STEPHANE BAELE, KATHARINE BOYD AND TRAVIS COAN

EXTREMIST PROSE AS NETWORKS

USING CO-OCCURRENCE NETWORKS TO MONITOR EXTREMIST COMMUNICATIONS

A well-developed body of research demonstrates that network analysis offers considerable insight by studying the links between political actors, groups, and entities. At the same time, there is also a growing body of work that focuses more specifically on extremist networks. While much of the published research on extremism centres on 'social networks'—i.e., networks of individuals—our research takes a different approach, focusing instead on the network of words and concepts embedded in extremist propaganda. The results suggest that viewing textual information through the lens of network analysis offers a powerful way to complement existing approaches to monitoring and analysing extremist communication.

There are a number of different ways to extract relational information from written communication, but the most common approach focuses on the concept of word (or phrase) 'co-occurrence'. This approach assumes that co-occurring words also tend to share a common meaning or represent a common theme. For instance, our research on Islamic State (IS) propaganda demonstrates the words 'American' and 'crusader' often co-occur, while also sharing a common reference to an important outgroup. Despite its simplicity, the idea of co-occurrence provides the foundation for many of the most commonly employed text mining algorithms.

Using the basic idea of word co-occurrence, it is relatively easy to develop a network-based approach to examining textual information. In a co-occurrence network, each of the words in a body of text represent the 'nodes' in the network, and any words that commonly appear close together represents an 'edge' connecting two nodes. The connections (or edges) in these networks can be 'undirected' (i.e., simple cooccurrence) or 'directed' (i.e., represent a dependency

relationship between words – for instance 'effeminate scholars' in IS prose). Importantly for researchers and practitioners, text networks offer a flexible and relatively quick way to analyse large quantities of extremist propaganda by drawing on recent advances in the field of network science.

To gain a better sense of how this analysis works in practice, consider a co-occurrence network for the IS magazine, Dabiq (issues 1-15). A visual representation of the network is provided in Figure 1. Through our research, we observe five major 'communities' or themes: 'outgroup' (portrayals of the identity and evil actions of IS' perceived outgroups), 'jihad' (descriptions of IS' war against these enemies), 'Caliphate' (depictions on the perfect life and mores in the established Caliphate), 'legitimation' (discussions of the deeds and sayings of Muhammad and his companions in ways that legitimise the

establishment and actions of IS), and 'sins' (debates on the major sins such as 'kufr' and 'shirk', associated with the 'corrupt' 'Western' and 'Christian' world). These themes are closely connected through important bridge words (e.g., 'hijrah' between 'outgroup' and 'Caliphate', 'Dabiq' between 'outgroup' and 'legitimation,' and so on) and thus create a coherent system of meaning.
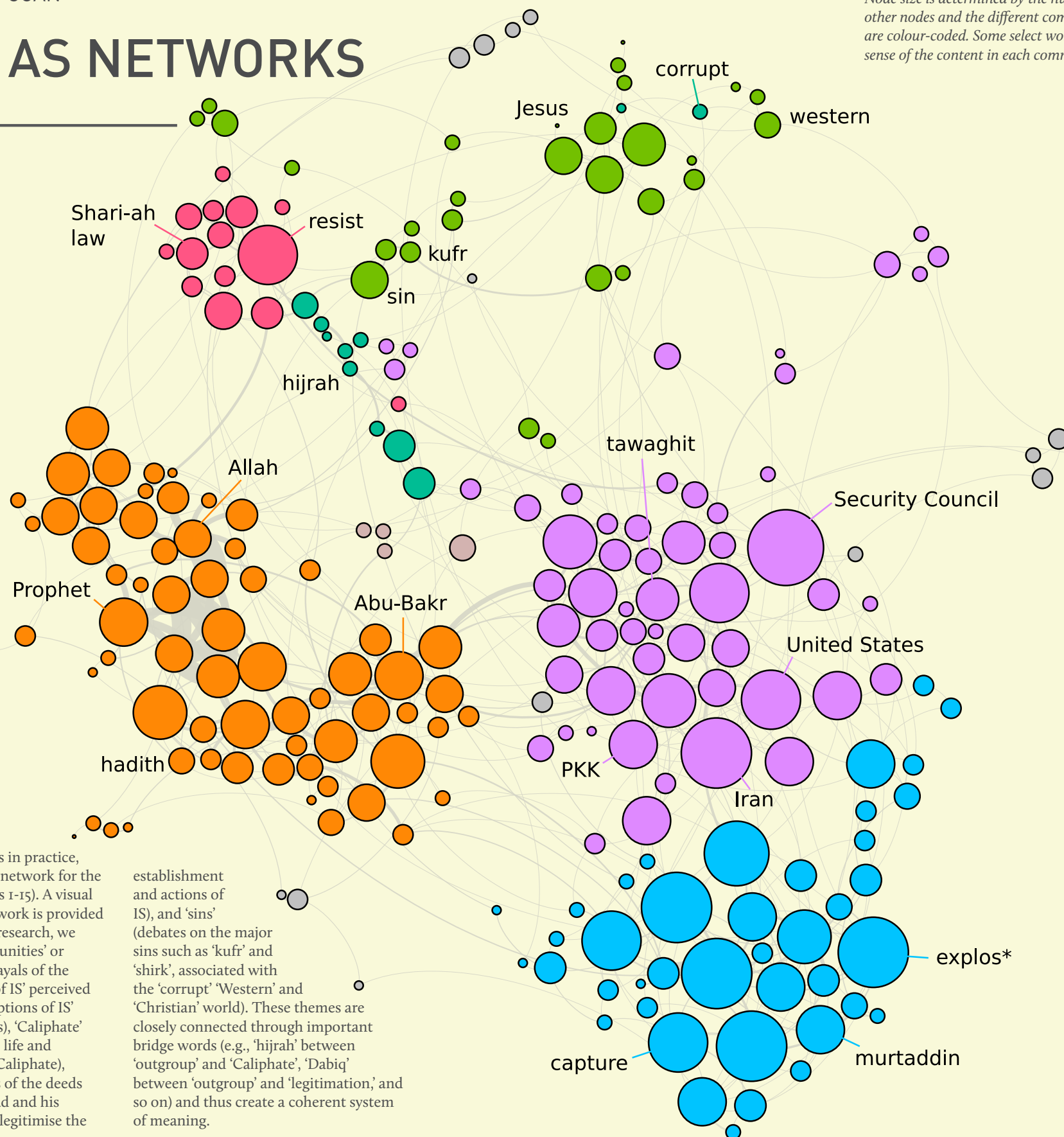

Figure 1: Co-occurrence network of Dabiq magazine, issues 1-15. Node size is determined by the number of connections it has with other nodes and the different communities within the network are colour-coded. Some select words are labelled to give a general sense of the content in each community.

Communities
- Outgroup
- Jihad
- Legitimation of IS
- Caliphate
- Sin
- Hijrah
- Other

Moreover, while Figure 1 provides a high-level representation of Islamic State communication via Dabiq, one may easily focus on important sub-networks to reveal a more nuanced look at salient topics and themes.

Overall, our research suggests that text network analysis provides a useful complement to existing qualitative and quantitative approaches to analyse extremist communication. These methods are scalable in the sense that they allow the analyst to study large bodies of text in a fast and reliable way. These techniques are also portable in the sense that they may be applied to any corpus (e.g., far-right prose on the internet) with only minor modifications and adaptations.

For security practitioners, network approaches to word co-occurrence provide a quick, easily deployed way to analyse and visualise the underlying structure of content produced by groups that pose a threat to national security. For instance, if a brand new group emerges as a 'splinter' from a previous violent extremist group, then the two word co-occurrence networks can be compared in order to identify points of difference, and from there the implications for threat assessment.

Stephane Baele, Katharine Boyd and Travis Coan are based at the University of Exeter and work on a CREST-funded project exploring Islamic State media.