Kinda SUS

Etapas 5 - Análise Final
MC 536 - 2020s2

Felipe Hideki - 196767
Julio Kiyoshi - 200483
Matheus Fernandes - 222228

Introdução

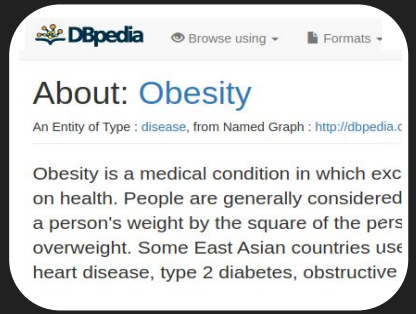
O conjunto de diversos fatores, como o consumo de alimentos altamente calóricos e processados e uma rotina apertada e sedentária tornou a obesidade um dos grandes problemas enfrentados pelos EUA.

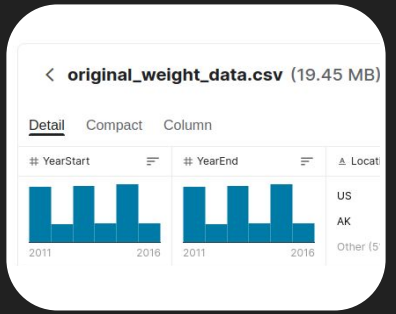
Devido a sua relevância, escolhemos essa questão como tema do nosso projeto, onde serão utilizados diversos dados para entender melhor as causas da obesidade e sua relevância para a incidência de doenças cardíacas.

Cronologia da construção das análises e dos bancos de dados

Etapa 2

- Seleção do tema.
- Seleção das bases de dados.
- Inicialmente, 4 bases foram selecionadas, uma em grafo, uma em json e duas em tabela (csv e xls).





Etapa 3
Primeiros Modelos e Análises

Etapa 3

- Nesta etapa, as bases de dados “Heart Disease Mortality” e “Obesity Stats” foram filtradas e integradas para a criação de uma nova base, a qual foi nomeada “Análise Obesidade”.
- Nesta nova base, valores de obesidade e mortalidade por doenças do coração nos EUA para homens e mulheres de cada Estado são expostos, de modo a tentar visualizar uma correlação entre esses dois fatores.

Etapa 3

- Extração de dados feita usando python em notebooks disponíveis no github.
- Análise dos dados usando a biblioteca SQLite no Python, também em um notebook.
- URLs:
  analiseObesidade.ipynb
  extracaoHeartDisease.ipynb
  extracaoObesityStats.ipynb

Etapa 4
Segundo Modelo e Análise

Etapa 4

- Nesta etapa, as bases de dados “Prevalence of obesity among adults” e “Global Dietary Database” foram filtradas e integradas em uma nova base, a qual foi nomeada de “Obesidade e Nutrição Global”.
- Esta nova base é um grafo que foi criado usando a linguagem Cypher no Sandbox do Neo4j.
- Foram, então, relacionados dados entre ingestão nutricional e porcentagem de obesidade para cada país.

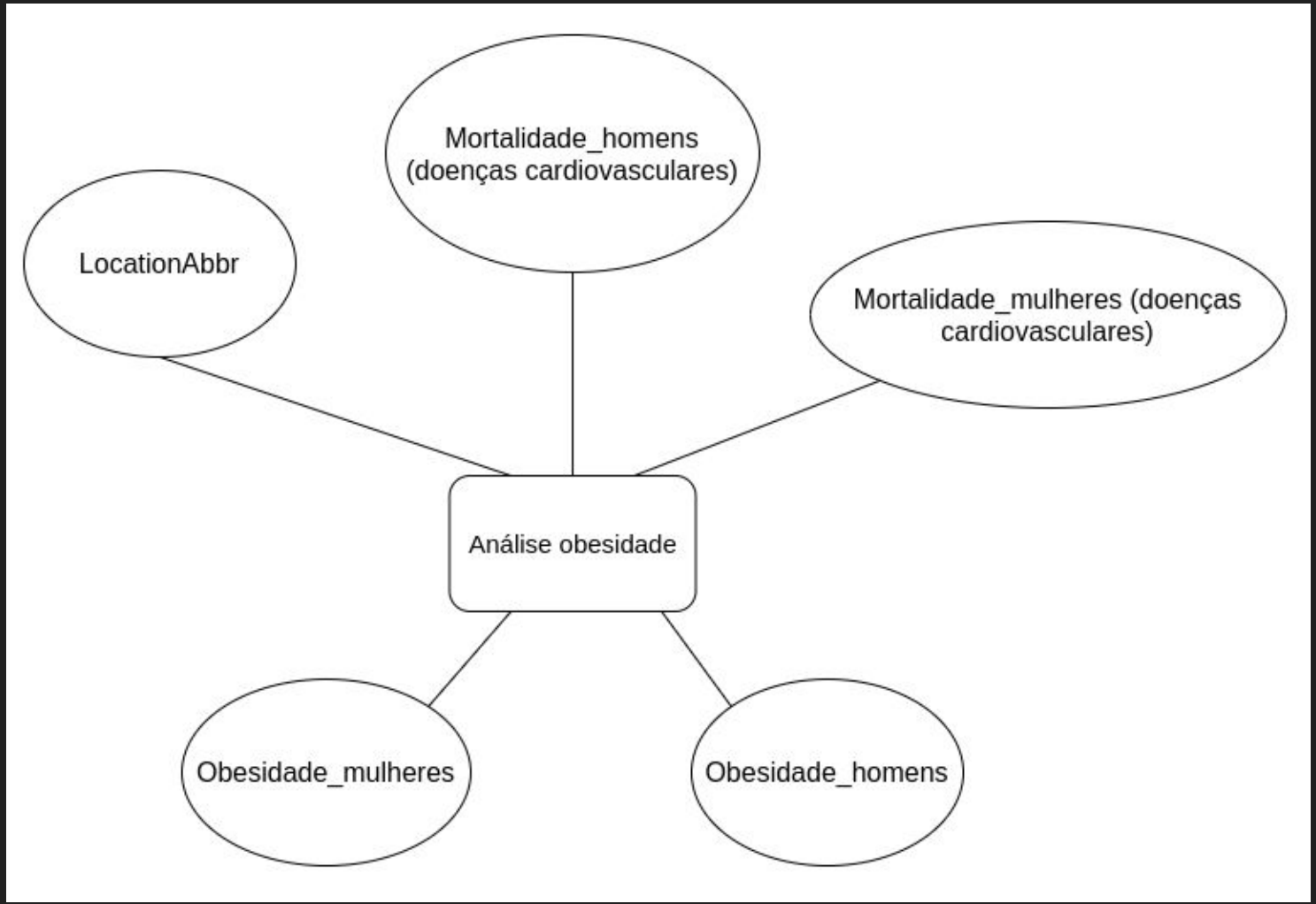
Etapa 4

- Extração de dados feita usando python em notebooks disponíveis no github.
- O dataset da etapa foi um grafo cujas queries foram feitas usando Neo4j.
- URLs:

  queries.md
  extracaoGlobalObesity.ipynb
  extracaoConsumoAlimentos.ipynb

Etapa 5 - Análise Final

Modelos Conceituais


Mortalidade_homens
(doenças cardiovasculares)
LocationAbbr
Mortalidade_mulheres (doenças
cardiovasculares)
Análise obesidade
Obesidade_mulheres
Obesidade_homens

Modelos Lógicos

- Análise obesidade(LocationAbbr, Mortalidade_homens, Mortalidade_mulheres, Obesidade_homens, Obesidade_mulheres)

Obesidade nos EUA

- Nesta etapa, as bases de dados “Heart Disease Mortality” e “Obesity Stats” foram filtradas e integradas para a criação de uma nova base, a qual foi nomeada “Análise Obesidade”.
- Nesta nova base, valores de obesidade e mortalidade por doenças do coração nos EUA para homens e mulheres de cada Estado são expostos, de modo a tentar visualizar uma correlação entre esses dois fatores.

Perguntas

- Existe diferença considerável nas taxas de obesidade, mortalidade e sedentarismo entre homens e mulheres?
- Os estados com os maiores níveis de obesidade são os também os com maiores níveis de mortalidade por doença cardíaca e sedentarismo?
- Existe correlação visível entre obesidade e mortalidade por doença cardíaca?

Queries

Query 11: Médias de Obesidade, Mortalidade por doença cardíaca e sendetarismo entre homens e mulheres.

```
query11 = pd.read_sql("""
  SELECT query1.LocationAbbr, query1.Mortalidade_homens, query1.Mortalidade_mulheres, query1.Obesidade_homens, query1.Obesidade_mulheres, query9.Average Homens_seden
  FROM query1, query9, query10
  WHERE query9.LocationAbbr = query1.LocationAbbr AND query10.LocationAbbr = query1.LocationAbbr
  GROUP BY query1.LocationAbbr
  ORDER BY query1.LocationAbbr
""", conn)
c.execute("DROP TABLE IF EXISTS query11")
query11.to_sql('query11', conn)
query11
```

	LocationAbbr	Mortalidade_homens	Mortalidade_mulheres	Obesidade_homens	Obesidade_mulheres	Homens_sedentarios	Mulheres_sedentarias
0	AK	350.9	196.7	29.2	27.5	22.8	21.7
1	AL	540.9	352.6	30.9	34.0	26.9	35.6
2	AR	518.5	334.1	35.2	34.1	32.2	36.4
3	AZ	344.9	211.5	26.5	27.1	23.3	27.0
4	CA	499.9	227.5	25.1	23.1	21.0	21.8
5	CO	309.4	197.1	20.8	21.8	17.3	18.6
6	CT	369.4	230.8	25.5	24.4	23.2	26.5
7	DE	401.8	265.8	30.6	31.5	25.0	30.3
8	FL	373.0	227.0	27.5	25.3	25.3	30.0

Query 15: Médias Gerais de Obesidade, Mortaliadade por doenças cardíacas e Sedentarismo para homens e mulheres.

```
query15 = pd.read_sql("""
  SELECT AVG(query2.Average) Obesidade_homens, AVG(query3.Average) Obesidade_mulheres, AVG(query2.Data_Value) Mortalidade_homens, AVG(query3.Data_Value) Mortalidade_mulheres, AVG(query9.A
  FROM query2, query3, query9, query10
  GROUP BY query2.Gender, query3.Gender, query9.Gender, query10.Gender
""", conn)
c.execute("DROP TABLE IF EXISTS query5")
query15.to_sql('query15', conn)
query15
```

	Obesidade_homens	Obesidade_mulheres	Mortalidade_homens	Mortalidade_mulheres	Homens_sedentarios	Mulheres_sedentarias
0	29.043137	28.458824	416.398039	259.943137	24.835294	27.464706

Query 6: Estados onde as Médias da Obesidade é maior do que a Média Geral.

```
[ ] query6 = pd.read_sql("""
      SELECT query2.LocationAbbr, query2.Average Obesidade_homens, query3.Average Obesidade_mulheres, query2.Data_Value Mortalidade_homens, query3.Data_Value Mortalidade
      FROM query2, query3, query5
      WHERE query2.LocationAbbr = query3.LocationAbbr AND query2.Average > query5.Obesidade_homens AND query3.Average > query5.Obesidade_mulheres
      ORDER BY query2.LocationAbbr
    """, conn)
    c.execute("DROP TABLE IF EXISTS query6")
    query6.to_sql('query6', conn)
    query6
```

	LocationAbbr	Obesidade_homens	Obesidade_mulheres	Mortalidade_homens	Mortalidade_mulheres
0	AL	30.9	34.0	540.9	352.6
1	AR	35.2	34.1	518.5	334.1
2	DE	30.6	31.5	401.8	265.8
3	GA	29.2	31.5	434.3	355.5
4	IA	32.4	30.1	410.4	251.7
5	IN	31.1	32.6	456.5	284.6
6	KS	30.5	29.4	391.2	236.5
7	KY	33.3	33.1	500.1	310.4
8	LA	31.1	35.0	522.2	330.4

Query 7: Estados onde a Mortalidade por doenças do coração é maior do que a média geral.

```
[ ]  query7 = pd.read_sql("""
       SELECT query2.LocationAbbr, query2.Average Obesidade_homens, query3.Average Obesidade_mulheres, query2.Data_Value Mortalidade_homens, query3.Data_Value Mortalidade
       FROM query2, query3, query5
       WHERE query2.LocationAbbr = query3.LocationAbbr AND query2.Data_Value > query5.Mortalidade_homens AND query3.Data_Value > query5.Mortalidade_mulheres
       ORDER BY query2.LocationAbbr
     """, conn)
     c.execute("DROP TABLE IF EXISTS query7")
     query7.to_sql('query7', conn)
     query7
```

	LocationAbbr	Obesidade_homens	Obesidade_mulheres	Mortalidade_homens	Mortalidade_mulheres
0	AL	30.9	34.0	540.9	352.6
1	AR	35.2	34.1	518.5	334.1
2	GA	29.2	31.5	434.3	355.5
3	IL	28.3	30.5	423.3	261.0
4	IN	31.1	32.6	456.5	284.6
5	KY	33.3	33.1	500.1	310.4
6	LA	31.1	35.0	522.2	330.4
7	MD	27.8	28.8	420.8	264.7
8	MI	31.1	31.9	482.1	311.3
9	MO	30.7	30.2	473.3	298.7
10	MS	33.8	36.4	564.3	361.9

Query 13: Estados em que as Médias de Sedentarismo são maiores do que as médias gerais.

```
query13 = pd.read_sql("""
  SELECT query11.LocationAbbr, query11.Obesidade_homens, query11.Obesidade_mulheres, query11.Mortalidade_homens, query11.Mortalidade_mulheres, query11.Homens_sedentarios, query11.Mulheres_
  FROM query11, query12
  WHERE  query11.Homens_sedentarios > query12.Homens_sedentarios AND query11.Mulheres_sedentarias > query12.Mulheres_sedentarias
  ORDER BY query11.LocationAbbr
""", conn)
c.execute("DROP TABLE IF EXISTS query13")
query13.to_sql('query13', conn)
query13
```

	LocationAbbr	Obesidade_homens	Obesidade_mulheres	Mortalidade_homens	Mortalidade_mulheres	Homens_sedentarios	Mulheres_sedentarias
0	AL	30.9	34.0	540.9	352.6	26.9	35.6
1	AR	35.2	34.1	518.5	334.1	32.2	36.4
2	DE	30.6	31.5	401.8	265.8	25.0	30.3
3	FL	27.5	25.3	373.0	227.0	25.3	30.0
4	IA	32.4	30.1	410.4	251.7	28.8	28.2
5	IN	31.1	32.6	456.5	284.6	28.8	33.1
6	KY	33.3	33.1	500.1	310.4	28.6	31.7
7	LA	31.1	35.0	522.2	330.4	28.7	35.5
8	MO	30.7	30.2	473.3	298.7	27.9	28.6
9	MS	33.8	36.4	564.3	361.9	33.5	42.3
10	NC	28.6	30.2	404.6	248.6	25.1	28.0
11	ND	33.7	28.0	361.2	231.6	27.5	27.7
12	NY	24.7	26.1	442.0	291.0	24.9	28.4
13	OH	30.0	30.8	459.2	288.4	27.2	29.6
14	OK	32.9	32.2	537.4	355.7	31.8	34.1

Query 14: Interseção entre os estados que possuem taxas de obesidade, mortalidade e sedentarismo acima da média.

```
query14 = pd.read_sql("""
  SELECT query13.LocationAbbr, query13.Obesidade_homens, query13.Obesidade_mulheres, query13.Mortalidade_homens, query13.Mortalidade_mulheres, query13.Homens_sedentarios, query13.Mulheres_
  FROM query8, query13
  WHERE  query13.LocationAbbr = query8.LocationAbbr
  ORDER BY query13.LocationAbbr
""", conn)
c.execute("DROP TABLE IF EXISTS query14")
query14.to_sql('query14', conn)
query14
```

	LocationAbbr	Obesidade_homens	Obesidade_mulheres	Mortalidade_homens	Mortalidade_mulheres	Homens_sedentarios	Mulheres_sedentarias
0	AL	30.9	34.0	540.9	352.6	26.9	35.6
1	AR	35.2	34.1	518.5	334.1	32.2	36.4
2	IN	31.1	32.6	456.5	284.6	28.8	33.1
3	KY	33.3	33.1	500.1	310.4	28.6	31.7
4	LA	31.1	35.0	522.2	330.4	28.7	35.5
5	MO	30.7	30.2	473.3	298.7	27.9	28.6
6	MS	33.8	36.4	564.3	361.9	33.5	42.3
7	OH	30.0	30.8	459.2	288.4	27.2	29.6
8	OK	32.9	32.2	537.4	355.7	31.8	34.1
9	TN	32.8	34.6	495.8	315.8	34.7	39.5
10	WV	35.4	34.9	469.9	311.2	29.4	33.4

Obesidade:
Mundo vs. EUA

- Nesta etapa, as bases de dados “Prevalence of obesity among adults” e “Global Dietary Database” foram filtradas e integradas em uma nova base, a qual foi nomeada de “Obesidade e Nutrição Global”.
- Esta nova base é um grafo que foi criado usando a linguagem Cypher no Sandbox do Neo4j.
- Foram, então, relacionados dados entre ingestão nutricional e porcentagem de obesidade para cada país.

Perguntas

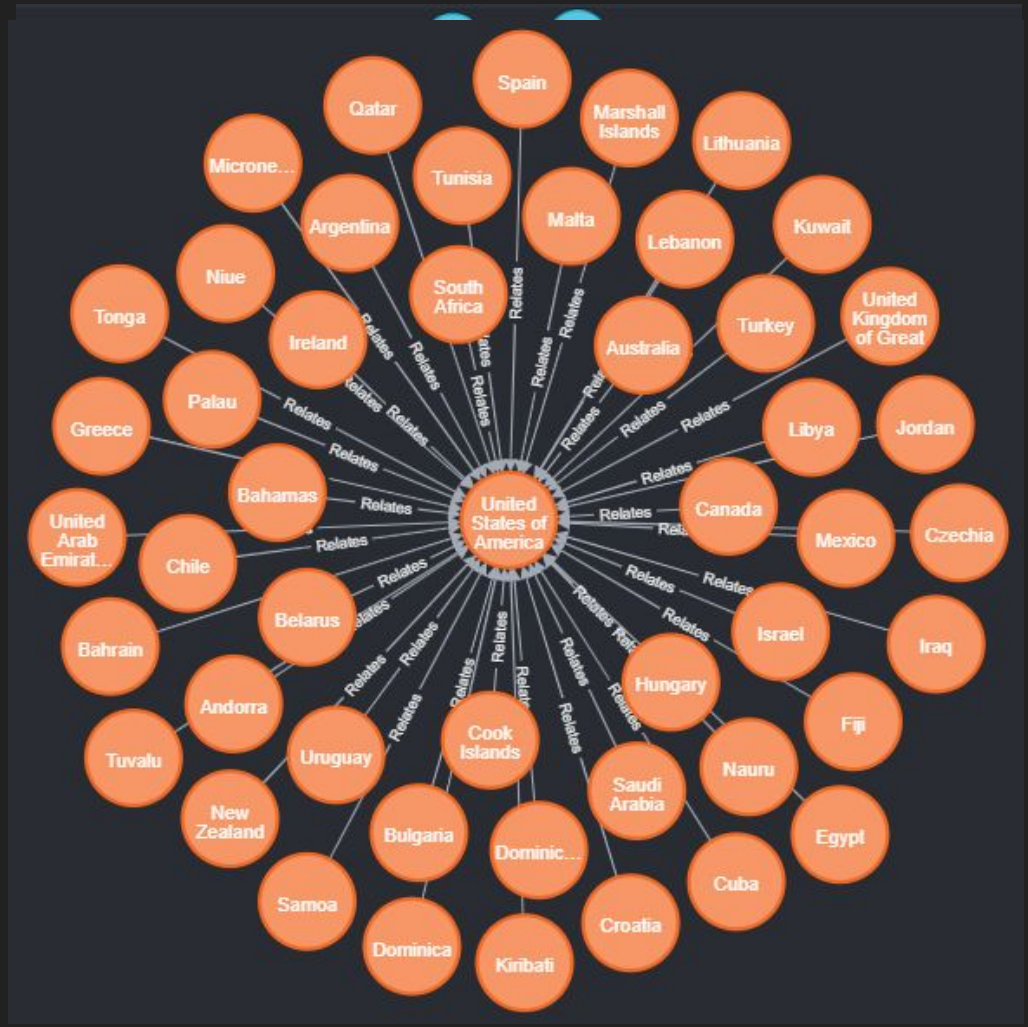
- Como se dá a relação entre os EUA e outros países ao comparar o consumo de alimentos (tipo e quantidade) e as taxas de obesidade?

- Como se dá a organização global nesse mesmo aspecto?

Queries

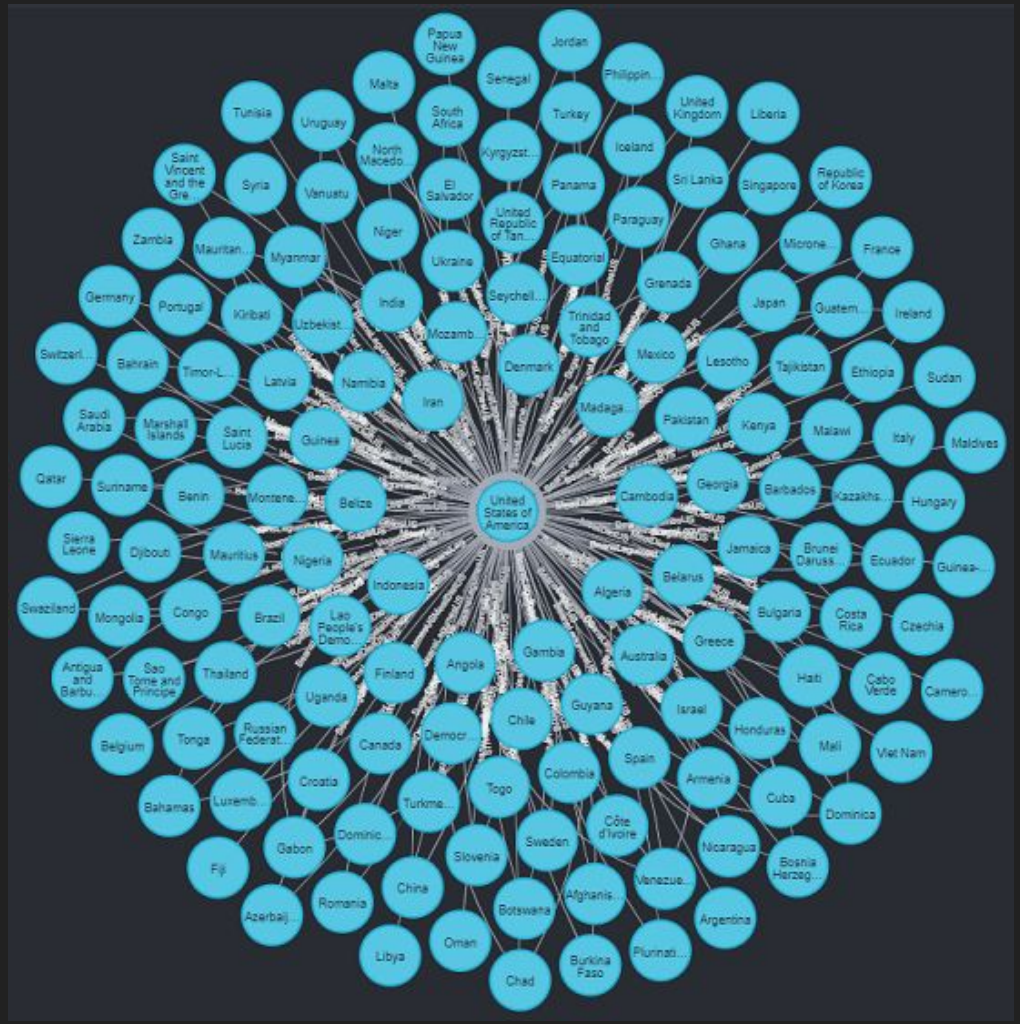
Grupo 1

Nesse primeiro conjunto de queries são agrupados países que possuem uma taxa de obesidade semelhante aos EUA, país alvo da nossa análise, e ao Japão, um dos países com menor taxa de obesidade.



Grupo 2

Aqui foram agrupados os países que possuem um consumo alimentos semelhantes aos EUA, com o objetivo de, em conjunto com as queries do grupo 1, visualizar a possível correlação entre os dois grupos.

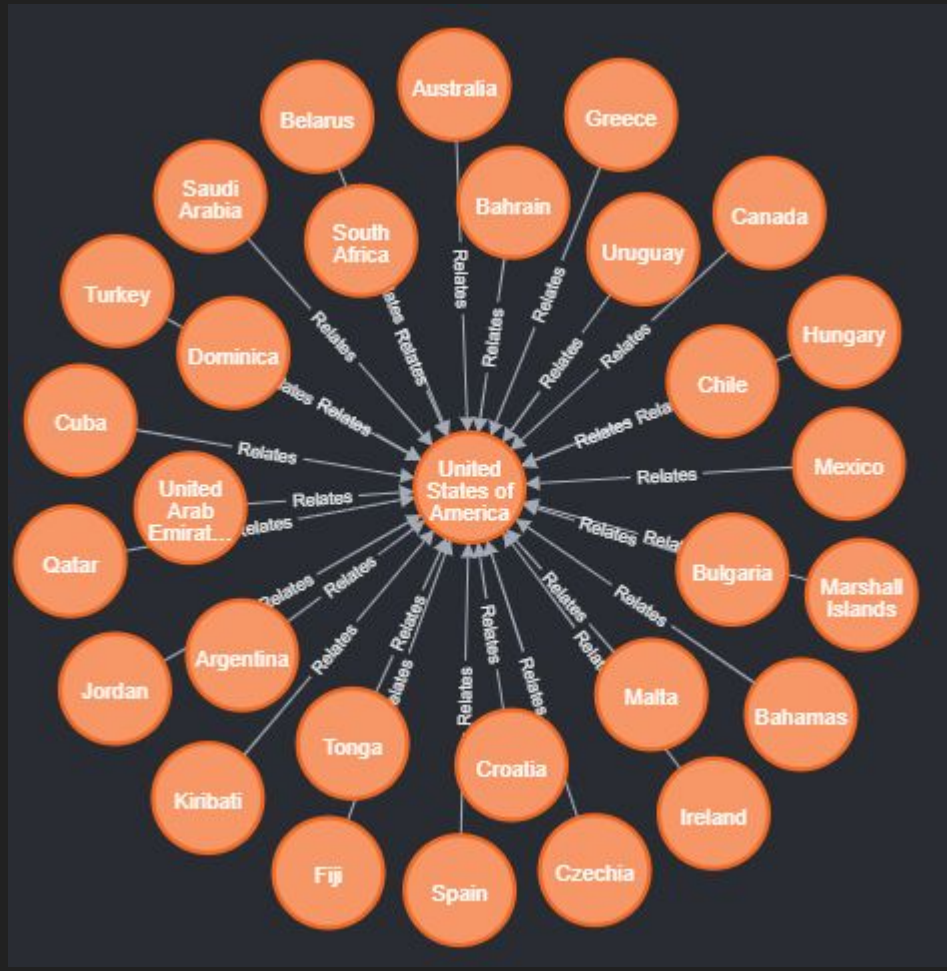


Grupo 2

Países com consumo de alimentos não considerados contribuintes para a obesidade semelhante aos EUA.

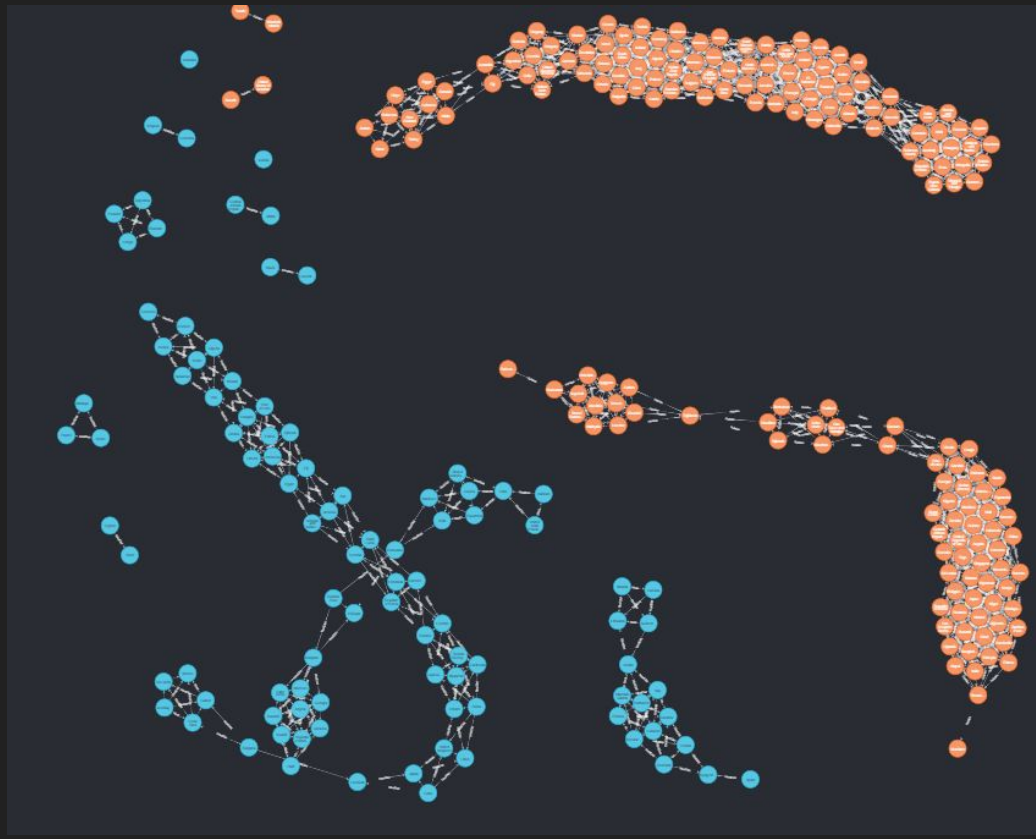
Grupo 2

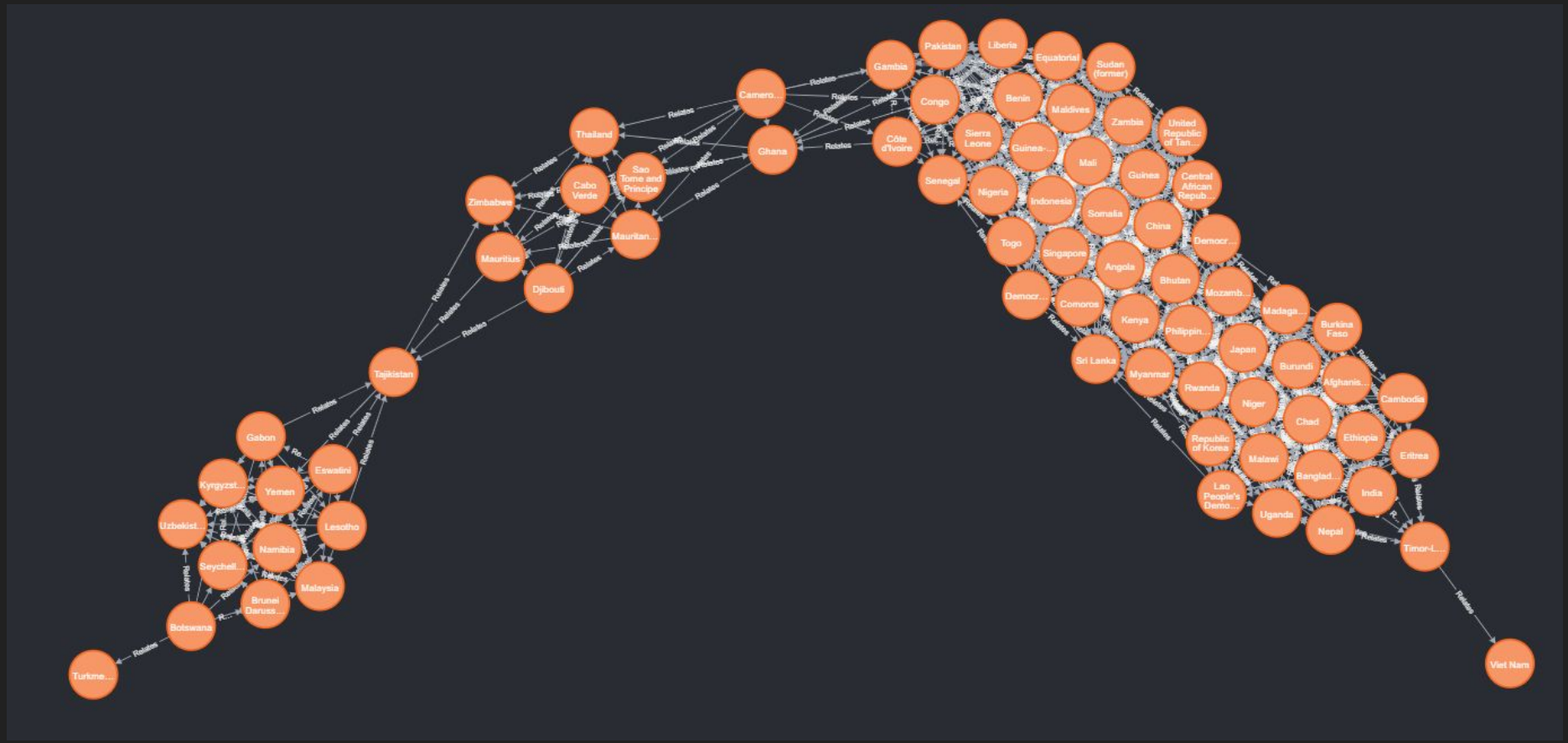
Interseção entre os países com taxa de obesidade e consumo de alimentos considerados contribuintes para a obesidade semelhantes aos EUA.

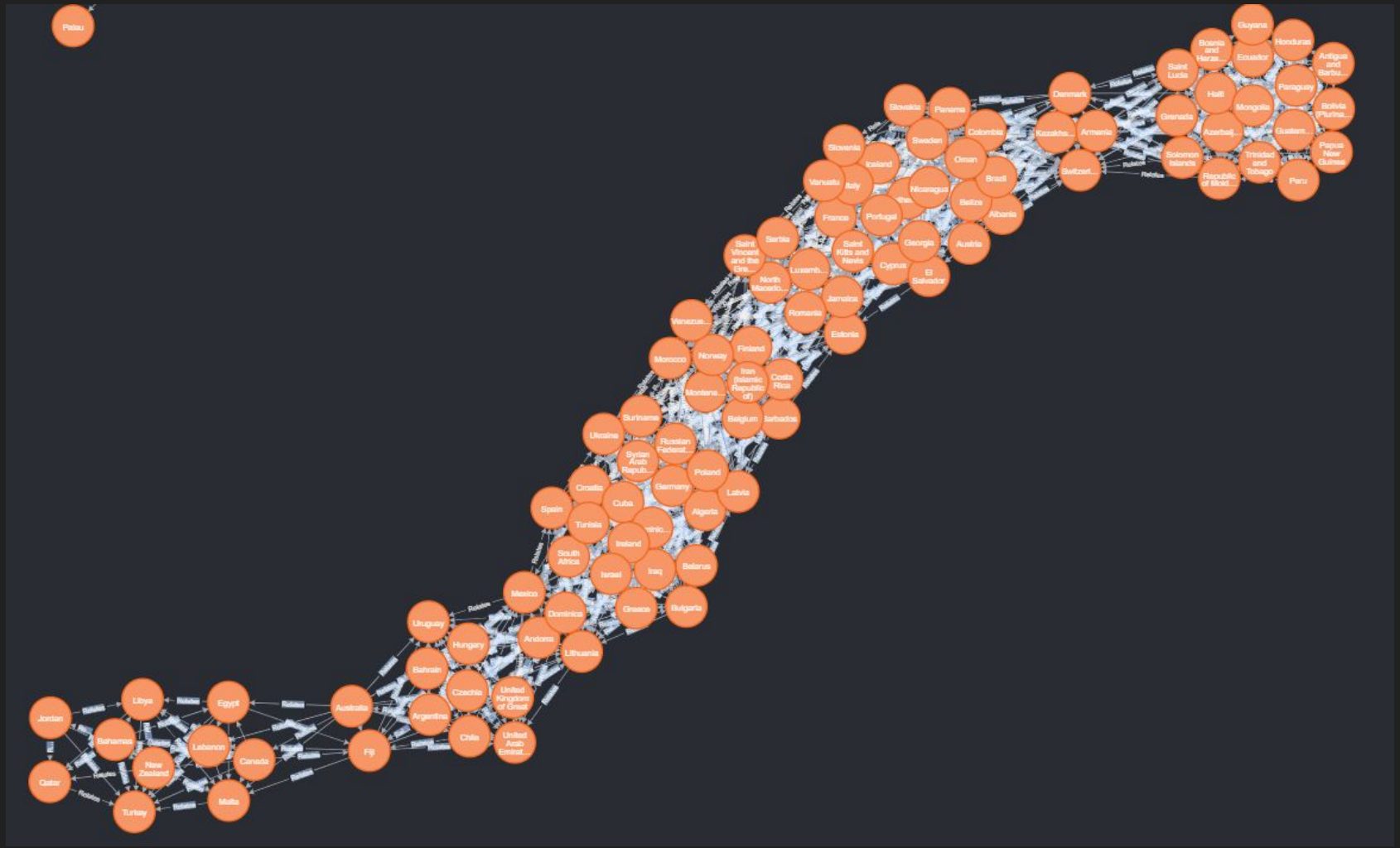


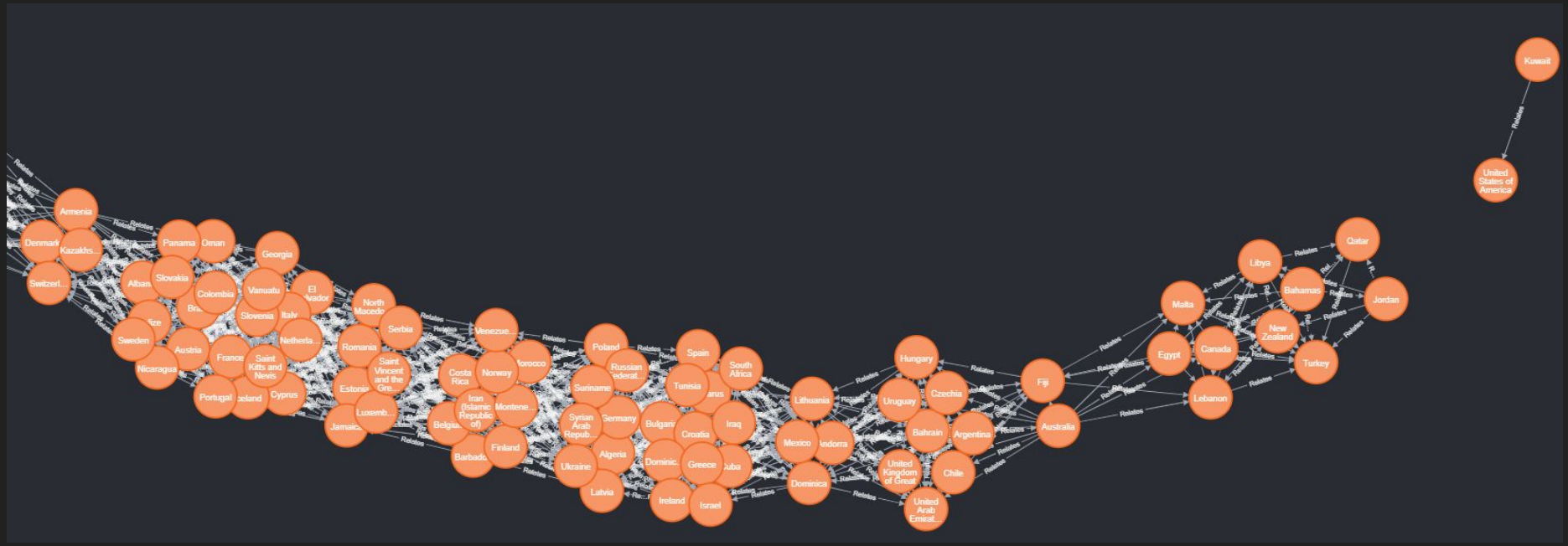
Grupo 3

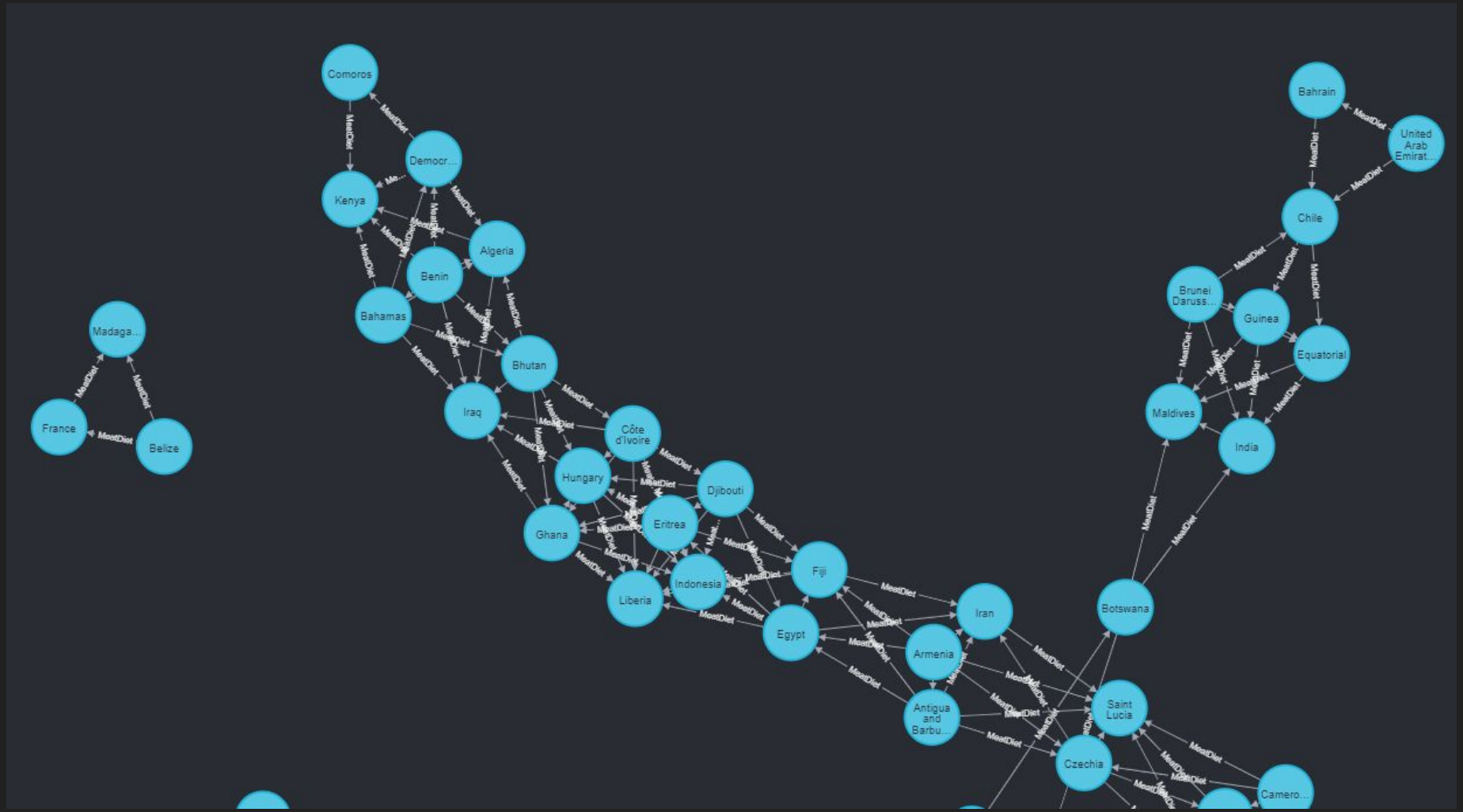
Neste grupo de queries foram criados um grafo que relaciona os países que possuem taxa de obesidade semelhante entre si e outro que agrupa os países com consumo de carne vermelha parecido. Espera-se criar com isso uma visualização mais geral entre esses grupos.

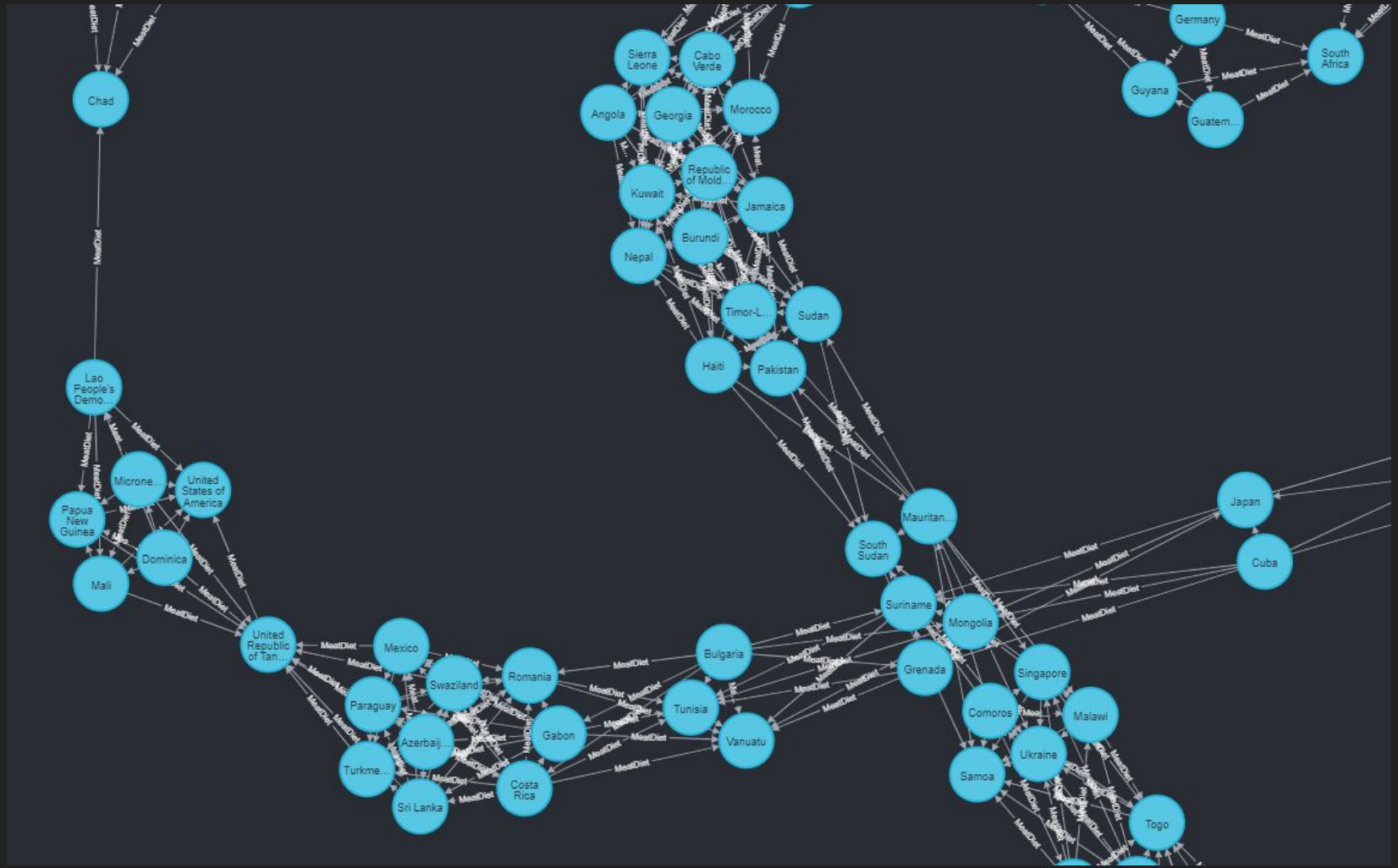



Pakistan
Liberia
Equatorial
Sudan (former)
Gambia
Camero...
Congo
Benin
Maldives
Zambia
United Republic of Tan...
Thailand
Ghana
Côte d'Ivoire
Sierra Leone
Guinea-...
Mali
Guinea
Central African Repub...
Zimbabwe
Cabo Verde
Sao Tome and Principe
Senegal
Nigeria
Indonesia
Somalia
China
Democr...
Mauritan...
Mauritius
Djibouti
Togo
Singapore
Angola
Bhutan
Mozamb...
Democr...
Comoros
Kenya
Philippin...
Madaga...
Burkina Faso
Sri Lanka
Myanmar
Japan
Burundi
Afghanis...
Cambodia
Rwanda
Niger
Chad
Ethiopia
Tajikistan
Gabon
Republic of Korea
Malawi
Eritrea
Eswatini
Kyrgyzst...
Yemen
Lao People's Demo...
Banglad...
India
Uzbekist...
Lesotho
Uganda
Nepal
Timor-L...
Namibia
Seychell...
Malaysia
Brunei Darussa...
Botswana
Turkme...
Viet Nam
Relates


Palau
Guyana
Honduras
Bosnia and Herze...
Ecuador
Antigua and Barbu...
Saint Lucia
Haiti
Paraguay
Denmark
Mongolia
Bolivia Plurina...
Slovakia
Panama
Grenada
Azerbaij...
Guatem...
Colombia
Kazakhs...
Armenia
Slovenia
Sweden
Solomon Islands
Trinidad and Tobago
Papua New Guinea
Iceland
Oman
Switzerl...
Republic of Moldi...
Peru
Vanuatu
Brazil
Nicaragua
Belize
Albania
France
Portugal
Serbia
Georgia
Austria
Saint Kitts and Nevis
Cyprus
El Salvador
Saint Vincent and the Gre...
Luxemb...
North Macedo...
Jamaica
Venezue...
Romania
Estonia
Finland
Morocco
Norway
Iran (Islamic Republic of)
Costa Rica
Montene...
Barbados
Belgium
Suriname
Ukraine
Russian Federat...
Syrian Arab Republ...
Poland
Croatia
Germany
Latvia
Spain
Cuba
Algeria
Tunisia
Ireland
South Africa
Israel
Iraq
Belarus
Mexico
Dominica
Greece
Bulgaria
Uruguay
Andorra
Hungary
Lithuania
Bahrain
Czechia
United Kingdom of Great
Argentina
Chile
United Arab Emirat...
Australia
Fiji
Libya
Egypt
Jordan
Bahamas
Lebanon
Canada
New Zealand
Qatar
Malta
Turkey


Kuwait
Relates
United States of America
Armenia
Denmark
Kazakhs...
Switzerl...
Albani...
Slovakia
Panama
Oman
Georgia
Colombia
Vanuatu
El Salvador
Slovenia
Italy
Netherla...
Sweden
Austria
France
Saint Kitts and Nevis
Nicaragua
Portugal
Iceland
Cyprus
Estonia
Luxemb...
Jamaica
Romania
North Macedo...
Serbia
Saint Vincent and the Gre...
Belgium
Barbados
Costa Rica
Norway
Morocco
Iran (Islamic Republic of)
Montene...
Finland
Venezue...
Poland
Russian Federat...
Suriname
Syrian Arab Repub...
Germany
Ukraine
Algeria
Latvia
Spain
South Africa
Tunisia
Belarus
Bulgaria
Croatia
Iraq
Dominic...
Greece
Cuba
Ireland
Israel
Lithuania
Mexico
Andorra
Dominica
Hungary
Uruguay
Czechia
Bahrain
Argentina
United Kingdom of Great...
Chile
United Arab Emirat...
Fiji
Australia
Malta
Egypt
Canada
Lebanon
Libya
Bahamas...
New Zealand
Turkey
Qatar
Jordan
Relates


Comoros
Democr...
Kenya
Algeria
Benin
Bahamas
Bhutan
Iraq
Côte d'Ivoire
Hungary
Djibouti
Eritrea
Ghana
Liberia
Indonesia
Fiji
Egypt
Iran
Armenia
Antigua and Barbu...
Saint Lucia
Czechia
Botswana
Maldives
India
Brunei Daruss...
Guinea
Equatorial
Chile
Bahrain
United Arab Emirat...
Madaga...
France
Belize
MeatDiet


Chad
Sierra Leone
Cabo Verde
Angola
Georgia
Morocco
Germany
South Africa
Guyana
Guatem...
Republic of Mold...
Kuwait
Jamaica
Burundi
Nepal
Timor-L...
Sudan
Haiti
Pakistan
Lao People's Demo...
Microne...
United States of America
Papua New Guinea
Dominica
Mali
Mauritan...
South Sudan
Japan
Cuba
Surname
Mongolia
Grenada
Singapore
Comoros
Malawi
Ukraine
Samoa
Togo
United Republic of Tan...
Mexico
Swaziland
Paraguay
Azerbaij...
Turkme...
Sri Lanka
Costa Rica
Gabon
Romania
Bulgaria
Tunisia
Vanuatu
MeatDiet

Grupo 4

Para este grupo foi aplicado o conceito de community para agrupar os países com taxa de obesidade semelhante, baseado nas relações criadas no grupo 3.

Ao lado se tem o grafo resultante, feito no software Gephi, com nós coloridos a partir da comunidade a que pertencem e de tamanho proporcional à sua taxa de obesidade.