

Programming Assignment 5

Friendship Graph Algorithms

In this assignment, you will implement some useful algorithms that apply to friendship graphs of the Facebook kind.

Worth 80 points = 8% of your course grade

Posted Wed, Nov 21

Due Fri, Dec 7, 11:00 PM (**WARNING!! NO GRACE PERIOD**)

Extended deadline (with ONE time free extension pass): Mon, Dec 10, 11:00 PM (**NO GRACE PERIOD**)

You get ONE free extension pass for assignments during the semester, no questions asked. There will be a total of 5 assignments this semester, and you may use this one free extension pass for any of the 5 assignments.

A separate Sakai assignment will be opened for extensions AFTER the deadline for the regular submission has passed. The extension will be 3 days (72 hours). If/when you choose to use the one-time extension, you don't need to ask for permission - just drop your submission in the extension assignment. And you can do this even if you dropped something in the regular submission, and later on decided to use the extension - in this case, only the extension submission will be graded, the regular submission will be ignored.

-
- You will work on this assignment individually. Read [DCS Academic Integrity Policy for Programming Assignments](#) - you are responsible for abiding by the policy. In particular, note that "**All Violations of the Academic Integrity Policy will be reported by the instructor to the appropriate Dean**".

- **IMPORTANT - READ THE FOLLOWING CAREFULLY!!!**

Assignments emailed to the instructor or TAs will be ignored--they will NOT be accepted for grading. We will only grade submissions in Sakai.

If your program does not compile, you will not get any credit.

Most compilation errors occur for two reasons:

1. You are programming outside Eclipse, and you delete the "package" statement at the top of the file. If you do this, you are changing the program structure, and it will not compile when we test it.
2. You make some last minute changes, and submit without compiling.

To avoid these issues, (a) **START EARLY**, and give yourself plenty of time to work through the assignment, and (b) **Submit a version well before the deadline so there is at least something in Sakai for us to grade**. And you can keep submitting later versions (up to 10) - we will accept the **LATEST** version.

-
- [Background](#)
 - [Algorithms](#)
 - [Implementation](#)
 - [Submission](#)

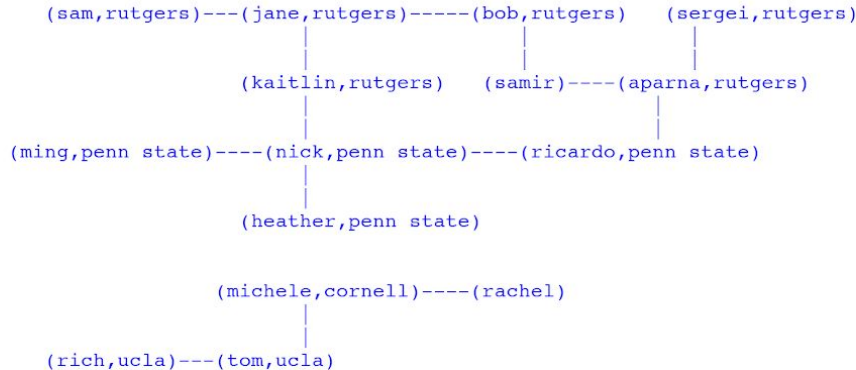
Background

In this program, you will implement some useful algorithms for graphs that represent friendships (e.g. Facebook). A friendship graph is an undirected graph without any weights on the edges. It is a simple graph because there are no self

loops (a self loop is an edge from a vertex to itself), or multiple edges (a multiple edge means more than edge between a pair of vertices).

The vertices in the graphs for this assignment represent two kinds of people: students and non-students. Each vertex will store the name of the person. If the person is a student, the name of the school will also be stored.

Here's a sample friendship graph:



Note that the graph may not be connected, as seen in this example in which there are two "islands" or cliques that are not connected to each other by any edge. Also see that all the vertices represent students with names of schools, except for `rachel` and `samir`, who are not students.

Algorithms

1. Shortest path: Intro chain

sam wants an intro to aparna through friends and friends of friends. There are two possible chains of intros:

sam--jane--kaitlin--nick--ricardo--aparna

or

sam--jane--bob--samir--aparna

The second chain is preferable since it is shorter.

If **sam** wants to be introduced to **michele** through a chain of friends, he is out of luck since there is no chain that leads from **sam** to **michele** in the graph.

Note that this algorithm does NOT have any restriction on the composition of the vertices: a vertex along the shortest chain need NOT be a student at a particular school, or even a student. In other words, this algorithm is not about students, let alone students at a particular school. So, for instance, you may need to find the shortest path (intro chain) from **nick** to **samir**, which will be:

nick--ricardo--aparana--samir

which consists of two penn state students, one rutgers student, and one non-student.

2. Cliques: Student cliques at a school

Students tend to form cliques with their friends, which creates islands that do not connect with each other. If these cliques could be identified, particularly in the student population at a particular school, introductions could be made between people in different cliques to build larger networks of friendships at that school.

In the sample graph, there are two cliques of students at rutgers:

```

(sam,rutgers)---(jane,rutgers)----- (bob,rutgers)    (sergei,rutgers)
                |                               |
                (kaitlin,rutgers)              (aparna,rutgers)

```

Note that in the graph these are not islands since **samir** connects them. However, since **samir** is not a student at rutgers, it results in two cliques of rutgers students that don't know each other through another rutgers student.

At penn state, there is a single clique of students:

```

(ming,penn state)----(nick,penn state)----(ricardo,penn state)
                    |
                    (heather,penn state)

```

Also, a single clique of students at ucla:

```

(rich,ucla)---(tom,ucla)

```

And a single clique of students at cornell:

```

(michele,cornell)

```

3. Connectors: Friends who keep friends together

If **jane** were to leave rutgers, **sam** would no longer be able to connect with anyone else--**jane** was the "connector" who could pull **sam** in to hang out with her other friends. Similarly, **aparna** is a connector, since without her, **sergei** would not be able to "reach" anyone else. (And there are more connectors in the graph...)

On the other hand, **samir** is not a connector. Even if he were to leave, everyone else could still "reach" whoever they could when **samir** was there, even though they may have to go through a longer chain.

Definition: In an undirected graph, vertex **v** is a connector if there are at least two other vertices **x** and **w** for which every path between **x** and **w** goes through **v**.

For example, **v=jane**, **x=sam**, and **w=bob**.

Finding all connectors in an undirected graph can be done using DFS (depth-first search), by keeping track of two additional quantities for every vertex **v**. These are:

- **dfsnum(v)**: This is the dfs number, assigned when a vertex is visited, dealt out in increasing order.
- **back(v)**: This is a number that is initially assigned when a vertex is visited, and is equal to **dfsnum**, but can be changed later as follows:
 - When the DFS backs up from a neighbor, **w**, to **v**, if **dfsnum(v) > back(w)**, then **back(v)** is set to **min(back(v), back(w))**
 - If a neighbor, **w**, is already visited then **back(v)** is set to **min(back(v), dfsnum(w))**

When the DFS backs up from a neighbor, **w**, to **v**, if **dfsnum(v) ≤ back(w)**, then **v** is identified as a connector, IF **v** is NOT the starting point for the DFS.

If **v** is a starting point for DFS, it can be a connector, but another check must be made - see the examples below. The examples don't tell you how to identify such cases--you have to figure it out.

Here are some examples that show how this works.

- Example 1: (B is a connector)

```

A---B---C

```

Neighbors for a vertex are stored in adjacency linked lists like this:

```

A: B

```

```

ricardo|y|penn state
kaitlin|y|rutgers
samir|n
aparna|y|rutgers
ming|y|penn state
nick|y|penn state
bob|y|rutgers
heather|y|penn state
rachel|n
rich|y|ucla
tom|y|ucla
sam|jane
jane|bob
jane|kaitlin
kaitlin|nick
bob|samir
sergei|aparna
samir|aparna
aparna|ricardo
nick|ricardo
ming|nick
heather|nick
michele|rachel
michele|tom
tom|rich

```

The first line has the number of people in the graph (15 in this case).

The next set of lines has information about the people in the graph, one line per person (15 lines in this example), with the '|' used to separate the fields. In each line, the first field is the name of the person. Names of people can have any character except '|', and are case *insensitive*. The second field is 'y' if the person is a student, and 'n' if not. The third field is only present for students, and is the name of the school the student attends. The name of a school can have any character except '|', and is case *insensitive*. **No two people will have the same name.**

The last set of lines, following the people information, lists the friendships between people, one friendship per line. Since friendship works both ways, any friendship is only listed once, and the order in which the names of the friends is listed does not matter.

You will complete the following static methods in the `Friends` class, to implement the three algorithms in the previous section. (All of these methods take a `Graph` instance as a parameter, aside from other possible inputs detailed below.)

Methods

1. (25 pts) `shortestPath`

- Input: Name of person who wants the intro, and the name of the other person. For instance, inputs could be "sam" and "aparna" for the graph in the `Background` section. (Neither of these, nor any of the intermediate people in the chain, are required to be students, in the same school or otherwise.)
- Result: The shortest chain (list) of people in the graph starting at the first and ending at the second, returned in an array list.

For example, if the inputs are `sam` and `aparna` (`sam` wants an intro to `aparna`), then the shortest chain from `sam` to `aparna` is `[sam, jane, bob, samir, aparna]`

(This represents the path `sam--jane--bob--samir--aparna`)

If there is more than one shortest path, ANY of them is acceptable.

If there is no way to get from the first person to the second person, then the returned list is empty (null or zero-sized array list).

2. (20 pts) `cliques`

- Input: Name of school for which cliques are to be found, e.g. "rutgers"

- Result: The names of people in each of the cliques, in any order, returned in an array list of array lists. Moreover, the cliques themselves could be in any order in the top level array list.

For the example cited in the **Cliques** part of the **Algorithms** section above, with input `rutgers`, the result is:

```
[[sam,jane,bob,kaitlin],[sergei,aparna]]
```

In other words, an array list that has two cliques, each of which is an array list.

The names in the clique array list can be in any order. So, the same cliques could have been returned as:

```
[[jane,sam,kaitlin,bob],[aparna,sergei]]
```

and it would be correct.

The cliques themselves can be in any order within the top level array lists, so the following variation (for example) is also acceptable:

```
[[sergei,aparna],[sam,jane,bob,kaitlin]]
```

However, names must not be repeated in a clique.

If there are no students in the input school, the result is empty (null or zero-sized array list).

3. (35 pts) connectors

- Input: None
- Result: Names of all connectors, in any order, returned in an array list.

In the sample friendship graph of the **Background** section, the connectors list is `[jane,aparna,nick,tom,michele]`. Any other permutation of the names in the list is fine, since the order does not matter.

Names in the list must not be repeated.

Implementation Rules

Do NOT change ANY of the contents of `Graph.java`, `Queue.java`, and `Stack.java`.

In `Friends.java`, you may NOT MAKE ANY CHANGES EXCEPT to (a) fill in the body of the required methods, or (b) add private helper methods. Otherwise, your submission will be penalized.

Note: While you may find them useful, you are not required to use either of the `Stack` or `Queue` classes that are imported in `Friends.java`. Also, `java.util` defines a `Stack` class, but it will be overridden by the `structures.Stack` class that is imported.

Submission

Submit your `Friends.java` file.