Inferring Topical Attributes of Users in the Twitter Social Network

Saptarshi Ghosh

Post-doctoral researcher

Max Planck Institute for Software Systems, Germany

socialnetworks.mpi-sws.org

Dynamics On and Of Complex Networks 2014

Joint work with …

MPI-SWS, Germany

- Md. Bilal Zafar

- Juhi Kulshrestha

- Mainack Mondal

- Krishna Gummadi

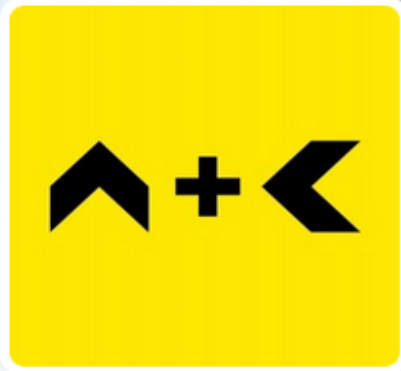## IIT Kharagpur, India

- Parantapa Bhattacharya

- Naveen Sharma
(currently at Univ. Washington)

- Niloy Ganguly

Profile of a Twitter user

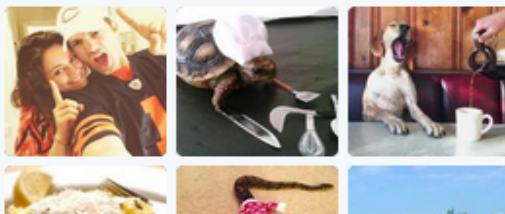| TWEETS	PHOTOS/VIDEOS	FOLLOWING	FOLLOWERS	FAVORITES	
8,483	401	721	16.4M	111	More ∨

ashton kutcher ✓
@aplusk

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

📍 Los Angeles, California
🔗 facebook.com/Ashton
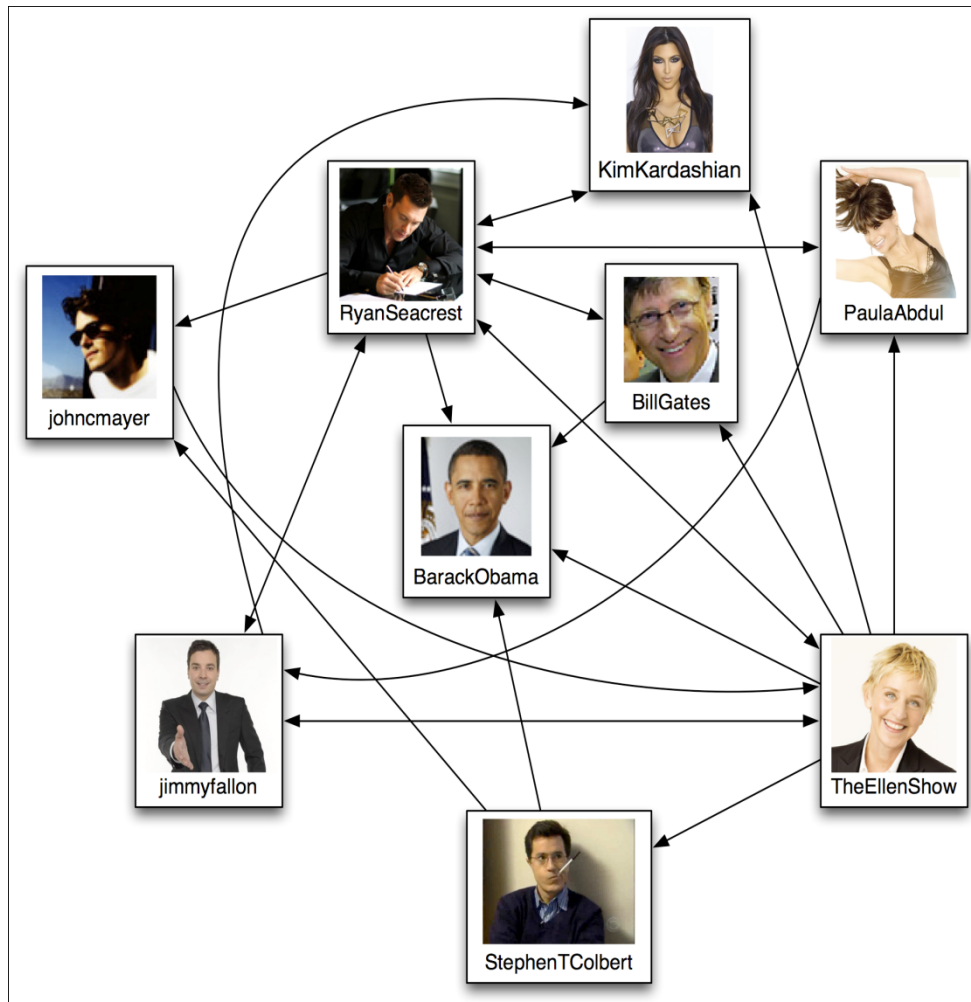🕐 Joined January 2009

📷 401 Photos and videos

Tweets   Tweets & replies

ashton kutcher @aplusk · 17h
Technologists' Duty Can Go Beyond For-Profit Industries techcrunch.com /2014/09/13/tec…

↩ ⟲ 76 ★ 171 ⋯   View summary

ashton kutcher @aplusk · Sep 11
Week2 fantasy start. Eric Decker or Golden Tate?

↩ ⟲ 69 ★ 230 ⋯

ashton kutcher @aplusk · Sep 10
There is more clapping and cheering a the apple keynote than the DNC

↩ ⟲ 120 ★ 415 ⋯

Twitter social network

- u → v: user u follows user v

- u subscribes to the tweets posted by v

- u is a follower of v

- v is a following of u

Twitter – an information network

- "We're not a social network, we're an information network" – Michael Abbott, VP Engineering, Twitter

- Twitter slogan
    - What are you doing? → changed to →
    - Discover what's happening right now, anywhere in the world

- Millions of users rely on Twitter to discover real-time content on various topics

Discovering information in Twitter

- Plethora of information
  - Over 600 million users
  - Over 400 million tweets posted daily

- Infeasible for an individual user to discover interesting information on her own

Our research

- Objective: enable Twitter users find information relevant to their interests

- How to locate interesting information on a topic?
  - Identify topical experts

- Who should be provided with information on a topic?
  - Identify users who are interested in the topic

Identifying topical experts

Inferring topical interests of users

Application: Interaction among experts and interested users

Features used by prior approaches

Our proposal …

- Utilize social annotations to infer topics of expertise of popular Twitter users
  - How does the Twitter crowd describe a user?

- Social annotations obtained through Twitter Lists
  - A feature by which one can create a named group containing some group related users

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 ▾   Followers   **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
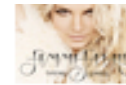*Written by Tanya Steel and the Epicurious editorial staff*

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 mons†er

Using Lists to infer topics for users

- If user *u* is an expert / authority in a certain topic
    - *u* is likely to be included in several Lists
    - List names / descriptions provide valuable semantic cues to the topics of expertise of *u*
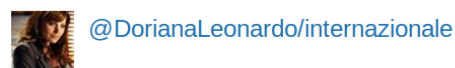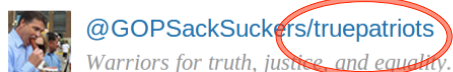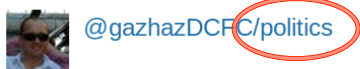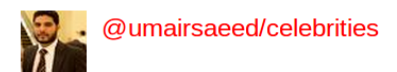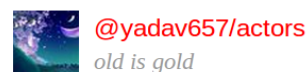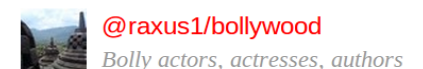
Topics inferred from Lists

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
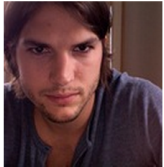
ashton kutcher ✔
@aplusk Los Angeles, California
I make stuff, actually I make up stuff, stories mostly, collaborations of thoughts, dreams, and actions. Thats me.
http://www.facebook.com/Ashton

celebs, actors, famous, movies, comedy, funny, music, hollywood, pop culture

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

Who-is-who service for Twitter

http://twitter-app.mpi-sws.org/who-is-who/

Inferring Who-is-who in the Twitter Social Network

ACM Workshop on Online Social Networks 2012,

ACM Computer Communication Review 2012

Identify topical experts: methodology

- Given a query (topic)

- Identify experts on the topic using Lists

- Rank identified experts w.r.t. given topic
  - Relevance of expert to topic
  - Popularity of expert

Cognos

- Search system for topical experts in Twitter

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

Cognos: Crowdsourcing Search for Topic Experts in Microblogs, **ACM SIGIR Conference 2012**

Cognos results for "politics"

Cognos results for "stem cell"

Evaluation of Cognos

- Cognos compared with Twitter Who To Follow

- Evaluator shown top 10 results (anonymized) by both systems
    - Evaluator judges which is better / both good / both bad
    - Judgment by majority voting

- Cognos results judged equally good or better for 60% topics

Identifying topical experts

Inferring topical interests of users

Application: Interaction among experts and interested users

Inferring Topics of Interest of a User

- Prior attempts
  - Use NLP on tweets posted by a user or received by a user
  - Limitation: tweets often contain day-to-day conversation

- Our approach
  - If a user is subscribing to several experts on a topic, she is very likely to be interested in that topic
  - Identify which topical experts a user follows, and infer the topics of expertise of those experts

Who likes what

- System to infer topical interests of users in Twitter

http://twitter-app.mpi-sws.org/who-likes-what/

Inferring User Interests in the Twitter Social Network,
ACM Conference on Recommender Systems (RecSys) 2014
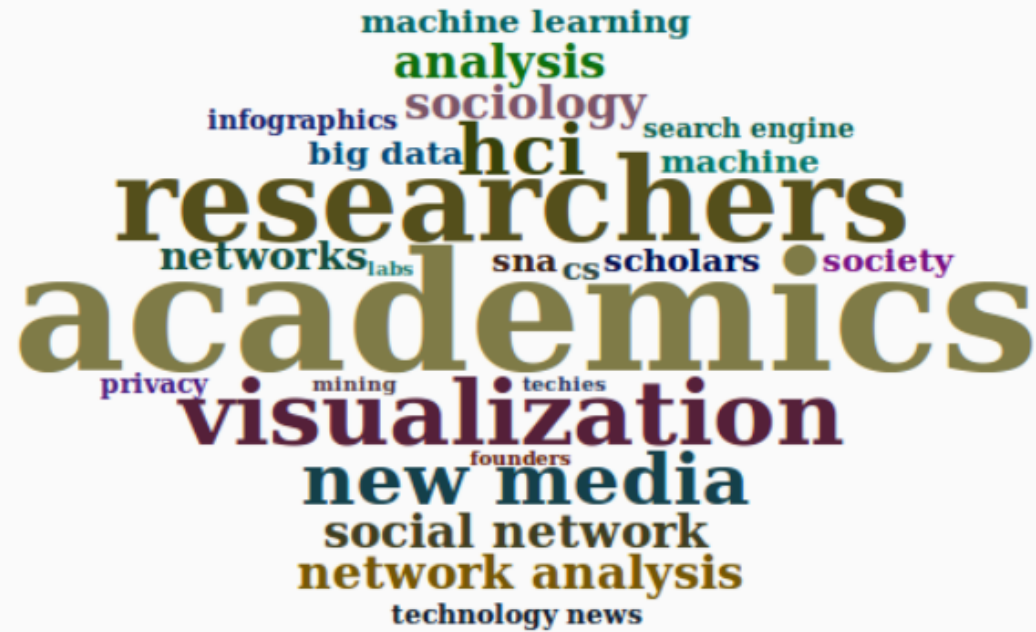
Topical interests of Lada Adamic

Topical interests of Duncan Watts

Evaluation of the approach

- Accuracy
  - Compared with topic modeling (Labeled LDA) on tweets posted / received by a user
  - Topics inferred by proposed approach judged significantly better by human volunteers

- Scalability
  - Enables inference of interests of more than 77% of all users in Twitter

Interests not same as expertise
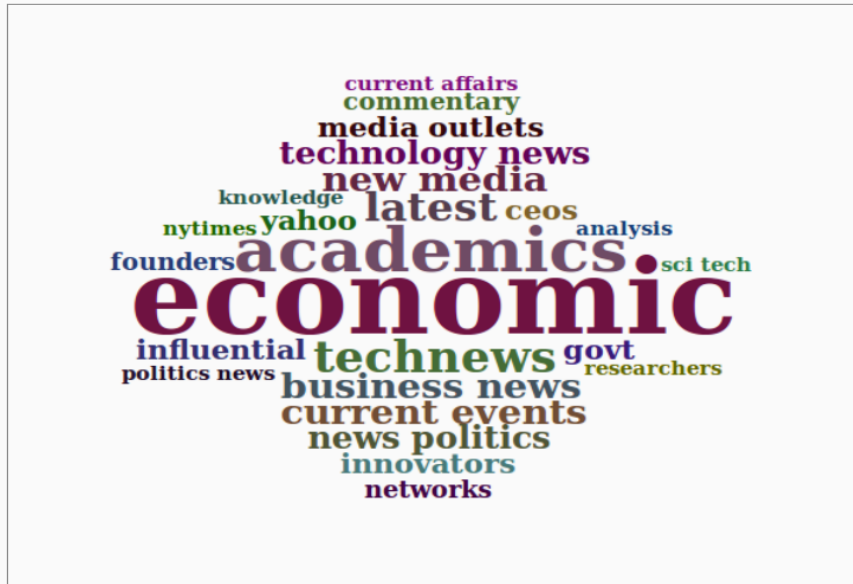
Topical interests of Duncan Watts

Topical expertise of Duncan Watts

Identifying topical experts

Inferring topical interests of users

Application: Interaction among experts and interested users

Topical groups in Twitter

Topical Groups = Experts + Seekers

Experts: Users who have expertise on the topic

Seekers: Users who are interested in the topic

@BarackObama
Expert on Politics

@BarackObama
Seeker on Basketball

Identifying topical groups at scale

- Crawled data for first 38 million users in Twitter

- 88 million Lists, 1.5 billion social links

- Identified 36 thousand topical groups

Deep Twitter Diving: Exploring Topical Groups in Microblogs at Scale, ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW) 2014

Diversity: Topics and Group Size

No. of seekers	Number of experts					
	< 100	100 − 500	500 − 1K	1K − 5K	5K − 10K	> 10K
< 1K	(5416) *geology, karate, malaria, neurology, tsunami,* psychiatry, radiology, pediatrics, dermatology, dentistry	(132) volleyball, philosophers, tarot, perfume, florists, copywriters, taxi, esperanto				
1K − 5K	(915) *biology, chemistry, swimmers,* astrophysics, multimedia, semiconductor, renewable-energy, breast-cancer, judaism	(428) *painters, astrology, sociology, geography, forensics,* anthropology, genealogy, archaeology, gluten, diabetes, neuroscience	(17) architects, insurance, second-life, police, progressives, creativity			
5K − 10K	(166) *malware,* gnu, robot, chicago-sports, gospel-music, space-exploration, wall-street	(202) horror, agriculture, atheism, attorneys, furniture, art-galleries, ubuntu	(34) *psychology,* poetry, catholic, hospitals, autism, jazz	(2) coffee, dealers		
10K − 50K	(174) ipod, ipad, virus, Liverpool-FC, choreographers, heavy-metal, backstreet-boys, world-cup,	(312) *olympics, physics, theology, earthquake,* opera, makeup, Adobe, wrestlers, typography, american-idol	(146) *tennis, linux, astronomy,* yoga, animation, manga, doctors, realtors, wildlife, rugby, forex, php, java,	(67) *law, history, beer, golf,* librarians, theatre, military, poker, conservatives, vegan		
50K– 100K	(7) bbc-radio, UK-celebs, christian-leaders, superstars	(61) *hackers, programmers,* bicycle, GOP, fantasy-football, NCAA, wwe, sci-fi	(35) *medicine, cyclists,* investors, recipes, NHL, xbox, triathlon, Google	(37) hotels, museums, hockey, architecture, charities, weather, space		
> 100K	(3) headlines, brits	(49) pop-culture, gospel, BBC, reality-tv, bollywood	(58) *religion,* actresses, gadgets, graphic-design, directors, lifestyle, gossip, commentators, youtube	(140) *books, government, comedy, environment, baseball, soccer, hollywood, iphone, economics,* money	(25) *fashion, education, wine,* photography, radio, restaurants, science, SEO	(17) *music, tech, business, politics, food, sports, celebs, health,* media, bloggers, travel, writers

A Small Number of Very Popular Groups

No. of seekers	Number of experts					
	< 100	100 – 500	500 – 1K	1K – 5K	5K – 10K	> 10K
< 1K	(5416) *geology*, *karate*, *malaria*, *neurology*, *tsunami*, psychiatry, radiology, pediatrics, dermatology, dentistry	(132) volleyball, philosophers, tarot, perfume, florists, copy-writers, taxi, esperanto				
1K – 5K	(915) *biology*, *chemistry*, *swimmers*, astrophysics, multi-media, semiconductor, renewable-energy, breast-cancer, judaism	(428) *painters*, *astrology*, *sociology*, *geography*, *forensics*, anthropology, genealogy, archaeology, gluten, diabetes, neuroscience	(17) architects, insurance, second-life, police, progressives, creativity			
5K – 10K	(166) *malware*, gnu, robot, chicago-sports, gospel-music, space-exploration, wall-street	(202) horror, agriculture, atheism, attorneys, furniture, art-galleries, ubuntu	(34) *psychology*, poetry, catholic, hospitals, autism, jazz	(2) coffee, dealers		
10K – 50K	(174) ipod, virus, Liver, choreographers, metal, backstr, world-cup,	books, government, comedy, environment, baseball, soccer, hollywood, iphone, economics, money	fashion, education, wine, photography, radio, restaurants, science, SEO	music, tech, business, politics, food, sports, celebs, health, media, bloggers, travel, writers		
50K–100K	(7) bbc-radio, celebs, leaders, supers		mentators, youtube	*iphone*, *economics*, money	restaurants, science, SEO	media, bloggers, travel, writers
> 100K	(3) headlines,					

Thousands of Specialized Niche Groups

No. of s...				Number of experts			5K – 10K	> 10K
<	geology, karate, malaria, neurology, tsunami, psychiatry, radiology, pediatrics, dermatology, dentistry			volleyball, philosophers, tarot, perfume, florists, copywriters, taxi, esperanto				
5K – 1...	biology, chemistry, swimmers, multimedia, astrophysics, semi-conductor, renewable-energy, breast-cancer, judaism			painters, astrology, sociology, geography, forensics, anthropology, genealogy, archaeology, gluten, diabetes, neuroscience				
> 100K	(3) headlines, brits	(49) pop-culture, gospel, BBC, reality-tv, bollywood	(58) *religion*, actresses, gadgets, graphic-design, directors, lifestyle, gossip, commentators, youtube	(140) *books, government, comedy, environment, baseball, soccer, hollywood, iphone, economics, money*		(25) *fashion, education, wine,* photography, radio, restaurants, science, SEO	(17) *music, tech, business, politics, food, sports, celebs, health,* media, bloggers, travel, writers

Detecting topical groups

- We followed a semantic approach to identify topical groups

- Could community detection algorithms be used on the social network to detect them?

- Applied BGLL / Louvain algorithm on the Twitter social network to identify communities

Detecting topical groups

- BGLL largely unable to detect topical groups, especially the smaller ones (on niche topics)

- Topical groups do not have good structural quality on which most community detection algorithms rely

- Difference explained by a sociological theory on how groups form in a social network

Why do groups form?

Common Identity and Bond Theory

- Prentice, Personality and Social Psychology Bulletin, 1994

Identity Based Groups

Low Reciprocity

Low Personal Interactions

High Topicality of discussions

Examples: audience at a conference, topical groups

Bond Based Groups

High Reciprocity

High Personal Interactions

Low Topicality of discussions

Examples: family, personal friends

Summary

- How to identify topical expertise / interests of users?
    - Proposed crowdsourcing based methods

- Identified topical groups in Twitter at scale

- Ongoing work
    - Developing search / recommender systems to help users find important content relevant to their interests

Thank You !