NOVOMATIC

Recommender Systems

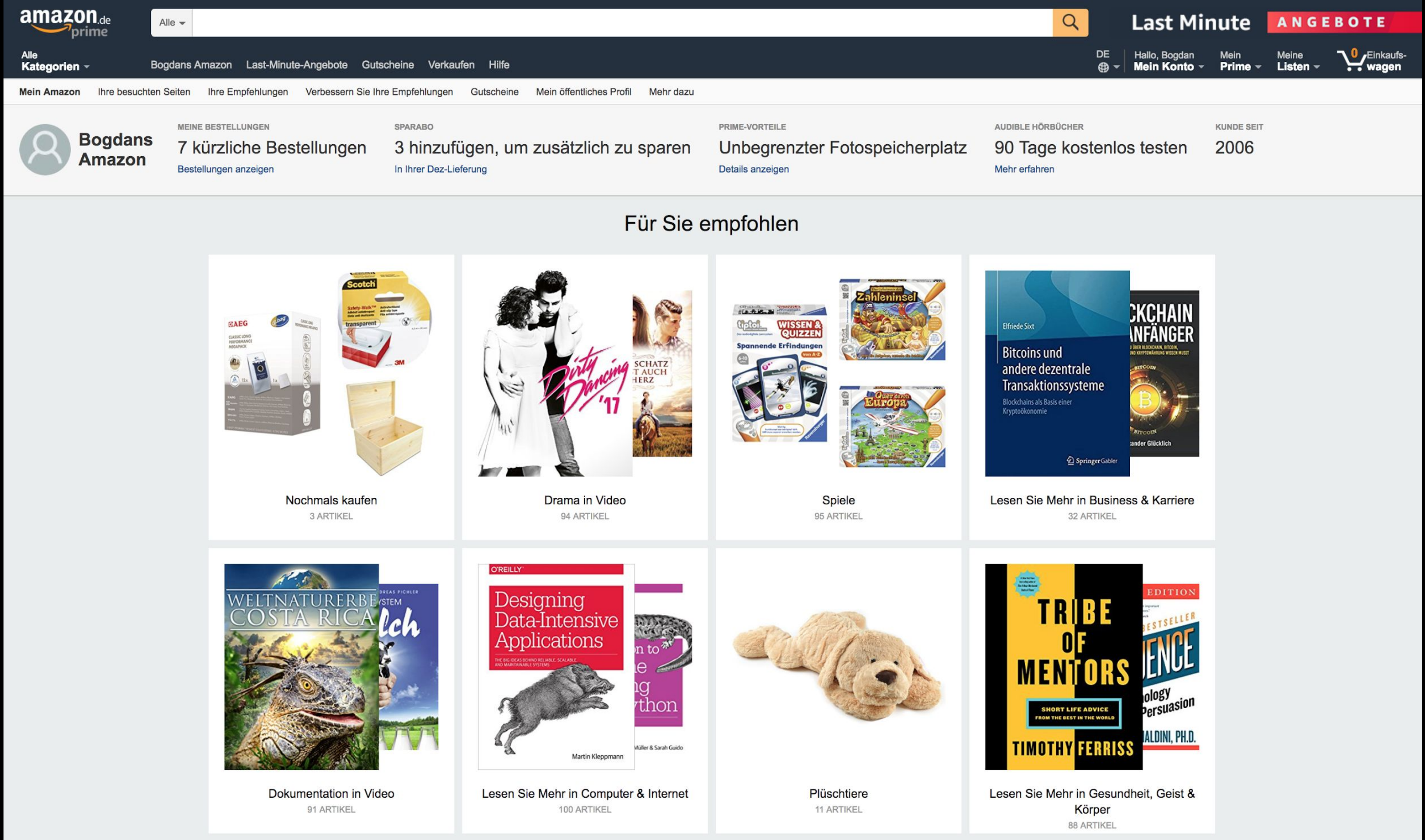
An Introduction

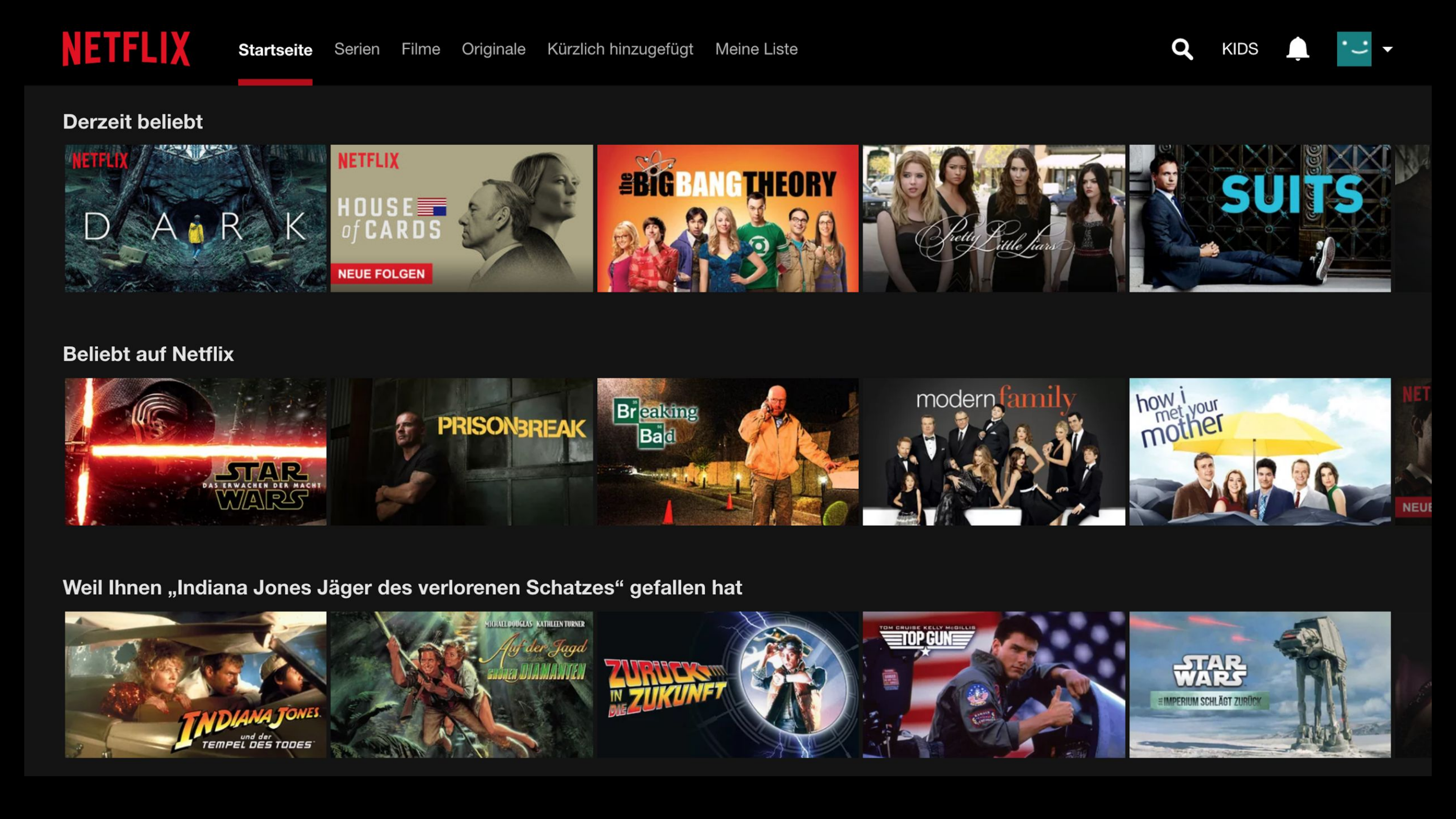
NOVOMATIC

Outline

- What are Recommender Systems
- Problem Formulation
- Algorithms
  - Collaborative Filtering
- Where to go next

What are Recommender Systems


amazon.de prime
Alle
Last Minute ANGEBOTE
Alle Kategorien
Bogdans Amazon
Last-Minute-Angebote
Gutscheine
Verkaufen
Hilfe
DE
Hallo, Bogdan Mein Konto
Mein Prime
Meine Listen
0 Einkaufs-wagen
Mein Amazon
Ihre besuchten Seiten
Ihre Empfehlungen
Verbessern Sie Ihre Empfehlungen
Gutscheine
Mein öffentliches Profil
Mehr dazu
Bogdans Amazon
MEINE BESTELLUNGEN
7 kürzliche Bestellungen
Bestellungen anzeigen
SPARABO
3 hinzufügen, um zusätzlich zu sparen
In Ihrer Dez-Lieferung
PRIME-VORTEILE
Unbegrenzter Fotospeicherplatz
Details anzeigen
AUDIBLE HÖRBÜCHER
90 Tage kostenlos testen
Mehr erfahren
KUNDE SEIT
2006
Für Sie empfohlen
Nochmals kaufen
3 ARTIKEL
Drama in Video
94 ARTIKEL
Spiele
95 ARTIKEL
Lesen Sie Mehr in Business & Karriere
32 ARTIKEL
Dokumentation in Video
91 ARTIKEL
Lesen Sie Mehr in Computer & Internet
100 ARTIKEL
Plüschtiere
11 ARTIKEL
Lesen Sie Mehr in Gesundheit, Geist & Körper
88 ARTIKEL
Dirty Dancing '17
Bitcoins und andere dezentrale Transaktionssysteme
Designing Data-Intensive Applications
TRIBE OF MENTORS
TIMOTHY FERRISS
WELTNATURERBE COSTA RICA


NETFLIX
Startseite
Serien
Filme
Originale
Kürzlich hinzugefügt
Meine Liste
KIDS
Derzeit beliebt
NETFLIX
DARK
NETFLIX
HOUSE of CARDS
NEUE FOLGEN
the BIG BANG THEORY
Pretty Little Liars
SUITS
Beliebt auf Netflix
STAR WARS
DAS ERWACHEN DER MACHT
PRISON BREAK
Breaking Bad
modern family
how i met your mother
Weil Ihnen „Indiana Jones Jäger des verlorenen Schatzes“ gefallen hat
INDIANA JONES
und der
TEMPEL DES TODES
MICHAEL DOUGLAS KATHLEEN TURNER
Auf der Jagd
nach dem
GRÜNEN DIAMANTEN
ZURÜCK IN DIE ZUKUNFT
TOM CRUISE KELLY McGILLIS
TOP GUN
STAR WARS
IMPERIUM SCHLÄGT ZURÜCK

Why do we need recommenders?

- Offerings from "The Long Tail[1]" must be presented to the right customers
- This should improve
  - User engagement
  - Retention rate
  - Conversion rate
  - Revenue
- Business value[2]
  - Netflix: 2/3 of the movies watched are recommended
  - Google News: recommendations generate 38% more click-through
  - Amazon: 35% sales from recommendations

---

[1] C.Anderson, "The Long Tail", 2008
[2] X.Amatriain lectures from 2014

Problem Formulation

Traditional Definition

Estimate an utility function that predicts how a user will like an item, based on:

- Past behavior
- Relations to other users
- Item similarity
- Context
- ...

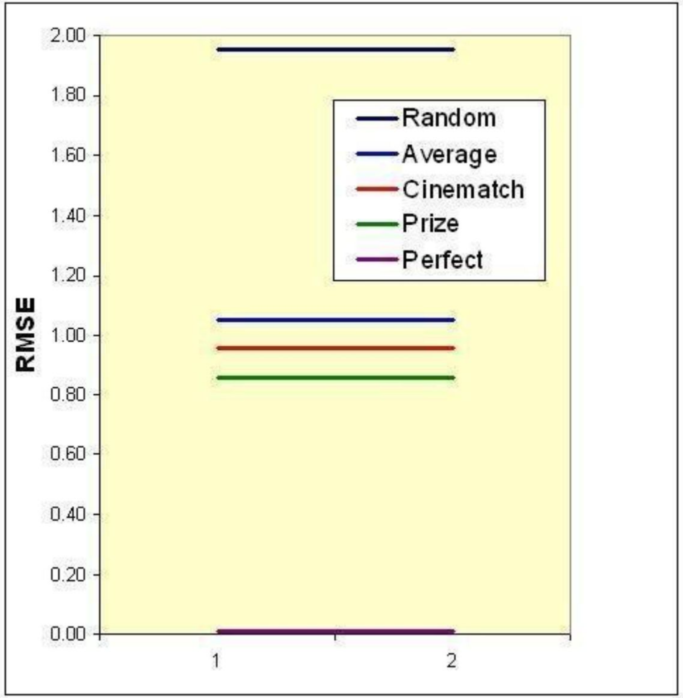
Target Variable

- User preferences
  - Explicit ratings
  - 1-5 stars ratings
- Purchase behavior
  - Implicit ratings
  - Yes / No
- Media consumption behavior
  - E.g. movies, songs or games
  - Implicit ratings
  - How many times played

Algorithms

The Netflix Prize

- 2006 - 2009
  - 1 million USD for 10% improvement of RMSE
  - ~ 100 million ratings
  - ~ 500 000 users
  - ~ 18 000 movies
- First conclusion[1] :
  - it is really extremely simple to produce "reasonable" recommendations and extremely difficult to improve them



[1] X.Amatriain lectures from 2014

Algorithm Types

- Non-personalized Recommendations
  - Popularity based
- Content-Based Filtering
  - Item features necessary
  - Each user is predicted independently
- Collaborative Filtering
  - Features are learned
  - All user ratings help improve the prediction
- ...

Collaborative Filtering

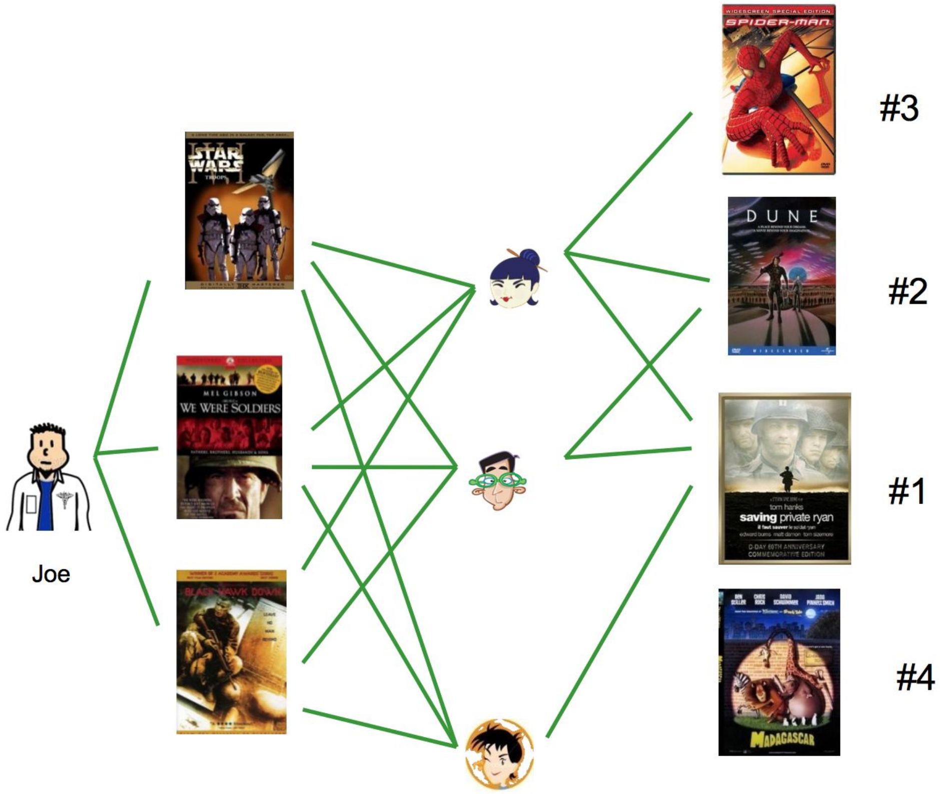
Ratings Matrix

2			4	5	
5		4			1
		5		2	
	1		5		4
		4			2
4	5		1		

CF Algorithms

- Neighborhood methods
  - User Similarity
  - Item Similarity
- Matrix Factorization

User Similarity - Concept



User Similarity - Implementation

2			4	5		NA
5		4			1	0.87
		5		2		1
	1		5		4	-1
		4			2	
4	5		1			NA

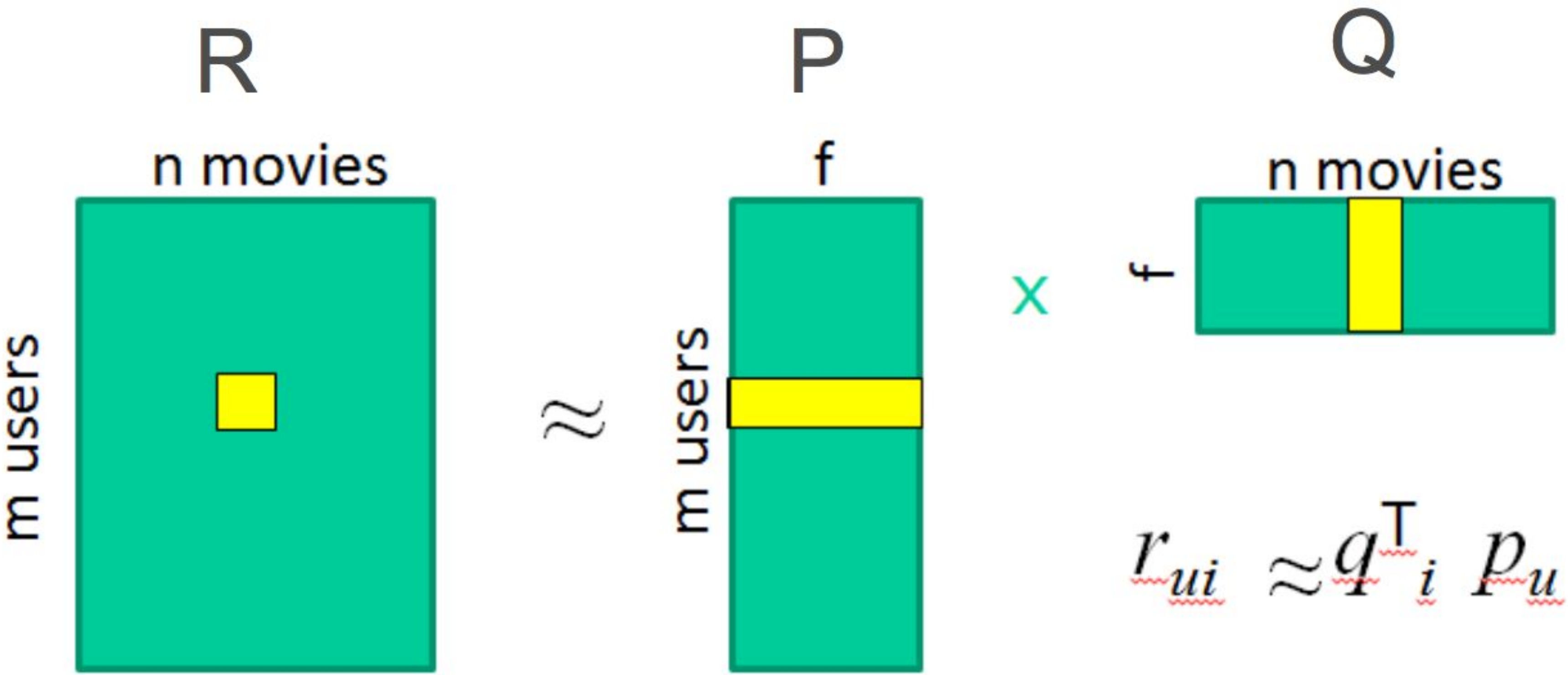
Item Similarity - Implementation

2			4	5	
5		4			1
		5		2	
	1		5		4
		4			2
4	5		1		
-1	-1	0.86	1	NA	

Matrix Factorization

2			4	5	
5		4			1
		5		2	
	1		5		4
		4			2
4	5		1		

Matrix Factorization



$$r_{ui} \approx q_i^T p_u$$

Problems of CF Methods

- Cold start
  - Changing user base
  - Changing inventory
- Matrix Factorization
  - Assumption about missing ratings
  - Retrain model when new ratings appear
- Neighborhood Methods
  - Similarity computation
  - Similarity matrix storage

Where to go next

Links

- Xavier Amatriain Lecture
  - http://technocalifornia.blogspot.co.at/
  - http://technocalifornia.blogspot.co.at/2014/08/introduction-to-recommender-systems-4.html
- Alex Smola Lecture
  - http://alex.smola.org/teaching/berkeley2012/recommender.html
- ACM RecSys Conference
  - https://recsys.acm.org/

Public Datasets

- MovieLens
  - https://grouplens.org/datasets/movielens/
- Netflix Prize
  - https://www.kaggle.com/netflix-inc/netflix-prize-data
- ACM RecSys Challenges
  - https://recsys.acm.org/

Libraries

- MyMediaLite
  - .NET, command line
- Surprise
  - Python
- Spark MLlib
  - Python, Scala, Java, R
- GraphLab Create
  - Python
- Lenskit
  - Java

Thank you!

bpirvu@novomatic.com
jwilms@novomatic.com

NOVOMATIC