
Cognos: Crowdsourcing Search for Topic Experts in Microblogs

Saptarshi Ghosh, Naveen Sharma,
Fabricio Benevenuto, Niloy Ganguly, Krishna Gummadi
IIT Kharagpur, India; UFOP, Brazil; MPI-SWS, Germany

Cognos: <http://twitter-app.mpi-sws.org/whom-to-follow/>

Topic experts in Twitter

- ❑ Twitter is now an important source of current news
 - ❑ 500 million users post 400 million tweets daily
 - ❑ Real-time search, news during natural calamities, ...
- ❑ Quality of tweets posted by different users vary widely
 - ❑ News, pointless babble, conversational tweets, spam, ...
- ❑ Challenge: to find topic experts
 - ❑ Sources of authoritative information on specific topics

Identifying topic experts in Twitter

- ❑ Existing approaches
 - ❑ Research studies: Pal [WSDM 11], Weng [WSDM 10]
 - ❑ Application systems: Twitter Who-To-Follow, Wefollow, ...
- ❑ Existing approaches primarily rely on information provided by the user herself
 - ❑ Bio, contents of tweets, network features e.g. #followers
- ❑ We rely on “wisdom of the Twitter crowd”
 - ❑ How do others describe a user?

Challenges in designing search system for topic experts in Twitter

- ❑ How to infer topics of expertise of an individual Twitter user?
- ❑ How to rank the relative expertise of users identified as experts on a topic?
- ❑ How to keep the system up-to-date as thousands of new users join Twitter daily?

Challenge 1

**How to infer topics of expertise of
Twitter users?**

Twitter Lists

- ❑ A feature to organize tweets received from the people whom a user is following
- ❑ Create a List, add name & description, add Twitter users to the list
- ❑ Tweets from all listed users will be available as a separate List stream



Pete Cashmore

@mashable NYC / SF

Breaking social media, tech and digital news and analysis from Mashable.com, the top resource and guide for all things web.

Updates from @mashable staff.

<http://mashable.com>

Tweets Favorites Following **Lists**

mashable's lists



@mashable/news

A curated list of news organization's Twitter accounts.



@mashable/tech

Experts and sources to keep up with the latest in tech.



@mashable/design

Tweets and tips from designers.



@mashable/food

Love food? Here are chef's, cooks and others in food to follow



@mashable/celebrity

Celebrities on Twitter.



@mashable/journalism

Journalists interested in the future of news media.



@mashable/music

Musicians on Twitter.



nytimes The New York Times

Where the Conversation Begins. Follow breaking news, NYTimes.com home page articles, special features and more.



101Cookbooks 101 Cookbooks

Heidi Swanson from 101Cookbooks.com - Healthy, vegetarian recipes made from natural foods and seasonal produce.



epicurious epicurious

Written by Tanya Steel and the Epicurious editorial staff



LATimesfood LA Times Food

News, recipes + reviews from the LA Times Food staff, test kitchen + Daily Dish blog, by @renelynch.



TylerFlorence Tyler Florence

Chef, Restaurateur, Wine Maker, Cookbook Writer, Shop Keep, Product Designer, Dad.



It's Britney Bitch!



ladygaga Lady Gaga

mother monster

Mining Lists to infer expertise

ashton kutcher ✓
@aplusk

Movies TV by Patty Holmes
59 members

Hollywood by mpc2000
1 members

Stars by Nadine Schultz
Meine Lieblingsstars
4 members

Entertainment by Al Royce
54 members

Celebrities by Ben Elcomb
142 members

Celebs by KING5 Photog Jim
126 members

Hollywood by Praesidian
17 members

- ❑ Collect Lists containing a given user U
- ❑ Identify U's topics from List meta-data
 - ❑ Basic NLP techniques
 - ❑ Extract nouns and adjectives
- ❑ Extracted words collected to obtain a topic document for user
[movies tv hollywood stars entertainment
celebrity hollywood ...]

Details in paper

Dataset

- ❑ Collected Lists of 55 million Twitter users who joined before or in 2009
- ❑ All analyses consider 1.3 million users who are included in 10 or more Lists

Topics extracted from Lists



ChuckGrassley ✓

@ChuckGrassley Iowa

U.S. Senator born, raised and still living in New Hartford, IA.

<http://facebook.com/grassley>

<http://www.youtube.com/SenChuckGrassley>

<http://grassley.senate.gov>

politics, senator, congress, government,
republicans, iowa, gop, conservative



Claire McCaskill ✓

@clairecmc Missouri/ Washington DC

<http://twitter.com/clairecmc>

politics, senate, government, congress,
democrats, Missouri, progressive, women



The Linux Foundation

@linuxfoundation San Francisco, CA

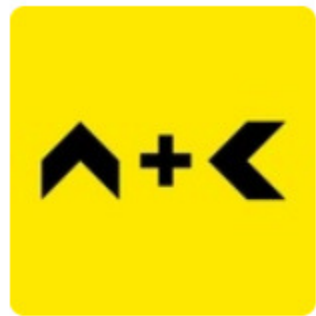
A nonprofit consortium dedicated to fostering the growth of
Linux.

<http://www.linux-foundation.org/>

linux, tech, open, software, libre, gnu,
computer, developer, ubuntu, unix

Cognos: <http://twitter-app.mpi-sws.org/whom-to-follow/>

Lists vs. other features



ashton kutcher ✓

@aplusk

I make stuff, actually I make up stuff, stories mostly, collaborations of thoughts, dreams, and actions. Thats me.

Los Angeles, California · <http://www.facebook.com/Ashton>

Profile bio

love, daily, people, time, GUI, movie,
video, life, happy, game, cool

Most common
words from tweets

celeb, actor, famous, movie, stars,
comedy, music, Hollywood, pop culture

Most common
words from Lists

Lists vs. other features



Profile bio

Fallon, happy, love, fun, video, song,
game, hope, #fjoln, #fallonmono

Most common
words from tweets

celeb, funny, humor, music, movies,
laugh, comics, television, entertainers

Most common
words from Lists

Cognos vs. Twitter Who-To-Follow

- ❑ Top 20 WTF results for 200 queries → 3495 users
- ❑ Do topics inferred from Lists contain that topic for which a user was returned by Twitter WTF?
- ❑ For 83.4%, yes
- ❑ From among the rest 16.6%, manual inspection of a random sample shows two major cases

Case 1 – topics inferred from Lists include semantically similar words, but not exact query-word

Restaurant *dineLA* for query “dining”

Topics from Lists – food, restaurant, recipes, Los Angeles

Space explorer *HubbleHugger77* for query “hubble”

Topics from Lists – science, tech, space, cosmology, NASA

Case 2 – results by Twitter unrelated to query

Comedian *jimmyfallon* for query “astrophysicist”

Topics from Lists – celebs, comedy, humor, actor

Web developer *ScreenOrigami* for query “origami”

Topics from Lists – web design, html, designers

Inferring expertise: Summary

- ❑ List-based method provides accurate & comprehensive inference of topics of expertise of Twitter users
- ❑ In many cases, more accurate than existing approaches that utilize profile information or tweets

Challenge 2

How to rank experts on a topic?

Ranking experts

- ❑ Used a ranking scheme solely based on Lists
- ❑ Two components of ranking user U w.r.t. query Q
 - ❑ Relevance of user to query – cover density ranking between topic document T_U of user and Q
 - ❑ Popularity of user – number of Lists including the user

Topic relevance(T_U, Q) \times log(#Lists including U)

Cognos

- ❑ Search system for topic experts in Twitter
- ❑ Given a query (topic)
 - ❑ Identify experts on the topic using Lists
 - ❑ Rank identified experts

politics

Search

Location Filter : Worldwide

Search results for "politics"



BarackObama : Barack Obama ✓

This account is run by #Obama2012 campaign staff. Tweets from the President are signed -BO.



whitehouse : The White House ✓

Official WH twitter account. Comments & messages received through official WH pages are subject to the PRA and may be archived. Learn more <http://wh.gov/privacy>



algore : Al Gore ✓



maddow : Rachel Maddow MSNBC ✓

I see political people...



billmaher : Bill Maher ✓

The Official Bill Maher Twitter



nprpolitics : NPR Politics

Political coverage and conversation from @NPRnews

Cognos
results for
"politics"

Cognos: <http://twitter-app.mpi-sws.org/whom-to-follow/>

stem cell

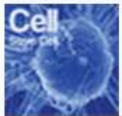
Location Filter : Worldwide

Search results for "stem cell"



StemCellTracker : Stem Cell Tracker

The #1 resource for up-to-date stem cell news.



CellStemCell : Cell Stem Cell

A monthly journal at the forefront of stem cell science and policy. Editor: Dr. Deborah Sweet



StemCellNetwork : Stem Cell Network

SCN supports cutting-edge projects that translate Canadian stem cell research discoveries into new and better treatments.



iPSCellNews : Stem Cell (iPS) News

Latest updates about reprogramming & induced pluripotent stem cell research. -Jordan Kho, PhD student at Baylor College of Medicine-



ATStemCell : All Things Stem Cell

Blog discusses stem cells in multifaceted manner: history, apps, probs, news, & more. Run by PhD stem cell grad student/science writer- Need something written?

Cognos
results for
"stem cell"

Cognos: <http://twitter-app.mpi-sws.org/whom-to-follow/>

Evaluation of Cognos

- System deployed and evaluated 'in-the-wild'
- Evaluators were students & researchers from the three home institutes of authors


User-evaluation of Cognos

Please rate the top 10 results for "science news"



bbcscitech : BBC SciTech



Set up by @mario, supported by backstage.bbc.co.uk

Relevant  



Reuters_Science : Reuters Science News



From newly charted astronomical anomalies at the far reaches of the universe to the rise of nanotechnology, nobody covers science like Reuters.com.

Relevant  



newscientist : New Scientist

New Scientist is the world's leading science and technology weekly

Relevant  



NatGeo : National Geographic

Since 1888, we've traveled the Earth, sharing its amazing stories with new generations. Official Twitter account of National Geographic.

Relevant  



science : science

Science news from ScienceNewsBlog.com.

Relevant  



NASA : NASA

Relevant  

Cognos: <http://twitter-app.mpi-sws.org/whom-to-follow/>

Sample queries for evaluation

Enter your query

Search

Location Filter : Worldwide

Sample Queries

News : politics sports entertainment science technology business

Journalists : politics sports entertainment science technology business

Politics : conservative news liberal politicians USA / German / Brazilian / Indian politicians

Sports : F1 baseball soccer poker tennis NFL NBA Bundesliga LA Lakers

Entertainment : celebrities movie reviews theater music

Hobbies : hiking cooking chefs traveling photography

Lifestyle : wine dining book clubs health fashion

Science : biology astronomy computer science complex networks

Technology : iPhone mac linux cloud computing

Business : markets finance energy

Cognos: <http://twitter-app.mpi-sws.org/whom-to-follow/>

Evaluation results

- ❑ Overall 2136 relevance judgments
 - ❑ 1680 said relevant (78.7%)
- ❑ Large amount of subjectivity in evaluations
 - ❑ Same result for same query received both relevant and non-relevant judgments
 - ❑ E.g., for query “cloud computing”, Werner Vogels got 4 relevant judgments, 6 non-relevant judgments

Evaluation results

- ❑ Considered only the results evaluated at least twice
- ❑ Result said to be relevant if voted relevant in the majority of evaluations
- ❑ Mean Average Precision considering top 10 results: 93.9 %

Details in paper

Cognos vs. Twitter Who-To-Follow

Please compare the anonymized search results for "iphone"

A is better B is better Both are equally good Both are equally bad

Search Engine A



iPhone_News : iPhone News
iPhone news and notes from around the web.



AppStore : App Store
Follow us for official App Store tweets including our featured apps, exclusive offers, and more.



iphone_dev : iPhone Dev
The iPhone Dev Team. We liberate your iPhones!



freeiphoneapps : Free iPhone Apps
We're currently giving away the new iPad! We also give away lots of cool paid iPhone, iPad, and Mac applications on a daily basis. FOLLOW US.



iphonhackx : iPhone Hacks
Unlock the potential of your iPhone, iPad and iPod Touch

Search Engine B



p0sixninja : Joshua Hill
iPhone Hacker



iH8sn0w : iH8sn0w
I'm the guy that made f0recast, iREB, iFaith, and sn0wbreeze. I also make #pie and dj on the side.



chronicdevteam : Chronic Dev
Hax



MuscleNerd : MuscleNerd
iPhone hacker



iPhone_News : iPhone News
iPhone news and notes from around the web.

Cognos: <http://twitter-app.mpi-sws.org/whom-to-follow/>

Cognos vs. Twitter Who-To-Follow

- ❑ Considering 27 distinct queries asked at least twice
- ❑ Judgment by majority voting

- ❑ Cognos judged better on 12 queries
 - ❑ Computer science, Linux, Mac, Apple, Ipad, Internet, Windows phone, photography, political journalist, ...

- ❑ Twitter Who-To-Follow judged better on 11 queries
 - ❑ Music, Sachin Tendulkar, Anjelina Jolie, Harry Potter, metallica, cloud computing, IIT Kharagpur, ...

Results for query music

De-anonymized search results for "music"

Search Engine A : Our results



katyperry : Katy Perry

i kissed a girl AND diddled her skittle.



ladygaga : Lady Gaga

mother mons^{ter}



taylorswift13 : taylorswift13



jtimberlake : Justin Timberlake

Official Justin Timberlake Twitter.



Pink : P!nk

it`s all happening

Search Engine B : Twitter results



iTunesMusic : iTunes Music

Official music updates for the U.S. iTunes Store including new releases, pre-orders, iTunes LP, exclusive offers and more.



guardianmusic : Guardian music

Squashing music into 140 characters since 2008



yahoo_music : Yahoo! Music

The official Twitter account of Yahoo! Music. We tweet about music news, concerts, performances, videos, and all the things that make us yodel!



SonyMusicGlobal : Sony Music Global

The home of Sony Music on Twitter!



CountryMusic : Country Music Associ

Official Tweet of the Country Music Association (CMA) managed by @bennett49r & @chappedman.

Cognos: <http://twitter-app.mpi-sws.org/whom-to-follow/>

Challenge 3

How to keep the system up-to-date by discovering newly joining experts?

Scalability problem

- ❑ Twitter now has around 500 million users
- ❑ 740K new users join daily

- ❑ Twitter restricts crawling through API
 - ❑ Only 3.6K API calls per day per IP

- ❑ Brute-force crawl of all users is infeasible

Solution

- ❑ Only 1.1% users are listed 10 or more times
 - ❑ If experts can be identified efficiently, possible to crawl their Lists
- ❑ Used hubs to identify authorities / experts
 - ❑ Hubs – users who selectively List many experts
 - ❑ Identify hubs using HITS, crawl Lists created by top hubs
 - ❑ 50% of users listed by top 2% hubs listed 10 or more times

Details in paper

Conclusion and Future work

- ❑ Developed and deployed Cognos
 - ❑ Uses crowdsourced Lists to infer topics of expertise as well as to rank experts
 - ❑ Competes favorably with Twitter Who-To-Follow and state-of-the-art research system
- ❑ Lists vital in searching for topic experts in Twitter
- ❑ Future work
 - ❑ Cognos still under development
 - ❑ Make the inference methodology robust against List spam

portland

Location Filter : Worldwide

Search results for "portland"



Powells : Powell's Books

The legendary independent bookstore in Portland, Oregon.



MayorSamAdams : Mayor Sam Adams

Mayor of Portland, Oregon, focused on public safety, education, jobs and economic development, arts, transportation, planning and sustainability.



travelportland : Travel Portland

We love all things Portland - the official account for Travel Portland. Questions? Use #inpx in your tweet. We will help.



Oregonian : The Oregonian

News updates from the largest news org in the Pacific NW. See our lists for reporters and other accts to follow: <http://ORne.ws/TWITlist>. Managed by @AliManzano



PdxPipeline : PDX Pipeline

Portland Music, Art, Events, Entertainment, Sustainability. Contact us: pdxpipeline@gmail.com. Links: <http://PDXPipeline.com> & <http://facebook.com/pdxpipeline>



PDXfoodcarts : FoodCartsPortland

We document all the wonderful food carts in Portland Oregon

Thank You

Questions / suggestions?

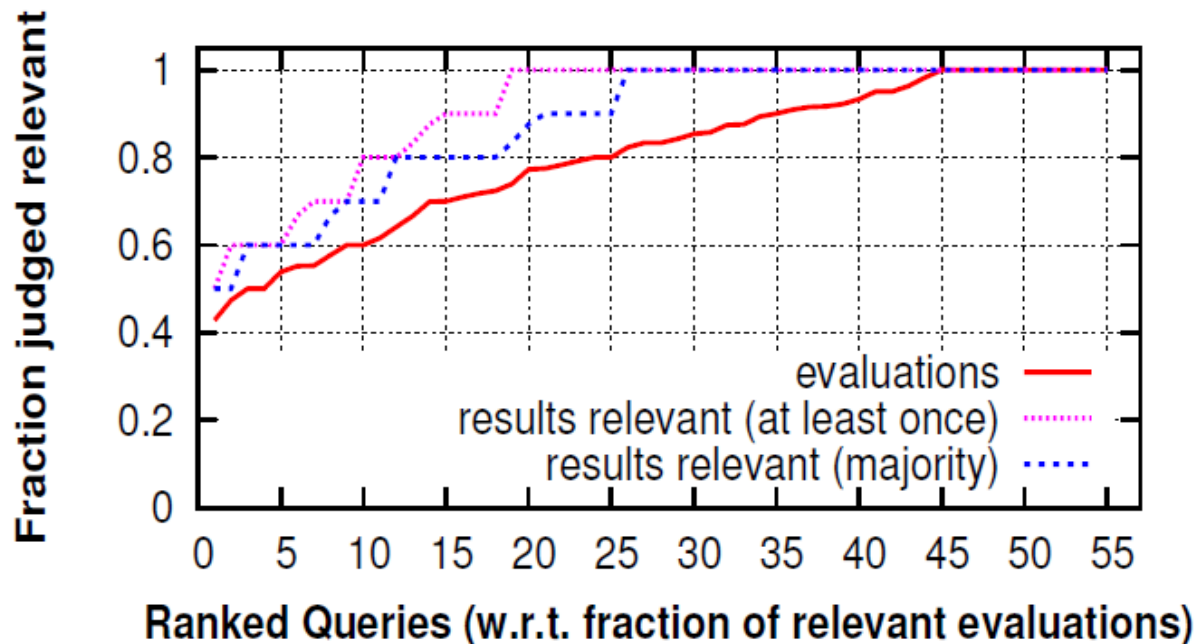
Cognos: <http://twitter-app.mpi-sws.org/whom-to-follow/>



Backup slides

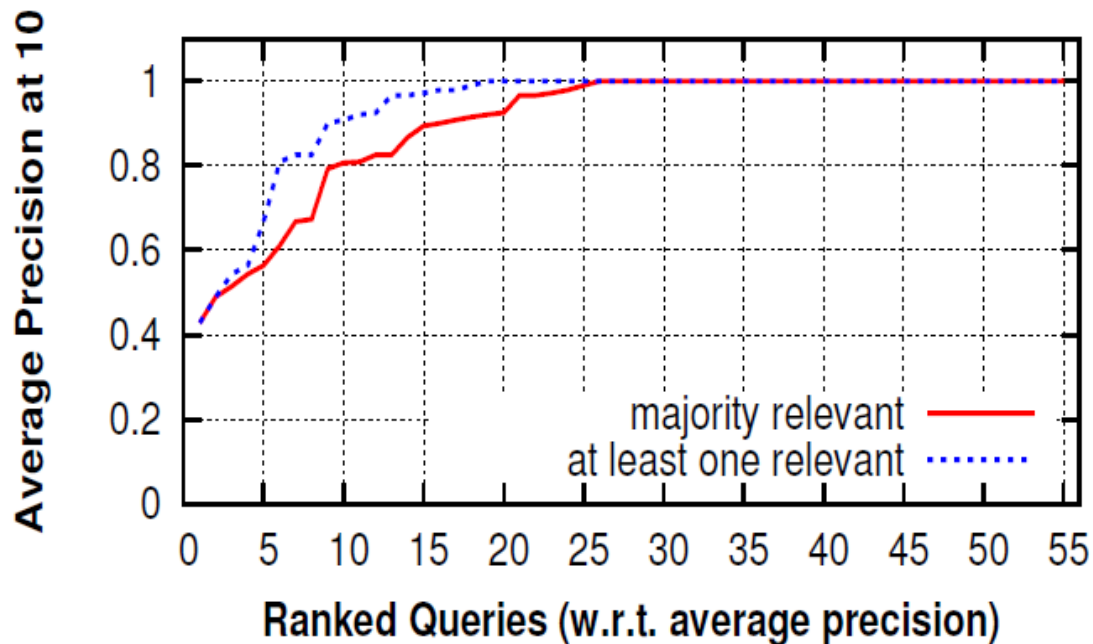
Evaluation results (contd.)

- For each query, measure
 - What fraction of evaluations judged a result as relevant?
 - What fraction of top 10 results judged relevant in (i) at least one evaluation? (ii) the majority of evaluations?



Evaluation results (contd.)

- Average precision at 10 (for top 10 results)
 - Result is relevant if so judged in majority of evaluations
 - Result is relevant if so judged in at least one evaluation



MAP scores:

Case 1: 0.905

Case 2: 0.939

How many experts in Twitter?

- ❑ Estimate number of highly listed experts
 - ❑ Twitter assigns integer userids starting from 1
 - ❑ 300K random integers in [1, 465M]
 - ❑ Distribution of experts in random sample expected to be similar to distribution in global population
 - ❑ Only 1.1% listed 10 or more times, 0.12% listed 100 or more times
 - ❑ If experts can be identified efficiently, possible to crawl their Lists
-

Identifying experts efficiently

- ❑ Hubs – users who follow many experts and add them to Lists
 - ❑ Identified top hubs in social network using HITS
 - ❑ Crawled Lists created by top 1 million hubs
 - ❑ Top 1M hubs listed 4.1M users
 - ❑ 2.06M users included in 10 or more Lists (50%)
 - ❑ Only 1.13% of the entire population estimated to be included in 10 or more Lists
 - ❑ Discovered 65% of the estimated number of experts listed 100 or more times
-

Identifying experts efficiently (contd.)

- ❑ More than 42% of the users listed by top hubs have joined Twitter after 2009
 - ❑ Discovered several popular experts who joined within the duration of the crawl
 - ❑ All experts reported by Pal et. al. discovered
 - ❑ Discovered all Twitter WTF top 20 results for 50% of the queries, 15 or more for 80% of the queries
-

Use Lists to infer expertise: intuition

- If U is an expert on a certain topic
 - U will be included in several Lists created by others
 - List names / descriptions indicate topics of expertise of U



Barack Obama ✓
@BarackObama Washington, DC

9,245,117
Followers

149,504
Listed



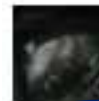
@flyingpackman/politics



@RobertKunz/celebrities



@marcelbedard/personalities



@LisaHathaway/government



@matthew_paul/politicians

Cognos: <http://twitter-app.mpi-sws.org/whom-to-follow/>

Mining Lists to infer expertise

- ❑ Collect Lists which contain a given user U
- ❑ Identify U 's topics from List names and descriptions
 - ❑ NLP – stop word removal, case folding, ...
 - ❑ Identify frequently occurring nouns and adjectives
- ❑ Gives for a user: a topic vector
 - ❑ Weight for a topic: frequency of occurrence of the topic in List meta-data

Evaluating topics inferred from Lists

- ❑ Evaluated through a user-survey
 - ❑ An evaluator shown 30 top topics for a set of Twitter users
 - ❑ Are the topics (i) accurate (ii) informative?
- ❑ 93% evaluations said both accurate and informative
 - ❑ Every user who received negative evaluations received more positive evaluations than negative ones

Details in “Inferring Who-is-Who in the Twitter Social Network”,
Workshop on Online Social Networks (WOSN) 2012

Lists vs other features



Barack Obama ✓

@BarackObama

This account is run by #Obama2012 campaign staff. Tweets from the President are signed -bo.

Washington, DC · <http://www.barackobama.com>

Profile bio

health, visit, American, vote, event,
Iowa, debate, reform, President

Most common
words from tweets

politics, celeb, government, famous,
President, news, leaders, USA

Most common
words from Lists

Comparison with Pal et. al.

- ❑ Pal et. al. give top 10 results for 3 queries – “iphone”, “oil spill”, “world cup”
- ❑ List-based method infers corresponding topic for 21 out of these experts
- ❑ For the rest – bios reveal these users are not specifically related to query-topic

Comparison with Pal et. al.

- Top results by Pal for whom topics inferred from Lists do not contain the query-topic

User	Extracts from Bio
Query: iphone	
teedubya macTweeter	Social Strategy Shaman, SEO <i>Account no longer exists in Twitter</i>
Query: oil spill	
Reuters CBSNews TIME huffingtonpost	latest news from around the world official Twitter feed of CBS News Breaking news and current events The Internet Newspaper
Query: world cup	
nikegoal Flipbooks channel4news	marketing, music, education, sport News, Random Information exclusive stories & breaking news

Comparing Cognos with Pal et. al.

Cognos results			Results by Pal <i>et. al.</i>		
User	Extracts from bio	followers	User	Extracts from bio	followers
Query: oil spill					
BP_America	BP America	35,505	NWF	National Wildlife Federation	76,796
TheOilDrum	energy, peak oil, sustainability	26,257	TIME	Breaking news, current events	3,231,359
GOHSEP	Emergency Preparedness	5,295	huffingtonpost	The Internet Newspaper	1,574,848
usoceangov	National Ocean Service	37,866	NOLAnews	Latest news and updates	29,433
USCG	US Coast Guard	20,513	Reuters	Latest news	1,491,852
Query: iphone					
p0sixninja	iPhone Hacker	127,631	macworld	Mac, iPod, iPhone experts	182,248
iH8sn0w	made f0recast, iREB, iFaith	105,015	Gizmodo	Technologies that change	347,667
chronicdevteam	Hax	107,541	macrumorslive	Updates from Apple events.	170,813
MuscleNerd	iPhone hacker	330,625	macTweeter	<i>Account no longer exists in Twitter</i>	
iPhone_News	iPhone news and notes	153,024	engadget	Twitter account of Engadget	419,583

Cognos: <http://twitter-app.mpi-sws.org/whom-to-follow/>

Tag clouds for well-known users

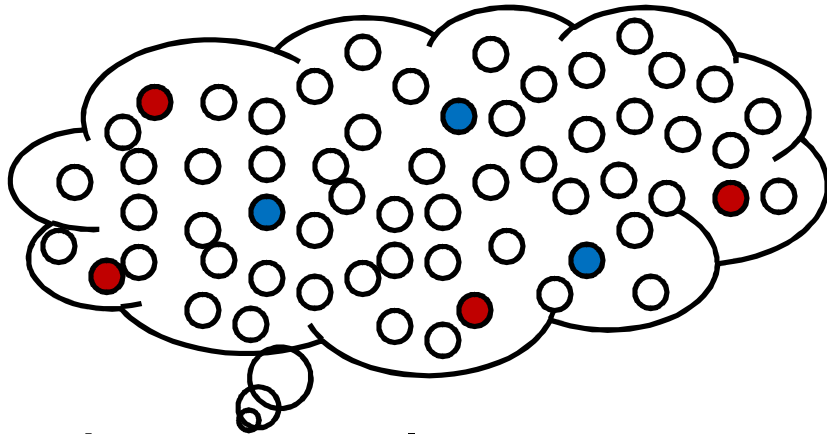


Barack Obama



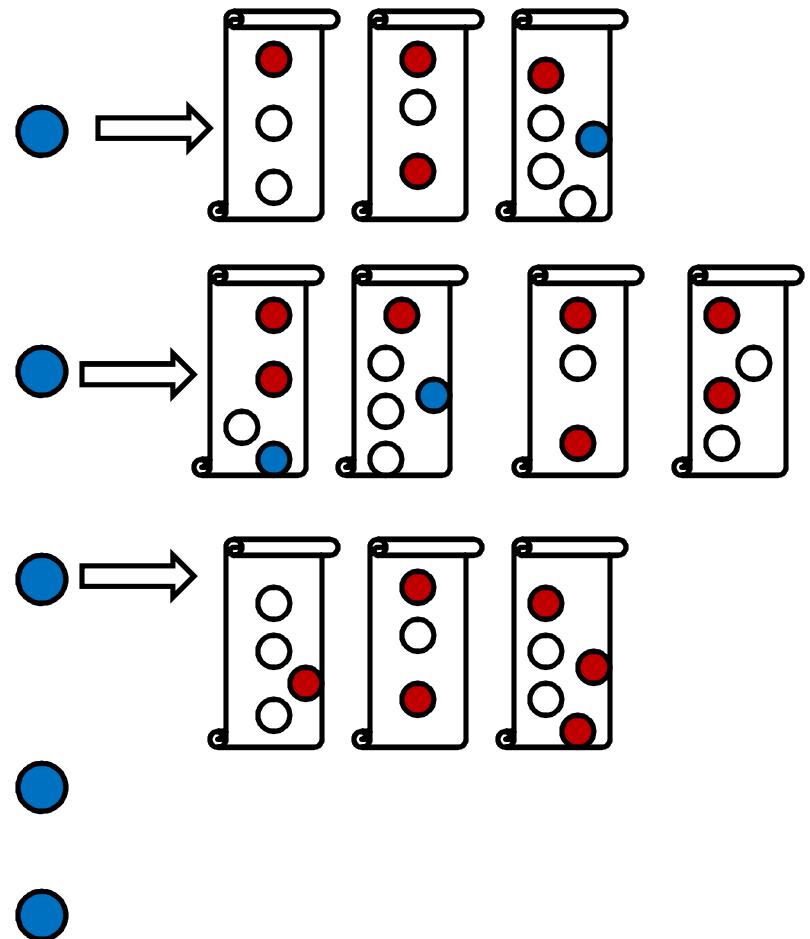
Lance Armstrong

Discovering experts using hubs



Random sample or
exhaustive crawl

- Hubs
- Experts
- Random user



Evaluation results

- ❑ Result said to be relevant if
 - ❑ Case 1: judged relevant in at least one evaluation
 - ❑ Case 2: judged relevant in the majority of evaluations
- ❑ All results evaluated at least twice judged relevant
 - ❑ Case 1: for 40 out of 42 queries
 - ❑ Case 2: for 33 out of 42 queries
- ❑ Mean Average Precision at 10
 - ❑ Case 1: 0.905, Case 2: 0.939

Details in paper

Existing approaches to find topic experts

The image shows a screenshot of Ashton Kutcher's Twitter profile. The profile header includes a yellow profile picture with a black arrow pointing up, a plus sign, and a black arrow pointing left. The name "ashton kutcher" is circled in red, along with the handle "@aplusk" and the bio "I make stuff, actually I make up stuff, stories mostly, collaborations of thoughts, dreams, and actions. Thats me." The bio is also circled in red. The location "Los Angeles, California" and a Facebook link are visible. The "Follow" button is present. Statistics show 7,561 tweets, 700 following, and 11,722,424 followers, with the follower count circled in red. Below the profile is a "Follow ashton kutcher" form with fields for full name, email, and password, and a "Sign up" button. To the right, a list of tweets is shown. The first tweet, posted 19 hours ago, is circled in red: "Everything you ever wanted to know about sushi bit.ly/Nmt5MJ". Other tweets include one about Olympic updates, a congratulatory message to @pranklab and @machinima_prime, and a tweet about "The Facts" from thrashlab.com.

Identify and rank topic experts utilizing

- User profile
- Contents of tweets
- Connectivity in social network

Cognos: <http://twitter-app.mpi-sws.org/whom-to-follow/>