Aprendizado Não-Supervisionado na Saúde

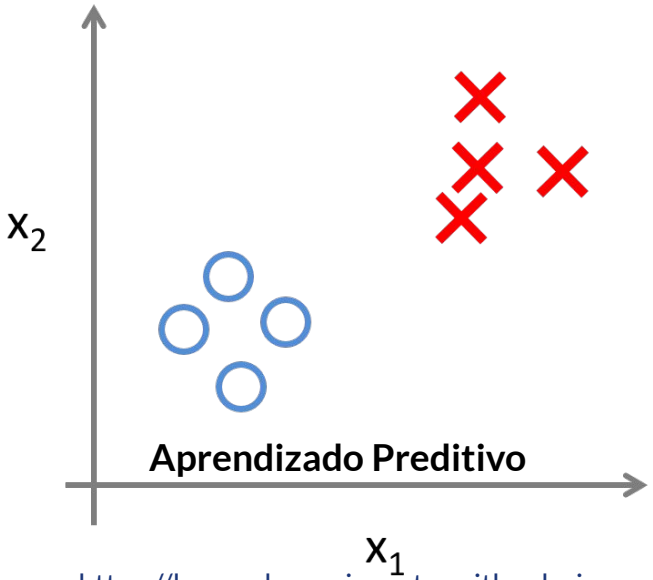
Henrique Dias - PUCRS
Grupo de Inteligência Artificial na Saúde

Slides: **http://goo.gl/85hzcX** (download pdf)

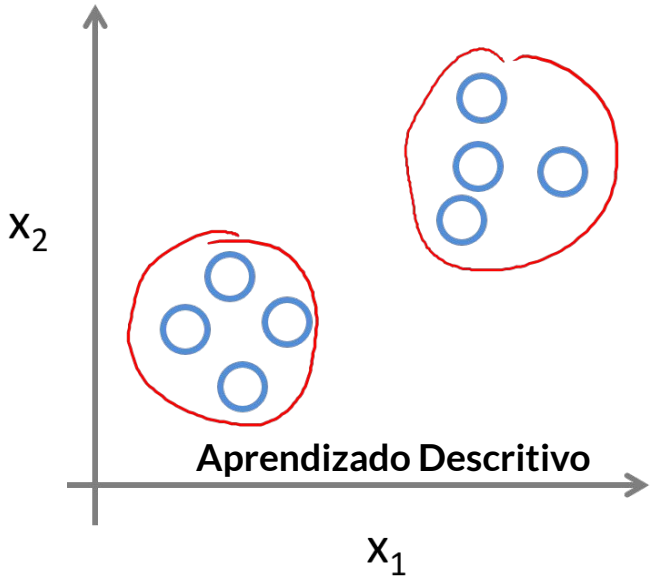
meetup
nubank
Janeiro 2020

Recapitulando...

Supervised Learning

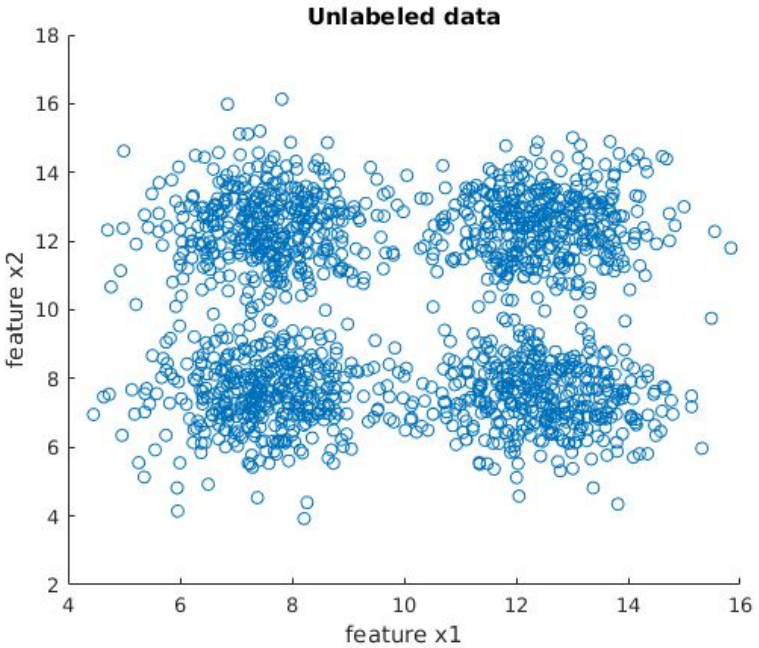


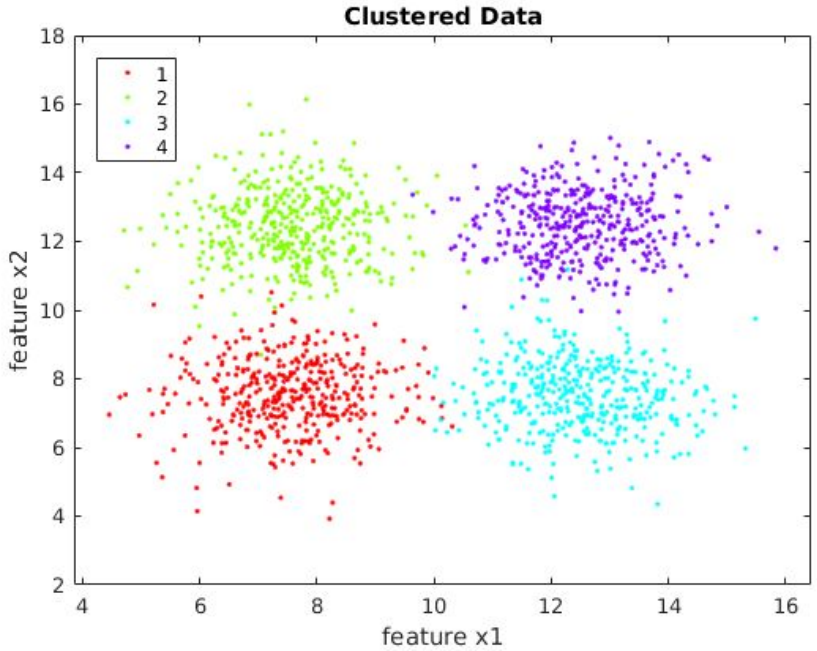
Unsupervised Learning



Imagens: https://leonardoaraujosantos.gitbooks.io

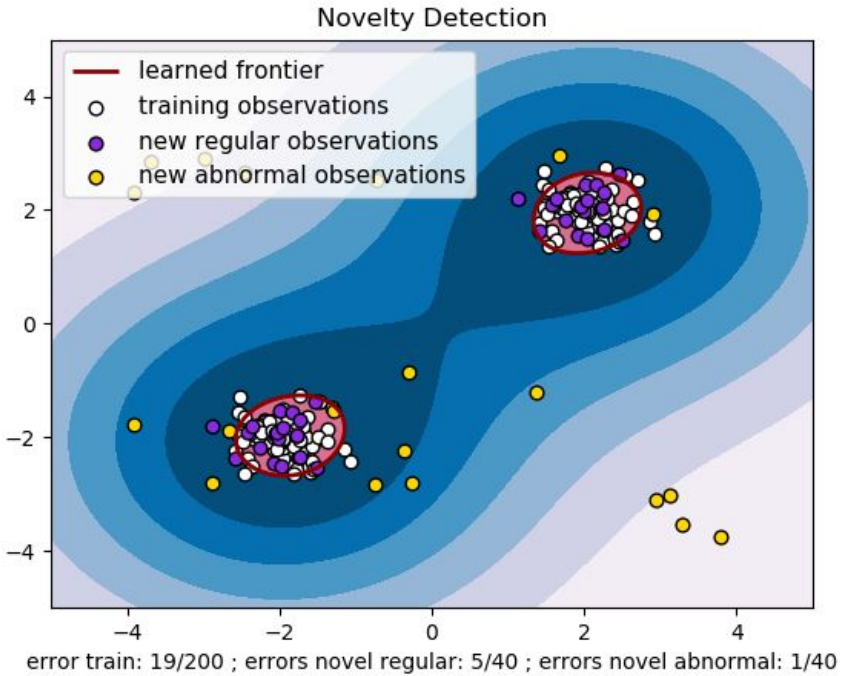
Não-Supervisionado (agrupamento)

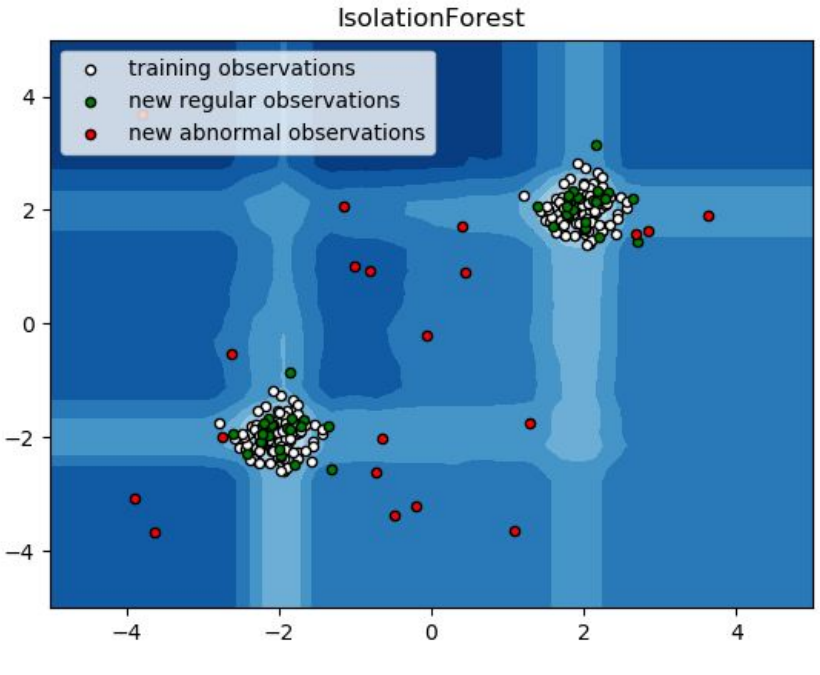




Imagens: https://leonardoaraujosantos.gitbooks.io

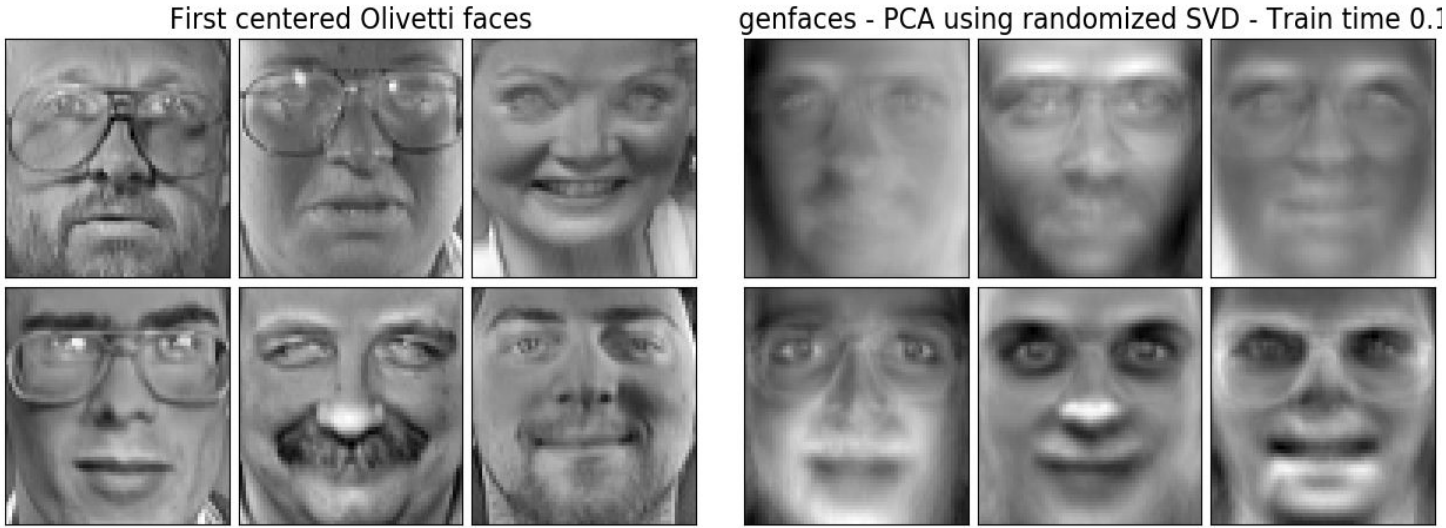
Não-Supervisionado (detecção de anomalia)





Imagens: http://scikit-learn.org/stable/modules/outlier_detection.html

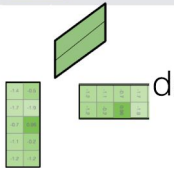
Não-Supervisionado (reduc. de dimensionalidade)



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

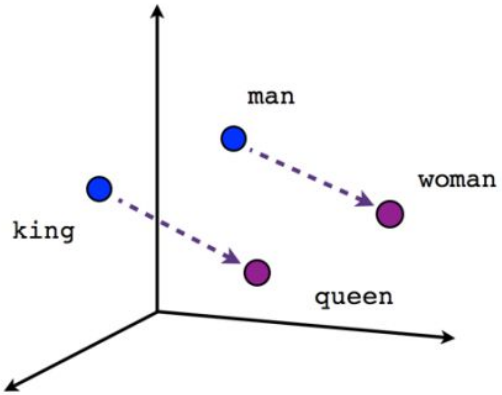
Imagens: http://scikit-learn.org/stable/modules/decomposition.html

Não-Supervisionado (word weight)

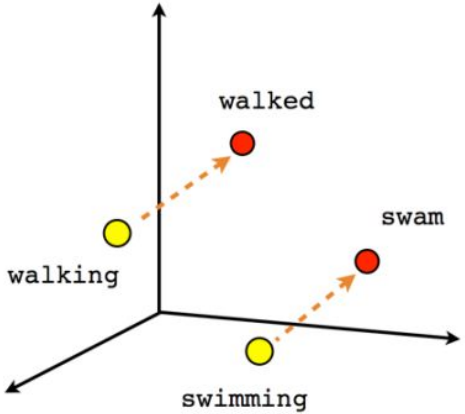
TF-IDF	$M_{ij} = \frac{\#\ of\ word\ i}{\#\ of\ word\ i\ in\ doc\ j}$	 document-word
LSI / LSA	$SVD(M_{ij})$	 document-word
LDA	LSI + Dirichlet Prior	 document-word
word2vec	$M_{ij} = \frac{\#(i,j)/n_{pairs}}{\#(i)/n_{words}\#(j)/n_{words}}$ $SVD(M_{ij})$	 word-word
word2tensor	$M_{ijk} = \frac{\#(i,j,k)/n_{triplets}}{\#(i)/n_{words}\#(j)/n_{words}\#(k)/n_{words}}$ $SVD(M_{ijk})$	 document-word-word

Imagens: https://multithreaded.stitchfix.com/blog/2017/10/18/stop-using-word2vec/

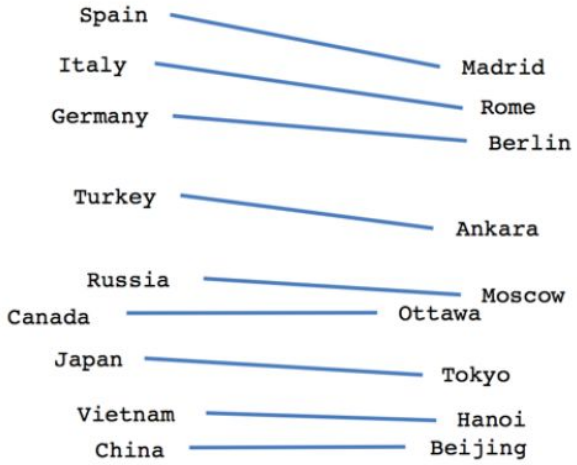
Não-Supervisionado (word embeddings)



Male-Female

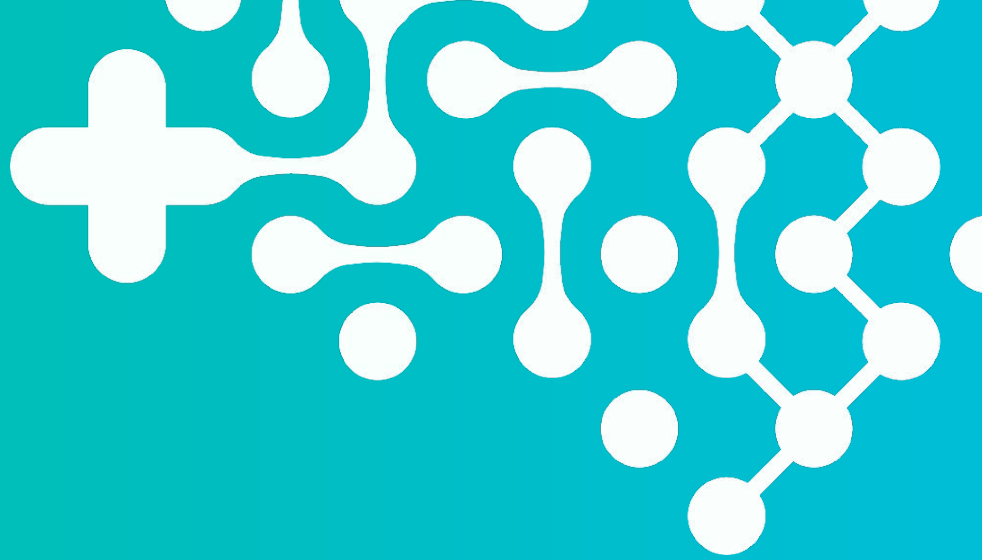


Verb tense



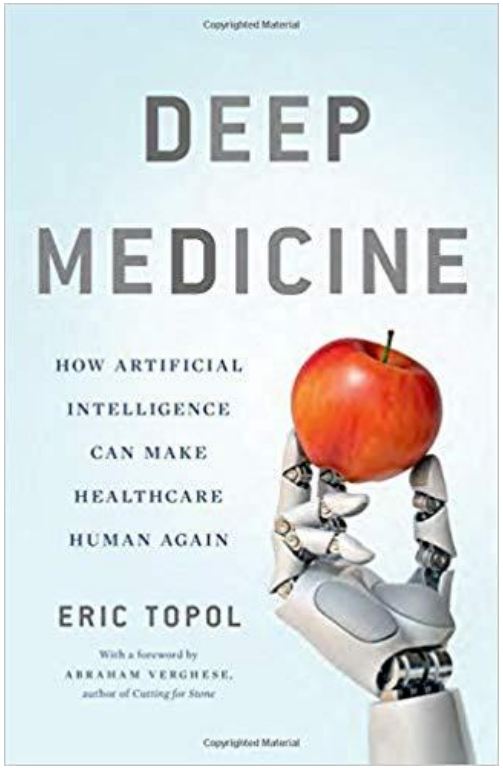
Country-Capital

Imagens: https://towardsdatascience.com/deep-learning-4-embedding-layers-f9a02d55ac12, Mikolov 2013



… e na saúde?

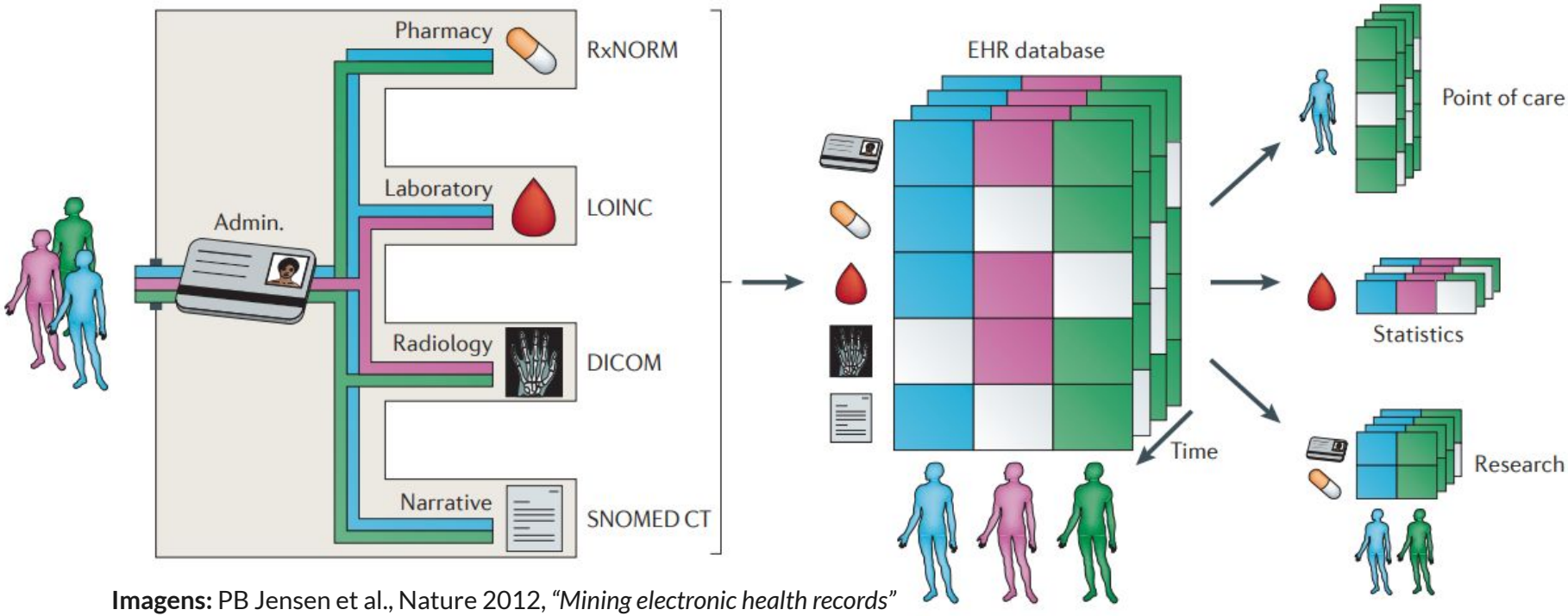
Inteligência Artificial na Saúde



"How Artificial Intelligence can make Healthcare Human again"

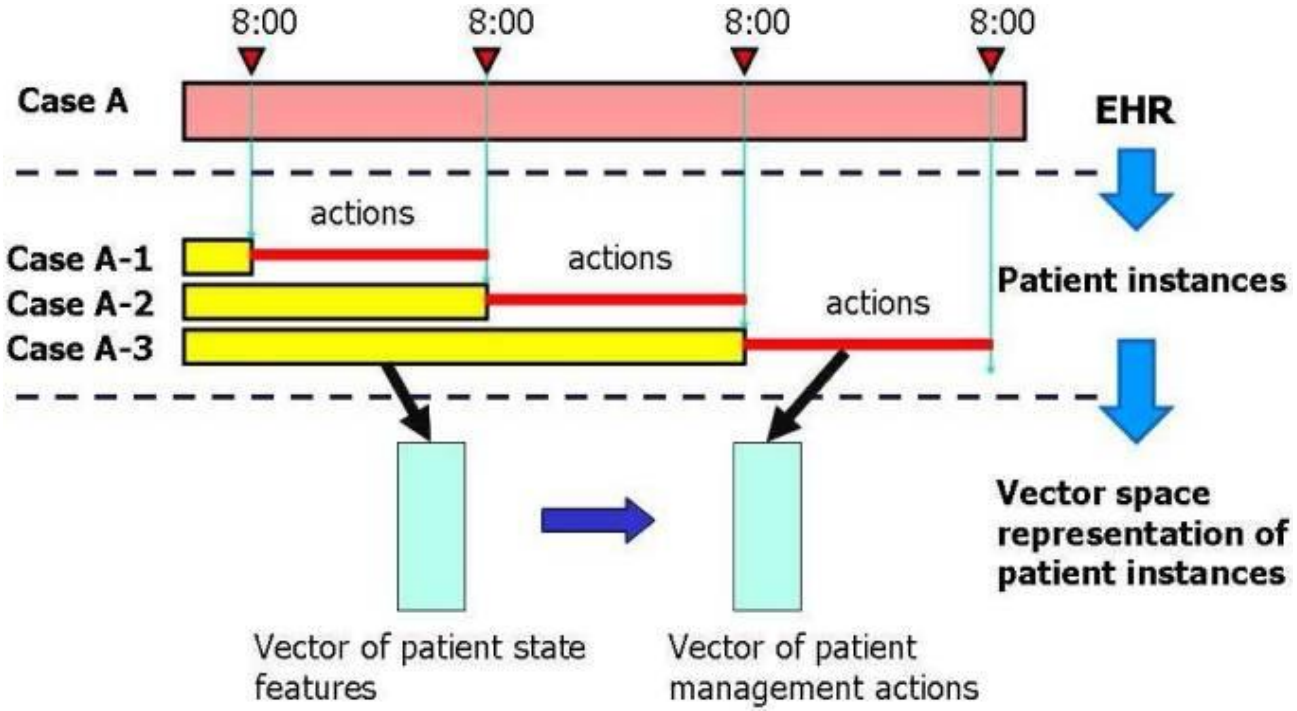
Eric Topol

Dados Básicos de Saúde



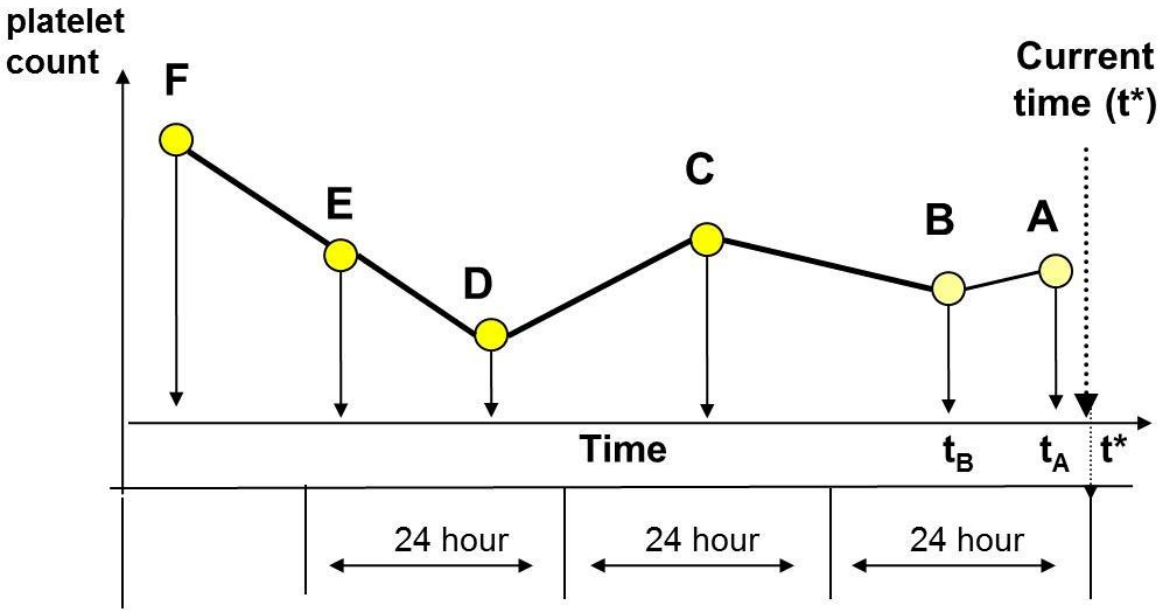
Imagens: PB Jensen et al., Nature 2012, *"Mining electronic health records"*

Procedimentos Fora do Padrão



Imagens: M Hauskrecht et al., JBI 2013, *"Outlier detection for patient monitoring and alerting"*

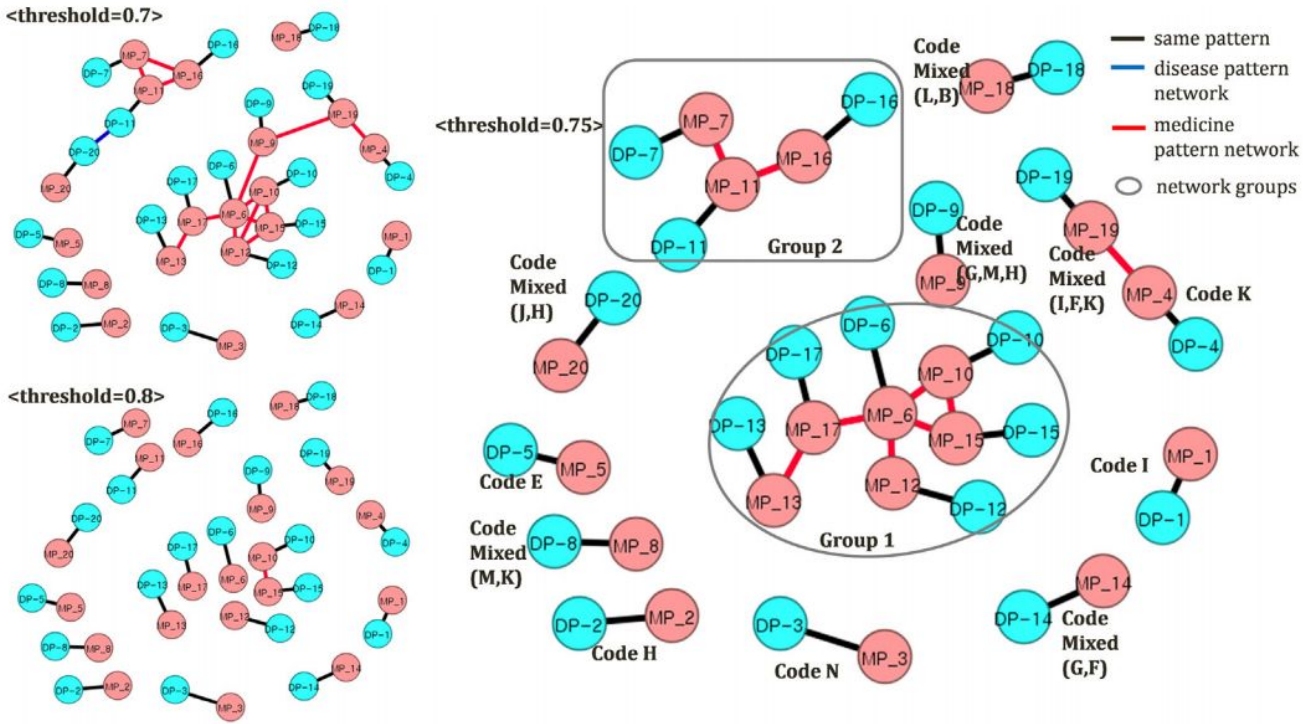
Exames Fora do Padrão



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
…

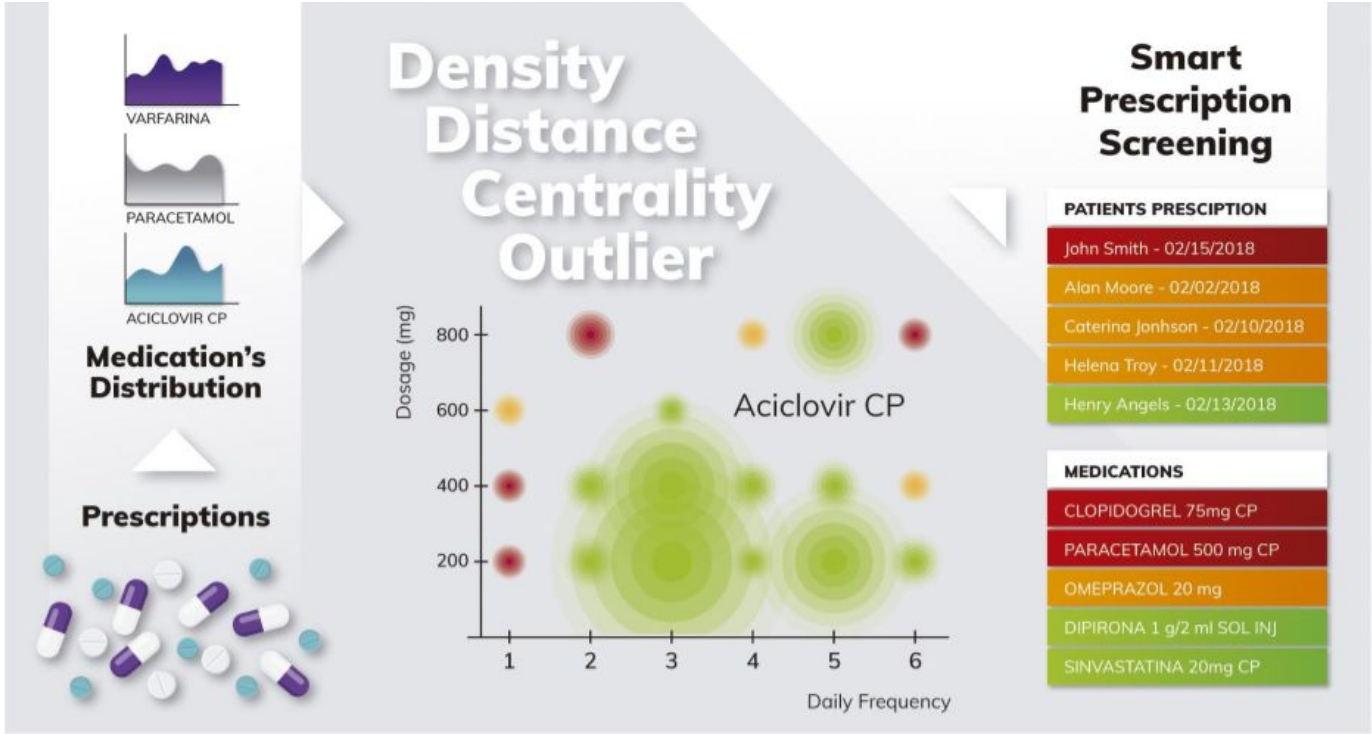
Imagens: M Hauskrecht et al., JBI 2013, *"Outlier detection for patient monitoring and alerting"*

Prescrições Padrão



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

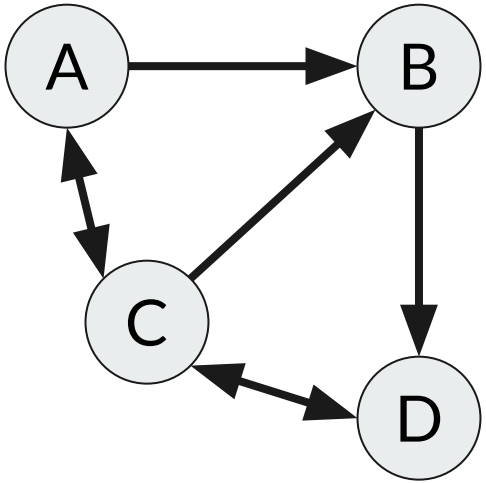
Prescrições Fora do Padrão



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

Aprendizado Não-Supervisionado e o PageRank

Google PageRank

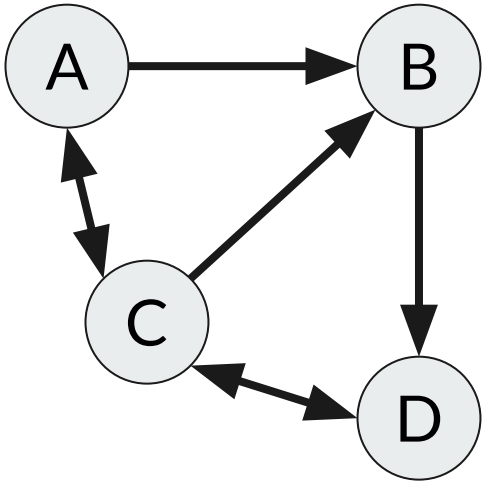


Problema de Autovetor ou Passeio Aleatório

- PageRank
- HITS
- SALSA

Calculadora: http://faculty.chemeketa.edu/ascholer/cs160/WebApps/PageRank/, Page & Brin ‘99

Google PageRank

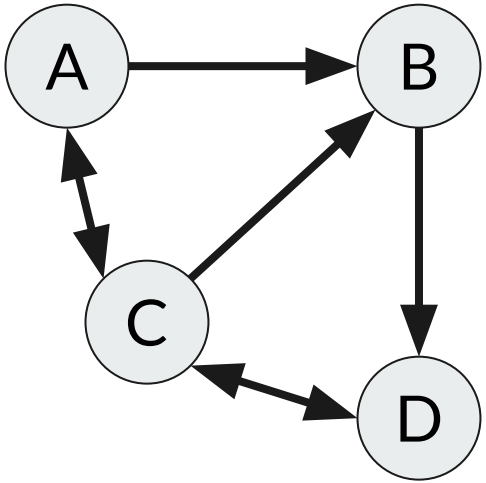


Web como um Grafo:

- Páginas são vértices
- Links são arestas

Calculadora: http://faculty.chemeketa.edu/ascholer/cs160/WebApps/PageRank/, Page & Brin ‘99

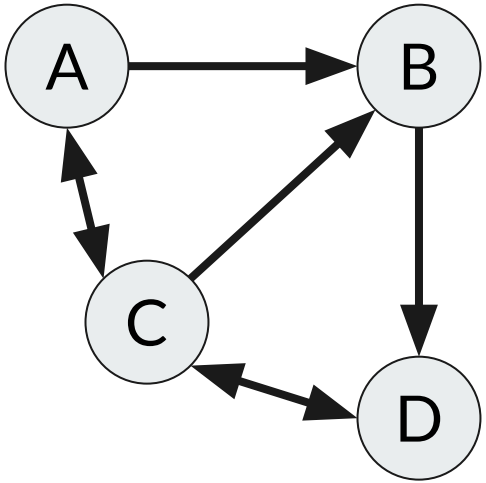
Google PageRank



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

Calculadora: http://faculty.chemeketa.edu/ascholer/cs160/WebApps/PageRank/, Page & Brin '99

Google PageRank

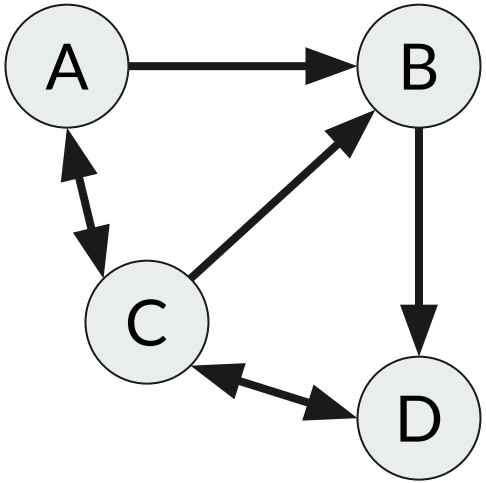


	step 0
A	1
B	1
C	1
D	1
ε	-

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

Calculadora: http://faculty.chemeketa.edu/ascholer/cs160/WebApps/PageRank/, Page & Brin '99

Google PageRank

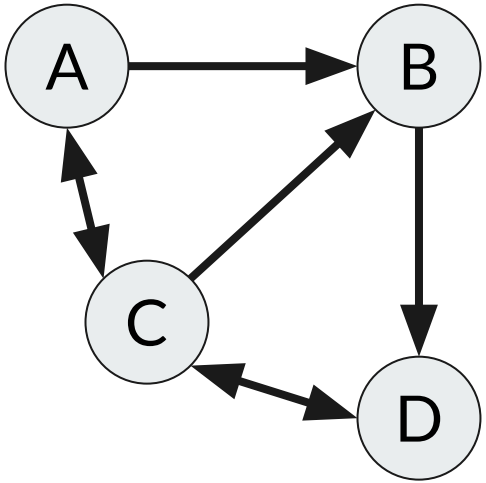


	step 0	step 1
A	1	0.33
B	1	0.83
C	1	1.50
D	1	1.33
ε	-	0.41

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

Calculadora: http://faculty.chemeketa.edu/ascholer/cs160/WebApps/PageRank/, Page & Brin '99

Google PageRank

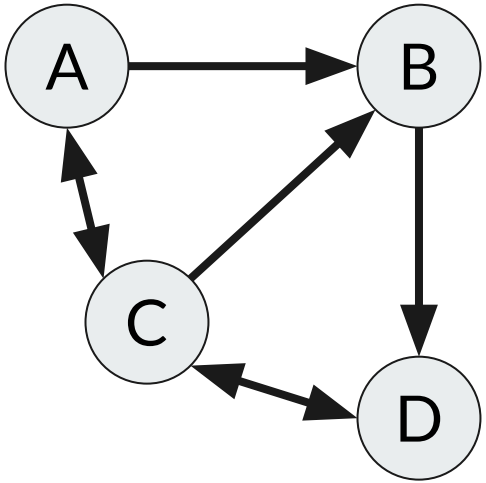


$$PR(u) = \sum \frac{PR(v)}{N_v}$$

	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

Calculadora: http://faculty.chemeketa.edu/ascholer/cs160/WebApps/PageRank/, Page & Brin '99

Google PageRank

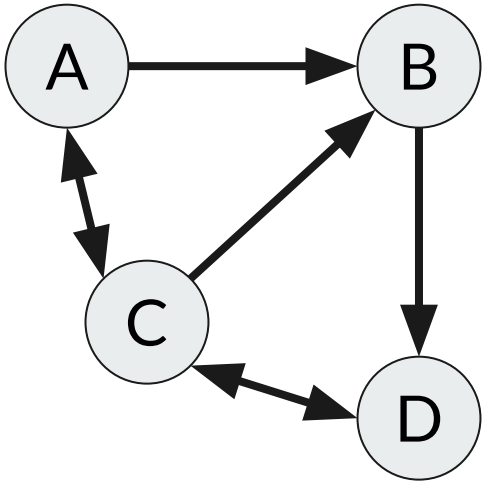


$$PR(u) = \sum \frac{PR(v)}{N_v}$$

	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

Calculadora: http://faculty.chemeketa.edu/ascholer/cs160/WebApps/PageRank/, Page & Brin '99

Google PageRank

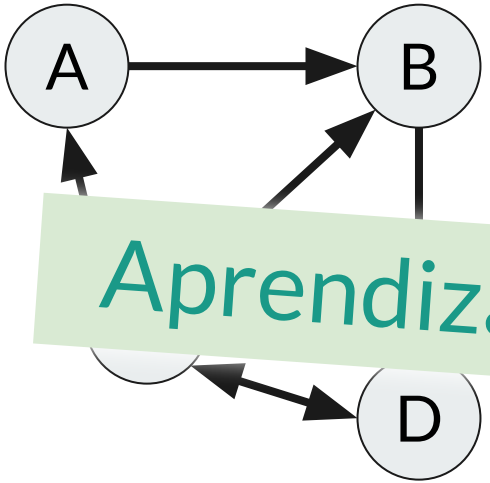


$$PR(u) = \sum \frac{PR(v)}{N_v}$$

	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	

Calculadora: http://faculty.chemeketa.edu/ascholer/cs160/WebApps/PageRank/, Page & Brin '99

Google PageRank

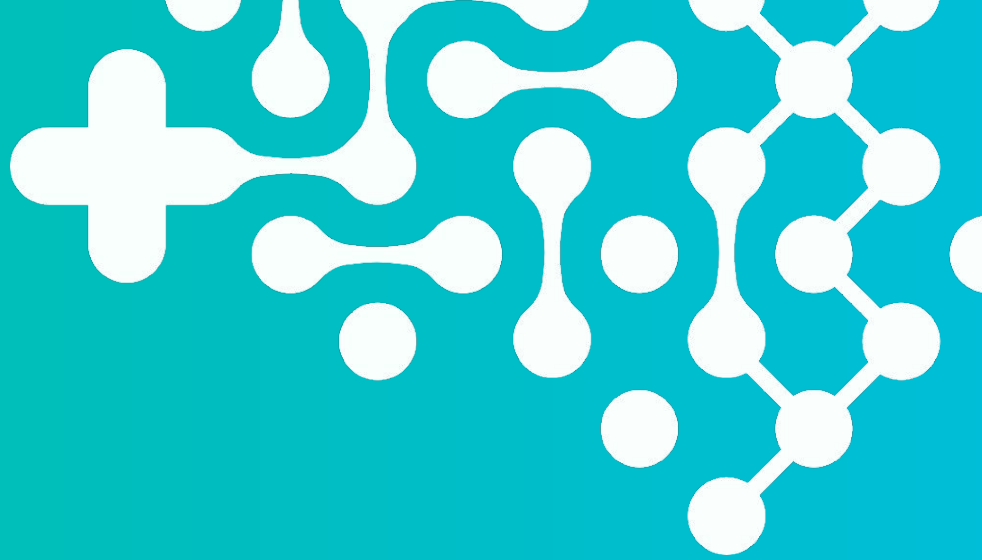


	step 0	step 1	step 2	step 3	PageRank
A	1	0.33	0.50	0.49	4
C	1	1.50			3
D	1	1.33	1.33	1.15	2
ε	-	0.41	0.08	0.09	

Aprendizado Não-Supervisionado

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

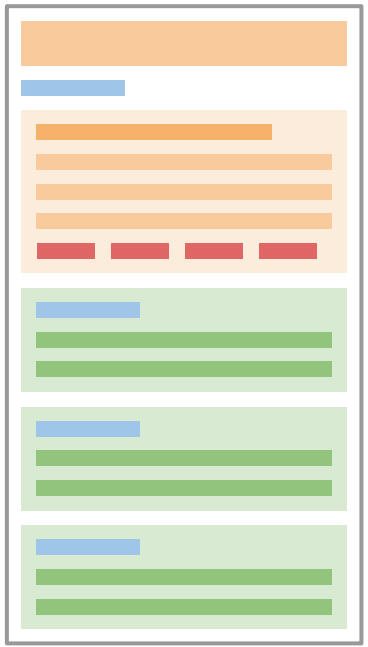
Calculadora: http://faculty.chemeketa.edu/ascholer/cs160/WebApps/PageRank/, Page & Brin '99

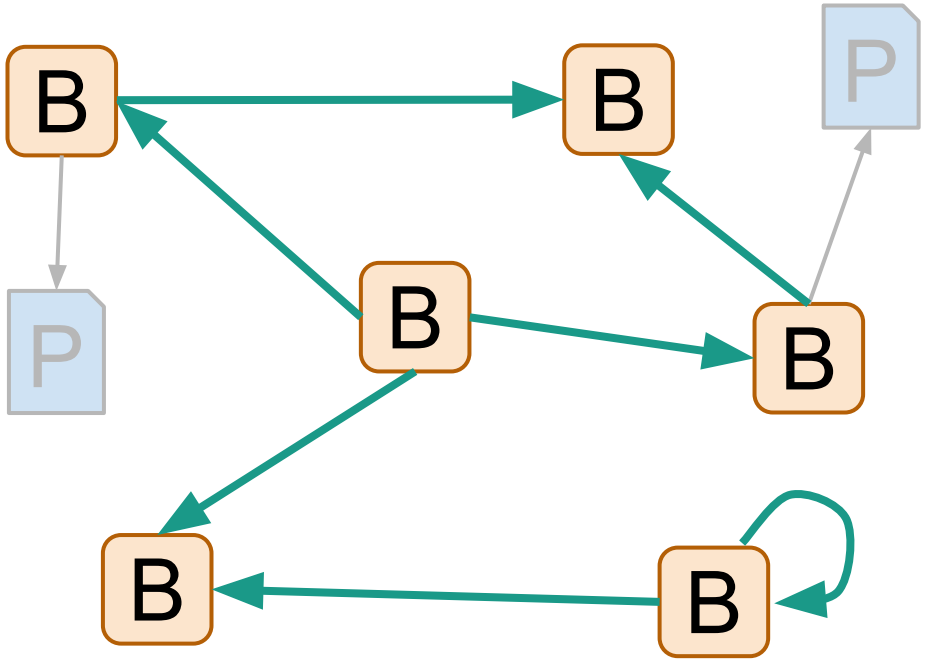


PageRank

e outras aplicações

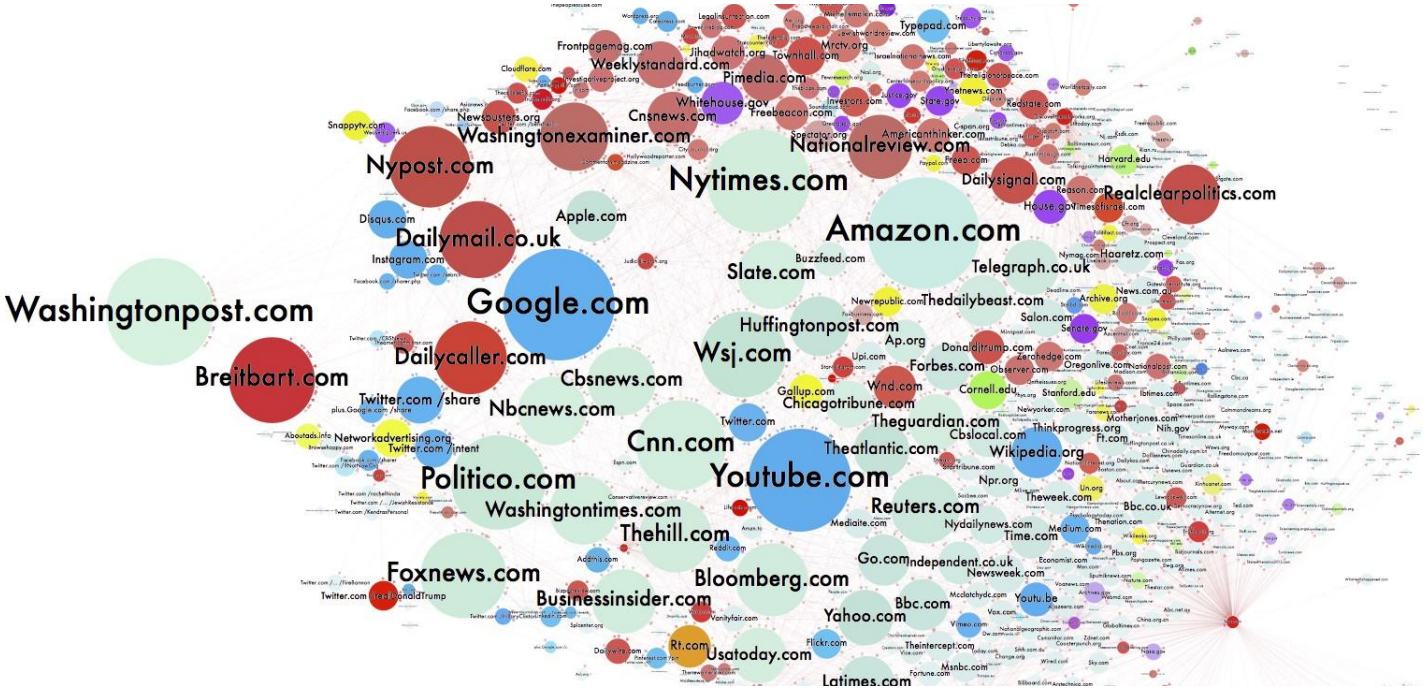
Autoridades em Blogs





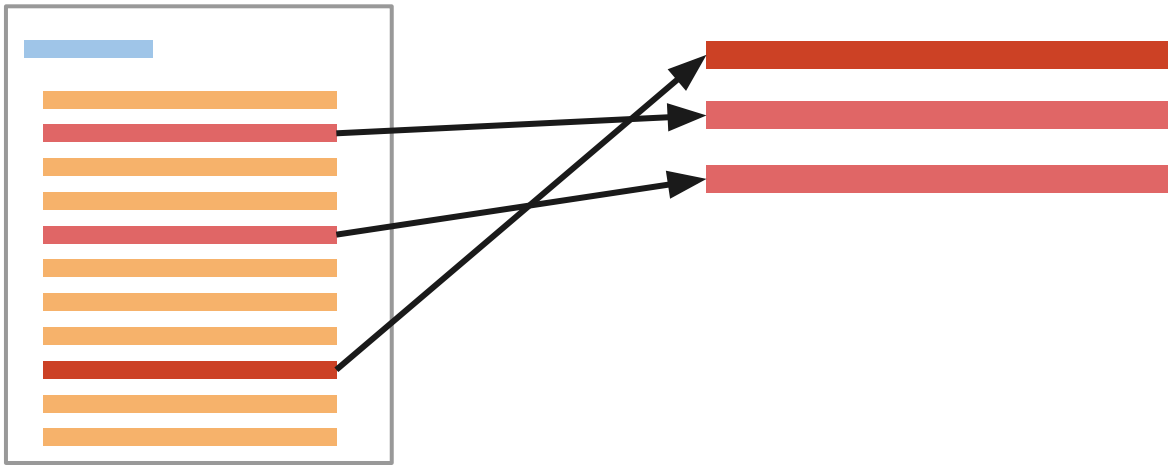
Fonte: Identificação de autoridades em tópicos na blogosfera brasileira, UFRGS'13

Identificando Fake News



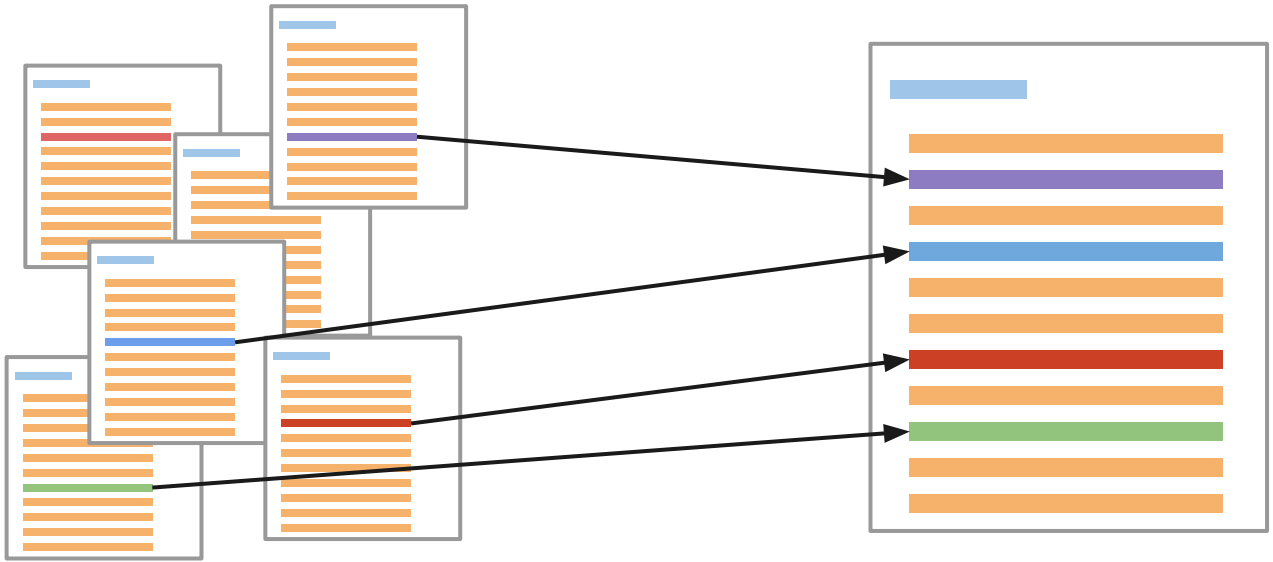
Fonte: DistrustRank: Spotting False News Domains, WebSci'18

Sumarização de Texto



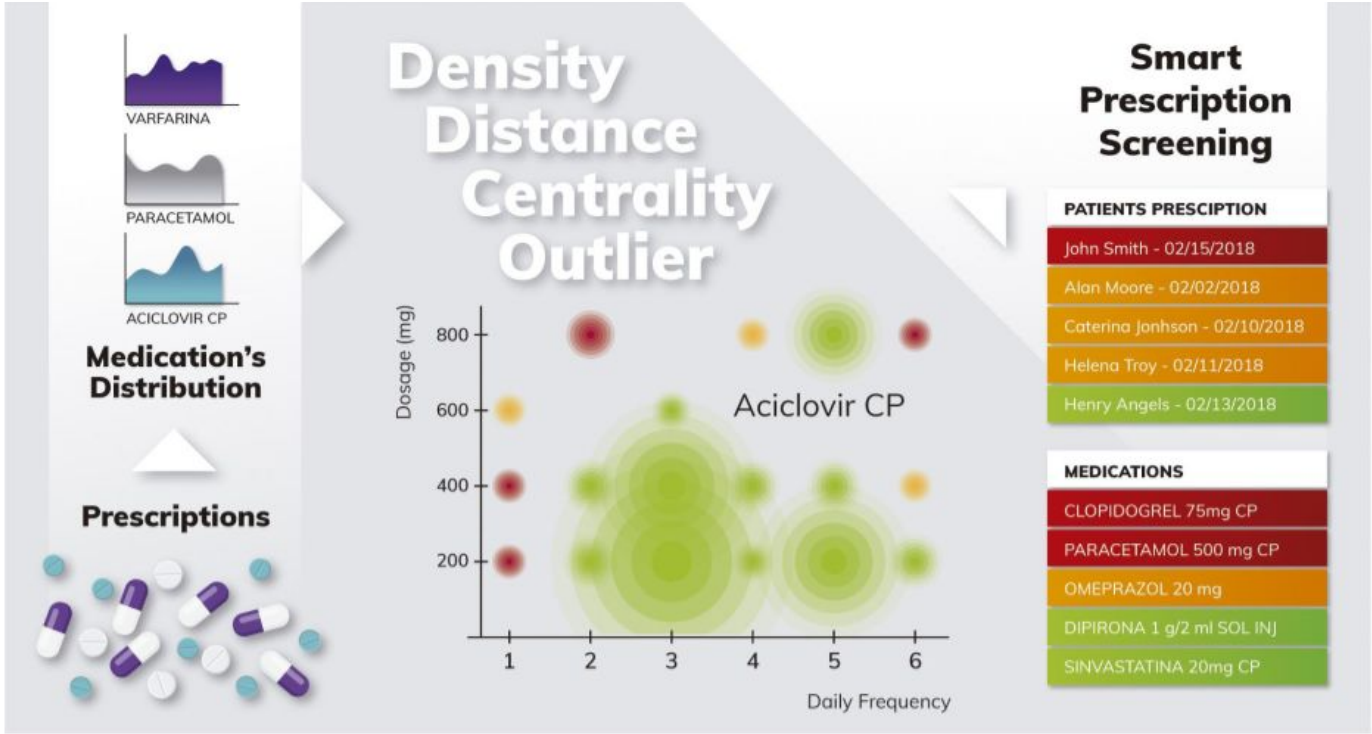
Fonte: LexRank: Graph-based Lexical Centrality as Salience in Text Summarization, JAIR'04

Ordenamento de Revisões



Fonte: MRR: an unsupervised algorithm to rank reviews by relevance, WI'17

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*

MEDICATION WITHOUT HARM
Global Patient Safety Challenge

World Health Organization

who.int/patientsafety/medication-safety

* além de dano ou morte dos pacientes

Erros de Medicação - OMS

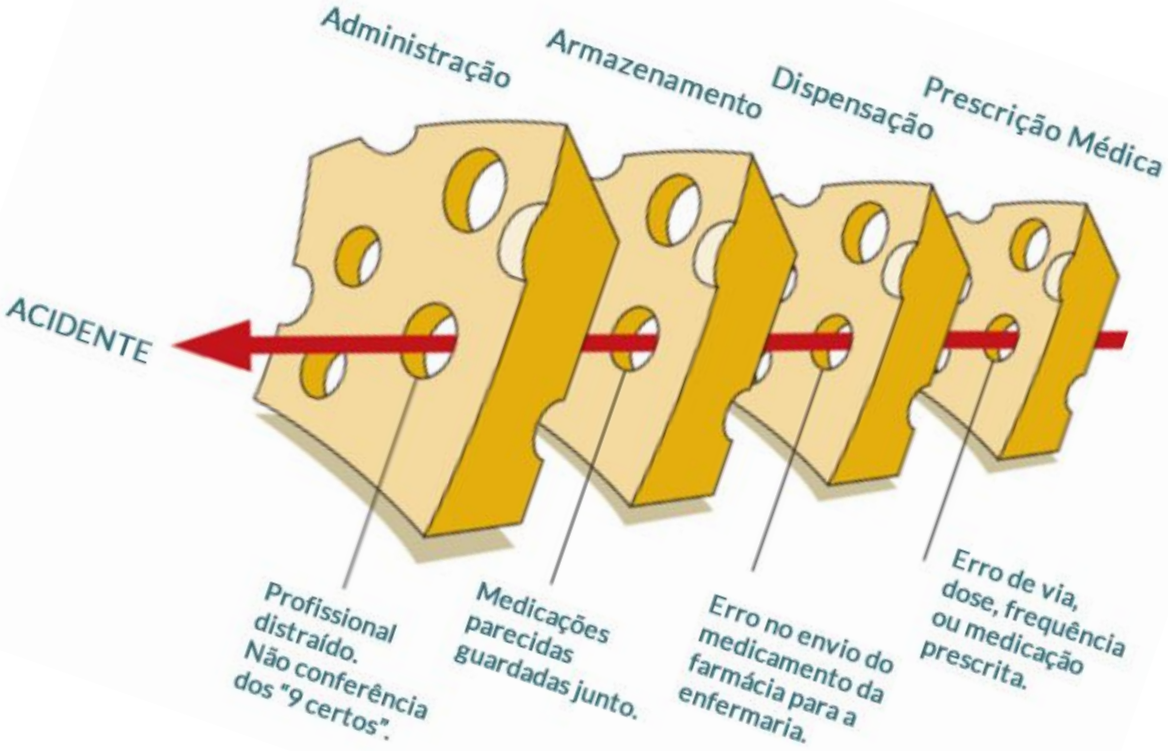
Prescrição

Triagem

Dispensação

Administração

Teoria do Queijo-Suiço



Erros de Medicação - OMS

Prescrição

Triagem

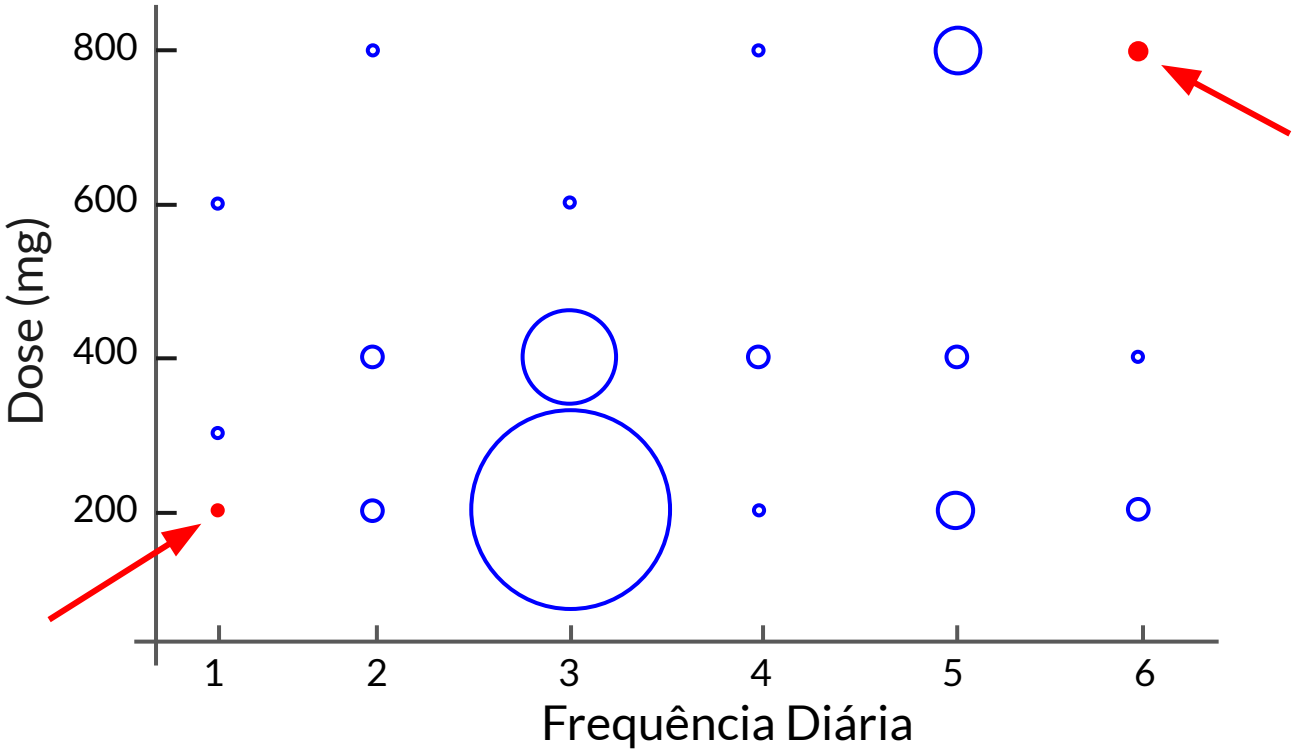
Dispensação

Administração

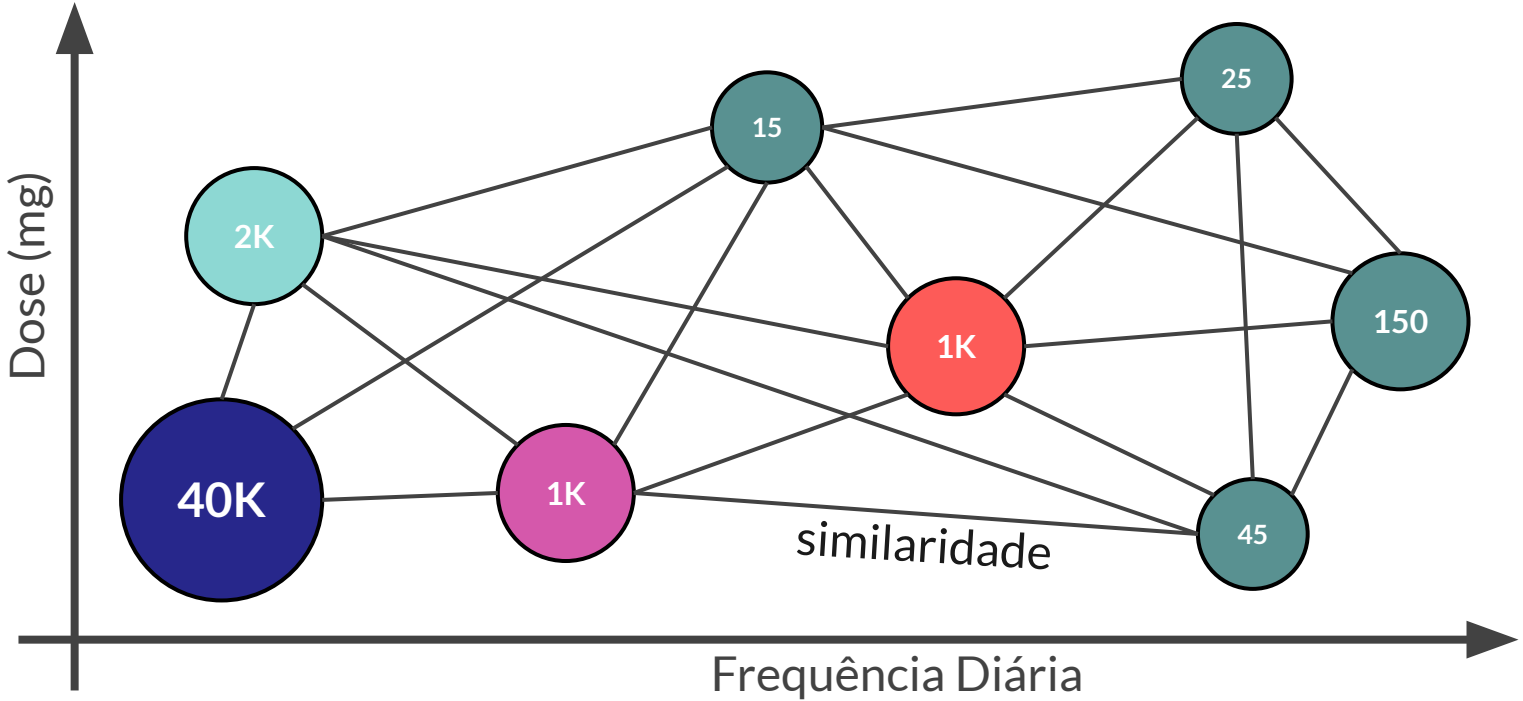
Triagem Farmacêutica

?

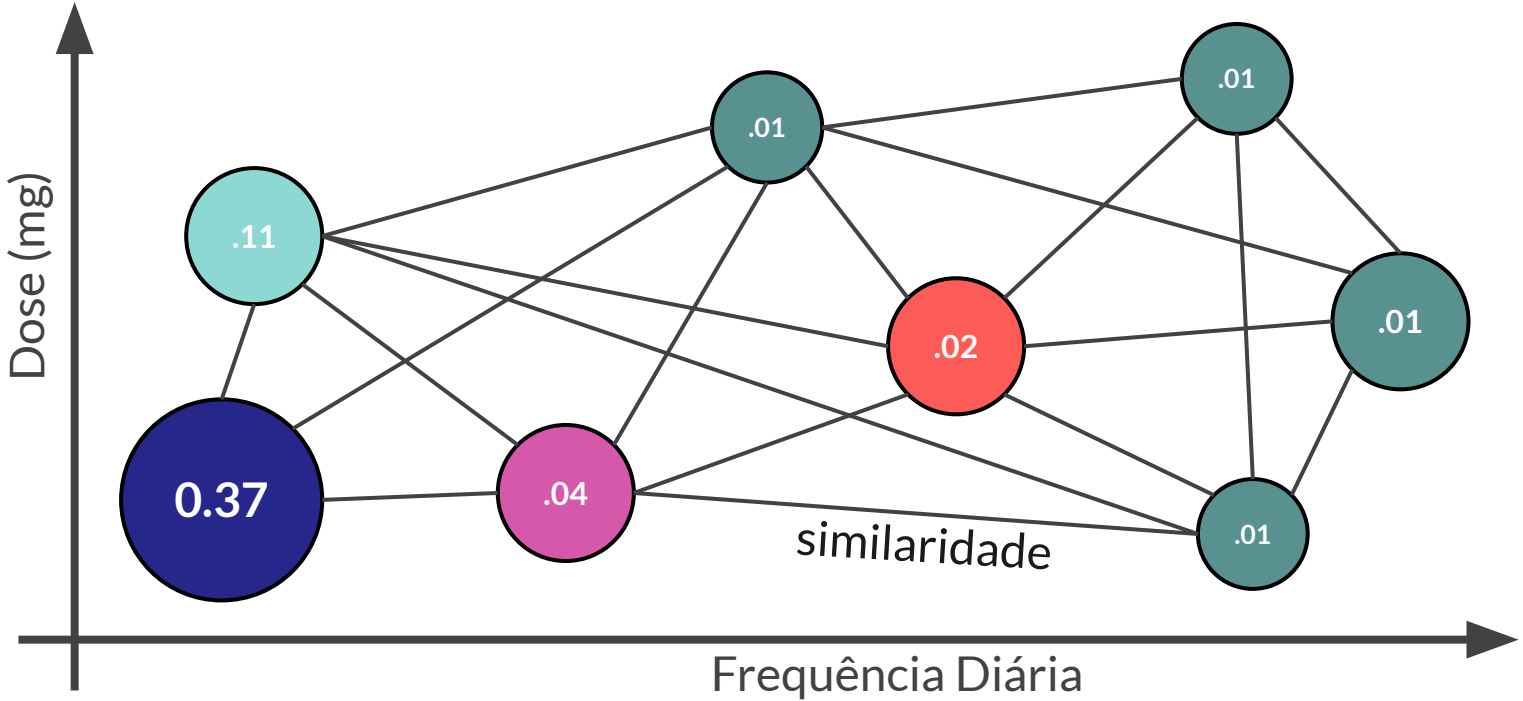
Detecção de Outliers (Aciclovir CP)



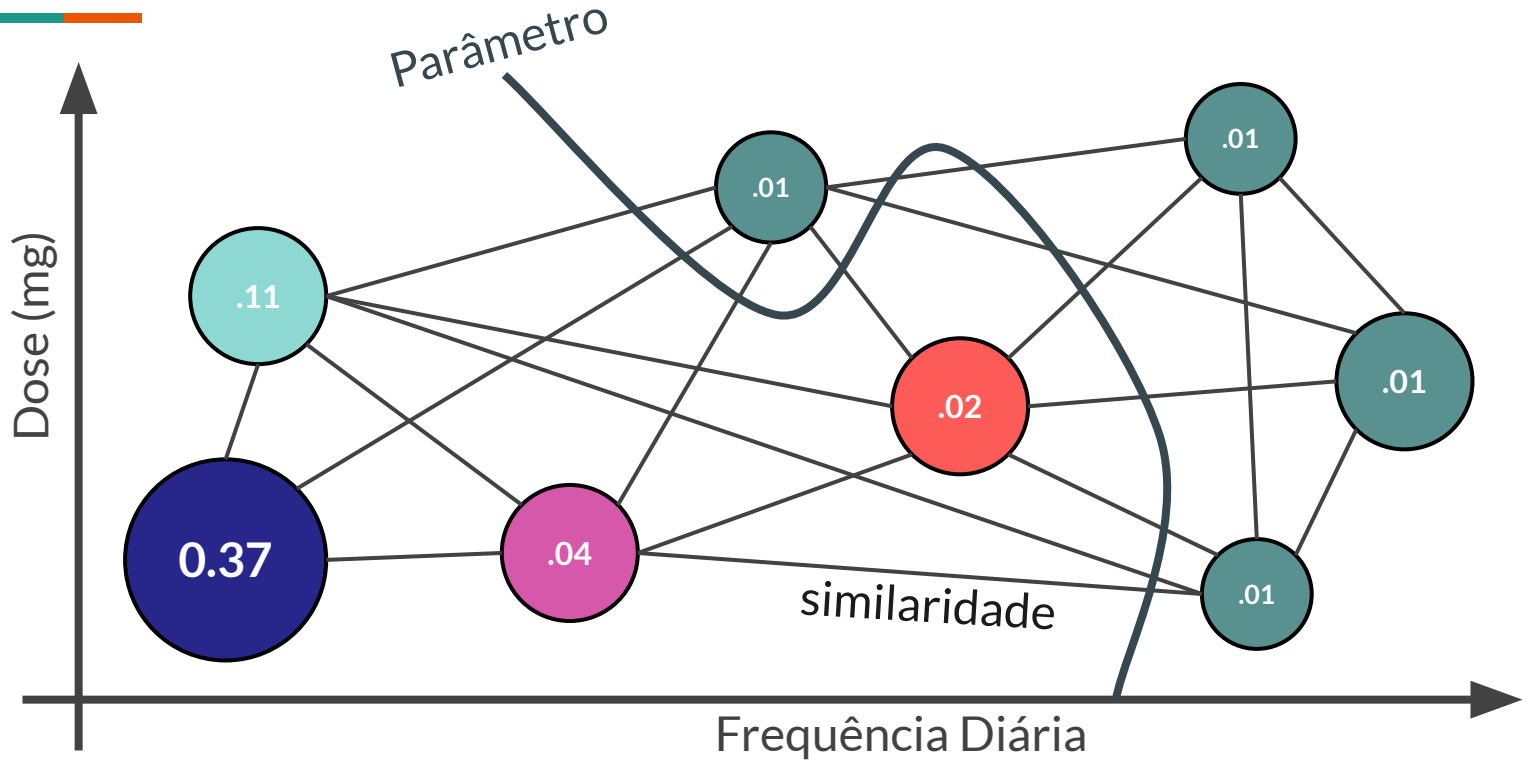
Detecção de Outliers (grafo de prescrições)



Detecção de Outliers (grafo de prescrições)



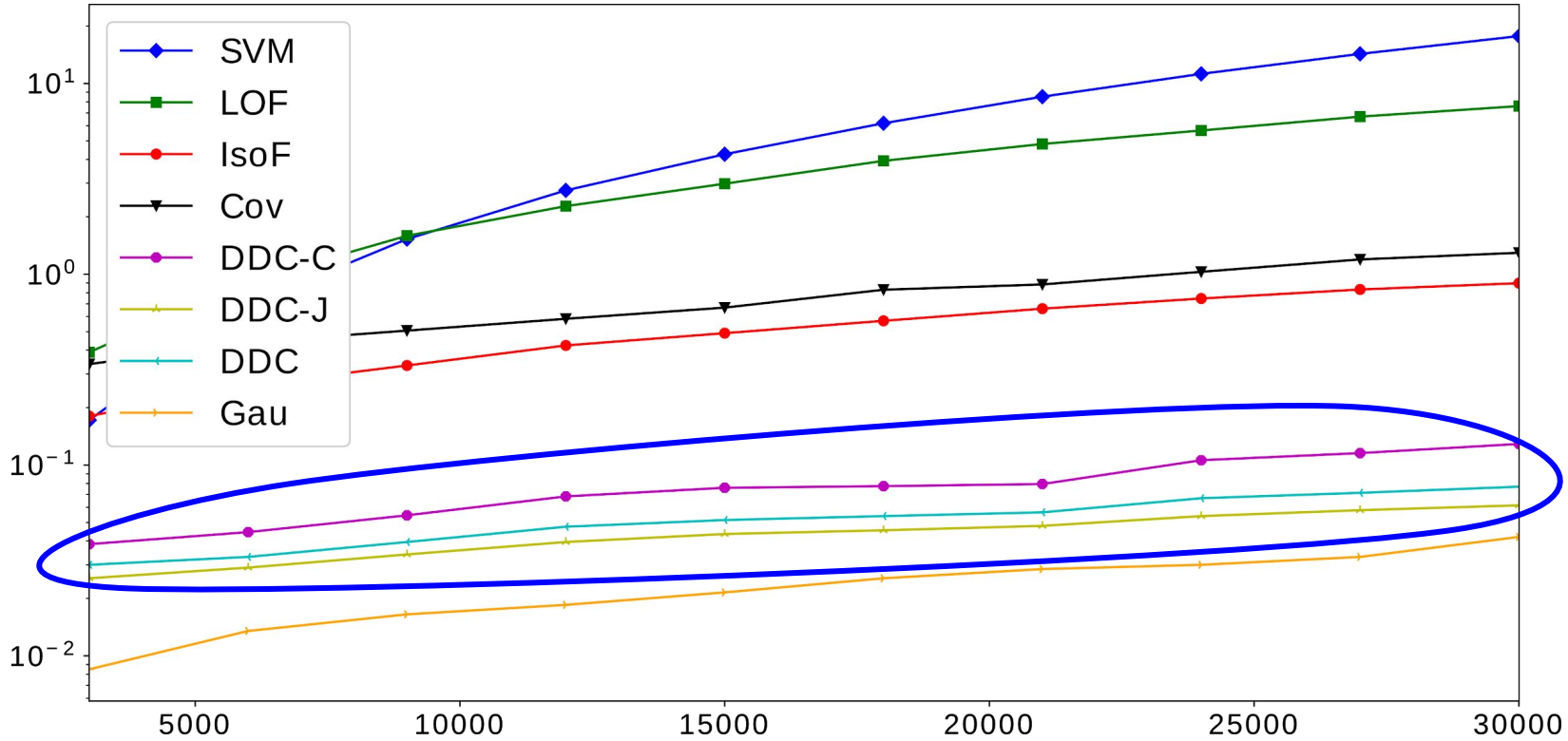
Detecção de Outliers (grafo de prescrições)



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

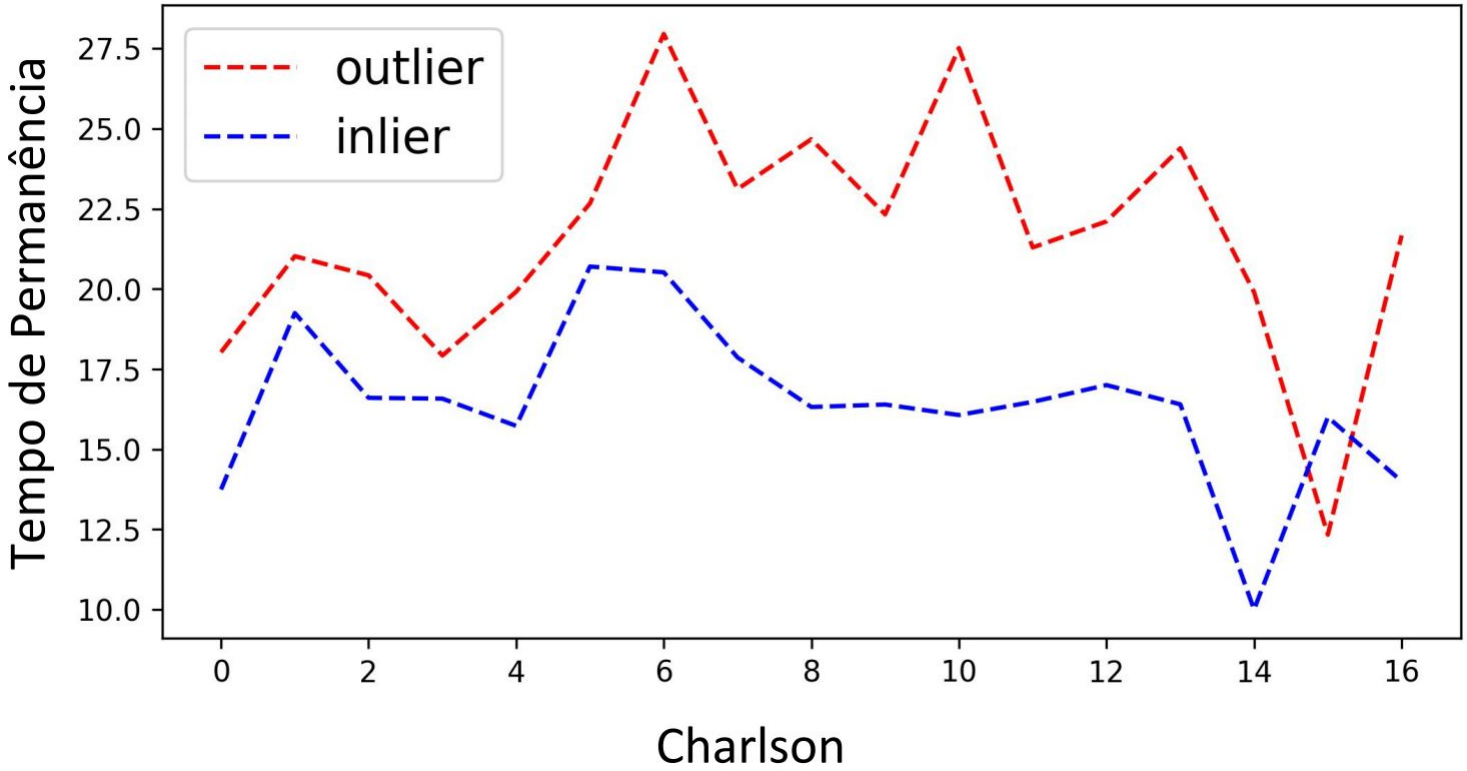
Frequência:

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

Apresentação:

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

Tempo de Internação



Repetibilidade

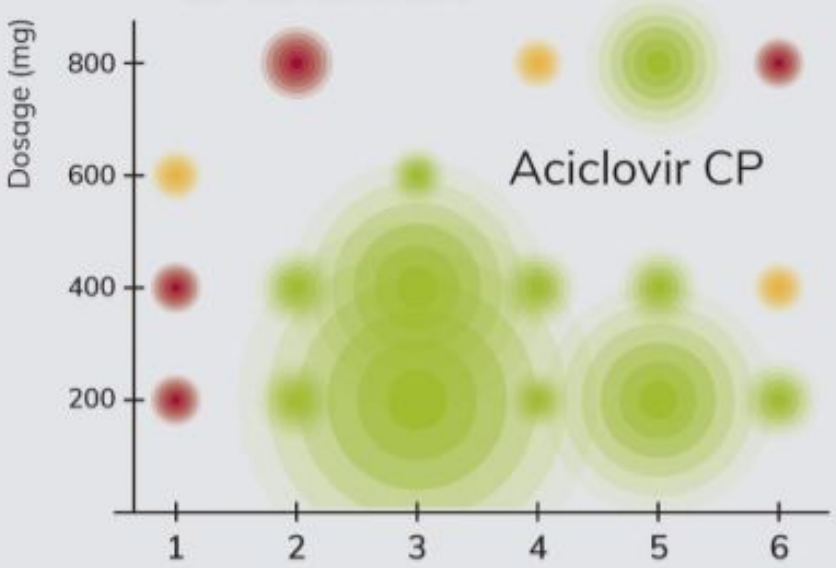
Replicabilidade

Reprodutibilidade

Medication's Distribution

Prescriptions

Density Distance Centrality Outlier



Smart Prescription Screening

PATIENTS PRESCIPTION

- 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

https://github.com/nlp-pucrs/prescription-outliers

PUCRS

GHC
Grupo Hospitalar Conceição

Henrique Dias

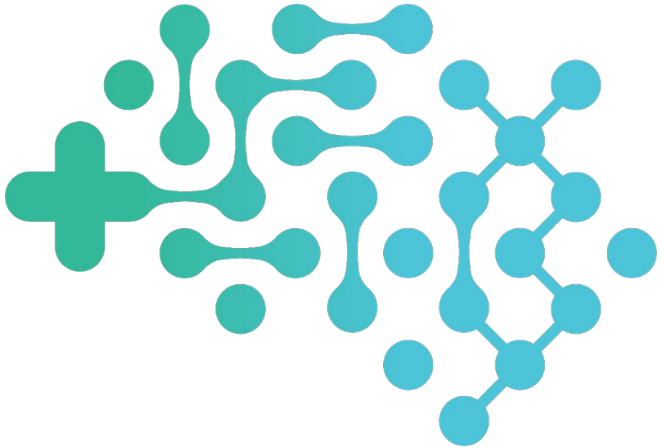
PhD* Ciência da Computação

Ana Helena Ulbrich

PhD Farmacêutica

PUCRS

GHC
Grupo Hospitalar Conceição



Grupo de
Inteligência Artificial
na Saúde

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

Hospital Ernesto Dornelles

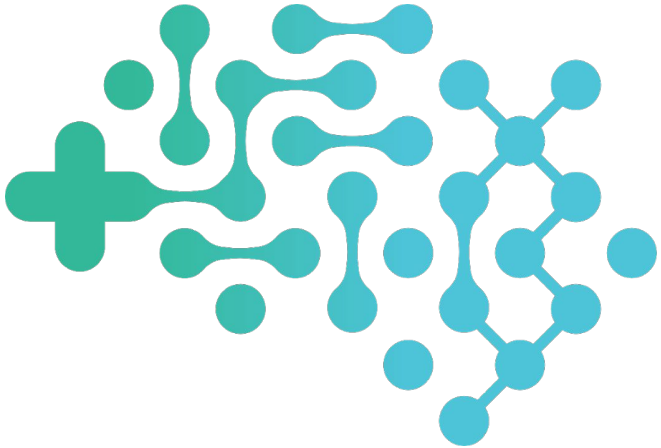
HOSPITAL MOINHOS DE VENTO

SANTA CASA DE MISERICÓRDIA PORTO ALEGRE

HOSPITAL MÃE DE DEUS
SISTEMA DE SAÚDE MÃE DE DEUS

SOCIEDADE BENEFICENTE DE SENHORAS
SÍRIO·LIBANÊS

HOSPITAL SÃO LUCAS DA PUCRS



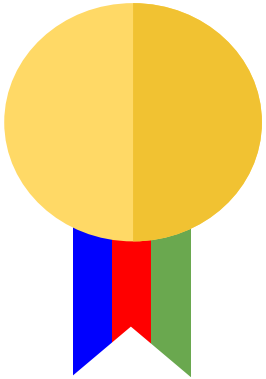
Grupo de Inteligência Artificial na Saúde

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

Premiações

PRÊMIO EMPREENDA SAÚDE

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

Gratuito para Leitos SUS

Leitos Privados (receita)

Dedução Fiscal (doações)

SUS

Hospitais Parceiros

SANTA CASA
DE MISERICÓRDIA
PORTO ALEGRE

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

HOSPITAL
MÃE DE DEUS

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

Voluntários 30+

noharm.ai

Henrique
Infra / Backend / ETL

Ana Helena
Farmácia Clínica

Fábio
Médico / PM / UX

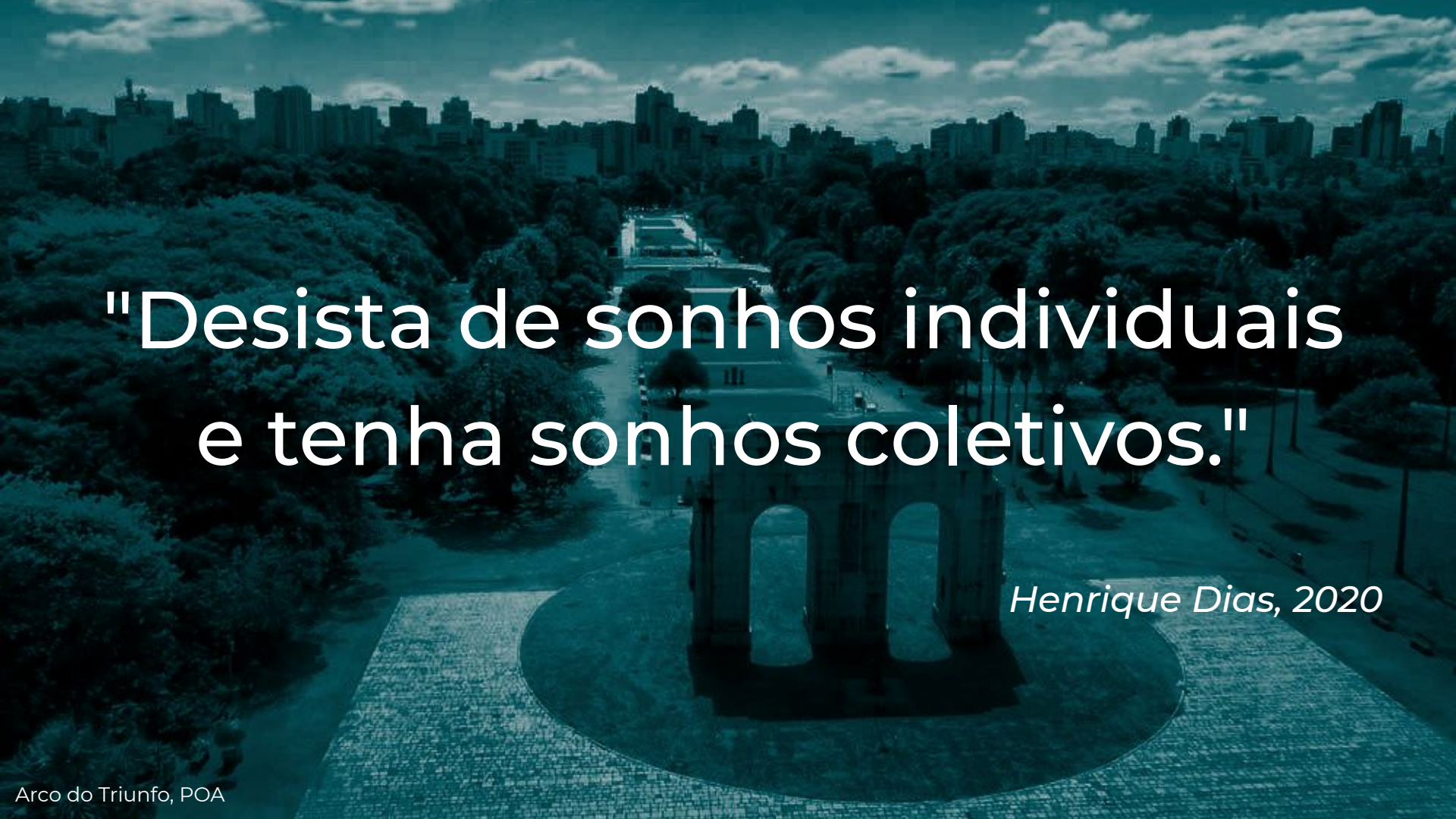
Marcelo
Backend / Mobile

Rodrigo
Comercial

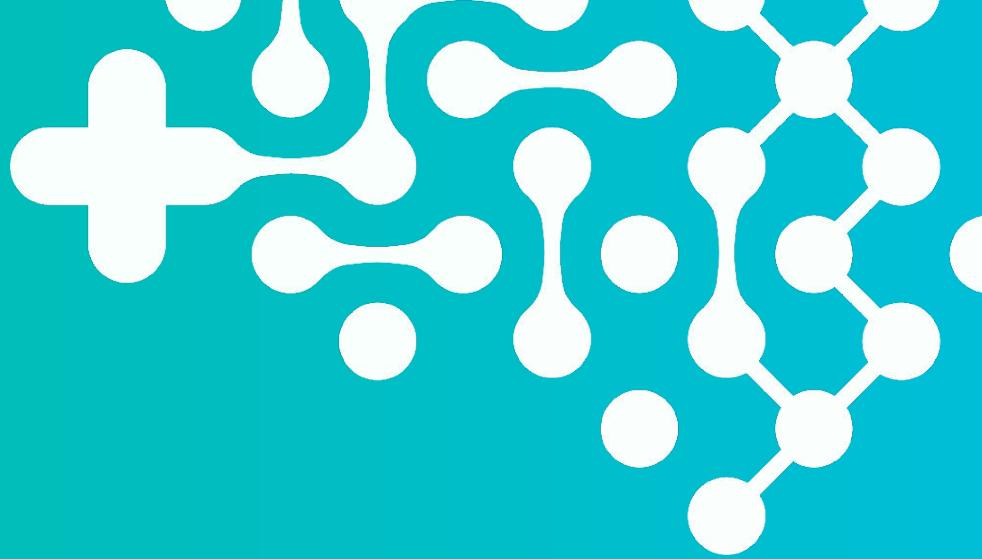
Augusto
ETL

Júlia
ETL / MV

Gabrielli
Farmácia Clínica



Arco do Triunfo, POA

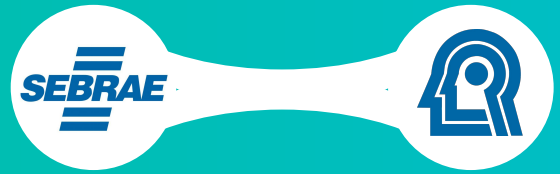


Obrigado!
Dúvidas?

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

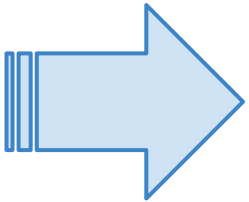
CBMS'18: https://github.com/nlp-pucrs/cci-regression

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

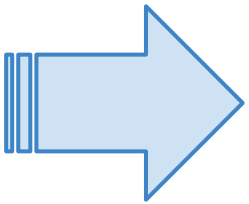
BIBE'19: https://github.com/nlp-pucrs/fall-detection

Infraestrutura

AWS RDS
Schemas
(bridge)

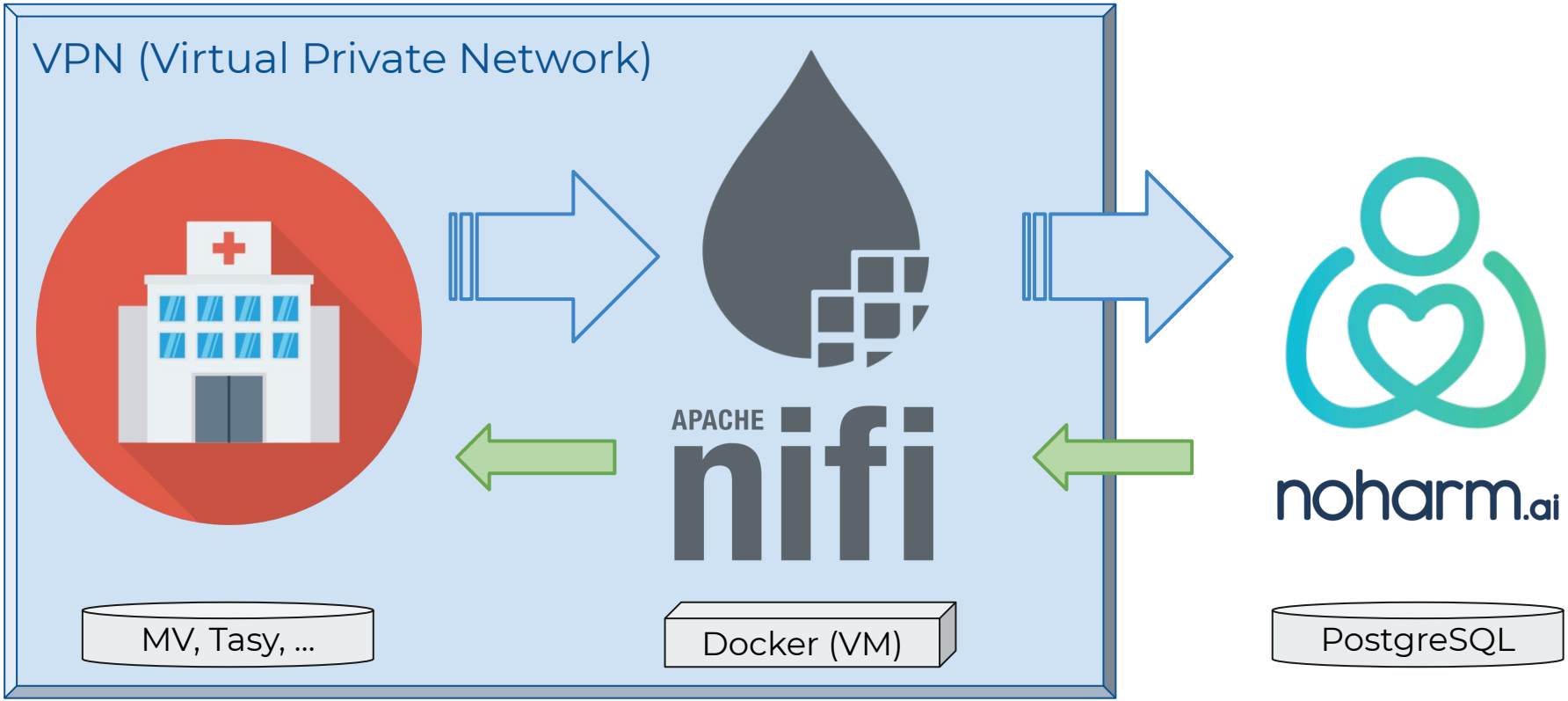


AWS Lambda
Python
Zappa



Web, AWS S3
React

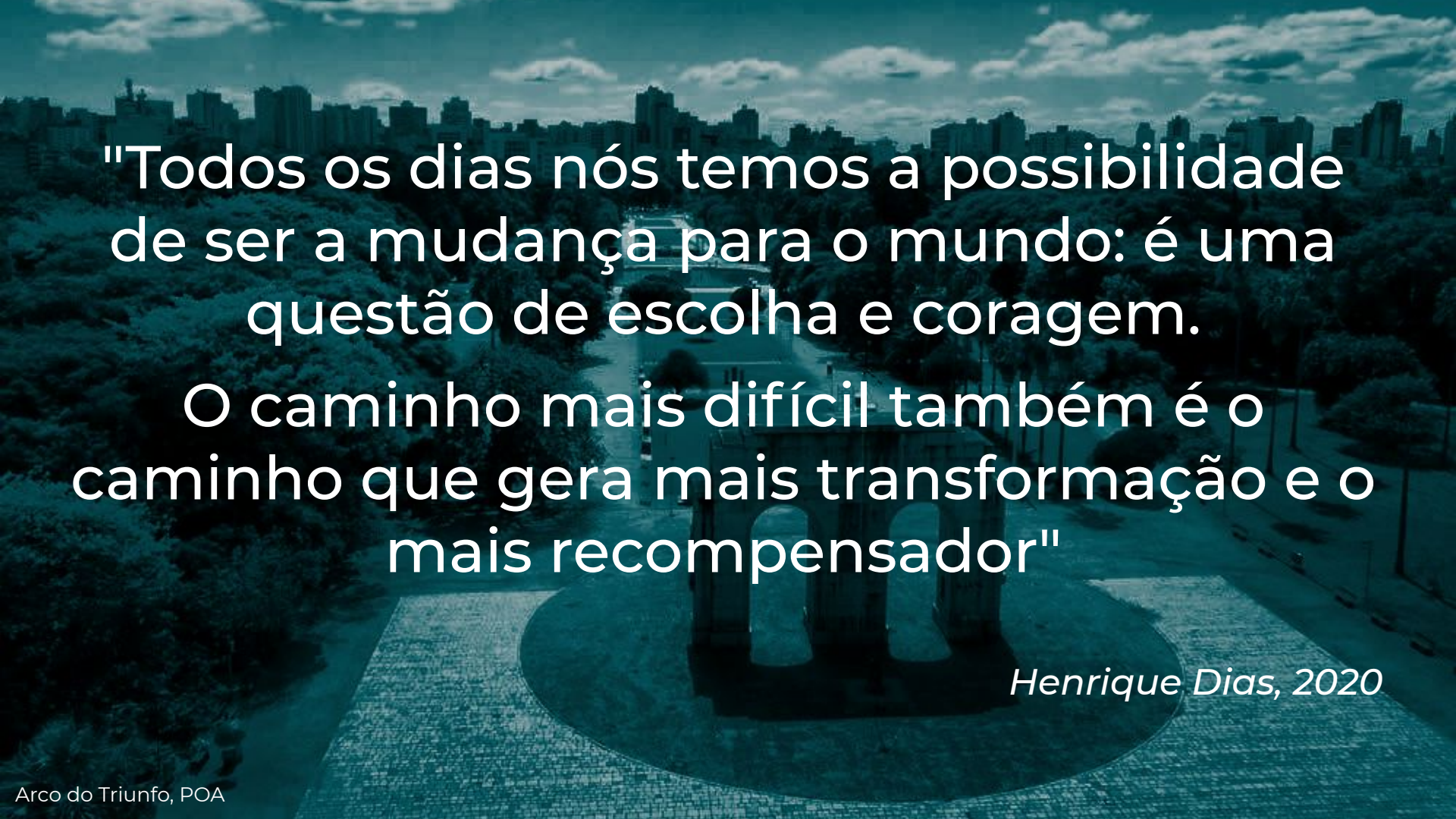
Integração de Dados



Inteligência Artificial na Saúde no Brasil

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


"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
Arco do Triunfo, POA