

MC536 - Projeto 2

Equipe VIRUS

Pietro Grazzioli Golfeto - 223694

Lucas Gabriel Monteiro da Costa - 183967

Guilherme de Luigi Mocelim de Oliveira - 223325

Sara Beatriz da Silva Oliveira - 231288

Vitor Rodrigues Zanata da Silva - 231718

Luiz Fernando Lima Leite - 248405



Motivação e Contexto

A segunda versão do projeto traz uma análise que une o poder de bancos de dados relacionais aos baseados em grafos. Por meio dessa conexão e do uso de fontes com diferentes enfoques, tornamos possível uma investigação da complexidade camuflada no cotidiano alimentar.

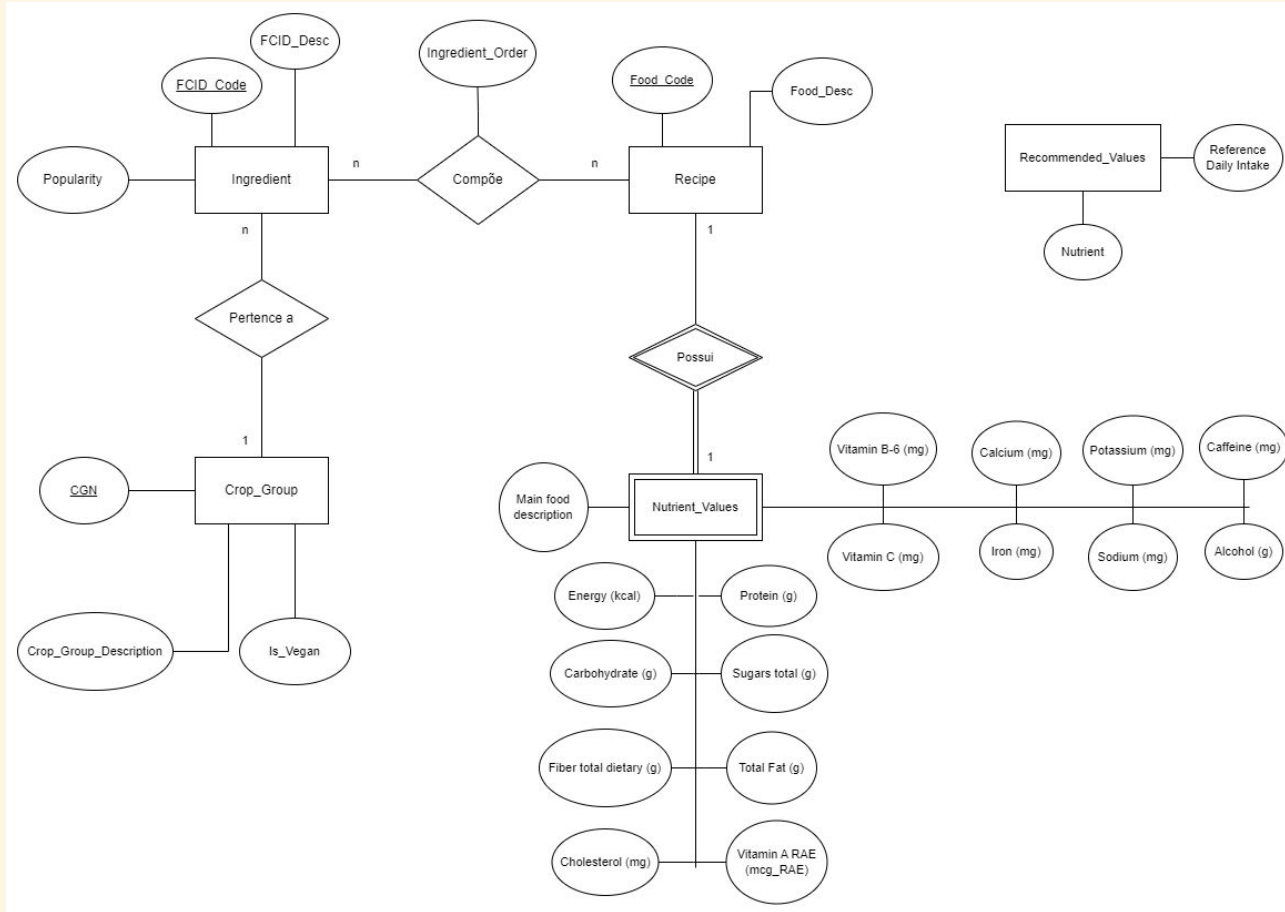
Para nos aprofundarmos nesse trabalho, arquitetamos uma rede de questionamentos que dialogam entre si, construindo passo a passo um painel que explora desde a identificação de relações não triviais entre alimentos até os efeitos que diferentes nichos de consumo possuem sobre as pessoas.



Bases de Dados Utilizadas

Título da Base	Link	Descrição
FCID - Food Commodity Intake Database	https://fcid.foodrisk.org/	Base de dados com receitas e ingredientes
FNDDS - Food and Nutrient Database for Dietary Studies	https://www.ars.usda.gov	Base de dados com valores nutricionais de receitas

Modelo Entidade Relacionamento



Modelo Lógico Relacional

Recipe(_Food_Code_, Food_Desc)

Crop_Group(_CGN_, Crop_Group_Description, Is_Vegan)

Ingredient(_FCID_Code_, _Crop_Group_, FCID_Desc, Popularity)

Crop_Group: chave estrangeira -> Crop_Group

Recipe_Composition(_Food_Code_, _FCID_Code_, Ingredient_Order)

Food_Code: chave estrangeira -> Recipe

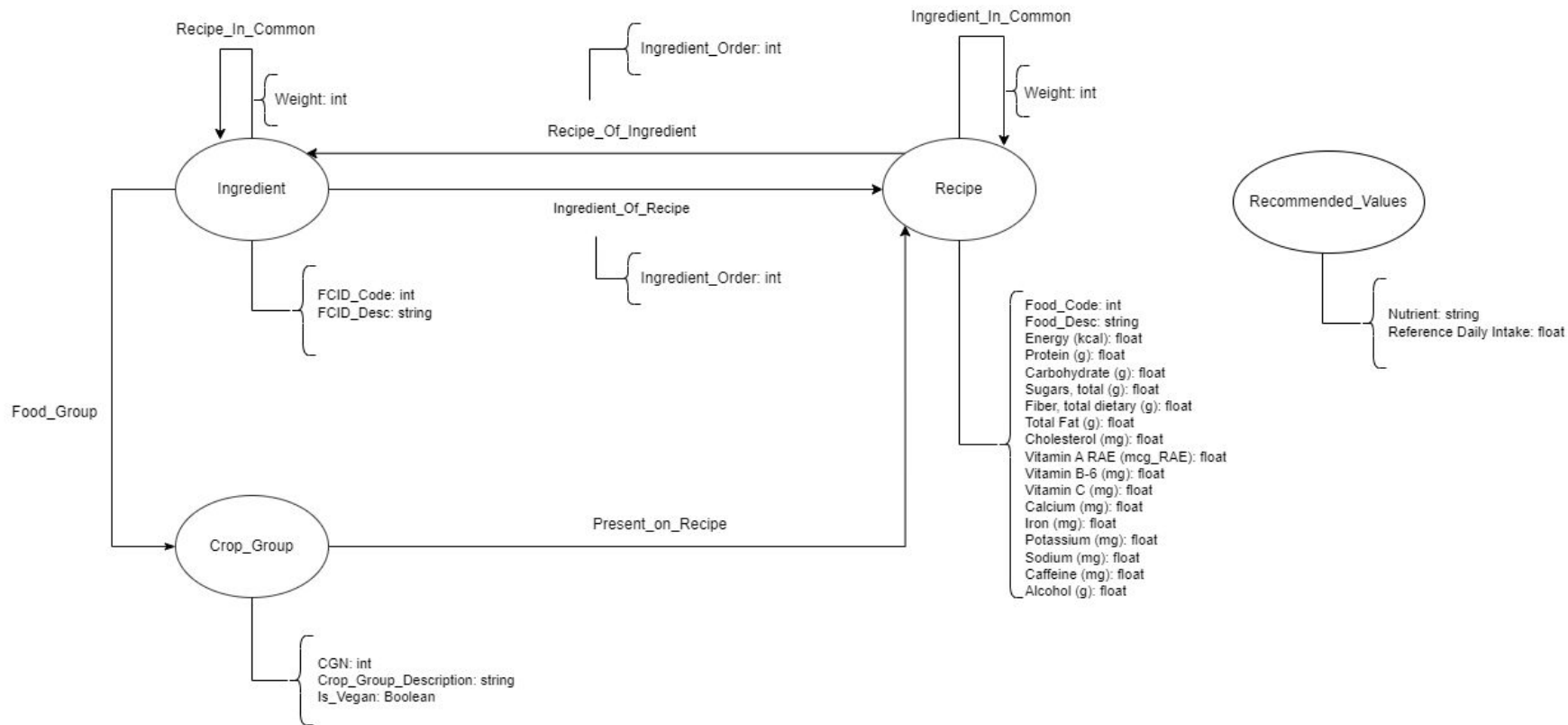
FCID_Code: chave estrangeira -> Ingredient

Nutrient_Values(_Food_Code_, Main food description, Energy (kcal), Carbohydrate (g), Protein (g), Sugars total (g), Total Fat (g), Cholesterol (mg), Fiber total dietary (g), Vitamin A RAE (mcg_RAE), Vitamin B-6 (mg), Vitamin C (mg), Calcium (mg), Iron (mg), Potassium (mg), Sodium (mg), Caffeine (mg), Alcohol (g))

Food_Code: chave estrangeira -> Recipe

Recommended_Values(Nutrient, Reference Daily Intake)

Modelo Lógico de Grafos



Perguntas de Pesquisa/Análise Relacional

Pergunta/Análise 1

- Analisar o quão "diet" é uma receita baseado no percentual diário de determinados indicadores nutricionais (gorduras totais, sódio, açúcar). Quais as receitas mais diet baseadas nesse critério? E quais possuem maior teor de açúcar, gordura e sódio?
 - Primeiro, foi criada uma view estendendo a tabela receitas, adicionando uma nota para essa receita. Essa nota é calculada pela média dos três níveis (nível de gordura, sódio e açúcar), por meio da fórmula:

Nível = (Quantidade do nutriente presente em 100 g da receita)/(Valor recomendado para 100g)

Perguntas de Pesquisa/Análise Relacional

```
CREATE VIEW Receitas_Diet AS
  SELECT R.Food_Code, R.Food_Desc,
         N.Total_Fat/6 AS Nivel_Gordura, N.Sugars_total/15 AS Nivel_Acucar, N.Sodium/600 AS
         Nivel_Sodio, (N.Total_Fat/6 + N.Sugars_total/15 + N.Sodium/600)/3 AS Nota
  FROM FCID_Food_Code_Description R, FNDDS_Nutrient_Values N
  WHERE R.Food_Code = N.Food_code AND N.Sugars_total/15 > 0.05 AND N.Sodium/600
  > 0.05 AND N.Total_Fat/6 > 0.05 AND N.Caffeine = 0 AND N.Alcohol = 0
```

Foram desconsideradas receitas contendo cafeína e álcool, bem como as com valores muito insignificantes dos nutrientes usados para essa análise. Para calcular a nota, foi tirada uma média entre os indicadores dos três nutrientes, sendo as receitas diet aquelas com as menores notas.

```
SELECT * FROM Receitas_Diet ORDER BY Nota LIMIT 10;
```

```
SELECT Food_Code, Food_Desc, Nivel_Gordura FROM Receitas_Diet ORDER BY Nivel_Gordura DESC LIMIT 10;
```

```
SELECT Food_Code, Food_Desc, Nivel_Acucar FROM Receitas_Diet ORDER BY Nivel_Acucar DESC LIMIT 10;
```

```
SELECT Food_Code, Food_Desc, Nivel_Sodio FROM Receitas_Diet ORDER BY Nivel_Sodio DESC LIMIT 10;
```


Receitas diet

Food_Code	Food_Desc	Nivel_Gordura	Nivel_Acucar	Nivel_Sodio	Nota
72201100	Broccoli, raw	0.056	0.093	0.06	0.069
75109550	Cilantro, raw	0.086	0.058	0.076	0.073
75100750	Artichoke, Jerusalem, raw	0.056	0.066	0.1	0.074
75119000	Parsley, raw	0.131	0.056	0.093	0.093
72122211	Mustard greens, cooked, from fresh, fat not added in cooking	0.075	0.094	0.243	0.137

Receitas com alta gordura

Food_Code	Food_Desc	Nivel_Gordura
42109100	Macadamia nuts, roasted	12.68
81301020	Lemon-butter sauce	12.333
83100200	Salad dressing, NFS, for sandwiches	10.700
81302010	Hollandaise sauce	9.663
83102000	Caesar dressing	9.641

Receitas com alto teor de açúcar

Food_Code	Food_Desc	Nivel_Acucar
92900110	Fruit-flavored beverage, dry concentrate, with sugar, not reconstituted	6.476
92900300	Fruit-flavored thirst quencher beverage, dry concentrate, not reconstituted	6.476
11825000	Whey, sweet, dry	4.963
12210430	Cream substitute, fat free, powder	4.893
12210440	Cream substitute, fat free, flavored, powder	4.893

Receitas com alto teor de sódio

Food_Code	Food_Desc	Nivel_Sodio
41205100	Black bean sauce	9.738
41420110	Miso (fermented soybean paste)	6.213
75236500	Yeast extract spread	5.633
23321900	Venison/deer jerky	4.925
21602000	Beef, dried, chipped, uncooked	4.65

Perguntas de Pesquisa/Análise Relacional

Pergunta/Análise 2

- Analisar o quão balanceada é uma receita com base na quantidade de nutrientes para os quais essa receita possui valores dentro da faixa recomendada, considerando todos os nutrientes do banco (proteínas, carboidratos, vitaminas, etc). Quais as receitas mais balanceadas de acordo com esse critério?
 - Foi criada uma view estendendo as receitas, contendo uma coluna para cada um dos nutrientes. Essas colunas possuem valores booleanos (1 caso a receita possua uma quantidade considerada aceitável para aquele nutriente, i.e., entre 5% e 15% do valor diário, e 0 do contrário). Em outras palavras, atribuímos 1 se:

$(\text{quantidade do nutriente}) < 0.15 * (\text{valor diário}) \text{ AND } (\text{quantidade do nutriente}) > 0.05 * (\text{valor diário}).$

Perguntas de Pesquisa/Análise Relacional

Pergunta/Análise 2

```
CREATE VIEW Receitas_Grau AS
  SELECT R.Food_Code, R.Food_Desc,
         CASE
           WHEN N.Energy < 0.15*Rec.Energy AND N.Energy > 0.05*Rec.Energy THEN 1
           ELSE 0
         END AS Energy_level,
         CASE
           WHEN N.Protein < 0.15*Rec.Protein AND N.Protein > 0.05*Rec.Protein THEN 1
           ELSE 0
         END AS Protein_level,
         ...

         CASE
           WHEN N.Sodium < 0.15*Rec.Sodium AND N.Sodium > 0.05*Rec.Sodium THEN 1
           ELSE 0
         END AS Sodium_level
  FROM FCID_Food_Code_Description R, FNDDS_Nutrient_Values N, Recommended_Nutritional_Values Rec
  WHERE R.Food_Code = N.Food_code
```

Perguntas de Pesquisa/Análise Relacional

Pergunta/Análise 2

Em seguida, criamos outra view contendo agora um atributo nota, dado pela soma das colunas individuais de cada ingrediente. Essa nota define o quão balanceada cada receita é.

```
CREATE VIEW Receitas_Balanceadas AS
  SELECT R.Food_Code, R.Food_Desc, RG.Energy_level + RG.Protein_level + RG.Carbohydrate_level
    + RG.Sugars_total_level + RG.Fiber_total_dietary_level + RG.Total_Fat_level + RG.Cholesterol_level +
    RG.Vitamin_A_RAE_level + RG.Vitamin_B6_level + RG.Vitamin_C_level + RG.Calcium_level + RG.Iron_level
    + RG.Potassium_level + RG.Sodium_level AS Nota
  FROM FCID_Food_Code_Description R, FNDDS_Nutrient_Values N, Receitas_Grau RG
  WHERE R.Food_Code = N.Food_code AND R.Food_Code = RG.Food_Code;

SELECT * FROM Receitas_Balanceadas ORDER BY Nota DESC LIMIT 10;
```

Receitas balanceadas


Food_Code	Food_Desc	Nota
71508030	White potato, stuffed, baked, peel eaten, stuffed with chili	13
52304010	Muffin, wheat bran	12
27446350	Oriental chicken or turkey garden salad (chicken and/or turkey, lettuce, fruit, nuts), no dressing	11
27446355	Oriental chicken or turkey garden salad with crispy noodles (chicken and/or turkey, lettuce, fruit, nuts)	11
58106755	Pizza with meat and fruit, regular crust	11
71507020	White potato, stuffed, baked, peel not eaten, stuffed with cheese	11
71507030	White potato, stuffed, baked, peel not eaten, stuffed with chili	11
71508020	White potato, stuffed, baked, peel eaten, stuffed with cheese	11
74203010	Tomatoes, NS as to form, scalloped	11
13135010	Ice cream sandwich, made with light chocolate ice cream	10

Perguntas de Pesquisa/Análise em grafos



Pergunta/Análise 3

Que receitas possuem o maior número relativo de ingredientes compartilhados

- Com a finalidade de analisar a inter-relação entre as diferentes receitas com base em sua composição atômica de ingredientes, possibilitando assim uma métrica de paridade e igualdade entre as receitas, fez-se necessário a utilização de queries para buscar as receitas procuradas com base no ingrediente em comum e em seguida juntar as receitas com uma aresta com peso relativo a semelhança entre ambas. As mais similares são logo então retornadas.
 - Desse modo é possível criar comunidades de receitas nas quais a ligação representa uma relação de equivalência, sendo possível avaliar receitas que vão certos conjuntos de ingredientes e iterativamente ir-se trocando um ou mais ingredientes da composição por outros alternativos, mas ainda assim mantendo as propriedades principais da receita original (ao menos até certo ponto)
- 

Perguntas de Pesquisa/Análise em grafos

Implementação

MATCH(r1:Recipe)

MATCH(r2:Recipe)

WHERE r1.Food_Code <> r2.Food_Code

MATCH(r1)-[:Recipe_Of_Ingredient]->(:Ingredient)-[:Ingredient_Of_Recipe]->(r2)

Merge (r1)<-[p:Parity]->(r2)

ON CREATE SET p.weight = 0

ON MATCH SET p.weight = p.weight+1

MATCH x = ()-[p:Parity]->()

WHERE p.weight>5

RETURN p

Perguntas de Pesquisa/Análise em grafos



Pergunta/Análise 4

- Quais os crop groups mais centrais, com base no número de receitas?
 - Foi necessário criar um atributo "Centrality" nos nós que representam os Crop Groups. Assim, sempre que encontrarmos uma Receita que possui um ingrediente de um determinado Crop Group, esse atributo é incrementado. Em seguida, podemos seleccionar apenas os nós de Crop Group com uma "Centrality" alta.



Perguntas de Pesquisa/Análise em grafos

Pergunta/Análise 4

No exemplo, estão apenas os nós onde a centralidade é maior que 1000.

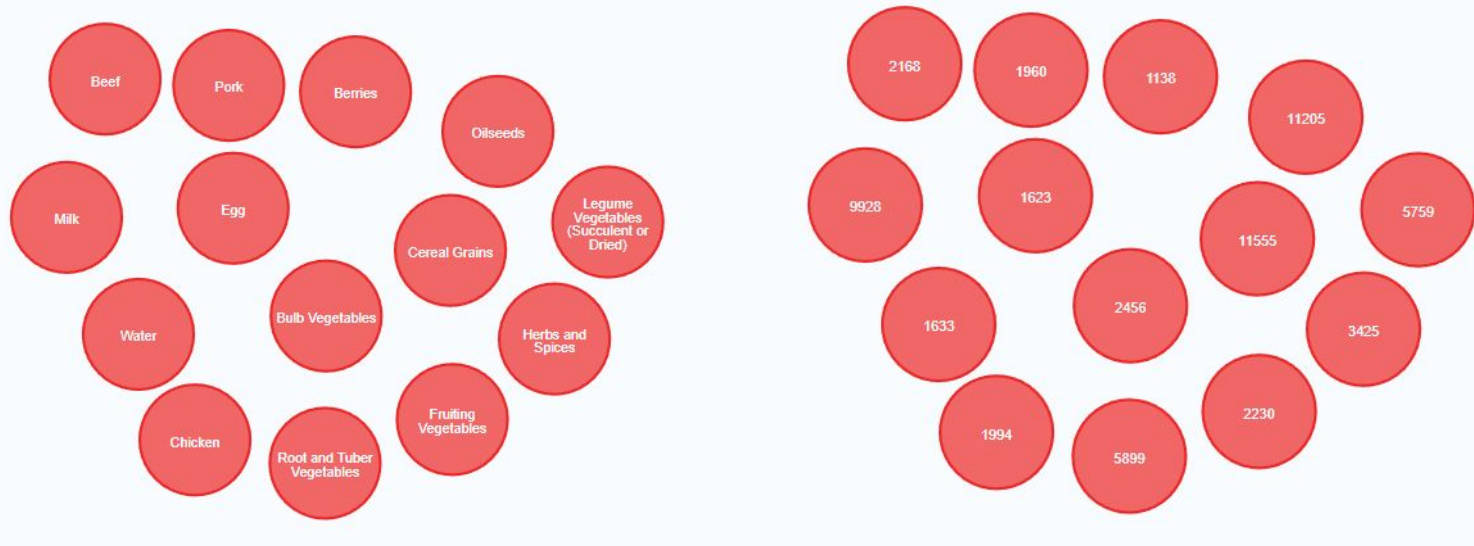
```
MATCH(g:Crop_Group)  
SET g.Centrality = 0
```

```
MATCH(r:Recipe)-[:Recipe_Of_Ingredient]->(i:Ingredient)-[:Food_Group]->(g:Crop_Group)  
MERGE(r)-[x:Has_Group]->(g)  
ON CREATE SET g.Centrality=g.Centrality+1  
ON MATCH SET g.Centrality=g.Centrality+1
```

```
MATCH(g:Crop_Group)  
WHERE g.Centrality>1000  
RETURN g
```

Perguntas de Pesquisa/Análise em grafos

Pergunta/Análise 4



Agradecemos sua atenção!

Equipe VIRUS

Subgrupos BOYS e GATOS

