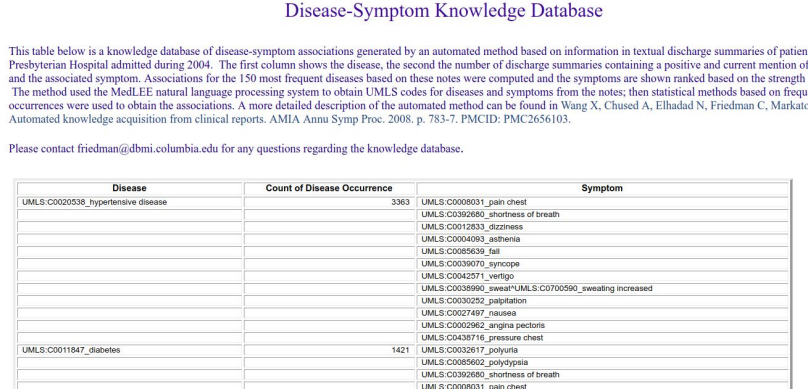
Etapa 3

Análise de dados a partir de grafos

Banco de dados

1) Disease-Symptom Knowledge Database



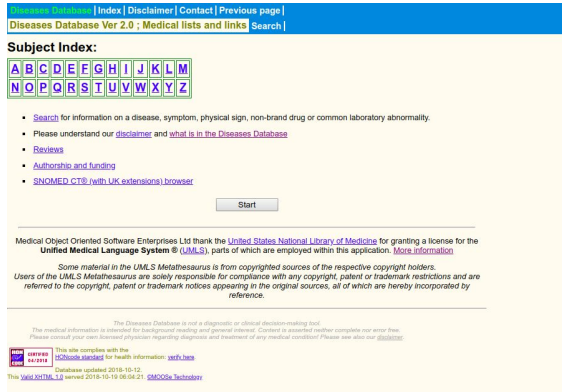
Disease-Symptom Knowledge Database

This table below is a knowledge database of disease-symptom associations generated by an automated method based on information in textual discharge summaries of patien
Presbyterian Hospital admitted during 2004. The first column shows the disease, the second the number of discharge summaries containing a positive and current mention of
and the associated symptom. Associations for the 150 most frequent diseases based on these notes were computed and the symptoms are shown ranked based on the strength
The method used the MedLEE natural language processing system to obtain UMLS codes for diseases and symptoms from the notes; then statistical methods based on frequ
occurrences were used to obtain the associations. A more detailed description of the automated method can be found in Wang X, Chused A, Elhadad N, Friedman C, Markatc
Automated knowledge acquisition from clinical reports. AMIA Annu Symp Proc. 2008. p. 783-7. PMCID: PMC2656103.

Please contact friedman@dbmi.columbia.edu for any questions regarding the knowledge database.

Disease	Count of Disease Occurrence	Symptom
UMLS:C0020538_hypertensive disease	3363	UMLS:C0008031_pain chest
		UMLS:C0392680_shortness of breath
		UMLS:C0012833_dizziness
		UMLS:C0004093_asthenia
		UMLS:C0085639_fall
		UMLS:C0039070_syncope
		UMLS:C0042571_vertigo
		UMLS:C0038990_sweat^UMLS:C0700590_sweating increased
		UMLS:C0030252_palpitation
		UMLS:C0027497_nausea
		UMLS:C0002962_angina pectoris
		UMLS:C0438716_pressure chest
UMLS:C0011847_diabetes	1421	UMLS:C0032617_polyuria
		UMLS:C0085602_polydypsia
		UMLS:C0392680_shortness of breath
		UMLS:C0008031_pain chest

2) Diseases Database



Diseases Database | Index | Disclaimer | Contact | Previous page |

Diseases Database Ver 2.0 ; Medical lists and links Search |

Subject Index:

A B C D E F G H I J K L M

N O P Q R S T U V W X Y Z

- Search for information on a disease, symptom, physical sign, non-brand drug or common laboratory abnormality.
- Please understand our disclaimer and what is in the Diseases Database
- Reviews
- Authorship and funding
- SNOMED CT® (with UK extensions) browser

Start

Medical Object Oriented Software Enterprises Ltd thank the United States National Library of Medicine for granting a license for the **Unified Medical Language System ®** (UMLS), parts of which are employed within this application. More information

Some material in the UMLS Metathesaurus is from copyrighted sources of the respective copyright holders. Users of the UMLS Metathesaurus are solely responsible for compliance with any copyright, patent or trademark restrictions and are referred to the copyright, patent or trademark notices appearing in the original sources, all of which are hereby incorporated by reference.

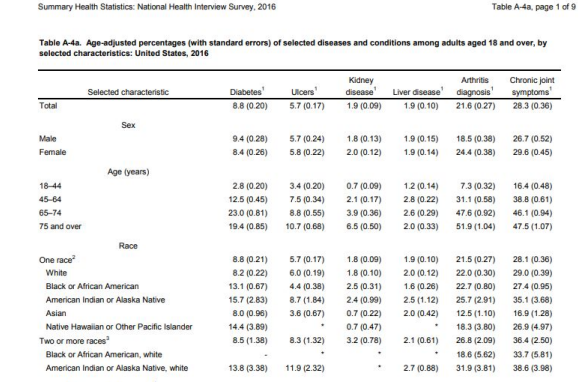
The Diseases Database is not a diagnostic or clinical decision-making tool.
The medical information is intended for background reading and general interest. Content is asserted neither complete nor error free.
Please consult your own licensed physician regarding diagnosis and treatment of any medical condition! Please see also our disclaimer.

This site complies with the HONcode standard for health information: verify here

Database updated 2018-10-12 /

This Valid XHTML 1.0 served 2018-10-19 08:04:21 - GMOOSe Technology

3) National Health Interview Survey

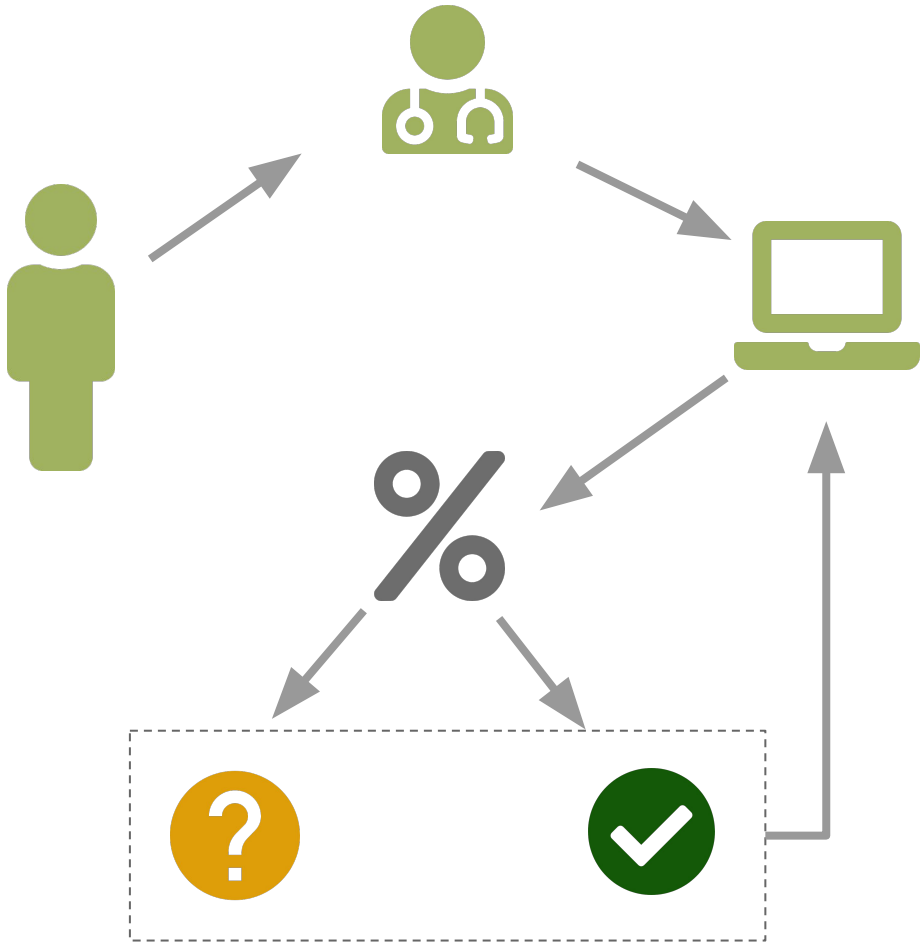


Summary Health Statistics: National Health Interview Survey, 2016 — Table A-4a, page 1 of 9

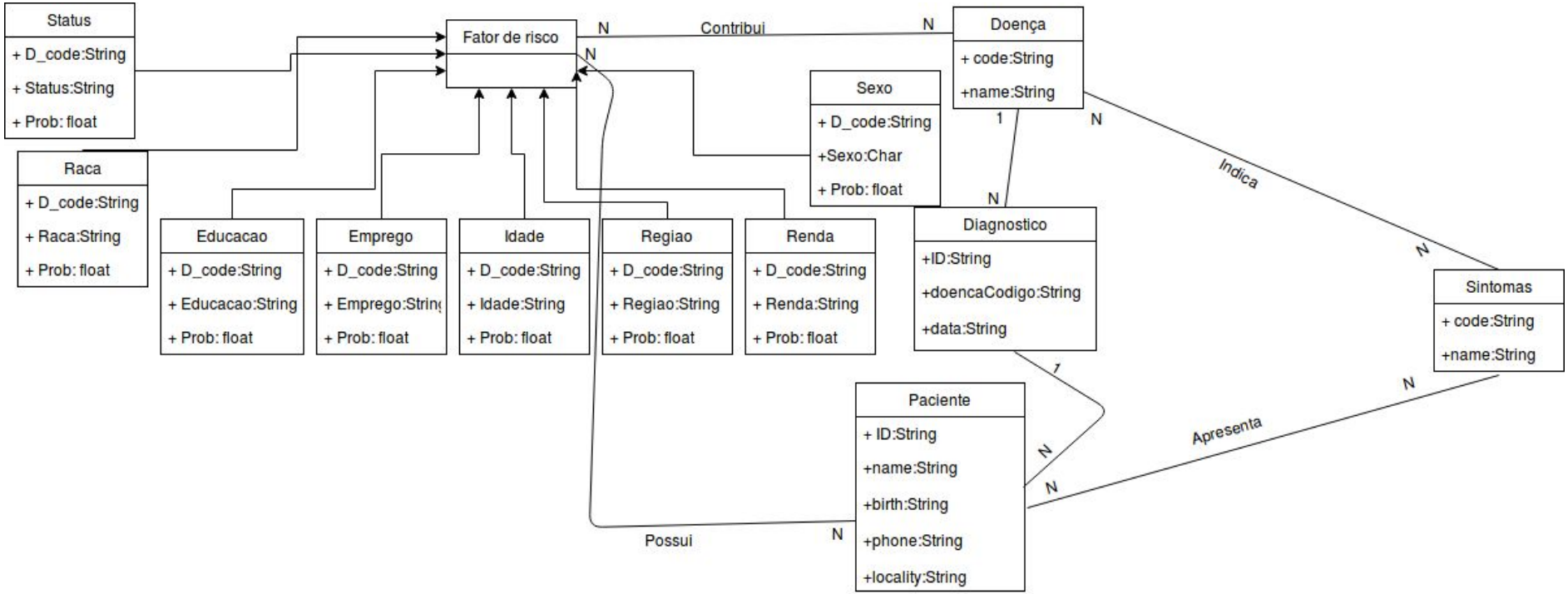
Table A-4a. Age-adjusted percentages (with standard errors) of selected diseases and conditions among adults aged 18 and over, by selected characteristics: United States, 2016

Selected characteristic	Diabetes[1]	Ulcers[1]	Kidney disease[1]	Liver disease[1]	Arthritis diagnosis[1]	Chronic joint symptoms[1]
Total	8.8 (0.20)	5.7 (0.17)	1.9 (0.09)	1.9 (0.10)	21.6 (0.27)	28.3 (0.36)
Sex						
Male	9.4 (0.28)	5.7 (0.24)	1.8 (0.13)	1.9 (0.15)	18.5 (0.38)	26.7 (0.52)
Female	8.4 (0.26)	5.8 (0.22)	2.0 (0.12)	1.9 (0.14)	24.4 (0.38)	29.6 (0.45)
Age (years)						
18–44	2.8 (0.20)	3.4 (0.20)	0.7 (0.09)	1.2 (0.14)	7.3 (0.32)	16.4 (0.48)
45–64	12.5 (0.45)	7.5 (0.34)	2.1 (0.17)	2.8 (0.22)	31.1 (0.58)	38.8 (0.61)
65–74	23.0 (0.81)	8.8 (0.55)	3.9 (0.36)	2.6 (0.29)	47.6 (0.92)	46.1 (0.94)
75 and over	19.4 (0.85)	10.7 (0.68)	6.5 (0.50)	2.0 (0.33)	51.9 (1.04)	47.5 (1.07)
Race						
One race[2]	8.8 (0.21)	5.7 (0.17)	1.8 (0.09)	1.9 (0.10)	21.5 (0.27)	28.1 (0.36)
White	8.2 (0.22)	6.0 (0.19)	1.8 (0.10)	2.0 (0.12)	22.0 (0.30)	29.0 (0.39)
Black or African American	13.1 (0.67)	4.4 (0.38)	2.5 (0.31)	1.6 (0.26)	22.7 (0.80)	27.4 (0.95)
American Indian or Alaska Native	15.7 (2.83)	8.7 (1.84)	2.4 (0.99)	2.5 (1.12)	25.7 (2.91)	35.1 (3.68)
Asian	8.0 (0.96)	3.6 (0.67)	0.7 (0.22)	2.0 (0.42)	12.5 (1.10)	16.9 (1.28)
Native Hawaiian or Other Pacific Islander	14.4 (3.89)	*	0.7 (0.47)	*	18.3 (3.80)	26.9 (4.97)
Two or more races[3]	8.5 (1.38)	8.3 (1.32)	3.2 (0.78)	2.1 (0.61)	26.8 (2.09)	36.4 (2.50)
Black or African American, white	-	*	*	*	18.6 (5.62)	33.7 (5.81)
American Indian or Alaska Native, white	13.8 (3.38)	11.9 (2.32)	*	2.7 (0.88)	31.9 (3.81)	38.6 (3.98)

Requisitos



Modelo conceitual



Relações

Paciente.csv

ID	nome	nascimento	celular	localidade
pat002X66	Susan Kratz	26/03/1991	8889256	South
pat0044ON	Bong Williams	06/03/1975	8779548	Midwest
pat004KOX	Denny Benner	21/12/1955	9308805	Northeast
pat008GZH	Georgette Moseley	04/04/1929	8012333	South
pat0097IR	Stella Smith	10/11/1951	8612152	West
pat009J6P	Charles Hernandez	07/10/1904	9179356	West
pat00CIZQ	Raul Mohmed	10/05/1966	9177464	West
pat00D671	Julie Gayton	23/10/1924	8827791	Midwest
pat00FCFE	Violet Carrington	17/01/1900	8748114	South

33724 records in your data have been successfully parsed!

Propostas

- Categorizar os elementos do grafo
- Visualização de padrões através das relações de coocorrências
- Traçar possíveis diagnósticos através de caminhos

Modelo lógico -> Grafo

Estruturas básicas:

Doenca (codigo,nome)

codigo	nome

Sintomas (codigo,nome)

codigo	nome

Paciente(ID, nome,celular)

ID	nome	celular

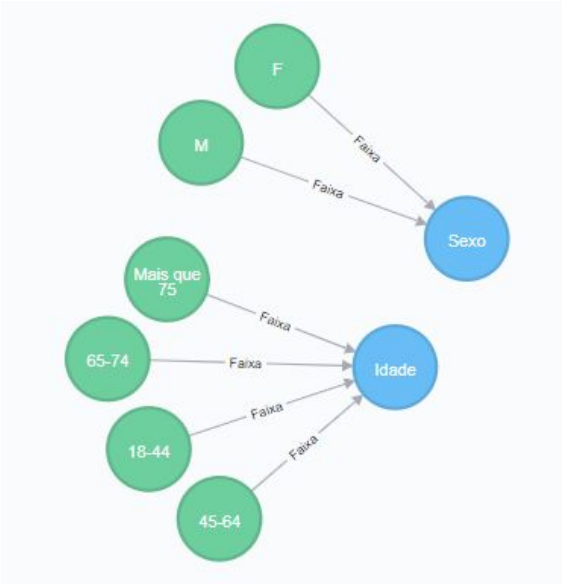
Qualificação dos fatores de risco:

Sexo (D_code,Sexo,Prob)
CHE: D_code -> Doenca (codigo)

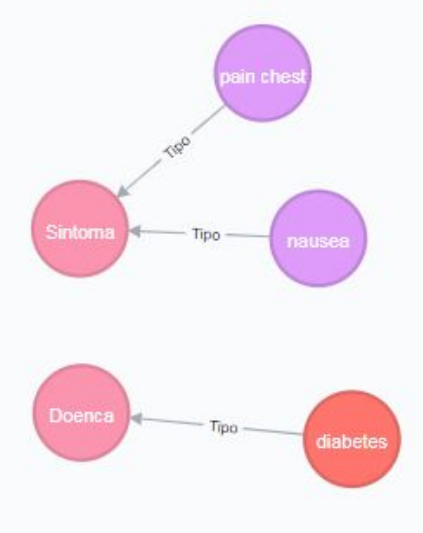
D_code	Sexo	Prob

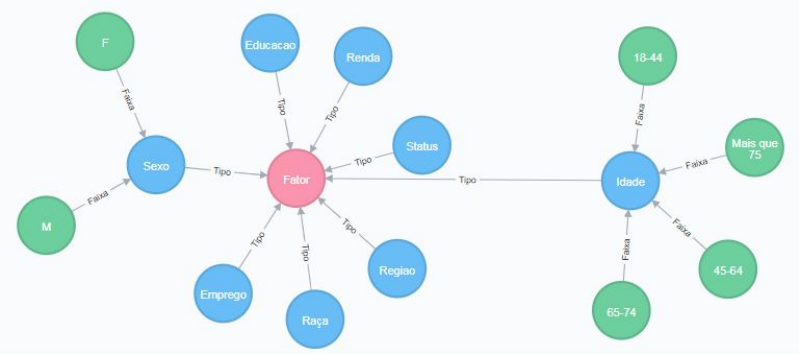
Idade (D_code,Idade,Prob)
CHE: D_code -> Doenca (codigo)

D_code	Idade	Prob

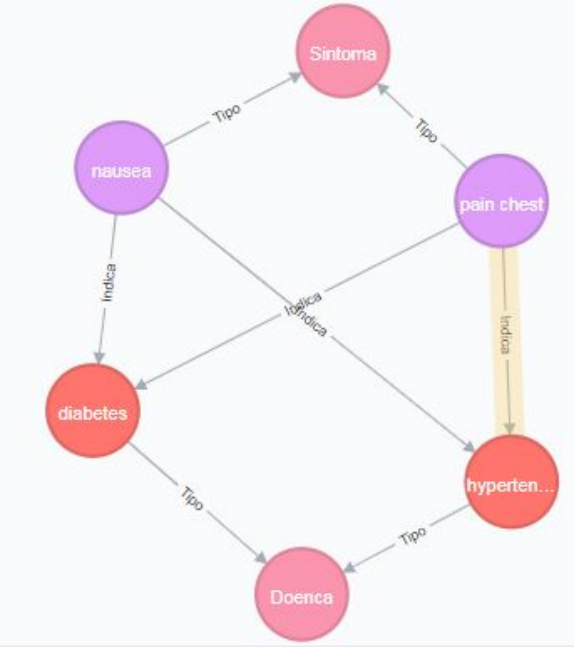


Categorias

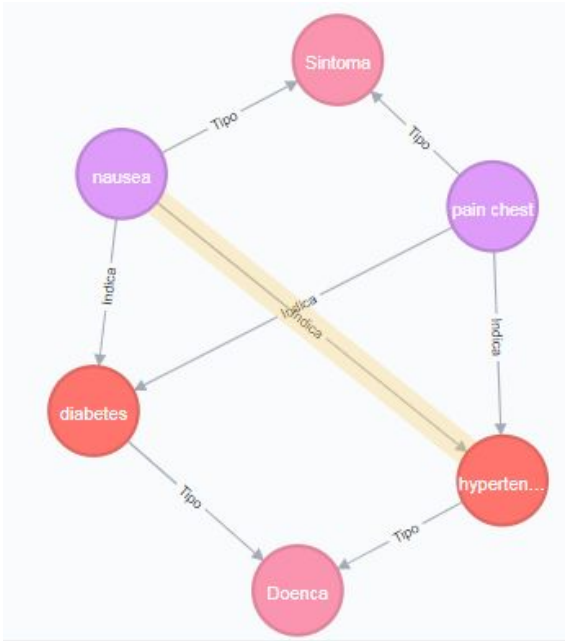

pain chest
Tipo
Sintoma
Tipo
nausea
Doenca
Tipo
diabetes


F
Faixa
Educacao
Tipo
Renda
Tipo
18-44
Faixa
Sexo
Tipo
Fator
Tipo
Status
Tipo
Idade
Faixa
Mais que 75
M
Faixa
Emprego
Tipo
Raça
Tipo
Regiao
Tipo
Faixa
65-74
Faixa
45-64

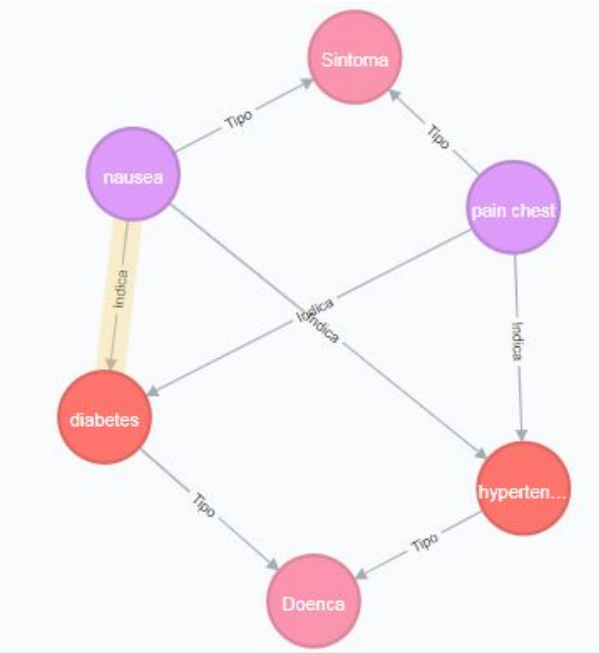
Coocorrências



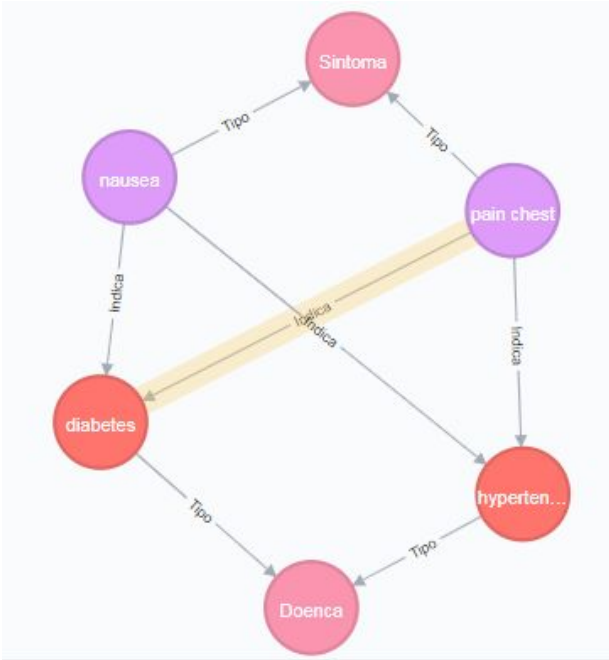
<id>: 19 **Ocorrencias:** 3000



<id>: 74 **Ocorrencias:** 1000

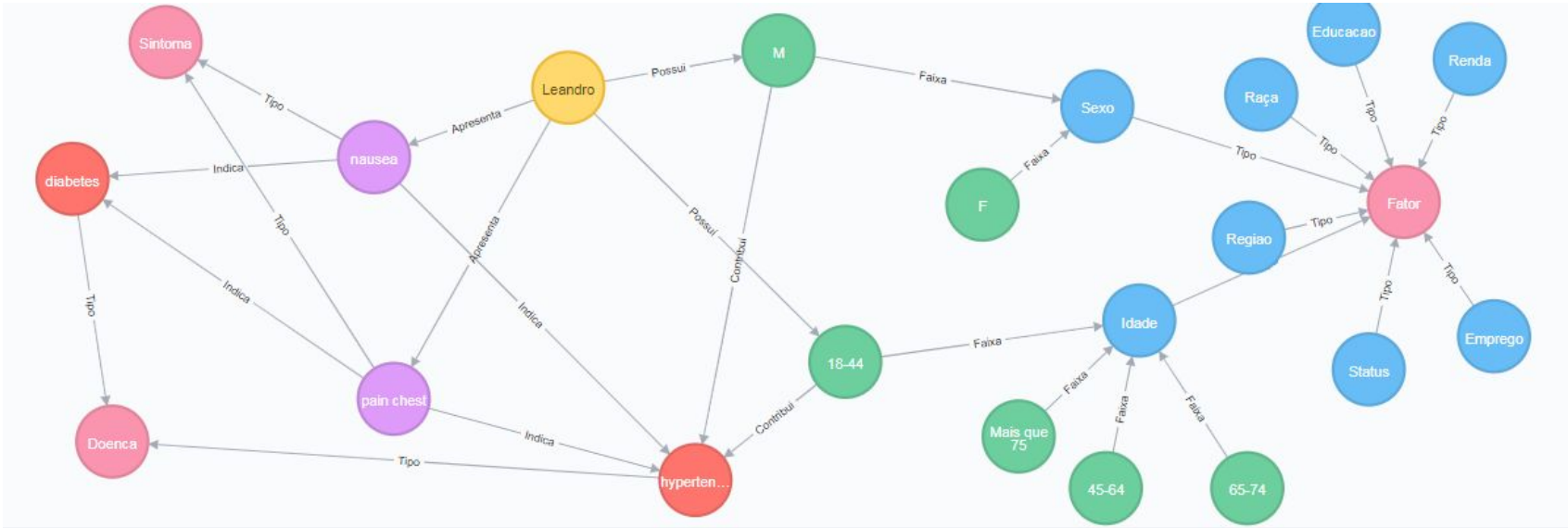


<id>: 18 **Ocorrencias:** 500



<id>: 57 **Ocorrencias:** 300

Caminhos


Sintoma
Leandro
M
Possui
Faixa
Sexo
Educacao
Renda
Raça
Tipo
Apresenta
nausea
diabetes
Indica
F
Fator
Regiao
Possui
Contribui
Apresenta
Indica
18-44
Idade
Emprego
Status
pain chest
Doenca
Contribui
Mais que 75
45-64
65-74
hyperten...