

Web Science – Investigating the Future of Information and Communication #Science Identifying and Analyzing Researchers on Twitter

Robert Jäschke Asmelash Teka Hadgu

#science, Robert Jäschke



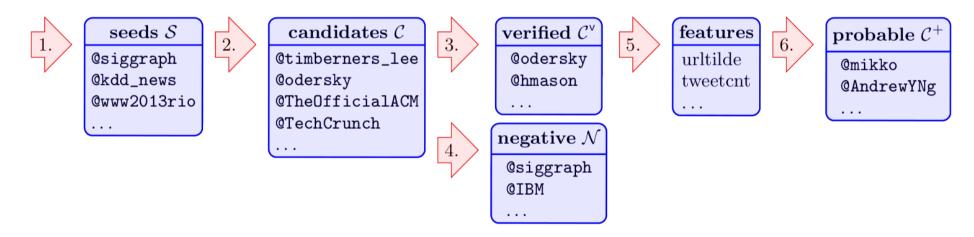


Agenda

Motivation







Results

chief developer intéreste





Twitter is ...

- a communication platform,
- a social network,
- a system for resource sharing
- ... which **researchers** use ...
- to connect with other researchers,
- to announce calls for papers,
- to communicate and discuss,
- to stay up-to-date,
- etc.

TF	REC, informatio
4.987 TWEETS	193 Folgt
ESWC	ESWC Confe #eswc2014 se deadlines: ab conferences. Retweetet v Schließen
	4 RETWEETS
	7:17 AM - 18 No
	Jure Leskov Computing In <u>Schließen</u>
	3 Retweets
	8:30 AM - 1 Nov

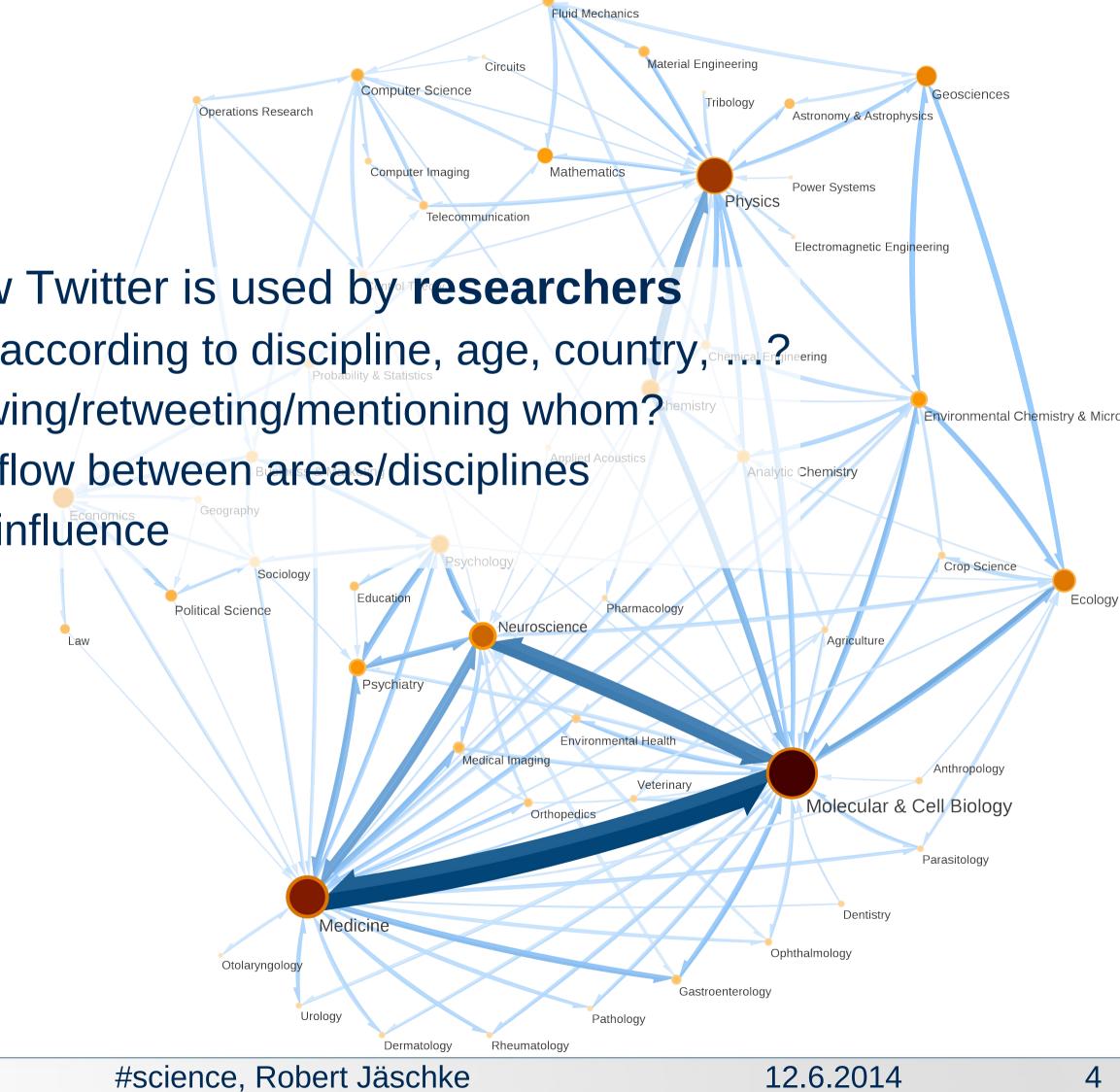




Goals

Understand how Twitter is used by researchers

- Differences according to discipline, age, country, and a series
- Who's following/retweeting/mentioning whom?
- Information flow between areas/disciplines
- Impact and influence





Goals

Improve retrieval and discovery of scientific content

- Researchers, topics, publications, conferences, ...
- Trends, developments over time
- Personalized recommendations



source: twitter.com

#science, Robert Jäschke

content es, ...





Goals

Transfer peer review to social media

What do researchers regard as important?

tric com		Haniar	Dob Dob	ort löo	oble	
5 blogs 527 tweeters 8 Facebook users 8 Google+ users	MONITOR	0	el awards to		oroariou	s science
4 news outlets		2012 pri	zes honor funn	y research wi	th a serious	s twist
Mentioned by	Science NOW Ig Nobels Go To Studies of Swaying Pony Science/AAAS					
show more						
Is one of the highest ever scores in this journal (ranked #2 of 7,374)			eht der Eiffeltu chwnze? Die Ig			siert im Hirn eines tote
Score in context	DER SPIEGEL	Pferde	kte Wissen schwanz-W		Nobelpre	is für Sprachsto
	So far Altmetric I	nas seen 6 s	stories from 4 o	outlets.		
507	News Blogs	Twitter	Facebook	Google+	Score	Demographics
	The shap fiber bund	e of a dles	ponyto	ail a <mark>nd</mark>	the s	statistical p

source: altmetric.com

#science, Robert Jäschke

physics of hair

Help

opper und

oten Fisches, warum wippen

2012-09-21T14:25:00+01:00

ly) Exploding Colons

2012-09-21T22:00:00+01:00



Challenges

- Data acquisition
 - Tweets and users from Twitter
 - Ground truth to train and evaluate algorithms
- Identifying researchers
 - One class problem: finding good counterexamples is difficult
- Brevity of tweets
 - How to extract meaning from 140 characters?
- Identifying and classifying scientific tweets
 - What is a scientific tweet?
- Ranking scientific content
 - How to evaluate a ranking?





Related Work

- Twitter directories (e.g., Wefollow, Twellow, JustTweetIt)
- User classification:
 - D. Rao, D. Yarowsky, A. Shreevats, and M. Gupta. Classifying latent user attributes in **Twitter (2010)**
 - M. Pennacchiotti and A.-M. Popescu. Democrats, republicans and starbucks afficionados: user classification in Twitter (2011)
- Scholars on Twitter:
 - J. Priem and B. Hemminger. Scientometrics 2.0: New metrics of scholarly impact on the social web (2010)
 - J. Letierce, A. Passant, J. Breslin, and S. Decker. Understanding how Twitter is used to widely spread scientific messages (2010)
 - K. Weller, E. Dröge, and C. Puschmann. Citation analysis in Twitter: Approaches for defining and measuring information flows within tweets during scientific conferences (2011)
 - G. Eysenbach. Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact (2011)
- Typically: focus on *tweets*, not users

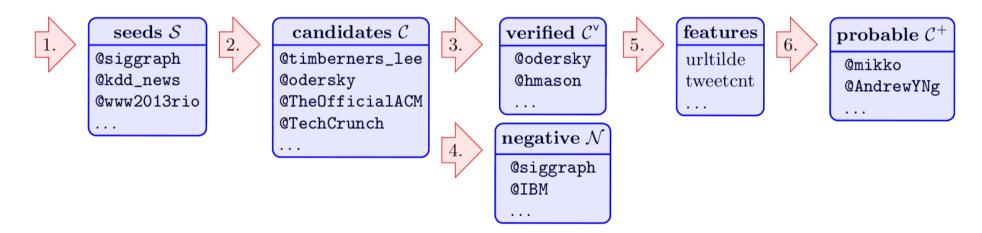


Agenda

Motivation



Approach



Results

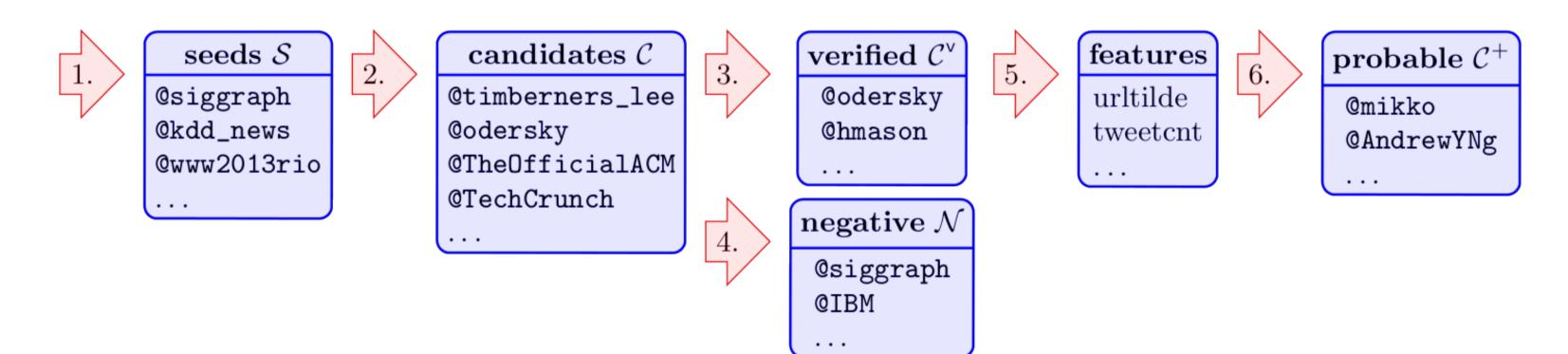
chief developer semantic intéreste

#science, Robert Jäschke





Approach



As a first step, we

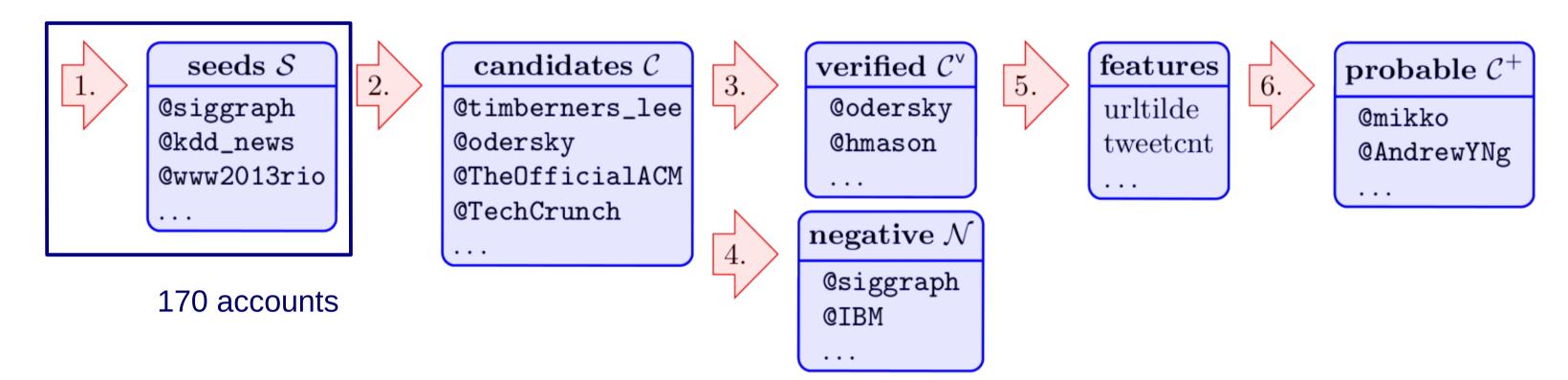
- focused on computer science
- developed a pipeline to *identify* researchers
- analyzed their age, popularity, influence, and social network

12.6.2014

10



Approach



Finding good **seeds**:

- requirements: small set, good coverage, followers likely scientists
- solution: Twitter accounts of computer science conferences
- started with a list from Wikipedia¹, searched for Twitter accounts
- 170 accounts for 98 conferences

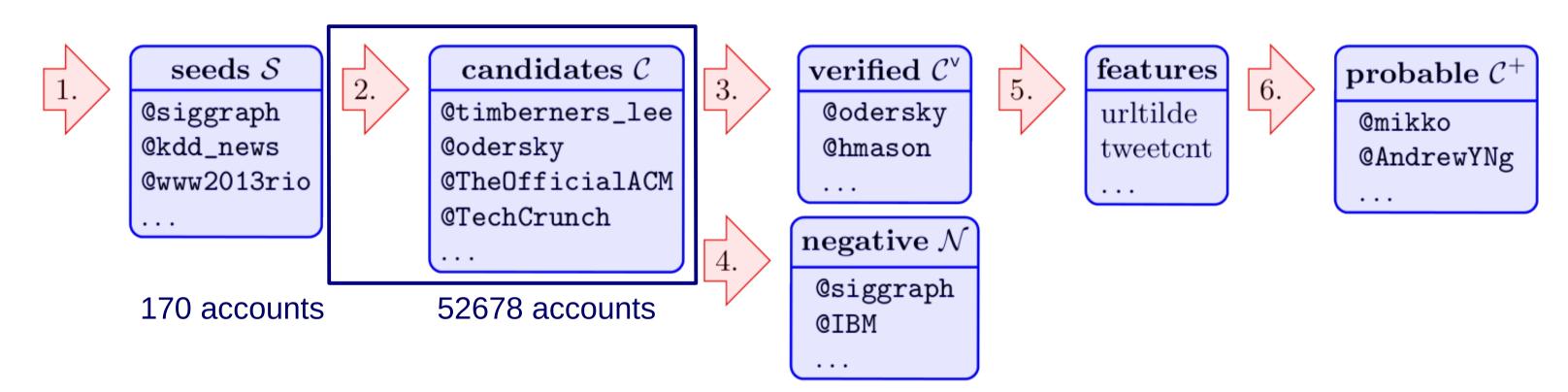
1: http://en.wikipedia.org/wiki/List_of_computer_science_conferences

#science, Robert Jäschke

vers likely scientists e conferences or Twitter accounts



Approach



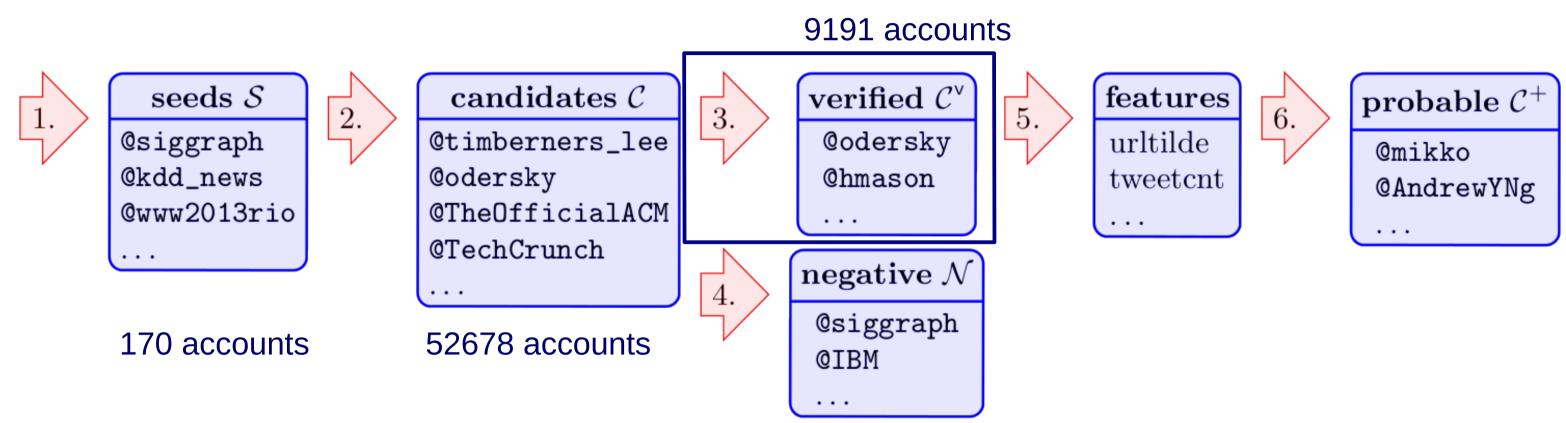
Generating candidates:

- follower, friends, retweeter of the seeds
- recursive approach possible but reduces precision
- 52678 accounts, mostly interested in one conference (83%)

ision ference (83%)



Approach



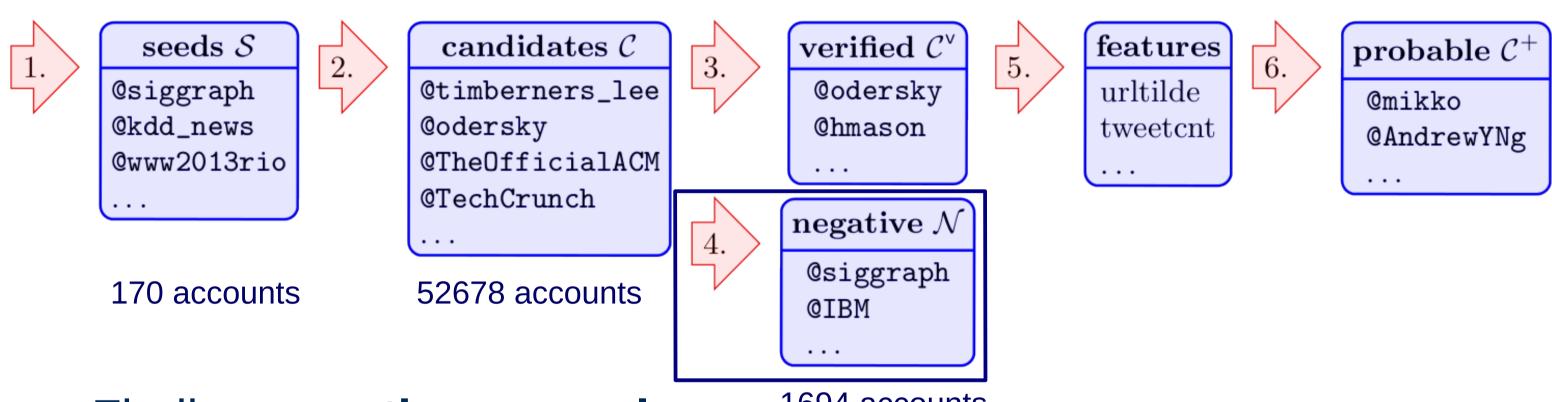
Verifying candidates with ground truth:

- using computer science publications as evidence
- matching against 1304283 author names from DBLP
- matching: string matching of real name, ignoring duplicates
- manual validation of 150 verified accounts: 73% accuracy





Approach



Finding **negative examples**:

1694 accounts

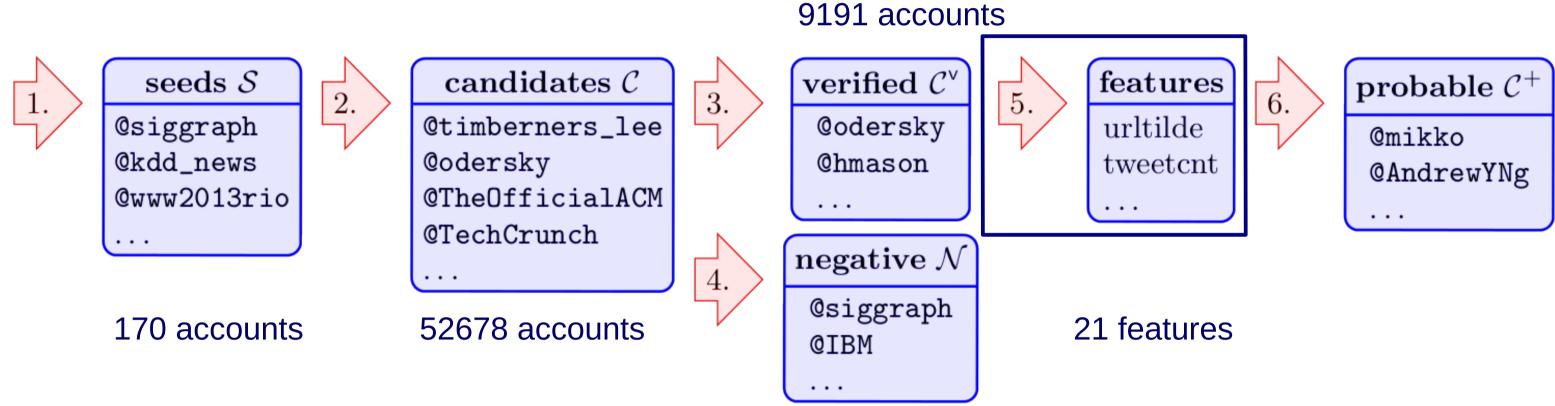
9191 accounts

- challenging task: most users are not researchers
 - How to get a representative sample?
- randomly collected users from the Twitter stream
- removed candidates, their followers and friends
- added seeds and large companies





Approach



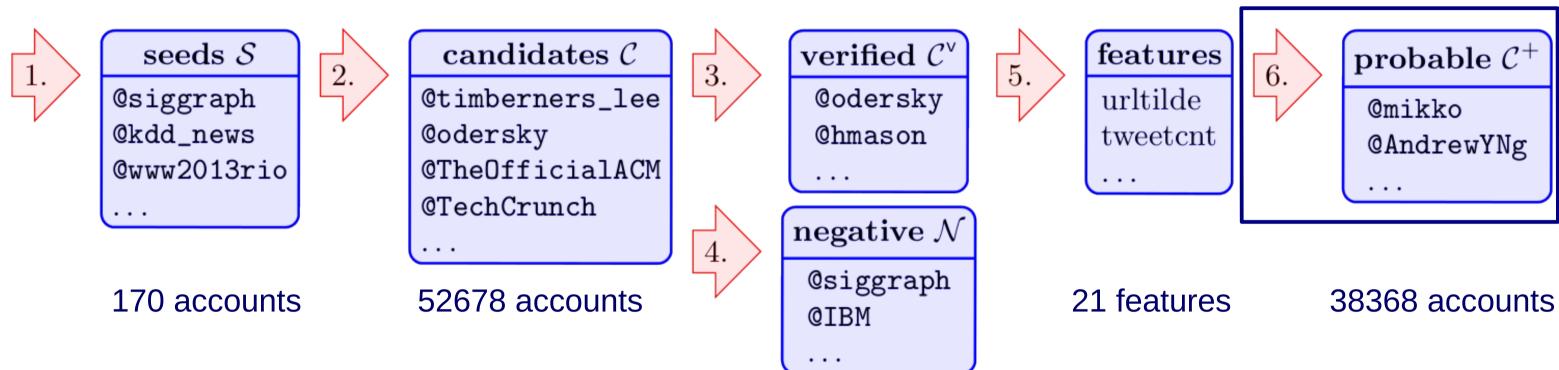
Extracting **features**:

1694 accounts

- Which features can separate researchers from other users?
- profile (#tweets, #followers, website set, bio keywords, etc.) and
- content (#tweets with URLs, #scientific tweets, etc.) features,
- no network (#followed seeds, etc.) features



Approach



Classifying candidates:

1694 accounts

9191 accounts

- stratified 10-fold cross validation (2000 random cand. + neg. ex.)
- Random Forest: F1 of 0.94
- Baseline (SVM on Bag of Words): F1 of 0.93
- 38368 positive candidates, 5015 negative candidates

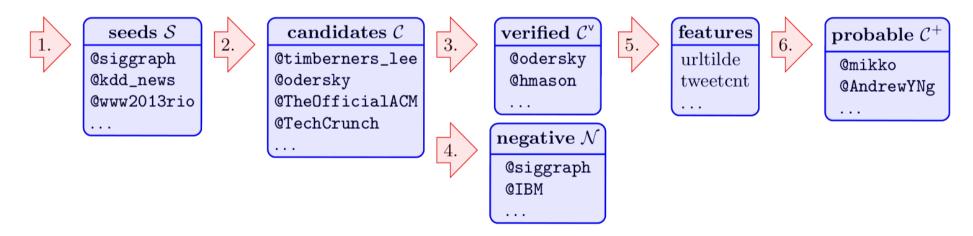


Agenda

Motivation







Results

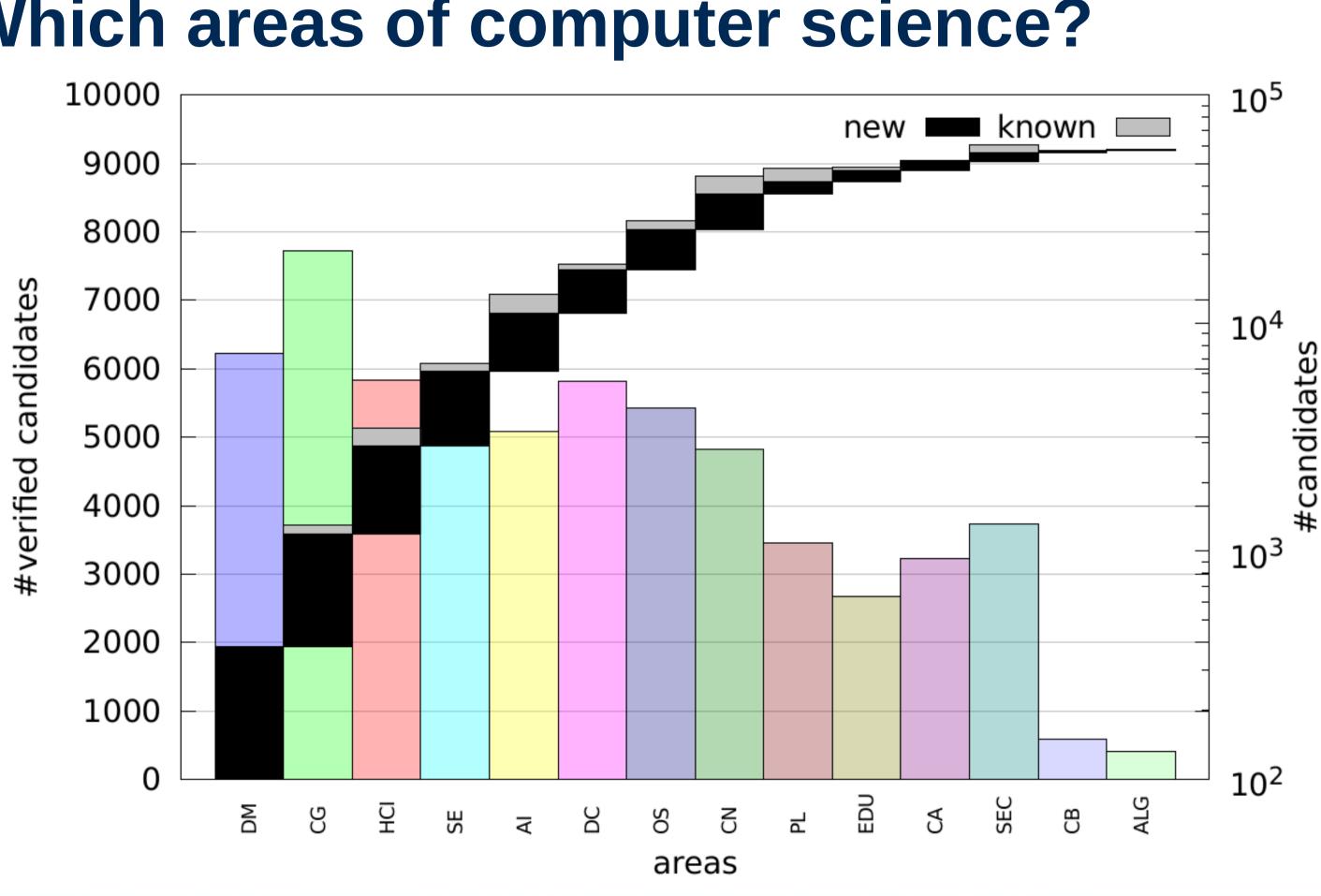
chief developer intéreste

#science, Robert Jäschke





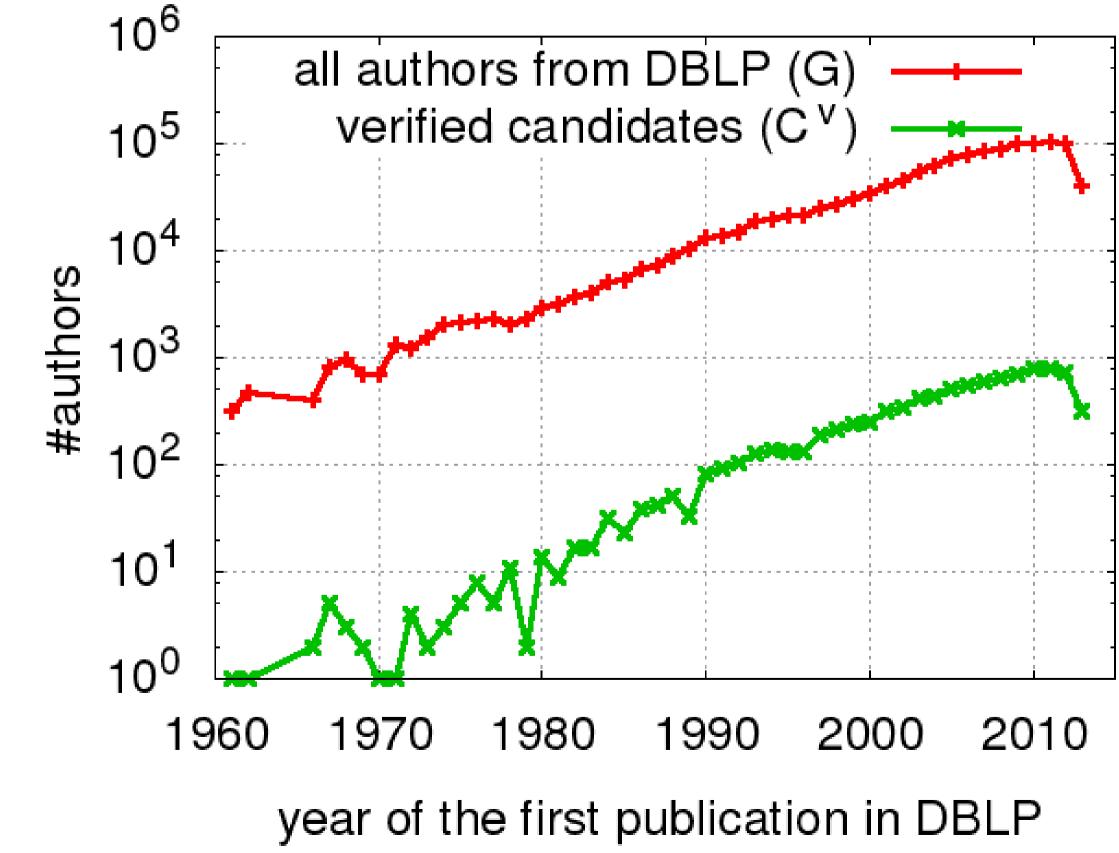
Which areas of computer science?



#science, Robert Jäschke



Are researchers on Twitter younger?

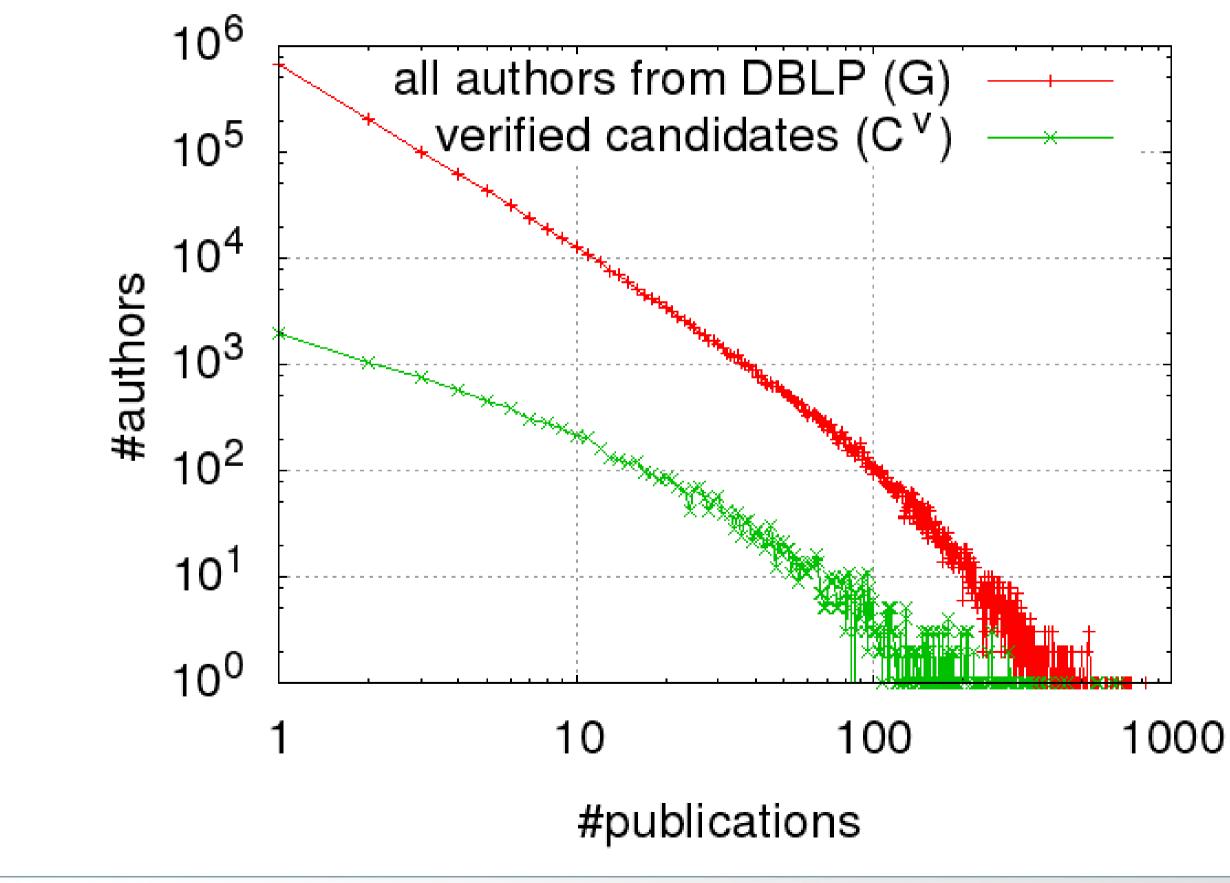


[#]science, Robert Jäschke





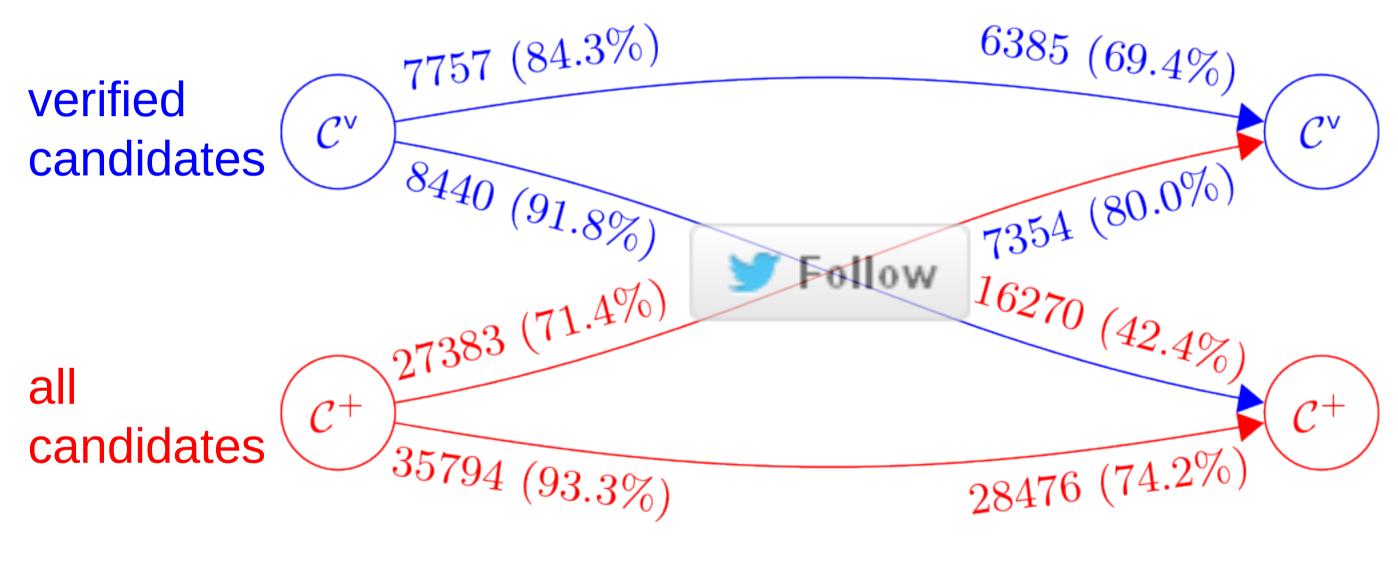
Are they more productive?



#science, Robert Jäschke



How are they connected with each other?



(a) Who follows whom?

in general, the order of activity is follow, mention, retweet

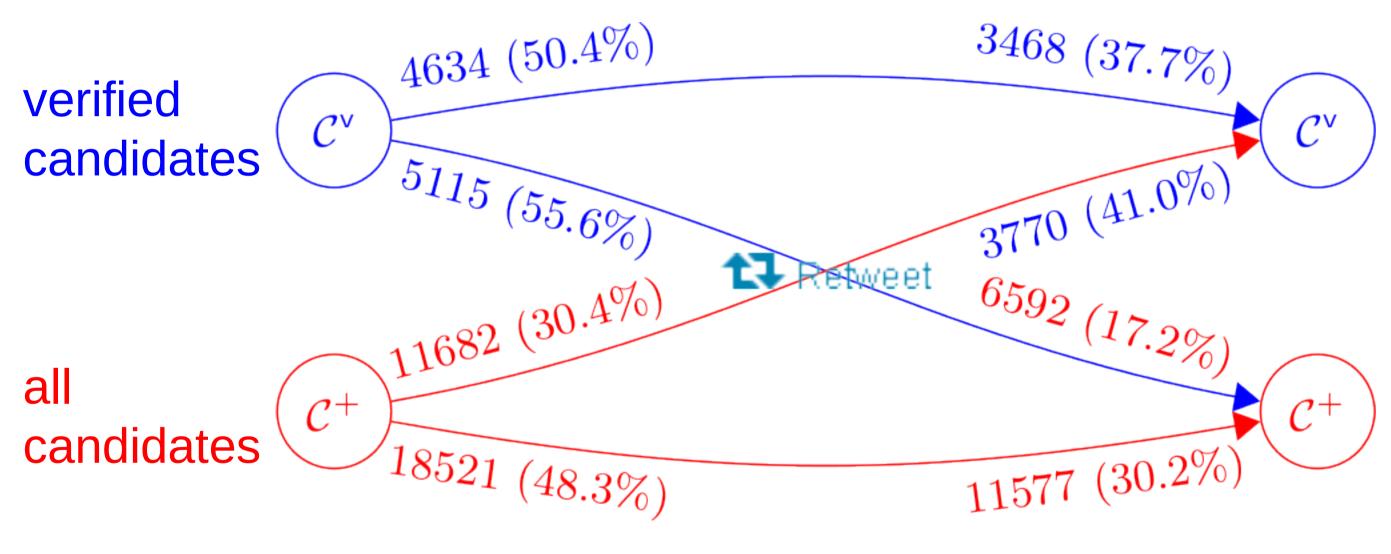
#science, Robert Jäschke

verified candidates

all candidates



How are they connected with each other?



Who retweets whom? (b)

in general, the order of activity is follow, mention, retweet

#science, Robert Jäschke

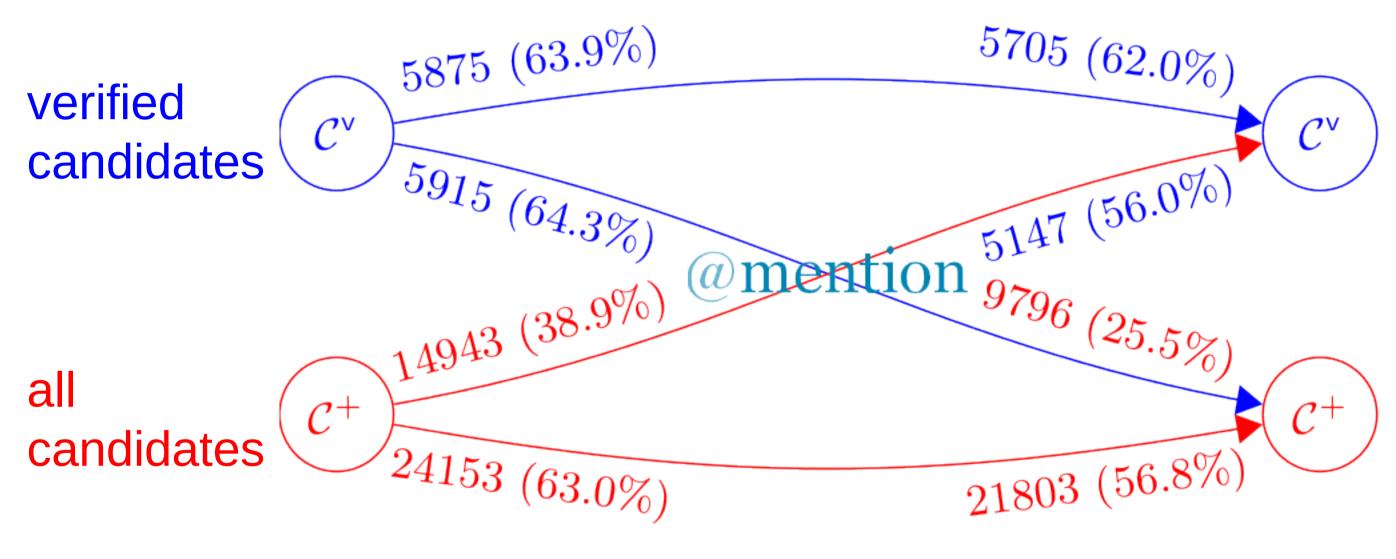
verified candidates

all candidates





How are they connected with each other?



(c) Who mentions whom?

in general, the order of activity is follow, mention, retweet

#science, Robert Jäschke

verified candidates

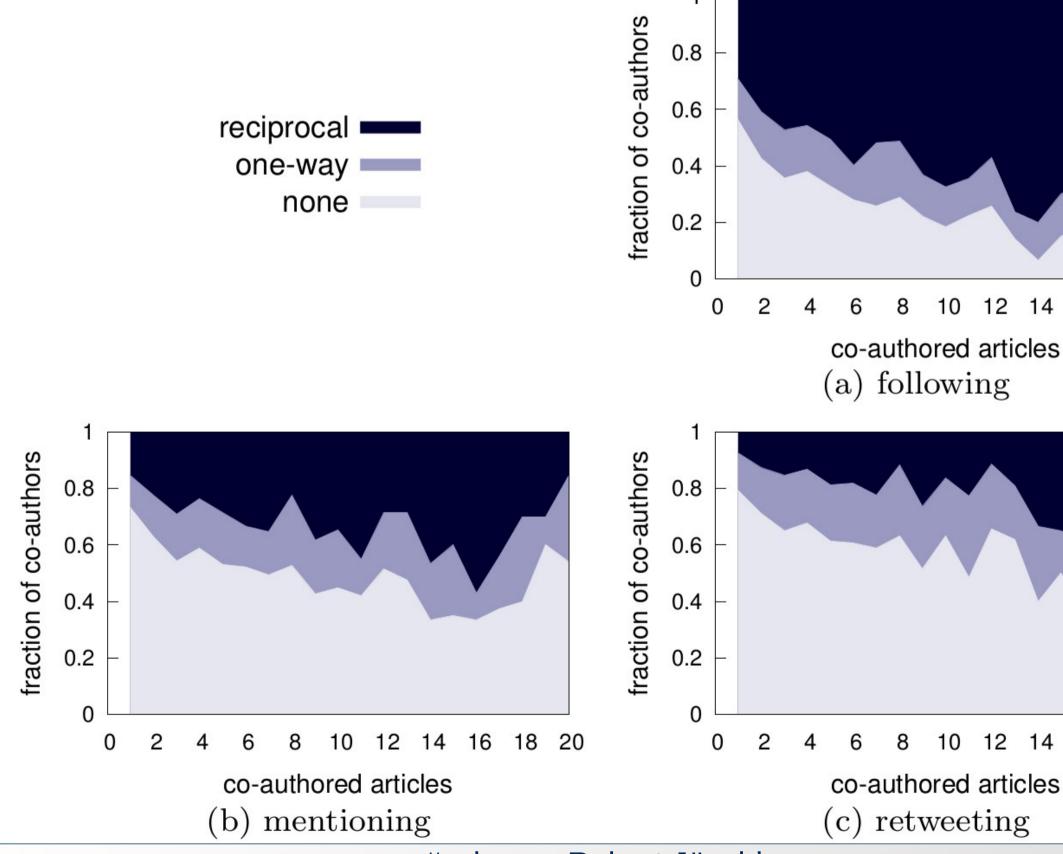
all candidates



Lorschungszent

Web Science – Investigating the Future of Information and Communication

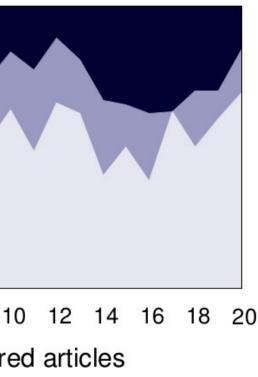
How does closer scientific collaboration affect interaction on Twitter?



#science, Robert Jäschke



21 12 4 16 18







Who are the most influential researchers?

			ranking by				
	screen name	name	r	f	m	р	
_	@timoreilly	Tim O'Reilly	1	2	1	16	
	@billgates	Bill Gates	2	1	2	1	
S	@hmason	Hilary Mason	3	9	3	2	
ate	@zephoria	Danah Boyd	4	7	6	24	
	@csoghoian	Christopher Soghoian	5	51	12	5	
j D	@doctorow	Cory Doctorow	6	16	4	2	
candidates	@ioerror	Jacob Appelbaum	7	30	7	5	
	@mattmight	Matthew Might	8	47	16	34	
verified	@kentbeck	Kent Beck	9	18	17	35	
ij	@mattcutts	Matt Cutts	10	15	9	2	
/el	@timberners_lee	Tim Berners-Lee	11	3	5	35	
	@codepo8	Christian Heilmann	12	87	14	1	
	@mattblaze	Matt Blaze	13	60	25	72	
	@digiphile	Alex Howard	14	42	13	1	
			retu	toll	ma		
Weet W Chtion Dlin							
				-(ation	
@digiphile Alex Howard 10 00 20 12 @digiphile Alex Howard 14 42 13 1 /@two follow follow follow follow follow follow #science, Robert Jäschke 12.6.2014 12.6.2014				.2.6.2014 25			



Who are the most influential researchers?



source: twitter.com

CS prof

#science, Robert Jäschke

top 200 influential researchers other researchers systems scientist

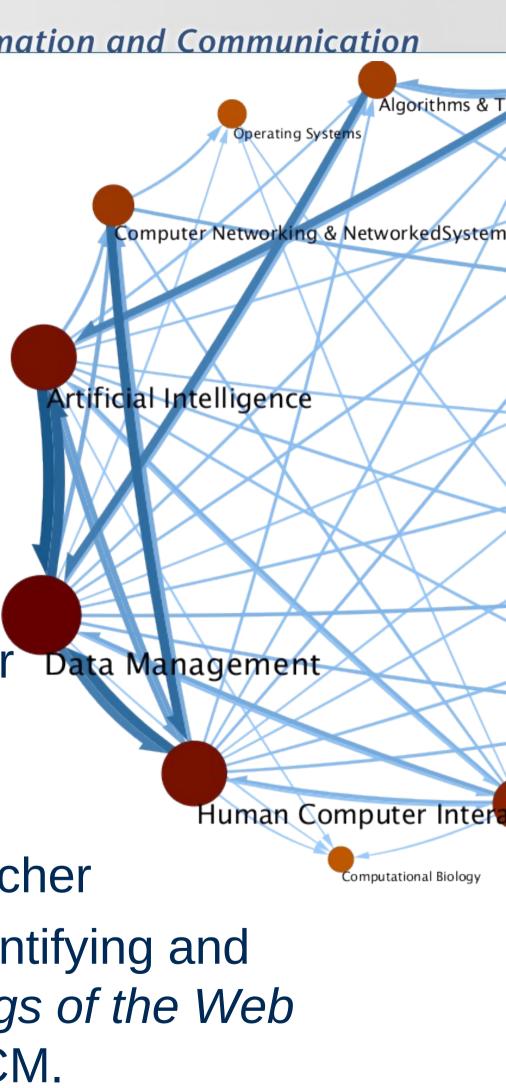




Outlook

- improve matching accuracy
- analyze topics & interests of users
- social network analysis
- transfer to other disciplines
- build a web directory of researchers on Twitter Data Management

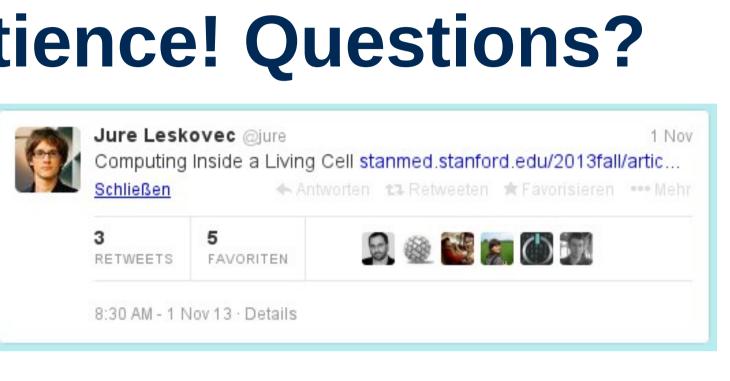
- dataset: https://github.com/L3S/twitter-researcher
- paper: Hadgu, A.T. & Jäschke, R. (2014), Identifying and Analyzing Researchers on Twitter. *Proceedings of the Web Science Conference*, New York, NY, USA: ACM.



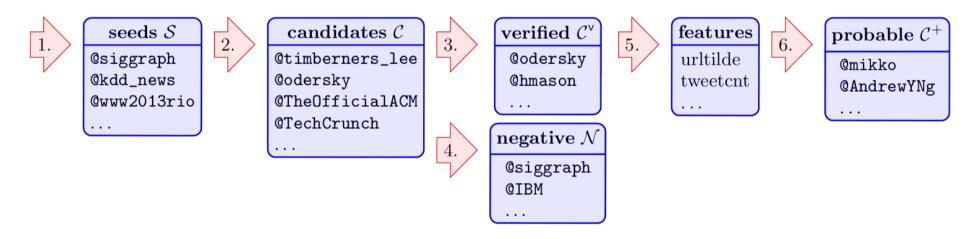


Thanks for your patience! Questions?

Motivation







Results

chief develope

