pectral Clustering and Transition Paths Analysis of Karate Club Network

Wei Cheng

wchengad@connect.ust.hk

May 20, 2019

Introduction

Networks have widespread its application for describing structures in a variety of fields, including social sciences, biology, economics and engineering.

- graph : degrees, paths, connectivities, etc.
- geometry structures in manifold learning theory
- topological structures in persistent homology

General problems on random walks of finite Markov chains on graphs

- How to find the optimal bi-partition or clustering on the graph?
- How to analyze the relative importance of each node on bridging the clusters?

Overview

1. Introduction
2. Methodology
   - Fiedler Theory and Spectral Clustering
   - Stationary Distribution of Markov Chain
   - Committor Function
   - Reactive Current and Effective Current
   - Transition Current
3. Evaluation on Karate Club network
4. Evaluation on More Networks
   - The Social Network of Les Misrables
   - LAO Protein Binding Transition Network

Fiedler Theory and Spectral Clustering

Let $G = (V, E)$ be an undirected, unweighted simple 1 graph. Then the graph Laplacian L and normalized graph Laplacian $\mathcal{L}$ can be defined as

$$L = D - A \qquad \mathcal{L} = D^{-1/2}(D - A)D^{-1/2} \tag{1}$$

Fiedler theory

$$\mathcal{L}v_i = \lambda_i v_i,\ \ v_i \neq 0,\ \ i = 0, \ldots, n-1 \tag{2}$$

For the second smallest eigenvector v_1, or call Cheeger vector

$$\begin{aligned} V_0 &= \{i : v_1(i) < 0\}, \\ V_1 &= \{i : v_1(i) > 0\}, \\ V_u &= V - V_0 - V_1. \end{aligned} \tag{3}$$

- $\#\{i, \lambda_i = 0\} = \#\{\textit{connected components of } G\}$;
- If G is connected, then both V_0 and V_1 are connected. $V_0 \cup V_u$ and $V_1 \cup V_u$ might be disconnected if $V_u \neq \varnothing$.

Stationary Distribution of Markov Chain

Given a graph $G = (V, E)$ and the row-Markov matrix $P = D^{-1}A$, the stationary distribution ^{T} is

$$\pi^T P = \pi^T \tag{4}$$

such π is invariant/equilibrium distribution.

- If P is primitive, then the largest eigenvalue λ with $|\lambda| = 1$ is unique w.r.t

$$\lim_{t\to\infty} \pi_0^T P^k = \pi^T, \ \forall \pi_0 \geq 0, \ 1^T \pi_0 = 1 \tag{5}$$

  This means when we take powers of P , i.e.P^k, all rows of P^k will converge to the stationary distribution π^T . Such a convergence only holds when P is primitive.

- If P is irreducible, then π is unique.

Committor Function

The reactive trajectories are those part of the equilibrium trajectory that the system is going from V_0 to V_1 . Let the hitting time of V_l be

$$\tau_i^k = inf\{t \geq 0 : x(0) = i, x(t) \in V_k\}, k = 0, 1. \tag{6}$$

The defined *committor* function satisfies the following Laplacian equation with Dirichlet boundary conditions

$$\begin{gathered}(Lq)(i) = [(IP)q](i) = 0, i \in V_u \\ q_{i \in V_0} = 0,\ q_{i \in V_1} = 1.\end{gathered} \tag{7}$$

which is in the following formulation

$$q_i = Prob(\tau_i^1 < \tau_i^0) = \begin{cases} 1, & x_i \in V_1 \\ 0, & x_i \in V_0 \\ \sum_{j \in V} P_{ij} q_j, & i \in V_u \end{cases} \tag{8}$$

Reactive Current and Effective Current

For any stationary trajectory, we call each proportion of the trajectory from A to B a $AB-$ *reactive* trajectory. We also define the reactive current from A to B of a directed edge ij as

$$J(ij) = \begin{cases} \pi(i)[1-q(i)]P_{ij}q(j), & i \neq j \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

The reactive current gives a clue for us to measure the importance of an edge and a node among all $AB-$ *reactive* trajectories. In particular, we define the effective current of an edge ij as

$$J^{+}(ij) = \max\{J(ij) - J(ji), 0\}. \tag{10}$$

Transition Current

For a node $i \in V$, we define the transition current through i as

$$T(i) = \begin{cases} \sum_{j \in V} J^+ ij, & x_i \in A \\ \sum_{j \in V} J^+ ji, & x_i \in B \\ \sum_{j \in V} J^+ ij = \sum_{j \in V} J^+ ji, & i \in V - A - B \end{cases} \tag{11}$$

Obviously, a node with high transition current through it plays a key role in the transition from A to B. Similarly, we can adopt this approach to identify the key nodes who bridge two communities of nodes. For example, in the thresholding scheme V_0 and V_1 we mentioned previously, the transition function

$$T(i) = \sum_{j \in V} J^+(ij), \ \ i \in V_0 \tag{12}$$

measures the contribution of every node in V_0 in the connection of V_0 and V_1 .

Karate Club network

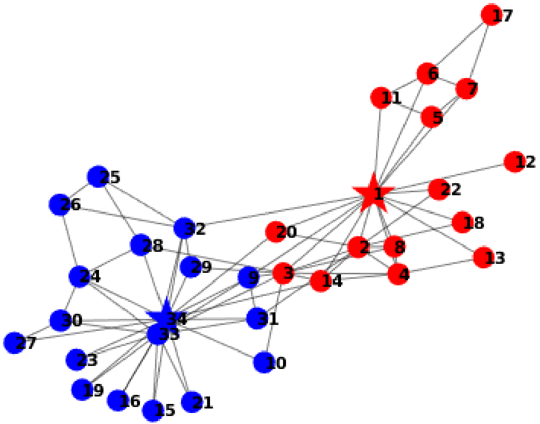
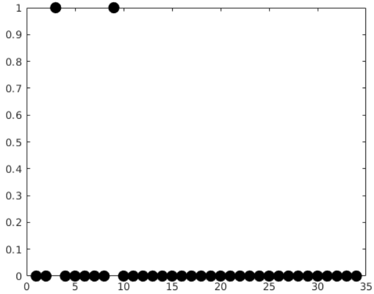
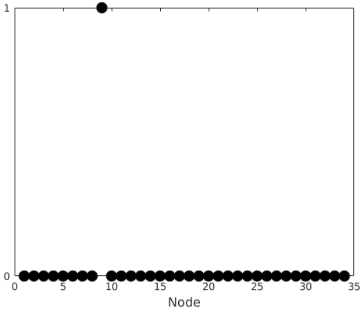


Figure: The Zachary's Karate Club network. Node highlighted with stars are $A = \{1\}, B = \{34\}$ are local minimal, nodes are distinguished by red and blue color into two clusters.

Spectral Clustering vs. Committor Fuction



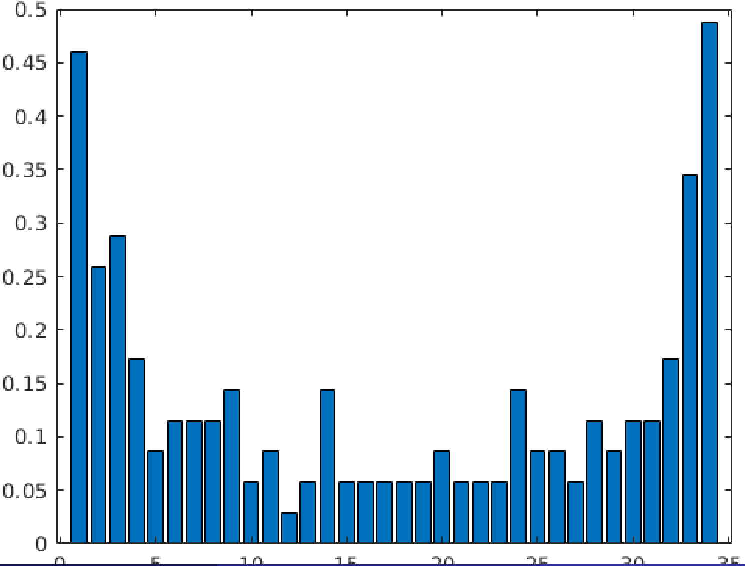
(a) Difference between spectral clustering with true fission.



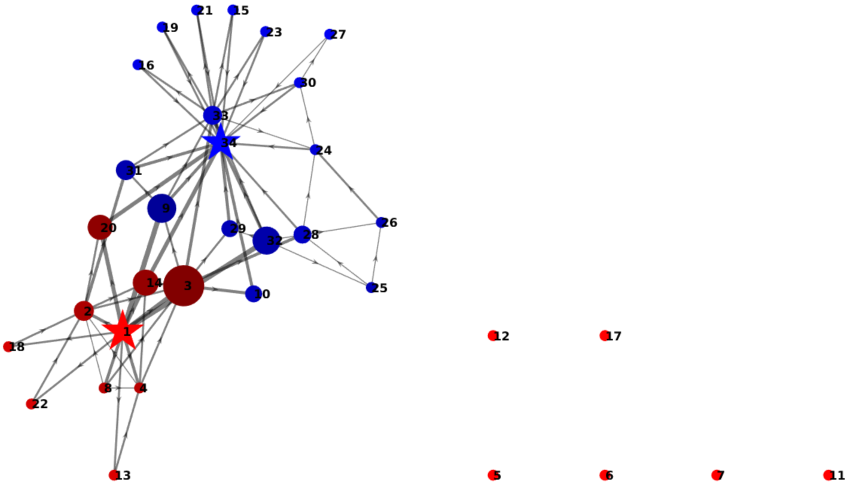
(b) Difference between *committor* function clustering with true fission.

Figure: Clustering results compared to true fission in spectral clustering and path transition analysis.

Markov Chain and Stationary Distribution



Effective Current and Transition Current



The Social Network of Les Misrables

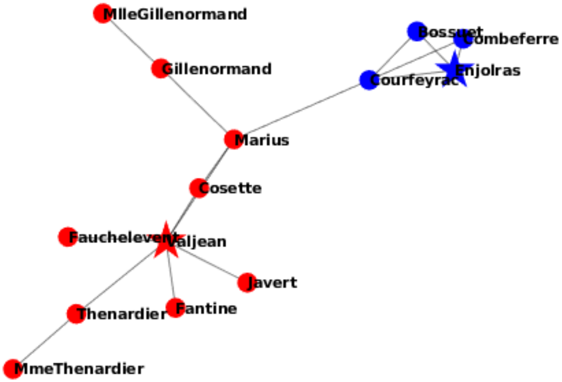


Figure: The subnetwork of the Les Misrables social network. Two local minima, Enjoras and Valjean are highlighted with star markers, and the node color is coded by the same clustering result on two different clustering methods.

The Social Network of Les Misrables

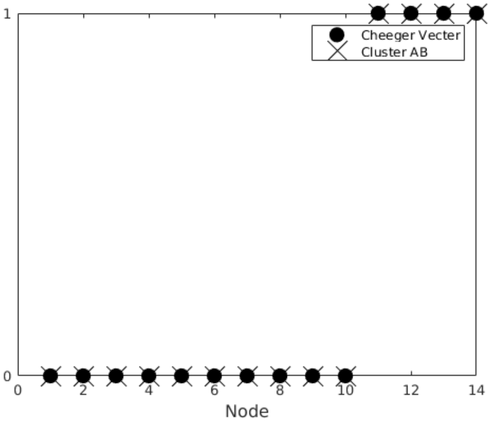
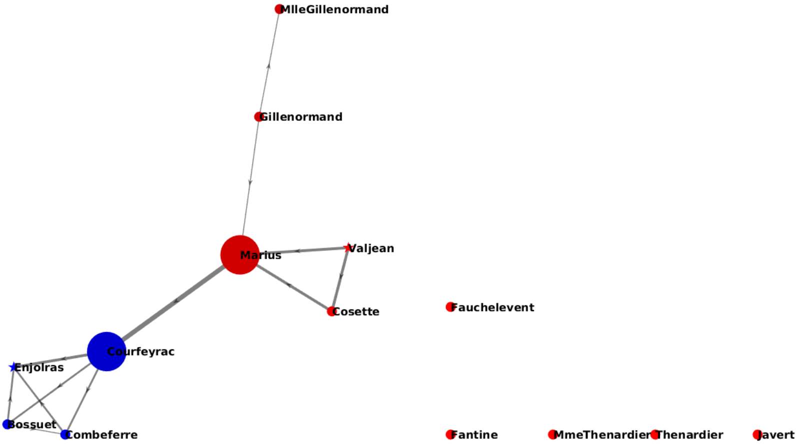


Figure: Clustering results using spectral clustering and *committor* function clustering.

The Social Network of Les Misrables



LAO Protein Binding Transition Network

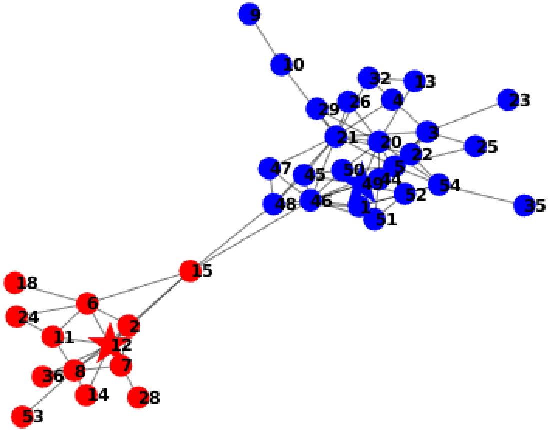


Figure: Undirected graph visulization of LAO network, two local minimas are found are thresholding and dropping isolated states. The transition states are clustered into two clustering using spectral clustering and *committor* function.

LAO Protein Binding Transition Network

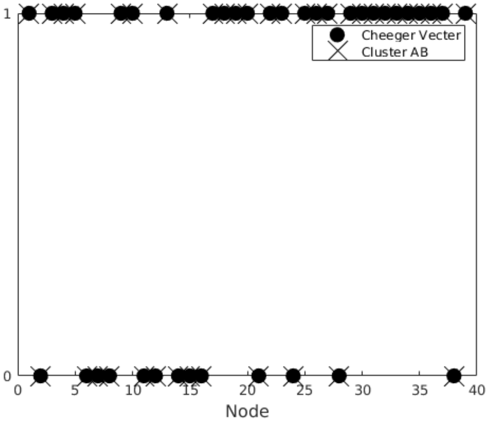
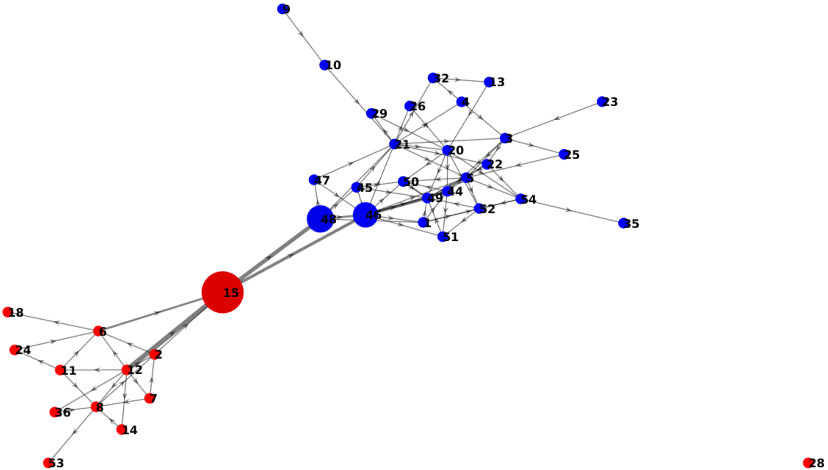


Figure: Clustering results by applying spectral clustering and *committor* function on the preprocessed graph.

LAO Protein Binding Transition Network



Thanks for watching