Journal of Communication, 2022, **00**, 1–14
https://doi.org/10.1093/joc/jqac030
Original Article

Emerging hybrid networks of verification, accountability, and institutional resilience: the U.S. Capitol Riot and the work of open-source investigation

Stephen D. Reese* and Bin Chen

School of Journalism & Media, University of Texas at Austin, Austin, TX, USA
*Corresponding author: Stephen D. Reese. Email: Steve.reese@utexas.edu

Abstract

The violent spectacle of the January 6, 2021, U.S. Capitol riot provides a case study of how online open-source investigation helped identify members of the mob and provide deeper understanding of the day's events. Considering this form of investigation as an emerging network for the hybrid institution of journalism, an assemblage extending beyond the newsroom, this study takes a mixed-method, networked ethnography approach to map out this open-source space. Using a network analysis of open-source investigators based on Twitter data shows that the recently emerged Bellingcat organization and similar open-source groups serve a liaison function, bridging professional journalists and largely anonymous citizen "sleuthers." Using this network as a guide, depth interviews were conducted with key participants, showing that the community was organized around an accountability ethos and a methodology of verification. We argue that this interlocking network has potential to strengthen the resilience of the journalistic institution, build citizen trust and resist politicized historical revisionism.

Keywords: assemblage, news institution, network analysis, hybrid media system, journalism, open-source intelligence

On January 6, 2021, a rally of Trump supporters turned violent when directed at the Capitol building during the certification of the presidential election. The mob, which included paramilitary members, broke through the out-numbered lines of police and threatened members of Congress, even the vice-president presiding over the session. Order was eventually restored, but tight security precautions remained along with many questions. In all, the event represents the culmination of a massive misinformation campaign that threatened to derail the American democratic process itself.[1] Journalists and law enforcement officials made an urgent effort to identify the participants, leading to criminal charges in an effort still underway—represented most visibly by the extensive work of the Congressional Select Committee appointed to investigate the attack. A more complete understanding of that day's events was aided by parallel streams of information, including from on the ground, professional news gathering teams, but also well documented by the participants themselves, via Facebook, Twitter, Parler, and other social media—leaving a massive trove of incriminating information available online for journalists to process, with the help of an ad hoc alliance of news professionals, experts, data transparency activists, and citizens. This was a new kind of citizen journalism based on open-source intelligence (OSINT), an established and growing field with roots in government intelligence and the defense community, also called digital forensics, or, more broadly, *online open-source investigation*. Unlike the rioters who went online to celebrate breaching the Capitol, citizen investigators (often called "sleuthers") sought to hold them accountable.

As a high-profile and complex event, the riot provides a case study of how a diverse mix of online open-source researchers—citizen, professional, and other experts—came together to carry out a coordinated form of investigation. In this study, we want to explore the implications of that hybrid structure for the institution of journalism and its practices, as these elements come together to promote a form of innovation with vital political implications. More broadly, how is such an *assemblage* brought together around core institutional values of the press, including the goal of providing accountability? In a mixed methods approach, we seek to better understand the participants, professional and otherwise, involved around a common goal. Open-source investigators are naturally active on social media, and based on these interactions we conduct a network analysis to help identify the most influential contributors, who are then further described and interviewed: about their routines, collaborations, and values. The Capitol insurrection was a shock to the system, starkly revealing the threat from violent extremists against political institutions, including the press. In the face of these attacks, we must identify those practices that have potential to contribute to institutional resilience, greater transparency of newsgathering, and public trust in the process. Was this a fleeting coming together of social actors and interests or does it indicate some more lasting features of a new hybrid institution? To the extent it does, an emphasis on accountability and ground-level empirical evidence in newly tangible ways may help recuperate the core values of journalism and help alleviate the disconnect with the citizens it serves.

Theoretical framework and perspective

The major news organizations provided extensive coverage of the Capitol riot, assisted by the journalistic ecosystem that now incorporates a diverse mix of players, including traditional commercial, non-profit, and digital-native. We go

Received: 4 December 2021. **Revised:** 8 August 2022. **Accepted:** 9 August 2022

beyond this focus on professional news coverage to consider a more diverse set of inter-dependent elements cutting across social fields, which we characterize as an assemblage. Indeed, this means taking a broader view of the institutional press, beyond the static analysis of specific organizations or particular group of professionals. This is more compatible with the definition proposed by Reese (2021) of the institution, as a "complex social structure—formed by an interlocking network of rules and activities, roles, technologies, norms, and collective frames of meaning–which work together to sustain its coherence, endurance, and value." That means considering how the boundaries of the press-public division have opened up in new ways, with the work of open-source investigation—of particular interest here—taking place within a new kind of institutional alignment. In other words, traditional journalism is not just *normalizing* or *domesticating* new developments and innovations into its institutionalized practices, but a process of mutual adaptation helps produce new communities altogether that transcend professional news practices. Norms are developing around this work, but to what extent are emerging open-source practices coordinated around a shared ethos that serves institutional values?

Online open-source investigation

Although the analysis of online digital evidence has not found a consistent label or well-defined boundaries, the term *open-source intelligence* (OSINT) is relatively well accepted in the government world. OSINT has grown beyond its intelligence roots to include traditional organizations and decentralized alliances of activists, coming together to document and understand events around, for example, human rights, war crimes, and environmental disputes. When criminal cases are involved, open-source blends into the interpretive work of "forensic journalism." The New York Times, for example, says its visual investigation team "combines traditional reporting with digital sleuthing and the *forensic* analysis of visual evidence to find truth." As perhaps the most prominent open-source team, the Netherlands-based Bellingcat came to fame with its investigation of the downing of a Malaysian Airlines flight in 2014. Preferring the label "online open-source investigation," founder Eliot Higgins says his work has been developing a new and globally distributed field, connecting journalism, advocacy, and crime research, "…where advanced technology, forensic research, journalism, investigations, transparency and accountability come together."[2] He calls his team "the most innovative practitioners of open-source intelligence and online journalism in the world" (Higgins, 2021).

This work goes beyond what once was labeled *citizen journalism*, a practice often seen as an alternative to professional authority and corporate media hegemony. That dichotomous distinction between professional and citizen, however, does not capture the new, more expansive and coordinated formations of interest here. This kind of investigation does not operate independently from, but *combines with*, professionals to form more of a loose partnership in the service of shared values, and by helping democratize journalistic work open-source methodology brings a greater collaborative ethos. We will refer to citizen online investigators here simply as "sleuthers," a common term even if it implies a less serious, amateurish image than the professionalism often found in this community.

From boundaries to assemblage

In mapping this space, we put less emphasis on the traditional professional boundary concerns with who's in and who's out. We instead take a more inclusive perspective on how it all adds up to a larger structure, and how connections are formed around institutional values. Carlson and Lewis (2015) observe that "Whereas boundaries stress limits, networks accentuate connections," suggesting greater emphasis on "how groups form allegiances that work across boundaries" (p. 7). That suggests looking not at boundary limitations but institutional expansion, and how adjacent fields have helped import their values into the profession. Ananny and Crawford (2014), for example, studied the work of mobile news app designers, who although not regarding themselves as journalists did value transparency, also a feature of hacker culture tapped for help by news organizations (Lewis & Usher, 2013). Chadwick and Collister (2014) showed how a news organization incorporated new technological tools, expanding its boundaries in the service of institutional renewal as it reported on the Edward Snowden 2013 leak of National Security Agency material. These approaches have helped expand research beyond organizational boundaries but still largely take a newsroom-centered vantage point.

We want to consider the broader and more multi-layered perspective of *assemblage*—an alignment of individuals, technologies, and norms—and how it comes together to yield a new formation, emphasizing process over fixed structures and aggregates, networks of connection rather than predefined forms. This directs attention in our case to how open-source work after January 6 took place around, and was held together by, a dynamic structure of professional work, key civil society actors and experts, citizen sleuthers, and the technological tools that made it all possible. A surge of journalism and media research, just in the last decade, has begun invoking this inter-disciplinary assemblage concept, to capture connections that cut across traditional organizational containers. Hermida (2020), for example, adopts this more expansive sense in describing the elements beyond formal institutions shaping the gatekeeping process. Assemblage thinking follows closely the spirit of Chadwick's (2013) hybrid media system and Latour's (2005) Actor Network Theory (ANT), by emphasizing the heterogeneous, loosely coupled, dynamic, interactive, even ephemeral quality—a mix, not of pre-existing formal associations, but of social actors and technological platforms that come into being and become consequential when "activated." Three major applications of this concept can be observed.

One has been toward greater cross-media inclusion, "assembling" a mix of legacy, hyper-partisan, and social media accounts, with Harder et al. (2016) interpreting Chadwick's *information cycle* as a complex "assemblage" that requires accounting for interactions across different media platforms. Other studies have emphasized a second feature: social heterogeneity. McGregor and Mourão (2016), for example, adopt an "assemblage-oriented" perspective in their study of socially networked political conversations based on fluid and dynamic, loosely linked individuals and groups. Broadening this perspective to include media, Langer and Gruber's (2021) study of UK scandal coverage combines new actors and platforms with legacy media and actors—elites, advocacy organizations, digitally enabled crowds, in dynamic interaction. Russell and Waisbord (2017) similarly examine

links among legacy news, platforms, professionals, citizens, and special interest groups.

A third distinctive, more Latourian, application combines human and material elements. Protests in Hong Kong become a form of "news assemblage," when including both online and physical elements, hybrid forms of organization combining with journalists (Wong & Wright, 2020). More often, these material elements are technological. Papacharissi (2015), for example, refers to assemblages as opening the way to include technological affordances into a socio-cultural context, while Uusitalo and Valaskivi (2020) treat the news organization itself as an "assemblage": a container for human capabilities, equipment, and affordances. Rather than treating them as just tools, Carlson (2018) follows Ananny (2016) in locating algorithms (embodiments of journalistic judgment) within their broader "sociotechnical assemblages" that to networks of actors and practices add norms.

From assemblage to institution

This brings us back to Reese's definition earlier, and whether the concepts of assemblage and institution are compatible. They are, to the extent that an institution is regarded not as just a loosely coupled configuration of *firms* that contribute to news *production*, but rather a new networked space. Mohr and White (2008), for example, define institutions as "linkage mechanisms," composed of "inter-locking relational networks," while Ananny (2018, p. 60) similarly defines the press infrastructure as a system of "loosely coupled arrays of standardized elements" coming together to produce an institution. Perhaps the configuration of interest here has yet to be firmly established and well-routinized, which seems at odds with the stability that institutions imply. That is no doubt why, as Anderson and Kreiss (2013) have observed, the tools of assemblage-oriented actor-network theory are largely absent from studies of institutional politics, but those tools are particularly relevant to moments like this, when new alliances are developing all the time. Mapping the elements of a new investigatory assemblage helps reveal the possibilities for more stable formations.

In addition, we suggest that an assemblage can have institutional qualities to the extent it orients around certain values and norms. The Times visual investigations unit's Malachy Browne, for example, describes this work as accountability-driven, and Bellingcat has taken the lead in advocating for transparency, openness, and collaboration, which when applied to the Capitol riot investigation has been called a "digital ecosystem of accountability."[3] The website "Faces of the riot," for example, featured material from the right-wing-friendly social media site Parler, which was scraped by anonymous hackers and made available to sleuthers. One of the hackers (donk_enby) said, "I hope that it can be used to hold people accountable…"[4] And, shared ethical norms are also developing around this work. The University of Toronto Citizen Lab's John Scott-Railton argued that digital forensics, while a powerful method, carries risks, "Which is why it was so important in this case to not make the name (of a perpetrator) public immediately but to refer it to the F.B.I."[5] A protocol for this kind of identification from more established players encourages citizens to not crowdsource on their own.[6]

Capitol open-source investigation

This space of open-source investigation involved three main groups: professional journalists, largely anonymous citizen sleuthers, and research centers pursuing open-source projects—all participants in what some of them refer to as the "community." First, professional news organizations participated depending on their commitment to this style of reporting, which can be time and labor intensive, but noteworthy efforts have begun receiving professional recognition. The Times visual investigation team helped piece together the sequence of the day's events, showing, for example, how a woman was fatally trampled,[7] as did Scripps-owned Newsy in its partnership with Bellingcat.[8] The Washington Post, Huffington Post, and ProPublica also did major projects. Second, largely anonymous citizen "sleuthers" shared their work with journalists and coordinated among themselves, breaking news and leading to arrests. Finally, non-profit organizations played a role. Some open-source specialists were part of larger think tanks, while Bellingcat has been specifically dedicated to this kind of investigation. Other cooperation developed between smaller specialized organizations, experts, hackers, and providers of search and archiving technology. So, this is a highly heterogeneous example of what we will call an investigatory assemblage, including, but certainly not limited to, elements within the mainstream press.

The Capitol riot helps focus the scope of this project, but the challenge of understanding such a complex space remains. So, the research questions for this study concern how to map out what we more broadly consider a hybrid institutional structure. First, regarding the assemblage itself and its key elements, we want to know:

> RQ1: Who were the major participants in the open-source investigation of the Capitol riot?
> RQ2: What were the connections of interaction and working collaborations among these participants?

Next, we focus on the practices linking these elements, what they mean to the participants, and the larger ethos that supports them. In a larger sense, we want to know what kind of institutional qualities this structure has of relevance to journalism. That is:

> RQ 3: What common practices across journalistic, open-source organizations, and citizens promote coordination of work?
> RQ 4: What kinds of shared values, norms, and motivations characterize and are supported by this open-source space?

Methods

In a mixed-methods design, we use network analysis to address the first two research questions dealing with the structure of the open-source space, and in-depth, semi-structured interviews to address the other two questions, regarding motivations, values, and practices of the participants. Taken together these methods yield a form of *networked ethnography*, which uses network analysis to identify strategic social actors (e.g., Howard, 2002; Robinson & Anderson, 2020), who can be examined more closely with qualitative methods.

Network analysis

Twitter data provide a number of potential links with which to construct a meaningful network, including direct retweets

(RT) or the @mention of other users, as well as hashtags (e.g., Murthy, 2018). Both journalists and open-source investigators are active on Twitter as a means of coordinating and disseminating their work, but beyond its function as a means of communication we use the links that can be constructed using the platform's affordances as a mapping strategy to reveal the relative prominence of participants and their relationships. A purposive sample of tweets was gathered during the two-week period between January 6th and 20th, starting with 36 users identified by carefully tracking the popular press and social media over that time, using Tweetdeck to monitor the most active accounts and what they were reporting. This soon revealed those doing the most active open-source investigation work about the Capitol riot, confirmed by the most prominent hashtags (Table 2), which included the earliest sleuthing team (#seditionhunters), the most prominent early perpetrator targets (#scallops and #bullhornlady), and most targeted group (#oathkeepers). As a form of snowball sampling, these 36 users were more closely examined to determine who they retweeted relevant to the investigation, with an RT considered more serious than a more casual "liking" engagement, and the stronger measure of endorsement and certification. This step yielded an additional 37 accounts, for a total core group of 73 (henceforth "users"). We first constructed an RT network (Figure 1), based on retweets by that core group, yielding an *N* of 14,378 with 4,398 unique users.[9] We also identified the hashtags included in that same group of retweets, yielding a total of 823 hashtags, with the top 50 shown in Table 2.

Next, an *actor-topic network* was constructed that uses these hashtags to add an additional dimension to the retweets. Taking the individual RT-tweet "package" as the unit of analysis (including the *retweeting* user from the core group and *i* users), we examined the co-occurrences within that unit involving four kinds of nodes: the *retweeting* user; the *retweeted* author-user; other user, if any, @mentioned in the RT; and #hashtags, if any, included in the RT. Our approach follows Hellsten and Leydesdorff (2020), who advocate combining these data of actors and what they call "topics" into a *socio-semantic network*, a mapping that reflects Latour (1996) in treating both people and forms of message content as "actants." Although *topics* can be derived from various measures of content, their study and ours are restricted to using #hashtags, as a succinct form of semantic emphasis. People orient around these message elements (e.g., #seditionhasconsequences), which in this case are more often quite practical means of carrying out specific tasks of identifying perpetrators (e.g., #bullhornlady). We are less interested here in the precise semantic qualities of a message, which would require closer inspection, than in knowing who is involved with whom (and what)—for which hashtags provide useful clues.

Tracing consequential elements reflects both the logic of network analysis and ANT, "...how a given element becomes strategic through the number of connections it commands, and how it loses its importance when losing its connections" (Latour, 1996, p. 372). We essentially are adding to their suggested 3-mode network (tweet author-user, addressee/mention, hashtag) by incorporating the RT relationship.[10] This creates a non-directional, co-occurrence network among hashtags and two kinds of users (retweeting, or originating, users and users either mentioned or retweeted). Thus, users are connected to the extent that they engage not only with each other but around similar topics, as expressed here by the hashtags. In a relatively small network of the relevant open-source actors and a restricted two-week time frame, this strategy not only reflects our heterogenous assemblage mentality but optimizes the connections available for an interpretable network. (Figure 3 in the online supplementary file shows an example of a coded tweet.)

Interviews

Turning to the interview stage of this project, the loose confederation of open-source investigators makes it impossible to create a more formal sample frame, so the strategy of networked ethnography was used to identify the most prominent users (Twitter accounts) and target them for semi-structured interviews to explore the work and norms of the open-source community. We also made an effort to ensure representation from the three groups of investigators prominent (as the results will show) in the network: professional journalists, members of open-source-oriented organizations, and (usually) anonymous sleuthers. The greatest challenge was posed by this third category's anonymity, an understandable protective measure given that some of their targets involve paramilitary organizations (e.g., Oathkeepers). Many of the accounts were no longer active and others non-responsive, but with enough persistence via Twitter direct messaging (DM) and email to explain the project and request their cooperation one replied and was able to then vouch for the first author with others in the community, leading to interviews with three major group leaders.

Identifiable users with organizational affiliations were, of course, easier to contact, producing two interviews with journalists at major national news organizations and two with smaller online outlets. Four interviews were conducted with members of a major open-source organization and two with open-source researchers at non-profit centers. Given their prominence in the network and our emphasis on the early phase of the investigation, we considered this (total of 13) a sufficient number for further analysis. (Two other sleuthers replied through DMs, Signal, and even email with more limited responses.). Interviews lasting on average one hour were conducted in September and early October of 2021 (and one major original sleuther in late January, 2022), by the first author via Zoom, with the transcription feature providing a largely usable record after some further editing. The questions focused on how they became involved in open-source work, particularly around the Capitol riot, the nature of the working relationships among participants in the community (especially between professional journalists and citizens), including with law enforcement, how they perceived their own role in the investigation, and (especially for the non-journalists) their motivations. Transcripts were reviewed multiple times using a constant comparative method (Glaser & Straus, 1967) to identify codes, concepts, and finally themes. Key details were also extracted that helped better understand the chronology of the investigation, and these were added to what was already known from a thorough reading of news accounts of January 6th and the contributions of open-source practices.

Results

Our mixed-method strategy starts with the network analysis to find meaningful patterns that can be probed more deeply using targeted interviews. This cross-referencing provides a

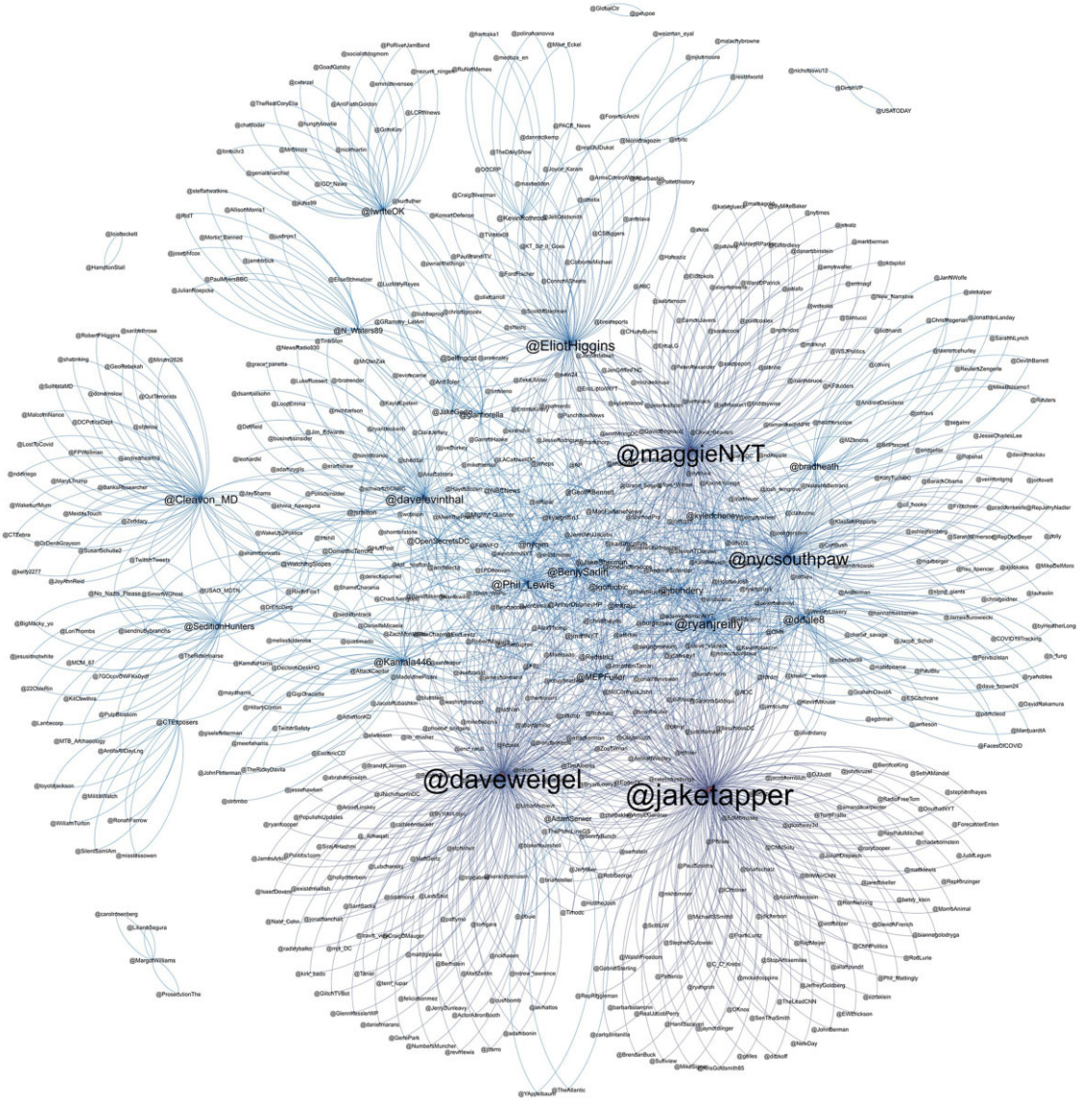

Figure 1. Retweet Network and Sample of Open-source Accounts.

Notes. The RT network is visualized with undirected ties. Ties with RT times smaller than 2 are filtered out. Node size and label size are scaled based on a node's degree centrality. See high-resolution online supplementary file and at https://osf.io/69zbp/

fuller and mutually reinforcing understanding of the dynamics behind the patterns found in social media data.

Network analysis

In our network analysis, importance is reflected by the extent to which an account both retweeted posts from others and was retweeted by them, and a ranking of these top accounts (or *nodes*) is shown in Table 1.[11] Not unexpectedly, professional journalists, especially at major news organizations, are at the core of this network (e.g., Maggie Haberman, NYTimes; Dave Weigel, Washington Post; and Jake Tapper, CNN), including contributors to news sites (e.g., @nycsouthpaw, a New York-based attorney). They tweet as a routine part of their work (e.g., Tapper has 260k tweets), but so do open-source organizations, represented most prominently here by Bellingcat and its leader, Elliot Higgins, along with Citizen Lab's Scott-Railton. Only a few anonymous investigators, such as @seditionhunters, make the top 20.

The same data visualized in Figure 1 show the most influential members, with the strength of ties between users proportional to the number of RTs between them, we required that users retweet another more than twice in order to be considered linked (which reduces and simplifies the network while still retaining most of the data), yielding a total network of 615 nodes. It shows that beyond the prominent positions of the major news organizations, Higgins occupies a key location near the center of the network, along with other Bellingcat contributors. As another way to describe this community, we ranked the top hashtags used in the retweets by the core open-source group (Table 2). Some are associated with certain accounts (e.g., #seditionhunters), while others help expedite the search for specific rioters (e.g., #bullhornlady, #oathkeepers). And the prominence of #fbi and #fbiwfo (Washington field office) showed the inclusion of law enforcement as an important consumer of open-source work product.

Table 1. Top-20 Accounts with Highest In-degree, Out-degree, and Total Degree in the RT Network

Nodes	In-degree	Nodes	Out-degree	Nodes	Total degree
@maggieNYT[a]	0.03	@jaketapper[a]	0.24	@jaketapper[a]	0.26
@igorbobic[a]	0.03	@maggieNYT[a]	0.20	@maggieNYT[a]	0.23
@ryanjreilly[a]	0.02	@daveweigel[a]	0.20	@daveweigel[a]	0.22
@MEPFuller[a]	0.02	@nycsouthpaw	0.10	@ryanjreilly[a]	0.10
@Redistrict[a]	0.02	@ryanjreilly[a]	0.08	@nycsouthpaw	0.10
@oneunderscore[a]	0.02	@EliotHiggins[b]	0.07	@EliotHiggins[b]	0.08
@daveweigel[a]	0.02	@davelevinthal[a]	0.06	@davelevinthal[a]	0.07
@chrislhayes[a]	0.02	@ddale8[a]	0.05	@BenjySarlin[a]	0.05
@jsrailton[b]	0.02	@Cleavon_MD	0.04	@ddale8[a]	0.05
@jaketapper[a]	0.01	@Kamala446	0.04	@Cleavon_MD	0.05
@BenjySarlin[a]	0.01	@BenjySarlin[a]	0.04	@bradheath[a]	0.05
@bellingcat[b]	0.01	@bradheath[a]	0.04	@Kamala446	0.05
@JakeSherman[a]	0.01	@Phil_Lewis[a]	0.04	@Phil_Lewis_[a]	0.04
@kaitlancollins[a]	0.01	@SeditionHunters	0.03	@IwriteOK[a]	0.04
@AlexThomp[a]	0.01	@IwriteOK[a]	0.03	@SeditionHunters	0.04
@MacFarlaneNews[a]	0.01	@kyledcheney[a]	0.02	@kyledcheney[a]	0.03
@kyledcheney[a]	0.01	@gianfiorella[b]	0.02	@igorbobic[a]	0.03
@jmartNYT[a]	0.01	@jbendery[a]	0.02	@bellingcat[b]	0.03
@AricToler[b]	0.01	@bellingcat[b]	0.01	@jbendery[a]	0.02
@jdawsey1[a]	0.01	@OpenSecretsDC[b]	0.01	@gianfiorella[b]	0.02

Note. Degrees presented in the table are weighted. In-degree is measured by the frequency of an actor being retweeted by others. A higher score of in-degree indicates greater influence as the information source. Out-degree is measured by frequency of an actor retweeting others. A higher score of out-degree indicates greater influence as the information disseminator. Total degree is sum of the in- and out-degree.

[a] Professional journalist.

[b] Bellingcat or open-source organization.

Table 2. Top 50 Hashtags most used in Retweets by the Core Open-source Group of 73

Rank	Hashtag	Rank	Hashtag
1	#seditionhunters	26	#slickback
2	#capitolriots	27	#seditionhunter
3	#capitolriot	28	#seattleuwguy
4	#helmetboy	29	#extinguisherman
5	#seditiontrack	30	#billsmafia
6	#scallops	31	#walkertexasfascist
7	#bullhornlady	32	#dcriots
8	#seditionvids	33	#gasen
9	#fbi	34	#cnnsotu
10	#pinkhatlady	35	#hathag
11	#dcterrorists	36	#threepercentsheriff
12	#seditionhasconsequences	37	#whitejeansguy
13	#oathkeepers	38	#canadagooseman
14	#covid19	39	#fleecenazi
15	#capitol	40	#fingerman
16	#individual4	41	#thepinman
17	#catsweat	42	#earpiece
18	#stripesguy	43	#spazzo
19	#caveman	44	#thecommander
20	#baldeagle	45	#dc
21	#breaking	46	#bidenclimate
22	#capitolbuilding	47	#eyegouger
23	#capitolbreach	48	#emonazi
24	#4	49	#ironworkerguy
25	#fbiwfo	50	#identifytheterrorists

Notes. 823 hashtags identified in total. The hashtags are ranked based on their occurrence frequency in the document-feature matrix of RT data.

Next, we combined these data from the top users and hashtags to create a deeper co-occurrence network (described in the "Methods" section), not differentiating among these nodes but treating all as Latourian actants. We selected the top 50 #hashtags included in the original RT network, along with the top 50 @mentions or retweeted users, and added this data to the original core group of 73 accounts, mapping the co-occurrence network among these actants. We calculated the modularity (see Blondel et al., 2008) in Gephi to detect clusters by setting the parameter resolution = 1.5 (see Lambiotte et al., 2008), rather than the default 1, yielding 3 major clusters.[12] "We were searching for broad patterns rather than subtle nuances, which would be more difficult to detect given the relatively small size of our dataset. So, we found the community structure was more clearly interpretable when the resolution parameter was set to 1.5 yielding three communities, with a modularity score at 0.340 (which varied little with lower resolutions that produced more communities)."

The size of the nodes is based on frequency of co-occurrences, suggesting their relative importance, and cluster-groups are calculated such that the strength of links among members (edges) are greater than what would be expected by chance.[13] Within these groups, nodes with a higher degree-centrality are located closer to the core of the cluster's space, reflecting how well-connected they are to the other group members, both directly and indirectly. We opted for degree-centrality as the most common, intuitive and straightforward centrality metric of social and semantic networks, reflecting Latour's reference cited earlier to "connections," and more easily interpretable for our socio-semantic network. Table 3 shows the rankings from the 3-community results, displaying the top-20 most important nodes by community cluster, using degree-centrality, a common measure of a node's influence in the network. Figure 2 shows the resulting structure, as a well-integrated community that will guide our further investigation into the strategies and motivations behind the interactions revealed on Twitter.[14]

In this representation, the major news organizations still occupy influential positions, but adding the hashtag nodes flattens the original retweet network that was dominated by the elite journalistic hierarchy, with citizen sleuthers and open-source organizations now showing greater prominence.

Table 3. Top-20 Accounts in the 3-mode Network with Highest Degree-Centrality by three Communities

Cluster = 0 (middle)	Degree Centrality	Cluster = 1 (left)	Degree Centrality	Cluster = 2 (right)	Degree Centrality
@jsrailton[a]	1.00	@seditionhunters	0.88	@nycsouthpaw[b]	0.55
@bellingcat[a]	0.96	@no_nazis	0.84	@jaketapper[b]	0.49
@fbiwfo	0.75	@cleavon_md	0.84	@ryanjreilly[b]	0.47
@fbi	0.73	#seditionhunters	0.79	@igorbobic[b]	0.38
@kamala446	0.62	@riverfox1	0.63	@kyledcheney[b]	0.38
@eliothiggins[a]	0.49	#capitolriots	0.59	@ddale8[b]	0.38
#capitol	0.47	#capitolriot	0.56	@daveweigel[b]	0.37
@ronanfarrow[b]	0.40	@domesticterror2	0.53	@maggienyt[b]	0.34
@evanchill[a]	0.40	@watchingslopes	0.52	@mepfuller[b]	0.32
@ctexposers	0.36	@beebthesimp	0.51	@phil_lewis_[b]	0.30
#fbi	0.33	@capitolhunters	0.45	@jakesherman[b]	0.30
@gianfiorella	0.29	#scallops	0.42	@bradheath[b]	0.27
#fbiwfo	0.27	#bullhornlady	0.38	@nycjim[b]	0.27
@corycullington	0.27	#stripesguy	0.37	@davelevinthal[b]	0.26
@arictoler[a]	0.27	@sendnudybranchs	0.37	@jbendery[b]	0.26
@n_waters89	0.27	#baldeagle	0.36	@kylegriffin1	0.25
#dcriots	0.26	#slickback	0.34	@benjysarlin[b]	0.25
@k2thesky	0.26	#extinguisherman	0.34	#breaking	0.21
@jakegodin	0.25	@seditiontrack	0.34	@annalecta[a]	0.21
@iwriteok	0.23	@simonwghost	0.32	@opensecretsdc[a]	0.21

Notes. Degree centrality of a node measures how many connections it has. A higher score indicates more connections. The degree centrality in the table is scaled.
[a] Bellingcat or open-source organization.
[b] Professional journalist.

Journalists cluster with other journalists, and anonymous sleuthers (e.g., @seditionhunters) cluster among themselves in two distinct groups. The journalist cluster (represented in red) consists most prominently of professionals from legacy news organizations, but also from HuffPost (@ryanjreilly), Daily Beast (@mepfuller), and Politico (@kyledcheney), with the Center for Responsive Politics (@opensecretsdc.org) also included but more marginal. A bridging cluster (represented in blue) includes Bellingcat and, less prominently, its Eliot Higgins, but also Scott-Railton and two sleuthers: Capitol Terrorist Exposers (@ctexposers) and @kamala446, who also play a bridging role. Also of interest here are the strong linkages in this bridging group to other institutional actors, including law enforcement (@fbi and @fbiwfo). Interestingly, the prominence of @cleavon_md, an Arizona emergency room doctor-activist is the rare example of a non-journalist, non-open-source organization, non-anonymous sleuther, but it does suggest a certain egalitarian quality to this space, where anyone can potentially rise to prominence if deemed credible by the community.

Interview results

Adding insights from prominent members of the network provides a better understanding of what lies behind that structure. In the analysis that follows, we identify key themes that explain the connections, motivations, and values binding this diverse group of participants together. Regarding the field itself, we highlight first its fluid boundaries and then the cross-sector connections that allow coordination and collaboration. We then identify four key dimensions of open-source institutional logic: (a) archive-first mentality, (b) accountability ethos, (c) open-source epistemology, and (d) gamification.

Fluid boundaries

Open-source practices are, of course, shared across the three general communities reflected in the network analysis and reviewed above: journalistic, research organization, and citizen. But the deep interconnections reflect the fluid boundaries of open-source work, with circulation of personnel among fields, including hiring from one into the other (e.g., a Times visual unit journalist was hired from the Berkeley Human Rights Center, as was a Bellingcat investigator). As a matter of professional practice, open-source is merging with traditional journalism, with one informing and complementing the other. Groups like Bellingcat incorporate journalistic elements in its reports, but even if they do not its technical investigations will attract attention from journalists who themselves will add that dimension, as they enter into mutually symbiotic relationships to advance investigations.

Certainly, open-source has increasingly been incorporated into newsrooms, as evidenced by the many January 6th reporting projects, and, according to one reporter at a large organization, become accepted as "real journalism."

> J1: You know, we're not necessarily trying to compete with reporters on the ground, because, frankly, it would be a mistake to say we can do everything with open source… but I think it would also be a mistake to say that you get the full picture from being on the ground. Open-source doesn't happen in a vacuum: we do interviews, we call people that were there, we call law enforcement.

Newsroom-based investigators vary a good deal in their position within the larger organization. One, with a small news site, contrasted his "more nimble and flexible" operation with the "bureaucratic juggernaut" of the Times.

A number of other non-profit and university research centers, many of them DC-based, have a long involvement with OSINT-oriented investigation (as one researcher put it, "Bringing clarity to contested information environments"), and several of their affiliates devoted time to the investigation. Besides Bellingcat, the most active center-based participant was Citizen Lab's John Scott-Railton, who was reported to be instrumental in providing a "template" for the work of what

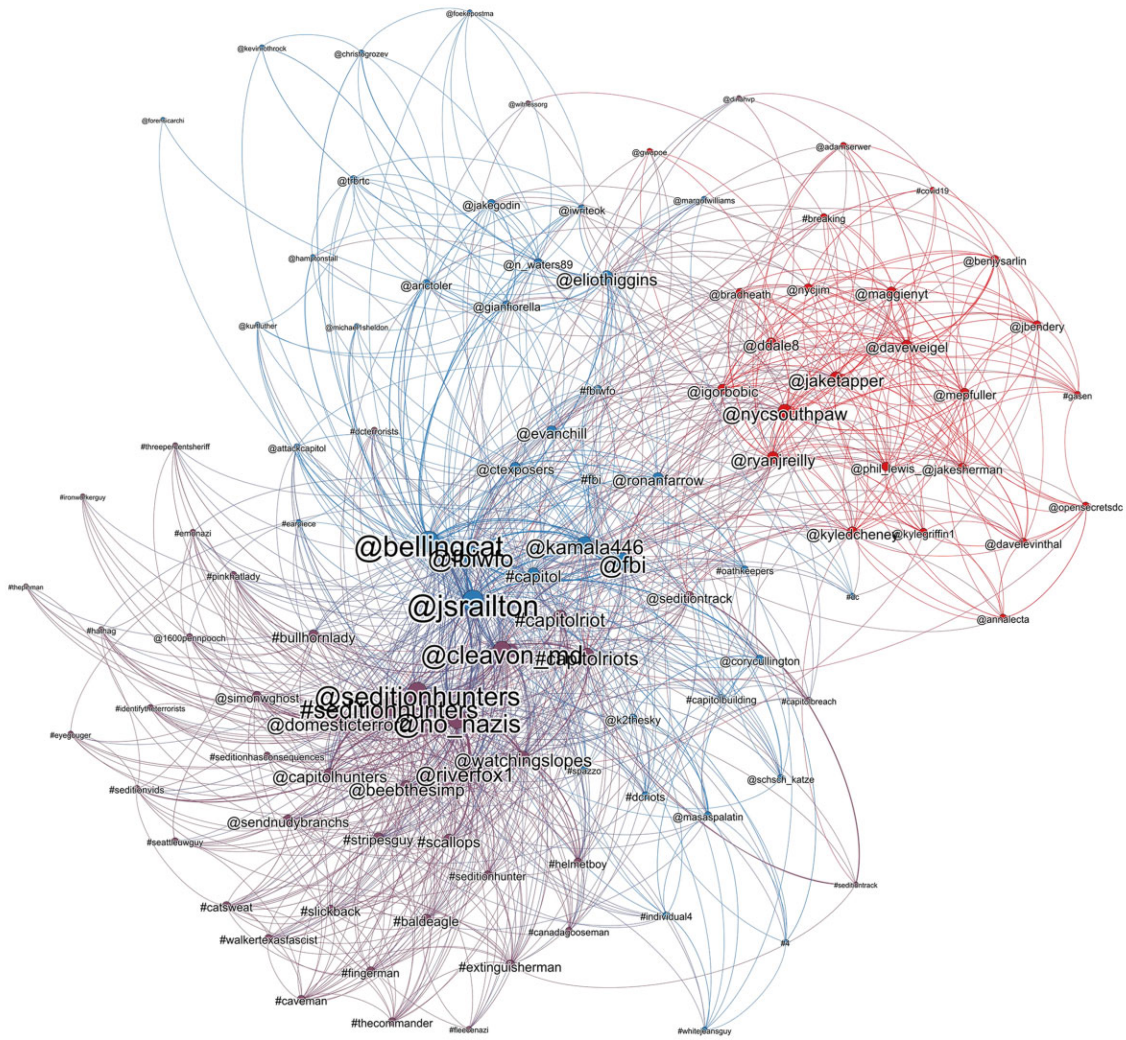

Figure 2. Users–Mention–Hashtags Network of Open-source Investigation on Twitter (3 clusters)

Notes. Co-occurrence network of 73 open-source investigation accounts, top 50 mentions and top 50 hashtags in their retweets from January 6 to January 20, 2021. Modularity scores are computed in *Gephi* (Blondel et al., 2008) and resolution is set to 1.5 (see Lambiotte et al., 2008). See high-resolution online supplementary file and at https://osf.io/69zbp/

one journalist described as "like wrangling cats…a sprawling bunch of different anonymous accounts and identities." A sleuther leader agreed:

> S1: Scott-Railton had the genius to pull it all together, because he realized the potential of crowdsourcing and wanted to be sure that it was done with good practices…So, I consider him one of the keys to keeping it together, and then out of that this whole ecosystem came together.

Bellingcat's relatively flat organizational structure enables its team to move quickly on ideas, while anonymous "sedition hunter" groups obviously have a much more "democratic" approach to publishing information, with minimal editorial oversight. A number of these anonymous accounts were prominent in the network, exemplified by @seditionhunters, @ctexposers (Capitol Terrorist Exposers), and Deep State Dogs, aka @1600PennPooch, which is coordinated by @forrest_rogers, one of the few of them who is public. Anonymity, however, need not mean lacking seriousness of purpose and professionalism, as these sleuthers provided important coordination and significant research skills. One anonymous Twitter account, for example, is said by one journalist to belong to one of the "best geo-locators on the planet." Another developed a map of the Capitol riot, searchable by keywords tagged to videos, which caught the attention of Bellingcat.

Coordination and collaboration

Online investigations can harness the power of crowdsourcing, but this contribution needs to be systematically managed, a function baked into the work of Bellingcat, which has brought this innovation into the journalistic field. As one of its researchers put it:

> OS1: One of the great things about open-source research is that the entry level to this field is pretty open. Like, you

> don't have to have a press badge from CNN to do the work, right? And so the more hands you get on it, again from people who are (doing it) for the right reasons, the better.

In contrast, one major newspaper solicited citizen input, which ended up not being useful when poorly planned and flooded with trolls.

> J2: It was interesting, and we certainly in the beginning got a lot of tips. There was a lot of negative, really negative feedback and a lot of vulgarity from people who asked us why we weren't trying to do the same thing for Portland (protests for racial justice), for example.

Indeed, Bellingcat can appeal to citizens in a way that traditional newsrooms cannot, using Google forms to standardize input data, soliciting around specific questions (e.g., tweets and videos pertaining to police violence against journalists at a BLM protest), and relying on a smaller group of volunteers to further process the material. This explains why Bellingcat occupies an important niche space in this information ecosystem—shown as a liaison in the network structure—acting as a verifier and interpreter of information for journalists, while professionalizing those techniques and promoting them to news consumers—helping them better verify what they're seeing online. (One sleuther specifically mentioned learning from Bellingcat.) As one center-based investigator observed: "You can say, no, actually this video is not from the Capitol riot. We've determined that it's actually from this other event, and here's how we can prove it." The entry of more amateur investigators into this space makes some kind of coordination function even more important, as one think-tank researcher warned:

> OS2: Opportunists can devalue the field, if you start getting more and more people who will just post something really quick, maybe even just be the first person to reach a conclusion. And then you'll be wrong, and people would be, like, all right, well, why do we trust these open-source researchers if they're just these goobers on the Internet?

This open-source community is highly collaborative in working toward a common goal. As a practical matter, the "seditionhunters" hashtag allowed sleuthers to find each other early on, with initial Zoom meetings giving way to more secure (and private) coordination via platforms like Discord. From there, ideas for investigation were (and are) pitched back and forth, with some leading to joint projects, some formally with news organizations using specific contractual agreements. These close collaborative relationships bind this community together but are not easy to disentangle. Investigators may pool their insights informally through Slack channels and other contacts, and only later realize they have to sort out credit for publication or contractual paperwork required by their newsroom. Groups co-publish and collaborate with major news organizations as a way of drawing greater attention and credibility to their work. In the case of Bellingcat, these relationships have been deepened by the training workshops it conducts: promoting the methodology, building trust, and creating a well-established network for collaboration. As one researcher put it, "I see those (news) organizations and I think, yeah, like if there's a way that we can help with the digital and you guys do that sort of the legwork, that's often a fertile ground for collaboration." This relationship works both ways, between sleuthers and journalists, who also share images to help with identification. Non-profit think tanks have always had a close relationship with journalists, sharing expertise and in turn gaining visibility for their work, and this continues with similar overlap in the open-source space. As one put it, "It's definitely a symbiotic relationship."

Open-source work connects journalists with sleuthers, but not without some difficulty. Given the diversity of the participants and still emerging open-source practices, the values of objectivity were not completely shared. A sleuther team leader who *was* comfortable with the press understood that many of his group were not.

> S1: I would love for the press to be able to call @seditionhunters or all these other people and have conversations with them, but they (sleuthers) won't do it. Don't forget, they don't trust the press. I mean, why should they? They're gamers, they know how to do this, and they did. They're not individuals who have worked with journalists, who would even know how to contact the journalist.

And yet, once these collaborations get going, they can, he suggested, engender greater trust.

> S1: It's good to see that the citizens are beginning to trust journalists and that the journalists are earning that trust. It's a great experience for a citizen just to see that their words and their information are being accurately reported.

Another sleuther expressed a similar view, saying, "I am usually pretty skeptical about major media, you know. Sure, I was a pretty critical reader but right now I'm the fiercest defender–I love them." But even with the similarities in work routines, citizens do not always have the same view of when traditional journalistic standards of evidence are met, as noted by one reporter:

> J2: The tricky thing about working with people who aren't trained in journalism is that you've got to watch out for biases and stuff, people who are like, "Oh, I know it was them, it has to have been them, they did this, or, like, they organized this whole event, absolutely." And it's like, well, actually the evidence doesn't (show) the Oath Keepers and the (Proud) Boys were there.

The partnering between sleuthers and journalists varies greatly, with some professionals treating them like any other anonymous source, while others enter into more co-equal relationships, a practice which some of the more organized sleuthers find deeply rewarding. As one put it:

> S2: In the beginning everybody wanted something from me, but that gave me the luxury to track my own people since I'm the contact person, but it goes so far that I got a relationship with a Washington Post journalist. And then we sleuth together. It's amazing, and like with New York (Times) visual investigations, they were 24-7, they were part of the group. There is no difference, like literally a guy

with a Pulitzer Prize, or we amateurs, there's no difference. So, and I really gotta say, there is no competitiveness.

From the newsroom side, some reporters try to facilitate this trust, assuring sleuthers that they themselves come out of the same background and "get it."

J1: I do think there's a little bit of tension, and these anonymous accounts doing a lot of work getting picked up by big media outlets. And how does accreditation work, do you pay people? All of that. And it is unfortunate, sometimes people who are in the open-source community, they put work out there and it's, for lack of a better word, stolen. We're trying to find a way to work in harmony, because we want to recognize, you know, @CTExposers found this Oath Keeper...

In spite of some suspicion of the press, sleuthers recognize the value of having a high-profile journalist bring attention to the case from law enforcement and adding that extra level of authentication (The New Yorker's Ronan Farrow being one of the more notable). Another journalist was described by a sleuther as "embedding" with them (online), helping them problem-solve and gaining their trust:

S3: So it's multi-layers of trust that I have for him, to trust his discretion. I trust his ability to independently evaluate evidence and make a go or no-go call. Or "we need more evidence before we can go" call.

Archive-first, analyze later

The logic of open-source emphasizes a certain brute-force empiricism of induction from large bodies of evidence, especially in the early stages of contentious events—made more critical by the likelihood that material will be lost. That helps explain the realistic need for collaboration, crowd-sourcing, and reliance on a larger open-source community. The method prioritizes raw evidence and seeks to preserve what one called the "primary materials of the internet," to counter those revisionists who would try to change the record. After January 6th this drove a quick search for as much relevant, riot-related materials as possible, even before any analysis, using Twitter to solicit material, collecting tweets in threads, and putting suspects in publicly available spreadsheets. According to one journalist who has worked closely with an open-source organization:

J2: There were live streams coming out of what was unfolding, and we're really close (to the Capitol), so that's when that kind of kicked into open-source collection mode. And so they started trying to scrape links... I think this is how a lot of open-source works, especially in breaking news situations. You try to collect as much as you can.

This "journalistic" response to breaking news was shared by an open-source investigator:

OS1: ...there comes a time in every kind of breaking news situation where like a switch goes off in our heads. Where you go from, I'm watching this as a consumer, like as an individual I'm thinking, oh my God, I can't believe this is happening, and then you go, okay, wait now, I'm a really good investigator...there's work to be done, right? Let's just save as much of the stuff as possible.

This "archive-first" mentality was also emphasized by other sources from open-source organizations.

OS3: We've had a lot of experience with these kind of events and it's almost always true that a lot of the content will end up being removed at some point or another, and so we made an attempt to preserve as much as possible....So that was very much the kind of initial reaction, and then later on we started to analyze, putting together the videos into a comprehensible timeline ...and so, yeah, initially collect and preserve and then start the analysis, sorting and verification.

The accountability ethos

The diverse array of investigators all had their reasons for getting involved, but some important motivations emerged, especially for the volunteer sleuthers. The spectacular nature of the Capitol riot must be taken into account, with many around the world outraged by what they saw. As one sleuther group leader put it: "I felt the responsibility, and I was just propelled by rage at what I saw....all these violent actions done with impunity, without any fear of repercussions just continues to baffle me."

For others, many directly affected by conspiracy-believing family members, participating in this work gave them a sense of efficacy in responding to a politics out of control. As one investigator put it: "A lot of people felt kind of powerless, right? Because we're watching these crazy videos on CNN... I think that people were trying to find a way to do something, even if it was something as simple as scraping a video." A large-newsroom journalist agreed with this assessment, although she didn't grant sleuthers much greater altruistic credit than that: "For me, it would seem to me that a lot of them it's just anti-Trump emotion, they were out to get anyone who was partaking in that circus on January 6."

Others in the think-tank world came to the Capitol investigation as a chance to offer their specialized skills and topical expertise, as did some sleuthers. As one journalist observed:

J1: A lot of (OSINT) people are just very technically focused, like some only look at guns, and they say, I know exactly what kind of gun that is...It's almost like a hobby, but also he's just spent so much time that he's built such an expertise...he's somebody you can go to.

This resembled the career trajectory of other researchers, one who was driven even before coming to his organization by a particular subject matter interest:

OS1: I was doing it almost stubbornly, like nobody else was paying attention to me, and I didn't care because I really cared about it...If you ask 20 of us, you'll get 20 different answers, but there's like a stubborn passion for an issue that drives you to do this kind of work.

The idea of accountability, however, is a recurring theme across the board. That meant not only holding perpetrators accountable but also helping compensate for the limited attention span of the media. For one sleuther leader, that

accountability, as suggested in the network analysis, also aligned open-source work with law enforcement as the ultimate accountability mechanism:

> S1: I felt that we would lose momentum if we did not start providing them (the press) information. They were not doing it themselves… you don't have many New York Times visual investigative teams, so my fear was that if we didn't keep it in people's faces that they would move on to something else, and I also knew that it would keep the police at the FBI in check…

Indeed, the law enforcement officials, who are main consumers of open-source evidence, have gained greater appreciation for the power of trained citizen work to help make arrests. The FBI actively sought tips from citizens, as officials monitored the work of sleuthers along with everyone else, using the same nicknames attributed to rioters by researchers. In fact, mentioning sleuther work in publicly posted court documents gave the work greater legitimacy and new momentum, as identifications were resolved and fed back into common data-bases. One justice-beat reporter called the periodic release of evidence by the courts and Department of Justice a "goldmine," where investigators could say, "Oh, there's that person, and they can upgrade the chart, they can add that to the list. You may find somebody, you know, throwing something at officers or something that then gets added in…"

Another reporter recognized that the relationship can still be fraught, especially knowing now that some officials themselves participated in the riot: "A lot of people in the community are still very skeptical of law enforcement, for a wide variety of reasons, and because a lot of times the things we're investigating are actions by those exact governments and law enforcement agencies." Again, this relationship worked better when information-sharing was targeted and coordinated among the sleuthers. Flooding the FBI with tips that strained its capacity made it difficult for more well-vetted findings to be prioritized, as recalled by one sleuther.

> S2: In the beginning, there are a lot of overzealous people, and they say this guy looks suspicious. And I said, I cannot do anything with it, and then they say, yes, but we share this to the Feds regardless. Yes, that's why *we* can't get through!

Whatever the motivation, the practices have coalesced into something more "institutionalized" around a spirit of accountability, as expressed by one of the original sleuthers:

> S3: It's like you can't unring a bell. And this open-source type investigation to protect democracy when it's attacked, that bell has rung. We now know how to do that, there's hundreds of us that have systematically come together and develop an architecture framework to do this.

"You need receipts!": open-source epistemology

Most importantly, this networked assemblage is bound together by the logic of open-source methodology. We mentioned the brute-force, archive-first practices, but this logic represents a larger attitude toward knowledge production. The stance toward empirical evidence was succinctly summarized in one sleuther remark: "If you believe in the Big Lie, you need receipts!" This allows investigators to coordinate and harmonize their work using a common methodology, whether they have a professional background in journalism or a more activist agenda. Open-source researchers may migrate from activism to journalism, but the methodology remains the same: whether at Berkeley, Amnesty International, or Human Rights Watch; whether doing reports for the UN or writing a news story. And this allows movement in personnel and work product across those boundaries. The methodology also provides the basis for trust in anonymous sleuthers and for collaboration, as two different journalists agreed:

> J1: Again, the beauty of going back to the methodology is that, despite all of these reasons that people may have for doing the work, personal backgrounds, the way they approach it and what they hope to get out of it, at the end of the day…..it's really easy for us to reproduce their methodology because of the open-source nature of the work. Anything that @seditionhunters is using we also have access to, and that ability to recreate the process and say, "Oh, we can get the same conclusion." Then you start to kind of build up these networks of trust, even though you don't know the name of the person.
>
> J3: Even though I don't know their identity, I can just check their work, but they're very reliable sources of information, and they have this extensive track record, where it's like, yeah, they're good. They've never steered me wrong.

Open-source journalists see the value of this "show your work" methodology in their own projects. As an investigator at a large newsroom suggested, the reliance on, for example, visual forensics techniques helped explain the appeal of its investigations.

> J1: What we're telling you, you can check yourself, right? You have access to all the videos we do, and it creates a certain level of trust that you might not get when, you know, a reporter in DC reports something, and they said that they got it from an anonymous source in the Pentagon.

Even though Bellingcat may be associated with more activist-oriented causes, the methodology is prioritized above all, and a source there claimed it was effective in rendering their work useful, regardless of end consumer.

> OS3: We're also able to mitigate our own biases and present information in a way that, especially for contentious subjects, if not acceptable, is certainly verifiable by both sides. Whereas I think if you go to the more extreme ends of the political spectrum, people who try and use these kinds of techniques frequently let those kind of biases and their own preconceptions influence their analysis.

"It's like a puzzle": gamification of open-source

Finally, in addition to the more institutionally serious dimensions of accountability and verification, the logic of gaming must not be overlooked. Coming up across multiple interviews and deeply embedded in the work, it means solving a puzzle that many find rewarding, appealing to both

journalists and sleuthers alike. Many of the citizen investigators, for example, shared information and coordinated their efforts via Discord, a "chat app" used originally by online gamers, and Higgins himself recalls that before his Bellingcat days he was an obsessive online gamer who organized an international network. Gamification logic fits online investigation, which requires coordinating the insights of a dispersed global team of "players," who are tech-savvy and have the patience to be on screens for hours on end, monitor the internet and scour digital traces for clues. One sleuther acknowledged these rewards of the investigation "game."

> S2: Of course, it is like a puzzle, right? So, when I do Sudoku or I get like a terrorist arrested. . .I am like, oh my God, I found this guy, right? It's rewarding in a way, but it's definitely not what drives me, but I think there's a lot of overlap. (When the #bullhornlady was identified) that was like the big wow because we just got the white whale, and that got everybody even more excited, because after three weeks bullhorn lady had been ID'd. That just kind of kept it going as a game.

One self-taught, open-source news professional mentioned something similar. When military airstrike footage would be released, "I would just kind of for fun try to see if I can match that up with wherever it was on Google Earth, and it kind of gives you a little dopamine rush. Oh yeah, actually these buildings match up."

Discussion

In this study, we have tried to better understand the work of open-source investigation on the Capitol riot: what the components of this assemblage were and how they were connected, in an emerging and fluid new field that cuts across journalism, activism, and civil society organizations. The stunning images of a mob overwhelming police lines and breaching the symbol of democracy led to a massive amount of reporting using open-source techniques, work that extended far beyond the boundaries of professional news organizations. Using Twitter data to help discern this networked assemblage, we identified a prominent group of journalists and citizen sleuthers in distinct clusters, bridged by another group represented by open-source organizations, most notably Bellingcat, which guided citizen work and imposed frameworks for the identification process to proceed efficiently and more ethically. Rather than pose an alternative social space of "connective action" (e.g., Bennett & Segerberg, 2013), we show that open-source work allows this assemblage model to build on the traditional norms and practices of investigative journalism. Sleuthers may have their own ideological motivation, but without their skills and commitment to the larger accountability project they would not have found a place in this structure.

We regard the mixed-method combination of network analysis and depth interviews, built around a case study, to be a major contribution in making sense of these far-flung structures. In combining (mainly) the social and semantic, our use of an *actor-topic* network approach reflects the Latourian emphasis on *actants*, of consequential elements, and the idea of an institution as a "linkage mechanism." Guided by Reese's (2021) definition of an institution, we looked for whether the "interlocking network" found in this assemblage shows a set of stable practices and shared "collective frames of meaning" which "work together to sustain its coherence, endurance, and value." We think it does. We argue that it suggests a hybrid institution, combining different logics and de-emphasizing any fixed boundaries of journalism. With its dedication to social justice issues, the logic of activism may not be completely compatible with traditional objective reporting, where practitioners may doubt its neutrality, but it merges with and reinforces journalism to the extent it relies on the same methodology of verification.

Responses showed a shared accountability-driven ethos, and a unifying epistemological dedication to standards of empirical evidence, with the potential to enliven the practice of journalism and enlist a diverse group of participants in the work of transparency-based reporting. Citizen sleuthers are not sufficient to carry out this kind of investigation on their own but still need journalists to help broker the massive amounts of digital information. At the same time, news organizations could not accomplish what they did without the crowdsourcing and contributions of citizen investigators. But the "wisdom of the crowd" does not nearly capture how the work of this distributed yet integrated division of labor is accomplished, where highly skilled (and those willing to quickly learn) open-source professionals coordinate with amateur sleuthers. Certainly, we should also note that the economics of journalism play a role in promoting this assemblage, with newsroom budget cuts making unpaid contributors more welcome. That is particularly true in this case, where investigations can be easily supplemented with results from outside the newsroom—not the submission of free-lance stories, but building on a pyramiding of individual discoveries made possible by mutual collaboration.

We acknowledge some study limitations. Social media traces are only proxy for actual engagement and interest, but in this case, we thought them sufficient to help guide further qualitative interviews. Twitter, of course, is not universally adopted, but it's where journalists hang out and open-source investigators go to advance their work, which in turn is pushed out via those same social and other media. The sample of users and hashtags was not intended to capture the entire network of reporting during this period but to strategically highlight one core assemblage for additional probing. Gaining cooperation from anonymous (and even other) sources took several cycles of outreach that limits our sample,[15] but we were able to interview strategically important representatives from the main sectors, and coupled with the network analysis they give a consistent picture of the key themes.

We did not intend to capture the entirety of the media ecosystem documenting the Capitol riot, which included a number of intrepid professional broadcast and print reporters on the scene. The work of open-source is carried out at a certain, often safer, remove, with the results of the investigatory assemblage fed back into the media system for dissemination. These kinds of media forms are mixing in interesting hybrids: print media incorporate video into investigations, while open-source work relies in part on the images collected by reporters on the ground to produce a more detailed and comprehensive record of the event than could be produced by any one person online.

Finally, we do not have space to explore the ideological dimension that runs through this case. Given that the riot was inspired by such a polarizing figure as Trump, any opposition to him and his followers can take on a liberal cast. This was a

question put to our sources: could the tools of open-source investigation, often associated with human rights and other social activist causes, be just as appealing to right-wing groups? We suspect the open-source epistemology works against other political stripes, which although they certainly engage in targeted "doxing" of opponents are more attuned to authoritarian assertion and tribal reinforcement, not spreadsheets and data. To use Oliver and Wood's (2018) distinction, open-source is a *rationalist* enterprise, not an *intuitionist* one more associated with the political right. The politically agnostic tools of digital technology were used by some to help assemble an anti-democratic assault and by others to hold them accountable. We can still, however, examine what kind of values those structures support.

So, in spite of the shock to the system symbolized by the Capitol riot, we end on an explicitly normative and positive note. Although not a panacea, open-source investigation has great potential to build trust between journalism and citizens, particularly as organizations like Bellingcat have developed that serve as a liaison between them. The methodology of verification involves an array of people in the work of investigation and gives them a stake in the outcome. Even now, the reality of January 6th has been challenged and collective memory degraded, making evidence-based reporting more important than ever to combat politicized revisionism. This was an analysis of one event, and a spectacular one at that, but we hope the values that it highlights, both political and journalistic, will add to our understanding of how this kind of innovation works, as professional newswork gives way to more far-flung forms of coordinated investigation—yet still guided by a common institutional ethos.

Supplementary material

Supplementary material is available at *Journal of Communication* online.

Notes

1. The characterization of January 6 has been contested, variously called a "riot," "storm," "siege," "attack," or "insurrection," and participants labelled a "mob," "rioters," or "protesters." We use the term "riot" here for the event and "mob" or "rioters" for the participants, the terms used most consistently by media across the political spectrum ("protesters" is clearly a right-wing preference).
2. https://www.bellingcat.com/about/
3. https://www.newyorker.com/news/news-desk/an-air-force-combat-veteran-breached-the-senate
4. https://www.cnet.com/news/website-features-faces-from-parlers-capitol-riot-videos/
5. "An Air Force Combat Veteran Breached the Senate," by Ronan Farrow, newyorker.com, January 9, 2021
6. https://www.technologyreview.com/2021/01/14/1015931/how-to-be-an-ethical-online-investigator-activist/
7. https://www.nytimes.com/2021/01/15/us/rosanne-boyland-capitol-riot-death.html
8. https://www.newsy.com/stories/newsy-bellingcat-how-a-violent-mob-breached-the-capitol/
9. We note using this "snowball" approach carries a sampling bias in that the second group is not independent of the first. But we were mainly interested in the vastly larger network of users yielded from that combined core group, not comparisons within it.
10. As Hellsten and Leydesdorff (p. 5) note, one can go beyond the ontologically flat ANT in the sense that the users can "author" a hashtag, but it cannot "author" the users. Nevertheless, this approach shows what we are seeking: how users are connected with respect to their involvement with the same hashtag/topic.
11. Top 20 accounts with the highest in-degree/out-degree/total degree are based on the retweeting network, calculated with igraph R package (https://igraph.org/r/).
12. Gephi uses the Louvain (or Multilevel) method, found to be one of the best performing of eight algorithms in a test of detecting communities by taking both accuracy and computing time into account (Yang et al., 2016). This unsupervised method requires researchers to select the number of communities by considering different resolutions. In Gephi, a lower resolution produces more communities, and higher resolution means fewer, which affects only the number of communities, not the community structure (i.e., small communities will join and form a bigger community with increasing resolution). As a robustness check, we also used the widely used Walktrap method in igraph R package to do community detection for the actor-topic network, yielding results very close to the those observed from the Louvain method.
13. The greater the presence of these groups, the greater the "modularity," a function available in the Gephi network visualization software.
14. Note that because the first retweet network formed visually clear clusters, intuitively obvious, around prominent journalists, we did not perform a community detection check on it, given our main focus on the actor-topic network.

References

Ananny, M. (2018). *Networked press freedom: Creating infrastructures for the public right to hear*. MIT Press.

Ananny, M., & Crawford, K. (2014). A liminal press. *Digital Journalism*, *3*(2), 192–208. https://doi.org/10.1080/21670811.2014.922322

Ananny, M. (2016). Toward an ethics of algorithms: Convening. *Observation, Probability, and Timeliness*, *41*(1), 93–117. https://doi.org/10.1177/0162243915606523

Anderson, C. W., & Kreiss, D. (2013). Black boxes as capacities for and constraints on action: Electoral politics, journalism, and devices of representation. *Qualitative Sociology*, *36*(4), 365–382. https://doi.org/10.1007/s11133-013-9258-4

Bennett, W. L., & Segerberg, A. (2013). *The logic of connective action: Digital media and the personalization of contentious politics*. Cambridge University Press.

Blondel, V., Guillaume, J., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, P10008, *2008*(10).

Carlson, M. (2018). Automating judgment? Algorithmic judgment, news knowledge, and journalistic professionalism. *New Media & Society*, *20*(5), 1755–1772. https://doi.org/10.1177/1461444817706684

Carlson, M., & Lewis, S. (2015). *Boundaries of journalism: Professionalism, practices and participation*. Routledge.

Chadwick, A. (2013). *The hybrid media system: Politics and power*. Oxford University Press.

Chadwick, A., & Collister, S. (2014). Boundary-drawing power and the renewal of professional news organizations: The case of the guardian and the Edward Snowden National Security agency leak. *International Journal of Communication*, *8*, 2420–2441.

Glaser, B., & Straus, A. (1967). *The discovery of grounded theory*. Aldine.

Harder, R. A., Paulussen, S., & Van Aelst, P. (2016). Making sense of twitter buzz: The cross-media construction of news stories in election time. *Digital Journalism*, *4*(7), 933–943. https://doi.org/10.1080/21670811.2016.1160790

Hellsten, I., & Leydesdorff, L. (2020). Automated analysis of actor–topic networks on twitter: New approaches to the analysis of

socio-semantic networks. *Journal of the Association for Information Science and Technology*, *71*(1), 3–15. https://doi.org/10.1002/asi.24207

Hermida, A. (2020). Post-publication gatekeeping: The interplay of publics, platforms, paraphernalia, and practices in the circulation of news. *Journalism & Mass Communication Quarterly*, *97*(2), 469–491. https://doi.org/10.1177/1077699020911882

Higgins, E. (2021). *We are Bellingcat: Global crime, online sleuths, and the bold future of news*. Bloomsbury.

Howard, P. N. (2002). Network ethnography and the hypermedia organization: New media, new organizations, new methods. *New Media & Society*, *4*(4), 550–574.

Lambiotte, R., Delvenne, J., & Barahona, M. (2008). Laplacian dynamics and multiscale modular structure in networks. *ArXiv*, *Preprint*, *0812.1770*.

Langer, A. I., & Gruber, J. B. (2021). Political agenda setting in the hybrid media system: Why legacy media still matter a great deal. *The International Journal of Press/Politics*, *26*(2), 313–340. https://doi.org/10.1177/1940161220925023

Latour, B. (1996). On actor-network theory: A few clarifications. *Soziale Welt*, *47*(4) 369–381.

Latour, B. (2005). *Reassembling the social: An introduction to actor-network theory*. Oxford University Press.

Lewis, S., & Usher, N. (2013). Open source and journalism: Toward new frameworks for imagining news innovation. *Media, Culture & Society*, *35*(5), 602–619. https://doi.org/10.1177/0163443713485494

McGregor, S. C., & Mourão, R. R. (2016). Talking politics on twitter: Gender, elections, and social networks. *Social Media and Society*, *2*(3) 1–14. https://doi.org/10.1177/2056305116664218

Mohr, J., & White, H. (2008). How to model an institution. *Theory and Society*, *37*(5), 485–512.

Murthy, D. (2018). *Twitter*. Polity.

Oliver, E., & Wood, T. (2018). *Enchanted America: How intuition and reason divide our politics*. University of Chicago Press.

Papacharissi, Z. (2015). Toward new journalism(s): Affective news, hybridity, and liminal spaces. *Journalism Studies*, *16*(1), 27–40. https://doi.org/10.1080/1461670X.2014.890328

Reese, S. D. (2021). *The crisis of the institutional press*. Polity.

Robinson, S., & Anderson, C. W. (2020). Network ethnography in journalism studies: A mixed-method approach to studying media ecologies. *Journalism Studies*, *21*(7), 984–1001. https://doi.org/10.1080/1461670X.2020.1720519

Russell, A., & Waisbord, S. (2017). The Snowden revelations and the networked fourth estate. *International Journal of Communication*, *11*, 858–878.

Uusitalo, N., & Valaskivi, K. (2020). The attention apparatus: Conditions and affordances of news reporting in hybrid media events of terrorist violence. *Journalism Practice*, *16(17)*, 1–19. https://doi.org/10.1080/17512786.2020.1854052

Wong, S. C., & Wright, S. (2020). Hybrid mediation opportunity structure? A case study of Hong Kong's Anti-National Education Movement. *New Media & Society*, *22*(10), 1741–1762. https://doi.org/10.1177/1461444819879509

Yang, Z., Algesheimer, R., & Tessone, C. (2016). A comparative analysis of community detection algorithms on artificial networks. *Scientific reports*, *6*, 30750. https://doi.org/doi.org/10.1038/srep30750