



Felicidade: fatores socioeconômicos

Grupo PsychHealth



Integrantes

- Christian Massao Konishi, 214570
- Gabriel de Freitas Garcia, 216179
- Andrey Torres de Lima, 231442



Aviso

- Nenhum integrante é psicólogo ou tem formação em alguma área relacionada.
- Iremos analisar padrões, comportamentos, correlações, mas não podemos estabelecer conclusões sobre a causalidade dos fatos.



Motivação

- Queremos ter uma visão global de como os países ditos mais felizes se comportam.
- O que eles têm em comum?
- Que fatores costumam estar mais associados à felicidade da população?
- Podemos aprender algo com isso?



Nossas bases de dados

- World Happiness Report
- Demographic and socio-economic
- International Labour Organization (ILOSTAT)



World Happiness Report (WHR)

- Relatório anual sobre a felicidade dos países.
- Além disso, o relatório correlaciona alguns indicadores por conta própria, disponibilizando esses dados.
- A felicidade é definida por uma pontuação, obtida através de amostras da população e a percepção do “quão boa é sua vida”.
- Mede a percepção do próprio povo em relação à sua vida.
- Valores podem ser obtidos em csv's.



Unesco

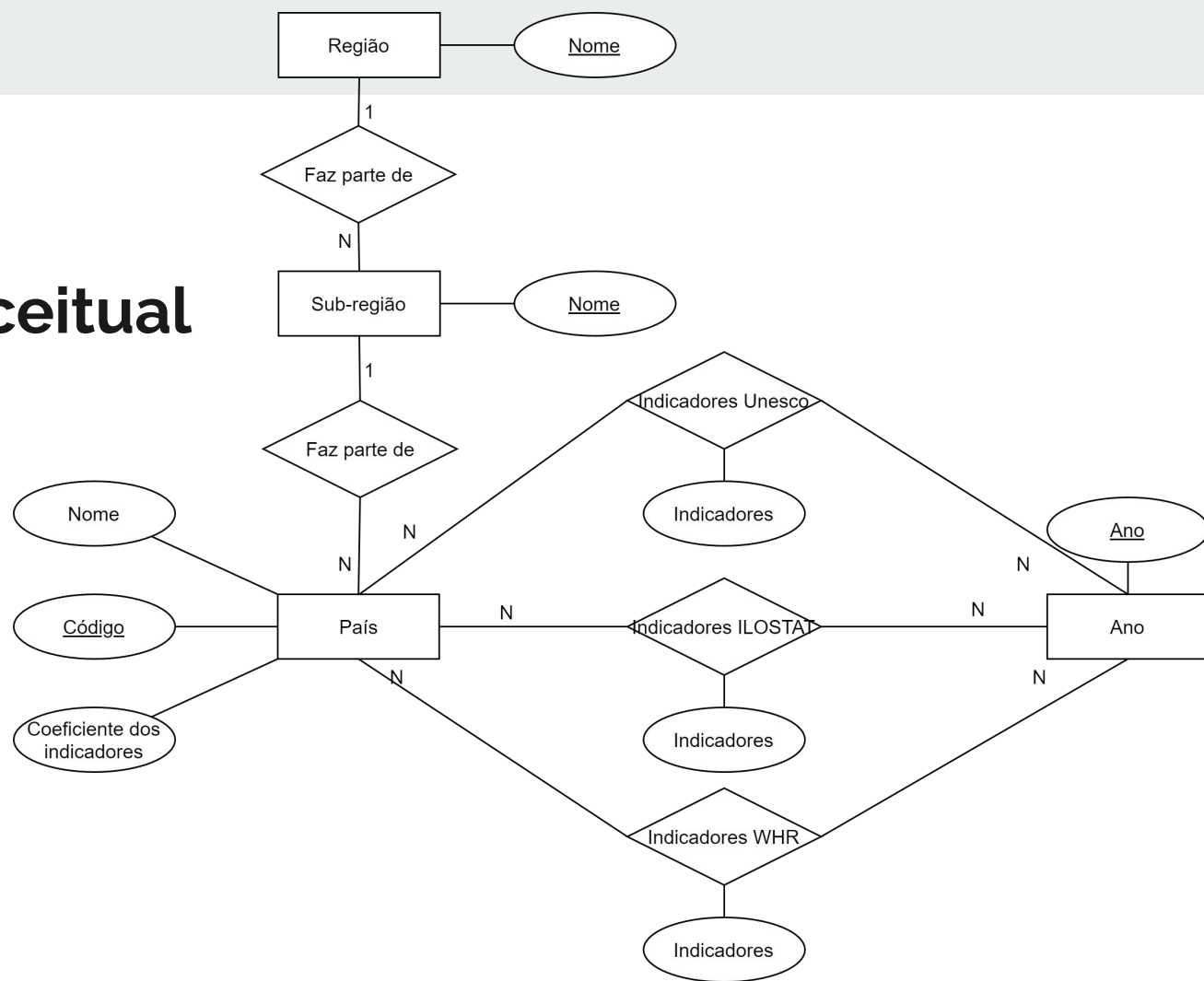
- Base de dados diversificada que fornece dados de diversos países
- O relatório nos traz informações relevantes sobre o estilo de vida da população, como índice de população rural, expectativa de vida, PIB, entre outros dados impactam o dia-dia da sociedade
- Dados gerais sobre os países sem muito foco no pessoal.
- Os dados podem ser filtrados também pelo ano que foram coletados
- Valores podem ser obtidos e filtrados para arquivos csv's, mas originalmente estavam em xml.



ILOSTAT

- Relatório com dados variados de diversos países.
- Relatórios que explicitam dores da população como o índice de desemprego, trabalho infantil, pobreza e fatores comuns aos países.
- Tem uma visão geral do povo da nação e não tão pessoal.
- Os dados podem ser filtrados também pelo ano que foram coletados
- Valores podem ser obtidos e filtrados para arquivos csv's.

Modelo Conceitual



Modelo Lógico

País

<u>Código alpha-3</u>	Código alpha-2	Nome	Coeficientes dos indicadores	Faz parte de (sub-região)
BRA	BR	Brazil	0.756 (valor ilustrativo)	Latin America and the Caribbean
...

Sub-região

<u>Nome</u>	Faz parte de (região)
Latin America and the Caribbean	Americas
...	...

Região

<u>Nome</u>
Americas
...

Indicadores Unesco

<u>Código país</u>	<u>Ano</u>	Indicador 1	Indicador 2	...
BRA	2015	0.56	0.87	...
...

Descrição Unesco

código indicador	descrição
DEMO_H1	Demographic
...	...

Indicadores ILOSTAT

<u>Código país</u>	<u>Ano</u>	Indicador 1	Indicador 2	...
BRA	2015	0.56	0.87	...
...

Indicadores WHR

<u>Código país</u>	<u>Ano</u>	Indicador 1	Indicador 2	...
BRA	2015	0.56	0.87	...
...



Processamento e integração das bases

- Os dados foram retirados das bases originais e processados usando Python.
- Por meio de Scripts Python foram selecionados os dados de interesse e também tornamos os dados combináveis entre as bases de dados, optando por usar tabelas CSV como representação intermediária. Uma das bases possuía a representação dos países totalmente destoantes das outras, sendo necessário até mesmo manipulá-la manualmente.
- Utilizando-se dessas tabelas fizemos queries SQL para combinar as tabelas e utilizá-las para análise. Existia uma tabela para cada ano e foi necessário combiná-las para se adequarem ao modelo lógico pré estabelecido.



Análises realizadas

- Primeiro foi observado se existe alguma relação entre os indicadores sócio-econômicos e índice de felicidade, notadamente, alguns dos índices analisados mostraram alguma espécie de relação. Observe na tabela abaixo:

Indicador	Países com felicidade acima da média	Países com este indicador e felicidade acima da média	Relação
Pib per capita	50,34%	76,19%	Direta
Expectativa de vida	50,00%	74,41%	Direta
População rural	48,95%	24,29%	Inversa



Query

```
CREATE VIEW HAPPYRURAL AS
SELECT W.COUNTRYCODE, W.ANO, W.HAPPINESSSCORE, U.SP_RUR_TOTL_ZS AS PROPRURAL
FROM WHR W, UNESCO U
WHERE
    W.COUNTRYCODE = U.LOCAL AND W.ANO = U.ANO AND W.ANO = 2016
    AND U.SP_RUR_TOTL_ZS IS NOT NULL;
```

```
CREATE VIEW ABOVERURAL AS
SELECT COUNTRYCODE, PROPRURAL
FROM HAPPYRURAL
WHERE PROPRURAL > (SELECT AVG(PROPRURAL) FROM HAPPYRURAL)
ORDER BY PROPRURAL;
```

```
CREATE VIEW ABOVEHAPPY AS
SELECT COUNTRYCODE, HAPPINESSSCORE
FROM HAPPYRURAL
WHERE HAPPINESSSCORE > (SELECT AVG(HAPPINESSSCORE) FROM HAPPYRURAL)
ORDER BY HAPPINESSSCORE;
```

```
-- TOTAL DE PAISES EM 2016
```

```
SELECT COUNT(COUNTRYCODE)
FROM HAPPYRURAL;
```

```
-- TOTAL DE PAISES COM FELICIDADE ACIMA DA MEDIA EM 2016
```


```
SELECT COUNT(COUNTRYCODE)
FROM ABOVEHAPPY;
```

```
-- TOTAL DE PAISES COM POPULACAO RURAL (%) ACIMA DA MEDIA EM 2016
```

```
SELECT COUNT(COUNTRYCODE)
FROM ABOVERURAL;
```

```
-- TOTAL DE PAISES COM FELICIDADE E EXPECTATIVA DE VIDA ACIMA DE MEDIA EM 2016
```

```
SELECT COUNT(H.COUNTRYCODE)
FROM ABOVEHAPPY H, ABOVERURAL R
WHERE H.COUNTRYCODE = R.COUNTRYCODE;
```

- 
- Após isso fizemos uma análise para descobrir se há alguma semelhança entre os países e se países semelhantes possivelmente formam comunidades.
 - Para isso calculamos a correlação de Pearson entre os indicadores e o índice de felicidade.
 - A correlação pode ser calculada para o caso discreto da seguinte forma:

$$\rho = \frac{\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)}{\sigma_X \sigma_Y}$$


- $\mathbb{E}(X)$ é o valor esperado da variável X e σ_X é o desvio padrão de X .
- O valor diz se existe um comportamento similar em duas curvas, de maneira adimensional, variando de -1 a 1, sendo 0 sem correlação.




Query

```
-- felicidade e gdppercapita
CREATE VIEW H_GDPPERCAPITA2016 AS
SELECT
    (AVG(W.HAPPINESSSCORE*W.GDPCAPITA) - (AVG(W.HAPPINESSSCORE)*AVG(W.GDPCAPITA))) /
    (STDDEV_POP(W.HAPPINESSSCORE) * STDDEV_POP(W.GDPCAPITA))
FROM WHR W
WHERE W.ANO=2016
;
```

Obtivemos os seguintes resultados:

Indicador	Correlação com a pontuação de felicidade
PIB per capita	0,7954
 Percepção de corrupção	0,4164
Liberdade para fazer escolhas de vida	0,5849
Generosidade	0,1679
Expectativa de vida	0,7702
População rural	-0,7241
Mortalidade infantil	-0,6751
Crescimento do PIB	-0,1402
Taxa de desemprego	-0,2856
População abaixo da linha de pobreza (%)	-0,4500
PIB atual	0.2048

- 
- A partir desses dados nós criamos e utilizamos uma fórmula que chamamos de fator de semelhança.

$$S_{country} = \frac{\sum_i corr(i, f) \cdot v_i}{\sum_i |corr(i, f)|}$$


- Sendo i um indicador do país, f , a pontuação de felicidade, v_i , o valor do indicador normalizado entre 0 e 1 e $corr(i, j)$ a correlação de i e j .
- Os resultados podem ser vistos nos próximos slides.


LOCAL	INDICE
URY	0.29759078104632536
CHL	0.263485744975172
BRA	0.2513489850382742
CRI	0.2491978063627656
MYS	0.24735623755776304
MEX	0.24489444403907762
BLR	0.23640021523011157
PAN	0.22551308486506344
RUS	0.2232497358191747
COL	0.21490571781117038
THA	0.21214041543528203

LOCAL	INDICE
SUR	0.20960572019511722
DOM	0.20474037563067324
TUR	0.20186311337765014
SLV	0.18212256855438205
PRY	0.17367432612733918
MNG	0.17211350605782952
MUS	0.17188645176405853
BLZ	0.17102279418460686
IRN	0.16742449196882325
BOL	0.1647498983528386
MNE	0.16422512676501544

LOCAL	INDICE
GEO	0.16260659164421745
GTM	0.1579282962835596
LKA	0.15536194423083788
ALB	0.1536088396597242
IDN	0.15118174587651073
VNM	0.15013072204013694
PHL	0.1494753008388515
TUN	0.14672934646994937
IRQ	0.1367364421823983
SRB	0.13514453703713175
MKD	0.1350798920896189

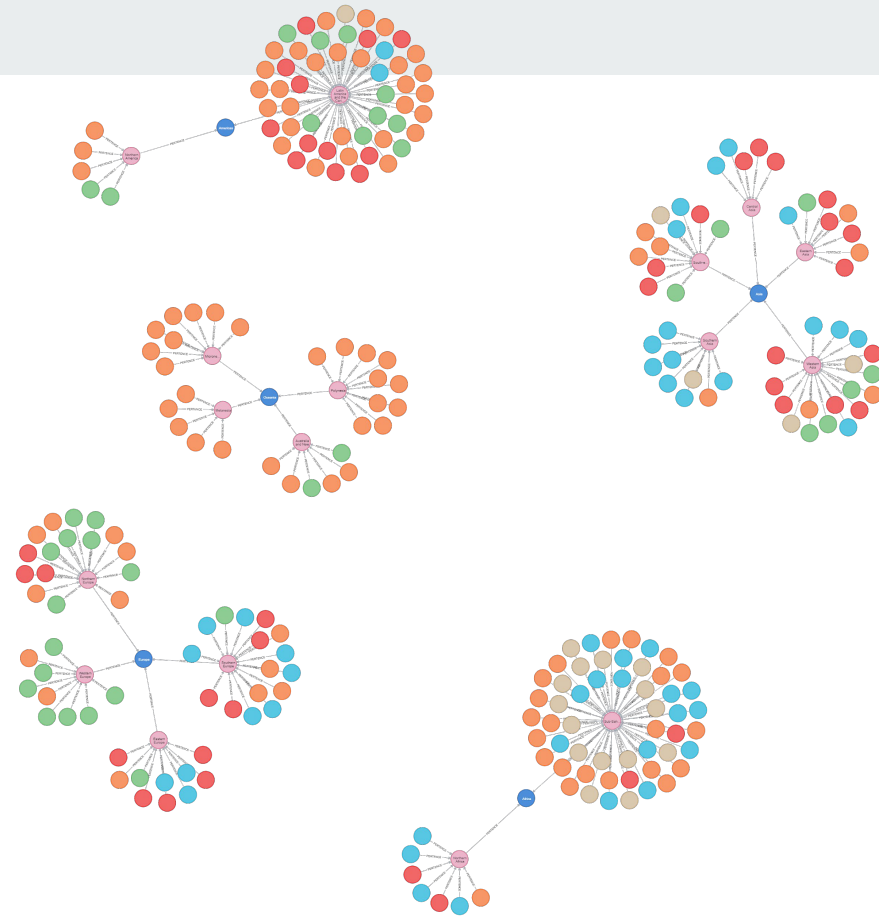
LOCAL	INDICE
UKR	0.13359304065916516
ZAF	0.11214291145322065
ARM	0.11091979672703528
EGY	0.09760336217878798
KGZ	0.09283418299754365
NAM	0.09174886561396203
BIH	0.0884179600634975
KHM	0.08598591380340866
BGD	0.07144568655534829
KEN	0.022068705825141855
MLI	-0.06668138970140604
LBR	-0.04998580940340341

- 
- Com estes fatores calculados pudemos usar o Cypher para produzir um grafo que liga os países semelhantes.
 - O critério de semelhança é $|F_i - F_j| < 0,01$, onde F_i é o fator de semelhança do país i e F_j é o fator de semelhança do país j , e este valor 0,01 foi definido empiricamente. Para valores maiores uma boa parte dos países estaria na mesma comunidade e para valores menores, o grafo seria composto basicamente por várias componentes conexas desconectadas contendo apenas um vértice, ambos resultados impossibilitando a análise desejada

- 
- A partir dessas associações utilizamos o Algoritmo de Louvain para detectar possíveis comunidades formadas entre os países, a representação gráfica dos resultados e grafos construídos pode ser vista nos próximos slides.

Grafo de regiões, sub-regiões e países

- Os países são as folhas
- Verde: índice de felicidade $\geq 0,75$
- Vermelho: $0,75 > \text{índice} \geq 0,5$
- Azul claro: $0,25 \geq \text{índice} > 0,5$
- Bege: índice $< 0,25$
- Laranja: sem dados



Não parece haver uma relação clara entre felicidade e subregião, então essa linha de pesquisa foi abandonada em favor das análises com fator de semelhança.



Query Cypher

```
LOAD CSV WITH HEADERS FROM 'https://raw.githubusercontent.com/Chris  
CREATE (:PAIS {CODE: line.ALPHA3});  
CREATE INDEX ON :PAIS(CODE);
```

```
LOAD CSV WITH HEADERS FROM 'https://raw.githubusercontent.com/Chris  
MATCH (N:PAIS {CODE: LINE.COUNTRYCODE})  
SET N.HAPPY = toFloat(LINE.HAPPINESSSCORE);
```

```
MATCH (N:PAIS)  
WHERE N.HAPPY >= 0.75  
SET N:HAPPY;
```

```
MATCH (N:PAIS)  
WHERE N.HAPPY >= 0.5 AND N.HAPPY < 0.75  
SET N:SOMEWHATHAPPY;
```

```
MATCH (N:PAIS)  
WHERE N.HAPPY >= 0.25 AND N.HAPPY < 0.5  
SET N:SOMEWHATNOTHAPPY;
```

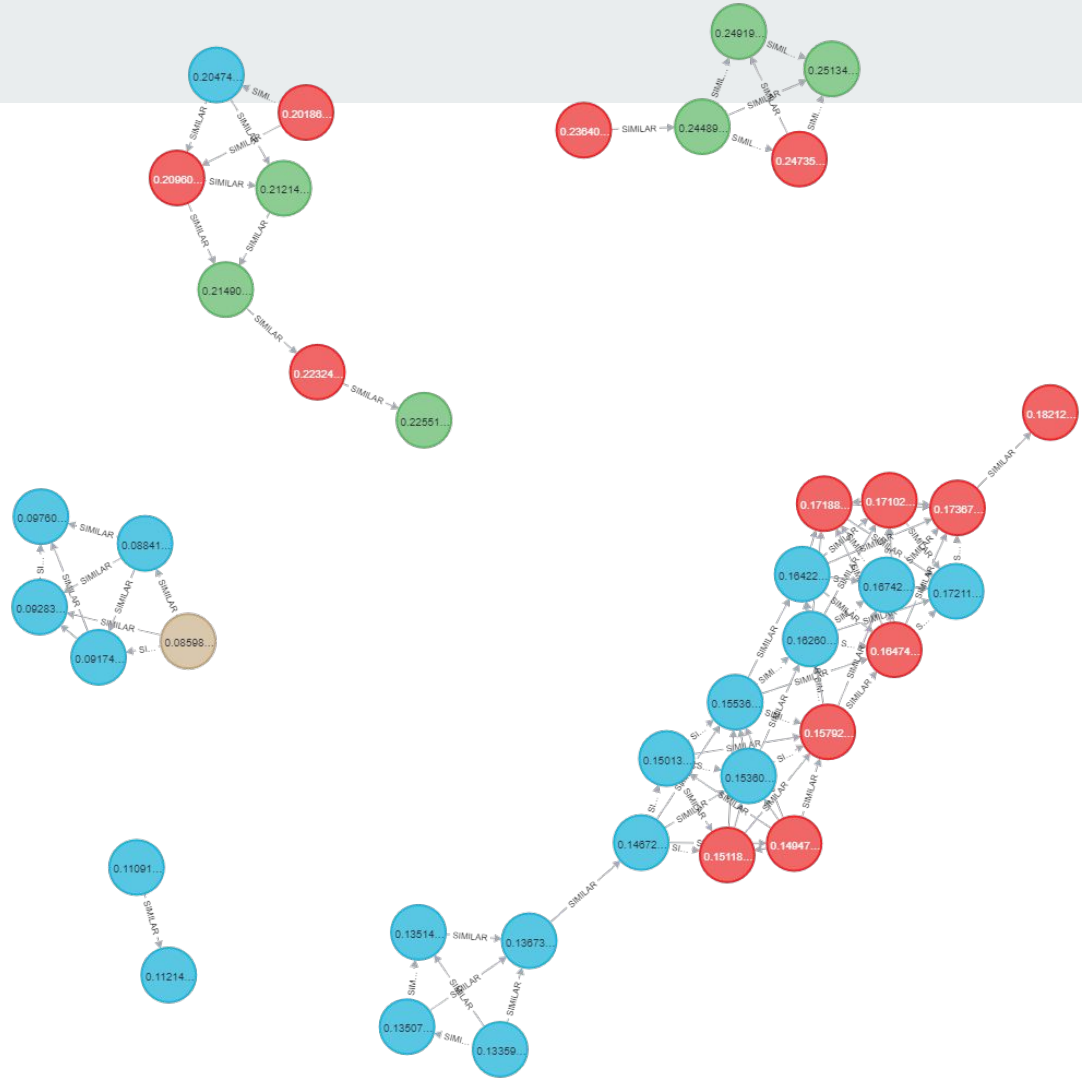
```
MATCH (N:PAIS)  
WHERE N.HAPPY < 0.25  
SET N:NOTHAPPY;
```

```
LOAD CSV WITH HEADERS FROM 'https://raw.githubusercontent.com/Chris  
MATCH (P:PAIS {CODE: LINE.LOCAL})  
SET P.INDICE = toFloat(LINE.INDICE);
```

```
MATCH (N:PAIS)  
MATCH (O:PAIS)  
WHERE N <> O AND N.INDICE > O.INDICE - 0.01 AND N.INDICE < O.INDICE  
CREATE (N)-[:SIMILAR]->(O);
```

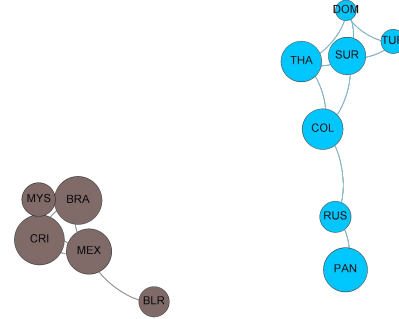
Grafo de semelhança visualizado no Cypher

- Verde: índice de felicidade $\geq 0,75$
- Vermelho: $0,75 > \text{índice} \geq 0,5$
- Azul: $0,25 \geq \text{índice} > 0,5$
- Bege: índice $< 0,25$
- Número no vértice: fator de semelhança

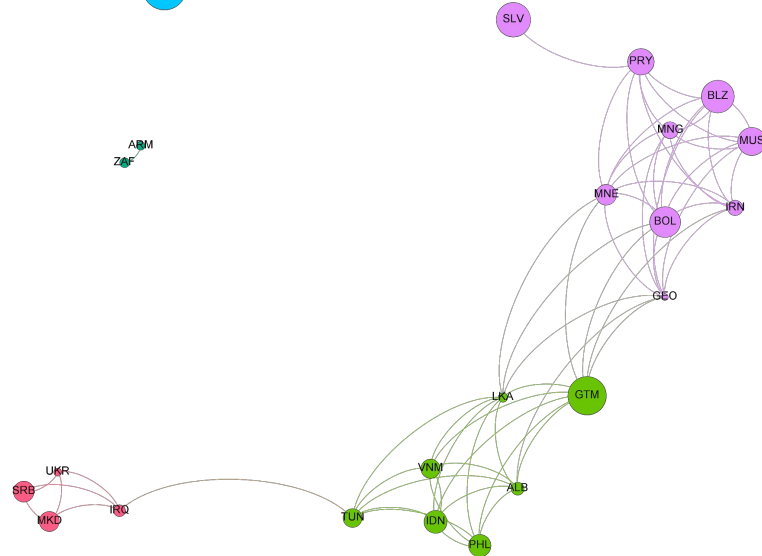


Visualização de comunidades com o Gephi

- Cada cor é uma comunidade.
- O tamanho do nó é proporcional ao seu índice de felicidade.

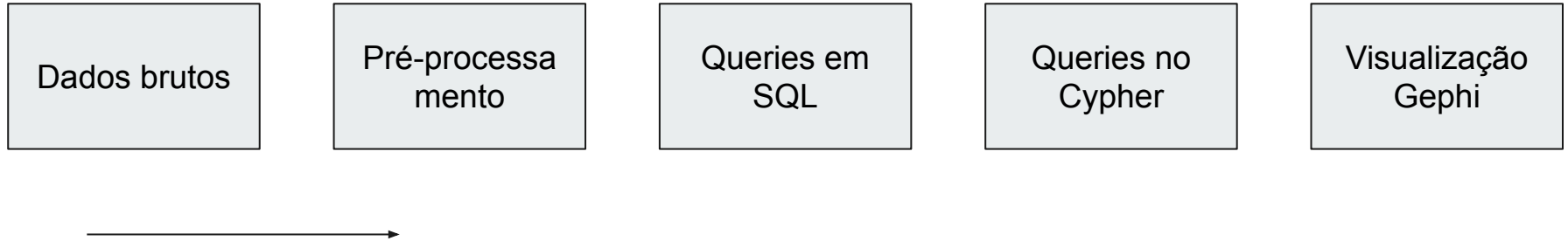


ABM
ZAF






Resumo da jornada






Conclusão

- A partir do grafo de semelhança construído foi possível perceber que países na mesma comunidade de semelhança possuíam níveis próximos de felicidade.
- Desta forma, este valor que criamos poderia ser usado como um possível indicador indireto da felicidade de um país. No entanto não existe um conjunto de testes para comprovar esta hipótese.

- 
- Também é possível perceber com base no grafo de localização geográfica que estas comunidades não possuem relação alguma com a localização de um país, apenas, provavelmente, com seus fatores socioeconômicos, que foi o que utilizamos em nossa análise. Por exemplo, Brasil, Belarus, Malásia, Costa Rica, México e Panamá estão numa mesma comunidade,

- 
- Um fato importante a se destacar é que o índice de felicidade do World Happiness Report é um resultado subjetivo que depende basicamente da percepção de uma amostra da população de um país sobre sua própria felicidade, então muitos outros fatores baseados em outras evidências empíricas foram deixadas de fora de nossa análise que poderiam alterar alguns de nossos resultados, no entanto, isso não invalida os resultados interessantes que foram produzidos em nossa análise.