From Playboy Cricketer to Islamist Politician?
An Analysis of Imran Khan's Tweets Pre and Post Aug 2018

By: Maryam Khalid Shah

Background

- Former captain of Pakistan's cricket team
- Oxford graduate
- Ex-husband of a British billionaire tycoon's daughter
- Friend of Princess Diana's
- Self-proclaimed liberal

Background

- Elected as Pakistan's Prime Minister in 2018 (removed in Apr 2022)

- As Prime Minister, he stressed on Islam being the foundation of the Pakistani state and society.

- Political opportunist?

Theory of Interest

- "Religious/Islamic tweets" → tweets containing Islamic terms
- Including:
  - Tweets about religion/Islam
  - Tweets not about religion, but using Islamic terms

- Research questions:
  - *Has the number of Imran Khan's tweets that contain religious (Islamic) elements increased over time?*
  - *Have Imran Khan's (religious) tweets become <u>more</u> religious over time?*

Data Collection

- snscrape
- December 2014 to May 2022 (45 months pre and 45 months post August 2018)
- Total tweets = 5259
- English tweets = 3911
- All English tweets combined contained 621,423 words

Data Cleaning

- Removed:
  - Numbers
  - URLs
  - Mentions
  - Urdu hashtags
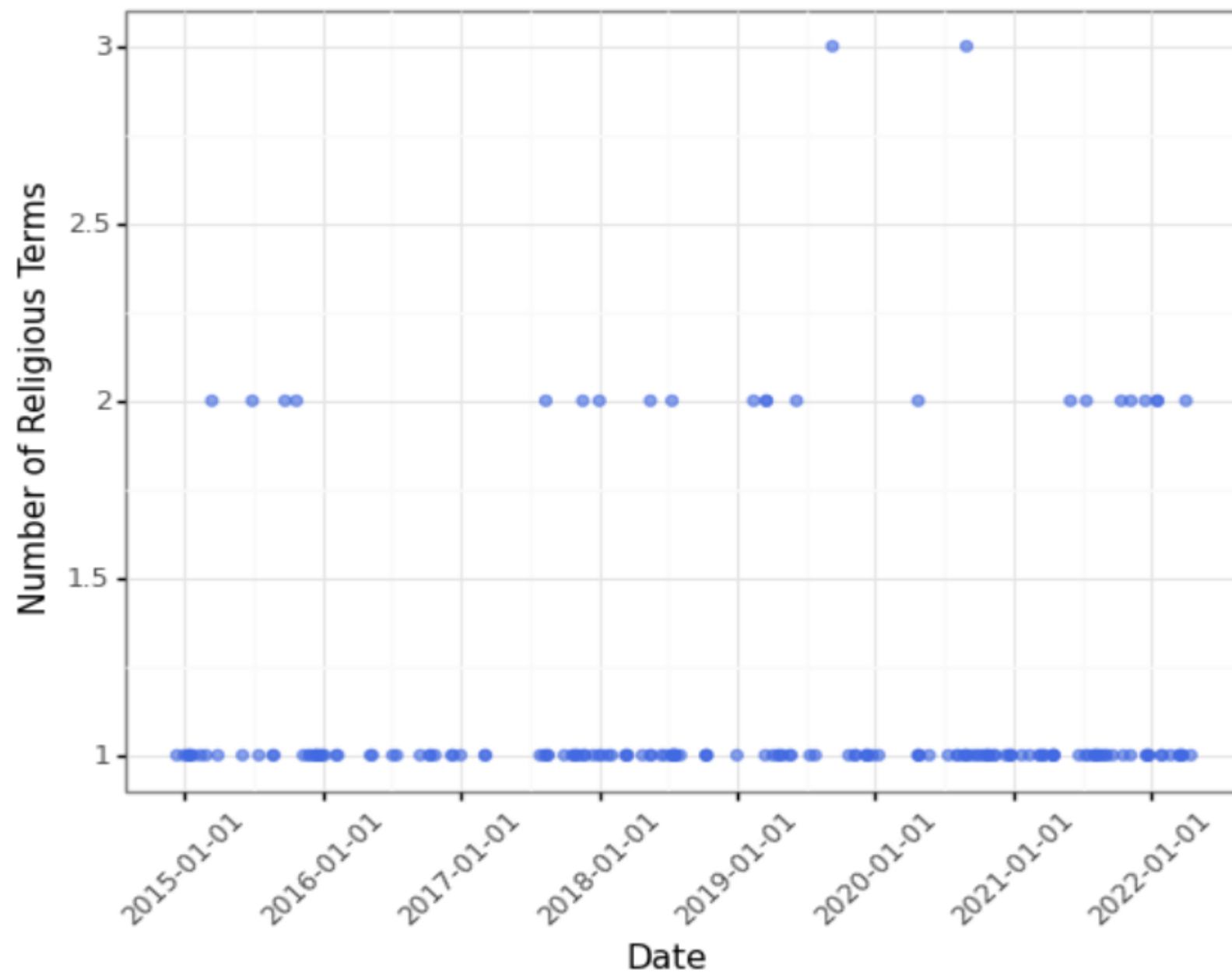 - Roman Urdu words

- Expanded abbreviations

جلسہ

↑

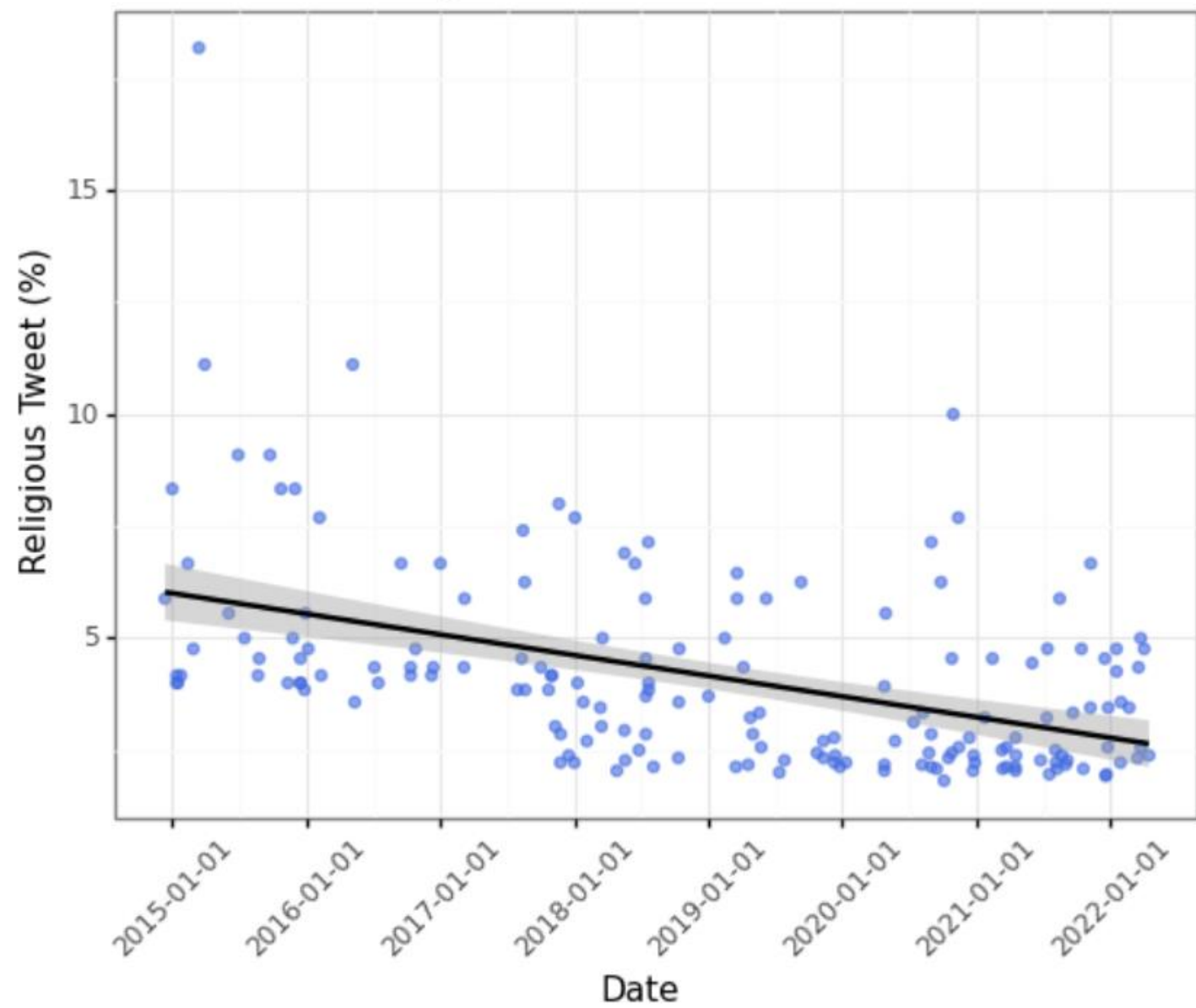"Thank you for coming to the *jalsa*"

Religious/Islamic Terms

- Tweets manually scanned to identify Islamic terms.
- These include: 'martyrs', 'karbala', 'imam hussain', 'prophet', 'baatil', 'inshaallah', 'riyasat-i-madina' and more.
- List used to guide BERTopic

Data Exploration

"Religious" Tweets Over Time
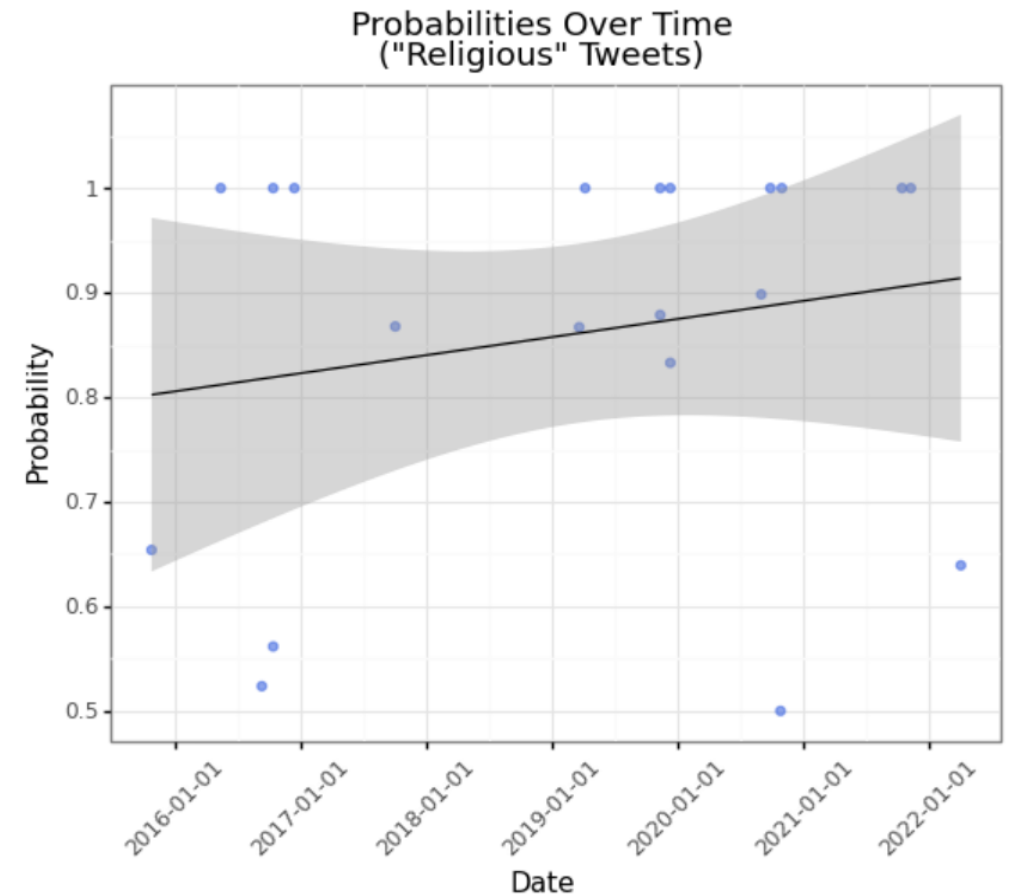
"Religious" Tweets Over Time

Guided Topic Modeling

- **Guided Topic Modeling is a collection of techniques that guides the topic modeling approach by setting a number of seed topics that the model will converge to.**

- Created a seed topic list containing religious/Islamic terms that was passed through the model.

- 59 topics were created, with the highest frequency of a topic being 374.

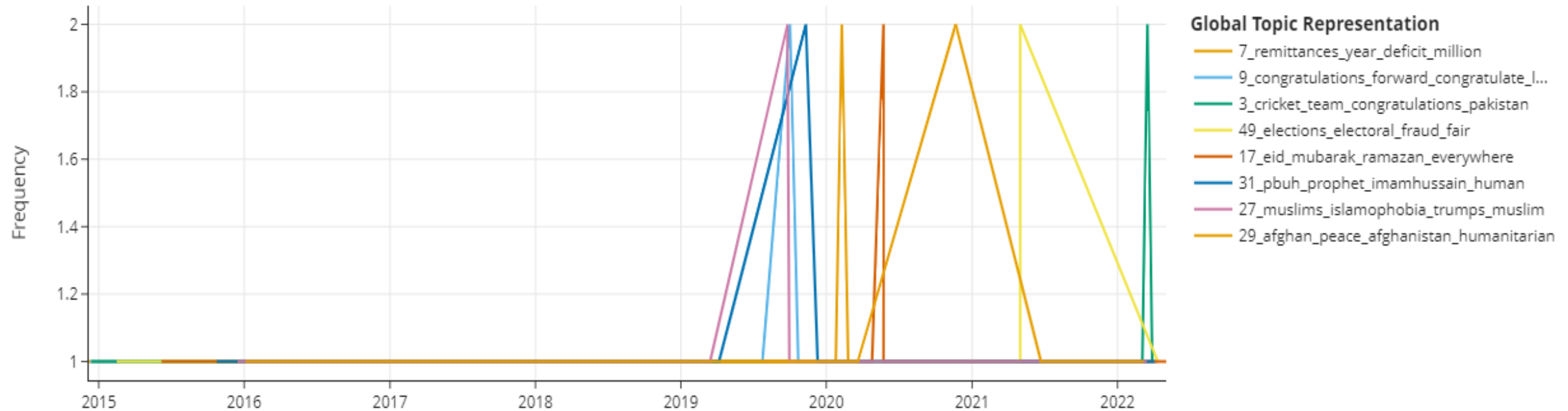- The topic of interest was 30th in the list, with a frequency of 22.

Guided Topic Modeling

- Looked at only those tweets where the dominant topic was the one of interest – less than 30.

- The probability of a tweet having the 'religious' topic increased over time.

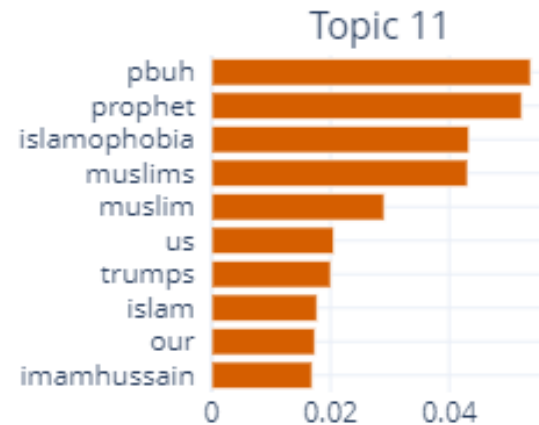Probabilities Over Time
("Religious" Tweets)

Dynamic Topic Modeling

Topic Modeling

- As suspected, without guiding the topic model, it created one topic that captured all Islamic terms.

Topic Word Scores

Text Classification

- Given the content of Imran Khan's tweets, can we predict whether his tweet would be "religious"/Islamic (based on our 'definition')?

- Comparing support vector machine classifiers (SVC and LinearSVC)

- Under-sampled the majority class

Classifier	Accuracy	Precision	F1 Score	Recall
SVM	0.85	0.87	0.86	0.85
Linear SVM	0.90	0.93	0.90	0.87

Was Success Achieved?

- Creating a corpus of Imran Khan's tweets across 8 years with their corresponding dates

- Cleaning the corpus and getting it ready for analysis

- Using guided topic modeling correctly

- Being able to answer the 2 main questions I have using my topic modeling results.