Dealing with online falsehoods in the digital age: Experiences from Germany

Authors

This written representation is authored by Prof. Dr. Simon Hegelich and Morteza Shahrezaye. Simon Hegelich is professor for political data science at the Bavarian School of Public Policy at the Technical University of Munich. Morteza Shahrezaye is working as researcher at the professorship for political data science. Morteza Shahrezaye is willing to appear before the Committee to give evidence if required.

The philosophical foundation of the public discourse and why it matters

Today we are witnessing a disruptive change of the public sphere, caused by fundamental changes in the technical infrastructure of communication. The political discourse in democracies is *conceptionally* separated in a private sphere and a public sphere. In the private sphere, citizens follow their private interests and motivations. They are bourgeois. In contrast, the public sphere is about the commonweal, or the general welfare or public good. The engagement in this sphere creates the citoyen. This differentiation is „conceptional" in two distinct ways: On the one hand, it is just a concept. In every real life political discourse there will always be both aspects. Every citizen is bourgeois and citoyen at the same time. While following their own pursuit of happiness people try to contribute to the public interest as well. For example Facebook is emphasizing, its mission is „to connect people" and not just to make profits. And people serving the public interest – may it be as voters or on the highest levels of public service – naturally are following private motivations as well. At the same time „conceptional" means that the separation of private and public sphere is formative for democracy. To judge everything from the perspective of the higher common good would also mean the end of personal freedom. But unlimited personal freedom would destroy society. The radical liberal dream of an „invisible hand" as the only mechanism of a society is a dystopia where not only the weak are dominated by the strong, but where private interest is destroying its own foundation. At the same time, there is an ongoing integration of private aspects in the public sphere and vice versa. The idea of free speech for example means that a private person has every right to comment publicly any kind of political event. All democratic societies therefore organize a specific process of separation and integration of the public and private spheres. These processes are materialized in laws, institutions, culture and conventions, and in the technical means of communication.

For the actual political discourse about online falsehoods, the importance of the conceptional character of the separation/integration of the public and the private sphere cannot be overestimated. A change in the technical means of communication is necessarily a fundamental change of democracy. Historical examples are the invention of the printing press, advertising pillars, daily newspapers, radio and television: All of these techniques changed the democratic discourse. And the actual move towards social media as channel of political communication might be even more disruptive, as will be discussed later.

The separation/integration of the private and public sphere are linked to the philosophical question of subjectivity and objectivity. What is right from a private perspective might be wrong from a public point of view and vice versa. And what is to be considered right on the public level is in a democracy the result of the integration of contradicting private interests into the public sphere. Therefore, there is a natural right to be wrong. A law that sanctions falsehoods is necessarily damaging democracy, because it has to rely on an a priori decision of what is to be considered right and wrong, and this distinction destroys the separation/integration of the public and private sphere. We will discuss this aspect in regards to the German Netzwerkdurchsetzungsgesetz (Network Enforcement Law).

The digital revolution: Techno-social disruptions

Without going into detail the digital revolution that we are witnessing today can be summarized in three main points:

1. Access to the (mobil) internet at any time from anywhere is everyday life's reality (at least in industrialized countries). The amount of information that is public has exploded. What has to be considered the public sphere is growing exponentially.

2. Social media is the new channel of private communication. By connecting more and more people and by offering channels to address the public directly (like Twitter, open Facebook pages, blogs etc.), the former separation/integration of the public and private sphere is redefined.

3. Decisions about what should be public, and to which extent are more and more made by algorithms because the exponential growth in data cannot be handled in any other way. One side aspect of this shift is that institutions that used to safeguard the former separation/integration of the public and private sphere (like media and to some extant political parties) are losing influence.

The digital revolution of the public and private sphere is exceptional in at least two dimensions: Never before has the political communication of so many people changed in such a short time. We strongly believe that we are just at the beginning of this process and that the exponential growth of the last years will continue (at least for the next years).

The politicization of social media

As has been discussed above, the digital revolution is per se a political transformation. In addition we are witnessing today a strong movement for bringing more genuine political content into social media. This development is to some extent

just caused by the growth of social media services. The more content is shared, the more political content will also be discussed. Besides this "natural" growth of the absolute amount of political content both, social media companies and political institutions try to raise the relative share of political content. For social media companies, political communication is just a new market. For political institutions, social media provides a new way to reach out to private individuals in their function as "the public" (as electorate, supporters etc.). This leads to an enormous misfit in design: Social media has not been designed for political communication. The idea to connect private persons and steadily increase their outreach works, because this kind of communication is made so convenient. It takes just a click to offer a friendship, to like a post or to show your support. The whole communication is guided by private affinity and emotions. But political discourses should not be convenient. In democracies, politics should be the result of debates, which are often arduous, because a compromise between legitimated interests has to be found. It is therefore easy to argue that social media is by design not the right place for political communication. Given the above, social media will not disappear from the political discourse. The opposite is to be expected: The importance of social media for political communication is growing. Either we learn how to use these platforms in a way that fits better to what we are used to know as political debating culture or this culture will fundamentally change.

Online manipulation of public opinion

Along with the politicization of social media come the attempts to use these platforms in a manipulative way. Some studies even assume that the political opinion might be fundamentally changed by orchestrated attempts in the social network. Taking into consideration how difficult it is – no matter by which means – to change the political attitude of individuals, we think that these fears are exaggerated. It is very unlikely that anyone is changing his or her mind on important political issues just because of some suspicious accounts in social media. Nevertheless, it is a fact that in every political discourse, manipulative attempts with social bots (fake accounts controlled by a software), trolls (users engaging in debates to destroy them) and hyperactive users (accounts spending much more time than normal users to up- or down-vote specific content) can be observed. These manipulative attempts can have a short term and a long term effect. Firstly, it is very easy to create the impression that a specific opinion is very (un)popular online. Normal users are not so heavily engaged in political debates, and most of the time they stay passive. If someone – with the help of automation or with the help of others – is systematically engaging in a political discussion with the goal to raise the amount of posts, hashtags, likes etc. the debates becomes relative popular. Journalists, politicians or normal citizens might fall for these wrong trends and comment on them and thereby making them even popular. In addition, the algorithms steering who is seeing what on social media can take the strong activity in a debate as signal, and thus to distribute it even more widely. Again, the main danger is not that the distribution of these messages will lead to a change of mind by many people. But anyone who is monitoring what is going on on social media might get a wrong impression and make bad decisions. We are not able to prove it, but from the data we analyzed, we had the impression that the turn

in the public debate about the refugee situation in Germany might have been effected by such manipulative attempts. While many people in the real world were very optimistic and trying to do their best to help the refugees, the social media platforms were flooded with negative comments. What we know today is that people from the political right were using all kinds of online manipulation techniques to create this negative trend. What we do not know is, if any politician really fell for it and tried to serve this "mood of the people in the internet".

The second danger of the manipulative attempts is that in the long run, tensions already existing in society are amplified resulting in polarization. Some well known effects from social behavior studies – like homophily (the tendency to connect with like-minded others) – seem to point in this direction. From our own empirical work on Facebook and Twitter debates, we would argue that there is already a measurable effect of polarization caused by uneven distribution of information in these networks. But measuring these effects is an ongoing research project, and the reality of social networks is definitively more complex than simple explanations like filter-bubbles or echo-chambers. There seems to be a big variance in network structures and therefore it is very dangerous to transfer findings from one context (e.g. a national political debate) to another. Fact is that the now floating separation/integration of the public and private sphere makes it easy to target politicians and institutions in a slanderous way. Because social media is not one public sphere but a quite complex network of subgraphs negative campaigns might be at the same time very popular as well as nearly invisible for outsiders.

Who is trolling whom?

Based on leaks and revelations of official documents (most important the files released by Edward Snowden) it is very likely that the idea of manipulation of public online discourses was first systematically developed by western secret services like the NSA and GCHQ. Due to the ongoing discussion about Russia manipulating the US election, and considering, how easy manipulative techniques can be applied online, it is very unlikely that any secret service of sufficient size is not already engaging in this game. But again, engagement should not be confused with effects. For example the quantity of social media posts that are now connected to Russian interference (and this connection is in many cases quite questionable), seems to be very tiny when compared with the online efforts of the Trump- and Clinton-campaigns.

In addition, since the early days of "hacktivism" groups like anonymous, it is clear that online manipulation can be done by none state actors as well. In our own data studies that are focused only on Germany, the only group we can regularly connect to manipulative attempts are right wing nationalists, although we do not claim to have a comprehensive overview, Thus other teams with other methods might come up with different results.

What should be done?

We believe that the digital revolution with its effect on the separation/integration of the public and private sphere, is already changing democracy in a fundamental way.

Therefore, regulations and laws dealing with this new phenomenon seem to be necessary, not to stop this development – which is probably already impossible – but to steer it in favor of society. Policy makers face at least two very difficult challenges: As described above, the digital revolution is rapidly and ongoing transforming the public and private sphere. Any governmental action has to deal with the situation that the halflife of any solution might be very short. In addition, the separation/integration of the public and private sphere is a core element of democracies. There will always be a trade-off between personal freedom and public interest.

The German Government has introduced the NetDG (Network Enforcement Act). The first challenge is quite successfully addressed by this law: It was made very fast (first reading in May, 2017, passed end of June, 2017), it includes several mechanisms to monitor its effects (like a voluntary self regulation of the social media companies) and the politicians seem to be quite open to change this law quickly, in case it would be necessary.

If the law is interfering too strong into free-speech, is an ongoing controversial debate in Germany. One very critical point is that the social media companies themselves are responsible to delete content that is "obviously against the law" within 24 hours after notice. Many people argue that the judiciary should be the only institution deciding what is against the law. In addition, there is the fear that the social media platforms will delete many questionable posts just to avoid fines. But it should also be noted, that the NetDG is not a law against falsehoods. The question of right or wrong is mostly irrelevant to judge if something is against the law. The NetDG is strongly focused on hate speech instead.

The social media companies only have to act in case of any complaint. As long as nobody is flagging a post, there is no need to take it down. The users' experience has somehow priority. But in the political context, this system could turn out to be a weak point, because it offers a new target for online manipulation. Political opponents might systematically flag posts they do not agree with and thereby force the platforms to take these posts down. In addition, populists may claim that their posts where banned to increase distrust in institutions. Both is already happening in Germany.

Based on the arguments presented here, a way to find a better solution for the trade-off between private freedom and public interest could be to differentiate stronger between the authorship and the distribution of illegal content. We think social media companies should not be responsible for illegal content that is not produced by them but by the 5/6 users. On the other hand, the core business of social media companies is the distribution of users' content. This distribution could be regulated for example with fines for each additional view an illegal post has got, because of its algorithmic distribution on the platform. To be responsible for sharing illegal content would also apply to users who are distributing such posts via shares or retweets etc. Of course, the trade-off between private freedom and public interest is not vanishing with this proposals. But it might be a way to fine-tune it in a more appropriate way. In a nutshell, there might be many things that should be allowed to be said, at least until it is proven that they are illegal. This is part of the private sphere and it is a core

element of democracy. But this does not mean that there is a right to get all these messages distributed to the public sphere. Everyone who is engaged in public discourses bares a responsibility that goes beyond their individual existing in the private sphere.