



IMPERIAL COLLEGE LONDON

DEPARTMENT OF COMPUTING

Algorithms for Graph Partitioning

Author:
Shahrokh Shahi

Supervisor:
Dr. Mahdi Cheraghchi
Bashi Astaneh

Submitted in partial fulfillment of the requirements for the MSc degree in
Computing Science of Imperial College London

September 2016

Abstract

Graph partitioning is a fundamental algorithmic problem in combinatorics motivated by applications such as clustering and community detection in social networks as well as theoretical importance in particular in spectral methods and the Unique Games Conjecture in computational complexity.

The main theme in this problem is the following: Given a graph consisting of a collection of “loosely connected dense subgraphs”, design an efficient algorithm to detect, either exactly or approximately, the underlying dense subgraphs. There are many variations of the problem, ranging from information theoretic possibility of discovering the communities to efficient algorithms for doing so as well as “local walk” and distributed models for the algorithmic task.

In this research, a systematic study of the major techniques and discoveries in this area has been conducted with an emphasis on the methods based on the spectral graph theory. In the last years, spectral graph partitioning approaches have become very popular and there has been a growing interest in their applications, mainly on account of their efficiency and mathematical elegance. Therefore, the main concepts of the spectral partitioning algorithms are comprehensively discussed in this research and some novel applications of these methods have been concluded at the end.

Contents

1	Introduction	1
1.1	Motivation and Objectives	1
1.2	Graph Partitioning Applications	2
1.2.1	Parallel Processing	2
1.2.2	Complex Networks	2
1.2.3	Image Processing	3
1.2.4	VLSI Physical Design	4
1.2.5	Clustering	4
1.2.6	Community Detection	5
1.3	Report Outline	6
2	Preliminaries	7
2.1	Overview	7
2.2	A Review of Graph Theory	7
2.2.1	Basic Concepts	7
2.2.2	Subgraphs	8
2.2.3	Walk, Path, Trail, Circuit, and Circle	8
2.2.4	Connected Graphs	9
2.2.5	Specific Classes of Graphs	9
2.2.6	Graph Matrices	10
2.3	A Review of Linear Algebra	10
2.3.1	Basic Definitions	10
2.3.2	Eigenvalues and Eigenvectors	10
2.3.3	Orthogonality	11
2.3.4	Spectral Theorem for Real Symmetric Matrices	11

2.3.5	Graph Spectrum	12
2.3.6	Complete Graph	12
2.3.7	Bipartite Graph	12
2.3.8	Positive Semidefinite Matrix	13
2.3.9	Connectedness	13
3	Graph Partitioning Approaches	15
3.1	Overview	15
3.2	Graph Partitioning Problem	15
3.2.1	Definition	15
3.2.2	Problem Complexity	16
3.3	Global Algorithms	16
3.3.1	Exact Algorithms	16
3.3.2	Spectral Partitioning	17
3.3.3	Graph Growing	17
3.3.4	Flows	17
3.3.5	Geometric Partitioning	18
3.3.6	Stream Graph Partitioning	19
3.4	Improved Heuristic Methods	19
3.4.1	Node-swapping Local Search	20
3.4.2	Extension to k-way Local Search	21
3.4.3	Tabu Search	22
3.4.4	Flow Based Improvement	23
3.4.5	Bubble Framework	23
3.4.6	Random Walks and Diffusion	24
3.5	Multilevel Graph Partitioning	25
3.5.1	Coarsening Approaches	27
3.6	Summary	28

4	Spectral Graph Partitioning	29
4.1	Overview	29
4.2	Definitions	30
4.3	Scope	32
4.4	The Laplacian Matrix	32
4.5	Graph Partitioning as Constrained Quadratic Optimization	33
4.6	Bounds on the Weight of a Bisection	34
4.6.1	Rayleigh Quotient	34
4.7	Spectral Graph Partitioning	36
4.7.1	The Relaxed Optimization Problem	36
4.7.2	The Spectral Partitioning Algorithm without Vertex Masses	37
4.7.3	Vertex Masses	38
4.7.4	The Spectral Partitioning Algorithm with Vertex Masses	39
4.8	Unbalanced Cuts	39
4.8.1	Bounds on the Weight of an Unbalanced Cut	40
4.9	Cheeger's Inequality	40
4.9.1	Normalized Matrices	40
4.9.2	The Theorem (Cheeger's Inequality)	41
4.9.3	Notes on the Cheeger's Inequality	43
4.10	Maximum Cut Problem	44
4.10.1	Last Eigenvalue	44
4.10.2	Maximum Cut	44
4.10.3	Theorem (Trevisan)	45
4.10.4	Approximation Algorithm	45
4.11	More Eigenvalues	46
4.11.1	Small-set Expansion	46
4.11.2	Multi-Partitioning	47
4.12	k -way Partitioning	52
4.12.1	Vector Partitioning	52
4.12.2	k -Subgraph Partitions	54
4.13	Summary	55

5	Spectral Graph Clustering	57
5.1	Overview	57
5.2	Graph Cut and Problems	57
5.2.1	Minimum Cut Problem	57
5.2.2	Minimum Ratio Cut Problem	58
5.2.3	Minimum Normalized Cut Problem	58
5.2.4	Min-max Cut Problem	59
5.2.5	Modularity Maximization Problem	59
5.3	Spectral Clustering Algorithms	60
5.3.1	Two-way Partitioning Algorithms	60
5.3.2	k -way Partitioning Algorithms	61
5.4	Summary	64
6	Practical Application: Group Testing	65
6.1	Overview	65
6.2	Graph-Constrained Group Testing	65
6.2.1	Network Tomography	65
6.2.2	Problem Statement	66
6.3	Expander Graphs	67
6.4	Partitioning into Expanders	68
6.4.1	Definitions	68
6.4.2	Theorems	69
6.5	Extension to the Group Testing Scheme	69
6.6	Summary	71
7	Programming and Visualization	73
7.1	Overview	73
7.2	Available Packages	73
7.3	MATLAB Implementation	74
7.3.1	Generating Random Graphs	76
7.3.2	MATLAB Toolbox	78
7.4	Evaluation	79
7.4.1	Example1: Bisection Problem	79

8	Conclusions and Future Work	83
8.1	Summary of Achievements	83
8.2	Future Work	84
	Bibliography	85
	Appendices	94
A	Notations and Symbols	95
B	User Guide	97