



CS 6474/4803 Social Computing: Bridging the Offline and the Online: Language

Munmun De Choudhury

munmund@gatech.edu

Week 5 | February 9, 2026

Start Working on and Submitting
your Class Reading Assignments for
early feedback!

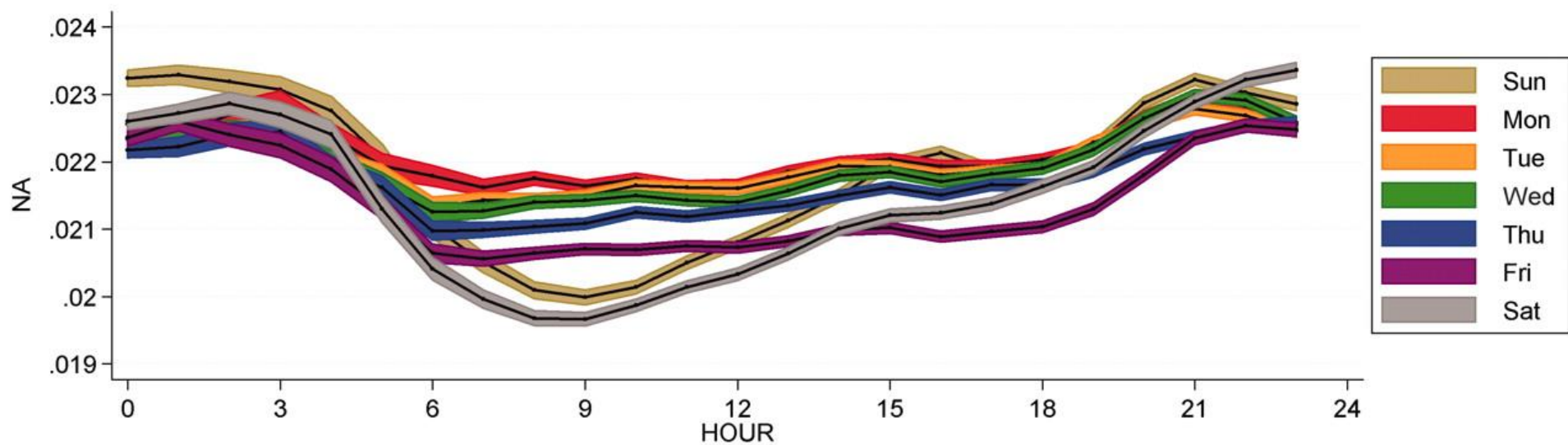
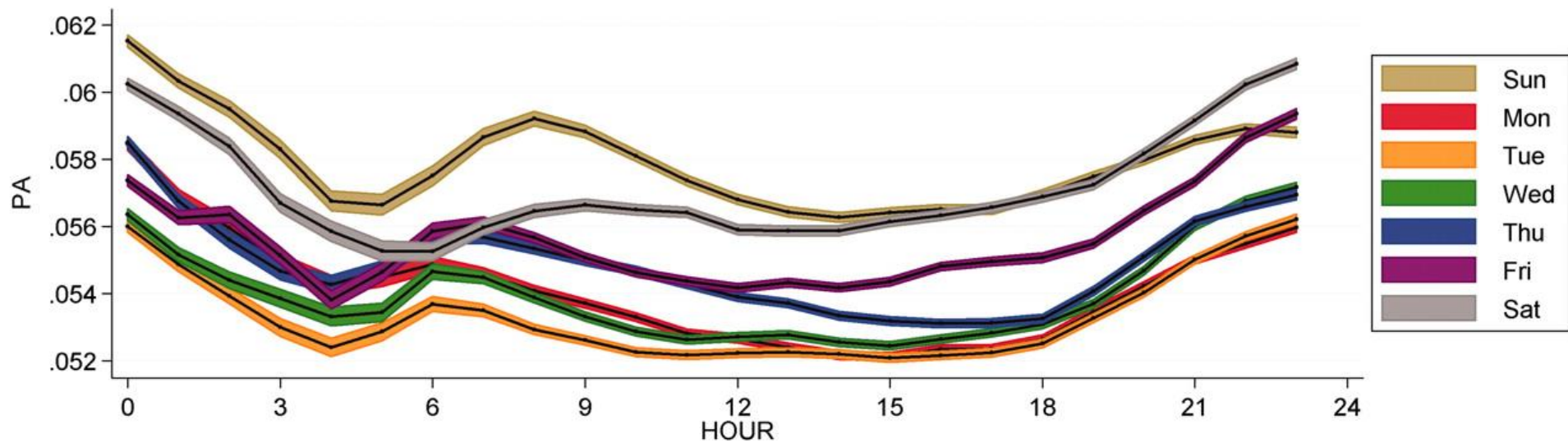


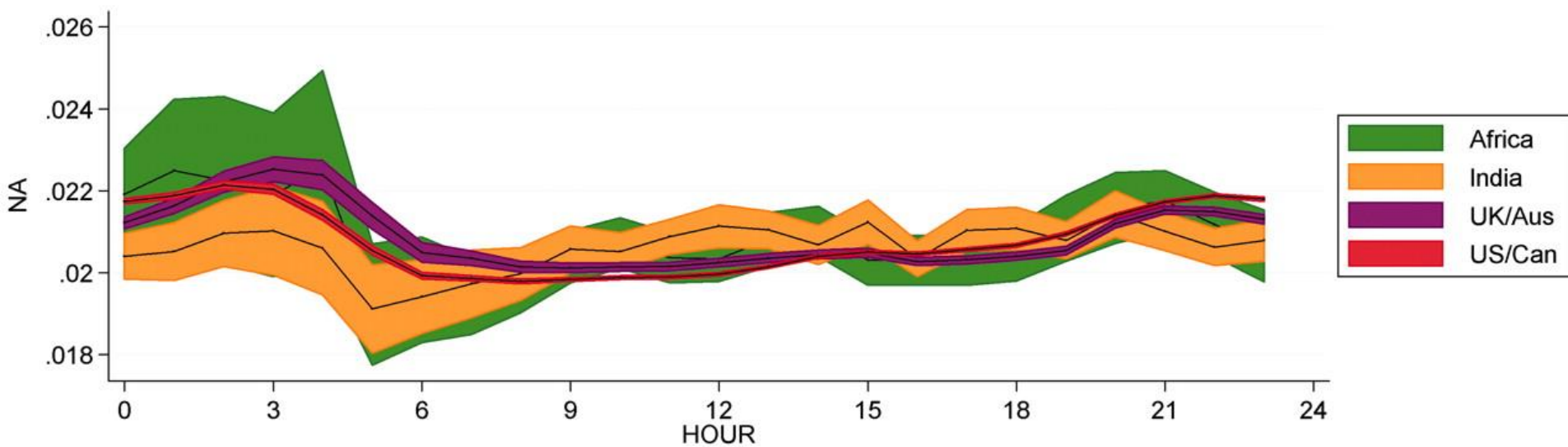
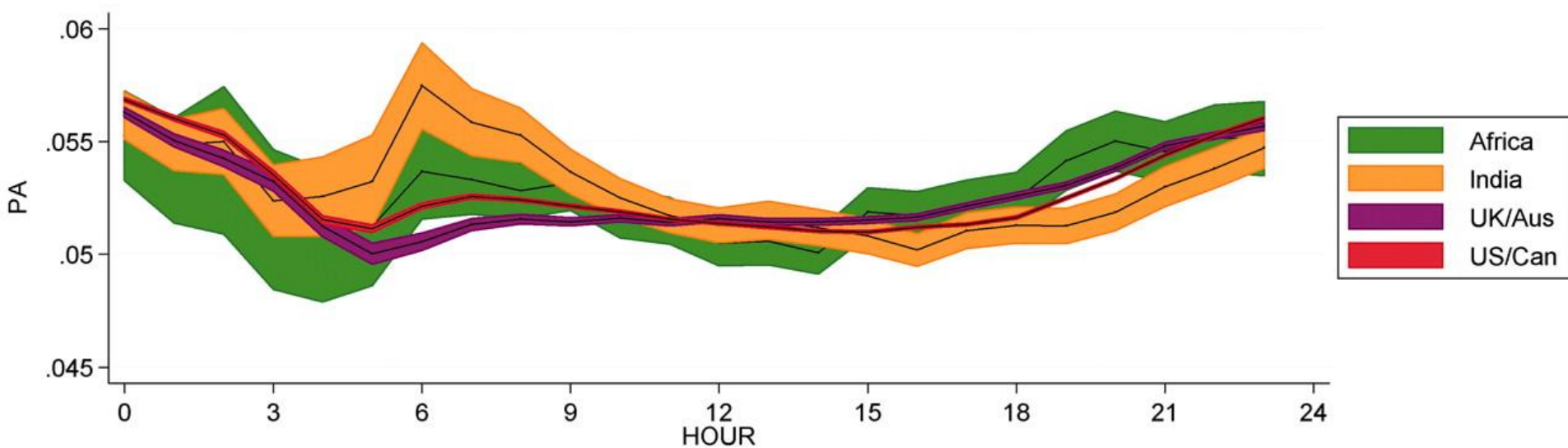
Language is the most common and reliable way for people to translate their internal thoughts and emotions into a form that others can understand. Words and language, then, are the very stuff of psychology and communication -- Tauszczik & Pennebaker

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.





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) Reddit 2) Instagram 3) Snapchat?

Class Exercise

An important aspect of studying emotion and mood with social media like Twitter is that we have no knowledge if the displayed emotion is truly the emotion experienced by the respective individuals at the moment in time when a tweet was shared. That is, when a tweet says “So happy that the weather is cooling down”, was the person really feeling “happy” at that time?

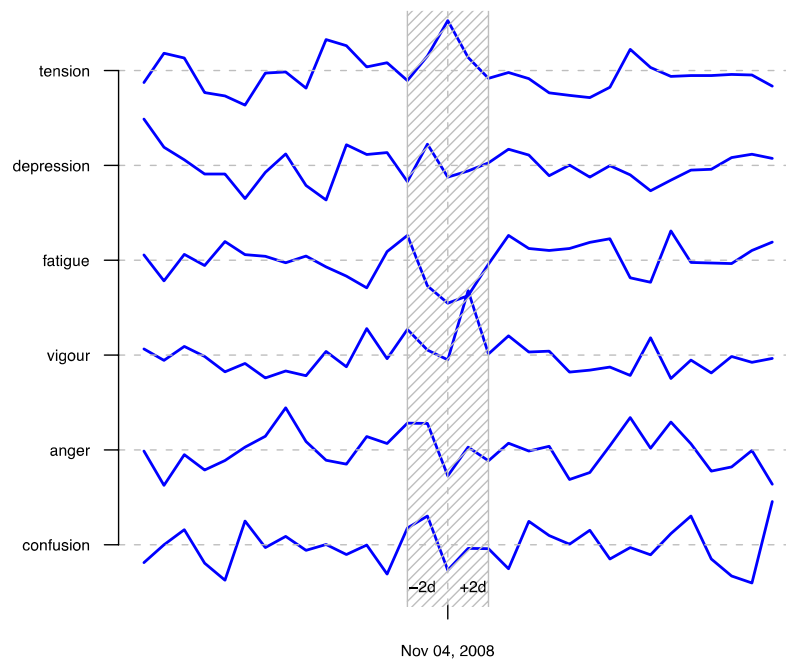
This exercise will explore your ideas around going about assessing to what extent social media emotion and real emotion are consistent, if at all. Specifically, you need to present a study design, involving data analysis, to examine this question. You need to:

1. Propose how you would measure true emotion of a person.
2. Propose how you would assess the relationship of an individual's true emotion measured in step #1 and their manifested emotion on social media.
3. What do you expect to find based on step #2? Why?

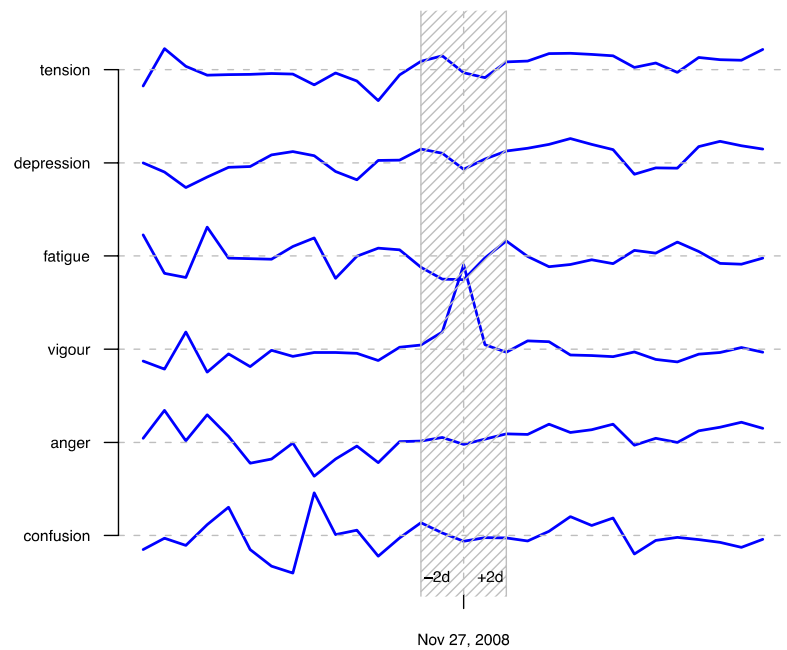
Why is measuring mood useful?
Some examples follow...

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

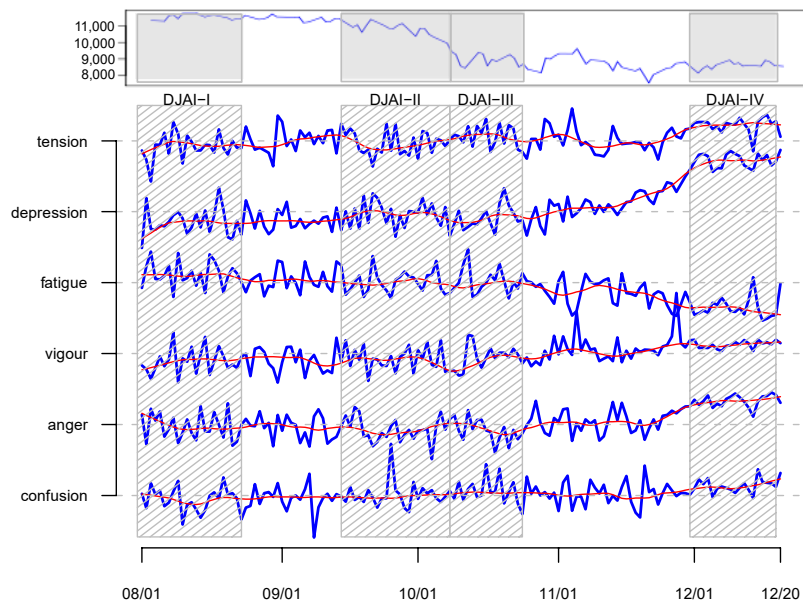
- Examine how Twitter moods reflect social, political, and economic events



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

Some illustrative examples of average happiness we obtained for individual words are:

$$h_{\text{avg}}(\text{laughter}) = 8.50,$$

$$h_{\text{avg}}(\text{food}) = 7.44,$$

$$h_{\text{avg}}(\text{reunion}) = 6.96,$$

$$h_{\text{avg}}(\text{truck}) = 5.48,$$

$$h_{\text{avg}}(\text{the}) = 4.98,$$

$$h_{\text{avg}}(\text{of}) = 4.94,$$

$$h_{\text{avg}}(\text{vanity}) = 4.30,$$

$$h_{\text{avg}}(\text{greed}) = 3.06,$$

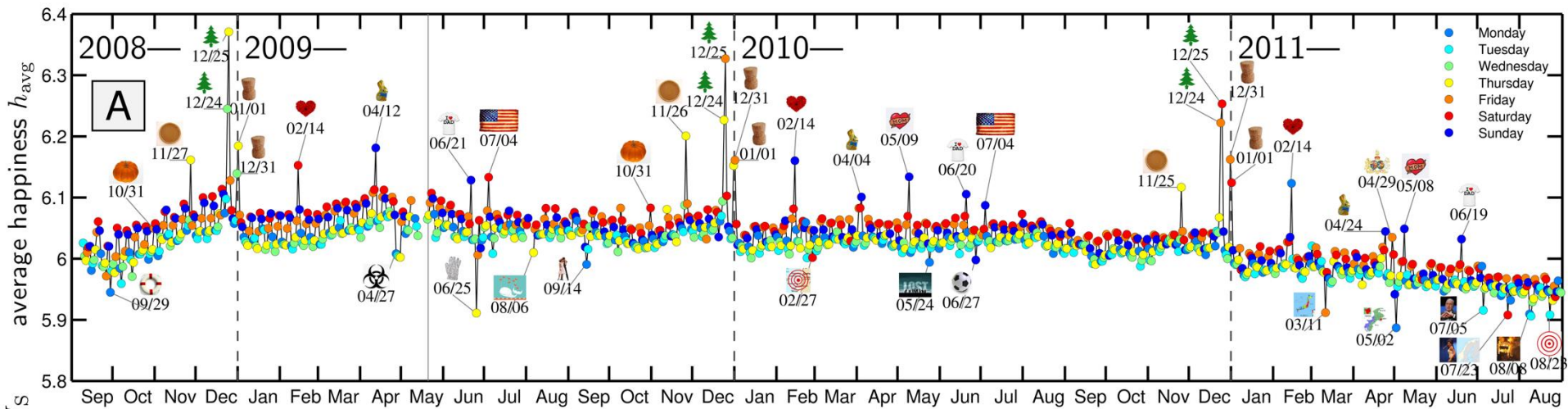
$$h_{\text{avg}}(\text{hate}) = 2.34,$$

$$h_{\text{avg}}(\text{funeral}) = 2.10,$$

$$\text{and } h_{\text{avg}}(\text{terrorist}) = 1.30.$$

where f_i is the frequency of the i th word w_i for which we have an estimate of average happiness, $h_{\text{avg}}(w_i)$, and $p_i = f_i / \sum_{j=1}^N f_j$ is the corresponding normalized frequency.

As this small sample indicates, we find the evaluations are sensible with neutral words averaging around 5.



But language based
inferences can be biased!

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

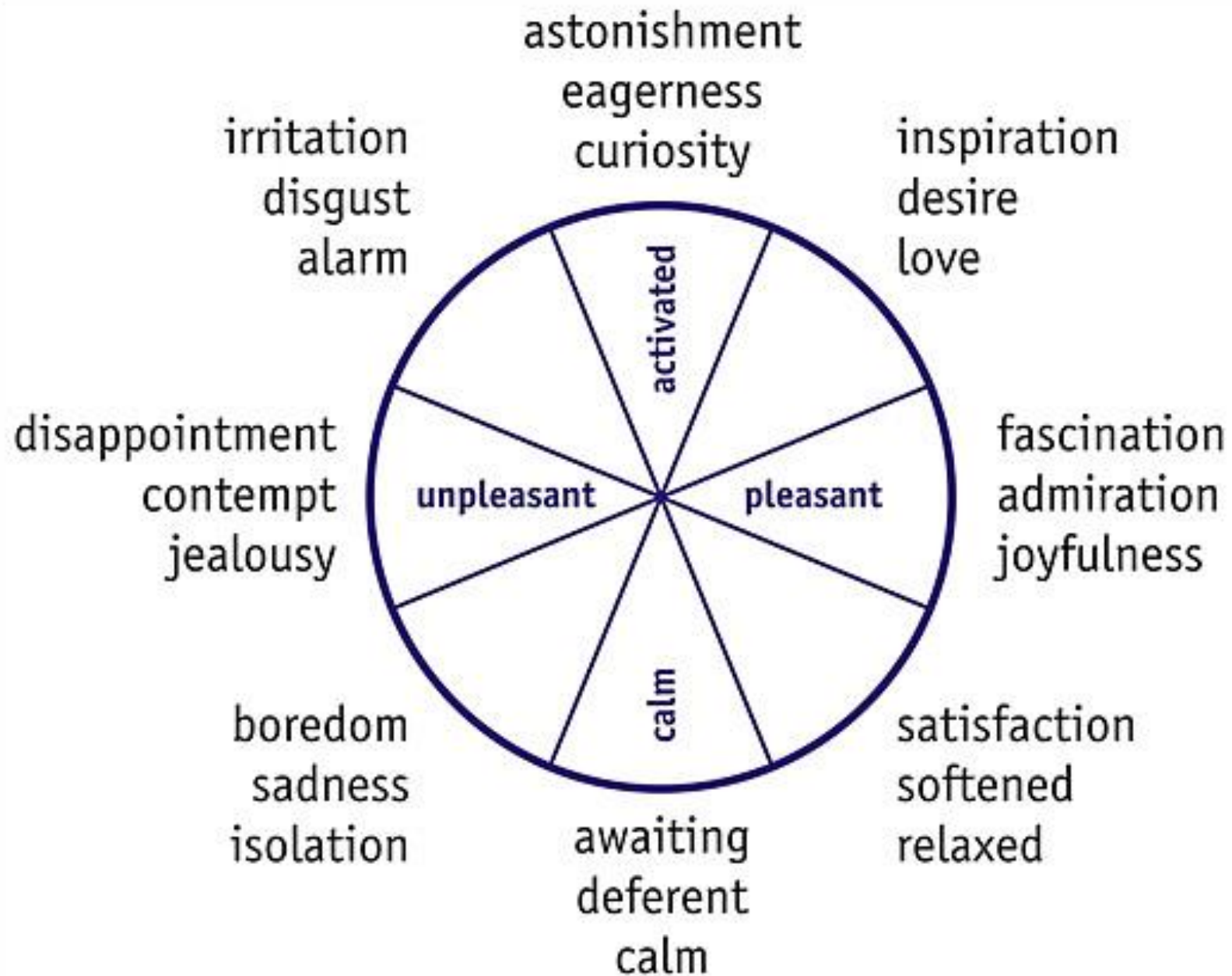
Jonathan Mellon¹ and Christopher Prosser²

Abstract

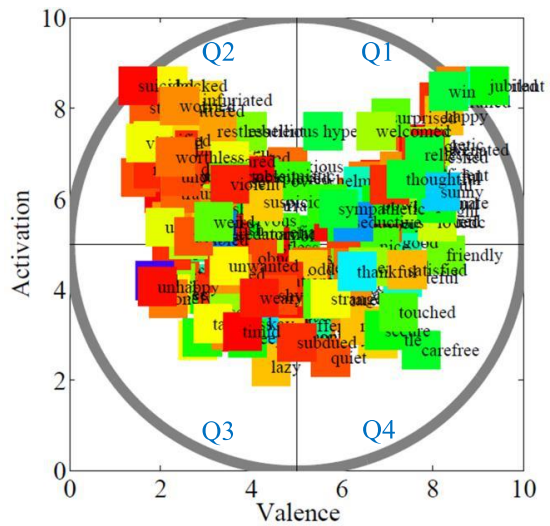
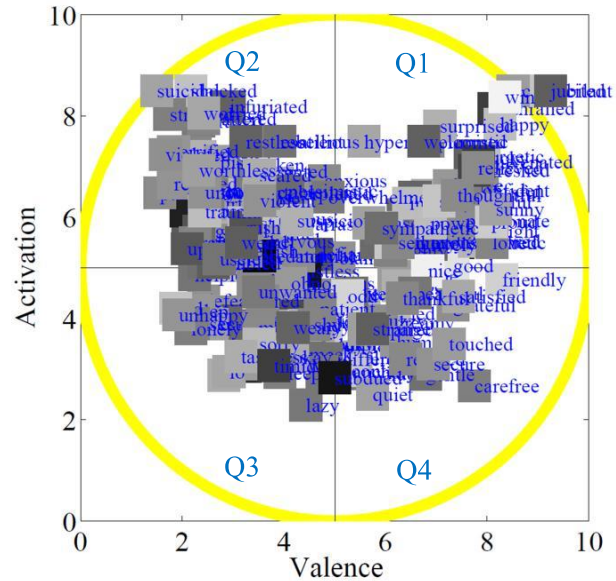
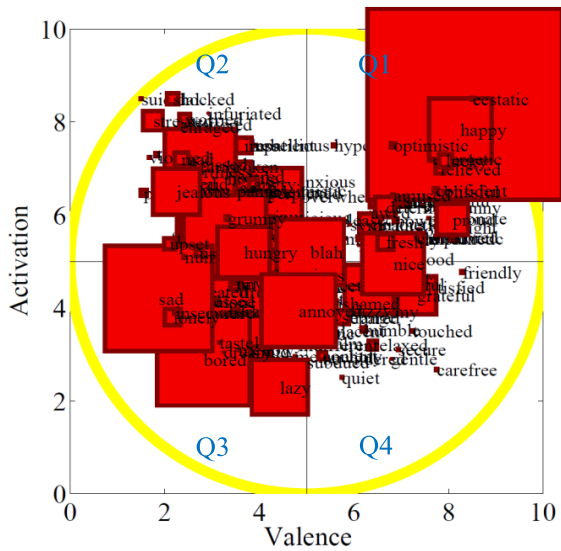
A growing social science literature has used Twitter and Facebook to study political and social phenomena including for election forecasting and tracking political conversations. This research note uses a nationally representative probability sample of the British population to examine how Twitter and 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

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

Richer linguistic
representations?



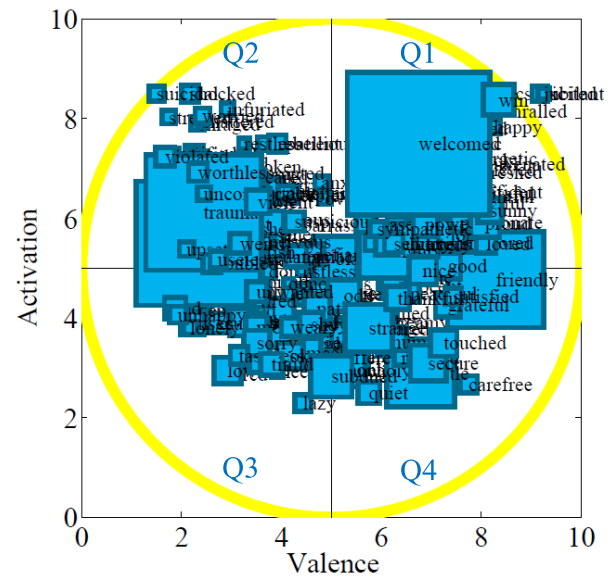
Not All Moods Are Created Equal! Exploring Human Emotional States in Social Media, by De Choudhury, Counts, and Gamon 2012



Less social –
lower followers

Most social

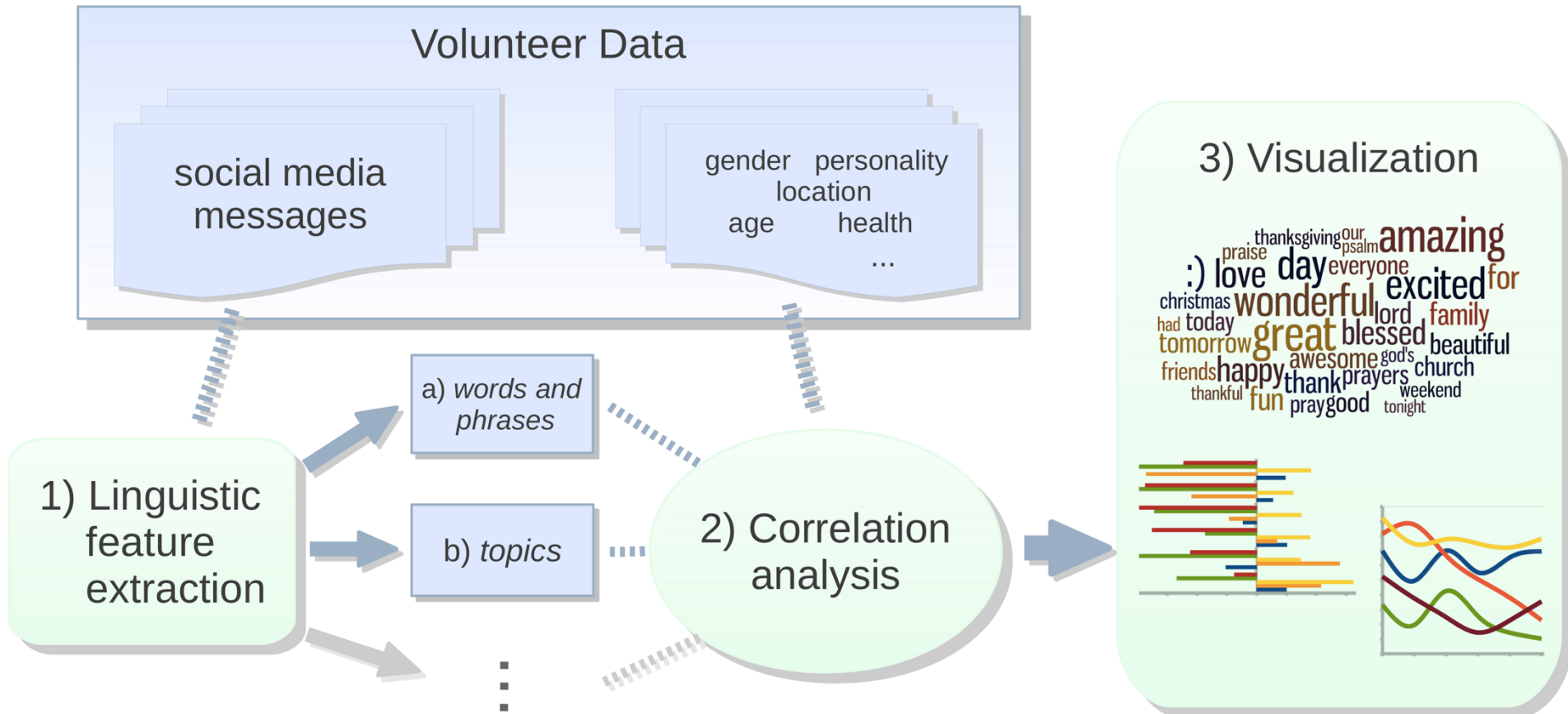
Less social –
lower followers



Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach, by Schwartz et al 2013

- 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

Schwartz et al 2013



Summary

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

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.,

Character n -grams

Quantifying Mental Health Signals in Twitter

Glen Coppersmith Mark Dredze Craig Harman

Human Language Technology Center of Excellence

Johns Hopkins University

Baltimore, MD, USA

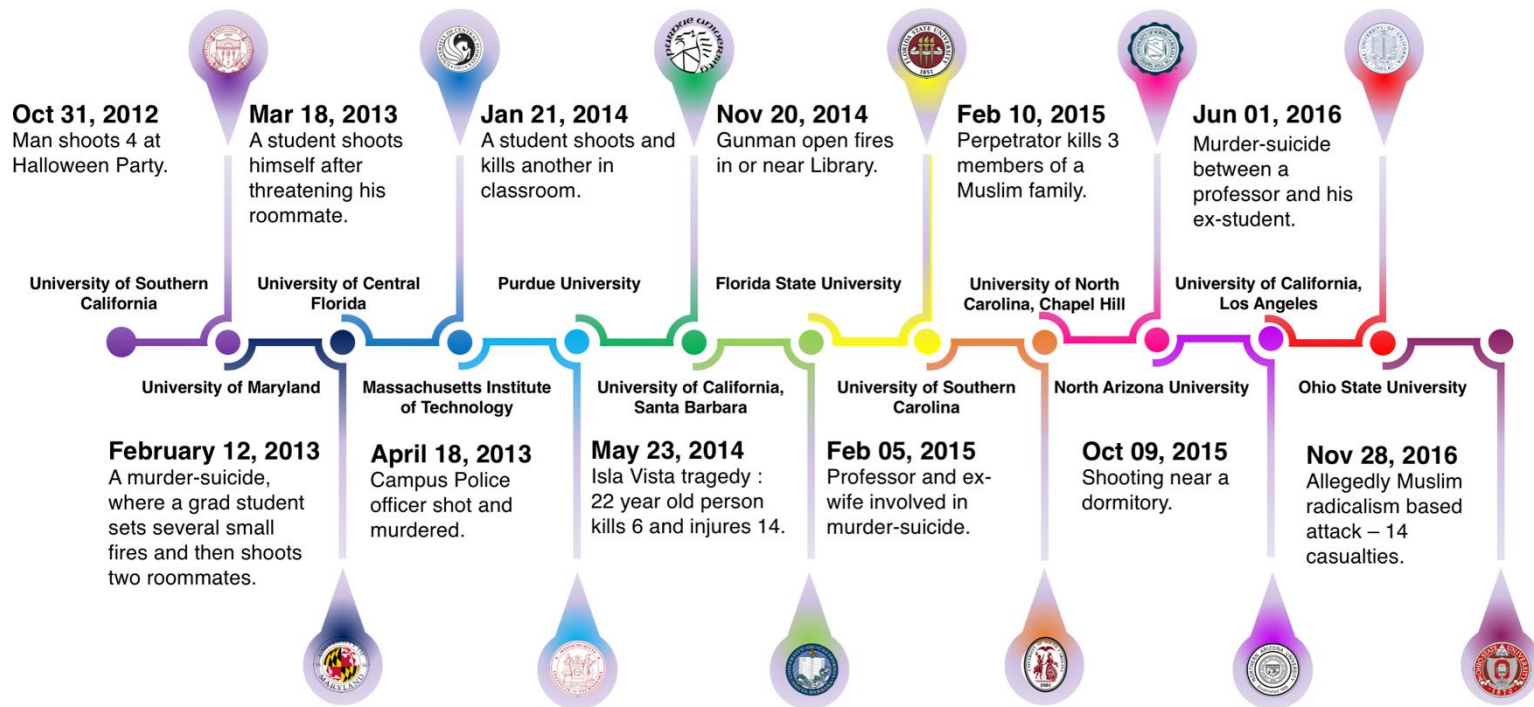
Abstract

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 (Boydston 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-

Modeling Stress with Social Media Around Incidents of Gun Violence on College Campuses

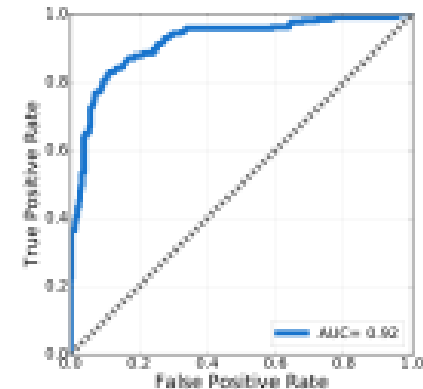
Acute Stress in College Campuses Increases Following Gun Violence Incidents



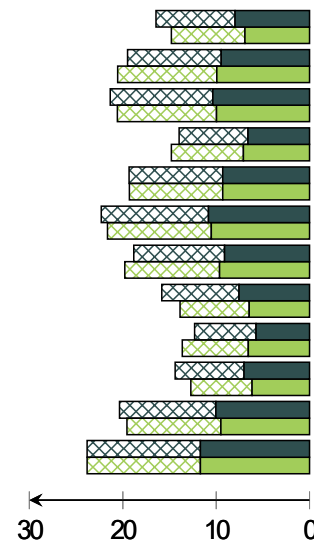
12 universities; 12 incidents over 5 years (2012-2016); 113,337 posts

Stress Classifier

Metric	mean	stdev.	median	max.
Accuracy	0.82	0.11	0.78	0.90
Precision	0.83	0.14	0.77	0.92
Recall	0.82	0.09	0.78	0.88
F1-score	0.82	0.11	0.79	0.89
ROC-AUC	0.90	0.08	0.78	0.95

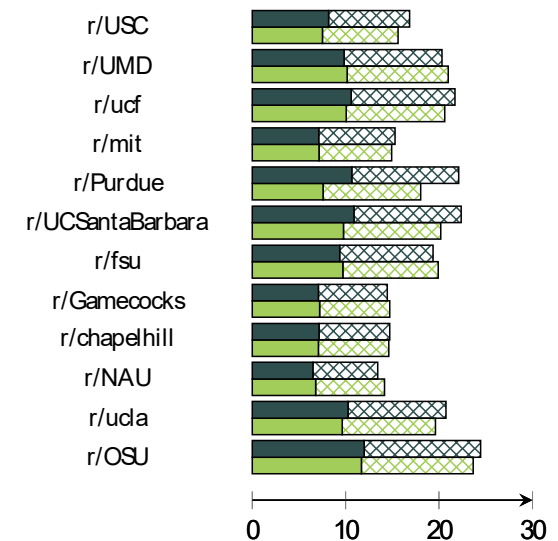


Feature	p	log(score)	Feature	p	log(score)
stress	***	9.63	thank	***	6.20
try	***	7.46	meet	***	6.17
work	***	7.20	life	***	6.07
anxiety	***	7.05	sleep	***	6.03
meditation	***	6.88	problems	***	5.98
help	***	6.81	control	***	5.95
focus	***	6.62	job	***	5.89
luck	***	6.62	good	***	5.87
breathing	***	6.44	health	***	5.87
techniques	***	6.33	week	***	5.86
feel	***	6.30	minutes	***	5.83
exercise	***	6.30	doctor	***	5.83
time	***	6.25	mental	***	5.83
play	***	6.23	relax	***	5.72
body	***	6.21	stressful	***	5.67



Before (log(number of posts))

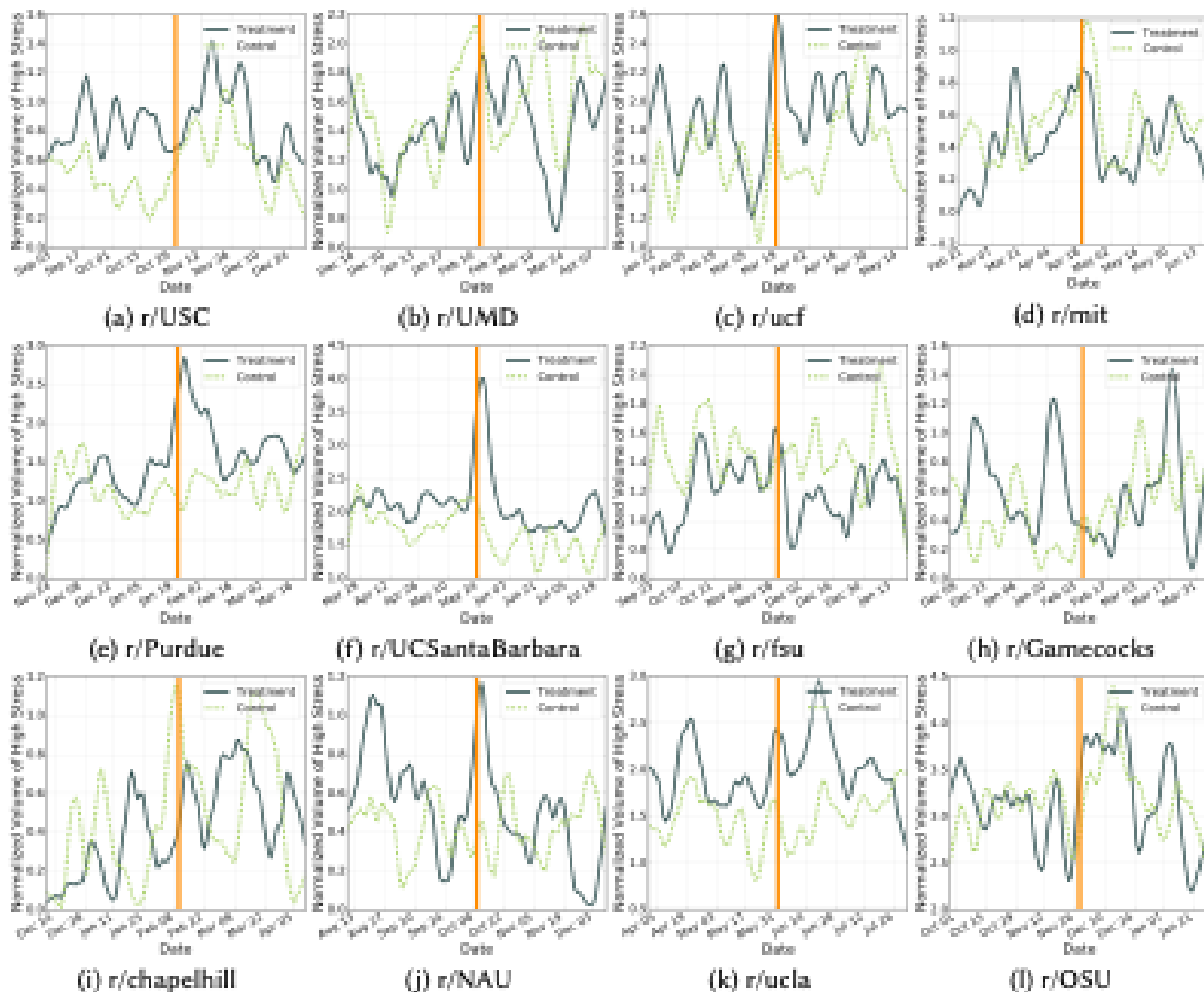
Control HS Control LS



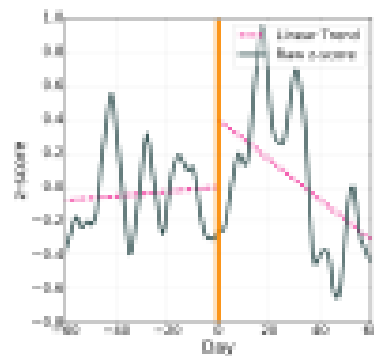
After (log(number of posts))

Treatment HS Treatment LS

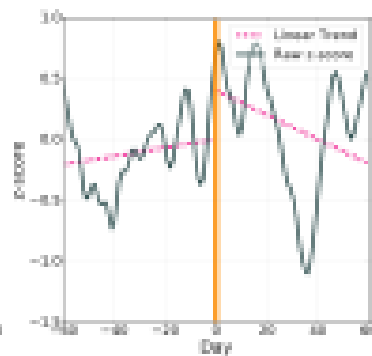
Temporal and Linguistic Patterns of Stress



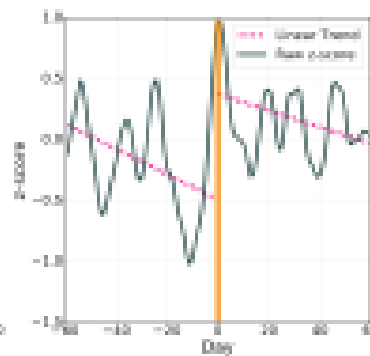
Temporal and Linguistic Patterns of Stress



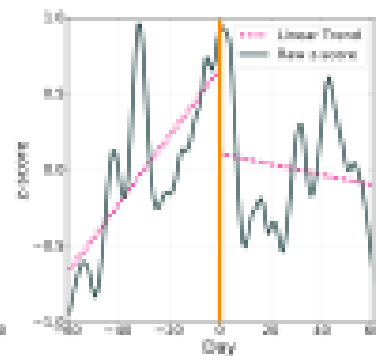
(a) r/USC



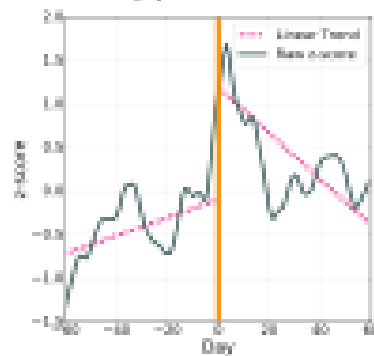
(b) r/UMD



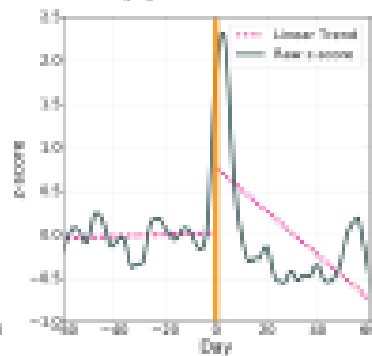
(c) r/ucf



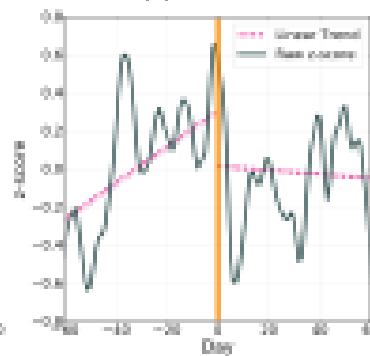
(d) r/mit



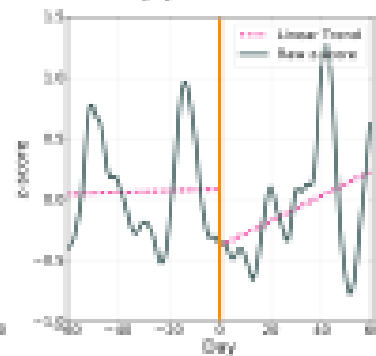
(e) r/Purdue



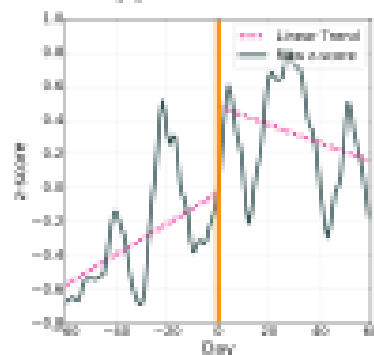
(f) r/UCSantaBarbara



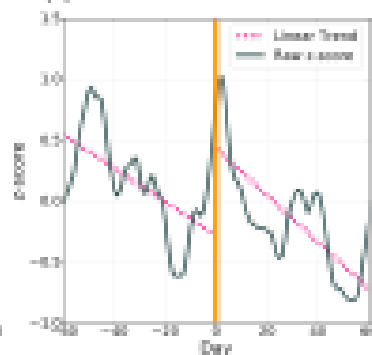
(g) r/fsu



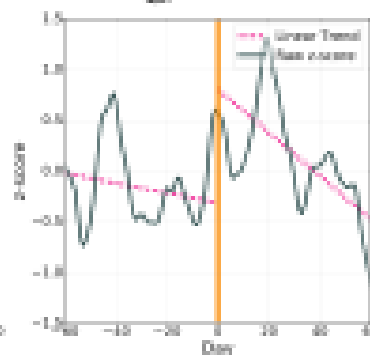
(h) r/Gamecocks



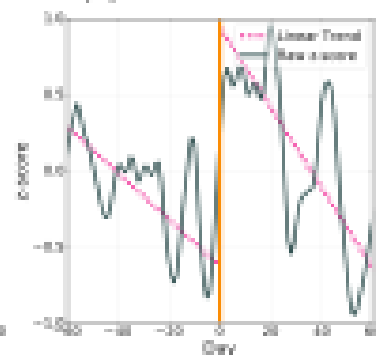
(i) r/chapelhill



(j) r/NAU

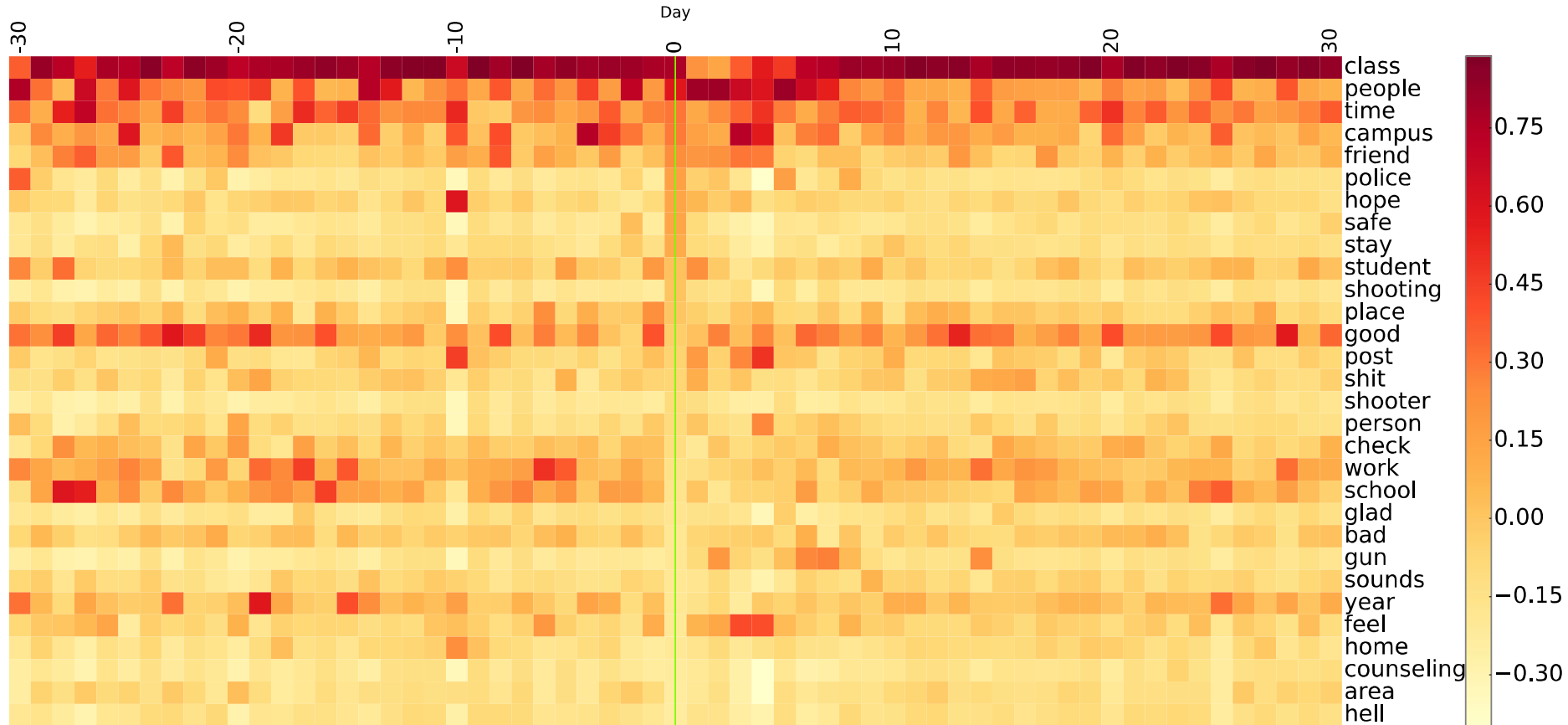


(k) r/ucla



(l) r/OSU

Temporal and Linguistic Patterns of Stress



Open Questions and Challenges

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

NEW RESEARCH IN

Physical Sciences

Social Sciences

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 massive-scale emotional contagion through social networks](#)

[Correction for Kramer et al., Experimental evidence of massive-scale emotional contagion through social networks](#)

Article

Figures & SI

Info & Metrics

PDF

Significance

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

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.

Generative AI both enables as well as exacerbates what we can learn from social media language

Semantics derived automatically from language corpora necessarily contain human biases

Aylin Caliskan¹, Joanna J. Bryson^{1,2}, and Arvind Narayanan¹

¹Princeton University

²University 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.

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

Tolga Bolukbasi¹, Kai-Wei Chang², James Zou², Venkatesh Saligrama^{1,2}, Adam Kalai²

¹Boston University, 8 Saint Mary's Street, Boston, MA

²Microsoft 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*



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