

Arnav Shirodkar
CMSC 201 – Lab 11
as9086@bard.edu
15/5/20

Assignment: Using the java files provided and the IMDB database, answer the following questions about Kevin Bacon's Hollywood presence and continue exploring the data set as much as possible.

Collaboration Statement: I worked on this alone

Note: I exclusively used the full IMDB file, movies2017-309K.txt for this lab. In the case where code has been pasted, the additions to the code have been highlighted in Red.

How many movies has Kevin Bacon acted in? (use symbolgraph.java)

I made the following adjust to SymbolGraph.java to address this question:

```
....
String source = StdIn.readLine();
    if (sg.contains(source)) {
        int s = sg.index(source);
        int count = 0;
        for (int v : graph.adj(s)) {
            StdOut.println(" " + sg.name(v));
            count++; //increments for every movie connected to Kevin Bacon
        }
        StdOut.println(count);
    }
    else
        ....
```

Since Kevin Bacon is listed as *Bacon, Kevin (I)* in the file, we provide that as the argument source and the value returned for count was 74, which is the total number of movies Kevin Bacon has starred in as of when the file was created.

How many actors have co-starred with Kevin Bacon?

Based on the structure of the data input, we know for sure that each "movie" vertex only has "actor" vertices connected to it, and vice-versa. Therefore, for each movie Kevin Bacon stars in, we can use a nested for-loop to find the number of actors vertices connected to each movie vertex (i.e. the number of actors he worked with in the movie) and deduce the number of actors who have co-starred with him. However, we must also avoid the unnecessary addition of duplicates where Kevin Bacon co-stars with an actor in more than one movie.

Therefore, we can create a function that creates a new ST <string, integer> that holds the name of each actor that has co-starred with Kevin Bacon.

The function is as follows:

```
public int findCoStars(String actor){
    costar = new ST<String, Integer>();
    int s = this.index(actor); // find the index of the actor from the constructed graph
    for(int v : graph.adj(s)){ // loop for all movies connected to argument "actor" vertex
        for(int x : graph.adj(v)){
            // nested loop for all "actor" vertices connected to each movie
            if(!costar.contains(this.name(x))){
                costar.put(this.name(x),1); //create symbol table with each costar
            }
        }
    }
    int count = 0;
    for(String x : costar){
        count++; // iterate through the ST to find total number of co-stars
    }
    return count;
}
```

Using the above method, we can find that Kevin Bacon has co-starred with 3567 other actors. However, we have not excluded Kevin Bacon's own vertex from the tally, so the actual number should be 3566. We can check the viability of this value by running a nested loop test with a count increment for each "actor" vertex connected to each of the movies Kevin Bacon has starred in, without accounting for duplicates. This gives us the value 3854, and since the value taking duplicates into account is smaller, we can partially guarantee its accuracy.

Find all the actors who have co-starred with Kevin Bacon at least twice

The method we introduced earlier can be easily repurposed to answer this question. We can alter it such that it not only stores the names of all of Kevin Bacon's co-stars, but also holds the number of times they have each worked with him.

```
public void findCoStars(String actor){
    costar = new ST<String, Integer>();
    int s = this.index(actor);
    for(int v : graph.adj(s)){
        for(int x : graph.adj(v)){
            if(!costar.contains(this.name(x))){
                costar.put(this.name(x),1);
            }
        }
    }
}
```

```

    } else if (costar.contains(this.name(x))) {
        costar.put(this.name(x), costar.get(this.name(x)) + 1);
        // increment for co-stars who appear more than once
    }
}
} for (String x : costar) {
    if (costar.get(x) > 1) StdOut.println(x);
    // print the names of co-stars who have worked with Kevin Bacon at least twice
}
}

```

The list of actors who have co-starred with Kevin Bacon at least twice is included in the appendix below. Like the earlier question, Kevin Bacon's own name still appears in this tally as no exclusion case in the for-loop has been specified. Given that he has starred in 74 movies, the integer value against his name in the ST should also be 74.

What pairs of actors have co-starred together most often? (refer to AlteredSG . java)

This scenario is computationally difficult to handle as in the massive data set, we must take each actor as our Kevin Bacon, check all of their co-stars and find out whom they have worked with the most, afterwards we repeat this for every actor in the set.

To address this, I created a new class AlteredSG.java. As the graph and symbol table is being constructed, this class creates an additional ST labelled as *tracker* which holds all the indices of actor vertices as keys with a binary value 0 or 1 (where 0 represents an unmarked actor, and 1 represents a marked actor). A second ST is created known as *maxLog*, holding string keys and integer values; this will be elaborated upon later.

The AlteredSG class also has two additional methods: "storeCoStarMax" and "findMaxPair".

Like before, "StoreCoStarMax" takes the index of an actor vertex as an argument, using it to find all of the actor's costars as well as the number of times they have worked together, storing it in an ST called *costar*. We then find the entry for the co-star most frequently worked with and adds this max entry into the *maxLog* ST, concatenating the string key with the name of the actor whose costars we are currently finding (the "argument"). Then, using the argument index, we change the value for our actor vertex in the *tracker* ST to 1 so that it is marked, ensuring that it does not get checked again.

We run this for all actor vertexes stored in the *tracker* ST. As the number of marked actors increases, the search becomes faster for each subsequent actor.

The function "findMaxPair" then finds the max value from *maxLogST* and prints out all the keys that have this value, finding all the pairs of actors that have co-starred most frequently as well as the number of times they have costarred.

Using this, I managed to deduce that the pair "Flowers, Bess/Miller, Harold (I)" have co-starred the most frequently, having done so 233 times.

Degrees of Kevin Bacon (use `baconhistogram.java` and `degreesofseparation.java`)
Find the Bacon# of all the actors in the graph (represented as a histogram)

Bacon#	Number of Actors
0	1
1	3566
2	409658
3	1097885
4	146885
5	7070
6	627
7	66
8	7
9	12
INF	99129

Using `BaconHistogram.java`, we get the following histogram representing the distribution of Bacon numbers of all the actors in the graph. The number of actors who have co-starred with Kevin Bacon is again verified here (all actors with Bacon number = 1). We can observe that a majority of the actors have a Bacon Number of 3. The program already takes into account that the graph is a mixture of both movie and actor vertices, using `bacon/2` to find the bacon numbers for actors. The full list of each actor and their bacon number is included in the list below.

Who has the largest and second largest Bacon #'s?

Based on how `BaconHistogram.java` was written, it computes the histogram of Kevin Bacon numbers - 100 for infinity, initialising an array of size 100 and taking anything with a larger bacon value as exceeding "MAX_BACON". The program also prints out the names and bacon number of actors with a bacon number between 7 and the maximum.

The largest 2 bacon numbers for the graph is 9 and 8, belonging to these actors

- 8 – Arthi, Antara Sarker
- 8 – Chowdhury, Iftikhar Ahmed
- 8 – Sarker, Jharna
- 8 – Tathoi
- 8 – Al-Eyaf, Abdullah
- 8 – Al-Gannas, Ibrahim
- 8 – Areeshi, Bassam
- 9 – Angsu, Tahmid
- 9 – Ophelia (VI)
- 9 – Sunny, Arman Haque

9 - Astrian, Adre
9 - Khan, Wazi Ullah
9 - Mojid
9 - Mridha, Wasi Ur
9 - Murshed, Niaz
9 - Piu, Adrita Priyoti
9 - Reman, Asif
9 - Rifat, Nur
9 - Riyon, Aeiub

Center of the Hollywood Universe

What is Kevin Bacon's Hollywood number? Are there other actors with better Hollywood numbers?

The Hollywood number is computed by finding the average bacon number of all the other actors in the data set. We can find this value for Kevin Bacon by again using BaconHistogram.java and modifying it to return the Hollywood number as a statistic, excluding those who's bacon number exceeds "MAX_BACON".

```
// print out histogram - even indices are actors
for (int i = 0; i < MAX_BACON; i += 2) {
    if (hist[i] == 0) break;
    StdOut.printf("%3d %8d\n", i/2, hist[i]);
}
StdOut.printf("Inf %8d\n", hist[MAX_BACON]);

// using the histogram values, calculate the hollywood number
double actorTotal = 0;
double baconTotal=0;
for(int i=0; i< MAX_BACON; i += 2){
    actorTotal += hist[i];
    baconTotal += (i/2)*hist[i];
}

StdOut.println(baconTotal/actorTotal);
```

In this situation, I have left the actorTotal and baconTotal variables as double and manually round them off to 3 significant figures. From the program, the returned value for Kevin Bacon's hollywood number is 2.84. Visually, this tallies with the histogram as it's closest integer is 3, which is also the mode bacon number.

Using the data from the histogram, we can deduce that this data set includes 1,764,905 actor vertices: trying to use a brute force method to examine the centrality of every actor would take very long. However, with a few guesses I found some actors who have a better Hollywood number than Kevin Bacon.

2.70 – Samuel L. Jackson

2.79 – Tom Hanks

2.76 – Dennis hopper

It is also of interest to note that for several of these actors with better Hollywood numbers, the names of actors and movies with large “bacon numbers” that were printed were almost identical to the actors in the list above (when we found the actors with the largest bacon numbers.) Furthermore, the number of actors that fell into category of having bacon numbers in the “INF” range is also almost always 99129 (the value we found for the “INF” range in Kevin Bacon’s case.) This holds true for hollywood centric actors as well as other actors like **“Riyon, Aeiub”** and **“Sifrin, R.”** This is interesting because it suggests that this data set might not be one big clustered interconnected graph; instead it might be one very large cluster of vertices with a much smaller cluster next to it that is largely disconnected.

Another interesting observation is that I have been unable to find an actor that is connected to any other actors beyond 10 degrees of separation, as the range of bacon numbers for the histogram should technically go all the way to 100, but never does so.

Appendix

List of actors who have co-starred with Kevin Bacon at least twice is as follows:

Alabiso, Giovanni
Alejandro, Charlie
Alejandro, Mel
Bacon, Kevin (I)
Barkin, Ellen
Beach, Michael (I)
Becker, Gerry
Belushi, John
Belzer, Richard
Bergeron, Susan
Berkeley, Xander
Bishop, John (I)
Blake, Geoffrey
Boston, David (IV)
Bratt, Benjamin (I)
Bronk, Paul
Brutsche, Mark
Burt, Rick
Burzenski, Mark
Byrd, Eugene
Cafiso, Matteo
Cali, Joe (III)
Candy, John
Cavestani, Frank

Christopher, Tory
Clark, Spencer Treat
Connell, Angel
Corazzini, Jeffrey
Costa, Stevie
Costner, Kevin
Craven, Matt (I)
Cruise, Tom
Darga, Christopher
Dash, Nick
De Beck, David
Di Stasio, Mike
DiMino, Sal
Dillon, Matt (I)
Dodds, K.K.
Douglas, Illeana
Drew, Boyd
Dunn, Kevin (I)
Eklund Jr., Dicky
Erwin, Bill
Eudene, Larry
Falvo, Mark
Farmer, Lonnie (I)
Feruzi, Mugisha
Fishburne, Laurence
Fluker, Keith
Flynn, Steve (III)
Fogarty, Shawn
Ford, Jim (III)
Franchi, John (I)
Garcia, Naheem
Gathegi, Edi
Gerety, Peter
Gilborn, Steven
Gillies, Suzanne
Gooding Jr., Cuba
Goodman, John (I)
Gordon, Jared M.
Gress, Googy
Grover, Deborah
Grunberg, Greg
Guest, Christopher (I)
Gunderson, Josh (II)
Gunton, Bob
Hall, London
Hankin, Larry
Hanks, Tom

Harden, Marcia Gay
Harkins, Michael (I)
Hecker, Gary A.
Hiou, Arthur
Hoffman, Dustin
Hounsou, Djimon
Howard, Clint
Howard, Rance
Hulce, Tom
Imbergamo, Frankie
Iwanicki, Stacey Forbes
Jarret, Gabriel
Johnston, J.J. (I)
Jonsdottir, Berglind
Kay, Dominic Scott
Kelly, Colleen (XXII)
Kemp, Tom (I)
Kenney, Douglas (I)
Kent, J Parker
Kidd, Rhet
Kinney, Terry
Kyle, Stephen (I)
Laperriere, Lauren
Leigh, Jennifer Jason
Lenehan, Nancy
Lezama, Sydney
Luise, Charles
MacLean, Paige
MacVittie, Bruce
Mariano, Tom
Markell, Jodie
Martin, Steve (I)
Mason, Marsha (I)
Matheson, Tim (I)
Mathews, John M.
Mazareas, Jamie
Mazzello, Joseph
McCann, Sean (I)
McClure, Marc (I)
McClurg, Edie
McDermott, Erica
McGill, Bruce
McKean, Michael (I)
McRobbie, Peter
Meskimen, Jim (I)
Milder, Andy
Miller, Chris (III)

Molinari, Anthony
Morrison, Jennifer (II)
Morse, Peter (VII)
Morshower, Glenn
Murray, Bill (I)
Newcomb, Lulie (I)
Nixon, Richard
O'Halloran, Johnny
Oldman, Gary (I)
Oliveira, Joseph (III)
Pacheco, Richard (II)
Paxton, Bill
Pierce, Wendell (I)
Pieterse, Sasha
Pitt, Brad
Platt, Oliver
Powell, Rickland
Prochaska, Bobby
Putney, Duncan B.
Quinlan, Kathleen
Radakovic, Marinko
Rankin, Steve (I)
Reiser, Paul (I)
Renfro, Brad
Replogle, Stew
Ritz, James (I)
Robbins, Tim (I)
Rooker, Michael (I)
Sanders, Jay O.
Sedgwick, Kyra
Senzy, Arthur
Sinise, Gary
Slade, Max Elliott
Smith Jr., Eddie Bo
Spano, Joe
Stapleton, Joe (II)
Stern, Daniel (I)
Stolte, Christian
Strathairn, David
Streep, Meryl
Struffolino, David
Sutherland, Donald (I)
Sutherland, Kiefer (I)
Suvari, Mena
Swanson, Michael Steven
Tabakin, Ralph
Thornton, Billy Bob

Thorpe, Bill (I)
Tomei, Marisa
Turner, J.T. (I)
Vaughn, Ned
Vezina, George J.
Villamizar, Nicolas
Vince, Pruitt Taylor
Wahlberg, Mark (I)
Walsh, J.T.
Wendt, Mike
Whaley, Frank (I)
White, Brian A.
White, Jamie Christopher
Wilder, Bates (I)
Wilson, Luke (I)
Windom, William
Winifred, Dora
Woodruff Jr., Tom
Wooten, Darryl
Young, Ron G.