

Studies in Conflict & Terrorism

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/uter20

Socio-Semantic Network Analysis for Extremist and Terrorist Online Ecosystems

Stephane J. Baele & Lewys Brace

To cite this article: Stephane J. Baele & Lewys Brace (24 Mar 2025): Socio-Semantic Network Analysis for Extremist and Terrorist Online Ecosystems, Studies in Conflict & Terrorism, DOI: 10.1080/1057610X.2025.2480622

To link to this article: https://doi.org/10.1080/1057610X.2025.2480622

© 2025 The Author(s). Published with license by Taylor & Francis Group, LLC

View supplementary material

Published online: 24 Mar 2025.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at https://www.tandfonline.com/action/journalInformation?journalCode=uter20

STUDIES IN CONFLICT & TERRORISM
https://doi.org/10.1080/1057610X.2025.2480622

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Socio-Semantic Network Analysis for Extremist and Terrorist Online Ecosystems

Stephane J. Baele[a] and Lewys Brace[b]

[a]ISPOLE Institute, University of Louvain, Belgium; [b]Centre for Computational Social Sciences, University of Exeter, UK

ABSTRACT

How to best chart and analyse extremist digital ecosystems? This paper proposes to complement standard mapping methods based on URL outlinks, which are typically deployed but present several critical flaws, with socio-semantic network analysis. Adapting bi-nodal socio-semantic network principles to the specificities of extremist and terrorist digital communications, we put forward a simple yet efficient method for generating informative networks based on ideological or thematic proximity rather than URL connections. We use three datasets of various sizes and nature (French far-right websites, British far-right websites, US far-right Telegram channels) to compare traditional URL-based vs. socio-semantic networks, demonstrating how the latter bypasses the flaws of the former and offers significant advantages in multi-methods research designs. This empirical validation of our methodological proposition unfolds several new observations concerning the contemporary Western far-right digital ecosystem, highlighting specificities and commonalities of its French, British, and American variants.

ARTICLE HISTORY
Received 10 December 2024
Accepted 12 March 2025

Mapping Extremist Online Ecosystems

The past three decades have witnessed the relentless expansion of online extremism, growing from the confidential bulletin board systems of the mid-1980s to the vast and sprawling digital environments of today.[1] Societies now face an information and communication landscape contaminated by hundreds of thousands, if not millions, of extremist digital spaces of all kinds (social media accounts, websites, wiki structures, e-commerce, dating platforms, image-boards, private chatrooms, etc.) representing multiple radical ideologies and subcultures. Some are piloted by terrorist organizations or influencers promoting radical worldviews and/or political violence, others by states or state proxies; some parts grow organically, others are strategically established.[2] This severe challenge, which has multiple ramifications (e.g. exacerbation of political polarization and radicalization, emergence of "stochastic" terrorism,[3] electoral destabilization

CONTACT Lewys Brace l.brace@exeter.ac.uk Centre for Computational Social Sciences, University of Exeter, UK.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/1057610X.2025.2480622.
© 2025 The Author(s). Published with license by Taylor & Francis Group, LLC
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

and democratic backsliding, fuelling of geopolitical tensions), has become a field of scientific study in its own right, where one of the baseline priorities has always been to detect as many relevant digital spaces as possible while simultaneously assessing their respective importance and identifying their connections.

As such, network analysis (combined with preliminary web-crawling and scraping)[4] has proven to be a key tool allowing for the mapping of complex and fast-moving dynamic extremist ecosystems on the internet and social media. A standard network graph displays individual online spaces (e.g. Facebook accounts, Telegram channels, forums, extremist blogs) as *nodes*, as well as the links (called *edges*) connecting them. Invariably, these edges represent the URL outlinks connecting the sites: an directional edge is drawn between the nodes symbolizing online spaces A and B if A contains a clickable link to B. Typically, the thickness of an edge depends on the absolute number of corresponding outlinks shared between two spaces (thick edges signifying multiple links from A to B), and the size of a node is relative to the online space's importance (usually measured using a proxy measure of network influence such as in-degree, eigenvector centrality, or PageRank). Link thickness and node size are intuitive visual markers allowing for quick identification of recurring connections and influential sites. The structure of the network – its overall layout and the "communities" one can detect in it – therefore depends on how many URL outlinks are found and which nodes they connect together, which in turn feeds a display algorithm (e.g. ForceAtlas, OpenOrd).

This straightforward methodology has several advantages and has proven to be useful for mapping small to very large extremist digital ecosystems and communication channels, anchoring the growing scientific understanding of the phenomenon on solid bases, as well as empowering Counter Violent Extremism (CVE) and Counter Terrorism (CT) stakeholders with pertinent knowledge of the reality they confront. Stemming from pioneering works like Caiani and Parenti's mapping of Italian far-right websites,[5] outlinks-based network analysis has, for instance, allowed scholars to chart multiple "biotopes"[6] constitutive of the now immense far-right digital ecosystem, and sometimes to compare their structures.[7] For example, Lee offered a presentation of the small counter-jihad blogs "nebula" (which proved to be Anders Breivik's main influence ahead of his 2011 terrorist attack),[8] while more recent studies of "alt-tech" social media platforms like Gab[9] or Telegram[10] plot tens of thousands of digital spaces, sometimes in dynamic chronological ways allowing for time-series analysis.[11] Other ideologies have similarly been explored, and their structures and key actors revealed; for instance, Brace and colleagues have mapped the "incelosphere" and its connections with neighbouring online milieux,[12] and Berger and Morgan uncovered the Islamic State's multidimensional propaganda strategy at its peak.[13] A key advantage of this method is that it provides a basis for snowballing, enabling researchers to discover links, among URLs collected in known online spaces, to previously unknown spaces that ought to be included in the analysis. This "discovery" purpose is particularly important for specific tasks like measuring transnational ties between national extremist ecosystems, as done for example by Törnberg and Nissen.[14] In sum, mapping radical digital ecosystems using URL outlinks has by now become a standard exercise in extremist/terrorism research (and media studies more broadly), which is an excellent scholarly development given the scale of the problem. The method is remarkably efficient: it

rests on straightforward techniques yet delivers valuable, intuitively understandable information.

However, this method also presents several shortcomings, which at times prove to be critical. Three of them stand out. First, an outlinks-based network is not a direct representation of the ecosystem under scrutiny: URL outlinks indeed provide only indirect proxies of the attributes and dynamics *that really matter*. Besides ideological alignment and partnership, outlinks from A to B can indeed have widely different meanings – when they actually mean anything at all.[15] They can represent news information sharing, pointers to "evidence" in conspiratorial narratives, doxing and harassment attempts, or even ideological opposition. As such, while ideological community detection typically features in extremist network maps, their inference from URL links is imperfect and regularly ensues in misleading maps and statistics, especially in overly bloated networks containing myriads of irrelevant digital spaces.

Second, and related to the first consideration, the deployment of automated crawling and scraping techniques in collecting network-relevant data from today's vast online ecosystems indeed inevitably produces large quantities of "borderline" digital spaces such as "alternative" news sites, political movements, and parties situated on the fence between extremism and the "mainstream" (to take the far-right example: *Breitbart* in the US, *Spiked* in the UK, *Valeurs Actuelles* in France). This situation, both a symptom and a cause of the increasingly porous nature of the boundary between extremist and mainstream media,[16] begs the question of where exactly to prune the network to keep only nodes and edges deemed relevant: what criteria ought to be chosen to determine whether site X or Y is "extremist" or not, "relevant" or not, and how to implement such criteria in an impartial way, especially on big databases uncombable by human beings? These questions – which URL-based networks cannot, alone, solve – have only recently begun to receive satisfying answers thanks to AI classifiers,[17] but these models themselves come with their own quandaries and have to be run in addition to network analysis; they are, what's more, yet to be implemented in non-testing, genuinely substantive empirical studies.

Third and finally, when building a URL-based network from a starting list of relevant online spaces, it is frequent that some of them are not connected to others *via* outlinks. In such cases, the nodes representing them will be detached and drift away from the network, with no information to be gathered about them. However tempting it might be to manually remove them, these online spaces are often stand-alone sites hosting the most extremist content. Again, generating a network resting on something else than outlinks has the potential to keep and present all information contained in the dataset. We therefore contend that a new way of generating networks representing extremist and terrorist digital ecosystems is needed that overcomes these three weaknesses in one go, and which therefore does not rely on outlinks.

In this paper, we propose such a method. We import and adapt to extremist digital ecosystems a simple yet highly useful method for generating more meaningful networks based directly on ideological proximity, offering clear ideological distance metrics on which to base pruning criteria and community detection, and integrating all online spaces including those without formal URL connections. Initially designed to study bibliometrics, this method will, we hope, reinforce the toolkit of digital extremism and terrorism research as well as neighbouring fields in communication and social

media studies, such as disinformation campaigns, polarized information systems, or foreign influence in national media landscapes (e.g. Russian influence in the US, Europe, or the Sahel). We do not claim that this approach is particularly sophisticated. To the contrary, just like the classic outlinks-based network analysis, we see its simplicity from an Occam's razor perspective as an asset allowing for highly efficient research. Similarly, we do not suggest that it should become the go-to method, replacing URL-based networks; rather, we add a new tool to be used, when pertinent, alongside or in combination with existing ones (we come back to this idea in the conclusion).

We proceed in two steps. First, we explain the basic principles of the method, which operationalizes a bi-partite (bi-nodal, that is, with two types of nodes) network[18] best described as a *social-semantic* one (i.e. integrating both social and semantic information), and we detail the steps of a typical workflow. Second, we demonstrate the usefulness of this method by applying one variant to three different far-right digital ecosystems: British and French far-right websites and US Telegram channels. For each, we also construct outlinks-based networks to allow for comparison, and we highlight a series of findings on the structure and ideological fabric of these ecosystems. A brief conclusion summarizes the key points from these two sections, including a comparative perspective on the empirical cases, and opens up avenues for further methodological and empirical research.

Creating Social-Semantic Networks of Extremist and Terrorist Digital Ecosystems

Principles

The idea of generating a network based on, and conveying information about, the linguistic content of a corpus is not new. Networks featuring words as nodes and edges representing frequency of co-occurrence have usefully been built across the social sciences to analyse major themes, worldviews, and narratives in different kinds of texts (i.e. speeches, literary works, traditional and social media, etc.),[19] including for the comparison of different parts of bigger composite corpora.[20] Text-analysis/NLP (Natural Language Processing) coding packages and software now routinely integrate the constitution of the co-occurrence matrixes[21] (typically at sentence or paragraph levels) underpinning these semantic networks. Seeking to ground their analyses on other linguistic metrics than co-occurrence statistics, some scholars have built semantic networks on the results of diverse NLP tools.[22] Occasionally, semantic networks have fruitfully been implemented in extremism and terrorism studies, using co-occurrence or other linguistic metrics; for example, Baele and colleagues used syntactic parsing to generate the semantic networks of neo-Nazi and ISIS magazines,[23] providing an easily understandable overview of the main themes and claims covered in these publications.

Here, we build on this literature to present a method that helps comprehend, in a single visualization and related set of statistics, the relationships between the spaces/participants of a given extremist digital ecosystem on the basis of their respective ideological inclinations. More specifically, the aim is to construct "binodal" networks whereby two types of nodes are plotted together on a vector field representing semantic

closeness/distance: one type of node represents the communicating actors (e.g. social media accounts, websites, posters of a forum), the other one the linguistically expressed topics or themes deemed relevant (e.g. major ideologies, narratives). In such a field, the position of each actor's node depends on its semantic similarity to the various semantic nodes (not on URL connections), meaning that its location is also relative to other actors' own positions vis-à-vis these linguistic nodes. As a result, semantically similar actors are pulled closely together in the network, and distanced from semantically different ones. For example, in our empirical illustrations below, where semantic nodes represent the main sub-ideologies of the "far-right", online spaces sharing a common anti-Semitic language cluster together, while Islamophobic ones are grouped together further in the network; online spaces combining anti-Semitism and Islamophobia are then situated in the area between the two semantic nodes.

The technical basis for this approach is not new: we build here on a binodal "topic network" methodology described in the field of bibliometrics and scientometrics by Lee, Jung, and Song, who wanted to highlight in a single representation scholars' privileged methods and objects of analysis in Communication and Media Studies.[24] This particular study was itself inspired by earlier attempts to create meaningful "social-semantic networks", that is, graphs based on both "a social network and a socio-semantic network, i.e. an epistemic network featuring agents, concepts and links between agents and between agents and concepts",[25] thereby offering "a socio-semantic perspective for the joint study of social networks and discourse similarity".[26] Variations of such "social-semantic networks", including adaptations of Lee, Jung and Song's model, have already been successfully applied to study the different topics covered in the speeches delivered by various political actors at the UN,[27] demonstrating the pertinence and efficiency of the tool for political science and security studies. In the following paragraphs, we further develop this line of research by offering an adapted version for the study of extremist and terrorist digital ecosystems.

Worflow

We tailor this approach and algorithms to the study of extremist and terrorist digital networks by proposing the following, five-steps workflow (shown in Figure 1

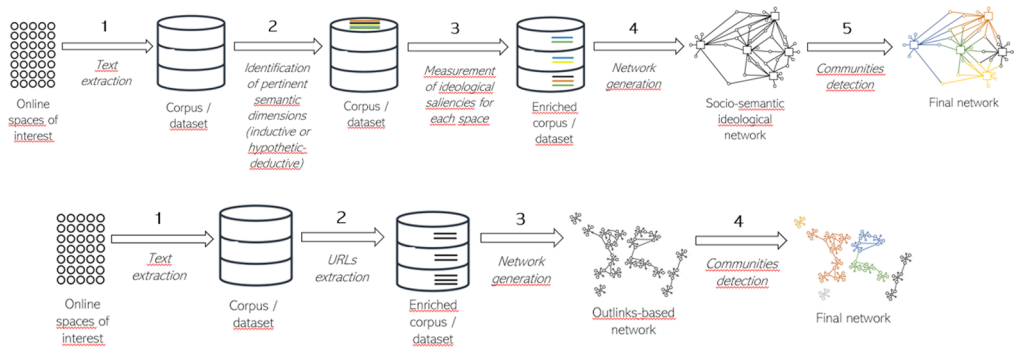


Figure 1. Workflow for creating socio-semantic networks of extremist and terrorist digital spaces (above), compared to workflow for creating standard URL-based networks (below).

– including comparison with the typical workflow behind URL-based networks). We offer in Appendix 1 the full Python code of the computational steps involved in this workflow; the code is carefully annotated to empower researchers with lower computational skills levels to implement it (supplementary material).

First, digital spaces of interest are identified, and their content is collected in a corpus/dataset using standard crawling and scraping tools. Our illustrative cases investigated below use three corpora containing the text content of 107 French (corpus 1) and 99 British (corpus 2) websites identified as "far-right" by all scientific articles mapping these spaces (finite corpus), and a corpus of 425 US far-right Telegram channels. The text content is collected with meta-data in a well-ordered dataset. Because we offer a comparison between URL-based and socio-semantic networks, we also extracted all URLs from each of these corpora, yielding 11,843 additional US Telegram channels, 7,271 French websites, and 1,476 British websites.[28]

Second, after standard corpus cleaning, relevant semantic dimensions are established. Depending on the research design, this can be done in an inductive way with a topic model or similar tool, or hypothetically-deductively using dictionary/lexicon-types of tools containing theoretically relevant terms (a range of different procedures are available here, such as selecting specific words from the corpus frequency list and assigning them to semantic categories they characterize). Tools like seeded topic models, AI classifiers, or word embedding models exist that combine inductive and deductive logics. There is no inherently "right" or "wrong" choice when opting for one of these logics and tools: each research question and case brings about its own most pertinent one. While early developers of socio-semantic networks used inductive LDA topic models,[29] we opted below to analyse our illustrative cases with a deductive approach geared at positioning each online space according to the main far-right (sub-)ideologies usually mentioned in the literature; following an earlier idea by Froio in her network analysis of the French far-right websites,[30] we approximated these ideologies by looking at which out-group is vilified in language: anti-Semitism, Islamophobia, anti- left-wing liberalism, anti-LGBTQ, and anti-black and people of color. More precisely, from each corpus (French websites, British websites, US Telegram), we selected from the frequency list all words labelling these groups (often in derogatory or dehumanizing ways, like "kike" for Jews or the n-word for black individuals) and words unambiguously representative of these sub-ideologies (e.g. "Zionist Occupied Government" or "ZOG", for anti-Semitism), to construct five distinct and non-overlapping ideological dictionaries/lexicons[31] which are provided in Appendix 1, clearly highlighted within the code lines (supplementary material). As such, our method takes up Froio's idea to blend network analysis with ideological content analysis, but departs from it in a critical way: while she built an URL-based network first and then qualitatively assigned a dominant ideology to each site, we directly base network connections and distances on ideological proximity.

Third, the score of each online space is computed for each semantic dimension: the aim is to have a metric, *for every participant* of the corpus, expressing the saliency *of each semantic dimension*. Again, different techniques exist to do so; for our illustrative cases, we calculated the saliency (expressed as the ratio of dictionary words per 1,000 words) of each ideological dictionary for each website (France and Britain) and each Telegram Channel (US). The same step would have been applied to topics

found by the unsupervised or seeded topic models, should a more inductive approach have been preferred. In cases where researchers seek to use the method as a basis to prune irrelevant content from large, snowballed, networks, digital spaces obtaining null or very small scores on all the semantic dimensions would already be dismissed at this stage: offering a transparent and bias-free basis for pruning is a key advantage of this workflow.

Fourth, this data (all semantic scores for all online spaces) are treated as an edge-list matrix dataset for the constitution of a network, whereby each row of the data set specifies a semantic node – site node dyad, along with the *weight* of the edge between them; this edge weight representing the saliency of the semantic dimension within the corpus of the site. This dataset serves as the basis for the generation of the bi-nodal network which features all selected online spaces as one type of node and the five semantic dimensions (in our case, the five far-right sub-ideologies) as the other type of nodes. An algorithm is chosen to plot nodes representing online spaces more or less distantly from ideological nodes (and thus from each other) according to their scores on the dictionaries representing these ideological nodes. For example, some British websites featuring in the next section cluster closely around the anti-Semitism ideological node even though some have no URL outlinks connecting them, because they share similar levels of anti-Semitic language. This technique solves a problem Froio acknowledged in her attempt to enrich network analysis with ideological identification: that of having the majority of the far-right sites she studied castigating multiple outgroups and thus demonstrating different sub-ideologies at varying levels.[32] Our method acknowledges that many websites inevitably contain, at least to a certain degree, several sub-ideologies: the position of an online space on the graph depends on the *relative intensity* found, for each sub-ideology, in its content.

Fifth, basic network statistics are computed (e.g. degree, centrality, modularity, density), and on that basis communities are identified with the most pertinent algorithm to highlight the most semantically coherent clusters. These communities will not necessarily neatly correspond to groupings around the semantic nodes: this will only be the case if the technique for measuring the chosen semantic dimension is highly discriminatory when it comes to categorizing online space nodes (which in turn depends on both the choice of semantic dimension and the features of the online ecosystem). In our illustrative examples below, we use modularity classes and the Louvain algorithm[33] for identifying far-right ideological clusters. Other options are available, again depending on the specific research questions. For large, typically snowballed networks, a second stage of pruning could be implemented here, whereby far-off sites belonging to no community could be qualitatively examined and perhaps pruned out.

To summarize, these five steps produce a socio-semantic network that positions actors according to their semantic profiles (using a pertinent linguistic marker, in our case below ideological content approximated by dictionary scores) and identifies communities based on semantic similarity (in our case, ideological resemblance). One can see how such an approach alleviates the three main shortcomings of outlinks-based network analysis of extremist digital ecosystems: communities are directly and more pertinently identified, less relevant online spaces appear more clearly and can be pruned out in one go on a more objective basis, and online spaces that have no formal URL

connection to others feature in the network in an informative way. Moreover, this method also presents distinct advantages. First, it unlocks the possibility of visualizing the ideological evolution of a given extremist ecosystems across time (hence answering questions like "is anti-Semitism increasing?" in a rigorous and falsifiable way). Second, it can serve as a tool for comparative studies of digital ecosystems, evidencing differences in ideological fabric or other semantic dimensions; when analysing our three empirical cases below, we offer comparative insights that illustrate this advantage. Third, this type of network can be constructed for digital spaces whose technical features prevent the collection of some or all outlinks. For example, alt-tech social media Gab – a key hub for, among others, neo-Nazis – took the decision to only allow paying "pro" users to see accounts' followers and connections, making traditional network data collection much harder. Similarly, Facebook offers the possibility to set up groups without listing their members – an affordance consistently used by radical anti-Western Sahelian groups backed by Russia.

Illustrative Cases: French and British Far-Right Websites, US Extreme-Right Telegram

French Websites

We first examine the French case, which renders a network of 107 nodes. Using the classic outlinks-based method, we obtain Figure 2; whereby only 56 websites appear together in a URL network, 51 having no URL ties and are thus displayed randomly

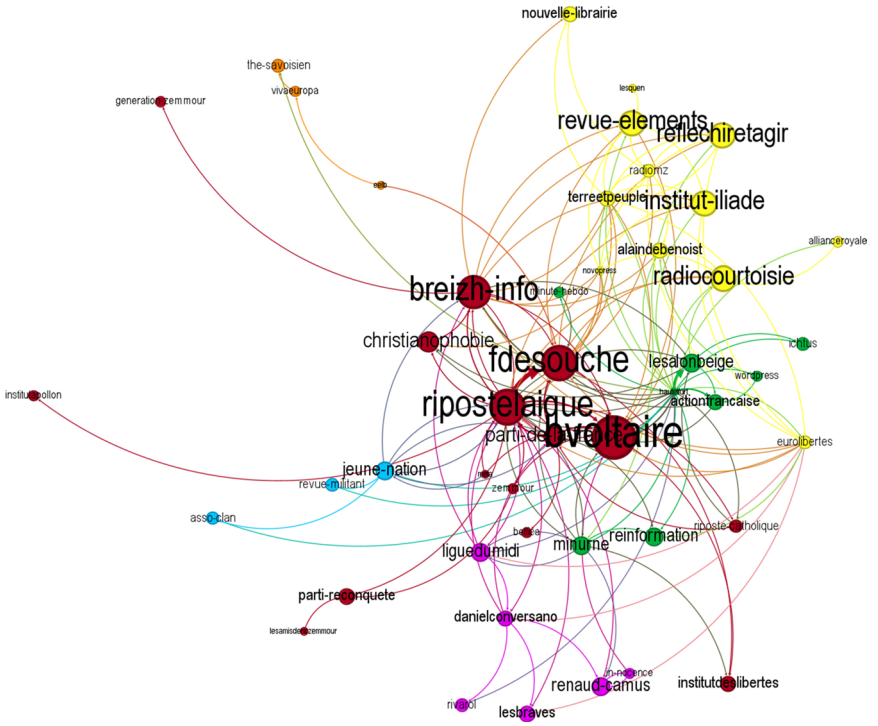


Figure 2. French websites URL network. Node size proportional to in-degree score.

far in the peripheries of the graph (and therefore not displayed below). The dataset contains well-known websites like *FdeSouche* or *Riposte Laïque*, whose influence is correctly approximated by the larger amount of incoming edges they receive (in network vocabulary, this is called "in-degree", which is used here to determine node size in our display). Far-right politician Eric *Zemmour*, his *Parti Reconquête*, and the informal support group *Les Amis d'Eric Zemmour* are unsurprisingly linked together, as are extremist influencer Daniel *Conversano* and his offshoot organization *Les Braves*. The classic URL method therefore reveals some useful information, including strategically set up webs of interconnected online spaces. However, the network is quite dense and its communities – regardless of the chosen community-detection algorithm – weakly differentiated are hardly meaningful. One reason for this is the high number of (sometimes important, from an expert point of view) disconnected websites that do not belong to any community and for whom information is lost. In sum, the network is hardly useful for expert researchers and practitioners. One understands Froio's abovementioned decision to replace, in her network of the French far-right comprising many of the websites included here, algorithm-detected communities by qualitative assignments of sites' dominant ideology.

While the small, seed list-only network is not very informative, URL-based methods have the advantage of snowball-detecting URLs to online spaces beyond the initial dataset. In this case, we were able to generate an expanded URL-based network reaching no less than 7.271 nodes (and 7.326 edges connecting the initial websites to themselves and the newly found online spaces). Because this graph is secondary to the direct comparison of URL-based versus socio-semantic methods, we display it outside the manuscript in Appendix 2 (figure A, supplementary material). On the one hand, this expanded version offered rich information; chiefly, it now includes online spaces of critical importance in the French context yet surprisingly not contained in the initial list, such as *Génération Identitaire*, *Réfléchir et Agir*, *Synthèse Nationale*, or Alain Soral's *Egalité et Reconciliation* and *KontreKulture* (the graph reveals their high in-degree scores, rightly approximating their genuine importance). The URL-based method therefore fulfils what we consider to be its main purpose: the identification of new relevant digital spaces and a fair approximation of their relative influence. As such, this technique remains important to verify the comprehensiveness of available datasets, especially in online environments as quickly evolving as extremist ones (where organizations are prescribed, change their names, move to other web domains, splinter, etc.). On the other hand, however, the network also presents all the problems identified above. First, it is polluted by a large amount of non-far-right websites such as mainstream news sites (e.g. *Le Figaro*, *Le Parisien*, *Le Monde*) and official government websites (e.g. *Légifrance*, *Europarl*), meaning that an extensive pruning – necessarily carried out manually if no objective criteria allows for automated filtering, as detailed below – is necessary before being able to truthfully present it as "the French far-right digital ecosystem". As a result, and just like with the small seed-list network, the communities identified by the algorithms have very limited relevance, if any, in spite of their deceptively "nice" appearance in the graph. For instance, three websites run by pagan white-supremacist Daniel Conversano are grouped in the blue-coloured community alongside a fundamentalist Christian movement (*Riposte Catholique*) and *Bruno Gollnisch* (a prominent Rassemblement National/

Front National politician), as well as with news sites like *Forbes* and the *Daily Telegraph*. Conversely, multiple sites ideologically aligned with Conversano's websites are spread in other communities.

Figure 3 shows our socio-semantic network of the initial dataset.[34] Here, only a handful of websites are missing – and only because their text content was too small for the dictionaries to run (or because, in one case, most of the text was actually in English). The "disappearance" of sometimes important online spaces is avoided. In contrast to the outlinks-based method, this new approach also produces a clear and truthful representation of the French far-right online ecosystem directly based on its ideological fabric. Examining the underlying data of dictionary scores as well as the distances in the graph, one can indeed appreciate how accurate the ideological positioning is for each website. For instance, *Riposte Catholique* appears somewhere midway between the anti-LGBT and Islamophobia ideological nodes, with a lower anti-left elites score pulling it towards that node – this corresponds to the hardline catholic site's major ideological lines. *Les Localistes* only articulates an anti-Immigrants discourse, totally devoid of references to Islam and Muslims, Jews, LGBT, and the left-liberal elites; it is thus only connected to the anti-Immigrants node, far from all other ones – again, this fits the ideational leaning of the platform. Interestingly, Eric Zemmour's *Parti Reconquête*

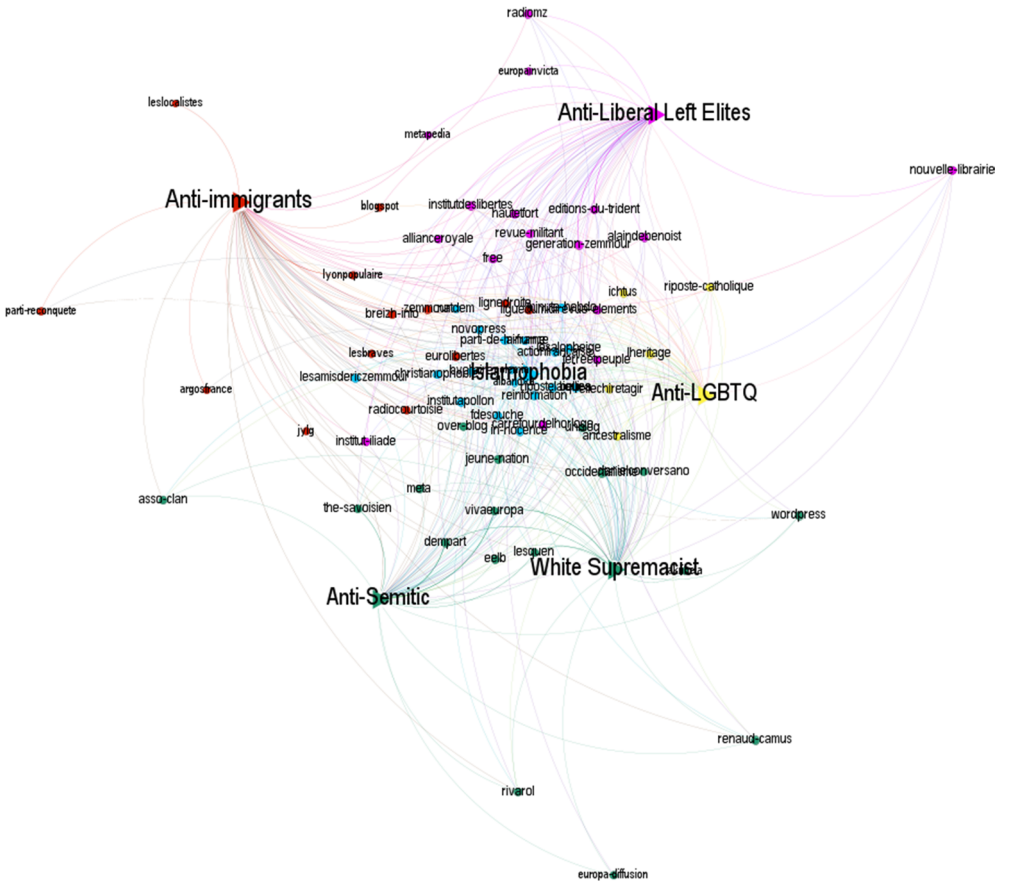


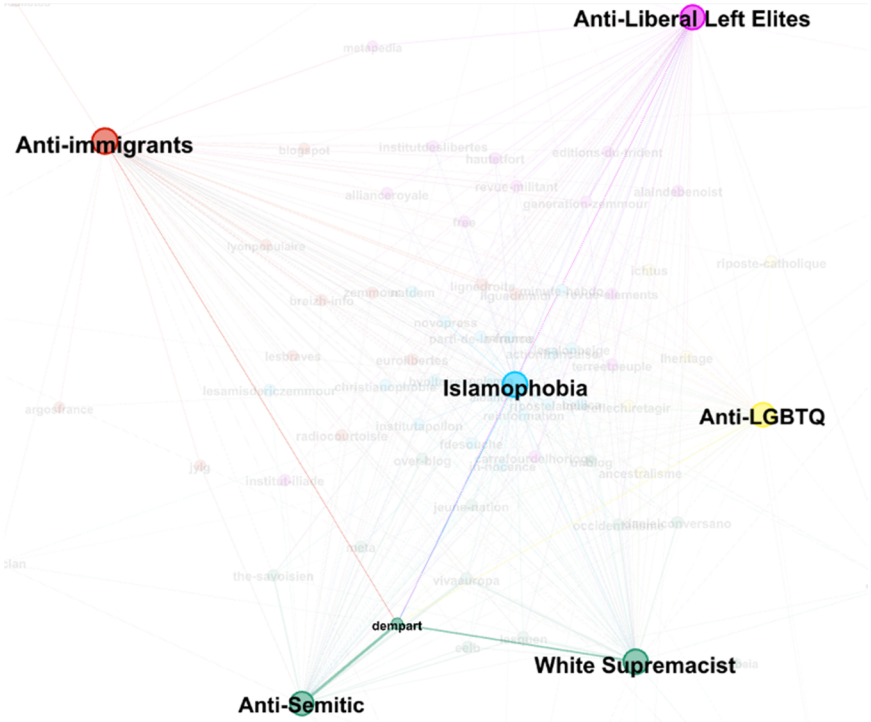
Figure 3. French websites socio-semantic network.

scores low on anti-Immigrants and Islamophobia, which is intuitively surprising; however, his informal supporter groups (e.g. *Les Amis d'Eric Zemmour*) contain more extreme language that anchors them firmly around the centre of the network, indicating that the official website proposes a more polishes, acceptable discourse while non-official sites more accurately reveal the underlying discourse. In this graph, meaningful communities appear more clearly, and closely reflect the sub-ideologies at the exception of anti-Semitic and white supremacists which are mathematically undistinguishable (red = anti-immigrants, pink = anti-liberal left elites, blue = Islamophobia, yellow = anti-LGBTQ, and green = anti-semitic and white supremacist).

Looking at the whole ecosystem level, two major findings appear. First, the French far-right ecosystem of websites is centred around Islamophobia. This Islamophobic core is due to a series of websites having extremely high Islamophobia scores demonstrating an intense anti-Muslims language (e.g. *Alba Nova* at 0.85, *Riposte Laïque* at 0.55, or *Réinformation* at 0.5), a broader range of websites presenting a more eclectic profile around repeated Islamophobic narratives (e.g. *Novopress*, *Parti de la France*), and recurrent Islamophobic themes found in the majority of websites presenting another primary ideological profile (e.g. *Démocratie Participative*, cf. below). This is a distinct characteristic of the French case, not found as we will see in the British and American ones, that has already been stressed in the existing literature. For example, Benveniste and Pingaud observed that Islamophobia has risen in the French far-right and became a dominant "new way of observing French society through the lens of Islamic issues" thanks to digital activists' links to foreign digital milieux such as the counter-Jihad nebula evoked above that is being built in Europe.[35] Similarly, Froio emphasised the centrality of anti-Islam in the French "faschosphere", detailing how it comes with different frames (e.g. from a hardline catholic stance, from a racial perspective, etc.),[36] and showed in another article that Muslims are identified as an "other" in almost all French far-right websites, while Jews and LGBTQs are only "othered" in a minority of them.[37]

Second, anti-Semitism and white supremacism appear very close together – as evoked above, the community-detection algorithm actually groups them within a single cluster because of a handful of websites combining extreme anti-Semitism and white supremacism (*Démocratie Participative* [a.k.a. *DemPart*], its associated forum *Europe Ecologie Les Bruns* [a.k.a. *EELB*], *Viva Europa*, and *Lesquen*). These findings complement previous, URL-based studies of segments of the French far-right digital ecosystem, such as Macdonald and colleagues' snowballing analysis from Rassemblement National's Twitter account,[38] which offer useful information on major nodes of influences within the ecosystem and relationships towards its neighbouring milieux but falls short on offering a meaningful ideological map.

We propose to focus on one of these websites – the infamous *DemPart* – to illustrate in Figure 4 how semantic scores determine position in the network. We see how this online space's predominantly anti-Semitic (score of 0.610, the highest of our corpus), then white supremacist (score of 0.353) language situates it very close to the node representing the former ideology and a bit further to the one signifying the latter, while the lower (yet sill high comparatively to many websites) scores on the other dimensions pull it somewhat higher in the graph.



Ideology node	Edge strength with *DemPart* node
Anti-Semitic	0.610
White supremacist	0.353
Islamophobia	0.117
Anti-Immigrants	0.076
Anti-LGBTQ	0.057
Anti-liberal left elites	0.057

Figure 4. Situation of *Démocratie Participative* (*DemPart*) in the French network.

British Websites

We now turn to the British network, initially made of the 99 nodes appearing in Figure 5. This network has the same strengths and weaknesses as the French one: while it does offer some insights on established connections (such as the one between *Patriotic Alternative* and *Radio Albion*), its communities are not meaningful, a series of sites (some of them included in the figure, most are not) are disconnected from the network and thus offer no information, and websites receive a fictitious central/peripheral role (such as conspiracy theorist *David Icke*, or the borderline information site *Spiked*).

If we expand this network by displaying all online spaces outlinked from this initial list, thousands of other websites and social media channels appear (reaching 1,476 nodes and 1,813 edges), offering much richer information (the graph is located in Appendix 2 – figure B, supplementary material). On the one hand, this new graph brings in important British or English-speaking far-right sites surprisingly not initially

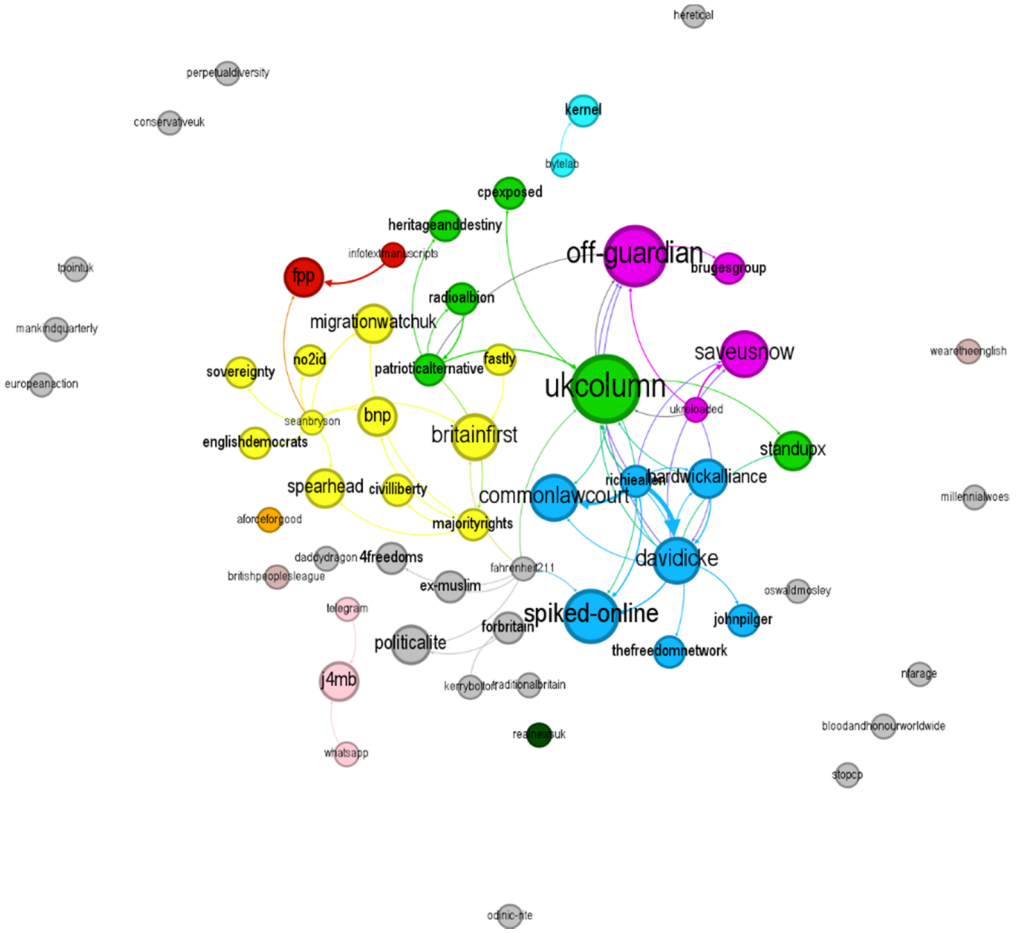


Figure 5. British websites URL network. Node size proportional to in-degree score.

listed,[39] such as the *English Defence League*, *Britain First*, the *Occidental Observer*, *Jihad Watch*, or *Stormfront*. On the other hand, it also includes large amounts of non-pertinent sites, including many national or local news outlet (e.g. *BBC*, *ABC News*, *ITV*), some of them even with left-leaning affinities (e.g. the *Guardian*, *The Atlantic*), as well as International Organizations (IOs) and NGOs working against extremism (e.g. *Europol*, *Hope Not Hate*, *ISD Global*). Displaying such an unfiltered network, as we do in the Appendix (supplementary material), is therefore fundamentally misleading even if the resulting graph seems informative at first sight (with neatly distinct clusters). It may be tempting to remove all nodes only outlinked to once, with the idea that they would represent less pertinent sites. However, a good amount of these sites represent the darkest corners of the ecosystem and are actually more relevant than those outlinked to a lot (i.e. mainstream news outlets). Moreover, the network contains borderline online spaces such as right-wing reactionary or contrarian-libertarian commentary outlets (like *Quillette*) and conspiracy theory sites dedicated to health (e.g. *Virus Myth*, *Coronavirus Yellow Card*, *UK Med Freedom*). These latter, mostly Covid-centred sites appear mostly in the yellow-coloured community (mainly because of extensive outlinking done by conspiracy theorist David

Icke), yet they are mixed with hardline far-right sites such as *Genocide Watch*, and with irrelevant online spaces clearly pulled into the network because of adversarial outlinking (e.g. *Gates Foundation*, *World Health Organization*, etc.). As a result, the detected communities are only, at best, superficially meaningful.

In contrast, the socio-semantic network quite accurately shows coherent ideological clusters. Our dictionaries indeed appear to measure ideological saliency well. For example, the most Islamophobic website appears to be *The Religion of Peace* (an ironic title masking hardline anti-Islam "documentation") and *Ex-Muslim*, and the highest-scoring sites for white supremacism are *Blood and Honour Central* and the site dedicated to Patriotic Alternative's leader Mark Collett's book *The Fall of Western Man*. For the anti-Immigrant ideology, *Migration Watch* arrives first. As a result, meaningful communities appear, such as the white supremacist cluster coalescing the likes of *Patriotic Alternative*, *Radio Albion*, *Heritage & Destiny*, *Candour*, *We Are The English*, or the *British Movement*, or the anti-liberal left elites one adequately containing *Nigel Farage*, *Unite Against Communism*, or *Spiked*. The grafting of this anti-establishment cluster to those more traditionally associated with the far-right ideology, with several sites combining this type of language with, say, white supremacist repertoires, is one more piece of evidence of the "mainstreaming" of far-right narratives and perceptions into the propositions and discourses of (sometimes major) political parties following Brexit.[40] Anti-Semitism seems to occupy a fringe position in the British landscape, with few sites scoring high and most of them having white supremacy as a second (yet more distant) score (e.g. *Heretical*, *European Action*).

This network helps understand the ideological differences between key players in a contemporary British far-right online and offline ecosystem in constant reorganization since the implosion of the BNP and the arrival of new types of movements like the English Defence League (we see in the network *Patriotic Alternative*, the *National Front*, *Britain First*, the *Reform Party*, *Nigel Farage*, etc.). It also provides a metrics-based reason to prune out websites initially thought to be relevant but actually containing almost zero semantic markers of far-right ideologies (e.g. *Odinic-rite*). Focusing on websites offers a different perspective on the state of the British far-right than looking at social media; Bovet and Grindrod's recent analysis of the UK far-right ecosystem on Telegram, one of the rare studies of the British right-wing extremism digital landscape, indeed produces a rather different picture.[41] It includes a lot of foreign far-right channels, chiefly from the US (many of them included in our own study below) and Russia, which has the advantage of revealing the porosity of national far-right ecosystems to transnational influence,[42] but comes at the price of less meaningful community detection and loss of grip on the British idiosyncrasies; this illustrates both the strengths and the weaknesses associated with unfiltered URL-based approaches (Figure 6).

US Telegram Channels

Finally, we examine the US case. The initial dataset contains 425 nodes connected by 958 edges, yet (once again) the network displayed in Figure 7 does not show them all because most Telegram channels of the initial list do not have outlinks to the group of connected channels. Although many of them have obviously extremist names, their off-network position nonetheless means information regarding them is lost. While

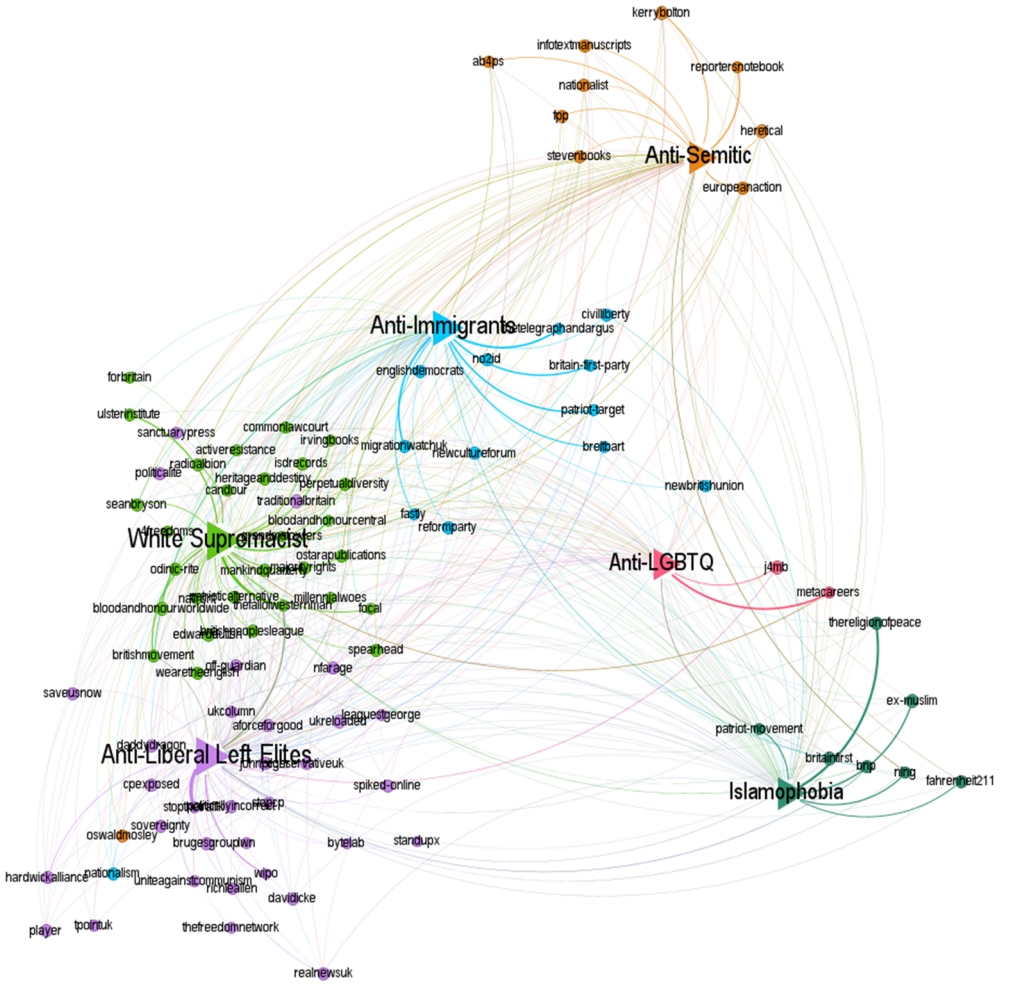


Figure 6. British websites socio-semantic network.

some of the clusters seem coherent (such as the green-coloured community, which groups distinctly white supremacist channels), most are not. Here the problem is that most clusters contain diverse nuances of a dominant MAGA-aligned movement (with variable nuances of racism, conspiratorial ideation, and anti-establishment criticisms) – which is, however, a major finding in itself about the now firmly established symbiosis between Trump supporter groups and preceding extreme right movements. For example, at first sight the orange-coloured community connects together loosely pro-Trump borderline conspiratorial channels like *Red Voice Media*, *Election Evidence*, or *Audit The Vote PA*, but multiple channels proposing very similar narratives (like the "alternative news source" *Breaking 911*) appear far out in the network.

The extended network built on the basis of this initial list (Appendix 2 – figure C, supplementary material), containing no less than 11,843 nodes and 4,339 edges, adds key players in the US far-right landscape; including John Solomon's *Just the News* or *Gateway Pundit*, as well as borderline news sites *NewsMax* and *Breitbart* – and Donald Trump's own channel. The graph also reveals where in the digital realm the US far-right

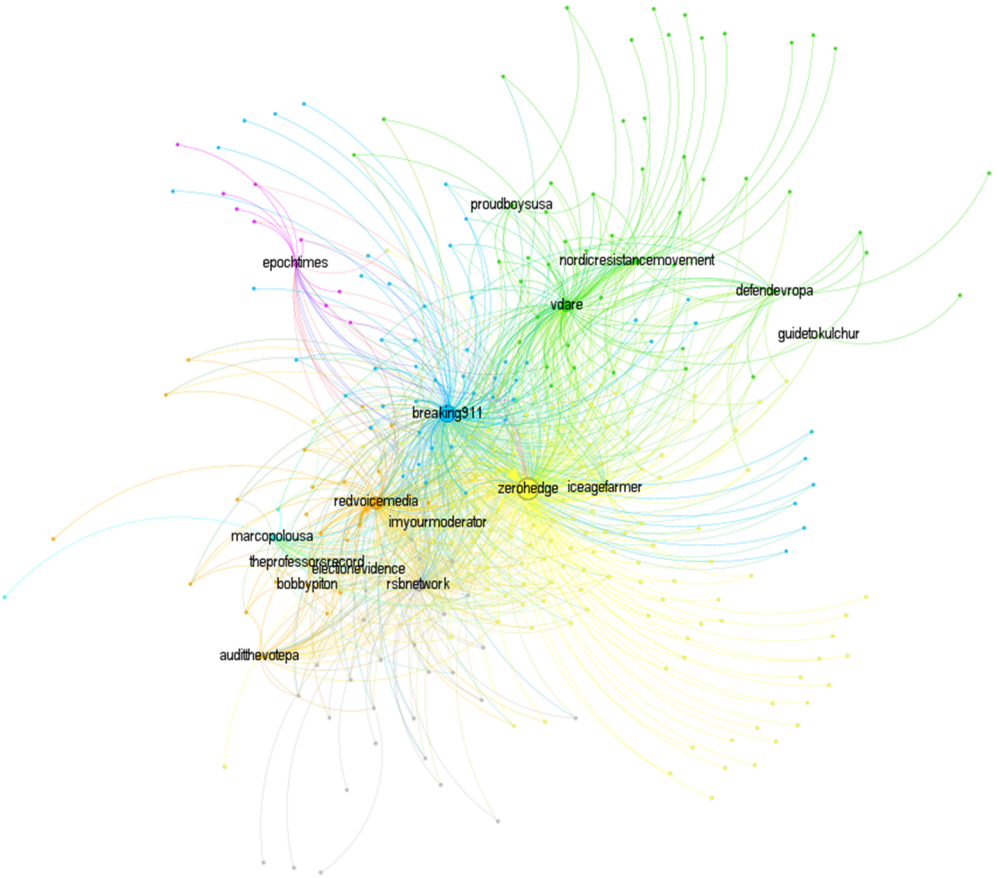


Figure 7. US telegram URL network.

is organized outside Telegram, with multiple accounts from "alt-tech" platforms being linked to (e.g. *Gab*, *Bitchute*, *Odyssee*, or *Truth Social*). Like its French and British counterparts, the network *seems* to be organized around very distinct communities. Some are relatively meaningful; for example, the orange-coloured cluster gathers some of the most brazenly extremist channels (e.g. 4chan, or the longstanding anti-Semitic *Vanguard News Network*). However, again like the French and British graphs, this broader US network is also full of non-pertinent (mainstream and local news channels/sites, government websites) or borderline (e.g. deep conspiracy, alt-right) online spaces, both contributing to prevent a meaningful mapping and identification of communities. The distinct appearance of the clusters in the graph is mostly an illusion.

Implementing the socio-semantic approach yields better insights (see Figure 8), because once more the dictionaries appear to measure ideological saliency quite accurately. For instance, the channels with the highest scores for Anti-Semitism are called *Israel WC2*, *Andrew Joyce Archive*, *Jewish Contributions*, and *Spooky Ghost 88*;[43] the channels having the highest scores for white supremacism are named *White Person Fired*, *The Noticer*, and *White Lives Matter Canada SWO*. These two ideologies appear very close together at the centre of the graph, revealing a core of channels conveying a dual anti-Semitic and white supremacist discourse; on the other hand, islamophobia and anti-immigrant

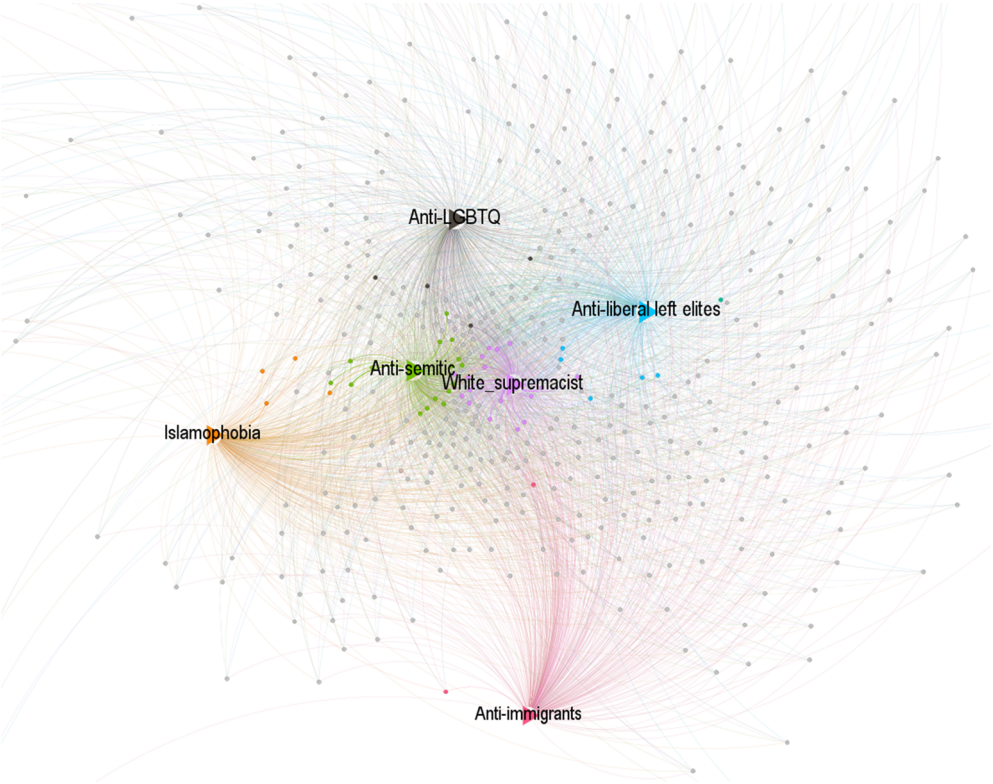


Figure 8. US telegram socio-semantic network.

dictionaries generally get lower scores, and have therefore been pushed to the periphery of the network because of thin edges with most channels. The centrality of white supremacism and anti-Semitism, and their closeness to anti-liberal left elites content, is a key feature of the US landscape, and one that stands in sharp contrast with the French and British cases. Moreover, this socio-semantic network provides a bias-free basis for pruning out irrelevant channels; for example, *DVA Channel* scores 0.0002 on White Supremacism, 0.00005 for anti-LGBTQ, and 0.00003 for Anti liberal left elites – negligible scores pushing it far off the graph (it does not even feature in Figure 8), suggesting that its inclusion in the initial list was not warranted. This case illustrates that implementing socio-semantic networks can serve to correct unwarranted inclusion of online spaces within the "far-right" category by earlier research.

Conclusions and Way Forward

How to best map and analyse extremist digital ecosystems and similar communications? As these environments continue to grow, and as the mechanistic URL-based networks typically used in scholarship present serious limitations, this article proposed a novel approach adapted from existing applications in other fields (bibliometrics, International Relations). We aimed to empower extremism and terrorism studies, and more broadly security studies and research information operations, with an efficient method for generating network maps of digital ecosystems without relying on URL connections. Building

bi-nodal socio-semantic networks bypasses the major weaknesses of URL-based network analysis, displaying online sites according to directly relevant criteria, adequately detecting meaningful communities, offering rigorous data-based grounds for assessing relevance and pruning, and allowing for the mapping of platforms with less transparency on connections. Recognizing the specific usefulness of the URL-based approach as a discovery tool for expanding the empirical horizon – something socio-semantic networks cannot do – we advocate for the systematic combination of these two different ways of generating networks, and underline the importance of qualitative, substantial expertise in the preliminary and interpretation steps of this type of research. Augmenting the methodological toolkit available to study digital extremism is of crucial importance as theoretical developments anchored around ecological notions are booming and call for novel empirical approaches,[44] and more generally participates to recent drives to improve a field still characterized by highly unbalanced, deficient methodological practices.[45]

The three empirical cases explored as illustrative examples of this method demonstrate the benefits of this method: they allowed for the generally accurate identification of distinct (yet interconnected) ideological sub-systems – empirically validating Baele, Brace and Coan's conceptualization of the far-right online ecosystem as constituted by distinct yet porous ideological "biotopes".[46] Edifying national idiosyncrasies appeared from this exploration, with the French network standing out as organized around widespread Islamophobia (confirming earlier work on that national case), the US one containing a central, closely connected white-supremacist – anti-Semitic core, and the British one offering more distinct ideological communities within which the most well-known actors provide anchor and anti-establishment initiatives and leaders graft onto some of the digital milieux more traditionally considered to be "extremist".

It remains to be seen, however, whether this methodology works equally effectively for other extremist ecosystems such as the salafi-jihadist one or the "incelosphere". Each ecosystem is indeed characterized by a highly specific structure and particular practices and dynamics inherited from a genealogy shaped by differing levels and types of state/intelligence disruptions, geographical idiosyncrasies, varying abilities or willingness to take up new technological affordances, and the like. As a result, ecosystems rely to varying degrees on outlinking, are more or less stable across time, and favour certain platforms and services over others, making URL-based and socio-semantic networks more or less useful. We foresee that the far-right online ecosystem might actually be the one where our method is the most insightful, but only further research could precisely answer these questions.

Indeed our methodological contribution and accompanying empirical findings should first and foremost be understood as a starting point from which to think about other innovative ways to generate the most useful networks for particular research projects. A similar method could for example be based on *visual* content instead of text, with communities constituted by the salience, in included online spaces, of particular types of visual tropes (e.g. anti-Semitic memes, neo-Nazi symbols, etc.) identified as "iconographic" (and not semantic any more) nodes; AI object detection models and image-to-text (caption) models now allow for this type of innovative methodologies particularly well-suited for our "pictorial age". Another possibility would be to use linguistic measures geared at identifying markers of non-semantic processes – for example linguistic markers associated with higher endorsement of violence, emotional categories or

intensity, or revealing synthetic (fake AI) content – or even meta-data informing us on dimensions such as geography or time (e.g. newer sites clustering together). As such, the more general argument of this article is that network analysis should not be confined to one set of underlying data and standard practices; the stakes of today's extremist digital communications, and their dramatic impact on societies, are simply too high for not leveraging all available research tools in our hands.

Notes

1. For a comprehensive account of the history of, for instance, online right-wing extremism, read Maura Conway, Ryan Scrivens, and Logan Macnair, "Right-Wing Extremists' Persistent Online Presence: History and Contemporary Trends," *ICCT Policy Briefs* (2019), https://icct.nl/publication/right-wing-extremists-persistent-online-presence-history-andcontemporary-trends/.
2. Stéphane Baele, Lewys Brace, and Travis Coan, "Uncovering the Far-Right Online Ecosystem: An Analytical Framework and Research Agenda," *Studies in Conflict & Terrorism* 46, no.9 (2020): 1599–1623.
3. For a thorough presentation and discussion of "stochastic terrorism", see James Angove, "Stochastic Terrorism: Critical Reflections on an Emerging Concept," *Critical Studies on Terrorism* 17, no.1 (2024): 21–43.
4. That is, the (usually automated) exploration of online spaces connected together by HTML links and collection of their content into a single database/corpus.
5. Manuela Caiani and Lisa Parenti, "The Dark Side of the Web: Italian Right-Wing Extremist Groups and the Internet," *South European Society & Politics* 14, no.3 (2009): 273–294.
6. Baele, Brace, and Coan, "Uncovering the Far-Right Online Ecosystem".
7. As Caiani and Parenti themselves did in a chapter of their seminal book dedicated to the far-right internet: Manuela Caiani and Linda Parenti, *European and American Extreme Right Groups and the Internet* (London: Routledge, 2013).
8. Benjamin Lee, "A Day in the 'Swamp': Understanding Discourse in the Online Counter-Jihad Nebula," *Democracy & Security* 11, no.3 (2015): 248–274.
9. Julian Droogan, Lise Waldek, Brian Ballsun-Stanton, and Jade Hutchinson, *Mapping a Social Media Ecosystem: Outlinking on Gab & Twitter Amongst the Australian Far-right Milieu* (Washington, DC: RESOLVE Network, 2022).
10. Aleksandra Urman and Stefan Katz, "What They Do in the Shadows: Examining the Far-right Networks on Telegram," *Information, Communication & Society* 25, no.7 (2020): 904–923; Alexandre Bovet and Peter Grindrod, "Organization and Evolution of the UK Far-Right Network on Telegram," *Applied Network Science* 7, no.76 (2022): 76.
11. For an excellent example of dynamic URL-network analysis, focusing on German far-right and conspiratorial Telegram channels, done by the *Bundesarbeitsgemeinschaft Gegen Hass im Netz* ("Federal Working Group on Online Hate"), see https://dashboard.bag-gegen-hass.net/.
12. Lewys Brace, Stéphane Baele, and Debbie Ging, "Where do 'Mixed, Unclear, and Unstable' Ideologies Come From? A Data-driven Answer Centred on the Incelosphere," *Journal of Policing, Intelligence & Counter Terrorism* 19, no.2 (2023): 103–124.
13. J.M. Berger and Jonathon Morgan, "The ISIS Twitter Census: Defining and Describing the Population of ISIS Supporters on Twitter," *The Brookings Project on U.S. Relations with the Islamic World Analysis Papers* 20 (2015). See also, more recently, Stuart Macdonald, and Sean McCafferty, "Online Jihadist Propaganda Dissemination Strategies," *VOX-Pol Reports* (2024).
14. Anton Törnberg and Anita Nissen, "Mobilizing Against Islam on Social Media: Hyperlink Networking Among European Far-right Extra-parliamentary Facebook Groups," *Information, Communication & Society* 26, no.15 (2022): 2904–2922.
15. For example, they are often links to a generic ad, or to a local restaurant, or to a sponsored commercial partner.

16. Read for example Teresa Völker and Daniel Saldivia Gonzatti, "Discourse Networks of the Far Right: How Far-Right Actors Become Mainstream in Public Debates," *Political Communication* 41, n.3 (2024): 353–372.
17. For example Mayur Gaikwad, Swati Ahirrao, Shraddha Phansalkar, and Kettan Kotecha, "Online Extremism Detection: A Systematic Literature Review With Emphasis on Datasets, Classification Techniques, Validation Methods, and Tools," *IEEE Access* 9 (2021): 48364–48404; Mohammad Fraiwan, "Identification of Markers and Artificial Intelligence-based Classification of Radical Twitter Data," *Applied Computing & Informatics* (2022), online before print, https://doi.org/10.1108/ACI-12-2021-0326.
18. On this logic see for example Stephen Borgatti and Martin Everett, "Network Analysis of 2-Mode Data," *Social Networks* 19, no.3 (1997): 243–269.
19. For a detailed introduction, methodological discussion, and multiple examples, read Elad Segev, *Semantic Network Analysis in Social Sciences* (Abingdon: Routledge, 2022).
20. See for example Philip Drieger, "Semantic Network Analysis as a Method for Visual Text Analytics," *Procedia – Social and Behavioral Sciences* 79 (2013): 4–17; Stefano Bruzzese, Wasim Ahmed, Simone Blanc, Filippo Brun, "Ecosystem Services: A Social and Semantic Network Analysis of Public Opinion on Twitter," *International Journal of Environmental Research & Public Health* 19, no.22 (2022): 15012.
21. A co-occurrence matrix is a table that shows how often and how closely words appear together in a given text.
22. For instance the "affective" network of Shakespeare's plays, based on sentiment analysis, proposed by Eric Nalisnick and Henry Baird, "Extracting Sentiment Networks from Shakespeare's Plays," *12th International Conference on Document Analysis and Recognition* (2013): 758–762.
23. Stéphane Baele, Katharine Boyd, Travis Coan, and Elahe Naserian, "Lethal Words: An Integrated Model of Violent Extremists' Language," *Studies in Conflict & Terrorism* (2023), online before print, https://doi.org/10.1080/1057610X.2023.2213963.
24. Keeheon Lee, Hyojung Jung, and Min Song, "Subject-method Topic Network Analysis in Communication Studies," *Scientometrics* 109 (2016): 1761–1787.
25. Camille Roth and Jean-Philippe Cointet, "Social and Semantic Coevolution in Knowledge Networks," *Social Networks* 32, no.1 (2010): 16.
26. Johanne Saint-Charles and Pierre Mongeau, "Social Influence and Discourse Similarity Networks in Workgroups," *Social Networks* 52 (2018): 230.
27. Stefano Gurciullo and Slava Mikhaylov, "Topology Analysis of International Networks Based on Debates in the United Nations," *arXiv:1707.09491v1* (2017); Steffen Eckhard, Ronny Patz, Mirco Schönfeld, and Hilde van Meegdenburg, "International Bureaucrats in the UN Security Council Debates: A Speaker-topic Network Analysis," *Journal of European Public Policy* 30, no. 2 (2021): 214–233.
28. The totals were actually much bigger, but we discarded online spaces only outlinked once, which were overwhelmingly irrelevant.
29. Lee, Jung and Song, "Subject-method Topic Network Analysis"; Gurciullo and Mikhaylov, "Topology Analysis of International Networks".
30. Caterina Froio, "L'Altérité sur les Sites Web des Extrêmes Droites en France," *Réseaux* 2, n.202-203 (2017): 39-78.
31. We actually have twice as many dictionaries: five in French, five in English – see Appendix 1 (supplementary material).
32. Caterina Froio, "L'Altérité"; Caterina Froio, "Race, Religion, or Culture? Framing Islam between Racism and Neo-Racism in the Online Network of the French Far Right," *Perspectives on Politics* 16, no. 3 (2018): 696–709.
33. Vincent Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics* 10 (2008): P10008.
34. Data for this figure and all other ones included in the paper is provided in CSV format in an online appendix to this paper.

35. Annie Benveniste and Etienne Pingaud (2016) Far-Right Movements in France: The Principal Role of Front National and the Rise of Islamophobia. In: *The Rise of the Far Right in Europe*, edited by Gabriella Lazaridis, Giovanna Campani and Annie Benveniste, editors (London: Palgrave Macmillan, 2016).
36. Froio, "Race, Religion, or Culture".
37. Caterina Froio, "L'Altérité".
38. Stuart Macdonald, Kamil Yilmaz, Chamin Herath, J.M. Berger, Suraj Lakhani, Lella Nouri, and Maura Conway, *The European Far-right Online: An Exploratory Twitter Outlink Analysis of German & French Far-Right Online Ecosystems* (Washington, DC: RESOLVE Network, 2022).
39. Again, this is due to the contrast between a fast-moving extremist digital environment and a slow-paced academic research and publishing system – here aggravated by the scarcity of empirical studies focusing on the British far-right online ecosystem.
40. Read for example Owen Worth, "The Great Moving Boris Show: Brexit and the Mainstreaming of the Far Right in Britain," *Globalizations* 20, n. 5 (2022): 814–828.
41. Bovet and Grindrod, "Organization and Evolution of the UK Far-Right Network on Telegram."
42. As stated above, URL detection and networks remain key to study these transnational ties.
43. Andrew Joyce is an anti-Semitic Holocaust-denying author, and "88" is a common neo-Nazi shorthand for "Heil Hitler".
44. For a presentation and appraisal, read Jade Hutchinson, Julian Droogan, Lise Waldek, and Brian Ballsun-Stanton, *Violent Extremist & REMVE Online Ecosystems: Ecological Characteristics for Future Research & Conceptualization* (Washington, DC: RESOLVE Network, 2022).
45. See for example Bart Schuurman, "Research on Terrorism, 2007-2016: A Review of Data, Methods, and Authorship," *Terrorism & Political Violence* 32, no. 5 (2020): 1011–26; Elanie Rodermond and Frank Weerman, "The Strengths and Struggles of Different Methods of Research on Radicalization, Extremism, and Terrorism," *Studies in Conflict & Terrorism* (2024), online before print.
46. Baele, Brace, and Coan, "Uncovering the Far-Right Online Ecosystem."

Acknowledgements

The authors want to thank Travis Coan, Elahe Naserian, and other collaborators of the NORFACE "ExID" consortium for assistance in data collection, as well as John Carbone for comments on an earlier iteration of the manuscript. Scraping of the websites as well as the organization of the content in a dataset followed the project's detailed ethics and data management plan, which included storage in secure firewalled servers as well as strict guidelines for content collection.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

The data for this article was collected during the "Extreme Identities" ("ExID") which is financially supported by UKRI-ESRC [2020/SGW/00884431] involved in the NORFACE Joint Research Program Democratic Governance in a Turbulent Age, itself co-funded by the European Commission through Horizon 2020 under grant agreement [No. 822166].

ORCID

Stephane J. Baele http://orcid.org/0000-0002-3632-0888
Lewys Brace http://orcid.org/0000-0002-7075-1243