The Rand and block distances of pairs of set partitions

Frank Ruskey[1⋆] and Jennifer Woodcock[1]

Dept. of Computer Science, University of Victoria, CANADA

Abstract. The *Rand distances* of two set partitions is the number of pairs $\{x, y\}$ such that there is a block in one partition containing both x and y, but x and y are in different blocks in the other partition. Let $R(n, k)$ denote the number of distinct (unordered) pairs of partitions of n that have Rand distance k. For fixed k we prove that $R(n, k)$ can be expressed as $\sum_j c_{k,j} \binom{n}{j} B_{n-j}$ where $c_{k,j}$ is a non-negative integer and B_n is a Bell number. For fixed k we prove that there is a constant K_n such that $R(n, \binom{n}{2} - k)$ can be expressed as a polynomial of degree $2k$ in n for all $n \geq K_n$. This polynomial is explicitly determined for $0 \leq k \leq 11$. The *block distance* of two set partitions is the number of elements that are not in common blocks. We give formulae and asymptotics based on $N(n)$, the number of pairs of partitions with no blocks in common. We develop an $O(n)$ algorithm for computing the block distance.

1 Introduction and Motivation

In statistics, particularly as it is applied to cluster analysis, it is sometimes useful to have a measure of the difference between two set partitions [4]. The Rand distance is one such measure, and was introduced in Rand [8]. In this paper we will initiate a combinatorial study of the properties of the Rand distance, taken over all unordered pairs of partitions of an n-set. We will also introduce another measure, which we call the block distance, and determine some of its properties. For example, we will determine an exact expression for the number of pairs of partitions that have no blocks in common. Furthermore, we will show how to compute the block distance efficiently.

The *Rand distance* of two set partitions is the number of unordered pairs $\{x, y\}$ such that there is a block in one partition containing both x and y, but x and y are in different blocks in the other partition. We use $\mathcal{R}(P, Q)$ to denote the Rand distance between two set partitions P and Q. For example, $\mathcal{R}(\{\{1, 2\}, \{3\}\}, \{\{1\}, \{2, 3\}\}) = 2$ (the pairs are $\{1, 2\}$ and $\{2, 3\}$) and $\mathcal{R}(\{\{1, 2, 3\}\}, \{\{1\}, \{2\}, \{3\}\}) = 3$ (the pairs are $\{1, 2\}$, $\{1, 3\}$, and $\{2, 3\}$). In general, if P and Q are partitions of an n-set, then $0 \leq \mathcal{R}(P, Q) \leq \binom{n}{2}$.

Let $R(n, k)$ be the number of distinct (unordered) pairs of partitions of an n-set that have Rand distance k. See Table 2 in Section 3.1. This table was computed from exhaustive computer listings of all partitions of $\{1, 2, \ldots, n\}$ up

⋆ Research supported in part by NSERC

to $n = 11$. The column sums are $\binom{B_n}{2}$. Note that the numbers for fixed n are not unimodal in general.

We define the *block distance* $\mathcal{B}(P, Q)$ between two partitions of n as the number of elements in the blocks that are not common to both P and Q. For example,

$$\mathcal{B}(\{\{1,2\},\{3\}\},\{\{1\},\{2\},\{3\}\}) = 2$$

since the only block that is common to both partitions is $\{3\}$ and there are 2 elements in the remaining blocks. By $B(n, k)$ we denote the number of pairs of partitions of n that have block distance k. See Table 1 in Section 2. The Rand distance can be cleverly computed using a linear number of arithmetic operations; see Filkov and Skiena [2] and we will show that the block distance is also efficiently computable.

Organizationally, we will finish this section by giving some background on set partitions. In the succeeding two sections, we discuss first the block distance and then the Rand distance. The focus is mainly on the elucidation of some enumerative results along with a clever $O(n)$ algorithm for computing the block distance.

1.1 Background on set partitions

A *partition* of a set S is collection of disjoint subsets of S, say $\{S_1, S_2, \cdots, S_k\}$ whose union is S. Each S_i is referred to as a *block*. The number of partitions of an n-set into k blocks is the Stirling number (of the second kind), which is denoted as $\left\{ {n \atop k} \right\}$. We use $[n]$ to denote $\{1, 2, \ldots, n\}$.

In the computer, partitions are usually represented by *restricted growth strings*. We assume that the blocks of a partition X are numbered $S_1, S_2, \ldots, S_k$ according to the size of the smallest element in each block. That is, S_1 contains 1, S_2 contains the smallest element not in S_1, and so on. Then the restricted growth string $r[1..n]$ of X is defined by taking $r[i]$ to be the distance of the block containing i. The Gray code algorithms for generating restricted growth strings developed in [9] and discussed in [5] were used to generate the numbers in Tables 1 and 2; as each string was generated the $O(n)$ algorithms for computing the Rand distance and the block distance were applied.

The n-th Bell number, B_n, is the total number of partitions of an n-set, irrespective of block size. Thus $B_n = \sum_k \left\{ {n \atop k} \right\}$. The exponential generating function (egf) of the Bell numbers is well-known (e.g., Stanley [10], pg. 34)) to be

$$B(z) = \sum_{n \geq 1} B_n \frac{z^n}{n!} = e^{e^z - 1}. \tag{1}$$

The number of pairs of partitions is $\binom{B_n}{2}$. For $n = 1, 2, 3, \ldots, 10$ these numbers are

$$0, 1, 10, 105, 1326, 20503, 384126, 8567730, 223587231, 6725042325.$$

They give the row sums in Tables 1 and 2.

We use several times a generalization of the fact that if $f(z) = \sum_{n\geq 0} f_n z^n/n!$ is the egf of a sequence f_n, then $zf(z)$ is the egf of the sequence nf_{n-1}. See Knuth, Graham, Patashnik [3], page 350. Furthermore, for $k \geq 0$,

$$z^k f(z) = \sum_{n\geq k} n(n-1)\cdots(n-k+1) f_{n-k} \frac{z^n}{n!}$$
$$= k! \sum_{n\geq 0} \binom{n}{k} f_{n-k} \frac{z^n}{n!}. \tag{2}$$

Thus $k!\binom{n}{k} f_{n-k}$ is the n-th coefficient of $z^k f(z)$.

2 The Block Distance

Recall that the *block distance* $\mathcal{B}(P, Q)$ of two partitions of n is the number of elements in the blocks that are *not* common to both P and Q, and that $B(n, k)$ is the number of pairs of partitions of n that have block distance k. See Table 1.

$n\backslash k$	2	3	4	5	6	7	8	9
2	1							
3	3	7						
4	12	28	65					
5	50	140	325	811				
6	225	700	1950	4866	12762			
7	1092	3675	11375	34062	89334	244588		
8	5684	20384	68250	227080	714672	1956704	5574956	
9	31572	119364	425880	1532790	5360040	17610336	50174604	148332645

Table 1. The values of $B(n, k)$ for $1 \leq k \leq n \leq 9$.

Let $N(n) = B(n, n)$; this is the number of unordered pairs of partitions that have no blocks in common. The numerical values of $N(n)$, for $0 \leq n \leq 10$, are

$$0, 0, 1, 7, 65, 811, 12762, 244588, 5574956, 148332645, 4538695461.$$

Determining $N(n)$ for $i = 1, \ldots, n$ is sufficient to determine $B(n, k)$ since, by direct combinatorial considerations,

$$B(n, k) = N(k)\binom{n}{k} B_{n-k}. \tag{3}$$

We also note that

$$\binom{B_n}{2} = \sum_{k=0}^{n} B(n, k) = \sum_{k=0}^{n} N(k)\binom{n}{k} B_{n-k}. \tag{4}$$

Letting $N(z)$ be the egf of the $N(n)$ numbers, from (4) we obtain the equation

$$P(z) := \sum_{n\geq 0} \binom{B_n}{2} \frac{z^n}{n!} = N(z)e^{e^z-1}.$$

And thus

$$N(z) = P(z)e^{1-e^z}. \tag{5}$$

The egf e^{1-e^z} is known; it is the egf of the "complementary Bell numbers" (OEIS A000587). The complementary Bell numbers, C_n, for $n = 0, 1, 2, \ldots, 14$ are

$$1, -1, 0, 1, 1, -2, -9, -9, 50, 267, 413, -2180, -17731, -50533, 110176.$$

It is known that

$$C_n = \sum_{k=0}^{n} (-1)^k \left\{ {n \atop k} \right\}.$$

Thus, from (5) we get a "closed-form" formula for $N(n)$, namely

$$N(n) = \sum_{j=0}^{n} \binom{n}{j} C_j \binom{B_{n-j}}{2} = \sum_{j=0}^{n} \binom{n}{j} \binom{B_{n-j}}{2} \sum_{k=0}^{j} (-1)^k \left\{ {j \atop k} \right\}.$$

2.1 Linear Time Algorithm to Compute the Block Distance

In this subsection we present a linear time algorithm to compute the block distance of two partitions.

Closely related to the restricted growth string, we define the *block string*, $b[1..n]$, of P as follows: $b[i]$ is the smallest element in the block containing i. Every block string has the characterizing property that $b[1] = 1$, and for $i > 1$,

$$b[i] \in \{i,\ b[1], b[2], \ldots, b[i-1]\}.$$

It is relatively simple to convert a restricted growth string into the corresponding block string in $O(n)$ time.

The following code takes as input a restricted growth function $r[1..n]$ and returns the corresponding block string $b[1..n]$. It uses a temporary array $m[1..n]$ that maintains the invariant $b[i] = m[r[i]]$.

for $i \in \{1, 2, \ldots, n\}$ **do** $m[i] := 0$;
for $i := 1, 2, \ldots, n$ **do**
  if $m[r[i]] = 0$ **then** $m[r[i]] := i$;
  $b[i] := m[r[i]]$;

Before describing the algorithm for computing the block distance, we encourage the reader to consider the following small example. Suppose

$$P = \{1\}\{2\}\{3,4\}\{5,7\}\{6\}, \qquad Q = \{1,2\}\{3,4,6\}\{5,7\}.$$

Then the restricted growth strings for P and Q are

$$rP = 1, 2, 3, 3, 4, 5, 4, \qquad rQ = 1, 1, 2, 2, 3, 2, 3$$

and the block strings are

$$p = 1, 2, 3, 3, 5, 6, 5, \qquad q = 1, 1, 3, 3, 5, 3, 5.$$

Comparing the elements, we find that the blocks labelled 1, 2, 3, and 6 are not common to P and Q and that the block labelled 5 is common to P and Q. Since there are 5 elements in blocks 1, 2, 3, and 6, the block distance of P and Q is 5.

The algorithm maintains a boolean array $C[1..n]$ with the property that, upon termination, $C[i]$ is true if i is in a block common to P and Q, and is false otherwise. The block distance is thus equal to the number of entries in this array that are false.

The algorithm makes two passes over p, one pass over q, and one pass over C. Consider $p[i]$ and $q[i]$; there are three mutually exclusive cases: (a) $p[i] \neq q[i]$ and i is not in a common block, (b) $p[i] = q[i]$ and i is in a common block, and (c) $p[i] = q[i]$ and i is not in a common block. (Because we are using the block string and not the restricted growth string, it is not possible that $p[i] \neq q[i]$ and i is in a common block.) In the first pass we test only for case (a). In the second pass we (indirectly) distinguish cases (b) and (c).

The key observation is this: If i is *not* in a common block and $p[i] = q[i]$, then there is some value $j \neq i$ such that j is in the same block as i in P but is in a different block than i in Q, or vice-versa. In other words, $p[i] = p[j] \neq q[j]$ or $p[j] \neq q[j] = q[i]$. Thus, in the first pass $C[p[j]]$ and $C[q[j]]$ were set to false.

So on the second pass, we test whether $C[p[i]]$ is false to determine whether i is in a common block or not. On the final pass, we find the block distance by counting the number of false values in C. Below is the code in detail.

```
for i ∈ {1, 2, ..., n} do C[i] := true
for i := 1, 2, ..., n do
    if p[i] ≠ q[i] then C[p[i]] := C[q[i]] := false;
for i := 1, 2, ..., n do
    if ¬C[p[i]] then C[i] := false;
c := 0;
for i ∈ {1, 2, ..., n} do
    if ¬C[i] then c := c + 1;
return(c);
```

3 Results on the Rand distance

Now recall that $\mathcal{R}(P, Q)$ is the number of unordered pairs $\{x, y\}$ such that there is a block in one partition containing both x and y, but x and y are in different blocks in the other partition, and that $R(n, k)$ is the number of distinct (unordered) pairs of partitions of an n-set that have Rand distance k. See Table 2. Let $R(n)$ be the sum of the Rand distance over all unordered pairs of partitions.

$k\backslash n$	2	3	4	5	6	7	8	9	10	11
1	1	3	12	50	225	1092	5684	31572	186300	1163085
2		6	30	150	780	4200	23772	141624	887220	5835060
3		1	32	280	1720	10885	69272	452508	3060360	21482340
4			24	300	3360	25200	183960	1341648	9883440	74471760
5			6	240	3426	42672	391356	3266172	26969040	222185304
6			1	220	4100	56889	696178	7234374	67288830	612903720
7				60	2400	60165	941088	12259368	141778440	1469224350
8				15	2700	57750	1182888	18992502	256463820	3164268690
9				10	1075	46585	1150520	23324140	399874640	5762811670
10				1	471	31374	1165416	28129626	547907454	9538994388
11					150	24528	815640	26605908	670419540	13513772745
12					35	14140	780570	26190612	742419510	18112131840
13					45	4725	413840	21568932	744780330	20675910420
14					15	1890	369180	17119818	701747010	23653643310
15					1	1302	178080	13040280	607809750	22677991578
16						252	115780	8948079	520591950	22923998460
17						210	43512	6244308	377521875	19287053775
18						140	20734	3679032	312082260	17554312490
19						105	6860	2431044	198307620	13495597225
20						21	7098	1250109	158606532	11143736604
21						1	3508	640908	87210930	8029798920
22							574	315828	63688410	6035010960
23							840	197568	33243120	4254456690
24							665	57288	25703205	2872892550
25							476	46116	11343906	1924619235
26							210	30366	6764940	1215058680
27							28	25732	3272500	789847190
28							1	7695	2003805	453548480
29								4104	1532340	306871290
30								2226	757080	177358500
31								3780	211410	112440900
32								2205	212625	53211510
33								1344	198345	35497935
34								378	138600	16793040
35								36	82512	13781493
36								1	21080	10664335
37									16200	6744100
38									15750	2483415
39									14910	1445565
40									13545	802164
41									7245	1320165
42									3270	860640
43									630	580965
44									45	215325
45									1	104313
46										62205
47										103950
48										70455
49										74250
50										45045
51										21945
52										7095
53										990
54										55
55										1

Table 2. The values of $R(n,k)$ for $2 \le n \le 11$ and $1 \le k \le 55$. This table is inverted in the sense that k increases down columns and n varies along the columns.

Theorem 1.

$$R(n) = \sum_{k=0}^{\binom{n}{2}} k \; R(n,k) = \binom{n}{2} B_{n-1}(B_n - B_{n-1}).$$

Proof. Choose a pair $\{x, y\}$. The number of partitions in which this pair appears in the same block is B_{n-1}. The number of partitions in which this pair appears in different blocks is the difference $B_n - B_{n-1}$. Thus in total, each pair contributes $B_{n-1}(B_n - B_{n-1})$ to the sum. Since there are $\binom{n}{2}$ ways to choose a pair, the proof is finished. □

The average value of the Rand distance is thus

$$\frac{R(n)}{\binom{B_n}{2}} = \frac{n(n-1)B_{n-1}(B_n - B_{n-1})}{B_n(B_n - 1)}.$$

Since the Bell numbers grow exponentially,

$$\frac{R(n)}{\binom{B_n}{2}} \sim n^2 \frac{B_{n-1}}{B_n},$$

which experimentally appears to be $\Theta(n \log n)$.

3.1 Determining $R(n,k)$ for small values of k

We now consider $R(n,k)$ for small values of k.

Clearly $R(n, 0) = 0$.

Theorem 2. *For all* $n \geq 1$,

$$R(n, 1) = \binom{n}{2} B_{n-2}.$$

Proof. The only way that the Rand distance can be 1 is if there is a block $\{x, y\}$ in one partition and two blocks $\{x\}, \{y\}$ in the other, and all other blocks in one partition are present in the other. There are $\binom{n}{2}$ ways to choose the pair and B_{n-2} ways to determine the other blocks. □

Corollary 1. *The egf of the* $R(n,1)$ *numbers is*

$$\sum_{n\geq 1} R(n,1)\frac{z^n}{n!} \;=\; \frac{z^2}{2}B(z) \;=\; \frac{z^2}{2}e^{e^z-1}.$$

Proof. Apply (2) with $k = 2$. □

The previous two results were warm-ups for the more technical results that follow.

Theorem 3. *For fixed k, there are non-negative integer constants $c_{k,j}$ such that, for all $n \geq 1$,*

$$R(n,k) = \sum_{j=\lceil (1+\sqrt{1+8k})/2 \rceil}^{2k} c_{k,j} \binom{n}{j} B_{n-j}.$$

Proof. Any two partitions P and Q will have a largest subpartition X that is common to both P and Q. Thus $\mathcal{R}(P,Q) = \mathcal{R}(P \setminus X, Q \setminus X)$.

In the sum above j represents $n - |X|$, given that $\mathcal{R}(P,Q) = k$. The lower bound in the summation follows from the fact that the maximum Rand distance between two partitions of n is $\binom{n}{2}$ and thus $k \leq \binom{j}{2}$. Solving the implied quadratic yields $j \geq (1 + \sqrt{1+8k})/2$, which gives us the lower bound. We hereafter use $\alpha = \lceil (1 + \sqrt{1+8k})/2 \rceil$ for ease of reading.

For the upper bound, consider two partitions P and Q of an j-set that have no block in common, and have Rand distance k. We claim that $\geq \lceil j/2 \rceil$. Consider some arbitrary integer $x \in \{1, 2, \ldots, j\}$. Since P and Q have no common blocks, there is some integer y that is in the same block as x in one partition, and in another block in the other partition. Thus we have j distinct ordered pairs (x, y), one for each different value of x. At least $\lceil j/2 \rceil$ of them have to be distinct as unordered pairs, and each such unordered pair contributes 1 to the Rand distance. Thus $k \geq \lceil j/2 \rceil$ as claimed. From this it follows that $j \leq 2k$, which is the upper bound in the sum above. □

Theorem 4. *For all $n \geq 1$,*

$$R(n,2) = 6\binom{n}{3}B_{n-3} + 6\binom{n}{4}B_{n-4}.$$

Proof. Theorem 3 tells us that

$$R(n,2) = c_{2,3}\binom{n}{3}B_{n-3} + c_{2,4}\binom{n}{4}B_{n-4}.$$

From the $k = 2$ row of Table 2 we then have the following two equations.

$$R(3,2) = 6 = c_{2,3}\binom{3}{3}B_0 + c_{2,4}\binom{3}{4}B_{-1} = c_{2,3} \quad \text{and}$$

$$R(4,2) = 30 = c_{2,3}\binom{4}{3}B_1 + c_{2,4}\binom{4}{4}B_0 = c_{2,3}4 + c_{2,4}.$$

This system of equations can be solved to obtain $c_{2,3} = c_{2,4} = 6$. □

Corollary 2. *The egf of the $R(n,2)$ numbers is*

$$\sum_{n \geq 1} R(n,2)\frac{z^n}{n!} \;=\; \left(z^3 + \frac{z^4}{4}\right)B(z) \;=\; \left(z^3 + \frac{z^4}{4}\right)e^{e^z - 1}.$$

In a similar fashion we can solve systems of linear equations to obtain the following theorems and corollaries.

Theorem 5. *For all* $n \geq 1$,

$$R(n,3) = \binom{n}{3} B_{n-3} + 28\binom{n}{4} B_{n-4} + 120\binom{n}{5} B_{n-5} + 60\binom{n}{6} B_{n-6}.$$

Corollary 3. *The egf of the* $R(n,3)$ *numbers is*

$$\sum_{n\geq 1} R(n,3)\frac{z^n}{n!} = \left(\frac{z^3}{6} + \frac{7z^4}{6} + z^5 + \frac{z^6}{12}\right) B(z) = \left(\frac{z^3}{6} + \frac{7z^4}{6} + z^5 + \frac{z^6}{12}\right) e^{e^z-1}.$$

Theorem 6. *For all* $n \geq 1$, *the value of* $R(n,4)$ *is*

$$24\binom{n}{4} B_{n-4} + 180\binom{n}{5} B_{n-5} + 1560\binom{n}{6} B_{n-6} + 2520\binom{n}{7} B_{n-7} + 840\binom{n}{8} B_{n-8}.$$

Corollary 4. *The egf of the* $R(n,4)$ *numbers is*

$$\sum_{n\geq 1} R(n,4)\frac{z^n}{n!} = \left(z^4 + \frac{3z^5}{2} + \frac{13z^6}{6} + \frac{z^7}{2} + \frac{z^8}{48}\right) B(z).$$

Intuitively, $c_{k,j}$ is the number of pairs of j-element set partitions with no common blocks that have Rand Distance k. We summarize the known values of $c_{k,j}$ in Table 3. Although we don't know the value of $c_{k,j}$ in general, we can determine a few specific infinite sequences, which are given in the next lemma.

Lemma 1. *For all* $k \geq 1$,

$$c_{k,2k} = \frac{(2k-1)!}{(k-1)!} \qquad \textit{and} \qquad c_{k,\alpha} = R(\alpha, k).$$

Proof. For a pair of $2k$ element set partitions P and Q to have Rand distance k with no common blocks, the $2k$ elements must be paired, and each pair of elements is a block in either P or Q. Further, if $\{a,b\}$ is a block in set P then set B contains the singleton blocks $\{a\}$ and $\{b\}$ and vice versa. Since the order of the blocks doesn't matter, we can assume the blocks (pairs) are sorted by their smallest elements. So, for $i = 1, 2, \ldots, k$, once we have chosen the elements for blocks $1, 2, \ldots, i-1$, the first element in block i must be the smallest remaining element and there are $2k - (2(i-1)+1) = 2k - 2i + 1$ choices for the second element in block i. Thus the number of ways to pair the elements is

$$\prod_{i=1}^{k} (2k - 2i + 1) = \frac{(2k-1)!}{2^{k-1}(k-1)!}$$

If we assume, without loss of generality, that a pair, say $\{a,b\}$, is in partition P, then there are 2^{k-1} unique ways to distribute the remaining pairs between P and Q. So we have

$$c_{j,2k} = \frac{(2k-1)!}{2^{k-1}(k-1)!} 2^{k-1} = \frac{(2k-1)!}{(k-1)!}.$$

$k\backslash j$	2	3	4	5	6	7	8	9	10	11
1	**1**									
2		**6**	6							
3		**1**	28	120	60					
4			**24**	180	1560	2520	840			
5			**6**	210	1986	18900	63840	60480	15120	
6			**1**	215	2780	28224	253246	1340640	2520000	1663200
7				**60**	2040	43365	463128	3998736	26878320	82328400
8				**15**	2610	38850	721728	8575200	74028240	554843520
9				**10**	1015	39060	778400	13061020	172444150	1568364600
10				**1**	465	28077	914480	17680572	270474480	3714220092
11					**150**	23478	619416	19277748	407335320	6281694045
12					**35**	13895	667450	19168422	482217540	10078945140
13					**45**	4410	376040	17848152	529667460	12553128060
14					**15**	1785	354060	13798458	530778780	15995950740
15					**1**	1295	167664	11437644	477563400	16021896264
16						**252**	113764	7906059	431141400	17216673870
17						**210**	41832	5852700	315103995	15141561930
18						**140**	19614	3492426	275308740	14124874940
19						**105**	6020	2369304	174009780	11315379955
20						**21**	6930	1186227	146107962	9400242852
21						**1**	3500	609336	80801970	7071057840
22							**574**	310662	60530130	5334533160
23							**840**	190008	31267440	3888920970
24							**665**	51303	25130325	2590267020
25							**476**	41832	10882746	1799914809
26							**210**	28476	6461280	1140678990
27							**28**	25480	3015180	753854310
28							**1**	7686	1926855	431506790
29								**4104**	1491300	290015550
30								**2226**	734820	169030620
31								**3780**	173610	110115390
32								**2205**	190575	50872635
33								**1344**	184905	33316140
34								**378**	134820	15268440
35								**36**	82152	12873861
36								**1**	21070	10432455
37									**16200**	6565900
38									**15750**	2310165
39									**14910**	1281555
40									**13545**	653169
41									**7245**	1240470
42									**3270**	824670
43									**630**	574035
44									**45**	214830
45									**1**	104302
46										**62205**
47										**103950**
48										**70455**
49										**74250**
50										**45045**
51										**21945**
52										**7095**
53										**990**
54										**55**
55										**1**

Table 3. Known values of $c_{k,j}$ for $2 \le j \le 11$. The bold value at the beginning of each row is $c_{k,\alpha} = R(\alpha, k)$.

Since $B_i = 0$ when $i < 0$, $B_0 = 1$, and $\binom{i}{i} = 1$,

$$R(\alpha, k) = \sum_{j=\alpha}^{2k} c_{k,j} \binom{\alpha}{j} B_{\alpha-j} = c_{k,\alpha}.$$

□

Lemma 2. *For all* $j \geq 1$*,*

$$c_{\binom{j}{2},j} = R\left(j, \binom{j}{2}\right) = 1.$$

For all $j \geq 4$*:*

$$c_{\binom{j}{2}-1,j} = R\left(j, \binom{j}{2} - 1\right) = \binom{j}{2}.$$

For all $j \geq 5$*:*

$$c_{\binom{j}{2}-2,j} = R\left(j, \binom{j}{2} - 2\right) = \binom{\binom{j-1}{2}}{2}.$$

For all $j \geq 2 + x$*:*

$$c_{\binom{j}{2}-x,j} = R\left(j, \binom{j}{2} - x\right).$$

Proof. Omitted in this extended abstract. □

3.2 The numbers $R(n, \binom{n}{2} - k)$ for small k

We now consider the numbers at the bottom of the columns in Table 2. Clearly $R(n, \binom{n}{2}) = 1$ (the pair is $\{1, 2 \ldots n\}$ and $\{1\}\{2\} \ldots \{n\}$).

Theorem 7. *For all* $n \geq 4$*,*

$$R(n, \binom{n}{2} - 1) = \binom{n}{2}, \text{ and } R(3, 2) = 6.$$

Proof. For $n \geq 4$, the two partitions are the full set $\{1, 2, \ldots, n\}$ and the partition consisting of one pair and $n - 2$ singleton sets. □

Theorem 8. *For all* $n \geq 5$*,*

$$R(n, \binom{n}{2} - 2) = \binom{\binom{n-1}{2}}{2} = \frac{1}{8}n(n-1)(n-2)(n-3),$$

and $R(3, 3) = 3$*,* $R(4, 5) = 24$*.*

Proof. For $n \geq 5$ the two partitions are the full set $\{1, 2, \ldots, n\}$ and the partition consisting of two pairs and $n - 4$ singleton sets. The order of the two pairs does not matter so we have

$$R(n, \binom{n}{2} - 2) = \frac{1}{2}\binom{n}{2}\binom{n-2}{2},$$

which can be shown to be equal to the two values given in the statement of the theorem. □

The numbers in Theorems 7 and 8 are a shifted versions of OEIS A000217 and OEIS A050534, respectively.

Theorem 9. *For all $n \geq 6$,*

$$R(n, \binom{n}{2} - 3) = \frac{1}{6}\binom{n}{2}\binom{n-2}{2}\binom{n-4}{2} + \binom{n}{3},$$

and $R(4, 3) = 32$, $R(5, 7) = 60$.

Proof. For $n \geq 5$ the two partitions are either the full set $\{1, 2, \ldots, n\}$ and the partition consisting of three pairs and $n - 6$ singleton sets, or the full set $\{1, 2, \ldots, n\}$ and the partition consisting of one triple and $n - 3$ singleton sets. □

Theorem 10. *For fixed k there is a constant K_k such that $R(n, \binom{n}{2} - k)$ is a polynomial of degree $2k$ in n for all $n \geq K_k$.*

Proof. Omitted in this extended abstract. □

4 Acknowledgements

We wish to thank Rod Canfield for helpful discussions.

References

1. E.R. Canfield, *Engel's inequality for the Bell numbers*, Journal of Combinatorial Theory, Series A, 72 (1995) 184–187.
2. V. Filkov and S. Skiena, *Integrating Microarray Data by Consensus Clustering*, International Journal on Artificial Intelligence Tools, 13(4): 863–880, 2004.
3. R.L. Graham, D.E. Knuth, and O. Patashnik, *Concrete Mathematics*, Addison-Wesley, 1989.
4. L. Hubert, *Comparing Partitions*, Journal of Classification, 2 (1985) 193–218.
5. D.E. Knuth, *The Art of Computer Programming, Volume 4: Combinatorial Algorithms, Part 1*, Addison-Wesley, Reading, MA, 2011.
6. L. Moser and A. Wyman, *An asymptotic formula for the Bell numbers*, Transactions of the Royal Society of Canada III, 49 (1955) 49–54.

7. A. O. Munagi, *Set Partitions with Successions and Separations*, Int. J. Math and Math. Sc. 2005, no. 3 (2005), 451–463.
8. W. Rand, *Objective criteria for the evaluation of clustering methods*, J. American Statistical Assoc., 66 (336): 846–850, 1971.
9. F. Ruskey, *Simple combinatorial Gray codes constructed by reversing sublists*, 4th ISAAC (International Symposium on Algorithms and Computation), Lecture Notes in Computer Science (LNCS), #762 (1993) 201–208.
10. Richard R. Stanley, *Enumerative Combinatorics, Volume 1* Wadsworth, 1986.