Crowdsourced Affinity: A Matter of Fact or Experience

Chun Lu, Milan Stankovic, Filip Radulovic, Philippe Laublet

ESWC 2017

Portoroz, Slovenia

About

Comparison of the usage of **knowledge graph** and **folksonomy** for user-entity **affinity assessment** with two studies within a **travel** destination recommendation scenario

Outline

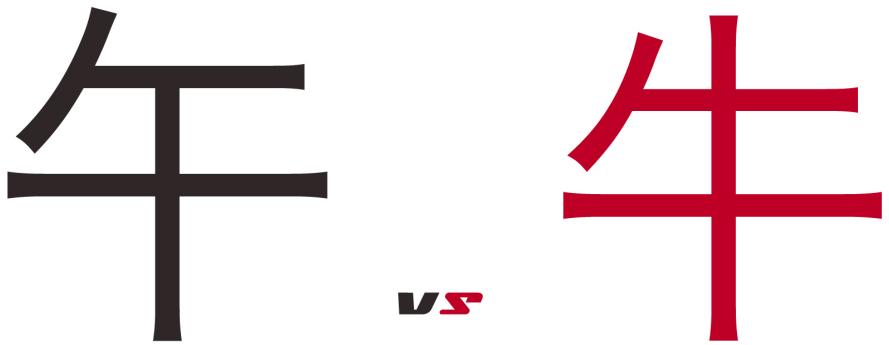
- Introduction about user-entity affinity
- Knowledge graph V.S. Folksonomy
- Gold standard study
- Semantic Affinity Framework
- User study
- Take-away

User-entity affinity

- Likelihood of a user to be attracted by an entity (book, film, artist) or to perform an action (click, purchase, like, share) related to an entity
- Big impact from both economic and user experience point of view in user-centric information systems:
  - Online advertising
  - Exploratory search
  - Recommendation

Affinity assessment technique

- Content-based: users have higher affinity with entities that are similar to the ones with which they had positive interactions in the past
- Entity similarity
- Boosted by Knowledge graphs and Folksonomies with large amount of data about entities



Knowledge graph V.S. Folksonomy

	Knowledge graph	Folksonomy
Collaborative	Creation of large public knowledge graphs like DBpedia and Wikidata	Annotation and categorization of entities with free folksonomy tags
Structure	Formal underlying ontology	Loose structure
Nature	Factual (e.g dbr:Jumanji, dbo:starring, dbr:Robin_Williams)	Experience and intersubjectivity (e.g nostalgic, not funny, natural disaster)

Research question

- Which data space contributes better to the user-entity affinity assessment?
- Unanswered in the literature
- Proliferation: linked data cloud, hastags, instagram, flickr, mendely
- Two experiments within a travel destination recommendation scenario:
  - Gold standard study
  - User study

Why travel domain?

- More than 80% of people do travel planning online
- 68% of travelers begin searching online without having a clear travel destination in mind.
- Recommender systems can help travelers find more efficiently destinations (cities) in affinity with them.

Gold standard study

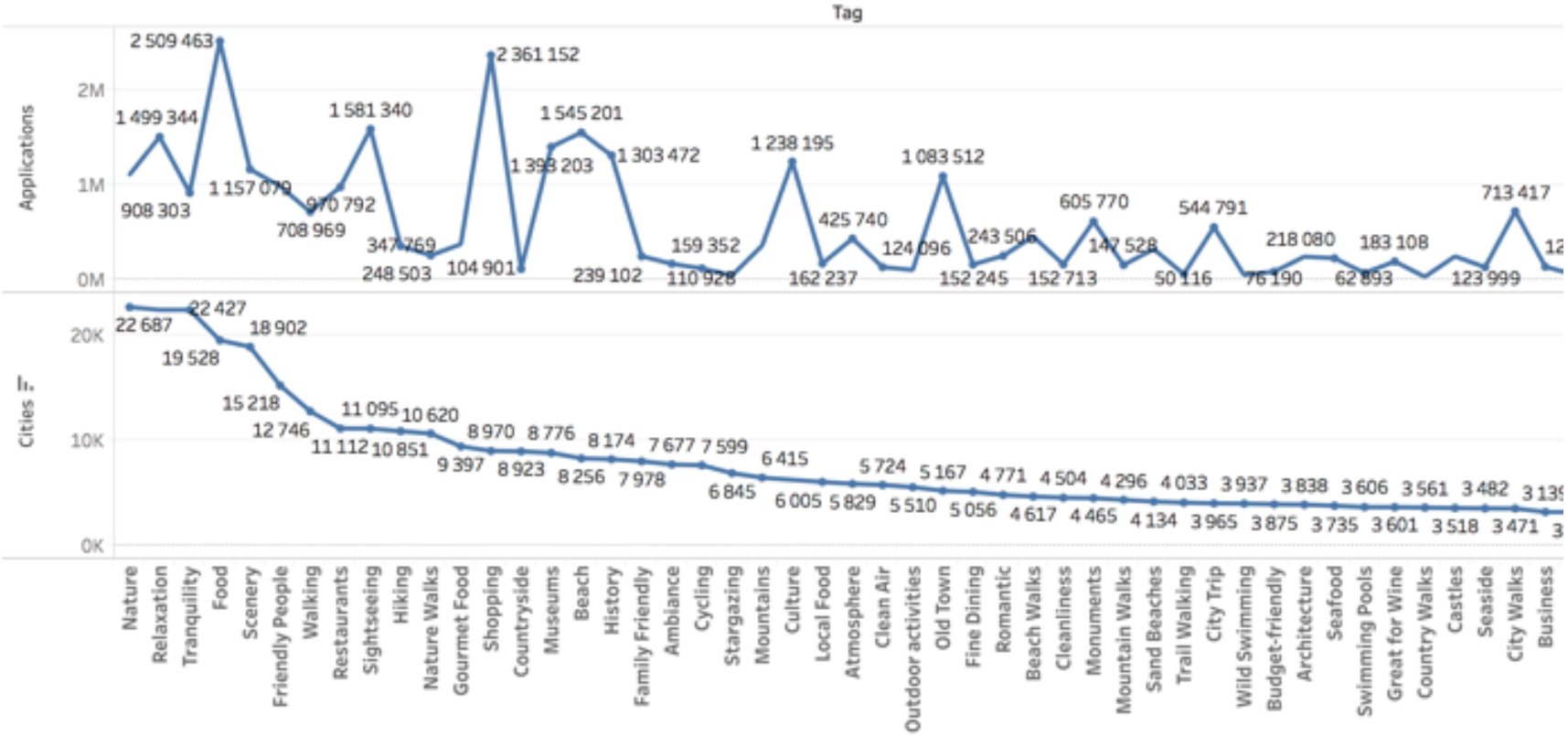
- Dataset: processed YFCC 100M (100 million of geotagged photos and videos published on Flickr)
  - Users with travel sequences
  - dbr:Munich -> dbr:Stockholm -> dbr:New_York_City



# users	3878
# cities	705
avg # cities per user	5.27

Folksonomy engineering

- Crawled from a collaborative travel platform
- 234 tags (e.g kayaking, great for wine, people watching)
- 26,237 cities
- 154 countries
- TF-IDF
- Cosine similarity



Knowledge graph engineering

- Property selection
- Feature retrieval
- Feature cleaning
- Jaccard measure

inbound		outbound	
dbo:birthPlace	dbo:broadcastArea	dbo:isPartOf	dbo:part
dbo:location	dbo:nearestCity	dbo:country	dbo:twinTown
dbo:deathPlace	dbo:ground	dbo:timeZone	dbo:saint
dbo:city	dbo:foundationPlace	dbo:Mayor	dbo:district
dbo:capital	dbo:assembly	dbo:region	dct:subject
dbo:hometown	dbo:restingPlace	dbo:province	(skos:broader)
dbo:recordedIn	dbo:place	dbo:leaderName	
dbo:residence	dbo:locationCity		
dbo:headquarter			

Protocol

- Candidates: KG & FOLK
- Common affinity calculation:

$$affinity(u, c_i) = \frac{\sum_{c_j \in profile(u)} Sim(c_i, c_j)}{|profile(u)|}$$

- All but n strategy, n=1, 5.27 cities per profile
- Top-10, -20 et -30 recommendations

Metrics

- Success $$Success = \frac{\sum_{u \in U} rel_{g,u}}{|U|} \text{ where } rel_{g,u} = \begin{cases} 1, & if\ ground\ truth\ g\ is\ in\ top - N \\ 0, & otherwise \end{cases}$$

- Mean Reciprocal Rank (MRR) $$MRR = \frac{1}{|U|} \sum_{u \in U} \frac{1}{rank_u}$$

- Intra-List Similarity (ILS) (*dbo:country, dct:subject*)

$$ILS_u@N = \sum_{i \in L_u^N} \sum_{j \in L_u^N} \frac{sim(i,j)}{|pairs|}$$

$$ILS@N = \frac{1}{|U|} \sum_{u \in U} ILS_u@N$$

- Novelty (long-tail cities) $$Novelty@N = \frac{number\ of\ recommended\ long - tail\ cities}{N * |U|}$$

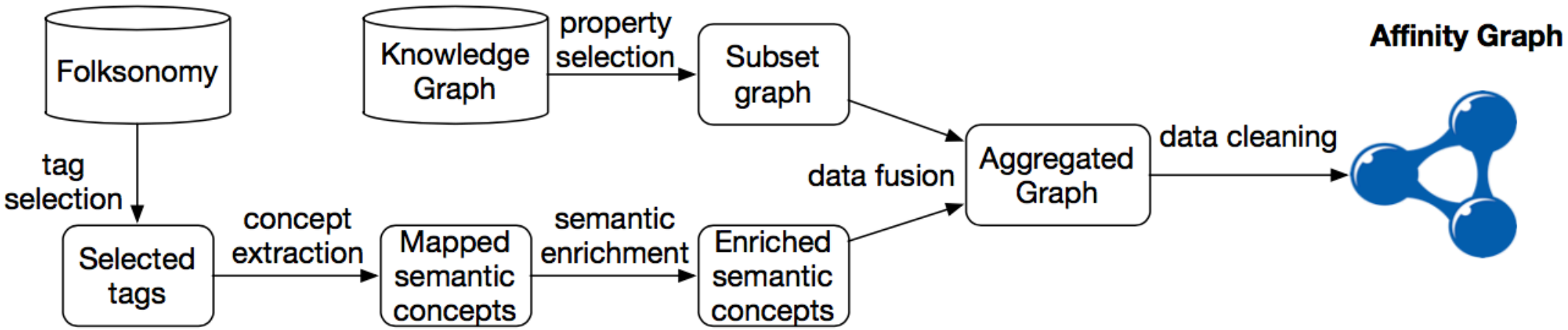
Results

	Top-10		Top-20		Top-30	
	KG	FOLK	KG	FOLK	KG	FOLK
Success	**0.232**	0.06	**0.33**	0.116	**0.386**	0.166
MRR	**0.047**	0.003	**0.047**	0.003	**0.047**	0.003
ILS	0.257	**0.089**	0.208	**0.072**	0.176	**0.065**
Novelty	0.717	**0.824**	0.722	**0.772**	0.723	**0.755**

- KG > FOLK on Success & Mean Reciprocal Rank
- FOLK < KG on Intra-List Diversity and Novelty
- Complementarity between KG and FOLK

Semantic Affinity Framework

- Integrate, aggregate, enrich and clean entity data from knowledge graphs and folksonomies



Content-based explanation

- Content-based explanation
  - E.g. Recommending *Ljubljana* because of *dbc:Capitals_in_Europe*
- Most common features shared by entities in user profile
- Explanation diversification with regards to the properties

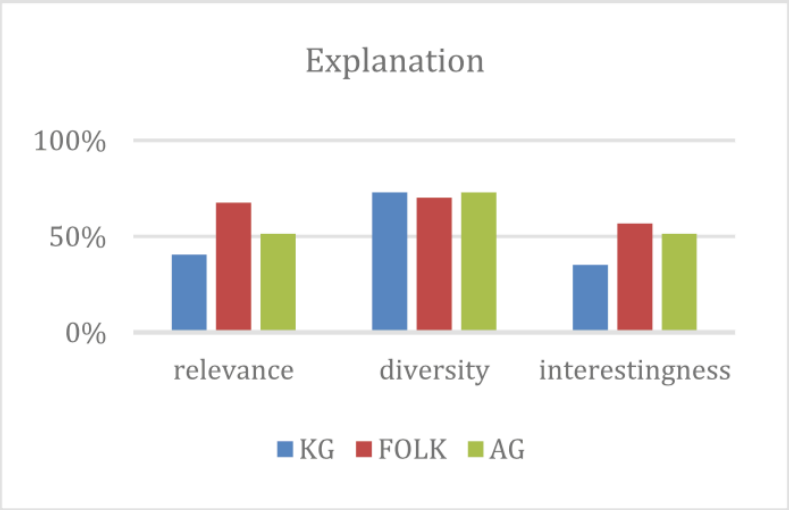
User study

- Comparing 3 approaches Affinity graph (AG) with KG and FOLK on recommendation and explanation tasks
- Protocol: Simulation -> Browsing -> Submitting cities -> Rating recommendations and explanations on 5-point Likert scale on relevance, diversity, novelty/interestingness

You submitted:	You might like:	We recommend you:
dbr:Rome	dbc:Clothing	dbr:The_Hague
dbr:Florence	dbr:Food	dbr:Haarlem
dbr:Amsterdam	dbr:David_de_Haen	dbr:Naples
	dbr:Italy	dbr:Milan
	dbr:History	dbr:Turin

- 37 participants, 25-38 years old
- Metric: percentage of positive ratings (4 or 5)

Results



- Findings on recommendation confirmed
- On explanation
  - AG boosted by FOLK which is most appreciated
  - Skeptical about KG
    - Too general e.g dbc:Leisure
    - Hard to understand e.g. dbr: China_Record_Corporation

Take-away

- Comparison of the usage of knowledge graph and folksonomy for user-entity affinity assessment with two studies within a travel destination recommendation scenario
- Knowledge graph for better accuracy
- Folksonomy for better diversity and novelty
- Semantic affinity framework to harvest respective advantages
- Affinity graph yields equitable performance for both recommendation and explanation tasks

Thank you!

PARIS SORBONNE UNIVERSITÉ

Sépage PARIS

THE FUTURE OF TRAVEL IS HERE

VISIT OUR NEW WEBSITE AT SEPAGE.COM