HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Why not to use popularity scores from platforms

The hidden biases of YouTube data

Merja Mahrt

General Online Research, March 6-8, 2019

Biases in platform data

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Issues in CSS / big data literature

- Meaning of "likes," "retweets," etc. unclear (boyd & Crawford, 2012; Lomborg & Bechmann, 2014; Ruths & Pfeffer, 2014)
- "Hidden biases" due to social inequalities of usage (Crawford, 2013)
- Intransparent algorithms in platforms
  - E.g., Twitter's "trending" algorithm (Gillespie, 2012)

Biases in platform data

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Background: Research on digital fragmentation

- Fears about "echo chambers" (Sunstein, 2007) and "filter bubbles" (Pariser, 2011)
- Main factors
  - More content available online
  - More selective user behavior
  - Selection through platform design
- How to determine the **reach of online content**?

Digital fragmentation

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

A multi-method approach

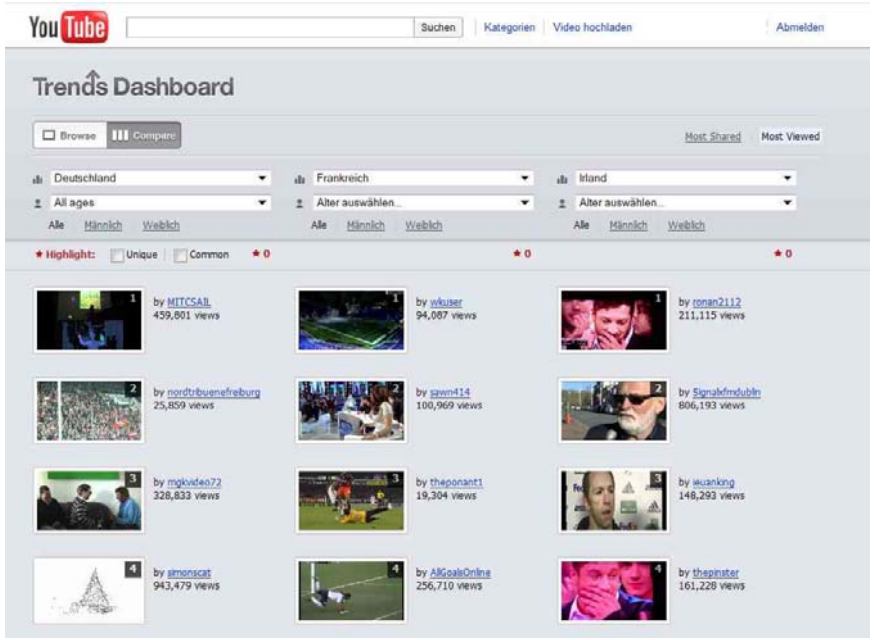
- Platform data
  - Content structures
  - Popularity scores
- Usage data
- Survey

Popularity of YouTube content

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Platform information: Popularity scores

- Content analysis
  - YouTube daily top 10
  - “Trends” in Germany
    - Trends Dashboard (now defunct)
  - March through August, 2014
- 1,164 unique videos



Top 10 YouTube videos

1,165 videos from German top 10, 03-08/2014

Digital fragmentation

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Popularity of YouTube content

- Survey
  - Online access panel (Respondi)
  - First week of September, 2014
  - 1,665 respondents, aged 18-69
  - Representative for German online users
- YouTube use
  - Reach of online video platforms
  - Recognition of popular YouTube videos

Popularity of YouTube content

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Survey

- Video platforms highly popular
  - 50% of online users access video platforms at least once per week
  - [ARD/ZDF Online Study 2014: 45%; Koch & Liebholz, 2014]
- (Almost) daily users of video platforms
  - +24% aged 18-29
  - +11% male
  - +13% highschool graduates [Abitur]

Popularity of YouTube content

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Survey

- Popularity of YouTube videos

Popularity of YouTube content

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

YouTube usage behavior

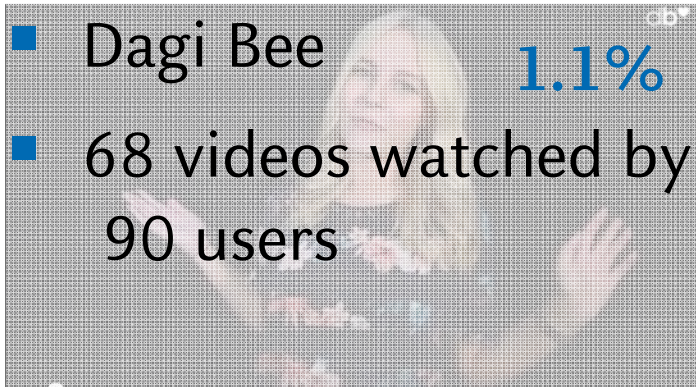
- Clickstream data
  - Nielsen panel of 54,790 German users
  - YouTube usage in June, 2014
    - 8,147 users
    - 433,235 video views
    - 244,925 unique videos
      - 87% viewed by one user each
    - 109,093 channels
      - 80% viewed by one user each

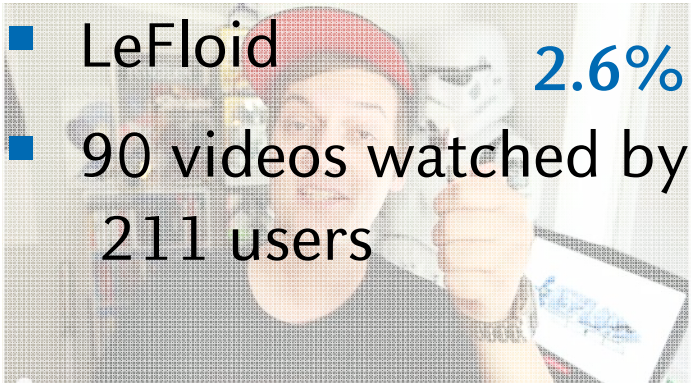
Popularity of YouTube content

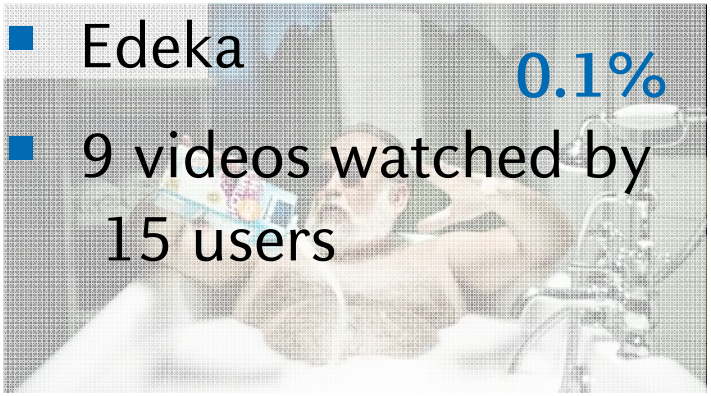
HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Clickstream data

- Popularity of YouTube channels, June, 2014





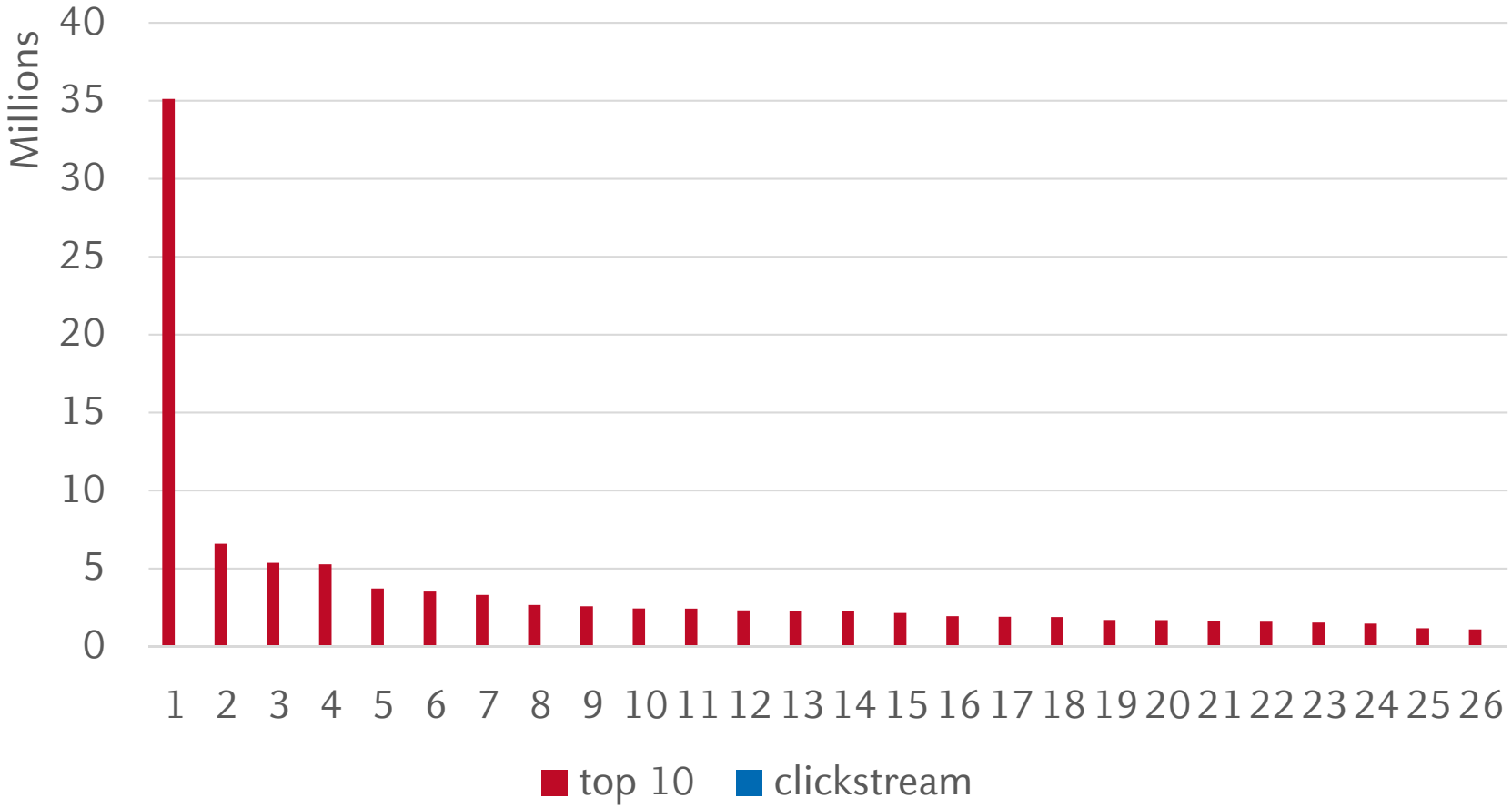


Popularity of YouTube content

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Comparing platform and clickstream data

- 26 videos from top 10 with 1 million+ views (June, 2014)

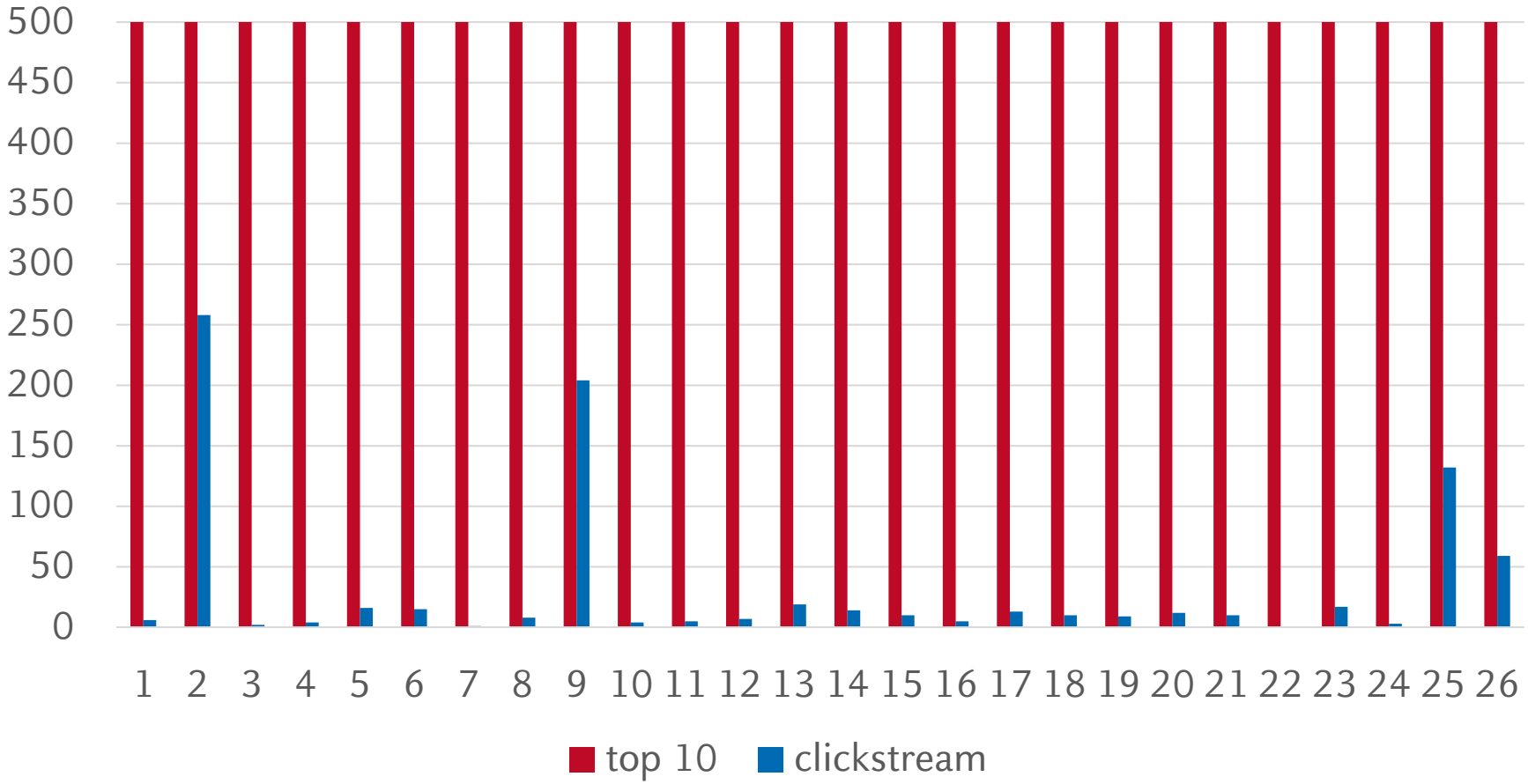


Popularity of YouTube content

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Comparing platform and clickstream data

- 26 videos from top 10 with 1 million+ views (June, 2014)



Conclusion

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Why not to use popularity scores

- Popularity scores and clickstream data capture different aspects of YouTube "use"—with different biases
- Over- and underrepresentation (age)
  - Validity of age in clickstream data doubtful
  - Small (sub)samples and idiosyncratic YouTube preferences
  - Limitations of assessing use on content level with survey
- (How) does YouTube favor advertisers and/or influencers?

- Which dataset captures reach of YouTube content best?
  - ...to ultimately assess digital fragmentation?

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Thank you

Merja Mahrt

mahrt@hhu.de

Out now: *Beyond filter bubbles and echo chambers. The integrative potential of the Internet.* Berlin: Digital Communication Research.
www.digitalcommunicationresearch.de/v5/

References

boyd, d., & Crawford, K. (2012). Critical questions for big data. Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society, 15*(5), 662-679.

Crawford, K. (2013, April 1). The hidden biases in big data. *Harvard Business Review*. Retrieved from https://hbr.org/2013/04/the-hidden-biases-in-big-data

Gillespie, T. (2012). Can an algorithm be wrong? *Limn*, *n.v.*(2). Retrieved from http://limn.it/can-an-algorithm-be-wrong/

Koch, W., & Liebholz, B. (2014). Bewegtbildnutzung im Internet und Funktionen von Videoportalen im Vergleich zum Fernsehen. *Media Perspektiven, n.v.*(7-8), 397-407.

Lomborg, S., & Bechmann, A. (2014). Using APIs for data collection on social media. *The Information Society*, *30*(4), 256-265.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. London, UK: Penguin.

Ruths, D., & Pfeffer, J. (2014). Social media for large studies of behavior. *Science, 346*(6213), 1063-1064.

Sunstein, C. R. (2007). *Republic.com 2.0*. Princeton, NJ: Princeton University Press.