IN THE GEARS OF THE OPINION MACHINE

TEXT: PETER HERGERSBERG

Nowadays, political debates often turn into verbal brawls – especially on social media. In order to counteract this, Eckehard Olbrich and Sven Banisch of the Max Planck Institute for Mathematics in the Sciences in Leipzig and Philipp Lorenz-Spreen of the Max Planck Institute for Human Development are investigating how polarization occurs and how opinion formation in groups works.

To explain why he studies political developments, mathematician Sven Banisch describes an experiment with rats. Biologists observed that the same ratio of cooperating, solitary, and bullied animals always emerged in small groups of the rats. The same order also emerged when animals of only one type, (e.g. solitary animals) where brought together in a new miniature society. "I want to understand such forms of social self-organization," says Banisch.

The formation of a consensus or else polarized viewpoints in a debate also involves a self-organizing process. Experts refer to the dynamics of opinion, which is something being studied at the Odycceus (Opinion Dynamics and Cultural Conflict in European Space) project, which Banisch initiated in collaboration with Eckehard Olbrich, who heads a research group at the Max Planck Institute in Leipzig. In this four-year project (2017–2021), which Olbrich is now coordinating, researchers from eight institutions in six countries have been studying, among other things, how we position ourselves in controversies or how populism drives opinions to the edges of the political spectrum, particularly now that online and especially social media are shaping political disputes. Debates in social media are rarely moderated and often degenerate into a heated verbal exchange.

"Digitalization," as Olbrich explains, "means that we have much more and faster-moving information at our disposal. And comment functions and social media mean that even more people have the opportunity to express an opinion. It's almost impossible to follow debates in detail and understand why conflicts arise." That is why some of the Odycceus partners have developed mathematical tools to help manage the flood of information. "We hope to help make debates more transparent and disputes more objective," says Olbrich.

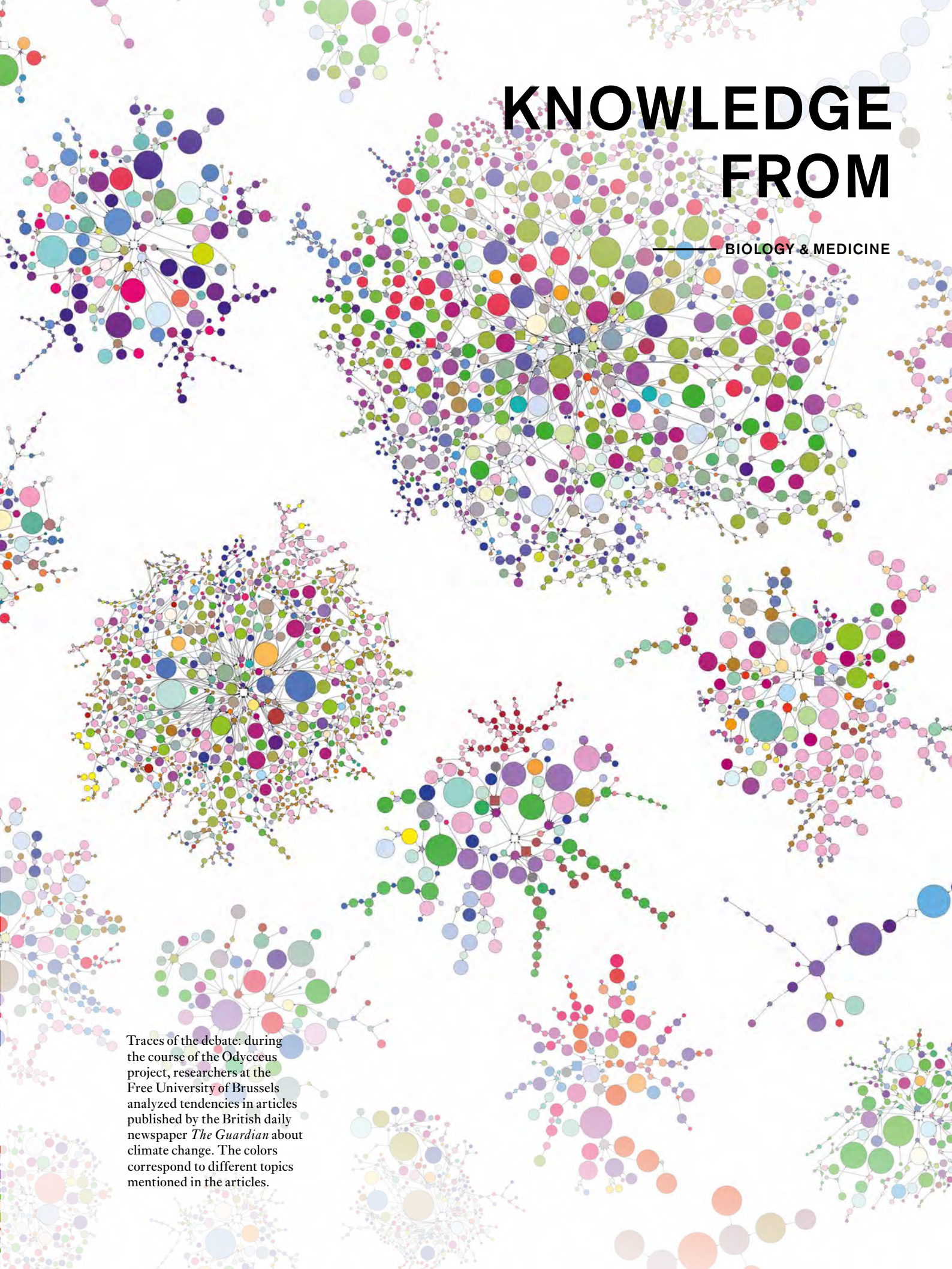
To this end, his colleague Sven Banisch is developing models that depict the conditions under which a debate leads either to a common viewpoint or to irreconcilable differences. The mathematical parameters and starting conditions with which the models realistically reproduce these processes enable the researchers to conclude which social factors play a role. Banisch and his colleagues use empirical data to verify the models. This works particularly well with experiments conducted under defined conditions. Together with economist

→

GRAPHIC: VU BRUSSEL/MICHAEL ANSLOW

KNOWLEDGE FROM

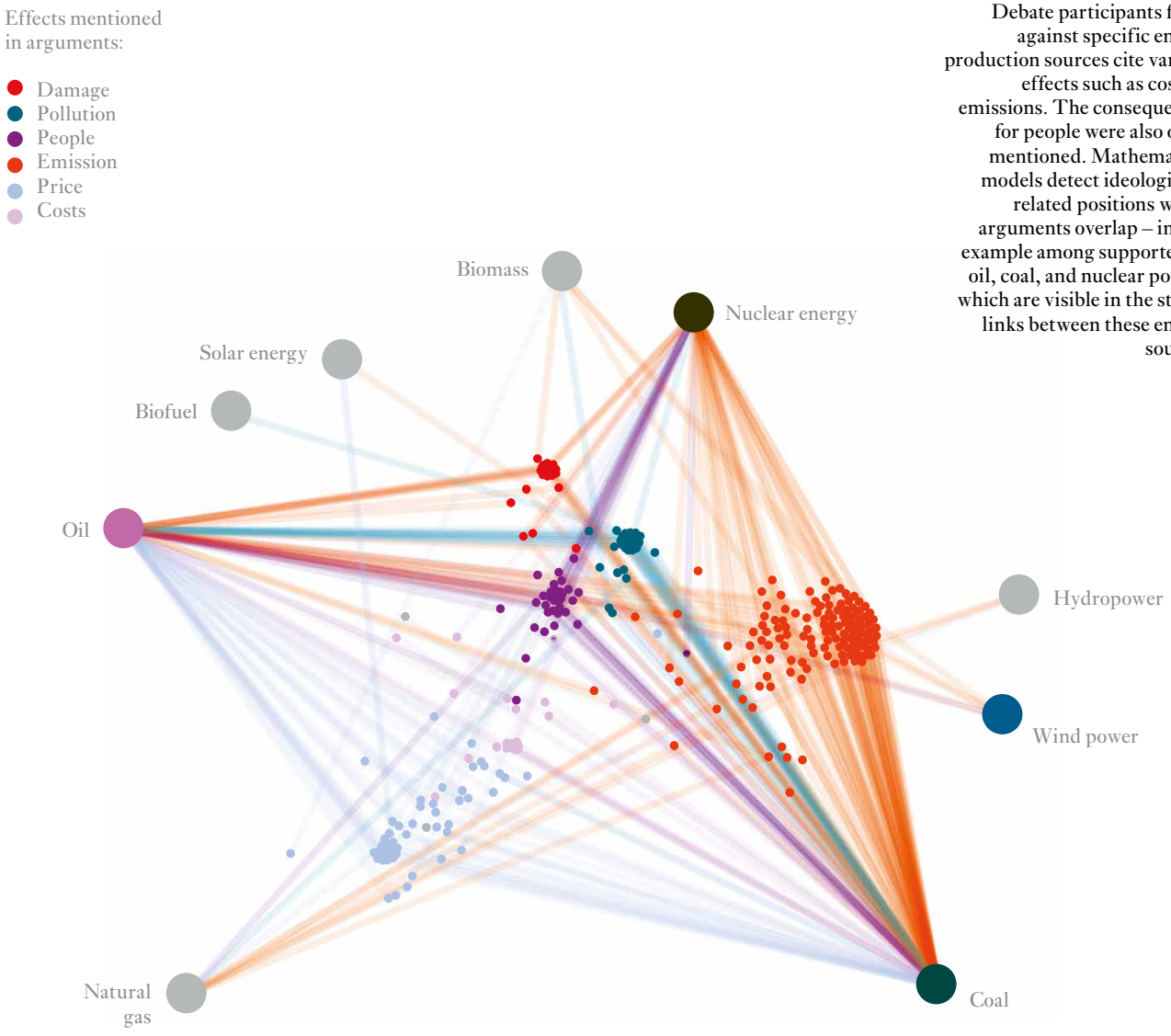
BIOLOGY & MEDICINE



Traces of the debate: during the course of the Odycceus project, researchers at the Free University of Brussels analyzed tendencies in articles published by the British daily newspaper *The Guardian* about climate change. The colors correspond to different topics mentioned in the articles.

Hawal Shamon of the Research Center Jülich, Banisch investigated how strongly the biases held by various test subjects towards different energy sources determines the formation of opinion within a group. This cognitive bias causes ardent supporters of coal or wind power to find the arguments in favor of their preferred energy source to be more compelling, which is hardly surprising. However, the collective impact is surprising, as a strong cognitive bias can result in polarization within a group. However, a different effect occurs if people simply prefer one form of energy less strongly: "We were also surprised by the fact that when the bias was weak, the group quickly agrees on a stance," says Banisch. This was the case with energy sources such as biomass and gas, which are not discussed as prominently and heatedly in public. Groups also reach a consensus in the complete absence of biases, although the consensus remains undecided between pro and con and also takes a long time to reach. "Our hope is that debates will be more constructive if we demonstrate the effects of evolved cognitive apparatus at the collective level," says Banisch.

This bias cannot even be overcome through discussions with others – on the contrary. This was demonstrated by a team in which Philipp Lorenz-Spreen was involved. The physicist, who works at the Max Planck Institute for Human Development in Berlin, relies on models as well as data from social media. Together with German-Italian partners, he has shown how social discourse can divide opinions. The more avidly users post comments, the more extreme their views become. In this way, the researchers reproduced the divide that emerged on Twitter in rela-



Debate participants for or against specific energy production sources cite various effects such as costs or emissions. The consequences for people were also often mentioned. Mathematical models detect ideologically related positions where arguments overlap – in this example among supporters of oil, coal, and nuclear power – which are visible in the strong links between these energy sources.

GRAPHIC: SVEN BANISCH/MPI FOR MATHEMATICS IN THE SCIENCES

tion to three issues that dived opinions in the U.S. – Obamacare, gun control, and abortion – although the positions predicted by the model were less widely divergent. "The initial hope that the Internet could facilitate constructive discourses between people with different opinions has not been fulfilled," says Lorenz-Spreen. The disintegrating effect of the discourse is due to psychology: we find it difficult to put up with opinions that differ from our own. We tend to be social homophiles and prefer to discuss things with like-minded people.

Discourse promotes polarization

This creates echo chambers in which one opinion clearly predominates and people can push each other toward increasingly extreme views, which do not even need to be supported by substantive arguments; pro and contra comments are sufficient: "No way is that possible" or "Yes, I totally agree" are among the more civilized statements; support can even be shared by simply clicking the "like" button. Social scientists refer to such comments as social feedback. Banisch and Olbrich have shown that posts in favor of or against a particular viewpoint rapidly help to create unity in a group. Again, the reason is homophily: most people enjoy pats on the back and fear slaps in the face even if these are only verbal, which is why group members increasingly rally behind a particular opinion which may only have been expressed as a tendency at first. This makes perfect sense from an evolutionary point of view, as this is how our early ancestors quickly reached decisions when it came to spontaneously gathering the group for hunting and taking up arms or fleeing in the face of hostile hordes. "But in larger groups," Banisch explains, "social feedback on differences of opinion quickly leads to polarization and the creation of echo chambers." The extent of the social consequences of the echoed opinions is still under debate. Studies from the U.S., for example, show that even the most active opinion leaders rarely spend time exclusively in echo chambers. Most people also consume other media – even if the reporting might be as unbalanced as Fox News.

SUMMARY

Researchers at the Max Planck Society want to uncover the mechanisms that lead to the polarization of debates in social media. They initiated the Odycceus project to counteract this.

Bias and social discourse can lead to polarization, in the course of which a loud minority can silence the quiet majority.

Artificial intelligence can help to deduce opinions and lines of argumentation from texts.

Analysis of Tweet battles

The fact that the societal implications of the modeled effects are not yet clear is not the only reason to treat the model results with caution – even by the modelers themselves: "We simulate mechanisms of opinion formation – but not real behavior," says Lorenz-Spreen. Real behavior is also determined by environmental influences and individual differences, which the current models do not consider. Olbrich also thinks that the simulations could be more realistic: "There's still a lot of room for improvement." Nevertheless, the model calculations are helpful even in their current form. "They help us develop theories about how opinion formation works," says Banisch.

For example, a theory about how opinions on various topics result in a closed view of the world and why camps are formed in the process. Why, for example, are proponents of rigid climate protection often also more open to immigration – and vice versa? Banisch and Olbrich were able to use a model to illustrate this with an example relating to energy policy where advocates of coal-fired and nuclear power are comparatively close whilst the distance between them and advocates of renewable energies is significantly greater. This is because some arguments such as the need for a stable energy supply or the opinion that wind turbines and solar parks disfigure the landscape speak in favor of both coal and nuclear power. In contrast, there is less overlap with the arguments in favor of renewable energy. So, different ideologies are formed on the different argumentative foundations.

The team also uses empirical data to research the political clashes between ideological camps. Social networks, especially Twitter, are ideally suited for this purpose because they contain a wide range of information. In the case of Twitter, conflict lines and alliances can be identified by who retweets which tweets (i.e. shares them with their own followers) and who replies to whom. The team has therefore written software that creates and visualizes networks of Twitter users and their tweets. Each node represents a user, and each line between two nodes represents a retweet or a reply. The program also arranges strongly linked nodes close to each other. To enable interested parties to analyze debates themselves, the researchers have put the software online under the name Twitter explorer along with instructions on how to install it. They then used the algorithms to examine two tweet battles in more detail: one about the 2019

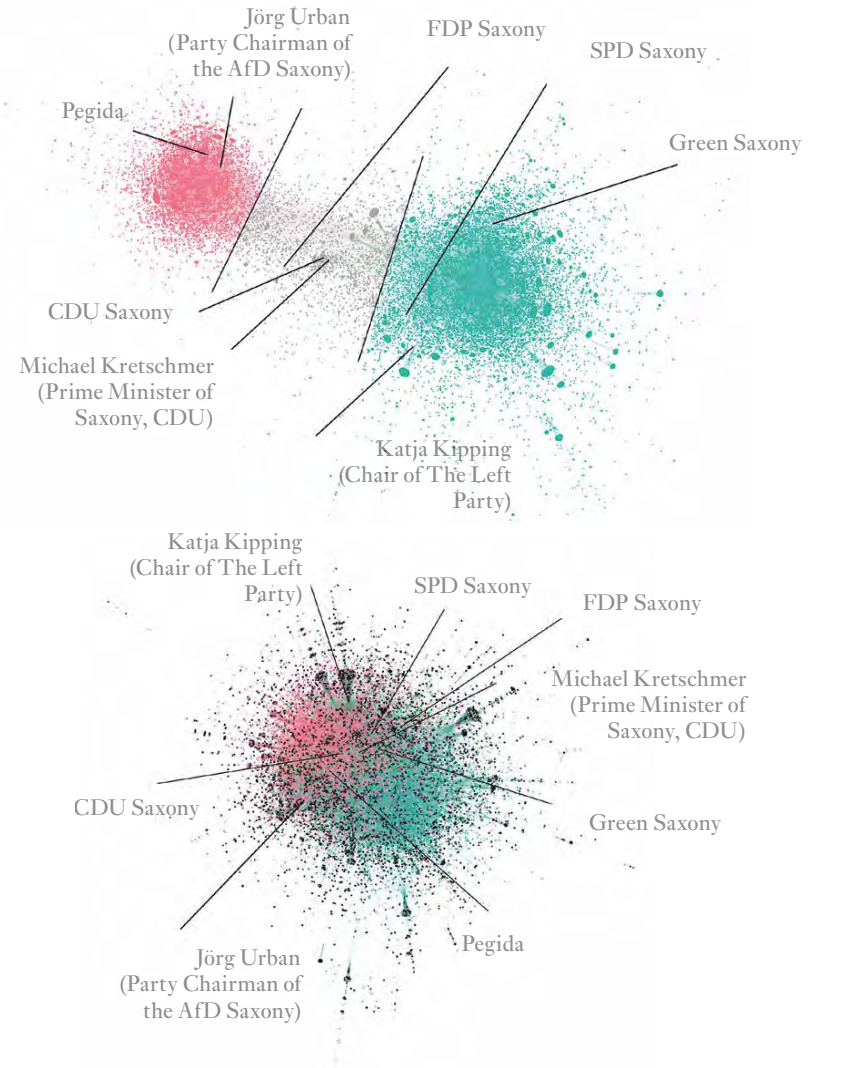
→

state elections in Saxony and one about riots in Leipzig on New Year's Eve 2019/20. In both cases, two poles of opinion formed in the retweet network along with a much weaker area in between. But the poles were also occupied to varying degrees. The majority group, which included politicians from the SPD, the Left Party, and the Greens and certain media outlets such as MDR Sachsen and Bild Leipzig retweeted each other's post more frequently and had almost three times as many followers as the minority group, which consisted mainly of representatives of the AfD and Pegida (very right wing paties). The middle ground was occupied by politicians from the CDU and FDP as well as other media such as MDR aktuell or Bild Dresden.

Psychological spiral of silence

A completely different picture emerged for the network of replies: the two clusters now seemed to be virtually wedged into each other – the right-leaning minority in particular sought confrontation. It responded as often as the majority, which was three times larger. This is where psychology comes into play. Because most people fear criticism – especially insults – they refrain from making statements in the case of doubt. "This can lead to a spiral of silence," says Felix Gaisbauer, a doctoral candidate at the Max Planck Institute for Mathematics in the Sciences: "A loud minority can silence a quiet majority to such an extent that the public may perceive it as a majority." The main danger arises when the traditional media portrays events on Twitter, for example, thus providing the minority with an even bigger platform on which to vent their outrage.

To make disputes more constructive, the Odycceus partners are developing mathematical tools for text analysis. "We want to extract opinions, lines of argumentations, and how people make causal connections from texts and process them in such a way that they are easier to comprehend," says Olbrich. The researchers are relying on artificial intelligence (AI), or more precisely, on machine learning. For example, a team led by Katrien Beuls, and Tom Willaert of the Free University of Brussels has developed an opinion facilitator that recognizes causal relationships in texts and catalogs causes and effects. As an example, the group applied the tool to articles on climate change that appeared in the Guardian newspaper in England. It shows whether different causes are discussed for an effect such as global warming (e.g., human greenhouse gas emissions or increased solar activity). Olbrich also specializes in machine text comprehension. In doing so, he combines his mathematical skills with his own political interest. He is working specifically on computational rules – primarily topic models – that are designed to extract opinions from



Visibly divided: supporters of an AfD-affiliated (pink) and a more left-liberal (turquoise) camp mainly shared like-minded posts about the 2019 elections in Saxony. This appears as polarization in the retweet network (above). There are fewer users (gray) between the two camps. In the reply network (below), the two clusters move closer together because it was mainly the AfD-affiliated faction that replied to tweets published by the opponents.

GRAPHIC: FELIX GAISBAUER, ARMIN POURNAKI, SVEN BANISCH, ECKEHARD OLBRICH – IDEOLOGICAL DIFFERENCES IN ENGAGEMENT IN PUBLIC DEBATE ON TWITTER

press reports, posts, and other online publications. To do this, the programs must first recognize the topics – even if they are not explicitly named. The algorithms learn from training data (e.g., how likely a text is to be about mobility if it contains terms such as traffic, train, highway, and airport) after which they recognize such correlations independently, even for new topics such as the environment or law and order.

Olbrich based the formulas on the election programs published by political parties in Germany and other European countries as well as the U.S. in the past decades. The Berlin Social Science Center (WZB) has digitalized the political declarations of intent in the Manifesto project. "We can identify topics quite well using the topic models," says Olbrich. This helps to compare the programs. The goal now is for the AI program to recognize attitudes toward the relevant issues, whereby the researchers are always trying to understand how the algorithms arrive at their results, which is by no means self-evident: "Practical users, such as the Google team, are much more advanced when it comes to text interpretation," says Olbrich. But it is difficult to discern which criteria the algorithms use to assign meanings. It remains unclear whether the algorithms actually recognize a real connection or only construct it. Clearly, the latter can distort debates just as much as fake news can.

Quality inspection through nudges and boosts

The Leipzig team also wants to use mathematical tools for text analysis to help answer questions that the social sciences are currently working on. This was also a major motivation for Olbrich to initiate the Odycceus project in 2015. Specifically, the aim was to uncover the background of the Pegida movement, which gained significant popularity after 2014. Were the demonstrations against Muslim citizens just another right-wing populist movement? Can Pegida still be classified in the traditional left-right scheme, which is strongly influenced by economic criteria, with the proponents of the free market on one hand and those who argue for more government intervention and redistribution on the other? Many sociologists currently tend to identify the line of conflict along cultural differences associated with open-mindedness and patriotism. Olbrich is working on using algorithms to deduce this reconfiguration of the political space from relevant texts.

Social media can act as a lubricant for such a transformation. One example is the rift that Donald Trump's Twitter rage has continued to widen for years whereby his most important tool consists of alternative facts. If fake news could be more easily identified as such and less easily disseminated, the common factual basis that many debates currently lack could re-emerge. Lorenz-Spreen would therefore like to encourage social media users to pay more attention to the robustness of the claims people make. Together with an international team led by Ralph Hertwig, Director at the Max Planck Institute for Human Development, he has developed proposals based only on external characteristics rather than substantive criteria: a content presentation that makes its credibility easier to recognize based on the sources or senders is just as much a part of this as attempts to slow down the sharing of posts, for example, by requiring additional clicks if users want to forward a message without having read it. Social scientists refer to such interventions as "nudges". "We are fully aware that nudging can be paternalistic," says Lorenz-Spreen, "so, it would therefore always have to be made transparent that these are nudges."

Boosts, which refer to the ability to judge the quality of a piece of information, are less paternalistic intervention mode. Navigating through a decision tree with corresponding (again external) cues before or after reading could give users such a systematic quality check boost. "Boosts require a relatively large motivation. But unlike nudging, they are likely to work even when they are removed," says Lorenz-Spreen. However, all of the team's proposals would have a profound impact on the business model of YouTube, Twitter, and similar platforms, as they are likely to slow down the dissemination of information thereby depriving the platforms of attention "which," as Lorenz-Spreen explains, "means less profit." The only way out he sees is for users themselves to demand more transparency and, for example, to migrate to alternative platforms that could work along the same lines as Wikipedia. "In the long term, I can even imagine public institutions operating such platforms along the lines of public media," says Lorenz-Spreen. However, for now it does not work without rules.

The European Commission has demonstrated its openness to regulation with the Digital Service Act, which aims to establish security and liability rules for digital services. Germany too has already taken a step against criminal digital content with the Network Enforcement Act. Whichever additional standards should be applied to social media is still open to public negotiation. "Regulating digital presences down to the smallest detail is difficult. But design details definitely play a big role," says Lorenz-Spreen. Going forward, he would also like to work with Olbrich's team to study which measures could lead to a more careful handling of information arising from social media, because although the Odycceus project is now coming to an end, the road to a fair, fact-based exchange of opinions is still long – and not just in the digital world.