ANIMAL SPIRITS

ECB & IIF Workshop on using big data for forecasting and statistics, Frankfurt am Main, April 2014

DISCUSSANT: JOHAN BOLLEN

PAPERS:

1) NEWS AND NARRATIVES IN FINANCIAL SYSTEMS: EXPLOITING BIG DATA FOR SYSTEMIC RISK ASSESSMENT BY RICKARD NYMAN ET AL.

2) BELGIAN ECONOMIC POLICY UNCERTAINTY INDEX: IMPROVEMENT THROUGH TEXT MINING BY ELLEN TOBBACK ET AL.

ANIMAL SPIRITS

FEAR AND GREED

ECONOMIC AND MARKET INDICES OF ANIMAL SPIRITS

Surveys:

- **Consumer confidence indices**

- **Well-being indicators**

- **Investor sentiment, cf. Huina Mao's in the next session**

Econometric and financial indicators:

- **Volatility or "fear" index**

- **Misery index (inflation and unemployment)**

- **The markets themselves (cf. Prechter)**

A Global Projection of Subjective Well-being: The First Published Map of World Happiness
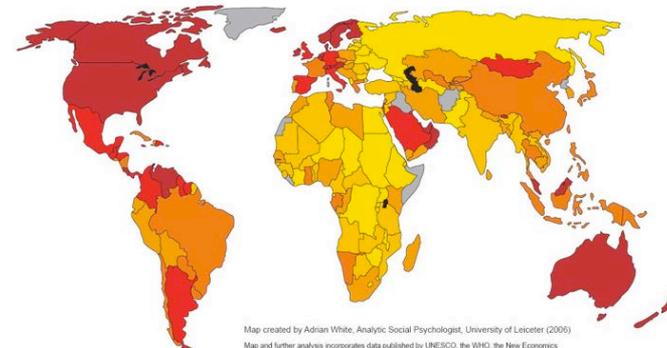
Map created by Adrian White, Analytic Social Psychologist, University of Leicester (2006)
Map and further analysis incorporates data published by UNESCO, the WHO, the New Economics Foundation, the Veenhoven Database, the Latinbarometer, the Afrobarometer, the CIA, and the UN Human Development Report.

Happy — Average — Unhappy

Cartographic Unit • University of Leicester
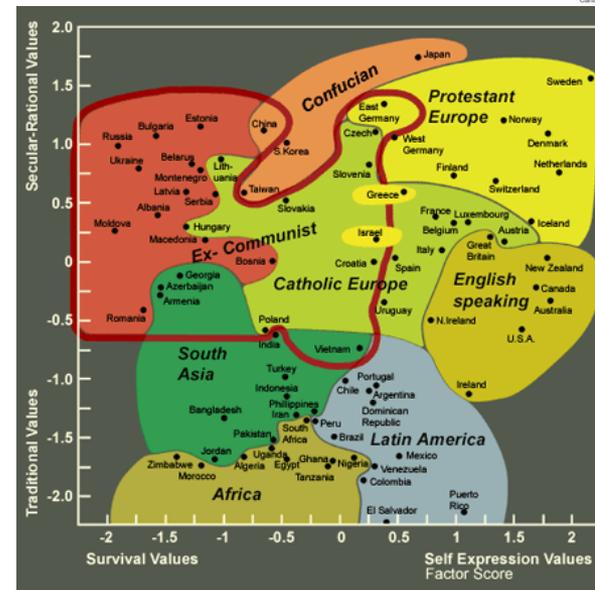
SOCIAL MEDIA
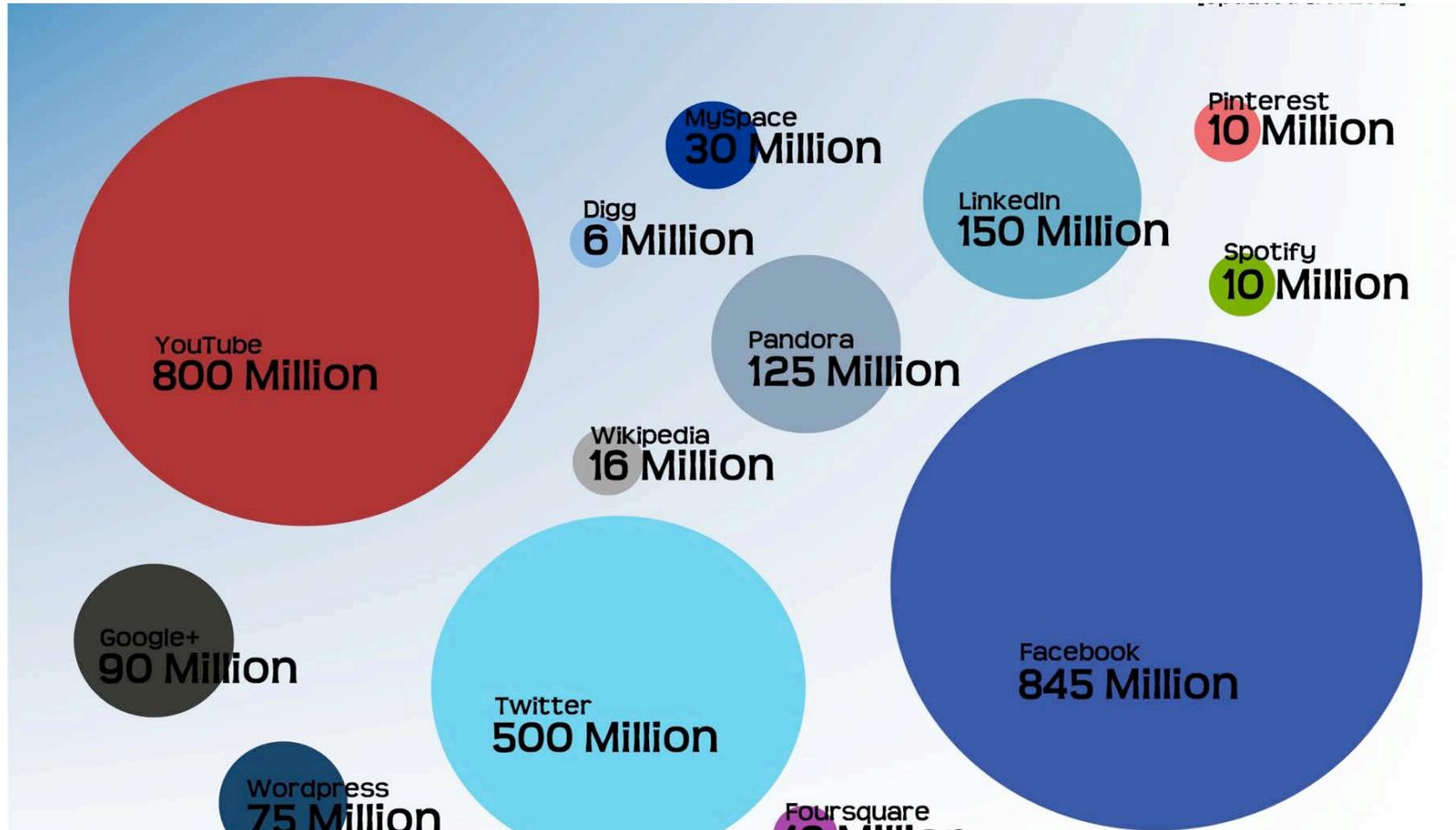
BIG DATA AND SOCIAL MEDIA

SOCIAL MEDIA DATA ANALYTICS
ADOPTION BY INDUSTRY (THE "S-CURVE")
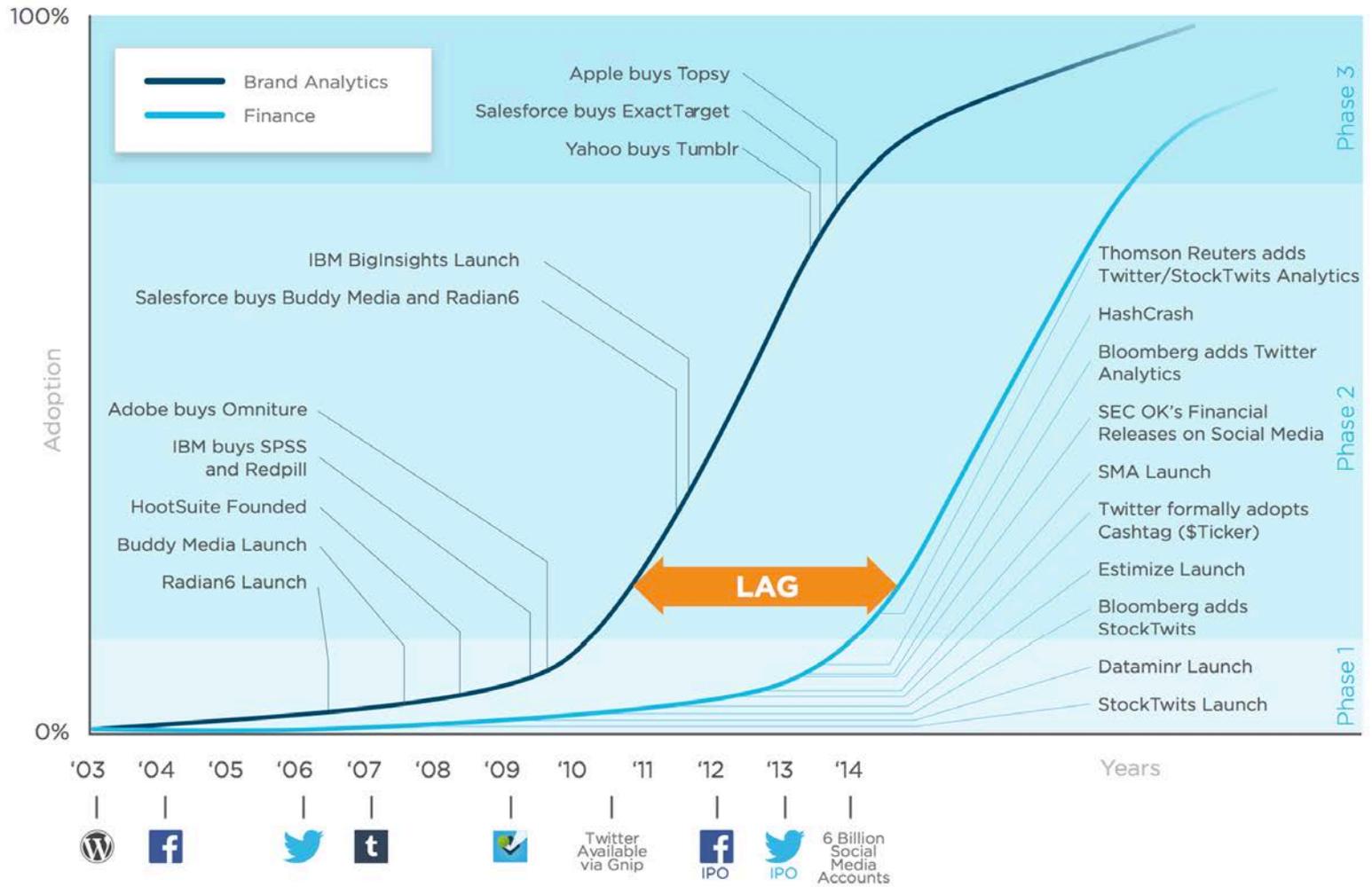
TEXT ANALYTICS

"semantic analysis of financial text (Tetlock 2007; Tetlock, Saar-Tsechansky, and Macskassy 2008; Loughran and McDonald 2011; Devitt and Ahmad 2007; Zhang and Skiena 2010)…analyzing traditional financial documents (e.g. news or corporate reports), … financial sentiment from blogs (Gilbert and Karahalios 2010), investor message boards (Antweiler and Frank 2004), Twitter (Bollen, Mao, and Zeng 2011), as well as search engine (Da, Engelberand, and Gao 2011) " -- from Huina Mao et al (2014)

Sentiment analysis:

Natural Language processing (n-grams) for reviews (Nasukawa, 2003), topics (Yi, 2003), Support Vector Machines: text classification (positive vs. negative) using pre-classified learning sets: Gamon (2004), Pang (2008), Blogs, web sites: mixed approaches. Mishne (2006), Balog (2006), Gruhl (2005), LWIC Lexicons: Dodds/ANEW, Stanford CoreNLP Sentiment

NOWCASTING AND FORECASTING

- **Box office receipts from Twitter chatter: Asur (2010)**

- **Google trends: flu prediction**

- **Predicting consumer behavior from search query volume (Goel, 2010)**

- **Predicting elections: Tumasjan (2010), O'Connor (2010), DiGrazia (2013)**

- **Contagion of "Loneliness" and happiness in social networks (Cacioppo, 2010 - Bollen, 2011)**

- **Market prediction:** Tobias Preis (2013), Mao (2011, 2012), Yigitcan Karabulut (2011, 2014), Gilbert (2011), Sul, Dennis and Yuan (2014)

GLOBAL MOOD TIME SERIES
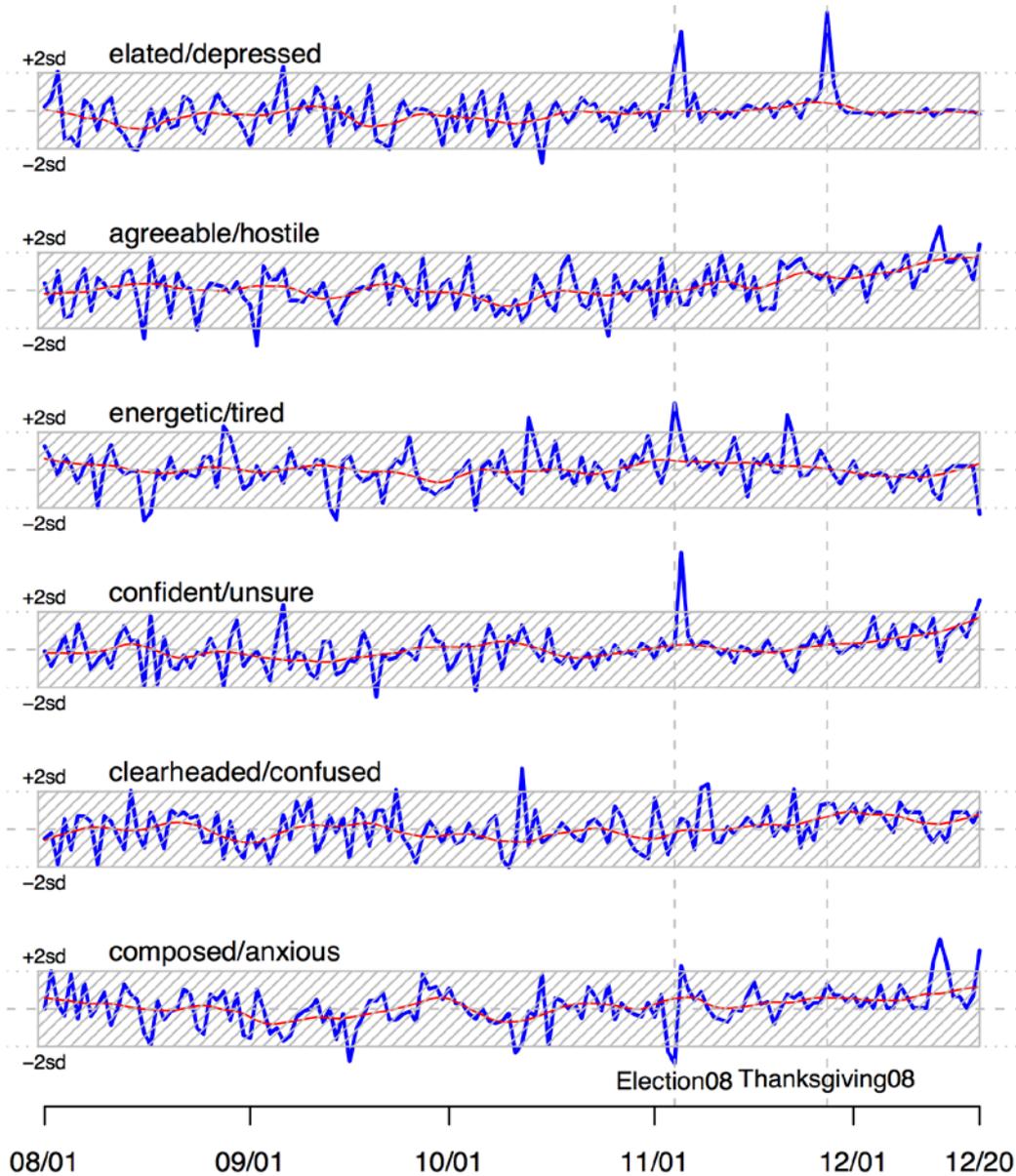
MARKET CORRELATION?
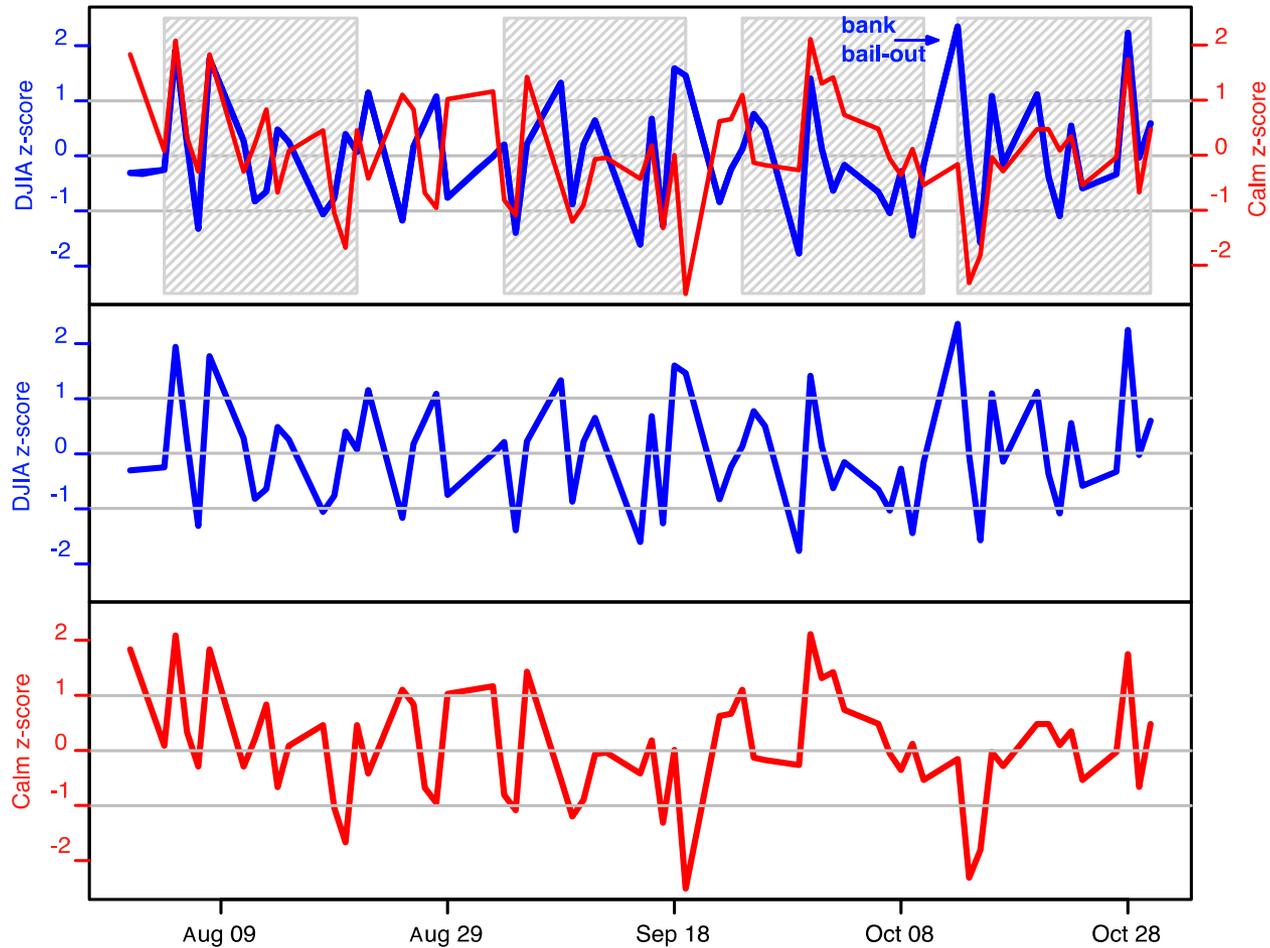
TOBBACK ET AL.
BELGIAN ECONOMIC POLICY UNCERTAINTY INDEX

Objectives:

Improve on existing indicator of Economic Policy Uncertainty using text mining of Flemish newspapers

Approach:

- **Data: 210,000 articles in Flemish newspapers (1998-2014)**

- **Analysis:**

  1. **Term (uncertainty lexicon) freq time series (Baker 2013)**
  2. **Modality in text:** detect occurrence frequencies of selected "hedges" in paragraphs that contain "economy"
  3. **Article classification** using SVM into EPU vs. non EPU: # articles/month

TOBBACK ET AL.
BELGIAN ECONOMIC POLICY UNCERTAINTY INDEX

Outcomes:

- **New indicators of economic policy uncertainty**

- **Face/Cross-validation against other uncertainty indicators:**

    - Indicates periods of major uncertainty

    - 0.88 correlation with the spread of Belgian 10-year gov. paper vs. German bund (only 0.66 with existing indicator)

    - Correlation to uncertainty-related questions (3) of DG ECFIN industry and consumer survey:

        - Little gain in correlation between SVM and naïve method

        - Best correlation: EPU modality vs. "producer uncertainty"

TOBBACK ET AL.
BELGIAN ECONOMIC POLICY UNCERTAINTY INDEX

Comments:

- **Economic policy uncertainty:** interesting addition to "social mood" research

- **Data:** Use of extensive and varied newspaper archive

- **Analysis:**

  - Use of supervised learning technique to define Economic Policy Uncertainty "extensionally" and trace back to highly loaded features

 - Comparison of effectiveness of lexicon and machine learning techniques

SOME THOUGHTS

Comments:

- **Are Belgian newspaper articles proper data to evaluate Economic Policy Uncertainty:** should they not be the least uncertain? Ideological bias.

- **Economic Policy Uncertainty:** rather abstract concept that is boiled down to bag of words or features

    - confounding general social mood (hedging) and actual policy uncertainty?
    - Belgium is a rather stable economy with few significant policy changes. Poor test case?

- **Many of "most discriminating words" are event-related:** "greek", "Rupo", "Rompuy" -> overfitting of news cycle?

- **"Improvement" =** better correlation with existing measures? Now-casting?

- **Existing psycho-social models of human emotion?** In Bollen (2011) we use a 6 dimensional mood model that includes "energy" "clearheadedness", "confusion", etc. Connect to existing science on human emotion.

- **Use of language models and other data**

    - N-grams: map dependencies between terms and features
    - Google trends?

GOOGLE 4-GRAMS

decision making under uncertainty
Cape economy faces uncertainty
forecasters with great uncertainty
projects looks increasingly uncertain
Mortgage Basics takes uncertainty
surgery carries some uncertainty
risks associated with uncertainties
chip makers still uncertain
shops pages with uncertain
rules under model uncertainty
Orleans colleges face uncertain
bowl fate still uncertain
monetary policy with uncertain
Home phones face uncertain
Internet shops with uncertain
Defense review addresses uncertain
reduced this projection uncertainty
seafood industry remains uncertain
quake households face uncertain
immigration status still uncertain
Middle East future uncertain
dysfunction drugs remains uncertain

STAATSCHULD EN ONRUST
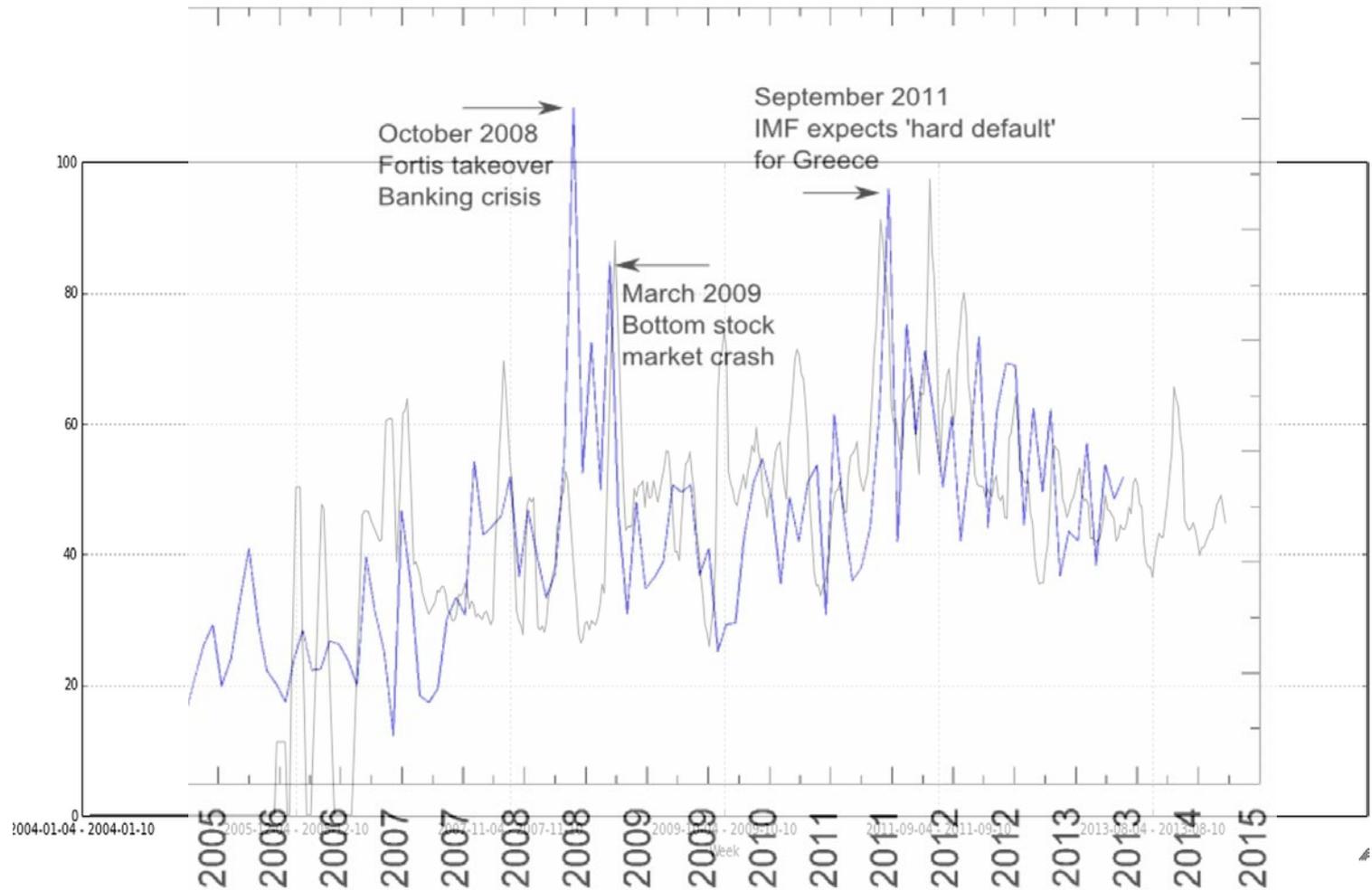
NYMAN ET AL "NEWS AND NARRATIVES"

Objectives:

Operationalize "conviction narratives" to construct quantitative metrics that capture longitudinal sentiment and consensus in market-relevant news and commentary archives

Approach:

- **Data: Bank of England Commentary (26x12x10), Broker news reports (14), Reuters news articles (12M), span ~ 2000-2013**

- **Analysis:**

    1. **Lexicon-based sentiment analysis** method to measure excitement vs. anxiety using 2 x 150 terms
    2. **Entropy of doc/topic clustering** to measure consensus vs. disagreement

NYMAN ET AL "NEWS AND NARRATIVES"

Outcomes:

- Lower anxiety & excitement levels before financial crisis of 2008, sharp uptick in anxiety during

- Correlation between Broker sentiment and Michigan Consumer Confidence index (~0.7 at a 3 week lead)

- Bank of England commentary sentiment vs. VIX: ~-0.6

- Consensus/disagreement in Reuter's London news ~ onset of financial crises: growing consensus at low anxiety values antecedent to crisis

NYMAN ET AL "NEWS AND NARRATIVES"

Comments:

- **Starting point is social theory: uncommon in comp. soc. sci**

- **Use of multiple relevant sources of text and news data**

- **Confirms earlier findings on influence of fear/energy/excitement on market (Bollen, 2011)(Gilbert 2011), Mao (2011), Preis(2013)**

- **Models follows existing "circumplex" models of human mood (valence/arousal, POMS, etc)**

- **Extensive cross-validation against existing indicators (VIX)**

- **Additional dimension: not just sentiment, but narrative consensus**

SOME THOUGHTS

- **News and commentary text:** nowcasting or post-hoc casting?
- **Lexicon method:**
    - analysis of text valence, not necessarily event or commentary valence
    - False negatives (no lexicon match) & false positives (incorrect lexicon match)
- **Operationalization alert:** Narrative is not bag of words, but semantics over time, knowledge modeling?
- **Choice of theory is equally an arbitrary choice (social science):** confirmation bias.
- **Heeding Popper:** What's a sufficiently significant drop or increase to falsify hypothesis?
- **Neglect of social network effects in public fear and excitement**
- **Cross-validation:** correlation with existing metrics, but what's unique value/validity of new indicators?

BOTH PAPERS

- **Acknowledgement of multi-dimensional sentiment/mood:** not just pos/neg, but also fear, uncertainty, confusion, energy, etc. congruent with existing NLP and psychological models

- **Dispersion of sentiment/valence across community, "consensus":** crucial indicator

- **Tapping into wisdom of the crowd effect, connections to social psychology and computational social science:**

  - WoCs works with independent judgments and wide range of estimates
  - Sentiment=estimate, consensus=moments of the distribution
 - Must include model of social network effects and social influence

GALTON'S FAT OX

GALTON, OXES, AND VOTES

A few other historic results in the history of crowd estimation and voting…

Galton (1907), Nature, 1949(75):450-451

- "In these democratic days, any investigation into
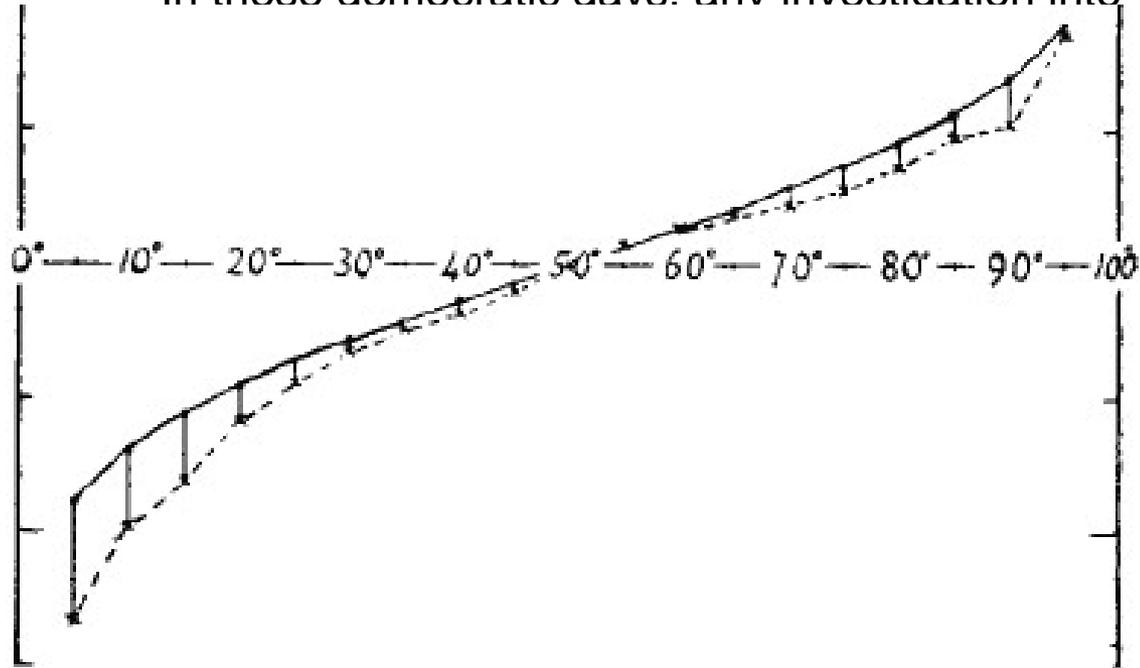
1822-1911

The continuous line is the normal curve with p.e.=37.
The broken line is drawn from the observations.
The lines connecting them show the differences between the observed and the normal.
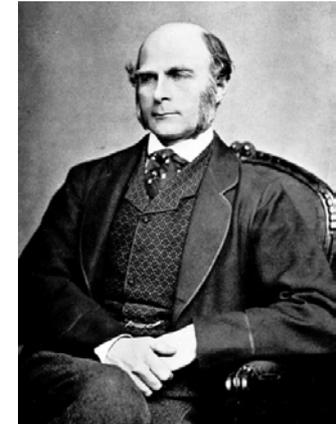
SOCIAL INFLUENCE VS. WISDOM OF THE CROWD

Salganik et al (2006) Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market, Science, 311

Lorenza et al (2011) How social influence can undermine the wisdom of crowd effect. 10.1073/pnas.1008636108 PNAS May 16

Social influence distorts and biases outcome by narrowing diversity, dispersion of opinion – social media and news??

HOMOPHILY (& CONTAGION) IN SOCIAL NETWORKS
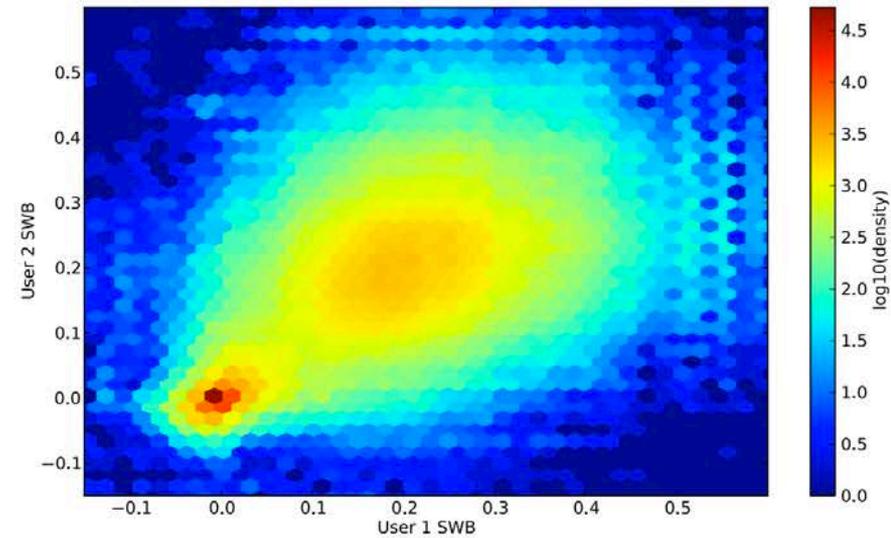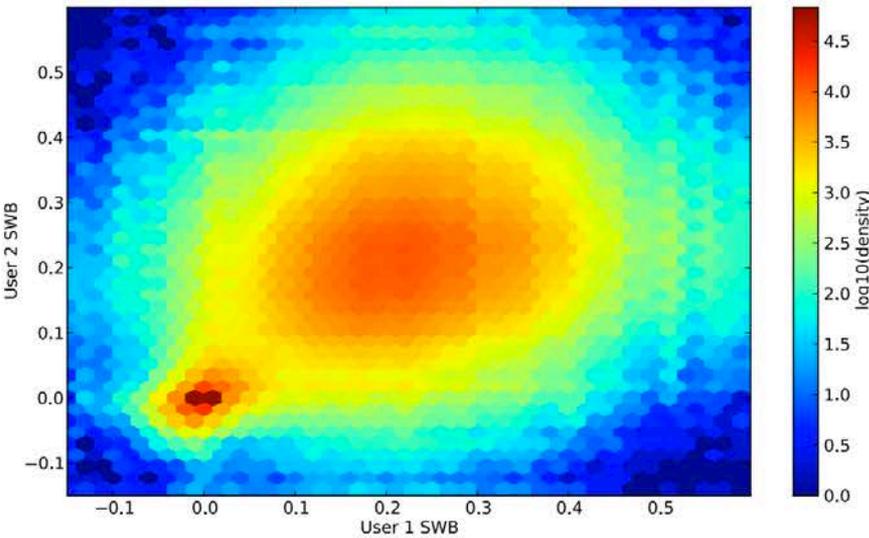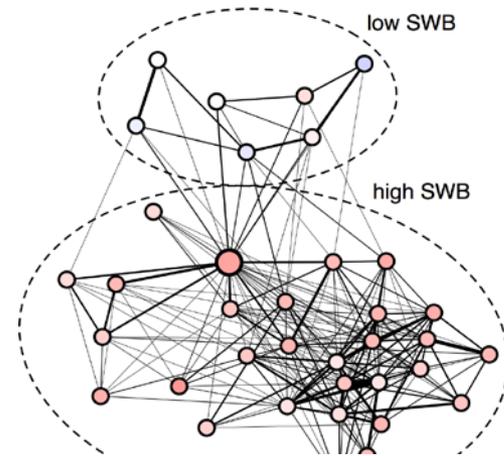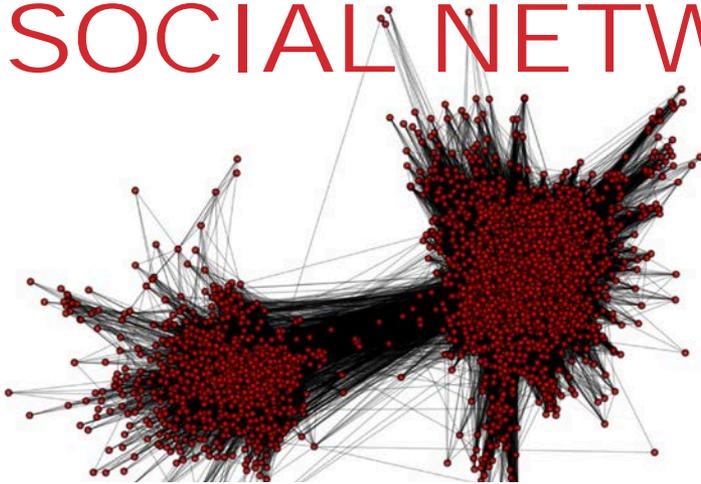
Johan Bollen, Bruno Gonçalves, Guangchen Ruan & Huina Mao. Happiness is assortative in online social networks. Artificial Life, Summer - 17(3), 237-251 (2011)