



# Universidade Federal de Pernambuco

Graduação em Ciência da Computação

Centro de Informática

2015.1

---

Avaliando modelos de clustering colaborativo aplicado à  
dados relacionais

---

Proposta de Trabalho de Graduação

**Aluno:** Diogo Philippini Pontual Branco ([dppb@cin.ufpe.br](mailto:dppb@cin.ufpe.br))

**Orientador:** Francisco de Assis Tenório de Carvalho ([fatc@cin.ufpe.br](mailto:fatc@cin.ufpe.br))

Recife, 24 de março de 2015

## Sumário

Contexto .....	2
Objetivos .....	2
Cronograma.....	3
Possíveis avaliadores.....	3
Referências.....	4
Assinaturas .....	5

## Contexto

Clustering é uma tarefa essencial e frequentemente usada em reconhecimento de padrão, mineração de dados, visão computacional e bioinformática. Seu objetivo é organizar um conjunto de itens em clusters tal que itens contidos num cluster possuem alto grau de similaridade, enquanto itens pertencentes à clusters diferentes possuem alto grau de dissimilaridade [2].

Cada objeto (ou item) pode ser representado como dados relacionais, onde cada par de objeto é representado por um relacionamento. O caso mais comum é quando se tem uma matriz de dissimilaridades de dados, onde dada posição da matriz representa a dissimilaridade entre dois objetos [2]. Esses objetos podem ser (e geralmente são) complexos, tais como objetos multimídia, o que faz com que o relacionamento entre objetos possa ser descrito por múltiplas matrizes de (dis)similaridade [1].

[1] propõe um algoritmo para clustering e agregação de dados relacionais (*clustering and aggregation of relational data*, CARD). Além de particionar os dados em clusters significativos, CARD agrega distâncias par a par advindas de múltiplas matrizes relacionais e aprende um peso de relevância para cada matriz em cada cluster. O CARD possui duas versões: o  $CARD_F$  e o  $CARD_R$ ; sendo baseados nos algoritmos RFCM (NERF) e FANNY respectivamente.

[2] propõe outros algoritmos para clustering e agregação de dados relacionais dados por múltiplas matrizes de dissimilaridades. Nesse trabalho, faz-se a comparação de um dos algoritmos propostos, o MFCMdd-RWL-P, com o NERF e o  $CARD_R$  e constata que o MFCMdd-RWL-P possui um melhor desempenho.

## Objetivos

O objetivo desse trabalho é implementar os algoritmos  $CARD_F$  [1] e MFCMdd-RWL-S [2], aplicá-los à uma base de dados relacional e comparar os resultados. Os dados relacionais poderão ser gerados, a partir de uma base já existente, ou pode-se simplesmente usar uma base de dados relacionais já gerados e disponíveis. A comparação dos resultados será feita através de medidas estatísticas que avaliarão o desempenho de cada algoritmo.

## Cronograma

	Abril				Maio				Junho				Julho			
<b>Implementação, teste e validação do CARD<sub>F</sub> e do MFCMdd-RWL-S</b>																
<b>Gerar dados relacionais</b>																
<b>Aplicar os algoritmos aos dados e verificar resultados</b>																
<b>Elaboração do relatório</b>																
<b>Elaboração da apresentação</b>																

## Possíveis avaliadores

- Francisco de Assis Tenório de Carvalho ([fatc@cin.ufpe.br](mailto:fatc@cin.ufpe.br))
- Sergio Ricardo de Melo Queiroz ([srmq@cin.ufpe.br](mailto:srmq@cin.ufpe.br))

## Referências

[1] Hichem Frigui, Cheul Hwang, and Frank Chung-Hoon Rhee. 2007. Clustering and aggregation of relational data with applications to image database categorization. *Pattern Recogn.* 40, 11 (November 2007), 3053-3068. DOI=10.1016/j.patcog.2007.02.019.

[2] Francisco de A.T. de Carvalho, Yves Lechevallier, Filipe M. de Melo, Relational partitioning fuzzy clustering algorithms based on multiple dissimilarity matrices, *Fuzzy Sets and Systems*, Volume 215, 16 March 2013, Pages 1-28, ISSN 0165-0114, <http://dx.doi.org/10.1016/j.fss.2012.09.011>.

## **Assinaturas**

Trabalho de graduação: Avaliando modelos de clustering colaborativo aplicado à dados relacionais

Aluno: Diogo Philippini Pontual Branco

Orientador: Francisco de Assis Tenório de Cavalho

Universidade Federal de Pernambuco

Graduação em Ciência da Computação

Centro de Informática

2015.1

---

Francisco de Assis Tenório de Carvalho

**Orientador**

---

Diogo P. P. Branco

**Orientando**