

Recomendación de contactos en redes sociales mediante métodos de búsqueda

Javier Sanz-Cruzado

Information Retrieval Group

Universidad Autónoma de Madrid

Coautores



Pablo Castells
Universidad
Autónoma de Madrid



Craig Macdonald
University of Glasgow



Iadh Ounis
University of Glasgow

1. Introducción
2. Búsqueda y recomendación
3. Adaptación a la recomendación de contactos
4. Adaptación avanzada
5. Conclusiones

Introducción

Sistemas de recomendación

- ♦ **Objetivo:** Ayudar a los usuarios a descubrir nuevos ítems, a partir de sus interacciones pasadas



- ♦ **Dominios de aplicación**

- Contenido audiovisual:
Netflix, Spotify...
- Comercio electrónico:
Amazon, eBay..
- Publicaciones académicas:
Google Scholar, Mendeley
- **Redes sociales**

Redes sociales on-line

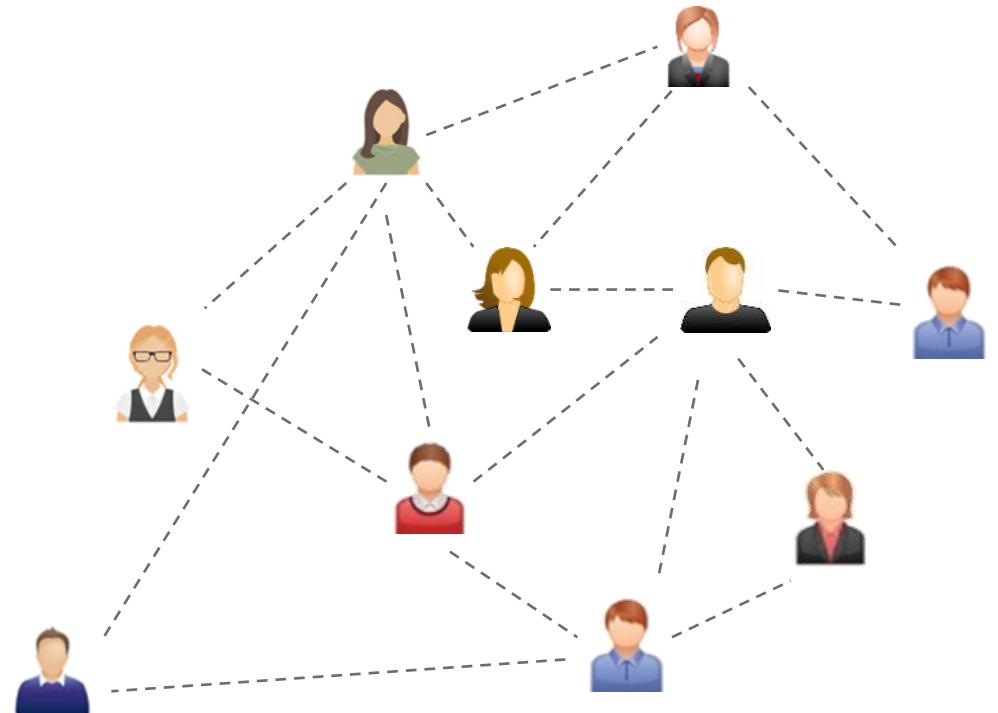


- ◆ Trasladar círculos sociales a la red
- ◆ Comunicación

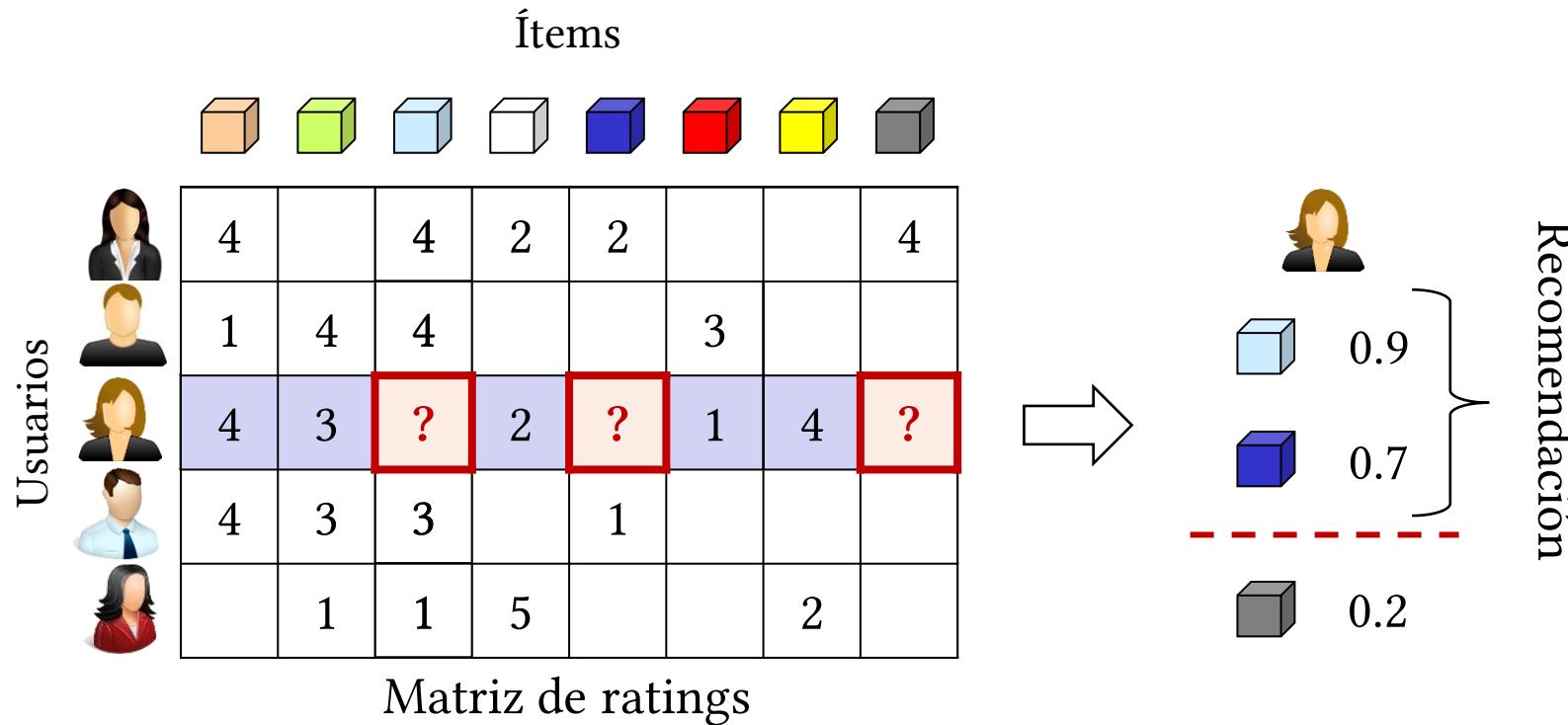
- ◆ Descubrir nuevas personas

- Intereses comunes
- Amistad
- Nuevas perspectivas

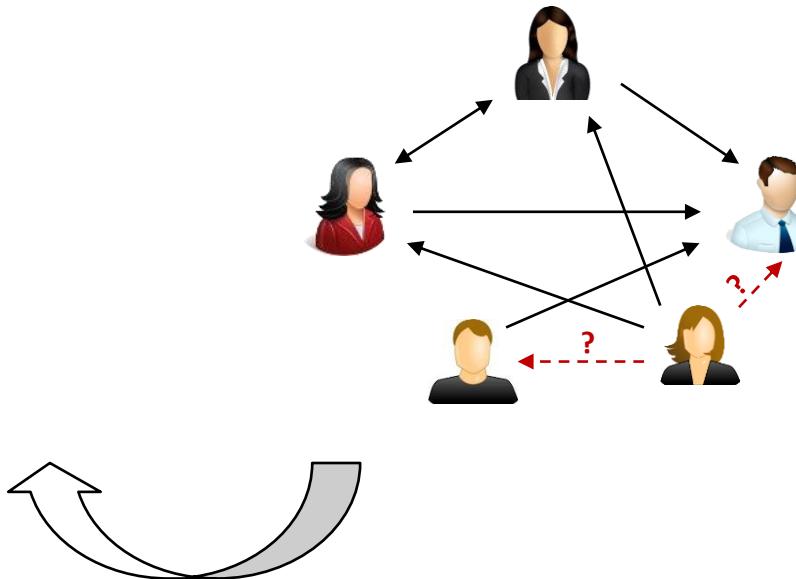
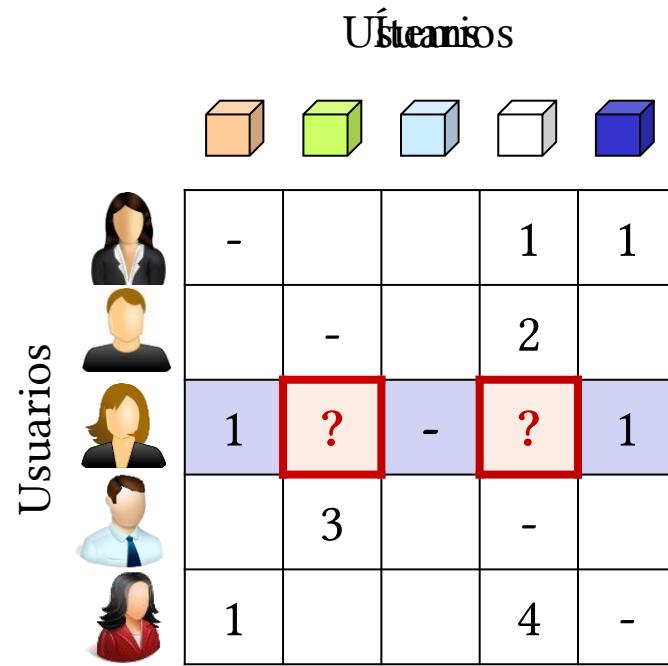
Recomendación de contactos



La tarea de recomendación



Recomendación de contactos



- ◆ Ítems = usuarios
- ◆ Disponibilidad de relaciones sociales
- ◆ Matriz de ratings = matriz de adyacencia

Ejemplos de recomendación de contactos



Who to follow

-  **David Maxwell** 
@maxwelld90 [Follow](#)
-  **Yoan Chabot**
@yoan_cha... Follows you [Follow](#)
-  **Richard McCreadie**
@richardm_ [Follow](#)

Friends

People you may know

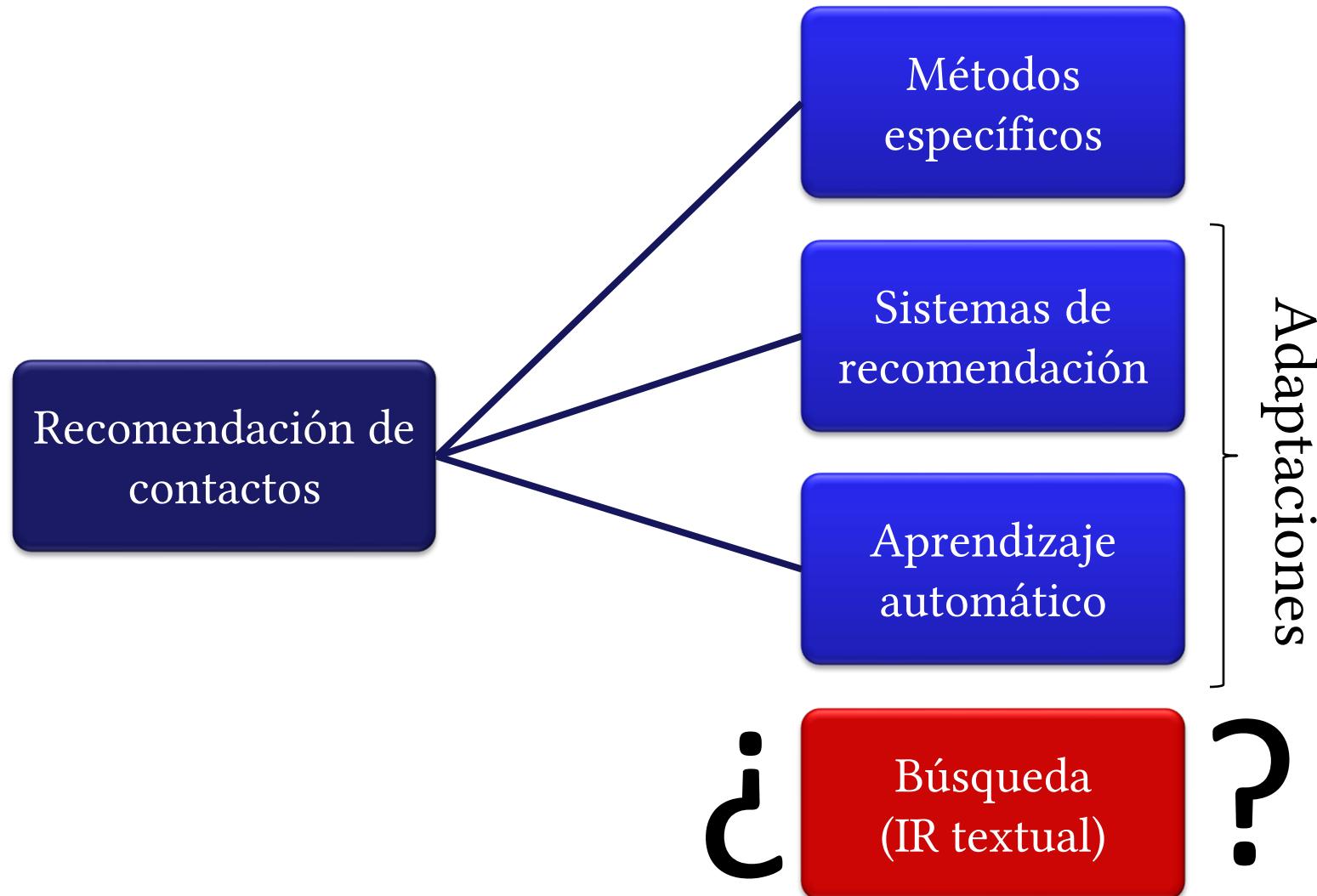
-  **Lorena Gil Monterroso**
2 mutual friends [Add Friend](#) [Remove](#)
-  **Rodri De Blas García**
 16 mutual friends [Add Friend](#) [Remove](#)

¿Por qué la recomendación de contactos?

- ◆ Características particulares
 - Desarrollo de nuevos algoritmos
 - Uso de propiedades de la red social

- ◆ Creación de nuevos enlaces
 - Principal recurso de las redes sociales
 - Canales de comunicación
 - Fuente de información
 - Promueven el uso de la red

Motivación

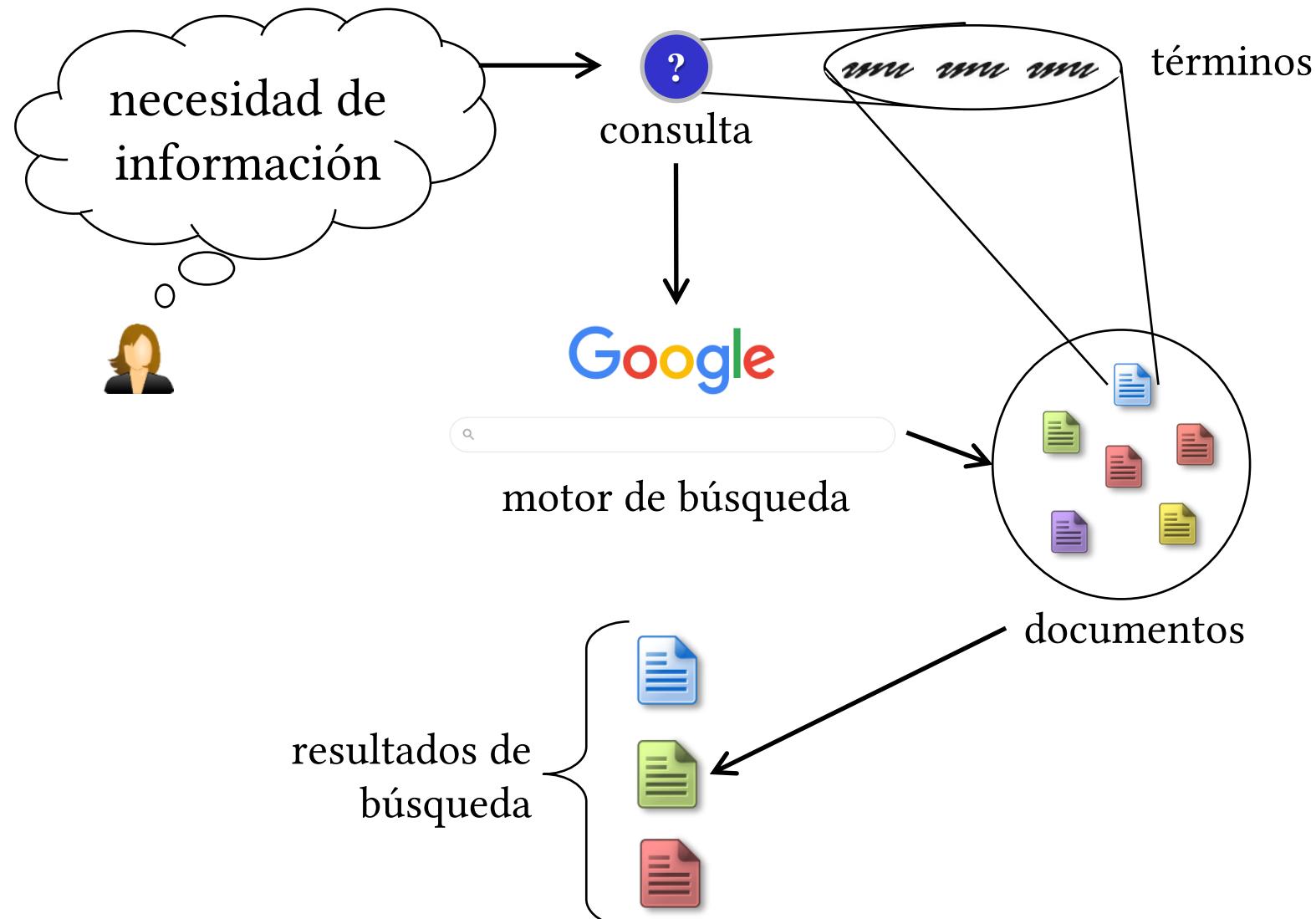


(Sanz-Cruzado & Castells, 2018)

12

Búsqueda y recomendación

Búsqueda textual (recuperación de información textual)



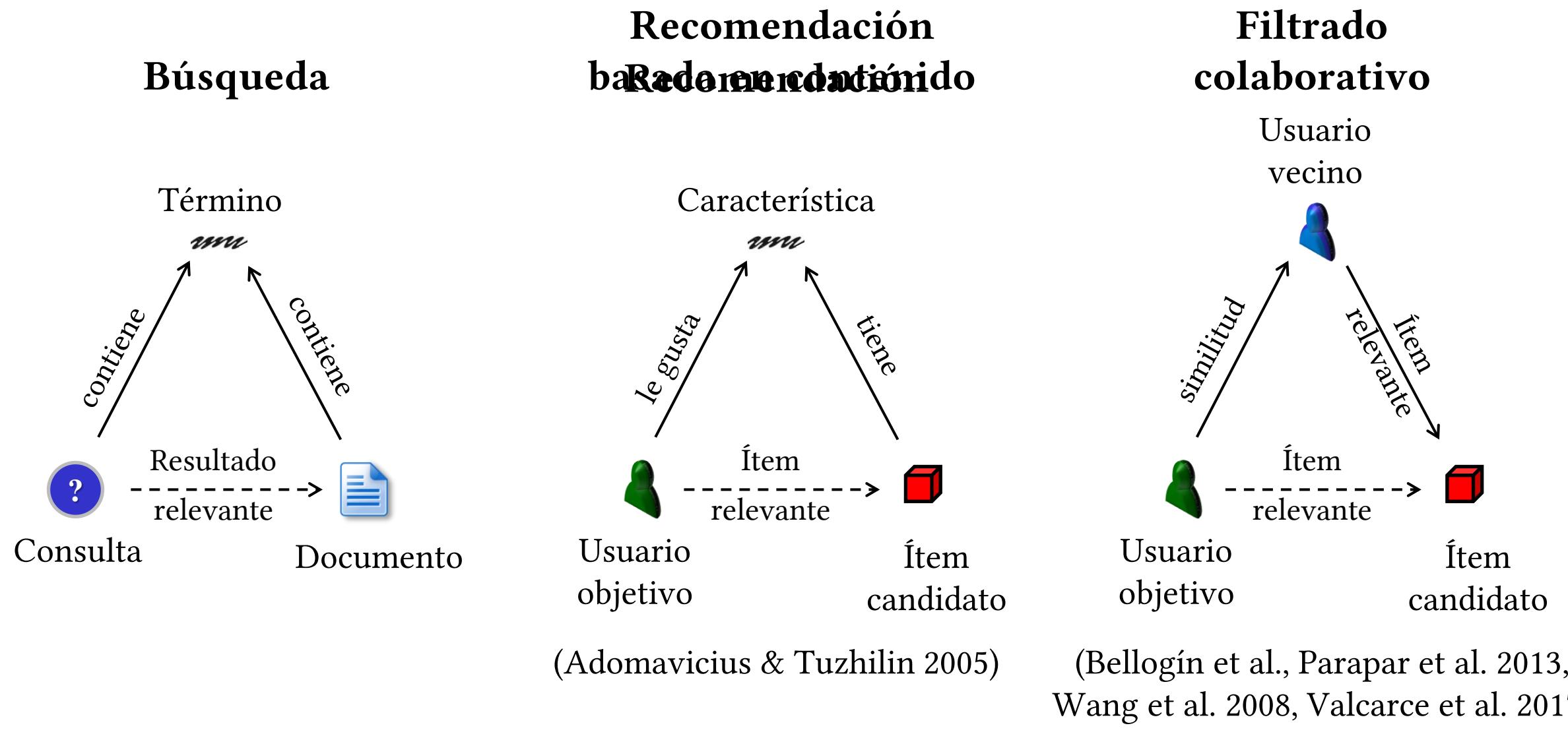
(Baeza-Yates & Ribeiro-Neto 2010)

14

- ◆ Filtrado de información (Belkin & Croft 1992):
 - Seleccionan elementos de espacios masivos
 - Solucionan la sobrecarga de información
- ◆ Solucionan una necesidad de información de los usuarios:
 - Explícita (mediante consulta) en búsqueda
 - Implícita (de descubrimiento) en recomendación
- ◆ Problemas de ranking
 - Identificar un conjunto pequeño de elementos
 - Aquellos de mayor interés para los usuarios.

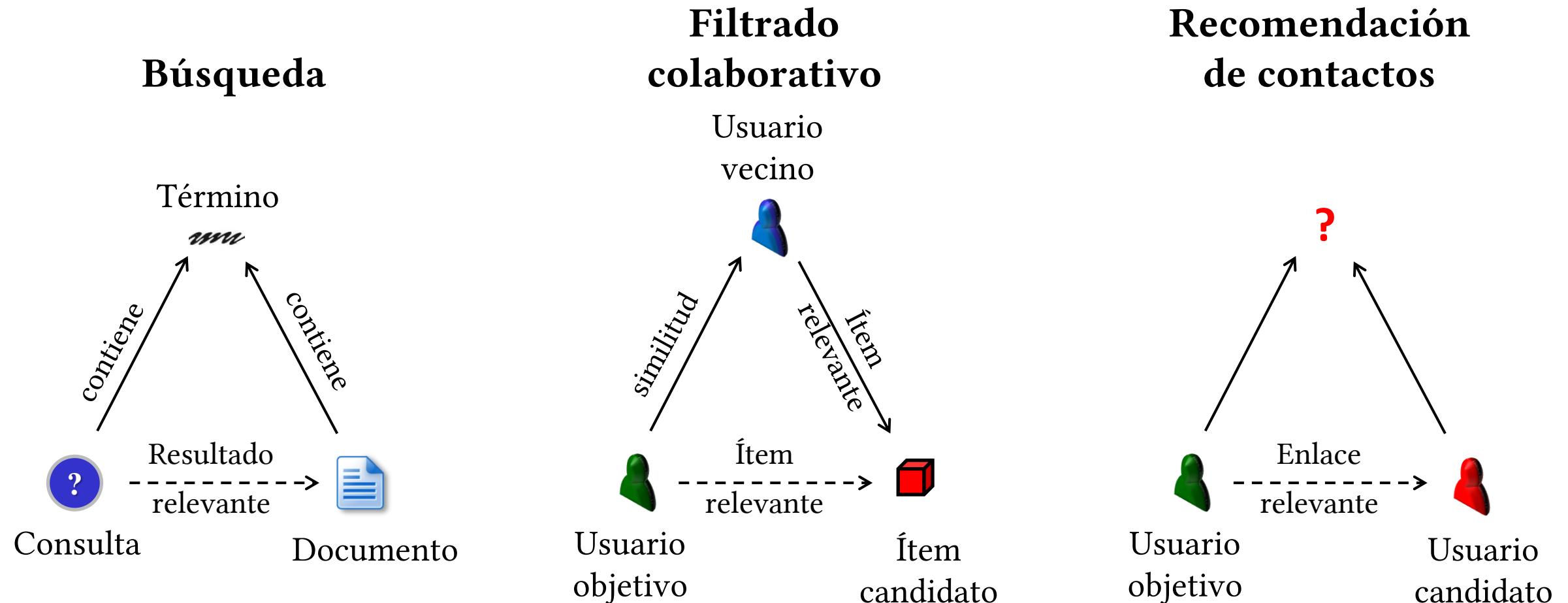
¿Podemos usar algoritmos
de búsqueda en
recomendación?

Búsqueda vs. recomendación

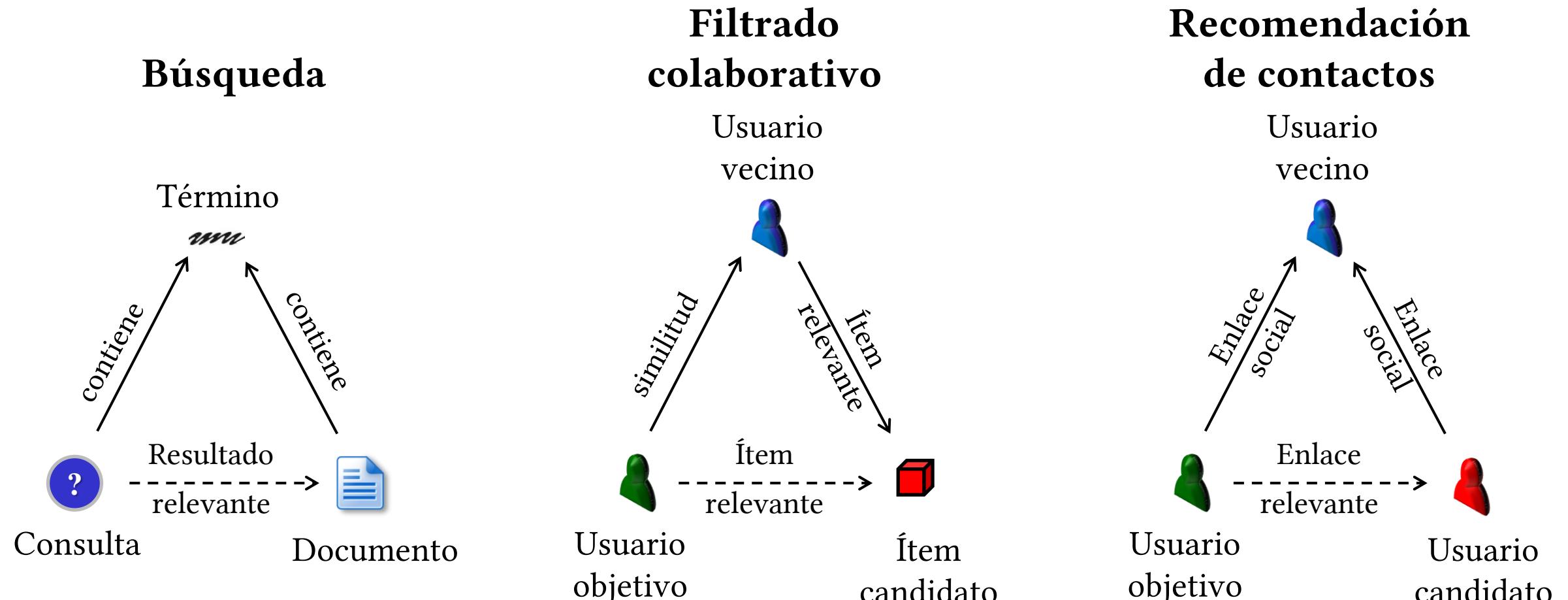


Adaptación a la recomendación de contactos

Búsqueda vs. filtrado colaborativo vs. recomendación de contactos



Búsqueda vs. filtrado colaborativo vs. recomendación de contactos



(Hannon et al. 2010)

(Sanz-Cruzado et al. 2020)

19

Un ejemplo: BM25

Búsqueda textual:

(Robertson et al., 2009)

$$f_q(d) = \sum_{t \in d \cap q} \frac{(k + 1) \text{ freq}(t, d)}{k \left(1 - b + \frac{b|d|}{\text{avg}_{d'}|d'|} \right) + \text{ freq}(t, d)} \text{ RSJ}(t)$$

$$\text{RSJ}(w) = \log \frac{|D| - |D_t| - 0.5}{|D_t| - 0.5}$$

donde

- ◆ d : documento $\longrightarrow \Gamma^d(v)$: usuario candidato
- ◆ q : consulta $\longrightarrow \Gamma^q(u)$: usuario objetivo
- ◆ $t \in d \cap q$: término $\longrightarrow t \in \Gamma^q(u) \cap \Gamma^d(v)$: usuario vecino
- ◆ D : conjunto de documentos $\longrightarrow \mathcal{U}$: conjunto de usuarios
- ◆ D_t : documentos que contienen t $\longrightarrow \Gamma_{\text{inv}}^d(t)$: usuarios v con t en $\Gamma^d(v)$
- ◆ $\text{ freq}(t, d)$: frecuencia de t en d $\longrightarrow w^d(t, v)$: peso del enlace
- ◆ $|d|$: longitud del documento d $\longrightarrow \text{len}^l(v) = \sum_{x \in \Gamma^l(v)} w^l(x, v)$

Un ejemplo: BM25

Búsqueda textual:

$$f_q(d) = \sum_{t \in d \cap q} \frac{(k + 1) \text{ freq}(t, d)}{k \left(1 - b + \frac{b|d|}{\text{avg}_{d'}|d'|} \right) + \text{ freq}(t, d)} \text{ RSJ}(t)$$

$$RSJ(w) = \log \frac{|D| - |D_t| - 0.5}{|D_t| - 0.5}$$

Recomendación de contactos:

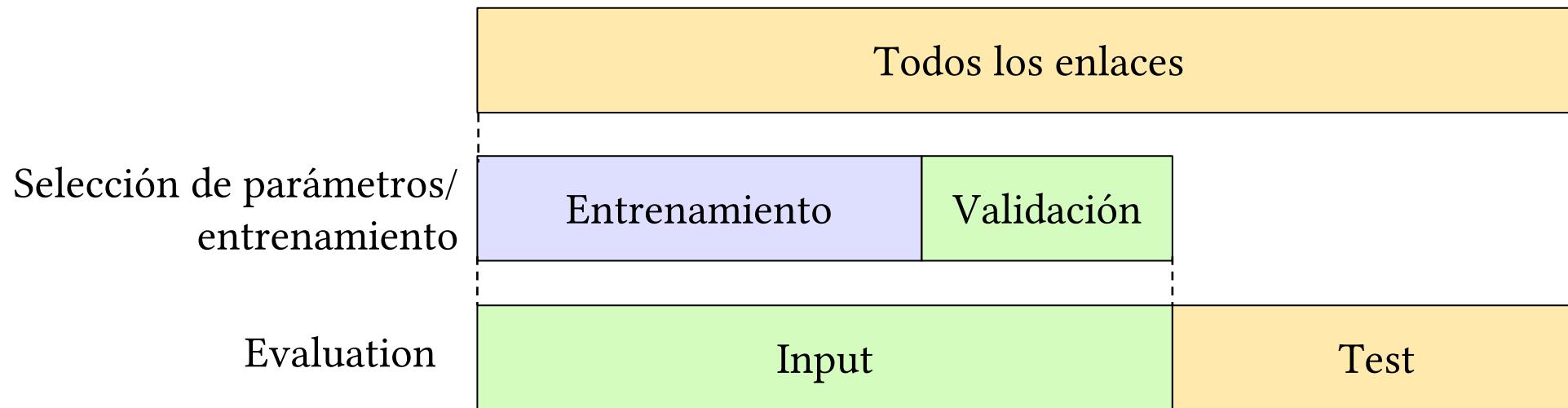
$$f_u(v) = \sum_{t \in \Gamma^q(u) \cap \Gamma^d(v)} \frac{(k + 1) w^d(t, v) \text{ RSJ}(t)}{k \left(1 - b + \frac{b \cdot \text{len}^l(v)}{\text{avg}_{v'}(\text{len}^l(v'))} \right) + w^d(t, v)}$$

$$RSJ(t) = \log \frac{|U| - |\Gamma_{\text{inv}}^d(t)| + 0.5}{|\Gamma_{\text{inv}}^d(t)| + 0.5}$$

Configuración experimental

- ◆ Evaluación offline
- ◆ Datos de Twitter y Facebook
- ◆ **Twitter:**
 - 2 muestras:
 - **1 month:** Todos los tweets entre el 19 de junio y el 19 de julio de 2015
 - **200 tweets:** 200 últimos tweets de los usuarios antes del 2 de agosto de 2015
 - 2 redes / conjunto de datos:
 - Red de interacción: $(u, v) \in E$ si u menciona/retuitea v
 - Red de seguimiento
- ◆ **Facebook:**
 - De Stanford Large Network Dataset Collection
 - Unión de 10 redes ego

- ◆ Partición:



- ◆ Selección de hiperparámetros: búsqueda en rejilla (nDCG@10)
- ◆ Evaluación mediante métricas de ranking: nDCG@10, MAP@10

Estadísticas de los conjuntos de datos

	Twitter 1-month		Twitter 200-tweets		Facebook
	Interacción	Seguimiento	Interacción	Seguimiento	
# usuarios	9,528	9,770	9,985	9,964	4,039
# enlaces (input)	170,425	645,022	104,866	427,568	56,466
# enlaces test	54,335	81,110	21,598	98,519	17,643
Dirigido	✓	✓	✓	✓	✗
Con pesos	✓	✗	✓	✗	✗
Tipo de partición	Temporal	Temporal	Temporal	Temporal	Aleatoria

- ◆ Búsqueda (IR):
 - **Probability ranking principle:** BM25, BIR, Extreme BM25
 - **Modelos de lenguaje:** Query likelihood (QLJM, QLD, QLL)
 - **Divergence from randomness:** PL2, DFRee, DFReeKLIM, DLH, DPH
 - Modelo vectorial (VSM)
- ◆ Filtrado colaborativo
 - kNN basado en usuario / item (similitud por coseno)
 - Implicit matrix factorization (iMF)
- ◆ Métodos específicos
 - **Amigos de amigos:** Adamic-Adar, MCN, Jaccard, similitud por coseno
 - **Paseos aleatorios (*random walks*):** Personalized PageRank, Money,...
 - **Basados en caminos:** Local Path Index, Katz...
- ◆ Comprobación: Recomendación aleatoria y por popularidad

Resultados (nDCG@10)

Algoritmo	200-tweets		Facebook
	Interacción	Seguimiento	
BM25	0.1097	0.1159	0.5731
BIR	0.1004	0.1140	0.572
PL2	0.0983	0.1166	0.5712
VSM	0.0425	0.0787	0.5237
iMF	0.1035	0.1329	0.5210
kNN usuario	0.0954	0.1297	0.5457
kNN ítem	0.0724	0.1205	0.4542
Adamic-Adar	0.0997	0.1140	0.5746
MCN	0.0948	0.1110	0.5585
Pers. PageRank	0.0630	0.0843	0.5891
Coseno	0.0480	0.0768	0.4943
Popularidad	0.0422	0.0397	0.0523
Aleatorio	0.0003	0.0018	0.0030

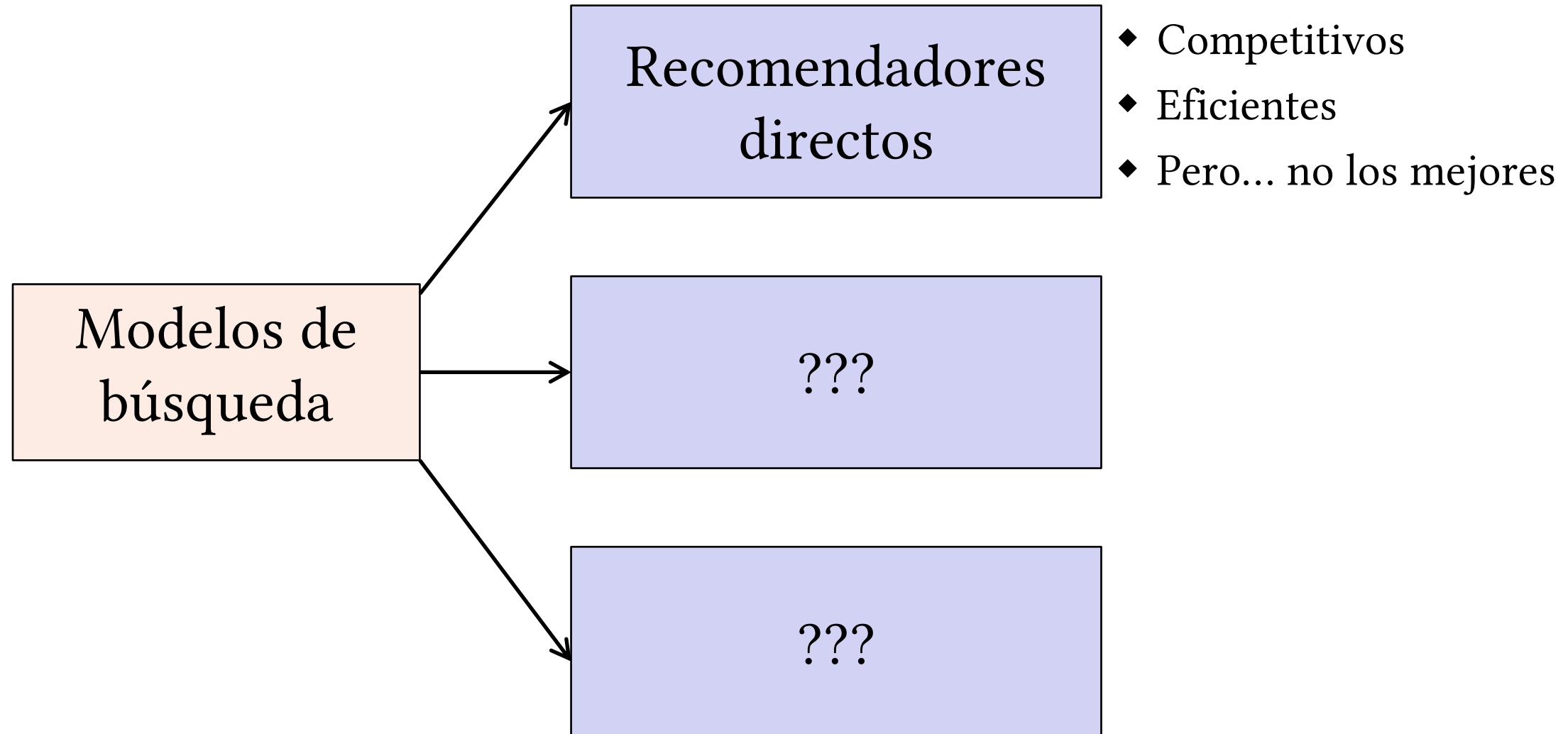
◆ Los algoritmos de IR son efectivos

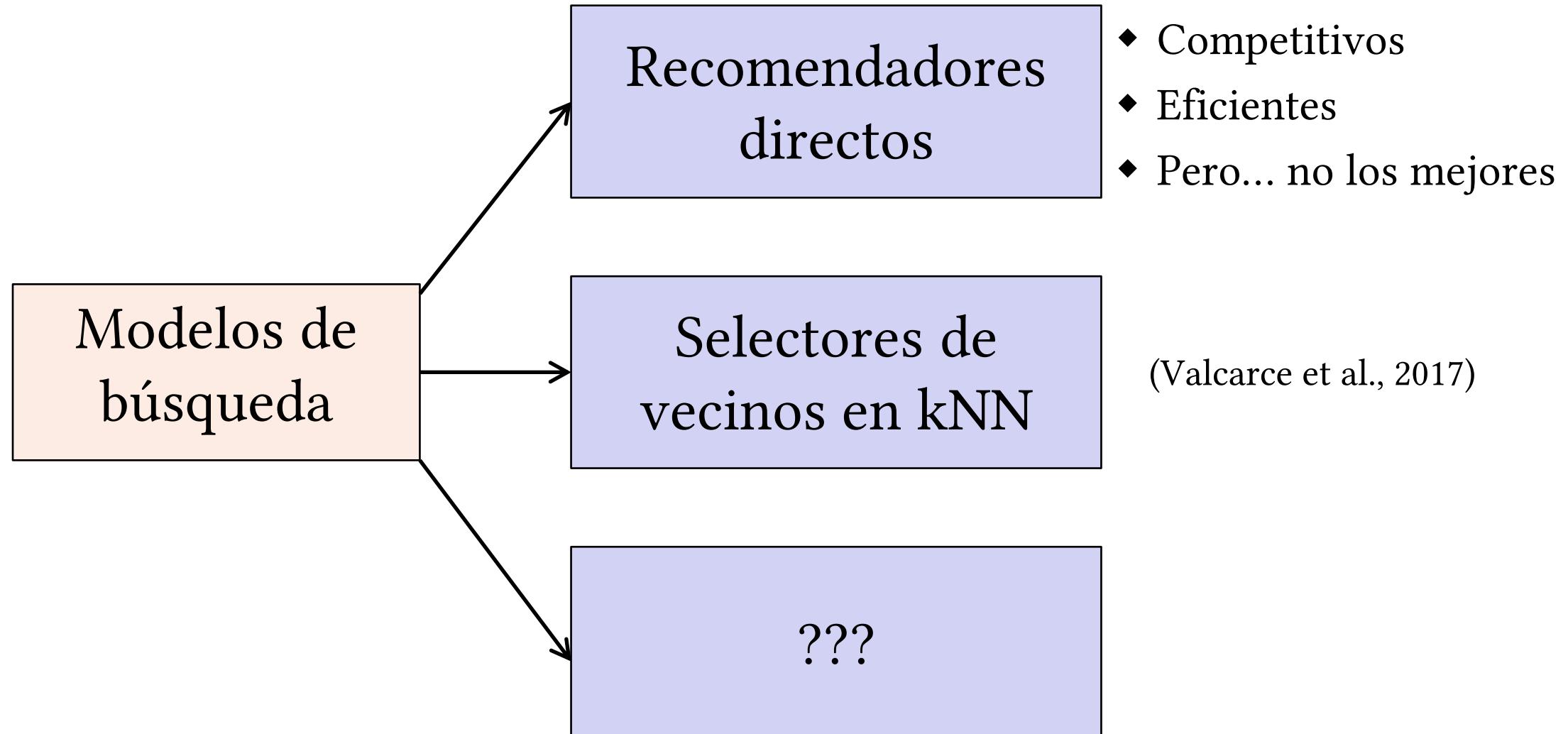
- BM25 entre el top 5
- **Mejor en:** 200-tweets interacciones
- VSM el peor modelo de búsqueda

◆ Resto de algoritmos

- Implicit MF es el mejor.
- Adamic-Adar y MCN son competitivos.
- Jaccard/coseno no son muy competitivos.
- El resto parece depender mucho de la red.

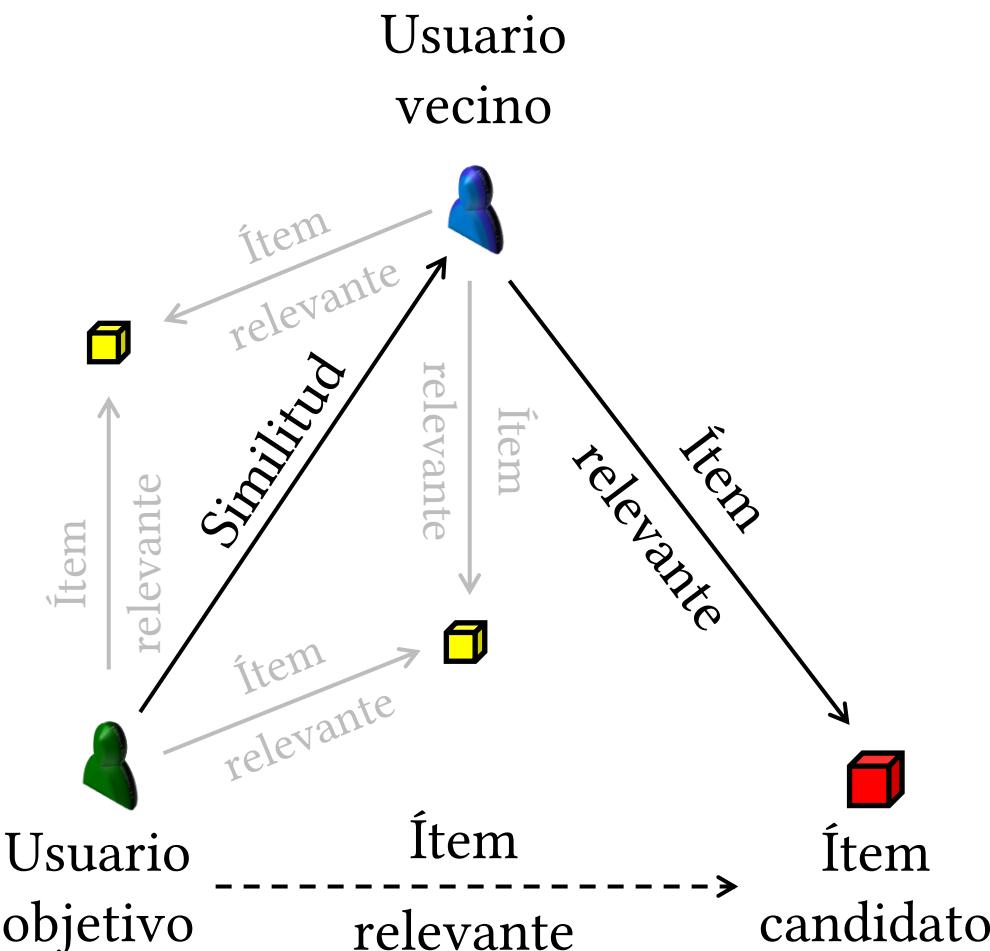
Adaptación avanzada



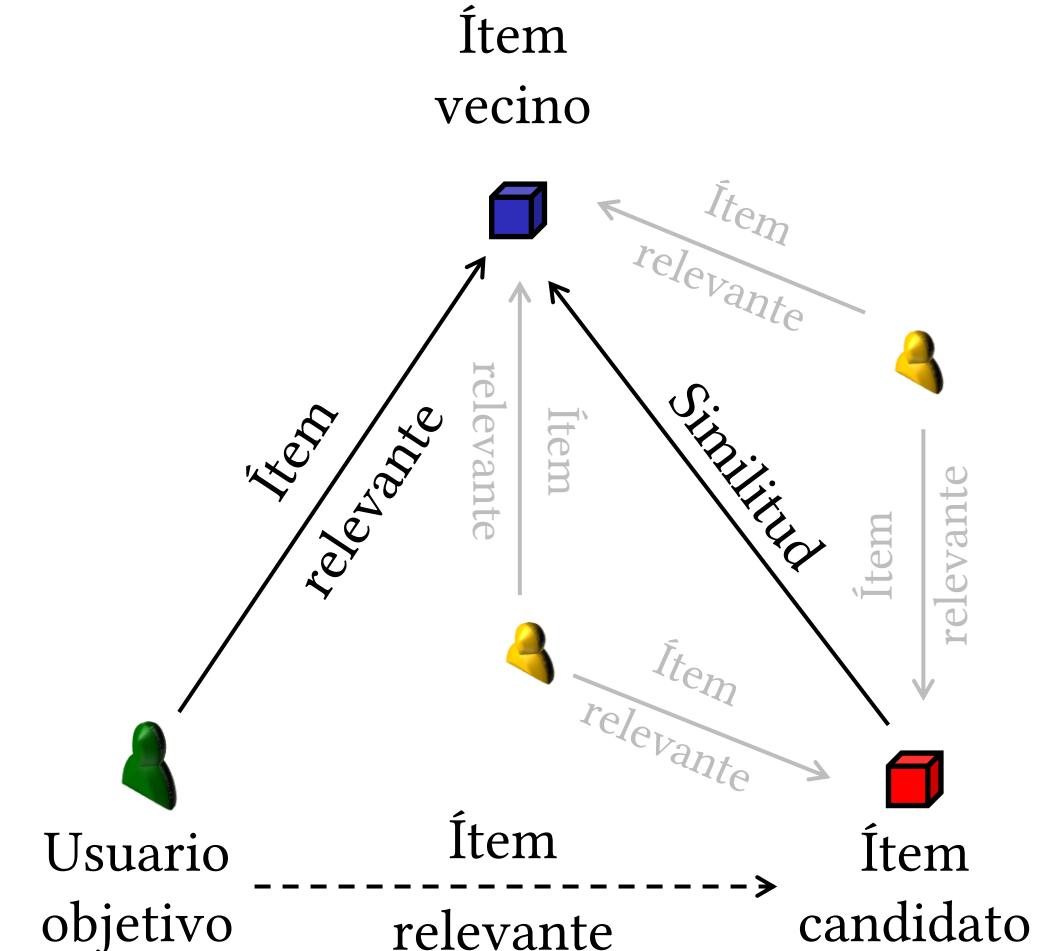


(Ning et al. 2015)

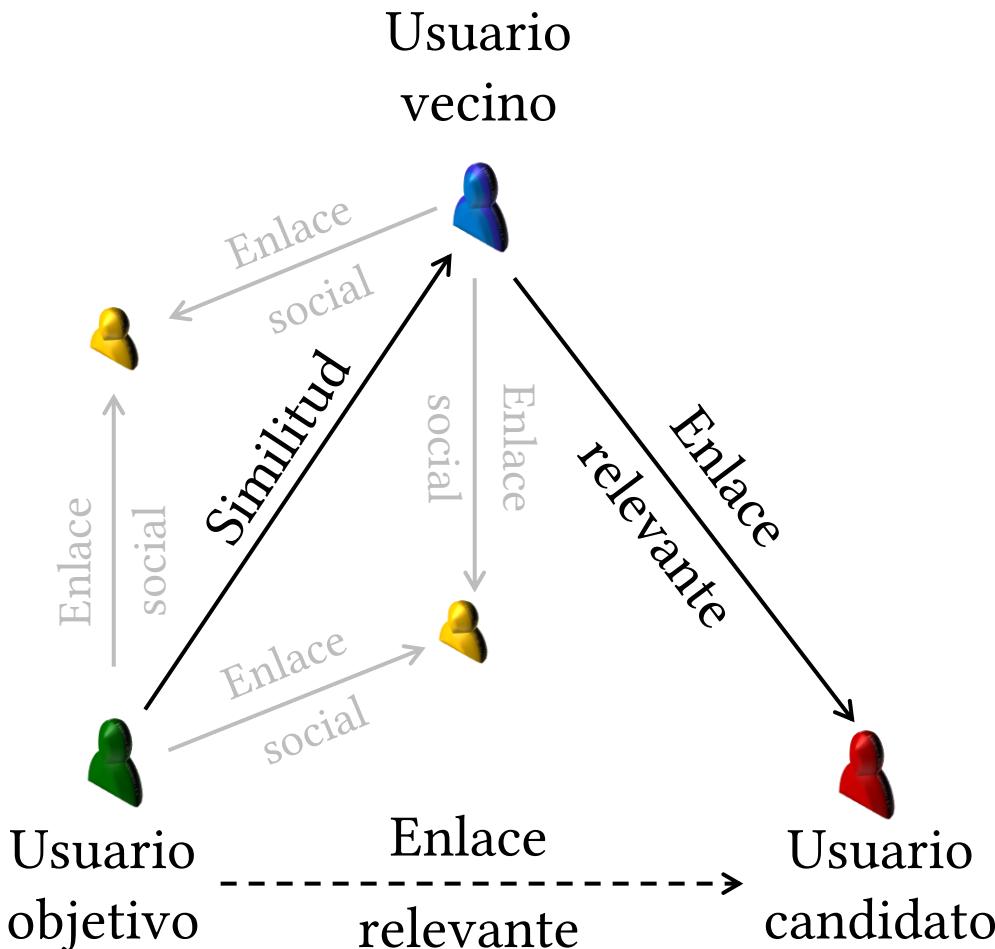
Basado en usuarios



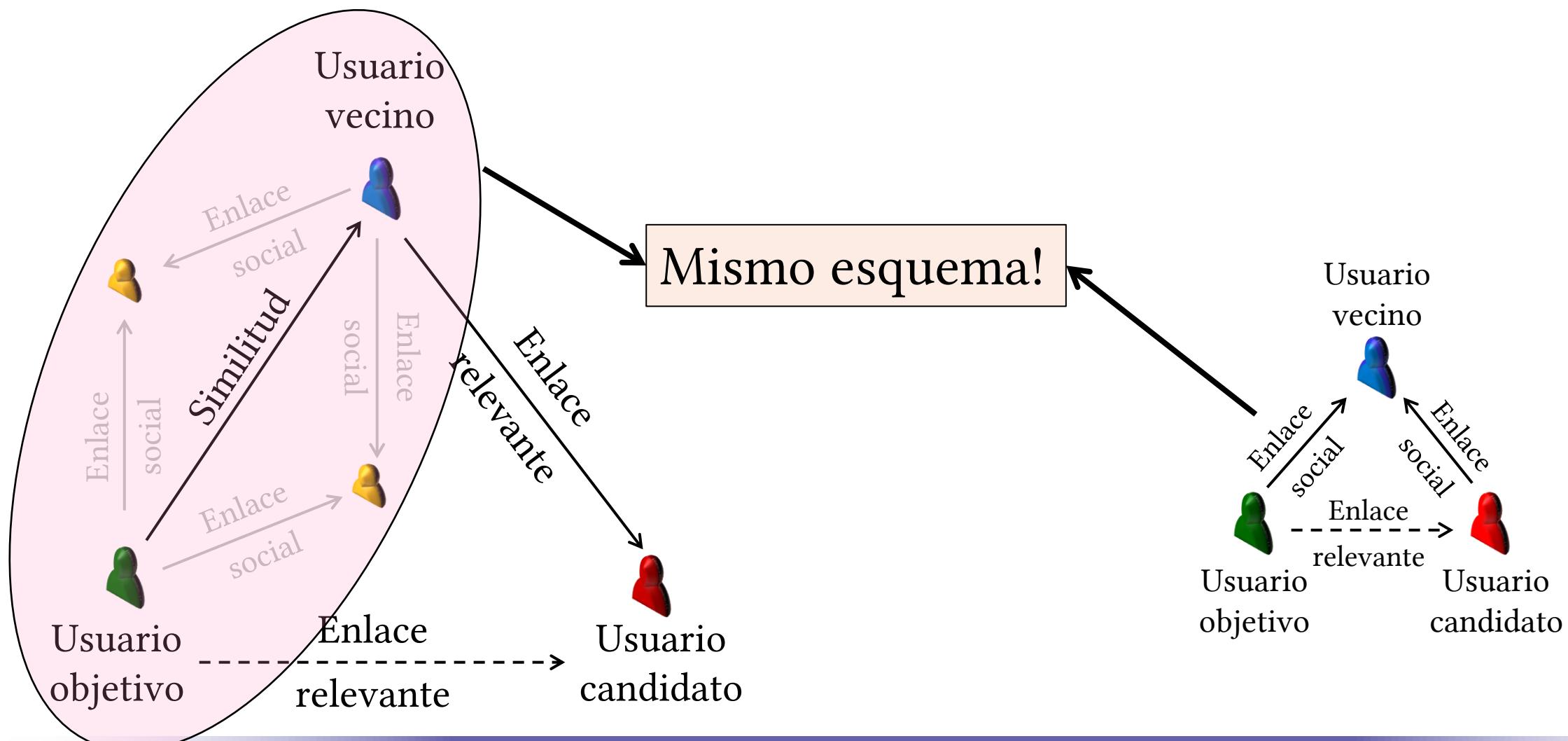
Basado en ítems



kNN en recomendación de contactos



$$f_u(v) = \sum_{t \in N(u)} sim(u, t) \cdot w(t, v)$$

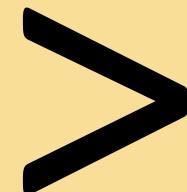


Métodos de selección de vecinos

- ◆ Algoritmos basados en amigos de amigos como similitudes
 - Adamic Adar, MCN, Jaccard
 - Todos los modelos de IR
- ◆ Volvamos a los experimentos anteriores...

Algoritmo	Twitter 1-month		Twitter 200-tweets		Facebook
	Interacción	Seguimiento	Interacción	Seguimiento	
kNN basado en usuario	0.1367	0.1413	0.0954	0.1297	0.5457
kNN basado en ítem	0.1174	0.1296	0.0724	0.1205	0.4542
Coseno	0.0393	0.0497	0.0480	0.0768	0.4943

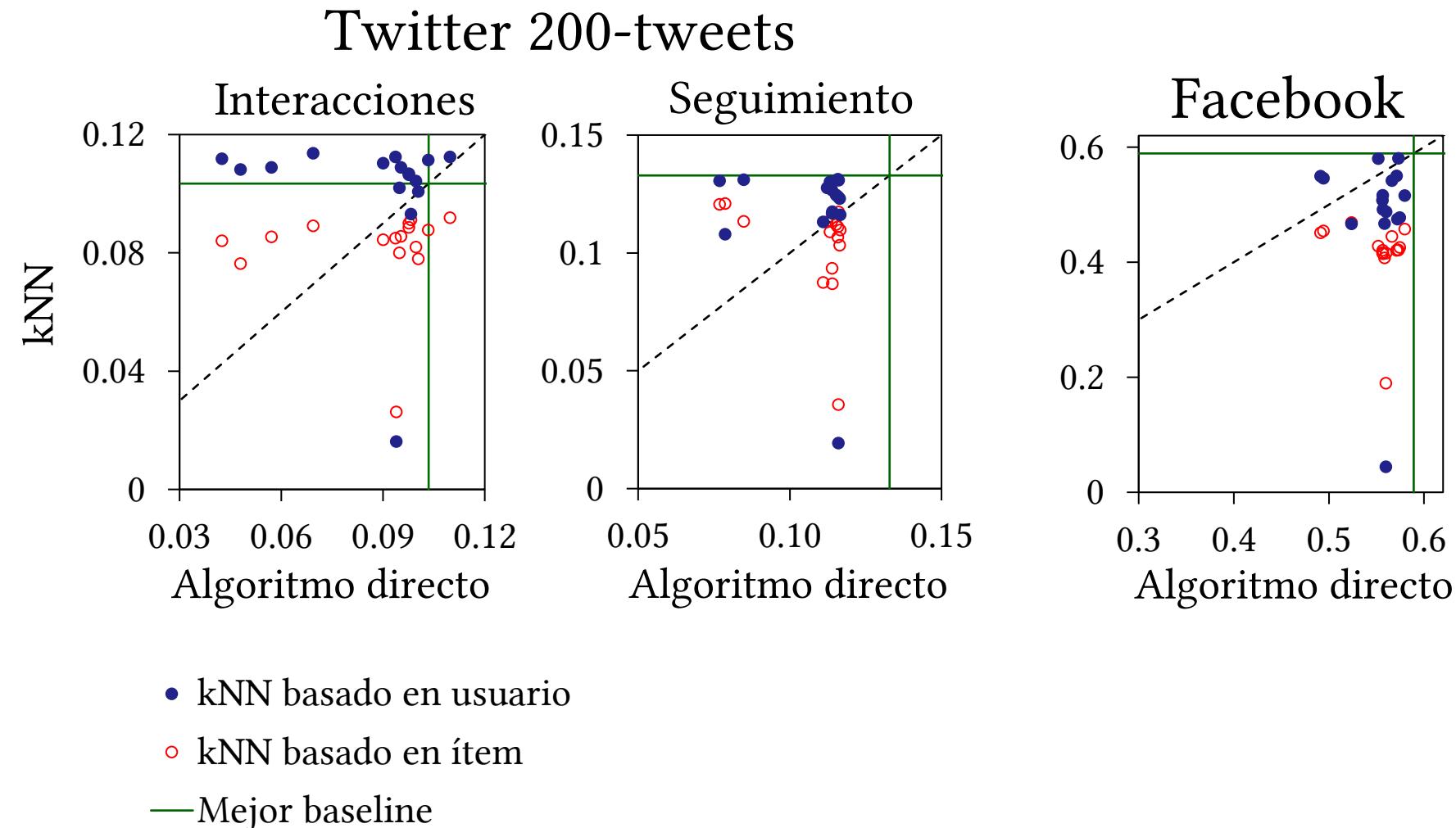
kNN basado en usuario / ítem
(similitud por coseno)

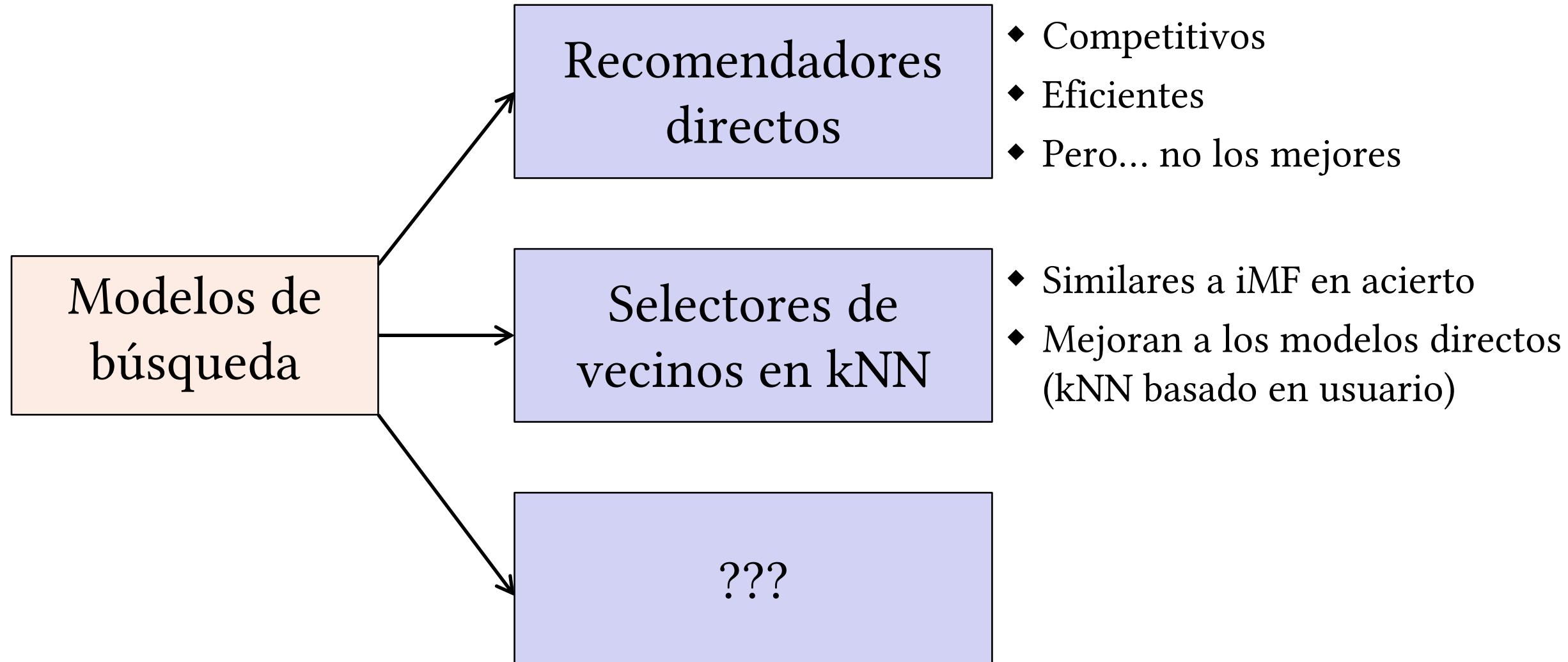


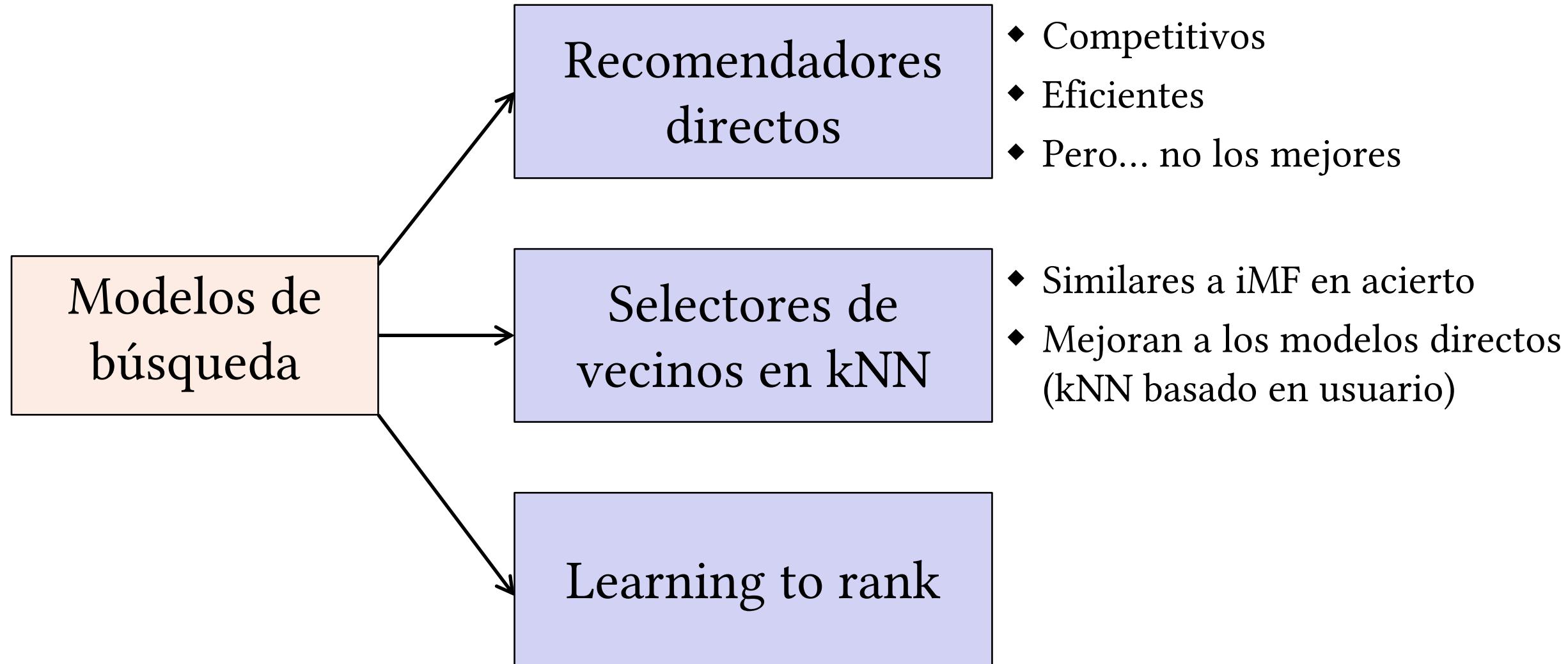
Similitud por coseno
como recomendador

Entonces, ¿y si utilizamos los modelos de IR?

Resultados kNN + IR (nDCG@10)

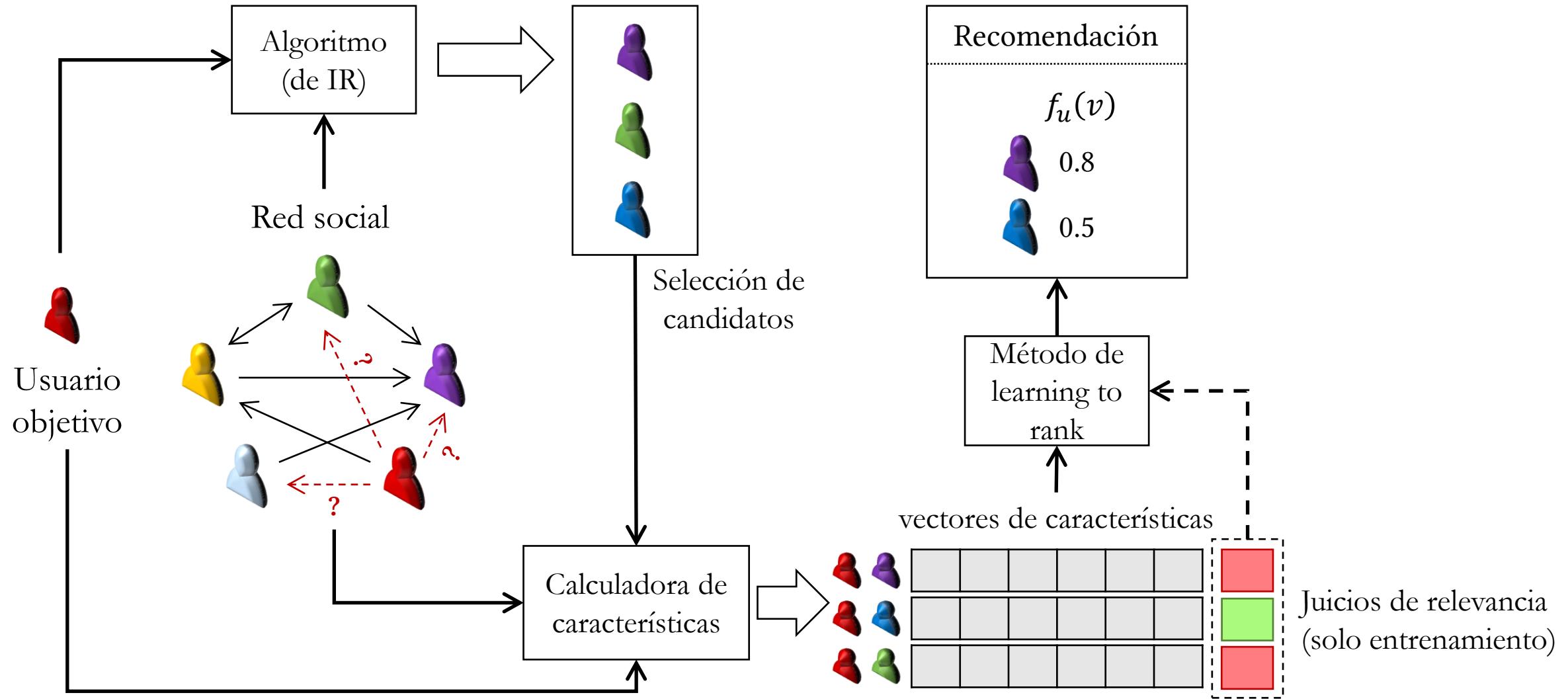






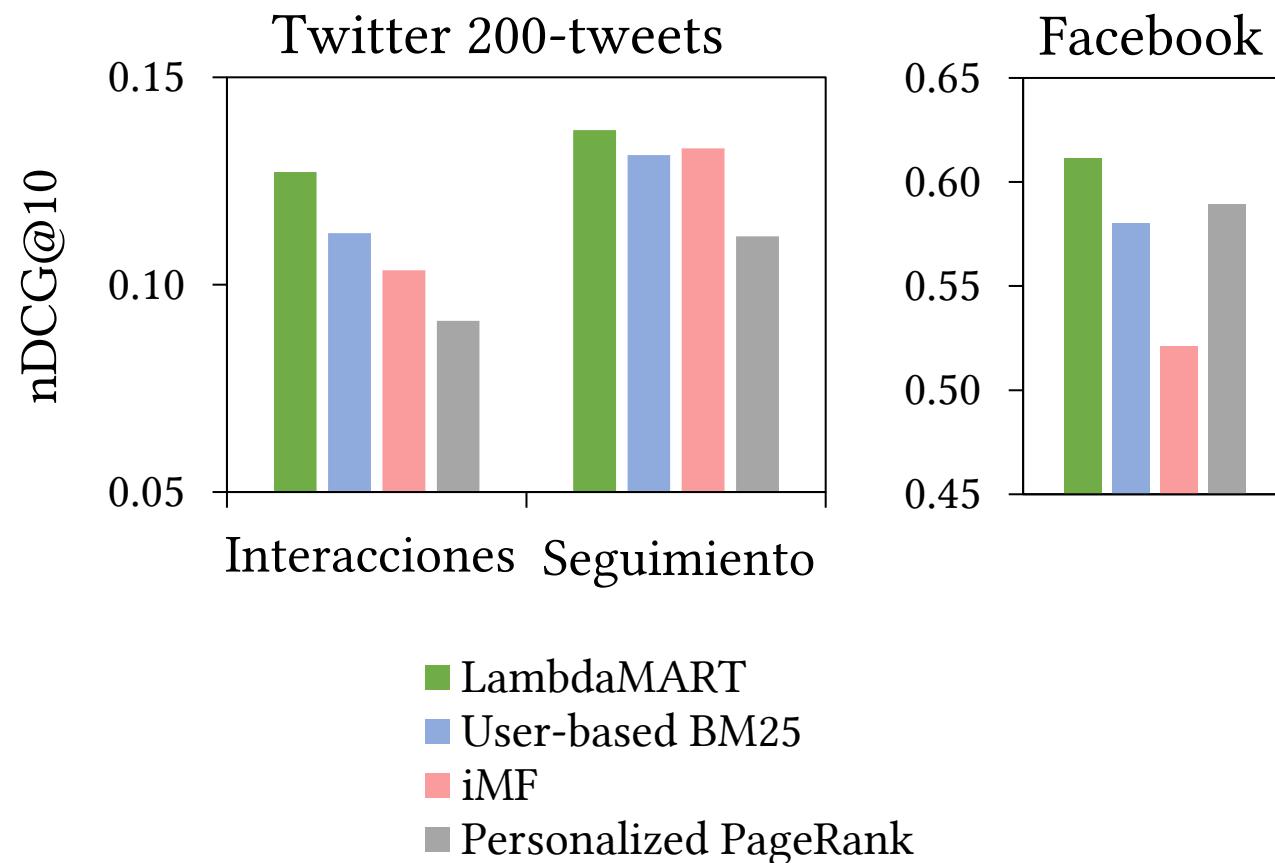
- ◆ Algoritmos supervisados
- ◆ Basados en vectores de características
- ◆ Muy efectivos en IR (Liu 2007)
- ◆ Aprender a ordenar documentos (ítems, usuarios...)

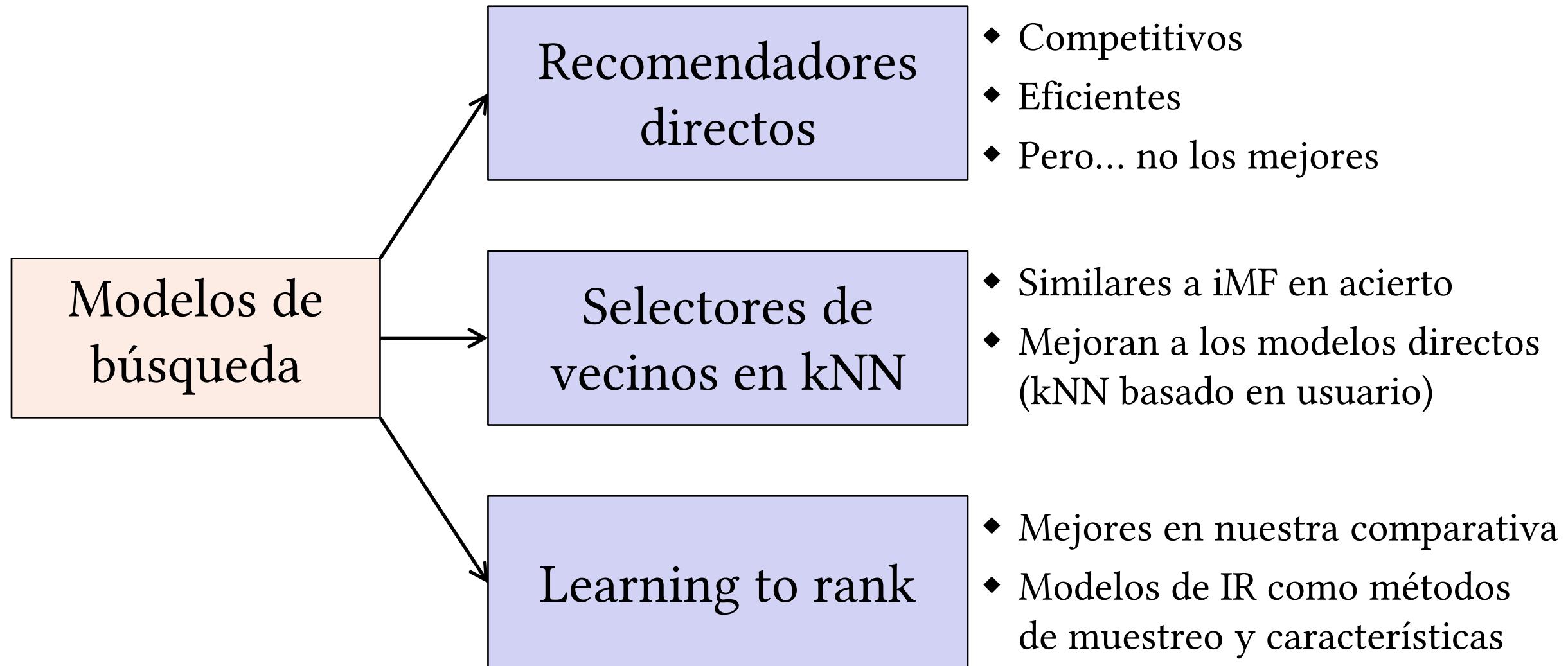
¿Cómo funciona?



- ◆ **Método:** LambdaMART (Burges 2010, Ganjisaffar et al. 2011)
 - Árboles de regresión
 - Muy utilizado en IR
- ◆ **Características:**
 - Modelos de IR
 - Algoritmos de amigos de amigos (AdA)
 - kNN basado en usuario / ítem+ IR / AdA
- ◆ **Selección de candidatos:** utilizar modelos de búsqueda

LambdaMART mejora a los baselines más efectivos





Conclusiones

Conclusiones

- ◆ La recomendación de contactos y la búsqueda están muy relacionadas.
- ◆ Podemos utilizar modelos de búsqueda para recomendar contactos.
- ◆ Los modelos de IR son competitivos y eficientes (BM25).
- ◆ Funcionan mejor para seleccionar vecinos en kNN.
- ◆ Las técnicas de learning to rank mejoran el acierto de los mejores algoritmos del estado del arte.
- ◆ Los modelos de IR son efectivos en tres papeles en recomendación de contactos
 - Recomendadores directos.
 - Selectores de vecinos en kNN.
 - Métodos de muestreo y características en learning to rank.

- ◆ Experimentos on-line
- ◆ Adaptación de otras técnicas de búsqueda
 - Deep learning
 - Reformulación de consultas
 - Relevance feedback
- ◆ Recomendación interactiva
 - Efectividad en varios pasos.
 - Comparación con bandidos multi-brazo.
 - Efectos en las propiedades de red.

Referencias

Adomavicius, G., Tuzhilin, A. **Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions.** *IEEE TKDE* 17(6), 2005.

Baeza-Yates, R., Ribeiro-Neto, B. **Modern Information Retrieval: the concepts and technology behind search**, 2nd edition.

Belkin, N., Croft, B. **Information filtering and information retrieval: two sides of the same coin?** *Communications of the ACM* 35(12), pp. 29-38, 1992.

Bellogín, A., Wang, J., Castells, P. **Bridging memory-based collaborative filtering and text retrieval.** *Information Retrieval* 16(6), pp. 697-724, 2013.

Burges, C. **From RankNet to LambdaRank to LambdaMART: an overview.** *Microsoft Technical Report*, 2010.

Ganjisaffar, Y., Caruana, R., Lopes, C. **Bagging gradient-boosted trees for high precision, low variance ranking models.** *SIGIR* 2011, Beijing, China, pp. 85-94, 2011.

Guy, I. **People recommendation on social media.** *Social Information Access*, pp. 570-623, 2018.

Referencias (II)

Hannon, J., Bennet, M., Smyth, B. **Recommender Twitter users to follow using content and collaborative filtering approaches.** *ACM RecSys 2010*, Barcelona, Spain, pp. 199-206, 2010.

Liu, T. **Learning to rank for information retrieval.** *Foundations and Trends in Information Retrieval* 3(3), pp. 225-331, 2007.

Ning, X., Desrosiers, C., Karypis, G. **A comprehensive survey of neighborhood-based recommendation methods.** *Recommender Systems Handbook*, 2nd edition, pp. 37-76, 2015.

Parapar, J., Bellogín, A., Castells, P., Barreiro, A. **Relevance-based language modelling for recommender systems.** *Information Processing & Management* 49, pp. 966-980, 2013.

Robertson, S., Zaragoza, H. **The Probabilistic Relevance Framework: BM25 and beyond.** *Foundations and Trends in Information Retrieval* 3(4), pp. 333-389, 2009.

Sanz-Cruzado, J., Castells, P. **Contact recommendations in social networks.** *Collaborative Recommendations: Algorithms, Practical Challenges and Applications*, pp. 519-569, 2018.

Referencias (III)

Sanz-Cruzado, J., Castells, P., Macdonald, C., Ounis, I. **Effective contact recommendation in social networks by adaptation of Information Retrieval models.** *Information Processing & Management* 57(5), 102285, 2020.

Sanz-Cruzado, J., Macdonald, C., Ounis, I., Castells, P. **Axiomatic analysis of contact recommendation methods in social networks: an IR perspective.** *ECIR 2020*, Online, pp. 175-190, 2020.

Sanz-Cruzado, J., Castells, P. **Information retrieval models for contact recommendation in social networks.** *ECIR 2019*, pp. 148-163, 2020.

Valcarce, D., Parapar, J., Barreiro, A. **Axiomatic analysis of language modelling of recommender systems.** *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 25, pp. 113-127, 2017.

Wang, J., Robertson, S., de Vries, A.P., Reinders, M.J.T. **Probabilistic relevance ranking for collaborative filtering.** *ACM TOIS*, 26, article 16. 2008.

Gracias por vuestra atención!

¿Alguna pregunta?

Javier Sanz-Cruzado

Twitter: @JavierSanzCruza

E-mail: javier.sanz-cruzado@uam.es

Web: <https://javiersanzcruza.github.io>



49