

WORKSHOP: SOCIAL MEDIA DATA

Taehee Kim

University of Oldenburg

19th Nov 2021, University of Konstanz

TAEHEE KIM

- Post-doc researcher
- Ph.D. Political Science, Keio University, Japan
- M.Sc. Social and Economic Data Analysis, University of Konstanz
- Hate speech spread in social media
- Conspiracy theory and extremists' movement in social media

Digital trace data workshop

- Gesis 2020 (R), [Github rep.](#)
- ECPR winter school 2019, 2020 (Python)

FELIX LORENZ

- Studying Politics- and Administration in the Master's programme.
- Currently writing my Master Thesis on whether and how Tweets of candidates can impact the media agenda during election times.
- I'm working as a student assistant for the MethodsHub under Thomas Woehler at the Cluster

THE DIGITALIZATION AND DATA REVOLUTION FOR THE SOCIAL SCIENCES

- Nowadays, diverse kinds of digital trace data with a large volume have become available to researchers.
- It has a great potential for new approaches to social science questions.
- Vast textual corpora
- Alternatives to telephone or face-to-face surveys
- Users's interactions with various digital devices or services

AVAILABLE SOURCES



- Twitter: APIs are available, open with some restrictions. Provides rich data, text, network, images, geo tag etc.
- Facebook: Largely restricted unless you work with them.
- Instagram: Great for research using images. You need to scrape.
- Youtube: APIs are available, open.
- Online communities, e.g., reddit, Wikipedia etc..



This workshop introduces Twitter API and provides multiple examples of collecting and analyzing Twitter data in R. More concretely we work on extracting meaningful information from Twitter data, exploring retweet network, conducting simple text analysis.

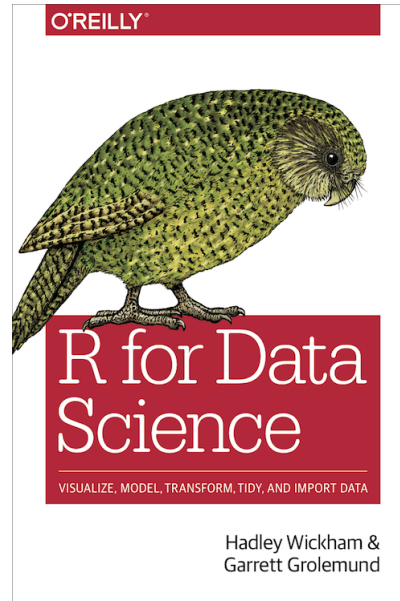
THE LANGUAGE



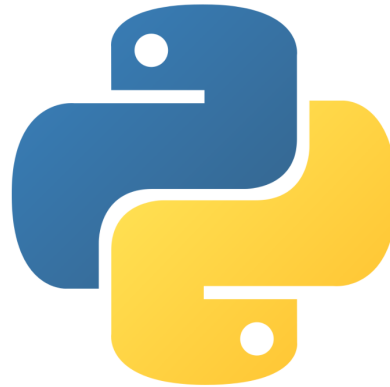


...with **rtweet** package.

PREPARATION



- R and RStudio in your laptop
- Base R + tidyverse
- Wickham and Grolemund. 2017. [R for Data Science](#). O'Reilly, especially **Data transformation** chapter.
- For other self-learning resources, check [Github rep](#).



In case you want to use Python..

Juergens, Pascal and Andreas Jungherr. 2016. "A Tutorial for Using Twitter Data in the Social Sciences: Data Collection, Preparation, and Analysis.

[SSRN2710146](#)



- Python for data collection and management
- R for analysis and visualization
- Juergens, Pascal and Andreas Jungherr. 2016.
- ZC Steinert-Threlkeld. 2017. Twitter as Data. Cambridge University Press

TWO DAY (12 HOURS) WORKSHOP

17/18 FEBRUARY 2022

	17th	18th
9:00 - 10:30	Session 1	Session 5
10:45 - 12:15	Session 2	Session 6
Lunch Break		
13:00 - 14:30	Session 3	Session 7
14:45 - 16:15	Session 4	Session 8

Room: Y213



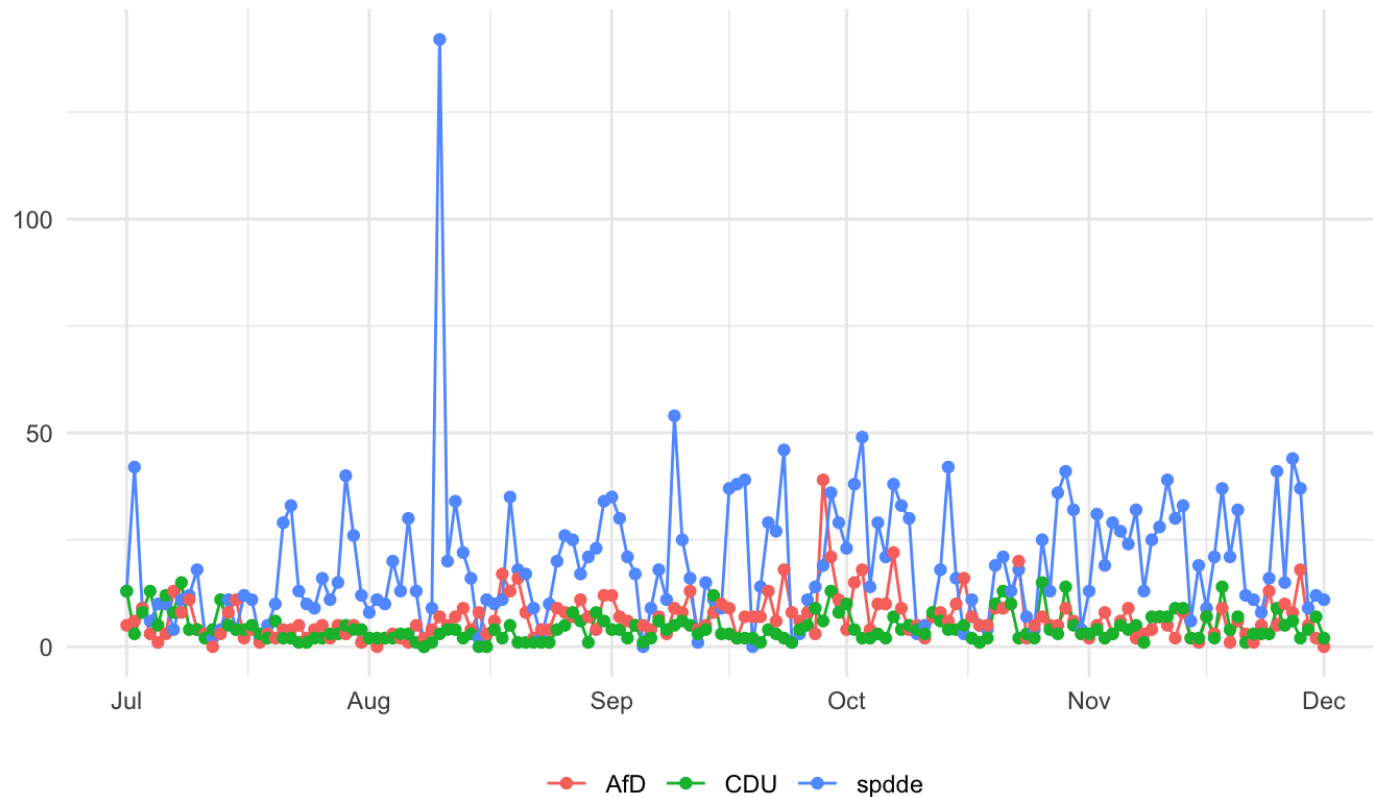
Session 1. Introduction to Twitter APIs

Session 2. Collecting Tweets 1: Rest APIs

Session 3. Collecting Tweets 2: Streaming APIs

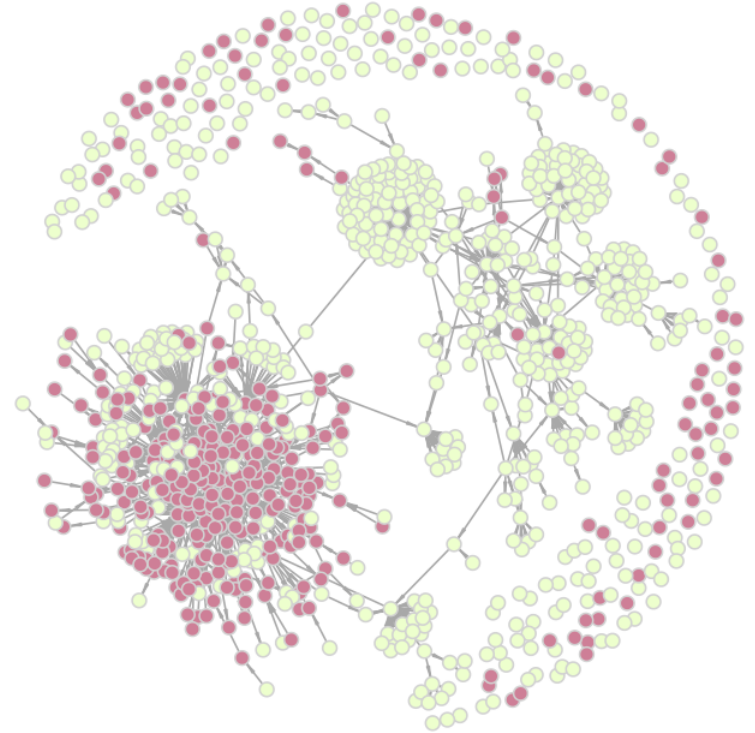
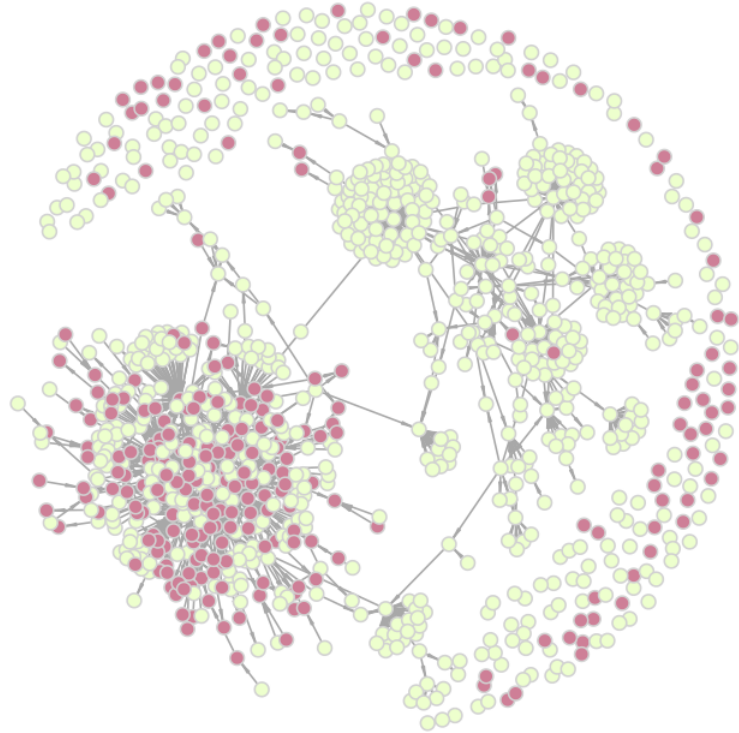
Frequency of Twitter statuses posted by AfD, CDU and SPD

Twitter status (tweet) counts aggregated by day from July

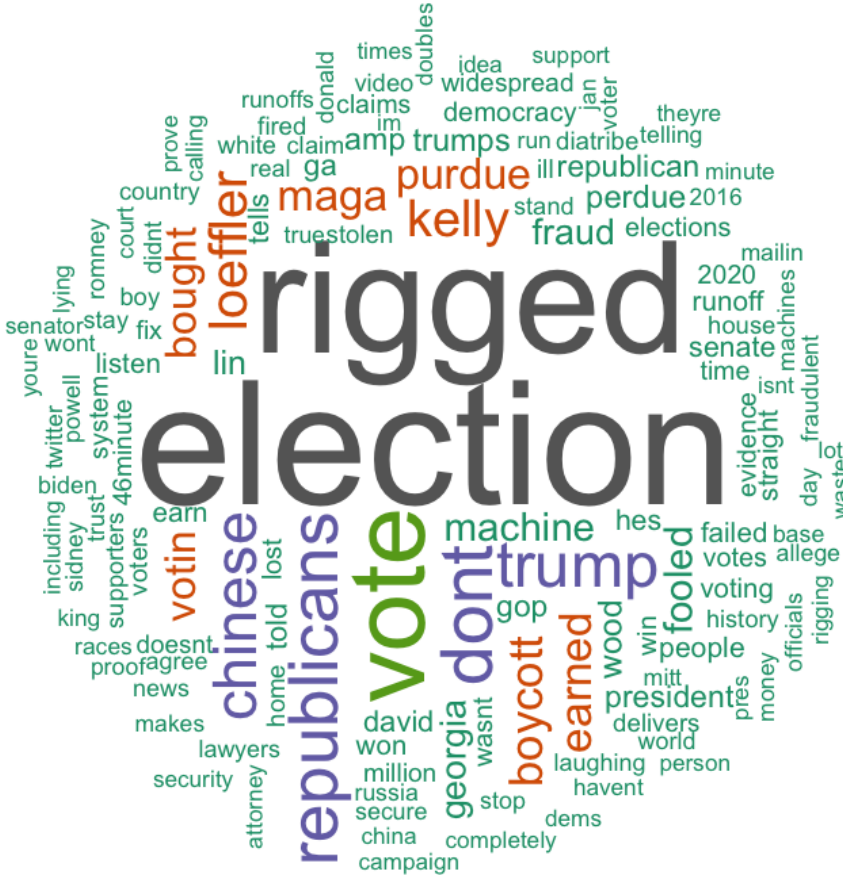


Source: Data collected from Twitter's REST API via rtweet

Session 4. Exploring Twitter data (summary statistics)



Session 5. Twitter user retweet network: Centrality, Cluster detection



Session 6. Simple text analysis: Word cloud, Topic model

- Session 7. Youtube: What we can collect and how we can utilize
- Session 8. Discussion about quality of digital trace data

WORKSHOP GITHUB REPOSITORY

You can download materials from here:

<https://github.com/thkim0321/konstanz2021socialmedi>

TWITTER'S API

" APIs are the way computer programs talk to each other so that they can request and deliver information. ([About Twitter's APIs](#)) "

You need to create Twitter applications, so called 'apps', which you need to access Twitter APIs. There are two options for you...

1. Prepare your own developer account
2. Send me your @handle of Twitter account by **10th February**

[More detail](#)

- + We will use Standard APIs v1
- + Do not use [Academic Research Access](#)