J. RECREATIONAL MATHEMATICS, Vol. 33(3) 194-203, 2005-2006

HOW KEN GRIFFEY, JR. IS LIKE KEVIN BACON OR, DEGREES OF SEPARATION IN BASEBALL

MICHAEL R. HUBER
STEVEN B. HORTON
Department of Mathematical Sciences
United States Military Academy
West Point, NY 10996

What do baseball and the movie industry have to do with mathematics? There are many situations in which it is interesting to model relationships between members of a finite set. The Oracle of Bacon (www.cs.virginia.edu/oracle [1]) is a Website that helps visitors find a "path" of motion pictures and actors that links a specified actor to Kevin Bacon. If you confer with the Oracle of Bacon, you can find that just about every motion picture actor has a link to any other actor. In fact, Babe Ruth, the New York Yankees baseball star who actually had a few acting parts, has a Bacon Number of 3. According to the Oracle of Bacon, Babe Ruth was in *The Pride of the Yankees* (1942) with Teresa Wright; Teresa Wright was in *Somewhere in Time* (1980) with JoBe Cerny; and JoBe Cenry was in *Novocaine* (2001) with Kevin Bacon. According to its Website, the object of the Oracle is to start with any actor or actress who has been in a movie and link him or her to Kevin Bacon in the smallest number of connections possible. Let's call each of these connections a Degree of Separation.

Interestingly, there is also an Oracle of Baseball, found at www.baseball-research.com [2], where you can discover how many Degrees of Separation a player in the Major Leagues is from another player. Two players are related if they were on the same team at the same time. For example, Babe Ruth and Roger Clemens, two former Yankees who came to New York after successful pitching careers in Boston, have 5 Degrees of Separation, even though they played over 50 years apart form each other. One such set of connections (or path) is:

Babe Ruth played with Red Ruffing for the 1932 New York Yankees.
Red Ruffing played with Phil Rizzuto for the 1942 New York Yankees.

Phil Rizzuto played with Ralph Terry for the 1956 New York Yankees.
Ralph Terry played with Tom Seaver for the 1967 New York Mets.
Tom Seaver played with Roger Clemens for the 1986 Boston Red Sox.

There are several other paths that link Ruth to Clemens. Some of them involve other legendary baseball players such as Willie Mays, Leo Durocher, Frank Robinson, and Brooks Robinson.

1. Bacon Numbers, Graphs, and Adjacency Matrices

Mathematicians can use graphs to model these aforementioned relationships. In this sense of the word, a *graph* is a set of elements called *vertices*, and a set of pairs of vertices called *edges*. We write $G = (V, E)$, where G is the graph, V is the vertex set, and E is the edge set, to express this idea. We will only consider graphs where the set of vertices is finite. Graphs are ideally suited to model relationships between objects.

One way to consider the Oracle of Bacon is to think of it as a graph. The actors are represented by vertices, and whenever two actors appeared in some movie, there is an edge in the graph between the vertices corresponding to those actors. For example, the Oracle of Bacon graph would contain an edge between vertices corresponding to Babe Ruth (the actor) and Teresa Wright, because of their appearance in *The Pride of the Yankees*. But there is not an edge between, say, Rudolph Valentino and Matt Damon because they never made a movie together. An actor's *Bacon number* is the length of a shortest path in this graph from the actor's vertex to Kevin Bacon's vertex (by the way, we define Kevin Bacon's Bacon number to be 0).

From this graph, one could in theory determine the Bacon number of any actor. Unfortunately, doing this by hand is impractical. The graph that underlies the Oracle of Bacon is quite large—there are hundreds of thousands of actors in the Oracle's database. The number of edges is even larger. Just writing it down would take a long time and a lot of paper. Even if you could put the graph on a piece (or several pieces) of paper, it would be hard to find the shortest paths between specified vertices without the help of technology.

Fortunately, there is a way to get your computer to help you. A graph is often represented in a computer using an *adjacency matrix*. An adjacency matrix A of a graph G as a square $(n \times n)$ matrix defined by

$$a_{ij} = \begin{cases} 1 & \text{if there is an edge between vertices } i \text{ and } j \\ 0 & \text{otherwise} \end{cases}.$$

Here, a_{ij} represents the matrix element of A that is in the i^{th} row and the j^{th} column.

In essence, we represent the relationships between actors using a graph, and we in turn represent the graph using an adjacency matrix, with we can store in a

computer. Once we have done this, we will take advantage of an interesting fact form linear algebra that allows us to compute the Bacon number for any actor. Here is the fact: if A is an adjacency matrix for a graph G, then the matrix A^k tells us how many paths of length k exist in G between two specified vertices, where A^k is just k copies of A multiplied together using matrix multiplication. In particular, the entry in the i^{th} row and the j^{th} column of A^k is the number of paths of length k between vertex i and vertex j. We will denote this using $a_{ij}^{[k]}$.

This fact lets us compute Bacon numbers. We just have to look at the appropriate entry of A^k for successive values of k. In particular, if we let I be the vertex in G corresponding to Kevin Bacon, and let j be the vertex corresponding to some other actor whose Bacon number we seek, we simply check to see if $a_{ij}^{[1]} > 0$. If it is, then the actor's Bacon number is 1, and otherwise we check to see if $a_{ij}^{[2]} > 0$. If so, then the Bacon number is 2, otherwise we increase k again. We could write down this idea in a more formal way as follows:

Procedure *Bacon Number*

Step 0. Let I be the vertex corresponding to Kevin Bacon and j be the vertex corresponding to another actor; we'll call him or her "actor j."

Step 1. If actor j is Kevin Bacon, then you can stop; his Bacon number is 0. Otherwise, set $k = 1$ and go to the next step.

Step 2. If $a_{ij}^{[k]} > 0$, then actor j's Bacon number is k. Otherwise, increment k (that is, set $k = k + 1$) and repeat this step.

One potential problem with Procedure *Bacon Number* is that it might never end. If actor j has no connection to Kevin Bacon, then this procedure will go on computing values of $a_{ij}^{[k]}$ for successively higher values of k forever, at least in theory. But in practice it seems that most actors who have been in movies have a connection to Kevin Bacon. Those that do typically have Bacon numbers that are fairly small—in fact, the average Bacon number of all 739980 "linkable" actors is (at a recent viewing of the Website) 2.954, and the largest Bacon number of any such actor is 8.

2. A Bacon-like Relation for the 500+ Home Run Club

Inspired by an online article by Jayson Stark, senior baseball writer for *ESPN.com* [3, 4], the authors decided to consider hitting milestones and see if there was a connection between the 500+ Home Run Club members. On the verge of Roger Clemens' 300th career victory, Stark had listed the *Six Degrees of Separation* connecting Clemens to Grover Cleveland Alexander, the third pitcher in modern baseball history to win his 300th game in a career beginning after 1900. Although Stark used pitchers who had not won 300 games as part of the connection (otherwise it would be impossible), the natural question arose if the chain of connections could be established using only members of a specific (in this case,

500 homers) club. Baseball fans will be happy to note that Stark's chain of players connecting Clemens to Alexander did include Hall of Famer Warren Spahn.

Rather than players with 300 or more career wins, we will initially consider only players with 500 or more career home runs [5-7]. That is, our graph will have a vertex for each player who hit at least 500 homers in his career. We will consider two players connected if a player hit his first career home run on the same date that another club member hit a home run. Table 1 summarizes the data relevant to this scenario.

Perhaps somewhat surprisingly, most members of the 500+ home run club are connected to each other. For example, Sammy Sosa is connected to Mickey Mantle by Five Degrees of Separation, shown as follows:

On the day Sosa hit his first home run, Eddie Murray also homered.
On the day Murray his his first home run, Mike Schmidt also homered.
On the day Schmidt his his first home run, Harmon Killebrew also homered.
On the day Killebrew hit his first home run, Ernie Banks also homered.
On the day Banks hit his first home run, Mickey Mantle also homered.

Table 1. 500+ Home Run Club Data

Player	First HR	Who Else Hit a Home Run That Day?	Total HRs*
Sammy Sosa	21-Jun-89	Murray	588
Ken Griffey Jr.	10-Apr-89	Murray, Palmiero	536
Rafael Palmiero	09-Sep-86	McGwire, Schmidt	569
Mark McGwire	25-Aug-86		583
Barry Bonds	04-Jun-86		708
Eddie Murray	18-Apr-77	Schmidt	504
Mike Schmidt	16-Sep-72	Jackson, Killebrew, McCovey (2)	548
Reggie Jackson	17-Sep-67		563
Willie McCovey	02-Aug-59		521
Frank Robinson	28-Apr-56	See Note below**	586
Harmon Killebrew	24-Jun-55	Aaron, Banks, Mathews	573
Hank Aaron	23-Apr-54	Banks	755
Ernie Banks	20-Sep-53	Mantle	512
Eddie Mathews	19-Apr-52	Mays	512
Willie Mays	28-May-51		660
Mickey Mantle	01-May-51		536
Ted Williams	23-Apr-39		521
Mel Ott	18-Jul-27		511
Jimmie Foxx	31-May-27	Ruth (2, in each game of DH)	534
Babe Ruth	06-May-15		714

*The number of home runs is current as of the end of the 2005 season.

**According to the 29 April 1956 edition of *The New York Times*, on the same day that Frank Robinson hit his first home run for the Cincinnati Reds, the Yankees' Mickey Mantle hit a "home run" against the Boston Red Sox, but the umpire ruled it a triple, as it bounced from the fourth row of the stands back onto the field. Manager Casey Stengel was ejected from the game after unsuccessfully arguing the call. The Yankees filed a formal protest, but the call remained. With Mantle's official triple, Frank Robinson remains unconnected to any other slugger with 500 home runs.

Sammy Sosa also has Five Degrees of Separation to Willie Mays, through Eddie Murray, Mike Schmidt, Harmon Killebrew, and Eddie Mathews. In total, Sammy Sosa is connected to ten other members of the 500+ home run club.

On June 20, 2004, Ken Griffey, Jr., hit his historic 500th home run and became the newest member of the Club, and he thus established connection to 13 other Club members (Sosa, Murray, Palmiero, McGwire, Schmidt, Jackson, Killebrew, McCovey, Aaron, Banks, Mathews, Mantle, and Mays).

In terms of recent members of the 500+ Home Run Club, Barry Bonds is not connected to any other member. Contributing to Bonds' isolation could be the relatively small number of 500+ Home Run Club members playing in 1986 and the low number of home runs hit by those players in that year. In 1986, Reggie Jackson hit 18 home runs, Eddie Murray only hit 17, and Mike Schmidt hit 37 (to lead the National League). On the day when Barry Bonds hit his first home run (June 4, 1986), Jose Canseco (462 career home runs) also homered, and Dave Winfield (465 career home runs) homered twice. But even in the 400+ Home Run Club, this is as far as it goes; no other 400+ career home run hitters are connected to any of these three players. However, including players with 400+ home runs, Canseco is only connected to Winfield (since Winfield homered on the same day than Canseco his his first), and Winfield is connected to no one else with at least 400 career home runs. Figure 1 shows a graph of the hitters in the 500+ Home Run Club.

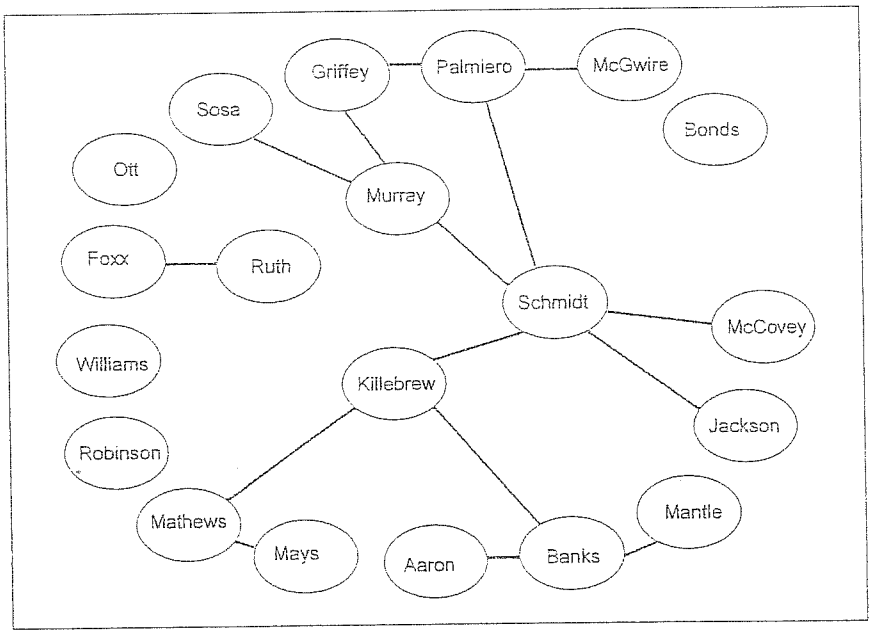

Figure 1. The 500+ Home Run Club.

If we remove the vertices that have no neighbors, and add a numerical label to each remaining vertex, we obtain the graph shown in Figure 2. These labels will correspond to the row/column number in the adjacency matrices that appear later. Notice that Ruth and Foxx are not considered isolated vertices since they are adjacent to each other.

The adjacency matrix for the graph shown in Figure 2 is shown in Figure 3.

If we square the adjacency matrix, the (i,j) entry of A^2 is equal to the number of paths of length 2 between vertex i and vertex j. In this case, A^2 is given by the adjacency matrix shown in Figure 4.

Observe that $a_{2,6}^{[2]}=2$. The (2,6) entry, corresponding to Griffey-Schmidt, is equal to 2, meaning that there are two paths of length 2 between the corresponding vertices. There are also two paths of length 2 between Palmiero and Murray (the (3,5) entry). Similarly, by cubing the adjacency matrix we can see that there are two paths of length 3 between Griffey and Schmidt, since $a_{2,6}^{[3]}=2$. Alternatively, you can convince yourself of this by inspecting the graph shown in Figure 2. Further, if we sum the elements of any row of the original adjacency matrix, we find that the sum equals the same number on the corresponding main diagonal of A^2.

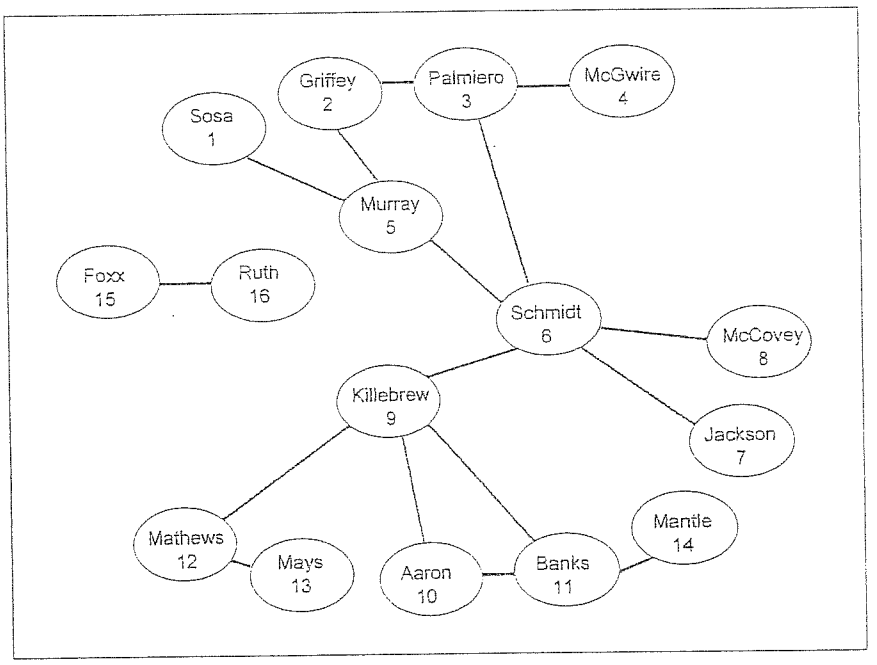

Figure 2. The 500+ Home Run Club; labeled and without isolates.

$$\begin{pmatrix}
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{pmatrix}$$

Figure 3. Adjacency Matrix for the 500+ Home Run Club.

$$\begin{pmatrix}
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 2 & 0 & 1 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 3 & 0 & 2 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 2 & 0 & 3 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 2 & 0 & 1 & 0 & 5 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 4 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 2 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}$$

Figure 4. A^2, the Square of the Adjacency Matrix.

What club membership rules should we change to ensure that each batter is connected to every other batter? Since A^k gives us the number of paths of length exactly k between each pair of players, we can obtain the number of paths of length k or less between each pair of players by computing $A + A^2 + \cdots + A^k$. A natural question is what is the smallest value of k such that $A + A^2 + \cdots + A^k$ has no zero entries, meaning that all vertices are connected by a path of length k or less? For the 500+ Home Run Club, that value of k is 5, as long as we are willing to ignore the zeroes in the last two rows and columns corresponding to Ruth and Foxx. Of course if we don't ignore Ruth and Foxx, there is no k that is big enough. (See Figure 5.)

If we relax the requirements for club membership to a number well below 500, we can connect all of these sluggers to Babe Ruth. However, the minimum number of home runs would be below 400. Future research could establish this number exactly. As an appetizer, consider the following:

> On the day Mickey Mantle hit his first home run, Stan Musial also homered.
> On the day Willie Mays hit his first home run, Stan Musial also homered.

Hall of Famer Stan Musial had 475 career home runs.

> On the day Stan Musial hit his first home run, Joe DiMaggio also homered.
> On the day Ted Williams hit his first home run, Joe DiMaggio also homered.

$$\begin{pmatrix}
4 & 6 & 17 & 2 & 20 & 9 & 9 & 9 & 12 & 2 & 1 & 1 & 1 & 1 & 0 & 0 \\
6 & 12 & 37 & 6 & 37 & 18 & 18 & 18 & 24 & 4 & 2 & 2 & 2 & 2 & 0 & 0 \\
17 & 37 & 19 & 20 & 17 & 67 & 9 & 9 & 13 & 13 & 13 & 13 & 1 & 1 & 0 & 0 \\
2 & 6 & 20 & 4 & 17 & 9 & 9 & 9 & 12 & 2 & 1 & 1 & 1 & 1 & 0 & 0 \\
20 & 37 & 17 & 17 & 19 & 67 & 9 & 9 & 13 & 13 & 13 & 13 & 1 & 1 & 0 & 0 \\
9 & 18 & 67 & 9 & 67 & 40 & 40 & 40 & 70 & 20 & 11 & 11 & 10 & 10 & 0 & 0 \\
9 & 18 & 9 & 9 & 9 & 40 & 6 & 6 & 10 & 10 & 10 & 10 & 1 & 1 & 0 & 0 \\
9 & 18 & 9 & 9 & 9 & 40 & 6 & 6 & 10 & 10 & 10 & 10 & 1 & 1 & 0 & 0 \\
12 & 24 & 13 & 12 & 13 & 70 & 10 & 10 & 36 & 35 & 34 & 34 & 7 & 7 & 0 & 0 \\
1 & 2 & 12 & 1 & 12 & 10 & 9 & 9 & 28 & 12 & 7 & 7 & 6 & 6 & 0 & 0 \\
2 & 4 & 14 & 2 & 14 & 21 & 11 & 11 & 41 & 24 & 17 & 15 & 8 & 11 & 0 & 0 \\
1 & 2 & 13 & 1 & 13 & 11 & 10 & 10 & 34 & 14 & 8 & 10 & 10 & 7 & 0 & 0 \\
1 & 2 & 1 & 1 & 1 & 10 & 1 & 1 & 7 & 7 & 7 & 10 & 3 & 1 & 0 & 0 \\
1 & 2 & 2 & 1 & 2 & 11 & 2 & 2 & 13 & 11 & 11 & 8 & 2 & 4 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 3 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 2
\end{pmatrix}$$

Figure 5. $A + A^2 + \cdots A^k$, where $k = 5$.

Hall of Famer Joe DiMaggio had 361 career home runs.

On the day Joe DiMaggio hit his first home run, Al Simmons also homered.

Hall of Famer Al Simmons had 307 career home runs.

On the day Al Simmons hit his first home run, Babe Ruth also homered.
On the day Met Ott hit his first home run, Lou Gehrig also homered.
On the day Jimmie Foxx hit his first home run, Lou Gehrig and Babe Ruth both homered.

Hall of Famer Lou Gehrig had 493 career home runs.

This would bring the "magic" number down to 307 (Simmons' career total), but it would tie all of the 500+ Home Run Club sluggers together with other Hall of Fame sluggers (except for Frank Robinson and Barry Bonds; Frank Robinson gets connected if we consider the Mantle triple ...). The new graph that also includes vertices for the players mentioned above is shown in Figure 6.

There is a rich tradition among baseball fans of gathering and analyzing obscure statistical information. If the ideas described in this article are interesting to you, you are probably a baseball fan! How connected to the great sluggers is your favorite home run hitter?

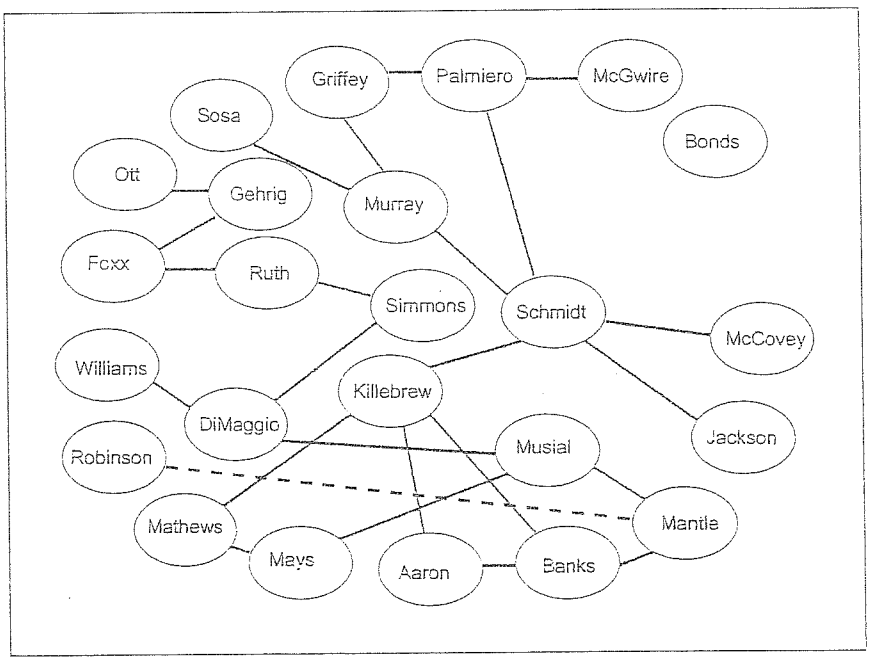

Figure 6. The 500+ Home Run Club Plus Four.

References

1. P. Reynolds, *The Oracle of Bacon*, located on the Web at www.cs.virginia.edu/oracle.
2. P. Mathews and S. Forman, *The Oracle of Baseball*, located on the Web at www.Baseball-Reference.com.
3. J. Stark, *Clemens, Clemens ... and More Clemens*, an article posted on *ESPN.com* on Saturday, June 7, 2003, located on the Web at http://espn.go.com/mlb/columns/stark_jayson/1564647.html.
4. J. Stark, personal correspondence, June 9, 2003.
5. B. McConnell and D. Vincent, *SABR Presents The Home Run Encyclopedia*, Macmillan, Inc., 1996. (This amazing book lists the dates when each slugger hit his first Major League home run.)
6. www.Retrosheet.org. (The authors were able to check box scores online for all sluggers who hit home runs after 1972. The online box scores for 1967 were not complete at the time of writing.)
7. *The New York Times*. (The authors used the *Times* on microfilm to check box scores before 1972.)