S C I E N C E ▪ P A S S I O N ▪ T E C H N O L O G Y

Temporal Effects on Hashtag Reuse in Twitter:

A Cognitive-Inspired Hashtag Recommendation Approach

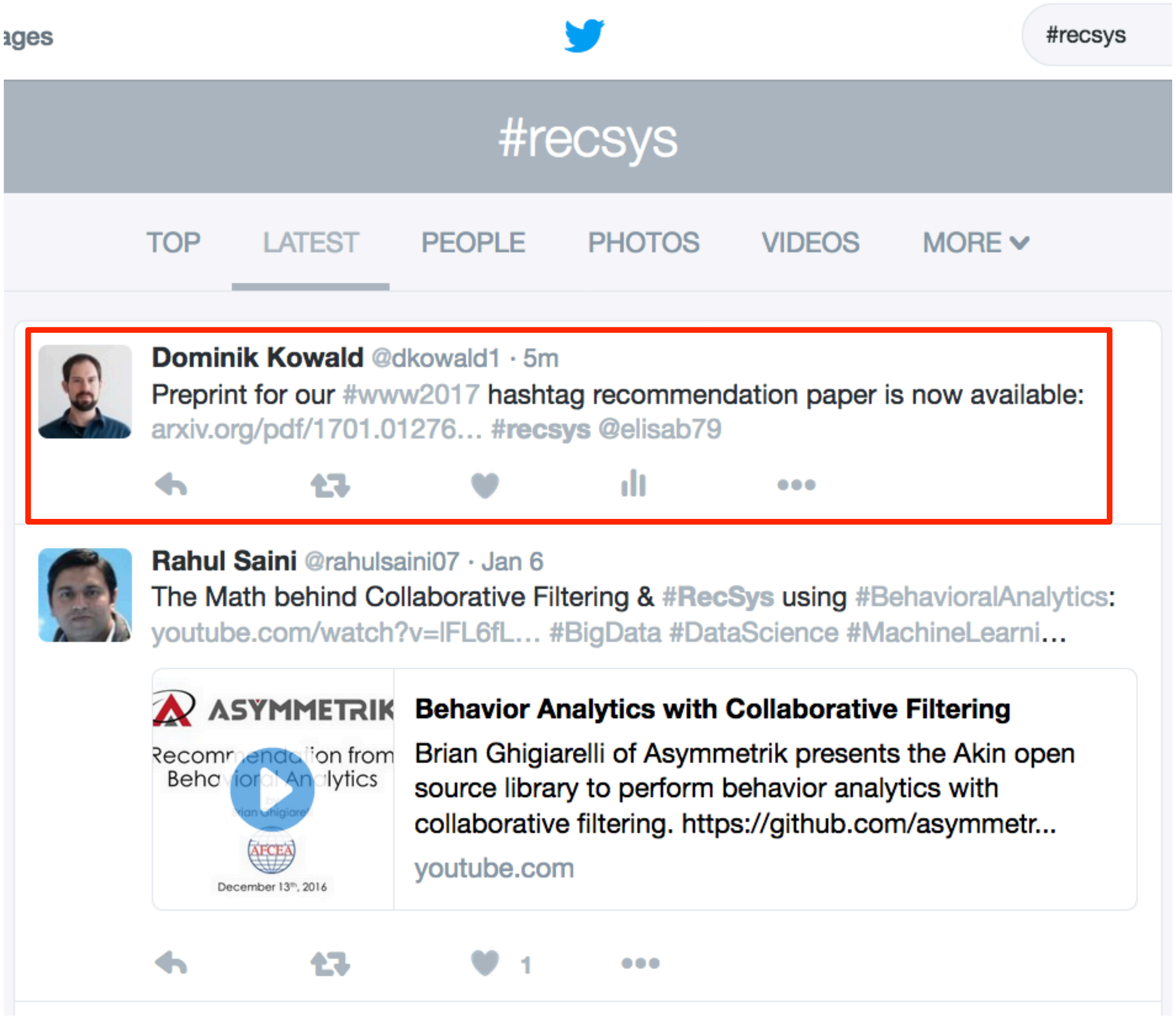
Dominik Kowald, Subhash Pujari & Elisabeth Lex
Know-Center & Graz University of Technology (Austria)

WWW’17, Perth, Australia
April, 7th, 2017

Motivation

- Microblogging platform ***Twitter***
  - Post messages (***tweets***) with max 140 characters
  - Subscribe to tweets of other users (***followees***)
    - Other users subscribe to your tweets (***followers***)
  - Contextualize tweets with freely-chosen keywords (***hashtags***)
  - Hashtags can be searched to receive content of a **specific topic** or event (e.g., *#recsys*)

Motivation (II)



Hashtag Recommendations in Twitter

- ***Scenario 1***: Hashtag rec. w/o current tweet
  - For a given user *u*, **predict the set of hashtags *u* will use next**
  - **Foresee the topics** a user will tweet about
- ***Scenario 2***: Hashtag rec. w/ current tweet
  - For a given user *u* and tweet *t*, **predict the set of hashtags *u* will use to annotate *t***
  - Support a user in **finding descriptive hashtags**
- **We propose an approach for both scenarios**

Our Previous Work: Tag Recommendations based on a Model of Human Memory

- Support users in social bookmarking systems with **tag recommendations** [Kowald et al., 2014)
- **Base-Level Learning (BLL**) equation of the cognitive architecture ACT-R [Anderson et al., 2004]
  - Quantifies the **usefulness of information** (e.g., a word or tag) in human memory

$$B_i = ln(\sum_{j=1}^{n} \boxed{t_j^{-d}})$$

- *Can we also use it for **hashtag recommendations**?*

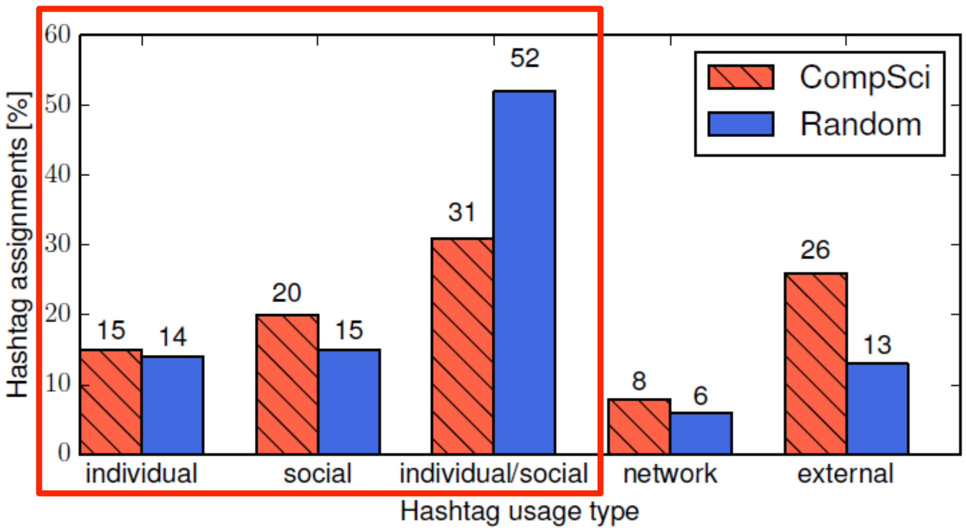
Datasets

- 2 datasets: ***CompSci*** **and** ***Random***
- **Crawling strategy**
  - (i) Crawl seed users [Hadgu & Jäschke, 2014]
  - (ii) Crawl followees
  - (iii) Crawl tweets
  - (iv) Extract hashtag assignments

Dataset	$\lvert U_S \rvert$	$\lvert U \rvert$	$\lvert T \rvert$	$\lvert HT \rvert$	$\lvert HTAS \rvert$
CompSci	2,551	91,776	5,649,359	1,081,403	9,161,842
Random	3,466	127,112	8,157,702	1,507,773	13,628,750

Hashtag Reuse Types

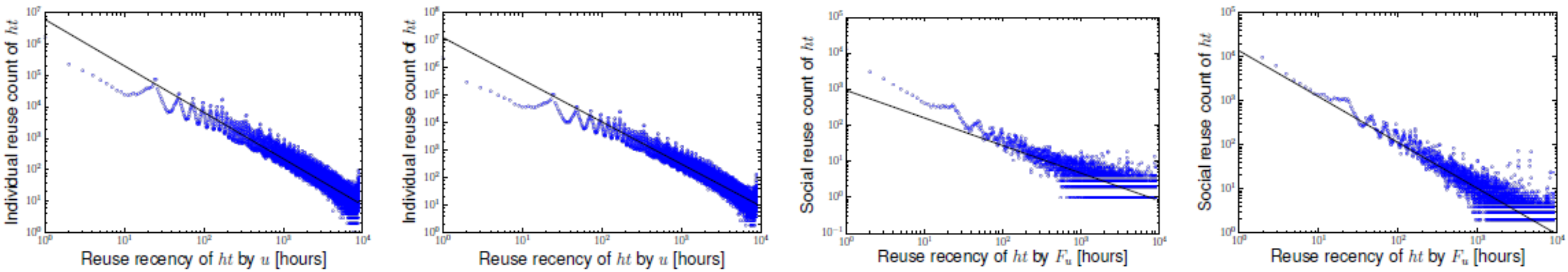
- *How are people **reusing hashtags** in Twitter?*



- 66% and 81% of hashtag assignments can be explained by **individual** or **social** hashtag reuse

Temporal Effects on Hashtag Reuse

- *Do temporal effects have an **influence on individual and social hashtag reuse**?*



(a) Individual hashtag reuse *CompSci* dataset ($R^2 = .883$)
(b) Individual hashtag reuse *Random* dataset ($R^2 = .894$)
(c) Social hashtag reuse *CompSci* dataset ($R^2 = .689$)
(d) Social hashtag reuse *Random* dataset ($R^2 = .771$)

- People tend to reuse hashtags that were **used very recently** by their own or by their followees

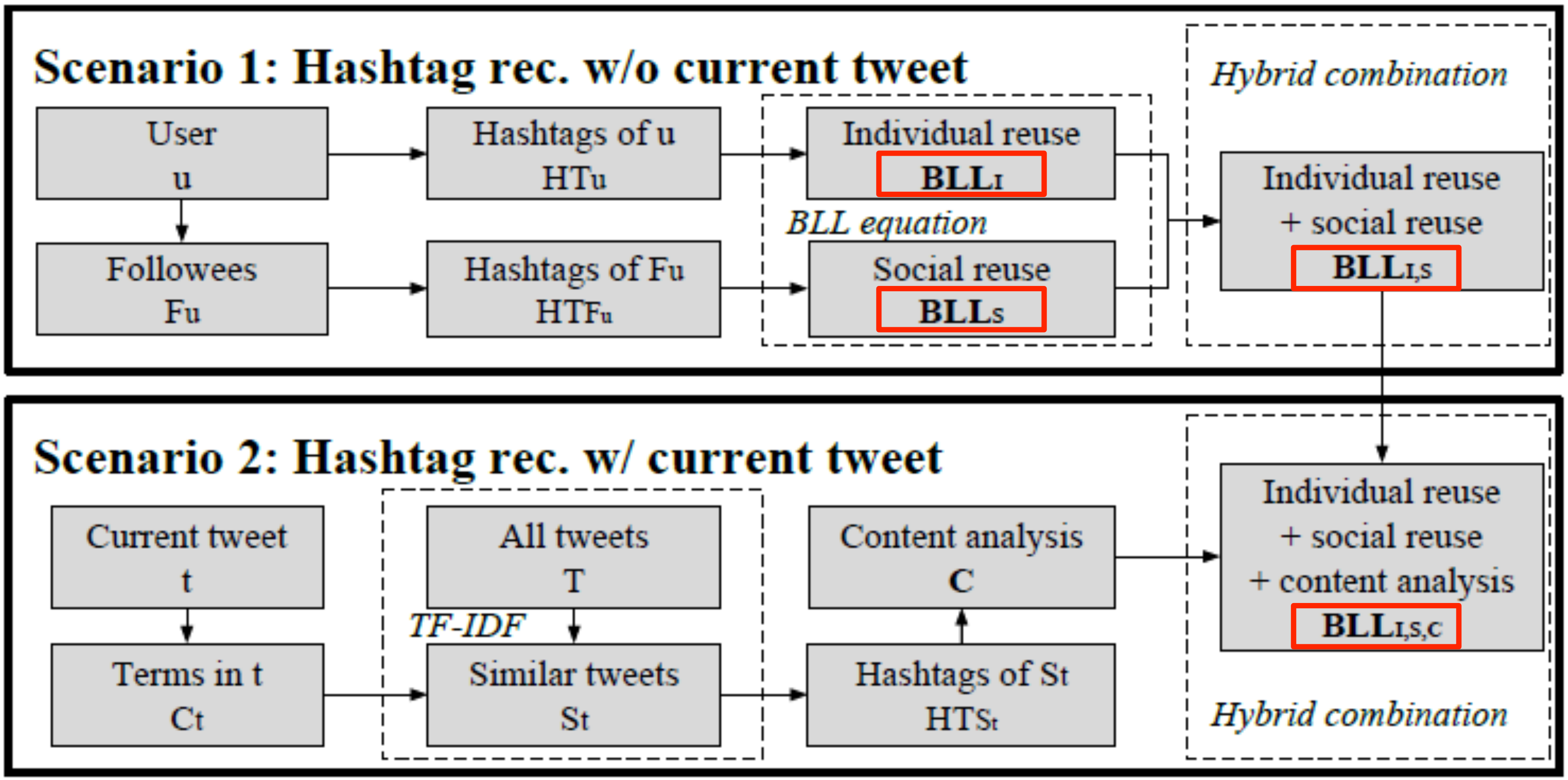
Temporal Effects on Hashtag Reuse (II)

- *Is a **power or an exponential function** better suited to model this time-dependent decay?*
  - Log-likelihood ratio test [Clauset et al., 2009]

Dataset	Parameter	Individual ht reuse	Social ht reuse
CompSci	x_{min}	141	1
	α	1.699	1.242
	R	**188**	**164**
Random	x_{min}	141	1
	α	1.723	1.269
	R	**235**	**294**

- The time-dependent decay of hashtag reuse follows a **power-law distribution** → BLL equation (*d* → *α*)

A Hashtag Recommendation Approach using the BLL Equation



Evaluation

- Evaluation protocol
  - For each seed user, put **most recent tweet into test set** → the rest is used for training
- Evaluation metrics
  - Precision, Recall, F1-score, MRR, MAP, nDCG
- Baseline algorithms
  - MostPopular (MP), MostRecent (MR), FolkRank (FR), Collaborative Filtering (CF), SimRank (SR), **TemporalCombInt (TCI)** [Harvey & Crestani, 2015]
- **TagRec** open-source framework: https://github.com/learning-layers/TagRec

Results (*Scenario 1*)

- *Can we predict the hashtags of **a given user** using the BLL equation?*

Dataset	Metric	MP_I	MR_I	BLL_I	MP_S	MR_S	BLL_S	MP	FR	CF	$BLL_{I,S}$
CompSci	F1@5	.086	.098	**.101**	.022	.076	**.118**	.006	.083	.099	**.153*****
	MRR	.136	.188	**.193**	.032	.122	**.187**	.007	.130	.163	**.268*****
	MAP	.143	.195	**.202**	.033	.128	**.205**	.007	.136	.169	**.285*****
	nDCG	.175	.218	**.225**	.046	.154	**.235**	.012	.169	.196	**.324*****
Random	F1@5	.160	.169	**.175**	.072	.103	**.138**	.012	.159	.165	**.208*****
	MRR	.261	.300	**.314**	.109	.159	**.220**	.023	.260	.278	**.361*****
	MAP	.279	.315	**.335**	.116	.171	**.240**	.024	.279	.296	**.389*****
	nDCG	.323	.352	**.370**	.144	.205	**.280**	.035	.324	.333	**.434*****

- **BLL_I > MP_I, MR_I**
- **BLL_S > MP_S, MR_S**
- **$BLL_{I,S}$ > MP, FR, CF**

Results (*Scenario 2*)

- *Can we predict the hashtags of a **given user and a given tweet** using the BLL equation?*

Dataset	Metric	SR	TCI	$BLL_{I,S,C}$
CompSci	F1@5	.139	.182	**.200***
	MRR	.264	.334	**.395*****
	MAP	.283	.354	**.417*****
	nDCG	.299	.385	**.446****
Random	F1@5	.181	.243	**.261***
	MRR	.341	.436	**.489****
	MAP	.374	.472	**.530****
	nDCG	.388	.507	**.562****

- **TCI, $BLL_{I,S,C}$ > SR**
- **$BLL_{I,S,C}$ > TCI**
- ***Random* dataset > *CompSci* dataset**
  - More ***external*** hashtags in *CompSci* dataset

Conclusion

- **Temporal effects** have an important influence on individual and social hashtag reuse
- A **Power function** is better suited to model this time-dependent decay than an exponential one
- The **BLL equation** provides a suitable model for personalized hashtag recommendations
  - Without ($BLL_{I,S}$) and with the current tweet ($BLL_{I,S,C}$)
- ***Future Work***
  - Incorporate **social connections** (e.g., edge weight)
  - Use additional knowledge source to cope with ***external*** hashtags (e.g., trending hashtags)

Thank you for your attention!

Do you have questions?

Dominik Kowald

- Mail: *dkowald [AT] know-center.at*
- Web: *www.dominikkowald.info*

Subhash Chandra Pujari

- Mail: *subhash.pujari [AT] gmail.com*

Elisabeth Lex

- Mail: *elisabeth.lex [AT] tugraz.at*
- Web: *www.elisabethlex.info*

References

- **[Anderson et al., 2004]** J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin. An integrated theory of the mind. Psychological review, 111(4):1036, 2004.
- **[Clauset et al., 2009]** A. Clauset, C. R. Shalizi, and M. E. Newman. Power-law distributions in empirical data. SIAM review (SIREV), 51(4):661-703, 2009.
- **[Hadgu & Jäschke, 2014]** A. T. Hadgu and R. Jäschke. Identifying and analyzing researchers on twitter. In Proc. of WebSci '14, pages 23-30, New York, NY, USA, 2014.
- **[Harvey & Crestani, 2015]** M. Harvey and F. Crestani. Long time, no tweets! Time-aware personalised hashtag suggestion. In Proc. Of ECIR'15, pages 581-592. Springer, 2015.
- **[Kowald et al., 2014]** D. Kowald, P. Seitlinger, C. Trattner, and T. Ley. Long time no see: The probability of reusing tags as a function of frequency and recency. In Proc. of WWW '14 companion, pages 463-468. ACM, 2014.

Appendix: Formalization of Hashtag Recommendations using the BLL Equation

Modeling hashtag reuse:

$$B_I(ht, u) = \ln(\sum_{j=1}^{n} (TS_{ref} - TS_{ht,u,j})^{-d_I})$$

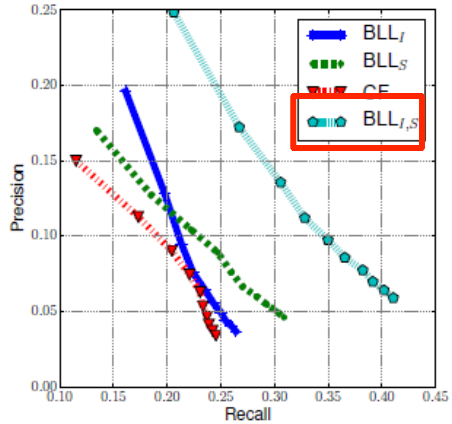
Combining individual and social hashtag reuse (**BLL$_{I,S}$**):

$$B_{I,S}(ht, u) = \beta \underbrace{\sigma(B_I(ht, u))}_{BLL_I} + (1 - \beta) \underbrace{\sigma(B_S(ht, u))}_{BLL_S}$$

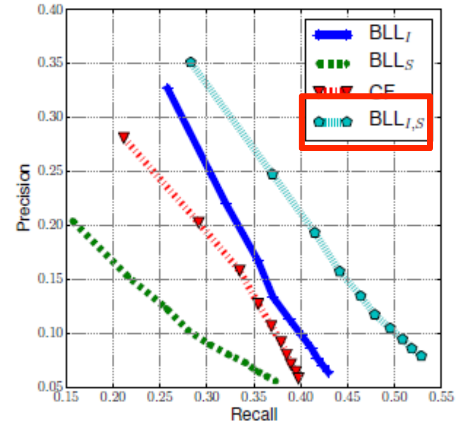
Combining BLLi,s with TF-IDF (**BLL$_{I,S,C}$**):

$$\widetilde{HT}_{u,t} = \underset{ht \in \overline{HT}_{u,t}}{\overset{k}{\arg\max}} (\lambda \underbrace{B_{I,S}(ht, u)}_{BLL_{I,S}} + (1 - \lambda) \underbrace{\sigma(CB(ht, t))}_{C})$$

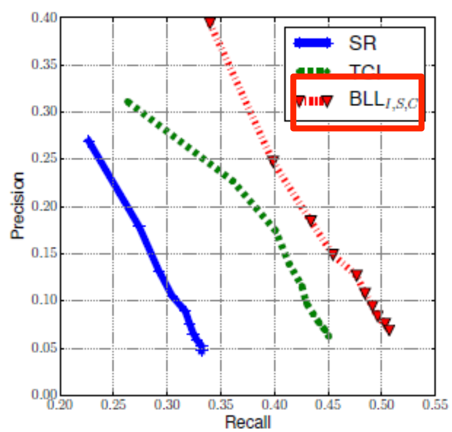
Appendix: Results (Precision / Recall Plots)



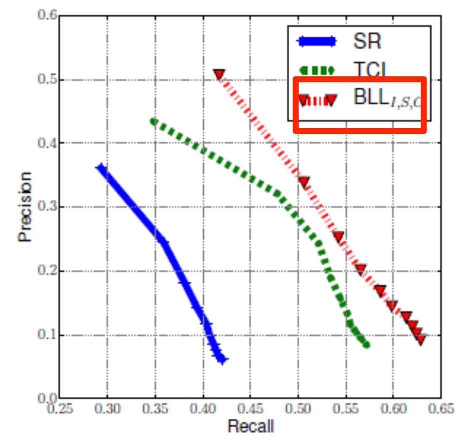
(a) *Scenario 1*: Hashtag recommendation w/o current tweet *CompSci* dataset



(b) *Scenario 1*: Hashtag recommendation w/o current tweet *Random* dataset



(c) *Scenario 2*: Hashtag recommendation w/ current tweet *CompSci* dataset



(d) *Scenario 2*: Hashtag recommendation w/ current tweet *Random* dataset