CSCI 3412 – Algorithms
Spring 2018 – Dr. Williams
Program 3
Due Wednesday, April 11

For this programming assignment, we will be using a data set from the Social Characteristics of the Marvel Universe web page(s). Specifically, the input file (porgat.txt) defines the social network we will examine.

Here is what the input file looks like.

```
*Vertices 19428 6486
1 "24-HOUR MAN/EMMANUEL"
2 "3-D MAN/CHARLES CHAN"
3 "4-D MAN/MERCURIO"
4 "8-BALL/"
5 "A"
6 "A'YIN"
7 "ABBOTT, JACK"
8 "ABCISSA"
9 "ABEL"
10 "ABOMINATION/EMIL BLO"
…
6484 "STORMER"
6485 "TIGER WYLDE"
6486 "ZONE"
6487 "AA2 35"
6488 "M/PRM 35"
6489 "M/PRM 36"
6490 "M/PRM 37"
6491 "WI? 9"
6492 "AVF 4"
6493 "AVF 5"
…
19426 "AA2 30"
19427 "AA2 20"
19428 "AA2 38"
*Edgeslist
1 6487
2 6488 6489 6490 6491 6492 6493 6494 6495 6496
3 6497 6498 6499 6500 6501 6502 6503 6504 6505
4 6506 6507 6508
5 6509 6510 6511
6 6512 6513 6514 6515
7 6516
8 6517 6518
9 6519 6520
10 6521 6522 6523 6524 6525 6526 6527 6528 6529 6530 6531 6532 6533 6534 6535
10 6536 6537 6538 6539 6540 6541 6542 6543 6544 6545 6546 6547 6548 6549 6550
10 6551 6552 6553 6554 6555 6556 6557 6558 6559 6560 6561 6562 6563 6564 6565
…
6484 18709
6485 15336
6486 15336
```

The “`*Vertices 19428 6486`” line says there are 19,428 total vertexes in the input file and that the first 6,486 of them are character vertexes and the remaining are comic book vertexes. Each vertex line contains the vertex number and a string with the character name or comic book title. The “`*Edgeslist`” line begins the edge list. Each edge line contains a source (character) vertex and one or more (apparently up to 15) target (comic book) vertexes.

Part I – Representing the Social Graph

Since the graph is a fixed size (and not too large) and all of the references are already integer indexes, an array representation is the most natural to use. To represent the information in the input file, we can use arrays of strings to hold the vertex names – character names and comic book titles – and another two dimensional array of Booleans to hold the edges. [Due to the nature of the graph, there are large blocks of it that will always be false and it is symmetric. Therefore, we could only store part of the graph if we wanted to save space.]

The data in the input file uses one set of integers for both characters and comic books. You can keep these values or separate them into separate integer ranges for characters and comic books. [Assuming you are using C++, you might also want to change them into zero-based ranges.]

So, you need to be able to read and parse the input file and create the array of vertex names and the edge matrix.

Part II - Marvel Universe looks almost like a real social network

There is a paper on arXiv.org titled Marvel Universe looks almost like a real social network that looks at the structure of this network. Here, two characters are considered linked if they jointly appear in some comic book.

We can make a collaboration matrix, which is a square (or diagonal if we prefer) matrix of integers where each cell (i, j) is a count of the number of comic books in which the pair of characters have jointly appeared. We can compute this by taking the ith and jth row (or column), anding them together to get a vector of Booleans where each true value is an instance where the two characters appear in the same comic book, then summing the number of true values as the value of the cell (i, j). If you apply this to every pair of characters, then the diagonal will contain the counts of the number of comics that each character occurs. Using this collaboration matrix, we can calculate the total number of collaborations and the number of collaborating pairs of characters. [Be careful to exclude the diagonal from the counts – we don't consider a person as collaborating with themselves – and symmetry – if x collaborates with y then that is stored twice (so essentially, we divide the final count(s) by 2.]

Part III – Six degrees of Spider Man

Six degrees of Kevin Bacon is a game based on the “six degrees of separation” concept, which posits that any two people on Earth are six or fewer acquaintance links apart. The idea eventually morphed into this game where movie buffs challenge each other to find the shortest path between an arbitrary actor and Kevin Bacon. Therefore, the Bacon number of an actor or actress is the degrees of separation he or she has from Kevin Bacon, as defined by the game.

For this programming assignment, we want to develop a program to compute the Spider Man number for any arbitrary Marvel character.

The computation of a Spider Man number for character X is a shortest path algorithm, applied to the Marvel social graph:

- Spider Man himself has a Spider Man number of 0.
- Those Marvel characters who have appeared in a comic book with Spider Man have a Spider Man number of 1.
- If the lowest Spider Man number of any Marvel character with whom X has appeared in any comic book is N, X's Spider Man number is N+1.

Part IV – Specifics

Each student will write a program to compute the Spider Man numbers of a set of Marvel characters. Specifically, the program should take as input an arbitrary number of Marvel characters – either as integer indexes (easiest) or strings – and for each of these characters, print the characters name and their Spider Man number.

I will provide you a list of names for test input.