Resumen

Aquí va el resumen...

ABSTRACT

Aquí inicia el abstract...

AGRADECIMIENTOS

Aquí van los agradecimientos...

Contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

El capítulo 1 inicia aquí...

1.1 Presentation

1.2 Objectives

1.2.1 General objective

1.2.2 Particular objective

1.3 Justification

1.4 Limitations and delimitations of the project

1.5 Research Problem

1.6 Hypothesis

1.7 Project organization

CHAPTER 2

THEORY AND CONCEPTUAL FRAMEWORK

2.1 Preliminaries

Description of concepts

Definition 2.1.1. *orthogonal complement*

Theorem 2.1.1. *For every $n \times n$ symmetric real matrix, the eigenvalues are real and the eigenvectors can be chosen real and orthonormal.*

Theorem 2.1.2 (Courant-Fisher Formula)**.** *Let A be an $n \times n$ real symmetric matrix with eigenvalues $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ and corresponding eigenvectors $v_1, v_2, \ldots, v_n$. Then*

$$\lambda_1 = \min_{\|x\|=1} x^T A x = \min_{x \neq 0} \frac{x^T A x}{x^T x},$$

$$\lambda_2 = \min_{\substack{\|x\|=1 \\ x \perp v_1}} x^T A x = \min_{x \neq 0} \frac{x^T A x}{x^T x},$$

$$\lambda_n = \lambda_{max} = \max_{\substack{\|x\|=1 \\ x \perp v_1}} x^T A x = \max_{\substack{x \neq 0 \\ x \perp v_1}} \frac{x^T A x}{x^T x}.$$

In general, for $1 \le k \le n$, let S_k denote the span of $v_1, v_2, \ldots, v_k$ (with $S_0 = \{0\}$). Then

$$\lambda_k = \min_{\substack{\|x\|=1 \\ x \in S_{k-1}^{\perp}}} x^T A x = \min_{\substack{x \neq 0 \\ x \in S_{k-1}^{\perp}}} \frac{x^T A x}{x^T x}.$$

Proof. Let $A = Q\Lambda Q^T$ be the eigenvalue decomposition of A, where Q is an orthogonal matrix whose columns are eigenvectors of A, and Λ is a diagonal matrix whose entries are the eigenvalues of A. ■

2.2 Graphs and Laplacian Matrix

For the rest of the chapter, let $G = (V,E)$ be an undirected graph, where $V = \{v_1, v_2, ..., v_n\}$ is the non-empty set of nodes (or vertices) and E is the set of edges, composed by pairs of the form (v_i, v_j), where $v_i, v_j \in V$. Let $w : E \to R_{\geq 0}$ be a weight function and define $w_{ij} = w(v_i, v_j)$, for $1 \leq i, j \leq n$, with $w_{ij} = 0$ if there is not an edge connecting the nodes v_i and v_j.

The *weighted adjacency matrix* of the graph is the matrix defined by $W = [w_{ij}]_{n\times n}$

The *degree of a vertex* $v_i \in V$ is defined as

$$d_i = \sum_{j=1}^{n} w_{ij}.$$

The *degree matrix* D is defined as the diagonal matrix with the degrees $d_1, d_2, ..., d_n$ on the diagonal.

The unnormalized graph *Laplacian matrix* L is defined as

$$L = D - W$$

The normalized Laplacian matrix L_{sym} is defined as

$$L_{sym} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$$

Proposition 2.2.1 (Some properties of L)**.** *The matrix L, as defined above, satisfies the following properties:*

1. *For every vector $x = (x_1, x_2, ..., x_n) \in \mathbb{R}$ we have*

$$x^T L x = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \left(x_i - x_j\right)^2$$

2. *L is symmetric and positive semi-definite*

3. *L has n non-negative, real-valued, eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$*

4. *The smallest eigenvalue L is 0, the corresponding eigenvector is the constant one vector $\mathbb{1}$.*

Proof. Here is your proof ■

Definition 2.2.1. *Given a graph $G = (V, E)$, a partition of G is a collection of k subsets $P_1, P_2, ..., P_k \subset V$ such that:*

1. $P_i \cap P_j = \emptyset$ *for* $i \neq j$, *where* $i, j \in \{1, 2, ..., k\}$
2. $\cup_{i=1}^{k} P_k = V$

For a collection $S \subset V$ of vertices, we define the *edge boundary* $\partial(S)$ to consist of all edges in E with exactly one endpoint in S, that is,

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

For two collections of vertices $A, B \subset V$ consider the following quantities related to the edges of

$$W(A, B) := \sum_{v_i \in A} \sum_{v_j \in B} w_{ij}$$

and

$$\textsc{Vol}(A) := \sum_{v_i \in A} d_i$$

In order to measure the quality of the partition we introduce the following we want to agroupe by similarity so its natural to Cut value of that partition

$$\textsc{Cut}(P_1, P_2, ..., P_k) := \frac{1}{2} \sum_{i=1}^{k} W(P_i, \overline{P_i})$$

solve the mincut problem

The next consider two different ways of measuring the size of the partitions

$$\begin{aligned}\textsc{RatioCut}(P_1, P_2, ..., P_k) &:= \frac{1}{2} \sum_{i=1}^{k} \frac{W(P_i, \overline{P_i})}{|P_i|} \\ &= \sum_{i=1}^{k} \frac{\textsc{Cut}(P_i, \overline{P_i})}{|P_i|}\end{aligned}$$

$$\begin{aligned}\textsc{NormCut}(P_1, P_2, ..., P_k) &:= \frac{1}{2} \sum_{i=1}^{k} \frac{W(P_i, \overline{P_i})}{\textsc{Vol}(P_i)} \\ &= \sum_{i=1}^{k} \frac{\textsc{Cut}(P_i, \overline{P_i})}{\textsc{Vol}(P_i)}\end{aligned}$$

The spectral method

1. Let v denote the second smallest eigenvector of $\mathscr{L}$. Sort the vertices i of G in increasing order of v_i. Let the resulting ordering be $v_1 \leq v_2 \leq \cdots v_n$
2. For each i, consider the cut induced by $\{1,2,...,i\}$ and its complement. Calculate its conductance.
3. Among these $n-1$ cuts, choose the one with minimum conductance.

Cheeger's inequality

For a graph $G = (V, E)$ the *conductance* or *Cheeger ratio* of a set $S \subset V$ is the ratio of the fraction of edges in the cut $(S, \overline{S})$ o the volume of S,

$$\phi(S) = \frac{E(S,\overline{S})}{\textsc{Vol}(S)}$$

The *conductance* or *Cheeger constant* of a graph G is denoted by

$$\phi(G) = \min_{S} \phi(S)$$

Theorem 2.2.1. *In a graph G, the Cheeger constant $\phi(G)$ and the spectral gap λ_G are related as follows:*

$$2\phi(G) \geq \lambda_G \geq \frac{\alpha_G^2}{2} \geq \frac{\phi(G)^2}{2}$$

where α_G^2 is the minimum Cheeger ratio of subsets S_i consisting of vertices with the largest i values in the eigenvector associated with $lambda_G$, over all $i \in [n]$

Generalization to many partitions

1. Perform eigenvalue decomposition to find the eigenvectors of L_{sym}.
2. Select the k largest eigenvectors $e_1, e_2, ..., e_k$ of L_{sym} associated to the largest eigenvalues $\lambda_1, \lambda_2, ..., \lambda_k$
3. Form the matrix Y from the matrix $X = [e_1, e_2, ..., e_k]$ given by
$$Y_{ij} = \frac{X_{ij}}{\left(\sum_j X_{ij}^2\right)^{\frac{1}{2}}}$$
4. Treating each row of Y as a point in $\mathbb{R}^k$, cluster them into k clusters using $K-means$
5. Finally, assign the original vertex to cluster j if and only if row i of the matrix was assigned to cluster j

2.2.1 The graph partitioning problem

2.2.2 Spectral partitioning and Normalized Cut

2.3 Literature review

2.3.1 Graph Convolutional Neural Networks and GraphSAGE

2.3.2 Generalizable Approximate Graph Partitioning (GAP) Framework

2.3.3 PinSAGE and Markov Chain Negative Sampling (MCNS)

CHAPTER 3

Proposed solution (Graph Partitioning for Large Graphs)

CHAPTER 4

EXPERIMENTAL RESULTS

CHAPTER 5

CONCLUSION

Las conclusiones y el trabajo a futuro inicia aquí...

5.1 Contributions

5.2 Recommendations and future work

APPENDIX A

ANALISIS

El apéndice inicia aquí.

BIBLIOGRAPHY

[1] Datos de la publicación.