Visualization with Euler Diagrams

Case for Support

Peter Rodgers, John Howse, Andrew Fish

1 Track Record

This is a collaborative proposal between the Universities of Kent and Brighton. The team is ideally positioned to carry out the proposed programme of research, bringing together leading researchers in layout, visual tools for diagrams and diagrammatic reasoning. The investigators have current experience of collaborating on the EPSRC grant "Reasoning with Diagrams" [GR/R63509, GR/R63516] (RWD) on which Dr Rodgers was principal investigator at Kent and Prof Howse was principal investigator at Brighton. The reviewers' grades for this grant are not yet available.

The team has taken a lead in the research of Euler diagrams, producing the first work on visualization techniques [10,13,14]. In addition they hosted the first international workshop on Euler diagrams (Euler Diagrams 2004), at the University of Brighton, co-chaired by Dr Rodgers.

The team have developed further international collaborations in this field with the help of an EPSRC grant "Visiting Researcher and Travel for Euler Diagram Research" [GR/T28874/01], on which Dr Rodgers and Prof Howse were investigators. This enabled an exchange of researchers between the UK and the University of Victoria, Canada, where much of the recent important work on area proportional Euler diagrams has been performed. The two overall reviewer grades for this grant were "Outstanding" and "Tending to Outstanding".

Computing Laboratory, University of Kent

The Computing Laboratory's research activities are organized into large and active research groups. Dr Rodgers is a member of both the Theoretical Computer Science Group and the Applied and Interdisciplinary Research Group, as well as coordinator of the Graphs Special Interest Group. In addition, there is significant ongoing research related to this proposal in other groups. The Systems Engineering Research Group has considerable experience of specialization in manipulating diagrams for software engineering and there is also a strong interest in diagrammatic visualization in the Network and Distributed Systems Research Group.

Dr Peter Rodgers (Principal Investigator) is a Senior Lecturer at the University of Kent. His main research interests include the development of graph drawing systems and diagram display technology. He has contributed to much of the recent layout effort on Euler diagrams [1,7,10,13] and has performed empirical studies into the comprehension of Euler diagrams [5]. He is responsible for a number of graph drawing systems and diagram visualization tools. He also has conducted research in applied graph transformation systems [3,4]. He was an investigator on an EPSRC project exploring information visualization technology [GR/R59502]. Dr Rodgers is also lead scientist at Kent on the SegraVis "Syntactic and Semantic Integration of Visual Modelling Techniques" Framework 5 Research Training Network.

He was principal investigator on the successfully completed fast stream grant "Graph Drawing by Graph Rewriting" [GR/M23564], which resulted in an innovative transformation based diagrammatic language for visualizing graph drawing [2,3,4]. It gained positive assessments after completion, with both assessments giving the maximum 'Excellent' rating for management and use of resources, and both assessments indicating that it provided a 'significant contribution to the field'. Other recent work in the area of graph layout methods includes work on force directed approaches [10,12] as well as empirical studies in graph understanding [6]. He has developed a genetic algorithm based graph drawing method, using a new form of graph based crossover operation [9] and has also developed novel graph based visualizations for natural language processing data [8].

School of Computing, Mathematical and Information Sciences, University of Brighton

The host department at Brighton contains several productive research groups. The Visual Modelling Group (VMG), which Prof Howse leads, is acknowledged as internationally leading in the diagrams field. The VMG collaborates with other research groups in the University. Specifically relevant to this proposal is collaborative work between the Interactive Technologies Group and the VMG which includes the joint supervision of a doctoral student on the usability of diagrammatic notations, particularly those based on Euler diagrams.

Prof John Howse (Principal Investigator) is Professor of Mathematics and Computation. His research focuses on the development, formalization and application of formal visual modelling notations [14,15,19,34], and on the development of diagrammatic reasoning systems [16,17,21,23]. He has taught courses to industry and presented conference tutorials in visual modelling and diagrammatic reasoning [20]. He has won Best Paper Award for "Generating Euler Diagrams" [14] at the Diagrams 2002 conference. This is the first paper to describe an automatic visualization mechanism for Euler diagrams.

He was principal investigator on RWD and has been an investigator on the EPSRC projects "Precise Visual Patterns for the Evolutionary Migration of Legacy Systems to Reusable Components" [GR/M02606], "Formal Underpinnings of Object Technology" [GR/K67304], and "Developing and Using Formal Models of Inheritance" [GR/H16629]. He is General Chair for the IEEE Symposium on Visual Languages and Human-Centric Computing 2006, is on the programme committee for various international conferences, and is on the steering committee for the Diagrams conference series.

Dr Andrew Fish (Named Researcher) is a Research Fellow and a member of the VMG. He was an RA on RWD at the University of Brighton. He has a PhD in Geometric Topology from the University of Warwick. His research interests lie in the area of formal diagrammatic systems, in areas such as knot theory [36], formal specification and reasoning [24,29,32,34,35] and graph theory [30]. As well as this work in the application of Euler diagrams, he has published work on their representation [33] and conducted empirical studies of their use [37]. He has also performed a cognitive based study of diagrammatic notations [31] and integrated theoretical and empirical research on human understanding of diagrams [38]. His unique expertise in both diagrammatic systems and topology make him ideal for the proposed project.

Related Publications

The following list contains selected publications from the project team; a further list of general references can be found at the end of the case for support.

1. **P. Rodgers**, P. Mutton and J. Flower. Dynamic Euler Diagram Drawing. In Proceedings IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC'04), pages 147-156. IEEE, 2004.
2. **P. Rodgers**. A Graph Rewriting Programming Language for Graph Drawing. In Proceedings of the 14th IEEE Symposium on Visual Languages (VL98), Halifax, Nova Scotia, Canada. IEEE Computer Society Press, September 1998.
3. **P. Rodgers** and N. Vidal. Pragmatic graph rewriting modifications. In Proceedings of the 1999 IEEE Symposium on Visual Languages. IEEE Computer Society, September 1999.
4. **P. Rodgers** and P. King. A Graph Rewriting Visual Language for Database Programming. The Journal of Visual Languages and Computing, 8(6):641-674. Academic Press, 1997.
5. P. Benoy and **P. Rodgers**. Evaluating the Comprehension of Euler Diagrams. Proc. Euler Diagrams 2005.
6. J. Bovey, P. Benoy, and **P. Rodgers**. Using Games to Investigate Movement for Graph Comprehension. Proc AVI 2004, pages 71-79. ACM, May 2004.
7. S. Chow and **P. Rodgers**. Constructing Area-Proportional Venn and Euler Diagrams with Three Circles. Proc. Euler Diagrams 2005.
8. R. Gaizauskas, **P. Rodgers** and K. Humphreys. Visual Tools for Natural Language Processing. Journal of Visual Languages and Computing, 12(4): 375-412. Academic Press, 2001.
9. M. Hobbs and **P. Rodgers**. Representing space: A hybrid genetic algorithm for aesthetic graph layout. In FEA'98 Frontiers in Evolutionary Algorithms, Proceedings of JCIS'98. The Fourth Joint Conference on Information Sciences, vol 2, pp. 415-418, Oct 1998.
10. P. Mutton, **P. Rodgers** and J. Flower. Drawing Graphs in Euler Diagrams. Proc Diagrams 2004. LNAI 2980, pp. 66-81. Springer Verlag 2004.
11. P. Mutton and **P. Rodgers**. Spring Embedder Preprocessing for WWW Visualization. Proc. IEEE Information Visualization, pp. 744-749. 2002.
12. P. Mutton and **P. Rodgers**. Demonstration of a Preprocessor for the Spring Embedder. GD2002. LNCS 2528. pp. 374-375. 2002.
13. J. Flower, **P. Rodgers** and P. Mutton. Layout Metrics for Euler Diagrams. Proc. IEEE Information Visualization (IV03). pp. 272-280. 2003.
14. J. Flower and **J. Howse**. Generating Euler Diagrams, Proc. Diagrams 2002**,** LNAI 2317, Springer Verlag, 61-75. 2002
15. **J. Howse**, F. Molina, S.-J.Shin and J. Taylor. On Diagram Tokens and Types. Proceedings of Diagrams 2002. LNAI 2317, Springer-Verlag, 146-160. 2002.
16. **J. Howse**, F. Molina, J. Taylor, S. Kent and J. Gil. Spider Diagrams: A Diagrammatic Reasoning System, Journal of Visual Languages and Computing **12**, 299-324. 2001.
17. **J. Howse**, G. Stapleton and J. Taylor. Spider Diagrams. LMS Journal of Computation and Mathematics, 8 pp 145-194. 2005.
18. **J. Howse** and B. Meyer. Conjunction Labels in Euler Diagrams. Euler Diagrams 2004.
19. J. Gil, **J. Howse** and E. Tulchinsky. Positive semantics of projections, Journal of Visual Languages and Computing. Vol. 13, No. 2, Apr 2002, 197-227. 2001.
20. J. Gil, **J. Howse** and S. Kent. Advanced Visual Modelling: Beyond UML, tutorial presented at ICSE, TOOLS, ECOOP, VL/HCC conferences 2000-2002.
21. G. Stapleton, **J. Howse** and J. Taylor. A Constraint Diagram Reasoning System. In Proceedings of Distributed Multimedia Systems, International Conference on Visual Languages and Computing (VLC '03). pp. 263-270, Miami, USA, 2003.
22. G. Stapleton, **J. Howse**, J. Taylor. A Decidable Constraint Diagram Reasoning System. Journal of Logic and Computation 15(6) 975-1008. 2005.
23. G. Stapleton, **J. Howse**, J. Taylor, S. Thompson. What Can Spider Diagrams Say? Proc. Diagrams '04, Int. Conference on the Theory and Application of Diagrams, pp 112-127. 2004.
24. G. Stapleton, S. Thompson, **A. Fish**, **J. Howse**, J. Taylor. A New Language for the Visualization of Logic and Reasoning. Proc. Visual Languages and Computing 2005. pp. 263-170.
25. J. Flower, **J. Howse**, J Taylor. Nesting in Euler diagrams: syntax, semantics and construction, Journal of Software and Systems Modeling, pp 55-67, 2003.
26. **J. Howse**, S. Schuman. Precise Visual Modelling: a case study. Software and Systems Modelling Journal, Vol.4, No.3, pp 310-325, 2005.
27. G. Stapleton, S. Thompson, **J. Howse**, J. Taylor. The Expressiveness of Spider Diagrams. Journal of Logic and Computation 14(6), 857-880. 2004
28. G. Stapleton, **J. Howse**, J. Taylor, S. Thompson. The Expressiveness of Spider Diagrams Augmented with Constants. Proc. VL/HCC'04, Sept. 2004, pp 91 - 98.
29. **A. Fish**, **J. Howse**. Towards a Default Reading for Constraint Diagrams. In Proc. Diagrams '04, International Conference on the Theory and Application of Diagrams, Cambridge, March 2004, pp 51-65, LNAI 2980.
30. **A. Fish, J. Howse**. Computing Reading Trees for Constraint Diagrams. In Proceedings of Applications of Graph Transformations with Industrial Relevance, pp 260-274, LNCS 3062, Springer, 2003.
31. **A. Fish**, **J. Howse**, G. Taenzer and J. Winkleman. Two visualizations of OCL: A comparison. University of Brighton Technical Report. 2005.
32. **A. Fish**, J. Flower, **J. Howse**. The Semantics of Augmented Constraint Diagrams. Journal of Visual Languages and Computing16(6) pp. 541-573 2005.
33. **A. Fish**, J. Flower. Abstractions of Euler Diagrams. Proc. Euler Diagrams 2004, ENTCS 134, pp. 77-101, 2005.
34. **A. Fish**, J. Flower and **J. Howse**. A Reading Algorithm for Constraint Diagrams. In Proceedings IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC'03), pp 161-168. IEEE, 2003.
35. **A. Fish**, J. Flower. Investigating Reasoning Rules for Constraint Diagrams. In Proc. of Visual Languages and Formal Methods 2004, ENTCS 127, pp. 53-69, 2005.
36. **A. Fish**, E. Keyman. Jones Polynomial Invariants. Journal of Knot Theory and its Ramifications 15(3), pp. 1-12, World Scientific Publishing Company, 2006.
37. **A. Fish** and J. Masthoff. An Empirical Study into the Default Reading of Constraint Diagrams. Proc. IEEE Symp. on Visual Languages and Human-Centric Computing 2005. pp 287-289.
38. C. John, **A. Fish**, **J. Howse**, and J. Taylor. Exploring the Notion of Clutter in Euler Diagrams. Accepted for Diagrams 2006.

2 Proposed Research

2.1 Summary and Motivation

This research proposal will revolutionize the field of Euler diagram drawing. We will create a substantial body of theory which will enable the widespread use of Euler diagram visualizations. The project will develop original techniques to allow the practical layout of diagrams that currently cannot be drawn automatically. A major outcome will be the production of a usable software tool exemplifying the new layout techniques.

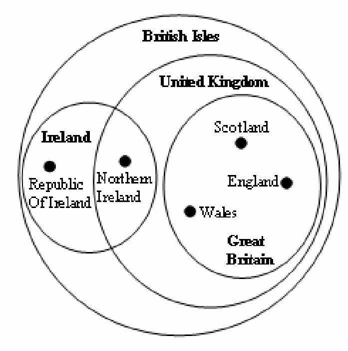

Figure 1: British Isles Euler Diagram by Sam Hughes

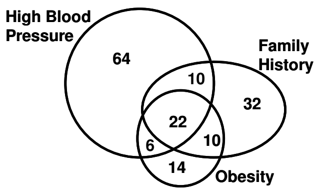

Figure 2: Heart Disease Data

Euler diagrams (a generalization of Venn diagrams) are an attractive method for visualization information because they represent exclusion, containment and intersection of sets in an intuitive manner, see Figures 1 and 2. In addition, they have the advantage of a rigorous mathematical basis [14,15]. However, despite their benefits as a visualization method, the practical use of Euler diagrams has been held back because users must lay out these diagrams by hand. Manual layout places a heavy burden on users and is impractical for Euler diagrams involving many sets.

Relevance and Significance of Euler Diagram Drawing Whilst diagrams of the type developed by Euler have been in use for centuries, automated Euler diagram drawing is a new field. Since the first work appeared in 2002 (much of which was developed by members of this team [10,13,14]), two International Euler Diagram Workshops have been held: Brighton in 2004 and Paris in 2005.

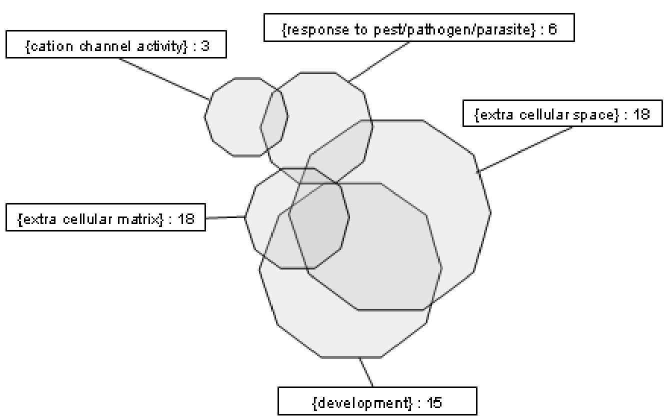

Figure 3: GoMiner Visualizing Genetic Set Relations in a Gene Ontology Database [51]

Euler diagrams are used in visualizing data of all sorts, particularly data associated with medical experiments or biological processes, for example [39,51], see Figures 3 and 4. They are also used to visualize data from artificial intelligence applications [55], see Figure 5. Other application areas apply Euler diagrams to visualize complex hierarchies. For example, VENNFS [43] visualizes file system organization by allowing files to appear in more than one directory and [59] proposes using Euler diagrams to visualize large databases involving multiple classifications.

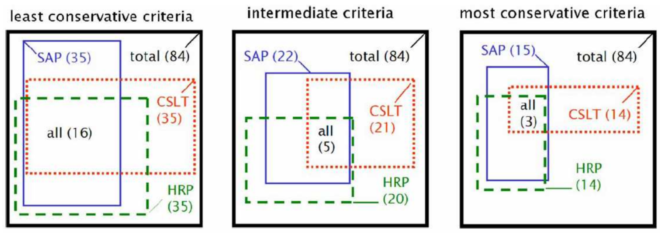

Figure 4: Agreement Between Glaucoma Tests [39]

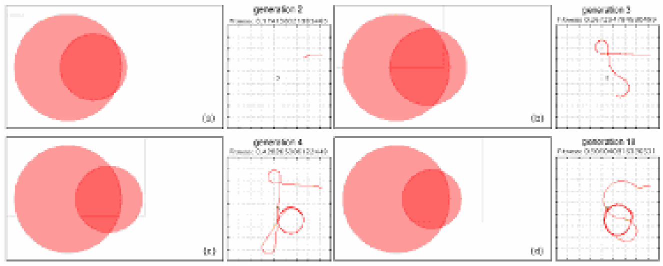

Figure 5: Robot Trajectory Data [55]

A wide variety of diagrammatic reasoning systems are based on Euler diagrams [17,22,47,57,58] as are many of the UML notations. The project team have been involved in developing the constraint diagram notation [21,32], see Figure 6, which is based on Euler diagrams enhanced with graphs. These diagrams have been used to write constraints on network services in the telecoms domain and for meta-modelling [26,52].

Euler diagrams are closely related to extended graph notations such as hypergraphs and higraphs [48,49]. These are graphs with some notion of containment and are widely used. For example, several UML diagram types use variants of graph notation extended with containment.

There are widely available software tools for visualizing Euler diagrams. Microsoft PowerPoint 2003 includes a 'Venn diagram' option as one of its six built in diagram types; in fact, this option produces Euler diagrams when the number of circles increases beyond three. The site venndiagrams.com has a database of over 10,000 diagrams created by users of its online application, which visualizes 3-set Venn diagrams. However, these tools are heavily restricted in the diagrams that can be created and include little automated layout.

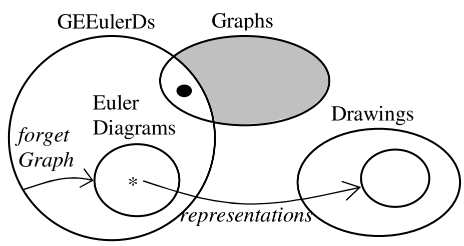

Figure 6: A visual logic statement [21]

Key Issues in Euler Diagram Drawing The lack of a general Euler diagram layout method is due to difficult unsolved theoretical issues concerning the generation of a drawn (concrete) diagram from an abstract description. For various classes of abstract descriptions a concrete diagram can only be drawn by relaxing various well-formedness conditions (see Section 2.2 for more details). This project will solve many of the outstanding problems and develop a general theory of the drawability of Euler diagrams.

For usability reasons, it is not enough to simply draw a diagram in the plane; the drawing must also be easy to understand. We will develop a set of novel automatic layout methods, targeted at the application domains. Combined with the software developed from the theory work, these methods will allow the generation of a user-friendly visualization from a wide range of data sources.

Our Approach This project will develop new general Euler diagram drawing methods as well as examining the two common classes of area proportional diagrams and sparse diagrams:

- Although the work in laying out Euler diagrams has been started [1,10,13,14] there are currently severe restrictions on which abstract descriptions can be drawn, on how effective the layout is for use in applications, and on how quickly the layout can be generated. This project will develop the theory required to allow all abstract descriptions to be visualized and produce efficient techniques that can be used to improve the understandability of the resulting diagrams.
- An area proportional Euler diagram has regions whose areas are proportional to the cardinalities of the sets represented. These are popular for data visualization in biological and medical studies (for example, see Figures 2 and 3). However, no tool can currently perform the task of accurately laying out area proportional Euler diagrams with more than two closed curves, using ellipses. We will develop techniques to draw area proportional Euler diagrams with circles, ellipses and other geometric shapes for three and four sets. We will also examine the theoretical limits on what combinations of set sizes can be drawn exactly and produce methods to draw approximations where exact solutions are not possible.
- Sparse Euler diagrams contain few set intersections in proportion to the number of closed curves. For example, sparse Euler diagrams are often used when visualizing multiple file classifications [43]. As with area proportional diagrams, ideally simple geometric shapes will be used in the Euler diagrams. We will identify levels of sparseness for which diagrams can be drawn using certain geometric shapes. We will develop rapid layout methods to avoid significant feedback delay, a strong requirement for the interactive applications envisaged.

There is a long record of successful research in the automatic layout of graphs and Venn diagrams [41,50,56], whereas the layout of Euler diagrams has not yet been widely investigated. The research to this point in automated Euler diagram layout has necessitated a combination of theoretical research alongside the implementation of layout methods. We intend to continue this two pronged approach, and are ideally placed to perform the work, with a strong basis of theoretical work at Brighton allied with a track record of developing diagram drawing software at Kent.

Adventure in Research The project outlined in this proposal includes significant novelty. The area of Euler diagram visualization is new and developing, with the first work appearing within the last five years.

Classifying which abstract diagrams can be drawn under each set of well-formedness conditions and generating concrete diagrams are ambitious tasks. There are many hard topological and combinatorial problems to overcome. General results are feasible and will revolutionize the field of Euler diagram drawing. Even solutions to these problems for some classes of abstract diagrams would be very beneficial and will facilitate the drawing of commonly used diagrams. Investigating different diagram abstractions is innovative and will provide new insights.

The approaches to Euler diagram layout we will take are novel. The force-based models we will develop have not previously been considered for Euler diagrams. Our pattern-based technique is original, and such an approach has not been used for visualizing structures such as graphs or Venn diagrams.

A general solution for area proportional diagrams, for up to four contours, is challenging but likely to be successful. This research will involve the solution of hard geometric and constraint satisfaction problems. It will classify which area proportional diagrams are drawable with particular geometric shapes.

The theory of sparse diagrams is new and exciting. The relationship between sparseness and drawability under particular well-formedness conditions has not previously been explored, and will produce interesting research challenges, the solution of which will encourage large scale take up of Euler diagrams in many application areas.

Overall, the results and techniques developed will generate a significant advance in Euler diagram visualization and will stimulate further research beyond the lifetime of the project.

2.2 Background

An Euler diagram is shown in Figure 1. The four *contours* represent sets. A contour is a closed curve in the plane. The interiors of the contours are called *regions*, and the minimal non-empty intersections of these regions are called *zones*. Each zone can be described by the contours that contain the zone and those that exclude the zone. A Venn diagram [56] is an Euler diagram where all possible set intersections appear as zones. For instance, Figure 2 shows a Venn diagram with three contours. As the number of contours gets larger, Venn diagrams become increasingly cluttered; this is less a problem for Euler diagrams, which only need to show the required zones, for instance, the Euler diagram in Figure 7a does not have the zone formed from the intersection of all three sets.

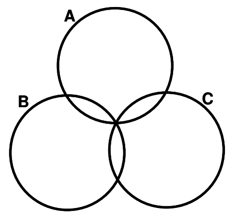

Figure 7a: A triple point

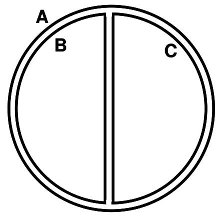

Figure 7b: Concurrent contours

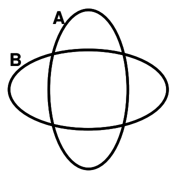

Figure 7c: Disconnected zones

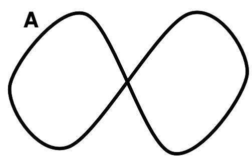

Figure 7d: A self intersecting contour

We make a distinction between a *concrete* diagram, one that is drawn in the plane, and an *abstract* diagram, which is a mathematical description of the zones in a concrete diagram. In essence, an abstract diagram is a list of sets of contours which contain zones. For example, $\{\{A\}, \{B\}, \{C\}, \{A,B\}, \{A,C\}, \{B,C\}\}$ is the abstract diagram for the concrete diagram in Figure 7a.

Well-formedness conditions are usually imposed on concrete Euler diagrams. For example, contours being restricted to particular shapes, such as ellipses; no triple points (which prevents more than two contours intersecting at any point); no concurrent contours; no disconnected zones (from which it follows that all zones are connected components of the plane); allowing only simple contours (self intersecting contours are non-simple). See Figure 7 for some examples of Euler diagrams breaking various well-formedness conditions.

An abstract diagram is *drawable* under a given set of well-formedness conditions provided that there is a well-formed concrete diagram with that abstraction. Different choices of well-formedness conditions give different classes of drawable abstract diagrams – classifying which abstract diagrams are drawable under various well-formedness conditions is a rich area for new theoretical research. It is known that any abstract diagram can be drawn when no well-formedness conditions are imposed. However, imposing some well-formedness conditions is important for diagram comprehension, the choice of which is dependent on the user. For example, in Figure 7b, had the contours really been drawn concurrently, it would have been impossible to distinguish them.

Apart from the investigators' work, there have been other efforts to automatically draw Euler diagrams. In [42] an Euler diagram drawing system is presented that generates area proportional diagrams with either two circles (see Figure 5) or with up to four contours in a rectilinear layout (see Figure 6). In [59] an initial concrete representation method is presented that draws Euler diagrams, but at the expense of some key well-formedness conditions. GoMiner [51] (see Figure 3) visualizes area proportional Euler diagrams using circles, with an optimizing technique to place the circles in the diagram, but because of the restriction to circles, some required zones do not appear, and others appear that should not be there.

A key problem is to find an algorithm which produces a concrete diagram from an abstract diagram under any given well-formedness conditions. In [14] an algorithm is presented for one choice of well-formedness conditions, under which many abstract diagrams cannot be drawn (such as those in figures 7a and 7b). For practical purposes it is essential that an algorithm is developed for the general case.

A concrete diagram produced using any of the above methods is typically not satisfactory as a usable visualization. A metric based layout technique has been developed in order to improve the aesthetic appearance of concrete diagrams [13], including Euler diagrams that have graphs [10]. However, the method scales poorly and does not produce a good layout in all cases.

Graph drawing is the process of automatically laying out a graph on a screen or page so that it has a good appearance. There are parallels between graph drawing and Euler diagram drawing, and we expect some techniques from graph drawing can be adapted for Euler diagram drawing. Every Euler diagram has a *topological dual graph*, which is formed by placing a node in each zone and edges are placed between nodes in adjacent zones. The topological dual graph is often used to generate Euler diagrams and needs laying out using graph drawing methods. A good layout of the topological dual graph assists with generating an aesthetically pleasing Euler diagram. There are various techniques to achieve a good graph drawing, some of which depend on the graph belonging to particular classes, for example that it is a tree, has a notion of flow, or that it is planar; see [50] for a summary of such techniques. These restricted graphs can typically be drawn more easily than general graphs. Other methods work on more general graphs such as force directed techniques, most of which are based on Eades' Spring Embedder [44], which can produce a reasonable layout in a quick time. Techniques pioneered by Purchase [53,54] have been used to empirically evaluate the effectiveness of layout methods and metrics.

2.3 Program and Methodology

2.3.1 Aims and Objectives

The aim of the project is to facilitate the large scale take up of Euler diagrams by developing a body of theoretical research and producing visualization software tools. The specific objectives are:

1. Develop the theory of general Euler diagrams and produce novel general Euler diagram drawing techniques using a *pattern-based approach* and a *force-based approach*.
2. Comprehensively analyse the theory behind three- and four-contour area proportional Euler diagrams, and implement software for visualizing them.
3. Develop the theory of sparse Euler diagrams and apply it to drawing such diagrams.
4. Develop a software system allowing access to the tools and techniques developed in the project, both as an executable tool and as a library for use within other applications.

2.3.2 Developmental Methodology

From our experience, a rapid prototyping methodology is the most appropriate software engineering technique for this kind of project. This will be aided by the results from the recently completed project of the investigators, exploring reasoning with diagrams, which supported the initial work on Euler diagram layout. Hence a rapid start to the project is envisaged, with original work being performed from the beginning.

2.3.3 Workplan

This work plan should be read in conjunction with the project plan, given on Page 9. The project is divided up into five major work packages relating to the four objectives given in Section 2.3.1, together with an evaluation work package. Dissemination is assumed to be an integral part of each work package. The effort is shared between the two sites, with, broadly, the theory work performed at Brighton and the implementation work performed at Kent. Evaluation will be performed at both sites. The project is planned to start at the same time at the two sites, ensuring that all project members can communicate about each work package. In this project, theory and tooling are strongly interlinked. The tools will be built on the specification given by the theory, but the development of tools will feedback into the theory. Furthermore, the evaluation of theory and tools will inform their development.

Work Package 1: General Euler Diagrams

This work package, the largest in the project, will address the key open issues that prevent the wider application of Euler diagrams. As with graph drawing, different layout methods are appropriate to different types of Euler diagram. We will develop methods that will work with any combination of well-formedness conditions to ensure that all abstract diagrams can be drawn. Furthermore, we will produce methods that generate very effective drawings for commonly used classes of Euler diagrams.

Work Package 1.1: Embedding Methods

Previous work on embedding topological dual graphs in the plane to produce Euler diagrams will be significantly extended to enumerate the set of topologically distinct planar embeddings. A decision mechanism for indicating a preferred embedding, based on graph characteristics, will also be defined.

Whilst ambitious, the possible utility of the results of this work outweigh the risks involved. Furthermore, partial results (limiting the number of contours, or considering special cases) are expected as these can build on existing Venn diagram research, for example the number of topologically distinct Venn diagrams has been determined for up to three contours [40], and for up to five contours when the Venn diagram is well-formed in the sense of having no disconnected zones or multiple points [46]. Restricted results of this form would be sufficient for generating drawings of many commonly encountered Euler diagrams.

Work Package 1.2: Well-formedness

Existing work on generating diagrams identifies some abstract diagrams as undrawable subject to (very restrictive) well-formedness conditions (see Section 2.2). It is desirable to relax these conditions in a manner that depends on user requirements. We will classify diagrams that can be drawn under various well-formedness conditions and identify graph-theoretic properties of dual graphs that encapsulate subsets of the well-formedness conditions. Associated with an abstract diagram is an abstract dual graph, see [14] for details. Of particular interest are those properties that are preserved under the algorithm used to transform an abstract graph into a topological dual graph. However, these are difficult problems and it may not be possible to produce complete solutions. Therefore, we will also investigate alternative abstractions (for example, viewing a diagram as a sequence in which the contours are added one at a time) which may help to identify certain properties more easily.

There exist abstract diagrams that have various concrete visualizations satisfying different well-formedness conditions. For example, it is possible to draw some abstract diagrams either with a triple point or with disconnected zones. Using current techniques, when an abstract diagram is deemed "undrawable" subject to the restrictive conditions, the algorithm terminates without resolving the problem. This work will be extended, allowing users a choice of well-formedness conditions under which the abstract diagram can be drawn. Currently no software system allows all abstract Euler diagrams to be drawn. Along with WP 1.1, this work will mean that all abstract diagrams are automatically drawable by the software, when implemented in WP 1.6.

Work Package 1.3: Layout Using Diagram Patterns

This work package will develop a pattern based layout technique for Euler diagram drawing within the software. The method will rely on having a known layout for a set of diagrams (a library of examples). From this library a diagram is selected and subsequently modified to produce the required diagram. A principal advantage of this method is that no concrete representation algorithm is required to generate an initial layout. The library will include some Venn diagrams, as well as frequently-used Euler diagrams. The modifications to the library diagram may include both removal and addition of zones and contours. Various strategies are possible when modifying diagrams selected from the library; it may be possible to make a small local change to a single contour, although larger changes may be needed.

This technique will require the implementation of a structure comparison algorithm to determine "how far apart" abstract diagrams are, both in order to choose the best initial diagram to start the layout method, and to decide what further changes are required to transform the initial diagram into the required diagram. The pattern-based approach will be extended, incorporating the theory of nested diagrams [25]. A further extension is to combine library diagrams in more complex ways to produce the required diagrams.

The pattern-based approach will allow the efficient generation of many diagrams. However, it is unlikely that all abstract diagrams could be generated by this method: the library would need to be very large and developing a complete set of transformation rules is unlikely to be feasible. Furthermore, there is a balance to be struck between the size of the library and the efficiency benefits that this approach will bring over the general case.

Work Package 1.4: Force Directed Layout Approach

This work package will develop various software force models for improving Euler diagram layout. The work here is inspired by force directed graph drawing techniques, which typically apply forces between nodes, or between edges and nodes. When applying a force model to Euler diagrams, there are various types of attraction and repulsion that might be used. The first investigation, in the case where the contours are polygons, will look at forces between vertices of the polygons, and the edges joining the vertices. We will also investigate forces acting between these vertices and edges, forces acting on the centre of the contours, and the centre of zones. We will attempt to obtain a force model that nicely lays out the diagram and maintains the abstract description. We will be able to achieve this for a substantial subset of diagrams. However, we cannot rule out the need to maintain the abstract description by checking the result of force applications. The force directed method requires an initial concrete diagram, such as those generated in WP 1.1 and 1.3.

Work Package 1.5: Theoretical Evaluation

The pattern-based approach requires diagrams from a library to be adapted in order to create a drawing of a given abstract diagram. The comprehensiveness of the library of examples and the diagrams that can be generated from it can be established combinatorially. We can count the total number of diagrams satisfying certain conditions (for example, having a fixed number of contours and under certain well-formedness conditions) and compare this with the number of diagrams with such properties that we can generate from our library.

This combinatorial approach will also be used to inform the development of the library and to indicate the necessity for new diagram generation operations within the library. Also, if two library examples present themselves as possible starting points, theoretical work will justify why one should be chosen over another.

Theoretical results for the force-based approach will evaluate which kinds of forces can be guaranteed to terminate, and which can be guaranteed not to change the abstract description of the diagram.

Work Package 1.6: Theory Implementation

We will implement the theoretical techniques developed in WP 1.1 and 1.2, including facilities to allow users to select well-formedness conditions. The theoretical results obtained in WP 1.5 will be implemented and used to provide feedback to users.

Work Package 2: Area Proportional Euler Diagrams

This work package will concentrate on Euler diagrams of up to four contours, drawn with shapes such as circles and ellipses. As these area proportional diagrams are typically used in statistical visualization, it is important that the information contained in the diagram is easily accessible; the restrictions we place on the number and shape of the contours ensure this is the case. We will develop a complete set of theoretical results for well-formedness, as well as implement layout methods for all cases of diagram.

Work Package 2.1: Area Proportional Theory

It is known that any area proportional Euler diagrams can be drawn with two circles [42]. However, there are combinations of zone areas using three contours that cannot be drawn with convex contours. A large number of cases can be drawn with circles or ellipses and this work package will classify which area proportional diagrams can be drawn with such shapes. The analysis will be difficult because of the nature of the constraints involved. There are no known (non-trivial) results for area proportional diagrams with four contours, and we will extend the work to this case.

Work Package 2.2: **Area Proportional Implementation**

This work package will implement the layout of area proportional Euler diagrams of up to four contours. Using the theory produced in WP 2.1, diagrams will be drawn with circles or ellipses where possible. In addition, an approximation mechanism for diagrams that cannot be laid out exactly with ellipses will be developed, as in many cases an impression of relative size is all that is required.

The software developed in this WP will be included in the software environment and released as a stand-alone applet on the web. This will enable easy access to the software for anyone wishing to visualize statistical information using area proportional Euler diagrams.

Work Package 3: Sparse Diagrams

This work package is designed to assist the use of Euler diagrams in applications such as multiple category file browsing, database queries, software modelling and data visualization. These areas typically make use of sparse Euler diagrams. Often these applications are dynamic, with the diagrams changing as information is updated.

Work Package 3.1: Sparse Diagrams Theory

This work package will establish how Euler diagrams sparseness is related to properties, such as being drawn with circles or ellipses. The package also includes a dynamic layout component: transformation rules will be developed to update a diagram as information changes maintaining as much of the original diagram as possible.

Work Package 3.2: Sparse Diagrams Implementation

This work will involve implementing an example system, displaying sparse diagrams with circles and ellipses. The software must cope with the dynamic layout of Euler diagrams, and will use concepts from dynamic graph drawing. An important feature of good dynamic layout systems is that they avoid major changes, where possible, to the diagram visualization when the abstract description is altered, because the user has invested time in understanding the current layout. However, current Euler diagram drawing mechanisms may produce very different layouts for diagrams with a similar abstract description. We will develop methods to adapt layouts incrementally as the abstract description changes.

Work Package 4: Visualization Software Environment

This work package concentrates on the software infrastructure aspects of the project.

Work Package 4.1: Initial Software Development

This involves the implementation of a skeleton display environment, providing user interaction, including diagram editing, the diagram generation algorithm in [14], and facilities to support the layout methods. The software will include the existing metric based layout technique [13] and area proportional layout [7,42] for benchmarking.

Work Package 4.2: Software Finalization

This work package is designed to provide time to deliver the software, packaging it in a suitable manner to allow simple access by users and to allow for further development by interested parties.

Work Package 5: Evaluation

The research team has experience in empirical evaluation in areas such as: Euler diagram visualization [5,38], the application of Euler diagrams [37] and other diagrammatic systems [6]. In addition, the team has close collaboration with internationally respected empiricists: Prof Peter Cheng at Sussex, with whom Prof Howse will be supervising a PhD student in the area of empirically evaluating Euler diagrams and Dr Helen Purchase at Glasgow is planning to spend a significant part of her 2008 sabbatical at Kent. Dr Fish is collaborating with Dr Judith Masthoff, Aberdeen, on empirical studies evaluating diagram comprehension [37].

Task based evaluation will be performed, obtaining quantitative data (for example, time taken and success rates) and qualitative data (obtained from questionnaires) from around 30-50 subjects per study; typically, subjects will be drawn from the local student cohort. The tasks will involve the interpretation of Euler diagrams, drawn by the tool, and the analysis of the data contained within. For example, the correct identification of statistical data in area proportional diagrams, and the correct identification of the zones that appear in Euler diagrams.

In-depth studies of the use of the software for generating diagrams in an application context will be performed. They will follow a user through an example, the creation and modification of an Euler diagram in a modelling exercise. The studies will be more focussed, using a smaller number of expert users, around 6-8 subjects per study, mostly drawn from professionals working in the team’s universities. Feedback will be obtained from questionnaires and think aloud protocols.

We will target the application areas in our studies. For example, area proportional Euler diagrams are used to visualize medical and biological data; a task-based approach will be performed to measure the effect of the use of different geometric shapes for contours to see which are most effective in communicating information.

User tests will investigate the effect of various well-formedness conditions on diagram interpretation. We will use the findings to prioritize theoretical work.

2.4 Relevance to Beneficiaries

The output of this project will be the production of important theoretical results and the development of usable software for visualizing Euler diagrams. The creation of this software environment will have significant impact, facilitating the widespread use of Euler diagrams in diverse application areas. Users will have access to software that produces visually appealing Euler diagrams. The software will be produced in such a way that users and researchers can easily access and modify the code which will increase the impact of the research well beyond the lifetime of the project.

Information Visualization Users

People wanting to use Euler diagrams to visualize information such as statistical data, multiple hierarchies or logic will, as a result of this project, have access to powerful visualization software that displays diagrams to the required specification.

Euler Diagrams Research Community

This community will get an important boost to the research in this area. They will be able to build on significant new theoretical results and have access to novel layout and analysis tools. In addition, the empirical studies will inform researchers about the qualities that should be exhibited by Euler diagrams to enhance their usability.

2.5 Dissemination and Exploitation

The usual channels will be used for publication including conferences: Euler Diagrams, Diagrams, GD, AVI, HCI, IV, VL/HCC, VLC, ICGT, SoftVis and VisSym; and journals: Journal of Visual Languages and Computing, Journal of Graph Algorithms and Applications, Information Visualization Journal, IEEE Transactions on Visualization and Computer Graphics, the Electronic Journal of Combinatorics and the European Journal of Combinatorics. Once the software has been sufficiently developed, we will make the software freely available under an appropriate license in line with standard academic practice.

Further dissemination will be possible through the organisation of workshops and tutorials. A web site will be produced and maintained allowing the download of publications and software resulting from the project. Appropriate software will also be presented as Java applets to encourage use.

2.6 Justification of Resources

Staff There is a large amount of both innovative and detailed work in this project, which will require full time effort at each site for three years The challenges of the work require researchers with high levels of technical ability and research flair and we will therefore need to employ the two named research assistants. Support is requested for a fixed-term 20% FTE employment of a project management assistant at Kent and 10% of an administrator at Brighton. The requested new, project specific administrative support will aid project organization, enabling the team to devote as much time as possible to the research and management of the project.

Investigators The Principle Investigators will be involved in: the supervision and management of the Research Associate at their sites (60 hours p.a.), project management, such as attendance at project meetings (120 hours p.a.), reporting and research (260 hours p.a.) for a total time of 440 hours p.a. The salaries included in the proposal have been calculated using pay scales currently available to the institutions. New pay structures will be introduced during the lifetime of this project and, as agreed with RCUK, reimbursement of any net cost increases will be expected at the reconciliation stage.

Equipment A laptop and peripherals for the RA at each site. We consider laptops to be standard equipment for postdoctoral researchers and their productivity benefits justify their relatively low cost.

Consumables Project specific software and computing supplies are requested at rate of £2000, to include upgrades over the course of the project, and £500 for books at each site.

Payments to Experimental Subjects A payment of £10 per hour is requested to pay experimental subjects (estimated for a total of 200 subject hours for five studies), see WP 5.

Travel We expect the project to be successful, and therefore to generate a significant number of papers. The travel budget reflects this. Funds are requested at the rate of one international (£1500 per trip) and one European conference (£800 per trip) per person per annum. The investigators collaborated successfully on the Reasoning with Diagrams EPSRC grant, where the frequency of trips between sites averaged one a month, with a full project meeting every 6 months and we expect to maintain this successful working pattern. Funds have been requested for travel between the two sites at £60 per person per trip.

We also request funds for travel to utilize relationships with internationally respected collaborators. Trips in the UK are planned to empirical researchers Dr. Purchase in Glasgow and Dr Masthoff in Aberdeen at a cost of £300 each. Internationally, trips are planned to Mr Chow and the Combinatorial Algorithm Research Group, University of Victoria, Canada, with whom Dr Rodgers has previously collaborated [7] and to Dr Meyer, Monash University, Australia, with whom Prof Howse has worked on reducing clutter in Euler diagrams [18]. These international trips are estimated to cost £1500 each.

References

39. P. H. Artes and B. C. Chauhan. Longitudinal changes in the visual field and optic disc in glaucoma. Progress in Retinal and Eye Research, Volume 24, Issue 3, May 2005, Pages 333-354.
40. K. B. Chilakamarri, P. Hamburger and R. E. Pippert. Analysis of Venn Diagrams Using Cycles in Graphs, Geometriae Dedicata, vo1. 82 pp. 193-223. 2000.
41. I.G. Tollis, G. Di Battista, P. Eades, R. Tamassia. Graph Drawing: Algorithms for the Visualisation of Graphs. Prentice-Hall. 1999.
42. S. Chow and F. Ruskey. Drawing Area-Proportional Venn and Euler Diagrams. Proceedings of GD2003. LNCS 2912, pp.466-477, Springer Verlag, 2003.
43. R. De Chiara, U. Erra and V. Scarano. VENNFS: A Venn-Diagram File Manager. Proc. IEEE Information Visualization (IV03). pp. 120-126. 2003.
44. P. Eades. A Heuristic for Graph Drawing. Congressus Numerantium 42. pp. 149-60. 1984.
45. J. Griggs, C. E. Killian and C. D. Savage. Venn Diagrams and Symmetric Chain Decompositions in Boolean Lattice. Electronic Journal of Combinatorics 11(1). 2004.
46. P. Hamburger and R. E. Pippert. Simple, Reducible Venn Diagrams on Five Curves and Hamiltonian Cycles. Geometriae Dedicata, vo1. 68 pp. 245-262. 1997.
47. E. M. Hammer. *Logic and Visual Information*, CSLI Publications. 1995.
48. David Harel. On Visual Formalisms. Communications of the ACM. 31(5). pp. 514-530. 1988.
49. Higraph web page: http://db.uwaterloo.ca/~gweddell/higraph/higraph.html.
50. M. Kaufmann and D. Wagner. Drawing Graphs: Methods and Models, LNCS 2025. 2001.
51. H. A. Kestler, A. Müller, T.M. Gress and M. Buchholz. Generalized Venn diagrams: a new method of visualizing complex genetic set relations. Bioinformatics 21(8) 2005.
52. S.-K. Kim and D. Carrington, Visualization of Formal Specifications, 6th Asia Pacific Software Engineering Conference, 102-109 IEEE 1999.
53. H. C. Purchase. Performance of Layout Algorithms: Comprehension, not Computation, J. Visual Languages and Computing, 9, pp. 647-657, 1998.
54. H. C. Purchase. Evaluating Graph Drawing Aesthetics: Defining and Exploring a New Empirical Research Area, Computer Graphics and Multimedia. DiMarco, J. (ed), Idea Group Publishing, pp 145-178, 2004.
55. T. Quick, C. Nehaniv, K. Dautenhahn and G. Roberts. University College London Research Note RN/05/29.
56. F. Ruskey. A Survey of Venn Diagrams. The Electronic Journal of Combinatorics. March 2001.
57. S-J Shin. *The Logical Status of Diagrams*. CUP. 1994.
58. N. Swoboda and G. Allwein. Using DAG Transformations to Verify Euler/Venn Homogeneous and Euler/Venn Hetrogeneous Rules of Inference. Journal of Software and Systems Modeling 3:2 pp. 136-149. 2004.
59. A. Verroust and M.-L. Viaud. Ensuring the Drawability of Extended Euler Diagrams for up to 8 Sets. Proc.. Diagrams 2004. LNAI 2980. Springer Verlag. pp. 128-141. 2004.
60. D. Winterstein, A. Bundy, C. Gurr, and M. Jamnik. Using Animation in Diagrammatic Theorem Proving. In Diagrams 2002. LNAI 2317. pp. 46-61. Springer-Verlag. 2002.

Diagrammatic Project Plan