

Parameterless clustering by dynamic tree-cutting

Simon Lehmann Knudsen
simkn15@student.sdu.dk
ECTS: 10
04/09-2017 - 31/01-2018
5th semester in Computer Science

31/01-2017

Contents

1	Abstract	3
2	Introduction	4
2.1	What is clustering?	4
2.2	The clustering problem	4
2.2.1	Determining parameters	4
3	Background	5
3.1	Transitivity Clustering	5
3.2	Hierarchical Clustering	5
3.3	Cluster Validation	5
3.4	GAP Statistics	5
3.5	Multidimensional Scaling	5
4	Method	6
4.1	Assessing best clustering	6
4.2	Randomization approach 1	6
4.3	Randomization approach 2	6
4.4	Randomization approach 3	6
4.5	Randomization approach 4	6
4.6	Results for randomizations	6
4.7	Short about the implementation / tools	6
5	Conclusion	7

1 Abstract

2 Introduction

2.1 What is clustering?

2.2 The clustering problem

2.2.1 Determining parameters

3 Background

3.1 Transitivity Clustering

3.2 Hierarchical Clustering

3.3 Cluster Validation

3.4 GAP Statistics

3.5 Multidimensional Scaling

4 Method

4.1 Assessing best clustering

4.2 Randomization approach 1

4.3 Randomization approach 2

4.4 Randomization approach 3

4.5 Randomization approach 4

4.6 Results for randomizations

4.7 Short about the implementation / tools

5 Conclusion