KEVIN BACON AND GRAPH THEORY

Brian Hopkins

ADDRESS: Department of Mathematics, Saint Peter's College, Jersey City NJ 07306 USA. bhopkins@spc.edu.

ABSTRACT: The interconnected world of actors and movies is a familiar, rich example for graph theory. This paper gives the history of the "Kevin Bacon Game" and makes extensive use of a Web site to analyze the underlying graph. The main content is the classroom development of the weighted average to determine the best choice of "center" for the graph. The article concludes with additional student activities and some responses to the material.

KEYWORDS: Cinema, finite mathematics, graph theory, popular culture, six degrees of separation, weighted averages.

1 INTRODUCTION

Graph theory is the mathematics of connections. It has wide applications to large, interconnected systems: transportation networks, epidemiology, and the Internet, to name just a few. But we teach graph theory with pictures of a handful of dots and lines. There is one large system that is easy to work with, thanks to a Web site run by the University of Virginia, Department of Computer Science. The Oracle of Bacon at Virginia [6] uses the Internet Movie Database [3], which documents almost all of cinematic history. This is a good tool for illustrating complete subgraphs, connected components, and the distance between vertices. There is also a nice application of weighted averages. I have used this material in freshman finite mathematics classes and mathematics major courses that cover graph theory; students always respond enthusiastically.

2 SIX DEGREES

One notion of connection in popular culture is "six degrees of separation." This traces back to a 1967 article by Stanley Milgram, a social psychologist better known for experiments on obedience to authority. "The Small-World Problem" [5] reports an experiment where Milgram gave letters to residents of Omaha, Nebraska that were intended for a stockbroker in a Boston suburb. The letters could only be sent to someone the sender knew on a first-name basis, who would forward the letter, etc. Milgram collected the letters that made it to the stockbroker and tallied how many links there were. Surprisingly, the average was only 5.2. The popularized version of this concept claims that there are only six links between any two people on earth; sociologists continue to speculate on this notion [4]. Milgram also reports on patterns in the paths of the successfully delivered letters; over half came through the same three individuals in the Boston area. Malcolm Gladwell's "Six Degrees of Lois Weisberg" [2] is an entertaining article exploring these ideas and their ramifications in society.

There are several problems with six degrees of separation in the context of graph theory. Defining a graph requires specifying its vertices and edges. The vertices of the "six degrees graph" are people, but the edge criterion is vague: should your vertex be connected to the vertices of everyone you have met? your acquaintances? just your friends? Even if we settled what relationship merits an edge in the graph, we do not have access to a significant amount of data. No database lists all of my friends, for instance. Finally, even if there were such a database, it would be unwieldy: a graph with over 6 billion vertices is too big for current computational analysis.

3 CINEMA AND THE ORACLE

Fortunately, there is a well-defined system of connections documented in a database which is large enough to be interesting, but well within the range of current computational power. The Internet Movie Database (IMDb below) has information on roughly 620,000 movie actors from around the world and approximtely 280,000 films dating back to 1891. To make this a graph, make a vertex for each actor and use an edge to connect the vertices corresponding to two actors who were in the same film. For instance, Tom Cruise and Kevin Bacon were both in *A Few Good Men,* which we represent as in Figure 1.

There will be an edge corresponding to *A Few Good Men* between each pair of actors from that film, so the "cinema graph" will have many more

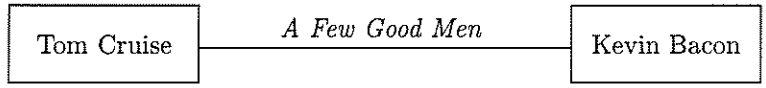

Figure 1. A small part of the cinema graph.

than 280,000 edges. In other words, the subgraph of vertices for actors in *A Few Good Men* and all edges for that film make a complete graph. Many more graph theoretic aspects of the cinema graph are discussed in Duncan Watt's *Small Worlds* [9], which also considers the Western States Power Grid and the neural network of the *Caenorhabditis elegans*, a worm.

In January 1994, Craig Fass, Brian Turtle, and Mike Ginelli, students at Albright College, came up with a cinematic variant of six degrees of separation. The object of the "Kevin Bacon Game" is to link an actor to Kevin Bacon in as few films as possible. For example, Nicole Kidman has (to date) never been in a film with Kevin Bacon, but she was in *Eyes Wide Shut* with Tom Cruise, who we know is 1 away from Kevin Bacon. In the language of graph theory, the distance between Nicole Kidman and Kevin Bacon in the cinema graph is 2. The game, also known as "Six Degrees of Kevin Bacon," became popular among movie buffs [1].

If your knowledge of cinema is less than encyclopedic, then you can cheat in the Kevin Bacon Game with a very helpful interface for the IMDb. The Oracle of Bacon at Virginia [6] is hosted by Department of Computer Science at the University of Virginia. Enter the name of an actor, and it will quickly search the database to find that person's distance to Kevin Bacon, reporting back intermediaries and the associated films. Typing "Nicole Kidman" into the Oracle gives a length 2 connection different from above: she was in *The Peacemaker* with Steve Altes, who was in *Hollow Man* with Kevin Bacon. This illustrates the point that while distance is well-defined, there may be multiple paths of that minimal length between two vertices.

The Oracle is fast and user-friendly. It has a generous system for dealing with misspellings, offering likely candidates close to names not in the database. It will find the distance between any two actors, not just to Kevin Bacon. The Oracle provides links to the IMDb pages for actors and films and does several summary computations, as discussed below.

4 CLASSROOM PRESENTATION: WHO'S THE CENTER?

Most students have heard of six degrees of separation, and many are familiar with the Kevin Bacon Game or at least have seen a recent commercial based

on the idea [8]. After I have explained Milgram's work and the set-up for the cinema graph, we spend several minutes on the Oracle looking up the distances of various actors from Kevin Bacon. Students are surprised that the distances are so small and usually cannot come up with an actor more than 3 away.

But why, a student usually asks, of all possible actors, is it Kevin Bacon? The game's creators, Fass, Turtle, and Ginelli, explain that they had watched *Footloose* one snowy day. That night, there was a television advertisement for the upcoming film *The Air Up There,* also starring Kevin Bacon.

> Somehow it occurred to us that Kevin Bacon had been in so many different types of movies, you could connect a lot of unlikely people together through his work. Well, that idea took off. We found we could actually get *anybody* back to Kevin Bacon in a few steps. There could only be one explanation—Kevin Bacon was at the center of the entertainment universe.[1, pp. 13-14]

Is Bacon really the best choice for the game? What would it even mean to be the best choice?

We want a way to measure an actor's centrality in the cinema graph. Students often suggest using the number of co-stars, i.e., how many actors are at distance 1. For example, there are 1703 actors distance 1 from Kevin Bacon (so the full cinema graph has 1702 other edges connected to the Kevin Bacon vertex in Figure 1). There are 1527 actors distance 1 from Wesley Snipes. (All data were supplied by the Oracle in December 2003; the IMDb is updated frequently, so numbers used here may have changed by the time of publication.) So you might conclude that Kevin Bacon is better connected in the cinema graph than Wesley Snipes. But Snipes' fewer co-stars connect to more actors than Bacon's, that is, there are more actors distance 2 from Snipes than from Bacon. Likewise, there are more actors distance 3 from Snipes than from Bacon. How can these data be incorporated?

The weighted average does this well; the computation for Kevin Bacon is shown in Figure 2. (Notice that there are actors up to distance 8 from Kevin Bacon, so "Six Degrees of Kevin Bacon" is a misnomer.)

$$\frac{0\cdot 1+1\cdot 1703+2\cdot 135839+3\cdot 370657+4\cdot 90057+5\cdot 7154+6\cdot 922+7\cdot 94+8\cdot 2}{1+1703+135839+370657+90057+7154+922+94+2}$$

Figure 2. Computation for the average Kevin Bacon number.

So the average distance to Kevin Bacon is 2.948. The complete data for Kevin Bacon, Wesley Snipes, Anthony Quinn, and Rod Steiger are given in

Figure 3, along with the weighted averages. Notice that Snipes' weighted average is 2.923, slightly lower than Bacon's. Thus Wesley Snipes is, on average, more closely connected to other actors, so he is a better center.

actor	0	1	2	3	4	5	6	7	8	ave.
Bacon	1	1703	135,839	370,657	90,057	7154	922	94	2	2.948
Snipes	1	1527	137,327	381,486	79,463	5767	738	117	3	2.923
Quinn	1	3539	219,774	336,525	42,468	3594	490	38	0	2.710
Steiger	1	2692	240,487	322,786	36,391	3533	487	51	1	2.669

Figure 3. Table giving number of actors at each distance and weighted averages.

As with Bacon and Snipes, there's a similar relationship between Anthony Quinn, who has more co-stars, and Rod Steiger, who has a lower weighted average. It may surprise you that Rod Steiger is the best connected actor by this measure. He was both a leading man and character actor in 50 years of films, performing in musicals, dramas, comedies, horror films, independent U.S. films, and European films. There are better known actors, but they have not performed as long in so many genres. The top 1000 best connected actors are listed on the Oracle; Kevin Bacon just misses that distinction.

As you compute the weighted averages, you see that the denominator in all four cases is 606,429. The four actors are all connected to each other, so they are all connected to the same large set of actors. But the IMDb includes some 620,000 actors; where are the rest? Not all actors connect to the Hollywood mainstream. Type "Olga Holts" into the Oracle, for example, and it will report a distance of infinity from Kevin Bacon. The cinema graph has a large connected component of 606,429 vertices with the remainder falling into several small connected components. Most components consist of only a few vertices, but Olga Holts' vertex is part of a component of 29 corresponding to Estonian silent film actors of the 1920s.

5 OTHER ACTIVITIES AND STUDENT RESPONSES

In order to reinforce student understanding of the cinema graph's structure, I ask them in homework or a collaborative project to draw the connected component that includes Olga Holts' vertex. The Oracle will list all of an actor's co-stars, and switching over to the IMDb provides complete data

on the four films connecting these 29 actors. The drawing consists of four overlapping complete graphs.

The IMDb is updated regularly, so the cinema graph is dynamic. Homework or project questions addressing this are "Can a deceased actor's Kevin Bacon number decrease? If so, how low can it go?" In more advanced settings, I explain that the diameter (the maximum distance) of the cinema graph is 14 and ask "Over time, do you think the diameter of the cinema graph decreases, stays the same, or increases? What factors are relevant to the diameter?" One student suggested that the diameter decreases, explaining that there are new films (which can decrease the diameter) without new actors, while new actors (which can increase the diameter) can only be introduced in new films. In fact, the diameter of the cinema graph has decreased over time [7].

Also, I offer extra credit for finding an actor with (finite) distance 5 or more from Kevin Bacon (with no credit for students who turn in names from the same film). The IMDb allows for searches by genre, language, and decade, so this is more about database work than obscure film knowledge. Most students can find an actor with distance 5 from Kevin Bacon; recently a student found one of the two Kevin Bacon 8's!

Student response has been very favorable. Graph theory students enjoyed exploring a "real graph" of significant size. Working with a graph larger than one I could draw on the board helped them realize the subtlety of various definitions and the need for efficient search algorithms. Finite mathematics students liked the application of weighted averages to deciding which actors are better connected. The discussion of the Kevin Bacon Game, and basic graph theory in general, engaged several finite mathematics students who had not previously participated at a high level. One such student said, "This isn't math!" I was very happy for her to see mathematics other than formulas and variables.

ACKNOWLEDGEMENTS

This article is based on a talk I gave in the Using Popular Culture in the Mathematics and Mathematics Education Classroom session of MathFest 2002, held in Burlington, Vermont. I would like to thank the organizers, Sarah Greenwald and Andrew Nestler, for accepting my talk and encouraging the development of this article. Suggestions from the anonymous reviewers were very helpful.

REFERENCES

1. Fass, Craig, Mike Ginelli, and Brian Turtle. 1996. *Six Degrees of Kevin Bacon.* New York: Plume/Penguin Books.

2. Gladwell, Malcolm. 1999. Six Degrees of Lois Weisberg. *The New Yorker.* 74(41): 52-63. (Also available through `http://gladwell.com/`.)

3. *The Internet Movie Database.* `http://www.imdb.com/`. Maintained by Internet Movie Database, Ltd., an Amazon.com company.

4. Kleinfeld, Judith. 2002. The Small World Problem. *Society.* 39(2): 61-66.

5. Milgram, Stanley. 1967. The Small-World Problem. *Psychology Today.* 1(1): 60-67.

6. *The Oracle of Bacon at Virginia.* `http://www.cs.virginia.edu/oracle/` Maintained by Patrick Reynolds, hosted by the University of Virginia Department of Computer Science.

7. Reynolds, Patrick. 2003. Private communication.

8. Visa Check Card commercial. 2002. Six Degrees of Identification. Superbowl XXXVI.

9. Watts, Duncan. 1999. *Small Worlds: The Dynamics of Networks between Order and Randomness.* Princeton NJ: Princeton University Press.

BIOGRAPHICAL SKETCH

Brian Hopkins is an assistant professor of mathematics at Saint Peter's College, a Jesuit liberal arts college in Jersey City, New Jersey. He received a BS in Mathematics and a BA in Philosophy from the University of Texas in 1990, and a PhD in Mathematics from the University of Washington in 1997. His research interests range from representation theory of Lie algebras through algebraic combinatorics to graph theory. He also enjoys choral singing, poetry, and New York City.