Good Friends, Bad News

How emotions influence virality in Twitter

Lars Kai Hansen
www.imm.dtu.dk/~lkh
lkh@imm.dtu.dk

DTU Informatics
Department of Informatics and Mathematical Modeling

cognitive systems

7 faculty, 6 postdocs, 17 graduate students
Total grant volume 2006-2011 approx 7 mill USD

machine mearning
- adaptive systems, automated text analysis, pattern classification,

signal processing
- detection of weak signals
- blind signal separation

cognitive science
- psycho-physics e.g. audio-visual integration
- computational cognitive modelling

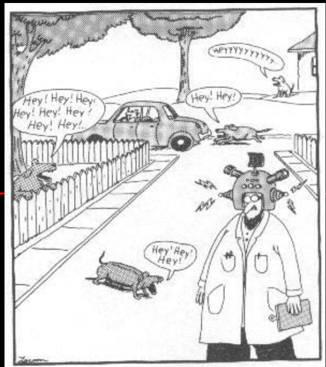
systems neuroscience
- mental state decoding
- meta-analysis/neuroinformatics

datamining
- search engines, meta-analysis
- wikipedia research

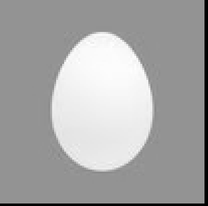
mobile services
- Communication, mobile web apps, Mobile phone as sensor system

digital economy
- web 2.0, web 3.0
- social networks, web 2.0, user generated content
- digital media semantics
- latent semantics, automated text analysis, meta-data analysis

human computer interfaces
- Small interfaces



Donning his new canine decoder, Professor Schwartzman becomes the first human being on Earth to hear what barking dogs are actually saying.



Outline

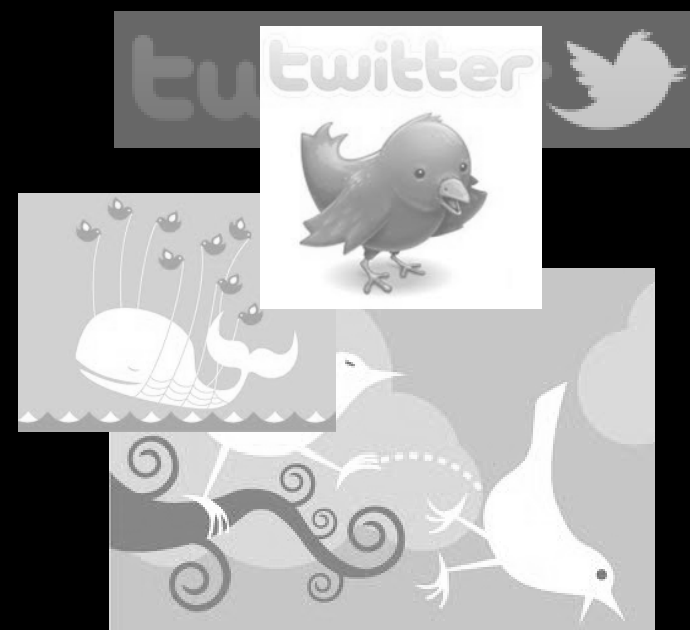
Social media outlets are instrumental for the new, data driven science of human behavior

- Digital media modeling: Mind reading
- Quantifying subjectivity: Affective computing
- Sentiment analysis -in text
- Sentiment analysis mind reading

The sentiment/virality paradox

- Social interaction is positive
- Twitter as a laboratory
- Analyzing virality in Twitter
- Paradox resolved

Conclusions

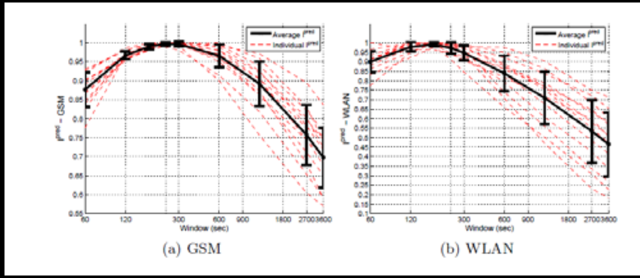


Human predictability

C. Song, Z. Qu, N. Blumm, A.-L. Barabási
Limits of Predictability in Human Mobility
Science 327, 1018-1021 (2010).

<u>A new picture of human predictability is emerging</u>

- **Barabasi group’s study of mobility (70-93% /hour)**
- **Individuals are predictable (intersubject variability is huge … long tail … “power law”)**
- **DTU mobile phone context tracking (80-90% /4 min)**

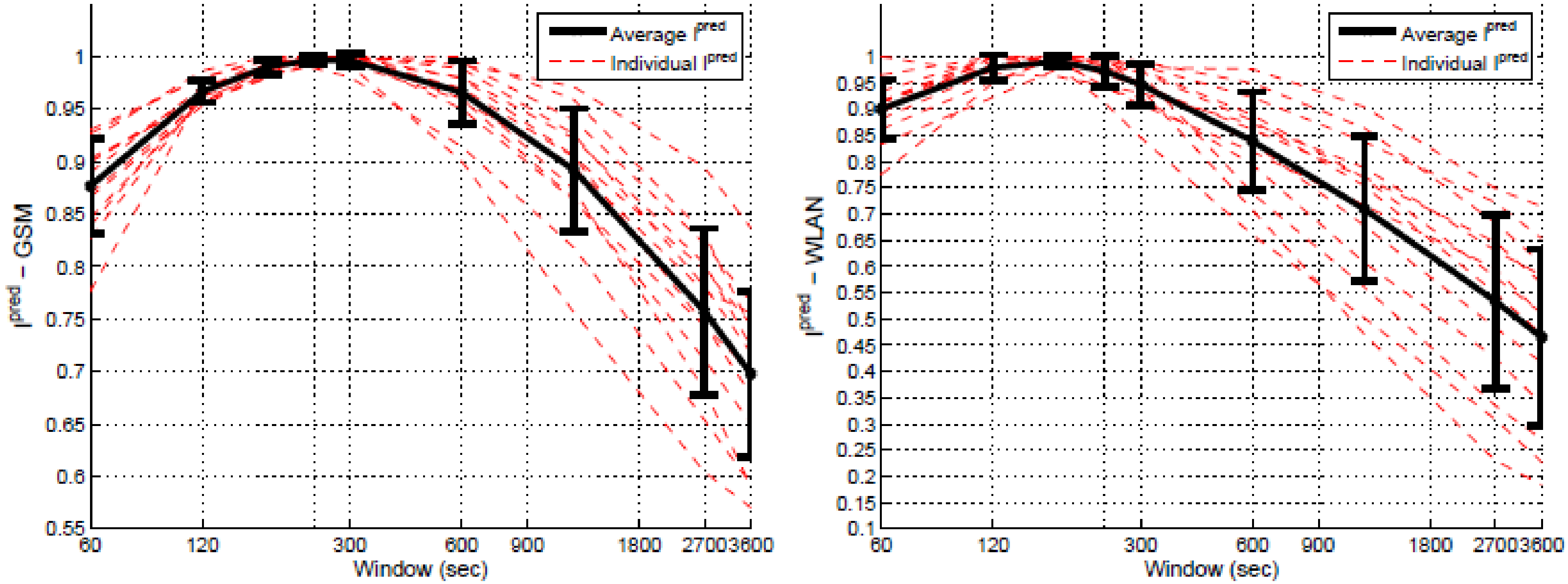


(a) GSM (b) WLAN

<u>Why is this important?</u>

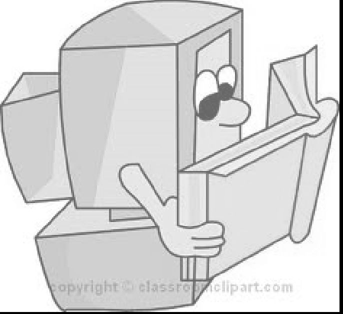
- **Curiosity: Understanding the human brain**
- **Engineering: New services/products are based on human predictability (vz. the Google paradigm)**

Human predictability



(a) GSM (b) WLAN

B.S. Jensen, J.E. Larsen, K. Jensen, J. Larsen, L.K. Hansen: Estimating Human Predictability From Mobile Sensor Data In Proc. IEEE International Workshop on Machine Learning for Signal Processing MLSP (2010).
B.S. Jensen, J.E. Larsen, K. Jensen, J. Larsen, L.K. Hansen: Proc. 21st European Conference on Machine Learning, Mining Ubiquitous and Social Environments Workshop. Barcelona, Spain (2010).



Digital media modeling: Mind reading

…is largely based on text and tagging

Main approaches to text analysis in computers

- **<u>Computational linguistics (NLP)</u>**
  - **Rule based, tight statistical models**
  - **Precise / high specificity**
  - **Challenged by informal/innovative text**

- **<u>Statistical learning</u>**
  - **Flexible / unsupervised**
  - **High sensitivity –informal OK**
  - **Challenged in the quantitative/opionion/precision**

Trick: Vector space representation

- Abstract representation - can be used for all digital media
- Document is represented as a point in a high-dimensional ”feature space” – document similarity ~ spatial proximity

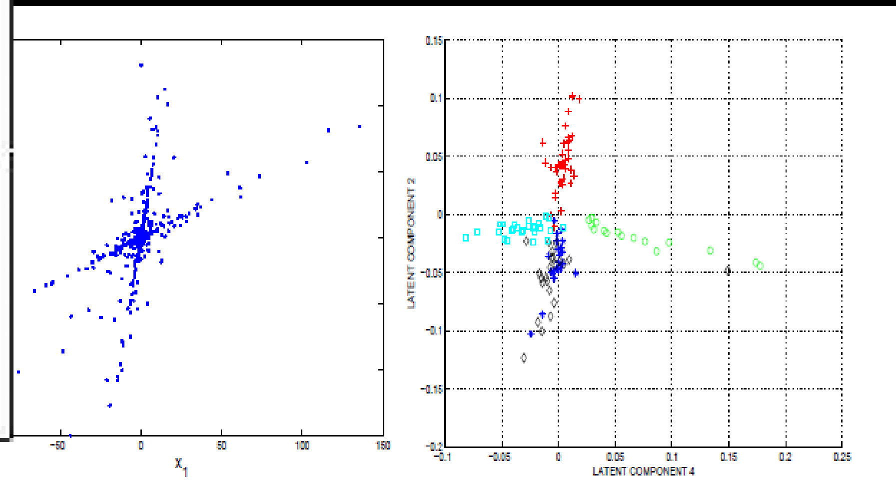
- General features or “events”
- Social network: Social event involving a set of nodes
- Text: Bag of words (Term/keyword histograms),
- Image: Color histogram, texture measures, ”bag of features”
- Video: Object coordinates (tracking), active appearance models
- Sound: Spectrograms, cepstral coefficients, gamma tone filters

Document features are correlated, the pattern of correlation reflects ”associations”. Associations are context specific

Contexts can be identified unsupervised fashion by their feature associations = Latent semantics

“Bag of Words”

Terms	Documents								
	c1	c2	c3	c4	c5	m1	m2	m3	m4
computer	1	1	0	0	0	0	0	0	0
EPS	0	0	1	1	0	0	0	0	0
human	1	0	0	1	0	0	0	0	0
interface	1	0	1	0	0	0	0	0	0
response	0	1	0	0	1	0	0	0	0
system	0	1	1	2	0	0	0	0	0
time	0	1	0	0	1	0	0	0	0
user	0	1	1	0	1	0	0	0	0
graph	0	0	0	0	0	0	1	1	1
minors	0	0	0	0	0	0	0	1	1
survey	0	1	0	0	0	0	0	0	1
trees	0	0	0	0	0	1	1	1	0

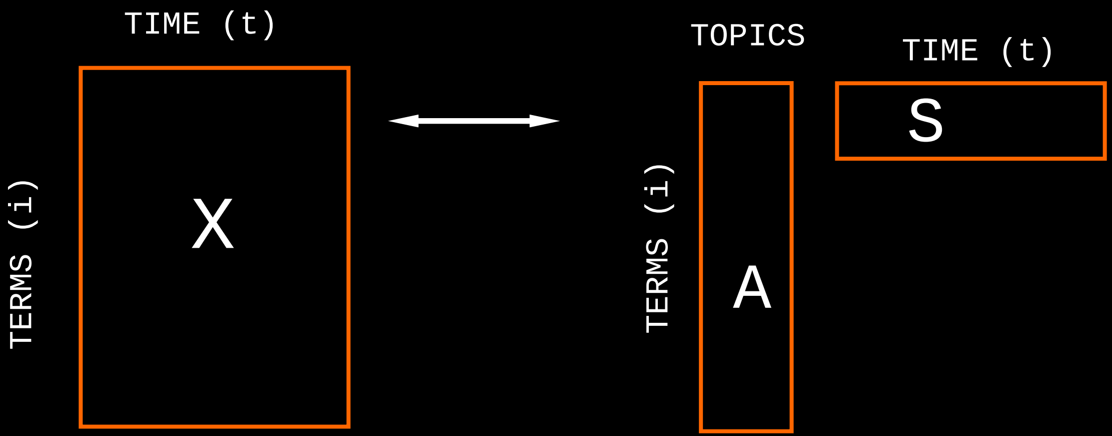


Very efficient for detection of context, classification and topic modeling.
Often leads to very high-dimensional learning problems

G. Salton , A. Wong , C. S. Yang, A vector space model for automatic indexing, Communications of the ACM, v.18 n.11, p.613-620, Nov. 1975

Factor models

- Represent a datamatrix by a low-dimensional approximation, eg. stream of text

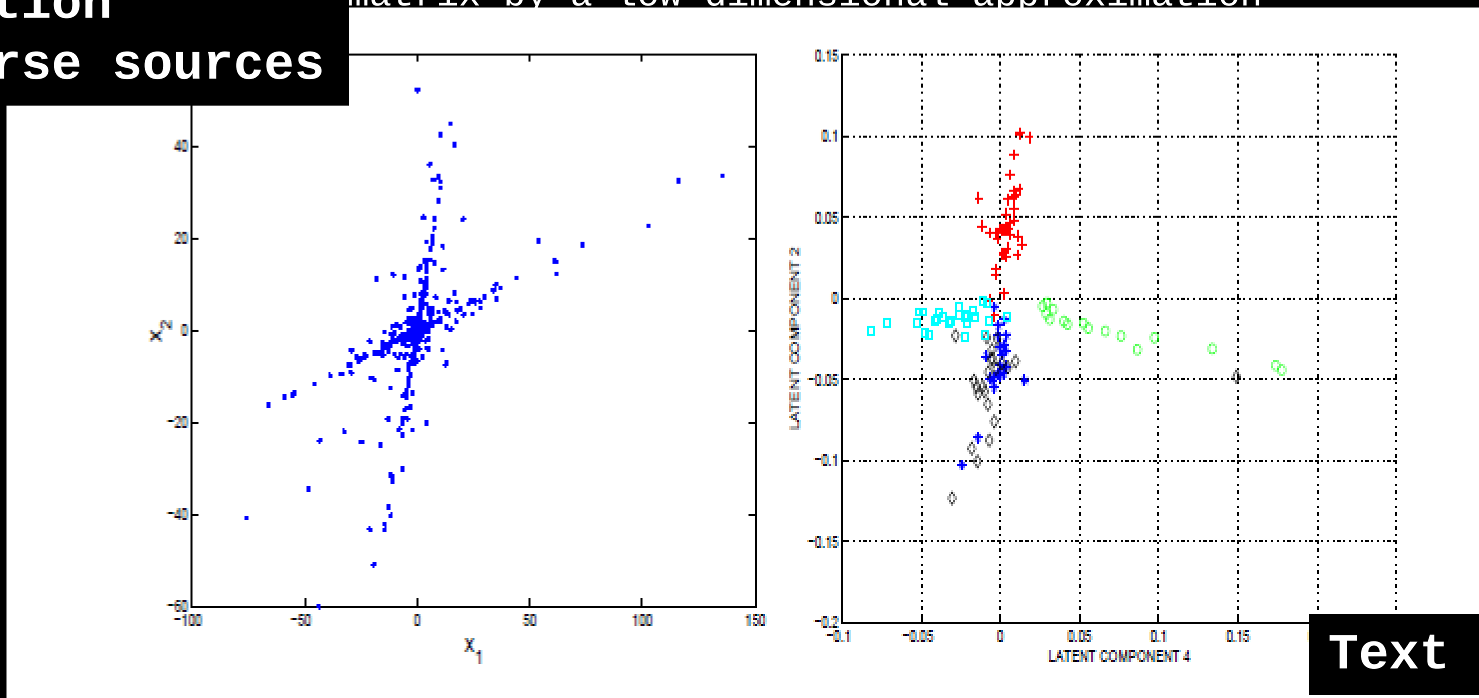


$$X(i,t) \approx \sum_{k=1}^{K} A(i,k)S(k,t)$$

Factor models

Simulation w/ sparse sources

matrix by a low-dimensional approximation



Text data

$$X(i,t) \approx \sum_{k=1}^{K} A(i,k)S(k,t)$$

CASTSEARCH - CONTEXT BASED SPEECH DOCUMENT RETRIEVAL

Lasse Lohilahti Mølgaard, Kasper Winther Jørgensen, and Lars Kai Hansen

Informatics and Mathematical Modelling
Technical University of Denmark Richard Petersens Plads
Building 321, DK-2800 Kongens Lyngby, Denmark

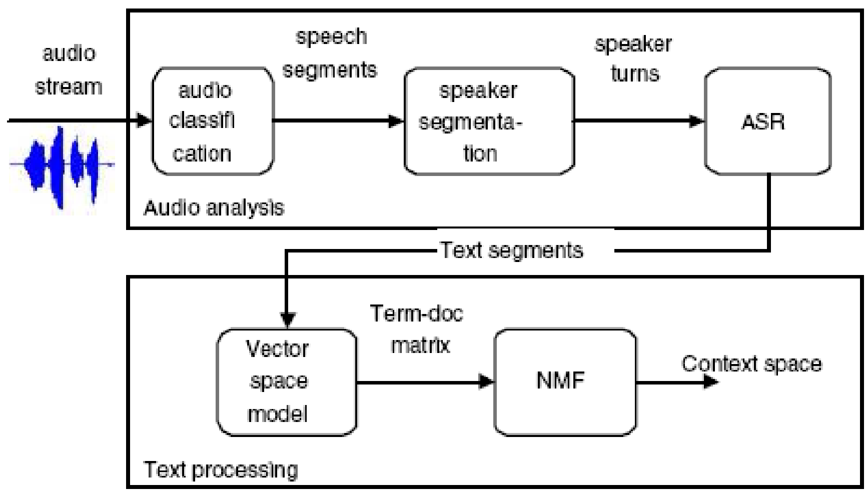
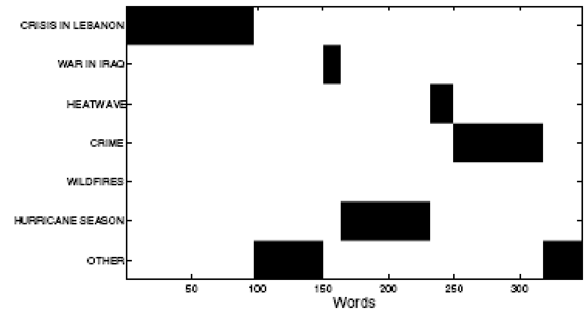
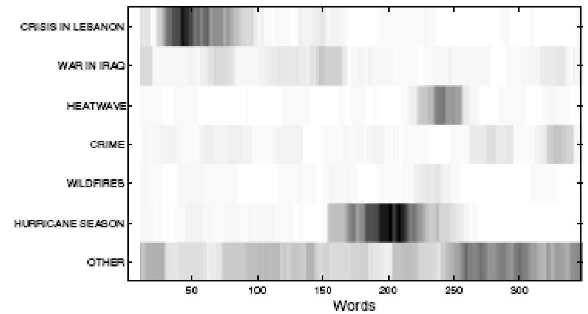


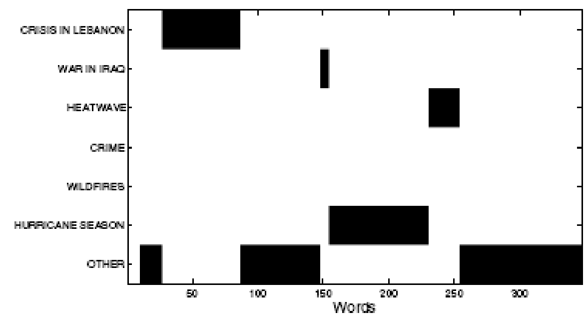
Fig. 1. The system setup. The audio stream is first processed using audio segmentation. Segments are then using an automatic speech recognition (ASR) system to produce text segments. The text is then processed using a vector representation of text and apply non-negative matrix factorization (NMF) to find a topic space.



(a) Manual segmentation.

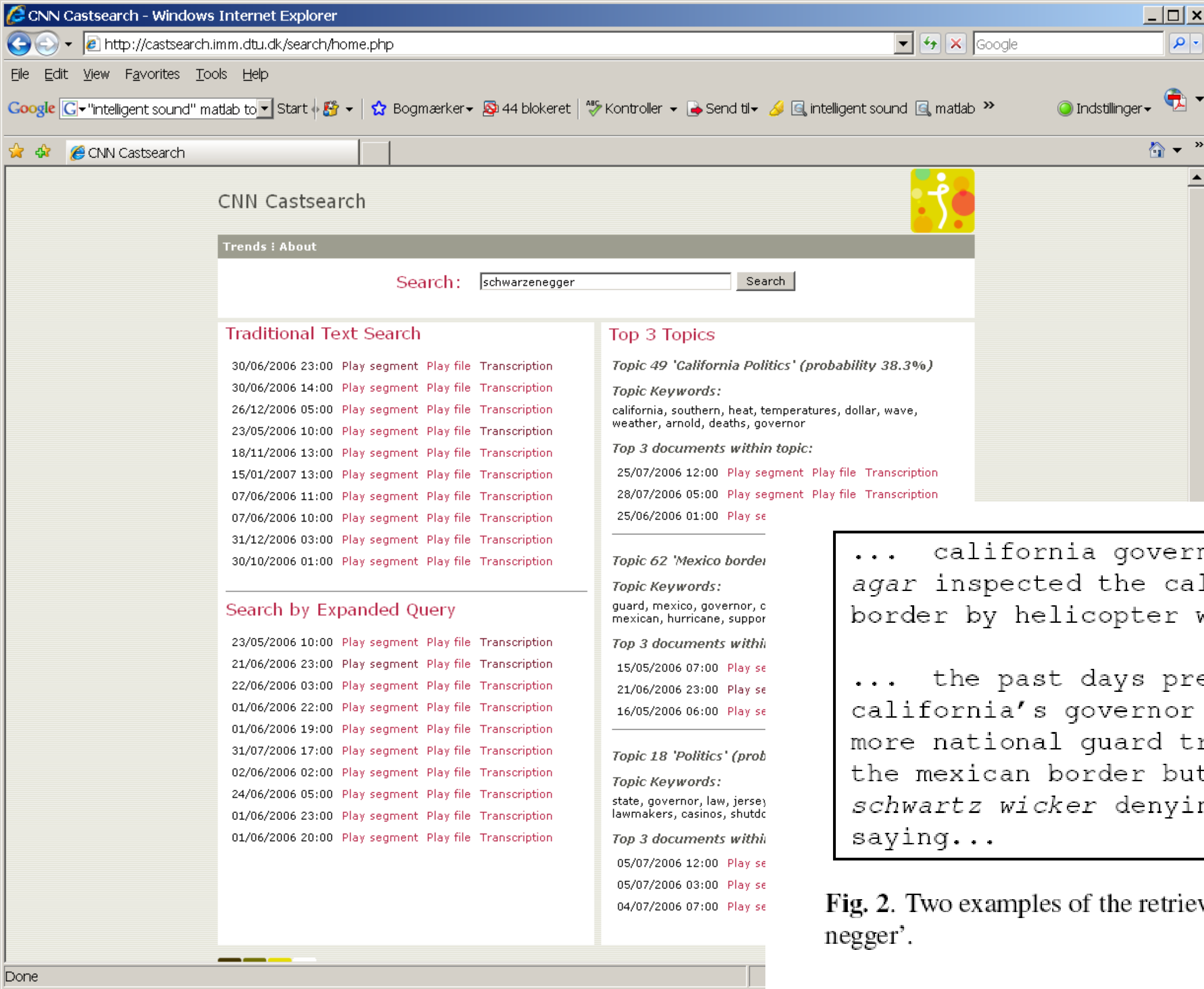


(b) $p(k|d^*)$ for each context. Black means high probability.



(c) The segmentation based on $p(k|d^*)$.

Fig. 3. Figure 3(a) shows the manual segmentation of the news show into 7 classes. Figure 3(b) shows the distribution $p(k|d^*)$ used to do the actual segmentation shown in figure 3(c). The NMF-segmentation is in general consistent with the manual segmentation. Though, the segment that is manually segmented as 'crime' is labeled 'other' by the NMF-segmentation



... california governor arnold's *fortson agar* inspected the california mexico border by helicopter wednesday to see ...

... the past days president bush asking california's governor for fifteen hundred more national guard troops to help patrol the mexican border but governor orville *schwartz wicker* denying the request saying...

Fig. 2. Two examples of the retrieved text for a query on 'schwarzenegger'.

castsearch.imm.dtu.dk

K.W. Jørgensen, L.L. Mølgaard, L.K. Hansen, Unsupervised Speaker Change Detection for Broadcast News Segmentation, Eusipco, 2006
L.L. Mølgaard, K.W. Jørgensen, L.K. Hansen, Castsearch - Context Based Spoken Document Retrieval, ICASSP, 2007

MiRocket

Time frequency analysis pipeline:
MFCC’s @ 30 ms windows

Temporal integration and genre classification at 1000ms

Music recommendation in 12-D genre space

A. Meng, P. Ahrendt, J. Larsen, L.K. Hansen: *Temporal Feature Integration for Music Genre Classification. IEEE* Transactions on Audio and Speech and Language Processing 15(5): 1654-1664 (2007)

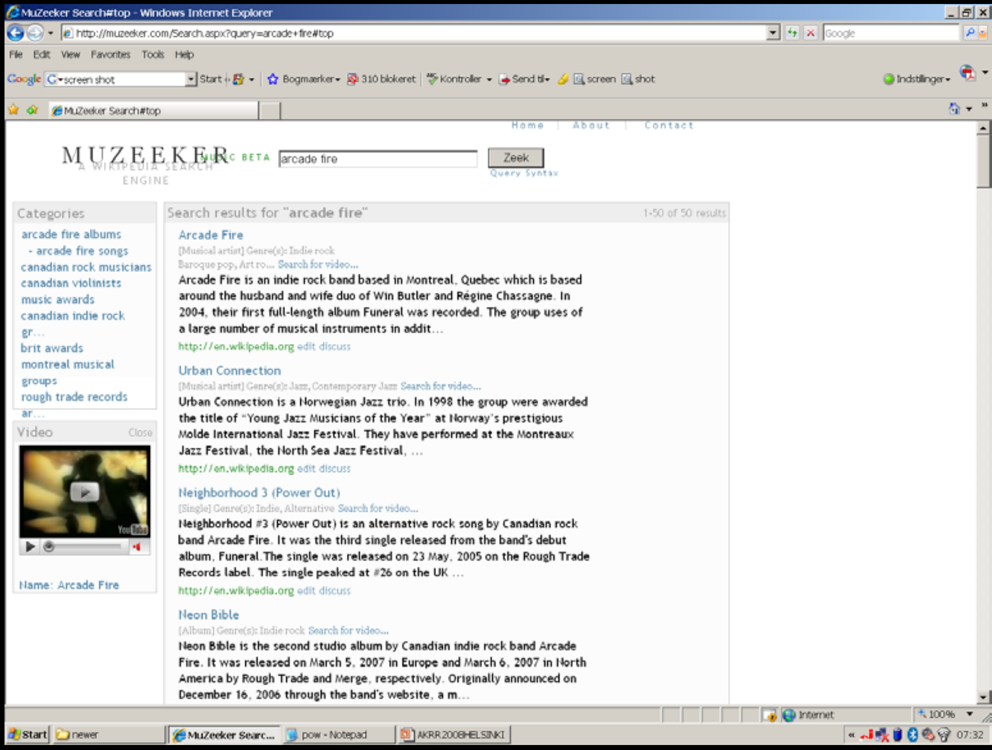
T. Lehn-Schiøler, J. Arenas-García, K.B. Petersen and L.K. Hansen: *A Genre Classification Plug-in for Data Collection. Proc. 7th Intl. Conf. on Music Information Retrieval, ISMIR 2006, pp. 320-321, Victoria, Canada, Oct.* (2006).

L.K. Hansen, T. Lehn-Schiøler, K.B. Petersen, J. Arenas-Garcia, J. Larsen, and S.H. Jensen: *Learning and clean-up in a large music database. EUSIPCO 2007, European Conference on Signal Processing, Poznan (2007).*

muzeeker

- Wikipedia based common sense
- Wikipedia used as a proxy for the music users mental model
- Implementation: Filter retrieval using Wikipedia’s article/ categories

muzeeker.com



S. Halling, M.K. Sigurdsson, J.E. Larsen, S. Knudsen, L.K. Hansen: *MuZeeker: A domain SpecificWikipedia-based Search Engine. In Proc. First International Workshop on Mobile Multimedia Processing. Tampa, USA (2008).*

J.E. Larsen, S. Halling, M. Sigurdsson and L.K. Hansen: *MuZeeker - Adapting a music search engine for mobile phones. To appear in Springer Lecture Notes in Computer Science ‘Mobile Multimedia Processing: Fundamentals,* Methods, and Applications’, Selected papers from First International Workshop on Mobile Multimedia Processing, Tampa, USA. (2010).

Outlook – the future of mind reading

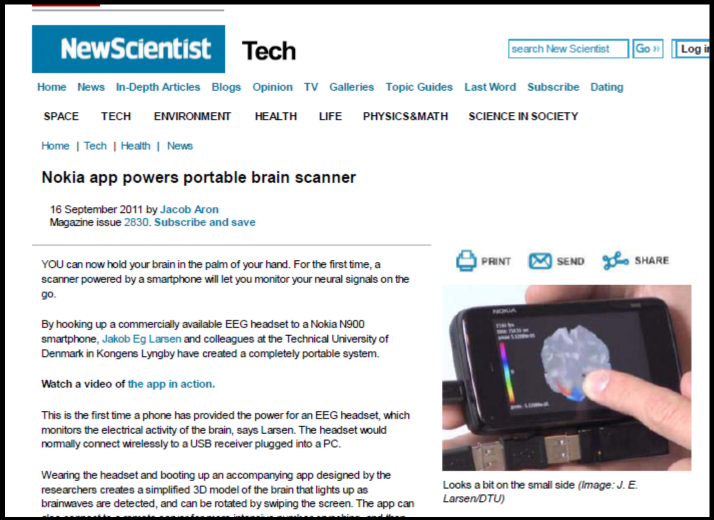
Ecological brain imaging

Continuous mental state
decoding in the wild

24/7 monitoring

EEG real time 3D imaging for bio-feedback

Fig. 1. Handheld brain scanner components. Emotiv EPOC wireless EEG headset (1), Emotiv Receiver module with USB connector (2), USB connector and adapter (3+4), and Nokia N900 mobile phone. The total cost of the system is less than USD1000.



NewScientist Tech

search New Scientist Go »

Home News In-Depth Articles Blogs Opinion TV Galleries Topic Guides Last Word Subscribe Dating

SPACE TECH ENVIRONMENT HEALTH LIFE PHYSICS&MATH SCIENCE IN SOCIETY

Home | Tech | Health | News

Nokia app powers portable brain scanner

16 September 2011 by Jacob Aron
Magazine issue 2830. Subscribe and save

PRINT SEND SHARE

YOU can now hold your brain in the palm of your hand. For the first time, a scanner powered by a smartphone will let you monitor your neural signals on the go.

By hooking up a commercially available EEG headset to a Nokia N900 smartphone, Jakob Eg Larsen and colleagues at the Technical University of Denmark in Kongens Lyngby have created a completely portable system.

Watch a video of the app in action.

This is the first time a phone has provided the power for an EEG headset, which monitors the electrical activity of the brain, says Larsen. The headset would normally connect wirelessly to a USB receiver plugged into a PC.

Wearing the headset and booting up an accompanying app designed by the researchers creates a simplified 3D model of the brain that lights up as brainwaves are detected, and can be rotated by swiping the screen. The app can

Looks a bit on the small side *(Image: J. E. Larsen/DTU)*

hyposafe®

Good Friends, Bad News - Affect and Virality in Twitter

Lars Kai Hansen, Adam Arvidsson, Finn Aarup Nielsen,
Elanor Colleoni, and Michael Etter

DTU Informatics, Technical University of Denmark,
DK-2800 Lyngby, Denmark, Email: lkh@imm.dtu.dk fn@imm.dtu.dk
University of Milan, via Conservatorio Milan, Italy,
Email: adam.arvidsson@unimi.it
Copenhagen Business School, DK-2000 Frederiksberg, Denmark
Email: elc.ikl@cbs.dk me.ikl@cbs.dk

L.K. Hansen, A. Arvidsson, F.A. Nielsen, E. Colleoni, M. Etter. *Good Friends, Bad News - Affect and Virality in Twitter. In Proc. International Workshop on Social Computing, Network, and Services (SocialComNet 2011).* Springer Communications in Computer and Information Science 185: 34-43 (2011))

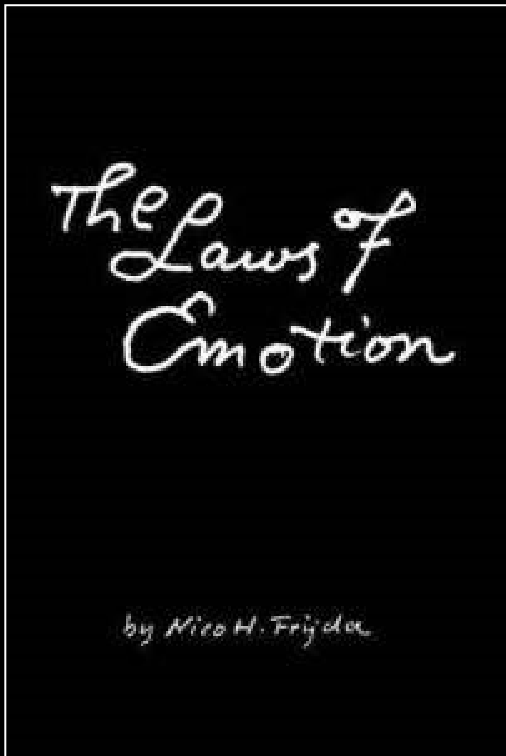
Quantifying subjectivity
“Affective computing”

Affective computing is research in systems that can recognize, interpret, process, and simulate human emotion

Emotions are omnipresent and extremely important to communication / opinion formation / intent reading etc

Psychology of emotion is well developed but still far from complete…

7. The law of hedonic asymmetry: *“Negative emotions last longer, positive emotions fade”*

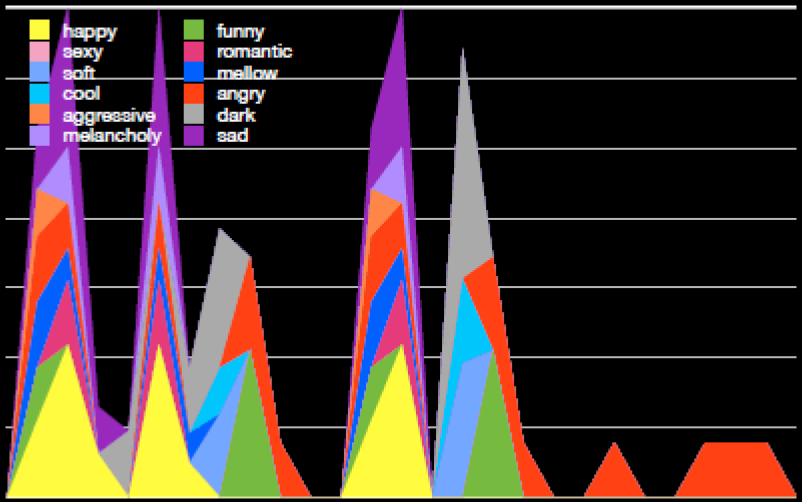


Sentiment detection

A step towards understanding subjectivity, opinion

Important to many services
- Recommender (Amazon reviews)
- Information navigation, e.g. navigating music

Oasis ”Wonderwall”
Emotional content in the song lyrics through time



M.K. Petersen: Modeling media as latent semantics based on cognitive components (Ph.D. Thesis, DTU,2010)

Text sentiment detection methods

<u>WordNet</u>

Use the linguistic structure (network of word relations) to compute the distance to “good” or “bad”

<u>Supervised learning “seing the stars”</u>

Learn a predictive model based on labeled data, e.g. product reviews (text + # stars)

<u>Curated word list</u>

Design a list of domain relevant keywords with emotional annotation, e.g. the generic list ANEW

Defining virality

Virality lacks a formal definition,

e.g., Wiktionary says: *(advertising, marketing) “The state or condition of being viral; tendency to spread by word of mouth.”*

We define virality statistically as the probability that a message is passed on in the network.

In Twitter this means ***retweeted,*** in other media different. Eg. in news papers one can *email* articles to friends (NYT)

Analysis of retweeting (Suh et al, 2010)

TABLE I. TWEET VARIABLES

URL	# of URLs in a tweet
Hashtag	# of hashtags in a tweet
Mention	# of usernames specified in a tweet excluding ones used for making a retweet (e.g. via @username, RT: @username)
Follower	# of users who follows the author of a tweet
Followee (Friend)	# of friends that the author is following
Day	# of days since the author of a tweet created the Twitter account
Status	# of tweets made by the user since the creation of the account
Favorite	# of favorited tweets by a user
Retweet	# of retweets recorded for a given tweet

TABLE IV. GENEARALIZED LINEAR MODEL

	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**
(Intercept)	-4.42000	0.146400	-30.19	**0.0000***
Days	0.00122	0.000296	4.12	**0.0000***
HashtagOrNot	1.32800	0.160300	8.28	**0.0000***
MentionOrNot	-0.29490	0.166800	-1.77	0.0771
URLOrNot	0.76360	0.150900	5.06	**0.0000***
Followee	0.00006	0.000020	2.85	**0.0043***
Follower	0.00002	0.000005	3.82	**0.0001***
Status	-0.00002	0.000009	-1.71	0.0876
Favorite	-0.00004	0.000163	-0.26	0.7987

B. Suh, L. Hong, P. Pirolli, E.H. Chi: *Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network.* In Proc IEEE SocialCom, 2010

The Sentiment / Virality paradox

Seemingly conflicting observations on virality

i) Bad news is good news. Negative sentiment propagates in news media

(Galtung & Ruge, 1965)

ii) Most social networks including Twitter have a dominant positive sentiment

iii) emailing from NYT is dominated by positive sentiment articles (Berger & Milkman, 2010)

We will go to Twitter to find out!

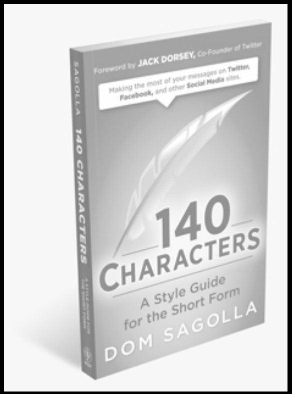
Twitter - a digital behavior lab

“Microblog”: Conceived as SMS/Texting Short Text Messaging (140 characters) for the Internet

Jack Dorsey, Biz Stone and a group of engineers defined Twitter in March 2006, and launched it in July 2006.

300 mill. users, 300 mill. messages pr day

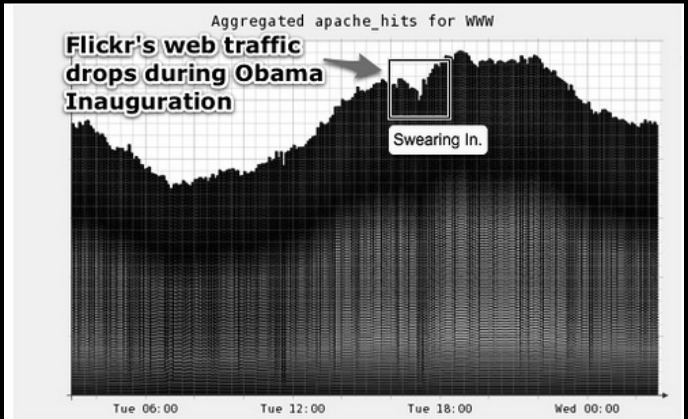
1.6 billion search queries per day



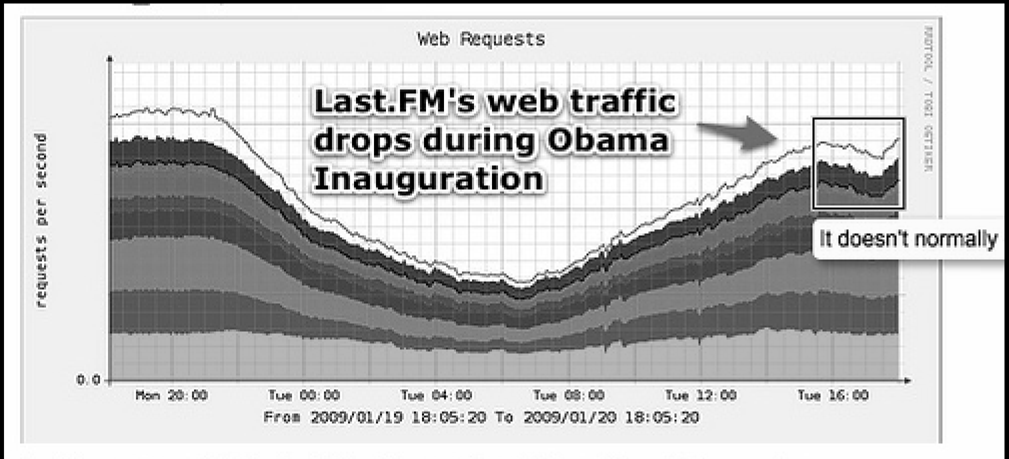
Dom Sagolla: 140 Characters (Wiley, 2009)

Web of time....the blog pulse

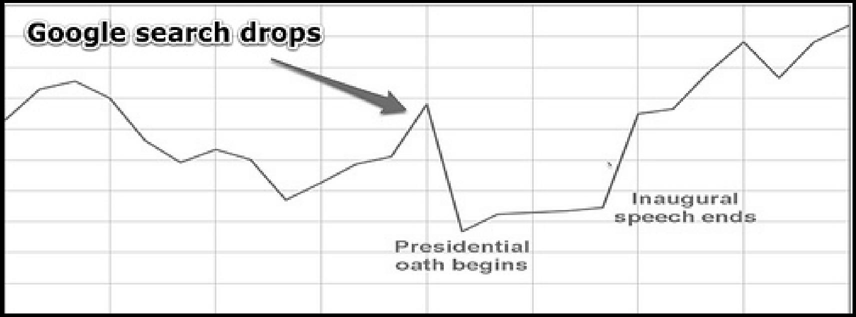
Jesse Robbins O'reilly's radar Feb 08, 2009

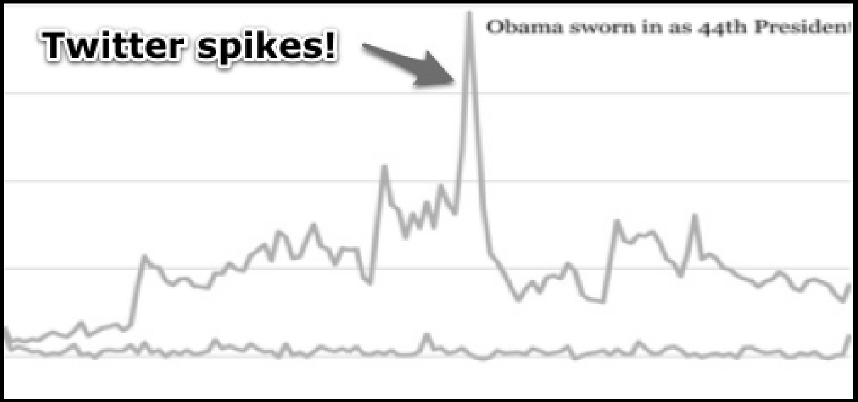


During the actual swearing in, we see a little dip while people watch the live inauguration coverage.



Numbers removed obviously, this is dip we suffered during Obama's Inauguration.





Twitter is a news medium

Obama inauguration speech,

Egypt upraise tag #jan25

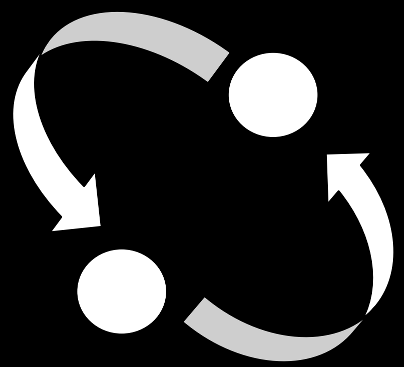
Tsunami, Japan phone system collapsed, but net still worked and Tweeting soared

http://mashable.com/2011/03/11/japan-tsunami/

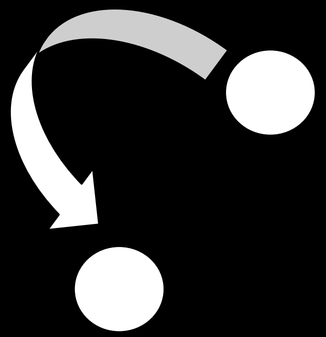
Twitter is a social medium

Twitter conceived as Internet SMS/texting – keeping friends updated

Symmetry of connections



Friendship is symmetric



News / interest graph may be asymmetric

So is Twitter a news medium or a social network?

Kwak et al. (2010)

"…Twitter shows a low level of reciprocity; 77.9% of user pairs with any link between them are connected one-way, and only 22.1% have reciprocal relationship between them.

"…Previous studies have reported much higher reciprocity on other social networking services: 68% on Flickr [4] and 84% on Yahoo! 360.

"…Moreover, 67.6% of users are not followed by any of their followings in Twitter. We conjecture that for these users Twitter is rather a source of information than a social networking"

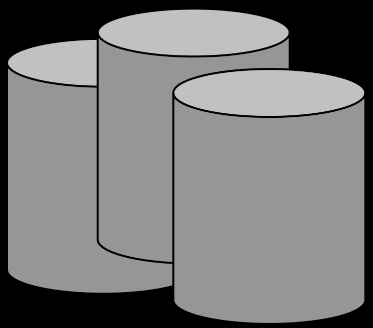
H Kwak, C Lee, H Parkand, S Moon. What is Twitter, a social network or a news media? In WWW '10: Proceedings of the 19th international conference on World wide web, p. 591–600, New York, NY, USA, 2010. ACM

Understanding virality and sentiment in Twitter

Our research questions

- Q1: How accurately can text be characterized as ‘news’?
- Q2: How big a fraction of Twitter is news?
- Q3: If Twitter is a news medium, does negative sentiment influence virality?
- Q4: Does sentiment influence retweet probability diffentially in news and social messages?

Databases for this study



Brown corpus

a general text corpus with a known mixture of news/nonnews documents. The corpus consist of 47134 sentences.

RANDOM Twitter sample

348862 **tweets collected September 9-14, 2010. The Tweets were randomly sampled following the ’Spritzer’ protocol.**

COP15 Twitter sample

complete set of tweets for a specific news event COP15 2009 UN Climate Change Conference, Denmark Dec 7-18. ***207782*** **tweets downloaded December 1-31 with the term/tag “cop15”.**

News detection (Q1: Browne corpus)

$$p(\text{news}|\boldsymbol{w}) = \frac{p(\text{news})p(\boldsymbol{w}|\text{news})}{p(\text{news})p(\boldsymbol{w}|\text{news}) + p(\neg\text{news})p(\boldsymbol{w}|\neg\text{news})}$$
$$= \left(1 + \frac{p(\neg\text{news})\prod_{d=1}^{D} p(w_d|\neg\text{news})}{p(\text{news})\prod_{d=1}^{D} p(w_d|\text{news})}\right)^{-1}$$

We posit a simple model (Naive Bayes) to estimate the probability of carrying the news label, given the bag of words features (D=10000). The accuracy on test data is 84%

News detection (Q2: Twitter news?)

We apply the trained NB classifier to the two Twitter samples

A tweet is declared news if p(news | w) > 0.5

RANDOM data

Rate of news: 22.3%

COP15 data

Rate of news: 30.3%

Sentiment detection in Twitter (Q3+Q4 How does sentiment affect retweeting?)

Now we are interested in estimating the probability of retweet and to infer the importance of the features.

We use the generalized linear model as in Suh et al., it provides standard scores for deletion of individual features (saliency)

$$p(RT|\boldsymbol{f}) = \left(1 + e^{-\sum_{i=0}^{F} \beta_i f_i}\right)^{-1}$$

Sentiment detection in Twitter (Q3+Q4 How does sentiment affect retweeting?)

GLM Features: Negative sentiment (*news), mention, url, hash-tag

RANDOM data set

- There is no significant dependency on presence of negative sentiment.
- Negative AND news, is retweeted more.
- Confining to tweets that have a non-zero sentiment retweeting strongly suppressed by negative sentiment

COP15 data set (a news event)

- For all tweets in this sample there significant effect that negative sentiment promotes retweeting. Also seen in the sentimental tweets.

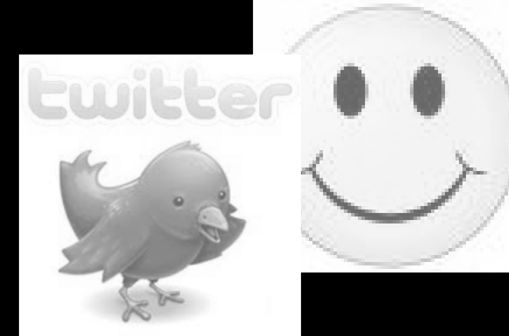
Sentiment detection in Twitter (Q3+Q4 How does sentiment affect retweeting?)

GLM

RAN

	COP15		RANDOM	
	All	**English**	**All**	**English**
N	147,041	136,262	335,236	106,719
Rate News	0.303	0.305	0.226	0.233
t(Negative)	4.8[illegible]	**4.649**	2.775	-0.024
t(Negative_newsness)	2.275	**1.471**	-6.0[illegible]	**3.904**

Conclusions



We may train computational models to partially understand social media behaviors

In general propagation in social media is enhanced by positive content.

Twitter is both a social and a news medium

Virality can be defined through the retweet probability (probability of message being passed on).

Likelihood of retweet is increased by negative content in newsy posts, in line with general news media theory

Contact

Prof. Lars Kai Hansen

DTU Informatics
Technical University of Denmark
Richard Petersens Plads B321
DK-2800 Kgs. Lyngby

lkh@imm.dtu.dk
www.imm.dtu.dk/~lkh

Tel.: +45 45253889