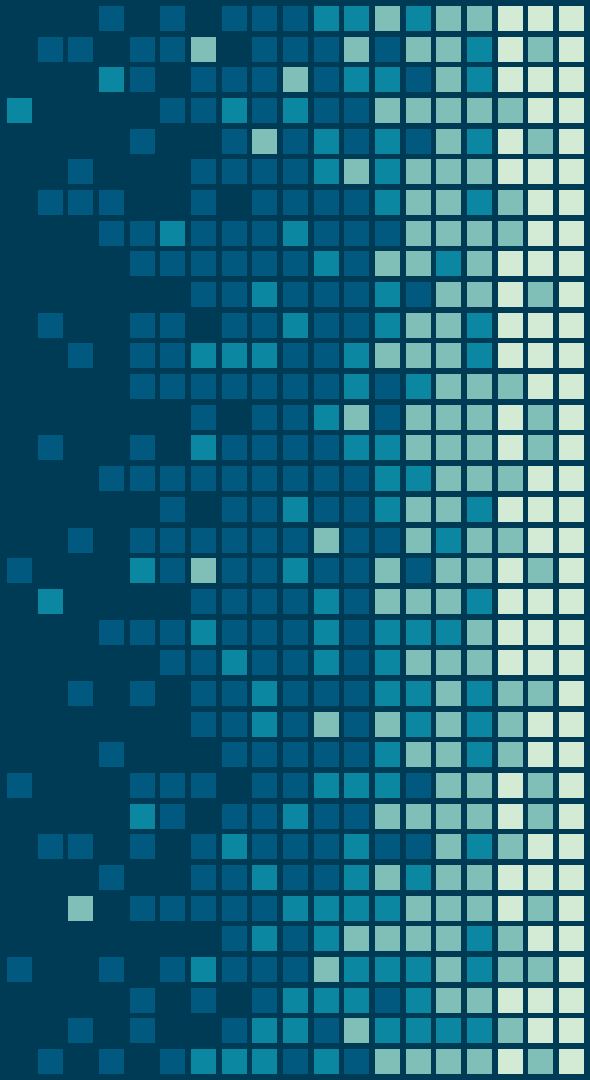
SISTEMA DE RECOMENDAÇÃO COM NEO4J + SURPRISE



MORVANA BONIN

- Analista de Desenvolvimento na KingHost,
- Estudante em Análise e Desenvolvimento de Sistemas
- Fundadora e uma das coordenadores do Meetup Microservices de Poa
- Entusiasta de Machine Learning
- Fã de animes e mangás.

Sistemas de Recomendação

The most successful and widespread application of machine learning technologies in business


BUSCA
FILTRO
RECOMENDAÇÃO

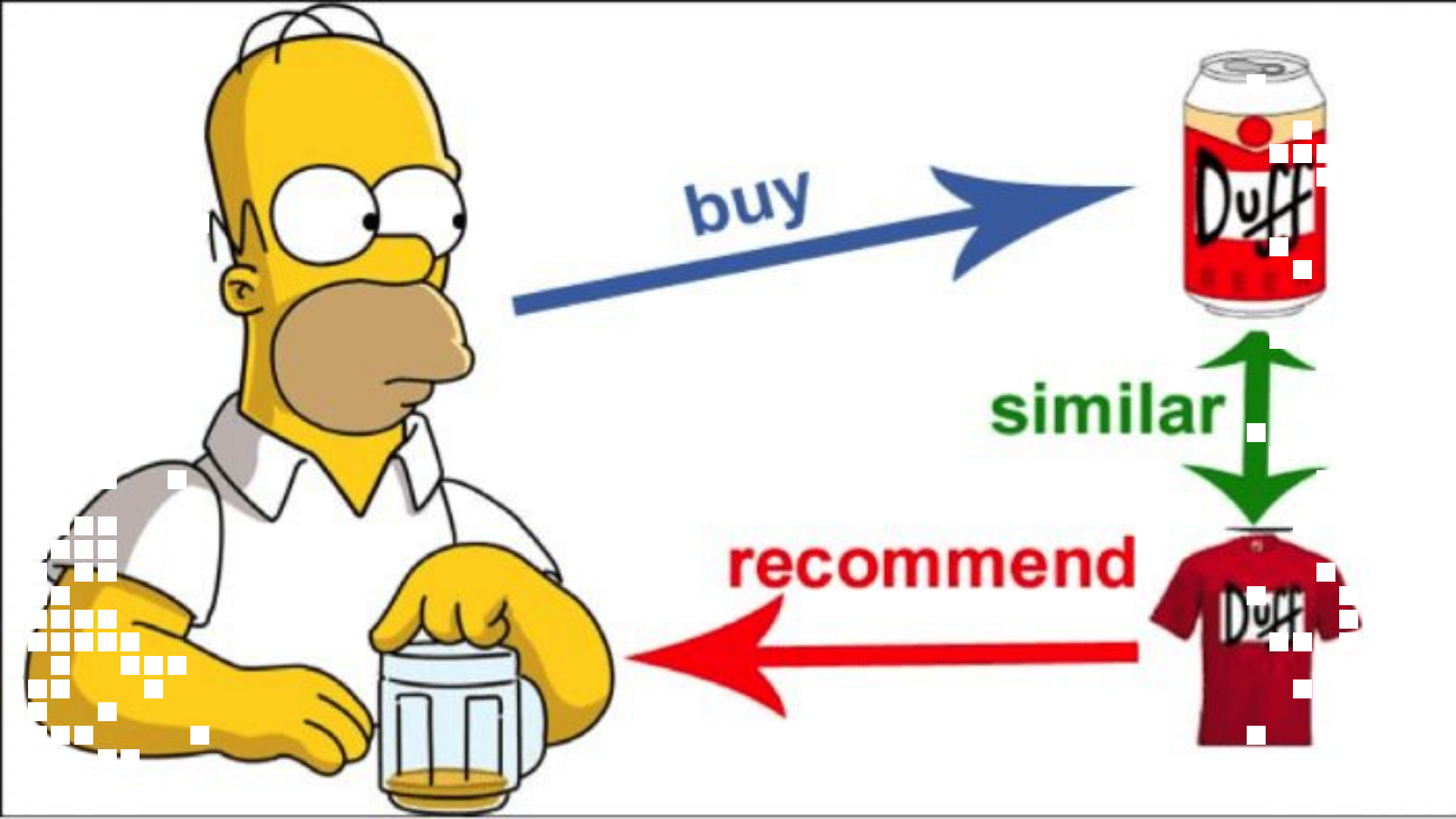
São classificados nos tipos

- *Sistemas baseados em* ***filtragem de conteúdo***
- *Sistemas baseados em* ***filtragem colaborativa***
- *Sistemas* ***híbridos***

Content-based systems

examine properties of the items recommend

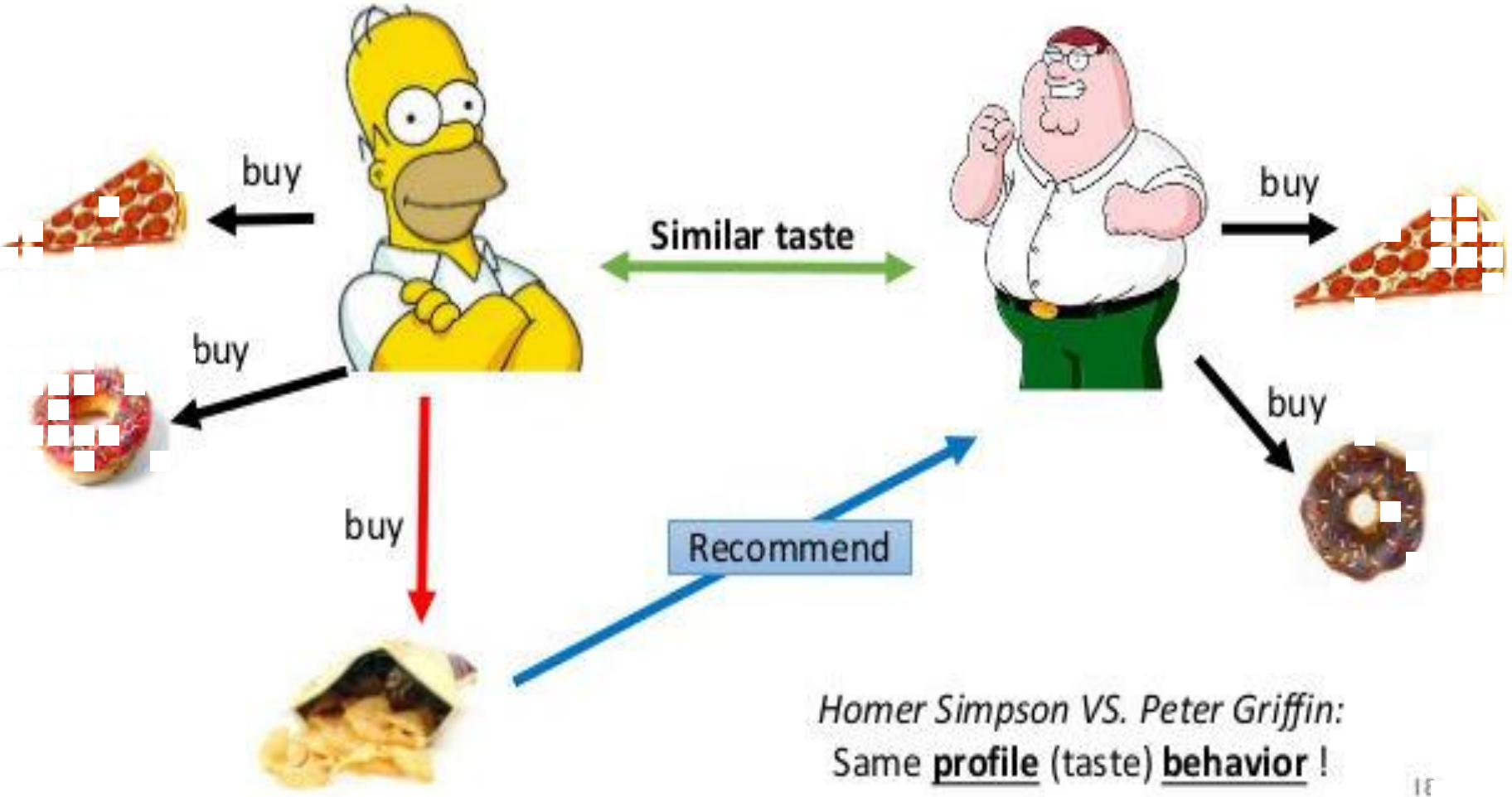
- Examina as propriedade dos itens recomendados
- Fazem a sugestão semelhantes aos que o usuário demonstrou interesse no passado
- Ou sugestão sobre as configurações de preferências do usuário.


buy
Duff
similar
recommend
Duff

Collaborative-filtering systems

recommend items based on similarity measures between users and/ or items

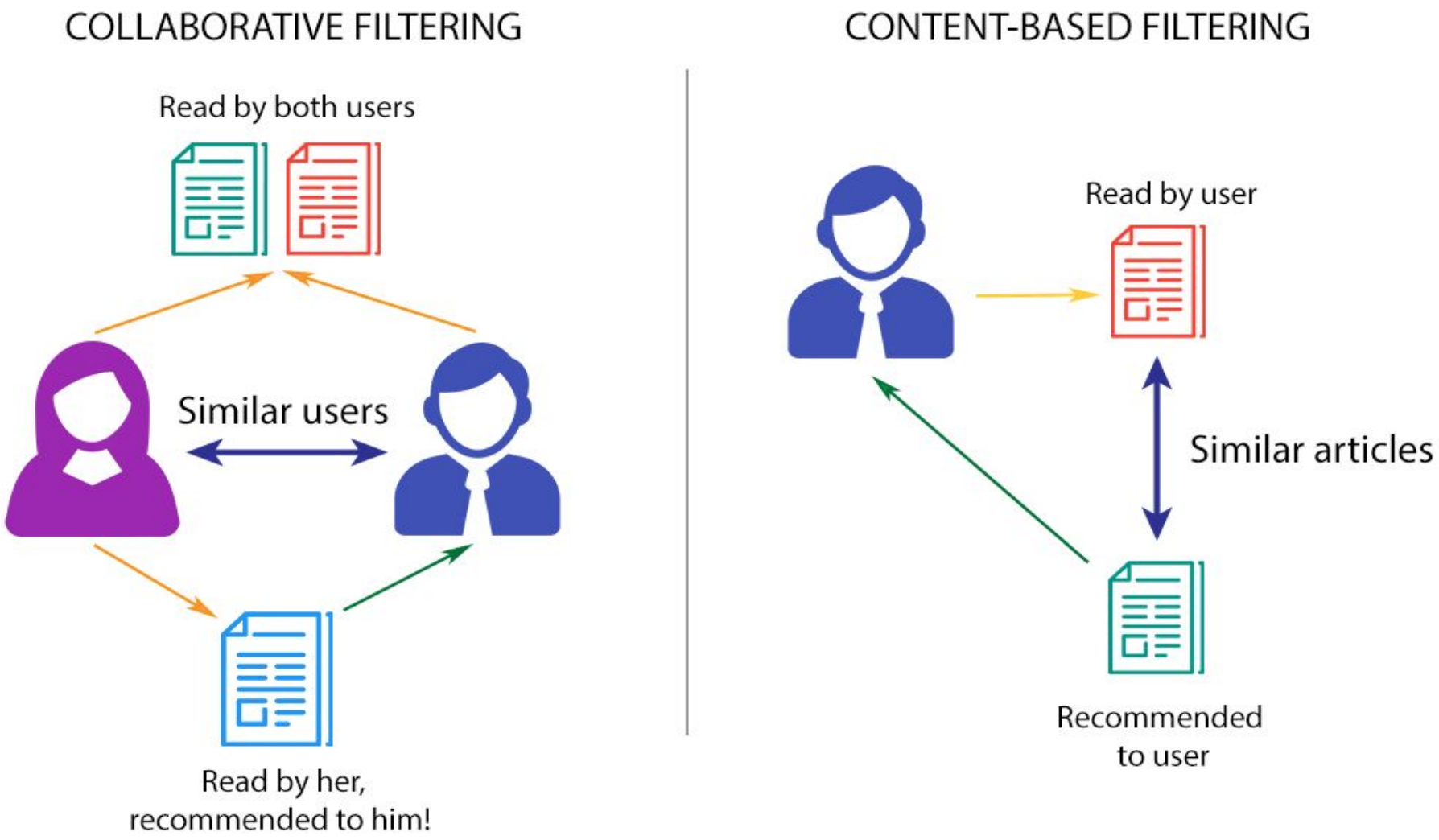
- Recomendação baseada na similaridade medida entre usuários e/ou itens.
- Essa medição pode ser uma escala de pontuação baseado em estrelas
  - ícone caracterizando gostei e não gostei (avaliação binária)
  - através da postagem de comentários sobre o item.


buy
buy
buy
Similar taste
buy
buy
Recommend
Homer Simpson VS. Peter Griffin:
Same profile (taste) behavior !

Hybrid systems

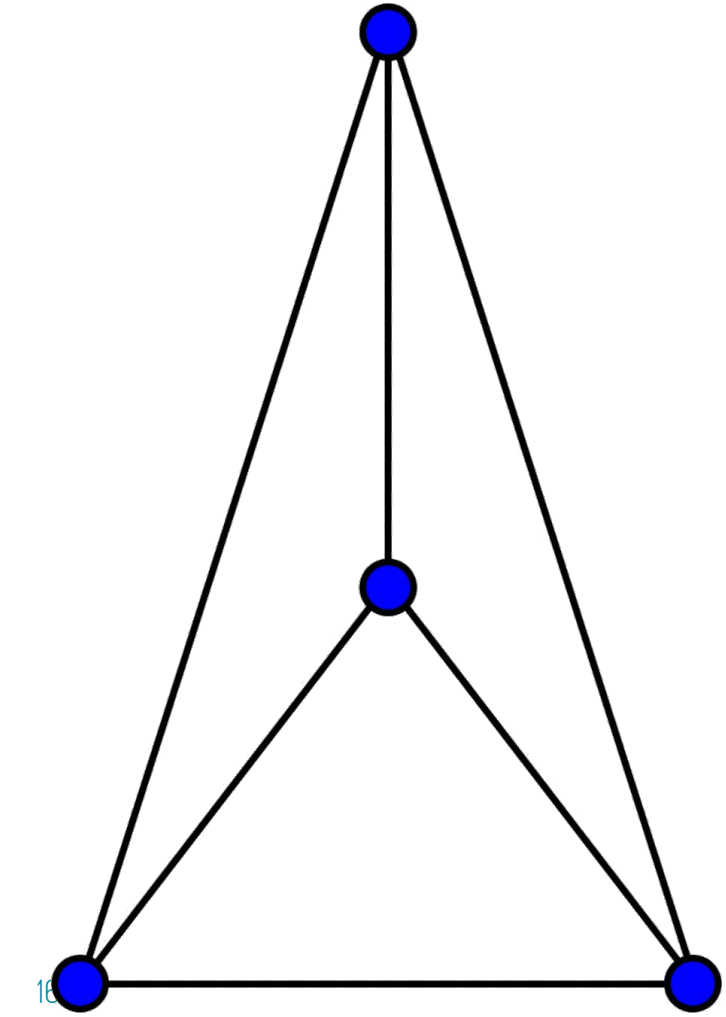
both content-based filtering and collaborative filtering have there strengths and weaknesses

- Fortalecem as vantagens e minimizam as principais desvantagens da filtragem baseada em conteúdo e filtragem colaborativa
- Combinam diferentes métodos.
- Ajuda no chamado cold-start

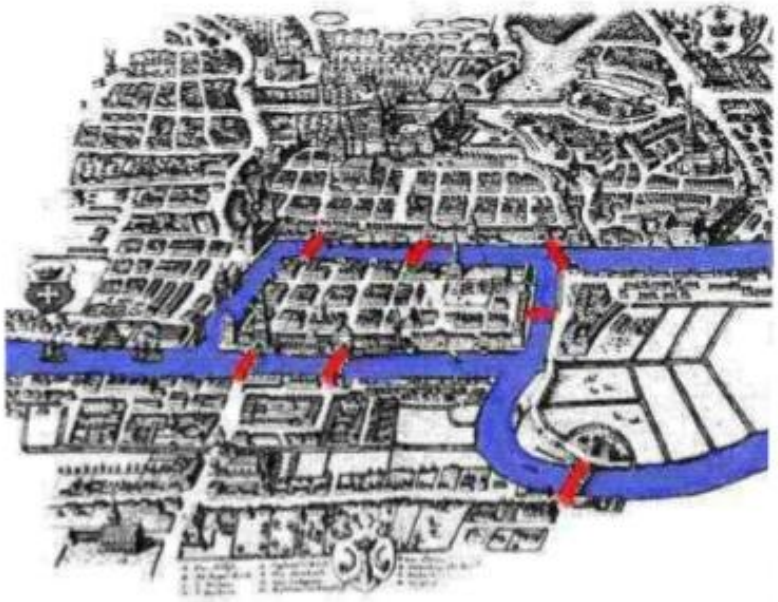

COLLABORATIVE FILTERING
Read by both users
Similar users
Read by her,
recommended to him!
CONTENT-BASED FILTERING
Read by user
Similar articles
Recommended
to user

Neo4j

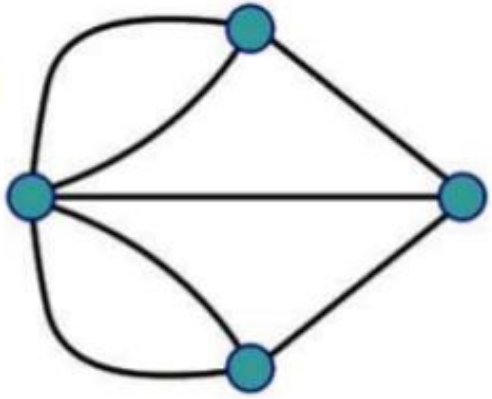
Graph Databases for connected data



um conjunto de vértices e arestas que se ligam em pares de vértices distintos.

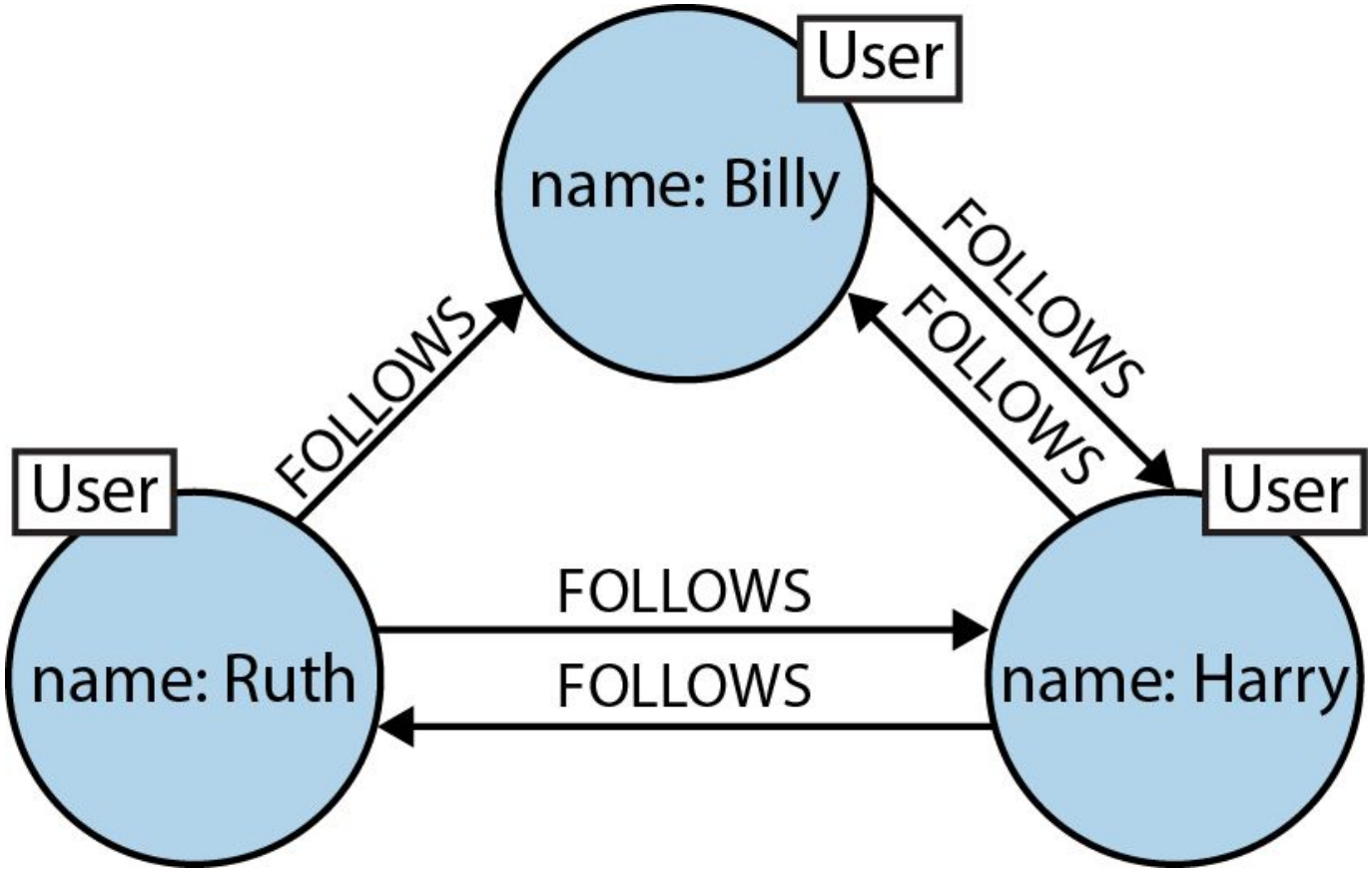


1736
Leonhard Euler

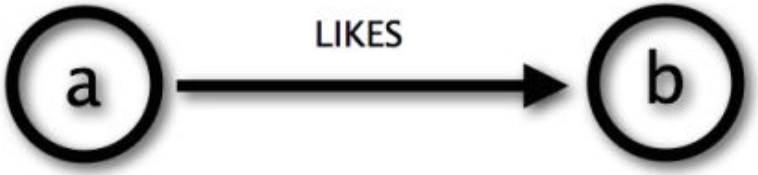


As pontes de Königsberg

Muitas situações do mundo real podem ser convenientemente descritas por meio de diagrama.


User
name: Billy
FOLLOWS
FOLLOWS
FOLLOWS
User
User
FOLLOWS
name: Ruth
FOLLOWS
name: Harry

Cypher using relationship 'likes'

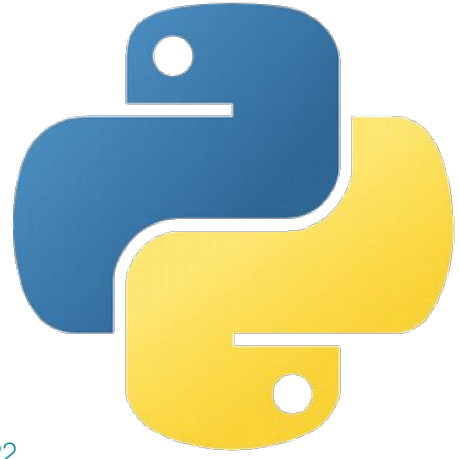


Cypher

```
(a) -[:LIKES]-> (b)
```

Surprise Python

A scikit building and analyzing recommender systems.



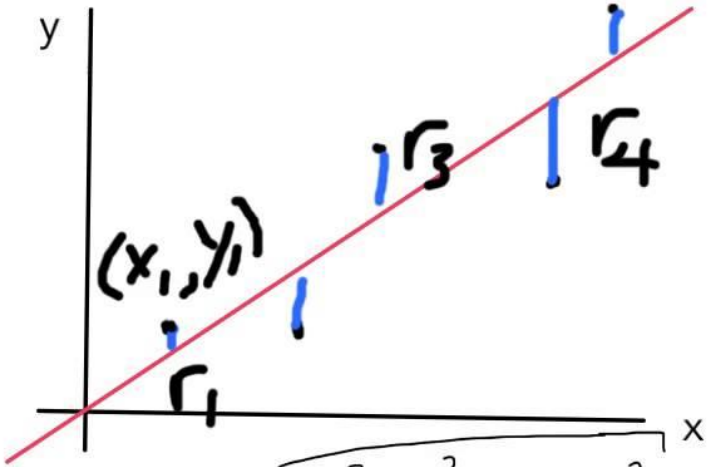
é um scikit Python para criar e analisar sistemas de recomendação.

- Implantação de vários algoritmos de recomendação
- Documentação rica e detalhada.
- Benchmark dos algoritmos sendo uma delas a medida RMSE.

Movielens 100k	RMSE	MAE	Time
SVD	0.934	0.737	0:00:11
SVD++	0.92	0.722	0:09:03
NMF	0.963	0.758	0:00:15
Slope One	0.946	0.743	0:00:08
k-NN	0.98	0.774	0:00:10
Centered k-NN	0.951	0.749	0:00:10
k-NN Baseline	0.931	0.733	0:00:12
Co-Clustering	0.963	0.753	0:00:03
Baseline	0.944	0.748	0:00:01
Random	1.514	1.215	0:00:01

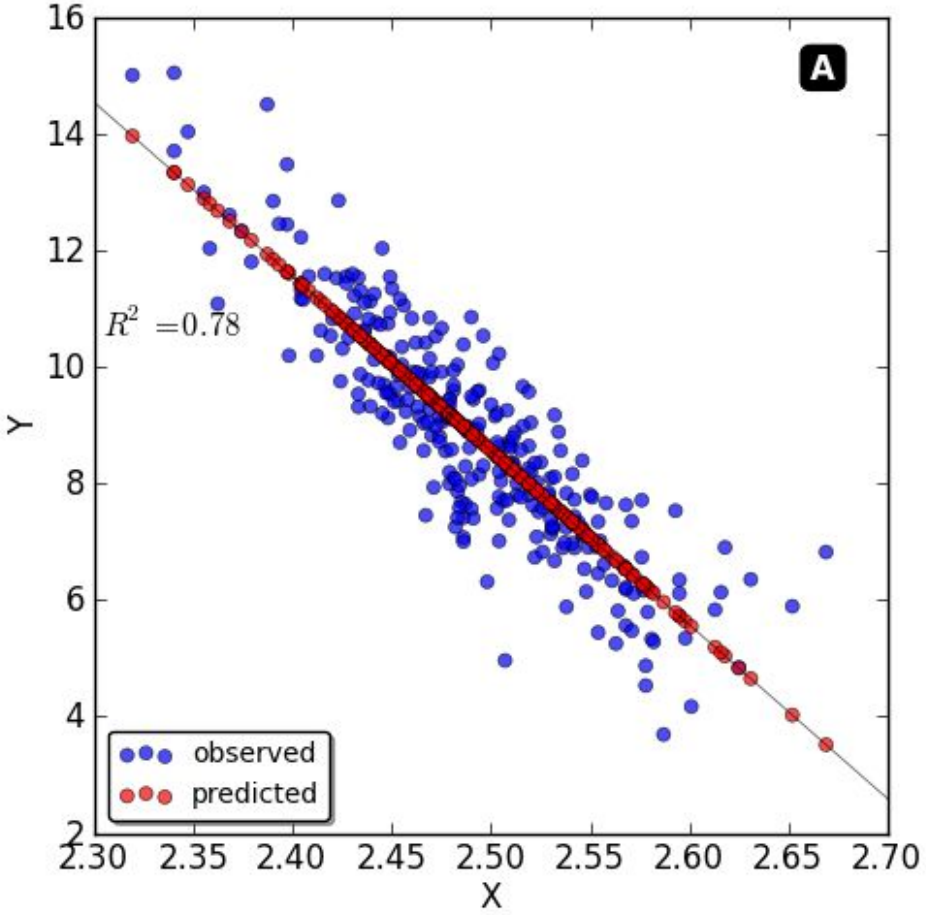
Movielens 1M	RMSE	MAE	Time
SVD	0.873	0.686	0:02:13
SVD++	0.862	0.673	2:54:19
NMF	0.916	0.724	0:02:31
Slope One	0.907	0.715	0:02:31
k-NN	0.923	0.727	0:05:27
Centered k-NN	0.929	0.738	0:05:43
k-NN Baseline	0.895	0.706	0:05:55
Co-Clustering	0.915	0.717	0:00:31
Baseline	0.909	0.719	0:00:19
Random	1.504	1.206	0:00:19

Root Mean Squared Error (RMSE)



$$RMSE = \sqrt{\frac{r_1^2 + r_2^2 + \cdots + r_n^2}{n}}$$

MAE – Mean Absolute Error



OBRIGADA!

Perguntas?

Meus contatos:
@morvanabonin
morvanabonin@gmail.com
morvana.bonin@king.host

REFERÊNCIAS

Estudo de Técnicas de Filtragem Híbrida em Sistemas de Recomendação de Produtos <http://www.cin.ufpe.br/~tg/2013-2/cemb.pdf>

Bridges of Königsberg and Graph Theory <http://www.mathscareers.org.uk/article/bridges-of-konigsberg-and-graph-theory>

Machine Learning for Recommender systems — Part 1 (algorithms, evaluation and cold start) <https://medium.com/recombee-blog/machine-learning-for-recommender-systems-part-1-algorithms-evaluation-and-cold-start-6f696683d0ed>

Mining of Massive Datasets <http://infolab.stanford.edu/~ullman/mmds/ch9.pdf>

Graph Databases Will Change Your Freakin' Life (Best Intro Into Graph Databases) <https://youtu.be/GekQqFZm7mA>

Neo4j o que? Uma visão prática do banco de dados orientado a grafos <https://neo4j.com/news/neo4j-o-que-uma-visao-pratica-do-banco-de-dados-orientado-a-grafos>

SciKits <http://scikits.appspot.com/>

Surprise Python <https://surprise.readthedocs.io/en/stable/getting_started.html>

Cypher Reference Card <https://neo4j.com/docs/pdf/neo4j-cypher-refcard-stable.pdf>

MAE and RMSE — Which Metric is Better? <https://medium.com/human-in-a-machine-world/mae-and-rmse-which-metric-is-better-e60ac3bde13d>