

Analysis: Twisty Little Passages

In general, we can't just teleport to every room, because N can be much larger than K . We could teleport to a randomly-chosen subset of the rooms, calculate the average degree (number of adjoining passages) of those rooms we've visited, and assume that this is a good estimate of the average degree of all the rooms. The number of passages is half the sum of the room degrees (since each passage connects two rooms.)

When can this approach yield a poor estimate? If a small subset of the rooms have a degree much higher or much lower than the median degree, then we might not visit any of them, and our estimate would be too high or too low.

Cases where most rooms have high degree and a small number of rooms have low degree are not a problem, because we are judged on relative error, and the relative error in this case is low. If the average degree is 100 but we estimate that it is 103, we would only be 3% from the answer. But if the average degree is 5 and we estimate that it is 2, we would be 60% away from the answer!

So consider a troublesome case where most rooms have low degree, and a small set of rooms have high degree — few enough that we are unlikely to find one of them by teleporting randomly. If these rooms contribute a significant fraction of the total number of passages, then they must connect to a significant fraction of the total set of rooms. So we can find these with high probability by repeatedly teleporting to a random room and then walking through a random passage with the "W" command.

Consider a series of rounds where we alternate between teleporting to a random room with the "T" command and walking through a random passage with the "W" command. We cannot simply extrapolate the average degree for the whole cave from the average degree of all the rooms we've seen — the rooms we reach with "W" commands are not a uniform sample. We can, however, use the average degree of just the rooms we visited with the "T" command as our estimate for all the rooms we haven't visited, and then add the known degrees to that. This is sufficient to solve the problem.

Another solution is to use an alternating series of "T" and "W" commands as above, and then use a technique called [importance sampling](#). Each visit to a room gives us a sample of the average degree, but each room does not have an equal chance of being visited in any given sample. By weighting each sample appropriately, we can compute a weighted average which serves as an unbiased estimate of the total. The weights are chosen to compensate for the non-uniform probabilities of visiting each room.

When we randomly choose a room and visit it via a "T" command, each room had an equal chance of being chosen; we give this sample a weight of 1. When we visit a room via a "W" command, each room did not have an equal chance of being visited. We need to calculate weights for these samples, such that the expected weight for each room (the probability of visiting that room via a "W" command, multiplied by the expected weight we assign for such visits) is $1/N$. This will give our final estimate the correct expected value.

Consider a sample where we were previously in a room R_1 with degree A , and then we issued a "W" command and walked into a room R_2 with degree B . The probability of us being in R_1 after the last "T" command was $1/N$. The probability of following the passage to R_2 was $1/A$, because R_1 had A passages connected to it. So the overall probability was $1/(AN)$. We assign this sample the weight A/B , so that the contribution to the expected weight for room R_2

is $1/(BN)$. After we sum over all B ways we could arrive at R_2 , we get a total expected weight of $1/N$ for R_2 , as required.

As an example, consider the following interaction:

```

T 1
1 1
W
3 2
T 2
2 1
W
3 2
T 3
3 2
W
1 1

```

We get samples of degrees 1, 2, 1, 2, 2, 1 with weights 1, 1/2, 1, 1/2, 1, 2, and the weighted average degree is 8/6.