

# MC536 - Grupo 9 - Dieta de frutas

---

**Estevam Souza Machado** **140596**

**Farzin Shams** **139424**

**Guilherme Augusto Sakai Yoshike** **138446**

**Pedro Correa Bueno de Castro** **118355**



# Definição do Problema

---

- Proposta inicial:
  - Compor uma dieta com porções de alimentos distintos que suprissem as necessidades nutricionais diárias de vitaminas e minerais por faixa etária.



# Definição do Problema

- Bases encontradas:

## Faixas etárias -> Vitaminas e minerais

Age	Vitamin A See Report		Folate (Vitamin B-9) See Report		Vitamin C See Report		Vitamin D See Report		Calcium See Report		Iron See Report	
	RDA <sup>1</sup>	Upper Limit <sup>2</sup>	RDA <sup>3</sup>	Upper Limit <sup>4</sup>	RDA	Upper Limit	RDA <sup>5</sup>	Upper Limit	RDA	Upper Limit	RDA	Upper Limit
1 - 3	1,000 IU	2,000 IU	150 mcg	300 mcg	15 mg	400 mg	600 IU	2,500 IU	700 mg	2,500 mg	7 mg	40 mg
4 - 8	1,300 IU	3,000 IU	200 mcg	400 mcg	25 mg	650 mg	600 IU	3,000 IU	1,000 mg	2,500 mg	10 mg	40 mg
9 - 13	2,000 IU	5,666 IU	300 mcg	600 mcg	45 mg	1,200 mg	600 IU	4,000 IU	1,300 mg	2,500 mg	8 mg	40 mg
14 - 18	1,000 IU	9,333 IU	400 mcg	800 mcg	75 mg (m) 65 mg (f) 80 mg (preg) 115 mg (lact)	1,800 mg	600 IU	4,000 IU	1,300 mg	2,500 mg	11 mg (m) 15 mg (f) 27 mg (preg) 10 mg (lact)	45 mg
Adult	3,000 IU (m) 2,300 IU (f)	10,000 IU	400 mcg 600 mcg (preg)/ 500 mcg (lact)	1,000 mcg	90 mg (m) 75 mg (f) 85 mg (preg) 120 mg (lact)	2,000 mg	600 IU (51- 70 years) 800 IU (71+ years)	4,000 IU	1,000 mg (to 50 years) 1,200 mg (51+ years)	2,500 mg	8 mg (m) 18 mg (f 19 to 50 years) 8 mg (f 51+ years) 27 mg (preg) 9 mg (lact)	45 mg

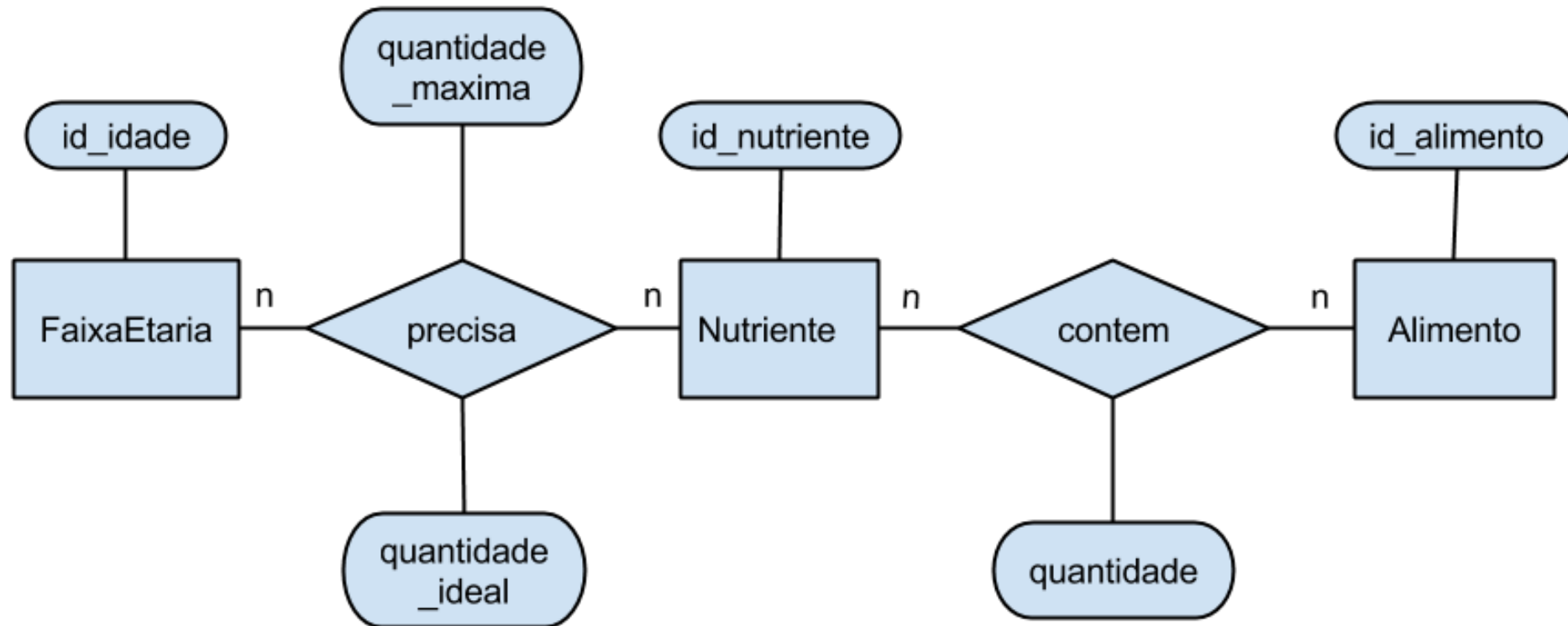
## Alimentos -> Nutrientes

Pears, values per 100g		
Minerals		
Calcium, Ca	mg	9
Iron, Fe	mg	0.18
Zinc, Zn	mg	0.10
Vitamins		
Vitamin C, total ascorbic acid	mg	4.3
Thiamin	mg	0.012
Riboflavin	mg	0.026
Niacin	mg	0.161

# Definição do Problema

---

- Modelo ER proposto para a resolução:



# Definição do Problema

---

- **Complicações:**

- Base de alimentos muito extensa
- Base de nutrientes é diferente entre as bases
- Necessidades ideal (mínima) e máxima diárias
- Gênero

- **Reformulação do problema:**

- Frutas e sucos
  - Intersecção das bases de nutrientes
  - Apenas a necessidade ideal diária para cada faixa etária
  - Gênero masculino
-

# Extração de dados e Data Cleaning

---

- Ferramentas:

Python



BeautifulSoup





# Extração de dados e Data Cleaning

- Faixas etárias -> vitaminas e minerais

Recommended Daily Intakes and Upper Limits for Common Nutrients

Age	Vitamin A See Report		Folate (Vitamin B-9) See Report		Vitamin C See Report		Vitamin D See Report		Calcium See Report		Iron See Report	
	RDA <sup>1</sup>	Upper Limit <sup>2</sup>	RDA <sup>3</sup>	Upper Limit <sup>4</sup>	RDA	Upper Limit	RDA <sup>5</sup>	Upper Limit	RDA	Upper Limit	RDA	Upper Limit
1 - 3	1,000 IU	2,000 IU	150 mcg	300 mcg	15 mg	400 mg	600 IU	2,500 IU	700 mg	2,500 mg	7 mg	40 mg
4 - 8	1,300 IU	3,000 IU	200 mcg	400 mcg	25 mg	650 mg	600 IU	3,000 IU	1,000 mg	2,500 mg	10 mg	40 mg
9 - 13	2,000 IU	5,666 IU	300 mcg	600 mcg	45 mg	1,200 mg	600 IU	4,000 IU	1,300 mg	2,500 mg	8 mg	40 mg
14 - 18	1,000 IU	9,333 IU	400 mcg	800 mcg	75 mg (m) 65 mg (f) 80 mg (preg) 115 mg (lact)	1,800 mg	600 IU	4,000 IU	1,300 mg	2,500 mg	11 mg (m) 15 mg (f) 27 mg (preg) 10 mg (lact)	45 mg
Adult	3,000 IU (m) 2,300 IU (f)	10,000 IU	400 mcg (m) 600 mcg (preg)/ 500 mcg (lact)	1,000 mcg	90 mg (m) 75 mg (f) 85 mg (preg) 120 mg (lact)	2,000 mg	600 IU (51- 70 years) 800 IU (71+ years)	4,000 IU	1,000 mg (to 50 years) 1,200 mg (51+ years)	2,500 mg	8 mg (m) 18 mg (f) 19 to 50 years) 8 mg (f 51+ years) 27 mg (preg) 9 mg (lact)	45 mg

Source: Dietary Reference Intakes Tables and Application [td.reviewTable](#) 68px x 154px National Academy of Sciences,

```

<center>
"
        600 IU (51- 70 years)"
<br>
"
        800 IU (71+ years)
"
</center>

<center>
"75 mg (m)"
<br>
"
        65 mg (f)"
<br>
"
        80 mg (preg)"
"
        115 mg (lact)"
</center>

```

# Extração de dados e Data Cleaning

---

- Faixas etárias -> vitaminas e minerais
    - Desafios:
      - Tags não padronizadas
      - Caracteres inválidos
      - Valores e unidades na mesma célula
      - Informações adicionais não padronizadas
      - Caracteres de tabulação, carriage return, newline, etc.
      - (IU, mg, mcg) -> mcg
-



# Extração de dados e Data Cleaning

---

- Faixas etárias -> vitaminas e minerais

```
#etapa que tira os caracteres inuteis dos valores das tabelas
col_names_cleaned = re.sub('(See Report).*', '', col_names[0].text)
col_names_cleaned = re.sub('\(.*', '', col_names_cleaned)
col_names_cleaned = re.sub(' *$', '', col_names_cleaned)
cells_mg = cols[col + (col - 1)].find_all("center") #tabela alinhada, calculo para obter os dados referentes a cada nutriente
cells_mg_cleaned = re.sub('( )*', '', cells_mg[0].text)
cells_mg_cleaned = re.sub('(\r\n)|(\n)|(\t)|(\s).*', '', cells_mg_cleaned)
cells_mg_cleaned = re.sub('(g)+.*', 'g', cells_mg_cleaned)
cells_mg_cleaned = re.sub('(IU)+.*', 'IU', cells_mg_cleaned)
cells_mg_cleaned = re.sub(',', '', cells_mg_cleaned)
valoridealprocura = re.findall('[0-9]+[.]*[0-9]*', cells_mg_cleaned)
valorideal = valoridealprocura[0]
unidadeideal = re.sub('[0-9]+[.]*[0-9]*[ ]*', '', cells_mg_cleaned)
multideal = 1
if unidadeideal == 'mg':
    multideal = 1000
elif unidadeideal == 'g':
    multideal = 1000000
elif unidadeideal == 'IU':
    if col_names_cleaned == 'Vitamin A':
        multideal = 0.3
    elif col_names_cleaned == 'Vitamin C':
        multideal = 50
    elif col_names_cleaned == 'Vitamin D':
        multideal = 0.025
    elif col_names_cleaned == 'Vitamin E':
        multideal = 0.667
```

# Extração de dados e Data Cleaning

- Alimentos -> nutrientes

09006	Apples, raw, without skin, cooked, microwave	Fruits and Fruit Juices
09003	Apples, raw, with skin	Fruits and Fruit Juices
09403	Apricot nectar, canned, with added ascorbic acid	Fruits and Fruit Juices
09036	Apricot nectar, canned, without added ascorbic acid	Fruits and Fruit Juices
09029	Apricots, canned, extra heavy syrup pack, without skin, solids and liquids	Fruits and Fruit Juices
09025	Apricots, canned, extra light syrup pack, with skin, solids and liquids	Fruits and Fruit Juices
09357	Apricots, canned, heavy syrup, drained	Fruits and Fruit Juices

## Pears, values per 100g

### Minerals

Calcium, Ca	mg	9
Iron, Fe	mg	0.18
Zinc, Zn	mg	0.10

### Vitamins

Vitamin C, total ascorbic acid	mg	4.3
Thiamin	mg	0.012
Riboflavin	mg	0.026
Niacin	mg	0.161

# Extração de dados e Data Cleaning

---

- Alimentos -> nutrientes
    - Desafios:
      - Iteração a muitas páginas.
      - Muitos alimentos do mesmo tipo.
      - Tag com problema de sintaxe.
      - Nomes de nutrientes em formatos diferentes dos da base anterior.
      - Intersecção dos nutrientes nas bases.
      - Padronização de unidades.
-

# Extração de dados e Data Cleaning

---

- Alimentos -> nutrientes

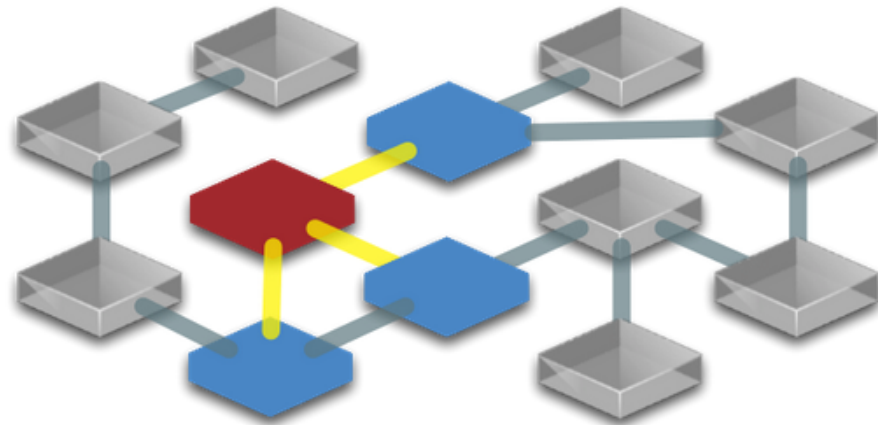
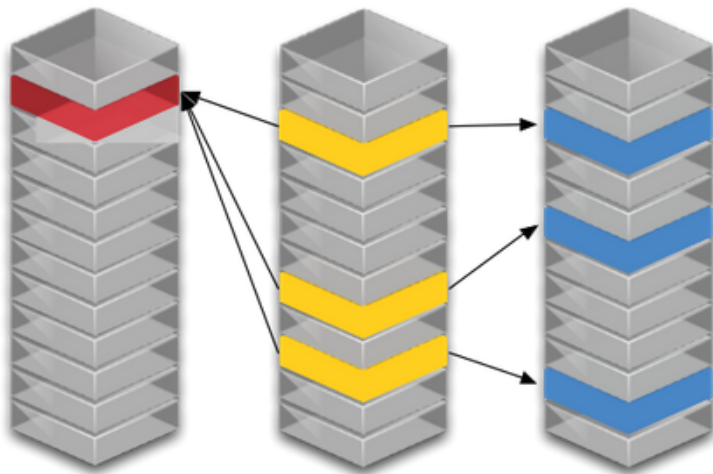
```
vetor_alimentos = []
vetor_nutrientes = []
nutriente_existente = False

#iteracao nas paginas dos alimentos
for page in range(0,14):
    url = "http://ndb.nal.usda.gov/ndb/?format=&count=&max=25&sort=fd_s&fg=Fruits+and+Fruit+Juices&man=&lfacet=&qlookup=&offset="+str(page*25)
    +"&order=asc"
    req = requests.get(url)
    soup = BeautifulSoup(req.content, 'lxml')
    name_aux = ""
    col_aux=""

    # tabela com a lista de alimentos
    table = soup.find_all("div", {"class": "wbox"})
    lines = table[0].find_all("tr")
```

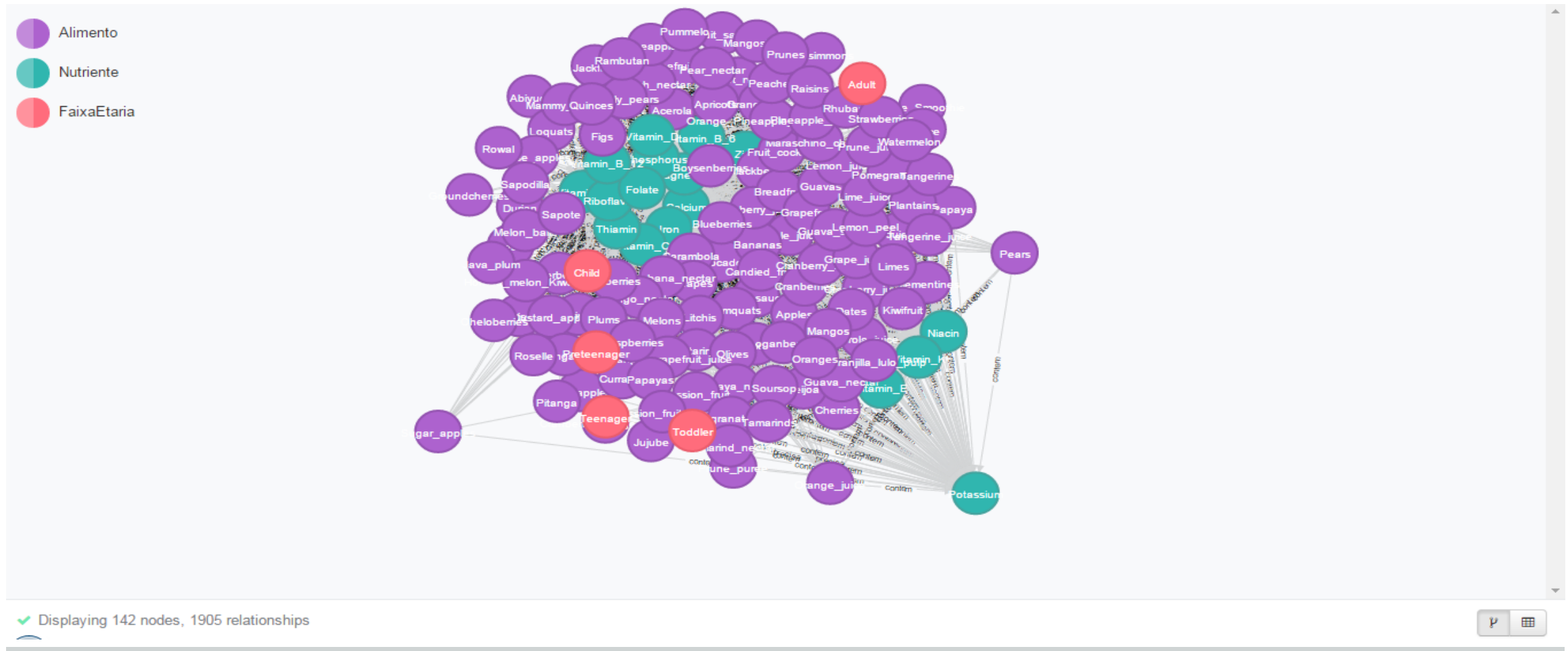
# Criação do banco

---



# Criação do banco

- Alimento, Nutriente, FaixaEtaria e relações



# Criação do banco

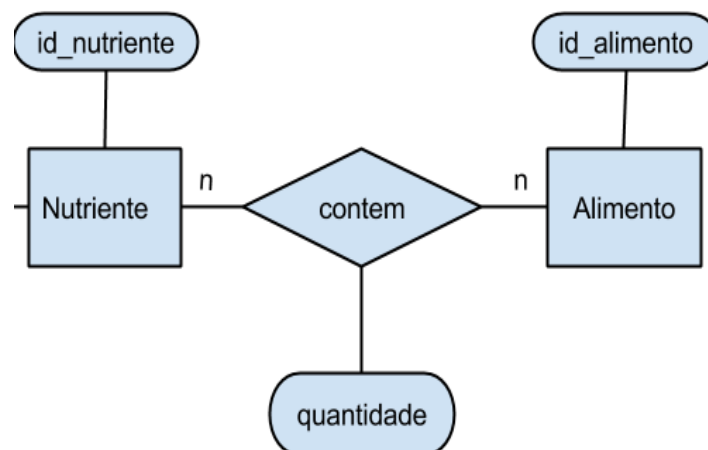
- Relacionamento entre Alimento e Nutriente

```
mysql> select * from Nutriente;
```

id_nutriente
Calcium
Folate
Iron
Magnesium
Niacin
Phosphorus
Potassium
Riboflavin
Thiamin
Vitamin A
Vitamin B-12
Vitamin B-6
Vitamin C
Vitamin D
Vitamin E
Vitamin K
Zinc

```
mysql> select * from Alimento;
```

id_alimento
Abiyuch
Acerola
Acerola juice
Apple juice
Apples
Applesauce
Apricot nectar
Apricots
Avocados
Bananas
Blackberries
Blackberry juice
Blueberries
Boysenberries
Breadfruit
Candied fruit
Carambola
Carissa
Cherimoya
Cherries
Clementines
Crabapples
Cranberries
Cranberry juice
Cranberry sauce
Cranberry-orange relish
Currants
Custard-apple
Dates
Durian
Elderberries
Feijoa
Figs
Fruit cocktail
Fruit salad
Gooseberries



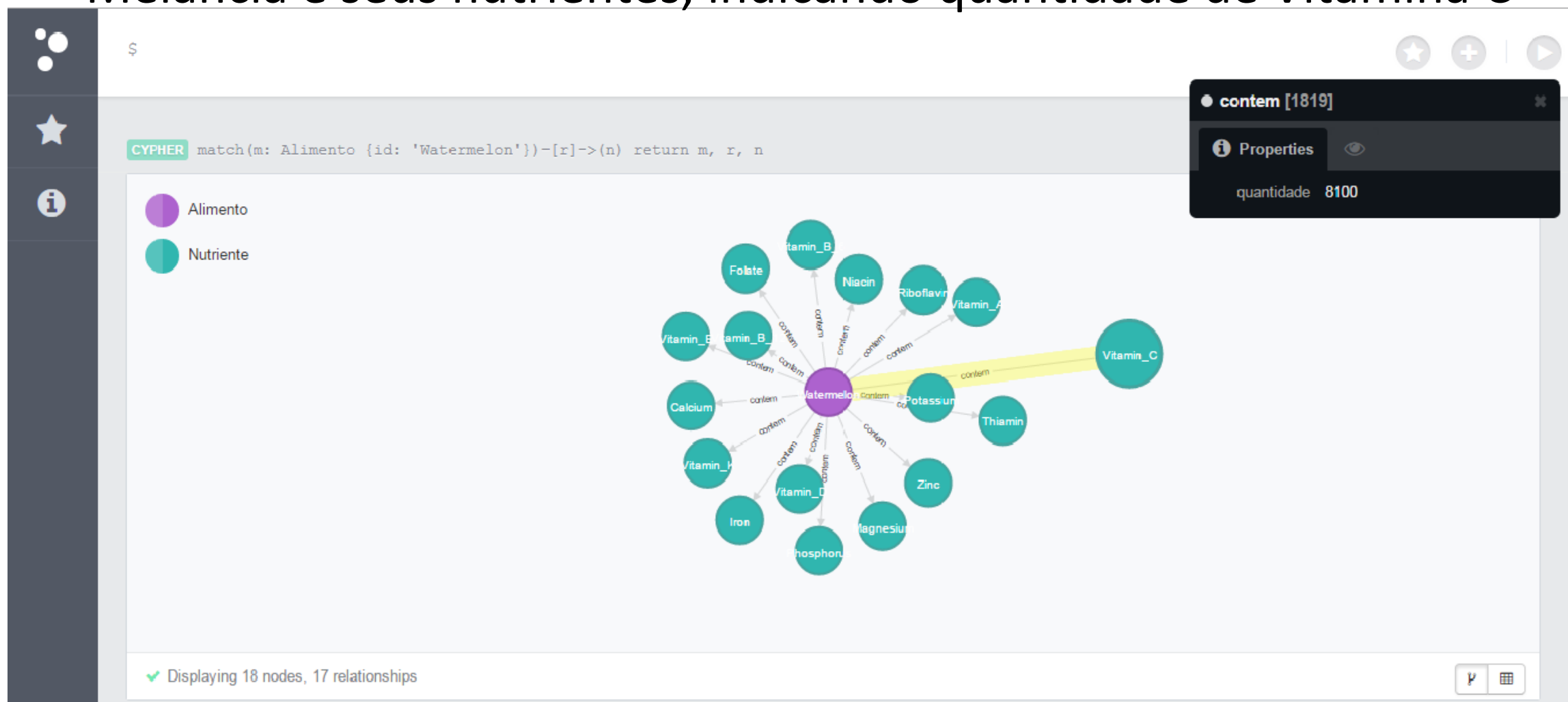
Tangerines	Iron	270.0
Tangerines	Magnesium	11000.0
Tangerines	Niacin	445.0
Tangerines	Phosphorus	10000.0
Tangerines	Potassium	133000.0
Tangerines	Riboflavin	29.0
Tangerines	Thiamin	82.0
Tangerines	Vitamin A	43.0
Tangerines	Vitamin B-12	0.0
Tangerines	Vitamin B-6	42.0
Tangerines	Vitamin C	34200.0
Tangerines	Vitamin D	0.0
Tangerines	Vitamin E	100.0
Tangerines	Vitamin K	0.0
Tangerines	Zinc	510.0
Watermelon	Calcium	7000.0
Watermelon	Folate	3.0
Watermelon	Iron	240.0
Watermelon	Magnesium	10000.0
Watermelon	Niacin	178.0
Watermelon	Phosphorus	11000.0
Watermelon	Potassium	112000.0
Watermelon	Riboflavin	21.0
Watermelon	Thiamin	33.0
Watermelon	Vitamin A	28.0
Watermelon	Vitamin B-12	0.0
Watermelon	Vitamin B-6	45.0
Watermelon	Vitamin C	8100.0
Watermelon	Vitamin D	0.0
Watermelon	Vitamin E	50.0
Watermelon	Vitamin K	0.1
Watermelon	Zinc	100.0

1820 rows in set (0.00 sec)



# Criação do banco

- Melancia e seus nutrientes, indicando quantidade de Vitamina C

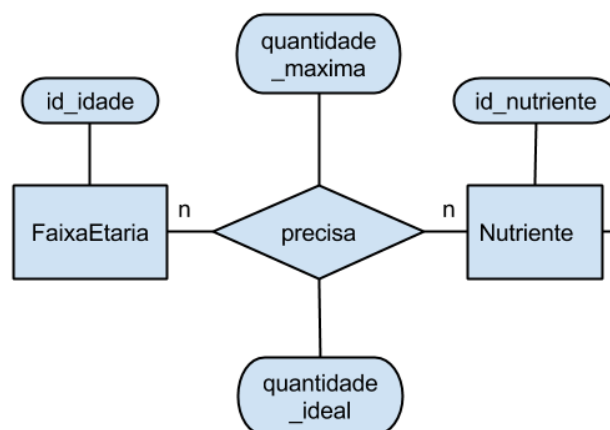


# Criação do banco

- Relacionamento entre FaixaEtaria e Nutriente

```
mysql> show tables;
+-----+
| Tables_in_projetoBD |
+-----+
| Alimento              |
| Alimento_Nutriente    |
| FaixaEtaria           |
| FaixaEtaria_Nutriente |
| Nutriente              |
+-----+
```

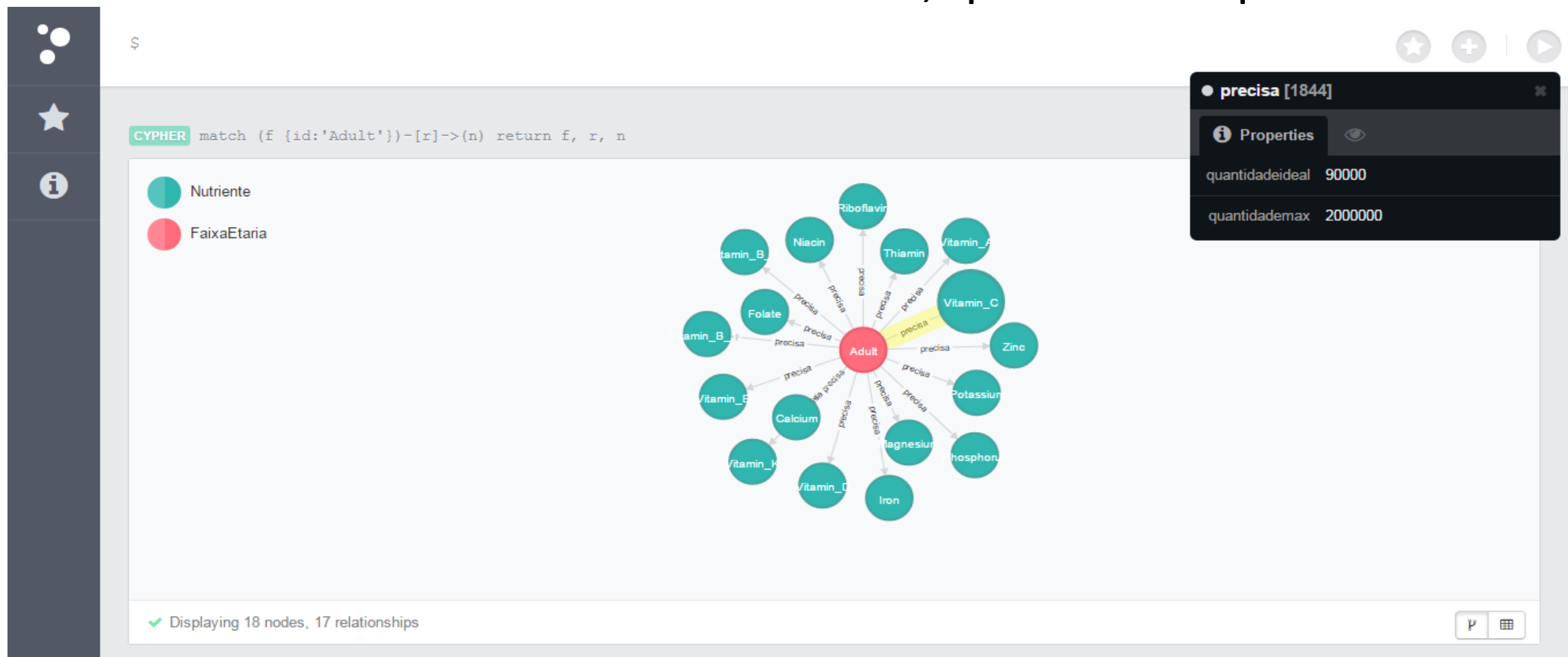
```
mysql> select * from FaixaEtaria;
+-----+
| id_idade |
+-----+
| Adult    |
| Child    |
| Preteenage |
| Teenager |
| Toddler  |
+-----+
```



```
mysql> select * from FaixaEtaria_Nutriente;
+-----+-----+-----+-----+
| id_nutriente | id_idade | quantidade_ideal | quantidade_maxima |
+-----+-----+-----+-----+
| Calcium      | Adult    | 1000000.0        | 2500000.0         |
| Folate       | Adult    | 400.0            | 1000.0            |
| Iron         | Adult    | 8000.0           | 45000.0           |
| Magnesium    | Adult    | 400000.0         | 350000.0          |
| Niacin       | Adult    | 16000.0          | 35000.0           |
| Phosphorus   | Adult    | 700000.0         | 4000000.0         |
| Potassium    | Adult    | 4700000.0        | NULL              |
| Riboflavin   | Adult    | 1300.0           | NULL              |
| Thiamin      | Adult    | 1200.0           | NULL              |
| Vitamin A    | Adult    | 900.0            | 3000.0            |
| Vitamin B-12 | Adult    | 2.4              | NULL              |
| Vitamin B-6  | Adult    | 1300.0           | 100000.0          |
| Vitamin C    | Adult    | 90000.0          | 2000000.0         |
| Vitamin D    | Adult    | 15.0             | 100.0             |
| Vitamin E    | Adult    | 22.0             | 733.7             |
| Vitamin K    | Adult    | 120.0            | NULL              |
| Zinc         | Adult    | 11000.0          | 40000.0           |
| Calcium      | Child    | 1000000.0        | 2500000.0         |
| Folate       | Child    | 200.0            | 400.0             |
| Iron         | Child    | 10000.0          | 40000.0           |
| Magnesium    | Child    | 130000.0         | 110000.0          |
| Niacin       | Child    | 8000.0           | 15000.0           |
| Phosphorus   | Child    | 500000.0         | 3000000.0         |
| Potassium    | Child    | 3800000.0        | NULL              |
| Riboflavin   | Child    | 600.0            | NULL              |
| Thiamin      | Child    | 600.0            | NULL              |
| Vitamin A    | Child    | 390.0            | 900.0             |
| Vitamin B-12 | Child    | 1.2              | NULL              |
+-----+-----+-----+-----+
```

# Criação do banco

- 'Adulto' e seus nutrientes necessários, quantidades para Vitamina C



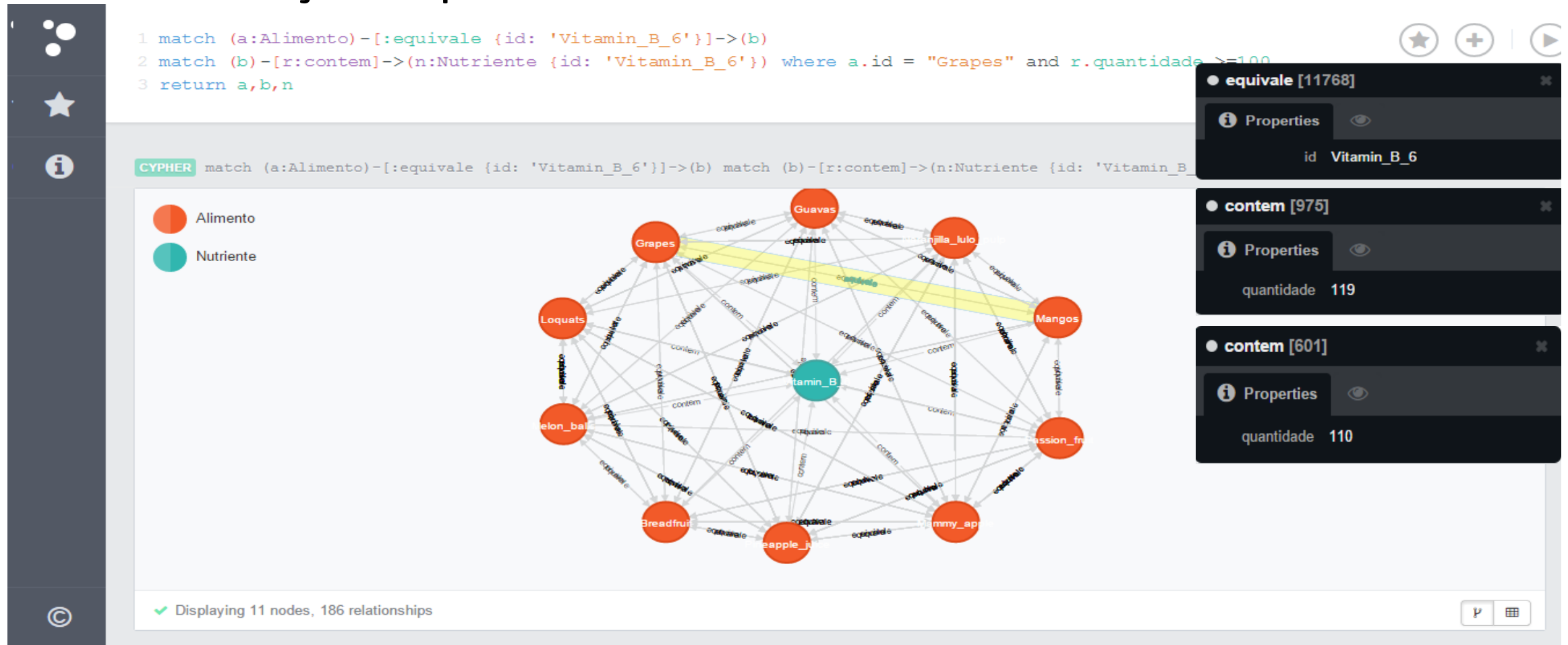
# Análise de Dados e Aplicações

---



# Análise de Dados e Aplicações

- Nova relação 'equivale'



# Análise de Dados e Aplicações

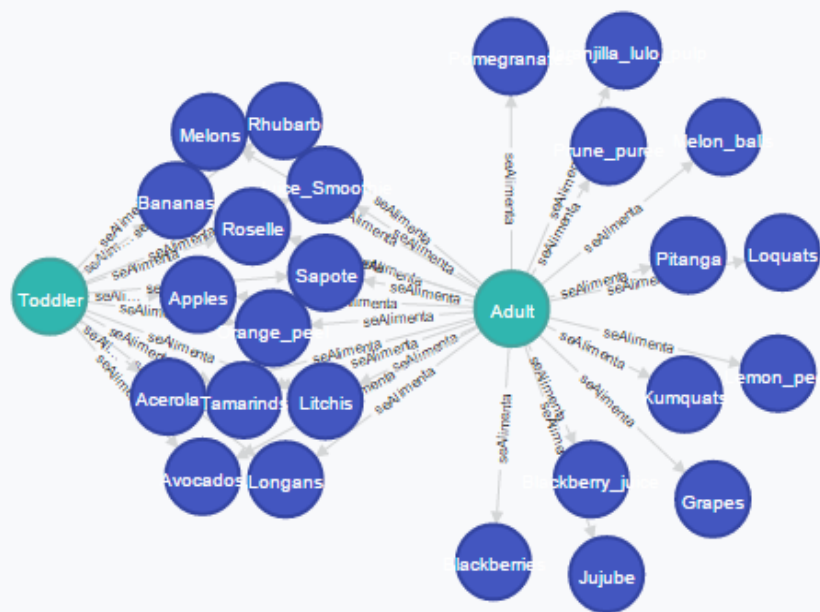
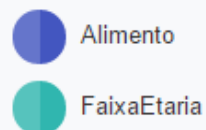
---

- Criação da lista de alimentos para cada FaixaEtaria
  - 1ª abordagem
    - Teste de combinações -> menor lista
      - Resultado: tempo de execução inviável
      - Algumas conclusões: Vitamina D, Vitamina B 12 e Zinco não supridos bem por frutas
  - 2ª abordagem
    - Alimento ideal para cada nutriente
    - Nutrientes não supridos -> adição dos melhores alimentos para cada até suprí-los
      - Resultado: lista encontrada! -> relações seAlimenta criada, entre as faixas etárias e os alimentos que compõem sua lista encontrada. Porém, listas de dietas com mais de 2kg de alimentos. Por quê?

# Análise de Dados e Aplicações

- Relações entre Toddler e Adult e os alimentos da dieta encontrada

CYPHER MATCH (f:FaixaEtaria)-[]->(a:Alimento) WHERE f.id = 'Toddler' OR f.id = 'Adult' RETURN f, a



✓ Displaying 27 nodes, 38 relationships



# Análise de Dados e Aplicações

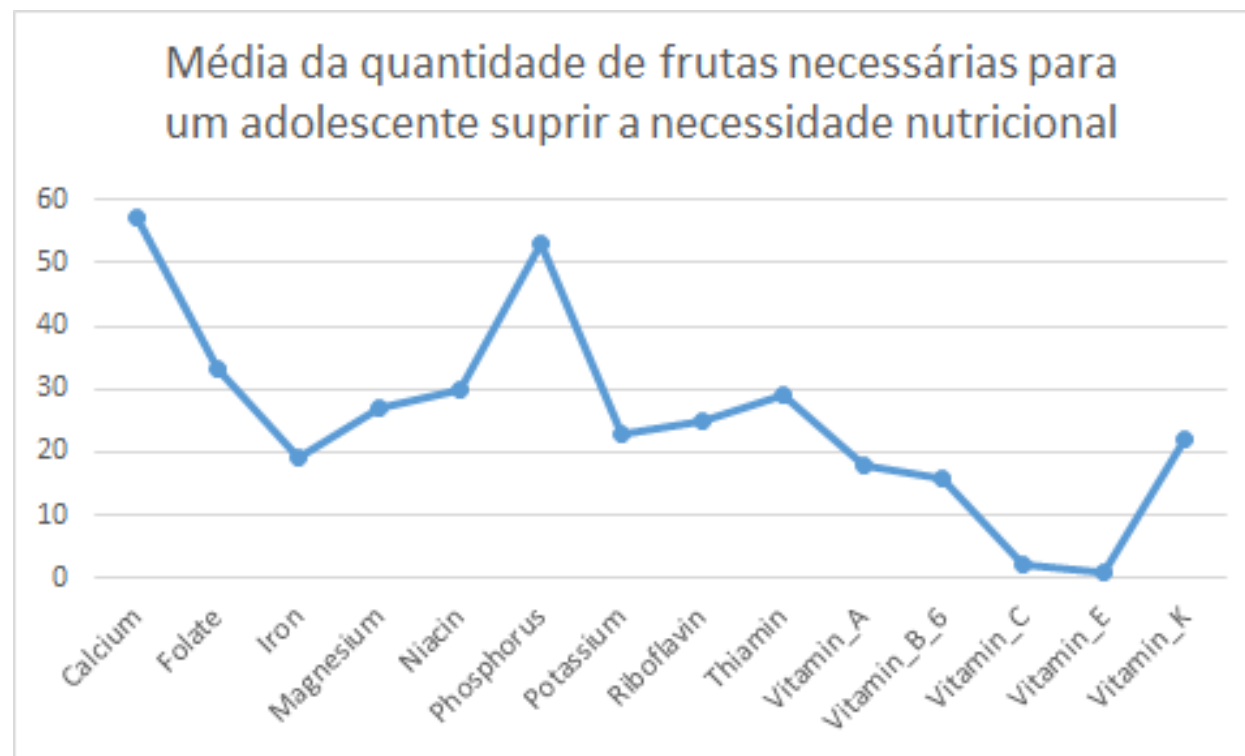
---

- Cálculo da média do número de alimentos necessários para suprir cada nutriente
    - Query da média das quantidades de cada nutriente nos alimentos
    - Query das quantidades ideais de cada nutriente para uma FaixaEtaria
    - Para cada nutriente divide-se a quantidade ideal pela média por alimento
      - Resultado: quantos alimentos, na média, são necessários para suprir cada nutriente
-

# Análise de Dados e Aplicações

---

- Cálculo da média do número de alimentos necessários para suprir cada nutriente



# Análise de Dados e Aplicações

---

- Cálculo da média do número de alimentos necessários para suprir cada nutriente
  - Conclusão
    - Frutas e sucos não são suficientes para todos esses nutrientes!
    - São suficientes, no entanto, para suprir Vitaminas C e E