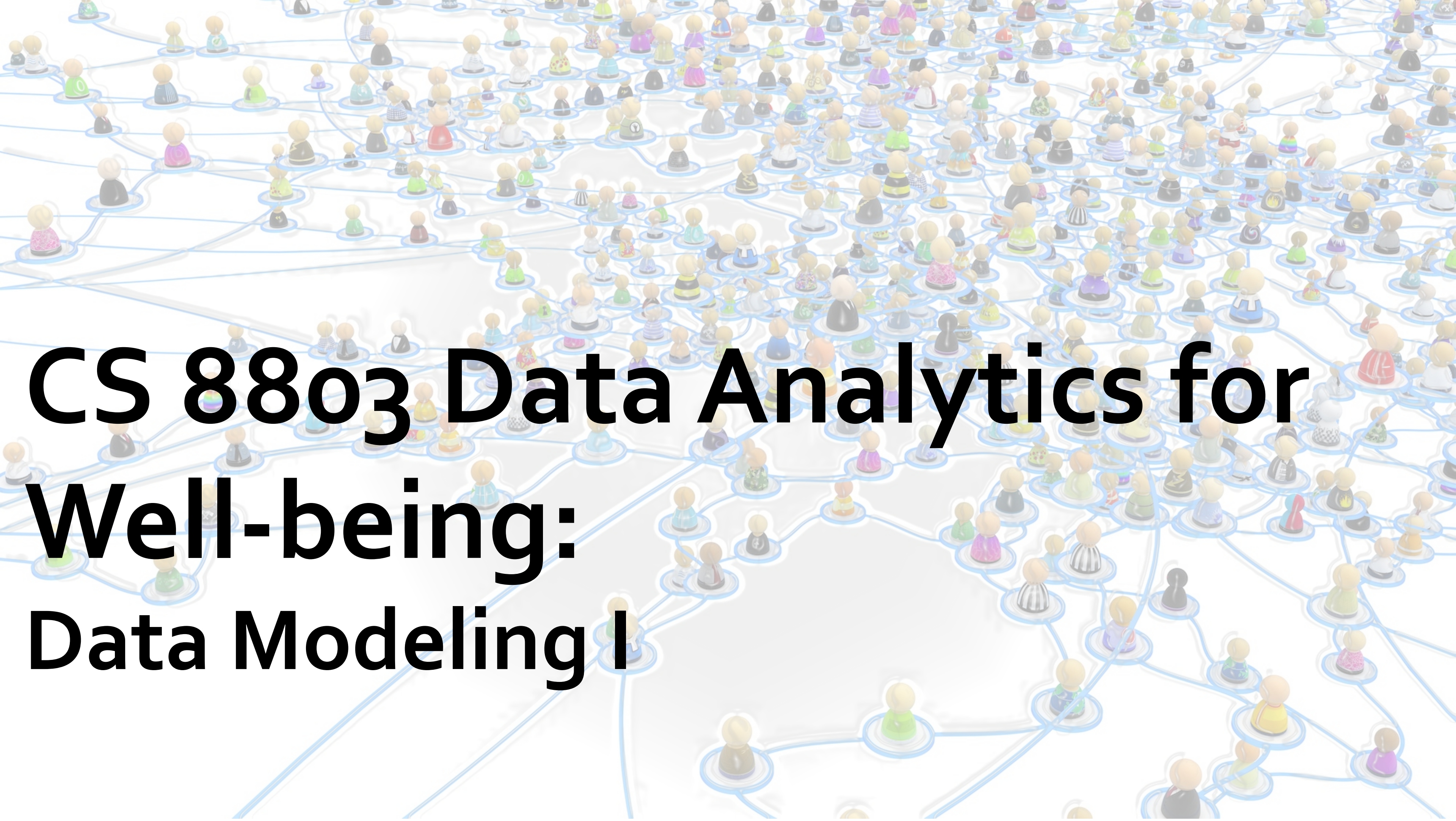


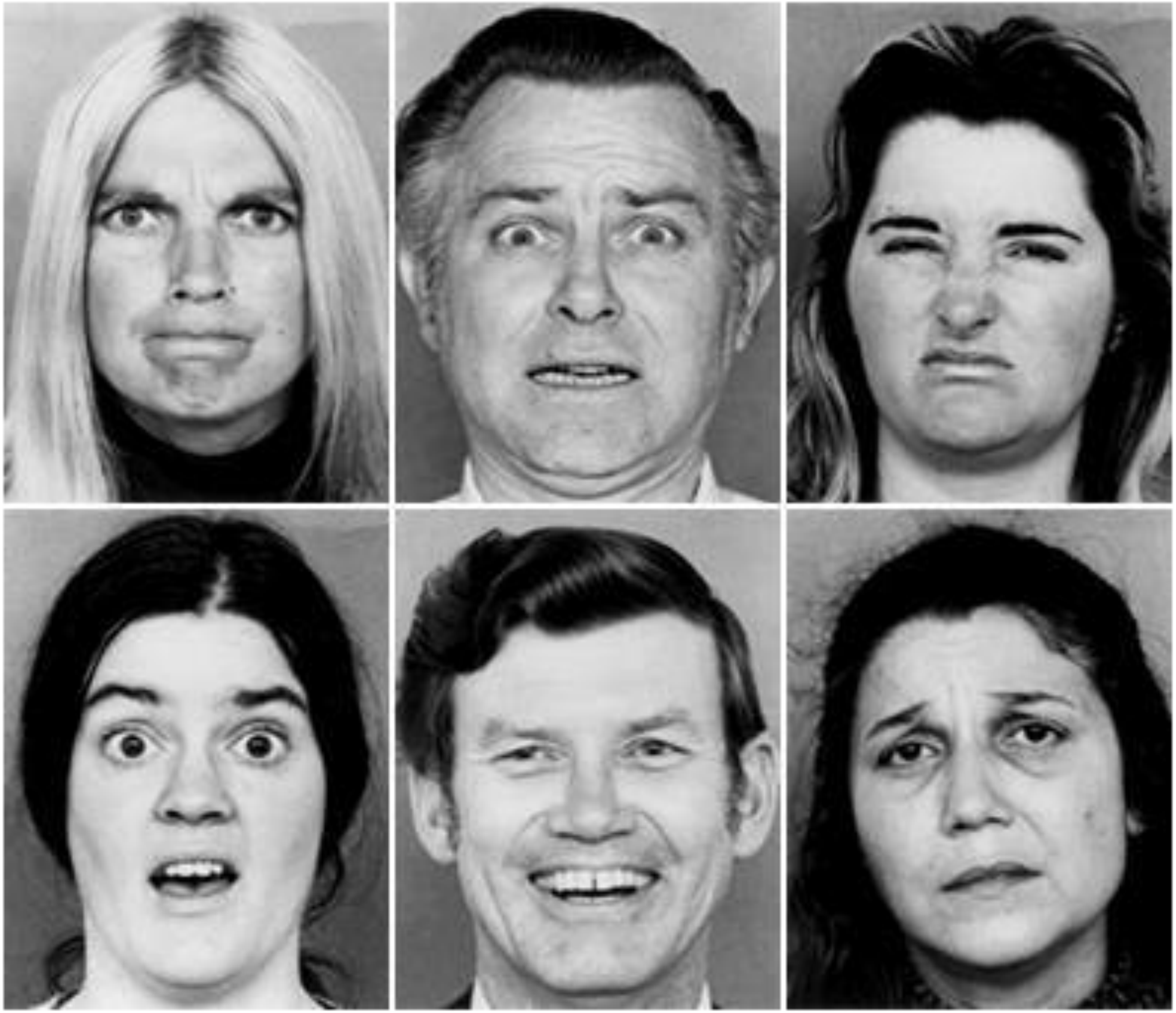
CS 8803 Data Analytics for Well-being:

Data Modeling I

Munmun De Choudhury

munmund@gatech.edu

Week 4 | February 1, 2016



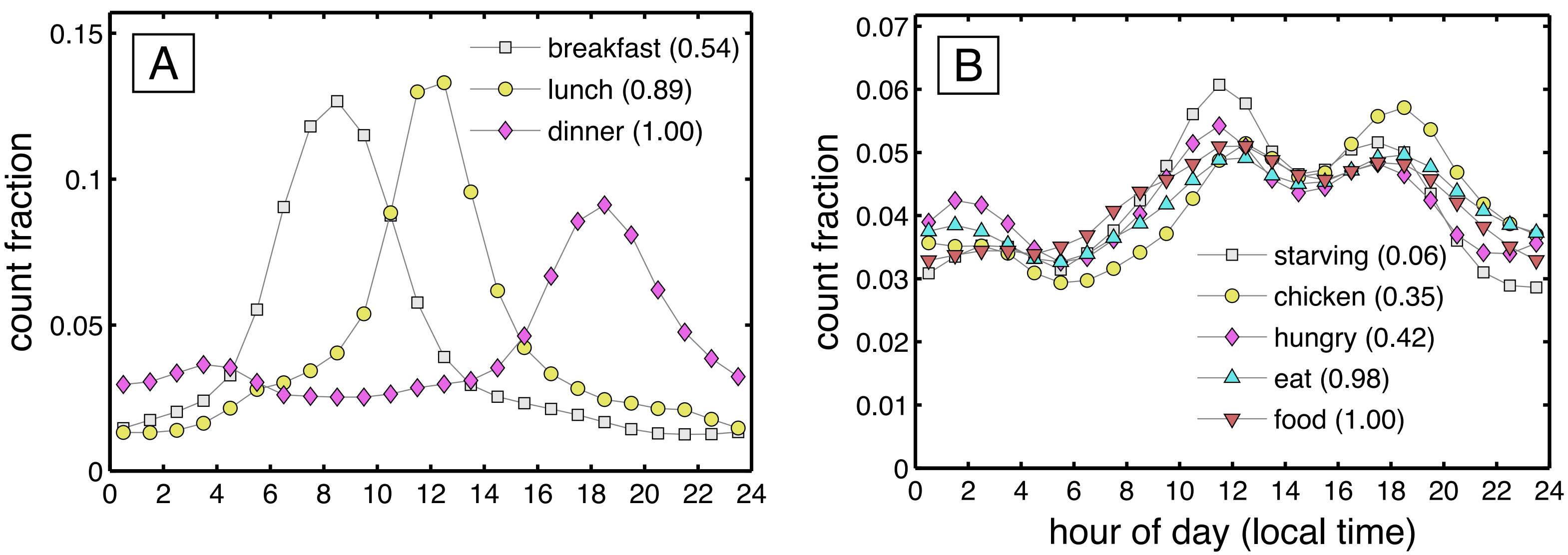
Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter

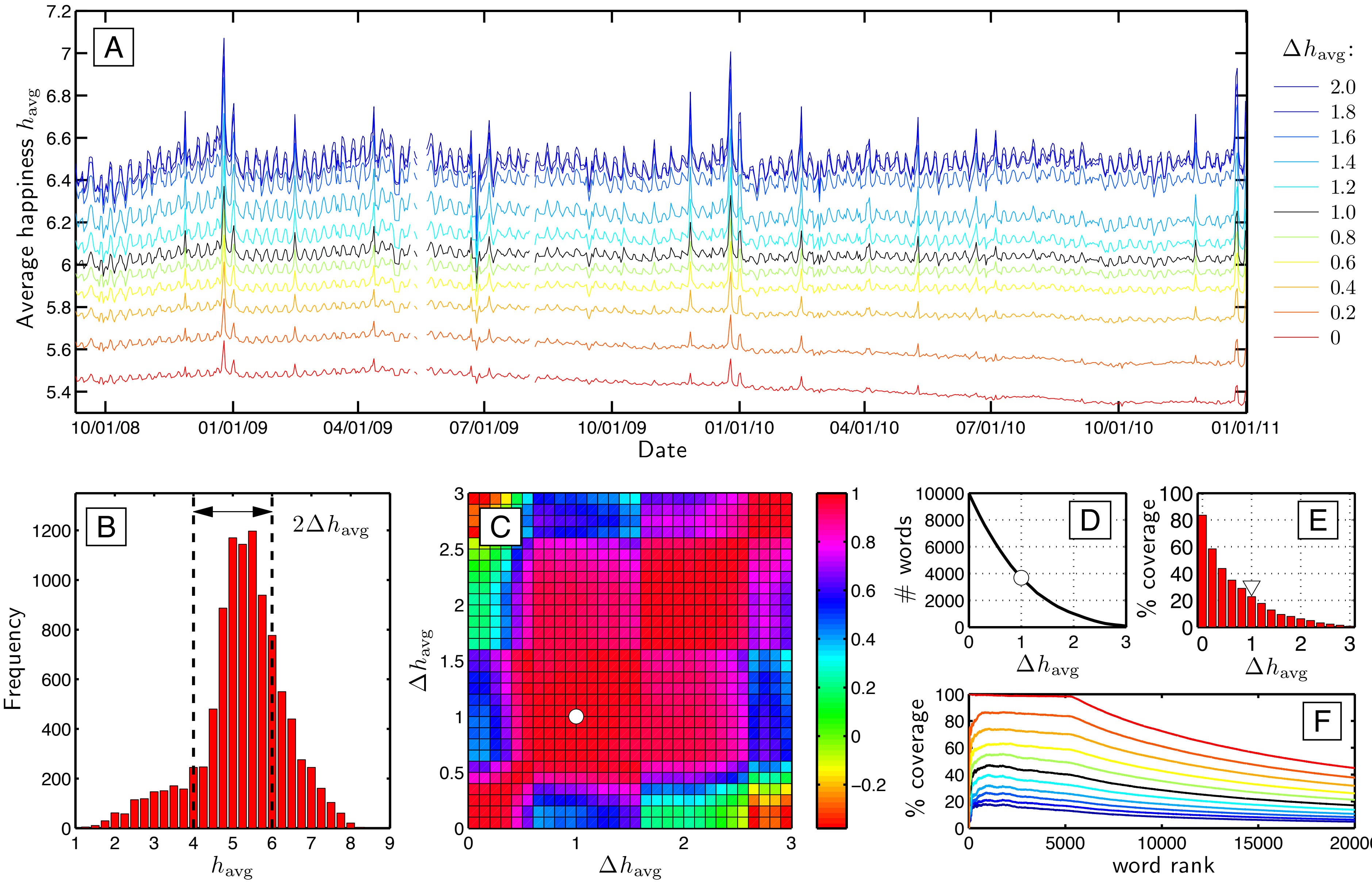
Main Idea

- The article says the following about the web: “a collective, open recording of an enormous number of transactions, interactions, and expressions, marking a clear transition in our ability to quantitatively characterize, and thereby potentially understand, previously hidden as well as novel microscale mechanisms underlying sociotechnical systems”
- Method (a hedenometer): use of word frequency distributions combined with independently assessed numerical estimates of the ‘happiness’ of over 10,000 words obtained using Amazon’s Mechanical Turk

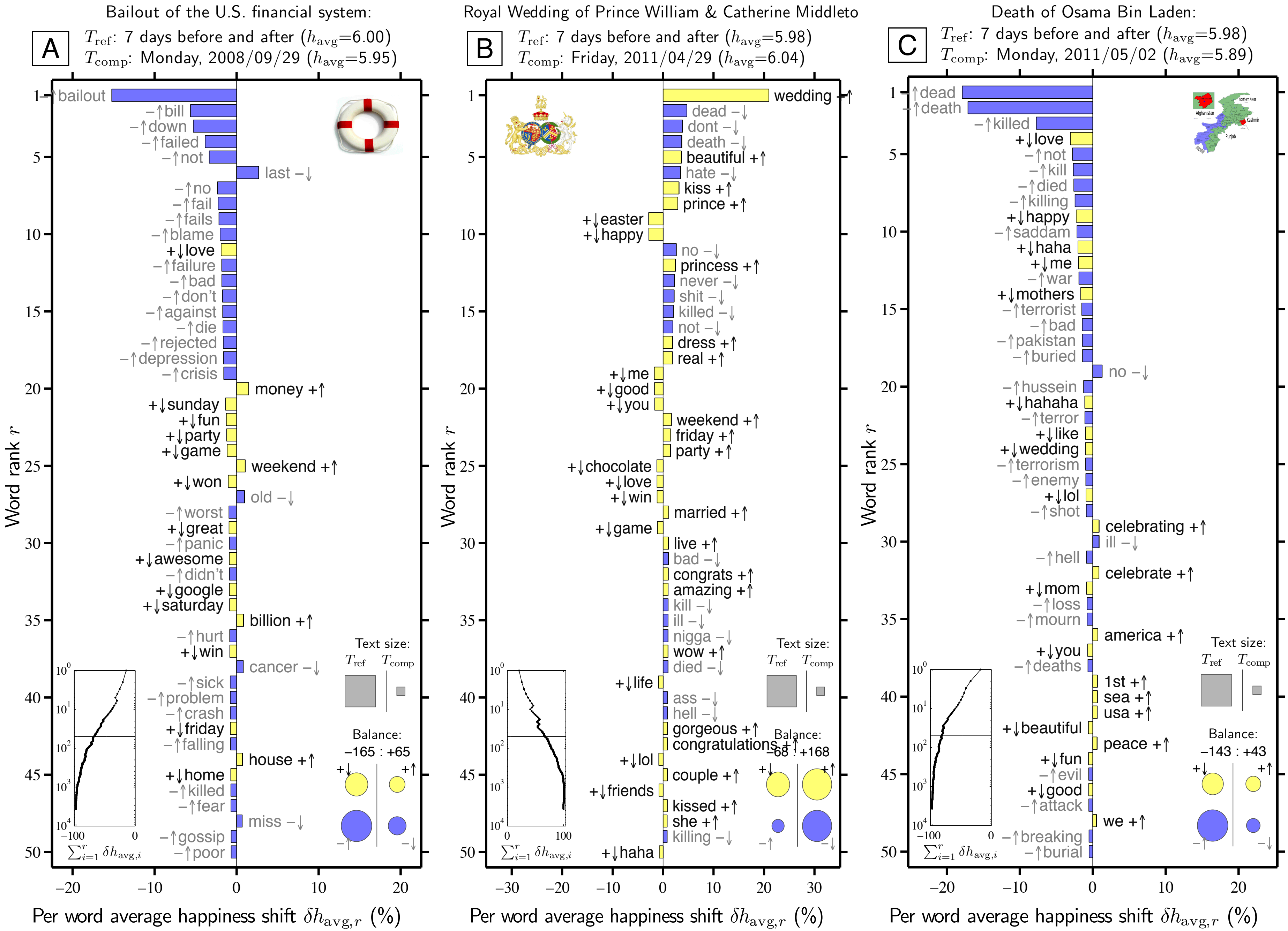
Contributions and Findings

- The paper explores happiness as a function of time, space, demographics, and network structure.
- Examine temporal variations in happiness including: the overall time series; regular cycles at the scale of days and weeks; time series for subsets of tweets containing specific keywords; and detailed comparisons between texts at the level of individual words.

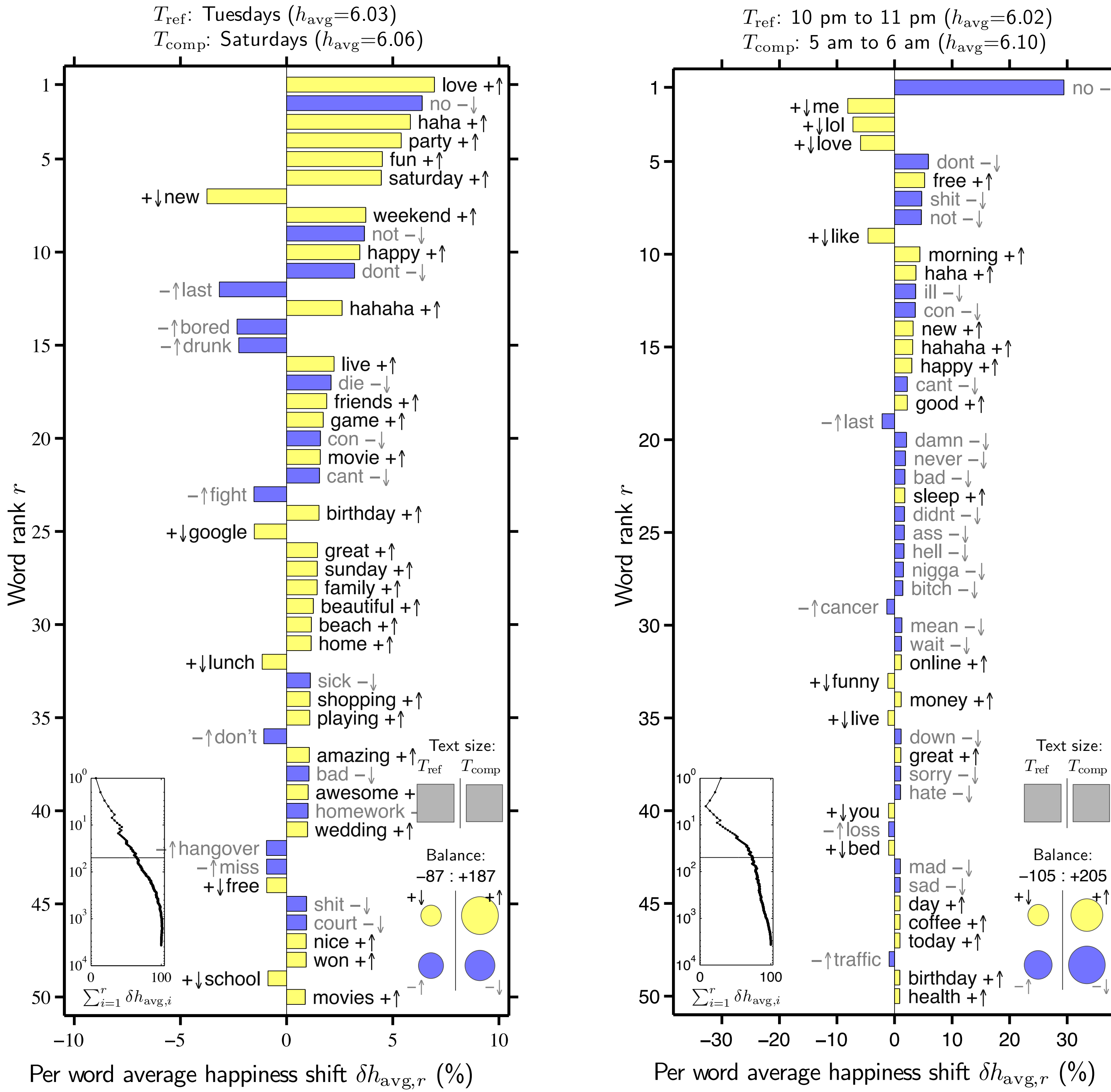




A
Average happiness h_{avg}
Date
10/01/08
01/01/09
04/01/09
07/01/09
10/01/09
01/01/10
04/01/10
07/01/10
10/01/10
01/01/11
Δh_{avg}:
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
B
Frequency
$2\Delta h_{\text{avg}}$
h_{avg}
C
Δh_{avg}
Δh_{avg}
D
words
Δh_{avg}
E
% coverage
Δh_{avg}
F
% coverage
word rank



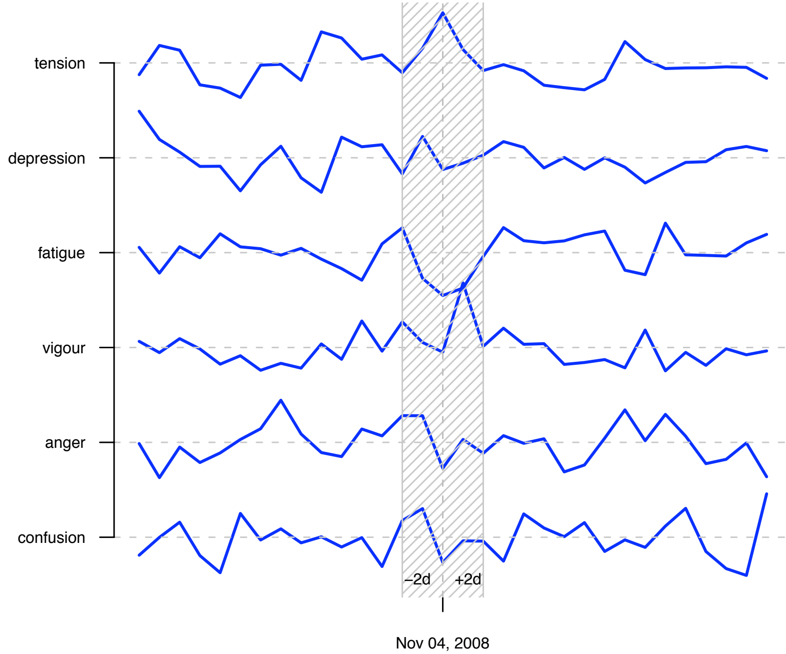
Bailout of the U.S. financial system:
A
$T_{\rm ref}$: 7 days before and after ($h_{\rm avg}$=6.00)
$T_{\rm comp}$: Monday, 2008/09/29 ($h_{\rm avg}$=5.95)
−↑bailout
−↑bill
−↑down
−↑failed
−↑not
last −↓
−↑no
−↑fail
−↑fails
−↑blame
+↓love
−↑failure
−↑bad
−↑don't
−↑against
−↑die
−↑rejected
−↑depression
−↑crisis
money +↑
+↓sunday
+↓fun
+↓party
+↓game
weekend +↑
+↓won
old −↓
−↑worst
+↓great
−↑panic
+↓awesome
−↑didn't
+↓google
+↓saturday
billion +↑
−↑hurt
+↓win
cancer −↓
−↑sick
−↑problem
−↑crash
+↓friday
−↑falling
house +↑
+↓home
−↑killed
−↑fear
miss −↓
−↑gossip
−↑poor
Text size:
$T_{\rm ref}$ $T_{\rm comp}$
Balance:
−165 : +65
Royal Wedding of Prince William & Catherine Middleto
B
$T_{\rm ref}$: 7 days before and after ($h_{\rm avg}$=5.98)
$T_{\rm comp}$: Friday, 2011/04/29 ($h_{\rm avg}$=6.04)
wedding +↑
dead −↓
dont −↓
death −↓
beautiful +↑
hate −↓
kiss +↑
prince +↑
+↓easter
+↓happy
no −↓
princess +↑
never −↓
shit −↓
killed −↓
not −↓
dress +↑
real +↑
+↓me
+↓good
+↓you
weekend +↑
friday +↑
party +↑
+↓chocolate
+↓love
+↓win
married +↑
+↓game
live +↑
bad −↓
congrats +↑
amazing +↑
kill −↓
ill −↓
nigga −↓
wow +↑
died −↓
+↓life
ass −↓
hell −↓
gorgeous +↑
congratulations +↑
+↓lol
couple +↑
+↓friends
kissed +↑
she +↑
killing −↓
+↓haha
Text size:
$T_{\rm ref}$ $T_{\rm comp}$
Balance:
−68 : +168
Death of Osama Bin Laden:
C
$T_{\rm ref}$: 7 days before and after ($h_{\rm avg}$=5.98)
$T_{\rm comp}$: Monday, 2011/05/02 ($h_{\rm avg}$=5.89)
−↑dead
−↑death
−↑killed
+↓love
−↑not
−↑kill
−↑died
−↑killing
+↓happy
−↑saddam
+↓haha
+↓me
−↑war
+↓mothers
−↑terrorist
−↑bad
−↑pakistan
−↑buried
no −↓
−↑hussein
+↓hahaha
−↑terror
+↓like
+↓wedding
−↑terrorism
−↑enemy
+↓lol
−↑shot
celebrating +↑
ill −↓
−↑hell
celebrate +↑
+↓mom
−↑loss
−↑mourn
america +↑
+↓you
−↑deaths
1st +↑
sea +↑
usa +↑
+↓beautiful
peace +↑
+↓fun
−↑evil
+↓good
−↑attack
we +↑
−↑breaking
−↑burial
Text size:
$T_{\rm ref}$ $T_{\rm comp}$
Balance:
−143 : +43
Word rank r
Per word average happiness shift $\delta h_{{\rm avg},r}$ (%)
$\sum_{i=1}^{r} \delta h_{{\rm avg},i}$



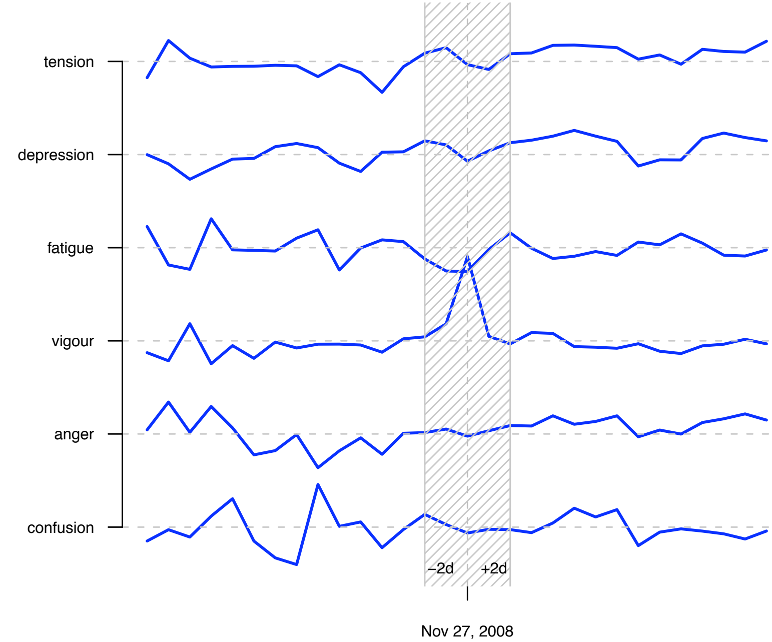
$T_{\rm ref}$: Tuesdays ($h_{\rm avg}$=6.03)
$T_{\rm comp}$: Saturdays ($h_{\rm avg}$=6.06)
love +↑
no −↓
haha +↑
party +↑
fun +↑
saturday +↑
+↓new
weekend +↑
not −↓
happy +↑
dont −↓
−↑last
hahaha +↑
−↑bored
−↑drunk
live +↑
die −↓
friends +↑
game +↑
con −↓
movie +↑
cant −↓
−↑fight
birthday +↑
+↓google
great +↑
sunday +↑
family +↑
beautiful +↑
beach +↑
home +↑
+↓lunch
sick −↓
shopping +↑
playing +↑
−↑don't
amazing +↑
bad −↓
awesome +
homework
wedding +↑
−↑hangover
−↑miss
+↓free
shit −↓
court −↓
nice +↑
won +↑
+↓school
movies +↑
Text size:
$T_{\rm ref}$
$T_{\rm comp}$
Balance:
−87 : +187
Word rank r
Per word average happiness shift $\delta h_{{\rm avg},r}$ (%)
$\sum_{i=1}^{r} \delta h_{{\rm avg},i}$
$T_{\rm ref}$: 10 pm to 11 pm ($h_{\rm avg}$=6.02)
$T_{\rm comp}$: 5 am to 6 am ($h_{\rm avg}$=6.10)
no −↓
+↓me
+↓lol
+↓love
dont −↓
free +↑
shit −↓
not −↓
+↓like
morning +↑
haha +↑
ill −↓
con −↓
new +↑
hahaha +↑
happy +↑
cant −↓
good +↑
−↑last
damn −↓
never −↓
bad −↓
sleep +↑
didnt −↓
ass −↓
hell −↓
nigga −↓
bitch −↓
−↑cancer
mean −↓
wait −↓
online +↑
+↓funny
money +↑
+↓live
down −↓
great +↑
sorry −↓
hate −↓
+↓you
−↑loss
+↓bed
mad −↓
sad −↓
day +↑
coffee +↑
today +↑
−↑traffic
birthday +↑
health +↑
Text size:
$T_{\rm ref}$
$T_{\rm comp}$
Balance:
−105 : +205
Word rank r
Per word average happiness shift $\delta h_{{\rm avg},r}$ (%)
$\sum_{i=1}^{r} \delta h_{{\rm avg},i}$

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

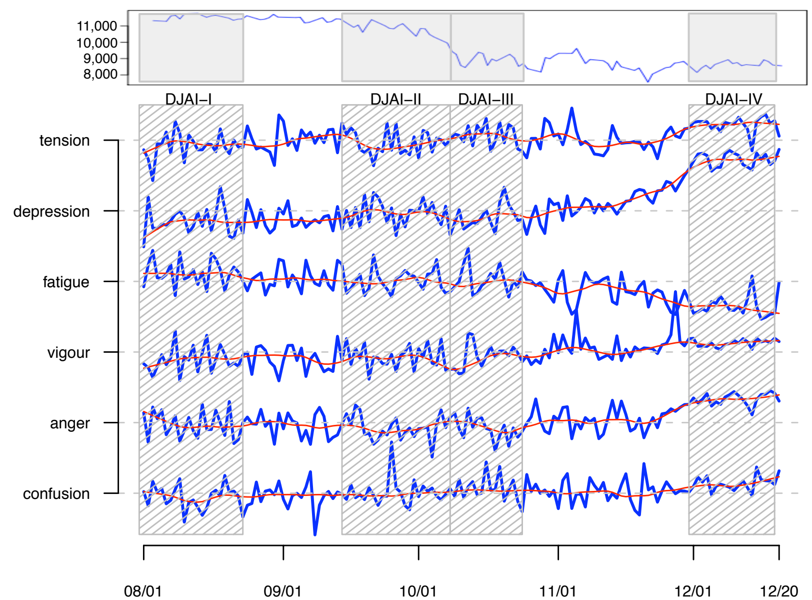
- Examine how Twitter moods reflect social, political, and economic events
- Use POMS (profile of mood states) for detecting mood-indicative twitter posts.
  - POMS dimensions: tension, depression, anger, vigor, fatigue and confusion
- Investigate how a six vector representation of moods deviates during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



2008 Presidential elections

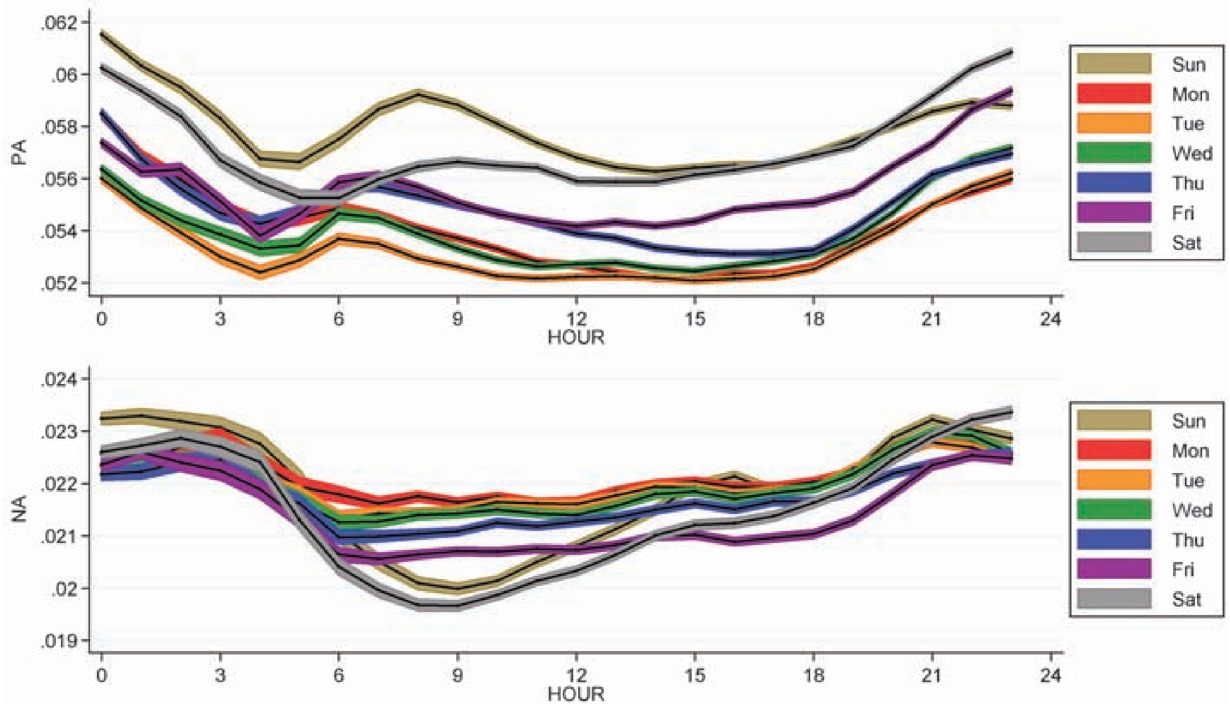


Thanksgiving

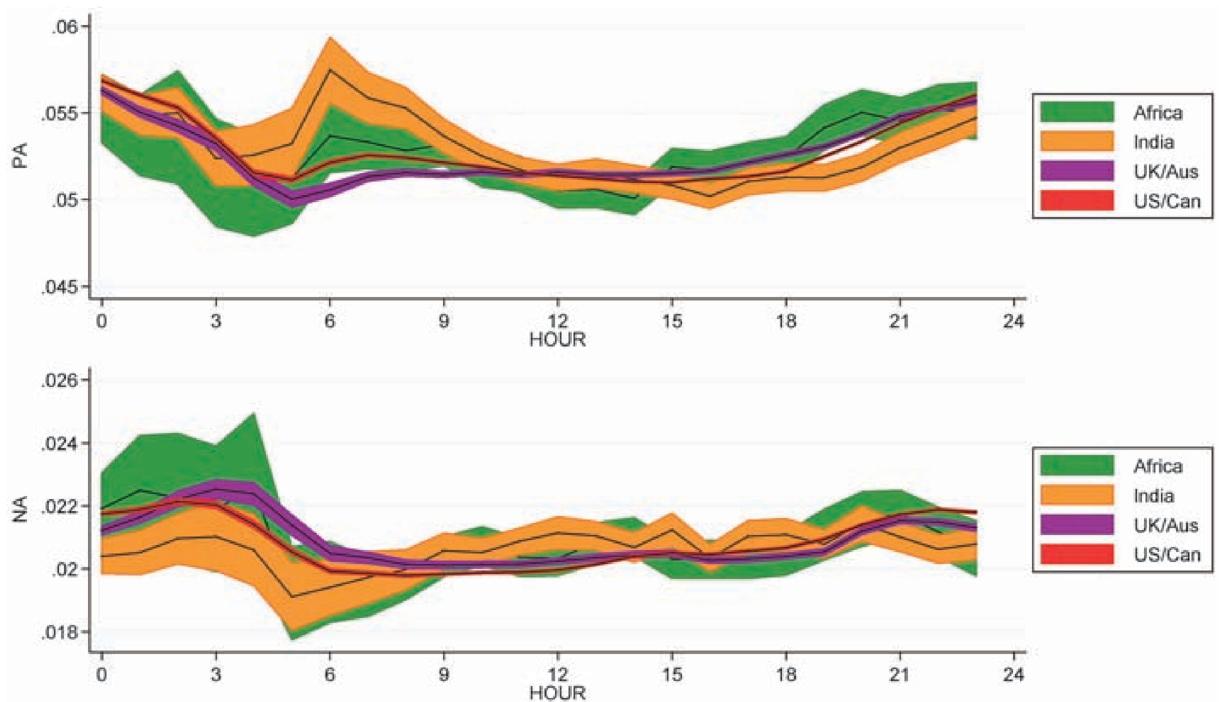


Diurnal and Seasonal Mood Vary with Work, Sleep, and Day length Across Diverse Cultures

- One of the early works examining relationship between social media mood and behavior and psychological theories.
- 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.

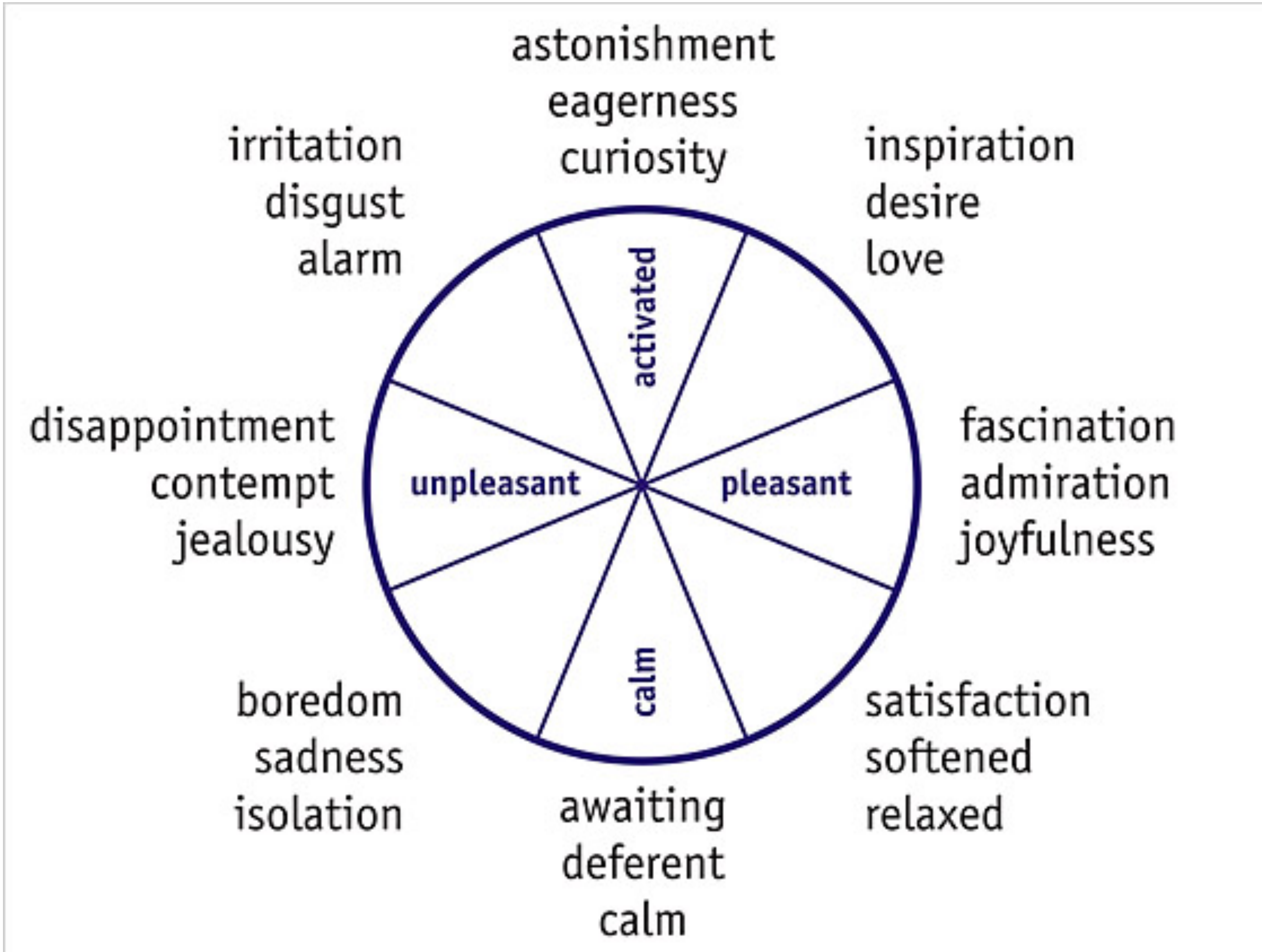


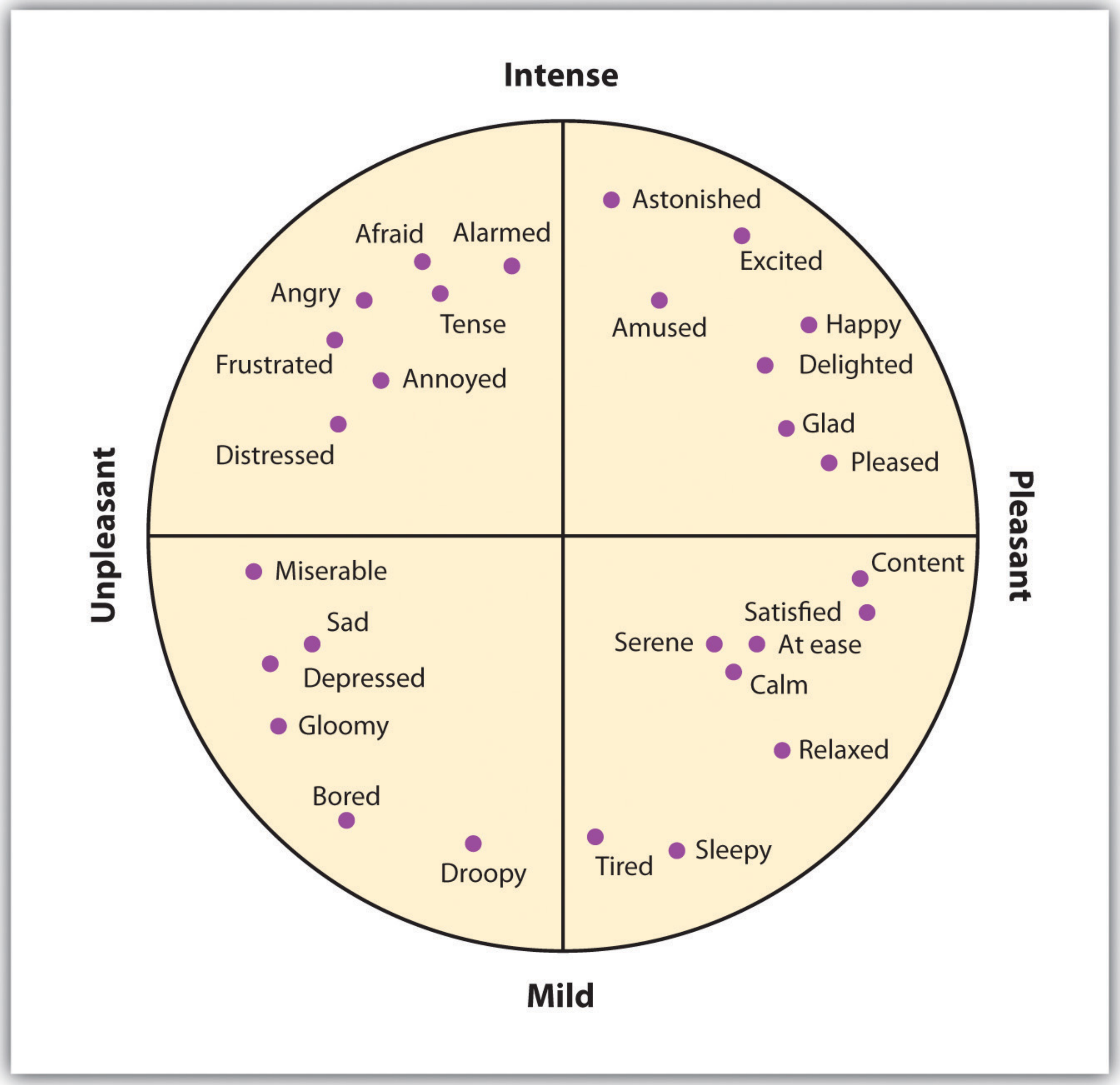
PA
.062
.06
.058
.056
.054
.052
HOUR
0
3
6
9
12
15
18
21
24
Sun
Mon
Tue
Wed
Thu
Fri
Sat
NA
.024
.023
.022
.021
.02
.019
HOUR
Sun
Mon
Tue
Wed
Thu
Fri
Sat



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

Is measurement of positive and negative emotion sufficient for assessing the well-being of populations?


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



Intense
Unpleasant
Pleasant
Mild
Afraid
Alarmed
Angry
Tense
Frustrated
Annoyed
Distressed
Astonished
Excited
Amused
Happy
Delighted
Glad
Pleased
Miserable
Sad
Depressed
Gloomy
Bored
Droopy
Content
Satisfied
Serene
At ease
Calm
Relaxed
Tired
Sleepy

The hedenometer algorithm uses ratings from Amazon’s Mechanical Turk on words obtained from music lyrics, Twitter, NY Times and Google Books. Are there limitations to this rating gathering approach?

Twitter is used by millions, but could it also have bias?

Dictionary approach of mood detection: what is its limitation?

True emotion versus displayed emotion on social media: how would you tackle this issue?

People use social media for all kinds of reasons and purposes. Would that affect the moods they express?

Would “self-presentation”, “social comparison” or identity impact the kinds of moods shared?

Can social media manifested emotion have a cultural, demographic, or geographical bias?

Could the moods of certain Twitter users be more “important” than others? (Hint: influencers and contagion)

What are some of the other aspects, not considered in the papers, they may impact mood? (Hint: Aristotle said: "man is a social animal")

One possible application is to study Twitter moods during important events, and how they impact each other. However can public displays of mood from others impact our opinions?