Subgraph Counting Identities and Ramsey Numbers

To the fond memory of Paul Erdős

Brendan D. McKay+
Department of Computer Science
Australian National University
Canberra, ACT 0200, Australia
bdm@cs.anu.edu.au

Stanislaw P. Radziszowski*
Department of Computer Science
Rochester Institute of Technology
Rochester, NY 14623, USA
spr@cs.rit.edu

Abstract.

For each vertex v of a graph G, we consider the numbers of subgraphs of each isomorphism class which lie in the neighbourhood or complementary neighbourhood of v. These numbers, summed over v, satisfy a series of identities that generalise some previous results of Goodman and ourselves. As sample applications, we improve the previous upper bounds on two Ramsey numbers. Specifically, we show that $R(5,5) \leq 49$ and $R(4,6) \leq 41$. We also give some experimental evidence in support of our conjecture that $R(5,5) = 43$.

1. Introduction.

We shall only consider graphs without multiple edges or loops. For $s, t, n \geq 1$, an (s,t)-graph is a graph without cliques of order s or independent sets of order t, and an (s,t,n)-graph is an (s,t)-graph of order n. Similarly, an (s,t,n,e)-graph is an (s,t,n)-graph with e edges. Let $\mathcal{R}(s,t)$, $\mathcal{R}(s,t,n)$ and $\mathcal{R}(s,t,n,e)$ denote the set of all (s,t)-graphs, (s,t,n)-graphs and (s,t,n,e)-graphs, respectively. The Ramsey number $R(s,t)$ is defined to be the least $n > 0$ such that there is no (s,t,n)-graph.

A regularly updated survey of the most recent results on this subject can be found in [20].

In Section 2, we derive some identities involving subgraph counts, which form the basis of our approach. In Section 3, we show that $R(5,5) \leq 49$, which improves over the previous bound of 50 [17]. Nevertheless, the correct value is more likely to be 43, for the reasons we give in Section 4. Finally, in Section 5, we show that $R(4,6) \leq 41$ by linear programming methods. Comprehensive surveys of the history of $R(5,5)$ and $R(4,6)$ will be given in the appropriate sections.

+ Supported in part by ARC grant ANU-F95003 and by the Computer Science Department of the Rochester Institute of Technology

* Supported in part by NSA grant MDA904-94-H-2009

2. Subgraph identities.

For two graphs J and G, let $s(J, G)$ denote the number of induced subgraphs of G that are isomorphic to J. It will be convenient to permit both J and G to be the "graph" K_0, which has no vertices or edges. In this case we define $s(K_0, G) = 1$ for all G and $s(J, K_0) = 0$ for all $J \neq K_0$.

A summary of much of what is known about this "algebra of subgraphs" can be found in [12]. For our purposes, the following theorem is important.

Theorem 2.1.
(a) For each disconnected graph J, there is a sequence of connected graphs $J_1, J_2, \ldots, J_k$ and a polynomial p_J with rational coefficients such that

$$s(J, G) = p_J\bigl(s(J_1, G), s(J_2, G), \ldots, s(J_k, G)\bigr)$$

for every graph G.
(b) There is no sequence of nonisomorphic connected graphs $J_1, J_2, \ldots, J_k$ and nonzero polynomial p such that

$$p\bigl(s(J_1, G), s(J_2, G), \ldots, s(J_k, G)\bigr) = 0$$

for all graphs G.

Proof. Part (a) was proved by Whitney [26], while part (b) follows from a considerably stronger result of Erdős, Lovász and Spencer [3]. ∎

We will need a particular case of part (a) of this theorem, stated as Lemma 2.1 below. For $m \geq 0$ and $0 \leq j \leq m$, define the graphs $T_{m,j}$ as follows. For $m = 0$, define $T_{0,0} = K_1$. For $m > 0$, $T_{m,0}$ is the disconnected graph $K_m \cup K_1$, and for $j \geq 0$, $T_{m,j+1}$ is formed by adding one edge to $T_{m,j}$. It is easy to see that this defines $T_{m,j}$ uniquely up to isomorphism and that $T_{m,m} = K_{m+1}$.

Lemma 2.1. *Suppose G is a graph with n vertices. Then, for $m \geq 0$,*

$$(n - m)s(K_m, G) = \sum_{j=0}^{m} \beta_{m,j}\, s(T_{m,j}, G), \tag{1}$$

where

$$\beta_{m,j} = \begin{cases} m + 1, & \text{if } j = m; \\ 2, & \text{if } j = m - 1; \\ 1, & \text{if } 0 \leq j \leq m - 2. \end{cases}$$

Proof. Since $n = s(K_1, G)$ and $T_{m,0}$ is the only disconnected graph appearing here, this is an special instance of Theorem 2.1 (a).

The cases $m = 0, 1$ are easy to check, so we can assume $m \geq 2$. Both sides of (1) count the number of subgraphs of the form $K_m \cup K_1$, induced or not. The left side of (1) is obvious in this context. For the right side, consider the number j of edges that join the K_m to the K_1. These $m+1$ vertices induce a subgraph $T_{m,j}$. Finally, note that each subgraph $T_{m,j}$ can arise in $s(K_m, T_{m,j}) = \beta_{m,j}$ such ways. ∎

For $m = 2$, Lemma 2.1 becomes

$$(n-2)s(K_2, G) = s(T_{2,0}, G) + 2s(T_{2,1}, G) + 3s(K_3, G),$$

which is equivalent to Goodman's identity [7].

We will find it convenient to adopt the following notational conventions. If G is a graph, then VG and EG are its vertex set and edge set, respectively. If $v \in VG$ and $W \subseteq VG$, then $N_G(v, W) = \{w \in W \mid vw \in EG\}$. The subgraph of G induced by W will be denoted by $G[W]$. Also define the induced subgraphs $G_v^+ = G[N_G(v, VG)]$ and $G_v^- = G[VG - N_G(v, VG) - \{v\}]$.

Lemma 2.2. *Let J and G be graphs.*
(a) If J has $k \geq 1$ vertices of degree $|VJ| - 1$, then

$$k\, s(J, G) = \sum_{v \in VG} s(J', G_v^+),$$

where J' is the result of removing from J a vertex of degree $|VJ| - 1$.
(b) If J has $k \geq 1$ vertices of degree 0, then

$$k\, s(J, G) = \sum_{v \in VG} s(J'', G_v^-),$$

where J'' is the result of removing from J a vertex of degree 0.

Proof. In case (a), each subgraph isomorphic to J lies in $\{v\} \cup N_G(v, VG)$ for exactly k vertices v, so both sides of the identity count induced subgraphs isomorpic to J with a vertex of maximum degree distinguished. Case (b) is similar. ∎

Each of the subgraphs involved in Lemma 2.1 matches one of the types considered by Lemma 2.2. This yields a family of identities involving those functions. Let $\delta_{i,j}$ denote the Kronecker delta.

Theorem 2.2. *For $m \geq 1$, every graph G satisfies*

$$\sum_{v \in VG} s(K_m, G_v^-) = \sum_{v \in VG} \Big(\big(n/m - s(K_1, G_v^+) + m - 2\big) s(K_{m-1}, G_v^+) + (m-1)s(K_m, G_v^+) + \sum_{j=1}^{m-2} \frac{(1 + \delta_{j,m-2})j}{j+1} s(T_{m-1,j}, G_v^+) \Big).$$

Proof. The case $m = 1$ is easy to check directly, so we will assume $m \geq 2$.

From Lemma 2.2, using (b) for $j = 0$ and (a) for $j > 0$, we have

$$s(T_{m,j}, G) = \begin{cases} \dfrac{1}{1 + \delta_{m,1}} \displaystyle\sum_{v \in VG} s(K_m, G_v^-), & \text{for } j = 0; \\ \dfrac{1}{j + \delta_{j,m}} \displaystyle\sum_{v \in VG} s(T_{m-1,j-1}, G_v^+), & \text{for } 1 \leq j \leq m. \end{cases}$$

Applying Lemma 2.2 (a) for $J = K_m$, we can substitute into Lemma 2.1 to obtain

$$\frac{n-m}{m} \sum_{v \in VG} s(K_{m-1}, G_v^+) = \sum_{v \in VG} s(K_m, G_v^-) + \sum_{j=1}^{m} \frac{\beta_{m,j}}{j + \delta_{j,m}} \sum_{v \in VG} s(T_{m-1,j-1}, G_v^+). \tag{2}$$

All the subgraphs appearing as the first argument of $s(\,)$ in (2) are connected except $T_{m-1,0}$. Using Lemma 2.1 again, we have that

$$s(T_{m-1,0}, G_v^+) = \frac{1}{\beta_{m-1,0}} \Big((s(K_1, G_v^+) - m + 1) s(K_{m-1}, G_v^+) - \sum_{j=1}^{m-1} \beta_{m-1,j} s(T_{m-1,j}, G_v^+) \Big).$$

Substituting into (2) and collecting similar terms gives the desired identity. ∎

The case of $m = 1$ is elementary, and the case of $m = 2$ is equivalent to Goodman's identity. Though less obvious, the identity for $m = 3$ can be derived from Lemma 2 of [15]. The later identities are new as far as we know.

It is interesting to consider the question of the completeness of Theorem 2.2. That is, what other identities of similar form are there? We have explored this question by experimental means. Consider identities with the general form

$$\sum_{v \in VG} p(G_v^+, G_v^-) = 0,$$

where p is a polynomial in items of the form $s(J, G_v^+)$ and $s(J, G_v^-)$ for some family of connected graphs J. The coefficients can be arbitrary functions of $n = s(K_1, G)$. The restriction to connected J is justified by Theorem 2.1. We further forbid the term $s(K_1, G_v^-)$, as it can be replaced by $n - 1 - s(K_1, G_v^+)$.

Define the *degree* of p to be the maximum total number of vertices appearing (as the first argument of s) in a single term of p. Our experiment was to take large numbers of random graphs of the same order, and count the numbers $s(J, G_v^+)$ and $s(J, G_v^-)$ for each vertex v and small connected graph J. Then we formed a matrix of values of the possible terms of p, up to some fixed degree with one row per graph and one column per term. The rank of this matrix, and linear relationships between the columns, tell us about identities satisfied by the set of graphs we have chosen. In particular, linear independence can prove the nonexistence of particular types of identity for these graphs and hence for all graphs. For example, we have established:

Lemma 2.3. *The only identities of degree at most 6, in which p can be separated as $p(G_v^+, G_v^-) = p_1(G_v^+) + p_2(G_v^-)$, are those of Theorem 2.2 and their linear combinations.* ∎

If p does not have to separate in the manner of the lemma, we suspect that further identities exist. For example, the following identity of degree 4 holds for such a large number of random graphs (many thousands) that we conjecture it to hold always. Let P_k and C_k denote the path and cycle of length k, respectively.

Conjecture 1. *For every graph G, $\sum_{v \in VG} \bigl(p_1(G_v^+) + p_2(G_v^-) + p_3(G_v^+, G_v^-)\bigr) = 0$, where*

$$\begin{aligned} p_1(X) &= n(n-3)s(K_1,X) - (n^2+2n-6)s(K_1,X)^2 + 3ns(K_1,X)^3 \\ &\quad - 2s(K_1,X)^4 + 2(n^2+n-8)s(K_2,X) - 12s(K_2,X)^2 \\ &\quad - 12(n-1)s(K_1,X)s(K_2,X) + 12s(K_1,X)^2 s(K_2,X) + 72s(C_4,X) \\ &\quad + 12(n-2)s(K_3,X) + 24s(K_{1,3},X) + 24s(P_4,X) + 24s(T_{3,1},X) \\ &\quad + 12(n+2)s(P_3,X) - 24s(K_1,X)s(P_3,X) + 32s(T_{3,2},X), \\ p_2(Y) &= 4s(K_2,Y)^2 - 12s(K_{1,3},Y) - 8s(C_4,Y) - 8s(T_{3,1},Y) \\ &\quad - 24s(T_{3,2},Y) + 2(n-8)s(P_3,Y), \end{aligned}$$

and

$$\begin{aligned} p_3(X,Y) &= 4s(K_1,X)s(P_3,Y) - 2(n-2)s(K_1,X)s(K_2,Y) \\ &\quad + 4s(K_1,X)^2 s(K_2,Y). \quad ∎ \end{aligned}$$

We also have a tentative identity of degree 5, but it is even more complicated. We expect that there is a rich theory of such identities, but we have merely scratched the surface.

3. A proof that $R(5,5) \le 49$.

A history of the known bounds on $R(5,5)$ is presented in Table I. The initials "LP" refer to linear programming techniques.

Our theorem that $R(4,5) = 25$ [17] implies immediately that $R(5,5) \le 50$. Moreover, it implies that any $(5,5,49)$-graph G must be regular of degree 24, with each G_v^+ being a $(4,5,24)$-graph and each G_v^- being the complement of a $(4,5,24)$-graph. (Note that $\bar{G}$, the complement of G, is also a $(5,5,49)$-graph.) Applying the case $m = 2$ of Theorem 2.2, we find

$$\sum_{v \in VG} s(K_2, G_v^-) = 588 + \sum_{v \in VG} s(K_2, G_v^+).$$

year	reference	lower	upper	comments
1965	Abbott [1]	38		quadratic residues in Z_{37}
1965	Kalbfleisch [9]		59	pointer to a future paper
1967	Giraud [6]		58	combinatorics & LP
1968	Walker [24]		57	combinatorics & LP
1971	Walker [25]		55	combinatorics & LP
1973	Irving [8]	42		sum-free sets
1989	Exoo [4]	43		simulated annealing
1992	McKay & Radziszowski [15]		53	$(4,4)$-graph enumeration & LP
1994	McKay & Radziszowski [16]		52	LP & computation
1995	McKay & Radziszowski [17]		50	implication of $R(4,5)=25$
1995	McKay & Radziszowski		49	this paper

Table I. The history of bounds on $R(5,5)$.

Since also $s(K_2, G_v^-) = \binom{24}{2} - s(K_2, \bar{G}_v^+)$, we have that

$$\sum_{v \in VG} \left(s(K_2, G_v^+) + s(K_2, \bar{G}_v^+) \right) = 12936.$$

However, from the computations reported in [17] we know that $(4,5,24)$-graphs have at most 132 edges, and that there are no such graphs with maximum degree greater than 11. This leaves only graphs regular of degree 11, which gives the following key lemma.

Lemma 3.1. *Let G be a $(5,5,49)$-graph. Then, for each vertex v, G_v^+ and $\bar{G}_v^+$ are $(4,5,24,132)$-graphs which are regular of degree 11.* ∎

It is possible to derive some reasonably strong restrictions on those $(4,5,24,132)$-graphs which might fit into a $(5,5,49)$-graph, but we decided to aim instead to find all $(4,5,24,132)$-graphs. Two such graphs were found previously by Thomason [23], under the stronger conditions of both regularity and a constant number of triangles on each edge. These are the graphs H_1 and H_2 given in Figure 1.

Both H_1 and H_2 are vertex-transitive, so for information we give their automorphism groups. Define

$$\begin{aligned}
g_1 &= (0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 10\ 11)(12\ 13\ 14\ 15\ 16\ 17\ 18\ 19\ 20\ 21\ 22\ 23),\\
g_2 &= (0\ 12)(1\ 17)(2\ 22)(3\ 15)(4\ 20)(5\ 13)(6\ 18)(7\ 23)(8\ 16)(9\ 21)(10\ 14)(11\ 19),\\
g_3 &= (1\ 11)(2\ 10)(3\ 9)(4\ 8)(5\ 7)(13\ 23)(14\ 22)(15\ 21)(16\ 20)(17\ 19),\\
g_4 &= (0\ 12)(1\ 23)(2\ 22)(3\ 21)(4\ 20)(5\ 19)(6\ 18)(7\ 17)(8\ 16)(9\ 15)(10\ 14)(11\ 13).
\end{aligned}$$

Then $\mathrm{Aut}(H_1) = \langle g_1, g_2, g_3 \rangle$, of order 48, and $\mathrm{Aut}(H_2) = \langle g_1, g_4 \rangle$, of order 24.

H_1

```
0 1 0 0 1 0 0 0 1 0 0 1  0 1 0 0 1 1 1 1 1 0 0 1
1 0 1 0 0 1 0 0 0 1 0 0  1 0 1 0 0 1 1 1 1 1 0 0
0 1 0 1 0 0 1 0 0 0 1 0  0 1 0 1 0 0 1 1 1 1 1 0
0 0 1 0 1 0 0 1 0 0 0 1  0 0 1 0 1 0 0 1 1 1 1 1
1 0 0 1 0 1 0 0 1 0 0 0  1 0 0 1 0 1 0 0 1 1 1 1
0 1 0 0 1 0 1 0 0 1 0 0  1 1 0 0 1 0 1 0 0 1 1 1
0 0 1 0 0 1 0 1 0 0 1 0  1 1 1 0 0 1 0 1 0 0 1 1
0 0 0 1 0 0 1 0 1 0 0 1  1 1 1 1 0 0 1 0 1 0 0 1
1 0 0 0 1 0 0 1 0 1 0 0  1 1 1 1 1 0 0 1 0 1 0 0
0 1 0 0 0 1 0 0 1 0 1 0  0 1 1 1 1 1 0 0 1 0 1 0
0 0 1 0 0 0 1 0 0 1 0 1  0 0 1 1 1 1 1 0 0 1 0 1
1 0 0 1 0 0 0 1 0 0 1 0  1 0 0 1 1 1 1 1 0 0 1 0

0 1 0 0 1 1 1 1 1 0 0 1  0 0 0 0 1 1 0 1 1 0 0 0
1 0 1 0 0 1 1 1 1 1 0 0  0 0 0 0 0 1 1 0 1 1 0 0
0 1 0 1 0 0 1 1 1 1 1 0  0 0 0 0 0 0 1 1 0 1 1 0
0 0 1 0 1 0 0 1 1 1 1 1  0 0 0 0 0 0 0 1 1 0 1 1
1 0 0 1 0 1 0 0 1 1 1 1  1 0 0 0 0 0 0 0 1 1 0 1
1 1 0 0 1 0 1 0 0 1 1 1  1 1 0 0 0 0 0 0 0 1 1 0
1 1 1 0 0 1 0 1 0 0 1 1  0 1 1 0 0 0 0 0 0 0 1 1
1 1 1 1 0 0 1 0 1 0 0 1  1 0 1 1 0 0 0 0 0 0 0 1
1 1 1 1 1 0 0 1 0 1 0 0  1 1 0 1 1 0 0 0 0 0 0 0
0 1 1 1 1 1 0 0 1 0 1 0  0 1 1 0 1 1 0 0 0 0 0 0
0 0 1 1 1 1 1 0 0 1 0 1  0 0 1 1 0 1 1 0 0 0 0 0
1 0 0 1 1 1 1 1 0 0 1 0  0 0 0 1 1 0 1 1 0 0 0 0
```

H_2

```
0 1 1 0 1 0 0 0 1 0 1 1  1 0 0 0 1 1 0 0 0 1 0 1
1 0 1 1 0 1 0 0 0 1 0 1  1 1 0 0 0 1 1 0 0 0 1 0
1 1 0 1 1 0 1 0 0 0 1 0  0 1 1 0 0 0 1 1 0 0 0 1
0 1 1 0 1 1 0 1 0 0 0 1  1 0 1 1 0 0 0 1 1 0 0 0
1 0 1 1 0 1 1 0 1 0 0 0  0 1 0 1 1 0 0 0 1 1 0 0
0 1 0 1 1 0 1 1 0 1 0 0  0 0 1 0 1 1 0 0 0 1 1 0
0 0 1 0 1 1 0 1 1 0 1 0  0 0 0 1 0 1 1 0 0 0 1 1
0 0 0 1 0 1 1 0 1 1 0 1  1 0 0 0 1 0 1 1 0 0 0 1
1 0 0 0 1 0 1 1 0 1 1 0  1 1 0 0 0 1 0 1 1 0 0 0
0 1 0 0 0 1 0 1 1 0 1 1  0 1 1 0 0 0 1 0 1 1 0 0
1 0 1 0 0 0 1 0 1 1 0 1  0 0 1 1 0 0 0 1 0 1 1 0
1 1 0 1 0 0 0 1 0 1 1 0  0 0 0 1 1 0 0 0 1 0 1 1

1 1 0 1 0 0 0 1 1 0 0 0  0 1 1 0 1 0 0 0 1 0 1 1
0 1 1 0 1 0 0 0 1 1 0 0  1 0 1 1 0 1 0 0 0 1 0 1
0 0 1 1 0 1 0 0 0 1 1 0  1 1 0 1 1 0 1 0 0 0 1 0
0 0 0 1 1 0 1 0 0 0 1 1  0 1 1 0 1 1 0 1 0 0 0 1
1 0 0 0 1 1 0 1 0 0 0 1  1 0 1 1 0 1 1 0 1 0 0 0
1 1 0 0 0 1 1 0 1 0 0 0  0 1 0 1 1 0 1 1 0 1 0 0
0 1 1 0 0 0 1 1 0 1 0 0  0 0 1 0 1 1 0 1 1 0 1 0
0 0 1 1 0 0 0 1 1 0 1 0  0 0 0 1 0 1 1 0 1 1 0 1
0 0 0 1 1 0 0 0 1 1 0 1  1 0 0 0 1 0 1 1 0 1 1 0
1 0 0 0 1 1 0 0 0 1 1 0  0 1 0 0 0 1 0 1 1 0 1 1
0 1 0 0 0 1 1 0 0 0 1 1  1 0 1 0 0 0 1 0 1 1 0 1
1 0 1 0 0 0 1 1 0 0 0 1  1 1 0 1 0 0 0 1 0 1 1 0
```

Figure 1. Adjacency matrices for H_1 and H_2.

From now on, H will denote a $(4,5,24,132)$-graph. Since H is 11-regular, it is easy to see that $s(K_2,H_v^-)=s(K_2,H_v^+)+11$ for each v. Thus, we can find H by "gluing" together some $X\in\mathcal{R}(3,5,11,e)$ and $Y\in\mathcal{R}(4,4,12,e+11)$ for some e. The number of possibilities is listed in Table II.

e	$\lvert\mathcal{R}(3,5,11,e)\rvert$	$\lvert\mathcal{R}(4,4,12,e{+}11)\rvert$
15	1	8
16	6	177
17	19	1906
18	31	13332
19	30	58131
20	13	163757
21	4	302088
22	1	370368

Table II. Numbers of potential parts of $(4,5,24,132)$-graphs.

Theorem 2.2 can help us to reduce the number of possibilities somewhat.

Lemma 3.2. *For some v, $s(K_2,H_v^+)\ge 19$.*

Proof. For $15\le e\le 18$, the right side of Theorem 2.2 is at most 9 for every graph in $\mathcal{R}(3,5,11,e)$, but the left side is at least 10 for every graph in $\mathcal{R}(4,4,12,e+11)$. (These numbers were directly computed from the graphs themselves.) Hence no combination of such graphs can satisfy the identity. ∎

Given Lemma 3.2, we can construct all of $\mathcal{R}(4,5,24,132)$ using the methods described in [17], but there are many more cases to process and they are more difficult computationally. Fortunately, we can take advantage of the regularity to improve the efficiency of the search.

To describe the improved search, it is necessary to summarise the setting from [17]. That paper should be consulted for more details.

Suppose we have a particular $X \in \mathcal{R}(3,5,11)$ and $Y \in \mathcal{R}(4,4,12)$ and we wish to build them into $H \in \mathcal{R}(4,5,24,132)$. We need to choose the edges between X and Y. A *feasible cone* is a subset of VY that covers no clique of order 3. To avoid cliques of order 4, the neighbourhood in Y of each vertex in X must be a feasible cone. The set of all feasible cones can be packed into a smaller number of *intervals* of feasible cones, which are sets of cones of the form $[B,T] = \{W \mid B \subseteq W \subseteq T\}$.

Suppose $m = |VX|$. If $C_0,\ldots,C_{m-1}$ are feasible cones, then $F(X,Y;C_0,\ldots,C_{m-1})$ denotes the graph H with vertex v such that $H_v^+ = X$, $H_v^- = Y$ and $N_H(i, VY) = C_i$ for $0 \le i \le m-1$. Similarly, if $I_0,\ldots,I_{m-1}$ are intervals, then $\mathcal{F}(X,Y;I_0,\ldots,I_{m-1})$ consists of all $(4,5,24,132)$-graphs $F(X,Y;C_0,\ldots,C_{m-1})$ such that $C_i \in I_i$ for $0 \le i \le m-1$. The primary tool is a set of *collapsing rules*, which take as an argument a sequence $(X,Y;I_0,\ldots,I_{m-1})$ and return a sequence $(X,Y;I'_0,\ldots,I'_{m-1})$ such that $I'_i \subseteq I_i$ for $0 \le i \le m-1$ and $\mathcal{F}(X,Y;I'_0,\ldots,I'_{m-1}) = \mathcal{F}(X,Y;I_0,\ldots,I_{m-1})$. A collapsing rule is also permitted to generate the special event FAIL if $\mathcal{F}(X,Y;I_0,\ldots,I_{m-1}) = \emptyset$.

Four collapsing rules are given in [17]. If we have restrictions on the size of feasible cones, we can add some more rules.

Define two functions $K,T : 2^{VY} \to 2^{VY}$ such that, for $W \subseteq VY$,

$$\begin{aligned} K(W) &= \bigcap \bigl\{ \{x,y\} \bigm| x,y \in W \text{ and } \{x,y\} \in EH \bigr\}; \\ L(W) &= \bigcap \bigl\{ \{w,x,y,z\} \bigm| w,x,y,z \in W \text{ are distinct and } \{w,x\},\{y,z\} \in EH \bigr\}, \end{aligned}$$

with the understanding that the value of the intersection is VY if it has no arguments. These functions can be precomputed quickly for all $W \subseteq VY$ using simple recurrences.

Suppose that for each $u \in VX$, C_u is required to satisfy $l_u \le |C_u| \le h_u$. Let the corresponding interval be $I_u = [B_u, T_u]$. Then we can define the following rules.

(a) Suppose $u \in VX$.
if $|B_u| > h_u$, then FAIL.
if $|B_u| = h_u$, then $T_u := B_u$

(b) Suppose $u \in VX$.
if $|T_u| < l_u$, then FAIL
if $|T_u| = l_u$, then $B_u := T_u$

(c) Suppose $\{u, v\} \in EX$ and $|T_u| = l_u + 1$.
if $K(B_v \cap T_u) = \emptyset$, then FAIL
else $B_u := B_u \cup \big(T_u - K(B_v \cap T_u)\big)$

(d) Suppose $\{u, v\} \in EX$, $|T_u| = l_u + 1$, and $|T_v| = l_v + 1$.
if $|L(T_u \cap T_v)| \leq 1$, then FAIL
else $B_u := B_u \cup \big(T_u - L(T_u \cap T_v)\big)$

Lemma 3.3. *Rules (a)–(d) are valid collapsing rules.*

Proof. Rules (a) and (b) are an obvious application of the size restrictions.

Suppose $\{x, y\} \in EY$, $x, y \in B_v \cap T_u$ and $|T_u| = l_u + 1$. We can't have that $x, y \in C_u$ because then $\{u, v, x, y\}$ is a clique, so we must have one of x, y missing from C_u and all the rest of T_u equal to C_u (or else $|C_u| < l_u$).

Extending the same argument, we see that exactly one element of $K(B_v \cap T_u)$ must be avoided and the rest of T_u included. This is rule (c).

Suppose $\{w, x\}, \{y, z\} \in EY$, where w, x, y, z are distinct elements of $T_u \cap T_v$, $|T_u| = l_u + 1$, and $|T_v| = l_v + 1$. As before, exactly one of w and x, and exactly one of y and z, are not in $C_u \cap C_v$. The restrictions on the sizes of T_u and T_v imply that each of C_u and C_v are missing one of $\{w, x, y, z\}$ (but not the same one) and so must equal all of the rest of T_u and T_v, respectively. Applying this idea simultaneously to all pairs of edges $\{w, x\}, \{y, z\}$ gives rule (d). ∎

The method by which these collapsing rules were built into a search procedure was the same as in [17], so we will not repeat it. Several implementations were made and compared at intermediate points on a large number of examples. Then the fastest was run to completion, establishing the following theorem.

Theorem 3.1. *The only two* $(4, 5, 24, 132)$*-graphs are those in Figure 1.* ∎

Theorem 3.2. $R(5, 5) \leq 49$.

Proof. If there exists a $(5, 5, 49)$-graph G, then by Lemma 3.1 and Theorem 3.1 we know that G_v^+ and $\bar{G}_v^+$ are one of H_1 and H_2. Consider the identity of Theorem 2.2 applied to G for $m = 4$.

The relevant subgraph counts are as follows.

$$s(K_2, H_1) = s(K_2, H_2) = 132; \quad s(K_3, H_1) = s(K_3, H_2) = 176$$
$$s(K_4, H_1) = s(K_4, H_2) = 0; \quad s(T_{3,1}, H_1) = s(T_{3,1}, H_2) = 1584$$
$$s(T_{3,2}, H_1) = s(T_{3,1}, H_2) = 792$$
$$s(K_4, \bar{H}_1) = 144; \quad s(K_4, \bar{H}_2) = 138.$$

The terms on the right side of the identity are 132 for both H_1 and H_2, but the terms on the left side are 144 and 138 for the two possible subgraphs. Thus the identity cannot be satisfied and we have a contradiction.

The fact that H_1 and H_2 cannot be built into a $(5,5,49)$ graph was previously proved by Thomason [23]. ∎

4. What is $R(5,5)$?

The effort required to bring the upper bound on $R(5,5)$ down to 49 was considerable, but still it is a long way from the best lower bound of 43. In this section we explain why we believe that the correct value is closer to the lower end of this range. In fact, together with Geoff Exoo, we make the following strong conjecture:

Conjecture 2. $R(5,5) = 43$. ∎

We further conjecture, though this time with Geoff's dissent, that the number of $(5,5,42)$-graphs is precisely 656.

The same set of 656 $(5,5,42)$-graphs, consisting of 328 graphs and their complements, was found by several paths. Firstly, we took a few known $(5,5,42)$-graphs found by Exoo, removed three vertices from them in all possible ways, then extended the resulting $(5,5,39)$-graphs back to $(5,5,42)$-graphs using a variation of the one-vertex extension algorithm given in [17]. This process was repeated until no further $(5,5,42)$-graphs were found.

Needless to say, we checked that none of these 656 graphs can be extended to $(5,5,43)$-graphs.

The second construction method was devised and coded by Geoff Exoo. Starting with a random graph on 30 vertices, edges are inserted or deleted using the simulated annealing rules until a $(5,5,30)$-graph is obtained. Then an extra vertex is appended randomly and the new graph adjusted in the same way to make a $(5,5,31)$-graph. This process is repeated until finally a $(5,5,42)$-graph is obtained. The search is very difficult, and at most several $(5,5,42)$-graphs per day are generated, but we ran it on many computers for a very long time, making 5812 $(5,5,42)$-graphs altogether. The result was that each of the 656 known $(5,5,42)$-graphs was constructed at least once, but no new graphs were found.

A third construction method, using a similar incremental structure but with tabu search instead of simulated annealing, constructed hundreds of $(5,5,42)$-graphs but none were new. A number of attempts to bias the search away from where the known graphs are were unsuccessful in finding anything new. Finally, more than one decade of cpu time was expended in searching the neighbourhoods of the known $(5,5,42)$-graphs, defined by the numbers of common edges or the size of common subgraphs. For example, 100 random 36-vertex subgraphs were formed and extended to 42 vertices in all possible ways, making over 65 million $(5,5,42)$-graphs that were all isomorphic to the known graphs.

The fact that several independent processes that start with a random graph repeatedly find only the known $(5,5,42)$-graphs leads us to strongly suspect that our collection of $(5,5,42)$-graphs is complete. It is not possible to put this belief on a quantitative level, but as a mere illustration suppose that there were in fact 658 $(5,5,42)$-graphs (one extra and its complement) and that Exoo's program generates $(5,5,42)$-graphs uniformly at random (an unlikely proposition). Then after 5812 trials our chance of not discovering the extra graphs is $(656/658)^{5812} \approx 2.5 \times 10^{-7}$.

We wish to encourage our readers to devise further heuristic searches for $(5,5,42)$-graphs, to support this evidence. In fact, we propose the construction of $(5,5,42)$-graphs as a challenging benchmark for heuristic search methods.

For completeness, we give some information on the known $(5,5,42)$-graphs, restricting our counts to those with fewer edges than their complements. Of these 328 graphs, 212 have trivial automorphism groups and the others have a single nontrivial involution without fixed points. The number of edges ranges from 423 to 430, with the number of graphs in each class being 1, 7, 29, 66, 89, 77, 43, and 16, respectively. (Note the bimodal nature of this distribution when the complements are included.) All the vertices have degrees between 19 and 22, inclusive. The graphs themselves are available from the authors.

All the isomorphism and automorphism computations required for this paper were performed by the first author's program `nauty` [13]. Distribution of tasks across a workstation network was performed with the help of `autoson` [14].

5. A proof that $R(4,6) \leq 41$

A summary of the history of bounds on $R(4,6)$ can be found in Table III. In this section we will show how the identities from Section 2 and some data from [17] imply that $R(4,6) \leq 41$.

First, some words about linear programming. The great majority of available linear programming codes employ floating point arithmetic and are subject to the usual questions of correctness and accuracy that inexact arithmetic implies. The linear programs that arise in our work are not exceptionally large, but often have properties (such as high-dimensional optimum facets) that give trouble to floating point codes. Some exact implementations are available, for example in the symbolic algebra package Maple [2], but they are quite slow in operation.

We have taken a hybrid approach to these problems, helped by the fact that there are usually exact solutions to our linear programs which are rational points with small common denominators. Firstly, the routine `E04MBF` from the `NAG` library [19] is called to obtain an approximate solution. Sometimes it is necessary to apply it to the dual

year	reference	lower	upper	comments
1965	Kalbfleisch [9]	30		cyclic graph, not presented
1965	Kalbfleisch [9]		47	bound announced, not derived
1966	Kalbfleisch [10]	34		cyclic graph
1967	Kalbfleisch [11]		46	edge counting
1968	Walker [24]		45	combinatorics & LP
1971	Walker [25]		44	combinatorics & LP
1993	Exoo [5]	35		simulated annealing
1994	McKay & Radziszowski [16]		43	$(4, 4)$-graph enumeration & LP
1995	McKay & Radziszowski		41	this paper

Table III. The history of bounds on $R(4, 6)$.

program, or to apply it repeatedly with different starting points. When tentative approximate feasible points in both the primal and dual programs are found, they are converted to rational points by guessing a common denominator (using continued fractions). These guessed feasible points are then tested for actual feasibility using the original inequalities and exact arithmetic. If this test succeeds, we have proven the optimality of the solution. To guard against gross errors, all linear program solutions were compared to the approximate solutions given by LINDO [22].

Note that strictly speaking we are dealing with integer linear programs, not rational linear programs. However, in our experience, it is rare for there not to be an integer feasible point with objective equal to the rounded value of the rational optimum. The exceptional cases have no importance that we know of, so we will not attempt to present them here.

We will now describe our approach, in terms of a linear program $LP(s, t, n)$ for an (s, t, n)-graph G. This is similar to, but more general than, linear programs we have defined previously [15, 16].

For convenience, for any graph X, define the functions $v(X) = s(K_1, X)$, $e(X) = s(K_2, X)$, $t(X) = s(K_3, X)$ and $p(X) = s(T_{2,1}, X)$. Then we can write cases $m = 2, 3$ of Theorem 2.2 as

$$\sum_{v \in VG} 2e(G_v^-) = \sum_{v \in VG} g_2(G_v^+, |VG|) \tag{$I2$}$$

and

$$\sum_{v \in VG} 3t(G_v^-) = \sum_{v \in VG} g_3(G_v^+, |VG|), \tag{$I3$}$$

where

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

Suspose we have bounds as follows:

(a) $d' \le n - R(s,t{-}1)$ and $d'' \ge R(s{-}1,t) - 1$.

(b) $e_1'(i) \le e(X) \le e_1''(i)$ for every $(s{-}1,t,i)$-graph X.

(c) $e_2'(i) \le e(X) \le e_2''(i)$ for every $(s,t{-}1,i)$-graph X.

(d) $t'(i,j) \le t(X) \le t''(i,j)$ for every $(s,t{-}1,i,j)$-graph X.

(e) $g_3'(i,j) \le g_3(X,n) \le g_3''(i,j)$ for every $(s{-}1,t,i,j)$-graph X.

The variables of $LP(s,t,n)$ are as follows.

(i) n_i is the number of vertices of G having degree i, for $d' \le i \le d''$.

(ii) $g_{i,j}$ is the number of vertices v of G such that $v(G_v^+) = i$ and $e(G_v^+) = j$, for $d' \le i \le d''$ and $e_1'(i) \le j \le e_1''(i)$.

(iii) $h_{i,j}$ is the number of vertices v of G such that $v(G_v^-) = i$ and $e(G_v^-) = j$, for $n - d'' - 1 \le i \le n - d' - 1$ and $e_2'(i) \le j \le e_2''(i)$.

The constraints of $LP(s,t,n)$ are as follows. In each case, the sums are taken over all values of the summation indices for which the summand exists.

(A) $\sum_i n_i = n$.

(B) $\sum_j g_{i,j} = n_i$, for $d' \le i \le d''$.

(C) $\sum_j h_{n-i-1,j} = n_i$, for $d' \le i \le d''$.

(D)

$$2\sum_{i,j} j h_{n-i-1,j} = \sum_{i,j} \bigl(i(n-2i)+2j\bigr) g_{i,j}.$$

(E')

$$3\sum_{i,j} t'(i,j) h_{n-i-1,j} \le \sum_{i,j} g_3''(i,j) g_{i,j}.$$

(E'')

$$3\sum_{i,j} t''(i,j) h_{n-i-1,j} \ge \sum_{i,j} g_3'(i,j) g_{i,j}.$$

The correctness of (A), (B) and (C) is clear from the interpretation of the variables. Equation (D) is just identity $(I2)$. The two inequalities (E') and (E'') are a consequence of identity $(I3)$, comparing lower bounds for one side against upper bounds for the other.

Let us apply our linear programs to show that there are no $(4,6,41)$-graphs. Since $R(3,6) = 18$ and $R(4,5) = 25$, we can take $d' = 16$ and $d'' = 17$. The $(3,6)$-graphs are known completely [18, 21], so we can find best values of e_1', e_1'', g' and g''. However, the values e_2', e_2'', t' and t'' depend on the $(4,5,23)$-graphs and $(4,5,24)$-graphs, of which our knowledge is incomplete. Hence we begin by constructing the linear programs

$LP(4,5,23)$ and $LP(4,5,24)$. Using the fact from [17] that $(4,5,24)$-graphs have at most 132 edges, we find the bounds $e_2'(23) = 98$, $e_2''(23) = 130$, $e_2'(24) = 109$ and $e_2''(24) = 132$. Bounds on $t(X)$ for $(4,5)$-graphs can be found in Table IV.

$n = 23$				$n = 24$			
e	t	e	t	e	t	e	t
98	90-98	115	106-153	109	112-112	126	130-168
99	88-104	116	109-154	110	111-117	127	133-169
100	87-109	117	113-156	111	110-121	128	135-170
101	85-113	118	116-157	112	110-125	129	138-172
102	83-117	119	119-158	113	109-128	130	142-173
103	82-121	120	122-160	114	108-131	131	146-174
104	82-125	121	126-161	115	108-135	132	176-176
105	84-128	122	129-162	116	107-138		
106	85-131	123	133-164	117	107-142		
107	87-134	124	138-165	118	109-145		
108	89-137	125	142-166	119	112-148		
109	91-141	126	147-168	120	114-151		
110	92-142	127	153-169	121	117-154		
111	94-144	128	159-170	122	119-156		
112	97-146	129	166-172	123	122-159		
113	100-149	130	172-173	124	125-162		
114	103-151			125	127-165		

Table IV. Bounds on the number of triangles in $(4,5,n,e)$-graphs.

Having the values in Table IV, we can construct $LP(4,6,41)$. It is infeasible, which demonstrates the following theorem.

Theorem 5.1. $R(4,6) \leq 41$. ∎

It is perhaps worth noting that exactly the same result is obtained without constraint (E''). This is also true of the lower bounds in Table IV, which are those needed for constraint (E'), but not for the upper bounds.

Unfortunately, the linear program $LP(4,6,40)$ has many feasible points, so the existence of a $(4,6,40)$-graph remains a possibility. However, we note that the result $R(4,6) \leq 40$ would follow if it was known that $(4,5,22)$-, $(4,5,23)$- and $(4,5,24)$-graphs had at least 93, 105 and 113 edges, respectively. These bounds are quite likely to hold, but we have not proved them.

Concerning the exact value of $R(4,6)$, we expect that the current lower bound of 35 is correct. However, our evidence for this is less persuasive than for our similar feelings about the conjecture that $R(5,5) = 43$. We have 30 $(4,6,34)$-graphs so far, produced by making modifications to some graphs provided by Exoo, and proved that there are no others sharing a 31-vertex induced subgraph with one of these 30. However we have not performed any major heuristic searches.

Finally, we give some information on the known $(4,6,34)$-graphs. Of these 30 graphs, 13 have trivial automorphism groups and the others have a single nontrivial involution with 8 fixed points. The number of edges ranges from 222 to 227, with the number of graphs in each class being 2, 4, 8, 10, 5, and 2, respectively. The graphs themselves are available from the authors.

References.

[1] H. L. Abbott, Some Problems in Combinatorial Analysis, Ph. D. Thesis, University of Alberta, Edmonton, 1965.

[2] B. W. Char, K. O. Geddes, G. H. Gonnet, M. B. Monagan and S. M. Watt, *MAPLE Reference Manual*, WATCOM Publications Limited, Waterloo, 1990.

[3] P. Erdős, L. Lovász and J. Spencer, Strong independence of graphcopy functions, in Graph Theory and Related Topics (J. A. Bondy and U. S. R. Murty, eds.) Academic Press, New York 1979, 165–172.

[4] G. Exoo, A lower bound for $R(5,5)$, *J. Graph Theory*, **13** (1989) 97–98.

[5] G. Exoo, On the Ramsey numbers $R(4,6)$, $R(5,6)$ and $R(3,12)$, *Ars Combinatoria*, **35** (1993) 1–10.

[6] G. R. Giraud, Une majoration du nombre de Ramsey binaire-bicolore en (5,5), *C. R. Acad. Sc. Paris*, **265**, (1967) 809–811.

[7] A. W. Goodman, On sets of acquaintances and strangers at any party, *Amer. Math. Monthly*, **66** (1959) 778–783.

[8] R. W. Irving, Contributions to Ramsey Theory, Ph. D. thesis, University of Glasgow, 1973.

[9] J. G. Kalbfleisch, Construction of special edge-chromatic graphs, *Canadian Math. Bulletin*, **8** (1965) 575–584.

[10] J. G. Kalbfleisch, Chromatic Graphs and Ramsey's Theorem, Ph.D. thesis, University of Waterloo, January 1966.

[11] J. G. Kalbfleisch, Upper bounds for some Ramsey numbers, *J. Combinatorial Theory*, **2** (1967) 35–42.

[12] B. Kocay, An extension of Kelly's lemma to spanning subgraphs, *Congressus Numerantium*, **31** (1981) 109–120.

[13] B. D. McKay, nauty users' guide (version 1.5), Technical report TR-CS-90-02, Computer Science Dept., Australian National University, 1990.

[14] B. D. McKay, Autoson, a distributed batch system for UNIX workstation networks (Version 1.3), *Technical Report* TR-CS-96-03, Computer Science Department, Australian National University, 1996; `http://cs.anu.edu.au/~bdm/autoson`.

[15] B. D. McKay and S. P. Radziszowski, A new upper bound for the Ramsey number $R(5,5)$, *Australasian J. Combinatorics*, **5** (1992) 13–20.

[16] B. D. McKay and S. P. Radziszowski, Linear programming in some Ramsey problems, *J. Combinatorial Theory, Ser. B*, **61** (1994) 125–132.

[17] B. D. McKay and S. P. Radziszowski, $R(4,5) = 25$, *J. Graph Theory*, **19** (1995) 309–322.

[18] B. D. McKay and K. M. Zhang, The value of the Ramsey number $R(3,8)$, *J. Graph Theory*, **16** (1992) 99–105.

[19] Numerical Algorithms Group, NAG Fortran Library, Oxford, 1993.

[20] S. P. Radziszowski, Small Ramsey numbers, *Electronic J. Combinatorics*, Dynamic Survey 1, 1994–6;
`http://ejc.math.gatech.edu:8080/Journal/Surveys/index.html`.

[21] S. P. Radziszowski and D. L. Kreher, On $R(3,k)$ Ramsey graphs: theoretical and computational results, *J. Combinatorial Math. and Combinatorial Computing*, **4** (1988) 37–52.

[22] L. E. Schrage, User's Manual for LINDO, The Scientific Press, Palo Alto, 1981.

[23] A. G. Thomason, On finite Ramsey numbers, *European J. Combinatorics*, **3** (1982) 263–273.

[24] K. Walker, Dichromatic graphs and Ramsey numbers, *J. Combinatorial Theory*, **5** (1968) 238–243.

[25] K. Walker, An upper bound for the Ramsey number $M(5,4)$, *J. Combinatorial Theory*, **11** (1971) 1–10.

[26] H. Whitney, The colouring of graphs, *Ann. Math.*, **33** (1932) 688–718.