Twitter as an Indicator of Censorship

Ravindra Mounika

Master thesis in Computer Science

Widespread use of the Internet has spurred up the utilization of Online Social Networks making it one of the most popular sources of news and information. Due to its public nature and availability to a wide audience, countries across the world have imposed limitations on social media access on their citizens across the borders based on what the regimes regard as inappropriate. A third of the world's countries have restricted social media access since 2015. We use one such platform, Twitter as a tool to learn which content countries are interested in blocking. The goal of this research is to measure the censorship landscapes in two distinct countries from different regions of the world, India and Germany, through insights from Twitter data. We start with a pre-collected dataset of censored tweets from archive.org for topical and temporal analysis to get a better understanding of censored content. We analyze the patterns of tweet and account censorship by the governments in our case studies. We observe that India mostly is concerned with people tweeting about Kashmir and censors tweets revolving around this topic. On the contrary, In Germany, we observe that most censored tweets and accounts seem to be promoting extreme right content as the hashtags contain the themes of white supremacy and Islamophobia. We introduce a machine learning approach to predict tweet and account censorship. Our results reveal that India censors at the account level and the prediction of account censorship yields an AUC of 88 using Random Forest and a combination of features like user profile features, stance-based features, textual features making it a feasible task.

Prof. Philippe Cudré-Mauroux