Graph Analytics

Modeling Chat Data using a Graph Data Model

Graph analytics approach is used to simulate chat data as it relates to the Catch the Pink Flamingo game. Currently this chat data is purely numeric, no text. Analytically, it can still serve a useful purpose in revealing certain types of behaviours which can only be observed within a graph analytics context.

Creation of the Graph Database for Chats

Steps taken for creating the graph database. As part of these steps

i) Schema of the 6 CSV files:

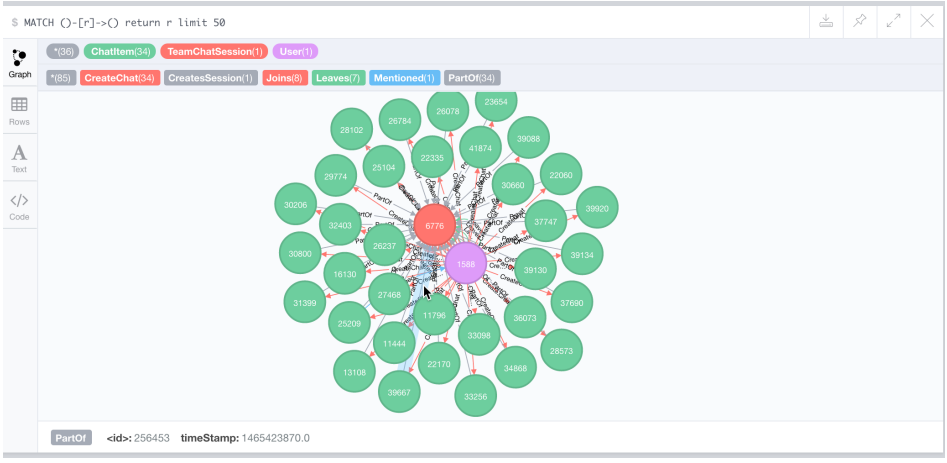
1. chat_create_team_chat.csv: userid, teamid, TeamChatSessionID, timestamp
2. chat_item_team_chat.csv: userid, teamchatsessionid, chatitemid, timestamp
3. chat_join_team_chat.csv: userid, TeamChatSessionID, teamstamp
4. chat_leave_team_chat.csv: userid, teamchatsessionid, timestamp
5. chat_mention_team_chat.csv: ChatItem, userid, timeStamp
6. chat_respond_team_chat.csv: chatid1, chatid2, timestamp

ii) The loading process and a sample LOAD command:

1. Locate the chat-data folder you downloaded.
2. Start the Neo4J graph database in your Browser.
3. Load the CSV data files into Neo4J.

```
LOAD CSV FROM "file:///chat-data/chat_create_team_chat.csv" AS row
MERGE (u: User {id: toInt(row[0])}) MERGE (t: Team {id: toInt(row[1])})
MERGE (c: TeamChatSession {id: toInt(row[2])})
MERGE (u)-[:CreatesSession{timeStamp: row[3]}]->(c)
MERGE (c)-[:OwnedBy{timeStamp: row[3]}]->(t)
;
```

iii) A screenshot of some part of the graph you have 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.

Query:

Match p = (i1) - [:ResponseTo*] -> (i2)

return length(p)

order by length(p) desc limit 1

Result:

length(p) = 9

Query 2:

match p = (i1)-[:ResponseTo*]->(i2)

where length(p) = 9

with p

match (u)-[:CreateChat]->(i)

where i in nodes(p)

return count(distinct u)

Result:

count(distinct(u)) = 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

No, none of the chattiest users are part of any of the chattiest teams.

How Active Are Groups of Users?

Most Active Users (based on Cluster Coefficients)

User ID	Coefficient
394	0.9167
2067	0.7679
209	0.9524