

# Text Mining Islamic State's Online Propaganda

Stephane Baele, Katharine Boyd, Travis Coan

## CREST project overview:

### Central research question:

How does Islamic State (IS) online propaganda demonstrate dynamics of radicalisation?

### Primary aims of the project:

- 1) Develop tools to harvest and analyse online content.
- 2) Use these tools to examine how propaganda taps into known dynamics of radicalisation (individual cognitive factors, intergroup conflict).
- 3) Examine if the content structure, logic, and language differs across type of online propaganda outputs (magazines, videos, photos).

## Methodologies employed:

### Video-to-text transcriptions & validation

We used IBM Watson to transcribe English and Arabic audio of IS videos. We then conducted validation analysis to assess the accuracy of the transcriptions in both languages

### Dictionary-based content analysis

We identified words associated with specific topics in IS content (e.g., violent jihad, group polarisation, victimisation, etc.) to conduct content analysis chronologically within platforms (e.g., magazines, videos) and between platforms.

### Linguistic networks analysis

We used word or phrase 'co-occurrence' to develop a network-based approach to analysing IS propaganda.

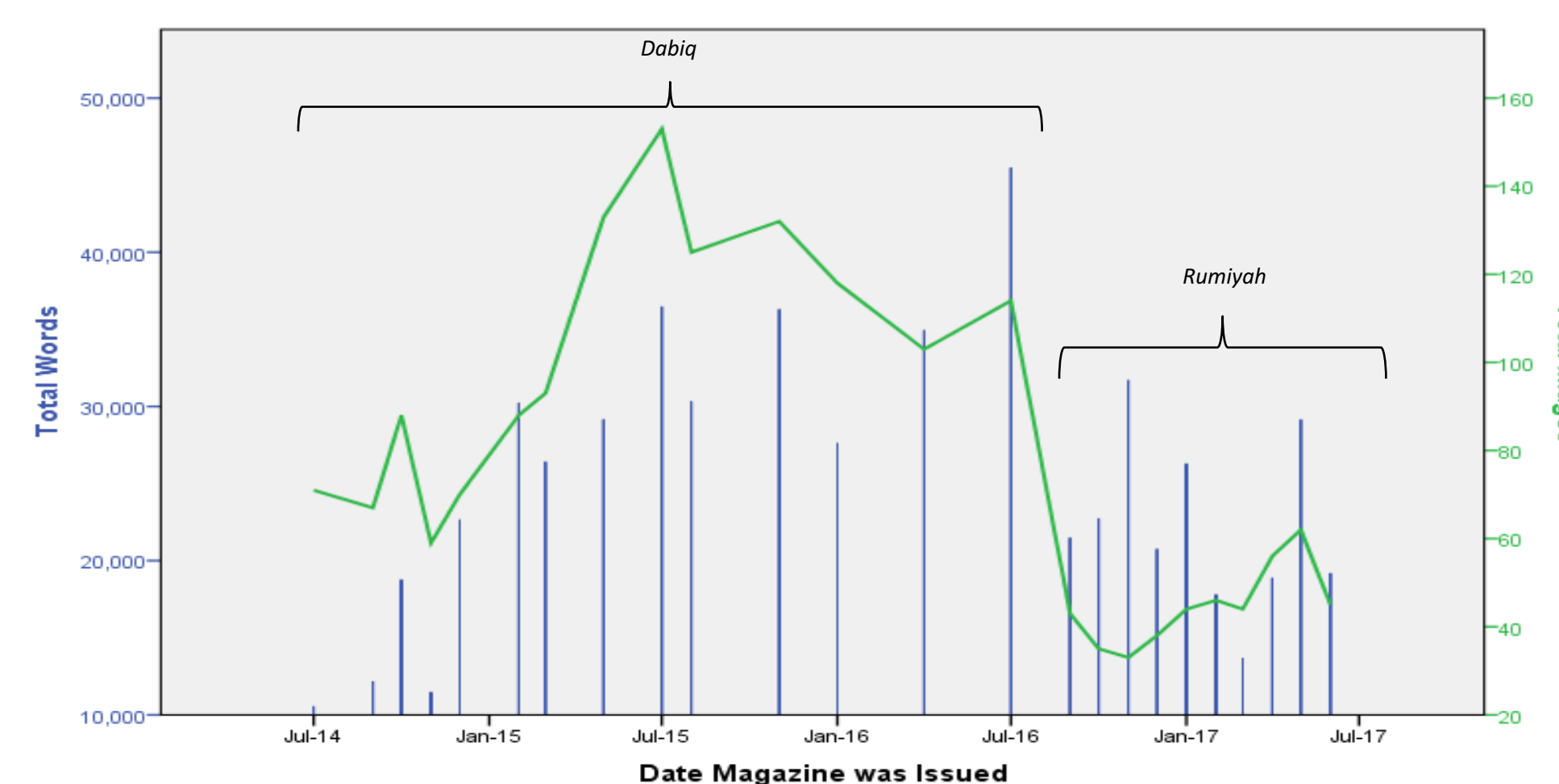
### Visual coding

We developed a methodology for coding extremist image content and used this method to analyse images in IS' *Dabiq* and *Rumiyah* magazines.

### Project outputs:

- Paper on images used in IS magazines
- Paper on language use in IS videos
- Paper on language use across IS propaganda platforms
- Edited book on IS propaganda
- Research note for UK parliament
- Short policy brief note.

## Magazine content



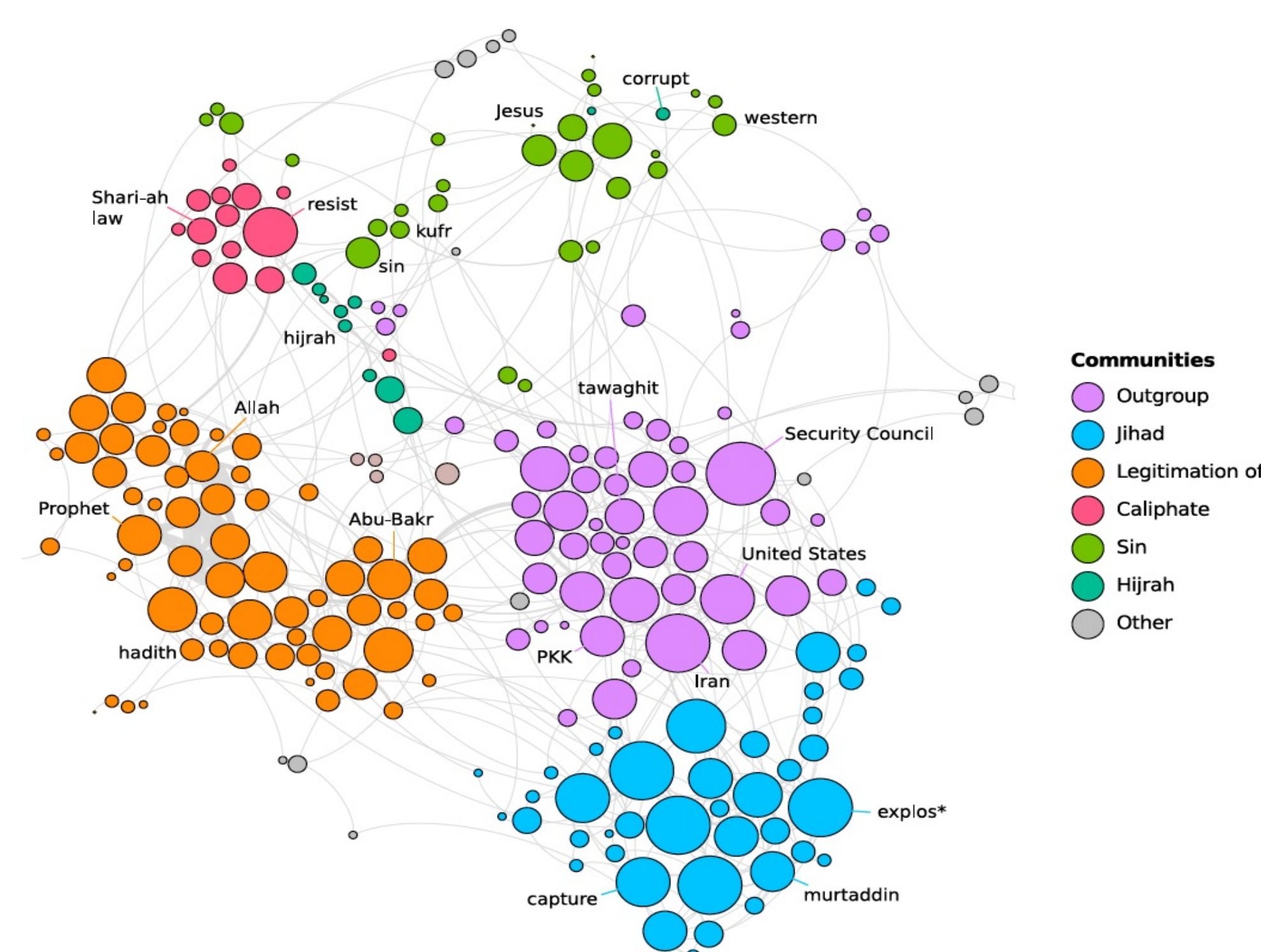
Depiction of when *Dabiq* and *Rumiyah* magazines were released and the word and image count in each issue.

### Images

- Over 1,930 images in 25 issues
- IS magazines feature 'solution' more than 'crisis' images across issues
- The most frequent narrative featured in IS images is violent jihad, followed by outgroup plots
- IS uses comparable proportion of gruesome images in *Dabiq* and *Rumiyah*.

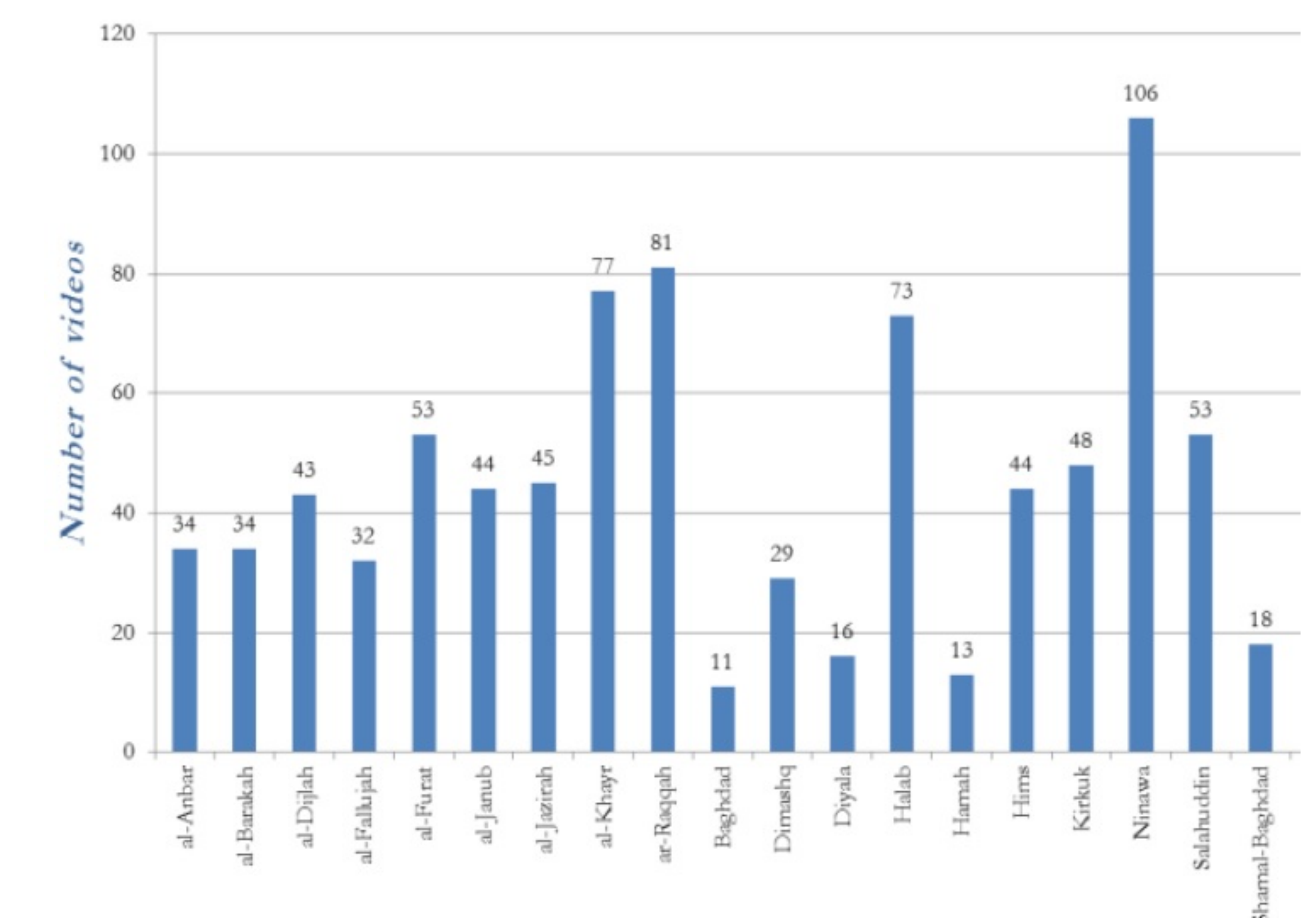
### Language

- Dictionary-based content analysis shows polarised language used to describe distinct outgroups ('*Kuffar*', '*taghut* regimes', 'apostates', 'crusaders' and 'Americans' (as well as 'Jews'), and shi'as) with greater frequency referencing the near enemy than the far enemy.
- Network analysis shows communities with the size of the circle representing the number of connections the word has with other words. These communities relate to specific IS narratives.



## Video content

- 1,278 IS videos between Dec 2014 - Jan 2017
- Majority (over 850 videos) produced by core province (*wilayah*), followed by the central office (over 250), then periphery provinces - showing the de-centralised character of IS video propaganda.



Number of videos released by each core province.

- Most videos are 3-10 minutes, some more than a half hour, few over an hour.
- The number of videos is decreasing over time across media centres.

## Discussion

- Outputs connect findings with mechanisms of radicalisation showing how language and images contribute to: polarisation, provoking emotions that elicit action, contributing to 'crisis' and 'solution' narratives, and encouraging viewers with risk and status.
- The linguistic network analysis methods are scalable to study large bodies of text in a fast, reliable way that allows security practitioners to quickly and easily analyse and visualise the underlying structure of extremist group propaganda.
- The visual methods developed can be used to make comparisons of propaganda strategies across extremist groups.