

# Using Radical Environmentalist Texts to Uncover Network Structure and Network Features

Sociological Methods & Research  
2019, Vol. 48(4) 905-960  
© The Author(s) 2017  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/0049124117729696  
journals.sagepub.com/home/smr



Zack W. Almquist<sup>1,2</sup> and Benjamin E. Bagozzi<sup>3</sup>

## Abstract

Radical social movements are broadly engaged in, and dedicated to, promoting change in their social environment. In their corresponding efforts to call attention to various causes, communicate with like-minded groups, and mobilize support for their activities, radical social movements also produce an enormous amount of text. These texts, like radical social movements themselves, are often (i) densely connected and (ii) highly variable in advocated protest activities. Given a corpus of radical social movement texts, can one uncover the underlying network structure of the radical activist groups involved in this movement? If so, can one then also identify which groups (and which subnetworks) are more prone to radical versus mainstream protest activities? Using a large corpus of British radical environmentalist texts (1992–2003), we seek to answer these questions through a novel integration of network discovery and unsupervised topic modeling. In doing so, we apply classic network descriptives (e.g., centrality measures) and more modern

---

<sup>1</sup> Department of Sociology, School of Statistics, University of Minnesota, Minneapolis, MN, USA

<sup>2</sup> Minnesota Population Center, University of Minnesota, Minneapolis, MN, USA

<sup>3</sup> Department of Political Science and International Relations, University of Delaware, Newark, DE, USA

## Corresponding Author:

Zack W. Almquist, Department of Sociology, School of Statistics, University of Minnesota, 909 Social Sciences Building, 267 19th Avenue South, Minneapolis, MN 55455, USA.

Email: almquist@umn.edu

statistical models (e.g., exponential random graph models) to carefully parse apart these questions. Our findings provide a number of revealing insights into the networks and nature of radical environmentalists and their texts.

## **Keywords**

text networks, text analysis, topic modeling, social networks, environmental networks

Radical activist groups are of great interest to social movement scholars as well as sociologists and social scientists more generally. Indeed, these activist groups—which in some cases encompass leftist-oriented radical organizations such as Earth First! (EF!), the Earth Liberation Front (ELF), and the Anarchist Federation (AF), and in other cases encompass right-wing extremist groups such as the White Wolves, Aryan Nations, and Combat 18 (C18)—have now been extensively studied, especially within the contexts of advanced developed democracies in Europe, North America, and elsewhere (Blee and Creasap 2010; della Porta 1995; Downing 2001; Plows, Wall, and Doherty 2004; Wall 1999a). The use of illegal or controversial protest tactics by radical social movements—including (nonviolent) direct action, related forms of property sabotage,<sup>1</sup> protest camps,<sup>2</sup> and terrorism—has similarly garnered attention for such movements within the media and general public and among those concerned with domestic terrorism (Ackerman 2003; Chalecki 2002; Leader and Probst 2003; Sánchez-Cuenca and de la Calle 2009). These features have together led social scientists and social movements scholars to pursue a number of compelling studies of the life cycles, tactics, ideologies, and influences of radical activist groups (e.g., Blee and Creasap 2010; Foreman and Haywood 1993; Kitschelt 1986; Seel 1997; Wall 1999b).

Yet, extant research on radical social movements has also been limited by a paucity of data pertaining to the strategies, linkages, and agendas of these organizations. This deficiency is unsurprising. Many radical movements are short-lived or are highly volatile in their ideology, strategies, and membership (Smith and Dampousse 2009; Simi and Futrell 2015). The groups involved in these movements are also typically nonhierarchical in structure and anonymous in membership (Fitzgerald and Rodgers 2000; Joossee 2007), which limits researchers' abilities to identify and compare membership structures and patterns across groups. By virtue of occupying the fringes of the society, the viewpoints of radicals, and the literature they produce, commonly fail to make it into mainstream media sources or public archives. Perhaps most notably, the

ideology of radical activist groups, the illegal (or antigovernment) protest tactics that they often favor, and past countermovement efforts led by government actors have together given radical activist groups the incentives to conceal, obfuscate, and misrepresent their membership and ties with other radical groups as well as their favored tactics and ideologies (Plows et al. 2001, 2004; Simi and Futrell 2015). As a consequence of these tendencies, research on radical activism has been largely limited to qualitative case studies and small-*N* research. While these qualitative studies are insightful, we believe that a more systematic quantitative analysis of radical social movements could serve as a useful compliment to existing approaches.

Further, we contend that recent advances in automated content analysis, and social network analysis,<sup>3</sup> may together provide researchers with a means to do so. Regarding automated content analysis, we note that while radical groups are indeed guarded in membership, ideology, protest strategies, and collaborative efforts, they nevertheless communicate extensively with group members, potential recruits, and like-minded extremist groups through self-produced literature and periodicals (e.g., newsletters, websites, and magazines). For example, the U.S.-based *Earth First! Journal* has published approximately eight issues a year since 1981, whereas recent estimates by the Simon Wiesenthal Center (2008) have identified roughly 8,000 problematic hate/terrorist websites and related Internet postings. These texts are typically typeset and distributed by radical groups themselves and serve to discuss, organize, and promote the goals, concerns, tactics, and past and future activities of radical groups. We thus believe that a specific subset of automated content analysis techniques known as topic models, which allow one to systematically unpack a corpus of text documents by identifying words and phrases that commonly group together across documents, can help researchers to uncover the latent topics, or common themes, that arise across radical environmentalist texts. To the extent that these texts speak for radical groups, which they commonly profess to do,<sup>4</sup> the topics identified by topic modeling can be viewed as representing the ideology, tactics, and foci of radical groups themselves. As such, whereas the texts produced by radical groups offer a window into their activities and interests, topic models provide us with a means of systematically accessing this window.

Social network analysis (SNA) offers an additional, and complementary, suite of tools for the systematic study of radical groups. Indeed, group interaction, cohesion, and coordination have long been of interest to the social science community and lie at the heart of radical group behaviors. Over the last century, the development of methods and theories to both describe and predict such interaction has been largely developed under the umbrella of

SNA. This theory includes formal mathematical models for describing complex relationships between entities (e.g., organizational coordination), and such methods have now been widely applied to studies of (radical) social movements (Ackland and O’Neil 2011; Caiani and Wagemann 2009; Hoffman and Bertels 2009; O’Callaghan et al. 2013; Saunders 2007b). For such movements, SNA offers researchers the ability to identify the ties that exist between and within various radical groups, allows scholars to quantify the centrality of various actors in a group network, and can potentially be used to shed light on the protest strategies that different networked group pairs choose to use with one another. However, while this technique has much potential for understanding radicalism, it has been employed rarely in this area of research due to the complexity of gathering high-quality network data for the covert actors described above. A key contribution of our article rests in the development of a means to overcome this limitation: We argue below, and show, that quantitative text analysis techniques can help to rectify data limitation challenges for social network analyses of covert groups.

Our article does so by combining topic modeling and SNA in novel ways—and at multiple stages of one’s analysis—so as to allow researchers to uncover the tactics *and* interactions of clandestine social movements and groups. In these respects, we develop a methodological “road map” for future scholars to use in their own applications and provide a number of important substantive insights into the networks and behaviors of contemporary radical leftist groups. Specifically, our proposed approach details how one can combine the use of co-occurrence counts, social network statistics, and structural topic models (STM) to extract an environmental group network—and the full set of protest strategies used within that network—from a corpus of self-produced (and often highly variable) radical environmentalist texts. We then show how to statistically associate one’s identified protest strategies with various subcomponents of the group network itself, which can be directly used in validation or hypothesis testing. In full, our approach enables us to explain why, when, and how radical groups may cooperate with one another in order to achieve their respective aims as well as the manners in which these groups choose to cooperate. To fully illustrate this approach, and the insights that arise from it, we apply our proposed methods to a novel corpus of radical UK environmentalist texts from the 1990s and 2000s, which yields a number of intriguing and theoretically consistent results.

Specifically, we extract a network of 143 UK leftist and environmentalist groups—with 134 co-occurrence relations between these groups—from our radical UK environmentalist texts. While regional UK environmentalist groups comprise the bulk of this network, centrality measures reveal that a

small number of broader radical leftist (e.g., anarchist, antiglobalization, or worker's rights) groups feature most prominently within our network's connected component. This finding may help to explain the decline in the (UK) radical environmental protest movement during the late 1990s and early 2000s, as the centrality of nonenvironmentalist actors within our identified network suggests that the persistence of this network, and the groups therein, may have primarily rested on broader support for nonenvironmental leftist groups and views (e.g., anarchism) which was also waning during this time period.<sup>5</sup> Our analysis of these texts also identifies clear tactics of direct action, protest camps, international terror, and mass protest, and we find these tactics to be statistically associated with network group pairs and clusters in revealing manners that are also consistent with extant understandings of these groups. These findings help to validate our approach and suggest that systematic information on group ties, clusters, and shared strategies can be collected from radical group texts in an automated manner, thereby demonstrating how text and SNA can be combined to improve our understanding of covert groups.

This article proceeds as follows: We first cover a brief literature review of automated text and text-based networks, followed by a presentation of our sample of radical UK environmentalist texts, and describe how we extract a number of network and text analysis inputs from these raw texts. Next, we derive the complete network of radical groups from our radical environmentalist texts and discuss this network's features. We then identify the group tactics, or protest strategies,<sup>6</sup> that underlie our radical environmentalist texts and examine how these strategies vary in relation to the radical activist groups found in our network's connected component. Next to last, we discuss some novel ways to employ our automated text analysis of network discovery methods, and finally, we conclude with a discussion of our results and their implications.

## Literature Review

In the following subsection, we discuss briefly the use of automated text models in sociology and the social sciences broadly and the use of text-based networks within these social science communities.

### *Automated Text Analysis*

Recently, automated text methods have become a popular tool within sociology and the social sciences more broadly (see, e.g., Mohr and Bogdanov

2013; Roberts et al. 2014). Blei, Ng, and Jordan (2003) introduced Latent Dirichlet Allocation (LDA) models for text/topic modeling in machine learning—and recent developments in available software have begun to bring these methods to the forefront of both qualitative and quantitative analysis due to these methods' scalability and flexibility (see, e.g., Roberts, Stewart, and Tingley 2015). This can be seen through the special issue in *Poetic's* (Mohr and Bogdanov 2013) on topic models in cultural sciences and key work in political science such as that for unstructured survey responses (Roberts et al. 2014) or analyses of media (Grimmer 2012). This work has largely focused on using automated text methods for extracting central themes from big and small data. Important developments in calibration (e.g., Grimmer and Stewart 2013) and interpretation (Bonilla and Grimmer 2013; DiMaggio, Nag, and Blei 2013; Jockers and Mimno 2013; Marshall 2013; McFarland et al. 2013; Miller 2013; Mohr et al. 2013; Tangherlini and Leonard 2013) have similarly led to key insights and improvements in our understanding of social phenomena and a larger acceptance of these methods in practice within the social science communities. Finally, while the STMs methods discussed below have gained traction in political science, they have not yet been readily introduced to the sociological community which has primarily focused on classic LDA methods. In the following sections, we employ a combination of automated text analysis and extraction methods, data cleaning, and STM to build and qualitatively assess automated network discovery methods from automated text models. Later, we discuss how to employ these methods to pull apart a complex multiplex network.

### *Text-Based Networks*

Interest in building networks out of text has a rich history in sociology, extending from such classic works by Padgett and Ansell (1993), who built social/political networks from accounts of historical data of 1400s Italy. Other major examples include networks built out of ethnographic data such as group observation studies (e.g., Freeman and Webster 1994) and networks built from surveys, card sorts, and other methods (e.g., Marsden 1990; Romney, Weller, and Batchelder 1986). Citation networks are among the most famous form of text-based network analysis and have a rich literature in the bibliometric field. Examples in sociology include comparing the citation networks of North America and Europe via organizational journals (Üsdiken and Pasadeos 1995) or cocitation of biological science papers (Mullins et al. 1977). This work also draws heavily upon the exchange theory literature (e.g., Burris 2004). Other novel uses of text-based networks include Bearman

and Stovel's (2000) work on narrative networks, and Martin's (2000) historical narrative work on the use of animals in stories and their relations to class division and labor. However, traditional methods for building networks out of text are labor-intensive and have the potential for coding effects (Marsden 1990; Romney et al. 1986). In response, more recent work in *automated* text analysis has been implemented by sociologists and social scientists, such as Adams and Light (2014) and Light (2016), work on understanding interdisciplinary networks in the medical community. In the field environmental sociology, Farrell (2016) has likewise employed automated topic models augmented with networks built out of affiliation networks, in a similar fashion, to the networks built for the subfield of *corporate interlock* networks (Mintz and Schwartz 1981). Also, in the area of environmental sociology, Farrell (2015) has explored environmental activist conflicts through automated text analysis and social networks.

## Sample and Background

In implementing our proposed methodological approach, we focus on a single UK radical environmentalist publication, known as the *Do or Die* (DoD) magazine. In addition to the substantive justifications provided below, we chose to focus on this publication—as opposed to a broader collection of publications produced by different groups, across different countries, or over larger periods of time—in order to ensure that our underlying documents, networks, and identified groups were as comparable as possible. DoD was published semiannually during the years 1992 to 2003 by an anonymous collective of radical British environmentalists based, in large part, in Brighton, United Kingdom. At the time, the publication referred to itself as *the voice of the UK EF!* movement, although its editorial collective have since suggested that the publication was rather only *a voice of the movement* (*Earth First! Journal* 2006). As stated by an editor, DoD more generally “pushed a green anarchist, direct action perspective [. . .] gave publicity to sabotage and had a no compromise attitude [and was] largely aimed at a few hundred people in the UK eco scene,” although it was also widely read by more traditional anarchists and conservationists (*Earth First! Journal* 2006).

In a total of 10 issues, DoD was highly variable in structure and topics covered. Each issue had roughly 50–60 uncredited voluntary contributors, and the publication on the whole grew over time from a 20-page “zine” (issue 1) into a 343-page “massive book” (issue 10; *Earth First! Journal* 2006). Individual contributions (hereafter “stories”) included but were not limited to journalistic accounts of environmental protests in the United Kingdom or

abroad; tutorials on sabotage, incarceration, or police interrogation; book reviews; and calls to arms over specific environmental crises. As is common across radical environmental zines, cartoons, images, and unstructured text were interspersed throughout these contributions. Ultimately, DoD ceased production in 2003 because its contributors wanted to pursue other activities, rather than for financial reasons (*Earth First! Journal* 2006). With respect to the UK EF! movement that DoD sought to speak for, DoD's editors note that, during its publication run, their preferred "green anarchist perspective went from being the minority to the majority perspective within EF! in the UK [which] resulted in lots of people dropping much of the non-violent pacifist ideology, moving more towards an anarchist position and supporting sabotage actions" (*Earth First! Journal* 2006).

Scholarly accounts of the UK EF! movement largely support this latter contention. The UK EF! movement began in Spring of 1991 as a radical environmental advocacy group with the expressed goals of defending the environment, confronting those who threaten it, and promoting lifestyles that were in balance and harmony with nature (Purkis 1998:201). As such, the first UK EF! actions were heavily influenced by the previously established US EF! movement, as well as by the UK peace movement, and were initially directed toward blockades of the Dungeness nuclear power plant and the M. V. *Singa Wilstream* (a ship carrying tropical timber) on the Thames Estuary. These preliminary tactics of blockade then quickly evolved into larger scale efforts toward mass action, anti-roads protests (in turn spurring new environmental movements and groups), and later into additional issue areas such as opposition to airport construction, genetics, and housing on greenfield sites (Wall 1999b). Furthermore, by the mid-1990s, these strategies of nonviolent direct action (NVDA) were increasingly observed alongside tactics of active sabotage, as both moderate and militant groups grew to coexist within the UK EF! during this period (Wall 1999b:88). The EF! movement and a variety of additional groups that have arisen from it—such as the oft-terrorist-labeled ELF—remain active to this day, albeit with considerably lesser prominence.

There is good reason to view the UK EF! and its DoD publication as embedded in a larger network of radical British groups and shared protest strategies. For example, Seel (1997:177) argues that the UK EF! culture "serves as an informal network from which more publicized protests spring. EF! is embedded in the same protest milieu as Reclaim the Streets, Corporate Watch, The Defiance Alliance, Road Alert!, The Land is Ours, and Justice? which prints Schnews." He goes on to point out that



[i]ndividuals often take part in several of these groups while being tied to none [and] EF! overlaps with these other disorganisations to such an extent that it is not really distinguishable from them. It is best understood [...] as a way of *linking* activists within the green protest scene. (p. 177, emphasis added)

Wall (1999b) similarly discusses the importance of networking to the UK EF!'s initial organization and recruitment, subsequent mobilization, and eventual fragmentation. In these regards, the UK EF!'s network greatly facilitated the adoption and diffusion of protest strategies from and between other groups, wherein

[n]etworking brought with it the influence of the ideologies, issues foci, repertoires of action, strategies and articulations of existing green networks [...], mass NVDA repertoires were derived from the peace and Australian rainforest movements. Repertoires of sabotage, in turn, were derived from animal liberation militants and to a lesser extent EF! (US) (Foreman and Haywood 1993). (Wall 1999b:90)

Taken together, and given the near-preeminent role of EF! within the broader UK radical environmental movement during this era, these features ensure that DoD encompasses both the core and periphery of the UK radical environmental movement over a relatively large period of time. We believe that this makes DoD an ideal text corpus for extracting and analyzing the underlying network of radical UK (environmental) groups and their strategies during the 1990s and 2000s.

Acquiring the entire 10-issue corpus of DoD publications was a challenging task. Given the relative obscurity and somewhat dated nature of this publication, comprehensive records of DoD do not exist in machine readable form within any single library or repository. Fortunately, issues 5–10 and issue 1 are fully transcribed and archived in machine readable format on the now inactive, online archive “eco-action.org,” which was a website used primarily for documenting UK environmental activism during the 1990s (Eco-Action.org, Nd). The entries are stored in HTML at the story level (separately for each issue), and these individual stories were web-scraped for use in the analysis below. pdf images of the four remaining issues (issues 2–4)—as well as pdf images of the additional web-scraped issues mentioned above—were then obtained and downloaded from a second online archive of radical environmental activities known as the “Talon Conspiracy” (2014). These pdf-image files, which occasionally exhibit low image quality,<sup>7</sup> were then converted to machine readable text using optical character recognition (OCR) software.

Through the above approach, we obtained or created (i) pdf-image files of each DoD issue, (ii) OCR'd versions of each of these 10 issues (at the publication-page level) stored in txt and rtf formats, and (iii) cleaned web-scraped versions (in txt format and at the publication-story level) for issues 5–10 and 1. Given the higher quality of our web-scraped versions, we use these versions for issues 5–10 and 1 in our analysis below and then employ the OCR'd versions for the remaining issues (2–4).<sup>8</sup> Importantly, while our reliance on both OCR'd and web-scraped text for the creation of our final corpus may give rise to comparability issues, the fact that our OCR'd text corresponds to three early issues of DoD helps to minimize these concerns. Specifically, while the 10-issue DoD corpus is comprised of 1,565 total magazine pages, the subtotal for our three OCR'd issues amounts to only 136 pages. The remaining 1,429 pages are in web-scraped format, implying that only 8.7 percent of our final corpus was derived from OCR'd text.

## **Preprocessing**

Our analysis requires that we create two intermediate input quantities from the text files described above: (i) a list of relevant (radical) UK environmental groups and (ii) a corpus of fully preprocessed text “documents.” The former is used for identifying the pairs of radical groups that co-occur within our text corpus and the underlying network of radical groups that arises from these co-occurrences. The latter is used for estimating the latent strategies that are discussed across our corpus and the variation in these latent strategies vis-à-vis our identified group network. We directly derive each intermediate input quantity from our text sample of DoD issues 1–10 and discuss each process in turn below, beginning first with our radical UK group list.

We used information contained within DoD itself to identify and construct our list of relevant UK groups. The final three to four pages of each DoD issue provide comprehensive contact information (e.g., names and addresses) for a wide range of radical leftist groups, typically separated by international and domestic (UK) groups as well as by group type (e.g., UK ecological direct action groups and other UK contacts). UK ecological direct action groups include regional EF! groups (e.g., Leeds EF!) as well as environmental protest groups focused on more specific environmental threats such as “Reclaim the Valleys,” an antimining protest movement in South Wales. The groups listed under “other UK contacts” ranged from radical bookshops (e.g., 56a Infoshop) to more traditional anarchist groups (e.g., AF) to left-wing UK organizations focused on nonenvironmental issues (e.g., Campaign Against Arms Trade). From these lists, we constructed a master list that included all

UK ecological direct action groups and all other UK contact groups and then standardized our final group list to account for alternate spellings and capitalizations of each group name. Further below, we then use this UK environmental group list to identify the instances of proximate co-occurrence of each group pair within the DoD text, which is in turn used to construct our radical environmental group network. In this respect, our analytic framework follows past scholarship (e.g., Saunders 2007b) in analyzing our environmental group network through a relational approach—that allows the group interactions and behaviors within our texts to inform our network—rather than through a positional approach that defines our network based upon a (subjective) classification of our groups into specific issue areas (or categories) *a priori*.

The final list of 143 UK (environmental) groups and organizations—along with descriptions of each group or entity—is provided in Online Supplemental Materials. Importantly, while many of our 143 identified groups have historically advocated for issues that fall outside of the environmental arena (e.g., anarchism), we continue to characterize all 143 groups as (radical) environmental groups in our ensuing discussions, given that all groups in our sample were (i) at least tangentially involved in environmental issues and (ii) listed as key contacts, and frequently referenced within the texts, of a radical environmentalist publication. While we believe this to be most consistent with our relational approach, however, we do take care to note those instances where a specific group's focus extended beyond radical environmentalism. Likewise, our final 143-group list also contains a number of UK groups whose tactics did not traditionally encompass radical environmental direct action. Rather than arbitrarily remove these potentially nonradical groups from our UK-group list *ex post*, we maintain these groups in our sample and use this variation (in known group strategies) to validate our protest strategy extraction methods below by assessing whether our unsupervised text analysis approach is indeed correctly identifying those groups that have been known to use, or not use, radical environmental direct action tactics.

We next turn to the second intermediate task mentioned above: the creation of a set of fully processed “documents.” Because the unsupervised topic modeling techniques used below require that we apply these methods to a collection of text documents, we must first define what a standard document should be for this stage of our text analysis. Based upon our substantive knowledge of the DoD corpus, as well as previous applications of topic models to social science texts, plausible document designations include each individual DoD issue, each individual page of text within our DoD sample, individual sentences or paragraphs, or each individual story entry within the

corpus. Extant social science research has applied similar topic models to those used below to documents ranging in size from individual tweets (Barberá et al. 2014) to individual books (Blaydes, Grimmer, and McQueen 2015). Others have used more arbitrary text breaks to define documents such as page breaks, sentences, multisentence sequences, or paragraphs (Bagozzi and Schrodt 2012; Brown 2012; Chen et al. 2013). Hence, for the methods applied below, a great deal of flexibility can be afforded in defining one's documents.

However, given that topic models are inherently designed to analyze *large collections of text* (Grimmer and Stewart 2013), the limited number of actual DoD issues (i.e., 10) precludes us from treating the individual DoD issues as our documents of interest. This limitation, and the broader formatting idiosyncrasies of DoD, accordingly made the designation of (comparable) documents for this corpus uncharacteristically difficult. To this point, recall that the individual contributions within DoD were often inherently unstructured and variable, with some entries following traditional paragraph form, others being more experimental and artistic in format, and still others following a simple bullet-point style. Moreover, the stories and pages within DoD often have text boxes arranged in different paragraph and column formats, with individual story entries ranging from 21,000+ words to 140 words (or less). Complicating this further was the fact that the OCR'd documents in our sample (issues 2–4), which were OCR'd at the page level, often lost information on paragraph breaks and story breaks whereas our web-scraped documents (issues 1 and 5–10) were archived at the story level, without complete information on page-breaks.

In light of these inconsistencies, the most defensible method for dividing the DoD corpus into text documents of comparable length is the use of multisentence sequences. Individual sentence-length documents were deemed far too short for identifying group co-occurrences or underlying mixtures of topics. By contrast, the use of consecutive sequences of sentences has past justification both in the estimation of underlying topics and in the identification of actors' co-occurrences for the extraction of social networks from text (Bagozzi and Schrodt 2012; Chang, Boyd-Graber, and Blei 2009; Culotta, Bekkerman, and McCallum 2004; Davidov, Rappoport, and Koppel 2007). In essence, this approach entails that one link together non-overlapping<sup>9</sup> series of  $n$  consecutive sentences, treating each resultant sentence sequence as an individual document. To do so, we start from the first DoD issue, and iterated through all 10 issues (in order), combining consecutive sentences by our designated  $n$ , and incrementally treating each subsequent  $n$ -sentence sequence as a new document. While somewhat arbitrary,

this process helps immensely in ensuring that each document is of comparable size and content, as opposed to story-level breaks (which would entail documents ranging from approximately 100 words to others with over 20,000 words) or page-level breaks (which are equally arbitrary and likely too large for credible co-occurrences). As elaborated upon below, we specifically use text documents derived from 12-sentence sequences in our primary analysis but also explore 6- and 18-sentence sequences in order to ensure that our results are not overly sensitive to any single sentence sequence length.

Before constructing these sentence sequence documents, we first removed the aforementioned UK (and international) group contact lists from the final pages of each DoD issue. While we use these contacts for identifying our groups of interest, we wish to avoid treating this text as actual content text during either the extraction of group co-occurrence information or the topic modeling stages of our analysis, given that the contact lists included within DoD provide little to no surrounding content text aside from the listed names and contact information for each group. After dividing our remaining corpus into 12-sentence documents, we then processed each document's text to remove all punctuation, numbers, and stopwords. The removal of these character sets is standard preprocessing for our intended topic model techniques (Bagozzi and Berliner 2017; Roberts et al. 2014). We then omitted sparse terms that did not occur in at least 1 percent of the documents in our corpus. Finally, and as is also common in topic modeling applications, all remaining words were converted to lower case, stemmed, and restructured into a document-term corpus before analysis. Altogether, these preprocessing steps created a corpus with 3,210 unique documents and 2,082 unique word stems. Further below, these preprocessed documents are used within topic models to discover the latent strategies that underlie the DoD corpus and to examine how these strategies vary in relation to our (separately derived) radical UK group network.

The tasks associated with extracting our radical group network, together with our topic model analysis, helped to inform our decision to use 12-sentence sequences as our primary document length. For our topic model analysis, the choice of 12-sentence sequences yielded a document length that was comparable to the shorter, and most common, stories and related text entries within the DoD corpus, while minimizing the aforementioned heterogeneity in story-length documents. The choice of 12-sentence sequences also allowed us to maximize the number, and validity, of our group co-occurrence indicator, which as mentioned above serves as the primary input for our radical group network. Here, the co-occurrence indicator reports whether or not a pair of groups from our final group list were each named within the

same document in our corpus. As such, the length of a document directly affects our identified co-occurrences and is tied to an assumption of how proximate two groups must be in the text for us to consider them as co-occurring, and thus sharing a tie within our resultant network. The use of co-occurrences for building networks in this fashion is well established (Chang, Boyd-Graber, and Blei 2009; Culotta et al. 2004; Davidov et al. 2007), and using this literature as a guide, alongside information on the typical document length for our corpus, we believe 12-sentence sequences to be a reasonable balance between the identification of true co-occurrences and the avoidance of false co-occurrences.

Indeed, lowering this sentence sequence threshold by 50 percent to six-sentence sequences—which roughly corresponds to a paragraph’s length of English-language text—significantly reduces the number of identified co-occurrences and thus our identified network ties, potentially ensuring a substantial number of missed ties. Increasing the sentence sequence threshold by 50 percent, to 18-sentence sequences, instead increases the potential for false positives, as radical leftist groups that appear much further apart in the original DoD text will be considered as having a tie when they may in fact have been discussed in wholly separate contexts. We nevertheless explore networks constructed from 6- and 18-sentence documents in Online Supplemental Materials. The 6- and 18-sentence sequence results yield findings that are generally comparable to those obtained from our primary 12-sentence sequence documents, but which are also less coherent in terms of the underlying network identified and group clusters uncovered. These deficiencies are in line with the expectations outlined above, and further reinforce the choice of 12-sentence sequences as our primary documents within the analysis presented below, while also underscoring the more general stability of our network to sizable changes in sentence sequence length.

To construct our co-occurrence indicator, we iterate through every group pair included within our UK radical group list. For each pair ( $i$ ), we query each 12-sentence document ( $d$ ) to record whether ( $= 1$ ) or not ( $= 0$ ) group pair  $i$  co-occurred in document  $d$ . Importantly, for our documents  $d$ , we in this case use the *unprocessed* 12-sentence sequence documents, as opposed to the fully preprocessed documents discussed above, as decapitalization, stemming, and proper noun removal each undermine the appearance and identification of our target group names within our co-occurrence queries.<sup>10</sup> Furthermore, and immediately prior to implementing these co-occurrence queries, we applied a standardization routine to address typographical errors, and/or common variations, in the spellings of target groups within our unprocessed documents. This standardization approach first entailed that we

identify and record each common (and plausible) group name variation, abbreviation, and misspelling, which was done through a manual review of the group contact lists provided in DoD issues 1–10. With this set of group name variations in hand, we implemented a processing script that individually standardized each group name within our unprocessed documents, while also incorporating unique features into some group names so as to ensure that groups with closely overlapping names were not incorrectly counted as (co-)occurring in instances when only a similarly named group was mentioned in a document.

With our group names standardized in this fashion, we are able to implement our co-occurrence queries in a manner that maximizes precision and recall. From the results of these queries, we construct a binary  $i$  by  $d$  matrix, where 1s denote co-occurrences between groups (rows) in particular documents (columns) and 0s denote the absence of a group-pair co-occurrence for a given document. A total of 44 groups in our 143-group master list co-occurred with another group at least once across all documents, with a mean, standard deviation, median, and maximum number of co-occurrences (per group) of 1.13, 2.52, 0.00, and 19.00, respectively. Accordingly, we define any group pair that exhibited *at least one* co-occurrence across all documents as having a “tie,” and use this classification for (i) constructing our group network and, later, (ii) assessing variation in latent protest strategies in relation to these network ties and clusters thereof.

## Environmental Group Networks

Social network methods have a long history in the social sciences, dating back to the 1930s (Freeman 2004). Co-occurrence networks have been used effectively in the bioinformatics literature (e.g., Cohen et al. 2005), computer science and engineering (e.g., Matsuo and Ishizuka 2004), and the bibliometric literature (for a review, see King 1987). Social networks have been employed for related theoretical development in explaining social movements (e.g., Byrd and Jasny 2010) and in organizational (e.g., Burt 2000; Spiro, Almquist, and Butts 2016) activities. Further, one can represent a collection of individuals engaged in shared activity as single entity which engages in collaborative activities, for example, citation networks between blogs or disaster relief (e.g., see Almquist and Butts 2013; Almquist, Spiro, and Butts 2016). This work builds on this literature and extends it in several key dimensions. First, we directly incorporate text information to understand the diffusion of information, ideas, and activities which occur through these networks. Second, we explore the endogenous nature of the network and

concepts which create these complex social interactions. Third, we empirically validate these measures in the policy relevant setting of modern environmental movements.

Here, we adopt the social network nomenclature of representing a network as a mathematical object known as a *graph* ( $G$ ). A graph is defined by two sets,  $G = (E, V)$ : (i) an edge set ( $E$ ) which represents a relationship, such as friendship and (ii) a vertex set ( $V$ ) which represents an entity, such as an individual or organization. Note that the edge set is sometimes referred to as a *link* or *tie*, interchangeably, and that the vertex set is referred to as the *node* set or *actor* set, interchangeably, in much of the social network literature (Wasserman and Faust 1994). We will use the same conventions. The edge set can be either directed (e.g.,  $A \rightarrow B$ ) or undirected (e.g.,  $A \leftrightarrow B$ ). Further, for analysis purposes, it can be shown that a graph may be fully identified by its adjacency matrix. An adjacency matrix ( $Y$ ) is a square dyadic matrix where  $y_{ij}$  represents a relation (0 or 1) between actor  $i$  and actor  $j$ . In this case, one specifies a relationship is present with a 1 and absent with a 0. The size of the matrix ( $n$ ) is also the size of the network or the number of vertices present in the graph. Often in network analysis the diagonal (or self-tie) is considered either as “empty” or not applicable (NA) and ignored in the analysis.

The network literature is comprised of both theoretical and methodological insights. In the context of radical environmental group networks we are particularly interested in descriptive analysis (Almquist 2012; Wasserman and Faust 1994), metrics of power (also known as centrality indices, see Freeman 1979), and group analysis (i.e., community detection or clustering, see, e.g., Fortunato 2010; Mucha et al. 2010).

In the following subsections, we will outline the necessary literature and background for interpreting and using these statistics in our specific context. We will begin by carefully detailing the network of co-occurrence for our radical UK environmental groups, including visualization and basic descriptives. We will follow up the descriptive statistics with a node-level analysis of the individual actors and finish up with a group-level analysis of the network.

### *The Network and Descriptive Statistics*

Typically network analysis begins by defining the bounds of the network and the relation of interest (Wasserman and Faust 1994). The group network analyzed in this article consists of 143 radical environmental groups identified in the text analysis portion of this research. The edge set is defined, as mentioned in the fourth section, by at least one co-occurrence between a ( $i, j$ )

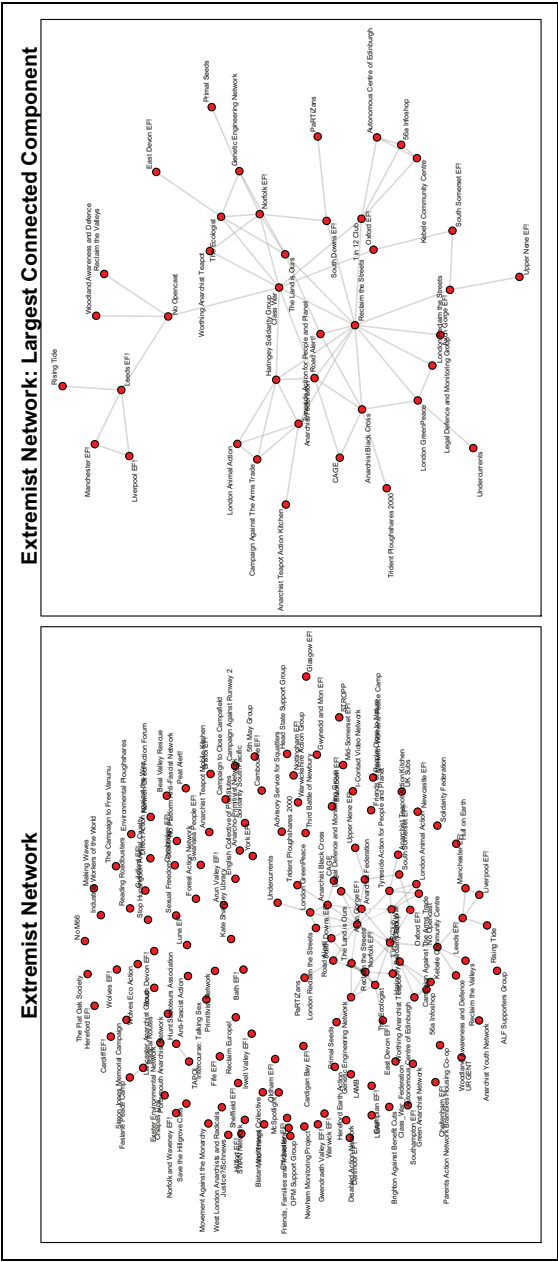


group in a given document. We treat this as a symmetric (or undirected) relation, such that  $y_{ij} = y_{ji}$ . To visualize the network and compute the descriptive statistics, we employ a specialized package in R (R Core Team 2015) dedicated to network analysis (i.e., the SNA package; see Butts 2008, for more details). In Figure 1, we plot the complete network of radical environmental groups along with the largest connected component. The largest connected component will be the focus of much of the analysis of the following sections. This is the largest group in which action, information, and so on, can flow.

This network is defined by 143 groups and 134 co-occurrence relations, providing a density of 0.00583.<sup>11</sup> Another important measure of the network is the degree distribution and the mean degree statistic, where degree of node  $v$  is number of ties connected to  $v$  ( $\text{deg}(v) = \sum_i y_{iv}$ ). This network is highly skewed with many groups having zero ties and the most active group having 13 ties with a mean degree of 0.94. However, if we observe only the connected component, we see that mean degree of this group is 3.25 ties. This network is heavily stratified between active members and nonactive members. Among active members, the upper-and right-most outer sections of our network's connected component are dominated by regional EF! groups and related environmental direct action groups, which then bleed into a series of leftist bookshops, community centers, and nonenvironmental leftist groups toward the bottom and outer left-most regions of the connected component in Figure 1. These initial findings bear a number of similarities to extant research. For example, our observance of a highly stratified network between active and nonactive members is consistent with Saunders (2007b) relational network analysis of London-based environmental organizations, which found that environmental radicals exhibited a small number of ties, relative to other environmental organizations, and did not collaborate with more mainstream conservationist groups. Hence, our broader network appears to be consistent with existing understandings of radical environmentalist networks. We use node-level characteristics to discuss the most central members of this network in the next section, which are often interpreted as influence or power in the network.

### *Centrality*

In SNA, the concept of node centrality is an important and classic area of study within the field (Freeman 1979). There exist within the social network literature a large set of centrality metrics and measures of power. We choose to



**Figure 1.** Network plot of the 12-sentence co-occurrence network. The left panel of the figure is the complete network including isolates, and the right panel of the figure is the largest connected component. Graph plotted using Fruchterman and Reingold's force-directed placement algorithm.

**Table 1.** Top 10 Actors for Degree Centrality.<sup>a</sup>

Organizations	Degree	Eigen	Betweenness	Closeness
Reclaim the Streets	13.000	.489	.213	.172
Class War	9.000	.360	.204	.158
Norfolk EF!	6.000	.214	.025	.127
Anarchist Black Cross	6.000	.237	.030	.126
Haringey Solidarity Group	6.000	.294	.040	.136
The Land is Ours	5.000	.223	.029	.127
Tyneside Action for People and Planet	5.000	.198	.032	.122
<i>The Ecologist</i>	5.000	.180	.033	.123
Anarchist Federation	5.000	.279	.023	.133
The 1 in 12 Club	5.000	.194	.064	.131

<sup>a</sup>Included are the actor’s eigen, betweenness, and closeness centrality measures.

focus on the four most popular (Freeman 1979; Wasserman and Faust 1994): degree, eigen, betweenness, and closeness centrality. Each captures a slightly different, but important aspect of node position within the network. Degree centrality is a straightforward measure of how connected a given individual actor is to all other actors in the network (Anderson, Butts, and Carley 1999). Eigen centrality (or eigenvector centrality) corresponds to the values of the first eigenvector of the adjacency matrix; this score is related to both Bonacich’s power centrality measure and page-rank and can be thought of as core-periphery measure (Anderson et al. 1999; Bonacich 1972). Betweenness is a (often normalized) measure of shortest path or geodesic distance between overall graph combinations; conceptually, a node with high-betweenness resides in a space with a large number of nonredundant shortest paths between other nodes. This can be interpreted as nodes which act as “bridges” or “boundary spanners” (Freeman 1977). The last centrality measure we consider is closeness which is another geodesic distance-based measure. Here we use the Freeman (1979) version which is defined for disconnected graphs and has largely the same properties as the classic closeness measure. We have tabulated the top-10 activist groups as ordered by degree in Table 1.

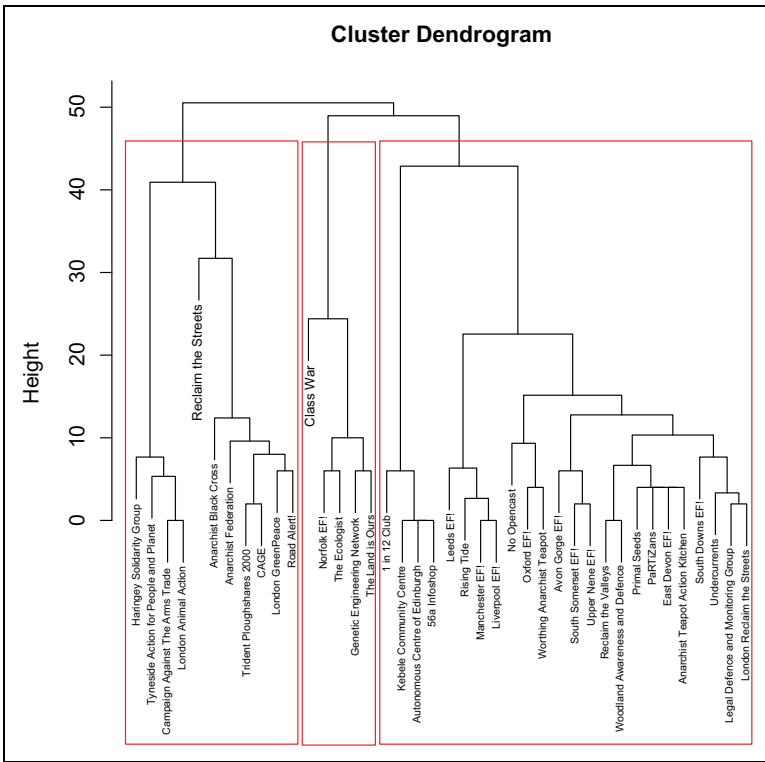
Generally speaking, an activist group who is high on one measure of centrality is also high on the other measures of centrality. The lowest correlation is between betweenness and closeness which correlate at 0.56, and the highest correlation is between degree and eigen centrality at 0.95. The ranking of nodes is of course not perfectly aligned even in the case of degree and eigen centrality. Consider the Haringey Solidarity Group, which is tied for third with Norfolk EF! and Anarchist Black Cross in degree but is clearly ranked second

in eigen centrality. In this case, the Haringey Solidarity Group's high levels of centrality across both metrics are unsurprising, given this group's long history, central London location, active participation in a wide array of crosscutting issue areas (including environmental sustainability, social justice, worker's rights, and antifascism, among others) and media profile (including a website, very widely distributed newsletter, and Twitter account). Another illustrative case is that of AF which is tied for fourth rank with several other groups in degree, but would be ranked third in eigen centrality,<sup>12</sup> with each high-centrality measure perhaps owing to this AF's uncharacteristically large membership size, reach, organizational development, and media presence.

The key finding of the centrality measures in Table 1 is that there are two groups—Reclaim the Streets and Class War—who very much stand out on all measures followed by several other groups who are highly active and then a separate set of groups with no collaboration at all. Like the cases of the Haringey Solidarity Group and AF discussed above, the centrality of Reclaim the Streets (and Class War) is consistent with extant knowledge of these groups. Reclaim the Streets was originally a UK-based antiroads NVDA group that grew out of the UK EF! movement and related protest camps but quickly established itself as a key actor within the (UK and international) antiglobalization movement (Chesters and Welsh 2011:67). The (international) scope of this group, its ties to the UK EF! movement and related UK direct action groups, and the frequent mass street protests that it led<sup>13</sup> all undoubtedly helped to establish Reclaim the Streets as a central actor within the radical UK environmentalist scene during the 1990s and early 2000s. Similarly, Class War is a UK anarchist group that rose to prominence in the early 1980s through a widely circulated self-produced newspaper, music records, and national conferences and later through its coordination of various actions and protests. Although Class War splintered into a number of regional groups during the late 1990s, its central role as an organizing and publicizing force for far left-activism during this period (and earlier) likely contributed to its high levels of centrality in our network.<sup>14</sup> That said, we need to follow up this node-level analysis with an attempt to better explain the larger macro dynamics of the network. We analyze this aspect of the network in the following section.

### *Clustering*

There are numerous clustering (e.g., community finding) algorithms employed in SNA; however, the results of these various methods tend to be highly correlated (Anderson et al. 1999). We use a well-documented and



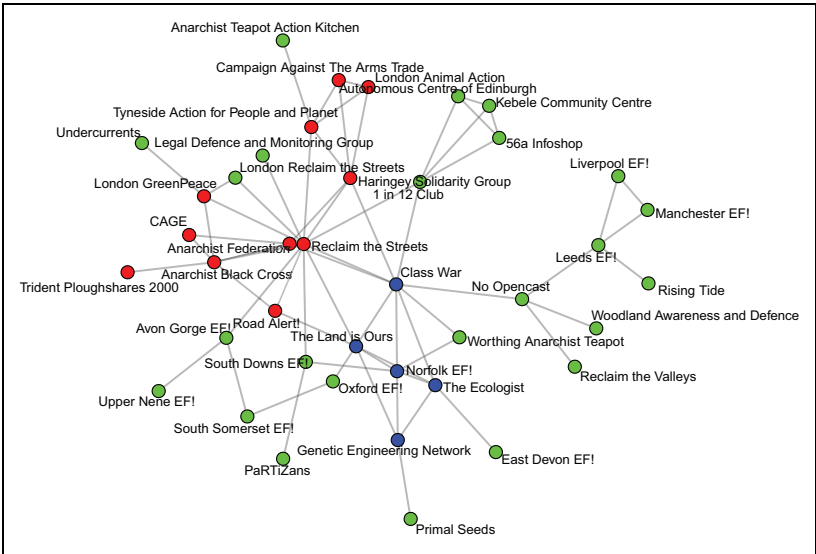
**Figure 2.** Dendrogram of the large cluster in the 12-sentence network. Computed using hamming distance with Ward’s D-decision criterion.

accepted method of Hierarchical cluster analysis on a set of distances. Because social networks are binary arrays it makes sense to use Hamming distance (Butts and Carley 2005), which measures the minimum number of substitutions required to change one bit into the other bit (0 or 1) in the case of a graph. Further, we employ the Ward algorithm for finding these clusters (Butts and Carley 2005). In SNA, this is equivalent to finding the structural equivalence of a set of nodes in a given network (for more details, see Butts 2008). Because the isolates (nodes with no connections) automatically form their own equivalence classes, we do not include them in our analysis. In Figure 2, one can see the dendrogram plot from the clustering algorithm. Through visual inspection (and statistical analysis), we settle on three core groups (see Table 2).

**Table 2.** Names of Radical Environmental Groups Attached to Each Cluster Assigned by the Dendrogram Clustering Method.

Cluster 1	Cluster 2	Cluster 3
Reclaim the Streets London GreenPeace	Norfolk EF! Genetic Engineering Network	Avon Gorge EF! East Devon EF!
Road Alert! Trident Ploughshares 2000 Tyneside Action for People and Planet	The Land Is Ours <i>The Ecologist</i> Class War	Leeds EF! Manchester EF! Oxford EF!
Anarchist Black Cross Campaign Against Arms Trade Haringey Solidarity Group Anarchist Federation CAGE		Reclaim the Valleys South Downs EF! South Somerset EF! Upper Nene EF! Legal Defence and Monitoring Group
London Animal Action		Liverpool EF! Woodland Awareness and Defence London Reclaim the Streets Undercurrents Anarchist Teapot Action Kitchen No Opencast Autonomous Centre of Edinburgh 56a Infoshop Kebele Community Centre The 1 in 12 Club PaRTiZans Primal Seeds Rising Tide Worthing Anarchist Teapot

To better visualize these clusters, we replot the connected component network colored by the individual clusters in Figure 3. Based upon the individual characteristics of our connected component groups, the three identified clusters reveal a number of interesting and theoretically consistent patterns. We highlight a number of these patterns here and evaluate our contentions more explicitly within our Text Analysis section further below. Cluster 1 contains leftist environmentalist groups focused on a wide variety



**Figure 3.** Network plot of the connected component of the 12-sentence network colored for cluster assignment.

of issue areas, including antinuclear protest (Campaign Against Arms Trade; London GreenPeace; and Trident Ploughshares 2000), environmental direct action (Tyneside Action for People and Planet, London Animal Action), antiroads agendas (Reclaim the Streets, Road Alert!), and opposition to excessive government control (AF, Anarchist Black Cross, Haringey Solidarity Group, and CAGE). Notably absent from this cluster are any regional EF! groups or community centers. Given this variation, and the high-centrality measures of many of the groups included in cluster 1, we suspect that cluster 1 primarily identifies radical environmental and leftist groups that coordinate in omnibus street protests and direct actions. Indeed, a recent Reclaim the Streets protest included a workshop led by London Campaign Against Arms Trade and stalls hosted by Haringey Solidarity Group and Animal Rescue (Reclaim the Streets, Nd.c).

Cluster 2 is smaller and more ambiguous. One can note that the groups contained in cluster 2 are generally environmental in nature and range from direct action-oriented groups (Norfolk EF!, The Land is Ours) to organizations primarily focused on information dissemination (*The Ecologist*, Genetic Engineering Network, and Class War). We therefore believe that this cluster

identifies groups coordinating over ecotage and related direct action activities or coverage thereof. More generally, this cluster, as well as clusters 1 and 3, appears to incorporate groups arising from multiple (environmental) issue areas and/or ideologies. This is consistent with the levels of positional incoherence observed within positional blocks in past network analyses of European environmental groups and organizations (Diani 1995, 2002; Saunders 2007b:232-3) and underscores the utility of our relational approach.

Cluster 3 is the largest identified cluster and contains a majority of our regional EF! groups, several related groups known for occupation/camp-based protests (e.g., Reclaim the Valleys, No Opencast), and a large number of community/organizing centers, among others. Given the common usage of direct action tactics by a majority of these groups, as well as their frequent mutual participation in large-scale protest camps, we suspect that this cluster is identifying groups that coordinated together via these more targeted direct action activities as well as the venues where such activities were organized. An example of such behavior with respect to protest camps targeting opencast mining comes from DoD itself, which notes that “Leeds EF! took action in early 1995 targeting an opencast site in Yorkshire, and at about the same time, Welsh activists were setting up camps at Selar and Brynhenllys sites near Swansea. These actions, among others, led to an increasing alliance between EF!(ers) and the No Opencast campaign” (DoD 1998).

## Radical Environmental Group Tactics

Our identified list of UK radical environmental groups, and group network clusters, includes some groups (e.g., *The Ecologist*) that neither engaged in nor advocated for acts of violence, sabotage, or property destruction, whereas others, including a number of our identified regional EF! groups, routinely did so. Hence, we next seek to discover the underlying protest tactics that are pursued by environmental groups within this network’s connected component and to evaluate whether these methods can accurately identify which groups actively pursued radical protest, and which did not, in an unsupervised fashion. To do so, we apply unsupervised topic models to our preprocessed text documents, so as to simultaneously (i) uncover the latent themes or “topics” that are discussed across documents and (ii) associate these topics with the group ties and clusters that we discussed above. Topic models allow one to recover the former quantity by treating one’s documents as a combination of multiple overlapping topics, each with a representative set of words. Latent topics are then estimated via a hierarchical model that treats each document as a mixture of underlying topics and returns the word



distributions that are most strongly associated with each topic across all documents. Topic model extensions that incorporate predetermined network information have largely focused on conditioning topic estimation upon network structures (and documents) linked by authorship/recipient (e.g., McCallum, Corrada-Emmanuel, and Wang 2005) or citations (e.g., Chang and Blei 2009; Dietz, Bickel, and Scheffer 2007), although Chang et al. (2009) and Krafft et al. (2014) have developed more flexible network-oriented topic models that infer topic descriptions and their relationships from documents that are indexed by a network's entities and/or entity pairs.

While the above models provide useful methods for discovering latent topics from networked documents, they each generally require that one's documents be fully indexed with relevant network information. This is unsurprising, as a primary goal of these models is to recover latent topics from networked texts. However, for the DoD corpus, the vast majority of our (12-sentence sequence) documents are not associated with specific network entities, entity pairs, or clusters. Indeed, only a small proportion of these documents contains relevant network information, which effectively precludes us from using the models discussed above to extract topic-based information concerning our network's underlying strategies and tactics. This is for two interrelated reasons. First, subsetting out *only* network-associated documents from our corpus for use in the above models would leave us with a sample size that is far too small for tractable topic model estimation. Second, subsetting our sample to only these network-associated documents would also sacrifice the wealth of information that is contained in our nonnetwork-associated documents. While these latter documents do not contain information directly associated with our network's features, they *do* contain extensive information pertaining to the latent topics underlying our corpus on the whole and thus help us to obtain more accurate estimates of our topics of interest.

Accordingly, we favor a more recently developed approach for the incorporation of external document-level information and structure into unsupervised topic model analyses known as the structural topic model (STM Roberts et al. 2014). The STM estimates latent topics in a similar hierarchical manner to that of the general discussion provided earlier, while also incorporating document-level information via external covariates into one's prior distributions for document topics or topic words. In this manner, one can use the STM to not only identify a set of shared latent topics across a corpus but also to evaluate potential relationships between document-level covariates and the prevalence of a given topic within and across documents. As such, the STM has been effectively used to estimate the effects of survey–respondent

characteristics upon variation in respondents' open-ended responses (Roberts et al. 2014) as well as the effects of country-level characteristics (e.g., regime type) upon the topical attention of U.S. State Department Country Reports on Human Rights Practices (Bagozzi and Berliner 2017). For our application, the STM's advantages intuitively lie in its ability to incorporate structural network information—namely, group tie presence and cluster member presence—as binary predictors of variation in attention toward different radical protest strategies and tactics across documents. This, in turn, allows us to estimate a set of topics across all documents and then to evaluate whether the presence of a given group tie (or cluster member) significantly increases the attention dedicated to a given topic.<sup>15</sup> When it does, we interpret this as evidence for shared strategies or tactics between groups within our network.

We specifically model our 12-sentence sequence document corpus as a function of the aforementioned external covariates using the *stm* version 1.1.3 R package (Roberts et al. 2015). As is the case for most unsupervised topic models, we must explicitly choose the number of topics to be estimated within the STM. Roberts et al. (2015) note that “[t]here is no right answer to the appropriate number of topics. [. . .] For small corpora (a few hundred to a few thousand) 5-20 topics is a good place to start.” As our corpus contains 3,210 documents, we rely on the above suggestions—as well as on the topics identified in similar analyses of international social texts (e.g., Bagozzi and Berliner 2017)—to select a topic number of 15 for our primary STM analysis. We further validate our choice of a 15-topic model both qualitatively and quantitatively in Online Supplemental Materials. In brief, we first show qualitatively in our Online Appendix that the topwords from our final 15-topic STM (after addressing the multimodality issues discussed below) are similar, but more relevant and interpretable than those identified in comparably estimated 10- and 20-topic models. Second, and in keeping with the topic number selection routines employed by Bagozzi and Berliner (2017), we then quantitatively derive aggregate exclusivity and semantic coherence metrics for each of our 10-, 15-, and 20-topic models so as to compare the performance of each candidate model more objectively. In this context, exclusivity quantifies how exclusive one's topwords are to each topic based upon a word's relative probabilities of association across topics, while semantic coherence measures the relative co-occurrence of our topics' topwords across our corpus and thus provides a sense of how internally consistent the topics are within each candidate model (Roberts et al. 2014). As discussed in our Online Appendix, we find that our 15-topic STM exhibits the best *joint performance* along both semantic coherence and exclusivity dimensions of comparison, whereas the 10- and 20-topic STMs perform

noticeably on at least one of these two metrics. We interpret these findings as further evidence in support of our choice of a 15-topic STM within our primary analysis.

Finally, in order to address multimodality concerns in the estimation of our final 15-topic model, we follow Roberts et al. (2014) and estimate a set of 50 separate 15-topic STMs, while employing different starting parameters for each initialization,<sup>16</sup> and store exclusivity and semantic coherence metrics for each model. We select a single model from this set of 50 models that maximizes the coherence and exclusivity of our topic word vectors. Using the results of this model, we examine the sets of 20 topwords that best characterize each identified latent topic. After establishing the semantic meaning of our 15 identified latent topics according to these topwords, we then derive a series of postestimation quantities that allow us to examine how topical prevalence varies in relation to group ties and group clusters.

## Topics

This section discusses the topics that we identify as underlying the DoD corpus. Recall that our STM identifies the 15 topics that best characterize our preprocessed 12-sentence sequence documents. Each topic represents an underlying word distribution, wherein each word in our corpus is given a posterior probability of assignment to that topic. For these word vectors, and consistent with Roberts et al. (2014), we derive the words most associated with each topic according to frequency exclusivity scoring metrics, which ensures that our reported topwords correspond to the word stems that are both most frequently assigned to a given topic and most exclusive in their assignment to that topic. It is up to us to substantively interpret the meaning of each topic based upon these probabilistic word assignments. To do so, we draw upon the 20 most frequent and exclusive words for each topic and present our 15 topics of interest—including our labels for each topic, the topic's (stemmed) topwords, and the topic's reference number—in Figure 4.

By and large, the 15 topics identified by the STM appear to correspond to meaningful constructs. We group the 15 identified topics into three general categories for discussion: “ideology,” “tactics,” and “general discourse.” Ideology represents the largest of these topic groupings and includes topics 2–4, 6, 8, 12, and 15. In line with our earlier discussions of the overarching goals of the DoD publication *Earth First! Journal 2006*, topic 2 is labeled Group Identity Debates and exhibits topwords pertaining to the ever-changing environmental perspectives and overarching goals of the UK EF! movement and debates therein. Topic 4 is labeled Eco-Literature and

Topic	Top 20 Words	Labels
1	one, made, anoth, everi, side, time, hand, mani, turn, left, open, togeth, enough, around, given, reach, great, strong, run, join	Inspirational Language
2	role, polit, organis, ideolog, non, activ, idea, movement, radic, mainstream, question, organi, opinion, individu, media, violent, problem, engag, potenti, posit	Group Identity Debates
3	oil, compani, mine, crop, farmer, fish, indigen, papua, corpor, generet, govern, zapatista, shell, bougainvill, peasant, mexico, engin, industri, western, indonesian	Neocolonialism
4	book, isbn, publish, box, http, zine, magazin, guid, press, copi, write, articl, send, email, read, page, list, web, review, mail	Eco-Literature
5	women, law, case, footbal, men, legal, game, evid, court, school, terrorist, privat, famili, properti, terror, intellig, record, team, agenc, custom	News and Culture
6	speci, habitat, restor, wildlif, peat, bird, ecosystem, biodivers, plant, soil, extinct, woodland, highland, conserv, hotspot, moor, natur, garden, biolog, wild	Species Conservation
7	let, away, will, back, voic, danger, keep, put, lake, look, etc, earth, thank, mother, frst, ever, fill, eye, die, sound	General Concern
8	forest, protect, log, water, environment, mountain, dam, north, nativ, area, east, island, region, rainforest, limber, ago, urban, northern, land, river	Land Conservation
9	police, arrest, cop, crowd, march, banner, vehicl, window, offic, smash, demo, confer, riot, hour, demonstr, mask, bank, pragu, car, camera	Violent Protest
10	action, direct, campaign, group, involv, network, sabotag, anti, reclaim, tactic, meet, success, event, sab, act, target, rts, opencast, media, aim	Direct Action/Ecolage
11	camp, evict, site, road, quarri, twyford, tunnel, sit, council, tree, climb, squat, set, hill, fence, hous, build, tarmac, built, trash	Occupation/Camps
12	cultur, must, societi, exist, life, learn, live, desir, civilis, understand, human, relationship, skill, can, experi, planet, process, sens, domin, alien	Sustainable Societies
13	pirat, prison, sentenc, black, panther, bomb, murder, jail, ship, imprison, kill, shepherd, sent, africa, gun, white, death, frame, tbi, trial	International Terror
14	know, dont, realli, want, thing, that, think, lot, someth, get, someone, sure, there, say, just, bad, they, didnt, ask, thought	Admonishments
15	capit, capitalist, labour, revolut, class, struggl, global, economi, union, counter, worker, social, econom, elit, resist, globalis, autonomi, wage, spanish, market	Anti-Capitalist Left

Figure 4. Topwords for 15-topic model.

encompasses topwords pertaining to the *publication* and *dissemination* of radical environmentalist viewpoints. The remaining ideology-based topics clearly identify ideological issues that environmentalists and radical leftists are either for (topic 6: Species Conservation; topic 8: Land Conservation; and topic 12: Sustainable Societies) or against (topic 3: Neo-Colonialism; topic 15: Anti-Capitalism). We take the identification of these highly plausible ideological topics as evidence to suggest that our STM is performing as expected for the corpus as hand.

The next grouping of topics is of most interest to our analysis and encompasses a number of radical protest tactics and strategies. In full, we identify four topics in Figure 4 that appear to pertain to such tactics: topic 9: Violent Protest; topic 10: Direct Action/ECOTAGE; topic 11: Occupation/Camps; and topic 13: International Terror. The Violent Protest topic encompasses topwords related to mass protest (“crowd,” “march,” “demonstr,” and “banner”) as well as to the potential violent consequences of such activities (“polic,” “arrest,” “cop,” “smash,” and “riot”). *Direct action/ecotage* contains both “direct” and “action” in its topwords in addition to “sabitag,” “act,” and “target.” The *occupation/camps* topic appears to relate primarily to tactics involving the semipermanent occupation of areas for protest or obstruction and includes words related to the strategy of occupation itself (“camp,” “evict,” “climb,” and “squat”) and words related to the areas that were typically targeted by this strategy in the UK during this time period (“road,” “quarri,” “twyford,” “tunnel,” and “tarmac”). Finally, *international terror* captures illegal international activities (“bomb,” “murder,” and “kill”) involving environmental and/or leftist radicals (“black,” “panther,” “shepherd,” and “ship”) and their aftermath (“prison,” “sentenc,” “jail,” “fbi,” and “trial”). As such, our documents and STM approach have together identified four fairly coherent, yet largely distinct, *tactics* used by radical environmental (and leftist) groups.

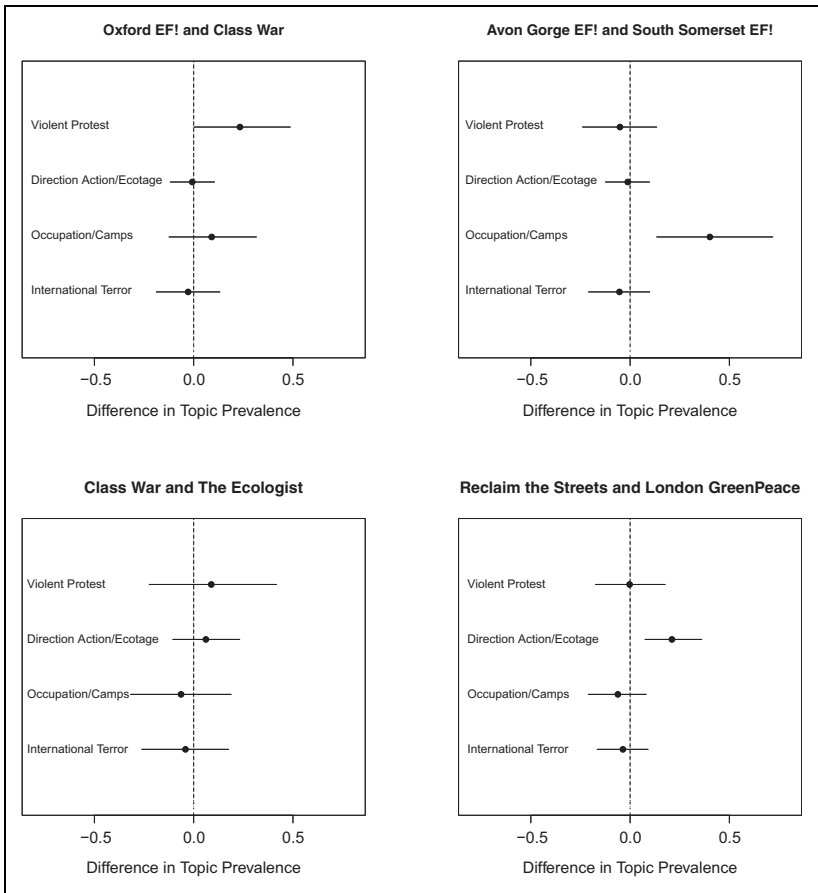
In addition, four of the latent topics in Figure 4 appear to relate to more general (environmental) media discourse. Topic 1 is fairly ambiguous and is labeled Inspirational Language. It contains a range of topwords that we suspect are intended to communicate optimism (“great,” “strong,” and “left”) and inclusiveness (“mani,” “togeth,” and “join”). Topic 5: News and Culture encompasses current events that are likely of general interest to DoD’s UK environmentalist readership, including family-related issues (“men,” “women,” and “school”), sports (“footbal,” “game,” and “team”), and legal/terrorism updates (“terrorist,” “intellig,” and “court”). Topics 7 and 14 are labeled General Concern and Admonishments and appear to, respectively, voice (i) overarching environmental concerns for the survival

of the planet and (ii) criticisms of those that do not support the (radical) environmentalist viewpoint. While we remove stopwords during preprocessing, topic 14 in particular seems to contain a large number of words that could be considered stopwords. We refrain from removing these additional stopwords in a post hoc fashion, given that they are not included in the primary English-language stopwords list that is commonly used in STM analyses (e.g., Bagozzi and Berliner 2017; Roberts et al. 2015, 2014). Finally, while we rely on the stopwords to define our topics above, note that we also use the STM to identify a sample of 10 highly representative documents for each topic, and have used these sample documents to qualitatively guide our topic labeling efforts.

### *Shared Group Tactics*

Focusing on the four tactic-based topics discussed above, we next evaluate whether these topics vary systematically in relation to the presence or absence of specific group ties and whether such variation is consistent with our extant theoretical expectations. To do so, we make use of our group-pair STM covariates—which indicate whether (or not) a connected group pair co-occurs within a given document—and specifically estimate the effect of a 0-to-1 change in each group pair's presence in a document on that document's attention to each tactic-based topic. We highlight and discuss a selection of our specific findings in this regard below, before summarizing our more general findings for all connected component group pairs in the identified network.

Figure 5 presents three sets of estimated effects—which correspond to three specific group pairs—to illustrate our findings pertaining to the association of particular groups with certain tactics and variation therein. Turning first to the pairing of Oxford EF! and Class War in Figure 5, we find that this group pair is positively associated with the use of violent protest tactics, but not significantly associated with tactics of direct action/ecotage, occupation/camps, or international terror. While we held no prior expectations for the shared tactics involving a pairing of Oxford EF! and Class War, these findings are intuitive. As alluded to above, Class War is a long-running anarchist movement that, among other activities, was known for its coordination and coverage of large-scale violent protests and riots during the 1980s and 1990s<sup>17</sup> Hence, in terms of its ties to an ecological direct action group such as Oxford EF!, it's likely that such ties would encompass coordination on tactics of violent protest as opposed to the more explicitly environmental direct action and protest-camp activities that regional UK EF! groups are more commonly known to participate in.



**Figure 5.** Estimated associations between tactics and group ties.

By contrast, we find in Figure 5 that the Avon Gorge EF! and South Somerset EF! group pair is not significantly related to tactics of violent protest nor to tactics of direct action/ecotage or international terror. However, our findings for the Avon Gorge EF! and South Somerset EF! groups *do* suggest that these two groups are more likely to be jointly associated with tactics of occupation/camps-based protests: The co-occurrence of these two groups in a given document is estimated by our STM to be associated with a roughly 40 percent increase in attention to occupation/camps-type tactics. While EF! groups are most commonly viewed as coordinating over direct

action and ecotage-type tactics, our finding in this case is again unsurprising. Indeed, one of the UK EF! movements' longest running campaigns during the 1990s was its use of occupation camps to protest the Somerset-based Whatley Quarry, which most notably involved the two regional EF! groups examined here (DoD 1996; Schneus 1995). As such, this result suggests that our tactic-based findings do indeed correspond to actual strategies of coordination between *specific* groups, rather than to more arbitrary overlapping associations of the individual (and unique) tactics associated with pairs of co-occurring groups within the DoD text.

Recall that some groups within our UK radical environmental groups list (and group network clusters) are generally known to *not* partake in radical protest strategies. We next seek to evaluate whether our proposed approach correctly identifies such groups as *nonradical*, even when one examines the ties between nonradical groups and more radical groups within an assigned group cluster. One such case is *The Ecologist*, a mainstream UK environmental journal that was nevertheless listed as a key UK-group contact within some issues of DoD and was earlier assigned to cluster 2 alongside more violent and radical groups such as Class War. The third subfigure in Figure 5 accordingly examines the associations of the Class War and *The Ecologist* group pair with each of our four identified protest tactics. As one would expect, we find in this case that the Class War and *The Ecologist* group pairing is not significantly associated with *any* of the four identified protest tactics. This implies that, within DoD, these two groups are not associated along radical protest dimensions and that our proposed method is capable of avoiding false positives with respect to the association of radical protest strategies with specific (nonradical) group pairs.

The final subfigure in Figure 5 highlights a group pair—that of Reclaim the Streets and London GreenPeace—that is positively associated with the direct action/ecotage tactic. This finding is understandable, given the extensive ties between these two groups<sup>18</sup> and the direct action agenda of London GreenPeace during the period under analysis. Moreover, whereas Reclaim the Streets is most commonly characterized as pursuing large-scale nonviolent protests and “street parties,” evidence suggests that these events are often cover or magnets for ecotage-oriented tactics as was widely observed during Reclaim the Street's guerilla gardening actions in Parliament Square on May Day 2000, as well as, for example, “during the 1996 Reclaim the Streets Party on the M41 (urban freeway) in West London, [wherein] the skirts of giant pantomime dames on stilts were used as cover for activists using pneumatic drills to break up the road” (Plows et al. 2004:205). Therefore, in sum, our Figure 5 findings highlight the useful information that can be extracted from our network using

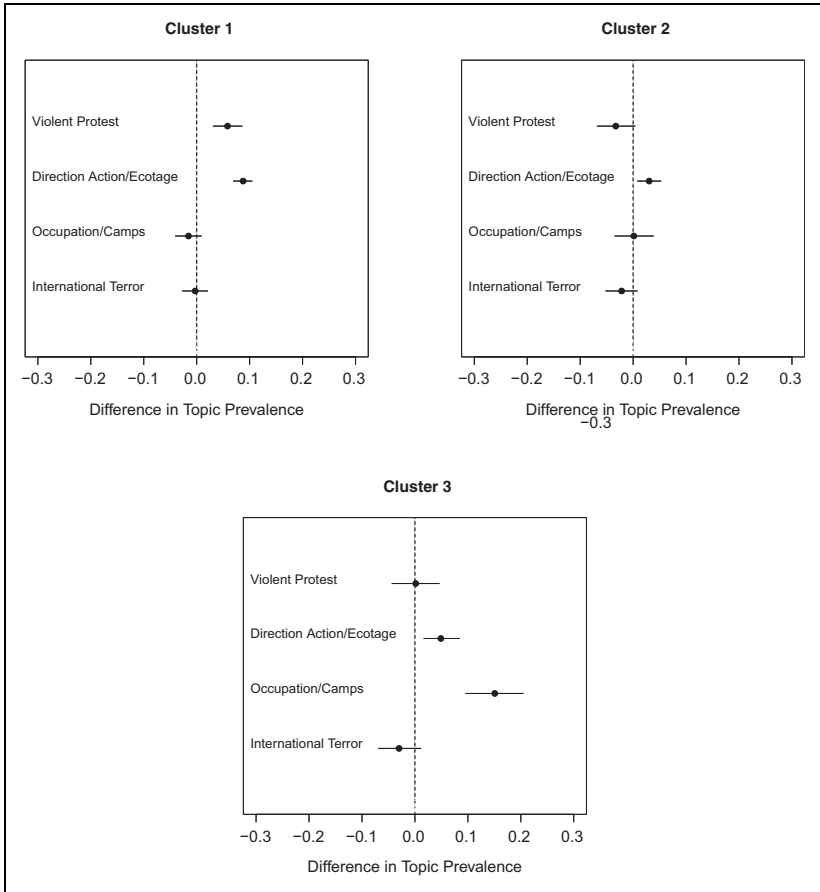


STMs and associated co-occurrence information. While these findings are far from definitive, extant theory suggests that they are consistent with our understanding of past UK environmental group behaviors, thereby highlighting and validating the utility of this approach.

More generally, we find that a very large number of the group pairs within the connected component of our network are significantly and associated with the direct action/ecotage topic, though we do not report each figure here to save space. This frequent association with direct action/ecotage is in line with extant studies and theory, as direct action and ecotage-type strategies are by and large the primary tactic used by the UK EF! groups examined here. At the same time, we find that a much smaller number of our overall connected component group pairs are significantly associated with either violent protest or occupation/camps. Furthermore, virtually no group pairs were significantly and positively associated with the international terror topic. Given our focus on UK rather than international extremist groups, and the predominately nonterrorist designations of the radical leftist groups in our sample, this latter null finding is reassuring. Nevertheless, an examination of whether international (environmental) extremist groups (which are also listed in the appendices to DoD publication) exhibit similar patterns of (non)association with these four tactics would make for an interesting future extension of the present analysis.

### *Cluster Tactics*

We next evaluate the extent to which our three connected component network clusters exhibit similar variation vis-à-vis the identified protest-strategy topics. These estimated effects, again with 95 percent confidence intervals, appear in Figure 6. Beginning first with cluster 1, we find in Figure 6 that the connected component groups included within this cluster are significantly more likely to be associated with tactics of violent protest and direct action/ecotage but are no more or less likely to be associated with the radical protest tactics of occupation/camps or international terror. The former findings suggest an increase in topical attention of roughly 5–10 percent per document and is highly consistent with our expectations. As discussed above, the groups included within cluster 1 tend to have high centrality within our overall network, are often located in the London area, each share a history of joint participation in large-scale omnibus protests and can be generally thought of as being involved in a diverse cross-section of radical leftist issues and concerns (e.g., Reclaim the Streets, Nd.c). We hence anticipated that these groups had been clustered together based upon their frequent



**Figure 6.** Estimated associations between tactics and clusters.

coordination over (violent) protest and direction action campaigns but not international terrorism or protest-camp activities—which is confirmed in Figure 6. This accordingly suggests that our clustering strategy is revealing at least some commonly shared group tactics.

Our findings for cluster 2 reinforce these conclusions. Recall that, while virtually all groups in cluster 2 are environmentalist in nature, some are direct action-oriented groups (Norfolk EF!, The Land is Ours), while others are predominately focused on reporting and disseminating information on environmentalism and environmental protest (*The Ecologist*, Genetic

Engineering Network, and Class War). We therefore did not expect the groups in Cluster 2 to uniformly be partaking in violent protest activities but rather anticipated that the active groups in cluster 2 (Norfolk EF! and The Land is Ours) would orient this cluster's tactics toward the strategies most commonly used by these two groups, specifically direct action/ecotage. In Figure 6, we indeed find that cluster 2 is significantly more likely to be associated with the direct action/ecotage topic (though the effect size is small in magnitude), but is perhaps less likely to be associated with violent protest, and is no more or less likely to be associated with occupation/camps or international terror. These findings are consistent with our expectations, assuming that *The Ecologist*, Genetic Engineering Network, and Class War are primarily reporting on and publicizing these direct action strategies and tactics, rather than actively participating in them.

In contrast to clusters 1 and 2, cluster 3 exhibits a positive and significant association with the occupation/camps topic. Specifically, when discussions pertaining to cluster 3's groups appear in DoD's text, we see a nearly 20 percent increase in a document's attention to occupation/camps. Similar to clusters 1 and 2, we also find a positive and significant association between cluster 3 and attention toward our direct action/ecotage topic. At the same time, we find no significant association for international terror or violent protest. Taken together, these findings are consistent with our expectations: This cluster encompasses a large number of regional EF! groups alongside several protest camp-oriented groups that often coordinate with EF! within protest camps or direct actions. A number of anarchist bookstores and community organizing centers, which often serve as organizing venues for EF! and related group activities, also appear in this cluster. We therefore expected that cluster 3's shared tactics would largely fall within the protest camps- and direct action/ecotage-type tactics, as these are the activities that cluster 3's groups are most often and likely to coordinate on—which is indeed the case in Figure 6. Hence, these findings further demonstrate that our combined network and text analysis strategy is able to uncover useful and theoretically consistent information with respect to radical protest tactics and strategies.

The results presented in Figure 6 also reveal a number of novel theoretical insights that together help to sharpen our understandings of social and environmental movements in the United Kingdom. For instance, while we find above that both clusters 1 and 2 are positively associated with our direct action/ecotage topic, we find that cluster 2 is marginally less likely to coordinate on violent protest, whereas cluster 1 is significantly *more* likely than not to coordinate on violent protest. This divergence, alongside the observed

absence of regional EF! groups in cluster 1 (and the presence of these groups in cluster 1) potentially identifies a split in the UK environmental movement at this time, with one subset eschewing violent tactics and another subset embracing violence. While a schism among the specific groups identified here, to our knowledge, has not been identified in past literature, our findings in these regards are consistent with accounts of broader tensions in protest tactics among radical environmental groups in the United Kingdom at this time<sup>19</sup> as well as with characterizations of the UK EF! movement as being “resolutely” nonviolent (Saunders 2007a:112). In these respects, our findings of a positive association with violent protest for cluster 1, but a potentially negative association with this topic for cluster 2, may also suggest that the broader radical leftist and anarchist groups in our sample—which happen to be among the most central groups in our identified network—have a uniquely violence-inducing effect on environmental groups’ protest strategies (i.e., in the case of cluster 1), though further research is needed to fully evaluate this claim.

### **Network Discovery/Classification via Topic Models**

A common problem in network analysis is that of network discovery or network classification of a given relation (see e.g., Wasserman and Faust 1994). This is especially a problem in the case of networks derived from text, such as co-occurrence networks. The methods employed in this article can be used for a two-fold analysis of this issue. First, the topic models can be engaged so as to allow for a qualitative understanding of the co-occurrence relationships (e.g., coordination or communication over direct action or media campaign). Second, these models can be employed as classifiers (here we use classifier in the manner computer scientists refer to the term, for a review, see Vapnik and Vapnik 1998, where a classifier is a method to identify categorical items of interest) to select out relations of interest. Say, for example, that we want to compare networks which exclusively represent the direct action/ecotage relation and/or networks which exclusively represent the eco-literature shared knowledge network relation. To demonstrate how this could be accomplished, we provide two different classification techniques. We derive a network based on the classification of a given relation by its statistical significance (direct action/ecotage), and we also derive a network based on the magnitude of the effect size (eco-literature). In the following subsections, we first describe the classification of each network. We follow these descriptions with a comparison of the two resulting networks. Lastly, we conclude with a statistical analysis of each network and a brief discussion of our substantive findings.

### Classification by Statistical Significance

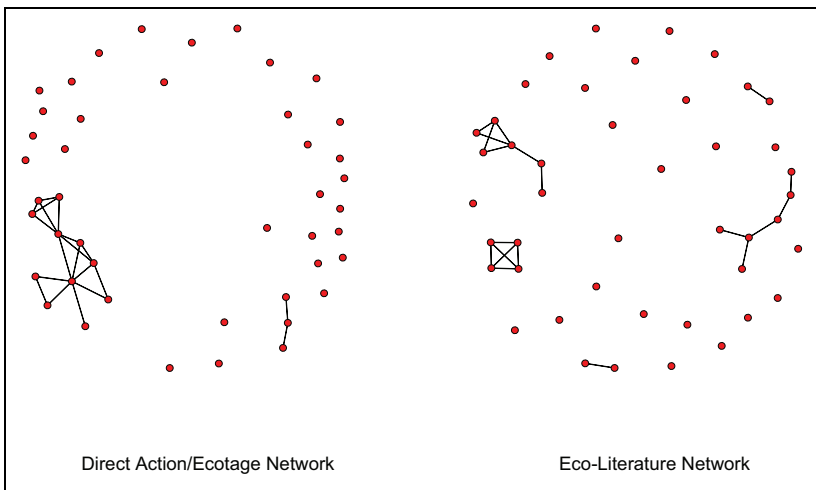
To identify which co-occurring groups in our sample exhibited direct action/ecotage-oriented relation, we iterated through each group pair in our co-occurrence sample and extracted from our primary STM model the mean estimates and bootstrapped 90 percent confidence intervals for the effect of each group pair’s presence in a document upon each topic’s prevalence. We then classified a group pair as a direct action/ecotage group pair based upon whether ( $= 1$ ) or not ( $= 0$ ) that group–group had a positive and statistically significant association with our direct action/ecotage topic. This exercise identified 20 of our 67 total group pairs as direct action/ecotage-oriented pairs. We then chose to limit the bounds of the network to only those actors who were actively engaged (i.e., had a relation in the larger co-occurrence network). Overall, this resulted in a network with a density of 0.021 and we visualize this with a standard network plot in Figure 7.

### Classification by Magnitude of the Effect Size

We followed a similar approach in classifying our group pairs as groups exhibiting ties through our eco-literature topic. Employing the same topic-indexed mean estimates discussed above, we considered a group pair to be an eco-literature pair, if the presence of that group pair in a given document (relative to its absence) was estimated by our primary STM model to exert a positive substantive effect of at least 0.20 (corresponding to at least a 20 percent increase in the prevalence of a chosen topic) with our eco-literature topic. This approach identified 20 group pairs of our 67 total group pairs as being eco-literature-coordinating group pairs. We then chose to limit the bounds of the network to only those actors who were actively engaged (i.e., had a relation in the larger co-occurrence network). Again, we can visualize this with a standard network plot in Figure 7.

### Comparison Direct Action/ Ecotage Network to Eco-Literature Network

The resulting networks are quite different even though their density is the same. The graph correlation (given graphs  $G$  and  $H$ , then  $\text{cor}(G, H) = \frac{\text{cov}(G,H)}{\sqrt{\text{cov}(G,G)\text{cov}(H,H)}}$ , where  $\text{cov}(G,H) = \sum \sum \frac{(Y_{ij}^G - \mu_H)(Y_{ij}^H - \mu_H)}{\binom{N}{2}}$ ; Anderson et al. 1999) is  $-0.02$ . The max degree in the direct action/ecotage network is 7 and is Reclaim the Streets (the same as in the larger co-occurrence network);



**Figure 7.** Direct action/ecotage-classified network and eco-literature-classified network plotted based on Fruchterman and Reingold's force-directed placement algorithm (Butts 2008).

however, the max degree in the eco-literature network is 4 and is *The Ecologist*, which is not unreasonable since we know this is a “group” that is primarily active via the media that it produces. We can further visualize this difference using a graph plot of the two networks, in Figure 7, where we can see that direct action/ecotage-classified network is largely one unified cluster and the eco-literature network is composed of several groups. For example, under the Walk Trap selected by a modularity score clustering algorithm (Gallos 2004)—ignoring isolates—we find three core groups for the direct action/ecotage network and five groups for the eco-literature classified network. Further, we can test substantive network behavior through statistical models of these social networks as discussed in the next subsection.

### *Statistical analysis of Direct Action/ECotage Network and Eco-Literature Network*

Here, we consider probabilistic models of social networks fit by maximum likelihood estimation (MLE). These models are derived from exponential random family models and have thus been referred to as exponential random-family graph models (ERGM). The probability model is typically written as, using the notation of Almquist and Butts (2014, forthcoming),

$$\Pr(Y = y|s, \theta) = \frac{\exp(\theta^T s(y))}{\sum_{y' \in Y} \exp(\theta^T s(y'))} I_Y(y), \quad (1)$$

where  $\Pr(\cdot)$  is the probability mass of its arguments,  $Y$  is the support of  $Y$ ,  $y$  is the realized adjacency matrix,  $s$  is the function of sufficient statistics,  $\theta$  is a vector of parameters, and  $I_Y$  is the counting measure on  $Y$ . This model can be interpreted in conditional log odds or as conditional edge probabilities (Handcock et al. 2008). These models can be fit using Markov chain Monte Carlo (MCMC)-MLE methods in the Handcock et al. (2008) software package and have been used extensively in the sociology literature (e.g., Goodreau, Kitts, and Morris 2009; Papachristos, Hureau, and Braga 2013; Wimmer and Lewis 2010). Here, we employ model selection procedures through the use of BIC criterion, which can be interpreted as a first-order hypothesis such as that performed in Almquist and Butts (2014). In the following subsections, we will analyze each network separately and interpret the findings substantively. However, we first will discuss a metric employed to control for topic similarity in the exponential random-family graph (ERG) models.

*Cosine similarity metric.* This metric, which corresponds to a cosine similarity measure for the group pairs in our sample, allows us to control for the overall topical similarity of each group pair ( $i, j$ ) within our ERGM. To create this measure, we must first recover a vector of group (rather than group pair) specific topic measures as input quantities. We do so by extracting a set of group-specific occurrence measures in a similar fashion to the manner in which we extracted our group-pair co-occurrence indicators from our 12-sentence sequence documents. We then estimate a new 15 STM that included these group-specific document occurrence indicators rather than our group co-occurrence indicators. The topics obtained from this modified STM are virtually identical to those discussed above and appear in Online Supplemental Materials. With these group-occurrence STM estimates in hand, we use our STM model's posterior probabilities of topical assignment to classify our 12-sentence sequence documents based upon their most "dominant topic" ( $f \in 1 : 15$ ), which we in turn use to generate the proportion of each group's associated documents ( $x_f^i$ ) arising from these 15 subsets of topic indexed documents. These input quantities allow us to fully derive a "shared topic" cosine similarity metric for each unique group pair in our sample in a

manner consistent with the cosine approach used by Fafchamps, Leij, and Goyal (2010), where for groups  $i$  and  $j$ , we can define cosine similarity as

$$w^{ij} = \frac{\sum_f x_f^i x_f^j}{\sqrt{(\sum_f (x_f^i)^2)(\sum_f (x_f^j)^2)}}, \quad (2)$$

where values range from 0 to 1, with lower values denoting groups that never co-occur in documents of the same dominant topic, and groups equal to 1 correspond to groups that only co-occur in documents of the same topic. As mentioned above, we use this measure as an edge covariate in our ERG model below so as to control for overall topic similarity between groups.

*Analysis of direct action/ecotage network.* In line with the typical ERGM (and larger statistical model literature), we begin with an edge model (or random graph model; Bayesian information criterion; BIC = 200.69) and then extend it to encompass more complex effects, such as edgewise-shared partners (a measure of clustering, Hunter and Handcock 2012, BIC = 197.42), and then follow up by controlling for topic similarity through our cosine similarity metric. Here, we follow Hunter and Handcock (2012) and use the geometrically weighted edgewise-shared partner distribution (GWESP)—a reparameterization of a statistic of Robins et al. (2007)—which is an alternative approach to counting triangles and can be interpreted as a measure of clustering. It can be considered in comparison or as an alternative measure to the census of triangles or the clustering coefficient for a graph (Hunter and Handcock 2012). Handcock et al. (2008) has implemented a parametric form of the GWESP statistic and the entire model is fit as a curved exponential family via MCMC-MLE, with convergence assessed using the Raftery and Lewis diagnostic (Handcock et al. 2008). We find that the lowest BIC score at this point is in the model which includes the cosine similarity metric with a BIC of 150.14 (see Table 3). We then follow Hunter, Goodreau, and Handcock's (2012) advice and examine the goodness-of-fit statistics (Figure 8) and find that on all metrics we seem to be doing quite well. We do notice that we could do a bit better on two-edgewise-shared partners and degree of 2 and 3 actors statistics, and this can be managed by adding in terms for two-edgewise-shared partners, degree 2, and degree 3; however, this lowers the BIC score and so we choose to interpret the original full model. The model is composed of an edge statistic (baseline interaction effect), a GWESP term (clustering/triadic closure), and the cosine similarity term (the metric of topic level relation). In Table 3, the baseline log odds of a relation is quite low at



**Table 3.** Output for ERG Model From Statnet (Handcock et al. 2008) for the Direct Action/Ecotage-classified Network.

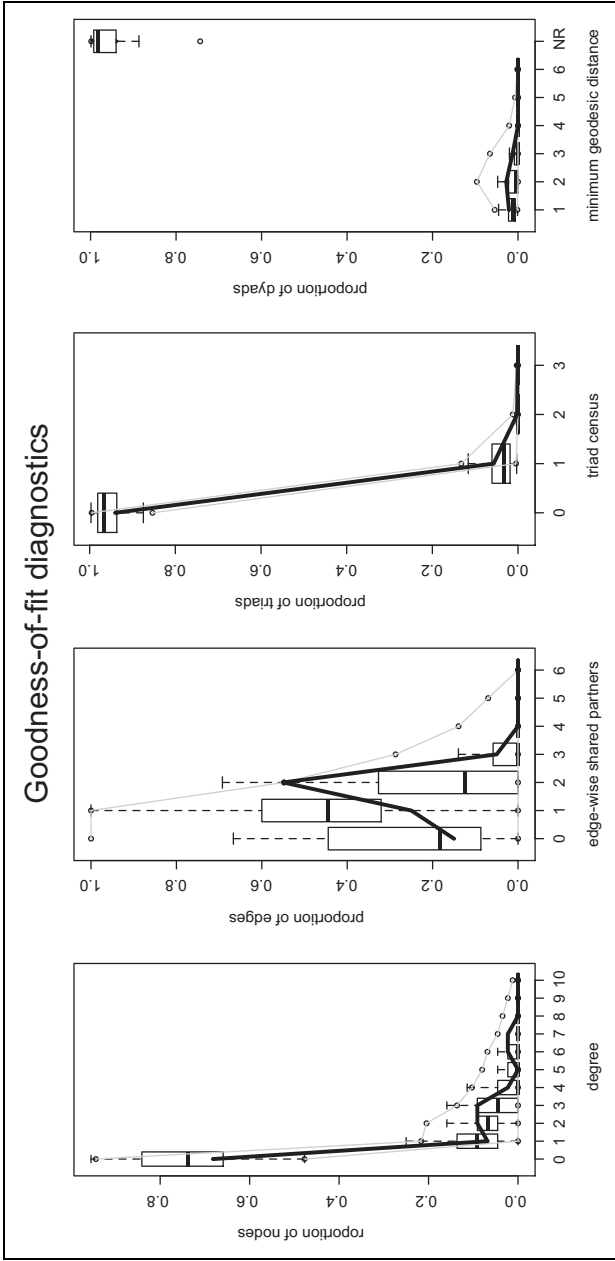
Parameters	Estimate	SE	p Value
Edges	-6.54928	.65565	< 1e - 04***
GWESP	2.63581	.54766	< 1e - 04***
GWESP( $\alpha$ )	0.34606	.09863	0.000472***
CosineMetric	1.79265	.56282	0.001495**

Note: BIC = 150.8722. SEs are as discussed by Hunter et al. (2012) and is based on the inverse Hessian at the approximated MLE and the p values for the parameters are obtained from t tests, with the data degrees of freedom taken to be the number of edge variables. GWESP = edgewise-shared partner distribution; SE = standard error; MLE = maximum likelihood estimations.

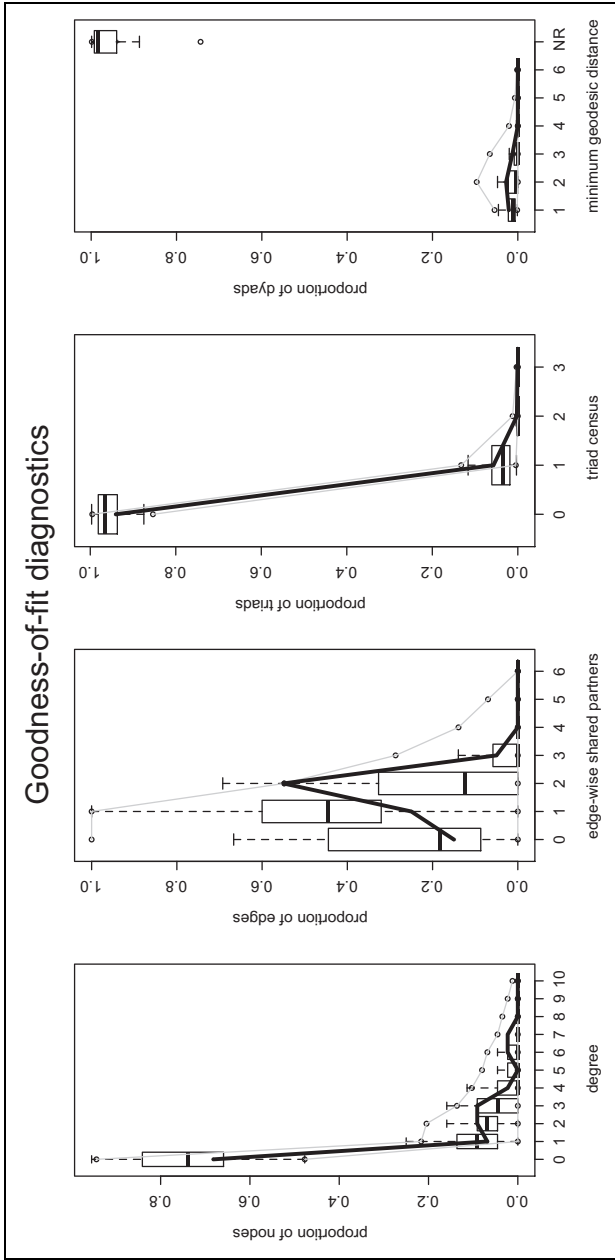
Significance codes:  $p \leq 0$  " \*\*\*";  $p < 0.001$  "\*\*";  $p < 0.01$  "\*";  $p < 0.05$  "."

-6.54, but each edgewise-shared partner results in a large increase, and a shared topical relation results in an increase of 1.79.

*Analysis of eco-literature network.* Again, in the typical fashion, we begin with a random graph model and then extend it to encompass more complex effects, such as edgewise-shared partners (a measure of clustering, Hunter and Handcock 2012) and then extend this by controlling for topic similarity through our cosine similarity metric. The GWESP statistic can be interpreted as before. Because the clustering in this network appears to be more localized than the direct action/ecotage network, we also employ a statistic for local triangles conditioned on the larger co-occurrence network (Schweinberger and Handcock 2015). We again follow the model building/initial hypotheses testing formulation through BIC selection criterion (following the logic of Almquist and Butts [2014]). We find that the lowest BIC score at this point is in the full model (BIC scores: 200.68, 255.6126, 251.2582, and 151.388). We then follow Hunter and Handcock’s (2012) advice and look at goodness-of-fit statistics (Figure 9) to find that our models perform commensurately along all metrics examined. We do notice that we could do a bit better on two-edgewise-shared partners and degrees of 2 and 3 actors statistics and this can be managed by adding in terms for two-edgewise-shared partners, degree 2, and degree 3; however, this lowers the BIC score and so we choose to interpret the original full model. The model is composed of an edge statistic (baseline interaction effect), a GWESP term (clustering/triadic closure), the cosine similarity term (the metric of topic level relation), and a statistic for local triangles (LocalTriangle). In Table 4, the baseline log odds of a relation



**Figure 8.** Direct action/ecotage-classified network goodness-of-fit plot based on Hunter et al. (2012).



**Figure 9.** Eco-literature-classified network goodness-of-fit plot based on Hunter et al. (2012).

**Table 4.** Table Output for ERG Model From Statnet (Handcock et al. 2008) for the Eco-literature-classified Network.

Parameters	Estimate	SE	p Value
Edges	-5.2238	0.5050	< 1e - 04***
GWESP	1.1075	0.4809	0.0215*
GWESP( $\alpha$ )	-0.3521	0.7063	0.6182
LocalTriangle	5.2787	1.2761	< 1e - 04***
CosineMetric	1.5247	0.8107	0.0603
BIC 152.8	Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1		

Note: BIC = 152.8. SEs are as discussed by Hunter et al. (2012) and is based on the inverse Hessian at the approximated MLE and the p values for the parameters are obtained from t tests, with the data degrees of freedom taken to be the number of edge variables. GWESP = edgewise shared partner distribution; SE = standard error; MLE = maximum likelihood estimations. Significance codes:  $p \leq 0$  "\*\*\*";  $p < 0.001$  "\*\*";  $p < 0.01$  "\*";  $p < 0.05$  ".".

is quite low at -5.22, but each edgewise-shared partner—with a log-odds effect of 1.1075—results in a large increase, and a shared topical relation results in an increase of 1.52. Notably, the LocalTriangle effect is quite large with a 5.28 log-odds effect for each local triangle an edge is embedded—implying that being in just one local triangle completely overcomes the baseline effect.

### Discussion and Summary of Topic Based Network Classification

We have found that it is possible to use our favored topic model approaches to carefully classify subnetworks from the core co-occurrence network for further analysis and comparison. These methods not only provide for a better qualitative understanding of a given edge without direct human coding but also allow for the identification of salient subnetworks within the larger network without the aid of human coders. Further, we can individually analyze these networks to uncover distinct substantive findings. For example, in doing so above, we found that the direct action/ecotage network can be largely explained through edgewise shared partner clustering and topic similarity, whereas the eco-literature network requires the extra information of local clustering in order to model it correctly. This suggests that there are distinct local groups within the eco-literature network, whereas the direct action/ecotage network appears to be much more cohesive. The latter finding lends support to past anecdotes of high inbreeding among radical environmental groups in the United Kingdom—which previously noted that

insufficient data on radical UK environmental groups limited systematic conclusions in these regards (Saunders 2007b:237)—as well as to findings concerning the tendencies of subversive and radical groups to disproportionately seek out closer and/or more like-minded collaborators when coordinating on potentially illegal or violent actions (Almquist and Bagozzi 2016). These findings also fit within our larger network and topic model analysis, which demonstrated distinct subgroups within the more mainstream groups as compared to the more subversive groups in the network. Hence, altogether, we find that this is a strong and clever technique for improving our understanding of networks and subnetworks through topic models and modern text analysis approaches.

## Summary

This article presents a suite of tools for the automated discovery of radical activist group networks and tactics from raw unstructured text and applies these methods to a collection of radical UK environmentalist publications. While radical groups are often covert in their networks, membership, and tactics, they produce a great deal of text during their efforts to publicize their concerns, communicate with like-minded groups, and mobilize support for their activities. The advent of the World Wide Web, along with more recent advancements in automated text analysis and SNA tools, enables scholars to take advantage of these self-produced texts in novel ways. As we show, such methods not only allow one to uncover detailed network information from unstructured environmentalist texts but can also identify the underlying tactics that are discussed in these texts and help to pair these tactics with one's uncovered network. In these respects, our findings contribute to the growing body of research that has sought to apply SNA tools to environmental advocacy groups (e.g., Ansell 2003; Ackland and O'Neil 2011; Saunders 2007b) and radical social movements (Almquist and Bagozzi 2016; Caiani and Wagemann 2009; O'Callaghan et al. 2013). Furthermore, our UK environmental group application provides scholars with a detailed guide for implementing these techniques and offers a number of insights into the network and tactics of radical leftist and environmentalist groups within the UK.

Specifically, we find that the UK environmental movement of the 1990s and early 2000s, while most commonly associated with the UK EF! organization, is actually embedded within a larger network of radical leftist groups and venues. Surprisingly, the most central members of this network share close ties to the EF! movement, but primarily advocate for a more general set of leftist ideals, including opposition to globalization, the promotion of

worker's rights, and anarchism. This is consistent with characterizations of leftist global social movements' increasing de-environmentalization, and subsequent reorientation toward social justice and economic issues during the late 1990s and onward (Buttel and Gould 2004). With respect to the UK specifically, scholars have often noted the decline of environmental protest in the UK during the late 1990s (Rootes 2000:49), which was shortly followed by the demise of the DoD publication itself. We believe that the highly central, but nonenvironmentalist, radical leftist groups that we identify—and variation in support for *their* respective movements and causes—may help to explain this decline. Additional research, however, is needed to fully explore this possibility. Lastly, we also find sound evidence to suggest that UK leftist groups pursue a range of coherent protest tactics and that these tactics vary considerably depending on the group pairs and clusters that one examines within the larger network. It is hoped that this variation will accordingly offer a new avenue for future theoretical tests concerning the timing, circumstances, and dynamics by which radical groups choose to favor one protest tactic over another.

### Authors' Note

Author list is composed in alphabetic order to denote equal authorship. All data and code are made available and can be found at doi:10.7910/DVN/PNK7AB, Harvard Dataverse, V1.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by an ARO YIP #W911NF-14-1-0577.

### Supplemental Material

Supplementary material for this article is available online

### Notes

1. That is, illegal property destruction (e.g., arson or monkey-wrenching/ecotage).
2. The organized occupation of physical spaces (e.g., sit-ins, peace camps, or tree-sits) to protest, call attention to, and/or obstruct an activity deemed threatening to a space or issue area.

3. Including classic social network descriptives, as well as more recent statistical tools for the network discovery (i.e., network classification) of a given relation as described in detail.
4. For example, the “Do or Die” (DoD) publication analyzed bears the phrase “The Voice of the British Earth First! Movement” on the cover of issues 1 and 2 and variants of this statement on the covers of later issues.
5. For example, Class War, one of the highly central, nonenvironmentalist groups that we identify in our analysis, had splintered into several smaller regional groups by the late 1990s.
6. We use “tactics” and “strategies” interchangeably.
7. Unfortunately, while low in quality, these pdf images are the best records of issues 2–4 available, with a Talon Conspiracy (2014) archivist noting that some of these specific issues were “very rare” and lamenting that earlier copies of this publication had been stolen from him or her by a “joint terrorism task force.”
8. To facilitate future research with these documents, this article also provides *all* original, optical character recognized, and web-scraped versions of the above files (in pdf, rtf, and txt forms) as Online Supplemental Materials.
9. That is, we use fixed windows to construct our sentence sequences, rather than rolling windows. We demonstrate that our co-occurrence and network results are robust to our choice of “starting sentence” for creating these sentence sequences in Online Supplemental Materials.
10. We do, however, continue to omit the actual contact lists included at the end of each DoD issue from the construction of our 12-sentence sequence documents for this step, as including this text within our document sample for co-occurrence identification would lead to a large number of false positives.
11. Graph density measures how many edges are in set  $E$  compared to the maximum possible number of edges between vertices in set  $V$  as defined earlier; in the case of an undirected graph with no self-ties, this can be computed as  $\left(|E| / \left(2 \cdot \binom{n}{2}\right)\right)$ , where  $|\cdot|$  is the operator that provides the size of a given set.
12. A metric distinguishing between the core of the network and the periphery of the network.
13. Indeed, Reclaim the Streets led at least 27 UK actions between May 1995 and August 1997 alone (Reclaim the Streets, Nd.b), often in coordination with other groups, such as the Liverpool Dockers (Reclaim the Streets, Nd.a).
14. Note that we maintained capitalization of group names during our co-occurrence procedure to ensure that mentions of the term “class war” were not treated as occurrences of the group “Class War.”
15. In our Online Supplemental Materials, we verify that our findings concerning topics, group ties, and cluster members are each robust to the inclusion of an

additional “time” covariate that measures each document’s corresponding DoD issue.

16. Note that the methods proposed by Roberts, Stewart, and Tingley (2016) provide an alternative model selection approach to that of our routine, which may prove especially useful to future researchers analyzing larger corpora than the corpus discussed here.
17. Such as, for example, the Trafalgar Square Poll Tax Riots.
18. Indeed, the latter claims to have helped to initiate the former (Mcsplight, Nd).
19. See, for example, Saunders (2007a:112) and the citation of Rootes (2002:33) therein as well as passages of DoD itself (e.g., DoD 1992:4-5).

## References

- Ackerman, Gary A. 2003. “Beyond Arson? A Threat Assessment of the Earth Liberation Front.” *Terrorism and Political Violence* 15:143-70.
- Ackland, Robert and Mathieu O’Neil. 2011. “Online Collective Identity: The Case of the Environmental Movement.” *Social Networks* 33:177-90.
- Adams, Jimi and Ryan Light. 2014. “Mapping Interdisciplinary Fields: Efficiencies, Gaps and Redundancies in HIV/AIDS Research.” *PLoS One* 9:1-13.
- Almquist, Zack W. 2012. “Random Errors in Egocentric Networks.” *Social Networks* 34:493-505.
- Almquist, Zack W. and Benjamin E. Bagozzi. 2016. “The Spatial Properties of Radical Environmental Organizations in the UK: Do or Die!” *PLoS One* 11:1-19.
- Almquist, Zack W. and Carter T. Butts. 2013. “Dynamic Network Logistic Regression: A Logistic Choice Analysis of Inter-and Intra-group Blog Citation Dynamics in the 2004 US Presidential Election.” *Political Analysis* 21:430-48.
- Almquist, Zack W. and Carter T. Butts. 2014. “Logistic Network Regression for Scalable Analysis of Networks with Joint Edge/Vertex Dynamics.” *Sociological Methodology* 44:273-321.
- Almquist, Zack W. and Carter T. Butts. Forthcoming. “Dynamic Network Analysis with Missing Data: Theory and Methods.” *Statistica Sinica*. doi:10.5705/ss.202016.0108.
- Almquist, Zack W., Emma S. Spiro, and Carter T. Butts. 2016. “Shifting Attention: Modeling Follower Relationship Dynamics among US Emergency Management-related Organizations during a Colorado Wildfire.” Pp. 93-112 in *Social Network Analysis of Disaster Response, Recovery, and Adaptation*, edited by A. J. Faas and Eric Jones. Philadelphia, PA: Elsevier.
- Anderson, Brigham S., Carter Butts, and Kathleen Carley. 1999. “The Interaction of Size and Density with Graph-level Indices.” *Social Networks* 21:239-67.
- Ansell, Christ. 2003. “Community Embeddedness and Collaborative Governance in the San Francisco Bay Area Environmental Movement.” Pp. 123-44 in *Social*



- Movements and Networks: Relational Approaches to Collective Action*, edited by Mario Dani and Doug McAdam. Oxford, England: Oxford University Press.
- Bagozzi, Benjamin E. and Daniel Berliner. 2017. "The Politics of Scrutiny in Human Rights Monitoring: Evidence from Structural Topic Models of U.S. State Department Human Rights Reports." *Political Science Research and Methods*:1-17. doi: 10.1017/psrm.2016.44.
- Bagozzi, Benjamin E. and Philip A. Schrodt. 2012. "The Dimensionality of Political News Reports." Paper presented at the EPSA Annual Convention, April 3–6, Berlin, Germany.
- Barberá, Pablo, Richard Bonneau, John T. Jost, Jonathan Nagler, and Joshua Tucker. 2014. "Is There Anybody Out There? The Effects of Legislators' Communication with Their Constituents." Paper presented at the ISA Annual Convention, London School of Economics and Political Science, London, England, September 21, pp. 1-24.
- Bearman, Peter S. and Katherine Stovel. 2000. "Becoming a Nazi: A Model for Narrative Networks." *Poetics* 27:69-90.
- Blaydes, Lisa, Justin Grimmer, and Alison McQueen. 2015. "Mirrors for Princes and Sultans: Advice on the Art of Governance in the Medieval Christian and Islamic Worlds." Working Paper. Accessed September 30, 2015 (<http://aalims.org/uploads/Mirrors%20Princeton.pdf>).
- Blee, Kathleen M. and Kimberly A. Creasap. 2010. "Conservative and Right-wing Movements." *Annual Review of Sociology* 36:269-86.
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3:993-1022.
- Bonacich, Phillip. 1972. "Factoring and Weighting Approaches to Status Scores and Clique Identification." *Journal of Mathematical Sociology* 2:113-20.
- Bonilla, Tabitha and Justin Grimmer. 2013. "Elevated Threat Levels and Decreased Expectations: How Democracy Handles Terrorist Threats." *Poetics* 41:650-69. Topic Models and the Cultural Sciences.
- Brown, Travis. 2012. "Telling New Stories about Our Texts: Next Steps for Topic Modeling in the Humanities." Paper presentation. Accessed September 30, 2015 (<http://mith.umd.edu/topicmodeling/wp-content/uploads/dh2012-slides.pdf>).
- Burris, Val. 2004. "The Academic Caste System: Prestige Hierarchies in PhD Exchange Networks." *American Sociological Review* 69:239-64.
- Burt, Ronald S. 2000. "The Network Structure of Social Capital." *Research in Organizational Behavior* 22:345-423.
- Buttel, Frederick H. and Kenneth A. Gould. 2004. "Global Social Movement(s) at the Crossroads: Some Observations on the Trajectory of the Anti-corporate Globalization Movement." *Journal of World-Systems Research* X:37-66.
- Butts, Carter T. 2008. "Social Network Analysis with SNA." *Journal of Statistical Software* 24:1-51.

- Butts, Carter T. and Kathleen M. Carley. 2005. "Some Simple Algorithms for Structural Comparison." *Computational & Mathematical Organization Theory* 11: 291-305.
- Byrd, Scott C. and Lorien Jasny. 2010. "Transnational Movement Innovation and Collaboration: Analysis of World Social Forum Networks." *Social Movement Studies* 9:355-72.
- Caiani, Manuela and Claudius Wagemann. 2009. "Online Networks of the Italian and German Extreme Right." *Information, Communication & Society* 12:66-109.
- Chalecki, Elizabeth L. 2002. "A New Vigilance: Identifying and Reducing the Risks of Environmental Terrorism." *Global Environmental Politics* 2:103.
- Chang, Jonathan and David M. Blei. 2009. "Relational Topic Models for Document Networks." Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AISTATS), Clearwater Beach, Florida, pp. 81-88.
- Chang, Jonathan, Jordan Boyd-Graber, and David M. Blei. 2009. "Connections between the Lines: Augmenting Social Networks with Text." Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, Paris, France, pp. 169-78.
- Chen, Zhiyuan, Arjun Mukherjee, Bing Liu, Meichun Hsu, Malu Castellanos, and Riddhiman Gosh. 2013. "Leveraging Multi-domain Prior Knowledge in Topic Models." *IJCAI* 13:2071-77.
- Chesters, Graeme and Ian Welsh. 2011. *Social Movements: The Key Concepts*. New York, NY: Routledge.
- Cohen, Aaron M, William R Hersh, Christopher Dubay, and Kent Spackman. 2005. "Using Co-occurrence Network Structure to Extract Synonymous Gene and Protein Names from MEDLINE Abstracts." *BMC Bioinformatics* 6:103.
- Culotta, Aron, Ron Bekkerman, and Andrew McCallum. 2004. "Extracting Social Networks and Contact Information from Email and the Web." Computer Science Department Faculty Publication Series, Paper 33. Accessed September 30, 2015 ([http://scholarworks.umass.edu/cs\\_faculty\\_pubs/33](http://scholarworks.umass.edu/cs_faculty_pubs/33)).
- Davidov, Dmitry, Ari Rappoport, and Moshe Koppel. 2007. "Fully Unsupervised Discovery of Concept-specific Relationships by Web Mining." Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, June 2007, Prague, Czech Republic, pp. 232-9.
- della Porta, Donatella. 1995. *Social Movements, Political Violence, and the State: A Comparative Analysis of Italy and Germany*. Cambridge, England: Cambridge University Press.
- Diani, Mario. 1995. *Green Networks: A Structural Analysis of the Italian Environmental Movement*. Edinburgh, England: Edinburgh University Press.
- Diani, Mario. 2002. "Britain Recreates the Social Movement: Contentious (and Not so Contentious) Networks in Glasgow." Paper for the Contentious Politics and the

- Economic Opportunity Structure Conference, University of Crete, October 17–19, Rethimno, Greece.
- Dietz, Laura, Steffen Bickel, and Tobias Scheffer. 2007. "Unsupervised Prediction of Citation Influences." Proceedings of the 24th International Conference on Machine Learning, ICML 2007, ACM, New York, NY, pp. 233-40.
- DiMaggio, Paul, Manish Nag, and David Blei. 2013. "Exploiting Affinities between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Arts Funding." *Poetics* 41:570-606.
- Do or Die (DoD). 1992. "International Temperate Day of Action."
- Do or Die (DoD). 1996. "Meanwhile Down in the West Country." Pp.17-22.
- Do or Die (DoD). 1998. "No Opencast!" Pp.23-32.
- Downing, John D. H. 2001. *Radical Media: Rebellious Communication and Social Movements*. Thousand Oaks, CA: Sage.
- Earth First! Journal. 2006. "Do or Die Editors Interview." [An interview conducted for the *American Earth First! Journal* and made available on eco-action.org]. Retrieved July 19, 2015 (<http://www.eco-action.org/dod/interview.htm>).
- Eco-Action.org. Nd. "Do or Die: Voices from the Ecological Resistance." *Eco.action.org*. ISSN 1462-5989/ISBN 0-9545662-0-3. Retrieved July 19, 2015 (<http://www.eco-action.org/dod/index.html>).
- Fafchamps, Marcel, Marco J. Leij, and Sanjeev Goyal. 2010. "Matching and Network Effects." *Journal of the European Economic Association* 8:203-31.
- Farrell, Justin. 2015. *The Battle for Yellowstone: Morality and the Sacred Roots of Environmental Conflict*. Princeton, NJ: Princeton University Press.
- Farrell, Justin. 2016. "Corporate Funding and Ideological Polarization about Climate Change." *Proceedings of the National Academy of Sciences* 113:92-97.
- Fitzgerald, Kathleen J. and Diane M. Rodgers. 2000. "Radical Social Movement Organizations: A Theoretical Model." *The Sociological Quarterly* 41:573-92.
- Foreman, Dave and Bill Haywood. 1993. *Ecodefense: A Field Guide to Monkey-wrenching*. Chico, CA: Abzug Press.
- Fortunato, Santo. 2010. "Community Detection in Graphs." *Physics Reports* 486:75-174.
- Freeman, Linton C. 1977. "A Set of Measures of Centrality based on Betweenness." *Sociometry* 40:35-41.
- Freeman, Linton C. 1979. "Centrality in Social Networks Conceptual Clarification." *Social Networks* 1:215-39.
- Freeman, Linton C. 2004. *The Development of Social Network Analysis: A Study in the Sociology of Science*. Vancouver, Canada: Empirical Press.
- Freeman, Linton C. and Cynthia M Webster. 1994. "Interpersonal Proximity in Social and Cognitive Space." *Social Cognition* 12:223.
- Gallos, Lazaros K. 2004. "Random Walk and Trapping Processes on Scale-free Networks." *Physical Review E* 70:046116.

- Goodreau, Steven M, James A. Kitts, and Martina Morris. 2009. "Birds of a Feather, or Friend of a Friend? Using Exponential Random Graph Models to Investigate Adolescent Social Networks." *Demography* 46:103-25.
- Grimmer, Justin. 2012. "A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases." *Political Analysis* 18: 1-35.
- Grimmer, Justin and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21:267-97.
- Handcock, Mark S, David R. Hunter, Carter T. Butts, Steven M. Goodreau, and Martina Morris. 2008. "Statnet: Software Tools for the Representation, Visualization, Analysis and Simulation of Network Data." *Journal of Statistical Software* 24:1548.
- Hoffman, Andrew J. and Stephanie Bertels. 2009. "Who is Part of the Environmental Movement? Assessing Network Linkages between NGOs and Corporations." Working Paper No. 1125, Ross School of Business, Ann Arbor, MI.
- Hunter, David R., Steven M. Goodreau, and Mark S. Handcock. 2012. "Goodness of Fit of Social Network Models." *Journal of the American Statistical Association* 103.481:248-258.
- Hunter, David R. and Mark S. Handcock. 2012. "Inference in Curved Exponential Family Models for Networks." *Journal of Computational and Graphical Statistics* 15: 565-583.
- Jockers, Matthew L. and David Mimno. 2013. "Significant Themes in 19th-century Literature." *Poetics* 41:750-69.
- Joosee, Paul. 2007. "Leaderless Resistance and Ideological Inclusion: The Case of the Earth Liberation Front." *Terrorism and Political Violence* 19:351-68.
- King, Jean. 1987. "A Review of Bibliometric and Other Science Indicators and Their Role in Research Evaluation." *Journal of Information Science* 13:261-76.
- Kitschelt, Herbert P. 1986. "Political Opportunity Structures and Political Protest: Anti-nuclear Movements in Four Democracies." *British Journal of Political Science* 16:57-85.
- Krafft, Peter, Juston Moore, Burce Desmarais, and Hanna Wallach. 2014. "Topic-partitioned Multinetwork Embeddings." Paper presented at the 2014 Northeast Political Methodology Meeting, New York City, USA, pp. 2807-2815.
- Leader, Stefan H. and Peter Probst. 2003. "The Earth Liberation Front and Environmental Terrorism." *Terrorism and Political Violence* 15:37-58.
- Light, Ryan. 2016. "Knowledge in Motion: The Evolution of HIV/AIDS Research." *Scientometrics* 107:1227-48.
- Marsden, Peter V. 1990. "Network Data and Measurement." *Annual Review of Sociology* 16:435-63.

- Marshall, Emily A. 2013. "Defining Population Problems: Using Topic Models for Cross-national Comparison of Disciplinary Development." *Poetics* 41: 701-24.
- Martin, John Levi. 2000. "What Do Animals Do All Day? The Division of Labor, Class Bodies, and Totemic Thinking in the Popular Imagination." *Poetics* 27:195-231.
- Matsuo, Yutaka and Mitsuru Ishizuka. 2004. "Keyword Extraction from a Single Document using Word Co-occurrence Statistical Information." *International Journal on Artificial Intelligence Tools* 13:157-69.
- McCallum, Andrew, Andrés Corrada-Emmanuel, and Xuerui Wang. 2005. "Topic and Role Discovery in Social Networks." Proceedings of the 19th International Joint Conference on Artificial Intelligence, IJCAI'05, Morgan Kaufmann, San Francisco, CA, pp. 786-91.
- McFarland, Daniel A., Daniel Ramage, Jason Chuang, Jeffrey Heer, Christopher D. Manning, and Daniel Jurafsky. 2013. "Differentiating Language Usage through Topic Models." *Poetics* 41:607-25.
- Mcspotlight. Nd. "London Greenpeace: A History of Peace, Protest and Campaigning." ([http://www.mcspotlight.org/people/biogs/london\\_grnpeace.html](http://www.mcspotlight.org/people/biogs/london_grnpeace.html)).
- Miller, Ian Matthew. 2013. "Rebellion, Crime and Violence in Qing China, 1722–1911: A Topic Modeling Approach." *Poetics* 41:626-49.
- Mintz, Beth and Michael Schwartz. 1981. "Interlocking Directorates and Interest Group Formation." *American Sociological Review* 46:851-69.
- Mohr, John W. and Petko Bogdanov. 2013. "Introduction—Topic Models: What They Are and Why They Matter." *Poetics* 41:545-69.
- Mohr, John W., Robin Wagner-Pacifci, Ronald L. Breiger, and Petko Bogdanov. 2013. "Graphing the Grammar of Motives in National Security Strategies: Cultural Interpretation, Automated Text Analysis and the Drama of Global Politics." *Poetics* 41:670-700.
- Mucha, Peter J., Thomas Richardson, Kevin Macon, Mason A. Porter, and Jukka-Pekka Onnela. 2010. "Community Structure in Time-dependent, Multiscale, and Multiplex Networks." *Science* 328:876-78.
- Mullins, Nicholas C., Lowell L. Hargens, Pamela K. Hecht, and Edward L. Kick. 1977. "The Group Structure of Cocitation Clusters: A Comparative Study." *American Sociological Review* 42:552-62.
- O'Callaghan, Derek, Derek Greene, Maura Conway, Joe Carthy, and Pádraig Cunningham. 2013. "Un-covering the Wider Structure of Extreme Right Communities Spanning Popular Online Networks." Proceedings of the 5th Annual ACM Web Science Conference, WebSci'13, ACM, New York, NY, pp. 276-85.
- Padgett, John F. and Christopher K. Ansell. 1993. "Robust Action and the Rise of the Medici, 1400-1434." *American Journal of Sociology* 98:1259-319.

- Papachristos, Andrew V., David M. Hureau, and Anthony A. Braga. 2013. "The Corner and the Crew: The Influence of Geography and Social Networks on Gang Violence." *American Sociological Review* 78:417-47.
- Plows, Alexandra, Derek Wall, and Brian Doherty. 2001. "From the Earth Liberation Front to Universal Dark Matter: The Challenge of Covert Repertoires to Social Movement Research." Paper for the Workshop on Democracy and Extremism. European Consortium for Political Research Joint Sessions, April 6-11, Grenoble, France.
- Plows, Alexandra, Derek Wall, and Brian Doherty. 2004. "Covert Repertoires: Ecotage in the UK." *Social Movement Studies: Journal of Social, Cultural and Political Protest* 3:199-219.
- Purkis, Jonathan. 1998. "Daring to Dream: Idealism in the Philosophy, Organization and Campaigning Strategies of Earth First!" Pp. 197-215 in *To Make Another World: Studies in Protest and Collective Action*, edited by Colin Barker and Paul Kennedy. Avebury, England: Aldershot.
- R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Reclaim the Streets. Nd.a. "If the World of the Liverpool Dockers Seems at Odds with That of the Street Party, Think Again . . ." Accessed September 30, 2015 (<http://rts.gn.apc.org/prop17.htm>).
- Reclaim the Streets. Nd.b. "A People's Festival." Retrieved May 5, 2014 (<https://www.j12.org/rts/rtседin.htm>).
- Reclaim the Streets. Nd.c. "Reclaim the Future 5—RTF 5." Retrieved May 5, 2014 (<http://rts.gn.apc.org/>).
- Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2015. *stm: R Package for Structural Topic Models*. R package version 1.1.0. Accessed July 30, 2015 <http://www.structuraltopicmodel.com>.
- Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2016. "Navigating the Local Modes of Big Data: The Case of Topic Models." Pp. 51-89 in *Computational Social Science: Discovery and Prediction*, edited by R. Michael Alvarez. New York, NY: Cambridge University Press.
- Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Gdarian, Bethany Albertson, and David Rand. 2014. "Structural Topic Models for Open Ended Responses." *American Journal of Political Science* 58:172-8.
- Robins, Garry, Tom Snijders, Peng Wang, Mark Handcock, and Philippa Pattison. 2007. "Recent Developments in Exponential Random Graph (p\*) Models for Social Networks." *Social Networks* 29:192-215.
- Romney, A. Kimball, Susan C. Weller, and William H. Batchelder. 1986. "Culture as Consensus: A Theory of Culture and Informant Accuracy." *American Anthropologist* 88:313-38.

- Rootes, Christopher. 2000. "Direct Action in British Environmentalism." Pp. 1-22 in *Direct Action in British Environmentalism*, edited by Brian Doherty, Matthew Paterson, and Benjamin Seel. London, England: Routledge.
- Rootes, Christopher. 2002. *The Transformation of Environmental Activism*, Final Report to European Commission DG XII on Contract ENV4-CT97-0514. European Commission, 200 pp. ISBN 1-902671-34-1.
- Sánchez-Cuenca, Ignacio and Luis de la Calle. 2009. "Domestic Terrorism: The Hidden Side of Political Violence." *Annual Review of Political Science* 12:31-49.
- Saunders, Clare. 2007a. "Comparing Environmental Movement Networks in Periods of Latency and Visibility." *Graduate Journal of Social Science* 1:109-139.
- Saunders, Clare. 2007b. "Using Social Network Analysis to Explore Social Movements: A Relational Approach." *Social Movement Studies* 3:227-43.
- Schnews. 1995. "Whatley Quarry?! Earth First! 60 Nicked in Superquarry Shutdown!" Issue 52. Accessed August 30, 2015 (<http://www.schnews.org.uk/archive/news52.htm>).
- Schweinberger, Michael and Mark S. Handcock. 2015. "Local Dependence in Random Graph Models: Characterization, Properties and Statistical Inference." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 77:647-76.
- Seel, Ben. 1997. "If Not You, Then Who? Earth First! in the UK." *Environmental Politics* 6:172-8.
- Simi, Pete and Robert Futrell. 2015. *American Swastika: Inside the White Power Movement's Hidden Spaces of Hate*. 2nd ed. London, England: Rowman & Littlefield.
- Simon Wiesenthal Center. 2008. *iReport: Online Terror + Hate: The First Decade*. <http://www.wiesenthal.com/atf/cf/%7BD5FD2AAC1-2ADE-428A-9263-35234229D8D8%7D/IREPORT.PDF>.
- Smith, Brent L. and Kelly R. Damphousse. 2009. "Patterns of Precursor Behaviors in the Life Span of an Environmental Terrorist Group." *Criminology & Public Policy* 8:172-8.
- Spiro, Emma S., Zack W. Almquist, and Carter T. Butts. 2016. "The Persistence of Division: Geography, Institutions, and Online Friendship Ties." *Socius: Sociological Research for a Dynamic World* 2:1-15.
- Talon Conspiracy. 2014. "Do or Die—The Complete Set!" *The Talon Conspiracy*, April 9. Retrieved July 19, 2015 (<http://thetalonconspiracy.com/2014/04/do-or-die-earth-first/>).
- Tangherlini, Timothy R. and Peter Leonard. 2013. "Trawling in the Sea of the Great Unread: Sub-corpus Topic Modeling and Humanities Research." *Poetics* 41: 725-49. Topic Models and the Cultural Sciences.
- Üsdiken, Behlül and Yorgo Pasadeos. 1995. "Organizational Analysis in North America and Europe: A Comparison of Co-citation Networks." *Organization Studies* 16:503-26.

- Vapnik, Vladimir Naumovich and Vladimir Vapnik. 1998. *Statistical Learning Theory*. Vol. 1. New York, NY: Wiley.
- Wall, Derek. 1999a. *Earth First! and the Anti-roads Movement: Radical Environmentalism and Comparative Social Movements*. New York, NY: Routledge.
- Wall, Derek. 1999b. "Mobilising Earth First! in Britain." *Environmental Politics* 8:81-100.
- Wasserman, Stanley and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. Vol. 8. Cambridge, UK: Cambridge University Press.
- Wimmer, Andreas and Kevin Lewis. 2010. "Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook1." *American Journal of Sociology* 116:583-642.

### Author Biographies

**Zack W. Almquist** is an assistant professor of sociology and statistics at the University of Minnesota. His main research focus is on understanding, modeling, and predicting the effects that space (geography) and time have on human interaction (e.g., communication, friendship, mentorship, and needle sharing) and social processes (e.g., information passing, knowledge acquisition, skill development, and disease transmission).

**Benjamin E. Bagozzi** is an assistant professor of political science and international relations at the University of Delaware. His current research interests lie in the study of political violence, environmental politics, and political methodology.