



max planck institut
informatik

What Computers Should Know, Shouldn't Know and Shouldn't Believe

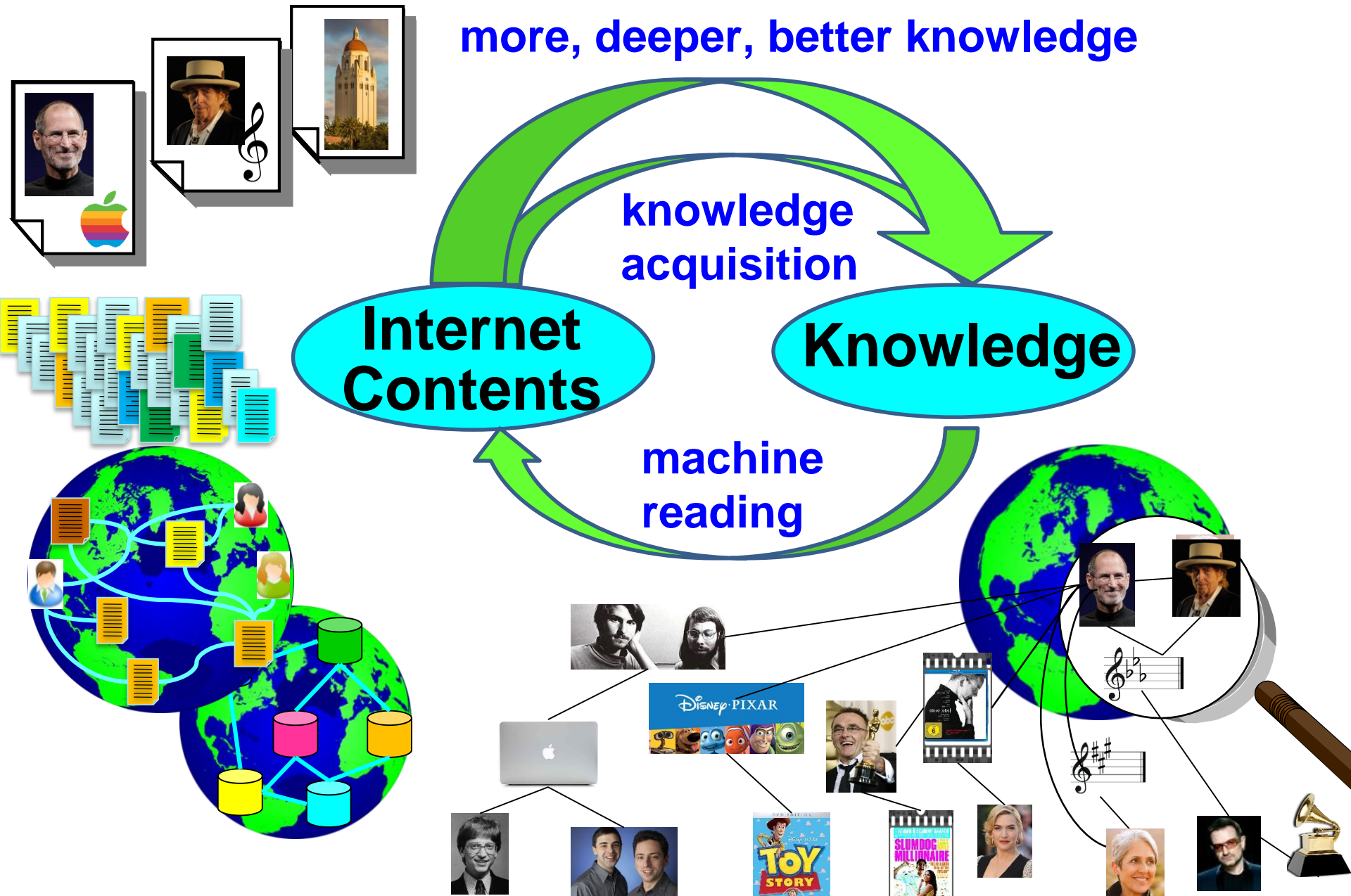
Gerhard Weikum

Max Planck Institute for Informatics and
Saarland University

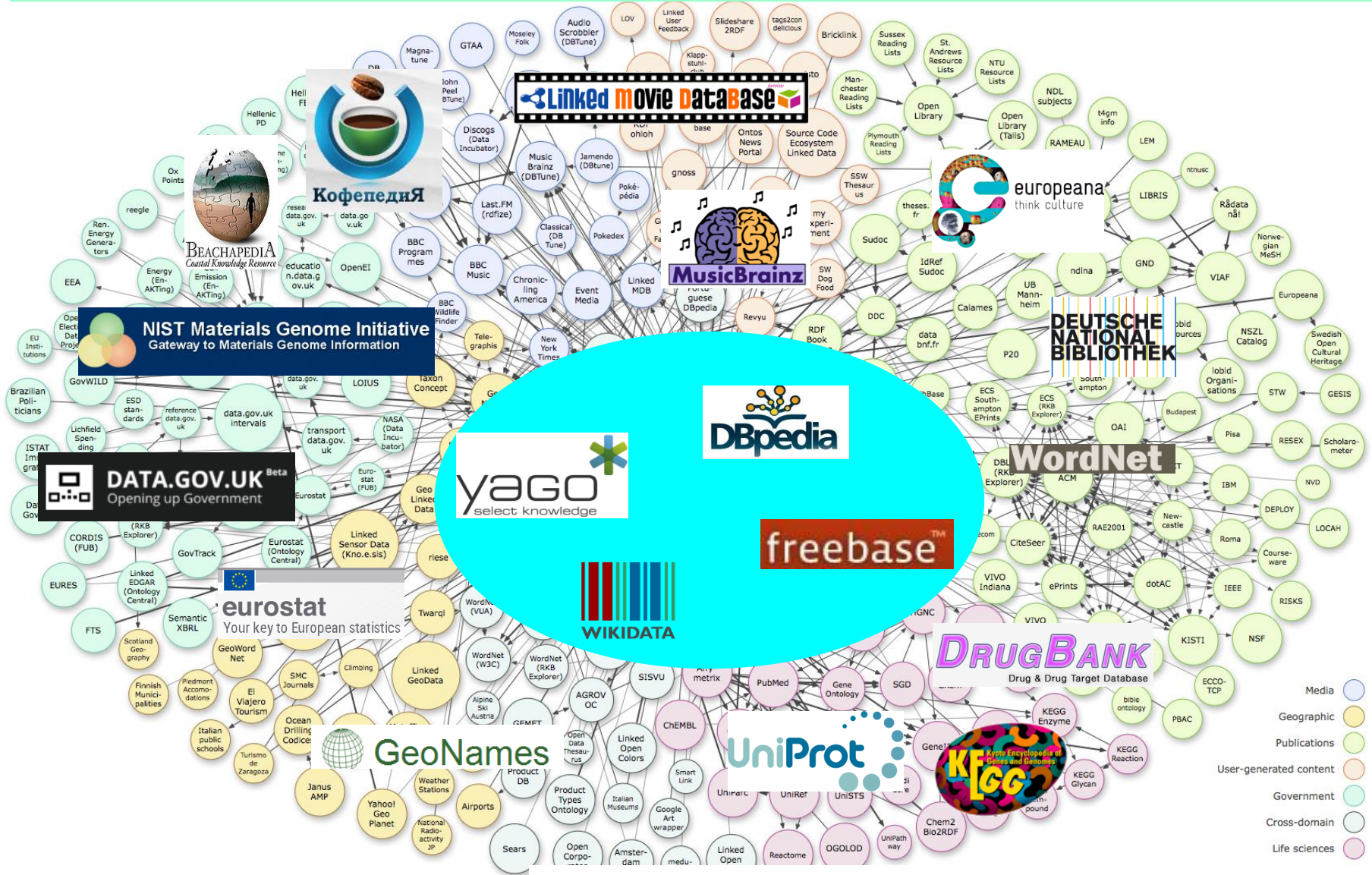
<http://mpi-inf.mpg.de/~weikum>

Turn Text & Data into Knowledge

more, deeper, better knowledge



> 100 Billion subject-predicate-object facts from > 1000 sources



Web of Knowledge

- 10M entities in 350K types
- 200M facts
- 100 languages
- >95% accuracy

yAGO
select knowledge

- 5M entities in 250 types
- 500M facts for 6000 relations

DBpedia

- 600M entities
- 20B facts

- 15 M entities
- 150 M facts

WIKIDATA

freebase

- 40M entities
- 1B facts for 4000 relations

Google
Knowledge
Graph

Applications:

- semantic search & question answering
- natural language understanding
- recommender systems
- text analytics, data cleaning,

Machine Knowledge for Answer Engines

**Precise and concise answers
for advanced information needs:**



properties of entity

★ **Nobel laureate who outlived two world wars and all his children?**

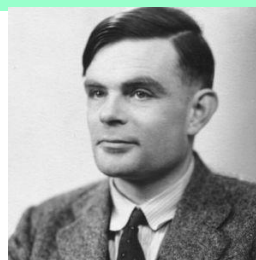
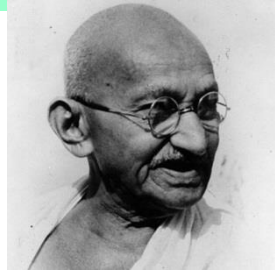
★ **Politicians who are also scientists?**



sets of entities

relationships between entities

★ **Commonalities & relationships among:
T.E. Lawrence, Gandhi, Alan Turing, June Carter?**



Machine Knowledge for Answer Engines

**Precise and concise answers
for advanced information needs:**

real applications

- ★ **Proteins that bind to the Zika virus?**
- Antidepressants that interfere with thyroid drugs?**
- Polymer materials for super-capacitators?**
- German philosophers influenced by William of Ockham?**
- Green politicians mentioned in Panama Papers?**

Outline

- ✓ **Introduction**
- ★ **What Computers Should Know**
- ★ **What Computers Shouldn't Know**
- ★ **What Computers Shouldn't Believe**
- ★ **Conclusion**

Missing Predicates & Salient Facts

Which **interesting facts** about an entity are in infoboxes?



Johnny Cash



Born J. R. Cash
February 26, 1932
[Kingsland, Arkansas, U.S.](#)

Died September 12, 2003
(aged 71)
[Nashville, Tennessee, U.S.](#)

Cause of death [Diabetes mellitus](#)

Resting place [Hendersonville Memory Gardens](#)

Occupation Singer-songwriter · guitarist · actor · author

Years active 1954–2003

Spouse(s) Vivian Liberto (m. 1954; div. 1966)
[June Carter](#) (m. 1968; her death 2003)

Military career	
Allegiance	 United States of America
Service/branch	 United States Air Force
Years of service	1950–1954
Rank	 Staff sergeant
Musical career	
Genres	Country, rockabilly, ^[1] rock and roll, gospel
Instruments	Vocals, guitar
Labels	Sun, Columbia, Mercury, American, House of Cash, Legacy Recordings
Associated acts	The Tennessee Three, The Highwaymen, June Carter Cash, The Statler Brothers, The Carter Family , Waylon Jennings
Website	johnnycash.com
Notable instruments	
Martin Acoustic Guitars ^[2]	

romanticRelationship
(Johnny Cash,
JuneCarter, 1961-2003)

featuredInMovie
(romanticRelationship(...),
WalkTheLine))

playedConcertAtLocation
(JohnnyCash,
SanQuentinStatePrison)

covered (JohnnyCash,
One by U2)

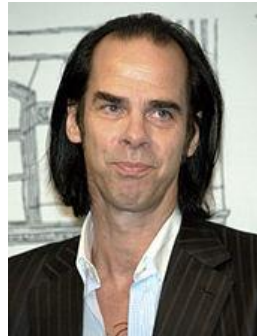
covered (JohnnyCash,
MercySeat by NickCave)

.....

not in any KB !

Missing Predicates & Salient Facts

Which **interesting facts** about an entity are in infoboxes?



Abattoir Blues / The Lyre of Orpheus

Nick Cave
& The Bad Seeds
Abattoir Blues / The Lyre of Orpheus



Studio album by Nick Cave and the Bad Seeds

Released 20 September 2004

Recorded March–April 2004 at Studio Ferber
in [Paris](#), France

Genre [Alternative rock](#)

Length 82:30

Label [Mute](#)

Producer [Nick Launay](#)

Nick Cave and the Bad Seeds chronology

Nocturama
(2003)

*Abattoir
Blues / The
Lyre of
Orpheus*
(2004)

*B-Sides &
Rarities*
(2005)

Singles from *Abattoir Blues / The Lyre of Orpheus*

- "Nature Boy"
Released: 6 September 2004
- "Breathless / There She Goes, My Beautiful World"
Released: 15 November 2004
- "Get Ready for Love"
Released: 14 March 2005

isOnAlbum

(Let the Bells Ring,
AbbattoirBlues)

lyricsAbout

(Let the Bells Ring,
JohnnyCash)

isOnAlbum

(O'Children,
AbbattoirBlues

playedInMovie

(O'Children,
HarryPotter 6)

not in any KB !

Spectrum of Digital Knowledge (1): School Education for Computers

taxonomic knowledge:

type (SteveJobs, entrepreneur), **subtypeOf** (entrepreneur, businessperson)

type (SteveJobs, inventor), **subtypeOf** (inventor, human)

type (SteveJobs, YogaPractitioner), **type** (SteveJobs, GratefulDeadFan)

long-tail
classes

factual knowledge:

hasFounded (SteveJobs, Apple), **CEO** (SteveJobs, Apple)

long-tail entities

hasDaughter (SteveJobs, LisaBrennan), **namedAfter** (AppleLisa, LisaBrennan)

hasFavoriteSong (SteveJobs, Imagine), **hasFavoriteSong** (SteveJobs, Truckin')

dated (SteveJobs, JoanBaez), **admired** (SteveJobs, BobDylan)

composed (JoanBaez, Diamonds&Rust), **lyricsAbout** (Diamonds&Rust, BobDylan)

sangAt (JoanBaez, memorialForSteveJobs)

long-tail relations

spatial & temporal knowledge:

diedOn (SteveJobs, 5-Oct-2011), **diedIn** (SteveJobs, Palo Alto)

happened (**hasFounded** (SteveJobs, Apple), Cupertino, 1976)

validDuring (**CEO** (SteveJobs, Apple), 1997-2011)

Spectrum of Digital Knowledge (2): Kindergarten and University

commonsense properties:

property (lemon, yellow), **property** (lemon, juicy), **property** (lemon, sour),
ability (fish, swim), **ability** (human, speak), **usedFor** (classroom, teaching),
maxHeight (human, 2.5 m), **maxLength** (snake, 10 m)

commonsense rules:

$\forall x: \text{human}(x) \Rightarrow (\exists y \text{ mother}(x,y)) \wedge (\exists z \text{ father}(x,z))$

$\forall x, y, z \text{ mother}(x,y) \wedge \text{mother}(x,z) \Rightarrow y = z$

domain-specific expert knowledge:

type (Ubiquinone-8, coenzyme), **expresses** (COQ8, Ubiquinone-8)
causes (lack of Ubiquinone-8, mitochondrial disorder)

Spectrum of Digital Knowledge (3): Learned in Life

socio-cultural and social knowledge:

invented (Internet, Cerf and Kahn, techies),

invented (Internet, Berners-Lee, common people),

invented (Internet, Al Gore, mid-aged democrat voters)

drink (beer, Germany), **drink** (wine, California), **drink** (lassi, India)

alleviates (ice, bruises), **alleviates** (eucalyptusOil, sinusitis)

belief knowledge:

believe (Ptolemy, **center** (world,earth), **believe** (Galileo, **center** (world, sun))

believe (Chinese, **badLuckNumber** (4)), **believe** (Germans, **badLuckNumber** (13))

believe (AustralianAborigines, **taboo** (photosOfDeadPeople))

process knowledge:

type (changeTire, mechanicalTask)

subtask (changeTire, loosenBolts), **subtask** (changeTire, liftCar),

requires (loosenBolts, spiderWrench), **requires** (liftCar, jack)

precedes (loosenBolts, liftCar)

Knowledge Gaps

Temporal and Spatial Knowledge

Long-Tail Knowledge (on types and entities)

Dynamic Knowledge (events, emerging entities)

Open-Ended Knowledge (relation types)

On-the-Fly Knowledge

Visual Knowledge (on types and long-tail entities)

Cultural Knowledge

Commonsense Knowledge

Social Knowledge

Intensional Knowledge

Negative Knowledge

Commonsense Knowledge: Not So Common

Every child knows that

apples are green, red, round, juicy, ...
but not fast, funny, verbose, ...

pots and pans are in the kitchen or cupboard, on the stove, ...
but not in the bedroom, in your pocket, in the sky, ...

children usually live with their parents

But: commonsense is rarely stated explicitly

Plus: web and social media have reporting bias

color of elephants ?

pink elephant: 0.9 Mio on Google

grey elephant: 0.4 Mio on Google



rich family: 27.8 Mio on Bing

poor family: 3.5 Mio on Bing

singers: 22.8 Mio on Bing

workers: 14.5 Mio on Bing

Pattern-based Harvesting of Commonsense Properties



Gerard
de Melo



Niket
Tandon



Cuong
Chu

Approach: Start with **seed facts**

hasProperty (apple, round)

hasAbility (dog, bark)

hasLocation (plate, table)

Learn **patterns** that express these relations, such as
X is very Y, X can Y, X put in/on Y, ...

Apply patterns to Web, books, N-grams corpora, image tags, etc.

→ **statistics**, semisupervised **learning**, constraint **reasoning**

hasColor (elephant, grey), **hasShape** (circle, round) ...

hasAbility (fish, swim), **hasAbility** (human, talk) ...

usedFor (book, learn), **usedFor** (computer, learn)

partOf (wheel, bike), **partOf** (wheel, car) ...

hasTemperature (oven, hot), **hasTaste** (chili, hot)

WebChild KB:

5 Mio. assertions

semantically typed

sense-disambiguated

[WSDM 2014]

Commonsense & Visual Contents

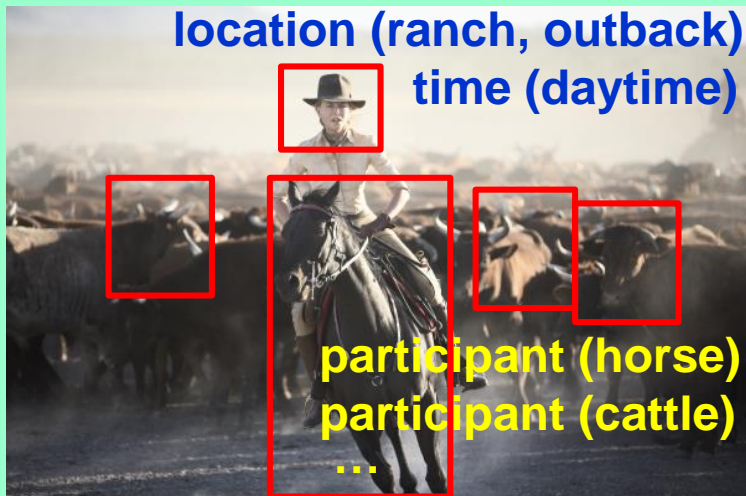
[N. Tandon et al.: WWW'15, CIKM'15, AAI'16, WWW'17]



Refined part-whole relations from web&books text and image tags

→ 6.7 Mio sense-disambiguated triples for physicalPartOf, visiblePartOf, hasCardinality, memberOf, substanceOf

driveHerd:



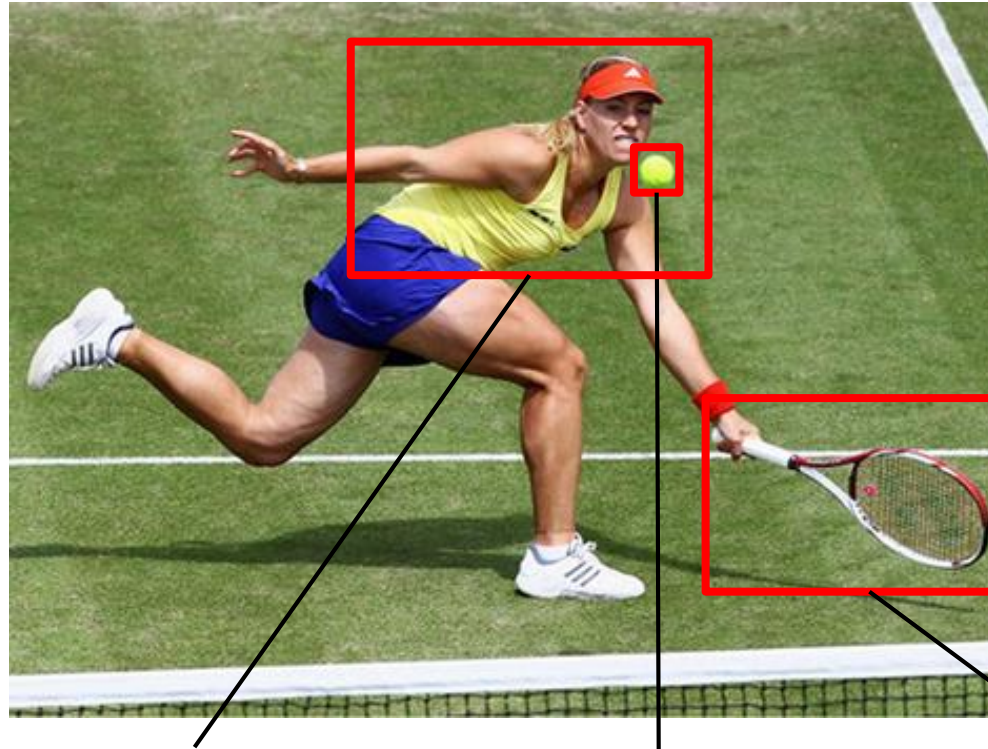
Activity knowledge from movie&TV scripts, aligned with visual scenes

→ 0.5 Mio activity types with attributes: location, time, participants, prev/next

Commonsense for Language Understanding

- **How-to queries:**
repair a bike tire, pitch a tent, cross a river, ...
- **Scene search** (over videos, books,):
romantic dinner, dramatic climb, ...
- **Question disambiguation:**
jordan weather forecast
Germany's keeper at the Euro ?
- **Sentiment analysis:**
the hottest place on earth
the hottest bar in town

Commonsense for Computer Vision



woman (o1) ? lemon (o2) racket (o3)

+ commonsense
knowledge:

hasColor (tennisball, yellow)
hasShape (tennisball, round)
occursAt (tennisball, tennis court)
occursAt (racket, tennis court)

⇒ tennisball(o2)

Commonsense for Image Description



He's been everywhere mate.

dog, man, woman, motorbike

→ *red dog in front of man and woman on a motorbike*

+ commonsense
knowledge:

→ *romantic Australia*

→ *the wanderer of
the Pilbara*



Outline

- ✓ **Introduction**
- ✓ **What Computers Should Know**
- ★ **What Computers Shouldn't Know**
- ★ **What Computers Shouldn't Believe**
- ★ **Conclusion**

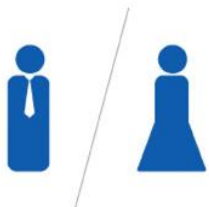
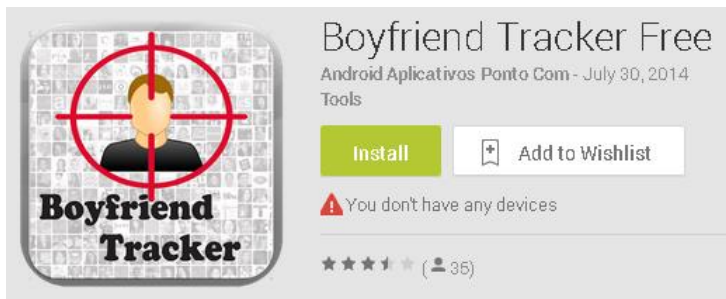
What Computers Know About **You** ... (But Shouldn't ?)



**Nobody interested
in your research?
We read your papers!**



The Dark Side of Knowledge



Applicant Data



Social Web Analysis



Social Intelligence Risk Score

Goal: Privacy Risk Advisor

Users are tracked, linked, targeted, rated ...
Algorithmic decisions made with this data

Searching 61+ Social Media Platforms

Gathering results for *marc spaniol*

91%

Your search history is 100% private. marc spaniol will NOT be notified about your search.

Results May Include

- ✓ Online Alias, Phone & Address
- ✓ Profiles from 60+ Social Networks
- ✓ Top Online Dating Site Profiles
- ✓ Social Profile Pics & Blogs
- ✓ Info Search Engines can't find
- ✓ And Much More!

Background Check Report on *marc spaniol*

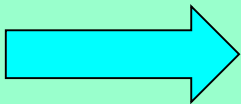
- ✓ Full Name
- ✓ Current Address
- ✓ Phone Numbers
- ✓ Prior Residences
- ✓ Relatives
- ✓ Aliases & AKAs
- ✓ Age & DOB
- ✓ Neighbors
- ✓ Property Records
- ✓ Bankruptcies
- ✓ Judgments & Liens
- ✓ Marriage & Divorce
- ✓ Birth & Death Records
- ✓ Misdemeanors
- ✓ Criminal Check
- ✓ Sex Offender Records

Limited Time Offer!
Report • Trial
\$19.95 \$9.95

continue

Established privacy models

- Data: single database
- Adversary: computationally powerful, but agnostic
- Goal: anonymity guarantee
- Measures: data coarsening, perturbation, limit queries



Today's user behavior & risks

- Data & User: multimodal, **social, agile, longitudinal**
- Adversary: **world knowledge & probabilistic inference**
- Goal: **alert & advise, bound risk**
- Measures: **guide user** on policies, selective anonymization

Ranking-based Risk Assessment



Asia
Biega

Krishna
Gummadi

Quora

Search for questions, people, and topics

FMLA Pregnancy Sex Medicine and Healthcare Jobs and Careers Career Advice

Should I tell prospective employers that I am pregnant?

I am being laid off from a large co in July for the birth of my child. I v to know since I just got a good re and telling them I will be available debating whether or not to tell the do not have to tell them, and they honest and simply making sure I

amy63628



★1

Fluoxetine symptoms??

Hi everyone, So on friday i was dignosed with clincial depression and PTSD and was put on fluoxetine starting at 2.5ml (to be increased gradually because i'm quite young) So yesterday i was really hyper but i was excited because i was going to see a concert later in the night so i'm going to put

2 replies 0 votes

Last reply 15 days ago



*social media,
online
communities,
review forums,
query logs,*

...

4 Answers

Exposure Risk [J.A. Biega et al.: SIGIR'16]:

1. Capture **sensitive topics** (depression, gambling, debts, ...) with **knowledge** on vocabulary & correlations
→ **latent topic models** (LDA, word2vec skip-gram)
2. Estimate user exposure per topic:
→ **quantitative risk measures**
3. **Rank users:** compute **top-k targets**
→ alert user, recommend selective anonymization

Outline

- ✓ **Introduction**
- ✓ **What Computers Should Know**
- ✓ **What Computers Shouldn't Know**
- ★ **What Computers Shouldn't Believe**
- ★ **Conclusion**

Social Knowledge: “Folk Wisdom”

Beliefs become like facts
in specific **socio-cultural contexts**

invented (computer, Eckert and Mauchley, USA)

invented (computer, KonradZuse, Germany)

invented (computer, AlanTuring, UK)

invented (computer, SteveJobs, young nerds)

invented (computer, DonaldTrump, twitter)

**Need to tap
riskier sources
(social media
posts etc.)**

**Knowledge distilled from subjective
behavior / experience of many people**

- common habits on food, drinks, social behavior
- experience with electric/diesel/green cars, on the road
- grassroots insight on health, symptoms, therapies, ...

Credibility and Trust: Where the Truth Lies



Subhabrata
Mukherjee



Cristian
Danescu



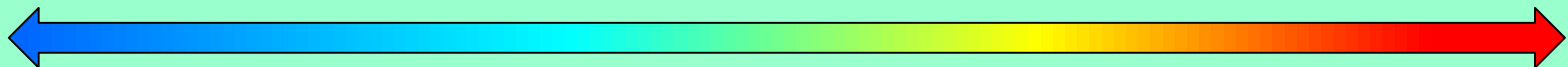
Kashyap
Popat



Jannik
Strötgen

Assess **credibility** of statements / claims on the Internet
and the **trustworthiness** of their sources

- Search results: *love affairs of Hillary Clinton ?*
- Biased news: *Merkel hates Greece*
- KB contents: *Berners-Lee & Al Gore invented the Internet*
- Social media: *Obamacare requires microchip implant*
- Health forums: *Xanax causes hallucinations*
- Press conference: *Trump proves that CNN has only fake news*



factsbias post-truth manipulation fake

→ **Polygraph for digital contents:** for IE, KB, analytics, users

Between Facts and Fake



Post-truth politics

From Wikipedia, the free encyclopedia

Post-truth politics (also called **post-factual politics**) is a [political culture](#) in which debate is framed largely by [appeals to emotion](#) disconnected from the details of [policy](#), and by the repeated assertion of [talking points](#) to which factual rebuttals are ignored. [Post-truth](#) differs from traditional contesting and falsifying of [truth](#) by rendering it of "secondary" importance. While this has been described as a contemporary problem, there is a possibility that it has long been a part of political life, but was less notable before the advent of the Internet. In the novel *Nineteen Eighty-Four*, [George Orwell](#) cast a world in which the state changes historic records daily to fit its [propaganda](#) goals of the day.

Political commentators have identified post-truth politics as ascendant in Russian, Chinese, American, Australian, British, Indian, Japanese and Turkish politics, as well as in other areas of debate, driven by a combination of the [24-hour news cycle](#), [false balance](#) in news reporting, and the increasing ubiquity of [social media](#).^{[1][2][3][4][5][6]} In 2016, "post-truth" was chosen as the [Oxford Dictionaries' Word of the Year](#),^[7] due to its prevalence in the context of that year's [Brexit referendum](#) and [U.S. presidential election](#).^{[8][9]}

This article is part of a series on
**Misinformation and
disinformation**



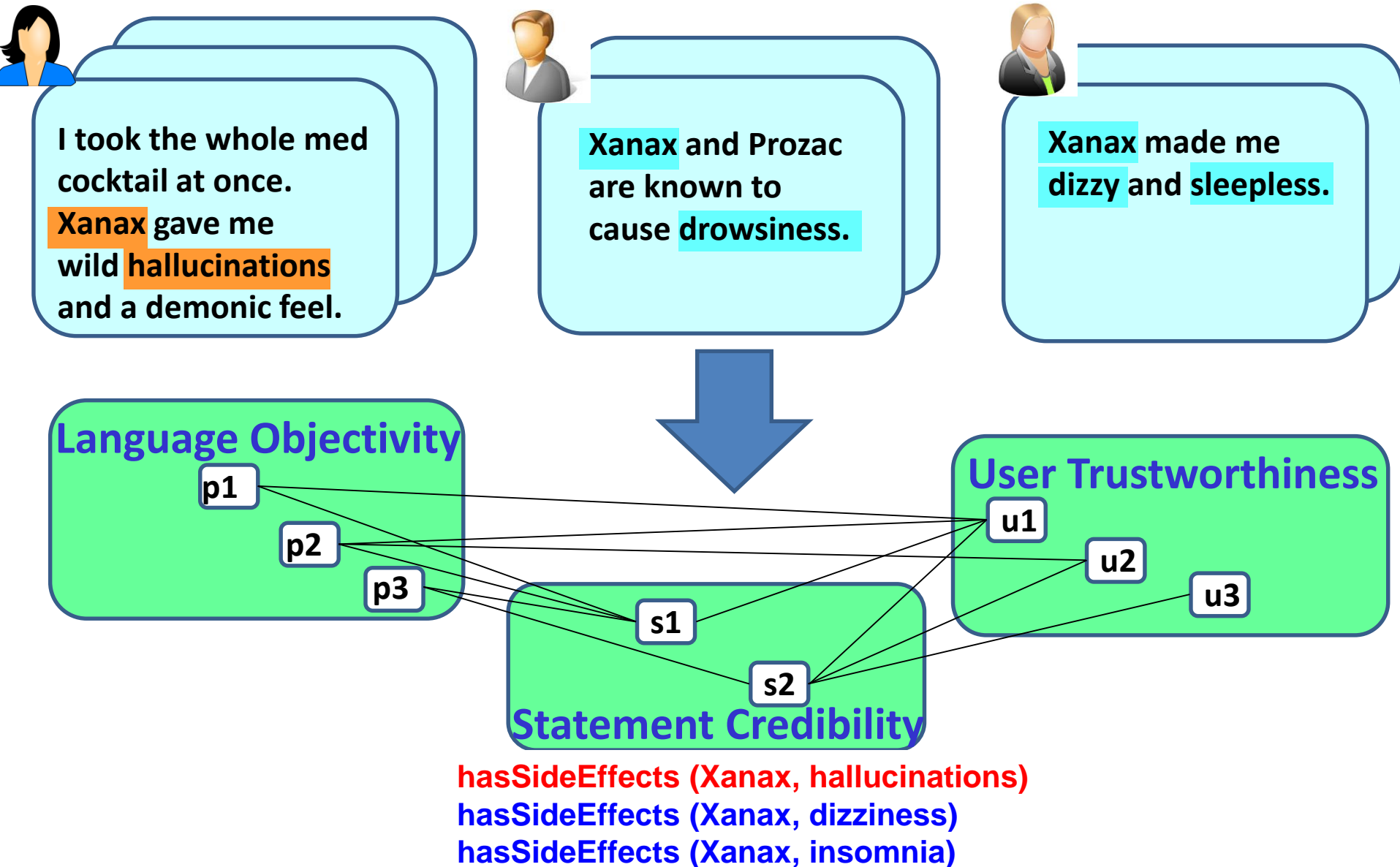
General information [hide]

[Big lie](#) • [Black propaganda](#) • [Circular sourcing](#) • [Deception](#) • [Dog-whistling](#) • [Doublespeak](#) • [Echo chamber](#) • [Euphemistic misspeaking](#) • [False flag](#) • [Filter bubble](#) • [Factoid](#) • [Fallacy](#) • [Fake news \(by country • online\)](#) • [Gaslighting](#) • [Half-truth](#) • [Hoax](#) • [Ideological framing](#) • [Internet manipulation](#) • [Media manipulation](#) • [Post-truth](#) • [Propaganda](#) • [Quote mining](#) • [Scientific fabrication](#) • [Social bot](#) • [Spin](#)

https://en.wikipedia.org/wiki/Post-truth_politics

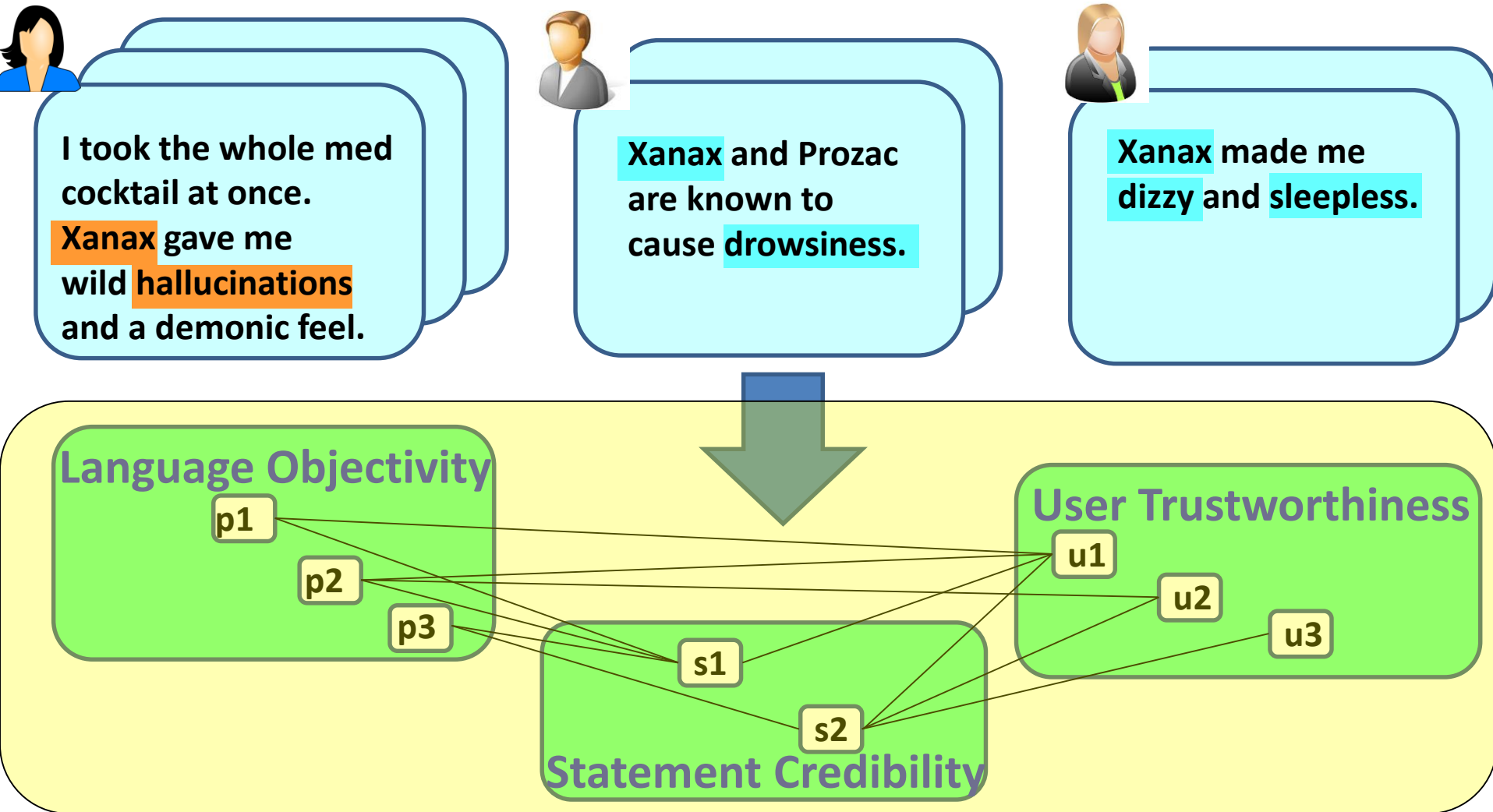
Credibility & Trust in Health Communities

[S. Mukherjee et al.: KDD'14, CIKM'15]



Credibility & Trust in Health Communities

[S. Mukherjee et al.: KDD'14, CIKM'15]



**joint reasoning with probabilistic graphical model
(semi-supervised heterogeneous CRF with EM-style learning)**

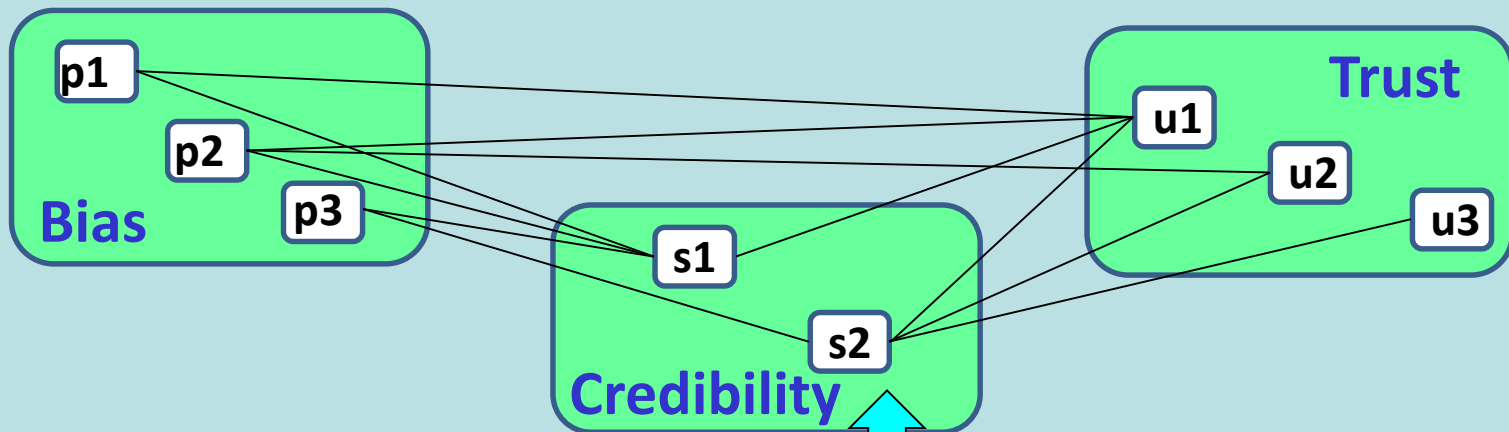
Semi-Supervised Heterogeneous MRF for Joint Inference on Posts, Users, Statements

Post language style:
modals, emotion adject.,
first person pronouns,
inferential conjunctions,
etc.

**Observable
Features**

User activities:
#posts, #replies,
#thanks received,
etc.

Joint Probabilistic Inference



Statement training labels: true,false

Experiments with Health Community Data

[S. Mukherjee et al.: KDD'14]

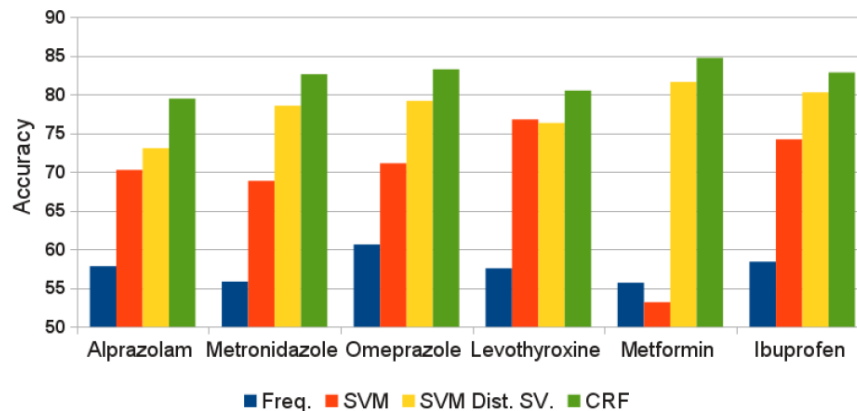
Data: healthboards.com

850,000 users, 4.5 mio. posts

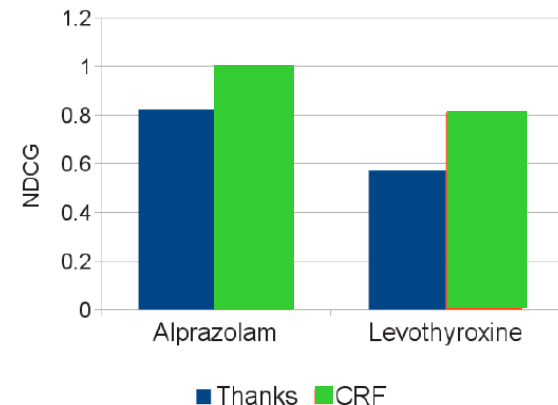
sampled 15,000 users, 2.8 mio. posts

**Ground-truth: expert data from MayoClinic portal
focused on 6 widely used drug families**

**Inferring rare side effects
of drug families**



**Inferring expert users
on drug families**



Credibility of Textual Claims

Jack Ma is the new owner of AC Milan

[K. Popat et al.: CIKM'16, WWW'17]

➡ True or False ?

search Web for evidence

GAME ON

Chinese investors want to buy the legendary AC Milan soccer club from Silvio Berlusconi

By Zheping Huang | May 10, 2016



calciomercato.com

ITALY'S FIRST FOOTBALL NEWS SITE



CM EXCLUSIVE: Alibaba and Whampoa not interested in buying AC Milan. Is Robin LI the right man?



28 April at 12:50

Throug Weibo, the Chinese Twitter, Jack Ma dismissed the rumors regarding the possibility to buy AC Milan's shares.

Last night, some sources reported of the interest of Li Ka Shing -owner of the Hutchinson Whampoa Limited- in buying the Lombard club's stake, but Li Ka -contacted by calciomercato.com- has denied such interest.

TIME

LATEST

MAGAZINE

VIDEOS



Legendary Italian Soccer Club AC Milan Is Now Chinese-Owned

SPORTS SOCCER

Legendary Italian Soccer Club AC Milan Is Now Chinese-Owned

Charlie Campbell / Beijing @charliecampbell
July 6, 2016

"Milan has n

Storied Italian
been sold to a
the club's own
Minister Silvio

CNBC

HOME U.S. NEWS MARKETS INVESTING TECH MAKE IT VIDEO

FINANCE BANKS INVESTING WALL STREET HEDGE FUNDS M&A INSURANCE

Alibaba founder Jack Ma laughs off AC Milan football club purchase rumors

CNBC.com staff | @CNBC
Thursday, 28 Apr 2016 | 9:17 PM ET

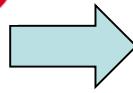
CNBC



JACK MA
SETTLES RUMORS

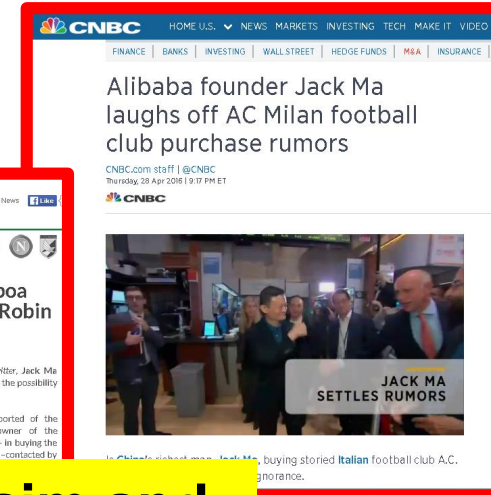
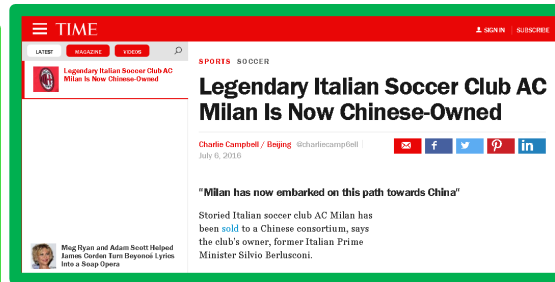
Credibility of Textual Claims

Jack Ma is the new owner of AC Milan

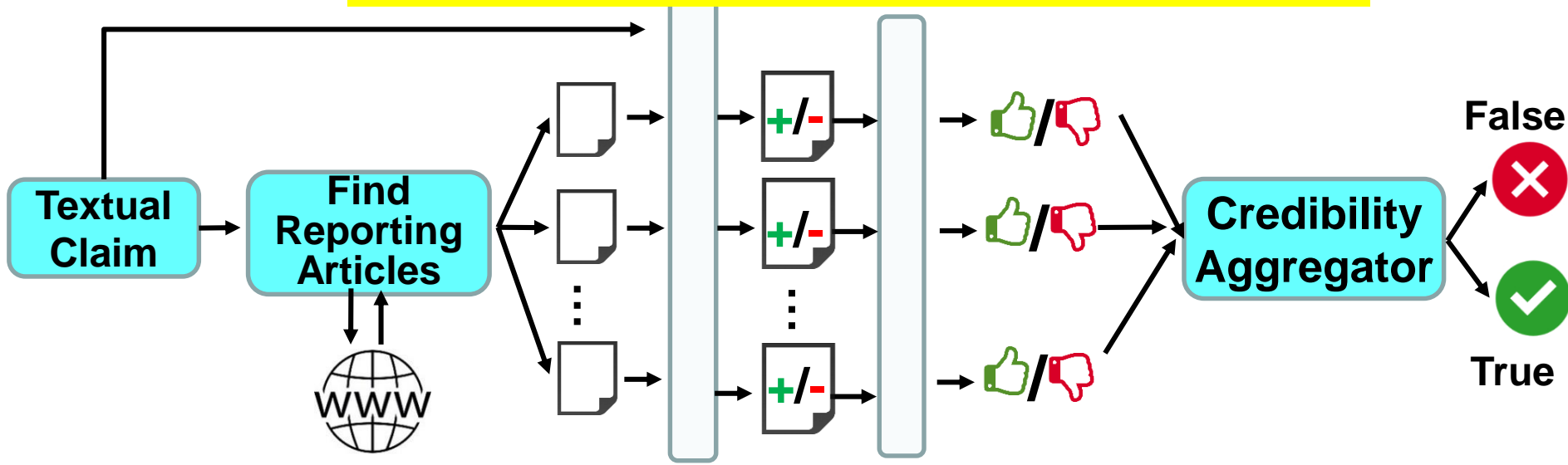


True or **False** !

[K. Popat et al.: CIKM'16, WWW'17]



joint reasoning on credibility of claim and authority of supporting/refuting sources



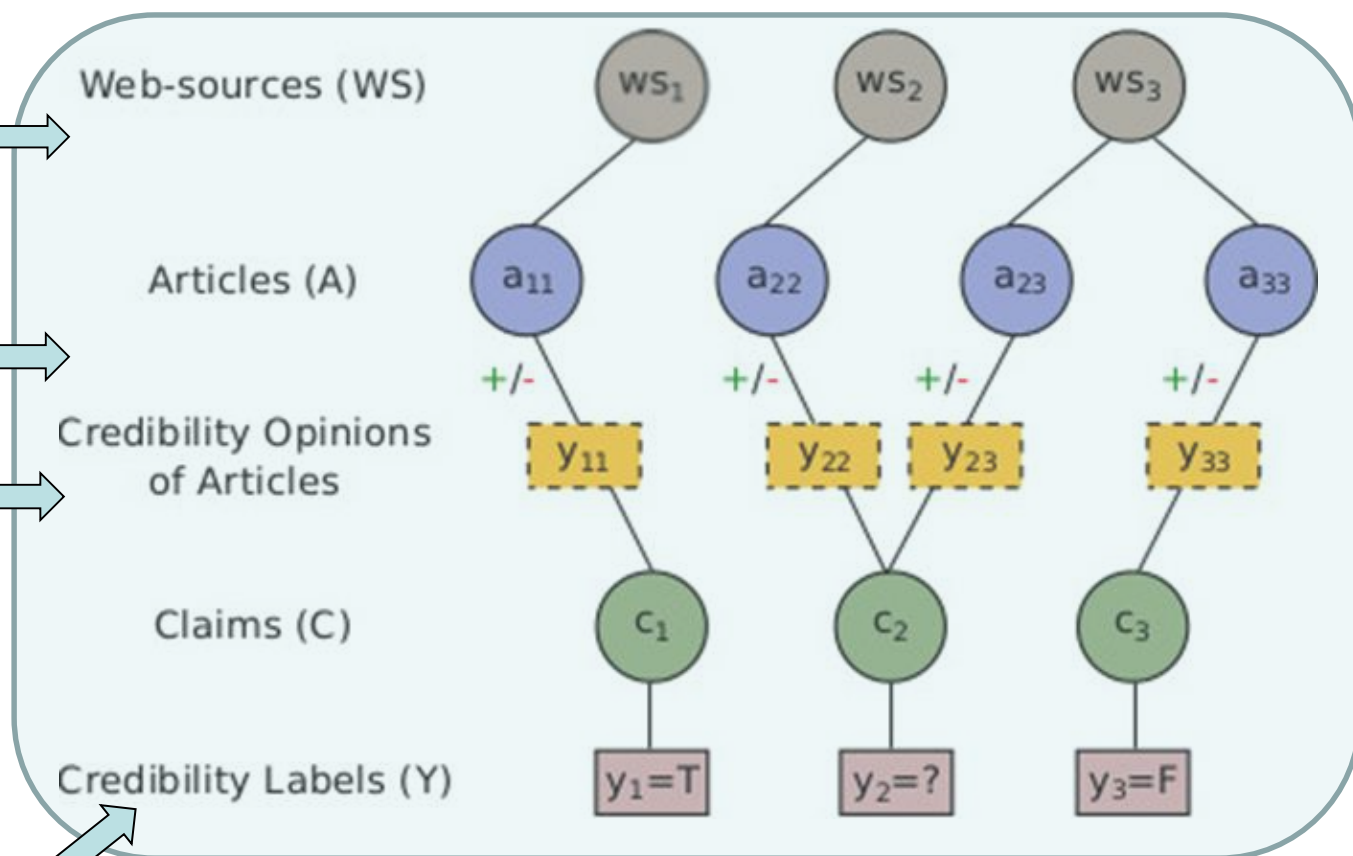
Distantly Supervised Model

[K. Popat et al.: WWW 2017]

Reliability
scoring
per source

Stance classifier
per article

Credibility
classifier
per article



Harness per-claim labels
from snopes.com
for training

CLAIM

A viral video shows a man punching a kangaroo in the face to save his dog.

RATING



TRUE

ORIGIN

In December 2016, a video went viral, purportedly showing a man punching a kangaroo in order to save a dog:

CLAIM

Australia is the first country to begin microchipping its citizens. [See Example\(s\)](#)

RATING



FALSE

ORIGIN

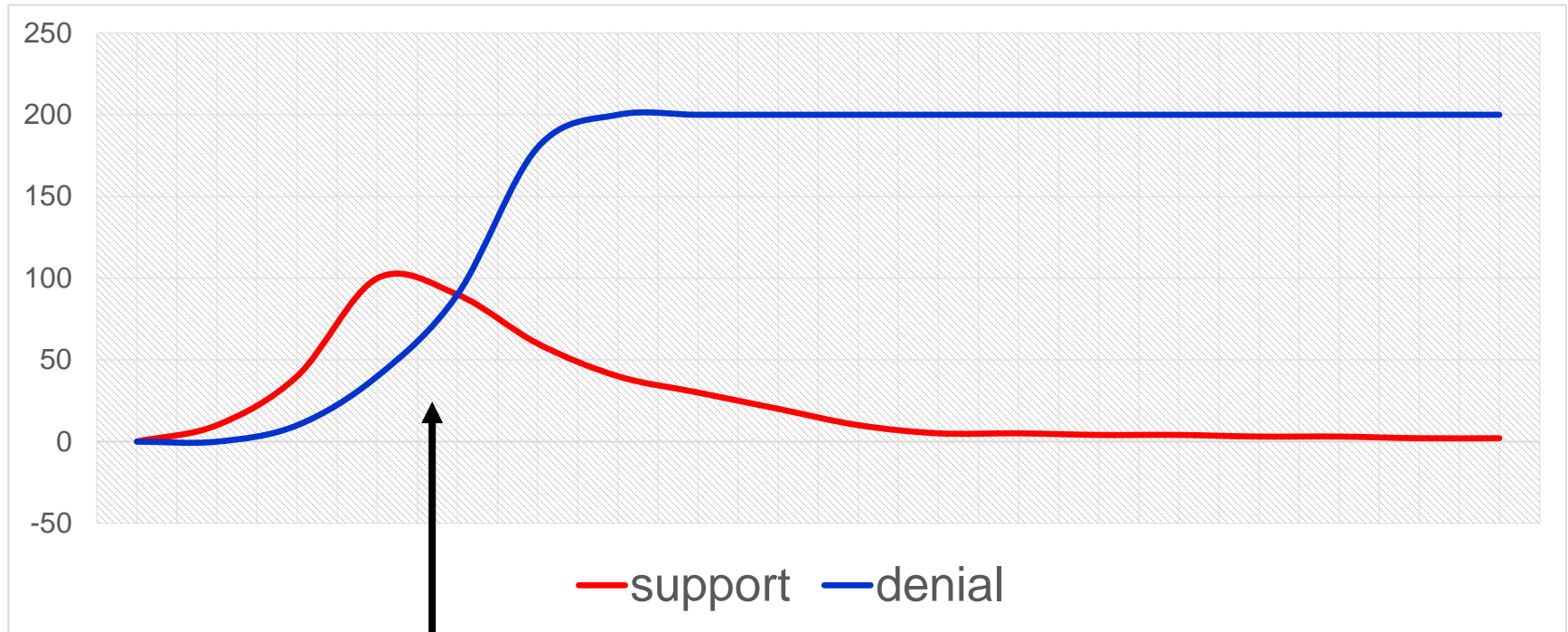
The prospect of governments forcing citizens to undergo microchip implants for tracking or other purposes is of perennial concern to the conspiracy-minded, so much so that any mention of chip implants in the news is enough to spark apocalyptic howls.

Web Evidence for Credibility Assessment

Claim	Assessment & Web Evidence
The use of solar panels drains the sun of energy.	False - Solar panels do not suck up the Sun's rays of photons. Just like wind farms do not deplete our planet of wind. These renewable sources of energy are not finite like fossil fuels.
A woman stabbed her boyfriend with a sharpened selfie stick because he didn't like her Instagram selfie.	False - A weird kind of story in heavy circulation online states ... No, the claim is not a fact.
Between 1988 and 2006, a man lived at a Paris airport.	True - Mehran Karimi Nasseri (born 1942) is an Iranian refugee who lived in the departure lounge of Terminal One in Charles de Gaulle Airport from 26 August 1988 until July 2006 ... His autobiography has been published as a book and was the basis for the 2004 Tom Hanks movie The Terminal.

Temporal Behavior of Claims & Evidence

[K. Popat et al.: CIKM'16, WWW'17]

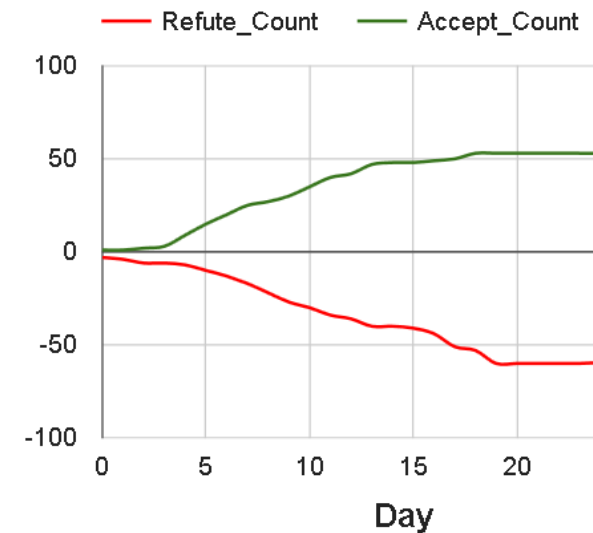
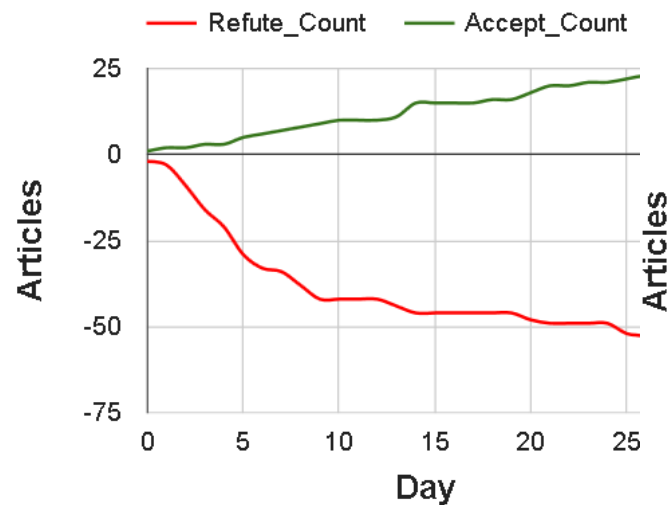
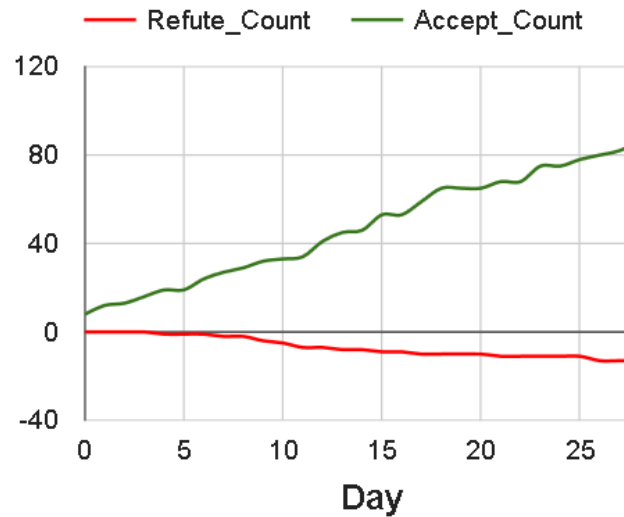


consider gradients

- **Model temporal patterns of spreading claims, support and denial**
- **Exploit for early assessment of credibility**

Temporal Footprint of Claims

[K. Popat et al.: WWW 2017]



The Centers For Disease Control confirmed that a patient in Dallas has tested positive for Ebola.



Actor Macaulay Culkin has died.



The iPhone 6 Plus will bend easily if placed in a pocket.

Research Opportunities:

- Understand & incorporate the influence of **attribution** in sources and **dissemination** across sources
- Consider **anonymous** sources (e.g. user posts in online communities)
- **Explain** veracity of claims in socio-cultural context

Outline

- ✓ **Introduction**
- ✓ **What Computers Should Know**
- ✓ **What Computers Shouldn't Know**
- ✓ **What Computers Shouldn't Believe**
- ★ **Conclusion**

Take-Home Message



more knowledge, analytics, insight



Computers shouldn't believe everything

Computers need commonsense

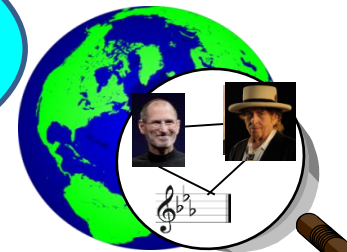


Internet Contents

knowledge acquisition

Digital Knowledge

machine reading



Computers need socio-cultural knowledge

You should know what computers know about YOU !

