

UNIVERSIDADE FEDERAL DE MINAS GERAIS  
PROPOSTA DE PFC

AGRUPAMENTO ESPECTRAL PARA PROBLEMAS DE  
GRANDES BASES DE DADOS

ENGENHARIA DE CONTROLE E AUTOMAÇÃO

ALUNA: DANIELLE AGUIAR CARNEIRO SILVA

ORIENTADOR: PROF. ANTÔNIO DE PÁDUA BRAGA

14 DE DEZEMBRO DE 2015

# Sumário

<b>List of Figures</b>	<b>2</b>
1 Descrição do Problema . . . . .	1
1.1 Motivação . . . . .	1
1.2 Aplicações do método . . . . .	1
1.3 O projeto . . . . .	1
2 Objetivos do Projeto . . . . .	1
3 Etapas . . . . .	2
4 Cronograma . . . . .	2
5 Recursos Necessários . . . . .	2
6 Assinaturas . . . . .	3
<b>Bibliography</b>	<b>3</b>

# Lista de Figuras

1	Diagrama de Gantt do PFC . . . . .	2
---	------------------------------------	---

# 1 Descrição do Problema

O agrupamento, mais conhecido como *clustering*, é uma das mais vastas técnicas para análise de dados exploratórias. O objetivo do *clustering* é dividir uma base de dados em grupos de forma que os pontos de um mesmo grupo sejam similares.

## 1.1 Motivação

Através da percepção visual humana pode-se analisar uma base de dados e facilmente segmentá-la em grupos coerentes, identificando comportamentos semelhantes. Portanto, é desejável, obter uma performance semelhante através do processamento computacional: ser capaz de classificar automaticamente dados em diferentes grupos a partir da análise dos mesmos.

Nos últimos anos, o agrupamento espectral tornou-se um dos mais populares algoritmo de agrupamento. Esse método realiza o agrupamento de dados baseado em autovetores da matriz de afinidade.

Apesar do sucesso do agrupamento espectral, ainda existem alguns problemas à serem resolvidos. Além de existir uma grande variedade de algoritmos que utiliza os autovalores de várias maneiras diferentes, em muitos casos não há garantia que eles irão agrupar os dados de maneira satisfatória.

## 1.2 Aplicações do método

As aplicações da técnica de agrupamento espectral vão desde estatística, engenharia da computação, biologia à ciências sociais ou psicologia.

Dentro da engenharia, pode-se destacar uma enorme gama de aplicações em aprendizado de máquinas, visão computacional e processamento de sinais.

## 1.3 O projeto

O projeto consistirá na escolha de algoritmos de agrupamento espectral que apresentem a melhor performance para grandes bases de dados e na implementação dos mesmos.

Dentre as bases de dados a serem utilizadas, pode-se destacar imagens.

# 2 Objetivos do Projeto

O objetivo desse trabalho final de curso é apresentar, implementar e validar algoritmos de agrupamento espectral que apresentem uma boa performance em para grandes bases de dados.

### 3 Etapas

O projeto será dividido nas seguintes etapas:

1. Revisão bibliográfica;
2. Definição da base de dados;
3. Definição dos algoritmos a serem implementados;
4. Implementação dos algoritmos;
5. Testes e validação;
6. Entrega e apresentação do PFC.

O cronograma do projeto será apresentado na próxima sessão.

### 4 Cronograma

O projeto final de curso se estende por um ano. Portanto, é indispensável criar um planejamento, a partir de um diagrama de Gantt para garantir o cumprimento dos prazos estabelecidos. A Figura 1 mostra o diagrama de Gantt inicial do PFC. Esse diagrama será atualizado e detalhado ao longo de todo o desenvolvimento do projeto.

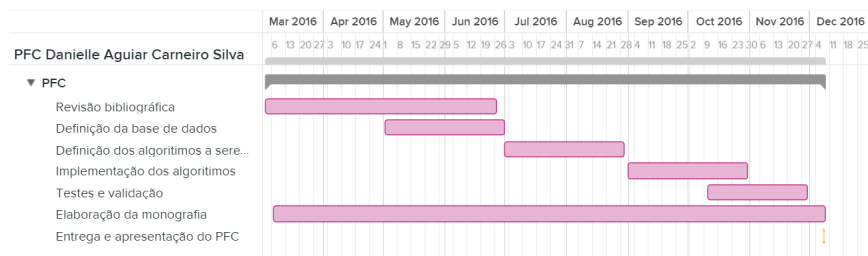


Figura 1: Diagrama de Gantt do PFC

### 5 Recursos Necessários

Para realização desse projeto, será necessário utilizar materiais bibliográficos que apresentem os conceitos de agrupamento espectral, além das bases de dados das quais deseja-se realizar o agrupamento. Além disso, serão utilizados softwares para processamento de dados, como MatLab ou R, por exemplo.

## 6 Assinaturas

Declaramos estar cientes dos compromissos assumidos pela aluna Danielle Aguiar Carneiro Silva no desenvolvimento do trabalho proposto.

---

Danielle Aguiar Carneiro Silva

---

Orientador: Prof. Antônio de Pádua Braga

# Referências Bibliográficas

- [NG06] Boaz Nadler and Meirav Galun. Fundamental limitations of spectral clustering. In *Advances in Neural Information Processing Systems*, pages 1017–1024, 2006.
- [NJW<sup>+</sup>02] Andrew Y Ng, Michael I Jordan, Yair Weiss, et al. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 2:849–856, 2002.
- [SLH90] Guy L Scott and Hugh Christopher Longuet-Higgins. Feature grouping by ‘relocalisation’ of eigenvectors of the proximity matrix. In *BMVC*, pages 1–6. Citeseer, 1990.
- [VM03] Deepak Verma and Marina Meila. A comparison of spectral clustering algorithms. *University of Washington Tech Rep UWCSE030501*, 1:1–18, 2003.
- [Wei99] Yair Weiss. Segmentation using eigenvectors: A unifying view. *International Conference on Computer Vision*, pages 975–982, 1999.