Sudoku strategies using graph theory

JEFF BROWN

Introduction

In this paper we discuss sudoku-solving strategies and how graph theory can be used to explain some of the advanced techniques. There are many websites that provide tutorials on solving sudoku puzzles. The sites [1] and [2] discuss the *xy*-chain technique, and the two explanations are quite different. We will define *xy*-chains as paths in a graph, and properties of the paths show why the technique works.

Sudoku puzzles have been used as teaching tools in a variety of disciplines, including chemistry [3], statistics [4], and mathematical proof techniques [5]. The material in this paper could be used to illustrate some fundamental notions in graph theory.

Previous applications of graph theory to sudoku have involved *k*-colouring of a graph whose vertices are all the puzzle cells [6]. Our approach is very different, involving relatively small graphs and techniques that can be used when solving a puzzle with pencil and paper.

Definitions

We consider traditional sudoku puzzles that consist of a 9 × 9 grid of cells where some cells have been assigned values in the range 1 to 9. The goal is to assign values to all the cells so that each of the nine values appear in every row, column, and 3 × 3 box.

		6		8		3		
	4	9		7		2	5	
			4		5			
6			3	1	7			4
		7				8		
1			8	2	6			9
			7		2			
	7	5		4		1	9	
		3		9		6		

FIGURE 1: Sudoku puzzle

Figure 1 shows a puzzle in which 32 cells have been assigned values. The nine 3 × 3 boxes are separated by bold lines.

We assume a puzzle has a *unique solution*. Figure 2 shows two partial puzzles. The puzzle on the left has no solution because there is no place for a 4 in the right-hand box. The puzzle on the right has multiple solutions because interchanging the 2 and 7 in columns 3 and 6 will yield a second solution.

https://doi.org/10.1017/mag.2024.68 Published online by Cambridge University Press

		4						
						1	2	3
					4			

5	6	2	4	1	7	3	9	8
9	3	1	8	5	6	7	2	4
8	4	7	9	3	2	5	6	1

FIGURE 2: No solution on left, multiple solutions on right

The rows, columns and boxes are called the *units* of the puzzle, and two cells are said to *intersect* if they are in the same unit. The ordered pair (x, y) denotes the cell in row x and column y.

The *candidates* of a cell are the puzzle values that can be assigned to that cell without creating a unit with a repeated value. For example, the cell (2, 6) in Figure 1 has candidates 1 and 3. All the other puzzle values have been assigned to cells that intersect cell (2, 6).

Basic techniques

Sudoku strategies all rely on rules that allow you to reduce the number of candidates. When the number of candidates for a cell is reduced to 1, then the value of that cell is known.

Naked sets

A *naked set* is a set of n locations in one unit such that the union of candidates for those locations has n values. The n candidates for the naked set must be the values of the n cells, so those values can be removed from the candidate lists of the other cells in the unit.

Figure 3 shows part of a puzzle where the three shaded cells form a naked set with candidates 1, 4, and 9. Remove those values from the candidate lists of the other cells in that row.

	2				
679	167	149	49	2	19
	5	8			

FIGURE 3: Naked triple

Hidden sets

A hidden set is a set of n puzzle values that are candidates in only n cells in a unit. The n values must appear in the n cells, so you can remove other candidates from those cells.

Figure 4 shows the top left box of a puzzle. Values 1 and 8 are candidates only in the shaded cells, so those cells must have values 1 and 8. Remove candidates 6 and 9 from the shaded cells.

https://doi.org/10.1017/mag.2024.68 Published online by Cambridge University Press

168	189	3
2	47	5
467	69	49

FIGURE 4: Hidden double 1,8

Locked candidate

Let I be the three cells of a box that intersect a row or a column. If a value v is a candidate in I, and it is not a candidate in any other cells in the box, then v must be the value of one of the cells in I, and you can remove v from the candidate lists of other cells in the row or column. If v is a candidate in I, and it is not a candidate in the other cells of the row or column, then again v must be the value of a cell in I, and you can remove v from the candidate lists of other cells in the box.

Figure 5 shows the intersection of a row with a box. The value 5 is a candidate in the intersection and it is not a candidate in the rest of the row. So, we may remove 5 from the candidate lists of the unshaded cells of the box.

			1	457				
1	679	8	56	567	3	4	2	679
			467	9	457			

FIGURE 5: Locked candidate 5

Graphs

A graph G consists of a set of vertices $V(G)$ and a set of edges $E(G)$, where each edge is associated with an unordered pair of vertices. We say that an edge *connects* the vertices associated with it, and two vertices connected by an edge are said to be *adjacent*. Graphs are often represented by drawing points for the vertices and line segments connecting the points for edges.

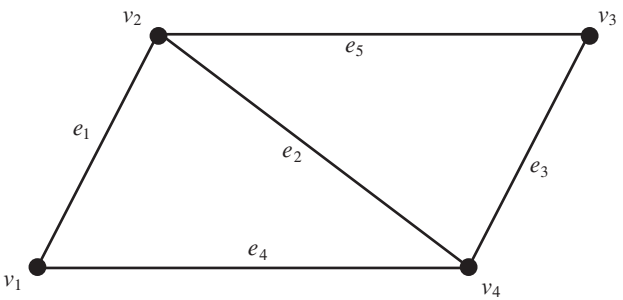

FIGURE 6: Simple graph

https://doi.org/10.1017/mag.2024.68 Published online by Cambridge University Press

Figure 6 shows a graph with four vertices and five edges. Vertex v_1 is adjacent to both v_2 and v_4.

All graphs discussed in this paper are *simple*, which means an edge cannot connect a vertex to itself, and any two vertices are connected by at most one edge. Let a and b be two distinct vertices. A path from a to b, called an *ab-path*, is a sequence of adjacent vertices $a = v_1v_2v_3\ldots v_n = b$ where no vertex appears more than once, and v_i is adjacent to v_{i+1} for $1 \leqslant i < n$. A *cycle* is like a path, except it begins and ends at the same vertex. The length of a cycle or path is the number of edges. A graph is *connected* if for every pair of vertices, a and b, there is an ab-path.

Bipartite graphs

A graph G is bipartite if there are two non-empty sets X and Y such that $V(G) = X \cup Y$, and every edge of G connects a vertex of X to a vertex of Y. X and Y are called the parts or partite sets of the graph, and $\{X, Y\}$ is called a *bipartition* of G.

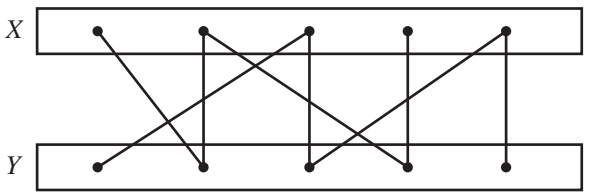

FIGURE 7: Bipartite graph

Figure 7 shows a bipartite graph with ten vertices. Note that this graph is not connected, it has two connected components.

Conjugate pair graphs

Let v be a puzzle value. If there are two cells in a unit that have v as a candidate, and v is not a candidate for any other cells of the unit, then the two cells form a *conjugate pair with respect to v*, and exactly one of the two cells will have the value v.

Definition: The conjugate pair graph with respect to v is the graph whose nodes are the cells that are in conjugate pairs, and two nodes are connected by an edge if they form a conjugate pair.

Figure 8 shows a conjugate pair graph with respect to 2. Note that this graph is not connected, it has two connected components, one of which contains just one conjugate pair (8,2) and (8,9).

https://doi.org/10.1017/mag.2024.68 Published online by Cambridge University Press

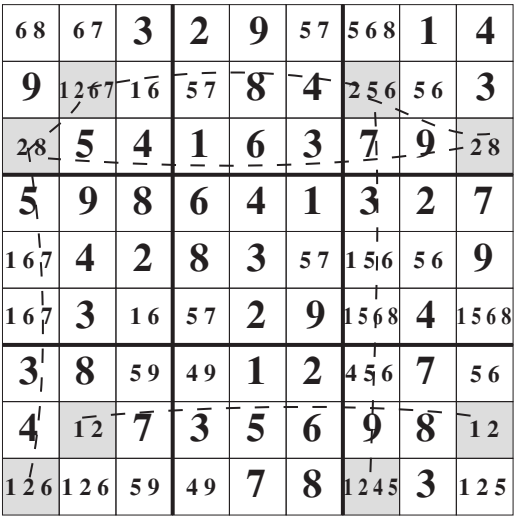

FIGURE 8: Conjugate pair graph with respect to 2

The next theorem is a well-known characterisation of bipartite graphs. See [7, p. 14] for a proof.

Theorem 1: A graph is bipartite if, and only if, it has no cycles of odd length.

Theorem 2: Let G be a conjugate pair graph with respect to v. Then G has a bipartition $\{X, Y\}$ such that either every cell in X has value v or every cell in Y has the value v.

Proof: We assume G is connected, as the following argument could be applied to connected components independently. Let $p = c_1c_2c_3\ldots c_n$ be a path in G with even length, so n is odd. Since adjacent cells in G form a conjugate pair, the cells c_i in path p alternate between having value v and not having value v. Since n is odd, then either both c_1 and c_n have value v, or neither of them do, and it follows that c_1 *and* c_n *are not adjacent in* G.

Suppose G has a cycle with odd length. Removing an edge from the odd cycle creates a path $p = c_1c_2c_3\ldots c_n$ where n is odd, so c_1 is not adjacent to c_n. However, the path p was created by removing an edge from a cycle, which means that c_1 is adjacent to c_n. This contradiction shows that G has no cycles of odd length, and by Theorem 1, G is bipartite. Let $\{X, Y\}$ be a bipartition of G.

Let a and b be two cells in X. Since G is connected, there is an ab-path, and since cells in a path alternate between being in X and being in Y, an ab-path has even length and hence an odd number of vertices. So, either both a and b have value v or neither do. Since this is true for any pair of cells in X,

https://doi.org/10.1017/mag.2024.68 Published online by Cambridge University Press

either all cells of X have value v or none do. The same is true for Y and we have proved Theorem 2.

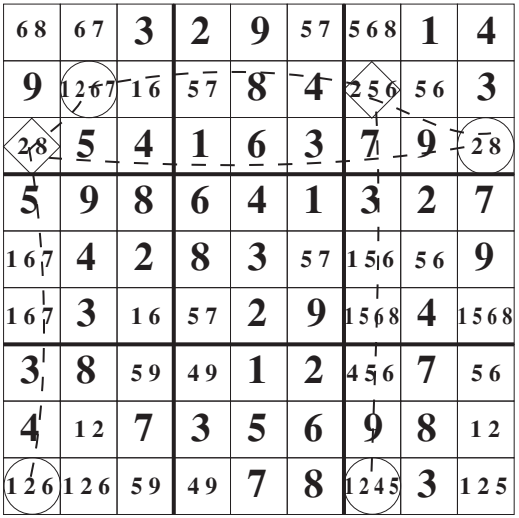

FIGURE 9: Bipartition

Figure 9 shows one connected component of a conjugate pair graph with respect to 2. One partite set is marked with circles and the other is marked with diamonds. Note that two of the cells with circles intersect, they are both on the last row. Therefore the cells with circles cannot have the value 2, and the cells with diamonds do have the value 2.

Two-candidate graph

In this section we consider another graph associated with a sudoku puzzle. The two-candidate graph has vertices that are cells with exactly two candidates, and cells are adjacent if they intersect.

Definition: Let $p = c_1c_2c_3\ldots c_n$ be a path in a two-candidate graph. The path p is a *linked path* if for $1 \leqslant i < n$ you can choose a candidate of c_i, called the link, so that the link of c_i is a candidate of c_{i+1}, and the link of c_i is not equal to the link of c_{i+1}. The candidate of c_n that is not the link of c_{n-1} is called the link of c_n.

https://doi.org/10.1017/mag.2024.68 Published online by Cambridge University Press

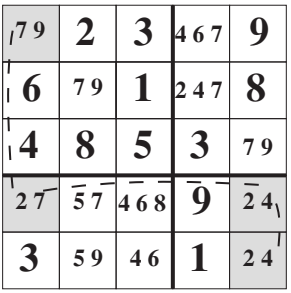

FIGURE 10: Linked path

Figure 10 shows a linked path from cell (1, 1) to cell (5, 5). The link of cell (1, 1) must be 7 because that is the only candidate that (1, 1) and (4, 1) have in common. Similarly, the link of (4, 1) must be 2. The last two cells in the path have two common candidates, but the link of (4, 5) must be 4 for the path to be linked. The link of (5, 5) is 2.

Cell	c_1	c_2	c_3	…	c_{n-2}	c_{n-1}	c_n
Link	a	c	d	…	g	h	i
Other	b	a	c	…	f	g	h

TABLE 1: Linked path from c_1 to c_n

Table 1 shows a view of a linked path from c_1 to c_n. The second row contains the links of the cells, and the third row shows the candidates that are not the links. The reverse of the linked path from c_n to c_1 is also a linked path, but all the links are switched, so the link of c_n will be h and the link of c_{n-1} will be g, and so on.

The reason linked paths are useful is that if the value of the first cell is its link, then the same is true for every cell in the path. For example, in the linked path of Figure 10, if the value of cell (1, 1) is 7, then the value of (4, 1) is 2, the value of (4, 5) is 4 and the value of (5, 5) is 2.

XY-chains

Definition: An *xy*-chain is a linked path from c_1 to c_n in which the link of c_n is the candidate of c_1 that is not the link. In Table 1, change the i under c_n to b and you have an *xy*-chain.

Cell	c_1	c_2	c_3	…	c_{n-2}	c_{n-1}	c_n
Link	a	c	d	…	g	h	b
Other	b	a	c	…	f	g	h

TABLE 2: Chain from c_1 to c_n

Table 2 shows an *xy*-chain from c_1 to c_n. Either c_1 or c_n will have the value b.

https://doi.org/10.1017/mag.2024.68 Published online by Cambridge University Press

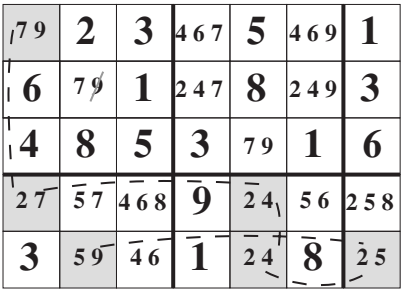

FIGURE 11: *XY*-Chain

Figure 11 shows an *xy*-chain where either the first or the last cell will have the value 9. We can eliminate 9 from the candidate list of cell (2, 2) because it intersects the first and last cells of the *xy*-chain.

References

1. Sudoku solver, XY-Chain, accessed November 2023 at: https://sudokusolver.app/xychain.html
2. Andrew Stuart, XY-Chains accessed November 2023 at: https://www.sudokuwiki.org/XY_Chains
3. T. D. Crute and S. A. Myers, Sudoku puzzles as chemistry learning tools, *Journal of Chemical Education* **84** (April 2007) pp. 612-613.
4. C. Brophy and L. Hahn, Engaging students in a large lecture: an experiment using sudoku puzzles, *Journal of Statistics Education*, **22**(1) (2014).
5. B. A. Snyder, Using sudoku to introduce proof techniques, *PRIMUS* **20**(5) (2010) pp. 383-391.
6. A. M. Herzberg and M. R. Murty, Sudoku squares and chromatic polynomials. *Notices of the AMS* **54**(6) (2007) pp. 708-717.
7. J. A. Bondy and U. S. R. Murty, *Graph theory with applications*, Macmillan (1976).

10.1017/mag.2024.68 © The Authors, 2024
Published by Cambridge University Press
on behalf of The Mathematical Association

JEFF BROWN
UNCW Mathematics and Statistics Department, 601 South College Road, Wilmington, NC 28403 USA
e-mail: *brownj@uncw.edu*