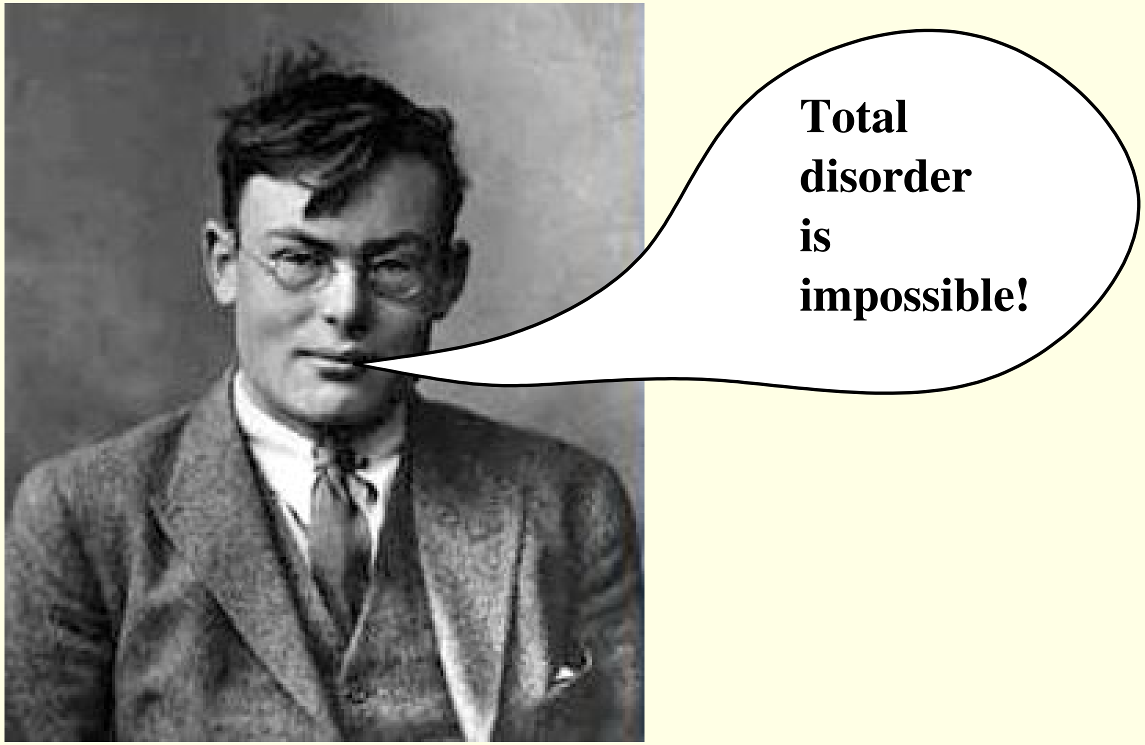
Graph Generation and Ramsey Numbers

Brendan McKay

Vigleik Angeltveit

Australian National University

Ramsey Theory

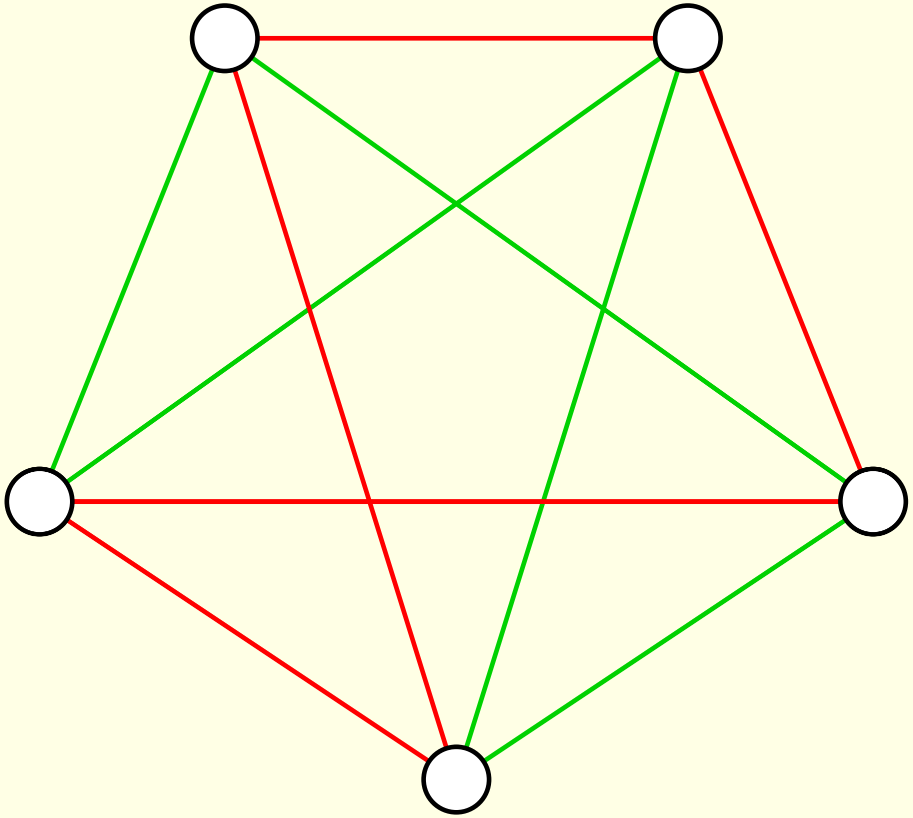


Frank Plumpton Ramsey (1903–1930)

Ramsey was an English logician who proved an minor lemma that eventually became the foundation of a large field of mathematics.

We are interested in applications in graph theory.

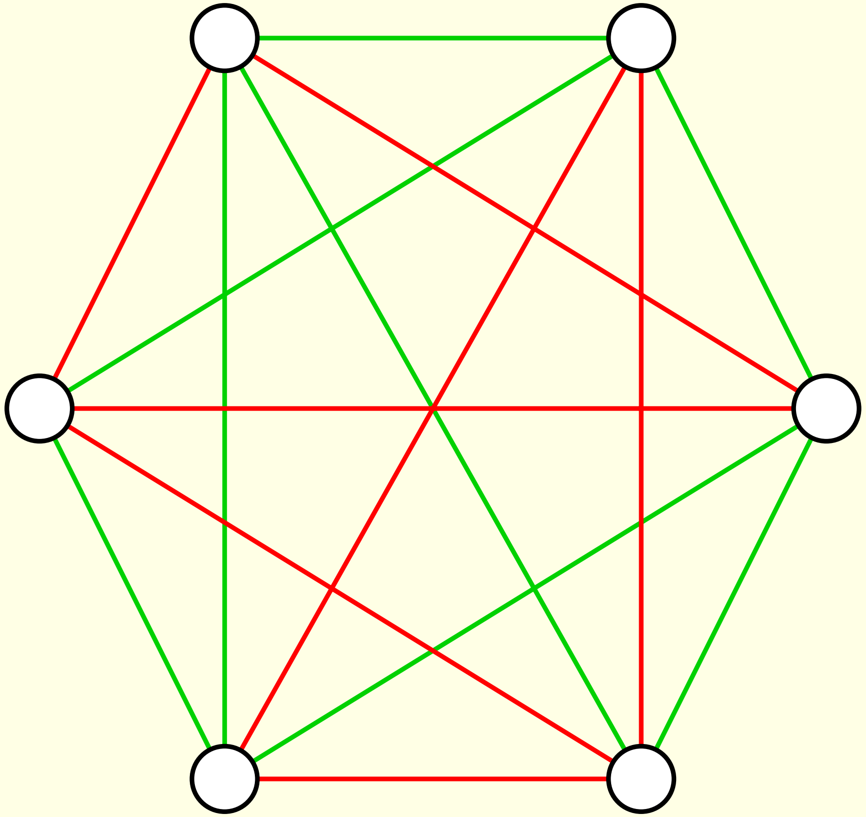
Ramsey Theory (continued)



It is possible to colour the edges of K_5 with two colours so that there is no monochromatic triangle.

Ramsey Theory (continued)

But with K_6, this is impossible.



We express this by writing $R(3, 3) = 6$.

Ramsey Theory (continued)

We can generalize the statement $R(3,3) = 6$ in multiple ways, but we will be interested in monochromatic cliques. A k-clique is a subgraph which is a complete graph with k vertices; e.g. a triangle is a 3-clique.

For integers s, t, an (s,t)-graph is a colouring of the edges of a complete graph with two colours, such that there is no s-clique with the first colour or t-clique with the second colour.

The Ramsey number $R(s,t)$ is the smallest $n > 0$ such there is no (s,t)-graph with n vertices.

It is easy to prove by induction that $R(s,t)$ exists for all s, t. That is, there are only a finite number of (s,t)-graphs for each s, t.

Example: (3,7)-graphs

n	$(3,7)$-graphs
1	1
2	2
3	3
4	7
5	14
6	38
7	106
8	402
9	1821
10	11217
11	88606
12	885319
13	10029294
14	113515631
15	1078421860
16	6856835524
17	21846684631
18	25078964384
19	7111066345
20	315293146
21	1118436
22	191

These counts are for isomorphism classes, where two graphs are isomorphic if they differ only in the names of the vertices.

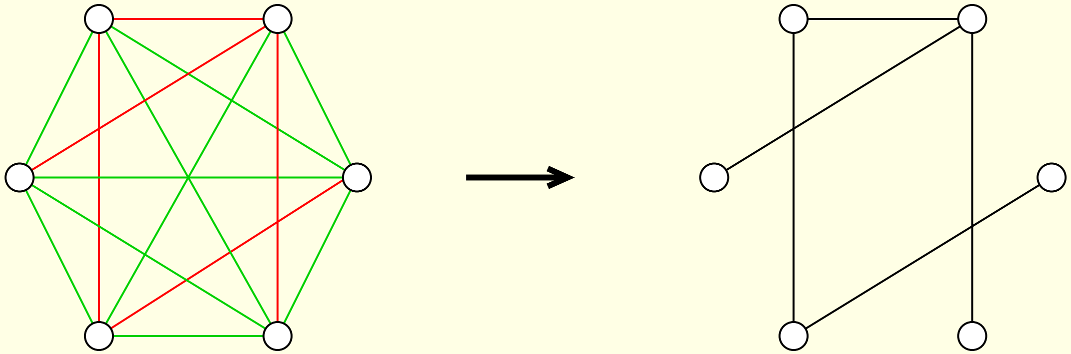
There are quite a lot of graphs here, which raises the question of how we found them all.

Small classes of Ramsey graphs

Recall:

For integers s, t, an (s,t)-graph is a colouring of the edges of a complete graph with two colours, such that there is no s-clique with the first colour or t-clique with the second colour.

Usually we just draw the edges of the first colour:



Then the definition reads:

For integers s, t, an (s,t)-graph is a graph with no s-clique and no independent t-set. (An independent set is a set of vertices with no edge between any two of them.)

Let's make all $(3,7)$-graphs

A $(3,7)$-graph has no triangle and no independent 7-set.

Since deleting a vertex from an $(3,7)$-graph leaves an $(3,7)$-graph, we can make all $(3,7)$-graphs by starting with K_1 and repeatedly adding one vertex in such a way that we don't create any triangles or independent 7-sets.

We can we do this efficiently?

Given a $(3,7)$-graph G, to extend it by one vertex v, we need to find all subsets $W \subseteq V(G)$ such that $G[W]$ does not contain an edge and $G[V(G) \setminus W]$ does not contain an independent set of size 6.

Example: Ramsey (3,7)-graphs (continued)

Given a $(3,7)$-graph G, to extend it by one vertex v, we need to find all subsets $W \subseteq V(G)$ such that $G[W]$ does not contain an edge and $G[V(G) \setminus W]$ does not contain an independent 6-set.

We use an old idea of Staszek Radziszowski. For $\emptyset \subseteq X \subseteq Y \subseteq V(G)$, define the interval $[X, Y] = \{W \subseteq V(G) \mid X \subseteq W \subseteq Y\}$.

We start with one interval $[\emptyset, V(G)]$, then for each edge or independent 6-set of G we break all our current intervals into smaller disjoint intervals so that no interval forms a triangle or independent 7-set with the edge or 6-set.

When this is complete, we have a set of disjoint intervals whose union is the set of all solutions.

This can be made very efficient using bit-vectors to implement sets.

The search for small Ramsey numbers

By complementation, we have $R(s,t) = R(t,s)$. Also, it is a simple exercise to find $R(1,t)$ and $R(2,t)$. So let's assume $3 \le s \le t$.

Here are all Ramsey numbers known exactly:

s	t=3	4	5	6	7	8	9
3	6	9	14	18	23	28	36
4		18	25				

Disgracefully, it is 26 years since the last exact value was found:
— $R(4,5) = 25$ by McKay and Radziszowski.

Apart from those few precise values, there are many far-apart bounds, for example

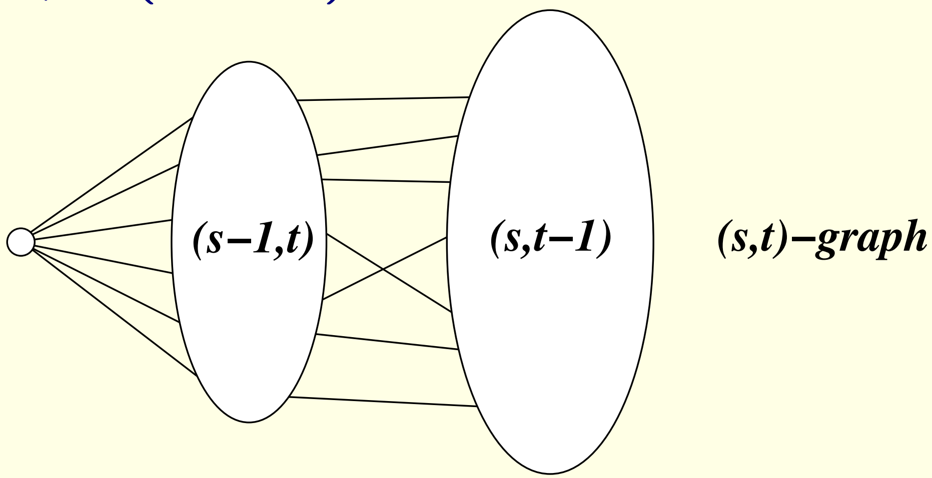
$$134 \le R(6,8) \le 495.$$

Small Ramsey numbers (continued)

Except for some asymptotic results using probabilistic methods, lower bounds are obtained by constructing examples:
If you find an (s, t)-graph with n vertices, then $R(s, t) \geq n + 1$.

The most recent such lower bounds for small s, t were found by Geoff Exoo: $R(3, 10) \geq 40$, $R(4, 6) \geq 36$ and $R(5, 5) \geq 43$, and many others by Exoo and Tatarevic such as $R(4, 8) \geq 58$.

Good upper bounds are much harder to prove. Here is a proof that $R(s, t) \leq R(s-1, t) + R(s, t-1)$:



$R(5,5) \le 48$

The most-sought unknown Ramsey number is $R(5,5)$.

Lower bound	Upper bound	Who
38		Abbot (1965)
	59	Kalbfleish (1965)
	58	Giraud (1967)
	57	Walker (1968)
	55	Walker (1971)
42		Irving (1973)
43		Exoo (1989)
	53	McKay & Radiszowski (1992)
	52	McKay & Radiszowski (1994)
	50	McKay & Radiszowski (1995)
	49	McKay & Radiszowski (1997)
	48	Angeltveit & McKay (2018)

Probably $R(5,5) = 43$.

There are 652 known $(5,5)$-graphs with 42 vertices.

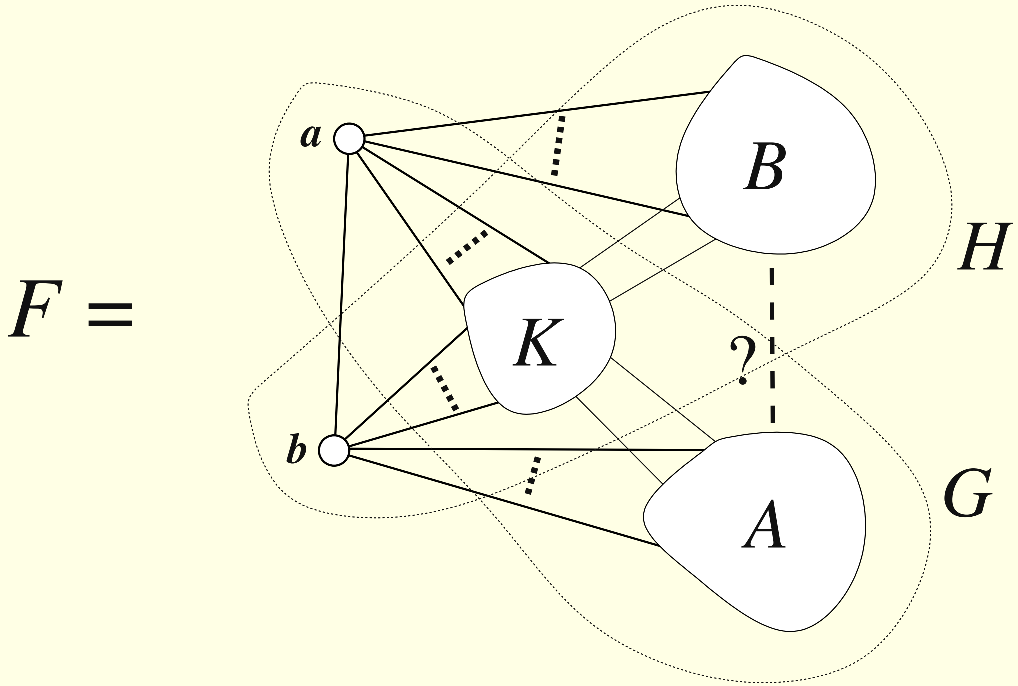
None of them extend to 43 vertices.

Moreover, many heuristic searches for $(5,5)$-graphs with 42 vertices keep finding the same ones over and over.

If $R(5,5) \geq 44$, we expect (but can't prove) that the number of $(5,5)$-graphs with 42 vertices would be extremely large.

$R(5,5) \leq 48$ (continued)

Consider a $(5,5)$-graph F with two adjacent vertices a, b of degree 24. Let the neighbourhoods of a and b, be H and G, respectively, and the intersection of their neighbourhoods be K.



Note that G, H are $(4,5)$-graphs, and K is a $(3,5)$-graph.

What remains is to fill in the edges between A and B.

$R(5,5) \le 48$ (continued)

1. Complete the catalogue of $(4,5)$-graphs on 24 vertices. McKay & Radziszowski found 350,904 in 1995. The complete set has 352,366 graphs.

2. Prove an easy lemma:
For $|V(F)| = 48$, a and b can be chosen such that $6 \le |V(K)| \le 11$.

3. Find all ways to overlap two $(4,5)$-graphs so that their intersection has 6–11 vertices. There are about 2 trillion ways.

4. For each of the 2 trillion cases, find all ways to add edges between parts A and B to form a $(5,5)$-graph. This is a satisfiability problem with about 650,000 solutions.

5. For each solution (650,500 graphs with 37 or 38 vertices), show that it cannot be extended to a $(5,5)$-graph with one more vertex.

$R(5,5) \leq 48$ (continued)

1. Complete the catalogue of $(4,5)$-graphs on 24 vertices. McKay & Radziszowski found 350,904 in 1995. The complete set has 352,366 graphs.

2. Prove an easy lemma:
For $|V(F)| = 48$, a and b can be chosen such that $6 \leq |V(K)| \leq 11$.

3. Find all ways to overlap two $(4,5)$-graphs so that their intersection has 6–11 vertices. There are about 2 trillion ways.

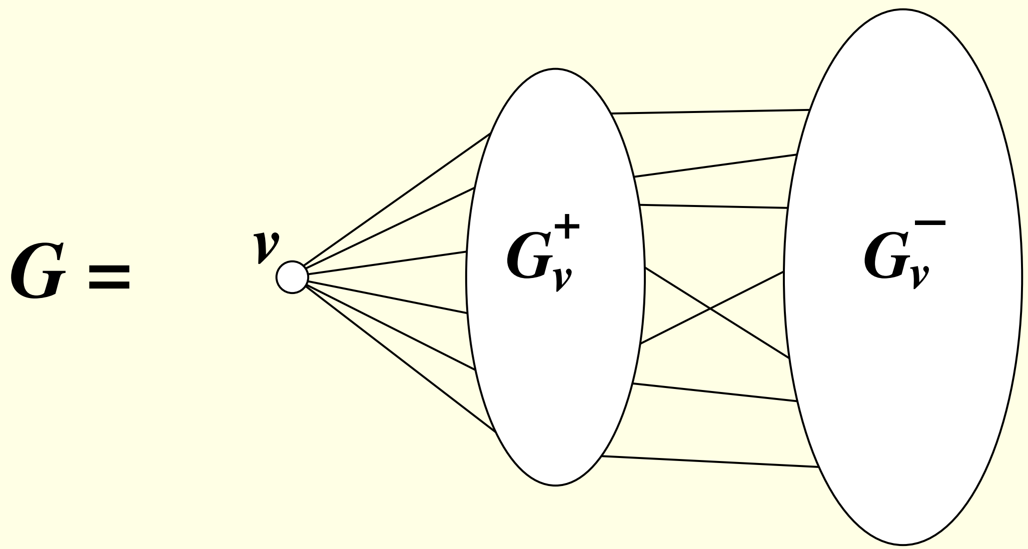
4. For each of the 2 trillion cases, find all ways to add edges between parts A and B to form a $(5,5)$-graph. This is a satisfiability problem with about 650,000 solutions.

5. For each solution (650,500 graphs with 37 or 38 vertices), show that it cannot be extended to a $(5,5)$-graph with one more vertex.

Conclusion: $R(5,5) \leq 48$.

Upper bounds on $R(s,t)$ by linear programming

For any graph G and vertex v, let G_v^+ be the subgraph induced by the neighbourhood of v, and let G_v^- be the subgraph induced by the complementary neighbourhood of v.



Recall that if G is an (s,t)-graph,

G_v^+ is an $(s{-}1,t)$-graph and G_v^- is an $(s,t{-}1)$-graph.

Upper bounds (continued)

For any graph X, define:

$v(X) =$ the number of vertices of X
$e(X) =$ the number of edges of X
$t(X) =$ the number of triangles in X
$p(X) =$ the number of induced paths of length 2 in X

$$g_2(X,n) = v(X)(n-2v(X)) + 2e(X)$$
$$g_3(X,n) = e(X)(n-3v(X)+3) + 6t(X) + 3p(X).$$

Then, if G has n vertices,

(I2) $2\sum_{v\in V(G)} e(G_v^-) = \sum_{v\in V(G)} g_2(G_v^+, n)$ (Goodman)

(I3) $3\sum_{v\in V(G)} t(G_v^-) = \sum_{v\in V(G)} g_3(G_v^+, n)$ (McKay & Radziszowski)

Linear programming (continued)

Identities (I2) and (I3) can be turned into linear programs for the existence of (s,t)-graphs.

The variables are
$g(i,j) =$ the number of neighbourhoods with i vertices and j edges
$h(i,j) =$ the number of complementary neighbourhoods with i vertices and j edges

The parameters $t(X)$ and $2t(X)+p(X)$ are not determined by the number of vertices and edges, but they are functions of $\{g(i,j),h(i,j)\}$ so we can find bounds on them by linear programs.

For small graphs, we can find better bounds on $t(X)$ and $2t(X)+p(X)$ by generating the graphs (for example, the $(3,7)$-graphs mentioned earlier).

Linear programming (continued)

To start the computation, we found precise bounds on $e(X)$, $t(X)$ and $2t(X) + p(X)$ for as many small graphs as possible. Then we used I2 and I3 to compute bounds for larger graphs by linear programming.

We used the exact simplex solver in the package GLPK by Andrew Makhorin.

McKay & Radziszowki used this approach in 1997 in finding $R(5,5) \leq 49$ and $R(4,6) \leq 41$.

For triangle-free graphs, I3 is a tautology, but I2 is still useful. Goedgebeur & Radziszowki used this in 2013 to improve the upper bounds for $R(3,t)$ for many t.

Linear programming (continued)

As the parameters grow, the numbers of variables and constraints in the linear programs grows into the millions. The number of linear programs which need to be to solved also grows into the millions. So there is a limit to how far we can go without some improvements.

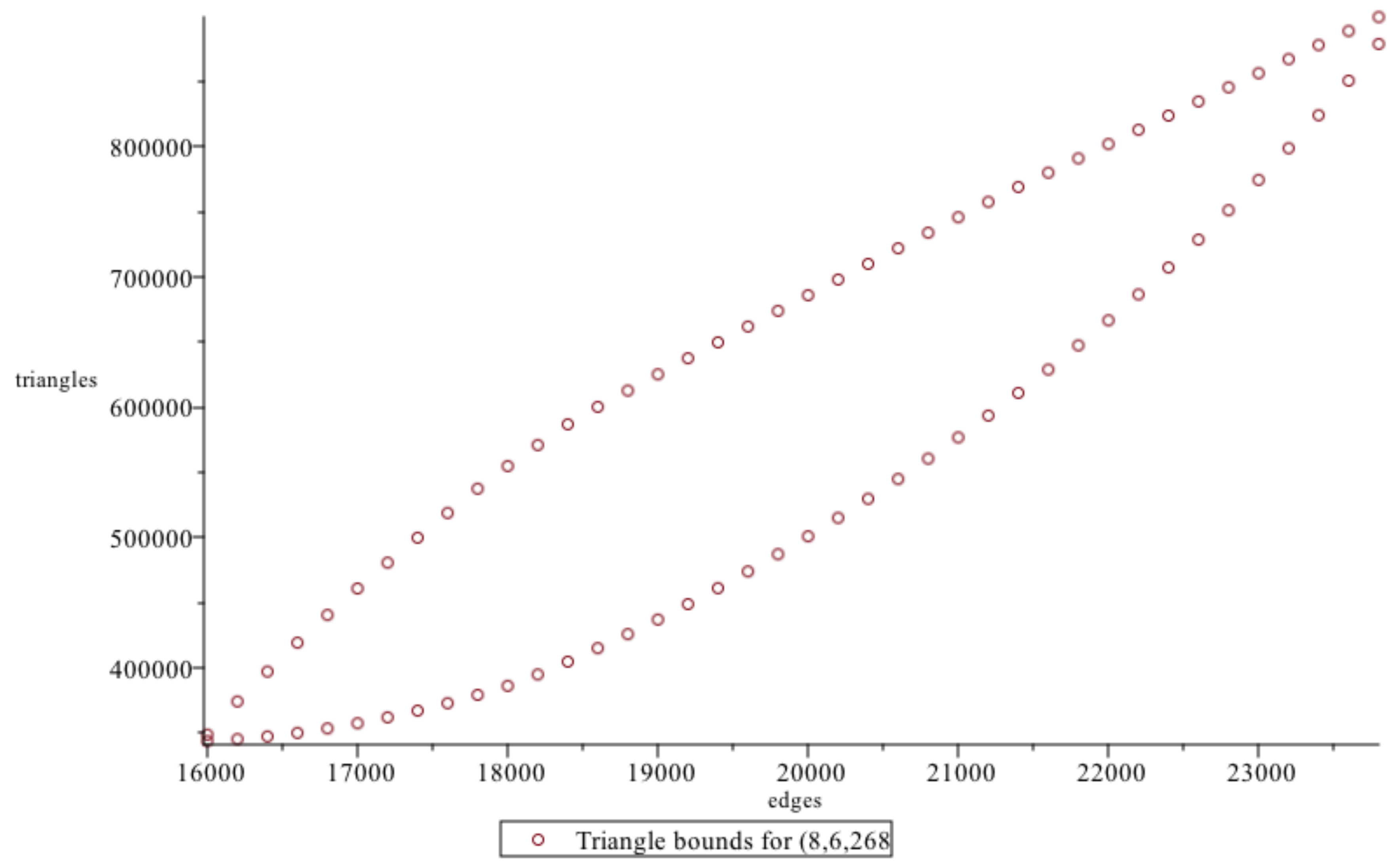
To get over this hurdle, we developed methods for reducing the number of variables and reducing the number of linear programs that are needed. These come at the cost of a slight degradation in accuracy.

To reduce the number of variables we show that, to some accuracy, some variables can be replaced by convex combinations of other variables. In this way we can often reduce the number of variables by a factor of 10–50.

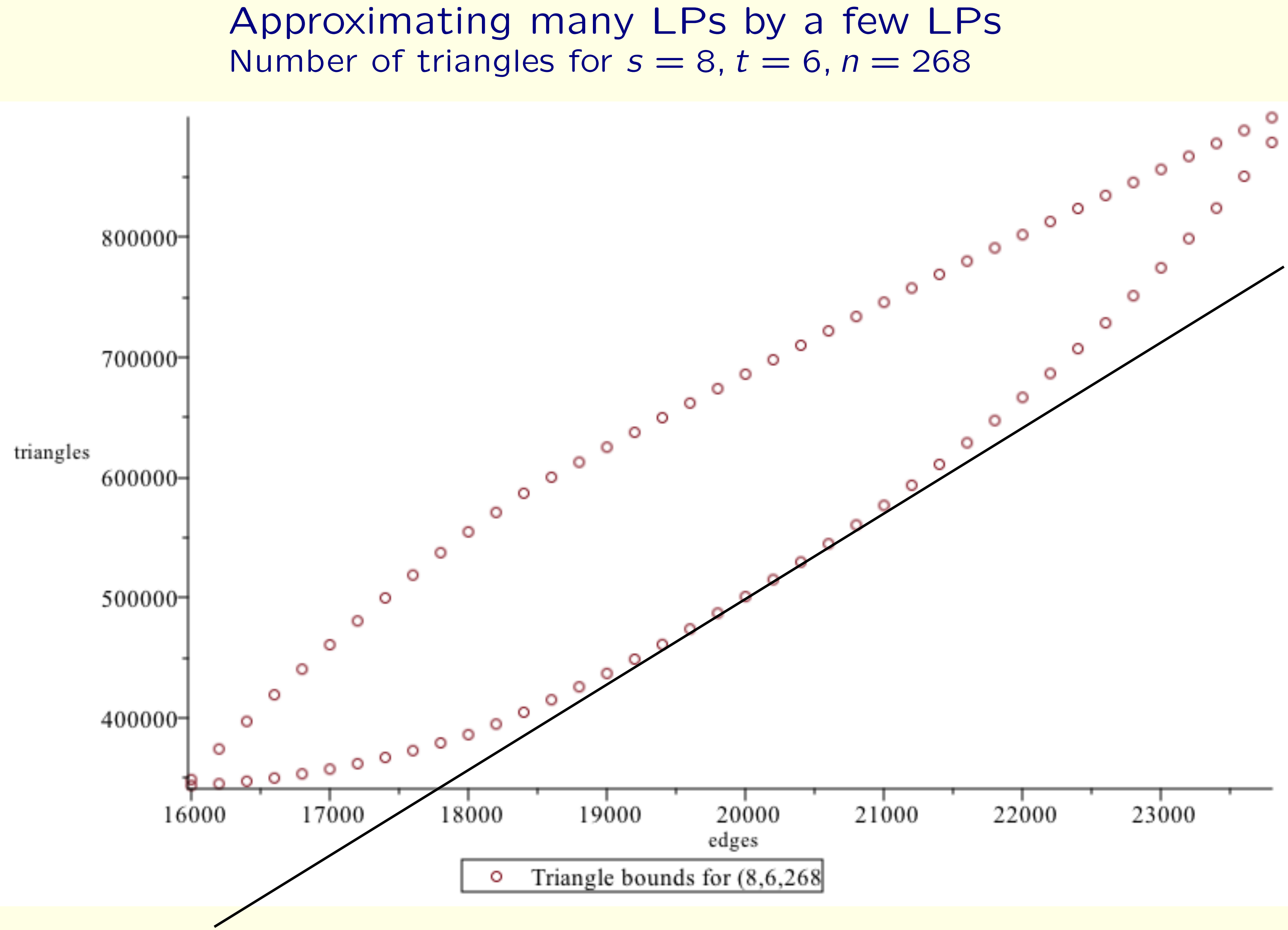
To reduce the number of linear programs, we used a method of successive approximation.

Linear programming (continued)

Approximating many LPs by a few LPs
Number of triangles for $s = 8$, $t = 6$, $n = 268$



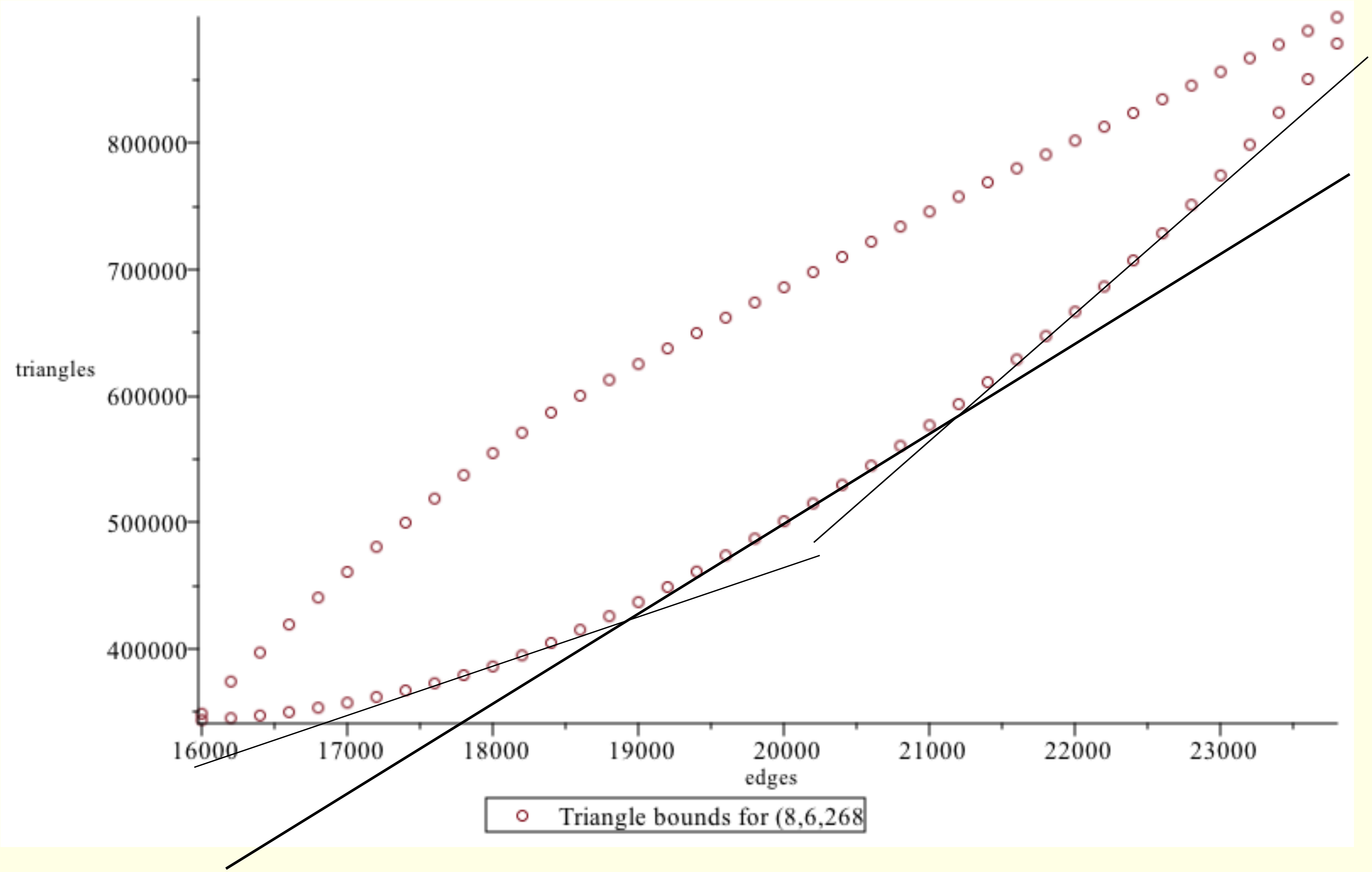
Linear programming (continued)



Linear programming (continued)

Approximating many LPs by a few LPs

Number of triangles for $s = 8$, $t = 6$, $n = 268$



Linear programming (examples of the results)

$R(s,t)$	lower bound	old upper bound	new upper bound
$R(4,6)$	36	41	40
$R(4,7)$	49	61	58
$R(4,8)$	59	84	79
$R(4,9)$	73	115	106
$R(4,10)$	92	149	136
$R(4,11)$	102	191	171
$R(4,12)$	128	238	211
$R(5,5)$	43	49	48
$R(5,6)$	58	87	85
$R(5,7)$	80	143	133
$R(5,8)$	101	216	194
$R(5,9)$	133	316	282
$R(6,6)$	102	165	161
$R(6,7)$	115	298	273
$R(7,7)$	205	540	497
$R(8,8)$	282	1870	1532

(Work in progress; please don't cite yet.)

A note on correctness

Since we are aiming to prove mathematical theorems, we want to minimize any chance of error. One nice property of linear programming is that the solution can be rigorously checked.

We are given vectors $\boldsymbol{b}$ and $\boldsymbol{c}$, and a matrix A. Define two linear programming problems:

Primal problem:
Find $\boldsymbol{x}$ to maximize $\boldsymbol{c}^T\boldsymbol{x}$, subject to $A\boldsymbol{x} \leq \boldsymbol{b}$ and $\boldsymbol{x} \geq 0$.

Dual problem:
Find $\boldsymbol{y}$ to minimize $\boldsymbol{b}^T\boldsymbol{y}$, subject to $A^T\boldsymbol{y} \geq \boldsymbol{c}$ and $\boldsymbol{y} \geq 0$.

Theorem: If $\boldsymbol{x}$ satisfies $A\boldsymbol{x} \leq \boldsymbol{b}, \boldsymbol{x} \geq 0$, and $\boldsymbol{y}$ satisfies $A^T\boldsymbol{y} \geq \boldsymbol{c}, \boldsymbol{y} \geq 0$, and $\boldsymbol{c}^T\boldsymbol{x} = \boldsymbol{b}^T\boldsymbol{y}$, then $\boldsymbol{x}$ and $\boldsymbol{y}$ are solutions to the primal and dual problems.

Ramsey hypergraphs

It is easy to prove by induction that, for any $j, k, s \geq 2$ there exists a least integer $n = R(j, k; s) \geq s$ with this property:

For every colouring of all the s-subsets of an n-set using two colours, either there a j-subset all of whose s-subsets have the first colour, or a k-subset all of whose s-subsets have the second colour.

The case $s = 2$ is just the classical Ramsey number.
For $s \geq 3$, only one non-trivial value is known:

$R(4, 4; 3) = 13$ (McKay and Radziszowski, 1991)

Ramsey hypergraphs (continued)

Say we have a set and we colour each triple (3-subset) with one of two colours.
A monochromatic quadruple is a set of 4 points such that the 4 triples in it all have the same colour.

Let's say that a colouring of all the triples with 2 colours is good if there is no monochromatic quadruple.

For example, here is a good colouring for 7 points:

123, 124, 134, 125, 135, 245, 345, 236, 146, 246, 346,
456, 237, 457, 167, 567
234, 235, 145, 126, 136, 156, 256, 356, 127, 137, 147,
247, 347, 157, 257, 357, 267, 367, 467

$R(4,4;3) = 13$ means there are good colourings for 12 points but not for 13 points.

Ramsey hypergraphs (continued)

Now we would like to answer two questions:

1. How many (inequivalent) good colourings are there for 12 points?
2. How close can we get to a good colouring for 13 points?

Let $\mathcal{R}(n, e)$ be the set of good colourings for n points, with e triples of the first colour. Since goodness is preserved by exchanging the colours, we can assume

$$e \le \frac{1}{2}\binom{n}{3}.$$

Therefore, Question 1 is answered by determining $\mathcal{R}(12, \le 110)$.

By averaging, given $G \in \mathcal{R}(12, \le 110)$, we can find a point v so that, if we remove v and all the triples that include v, we obtain a good colouring in $\mathcal{R}(11, \le 82)$.

Ramsey hypergraphs (continued)

Continuing in such manner, we find a construction path

$$\mathcal{R}(9, \leq 41) \to \mathcal{R}(10, \leq 59) \to \mathcal{R}(11, \leq 82) \to \mathcal{R}(12, \leq 110),$$

where "$\to$" means to add a new point and colour the new triples.

Using old programs from 1991, it was easy to find that $\mathcal{R}(9, \leq 41)$ contains 3,030,480,232 inequivalent colourings.

However, at this stage expansion by an extra point becomes much harder and we need to do it 3,030,480,232 times.

Ramsey hypergraphs (continued)

Consider extending $\mathcal{R}(9, \leq 41) \to \mathcal{R}(10, \leq 59)$.

Suppose the points 1–9 are given and the new point is "a".

There are 36 new triples $\{i, j, a\}$ $(1 \leq i < j \leq 9)$ that we need to colour: associate 0-1 variables $e_0, \ldots, e_{35}$ in any order (0 for the first colour and 1 for the second colour).

Avoiding monochromatic quadruples is equivalent to solving a set of 84 inequalities of the form

$$1 \leq e_u + e_v + e_w + e_x \leq 3,$$

one for each quadruple of the form $\{i, j, k, a\}$.

Ramsey hypergraphs (continued)

Consider extending $\mathcal{R}(9, \leq 41) \to \mathcal{R}(10, \leq 59)$.

Suppose the points 1–9 are given and the new point is "a".

There are 36 new triples $\{i, j, a\}$ $(1 \leq i < j \leq 9)$ that we need to colour: associate 0-1 variables $e_0, \ldots, e_{35}$ in any order (0 for the first colour and 1 for the second colour).

Avoiding monochromatic quadruples is equivalent to solving a set of 84 inequalities of the form

$$1 \leq e_u + e_v + e_w + e_x \leq 3,$$

one for each quadruple of the form $\{i, j, k, a\}$.

Each solution corresponds to a subset of $\{0, 1, \ldots, 35\}$ (the indices of the variables that are equal to 1).

Ramsey hypergraphs (continued)

For $\emptyset \subseteq B \subseteq T \subseteq \{0, 1, \ldots, 35\}$ define the interval

$$[B, T] = \{X \mid B \subseteq X \subseteq T\}.$$

We can use an interval to represent all the solutions lying in the interval, starting with $[\emptyset, \{0, 1, \ldots, 35\}]$.

Then, for each quadruple (i.e., for each inequality) each interval can be replaced by 0–3 disjoint subintervals that contain all the solutions respecting that inequality.

This can be implemented very quickly, using two machine words to represent each interval. After applying each quadruple, the solution appears as a set of disjoint intervals.

Ramsey hypergraphs (continued)

For each input colouring, the set of all solutions is obtained as the union of (typically) a few hundred intervals.

Unfortunately, the result is that $|\mathcal{R}(10, \leq 59)| > 10^{11}$, and the expansion to 11 points has even more variables and more inequalities.

It would be hopeless if we had to process them one at a time by the same method that we added the 10th point, but there is a better way.

We observe: Given a colouring for the triples in $\{1, \ldots, 9\}$, the constraints on an 11th point b arising from quadruples $\{i, j, k, b\}$ are the same as the constraints on the 10th point a arising from quadruples $\{i, j, k, a\}$, except that a is replaced by b.

Ramsey hypergraphs (continued)

This means that in adding the 11th point b we can start with the intervals giving all the solutions for the 10th point, rather than with an interval containing everything. We only need to further satisfy the quadruples of the form $\{i, j, a, b\}$.

This is much faster.

A similar argument allows the extension $\mathcal{R}(11, \leq 82) \to \mathcal{R}(12, \leq 110)$ to be quite fast (most colourings don't extend at all).

Result:

There are 434,714 nonisomorphic good colourings for 12 points.

Ramsey hypergraphs (continued)

Finally, we consider how close we can get to a good colouring for 13 points.

The same computational methods apply.

Result: We can colour all but two of the triples of a 13-set such that there is no monochromatic quadruple; but not all but one.

Question: If all triples are coloured, what is the least possible number of monochromatic quadruples?