



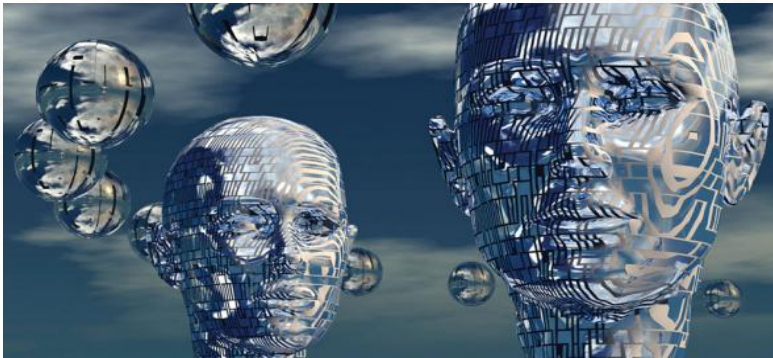
max planck institut
informatik

Information Veracity

Subhabrata Mukherjee
Max Planck Institute for Informatics

June 10, 2014

Online User Generated Data

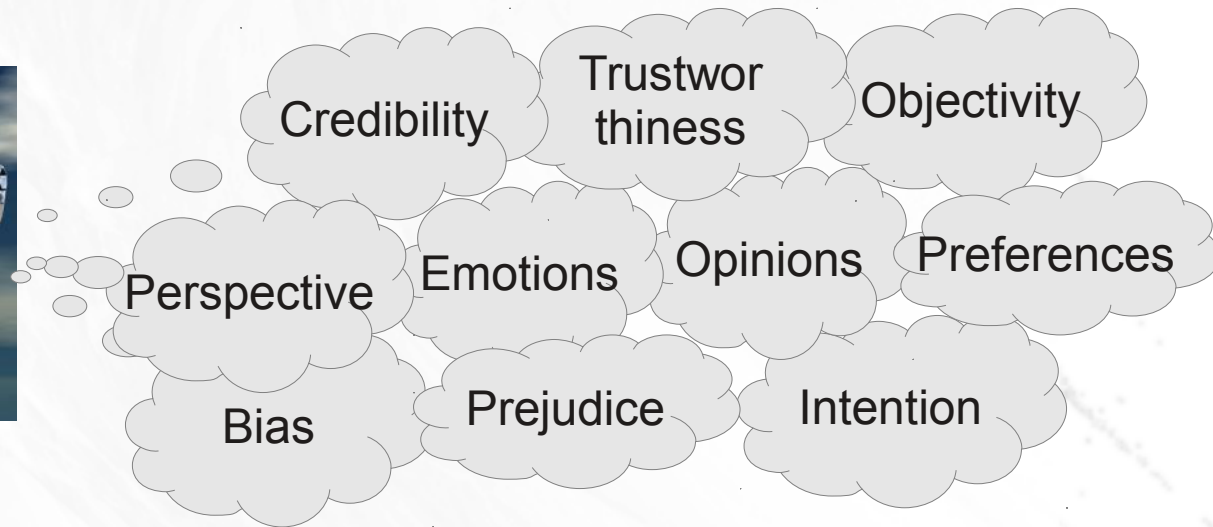
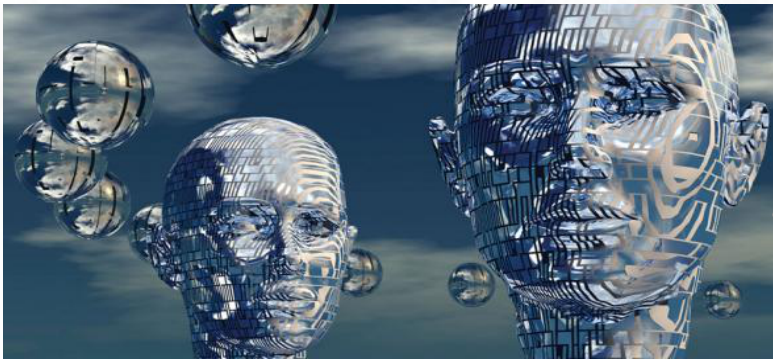


MAX-PLANCK-GESELLSCHAFT

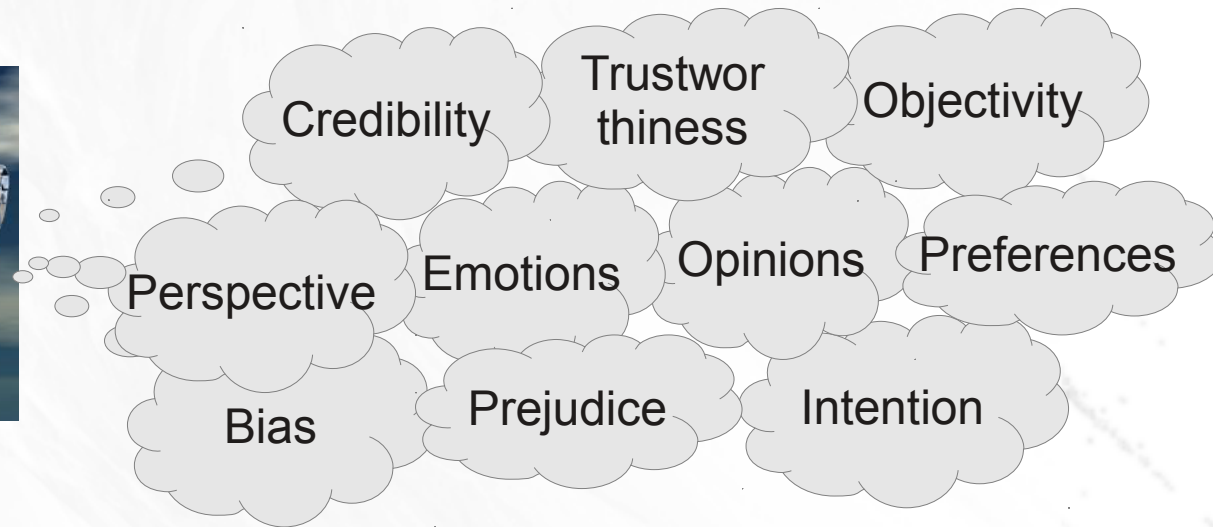
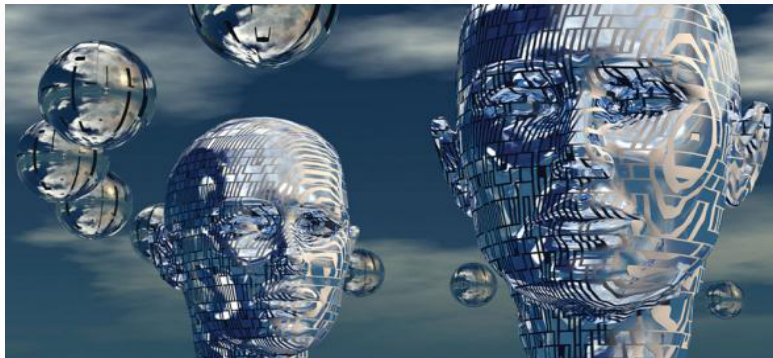


max planck institut
informatik

Online User Generated Data



Online User Generated Data



Veracity ?



MAX-PLANCK-GESELLSCHAFT

mpi max planck institut
informatik

Information Veracity

➤ Motivation

- 86% online users concerned about information *veracity*, more than 30% agree that information is not *trustworthy*
- 28% report frustration in retrieving relevant information, 40% concerned about quality and 20% did not understand



Information Veracity

- **Motivation**

- 86% online users concerned about information *veracity*, more than 30% agree that information is not *trustworthy*
- 28% report frustration in retrieving relevant information, 40% concerned about quality and 20% did not understand

- **Objective**

- Analyze *joint interaction* between user *trustworthiness*, *objectivity* of posts and *credibility* of statements therein



Information Veracity

➤ Motivation

- 86% online users concerned about information *veracity*, more than 30% agree that information is not *trustworthy*
- 28% report frustration in retrieving relevant information, 40% concerned about quality and 20% did not understand

➤ Objective

- Analyze *joint interaction* between user *trustworthiness*, *objectivity* of posts and *credibility* of statements therein

➤ Application

- Extracting drug side-effects from user posts in health community
- Identifying rumors, mis-information in news articles



Language vs. Credibility

Naive User

“I heard Xanax can have pretty bad side-effects. You may have peeling of skin, and apparently some friend of mine told me you can develop ulcers in the lips also. If you take this medicine for a long time then you would probably develop a lot of other physical problems. Which of these did you experience ?”



Language vs. Credibility

Naive User

“I heard Xanax can have pretty bad side-effects. You may have peeling of skin, and apparently some friend of mine told me you can develop ulcers in the lips also. If you take this medicine for a long time then you would probably develop a lot of other physical problems. Which of these did you experience ?”

Naive User

“I’ve had chronic depression off and on since adolescence. In the past I’ve taken Paxil (made me anxious) and Zoloft (caused insomnia and stomach problems, but at least I was mellow). I have been taking St. John’s Wort for a few months now, and it helps, but not enough. I wake up almost every morning feeling very sad and hopeless. As afternoon approaches I start to feel better, but there’s almost always at least a low level of depression there throughout the day. ”

Perspective, Time vs. Credibility



Katy @KatyinIndy · 29 Jun 2012

By March 2013, **Obamacare** requires you to like up to implant your **RFID chip** is.gd/RN7vQR [#cot](#) [#gop](#) [#eaparty](#) [#sgp](#) [#lot](#)

Expand

Reply Retweet Favorite More



Raw Milk Colorado @RawMilkColorado · 29 Jun 2012

@dmmckinney Hoax. metabunk.org/threads/497-De...

View conversation

Reply Retweet Favorite More



Bitter Pastor's Kid @Bitter_PK · 29 Jun 2012

@Hollivan @smokeeater1978 So apparently that's all a hoax... Will send more info as I find out more. metabunk.org/threads/497-De...

View conversation

Reply Retweet Favorite More

Rumor : Obamacare requires RFID microcip implant in all





MAX-PLANCK-GESELLSCHAFT




max planck institut
informatik

Perspective, Time vs. Credibility

 **Katy** @KatyinIndy · 29 Jun 2012
By March 2013, **Obamacare** requires you to like up to implant your **RFID chip** is.gd/RN7vQR #cot #gop #eaparty #sgp #lot
Expand Reply Retweet Favorite More

 **Raw Milk Colorado** @RawMilkColorado · 29 Jun 2012
[@dmmckinney](#) Hoax. metabunk.org/threads/497-De...
View conversation Reply Retweet Favorite More

 **Bitter Pastor's Kid** @Bitter_PK · 29 Jun 2012
[@Hollivan](#) [@smokeeater1978](#) So apparently that's all a hoax... Will send more info as I find out more. metabunk.org/threads/497-De...
View conversation Reply Retweet Favorite More

Rumor : Obamacare requires RFID microchip implant in all



MAX-PLANCK-GESELLSCHAFT

mp max planck institut
informatik

Perspective, Time vs. Credibility

 **Katy** @KatyinIndy · 29 Jun 2012
By March 2013, **Obamacare** requires you to like up to implant your **RFID**
c...
Ex... Authority of Katy (141K Followers) >> Raw Milk Colorado(1183) and Bitter Pastor's Kid (29) ★ Favorite ... More

 **Raw Milk Colorado** @RawMilkColorado · 29 Jun 2012
@dmmckinney Hoax. metabunk.org/threads/497-De...
View conversation Reply Retweet ★ Favorite ... More

 **Bitter Pastor's Kid** @Bitter_PK · 29 Jun 2012
@Hollivan @smokeeater1978 So apparently that's all a hoax... Will send more info as I find out more. metabunk.org/threads/497-De...
View conversation Reply Retweet ★ Favorite ... More

Rumor : Obamacare requires RFID microchip implant in all



MAX-PLANCK-GESELLSCHAFT

mp max planck institut
informatik

Perspective, Time vs. Credibility

 **Katy** @KatyinIndy · 29 Jun 2012
By March 2013, **Obamacare** requires you to like up to implant your **RFID**
c
Ex Authority of Katy (141K Followers) >> Raw Milk Colorado(1183) and Bitter Pastor's Kid (29) ★ Favorite ... More

 **Raw Milk Colorado** @RawMilkColorado · 29 Jun 2012

... More
more
info as I find out more. metabunk.org/threads/497-De...
View conversation Reply Retweet ★ Favorite ... More

Rumor : Obamacare requires RFID microcip implant in all



MAX-PLANCK-GESELLSCHAFT



max planck institut
informatik

PageRank (Network Features) vs. Credibility

☆ Whose Ox Is Gored?

 WALL STREET JOURNAL by JOHN FUND | MAR. 19, 2007 (News Analysis)

The media discover the former vice president's environmental exaggerations and hypocrisy.

The media are finally catching up with Al Gore. Criticism of his anti-global-warming franchise and his personal environmental record has gone beyond ankle-biting bloggers. It's now coming from the New York Times and the Nashville Tennessean... [Full Story »](#)

Now We Know... Obamacare Includes 3.8% Tax On Home Sales For All Americans

 FIRST THINGS.COM by JIM HOFT MAY. 13, 2010

"We Have to Pass the Bill So That You Can Find Out What Is In It" Remember that line by Speaker Pelosi... It will come back to haunt you repeatedly in the days ahead. Now that it's been a few weeks since ...

☆ Obama tops Bush at ducking reporters

 WASHINGTON TIMES by JOSEPH CURL | FEB. 22, 2010 (News Report)

President Obama, who pledged to establish the most open and transparent administration in history, on Monday surpasses his predecessor's record for avoiding a full-fledged question-and-answer session with White House reporters in a formal press conference. [Full Story »](#)

PageRank (Network Features) vs. Credibility

☆ Whose Ox Is Gored?

 **WALL STREET JOURNAL** by JOHN FUND | MAR. 19, 2007 (News Analysis)

The media discover the former vice president's environmental exaggerations and hypocrisy.


The media are finally catching up with Al Gore. Criticism of his anti-global-warming franchise and his personal environmental record has gone beyond ankle-biting bloggers. It's now coming from the New York Times and the Nashville Tennessean... [Full Story »](#)

Now We Know... Obamacare Includes 3.8% Tax On Home Sales For All Americans

 **FIRST THINGS.COM** by JIM HOFT MAY. 13, 2010

"We Have to Pass the Bill So That You Can Find Out What Is In It" Remember that line by Speaker Pelosi... It will come back to haunt you repeatedly in the days ahead. Now that it's been a few weeks since ...

☆ Obama tops Bush at ducking reporters

 **WASHINGTON TIMES** by JOSEPH CURL | FEB. 22, 2010 (News Report)

President Obama, who pledged to establish the most open and transparent administration in history, on Monday surpasses his predecessor's record for avoiding a full-fledged question-and-answer session with White House reporters in a formal press conference. [Full Story »](#)

Summary

- Network, generic IR word-level features not enough to measure credibility



Summary

- Network, generic IR word-level features not enough to measure credibility
- *Style, Content, Form* of data need to be analyzed



Summary

- Network, generic IR word-level features not enough to measure credibility
- *Style, Content, Form* of data need to be analyzed
- **Interaction**¹ between **users** (network, perspectives ...), **language** of posts (style, form, content) and **statements** (truthfulness, knowledge-base) influence information veracity

