

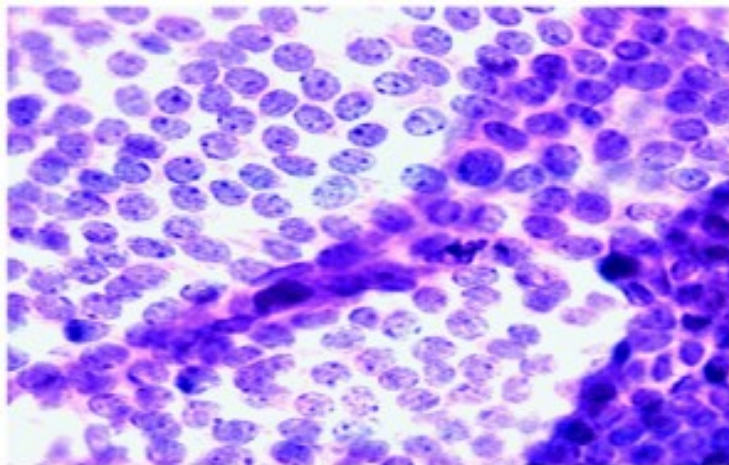


Trabalho MC536

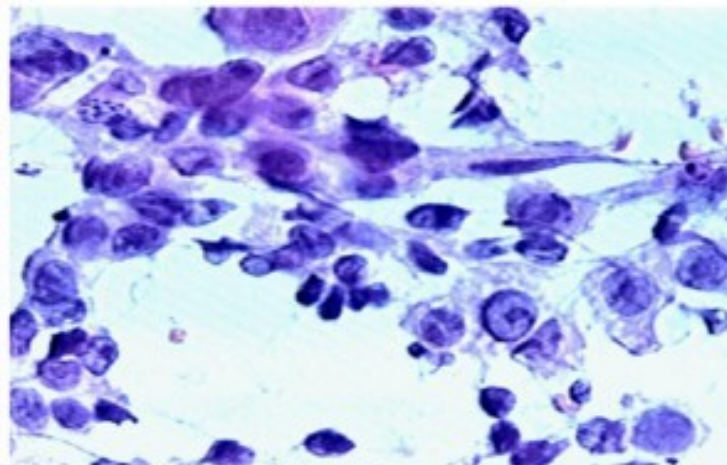
Matheus Rotta Alves (184403)
Douglas Matos Gomes

Overview

- Dataset sobre tumores na mama;
- Características escalares (raio, perímetro);
- Características visuais (compactude, simetria).



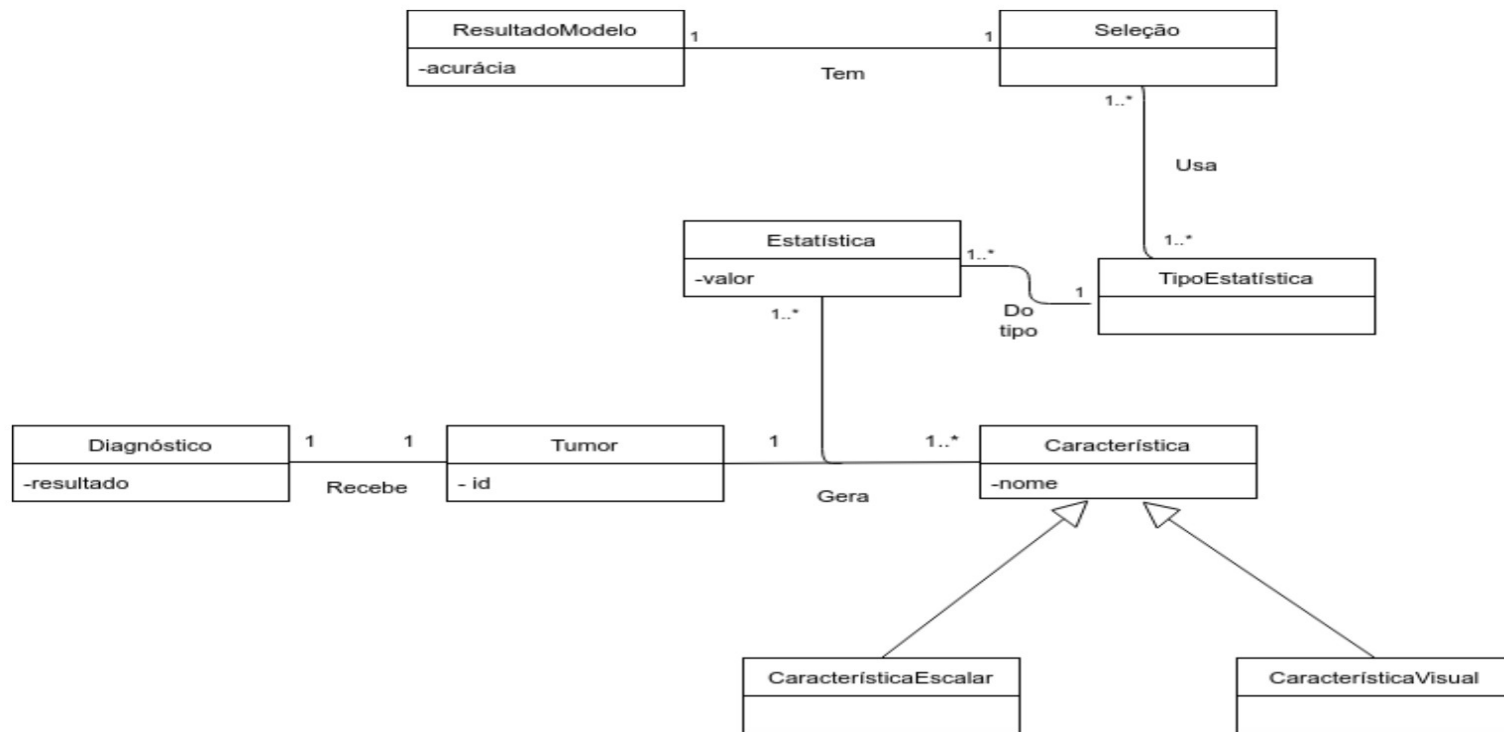
Smear with BENIGN diagnosis – uniform nucleus of cells, symmetrical, homogeneous, with areas within normal size



Smear with MALIGNANT diagnosis – nucleus of cells without uniformity, asymmetrical, not homogeneous (multiple sizes) and with areas above normal size

Modelo Conceitual

Segunda versão:



Modelo Lógico

- Tumor: (idTumor, diagnóstico);
- MédiaValoresEscalares: (idTumor, médiaRaios, médiaPerímetros, médiaÁreas);
- MédiaValoresVisuais: (idTumor, médiaTexturas, médiaSuavidades, médiaCompactudes, médiaConcavidades, médiaPontosCôncavos, médiaSimetria, médiaDimensãoFractal);
- Equivalentemente para desvio padrão e maiores valores.

Etapa 2 (análise com SQL)

- Buscou-se avaliar a capacidade de algumas características de separar o dataset em diagnósticos benignos ou malignos.
- Exemplo de query executada na etapa 2:

select t.Diagnostico, **count**(*)

from MediaValoresEscalares me, MediaValoresVisuais mv,
Tumor t

where t.Id = me.idTumor

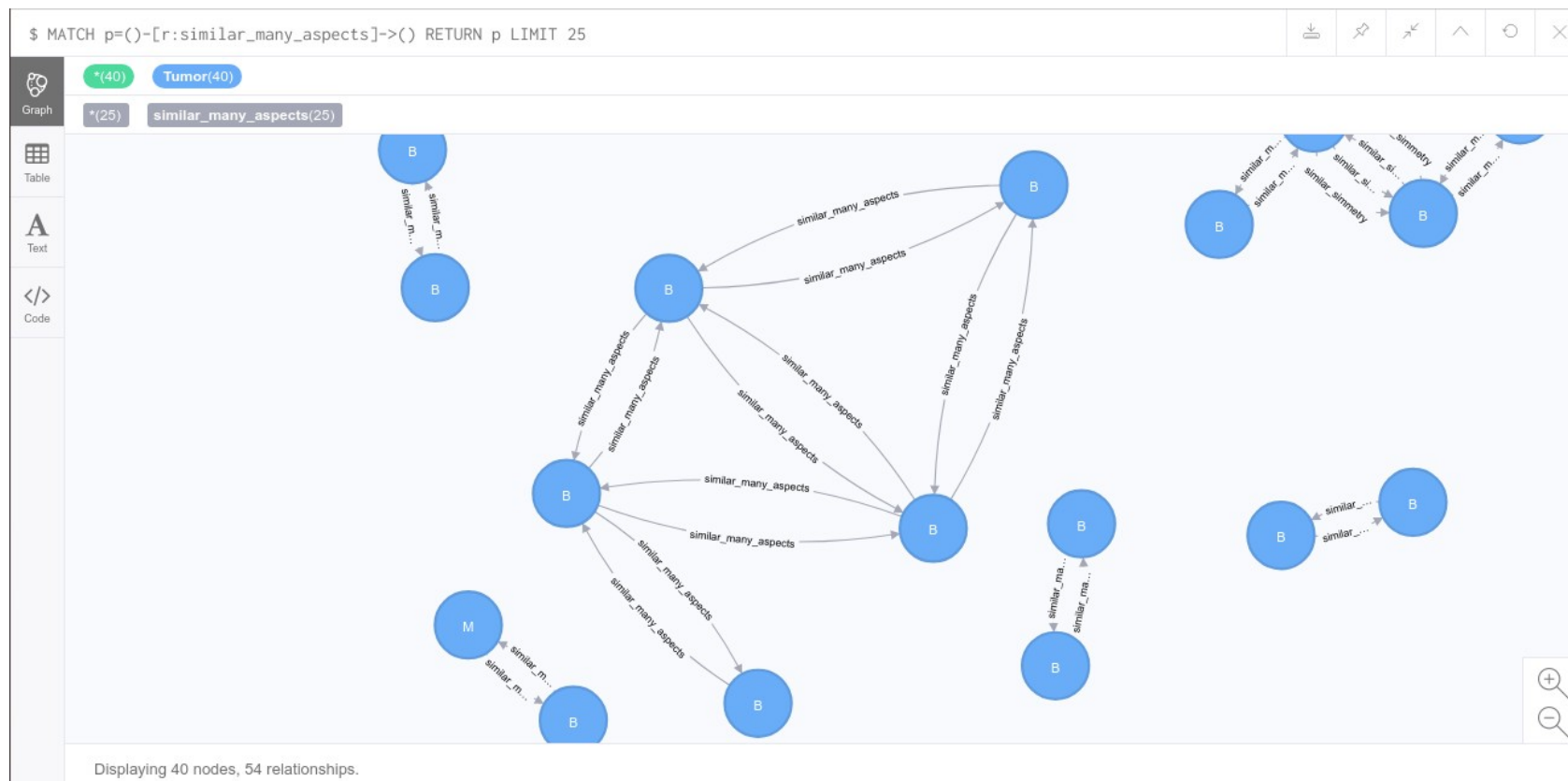
and mv.idTumor = me.idTumor

and me.mediaRaios > 15

and mv.mediaTexturas > 15 **group by** t.Diagnostico;

Etapa 3 (Análise com grafos - Neo4j)

- Buscou-se analisar visualmente a clusterização de tumores parecidos criando-se arestas entre nós semelhantes:



Etapa 3 (Análise com grafos - Neo4j)

- Exemplo de query executada na etapa 3 (cypher):

MATCH (a:mediaVisuais)

MATCH (b:mediaVisuais)

MATCH (e:mediaEscalares)

MATCH (f:mediaEscalares)

WHERE (**abs**(**toFloat**(a.mediaSimetria) - **toFloat**(b.mediaSimetria)) < 0.1

and **abs**(**toFloat**(a.mediaTexturas) - **toFloat**(b.mediaTexturas)) < 0.1

and **abs**(**toFloat**(a.mediaCompactudes) - **toFloat**(b.mediaCompactudes)) < 0.1

and **abs**(**toFloat**(e.mediaRaios) - **toFloat**(f.mediaRaios)) < 0.1

and a <> b and e <> f and a.idTumor = e.idTumor and b.idTumor = f.idTumor)

MATCH(c:Tumor)

MATCH(d:Tumor)

WHERE(c.Id = a.idTumor and d.Id = b.idTumor)

CREATE (c)-[r:similar_many_aspects]->(d)

Etapa 4 - (xQuery/XML)

- Foi necessário converter os dados de CSV para XML, e as queries foram executadas online no “<http://try.zorba.io/>”.
- Buscou-se novamente analisar clusterização, mas dessa vez apenas em pares, usando xQuery.



Etapa 4 - (xQuery/XML)

- Exemplo de query executada na etapa 4:

```
let $fichariodoc := doc('mydoc.xml')
```

```
for $i in ($fichariodoc//tumor)
```

```
for $j in ($fichariodoc//tumor)
```

```
where (abs(data($i/mean_symmetry) - data($j/mean_symmetry)) < 0.002 and  
abs(data($i/mean_concave_points) - data($j/mean_concave_points)) < 0.002 and  
($i != $j))
```

```
return
```

```
<dupla>
```

```
<primeiro> {data($i/diagnosis)} </primeiro>
```

```
<segundo> {data($j/diagnosis)} </segundo>
```

```
</dupla>
```

Etapa 4 - (xQuery/XML)

- Resultado:

```
<dupla><primeiro>B</primeiro><segundo>B</segundo></dupla><dupla><primeiro>M</primeiro><segundo>B</segundo></dupla><dupla><primeiro>B</primeiro><segundo>B</segundo></dupla><dupla><primeiro>B</primeiro><segundo>B</segundo></dupla><dupla><primeiro>B</primeiro><segundo>M</segundo></dupla><dupla><primeiro>B</primeiro><segundo>B</segundo></dupla><dupla><primeiro>B</primeiro><segundo>B</segundo></dupla><dupla><primeiro>B</primeiro><segundo>B</segundo></dupla><dupla><primeiro>B</primeiro><segundo>B</segundo></dupla><dupla><primeiro>B</primeiro><segundo>B</segundo></dupla><dupla><primeiro>B</primeiro><segundo>B</segundo></dupla><dupla><primeiro>B</primeiro><segundo>B</segundo></dupla>
```

- Ou, para conseguirmos ler (B – benigno, M - maligno):

- BB MB

- BB BB

- BM BB

- BB BB

- BB BB