This project tries to find whether friends' friends are also friends. The data set facebook_combined.txt was sourced from the Stanford University website, https://snap.stanford.edu/data/ego-Facebook.html, it has a total of 4039 nodes and 88,234 edges.

Two functions were defined. The first function is read_file, which reads the txt file and parses to construct the graph of the friends relationships and HashMap. The second function calculates the similarity between two sets. The main function uses those two functions to get the results.

Sample output:
Similarity between NodeIndex(3994) and NodeIndex(4006) : 0.08
Similarity between NodeIndex(3995) and NodeIndex(4005) : 0.22
Similarity between NodeIndex(3996) and NodeIndex(4004) : 0.29
Similarity between NodeIndex(3997) and NodeIndex(4003) : 0.10
Similarity between NodeIndex(3998) and NodeIndex(4002) : 0.08
Similarity between NodeIndex(3999) and NodeIndex(4001) : 0.15

The first line of the sample means for NodeIndex(3994) and NodeIndex(4006), the similarity is 0.08. Using the context of friends' relationships between node 3994 and node 4006, there are 8 mutual friends between them, but together in total they have 100 friends. Therefore the similarity score is 8/100 which is equal to 0.08.

Two test cases were also added to verify the correctness of the similarity methods.

This project has a dependency on the petgraph library, therefore in cargo.toml the following dependencies are defined.
[dependencies]
petgraph = "0.6.0"

The command to build the project:
cargo build --release

The command to run the project:
./target/release/my-project facebook_combined.txt jaccard

The project is in target/release directory and input file name is facebook_combined.txt, the jaccard refers to the jaccard index, which is a statistical matter used to compare the similarities and diversity of the sample set. It uses the intersection of two sets and divides the union of two sets as the similarity score. The result is a value between 0 and 1, where 0 means no similarity and 1 means that the two sets are identical.

The command to run the test cases:
cargo test