MLG:
Machine Learning with Graphs (*Strojno učenje na grafih*)

Assignment: Project proposal

Submission time: 20:17 **and date:** 7. February 2023

Submission Fill in and include this cover sheet with each of your assignments. It is an honor code violation to write down the wrong date and/or time. Assignments are due at 9:00am and should be submitted through Gradescope and eUcilnica. Students should check Piazza for submission details.

Late Periods Each student will have a total of *two* free late periods. *One late period expires the morning on the day before the next class.* (Assignments are usually due on Fridays, which means the first late period expires on the following Tuesday at 9:00am.) Once these late periods are exhausted, any assignments turned in late will be penalized 50% per late period. However, no assignment will be accepted more than *one* late period after its due date.

Honor Code We strongly encourage students to form study groups. Students may discuss and work on assignments in groups. However, each student must write down their solutions independently, i.e., each student must understand the solution well enough in order to reconstruct it by him/herself. Students should clearly mention the names of all the other students who were part of their study group. Using code or solutions obtained from the web (GitHub, Google, previous years etc.) is considered an honor code violation. We check all submissions for plagiarism. We take the honor code very seriously and expect students to do the same.

Name: Yon Ploj

Email: yp8702@student.uni-lj.si **SID:** 63200025

Study group: Team Kul

I acknowledge and accept the Honor Code.

(Signed) 

Playlist creation with GNNs

Alič Gregor **Drusany Jakob** **Piciga Aleksander** **Ploj Yon**

1 Introduction

Music streaming services have become increasingly popular over the past decade, with millions of users streaming music every day. However, one challenge faced by streaming services is creating dynamic playlists that are tailored to the users' preferences. In this project, we will explore how graph neural networks (GNNs) can be used to generate playlists that are tailored to a user's preferences.

2 Dataset

For this project, we will be using Spotify's Million Song Dataset, an open-source dataset containing information about over one million playlists. The MSD contains various metadata about the songs, such as artist name, album title, and song duration. The dataset contains around 1 million playlists, 2 million songs and 300 thousand artists. Since the dataset is way too large for us to process entirely, we are going to sample a subgraph by taking only the most popular songs and playlists (according to the highest node degree).

Main features of the dataset:

- Main Song features:
  - Song name
  - Album name
- Main Artist features:
  - ID
- Main Playlist features:
  - Name
  - Date of Modification
  - Whether it is collaborative
  - Followers
  - Duration

With most features we also get coresponding IDs which can be used for further data extraction via the Spotifies API.

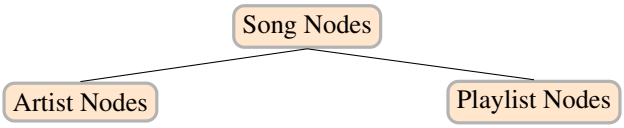
Main relationship features:

- Song to Playlist:
  - Position

3 Graph Structure

We will use this dataset to construct a heterogenous graph, with nodes representing songs, artists and playlists, and edges representing authorships/playlist inclusions. We will then use a GNN to generate playlists for each user based on their past recorded preferences - which can be viewed as a virtual playlist.

Base tripartite graph structure:



4 Graph ML Model

We will use GraphSAGE, a graph neural network architecture for semi-supervised node classification, for our GNN model. GraphSAGE is a neighbourhood-based graph neural network, where each node's representation is aggregated from its neighbours. This model is appropriate for our task because it can capture the similarities between songs and generate playlists that are tailored to a user's preferences.

Model will wrok by accepting n selected songs and generating additional k songs.