STEINHAUS GRAPHS AND PENDENT VERTICES

JOSIAH DAVIS

1. Introduction

What is a Steinhaus graph? First of all, when we speak of a *graph*, we are not referring to a mapping of points in the xy plane. Rather, we are referring to a set of *vertices* connected by *edges*. One of the common ways to represent a graph is pictorially, as in Figure 1. The numbered circles are vertices and the lines between them are edges. While a picture is nice to visualize a graph, what we call the *adjacency matrix* is more useful in definitions and theorems. The adjacency matrix of the graph in Figure 1 is shown in Figure 2.

Each vertex has its own row and column in the adjacency matrix. An edge exists between vertex i and vertex j if there is a one in the i^{th} row of the j^{th} column. Since the graphs that we are working with are undirected (that is, if vertex i is connected to vertex j, then vertex j is connected to vertex i), the adjacency matrix is symmetric about the diagonal. Because of this symmetry, we often symplify the notation for the adjacency matrix by keeping only the upper right triangle.

It is time to define a *Steinhaus graph* precisely. Let $T = a_{0,0}a_{0,1}...a_{0,n-1}$ be an n-long string of 0s and 1s beginning with 0. The *Steinhaus graph* generated by T has as its adjacency matrix the *Steinhaus matrix* $A = [a_{i,j}]$, where

$$a_{i,j} = \begin{cases} 0, & \text{if } 0 \le i = j \le n-1; \\ (a_{i-1,j-1} + a_{i-1,j})\text{mod}2, & \text{if } 0 < i < j \le n-1; \\ a_{j,i}, & \text{if } 0 \le j < i \le n-1. \end{cases}$$

More simply, a *Steinhaus graph* is a graph where the upper right triangle has the property that every entry is the boolean "Exclusive OR" result of the entry above it and the entry above and to the left of it. (Exception: the entries along the diagonal are always zero.) For instance, the adjacency matrix in Figure 2 adheres to this rule, so the corresponding graph is a Steinhaus graph.

A *pendent vertex* is a vertex with degree one, or a vertex that is incident to only one edge. For example, vertex 4 in Figure 1 is a pendent vertex, but none of the other vertices are pendent.

Steinhaus graphs have many interesting properties, yet there are many things about them that are not yet known. In [1], a formula was discovered for the total number of Steinhaus graphs on n vertices with at least one pendent vertex. Our research goal was to try to further characterize this result. Can we describe the the number of Steinhaus graphs on n vertices with exactly k pendent vertices? Let $P(n, k)$ be the number of Steinhaus graphs on n vertices with k pendent vertices. Our task was to find an explicit formula for it.

2. The Data

First, we needed to generate the data to analyze. This consisted of constructing a Java program to build the $P(n, k)$ table iteratively, starting at $P(0, 0)$ and expanding outward along both axes. The final version of the program was able to compute 325 rows of the table. (We could in fact compute more rows, but it would take longer and longer.) We had an independently computed data set that verified the first 80 rows of the table, so we have confidence that the algorithm for generating the table is correct. Figure 3 displays a portion of the $P(n, k)$ table.

The next phase of research was to analyze the patterns found in this table. The first thing we noticed is that the entries in the table separate into clearly visible bands. For instance, the line of 2's going down the diagonal is one such band, and another is the 2, 4, 2 pattern going down diagonally underneath the 2's. Furthermore, for a given column, the entries in one band are symmetric about the highest value (or the two highest in the case of a tie). For instance, 2, 4, 2 is symmetric about 4. Our efforts in the first phase of research were dedicated towards describing these bands formally.

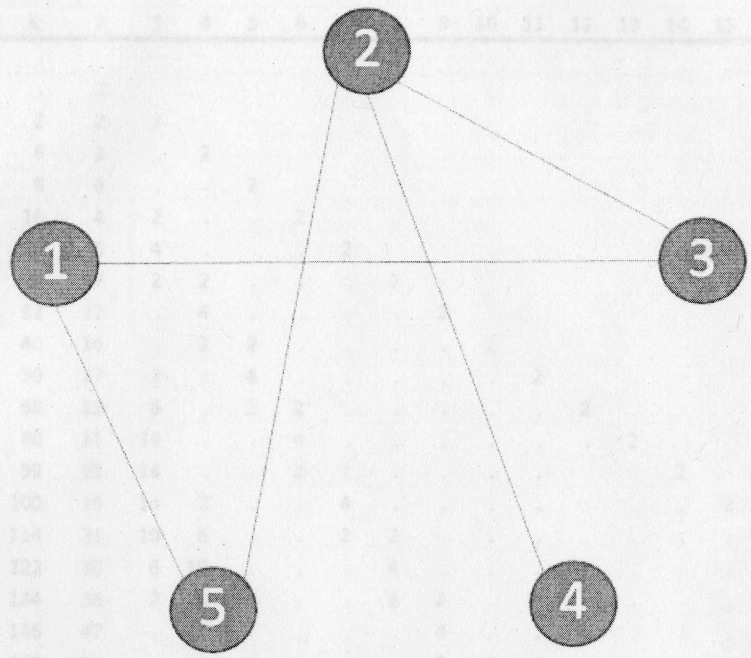

FIGURE 1. A Graph

Adjacency Matrix

	1	2	3	4	5
1	0	0	1	0	1
2	0	0	1	1	1
3	1	1	0	0	0
4	0	1	0	0	0
5	1	1	0	0	0

Adjacency Triangle

1	2	3	4	5
0	0	1	0	1
	0	1	1	1
		0	0	0
			0	0
				0

FIGURE 2. Adjacency Matrix and Adjacency Triangle

We decided to index the bands so that band 1 is the topmost, band 2 is the next highest, and so on. We wanted to discover the explicit pattern that the bands follow. Figure 4 shows a table with the values for the first 5 bands.

First of all, each band is unimodal; that is, the values increase steadily to a point, and then only decrease. We decided to do some differencing between successive terms to shed light on the pattern. Figure 5 shows a table of double differencing on the Fourth band. (The other bands yielded similar tables.)

Here it is easy to see that the first order differences are unimodal. From this table we concluded that each band divided naturally into 5 segments, or stages, based on the 5 different segments in the column of second order differences. Note that the zeros in the second order differences mean that the function is linear in n for those stages. Similarly, the other constants in the second order differences mean that the other stages are quadratic in n. We took points from these stages and were able to fit a quadratic to the entire stage. The five sections are based on the value J which is equal to 2^{j-2}. Here are the sections in terms of J:

(1) $1 \leq n < J$
(2) $J \leq n < 2J$
(3) $2J \leq n < 3J$
(4) $3J \leq n < 4J$
(5) $4J \leq n < 5J - 1$

Using this information regarding the sections of the band, we then defined a piecewise function $B(n, j)$ to return the n^{th} value from the start of the j^{th} band. That formula is desribed explicitly by the following:

	k:	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
n:	.																
3	.	3															
4	2	2	2														
5	6	3	.	2													
6	8	6	.	.	2												
7	16	4	2	.	.	2											
8	20	5	4	.	.	.	2										
9	28	7	2	2	.	.	.	2									
10	32	12	.	4	.	.	.	.	2								
11	40	16	.	2	2	.	.	.	.	2							
12	50	17	2	.	4	.	.	.	.	.	2						
13	68	12	6	.	2	2	.	.	.	.	.	2					
14	80	11	10	.	.	4	.	.	.	.	.	.	2				
15	92	11	14	.	.	2	2	.	.	.	.	.	.	2			
16	100	16	14	2	.	.	4	.	.	.	.	.	.	.	2		
17	114	21	10	6	.	.	2	2	.	.	.	.	.	.	.	2	
18	122	30	6	10	.	.	.	4	.	.	.	.	.	.	.	.	2
19	134	38	2	14	.	.	.	2	2	.	.	.	.	.	.	.	.
20	146	47	.	14	2	.	.	.	4	.	.	.	.	.	.	.	.
21	164	54	.	10	6	.	.	.	2	2	.	.	.	.	.	.	.
22	182	59	2	6	10	.	.	.	.	4	.	.	.	.	.	.	.
23	206	59	6	2	14	.	.	.	.	2	2	.	.	.	.	.	.
24	232	56	12	.	14	2	.	.	.	.	4	.	.	.	.	.	.
25	268	46	20	.	10	6	.	.	.	.	2	2	.	.	.	.	.
26	296	41	28	.	6	10	.	.	.	.	.	4	.	.	.	.	.
27	324	37	36	.	2	14	.	.	.	.	.	2	2	.	.	.	.
28	348	36	44	.	.	14	2	.	.	.	.	.	4	.	.	.	.
29	374	35	52	.	.	10	6	.	.	.	.	.	2	2	.	.	.
30	394	40	56	2	.	6	10	.	.	.	.	.	.	4	.	.	.
31	414	48	56	6	.	2	14	.	.	.	.	.	.	2	2	.	.
32	430	61	52	12	.	.	14	2	.	.	.	.	.	.	4	.	.
33	454	73	44	20	.	.	10	6	.	.	.	.	.	.	2	2	.
34	470	90	36	28	.	.	6	10	.	.	.	.	.	.	.	4	.
35	490	106	28	36	.	.	2	14	.	.	.	.	.	.	.	2	2

FIGURE 3. A Section of the $P(n,k)$ Table ("."s indicate zeroes)

$$B(n,1) = \begin{cases} 2 & \text{if } n = 1 \\ 0 & \text{if } n \geq 2 \end{cases}$$

$$B(n,2) = \begin{cases} 2 & \text{if } n = 1 \\ 4 & \text{if } n = 2 \\ 2 & \text{if } n = 3 \\ 0 & \text{if } n \geq 4 \end{cases}$$

If $j \geq 3$, then

$$B(n,j) = \begin{cases} n^2 + n & \text{if } 1 \leq n < J \\ -J^2 + (2n+1)J & \text{if } J \leq n < 2J \\ -9J^2 + 5(2n+1)J - 2(n^2+n) & \text{if } 2J \leq n < 3J \\ 9J^2 - (2n+1)J & \text{if } 3J \leq n < 4J \\ 25J^2 - 5(2n+1)J + n^2 + n & \text{if } 4J \leq n < 5J - 1 \\ 0 & \text{if } n \geq 5J - 1 \end{cases}$$

Now that we have an empirical formula for the contents of the bands, we would like to know where each new band begins. The bands follow diagonal patterns down the $P(n,k)$ table, but the question is, given a particular column number k and a band number j, can we calculate where that band begins? We decided to assign a function to this concept, $F(k,j)$. We studied the data for several different instances of k and

j	j: 1	2	3	4	5
1	2	2	2	2	2
2		4	6	6	6
3		2	10	12	12
4			14	20	20
5			14	28	30
6			10	36	42
7			6	44	56
8			2	52	72
9				56	88
10				56	104
11				52	120
12				44	136
13				36	152
14				28	168
15				20	184
16				12	200
17				6	212
18				2	220
19					224
20					224
21					220
22					212
23					200
24					184
25					168
26					152
27					136
28					120
29					104
30					88
31					72
32					56

FIGURE 4. Bands of the $P(n, k)$ Table

n	j: 4	1st order difference	2nd order difference
1	2		
2	6	4	
3	12	6	2
4	20	8	2
5	28	8	0
6	36	8	0
7	44	8	0
8	52	8	0
9	56	4	-4
10	56	0	-4
11	52	-4	-4
12	44	-8	-4
13	36	-8	0
14	28	-8	0
15	20	-8	0
16	12	-8	0
17	6	-6	2
18	2	-4	2

FIGURE 5. Double-Differencing on the Fourth Band of the $P(n, k)$ Table

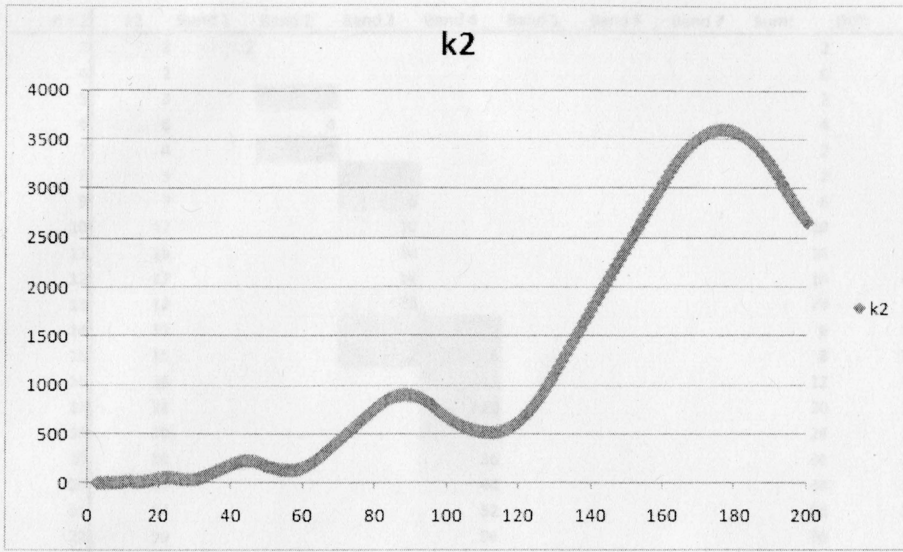

FIGURE 6. Graph of $P(n,2)$ vs. n

were able to come up with k-specific formulas. From there we were able to generalize to the following function, valid for for $k \geq 3$:

(1) $F(k,1) = k$
(2) $F(k,j) = 2^{(j-2)}(2k-1)+1$ (for $j \geq 2$).

Note that the function is designed so that $F(k,j)+1$ yields the n value at which the j^{th} band begins in the $P(n,k)$ table. (We included an offset of 1 so that the final formula would work out more nicely.)

Ultimately, we want to have a function $P(n,k)$ that takes an n value and a k value and produces the appropriate result. For $k > 2$, we know what the proper result is if we know which band we are in and how far into the band we are. $F(k,j)$ is designed so that $n - F(k,j)$ will tell how far into the band our number is. There is just one problem: we do not yet have a formula for computing the band number based on n and k. This was our next project, and we eventually came up with the following formula:

(1) $G(n,k) = 0$ if $n < k+1$
(2) $G(n,k) = 1$ if $n = k+1$
(3) $G(n,k) = \lfloor \log_2(\frac{n-2}{2k-1}) \rfloor + 2$ if $n > k+1$

Now we have all the pieces we need for $P(n,k)$ when $k > 2$. Putting it all together, we get the following formula in terms of B, F, and G:

$$P(n,k) = B(n - F(k, G(n,k), G(n,k))$$

In terms of B, n, and k only, the formula becomes

$$P(n,k) = B(n - 2^{\lfloor \log_2(\frac{n-2}{2k-1}) \rfloor}(2k-1)+1, \lfloor \log_2(\frac{n-2}{2k-1}) \rfloor + 2)$$

3. COLUMNS 1 AND 2

Now the question is, what happens when $k \leq 2$? The problem with these two columns is that no zeros ever appear. Thus, it is hard to tell where the different bands begin and end. One of the most exciting breakthroughs for us was when we conjectured that the first two columns were superpositions of multiple bands. Testing this out empirically came up with very intriguing results.

The dark blue line in Figure 6 is a graph of $P(n,2)$ vs. n, or, equivalently, the second column of the $P(n,k)$ table, which we will denote k_2. The first thing we noticed in the graph is that there are local maxima shaped roughly like bell curves. Furthermore, the value of these maxima increases with n, just like the bands. We wondered how similar these local maxima looked to the different bands, so we decided to superimpose the bands onto the graph of k_2. To do so, we took each peak in and matched it up with the largest value in the k^{th} band. The first two bands in the $P(n,k)$ table have only one value for the absolute

n = 2	k2	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7	Sum:	Diff:
3	3	2							2	1
4	2								0	2
5	3		2						2	1
6	6		4						4	2
7	4		2						2	2
8	5			2					2	3
9	7			6					6	1
10	12			10					10	2
11	16			14					14	2
12	17			14					14	3
13	12			10					10	2
14	11			6	2				8	3
15	11			2	6				8	3
16	16				12				12	4
17	21				20				20	1
18	30				28				28	2
19	38				36				36	2
20	47				44				44	3
21	54				52				52	2
22	59				56				56	3
23	59				56				56	3
24	56				52				52	4
25	46				44				44	2
26	41				36	2			38	3
27	37				28	6			34	3
28	36				20	12			32	4
29	35				12	20			32	3
30	40				6	30			36	4
31	48				2	42			44	4
32	61					56			56	5
33	73					72			72	1
34	90					88			88	2
35	106					104			104	2

FIGURE 7. The Splice Table

maximum, whereas all following bands have two values that tie for the maximum. Therefore, we decided to find the local maxima of $P(n,2)$ vs. n, and then keep track of the two highest values in each local maxima. In Figure 7, the local maxima (or the two highest values near a local maxima) are highlighted in yellow.

Figure 8 displays the graph of $P(n,2)$ vs. n, each band, then the sum of all of the bands together, which we have called a splice.

As you can see, the splice almost perfectly matches $P(n,2)$ vs. n and the two sequences only appear to diverge when the next band is unavailable. However, we could not really see the differences at this scale, so we made a separate chart just showing the differences.

First, the resulting error is small compared to the actual values of $P(n,k)$ which is very encouraging. Second, the sequence of error values is highly patterned. In fact, it turns out the sequence of error values can be described very concisely. We decided to call this sequence $D(n)$, and we noticed that $D(n)$ is equal to the number of 1's appearing in the binary expansion of n. For instance, all the powers of two yield a one, as you can see in Figure 9. Because powers of 2 are so important in Steinhaus graphs, it seems very certain that this pattern continues. We now have a way to describe $P(2,n)$ using the bands and this new sequence:

$$P(n,2) = \sum_{j=1}^{G(n,2)} B(n+1-F(2,j),j)) + D(n-1)$$

However, we needed to do more work to explicitly describe the starting point of each band in the splice. When we put the splice together originally, we looked for local maxima because those points were easily identifiable from the graph of $P(n,2)$ vs. n. In terms of defining a function, though, it is more helpful to know the start of each band. The function $F(k,j)$ which we defined earlier allows us to know the start of each band as long as $k \geq 3$, but we wanted to generalize the result further.

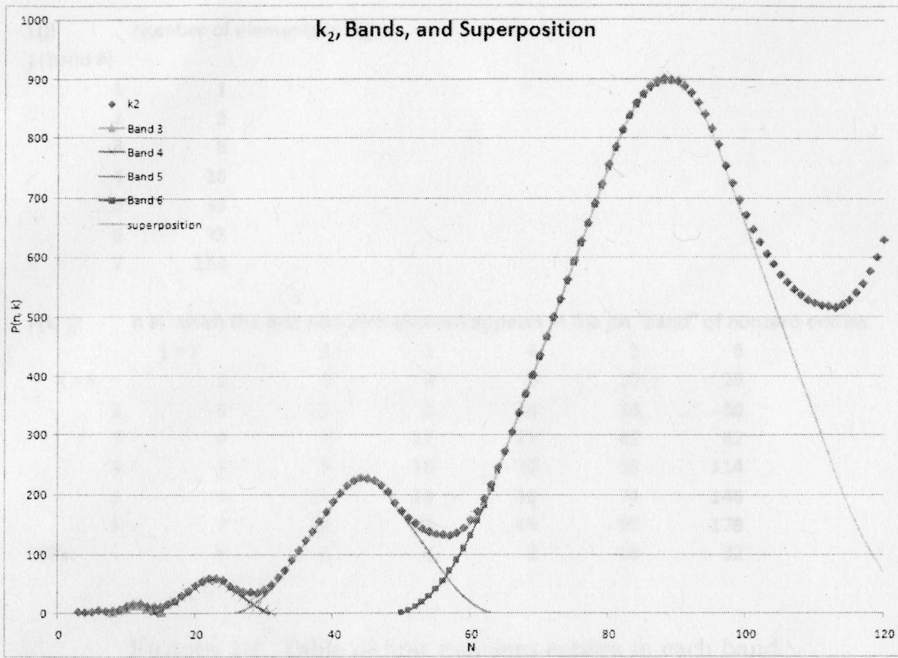

FIGURE 8. $P(n,2)$ vs. n with superimposed bands

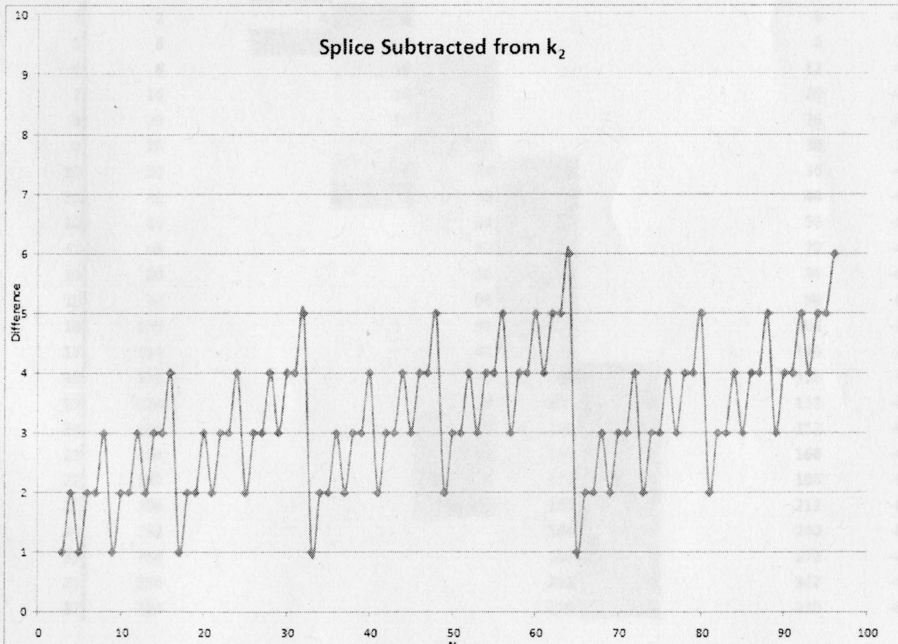

FIGURE 9. $P(n,2)$ vs. n - superimposed bands

Based on the positions of the bands in our k_2 splice table (see Figure 7), we discovered that the usual formula for $F(k,j)$ continues even when $k=2$; the start of the bands that we used in the splice coincides exactly with the values calculated using the formula for $F(k,j)$. Does the formula still work if $k=1$? We spliced the bands together as the formula dictated (see Figure 11), and the results were as shown in Figure 12.

Again, the actual sequence for $P(n,1)$ vs. n and the spliced values begin to diverge only when we run out of available bands. Figure 13 shows the sequence of differences ($P(n,1)$ vs. n - splice).

It turns out that this sequence is equal to $-2D(n)$, which is comforting because the same phenomenon seems to be at work as in $P(n,2)$ vs. n. With this information, we can now characterize $P(n,1)$, and combining this with the previous results, we can characterize $P(n,k)$ for all $n \geq 1$, for all $k \geq 1$.

L(j)	Number of elements in B(j)
j (band #)	
1	1
2	3
3	8
4	18
5	38
6	78
7	158

F(k, j): n at which the first non-zero element appears in the jth "band" of nonzero entries

	j = 1	2	3	4	5	6
k = 1	2	3	4	6	10	18
2	3	5	8	14	26	50
3	4	7	12	22	42	82
4	5	9	16	30	58	114
5	6	11	20	38	74	146
6	7	13	24	46	90	178
Diffs:	1	2	4	8	16	32

FIGURE 10. Table of first non-zero entries in each band

n = 1	k1	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7	Sum:	Diff:
2	0	2							2	-2
3	0		2						2	-2
4	2		4	2					6	-4
5	6		2	6					8	-2
6	8			10	2				12	-4
7	16			14	6				20	-4
8	20			14	12				26	-6
9	28			10	20				30	-2
10	32			6	28	2			36	-4
11	40			2	36	6			44	-4
12	50				44	12			56	-6
13	68				52	20			72	-4
14	80				56	30			86	-6
15	92				56	42			98	-6
16	100				52	56			108	-8
17	114				44	72			116	-2
18	122				36	88	2		126	-4
19	134				28	104	6		138	-4
20	146				20	120	12		152	-6
21	164				12	136	20		168	-4
22	182				6	152	30		188	-6
23	206				2	168	42		212	-6
24	232					184	56		240	-8
25	268					200	72		272	-4
26	296					212	90		302	-6
27	324					220	110		330	-6

FIGURE 11. Using the formula for $F(k, j)$ when $k = 1$ to determine how the bands are superimposed in k_1

$$P(n,k) = \begin{cases} \left(\sum_{j=1}^{G(n,k)} B(n - F(k,j), j)\right) - 2D(n-1) & \text{if } k = 1 \\ \left(\sum_{j=1}^{G(n,k)} B(n - F(k,j), j)\right) + D(n-1) & \text{if } k = 2 \\ B(n - F(k, G(n,k)), G(n,k)) & \text{if } k \geq 3 \end{cases}$$

4. Next Steps

Although empirically the formulas we have seem to work, we do not have concrete mathematical proofs of any of these formulas. One approach that might work would be to prove that the $P(n, k)$ formula holds when $k = 3$, and then use induction to prove it for higher k as well. The cases where $k = 1$ or $k = 2$ probably require an additional layer of complexity to deal with the superposition of bands.

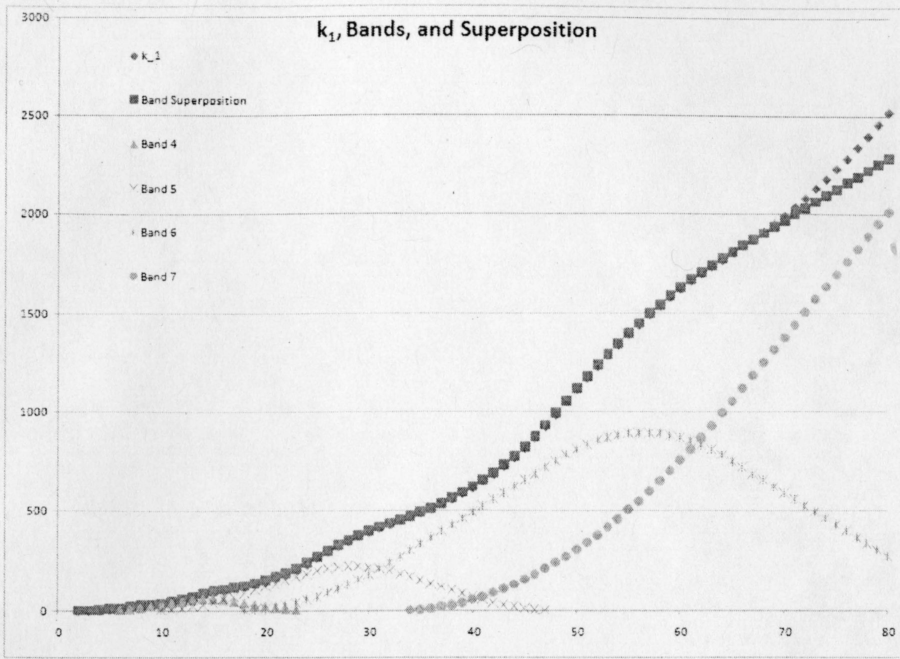

FIGURE 12. $P(n,1)$ vs. n with superposition of bands

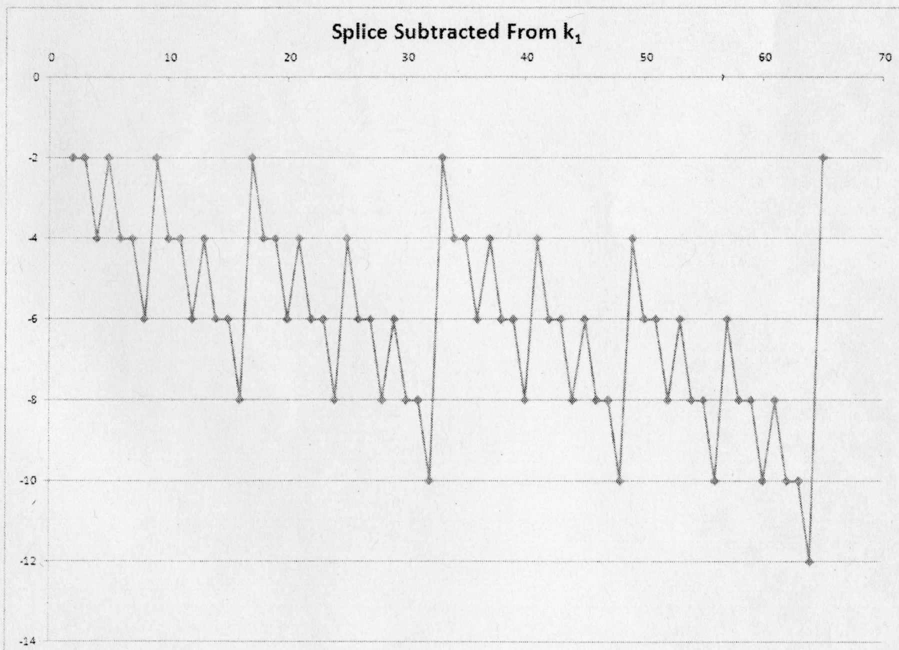

FIGURE 13. $P(n,1)$ vs. n - the superposition of bands

5. References

[1] Dongju Kim and Daekeun Lim, *2-connected and 2-edge-connected Steinhaus graphs*, Discrete Math. 256 (2002)

6. Acknowledgements

I would like to thank Professor Wayne Dymàček for inspiring me to do a thesis through his Graph Theory class, and then providing me with invaluable direction, insight, and support throughout the project.