Retweet Network of Anti-Corona Protests in Vienna '21-'22

Maximilian Paul
Marcus Röckl

Statistical Analysis of Social Networks

Freie Universität Berlin

Outline

1. Introduction
2. Graphics
3. Variable Importance and Results
4. Conclusion
5. References

Formal Problem Setting

- Tweets that were posted by using the hashtag *wXXXX* where *XXXX* indicates the respective date, i.e. 2021-11-20, 2021-12-11, 2022-01-15, 2022-02-11
- During the aforementioned dates, protests in Vienna against restrictions due to Corona pandemic took place
- To reduce noise, posts that were retweeted less than two times were excluded
- This procedure leaves us with 58,059 from originally 60,484 tweets
- and 15,815 from originally 17,725 users

Clustering

- Goal was to divide users in two groups, people who agree with protective measures against coronavirus and people who disagree (esp. vaccination)
- Different methods for clustering the data were considered, i.e.
  - *Edge betweenness* was working well on small graphs but computation time explodes for bigger networks
  - *Spin Glass* algorithm and
  - *Louvain* algorithm implemented in `igraph` package
  - *Mixed Membership Stochastic Block Model* approach from `mixedMem` failed
- Result using *Louvain* on the whole data set and conveying clusters to single protest was best
- Resolution was set to .15 which yielded two main and two negligible small clusters

Graphics

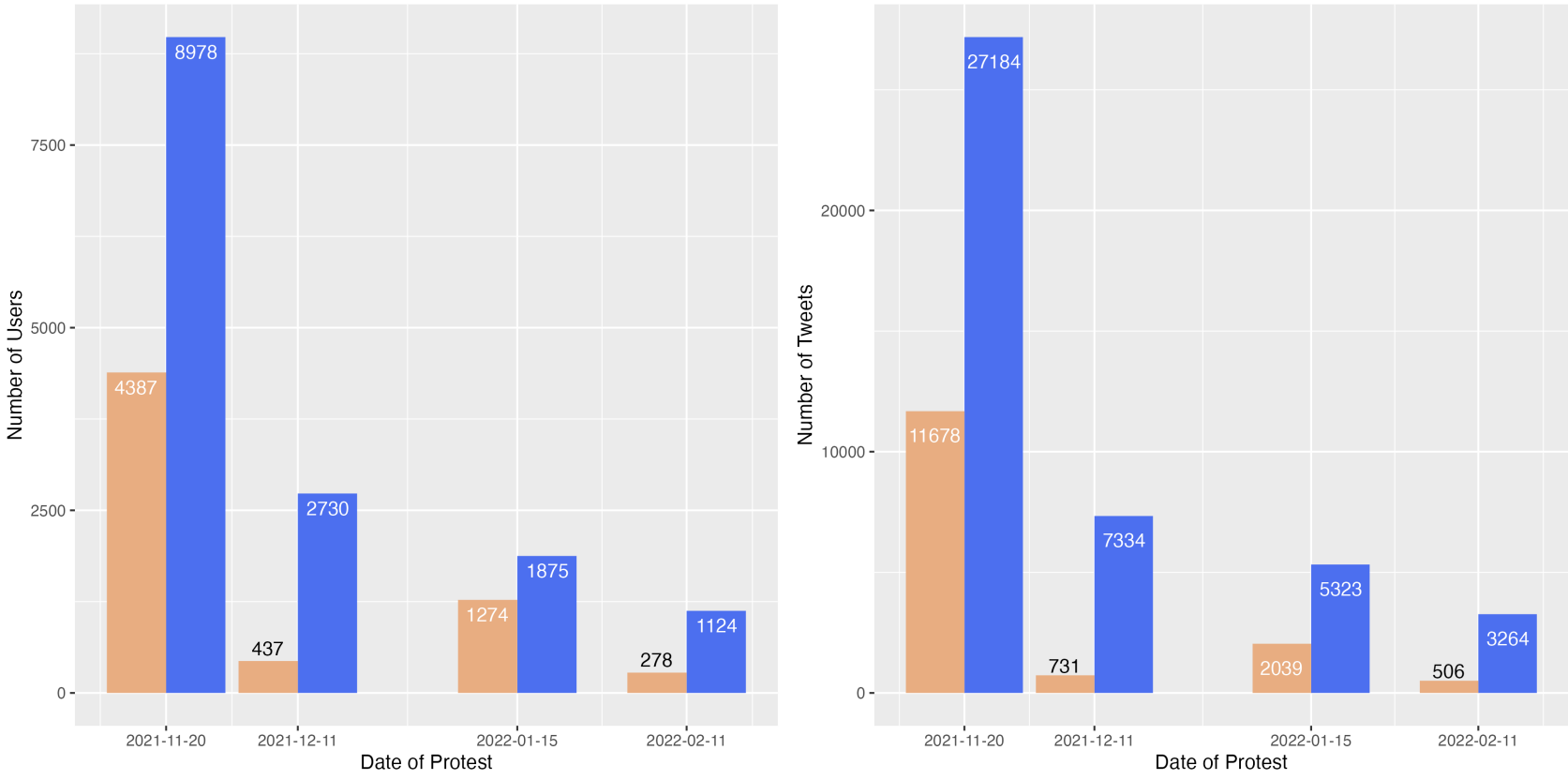


Figure 1: Number of users and tweets over time

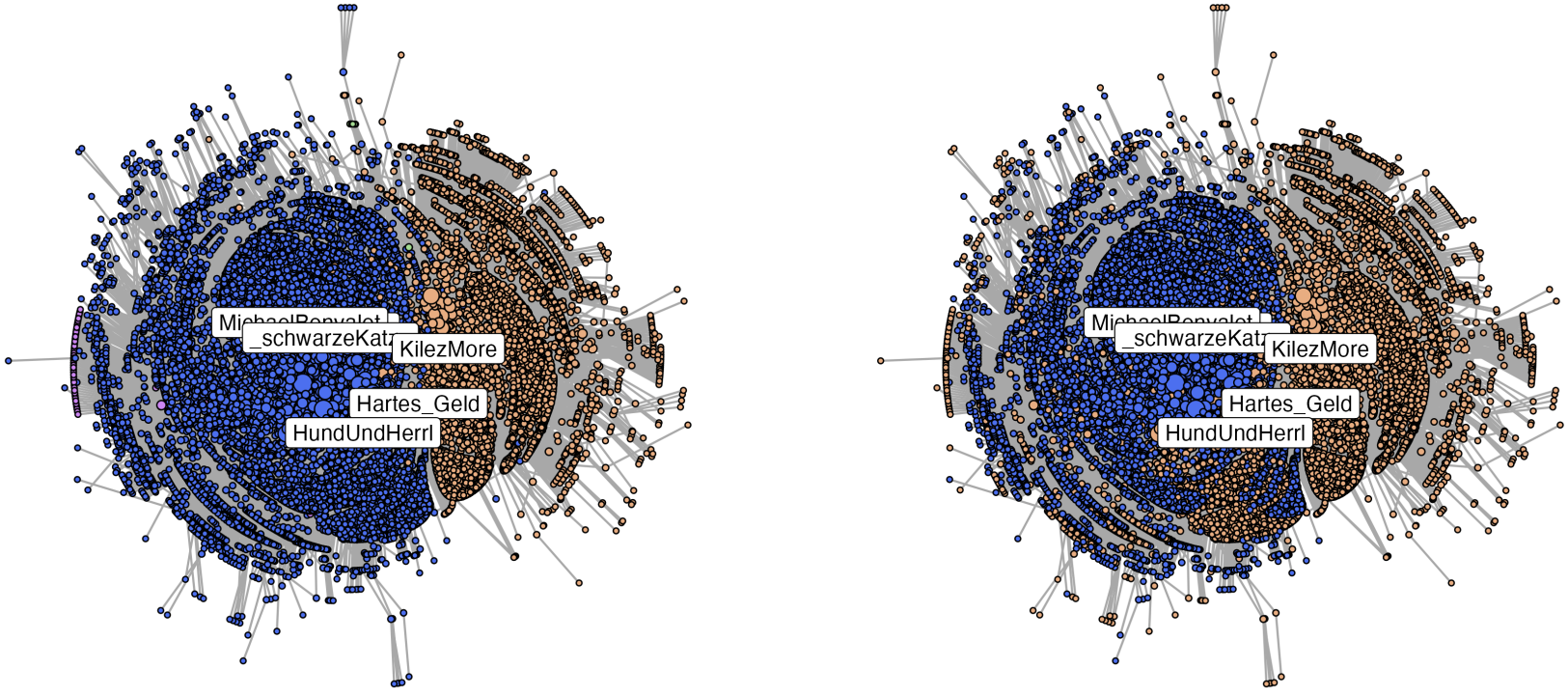
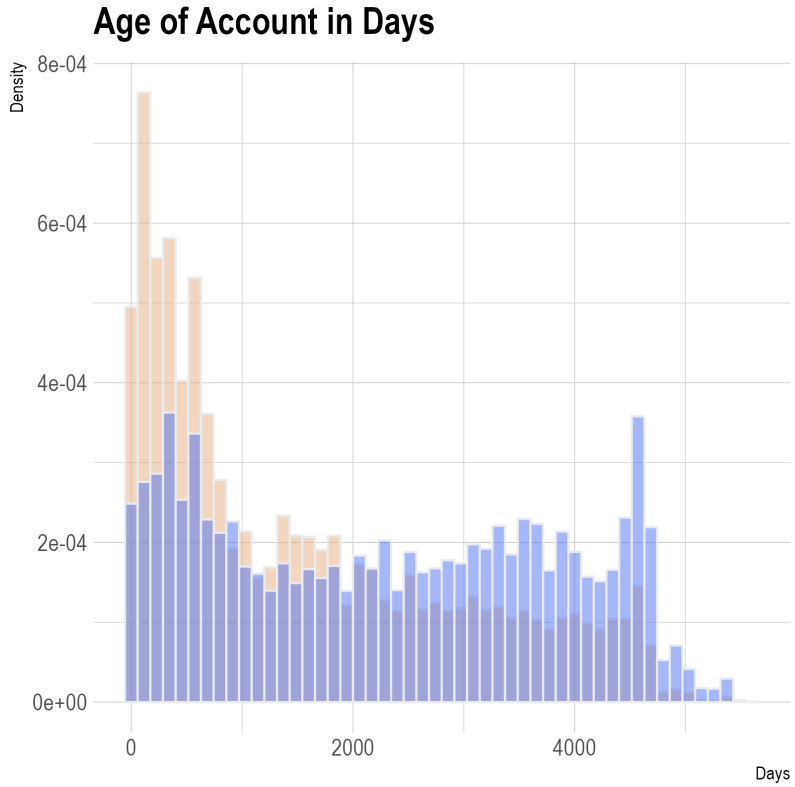


Figure 2: Full data set clustered using *Louvain* (left) and *Spin Glass* (right)



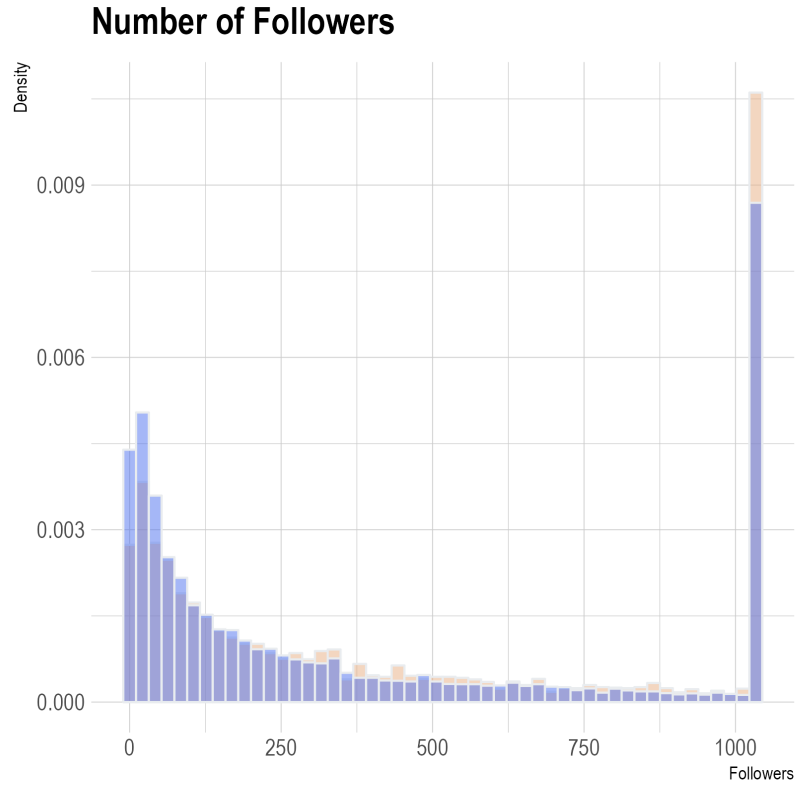


Figure 3: Histograms of non-lexical features for predicting group memberships

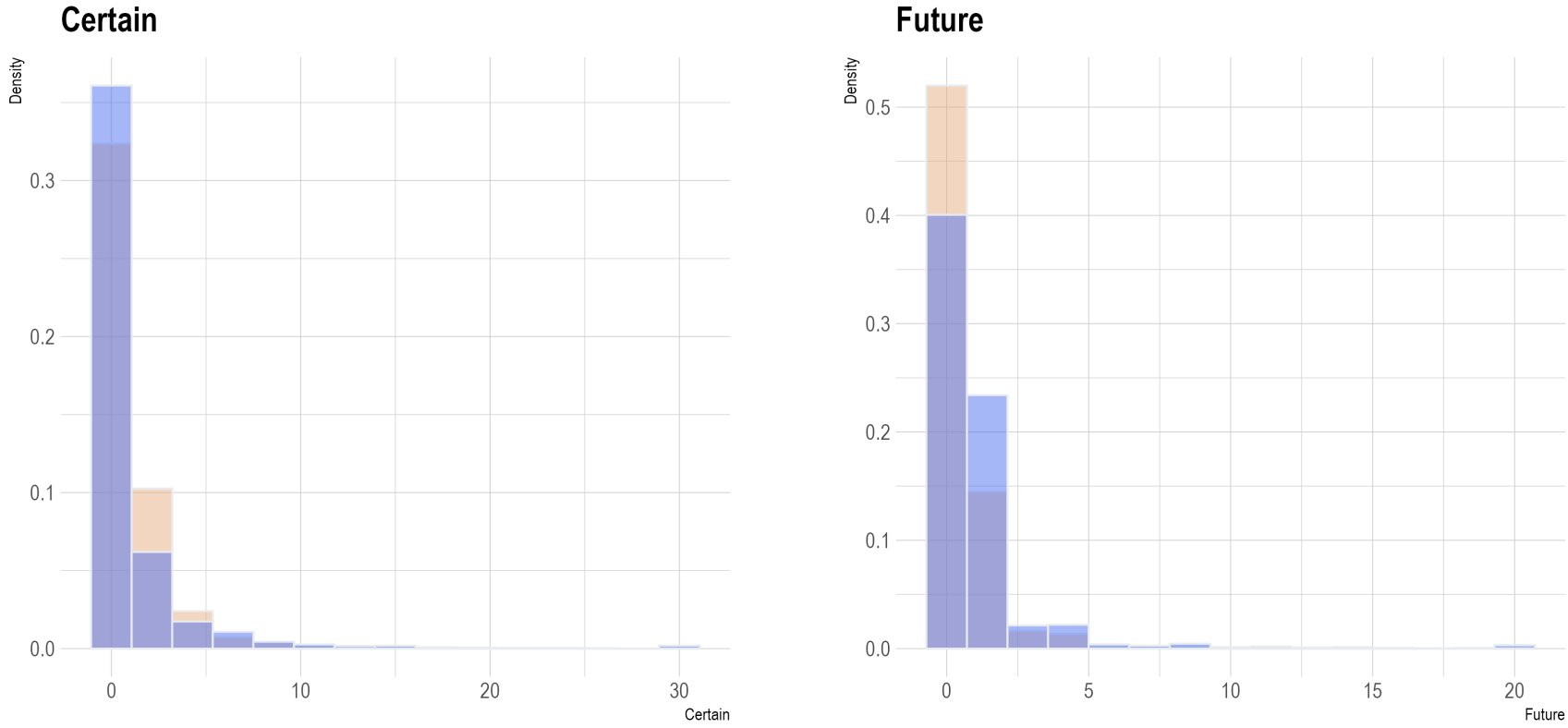


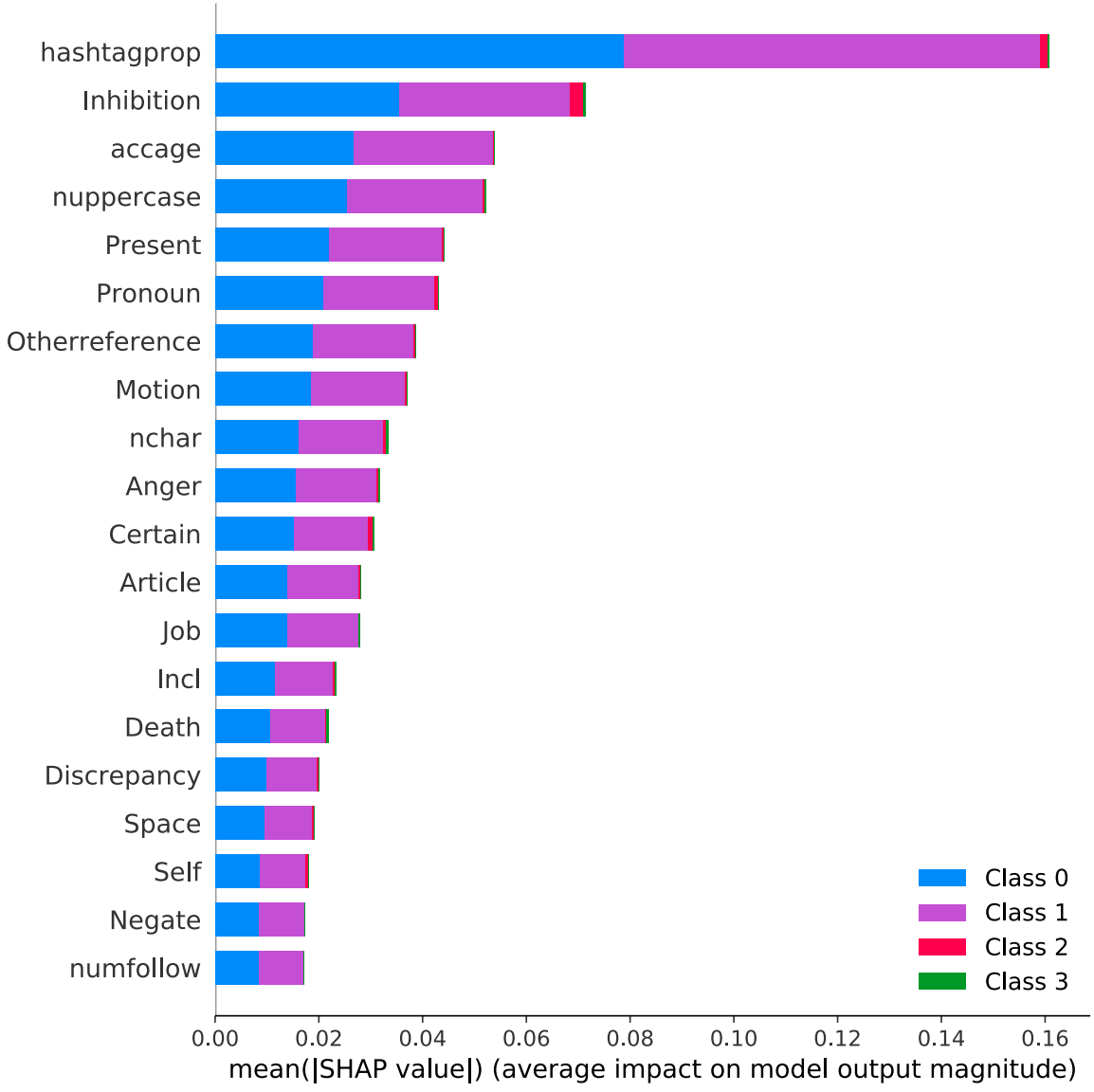
Figure 4: Histograms of lexical features for predicting group memberships

Feature extraction

- Baseline features for prediction based on Cossard et al. (2020):
  - *User (4):* account age in days, total number of tweets, number of followers, number of friends
  - *Twitter specific (3)* proportion of tweets that are retweets, have hashtags, or have a URL
  - *Lexical (2)* number of character, upper case characters
- Additionally: LIWC (linguistic inquiry and word count) German version
  - word list with 63 categories for automated word count analysis

Results

- In a baseline logistic regression, the prediction accuracy of individuals falling into clusters is around 70 %
- Main prediction was done with random forest classifier. Prediction results differ depending on clustering and noise reduction method.
  - If all data included, accuracy around 90 % (clustering: spinglass)
  - If all edges with a weight smaller than two excluded, accuracy around 93 % (clustering: spinglass)
  - If all posts that were retweeted once removed, accuracy around 95 % (clustering: louvain)
- Variable importance using SHAP yielded consistent results over all methods


hashtagprop
Inhibition
accage
nuppercase
Present
Pronoun
Otherreference
Motion
nchar
Anger
Certain
Article
Job
Incl
Death
Discrepancy
Space
Self
Negate
numfollow
Class 0
Class 1
Class 2
Class 3
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16
mean(|SHAP value|) (average impact on model output magnitude)

Conclusion and outlook

- Prediction of twitter-community regarding covid-protests on twitter using basic twitter features and lexical features of tweets yields high accuracy
- amount of hashtags used seems to be the most important predictor, lexical features seem to be of comparable importance overall
- Outlook: consider prediction for specific dates of protests
- Outlook: use SHAP local explanations to analyse most prominent actors in the network

References


Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre
"Fast unfolding of communities in large networks", Journal of Statistical Mechanics: Theory and Experiment. IOP Publishing (2008)
available on http://learningsys.org


Cossard, Alessandro, Gianmarco De Francisci Morales, Kyriaki Kalimeri, Yelena Mejova, Daniela Paolotti, and Michele Starnini
"Falling into the Echo Chamber: the Italian Vaccination Debate on Twitter", International AAAI Conference on Web and Social Media (2020)
available on arXiv:2003.11906v1

References


Wang, Y. Samuel and Elena A. Erosheva
"Fitting Mixed Membership Models using `mixedMem`" (2020)
available on cran.r-project.org