

The Shift of Incel Topics During the Pandemic

Anh V. Vu

1 September 2020

Executive summary

CCC COVID Briefing Papers are an ongoing series of short-form, open access reports aimed at academics, policymakers, and practitioners, which aim to provide an accessible summary of our ongoing research into the effects which the coronavirus pandemic (and government responses) are having on cybercrime.

Building on our work in [Briefing Paper #5](#), we examine the fluctuation, particularly in response to the pandemic, of the topics being discussed among the incel communities on [incels.co](#) and [incels.net](#). We find that discussion remains centralised around a small number of key topics but there is a sustained increase in the volume of posts, especially during the first few months of the strict lockdown period that started in March 2020. We also see significant uplifts in the popularity ranking of particular topics since the outbreak, particularly for ‘*crime*’, ‘*politics*’, ‘*ethnicity*’, and ‘*family*’. It is no surprise that the popularity of topic ‘*illness*’ shows an increase comparable to the increase in the overall volume of posts.

Hidden topic model

We use Latent Dirichlet Allocation (LDA), an unsupervised machine learning model to discover latent topics over the collected posts. While each post, as a document, is represented as a distribution of topics, these topics are hidden (unlabelled), and each is considered as a distribution of keywords. Prior to that, the collected data is refined by removing stop-words and subsequently, lemmatisation. We then assign the main theme of a post by the dominating topic – the one with highest probability observed in the estimated distribution. By training several models over more than five million posts, the number of topics is decided based on the coherence score of trained models – a 70-topic model was chosen in our experiment from a range of models spanning 10 to 100 topics. Note that although the hidden topics are not necessarily labelled, their implicit meaning can usually be inferred from the keyword distribution.

Volume of posts by assigned topic

The discourse of incels is centralised around key topics – the top five topics account for nearly 24% of all posts made. The lockdown has led to a considerable rise in post volume among most of the topics, particularly these dominant ones, with a dramatic but short-lived peak in April 2020 (see Figure 1). The most popular topic of all time is *topic-0*, which has increased the most during the peak, accounting for around 14% of posts – nearly 4.5 times higher than the second highest one. However, we find this topic is rather general and it is hard to infer the implicit meaning from the keyword distribution (top 20):

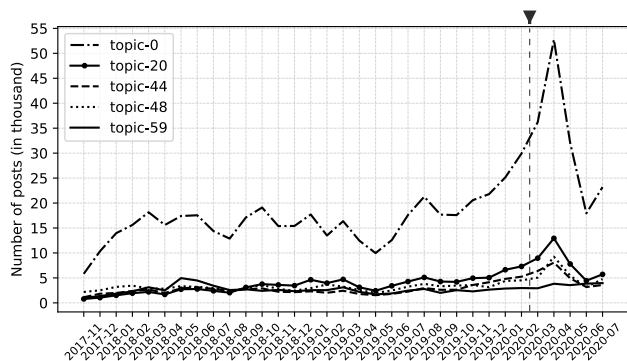


Figure 1: Number of posts for the top five topics

0.018 make + 0.016 create + 0.016 set + 0.014 build + 0.013 step + 0.011 learn + 0.011 order + 0.010 goal +
 0.009 thing + 0.009 require + 0.009 brain + 0.009 program + 0.008 computer + 0.008 part + 0.008 machine
 + 0.008 replace + 0.008 future + 0.008 plan + 0.008 level + 0.007 process

The shift of topic popularity

From these 70 topics, we manually selected a set of latent topics which showed a substantial increase of popularity ranking in the period after the WHO's declaration of a pandemic in March 2020. Figure 2 depicts the popularity ranking of five topics which showed particularly high increases. We observe that the timely topic 'illness' peaked in April 2020, 20 ranks higher than before the pandemic; however, it slipped back quickly afterwards to its previous position. We also see increases in discussions of

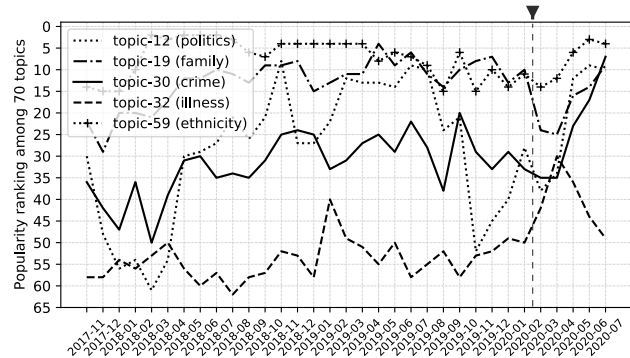


Figure 2: The popularity shift of topics

'family' – the posters presumably care about the health of their beloved ones, and worry about the pressure of coping with a difficult situation. When it comes to incel subcultures, 'ethnicity' is unsurprisingly consistently one of the top-tier topics, mostly ranked in the top 15 (with a high of second most discussed topic). The pandemic increased the popularity of this topic from rank 14 to 4. We see a clear link to the protests against racist violence in the USA. The topic mainly contains the following keywords:

0.228 white + 0.119 black + 0.079 ethnic + 0.078 race + 0.052 curry + 0.036 asian + 0.017 racist + 0.015 nigger + 0.012 indian + 0.011 half + 0.011 ricecel + 0.011 rice + 0.009 currycel + 0.009 noodlewhore + 0.008 mixed + 0.007 mix + 0.007 racism + 0.007 noodle + 0.006 whitecel + 0.005 tyrone

One of the most significant popularity increases is for 'crime', which appears 28 ranks higher compared to March 2020, jumping from 35 to 7. It mostly contains keywords related to violence and the justice system:

0.071 rape + 0.027 law + 0.019 prison + 0.019 crime + 0.019 victim + 0.018 murder + 0.017 police + 0.017 jail + 0.016 illegal + 0.014 violent + 0.014 legal + 0.013 case + 0.013 report + 0.013 violence + 0.012 abuse + 0.012 cop + 0.012 commit + 0.011 call + 0.011 criminal + 0.011 arrest

Another substantial increase is the 'politics' topic, also rising 28 ranks (from 38 to 10) in comparison to March 2020. This topic had been less popular since around end of 2019, so we believe it is likely that the rise is caused by the impending US election. It mainly contains keywords related to US politics, and to anti-semitic and anti-feminist conspiracy theories:

0.030 feminist + 0.028 support + 0.024 vote + 0.020 government + 0.016 jewish + 0.014 feminism + 0.013 liberal + 0.013 political + 0.012 power + 0.012 medium + 0.012 push + 0.011 anti + 0.011 group + 0.011 leftist + 0.010 state + 0.009 control + 0.009 movement + 0.009 politic + 0.009 free + 0.009 trump

Conclusions

The pandemic has not only intensified the volume of posts within the incel subcultures but also led to a considerable fluctuation of the topics being discussed within these online communities. Some of the topics have increased in popularity, as lockdowns undoubtedly led to difficulties in certain aspects of life, while the US political environment has had a clear effect in shifting discussions as well. Some topics appear to have been brought back from the past, after a long period of unpopularity.

At the Cambridge Cybercrime Centre we make our research data available to other academics, sometimes before we have looked at it ourselves! Researchers can be provided access to our 'CrimeBB' dataset of (26 and counting) underground cybercrime forums, our extensive collections of chat channel data, and our new collections of forums relating to online right-wing extremism and radicalisation. We can also share email spam and sensor data related to DDoS and IoT malware. All these collections are regularly updated and can be rapidly provided under license – for full details see: <https://cambridgecybercrime.uk>

The full set of CCC COVID Briefing Papers can be found at: <https://cambridgecybercrime.uk/COVID>

This work is licensed under CC BY 4.0. To view a copy of this licence visit: <https://creativecommons.org/licenses/by/4.0>