Good trumps evil in the battle against social media trolls

Mark Bridge Technology Correspondent

Good can triumph over hate on social media when trolls are challenged, research suggests.

Academics investigating antisemitism on Twitter found that most offending posts were ignored or shared within a small “echo chamber” of similar accounts whereas tweets challenging antisemitism were shared far more widely.

The Jewish charity the Community Security Trust commissioned the study from Cardiff University after reports of antisemitic incidents on the site spiked.

The researchers identified 2.9 million British tweets relating to Jews in the 12 months from October 2015, of which 15,575, or 0.6 per cent, were antisemitic.

The offending tweets peaked around the Brexit referendum and press coverage of antisemitism in the Labour party.

Researchers believe that many were designed to fuel social tensions.

Posts from community groups and news outlets were retweeted five to six times more often than the average tweet. Antisemitic posts were retweeted less often than average and over one to three days while positive, authoritative messages circulated for up to a week.

Matthew Williams, a co-author of the study, said that the findings should be a source of optimism because antisemitism was outweighed by positive content “which is present in greater amounts, lasts longer and spreads further than antisemitic content. There’s a common assumption that social media helps to spread hate but our research indicates that Twitter is self-regulating effectively as most people use the site responsibly and their voices are most influential.”

Tweets attacking the Jewish Labour MP Luciana Berger were retweeted a handful of times while her post about swapping was retweeted 13,802 times.

Leading article, page 31

Good Trumps Evil

An analysis of millions of tweets points to the decency of crowds

It takes more than bile to go viral. In fact, in the great Niagara of spin, salutation, half-formed thought and admonition that is social media, positivity lasts longest and hate dies fast. That is the reassuring finding of a study of nearly three million tweets about Jews analysed by a sophisticated algorithm and two academics from Cardiff University.

Only 0.7 per cent of the tweets were openly antisemitic. Many of these came from just 24 accounts, some of which were “bots” with minimal human input and all of which communicated most intensely with each other. Their tweets were shared, but for an average of just three days before drifting into digital obscurity. More thoughtful responses from Jewish organisations and mainstream media stayed in circulation from three to seven days.

Some will find this surprising, but it fits a pattern noticed by the same researchers in previous projects on anti-Islamic hate speech and Brexit-related abuse. Everything lasts for ever in the digiverse, of course, but, in the nuclear-waste vernacular, hate has a relatively short half-life.

There is a caveat: after spikes in malicious and negative tweets, the tone of online conversations seems to return quickly to normal, but it is in fact a new normal with what the researchers call a slightly higher “baseline of online hate”. In other words, the algorithm counts a few more provocative, needleling and outright abusive words per thousand tweets.

This is insidious, but it shouldn’t obscure the larger message that even as hackles and blood pressure rise online, good sense usually prevails. In the physical world the reverse is often true. In the pub, the bore can drown out the voice of reason by shouting louder. Perhaps we are learning that to type on Twitter is to etch our thoughts on the edifice of time. We are what we retweet.