

## Determining the key sources, speed and evolution of the immigration debate on Twitter in the UK

Andrea Nasuto<sup>1</sup>, Francisco Rowe<sup>1</sup>, Marzia Rango<sup>2</sup>  
<sup>1</sup>University of Liverpool, <sup>2</sup>IOM

### Background and Motivation

Xenophobia and racism harm society, disrupt human progress, create horrific mental and physical abuses, sustain existing forms of inequalities. Online social media are increasingly important in shaping public debate and they substantially affect humans' behaviors, including in the immigration public debate. Online social media have been recurrently pointed out as a source of uncontrolled racism and xenophobia. The COVID-19 pandemic has brought a largely untraced new wave of xenophobic content. These violent behaviors are hard to detect at scale and the structure of the UK online immigration debate is largely unknown. The 'virality' of the social media content has been a rightful major concern in the context of xenophobic messages but the speed at which they spread has been largely ignored. Previous studies have shown how small groups of people play a decisive role in spreading and generating toxic content, particularly misinformation but it is still unclear if this holds true for the public debate on immigration. Taming online hate speech is crucial to stop anti-immigration mental and physical acts of violence.

### Data and Methods

Drawing on a dataset of 220,870 tweets, the research studies the key spreaders, key producers, evolution and speed of the anti (pro) immigration content on Twitter between December 2019 and April 2020 in the UK.

The research uses a deep learning BERT text-classifier to identify tweets accordingly to their sentiment towards immigration. The classifier has AUC-ROC score equal to 0.82 and an average F1 score equal to 0.67. Anti (pro) immigration networks have been built using social network analysis (SNA) to analyze existing polarization patterns.

Using the Birdspotter Python library, the research identifies suspect bot users to track their impact across speed, spreading, production, tweets cascades and networks.

### Key Findings

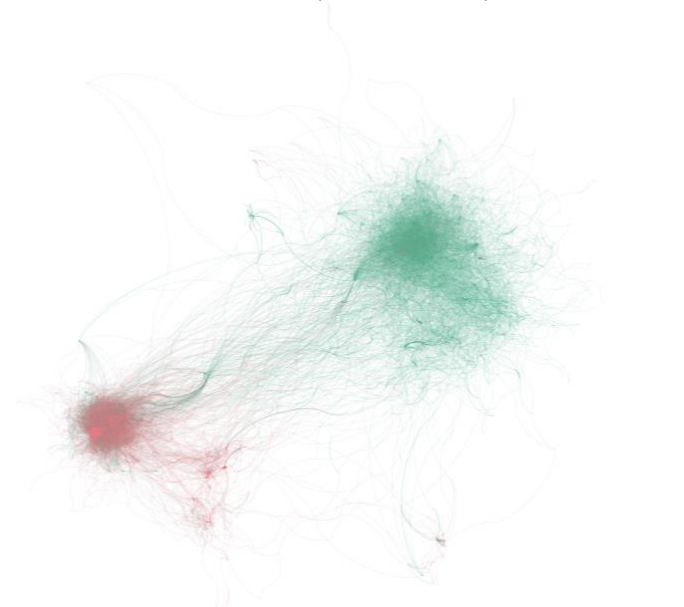
Results show that the public debate on immigration is largely polarized. The anti-

immigration network is more homogeneously connected while smaller in size than the pro-immigration community. It has few key nodes according to the overall number of connections established in the network (Figure 1). This is also underlined by looking at the leading 1% users by the number of retweets shared (key

spreaders) and tweets produced (key producers). A substantial fraction of the immigration-related tweets are generated by these key sources. These patterns are particularly stronger in the anti-immigration network. The top 1% anti-immigration key spreaders and producers create respectively 23.18% of the retweets and 21.36% of the total tweets in the community. Compared to the pro-immigration network, these percentages are considerably smaller, respectively 6.01% of the retweets and 11.69% of the tweets shared.

Figure 1 – Users Network Graph.

Retweets directed graph of anti-immigration (in red) and pro-immigration (in green). Each node is a user and edges are retweets between a source (user creating the original tweet) and a target (user retweeting). The size of the node is proportional to the number of connection (both in and out) each node has.



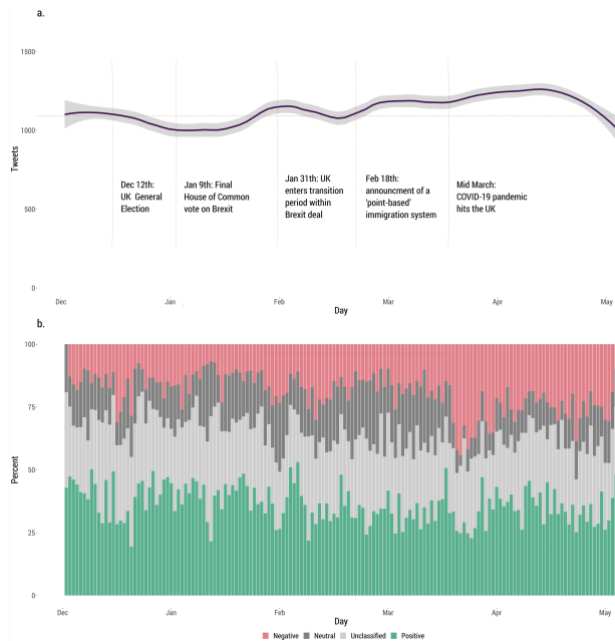
Negative immigration content spreads 1.66x faster and it is consistently shared more than positive immigration messages. Bots do not play a significant role in determining the content speed and size of the tweet cascades.

The spread of the COVID-19 in March 2020 is associated with an increase in the volume of anti-

immigration tweets but it is offset by an equal increase in pro-immigration content (Figure 2).

Figure 2 – Tweet fluctuations in the UK between December 1st 2019 and April 30th 2020.

A) Overall number of tweets. Smoothed conditional means are reported and were estimated via locally weighted scatterplot smoothing (loess) using a span of 0.3. B) Percentage of tweets by sentiment towards immigration as labeled by the BERT classifier.



causes including existing geographical polarization, social media algorithmic biases and the attention-seeking architecture of the online platforms.

Tweets on immigration increase accordingly to major events as these events seem to generate two different types of reactions: celebratory and opposite.

The findings are further investigated looking at the potential root causes grounded in social theories around online extremism on immigration. Particularly, the higher level of engagement seen in the key sources of content might suggest an interesting link between content consumption and polarization. Social media platforms rely on a designed attention-seeking mechanism to incentivize engagement which drives 'reactions' rather than deliberation.

### Value of the research

The research offers actionable insights to policymakers to contrast xenophobic extremism on social media and to design new policies to foster civil discussions on immigration.

Specifically, the study reveals that deactivating a very small group of anti-immigration extremists could largely disrupt xenophobic content. At the same time, an intervention to reduce the polarization of the immigration public debate needs to consider multiple potential