Solution to calculate the Bacon Number with Array Formulas in MS Excel

As promised, here we provide a working solution for the Bacon Number problem. Unfortunately we cannot record the solution in a nice video, so this is a written explanation.

What is the Kevin Bacon Number?

Kevin Bacon is an actor and has played in many movies. This Kevin Bacon Number problem solves the question how close an actor is to Kevin Bacon. Has a random actor played together with Kevin Bacon? Or did he play with somebody who played with Kevin Bacon? So there is an actor in between, or maybe there are two actors in between or three. In Hollywood there are actors who have 11 actors between Kevin Bacon and themselves.

By definition we state that Kevin Bacon has the Bacon Number 1. Everyone who played directly with Kevin Bacon has Number 2. Everyone who played with a Number 2 player has Number 3, etc. It is important to realise that each actor has just one Bacon Number, namely the highest ranked one (the lowest Number, numerically). So if you once played with a Number 2, then your Number is 3, even if you later on played with Number 4 and 5 actors. This also applies the other way around. If others with a higher Number play with you, their Number becomes your Number + 1; in this case 4. And that means that all their co-stars possibly also decrease their Bacon Number. So there is a feedback loop - The Bacon Number is never fixed, but always changing.

Secondly, also by definition, the Kevin Bacon Number is only calculated based on actors who played together in movies. So if somebody has Bacon Number 6, but he has a personal friend, who is personal friend of Kevin Bacon (so second line), or they are member of the same club, or share interests, none of that is taken into account for this problem. So only via movies and co-stars can a Bacon Number change.

How to solve the Bacon problem?

First we dive in to a very simple example; once we understand this one, we can upscale to a larger one.

In the workbook “Bacon Number - Original file.xlsx” we have 3 sheets: A sheet with actors (Table 1), a sheet with movies (Table 2) and a sheet listing which actor played in which movie (Table 3), each of which is shown below:

	A
1	**Full Name**
2	Kevin Bacon
3	Julianne Moore
4	Emma Stone
5	Denise Richards
6	Bill Murray
7	Scarlett Johansson
8	Joseph Gordon-Levitt

Table 1 – Actors on Actor sheet

	A
1	**Movie**
2	Crazy Stupid Love
3	Wild Things
4	Lost in Translation
5	Don Jon

Table 2 – Movies on Movie sheet

	A	B
1	**Name**	**Movie**
2	Kevin Bacon	Crazy Stupid Love
3	Julianne Moore	Crazy Stupid Love
4	Emma Stone	Crazy Stupid Love
5	Denise Richards	Wild Things
6	Kevin Bacon	Wild Things
7	Bill Murray	Wild Things
8	Scarlett Johansson	Lost in Translation
9	Bill Murray	Lost in Translation
10	Scarlett Johansson	Don Jon
11	Julianne Moore	Don Jon
12	Joseph Gordon-Levitt	Don Jon

Table 3 - Movie_Actor sheet, combining the actors and movies they played in.

So far, so good. This is what you have already seen multiple times in the videos. We can make the relations in such a small data set by hand as follows:

Kevin Bacon is Number 1 by definition. Then Julianne Moore, Emma Stone, Denise Richards and Bill Murray are Number 2 since they all play together with Kevin Bacon. Scarlett Johansson is related to Kevin Bacon via Bill Murray and Julianne Moore. The Bacon Number is the shortest connection to Kevin Bacon, so we need to check if the route via Bill is shorter, having a higher ranked Bacon Number, than Julianne. So we are looking for the minimum Bacon Number of Bill and Julianne. Since they both have Number 2, there is no difference. And Scarlett's number will be 2 + 1 = 3. Finally Joseph Gordon-Levitt is only connected to Kevin via Julianne and therefore gets Bacon Number 3.

As we see, for Scarlett Johansson there are two possible routes to Kevin Bacon, via Bill and via Julianne. The more actors and movies involved, the more possible routes you will find. However, we are looking for the shortest one, which gives us the highest ranked Bacon Number. This is exactly the challenge where the previous solutions failed. So how to solve this?

First we can indicate Kevin Bacon in Table 1. We add in cell B2 the formula =IF(A2="Kevin Bacon",1,999) and drag this down to B8 using the fill handle. The result is shown below in Table 4:

	A	B
1	**Full Name**	**IsBacon?**
2	Kevin Bacon	1
3	Julianne Moore	999
4	Emma Stone	999
5	Denise Richards	999
6	Bill Murray	999
7	Scarlett Johansson	999
8	Joseph Gordon-Levitt	999

Table 4 – Actor sheet

Now we go to the Movie_Actor sheet and add column C where we look up the Bacon Number of each actor based on the results of the analysis in the previous step. So we add in cell C2 the formula =VLOOKUP(A2,Actor!A:B,2,FALSE) and we drag this down to C12 using the fill handle. The result is shown in Table 5, below:

	A	B	C
1	**Name**	**Movie**	**Initial**
2	Kevin Bacon	Crazy Stupid Love	1
3	Julianne Moore	Crazy Stupid Love	999
4	Emma Stone	Crazy Stupid Love	999
5	Denise Richards	Wild Things	999
6	Kevin Bacon	Wild Things	1
7	Bill Murray	Wild Things	999
8	Scarlett Johansson	Lost in Translation	999
9	Bill Murray	Lost in Translation	999
10	Scarlett Johansson	Don Jon	999
11	Julianne Moore	Don Jon	999
12	Joseph Gordon-Levitt	Don Jon	999

Table 5 – Movie_Actor sheet

Now we see, for example, that 'Crazy Stupid Love' has Bacon Number 1 and Bacon Number 999. We are only interested in the lowest possible number, since this indicates the shortest path to Kevin Bacon.

So what we need is a function that looks up the minimum value belonging to a movie. Note that we cannot do this with a VLOOKUP function, since the VLOOKUP assumes that every movie only occurs once in the list. It is not taken into account that there are multiple hits, so it only returns the first hit and we never know if this is the minimum value. What we actually need is a function that takes into account all the possible hits and evaluates the minimum value. So a kind of conditional Minimum function, a Minimum if, a MINIF. However, Excel has a SUMIF, it does not have a MINIF function. Fortunately, we can now program this function ourselves by means of array formulas.

So what we want is for Excel to evaluate all the movies in column B, select all the Bacon Numbers in Column C that belong to a given movie and then determine the minimum Bacon Number. To do that, we go to the sheet "Movie" and add in cell B2 the formula =MIN(IF(A2=Movie_Actor!$B$2:$B$12,Movie_Actor!$C$2:$C$12,999)) and drag this formula to B5 (see Table 6). Don't forget to hit ctrl+shift+enter to enter this formula, which will add the outer curly brackets automatically.

The middle part of the formula, IF(A2=Movie_Actor!$B$2:$B$12,Movie_Actor!$C$2:$C$12,999) checks every cell in the range Movie_Actor!B2:B12 to see if this is the movie on the current row (in this example, row 2). If so, it returns the Bacon Number from column C in the Movie_Actor sheet, and if not it returns an unrealistically high value, namely 999. So this sub-step results in an array with 11 numbers entries/ numbers, being the Bacon Number of the movie if the movie name equals the

searched movie, and 999 for movies other than the searched one. In the outer part of the formula we determine the MIN value of the array returned by the inner part, so if we look again to C2:C12 in Table 5, we see that for the movies Crazy Stupid Love and Wild Things, this minimum value should be 1.

	A	B
1	**Movie**	**Step 1**
2	Crazy Stupid Love	1
3	Wild Things	1
4	Lost in Translation	999
5	Don Jon	999

Table 6 – Movie sheet

Now we are very close. We indicated Kevin Bacon in the Actor sheet, we assigned the Bacon Numbers of every actor to the movies in the Movie_Actor sheet, then we looked up for each movie the minimum Bacon Number and added this number to the Movie sheet.

So now we know that Crazy Stupid Love and Wild Things have Bacon Number 1. So the next thing is to feed this knowledge back to the Bacon Numbers of the actors. We know that Julianne Moore, Bill Murray, Emma Stone and Denise Richards played together with Kevin Bacon in these movies and therefore should have Bacon Number 2.

So we go back to the Movie_Actor sheet and we add an extra formula in the range D2:D12. Here we look up the Bacon Number of the movie from column B of the Movie sheet. The formula in D2 is =VLOOKUP(B2,Movie!$A$1:$B$5,2,FALSE)+1 and we drag this down using the fill handle. Note that we add 1 to the result of the VLOOKUP which increases the Bacon Number by 1 reflecting the step between Kevin Bacon and each actor. As you see in Table 7, all the actors who played in Crazy Stupid Love and Wild Things received Number 2, including Kevin Bacon (that are we going to repair later on).

	A	B	C	D
1	**Name**	**Movie**	**Initial**	**Return step 1**
2	Kevin Bacon	Crazy Stupid Love	1	2
3	Julianne Moore	Crazy Stupid Love	999	2
4	Emma Stone	Crazy Stupid Love	999	2
5	Denise Richards	Wild Things	999	2
6	Kevin Bacon	Wild Things	1	2
7	Bill Murray	Wild Things	999	2
8	Scarlett Johansson	Lost in Translation	999	1000
9	Bill Murray	Lost in Translation	999	1000
10	Scarlett Johansson	Don Jon	999	1000
11	Julianne Moore	Don Jon	999	1000
12	Joseph Gordon-Levitt	Don Jon	999	1000

Table 7 – Movie_Actor sheet

In the Actor sheet we now look up the Bacon Number of the actor from column D in the Movie_Actor sheet. Note that we cannot use a VLOOKUP function, because each actor may have multiple and different Bacon Numbers; compare Julianne Moore in cells D3 and D11 where she has Bacon Number 2 for the movie Crazy Stupid Love and Number 1000 for the movie Don Jon. So since we require the minimum value, we have to apply a similar formula as we did before in the movie sheet. We again want to have a minimum if function, to return the minimum Bacon Number belonging to a specific actor.

Therefore we apply in Actor!C2 the formula `=MIN(B2,MIN(IF(A2=Movie_Actor!$A$2:$A$12,Movie_Actor!$D$2:$D$12,999)))` and drag this down to C8 using the fill handle. Again, ensure you use ctrl+shift+enter.the results are shown below in Table 8:

	A	B	C
1	**Full Name**	**IsBacon?**	**Bacon step 1**
2	Kevin Bacon	1	1
3	Julianne Moore	999	2
4	Emma Stone	999	2
5	Denise Richards	999	2
6	Bill Murray	999	2
7	Scarlett Johansson	999	999
8	Joseph Gordon-Levitt	999	999

Table 8 – Actor sheet

The IF(A2=Movie_Actor!$A$2:$A$12,Movie_Actor!$D$2:$D$12,999) part of the formula returns an array of all the Bacon Numbers in Movie_Actor sheet column D where the actor name in the Movie_Actor sheet equals the actor name in A2, and 999 otherwise. So for Julianne Moore this part will return {999, 2, 999, 999, 999, 999, 999, 999 , 999, 1000, 999}. Now we take the minimum value of this array using MIN(IF(A2=Movie_Actor!$A$2:$A$12,Movie_Actor!$D$2:$D$12,999)) which results in the single number 2. Finally, we make sure that we did not find a lower Bacon Number in an earlier step. So we pick the minimum from the number in column B and the minimum number from the array. So in case of Julianne Moore this equates to =MIN(999, 2) and the result is 2. Similarly for Kevin Bacon we would end up with = MIN(1,2) which is equal to 1, which we know is correct!

Now we have one complete step, giving us all the Number 2 actors. I guess you already know how we can now find the actors with Bacon Number 3! Indeed, we are going to repeat all the sub-steps.

So first we add a column to Movie_Actor where in cell E2 we add the formula =VLOOKUP(A2,Actor!A:D,3,FALSE). We find the result is as shown below in Table 9:

	A	B	C	D	E
1	**Name**	**Movie**	**Initial**	**Return step 1**	**Step 2**
2	Kevin Bacon	Crazy Stupid Love	1	2	1
3	Julianne Moore	Crazy Stupid Love	999	2	2
4	Emma Stone	Crazy Stupid Love	999	2	2
5	Denise Richards	Wild Things	999	2	2
6	Kevin Bacon	Wild Things	1	2	1
7	Bill Murray	Wild Things	999	2	2
8	Scarlett Johansson	Lost in Translation	999	1000	999
9	Bill Murray	Lost in Translation	999	1000	2
10	Scarlett Johansson	Don Jon	999	1000	999
11	Julianne Moore	Don Jon	999	1000	2
12	Joseph Gordon-Levitt	Don Jon	999	1000	999

Table 9 – Movie_Actor sheet

Next we go to the Movie sheet and determine the minimum Bacon Number for each movie. So cell C2 gets the formula =MIN(IF(A2=Movie_Actor!$B$2:$B$12,Movie_Actor!$E$2:$E$12,999)) (remember to use ctrl+shift+enter). The result is shown below in Table 10:

	A	B	C
1	**Movie**	**Step 1**	**Step2**
2	Crazy Stupid Love	1	1
3	Wild Things	1	1
4	Lost in Translation	999	2
5	Don Jon	999	2

Table 10 – Movie sheet

Again we assign the Bacon Number of the movie to the movie_actor sheet by entering a formula in cell F2 and dragging it down to F12 using the fill handle. In cell F2 the formula reads: =VLOOKUP(B2,Movie!$A$1:$C$5,3,FALSE)+1 and we end up with the results shown below in Table 11:

	A	B	C	D	E	F
1	**Name**	**Movie**	**Initial**	**Return step 1**	**Step 2**	**Return step 2**
2	Kevin Bacon	Crazy Stupid Love	1	2	1	2
3	Julianne Moore	Crazy Stupid Love	999	2	2	2
4	Emma Stone	Crazy Stupid Love	999	2	2	2
5	Denise Richards	Wild Things	999	2	2	2
6	Kevin Bacon	Wild Things	1	2	1	2
7	Bill Murray	Wild Things	999	2	2	2
8	Scarlett Johansson	Lost in Translation	999	1000	999	3
9	Bill Murray	Lost in Translation	999	1000	2	3
10	Scarlett Johansson	Don Jon	999	1000	999	3
11	Julianne Moore	Don Jon	999	1000	2	3
12	Joseph Gordon-Levitt	Don Jon	999	1000	999	3

Table 11 – Movie_Actor sheet

Finally, we go the Actor sheet and in cell D2 apply the formula =MIN(C2,MIN(IF(A2=Movie_Actor!$A$2:$A$12,Movie_Actor!$F$2:$F$12,999))) and drag this down to D5. Finally, we have successfully found all the Bacon Numbers of the actors in this small example, as shown below in Table 12.

	A	B	C	D
1	**Full Name**	**IsBacon?**	**Bacon step 1**	**Bacon step 2**
2	Kevin Bacon	1	1	1
3	Julianne Moore	999	2	2
4	Emma Stone	999	2	2
5	Denise Richards	999	2	2
6	Bill Murray	999	2	2
7	Scarlett Johansson	999	999	3
8	Joseph Gordon-Levitt	999	999	3

Table 12 – Final result showing al Bacon Numbers in the Actor sheet.

Conclusion

This method is very cumbersome. The solution is not straightforward, but needs iterations. The Bacon Number can change at each step because of new found shorter connections. Furthermore, we loop over three tables in each step and connect those with VLOOKUP and “MINIF” functions. Finally, for every Bacon Step, we need extra columns to evaluate them. So the solution is far from optimal, since we have to know in advance what the maximum Bacon Number is and then add that much of columns to actually find it.

Other approaches are more efficient, for example databases such as Neo4j which are really designed to find relations like these, or scripts in programming languages such as Python to evaluate this differently. (Note that our Python solution is still very excel like and could be way more efficient if you do it completely outside of Excel).

A more efficient Excel approach for the above solution?

It is possible to make a more efficient solution completely within Excel and solve the Bacon Number for the actors in one single table, namely the Movie_Actor table, and with just two columns, relating to each other. We challenge you to do it yourself.

What you can do is define two columns: Column C is the Bacon Number of the actor and Column D is the Bacon Number of the movie. Then in Column C we enter a formula that says that if the actor in Column A is Kevin Bacon, the Bacon Number should be 1 by definition. If not, then it should be the lowest Bacon Number of the movies (plus one) in which this actor has played. In Column D we enter a formula that says that the Bacon Number of the movie equals the lowest Bacon Number of the actors who played in it, in other words, the minimum value in Column C for the rows that relate to that movie.

If we enter these formulas, we will get a circular reference error, since the solution of Column C is based on the solution of Column D and the Solution of Column D is based on the solution of Column C. This is only possible if you allow MS Excel to calculate the solution in an iterative manner. This is done in options → file → formulas → Enable Iterative Calculation, as shown in Figure 1, below. Make sure that you only allow Excel to calculate everything when you decide to by enabling manual calculations (also shown below in Figure 1). You can now tell Excel to run a calculation by hitting the F9 key.

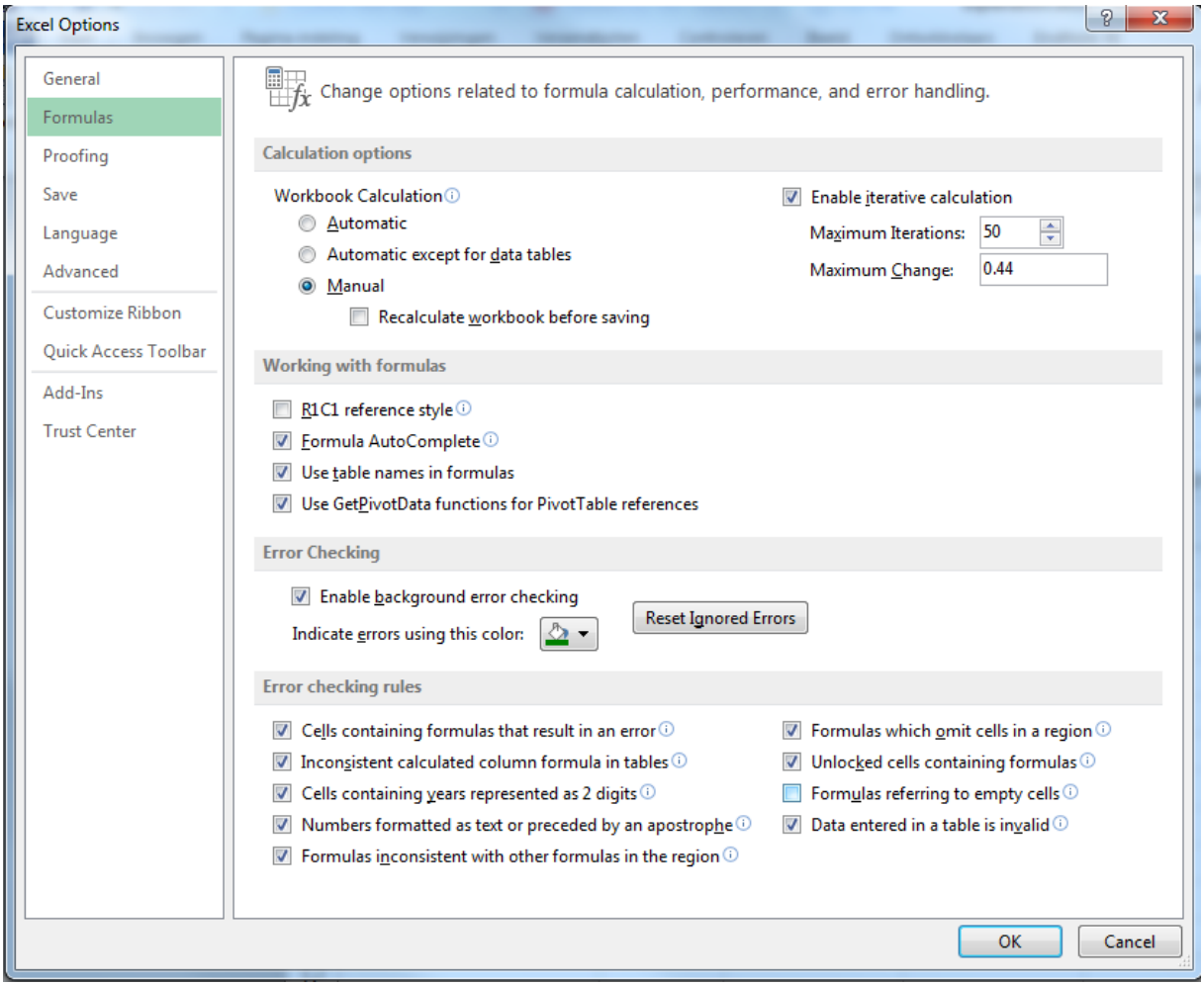

Figure 1 - Enable iterative calculations and make sure that you choose to only calculate the spreadsheet manually (by hitting F9).

Don't worry if you can't figure out this solution, as we have provided it in the workbook "Bacon Number - Alternative solution on Extended list.xlsx" so you can dive into it yourself. Make sure you have changed the settings shown in Figure 1 above, as these settings are not stored in the workbook. Be aware, if there are actors who are not connected by any number of steps to Kevin Bacon, their Bacon Number will increase indefinitely for each iteration. We did not solve this issue in this spreadsheet, as the formulas would become unreadable and incomprehensible. If you are up for a real challenge, try and solve the issue for yourself!

Good luck!