STUDYING MOOD VARIATIONS IN LONGITUDINAL TWITTER TIMELINES
APPLICATIONS TO THE DETECTION OF PSYCHOLOGICAL TRANSITIONS

JOHAN BOLLEN – JBOLLEN@INDIANA.EDU
INDIANA UNIVERSITY
SCHOOL OF INFORMATICS AND COMPUTING
CENTER FOR COMPLEX NETWORKS AND SYSTEMS RESEARCH
COGNITIVE SCIENCE PROGRAM
U. WAGENINGEN, NETHERLANDS – SPARCS INSTITUTE
Social Media Platforms by Total Number of Users

- YouTube: 800 Million
- Facebook: 845 Million
- Twitter: 500 Million
- Wikipedia: 16 Million
- Pandora: 125 Million
- Google+: 90 Million
- Wordpress: 75 Million
- Myspace: 30 Million
- Digg: 6 Million
- LinkedIn: 150 Million
- Spotify: 10 Million
- Pinterest: 10 Million
- Foursquare: 10 Million
Implementing the Macroscope vision


Psychology
Sociology
Biology
Science

Computational social science

GleamViz
OUR WORK: COLLECTIVE MOOD STATES
THE CROWD’S LIMBIC SYSTEM

Epictetus: “Men are disturbed, not by things, but by the principles and notions which they form concerning things”

Social mood:
- Social issues: public health, unrest, ...
- Economic issues: growth, market behavior

But how can you determine how people feel?
OUR MACROSCOPES

Network science
Large-scale social media data
Natural language processing
Sentiment and mood analysis

This talk:

1. Social mood & stock market prediction
2. Public mood: assortativity, contagion, & eigenmoods
3. Individual mood: longitudinal analytics, mental health
SENTIMENT ANALYSIS: TOOLS

- Lexicons: ANEW (Valence, Arousal, Dominance), OpinionFinder, SentiWordnet (Wordnet)

- Machine learning approaches: classification (positive, negative, neutral)
  - Naïve Bayesian classifiers: learning from “training set” which terms mark a particular mood, classify on that basis (bag of words)
  - Support Vector Machines (similar notion)
  - Semantic and grammatical analysis
  - Stanford CoreNLP Sentiment


Lexicon sentiment rating examples (ANEW):

“I'm totally coveting yer seafaring ways...my dream is oceans of bed...”
Arousal = 4.070, valence = 7.120, dominance= 6.205

“Feeling blue.. hoping I feel better before Christmas :("  
Arousal = 5.290, valence=7.280, dominance= 5.500

Analyze social media within hourly, daily time intervals
Study fluctuations

We Feel Fine
http://www.wefeelfine.org/
Moodviews
http://moodviews.com
Myspace: Thelwall (2009), FB: United States Gross National Happiness
http://apps.facebook.com/usa_gnh/, Mislove: pulse of the nation (http://www.ccs.neu.edu/home/amislove/twittermood/), Peter Dodd – hedonometrics:
http://arxiv.org/pdf/1101.5120v4

From: Dodd (2011) Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter
FINANCIAL MARKET PREDICTION

Collection of tweets: April 29, 2006 to December 20, 2008, 2.7M users

Subset: August 1, 2008 to December 2008 (9,664,952 tweets)

GPOMP: Based on Profile of Mood States, 6 dimensions of mood – Calm, Alert, Sure, Vital, Kind, Happy.

Evaluation

<table>
<thead>
<tr>
<th>MAPE (%)</th>
<th>I_{OF}</th>
<th>I_0</th>
<th>I_1</th>
<th>I_{1,2}</th>
<th>I_{1,3}</th>
<th>I_{1,4}</th>
<th>I_{1,5}</th>
<th>I_{1,6}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.95</td>
<td>1.94</td>
<td>1.83</td>
<td>2.03</td>
<td>2.13</td>
<td>2.05</td>
<td>1.85</td>
<td>1.79*</td>
<td>1.68</td>
</tr>
<tr>
<td>Direction (%)</td>
<td>73.3</td>
<td>73.3</td>
<td>86.7*</td>
<td>60.0</td>
<td>46.7</td>
<td>60.0</td>
<td>73.3</td>
<td>80.0</td>
</tr>
</tbody>
</table>

Google search 19 fear terms, e.g. recession

Cross-correlation US Unemployment rate vs. GIS

Cross-correlation DSI and Twitter Investor Sentiment


Chinese Financial Lexicon Construction (919,246 news headlines)

Positive seed words: 盈利(profit), 恢复(recover), 涨停(limit-up), 反弹(rebound), 优势(advantages) ...

Negative seed words: 跌停(limit-down), 亏(deficit), 失败(failure), 损失(loss), 稽查(inspect) ...

Testing
OTHER RESULTS


Computational models of consumer confidence from large-scale online attention data: crowd-sourcing econometrics.

**BUT, WHERE DOES ONLINE COLLECTIVE MOOD COME FROM?**

Measuring and averaging individual mood states =
- not really “collective” mood… sum of individual text sentiment
- Collective mood ~ emergent phenomenon, endogenous response, function of social network, response to drivers “internal” to community
- Median or average mood !~ variance, uncertainty, communities, language

**Very active research area:**
- Use of epidemiological models to model mood contagion (Ferrara et al)
- Agent-based models (Garcia, 2012)
- Role of homophily, preferential attachment, contagion, socio-economic factors, modeling uncertainty and community effects (our present work)
ROLE OF HOMOPHILY IN COLLECTIVE MOOD

“birds of a feather” in social networks:


- **Homophily**: tendency of individuals to associate with those of similar age, sex, religion, race, etc.
- **Heterophily**: associating with opposite or contrary features
- **Distinction**: Homophily vs features that it applies to.


**Note: homophily != contagion**

TWITTER MOOD ASSORTATIVITY

Nov. 20, 2008 to May 29, 2009, 4,844,430 user timelines, 129M tweets

Network parameter | Value
--- | ---
Nodes | 102,009 users
Edges | 2,361,547
Density | 0.000454
Diameter | 14
Avg. Degree: | 46.300
Avg. Clustering Coefficient | 0.262

Longitudinal “Subjective Well-being”

\[ S(u) = \frac{N_p(u) - N_n(u)}{N_p(u) + N_n(u)} \]

Pairwise SWB Assortativity

Neighborhood SWB Assortativity

IUNI --- jbollen@indiana.edu
RECENT WORK

Two lines of research:

1) Study spectrum of mood states
   1) Collective: “Eigenmoods”
   2) Individual: longitudinal timeline analysis

2) Applications to personal well-being and health
   1) Detecting mental health issues: depression features
   2) Critical transitions, early warning indicators
Average sentiment: language model x word sentiment value
Remove language effect, and find “eigenmood”.
Singular Value Decompositions of Mood Bin x time matrix:
- Decompose mood bin x time matrix: $M = U\Sigma V^t$
- Approximate without first singular value (language model)

Interesting correlations to social phenomena, for example Google sex searches, birth rates, and public eigenmoods
CRITICAL TRANSITIONS IN MENTAL HEALTH

Ingrid van de Leemput et al.


Applications to mental health: modeling depression as a critical transition in mood dynamics

2 populations: (1) not depressed (n=535), depressed (n=93). 6 consecutive days, 10 times a day (7:30 - 22:30). Monitoring of follow-up course depressive symptoms

Van de Leemput et al (2014), PNAS 111(1) Critical slowing down as early warning for the onset and termination of depression
DEPRESSION

With Ali Varamesh, Ingrid van de Leemput

Detect linguistic features of Depression in online social media users
Query: “I was diagnosed with depression today”

<table>
<thead>
<tr>
<th>Class</th>
<th>Users</th>
<th>Tweets</th>
<th>avg#tweets/user</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depressed</td>
<td>42</td>
<td>124,015</td>
<td>2,952</td>
</tr>
<tr>
<td>Non-Depressed</td>
<td>73</td>
<td>150,775</td>
<td>2,065</td>
</tr>
</tbody>
</table>

Feature extraction: Twitter Part-of-Speech Tagger from ARK project at CMU
24 different POS tags and about 25,448 tokens
User feature vectors: number of times a token or POS tag is used in a user’s timeline divided by the total number of user’s tweets.
Based on OneR test removal of “diagnosis” and “depression”
LONGITUDINAL MOOD DATA
CRITICAL TRANSITIONS

- About 700,000 timelines of individual Twitter users: June 2010 to June 2013
- Use of ANEW lexicon (13k terms: Valence, Arousal, Dominance), CRR U. Gent
- “I was diagnosed with depression today”

Bayesian updating using ANEW lexicon and Google n-grams as prior, with likelihood $P(D|M)$ from lexicon distribution.
CONCLUSION

Growth of “social” data is likely to be most significant in “ego”-related data vs. network data pur sang:

- significant longitudinal data since advent of social media 8 years ago
- Increasing use of personal monitoring devices
- Increasing focus on personal, well-being related data

Tremendous opportunities in:

- medicine, public health, forecasting, and possibly intervention strategies to prevent “hysteresis”.
- Study relations between individual features vs. network topology.
COLLABORATORS

Luis Rocha (IU)
Marten Scheffer (IU)
Huina Mao (ORNL)
Rion Coreia (IU)
Ian Woods (IU)
Ali Varamesh (IU)
Ingrid van de Leemput (U. Wageningen)
**READINGS**


Van de Leemput et al (2014), Critical slowing down as early warning for the onset and termination of depression, *PNAS 111*(1)