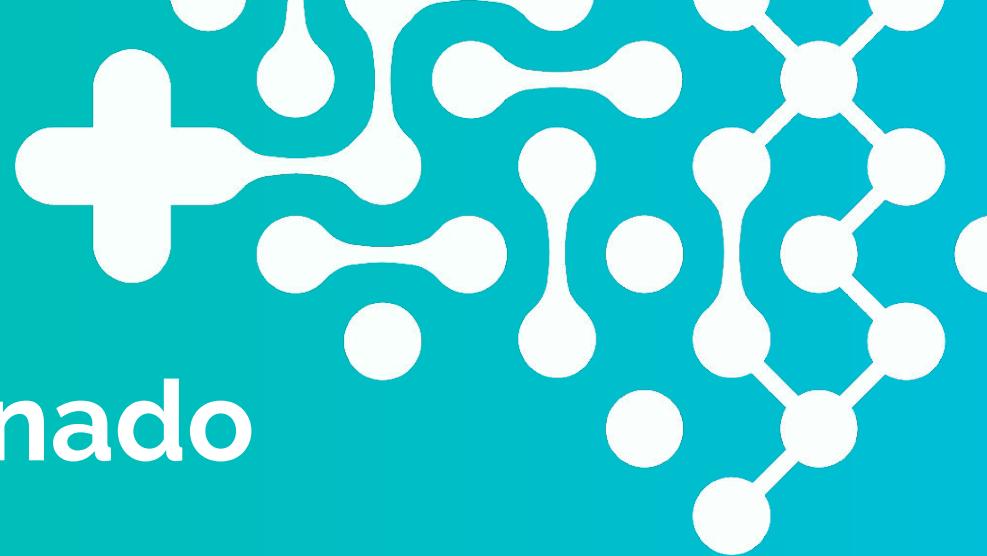


Aprendizado Não-Supervisionado na Saúde



Henrique Dias - PUCRS

Grupo de Inteligência Artificial na Saúde

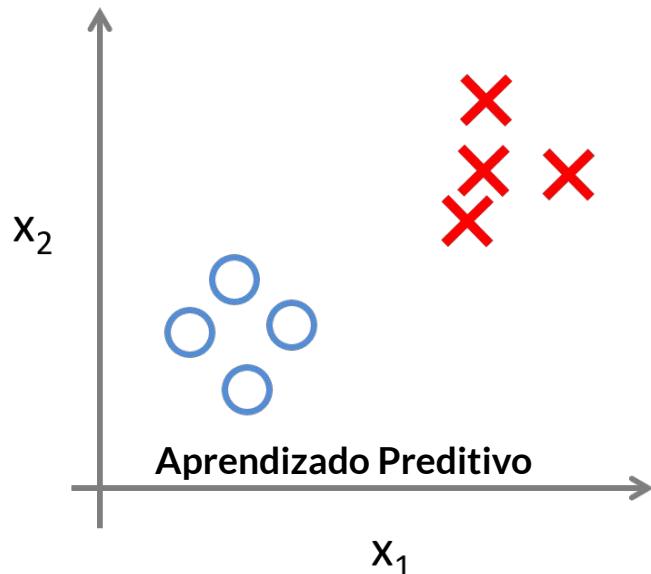
Slides: <http://goo.gl/85hzcX> ([download pdf](#))

meetup
 nubank
Janeiro 2020

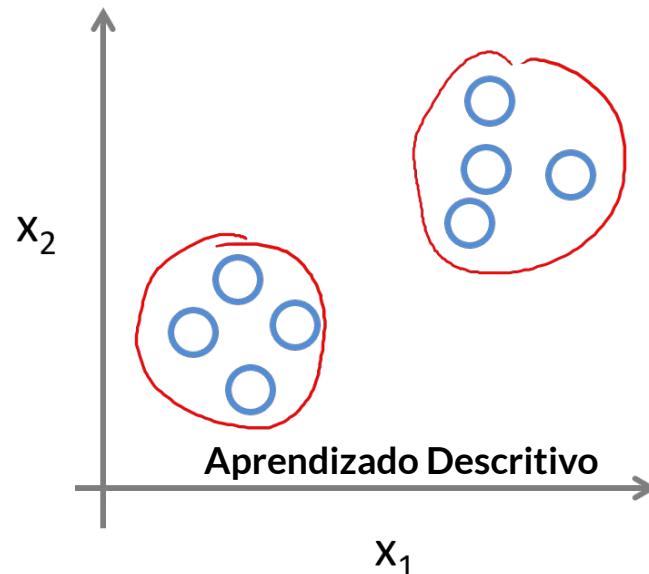


Recapitulando...

Supervised Learning



Unsupervised Learning





 alamy stock photo

JC3KYR
www.alamy.com

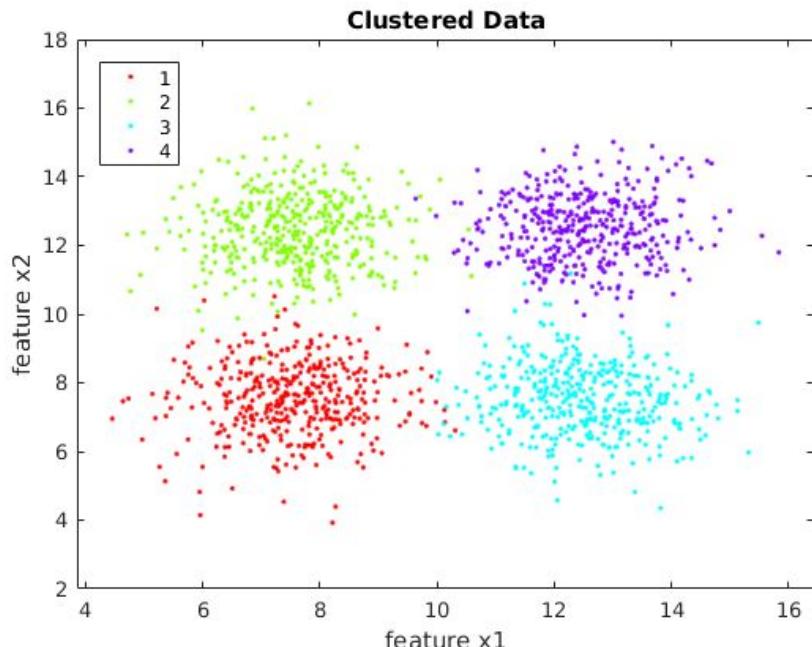
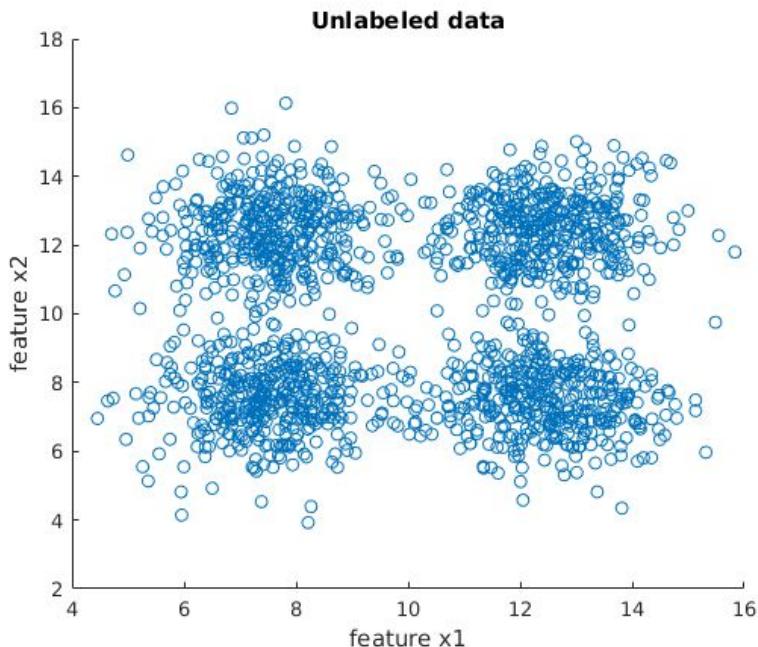


 alamy stock photo

E8NGXG
www.alamy.com

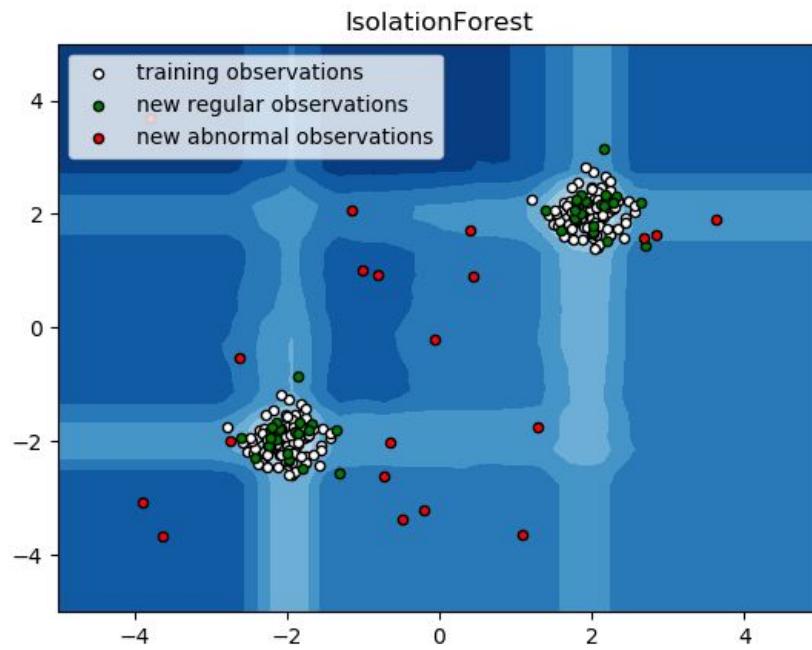
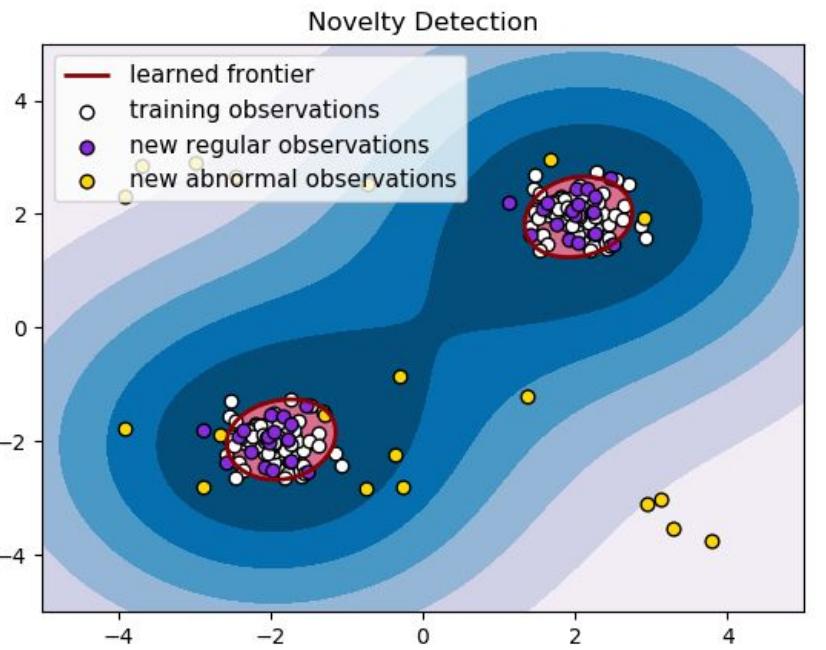


Não-Supervisionado (agrupamento)



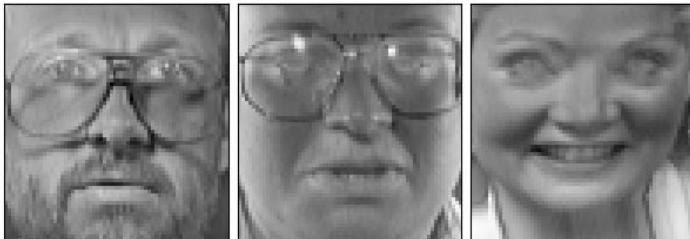


Não-Supervisionado (detecção de anomalia)

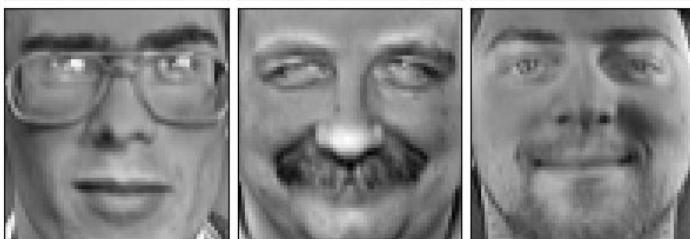


Não-Supervisionado (reduc. de dimensionalidade)

First centered Olivetti faces



genfaces - PCA using randomized SVD - Train time 0.1



PCA, NNMF, SVD, LDA, LSA, T-SNE



Não-Supervisionado (word weight)



$$M_{ij} = \frac{\# \text{ of word } i}{\# \text{ of word } i \text{ in doc } j}$$

0.5	-1.4	-1.4	-0.5	-1.4
-1.7	0.5	0.5	0.5	0.5
-0.5	0.5	0.5	-0.7	0.5
0.5	0.5	-1.8	0.5	-1.9
-1.1	0.5	0.5	0.5	0.5
	0.5	0.5	0.5	0.5
	0.5	0.5	0.5	0.5

document-word

$$SVD(M_{ij})$$

1.4	0.5	-1.4
-1.7	0.5	-0.5
-0.7	0.5	-1.8
-1.1	0.5	0.5
-1.2	0.5	0.5

1.4	0.5	-1.9	0.5	-1.9
-1.7	0.5	-0.7	0.5	-0.7
-0.7	0.5	-1.8	0.5	-1.8
-1.1	0.5	0.5	0.5	0.5
-1.2	0.5	0.5	0.5	0.5

document-word

LSI + Dirichlet Prior

0.5	0.5	0.5
0.5	0.5	0.5
0.5	0.5	0.5
0.5	0.5	0.5
0.5	0.5	0.5

0.5	0.5	0.5
0.5	0.5	0.5
0.5	0.5	0.5
0.5	0.5	0.5
0.5	0.5	0.5

document-word

$$M_{ij} = \frac{\#(i, j) / n_{pairs}}{\#(i) / n_{words} \#(j) / n_{words}}$$

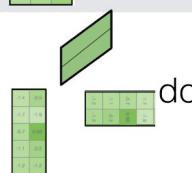
$$SVD(M_{ij})$$

-1.4	-0.5	-1.4
-1.2	-0.9	-0.5
-0.7	0.5	-1.9
-1.1	-0.2	0.5
-1.2	-0.2	0.5

word-word

$$M_{ijk} = \frac{\#(i, j, k) / n_{triplets}}{\#(i) / n_{words} \#(j) / n_{words} \#(k) / n_{words}}$$

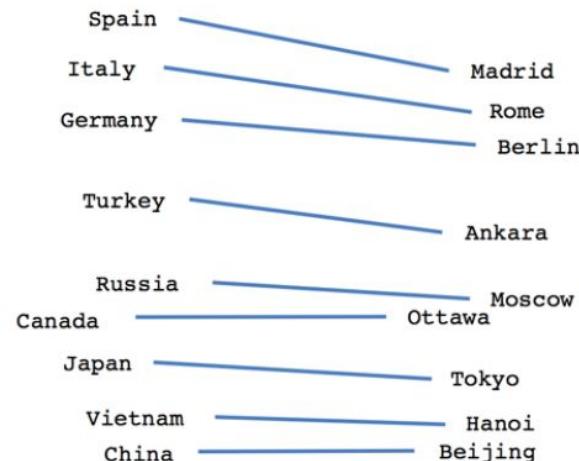
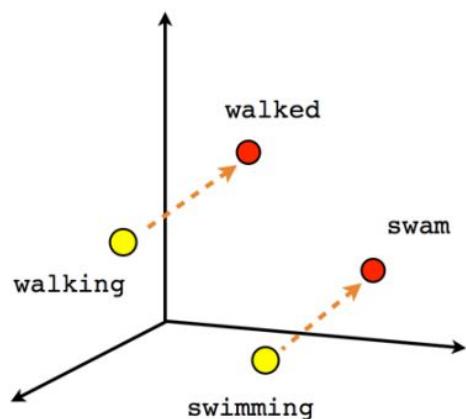
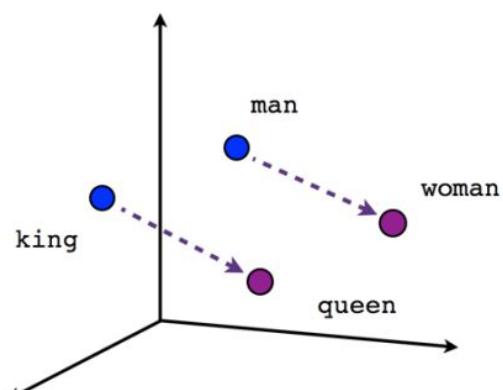
$$SVD(M_{ijk})$$

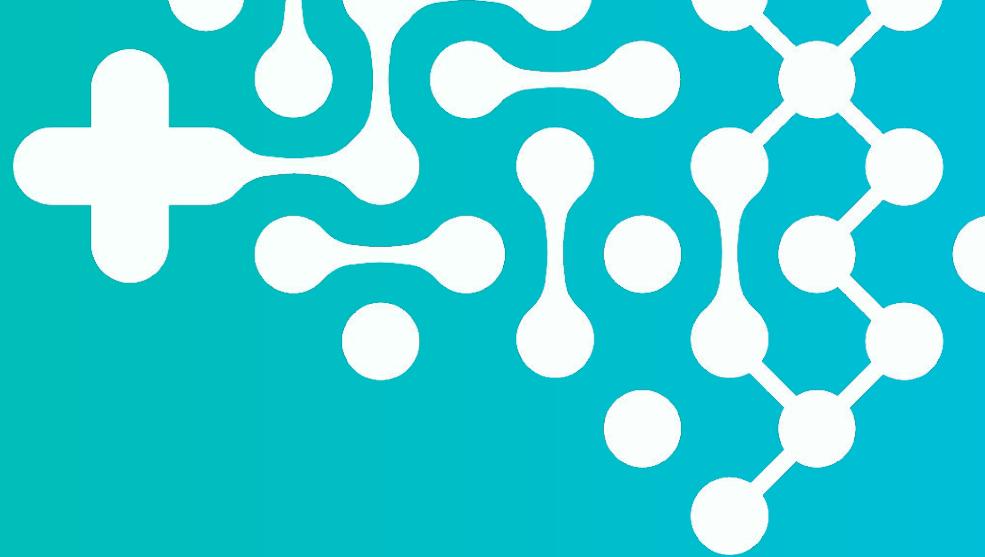


document-word-word



Não-Supervisionado (word embeddings)

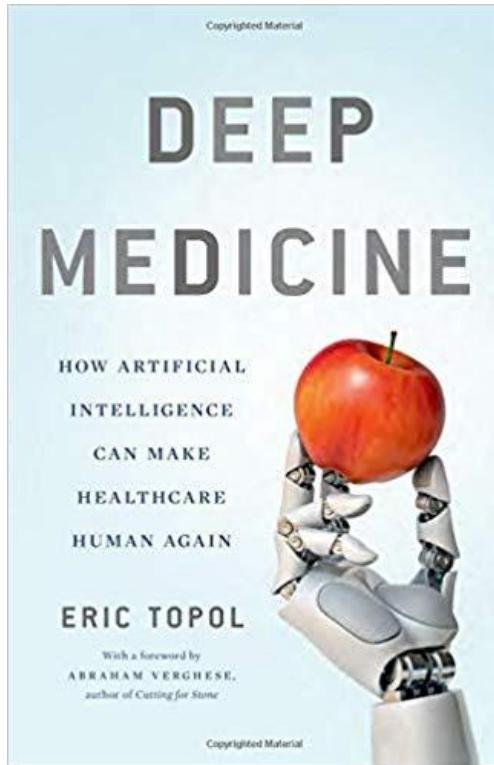




... e na saúde?



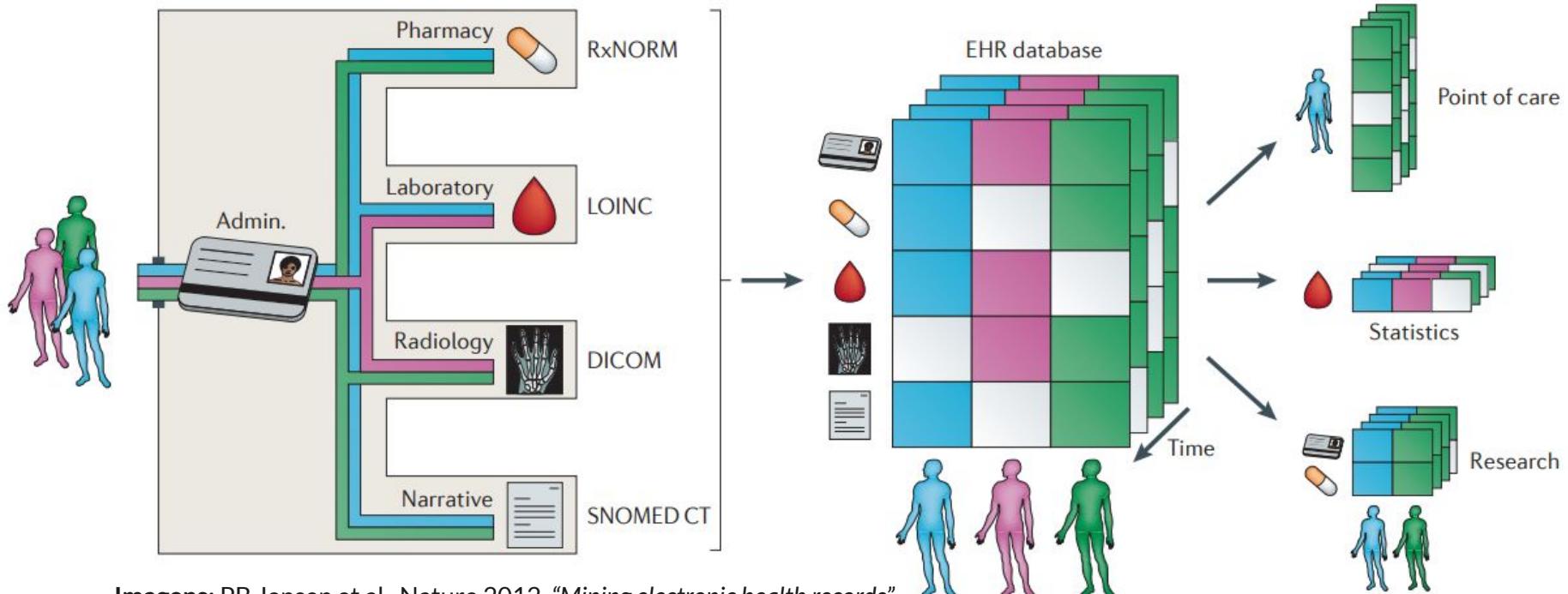
Inteligência Artificial na Saúde



“How Artificial Intelligence can make Healthcare Human again”
Eric Topol



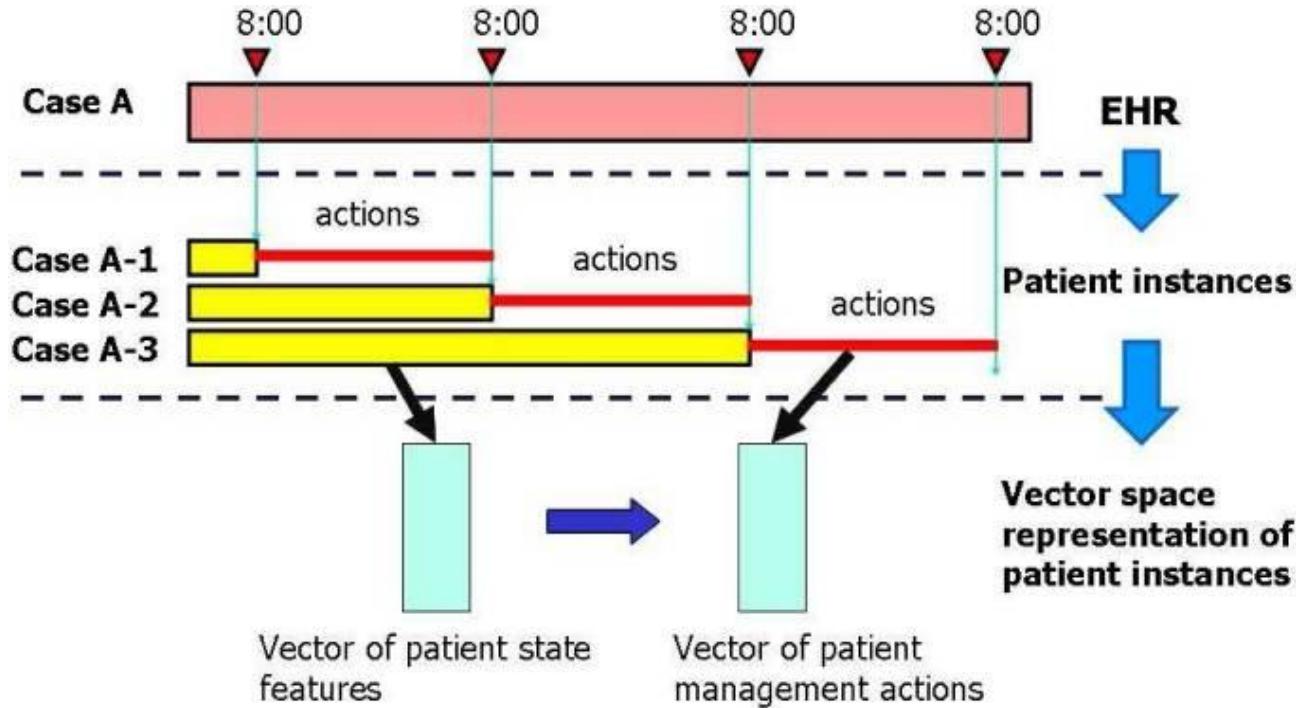
Dados Básicos de Saúde



Imagens: PB Jensen et al., Nature 2012, "Mining electronic health records"

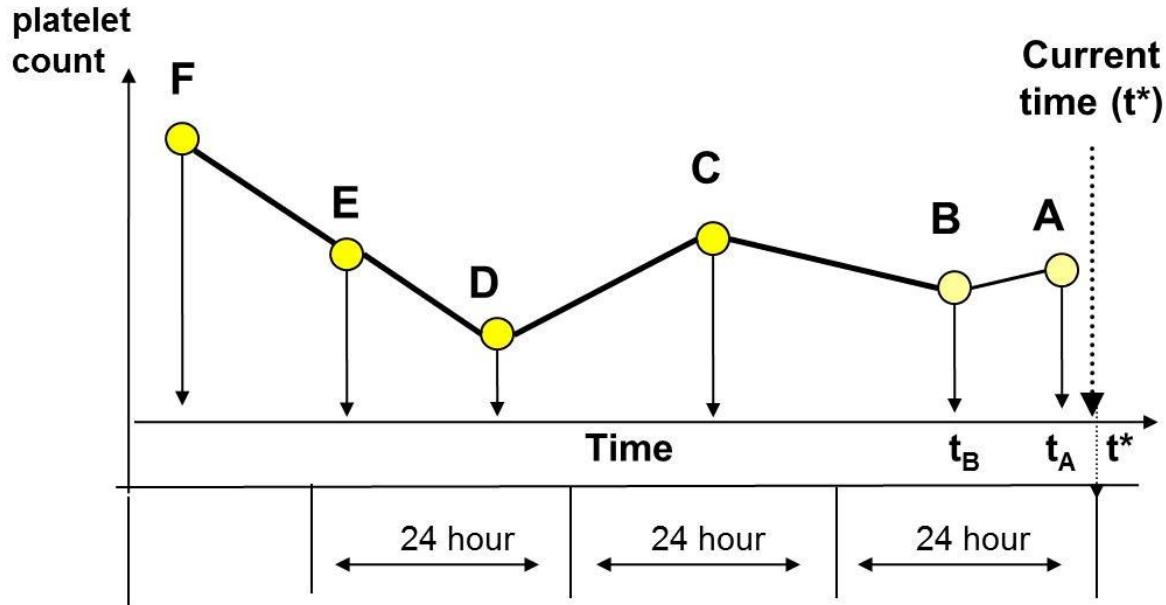


Procedimentos Fora do Padrão





Exames Fora do Padrão

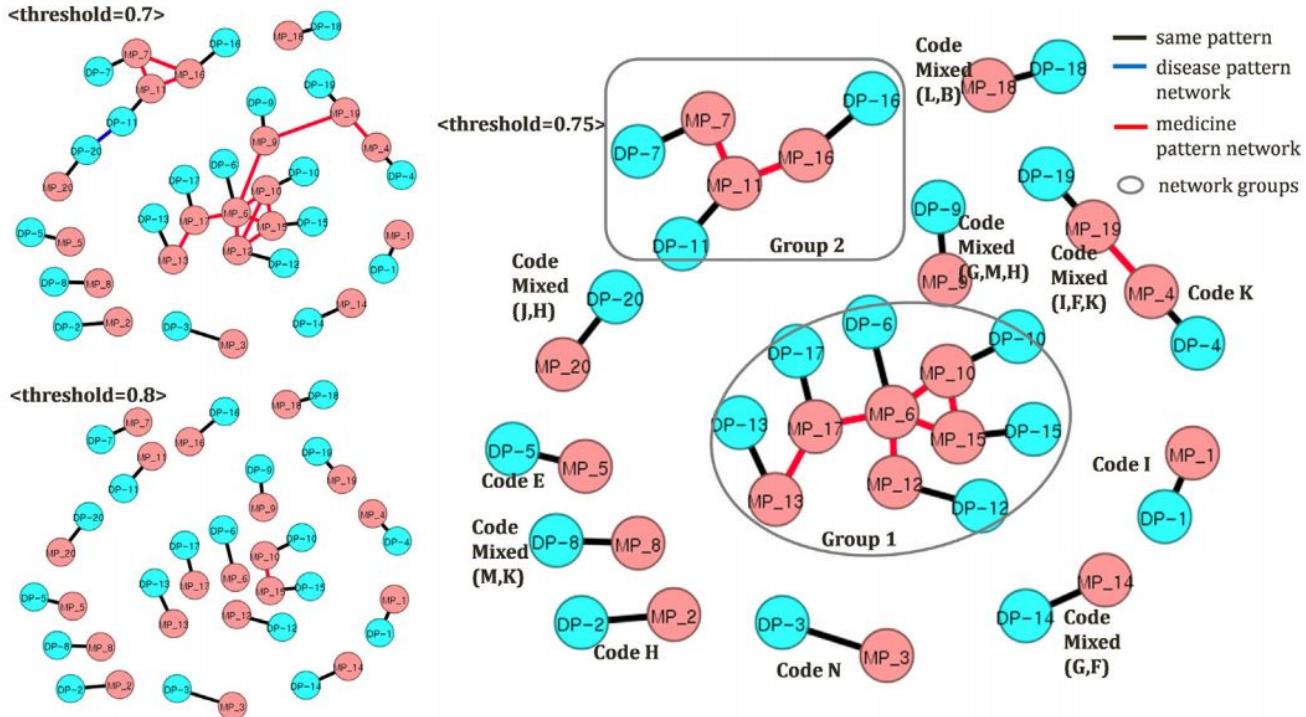


Temporal features:

Last value: A
Last value difference = A-B
Last % change = $(A-B)/B$
Last slope = $(A-B) / (t_A - t_B)$
Nadir = D,
Nadir difference = A-D,
Nadir% difference = $(A-D)/D$
...



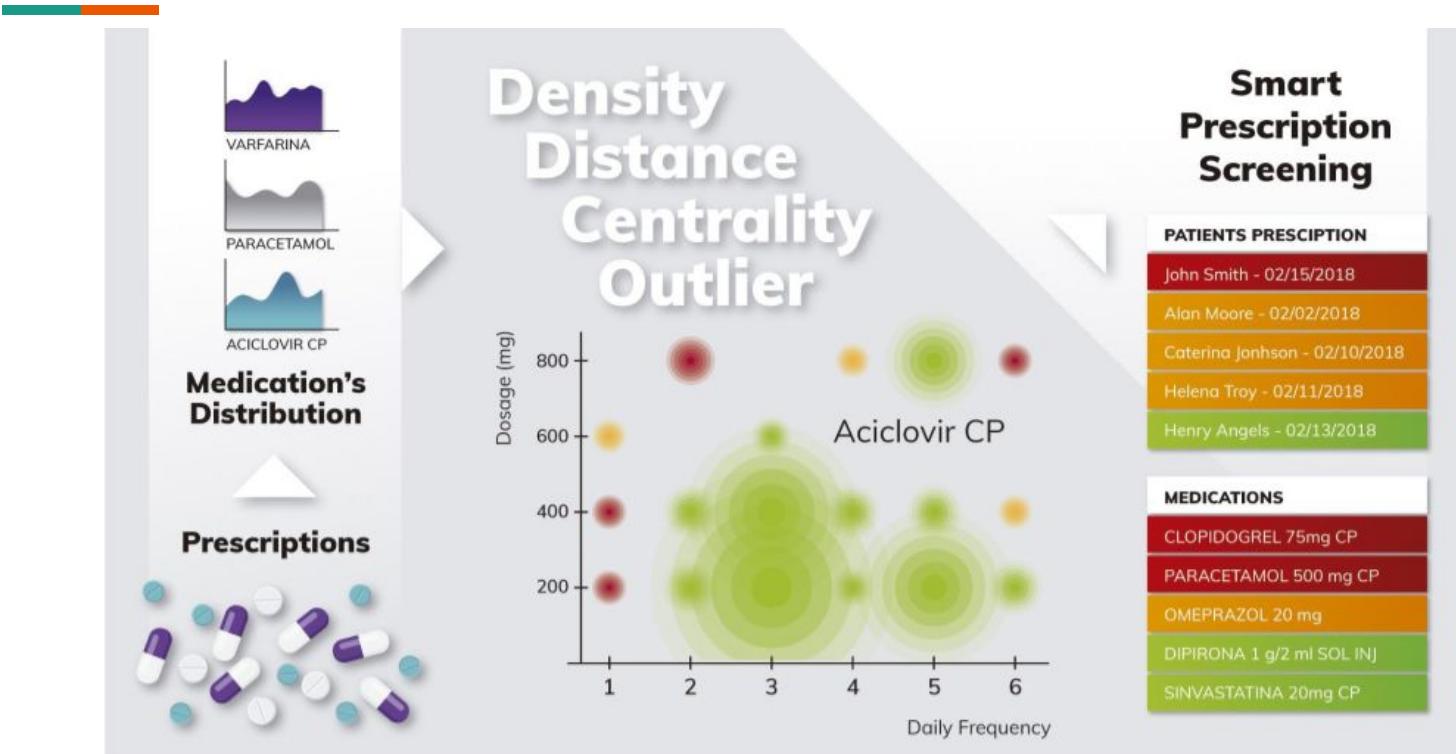
Prescrições Padrão



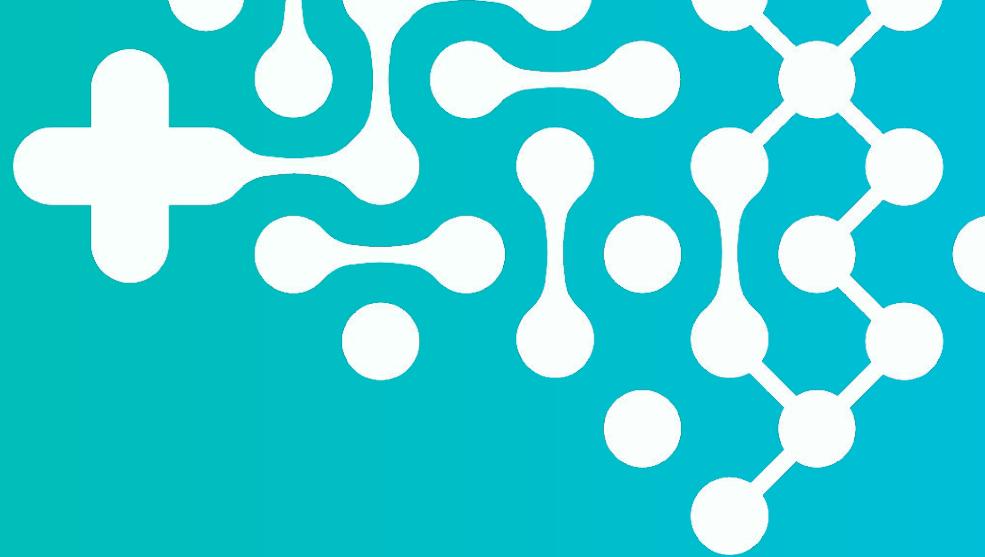
Imagens: S Park et al., JBI 2017, "Identifying prescription patterns with a topic model of diseases and medications."



Prescrições Fora do Padrão

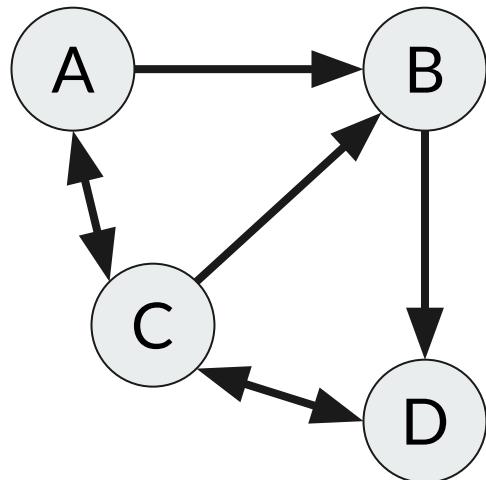


Imagens: HDP Santos et al., JBHI 2018, "DDC-Outlier: Preventing Medication Errors Using Unsupervised Learning"



Aprendizado Não-Supervisionado e o PageRank

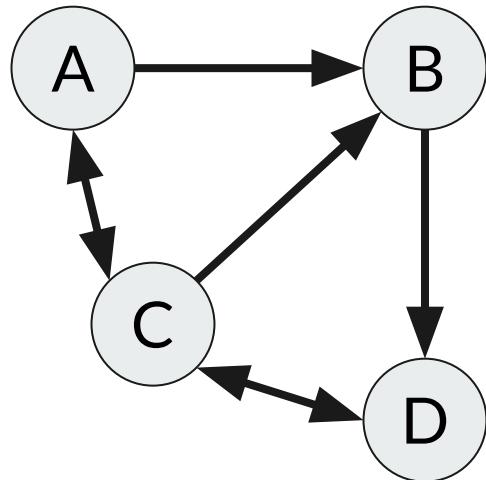
Google PageRank



Problema de Autovetor ou
Passeio Aleatório

- PageRank
- HITS
- SALSA

Google PageRank

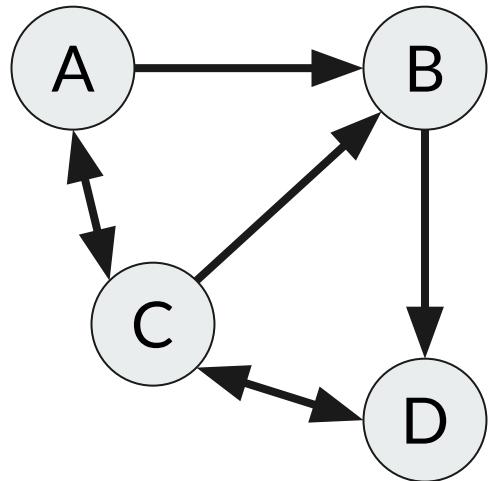


Web como um Grafo:

- Páginas são vértices
- Links são arestas



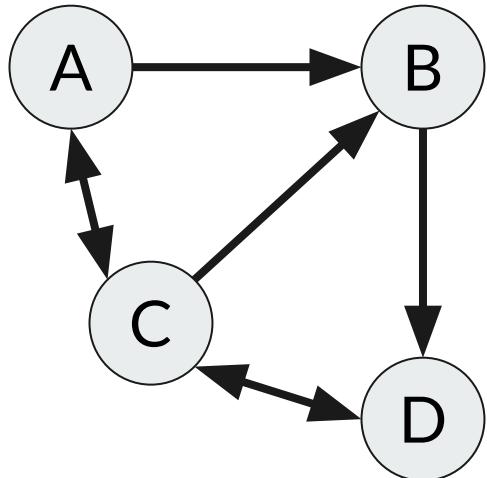
Google PageRank



$$\text{PR}(u) = \sum \frac{\text{PR}(v)}{N_v}$$



Google PageRank

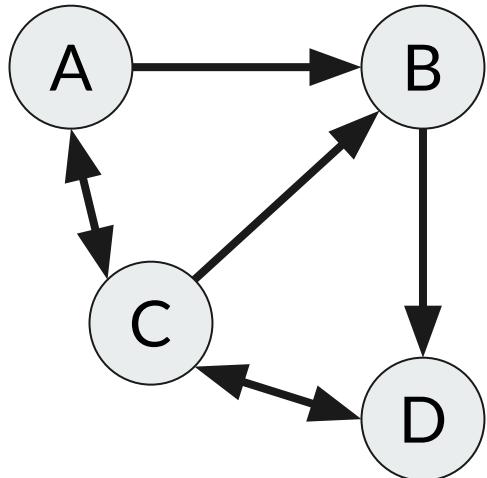


$$\text{PR}(u) = \sum \frac{\text{PR}(v)}{N_v}$$

	step 0
A	1
B	1
C	1
D	1
ϵ	-



Google PageRank

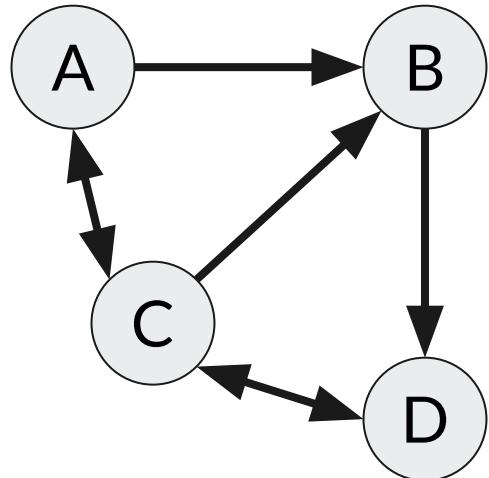


$$\text{PR}(u) = \sum \frac{\text{PR}(v)}{N_v}$$

	step 0	step 1
A	1	0.33
B	1	0.83
C	1	1.50
D	1	1.33
ϵ	-	0.41



Google PageRank

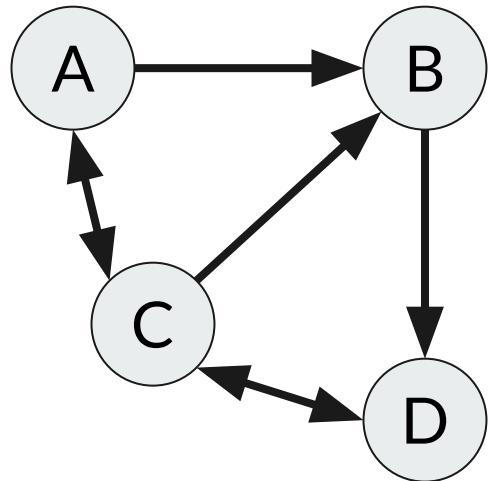


$$\text{PR}(u) = \sum \frac{\text{PR}(v)}{N_v}$$

	step 0	step 1	step 2
A	1	0.33	0.50
B	1	0.83	0.66
C	1	1.50	1.49
D	1	1.33	1.33
ϵ	-	0.41	0.08



Google PageRank

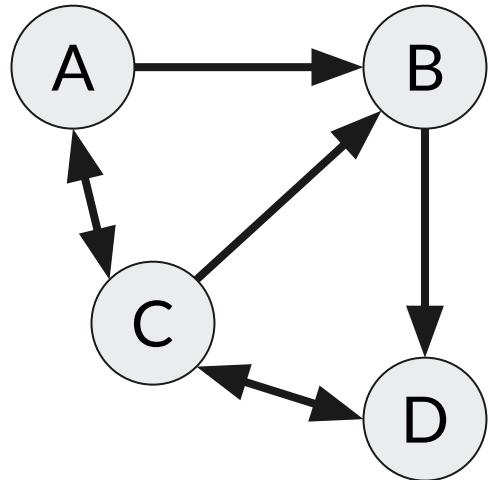


$$\text{PR}(u) = \sum \frac{\text{PR}(v)}{N_v}$$

	step 0	step 1	step 2	step 3
A	1	0.33	0.50	0.49
B	1	0.83	0.66	0.74
C	1	1.50	1.49	1.58
D	1	1.33	1.33	1.15
ϵ	-	0.41	0.08	0.09



Google PageRank

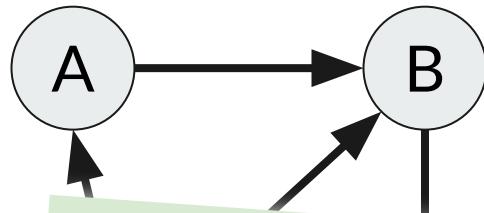


$$PR(u) = \sum \frac{PR(v)}{N_v}$$

	step 0	step 1	step 2	step 3	PageRank
A	1	0.33	0.50	0.49	4
B	1	0.83	0.66	0.74	3
C	1	1.50	1.49	1.58	1
D	1	1.33	1.33	1.15	2
ϵ	-	0.41	0.08	0.09	



Google PageRank

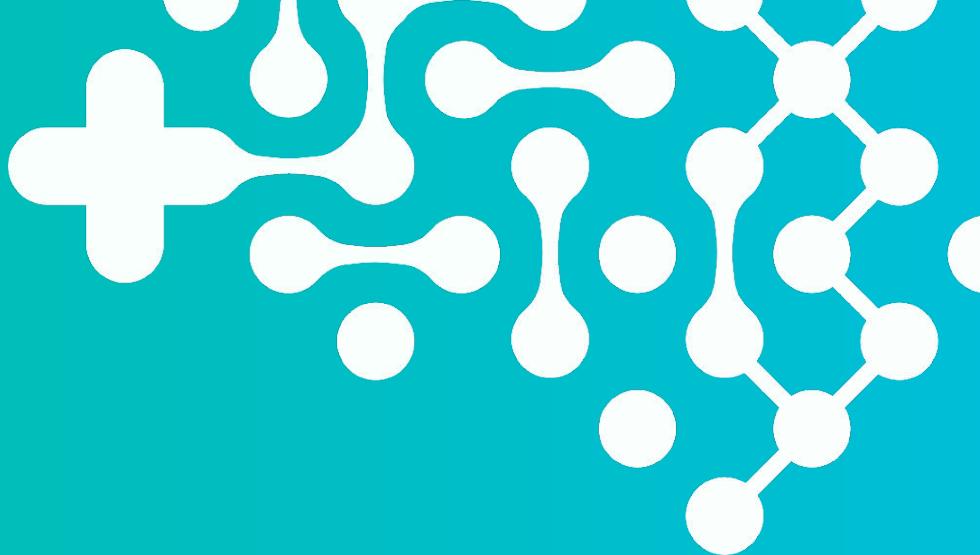


Aprendizado Não-Supervisionado

$$PR(u) = \sum \frac{PR(v)}{N_v}$$

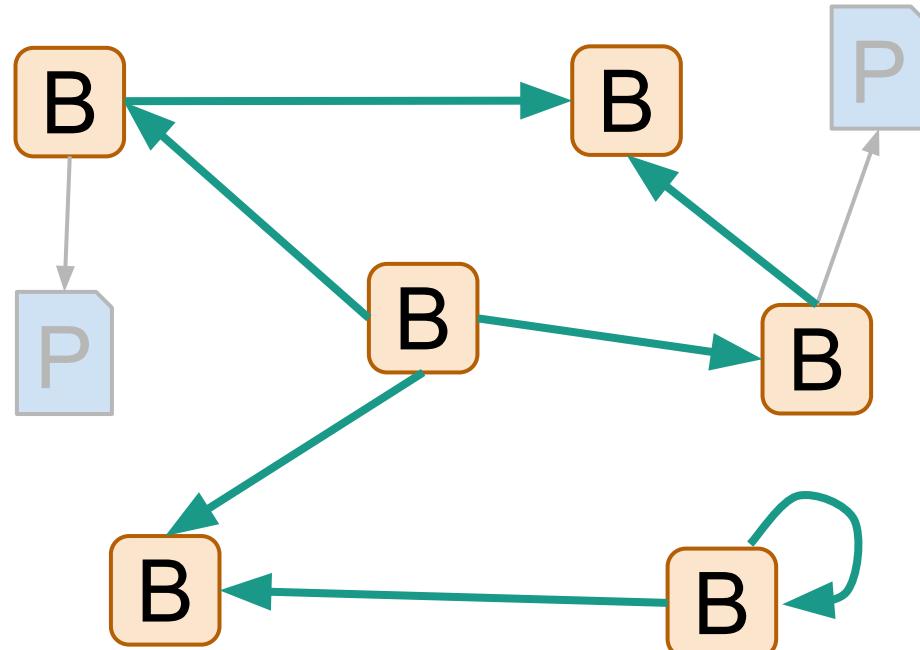
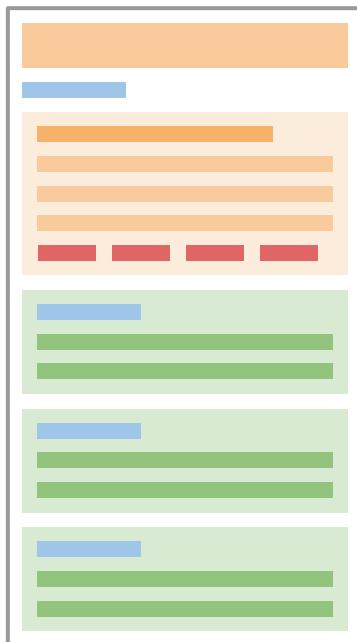
	step 0	step 1	step 2	step 3	PageRank
A	1	0.33	0.50	0.49	4
B	-	-	-	-	3
C	1	1.33	1.33	1.15	2
D	1	1.33	1.33	1.15	2
ϵ	-	0.41	0.08	0.09	

PageRank e outras aplicações



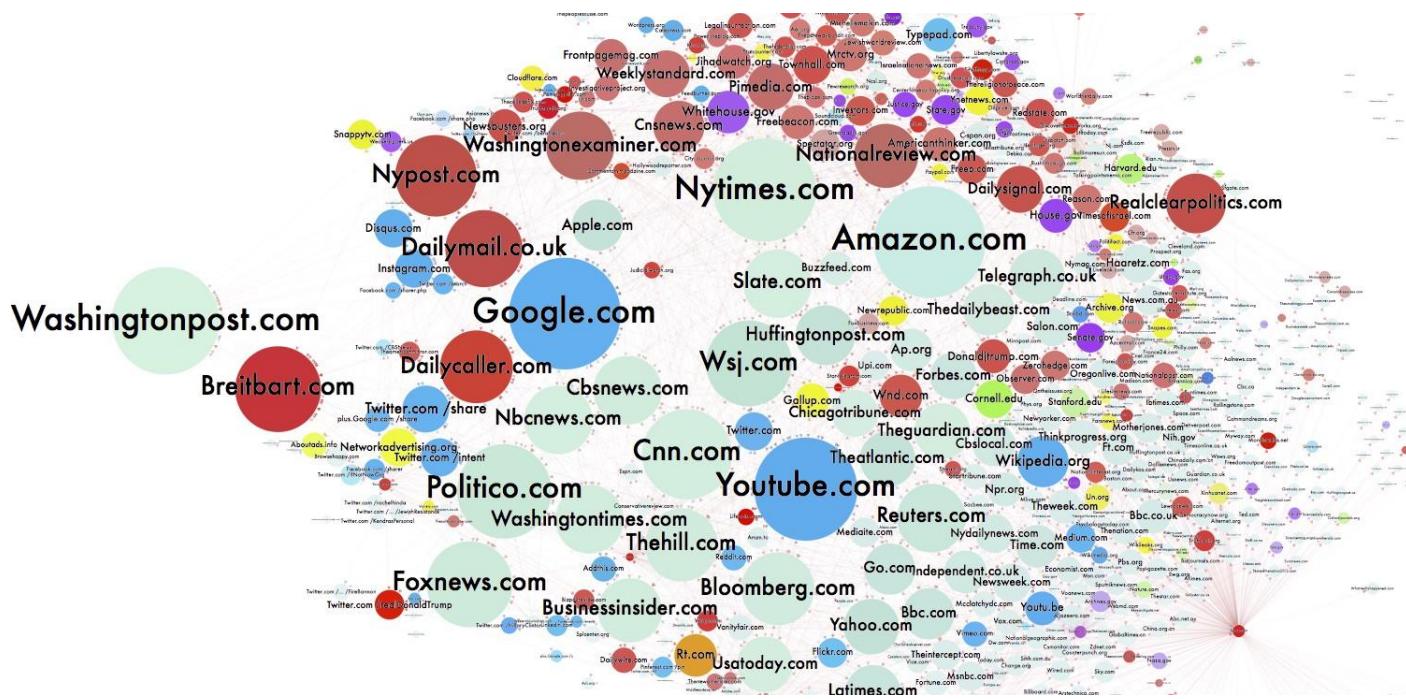


Autoridades em Blogs





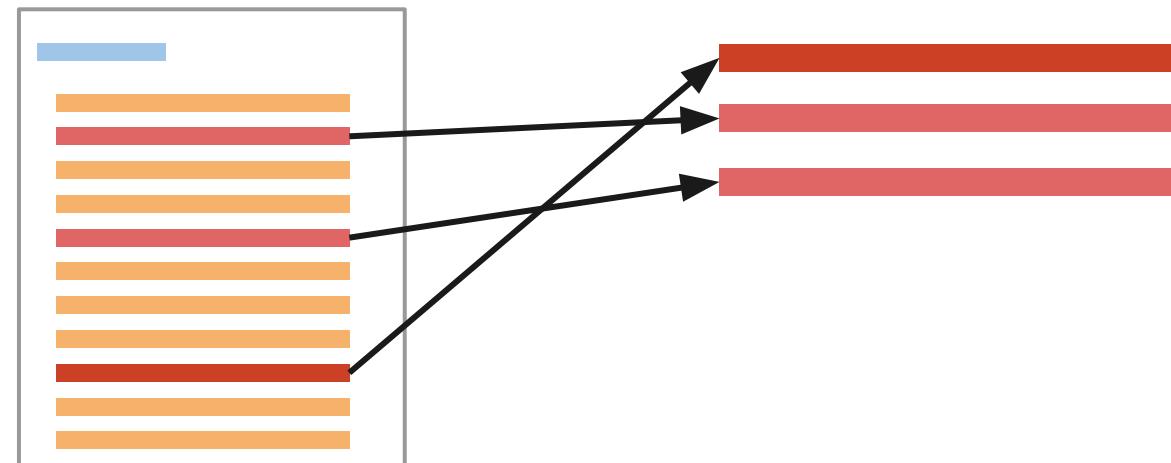
Identificando Fake News



Fonte: DistrustRank: Spotting False News Domains, WebSci'18

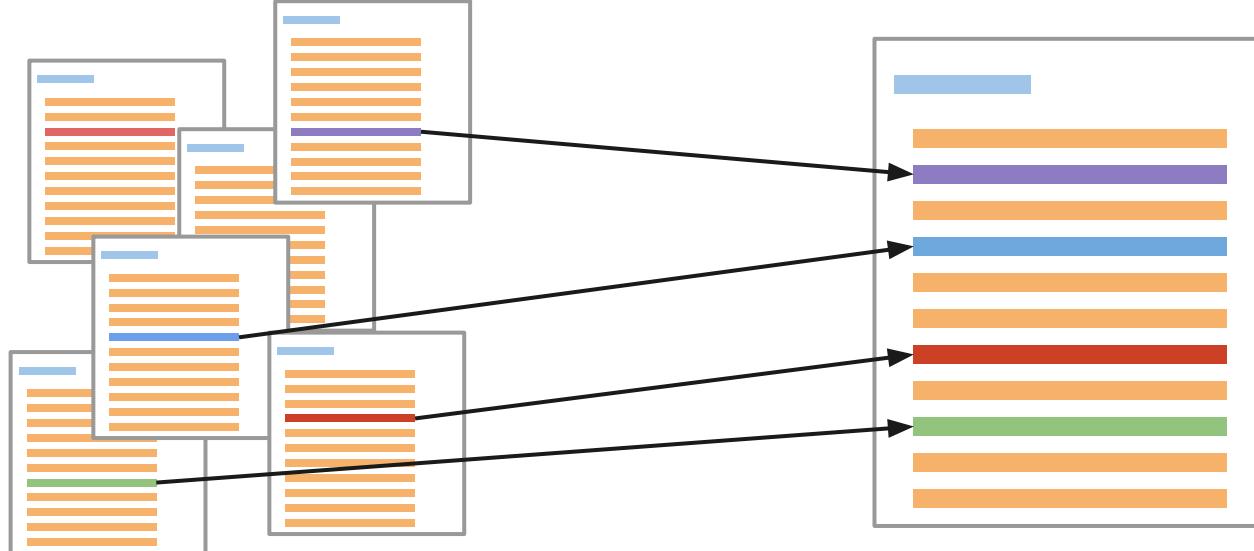


Sumarização de Texto



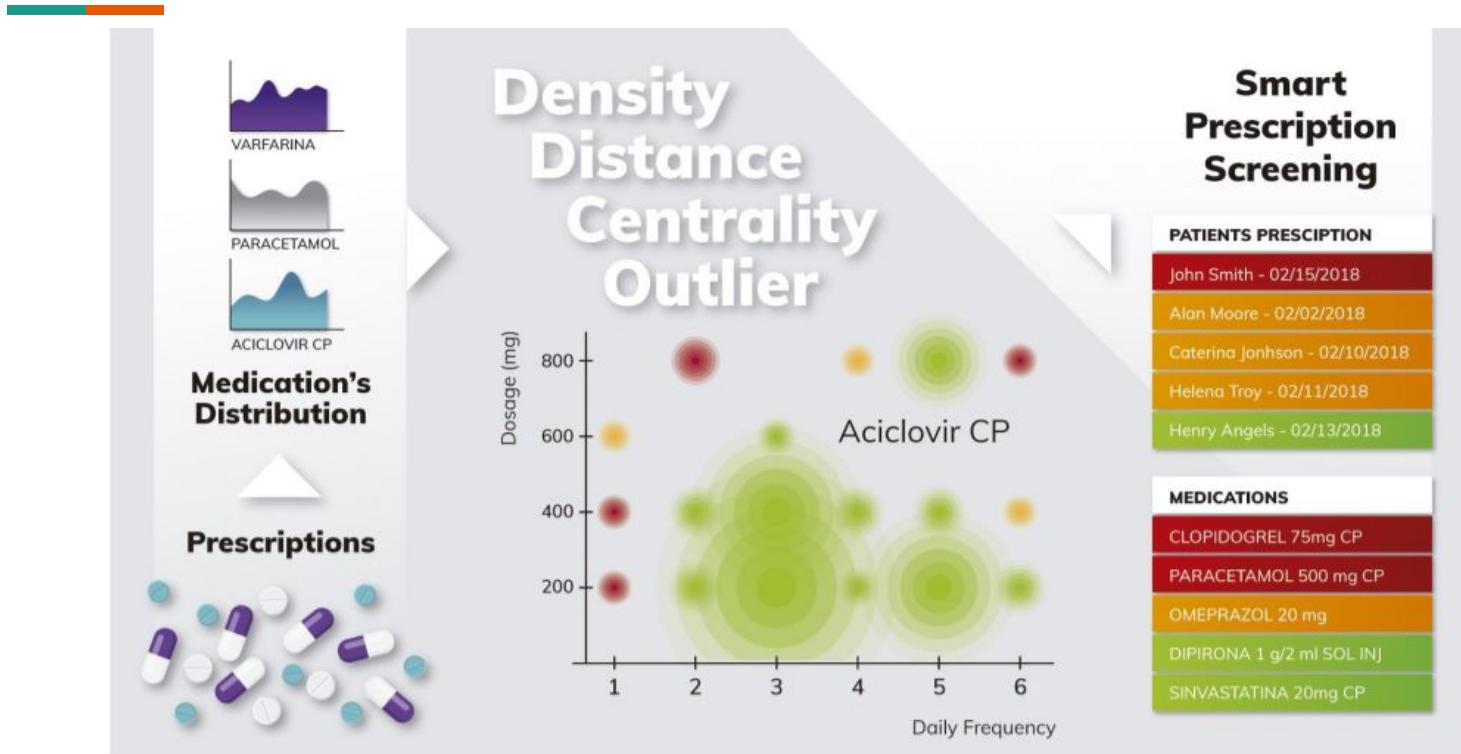


Ordenamento de Revisões





Prescrições Fora do Padrão

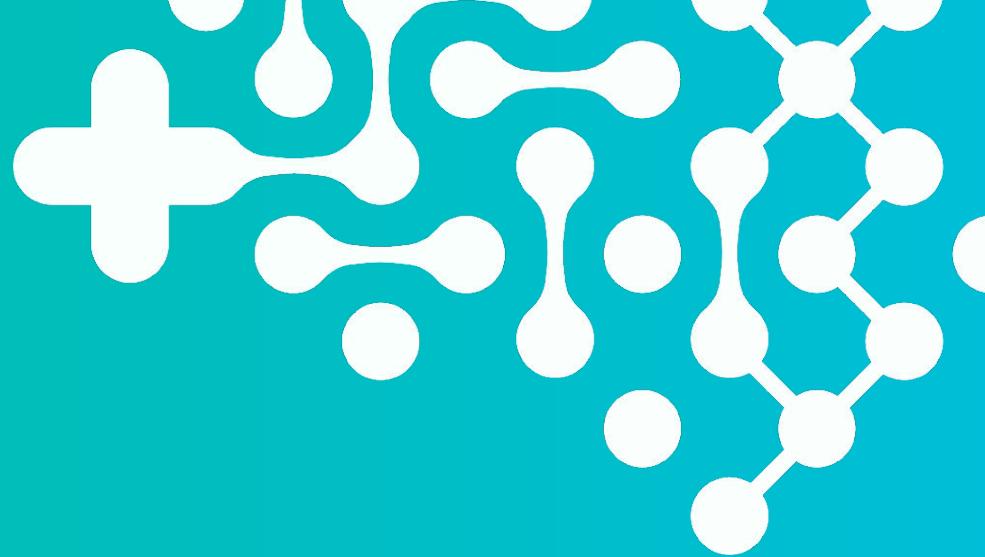


Fonte: [DDC-Outlier: Preventing Medication Errors Using Unsupervised Learning, JBHI 2018](#)



Tudo pode ser um Grafo

	Vértices	Arestas	PageRank
Web	páginas	links	relevância
Blogs	autores	comentários	autoridades
Textos	frases	similaridade	resumo
Reviews	texto	similaridade	qualidade
Prescrições	dose x freq	distância-Jacc	padrão



Prescrições Fora do Padrão

Erro de Medicação custam aprox.
\$42 bilhões de dólares anualmente*



World Health
Organization

who.int/patientsafety/medication-safety

* além de dano ou morte dos pacientes



Erros de Medicação - OMS

Prescrição

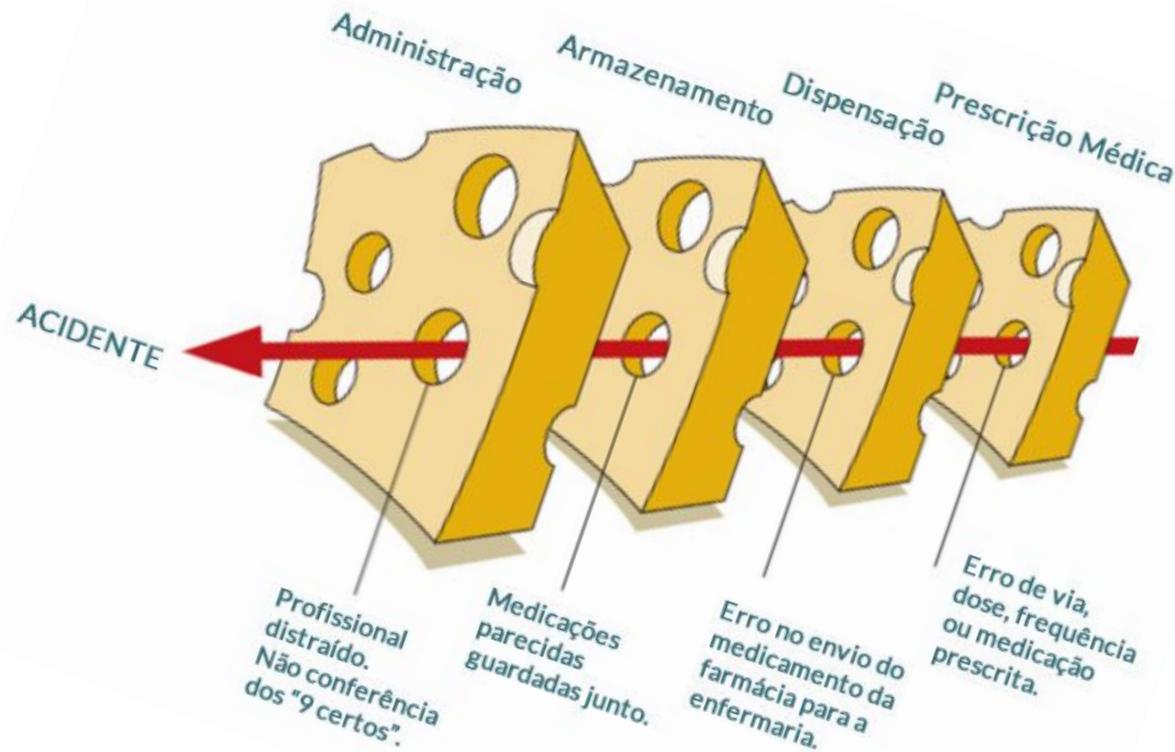
Triagem

Dispensação

Administração



Teoria do Queijo-Suiço





Erros de Medicação - OMS

Prescrição

Triagem

Dispensação

Administração

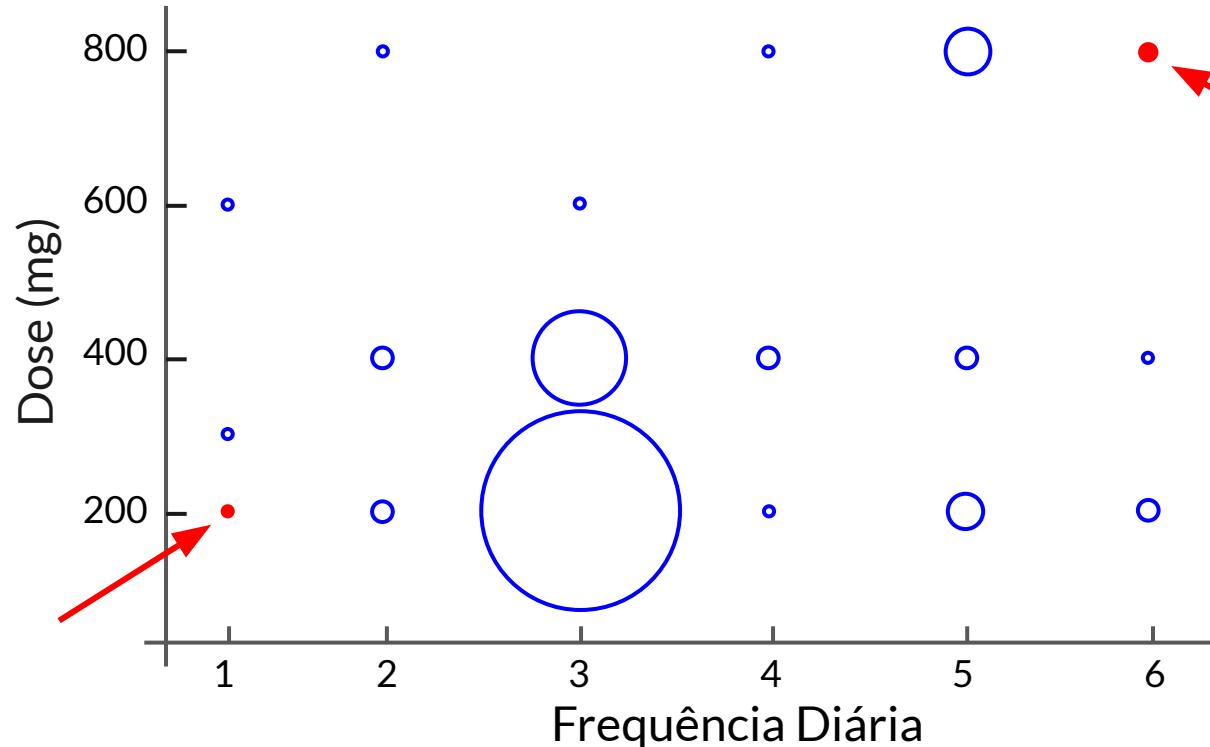


A large pile of green apples, with one red apple prominently placed in the center.

Triagem Farmacêutica

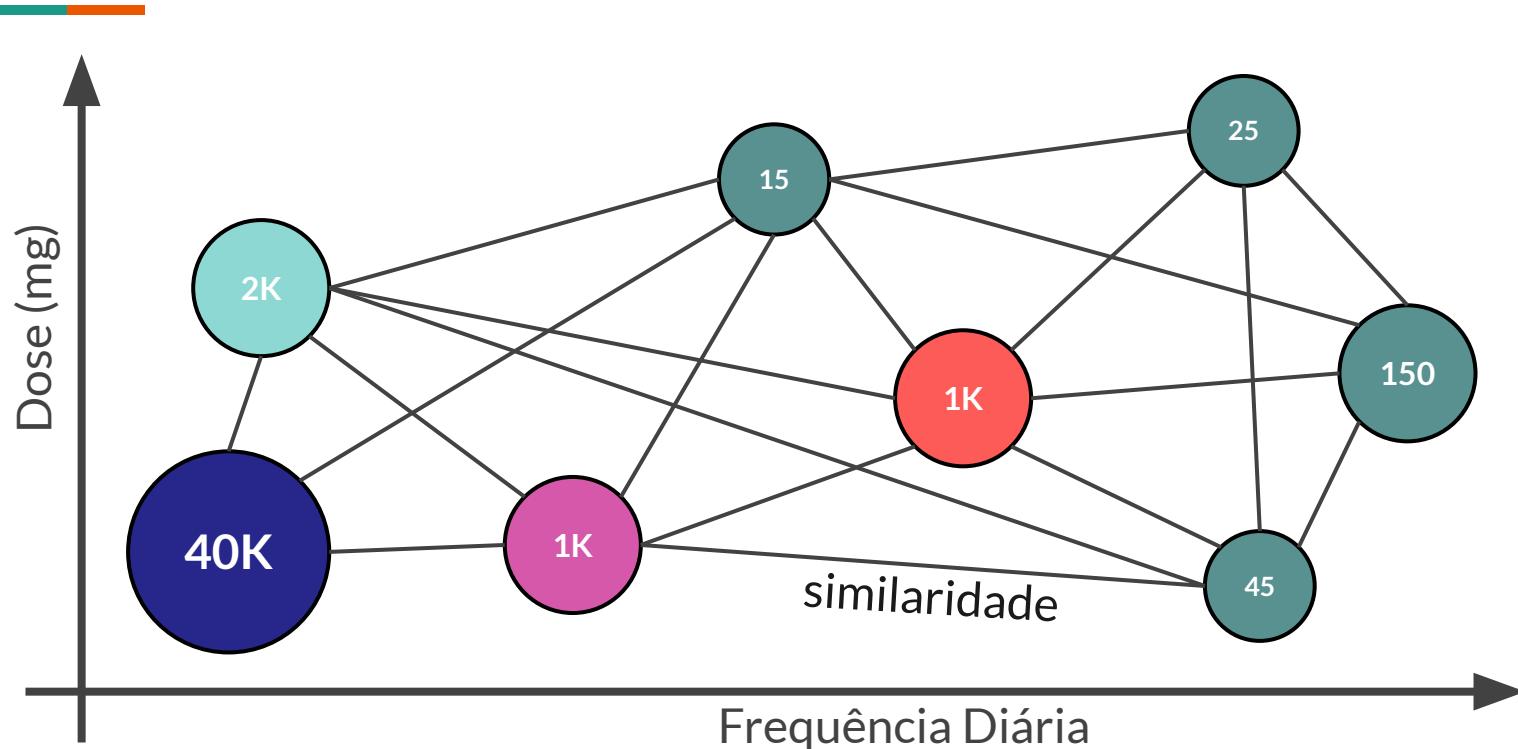


Detecção de Outliers (Aciclovir CP)



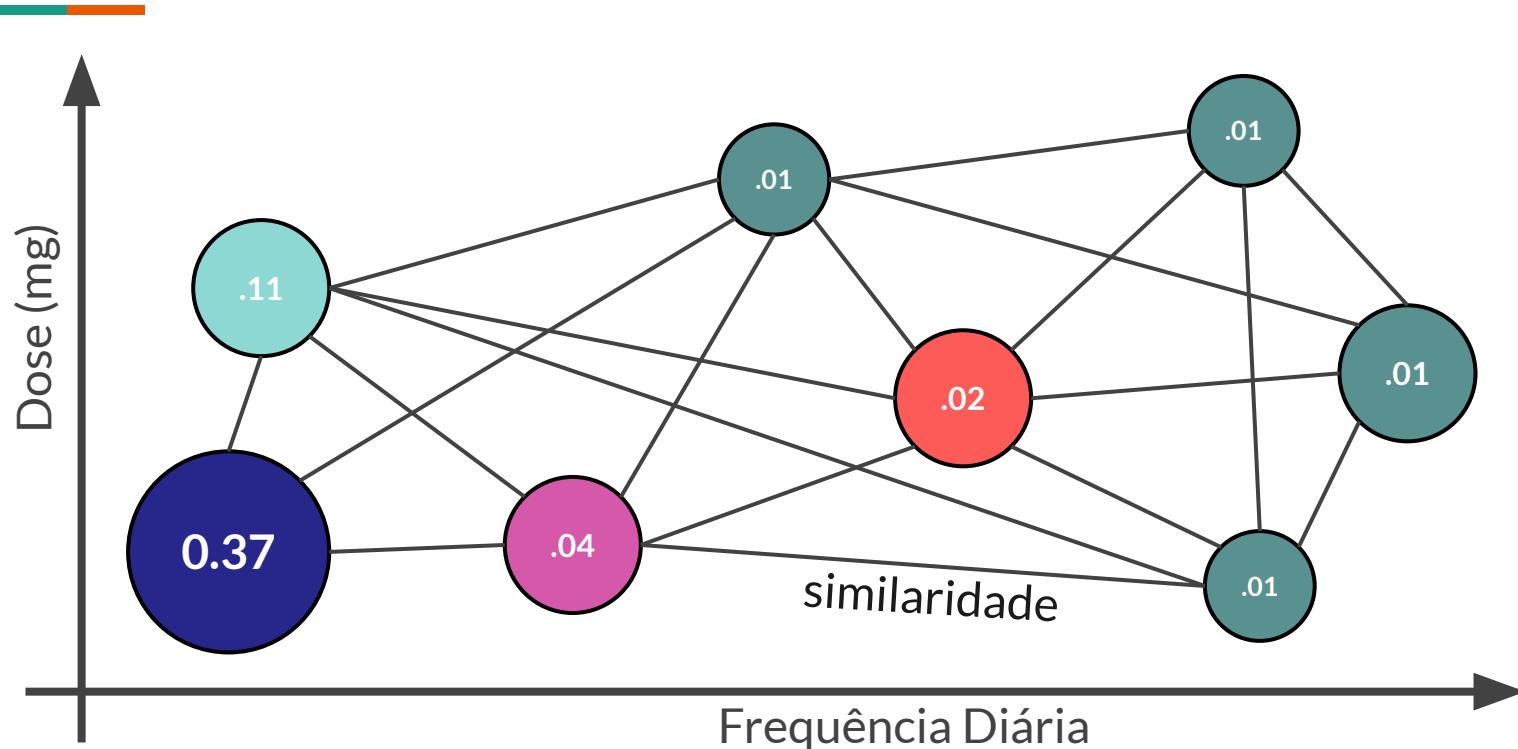


Detecção de Outliers (grafo de prescrições)



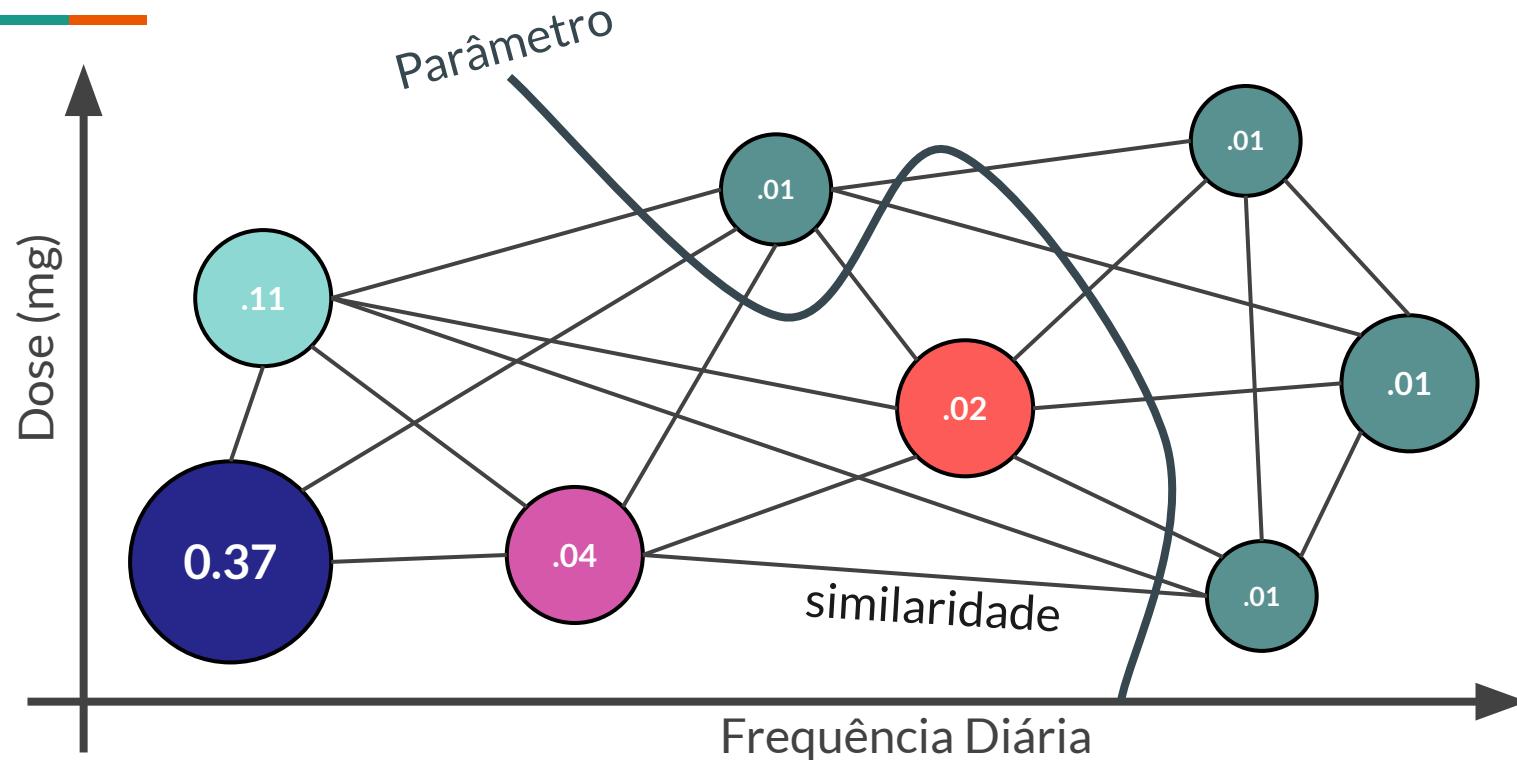


Detecção de Outliers (grafo de prescrições)





Detecção de Outliers (grafo de prescrições)

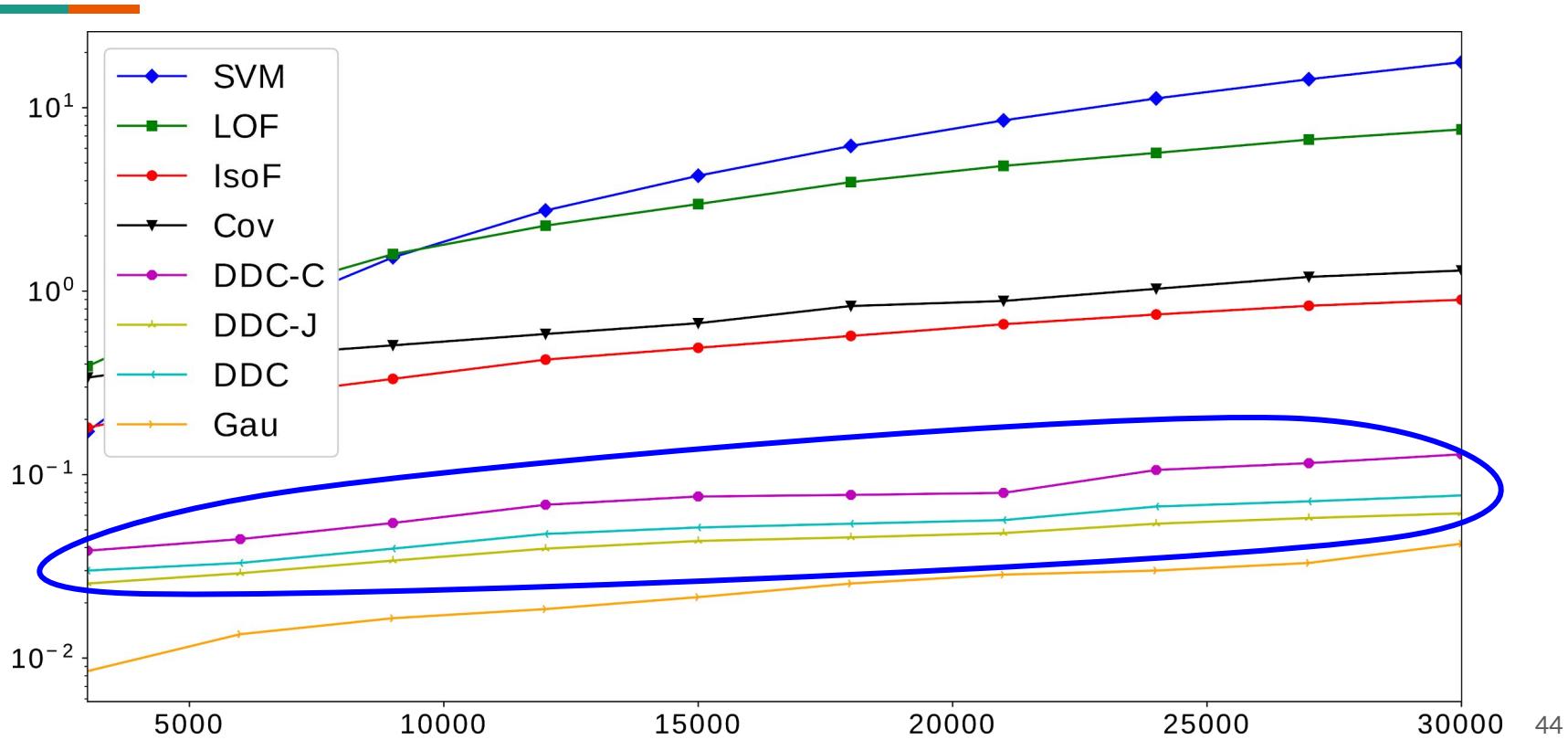




Comparação dos Algoritmos

Algoritmo	Recall	Precision	F-Measure	Top 3 ↑
DDC-J	0.90	0.61	0.68	31
Iso. Forest	0.91	0.52	0.61	26
DDC	0.86	0.50	0.58	21
SVM	0.94	0.39	0.48	20
DDC-C	0.72	0.54	0.51	19
Covariance	0.60	0.37	0.39	16
Gau	0.95	0.29	0.37	13
LOF	0.87	0.38	0.44	12

Detecção de Outliers (desempenho)





Identificando Prescrições Críticas

Medicamento	Dose	Freq	Score
CLOPIDOGREL 75mg CP	75 mg	8/8h	3
PARACETAMOL 500 mg CP	1000 mg	4/4h	2
OMEPRAZOL 20 mg	20 mg	AC	0
DIPIRONA 1 g/2 ml SOL INJ	1000 mg	6/6h	0
SINVASTATINA 20mg CP	20 mg	1x/noite	0



Erros Detectados (exemplos)

Overdose:

"Paracetamol 500mg CP" **1000mg 4/4h**

Frequência:

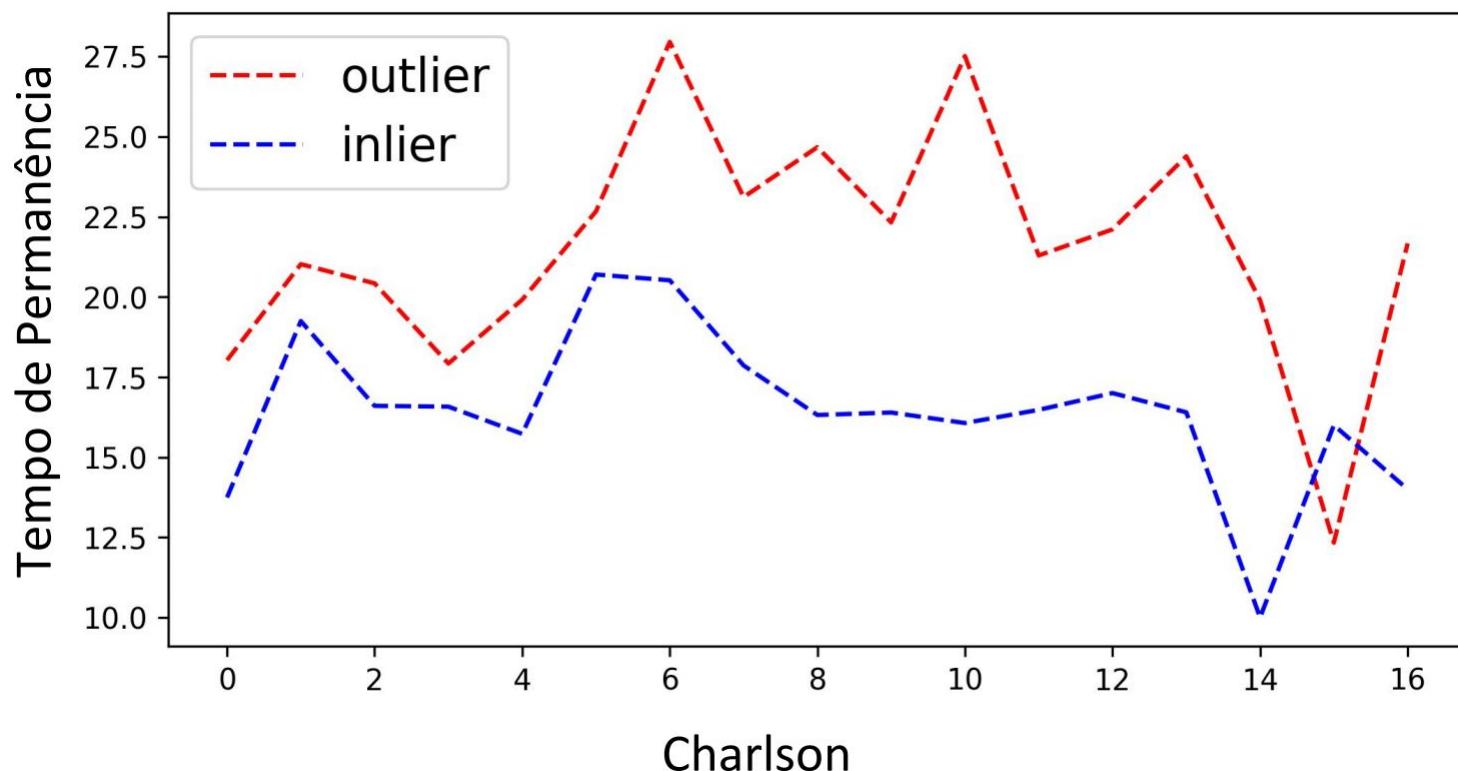
"Clopidogrel 75mg CP" **75mg 8/8h**

Apresentação:

"Hidralazina 50mg DG" **25mg 1x/dia**



Tempo de Internação



Smart Prescription Screening

PATIENTS PRESCRIPTION

John Smith - 02/15/2018

Alan Moore - 02/02/2018

Caterina Jonhson - 02/10/2018

Helena Troy - 02/11/2018

Henry Angels - 02/13/2018

MEDICATIONS

CLOPIDOGREL 75mg CP

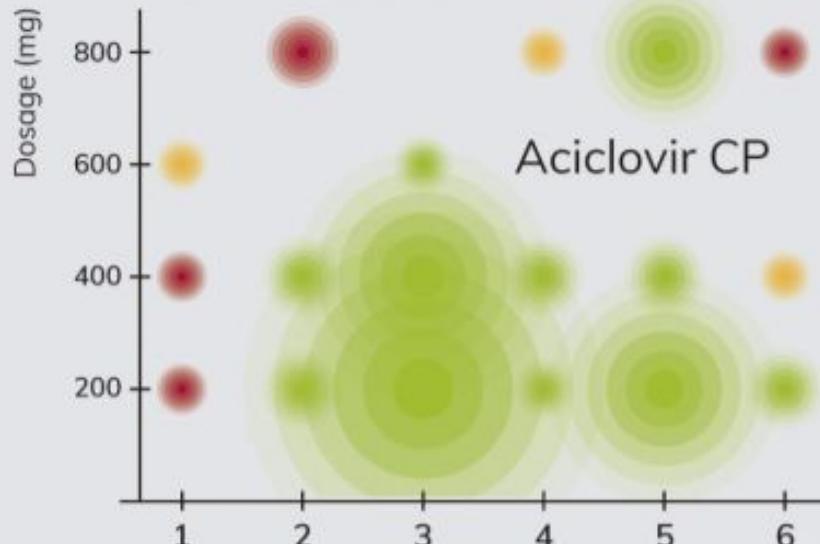
PARACETAMOL 500 mg CP

OMEPRAZOL 20 mg

DIPIRONA 1 g/2 ml SOL INJ

SINVASTATINA 20mg CP

Density Distance Centrality Outlier



Repetibilidade

Replicabilidade

Reprodutibilidade

Medication's Distribution

Prescriptions



PUCRS



Henrique Dias

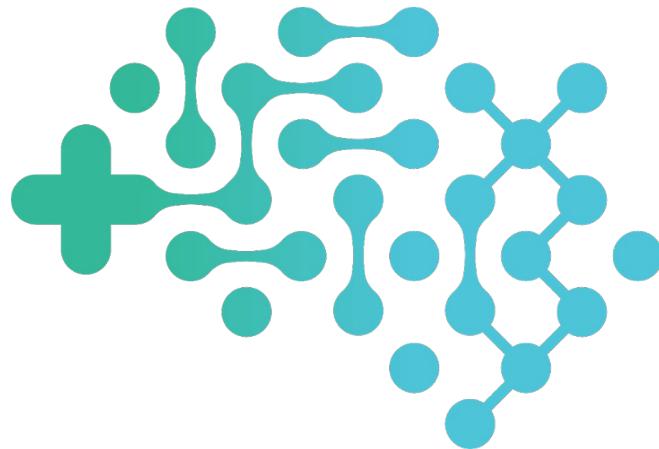
PhD* Ciência da Computação



Ana Helena Ulbrich

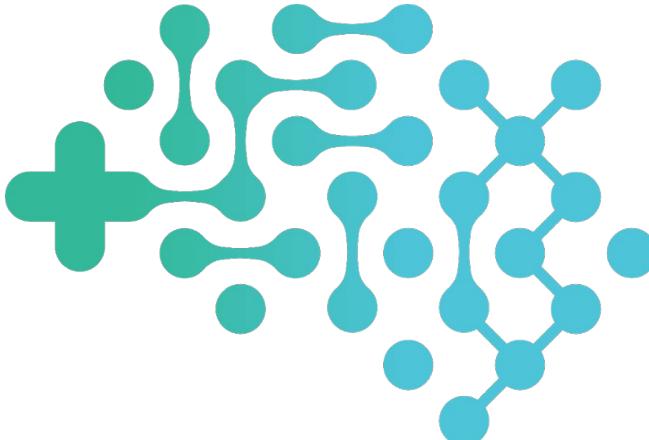
PhD Farmacêutica

PUCRS



Grupo de
Inteligência Artificial
na Saúde

<http://www.inf.pucrs.br/ia-saude/>



SÍRIO-LIBANÊS



Grupo de
Inteligência Artificial
na Saúde

<http://www.inf.pucrs.br/ia-saude/>



Projeto '17



Detecção Automática De Eventos Adversos Utilizando
Processamento De Linguagem Natural Nos Prontuários
Eletrônicos De Um Hospital Terciário

Henrique D. P. dos Santos, Ana Helena D. P. S. Ulbrich,
Graziella G. Baiocco

Publicação '18

J-BHI

Journal of Biomedical
and Health Informatics



IEEE

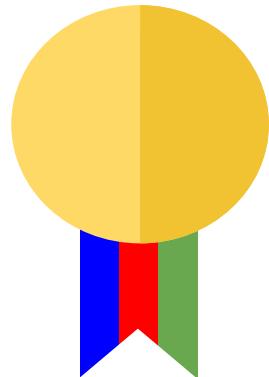
DDC-Outlier: Preventing Medication Errors Using Unsupervised Learning

Henrique D. P. dos Santos, Ana Helena D. P. S. Ulbrich,
Vinícius Woloszyn, Renata Vieira

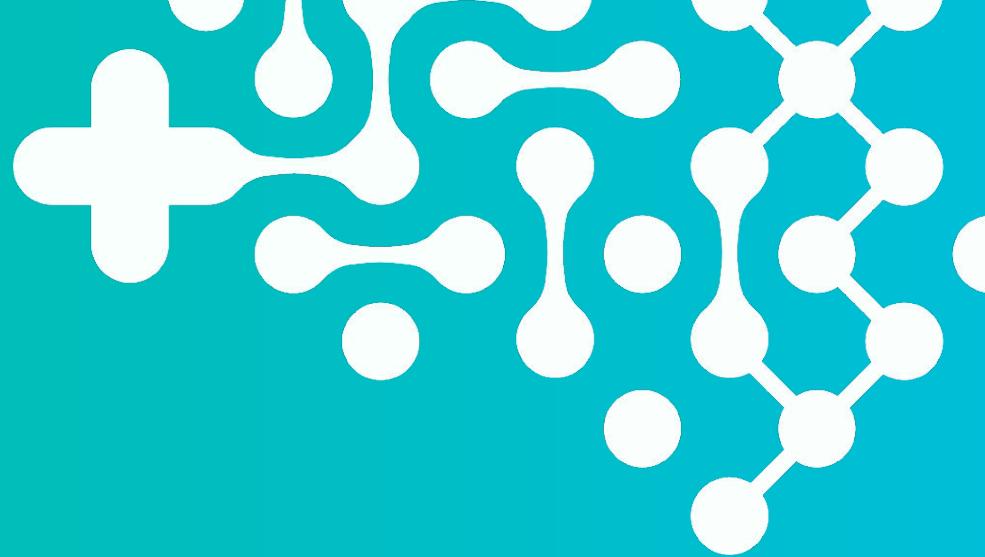
Premiações '18 e '19



Concurso de Inovação
na Saúde promovido
pelo Hospital
Sírio-Libanês e a
Fundação Everis
2018



Google
Research
Award
2018 e 2019



**De: Teoria
Para: Prática**



noharm.ai
CUIDANDO DOS PACIENTES



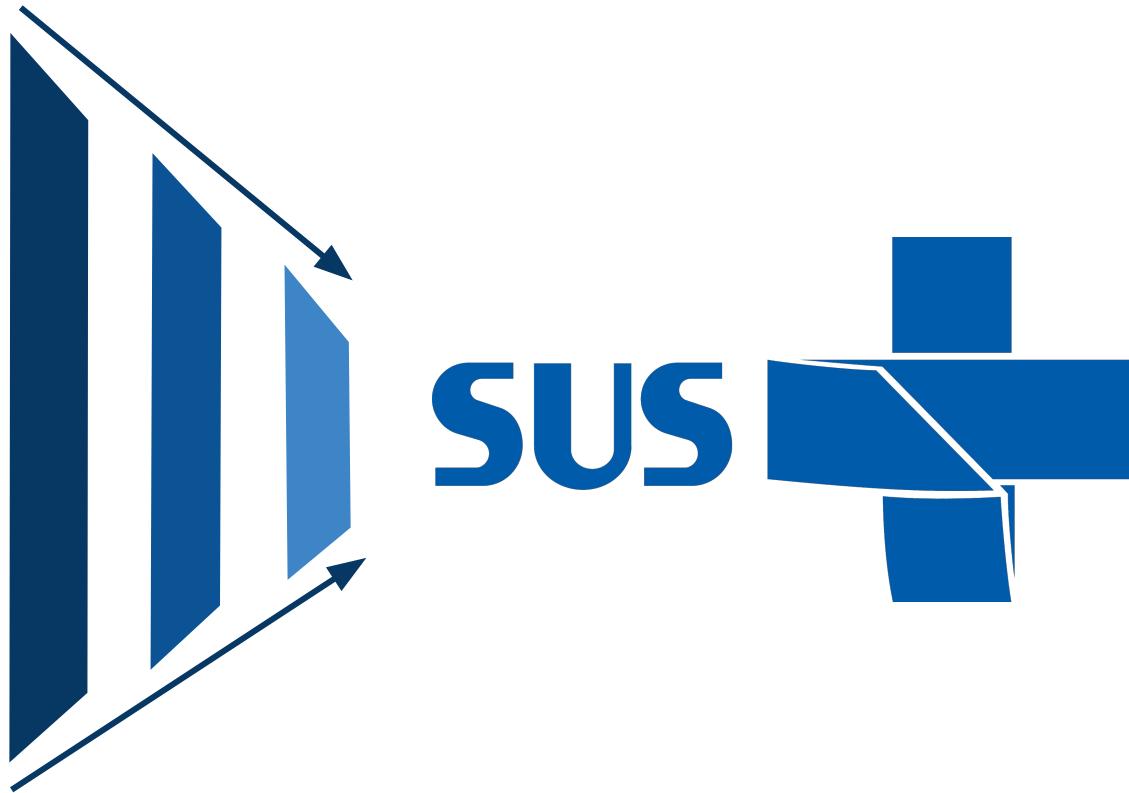
noharm.ai
CUIDANDO DOS PACIENTES

Startup sem fins lucrativos

<https://github.com/noharm-ai>

Leitos
Privados
(receita)

Dedução
Fiscal
(doações)





1260 leitos e 71 especialidades médicas, taxa de ocupação média de 84% e tempo médio de permanência de 6,7 dias



331 leitos e 49 especialidades médicas, taxa de ocupação média de 84% e tempo médio de permanência de 6,27 dias

Voluntários 30+



Henrique
Infra / Backend / ETL



Ana Helena
Farmácia Clínica



Fábio
Médico / PM / UX



Marcelo
Backend / Mobile



Rodrigo
Comercial



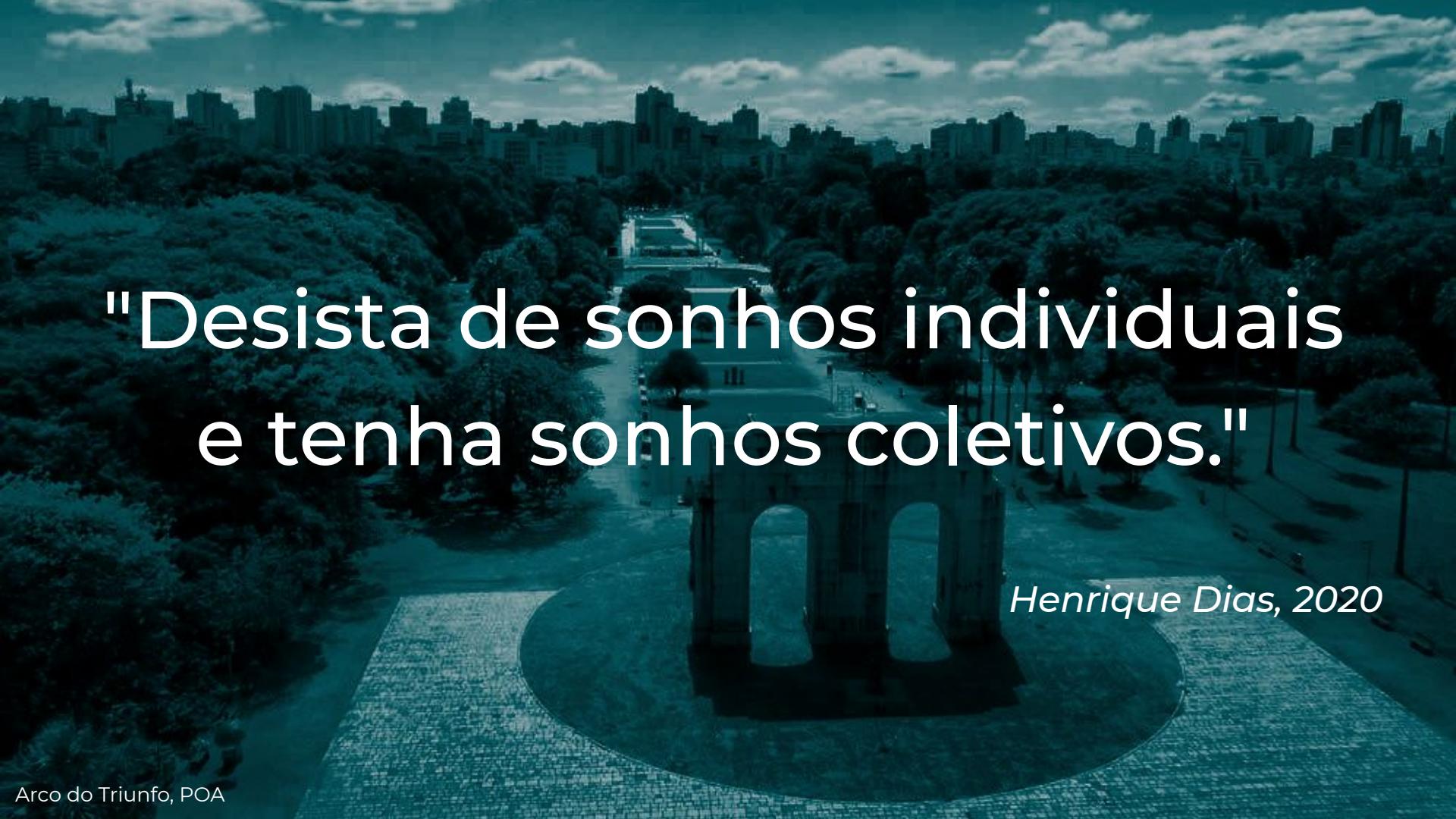
Augusto
ETL



Júlia
ETL / MV



Gabrielli
Farmácia Clínica

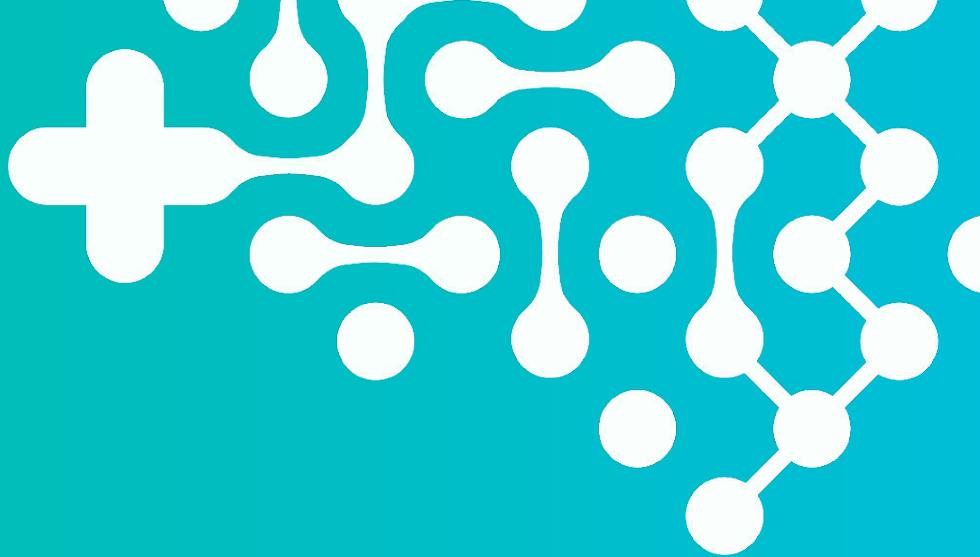


"Desista de sonhos individuais
e tenha sonhos coletivos."

Henrique Dias, 2020

Obrigado! Dúvidas?

Henrique Dias - PUCRS
henrique@gemeos.org



meetup
nubank
Janeiro 2020

Alinhamento dos Astros





Artigos Publicados

J-BHI'18: DDC-Outlier: Preventing medication errors

CBMS'18: Charlson comorbidity index Regression

AIH'18: MeSHx-Notes: Web System for Clinical Notes

BIBE'19: Fall Detection in EHR using Deep Learning



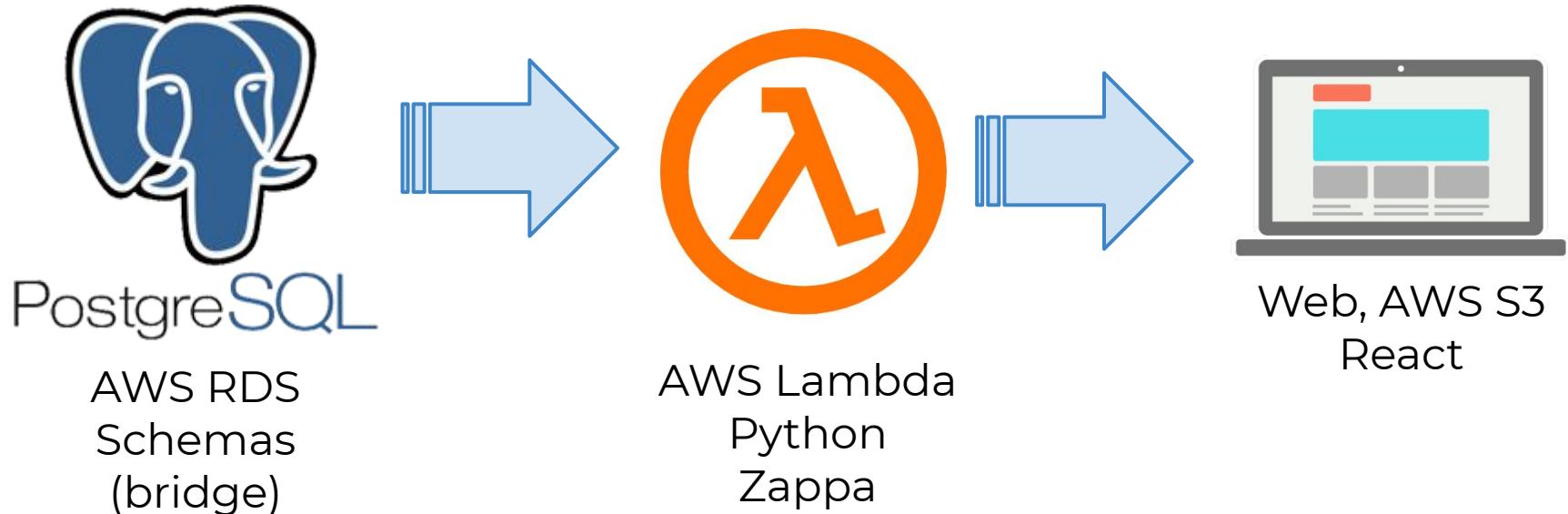
Código Aberto

J-BHI'18: <https://github.com/nlp-pucrs/prescription-outliers>

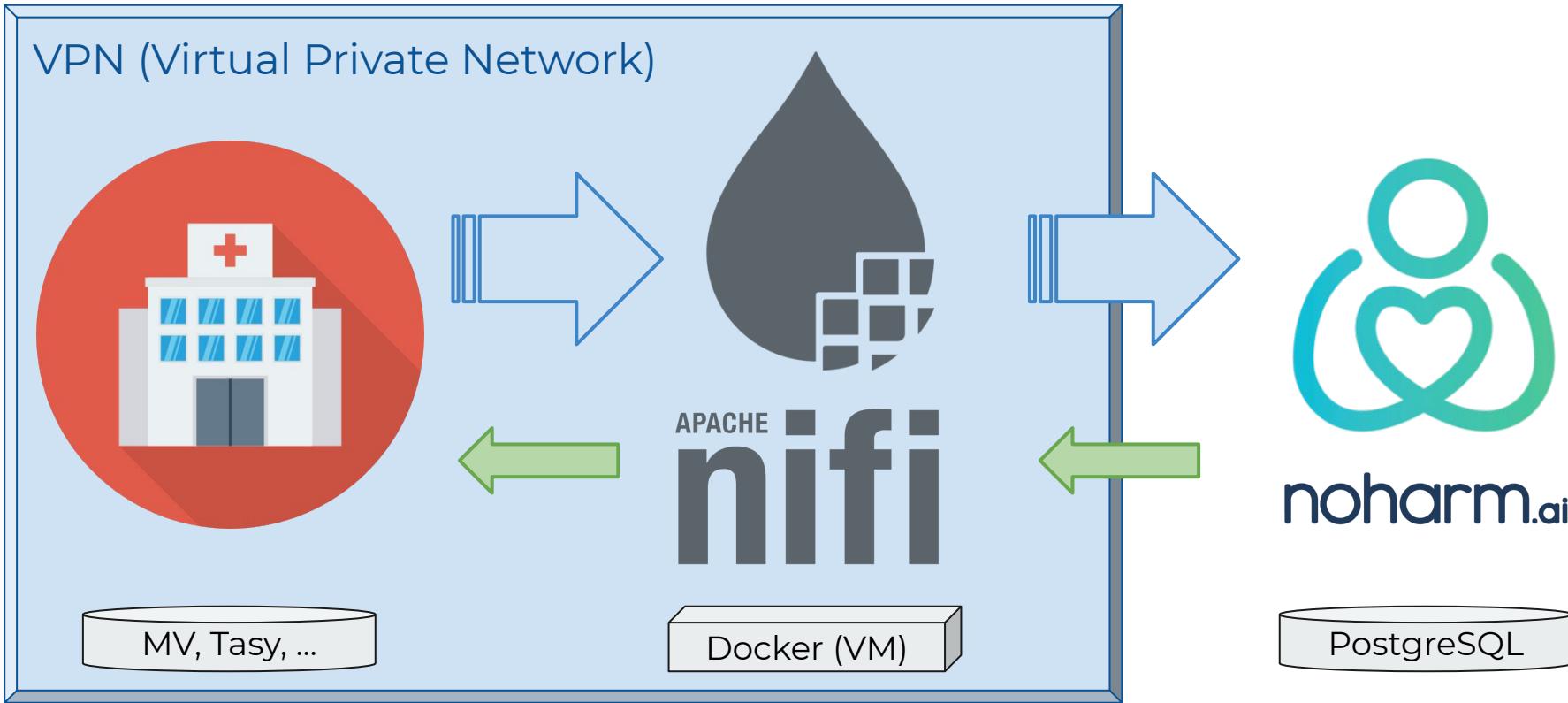
CBMS'18: <https://github.com/nlp-pucrs/cci-regression>

AIH'18: <https://github.com/nlp-pucrs/meshx-notes>

BIBE'19: <https://github.com/nlp-pucrs/fall-detection>



Integração de Dados





Inteligência Artificial na Saúde no Brasil



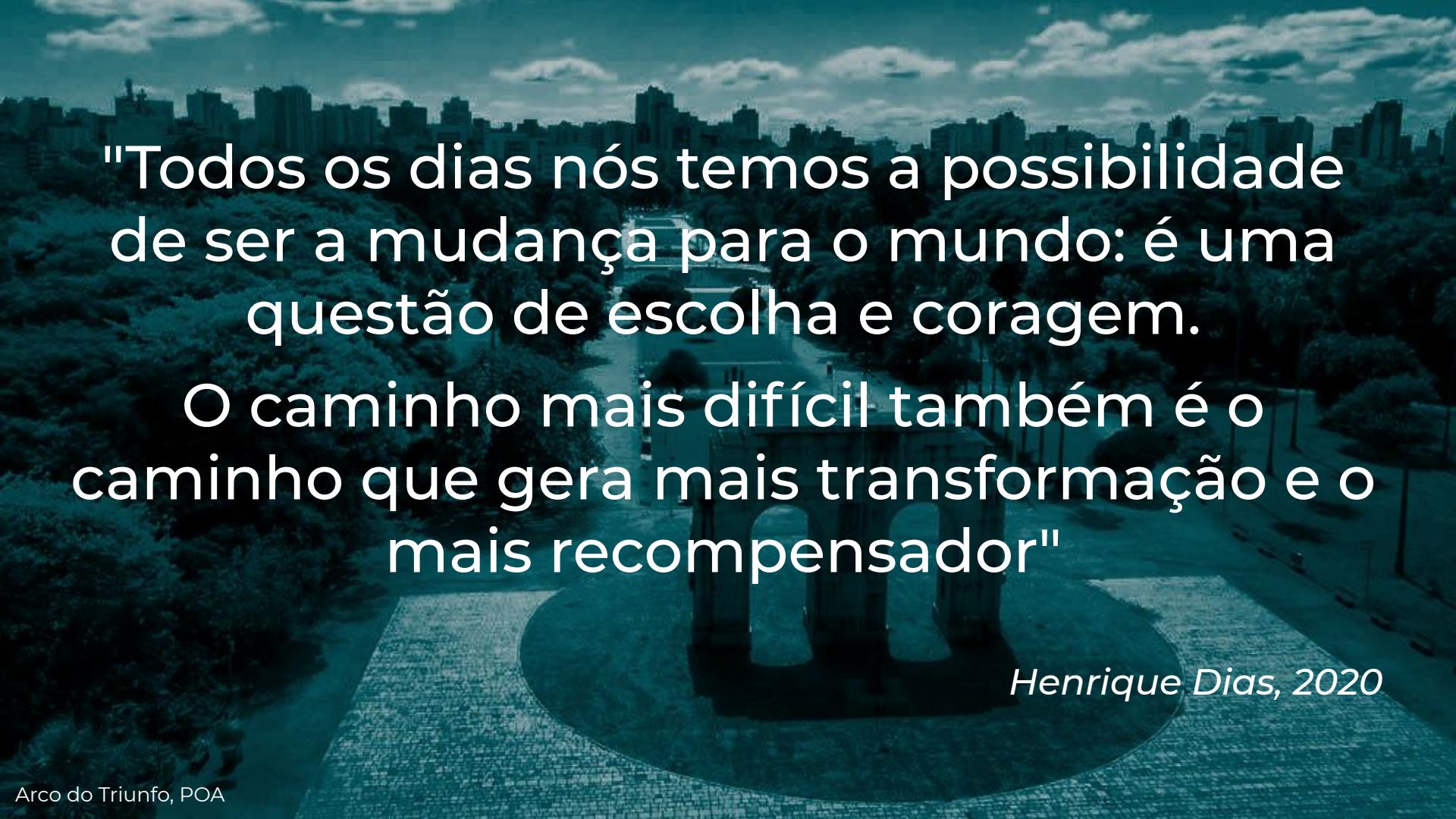
health
innova.**hub**





Outros Links

- [Oficina Gratuita de Machine Learning \(inscrições\)](#)
- [Ciência de Dados na Saúde \(vídeo\)](#)
- [Redes Neurais Self-Attention \(vídeo original\)](#)
- <http://henrique.gemeos.org>
- <http://noharm.ai>

The background of the slide is a photograph of a city skyline at dusk or night, viewed from an elevated perspective. In the foreground, a large, illuminated bridge spans a body of water. The city lights are visible through the trees and buildings in the distance.

"Todos os dias nós temos a possibilidade de ser a mudança para o mundo: é uma questão de escolha e coragem.

O caminho mais difícil também é o caminho que gera mais transformação e o mais recompensador"

Henrique Dias, 2020