AMSI VACATION RESEARCH SCHOLARSHIPS 2017-2018

Colouring intersection graphs of complete geometric graphs

Tim Banova

Supervised by Michael Payne

Monash University

Vacation Research Scholarships are funded jointly by the Department of Education and Training and the Australian Mathematical Sciences Institute.

Australian Government
Department of Education and Training

AMSI AUSTRALIAN MATHEMATICAL SCIENCES INSTITUTE

1 Introduction

Geometric graph theory is the study of graphs arising from configurations of geometric objects, such as points and lines in $\mathbb{R}^2$. A graph is an abstract network, consisting of specified objects and a relation indicating whether two objects are 'adjacent' to one another; they are valuable tools for studying the rich combinatorial structures that can arise from such configurations. Informally, we state the problem we studied: suppose one is given a set of n points in the plane such that no three points lie on a line. Draw every possible line segment between these points and colour these line segments so that any pair of segments which cross or are incident are coloured with distinct colours. For a given n, what is the largest number of colours required to colour the segments of an n-point set? Let $i(n)$ denote this number. The best current bounds are $\frac{4}{3}n \leq i(n) \leq Cn^{\frac{3}{2}}$ for some $C > 0$, by Langerman (private communication) and Araujo et al. (2005), respectively. We show that $\frac{3}{2}n - o(n) \leq i(n)$. As a secondary goal, we have also computed either exact of very tight bounds on $i(n)$ up to $n = 9$.

2 Definitions

A *graph* $G = (V, E)$ is a pair, where $V = V(G)$ denotes the set of *vertices* (or *nodes*) and $E = E(G) \subseteq [V]^2$ denotes the set of *edges* (or *links*). For a given graph G, a *proper k-colouring* is a function $f : V(G) \to \{1, 2, \ldots, k\}$ such that $f(u) \neq f(v)$ whenever $uv \in E(G)$; if such a k-colouring exists, we say that G is *k-colourable*. The *chromatic number* of a graph G, denoted $\chi(G)$, is the minimum integer such that G is k-colourable. Given a k-colouring of G, the *colour classes* $\{V_1, \ldots, V_k\}$ of a graph G are the subsets of $V(G)$ of pairwise non-adjacent vertices receiving the same colour. Lastly, a *matching* M is a subset of pairwise non-incident edges in G. More information regarding graph theoretic notions can be found in Diestel (2016).

Provided $g(x)$ is non-zero past a certain point, we say that $f(x) = o(g(x))$ ($f(x)$ is little-oh of $g(x)$) if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$, which is to say that $f(x)$ is much smaller than $g(x)$ as $x \to \infty$, denoted $f(x) << g(x)$. We also say that for $n \in \mathbb{N}$, $f(n) = O(g(n))$ ($f(n)$ is big-oh of $g(n)$) if there exists a constant $C > 0$ and $n_0 \in \mathbb{N}$ such that for any $n \geq n_0$, $f(n) \leq Cg(n)$.

A set of points $P \subseteq \mathbb{R}^2$ is in *general position* if no three points in P are *collinear*. All point sets are assumed to be finite. Given a set of points $P \subseteq \mathbb{R}^2$ in general position, a *geometric graph* is the graph with vertex set P and an edge set consisting of line segments between points in P; in general any subset of line segments may be included, but for our purposes, we will be interested in complete geometric graphs on n points, in which every possible line segment between any two points is considered an edge. For convenience, we denote a complete geometric graph by its underlying point set P.

For point set P and a pair of points $p, q \in P$, we define the *line segment* induced by p and q as $\overline{pq} = \{tp + (1-t)q : t \in [0,1]\}$. $S \subseteq \mathbb{R}^2$ is said to be *convex* if for any $p, q \in S$, $\overline{pq} \subseteq S$, which states that S contains every line segment between any two points. The *convex hull*, $\text{conv}(P)$, of P is the minimal convex set containing P; intuitively, we can view the boundary of a convex hull of a point set as a 'rubber band' stretched around the outermost points of P. We say that a set of points P is in *convex position* if the points of P are vertices of $\text{conv}(P)$.

Given a geometric graph P, we define the *intersection graph* of P to be the graph, $I(P)$, with all k-sets of points in P as vertices and two sets $X, Y \subseteq P$ are adjacent if and only if their convex hulls intersect; we limit ourselves to the $k = 2$ case and let $I(P) = I_2(P)$. More intuitively, for $k = 2$, we are defining a graph from a point set in which the vertices are the edges of the induced geometric graph with edges present whenever two edges are incident or cross at an interior point. With this in mind, we say that the geometric graph P is *k-intersection-colourable* if its corresponding intersection graph $I(P)$ is k-colourable. A k-intersection-colouring of P corresponds to a drawing in the plane of P in which edges that are incident or cross one another receive distinct colours (see Figure 1). Whenever 'colouring' is referred to in this report, we shall assume that we are colouring in this way, as opposed to the standard definition.

With the above observation in mind, we note that the colour classes of $I(P)$ correspond to non-crossing, non-incident matchings in P, so colouring the vertices of $I(P)$ amounts to decomposing the edges of P into matchings of this kind, which we refer to as *disjoint*.

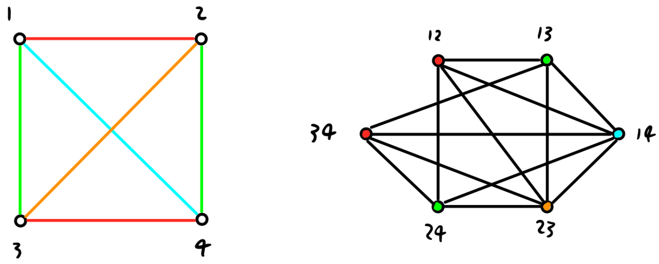

Figure 1: A point set P with a 4-intersection-colouring and its corresponding intersection graph $I(P)$

3 Problem and progress

We shall now formally state the problem. What is the maximum chromatic number of $I(P)$ over all sets P of n points in general position?

Let $i(n) := \max\{\chi(I(P)) : P \subseteq \mathbb{R}^2 \text{ is in general position, } |P| = n\}$ denote this number. Our purpose is to improve the current bounds of this function as n get arbitrarily large.

It has been shown that $i_c(n) = n$ and $i(n) = O(n^{\frac{3}{2}})$ in Araujo et al. (2005), where $i_c(n)$ is defined the same as $i(n)$, but limited to convex position point sets. As a consequence, $n \leq i(n)$. Our primary goal is to improve this lower bound. To show that $kn < i(n)$ for some constant $k > 0$, one needs to show that for any $n \in \mathbb{N}$ there is an n-point set P_n such that $I(P_n)$ cannot be coloured with kn colours. Doing so would imply that for any $n \in \mathbb{N}$, $\chi(I(P_n)) > kn$, from which we have that $kn < i(n)$, since $\chi(I(P_n)) < i(n)$. We considered this an achievable task, as finding upper bounds require very general arguments which were deemed too difficult to improve given the time frame. Our starting point was provided through private communication with Stefan Langerman, who provided the $\frac{4}{3}n$ lower bound with his "one-point-in-the-middle" example; we applied the same methods to a different example improving the lower bound.

Theorem 3.1. *$i(n) \geq \frac{3}{2}n - o(n)$*

The family of point sets we consider, $\{O_n\}_{n\in\mathbb{N}}$, consists of n points with $\lceil \log n \rceil$ points equally spaced around an inner circle and an odd number of $n - \log n$ points on an outer circle with the inner circle small enough so as to be enclosed within a cell induced by line segments

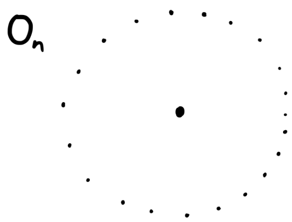

Figure 2: The point set, O_n

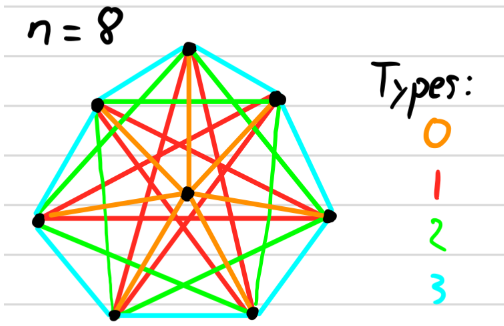

Figure 3: For $n = 8$, O_n is partitioned by type 3 edges

of a specific type, which we specify in the proof. We refer to vertices on the inner and outer circles as *inner* and *outer vertices*, respectively.

4 Proof of Theorem 3.1

Fix $n \in \mathbb{N}$ and consider the point set O_n, on which a complete geometric graph is induced. We want to show that $\chi(I(O_n)) \geq \frac{3}{2}n - o(n)$. We attribute this method to Stefan Langerman, which we suitably generalized for this example from a privately communicated "one point in the middle" example.

We say that an edge is of *type* 0 if it is connects the inner and outer circles. If $d \geq 1$ and $p, q \in V(O_n)$ are points on the outer circle, $pq \in E(O_n)$ is of *type* d if there are $\frac{n - \lceil \log n \rceil - 1}{2} - d$ points on the outer circle between p and q (see Figure 3). We define the *inner cell* of O_n to be the convex hull of the set of points induced by crossings for the type 1 edges of O_n. For simplicity, we shall refer to a type d edge as a *d-edge*.

Consider the edges of O_n. Clearly, O_n can have edges up to type $\frac{n-\lceil \log n \rceil -1}{2}$, and all edges must be assigned one type. We can also quickly verify that there are $n - \lceil \log n \rceil$ edges of type $d \geq 1$ and $\lceil \log n \rceil (n - \lceil \log n \rceil)$ edges of type 0.

Next, consider a maximal disjoint matching M of O_n; we classify the possible configurations of edges up to type 3 in M.

In the case that M does not contain a 0-edge, M must contain a 1-edge, by maximality. In this case, M can only contain single 2 and 3-edges, as any other configuration will contain an incidence or a crossing; this gives us configuration 7.

Now, suppose that M contains 0-edges. If M contains the maximum $\lceil \log n \rceil$ 0-edges, M may either contain one 2-edge or a 1-edge and a 3-edge; these cases give us configurations 5 and 6.

If M contains a single 0-edge, M may or may not contain a 1-edge. In the case that M contains a 1-edge, M must contain two 3-edges, as a 2-edge will cross a 1-edge when a 0-edge is in M; this gives us configuration 1. In the case that M contains no 1-edge, M must contain two 2-edges by maximality; this gives us configuration 2.

If M contains two 0-edges, a 1-edge cannot be present, as the presence of a 2-edge or a 3-edge induces a crossing with the 1-edge and including the 2 and 3-edges provides a larger matching; this gives us configuration 3.

If M contains three 0-edges, M may contain a 2-edge, a 1-edge and a 3-edge or two 3-edges. The first two cases are described by configurations 5 and 6, with the last case inducing configuration 4.

Lastly, if M contains more than three 0-edges, M can only contain one 2-edge or a 1-edge and a 3-edge, which is encapsulated by configurations 5 and 6, completing our classification.

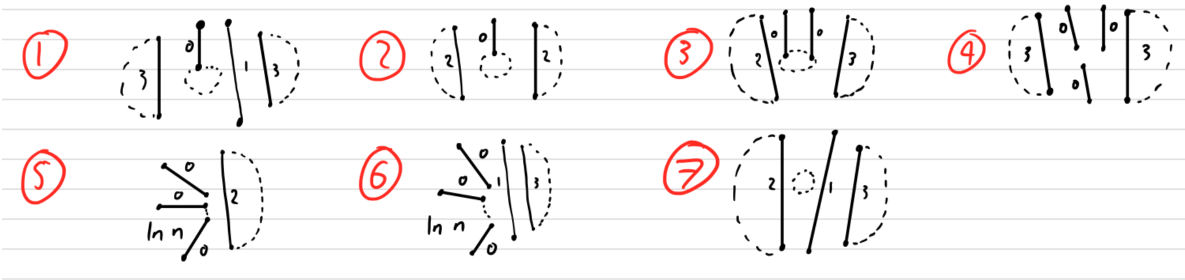

Figure 4: The possible configurations of edges in a maximal non-crossing matching of O_n up to type 3

For each of the 7 configurations, let c_i denote the number of colours using configuration i. Since O_n has $(\lceil \log n \rceil)(n - \lceil \log n \rceil)$ type 0 edges and $n - \lceil \log n \rceil$ edges of all other types, for each type the following constraints hold:

$$(0)\ c_1 + c_2 + 2c_3 + 3c_4 + \lceil \log n \rceil c_5 + \lceil \log n \rceil c_6 \geq \lceil \log n \rceil (n - \lceil \log n \rceil)$$

$$(1)\ c_1 + c_6 + c_7 \geq n - \lceil \log n \rceil$$

$$(2)\ 2c_2 + c_3 + c_5 + c_7 \geq n - \lceil \log n \rceil$$

$$(3)\ 2c_1 + c_3 + 2c_4 + c_6 + c_7 \geq n - \lceil \log n \rceil$$

Dividing both sides of equation (0) by $\log n$ and taking n to be large, we have the equivalent equation

$$(0')\ c_5 + c_6 \geq n - \lceil \log n \rceil$$

Scaling the coefficient vector by $\frac{1}{n - \lceil \log n \rceil}$ which affects the solution by a constant factor, we have the following constraints

$$(0'')\ c_5 + c_6 \geq 1$$

$$(1')\ c_1 + c_6 + c_7 \geq 1$$

$$(2')\ 2c_2 + c_3 + c_5 + c_7 \geq 1$$

$$(3')\ 2c_1 + c_3 + 2c_4 + c_6 + c_7 \geq 1$$

Minimizing the sum $\sum_{i=1}^{7} c_i$ (using lpsolve) subject to these constraints garners the solution.

$$c_1 = c_3 = c_4 = c_5 = c_7 = 0,$$
$$c_2 = \frac{1}{2},\ c_6 = 1$$

This implies that at least $\frac{3}{2}(n - \lceil \log n \rceil) = \frac{3}{2}n - o(n)$ colours are required to colour the edges of O_n. Thus, $\chi(I(O_n)) \geq \frac{3}{2}n - o(n)$. This completes the proof. $\square$

While we have shown that the lower bound holds, it is unclear at this stage as to whether such a colouring is achievable; we informally discuss the tightness of this lower bound. To verify that this bound is asymptotically tight for the family $\{O_n\}_{n\in\mathbb{N}}$, we show that O_n can have its edges coloured with $\frac{3}{2}n + o(n)$ colours.

We first make the following geometric observation which is quick to prove by a line-sweeping argument; for the $\lceil \log n \rceil$ points on the inner circle, there are $\lceil \log n \rceil$ disjoint segments to $\lceil \log n \rceil$ adjacent points on the outer circle.

Thus, the 0-edges are partitioned into $n - \lceil \log n \rceil$ classes of disjoint line segments, where there are no points on the outer circle nor inner circle between these lines; let $\{L_i\}_{1\leq i\leq n-\lceil \log n \rceil}$ denote these classes.

For $1 \leq i \leq n - \lceil \log n \rceil$, consider the 0-edge class L_i. We show that there is a disjoint matching of odd edges M_i containing every one of odd type up to $\frac{n-\lceil \log n \rceil - 1}{2}$ disjoint from L_i.

Fix an outermost edge $l_0 \in L_i$. So, one the endpoints of l_0 is an inner vertex, which is enclosed in the cell induced by 1-edges. If l_0 crossed every 1-edge enclosing the inner vertices, both its endpoints would not be contained in the inner cell and hence neither would the inner vertices, which contradicts our choice of l_0. Thus, there exists a 1-edge, l_1, which does not intersect l_0.

Starting from l_1, we sweep a line out along the odd edges to the outer circle to obtain a parallel class O_i of odd edges, none of which will intersect L_i; colour L_i and O_i with a new colour.

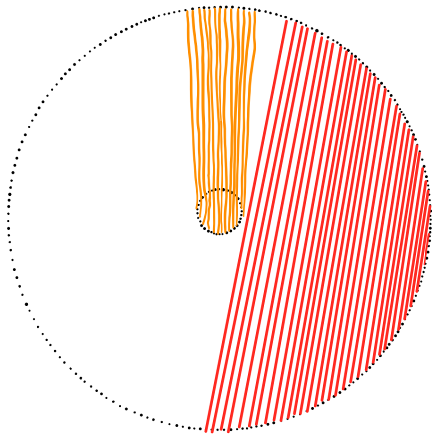

Figure 5: An artist's impression: L_i and O_i are indicated by orange and red, respectively

To colour the even edges, we can partition them into disjoint matchings with 2 of each type, thus requiring $\frac{n}{2}$ matchings to cover every edge, and hence $\frac{n}{2}$ new colours.

Lastly, we observe that there are 0-edge sets L_i containing crossing edges; to classify these edges, we make the following definition. Order the inner vertices in clockwise-order with labelling $\{p_1, p_2, \ldots, p_{\lceil \log n \rceil}\}$. A set of $\lceil \log n \rceil$ consecutive points $\{q_1, q_2, \ldots, q_{\lceil \log n \rceil}\}$ in clockwise order on the outer circle is a *good arc* if $\overline{p_1 q_1} \cap \overline{p_2 q_2} \cap \cdots \cap \overline{p_{\lceil \log n \rceil} q_{\lceil \log n \rceil}} = \emptyset$, which is to say that no crossings are induced; such a set is a *bad arc* if it is not a good arc (see Figure 5).

When identifying colourable 0-matchings, we take such a matching and shift it by one vertex on the outer circle; this approach classifies every 0-edge from each inner vertex. Consequently, this fixes the order in which the outer vertices are visible from inner vertices, inducing a form of compatibility between the good arcs and non-compatibility amongst bad arcs.

Good arcs are compatible in the sense that all vertices on a good arc are visible from the inner vertices in the specified order. Hence, as the matching shifts between good arcs around the circle, no crossings are induced. Bad arcs are incompatible amongst themselves and with respect to good arcs in the sense that this same shifting operation induces a variable number of crossings as the matching is shifted; in the worst case there are $\lceil \log n \rceil$ edges crossing, in which every edge crosses. To count the number of bad arcs, the inner vertices determine $\binom{\lceil \log n \rceil}{2} = O(\lceil \log n \rceil^2)$

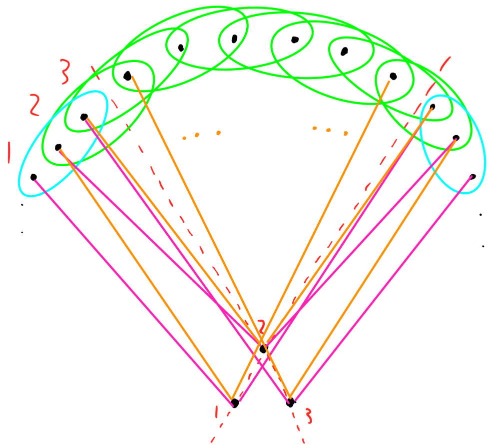

Figure 6: Blue sets indicate bad arcs, where pink edges indicate crossings

lines and each line determines $\lceil \log n \rceil$ bad arcs. Hence, there are $O(\lceil \log n \rceil^4)$ potential crossings amongst the 0-edges, which we colour with $O(\lceil \log n \rceil^4)$ new colours.

Having partitioned the edges in such a way we describe the following colouring algorithm. Firstly, for each $1 \leq i \leq n - \lceil \log n \rceil$, colour the non-crossing edges of L_i and all edges of O_i with the same colour. Secondly, colour the even with $\frac{n}{2}$ colours. Thirdly, colour the crossing edges of L_i all with new colours. Lastly, colour the edges amongst the inner vertices with all new colours.

The first step uses $n - \lceil \log n \rceil = n - o(n)$ colours, which corresponds to the number of $\lceil \log n \rceil$ adjacent points around the inner circle; this step colours every odd edge. The second step uses $\frac{n}{2}$ colours and colours every even edge. The third step colours all the remaining crossing 0-edges with new colours, of which there are $O(\lceil \log n \rceil^4) = o(n)$.

The last step colours the remaining edges amongst the inner circle, requiring $\lceil \log n \rceil = o(n)$ colours as they are in convex position.

Thus, $(n - o(n)) + \frac{n}{2} + o(n) + o(n) = \frac{3}{2}n + o(n)$ colours are sufficient to colour the edges of O_n, implying tightness of the lower bound for $\chi(I(O_n))$.

Table 1: Values of $i(n)$ computed up to $n = 9$

n	4	5	6	7	8	9
$i(n)$	4	5	6	7	≤ 9	≤ 11

5 Computational Work

A secondary objective was to compute the chromatic numbers of "all" point sets up to $n = 10$. While for any $n \in \mathbb{N}$ there are infinitely many n-point sets, we can group point sets into equivalence classes, of which there are finitely many, by the order type relation. The *order type* of a point set $\{p_1, p_2, \ldots, p_n\}$ in general position is a mapping that assigns to each ordered triple (i, j, k) in $\{1, \ldots, n\}$ the orientation (clockwise or counterclockwise) of the point triple (p_i, p_j, p_k). Point sets with the same order type encode the same crossing information, so the chromatic numbers of intersection graphs of order-type-equivalent points sets are the same.

In 2002, Aichholzer et al. (2002) created a point-set database containing every order-type equivalence class representative up to 11 points. Due to space and time constraints, we decided to compute $i(n)$ only up to $n = 9$. A motivation for this goal was to gain some intuition for constructing point set examples which were difficult to colour. It seems intuitive to take most points in convex position except for a few in the center, which gives many crossings and no obvious way of decomposing the edges sets into large colour classes; this intuition influenced Langerman's initial example, and our subsequent improvement. We wanted to confirm whether this intuition was justifiable by checking small examples unable to be coloured with a given number of colours; we hoped to find more pathological examples in Aichholzer's database, which would lend us insights into determining the classes of the worst-colourable point sets.

Table 1 lists the values of $i(n)$ computed or bounded. For $n = 8, 9$, these are not equalities, as there were examples in the database which could not be colored with the given number of colours in a reasonable time-frame, but were not demonstratively non-colourable; Having plotted some of these examples (as in Figure 6), they appear to confirm our intuition as they were mostly convex except for some points in the middle. Whether there is some more in-

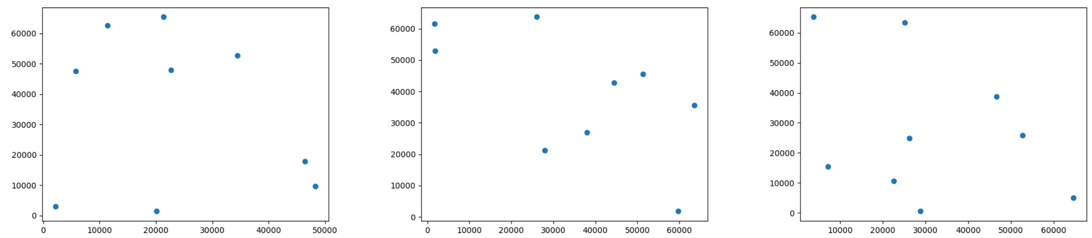

Figure 7: 9-point sets which could not be coloured with 10 colours in 60 seconds

variant structure in these example is unclear at this stage; supplementary analysis is required. We implemented a colouring program which, given a database of points and a number colours, would return a list of point sets which either could not be coloured with the specified number of colours or was unable to be coloured in a specified time frame. The main computational task was to define a procedure for translating a given point set into the equivalent SAT instance, encoding the edge-colouring problem for the corresponding complete geometric graph, which would then be checked using a SAT solver; Minisat (Een & Sorensson (2018)) was chosen for its Python interface, Satispy (Laszlo (2018)), and its options to time-out intractable instances. Figure 5 displays some 9-point sets which could not be coloured with 10 colours within 60 seconds.

6 Further Directions

We do not believe the current bounds on $i(n)$ are tight; to further improve the lower bound, a more careful analysis of intractable examples from our computational efforts seems fruitful. A possible direction involves the use of machine learning or pattern recognition algorithms, so that we might better characterize the structure of such point sets, using the examples found up to $n = 9$. Handling the $n = 10$ case is another direction, which would require the use of cluster computing due to space and time constraints. As for upper bounds, more powerful results from extremal graph theory are required. Bounding the chromatic number requires us to bound the number of independent sets, which in our case is the number of non-crossing matchings. A result of Toth (2000) tells us that geometric graphs with no $k+1$ pairwise disjoint (non-crossing and non-incident) edges contains at most 2^9k^2n edges, and using this result Araujo et al. (2005)

bounded the chromatic number by $O(n^{\frac{3}{2}})$. In this direction, the upper bound may be improved by improving the 2^9k^2n bound.

References

Aichholzer, O., Aurenhammer, F. & Krasser, H. (2002), 'Enumerating order types for small point sets with applications', *Order* **19**, 265–281.

Araujo, G., Dumitrescu, A., Hurtado, F., Noy, M. & Urrutia, J. (2005), 'On the chromatic number of some geometric type kneser graphs', *Computational Geometry: Theory and Applications* **32**, 59–69.

Diestel, R. (2016), *Graph Theory*, 5 edn, Springer-Verlag, Heidelberg.

Een, N. & Sorensson, N. (2018), 'Minisat'.
URL: *http://minisat.se/Main.html*

Laszlo, F. (2018), 'Satispy'.
URL: *https://github.com/netom/satispy*

Toth, G. (2000), 'Note on geometric graphs', *Journal of Combinatorial Theory, Series A* **89**, 126–132.