Weaving Linked Open Data into User Profiling on the Social Web

MultiA-Pro, Lyon, France, April 16th 2012

Fabian Abel, Claudia Hauff, Geert-Jan Houben, **Ke Tao**

Web Information Systems, TU Delft, the Netherlands

What we do: Science and Engineering for the Personal Web

domains: news social media cultural heritage public data e-learning

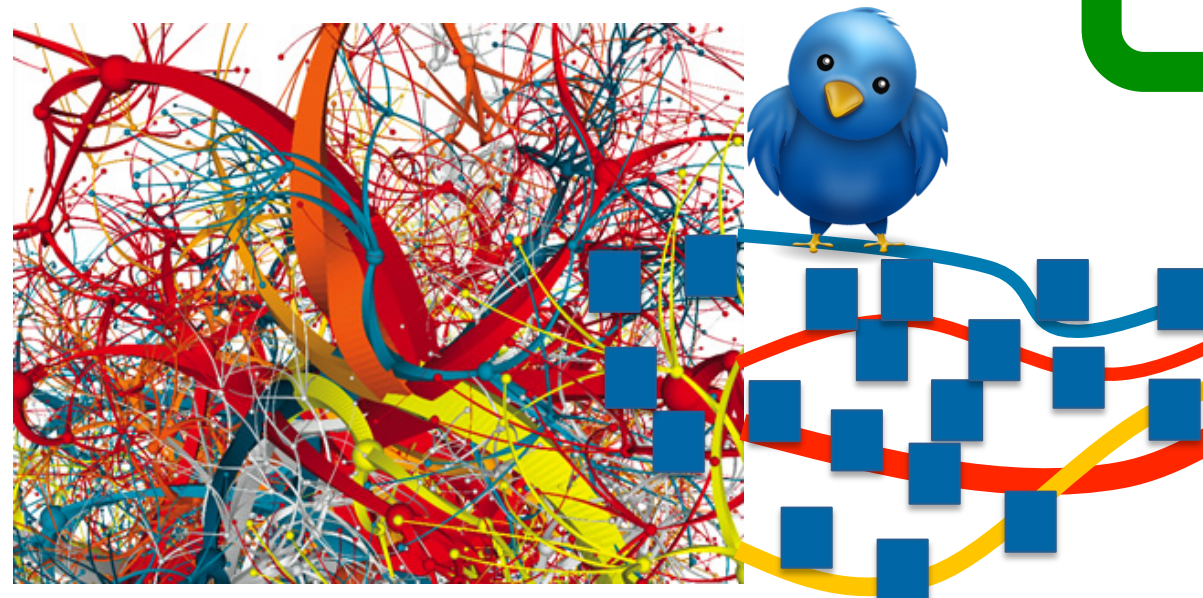| Personalized Recommendations | Personalized Search | Adaptive Systems |

recommending points of interest

Analysis and User Modeling

Semantic Enrichment, Linkage and Alignment

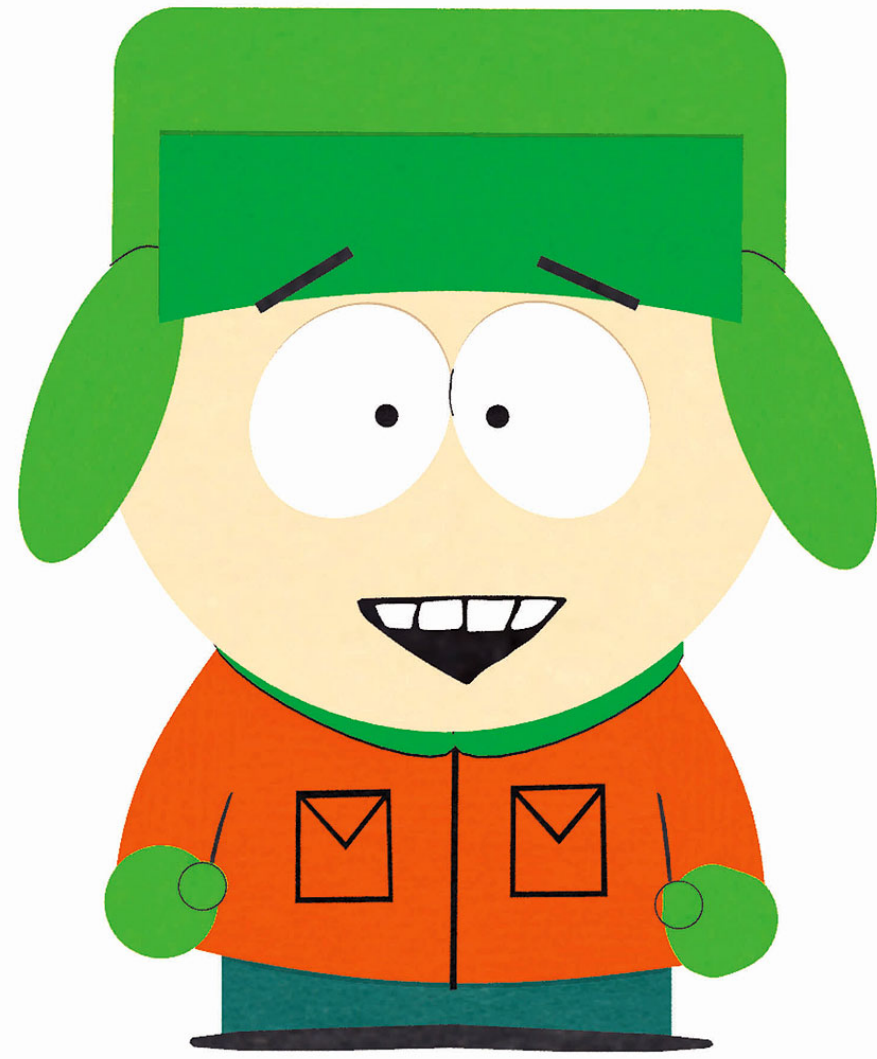*user/usage data*

Social Web

Kyle

Hometown: South Park, Colorado

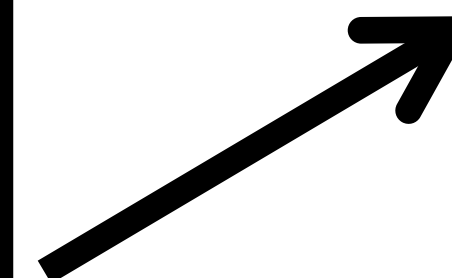# Kyle recently uploaded photos to Flickr
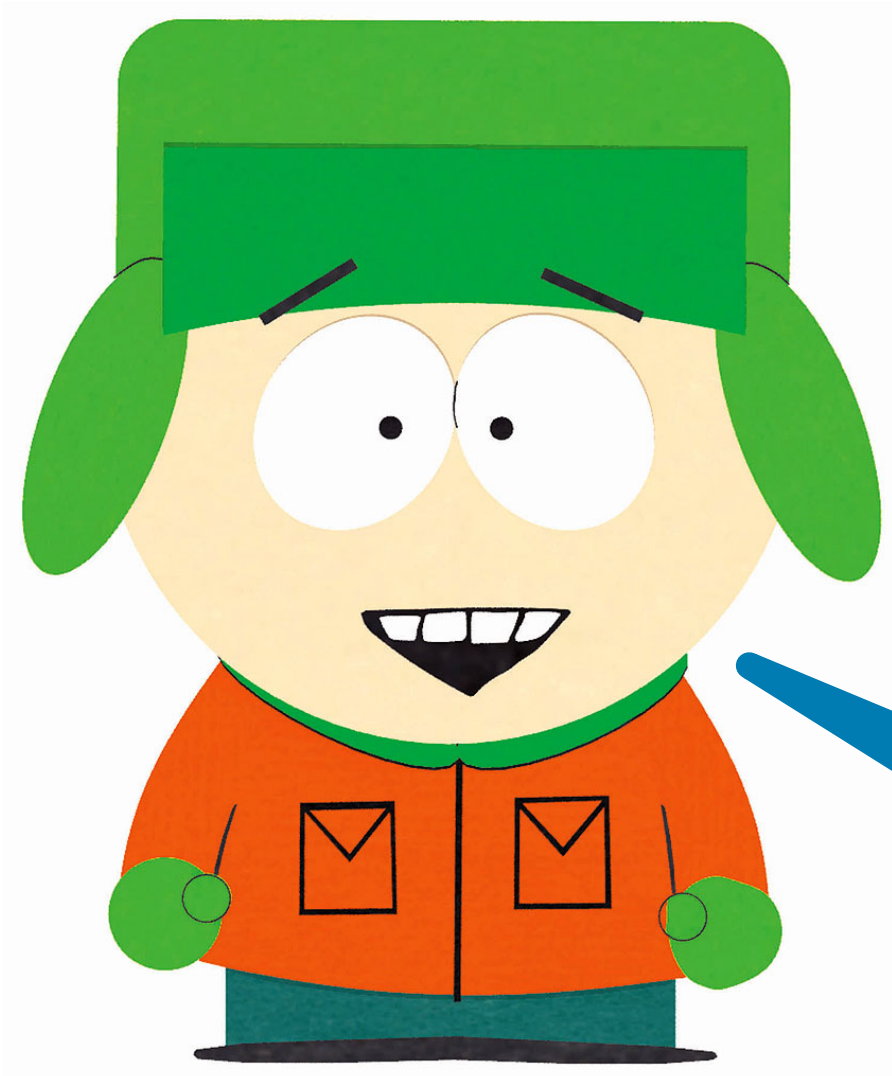

tags: delft,
geo: The H
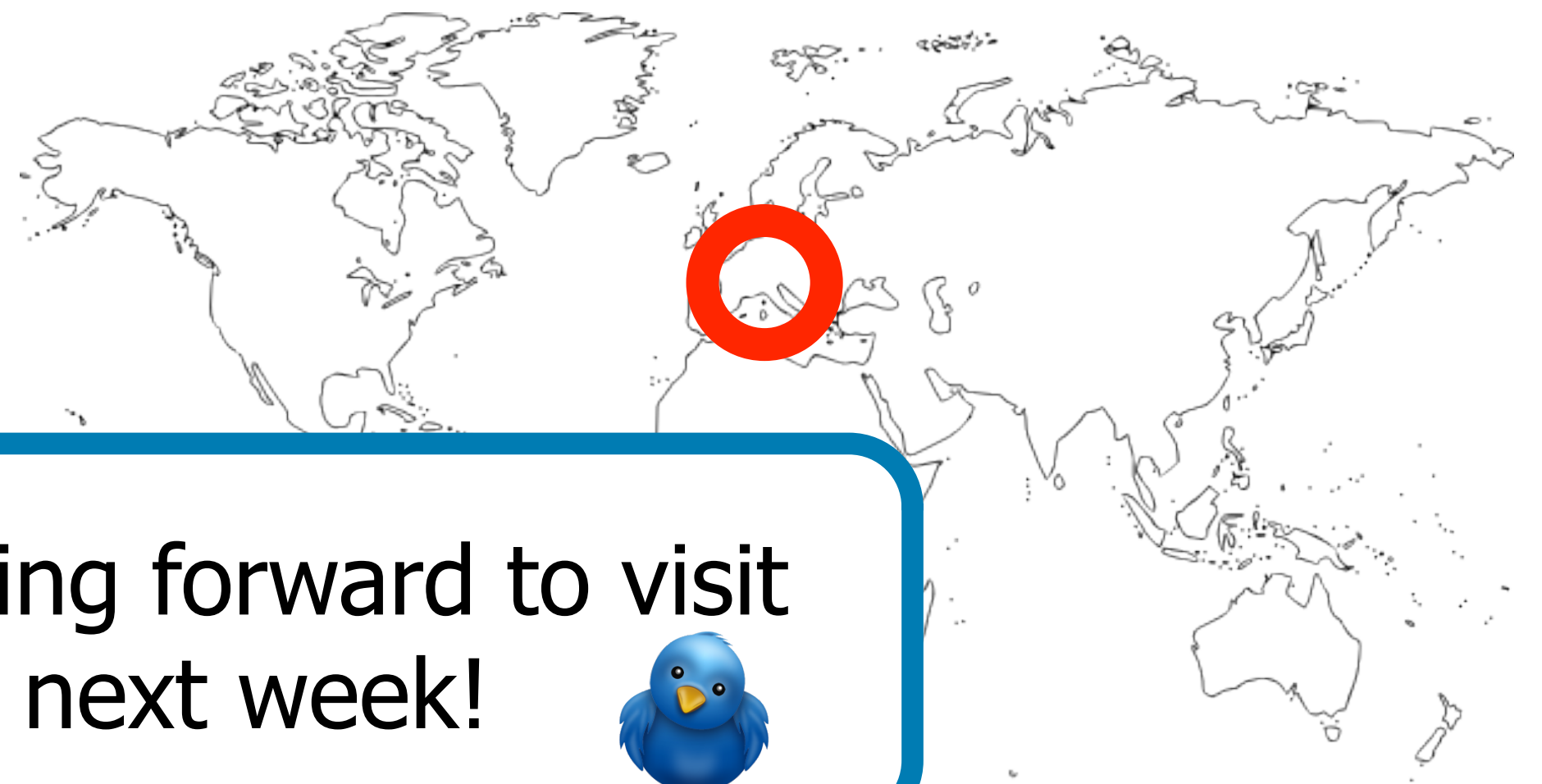
tags: girl with
pearl earring
geo: The Hague

flickr

During his trip to the Netherlands, he uploaded pictures to Flickr.
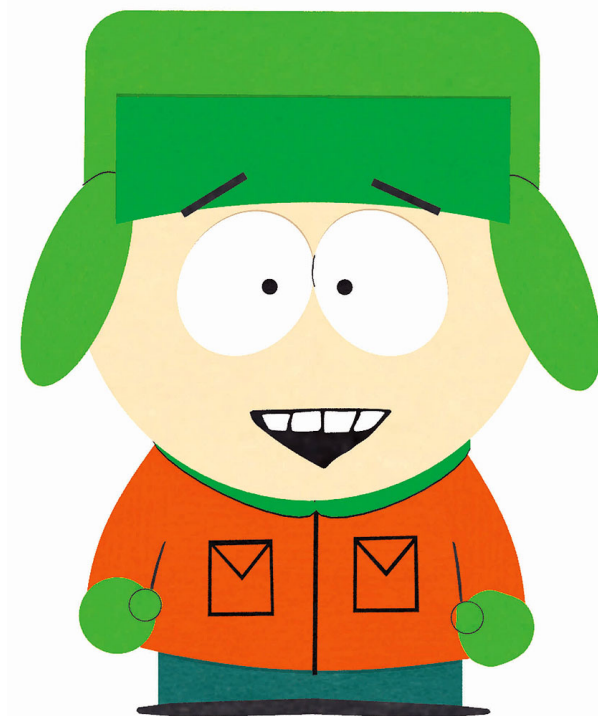
Kyle tweets about his upcoming trip

Looking forward to visit Paris next week!

Interests of Kyle?

- Given Kyle's Flickr and Twitter activities, can we infer Kyle's interests?

- Knowing that Kyle will visit Paris, France, can we recommend him places that might be interesting for him?

tags: delft,
geo: The H
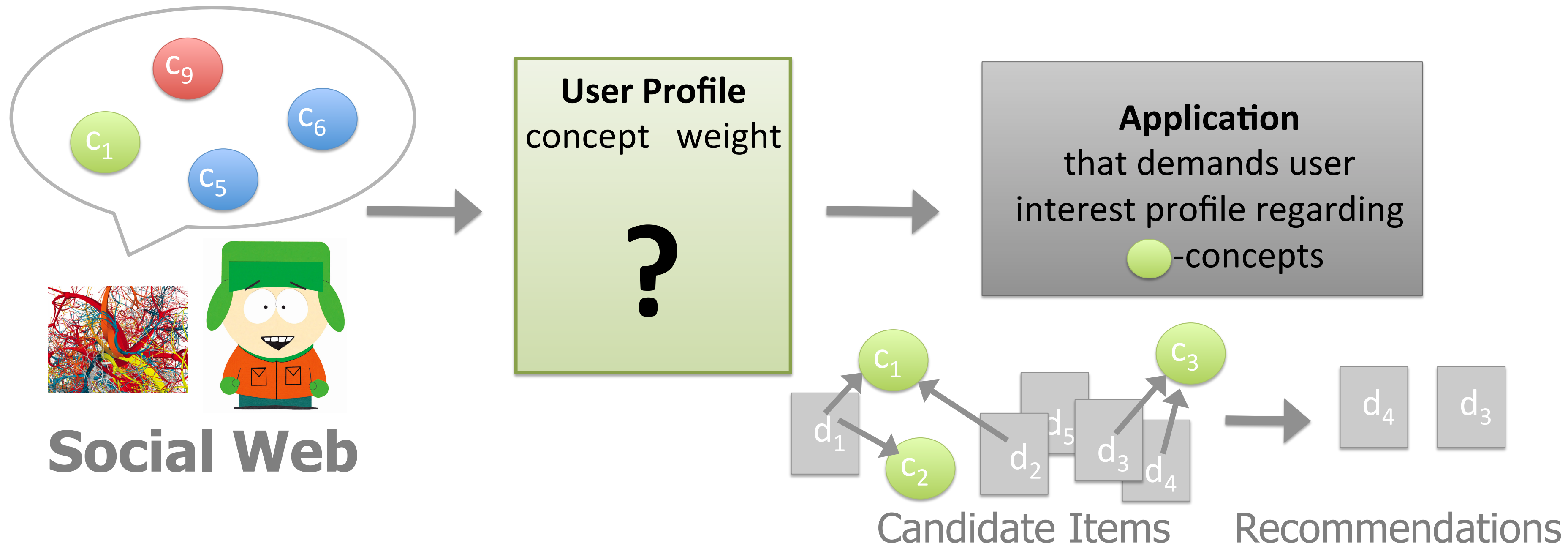
tags: girl with pearl earring
geo: The Hague
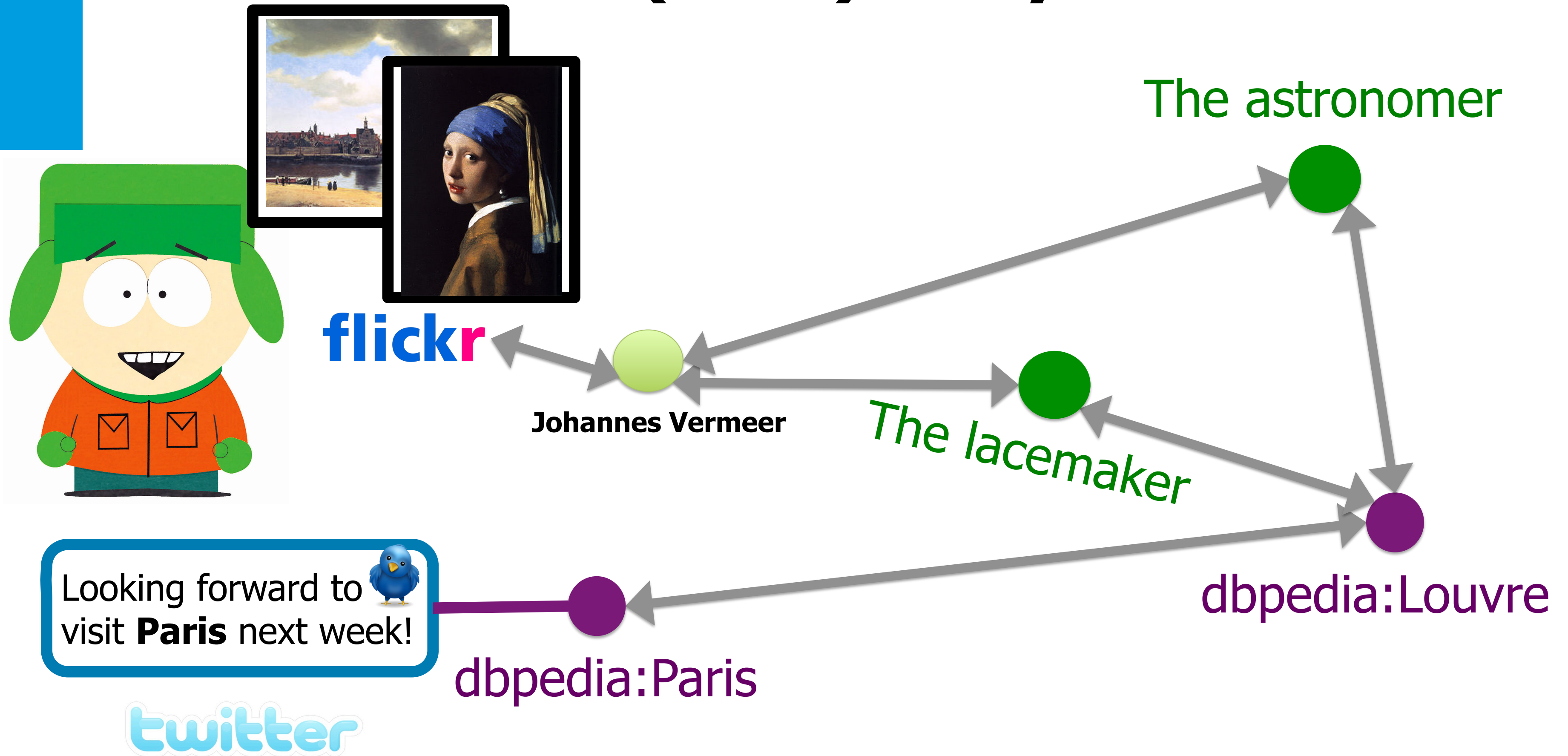
flickr

Looking forward to visit Paris next week!

twitter

Challenges

- How to create a meaningful profile that supports the given application?

 $\rightarrow$ *how to bridge between the Social Web chatter of a user and the candidate items of a recommender system?*

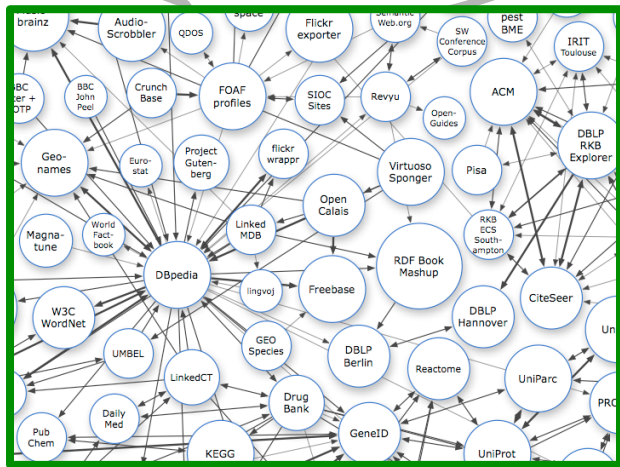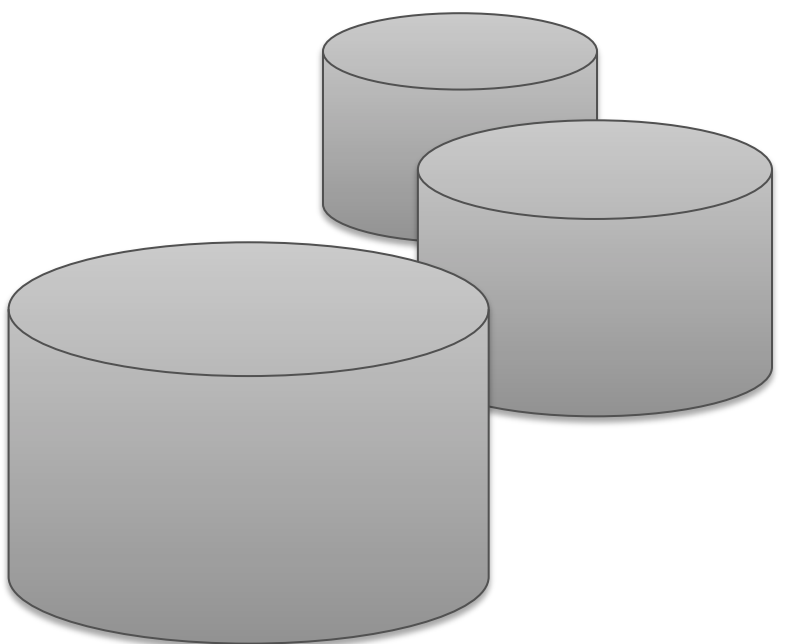# Challenge of Recommending Points of Interests (POIs) to Kyle

The astronomer

flickr

Johannes Vermeer

The lacemaker

Looking forward to visit **Paris** next week!

dbpedia:Paris

dbpedia:Louvre

twitter

LOD-based User Modeling

concepts that can be extracted from the user data

weighting strategies
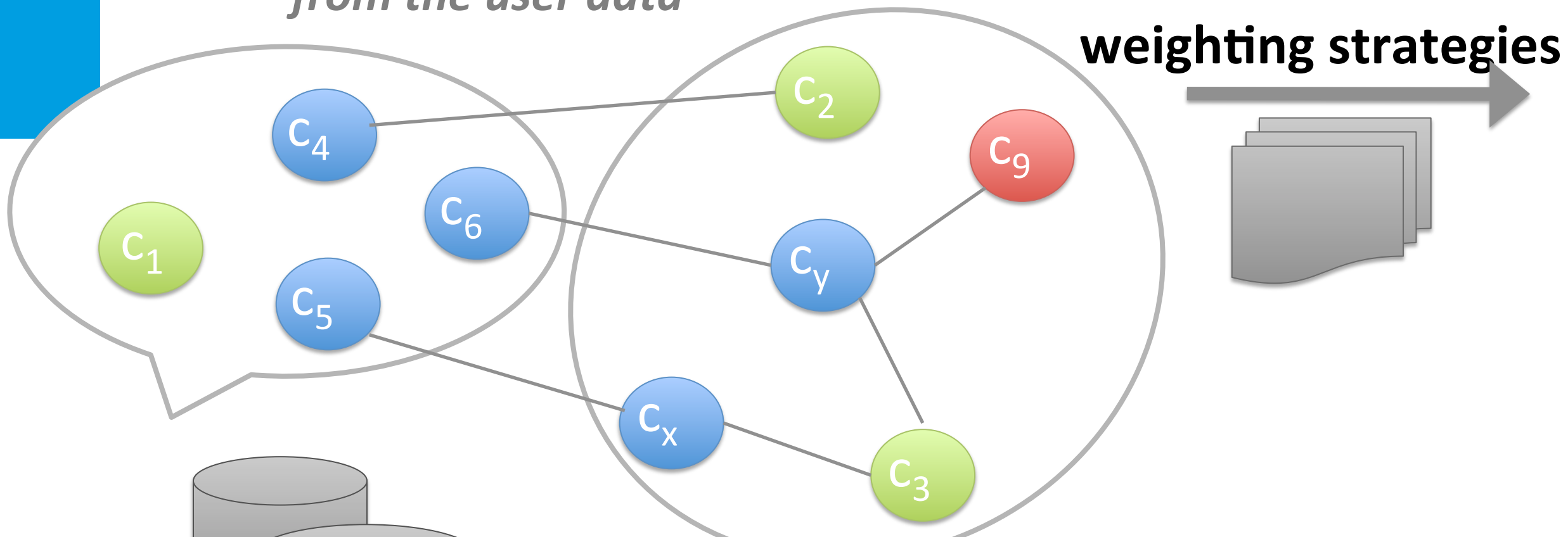
User Profile

concept	weight
c_1	0.4
c_2	0.1
c_3	0.2
...	...

user data

background knowledge (graph structures)

Social Web

Linked Data

Application
that demands user interest profile regarding
-concepts

User Modeling Building Blocks

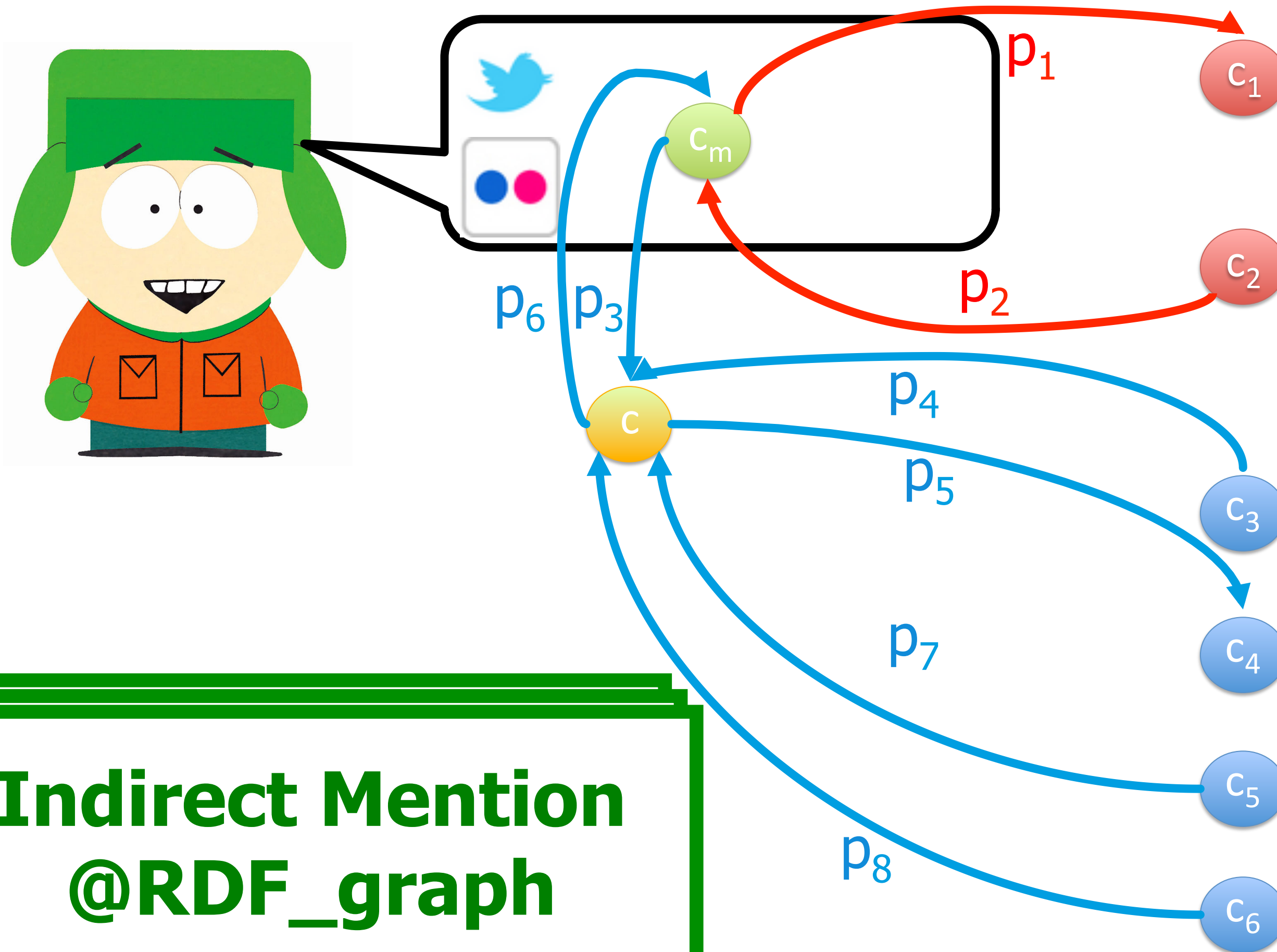# Strategies for exploiting the RDF-based background knowledge graph
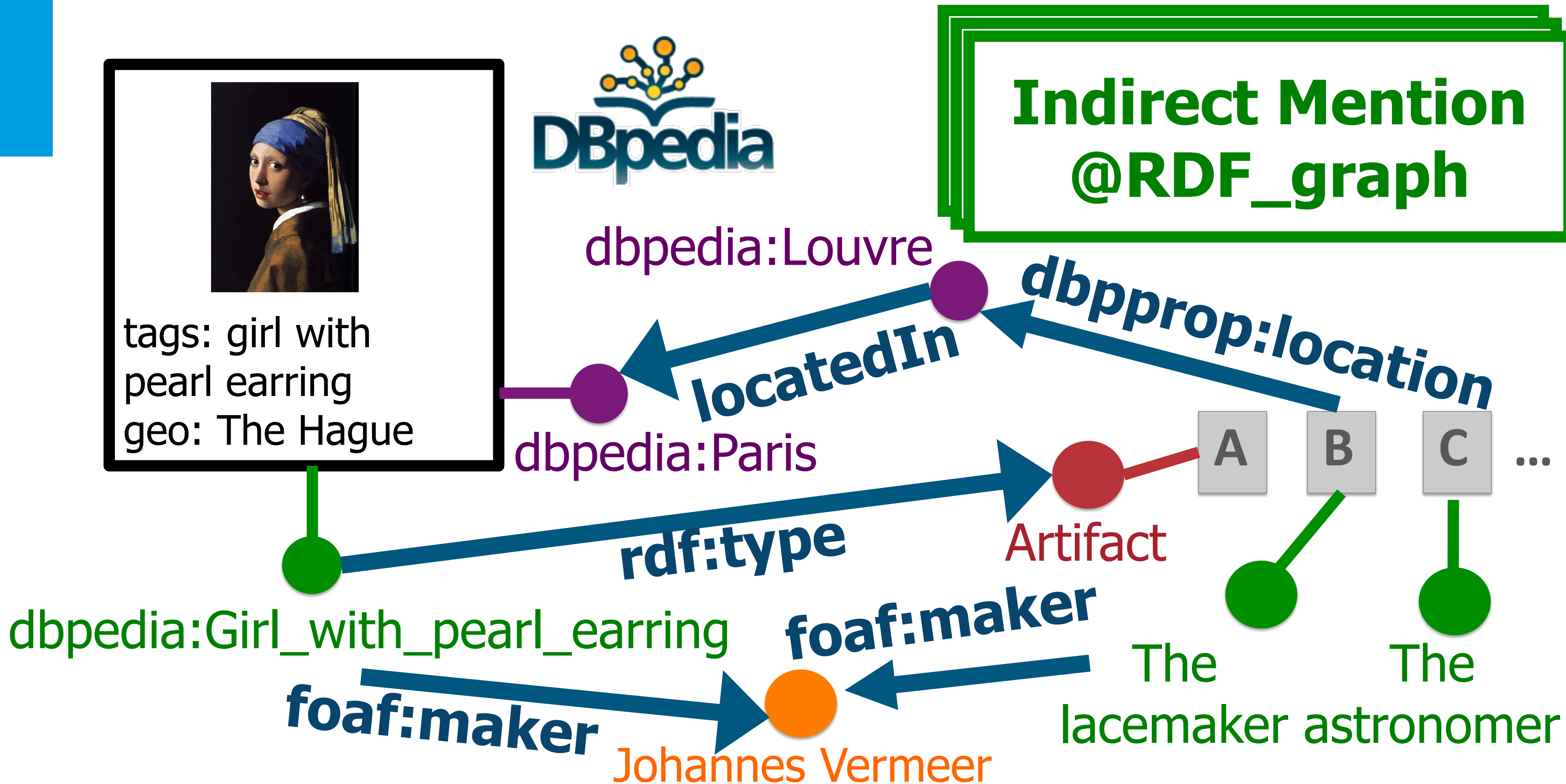
Indirect Mention @RDF_graph

tags: girl with pearl earring
geo: The Hague

dbpedia:Louvre

dbpedia:Paris

locatedIn

dbpprop:location

A B C ...

rdf:type

Artifact

dbpedia:Girl_with_pearl_earring

foaf:maker

foaf:maker
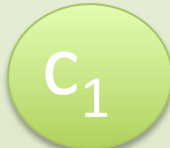
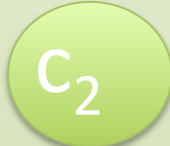Johannes Vermeer

The lacemaker astronomer

Strategies for exploiting the RDF-based background knowledge graph

Indirect Mention @RDF_graph

Weighting Scheme

User Profile

concept	weight
c_1	374
c_2	152
c_3	73
...	...

Weighting scheme: count the number of occurrences of a given graph pattern.

TUDelft

Source of User Data

- Twitter

- Flickr
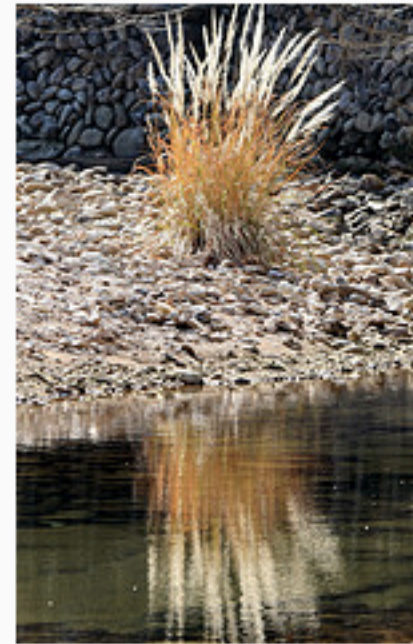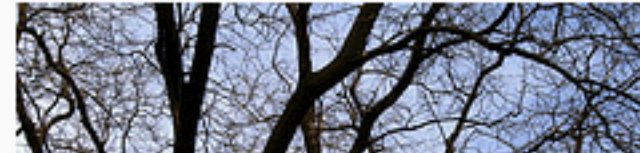
Mining the Geographic Origins of User Data

- Tweets or Flickr images posted by a GPS-enabled device;

Ke Tao @taubau   ← Reply   🗑 Delete   ⭐ Favorite   ·   Open
Next city: Paris. (@ Station Delft w/ 3 others) http://4sq.com/o2Xmzq
📍 📄 Vi

This photo was taken on January 28, 2012 in Beijing, CN, using a Nikon D700.

- Otherwise :   users assign to their ima

Zhangjiakou
Datong
Beiji
Qinhuangdao
Tangshan
@Yahoo!2010, ©NAVTEQ, CEInfoSystems 2009

Evaluation

TUDelft

Research Questions

1. How does the source of user data influence the quality in deducing user preferences for POIs?

2. How does the consideration of background knowledge from the Linked Open Data Cloud impact the quality of the user modeling?

3. What (combination of) user modeling strategies allows for the best quality?

Dataset

users	394
duration	11 months
tweets	2,489,088
location	11%
pictures	833,441
location	70.6%(within 10km)

twitter

flickr

Dataset Characteristics: Profile Sizes
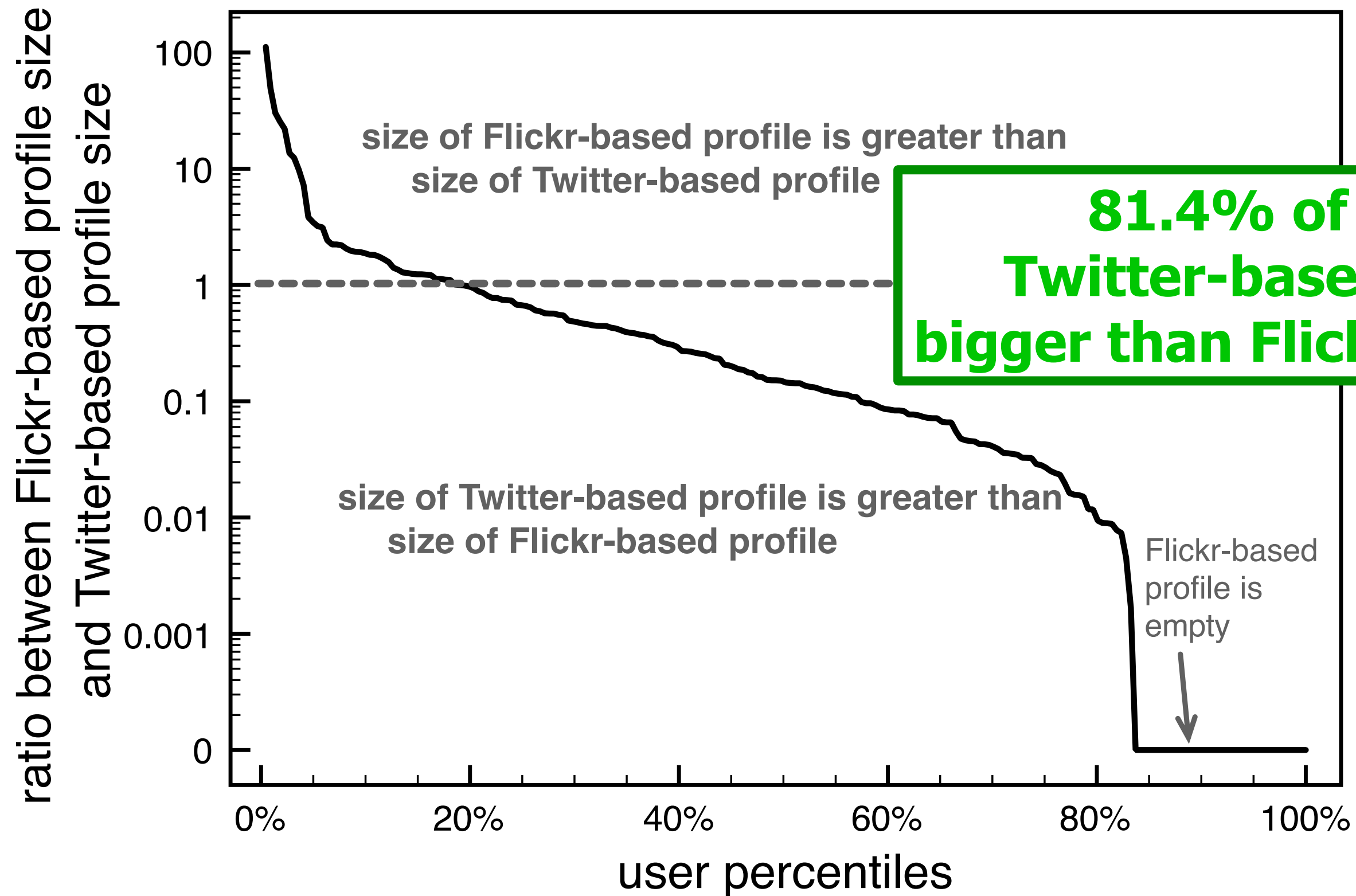
Experimental Setup

- **Task:**
  - = Recommending POIs
  - = Predicting POIs which a user will visit
- **Ground truth:**
  - split data into training data (= first nine months) and test data (= last two months)
  - POIs that the user visited in the last two months are considered as relevant
- **Metrics:**
  - **Precision@k, Recall@k and F-Measure@k:** precision, recall and f-measure within the top k of the ranking of recommended items

TUDelft

Results: Impact of User Data Source Selection
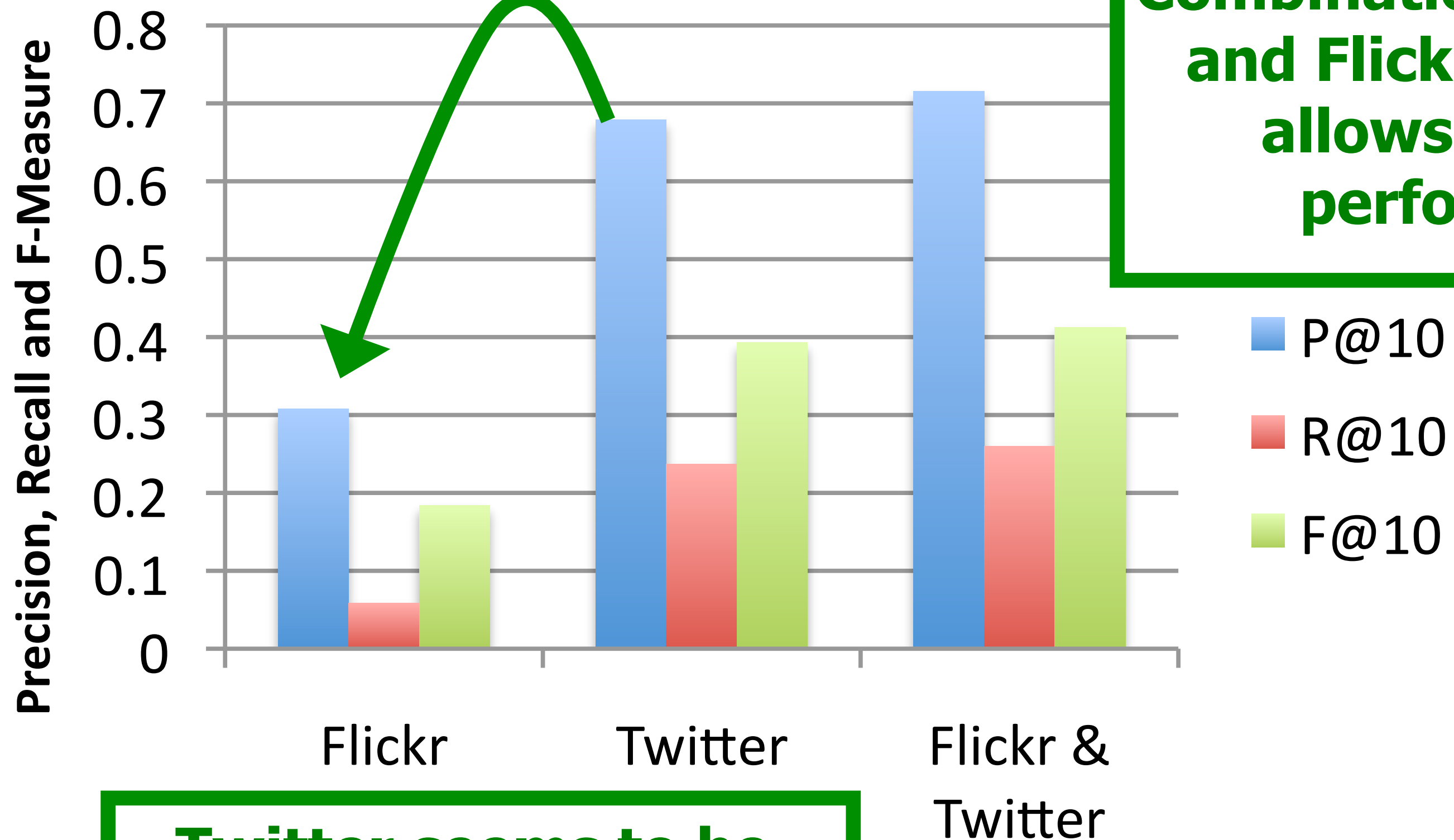
Combination of Twitter and Flickr user data allows for best performance

Twitter seems to be more valuable for the given application

Legend: P@10, R@10, F@10

Y-axis: Precision, Recall and F-Measure (0 to 0.8)

X-axis: Flickr, Twitter, Flickr & Twitter

Results: Impact of Strategies for Exploiting RDF-based Background Knowledge
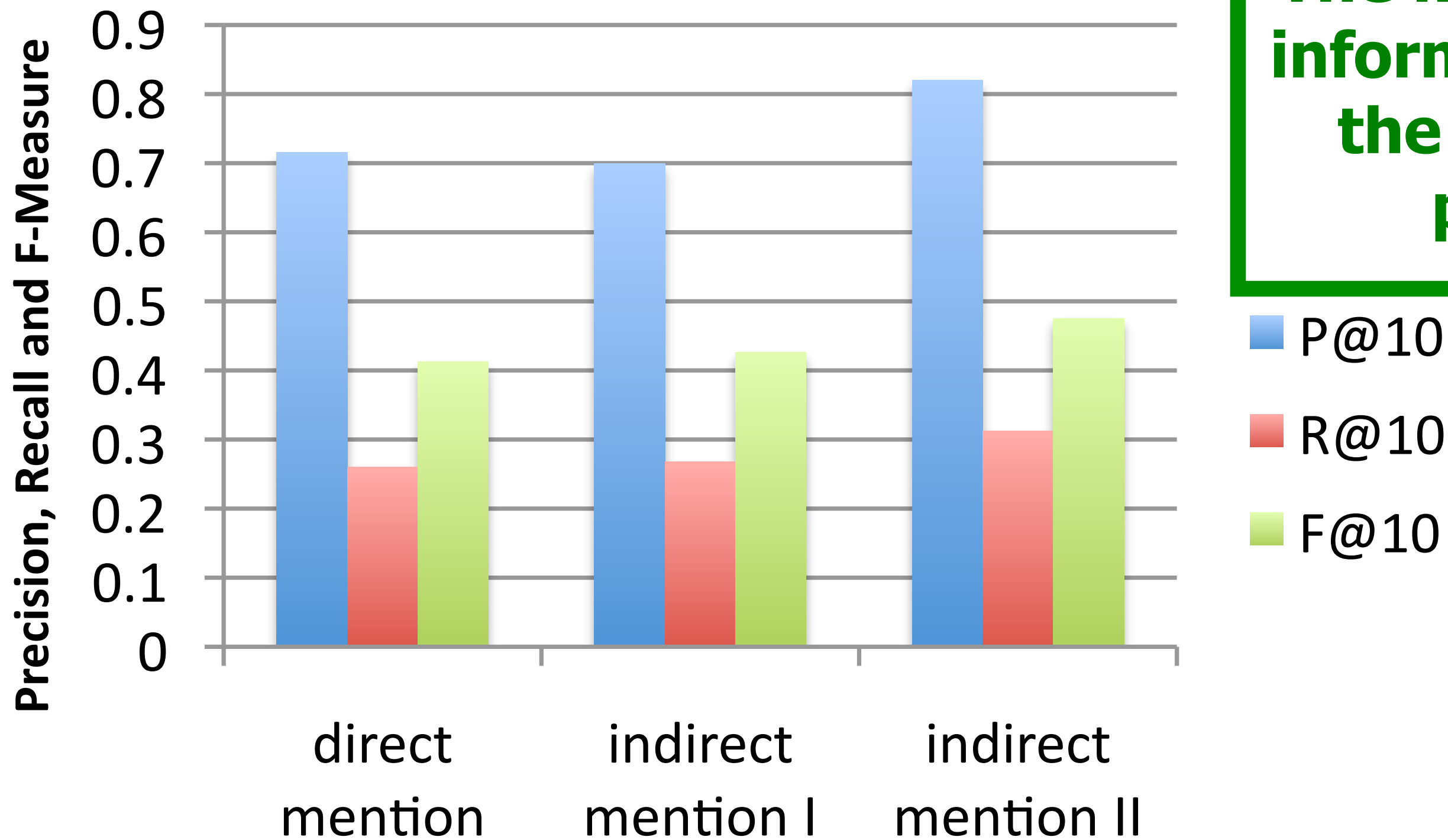
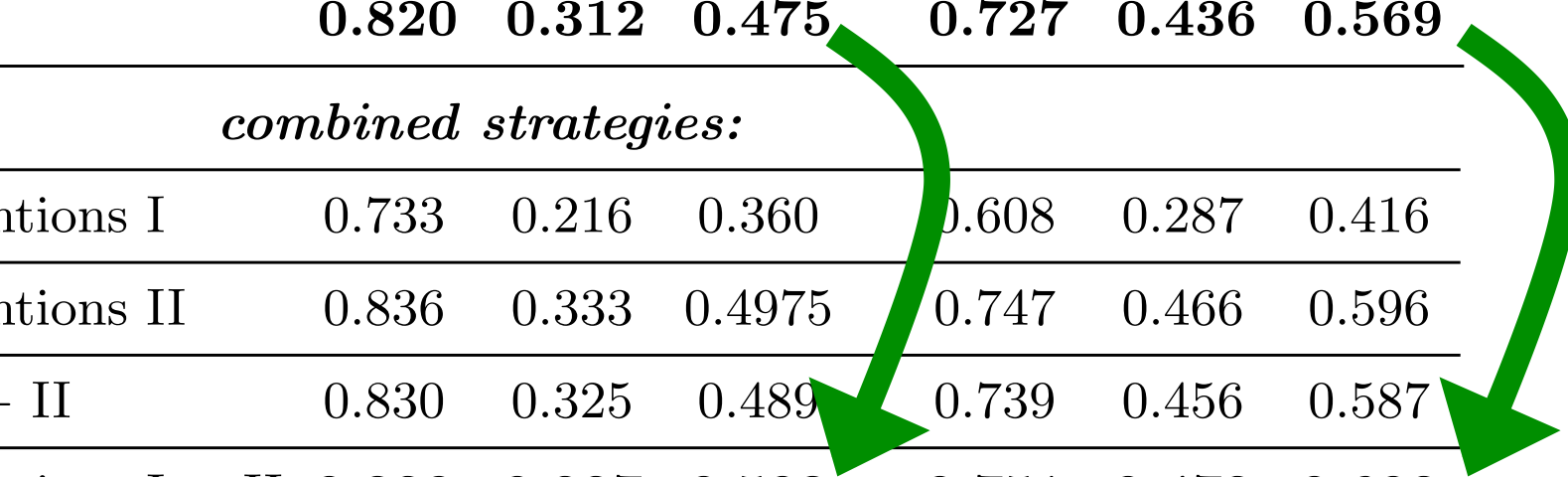The more background information the better the user modeling performance

Results: Combining different Strategies

strategy	P@10	R@10	F@10	P@20	R@20	F@20
indirect mentions II	**0.820**	**0.312**	**0.475**	**0.727**	**0.436**	**0.569**
combined strategies:						
direct & indirect mentions I	0.733	0.216	0.360	0.608	0.287	0.416
direct & indirect mentions II	0.836	0.333	0.4975	0.747	0.466	0.596
indirect mentions I + II	0.830	0.325	0.489	0.739	0.456	0.587
direct & indirect mentions I + II	**0.839**	**0.337**	**0.502**	**0.751**	**0.473**	**0.603**

Combining all background exploitation strategies improves the user modeling performance clearly

Conclusions

What we did:

- LOD-based User Modeling on the Social Web
- Different strategies for exploiting RDF-based background knowledge

Findings:

1. Combination of different user data sources (Flickr & Twitter) is beneficial for the user modeling performance
2. User modeling quality increases the more background knowledge one considers
3. Combination of strategies achieves the best performance

Future work:

- Investigate weighting schemes that weight the different RDF graph patterns for acquiring background knowledge differently

Thank you!

Slides : http://goo.gl/Zdg4K

Email: K.Tao@tudelft.nl
Twitter: @wisdelft @taubau
http://persweb.org