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 (*numberofaction*), the most recent time of action when a specific user listens to a specific song (*timeofaction*), 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 (*frequency*).
- 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.

Figure 1: ‘User’ node, ‘Song’ nodes and ‘Listen’ relationships

Algorithm (with scenario):

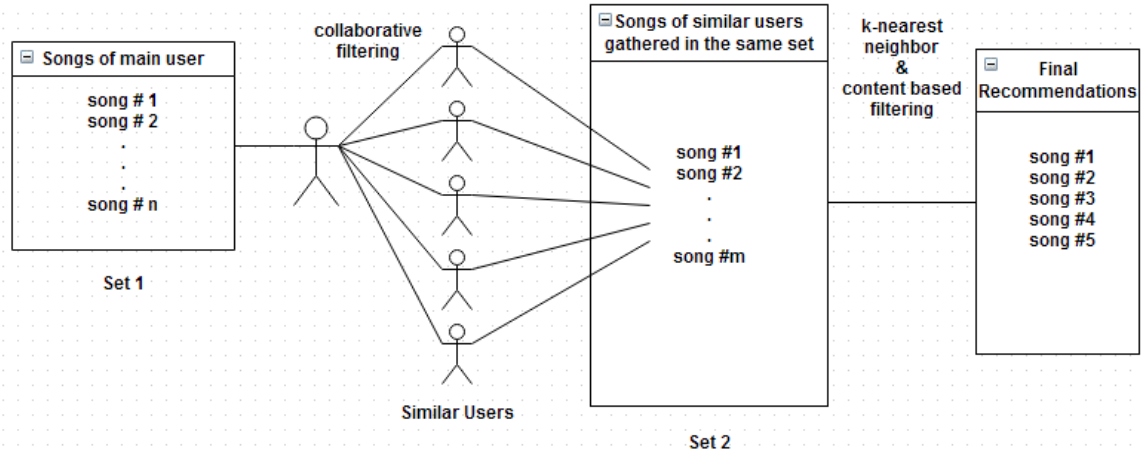


Figure 2: Recommendation algorithm

- User 2484378 is given to the *Recommend* function in order to get recommendations.
- The algorithm first finds five similar users to User 2484378. The similar users must have listened common songs with User 2484378 and rating value for these common songs must be above 0.6. Besides, the frequencies of 'LISTEN' relationships of the similar users and User 2484378 must be close to each other. The users that have listened the common songs only once are discarded from this set.
- Similar users are User 1667498, User 1660823, User 1608170, User 1612958, and User 1608461.
- All songs of the similar users, denoted by Set 2 (Figure 2), are compared to the five most frequently played songs of User 2484378. Weka K-nearest neighbor is used to find similarity over frequency and rating values during comparison. 30 songs in Set 2 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 with 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?

Cold Start Problem:

- Give recommendations to users for whom the system has very little information concerning the listening habits (i.e. newbie users)
- The algorithm first finds the most frequent genre among the songs of the cold start user.
- In the next step, the system selects the most frequently played five songs that belong to the most frequent genre found in the previous step.
- Cold start case is tested with User 2642298 (Figure 3).

Recommendation List - User ID: 2642298

#	Song Name	Performer Name	Album Name
1	A?k Nereden Nereye	Gripin	Yaln?z?l???n Çaresini Bulmu?lar
2	Seviyorsan ?nan?yorsan	Duman	darmaduman
3	Yürek	Duman	darmaduman
4	Eyvallah	Duman	darmaduman
5	Deli	Duman	darmaduman

Figure 3: Cold start recommendation

EVALUATION

DATA SET

User logs of 20, 30, 60 days are processed in Sound Tree to test with different sized databases.

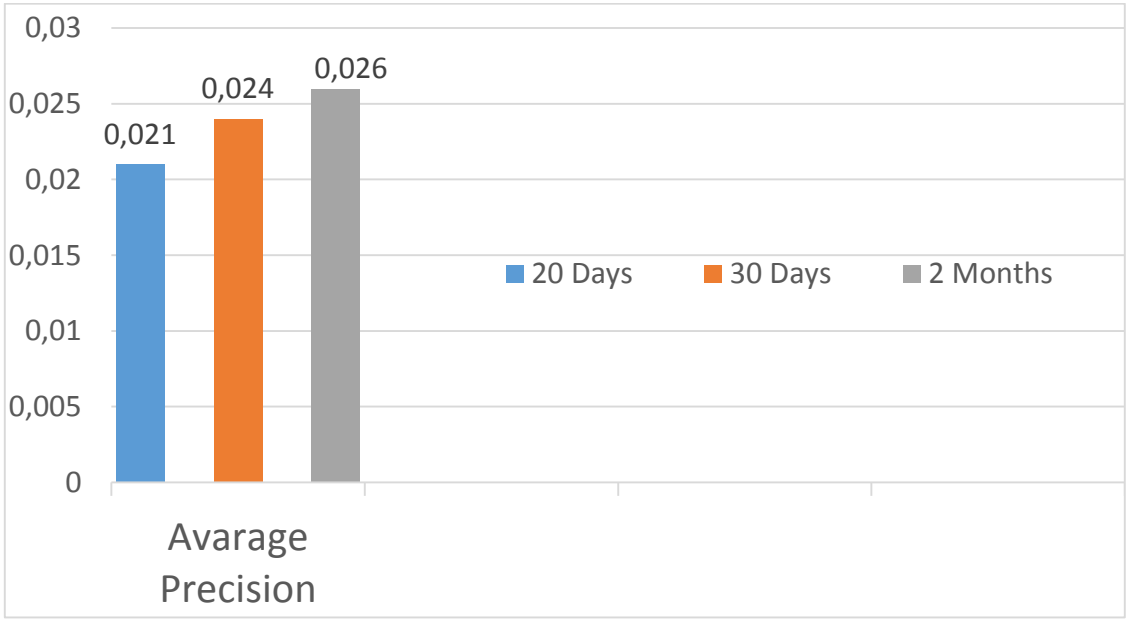
	Number of users	Number of songs	Average number of songs per user
Training Dataset (80 % of the dataset)	158542 176650 251378	1240034 1240034 1240034	22.83 30.68 55.41
Test Dataset (20 % of the dataset)	71242 82173 112967	1240034 1240034 1240034	14.12 18.78 26.56

Green: 20 days **Blue:** 30 days **Red:** 2 months

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



- Sound Tree evaluations are completed with databases of 20 days, 30 days and 2 months user logs.
- The precision result of each user in the test and training dataset is averaged in order to find an overall precision.