

Sistemas de recomendación

¿Es posible ayudar al usuario a encontrar información sin que la pida?

¿Cómo personalizar el proceso?

Sistemas de recomendación



Users who bought this also bought...



Related videos



Spotify®

Music discovery



People you may know...



Google now

Quora



Samsung
SMART TV

Google
News

Walmart

Filmaffinity



iTunes



Google play

eBay

Dailymotion

last.fm

PANDORA

Google AdSense

slideShare

Pinterest

• • •

Tipos de recomendación

- ◆ Personalizada vs. no personalizada
- ◆ Contextual vs. fuera de contexto
- ◆ Formas de envío / presentación
 - En la misma aplicación (online) vs. por mail / notificaciones (offline)
 - En primer plano vs. lateral
 - Solicitudada por el usuario vs. a iniciativa de la aplicación
- ◆ Tipo de objetos recomendados
 - Caso particular: otros usuarios en redes sociales

Recomendación personalizada

Secure | https://www.amazon.com/gp/yourstore/home/ref=nav_cs_yys

Recommended for you, Pablo

The screenshot shows a grid of recommended items on the Amazon homepage. The items are categorized into nine groups:

- Engineering Books:** Includes "The Digital Divide" and "The Internet Is Not Like Us". 31 items.
- Art & Photography Books:** Includes "Astra Taylor's The People's Platform: Taking Back Power and Culture in the Digital Age". 7 items.
- Self-Help Books:** Includes "The Go-Giver: A Little Story About a Powerful Business Idea". 24 items.
- Humor & Entertainment Books:** Includes "A Man for All Markets" by Edward O. Thorp. 23 items.
- Reference Books:** Includes "Blur: How to Know What's True in the Age of Information Overload" by Bill Kovach and Tom Rosenstiel, and "Blur: Standage". 21 items.
- Biographies & Memoirs:** Includes "The Undoing Project" by Michael Lewis and "The Undoing Project" by Steven Levy. 7 items.
- Movies & TV:** Includes movie posters for "Arrival", "Mildred Pierce", "Annie Hall", and "Forces of Nature". 53 items.
- Comedy in Video:** Includes a poster for "Annie Hall". 86 items.
- Romance in Video:** Includes a poster for "Forces of Nature". 26 items.

amazon

Recomendación no personalizada



Recomendación contextual



Contexto

Abbey Road (Remastered) [Enhanced, Limited Edition, Original Recording Remastered]
The Beatles | Format: Audio CD
4.5 stars (1,260 customer reviews) | Like (42)
Price: \$13.88 & eligible for FREE Super Saver Shipping on orders over \$25. [Details](#)
[Special Offers Available](#)
In Stock.
Ships from and sold by **Amazon.com**. Gift-wrap available.
Want it delivered Tuesday, April 12? Order it in the next 33 hours and 2 minutes, and choose One-Day Shipping at check
57 new from \$10.78 16 used from \$8.35 1 collectible from \$30.68

Customers Who Bought This Item Also Bought



Sgt. Pepper's Lonely
Hearts Club Band
(Remastered) ~ The
Beatles
4.5 stars (1,325)
\$14.49



Let It Be (Remastered)
~ The Beatles
4.5 stars (435)
\$14.66



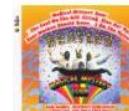
The White Album
(Remastered) ~ The
Beatles
4.5 stars (1,155)
\$18.99



Revolver (Remastered)
~ The Beatles
4.5 stars (923)
\$14.64



Rubber Soul
(Remastered) ~ The
Beatles
4.5 stars (725)
\$14.49



Magical Mystery
(Remastered) ~
Beatles
4.5 stars (531)
\$14.85



A Hard Day's Night
(Remastered) ~ The
Beatles
4.5 stars (307)
\$14.60



Please Please Me
(Remastered) ~ The
Beatles
4.5 stars (272)
\$14.69



Dark Side of the Moon ~
Pink Floyd
4.5 stars (1,360)
\$11.44



With The Beatles
(Remastered) ~ The
Beatles
4.5 stars (221)
\$14.55



The Beatles Mono Box
Set ~ The Beatles
4.5 stars (288)
\$207.82



Beatles For Sale
(Remastered) ~ T
Beatles
4.5 stars (270)
\$14.98



13:09 4G 55% Radio France

radio france

Radio France FM RadioFMApp 4.5★

France Radio Stations Online - Fr.. World Radio FM AM, Music and News 4.6★

Radios France - France Radio Sta... Radio Pro Free Apps - Bizzlead 5.0★

Recommended for you

JAZZ RADIO FRANCE 24 Jazz & Blues Music Radio 4.8★

FRANCE 24 FRANCE 24 Radios En Ligne 4.3★

Radios France Online radios 4.6★

Recomendación contextual / personalizada

The diagram illustrates the difference between contextual and personalized recommendations on YouTube across two platforms: a mobile phone and a desktop browser.

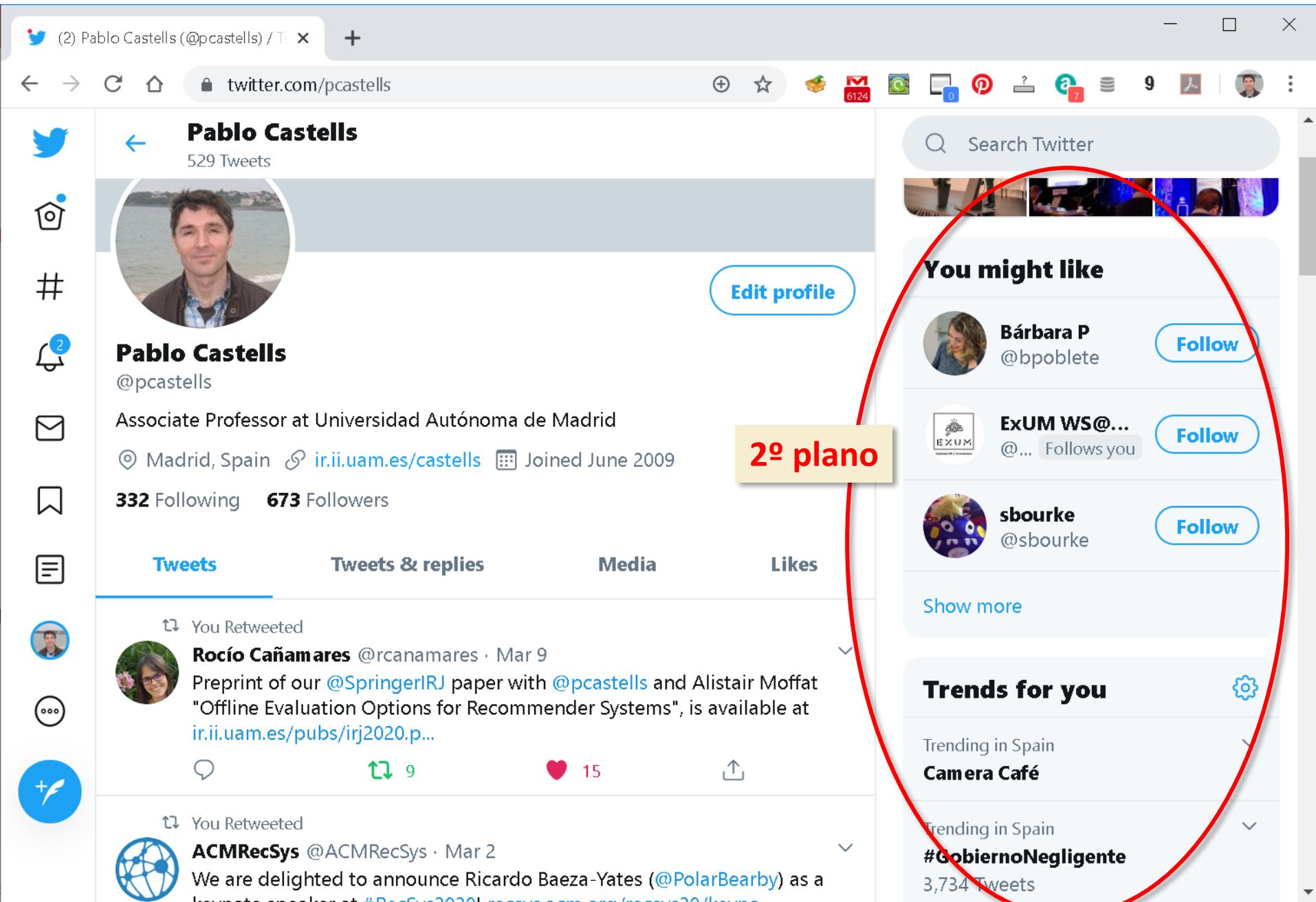
Contexto (Contextual Recommendations): This section shows a mobile phone screen displaying a video from "Real Time with Bill Maher" and a desktop browser showing a video by Michio Kaku. Both screens have blue circles highlighting their respective video content.

- Mobile Phone (Left):** Shows a video thumbnail for "Monologue: Repeal and Disgrace | Real Time with Bill Maher (HBO)". Below it, the "Up next" section lists "New Rule: Trump and the Long Con | Real Time with B..." and "Trump Vows to End the Nonexistent War on Coal: T...".
- Desktop Browser (Right):** Shows a video thumbnail for "Michio Kaku - Trappist-1 Solar System & Listener Questions". Below it, the "Up next" section lists various science-related videos like "Latest Secrets of The Solar System - Full BBC Documentary" and "M theory, String Theory, Parallel Worlds - Dr Michio Kaku".

Personalizada (Personalized Recommendations): This section shows a desktop browser screen where a red circle highlights a video thumbnail for "Alec Baldwin's Trump Returns to SNL in a State of lauren gillies Recommended for you". A red bracket labeled "Personalizada" points to this video.

1^{er} vs. 2^º plano (1st vs. 2nd tier): A large red bracket at the bottom right spans both the "Personalizada" and "Contexto" sections, indicating the hierarchy of recommendation levels.

2^º plano (2nd tier): A red bracket at the bottom right points to the "Personalizada" section, specifically highlighting the recommended video for Alec Baldwin.



Recomendación offline

How many milliseconds can a human remain alive exposed to the vacuum of space?

Quora Digest <digest-noreply@quora.com>
to pablo.castells [redacted]
Mar 12 (3 days ago)

Quora

Pablo's Digest

TOP STORIES FOR YOU

How many milliseconds can a human remain alive exposed to the vacuum of space?

Valentin Ghincolov, Amateur astronomer, and really passionate about Relativity and physics
Updated Jan 22

Oh, again so many misconceptions and myths. TL:DR: consciousness lost in 8-15 seconds. Death after 1 to 3 minutes Successful CPR resuscitation after up to 6 minutes. Details:... [Read More »](#)

In medieval times, how were fully armoured knights killed?

Jerimi Galligory, I read history books occasionally.
Written 15h ago

I'm going to assume that by "fully armoured" you mean "full plate armour." So, this: There are three different ways you can counter this guy. The first choice would be to pou... [Read More »](#)



How did your job prospects change after taking Andrew Ng's Deep Learning course?

Richard Ackon, Machine Learning Engineer at Kudobuzz (2018-present)

La tarea de recomendación (cont)

Confluencia de dos (o más) campos

- ◆ Se puede ver como un caso particular de IR
- ◆ Pero también un problema de clasificación / aprendizaje automático

El sistema...

1. **Observa** al usuario mientras éste realiza actividades
2. **Detecta** patrones de comportamiento, identifica indicios de interés del usuario
3. **Predice y sugiere** al usuario opciones de interés
 - Predice intereses futuros basados en indicios de intereses pasados

La tarea de recomendación (cont)

- ◆ Espacio de recuperación
 - Cualquier tipo de objeto útil que tenga sentido buscar: “ítems”
 - Documentos Web, productos (libros, películas, música, juegos, apps, ropa...), noticias, vídeos, servicios, eventos, ofertas de empleo, conversaciones, grupos, personas...
- ◆ Evidencia (indicios) de la necesidad de información
 - Típicamente consiste en interacción entre los usuarios y los objetos a recomendar, observada por el sistema: compras, inspección, play, click, like, puntuaciones, etc.
 - Pero puede ser cualquier otro tipo de actividad del usuario (la “traza digital”): búsquedas, navegación, posts, interacción social, etiquetas y comentarios asignados a ítems, etc.
 - Puede utilizarse también información propia del usuario: demográfica, geoposicional, etc.
 - Los algoritmos basados en contenido utilizan además datos de los ítems
- ◆ Dada toda esta evidencia, generar un ranking de ítems para cada usuario
 - Típicamente (pero no necesariamente) personalizado (i.e. distinto para cada usuario)

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2				4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

5 – Favorito

4 – Me gusta

3 – Normal

2 – No me gusta

1 – Pésima

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2				4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

- 5 – Favorito
 4 – Me gusta
 3 – Normal
 2 – No me gusta
 1 – Pésima

Notación: $r(Claudia, The Walking Dead) = 2$

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother	
Claudia	5	2	5	4	?	4	5 – Favorito
Ignacio	3	?	?	5	?	4	4 – Me gusta
Lucía	3	2	?	?	?	4	3 – Normal
Pablo	?	2	?	3	2	3	2 – No me gusta
Sergio	2	4	?	4	3	2	1 – Pésima

Recomendar:

- Estimar/predecir los valores desconocidos “?”
- Y con ellos generar rankings de ítems para cada usuario

Grey's Anatomy
 The Walking Dead
 Killing Eve
Peaky Blinders
 The Wire
 How I met your mother

	Grey's Anatomy	The Walking Dead	Killing Eve	The Wire	How I met your mother	
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2		?		4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

$$\hat{r}(Lucía, Peaky Blinders) = ?$$

Grey's Anatomy
 The Walking Dead
 Killing Eve
Peaky Blinders
 The Wire
 How I met your mother

	Grey's Anatomy	The Walking Dead	Killing Eve	The Wire	How I met your mother	
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2		?		4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

Opinión de expertos

$$\hat{r}(Lucía, Peaky Blinders) = ?$$

Grey's Anatomy
 The Walking Dead
 Killing Eve
Peaky Blinders
 The Wire
 How I met your mother

	Grey's Anatomy	The Walking Dead	Killing Eve	The Wire	How I met your mother	
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2		?		4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

Opinión de expertos

$$\hat{r}(Lucía, Peaky Blinders) = 5$$

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2		?		4
Pablo		2		3	2	3
Sergio	2	4		4	3	2
	13	10	5	16	5	17

Opinión de la mayoría:
suma de opiniones

$$\hat{r}(Lucía, Peaky Blinders) = r(Claudia, Peaky Blinders)$$

$$+ r(Ignacio, Peaky Blinders)$$

$$+ r(Pablo, Peaky Blinders)$$

$$+ r(Sergio, Peaky Blinders) = \mathbf{16}$$

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2		?		4
Pablo		2		3	2	3
Sergio	2	4		4	3	2
	4	4	1	4	2	5

Opinión de la mayoría:
número de opiniones

$$\hat{r}(Lucía, Peaky Blinders) = 1_{r(Claudia, Peaky Blinders) \neq \emptyset}$$

$$+ 1_{r(Ignacio, Peaky Blinders) \neq \emptyset}$$

$$+ 1_{r(Pablo, Peaky Blinders) \neq \emptyset}$$

$$+ 1_{r(Sergio, Peaky Blinders) \neq \emptyset} = 4$$

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2		?		4
Pablo		2		3	2	3
Sergio	2	4		4	3	2
	1	1	1	3	0	3

Opinión de la mayoría:
número de buenas opiniones

(p.e. “me gusta” \Leftrightarrow rating ≥ 4)

$$\hat{r}(Lucía, Peaky Blinders) = 1_{r(Claudia, Peaky Blinders) \geq 4}$$

$$+ 1_{r(Ignacio, Peaky Blinders) \geq 4}$$

$$+ 1_{r(Pablo, Peaky Blinders) \geq 4}$$

$$+ 1_{r(Sergio, Peaky Blinders) \geq 4} = 3$$

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2		?		4
Pablo		2		3	2	3
Sergio	2	4		4	3	2
	3.25	2.5	5	4	2.5	3.4

Opinión de la mayoría:
opinión promedio

$$\hat{r}(Lucía, Peaky Blinders) = r(Claudia, Peaky Blinders)$$

$$+ r(Ignacio, Peaky Blinders)$$

$$+ r(Pablo, Peaky Blinders)$$

$$+ r(Sergio, Peaky Blinders) = 16$$

4 = 4

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2		?		4
Pablo		2		3	2	3
Sergio	2	4		4	3	2
	13	10	5	16	5	17

Opinión de la mayoría:
suma de opiniones

$$\hat{r}(Lucía, Peaky Blinders) = r(Claudia, Peaky Blinders)$$

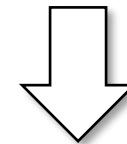
$$+ r(Ignacio, Peaky Blinders)$$

$$+ r(Pablo, Peaky Blinders)$$

$$+ r(Sergio, Peaky Blinders) = \mathbf{16}$$

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2		?		4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

Opinión de la mayoría
personalizada: suma ponderada



kNN basado en usuario

$$\hat{r}(\text{Lucía}, \text{Peaky Blinders}) = w_{\text{Lucía}, \text{Claudia}} r(\text{Claudia}, \text{Peaky Blinders})$$

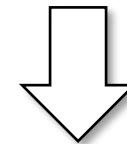
$$+ w_{\text{Lucía}, \text{Ignacio}} r(\text{Ignacio}, \text{Peaky Blinders})$$

$$+ w_{\text{Lucía}, \text{Pablo}} r(\text{Pablo}, \text{Peaky Blinders})$$

$$+ w_{\text{Lucía}, \text{Sergio}} r(\text{Sergio}, \text{Peaky Blinders})$$

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2		?		4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

Opinión de la mayoría
personalizada: suma ponderada



kNN basado en usuario

$$\begin{aligned}
 \hat{r}(\text{Lucía}, \text{Peaky Blinders}) &= \text{sim}(\text{Lucía}, \text{Claudia}) r(\text{Claudia}, \text{Peaky Blinders}) \\
 &\quad + \text{sim}(\text{Lucía}, \text{Ignacio}) r(\text{Ignacio}, \text{Peaky Blinders}) \\
 &\quad + \text{sim}(\text{Lucía}, \text{Pablo}) r(\text{Pablo}, \text{Peaky Blinders}) \\
 &\quad + \text{sim}(\text{Lucía}, \text{Sergio}) r(\text{Sergio}, \text{Peaky Blinders})
 \end{aligned}$$

Grey's Anatomy
 The Walking Dead
 Killing Eve
Peaky Blinders
 The Wire
 How I met your mother

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2		?		4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

kNN basado en usuario
 (coseno)

$$sim(Lucía, Claudia) \equiv \cos(Lucía, Claudia)$$

$$sim(Lucía, Ignacio) \equiv \cos(Lucía, Ignacio)$$

$$sim(Lucía, Pablo) \equiv \cos(Lucía, Pablo)$$

$$sim(Lucía, Sergio) \equiv \cos(Lucía, Sergio)$$

Grey's Anatomy
 The Walking Dead
 Killing Eve
Peaky Blinders
 The Wire
 How I met your mother

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2		?		4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

kNN basado en usuario
 (coseno)

$$\cos(Lucía, Claudia) = \frac{3 \cdot 5 + 2 \cdot 2 + 4 \cdot 4}{\sqrt{3^2 + 2^2 + 4^2} \sqrt{5^2 + 2^2 + 5^2 + 4^2 + 4^2}} = \mathbf{0.70}$$

$$\cos(Lucía, Ignacio) = \frac{3 \cdot 3 + 4 \cdot 4}{\sqrt{3^2 + 2^2 + 4^2} \sqrt{3^2 + 5^2 + 4^2}} = \mathbf{0.66}$$

$$\cos(Lucía, Pablo) = \frac{2 \cdot 2 + 4 \cdot 3}{\sqrt{3^2 + 2^2 + 4^2} \sqrt{2^2 + 3^2 + 2^2 + 3^2}} = \mathbf{0.58}$$

$$\cos(Lucía, Sergio) = \frac{3 \cdot 2 + 2 \cdot 4 + 4 \cdot 2}{\sqrt{3^2 + 2^2 + 4^2} \sqrt{2^2 + 4^2 + 4^2 + 3^2 + 2^2}} = \mathbf{0.58}$$

Grey's Anatomy
 The Walking Dead
 Killing Eve
Peaky Blinders
 The Wire
 How I met your mother

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2		?		4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

kNN basado en usuario
 (coseno)

$$\cos(Lucía, Claudia) = 0.70$$

$$\cos(Lucía, Ignacio) = 0.66$$

$$\cos(Lucía, Pablo) = 0.58$$

$$\cos(Lucía, Sergio) = 0.58$$

$$\hat{r}(Lucía, Peaky Blinders) =$$

$$0.70 \cdot 4 + 0.66 \cdot 5 + 0.58 \cdot 3 + 0.58 \cdot 4 = \mathbf{10.17}$$

Grey's Anatomy
The Walking Dead
Killing Eve
Peaky Blinders
The Wire
How I met your mother

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2		?	?	4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

kNN basado en usuario
(coseno)

$$\cos(Lucía, Claudia) = 0.70$$

$$\hat{r}(Lucía, Peaky Blinders) =$$

$$0.70 \cdot 4 + 0.66 \cdot 5 + 0.58 \cdot 3 + 0.58 \cdot 4 = 10.17$$

$$\cos(Lucía, Ignacio) = 0.66$$

$$\hat{r}(Lucía, The Wire) = 0.58 \cdot 2 + 0.58 \cdot 3 = \mathbf{2.92}$$

$$\cos(Lucía, Pablo) = 0.58$$

$$\cos(Lucía, Sergio) = 0.58$$

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2	?	?	?	4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

kNN basado en usuario
(coseno)

$$\cos(Lucía, Claudia) = 0.70$$

$$\hat{r}(Lucía, Peaky Blinders) =$$

$$0.70 \cdot 4 + 0.66 \cdot 5 + 0.58 \cdot 3 + 0.58 \cdot 4 = 10.17$$

$$\cos(Lucía, Ignacio) = 0.66$$

$$\hat{r}(Lucía, The Wire) = 0.58 \cdot 2 + 0.58 \cdot 3 = 2.92$$

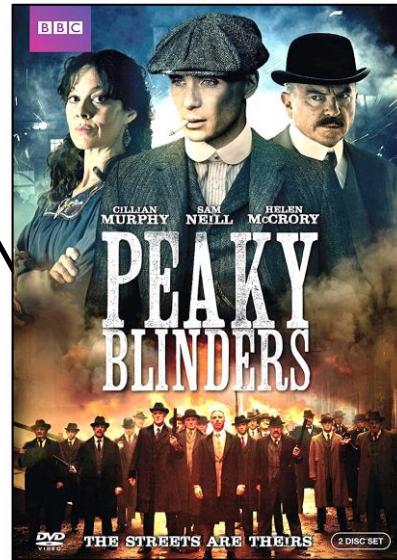
$$\cos(Lucía, Pablo) = 0.58$$

$$\hat{r}(Lucía, Killing Eve) = 0.70 \cdot 5 = \mathbf{3.5}$$

$$\cos(Lucía, Sergio) = 0.58$$

Grey's Anatomy
The Walking Dead
Killing Eve
Peaky Blinders
The Wire
How I met your mother

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2	?	?	?	4
Pablo		2		3	2	3
Sergio	2	4		4	3	2



kNN
ario

$$\cos(Lucía, Claudia) = 0.70$$

$$\hat{r}(Lucía, Peaky Blinders) =$$

$$0.70 \cdot 4 + 0.66 \cdot 5 + 0.58 \cdot 3 + 0.58 \cdot 4 = 10.17$$

$$\cos(Lucía, Ignacio) = 0.66$$

$$\hat{r}(Lucía, The Wire) = 0.58 \cdot 2 + 0.58 \cdot 3 = 2.92$$

$$\cos(Lucía, Pablo) = 0.58$$

$$\hat{r}(Lucía, Killing Eve) = 0.70 \cdot 5 = 3.5$$

$$\cos(Lucía, Sergio) = 0.58$$

Grey's Anatomy
The Walking Dead
Killing Eve
Peaky Blinders
The Wire
How I met your mother

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2	?	?	?	4
Pablo		2		3	2	3
Sergio	2	4		4	3	2



$$\cos(Lucía, Claudia) = 0.70$$

$$\cos(Lucía, Ignacio) = 0.66$$

$$\cos(Lucía, Pablo) = 0.58$$

$$\cos(Lucía, Sergio) = 0.58$$

$$\hat{r}(Lucía, Peaky Blinders) =$$

$$\frac{0.70 \cdot 4 + 0.66 \cdot 5 + 0.58 \cdot 3 + 0.58 \cdot 4}{|0.70 + 0.66 + 0.58 + 0.58|} = 4.03$$

$$\hat{r}(Lucía, The Wire) = \frac{0.58 \cdot 2 + 0.58 \cdot 3}{|0.58 + 0.58|} = 2.5$$

$$\hat{r}(Lucía, Killing Eve) = \frac{0.70 \cdot 5}{|0.70|} = 5$$

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2	?	?	?	4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

kNN normalizado
basado en usuario
(coseno)

$$\cos(Lucía, Claudia) = 0.70$$

$$\cos(Lucía, Ignacio) = 0.66$$

$$\cos(Lucía, Pablo) = 0.58$$

$$\cos(Lucía, Sergio) = 0.58$$

$$\hat{r}(Lucía, Peaky Blinders) =$$

$$\frac{0.70 \cdot 4 + 0.66 \cdot 5 + 0.58 \cdot 3 + 0.58 \cdot 4}{|0.70 + 0.66 + 0.58 + 0.58|} = 4.03$$

$$\hat{r}(Lucía, The Wire) = \frac{0.58 \cdot 2 + 0.58 \cdot 3}{|0.58 + 0.58|} = 2.5$$

$$\hat{r}(Lucía, Killing Eve) = \frac{0.70 \cdot 5}{|0.70|} = 5$$

Poco
fiable

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2		4		4
Ignacio	3			5		4
Lucía	3	2	?	?	?	4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

kNN normalizado
basado en usuario
(coseno)

$$\cos(Lucía, Claudia) = 0.70$$

$$\cos(Lucía, Ignacio) = 0.66$$

$$\cos(Lucía, Pablo) = 0.58$$

$$\cos(Lucía, Sergio) = 0.58$$

$$\hat{r}(Lucía, Peaky Blinders) =$$

$$\frac{0.70 \cdot 4 + 0.66 \cdot 5 + 0.58 \cdot 3 + 0.58 \cdot 4}{|0.70 + 0.66 + 0.58 + 0.58|} = 4.03$$

$$\hat{r}(Lucía, The Wire) = \frac{0.58 \cdot 2 + 0.58 \cdot 3}{|0.58 + 0.58|} = 2.5$$

$$\hat{r}(Lucía, Killing Eve) = \text{NaN} \leftarrow$$

Nuevo ítem

Grey's Anatomy
 The Walking Dead
 Killing Eve
Peaky Blinders
 The Wire
 How I met your mother

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2		?		4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

kNN basado en usuario
 (coseno)

$$\begin{aligned}
 \hat{r}(Lucía, Peaky Blinders) &= sim(Lucía, Claudia) r(Claudia, Peaky Blinders) \\
 &\quad + sim(Lucía, Ignacio) r(Ignacio, Peaky Blinders) \\
 &\quad + sim(Lucía, Pablo) r(Pablo, Peaky Blinders) \\
 &\quad + sim(Lucía, Sergio) r(Sergio, Peaky Blinders)
 \end{aligned}$$

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2		?		4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

kNN basado en ítem
(coseno)

$$\hat{r}(Lucía, Peaky Blinders) =$$

$$sim(Peaky Blinders, Grey's Anatomy) \quad r(Lucía, Grey's Anatomy)$$

$$+ sim(Peaky Blinders, The Walking Dead) \quad r(Lucía, The Walking Dead)$$

$$+ sim(Peaky Blinders, How I Met ...) \quad r(Lucía, How I Met ...)$$

Grey's Anatomy
 The Walking Dead
 Killing Eve
Peaky Blinders
 The Wire
 How I met your mother

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2		?		4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

kNN basado en ítem
 (coseno)

$$\cos(Peaky\ Blinders, Grey's\ ...) = \frac{4 \cdot 5 + 5 \cdot 3 + 4 \cdot 2}{\sqrt{4^2 + 5^2 + 3^2 + 4^2} \sqrt{5^2 + 3^2 + 3^2 + 2^2}}$$

$$\cos(Peaky\ Blinders, Walking\ ...) = \frac{4 \cdot 2 + 3 \cdot 2 + 4 \cdot 4}{\sqrt{4^2 + 5^2 + 3^2 + 4^2} \sqrt{2^2 + 2^2 + 2^2 + 4^2}}$$

$$\cos(Peaky\ Blinders, How\ I\ ...) = \frac{4 \cdot 4 + 5 \cdot 4 + 3 \cdot 3 + 4 \cdot 2}{\sqrt{4^2 + 5^2 + 3^2 + 4^2} \sqrt{4^2 + 4^2 + 4^2 + 3^2 + 2^2}}$$

Grey's Anatomy
 The Walking Dead
 Killing Eve
Peaky Blinders
 The Wire
 How I met your mother

	Grey's Anatomy	The Walking Dead	Killing Eve	The Wire	How I met your mother	
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2		?		4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

kNN basado en ítem
(coseno)

$$\cos(Peaky\ Blinders, Grey's\ ...) = 0.77$$

$$\hat{r}(Lucía, Peaky\ Blinders) = \\ 0.77 \cdot 3 + 0.70 \cdot 2 + 0.84 \cdot 4 = \mathbf{7.05}$$

$$\cos(Peaky\ Blinders, Walking\ ...) = 0.70$$

$$\cos(Peaky\ Blinders, How\ I\ ...) = 0.84$$

Grey's Anatomy
 The Walking Dead
 Killing Eve
 Peaky Blinders
The Wire
 How I met your mother

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	How I met your mother	
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2			?	4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

kNN basado en ítem
(coseno)

$$\cos(\text{The Wire}, \text{Grey's ...}) = 0.24$$

$$\begin{aligned} \hat{r}(\text{Lucía}, \text{Peaky Blinders}) &= \\ 0.77 \cdot 3 + 0.70 \cdot 2 + 0.84 \cdot 4 &= 7.05 \\ \hat{r}(\text{Lucía}, \text{Peaky The Wire}) &= \mathbf{4.11} \end{aligned}$$

$$\cos(\text{The Wire}, \text{Walking ...}) = 0.84$$

$$\cos(\text{The Wire}, \text{How I ...}) = 0.43$$

Killing Eve

Grey's Anatomy
The Walking Dead
Peaky Blinders
The Wire
How I met your mother

	Grey's Anatomy	The Walking Dead	Peaky Blinders	The Wire	How I met your mother	
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2	?			4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

kNN basado en ítem
(coseno)

$$\cos(Killing\ Eve, Grey's \dots) = 0.73$$

$$\cos(Killing\ Eve, Walking \dots) = 0.38$$

$$\cos(Killing\ Eve, How\ I \dots) = 0.51$$

$$\begin{aligned} \hat{r}(Lucía, Peaky\ Blinders) &= \\ 0.77 \cdot 3 + 0.70 \cdot 2 + 0.84 \cdot 4 &= 7.05 \\ \hat{r}(Lucía, Peaky\ The\ Wire) &= 4.11 \\ \hat{r}(Lucía, Peaky\ Killing\ Eve) &= \mathbf{4.99} \end{aligned}$$

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2	?			4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

kNN basado en ítem
(coseno)

$$\cos(Killing\ Eve, Grey's \dots) = \mathbf{0.73}$$

$$\cos(Killing\ Eve, Walking \dots) = \mathbf{0.38}$$

$$\cos(Killing\ Eve, How\ I \dots) = \mathbf{0.51}$$

$$\begin{aligned} \hat{r}(Lucía, Peaky\ Blinders) &= \\ 0.77 \cdot 3 + 0.70 \cdot 2 + 0.84 \cdot 4 &= 7.05 \\ \hat{r}(Lucía, Peaky\ The\ Wire) &= 4.11 \\ \hat{r}(Lucía, Peaky\ Killing\ Eve) &= \mathbf{4.99} \end{aligned}$$

Grey's Anatomy
The Walking Dead
Killing Eve
Peaky Blinders
The Wire
How I met your mother

	Grey's Anatomy	The Walking Dead	Peaky Blinders	The Wire	How I met your mother	
Claudia	5	2		4		4
Ignacio	3			5		4
Lucía	3	2	?			4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

kNN basado en ítem
(coseno)

$$\cos(Killing\ Eve, Grey's \dots) = \text{NaN}$$

$$\hat{r}(Lucía, Peaky\ Blinders) = \\ 0.77 \cdot 3 + 0.70 \cdot 2 + 0.84 \cdot 4 = 7.05$$

$$\cos(Killing\ Eve, Walking \dots) = \text{NaN}$$

$$\hat{r}(Lucía, Peaky\ The\ Wire) = 4.11$$

$$\cos(Killing\ Eve, How\ I \dots) = \text{NaN}$$

$$\hat{r}(Lucía, Peaky\ Killing\ Eve) = \text{NaN}$$

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2		4		4
Ignacio	3			5		4
Lucía	3	2	?			4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2		4		4
Ignacio	3			5		4
Lucía	3	2	?			4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

1. Métodos basados en contenido: kNN (p.e. Jaccard)

Comedy					1
Crime			1	1	1
Drama	1	1	1	1	1
Romance	1		1		
Thriller		1	1	1	1

Características: géneros, director, reparto...

$$Jaccard(Killing Eve, Grey's \dots) = 2/4 = 0.5$$

$$Jaccard(Killing Eve, Walking \dots) = 2/4 = 0.5$$

$$Jaccard(Killing Eve, How I \dots) = 1/5 = 0.2$$

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2		4		4
Ignacio	3			5		4
Lucía	3	2	?			4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

1. Métodos basados en contenido: kNN (p.e. Jaccard)

Comedy					1
Crime			1	1	1
Drama	1	1	1	1	1
Romance	1		1		
Thriller		1	1	1	1

$$\hat{r}(Lucía, Killing Eve) =$$

$$0.5 \cdot 3 + 0.5 \cdot 2 + 0.2 \cdot 4 = \mathbf{3.3}$$

$$Jaccard(Killing Eve, Grey's \dots) = 2/4 = \mathbf{0.5}$$

$$Jaccard(Killing Eve, Walking \dots) = 2/4 = \mathbf{0.5}$$

$$Jaccard(Killing Eve, How I \dots) = 1/5 = \mathbf{0.2}$$

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I...	Comedy	Crime	Drama	Romance	Thriller
Claudia	5	2		4		4					
Ignacio	3			5		4					
Lucía	3	2	?			4	4	0	5	7	2
Pablo		2		3	2	3					
Sergio	2	4		4	3	2					

↑ Lucía

					1
		1	1	1	
Drama	1	1	1	1	1
Romance	1		1		1
Thriller		1	1	1	

2. Métodos basados
en contenido:
centroides de usuario

$$\text{P.e.} \text{Lucía}[Drama] = 3 \cdot 1 + 2 \cdot 1 + 4 \cdot 0 = 5$$

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I...	Comedy	Crime	Drama	Romance	Thriller
Claudia	5	2		4		4					
Ignacio	3			5		4					
Lucía	3	2	?			4	4	0	5	7	2
Pablo		2		3	2	3					
Sergio	2	4		4	3	2					



					1	
Comedy						
Crime			1	1	1	
Drama	1	1	1	1	1	
Romance	1		1			1
Thriller		1	1	1	1	

2. Métodos basados en contenido: centroides de usuario

$$\hat{r}(Lucía, Killing Eve) = \cos(Lucía, Killing Eve)$$

$$= \frac{4 \cdot 0 + 0 \cdot 1 + 5 \cdot 1 + 7 \cdot 1 + 2 \cdot 1}{\sqrt{4^2 + 0^2 + 5^2 + 7^2 + 2^2} \sqrt{0^2 + 1^2 + 1^2 + 1^2 + 1^2}} = 0.72$$

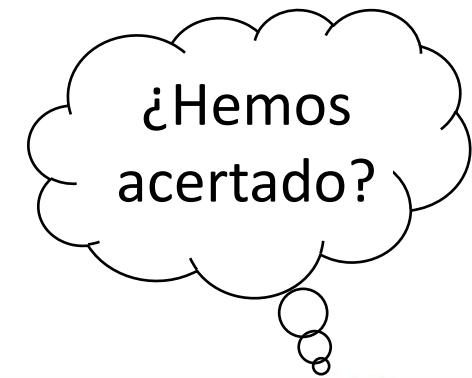
Grey's Anatomy
The Walking Dead
Killing Eve
Peaky Blinders
The Wire
How I met your mother

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4	2.9	4
Ignacio	3	2.91	4.5	5	3.1	4
Lucía	3	2	5	4.03	2.5	4
Pablo	3.18	2	4.12	3	2	3
Sergio	2	4	4.7	4	3	2

Preguntamos a los usuarios

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4	1	4
Ignacio	3	3	3	5	4	4
Lucía	3	2	5	5	2	4
Pablo	4	2	3	3	2	3
Sergio	2	4	2	4	3	2

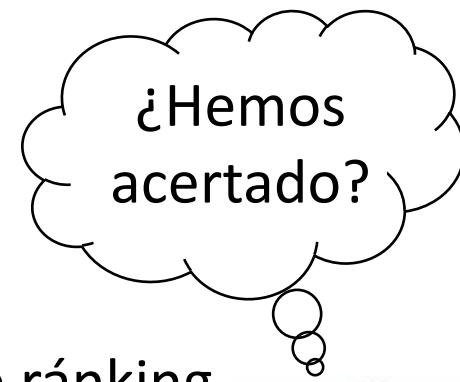
Métricas de error



Grey's Anatomy
The Walking Dead
Killing Eve
Peaky Blinders
The Wire
How I met your mother

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4	2.9	4
Ignacio	3	2.91	4.5	5	3.1	4
Lucía	3	2	5	4.03	2.5	4
Pablo	3.18	2	4.12	3	2	3
Sergio	2	4	4.7	4	3	2

¡Es muy costoso!



Preguntamos a los usuarios

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4	0	4
Ignacio	3	0	0	5	1	4
Lucía	3	2	1	1	0	4
Pablo	1	2	0	3	2	3
Sergio	2	4	0	4	3	2

Métricas de ranking
(por relevancia)



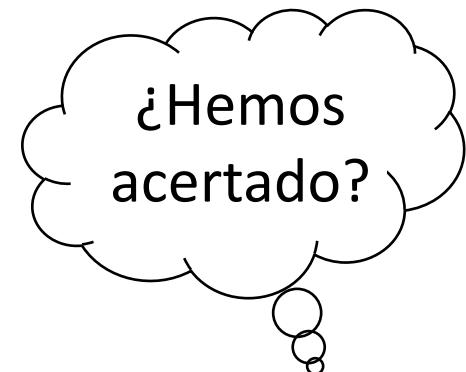
	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2				4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

¿Hemos
acertado?



	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2				4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

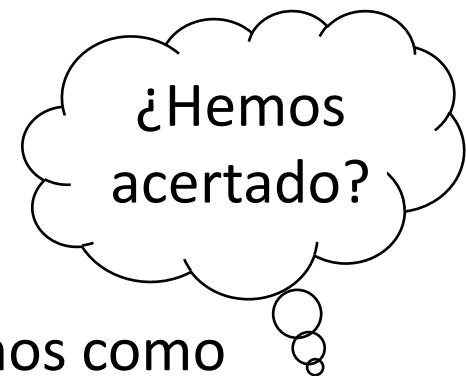
1. Elegimos unas celdas
(p.e. 20% al azar)



Grey's Anatomy
The Walking Dead
Killing Eve
Peaky Blinders
The Wire
How I met your mother

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	?	4		4
Ignacio	3			?		4
Lucía	3	?				4
Pablo		2		?	2	?
Sergio	?	4		4	?	2

1. Elegimos unas celdas (p.e. 20% al azar)
2. Se las ocultamos al algoritmo a evaluar



	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2				4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

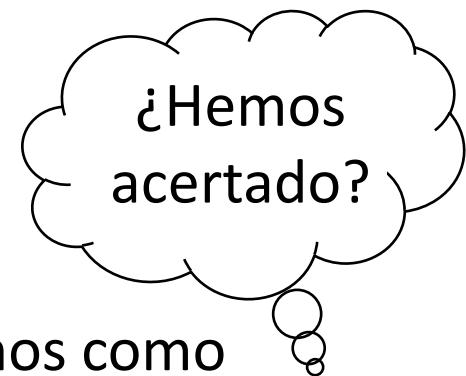
3. Las usamos como ratings de test (igual que antes)



Grey's Anatomy
The Walking Dead
Killing Eve
Peaky Blinders
The Wire
How I met your mother

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	?	4	?	4
Ignacio	3	?	?	?	?	4
Lucía	3	?	?	?	?	4
Pablo	?	2	?	?	2	?
Sergio	?	4	?	4	?	2

1. Elegimos unas celdas (p.e. 20% al azar)
2. Se las ocultamos al algoritmo a evaluar



	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2				4
Pablo		2		3	2	3
Sergio	2	4		4	3	2

3. Las usamos como ratings de test (igual que antes)



Grey's Anatomy
The Walking Dead
Killing Eve
Peaky Blinders
The Wire
How I met your mother

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	?	4	?	4
Ignacio	3	?	?	?	?	4
Lucía	3	?	?	?	?	4
Pablo	?	2	?	?	2	?
Sergio	?	4	?	4	?	2

1. Elegimos unas celdas (p.e. 20% al azar)
2. Se las ocultamos al algoritmo a evaluar

Datos de test
“gratis”

3. Las usamos como ratings de test (igual que antes)

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother
Claudia	5	2	5	4		4
Ignacio	3			5		4
Lucía	3	2				4
Pablo		2		3	2	3
Sergio	2	4		4	3	2



Partición entrenamiento/test

para evaluación offline

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	Hov
Claudia	5	2		4		4
Ignacio	3					4
Lucía	3					4
Pablo		2			2	
Sergio		4		4		2

“Entrenamiento”

			5			
Claudia						
Ignacio						
Lucía		2				
Pablo				3		3
Sergio	2				3	

“Test”

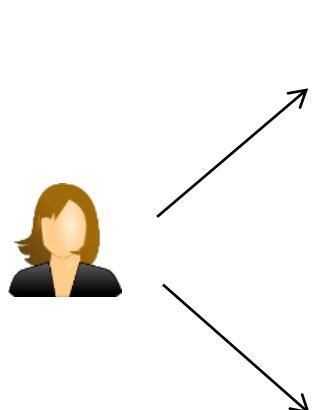
Datos de test
“gratis”



Algoritmos en detalle: fórmulas y variantes

kNN colaborativo

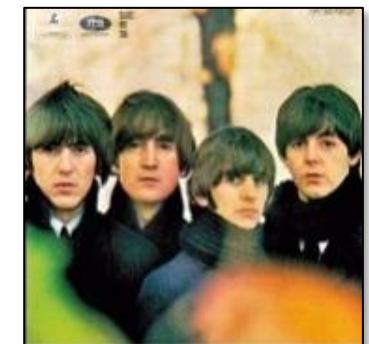
You like



They like



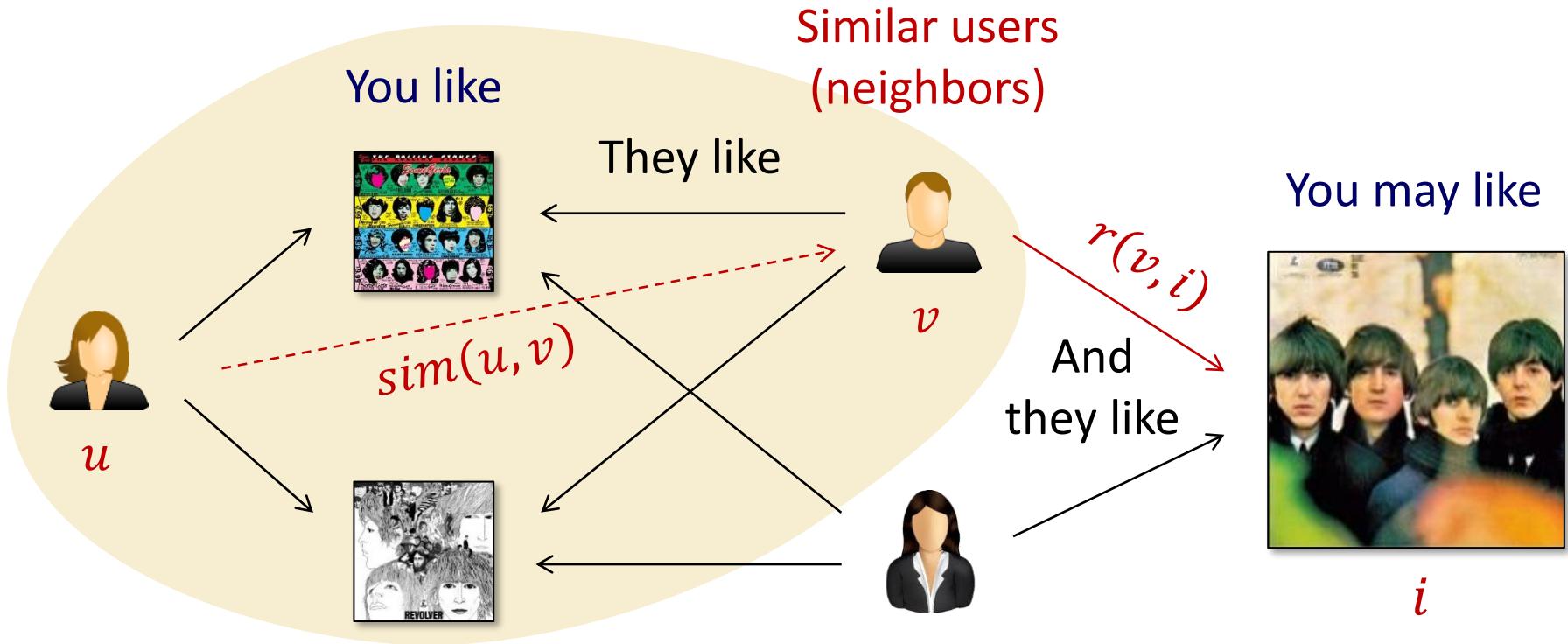
You may like



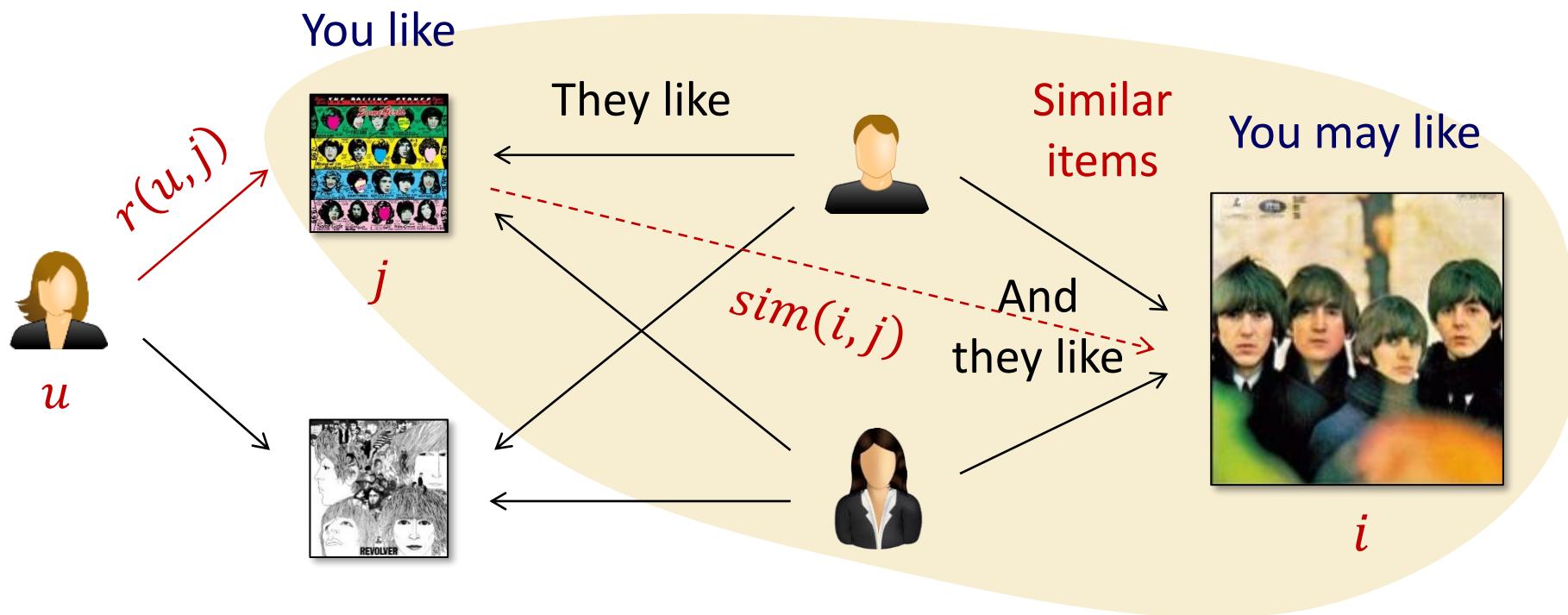
And
they like



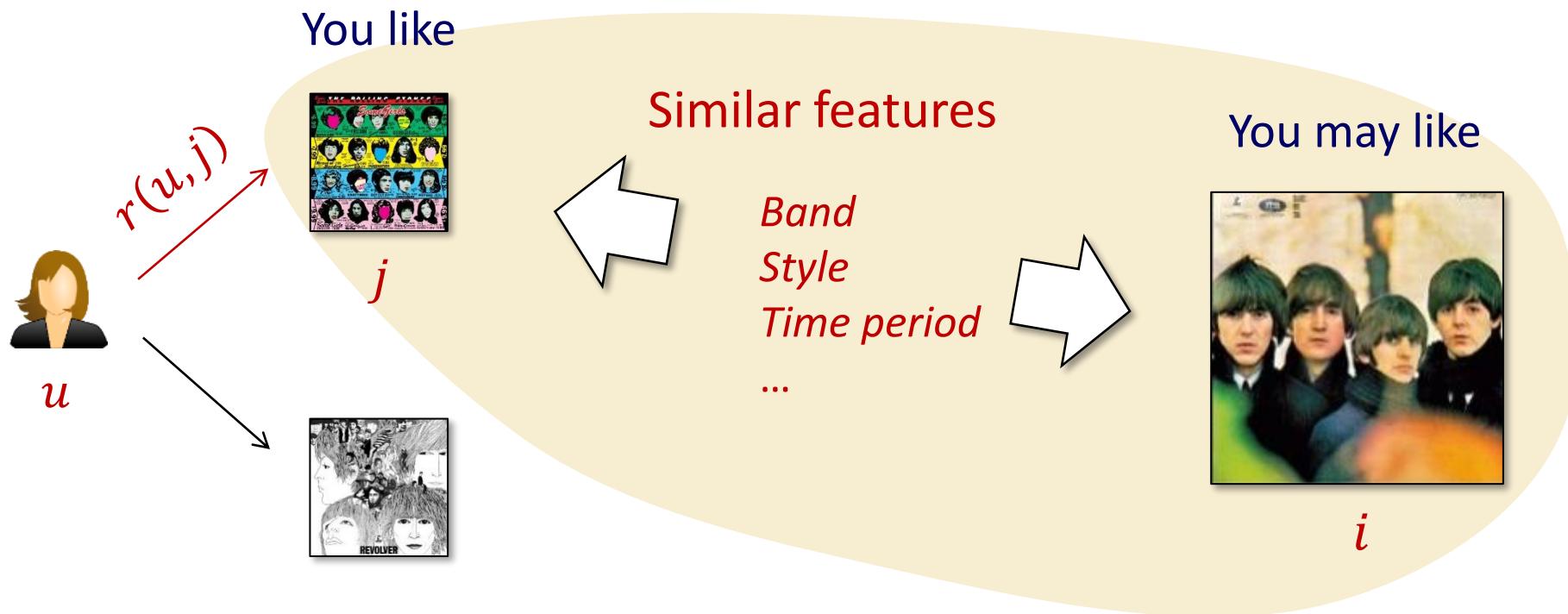
kNN colaborativo: basado en usuario



kNN colaborativo: basado en ítem



kNN basado en contenido



Predicción de ratings kNN

- Basado en usuario:

$$\hat{r}(u, i) = C \sum_{\substack{v \in N_k(u) \\ r(v, i) \neq \emptyset}} sim(u, v) r(v, i)$$

Desconocido
(a predecir)

- Basado en ítem:

$$\hat{r}(u, i) = C \sum_{r(u, j) \neq \emptyset} sim(i, j) r(u, j)$$

Vecindario $N_k(u) \equiv$ top k vecinos más similares a u

↑
Conocido

$$C = 1 \left/ \sum_{\substack{v \in N_k(u) \\ r(v, i) \neq \emptyset}} |sim(u, v)| \right.$$

↑
Necesario para
ratings en rango

$$C = 1 \left/ \sum_{r(u, j) \neq \emptyset} |sim(i, j)| \right.$$

Similitud entre usuarios

◆ Coseno

$$sim(u, v) = \frac{\sum_{\substack{i:r(u,i)\neq\emptyset \\ r(v,i)\neq\emptyset}} r(u, i)r(v, i)}{\sqrt{\sum_{i:r(u,i)\neq\emptyset} r(u, i)^2 \sum_{i:r(v,i)\neq\emptyset} r(v, i)^2}} \in [0,1]$$

◆ Correlación de Pearson

$$sim(u, v) =$$

$$\frac{\sum_{\substack{i:r(u,i)\neq\emptyset \\ r(v,i)\neq\emptyset}} (r(u, i) - \bar{r}_u)(r(v, i) - \bar{r}_v)}{\sqrt{\sum_{\substack{i:r(u,i)\neq\emptyset \\ r(v,i)\neq\emptyset}} (r(u, i) - \bar{r}_u)^2 \sum_{\substack{i:r(v,i)\neq\emptyset \\ r(u,i)\neq\emptyset}} (r(v, i) - \bar{r}_v)^2}}$$

↑
Puntuación promedio de u

↑
Ojo al detalle

$\in [-1,1]$

Por qué Pearson: sesgos de puntuación

	Grey's Anatomy	The Walking Dead	Killing Eve	Peaky Blinders	The Wire	How I met your mother	
Claudia	5	2	5	4		4	4
Ignacio	3			5		4	4
Lucía	3	2				4	3
Pablo		2		3	2	3	2.5
Sergio	2	4		4	3	2	3

$$\text{P.e. } \bar{r}_{Sergio} = \frac{2 + 4 + 4 + 3 + 2}{5} = \mathbf{3}$$

En rigor la correlación de Pearson toma la media restringida al solapamiento pero eso funciona peor

Similitud entre usuarios

◆ Coseno

$$sim(u, v) = \frac{\sum_{\substack{i:r(u,i)\neq\emptyset \\ r(v,i)\neq\emptyset}} r(u, i)r(v, i)}{\sqrt{\sum_{i:r(u,i)\neq\emptyset} r(u, i)^2 \sum_{i:r(v,i)\neq\emptyset} r(v, i)^2}} \in [0,1]$$

◆ Correlación de Pearson

$$sim(u, v) =$$

$$\frac{\sum_{\substack{i:r(u,i)\neq\emptyset \\ r(v,i)\neq\emptyset}} (r(u, i) - \bar{r}_u)(r(v, i) - \bar{r}_v)}{\sqrt{\sum_{\substack{i:r(u,i)\neq\emptyset \\ r(v,i)\neq\emptyset}} (r(u, i) - \bar{r}_u)^2 \sum_{\substack{i:r(u,i)\neq\emptyset \\ r(v,i)\neq\emptyset}} (r(v, i) - \bar{r}_v)^2}}$$

$$\in [-1,1]$$

- Nótese que Pearson puede ser negativo (es raro) y divisiones por cero
- Ojo con solapamientos muy pequeños, pueden dar sim falsamente alta

- ◆ Similitudes basadas en datos personales (p.e. demográficos)
- ◆ Y muchos más métodos..

Predictión de ratings kNN: variante centrada en la media

- ◆ Para compensar sesgos en la forma de puntuar de los usuarios, en la versión basada en usuario:

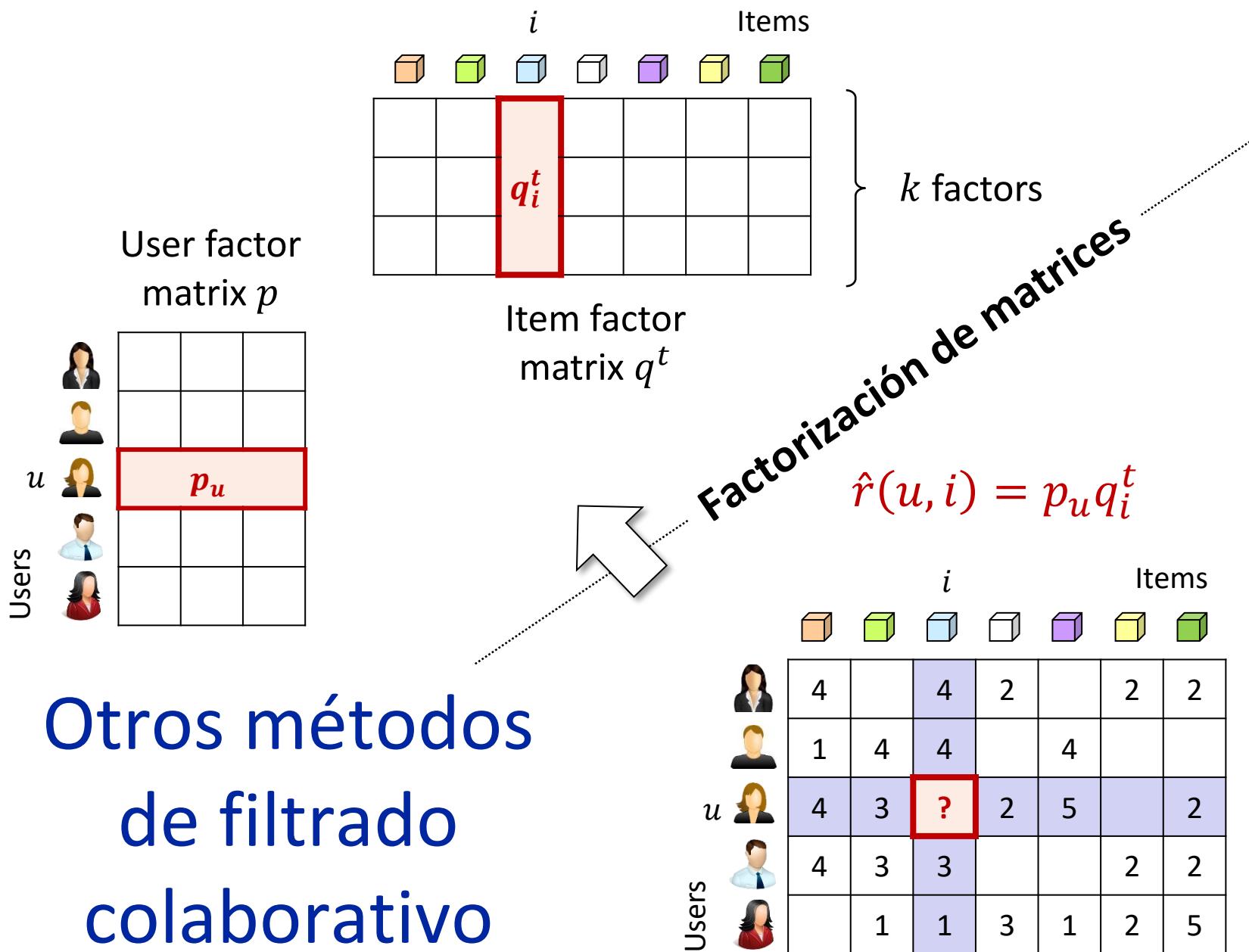
$$\hat{r}(u, i) = \bar{r}_u + c \sum_{\substack{v \in N_k(u) \\ r(v, i) \neq \emptyset}} sim(u, v)(r(v, i) - \bar{r}_v)$$

Puntuación promedio del usuario u ↗ Puntuación promedio del vecino v ↗

(obsérvese que con esta fórmula podrían salir ratings fuera de rango)

Similitud entre ítems

- ◆ Mismas funciones que para usuarios: Pearson, coseno...
 - Intercambiando simétricamente el papel de ítems y usuarios
- ◆ Se suele contemplar alguna variante más
 - P.e. coseno centrado en la media del usuario
- ◆ Si se utiliza similitud basada en contenido ya no es filtrado colaborativo, sino kNN basado en contenido
 - Por lo demás la estructura de kNN basado en ítem colaborativo y basado en contenido es la misma



Otros métodos de filtrado colaborativo

- ◆ Factorización de matrices
 - Obtener representación de usuarios e ítems u_f, i_f en un espacio de factores latentes $f \in \mathcal{F}$
 - Minimizando $\sum_{(u,i) \in training} \left[(r(u,i) - \sum_f u_f i_f)^2 + \lambda (\sum_f u_f^2 + \sum_f i_f^2) \right]$
(p.e. descenso por gradiente, mínimos cuadrados, etc.)
 - Finalmente $\hat{r}(u, i) = \sum_f u_f i_f$
- ◆ Clustering para kNN
- ◆ Random walks
 - En el grafo bipartito de usuarios + ítems, donde los ratings se interpretan como arcos
- ◆ Redes neuronales
- ◆ ...y muchos más

Punto débil de FC: data sparsity

- ◆ Usuarios e ítems que no son comparables entre sí
 - No tienen solapamiento, o muy pequeño (similitud poco fiable)
→ Exigir solapamiento p.e. 2-3 mínimo
 - Algun usuario (ítem) se queda sin vecinos \Rightarrow sin recomendaciones
- ◆ Los vecinos no siempre se pueden usar
 - Tienen que haber puntuado el ítem cuyo rating queremos predecir
Salvo que elijamos el vecindario distinto para cada ítem (asegurando k vecinos para cada ítem), pero es costoso y poco habitual seguir esta opción
- ◆ Estos problemas dominan en ítems nuevos / raros y usuarios nuevos / poco activos (arranque en frío)



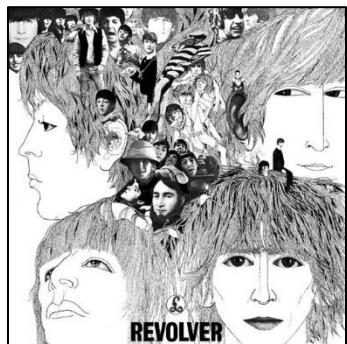
Valoración de filtrado colaborativo

- ◆ Puede funcionar con ítems totalmente opacos
- ◆ Buenos niveles de acierto
- ◆ Diversidad y novedad
 - Generalmente mayor que la recomendación basada en contenido
- ◆ Punto débil: arranque en frío, ítems poco conocidos, sparsity
 - Recomendación basada en contenido para ítems nuevos
 - Información estática del usuario para usuarios poco activos
 - Datos personales, cuestionarios, etc.
 - Para definir la similitud entre usuarios en filtrado colaborativo
 - Como características en clasificación o regresión basada en contenido

Recomendación basada en contenido

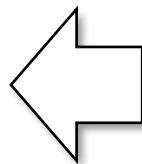
La intuición:

You like

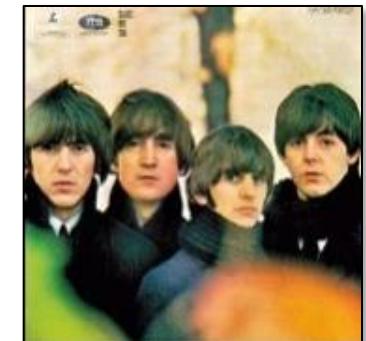
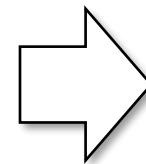


You may like

Similar features

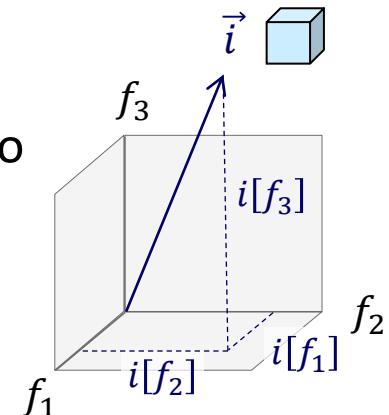


*Band
Style
Time period
...*

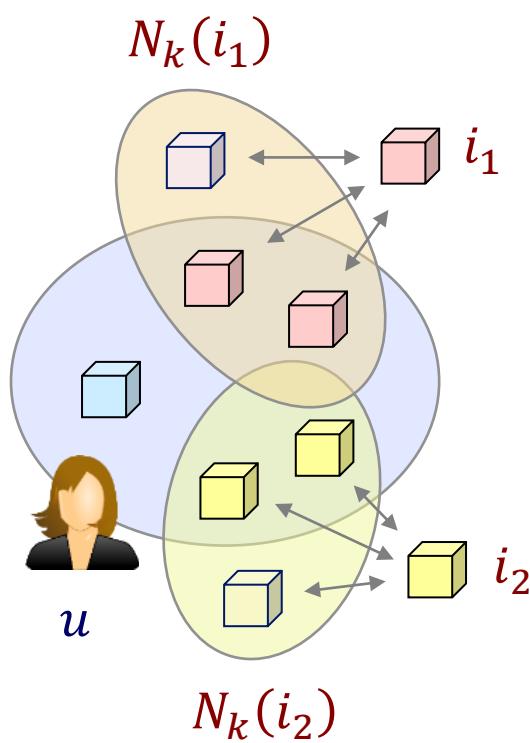


Recomendación basada en contenido

- ◆ Se recomienda a cada usuario **sin mirar a los demás**
- ◆ Se necesita un espacio de características de los ítems
 - “Datos” que describen los ítems, estructurados o no estructurados
 - P.e. metadatos del ítem (autor, lugar, idioma, categorías, etiquetas), palabras del texto asociado al ítem, etc.
 - Valores binarios, enteros, o reales
 - Los ítems se representan como vectores en este espacio
- ◆ Y una función de similitud en ese espacio
 - Por ejemplo, el coseno para características numéricas
 - O Jaccard para características binarias
- ◆ Dos métodos muy comunes: kNN y basado en centroides
 - Pero se pueden utilizar muchos otros basados en clasificación (donde los usuarios juegan en esencia el papel de clase)



Recom. basada en contenido: kNN



$$\hat{r}(u, i) = C \sum_{\substack{j \in N_k(i) \\ r(u, j) \neq \emptyset}} sim(i, j) r(u, j)$$

Función de ránking (y predicción de rating)

$$C = 1 / \sum_{\substack{j \in N_k(i) \\ r(u, j) \neq \emptyset}} |sim(i, j)|$$

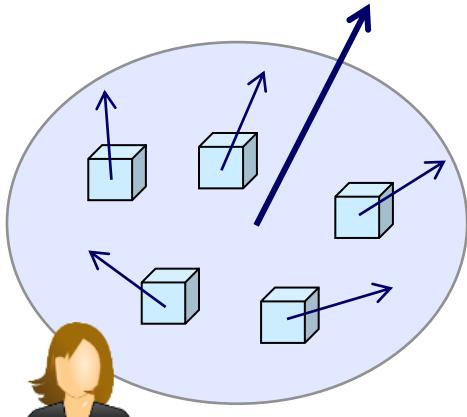
Opcional: para predecir ratings en rango

Adaptación de clasificación kNN

- ◆ En clasificación sería $r(u, j)$ binario
- ◆ Y además hacemos rankings de “instancias” (ítems) para cada “clase” (usuario) en lugar de lo contrario
- ◆ También se puede entender (es más natural) como regresión kNN

Recom. basada en contenido: centroides

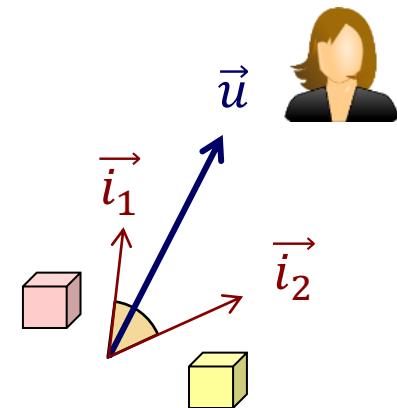
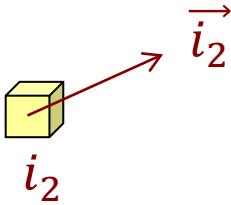
Vectores: características numéricas



Centroide

$$C = 1/|u| \text{ (opcional, p.e. para distancia euclídea)}$$

$$\vec{u} = C \sum_{j:r(u,j) \neq \emptyset} r(u,j) \cdot \vec{j}$$



$$sim(u, i) = \cos(\vec{u}, \vec{i})$$

Adaptación de clasificación Rocchio

- ◆ En clasificación sería $r(u, j)$ binario
- ◆ Normalización por C (para coseno no hace falta)
- ◆ Y además hacemos rankings de “instancias” (ítems) para cada “clase” (usuario) en lugar de lo contrario

Función de ránking

En resumen...

◆ Rocchio

$$u = C \sum_{j:r(u,j) \neq \emptyset} r(u,j) \cdot j$$

Función de ránking

Opcional p.e. para distancia euclídea

$C = 1/|u|$

◆ kNN

$$\hat{r}(u, i) = C \sum_{\substack{j \in N_k(i) \\ r(u,j) \neq \emptyset}} sim(i, j) |r(u, j)|$$

Generalización: originalmente
sería $r(u, j)$ binario

Opcional: para predecir ratings en rango

$C = 1 / \sum_{\substack{j \in N_k(i) \\ r(u,j) \neq \emptyset}} |sim(i, j)|$

Función de ránking (y predicción de rating)

Recomendación basada en contenido: valoración

- ◆ Capaz de recomendar ítems que no tienen ratings
 - Permite recomendar ítems nuevos y/o poco conocidos
 - No sirve en general para usuarios nuevos y/o poco activos (salvo NB o regresión usando otras características, datos o input de los usuarios)
- ◆ Aunque tiene un efecto de “encasillamiento” del usuario
 - Lo que se recomienda se parece a lo que el usuario ya ha visto
- ◆ Es un buen complemento de los métodos colaborativos en soluciones híbridas

Evaluación

- ◆ Metodología
 - Como en clasificación, separar ratings en entrenamiento y test (p.e. 80-20%)
 - Predecir ratings de test usando training y medir el nivel de acierto
 - Si el split es aleatorio, repetir n -fold (p.e. 5-fold para 80-20%)
 - También se utiliza split temporal, más realista respecto a un escenario natural
- ◆ Las métricas clásicas miden el **error de predicción**

$$\left. \begin{aligned} MAE &= \frac{1}{|test|} \sum_{(u,i) \in test} |\hat{r}(u, i) - r(u, i)| \\ RMSE &= \sqrt{\frac{1}{|test|} \sum_{(u,i) \in test} (\hat{r}(u, i) - r(u, i))^2} \end{aligned} \right\} 0 \leq MAE \leq RMSE \leq \max(r)$$

Evaluación

- ◆ El error de predicción no es la mejor métrica (evalúa regresión)
 - Correlaciona poco con la satisfacción del usuario en la práctica
 - Además, no se puede calcular el error para los métodos que no predicen ratings
- ◆ También se utilizan la curva ROC y el área bajo la curva (AUC) (evalúan clasificación)
- ◆ En general tienen más sentido **métricas de ranking**: todas las de IR ($P@k$, nDCG, MAP, etc.)
 - Tomando los usuarios como equivalentes a consultas
 - Usando los ratings de test como juicios de relevancia
- ◆ Métricas para otras dimensiones, p.e. novedad y diversidad
- ◆ En la industria: **tests A/B**

Ejemplo

	a	b	c	d	e	f	g	h
u_1 	4	5			5	2		4
u_2 			1	1		4	5	4

Ejemplo

Score del recomendador Rating de test

		$\hat{r}(u_1, i)$	$r(u_1, i)$
u_1			
a	5	4	
b	5	5	
c	4		
d	4		
e	4	5	

		$\hat{r}(u_1, i)$	$r(u_1, i)$
u_2			
a	5		
b	5		
c	5		
d	5	4	1
e	5		
f	5	4	
g	4		
h			

$$P@1 = \frac{1/1 + 0/1}{2}$$

$$P@2 = \frac{2/2 + 0/2}{2}$$

$$P@3 = \frac{2/3 + 1/3}{2}$$

etc.

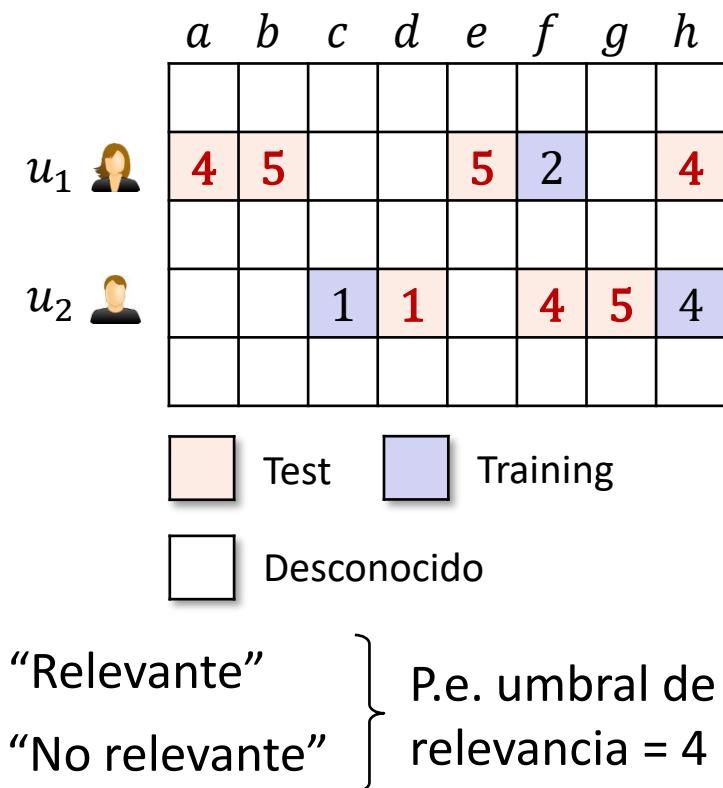
$$\text{Recall}@1 = \frac{1/4 + 0/2}{2}$$

$$\text{Recall}@2 = \frac{2/4 + 0/2}{2}$$

$$\text{Recall}@3 = \frac{2/4 + 1/2}{2}$$

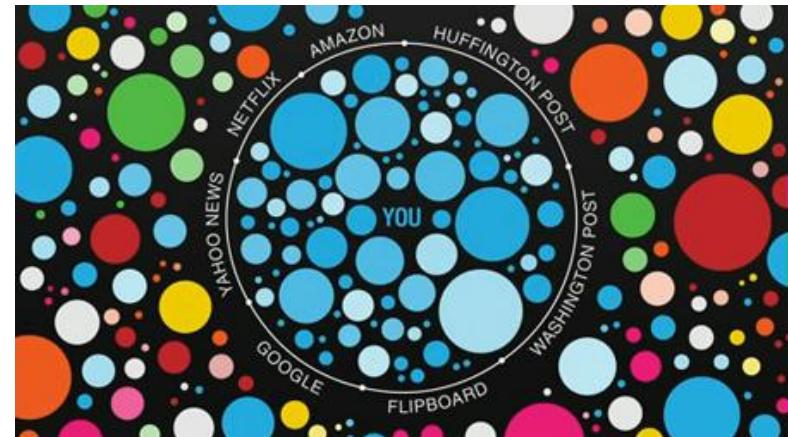
etc.

$$\text{RMSE} = \sqrt{[(5 - 4)^2 + (5 - 5)^2 + (4 - 5)^2 + (5 - 4)^2 + (5 - 1)^2]/5}$$



Precauciones con la personalización

- ◆ Dificultad intrínseca de la tarea → realismo, prudencia
- ◆ Privacidad → transparencia y respeto al usuario
- ◆ Intrusividad → evitar la pérdida de control para el usuario
- ◆ Utilidad vs. reacción del usuario
 - P.e. satisfacción vs. compulsividad
- ◆ El efecto “filter bubble”
 - Empobrecimiento, polarización
 - Diversidad y novedad



https://en.wikipedia.org/wiki/Filter_bubble

- ◆ Sesgos, discriminación
 - Problema similar en aprendizaje automático (fairness)