Measuring the Unmeasurable: Estimating True Population of Hidden Online Communities

DECISION: Accept

AUTHORS: Jonah Gibbon, Tina Marjanov, Alice Hutchings and John Aston.

Summary of Reviews

- Review 1: 0 (5)
- Review 2: 2 (3)
- Review 3: 1 (3)

Reviews

Review 1

Total score: 0

Overall evaluation: 0

Reviewer's confidence: 5

The paper seeks to measure two hidden population of online deviants (cyberoffenders and extremist forum members) by using a capture-recapture approach and a Bayesian model. The objective is sound as knowing the true size of these populations is useful from a theoretical and public policy perspective. However, the submission also suffers from a number of weaknesses that need to be addressed before it is accepted:

1. The paper ignores the criminological literature on the use of capture-recapture to measure hidden criminal population (in particular Bouchard and van der Heijden). These papers need to be acknowledged and their example followed: the outline their specific assumptions to apply a biology methodology to situations where human behaviours might differ.

2. The math is very detailed but often obscures more than it illuminates the subject: the application to cybercrime and the justification for certain decisions need to be stated a lot more plainly and connected to the subject matter.

3. The fact that registered users can be a good proxy for the unobserved population (section D and Table V) is surprising and needs to be discussed more thoroughly: what bother with such a complex methodology if a simple data extraction for registered users can produce roughly the same numbers? What have I missed here? If I understand correctly, the proposed solution is not very parsimonious.

4. The author needs to explain more plainly how the unobserved population differs from simple lurkers, which are a standard hidden population on all forums (even legal ones). If I did not understand correctly and the unobserved population is mainly made up of lurkers, this also needs to be mentioned, because it means that they are probably not very active members of this population and not very committed to contributing to the risk it poses to society. Either way, this needs to be clarified.

5. The conclusion needs to discuss less superficially what insights can be gained from this work: how can it inform public policy and how does it extend existing cybercrime scholarship?

Review 2

Total score: 2

Overall evaluation: 2

Reviewer's confidence: 3

The study estimates the hidden population behind online communities. To do so, the potential overlap between accounts in different forums is first estimated and second, a Bayesian model is used to estimate the hidden population, based on activity measures on these underground forums. The study concludes that the number of people participating in online discussions is 2 to 8.83 times larger than observed in cybercrime forums and 1.5-3.5 larger on extremist forums.

The study is comprehensive and well written. To be ready for publication, here are a few elements that the authors could consider adding/changing for better clarity:

- Most important: Throughout the text, it is unclear what the authors mean by "hidden" or "unobserved population". The authors mention that the "number of people participating in online discussion would be … higher". But if they participate in online forums, wouldn't we see their posts (implying participation)? Do the authors mean lurkers or individuals active on other forums or just cybercriminals at large? What the authors consider "hidden population" as opposed to "active population" needs to be better defined. This is the core of the paper and as is, I'm still uncertain what is truly estimated. Also, such explanation should be put forward in the abstract.

- The authors could better explain why the model detect a big increase in 2018-2019 for the hidden cybercrime population (figure 1). Is it because individual users are more active? More active forums (as shown in Fig.4). The trend is not seen for the observed population, so the estimation is a result of the model or the underlying data. What is the main explanation? Can this be explained and put forward next to the figure (and not later in the evaluation section).

- The authors put forward that their models show an increase in the hidden population during COVID for both populations, but the peaks are not readily apparent in figures 2 and 3. Could the author do a statistical test to confirm? The empirical support for this statement (which is also put forward in the abstract) is not extremely strong, could the authors be more extensive in their

demonstration? Otherwise, such result should be tampered down and maybe not put forward as a key result.

- The authors could better support their assumption that users will speak on many forums (to support the account overlap method). Why would they do so? Why wouldn't they just discuss on one forum? Also, what is the impact of such assumption on the model? If the model is recomputed assuming no account overlap, are the results similar?

- The concept "surface website" means not on the dark web? How does that decision impact the scope of the results?

Review 3

Total score: 1

Overall evaluation: 1

Reviewer's confidence: 3

This paper proposes a new approach to estimate the general population of online communities based on measurable online communities. The approach is broken into two steps. The first step involves identifying active users on multiple forums. This is achieved with setting conditions across features: content, time of posts, and names. Users meeting the threshold set for all three features are then considered to be the same users. The second step involves applying a Bayesian multiple systems estimation approach developed by Marique-Vallier. This approach is then used to estimate the size of the total population. To test the approach for estimation, the author(s) utilized the CrimeBB and ExtremeBB datasets from the Cambridge Cybercrime Centre. Specifically, the author(s) focused on forums between January 2012 and December 2022 due to the level of activities across forums. The results showed that the estimate on general population of cybercrimes is up to 8.3 times larger than the observed population, and 2.7 times larger for the general population of extremists. The author(s) then carried out evaluation to ensure that the validity of their findings.

This paper adds to current literature by introducing an innovative and feasible approach to provide more accurate estimates of the general population engaging in online criminal and deviant communities. This has implications for policy development and evaluation. For example, knowing that the extremist online communities on forums may be smaller than expected would have implications on resource allocations for law enforcement and intelligence communities.

Here are some comments on the paper:

1) For "Content Metric", the author(s) chose to focus on English-only posts due to the lack of existing tools to analyze posts in other languages. This is understandable. However, the author(s) need to be careful drawing conclusions on geographic regions based on the British and American English dialects. In Footnote 4, the author(s) did provide Citation [26] that splits English-speaking

nations into three different categories. However, Citation [26] is published in 1997 so it may be necessary to provide a more recent source.

2) A related comment on the issue is the treatment of posts where more than one language may be present. Did the author(s) remove these posts from the dataset, and if so, what percentage of total posts did these constitute? If these posts were included, how were the non-English portion analyzed?

3) The author(s) should consider an explanation for the results in Table II. With Sentence 1 and Sentence 2 in the first row, it was surprising to see that the CrimeBB similarity is lower than the ExtremeBB similarity since both sentences mentioned hacking. This is confusing, especially since the last sentence prior to the table states that "two content vectors were similar if the cosine similarity between them was greater than 1/2".

4) The author(s) should justify how relaxing the conditions from Citation [10] would help improve the identification of user overlaps across forums. The inclusion of content and time metrics would mean that only actively posting users on forums would be captured. For registered who do not actively participate, it is unclear how this group of users would factor into the Bayesian multiple system estimation approach.

5) The author(s) mentioned a sharp increase in the estimated population for cybercriminals in late 2018 and early 2019. The author(s) should identify possible events that may contributed to the spike. For example, would any events such as the increase of ransomware played a factor?

6) For extremist population, the author(s) discussed the peak in estimated population during the COVID-19 pandemic in March 2020 and remined high around late 2021. In Figure 3, however, the graph depicted a steady increase between 2016 to a speak in early 2015. The author(s) should consider providing an explanation on why the increase/spike may not be relevant (especially since this was near the time of the 2016 election in the United States, which was an event that has been shown to influence some of the extremist communities).

7) While speculating about the true extremist population being smaller than it might seem, the author(s) should take the introduction of non-mainstream platforms such as Gab and Parler into consideration.

8) The evaluation on the number of forums it takes for the estimates to plateau is excellent and further validate the results. It is comparable to sensitivity analyses perform for other statistical models.