Opinion Polarization on COVID-19 Measures: Integrating Surveys and Social Media Data

Markus Reiter-Haas, Beate Klösch, Markus Hadler and Elisabeth Lex
Institute of Interactive Systems and Data Science, Graz University of Technology and Department of Sociology, University of Graz

Motivation

Polarization in public opinion is a major issue for societies as high levels can promote adverse effects such as hostility and the spread of misinformation [1].

Opinion Polarization ~ Dispersion of Opinions

Survey Research

- Directly measures opinions using agreement scores
- Statistics of distribution [2]

Social Media Research

- Needs to extract opinions
- E.g., via content-based measures from text
- Often analyzed along predefined groups
- E.g., political affiliations
- Alternatively: derive network-based measures
- E.g., segregation in topology [3]

Integrating survey and social media data

- Emerging field [4]
- Merging of two research lines
- Limitations on comparability

Aim and Contribution

Bridge the perspectives on polarization research

→ Introduce a framework for analysis using an integrated data source

Experiment Design

Polarization analyzed regarding

- COVID-19 prevention measures (vaccination, mask-wearing, contact tracing)
- DACH region → Germany (D), Austria (A), Switzerland (CH)
- German speaking only
- Questionnaire in summer 2020
- Representative quota sample of Internet users
- Acquire Twitter handles of respondents

Framework

Creation of multiple datasets

- 3 data sources (survey, social media, integrated)
- 2 granularities each (full, comparable subset)
- → 6 perspectives
- Manual annotation for congruence analysis

Datasets on COVID-19 prevention measures	
Representative sample of survey respondents	(n = 2,560)
Respondents that use Twitter	(n = 705)
Tweets on COVID-19 in German	(n = 90,806)
Tweets in survey period	(n = 21,479)
Integrated survey respondents and Twitter accounts	(n = 79)
Integrated who tweeted about COVID-19	(n = 20)
Additionally: manually annotated tweets of integrated	(n = 221)

Trade-off: subsets are more comparable, but less data

E.g., via population characteristics or temporal information

Processing

- Subsets created by filtering data
- Sentiment analysis for tweets
- Descriptive analyses
- Manual annotation codebook: agreement per user account

Quantifying polarization

Should be applicable to all perspectives → defined on distributions

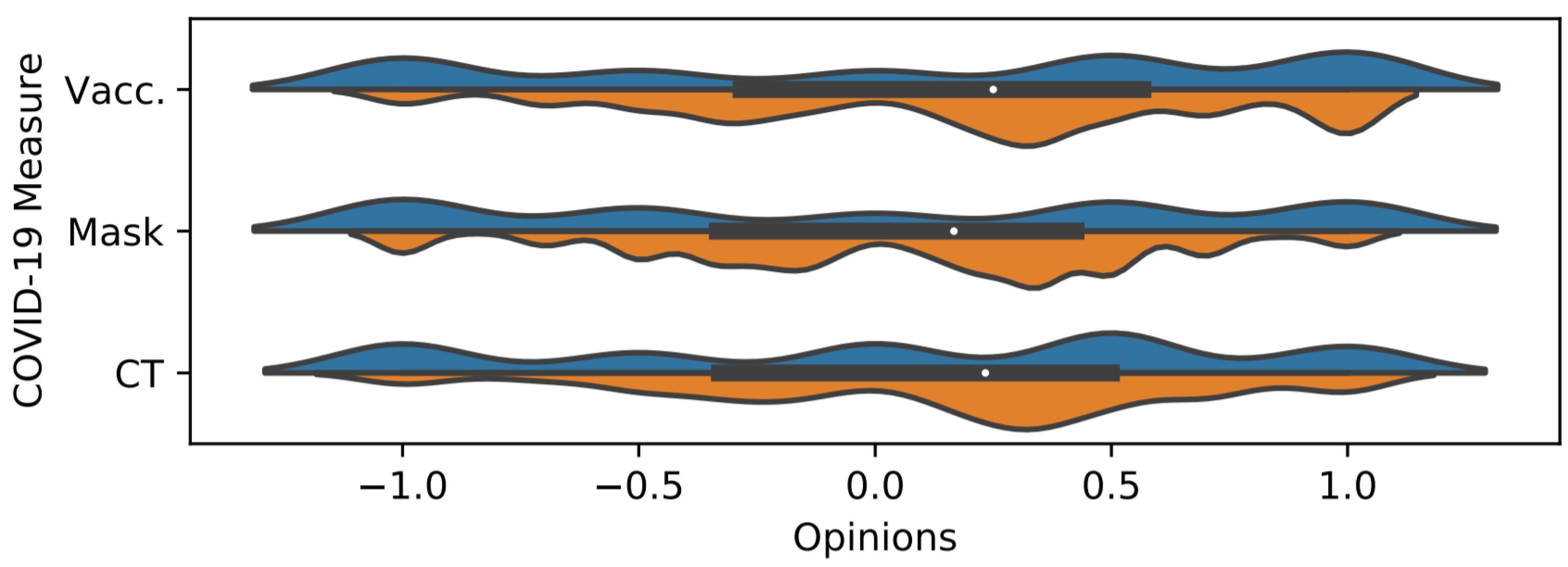
Bimodality coefficient β

$$\beta = \frac{(\gamma^2 + 1)}{\left(\kappa + 3\frac{(n-1)^2}{(n-2)(n-3)}\right)}$$

- Defined via skewness γ and excess kurtosis κ
- Sample size n as normalization factor

Findings on COVID-19 Prevention Measures

- Similarities in polarization effects between data sources
- High congruence in annotated tweets
- Vaccination is more polarizing (agreement: β = 0.67; sentiment: β = 0.49) compared to mask-wearing (β = 0.65; 0.44) and contact tracing (β = 0.59; 0.44)
- Less prevalence in Twitter users, i.e., respondents who use Twitter and tweet themselves → source of this phenomenon needs more investigation



Violinplot of opinions comparing normalized agreement in the full survey dataset (top; blue) to sentiment on the full Twitter dataset (bottom; orange). Vaccination (Vacc.) is more polarized compared to Mask wearing and contact tracing (CT) in both agreement and sentiment due to more extreme opinions skewed towards the positive side. Polarization in agreement is also higher compared to polarization in sentiment.

Conclusion and Future Work

Approach provides holistic view on polarization

Fills the gap in an emerging interdisciplinary field

Overall, similarities and congruences between survey and social media perspectives in analyzed topic

Potential biases due to consent in data collection

→ FW: study the characteristics of respondents and users

References

[1] Bessi A, Petroni F, Del Vicario M, Zollo F, Anagnostopoulos A, Scala A, et al. Viral misinformation: The role of homophily and polarization. In: Proceedings of the 24th International Conference on World Wide Web; 2015. p. 355–356.

[2] Bramson A, Grim P, Singer DJ, Berger WJ, Sack G, Fisher S, et al. Understanding Polarization: Meanings, Measures, and Model Evaluation. Philosophy of Science. 2017;84(1):115–159.

[3] Alamsyah A, Adityawarman F. Hybrid sentiment and network analysis of social opinion polarization. In: 2017 5th International Conference on Information and Communication Technology (ICoIC7). IEEE; 2017. p. 1–6.

[4] Stier S, Breuer J, Siegers P, Thorson K. Integrating Survey Data and Digital Trace Data: Key Issues in Developing an Emerging Field. Social Science Computer Review. 2020;38(5):503–516.

Route 63 project: Polarization in Public Opinion
https://socialcomplab.github.io/polarization/

TU Graz – Institute of Interactive Systems and Data Science
Inffeldgasse 16C, 8010 Graz, Austria, Tel.: +43 316 873-5624
office.isds@tugraz.at, https://www.tugraz.at/institutes/isds/home/