RECOMMENDATION ALGORITHM

Neo4j graph structure is as follows:

- There are User, Song, Album, Performer nodes and Listens relationships between User nodes and Song nodes.
- Listens relationship keeps the number of times a specific user listens to a specific song, the time of action when a specific user last listens to a specific song, the frequency which specifies how frequently a user listens to a specific song over a period of time (between the initial time of action and today).
- Other attributes of Listens relationship are shown in Figure 1.
- The user logs are used to create distinct User nodes and Listen relationships. Song, performer and album and relation files are also used to create Song, Performer and Album nodes.
- An example query that traverses the graph is as follows:

Figure 1: User node, Song nodes and Listen relationship

Algorithm:

- User 2484378 is given to the findNeighbor function in order to get recommendations.
- The algorithm first finds five similar users to User 2484378. The similar users must have listened to common songs with User 2484378 and rate for these common songs must be above 0.6. Besides, similar user common song listening frequencies must be close to the listening frequencies of User 2484378. The users that have listened to a common song only once are discarded from this set.
- Similar users are User 1667498, User 1660823, User 1608170, User 1612958, and User 1608461.
- Then all the songs of the similar users are compared to the most frequently played five songs of User 2484378. Weka K-nearest neighbor is used to find similarity over frequency and rating values during comparison. 30 songs closest to the first of the five songs of User 2484378 are found. Then this step is repeated for other four songs of User 2484378. A set of 150 songs are gathered. Then five songs are chosen among this set as the recommended songs. During this step the songs that occur more than three times in this set have priority and they are placed in the final set. Then the songs that have the same album, same performer or same genre as the songs of the User 2484378 have the priority. After all these cases are considered, the final set is filled with most frequently played songs of the set with 150 songs.
- As soon as the final set is formed, the recommendation list is returned.
- The list of recommended songs:
  Deli Ziynet Sali Sonsuz Ol
  ?laç Günce Koral Hayat A?ktan Geçer
  Ben Sen Olamam (Suat Ate?da?l? Remix) Emir Ben Sen Olamam
  Acele Deniz Seki ?effaf
  Güvercin Sezen Aksu Deniz Y?ld?z?

EVALUATION

DATA SET

October user logs (logs of 30 days) are used to create test and train sets. First 24 days of October logs are used for train dataset. Last 6 days of October logs are used for test dataset. This means that 80% of

the entire dataset belongs to the training set and 20% of the entire dataset belongs to the testing dataset.

Test Dataset:

- Number of users is 82173
- Number of songs is 1204033
- Average number of songs per user is 18.78 (Number of Listens relationship / number of users)
- Average number of listening per user is 25.67 ((Number of Listens relationship*number of action) / number of users)

Train Dataset:

- Number of users is 176650
- Number of songs is 1204033
- Average number of songs per user is 30.68 (Number of Listens relationship / number of users)

EVALUATION METRICS

- Precision metric is used: Successful recommendations/number of songs recommended. For each user five songs are recommended. Therefore this number is independent of the total number of songs that the user listened.

RESULTS

Method	Precision	Result
K-nearest neighbor	Tp /(Tp + Fp)	0.021

- After evaluated User 1667498 recommendations with precision metric, it is taken an evaluation result. Evaluation result is 0.2. It means 1 of the 5 song is in the test data and our achievement is 0.2 for that user.
- When looking overall, sum of precisions for all users / number of user gives overall achievement and it is 0.021.