July 15th 2019

Fabio Votta
Gutenbergstraße 130
70197 Stuttgart
+49 1742021316
fabio.votta@gmail.com
www.favstats.eu

Research Outline

PhD position in Political Science on the topic of Affective Polarization
University of Amsterdam

Echo chambers, affective polarization, and democracy

I think it's very important to combine both the study of off- and online environments. While echo chambers are often discussed in the context of online environments and more specifically social media platforms, offline echo chambers should be considered at least as much. However, focusing on both does not mean that the same methodology can necessarily be applied. There are different factors at play in online and offline echo chamber environments. One main distinctive feature of the online environment is the removal of personal identifiers that lead into a state of *deinviduation*. Festinger et al. define deindividuation as the state by which self-awareness and self-accountability are diminished when "individuals are not seen or paid attention to as individuals" (Festinger et al. 1952). Lee applies the concept of deindividuation effects to a Computer-Mediated Communication (CMC) setting and finds that polarisation flourishes in a CMC environment where individuating cues are absent (Lee 2006). I think this process is highly relevant when it comes to explaining affective polarization in social media interactions. Therefore it can be hypothesized that to the extent that social media inherits deindividuation effects polarization might be substantially amplified through this medium. A study that could be done in this regard might be observing behaviour of individuals with human vs. non-human profile pictures.

Existing Cross-Sectional Repeated Measures Survey Data

When I read the description for the PhD and the broader research project on *Affective polarization in Europe*, I was surprised that the European Social Survey (ESS), the European Value Study (EVS) and Eurobarometer did not come up at all which field some questions that might be relevant for the AFFPOL project. While to my knowledge all of the mentioned survey projects above do not include measures of affective polarization, they do ask about left-right self-placement. With fairly consistent questions asked biannually from 2002 to 2018 the ESS especially could be a good resource to study some trends over time. For example, one could identify and study individuals who live among similar and dissimilar minded ideological peers by subtracting their left-right placement from average left-right placement per subregion (these variables are asked in all ESS rounds). Once this is done, one can track the proportion of ideological enclaves and misfits over time to study how they perceive democracy or political trust.

Reducing Affective Polarization

While the current PhD project focuses on the causes and consequences of affective polarization in on- and offline environment, no time is spent on discussing possible remedies to it. Of course, it is important to establish the degree to which affective polarization is a problem in Europe, as this is very much an understudied phenomena. Nonetheless, I think it's important to think about possible remedies to affective polarization and see if the items that work in the US also work for Europe. For example, people who score higher on intellectual humility questionnaires are more likely to show interest in communicating with people who hold different views from oneself (Porter et al. 2018) and intellectual humility is negatively associated with fundamentalism and authoritarianism (Krumrei-Mancuso 2018). Validated measures of intellectual humility could be fielded in the European context as well, see Krumrei-Mancuso (2016).

Research Proposal

When the activist and journalist Eli Pariser gave a TED talk on algorithms, he coined a term that still dominates the debate today: the “filter bubble”. In his talk, Pariser feared for a future in which everyone lives in their own digital bubble and consumes only content ranked and served by opaque algorithms (cf. Pariser 2011). The concept of filter bubbles is often conflated with a related phenomenon that is illustrated by the metaphor of an “echo chamber” (cf. Sunstein 2001). The difference between the two concepts is that filter bubbles are a phenomenon of internet environments said to be caused by algorithmic prioritization of (ideologically) appealing content whereas echo chambers *also* emerge in face-to-face relationships when people are only exposed to opinions that they already hold. Although this phenomenon of selective exposure has been known for many years in the academic literature, the recent rise of social media platforms opened an opportunity to extend studies to the digital sphere. On social media, users are given the ability to browse, follow or otherwise make decisions on the content that they want to consume in ‘high choice media environments’ (Lelkes et al. 2017). Given the psychological tendency to reduce exposure to incongruent viewpoints (Tsang 2017; Frimer et al., 2017), users might self-select into online echo chambers while being aided by algorithms that prioritize what the user wants to see. The issue of online echo echambers described above raises the question of how homogenous digital environments might affect people's evaluation of ideological opponents that they have little or no contact with. In a study that measures the effects of exposure to congruent partisan content in the US and Israel, evidence suggests that exposure to supportive information increases affective polarization (Garett et al. 2014). I formulate the following hypothesis to test this relationship in a popular social media online environment (i.e. Twitter):

Individuals in online echo chambers (who interact mostly with their own ideological side) are more affectively polarized than individuals who are exposed to different political views.

Methodology

I propose the following course of action to test this hypothesis: First, collect data on Twitter from individuals who follow political accounts. Next, I estimate political ideology of Twitter users in order to identify individuals who mostly interact with similar others (i.e. individuals in echo chambers) and those who interact with a variety of ideological dispositions. Finally, collect tweets from identified users and use machine learning models to estimate 'toxic' language targeted at ideological opponents in order to assess affective polarization of users in the echo chamber vs. mixed condition. In the following, I will go into greater detail of how I plan to conduct this study:

As a first step, I will identify (official) Twitter accounts of politicians, parties, think tanks as well as political commentators and newspapers from various European countries. As a starting point, I will consider the R package LEGISLATOR, which tracks publically available data for 32,533 current and former elected politicians from nine countries' legislatures (Göbel et al. 2019). This database also includes social media handles and will serve as a main source for identifying politician accounts on Twitter (although one can reveal more accounts with additional research, especially on the local and regional levels). Next, I will collect a list of Twitter IDs for all followers of the previously identified political actors via the Twitter REST API. From each of these users I will retrieve their last 3200 tweets (as per Twitter API limit) and number of likes, retweets and replies of the collected tweets. However, as Pablo Barberá who conducted a similar study points out, we can expect that a large proportion of collected Twitter users who follow our target population will be bots, inactive or residents of a different country (Barberá 2015). To overcome this problem, I follow Barberá in only including users that: 1) have sent more than 100 tweets, 2) have tweeted at least once in the past six months, 3) have more than 25 followers, 4) are tweeting from within the country of interest, and 5) follow at least three political Twitter accounts within their country. It is important to note that this sample will therefore likely be biased towards more politically active and interested users. However, it can be assumed that such Twitter users will be more informative to extract political positioning from (Barberá 2015). In order to estimate the state of being in an echo chamber from the collected Twitter data, I first need to create a measure of ideology. To do this, I will expand upon the Bayesian Spatial Following model developed by Pablo Barberá which has succesfully modelled left-right political ideology in five European countries and the US by taking into acccount which political actors a given user is following (Barberá 2015). After validating the ideological measure with expert surveys (e.g. the Chapel Hill Expert Survey), I can start categorizing whether they are in echo chambers or not. For this to work I will look at 3200 latest tweets collected from the users and look at their interaction with content produced by users with similar and dissimilar political ideology. The right classification scheme will depend somewhat on the data but I propose the three following groups: *1) Echo chamber condition* in which users mostly interact with individuals who are ideologically similar to them, *2) mixed networks* in which users interact with individuals across the spectrum and *3) neutral networks* where users mostly interact with non-political content. Once the sample of Twitter users is classified, it is worth exploring some descriptive measures: how many individuals are part of echo chambers, where they only interact with their own side? How does that differ by country, party or political ideology? As a next step, we want to consider affective polarization of these Twitter users in echo chambers vs. mixed environments. To do that, I will use a machine learning classifier to identify toxic content produced or shared by the users which is targeted at the opposing ideological side. One avenue to explore is the Perspective API from Google which provides a machine learning model to score toxicity of a given text (via the R package PERSPECTIVE (Votta 2019)). Here, affective polarization can be measured as the toxicity of user content when it mentions an ideological opponent. For this, I would generate a list of terms and names associated with political actors for any given country context. However, the Perspective API so far only allows to score comments in English, French, Spanish and German which would limit the countries that could be studied this way. Finally, once I have a measure for affective polarization we can test our hypothesis to see whether people in echo chamber conditions are more affectively polarized towards their outgroups than those who are in more mixed conditions.

Limitations and possible Extensions

To conclude, I want to discuss a few limitations and possible extensions of this study. The current methodology seeks to collect data from millions of Twitter users, sampling only from users who follow political accounts. At the moment, it does not account for users that do not follow political accounts and I will thus not have measures on their ideological leaning. Given that I plan to collect followers of all the most important political actors in a given country, I hope to collect enough of a given country's political sphere to meaningfully cover all political input they might get on Twitter. A different limitation is that my methodology only looks at content that users interacted with which means I will not be able to assess what people *actually* see on their timeline and thus whether their experience on Twitter really resembles an echo chamber or not. Finally, using social media data has its up- and downsides. The upside of using social media data is that we would study people as they behave in the online environment without potentially priming them with our survey measures. However, self-report measures have their value as well as they are more controlled, especially compared to the computational methods to estimate ideological leaning and affective polarization proposed in this study which would need to be validated first. In order to alleviate this issue, I propose to combine this study of social media data with a survey of Twitter users, in a research desgin similar to Vaccari et al. (2016). The initial step would still be the same: first, collect followers of all relevant political Twitter accounts in any given country. After that, I would follow Vaccari et al. and draw a random sample of users that will be sent messages on Twitter to participate in an online survey. The survey would include questions on exposure to differing views, both online and offline, perceptions of democratic performance, as well as affective polarization towards ideological oppponents. I could then compare results of this survey with behavioural measures proposed in this study to test their validity and robustness.

Literature

Barberá, P. (2015). Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. Political Analysis, 23(1), 76-91.

Festinger, L., Pepitone, A., and Newcomb, T. (1952). Some consequences of deindividuation in a group. Journal of Abnormal and Social Psychology, 47, 382-389.

Frimer, J. ,Skitka, L. J., & Motyl, M.(2017). Liberals and conservatives are similarly motivated to avoid exposure to one another's opinions. Journal of Experimental Social Psychology, 72, 1-12.

Garrett, R. K., Gvirsman, S. D., Johnson, B. K., Tsfati, Y., Neo, R., & Dal, A. (2014). Implications of pro-and counterattitudinal information exposure for affective polarization. Human Communication Research, 40(3), 309-332.

Göbel, Sascha and Simon Munzert. (2019). legislatoR: Political, sociodemographic, and Wikipedia-related data on political elites. Source: https://github.com/saschagobel.

Krumrei-Mancuso, E. J., & Rouse, S. V. (2016). The development and validation of the comprehensive intellectual humility scale. Journal of Personality Assessment, 98(2), 209-221.

Krumrei-Mancuso, E. J. (2018). Intellectual humility's links to religion and spirituality and the role of authoritarianism. Personality and Individual Differences, 130, 65-75.

Lee, E. J. (2006). Deindividuation effects on group polarization in computer-mediated communication: The role of group identification, public-self-awareness, and perceived argument quality. Journal of communication, 57(2), 385-403.
Lelkes, Y., Sood, G., & Iyengar, S. (2017). The hostile audience: The effect of access to broadband internet on partisan affect. American Journal of Political Science, 61(1), 5-20.

Pariser, Eli. (2011). The Filter Bubble: What the Internet Is Hiding from You. The Penguin Group.

Polk, Jonathan, Jan Rovny, Ryan Bakker, Erica Edwards, Liesbet Hooghe, Seth Jolly, Jelle Koedam, Filip Kostelka, Gary Marks, Gijs Schumacher, Marco Steenbergen, Milada Vachudova and Marko Zilovic. (2017). "Explaining the salience of anti-elitism and reducing political corruption for political parties in Europe with the 2014 Chapel Hill Expert Survey data," Research & Politics (January-March): 1-9.

Porter, T. and Schumann, K. (2018). Intellectual humility and openness to the opposing view. Self and Identity, 17(2), 139-162.

Sunstein, C.R. (2001). Echo Chambers: Bush V. Gore, Impeachment, and Beyond. Princeton University Press.

Tsang, Stephanie Jean. (2017). Cognitive Discrepancy, Dissonance, and Selective Exposure. Media Psychology. 1-24. 10.1080/15213269.2017.1282873.

Vaccari, C., Valeriani, A., Barberá, P., Jost, J. T., Nagler, J., & Tucker, J. A. (2016). Of echo chambers and contrarian clubs: Exposure to political disagreement among German and Italian users of Twitter. Social Media+ Society, 2(3), 2056305116664221.

Votta, Fabio. (2019). peRspective: A wrapper for the Perspective API.
Source: https://github.com/favstats.