Link

www.hfp.tum.de
politicaldatascience.blogspot.de

Campaign 4.0 – **On the Trail of the Elusive Bot ...**

Online social networks have now become a part of political discourse, recently attracting a great deal of attention due to their role in the Brexit referendum and presidential elections in the US and France – not to mention the emotive tweeting of a certain US President. In the struggle to gain influence, intelligence services, terror organizations and political movements are increasingly deploying social bots, whose mass-messaging can polarize, mislead and unnerve other users. Simon Hegelich, Professor of Political Data Science at the Bavarian School of Public Policy within TUM, is working with his team to find ways of detecting these computer programs, helping to stop the originators in their tracks. Germany's parliamentary elections are the next big event.

Karsten Werth

Wahlkampf 4.0 – Meinungsmaschinen auf der Spur

Die sozialen Netzwerke im Internet sind wegen ihrer Rolle beim Referendum über den Brexit, bei den Präsidentschaftswahlen in den USA und in Frankreich, aber auch durch einen emotional twitternden US-Präsidenten ein viel beachteter Teil des politischen Diskurses geworden. Im Kampf um Einfluss setzen nicht nur Geheimdienste und Terrororganisationen, sondern auch politische Bewegungen vermehrt Social Bots ein, die durch ihre massenhaft verbreiteten Botschaften polarisieren, täuschen und verunsichern können. Simon Hegelich, Professor für Political Data Science an der Hochschule für Politik an der TUM, sucht mit seinem Team nach Wegen, diese Computerprogramme zu entdecken und dabei zu helfen, ihren Urhebern das Handwerk zu legen. Unter Einsatz neuer Methoden, einer Kombination aus IT-gestütztem Data-Mining und politikwissenschaftlicher Interpretation von großen Mengen von Nutzerdaten, kommen die Forscher immer komplexer werdenden Bots auf die Spur, die unter falscher Identität im Netz Stimmung machen. Im Jahr der Wahl zum deutschen Bundestag konzentrieren sich die Forscher auf die Debatte zur Flüchtlingskrise im sozialen Netzwerk Facebook. Sie haben bereits eindeutige Manipulationsversuche aus dem rechten politischen Lager festgestellt und arbeiten am Aufbau eines Monitoring-Systems, das dabei helfen kann, mehr Transparenz in die neuartige Form der politischen Willensbildung über die sozialen Medien zu bringen. □

Picture credit and Graphics: edlundsepp

"A people that no longer can believe anything cannot make up its mind. … And with such a people you can then do what you please." *Hannah Arendt*

It's a dream come true," acknowledges Simon Hegelich: "You pursue your research and suddenly everyone is interested in it. But the past few months have certainly been hectic." Specializing in social media forensics, he is now in high demand prior to Germany's 2017 election as an expert in media and politics. In one week in February alone, he had meetings in Berlin with parliamentary committees, the Ministry of Education and the Federal Press Office and also attended a panel debate held by public broadcaster ZDF as part of the Berlin International Film Festival. The interest in Hegelich's research into systematic manipulation via social networks is easy to understand, since it is becoming increasingly clear that this facet of our fast-evolving technology lifestyle is also heralding radical changes to the political process. In fact, Hegelich suspects Germany could be about to witness the last traditional election campaign – and possibly the dawn of a new era.

Direct communication – one to all – without any filters

"The democratic public space is undergoing disruptive change," emphasizes Hegelich. "In principle, social media platforms are enabling direct, one-to-all communication. This is shaking up the traditional role of political parties and the media in forming opinions. In historical terms, the situation is comparable to the invention of the printing press. That suddenly allowed information to be widely disseminated – and in a way no longer tied to the old frameworks of power and discourse." What is more, these new communication channels are emerging at a time of increasing anxiety about the future and loss of trust in politics and the media. So on one hand, we are facing political change due to factors such as the financial crisis, economic challenges in the US and upheaval in the European Union. This change would also be occurring without social media. On the other hand, though, we now also have technology that allows us to share our feelings in public. That certainly does nothing to calm the situation – in Germany, for instance, Internet users are becoming increasingly polarized. And, as all over the world, populists are also campaigning for votes online. These forces feed into each other, making it difficult to gain a clear picture of the situation here. As Hegelich puts it: "It is hard to pinpoint cause and effect. Did Twitter really help Trump into power? Or is Twitter only still in the market because Trump is such a prolific tweeter?"

Machine against machine

Another revolutionary factor is the fact that it's not just people who are active on social networks. Increasingly, they are being joined by machines. In 2015, Hegelich was able to show that Ukrainian ultra-nationalists were using a computer program to control 15,000 Twitter profiles, sending up to 60,000 tweets per day. Since then, he has delved even more deeply into the topic and is now investigating how robots fuel discussions and generate or strengthen opinions. Itself a player

in the cyber arms race, the research community is also benefiting from progress here. Thanks to new developments in machine learning, with algorithms autonomously detecting patterns in large data volumes and enabling them to be interpreted, political scientists can now harness the power of big data in their work. At the same time, the speed at which the Internet is evolving poses a real challenge for traditional research activities. While a researcher is busy submitting project proposals to apply for funding from the relevant institutions, the online environment continues to evolve. And by the time applications have been processed, staff hired and resources secured, it might look very different again. Two years from the original project idea, for instance, the social media platforms selected for study might have changed completely or lost a huge amount of impact. Hegelich's team is still small, but he is aiming to monitor social media around the German parliamentary elections in as much detail as possible. Its members are currently working to build capacity, gather data and program analysis tools. To date, there is no software for detecting social bots that could keep up with the rapid pace of development. Somewhat surprisingly, from a non-specialist perspective, the arsenal of Munich's social media forensics experts includes Raspberry Pis – small, single-board computers. These mini-computers are used to connect to the social networks. The researchers then conduct their analysis using advanced databases, such as Elasticsearch, and high-end servers. However, this data mining does not currently involve supercomputers such as you might find in the Physics faculty, for instance.

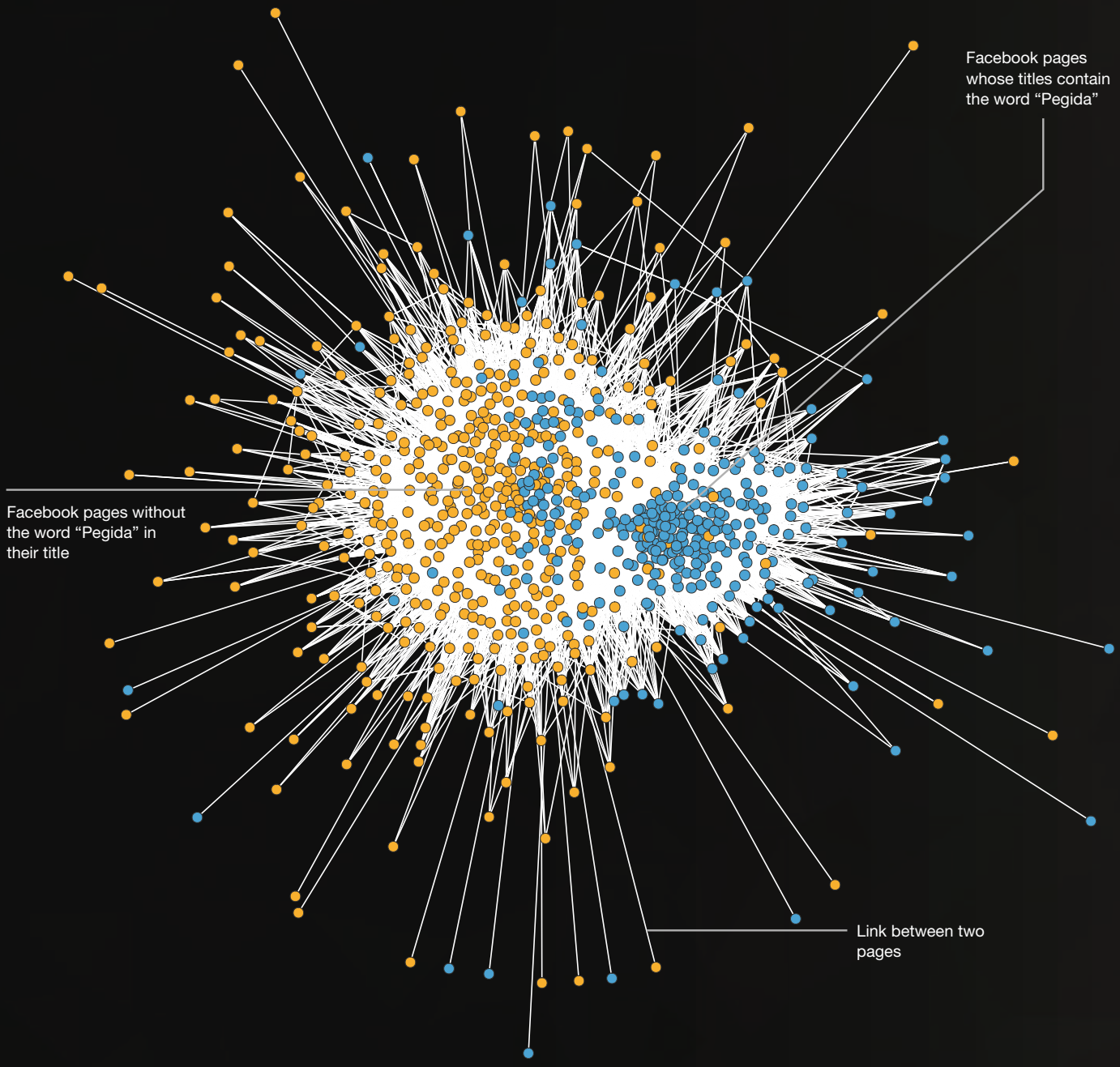
Political and data science team up

"We do a lot with programs we have written ourselves," explains Hegelich. "For instance, we are taking part in a fake news challenge – the task being to check 50,000 articles and see whether the headline matches the text." On Facebook it is often the case that real articles are shared with fake headlines so they are more likely to be circulated. That is why users should look not only at the Facebook preview, but also at the original source. Hegelich's team is trying to automate machine-based checking in these cases. When conducting research with social media data, the interplay between IT know-how and political expertise is particularly important. "If you just let the data speak for itself, the outcome is nonsense," underscores Hegelich. "So we look at bot mistakes from every angle. Otherwise you end up with theories like the one blaming bots for Brexit. It's really not that simple though. You don't just type 'Brexit' into the Internet and people are suddenly in favor of it." If you use machines for data analysis, continues Hegelich, you need to know what that means: "If I've developed an algorithm that classifies data correctly in 95 percent of cases, that sounds absolutely great. But if I'm using it to sort a billion posts per day, it will get a substantial number of them wrong. So it's not enough to look at just one aspect." ▷

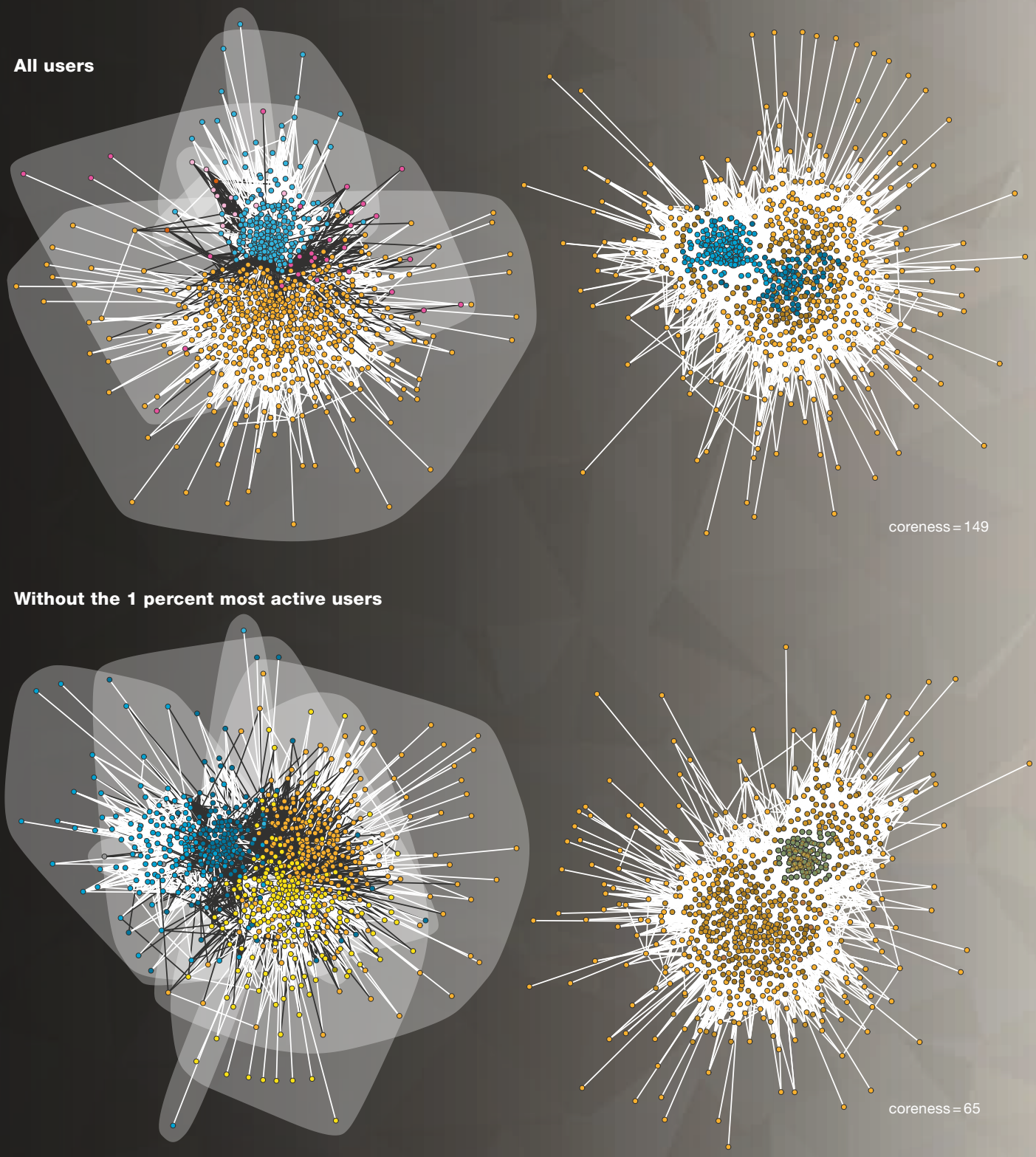
*Alessandro Bessi, Emilio Ferrara. Social bots distort the 2016 U.S. presidential election, online discussion. First Monday 21(11), 2016

Quote taken from The New York Review of Books, issue October 26, 1978; Graphics: ediundsepp (Source: Hegelich)

How hyperactive users affect social network structures



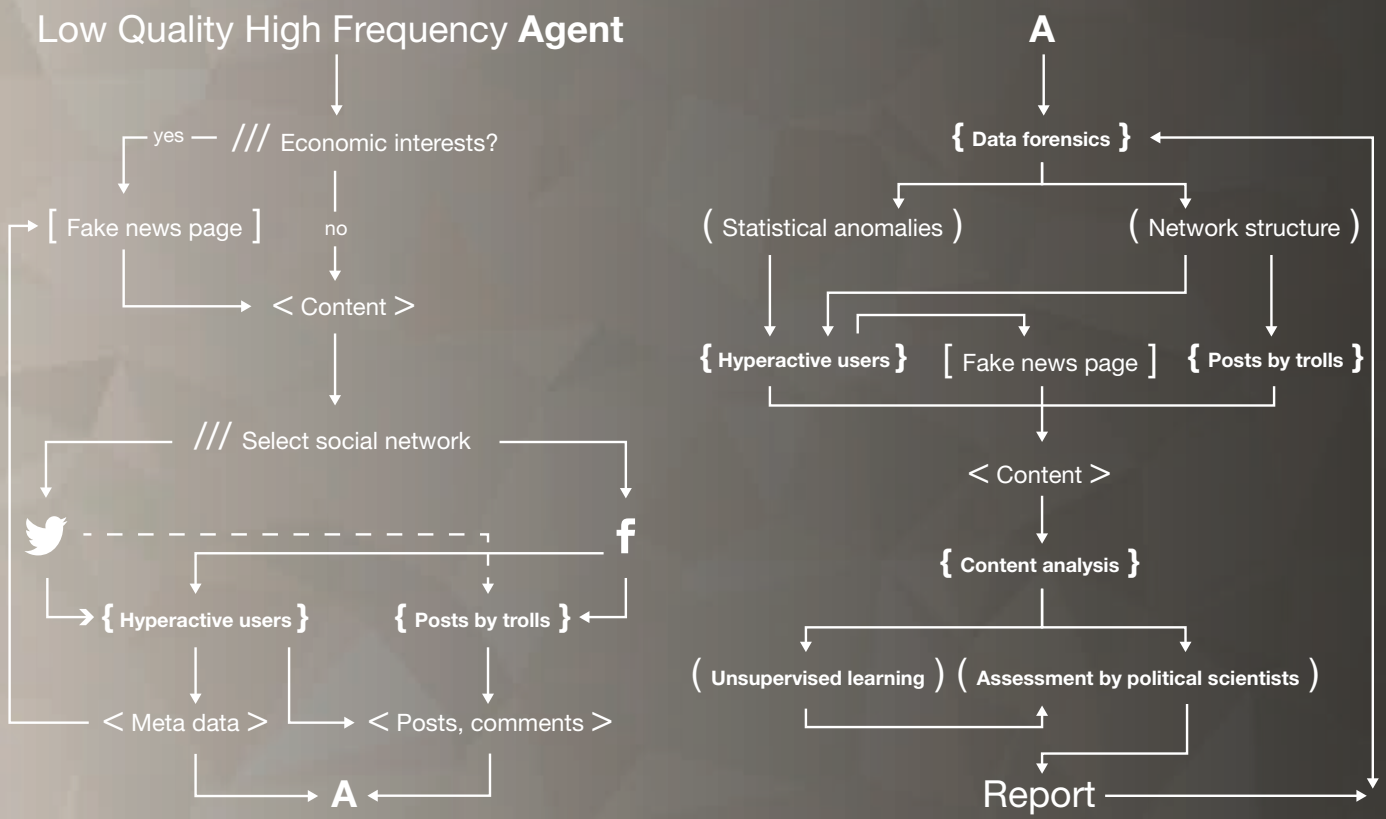
Can the structure of a social network be manipulated? This is an important question, because the algorithms based on which social networks display content to their users analyze the network structure. In 2016, Hegelich analyzed nearly 1,000 German Facebook pages (including posts, comments and likes) of all political parties, main media and of all pages with the word “Pegida” in the title. Pegida is an anti-immigrant initiative in Germany. Orange dots stand for “mainstream” pages, blue dots for pages which have “Pegida” in their title. The lines indicate that a user has linked these two pages. The resulting network is cleary separated into two clusters. The few blue dots within the orange network stand for “anti-Pegida” pages, which also use the word “Pegida” in their title.



Does the network structure for all users (top) change when the hyperactive users are eliminated (bottom)? The graphics on the left analyze the so-called statistical distance, which describes how strongly two pages are linked. For all users, the network is separated into a "mainstream" (orange dots) and a "right-wing" (blue dots) cluster. Without hyperactive users, this structure changes: Both clusters break up into two groups. The distance between "mainstream" and "Pegida" pages shrinks, indicating higher information flow between them.
The graphics on the right consider pages which are linked to a very high number of other pages. A coreness value of 149 means that they show only pages that are linked to 149 or more pages. Orange stands for a lower number of linked pages, blue for the highest number of connections. For all users, this network is clearly divided into a highly linked and a relatively weakly linked cluster. Without the hyperactive users, the two clusters are less clearly separated and the coreness value is much lower, indicating a less closely knit network altogether.

Graphics: ediundsepp (Source: Hegelich)

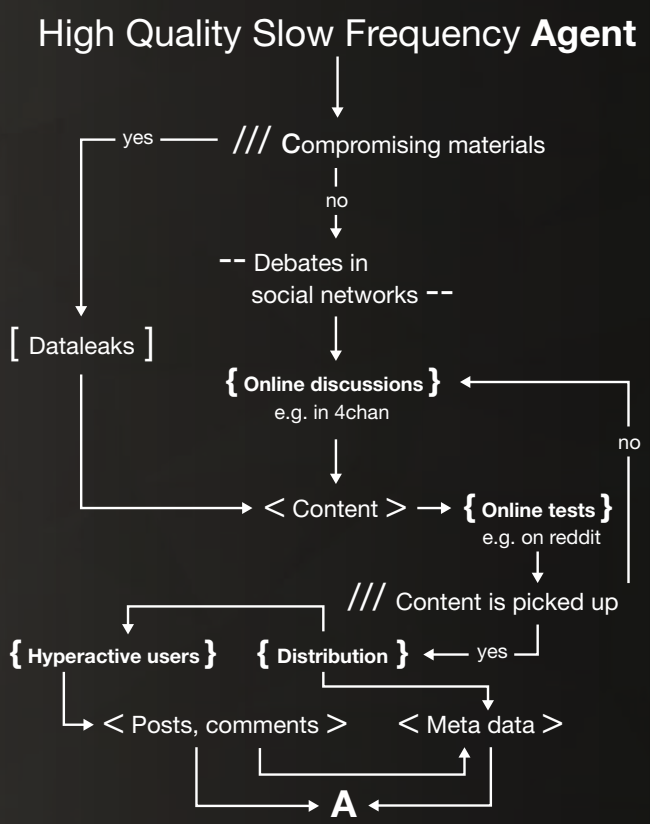
Opinion-forming tools in social media

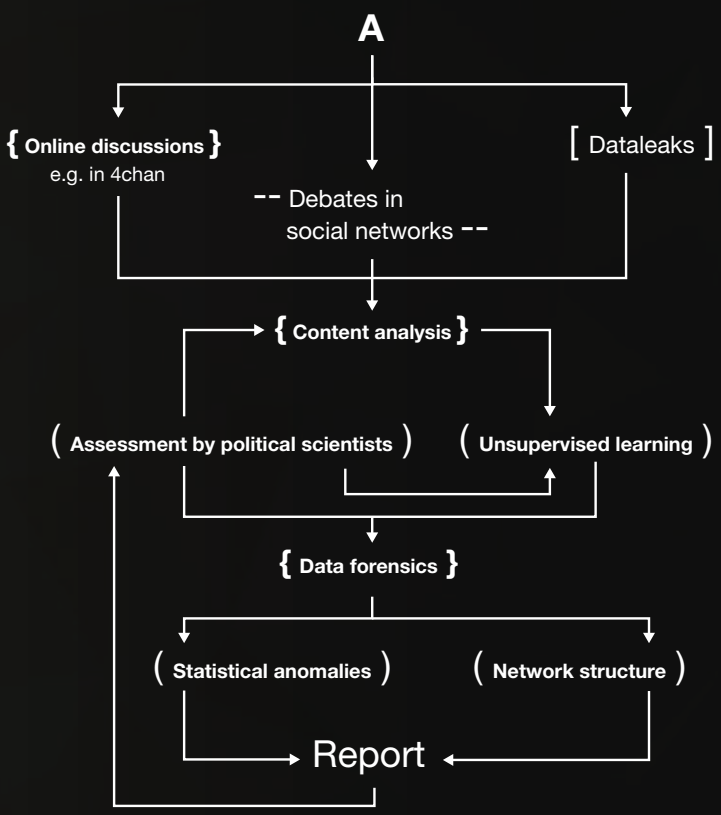


Left: Setting up a high number of relatively simple postings (Low Quality High Frequency Agents): A certain piece of content – either fake news or real news – is streamed into a selected network. Fake news pages are mostly set up in order to raise money from clicks on ads. Once streamed into the networks, the content gets distributed with the help of hyperactive users or trolls. The result (A) is a certain opinion that forms within the network.
Right: To uncover this manipulation, data scientists use statistical methods to find hyperactive users or trolls. Political scientists analyze the content of their posts and assess whether or not manipulation should be suspected.

Who or what are social bots?

The social bots under investigation by the TUM researchers are computer programs deployed for manipulative purposes that use fake accounts to emulate human identities and communication. They capture user data and spread targeted spam or political messages. In the US presidential race, for instance, bots sent out mass notifications that Donald Trump lied in the TV debate or that Hillary Clinton suffers from Parkinson's disease. They are particularly active on social networks using easily accessible application program interfaces (APIs). Twitter's low entry barriers mean it is a heavily used platform. It is currently at the center of international research, since it makes it relatively easy to procure the necessary data. However, Hegelich takes a critical view of this focus in relation to Germany, since Twitter has an extremely small following in comparison with Facebook there. Recently, Hegelich has been concentrating his own efforts on analysis of the refugee debate on Facebook. Prior to the upcoming German parliamentary elections, he intends to turn his attention to manipulation via social media that could influence the campaign. "We are also working to incorporate other platforms like 4chan, VKontakte and reddit in our analysis, but are reaching the limits of our resources there," he admits. Hegelich's previous studies, which analyzed over 30 million instances of Facebook activity in relation to the refugee issue, clearly show right-wing manipulation attempts. Hyperactive users – both human and automated – became apparent, systematically "liking" every post by Germany's right-wing populist AfD party. There are people spending eight hours a day writing hate-filled comments about refugees on the Internet. This huge engagement gives their views an exceptionally wide reach on Facebook. At some point, the site's algorithms





Left: Setting up conspiracy theories (High Quality Low Frequency Agents). A story to discredit the target is either sourced from data leaks or set up as fake news, tested in discussion platforms like reddit and then distributed via hyperactive users. As a result, a certain opinion forms within the network. **Right:** Such agents are difficult to uncover, because they cannot be detected with statistical tools early on. Political scientists watch sites which have been identified as at risk. If one suspects that a conspiracy theory is being launched, statistical tools help to assess how far it has already been distributed. A conspiracy theory can be made public and invalidated only if it is detected early enough.

kick in, ensuring that such an apparently popular topic is more and more visible to other users. And that achieves the aim of this type of purely statistical manipulation. According to Hegelich: “You look after the first 20,000 clicks yourself and hope it all takes off on its own from there. The topic then goes viral.”

Depending on their development level, social bots have varying capabilities when it comes to mimicking human identities. Simple bots recognize key terms such as “refugee” and respond to them by posting images or retweeting comments. These simple bots are currently the most prevalent on the Internet. Generating them takes little programming skill, and manuals and instructions are freely available online. Hegelich himself has published one such guide on his blog. “It’s about ten lines of code, and then you have a Twitter bot,” he ▷

Graphics: ediundsepp (Source: Hegelich); Picture credit: ediundsepp

confirms. More complex bots, meanwhile, can analyze the content of communications and engage in dialog. They cloak themselves with copied profile photos and follow a regular daily routine – just like the average human user.

Everything as a service

Social bots are also available to order. The business of creating fake digital identities is based on a value chain with globally distributed production. As Hegelich explains: "The slogan in Silicon Valley just now is 'everything as a service'. Sadly, on the Internet, that also extends to manipulation. You can buy 10,000 fake accounts for a few hundred dollars, for instance – and the software to run them can be purchased too. A highly sophisticated variant you could use to operate your 10,000 Twitter accounts costs around 500 dollars."

Not always political

Often, these purchases are not motivated by politics. If someone uses bots for promotional purposes, for instance, they automatically register them. And for each new registration, Twitter suggests users to follow. To make their human persona as believable as possible, the bots accept all these suggestions and thus end up following Donald Trump or Hillary Clinton, for instance. Hegelich is sure that many of the fake accounts attributed to the candidates have no political purpose at all: "And of course that distorts the picture. We need only recall the noise made by the media here about how many followers Trump has on Twitter. Yet these numbers tell us absolutely nothing!"

How do you spot a social bot?

The team's own initial studies have already yielded large data sets, which Hegelich is using for comparison purposes to establish how social bots behave. A computer program imitating human behavior will always generate recognizable patterns. The TUM analyses do show that bots have adopted a regular daily routine – they no longer post at night or every ten minutes, which is far too easy to detect, but now go to sleep and take lunch breaks. Over a longer time, though, it becomes evident that they are just as active over the weekend as during the week, for instance – unlike the average human user, who posts a lot less then. Patterns like this can be identified through data mining. As Hegelich describes: "We put all the data into the computer and say, 'These are bots and these are not, so now tell us how they differ'."

Uncertainty – the most dangerous type of manipulation

Another factor that contributes to uncertainty is the fact that rising social bot numbers increasingly cloud the topics and views of importance to voters. It remains unclear, for instance, whether people are generally expressing negative comments about refugees more frequently or whether this trend stems from a computer program. The US elections showed very clearly that the most dangerous type of manipulation deliberately sets out to create uncertainty. Hegelich clarifies: "What doesn't work is political conversion. I can't use social media to suddenly make a Democrat out of a Republican. But uncertainty works very well indeed. If I want to canvass for Trump, I don't need to promote him overtly. If I can manage to spread the impression that they're all lying, for instance, that's something that helps Trump more than Clinton. In Germany, that would help the right-wing AfD more than the mainstream SPD and CDU parties."

Is the clock ticking for the Internet as a human communication platform?

Taking social bots as an example, we see how digitalization invalidates the age-old assumption that ultimately, quantity is an indication of quality. This is now no longer the case, since even a message shared millions of times can be downright false. Growing lack of trust among users could be dangerous for a network like Twitter, since they would then move away from the platform in the end. As it stands, Twitter is currently reluctant to take serious action against bots for commercial reasons. In the main, however, Hegelich takes a fairly relaxed or even optimistic view of our digital future. As he sees it, everything is in flux and new technical solutions will continue to emerge – new networks and new rules leading to new ▷

Who is behind the bots?

Experts believe that social bots are now part of every political debate. The majority of bots deployed can currently prove detrimental in two ways: First, they are scalable – if you can run one, you can run a million, so the sheer volume can skew trends and divert the attention of large numbers of users to a given topic. Second, bots can contribute to polarizing opposing camps with hate messages.

With few exceptions, the originators of social bots have not yet been identified. The spectrum ranges from dubious PR agencies right through to organized cybercriminals. Many of those involved come from Eastern Europe, though are not necessarily – as often supposed – linked to the Russian government. Cybercrime is widespread in Eastern Europe, where there are many well-trained people who know they can make money in this way and that it is extremely difficult to bring legal action against them from abroad. Russia, for instance, would not extradite anyone to the US. However, the first large-scale programs for social network manipulation were developed by US intelligence services.

Graphics: ediundsepp (Source: Hegelich)

How to build a little spambot: Simple spambots can be programmed quite easily. Hegelich presented this very short script for a Twitter spambot (written in the open source programming language R) in his blog.

```
library(twitteR)

ckey <- “1234mykey”
csecret <- “1234mysecret”
atoken <- “45678mytoken”
asecret <- “56789othersecret”

setup_twitter_oauth(ckey, csecret, atoken, asecret)

LISTTopic <- twListToDF(searchTwitter(‘#BigData’, n=10))
View(LISTTopic)
LISTNames <- unique(LISTTopic$screenName)
text.examples <- c(“I am a bot, but I appreciate your work!”,
                   “Data is the new bacon!”,
                   “There are only 10 kinds of people: Those understanding binary code and others.”,
                   “Data is like people - interrogate it hard enough and it will tell you what you want to hear.”,
                   “Data that is loved tends to survive.”)
for(i in 1:length(LISTNames)){
  message.text <- paste0(“Hi @”,LISTNames[i], “ “,
                         text.examples[sample(length(text.examples),1)])
 print(nchar(message.text))
   try({
      updateStatus(message.text)
   }) -> temp
   if(class(temp)==”try-error”){
     print(‘Error!’)
     Sys.sleep(runif(1,50,100))
  }  else{
     print(paste0(‘i= ‘,i,’ (‘,LISTNames[i],’) is DONE!’))
     Sys.sleep(runif(1,10,22))
 }
}
```

Load library for Twitter communication

Get authentification for Twitter application programme interface (API)

Search for 10 latest tweets about big data

Use this list of text samples about big data

Answer each of the 10 tweets by writing: Hi, “your name”, the user’s name, “one of the text samples

Go to sleep on a regular basis to keep twitter from identifying you as a bot

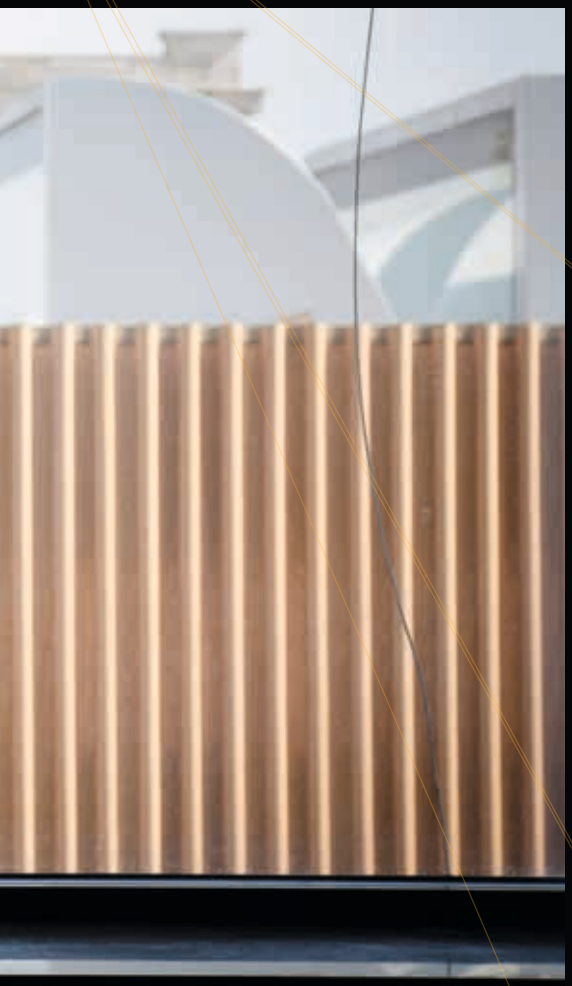
user behaviors. For him, the more urgent question is whether people get a handle on the current reality fast enough in relation to Germany's 2017 elections: "I am actually quite concerned about that. At the moment, we are dealing with social media structures that were absolutely not set up for political opinion-forming, but purely for business reasons. Facebook was intended to be a virtual friendship network – a feel-good environment for private users – and not an information medium. Large-scale, manipulative use of this kind of network by government organizations, for instance, was not part of the plan."

Responsibility of political parties

Since political opinion-forming is increasingly taking place online, Hegelich believes political parties have a responsibility to actively engage in these debates. However, he points out that, "Political campaigning on social medial is a gray area ethically speaking. There's a general lack of experience as to what conditions should apply here. Social bots concealing their presence are obviously not acceptable. But what about chatbots, which hold automated discussions with users about the party manifesto? And what about personalized campaign ads? These issues certainly call for a particularly transparent approach."

In Germany, politics is still blind – half a year before the elections, it remains oblivious to what is actually happening on social media. Hegelich thus advocates setting up a monitoring system. Especially during an election campaign, political players ought to know what is circulating publicly on social networks – especially when it comes to fake news. This is essential to have any chance of responding. A monitoring system could ensure early detection of mass-messaging with dubious content and inform the politicians and parties affected. "That, too, relies on machines," concludes Hegelich. "Monitoring of this type requires extensive automation – no-one can read the whole of Facebook and analyze the content from a political perspective."

Karsten Werth



“I actually think bots are a great technology that can be put to very good use. I could even envisage them being a valuable addition to a political campaign. But the moment a computer program tricks you into thinking it’s a real person, it then becomes highly problematic.”

Simon Hegelich

Prof. Simon Hegelich

A pioneer in political data science

Simon Hegelich has been Professor of Political Data Science at the Bavarian School of Public Policy since 2016. This professorship is the first of its kind in Germany. In his research, Hegelich blends political and computer science to pursue political data science. This entails both investigating the political relevance of issues surrounding digitalization and applying methods such as machine learning, data mining, computer vision and simulation to conventional aspects of political science. Hegelich studied political science at the University of Münster, completing both his doctoral and postdoctoral theses there. From 2011 through 2016, he was Director of the FoKoS interdisciplinary research center at the University of Siegen. Hegelich has been nominated for the 2017 German Research Foundation (DFG) Communicator Award, which recognizes excellence in communicating research findings to the public.

Picture credit: ediundsepp