A Fast Way to Extend Simple Venn Diagrams

Chris Winsor
CSC 482

Abstract: Using Venn's construction, a fast algorithm is developed to find a Hamilton cycle in the dual graph of simple Venn diagrams with a high success rate. This algorithm is then used to search through all simple Venn diagrams with six or less curves in an attempt to find a counter example or method of proof for Winkler's conjecture. This leads to some remarks on the structure of the dual graph, and directions on improving the algorithm.

1 Definitions and Notation

An *undirected graph* $G = (V, E)$ which will be referred to as a graph, consists of a set V of *vertices* and a set E of *edges* that correspond to unordered pairs of vertices. Two vertices are *adjacent* if there is an edge between them. A (u,v)-*path* is a sequence of unique vertices starting with vertex u and ending with vertex v in which consecutive vertices are adjacent. A *cycle* is a path that ends with the same vertex it starts with, sharing only that vertex. A *Hamiltonian* cycle is a cycle that uses every vertex in the graph. A graph is *connected* if there is a path between every pair of vertices. A graph $H = (V', E')$

is a *subgraph* of a graph $G = (V, E)$ if V' and E' are both subsets of V and E respectively. A *connected component* of a graph G is a connected subgraph that contains all vertices mutually reachable by a path in G and all edges between those vertices that are in the edge set of G.

A *planar* graph is a graph that can be drawn on the plane with no edges intersecting except at a vertex. Such a drawing is called a *planar embedding* of the graph. If you were to draw a planar embedding of a graph on a piece of paper and cut along the edges and vertices, then the resulting connected pieces of paper correspond to the *faces* of the graph. The *dual* of a planar graph is the graph where the vertices correspond to each face in the planar graph and for all edges in the planar graph, if that edge lies on two faces, then there is an edge in the dual between the two vertices in the dual that correspond with the two faces.

A *simple closed curve* is a connected curve that does not cross itself and ends at the same point a it starts. From now on, let curve refer to a simple closed curve. Suppose there are n curves $C_1, C_2, \ldots, C_n$ and let R_i refer to the region interior or exterior to the curve C_i. A region is *empty* if it contains no points within it. A region is *connected* if there exists a path between every pair of points in the region that is fully contained in the region. A *diagram* is a fixed position of all curves in the plane. Now consider the 2^n possible intersections $R_1 \cap R_2 \cap \cdots \cap R_n$. A diagram is *Euler* if every one of the 2^n intersections is connected. A diagram is an *Independent Family* if no intersection is empty. A diagram is *Venn* if it is Euler and no intersection is empty. A Venn diagram is *simple* if at most two curves intersect at any single point. An n-Venn diagram is a Venn diagram that is made with n curves. An n-Venn diagram is *extendible* if the addition of some new curve results in a $(n+1)$-Venn diagram.

2 History

Venn diagrams were introduced by and named after logician John Venn back in the late 1800s. Venn presented Venn diagrams as a visual representation of the relationship between multiple different classes of objects[8]. Venn diagrams have been used by logicians and set theorists ever since. One example of an early result relating to Venn diagrams is Venn's construction[9] which was used to show the existence of a simple n-Venn diagram for all $n > 1$. It goes like this:

1. Start with the basic 3-Venn in figure 1.

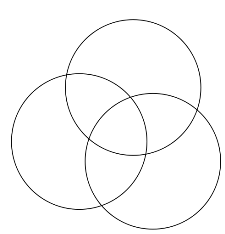

Figure 1: Starting 3-Venn[1]

2. Choose one curve and one vertex and then start encapsulating the curve starting just after the vertex till just before reaching the vertex. The red curve in figure 2 shows this.

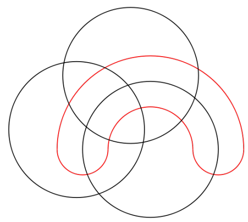

Figure 2: Resulting 4-Venn from Venn's construction[1]

3. Continue applying step two using the curve last added. Figure 3 and figure 4 shows the next two iterations.

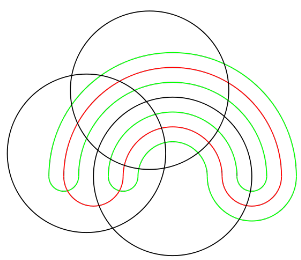

Figure 3: Resulting 5-Venn from Venn's construction[1]

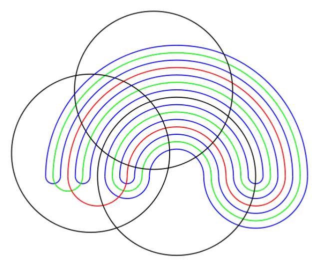

Figure 4: Resulting 6-Venn from Venn's construction[1]

The process of expansion in Venn's construction will be of great use in the proposed expansion algorithm. For an in-depth look at an overview on Venn diagrams, take a look at Frank Ruskey's survey[7].

For the history of Winkler's conjecture, we will start with Branko Grünbaum. Grünbaum defines Venn diagrams and independent

[1] `https://www.cs.mcgill.ca/~rwest/link-suggestion/wpcd_2008-09_augmented/wp/v/Venn_diagram.htm`

families, and examines some questions about Venn diagrams[3]. This sparks interest in formal Venn diagrams. Grünbaum and Peter Winkler worked together to find a simple Venn diagram using five triangles[5]. After, Winkler shows some results regarding Venn diagrams and their extensions which lead to the conjecture that every simple n-Venn diagram is extendible to a simple $(n + 1)$-Venn diagram[10]. Grünbaum took this conjecture and formed a similar conjecture by removing the simple condition in Winkler's conjecture[4]. Grünbaum's version of the conjecture was later proved by Kiran Chilakamarri, Peter Hamburger, and Raymond Pippert[1]. Chilakamarri, Hamburger, and Pippert also developed procedures for constructing Venn diagrams, showed simple Venn diagrams are 3-connected, and found all distinct Venn diagrams with three curves[2] and Hamburger and Pippert also answered several of Grünbaum's conjectures[6].

3 Conjectures, Theorems, and Algorithms

This section contains all conjectures, theorems, and algorithms that are used in this paper.

3.1 Conjectures

Conjecture 1 (Winkler's Conjecture)**.** *Every simple n-Venn diagram is extendible to a simple $(n+1)$-Venn diagram for $n > 1$.*

Conjecture 2. *The components after taking out the cycle found using Venn's construction, are partitioned into two sets each of which have exactly half the remaining vertices, and if a component of size k is in one of the partitions, then there is a component of size k in the other partition.*

3.2 Theorems

Theorem 1. *If the dual graph of a simple n-Venn diagram is Hamiltonian, than the simple n-Venn diagram is extendible to a simple $(n+1)$-Venn diagram.*

Proof. Suppose the dual graph G of a simple n-Venn diagram D is Hamiltonian. For any planar embedding of a graph, the edges of any cycle must form a simple closed curve since a planar embedding never has edges that cross another edge (simple) and a cycle forms a closed loop by definition (closed). Since the dual of a Venn diagram forms a planar embedding, the Hamilton cycle must form a simple closed curve. For each face in D, the Hamilton cycle would enter from one edge and leave on some other edge, partitioning the face into two regions. Since the Hamiltonian cycle is a simple closed curve, one of the two regions would be on the inside and the other region on the outside of the curve. This means that all the intersections of the n curves still exist and that all possible intersections of the n curves with the Hamiltonian cycle so D and the Hamiltonian cycle form a $(n+1)$-Venn diagram. This $(n+1)$-Venn diagram would be simple since the Hamiltonian cycle only crosses edges and therefore never intersects more than one curve at any time. □

Theorem 2. *The minimum size of the cycle in the dual graph of a simple n-Venn diagram found by copying Venn's construction is $4n - 4$ for $n > 1$.*

Proof. Take one curve C as the curve to follow. There are $n - 1$ curves left that can cross it. The least two curves can cross is twice because if they don't cross, then their intersection needed for being Venn is empty and if they cross once, they meet at a single point so their intersection is empty. To then minimize the number of crossings on C, every curve is made to cross exactly twice, giving $2(n - 1)$ crossing or equivalently, $2(n - 1)$ edges on C. For each of the $2(n - 1)$ edges on C, we can associate it with the two vertices in the dual corresponding to the two faces the edge lies on. No two of the edges on C can share the same face because if they were, the edge that splits them would lie on only one face. This is a problem since that face is on the inside and outside of the curve it is on which can't happen with simple closed curves. This shows that each of the $2(n - 1)$ edges on C add two vertices of the dual to the cycle with a total of $2 \cdot 2(n - 1) = 4n - 4$. □

This lower bound helps reveal how well this method is at finding a large cycle in the dual of a Venn diagram. For example, $n = 3$

gives a minimum size of 8 which is all vertices in the dual and for $n = 4$, it only misses 4 out of 16 vertices in the dual. This works well for small n but as n gets large, the amount missing gets quite large. For example, $n = 10$ guarantees only 36 out of 1024 vertices in the dual.

3.3 Algorithms

3.3.1 Face Walking

The face walking algorithm is a way of going through a planar graph to find all of its faces. The algorithm takes in a planar graph in adjacency list format with neighbours listed in clockwise order. It starts by taking an edge (u, v), giving it face 0, and moving to the edge (v, w) where w is the next vertex after u in the list of neighbours of v. This process continues till the next edge is back to (u, v). That finishes the first face. To continue, increment the face number and do this process over and over again starting with an edge that has not been assigned a face number yet. The algorithm finishes when the last edge has been given a face number.

3.3.2 Finding the Dual Graph

This algorithm is one way to find the dual graph of a planar graph in adjacency list format with neighbours listed in clockwise order. The idea is to use the face walking algorithm twice. The first face walk will give the vertex set and the face that each edge is used in.

On the second face walk when traversing an edge (u, v), add an edge between the face number of (u, v) and the face number of (v, u).

3.3.3 Walking a curve in a Venn diagram

This is identical to the face walking algorithm except when choosing the next edge (v, w) in the face, choose $(v, w + 1)$. That is, choose the vertex two away from u instead of one.

3.3.4 Breadth First Search

This algorithm is a very useful graph algorithm as it is very versatile. In this context it will be used to find all the connected components of a graph. The input is any graph. First, select a vertex to start with, mark it as used, add it to the component, and add it to a queue. Next, while the queue is not empty, dequeue a vertex v, add it to the component, and queue all unused vertices adjacent to v. When the queue is empty, check if there is an unused vertex, and if there is, make a new component and repeat the process done on the first vertex. This continues till all vertices have been used.

3.3.5 Naive Hamiltonian path

This is the simplest algorithm for finding a Hamiltonian path in a graph, but for small graphs with few edges it works fast enough. It is a basic recursive function that checks if the path so far covers all vertices. If not, it loops though each vertex u adjacent to the end of the path and calls itself with the path with u added to the end.

4 The Problem

4.1 Introduction

The problem that will be looked at is the conjecture mentioned earlier proposed by Peter Winkler. It states that every simple n-Venn diagram is extendible to a simple $(n+1)$-Venn diagram. Specifically, we will check if the conjecture holds for all simple 6-Venns generated by Ben Kinnett and Wendy Myrvold. This will be done by looking for Hamiltonian cycles in the dual graph of each simple 6-Venn. If we find that the dual graph of each simple 6-Venn is Hamiltonian, then using Theorem 1, it follows that each simple 6-Venn is extendible to a simple 7-Venn, so Winkler's Conjecture would hold for the 6-Venn case. Since determining if a graph is Hamiltonian is NP-Complete, to find a Hamilton cycle in all 3,430,404 simple 6-Venns, a way to reduce the computation will be needed. In this case, Venn's construction will be followed to find a main cycle in the dual graph guaranteed to have size at least $4n - 4$ by theorem 2. Then for the remaining vertices, it will be attempted to add them onto the main cycle till all vertices are used or there is no way to add onto the main cycle.

4.2 The Algorithm

This algorithm first of all requires an input of a Venn diagram as a planar embedding in adjacency list format with vertices in clockwise order. This is to make it easy to find all the faces and curves. The algorithm goes as follows:

1. Read in the Venn diagram.
2. Generate the Venn diagram's dual graph.
3. For each curve in the Venn diagram:
   (a) Find the main cycle in the dual by following the curve.
   (b) Use breadth first search to find each connected component of the dual minus the main cycle.
   (c) For each connected component, try to find a Hamiltonian path that starts and ends with vertices adjacent to two adjacent vertices in the main cycle.
   (d) If such a path is found, add it to the main cycle. If not, stop trying this curve and move to the next.
   (e) If all connected components get added, then the main cycle is a Hamilton cycle so return the cycle.
4. If no Hamilton cycle is found after trying each curve, then return a failure.

4.3 The Result

For these results, the algorithm was implemented using C++ and had its execution split between eight cores on an Intel i7 2600 processor. For the time of execution, the algorithm finished in 3564.68 seconds on the fastest core and 3597.15 seconds on the slowest core. The algorithm failed to find a Hamiltonian cycle in 356,520 out of the 3,430,404 simple 6-Venn diagrams which equates to a failure of 10.39%. For the three, four, and five Venn diagrams it found a Hamiltonian cycle every time in less than 0.002 seconds on just one core.

4.4 Where it Went Wrong

Running the algorithm on all three, four, and five Venn diagrams and finding a Hamiltonian cycle every time looks promising, but when restricted to only checking one curve, the problem cases start to show in the 5-Venns. The problem cases appear in two different forms. The first problem form is when a connected component does not have a Hamiltonian path. For example, this happens on one of the 5-Venns when following one of its curves. It has a connected component that looks like figure 5. The Hamiltonian path must start and end with vertices 6 and 30 but the 4-cycle in the centre prevents one of 21 and 22 from being reached. The other problem form is when a connected component has some Hamiltonian paths

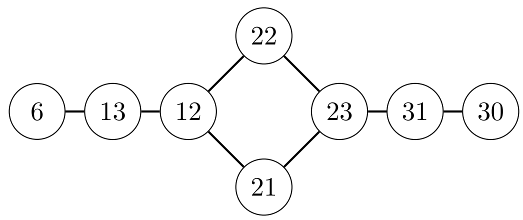

Figure 5: A problem case for finding a Hamilton path

but none of them can be attached to adjacent vertices in the main path. The main types of components that have this problem are the small components, especially size one components. This can be seen well in another of the 5-Venns. When following two of its curves, there becomes four components each of size one. This can be seen in the partial graph in figure 6 where vertex 7 is in its own component and is not adjacent to two adjacent vertices in the main cycle. This is reasonable since the vertex would need to form a 3-cycle with an edge of the main cycle and there are not many 3-cycles in the graph. Components each of size two on the other hand, tend to be easily incorporated because there tends to be many 4-cycles in the dual graph so when the component gets cut off, the main path almost always goes through the opposite side of one 4-cycle containing the component. Even though there were problem cases in the 5-Venns, at least one curve on each of them had no problem. This leads to what makes some of the 6-Venns fail on all curves. An example of a 6-Venn where this happens is given in figure 7. In this graph there is always at least four vertices in the dual that are split off

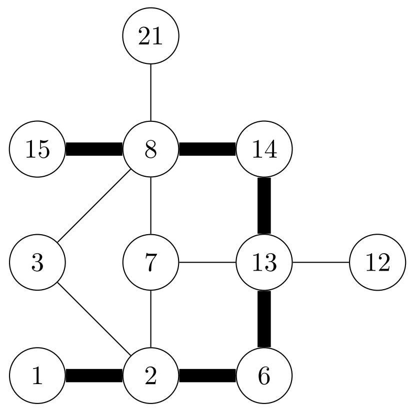

Figure 6: A problem case with isolated vertex 7. Bold edges correspond with the main cycle.

into their own component that can't be added. They all fail for the same reason illustrated in figure 6. It should be noted that every component that is size one that fails is degree three in dual. Degree three vertices have some forceful properties when choosing edges for Hamilton cycles. This may be why degree 3 vertices are the vertices that tend to cause problems.

4.5 Possible Solutions

Here are some of the possible ways to improve this algorithm. The first approach is to look into other results that guarantee the extendability of the Venn diagram. For example if a curve is traced and found that the resulting $n-1$ curves form a $(n-1)$-Venn diagram, then it is extendible so move on. This may only improve the run time because tracing that curve would likely find a Hamilton cycle with the algorithm. For the cases where the connected com-

0:1 2 3 4	21:11 20 35 22	42:28 52 53 43
1:0 4 5 6	22:12 21 35 23	43:29 42 44 30
2:0 6 7 3	23:12 22 36 13	44:33 43 53 54
3:0 2 8 9	24:13 36 37 25	45:34 54 55 56
4:0 9 10 1	25:13 24 38 26	46:34 57 58 35
5:1 11 12 6	26:14 25 38 39	47:37 58 59 38
6:1 5 7 2	27:14 40 41 28	48:39 59 60 61
7:2 6 13 14	28:15 27 42 29	49:39 61 50 40
8:3 14 15 16	29:15 28 43 30	50:41 49 61 51
9:3 17 18 4	30:16 29 43 31	51:41 50 55 52
10:4 18 19 20	31:16 30 32 17	52:42 51 55 53
11:5 20 21 12	32:17 31 33 19	53:42 52 54 44
12:5 11 22 23	33:19 32 44 34	54:44 53 55 45
13:7 23 24 25	34:20 33 45 46	55:45 54 52 51
14:7 26 27 8	35:21 46 36 22	56:45 61 60 57
15:8 28 29 16	36:23 35 37 24	57:46 56 60 58
16:8 15 30 31	37:24 36 47 38	58:46 57 59 47
17:9 31 32 18	38:25 37 47 26	59:47 58 60 48
18:9 17 19 10	39:26 48 49 40	60:48 59 57 56
19:10 18 32 33	40:27 39 49 41	61:48 56 50 49
20:10 34 21 11	41:27 40 50 51	

Figure 7: Adjacency list for the problem 6-Venn.

ponent has no Hamilton path, looking into the longest path in the connected component and replacing the section of the main cycle between the end points with that path may result in new connected components that can be added. Completing a similar process with the single vertex component may also be a solution. Whatever solution is found will at the very least make some modification on the main cycle, as that is what separates the dual graph into problem components. Another solution would be to follow the algorithm but with another process to find a main cycle to work off of.

5 Conclusion

5.1 Outcome

From what has been shown, a fast way to check the extendability of most simple Venn diagrams has been found, drastically reducing the number of Venn diagrams that need a slower algorithm to check. This will make computations that would normally take days to complete now only take hours. Also some progress has been made into the problems that come from following Venn's construction, so when trying to come up with a guaranteed algorithm or theoretical proof of Winkler's conjecture, some of the areas that need to be handled have been found.

5.2 Open Problems and Further Direction

From the progress of this paper there is still further work to do. The main open question is still a theoretical proof of Winkler's conjecture or a counter example. From what has been seen, it is likely that Winkler's conjecture is true. There are still some smaller questions to be answered. One is: what would be the guaranteed maximal size of Venn's construction cycle over all curves for arbitrary Venn diagram? Finding the minimum size of those cycles may help find some of the properties of the structure of the dual graph that splits the components up. Another question is conjecture 2. This was conjectured after seeing the component sizes when running the algorithm. It appears that this is caused by how the curves partition the regions and some symmetry between the inside and outside of the curve.

References

[1] Kiran B. Chilakamarri, Peter Hamburger, and Raymond E. Pippert. Hamilton cycles in planar graphs and Venn diagrams. *J. Combin. Theory Ser. B*, 67(2):296–303, 1996.

[2] Kiran B. Chilakamarri, Peter Hamburger, and Raymond E. Pippert. Venn diagrams and planar graphs. *Geom. Dedicata*, 62(1):73–91, 1996.

[3] Branko Grünbaum. Venn diagrams and independent families of sets. *Math. Mag.*, 48:12–23, 1975.

[4] Branko Grünbaum. Venn diagrams. I. *Geombinatorics*, 1(4):5–12, 1992.

[5] Branko Grünbaum and Peter Winkler. A simple Venn diagram of five triangles. *Math. Magazine*, 55(5):311, 1982.

[6] Peter Hamburger and Raymond E. Pippert. Simple, reducible Venn diagrams on five curves and Hamiltonian cycles. *Geom. Dedicata*, 68(3):245–262, 1997.

[7] Frank Ruskey. A survey of Venn diagrams. *Electron. J. Combin.*, 4(1):Dynamic Survey 5, 10 HTML documents (electronic), 1997.

[8] John Venn. On the diagrammatic and mechanical representation of propositions and reasonings. *London Edinburgh Dublin Philos. Mag. J. Sci.*, 9(1):1–18, 1880.

[9] John Venn. *Symbolic logic.* Chelsea Publishing Co., Bronx, N.Y., 1971. Reprint of the 1894 second edition, revised and rewritten.

[10] Peter Winkler. Venn diagrams: some observations and an open problem. In *Proceedings of the fifteenth Southeastern conference on combinatorics, graph theory and computing (Baton Rouge, La., 1984)*, volume 45, pages 267–274, 1984.