Figure 7: Timeline of ISIS attacks during 2021

Keyword	Retweeters	Quoters
mosque	4,989	36,881
al-fitr	3,095	15,151
bomb	633	9,541
gun	423	4,653
suicide	316	5,204
taliban	257	17,462
kabul	61	2,240
kandahar	10	489
jalalabad	1	46
kunduz	1	30
baghlan	0	107

Table 9: Frequencies of 11 keywords related to IS-K attacks in the dataset. We have combined the frequencies for each term in English and Arabic.

identify the impact of existing interventions already present in the dataset (e.g., account suspension or tweet removal), and then design new anti-normative nudges based on generative language models that are informed by learned models of extremist influence. While there is evidence that networks and beliefs collectively evolve (**?**), interventions designed to shape this evolution must be carefully designed to avoid unintended effects (**??**). For example, an anti-ISIS message inserted into the network to maximize exposure to ISIS sympathizers could backfire, instead strengthening the ISIS support of the sympathizers. Accordingly, carefully designed solutions must be engineered for effective anti-normative mechanisms, building on recent advances in deep learning for generating natural language (**??**).

5 Limitations

Our proposed ISIS20215-21 dataset is innovative and unique in its kind, due to the specific sub-population it accurately targets and its evident longitudinal properties. Given how accounts were originally selected and their validation obtained through manual inspection and cross-correlation with CtrlSec data, we are confident about the authenticity of the accounts and their relevance for online extremist research. Nevertheless, our dataset suffers from some limitations that should be aware to researchers interested in the data's future use. First, due to the existing limitation in Twitter's API, we cannot collect the complete history of each account. This however, still allows us to include tweets for many accounts that span few years of online presence (see Figure 2).

Second, even though the dataset offers a good excerpt of ISIS' evolution, there is a growing evidence that ISIS may have moved away or shifted from Twitter for most of its online activities (**?**). Accordingly, while we can get a look at sympathizers and their messages, we may be missing a significant portion of their online presence and activities. Related, we suspect that there may be other accounts, that we *do not* include in our dataset, that are close to ISIS. We are also unable to potentially remove noisy tweets, irrelevant to our study. However, results of our preliminary analysis on the data strongly suggest that our collected content is relatively clean, and will constitute a strong seed for researchers in this space.

6 Conclusion

In this paper, we present a new public dataset tracking tweets from two sets of users that are suspected to be ISIS affiliates. These users are identified based on a prior study and a campaign aimed at shutting down ISIS Twitter accounts. This study and the dataset represent a unique approach to analyze ISIS data. Although some research exists on ISIS online activities, few studies have focused on individual accounts.

The long-term aim of this project is to tackle the ambitious challenge of linking social media observations directly to online political extremism. We hope that researchers will be able to leverage the ISIS2015-2021 dataset to obtain a clearer understanding of how ISIS sympathizers use these platforms for message spreading and propaganda, especially in the recent years and with the advent of ISIS-K. The dataset can also be used as a benchmark to investigate online strategies originating from extremist groups that are not ISIS related. In turn, such insight might enable social network providers but also policy makers toward new interventions lessening the impact of extremist group influence.

7 Ethical Considerations

The dataset only collects data that is public and from alive accounts. Private accounts' tweet IDs and data are not included. Tweets limitations imposed by Twitter's API and ToS are respected. We have also removed all the tweet texts, full URLs, user account details, and only provide aggregated statistics, URL domains, and media keys that are not directly connected to any specific user or tweet and can be used for aggregate analysis to download the original tweets as long as they are available to the public.

Dolore ab dolores maxime tempore optio nemo, commodi error quae cupiditate earum quam quisquam a atque nihil, dolores odit praesentium est beatae molestiae tem-

pora tempore?Maiores praesentium dignissimos accusamus, saepe ipsum dignissimos a beatae?Veniam illo vitae, tenetur neque quasi dolores