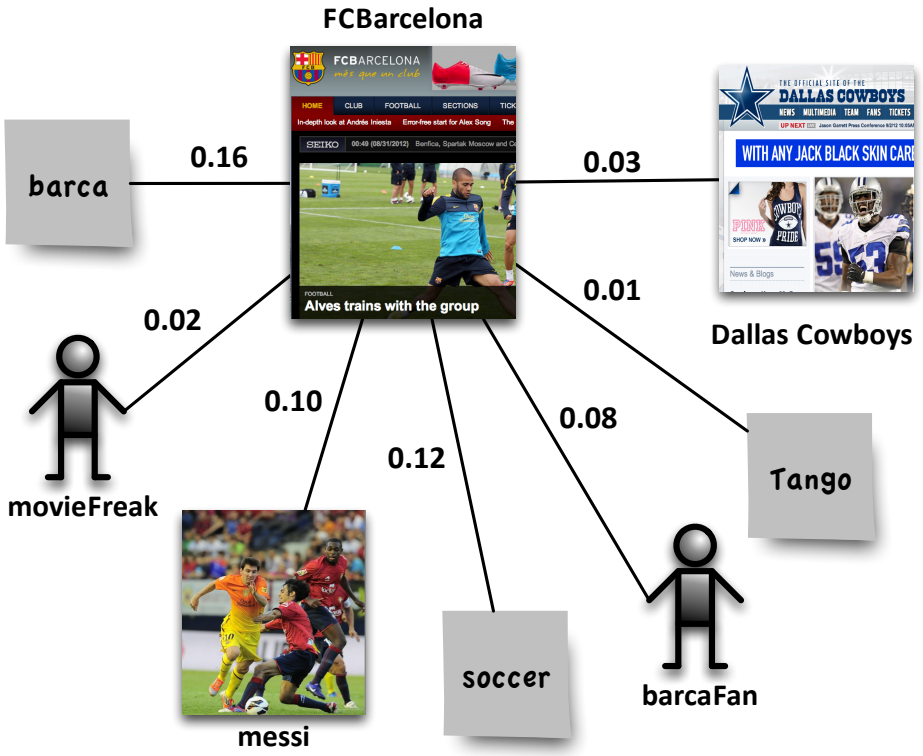
Context Determines Content

An Approach to Resource Recommendation in Folksonomies



Thomas Rodenhausen
Mojisola Anjorin
Renato Domínguez García
Christoph Rensing

Prof. Dr.-Ing. Ralf Steinmetz
KOM - Multimedia Communications Lab

© author(s) of these slides including research results from the KOM research network and TU Darmstadt; otherwise it is specified at the respective slide

Overview

TECHNISCHE UNIVERSITÄT DARMSTADT

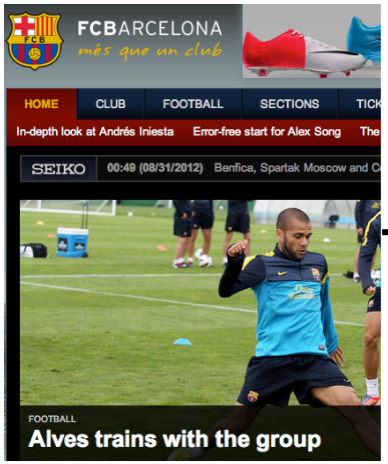
- Motivation
- Context of an Entity in a Folksonomy
- VSScore Framework
- Evaluation Methodology and Metrics
- Results
- Conclusion & Future Work

Challenge: Concept Drift

TECHNISCHE UNIVERSITÄT DARMSTADT

Concept drift is a challenge for graph-based ranking algorithms

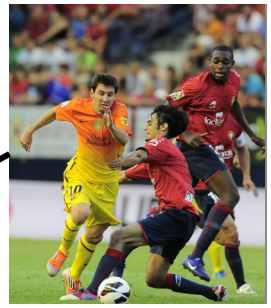
- e.g. Ambiguous tags can cause concept drift as a single tag might represent multiple semantic concepts



FC Barcelona Website

?

?



News about Messi

Dallas Cowboys' Website

Context of an Entity in a Folksonomy

TECHNISCHE UNIVERSITÄT DARMSTADT

Assumption on the Context of an Entity

- The context of an entity in a folksonomy describes the entity well

Assumptions on a Folksonomy's Content and Structure

[adapted from Abel 2011]

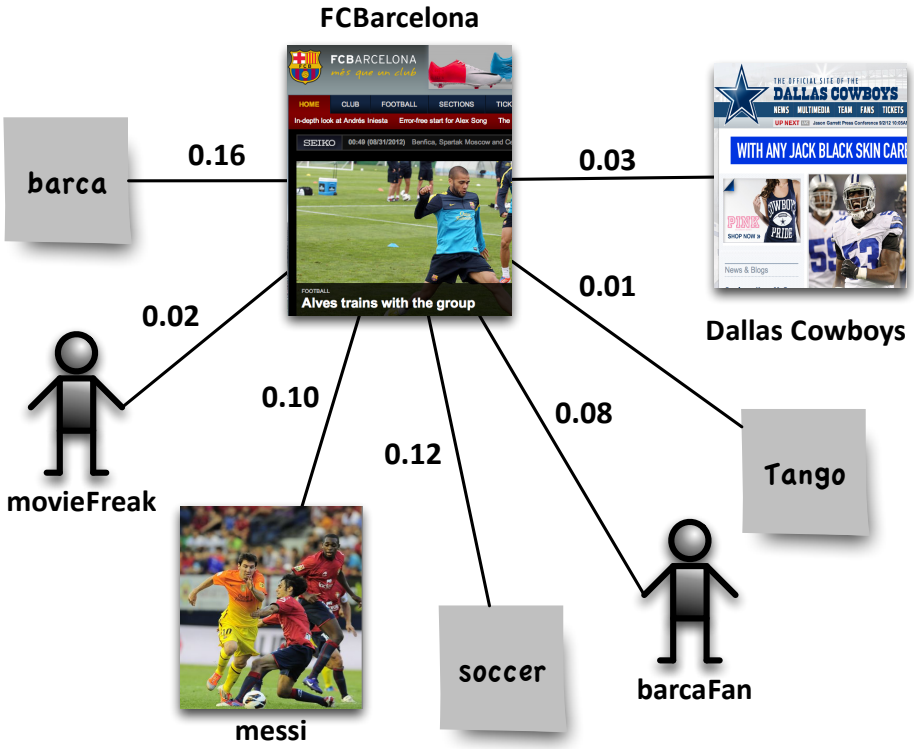
- Tags of a resource describe the resource's content well
- Tags of a user describe the user's interests well
- Resources of a user describe the user's interests well
- Resources of a tag describe the tag's semantic well
- Users of a tag describe the tag's semantic well
- Users of a resource describe the resource's content well

Context of an Entity in a Folksonomy

TECHNISCHE UNIVERSITÄT DARMSTADT

Assumption on the Context of an Entity *e* in a Folksonomy

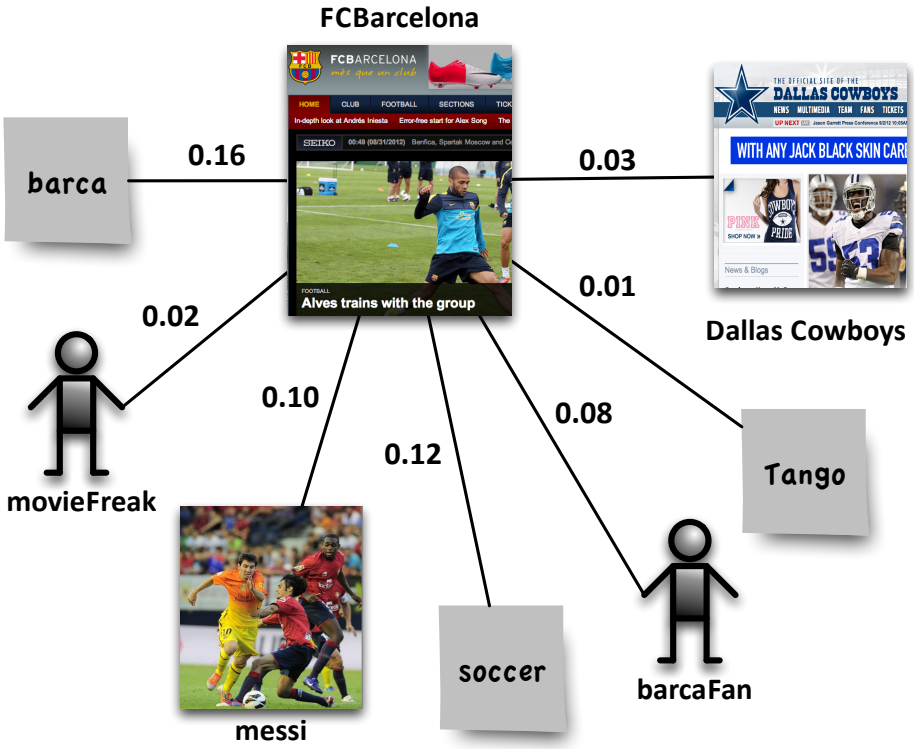
- The context of an entity *e* is given by the strength of relations between *e* and other entities in the folksonomy



Context of an Entity in a Folksonomy

Assumption on the Context of an Entity *e* in a Folksonomy

- A vector $\vec{s}_e$ created with a ranking algorithm (e.g. FolkRank) for a single query entity *e*, describes the relationship between *e* and other entities in the folksonomy well

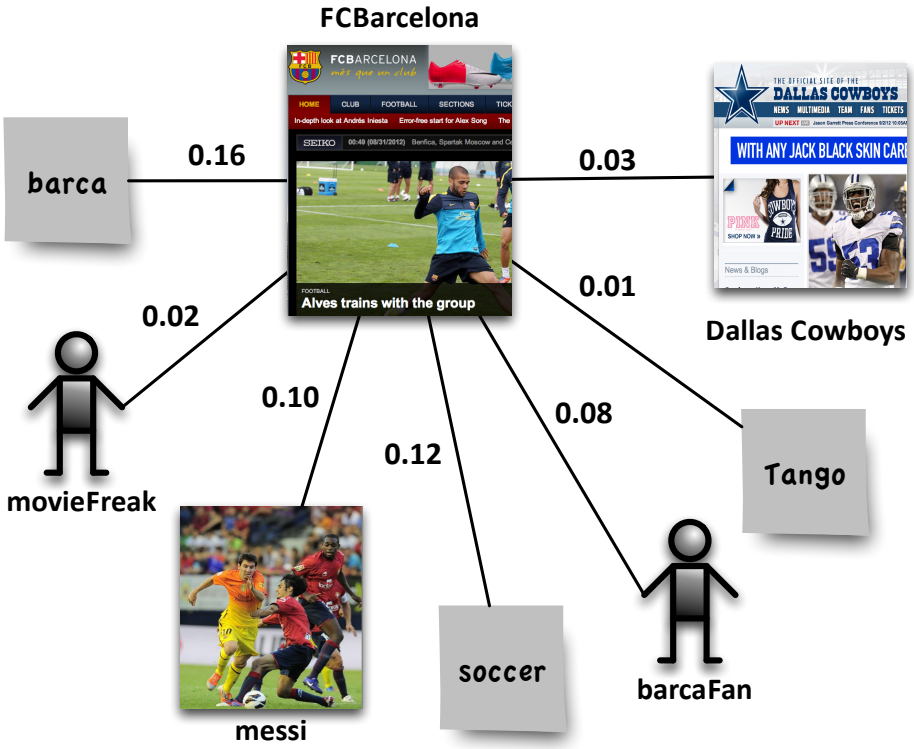


$$\vec{s}_{\text{FCBarcelona}} = \begin{pmatrix} 0.16 \\ 0.02 \\ 0.10 \\ 0.12 \\ 0.08 \\ 0.01 \\ 0.03 \\ \dots \end{pmatrix} \begin{matrix} \text{barca} \\ \text{movieFreak} \\ \text{messi} \\ \text{soccer} \\ \text{barcaFan} \\ \text{Tango} \\ \text{Dallas Cowboys} \\ \dots \end{matrix}$$

Context of an Entity in a Folksonomy

Assumption on the Context of an Entity *e* in a Folksonomy

- A vector $\vec{s}_e$ created with a ranking algorithm (e.g. FolkRank) for a single query entity *e*, describes the relationship between *e* and other entities in the folksonomy well



$$\vec{s}_{\text{FCBarcelona}} = \begin{pmatrix} 0.16 \\ 0.02 \\ 0.10 \\ 0.12 \\ 0.08 \\ 0.01 \\ 0.03 \\ \dots \end{pmatrix} \begin{matrix} \text{barca} \\ \text{movieFreak} \\ \text{messi} \\ \text{soccer} \\ \text{barcaFan} \\ \text{Tango} \\ \text{Dallas Cowboys} \\ \dots \end{matrix}$$

- Entities in a folksonomy represent semantic concepts
- Any entity (e.g. a query entity) can be represented by it's context

Vector Space Score (VSScore)

TECHNISCHE UNIVERSITÄT DARMSTADT

VSScore is a flexible framework which incorporates context-specific information into the recommendation process

- Based on the vector space model

Vector Space Score (VSScore)

TECHNISCHE UNIVERSITÄT DARMSTADT

VSScore is a flexible framework which incorporates context-specific information into the recommendation process

- Based on the vector space model
- Creates a vector representation of semantic concepts for each entity in the folksonomy using a ranking algorithm e.g. FolkRank [Hotho et al. 2006]

Vector Space Score (VSScore)

TECHNISCHE UNIVERSITÄT DARMSTADT

VSScore is a flexible framework which incorporates context-specific information into the recommendation process

- Based on the vector space model
- Creates a vector representation of semantic concepts for each entity in the folksonomy using a ranking algorithm e.g. FolkRank [Hotho et al. 2006]
- Applies the cosine similarity to calculate the distance between these vectors

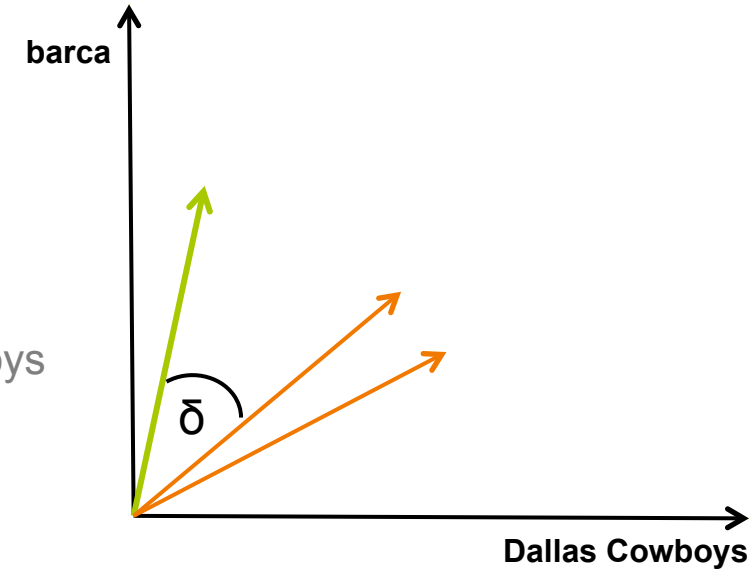
query-entity's context

$$\begin{bmatrix} 0.8 \\ \dots \\ 0.1 \end{bmatrix}$$

resources' contexts

$$\begin{bmatrix} 0.4 \\ \dots \\ 0.6 \end{bmatrix} \quad \begin{bmatrix} 0.3 \\ \dots \\ 0.7 \end{bmatrix}$$

barca
…
Dallas Cowboys



Overview

TECHNISCHE UNIVERSITÄT DARMSTADT

- Motivation
- Context of an Entity in a Folksonomy
- VSScore Framework
- Evaluation Methodology and Metrics
- Results
- Conclusion & Future Work

Evaluation Ranking Tasks

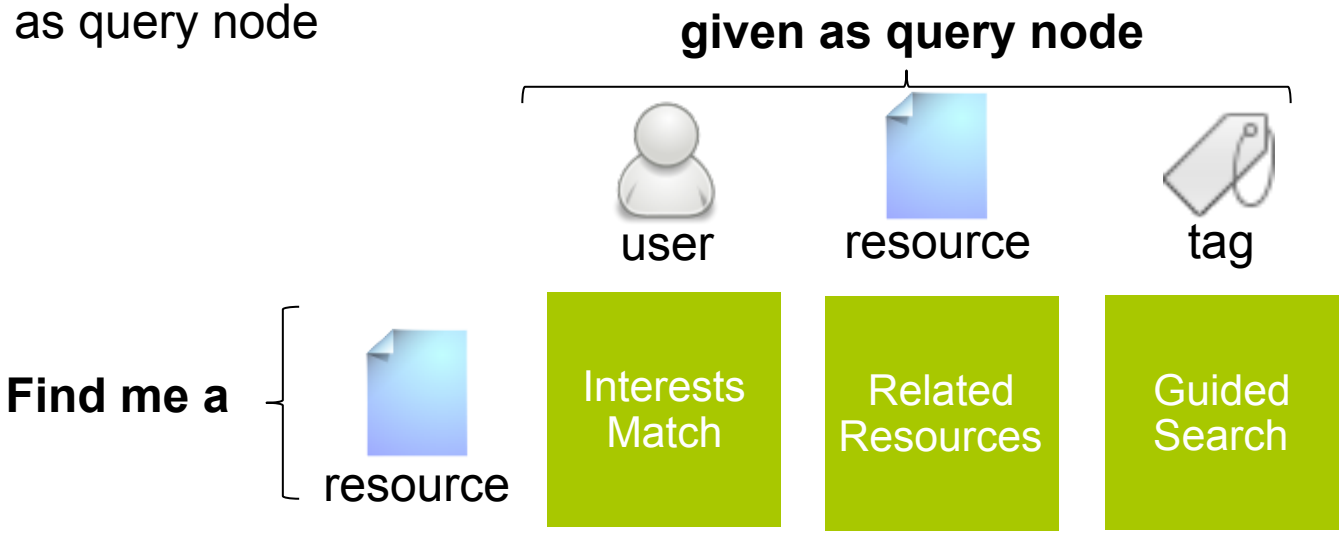
TECHNISCHE UNIVERSITÄT DARMSTADT

Interests Match

- User as query node

Guided Search

- Tag as query node

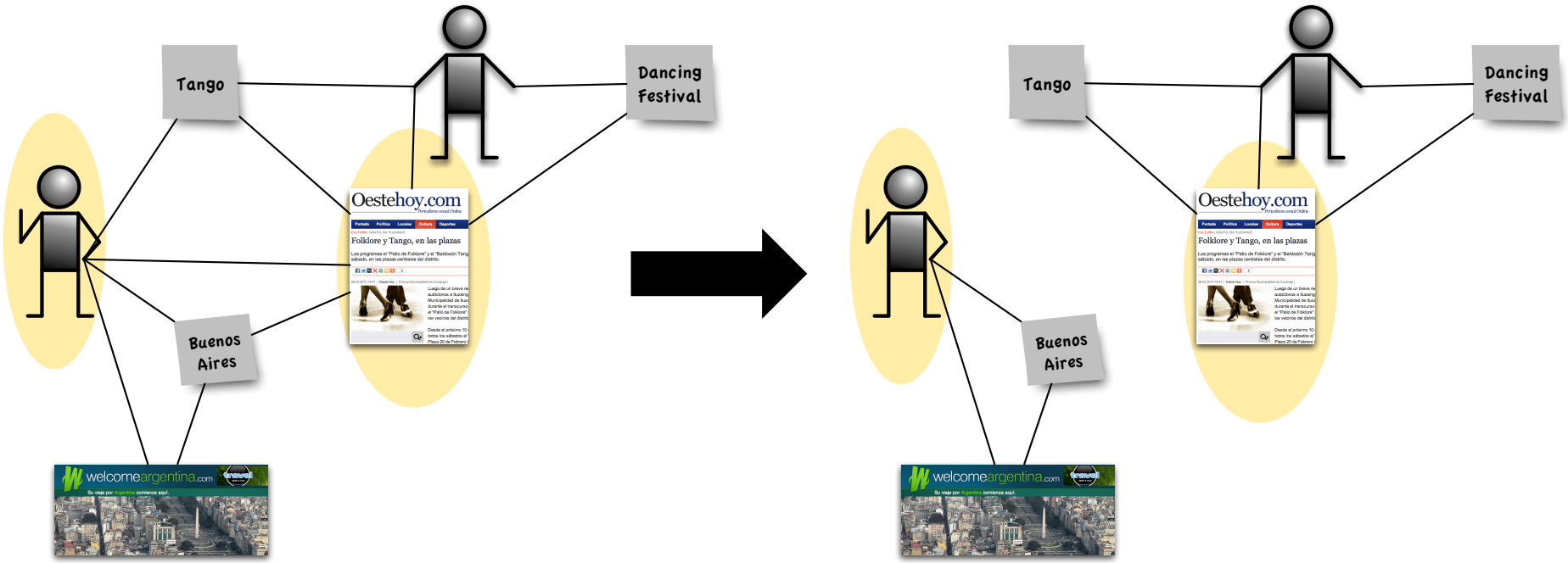


[adapted from Bogers 2009]

Evaluation Methodology: LeavePostOut

TECHNISCHE UNIVERSITÄT DARMSTADT

A post is a $P_{u,r}$= {(u,r,t)|(u,r,t) ∈ Y}

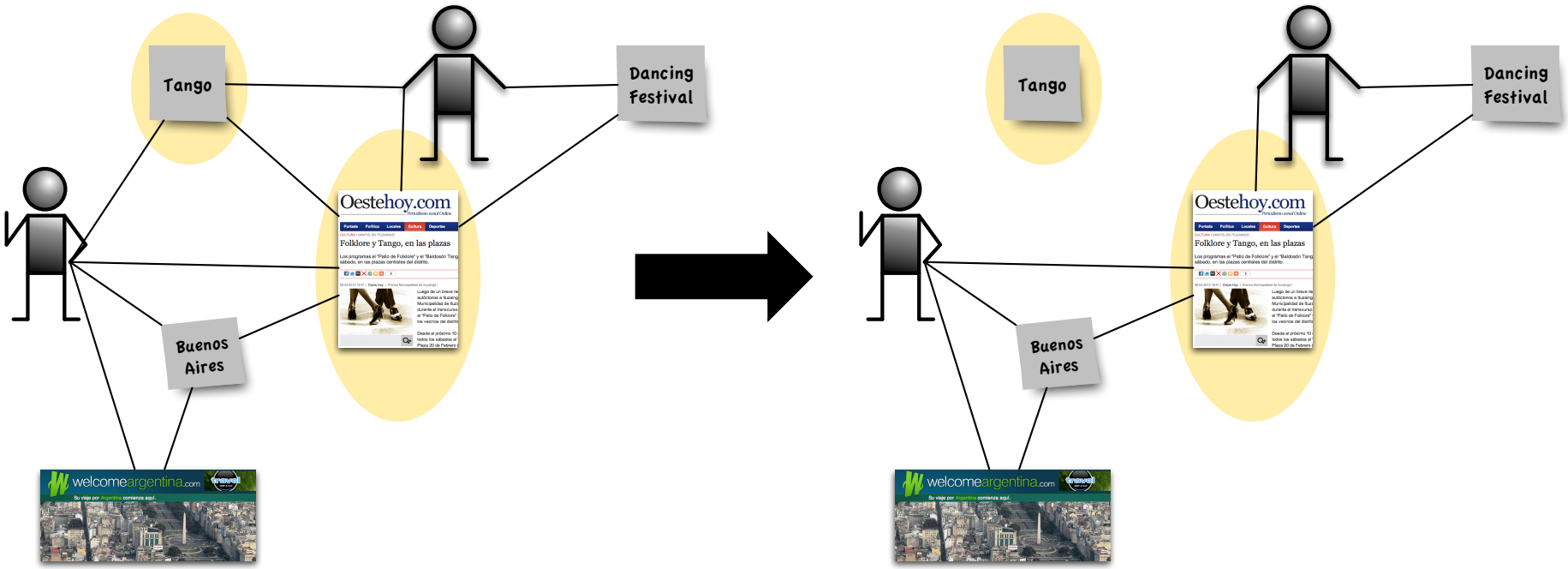


For LeavePostOut, the recommendation task with user as input is harder as with tag as input

[Jäschke et al. 2007]

Evaluation Methodology: LeaveRTOut

$RT_{r,t} = \{(u,r,t) | (u,r,t) \in Y\}$



For LeaveRTOut, the recommendation task with tag as input is harder as with user as input

Evaluation Metrics

TECHNISCHE UNIVERSITÄT DARMSTADT

Mean Average Precision:

$$\mathrm{MAP}(\mathrm{Q}) = \frac{1}{|\mathrm{Q}|} \sum_{\mathrm{j}=1}^{|\mathrm{Q}|} \frac{1}{\mathrm{m_j}} \sum_{\mathrm{k}=1}^{\mathrm{m_j}} \mathrm{Precision}(\mathrm{R_{jk}})$$

The mean of the Average Precision over several queries Q

[Manning et al 2008]

Mean Normalized Precision:

$$\mathrm{MNP}(\mathrm{Q}, \mathrm{k}) = \frac{1}{|\mathrm{Q}|} \sum_{\mathrm{j}=1}^{|\mathrm{Q}|} \frac{\mathrm{Precision_j(k)}}{\mathrm{Precision_{max,j}(k)}}$$

The mean of the normalized Precision at k (with respect to the maximal achievable Precision at k) over several queries Q

e.g. for LeavePostOut with k = 10, $\mathrm{Precision}_{\mathrm{max}}(k) = 1/10$

Evaluation Corpus

TECHNISCHE UNIVERSITÄT DARMSTADT

Bibsonomy corpus with p-core extraction at level 5 to reduce noise and to focus on the dense portion of the corpus

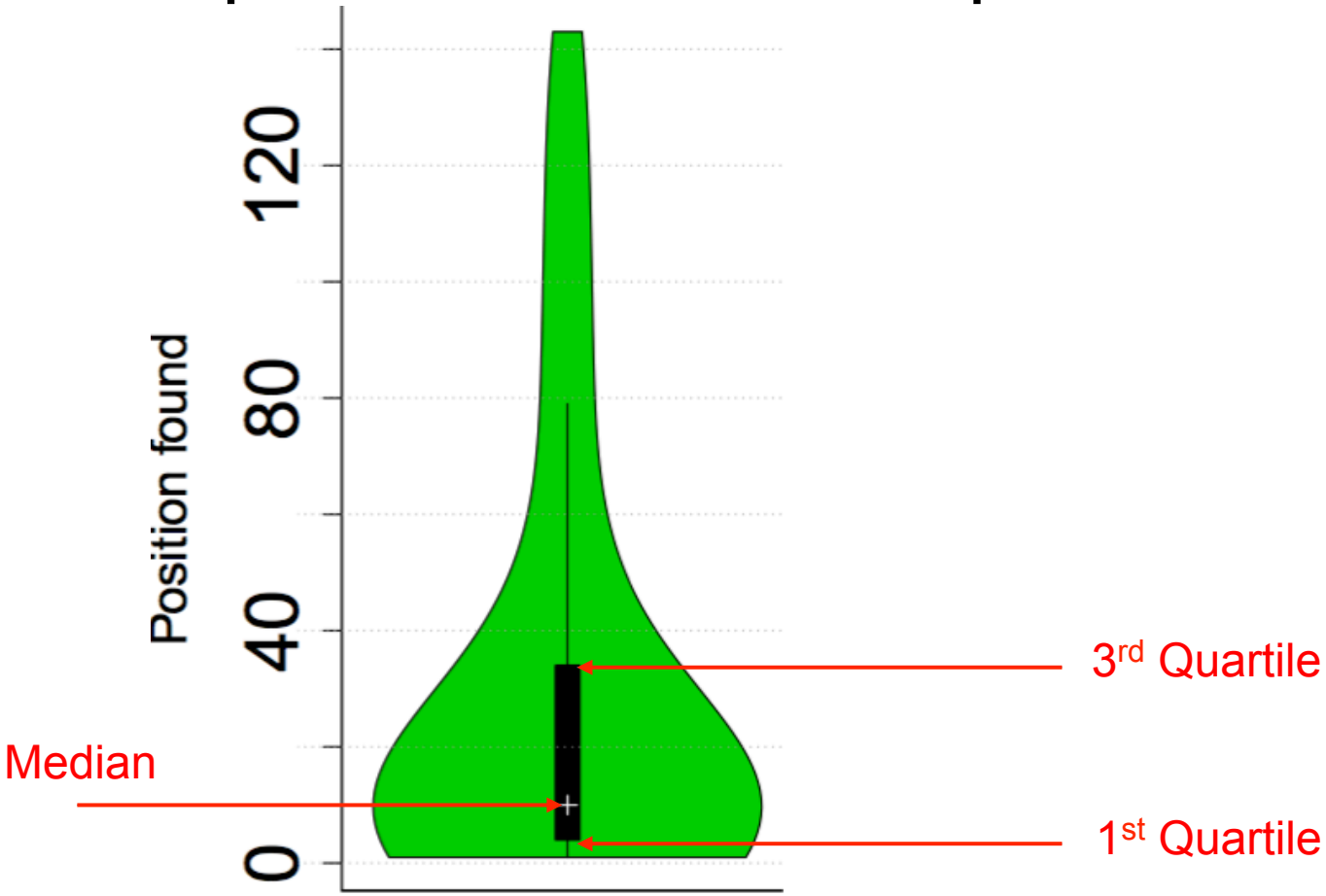
	Before	After
Users	7243	69
Bookmark resources	281550	9
Bibtex resources	469654	134
Tags	216094	179
Tag assignments	2740834	3269
Bookmark posts	330192	51
Bibtex posts	526691	959

FReSET – Domínguez García et al 2012
http://www.kom.tu-darmstadt.de/research-results/downloads/software/freset/

Visualization of Results with Violin Plots

TECHNISCHE UNIVERSITÄT DARMSTADT

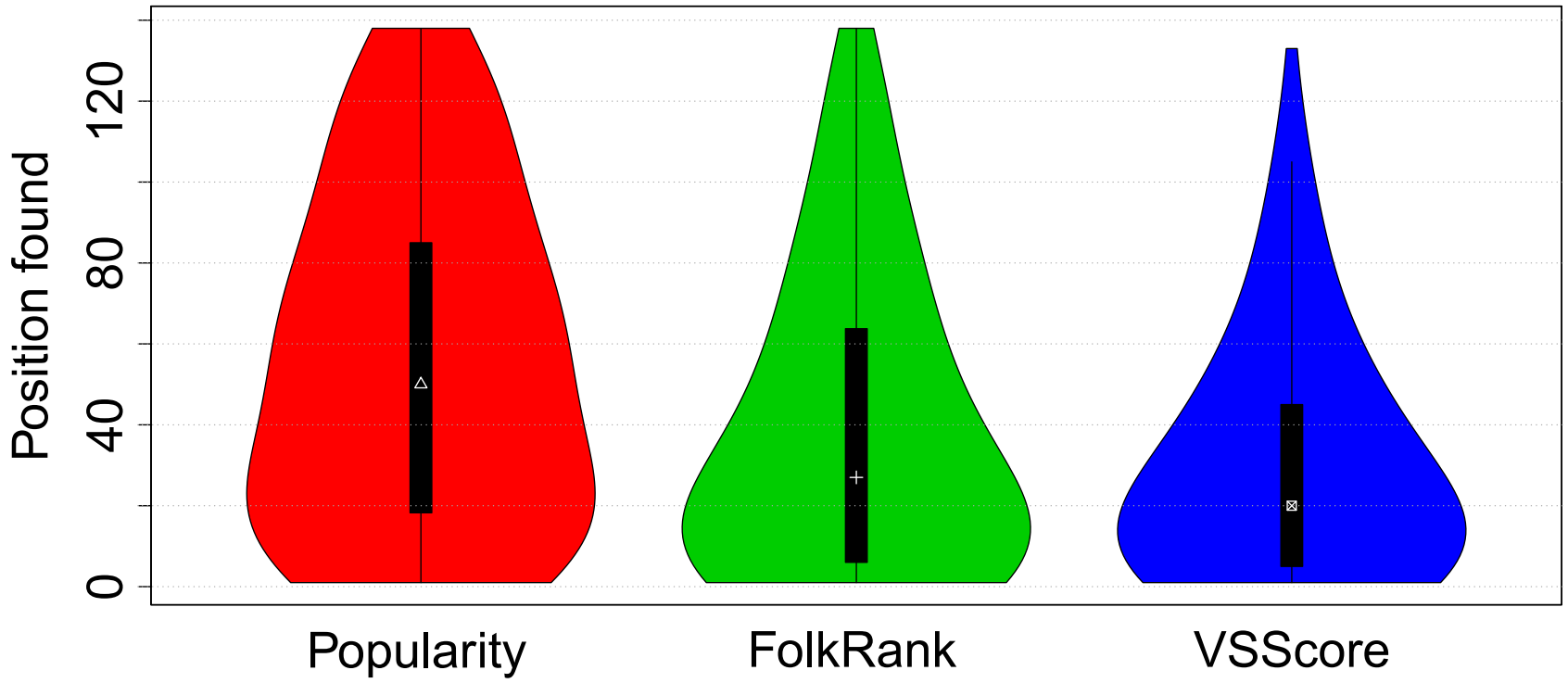
A violin plot is a combination of a box plot and a density trace



Evaluation Results for LeavePostOut

TECHNISCHE UNIVERSITÄT DARMSTADT

Evaluation results for recommendation task Interests Match

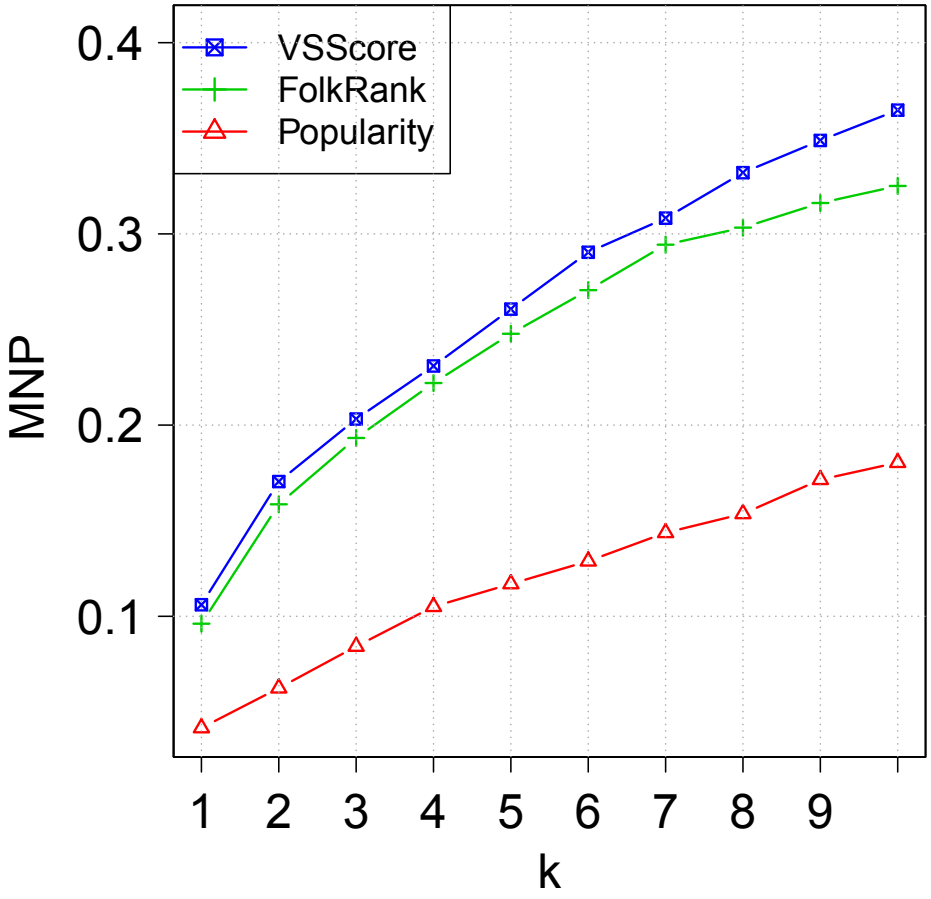


(Popularity of a resource is calculated as the sum of tags and users of a resource)

Evaluation Results for LeavePostOut

TECHNISCHE UNIVERSITÄT DARMSTADT

Evaluation results for recommendation task Interests Match

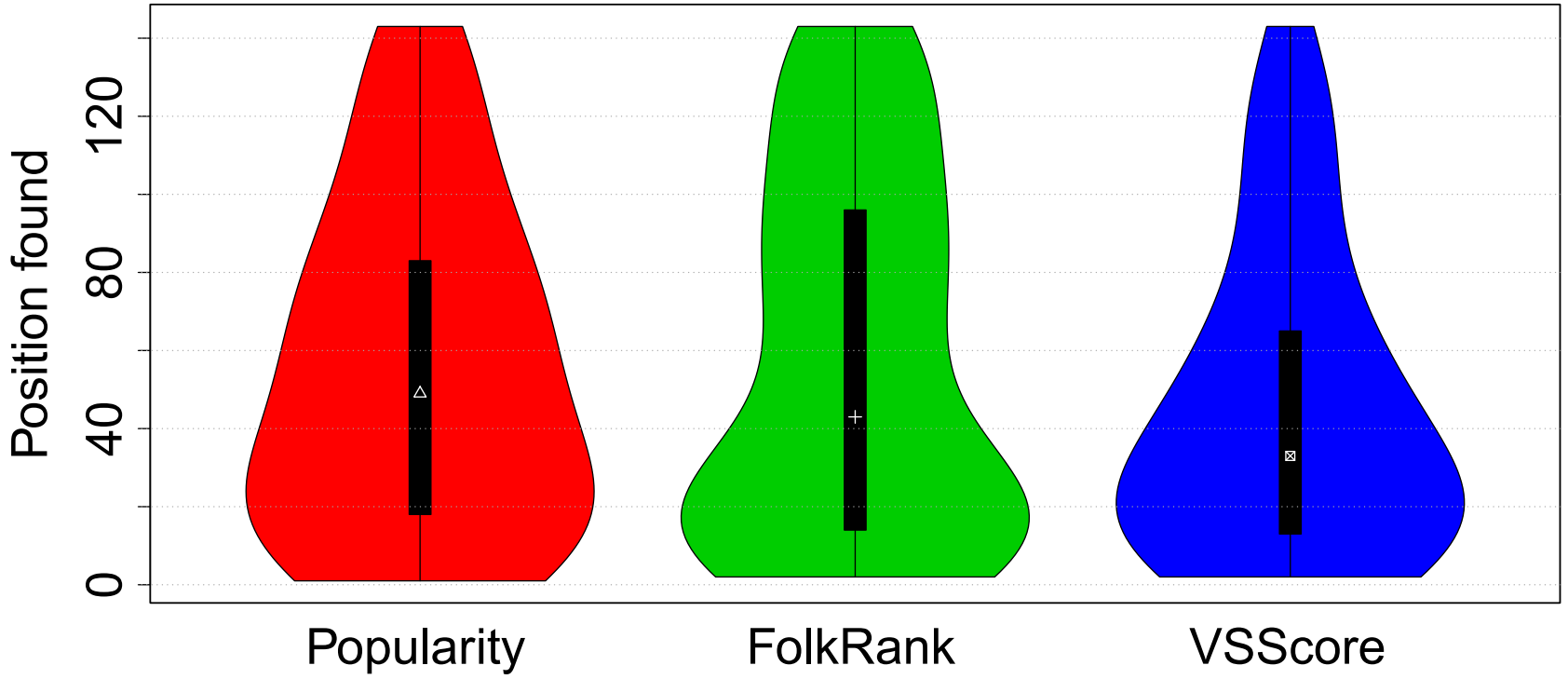


Approaches	MAP
VSScore	0.1972
FolkRank	0.1809
Popularity	0.0943

Evaluation Results for LeaveRTOut

TECHNISCHE UNIVERSITÄT DARMSTADT

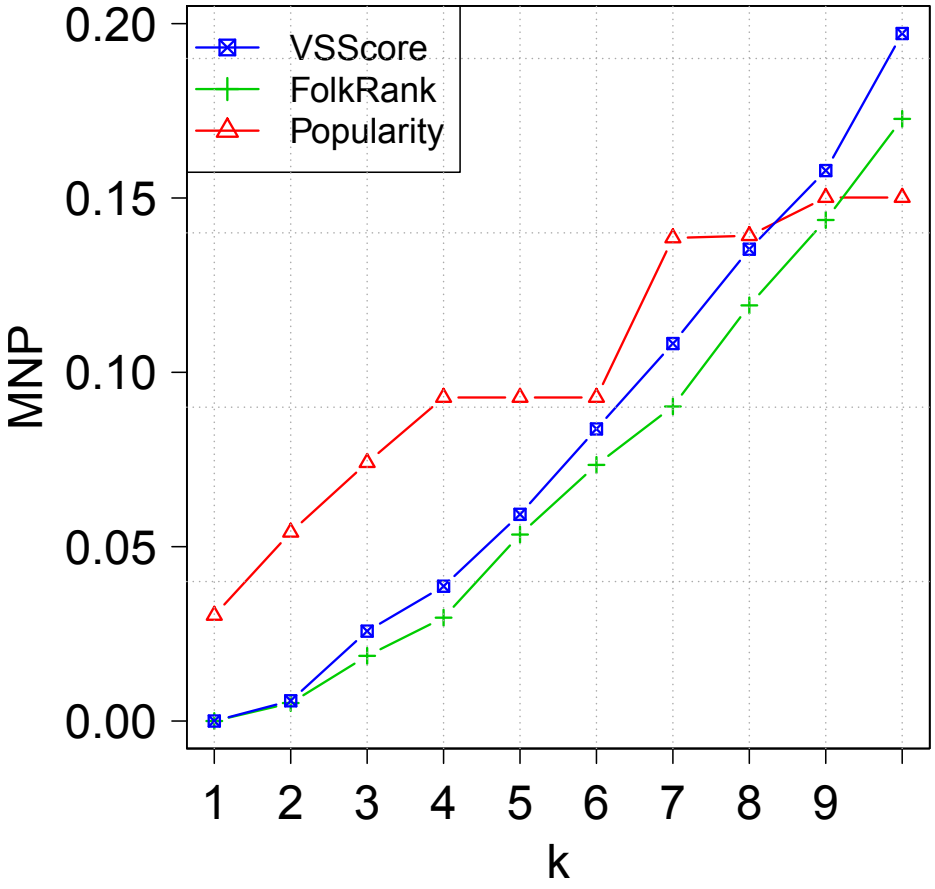
Evaluation results for recommendation task Guided Search



Evaluation Results for LeaveRTOut

TECHNISCHE UNIVERSITÄT DARMSTADT

Evaluation results for recommendation task Guided Search



Approaches	MAP
Popularity	0.0834
VSScore	0.0592
FolkRank	0.0529

Results for Statistical Significance Tests

TECHNISCHE UNIVERSITÄT DARMSTADT

Pairwise comparisons based on Average Precision with significance level of $p = 0.05$

- Scenario A: User-based resource recommendation
- Scenario B: Ranking of user‘s resources
- Scenario C: Tag-based resource recommendation

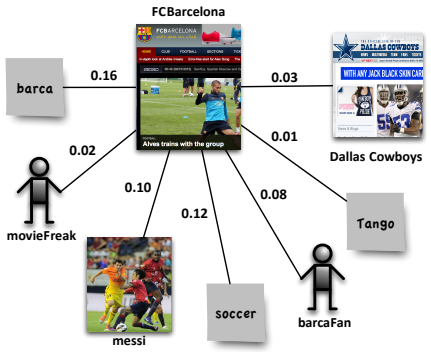
Methodology	Interests Match	Guided Search
LeavePostOut	VSScore^{A}	VSScore^{C}
LeaveNPostsOut	VSScore^{A}	FolkRank^{C}, VSScore^{C}
LeaveRTOut	FolkRank^{B}	VSScore^{C}
LeaveNRTsOut	FolkRank^{B}	VSScore^{C}

Wilcoxon signed-rank tests

Conclusion and Future Work

TECHNISCHE UNIVERSITÄT DARMSTADT

VSScore is a Framework leveraging context-specific information inherently found in a folksonomy for resource recommendation.

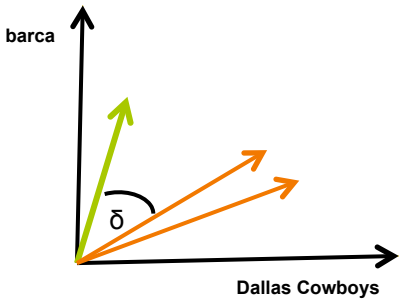


query-entity's context

$$\begin{bmatrix} 0.8 \\ \dots \\ 0.1 \end{bmatrix}$$

resources' contexts

$$\begin{bmatrix} 0.4 \\ \dots \\ 0.6 \end{bmatrix} \begin{bmatrix} 0.3 \\ \dots \\ 0.7 \end{bmatrix}$$



Limitations

- VSScore is computationally complex, therefore evaluations were performed on a limited corpus size

Future Work

- Reduce high-dimensional vector space to reduce computational complexity
- Evaluate on larger corpora from different domains
- Investigate further recommendation scenarios e.g. tag or user recommendation

Questions & Contact

TECHNISCHE UNIVERSITÄT DARMSTADT

Department of Electrical Engineering
and Information Technology
Multimedia Communications Lab - KOM

TECHNISCHE UNIVERSITÄT DARMSTADT

Dipl.-Inform. Mojisola Anjorin

Mojisola.Anjorin@KOM.tu-darmstadt.de
Rundeturmstr. 10
64283 Darmstadt
Germany

Phone +49 (0) 6151/166160
Fax +49 (0) 6151/166152
www.kom.tu-darmstadt.de