«Critical Social Media Analysis using Mixed Methods»

Data collection

Dr. Simon David Hirsbrunner, Michael Tebbe
Human-Centered Computing, Institute of Computer Science
Freie Universität Berlin

Session III, 19 Nov 2020

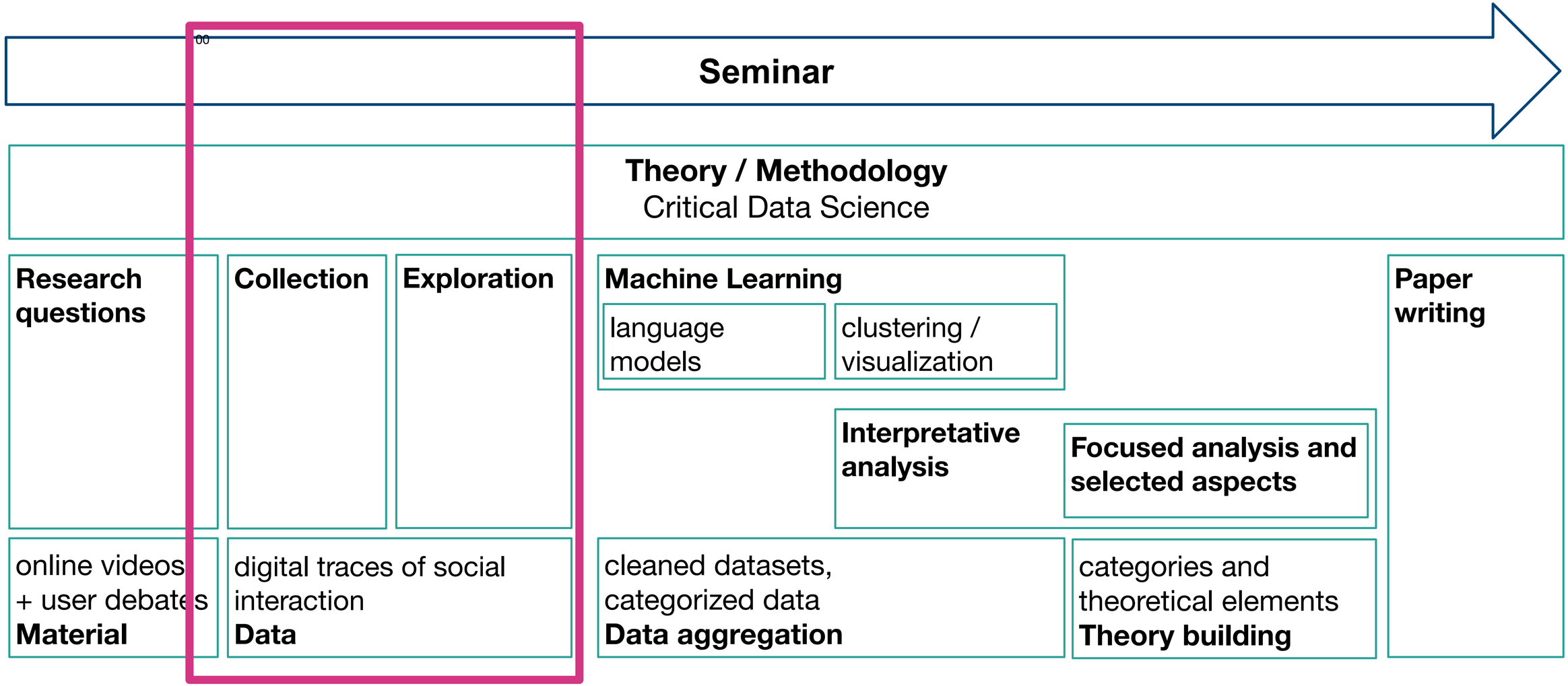
Recap last session

- What can(t) social media data (SMA) tell us about human behaviour?
  → SMD as digital traces of human interaction
  → SMD analysis as search and reading of traces

- Alignment of data, social categories (theory) and methods
  - Example of identifying media practices

- Controversy mapping, and different modes of inquiry in the analysis of social networks, digital platforms and media practices
  - Precautionary mode
  - Affirmative mode
  - Empiricist mode

Seminar progress / today



Plan for today

- Discuss the assignment

- Guidelines and preparatory steps for online research

- Tools for data collection
  - YouTube Data Tools
  - YouTube API
  - other platforms and tools

- Situating debates and data exploration with Gephi

- Collaborative collection of ideas and meeting peers

- Assignments

Controversy mapping

Assignment: how did it go? What was most challenging for you?

Epistemological precautions

Browser

- Clear cookies and other website data
- Or set up a research browser. https://youtu.be/bj65Xr9GkJM

→ However: how realistic is a “clean and depersonalized browser” and blind social media experience? Cultural embeddings of contemporary media experiences (co-curated by platform algorithms)

User profile

- Set up a research account

APIs

- Use APIs

Ethical and legal precautions

Things to consider

- Scientific soundness → open science (reproducibility, accountability, interpretability)
- Legal and ethical rights of data subjects, data privacy, GDPR
- platform guidelines (Terms of Use)
- challenges of online research in times of distributed, digital infrastructures
- Anonymization or pseudonymization sometimes not possible in qualitative research

Recommendations

- use open source software by default
- store data locally or on university server per default
- don’t share your data more than necessary for the research

More information: Association of Internet Researchers. „Internet Research: Ethical Guidelines 3.0“, 6. Oktober 2019. https://aoir.org/reports/ethics3.pdf

Don’t forget: take care of your own safety!

Digital Methods: research and tools

Digital Sociology / Media Research

→ see resources on Github / last slide of this presentation

Tool collections

- Digital Methods Initiative: https://wiki.digitalmethods.net/Dmi/ToolDatabase
- Center for Interdisciplinary Methodologies: http://blogs.cim.warwick.ac.uk/issuemapping/tools/
- SciencePo Médialab: https://medialab.sciencespo.fr/en/tools/

DMI Tools:

- YouTube Data Tools
- netvizz (Facebook): working anymore
  https://theconversation.com/facebooks-data-lockdown-is-a-disaster-for-academic-researchers-94533
- Twitter Capture and Analysis Toolset (TCAT)
- Instagram Tagnet Explorer
- ...

YouTube Data API

- one of a few APIs by YouTube (others: Analytics, Reports, Live Streaming)
- Resources: Channels, Comments, Users, Playlists…
- Intended Use Cases: Channel Management, Bots, Embedding Videos
- Issues:

1. only 100 items (comments) per request
2. Relevance sorting intransparent

https://developers.google.com/youtube/v3

YouTube Data API

- REST API (Create, Read, Update, Delete)
  - OAuth2 / API Key
  - send GET- and POST-Requests
  - Example: **CommentThreads: list**

https://www.googleapis.com/youtube/v3/commentThreads
?part=snippet,replies
&maxResults=100
&textFormat=plainText
&order=relevance
&videoId=aR5N2Jl8k14
&key=AIzaSyOHJd7d-Ho_AWr_OPqg


```
{
 "kind": "youtube#commentThreadListResponse",
 "etag": "G06jBwt1bH-UpjI102zqs41imrk",
 "nextPageToken":
"QURTSI9pMERySHBRaVNidFMxZWZIcE9oYkVCYUljV2doWHB2M0lHbWIlcFg4R0VKWkI1bmlVLVND
V0tPa2R2T0R1eVB0NHJYOTZjNXBoUQ==",
 "pageInfo": {
  "totalResults": 100,
  "resultsPerPage": 100
 },
 "items": [
  {
   "kind": "youtube#commentThread",
   "etag": "aKnEelfjKcze3ygrAyZbuDPb6Hs",
   "id": "Ugz0t8jjtWnDLZH4z2R4AaABAg",
   "snippet": {
    "videoId": "aR5N2Jl8k14",
    "topLevelComment": {
     "kind": "youtube#comment",
     "etag": "LLuo86m6QJJmx7usIXt8UtvmL8E",
     "id": "Ugz0t8jjtWnDLZH4z2R4AaABAg",
     "snippet": {
      "videoId": "aR5N2Jl8k14",
      "textDisplay": "Crazy and blinded scientists with their work...",
      "textOriginal": "Crazy and blinded scientists with their work…",
      "authorDisplayName": "Tadas K.",
      "authorProfileImageUrl":
"https://yt3.ggpht.com/ytc/yl08O57jheH4rgA=s48-c-k-c0xffffffff-no-rj-mo",
      "authorChannelUrl": "http://www.youtube.com/channel/UCM2ibvIskzL_qw55FwFA9VQ",
      "authorChannelId": {  "value": "UCM2ibvIskzL_qw55FwFA9VQ"  },
      "canRate": true,
      "viewerRating": "none",
      "likeCount": 0,
      "publishedAt": "2020-11-18T11:14:19Z",
      "updatedAt": "2020-11-18T11:14:19Z"
     }
    },
    "canReply": true,
    "totalReplyCount": 0,
    "isPublic": true
   }
  }
…]}
```

YouTube Data Tools

“A collection of simple tools for extracting data from the YouTube platform via the YouTube API v3.”

- YTDT modules: channel info, channel network, video list, video network, **video info and comments**

YTDT Tutorial by DMI on: https://youtu.be/sbErTW2MzCY

Rieder, Bernhard (2015). YouTube Data Tools (Version 1.11) [Software]. Available from https://tools.digitalmethods.net/netvizz/youtube/.

Data exploration

Gephi: https://gephi.org

Network analysis

- data exploration
- statistical analysis
- communication

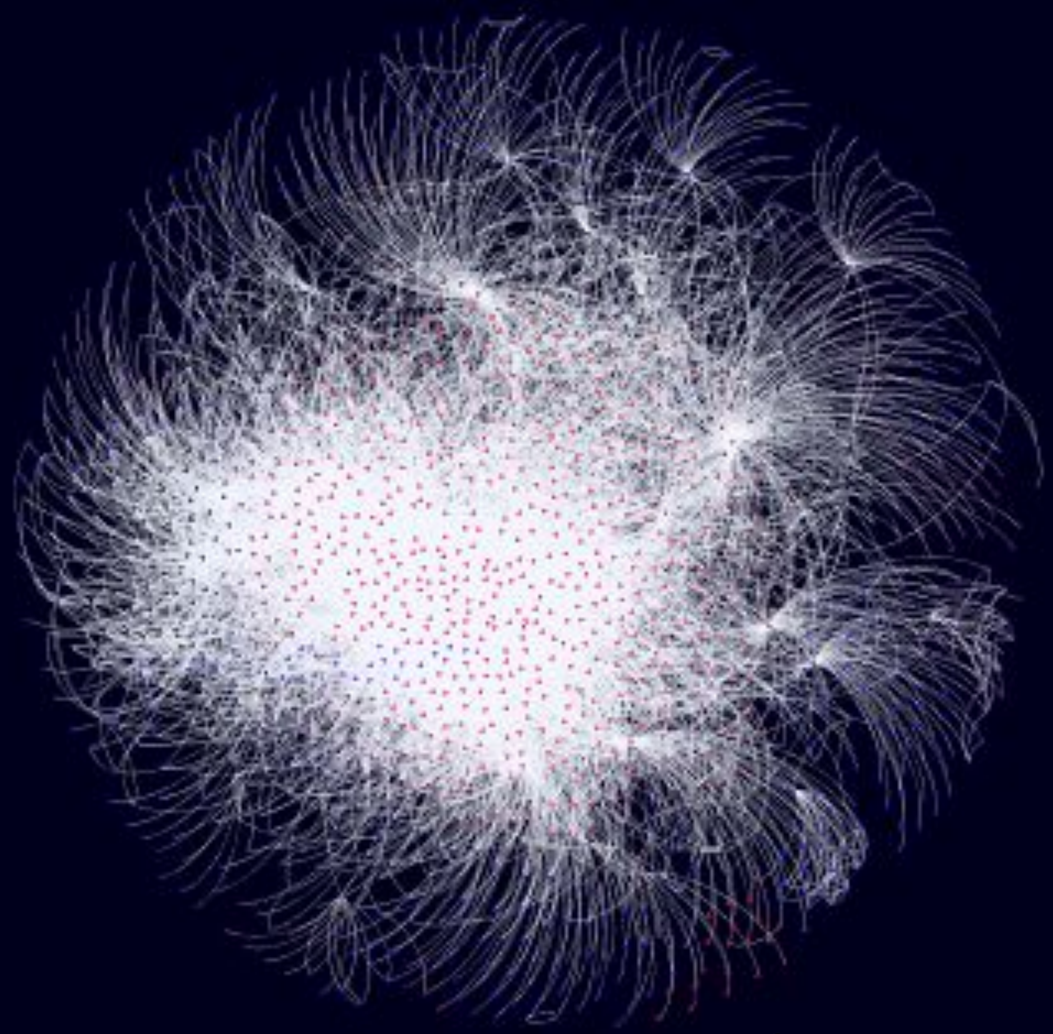


Image: YouTube channel network starting from search term ‘klimalüge’ extracted with the YTDT channel network module

Collaborative brainstorming and meeting peers

Go to: https://flinga.fi/s/FL249B5

Flinga is a tool for collaborative brainstorming and visualization

Tasks for collaborative collection of ideas and networking

- Use square post-it's to write prospective topics for further investigation and locate them on the canvas.
  - You can frame as question, topic or approach.
  - Positioning of post-its according to Mapping Controversy modes of inquiry
- Use the people symbol to create an avatar for you and add your name. Position yourself near one topic you find interesting
- If several people gather around one topic, we create a breakout room in WebEx, so you can discuss further and meet your peers. Instructors will drag a circle with the specification of the breakout room near your group.
- Join the indicated breakout room on Webex and discuss your subject.
  - Exchange email-addresses if you would like to collaborate in the future (e.g. for assignments and the seminar project)
  - Create an etherpad to document your discussions: https://pad.spline.inf.fu-berlin.de/

Assignments for next week

1 Reading assignment

- Read Paper: Pfeffer, Jürgen, Katja Mayer, und Fred Morstatter. „Tampering with Twitter's Sample API". EPJ Data Science 7, Nr. 1 (Dezember 2018): 1–21. https://doi.org/10.1140/epjds/s13688-018-0178-0.
- Answer questions on paper in Whiteboard:
  - What is the main problematic technical feature of the Twitter API described by the researchers?
  - How may this feature concretely influence the process and results of social media analysis (according to the researchers)

2 Analytical assignment

- Extract the data of one YouTube debate with the YTDT video list module https://tools.digitalmethods.net/netvizz/youtube/
- Explore your data:
  - Open your data as a spreadsheet
  - Sort the collected comments by:
    - Randomly (e.g. by User Names)
    - Date
    - Likes
- Answer the following question in Whiteboard
  - How do different sortings affect your interpretation of the data?
  - Compare to the first time you analyzed the comments directly on YouTube. They were sorted by relevance. How did this affect your understanding of the debate at hand?

What’s up next session?

Language models!

More resources

Association of Internet Researchers. „Internet Research: Ethical Guidelines 3.0“, 6. Oktober 2019. https://aoir.org/reports/ethics3.pdf

Bastos, Marco, und Shawn T. Walker. „Facebook’s Data Lockdown Is a Disaster for Academic Researchers“. The Conversation, 11. April 2018. http://theconversation.com/facebooks-data-lockdown-is-a-disaster-for-academic-researchers-94533.

Borra, Erik, und Bernhard Rieder. „Programmed Method: Developing a Toolset for Capturing and Analyzing Tweets“. Aslib Journal of Information Management, 19. Mai 2014. https://doi.org/10.1108/AJIM-09-2013-0094.

Bounegru, Liliana, Jonathan Gray, Tommaso Venturini, and Michele Mauri. “A Field Guide to ‘Fake News’ and Other Information Disorders.” SSRN Electronic Journal, 2018. https://doi.org/10.2139/ssrn.3097666.

Pfeffer, Jürgen, Katja Mayer, und Fred Morstatter. „Tampering with Twitter’s Sample API“. EPJ Data Science 7, Nr. 1 (Dezember 2018): 1–21. https://doi.org/10.1140/epjds/s13688-018-0178-0.

Rogers, Richard. Digital Methods. Cambridge, Massachusetts: The MIT Press, 2013.

———. “Digital Methods for Web Research.” In Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource, edited by Robert A Scott, and Stephen M Kosslyn, 2015.

———. Doing Digital Methods. 1. Edition. Thousand Oaks, CA: SAGE Publications Ltd, 2019. http://proxy.library.cornell.edu/login?url=http://onlinelibrary.wiley.com/book/10.1002/9781118900772.

Venturini, Tommaso, Liliana Bounegru, Jonathan Gray, and Richard Rogers. “A Reality Check(List) for Digital Methods.” New Media & Society 20, no. 11 (November 1, 2018): 4195–4217. https://doi.org/10.1177/1461444818769236.