

Aprendendo Histórias: Aprendizado de Máquina para Classificar Hiper-histórias Metabólicas

Larissa de Oliveira Penteado

Orientador: Roberto M. Cesar Jr

Coorientador: Ricardo Luiz de Andrade Abrantes

Apoio:



**FUNDAÇÃO DE AMPARO À PESQUISA
DO ESTADO DE SÃO PAULO**

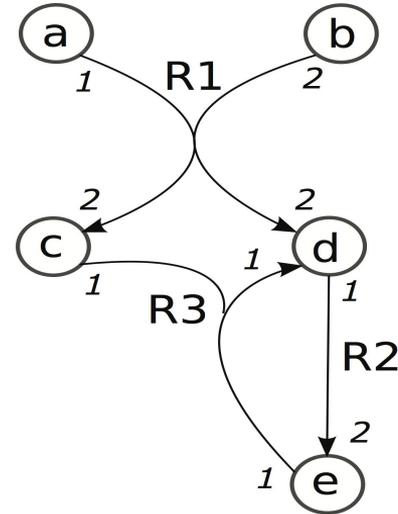
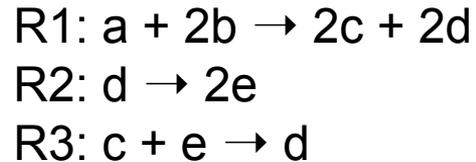
Processo:2017/04250-2

Conceitos Básicos

- Exposição levedura *S. cerevisiae* ao Cádmio(Cd^{2+})
- Rede metabólica
- Perfil metabolômico
- Hiper-histórias metabólicas

Totoro

- TOpological analysis of Transient metabOlic RespOnse
- Modelagem por hipergrafos



Dados utilizados

- Rede metabólica e perfil metabolômico da levedura *S.cerevisiae*
- Conjunto de dados com 268 hiper-histórias
- Conjunto de dados com 56.515.391 hiper-histórias.

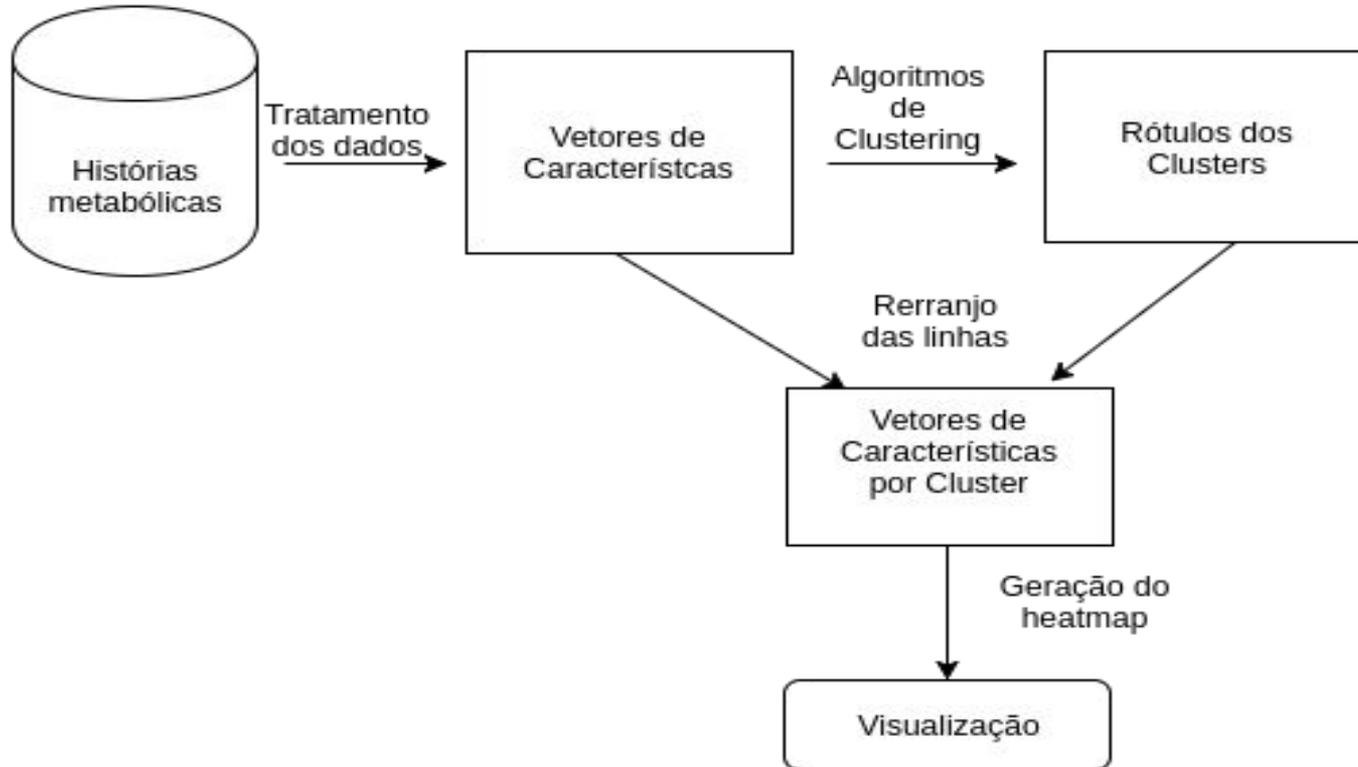
Algoritmos de Clustering

- K-Means
- Mini-Batch K-Means
- Affinity Propagation
- Spectral Clustering

Heurísticas

- Método da Silhueta
- Redução de dimensionalidade por *Principal Component Analysis-PCA*

Pipeline desenvolvido

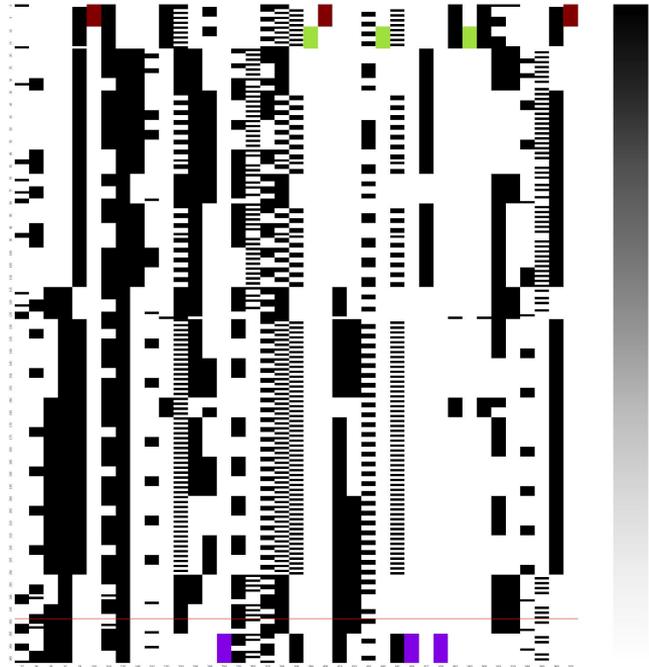


Resultados para o conjunto com 268 hiper-histórias

K-Means

Score de Silhueta: 0.660031

$k = 2$

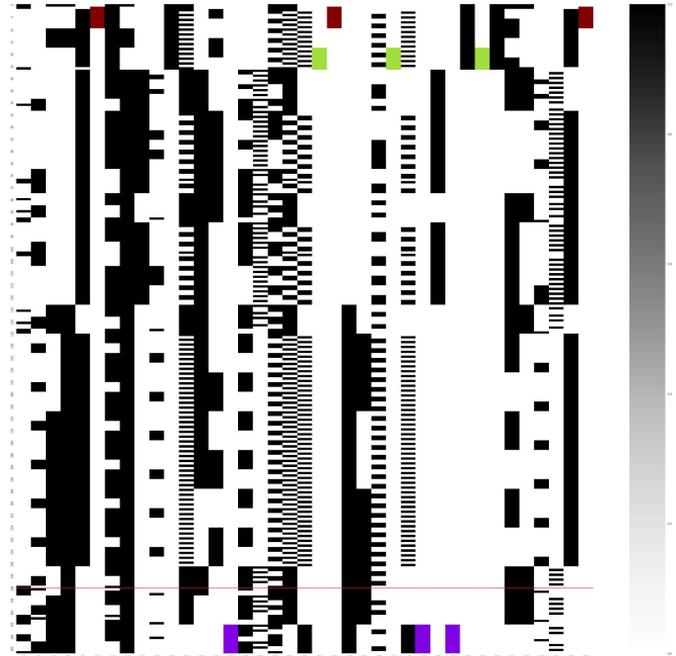


Resultados para o conjunto com 268 hiper-histórias

Mini-Batch K-Means

Score de Silhueta: 0.581661

$k = 2$

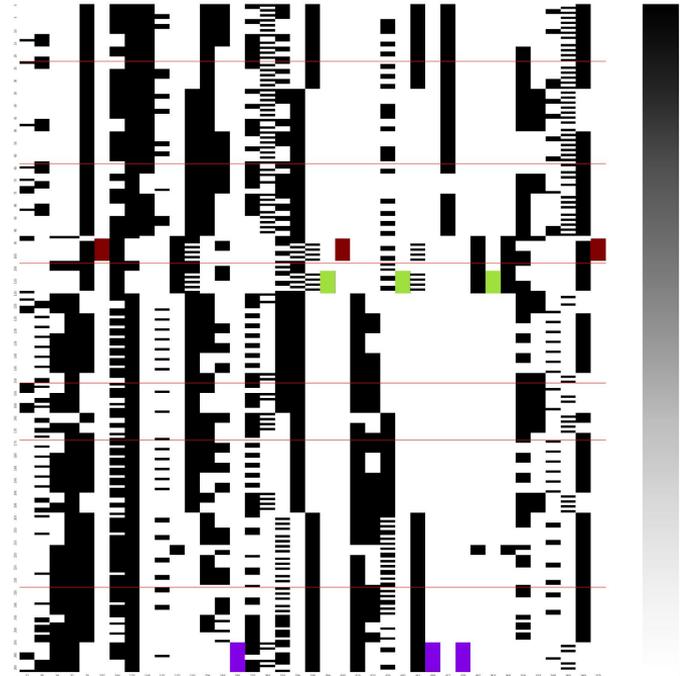


Resultados para o conjunto com 268 hiper-histórias

Affinity Propagation com
PCA

Score de Silhueta: 0.495615

$k = 7$



Análise dos Resultados

Os grupos encontrados correspondiam às vias de reações de:

Nicotinamida

Aminoácido purina

PPP (*Pentose Phosphate Pathway*)

Relacionados à via da Glutathiona

Resultados para o conjunto com 56.515.391 hiper-histórias

- Mini-Batch K-Means
- Ainda em análise
- Dificuldade: Muitos dados, problema de memória com o *Scikit-Learn*

Conclusão

- Reações relacionadas à desintoxicação da levedura em relação ao Cádmio (Cd^{2+})
- Identificação de grupos com diferentes funções biológicas
- Auxílio na compreensão dos dados obtidos pelo método *Totoro*

Trabalhos Futuros

- Análise dos resultados obtidos com o conjunto maior de hiper-histórias
- Métodos mais eficientes em relação à memória e ao tempo de execução

Perguntas?

Muito Obrigada!