DS210 Final Project

Resources:

- https://docs.rs/validator/latest/validator/ (https://docs.rs/validator/latest/validator/)
- https://www.rustadventure.dev/uploading-pokemon-data-from-a-csv-into-a-planetscale-sql-database/reading-csv-rows-into-rust-using-the-csv-crate (https://www.rustadventure.dev/uploading-pokemon-data-from-a-csv-into-a-planetscale-sql-database/reading-csv-rows-into-rust-using-the-csv-crate)
- https://github.com/malcolmvr/graphrs (https://github.com/malcolmvr/graphrs)
- https://docs.rs/petgraph/latest/petgraph/ (https://docs.rs/petgraph/latest/petgraph/)
- https://docs.rs/rustworkx-core/latest/rustworkx_core/ (https://docs.rs/rustworkx-core/latest/rustworkx_core/)
- https://www.nebula-graph.io/posts/how-to-compute-betweenness-centrality-aganist-nebula-graph (https://www.nebula-graph.io/posts/how-to-compute-betweenness-centrality-aganist-nebula-graph)
- lecture 25
- lecture 28
- lecture 30
- https://www.youtube.com/watch?v=FtBfVgcFgy8 (https://www.youtube.com/watch?v=FtBfVgcFgy8)
- https://docs.rs/petgraph/latest/petgraph/graph/struct.NodeIndices.html (https://docs.rs/petgraph/latest/petgraph/graph/struct.NodeIndices.html)
- https://docs.rs/petgraph/latest/petgraph/graph/struct.Graph.html (https://docs.rs/petgraph/latest/petgraph/graph/struct.Graph.html)

Report

Data set:721 Weight Training Workouts

-URL: https://www.kaggle.com/datasets/joep89/weightlifting (https://www.kaggle.com/datasets/joep89/weightlifting) (Kaggle) This data set provides the individual's weightlift routine with the types of exercises, the weight he lifted, the number of sets and reps he did, and the date of each workout for him. I am interested in working out, and I can consider myself a gym lover. As a gym lover, I am eager to have better performance in weightlifting and bodybuilding, and I believe it would be helpful for me to have better performance analyzing this data set. As analyzing the data set, I hope to identify the patterns in the training strategies, tips to get stronger, and some insights he improved his performance. I hope I can also analyze the mechanism he improved his performance and apply it to myself.

What Interesting thing I discovered

I discovered that the Hammer high row has the highest betweenness centrality of 213.61903558034524. It implies that hammer high row might be the most important and influential exercises in terms of their association with improved performance in that high betweenness centrality suggessts the node with the high betweenness centrality have considerable influence over the information passing between other nodes. However, still some other well known important exercises like squat and dead lift shows high centrality scores of 105.79072168385757 and 113.78611917095613, so I can assumme not only hammer high row, but also those two exercises are influential over other workout patterns associated to improved performance.

What algorithms I implemented/ what the project does/ how to run it

I implemented betweenness centrality measure algorithm based on Dijkstra's shortest path algorithm in my project. To compute the betweenness centrality, I needed to perform Dijkstra's algorithm to find shortest path between the vertices and then compute the betweenness centrality because my graph is weighted. To calculate betweenness centrality, I take every pair of the network and count how many times a node can interrupt the shortest paths between the two nodes of the pair;(n-1)(n-2)/2. With the shortest paths calculated by Dijkstra's algorithm, I computed betweenness centrality of each nodes from the start node. Therefore, for my dataset, the project reads a data set from csv file and constructs directed weighted graph using petgraph, and finds a node with the high betweenness centrality measure algorithm that can be assummed it has relatively crucial impact over other exercises in terms of their connections to other exercises using Dijkstra algorithm based betweenness centrality algorithm.

And the complextiy of my Dijkstra's algorithm would be O((E+V)logV) where E is the number of edges and V is the number of vertices vistied, so my graph has 9302 edges and 9303 nodes, O(18605)log9303), and for the betweennes algorithm, it would have O(V(E+V)logV, so O(9303(18605)log9303). And, overall, it would be O(V(E+V)logV), so O(9303(18605)log9303).

Conclusions/Findings

Do the highest centrality vertices match your intuition? What happens when you look at some meaningful subgraphs. Is this the case for them? What the output looks like?

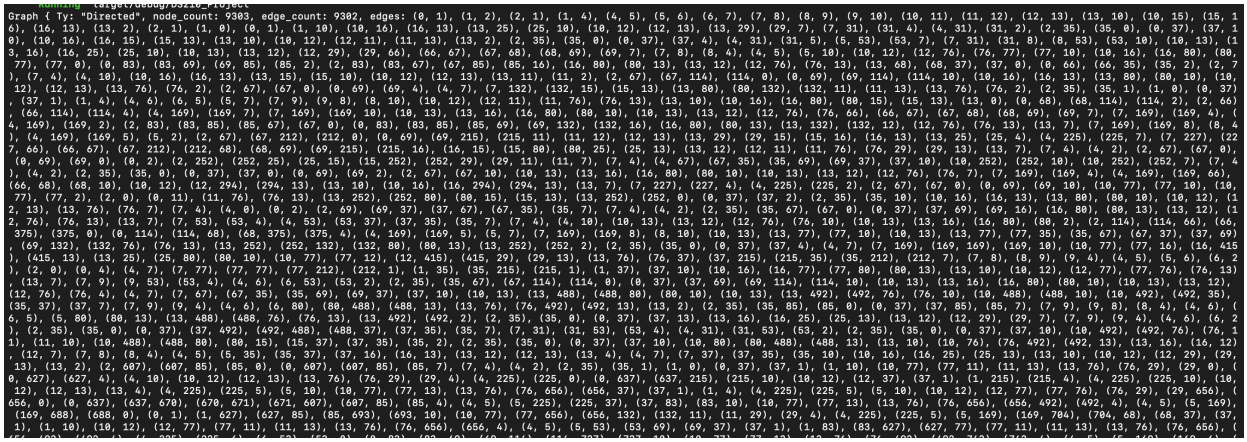
The higheset centrality vertices does not really match my intution. I expected deadlift or barbel squat to be the highest centrality vertices because it is well known that squat and deadlift helps to improve the overall performance of the weightlifting, but the Hammer high row was the vertice with the highest centrality vertices. However, When I look at some meaningful subgraphs and other centralities, deadlift, squat, bench press and some other exercises that I expected to have the highest centratlites are also located at top 10 highest centralities. It means they also might be influential to other exercises and a key exercise to help to improve the overall performance. So, I can say the centrality algorithm not perfectly, but certainly worked on my project and intuition.

Outcome

When I output shortest path from Dijskra algorithm and centrality scroes, there are some None and 0 values I think it is becasue some of my nodes names in the data set has the redundant name with the same weight since the kind of exercises are limited and the amount of weight lifted can not improve in short term.

Graph Nodes/Edges

Graph represented without weight in the form of integer with # of nodes and edges



Graph represented with weight in the form of String / I pasted just part of graph becasue my data set and graph are too large to paste all here.

```
Squat (Barbell) -> Romanian Deadlift (Barbell): 11
Romanian Deadlift (Barbell) -> Leg press (hinge ): 8
Leg press (hinge ) -> Bench Press (Barbell): 17
Bench Press (Barbell) -> Bicep Curl (Barbell): 8
Bicep Curl (Barbell) -> Hammer Row - Wide Grip: 3
Hammer Row - Wide Grip -> Hammer High Row - 1 Arm: 6
Hammer High Row - 1 Arm -> Squat (Barbell): 6
Squat (Barbell) -> Romanian Deadlift (Barbell): 11
Romanian Deadlift (Barbell) -> Leg press (hinge ): 7
Leg press (hinge ) -> Incline Bench Press (Barbell): 16
Incline Bench Press (Barbell) -> Incline Press (Dumbbell): 9
Incline Press (Dumbbell) -> Chin Up: 6
Chin Up -> Neutral Chin: 3
Neutral Chin -> Seated Cable Row (close Grip): 0
Seated Cable Row (close Grip) -> Squat (Barbell): 4
Squat (Barbell) -> Romanian Deadlift (Barbell): 9
Romanian Deadlift (Barbell) -> Leg press (hinge ): 8
Leg press (hinge ) -> Bench Press (Barbell): 20
Bench Press (Barbell) -> Seated Shoulder Press (Dumbbell): 8
Seated Shoulder Press (Dumbbell) -> Neutral Chin: 6
Neutral Chin -> Chin Up: 2
Chin Up -> Hammer seated row (CLOSE GRIP): 1
Hammer seated row (CLOSE GRIP) -> Hammer Row - Wide Grip: 7
Hammer Row - Wide Grip -> Bicep Curl (Barbell): 8
```

Dijkstra's alogrithm outcome

betweenness centrality/ descending order of centralities.

```
    Finished dev [unoptimized + debuginfo] target(s) in 0.66s
     Running `target/debug/DS210_Project`
Node Hammer High Row — 1 Arm: Centrality 213.61903558034524
Node Lat Pulldown Closegrip: Centrality 211.91221623413855
Node Hammer Decline Chest Press: Centrality 193.1545303170818
Node Bench Press (Barbell): Centrality 159.2308241037185
Node Curl Dumbbell: Centrality 140.27481082275904
Node Pull Up: Centrality 134.68937267069245
Node Hammer shoulder press: Centrality 127.76968687805453
Node Deadlift (Barbell): Centrality 113.78611917095613
Node Low Incline Dumbbell Bench: Centrality 113.50123456014268
Node Romanian Deadlift (Barbell): Centrality 112.98415064575411
Node Chin Up: Centrality 108.35076100049284
Node Squat (Barbell): Centrality 105.79072168385757
Node Incline Bench Press (Barbell): Centrality 104.74677572353151
Node Bent Over Row (Dumbbell): Centrality 102.55433173056916
Node Face pull: Centrality 99.06052368482077
Node Shoulder Press (Standing): Centrality 95.37905482806092
Node Overhead Press (Dumbbell): Centrality 91.64691343576955
Node Lateral Raise (Dumbbells): Centrality 91.29347101320433
Node Seated Shoulder Press (Dumbbell): Centrality 90.92399607587053
Node Hammer Row — Wide Grip: Centrality 88.47219449285924
Node Incline Press (Dumbbell): Centrality 85.2030796102678
```

Test

```
running 2 tests
test tests::test_centrality ... ok
test tests::test_construct_graph ... ok

test result: ok. 2 passed; 0 failed; 0 ignored; 0 measured; 0 filtered out; finished in 2.48s

crc-dot1x-nat-10-239-66-62:DS210_Project jadennoh$
```