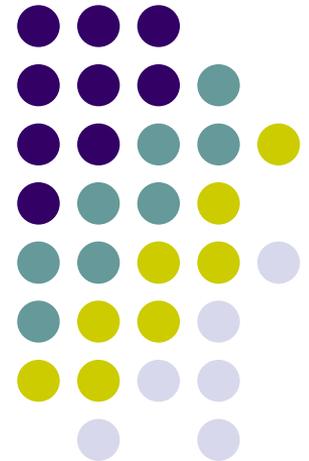
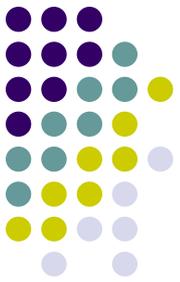


# *Applied Scaling & Classification Techniques in Political Science*

Lecture 4 (part 3)

Social Media as Data Generators

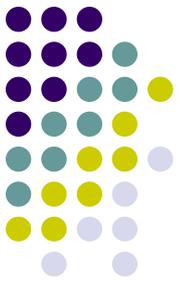




# Reference

- ✓ Barbera, Pablo, and Zachary C. Steinert-Threlkeld (2020). Social media as data generators. In Luigi Curini and Robert Franzese (eds.), *SAGE Handbook of Research Methods in Political Science & International Relations*, London, Sage, chapter 23
- ✓ Freelon, Deen. 2018. “Computational research in the post-API age.” *Political Communication*, 35: 665-668

# Social media

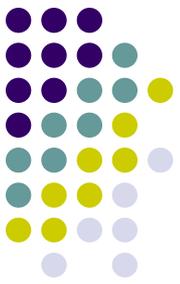


Citizens across the globe spend an increasing proportion of their daily lives on **social media websites**, such as Twitter, Facebook or Instagram

Their activities on the sites generate **granular, time-stamped footprints of human behavior and personal interactions**, sometimes with **longitude and latitude coordinates**

A sizable proportion of these **digital traces have to do with politics** – social media is an increasingly popular source of political news, as well as a forum for political debates where virtually every political candidate running for office is now present

# Social media

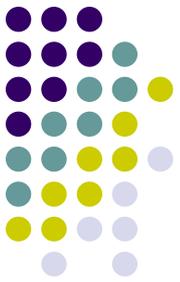


The data generated from these interactions is in many cases (more often in the past, still...) **freely available** for research purposes and provides a depth and breadth that was unimaginable even one decade ago

The high degree of **spatial and temporal granularity** allows the **study of behavior at low levels of aggregation** but also at a **more macro scale** and from a comparative perspective

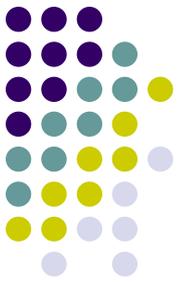
The fact that **human behavior is observed unobtrusively** also facilitates collecting data at a larger scale and reduces certain **types of biases** (on “hot” topics such as racism, terrorism approval, etc.)

# Social media



This set of advantages makes **social media data a new and exciting source of information** to study key questions about political and social behavior

# Social media



Two (main) types of studies **using social media data**:

- ✓ those where social media is being used merely as a **source of data**
- ✓ and those where the focus is on how social media is **transforming different political/social phenomena**

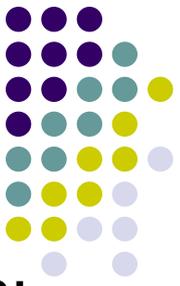
# Social media



The **first group** includes research that uses social media to  
(among the others):

- ✓ measure public opinion on several topics
- ✓ the ideology of citizens and elites
- ✓ the structure of social networks
- ✓ government censorship
- ✓ conflict dynamics
- ✓ elite rhetoric
- ✓ as well as research where social media sites are used as a new space to conduct affordable field experiments

# Social media



The **second set of studies** deals with questions such as:

- ✓ how social media platforms contribute to the success of collective action events
- ✓ how they are transforming election campaigns
- ✓ and whether their usage is contributing to greater political polarization and the spread of misinformation

# Social media

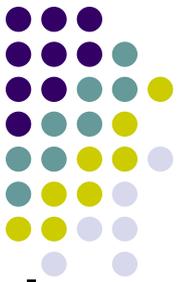


Employing social media...does it make any sense or is it **just noise?**

Despite obvious concerns about **representativeness of samples obtained from social media**, for some groups of actors **virtually the entire population** is present on social networking platforms

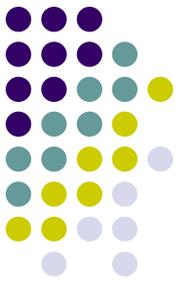
For example, over 85% of world leaders have active Twitter or Facebook accounts and virtually all Members of the U.S. Congress also maintain profiles on these sites, making it possible to make externally valid inferences about their communication strategies and rhetoric with data obtained from these accounts

# Social media



Moreover, representativeness of social media users is **only a concern** if the **behavior under study** is thought to vary according to variables by which **users on social media substantially differ** from the general population from which they are drawn

# Social media



The fact is that many well-known offline behaviors also occur online, supports the claim that representativeness issues matter but probably **not as much as feared**

Contrary in fact to the common view that characterizes social media interactions as **not occurring in the “real world”**, behavior on these sites **indeed mirrors offline behavior**, and thus it can help reveal the mechanisms that drive human behavior, not only on these platforms, but in people’s lives more generally

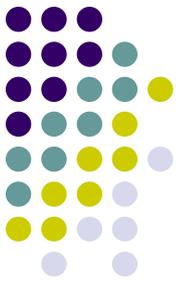
# Social media



**Social media users' ideology**, for example, is recoverable from the structure of their social network and maps onto offline estimates of ideology

Attitudinal homophily, diurnal activity, and geographic constraints occur offline and online, and Facebook users' scores on the Big Five personality traits (extraversion, neuroticism, conscientiousness, agreeableness, and openness to experience) are recoverable from their behavior on the site

# Social media

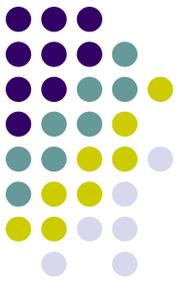


In other words...

...the preponderance of evidence suggests it is reasonable to expect that many, perhaps most, behaviors studied using social media are analogues to what would be observed offline, if it **were possible to observe these behaviors at scale offline**

It is rarely possible, of course, to observe these behaviors at scale offline

# Social media



Finally, also when representativeness is actually a real concern (such as in the attempts of doing an electoral forecasting from social media data for example), there is clear value in social media as a **complement to survey data**, both as an early indicator of changes in public opinion and a possible signal on un-pollled topics or areas