**Graph Analytics**

**Modeling Chat Data using a Graph Data Model**

**The chat data defines 4 nodes ( User, Team, TeamChatSession, ChatItem ) and following edges CreatesSession, OwnedBy, Joins, Leaves, CreateChat, part of, Mentioned, and InteractsWith.**

**Nodes are using properties “id” while the edges use timestamp as property to define when the actions were done.**

**Creation of the Graph Database for Chats**

Describe the steps you took for creating the graph database. As part of these steps

1. Write the schema of the 6 CSV files

**chat\_create\_team\_chat.csv :**

column 1 : User {id: Integer}

column 2 : Team {id: Integer}

column 3 : TeamChatSession {id: Integer}

column 4 : CreatesSession {timestamp}

column 5 : OwnedBy{timestamp}

**chat\_join\_team\_chat.csv :**

column 1 : User {id: Integer}

column 2 : TeamChatSession {id: Integer}

column 3 : Joins{timestamp}

**chat\_leave\_team\_chat.csv :**

column 1 : User {id: Integer}

column 2 : TeamChatSession {id: Integer}

column 3 : Leaves{timestamp}

**chat\_item\_team\_chat.csv :**

column 1 : User {id: Integer}

column 2 : TeamChatSession {id: Integer}

column 3 : ChatItem {id: toInteger}

column 4 : CreateChat{timestamp} and PartOf{timestamp}

**chat\_mention\_team\_chat.csv :**

column 1 : ChatItem {id: Integer}

column 2 : User {id: Integer}

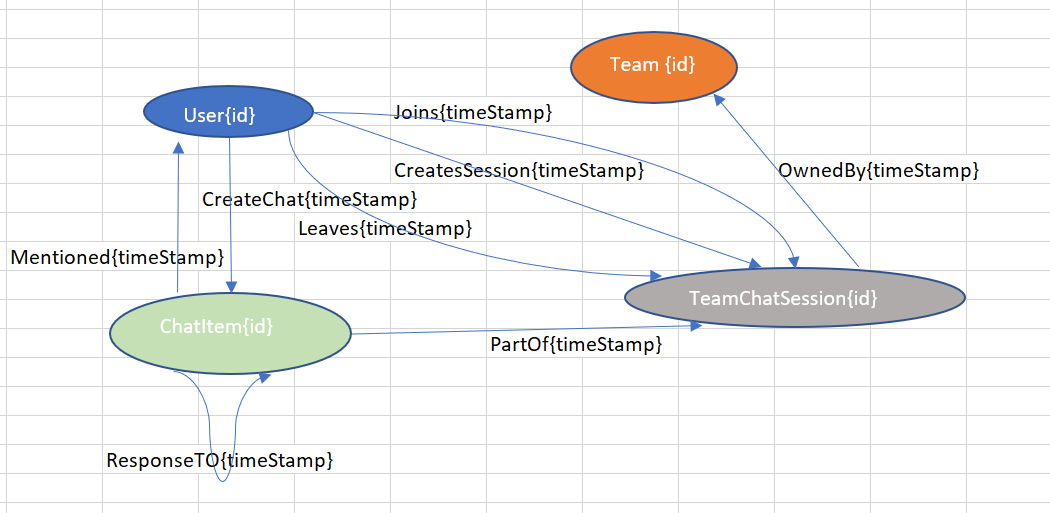
column 3 : Mentioned{timestamp}

**chat\_respond\_team\_chat.csv :**

column 1 : ChatItem {id: Integer}

column 2 : ChatItem {id: Integer}

column 3 : ResponseTO {timeStamp}



1. Explain the loading process and include a sample LOAD command

**1st line of command below specify the path of the CSV file to load . The CSV file contains 3 columns:**

**1st column is loaded as User on 2nd line of code, 2nd column is loaded as TeamChatSession at 3rd line . 3rd column contains timestamp is loaded at line 4.**

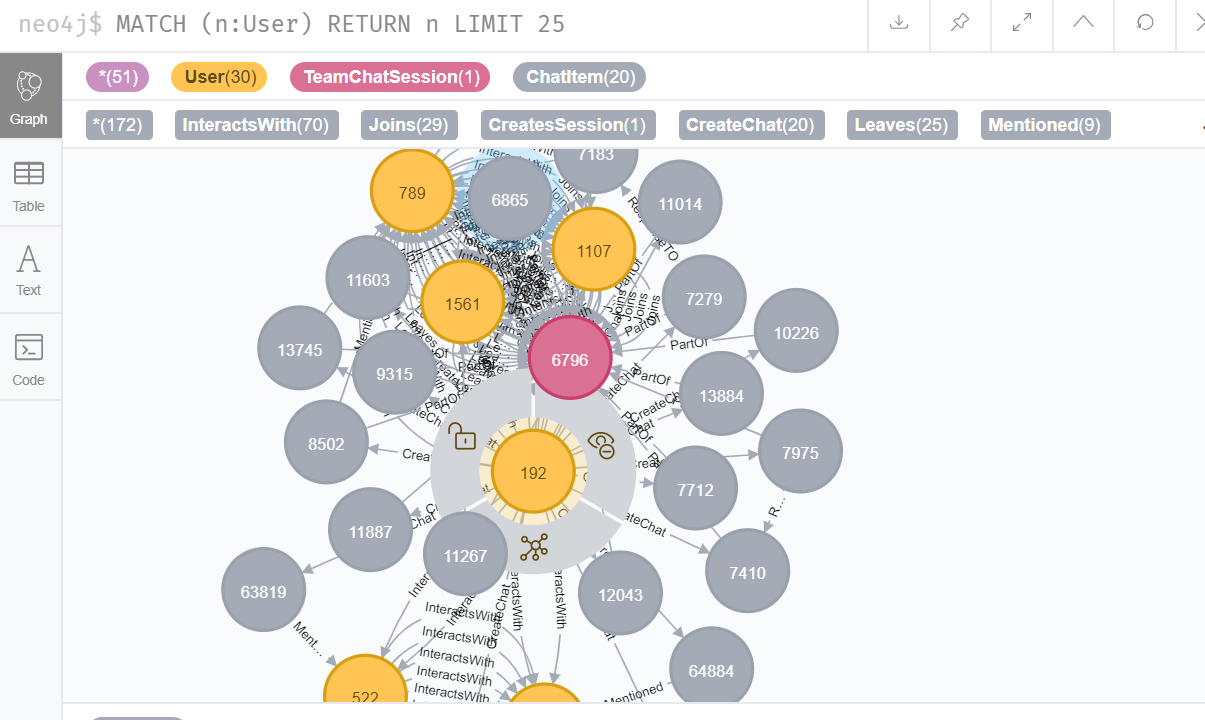
**LOAD CSV FROM "file:///C:chat\_data/chat\_join\_team\_chat.csv" AS row**

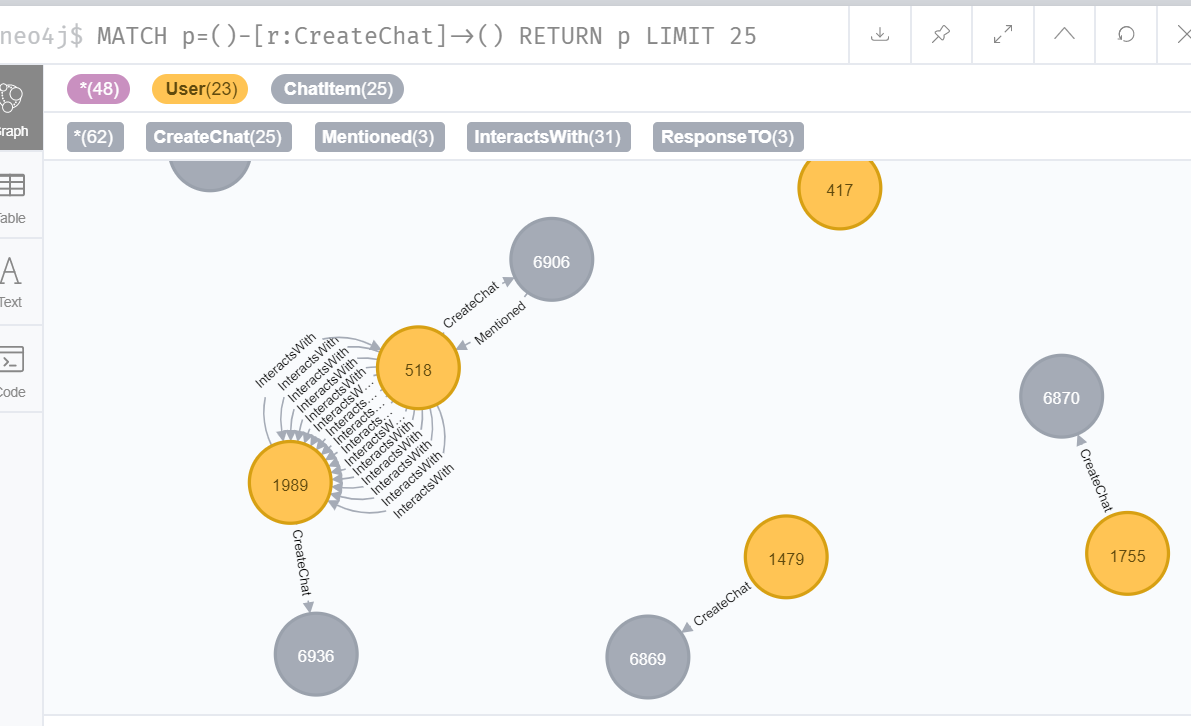
**MERGE (u:User {id: toInteger(row[0])})**

**MERGE (c:TeamChatSession {id: toInteger(row[1])})**

**MERGE (u)-[:Joins{timeStamp: row[2]}]->(c)**

1. 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. Below are two acceptable examples. The first example is a rendered in the default Neo4j distribution, the second has had some nodes moved to expose the edges more clearly. Both include examples of most node and edge types.





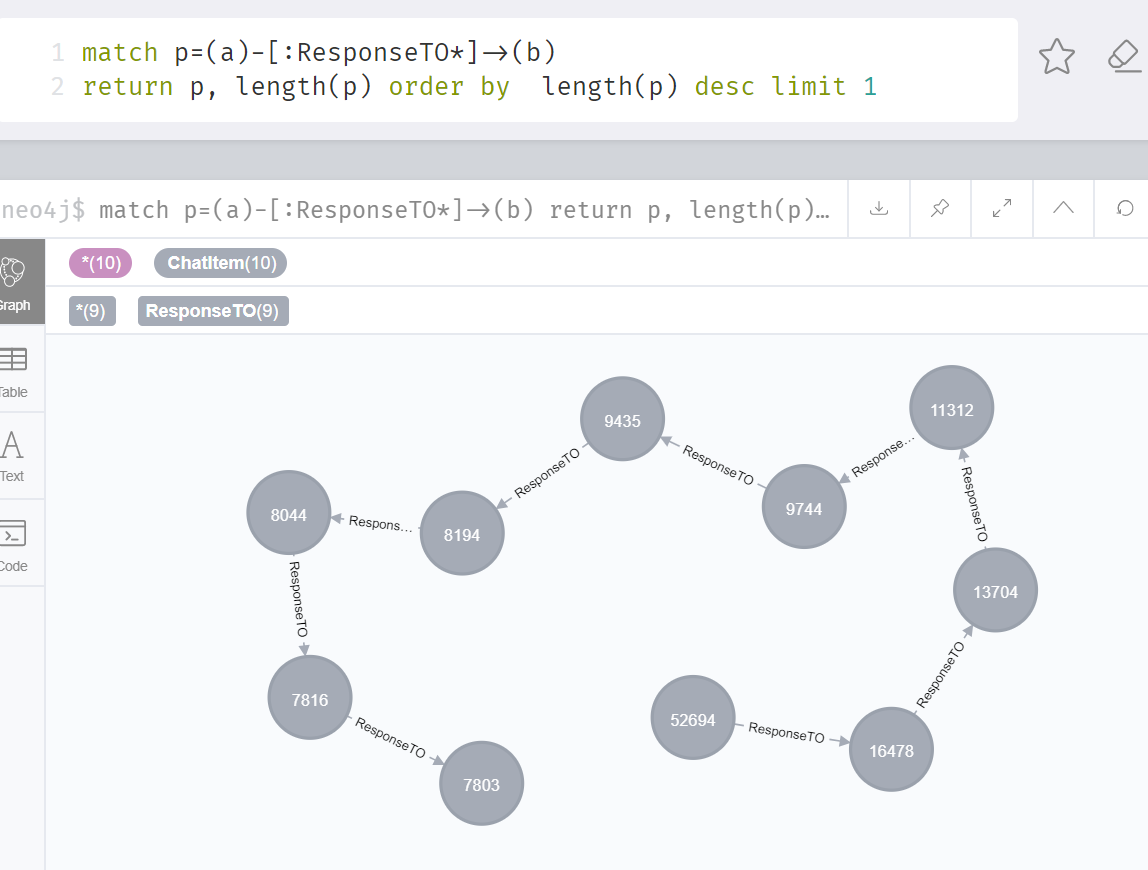
**Finding the longest conversation chain and its participants**

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.

**match p=(a)-[:ResponseTO\*]->(b)**

**return p, length(p) order by length(p) desc limit 1**

**=> 9 Chats involved in the longest conversation;**

****

**5 unique users in this conversation:**

****

**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**

|  |  |
| --- | --- |
| **Users** | **Number of Chats** |
| **394** | **115** |
| **2067** | **111** |
| **1087** | **109** |

**Chattiest Teams**

|  |  |
| --- | --- |
| **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.

**YES, User 999 that is part of top 10 chattiest users, belong to team 52 that is part of the top 10 chattiest team. We find all the users who joins a chatSession and corresponding Teams that own those TeamChatSession; Among those users, we search by User.id (top 10 chattiest user id) , and we can find their corresponding teams. Only user 999 teams is present in the top 10 chattiest team (team 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.

**The answer of this question is based on the new edges created ( InteractsWith ) to estimate of how “dense” the neighborhood of a node is.**

**A clustering coefficient is a score ranging from 0 (a disconnected user) to 1. For example, if the number of neighbors of a node is 5, then the clustering coefficient of the node is the ratio between the number of edges amongst these 5 neighbors(not counting the given node) and 5 \* 4 / 2 (all the pairwise edges that could possibly exist). Thus the denominator is k \* (k-1) / 2 if the number of neighbors of the node is k.**

**We get the list of neighbors and the number of neighbors of a node based on the “InteractsWith” edge.**

**Most Active Users (based on Cluster Coefficients)**

|  |  |
| --- | --- |
| **User ID** | **Coefficient** |
| 209 | 0.952 |
| 554 | 0.90 |
| **1087** | 0.8 |