

THE STATE OF CYBER HATE

Tamás Berecz and Charlotte Devinat

Bringing The Online In Line With
Human Rights

INACH

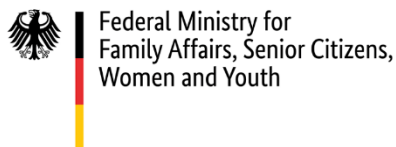
bringing the online in line with human rights



Supported by the Rights, Equality and Citizenship (REC) Programme of the European Union

Funded by

as part of the federal programme



Demokratie leben!

Legal Disclaimer

This publication has been produced with the financial support of the Rights, Equality and Citizenship (REC) Programme of the European Union. The contents of this publication are the sole responsibility of the International Network Against Cyber Hate and can in no way be taken to reflect the views of the European Commission.

Table of Contents

I. Introduction	3
II. Looking at the environment	4
A) Political Environment	4
The Dutch Elections and the Country’s Attitude Towards Hate Speech	4
The French Elections and the Country’s Attitude Towards Hate Speech	5
The German Elections and the Country’s Attitude Towards Hate Speech	6
The Austrian Elections and the Country’s Attitude Towards Hate Speech	7
The Hungarian Elections and the Country’s Attitude Towards Hate Speech	7
B) Policy and legal environment.....	9
The Code of Conduct and the monitoring exercises	9
The Communication by the EC.....	9
Summary.....	9
Social media.....	10
III. Looking at hate	11
A) Hate types	11
B) Ratio of complaints per online platform	12
C) Removal rates on all major platforms	15
D) Trends in removal rates	16
IV. Looking into the future	18
A) Plans of INACH for the next four years	18
B) Hopes for the future	19

I. Introduction

Our world is going through constant transformation, innovation and development, for better and, sadly, also for worse. Mentalities and values are continually being challenged, at times paving the path for positive shifts illustrated by people all around the world demonstrating on the streets for equality and risking or even losing their lives by doing so, or by campaigns, online or offline, joined by many to fight against discrimination. However, at the same time, major steps backwards are taken countering all the positivity. These steps backwards are illustrated by the rise of terrorism and extreme right in the West, the refugee crisis, the rebirth of antisemitism, coming from multiple sources, and of course the rise of internet hate, to name a few. The fight for human rights is simply in constant jeopardy.

This contradiction in our society becomes the most apparent when one spends some time analysing the online world. The internet gives numerous clues of where our world is heading by displaying the way in which our society is acting and reacting. This is the case as the Internet is such an intrinsic part of our everyday life, touching all of us and shifting our actions and thoughts. One aspect that is especially interesting and frightening at the same time is that behind the shield of computer screens people feel as though they are free to say anything they want, regardless if it is promoting discrimination or not. Indeed, many people who promote hate online might not do so without the protection of those screens, which is the cause of the creation of hate that, without internet, might not have had such an impact. Vice versa, by taking a look at the political and legal offline environment a deeper understanding can be reached of why society operates in such a way. It is undeniably obvious that the offline and online world go hand in hand, influencing one another constantly. Therefore, to understand one of them, one must look at the other as well.

Therefore, in this paper we will firstly look at the offline environments that shape cyber hate, like mentioned above, mainly politically and legally, in order to give an overview of the matter. This will then be followed by a summary of our findings from the previous year to give an extensive overview of the issue buttressed by in-depth data, done by reviewing the most prevalent hate types, platforms, trends in removals and such. Finally, we will go through our hopes for the future, of how we can better our fight against cyber hate and, quoting our motto, bring the online in line with human rights.

II. Looking at the environment

A) Political Environment

Hate speech and the fight against it do not exist in a vacuum, outside of everything. Every country's stance on hate speech and their efforts to curb it fundamentally depends on the ideological stance of their governments. There are several issues that steer a government's approach when it comes to policing hate speech. Very right-wing governments might decide to turn a blind eye towards it to appease their more extremist voters. More liberal governments might be hesitant to police speech altogether due to their stance on freedom of speech, fearing censorship above everything else. That is why there are so many approaches to hate speech in the European Union, even though most countries agree that hate speech is detrimental to minority communities and it is very corrosive to the institutions of democracy.

Several general elections took place in 2017 and early 2018. New governments have been formed in the Netherlands, France, Germany, Austria and Hungary. These elections will have had major impact on how these countries approach the policing hate speech both online and offline. Therefore, a short overview will be given in this chapter of the outcomes of these elections and the stances of these countries when it comes to tackling hate speech.

The Dutch Elections and the Country's Attitude Towards Hate Speech

The Dutch general elections took place in March 2017. They did not bring earth shattering changes to the government that run the country and yet they also changed everything. Mark Rutte's incumbent People's Party for Freedom and Democracy (VVD) won the elections and stayed on as Prime Minister. However, the Dutch left-wing, especially the Labour Party (PvdA) collapsed completely losing twenty-nine seats, whilst Geert Wilders' far-right Party for Freedom (PVV) gained five seats in Parliament compared to 2012 and came second.

The elections were followed by the longest government formation and coalition talks in Dutch history that resulted in a four-party-coalition of the VVD, the Democrats 66, the Christian Democratic Appeal and the Christian Union. All centrist parties mainly on the right side. Yet, Wilders' success looms strongly over the Dutch political sphere. His achievement pushes the whole spectrum towards the right and forces the ruling parties to implement policies that come from the far-right, even though in softer ways than the PVV would.¹

Thus, the Dutch government has been very ambivalent when it comes to policing hate speech online in the past two years. Although this ambivalence is coming from a liberal stance on freedom of speech, it is becoming harder and harder to draw a sharp line between liberal Dutch parties being faithful to their ideologies or pandering to the voters of radical parties. And, as far as the outcome goes, it almost does not

¹ https://en.wikipedia.org/wiki/Dutch_general_election,_2017

matter. In the end, the Dutch stance on policing hateful content online has become softer and softer as far as removal goes. The government and the Ministry of Justice is by far more concerned with shouting censorship and worrying about so-called over-removal of content than hindering the deluge of racist, misogynistic, antisemitic, homophobic, etc. hate. They did not criticise the Code of Conduct, but they were quite critical with the Commission's communication on tackling illegal content online.² Furthermore, they were highly critical of the new German law that put forth fines for social media companies that do not remove illegal content promptly.

Of course, it is not only the Dutch government and MEPs that were critical with these legal documents. However, the changes in the stance of the Dutch is definitely noteworthy. They never supported the harshest steps against cyber hate, but they have been moving further to the softer side in the past years. It is a question whether this is due to liberalism or due to a complete shift of the Dutch political spectrum towards the right.

The French Elections and the Country's Attitude Towards Hate Speech

The French presidential elections were held in April and May 2017. Emmanuel Macron won in the second round against Marine Le Pen by a landslide thanks to panic unifying the non-far-right voters once again, just like in 2002. This made Macron the youngest person ever to hold the presidency.³ After his success in the presidential elections, Macron's brand-new party (En Marche!) – established in 2016 – won the legislative elections too with a gigantic margin. His success underpinned the complete collapse of the traditional French left that only gained 45 seats in the National Assembly after losing 286 seats since 2012, and the dwindling on the traditional French right that also suffered major losses. Marine Le Pen's Front National (FN) only won 8 seats, a major rise since 2012, but still far away from what the party hoped or what the moderates feared.⁴

The failure of both Geert Wilders in the Netherlands and Marine Le Pen in France gave hope to centrists and leftists that Europe's not-so-slow slide towards the far-right can be stopped or at least slowed down. The fight against racism and especially online hate speech is very high on the new French government's agenda. It is clear that Macron stands besides Germany on this issue to a hundred per cent, moreover, he wants to even take the lead.

France has always had very strict laws that sanction hate speech. However, some of the vulnerable minorities had not been included and they had not addressed the online sphere as much as it would have been needed nowadays. Therefore, in August 2017, the government announced that they would include gender identity as a protected attribute to protect transgender people from offline and online hate speech.⁵ The French government has gone even further since then. In mid-March 2018, they announced that they were launching a plan, running from 2018 to 2020, that targets cyber hate and offline bullying related to racism and antisemitism. France also stated that they are working together with other EU members to find solution to these issues on a European level. This new initiative is aimed at social media companies, trying

² <https://marietjeschaake.eu/en/letter-to-the-commission-on-notice-and-action-procedures>

³ https://en.wikipedia.org/wiki/French_presidential_election,_2017

⁴ https://en.wikipedia.org/wiki/French_legislative_election,_2017

⁵ <https://www.lifesitenews.com/news/france-ramps-up-hate-speech-laws-for-transphobia-offenders>

to force them to detect, report and remove illegal content as rapidly as possible.^{6 7} Edouard Philippe, the Prime Minister, seemed especially incensed by the lack of adherence to existing French hate speech laws by the social media companies, criticising their stance and efforts harshly.⁸ These developments in the past months clearly show that France has the same stance on policing cyber hate as Germany. The government wants to solve the issue, or at least curb it, and it is not afraid to enact harsher laws to combat hateful online content. A very positive development indeed.

The German Elections and the Country's Attitude Towards Hate Speech

Just like the Dutch elections, the German federal elections did not bring major changes as far as the ruling parties go. Yet, its outcome was felt as a menacing tremor throughout Europe. Angela Merkel's Christian Democratic Union (CDU) won the popular vote together with the Christian Social Union (CSU), receiving 33 per cent of the votes. The Social Democratic Party (SPD) came second with 20 per cent. This outcome changed hardly anything since these parties had been governing the country for years and they received a well-cushioned majority again. However, Alternative for Germany (AfD), a far-right radical party came third (12.6 per cent), which made them the first far-right party to enter the Bundestag since the NSDAP. Furthermore, even though the CDU/CSU won, and the SPD came second, just like in the previous elections, both parties suffered major losses compared to the previous ballot, especially the SPD that achieved its worst result since the Second World War.

Evidently, a far-right party in the German parliament caused worry both nationally and internationally. These qualms were also not alleviated by the fact that the CDU/CSU and the SPD who had been governing the country in a grand coalition, did not want to go down the same road, citing irreconcilable differences. Therefore, Europe had a Germany with a large far-right faction in the Bundestag and practically no government.⁹ After the results, Merkel tried to bring together a coalition government with the Greens and the Free Democrats. The negotiations went on for months, but they ultimately failed when the Free Democrats pulled out for good, citing major differences in immigration and energy policies. That was when the SPD decided to come to the table and talks started between the parties in early 2018, which led to an agreement and another grand coalition that took office in March 2018, after six months of interregnum.¹⁰

When it comes to hate speech and the policing of cyber hate, the outcome of the elections was indeed worrisome at first glance. The major victory achieved by the AfD shocked everyone and the whole EU were worried that Germany would follow other European countries in sliding ever more rapidly towards the far-right. This would have meant a more lenient stance towards hate speech and maybe even backtracking of the achievements of the previous two years. Thankfully, the new grand coalition means that Germany kept its stance on cyber hate and the AfD has not been able to affect policy in this area as of yet. This is paramount since Germany has been the shining beacon in policing hate on the internet in the EU. The German government's influence was behind all EU policy changes in this field (Code of Conduct,

⁶ <https://aa.com.tr/en/europe/france-reveals-new-plan-to-counter-online-hate-speech/1093371>

⁷ <http://www.jpost.com/International/French-PM-France-to-get-tougher-on-social-media-hate-speech-545476>

⁸ <https://www.reuters.com/article/france-racism-socialmedia/france-to-get-tougher-on-social-media-hate-speech-pm-idUSL8N1R14G0>

⁹ https://en.wikipedia.org/wiki/German_federal_election,_2017

¹⁰ https://en.wikipedia.org/wiki/German_federal_election,_2017#Government_formation

Communication on Tackling Illegal Content Online), and they were the first (and so far, the last) to enact a law which sanctions social media companies with major fines if they do not remove illegal content from their platforms promptly. The law has received wide criticism from other countries in the EU, fearing over-removal, censorship and blanket removals by the companies. Yet, the German government has stuck by it and it will be very interesting to see it in action in the future. We will be able to gauge its effects in the next two years may they be positive or negative.

The German law might have received stern criticism, but it cannot be denied that Germany has been acting as the flagship for a strong and unwavering stance on policing hate speech online. Hopefully this will not change in the future as the EU needs a strong and unanimous will to curb the mainstreaming of cyberhate.

The Austrian Elections and the Country's Attitude Towards Hate Speech

The Austrian legislative elections brought minor changes as far as the winning parties go. The Social Democratic Party (SPÖ) lost its first place by gaining the same result as in the previous elections, because the Austrian People's Party (ÖVP) gained 15 more seats than in 2013 and the Freedom Party of Austria (FPÖ) – a far-right formation – gained 11 more seats. Since the SPÖ and the ÖVP had been governing in a grand coalition since 2013 and they did not want to continue on that road, the leader of the ÖVP, Sebastian Kurz, started coalition talks with the FPÖ and after a short negotiating period, the new government was announced in December 2017. This outcome means that the FPÖ became part of the Austrian Government for the second time after 1999.¹¹ It says a lot about the changes in the political sphere of the EU that, besides some minor whispers about worries, nobody really protested against a far-right party being in the government of the country; while in 1999 the EU put sanctions on Austria for the same.¹²

Even though Austria is not as active on the international podium as Germany when it comes to policing hate speech, the country has quite strict hate speech laws and its courts uphold them very thoroughly. This is very important, because hate speech has been at an all-time high in the country for the past years.^{13 14} That it is why it is also noteworthy that the Appeal Court in Vienna ruled in 2017 that Facebook must remove hate speech and it is not enough to simply block the content for Austrian IP addresses, a landmark decision that should lead all other European courts.¹⁵ We can simply hope that with the FPÖ in the government, the courts, the legislative and NGOs that combat cyber hate will not be pressured or politically attacked to soften their stance on hate speech.

The Hungarian Elections and the Country's Attitude Towards Hate Speech

The Hungarian general elections were won by Viktor Orbán's Fidesz by a landslide, giving him and his party a 2/3 majority in the Parliament. Just like in 2014, Jobbik came second, followed by the smaller centrist and leftist parties whose erosion continued after their total collapse in 2010.¹⁶

¹¹ https://en.wikipedia.org/wiki/Austrian_legislative_election,_2017

¹² http://www.demokratiezentrum.org/fileadmin/media/pdf/falkner_sanctions.pdf

¹³ <https://www.timesofisrael.com/anti-semitic-hate-speech-incidents-in-austria-at-all-time-high-watchdog-says/>

¹⁴ <https://www.thelocal.at/20161125/hate-speech-cases-soar-in-austria>

¹⁵ <https://www.npr.org/sections/thetwo-way/2017/05/08/527398995/austrian-court-rules-facebook-must-delete-hate-speech>

¹⁶ https://en.wikipedia.org/wiki/Hungarian_parliamentary_election,_2018

Orbán turned Fidesz into a far-right party in the past eight years and the country into a Potemkin democracy where all the institutions are there, but they do not function or serve his parties interests. Furthermore, since the start of the European refugee crisis in 2015, the Hungarian government launched a hate campaign against “migrants” that has been unheard of since the 1930s and 40s. Orbán knew that he needed an issue that could plant fear deep in the heart of his voters, otherwise Fidesz might have lost. His government was rocked by major corruption scandals, the country’s health, education and pension systems are in shambles, and they completely ignore the more than three million Hungarians that live in abject poverty, while favouring the upper middle class with tax cuts and other benefits. Hence, the refugee crisis was a godsend to Orbán. He could set the agenda completely and managed to eschew all uncomfortable topics during the campaign. All Fidesz and the government did was to spew hateful propaganda against refugees and George Soros, who, according to them, had a plan to settle hundreds of thousands of “migrants” in Hungary with the help of the opposition parties and NGOs that are all “Sorosists” and are all on Soros’ payroll. This tactic worked wonders for Fidesz. The fact that the party ran a permanent campaign using taxpayer’s money and practically controls most of the online and printed press in the country and especially in the countryside also helped.

Fidesz managed to get people to the ballot boxes who had never voted before, the party’s success in the poorest most uneducated pockets of the electorate cannot be denied. These people were fed government propaganda for years and the opposition was incapable of finding the cracks and getting through with a different message. Hence, most countryside electoral districts went to Fidesz with mindboggling margins. People were “scared to death from the evil Muslim brown people who wanted to take their jobs, their money, their culture, dilute their race and rape their women”.¹⁷ And since Fidesz said that they were the only ones who wanted to protect them from this evil that Fidesz itself created, these people voted in droves for Orbán and his party.^{18 19}

Fidesz is too preoccupied with building an autocracy using hate speech to stay in power than to actually see cyber hate as an issue. The country does not have a real stance on online hate speech, and even though it does have anti hate speech laws, they are very rarely enforced and often used against oppressed minorities like Roma people. The government is often openly homophobic²⁰, it has been refusing to ratify the Istanbul Convention²¹ and the “Stop Soros” campaign has very clear antisemitic undertones, even though Fidesz would deny that.

All in all, the Hungarian government acts very similarly towards hate speech as towards democracy, freedom of the press, the separation of powers and equality. They communicate very nice things about them, especially towards the EU and other powers outside of the country, but they erode and destroy them internally depending on their political needs and wants. Hence, it is highly unlikely that Hungary will be stepping up its efforts against hate speech or will be joining France and Germany to lead this fight. The

¹⁷ <http://www.dw.com/en/hungarys-orban-tells-germany-you-wanted-the-migrants-we-didnt/a-42065012>

¹⁸ <https://www.independent.co.uk/news/world/europe/refugees-muslim-invaders-hungary-viktor-orban-racism-islamophobia-eu-a8149251.html>

¹⁹ <https://budapestbeacon.com/viktor-orban-only-fidesz-can-protect-hungary-from-brussels-the-un-and-george-soros/>

²⁰ <https://www.independent.co.uk/news/world/europe/hungary-president-viktor-orban-anti-lgbt-hate-group-iof-homophobia-gay-europe-a7758166.html>

²¹ <https://dailynewshungary.com/ep-delegation-urges-hungary-ratify-istanbul-convention/>

government will do the minimum that they need to do to appease Germany, the country's biggest trade partner, and not to anger the EU as much as to lose EU funds. Otherwise, Orbán and his band of oligarchs will shy away from nothing to stay in power, may that be another taxpayer funded hate campaign, the total dismantling of the already tiny and weak Hungarian NGO sphere, or the appeasement of people's most base instincts.

B) Policy and legal environment

The Code of Conduct and the monitoring exercises

Regarding the European legal environment, there are two main advances that come to mind, both being discussed in previous INACH papers, namely the [Annual Report](#), [Strategic paper](#) and [Legislation paper](#). Firstly, there was the development of the Code of Conduct on Countering Illegal Hate Speech Online (2016) (CoC)²². This code embodies the main outcome of the EU Internet Forum. To sum it up, it simply allows, on the one hand, the public, organizations and anyone who fights against hate speech to inform the IT companies about the content in question, and on the other hand demands a stronger reaction from those companies, i.e. to react to the complaints within 24 hours and remove the content if necessary in a timely manner. The Social Media companies signed the CoC, promising to adhere to it and the Commission has been running so-called monitoring exercises with the help of NGOs to see whether the companies are actually adhering to the Code.

The Communication by the EC

The other notable advancement was the [Communication from the Commission](#) to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on “tackling illegal content online” and moving “towards an enhanced responsibility of online platforms” mentioned earlier on. INACH discussed this Communication in our [Strategic paper](#). This policy document basically illustrates the Commission's actions in taking the necessary steps towards a safer internet, whilst at the same time sidestepping the criticisms that the new German law received internationally.

Summary

We will not discuss these two examples in much depth in this paper as we have done so in other INACH paper as cited above, but, when evaluating the state of the legal environment, they each represent immense steps forward in the fight against cyber hate. Their existence shows that the matter is on the top of the EU's legal agenda, which gives hope for the future, notably the creation of ever more laws concerning cyber hate, that are sadly lacking to this day. Harmonization in the law, as well as on all other fronts is essential if we deem to have a chance at changing things for the better. Nonetheless, the most important aspect is what has yet to come, when looking at how effective these policies will become, and how much of a positive impact they will have. The coming years will be paramount, and hopefully bearers of great change.

²² European Commission and IT Companies announce Code of Conduct on illegal online hate speech, Brussels, 31 May 2016: http://europa.eu/rapid/press-release_IP-16-1937_en.htm

Social media

Now it is time to take a look at the steps forward that should be in focus by social media companies, namely Facebook, Twitter and YouTube. This is important to understand as the use of these companies' tools of reporting and strategies for tackling hate speech are paramount in the fight against cyber hate. Some information in this chapter is taken from INACH's paper titled "[How to improve reporting instances related to cyber hate phenomena](#)" that can be found on INACH's website. The paper makes many recommendations, so we will not get all of them in detail again, will only look at the main lines.

The first keyword that we will discuss is transparency. It's clear, those companies should always provide sufficient information on how they takedown cyber hate and on what basis. The companies do have guidelines regarding why they take down specific content, but that is not enough as they do not share how much content is actually taken down. This hinders the work of NGOs such as INACH and should be simply remedied. Indeed, we denoted that problem when in direct contact with Facebook during one of their workshop in Dublin a year ago, during which we asked them about their guidelines and on which criteria do they remove content, and they answered those questions; but when we went deeper and asked for examples, data, numbers concerning their claim, to prove what they were stipulating, they denied us the right to access any of that. During a second workshop in Dublin, in 2017, the Companies were much more forthcoming. They provided examples of what they removed and explained why they left some questionable content online. This was a major step into the right direction, yet NGOs still run into situation where their experts do not really understand why certain hateful content is not removed by Social Media companies. Thus, being even more transparent and open regarding the matter would help us and them as well to make a better job at a consistent removal rate. Moreover, in relation to the above, information is lacking concerning the number, credentials, and exact role of people working with the reporting systems. A solution to that could be that trusted flaggers should be made aware of who they are dealing with. Same goes for the way in which those people work and especially in regards to their community standards guidelines used for blocking or removal techniques. Lastly, the developments of software used to target hate speech is unknown to us. More knowledge of that would help organizations such as ours as well as any trusted flagger as it would enhance cooperation.

The other keyword, also noted in the aforementioned INACH paper, is simplification. What we noticed was that the reporting system at times was too complicated. Simplifying it would help trusted flaggers as well as social media users. The complications that surface when making a report (such as the long list of options, not making it clear which one fits best, and all of it taking very long) may discourage some to go forth with it. Examples of those complications were described in detail in the INACH paper but can be summarized as follows; a direct access for reporting a comment should be developed as well as the possibility to report multiple comments in one report, and finally users should be answered to on every report made. This would encourage users and trusted flaggers in many ways and would have a definite impact on the issue of cyber hate.

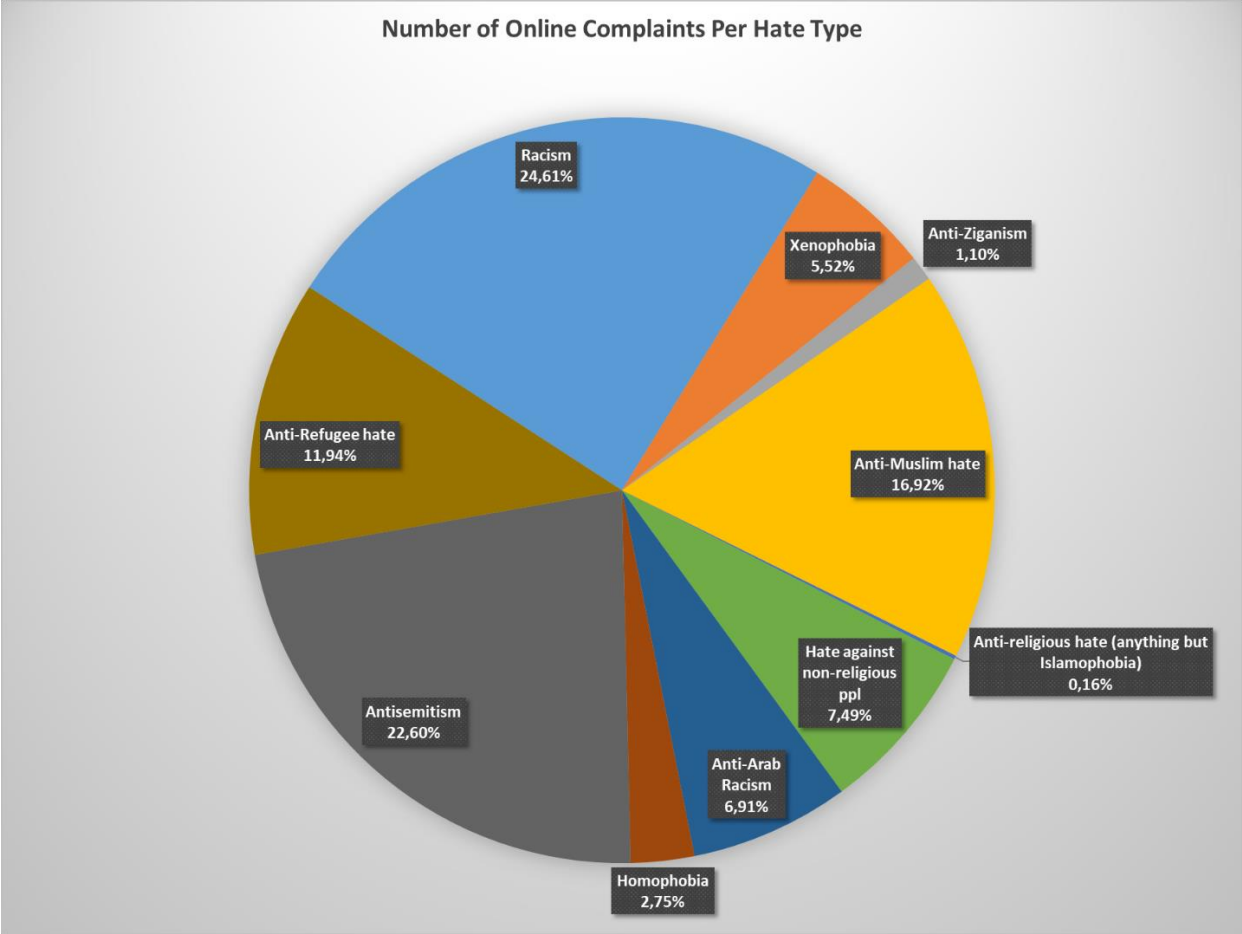
III. Looking at hate

A) Hate types

Based on our data collection, four hate types can be seen as predominantly prevalent in Europe, especially in Western Europe. These hate types were dominant all throughout the past year, and even though they might have changed places from one month or quarter to another, their place in the top four was never really in question. These hate types are the following: Racism (24.61%), antisemitism (22.60%), anti-Muslim hate (i.e. Islamophobia) (16.92%) and anti-refugee hate (11.19%). Some of these hate types are showing downward trends, but they are and have been far above all other monitored hate types throughout the year. We will call them the top four hate types.

Following the top four hate types, we have the bottom four hate types. These hate types are far below the top ones that make up more than 75 per cent of the data collected by INACH. None of the bottom four hate types reach 10 per cent and only two of them go above 5 per cent. These hate types are the following: anti-Arab racism (6.92%), xenophobia (5.52%), homophobia (2.75%) and anti-Roma hate (i.e. anti-Ziganism) (1.10%).

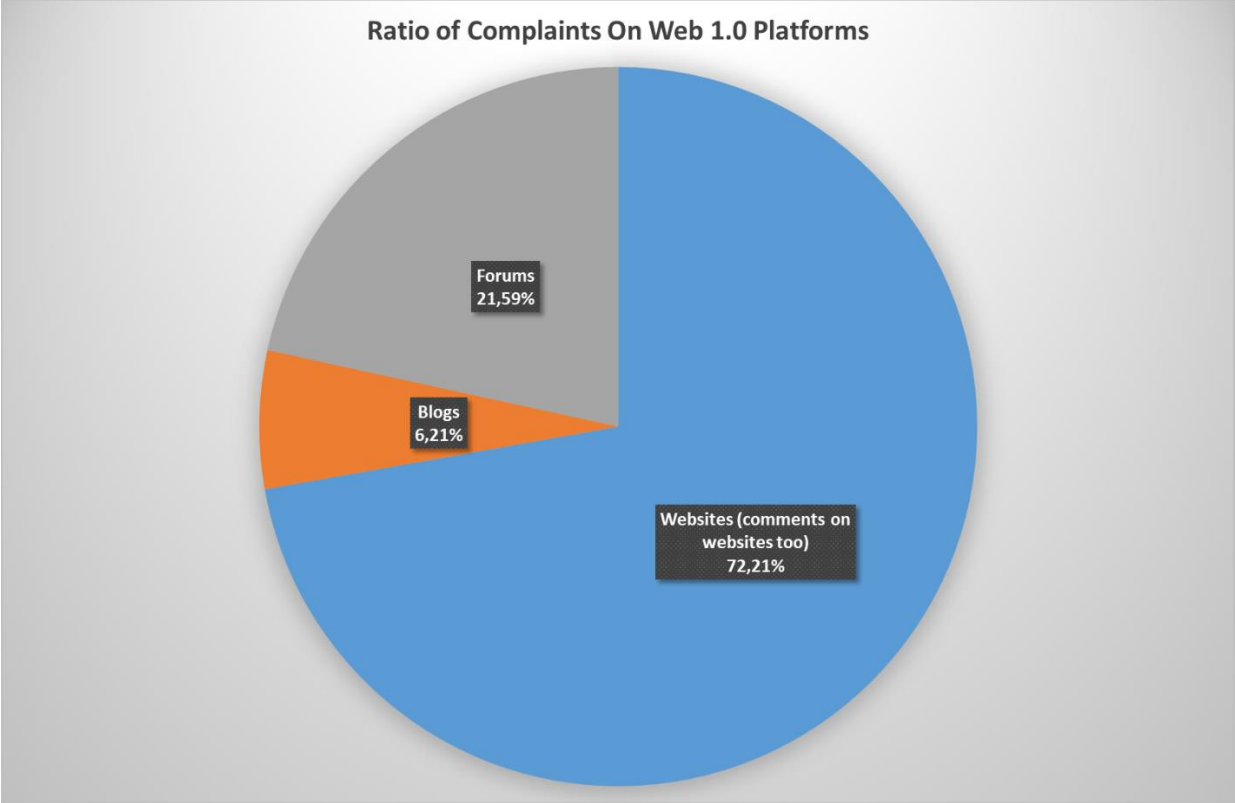
As one can see there are two other categories of hate left (hate against religious people [anything but Islamophobia] and hate against non-religious people). These categories were included per the request of jugendschutz.net, because they collect a lot of cases that fall into these categories. However, due to this peculiarity, more than 98 per cent of the data on these hate types came from Germany. Thus, it was decided to leave these categories out, because they do not represent these phenomena on a European level.



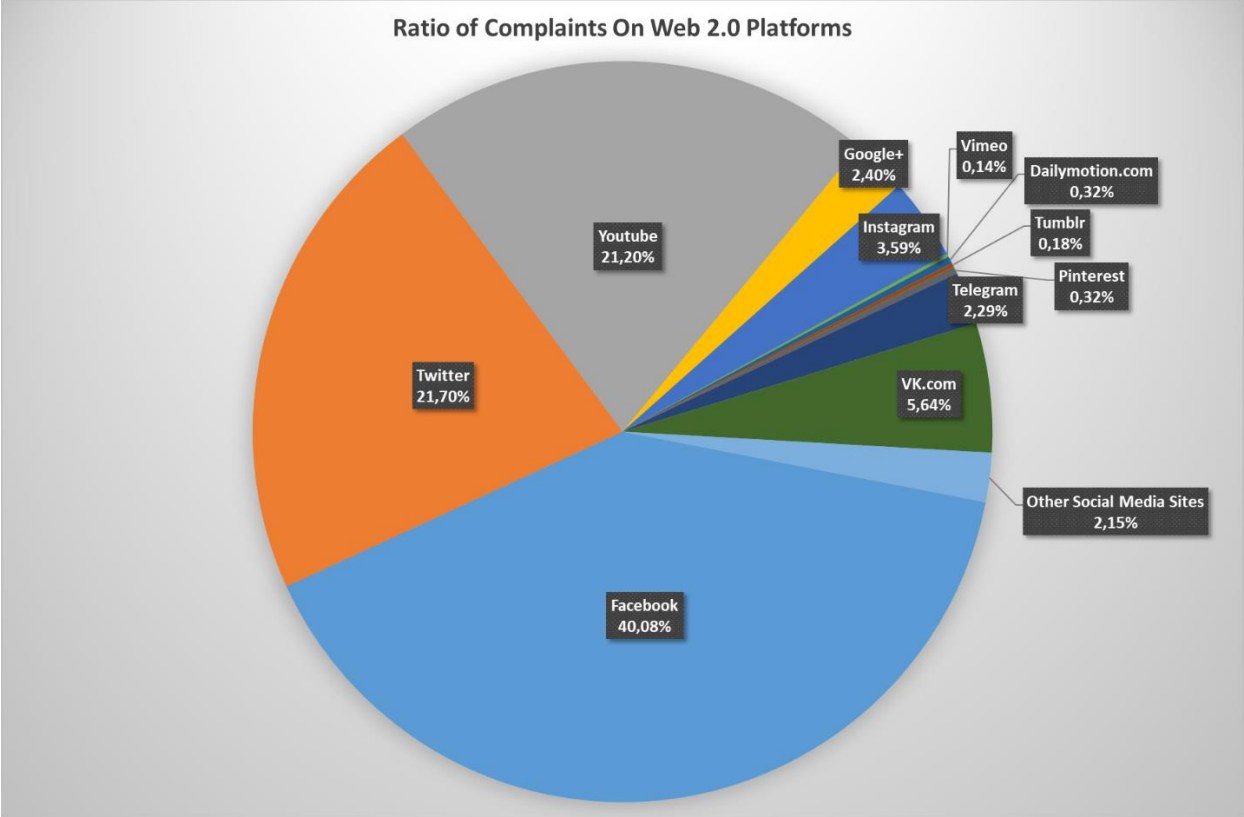
That being said, it is perfectly clear that general racism and antisemitism were the most neuralgic issues in the past year based on INACH’s data. Closely followed by Islamophobia and anti-Refugee hate. Anti-refugee hate is a special category that was created out of necessity, even though - as a hate type category - it had been virtually unseen before the so-called refugee crisis that started in 2015. Since then, however, it has clearly been a major issue, although a diminishing one.

B) Ratio of complaints per online platform

When it comes to platforms that online hate speech spreads on the most, social media is unbeatable. However, Web 1.0 platforms are still in the game. If we take out Web 2.0 platforms from the data pool, one can see that websites are a magnitude above forums and blogs. Almost three quarters of all complaints registered on Web 1.0 platforms by our partners were registered on websites. They are followed by forums (21.59%) and blogs (6.21%).

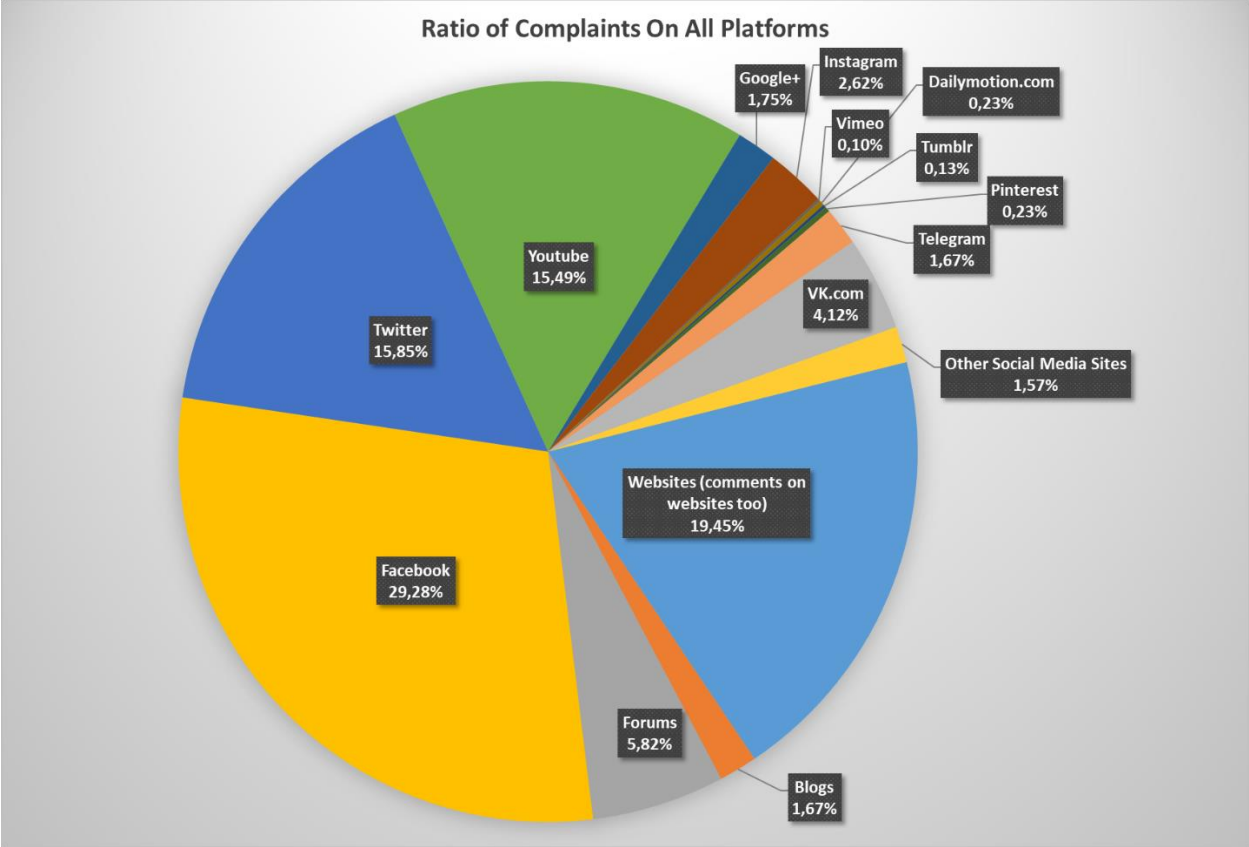


As far as social media platforms go, there is a clear triumvirate that rules the whole market, and therefore gives the biggest surface to cyber hate and extremist propaganda. The three main platforms are Facebook (40.08%), Twitter (21.70%) and YouTube (21.20%).



Almost 83 per cent of registered instances of cyber hate came from these platforms if Web 1.0 platforms are taken out of the data set. Facebook’s dominance is even more prevalent, since one can see that the ratio of instances of cyber hate registered on it is almost twice as big as the second platform, Twitter. The puissance of the triumvirate is underpinned by the fact that all other social media platforms are dwarfed by them quite literally when it comes to registered cases of cyber hate. The fourth largest number of cases were registered on VK.com, which is a meagre 5.64 per cent. Besides Telegram, Instagram and Google+ that are all above 2 per cent (mainly based on data coming from Germany and France), no other platform even reaches 1 per cent.

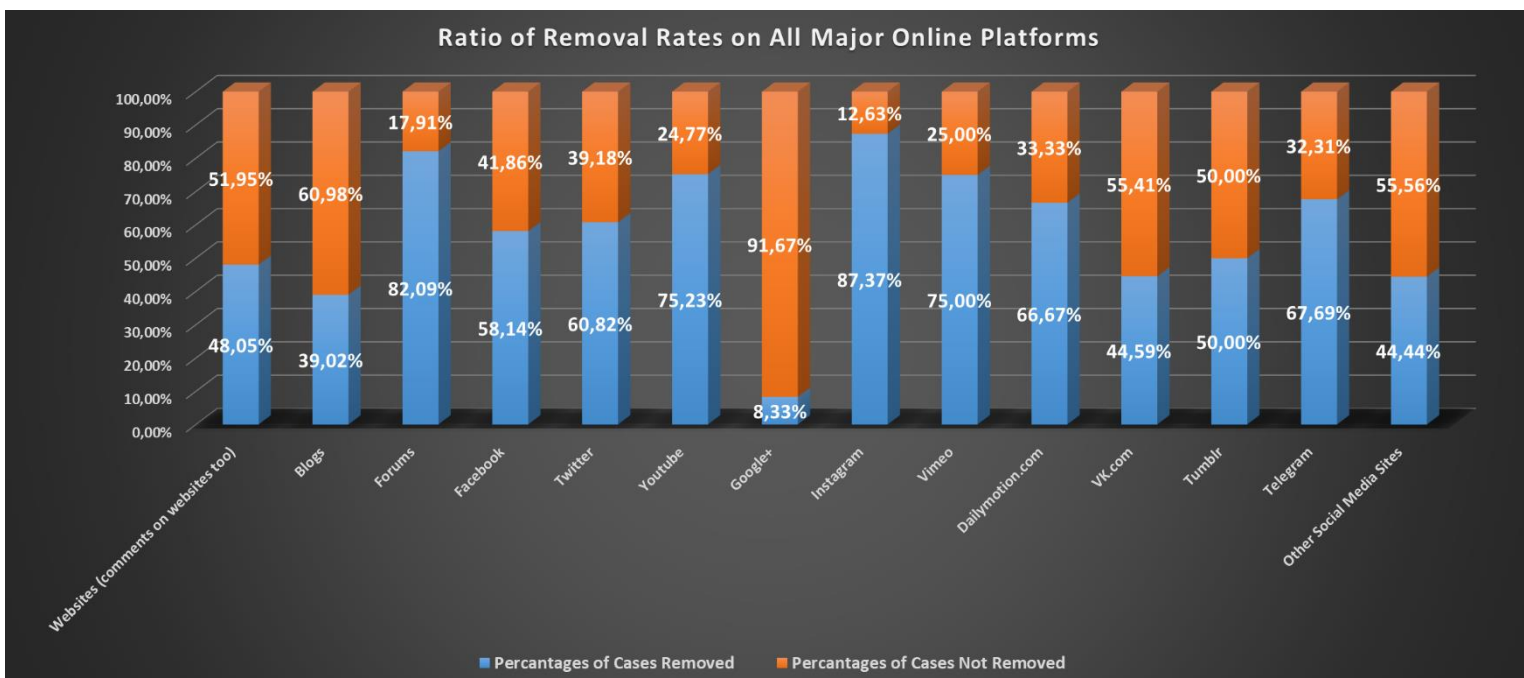
If the two data sets are combined, namely the ratio of cases registered on Web 1.0 and Web 2.0 platforms, it becomes perfectly apparent that social media dominates the online public sphere, and thus most instances of online hate speech are registered on these platforms.



Even with the data from Web 1.0 platforms added in, the three social media giants (Facebook, Twitter and YouTube) are responsible for more than 60 per cent of registered cases, insofar as they provided the online space for more than 60 per cent of online hate registered by INACH and its partners. This unparalleled dominance should also come with the highest level of social responsibility that can be possibly taken by multinational tech companies.

C) Removal rates on all major platforms

As far as classical online platforms go, getting hateful content removed from them is extremely hard. This definitely shows in the recorded removal rates. Websites removed less than half of the recorded questionable content, blogs removed even less (39,02%). However, forums stand out from the old guard with an 82,09 per cent removal rate. The reasons behind the abysmal removal rates on websites and blogs are probably twofold. Minor Web 1.0 platforms are not as prepared or well-funded enough to maintain an army of admins and moderators as social media companies. The second reason, which is also a major issue, is the fact that some of these platforms are specifically brought to life and maintained to give a surface for online hatred. Most of these are hosted on servers in the US and therefore it is almost impossible to get anything removed from them.



However, social media companies with all their money, data and manpower are also far away from perfect. INACH and its partners have to face massive issues due to vague policies and codes of conduct put forth by these companies. They also implement their own rules often highly arbitrarily. Moreover, their stance on different hate types or modes of online hate speech vastly differ from country to country, even though they are supposedly using the same rule book. That is why the numbers we see are not too far from being terrible. Facebook removed 58.14 per cent of cases, Twitter only 60.82 per cent and YouTube 75.23 per cent. These numbers are fairly low and show the great divide between NGOs that fight for a more inclusive online public sphere and social media companies that try to paint themselves as the knights in shining armour protecting free speech online. However, the fact is that these companies are money making machines first and foremost and, therefore, they resent the idea of spending more money to earn less money. And, essentially, that is what NGOs and some governments try to get these companies to do. Higher more people and devote more resources to remove content that - if left online - make them money.

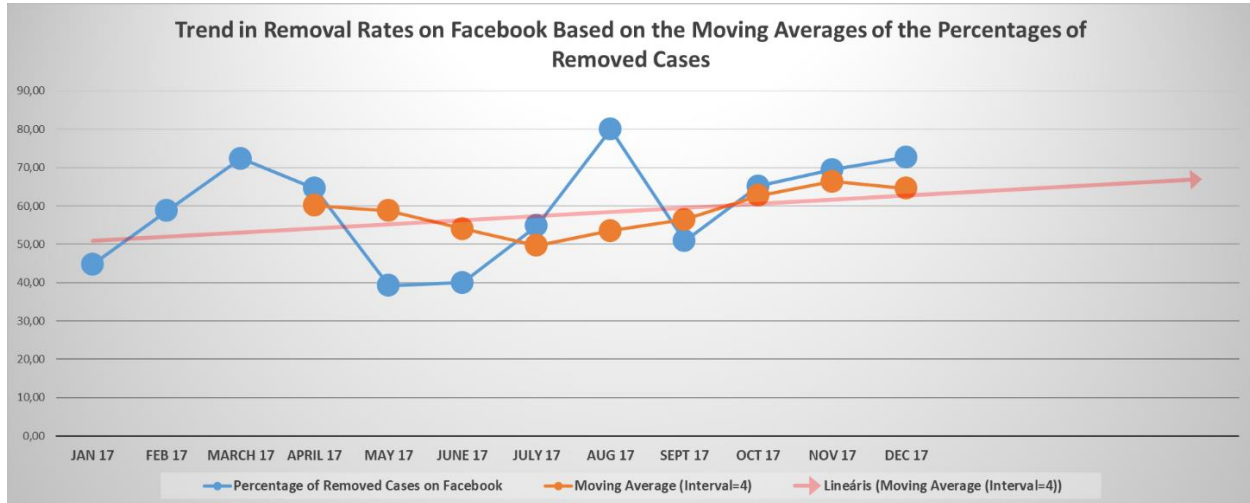
INACH, naturally, is not arguing that the removal rate should be a 100 per cent. But INACH has several member organisations that have dozens of experts working for them. These experts are well trained in the recognition of hate speech and international and national hate speech laws. Thus, when they approach these platforms to remove something they do that with the knowledge that the content is definitely hate speech AND most likely illegal. Still, there can be differences of opinion, but removal rates should most definitely reflect these facts and therefore they should be somewhere around 90 per cent (at least).

D) Trends in removal rates

Content removal is among the top goals of INACH and our partners. Cyber hate is corrosive, discriminative and more than capable of radicalising people. Hence, the removal of such content from social media sites is of paramount importance. However, since there is an obvious clash of human rights (between human

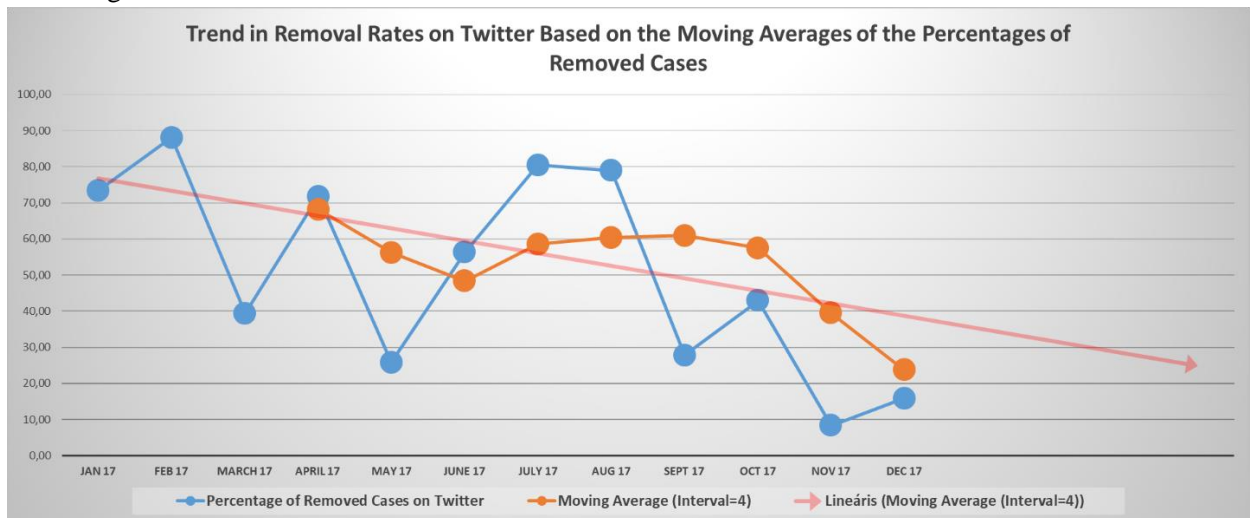
dignity, freedom from discrimination and freedom of speech) and a clash of interests between NGOs and social media companies. The removal rates on the major social media platforms are far from ideal.

Facebook's removal rates are mostly ok, but they still fluctuate immensely and a slight, but very steady downward trend can be observed on average. As one can see on the chart below

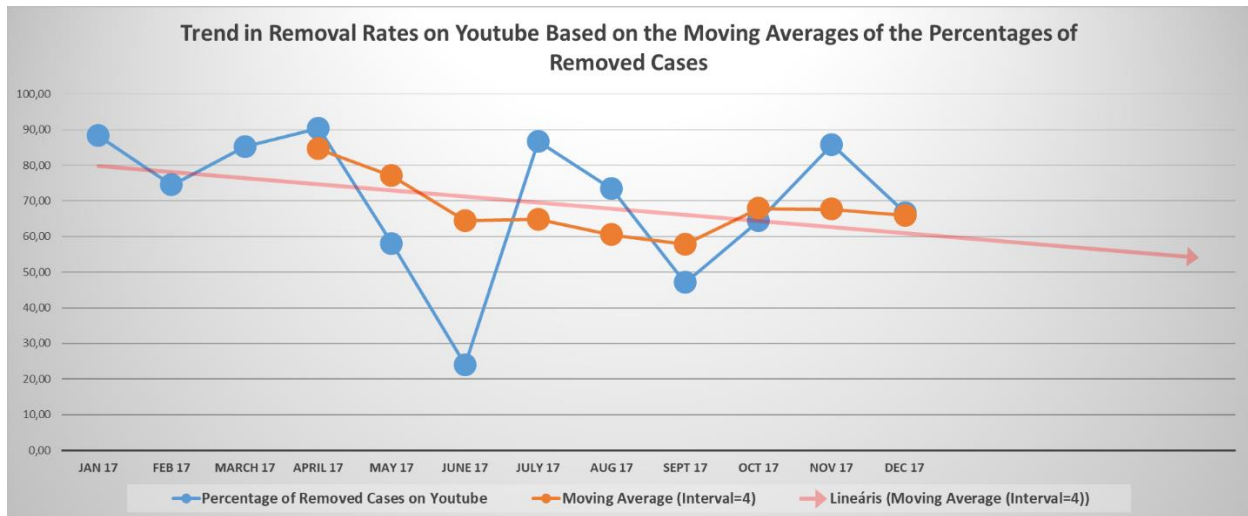


Facebook's highest removal rate was recorded in August 2017 at 80 per cent. This means that 80 per cent of complaints that our partners sent to the company were removed. However, the lowest removal rate was around 40 per cent and that was recorded in May 2017. With all this volatile fluctuation, Facebook is still the only major platform with a slight upward trend in removals. Hopefully this will not change in 2018.

Twitter is worse than Facebook in removals and they also present a steep downward trend. Even though their highest removal rate in February 2017 was at almost 90 per cent, their lowest one was at only 10 per cent in November 2017. A ratio much lower than the lowest of Facebook. Moreover, they remove less cases on average and their numbers fluctuate much more.



YouTube is far closer to Twitter than to Facebook when it comes to removal rates. Their numbers fluctuate in a very volatile manner, their lowest removal rates are between 20 and 45 per cent. The company was showing great promise in the beginning of 2017. Their removal rate was going up after a horrible last quarter in 2016, and they were the only one with an upward trend in the middle of 2017.



However, they went through a major dip in removals during the early summer of 2017 and then again in early autumn. Thus, all in all, after a very promising first four months, they managed to produce a downward trend in removals on average.

IV. Looking into the future

A) Plans of INACH for the next four years

INACH has big plans for the coming four years. These plans have been touched upon in previous publications, namely our Annual report and our Strategic paper, but will be summarized once more. One of our goals will be to monitor cyber hate, amongst other ways, through the monitoring exercises. We will collect and analyse data and present findings on a continuous basis. Monitoring activities will be further improved through the international online cyber hate database and complaints system. Relating to this will be the goal to develop this database that we created further. It will constantly be updated, debugged and evaluated when needed. This tool will bring new trends and problematic developments forward which can be a great asset for us in our argumentation with politicians on national and European levels, law enforcement agencies, educational institutions, civil society initiatives, support staff and policy makers of Social Media companies. Our aim is to get more organisations to use the database, firstly our member organisations, so that its potential could be achieved fully. Indeed, if this were to become the case, it would provide easy access to ample data and knowledge for those organisations, leading them, and us, to be better equipped in the fight against cyber hate. To achieve this goal, we will hold (and have already held) workshops and trainings to showcase how our database works, its advantages and how it can be used.

Linked to the aspect of workshops and trainings will be one of our other main goals; education. We will be looking at educating different target groups. For instance, we have already started working with the Dutch police on this matter, by looking at where the knowledge is lacking on their part, according to them, and by filling the gaps and creating a manual to enable them a better understanding of the issue of cyber hate and how to deal with it appropriately. That is one of the examples of how we are tailoring educational tools for

a specific target group by meeting their demands and needs. Other target groups include young adults, the corporate world and social media companies. Through education and knowledge aimed at every side of the “problem”, we can tackle the issue appropriately and on the one hand make (potential) victims know what their rights are and that they do not need to be powerless, and on the other hand we take a look at those who can help make a change and diminish the number of victims, and work hand in hand with the big companies whose actions have the biggest impact due to their reach. By doing this we will aim to improve removal rates from those online social media platforms by strengthening the cooperation within the network through launching joint (legal) actions and with IT companies and providing the necessary background information to the EC and other stakeholders. It’s simple, the more people understands what is hate speech, what its effects can be, and how to stop it from spreading, the more we have a chance at having a safer internet for everyone.

Moreover, as part of the educational aspect, other than the workshops and trainings, the development of counter-speech and counter-narrative modules will be pivotal. This will enable to counter online hate speech whilst armouring the public with enough knowledge to not become victims nor radicalised by internalizing racist and hateful views about members of minority communities. This can mainly be done through campaigns that we will spread online and offline. This has been proven in the past, counter-speech is greatly successful as it tackles the issue in another way that may suit people better. Instead of only looking at deleting hate speech, or perhaps having only purely negative reactions towards it, although legitimate, looking at it from the perspective of humour or with understanding as to why the person said such a thing, reaching out to those people, and, even better, addressing the fear and hate with kindness and love is a very valuable method.

Lastly, in order to reach our goal for education and to spread our message we will also work tremendously hard to persuade all of our members and national governments to develop such measures themselves as well. This ties up to our goal of broadening our network and obtaining more members, by of course still focusing on harmonizing the work of all our current members. This is also linked to the above point mentioned of developing the database and making sure progressively more organizations use it. It is undeniable, we are without a doubt, stronger together, and we can all learn from each other. As we all come from different countries, cultures and backgrounds, that is our strength. This will also become apparent during our annual conferences that will underpin our main goal: reaching unity against adversity.

B) Hopes for the future

As we can see throughout this paper many essential steps forward have been taken in the recent years, which gives hope for the future and shows that the fight against internet hate is on the top of the priority lists for those who can truly make a difference. Nevertheless, the issue will always be that, although the fight is on and battles are being won, the war is one that is never ending by nature. The internet is ever expanding, affecting the lives of more and more people every day, for better and for worse. The age groups that are affected are also becoming younger, namely children are being confronted with the internet at a younger age than they were before. Now, though the fight might seem without end, the right steps in the good direction can affect more and more people positively every day. If the mindset changes, then we have a chance. If people see internet hate as something they can oppose and not as something that they need to

learn to put up with, we have a chance. That is the first major hope for the future, the change in the mindset, which would give us a chance.

Moreover, more specific changes are also expected for the coming years, which would make that change in mindset easier to attain. We noticed that one of the main problems was that there is a major discrepancy between what material is concerned illegal or not. The root of the problem is clear, there is simply no universal definition of hate speech, every country has their own definition, and even those are not always very clear, and rarely mention anything about the online world. Our above-mentioned close relationship with the Dutch police for instance led us to understand that even though there are some laws that exist in the Netherlands regarding hate speech, the police is not really keeping up with them and especially not regarding the internet. When they are faced with a complaint for cyber hate they are in some cases not trained well enough to take appropriate measures. The manual we are developing might make matters easier, but this should be done on a broader scale. Law enforcement in each country should be better trained to deal with not only hate speech but specifically cyber hate. This of course goes hand in hand with the establishment of a universal definition of hate speech and more details regarding cyber hate especially. We therefore do hope that transnational institution and governments will move forward and continue in their right tracks in producing more treaties and legislation than they already have done (those existing ones can be found in our “Legislation related to cyber hate” on our website). This will certainly take time but starting on a smaller scale like we do with the police, by working with the material already available, would be a great step forward. Also, we do see that, as mentioned before, hate speech and especially cyber hate, is increasingly becoming a priority for all and definitely has an important spot on the political and legal agenda, leading us to be confident that, even though it might take a while, it is not unattainable.

Moreover, this issue can also be remedied by the platforms on which the hate festers. Indeed, if the mindset of the public were to be changed, the states were to make new laws, and law enforcers were to be trained, the last thing missing would be an effort to be made by those platforms. If those platforms were to stop with their inconsistencies, especially in their removal rates, amongst other things, and would stop interpreting their own rules subjectively that would help progress to be made immensely. Social media platforms need to have clear guidelines that make sense and they need to stand by them in a consistent manner. As well as the list of treaties and legislations that were mentioned above that were created by transnational institutions and governments showing progress is on the way, social media companies are also making progress, as we discussed in this paper, but we hope that this is only the beginning.

The hopes for the future are then threefold. There are three areas that need improvement, and they should each be improved individually and in respect to each other. They all work hand in hand, and will not be successful on their own, but only together. The transnational institutions and governments have to make rules clearer and focus on adding the issue of “cyber” in the hate speech discussion and law enforcement agencies should then be better trained to deal with such issues. This should be accompanied by efforts made by the platforms on which the hate is spread, especially social media companies should put more effort in fighting cyber hate in a consistent, efficient and fair manner. Lastly, the general mind set, specifically for the public, should shift. This can be helped on both sides of the coin, victims can be coached on how to defend themselves and not be defenceless anymore, and offenders should be taught that just because it is on the internet it does not make it ok. That is where campaigns, workshops, educational courses and such come into play, and where counter speech has a particularly important place.

To sum this up, a link should be made clear between the online and the offline world. What hurts online also affects offline. That is undeniable, but a hard truth for some to acknowledge. A majority of the people who spread hate would probably not dare to express such views and sentiments if they were standing right in front of the person they are attacking. This disconnect must change. If governments, law enforcement, social media platforms and the general public could all see that clearly and see that there is another way to make the world, online as much as offline, a safer and happier place that would be our hope for the future.