**Kedir Nasir Omer Date: Nov. 28/2022**

**UC SanDiego- Big Data Specialization**

**4. Graph Analytics with Chat Data Using Neo4j**

* **A Graph Analytics Approach to Simulated Chat Data**

# Modeling Chat Data using a Graph Data Model

A network based on user chat interactions makes up the graph model. Users on the same team can join and leave a chat session, which can be started by one of them. When a user creates a post, user interaction starts. Another user may be mentioned by a user. With a timestamp, every relationship between entities is recorded.

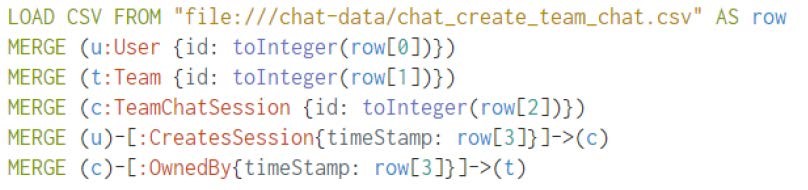
# Creation of the Graph Database for Chats

Describe the steps you took for creating the graph database.

## Write the schema of the 6 CSV files

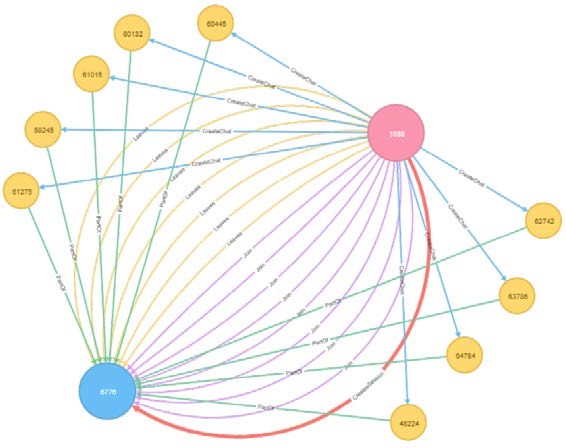
|  |  |
| --- | --- |
| chat\_create\_team\_chat.csv | userID teamID teamChatSessionID timestamp |
| chat\_join\_team\_chat.csv | userID teamChatSessionID timestamp |
| chat\_leave\_team\_chat.csv | userID teamChatSessionID timestamp |
| chat\_item\_team\_chat.csv | userID teamChatSessionID chatItemID timestamp |
| chat\_mention\_team\_chat.csv | chatItemID userID timestamp |
| chat\_respons\_team\_chat.csv | chatItemID\_1 chatItemID\_2 timestamp |

## Explain the loading process and include a sample LOAD command



The first line load the csv from the specific location one row at a time. From the second line to fourth, create the nodes for User, Team, TeamChatSession with a specific column converted to integer, this field is used by the id attribute. The fifth and sixth lines create CreatesSession and OwnedBy edges and link the nodes previously created. The edges have a timestamp property filled by the fourth column of schema.

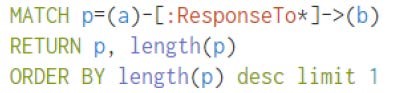
Present a screenshot of some part of the graph you have generated. The graphs must include clearly visible examples of most node and edge types.

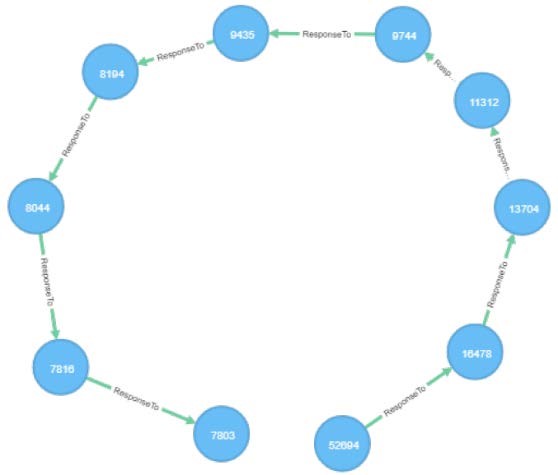


# Finding the longest conversation chain and its participants

Report the results including the length of the conversation (path length) and how many unique users were part of the conversation chain. Describe your steps. Write the query that produces the correct answer.

* How many chats are involved in it?





The longest conversation chain in the chat data has path length 9, therefore 10 chats are involved in it.

* How many users participated in this chain?



With 9 as longest path, count the number of distinct users who create ChatItem in this longest path. The query returns 5.

# Analyzing the relationship between top 10 chattiest users and top 10 chattiest teams

* Describe your steps from Question 2. In the process, create the following two tables. You only need to include the top 3 for each table. Identify and report whether any of the chattiest users were part of any of the chattiest teams.

## Chattiest Users

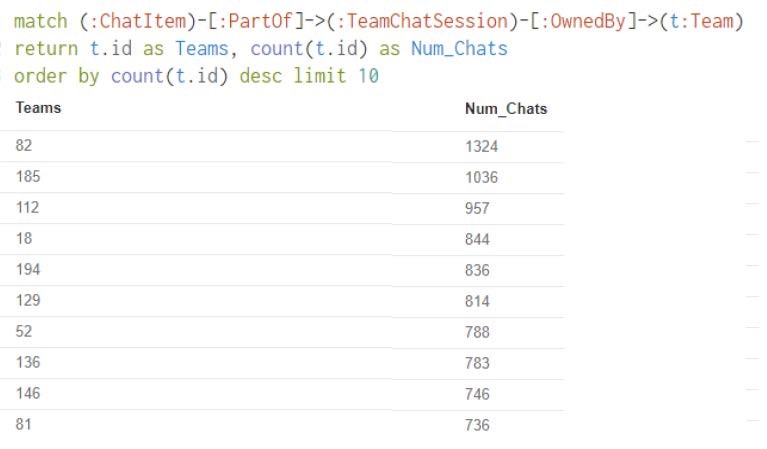
* Determine the number of chats created by a user from the CreateChat edge



|  |  |
| --- | --- |
| Users | Number of Chats |
| 394 | 115 |
| 2067 | 111 |
| 209 | 109 |

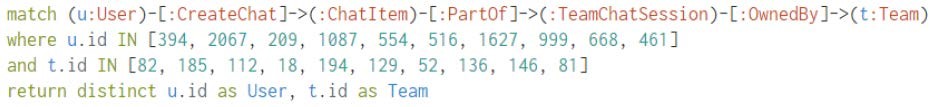
## Chattiest Teams

Match all ChatItem with a PartOd edge and connect them with a TeamChatSession node that have an OwnedBy edge connection them with any other node.



|  |  |
| --- | --- |
| Teams | Number of Chats |
| 82 | 1324 |
| 185 | 1036 |
| 112 | 957 |

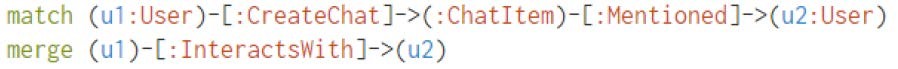
* Finally, present your answer, i.e. whether or not any of the chattiest users are part of any of the chattiest teams.



This query is used to investigate if the chattiest user is part of any chattiest team and it return one result, userID 999 is part of teamID 52. How Active Are Groups of Users?

Describe your steps for performing this analysis. Be as clear, concise, and as brief as possible. Finally, report the top 3 most active users in the table below.

## Connect mentioned users



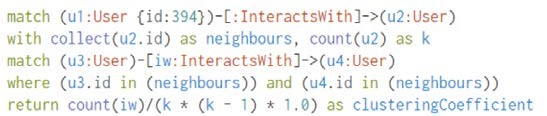
## Connect users’ responses with the chat creator



## Eliminate all self interaction



* Calculate the cluster coefficient.



## Most Active Users (based on Cluster Coefficients)

|  |  |
| --- | --- |
| User ID | Coefficient |
| 394 | 0.9167 |
| 2067 | 0.7679 |
| 209 | 0.9524 |