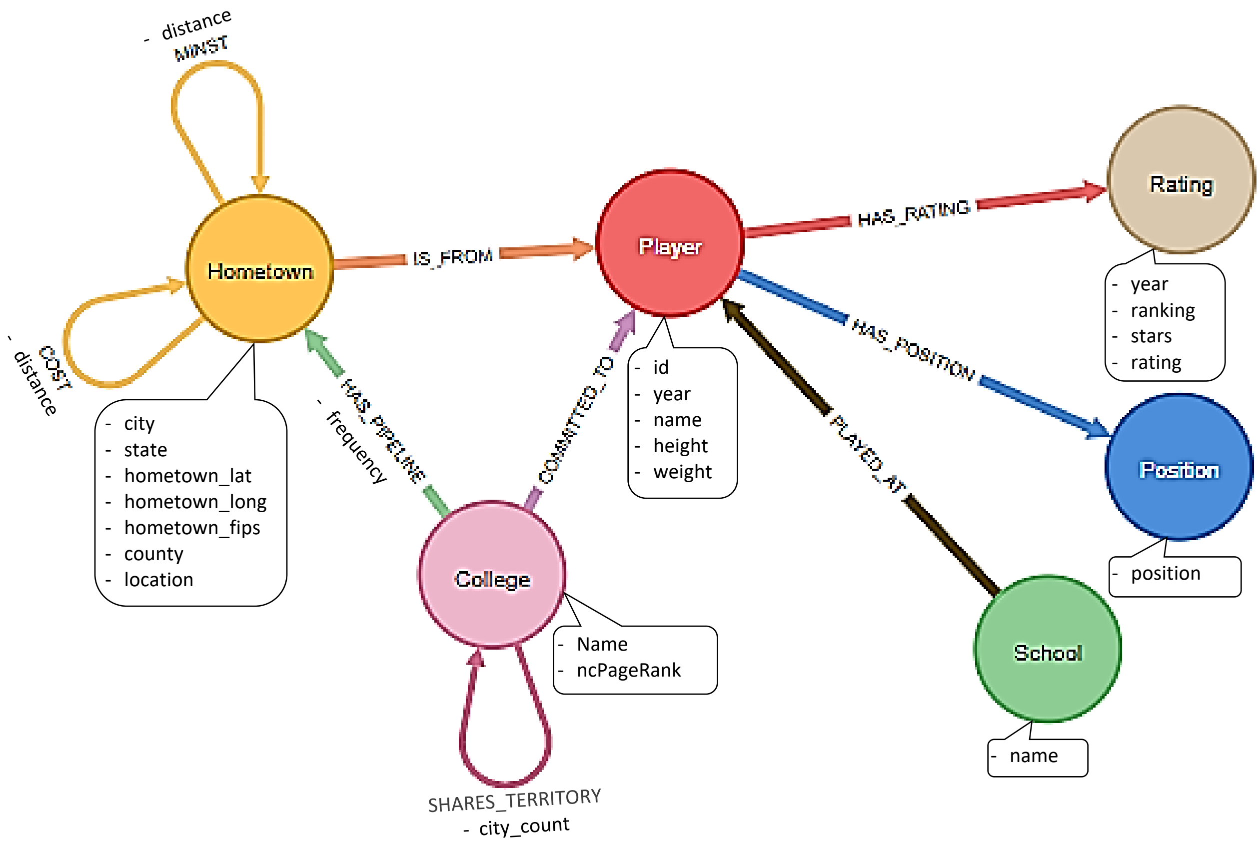
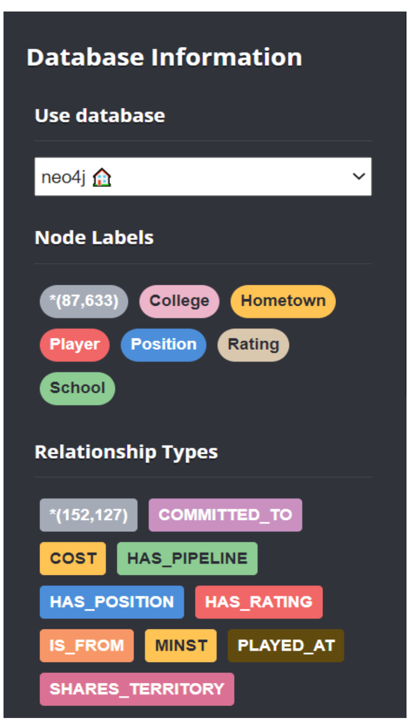
Joseph Wiens
jwiens@uncc.edu

Project Deliverable II | Graph Database Setup and Application of Algorithms

Section IV: Graph Data Model (Updated)





A number of updates to the data model was required to process queries and execute graph algorithms. First, the key variables were separated into node labels and assigned the appropriate properties. Combining the existing latitude and longitude properties, a new "location" property was added to define points, which would be referenced to calculate distance. The "ncPageRank" property on the "College" node is the score output of the PageRank algorithm of colleges recruiting at least one player from North Carolina

In addition, the relationships required maintenance and redundancies were removed from the model. New relationships include a 'HAS_PIPELINE' between the "College" node and "Hometown" node with a property called "frequency", representing the number of players recruited from a city. The 'SHARES_TERRITORY' relationship describes the number of cities where both universities recruited a player to find common player sources. The 'COST' relationship depicts the distance between two cities that are less than 35 kilometers away from one another, with "distance" as the relationship property name. Utilizing the distance attribute, the Minimum Spanning Tree algorithm was calculated and its results written into the model, signifying the 'MINST' relationship.

Section V: Graph Projections (Updated)

See Section VII: Graph Algorithms

Section VI: Cypher Queries

Which high schools had multiple players commit to Charlotte? Identifying which high schools have existing relationships with the Charlotte football program supports future recruiting efforts of potential athletes.

Figure 6.1. Frequent High School commits query

```
MATCH (n:Player)<-[:COMMITTED_TO]-(c:College {name: 'Charlotte'})
WITH n, c
MATCH (h:Hometown)-[:IS_FROM]->(n)<-[:PLAYED_AT]-(s:School)
WITH h.state AS state, h.city AS city, s.name AS high_school, count(c) AS charlotte_recruits
WHERE charlotte_recruits > 1
RETURN state, city, high_school, charlotte_recruits ORDER BY charlotte_recruits DESC
```

Table 6.1. Frequent High School commits results

State	City	High School	Number of Recruits
NC	Charlotte	Mallard Creek	3
FL	Orlando	Bishop Moore Catholic	2
GA	Buford	Buford	2
FL	Tallahassee	Tallahassee Leon	2
NC	Mooresville	Mooresville Senior	2
TN	Murfreesboro	Blackman	2
NC	Jamestown	Lucy Ragsdale	2
SC	Mount Pleasant	Wando	2
NC	Charlotte	Zebulon B. Vance	2
SC	Blythewood	Blythewood	2
NC	Matthews	Butler	2

Four local high schools had multiple players commit to Charlotte with Mallard Creek topping all secondary schools. Surprisingly, this list contains two Florida schools, more than any bordering state other than South Carolina. The Charlotte football program can leverage their connections to high schools to attract future athletes.

What are the average rankings and ratings of Charlotte commits? Determining the range of a typical Charlotte recruit's rating narrows the focus on realistic candidates, preventing the program from wasting effort and resources on unfeasible expectations.

For the seven years of recruiting classes dating 2015 and 2021, Charlotte's average player rated about 0.8 out of one. However, Charlotte's football program is still in its infancy, joining the top collegiate division in 2015. This likely posed a detrimental effect on recruiting top players; adjusting for more recent data would reflect a more accurate situation.

Figure 6.2. Charlotte recruiting statistics query

```
MATCH (n:Player)<-[:COMMITTED_TO]-(c:College {name: "Charlotte"})
WITH n
MATCH (n)-[:HAS_RATING]->(t:Rating)
RETURN percentileDisc(toFloat(t.rating), 0.25) AS lower_quartile_rating, AVG(toFloat(t.rating)) AS avg_rating,
percentileDisc(toFloat(t.rating), 0.75) AS upper_quartile_rating, percentileDisc(toFloat(t.ranking), 0.75) AS
lower_quartile_rank, AVG(toFloat(t.ranking)) AS avg_rank, percentileDisc(toFloat(t.ranking), 0.25) AS upper_quartile_rank
```

Table 6.2. Charlotte recruiting statistics results

Lower Quartile Rating	Average Rating	Upper Quartile Rating	Lower Quartile Rank	Average Rank	Upper Quartile Rank
0.7731	0.7949	0.8178	3,090	2,510	1,954

Evaluating the last three years from 2019 to 2021, the 49ers recruited much better than the aggregate performance of the past seven years back to 2015. Concentrating their efforts on a specific range of players, staff can be more efficient in utilizing and distributing their available resources.

Figure 6.3. Charlotte recruiting statistics from 2019 to 2021 query

```
MATCH (n:Player)←[:COMMITTED_TO]-(c:College {name: "Charlotte"})
WITH n
MATCH (n)-[:HAS_RATING]→(t:Rating)
WHERE t.year > '2018'
RETURN percentileDisc(toFloat(t.rating), 0.25) AS lower_quartile_rating, AVG(toFloat(t.rating)) AS avg_rating,
percentileDisc(toFloat(t.rating), 0.75) AS upper_quartile_rating, percentileDisc(toFloat(t.ranking), 0.75) AS
lower_quartile_rank, AVG(toFloat(t.ranking)) AS avg_rank, percentileDisc(toFloat(t.ranking), 0.25) AS upper_quartile_rank
```

Table 6.3. Charlotte recruiting statistics from 2019 to 2021 results

Lower Quartile Rating	Average Rating	Upper Quartile Rating	Lower Quartile Rank	Average Rank	Upper Quartile Rank
0.7831	0.8072	0.8301	3,096	2,387	1,689

Where are recruits in North and South Carolina that could realistically commit to Charlotte?
Recognizing and prioritizing talent rich areas reinforces an efficient recruitment plan, an important task in effective resource utilization. Local recruits are not only likely to be interested in staying close to home, but they also reduce budget strain with low traveling expenses.

Figure 6.4. Charlotte target areas query

```
MATCH (n:Player)←[:IS_FROM]-(h:Hometown)
WHERE h.state = 'NC' OR h.state = 'SC'
WITH n, h
MATCH (n)-[:HAS_RATING]→(t:Rating)
WHERE t.rating > '0.78' AND t.rating < '0.84'
RETURN h.state AS state, h.city AS city, count(n) AS num_players
ORDER BY num_players DESC
LIMIT 10
```

Table 6.3. Charlotte target areas results

State	City	Number of Target Players
NC	Charlotte	63
SC	Columbia	21
NC	Greensboro	20
NC	Cornelius	12
NC	High Point	10
NC	Fayetteville	10
NC	Raleigh	10
SC	Sumter	10
SC	Blythewood	10
SC	Anderson	10

Besides Charlotte, the most populous cities in the Carolinas are not always producing the most recruiting targets! Raleigh totaled ten recruits falling into the range of 0.78 to 0.84 rating since 2015, which is the same amount as the significantly smaller South Carolina towns of Blythewood, Anderson, and Sumter.

Section VII: Graph Algorithms

Pathfinding: Minimum Weight Spanning Tree

Creating an efficient itinerary to visit or scout players can be challenging, especially when recruits are located in a similar, but not close, proximity. Ideal for this scenario, a Minimum Weight Spanning Tree algorithm optimizes a connected route, ensuring that staff will travel in an efficient course.

For example, if a coach was heading from campus to High Point to see a recruit and they also wanted to visit every target player in cities on the way, this algorithm provides an optimal path to their destination through nine towns that contain players in our target rating range.

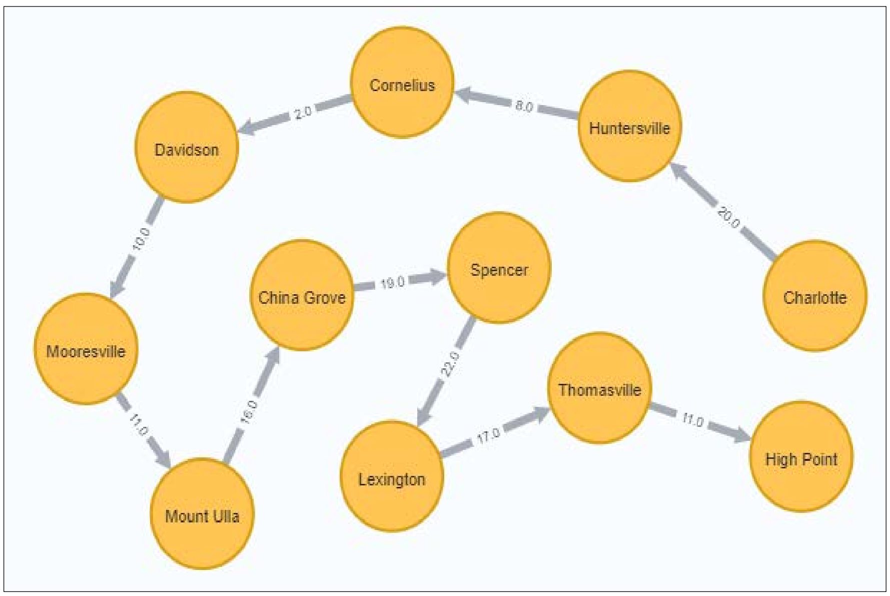
Figure 7.1. Code to call MSP algorithm

```
MATCH (n:Hometown{city:'Charlotte'})
CALL gds.alpha.spanningTree.minimum.write({
  startNodeId: id(n),
  nodeProjection: 'Hometown',
  relationshipProjection: {
    COST: {
      type: 'COST',
      properties: 'distance',
      orientation: 'UNDIRECTED'
    }
  },
  relationshipWeightProperty: 'distance',
  writeProperty: 'MINST',
  weightWriteProperty: 'cost'
})
YIELD createMillis, computeMillis, writeMillis, effectiveNodeCount
RETURN createMillis, computeMillis, writeMillis, effectiveNodeCount;
```

Table 7.1. MSP algorithm output

Source	Destination	Cost (km)
Charlotte	Huntersville	20
Huntersville	Cornelius	8
Cornelius	Davidson	2
Davidson	Mooresville	10
Mooresville	Mount Ulla	11
Mount Ulla	China Grove	16
China Grove	Spencer	19
Spencer	Lexington	22
Thomasville	High Point	11

Figure 7.2. A minimum weight spanning tree from Charlotte to High Point



Similarity: Jaccard

Since players often weigh several scholarship offers, Charlotte isn't the only team contending for a player's commitment. Recognizing the competition is an important step in order to differentiate Charlotte from other programs. The Jaccard Similarity algorithm will show which colleges received commitments from high schools that Charlotte players also attended.

Figure 7.3 Code to call MSP algorithm

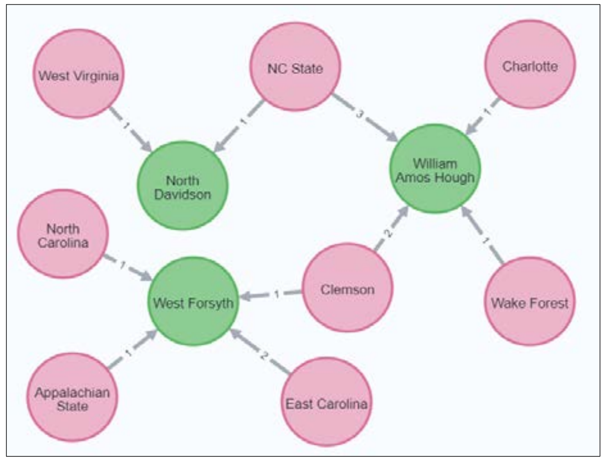
```
MATCH (c1:College {name: 'Charlotte'})-[:COMMITTED_TO]->(n)<-[:PLAYED_AT]-(school1)
WITH c1, collect(id(school1)) AS c1School
MATCH (c2:College)-[:COMMITTED_TO]->(n)<-[:PLAYED_AT]-(school2) WHERE c1 <> c2
WITH c1, c1School, c2, collect(id(school2)) AS c2School
RETURN c1.name AS charlotte, c2.name AS other_college,
gds.alpha.similarity.jaccard(c1School, c2School) AS similarity ORDER BY similarity DESC
```

Larger universities possess a significant presence in the high schools where Charlotte attempts to procure talent. To counter powerhouse programs, staff should promote unique selling points to attract players; like early playing time, proximity to a world class city, and post-football career opportunities. Also, staff can even emphasize negative aspects of other colleges, such as a program instability, rural location, or a poor fit into a team's coaching scheme.

Table 7.2. Jaccard Similarity algorithm output

College	Other College	Similarity
Charlotte	South Carolina	0.072
Charlotte	North Carolina	0.070
Charlotte	NC State	0.067
Charlotte	Wake Forest	0.065
Charlotte	Tennessee	0.056
Charlotte	Georgia State	0.051
Charlotte	East Carolina	0.051
Charlotte	Appalachian State	0.051
Charlotte	Coastal Carolina	0.050
Charlotte	Liberty	0.049

Figure 7.4. Sample projected similarity graph



Centrality: PageRank

Since only few programs can recruit nationally, a football program must ensure its locking down players from its own state. The PageRank algorithm will determine which colleges battling for players in the same cities hold the most influence over the state's geography.

Figure 7.5. Code to call and write PageRank algorithm

```
CALL gds.pageRank.write({
    nodeQuery: 'MATCH (n:College)-[r:HAS_PIPELINE]->(h:Hometown)
    WHERE h.state = $state RETURN DISTINCT id(n) AS id',
    relationshipQuery: 'MATCH (h:Hometown)<-[:HAS_PIPELINE]-(n)-[r:SHARES_TERRITORY]->(m)
    WHERE h.state = $state
    RETURN id(n) AS source, id(m) AS target, r.city_count as weight,type(r) as type',
    writeProperty: "ncPageRank",
    validateRelationships: false,
    parameters: { state: 'NC' }})
YIELD nodePropertiesWritten, createMillis, computeMillis, writeMillis, ranIterations
RETURN nodePropertiesWritten, createMillis, computeMillis, writeMillis, ranIterations;
```

Figure 7.6. Query to view top PageRank results and statistics

```
MATCH (c1:College)-[:HAS_PIPELINE]->(h:Hometown {state: 'NC'})
WITH c1, count(DISTINCT h.city) AS num_cities
MATCH (c1:College)-[r:SHARES_TERRITORY]->(c2:College)
RETURN c1.name, c1.ncPageRank, count(r) AS share_cnt, num_cities
ORDER BY c1.ncPageRank DESC LIMIT 10
```

North Carolina's flagship university possesses the most sway over the state's numerous cities of the 100 colleges that recruited a player in the state; however, smaller programs are outpacing more prominent schools like Wake Forest and Duke, which could be related to academic standards. All service academy players are on scholarship so Navy's inclusion isn't reflective of a typical college football program's scholarship constraints. While this does not account for the density of players by city, PageRank reveals that Charlotte is scouring the state for talent well in competitive areas.

Table 7.3. PageRank algorithm results and statistics

College	NC pageRank	Colleges Sharing City	City Count
North Carolina	2.076	88	35
Appalachian State	1.986	87	27
NC State	1.935	86	40
Charlotte	1.858	81	23
East Carolina	1.771	75	33
Wake Forest	1.664	79	25
Duke	1.587	74	19
Coastal Carolina	1.563	74	18
Old Dominion	1.535	71	15
Navy	1.507	70	12

Figure 7.7. Sample PageRank graph projection