CSIC 5011

Transition Paths of Karate Club Network

Xavier Plantaz, Department of Computer Science, The Hong Kong University of Science and Technology, xpfplantaz@connect.ust.hk

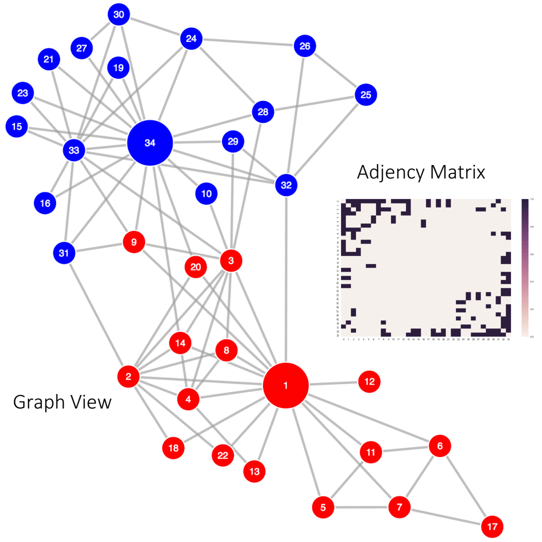
Abstract

When it comes to dimension reduction, we are facing a very complicated problem. The way the data is shaped makes the tools we used in regular geometric analysis obsolete. In this study, we focus on graph analysis using spectral clustering and transition path analysis using Markov chains. Visualizations are used along the way for sanity check and to make it more readable by the reader.

Dataset

The Dataset used for this study is the karate network, what is great with this dataset is that while it was recorded by a sociologist about the links between members in a college karate club outside of practice hours, the club splitted in two, about one half following the coach and the other half the president.

The code to clean and visualize the dataset can be found at https://github.com/xpfio/CSIC5011/blob/master/Miniprojects/Project%202%20-%20Cleaning.ipynb



Graph View

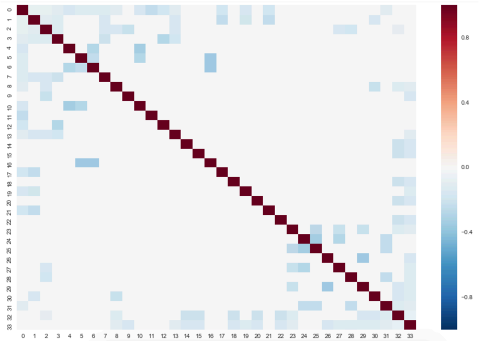
Adjency Matrix

Spectral Clustering

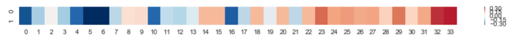
The code for Spectral Clustering can be found at https://github.com/xpfio/CSIC5011/blob/master/Miniprojects/Project%202%20-%20Spectral%20Clustering.ipynb

Normalized Laplacian Matrix

We focus on the normalized laplacian matrix as the second smallest eigenvalues will give us our **Fiedler Vector**.

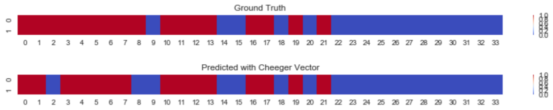


Normalized Laplacian Matrix

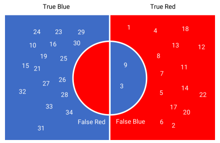


Fielder Vector

According to the sign, it give us the grouping:

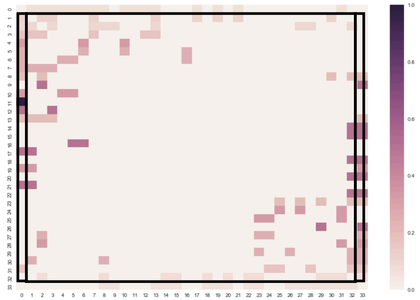


Which can also be visualized with the picture on the right.
The notebook contains extended tests and the use of scipy for comparision.



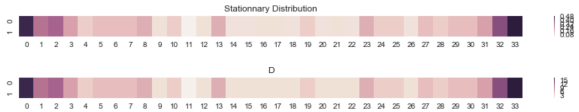
Transition Path Analysis

We compute the Markov chain transition matrix associated with the graph: https://github.com/xpfio/CSIC5011/blob/master/Miniprojects/Project%202%20-%20Transition%20path%20analysis.ipynb

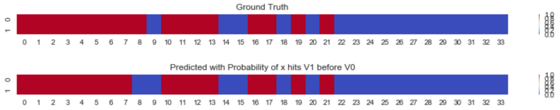


Transition matrix

And compute its stationary distribution associated with the eigenvalue 1. This stationary distribution is equivalent to the vector D, the invert of the sum of of the rows of the adjency matrix because the probability is given by $P_u^{(t)}(v) \xrightarrow{t\to\infty} \frac{d(v)}{2|E|}$.



For transition analysis, we set two points: 1 and 34, the coach and the president and compute the value of q(x), the probability of a member to hit 1 before 34, walking randomly on the graph. To do so, we extract the submatrix Q = A[1:33][1:33] and the vectors S, representing the transition from 1 or 34 to the rest of the graph, as represented on the previous figure, then q(x) = $(I-Q)^{-1}S(x)$. *Cutting the graph with probability higher or slower than 0.5, we get the cut* following and the graph on the right.



Effective flux

We compute effective flux associated to the previous probability and create a directed graph associated with the result.

Result