

Master Course in Data Mining and Knowledge Management  
Internship Report Presentation

# Cold Start Recommendations: A Non-negative Matrix Factorization Approach

Martin SAVESKI

Yahoo! Labs

July 9, 2013

Supervisors: Amin Mantrach, Yahoo! Labs  
Daniele Quercia, Yahoo! Labs  
Ricard Gavaldà, UPC

# Recommender Systems

- Designed to **suggest** users items of interest
- Widely spread: books, movies, video lectures ...

## Frequently Bought Together



## People who liked this also liked...



## Visitors who watched this lecture also watched...



**Markov Chain Monte Carlo**  
30189 views - Iain Murray, 2009



**Graphical Models**  
7120 views - Zoubin Ghahramani, 2009



**Bayesian or Frequentist, Which Are You?**  
12688 views - Michael I. Jordan, 2009



**Machine Learning, Probability and Graphical Models**  
38118 views - Sam Roweis, 2006

# Two Main Types of Recommender Systems

## Content-Based (CB):

- Use the properties of the items
- Build user profiles from the properties of the items liked before
- Suggest items to users with most similar profiles

## Collaborative filtering (CF):

- Use the preferences of the community (wisdom of the crowd)
- Recommend items that users with similar tastes like

# Two Main Types of Recommender Systems

## Content-Based (CB):

- Use the properties of the items
- Build user profiles from the properties of the items liked before
- Suggest items to users with most similar profiles

## Collaborative filtering (CF):

- Use the preferences of the community (wisdom of the crowd)
- Recommend items that users with similar tastes like

CF systems achieve the state-of-the-art performance  
(Tuzhilin & Adomavicius)

# The Cold Start

When a new item/user enters the system and no past information is available, no effective recommendations can be produced

# The Cold Start

When a new item/user enters the system and no past information is available, no effective recommendations can be produced

## Item cold-start

- No previous ratings are available
- Collaborative filtering is not an option

**Common solution:** Back-off to content-based recommendations

## User cold-start

- Visits from users who are not logged in
- Neither Content-Based, nor Collaborative Filtering is applicable

**Common solution:** Recommend the top- $k$  items (one-size-fits-all)

# The Cold Start

When a new item/user enters the system and no past information is available, no effective recommendations can be produced

## Item cold-start

- No previous ratings are available
- Collaborative filtering is not an option

**Common solution:** Back-off to content-based recommendations

## User cold-start

- Visits from users who are not logged in
- Neither Content-Based, nor Collaborative Filtering is applicable

**Common solution:** Recommend the top- $k$  items (one-size-fits-all)

**Practical Importance:**

Hundreds of new items and millions of visitors every day

# The Cold Start: Challenges

Research Questions:

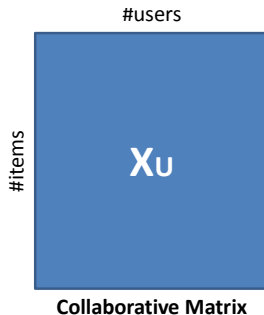
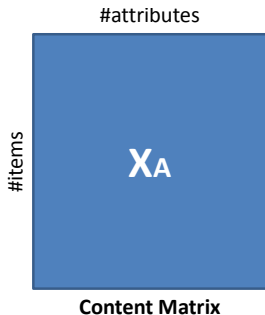
**Item cold-start** Can we combine the content and the collaborative information to outperform pure content-based recommenders?

**User cold-start** Can we use the context of the users, *i.e.*, location and time to perform better than the top- $k$  recommendations?



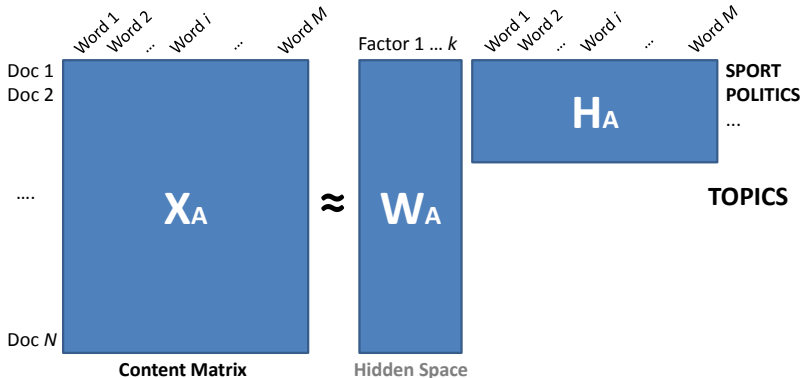
# Item Cold-start

# Data in Matrix Form



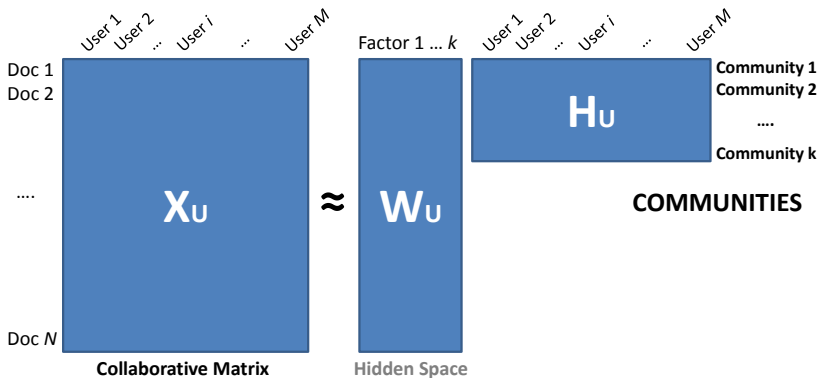
- Both are non-negative matrices
- New item corresponds to an empty row in  $X_U$

# Non-negative Matrix Factorization (NMF)



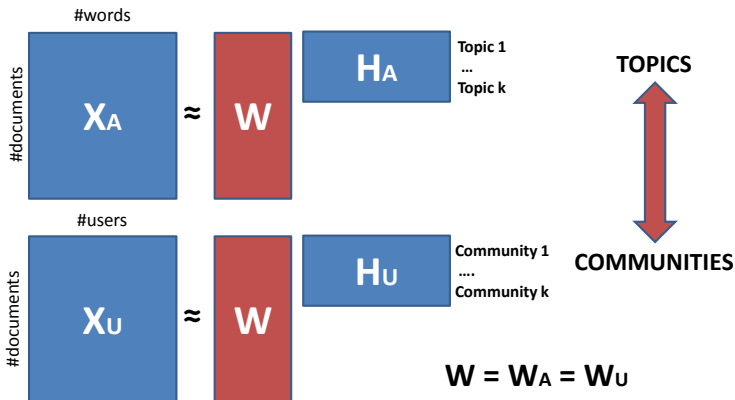
- NMF decomposes  $X_A$  in two lower-rank matrices  $W_A$  and  $H_A$

# Non-negative Matrix Factorization (NMF)



- Many CF approaches perform UV decomposition of  $X_U$  (NMF without non-negativity constraints)

# Joint Non-negative Matrix Factorization (JNMF)



At test time, given the properties of a new item,  $q_A$ :

- Project  $q_A$  in  $w$ , using  $H_A$
- Infer  $q_u$  as:  $q_u = wH_U$

# JNMF: Optimization

Non-convex optimization problem:

$$\min : J = \underbrace{\frac{1}{2}(\alpha \|\mathbf{X}_A - \mathbf{W}\mathbf{H}_A\|_F^2)}_{\text{Factorization of } \mathbf{X}_A} + \underbrace{(1 - \alpha) \|\mathbf{X}_U - \mathbf{W}\mathbf{H}_U\|_F^2}_{\text{Factorization of } \mathbf{X}_U} + \underbrace{\lambda (\|\mathbf{W}\|_F^2 + \|\mathbf{H}_A\|_F^2 + \|\mathbf{H}_U\|_F^2)}_{\text{Tikhonov Regularization}}$$

*s.t.*    $\mathbf{W} \geq 0, \mathbf{H}_A \geq 0, \mathbf{H}_U \geq 0$

Hyper-parameters:

- $k$ : number of factors
- $\alpha$ : importance of each factorization
- $\lambda$ : smoothness of the solution

Optimization Algorithms:

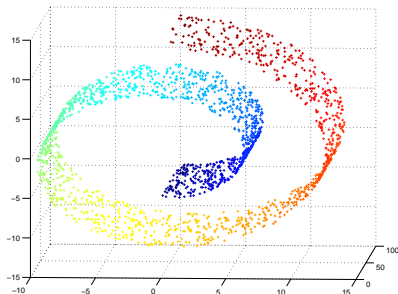
- ① Multiplicative Update Rules
- ② Alternating Least Squares with projections to 0

# Joint NMF with Graph Regularization (JNMF-GR)

**Assumption:** if two points  $x_i$  and  $x_j$  are close in the NN graph, then their latent representations  $w_i$  and  $w_j$  should also be close.

$$S = \frac{1}{2} \sum_{i,j=1}^n \|w_i - w_j\|^2 [\mathbf{NN}]_{ij} = \text{trace}(\mathbf{W}^T \mathbf{L} \mathbf{W})$$

Controlled with additional hyper-parameter



# Evaluation: Experimental Setup

Tasks:

- Item cold-start (implicit feedback)
- Email recipient prediction
- Author prediction

Datasets:

	Type	#documents	#users	vocabulary	Span
<b>Yahoo! News</b>	articles&comments	41K	650K	60K	40 days
<b>Enron</b>	email&recipients	36K	5K	12K-56K	10 mailboxes
<b>NIPS</b>	papers&authors	1.7K	2K	13K	13 years



# Evaluation: Experimental Setup (cont.)

## Baselines:

- Content Based Recommender (CB)
- Content Topic Based Recommender (CTB)
- Author Topic Model (ATM) - learns users' topical profiles

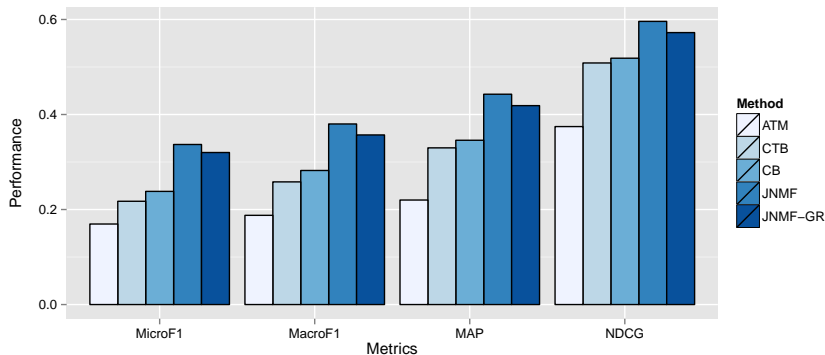
## Metrics:

- Macro and Micro F1
- Mean Average Precision (MAP)
- Normalized Discounted Cumulative Gain (NDCG)
- Ranking accuracy (recall oriented, suitable for implicit feedback)

## Protocol:

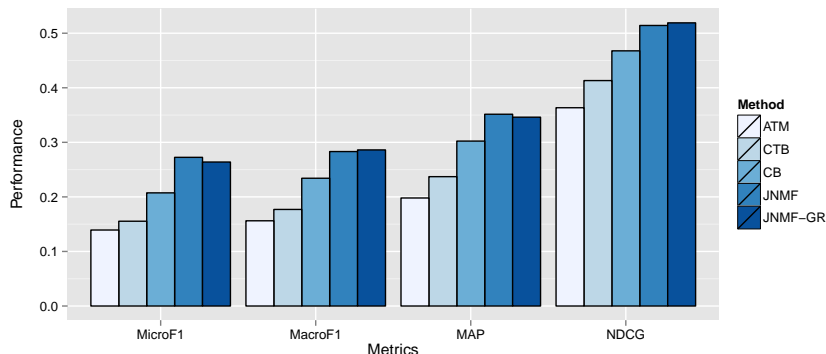
- Divide the data sets in chronological train/test folds
- Hyper-parameters tuned on independent set (part of train set)
- Differences evaluated with paired  $t$ -test

# Results: Author Prediction (NIPS)



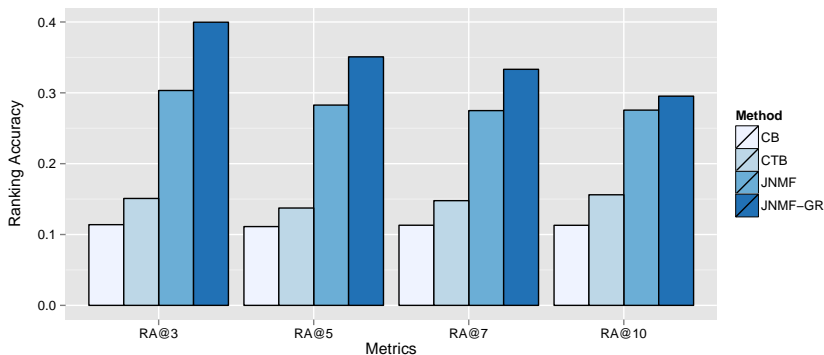
$JNMF \approx JNMF-GR > CB > CTB > ATM$

# Results: Email Recipient Prediction (Enron)



$JNMF \approx JNMF-GR > CB > CTB > ATM$

# Results: Item cold-start (Yahoo! News)



JNMF-GR > JNMF > CTB > CB

# User Cold-start

# User cold-start

No user history available  $\Rightarrow$  top- $k$  recommendations

# User cold-start

No user history available  $\Rightarrow$  top- $k$  recommendations

User context (IP address derived):

- Location
- Time

# User cold-start

No user history available  $\Rightarrow$  top- $k$  recommendations

User context (IP address derived):

- Location
- Time

Geography of user engagement in news platforms

- Specific locations  $\Leftrightarrow$  specific interests?
- Use these patterns to partially overcome the user cold-start?



# User cold-start

No user history available  $\Rightarrow$  top- $k$  recommendations

User context (IP address derived):

- Location
- Time

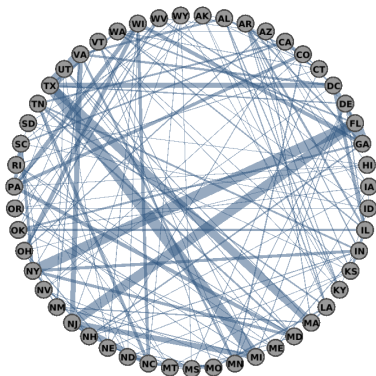
Geography of user engagement in news platforms

- Specific locations  $\Leftrightarrow$  specific interests?
- Use these patterns to partially overcome the user cold-start?

Yahoo! News US

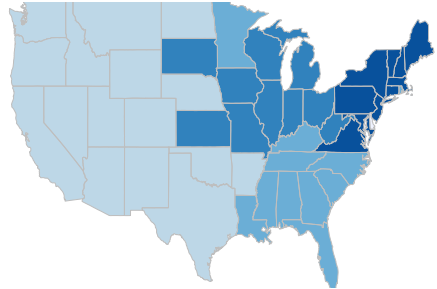
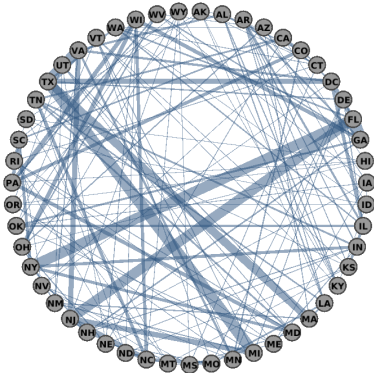
- 2 years
- 200K articles
- 41M comments

# State Commenting Graph



$\text{weight}(i, j)$ : # of times users  
from state  $i$  and  $j$  commented  
on the same article

# State Commenting Graph



$\text{weight}(i, j)$ : # of times users  
from state  $i$  and  $j$  commented  
on the same article

Four clusters of like-minded states

# The Time Zone Effect

## Hypothesis:

Users in the same time zone preferentially engage with the same articles

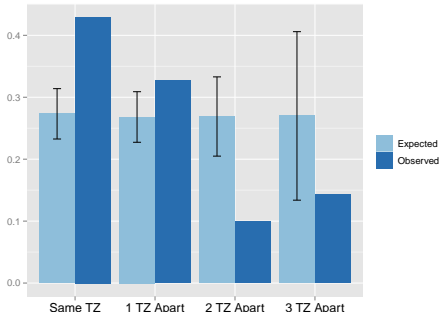
# The Time Zone Effect

## Hypothesis:

Users in the same time zone preferentially engage with the same articles

## Test:

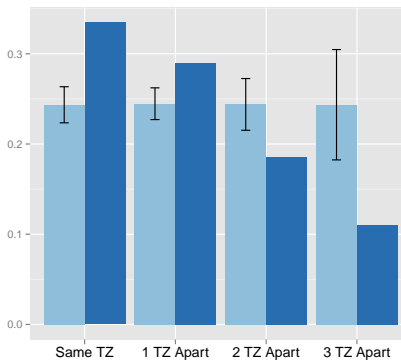
- 1 Measuring engagement  $k$  time zones apart
- 2 Random model: shuffling the time zone assignments
- 3 Comparison:



# What Makes the Time Zone Effect

## ① Interests in different topics

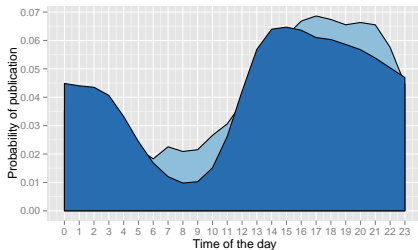
- Users in the same time zone are interested in the same topics



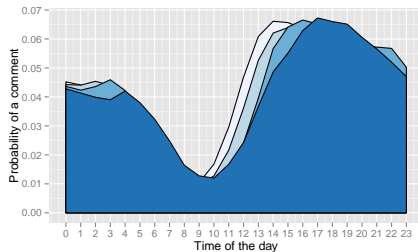
# What Makes the Time Zone Effect

## ② Better alignment with the publishing cycle

- Commenting cycles of some time zones meet the publishing cycle



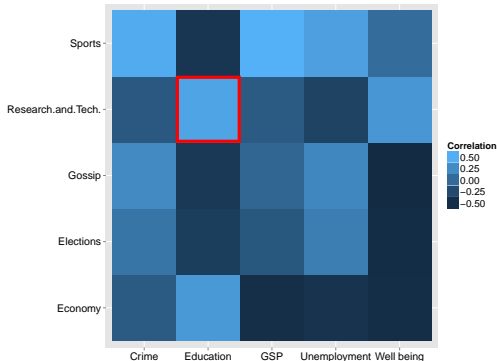
Source: Articles Comments



Time Zone: Eastern Central Mountain Pacific

# What Makes the Time Zone Effect

- ③ Differences in the socio-economical status of the users
  - E.g. Education  $\Rightarrow$  Research&Technology





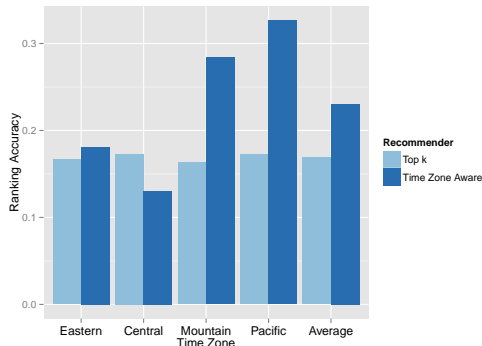
# Time-zone-aware recommender for user cold-start

**Recommend what is most popular in user's time zone**

**Evaluation:** top- $k$  vs. time-zone-aware recommendations

Subset of 1 month:

- >6M comments
- >35K articles
- >500K users



# To sum up

## Item cold-start

- Proposed two hybrid recommenders based on NMF
- Combining content + collaborative information outperforms pure content based recommendations

## User cold-start

- Studied the geography of user engagement in news platforms
- Making time-zone-aware recommendations leads to improved recommendation accuracy

# Future Work

## Time aware extension of the hybrid models

- Our tastes change over time

# Future Work

## Time aware extension of the hybrid models

- Our tastes change over time

## Exploiting power law behaviours

- Both words and users are power law distributed

# Future Work

## Time aware extension of the hybrid models

- Our tastes change over time

## Exploiting power law behaviours

- Both words and users are power law distributed

## Decoupling the geographic locality effect

- How much the time zone effect is solely due to time zones?

# Future Work

## Time aware extension of the hybrid models

- Our tastes change over time

## Exploiting power law behaviours

- Both words and users are power law distributed

## Decoupling the geographic locality effect

- How much the time zone effect is solely due to time zones?

# References

- Alexander Tuzhilin & Gediminas Adomavicius. *Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions*. IEEE Transactions on Knowledge and Data Engineering, 2005.

# Cold Start Recommendations: A Non-negative Matrix Factorization Approach

Martin SAVESKI

Yahoo! Labs

July 9, 2013





## Related Work

**Hybrid recommender:** Content-Boosted Collaborative Filtering

- Enriches the collaborative profiles by adding content-based scores
- In cold-start boils down to content-based recommendations

**Author Topic Model:**

- Extension of LDA
- Learns topical profiles for each author

**We compare with the ATM in the evaluation**

# Joint NMF: Multiplicative Update Rules

$$[\mathbf{W}]_{ij} \leftarrow [\mathbf{W}]_{ij} \cdot \frac{[\alpha \mathbf{X}_A \mathbf{H}_s^\top + (1 - \alpha) \mathbf{X}_U \mathbf{H}_u^\top]_{ij}}{[\alpha \mathbf{W} \mathbf{H}_A \mathbf{H}_s^\top + (1 - \alpha) \mathbf{W} \mathbf{H}_U \mathbf{H}_u^\top + \lambda_W \mathbf{W}]_{ij}},$$

$$[\mathbf{H}_A]_{ij} \leftarrow [\mathbf{H}_A]_{ij} \cdot \frac{[\alpha \mathbf{W}^\top \mathbf{X}_A]_{ij}}{[\alpha \mathbf{W}^\top \mathbf{W} \mathbf{H}_A + \lambda_{H_s} \mathbf{H}_A]_{ij}},$$

$$[\mathbf{H}_U]_{ij} \leftarrow [\mathbf{H}_U]_{ij} \cdot \frac{[(1 - \alpha) \mathbf{W}^\top \mathbf{X}_U]_{ij}}{[(1 - \alpha) \mathbf{W}^\top \mathbf{W} \mathbf{H}_U + \lambda_{H_u} \mathbf{H}_U]_{ij}}.$$

# Joint NMF: Alternating Least Squares

**Require:**  $k, \alpha, \lambda_W, \lambda_{H_s}, \lambda_{H_u}$

- 1: Initialize  $\mathbf{W}$  with random positive values
- 2: **repeat**
- 3:   Solve for  $\mathbf{H}_A$ :  $\alpha \mathbf{W}^T \mathbf{X}_A = (\alpha \mathbf{W}^T \mathbf{W} + \lambda_{H_s} \mathbf{I}) \mathbf{H}_A$
- 4:   Set all negative elements in  $\mathbf{H}_A$  to 0
- 5:   Solve for  $\mathbf{H}_U$ :  $(1 - \alpha) \mathbf{W}^T \mathbf{X}_U = [(1 - \alpha) \mathbf{W}^T \mathbf{W} + \lambda_{H_u} \mathbf{I}] \mathbf{H}_U$
- 6:   Set all negative elements in  $\mathbf{H}_U$  to 0
- 7:   Solve for  $\mathbf{W}$ :  $\alpha \mathbf{X}_A \mathbf{H}_s^T + (1 - \alpha) \mathbf{X}_U \mathbf{H}_u^T = \mathbf{W} [\alpha \mathbf{H}_A \mathbf{H}_s^T + (1 - \alpha) \mathbf{H}_U \mathbf{H}_u^T + \lambda_W \mathbf{I}]$
- 8:   Set all negative elements in  $\mathbf{W}$  to 0
- 9: **until** stopping condition

# JNMF with Graph Regularization: Details

New optimization problem:

$$\begin{aligned} \min : J = & \frac{1}{2}(\alpha \|\mathbf{X}_A - \mathbf{W}\mathbf{H}_A\|_F^2 + (1 - \alpha) \|\mathbf{X}_U - \mathbf{W}\mathbf{H}_U\|_F^2 + \beta \text{trace}(\mathbf{W}^T \mathbf{L} \mathbf{W}) \\ & + \lambda (\|\mathbf{W}\|_F^2 + \|\mathbf{H}_A\|_F^2 + \|\mathbf{H}_U\|_F^2) \\ \text{s.t. } & \mathbf{W} \geq 0, \mathbf{H}_A \geq 0, \mathbf{H}_U \geq 0 \end{aligned}$$

Additional hyper-parameters:

- $\beta$ : extent to which locality is imposed
- $p$ : number of nearest neighbours

Optimization Algorithms:

- 1 Multiplicative Update Rules
- 2 Alternating Least Squares with projections to 0
  - Updating  $\mathbf{W}$  requires solving Sylvester Equation

## JNMF-GR: Multiplicative Update Rules

$$[\mathbf{W}]_{ij} \leftarrow [\mathbf{W}]_{ij} \cdot \frac{[\alpha \mathbf{X}_A \mathbf{H}_s^\top + (1 - \alpha) \mathbf{X}_U \mathbf{H}_u^\top + \beta \mathbf{A} \mathbf{W}]_{ij}}{[\alpha \mathbf{W} \mathbf{H}_A \mathbf{H}_s^\top + (1 - \alpha) \mathbf{W} \mathbf{H}_U \mathbf{H}_u^\top + \beta \mathbf{D} \mathbf{W} + \lambda_W \mathbf{W}]_{ij}}$$

$$[\mathbf{H}_A]_{ij} \leftarrow [\mathbf{H}_A]_{ij} \cdot \frac{[\alpha \mathbf{W}^\top \mathbf{X}_A]_{ij}}{[\alpha \mathbf{W}^\top \mathbf{W} \mathbf{H}_A + \lambda_{H_s} \mathbf{H}_A]_{ij}}$$

$$[\mathbf{H}_U]_{ij} \leftarrow [\mathbf{H}_U]_{ij} \cdot \frac{[(1 - \alpha) \mathbf{W}^\top \mathbf{X}_U]_{ij}}{[(1 - \alpha) \mathbf{W}^\top \mathbf{W} \mathbf{H}_U + \lambda_{H_u} \mathbf{H}_U]_{ij}}$$

# JNMF-GR: Alternating Least Squares

**Require:**  $\mathbf{A}$ ,  $k$ ,  $\alpha$ ,  $\beta$ ,  $\lambda_W$ ,  $\lambda_{H_s}$ ,  $\lambda_{H_u}$

- 1: Compute the diagonal matrix  $\mathbf{D}$  as:  $\mathbf{D}_{ii} = \sum_j \mathbf{A}_{ij}$
- 2: Compute the Laplacian matrix  $\mathbf{L}$  as:  $\mathbf{L} = \mathbf{D} - \mathbf{A}$
- 3: Initialize  $\mathbf{W}$  with random positive values
- 4: **repeat**
- 5:   Solve for  $\mathbf{H}_A$ :  $\alpha \mathbf{W}^T \mathbf{X}_A = (\alpha \mathbf{W}^T \mathbf{W} + \lambda_{H_s} \mathbf{I}) \mathbf{H}_A$
- 6:   Set all negative elements in  $\mathbf{H}_A$  to 0
- 7:   Solve for  $\mathbf{H}_U$ :  $(1 - \alpha) \mathbf{W}^T \mathbf{X}_U = [(1 - \alpha) \mathbf{W}^T \mathbf{W} + \lambda_{H_u} \mathbf{I}] \mathbf{H}_U$
- 8:   Set all negative elements in  $\mathbf{H}_U$  to 0
- 9:    $\mathbf{C} = \alpha \mathbf{X}_A \mathbf{H}_s^T + (1 - \alpha) \mathbf{X}_U \mathbf{H}_u^T$
- 10:    $\mathbf{B} = \alpha (\mathbf{H}_A \mathbf{H}_s^T \otimes \mathbf{I}) + (1 - \alpha) (\mathbf{H}_U \mathbf{H}_u^T \otimes \mathbf{I}) + \beta (\mathbf{I} \otimes \mathbf{L}) + \lambda_W \mathbf{I}$
- 11:   Solve for  $\text{vec}(\mathbf{W})$ :  $\text{vec}(\mathbf{C}) = \mathbf{B} \cdot \text{vec}(\mathbf{W})$
- 12:    $\mathbf{W}$ :  $\mathbf{W} = \text{reshape}(\text{vec}(\mathbf{W}), n, k)$
- 13:   Set all negative elements in  $\mathbf{W}$  to 0
- 14: **until** stopping condition

# Percentile-ranking/Ranking accuracy

Recall based measure and is suitable when no reliable feedback:

- $rank_{u,i}$  is 0%, if most desirable article for  $u$
- $rank_{u,i}$  is 100%, if least least desirable

$$\overline{rank} = \frac{\sum_{u,i} comment_{u,i} \cdot rank_{u,i}}{\sum_{u,i} comment_{u,i}}$$

We covert the percentile ranking into ranking accuracy:

$$accuracy = \frac{50\% - \overline{rank}}{50\%}$$

- 1: best/ideal predictions
- 0: random predictions



# Normalized Discounted Cumulative Gain (NDCG)

Based on two assumptions:

- 1 highly relevant items are more useful
- 2 the lower the ranked position of a relevant item the less useful it is for the user

$$NDCG = \frac{DCG}{iDCG}, \quad DCG = rel_1 + \sum_{i=2}^n \frac{rel_i}{\log_2 i},$$

- $rel_i$  is the relevance of the document at rank  $i$
- $iDCG$  is the ideal  $DCG$ - the perfect ranking of the results

# Joint NMF with different hyper-parameters

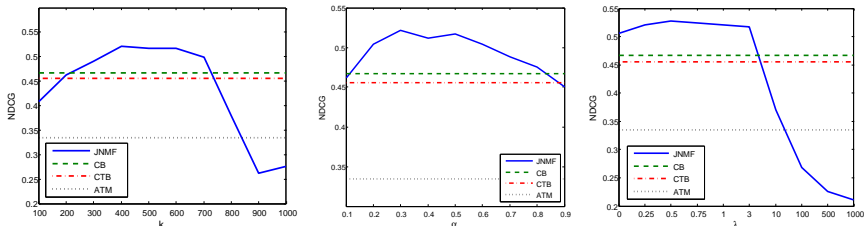


Figure: Behavior of the JNMF hyper-parameters.

# Time Zone Effect. Time Advantage

- Quantifying the advantage  $A$  a time zone might have within the next  $n$  hours
- The probability of commenting in that time zone given the availability of articles:

$$A = \sum_{i=0}^n \sum_{t=0}^{23} (p_t^a \cdot p_{t+i}^c) ,$$

- $p_t^a$  is the probability of a new article being published at time  $t$  (with  $t \in [0, 23]$ )
- $p_{t+i}^c$  is the probability of a comment at time  $t + i$ .