

TD Neo4j

Analyse de données Twitter

1 Cas d'étude Neo4j

En tant qu'analyste de données sociales sur le réseau Twitter, afin de comprendre les dynamiques en jeu sur le réseau, on souhaiterait connaître:

- quels sont les tweets les plus retweetés
- quels sont les utilisateurs les plus suivis
- quels sont les utilisateurs les plus mentionnés
- quels sont les hashtags les plus populaires

2 Modélisation

En considérant les données accessibles sur la plateforme Twitter modélisées sur la figure 3, quelle modélisation mettriez-vous en oeuvre pour répondre à ces objectifs ?

3 Interrogation

On suppose le graphe modélisé ci-dessus implanté dans Neo4j. Donnez les requêtes Cypher permettant d'obtenir :

- Q1. le nombre d'utilisateurs
- Q2. le nombre de tweets contenant le hashtag *WhiteRabbit*
- Q3. le nombre d'utilisateurs différents qui ont tweeté un tweet contenant le hashtag *AliceInWonderland*
- Q4. le nom des followers de *@Alice*
- Q5. le nom des utilisateurs suivis par *@Alice*
- Q6. le nom des utilisateurs ayant plus de 10 followers.
- Q7. Les utilisateurs étant à la fois followers et followees de *@Alice* (c'est à dire qui suivent et sont suivis par *@Alice*)
- Q8. les 10 tweets les plus populaires¹
- Q9. les 10 hashtags les plus populaires²

4 Comparaison au modèle relationnel

Comparez et discutez les réponses aux questions précédentes utilisant le langage Cypher à celles présentées ci-dessous en SQL (basées sur le modèle relationnel de la section 2).

```
[Q1] select count(*) from User;
```

¹Ici, la popularité se mesure au nombre de fois où le tweet est favori.

²Ici, la popularité se mesure au nombre de tweets dans lesquels ils apparaissent.

```

[Q2] select count(*) from Tweet_Hashtag
      where hashtag='WhiteRabbit';

[Q3] select count(distinct idUser)
      from Tweet T, Tweet_Hashtag TH
      where T.idTweet = TH.idTweet
      and hashtag='AliceInWonderland';

[Q4] select U2.screenName
      from User U1, User U2, User_Follow UF
      where U1.screenName='Alice'
      and U1.iduser = UF.idUser
      and UF.idUserFollow = U2.idUser;

[Q5] select U2.screenName
      from User U1, User U2, User_Follow UF
      where U1.screenName='Alice'
      and U1.iduser = UF.idUserFollow
      and UF.idUser = U2.idUser;

[Q6] select screenName
      from User, User_Follow UF
      where U.idUser = UF.idUser
      group by U.idUser, screenName
      having count(*) > 10;

[Q7] select U2.screenName
      from User U1, User U2, User_Follow UF
      where U1.screenName='Alice'
      and U1.iduser = UF.idUser
      and UF.idUserFollow = U2.idUser
      intersect
      select U2.screenName
      from User U1, User U2, User_Follow UF
      where U1.screenName='Alice'
      and U1.iduser = UF.idUserFollow
      and UF.idUser = U2.idUser;

[Q8] select *
      from (select *
            from tweet
            order by nbFavorites desc)
      where rownum < 11;

[Q9] select * from (select hashtag
                    from Tweet_Hashtag
                    group by hashtag
                    order by count(*) desc)
      where rownum < 11;

```