



Computational Social Science

Natural Language Processing: Jordan
Boyd-Graber
University of Maryland

INTRODUCTION

What is Social Science?

Study of Society and
Relationships Among Individuals

- Economics
- Sociology
- Psychology
- Political Science

Connection to NLP?

What is Social Science?

Study of Society and Relationships Among Individuals

- Economics
- Sociology
- Psychology
- Political Science

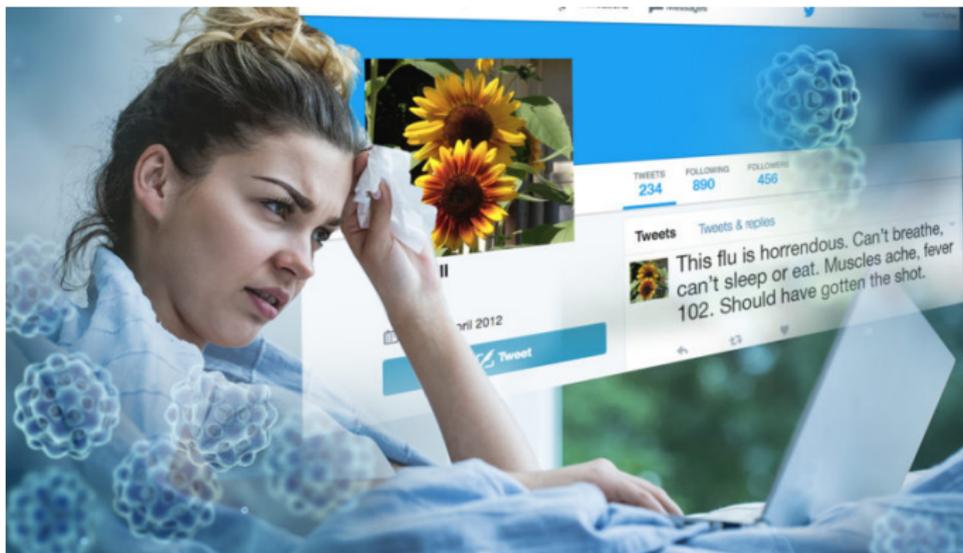
Connection to NLP? Many relationships happen through or are described by text!

Alternatives / Complements to Traditional Social Science

k.	JUN 21-25	B+	Suffolk University	500 LV	Wolf	49%	36%	Wagner	Wolf +13
d. 2*	JUN 23-24	B	YouGov	1,000 A	Approve	45%	49%	Disapprove	Disapprove +4
	JUN 22-24	B	YouGov	1,000 A	Approve	39%	50%	Disapprove	Disapprove +11
	JUN 20-24	C+	Rasmussen Reports/Pulse Opinion Research	1,500 LV	Approve	46%	52%	Disapprove	Disapprove +6
	JUN 18-24	B	Gallup	1,500 A	Approve	41%	55%	Disapprove	Disapprove +14
	JUN 20-21	B	YouGov	1,000 A	Approve	38%	51%	Disapprove	Disapprove +13
	JUN 16-19	B	Public Opinion Strategies	800 A	Approve	41%	47%	Disapprove	Disapprove +6
	JUN 22-24	B-	Morning Consult	1,989 RV	Democrat	44%	37%	Republican	Democrat +7
J-11	JUN 22-25	A+	Monmouth University	339 LV	Sherrill	45%	40%	Webber	Sherrill +5
J-11	JUN 22-25	A+	Monmouth University	406 V	Sherrill	40%	38%	Webber	Sherrill +2
k.	JUN 21-25	B+	Suffolk University	500 LV	Casey	47%	32%	Barletta	Casey +15
a.	JUN 21-25	A-	Quinnipiac University	1,153 RV	Kaine	54%	36%	Stewart	Kaine +18

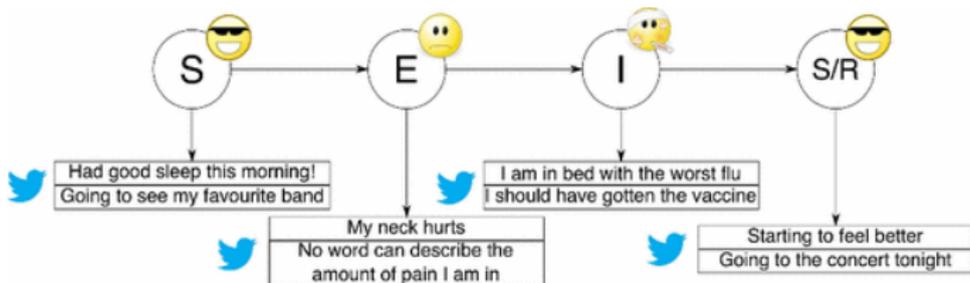
- Polls
- Surveys
- Voting patterns

Dataset: Social Media



- People share more readily
- Self-censorship on official channels
- Faster feedback

Dataset: Social Media



- People share more readily
- Self-censorship on official channels
- Faster feedback

Demography

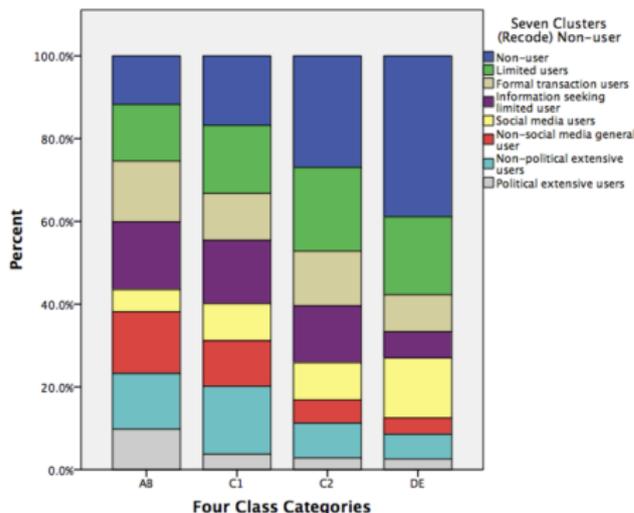
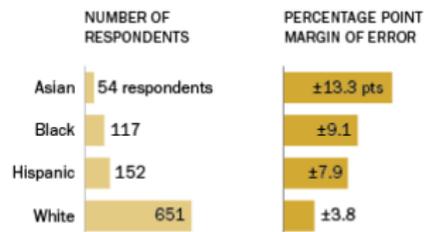


Figure 1: Types of internet user and NRS social class (Ofcom MLS 2012-13)

Bias in social media (and associated devices)

When sample sizes are smaller, margins of error are wider

In a hypothetical poll of 1,000 U.S. adults representative of the U.S. population ...

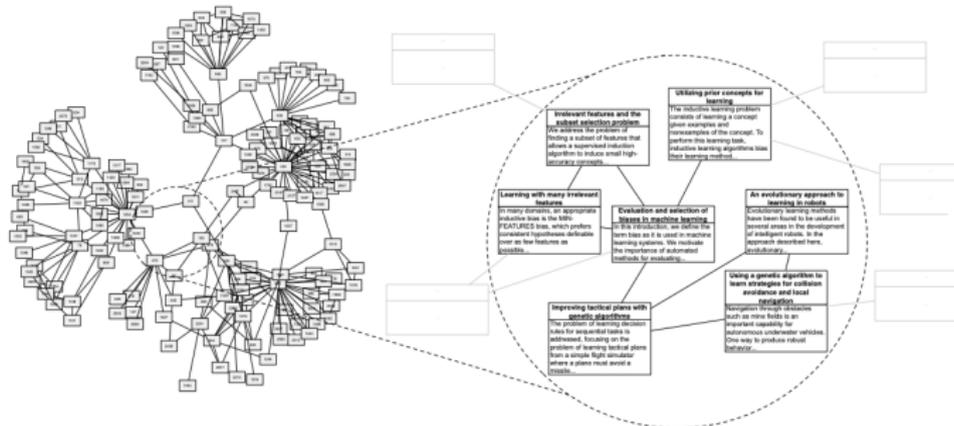


Note: Whites, black, Asian are single race, not Hispanic. Hispanics are any race. Margins of error assume a hypothetical survey estimate of 50% (for example, "50% approve of ...") at 95% confidence level, but with no effects of weighting or survey design adjustment.

PEW RESEARCH CENTER

But also a problem for traditional instruments

Dataset: Social Networks



- Predict links
- Predict when friendships end (and why)

More importantly ...

- Suicide risk
- Unhealthy habits
- Bullying

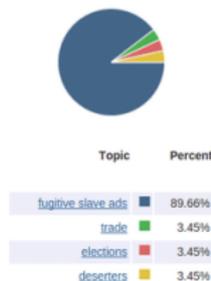
More importantly ...

- Suicide risk
- Unhealthy habits
- Bullying

Computational social science emphasizes interpretability, broader impacts, and captures what about language is unique to humans

Tool: Topic Models

TOPIC PROPORTIONS



From the Wed., Dec. 25, 1861 issue

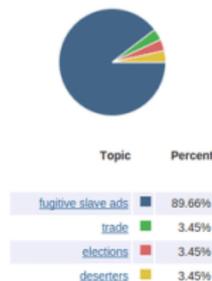
—Ranaway from the subscriber, on the 3d inst., my slave woman PARTHENA. Had on a dark brown and white calico dress. She is of a ginger-bread color; medium size; the right fore-finger shortened and crooked, from a whitlow. I think she is harbored somewhere in or near Duvall's addition. For her delivery to me I will pay \$10.

G. W. H. TYLER. de 6—ts

- Open-ended responses
- Questions you “forgot to ask”
- Help with interpretation
- Good for small amounts of data

Tool: Topic Models

TOPIC PROPORTIONS



From the Wed., Dec. 25, 1861 issue

—Ranaway from the subscriber, on the 3d inst., my slave woman PARTHENA. Had on a dark brown and white calico dress. She is of a ginger-bread color; medium size; the right fore-finger shortened and crooked, from a whitlow. I think she is harbored somewhere in or near Duvall's addition. For her delivery to me I will pay \$10.

G. W. H. TYLER. de 6—ts

- Open-ended responses
- Questions you “forgot to ask”
- Help with interpretation
- Good for small amounts of data
- Tension: prediction or interpretation

Qualitative Analysis (Grounded Theory)

Grounded Theory

Inductive coding of qualitative data to reach understanding

- low-quality models/theories and refining
- topics are “grounded” in specific word tokens
- Baumer et al. [2017]: compare themes
- model gives topic-specific “reading list”
- efficient, insulation from human biases.

Focus of today: Key Problems

- Sentiment Analysis
- Ideology Detection