

Base de datos No Relacionales

Martín Borba

`martin.borba@fing.edu.uy`

Mariana Bartesaghi

`mariana.bartesaghi@fing.edu.uy`

Facultad de Ingeniería — July 3, 2023

Introducción

El presente informe se basa en un trabajo previo realizado por nuestros excompañeros (Grupo8 Martín Ghazarian y Gianni Testa del curso 2022), donde se modelaron diferentes entidades del mundo del básquetbol en particular la liga NBA de estados unidos . El modelo inicial incluía las entidades de Jugadores (Players), Juegos (Games), Equipos (Team) y Temporadas (Season). Como parte de nuestro trabajo, hemos ampliado y mejorado este modelo inicial mediante la incorporación de nuevas entidades.

Adicionalmente, adaptaremos algunas de las consultas previas realizadas por nuestros excompañeros a nuestro nuevo modelo y también presentaremos nuevas consultas y modificaciones que hemos desarrollado para ampliar la funcionalidad del sistema y obtener información más detallada y relevante sobre los jugadores, juegos, equipos y temporadas de básquetbol.

Carga de datos

En el trabajo anterior, se llevó a cabo la carga de datos en dos etapas. La primera etapa implicó la carga manual a través de la interfaz de Neo4j, mientras que la segunda etapa se realizó de manera programática utilizando las bibliotecas de Python 'neo4j' y 'pandas'. Con el fin de ampliar el modelo del grafo y abordar datos más complejos, se tomó la decisión de cargar todos los datos de manera programática. Esto permitió incluir todo tipo de relaciones y datos, como el cálculo del tiempo jugado en los juegos, y también facilitó la reproducibilidad del proceso.

Para obtener el dato calculado de los juegos, se sumó el tiempo total jugado por los jugadores que participaron en cada juego, y luego se dividió esa suma por 10, ya que siempre hay 10 jugadores en el campo de juego. Esta métrica nos brinda una medida adicional del impacto y la contribución de los jugadores en cada juego, lo que nos permite evaluar su rendimiento de manera más precisa y detallada.

Dado el volumen abundante de datos, la carga puede volverse lenta, especialmente al cargar el archivo CSV "game_details" que contiene las estadísticas de todos los jugadores en todos los juegos. Observamos que el principal cuello de botella en la carga de estos datos utilizando la biblioteca Neo4j se producía en la conexión de red, ya que se realizaba una solicitud individual por cada fila del archivo CSV. Para abordar este problema, se modificó el proceso para realizar las solicitudes de red cada 100 filas.

Se implementó un enfoque donde se crea una transacción y se realiza un commit una vez que se han completado 100 consultas. Esto permitió reducir significativamente la cantidad de solicitudes de red y optimizar el rendimiento de la carga de datos. Al agrupar las consultas en lotes de 100, se mejoró la eficiencia y se redujo el tiempo total requerido para cargar los datos del archivo CSV. Toda la carga de datos así como las consultas de interés se encuentran en la notebook del repositorio [3].

Modelado

Para mejorar el modelo, hemos creado la entidad de Equipo-Temporada (TeamSeason) para optimizar las consultas relacionadas con las estadísticas y el rendimiento de los equipos en cada temporada. Esta entidad nos brinda la capacidad de obtener rápidamente información específica sobre los equipos en una temporada determinada.

Además, hemos introducido la entidad de Estadio (Arena), que representa los lugares físicos donde se llevan a cabo los juegos. Anteriormente se agregaba el dato de la arena dentro de cada juego, lo cual notamos es innecesario y es una clara duplicación de datos. Esta modificación nos permite tener en cuenta el factor del estadio en nuestro análisis y consultas sin tener que repetirlo en más de 26.000 juegos. El orden computacional no se ve afectado ya que desde un Juego cualquiera se puede obtener la Arena en que se jugó fácilmente accediendo por el camino de largo 3: (Game)-'[:HOME_TEAM]-(TeamSeason)-[:_season]-(Team)-[:HAS]-[Arena]'.

Otra mejora significativa que hemos realizado es la inclusión de la entidad de Conferencia (Conference). Esta entidad nos permite organizar y clasificar los equipos en grupos según su afiliación a una conferencia específica. Esto facilita el análisis y la comparación de equipos dentro de una misma conferencia. Existen 2 conferencias, la del oeste y la del este.

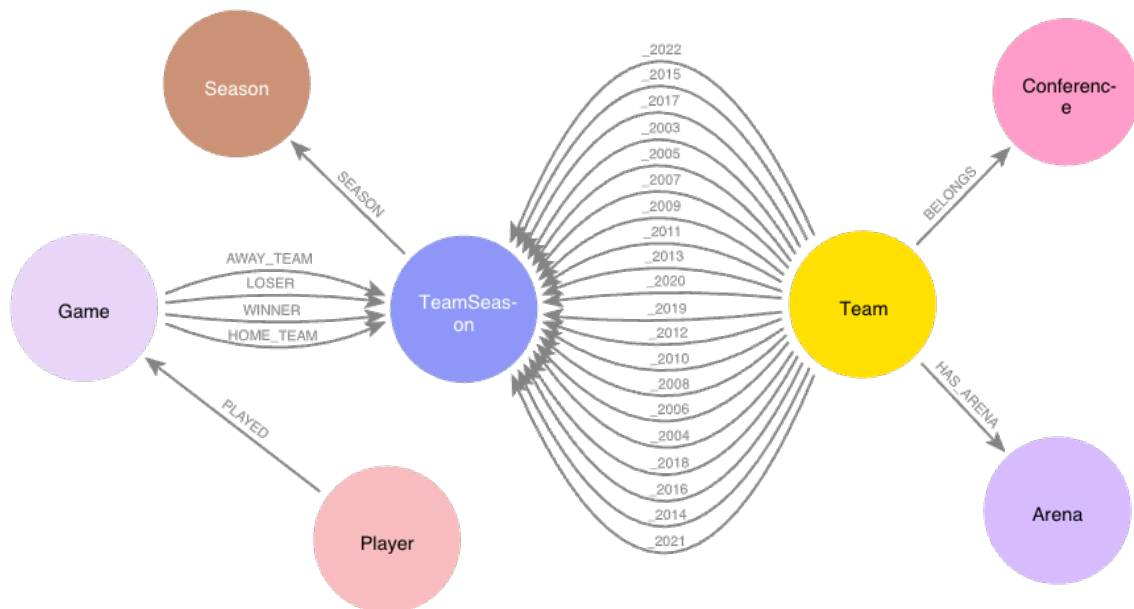


Figure 1: Visualización del modelo

Consultas de interés

Arenas

En la sección de consultas sobre Arenas, nos enfocamos en analizar el impacto de la ubicación física de los juegos en el rendimiento de los jugadores y equipos. Exploramos cómo el desempeño de un jugador puede variar en diferentes arenas y cómo la ventaja de jugar en casa puede afectar el rendimiento de un equipo en una temporada.

Todos los equipos tienen su propia arena, existe un caso particular el cuál es el de Los Angeles Lakers y Los Angeles Clippers los cuales comparten Arena 'Staples Center'. Es importante tener esto en consideración, ya que en esta arena se juegan el doble de juegos en promedio.

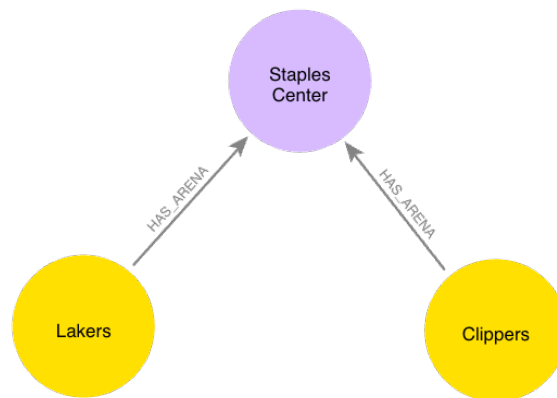


Figure 2: Arena compartida

Para esto realizamos dos consultas, la primera es la adaptación de la consulta realizada en el trabajo final de 2022, la cual obtiene un listado de los estadios donde LeBron James ha obtenido la mayor cantidad de victorias. Además decidimos agregarle estadísticas de interés adicionales para estos partidos de victoria. Específicamente, recopilamos información sobre los puntos anotados, los rebotes obtenidos y las asistencias realizadas por LeBron James.

```
1 MATCH (p:Player {name: "LeBron James"})-[pl:PLAYED]-(g:Game)
2 MATCH (g)-[w:WINNER]-(ts:TeamSeason)
3 MATCH (g)-[:HOME_TEAM]-(t:Team)-[:HAS_ARENA]-(a:Arena)
4 WHERE substring(ts.id, 0, 10) = pl.team
5 RETURN a.id AS Arena, sum(pl.PTS) as PTS, sum(pl.AST) as AST, sum(pl
6 ORDER BY PTS DESC
```

Nombre	Pts	Ast	Reb	Wins
Quicken Loans Arena	5344.0	1392.0	1452.0	192
Staples Center	1911.0	563.0	541.0	72
State Farm Arena	374.0	85.0	107.0	12
Capital One Arena	303.0	64.0	74.0	10
Little Caesars Arena	272.0	61.0	59.0	9
Scotiabank Arena	227.0	56.0	59.0	8
Fiserv Forum	221.0	62.0	50.0	7
Moda Center	218.0	70.0	64.0	7
Madison Square Garden	217.0	41.0	37.0	7
Bankers Life Fieldhouse	217.0	56.0	56.0	7
Barclays Center	211.0	50.0	59.0	7
AT&T Center	204.0	41.0	68.0	8
Toyota Center	195.0	54.0	58.0	7
Amway Center	192.0	44.0	50.0	7
FedExForum	189.0	46.0	55.0	6
Wells Fargo Center	187.0	44.0	40.0	7
Target Center	173.0	29.0	50.0	5
United Center	167.0	33.0	45.0	5
Spectrum Center	164.0	38.0	41.0	6
Chase Center	160.0	54.0	59.0	7
Golden 1 Center	156.0	37.0	28.0	5
Smoothie King Center	154.0	32.0	29.0	5
AmericanAirlines Arena	154.0	39.0	46.0	5
Vivint Smart Home Arena	153.0	43.0	39.0	5
TD Garden	137.0	36.0	32.0	5
Pepsi Center	133.0	48.0	45.0	5
American Airlines Center	112.0	22.0	35.0	4
Chesapeake Energy Arena	90.0	20.0	15.0	3
Talking Stick Resort Arena	88.0	28.0	24.0	4

Table 1: Estadísticas agregadas de las victorias de LeBron en cada estadio (2003..2008+2019..2022)

Sabiendo que tenemos la mayoría de estadísticas desde 2003 hasta 2022 de los partidos, podemos observar la mayor cantidad de victorias de LeBron por estadio se dieron en Staples Center(Los Angeles) y Quicken Loans Arena(Cleveland). Esto es esperable puesto LeBron comenzó jugando con los Cleveland Cavaliers, en total ha jugado 11 temporadas con este equipo y, por tanto, en su Arena como local. También jugó 4 temporadas en Miami y al día de hoy lleva 4 temporadas jugando con Los Angeles Lakers y está en transcurso su 5ta temporada con este equipo. Staples Center al ser compartida por 2 equipos, también suma estadísticas tanto jugando con los Lakers o los Clippers

Además, desarrollamos una nueva consulta que nos permite analizar la cantidad de partidos jugados en cada estadio, centrándonos en una temporada específica. Este enfoque resulta relevante ya que la cantidad de partidos disputados en un estadio puede ser un indicador de ventaja al jugar más partidos como local. A parte de la cantidad de partidos tenemos *Home* y *Away* wins, lo que representa la cantidad de victorias en ese estadio para los locales y visitantes respectivamente. Es relevante tener en cuenta la situación especial en la que los equipos utilizan el mismo estadio para sus partidos, lo cual se ve reflejado en que *Staples Center* aparece dos veces en la tabla, una vez por cada equipo de los Ángeles, en total se jugaron 89 partidos en esa Arena.

```

1 MATCH (t:Team)-[:_2021]-(ts:TeamSeason)-[:HOME_TEAM]-(g:Game)
2 OPTIONAL MATCH (ts)-[w:WINNER]-(g)
3 OPTIONAL MATCH (ts)-[l:LOSER]-(g)
4 MATCH (t:Team)-[]-(a:Arena)
5 WITH t.name as Team, a.id as Arena, count(g) AS gameCount,
6 count(w) as Wins_Local, count(l) as Wins_Away
7 ORDER BY gameCount DESC
8 RETURN Arena, Team, gameCount, Wins_Local, Wins_Away

```

Arena	Team	Game_Count	Wins_Local	Wins_Away
Chase Center	Warriors	56	45	11
TD Garden	Celtics	55	36	19
AmericanAirlines Arena	Heat	54	39	15
American Airlines Center	Mavericks	51	37	14
Talking Stick Resort Arena	Suns	50	39	11
FedExForum	Grizzlies	50	36	14
Fiserv Forum	Bucks	49	31	18
Wells Fargo Center	76ers	49	30	19
State Farm Arena	Hawks	46	30	16
Smoothie King Center	Pelicans	46	22	24
United Center	Bulls	46	30	16
Target Center	Timberwolves	46	29	17
Barclays Center	Nets	46	23	23
Scotiabank Arena	Raptors	46	27	19
Vivint Smart Home Arena	Jazz	46	32	14
Staples Center	Clippers	45	26	19
Pepsi Center	Nuggets	44	24	20
Staples Center	Lakers	44	21	23
Madison Square Garden	Knicks	44	20	24
AT&T Center	Spurs	44	18	26
Chesapeake Energy Arena	Thunder	44	13	31
Quicken Loans Arena	Cavaliers	44	25	19
Toyota Center	Rockets	43	12	31
Amway Center	Magic	43	13	30
Bankers Life Fieldhouse	Pacers	43	17	26
Moda Center	Trail Blazers	43	17	26
Golden 1 Center	Kings	43	18	25
Capital One Arena	Wizards	43	21	22
Little Caesars Arena	Pistons	43	15	28
Spectrum Center	Hornets	43	22	21

Temporadas

En el contexto del modelo de datos, se ha tomado la decisión de crear la entidad "TeamSeason" (Temporada de Equipo) basándose en el patrón de diseño utilizado en el artículo "*Modeling Airline Flights in Neo4j*" [1] para modelar los aeropuertos durante los 365 días del año. Este enfoque se ha adaptado para representar cada equipo en un año determinado.

La creación de la entidad "TeamSeason" se ha motivado por dos factores principales. En primer lugar, a diferencia del caso del número de días en un año, el número de temporadas en la NBA aumenta con el tiempo. Por lo tanto, tener una entidad separada para cada equipo en cada temporada permite darle cierta mantenibilidad al grafo en el tiempo a medida que se agreguen nuevas temporadas.

En segundo lugar, al separar los equipos en entidades individuales por temporada, se facilita la realización de consultas y análisis específicos para cada año y equipo en particular. Esto proporciona una mayor flexibilidad y permite obtener información contextualizada sobre el desempeño histórico de los equipos en diferentes temporadas de la NBA.

Se realizó una consulta sobre la cantidad de partidos ganados y perdidos por los equipos en una temporada específica, en este caso observamos que el equipo "Golden State" en la temporada 2015 tiene el mayor WinRate de 81% de la temporada, este es uno de los años en que el equipo se consagró campeón.

```
1 MATCH (t:Team)-[:_2015]-(ts:TeamSeason)-[w:WINNER]-(:Game)
2 WITH t, ts, COUNT(w) as Wins
3 MATCH (ts)-[l:LOSER]-(:Game)
4 WITH t.name as Team, ts, Wins, COUNT(l) as Losses
5 RETURN Team, Wins, Losses, (toFloat(Wins)/(Wins + Losses)) as WinRate
6 ORDER BY WinRate DESC
```

Las futuras consultas harán uso del modelado de Temporadas también, para reducir el orden computacional de las consultas.

Team	Wins	Losses	WinRate
Warriors	91	22	0.81
Spurs	75	23	0.77
Cavaliers	74	36	0.67
Thunder	70	35	0.67
Raptors	71	38	0.65
Clippers	58	36	0.62
Hornets	58	39	0.60
Celtics	54	39	0.58
Heat	59	45	0.57
Hawks	56	43	0.57
⋮	⋮	⋮	⋮

Table 2: Top 10 equipos en temporada 2015 basado en WinRate

Conferencias

Hemos ampliado el modelo, incluyendo una nueva entidad llamada "Conference". En la NBA, los equipos se dividen en dos conferencias principales: Este y Oeste. Cada equipo pertenece a una de estas conferencias y esta afiliación no cambia a lo largo del tiempo.

En esta sección, nos centraremos en el análisis de las conferencias Este y Oeste, explorando diferentes aspectos relacionados con su competitividad y desempeño. Para evaluar la competitividad, hemos utilizado un dato previamente calculado llamado "time_played", el cual representa la duración en segundos de los juegos. Los juegos pueden tener tres posibles valores: 2.880 (48 minutos), 3.180 (53 minutos), 3.480 (58 minutos) o hasta 3.780 (63 minutos). Cuanto más largo sea un juego, se considera que fue más competitivo, ya que pudo haber requerido tiempo extra (overtime) para determinar al ganador.

Además, hemos realizado otra consulta específica que analiza las victorias de los equipos de la conferencia del Este sobre los equipos del Oeste, y viceversa. A través de este análisis, intentaremos demostrar de alguna manera que existe una diferencia en la fortaleza o calidad entre las dos conferencias.

La siguiente consulta obtiene la cantidad de partidos agrupados por el tiempo empleado dentro de una conferencia (en este caso conferencia Este), tomando en cuenta sólo los partidos jugados entre equipos de una misma conferencia. Del mismo modo, se realizó la consulta también para la conferencia Oeste.

```
1 MATCH (s:Season) - [] - (ht:TeamSeason) - [:HOME_TEAM] - (g:Game) - [:AWAY_TEAM] -  
2   (at:TeamSeason) - [] - (:Team) - [] - (c:Conference)  
3 WHERE c.id='East' and s.id='2021'  
4 MATCH (ht) - [] - (:Team) - [:BELONGS] - (c)  
5 RETURN DISTINCT ROUND(g.time_played/60) as Minutes_played, count(g.  
6   time_played) as Game_count  
7 ORDER BY Minutes_played
```

Minutes_played	Game_count
48.0	434
53.0	17
58.0	3
63.0	1

Table 3: Partidos 2021 conferencia Este

Minutes_played	Game_count
48.0	438
53.0	18
58.0	0
63.0	1

Table 4: Partidos 2021 conferencia Oeste

Es interesante notar que la competitividad interna en conferencias Este y Oeste está casi igualada. En la temporada 2021 la conferencia Este necesitó de 5 overtimes más para definir un ganador en sus partidos que la conferencia Oeste. Por lo cual podríamos decir que en 2021 la competitividad interna de las conferencias estaba nivelada.

No obstante, el nivel de competitividad interno no tiene por qué ser indicador de la buena calidad de básquetbol de la conferencia. Como además queremos saber si una conferencia tiene mejor "calidad de básquetbol" que otra, nos interesa saber cuántas victorias tiene una conferencia sobre otra. Es decir, contabilizar solo los partidos inter-conferencias.

```

1 MATCH (s:Season)-[]-(tEast:TeamSeason)-[:HOME_TEAM|AWAY_TEAM]-(g:Game)
  -[:HOME_TEAM|AWAY_TEAM]-(tWest:TeamSeason)-[]-(cWest:
    Conference)
2 WHERE cWest.id='West' and s.id='2021'
3 MATCH (tEast)-[]-(cEast:Conference) where cEast.id='
  East'
4 OPTIONAL MATCH (g)-[eastWins:WINNER]-(tEast)
5 RETURN count(g) as Game_count, count(eastWins) as East_wins, count(g)-
  count(eastWins) as West_wins

```

Game_Counts	East_wins	West_wins
477	240	237

Table 5: Competitividad entre conferencias

La consulta retorna que la conferencia Oeste ganó el 52,5% de los partidos entre conferencias A pesar de que en la temporada de 2021 la conferencia Este mostró mayor competitividad interna, se percibe cierta superioridad de la conferencia Oeste sobre la del Este basándonos en las victorias.

Jugadores

En esta sección se aborda la entidad correspondiente a los jugadores. Se ha realizado una nueva consulta que calcula la máxima racha de victorias de cada jugador en los partidos en los que ha participado algún tiempo dentro de la cancha.

El cálculo de la máxima racha de victorias proporciona información valiosa sobre el desempeño de los jugadores en términos de su capacidad para contribuir a los resultados exitosos de su equipo. Esta métrica nos permite identificar y analizar la consistencia y el impacto de los jugadores en los momentos en que sus equipos han experimentado una racha ganadora.

Al considerar el tiempo de juego mínimo como criterio para incluir los partidos en el cálculo, nos aseguramos de que solo se tengan en cuenta aquellos encuentros en los que los jugadores han tenido alguna participación de tiempo sobre la cancha. Esto nos permite evaluar su influencia directa en el resultado final y capturar su contribución efectiva a las victorias de su equipo.

```
1 MATCH (p:Player)-[r:PLAYED]-(g:Game)-[:WINNER]-(ts:TeamSeason)-[]-(
   winningTeam:Team)
2 WHERE r.time_played>0
3 MATCH (ts)-[]-(s:Season) where s.id='2021'
4 WITH p, g, r, winningTeam
5 ORDER BY g.date ASC
6 WITH p, collect(r.team = winningTeam.id) AS hasWon
7 WITH p, reduce(
8     acc = [0], outcome IN hasWon |
9     CASE
10         WHEN outcome THEN [acc[0] + 1] + acc[1..size(acc)]
11         ELSE [0] + acc
12     END
13 ) AS streaks
14 UNWIND streaks as streak
15 RETURN p.name AS playerName, max(streak) AS longestWinningStreak
16 ORDER BY longestWinningStreak DESC LIMIT 10
```

Player	Longest_win_streak
Moses Moody	21
Devin Booker	20
Jae Crowder	18
Chris Paul	18
Cameron Johnson	18
Mikal Bridges	18
JaVale McGee	18
Landry Shamet	17
Bismack Biyombo	17
Cameron Payne	16

Se observa que el jugador con mayor racha de victorias participativa es "Moses Moody" el cual jugó para los Warriors en esa temporada. Se ven nombres fuertemente relacionados como Devin Booker y Chris Paul los cuales coinciden en pertenecer al equipo que tuvo la mayor racha de victorias en la temporada de 2021, Phoenix Suns con 18 victorias consecutivas como equipo.

Distancia

En el proyecto anterior, se diseñó una consulta para calcular la distancia entre jugadores en función de los juegos en los que ambos participaron. Sin embargo, en nuestro afán de ampliar la noción de distancia, hemos llevado esta idea un paso más allá. Ahora, no solo nos interesa la distancia entre jugadores en términos generales, sino que hemos introducido dos conceptos adicionales: la distancia rival y la distancia aliada.

La distancia rival se enfoca específicamente en establecer conexiones entre jugadores que han sido rivales en los juegos en los que participaron. Esto nos permite analizar y comprender mejor las interacciones y rivalidades entre jugadores a lo largo de los juegos.

De manera similar, la distancia aliada se centra en crear caminos entre compañeros de equipo, explorando la relación y la colaboración entre jugadores que han sido parte del mismo equipo en diferentes juegos.

Para lograr esto, hemos desarrollado consultas específicas que nos permiten calcular estas distancias y generar rutas que reflejen las conexiones entre jugadores en función de su relación rival o aliada.

```
1 MATCH path = (start:Player {name: 'LeBron James'}) -[:PLAYED*2..6] - (end
  :Player {name: 'Facundo Campazzo'})
2 WITH start, end, path, nodes(path) AS nodes, relationships(path) AS
  rels
3 WHERE all(i IN range(1, size(nodes)-1) WHERE
4   (labels(nodes[i])[0] = 'Player') OR (labels(nodes[i])[0] = 'Game'
5     AND (rels[i-1]).team <> (rels[i]).team)
6 )
7 RETURN path, length(path)
8 ORDER BY length(path) ASC
LIMIT 1
```

Observar que la noción de distancia aliada es exactamente la misma consulta, pero en la línea 4 se cambia el distinto "<>" por un símbolo de igualdad "=". Esto indica que los jugadores a través de los que se hace el camino sean compañeros de equipo para el juego que jugaron.

Adjuntamos resultados interesantes entre LeBron y Batista cómo se hizo en el anterior trabajo, tanto distancia rival como aliada.

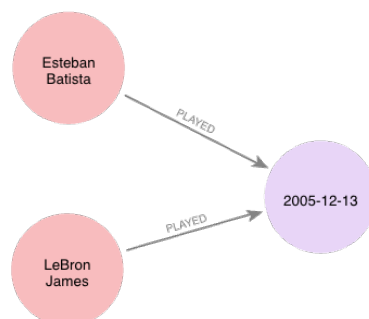


Figure 3: Distancia rival Batista - LeBron

Con el propósito de seguir analizando jugadores rioplatenses en la NBA, tomaremos como ejemplo al jugador Argentino "Facundo Campazzo", quien ha estado jugando en la liga desde la temporada 2020 hasta la actualidad. En nuestro análisis, nos enfocaremos en establecer un camino de conexión entre Campazzo el jugador Uruguayo Batista, quien jugó en la liga durante las temporadas 2005 hasta 2007. Al ejecutar la consulta para obtener el camino por rivales, nos encontramos con el famoso jugador "Carmelo Anthony" que disputó 19 temporadas en la NBA y pudo competir tanto contra el Uruguayo en 2005 como contra el Argentino en 2020.

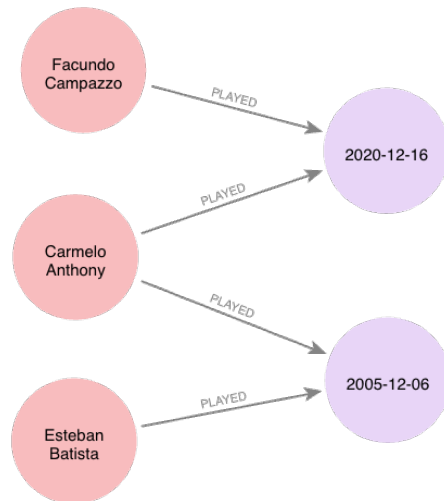


Figure 4: Distancia rival Batista - Campazzo

Sin embargo, es importante destacar que no podemos esperar encontrar un camino directo de longitud 2 entre estos dos jugadores, ya que no coincidieron en las mismas temporadas. Esto implica que nunca tuvieron la oportunidad de competir juntos en el mismo equipo ni enfrentarse como rivales en la cancha.

Finalmente, evaluamos la distancia aliados entre el LeBron James y el Uruguayo Batista.

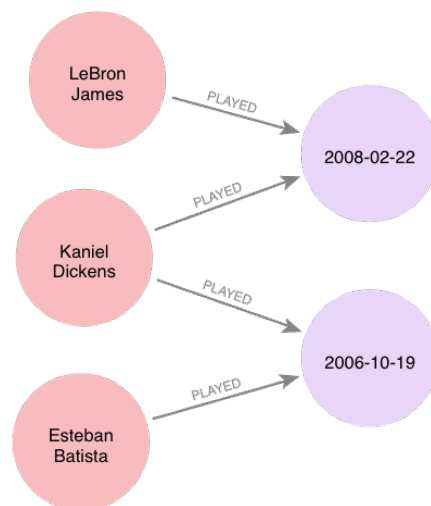


Figure 5: Distancia aliados de Batista - Lebron

Conclusión

En resumen, se han realizado varias mejoras y agregados que han enriquecido el modelo del grafo y han permitido realizar consultas más específicas y eficientes.

En primer lugar, se ha agregado el nodo "Arena" para representar las sedes de los juegos. Esta adición ha evitado la duplicación de información y ha permitido relacionar los juegos con sus respectivas arenas. Respecto a las Arenas se realizó un análisis de localidad sobre las mismas.

Además, se ha introducido la entidad "TeamSeason" para reducir la densidad de los nodos "Team" y mejorar el rendimiento de las consultas. Esta entidad representa la temporada de un equipo y agrupa la información específica de cada temporada.

Otro agregado importante ha sido el nodo "Conferencia", que permite realizar análisis y consultas basadas en la conferencia a la que pertenecen los equipos. Esta adición amplía las posibilidades de segmentación y estudio del rendimiento de los equipos según su conferencia. Se definió una noción de competitividad interna en las conferencias y se compararon las victorias de una conferencia sobre otra. Se observó que para el caso de la temporada 2021, ambas conferencias tienen un nivel de básquetbol y competitividad similar. Esto da a entender que las divisiones están equilibradas.

Además de estos cambios en la estructura del grafo, se han implementado consultas adicionales para explorar más a fondo los datos. Estas consultas incluyen el cálculo de los jugadores con rachas de victorias más largas en una temporada en los que participaron del juego.

Por último, se ha extendido la idea de la distancia entre jugadores, estas nociones de distancia han brindado una nueva perspectiva para analizar las interacciones y conexiones entre los jugadores en el contexto de los datos deportivos. Nos han permitido obtener información relevante sobre las relaciones y experiencias individuales de los jugadores en términos de los equipos con los que han jugado y los oponentes a los que se han enfrentado.

En conjunto, las mejoras y consultas adicionales implementadas en el modelo del grafo han enriquecido el análisis de los datos deportivos al considerar aspectos como las relaciones de los jugadores con sus compañeros y oponentes, así como la segmentación por temporada y conferencia. Estas mejoras han brindado nuevas perspectivas y posibilidades de análisis, permitiéndonos obtener información más precisa y significativa. No obstante, a pesar de las mejoras implementadas, es importante destacar que los datos utilizados en este trabajo presentan algunas limitaciones. Aunque se cuenta con información valiosa sobre estadísticas individuales de los jugadores, como asistencias, bloqueos y robos, no se dispone de datos relacionales que permitan identificar a quién se realizó la asistencia, a quién se bloqueó el tiro o a quién se robó la pelota. Con la información relacional de las asistencias se podría hacer un análisis de sinergia en un equipo.

References

- [1] De Marzi , Max (2015) *Modeling Airline Flights in Neo4j*, Max De Marzi Blog.
- [2] Lauga, Nathan. NBA Games - Dataset. <https://www.kaggle.com/datasets/nathanlauga/nba-games>
- [3] Borba, M. Repositorio del proyecto BDNR_G4. https://github.com/M-Borba/BDNR_G4