Elements of Spectral Graph Theory I

An Introduction to Isoperimetry

Nick Ryder

UC Berkeley Mathematics

March 27th, 2018

Berkeley
UNIVERSITY OF CALIFORNIA

Introduction

- Connections between spectral theory of matrices and graphs
- Give natural continuous notions of discrete properties of graphs
- These slides closely follow lecture notes of Dan Spielman

Clustering

Many times people want to study balanced cuts:

- Minimal cuts which split the graph into two near equal parts
- Hard to develop objective for

Instead we look at Isoperimetry:

- Cuts which have small size relative to the smaller cluster
- Objective easy to define for all subsets of graph

Isoperimetry

To formalize this we define the boundary

$$\partial(S) := \{(u, v) \in E \mid u \in S, v \notin S\}$$

and the isoperimetric ratio

$$\theta(S) := \frac{|\partial(S)|}{|S|}$$

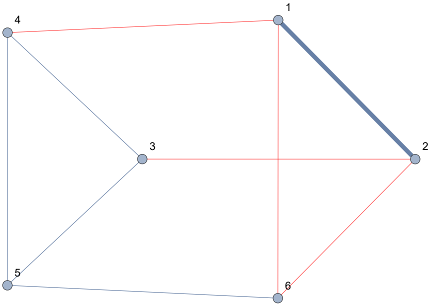
Then for a graph we define the isoperimetric number

$$\theta_G := \min_{|S| \leq \frac{n}{2}} \theta(S)$$

Berkeley
UNIVERSITY OF CALIFORNIA

Example for $\partial(S)$

$$\partial(S) := \{(u, v) \in E \mid u \in S, v \notin S\}$$

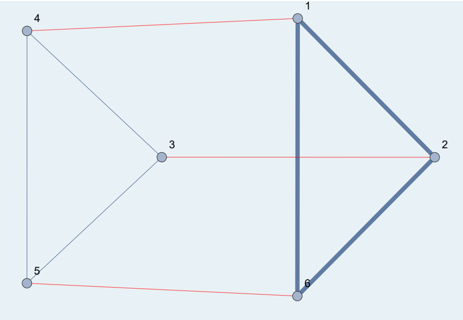


$$\theta(S) = \frac{4}{1} = 4$$

Berkeley
UNIVERSITY OF CALIFORNIA

Example for θ_G

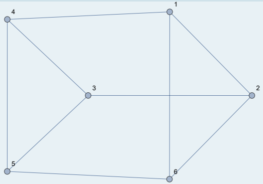
For this graph θ_G is achieved here:



$$\theta(S) = \frac{3}{3} = 1$$

Laplacian

Let G be a $d-$regular graph. The main objects we study are the adjacency matrix A_G and the Laplacian $L_G = dI - A_G$

G	A_G	L_G
	$\begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 3 & -1 & 0 & -1 & 0 & -1 \\ -1 & 3 & -1 & 0 & 0 & -1 \\ 0 & -1 & 3 & -1 & -1 & 0 \\ -1 & 0 & -1 & 3 & -1 & 0 \\ 0 & 0 & -1 & -1 & 3 & -1 \\ -1 & -1 & 0 & 0 & -1 & 3 \end{pmatrix}$

Laplacian as a Quadratic Form

We consider L_G as a quadratic form to obtain

$$x^T L_G x = \sum_{\{u,v\}=e\in G} (x(u) - x(v))^2$$

Properties of L_G:

- L_G is positive semidefinite
- $Null(L_G)$ is spanned by 1_S where $S \subset G$ is a connected component of G
- The multiplicity of the eigenvalue 0 is the number of connected components of G.
- For a connected graph, λ_2, the second smallest eigenvalue is the quantity of interest

Relating Isoperimetry to the Spectrum

We consider Rayleigh quotients of $\mathbf{y} \perp \mathbf{1}$ which by Courant Fischer bound λ_2

$$\mathcal{R}_G(\mathbf{y}) := \frac{\mathbf{y}^T L_G \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \qquad \min_{\mathbf{y} \perp \mathbf{1}} \mathcal{R}_G(\mathbf{y}) = \lambda_2$$

Given a indicator vector for a subset and projecting away from $\mathbf{1}$ we get

$$\mathcal{R}_G\left(\mathbf{1}_S - \frac{|S|}{|V|}\mathbf{1}\right) = \frac{|\partial(S)|}{|S|(1 - \frac{|S|}{|V|})} = \left(1 - \frac{|S|}{|V|}\right)^{-1} \theta(S)$$

Remark

This immediately gives the lower bound $\theta_G \geq \lambda_2/2$

Relaxation of Clustering Problem

We search for a relaxation of our problem using the relation between $\theta(S)$ and $\mathcal{R}_G(\mathbf{1}_S - \frac{|S|}{|V|}\mathbf{1})$

$$\begin{array}{c} \textbf{Minimize} \quad \theta(S) \\ S \subset V, |S| \leq |V|/2 \end{array} \xrightarrow{\text{Relaxation}} \begin{array}{c} \textbf{Minimize} \quad \mathcal{R}_G(\mathbf{y}) \\ \mathbf{y} \perp \mathbf{1} \end{array}$$

In order for this relaxation to succeed, we need a method of rounding vectors with small Rayleigh quotients to subsets with small isoperimetry.

Cheeger's Inequality

Theorem

Given any $\mathbf{y} \perp \mathbf{1}$, we have t such that $S_t = \{v \mid \mathbf{y}(u) \leq t\}$ satisfies

$$\theta(S_t) \leq \sqrt{2d\mathcal{R}_G(\mathbf{y})}$$

This implies

$$\theta_G \leq \sqrt{2d\lambda_2}$$

Remark

For irregular graphs G, one needs to use the conductance and normalized Laplacian in the statement of Cheeger's inequalities. In the case of regular graphs, they are related by a constant multiple to the isoperimetry and Laplacian.

Berkeley
UNIVERSITY OF CALIFORNIA

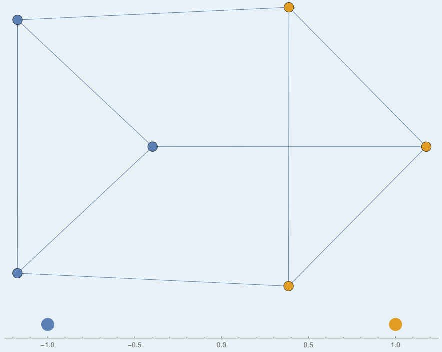
Algorithm for Clustering Embedding

To find a set S with low isoperimetry:

- Find vector $\mathbf{y}$ such that $\mathcal{R}_G(y)$ is small
- Hence $\mathbf{y}$ eigenvector for second smallest eigenvalue
- Sort entries of $\mathbf{y}$ and compute θ for all possible $S_t = \{v \mid y(u) \leq t\}$
- Guaranteed to find a set $S_t \subset V$ with $\theta(S_t) \leq \sqrt{2d\lambda_2}$

Can imagine this as embedding the graph G into $\mathbf{R}$ by using the second eigenvalue coordinates, then picking an appropriate threshold to split the vertices.

Example of Clustering



$$\begin{pmatrix} 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{pmatrix}$$

Summary

- Natural quantities associated to graphs can be related to the spectrum of the Laplacian
- Cheeger's Inequality gives an algorithm for clustering the graph with respect to Isoperimetry
- There exists a generalized Cheeger's Inequality for *k*-clustering