Module 3 of E0 259, Data Analytics, August 2019

Community Detection

Lectures

To be upated [Lecture 1-3](#) (Rajesh Sundaresan)

Data set

[Dolphins](#) with due thanks to Lusseau et al. (2003) for compiling the data, and to Mark Newman for creating the file.
Unzip the .zip file to see two files, dolphins.txt (description) and dolphins.gml which contains an undirected social network of frequent associations between 62 dolphins in a community off Doubtful Sound, New Zealand. You can use the NetworkX package in Python to read the file or any other similar tool.

[Two circles](#) with due thanks to Nihesh Rathod for generating the data.
The .csv file contains the x and y coordinates of several points.

Assignment 3

Due: 23:55 hrs, Monday 07 October 2019. Discussion is encouraged. But write your own code. Please comply with the ethics policy.

1. Identify the two communities in the dolphins data set by obtaining the Fiedler vector of the normalised Laplacian. Provide a visualisation of the outcome so that the names of the dolphins, the network, and the two communities are clearly visible. Indicate the approach you have taken in light of the fact that the two communities are of different sizes.

2. Use the Louvain method to identify the two communities. Provide a picture similar to the one you generated in question 1.

3. Follow the method outlined in slide 31 to generate the "adjacency" matrix. Choose the top three eigenvectors and use the k-means algorithm on the rows to arrive at your two clusters. Provide a picture of your outcome to demonstrate that the circles have been captured.

You can use existing Python packages (for eigenvectors, k-means, etc., but the spectral clustering algorithm implementation should be yours).