Background

IMT 547 - Social Media Data Mining and Analysis

7-Jan-2021 (Week 1, Day 2)

Today's Plan

1. Questions around pre-requisites and course
2. Prior Knowledge
3. Paper discussion
    - *Also, social media and yesterday's Capitol Hill incident*
4. Jupyter notebook setup lab

Prerequisites

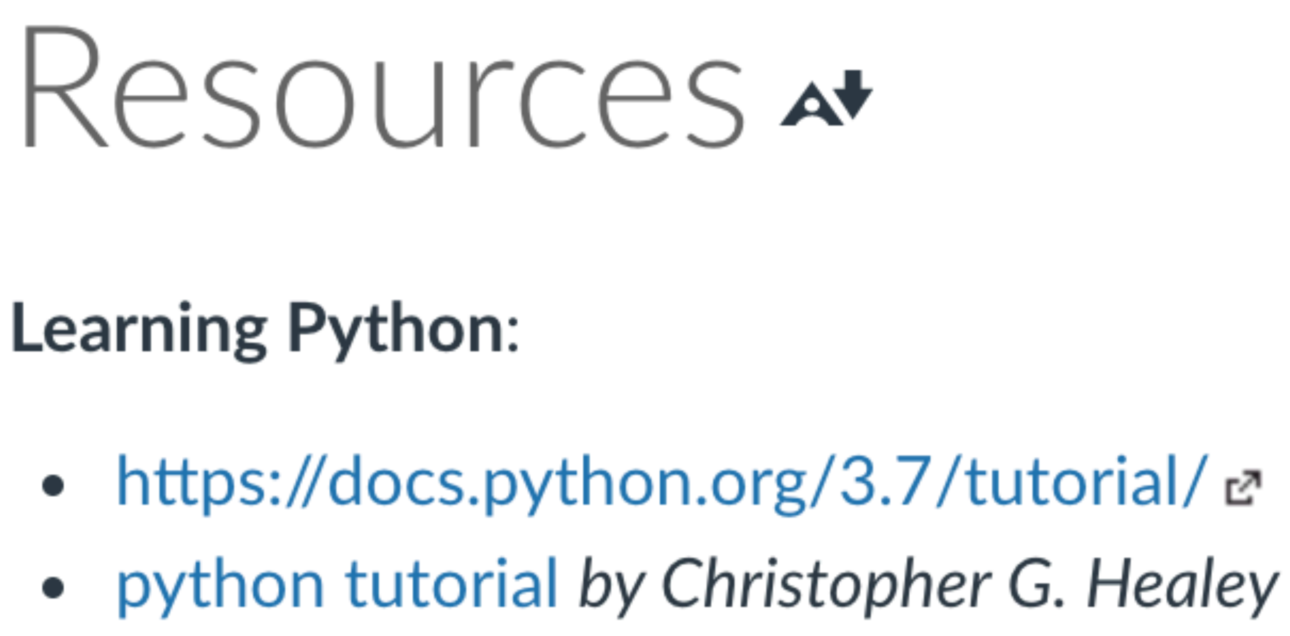
Prerequisites from syllabus

IMT573 or equivalent or permission of instructor. In terms of the required skills, students need to have basic knowledge of statistics and preliminary machine learning. An overview of the concepts and tools needed will be reviewed in class, however in-depth coverage of the fundamentals is not in the scope of this course. Students also need to be proficient in programming and comfortable programming in python.

If you are nervous about programming, best would be to take a programming courses before enrolling this one.

If you have never programmed in python, but have programmed in other languages, check:

https://canvas.uw.edu/courses/1434897/pages/resources



Resources

Learning Python:

- https://docs.python.org/3.7/tutorial/
- python tutorial *by Christopher G. Healey*

Social media mining and analysis

Learning Data Analytics is not a spectator sport!

- Type & run the **code**
  - in class (Jupyter notebook) for your labs,
  - for your problem sets individually
  - for your projects with your group
- Mess around with code until you understand what's going on
- Reference the documentation to learn more about

Learning Data Analytics in the context of Social Media

- Be willing and able to read research papers to have a basic understanding of the field
- Be able to connect to real-world happenings
- Since we will be dealing with real-world social media data and topics, at times the data that you see might be offensive, the topics that we will discuss might be controversial, even unpleasant. Before enrolling this course be sure that you are comfortable with this.

Few other things to keep in mind before deciding

- This course could be a lot of work. Every week there is something due!! *I am not comfortable with that.*
- There is in-class coding involved (aka labs). *I am not comfortable with that.*
- Those are some cool topics on the syllabus, but wait there is only so much that can be covered during class time. I need to do my own reading beyond just what is discussed in class. I need to read python documentation? *I am not comfortable with that.*
- Projects sound fun, but there are no free riders. Individual graded accountability would be built in despite being group projects
- I have already taken other heavy courses and won't be able to spend much time on this one
- Group project and team work is really not my thing
- You want the instructor to go over every new technology/tool that might be useful for your project
- You do not care about messy data and have no patience to clean & explore or throwaway and start again until you have derived useful insights.

Background survey

Paper discussion

Social media mining and analysis

Learning Data Analytics

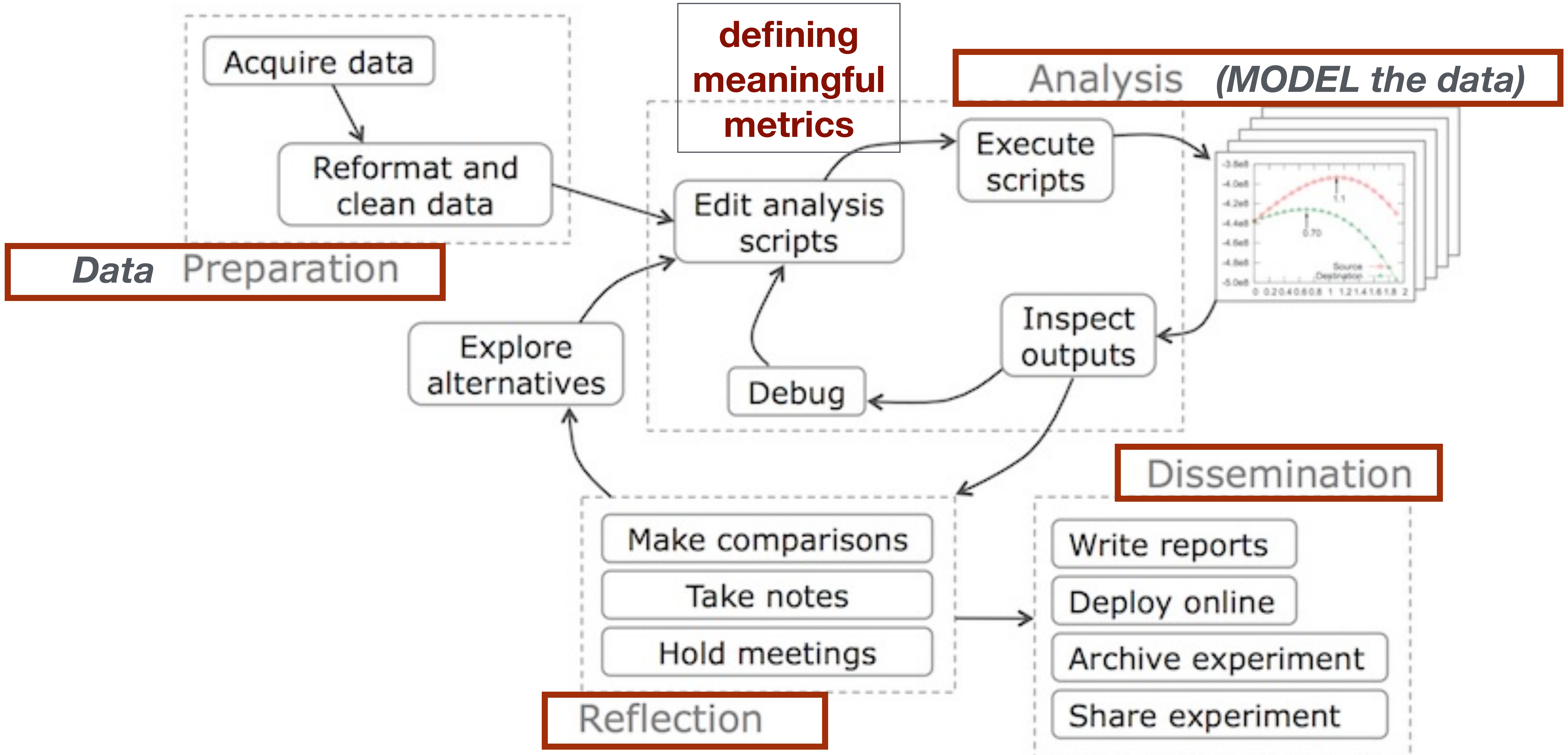
Data Science Workflow: Overview and Challenges, *CACM Blog, By Philip Guo*

Learning Data Analytics in the context of Social Media

Computational Social Science, *by Lazer et. al*

Data Science Workflow

Data Science Workflow: Overview and Challenges
By Philip Guo, *Communications of the ACM*



defining meaningful metrics

Analysis *(MODEL the data)*

Acquire data

Reformat and clean data

Data Preparation

Edit analysis scripts

Execute scripts

Inspect outputs

Debug

Explore alternatives

Dissemination

Make comparisons

Take notes

Hold meetings

Reflection

Write reports

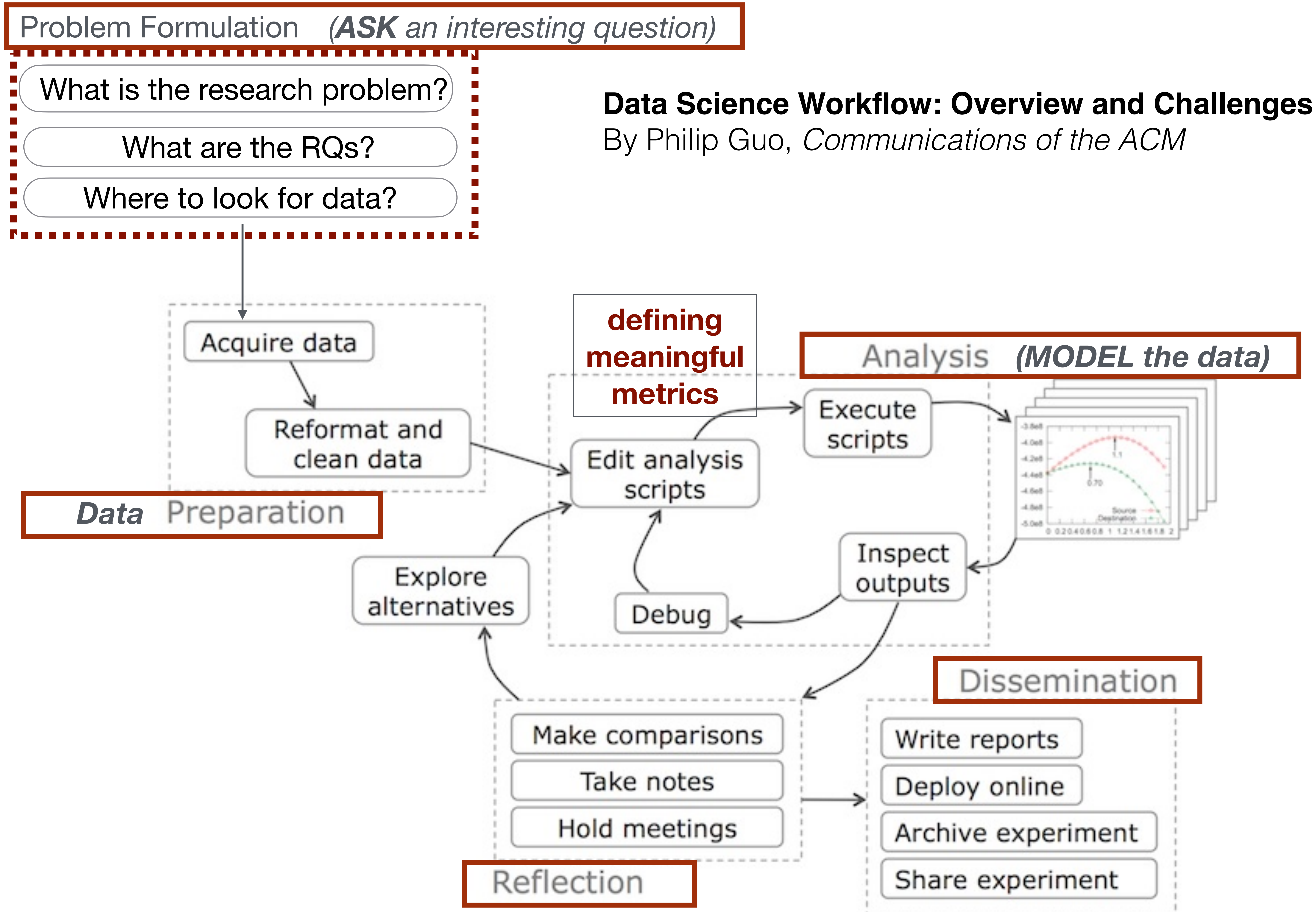
Deploy online

Archive experiment

Share experiment

Data Science Workflow

Data Science Workflow: Overview and Challenges
By Philip Guo, *Communications of the ACM*



Problem Formulation (***ASK** an interesting question*)

What is the research problem?

What are the RQs?

Where to look for data?

Acquire data

Reformat and clean data

Data Preparation

defining meaningful metrics

Edit analysis scripts

Execute scripts

Analysis ***(MODEL the data)***

Inspect outputs

Debug

Explore alternatives

Make comparisons

Take notes

Hold meetings

Reflection

Dissemination

Write reports

Deploy online

Archive experiment

Share experiment

Data Science Workflow

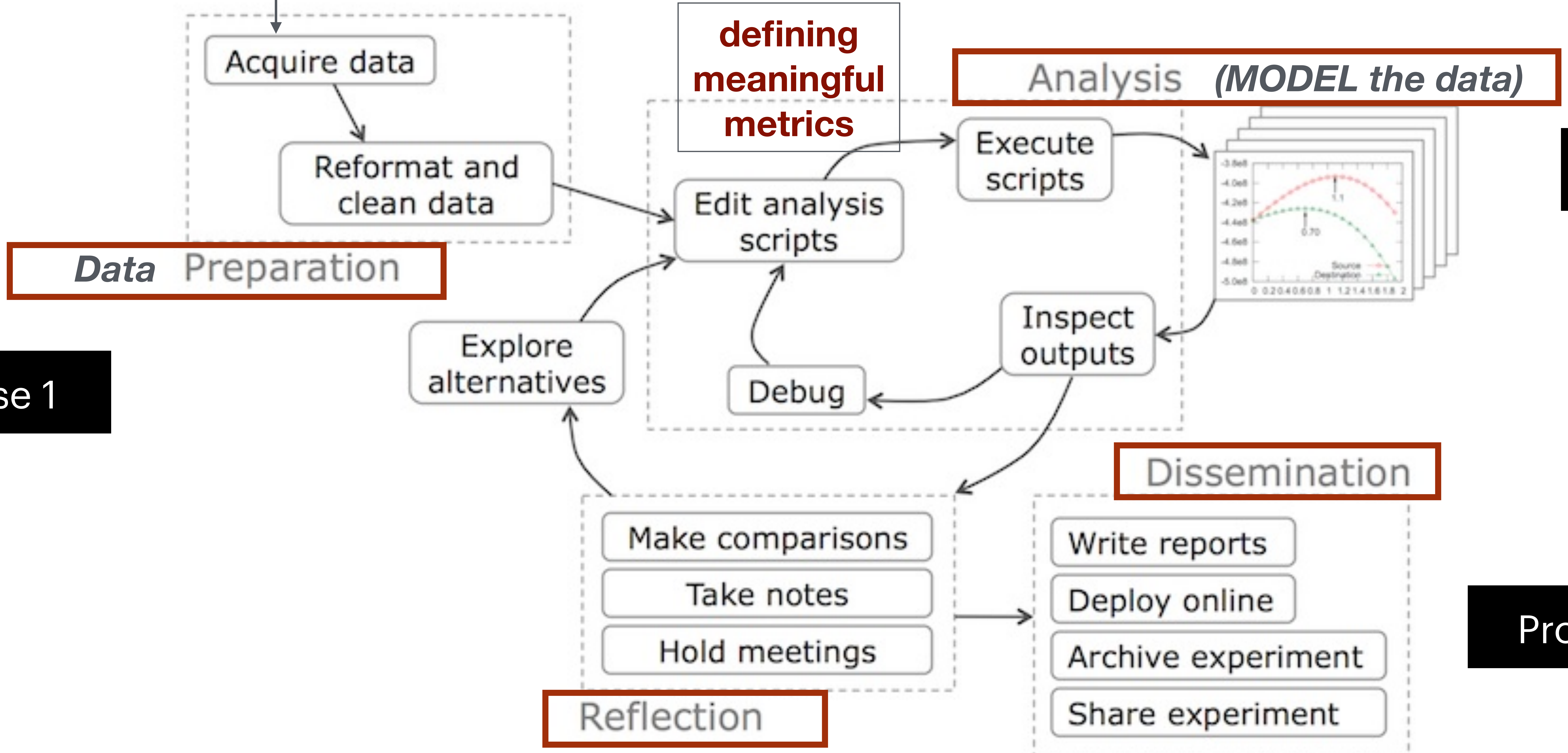
Problem Formulation (***ASK*** *an interesting question)*

What is the research problem?

What are the RQs?

Where to look for data?

Data Science Workflow: Overview and Challenges
By Philip Guo, *Communications of the ACM*



Acquire data

Reformat and clean data

Data Preparation

defining meaningful metrics

Analysis ***(MODEL the data)***

Edit analysis scripts

Execute scripts

Inspect outputs

Debug

Explore alternatives

Make comparisons

Take notes

Hold meetings

Reflection

Dissemination

Write reports

Deploy online

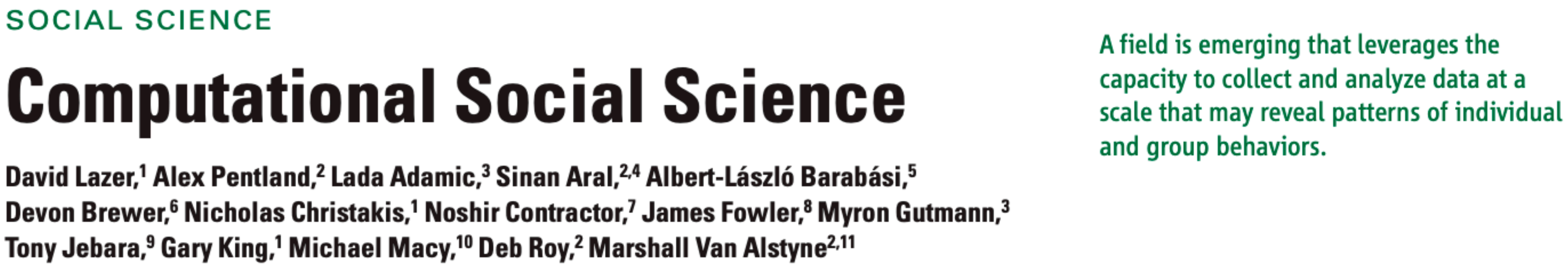
Archive experiment

Share experiment

Project Phase 1

Project Phase 2

Project Phase 3



SOCIAL SCIENCE

Computational Social Science

David Lazer,[1] Alex Pentland,[2] Lada Adamic,[3] Sinan Aral,[2,4] Albert-László Barabási,[5] Devon Brewer,[6] Nicholas Christakis,[1] Noshir Contractor,[7] James Fowler,[8] Myron Gutmann,[3] Tony Jebara,[9] Gary King,[1] Michael Macy,[10] Deb Roy,[2] Marshall Van Alstyne[2,11]

A field is emerging that leverages the capacity to collect and analyze data at a scale that may reveal patterns of individual and group behaviors.

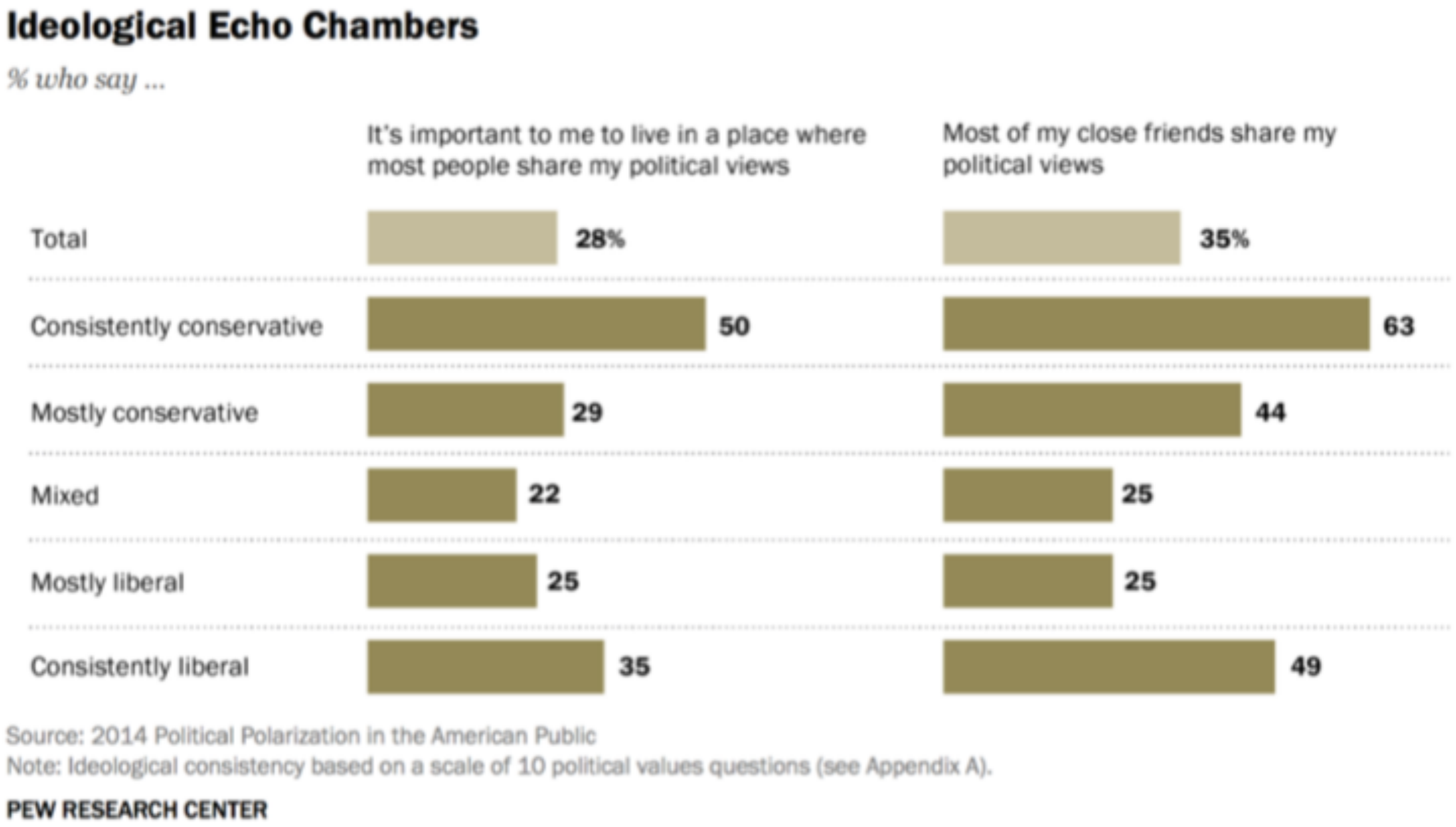
Computational techniques to answer Social Science questions

Data analytics to answer questions about human interactions happening on online social media platforms.

Example

Social Science Question:
Is there growing political polarization in the U.S.?

Traditional Social Science approach - ask **people**



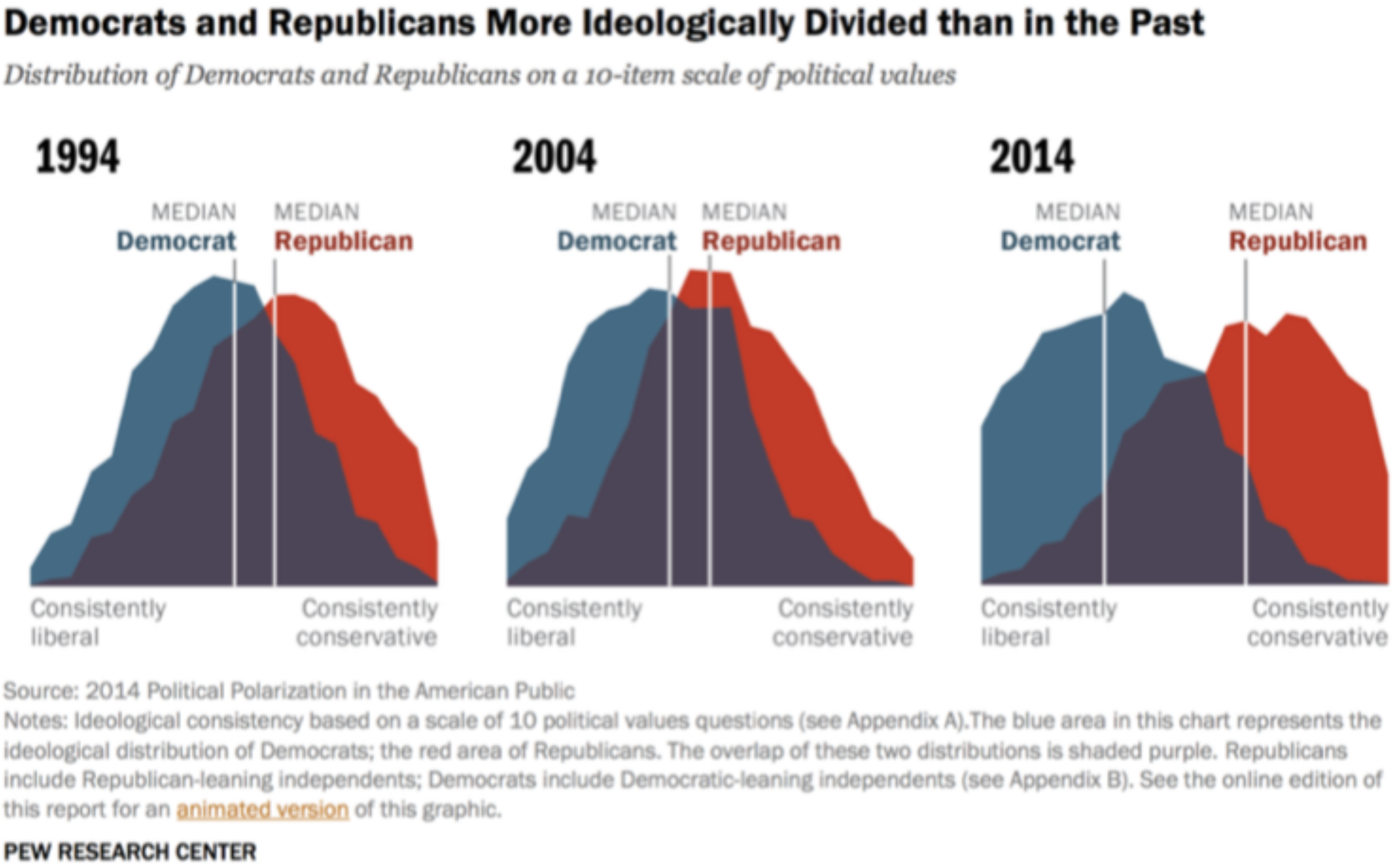
Ideological Echo Chambers

% who say ...

Source: 2014 Political Polarization in the American Public
Note: Ideological consistency based on a scale of 10 political values questions (see Appendix A).

PEW RESEARCH CENTER

Traditional Social Science approach - ask **people**



Democrats and Republicans More Ideologically Divided than in the Past

Distribution of Democrats and Republicans on a 10-item scale of political values

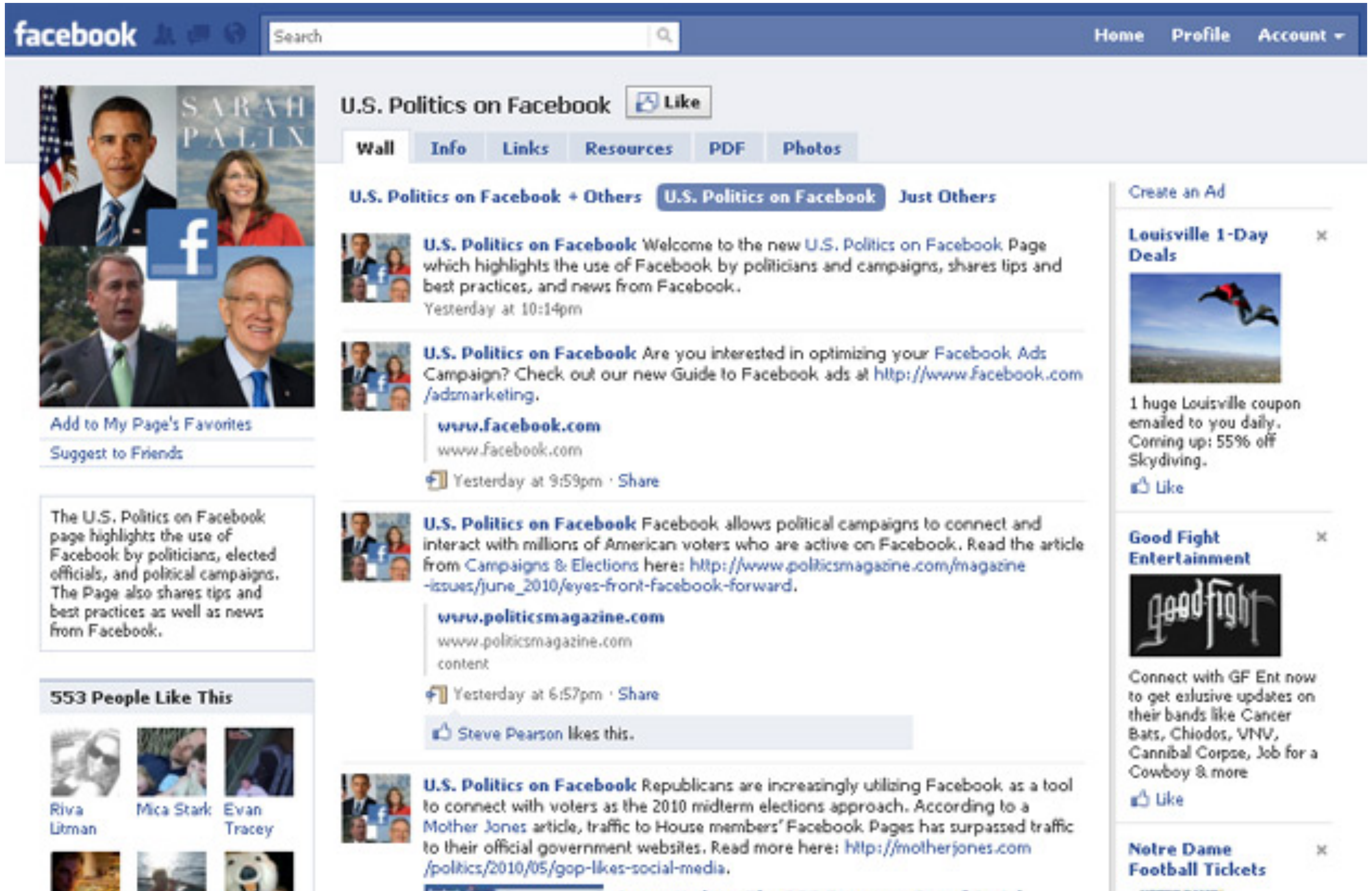
Source: 2014 Political Polarization in the American Public

Notes: Ideological consistency based on a scale of 10 political values questions (see Appendix A).The blue area in this chart represents the ideological distribution of Democrats; the red area of Republicans. The overlap of these two distributions is shaded purple. Republicans include Republican-leaning independents; Democrats include Democratic-leaning independents (see Appendix B). See the online edition of this report for an animated version of this graphic.

PEW RESEARCH CENTER

How would do it today?
What is the computational social science approach?

Analyze people’s Facebook interactions



The U.S. Politics on Facebook page highlights the use of Facebook by politicians, elected officials, and political campaigns. The Page also shares tips and best practices as well as news from Facebook.

U.S. Politics on Facebook Welcome to the new U.S. Politics on Facebook Page which highlights the use of Facebook by politicians and campaigns, shares tips and best practices, and news from Facebook.
Yesterday at 10:14pm

U.S. Politics on Facebook Are you interested in optimizing your Facebook Ads Campaign? Check out our new Guide to Facebook ads at http://www.facebook.com/adsmarketing.
www.facebook.com
www.facebook.com
Yesterday at 9:59pm · Share

U.S. Politics on Facebook Facebook allows political campaigns to connect and interact with millions of American voters who are active on Facebook. Read the article from Campaigns & Elections here: http://www.politicsmagazine.com/magazine-issues/june_2010/eyes-front-facebook-forward.
www.politicsmagazine.com
www.politicsmagazine.com
content
Yesterday at 6:57pm · Share
Steve Pearson likes this.

U.S. Politics on Facebook Republicans are increasingly utilizing Facebook as a tool to connect with voters as the 2010 midterm elections approach. According to a Mother Jones article, traffic to House members' Facebook Pages has surpassed traffic to their official government websites. Read more here: http://motherjones.com/politics/2010/05/gop-likes-social-media.

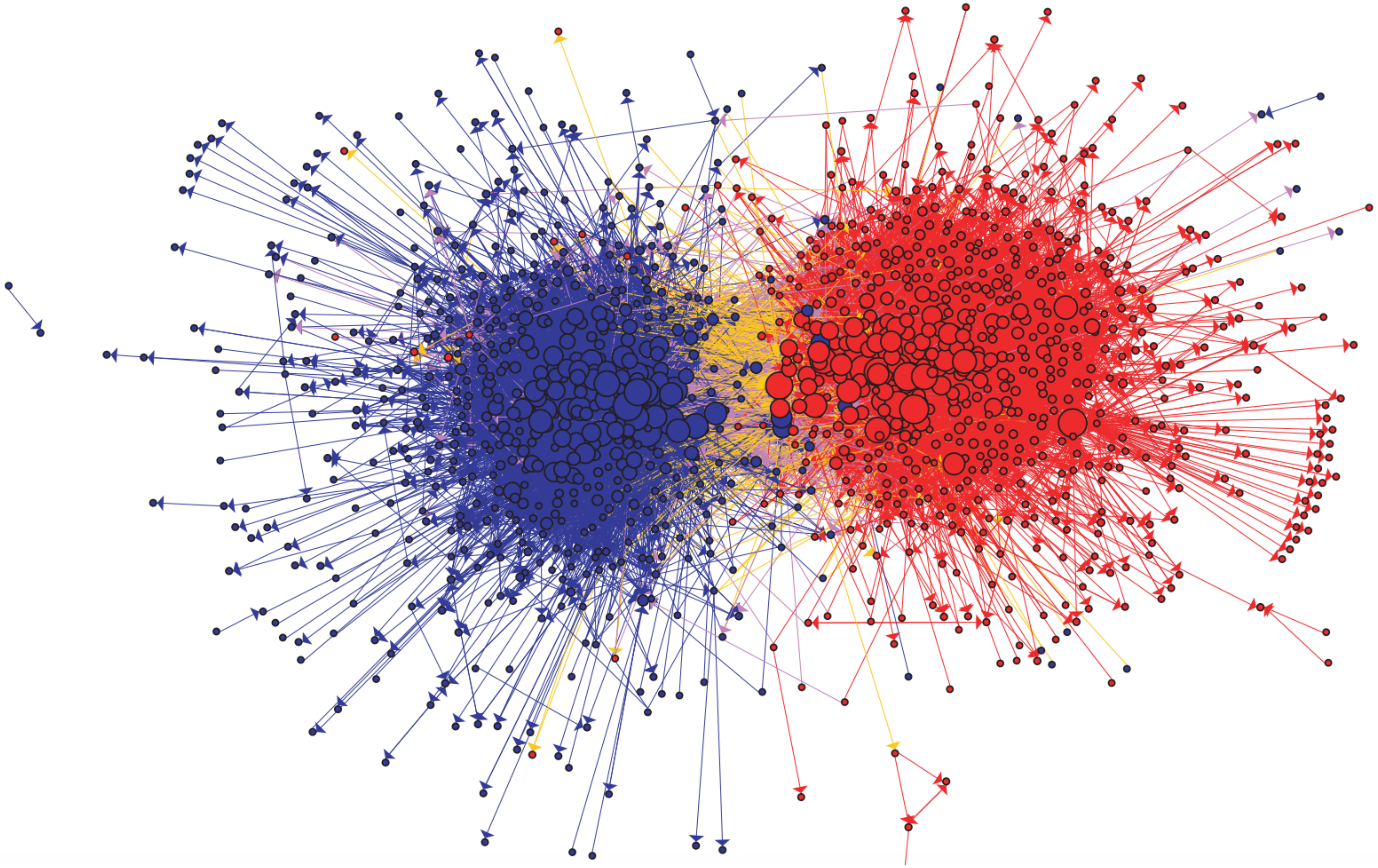
1 huge Louisville coupon emailed to you daily. Coming up: 55% off Skydiving.

Connect with GF Ent now to get exlusive updates on their bands like Cancer Bats, Chiodos, VNV, Cannibal Corpse, Job for a Cowboy & more

Community structure of political blogs
blog authors linked only to others of similar political leanings

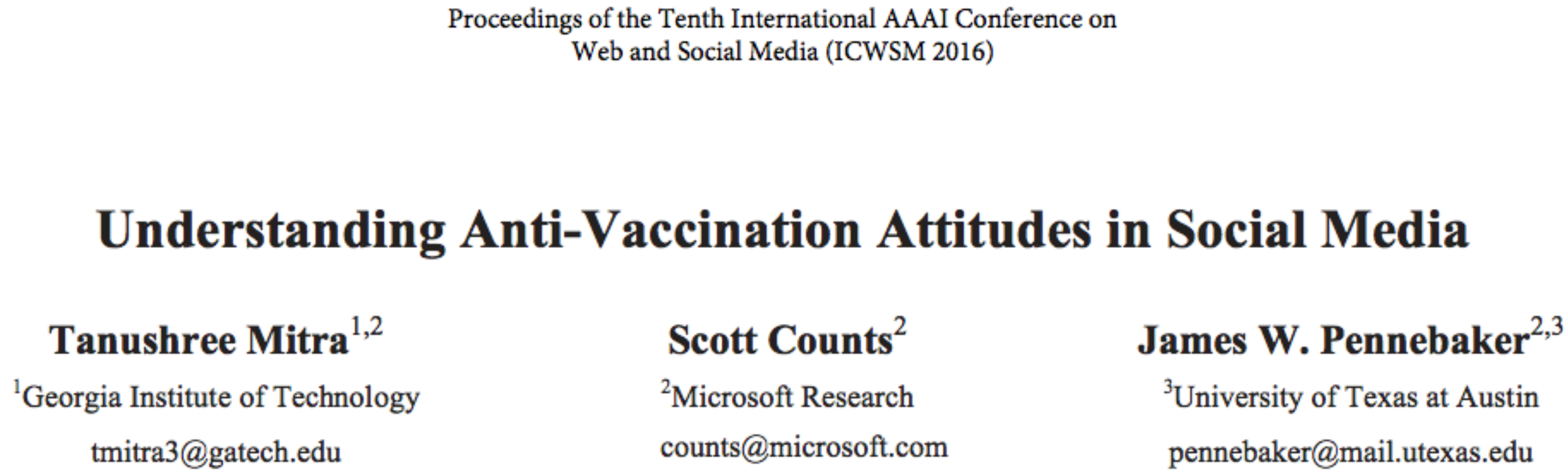
The colors reflect political orientation, red for conservative, and blue for liberal. Orange links go from liberal to conservative, and purple ones from conservative to liberal. The size of each blog reflects the number of other blogs that link to it.

What does this visualization suggest?



The Political Blogosphere and the 2004 U.S. Election, *Lada Adamic, Natalie Glance, [cited > 3000 times]*
ACM LinkKDD - KDD workshop (Knowledge Discovery & Data Mining)

Question answered with social media data:
Who are the anti-vaxxers?
What drives people to develop and perpetuate the anti-vaccination movement?



Proceedings of the Tenth International AAAI Conference on
Web and Social Media (ICWSM 2016)

Understanding Anti-Vaccination Attitudes in Social Media

Tanushree Mitra[1,2]
[1]Georgia Institute of Technology
tmitra3@gatech.edu

Scott Counts[2]
[2]Microsoft Research
counts@microsoft.com

James W. Pennebaker[2,3]
[3]University of Texas at Austin
pennebaker@mail.utexas.edu

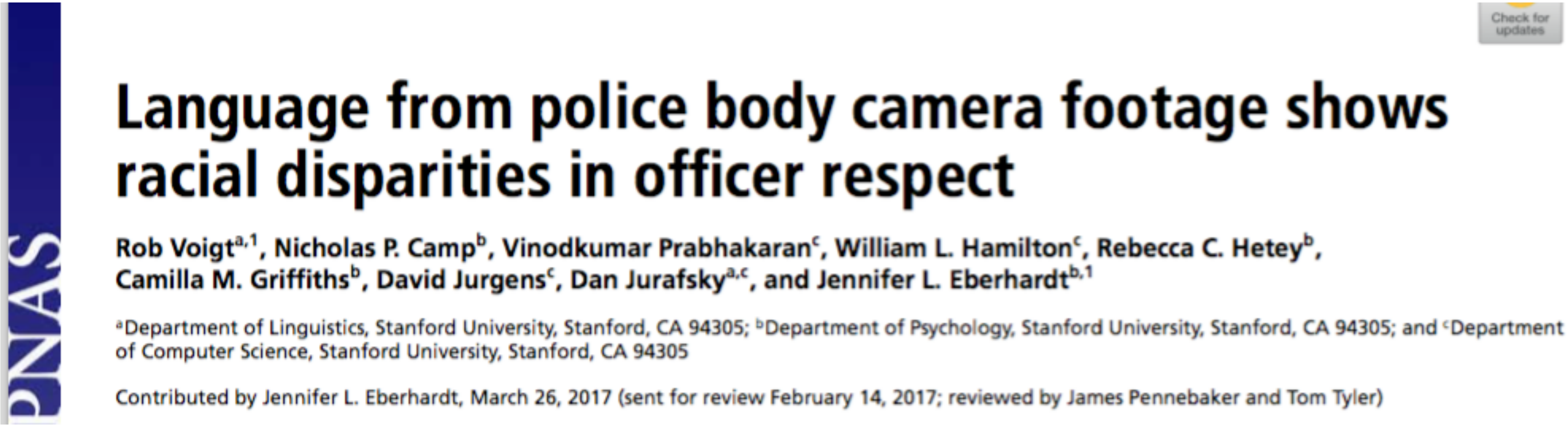
Topics anti-vaxxers talk about

> 3 million tweets from 32K users

SECRET GOVT.	CONSPIRACY	ORGANIC
cia	imf	gmo
ufo	laden	food
wtc	infowar	usda
secret	eugenic	organic
illuminati	conspiracy	chemical
homeland	bilderberg	monsanto
underground	dictatorship	genetically

[**Mitra, T.,** Counts, S., & Pennebaker, J.W., ICWSM 2016]

Social Science Question:
Do officers treat white community members with a greater degree of respect than they afford to blacks?

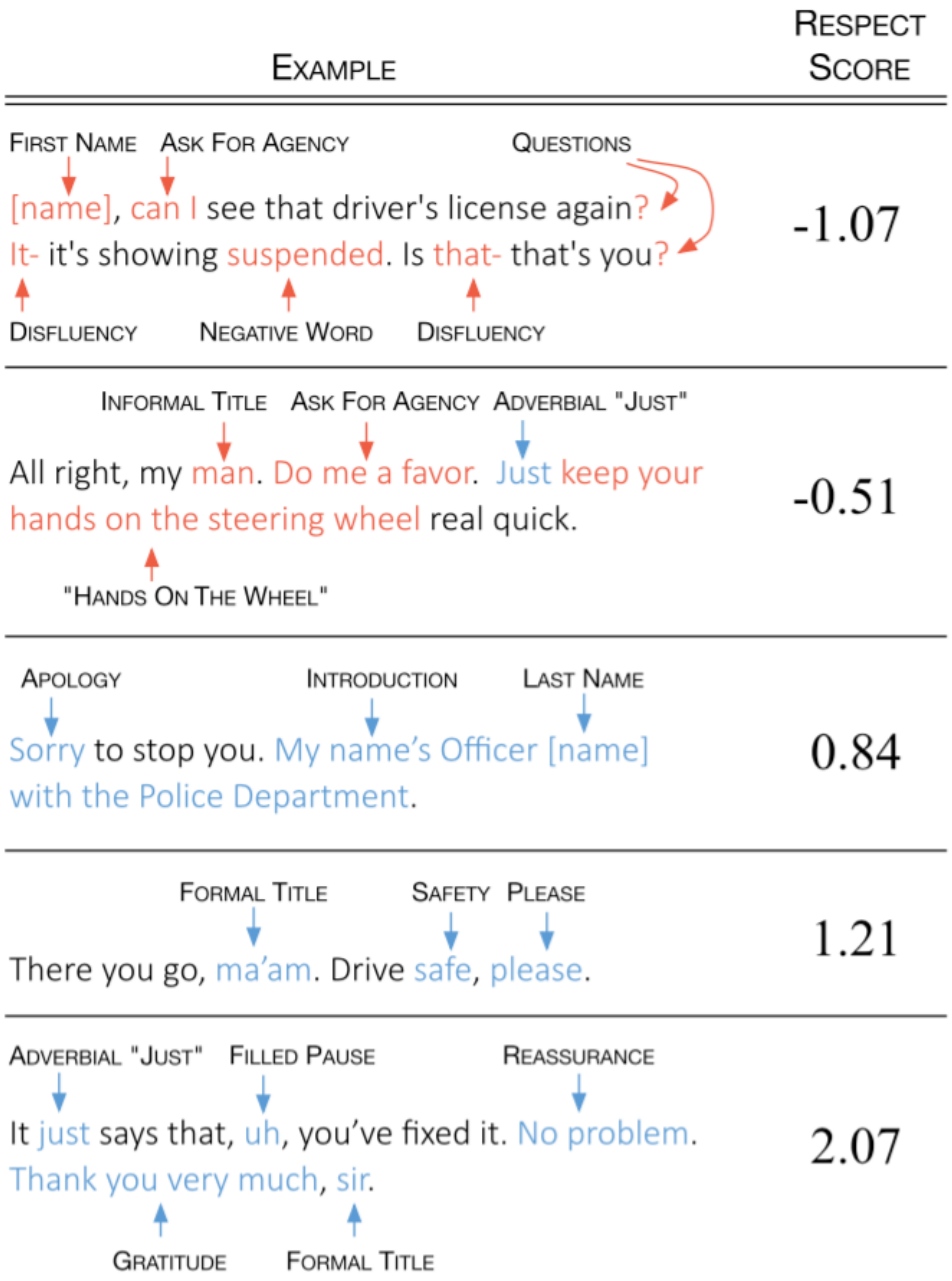


Language from police body camera footage shows racial disparities in officer respect

Rob Voigt[a,1], Nicholas P. Camp[b], Vinodkumar Prabhakaran[c], William L. Hamilton[c], Rebecca C. Hetey[b], Camilla M. Griffiths[b], David Jurgens[c], Dan Jurafsky[a,c], and Jennifer L. Eberhardt[b,1]

[a]Department of Linguistics, Stanford University, Stanford, CA 94305; [b]Department of Psychology, Stanford University, Stanford, CA 94305; and [c]Department of Computer Science, Stanford University, Stanford, CA 94305

Contributed by Jennifer L. Eberhardt, March 26, 2017 (sent for review February 14, 2017; reviewed by James Pennebaker and Tom Tyler)



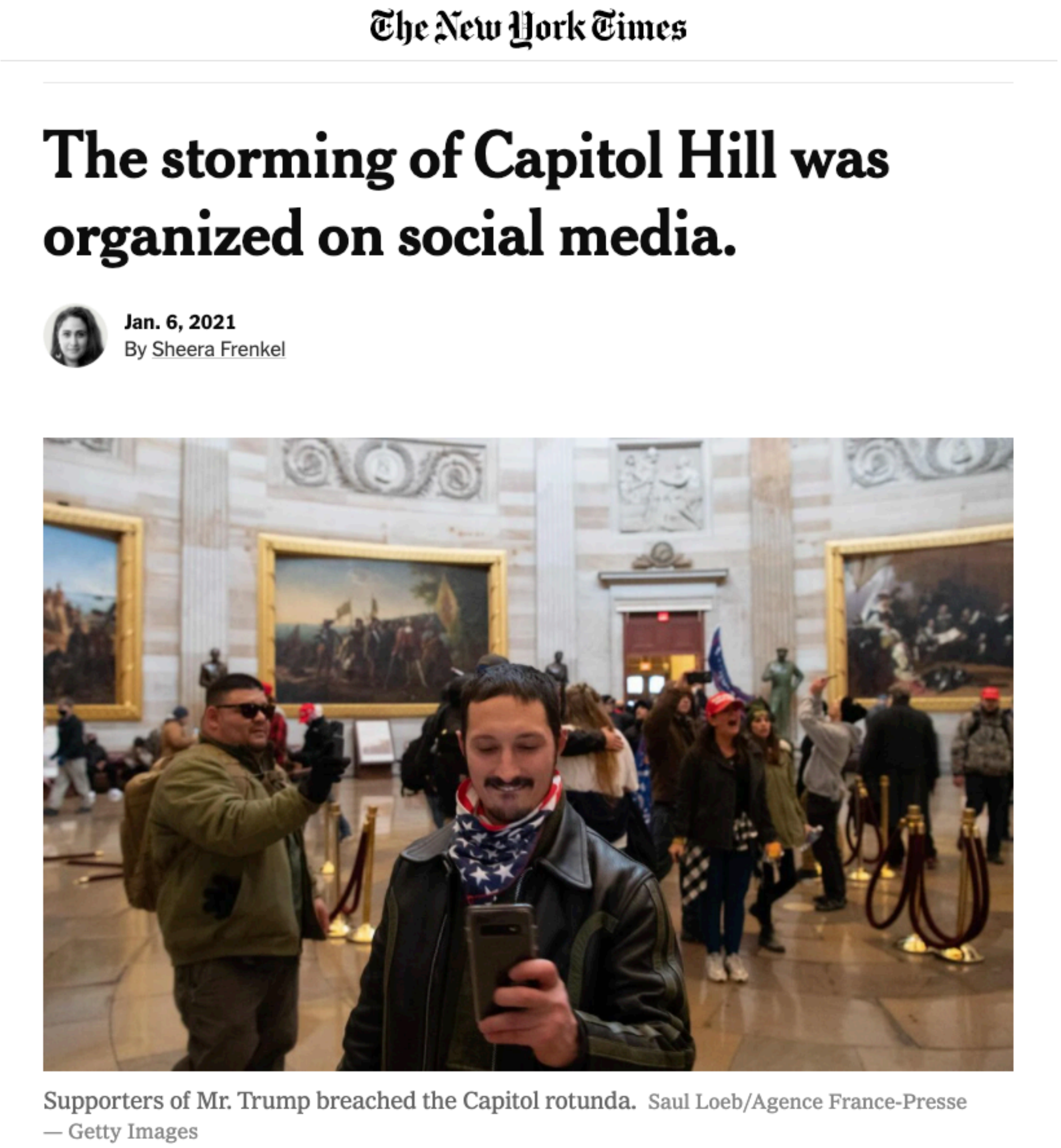
Example	Respect Score
[name], can I see that driver's license again? It- it's showing suspended. Is that- that's you?	-1.07
All right, my man. Do me a favor. Just keep your hands on the steering wheel real quick.	-0.51
Sorry to stop you. My name's Officer [name] with the Police Department.	0.84
There you go, ma'am. Drive safe, please.	1.21
It just says that, uh, you've fixed it. No problem. Thank you very much, sir.	2.07

Quote from the paper

‘These vast, emerging data sets on how people interact surely offer qualitatively new perspectives on collective human behavior.’

‘These vast, **emerging data sets** on **how people interact** surely offer qualitatively **new perspectives** on **collective human behavior**.’

- What social media platforms did you see this event being reported or discussed?
- Any new emerging platforms?
- Did you see any intriguing data (*posts, messages, videos….platforms where it was happening*)?
- Were you able to spot any perspectives about this collective behavior? *Thoughts about any intriguing observation, any thoughts on what you witnessed online?*



The New York Times

The storming of Capitol Hill was organized on social media.

Jan. 6, 2021

By Sheera Frenkel

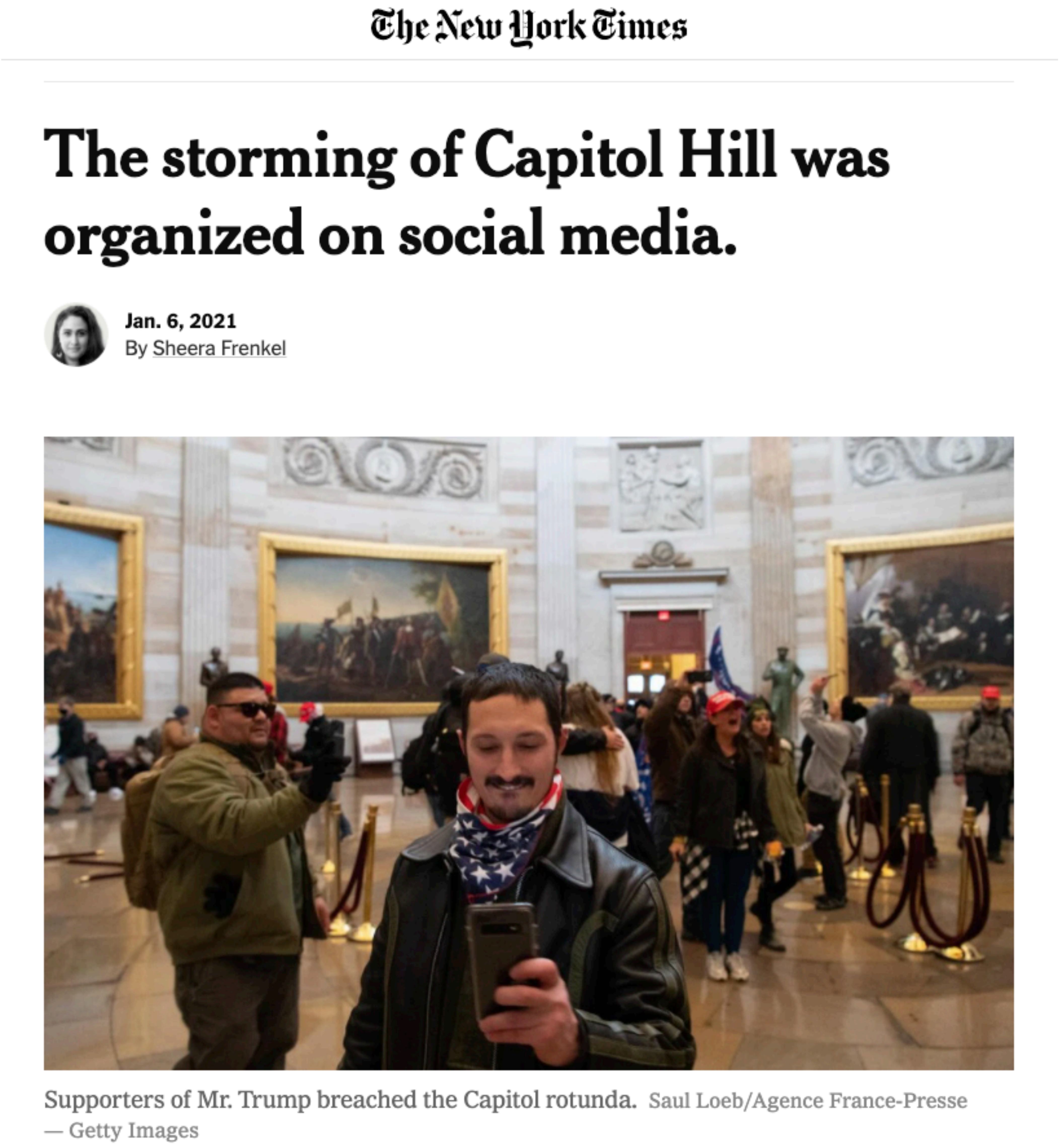
Supporters of Mr. Trump breached the Capitol rotunda. Saul Loeb/Agence France-Presse — Getty Images https://www.nytimes.com/2021/01/06/us/politics/protesters-storm-capitol-hill-building.html

On **social media sites** requested by the far-right, such as **Gab** and **Parler**, directions on which streets to take to avoid the police and which tools to bring to help pry open doors were exchanged in comments. At least a dozen people posted about carrying guns into the halls of Congress.

As Facebook and Twitter began to crack down groups like QAnon and the Proud Boys over the summer, **they slowly migrated to other sites** that allowed them to openly call for violence.

Calls for violence against members of Congress and for pro-Trump movements to retake the Capitol building have been circulating online for months. ….. fringe **movements** like QAnon and the Proud Boys, groups have openly **organized on social media networks** and recruited others to their cause.

On Wednesday, their **online activism became real-world violence**



The New York Times

The storming of Capitol Hill was organized on social media.

Jan. 6, 2021
By Sheera Frenkel

Supporters of Mr. Trump breached the Capitol rotunda. Saul Loeb/Agence France-Presse — Getty Images https://www.nytimes.com/2021/01/06/us/politics/protesters-storm-capitol-hill-building.html

On **social media sites** requested by the far-right, such as **Gab** and **Parler**, directions on which streets to take to avoid the police and which tools to bring to help pry open doors were exchanged in comments. At least a dozen people posted about carrying guns into the halls of Congress.

As Facebook and Twitter began to crack down groups like QAnon and the Proud Boys over the summer, **they slowly migrated to other sites** that allowed them to **openly call for violence**.

Calls for violence against members of Congress and for pro-Trump movements to retake the Capitol building have been circulating online for months. ….. fringe **movements** like QAnon and the Proud Boys, groups have openly **organized on social media networks** and recruited others to their cause.

On Wednesday, their **online activism became real-world violence**

Information exchange on social networks, asking questions to seek answers - a common phenomena on online networks

Moderation on one platform, people move to another

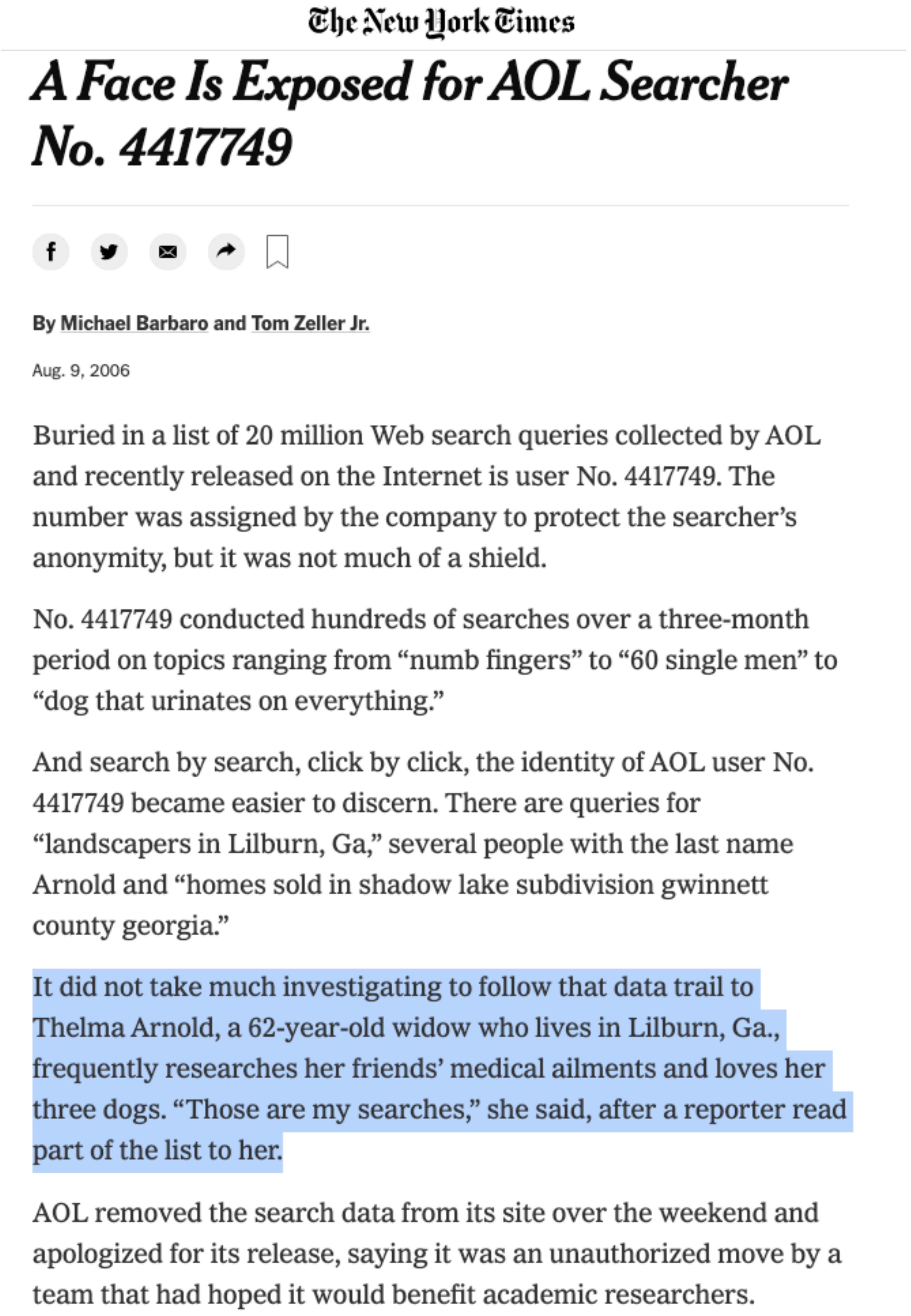
Social media companies ever changing moderation rules:

https://help.twitter.com/en/rules-and-policies/twitter-rules

https://www.facebook.com/communitystandards/

Quote from the paper

“Perhaps the thorniest challenges exist on the data side, with respect to access and privacy”



The New York Times

A Face Is Exposed for AOL Searcher No. 4417749

By Michael Barbaro and Tom Zeller Jr.

Aug. 9, 2006

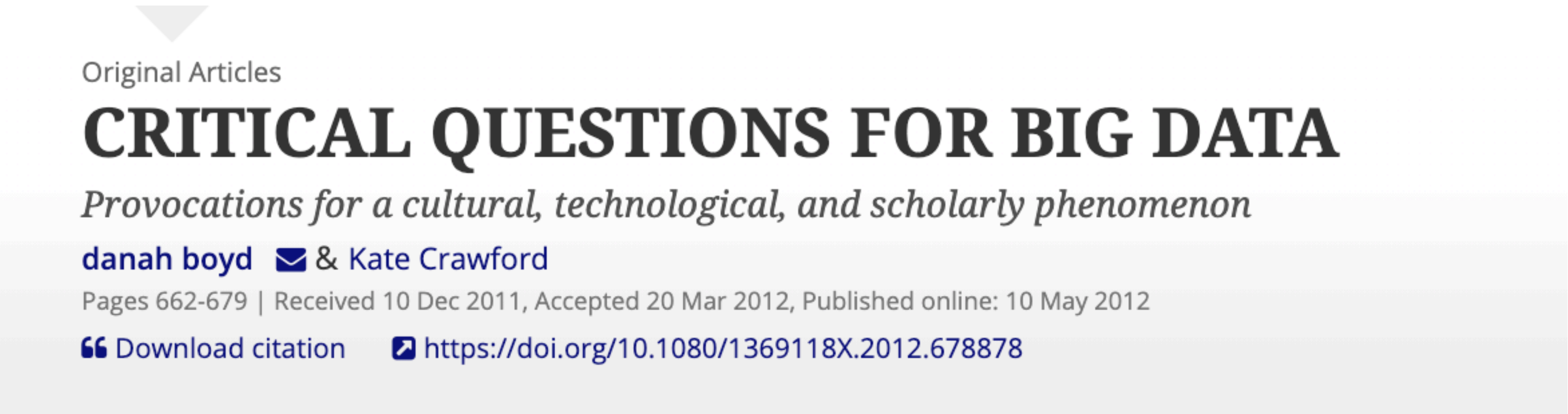
Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher’s anonymity, but it was not much of a shield.

No. 4417749 conducted hundreds of searches over a three-month period on topics ranging from “numb fingers” to “60 single men” to “dog that urinates on everything.”

And search by search, click by click, the identity of AOL user No. 4417749 became easier to discern. There are queries for “landscapers in Lilburn, Ga,” several people with the last name Arnold and “homes sold in shadow lake subdivision gwinnett county georgia.”

It did not take much investigating to follow that data trail to Thelma Arnold, a 62-year-old widow who lives in Lilburn, Ga., frequently researches her friends’ medical ailments and loves her three dogs. “Those are my searches,” she said, after a reporter read part of the list to her.

AOL removed the search data from its site over the weekend and apologized for its release, saying it was an unauthorized move by a team that had hoped it would benefit academic researchers.



Original Articles

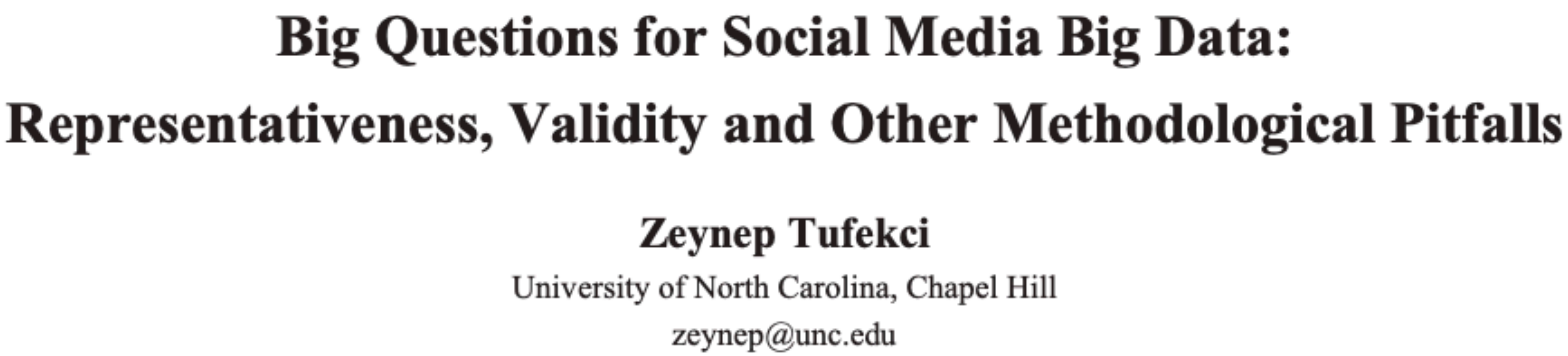
CRITICAL QUESTIONS FOR BIG DATA

Provocations for a cultural, technological, and scholarly phenomenon

danah boyd ✉ & Kate Crawford

Pages 662-679 | Received 10 Dec 2011, Accepted 20 Mar 2012, Published online: 10 May 2012

Download citation https://doi.org/10.1080/1369118X.2012.678878



Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls

Zeynep Tufekci
University of North Carolina, Chapel Hill
zeynep@unc.edu

the hidden influence of social networks

TED TALK: https://www.ted.com/talks/nicholas_christakis_the_hidden_influence_of_social_networks

BREAK

Be back at 9:45am

Lab

Before Next Class

Install:

- **anaconda python 3**:

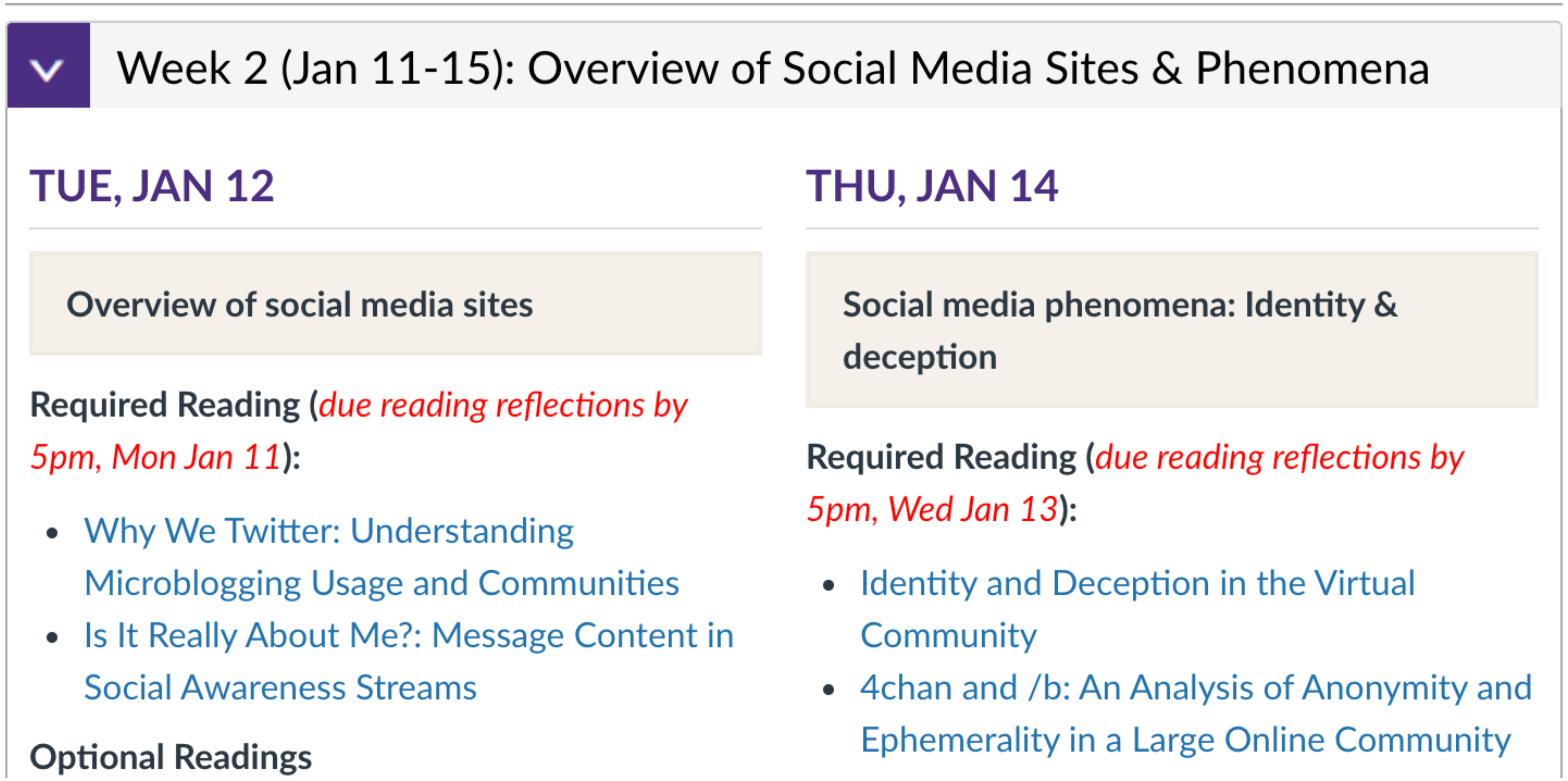
  https://www.anaconda.com/products/individual

- **jupyter notebook**, https://jupyter.org/install

Reader/Grader, Shikhar will hold office hour on Friday: 9:30am - 10:30am PT

Meet if you run into installation issues.

Next Week
1st reflection due by 5pm on Monday (no late days allowed on reflection, see late policy)



Week 2 (Jan 11-15): Overview of Social Media Sites & Phenomena

TUE, JAN 12

Overview of social media sites

Required Reading (*due reading reflections by 5pm, Mon Jan 11*):

- Why We Twitter: Understanding Microblogging Usage and Communities
- Is It Really About Me?: Message Content in Social Awareness Streams

Optional Readings

THU, JAN 14

Social media phenomena: Identity & deception

Required Reading (*due reading reflections by 5pm, Wed Jan 13*):

- Identity and Deception in the Virtual Community
- 4chan and /b: An Analysis of Anonymity and Ephemerality in a Large Online Community