Exploring Causal Relationships Among Emotional and Topical Trajectories in Political Text Data

Andreas Baumann
University of Vienna, Austria

Klaus Hofmann
University of Vienna, Austria

Bettina Kern
University of Vienna, Austria

Anna Marakasova
TU Wien, Austria

Julia Neidhardt
TU Wien, Austria

Tanja Wissik
Austrian Academy of Sciences, Vienna, Austria

Abstract

We explore relationships between dynamics of emotion (arousal and valence) and topical stability in political discourse in two diachronic corpora of Austrian German. In doing so, we assess interactions among emotional and topical dynamics related to political parties as well as interactions between two different domains of discourse: debates in the parliament and journalistic media. Methodologically, we employ unsupervised techniques, time-series clustering and Granger-causal modeling to detect potential interactions. We find that emotional and topical dynamics in the media are only rarely a reflex of dynamics in parliamentary discourse.

2012 ACM Subject Classification Computing methodologies → Lexical semantics; Computing methodologies → Discourse, dialogue and pragmatics; Information systems → Sentiment analysis

Keywords and phrases time-series clustering, Granger causality, topical stability, emotion, political discourse

Digital Object Identifier 10.4230/OASIcs.LDK.2021.38

Funding This research was funded by the Austrian Academy of Sciences (go!digital Next Generation grant, GDNG 2018-020) and by the City of Vienna (Digitaler Humanismus grant, MA7-737909/19).

1 Introduction

Political discourse is evidently associated with emotions [10]. As new topics emerge they are, for example, framed positively or negatively by political stakeholders in their communication, and these dynamics are received and perhaps even amplified by the media [18, 22]. In this paper, we explore to what extent shifts in the topics that political parties are associated with drive or are in fact driven by emotional dynamics. We do so in an explicitly exploratory way; after all, it is hard to evaluate causal relationships between topical and emotional dynamics. More concretely, we analyze time series that characterize dynamics of (i) emotional valence (ii) arousal and, (iii) topical stability, for three political parties in the Austrian parliament.

To tackle interactions among discourse in the parliament and in the media we investigate two corpora as part of our ongoing project DYLEN [1]: the ParlAT corpus of parliamentary speeches in Austria and the Austrian Media Corpus, covering both print and online media. Since we are interested in the dynamic aspects of the interaction between topical stability

3rd Conference on Language, Data and Knowledge (LDK 2021).
Editors: Dagmar Gromann, Gilles Sérasset, Thierry Declerck, John P. McCrae, Jorge Gracia, Julia Bosque-Gil, Fernando Bobillo, and Barbara Heinisch; Article No. 38; pp. 38:1–38:8
OpenAccess Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

and emotion, we adopt a diachronic approach, covering a period of 20 years. In our analysis, we first identify which of the variables display similar diachronic dynamics and subsequently map interactions among the variables in networks based on Granger causality [4, 28]. This allows us to assess, for example, whether parliamentary debates drive discourse in the media (or vice versa), and whether the emotions encoded in the language associated with one party are significantly driven by the dynamics in the language of another party. For example, it is not *a priori* clear whether emotions in the discourse by right-wing parties are a consequence of topical shifts by left-wing parties, or conversely, whether the former in fact drives topical changes in political discourse. More generally, it is interesting to investigate whether long-term dynamics in the media are at all related to dynamics in parliamentary discourse, or whether the two domains may better be described as dissociated spheres of political discourse. We argue that this data-driven exploratory approach has the potential of generating interesting hypotheses that can (and should) subsequently be evaluated in more detailed (qualitative) investigations.

The application of Granger causality to investigate the impact of sentiment (or more generally: emotion) in texts on variables of interest is not new. In particular, Granger causality was used to predict trends in economics and finance based on sentiment encoded in tweets [17, 11] or in newspaper articles [9]. In research on health-care, Granger-causal modeling revealed interactions among sentiment and the tendency to participate in medically related discussions on Reddit [2], as well as effects of anxiety dynamics in tweets on changes in social-interaction behavior [3]. On the structural level of language, Granger casuality was employed to analyze the relationship between syntactic change and frequency [16]. In our contribution, we focus on the interaction between emotion and topical shifts in the political context.

We first describe our data and how diachronic trajectories for topical stability and emotion estimates were derived. We then present the pipeline of our exploratory analysis resulting in Granger-causal networks. Finally, we briefly discuss our results as well as possible future directions of our research.

2 Data and time-series pre-processing

The data analyzed in our study comes from two different sources: first, the Austrian Corpus of Parliamentary Records (ParlAT; [25]), consisting of transcribed speeches in the Austrian parliament; second, the Austrian Media Corpus (AMC; [15]) consisting of Austrian print and online media. For the present analysis, both corpora were limited to the period from 1997 to 2016, thus covering two decades of political discourse in Austria. The two corpora differ considerably in size and structure. While ParlAT consists of 75 million tokens, AMC is much bigger covering 5.5 billion tokens. Even though in their current form the corpora do not allow us to track causal dynamics within the time frame of individual news cycles (e.g. interviews and opinion pieces by influential figures and "spin doctors" are not tagged separately in AMC), the two corpora combined provide the best available coverage of Austrian political discourse to explore questions on a broader temporal scale.

Both corpora were used to derive time series for three different variables and three different groups of individuals, namely the political parties FP ("Freiheitliche Partei Österreichs", Austrian freedom party; right-wing), VP ("Österreichische Volkspartei", Austrian people's party; conservatives), and SP ("Sozialdemokratische Partei Österreichs", Austrian social democrats). The three variables considered are (i) topical stability, (ii) valence, and (iii) arousal. Valence and arousal refer to two emotional dimensions; while valence measures whether a text is

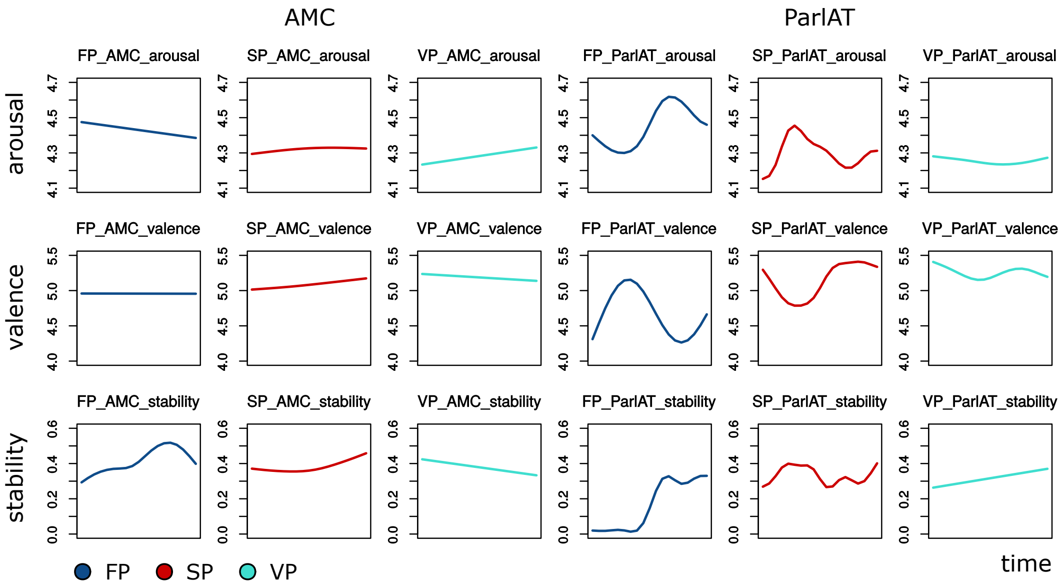

Figure 1 Trajectories for all corpora, metrics and targets (SP: red; FP: blue; VP: turquoise).

negative or positive, arousal measures the extent to which the text represents calm or agitated language [24, 21]. For valence and arousal, we used the time series computed in [7]. In a nutshell, these time series were determined by splitting both corpora into sub-corpora for each year, then, in each sub-corpus, extracting 200 words that are distinctive for each party (FP, VP, SP). Distinctive words were determined based on chi-squared statistics (see [5] for details). After that, a sentiment dictionary [8] was used to determine average valence and arousal, respectively, based on the sets of distinctive lexical items, in an unsupervised fashion. This was done separately for every single year.

To illustrate this approach, let us consider the three most positive and most negative words, respectively, associated with FP in ParlAT in two different years. In 2004, the most negative words are *Kriminalität* (“crime”, valence: 1.77), *Vorwurf* (“allegation”, 1.78), and *Opfer* (“victim”, 2.48), while the most positive words are *Freude* (“joy”, 8.29), *Erfolg* (“success”, 8.23), and *Gute* (“the good (one)”, 7.67). In contrast, the most negative words associated with FP in 2010 are *Versagen* (“failure”, 1.17), *Verfassungsbruch* (“constitutional violation”, 1.52), and *Arbeitsverweigerung* (“refusal to work”, 1.76); the most positive ones are *Familie* (“family”, 7.55), *Mut* (“courage”, 7.38), and *Wein* (“wine”, 7.04). Crucially, the most extreme words in 2004 show a higher valence than those in 2010. Differences at the 10^{-1} level already indicate a noticeable shift in this regard. In the time series of year-wise average valence/arousal scores over all 200 distinctive words, such shifts are encoded for the whole observation period.

In total, 12 time series were generated as described above (two corpora, three parties, two emotion scores). To eliminate noise, generalized additive models (GAM; [26]) were fitted to each time series. GAMs are suitable models for analyzing the time series at hand since they also capture non-linear dynamics (and, moreover, allow for factoring in autocorrelation). The predicted values of the fitted GAMs were then used for the present analysis. More details on the modeling procedure can be found in [7].

We measured topical stability of the parties over time by applying Jaccard index [6] to compare semantic neighbourhoods of the party names in two subsequent years. First, we computed word co-occurrence matrices with positive pointwise mutual information (PPMI) scores of each yearly subcorpus of the two corpora (AMC and ParlAT). The subcorpora are lemmatised, and only nouns, verbs and adjectives were considered. PPMI vector representation was preferred over state-of-the-art dense embeddings (i.e. word2vec or GloVe vectors) due to the fact that the latter resulted in largely linear dynamics (see below). Thus, for each party name in each year we extract semantic neighborhood which is represented by top-n semantically most similar words. The size of the neighbourhood is set to 50 words.[1] Next, each neighbourhood set is compared to the one from the previous year using Jaccard index which, in total, results in six additional time series.

Again, GAMs were fitted to each of the six time series in order to take care of noisy data, and the GAM estimates were added to our data set for further analysis. Thus, our final dataset consists of time series for 18 variables, each made up of scores for 20 years. The 18 resulting time series are shown in Figure 1. The dataset can be downloaded from `https://phaidra.univie.ac.at/o:1168825`. It can be seen, for example, that valence in FP discourse in the parliament seems to peak around 2004 and subsequently drops to obtain its minimum in 2010 (in [7] we argue that this might be a consequence of the transition of FP from government to opposition; an extra-linguistic factor which is certainly relevant, as one of the reviewers has pointed out as well).

3 Analysis

Our analysis unfolds in three steps: first, clustering of the time series described in the previous section; second, identification of clusters; third, computation of Granger-causal networks based on the clusters. Each of these steps will be explained in more detail in the following.

In order to cluster the time-series of emotional scores and topical stability, we first need a distance measure to assess the degree to which two time series are similar to each other. We opted for autocorrelation function (ACF) distance, which is derived by first computing the ACF for each time series and then, for each pair, computing the Euclidean distance between the two time series [12]. Thus, ACF distance treats those time series as similar which have a similar autocorrelation structure, i.e. which are characterized by similar degrees of changeability through time. This measure has multiple advantages for analyzing our data. First, it implicitly normalizes all variables, which is important since arousal, valence and topical stability operate on different scales. Second, it is invariant with respect to the orientation of the observed variable. That is, if a time series has, say, a W-shaped curve, then this time series and its vertically flipped M-shaped variant have the same ACF and are hence treated as similar. This is important for our analysis, since we also want to detect if downward trends in one variable are linked to upward trends in another variable. Since linear time series have identical and linearly decreasing ACFs, only non-linear time series were included in the analysis. This will be important for the causal analysis explained below, since the causal methods employed in this paper do not reasonably apply to pairs of linear time series.

We used ACF distance to derive a distance matrix for the remaining 12 variables (i.e. time series) in our dataset. We then applied hierarchical agglomerative clustering to this distance matrix to identify groups of similarly behaving time series. We used Ward linkage as a clustering criterion [27] and determined the optimal number of clusters through maximizing

[1] Experiments with the larger neighbourhood sizes did not show significant difference with respect to the current findings.

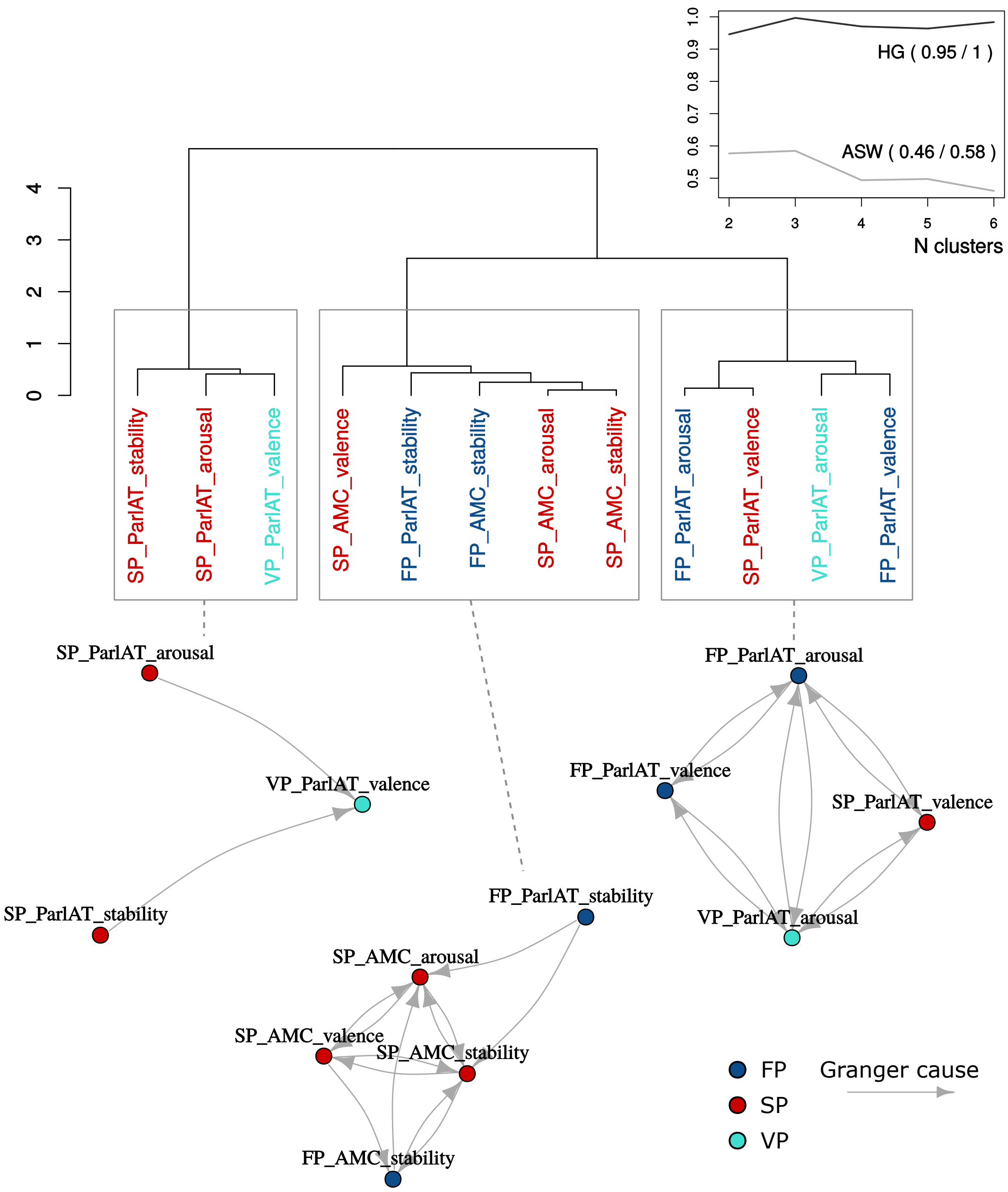

Figure 2 Top left: Hierarchical clustering of all time series (driven by ACF distance; complete linkage). Top right: Clustering quality measures average silhouette width (ASW) and Hubert's Gamma (HG). Bottom: Granger-causal networks for the clusters in the dendrogram. Each node represents a time series. Two nodes are linked if one node significantly Granger-causes another node. Arrows denote causal relationships pointing from cause to effect. Color code: SP: red; FP: blue; VP: turquoise.

average silhouette width (ASW), measuring homogeneity of clusters, and Hubert's Gamma (HG), which measures the extent to which the dendrogram reflects the original distance matrix [19]. A robustness analysis involving other linkage methods (single, average, complete, median, centroid) revealed that the final clusters are in fact invariant with respect to linkage selection. The optimal number of clusters in the dendrogram was computed as three. The resulting dendrogram and the corresponding measures for clustering quality are shown in Figure 2 (top).

After that, causal networks were computed for each cluster separately by means of Granger causality tests [4, 28]. Granger causality is a concept for modelling the causal relationship between two time series x_t and y_t. The underlying idea is to check whether predicting y_t is significantly improved by also considering past information of the former time series, i.e. x_{t-k} for some lag k (based on a Wald test). If this is the case then x_t is said to Granger cause y_t (but not necessarily vice versa). In the present analysis, we opted to only consider past information that goes back up to one time step (lag $k = 1$), i.e. one year, since we do not consider interactions among instances of discourse over more than one year plausible. More fine-grained time scales might plausibly allow for more than one time step. In each cluster in the dendrogram, a Granger test was computed for each pair of time series in that cluster, thereby considering both potential directions of causality. A significance threshold of $\alpha = 0.05$ was employed, which was Bonferroni-corrected through the overall number of Granger tests in that cluster (note, however, that employing a fixed threshold of $\alpha = 0.01$ yields exactly the same qualitative results; cf. e.g. [14]). Subsequently, a directed Granger causal graph was created for each cluster, in which two variables are linked (from cause to effect) if the p-value of their corresponding Granger test is below the respective significance threshold.

Figure 2 (bottom) shows Granger-causal networks for the three clusters. The left-most cluster displays causal relationships from SP stability and SP arousal to VP valence (all in ParlAT). The cluster in the middle shows all SP variables and FP stability in AMC being mutually connected and Granger-causally affected by FP stability in ParlAT. The final cluster shows mutually connected emotion variables (valence, arousal) for all parties in ParlAT.

4 Discussion and outlook

In this paper, we have shown how interactions among dynamics of emotion encoded in political discourse and topical changes, both across political parties and domains (parliamentary speeches; media) can be analyzed by means of time-series analysis and Granger-causal modeling, thus extending the application of Granger causality to the analysis of political text data.

Two observations can be made: First, it can be seen that topical stability and emotion are interconnected. However, we do not see a clear tendency that emotional shifts are driven by topical changes or vice versa. Second, the two domains, parliamentary discourse and media, seem to be rather disconnected. The only exception to this rule is topical stability of FP in the parliament which seems to affect dynamics in the media (both topical and emotional). This is interesting and tentatively suggests that changes in contributions to parliamentary discourse by right-wing politicians functions as an important driver of dynamics in the media. Conversely, it might be the case that right-wing discourse is more likely to be picked up and reflected on by Austrian media outlets (also if reporting on left-wing discourse) than the discourse produced by other parts of the political spectrum [23].

However, this observation has to be treated with caution. Evidently, our approach has many shortcomings. First, our diachronic data is rather coarsely grained. For analyzing time series in linguistic dynamics both a longer time span and shorter subperiods (e.g. months instead of years) are desirable to obtain more robust results. Second, the estimation of the variables investigated (valence, arousal, stability) was based on rather simple and straightforward methods, which were motivated by the large structural difference between the two underlying corpora (AMC vs. ParlAT). There is undoubtedly room for more sophisticated and reliable methods for emotion and topical change detection. Third, it is evident that

Granger causality is only one model of what is usually conceptualized as causality. It will be interesting to compare whether other methods for detecting causal relationships among time series, like Bayesian dynamic networks [13] or convergent cross mapping [20], produce similar outcomes. Still, we find that our exploratory approach generates stimulating hypotheses that deserve further investigation in future studies.

References

1 Andreas Baumann, Julia Neidhardt, and Tanja Wissik. DYLEN: Diachronic Dynamics of Lexical Networks. In *LDK (Posters)*, pages 24–28, 2019.

2 Giovanni Delnevo, Marco Roccetti, and Silvia Mirri. Modeling patients' online medical conversations: a granger causality approach. In *Proceedings of the 2018 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies*, pages 40–44, 2018.

3 Sarmistha Dutta, Jennifer Ma, and Munmun De Choudhury. Measuring the impact of anxiety on online social interactions. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 12, 2018.

4 C. W. J. Granger. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 1969. doi:10.2307/1912791.

5 Klaus Hofmann, Anna Marakasova, Andreas Baumann, Julia Neidhardt, and Tanja Wissik. Comparing lexical usage in political discourse across diachronic corpora. In *Proceedings of the Second ParlaCLARIN Workshop*, pages 58–65, 2020.

6 Paul Jaccard. The distribution of the flora in the alpine zone. 1. *New phytologist*, 11(2):37–50, 1912.

7 Bettina M. J. Kern, Klaus Hofmann, Andreas Baumann, and Tanja Wissik. Komparative Zeitreihenanalyse der lexikalischen Stabilität und Emotion in österreichischen Korpusdaten. In *Proceedings of Digital Lexis and beyond at OELT*, 2021.

8 Maximilian Köper and Sabine Schulte Im Walde. Automatically generated affective norms of abstractness, arousal, imageability and valence for 350 000 german lemmas. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2595–2598, 2016.

9 Jian Li, Zhenjing Xu, Lean Yu, and Ling Tang. Forecasting oil price trends with sentiment of online news articles. *Procedia Computer Science*, 91:1081–1087, 2016.

10 GE Marcus and N Demertzis. *Emotions in politics: The affect dimension in political tension.* Plagrave Macmillan Press, 2013.

11 Anshul Mittal and Arpit Goel. Stock prediction using twitter sentiment analysis. *Standford University, CS229*, 15, 2012. URL: http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf.

12 Pablo Montero, José A Vilar, et al. TSclust: An R package for time series clustering. *Journal of Statistical Software*, 62(1):1–43, 2014.

13 Judea Pearl. Graphical models for probabilistic and causal reasoning. In *Computer Science Handbook, Second Edition*. Springer, 2004. doi:10.1201/b16812-50.

14 Thomas V. Perneger. What's wrong with Bonferroni adjustments, 1998. doi:10.1136/bmj.316.7139.1236.

15 Jutta Ransmayr, Karlheinz Mörth, and Matej Ďurčo. Ii. amc (austrian media corpus)–korpusbasierte forschungen zum österreichischen deutsch. In *Digitale Methoden der Korpusforschung in Österreich*. Verlag der Österreichischen Akademie der Wissenschaften, 2017.

16 Malte Rosemeyer and Freek Van de Velde. On cause and correlation in language change: Word order and clefting in brazilian portuguese. *Language Dynamics and Change*, 11(1):130–166, 2021.

17 Jasmina Smailović, Miha Grčar, Nada Lavrač, and Martin Žnidaršič. Predictive sentiment analysis of tweets: A stock market application. In *International Workshop on Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data*, pages 77–88. Springer, 2013.

18 Stefan Stieglitz and Linh Dang-Xuan. Emotions and information diffusion in social media - Sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*, 2013. `doi:10.2753/MIS0742-1222290408`.

19 Matthias Studer. Weightedcluster library manual: A practical guide to creating typologies of trajectories in the social sciences with r. *LIVES Working papers*, 2013. `doi:10.12682/lives.2296-1658.2013.24`.

20 George Sugihara, Robert May, Hao Ye, Chih Hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. Detecting causality in complex ecosystems. *Science*, 2012. `doi:10.1126/science.1227079`.

21 Maite Taboada. Sentiment Analysis: An Overview from Linguistics, 2016. `doi:10.1146/annurev-linguistics-011415-040518`.

22 Joshua Tucker, Andrew Guess, Pablo Barbera, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan. Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature. *SSRN Electronic Journal*, 2018. `doi:10.2139/ssrn.3144139`.

23 Ineke Van Der Valk. Right-wing parliamentary discourse on immigration in france. *Discourse & Society*, 14(3):309–348, 2003.

24 Amy Beth Warriner, Victor Kuperman, and Marc Brysbaert. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 2013. `doi:10.3758/s13428-012-0314-x`.

25 Tanja Wissik and Hannes Pirker. ParlAT beta Corpus of Austrian Parliamentary Records. In *Proceedings of the LREC 2018 Workshop'ParlaCLARIN: LREC2018 workshop on creating and using parliamentary corpora*, pages 20–23, 2018.

26 Simon N Wood. *Generalized additive models: an introduction with R*. CRC press, 2017.

27 Mohammed J Zaki and Wagner Meira. *Data mining and analysis: Fundamental concepts and algorithms*. Cambridge University Press, New York, 2014.

28 Cunlu Zou and Jianfeng Feng. Granger causality vs. dynamic bayesian network inference: a comparative study. *BMC bioinformatics*, 10(1):1–17, 2009.