CS 6474/4803 Social Computing: Analyzing Language I

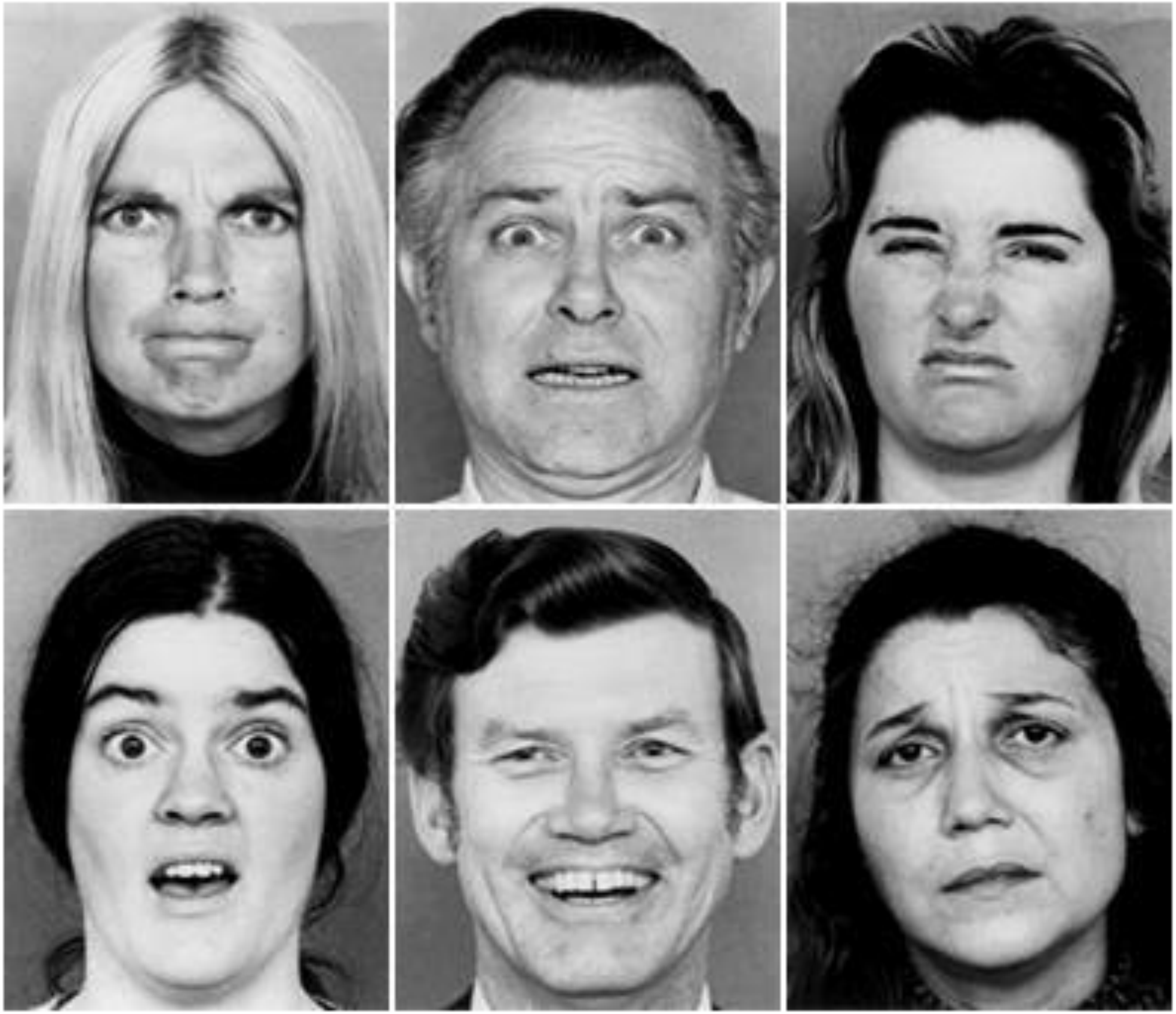
Munmun De Choudhury

munmund@gatech.edu

Week 7 | February 22, 2021

Assignment II

Slight Change in the Class Schedule

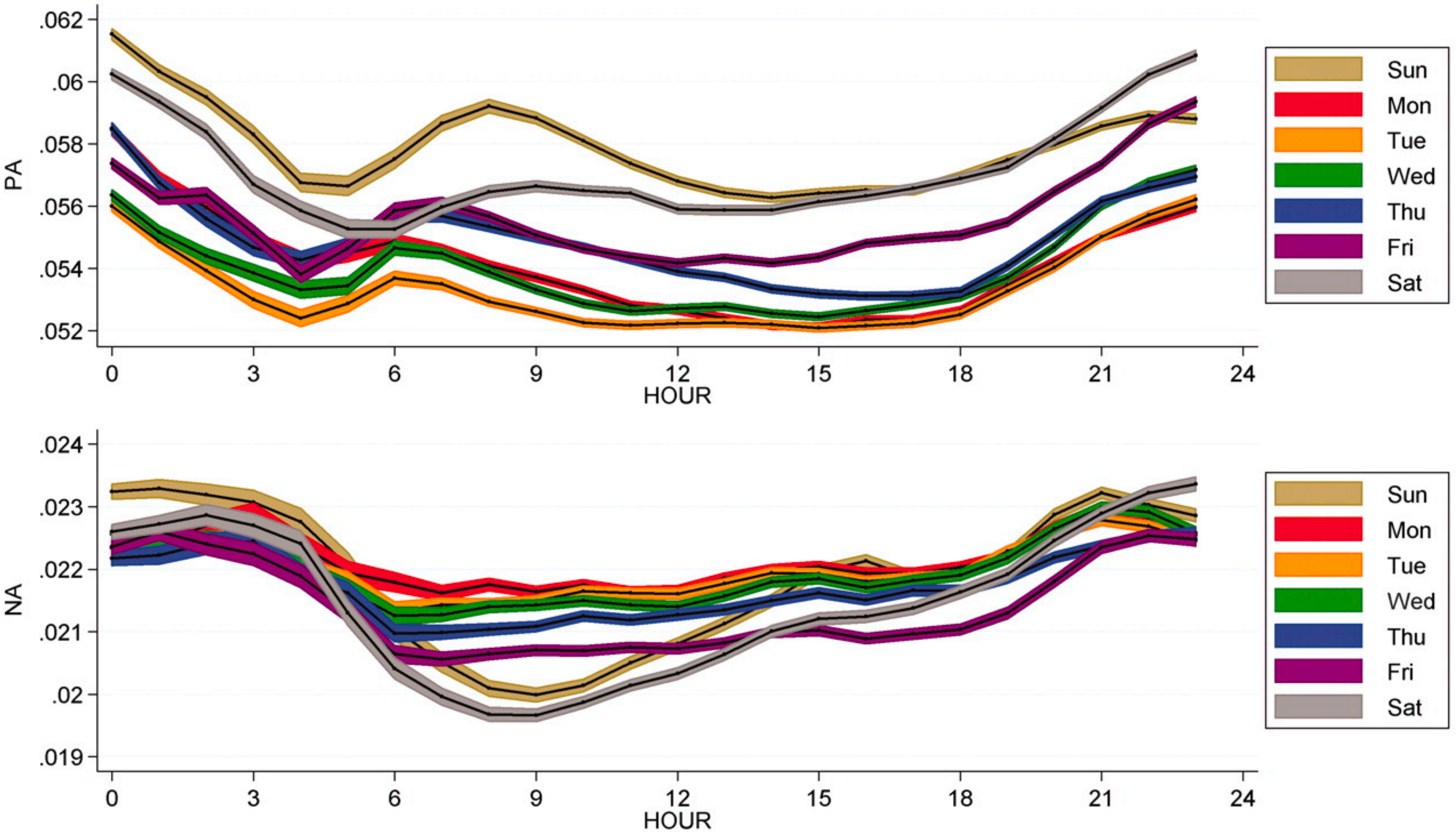


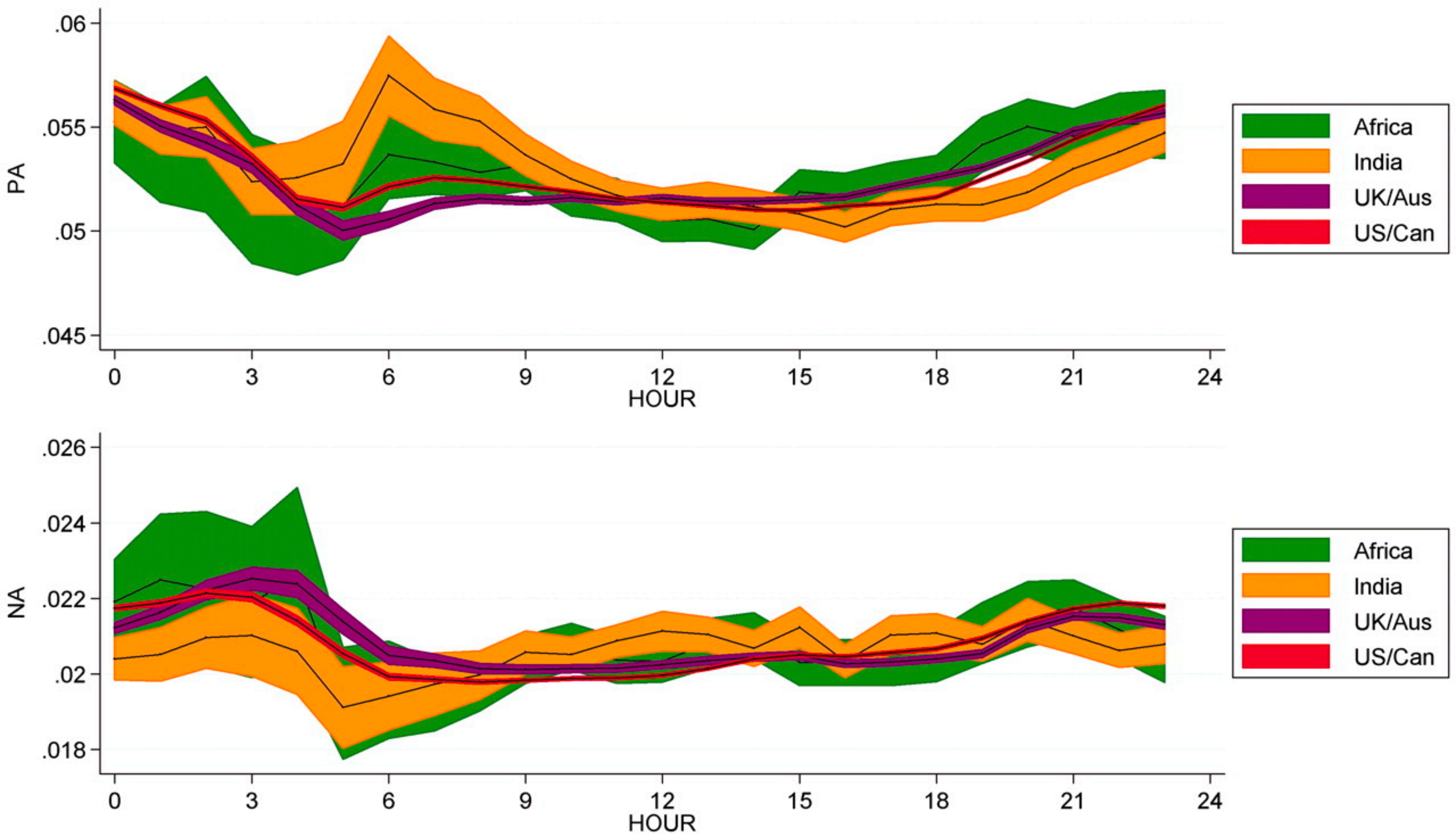
Language is the most common and reliable way for people to translate their internal thoughts and emotions into a form that others can understand. Words and language, then, are the very stuff of psychology and communication -- Tauszczik & Pennebaker

Diurnal and Seasonal Mood Vary with Work, Sleep, and Day length Across Diverse Cultures

Summary

- One of the early works examining relationship between social media mood and behavior and psychological theories.
- The potential of online social media to study individual behavior.
- Identify daily and seasonal mood variations and relate it to work, sleep and daylight.
- Validate circadian rhythms in humans.
  - PA spike in the morning, NA increases as the day progresses
- Measure positive affect and negative affect based on the lexicon LIWC.
- PA and NA are not mirror images of each other.


PA
.062
.06
.058
.056
.054
.052
0
3
6
9
12
15
18
21
24
HOUR
Sun
Mon
Tue
Wed
Thu
Fri
Sat
NA
.024
.023
.022
.021
.02
.019
0
3
6
9
12
15
18
21
24
HOUR
Sun
Mon
Tue
Wed
Thu
Fri
Sat



.06
.055
.05
.045
PA
0
3
6
9
12
15
18
21
24
HOUR
Africa
India
UK/Aus
US/Can
.026
.024
.022
.02
.018
NA
0
3
6
9
12
15
18
21
24
HOUR
Africa
India
UK/Aus
US/Can

Twitter is used by millions and both the papers extensively leverage this source of data in measuring mood and affect.

How does use of Twitter for this purpose address limitations in existing mood or affect measurement methods?

Twitter is used by millions and the paper extensively leverages this source of data in measuring mood and affect.

But could Twitter also have bias?

How do you expect the results relating to mood to be different if the paper used: 1) Facebook 2) Instagram?

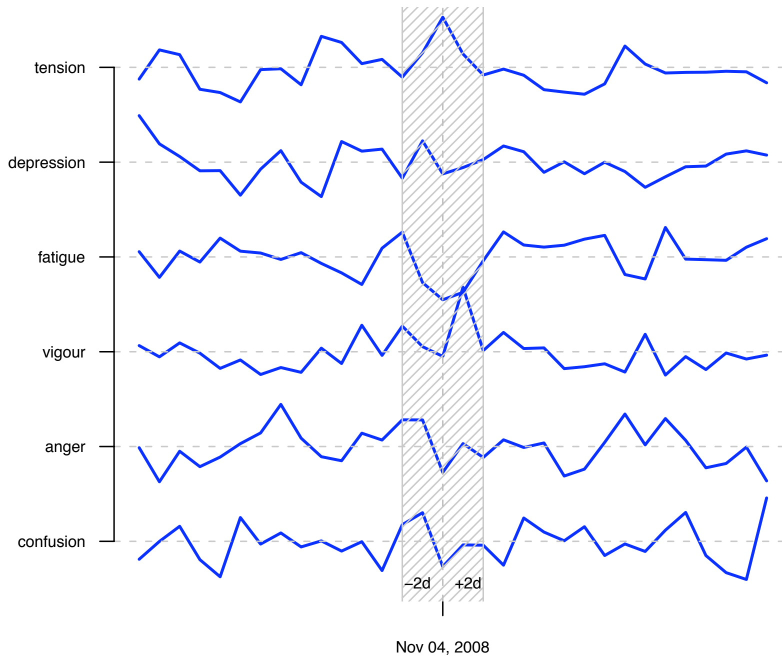
Could platform affordances impact specific moods and their manifestations on social media? How?

Class Exercise

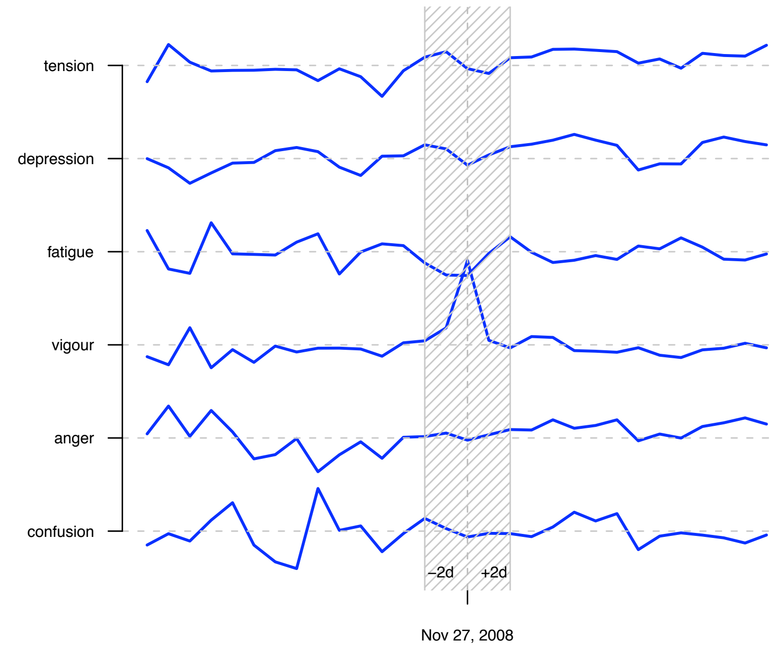
Why is measuring mood useful?
Some examples follow…

Modeling Public Mood and Emotion: Twitter Sentiment and Socioeconomic Phenomena – (Bollen, Pepe, Mao, 2010)

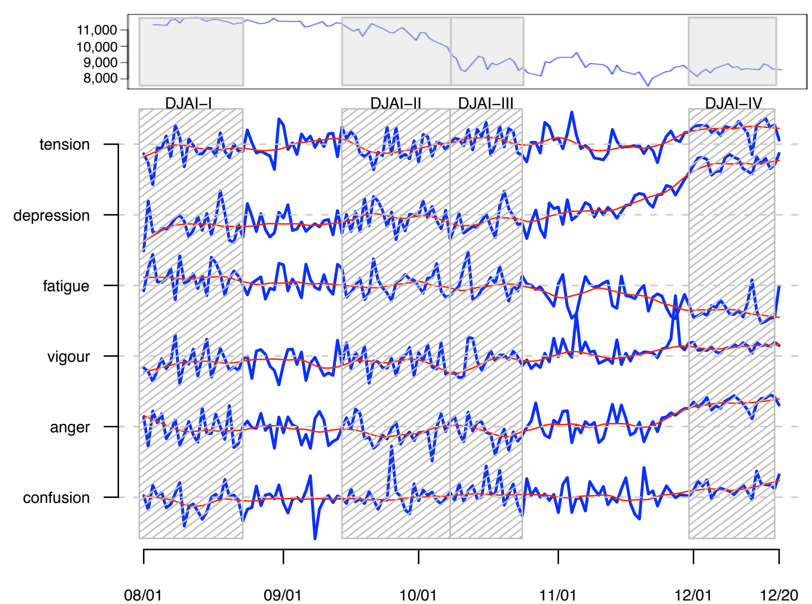
- Examine how Twitter moods reflect social, political, and economic events
- Use POMS (profile of mood states) for detecting mood-indicative twitter posts.
  - POMS dimensions: tension, depression, anger, vigor, fatigue and confusion
- Investigate how a six vector representation of moods deviates during different big scope events.
- High stress/tension during elections; excitement/vigor during thanksgiving.



2008 Presidential elections

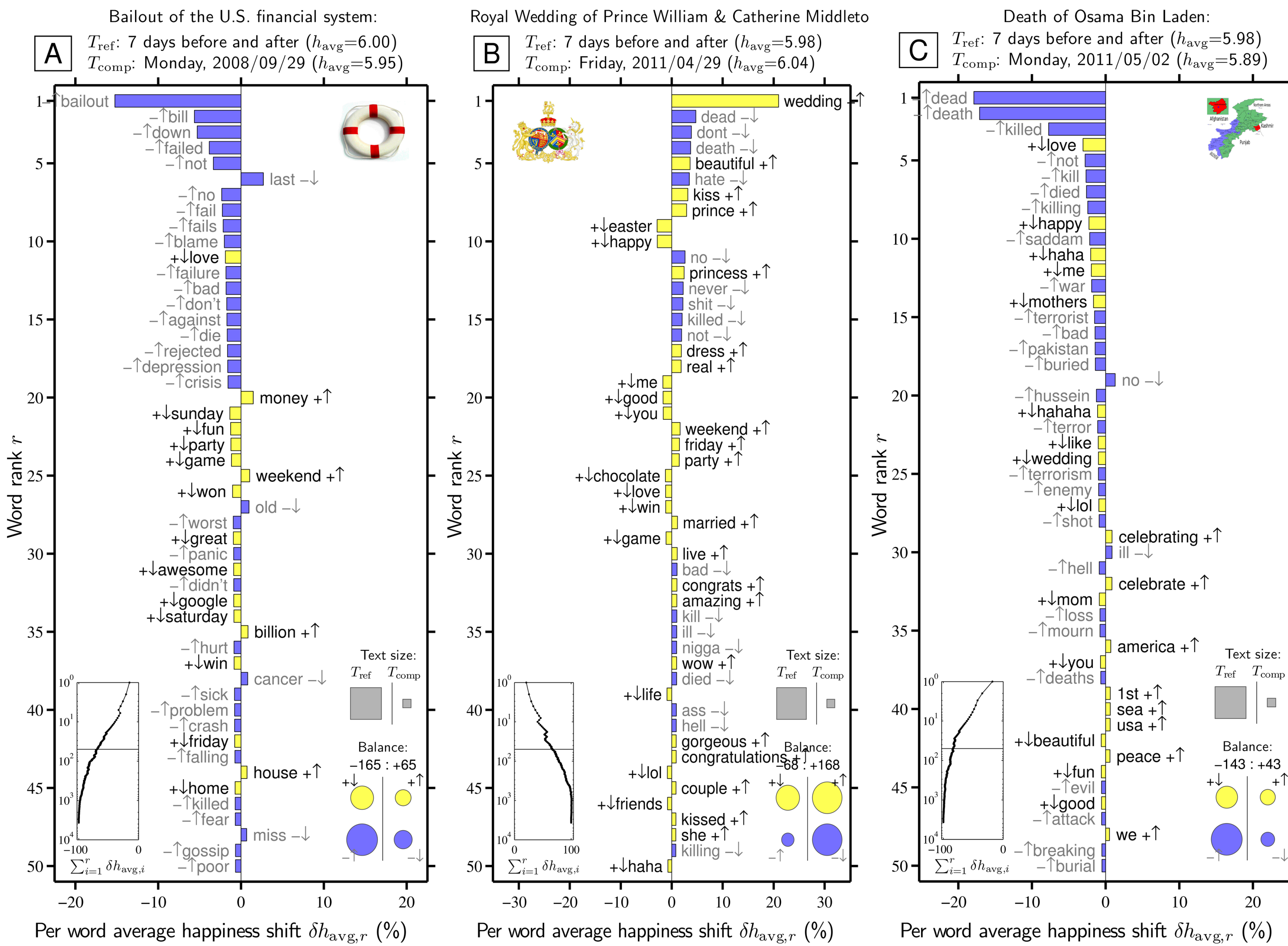


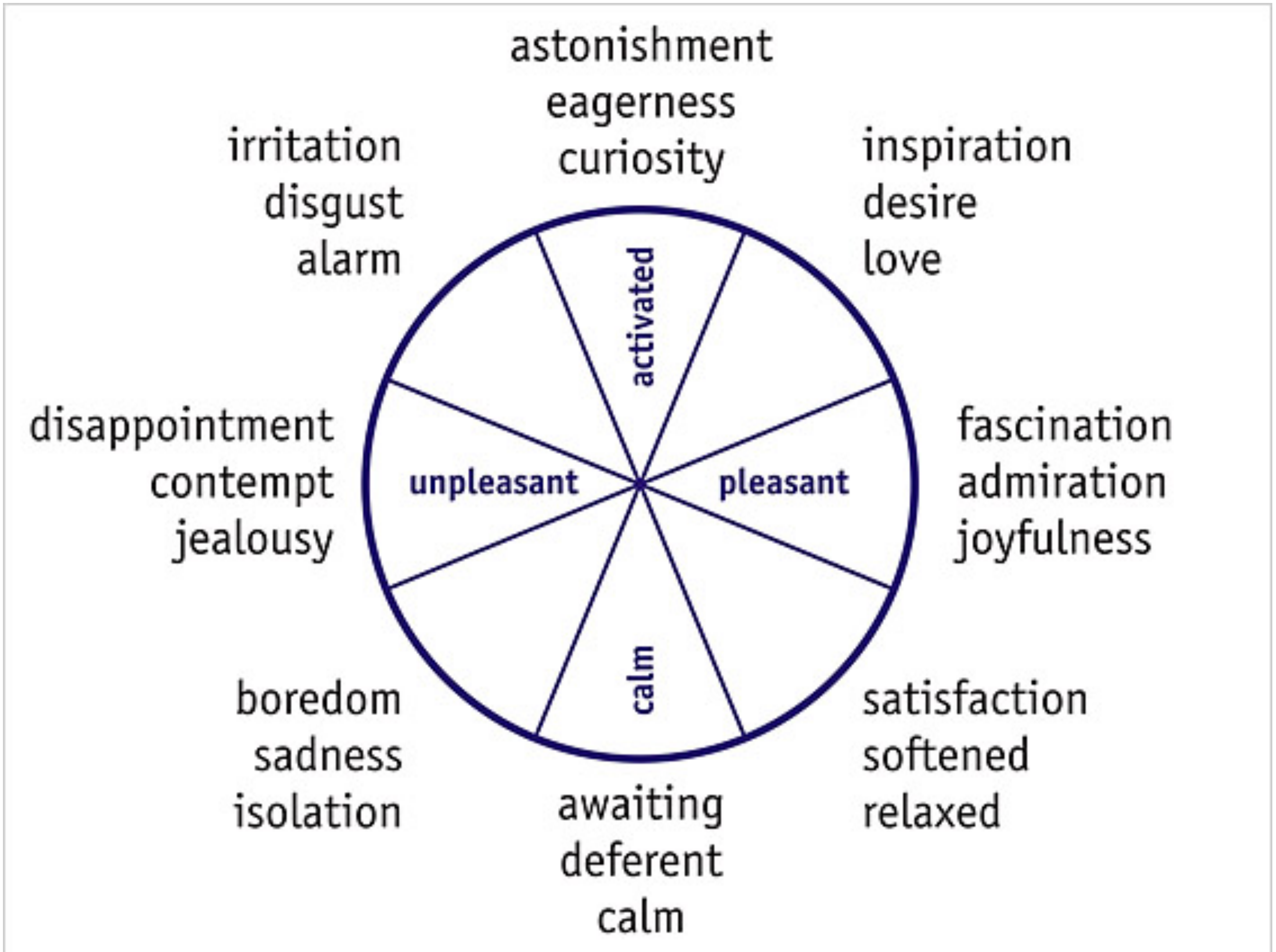
Thanksgiving



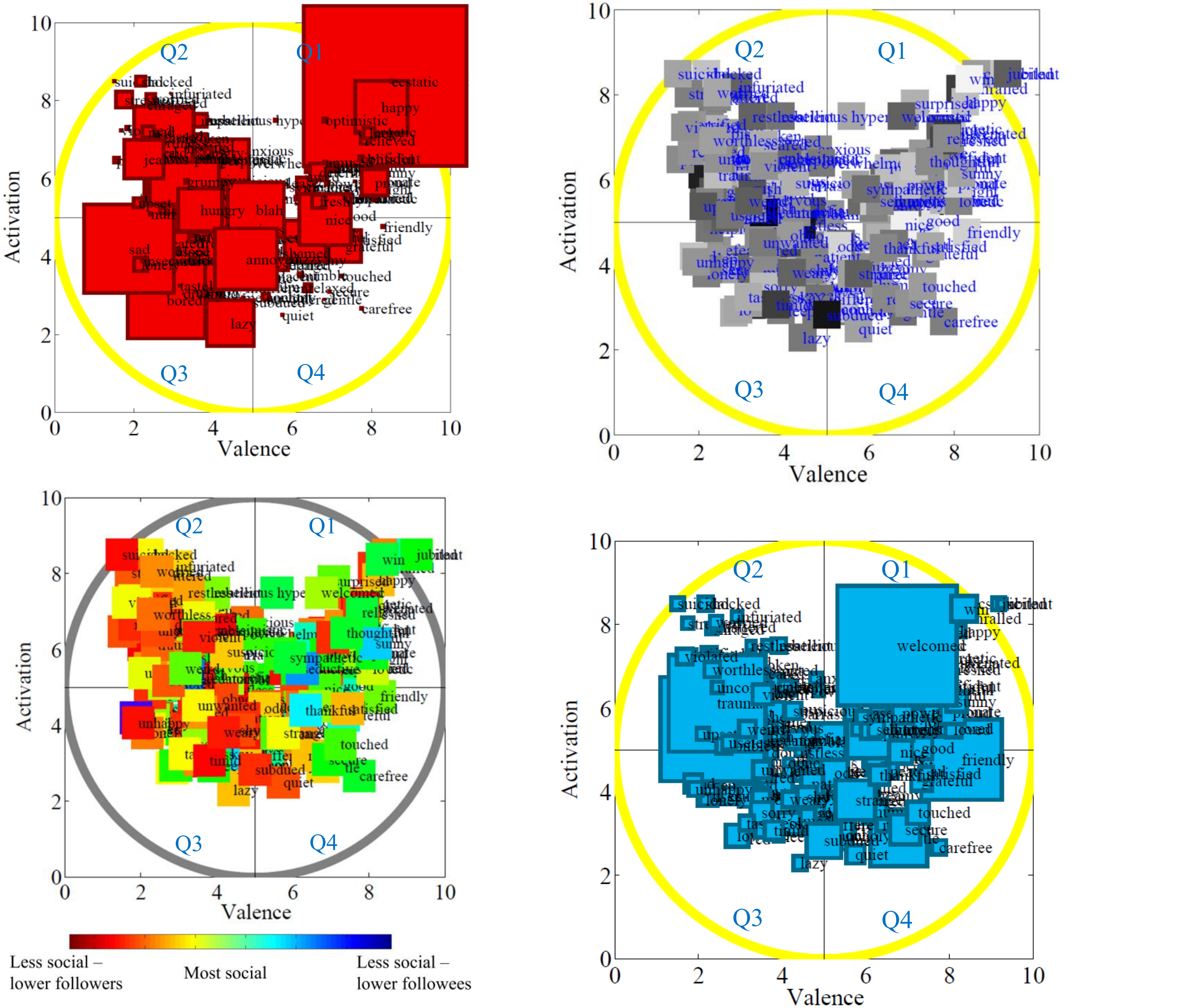
Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter

Peter Sheridan Dodds , Kameron Decker Harris, Isabel M. Kloumann, Catherine A. Bliss, Christopher M. Danforth




astonishment
eagerness
curiosity
inspiration
desire
love
fascination
admiration
joyfulness
satisfaction
softened
relaxed
awaiting
deferent
calm
boredom
sadness
isolation
disappointment
contempt
jealousy
irritation
disgust
alarm
activated
pleasant
calm
unpleasant

Not All Moods Are Created Equal! Exploring Human Emotional States in Social Media, by De Choudhury, Counts, and Gamon 2012



- **Emotion** – brief conscious experience characterized by intense mental activity and a high degree of pleasure or displeasure
- **Affect** – an instinctual reaction to stimulation occurring before the typical cognitive processes considered necessary for the formation of a more complex emotion
- **Mood** – emotional state. Moods differ from emotions or affects in that they are less specific, less intense, and less likely to be triggered by a particular stimulus or event
- **Sentiment** – attitude or opinion with respect to a specific topic, event or situation

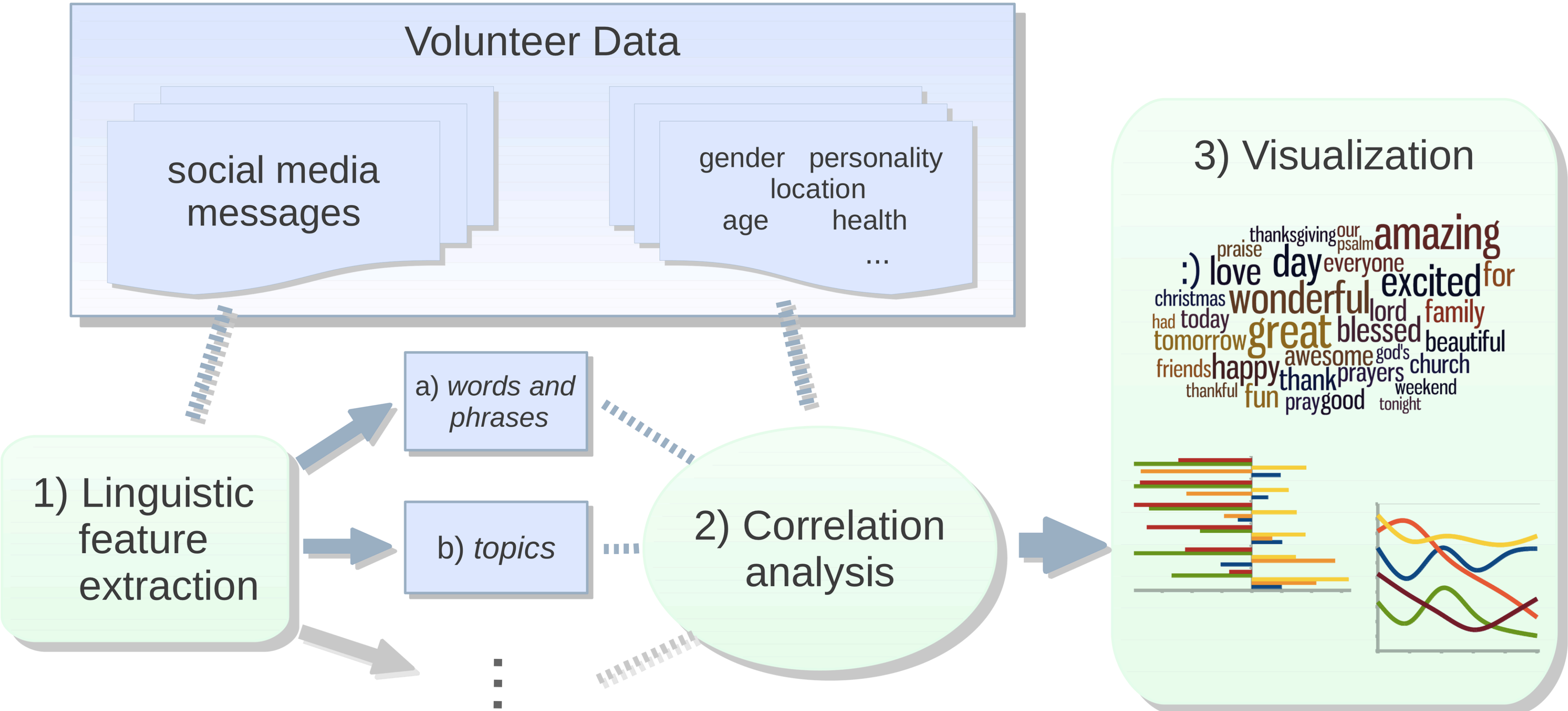
Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach

Summary

- Facebook data of 75K individuals
- Users took personality tests
  - Participants volunteered to share their status updates as part of the My Personality application, where they also took a variety of questionnaires
- Authors found found striking variations in language with personality, gender, and age
  - Use of an open vocabulary approach

- Results confirmed previously known social science findings, suggested new hypotheses, and showed sustained face validity

Summary



Volunteer Data
social media messages
gender personality
location
age health
...
3) Visualization
1) Linguistic feature extraction
a) words and phrases
b) topics
2) Correlation analysis

Summary

- Open Vocabulary: Differential Language Analysis
- Key characteristics:
  - Open-vocabulary – it is not limited to predefined word lists. Rather, linguistic features including words, phrases, and topics (sets of semantically related words) are automatically determined from the texts. (I.e., it is "data-driven".) This means DLA is classified as a type of open-vocabulary approach.
  - Discriminating – it finds key linguistic features that distinguish psychological and demographic attributes, using stringent significance tests.
  - Simple – it uses simple, fast, and readily accepted statistical techniques.

Summary

features	Gender accuracy	Age R	Extraversion R	Agreeableness R	Conscientious. R	Neuroticism R	Openness R
LIWC	78.4%	.65	.27	.25	.29	.21	.29
Topics	**87.5%**	**.80**	**.32**	**.29**	**.33**	**.28**	**.38**
WordPhrases	**91.4%**	**.83**	**.37**	**.29**	**.34**	**.29**	**.41**
WordPhrases + Topics	**91.9%**	**.84**	**.38**	**.31**	**.35**	**.31**	**.42**
Topics + LIWC	**89.2%**	**.80**	**.33**	**.29**	**.33**	**.28**	**.38**
WordPhrases + LIWC	**91.6%**	**.83**	**.38**	**.30**	**.34**	**.30**	**.41**
WordPhrases + Topics + LIWC	**91.9%**	**.84**	**.38**	**.31**	**.35**	**.31**	**.42**

accuracy: percent predicted correctly (for discrete binary outcomes). *R*: Square-root of the coefficient of determination (for sequential/continuous outcomes). *LIWC*: *A priori* word-categories from Linguistic Inquiry and Word Count. *Topics*: Automatically created *LDA* topic clusters. *WordPhrases*: words and phrases (n-grams of size 1 to 3 passing a collocation filter). Bold indicates significant ($p<.01$) improvement over the baseline set of features (use of *LIWC* alone).

doi:10.1371/journal.pone.0073791.t002

How is an open vocabulary approach more suitable for social media language data over closed vocabulary ones?

Character *n*-grams

Quantifying Mental Health Signals in Twitter

Glen Coppersmith **Mark Dredze** **Craig Harman**
Human Language Technology Center of Excellence
Johns Hopkins University
Balitmore, MD, USA

Abstract

The ubiquity of social media provides a rich opportunity to enhance the data available to mental health clinicians and researchers, enabling a better-informed and better-equipped mental health field. We present analysis of mental health phenomena in publicly available Twitter data, demonstrating how rigorous application of simple natural language processing methods can yield insight into specific disorders as well as mental health writ large,

In contrast, social media is plentiful and has enabled diverse research on a wide range of topics, including political science (Boydstun et al., 2013), social science (Al Zamal et al., 2012), and health at an individual and population level (Paul and Dredze, 2011; Dredze, 2012; Aramaki et al., 2011; Hawn, 2009). Of the numerous health topics for which social media has been considered, mental health may actually be the most appropriate. A major component of mental health research requires the study of behavior, which may be manifest in how an individual acts, how they com-

What to do about bad language on the internet

Jacob Eisenstein
jacobe@gatech.edu
School of Interactive Computing
Georgia Institute of Technology

Abstract

The rise of social media has brought computational linguistics in ever-closer contact with *bad language*: text that defies our expectations about vocabulary, spelling, and syntax. This paper surveys the landscape of bad language, and offers a critical review of the NLP community's response, which has largely followed two paths: normalization and domain adaptation. Each approach is evaluated in the context of theoretical and empirical work on

These examples are selected from celebrities (for privacy reasons), but they contain linguistic challenges that are endemic to the medium, including non-standard punctuation, capitalization, spelling, vocabulary, and syntax. The consequences for language technology are dire: a series of papers has detailed how state-of-the-art natural language processing (NLP) systems perform significantly worse on social media text. In part-of-speech tagging, the accuracy of the Stanford tagger (Toutanova et al., People use social media for all kinds of reasons and purposes. On Facebook in particular, people are heavily concerned about impression management.

Why do you think the assessments of personality are still accurate?

Why is gender and personality inference useful for social computing researchers and professionals?

Research Note

Research and Politics
July-September 2017: 1–9

DOI: 10.1177/2053168017720008
journals.sagepub.com/home/rap
SAGE

Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users

Jonathan Mellon[1] and Christopher Prosser[2]

Abstract

A growing social science literature has used Twitter and Facebook to study political and social phenomena including for election forecasting and tracking political conversations. This research note uses a nationally representative probability sample of the British population to examine how Twitter and Facebook users differ from the general population in terms of demographics, political attitudes and political behaviour. We find that Twitter and Facebook users differ substantially from the general population on many politically relevant dimensions including vote choice, turnout, age, gender, and education. On average social media users are younger and better educated than non-users, and they are more liberal

Experimental evidence of massive-scale emotional contagion through social networks

Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock

PNAS June 17, 2014 111 (24) 8788-8790; first published June 2, 2014 https://doi.org/10.1073/pnas.1320040111

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved March 25, 2014 (received for review October 23, 2013)

This article has corrections. Please see:

Editorial Expression of Concern: Experimental evidence of massivescale emotional contagion through social networks

Correction for Kramer et al., Experimental evidence of massive-scale emotional contagion through social networks

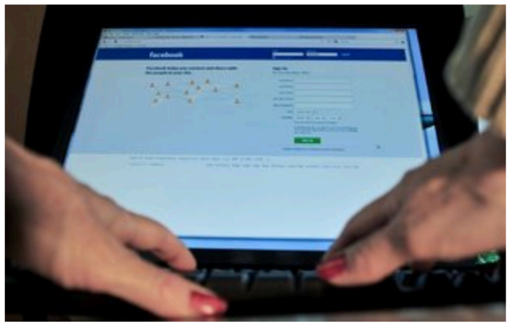
Significance

We show, via a massive (N = 689,003) experiment on Facebook, that emotional states can

TECHNOLOGY

Facebook Tinkers With Users' Emotions in News Feed Experiment, Stirring Outcry

By VINDU GOEL JUNE 29, 2014



Facebook revealed that it had altered the news feeds of over half a million users in its study.
Karen Bleier/Agence France-Presse — Getty Images

To Facebook, we are all lab rats.

Facebook routinely adjusts its users' news feeds — testing out the number of ads they see or the size of photos that appear — often without their knowledge. It is all for the purpose, the company says, of creating a more alluring and useful product.

But last week, Facebook revealed that it had manipulated the news feeds of over half a million randomly selected users to change the number of positive and negative posts they saw. It was part of a psychological study to examine how emotions can be spread on social media.

The company says users consent to this kind of manipulation when they agree to its terms of service. But in the quick judgment of the Internet, that argument was not universally accepted.

"I wonder if Facebook KILLED anyone with their emotion manipulation stunt. At their scale and with depressed people out there, it's possible," the privacy activist Lauren Weinstein wrote in a Twitter post.

On Sunday afternoon, the Facebook researcher who led the study, Adam D. I. Kramer, posted a public apology on his Facebook page.

Awareness of some of the risks

- Marginalized groups might be more marginalized in gender/personality inference because their language is less represented
  - LGBTQ / non-binary gender representation
- Unintended biases

Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings

Tolga Bolukbasi[1], **Kai-Wei Chang**[2], **James Zou**[2], **Venkatesh Saligrama**[1,2], **Adam Kalai**[2]
[1]Boston University, 8 Saint Mary's Street, Boston, MA
[2]Microsoft Research New England, 1 Memorial Drive, Cambridge, MA
tolgab@bu.edu, kw@kwchang.net, jamesyzou@gmail.com, srv@bu.edu, adam.kalai@microsoft.com

Abstract

The blind application of machine learning runs the risk of amplifying biases present in data. Such a danger is facing us with *word embedding*, a popular framework to represent text data as vectors which has been used in many machine learning and natural language processing tasks. We show that even word embeddings trained on Google News articles exhibit female/male gender stereotypes to a disturbing extent. This raises concerns because their widespread use, as we describe, often tends to amplify these biases. Geometrically, gender bias is first shown to be captured by a direction in the word embedding. Second, gender neutral words are shown to be linearly separable from gender definition words in the word embedding. Using these properties, we provide a methodology for modifying an embedding to remove gender stereotypes, such as the association between the words *receptionist* and *female*, while maintaining desired associations such as between the words *queen*

Semantics derived automatically from language corpora necessarily contain human biases

Aylin Caliskan[1], **Joanna J. Bryson**[1,2], **and Arvind Narayanan**[1]

[1]Princeton University
[2]University of Bath
*Address correspondence to aylinc@princeton.edu, bryson@conjugateprior.org, arvindn@cs.princeton.edu.

ABSTRACT

Artificial intelligence and machine learning are in a period of astounding growth. However, there are concerns that these technologies may be used, either with or without intention, to perpetuate the prejudice and unfairness that unfortunately characterizes many human institutions. Here we show for the first time that human-like semantic biases result from the application of standard machine learning to ordinary language—the same sort of language humans are exposed to every day. We replicate a spectrum of standard human biases as exposed by the Implicit Association Test and other well-known psychological studies. We replicate these using a widely used, purely statistical machine-learning model—namely, the GloVe word embedding—trained on a corpus of text from the Web. Our results indicate that language itself contains recoverable and accurate imprints of our historic biases, whether these are morally neutral as towards insects or flowers, problematic as towards race or gender, or even simply veridical, reflecting the *status quo* for the distribution of gender with respect to careers or first names. These regularities are captured by machine learning along with the rest of semantics. In addition to our empirical findings concerning language, we also contribute new methods for evaluating bias in text, the Word Embedding Association Test (WEAT) and the Word Embedding Factual Association Test (WEFAT). Our results have implications not only for AI and machine learning, but also for the fields of psychology, sociology, and human ethics, since they raise the possibility that mere exposure to everyday language can account for the biases we replicate here.

Social media is distorting the representation of women in Africa. Here's what can be done about it

DECEMBER 17, 2018

Gender and Jobs in Online Image Searches

Men are overrepresented in online image search results across a majority of jobs examined; women appear lower than men in such search results for many jobs

BY ONYI LAM, BRIAN BRODERICK, STEFAN WOJCIK AND ADAM HUGHES