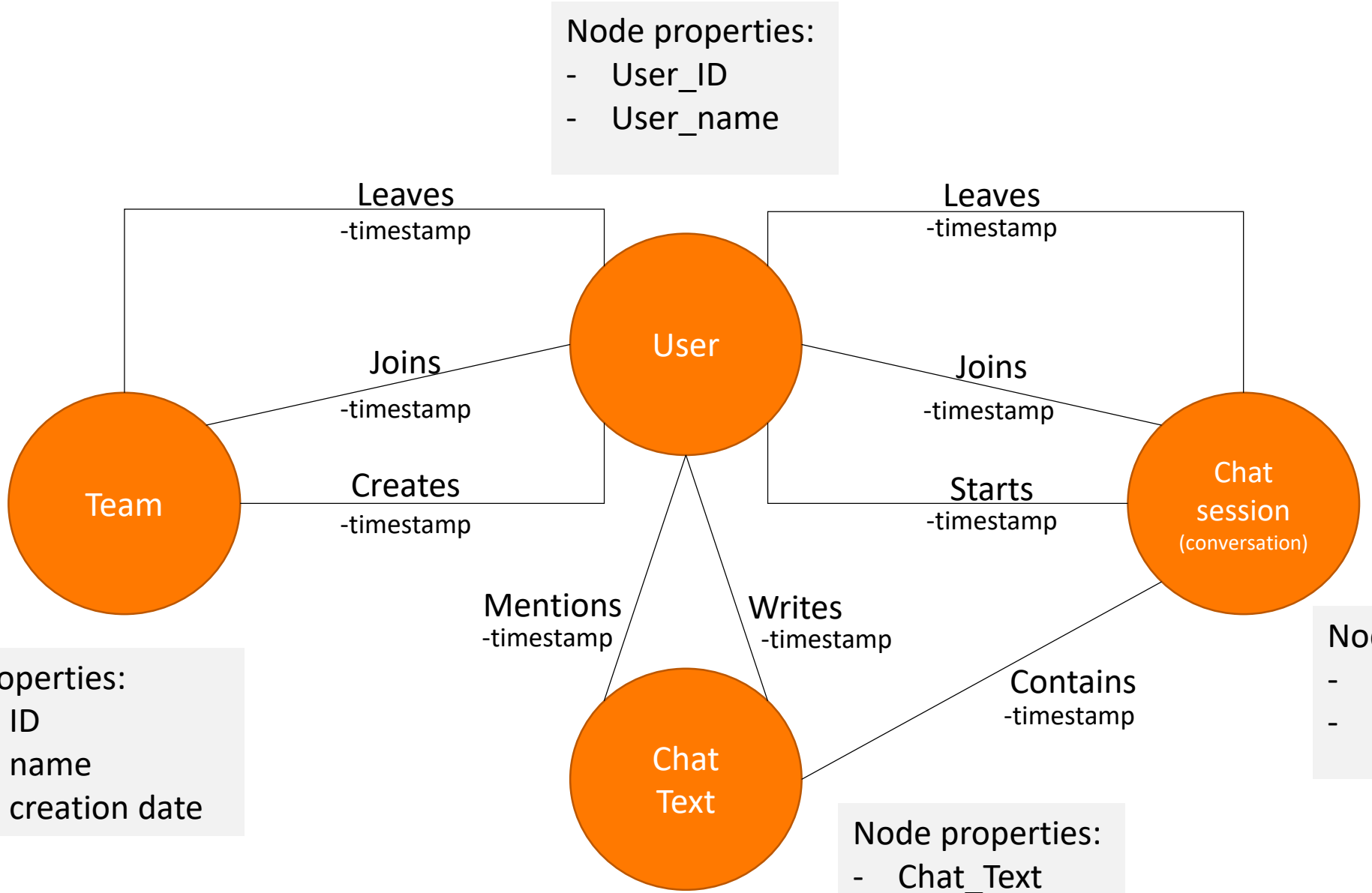
Catch the pink flamingo : Chat Data graph model



Q1: Which teams are having more conversations?

Assume that a conversation is a chat session and since we have created another node (Team), connected to the node User, therefore we can count the number of chat session by user and then group it by team to obtain which team having more conversations.

Q2: Do users chat more (or less) before they leave a team?

The property timestamp added to each action edge and the node Team, created as mentioned above in the data graph model, allows us to count the number of conversation when users has left a team and then compare it with the number of conversation (chat session) before the leave it.

Q3: What are the dominant terms (words) used in a chat session within a specific time period?

Since we have the timestamp property on the action edge writes between the nodes Chat session and Chat Text so we can analyze the occurrence of the terms used and retrieve the dominant terms using the vector space model.

Q4: Which users are most active in a specific chat session?

The most active users in a specific chat session are the users with the highest interactions (writes chat text, starts and joins chat session). So by calculating the count of these interactions by users we can retrieve them and highlight the most active users by chat session.

Q5: How many chat sessions is a user participating in at the same time?

While participation is not defined clearly, we assume that the participation in a session is when the user joins a chat session
And/or when he writes a chat text.
For the first case described above we aggregate the count of chat session by timestamp of joins and by users , then filter resultts greater than 1.
For the second case, we aggregate the count of writes by timestamp of writes and by users, thereby we retrieve the count of writes by user at a specific time then we filter results greater than 1.
If we consider that a participation is defined by the two cases at a time then we join the two results described above based on the user and the the timestamp of joins and writes.