Journal of Visual Languages and Computing 25 (2014) 518–532

Contents lists available at ScienceDirect

Journal of Visual Languages and Computing

journal homepage: www.elsevier.com/locate/jvlc

Set space diagrams ☆

Björn Gottfried

Centre for Computing Technologies, Universität Bremen, Germany

ARTICLE INFO

Article history:
Received 17 February 2014
Received in revised form
8 April 2014
Accepted 28 April 2014
Available online 9 May 2014

Keywords:
Diagrammatic reasoning
Set theory
Set space diagrams
Boolean algebra
Cardinalities
Euler circles
Venn diagrams

ABSTRACT

This paper introduces *set space diagrams* and defines their formal syntax and semantics. Conventional region based diagrams, like Euler circles and Venn diagrams, represent sets and their intersections by means of overlapping regions. By contrast, set space diagrams provide a certain layout that avoids overlapping geometrical entities. This enables the representation of a good deal of sets without getting diagrams which are cluttered due to overlapping regions. In particular, these diagrams can be employed for illustration purposes, e.g., for showing the laws of Boolean algebras. Additionally, cardinalities are represented and can be easily compared; inferences can be drawn to derive unknown cardinalities from a given knowledge base. The soundness of set space diagrams is shown with respect to their set-theoretic interpretation.

1. Introduction

This paper introduces a diagrammatic representation for set-theory. Indeed, a number of diagrammatic systems exist which represent sets and their relationships [33]. Most of them are based on Euler circles [8] and Venn diagrams [36,37]. What they all have in common is that they represent sets by regions in the two-dimensional plane while topological relations among regions represent subsets, set intersections, and disjoint sets. Therefore, they are altogether referred to as region based diagrams. These diagrammatic systems are probably the most widely used diagrams, since sets play a fundamental role in various areas: They are already deployed early on at school and serve as a useful tool, among others, for illustrating relationships between sets, syllogisms, and statistical data. Nonetheless, these diagrams are not indisputable as will be shown below.

There are two basic issues which arise when studying diagrammatic systems for the representation of sets. They concern the lack of diagrams to directly represent cardinalities and they concern the problem of clutter:

- Facing inspection tasks that require the consideration of cardinalities of sets, conventional diagrams show their confinements. The diagrammatic support of the comparison of cardinalities would be a major advantage, since cardinalities are one of the most fundamental characteristics of sets. For this purpose, it is necessary to encode them graphically instead of employing annotations alone. Just like the relationships between sets in terms of common elements, their relations regarding their cardinalities should be equally well represented by the diagrammatic system.
- Another issue is the resulting layout when depicting a number of sets and their relations. Being essential ingredient of conventional diagrams, overlapping regions depict intersecting sets. But overlapping regions bring in a fundamental disadvantage. Much of what has been called clutter in diagrams is attributed to overlapping objects. For instance, John et al. [17] state in the context of Euler diagrams that those diagrams where most pairs of contours intersect tend to appear more cluttered than those where most pairs are disjoint.

☆ This paper has been recommended for acceptance by Shi Kho Chang.
E-mail address: bg@tzi.de

http://dx.doi.org/10.1016/j.jvlc.2014.04.003
1045-926X/

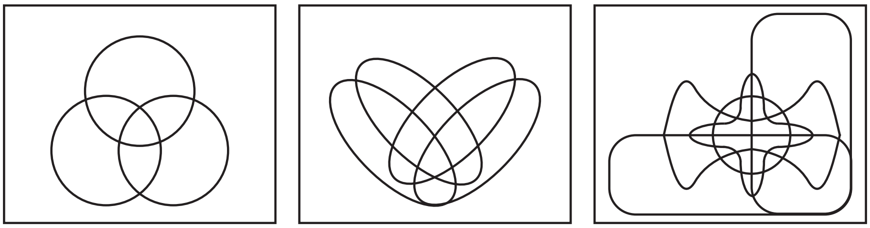

Fig. 1. Three Venn diagrams with three circles, four ellipses, and five curves for representing all set intersections for three, four, and five sets, respectively.

One approach to avoid clutter is the employment of linear diagrams, similar to those described by Cheng [3]. However, while Cheng focusses on the cognitive benefits of linear diagrams, the formal syntax of linear diagrams has not been introduced up to now. The same holds for the formal semantics, which should be defined with respect to the set-theoretic interpretation in order to show how linear diagrams compare to conventional diagrams. Therefore, this paper investigates linear diagrams with respect to their formal characteristics. In this sense, linear set space diagrams introduced in this paper can be conceived of as an extension of the work of Cheng [3]. Moreover, it shows that linear diagrams provide a direct graphical representation of cardinalities.

The body of this paper is structured as follows. At first, the new characteristics of linear diagrams mentioned above are discussed in relation to the state of the art. In the next section, the syntax of the new system is introduced, including the representation of sets, relations among sets, and diagrammatic operations for making inferences. Then, the set-theoretic semantics is adopted for these diagrams, and therewith, the illustration of the laws of Boolean algebras shows to be one of the applications of these diagrams. Afterwards, the representation of cardinalities is discussed. It is demonstrated how a problem can be solved by deploying set space diagrams together with specific inferences to derive unknown cardinalities from a given knowledge base. The paper is closed by a discussion and a summary.

2. State of the art

2.1. Complexity of depiction

Fig. 1 shows some examples of Venn diagrams with up to five sets. Already a Venn diagram with four sets is challenging to inspect. But then, its construction is yet more sophisticated, in particular since it needs to be ensured that no subset intersection is omitted. Taking more than five sets, ever more curved set depictions would be required in order to include all possible set intersections. Grunbaum [13] and Gil et al. [12] illustrate the complexity of their construction. By contrast, Euler circles have the advantage that they do only represent non-empty sets. Therefore, they are frequently much clearer.

All representations of sets that are based on curves in the two-dimensional plane share the property that intersections of curves define zones that represent intersections of sets. Such visualisations can be easily comprehended for two or three sets. However, Venn diagrams become quite complex for more sets, since the number of zones to be depicted grows exponentially by the number of sets involved. In the worst case, when all possible intersections are not empty, the same holds for Euler circles. It is even impossible to draw all these non-empty set intersections with circles for more than three sets [28], other curves than circles making the diagrams even more complex.

While algorithms exist in order to generate such diagrams [6], their inspection is just as difficult, at least for more than three sets. Although operations and their results can be visualised by shading the according intersections, those intersections first of all have to be determined. As Alper et al. [2] and Gil et al. [26] argue, both Euler circles and Venn diagrams are often effective but become cluttered when many sets intersect. Consequently parts of such diagrams are difficult to read. One solution in alleviating the comprehension of those diagrams is to variate the fill properties of zones. But Alper et al. [2] argue that it is unclear how those variations do actually improve the reading of diagrams.

Taking the right hand side diagram in Fig. 1, which is based on the idea of Edwards [5], one is already confronted with $2^5 = 32$ zones when taking five sets, all those zones being defined by a boundary spaghetti of intersecting lines. What makes it difficult to inspect the diagram is that the closed curves are arbitrarily laid out in the plane. As a consequence, diagrams based on two-dimensional curves lack a specific direction for systematically iterating through all zones.

By contrast, diagrams devised by Lewis Carroll look much neater, as shown in Fig. 2 with up to four sets (cf. [5]). But even for these diagrams it becomes difficult to access a specific subset at a glance the more sets there are in the diagram. This problem is common to all kinds of region based diagrams. As a matter of fact, the complexity of Euler based diagrams motivated the introduction of measures for their clutter [17].

2.2. Cardinalities

Besides the complexity of the depiction of region based diagrams, another issue concerns the representation of cardinalities. In fact, there are approaches to represent cardinalities by *area proportional diagrams* [4]. But their construction is even much more difficult, because the insides of arbitrarily complex curves have to stick to correct proportions that represent set cardinalities.

If specific cardinalities are difficult to represent, at least less than relations among them should be deducible. However, such qualitative relations are only clearly visible