

Taming the trolls of social media

Richard Harvey

IT Livery Company Professor of Information Technology, Gresham College

Professor of Computer Science, School of Computing Sciences,
University of East Anglia

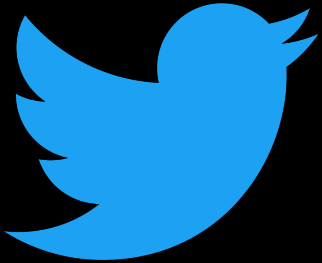
#richardwharvey

What is social media?

1. Web 2.0 Internet-based applications;
2. User-generated content is the lifeblood of social media;
3. Allows creation of user-specific profiles for a site or app designed and maintained by a social media service;
4. facilitate the development of social networks online by connecting a profile with those of other individuals and/or groups.

Obar, Jonathan & Wildman, Steven. (2015). Social Media Definition and the Governance Challenge: An Introduction to the Special Issue. SSRN Electronic Journal. 10.2139/ssrn.2647377.

Examples of Social Media



Critical feature of social media

- Anyone can join
- Anyone can publish anything to as many people will listen
- Instant
- No intermediaries or censorship

Social media buzzwords

Flame – flame war

Troll

Internet

Astroturf

Bot

Meme

Sentiment



factcheckUK
FROM CCHQ

Tweets 47.6K
Following 2,251
Followers 75.2K
Likes 1,953
Lists 6

Tweets Tweets & replies Media

factcheckUK Retweeted



Therese Coffey @theresecoffey · 41s

See where Labour actually runs the NHS - Wales. It's the only part of the UK where the NHS budget was cut - yes by a Labour government [#ILoveNHS](#) [#BackBoris](#) [#LeadersDebate](#)



factcheckUK

@CCHQPress

Fact checking Labour from CCHQ

Westminster

Joined July 2009

Photos and videos

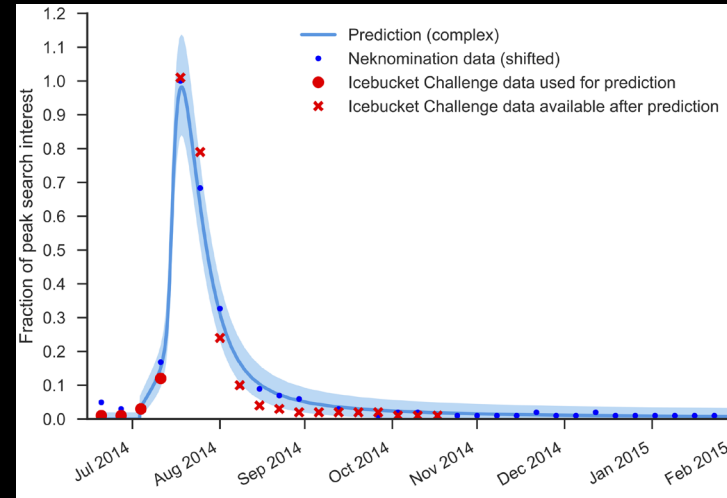
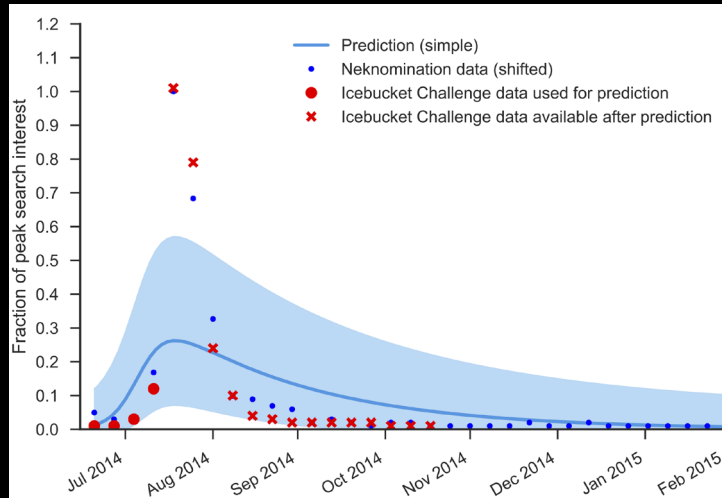


Want to take advantage of
the new Twitter feature?

It's simple – just look

Social media contagions

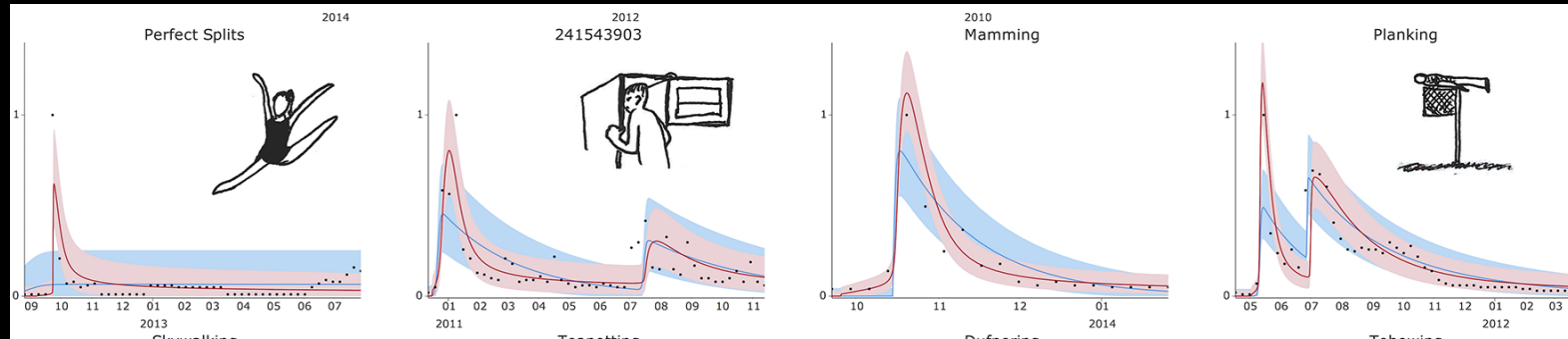
- Social media propagation is best modelled as an epidemic
- Sometimes as “complex contagion”



Sprague DA, House T (2017) *Evidence for complex contagion models of social contagion from observational data.* PLoS ONE 12(7): e0180802. <https://doi.org/10.1371/journal.pone.0180802>

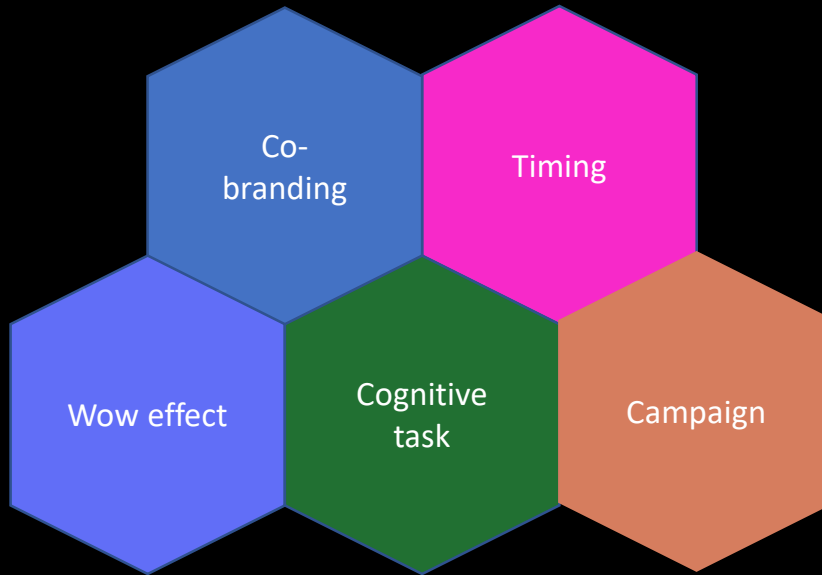
Social media contagions

- Social media propagation is best modelled as an epidemic
- Sometimes a “complex contagion”



Sprague DA, House T (2017) *Evidence for complex contagion models of social contagion from observational data*. PLoS ONE 12(7): e0180802. <https://doi.org/10.1371/journal.pone.0180802>

Five patterns of successful social media



Eismann T.T., Wagner T.F., Baccarella C.V., Voigt K.I. (2018) This Is How We Do It: Untangling Patterns of Super Successful Social Media Activities. In: Meiselwitz G. (eds) Social Computing and Social Media. User Experience and Behavior. SCSM 2018. Lecture Notes in Computer Science, vol 10913. Springer, Cham, https://doi.org/10.1007/978-3-319-91521-0_17

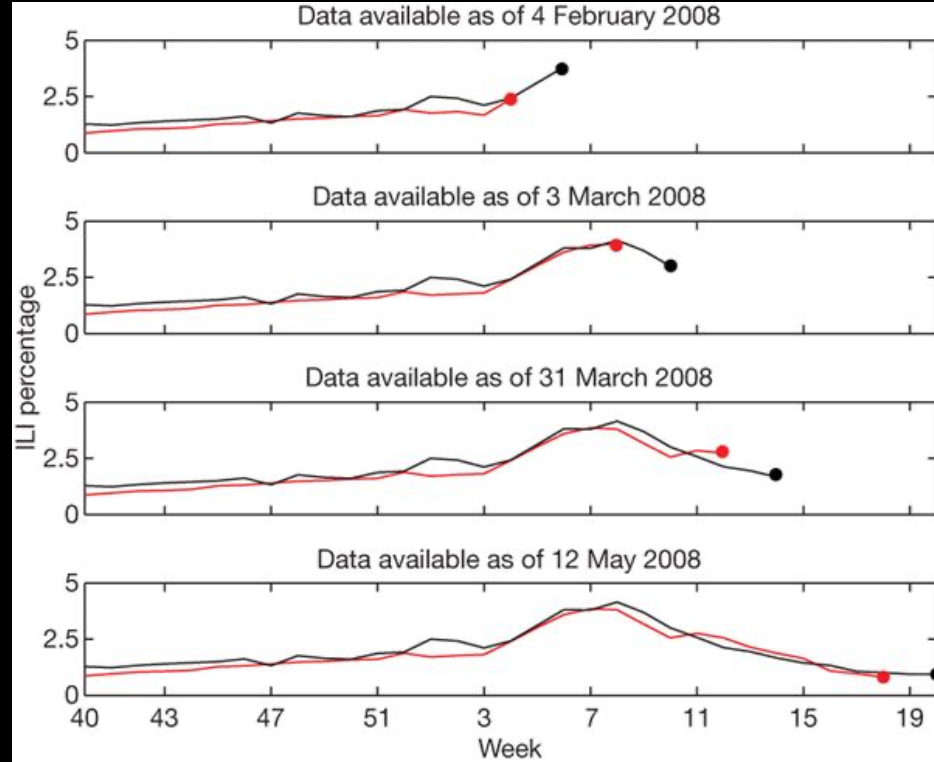
Sentiment analysis

n	Most negative n -grams	Most positive n -grams
1	bad; boring; dull; flat; pointless; tv; neither; pretentious; badly; worst; lame; mediocre; lack; routine; loud; bore; barely; stupid; tired; poorly; suffers; heavy;nor; choppy; superficial	touching; enjoyable; powerful; warm; moving; culture; flaws; provides; engrossing; wonderful; beautiful; quiet; socio-political; thoughtful; portrait; refreshingly; chilling; rich; beautifully; solid;
2	how bad; by bad; dull .; for bad; to bad; boring .; , dull; are bad; that bad; boring .; , flat; pointless .; badly by; on tv; so routine; lack the; mediocre .; a generic; stupid .; abysmally pathetic	the beautiful; moving.; thoughtful and; , inventive; solid and; a beautiful; a beautifully; and hilarious; with dazzling; provides the; provides.; and inventive; as powerful; moving and; a moving; a powerful
3	. too bad; exactly how bad; and never dull; shot but dull; is more boring; to the dull; dull, UNK; it is bad; or just plain; by turns pretentious; manipulative and contrived; bag of stale; is a bad; the whole mildly; contrived pastiche of; from this choppy; stale material.	funny and touching; a small gem; with a moving; cuts, fast; , fine music; smart and taut; culture into a; romantic , riveting; ... a solid; beautifully acted .; , gradually reveals; with the chilling; cast of solid; has a solid; spare yet audacious; ... a polished; both the beauty;
5	boring than anything else.; a major waste ... generic; nothing i hadn't already; ,UNK plotting;superficial; problem ? no laughs.; just horribly mediocre .; dull, UNK feel.; there's nothing exactly wrong; movie is about a boring; essentially a collection of bits	reminded us that a feel-good; engrossing, seldom UNK.; between realistic characters showing honest; a solid piece of journalistic; easily the most thoughtful fictional; cute, funny, heartwarming; with wry humor and genuine; engrossing and ultimately tragic.;
8	loud, silly, stupid and pointless.; dull, dumb and derivative horror film.; UNK's film, a boring, pretentious; this film biggest problem ? no laughs.; film in the series looks and feels tired; do draw easy chuckles but lead nowhere.; stupid, infantile, redundant, sloppy	shot in rich , shadowy black-and-white , devils an escapist confection that 's pure entertainment .; , deeply absorbing piece that works as a; ... one of the most ingenious and entertaining; film is a riveting , brisk delight .; bringing richer meaning to the story 's;

Richard Socher, Jeffrey Pennington, Eric H. Huang, Andrew Y. Ng, and Christopher D. Manning. 2011. *Semi-supervised recursive autoencoders for predicting sentiment distributions*. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '11). Association for Computational Linguistics, Stroudsburg, PA, USA, 151-161.

Google Flu Trends

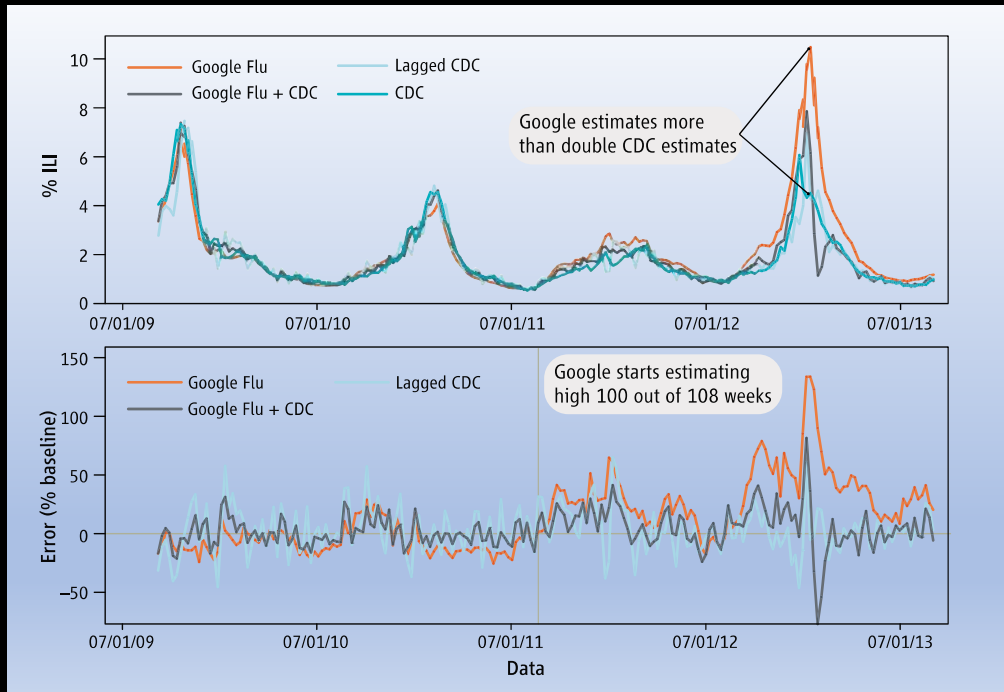
- Examined top-50 million queries over a five-year period
- Search for queries that correlate with Influenza-Like Illness (ILI) physician visits.



Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, et al. (2008) *Detecting influenza epidemic using search engine query data*. Nature 457: 1012–10155

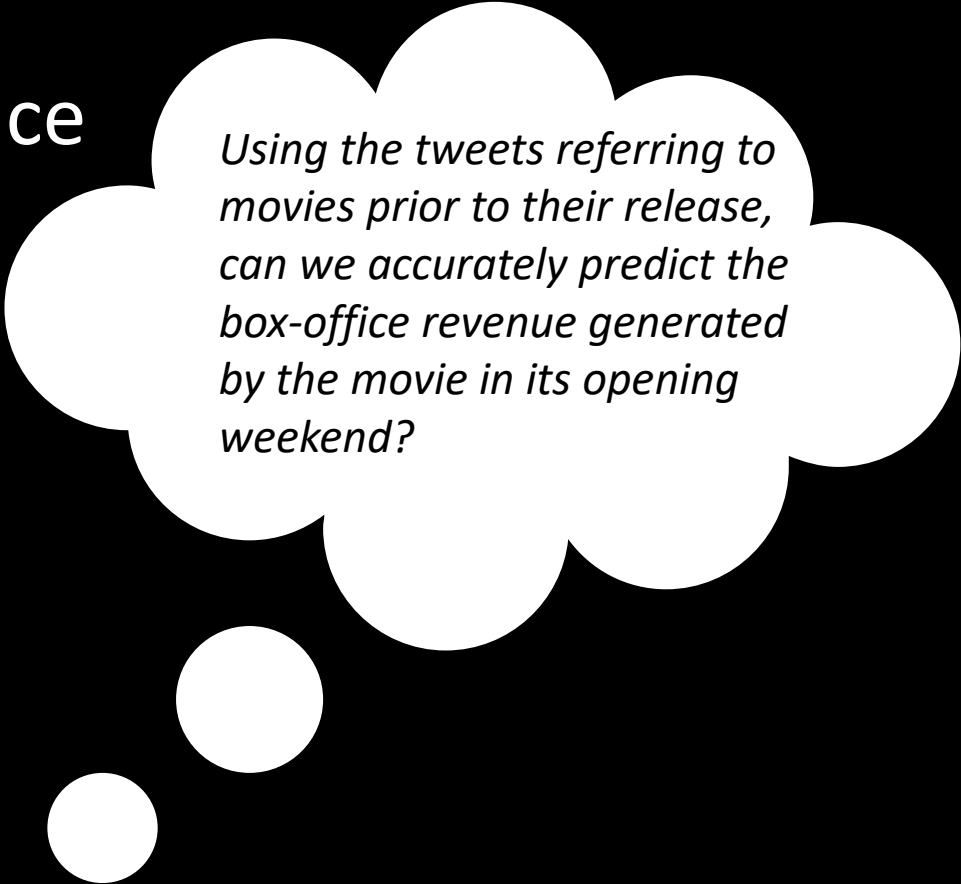
Google Flu Trends

- Consistently over-estimated
- Query terms never published
- Model over-fitted the data
- GFT closed in 2013



The Parable of Google Flu: Traps in Big Data Analysis, David Lazer, Ryan Kennedy, Garry King and Alessandro Vespignani, Science, 14 Mar 2014 : 1203-1205.

Twitter at the box office

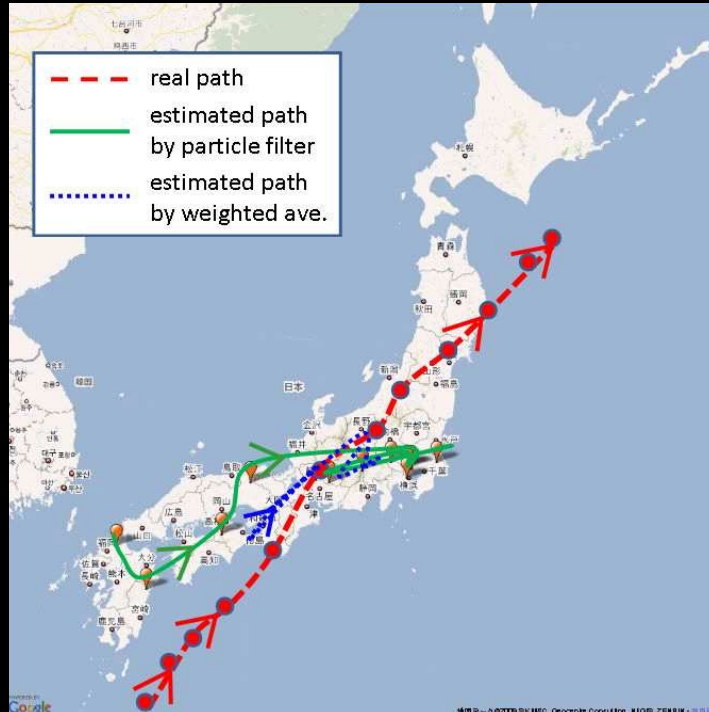


Using the tweets referring to movies prior to their release, can we accurately predict the box-office revenue generated by the movie in its opening weekend?

HSX versus twitter

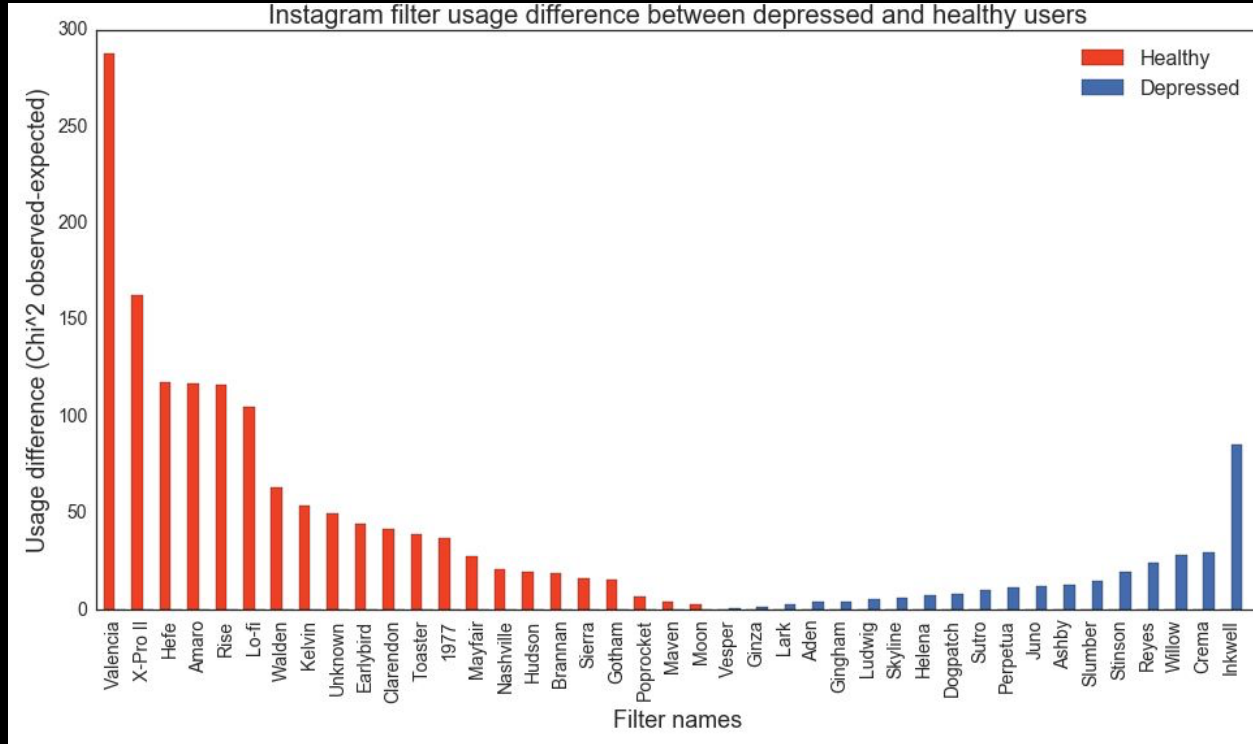
Hollywood Stock Exchange
(HSX) is the gold standard

Typhoon and earthquake tracking

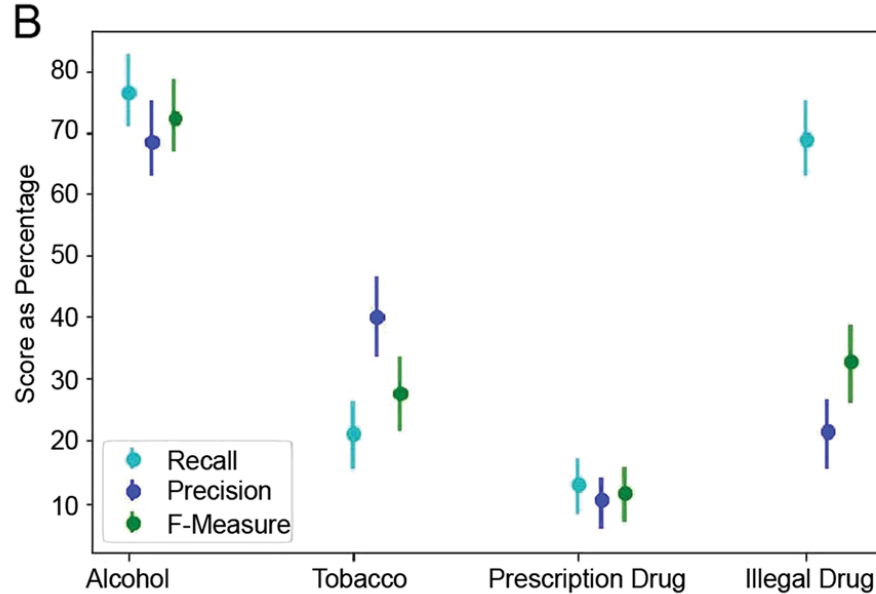


Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. 2010. *Earthquake shakes Twitter users: real-time event detection by social sensors*. In Proceedings of the 19th international conference on World wide web (WWW '10). ACM, New York, NY, USA, 851-860.

Instagram - depression



Instagram – alcohol use



Hassanpour, S., Tomita, N., DeLise, T. *et al.* Identifying substance use risk based on deep neural networks and Instagram social media data. *Neuropsychopharmacol* **44**, 487–494 (2019)



Moral panics

1. Someone, something or a group are defined as a threat to social norms or community interests;
2. The threat is then depicted in a simple and recognizable symbol/form by the media;
3. The portrayal of this symbol rouses public concern;
4. There is a response from authorities and policy makers;
5. The moral panic over the issue results in social changes within the community.

Techno panics

1. Some technologies are defined as a threat to social norms or community interests;
2. The threat is then depicted in a simple and recognisable symbol/form by the media;
3. The portrayal of this symbol rouses public concern;
4. There is a response from authorities and policy makers;
5. The techno panic over the issue results in social changes within the community.

Internet addictions and other flora and fauna

- Screen time
- Dunbar's number
- Social media addictions
- Instagram obsessions

Bad memes

Youtube: Muslim demographics: the truth Mar 2009
16.4M views



Bad memes

“8.1 children per Muslim family”

“Nine million Muslims in the U.S.”

“90% of immigration to the EU is Islamic”

“50% of newborns are Muslim”

“25% of Belgians are Muslim”

Youtube: BBC Muslim demographics: the truth Aug 2009

230k views

Misinformation

- Some misinformation tweets can be spotted
- Some users are predictable spreaders of misinformation



Reema Aswani, Arpan Kumar Kar, and P. Vigneswara Ilavarasan. 2019. *Experience: Managing Misinformation in Social Media—Insights for Policymakers from Twitter Analytics*. J. Data and Information Quality 12, 1, Article 6 (November 2019), 18 pages.



What next?

- Solutions to epidemic behaviour
 - Don't pass on the “disease” (self censorship)
 - Isolate the “infected people” (censorship)
 - Immunisation (education)
- More work needed!

Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence, Cook J, Lewandowsky S, Ecker,. 2017, PLOS ONE 12(5): e0175799.

What next?

- Determination and propagation of trust
- More work needed!

Eugene Agichtein, Carlos Castillo, Debora Donato, Aristides Gionis, and Gilad Mishne. 2008. *Finding high-quality content in social media*. In Proceedings of the 2008 International Conference on Web Search and Data Mining (WSDM '08). ACM, New York, NY, USA, 183-194. DOI: <https://doi.org/10.1145/1341531.1341557>

What next?

- Determination and propagation of trust
- More work needed!

Examples of significant features for trust

Answer length;

The number of words in the answer with a corpus frequency larger than c ;

The number of “thumbs up” minus “thumbs down” received by the answerer divided by the total number of “thumbs” s/he has received;

The entropy of the trigram character-level model of the answer;

The fraction of answers of the answerer that have been picked as best answers (either by the askers of such questions, or by a community voting);

The unique number of words in the answer;

Average number of abuse reports received by the answerer over all his/her questions and answers;

Average number of abuse reports received by the answerer over his/her answers;

The non-stopword word overlap between the question and the answer;

The Kincaid score of the answer;

The average number of answers received by the questions asked by the asker of this answer ;

Challenges of Social Media research

- Networks are a public good but in private hands
- Definitions are poor
- Fashions and fads are commonplace so usage patterns vary
- There is a constant moral panic about social media so dispassionate observers are rare

“Freedom of speech is not freedom of reach”

Next lecture:

“The digital university and other mythical creatures”

Tuesday 11th Feb 2020 at 18:00 (6pm) London
Time

www.gresham.ac.uk