# Genre Prediction and Character Community Detection in Movie Networks

## Team Members

First team member is the leader of the team.

1. Yucong Mo – A#02380278 (Grad Student)
2. Yingying,Deng – A# 02394960 (Grad Student)
3. Autumn,Cuellar – A#02393938 (Grad Student)

## TYPE

Implementation and Detection

## Description of the Project

In many good movies, the interactions or conflicts between characters never fail to fascinate the audiences. One interesting way to study these relations in a non-literary criticism manner is to address them as social networks mathematically. Each node in a movie network represents a character and the edges between the nodes mean these characters appear in the same scene. Besides just scene relationships, the networks have pre-defined communities between characters. Exploring such networks can help discover the principal features of a movie. Based on the networks and principal features, we are to train classifiers to predict a movie's genre, identify the social roles that the characters perform, and detect the character community.

As the project is still at the early stage, we are looking into applying different classification models such as Support Vector Machines (SVM) and Gaussian Mixture model etc., hoping to identify the dominant character(s)' properties to conform the genre of a movie. As for community detection, we will be looking at K-means and other algorithms to adapt a new algorithm.

## Relevant Papers

Zhou, Hermans, Karandikar, and Rehg's paper [2] presents the results of genre prediction based on the movie's trailer. The authors believe that the trailers should capture enough characteristics of a movie for deciding its genre. We agree with this insight, but if our algorithm turns out to be successful, we will only require a movie interaction graph, which can be easily produced from a script, instead of processing and storage of video files proposed in Zhou's paper.

Makis and Vikatos's paper [1] presents a community detection algorithm for characters in a movie. Similar to the dataset we are using, the authors constructed a social network from screenplays and added weights for scene sharing frequency. Their detection algorithm was a pre-existing modularity algorithm that was suitable for the network. This paper is a good starting point for this project. We can determine which features to use and get some insight on what makes a good detection algorithm. The paper's references may be helpful too.

## Dataset

This project will use the MovieGalaxies Database. It provides network graph data from about 773 films (1915–2012). The data includes individual network graph data in Graph Exchange XML Format and descriptive statistics on measures such as clustering coefficient, degree, density, diameter, modularity, average path length, the total number of edges, and the total number of nodes.

The data is available here: https://moviegalaxies.com/

## Expected Outcomes/Results

Using only the networks formed by character interactions, we are looking for an efficient classification model for genre prediction and an algorithm to detect characters' community effectively in each movie.

## References

[1] Makris, C. & Vikatos, P. (2016). Community Detection of Screenplay Characters. 475. 463-470. 10.1007/978-3-319-44944-9_40.

[2] Zhou, H., Hermans, T., Karandikar, A. V., & Rehg, J. M. (2010, October). Movie genre classification via scene categorization. In Proceedings of the 18th ACM international conference on Multimedia (pp. 747-750). ACM. Chicago