

# **Ranking & UB-CF**

## **IIC3633 - Sistemas Recomendadores**

Denis Parra

# TOC

## En esta clase

1. Actividades principales del semestre: Blog, Tarea 1, Presentacion, Proyecto Final
2. Definición y un poco de Historia
3. Ranking No Personalizado
4. User-Based Collaborative Filtering
5. Referencias

# Definición

Recommender Systems aim to help a user or a group of users in a system to select items from a crowded item or information space.  
(MacNee et. al 2006)

R. Burke tenía su propia definición, similar a esta, pero agregaba ...in a personalized way.



# ¿Por qué nos interesan los RecSys en estos días?

- Los Sistemas Recomendadores (RecSys) han ganado mucha popularidad en varios dominios y aplicaciones donde la gente debe tomar decisiones sobre una gran cantidad de información.



## El problema de recomendación formalizado (Adomavicius et al. 2007)

$$\forall c \in C, s'_c = \operatorname{argmax}_{s \in S} u(c, s)$$

$u : C \times S \rightarrow R$ , *funcion de utilidad*

$R$  : *conjunto recomendado de items*

$C$  : *conjunto de usuarios*

$S$  : *conjunto de items*

# 1. Un Poco de Historia



Recomendaciones estilo Amazon.com

## 1.1 En 1992 Xerox PARC Tapestry

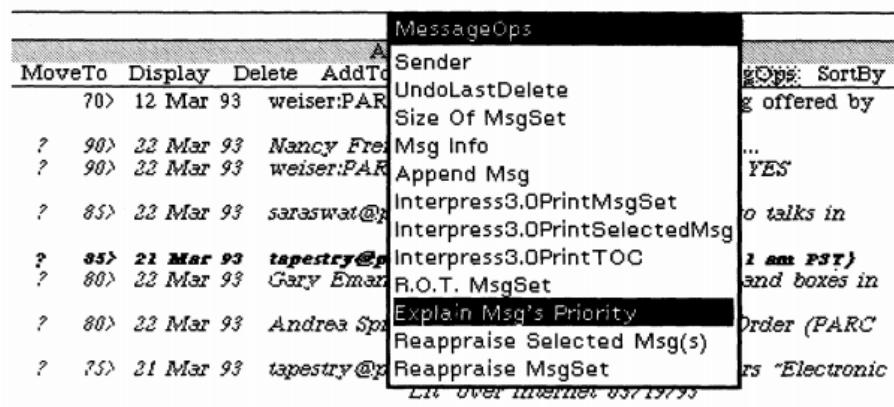


Figure 2. Requesting an explanation for a message's priority.

```
Annotations for message $ XNS-SMTP-Gateway:Parc:Xerox
appraiser terry$text:Bakersfield => priority 85
appraiser terry$Subject:Briefs<California => priority 55
appraiser terry$sender:tapestry => priority 10
```

Figure 3. An explanation of priorities assigned to a message by various appraisers.

am curious as to why this particular message was assigned priority 85. So I select the message by left-clicking on its summary, and then I click the "MsgOps" button at the top of the window. This produces a pop-up menu from which I select the "Explain Msg's Priority" option (see Figure 2.). The resulting textual explanation is shown in Figure 3.

Link to [PDF file](#)

## 1.2 MovieLens

Prediction or Rating ↴	Your Rating	Movie Information	Wish List
★★★★★	5.0 stars	<a href="#">Pink Floyd: The Wall (1982)</a> DVD info imdb flag Movie Tuner Drama, Musical  [add tag] Popular tags: social commentary   cult film   surreal	<input type="checkbox"/>
★★★★★	4.5 stars	<a href="#">Lives of Others, The (Das leben der Anderen) (2006)</a> DVD info imdb flag Movie Tuner Drama, Romance, Thriller - German  [add tag] Popular tags: disturbing   psychology   romance	<input type="checkbox"/>
★★★★★	Not seen	<a href="#">Shawshank Redemption, The (1994)</a> DVD info imdb flag Movie Tuner Crime, Drama  [add tag] Popular tags: based on a book   psychology   twist ending	<input type="checkbox"/>
★★★★★	Not seen	<a href="#">Godfather, The (1972)</a> DVD info imdb flag Movie Tuner Crime, Drama - English, Italian  [add tag] Popular tags: organized crime   based on a book   Oscar (Best Picture)	<input type="checkbox"/>
★★★★★	Not seen	<a href="#">One Flew Over the Cuckoo's Nest (1975)</a> DVD info imdb flag Movie Tuner Drama  [add tag] Popular tags: psychology   based on a book   Oscar (Best Picture)	<input type="checkbox"/>
★★★★★	Not seen	<a href="#">Usual Suspects, The (1995)</a> DVD info imdb flag Movie Tuner Crime, Mystery, Thriller - English, Hungarian, Spanish, French  [add tag] Popular tags: organized crime   twist ending   Kevin Spacey	<input type="checkbox"/>

Link to [Amatriain 2012](#)

# NetFlix Prize (2007-2009)

**Netflix Prize** COMPLETED

Home Rules Leaderboard Update

## Leaderboard

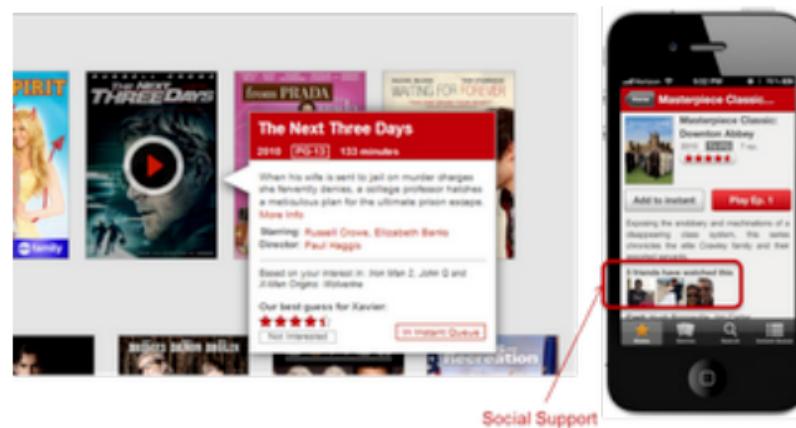
Showing Test Score. [Click here to show quiz score](#)

Display top  leaders.

Rank	Team Name	Best Test Score	% Improvement	Best Submit Time
<b>Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos</b>				
1	<a href="#">BellKor's Pragmatic Chaos</a>	0.8567	10.06	2009-07-26 18:18:28
2	<a href="#">The Ensemble</a>	0.8567	10.06	2009-07-26 18:38:22
3	<a href="#">Grand Prize Team</a>	0.8582	9.90	2009-07-10 21:24:40
4	<a href="#">Opera Solutions and Vandelay United</a>	0.8588	9.84	2009-07-10 01:12:31
5	<a href="#">Vandelay Industries!</a>	0.8591	9.81	2009-07-10 03:32:20
6	<a href="#">PragmaticTheory</a>	0.8594	9.77	2009-06-24 12:06:56
7	<a href="#">BellKor in BigChaos</a>	0.8601	9.70	2009-05-13 08:14:09
8	<a href="#">Dace</a>	0.8612	9.59	2009-07-24 17:18:43
9	<a href="#">Feeds2</a>	0.8622	9.48	2009-07-12 13:11:51
10	<a href="#">BigChaos</a>	0.8623	9.47	2009-04-07 12:33:59
11	<a href="#">Opera Solutions</a>	0.8623	9.47	2009-07-24 00:34:07
12	<a href="#">Bellkor</a>	0.8624	9.46	2009-07-26 17:19:11

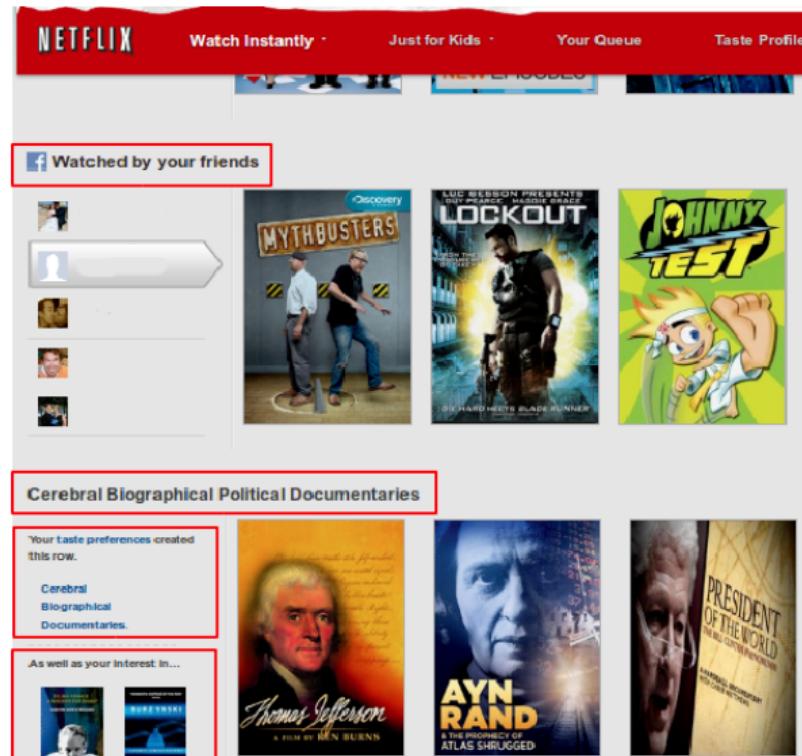
Rank	Team Name	Best Test Score	% Improvement	Best Submit Time
<b>Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos</b>				
1	<a href="#">BellKor's Pragmatic Chaos</a>	0.8567	10.06	2009-07-26 18:18:28
2	<a href="#">The Ensemble</a>	0.8567	10.06	2009-07-26 18:38:22
3	<a href="#">Grand Prize Team</a>	0.8582	9.90	2009-07-10 21:24:40

## 1.3 Netflix en 2012



Link to [Amatriain 2012](#)

## 1.3 Netflix en 2012 (continuación)



Link to [Amatriain 2012](#)

# Ranking no personalizado (Blog de Evan Miller, 2009)

1. Popularidad.
2. Score: (Ratings Positivos) - (Ratings Negativos)
3. Score: (Rating Promedio) = (Ratings Positivos)/(Total de Ratings)
4. Score: **Considerando Ratings positivos y negativos**, Limite inferior del Intervalo de Confianza del Wilson Score, para un parámetro Bernoulli.

$$\left( \hat{p} + \frac{z_{\alpha/2}^2}{2n} \pm z_{\alpha/2} \sqrt{\frac{[\hat{p}(1 - \hat{p}) + z_{\alpha/2}^2/4n]}{n}} \right) / (1 + z_{\alpha/2}^2/n).$$

Donde  $\hat{p}$  es la proporción (estimada) de ratings positivos,  $z_{\alpha/2}$  es el  $(1 - \alpha/2)$  cuantil de la distribución normal, y  $n$  el número de ratings.  $\alpha$ , también llamado nivel de significancia estadístico, generalmente se considera 95%.

# Clasificación(es)

1. Considerando los Datos usados
  1. Basado en Reglas (Rule-based)
  2. Basado en Contenido (Content-based)
  3. Filtrado Colaborativo (el usuario y sus vecinos)
2. Considerando el Modelo
  1. Memory-based (KNN)
  2. Model-based (Representación latente)

# Simplificación del Problema: Predicción de Ratings

Predict!

	Item 1	Item 2	...	Item m
User 1	1	5		4
User 2	5	1		?
...				
User n	2	5		?

- How good is my prediction?

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{r}_{ui} - r_{ui})^2}{n}}$$

# Filtrado Colaborativo basado en el usuario

Dos tareas son necesarias:

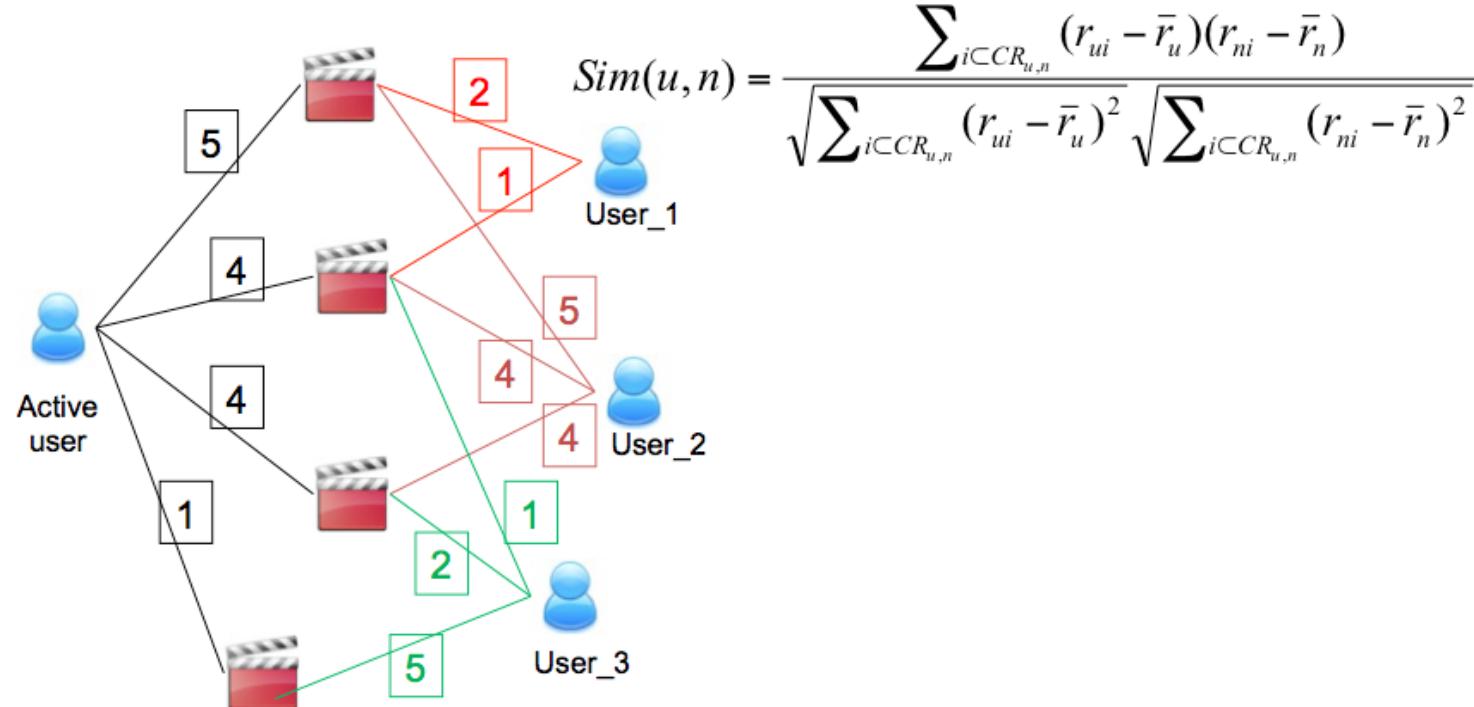
- KNN: Encontrar los K vecinos más cercanos (KNN) al usuario  $a$ :

$$\text{Similaridad}(a, i) = w(a, i), i \in K$$

- Predecir el rating que un usuario  $a$  dará a un ítem  $j$  :

$$p_{a,j} = \bar{v}_a + \alpha \sum_{i=1}^n w(a, i)(v_{i,j} - \bar{v}_i)$$

# Ejemplo: Correlación de Pearson

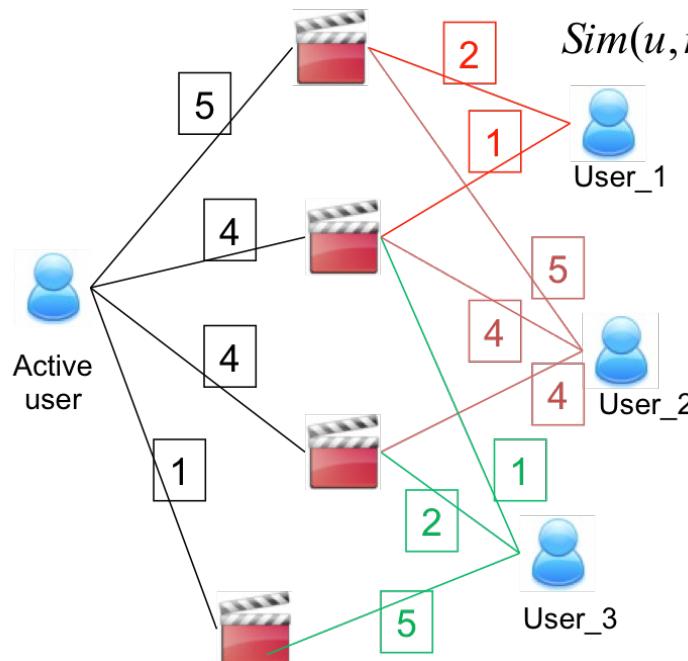


# Ejemplo: Correlación de Pearson

SOLUCION

17/24

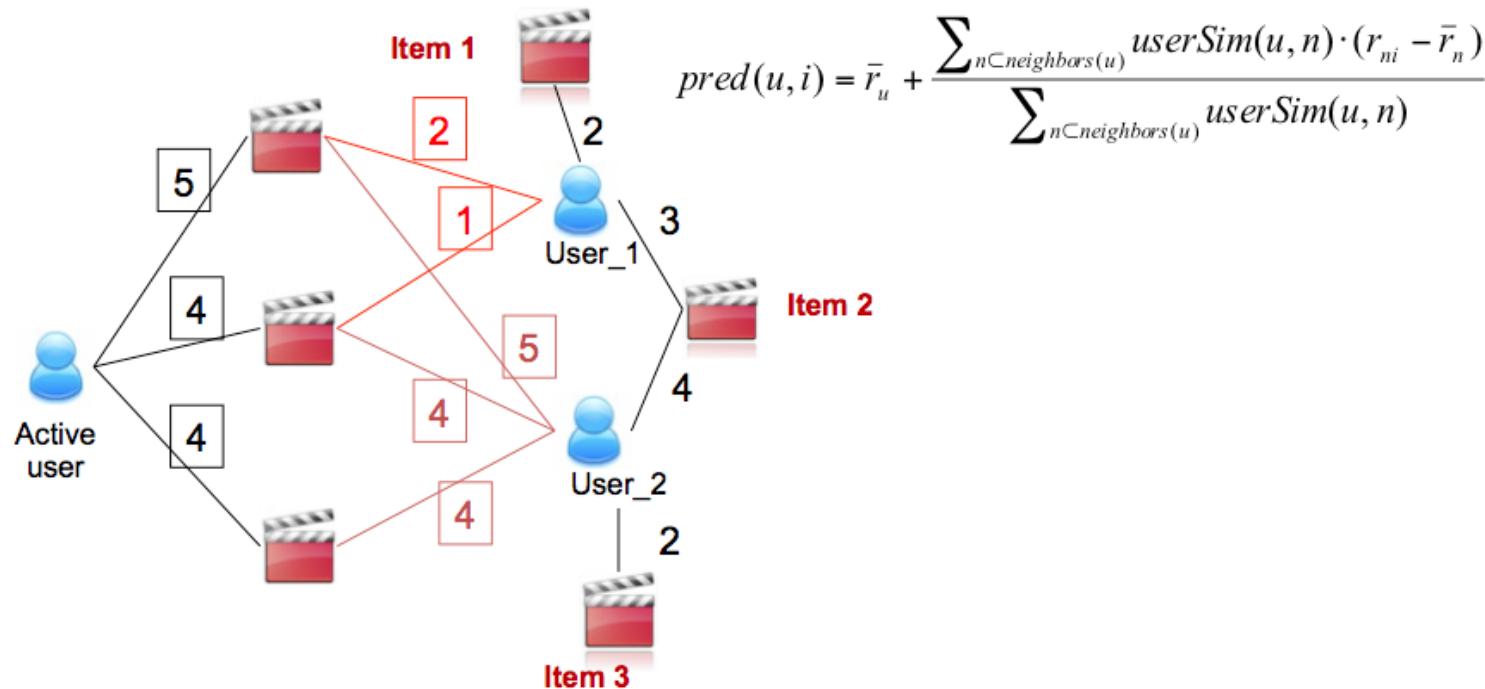
# Ejemplo: Correlación de Pearson



$$Sim(u, n) = \frac{\sum_{i \in CR_{u,n}} (r_{ui} - \bar{r}_u)(r_{ni} - \bar{r}_n)}{\sqrt{\sum_{i \in CR_{u,n}} (r_{ui} - \bar{r}_u)^2} \sqrt{\sum_{i \in CR_{u,n}} (r_{ni} - \bar{r}_n)^2}}$$

	<b>active user</b>
<b>user_1</b>	0.4472136
<b>user_2</b>	0.49236596
<b>user_3</b>	-0.91520863

## Ejemplo Paso 2: Predicción del rating

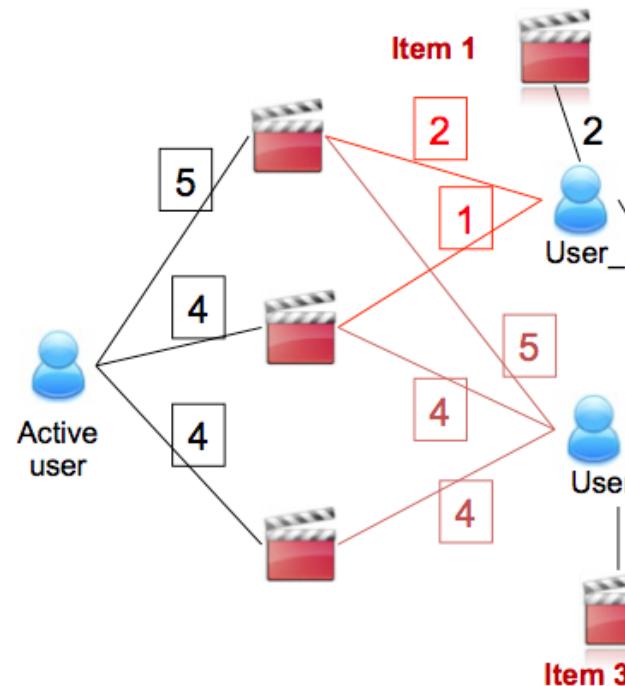


# Ejemplo Paso 2: Predicción del rating

SOLUCION

20/24

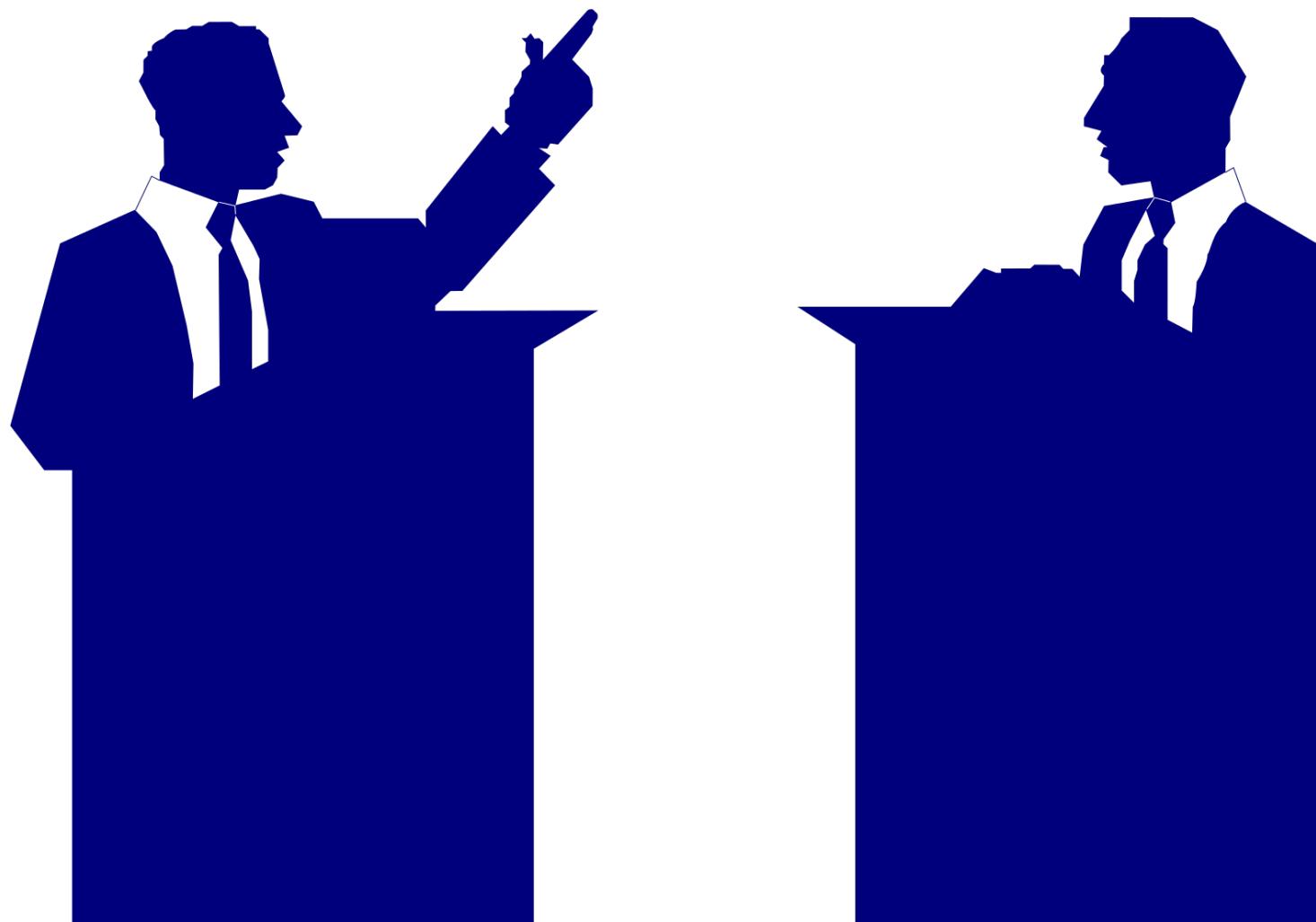
## Ejemplo Paso 2: Predicción del rating



$$pred(u, i) = \bar{r}_u + \frac{\sum_{n \in \text{neighbors}(u)} userSim(u, n) \cdot (r_{ni} - \bar{r}_n)}{\sum_{n \in \text{neighbors}(u)} userSim(u, n)}$$

	User_1	User_2	predicted interest
item_1	2	-	3.262013982
item_2	3	4	3.56331003
item_3	-	2	2.277268083

# ¿Discusión: cuáles son los pro y cons de este método?



22/24

# ¿Discusión: cuáles son los pro y cons de este método?

## PROS:

- Very simple to implement
- Content-agnostic
- Compared to other techniques such as content-based, is more accurate. There is also the Item KNN.

## CONS:

- Sparsity
- Cold-start
- New Item

# Referencias

- Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on*, 17(6), 734-749.
- Amatriain, X. (2013). Mining large streams of user data for personalized recommendations. *ACM SIGKDD Explorations Newsletter*, 14(2), 37-48.
- Miller, B. N., Albert, I., Lam, S. K., Konstan, J. A., & Riedl, J. (2003, January). MovieLens unplugged: experiences with an occasionally connected recommender system. In *Proceedings of the 8th international conference on Intelligent user interfaces* (pp. 263-266). ACM.
- Parra, D., & Sahebi, S. (2013). Recommender systems: Sources of knowledge and evaluation metrics. In *Advanced Techniques in Web Intelligence-2* (pp. 149-175). Springer Berlin Heidelberg.