



Pontificia Universidad Católica de Chile
Escuela de Ingeniería
Departamento de Ciencia de la Computación

PRESENTACIÓN CURSO SISTEMAS RECOMENDADORES

SISTEMAS RECOMENDADORES BASADOS EN GRAFOS

Estudiante: Juan Pablo Salazar Fernández

Profesor: Dr. Denis Parra Santander

05 de noviembre de 2015

› Artículos incluidos

1. **A Graph-based Recommender System for Digital Library.** Zan Huang, Wingyan Chung, Thian-Huat Ong, and Hsinchun Chen. In Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries (JCDL '02). ACM, New York, NY, USA, 65-73.
2. **A Graph-based model for context-aware recommendation using implicit feedback data.** Yao, W., He, J., Huang, G., Cao, J., & Zhang, Y. World Wide Web (2015) 18:1351–1371

› Agenda

1. **Introducción.**
2. **Incorporación de información de contexto.**
3. **Solución de Huang et. al. (2002)**
 - Problema a resolver.
 - Modelo utilizado.
 - Evaluación y resultados.
4. **Solución de Yao et. al. (2015)**
 - Problema a resolver.
 - Modelo Multi-Layer Context Graph (MLCG)
 - Evaluación y resultados.
5. **Conclusiones y comentarios.**

› Introducción

1. Aporte a problema de “information overload”.
2. Feedback explícito, matrices usuario-item relativamente densas.
3. Necesidad de mejorar resultados sin tener preferencias explícitas.
4. Feedback implícito.
 1. Información acerca del ítem.
 2. Información acerca del usuario.
 3. Información acerca de la transacción – contexto de decisión.

› Información de contexto.

1. Contexto del ítem.
2. Contexto del usuario.
3. Contexto de la decisión.



› Información acerca de los ítemes.

1. Descripción de cada producto.
2. Especificaciones.
3. Características.
4. Pertenencia a grupos / jerarquías / tipos.

› Información acerca de los usuarios.

1. Género.
2. Edad.
3. Educación.
4. Estado civil.
5. Número de hijos.
6. Edades de los hijos.
7. País, ciudad de residencia.
8. Trabajo.

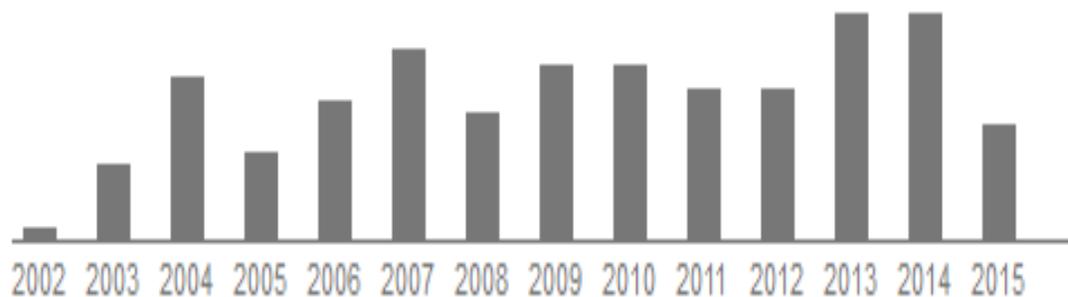
› Información acerca de la transacción.

1. Lugar / Actividad.
2. Día / hora.
3. Método de pago.
4. Método de envío / entrega.

A Graph-based Recommender System for Digital Library.

Zan Huang, Wingyan Chung, Thian-Huat Ong, and Hsinchun Chen. (2002).

A Graph-based Recommender System for Digital Library. Zan Huang, Wingyan Chung, Thian-Huat Ong, and Hsinchun Chen. (2002).



- 163 citas en Google Scholar (4/11/2015).
- 9 citas el 2015.
- Artículos más antiguos que lo citan (anteriores a 2007), son mucho más citados que los nuevos, varios con más de 100 citas.

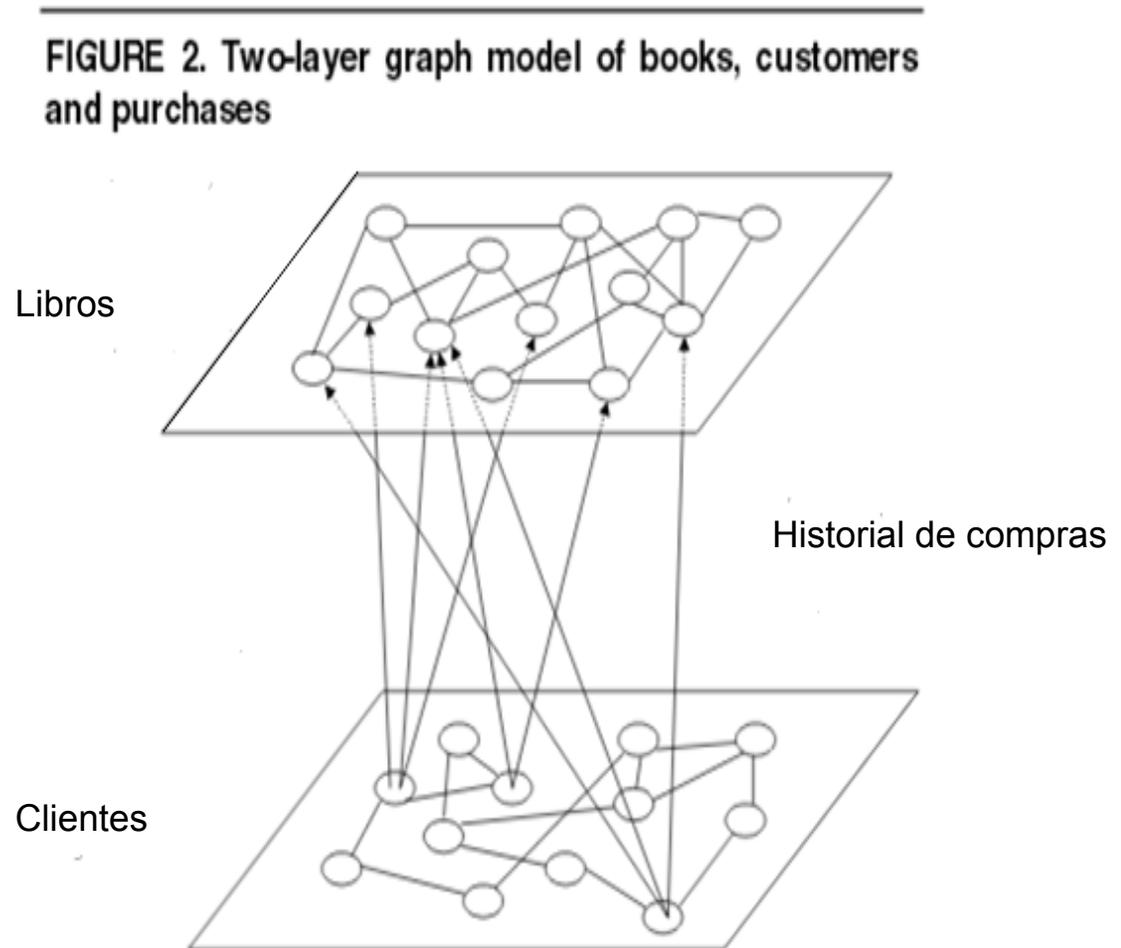
› Problema a resolver.

- Recomendación de libros a clientes de una librería digital de Taiwán.
- Integración de técnicas basadas en contenido con técnicas de filtrado colaborativo.
- Desarrollo de solución general, aplicable en otros contextos.

› Modelo Utilizado

Características:

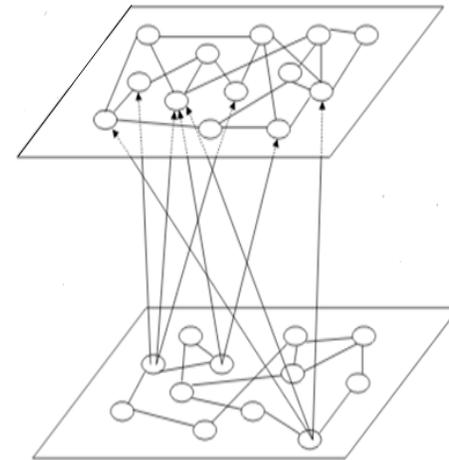
- Flexible.
- Integrador.
- Modular.



› Modelo - Representación

Dos etapas:

1. Creación de vectores de características para libros y clientes.
2. Construcción del grafo, calculando similitudes (aristas).



› Modelo – Representación – Etapa 1

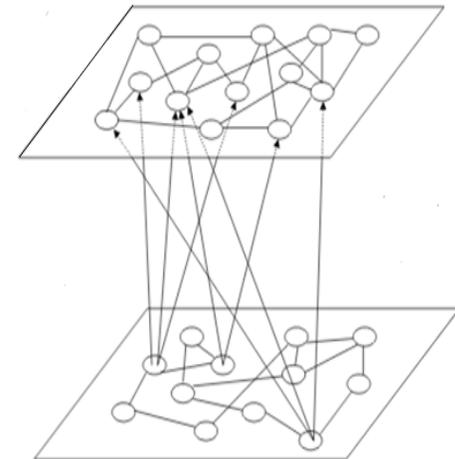
Representación de libros:

- Generación de listas de términos mediante método de información mutua. Análisis de co-ocurrencia de palabras y frases.

$$MI_c = \frac{f_c}{f_{left} + f_{right} - f_c}$$

Representación de clientes:

- Vector de características.



› Modelo – Representación – Etapa 2

- Cálculo de similaridad entre libros (basada en (TF-IDF) :

- Importancia de término en un libro.

$$d_{ij} = tf_{ij} \times \log [(N / df_j) \times w_j]$$

- Importancia del término en ambos libros.

$$d_{ijk} = tf_{ijk} \times \log [(N / df_{jk}) \times w_j]$$

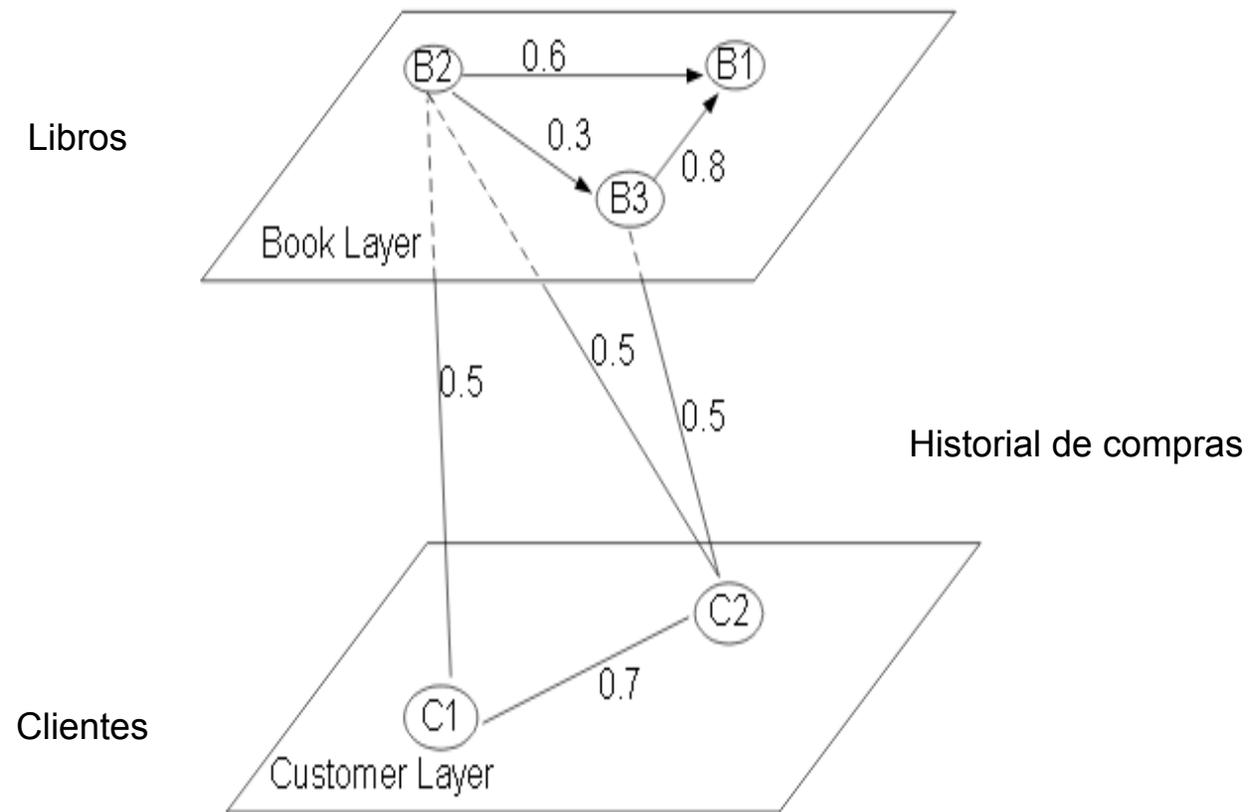
- Similaridad (asimétrica)

$$\text{Cluster Weight } (B_j, B_k) = \frac{\sum_{i=1}^n d_{ijk}}{\sum_{i=1}^n d_{ij}} \times \frac{\log(N / df_k)}{\log(N)}$$

$$\text{Cluster Weight } (B_k, B_j) = \frac{\sum_{i=1}^n d_{ikj}}{\sum_{i=1}^n d_{ik}} \times \frac{\log(N / df_j)}{\log(N)}$$

› Modelo – Resultado representación

FIGURE 3. Simplified two-layer graph example



› Recomendación en base al modelo

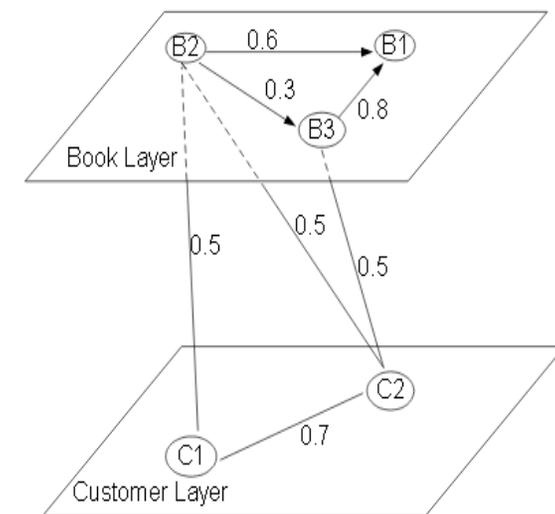
Low-degree association.

- Recomendación de libros más similares a los ya comprados por el cliente (1 salto).

High-degree association.

- Recomendación utilizando red de Hopfield (múltiples saltos).

FIGURE 3. Simplified two-layer graph example



› Recomendación en base al modelo

Red de Hopfield

- Inicialización: $u_i(0)=1$ para libros comprados anteriormente.

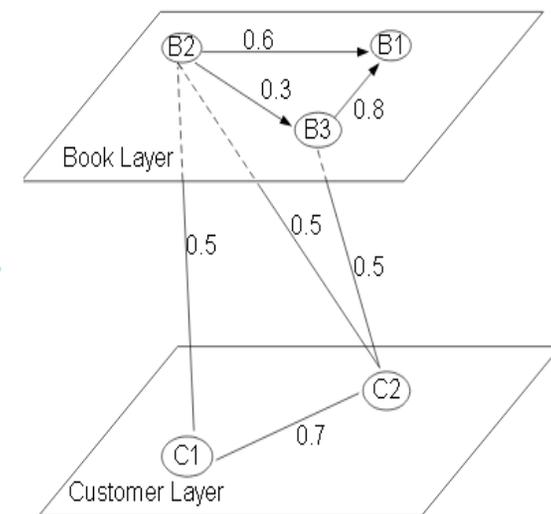
- Iteración:

$$\mu_j(t+1) = f_s \left[\sum_{i=0}^{n-1} t_{ij} \mu_i(t) \right], 0 \leq j \leq n-1$$

- Criterio de c

$$\sum_{j=0}^{n-1} |\mu_j(t+1) - \mu_j(t)| \leq \epsilon$$

FIGURE 3. Simplified two-layer graph example



› Evaluación y resultados

TABLE 1. Hold-out Testing Precision and Recall (sample size: 100)

| Approach | Low-degree association | | High-degree association | |
|------------------------|------------------------|--------|-------------------------|--------|
| | Precision | Recall | Precision | Recall |
| Content-based approach | 2.48% | 10.58% | 1.94% | 7.80% |
| Collaborative approach | 1.56% | 8.15% | 1.78% | 10.43% |
| Hybrid approach | 2.76% | 14.52% | 3.32% | 13.51% |



Hold-out (validación cruzada).

- 100 clientes, escogidos aleatoriamente.
- Para cada cliente:
 - 50% compras más antiguas, usadas para training.
 - 50% compras más nuevas, usadas para testing.

› Evaluación y resultados

TABLE 3. Subject Testing Precision and Recall

| Approach | Low-degree association | | High-degree association | |
|------------------------|------------------------|--------|-------------------------|--------|
| | Precision | Recall | Precision | Recall |
| Content-based approach | 16.0% | 36.3% | 18.3% | 38.1% |
| Collaborative approach | 7.0% | 27.7% | 6.0% | 20.7% |
| Hybrid approach | 10.3% | 31.6% | 7.0% | 34.4% |

Evaluación de usuarios expertos

- 2 ciudadanos taiwaneses con estudios de postgrado.

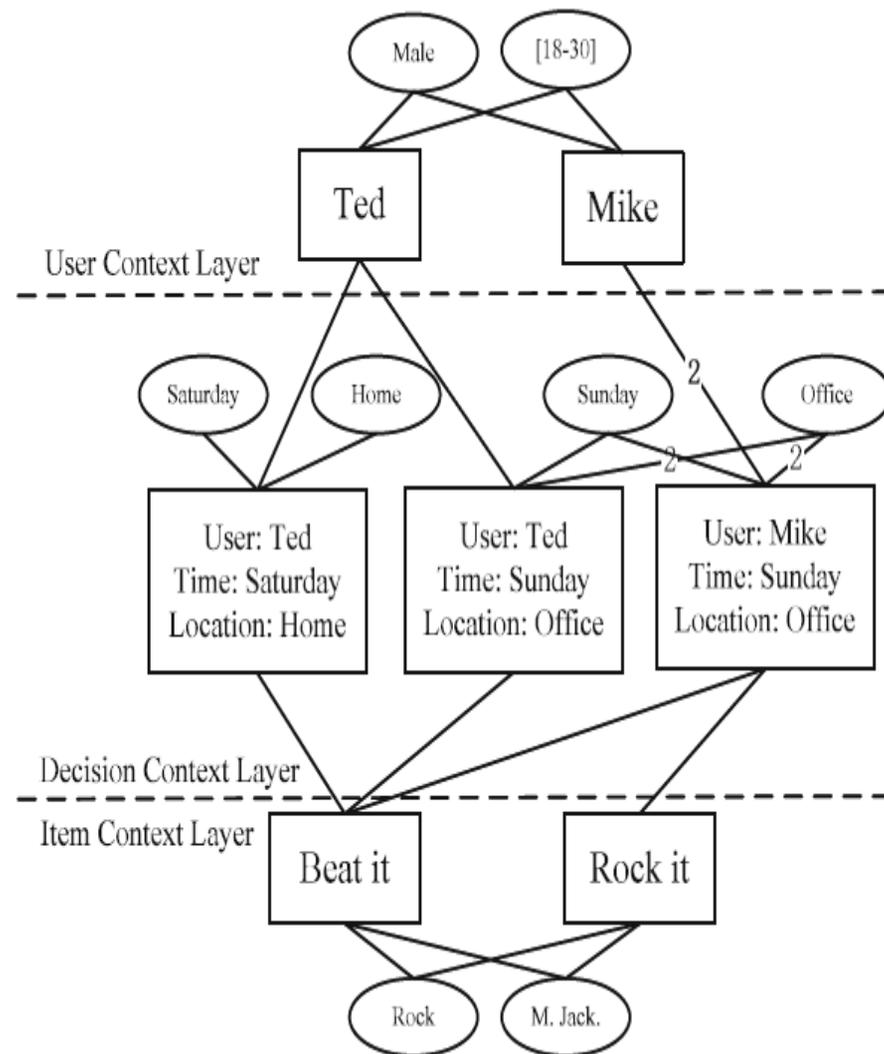
A Graph-based model for context-aware recommendation using implicit feedback data.

Yao, W., He, J., Huang, G., Cao, J., & Zhang, Y.

› Problema a resolver.

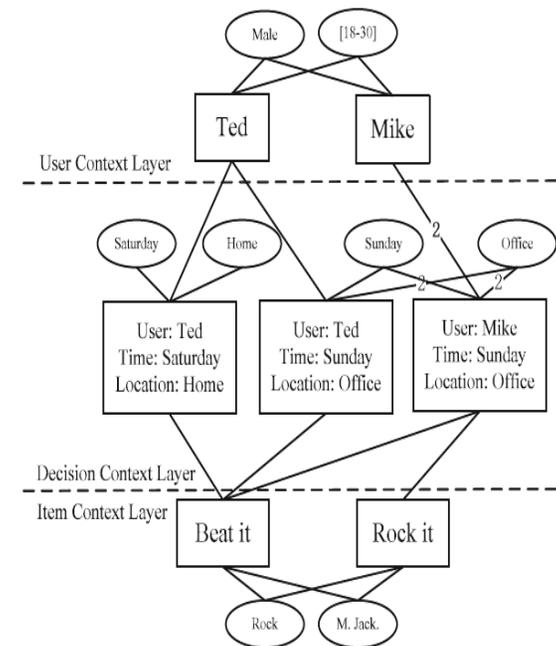
- Recomendación cuando se cuenta sólo con feedback implícito.
- Contexto del usuario, del ítem y de la decisión.
- Framework genérico de recomendación basado en grafos.

› Modelo Multi Layer Context Graph (MLCG)



› Tareas.

1. Construcción del grafo MLCG.
2. Asignación de pesos de los enlaces.
3. Generación del ranking personalizado (usuario, contexto).



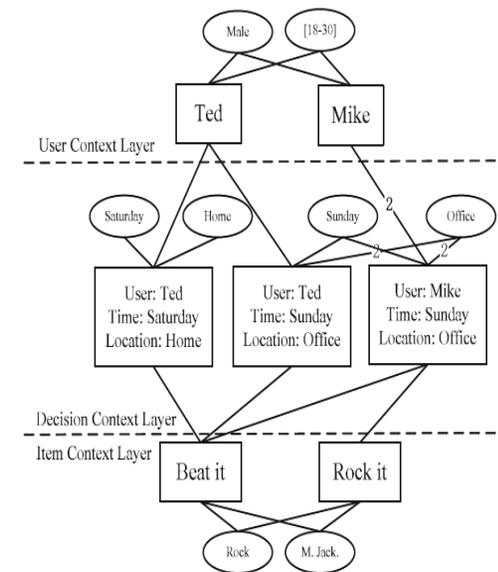
› 1. Modelo Multi Layer Context Graph (MLCG)

Algorithm 1 Construct a Multi-Layer Context Graph

Input: Set of users U and context C_U ; Set of Items I and context C_I ; Decision Context C_D ;
 Log record table $LogTable$, where each log is in the form of $\langle u, c_{d_1}, c_{d_2}, \dots, c_{d_{|C_D|}}, i \rangle$,
 $u \in U$, $i \in I$ and $c_{d_k} \in C_{D_k}$

Output: Multi-Layer Context Graph \mathcal{G}

- 1: Initialize a graph \mathcal{G} with *USER-Layer*, *ITEM-Layer* and *Decision-Layer*
- 2: $CreateLayer(C_U, U, USER-Layer)$
- 3: $CreateLayer(C_I, I, ITEM-Layer)$
- 4: **for** each context domain $c_d \in C_D$ **do**
- 5: **for** each context value $v \in c_d$ **do**
- 6: Create a decision context node for v on *Decision-Layer*
- 7: **end for**
- 8: **end for**
- 9: **for** each log record $log \in LogTable$ **do**
- 10: Create a decision node $v = \langle u, c_{d_1}, c_{d_2}, \dots, c_{d_{|C_D|}} \rangle$ on Decision Layer
- 11: Connect nodes $c_{d_1}, c_{d_2}, \dots, c_{d_{|C_D|}}$ and v
- 12: Connect user node u and v
- 13: Connect item node i node and v
- 14: **end for**
- 15: Return \mathcal{G}

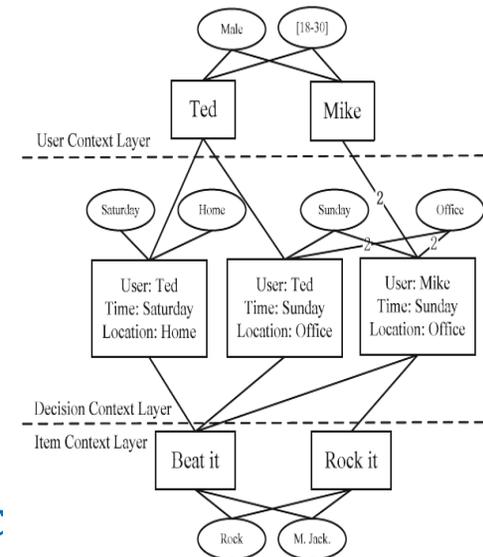


- › 2.1 Asignación de pesos a los enlaces
- › dentro de la misma capa.

$$w(j, k) = \begin{cases} \frac{\alpha}{n_c(j)} \frac{f(j, k)}{\sum_{t \in N_s(j)} f(j, t)} & \text{if } |N_d(j)| > 0 \\ \frac{co-occu(j, k)}{\sum_{t \in N_s(j)} co-occu(j, t)} & \text{if } |N_d(j)| = 0 \\ 0 & \text{otherwise,} \end{cases}$$

Do

- ,
- α es un parámetro.
- $n_c(j)$ representa la cantidad de elementos c
- $N_s(j)$ es el conjunto de nodos de la misma capa de j
- $N_d(j)$ es el resto de nodos conectados con j
- $f(j, k)$ es una función que mide la importancia de la relación.

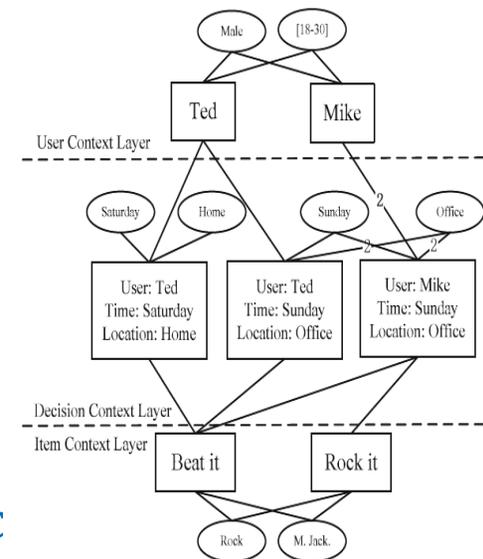


- › 2.2 Asignación de pesos a los enlaces
- › entre capas.

$$w(j, k) = \begin{cases} \frac{(1-\alpha)co-occu(j,k)}{\sum_{t \in N_d(j)} co-occu(j,t)} & \text{if } |N_s(j)| > 0 \\ \frac{co-occu(j,k)}{\sum_{t \in N_d(j)} co-occu(j,t)} & \text{if } |N_s(j)| = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Donde:

- j es un nodo
- α es un parámetro.
- $n \downarrow c(j)$ representa la cantidad de elementos c
- $N \downarrow s(j)$ es el conjunto de nodos de la misma capa de j
- $N \downarrow d(j)$ es el resto de nodos conectados con j



› 3. Cálculo de Ranking (Page Rank).

$$\mathbf{PR}(t + 1) = \mu \cdot \mathbf{M} \cdot \mathbf{PR}(t) + (1 - \mu) \cdot \mathbf{d},$$

Donde:

- M es la matriz de probabilidad de transición.
- μ es el factor de damping (normalmente 0.85).
- d es un vector, tal que $d_k = 1/n$, $k = 1..n$; n es el número de nodos.

Problema: Page Rank **no es personalizado, no depende del contexto.**

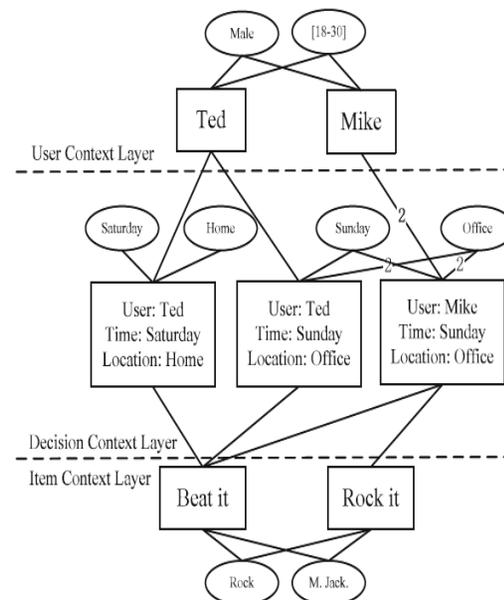
Context-aware Personalized Random Walk (CPRW)

$$\tilde{d}_j = \begin{cases} \frac{\lambda}{|\Phi|} & \text{if node } j \in \Phi \\ \frac{1-\lambda}{|c_d|+1} & \text{if node } j \in c_d \text{ or node } j = u \\ 0 & \text{otherwise,} \end{cases}$$

$$\mathbf{PR}(0)_j = \begin{cases} 1 & \text{if node } j \in c_d \text{ or node } j = u \\ 0 & \text{otherwise.} \end{cases}$$

Donde:

- $\mathbf{PR}(t+1) = \mu \cdot \mathbf{M} \cdot \mathbf{PR}(t) + (1 - \mu) \cdot \tilde{\mathbf{d}}$ sea hacer la recor
- c_d es el contexto de decisión.
- λ es un parámetro de ajuste del peso del contexto & preferencias.
- Φ es el conjunto de ítemes que el usuario u compró previamente.



› Context-aware Personalized Random Walk (CPRW)

Algorithm 2 Context-aware Personalized Random Walk (CPRW)

Input: Transition probability matrix M for MLCG \mathcal{G} , user u

Output: Top- K items.

¿Condición de parada?

1: Construct $\tilde{\mathbf{d}}$ for the active user u based on Equation 5

2: Initialize $\mathbf{PR}(0)$ based on Equation 6

3: $t \leftarrow 1$

4: **while** $|\mathbf{PR}(t) - \mathbf{PR}(t-1)| < \varepsilon$ **do**

5: $\mathbf{PR}(t+1) = \mu \cdot \mathbf{M} \cdot \mathbf{PR}(t) + (1 - \mu) \cdot \tilde{\mathbf{d}}$

6: **end while**

7: Rank item nodes based on \mathbf{PR} values

8: Return Top- K items



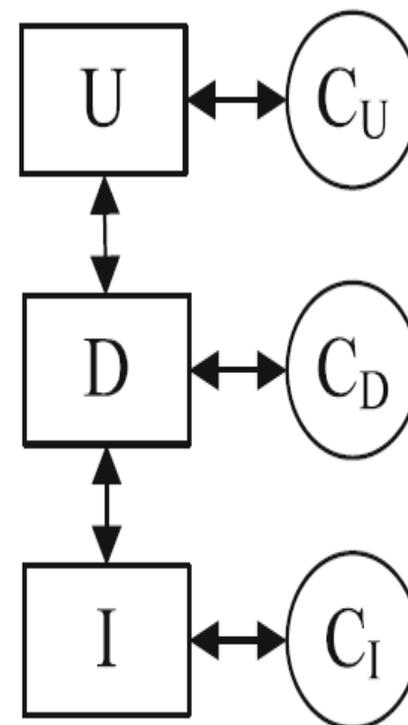
- › Semantic Path-Based Random Walk
- › (SPRW)

El Path Semántico representa el proceso de decisión, permite “dirigir la energía”, actúa como restricción de los caminos aleatorios.

Ejemplos:

Búsqueda del usuario: U – D – I - C_{UI} - I

Ítemes en promoción: I - C_{UI} - I - D - U



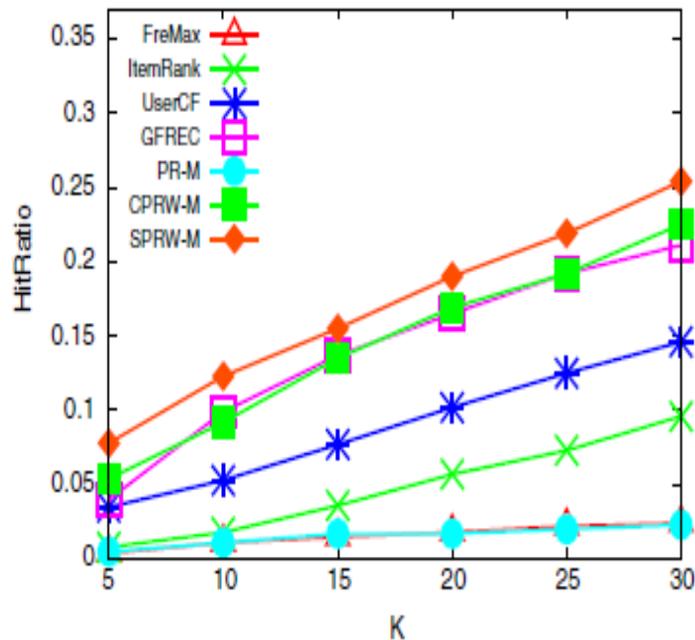
$$PR(t+1) = \underbrace{\mu \cdot M \cdot PR(t) + (1 - \mu - \nu) \cdot \tilde{d}}_{\text{CPRW}} + \nu \sum_{k=1}^{|\mathcal{P}|} w_k M_{\mathcal{P}_k} \cdot PR(t),$$

Matrices de transición de cada paso del Path Semántico

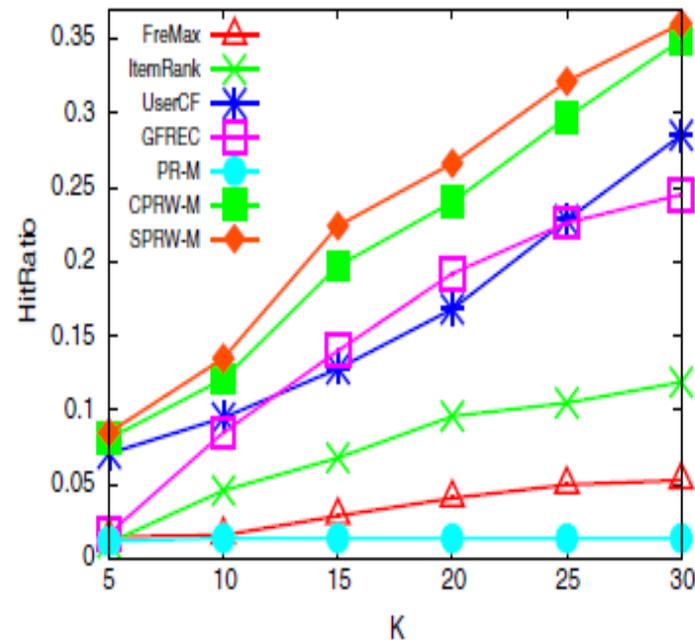
› Evaluación

- Datasets: Last.fm y CiteULike
- Métricas utilizadas:
 - HitRatio@K
 - Mean Reciprocal Rank (MRR@K)
 - Recall@K
- Métodos contra los que se compararon resultados:
 - **FreMax**: Popularidad.
 - **UserCF**: Filtrado Colaborativo UB.
 - **ItemRank**: Random Walk sobre grafo de ítemes.
 - **GFREC**: Método basado en grafo bipartito, similar al presentado al inicio.
 - **PageRank-M** (PR-M): Algoritmo original de Page Rank.

› Evaluación – HitRatio@K

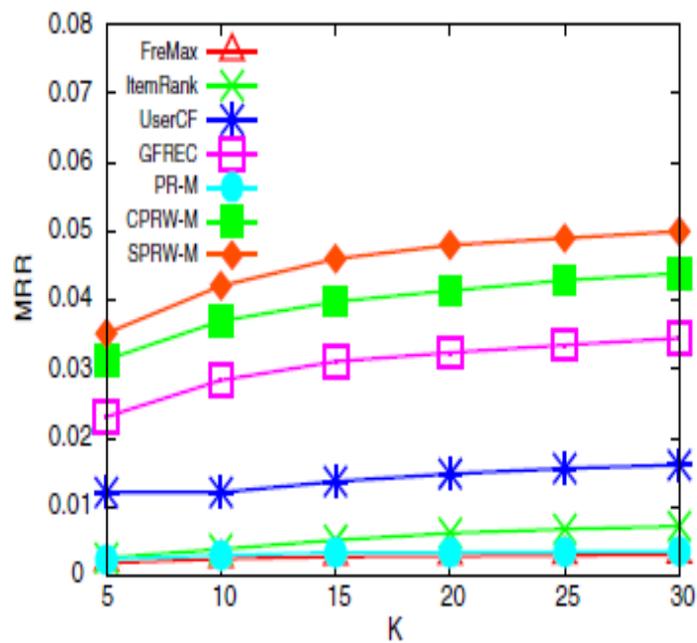


(a) Last.fm

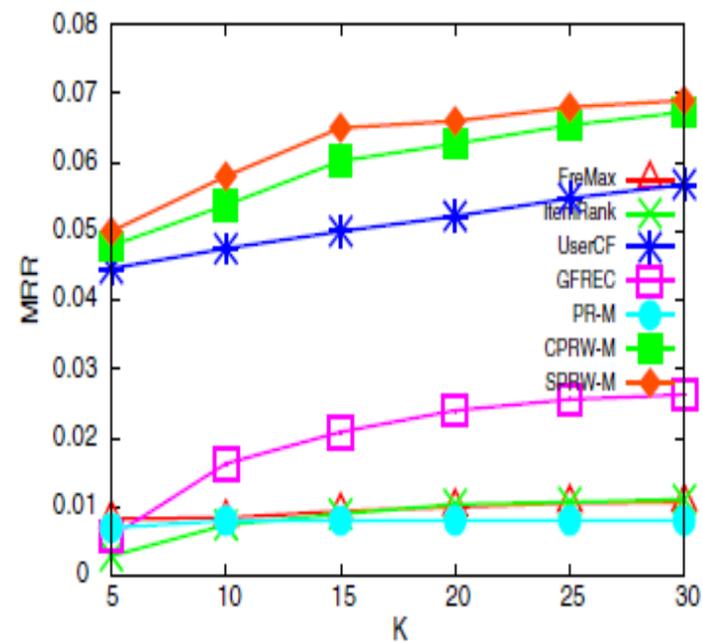


(b) CiteULike

› Evaluación – MRR@K

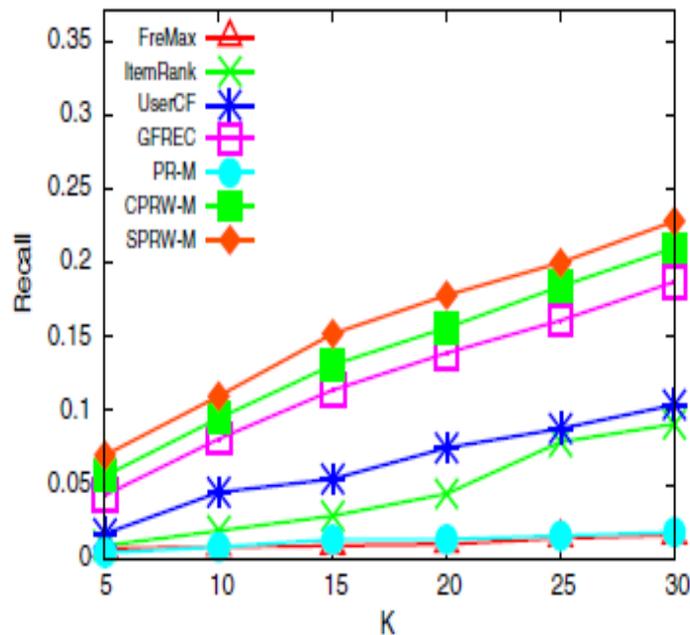


(a) Last.fm

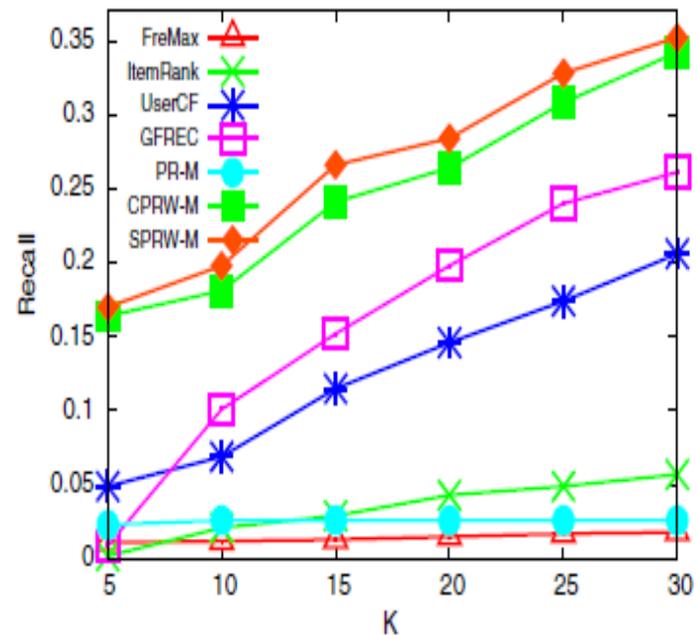


(b) CiteULike

› Evaluación – Recall@K



(a) Last.fm



(b) CiteULike

CONCLUSIONES Y COMENTARIOS

- **Primer modelo** presentado (2 capas) **intenta ser general**, pero fue probado sólo en un contexto específico (biblioteca virtual de Taiwán).
- **Segundo modelo** incorpora explícitamente el **contexto** de la decisión.
- **Modelo SPRW** (Semantic Path-Based Random Walk) se ve **promisorio**, por las posibilidades de ajustarlo al problema y por los resultados obtenidos, no obstante **no se analiza su escalabilidad**.
- Aun cuando el segundo artículo entrega **parámetros** utilizados en las pruebas, **no da directrices** respecto de cómo fijarlos ni hace un análisis de sensibilidad de éstos.
- Ambos artículos tienen **errores menores** en fórmulas y algoritmos, que dificultan su comprensión.

MUCHAS GRACIAS!
