely (because d is wellformed). Again, the two canonical paths $h_{s-\delta_s}$ and $h_{s+\delta_s}$ either give rise to the same path in the topological dual or they give rise to different paths. In the former case, no WF move arises. In the latter case, the two paths distinct paths in the topological dual, arising from $h_{s-\delta_s}$ and $h_{s+\delta_s}$ respectively, differ by either WF2 or WF3.

Since there are only a finite number of critical points, we have just described a sequence of WF-moves which demonstrate that paths in the topological dual which share beginning and end vertices are WF-related. Hence the topological dual is wellformed. □

The next theorem now establishes that from a wellformed spanning dual we can construct a wellformed 3D Euler diagram. It makes arguments about sequences of zones and surfaces that are passed through as we move along edges or paths in spanning duals. To illustrate, in Fig. 28 a point moving along the edge between *PQ* to *PQR* passes through the surfaces in sequence *P, Q, R, P, Q*. The point passes through the following zone sequence: *PQ, Q, ∅, R, PR* and *PQR*. The path in the chosen spanning tree from *PQ* to *∅* to *PQR* passes through a sequence of vertices corresponding to this same zone sequence. We also observe that each pair of topologically adjacent zones has a corresponding edge in the spanning dual.

Theorem 4.2 (*Wellformed drawability*). *Let D be a diagram description with superdual G. There exists a spanning dual, G′, of G such that G′ is wellformed iff there exists a wellformed 3D Euler diagram, d, that is a drawing of D.*

Proof. One direction is given by Theorem 4.1: if d exists then the topological dual passes the connectivity conditions and is wellformed. Since the topological dual is a spanning dual of G we are done.

For the other direction, we show that we can use $G′$ to construct d. First, we use $G′$ to construct a 3D Euler diagram using the approach described for Lemma 4.2.
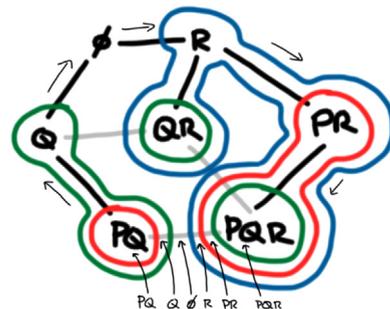
Fig. 28. Relationships between zones, surfaces, edges and paths.

This begins with an arbitrary embedding of G' in $\mathbb{R}^3$ and an arbitrary spanning tree T, with appropriately directed edges, of G'. The diagram construction used in the proof of Lemma 4.2 yields a 3D diagram, d', with description D which is entirely wellformed except, perhaps, for having non-unique labels.

Our task now is to demonstrate that any contour comprising more than one surface can be connected into a single surface. We will describe a sequence of adjustments to the diagram which maintain the zone set and which correspond to WF-moves. The changes will preserve the some of wellformedness conditions (no quadruple points, transverse crossings, no line concurrency and no surface concurrency). After these changes have been made, further transformations are required to produce a wellformed diagram with the required description.

The adjustments to the diagram are generated by consideration of each edge of $G' - T$ in turn. Consider an edge e in $G' - T$ and its incident vertices, v_1 and v_2; this edge is directed in G' from, say, v_1 to v_2. There are unique paths with no duplicated edges in T from v_1 to $\varnothing$ and from $\varnothing$ to v_2. Define p to be the composite path from v_1 to $\varnothing$ to v_2. The path p may pass along edges of T more than once. We observe the following facts:

1. Each edge of G' corresponds to a contour label (i.e. the label which is in the symmetric difference of its incident vertices).
2. The path p follows a sequence of edges, so it generates a sequence of contour labels $L_1, \ldots, L_{n-1}$.
3. Each vertex of G' corresponds to a zone.
4. The path p also follows a sequence of vertices, so it determines a sequence of zones, $z_1, \ldots, z_n$.
5. Consider a point in $\mathbb{R}^3$ moving along the edge e from v_1 to v_2. This point passes through the same sequence of zones, $z_1, \ldots, z_n$, that was determined by following the path p. Thus, the edge e can be associated with a sequence of zones. We also pass through a the sequence of $n-1$ surfaces, with the same sequence of labels $L_1, \ldots, L_{n-1}$ generated by the path p.
6. Each zone along the edge e can be associated with a path in T from v_1 to the vertex associated with that zone. These sequences of zones associated with e, each of which has a path in the tree to its vertex, will be useful later.

The theorem gives that G' is wellformed. We can apply this condition to two paths from v_1 to v_2: the path p constructed in the previous paragraph and the single-edge path going directly from v_1 to v_2. Since G' is wellformed, we can deduce that there is a sequence of WF-moves transforming p into the single-edge path, e. Each of these WF-moves will trigger a change to the 3D Euler diagram (we will describe these diagram changes in the next paragraph). The sequence of paths $p = p_0,\ p_1, \ldots, p_m = e$ corresponds to a sequence of 3D Euler diagrams $d = d_0, d_1, \ldots, d_m$. As we adjust the diagrams to convert d_{i-1} into d_i we will preserve the properties (a)–(c):

(a) Each diagram, d_i, has all the wellformed properties except for potentially having disconnected zones and non-unique labels.

(b) In each diagram, d_i, the sequence of G''s vertices given by path p_i is exactly the zone sequence we pass through along the embedded edge e.
(c) Each pair of topologically adjacent zones has a corresponding edge in G'.

Each WF-move transforms p_{i-1} to p_i in one of four ways:

1. *WF1.* We applied WF1, inserting an edge and reversing back along the same edge, which inserts $L_1 L_1$ into the contour label sequence.
2. *WF1 reversed.* We applied WF1 reversed, removing $L_1 L_1$ from the contour label sequence associated with p_{i-1} to give the contour label sequence associated with p_i.
3. *WF2.* We applied WF2, replacing L_2 with $L_1 L_2 L_1$ in the contour label sequence.
4. *WF2 reversed.* We applied WF2 reversed, replacing $L_1 L_2 L_1$ with L_2 in the contour label sequence.

We now describe the corresponding transformations on diagrams in order of simplicity, not the order of the WF-moves described above.

1. *WF1 reversed.* If we applied WF1 reversed, there is a sequence of vertices on path p_{i-1} which we call v, vL_1, v:
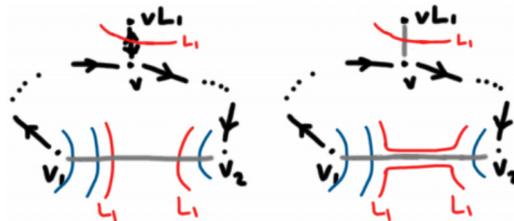
We can find the corresponding sequence of three zones along the embedded edge e. The three zones are partitioned along e by two pieces of the surface labeled L_1. Construct a tube along e joining the two pieces of the surface L_1 and cut holes in L_1 to attach the tube. The tube passes through no other surface. Adding the tube to the two pieces of surface labeled L_1 cannot split up a surface, and potentially joins together two different surfaces labeled L_1. All zones present before this change are present afterwards, and we have not disconnected any zones. We cannot have broken any of the wellformedness properties and we have reduced the sequence of zones along e in the required way, i.e. properties (a) and (b) hold. Any pair of zones which are adjacent after the move were also adjacent before, so we have preserved property (c), that is topologically adjacent zones have a corresponding edge in G'.
2. *WF2 reversed.* If we applied WF2 reversed, there is a sequence of vertices on path p_{i-1} which we call $v, vL_1, vL_1 L_2, vL_2$:
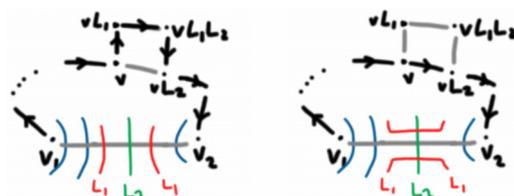
We can find a corresponding sequence of four zones along the embedded edge e. The four zones are partitioned along e by two pieces of the surface labeled L_1 and a piece of the surface L_2 in between. Construct a tube along e joining the two pieces of the surface L_1 and cut holes in L_1 to attach the tube. The tube passes transversely through the surface L_2:
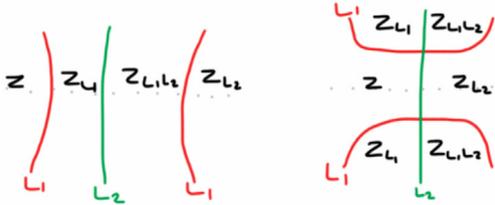
All zones present before this change are present afterwards, and we have not split any zone. Adding the tube to the two pieces of surface labeled L_1 cannot split up a surface, and potentially joins together two different surfaces labeled L_1. We cannot have broken any of the wellformedness properties and we have reduced the sequence of zones along e in the required way. Any pair of zones which are adjacent after the move were also adjacent before, so we have preserved the property that topologically adjacent zones have a corresponding edge in G'. Hence properties (a)–(c) all hold.

3. *WF1*. If we applied WF1, there is a vertex v of the path at which we insert the two extra edges both labeled L_1:
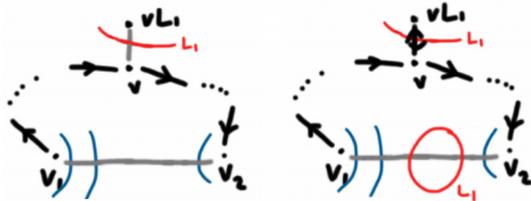
There is zone corresponding to v along the embedded edge e. In that zone, on the edge e, insert a sphere labeled L_1 small enough that it does not meet any other surfaces. Properties (a) and (b) hold in resulting diagram, but we have added a surface labeled L_1 and added a new minimal region for the pre-existing zone inside this new surface L_1. There is a (potentially new) topological zone adjacency in the diagram at the newly inserted surface, between the zones with descriptions corresponding to v and vL_1. This adjacency corresponds to the edge in G' from v labeled L_1. Hence property (c) also holds.

4. *WF2* If we applied WF2, there is an edge on the path v vL_2 which becomes three edges on the path of vertices that we call v vL_1 vL_1L_2 vL_2:
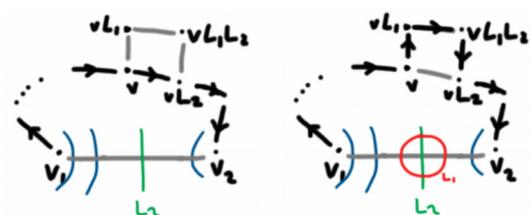
Between the zones associated with v and vL_2 along the edge e, there is a surface L_2. Add a sphere labeled L_1 which straddles L_2 along e. The resulting diagram has properties (a) and (b), but we have added a new surface labeled L_1 and added new minimal regions of the pre-existing zones labeled vL_1 and vL_1L_2. There are two (potentially new) topological zone adjacencies in the diagram at the newly inserted surface. These corresponds to the edges in G' between v to vL_1 and between vL_1L_2 and vL_2. Hence property (c) also holds.

After all the WF-moves have been processed, the diagram d_m has the property that each edge in G' intersects with one surface.

We have two remaining tasks: we must merge together any disconnected zones and we show that we have unique surface labels. We will merge zones by adding tubes between them. For each zone, there is a minimal region containing the corresponding embedded vertex – call that the *primary zone component*. Any other minimal regions of the zone will be called *secondary zone components*. Suppose that there exists a disconnected zone, z. Then there is a secondary zone component, z_s, of z which is topologically adjacent to some primary zone component, z'_p of some zone z'.
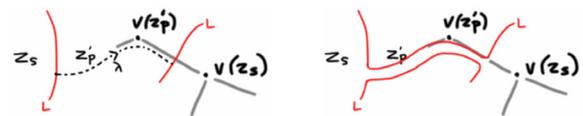
Every topological adjacency in d corresponds to an edge e in G', and every edge in G' has exactly one surface passing through it. The surface, S, labeled L say, between z_s and z'_p has the same label as the surface which crosses the corresponding edge of G'. Choose a point P on S between z_s and z'_p and construct a path, λ, in z'_p from P to the embedded vertex in z'_p (this is where we use the fact that z'_p is primary). Extend that path along the edge e (this is where we use the fact that topological adjacency corresponds to adjacency in G') until it meets the unique surface passing through e (this is where we use the fact that every edge of G' has only one surface through it). This path connects two surfaces with the same label, and if we add a tube along the path, joining the paths of the surfaces, we are joining the zone z_s to its primary counterpart. We have reduced the number of secondary zones and we can proceed by induction to eliminate every secondary zone. All zones are now connected. Call the resulting diagram d_{m+1}.

We now know that the diagram d_{m+1} passes all of the wellformed properties except for potentially having non-unique contour labels. We will finish the proof by showing that the diagram constructed so far does indeed have unique contour labels without needing any more adjustment. By construction, along any edge e, there is only one surface crossing e. This, with the connectivity conditions, is enough to deduce unique contour labels. Suppose, for a contradiction, that we have distinct surfaces labeled L and consider the collection of zones inside the contour labeled L. This corresponds to a connected subgraph of G', i.e. the

maximal subgraph, G_L, whose vertices include the label L. The vertices in this subgraph include all of the zone descriptions that include the label L. Choose two vertices, v_1 and v_2, of this subgraph that are in distinct surfaces and choose a path from v_1 to v_2. As we traverse this path we must, at some point, pass along an edge, e, that is crossed by a surface, S_L, labeled L in order to reach the vertex v_2. The surface S_L is the only surface that crosses the edge e but then one of the vertices incident with e does not contain the label L. This implies that it is not in the maximal subgraph G_L, contradicting its connectivity. Hence we have reached a contradiction, and the contour L must be comprise exactly one surface. Therefore, the diagram d_m is wellformed and has description D as required.   □

Returning to the example in Fig. 20, if we connect all of the contours using the process in the proof of Theorem 4.2 then we obtain the diagram Fig. 29, shown here as a cross-sectional slice.

As a consequence of Theorem 4.2, we can now demonstrate that the diagram description associated with the superdual in Fig. 22 cannot be drawn wellformed. To do so, we must demonstrate that there is no spanning dual that is wellformed. Fig. 23 identifies two paths, p_1 and p_2, in the superdual that are not WF-related. We demonstrated that p_1 is WF-related to p_3. Since the WF-related relation is transitive, it follows that p_2 and p_3 are also not WF-related. Now, suppose that there exists a spanning dual that is wellformed. Then this spanning dual cannot include all of the edges from p_2 and p_3 (or from p_1 and p_2), or it too would not be wellformed. It is straightforward to show that no edge can be removed from p_2 or from p_3 whilst ensuring that the connectivity conditions hold. Hence, there is no spanning dual that is wellformed. Thus, by Theorem 4.2, there does not exists a wellformed diagram with the given description.

4.4. Comparison with 2D

We now draw contrast with 2D Euler diagrams for each of the main results in the preceding subsections.

4.4.1. Comparison: everything is drawable

The first result in Section 4 demonstrated that all diagram descriptions could be drawn in 3D. Moreover, this could be achieved whilst ensuring that the unique labels property and the connected zones property were possessed by the resulting diagram. In 2D, all diagram descriptions can be drawn, but this requires the use of non-unique labels for curves. If we want to enforce the unique labels property then we must allow the curves
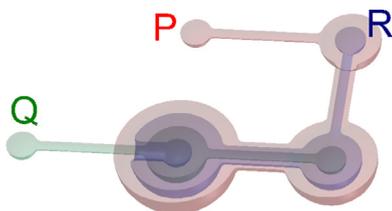
Fig. 29. The result of the construction process.

used in the Euler diagrams to be non-simple (i.e. they must be allowed to self-intersect) [9,24]. However, using either non-unique labels or self-intersecting curves brings with it usability weaknesses. It has been indicated, by empirical study, that in 2D using contours that comprise more than one curve or non-simple curves brings with it a cognitive burden [13], as is the case when other wellformedness properties do not hold. In addition, in the case of self-intersecting curves, there are different notions of what constitutes the interior [6] which could also result in usability problems. Thus, allowing either multiple label use or non-simple curves is less than desirable although it is an essential requirement in 2D for general drawability. We hypothesize that data sets which cannot be visualized in 2D without using non-unique labels may be more effectively visualized in 3D.

4.4.2. Comparison: conditions for wellformed drawability except for non-unique labels

In Lemma 4.2 we presented the main result concerning drawability with all wellformedness properties holding except for unique labels. In fact, the proof of this lemma readily translates to 2D. The first direction of the proof, showing that wellformed diagrams have a connected spanning dual is identical. The converse argument, that a connected spanning dual is sufficient for drawability, adapts as follows. Instead of drawing spheres around nodes, draw circles. Instead of fattening surfaces, simply fatten the simple closed curves. Thus, we have the following lemma for 2D Euler diagrams:

Lemma 4.3. *Let D be a diagram description. There exists a 2D Euler diagram, d, with description D such that d possesses all of the wellformedness properties except, perhaps, unique labels iff D has a spanning dual that is connected.*

4.4.3. Comparison: conditions for wellformed drawability

We presented two theorems on wellformed drawability, namely Theorems 4.1 and 4.2. The first of these theorems readily applies to the 2D case.[6] The necessary conditions for wellformed drawability in 3D are also necessary in 2D:

Theorem 4.3. *Let d be a wellformed 2D Euler diagram. The topological dual of d is wellformed.*

It is interesting that the proof of Theorem 4.1 which was about paths in $\mathbb{R}^3$ and wellformed 3D Euler diagrams, is virtually identical in the 2D cases, where we instead argue about paths in $\mathbb{R}^2$ and wellformed 2D Euler diagrams.

However, Theorem 4.2, which completely addresses 3D wellformed drawability, does not carry over to the 2D domain. The proof of Theorem 4.2 begins by embedding the spanning dual, G', in $\mathbb{R}^3$. That is not always possible in 2D, because it relies on G' being planar. The proof also uses tubes in 3D space to join together surfaces. This step cannot be applied in 2D without disconnecting the zone

---

[6] In 2D, the edges in the topological dual are also derived from zone topological adjacency; topologically adjacent zones in 2D have boundaries that share a curve.

through which the tube passes. Thus, we have indicated why the proof strategy does not extend to 2D; in fact, it can be shown that there are diagram descriptions that can be drawn wellformed in 3D that cannot be drawn well-formed in 2D [11].

The first paper on drawing Euler diagrams in 2D presented necessary and sufficient conditions for well-formed drawability [7]. These conditions include require-ment that there exists a spanning dual, G', that passes the connectivity conditions and is planar. It also requires that there is an embedding of G' that passes the so-called *face-conditions*; it is difficult to establish whether such a spanning dual G' exists. That original work did not con-sider any conditions on paths in the spanning duals, which is a core part of our work here. It will be interesting to explore whether there is an extension to our definition of a wellformed spanning dual that does not simply reuse the face-conditions, that can be used to provide a new set of necessary and sufficient conditions for wellformed drawability in 2D.

5. Conclusion

In this paper we have made several contributions. Initially, we discussed layout choices that arise and their impact on drawability. We demonstrated that there are more choices of topologically different layouts of Euler diagrams in 3D than in 2D. The detailed discussion around this centered on the Venn-3 example. We demonstrated that there is a rich variety of layouts of wellformed Euler diagrams, including Venn-3, in 3D, whereas there is essentially only one wellformed Venn-3 in 2D. In addition, we provided a classification of all wellformed Venn-3s drawn with simply connected surfaces.

The most significant results in this paper are centered around the drawability of diagram descriptions in 3D. We presented a series of drawability results, culminating in a set of conditions that are both necessary and sufficient for wellformed 3D drawability. The novelty of the approach lies, in part, by considering paths in spanning duals. Whilst spanning duals have been used for analyzing wellformed drawability in 2D, no previous research has considered the relationship between paths in these duals and the ability to draw wellformed diagrams. We not only demonstrated that paths in spanning duals must posses certain proper-ties (i.e. the WF-related condition must hold) for well-formed drawability in 3D but also established that it is a necessary condition in 2D. Thus, there is scope to extend the notion of a spanning dual being wellformed for 3D drawability by adding further conditions that are neces-sary and sufficient for wellformed 2D drawability. Identi-fying what further conditions are needed is an interesting avenue of future work.

Automated layout is important for the widespread application of Euler diagrams (2D or 3D) when using them for information visualization. As discussed in the introduc-tion, there are already a number of methods for drawing 2D Euler diagrams. Since 3D Euler diagrams can be used for visualizing information in all of the application areas where their 2D counterparts can be used, drawing algo-rithms are needed for 3D too. Indeed, further relevance of the results in this paper lies in their potential as a basis for the development of automated processes for generating 3D Euler diagrams: the proofs given in Section 4 all have a constructive flavour.

In order to show that a proposed algorithm is effective for drawing wellformed Euler diagrams, say, it is important to know the conditions that may be assumed about the diagram description for any example which is capable of being drawn wellformed; here, we have established neces-sary and sufficient conditions for wellformed drawability. We can then use that class of examples to compare the coverage of one drawing algorithm against another.

Future work includes conducting empirical studies to ascertain when 3D diagrams are better for users than 2D diagrams. We expect that in simple cases, such as Venn-3, there are unlikely to be significant differences in user understanding. However, cognitive benefits may arise for examples that are difficult to draw effectively in 2D but that can be drawn wellformed, say, in 3D. It will be interesting to establish the point at which 3D diagrams have a cognitive advantage over 2D diagrams, in terms of the richness of the represented data.

We also note that Euler diagrams can be augmented with additional syntax, such as graphs (whose nodes represent items in sets and the edges represent relation-ships) [10]. Here, the problem of finding an effective drawing is compounded by the additional syntax. Having more layout choice in 3D over 2D could be more impor-tant, from a cognitive perspective, in these informationally rich settings.

Appendix

To prove Theorem 3.2, we we will use P, Q and R to denote three contours that form *any* wellformed Venn-3 diagram whose surfaces are simply connected (i.e. sphere equivalents); and explicitly note that we are not assuming the Venn-3 lies in one of the classes – this is what we must prove. We will begin by providing some results about the sub-diagram made up of any two contours, P and Q. In Lemmas 5.1–5.4, the diagram $d = (\{P,Q\}, l)$ is the dia-gram obtained by removing R from any given wellformed Venn-3 with simply connected surfaces.

Lemma 5.1. *The diagram $d = (\{P,Q\}, l)$ has four zones, namely those with descriptions $\varnothing$, $\{P\}$, $\{Q\}$ and $\{P,Q\}$. In other words, d is a Venn-2 diagram.*

Proof. Eight zones are present in Venn-3, and when R is removed to create d, these eight zones merge in pairs to create four zones in d. □

Since we are considering wellformed Venn-3, each zone is connected but this is not necessarily a property of the diagram containing just P and Q. For example, in

Fig. 30, the zone Q becomes disconnected on the removal of R; the Venn-3 diagram in this figure is topologically equivalent, up to labeling, to the simple Venn-3 shown in Fig. 10. We now establish that any zone in an underlying Venn-2 diagram contains at most two minimal regions.

Lemma 5.2. *If a zone of $d = (\{P, Q\}, l)$ is disconnected then it has exactly two minimal regions.*

Proof. Suppose, for a contradiction, that a zone z in d comprises three (or more) minimal regions. Adding the third contour, R, to d in order to obtain a wellformed Venn-3 cannot join together any minimal regions of the zones of d, it can only further disconnect them. The three minimal regions of z in d become three or more minimal regions in Venn-3. Moreover, the minimal regions of z in d become two zones in Venn-3: one zone inside R and another zone outside R. But then one of these two zones in Venn-3 comprises at least two minimal regions, contradicting the wellformedness of the Venn-3. Hence, any disconnected zone has exactly two minimal regions. □

In fact, it can also be shown that if one zone comprises two minimal regions then all other zones are connected, which we establish in Lemma 5.3. In the proof of Lemma 5.3, we consider the intersections between the surfaces. The intersection points of two surfaces in a wellformed diagram form simple closed curves, because of the crossings property: at intersection points, surfaces cross transversely. For example, in Fig. 31, the intersection of the two surfaces that cross transversely form three simple closed curves, shown using dashed lines. It is therefore helpful to us to introduce some notation for describing the points that are on a pair (or more) of surfaces. Strictly speaking, two surfaces, P and Q, are functions with co-domain $\mathbb{R}^3$, but we will abuse notation and write $P \cap Q$ to mean the set of points to which both P and Q map; these are the points 'shared' by the surfaces. This use of set theory notation for describing the points on surfaces extends in the obvious way to more surfaces and other types of set theory operators.

Lemma 5.3. *If one zone of the diagram $d = (\{P, Q\}, l)$ is disconnected into exactly two minimal regions then all of the other zones are connected.*

Proof. The proof is by contradiction. Suppose that two zones, z_1 and z_2, each have two minimal regions. The proof splits into two cases: first we will handle the case where z_1 and z_2 are combinatorially adjacent and, secondly, we consider the case where they are not combinatorially adjacent.

If z_1 and z_2 are combinatorially adjacent then their four minimal regions combine to form all of the inside or all of the outside of some contour, S: if the zones are $\{P\}$ and $\{P, Q\}$ then they form the inside of P and if the zones are $\varnothing$ and $\{P\}$ form

Fig. 30. Removing contours can disconnect zones.
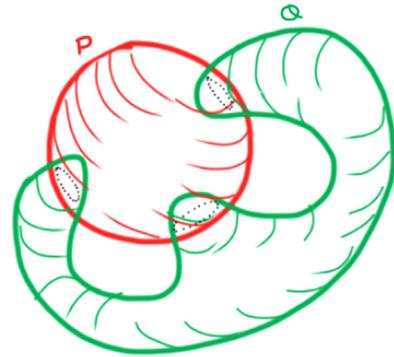
Fig. 31. Intersecting surfaces form simple closed curves.

the outside of Q, for example. In other words, the four minimal regions in z_1 and z_2 combine to make a connected set. The third contour R which combines with P and Q to make a wellformed Venn-3 diagram must not split any of these four minimal regions, since doing so would leave Venn-3 not wellformed (a zone would be disconnected). However, R must partition the four minimal regions of z_1 and z_2 into two pairs, such that: (a) each pair contains exactly one minimal region of each zone, (b) one pair contains two minimal regions inside R, and (c) the other pair contains two minimal regions outside R. That cannot be achieved without R being concurrent with the contour which already separates z_1 from z_2 in d. Hence, the resulting Venn-3 diagram would necessarily be not wellformed, reaching a contradiction.

If z_1 and z_2 are not combinatorially adjacent, we will again show that their minimal regions combine to make a connected set, with the four minimal regions touching along the intersection curves $P \cap Q$. Each minimal region in z_1 or z_2 must be bound partly by some of contour P and some of contour Q. Each of these minimal region must have part of an intersection curve of $P \cap Q$ on its boundary. Each intersection curve is on the boundary of both z_1 and z_2. Finally, because P and Q are each connected, we can find a path along P from any point on the boundary of any of z_1 and z_2's four minimal regions to any other point and the existence of such a path which shows that the four minimal regions are connected. The contour R must be added to create Venn-3 without splitting any one of these four minimal regions but partitioning the them so that two are inside R and two are outside R. The only way this partitioning can be achieved is by R passing through one of the intersection curves in $P \cap Q$, but not just across this curve at a point (making a legal triple point) but containing a continuous portion of the curve, making an illegal intersection between all three contours. That is, the resulting Venn-3 diagram is necessarily not wellformed, reaching a contradiction. Thus, in either case, there cannot be two zones each with two minimal regions. Hence, there is at most one zone comprising two minimal regions. □

Using the previous two lemmas, we are able to establish how the pair of surfaces P and Q intersect. Since we are dealing with wellformed Venn-3, we know that each maximal connected set of intersection points between P and Q is a simple closed curve. We are able to prove that their intersection is either 1 or 2 simple closed curves.
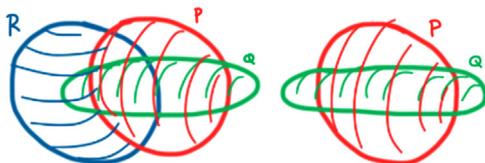
Lemma 5.4. *The two surfaces in the diagram* $d = (\{P, Q\}, l)$ *intersect in either* 1 *or* 2 *simple closed curves.*

Proof. Lemmas 5.2 and 5.3 tell us that *d*'s zones comprise, between them, four or five minimal regions. Now, the number of minimal regions can be counted by considering the number of simple closed curves in $P \cap Q$: there are $3 + n$ components where n is the number of simple closed curves. Hence, $n = 1$ or $n = 2$. □

Based upon what we have shown so far, Fig. 32 shows some possible Venn-2 diagrams obtained from a well-formed Venn-3 by removing a contour. Note that it is always possible to distort space so that one chosen contour is a sphere. If $P \cap Q$ is just one simple closed curve then we can distort space so that P and Q both appear as spheres (the standard Venn-2, i.e. the Venn-2 where both surfaces are actual spheres). However, if $P \cap Q$ is two simple closed curves then it is not always possible to unravel knots that can appear in the contours. That is, we cannot always distort space so that both contours become spheres at the same time, even though both surfaces are topologically equivalent to spheres.

With this understanding of how the Venn-2 diagram might appear, we are ready to state some results about the relationship between our final contour R and the contours P and Q. As an example, consider the Venn-3 diagram on the left of Fig. 33. In the Venn-2 diagram obtained by removing R, shown in the middle, there are three components of $P - Q$, labeled X_1, X_2 and X_3. The first of these, X_1, is an annulus and the other two are discs. Considering each of these in turn, we see that the surface R either intersects with X_i in a single curve component or not at all. Neither X_2 nor X_3 contains any part of R in the lefthand diagram, whereas $X_1 \cap R$ is a single simple closed curve; the right-hand diagram shows X_1 (P with two discs removed) intersecting with R. In the following lemma, just as we used the notation $P \cap Q$ to mean the set of points that lie on both P and Q, we write $P - Q$ to mean the set of points that lie on P but not in Q.

Lemma 5.5. *Given* $d = (\{P, Q, R\}, l)$, *if* X *is a connected component of* $P - R$ *(so* X *is like a disc or annulus) then* $Q \cap X$ *is either* $\varnothing$ *or a connected curve.*
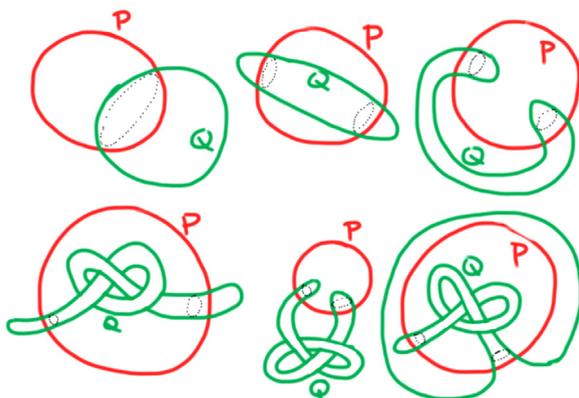
Fig. 32. Different drawings of Venn-2 that can be obtained from well-formed Venn-3.
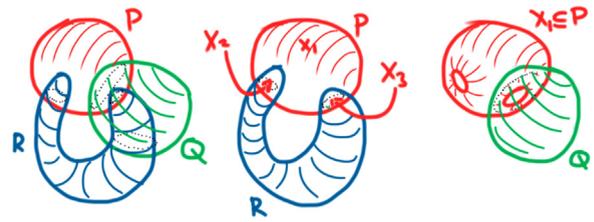
Fig. 33. Relationships between the intersections of surfaces.

Proof (*Sketch*). Assume, for a contradiction, that $Q \cap X$ comprises two or more connected curves. Suppose that one of those connected curves, c_1 say, is closed. Then c_1 bounds a disc, Y, on Q:
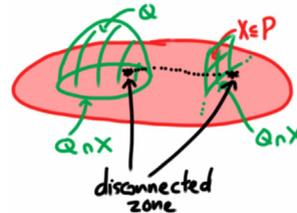
The zone between the disc Y and the component X has more than one minimal region, reaching a contradiction. Therefore, the connected curves in $Q \cap X$ cannot form closed curves. Thus, we can now assume that all of the curves in $Q \cap X$ are open. It can readily be shown that the end points of any curve in $Q \cap X$ are on the boundary of X. There are now several cases to consider relating to the topology of X. When X is a disc, the two connected curves in $Q \cap X$ are in the following configuration:
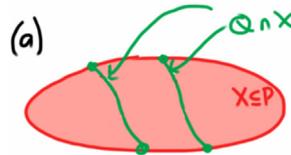
When X is an annulus, the two connected curves in $Q \cap X$ are in one of the following configurations:
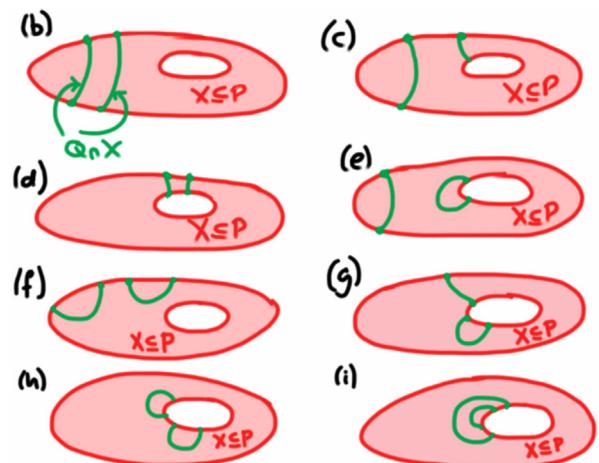
We prove that configuration (a) leads to a contradiction, with cases (b), (c), (d), and (e) being more straightforward. Since we are in case (a), the two open curves from $Q \cap X$ have endpoints on the unique boundary of X. Now, since the boundary of X is a simple closed curve that lies on R, there exists a connected component of R that is a disc with the same boundary as X. Choose such a connected component, Y say, of R:
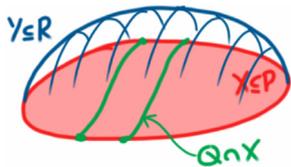
Since the end points of the curve components lie on the boundary of X, this implies that they also lie on R. Given any one of these end points, Q intersects with R in a simple closed curve including that point. Moreover, Lemma 5.4 tells us that $Q \cap R$ contains at most two simple closed curves. Therefore, each of the four end points must be joined to one of the other end points via a simple closed curve in $Q \cap R$. There are two ways in which this can be achieved:
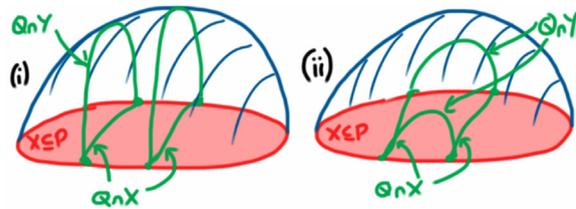
We proceed with these two sub-cases:

(i) In case (i), we have found two closed curves on the surface Q, each of which bounds a disc. It can be shown that these discs are either both inside $X \cup Y$ or both outside $X \cup Y$:
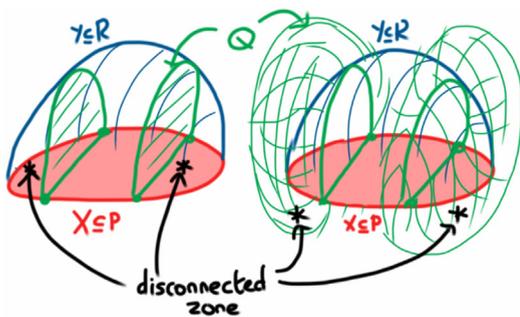
Such discs separate a zone into two components, reaching a contradiction.

(ii) In case (ii), we have found a single closed curve on the surface Q. This closed curve bounds a disc in Q. We will extend that disc along the surface of Q into the surrounding zones of the sub-diagram containing just P and R:
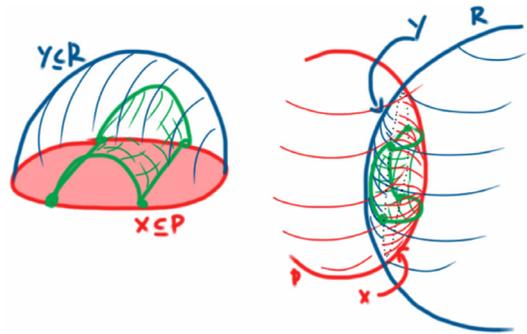
Define X' to be $P - X$ and Y' to be $R - Y$. We know that Q creates two curves in X which end at $P \cap Q$. There must be two curves on X' which end at those same four endpoints. If the four curves make two closed loops, we can identify two discs in Q and show that there are disconnected zones. So the four curves must make one closed loop, bounding a disc of Q. Similarly, the two curves where Q meets Y extend to two more curves where Q meets Y'. Those four curves on Q make a closed loop which bounds a disc. We have constructed three disc-patches on Q which must form a subset of the contour Q. The boundary of that part of Q is a closed loop. Since Q is simply connected, that patch must be a disc and simply connected:
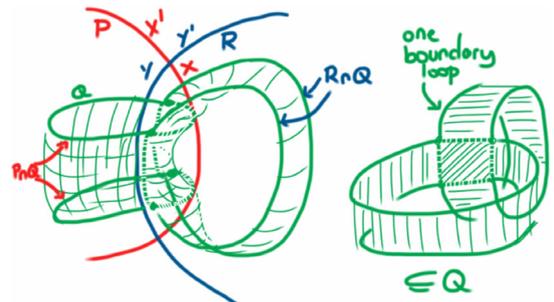
But the patch is not equivalent to a disc and we have reached a contradiction.

Hence, case (a) cannot arise. We leave the cases (b)–(e), which are more straightforward, to the reader. Therefore, $Q \cap X$ is either $\varnothing$ or a connected curve. □

Lemma 5.6. *Given* $d = (\{P, Q, R\}, l)$, *at most one of* $P \cap Q$, $P \cap R$ *and* $Q \cap R$ *comprise two simple closed curves.*

Proof. Assume, without loss of generality, that $P \cap Q$ and $P \cap R$ each have two simple closed curves. We proceed to derive a contradiction. The two simple closed curves of $P \cap R$ split the contour P into three pieces: two discs and an annulus. Given X, a component of $P - R$, Lemma 5.5 implies that $P \cap Q$ can only meet X in one or two connected curves. This ensures that each of the curves of $P \cap Q$ are contained entirely in one of the components of $P - R$. All four intersection curves form $P \cap Q$ and $P \cap R$ must be

disjoint on *P*. There are three possible arrangements of these four curves on the contour *P* and all of them generate a disconnected zone. We are not assuming that the intersection curves bound discs inside *P* but they either bound discs inside *P* or outside *P*, and these discs split a zone into two components, giving a contradiction. Hence at most one of $P \cap Q$, $P \cap R$ and $Q \cap R$ comprise two simple closed curves. □

Theorem 5.1 (*previously Theorem 3.2*). *Let* $d = (\{P, Q, R\}, l)$ *be a wellformed Venn-3 diagram whose surfaces are simply connected. Then d is in either class* 0*, class* 1*, class* 2 *or class* 3.

Proof. Consider what Lemma 5.6 tells us about the contours in Venn-3. We see that Lemma 5.6 provides us with two possibilities for the number of simple closed curves formed by intersecting surfaces:

1. *Case* 1: $P \cap Q$, $P \cap R$ and $Q \cap R$ all comprise one curve. In this case, each pair of contours forms a standard Venn-2 and all three curves form a standard Venn-3.
2. *Case* 2: $P \cap Q$ is one simple closed curve, $P \cap R$ is also one simple closed curve but $Q \cap R$ is two simple closed curves. In this case, the two contours *P* and *Q* form a standard Venn-2. Given this standard Venn-2, there are essentially two ways in which *R* intersects with *P* and *Q*, whilst ensuring that wellformedness is maintained:
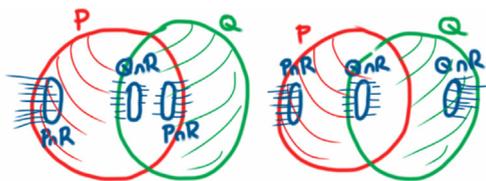
The surface *R* must join up these closed curves without creating further intersection points with *P* and *Q*. The only ways in which this can be done are:
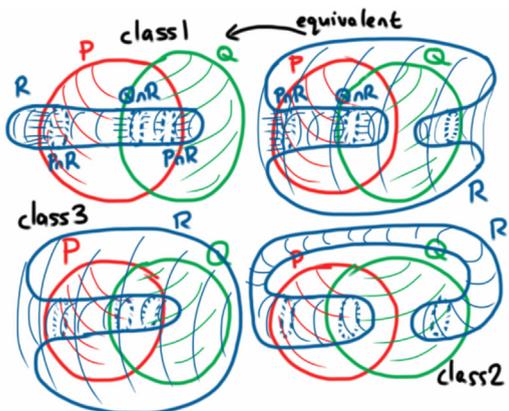
It is left as an exercise for the reader to confirm that the two Venn-3s marked as equivalent in this figure are topologically equivalent after label permutation. The labels *P*, *Q*, *R* on the lefthand diagram correspond

to the labels *R*, *P*, *Q* in the right-hand diagram. Here we show other examples from classes 1–3, where there are knots present:
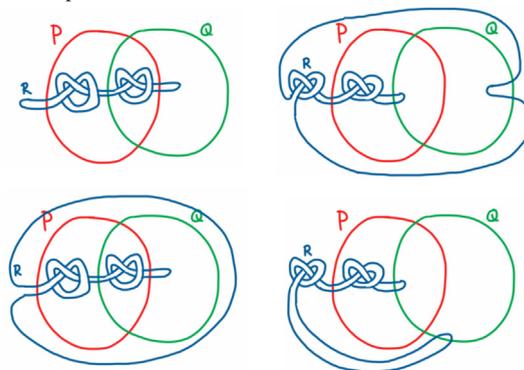
These three constructions in case 2 correspond exactly to classes 1–3, defined earlier and case 1 corresponds to class 0. Hence, we have proved our classification theorem. □

The following theorem allows us to determine the class to which a Venn-3 belongs:

Theorem 5.2 (*Topological properties*). *Let* $d = (\{P, Q, R\}, l)$ *be a wellformed Venn-3 diagram whose surfaces are simply connected.*

1. *Then d is a standard Venn-3 iff*
   (a) *all zones are simply connected,*
   (b) *all pair of contours intersect in one curve,*
   (c) *the removal of any contour results in a wellformed Venn-2 diagram, or*
   (d) *any pair of combinatorially adjacent two zones are topologically adjacent.*
2. *If d is a non-standard Venn-3 then there exists two topologically adjacent zones, z_1 and z_2, such that*
   (a) *z_1 and z_2 are not simply connected, but all other zones are simply connected,*
   (b) *the two contours that do not lie on both the boundary of z_1 and z_2 intersect in two simple closed curves and any other pair of contours intersect in exactly one simple closed curve,*
   (c) *either of the two contours that do not lie on both the boundary of z_1 and z_2 can be removed from Venn-3 to create a wellformed Venn-2 diagram, whereas the removal of the contour that lies on the boundary of both z_1 and z_2 creates a non-wellformed Venn-2 diagram (it has a disconnected zone), and*
   (d) *there exists a pair of combinatorially adjacent zones, z_3 and z_4, that are not topologically adjacent whereas any other pair of combinatorially adjacent zones are topologically adjacent; z_3 and z_4 are those with descriptions image(l) − des(z_1) and image(l) − des(z_2).*

The zones z_1 and z_2 allow us to classify any given wellformed Venn-3, *d*, whose surfaces are simply connected:
(a) *If z_1 and z_2 are contained in* 0 *and* 1 *contours respectively then d is in class* 2.

   (b) *If z_1 and z_2 are contained in 1 and 2 contours respectively then d is in class* 1.

   (c) *If z_1 and z_2 are contained in 2 and 3 contours respectively then d is in class* 3.

To follow these conditions in our example from Fig. 10, note that the zones P and PQ are not simply connected, the contours P and R meet in two curves, but the contours P and Q intersect in one curve and the contours Q and R intersect in one curve. We can remove contour P or Q leaving a wellformed Venn-2, but removing Q leaves a non-wellformed Venn-2. Finally, zones R and QR are combinatorially, but not topologically, adjacent.

References

[1] C.C. Adams, The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots, American Mathematical Society, 2004.

[2] M. Armstrong, Basic Topology, Springer-Verlag, 1979.

[3] S. Chow, F. Ruskey, Drawing area-proportional Venn and Euler diagrams, in: Proceedings of Graph Drawing 2003, Perugia, Italy, vol. 2912, Lecture Notes in Computer Science, Springer-Verlag, 2003, pp. 466–477.

[4] Anthony W.F. Edwards, Venn diagrams for many sets, New Scientist 7 (1989) 51–56.

[5] A. Fish, B. Khazaei, C. Roast, User-comprehension of Euler diagrams, Journal of Visual Languages and Computing 22 (5) (2011) 340–354.

[6] A. Fish, G. Stapleton, Formal issues in languages based on closed curves, in: Proceedings of the Distributed Multimedia Systems, International Workshop on Visual Languages and Computings, Knowledge Systems Institute, Grand Canyon, USA, 2006, pp. 161–167.

[7] J. Flower, J. Howse, Generating Euler diagrams, in: Proceedings of the 2nd International Conference on the Theory and Application of Diagrams, Springer, Georgia, USA, 2002, pp. 61–75.

[8] H. Kestler, A. Muller, T. Gress, M. Buchholz, Generalized Venn diagrams: a new method for visualizing complex genetic set relations, Bioinformatics 21 (8) (2005) 1592–1595.

[9] O. Lemon, I. Pratt, Spatial logic and the complexity of diagrammatic reasoning, Machine GRAPHICS and VISION 6 (1) (1997) 89–108.

[10] N. Riche, T. Dwyer, Untangling Euler diagrams, IEEE Transactions on Visualization and Computer Graphics 16 (6) (2010) 1090–1099.

[11] P. Rodgers, J. Flower, G. Stapleton, Introducing 3D Venn and Euler diagrams, in: 3rd International Workshop on Euler Diagrams, CEUR, 2012, pp. 92–106.

[12] P. Rodgers, L. Zhang, A. Fish, General Euler diagram generation, in: 5th International Conference on the Theory and Application of Diagrams, Springer, 2008, pp. 13–27.

[13] P. Rodgers, L. Zhang, H. Purchase, Wellformedness properties in Euler diagrams: which should be used? IEEE Transactions on Visualization and Computer Graphics 18 (7) (2012) 1089–1100.

[14] P. Rodgers, L. Zhang, G. Stapleton, A. Fish, Embedding wellformed Euler diagrams, in: 12th International Conference on Information Visualization, IEEE, 2008, pp. 585–593.

[15] F. Ruskey, A survey of Venn diagrams, Electronic Journal of Combinatorics, 1997. ⟨http://www.combinatorics.org/Surveys/ds5/VennEJC.html⟩.

[16] P. Simonetto, D. Auber, Visualise undrawable Euler diagrams, in: 12th International Conference on Information Visualization, IEEE, 2008, pp. 594–599.

[17] P. Simonetto, D. Auber, D. Archambault, Fully automatic visualisation of overlapping sets, Computer Graphics Forum 28 (3) (2009).

[18] E. Spanier, Algebraic Topology, Springer, 1994.

[19] G. Stapleton, J. Flower, P. Rodgers, J. Howse, Automatically drawing Euler diagrams with circles, Journal of Visual Languages and Computing 23 (3) (2012) 163–193.

[20] G. Stapleton, P. Rodgers, J. Howse, J. Taylor, Properties of Euler diagrams, in: Proceedings of Layout of Software Engineering Diagrams, EASST, 2007, pp. 2–16.

[21] G. Stapleton, P. Rodgers, J. Howse, L. Zhang, Inductively generating Euler diagrams, IEEE Transactions on Visualization and Computer Graphics 17 (1) (2011) 88–100.

[22] G. Stapleton, L. Zhang, J. Howse, P. Rodgers, Drawing Euler diagrams with circles: the theory of piercings, IEEE Transactions on Visualisation and Computer Graphics 17 (7) (2011) 1020–1032.

[23] J. Venn, On the diagrammatic and mechanical representation of propositions and reasonings, Philosophical Magazine (1880).

[24] A. Verroust, M.-L. Viaud, Ensuring the drawability of Euler diagrams for up to eight sets, in: Proceedings of 3rd International Conference on the Theory and Application of Diagrams, vol. 2980 of Lecture Notes in Artificial Intelligence, Springer, Cambridge, UK, 2004, pp. 128–141.

[25] L. Wilkinson, Exact and approximate area-proportional circular Venn and Euler diagrams, IEEE Transactions on Visualization and Computer Graphics 18 (2) (2012) 321–331.