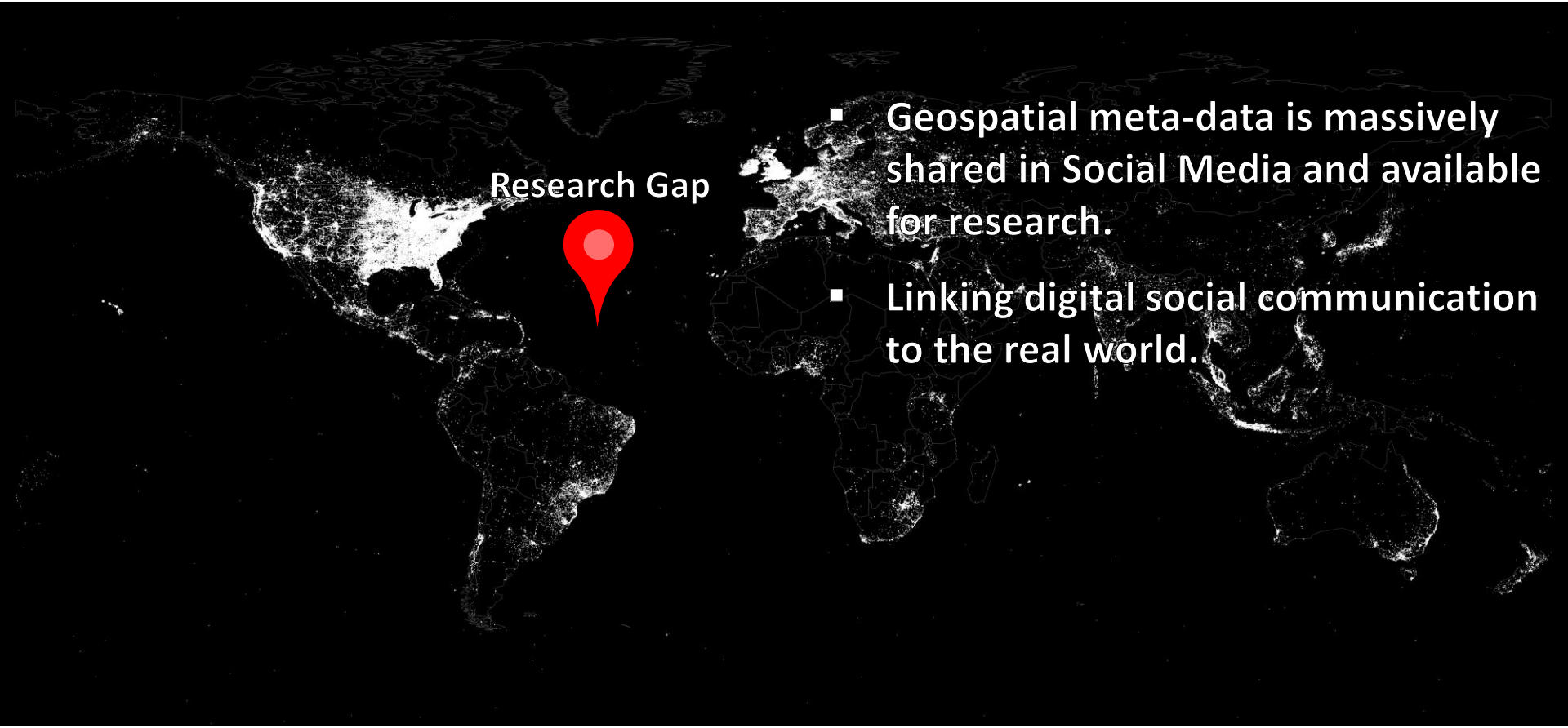
Social Sensors: Nutzung von Social Media Geotags für die Forschung

A Bridge Between Two Worlds!



- Geospatial meta-data is massively shared in Social Media and available for research.
- Linking digital social communication to the real world.

Availability of Geospatial Social Media Data

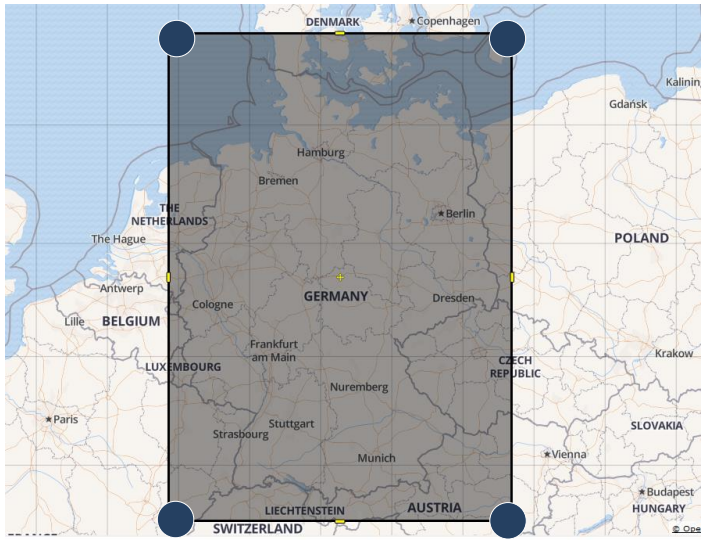
Instagram	Twitter	Facebook
~~Available Data~~ ~~• API closed for research purposes~~ ~~• Self-tagged place~~ ~~• Post coordinates~~	**Available Data** • Self-tagged place • Post coordinates • Self-tagged origin	**Available Data** • Only public data on pages, not personal timeline • Self-tagged place • Post coordinates
~~Users~~ ~~• 600M Users~~ ~~• ~5% of posts are geo-tagged~~	**Users** • 300M Users • ~5% of posts are geo-tagged	**Users** • 1.800M Users • 25% of users use geo-tads

Source: We Are Social, Facebook, Instagram, Twitter

How is the data retrieved?

Twitter Public 1% API

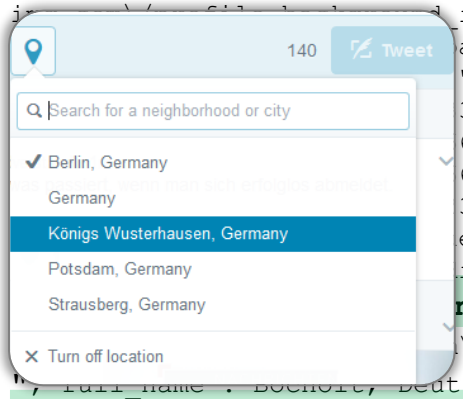
Bounding Box

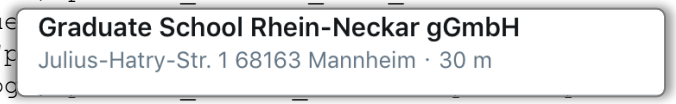


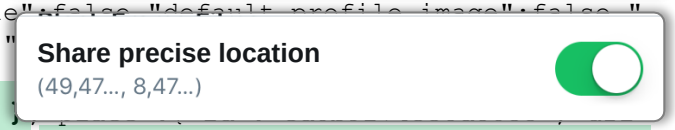
Python Script: Tweepy

Example Tweet

36%
(17% assignable)







{"created_at":"Sat Feb 25 11:57:59 +0000
2017","id":835458819257925632,"id_str":"835458819257925632
href=\"http:\/\/twitter.com\/download\/iphone\" rel=\"nofo
iPhone\u003c\/a\u003e","truncated":false,"in_reply_to_stat
n_reply_to_user_id":null,"in_reply_to_user_id_str":null,"
0260,"id_str":"239830260","name":"Yannick
Rieder","screen_name":"YRieder","location":"Mannheim.
rieder.de","description":"Music. Research HULK. Social Me
@Q_InsightAgency","protected":false,"verified":false,"foll
unt":18,"favourites_count":193,"statuses_count":406,"creat
2011","utc_offset":3600,"time_zone":"Vienna","geo_enabled":true,"lang":"de","contributors_enabled":false,
"is_translator":false,"profile_background_color":"C0DEED","profile_background_image_url":"http:\/\/pbs.tw
images\/229575451\/IMG_1557.JPG","profile_background_image_url_https":"https:
ackground_images\/229575451\/IMG_1557.JPG","profile_background_tile":false,"p
"profile_sidebar_border_color":"C0DEED","profile_sidebar_fill_color":"DDEEF6"
3","profile_use_background_image":true
6533738536960\/kVH1Ymgi_normal.jpg","p
6016653373853696O\/kVH1Ymgi_normal.jpg
39830260\/1455279564","default_profile":false,"default_profile_image":false,"
est_sent":null,"notifications":null},"
inates":
nates":[6.63926366,51.84070767]}
\/1.1\/geo\/id\/cd0b9274338d8055.json","place_type":"city","name":"Bocholt
","full_name":"Bocholt, Deutschland","country_code":"DE","country":"Deutschland","bounding_box":
{"type":"Polygon","coordinates":[[[6.472517,51.789875],[6.472517,51.917691],[6.721208,5
1.917691],[6.721208,51.789875]]]},"attributes":{}},"contributors":null,"is_quote_status":false,"re
tweet_count":0,"favorite_count":0,"entities":{"hashtags":[],"urls":[],"user_mentions":[],"symbols":[]},"f
avorited":false,"retweeted":false,"filter_level":"low","lang":"de","timestamp_ms":"1488023879242"}

Data processing

- Twitter database in CSV file format.
- Preprocessing and extracting of information and variables using statistical software packages *R* and *Stata.*
- Adding aggregated geographical levels (e.g., counties and states) to geolocalized tweets using Official Statistics *shape*-files.
- Enriching dataset with external information such as socio-economical data at county- and municipality-level.
- Visualizing maps using *R* and *carto.com.*

Application Possibilities

Locations

Identification and monitoring of places of interest and hotspots

- Local Events (e.g. demonstrations, concerts, natural disasters)
- Tourist destinations and leisure areas
- Locations of target groups

Analyzing mobility patterns and trajectories

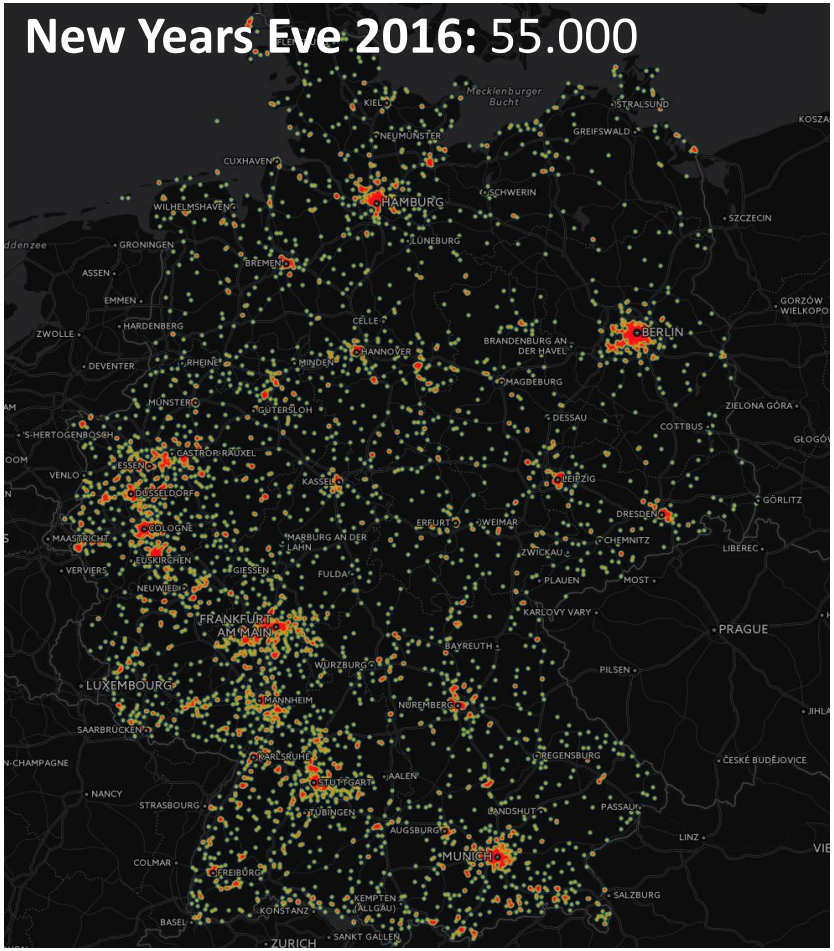
- Tracking movement of users (e.g. in a city, or on a festival site)
- Analyzing the places of residence of individuals attending specific events

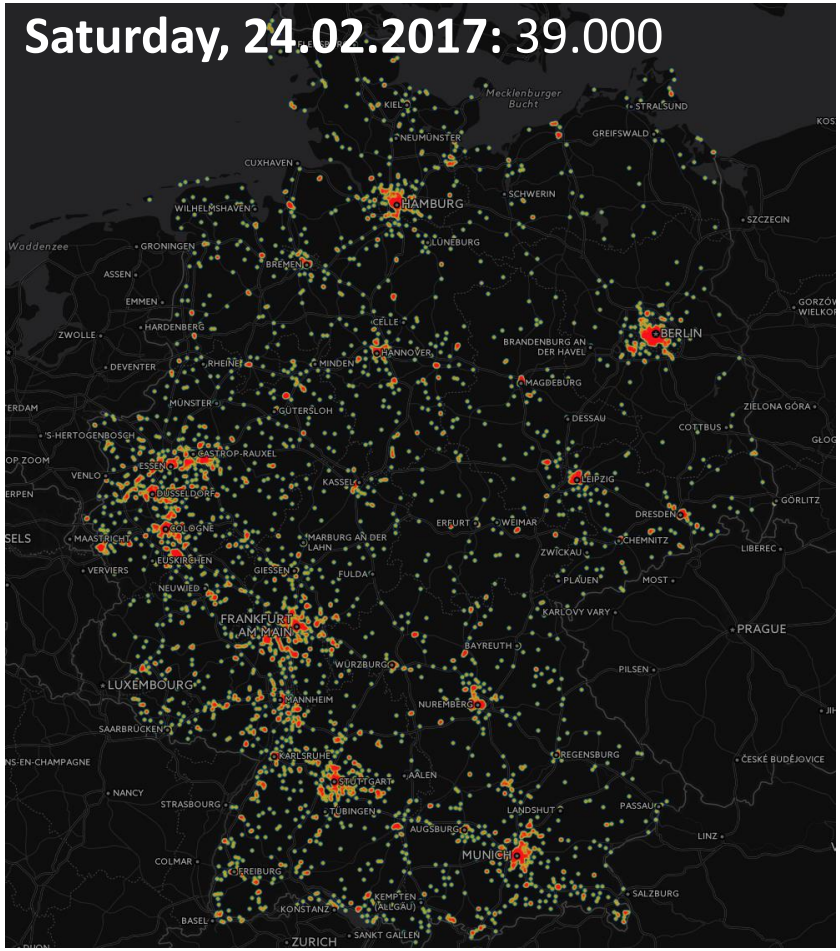
Topics / Content

Investigating the impact of geospatial, socio-economical and environmental structures on topics and opinion

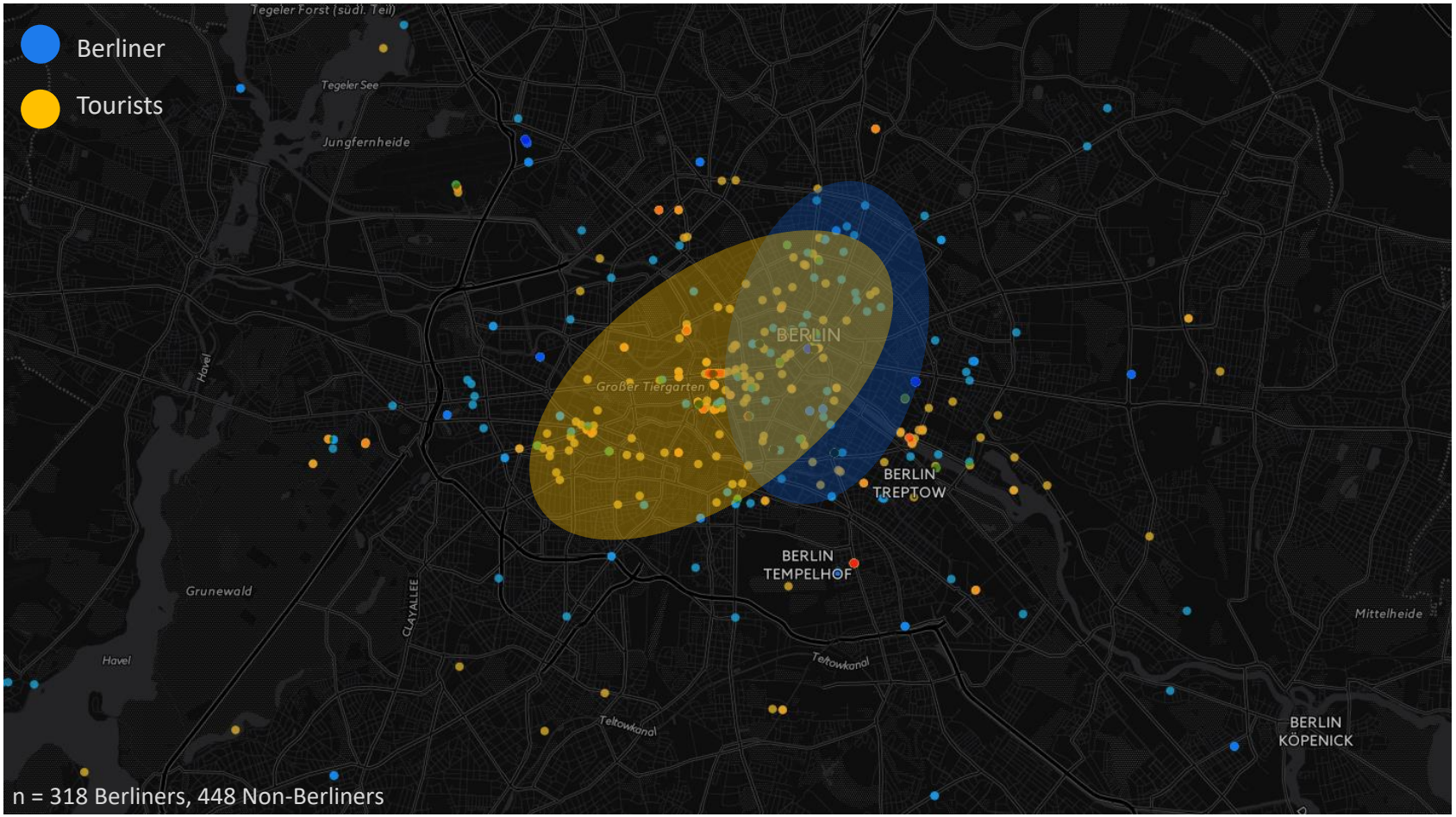
- Segment expressed opinions or attitudes across areas and locations
- Long-term influences of environmental changes on topics or opinions (construction of a freeway, opening of a park)

Buzz on New Year's Eve

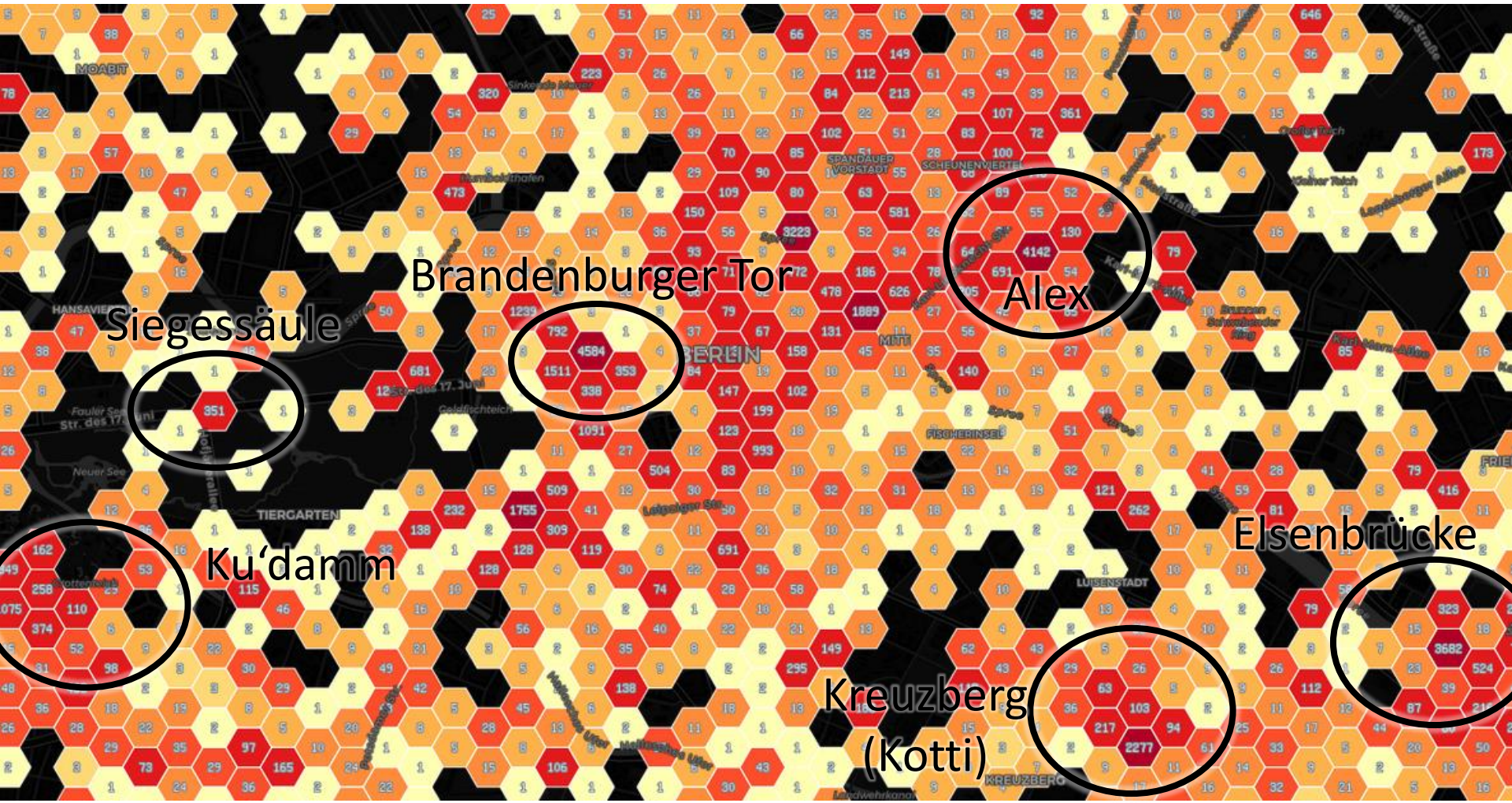




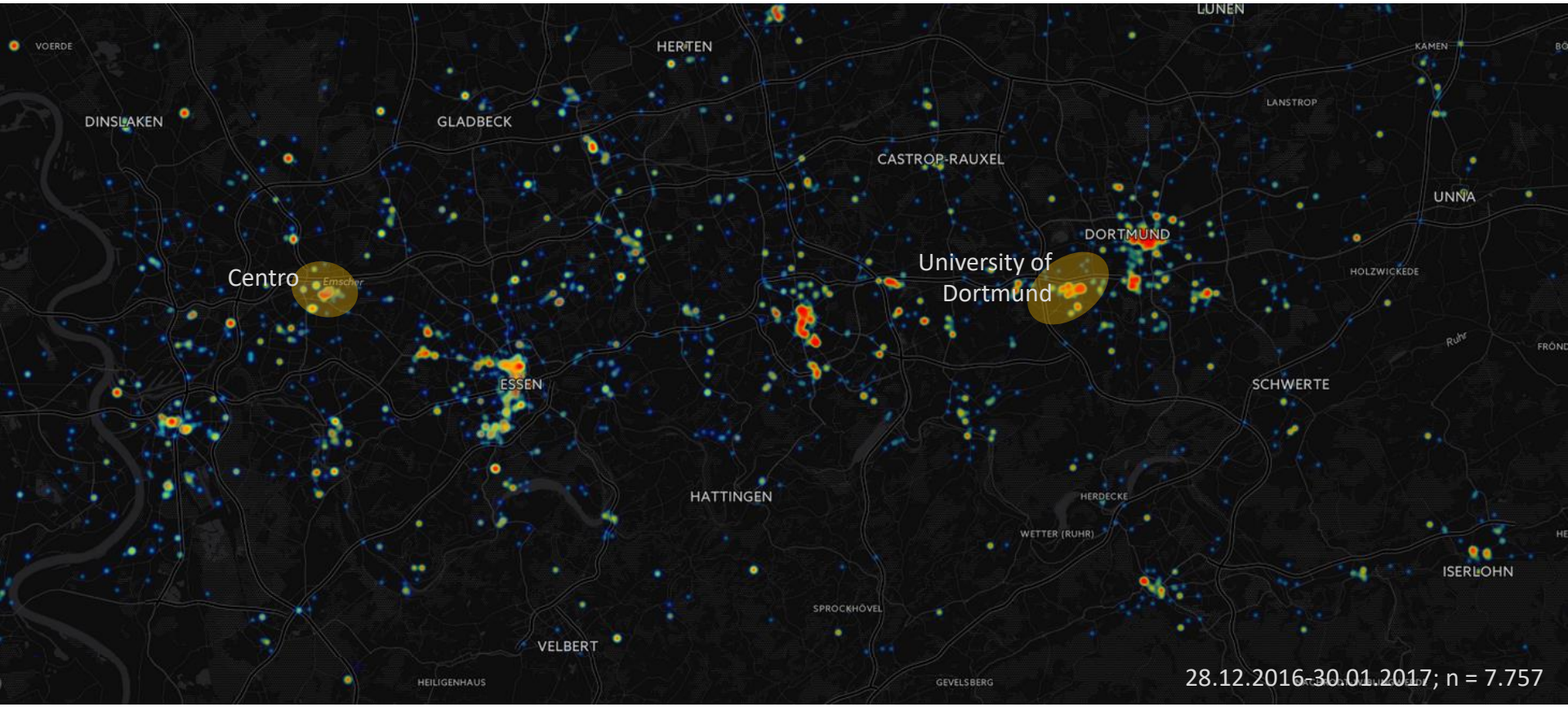
Berliners Vs. Tourists on New Year's Eve



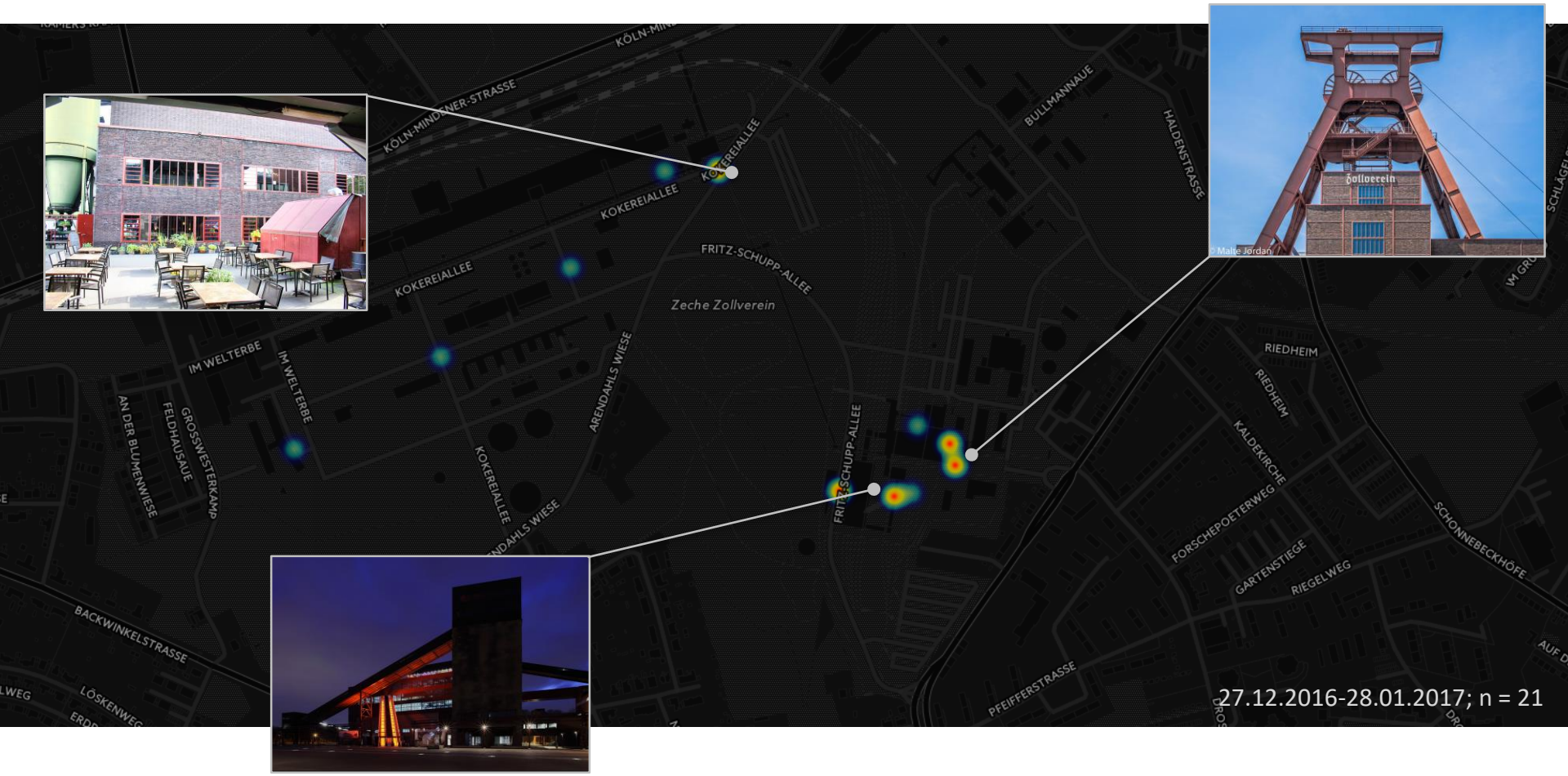
Instagram: New Year's Eve 2017



Ruhr Region: Night Vs. Day

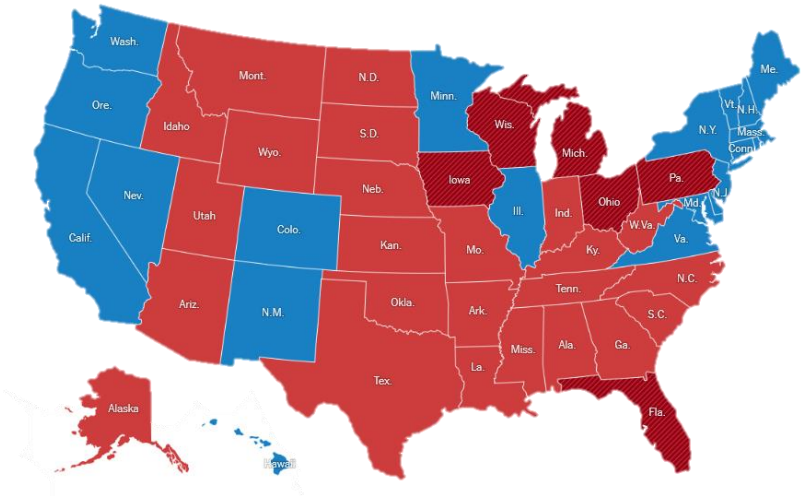


Hot-Spot Monitoring: Essen Zeche Zollverein



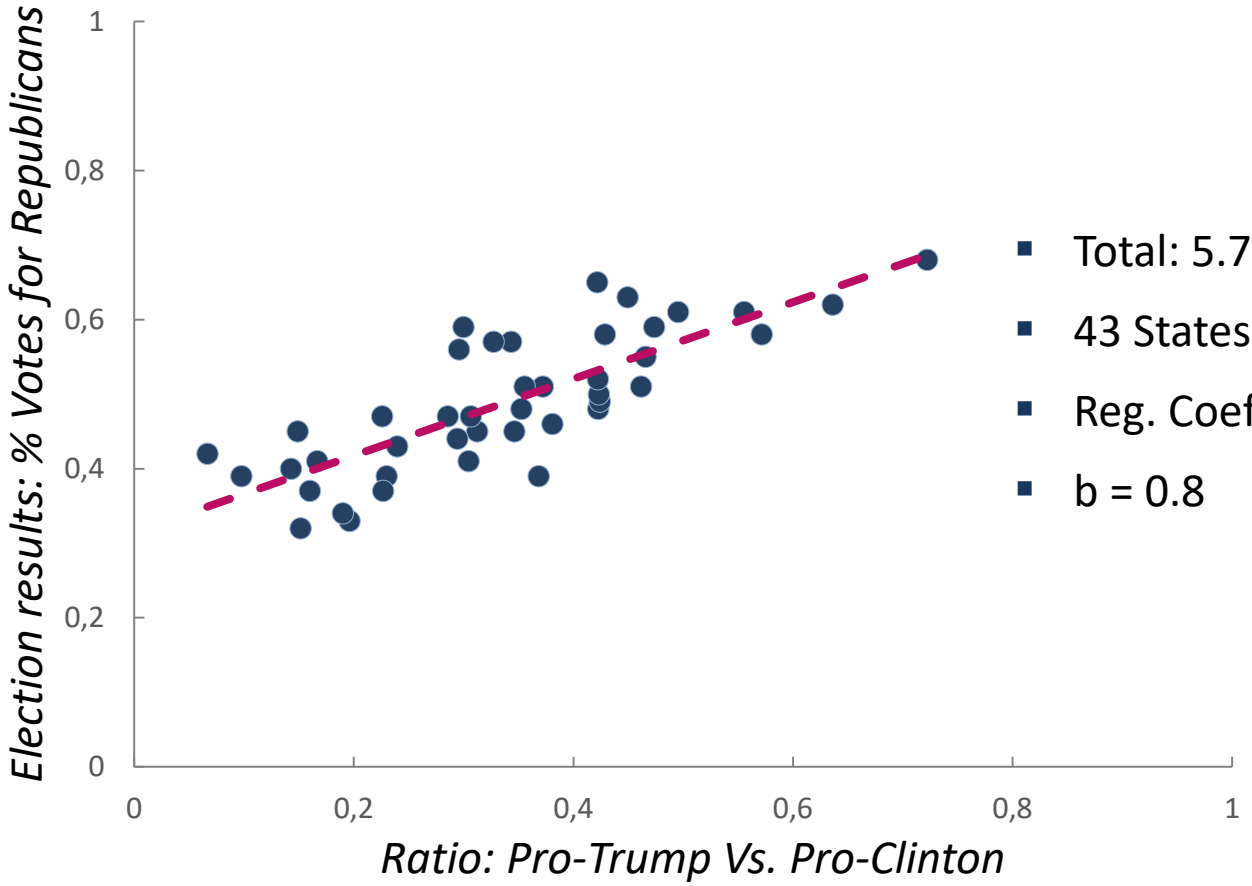
US Presidential Election 2016

- Collection of geo-localized tweets on Election Day (08.11.2016).
- Classifying users into Pro-Trump vs. Pro-Clinton by using relevant hashtags (e.g. #DrainTheSwamp, #NeverTrump).
- Comparing at state level:
  - Ratio: Pro-Trump Vs. Pro-Clinton
  - Election results: % Votes for Republicans



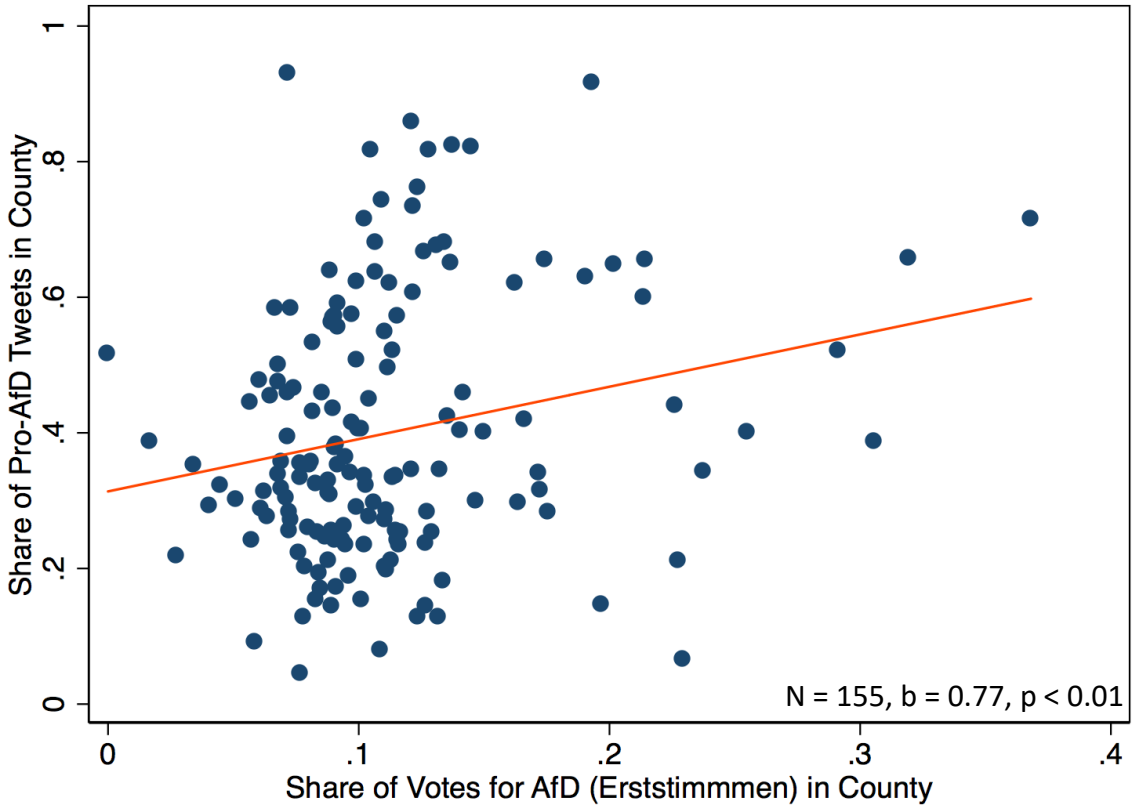
Source: NY TImes

US Presidential Election 2016



- Total: 5.724 Categorized Users
- 43 States (10-857 Users)
- Reg. Coefficient: 0,51 ($p < .0001$)
- b = 0.8

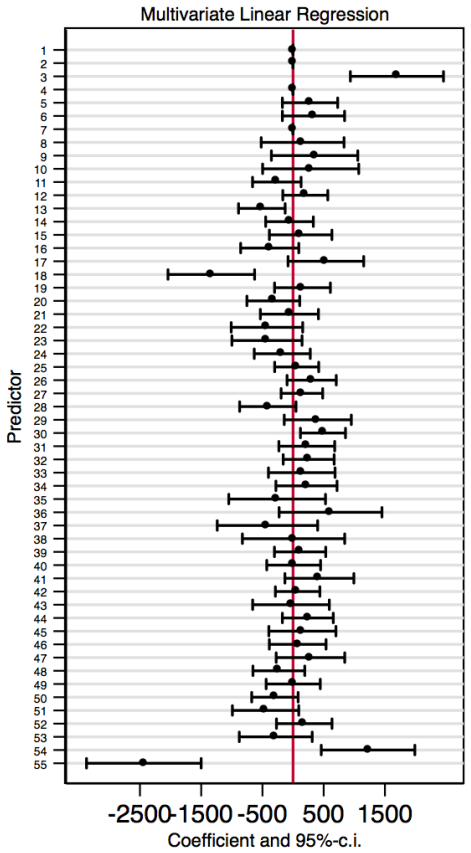
German Federal Election 2017



Regional Impact on Twitter Communication

- What types of regions are most active on Twitter?
- Adding socio-structural data at county level to our data
- INKAR Database: “Indikatoren und Karten zur Raum- und Stadtentwicklung” (Indicators and maps on spatial and urban development in Germany)
  - Variables such as dwelling type, share of students, share of foreigners, share of unemployed, building ground prizes, etc.
- Multivariate linear regression at county level (n = 390)
  DV = No. of tweets in county

Regional Impact on Twitter Communication



Adj. R^2= .66

Regression Results

- Only few variables do show a significant impact
- Obviously: More tweets in highly populated areas and cities
- Less Tweets in areas with a comparatively low share of party members ($b = -510$, $p < .01$)
- Less Tweets in areas with a high share of social benefit recipients ($b = -1334$, $p < .001$)
- More Tweets in areas with a higher share of students ($b = 488$, $p < .05$)

Summary

- Geospatial data is available and can be used for research purposes.
- Application possibilities:
  - Tracking behavior of users or user groups.
  - Longitudinal data allows to identify hotspots as well as patterns and changes over time.
  - Monitoring specific places/events.
  - Analyzing content across geospatial areas.

Limitations and Challenges

Consistency:

- Social Media is a dynamic field and thus permanently underlies changes in availability but also in usage structure and behavior.

Data availability:

- Dependency on the advent of geotagged communication.
- Filtering bots and commercial posts.

Bias:

- Unknown → Much more research needed.

Appendix

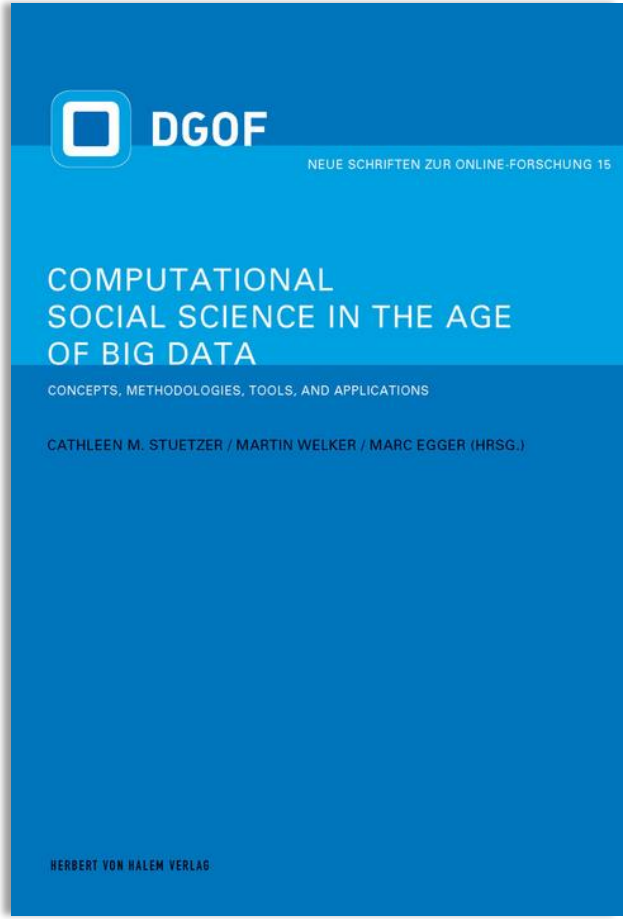
Literature

- Rieder, Y. & S. Kühne, 2018: Geospatial Analysis of Social Media Data – A Practical Framework and Applications Using Twitter. In: C.M. Stuetzer, M. Welker & M. Egger (Eds.), Computational Social Science in the Age of Big Data. Herbert von Halem Verlag, Köln.
- Rieder, Y. & S. Kühne, 2018: Analyzing Tweeters and Tweets During the 2017 German Federal Election. Presentation at the General Online Research Conference 2018 (GOR 2018), March 2018, TH Köln.
- Rieder, Y. & S. Kühne, 2017: Geospatial Analysis of Social Media Data – A Bridge Between Two Worlds. Presentation at the General Online Research Conference 2017 (GOR 2017), March 2017, HTW Berlin.

Thanks.

Yannick Rieder
Q | Agentur für Forschung
Senior Research Manager
yannick.rieder@teamq.de
@yrieder

Simon Kühne
Universität Bielefeld
Lehrstuhl für Methoden der empirischen Sozialforschung
simon.kuehne@uni-bielefeld.de



Contact

Yannick Rieder

Q | Agentur für Forschung

yannick.rieder@teamq.de

@yrieder

Simon Kühne

Universität Bielefeld

Lehrstuhl für Methoden der empirischen Sozialforschung

simon.kuehne@uni-bielefeld.de