Graph Analytics

Modeling Chat Data using a Graph Data Model

The purpose of this exercise is to identify the chattiest users and teams as well as the active groups of users. User interaction can be linked to the stickiness of the users.

Creation of the Graph Database for Chats

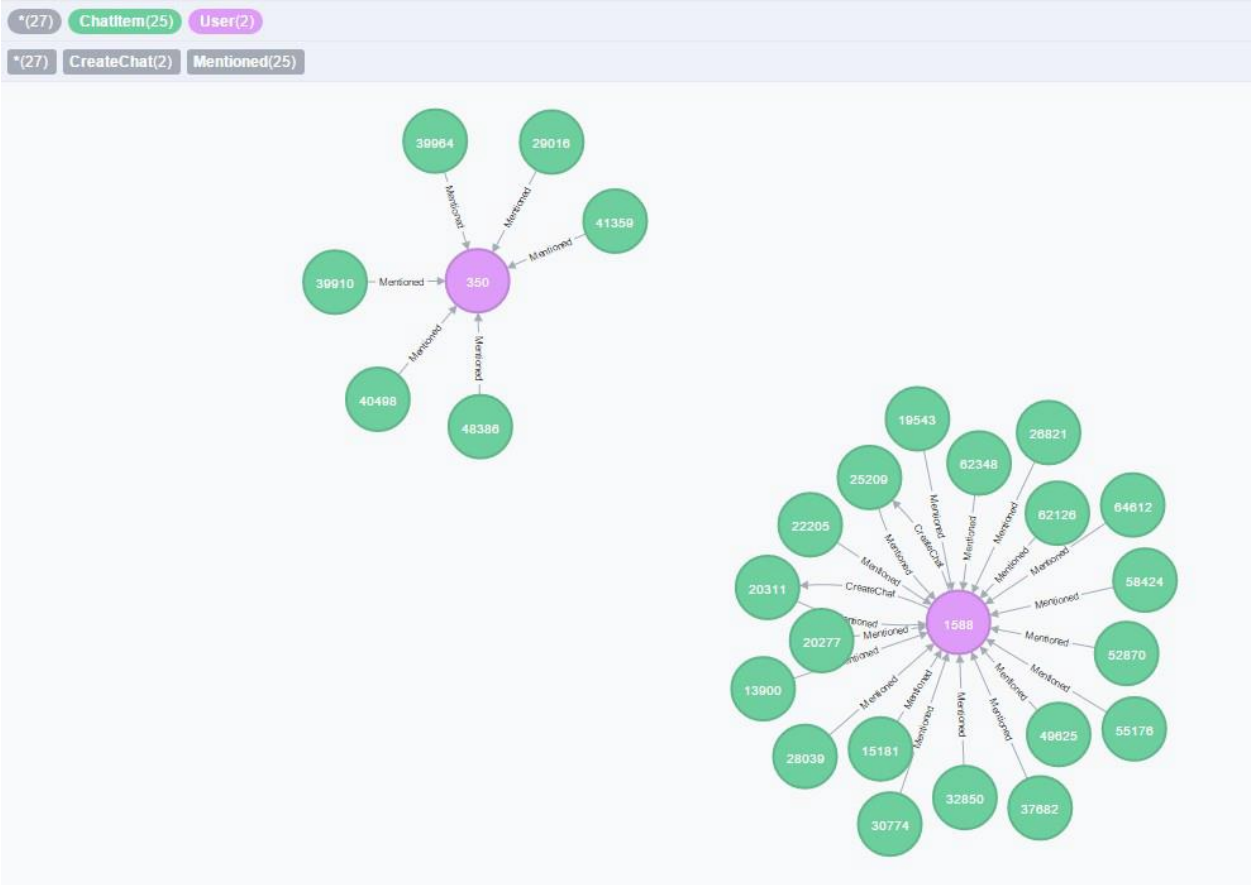
Describe the steps you took for creating the graph database. As part of these steps

i) chat_create_team_chat.csv (userid, teamid, TeamChatSessionId, timestamp)
chat_item_team_chat.csv (userid, teamid, ChatItemid,timestamp)
chat_join_team_chat.csv (userid, TeamChatSessionID, timestamp)
chat_leave_team_chat.csv (userid, chatid, timestamp)
chat_mention_team_chat.csv (ChatItem, userid, timestamp)
chat_respond_team_chat.csv (chatItemId1, chatItemId2)

ii) Date is loaded by identifying nodes (defined by MERGE command) and then loading in relationships with the nodes using time stamps.
Screenshot of LOAD command:

```
LOAD CSV FROM "file:///U:/chat/chat_leave_team_chat.csv" AS row
MERGE (u:User {id: toInt(row[0])})
MERGE (c:TeamChatSession {id: toInt(row{1})})
MERGE (u)-Leaves{timeStamp: row{2})})->(c)
```

iii) Present a screenshot of the graph generated:



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=(i)-[r:ResponseTo*]->()
with p
ORDER BY length(p) DESC
LIMIT 1
match (i:ChatItem)<-[r:CreateChat-(u:User)
where i in nodes(p)
return count(distinct u)
```

Unique number of users: 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

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.

No, the chattiest users are not part of the chattiest teams.

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.

Most Active Users (based on Cluster Coefficients)

User ID	Coefficient
394	1
2067	0.79
209	0.95
1087	0.8
554	0.91
1627	0.79
516	0.95
999	0.87
461	1
668	0.7