

## Shining a light on the dark corners of the web

Cybercrime researcher Gianluca Stringhini explains how he studies hate speech and fake news on the underground network 4chan.

Daniel Cressey

09 June 2017



Greg Kendall-Ball/Nature

Gianluca Stringhini.

Gianluca Stringhini spends his days in some of the shadier corners of the internet. As a cybercrime researcher at University College London, he has studied ransomware, online-dating scams and money laundering. In May, his team published two papers exploring how hate speech and fake news are spread around the Internet, focusing on the notorious but popular 4chan message boards.

In a conference-proceedings paper, the researchers analysed 8 million posts on 4chan's /pol/ ('politically incorrect') board, and traced how its users 'raid' other websites by posting inflammatory comments<sup>1</sup>. And in a preprint posted to the arXiv server<sup>2</sup>, they traced interactions between 4chan boards and other online communities, such as Twitter and Reddit, to examine how sites share links from known fake news sites, or from what the team calls 'alternative' news sources such as RT (formerly Russia Today). Stringhini talked to *Nature* about his research.

### **What made you decide to research 4chan?**

Nobody is really looking at these communities, but there is a lot of anecdotal evidence suggesting that they have an impact in the real world by spreading certain types of news. So we wanted to understand whether this is true, and to what extent they actually influence the rest of the web.

We started by just looking at 4chan. We selected /pol/, the politically incorrect board, which is where most alt-right users gather and discuss their world-views. We started by trying to understand the dynamics of these populations and this service. 4chan is very different from most other online sites in that it is both anonymous and its posts are ephemeral: they are deleted after a short while.

### **How did you go about it?**

We applied a number of techniques. We used a database containing hate words to understand what are the most prominent hate words, what is the incidence of hate speech and so on.

The percentage of /pol/ posts containing hate speech is 12%, whereas on Twitter it's 2%. It is reasonably higher, let's say. It's not perfect, because we used a keyword-based list, so we might actually be missing some hate speech that doesn't just fall into these pre-compiled categories. After understanding how this works, we started looking at how 4chan, and /pol/ in particular, influences the rest of the web.

### **And this is the subject of your paper<sup>1</sup> on 'raids' from 4chan to other websites? Was this something you already thought was happening?**

Yes. The limitations on what members of the research community have done so far are that they looked at the services in isolation. There is a lot of work towards understanding how attacks happen on Twitter, on YouTube, on Facebook. But there is not a lot of work on the source of these attacks, or their causes.

Because /pol/ is such a hateful platform, we saw empirically that often, people would post hyperlinks to YouTube videos that went against their world-views. They could be videos advocating for gender equality, feminism, tolerance. And then they would call for members to go and attack these people.

And so we would have a signal on 4chan that this link had been posted and people would be talking about it. And then we could see whether we could observe an effect on the YouTube comments to that video. We basically applied signal-processing techniques that have been used in radio signals to understand how synchronized these two signals are. There was a strong correlation between comments on YouTube spiking within the lifetime of a 4chan thread, and the amount of hate speech in those comments. This gave us evidence that these raids are really happening, and this will be grounds for future work. Now the question is, ‘So what?’ What do we do about it?

### **Can anything be done?**

This gives us an opportunity to identify videos that are at risk of being attacked. If YouTube only uses its own platform to identify raids, it can basically identify them as the raids are happening. But if it were looking at something else as well — an indicator that somebody is talking about this video in a hateful manner on a different platform — maybe it should start monitoring the comments more carefully. Or maybe, given that these threads on 4chan have a short lifespan, YouTube should disable comments on the video for the length of the lifespan.

### **In your paper on the arXiv<sup>2</sup>, you show that 4chan boards can influence the sharing of other news sources.**

Here, we studied whether, once an event happens on one Internet platform (say, a hyperlink to a piece of news), the same event happens on another platform. It will be the exact same news link being posted on /pol/ that then makes its way to Twitter, let’s say. We use a mathematical technique called ‘Hawkes-process modelling’, in which we can say with reasonable confidence that a particular event actually is related to the previous one that happened.

So we did this study, the first of its kind in tracing links between services. The idea here is that there has been quite some work on studying fake and alternative news. People look at how alternative news spreads on Twitter, for example; how people reshare it. But these services do not live in a vacuum — they’re part of the greater web. These places where alternative news stories are posted and they talk about them and they make up these crazy conspiracies and all of that: we wanted to understand whether this actually has an impact on the wider web.

What we found is that Twitter influences the other services a lot, which makes sense. Users of /pol/ and reddit will see news on Twitter, and then they will post those stories on their own boards and talk about them. But we also found that the opposite happens. To give you an example, we found that about 12% of the alternative news on worldnews — one of the main news boards on reddit — is coming from 4chan. And over 16% of the alternative news on the same board is coming from The\_Donald [a specific part of Reddit used by supporters of the US president].

### **Was it unpleasant reading all these posts?**

It's definitely a hateful place and quite unpleasant. It's not nice looking at it. My colleagues and I have some best practices: we advise whoever is working with us not to spend too much time continuously on the website, and to take breaks. We have this inside joke to every once in a while go and look at cat pictures.

*This interview has been edited for length and clarity.*

*Nature* doi:10.1038/nature.2017.22128

### **References**

---

1. Gonzalez-Bailon, S., Marwick, A. & Mason, W. (eds.) *ICWSM-17: 11th International Conference on Web and Social Media* (Association for the Advancement of Artificial Intelligence, 2017).
2. Zannettou, S. *et al.* Preprint at <https://arxiv.org/abs/1705.06947> (2017).