

Visualizing the Aggregation of Students' Unit Test Failures

Joe Hollingsworth

Rose-Hulman Institute of Technology

May 2023

Order of Topics

- JUnit test usage overview
- How to guess at what the students are thinking while problem solving
- Visualizing an aggregation of JUnit testing results

CSSE 220 Homework Assignments Overview

IDE: Eclipse

Package Explorer × JUnit

Order Assigned

- > Other Projects [csse220Wi22 master]
- ▼ HWs [csse220Wi22 master]
 - > HW1 [csse220Wi22 master]
 - > HW2DArray [csse220Wi22 master]
 - > HWIntroToUnitTesting [csse220Wi22 master]
 - > HWLinearLightsOut [csse220Wi22 master]
 - > HWMaps [csse220Wi22 master]
 - > HWQuizQuestions [csse220Wi22 master]
 - > HWRecursion [csse220Wi22 master]
 - > HWRefactoringInheritance [csse220Wi22 master]
 - > HWScenes [csse220Wi22 master]
 - > HWSinglyLinkedList [csse220Wi22 master]
 - > HWTeamGradebook [csse220Wi22 master]
 - > HWTwelveProblems [csse220Wi22 master]
- > Practice [csse220Wi22 master]
- > Sample [csse220Wi22 master]

12 programming homeworks
Collecting data on 4 assignments

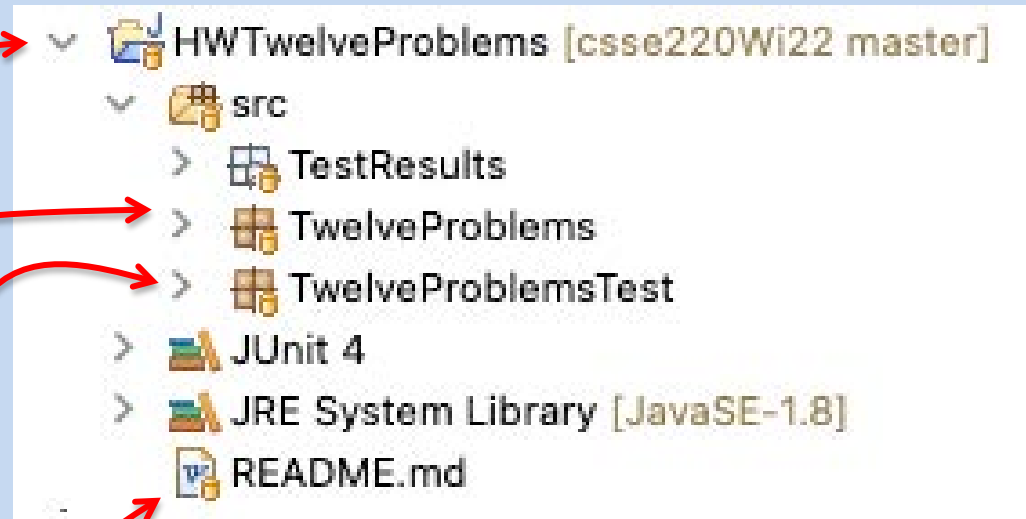
One Assignment Overview

Eclipse package

Students write code
in .java files

JUnit tests

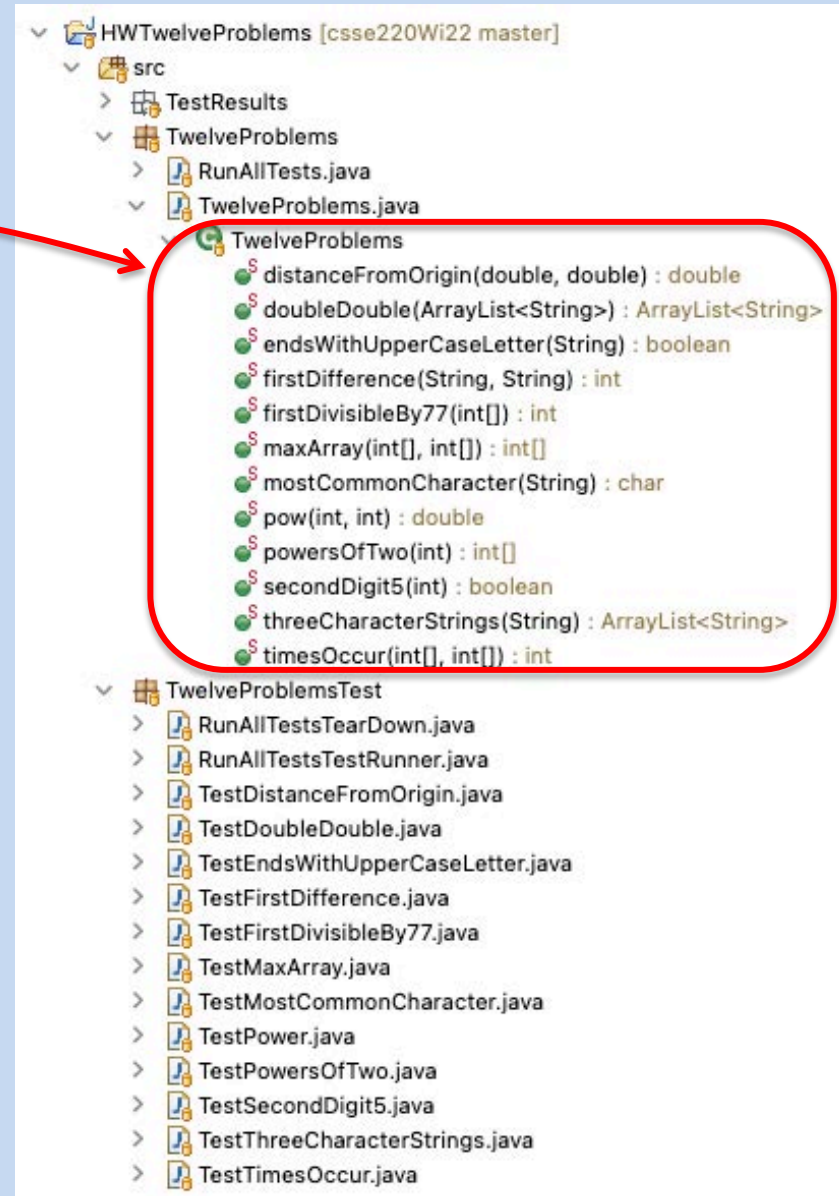
Assignment's
instructions



One Assignment Breakdown

The *12 problems*

There are 12 different operations that must be implemented

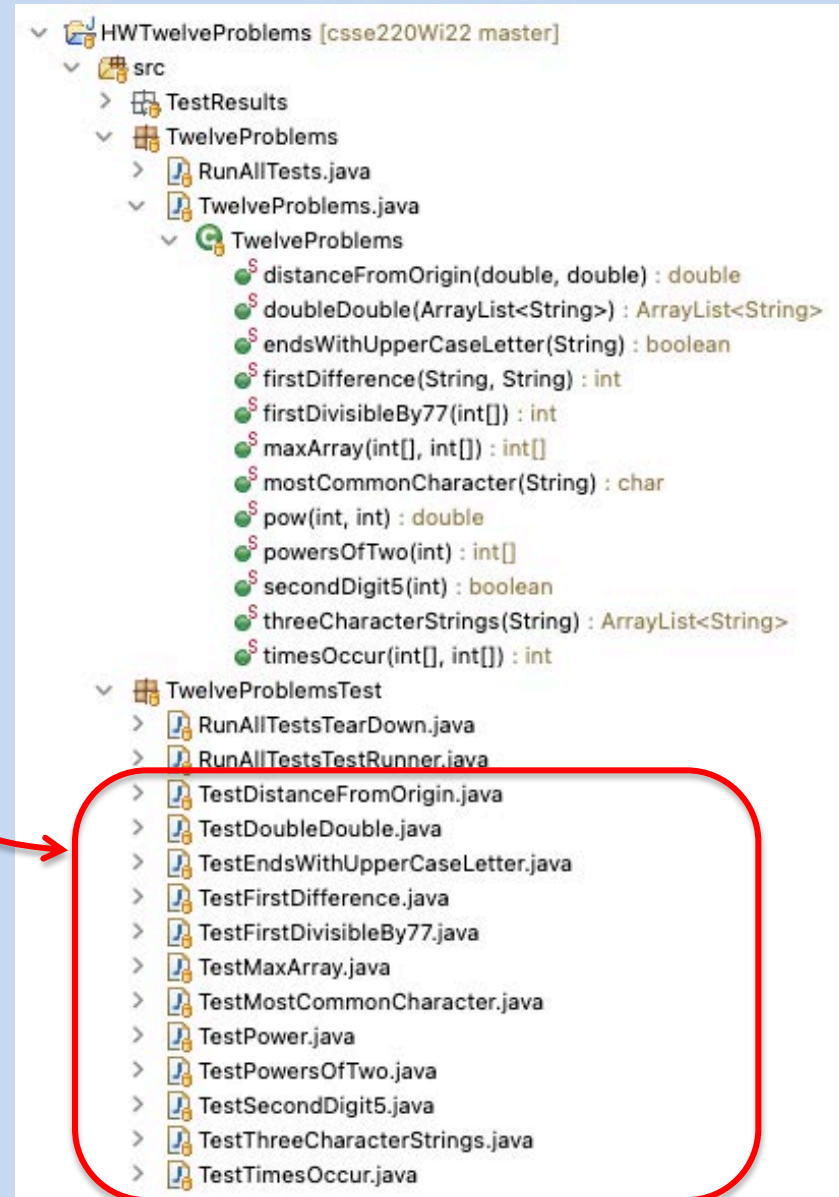


One Assignment Breakdown

The *12 problems*

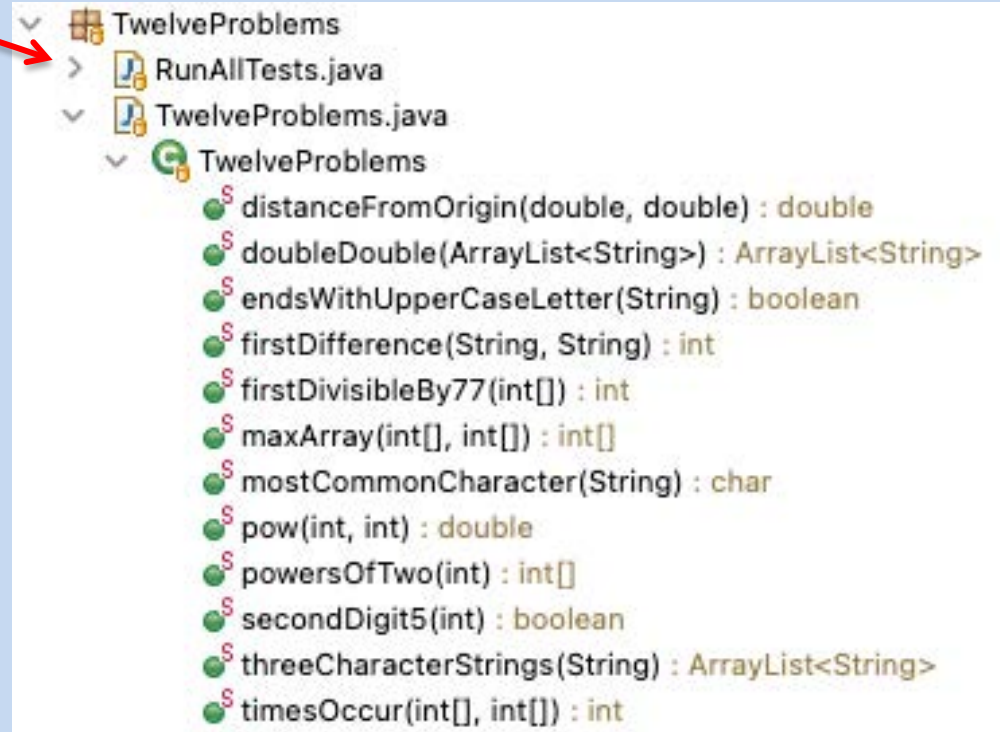
The JUnit tests for each of the 12 problems

Each JUnit test file contains multiple JUnit tests, i.e., concrete inputs with expected outputs



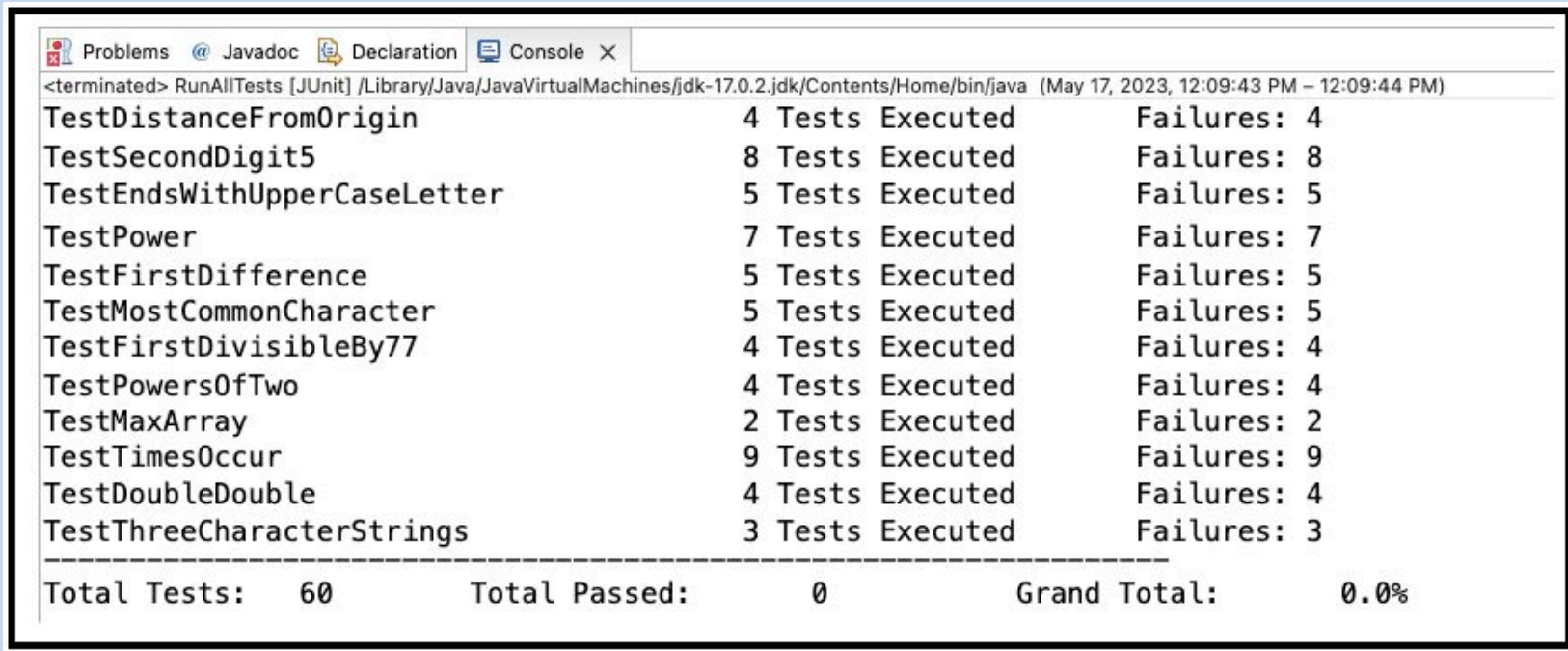
Executing JUnit Tests

A student runs this in order to to execute all the JUnit test files



Console Output Summarizes Results

The student sees the following summary in the IDE's Console output window



```
<terminated> RunAllTests [JUnit] /Library/Java/JavaVirtualMachines/jdk-17.0.2.jdk/Contents/Home/bin/java (May 17, 2023, 12:09:43 PM - 12:09:44 PM)
TestDistanceFromOrigin          4 Tests Executed      Failures: 4
TestSecondDigit5                8 Tests Executed      Failures: 8
TestEndsWithUpperCaseLetter     5 Tests Executed      Failures: 5
TestPower                        7 Tests Executed      Failures: 7
TestFirstDifference              5 Tests Executed      Failures: 5
TestMostCommonCharacter          5 Tests Executed      Failures: 5
TestFirstDivisibleBy77          4 Tests Executed      Failures: 4
TestPowersOfTwo                 4 Tests Executed      Failures: 4
TestMaxArray                    2 Tests Executed      Failures: 2
TestTimesOccur                  9 Tests Executed      Failures: 9
TestDoubleDouble                4 Tests Executed      Failures: 4
TestThreeCharacterStrings        3 Tests Executed      Failures: 3
-----
Total Tests:    60      Total Passed:    0      Grand Total:    0.0%
```

Assessment

- This JUnit approach supports assigning a grade
- Graders can run all these tests and record a score

Some Detractors to This Approach

- It makes it easy for grader/instructor to *not* look at individual submissions
- Even if grader/instructor inspects the submission, this is just a snapshot of the final submitted *draft*
- No way to see intermediate *drafts*

What If ...?

- We could collect and analyze the pass/fail results of:
 - one student's execution of all the JUnit tests for the code submitted for one operation
 - we call this a *trace*
- Or even better, an aggregation of all students' traces for one operation

What If ...?

- Maybe an analysis of these traces would reveal trends that:
 1. allow us to ascertain/guess what students are thinking while problem solving
 2. what problem solving strategies students are using
 3. what seems to be common obstacles, e.g., handling edge cases
 4. among others

Time for a Pilot Study

- Impetus for “What If” and for a Pilot Study:
 - Jason Hallstrom’s NSF grant proposal
 - Automate processing of think-aloud data while students solve BeginToReason problems
 - Attempting to gain access to what students are thinking
 - Bruce Weide’s suggestion that we use Hurtig’s visualization tool on data that we *can* collect

One Student's Trace

Culled From a Pilot Study

- Problem: *PowersOfTwo*
- 4 JUnit test cases
- Ethan ran the JUnit tests 5 times
 - The first 4 runs had at least one test case fail
 - The 5th run (last run) all 4 test cases passed

```
1 Nett Ethan, 1T 2T 3F 4F, 0
2 Nett Ethan, 1T 2T 3F 4F, 0
3 Nett Ethan, 1T 2T 3F 4F, 0
4 Nett Ethan, 1F 2F 3F 4T, 0
5 Nett Ethan, 1T 2T 3T 4T, 1
```

```
TestPowersOfTwo.java
├── TestPowersOfTwo
│   ├── testPowersOfTwoN01() : void
│   ├── testPowersOfTwoN02() : void
│   ├── testPowersOfTwoN03() : void
│   └── testPowersOfTwoN04() : void
```

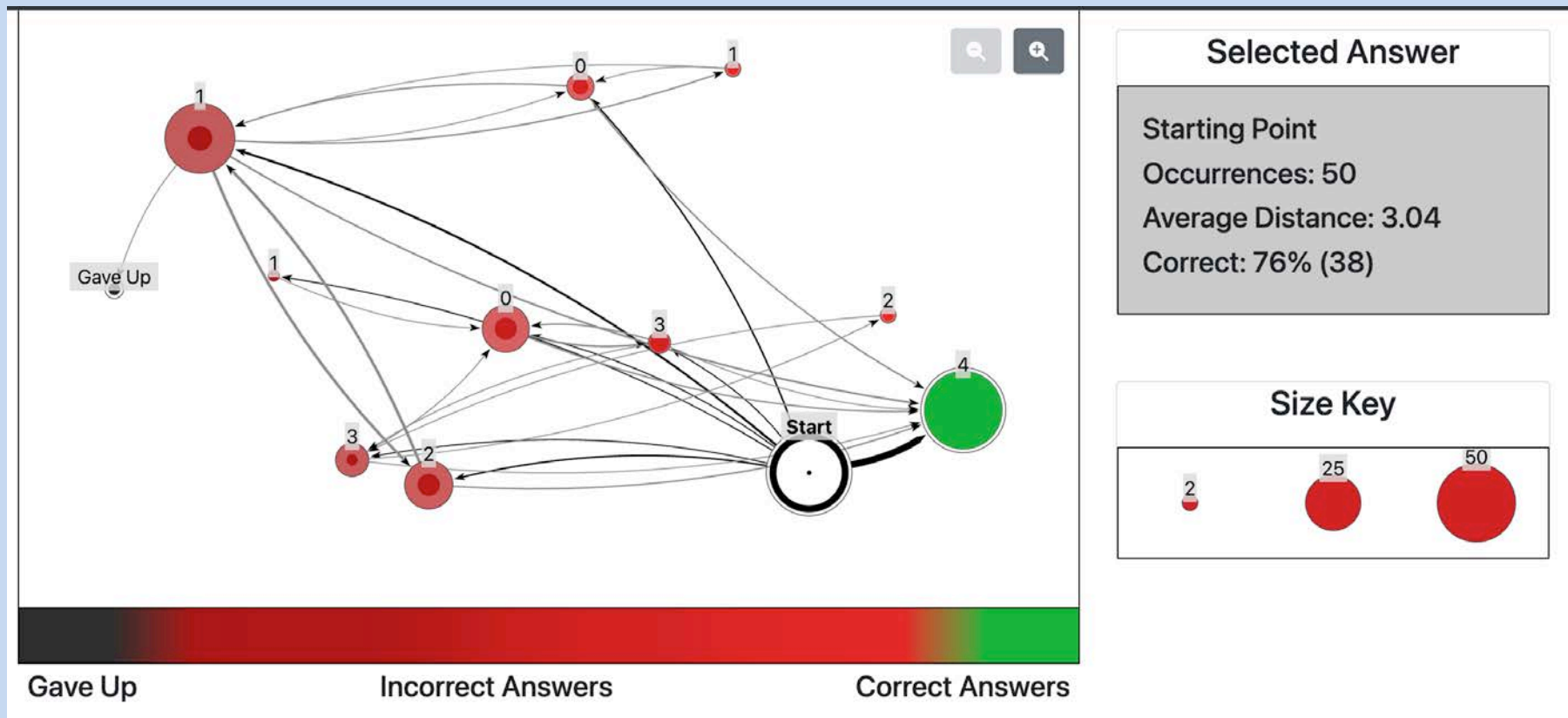

PowersOfTwo Test Cases

Test Powers of Two

1. `int[] a = { 1, 2, 4, 8 };`
`assertEquals(a, TwelveProblems.powersOfTwo(3));`
2. `int[] b = { 1, 2, 4, 8, 16, 32, 64, 128 };`
`assertEquals(b, TwelveProblems.powersOfTwo(7));`
3. `int[] c = { 1 };`
`assertEquals(c, TwelveProblems.powersOfTwo(0));`
4. `int[] d = {};`
`assertEquals(d, TwelveProblems.powersOfTwo(-200));`

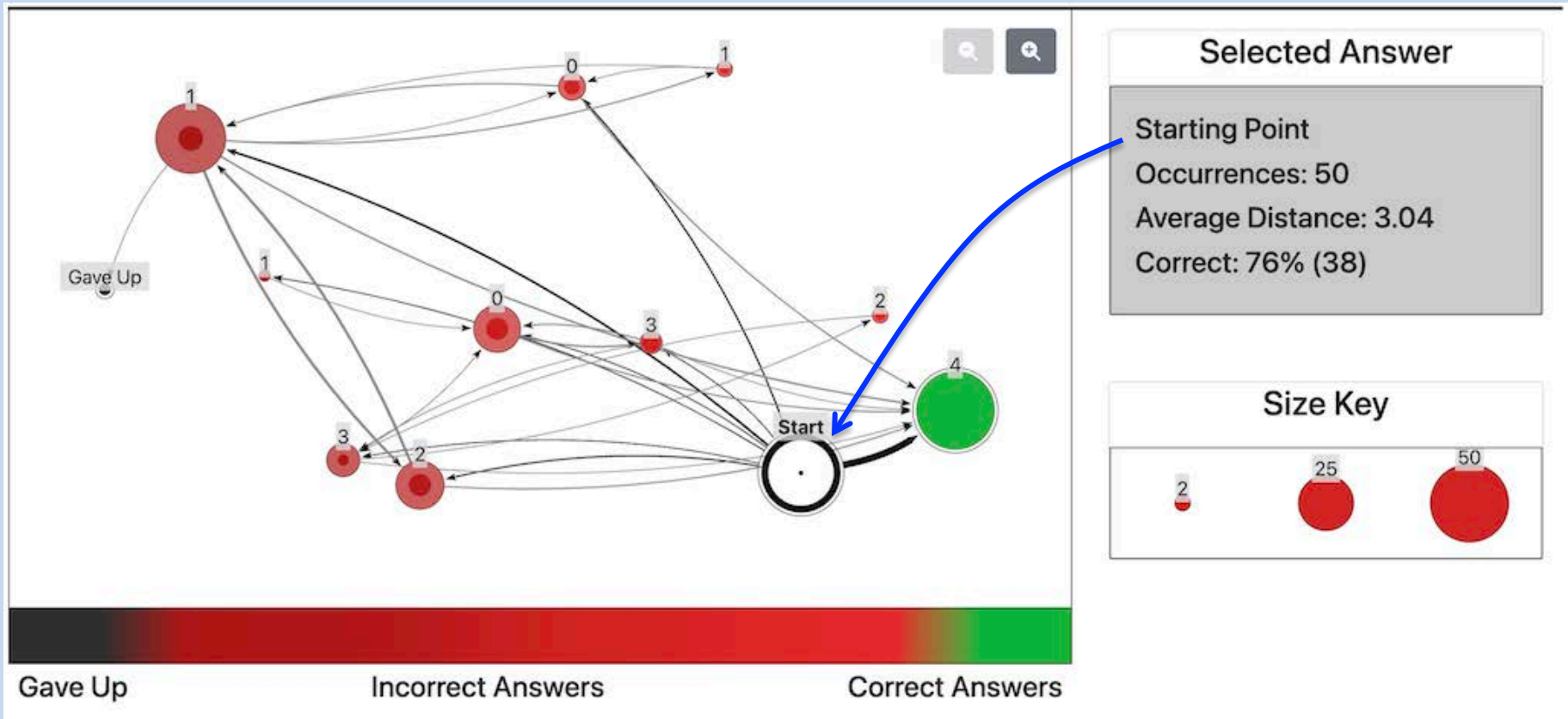
Interactive Visualization of a graph created from aggregated traces

- Tool was created by Nat Hurtig of Rose-Hulman



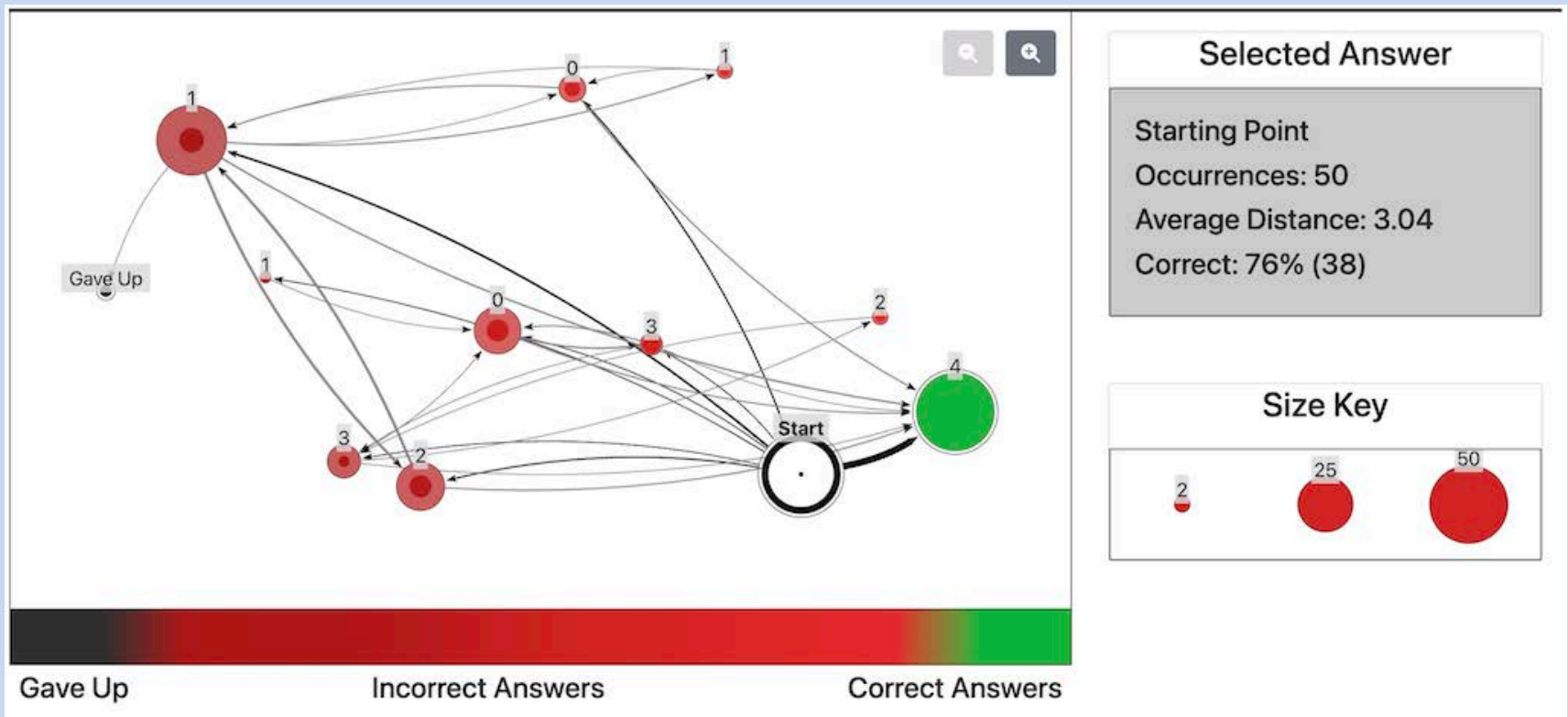
Graph Details #1

- Start state selected



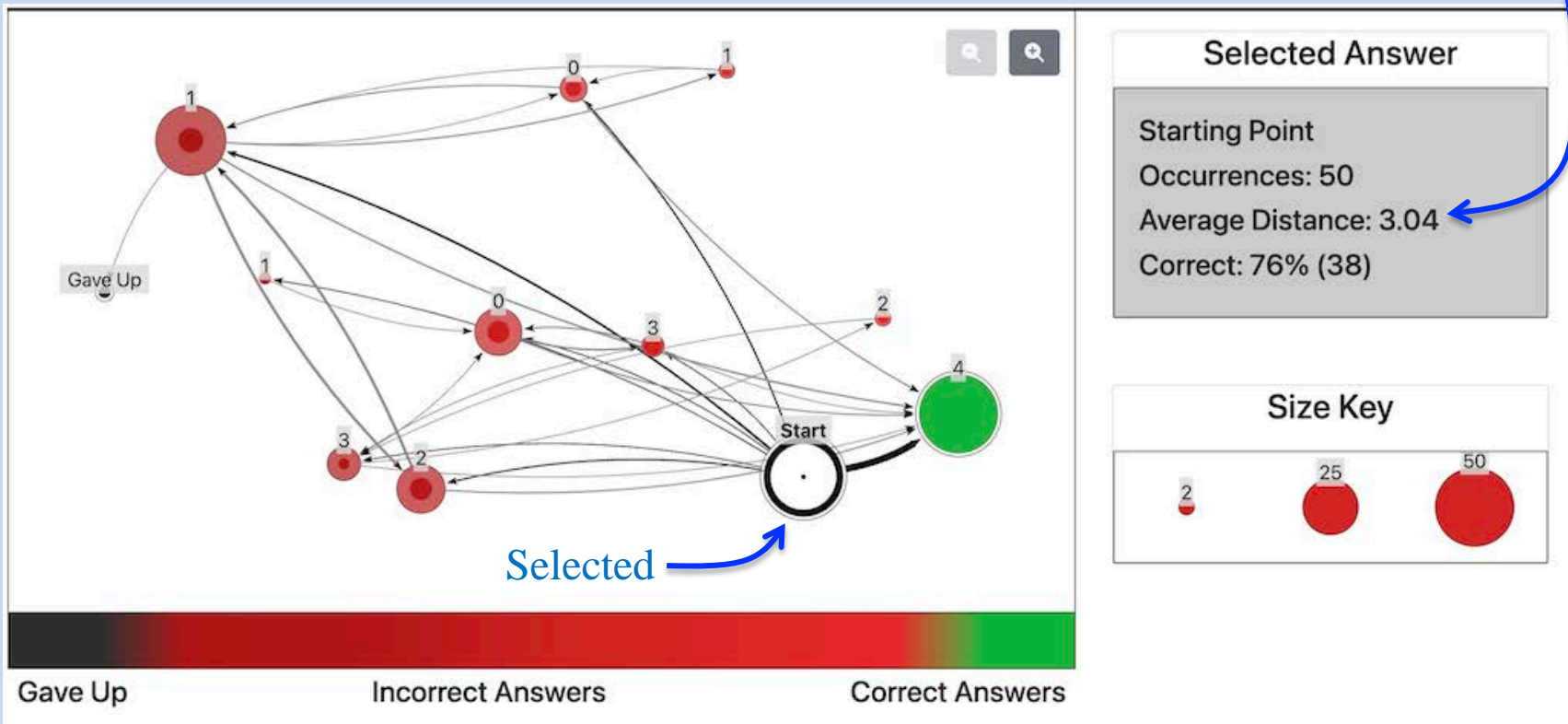
Graph Details #2

- Node color:
 - Red – one or more test cases failed
 - Green – all four test cases passed
- Node # represents how many passed



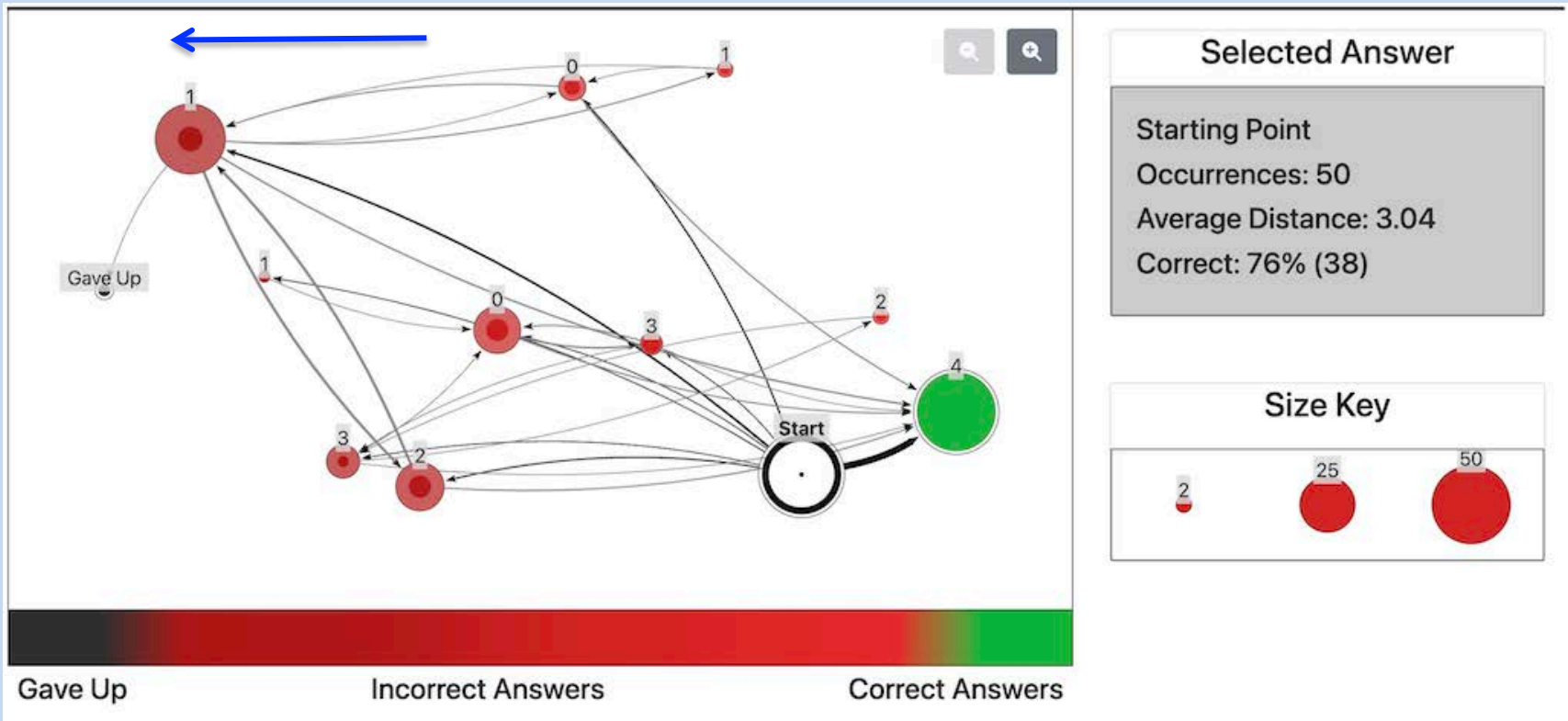
Graph Details #3

- *Average Distance* of a selected node:
 - average # of runs required to get to a green node



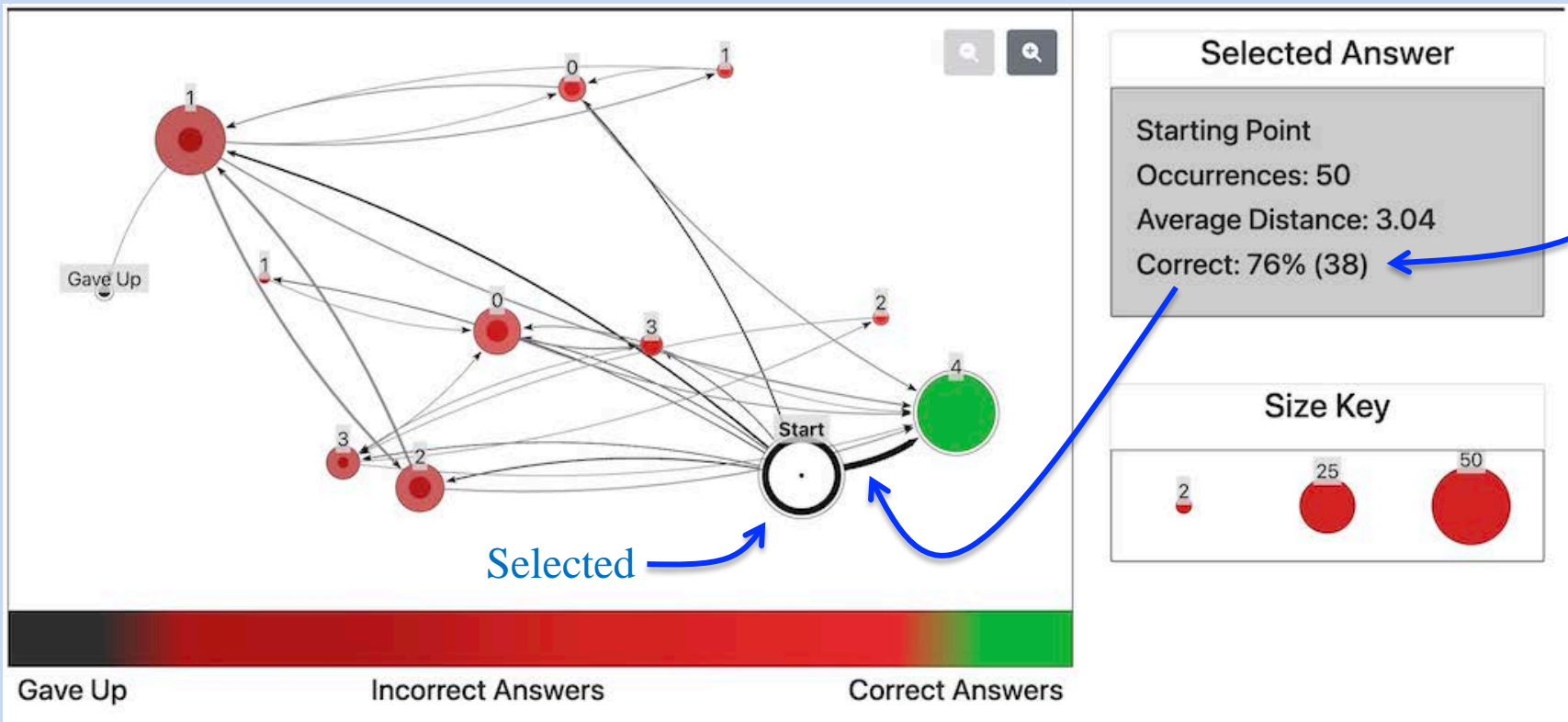
Graph Details #4

- Graph layout:
 - red nodes farther to the left have a greater average distance to a green node



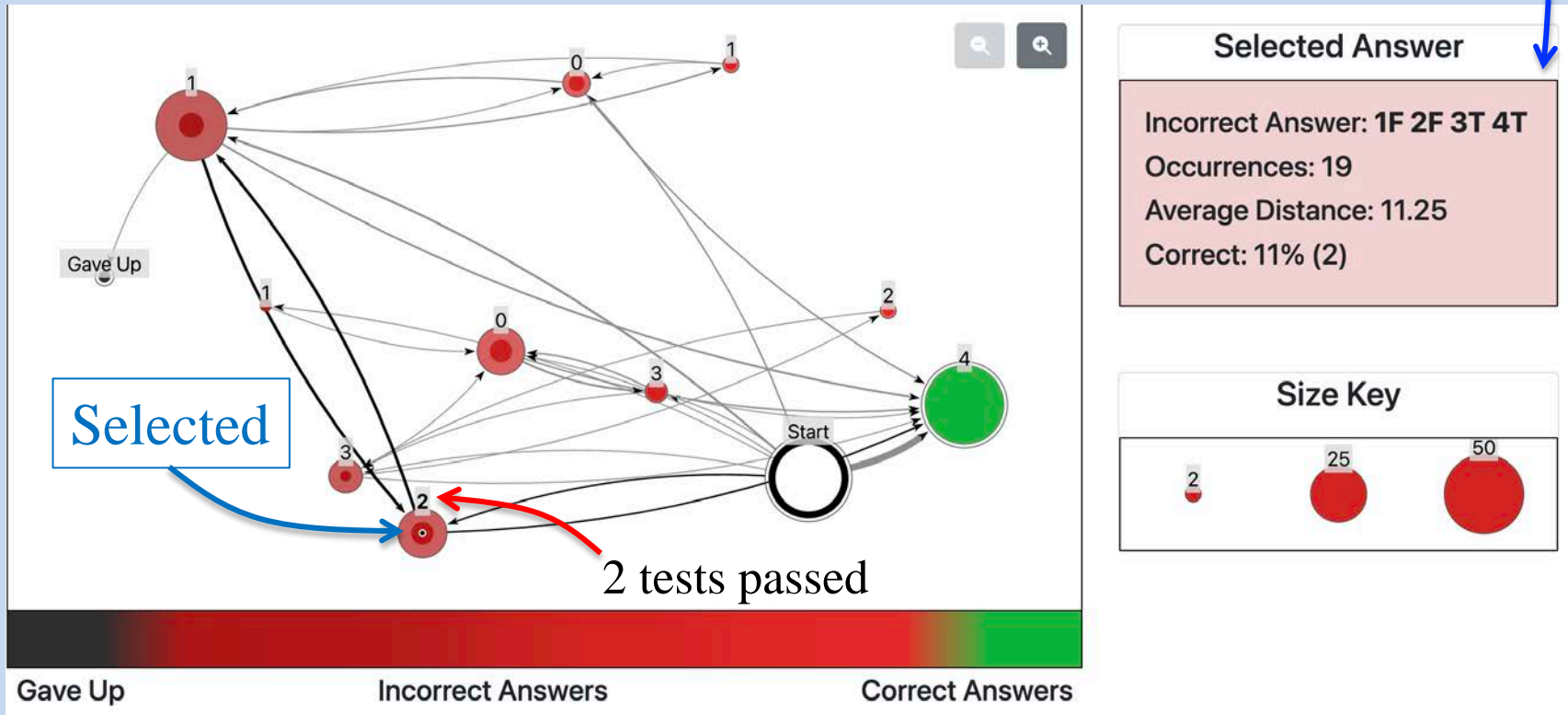
Graph Details #5

- *Correct* of a selected node:
 - % (#) of students that went to a green node on next run of tests



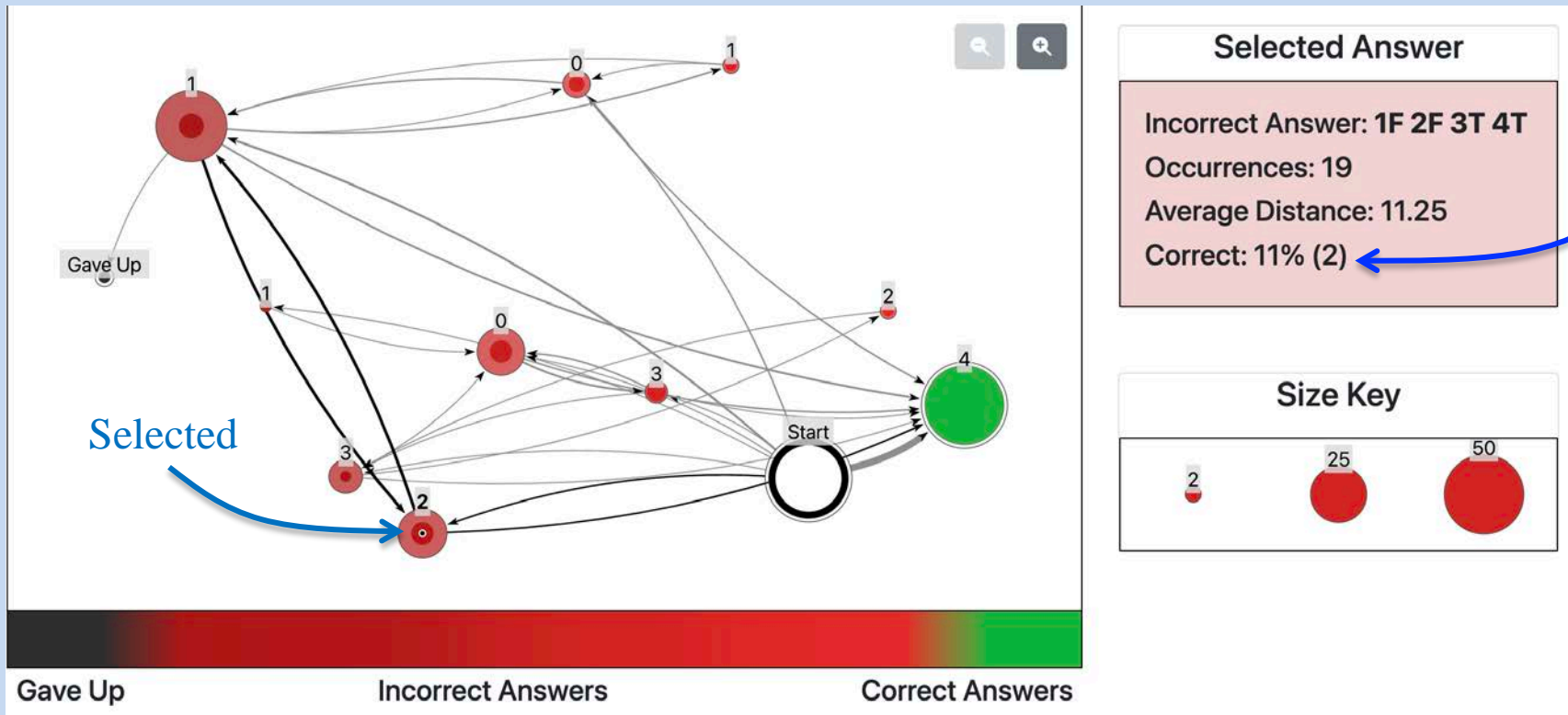
Graph Details #6

- A red node selected: 2 tests out of 4 passed



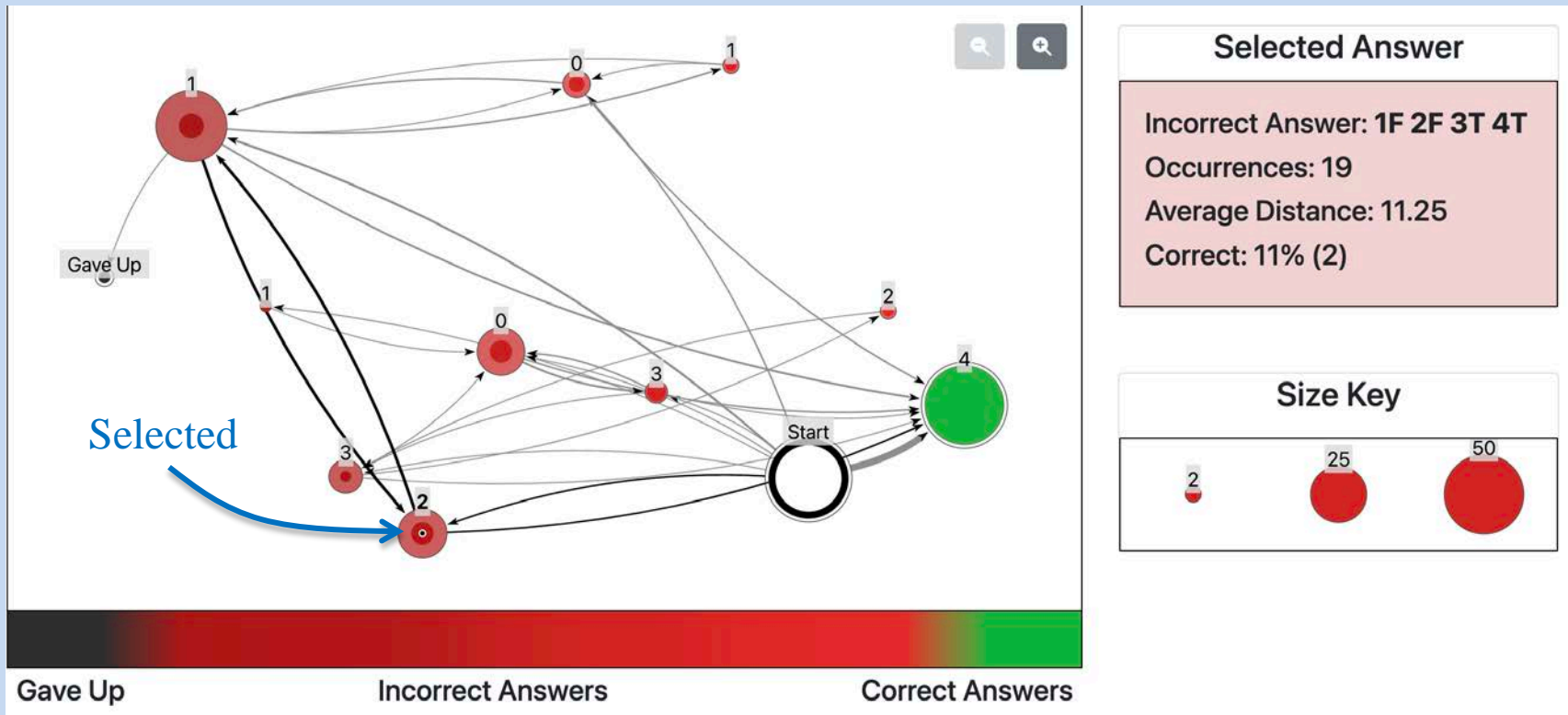
Graph Details #7

- 2 students got to 1T 2T 3T 4T in one more run



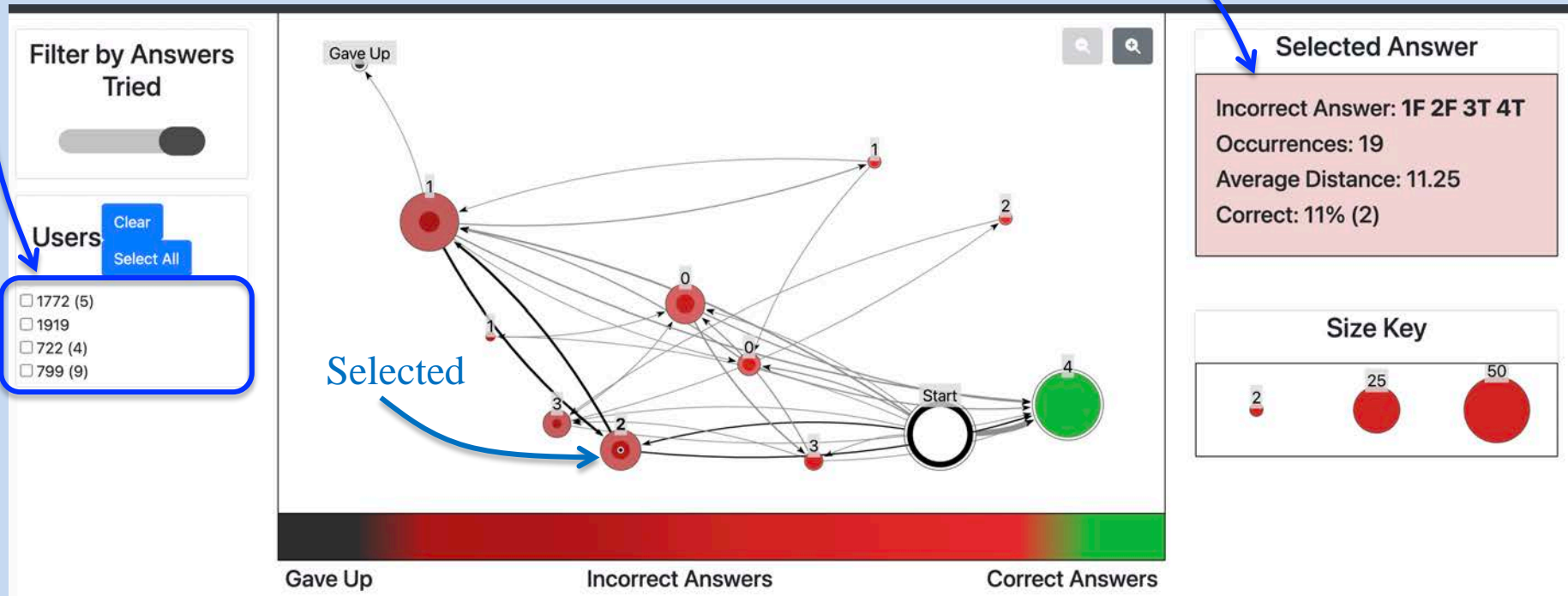
Graph Details #8

- From this node, ~11 more runs were required to get to a green node



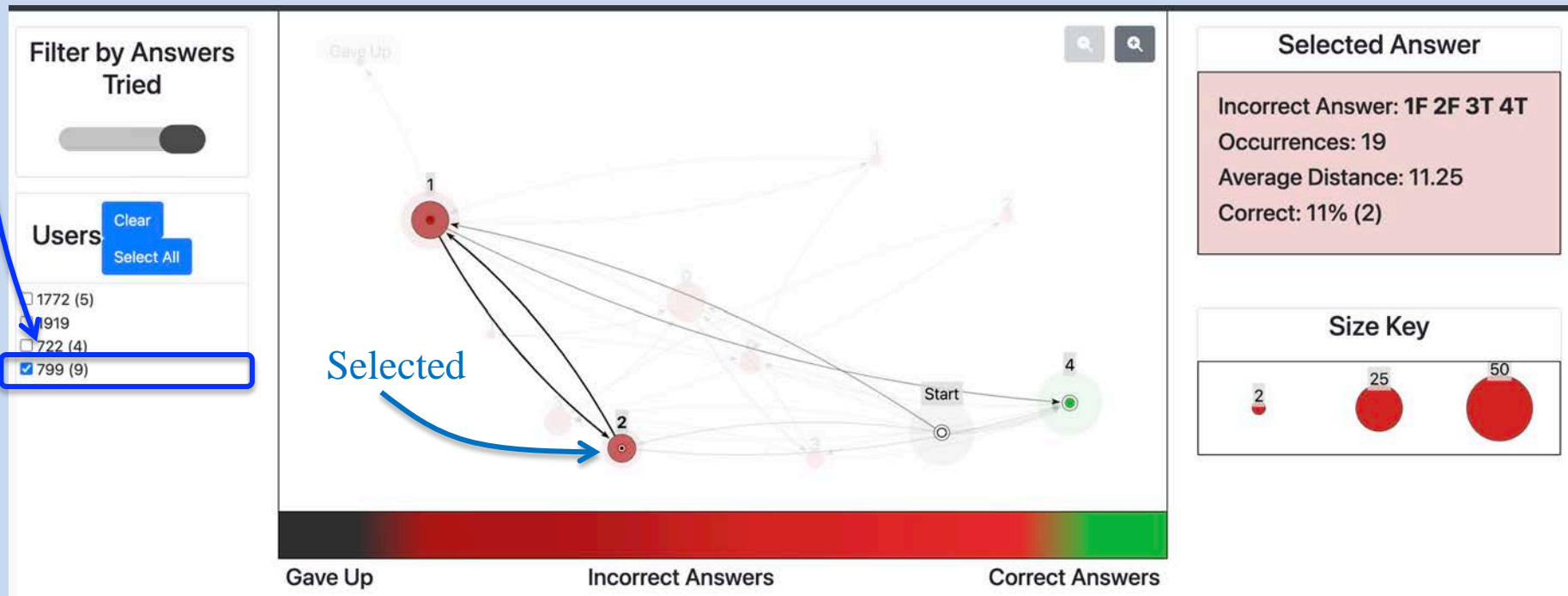
Graph Details #9

- Four students wrote code that caused this red node to be *visited*
- Examine the stats for this node



Graph Details #10

- Student 799:
 - had 9 test runs that caused this node to be visited
 - bounced back and forth between the “2” node and the “1” node
 - finally got to the green node from the “1” node



IRB Study Underway

- The pilot study (above) provided impetus for an IRB study
- This spring quarter:
 - 3 sections of ~25 students each of CSSE220
 - collecting trace data on 4 assignments:
 - HW2DArray, HWMaps, HWRecursion, HWTwelveProblems
 - We gain access to this data once grades have been submitted

Click on Some Graphs

- More nodes of the *PowersOfTwo* graph
- *TimesOccur* (and look at student 799)