U.S. Religious Landscape on Twitter

Lu Chen\textsuperscript{1,*}, Ingmar Weber\textsuperscript{2}, and Adam Okulicz-Kozaryn\textsuperscript{3}

\textsuperscript{1} Kno.e.sis Center, Wright State University, Dayton, OH, USA
chen@knoesis.org
\textsuperscript{2} Qatar Computing Research Institute, Doha, Qatar
iweber@qf.org.qa
\textsuperscript{3} Rutgers-Camden, Camden, NJ, USA
adam.okulicz.kozaryn@gmail.com

Abstract. Religiosity is a powerful force shaping human societies, affecting domains as diverse as economic growth or the ability to cope with illness. As more religious leaders and organizations as well as believers start using social networking sites (e.g., Twitter, Facebook), online activities become important extensions to traditional religious rituals and practices. However, there has been lack of research on religiosity in online social networks. This paper takes a step toward the understanding of several important aspects of religiosity on Twitter, based on the analysis of more than 250k U.S. users who self-declared their religions/belief, including Atheism, Buddhism, Christianity, Hinduism, Islam, and Judaism. Specifically, (i) we examine the correlation of geographic distribution of religious people between Twitter and offline surveys. (ii) We analyze users’ tweets and networks to identify discriminative features of each religious group, and explore supervised methods to identify believers of different religions. (iii) We study the linkage preference of different religious groups, and observe a strong preference of Twitter users connecting to others sharing the same religion.

1 Introduction

Religiosity is a powerful force shaping human societies, and it is persistent – 94% of Americans believe in God and this percentage has stayed steady over decades \cite{30}. It is important to study and understand religion because it affects multiple domains, ranging from economic growth \cite{1}, organizational functioning \cite{10} to the ability to better cope with illness \cite{3}. A key feature of any belief system such as religion is replication – in order to survive and grow, religions must replicate themselves both vertically (to new generations) and horizontally (to new adherents). The Internet already facilitates such replication. Traditional religions are likely to adapt to the societal and historic circumstances and take advantage of social media. Many churches and religious leaders are already using social networking sites (e.g., Twitter, Facebook) to connect with their believers.

* This work was done while the first author was an intern at Qatar Computing Research Institute.
While social networking and social media become important means of religious practices, our understanding of religiosity in social media and networking sites remains very limited. In this paper, we take a step to bridge this gap by studying the phenomenon of religion for more than 250k U.S. Twitter users, including their tweets and network information.

Twitter, because of its global reach and the relative ease of collecting data, is becoming a great treasure trove of information for computer and social scientists. Researchers have studied various problems using Twitter data, such as mood rhythms [14], happiness [12], electoral prediction [7], or food poisoning [8]. However, studies that explore the phenomenon of religion in social networking sites are still rare so far. To date, the most relevant study investigates the relationship between religion and happiness on Twitter [29]. It examines the difference between Christians and Atheists concerning the use of positive and negative emotion words in their tweets, whereas our work focuses on the religiosity of Twitter users across five major religions and Atheism. One recent study [26] addresses the prediction of users’ religious affiliation (i.e., Christian or Muslim) using their microblogging data, which focuses on building the classification model but not studying the phenomenon.

We collected U.S. Twitter users who self-reported their religions as Atheism, Buddhism, Christianity, Hinduism, Islam, or Judaism in their free-text self-description, and further collected their tweets and friends/followers. Our dataset comprises 250,840 U.S. Twitter users, the full lists of their friends/followers, and 96,902,499 tweets. In particular, we explore the following research questions in this paper:

1. **How does the religion statistics on Twitter correlate with that in the offline surveys?** Our correlation analysis shows that: (1) There is a moderate correlation between survey results and Twitter data regarding the distribution of religious believers of a given denomination across U.S. states, e.g., the macro-average Spearman’s rank correlation of all the denominations is \( \rho = .65 \). (2) Similarly, the fraction of religious people of any belief within a given U.S. state in surveys matches well with that of Twitter users referencing any religion in their profiles with a Pearson Correlation of \( r = .79 \) (\( p < .0001 \)).

2. **Whether or not do various religious groups differ in terms of their content and network? Can we build a classifier to accurately identify believers of different religions?** Specifically, (1) By looking at discriminative features for each religion, we show that users of a particular religion differ in what they discuss or whom they follow compared to random baseline users. (2) We build two classifiers that detect religious users from a random set of users based on either their tweets’ content or the users they follow, and we find that the network “following” features are more robust than tweet content features, independent of the religion considered.

3. **Does the in-group linkage preference exist in any particular religious denomination?** Our main findings include: (1) We find strong evidence of same-religion linkage preference that users of a particular denomination would have an increased likelihood to follow, be-followed-by, mention or retweet