Information Reliability

with Marked Temporal Point Processes

HUMAN-CENTERED MACHINE LEARNING

http://courses.mpi-sws.org/hcml-ws18/

MAX PLANCK INSTITUTE
FOR SOFTWARE SYSTEMS

Information gathering is an online activity

People can learn about a wide variety of topics in Wikipedia

People can get answers to their questions in Q&A sites

People can attend MOOCs to learn about a subject

People can be up to date with latest “news”

Opinionated, inaccurate, false facts

Wikipedia:List of hoaxes on Wikipedia

From Wikipedia, the free encyclopedia

Please do not attempt to create new hoaxes on Wikipedia; here is why.

03:21, 20 September 2016

Barack Hussein Obama II is a Kenyan politician

(Barack Obama's Wikipedia article)

The New York Times

As Fake News Spreads Lies, More Readers Shrug at the Truth

BBC

'Fake news': What's the best way to tame the beast?

BREITBART

12 FAKE NEWS STORIES FROM THE MAINSTREAM MEDIA

North Carolina For Donald Trump
October 14, 2016 ·

Like Page

Pope endorses Trump!
Game changer !!

http://endingthefed.com/pope-francis-shocks-world-endorses-...

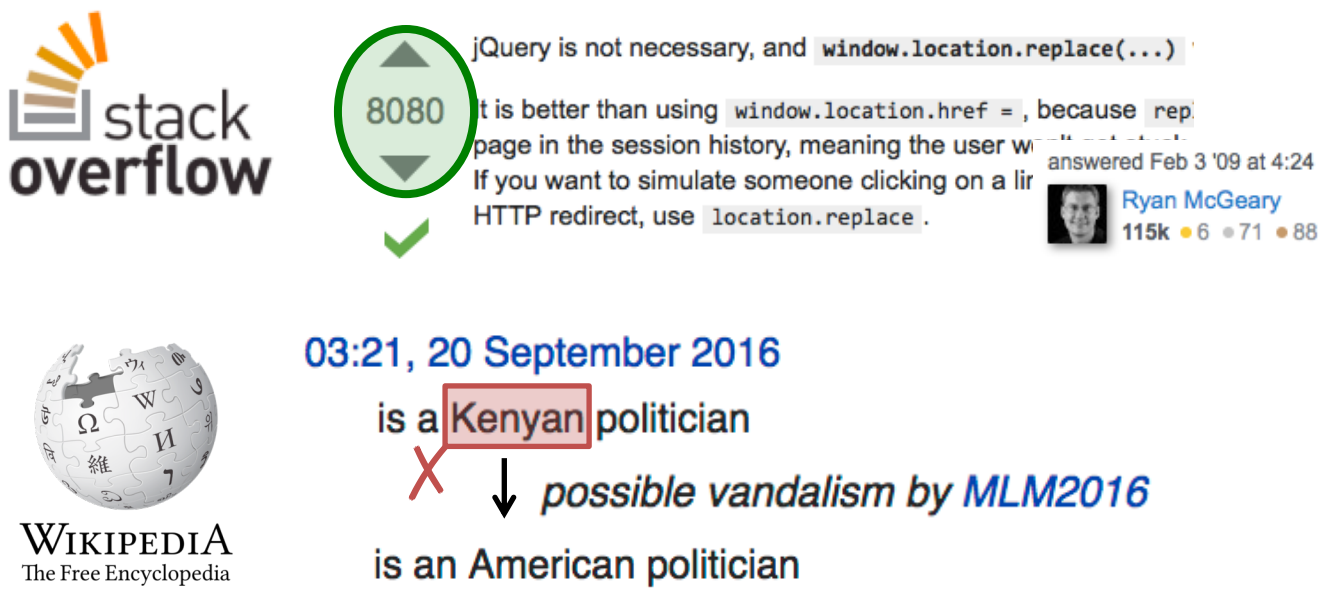
Pope Francis Shocks World, Endorses Donald Trump for President, Releases Statement

VATICAN CITY – News outlets around the world are reporting on the news that Pope Francis has made the unprecedented decision to endorse a US...

ENDINGTHEFED.COM

Solution: Resort to the crowd

Allow users to verify and/or refute information



Ubiquitous in Q&A, wikis…

Challenges

1. **Users** may be ***untrustworthy***
2. **Information** may be ***disputed***
3. **Users need to verify/refute**

Reliability & trustworthiness

Can we **quantify info reliability and source trustworthiness?**

Why this goal?

Showcase reliable information, fix unreliable information

Identify (un)trustworthy information sources

Increase information quality

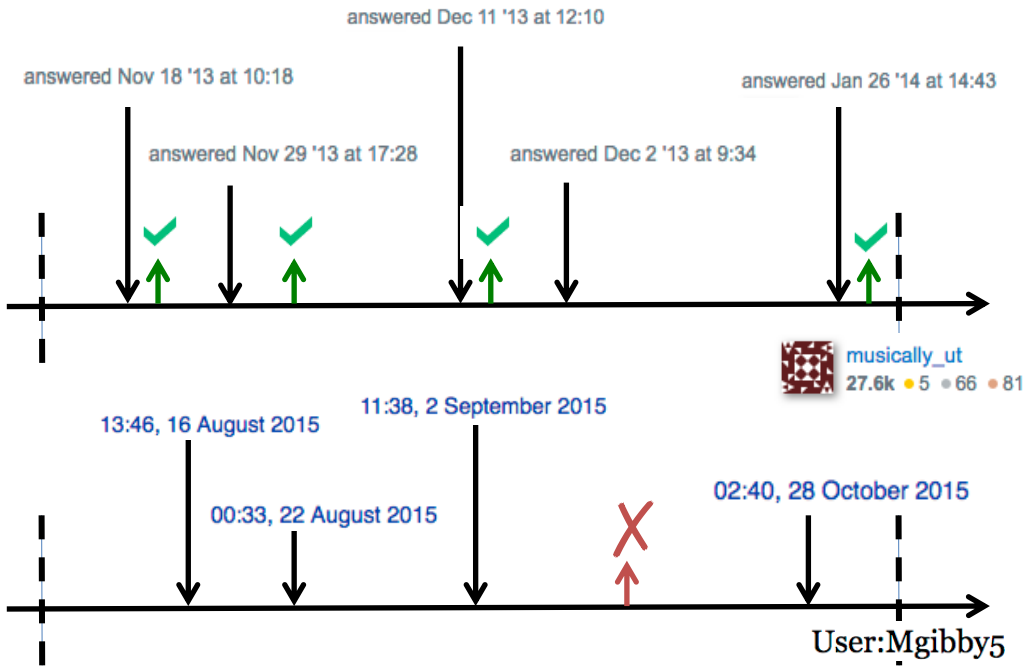
Information reliability: key, simple idea

A **source is trustworthy** if:

Its contributions are **verified more frequently**

and/or

Its contributions are **refuted more rarely**



Challenge

At a time t, a ***document*** **may be *disputed***

Verifications: rarer

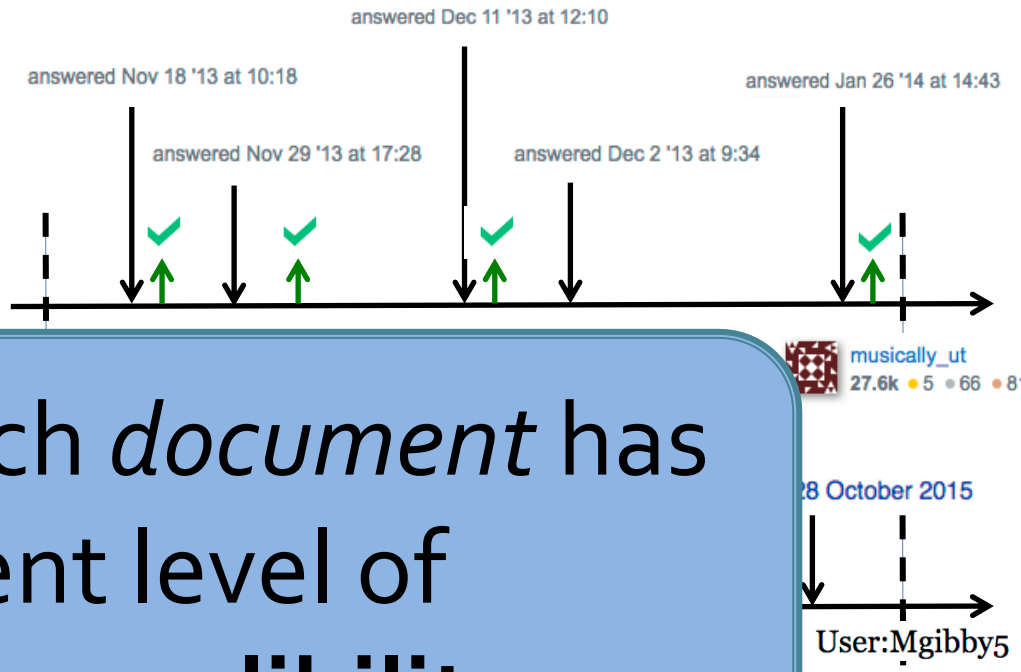
Refutations: more frequent

[Tabibian et al., WWW 2017]

Information reliability: key, simple idea

A **source is trustworthy** if:

Its contributions are
verified more frequently



Its c
refu

Over time, each *document* has a different level of ***inherent*** **unrealibility**

Challenge

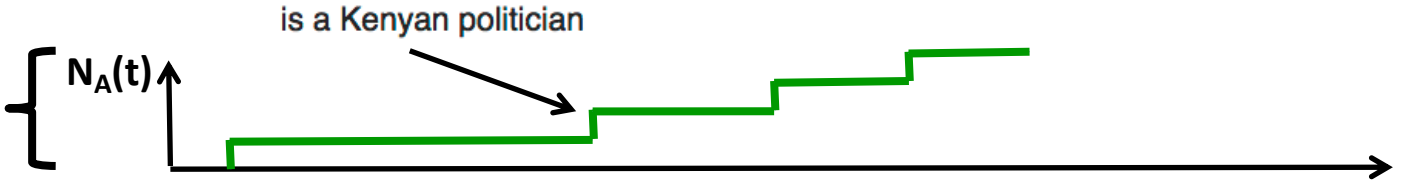
At a time t, a ***document*** **may be *disputed***

Verifications: rarer
Refutations: more frequent

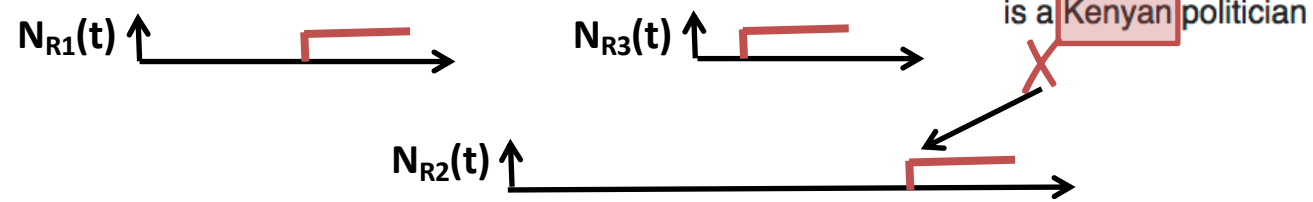
[Tabibian et al., WWW 2017]

Representation: temporal point processes

Statement additions
(one process per document)



Statement refutations
(one process per statement)




Barack Obama

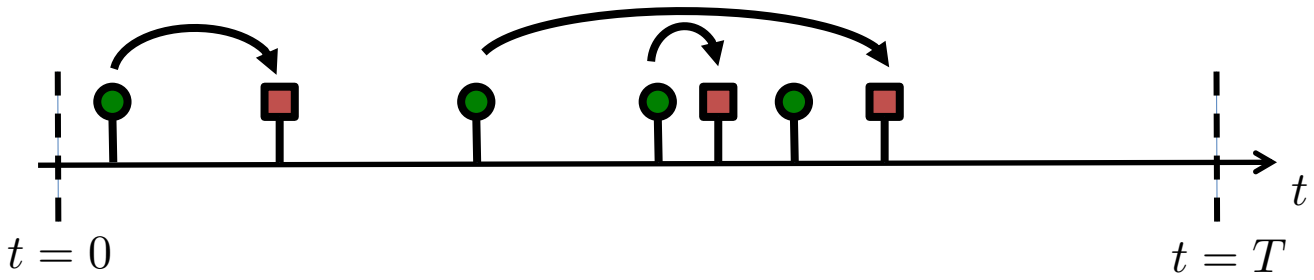
From Wikipedia, the free encyclopedia

"Barack" and "Obama" redirect here. For his father, see Barack Obama Sr. For other uses of "Barack", see Barack (disambiguation). (disambiguation).

Barack Hussein Obama II (
current President of the Unite
continental United States. Bo
was president of the *Harvard*
civil rights attorney and taugh
representing the 13th District
States House of Representat

Barack Obama: Revision history

03:41, 28 November 2016 Ranze (talk | contribs) . . (301,105 bytes) (+18) . . (E
03:32, 28 November 2016 Xin Deui (talk | contribs) . . (301,087 bytes) (-68) . . (
00:57, 28 November 2016 SporkBot (talk | contribs) m . . (301,155 bytes) (-37)
07:03, 27 November 2016 Saiph121 (talk | contribs) . . (301,192 bytes) (+25) . .



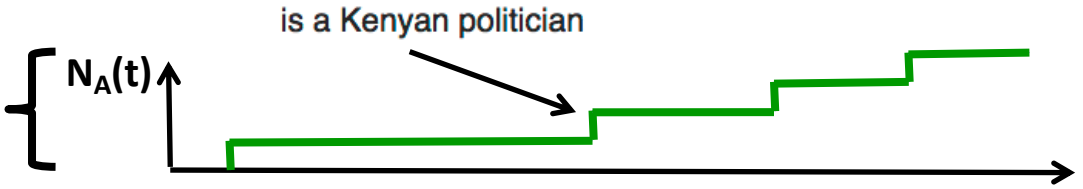
Statement: $e = (s, t, \tau)$

source — s
refutation time — τ
addition time — t

[Tabibian et al., WWW 2017]

Intensity of statement additions

Statement additions
(one process per article)



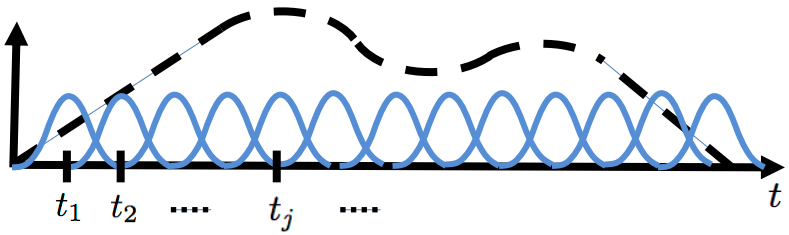
$$\lambda_d(t) = \underbrace{\sum_j \phi_{d,j} k(t - t_j)}_{\text{Article unreliability}} + \underbrace{\sum_{e_i \in \mathcal{H}_d(t)} \mathbf{w}_d^\top \gamma_{s_i} g(t - \tau_i)}_{\text{Effect of past refutations}}$$

Intensity or rate
(Statements per time unit)

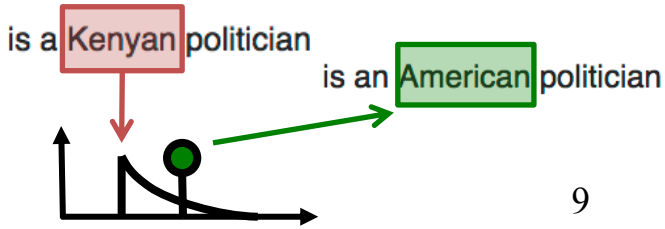
Article unreliability
(Mixture of Gaussians)

Effect of past refutations
(topic dependent; topic weight w_d)

Temporal evolution of the *intrinsic* reliability of the article

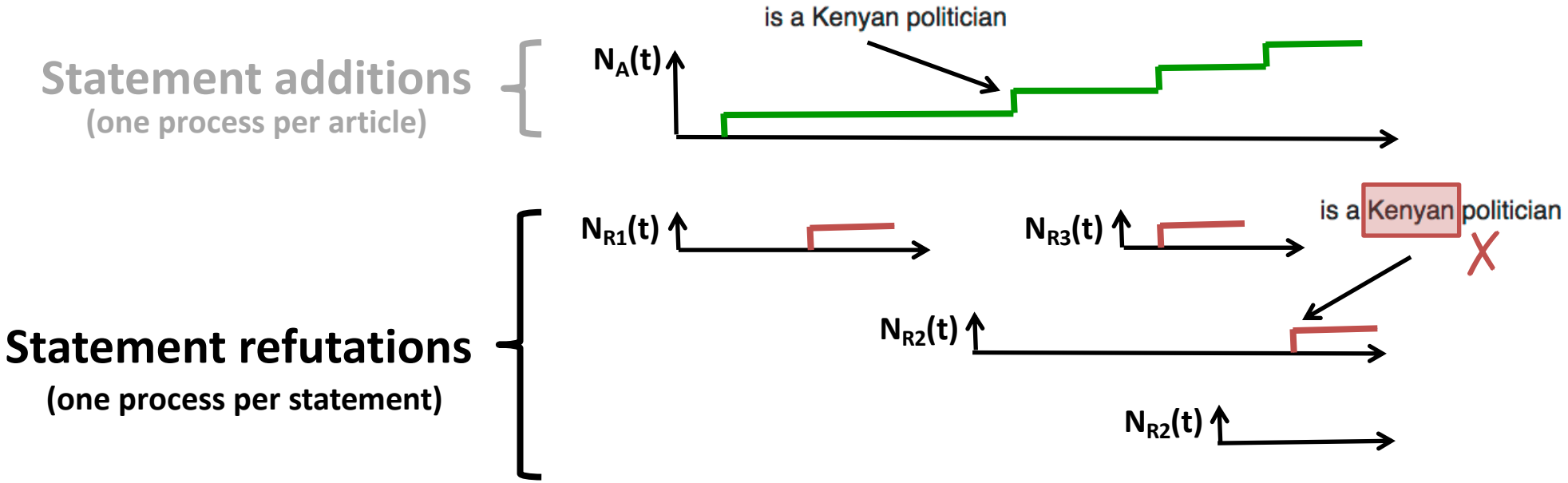


Refuted statements trigger the arrival of new statements to replace them



[Tabibian et al., WWW 2017]

Intensity of statement refutations



$$\mu_i(t) = (1 - N_i(t)) \left[\sum_j \beta_{d,j} k(t + t_i - t_j) + \mathbf{w}_d^\top \boldsymbol{\alpha}_{s_i} \right]$$

Intensity or rate
(Statements per time unit)

Refutations happen only once

Article unreliability
(Mixture of Gaussians)

Shared across statements of an article!

Source trustworthiness
(topic dependent; topic weight w_d)

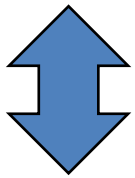
The higher the parameter α_{s_i}, the quickest an article gets refuted

[Tabibian et al., WWW 2017]

Model inference from event data

Conditional intensities

$$\{\lambda_d(t)\} \qquad \{\mu_i(t)\}$$



Events likelihood

$$\sum_{d=1}^{|\mathcal{D}|} \sum_{i:e_i \in \mathcal{H}_d(T)} \log \underbrace{p(t_i | \mathcal{H}_d(t_i), \phi_d, \{\gamma_s\}, \mathbf{w}_d)}_{\text{statements additions}} + \sum_{d=1}^{|\mathcal{D}|} \sum_{i:e_i \in \mathcal{H}_d(T)} \log \underbrace{p(\Delta_i | t_i, \beta_d, \{\alpha_s\}, \mathbf{w}_d)}_{\text{statements evaluations}}$$

Theorem. The **maximum likelihood** problem is **convex in the model parameters**.

[Tabibian et al., WWW 2017]

Wikipedia dataset

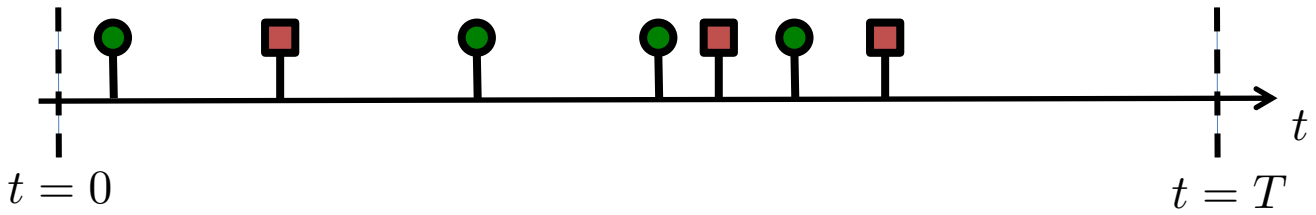
Complete edit history of Wikipedia up to July, 2014

→ **50k web sources** (with more than 10 additions)

→ **who were used in 100k articles** (with more than 20 additions)

by means of **10.4 million statement additions**

9 million statement refutations

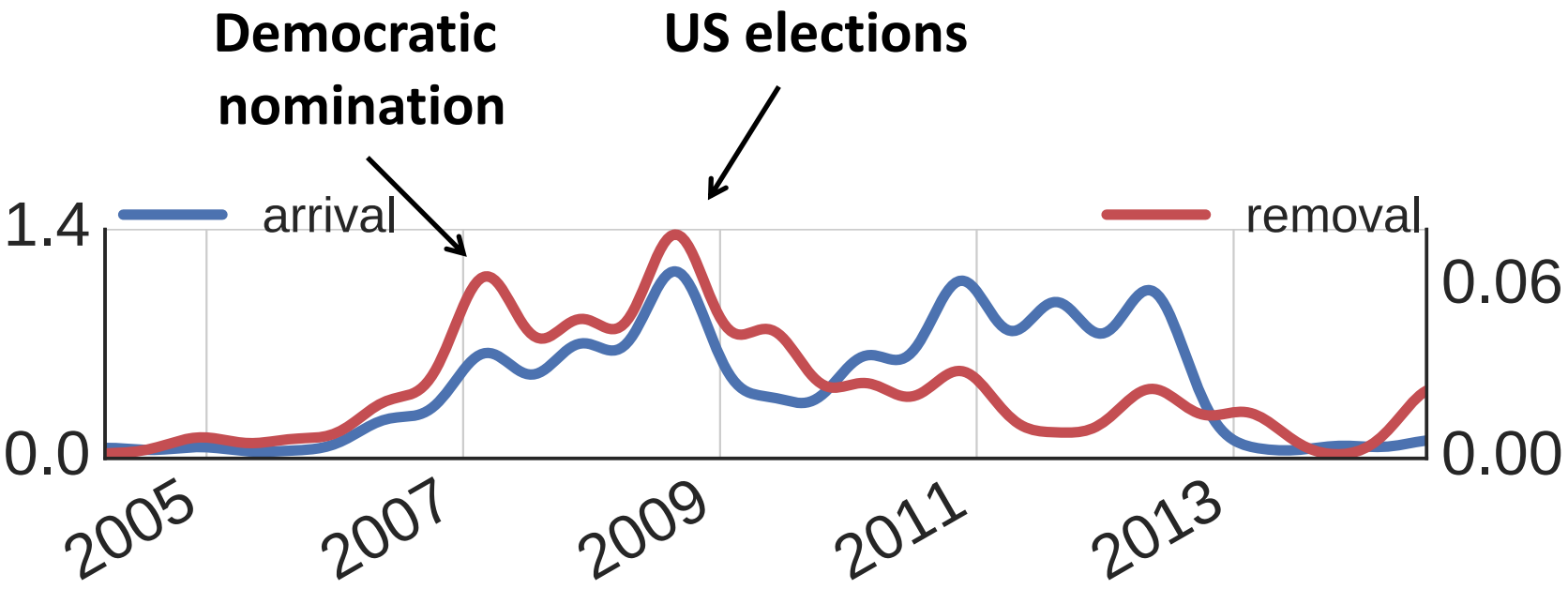


→ **10 topics**

[Tabibian et al., WWW 2017]

What can the model tell us about the **article unreliability**?

Barack Obama's biography



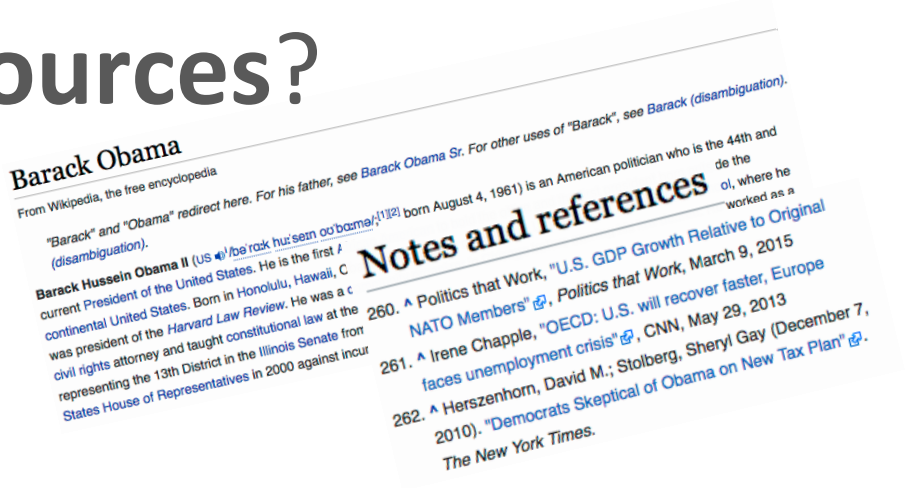
Inferred intrinsic unreliability

Peaks match noteworthy events

Difference between arrival and removal indicates controversy

[Tabibian et al., WWW 2017]

What about the **trustworthiness of the sources**?



Barack Obama

From Wikipedia, the free encyclopedia

"Barack" and "Obama" redirect here. For his father, see Barack Obama Sr. For other uses of "Barack", see Barack (disambiguation). (disambiguation).

Barack Hussein Obama II (US /bəˈrɑːk huːˈseɪn oʊˈbɑːmə/;[1][2] born August 4, 1961) is an American politician who is the 44th and current President of the United States. He is the first A de the continental United States. Born in Honolulu, Hawaii, C ol, where he was president of the *Harvard Law Review*. He was a c worked as a civil rights attorney and taught constitutional law at the representing the 13th District in the Illinois Senate fron States House of Representatives in 2000 against incur

Notes and references

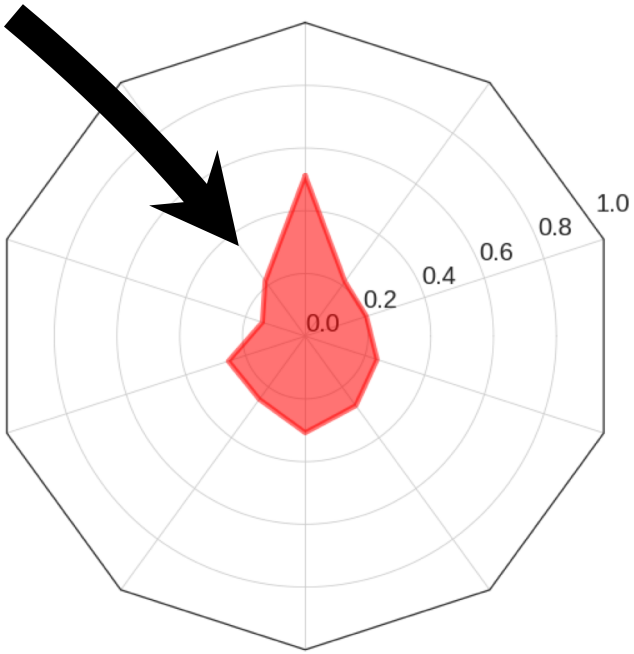
260. ^ Politics that Work, "U.S. GDP Growth Relative to Original NATO Members", *Politics that Work*, March 9, 2015

261. ^ Irene Chapple, "OECD: U.S. will recover faster, Europe faces unemployment crisis", CNN, May 29, 2013

262. ^ Herszenhorn, David M.; Stolberg, Sheryl Gay (December 7, 2010). "Democrats Skeptical of Obama on New Tax Plan". *The New York Times*.

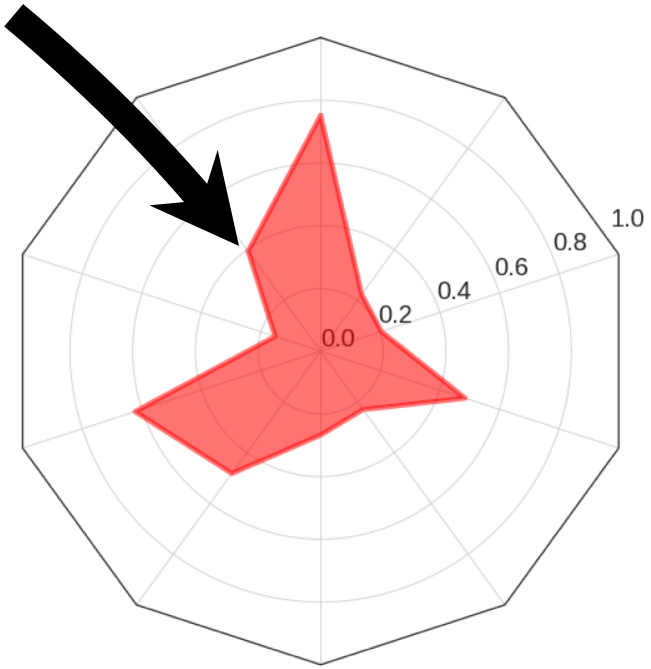
Source trustworthiness

Politics



bbc.co.uk

Politics



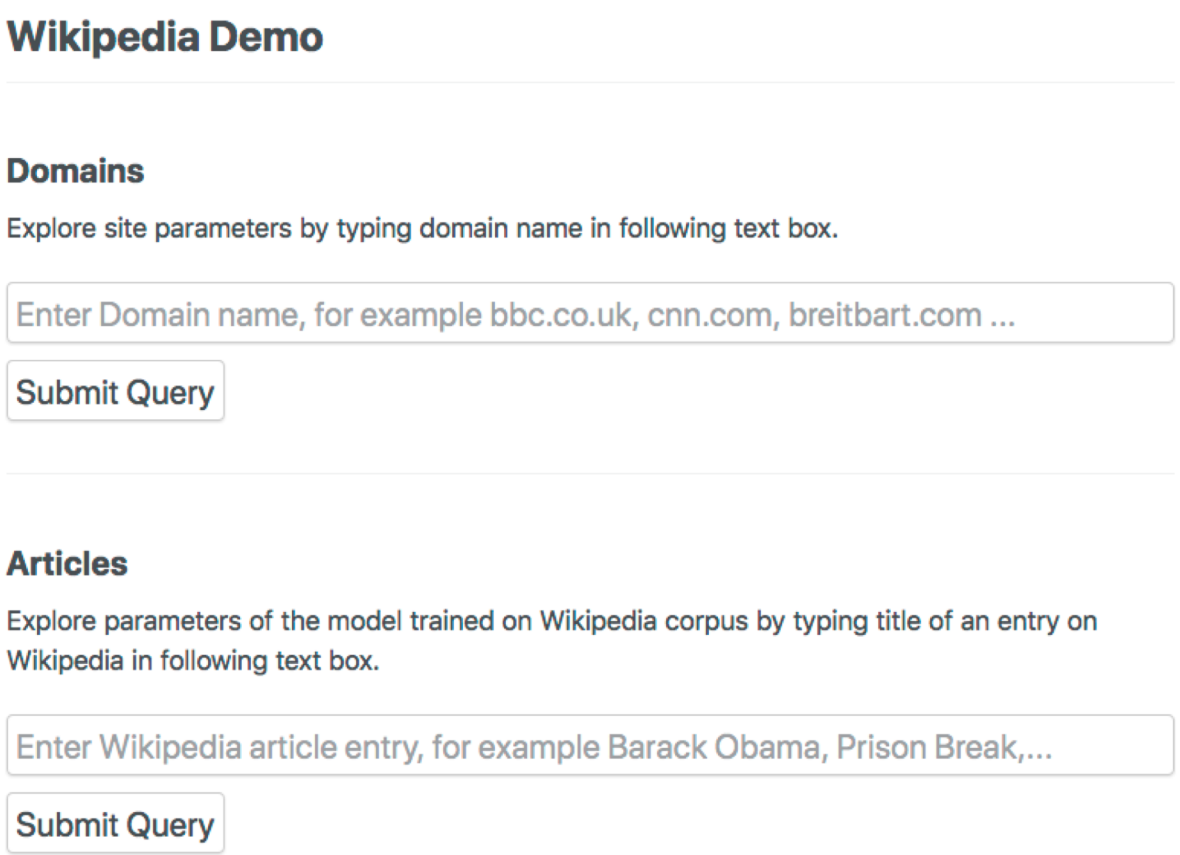
breitbart.com

Probability of refutation within 6 months in a *stable* Wikipedia article

[Tabibian et al., WWW 2017]

Demo for other pages and sources

http://btabibian.com/projects/reliability/



[Tabibian et al., WWW 2017]