Article

Digital Sousveillance: A Network Analysis of the US Surveillant Assemblage

Colin Burke

University of California, San Diego, USA
cmburke@ucsd.edu

Abstract

This paper introduces a new methodological approach to the study of surveillance that I call *digital sousveillance*—the co-optation of digital data and the use of computational methods and techniques to resituate technologies of control and surveillance of individuals to instead observe the organizational observer. To illustrate the potential of this method, I employ quantitative network analytic methods to trace the changes in and development of the vast network of public and private organizations involved in surveillance operations in the United States—what I term the “US surveillant assemblage”—from the 1970s to the 2000s. The results of the network analyses suggest that the US surveillant assemblage is becoming increasingly privatized and that the line between “public” and “private” is becoming blurred as private organizations are, at an increasing rate, partnering with the US government to engage in mass surveillance.

Introduction

The field of surveillance studies has long been concerned with the relationship between public and private organizations and how they carry out surveillance (Fyfe and Bannister 1996; Lyon 2001; Wakefield 2002). While it is a given that private entities play a crucial role in allowing the government to engage in surveillance, whether through “backdoors” (Crampton, Roberts, and Poorthuis 2014) or “revolving doors” (Hayes 2012), we know far less about how this vast assemblage of public and private organizations actually operates. When private corporations are implicated in helping government entities spy on their citizens, as was the case with several companies named in the documents released by Edward Snowden, the immediate reaction of companies is always an attempt to distance themselves from any association with government surveillance activities. The fear of being associated with government spying plays into the secretive nature of the surveillance industry, as companies fear that consumers will react negatively to such an association. Because of the tight-lipped nature of the US government, as well as the private corporations involved in surveillance activities, the study of what some have termed the “surveillance-industrial complex” has struggled to “unmask” the actors involved. Questions about *who* is involved in this assemblage, *to what extent* they are involved in carrying out surveillance activities, and *how* this contemporary form of public-private surveillance has emerged are still relatively unanswered.

This article aims to answer these unresolved questions and introduce a new methodological approach to the study of surveillance that I call *digital sousveillance*. To illustrate the potential of this approach, I employ quantitative network analysis to trace changes in the vast network of public and private organizations, or what I refer to as the “US surveillant assemblage,” involved in surveillance operations in the United States from the 1970s to the 2000s. Drawing on data from the Transparency Toolkit’s ICWatch database,¹ I

¹ <https://icwatch.wikileaks.org/>

demonstrate the potential of digital sousveillance as a critical research method, bringing together digital data and robust computational techniques to gather, visualize, and analyze this assemblage of public and private organizations. The results of the network analyses indicate that the US surveillant assemblage is becoming increasingly privatized and the line between “public” and “private” is becoming blurred as private organizations are, at an increasing rate, partnering with the US government to engage in mass surveillance. I conclude by outlining the limitations of these analyses and the dangers posed by the contemporary structure of the US surveillant assemblage.

Blurring Public and Private: The Surveillance-Industrial Complex

Drawing on Eisenhower’s age-old concept of “military-industrial complex,” Ben Hayes (2012) developed the term “surveillance-industrial complex.” While the surveillance-industrial complex does not amount to a comprehensive theory of surveillance, by employing this concept we can make important theoretical assumptions about the corrosive nature of the state-corporate nexus on political culture, democratic governance, and social control. First, it intimates the “revolving door” between those public entities which are officially tasked with security and those private actors that provide the new methods of surveillance and control that will enable them to do so. Second, it highlights the political and economic model that underpins these social relations. Lastly, it puts forth a critical understanding of the implications of this public-private nexus: that the surveillance-industrial complex promises to “deliver ever more pervasive, intrusive and effective surveillance technologies in perpetuity” (Hayes 2012: 167–168). The surveillance-industrial complex consists of many distinct actors and is strategically positioned at the center of many of the transformations in population control, policing, and intelligence gathering (Hayes 2012), while at the same time remaining mostly out of public sight and only ever showing itself in a way that fails to reveal the particularity of these public-private relations (van der Vlist 2017). The task of mapping this complex nexus of public and private is thus a challenge due to its secretive nature.

The partnership of the US government with private corporations to conduct surveillance is not a recent phenomenon. As far back as World War I, telegraph and cable companies like Western Union turned over all telegraphic communications to the earliest predecessor of the NSA, known as the “Cipher Bureau” (Bamford 2009). The leveraging of public-private partnerships for purposes of intelligence gathering continued up until the mid-1970s. It was around this time that much of the American public shifted from general trust in public institutions to dramatic distrust. Revelations surrounding the FBI and CIA, the Watergate episode, and other Nixon administration intrusions provided concrete examples of government abuse of power that made the public, as well as private companies, wary of intelligence operations. In light of the attacks on September 11, 2001, the perceived need for surveillance and other intelligence-gathering operations intensified. While solicitation and attempts to gather information from private entities for intelligence purposes may not have been received well by private actors in the decades following the 1970s, the opposite was true in the immediate aftermath of the 9/11 attacks. The perceived need for intelligence for counter-terrorism purposes was great and the private sector was ready to provide it.

Surveillance-Industrial Complex as Surveillant Assemblage

While the concept of the surveillance-industrial complex is useful in providing a conceptual framing of the public-private nexus regarding surveillance, it alone is relatively limited in its theoretical capacity to capture the complexity of surveillance as a phenomenon. I utilize Haggerty and Ericson’s (2000) concept of “surveillant assemblage” to move beyond the dualism of “public” and “private” and to highlight the multiplicities and nuances of these partnerships. Below, I outline the concept of surveillant assemblage and its operationalization in this article.

The origins of assemblage as an analytical concept can be located in the writings of Deleuze and Guattari (1987). The term has since been used by several scholars (Abrahamsen and Williams 2009; Collier and Ong 2005; Marcus and Saka 2006; Sassen 2006) to denote an understanding of structures that is not confined to

a distinct scale (such as local/global or micro/macro) (Bueger 2014). An assemblage may seem structural as “an object with the materiality and stability of the classic metaphors of structure, but the intent in its aesthetic use is precisely to undermine such ideas of structure” (Marcus and Saka 2006: 102). This does not mean disavowing the notion of structure completely (indeed, network analysis relies upon some semblance of structure) but rather acknowledging its dynamic and fluid nature. What some have called “assemblage thinking” thus represents an attempt to refuse totalities and embrace social life as a nonlinear, heterogeneous alignment of emerging and continuously moving parts (Bleiker 2014). This also means paying attention to the complex relationships between these multiple actors and the broader forces that impel them to act in the way they do (Lisle 2014: 72). In the case of surveillance, there is no single, centralized agency that coordinates the totality of surveillance systems and operations. What perhaps makes this assemblage so unique, and indeed powerful, is its ability to integrate discrete surveillance systems and actors.

Haggerty and Ericson (2000) draw on Deleuze and Guattari (1987) to describe what they call the “surveillant assemblage.” In speaking of *the* surveillant assemblage, however, they are not referring to a stable, fixed entity. Because it is “multiple, unstable and lacks discernible boundaries or responsible governmental departments,” the surveillant assemblage cannot be dismantled by just eliminating a technology or mode of surveillance, nor can it be confronted by focusing criticism on a single bureaucracy or institution (Haggerty and Ericson 2000: 609). Much surveillance research tends to concentrate on the capabilities of discrete technologies or social practices and emphasize how they cumulatively pose a threat to civil liberties. This research is also too often overly concerned with local sociotechnical instances of surveillance, in observing the propagation of what Latour (2005) calls “oligoptica”—durable but extremely narrow views of the broader whole (Murakami Wood 2013). It is thus necessary to recognize that the surveillant assemblage takes a variety of forms and therefore cannot be captured through one case study. This is because surveillance, as a phenomenon, is driven by the need to bring systems together, to combine different social practices and technologies and integrate them into a larger whole.

As Buchanan (2015) warns, however useful and analytically revealing assemblage theory may be, in practice the use of the concept of assemblage is often indistinguishable from that of an adjective, serving more to *name* than *frame* a problem. It is thus crucial to explicitly outline the specific ways in which assemblage, as a concept, contributes to this study both analytically and methodologically. As previously indicated, approaching surveillance as a multiplicity is an invitation to go beyond binaries and dualisms. Classifications such as state/non-state, human/non-human, and public/private are not explanatory frameworks, but rather distinctions that require explanations and attention to how they are enacted in surveillance discourses (Bueger 2014). Going “beyond dualisms,” however, does *not* mean dismissing the categories of “public” and “private” altogether, as they still serve as useful characterizations of the type of organizations that engage in surveillance. Indeed, the distinction between public and private is particularly important for surveillance given the differing legal, as well as social, expectations placed on public and private organizations when it comes to issues of privacy. Instead, conceptualizing this network as an assemblage means greater emphasis and attention to how these classifications are actually enacted in surveillance discourses and how these two mutually exclusive categories of actors interact in ways that may problematize their respective classifications.

Digital Sousveillance as Method

Representing an assemblage in an academic narrative always entails a political choice in the sense that it decentralizes power and authority away from the state (Bueger 2014). Assemblage thinking thus lends itself to critical analysis of the formations and multiplicities under study and acknowledges the complexity of surveillance and socio-technical objects, viewing them as entangled and used simultaneously as modes of exploitation as well as resistance (van der Vlist 2017). It thus aligns with the practice of what Mann, Nolan, and Wellman (2003: 19) call “sousveillance,” which seeks to resist dominant modes and structures of power within surveillance by inverting the organizational gaze to “watch the watchers.”

While social media data, and digital data in general, have been utilized for surveillance by public and private organizations alike (Brayne 2017; Greenwald 2014; Lyon 2014; Zuboff 2019), there have been few attempts to co-opt these data to surveil the state and its private partners. To this end, I build on the concept of sousveillance to introduce *digital sousveillance*—that is, the co-optation of digital data and the use of computational methods and techniques (e.g., network analysis) to resituate technologies of control and surveillance of individuals to instead observe the observer—as a method for studying surveillance.

The use of digital data and computational methods is nothing new for the social sciences; however, they have been largely underutilized in the field of surveillance studies with few exceptions (Introna and Gibbons 2009; van der Vlist 2017). Although not necessary for the act of sousveillance, pervasive digital technologies and the data they generate can make sousveillance more effective (Mann and Ferenbok 2013). Many of the reasons that state and private organizations use digital data and computational methods for surveillance are the same reasons why the field of surveillance studies should also consider using them. The most obvious advantage of using digital data for studying surveillance is their sheer size and depth. Not only do digital data allow for a larger sample, but they also often contain in-depth, relational information. The use of open-source data has the additional advantage of greater transparency and reproducibility. Perhaps most important to those studying a hidden phenomenon like surveillance is that digital data may also allow for access to information that would otherwise not be available. This study, for example, uses open-source social media data to trace the linkages between public and private organizations, something that, given the secretive nature of the surveillance-industrial complex, would be difficult to observe otherwise. Digital sousveillance, of course, does not constitute a rejection of “small data,” or qualitative methods of studying surveillance. Indeed, digital sousveillance methods are perhaps best used as a complement to these methods, allowing for exciting new pathways for interdisciplinary and collaborative studies of surveillance. The methodological toolkit of digital sousveillance is extensive, allowing for the use of a wide range of computational methods including robust statistical analyses, computational content analysis, machine learning, and network analysis. The digital sousveillance toolkit also has the advantage of accessibility, as many of these tools are available and achievable through free, open-source software and programming languages.

It is also important to consider the ethics of using digital sousveillance as a method for studying surveillance. First, there is *flexibility* in how digital data are interpreted and produced. Algorithms are not technical, self-contained objects but instead objects that embody the socio-political values and biases of their authors (Eubanks 2018; Noble 2018). This means moving beyond digital data as representationalist and towards seeing them as performative; the composition and interpretation of these bits of data in effect produce the life and body of the subject into “data doubles” (Matzner 2016; Raley 2013). As Cheney-Lippold (2017: 11) notes, “we are ourselves, plus layers upon additional layers of algorithmic identities.” Digital sousveillance requires considerable attention to how these algorithmic identities are produced and vary across different contexts. Second, those engaging in digital sousveillance also need to exercise *accountability*. As outlined by boyd and Crawford (2012), accountability is broader than concepts of privacy in that it applies even when there is little to no expectation of privacy. Professional standards and ethics do not disappear when working with publicly accessible data. The size, depth, and sensitive nature of digital data necessitate additional attention to the implications of their use. The ICWatch dataset, for instance, also contains sensitive data, such as individuals’ names, photographs, and geographical locations.² Those engaging in digital sousveillance need to be aware of the potential harm that can come from publishing and making these types of data more visible and accessible.³ Accountability is essential to prevent scholars of

² The inclusion of individuals with no actual surveillance-related affiliations who were inadvertently scraped in the original IC Watch dataset is also particularly problematic. These data underwent an extensive process of data cleaning, both manually and using computational techniques. A “fuzzy matching” algorithm was used to eliminate redundancies in the data, such as variations in organizational names (e.g., US Army, Army, United States Army, etc.). In addition, these data were manually audited to remove profiles incorrectly included in the dataset.

³ The open publication or presentation of these types of sensitive individual-level data are not only unnecessary for the study of surveillance but also raise considerable ethical issues due to the unintended harm they may cause. It also serves to incorrectly assign “blame” to particular individuals while lessening attention to the role of broader

surveillance from falling into the same harmful, intrusive behaviors and abuses of power that they are attempting to dismantle.

This study thus seeks to demonstrate the potential of digital sousveillance as a critical research method, bringing together robust computational analytic techniques to gather, visualize, and analyze this complex and often hidden assemblage of actors. Below, I detail the data and methods used to accomplish this task.

Data and Methodology

This article uses network analytic techniques to map the historical development of the US surveillant assemblage, and its public-private linkages, from 1970 to the 2000s. The primary sources of network data for this article are drawn from the Transparency Toolkit's ICWatch database. Transparency Toolkit is an organization that uses open-source data to bring greater public visibility to surveillance and possible human rights abuses. The data employed in this article are scraped from individual profiles on LinkedIn. The scraper collected public profile information based on a list of search terms that consisted of the names of surveillance programs identified in the Snowden documents. If an individual's profile contained terms or names from one of these programs, the scraper collected the entire profile, including job history (job title, company, start/end date, and description for each job), skills, educational history, and location/area. Information about the organizations mentioned in the profiles, including company size, industry, and the type of organization (e.g., government agency, public company, etc.) was added to the dataset.

These data were chosen in large part because they provide a glimpse into a phenomenon that is often kept hidden from public view and academic research. Up until the release of the Snowden documents in 2013, public knowledge about the extent and nature of US government surveillance was mostly based on speculation and what little information had been willingly revealed by the US government. Thus, little academic work has been done on the surveillance industry let alone on the ways and extent to which they interact with nation-states and other government entities (Murakami Wood 2013). The inaccessibility of US government intelligence organizations, as well as their private partners, means the avenues for the study of this phenomenon are incredibly limited. These sources are thus essential objects of study because they represent one of the only ways of studying US government surveillance directly.

The temporal scope of this analysis spans from 1970 until 2009. The primary network analysis will compare four distinct periods: 1970–1979, 1980–1989, 1990–1999, and 2000–2009. This comparison of networks allows for a greater understanding of how the US surveillant assemblage has developed over this period. The 1970s is an appropriate starting point because this time represents a crucial moment in the development of the US surveillant assemblage. As previously discussed, the 1970s marked a shift from general trust in public institutions to dramatic distrust. The expansion of partnerships and networks of intelligence gathering operations in recent decades stands in stark contrast to the doubt and skepticism of the 1970s. Thus, tracking these networks up to the contemporary moment is crucial to understanding the modern formation of this assemblage.

In total, 25,479 individual profiles⁴ are identified in the data. This represents a considerable sample as the total intelligence community workforce has been estimated to be around 183,000 people, of which 58,000

organizational and institutional structures that maintain and perpetuate mass surveillance. This paper attempts to protect the privacy of these individuals by focusing the analyses on the organizational level.

⁴ At the time of original collection, the ICWatch dataset contained over 27,000 individual profiles. This number was reduced to around 25,000 after cleaning the data and removing profiles that were inadvertently sampled. Additional individual profiles from new data sources (i.e., Indeed) have been added to the ICWatch dataset since the time the data used here were collected, though not all of these new data have clear relevance to surveillance. For instance, the portion of the new data from the "FBI/DHS hack" contains the names, titles, phone numbers and email addresses of individuals working for the FBI and Department of Homeland Security. The sample here thus represents a subset of the ICWatch dataset as currently composed.

are privately contracted (Shorrock 2016). As noted above, each of these individuals' profiles contained information on the organizations and companies presumed to be involved in US surveillance programs. These organizations and companies represent the primary unit of analysis. The network data were originally two-mode data, with organizations *indirectly* tied together based upon their shared affiliation with particular surveillance programs. These types of networks are often referred to as "affiliation networks" (Wasserman and Faust 1994). Connections in this two-mode network represent the tie between the surveillance program(s) mentioned in an individual's profile and the organization(s) that the individual worked for. For example, if an individual worked on the PRISM program, all organizations listed in the individuals' profile would be connected to the PRISM program. However, since the aim of this article is to map the linkages among organizations, this two-mode network (surveillance program-to-organization links) was collapsed into a one-mode network (organization-to-organization links). A connection in this network thus represents a tie between two organizations based on their shared affiliation with particular surveillance programs identified in individuals' profiles. For instance, if Individual A worked on PRISM and worked for Booz Allen and Individual B worked on PRISM and worked for the NSA, a connection is drawn between Booz Allen and the NSA.

It is important to acknowledge that the connections between these organizations are drawn based on an *assumed* shared affiliation with surveillance programs. One limitation of these data is that they do not contain information about the nature or direction of the relationship between these organizations nor do they detail the specific surveillance-related activities of these organizations. There are, therefore, several possibilities when it comes to the nature of the linkages between these organizations, such as engagement in actual spying on behalf of the government, financial relationships, research and development, or the exchange of surveillance-related goods and services. The type of connection between organizations is presumably dependent upon the type of organizations involved. Each actor, public or private, operates in light of its own logics, agendas, and local constraints (Ball et al. 2015; Haggerty and Gazso 2005). For instance, a university tied to the NSA may be involved in research and development whereas a telecommunications company may provide the NSA with access to communications data and infrastructure. Framing these partnerships as an assemblage requires acknowledging the multiplicity of the connections between organizations, meaning these connections also result in the linkage of physical (fiber-optic cables, cell towers, servers) and digital infrastructures (data, algorithms, code, software), technical objects, knowledge, and discourses.

Cytoscape, a network analysis software, was used to generate four network graphs. The visual layout used to map these networks is called the "Group Attributes Layout," which groups the nodes in a circle based on an attribute they have in common. In this case, the nodes are grouped based on whether they are public or private organizations. Node size was scaled based on the node's degree. In other words, larger nodes represent organizations that have a larger number of connections to other organizations. The nodes are color-coded with blue representing "public" entities and red representing "private" entities.⁵ To illustrate the extent and structure of public-private connections, ties or edges between nodes are also color-coded. Purple edges represent heterophilic ties (ties between public and private organizations) and grey edges represent homophilic ties (ties among public organizations or ties among private organizations). An edge-bundling algorithm was used to bundle edges with similar destinations and connections to create greater space within the graphs and render these connections clearer. In addition to the graphs, statistical measures of network connectivity and homophily/heterophily were calculated using Cytoscape and a Python package, NetworkX, for each of the four temporal periods to allow for a quantitative means of comparison and tracking of the changes to the US surveillant assemblage over time.

⁵ For the purposes of this analysis, "public" entities consist only of US government organizations. Public universities or similar institutions that are funded by the US government are considered "private" organizations in this analysis. This was done to clearly distinguish agents working on behalf of the state (e.g., US Military, NSA, CIA, etc.) to engage in surveillance from those that, although they are directly or indirectly funded by the state, tend to operate outside of the realm of government.

Results and Analysis

As noted above, the data are analyzed in four distinct sections representing each decade from the 1970s through the 2000s. In the following section, network graphs (Figures 1–4) are shown to visually illustrate the networks and to provide a literal “map” of the changes to this assemblage over time. These visualizations alone are insufficient to draw robust empirical conclusions. To compensate for this, network statistics and measures are provided to allow for a clear, empirical means of comparison between each network. I thus rely upon statistical network analyses, including measures of network structure as well as measures of homophily/heterophily for each decade, to support and expand upon the insights gained from the graphical network representations.

Mapping the US Surveillant Assemblage

The 1970s represented a crucial period for the US surveillance-industrial complex. In this decade the American public’s general trust in public institutions shifted to dramatic distrust. Numerous concrete examples of government abuse of power, such as revelations surrounding the FBI and CIA, the Watergate episode, and other Nixon administration intrusions, resulted in both private companies and the public experiencing increased wariness of intelligence operations. This led to a series of reforms in the mid-to-late 1970s, such as FISA, that were aimed at preventing further abuses by the US government and severely limiting their ability to surveil American citizens. Because these shifts occurred late in the 1970s, it is unlikely that they had an immediate effect on the ties between public and private organizations within this network. Thus, it might be expected that the 1970s network is quite heterophilic with extensive connections between public and private organizations, as was the case leading up to these changes.

Figure 1 displays the network graph for the years 1970 through 1979. Immediately apparent is that the most significant nodes in the network in terms of degree are government entities. This indicates that these particular government organizations are the most well-connected in this network. While private actors are present, they are, overall, smaller in terms of degree meaning that these private organizations had fewer numbers of connections within the network. Nonetheless, the notable presence of public-private ties (purple) in the graph evidences that, as hypothesized, these relations were quite common and certainly did not cease to exist despite some of the changes later in the decade mentioned above. There is also far less homophily (grey edges) among public organizations when compared to private organizations. Also, there are several clusters of public-private ties outside of the most central group of nodes in the two groups suggesting that this network is quite decentralized.

The 1980s are of particular interest because they represent the period we might expect to illustrate the earliest effects of the changes in policy and public opinion during the mid-to-late 1970s on the US surveillant assemblage. The 1980s continued the trend of increased concern and skepticism that characterized the late 1970s. In 1981, President Ronald Reagan signed Executive Order (EO) 12333 which required each intelligence agency to establish procedures for the collection of electronic communications using “the least intrusive collection techniques feasible within the United States or directed against United States persons abroad” (quoted in Donohue 2016: 12). This decade also witnessed the passage of several privacy laws governing private information, including the Privacy Protection Act (1980) and the Electronic Communications Privacy Act (1986). Public concerns about privacy threats in this decade had also increased dramatically with 48% of survey respondents reporting that they were “very concerned about threats to their personal privacy” compared to only 31% in the late 1970s (Kumaraguru and Cranor 2005; Raynes-Goldie 2010). While the findings of the Church Committee led to significant reform, many of the same flaws began to reappear by the mid-1980s as the executive branch and intelligence community found new ways to avoid legal obstacles and carry out mass surveillance (Murphy 2014). In particular, the revelations around the Iran-Contra affair in 1986 involving the Reagan administration and the CIA cast doubt upon the ability of Congress and the public to manage and oversee the intelligence community. As the Congressional Iran-Contra Joint Committee concluded, “secrecy was used not as a shield against our adversaries, but as a weapon against our own democratic institutions” (quoted in Schwarz and Huq 2008: 57).

Figure 1: Network graph, 1970–1979.

Figure 2 depicts the network graph for the years 1980–1989. Like the 1970s, the 1980s network contains larger, highly connected government organizations. Also, network ties between public and private organizations are the majority by a substantial margin. There appear to be a higher proportion of relations between private organizations and other private organizations compared to the 1970s network. Also, in contrast to the 1970s network, there are several well-connected government organizations and private organizations that seem to be more prominent in terms of node size and degree. The 1980s network thus appears to be more centralized regarding the number of connections among nodes than the previous network.

Figure 2: Network graph, 1980–1989.

The network for the 1990s (Figure 3) appears to be qualitatively different from that of the 1970s and 1980s. Beyond just the increase in the sheer number of organizations and ties within the network, there is an apparent shift in the prominence of private organizations within the US surveillant assemblage. While private organizations were reasonably small in terms of degree and importance in the 1970s and 1980s compared to government organizations, private organizations in the 1990s are on par with the significant, highly connected government actors in the network.

With the events of the 1980s and the status quo approach of the 1990s, the stage was already set for a dramatic expansion in the 2000s of public and private partnerships to carry out mass surveillance on a global scale. The terrorist attacks on September 11, 2001 gave the US government the immediate legal and public support necessary to expand law enforcement and executive powers to grow the US surveillant assemblage exponentially. The passage of the USA PATRIOT Act in the immediate aftermath of the 9/11 attacks, as well as subsequent laws billed (albeit falsely) as “surveillance reforms” such as the Protect America Act (2007) and the FISA Amendments Act (2008), dismantled the restrictions put into place in the 1980s.

Figure 3: Network graph, 1990–1999.⁶

The US surveillant assemblage continued to carry out mass surveillance through covert surveillance programs. One such program was PRISM, which allowed the NSA to collect private communications from the world’s largest internet companies including Google, Facebook, Microsoft, and Apple. PRISM was particularly significant because it allowed the NSA to obtain virtually anything it wanted from the servers of internet-based companies that hundreds of millions of people around the world now use as their primary means to communicate (Greenwald 2014). Although the NSA documents claimed the PRISM program was run with the assistance of the private companies many denied knowledge of any such program (Greenwald and MacAskill 2013).

⁶ This graph represents a sample of the original network. To present a more readable graph, only the top 5% of nodes in terms of degree are included in this visualization. Network statistics and degree distributions are, however, calculated using all nodes and edges.

The most striking feature of the network graph is the overwhelming presence of private organizations as the dominant nodes in terms of degree within the network. In contrast to the three previous networks, private organizations have overtaken government organizations as the most well-connected actors in the US surveillant assemblage. There is also clustering of public-private ties, most noticeable on the private side of the graph with two major groups of public-private ties clustered together. There also seem to be, proportionately speaking, fewer public-public and private-private ties than there were in previous decades. This would seem to indicate that connections between public and private organizations were more frequent during this period. In the following section, I rely upon statistical network analytic methods to support these graphical illustrations and to draw more direct empirical conclusions about the structure of the US surveillant assemblage from the 1970s to the 2000s.

Figure 4: Network graph, 2000–2009.⁷

Structure of the US Surveillant Assemblage

Table 1 displays several network statistics for the four periods examined here. These measures indicate the level of network connectivity. Multiple measures are provided to illustrate changes to various structural elements of the network and as a robustness check. The networks were analyzed as undirected networks—as organizations within these networks had no discernible directional relationship or tie with each other. Within a network, all nodes that are connected form a connected component. The number of connected components indicates the overall connectivity of a network. A lower number of components, therefore, suggests stronger connectivity among organizations, whereas a higher number of components suggests weaker connectivity among organizations. By this measure, the 1970s–1990s had weaker connectivity while the 2000s had robust connectivity between the various actors. The second measure, the clustering coefficient, denotes the extent to which nodes in a graph tend to cluster together and form “triangles” (connections between three mutually connected nodes). The clustering coefficient is relatively stable from the 1970s until the 1990s but increases noticeably in the 2000s. This would seem to indicate that organizations in the 2000s were more likely to form a tightly knit and highly connected group than in past

⁷ This graph also represents a sample of the original network. To present a more readable graph, only the top 1% of nodes (in terms of degree) are included here. Network statistics and degree distributions are, however, calculated using all nodes and edges.

decades. The network diameter represents the largest distance between two nodes. A smaller diameter indicates greater connectivity and centralization within the network (and vice versa). There was no real difference between the four networks for this measure. The fourth statistic, network centralization, measures how the network is distributed in terms of degree. Centralized networks have a value closer to one, whereas decentralized networks have values closer to zero. The centralization of the networks increased over time, going from 0.783 in the 1970s to 0.911 in the 1980s to 0.919 in the 1990s and to 0.976 in the 2000s. This is in line with previous measures and would seem to indicate again that the US surveillant assemblage becomes more centralized and highly-connected in the 2000s. Lastly, the characteristic path length gives the average distance between two connected nodes and is also used to indicate the connectivity of the network. A shorter path distance suggests greater connectivity within the network. The characteristic path length was relatively constant across the four networks.

	1970–1979	1980–1989	1990–1999	2000–2009
Nodes	743	2443	6961	31,015
Edges	10,189	95,939	708,517	1,048,575
Connected Components	13	9	14	1
Clustering Coefficient	0.838	0.856	0.834	0.966
Network Diameter	5	4	4	5
Network Centralization	0.783	0.911	0.919	0.976
Char. Path Length	2.010	1.996	1.988	2.004

Table 1: Measures of network connectivity.

The evidence provided in Table 1, overall, suggests that the US surveillant assemblage became more centralized and highly connected over time. Some measures, like network diameter and characteristic path length, are relatively constant. Others, such as the number of connected components, clustering coefficient, and network centralization, suggest that there is a considerable shift in the structure of the US surveillant assemblage over time. While the structure of this assemblage is useful in furthering our understanding, these measures are unable to uncover the extent of public-private partnerships and how that aspect of the assemblage has changed over time. In the next subsection, I discuss the measures of network homophily and heterophily used to evaluate the extent and likelihood of public-private linkages within each network.

Homophily/Heterophily in the US Surveillant Assemblage

Given the higher degree of connectedness within the US surveillant assemblage in recent years, it might be reasonable to expect that public and private organizations would be more likely to connect in the 2000s networks. To test this hypothesis, I ran a series of quantitative analyses of network homophily and heterophily (Table 2). The first statistical measure is called attribute assortativity. This is used to measure the level of correlation between connected nodes for the values of an attribute (categorical or scalar) of the nodes (organizations). Attribute assortativity ranges in value from -1 to 1 , with -1 representing a dissortative network, where nodes tend to connect to nodes with dissimilar attribute values, and 1 representing an assortative network, where nodes tend to link to nodes with similar attribute values. The attribute used for this measure was whether the organizational node was public or private. In this context, it measures whether organizations in these networks were likely to link with organizations that matched their value for the public/private variable. By this measure, the 1970s and 2000s were dissortative, whereas the 1980s and 1990s were assortative. In other words, the 1970s and 2000s networks had a higher tendency of linkages between public and private organizations compared to the 1980s and 1990s.

The second statistical measure used is degree assortativity. Similar to attribute assortativity, degree assortativity measures the tendency within networks for organizations to associate with organizations of a similar number of connections (degree). Taking these results together, all four networks were dissortative, meaning organizations tended to link with organizations with a different number of connections (e.g., highly connected organizations were connected with lowly connected organizations). Interestingly, the degree assortativity value for the 2000s is much more dissortative than the previous networks. This indicates that there is a much higher tendency in the 2000s for organizations with a smaller number of connections to link with organizations with a more substantial number of connections.

In addition to measures of assortativity, Table 2 displays the percentage of ties between public organizations, between private organizations, and between public and private organizations. The measure of primary concern here, of course, is the percentage of ties between public and private organizations. By this measure, the 2000s network had a higher percentage of public-private relations compared to earlier years. This supports the attribute assortativity measure in that the dissortative 1970s and 2000s networks had a higher percentage of public-private relationships when compared to the assortative 1980s and 1990s networks. Similarly, the odds-ratios of public-private ties were about twice as high for the 1970s (0.72) and 2000s (0.85) networks than they were for the 1980s (0.38) and 1990s (0.36) networks. The likelihood of public-private ties is thus highest for the 2000s network.

	1970–1979	1980–1989	1990–1999	2000–2009
Attribute Assortativity	–0.001	0.081	0.090	–0.064
Degree Assortativity	–0.171	–0.144	–0.112	–0.590
Public-Public Ties (%)	22%	21%	22%	17%
Private-Private Ties (%)	47%	58%	57%	47%
Public-Private Ties (%)	46%	38%	38%	48%
Odds-Ratio: Public-Private Tie	0.72	0.38	0.36	0.85

Table 2: Measures of network homophily/heterophily.

Overall, the measures of homophily and heterophily displayed in Table 2 seem to indicate that there were greater tendencies and likelihoods of public-private connections in the 1970s and 2000s networks compared to the 1980s and 1990s. This supports the graphical evidence provided in the previous section that suggested the 2000s network is quantitatively different from prior decades. The 2000s network is highly centralized, well-connected, and heterophilic—that is, the 2000s network had a greater tendency for public-private partnerships than previous years as well as a greater tendency for connections between nodes of differing degree values. Interestingly, network statistics and measures for the 1980s and 1990s were nearly identical, meaning the explosion of public-private partnerships in the 2000s was not something that built up over time in a linear fashion as might have been expected. This would seem to suggest that there is something unique about the 2000s that allowed for this to occur, the most obvious possibility being 9/11 and the various legal and institutional changes that occurred in its aftermath.

Digital Sousveillance of the NSA

While the above network analyses evidence the ways that the topology and overall structure of the US surveillant assemblage have shifted in recent decades, it is also crucial for sousveillance purposes to bring to light some of the specific connections between public and private organizations. In this section, I briefly examine the public and private connections of the NSA in the 2000s. The NSA was one of the most prominent organizations in the dataset with over 26,000 connections in the 2000s network alone. This is perhaps unsurprising given that the NSA likely engages in more surveillance than any government organization in the world (Greenwald 2014). Given the significance of the NSA's role in the US surveillant assemblage, it serves as a useful case for further analysis and as an example of the sousveillance potential

of these data. Exploring this specific case within the 2000s period also presents an opportunity to more closely examine the extensive level of privatization shown by the earlier network analyses.

Figure 5: NSA network graph, 2000–2009.

Figure 5 represents the ego network graph for the NSA during the 2000s era.⁸ Immediately apparent is that the closest ties to the NSA are the different branches of the US military. Outside of these organizations, however, the vast majority of the organizations connected to the NSA are private companies. Among the most prominent private actors within the NSA’s network are Leidos, Booz Allen, General Dynamics, L3 Technologies, SAIC, and CACI. This is perhaps unsurprising as these are some of the biggest corporations in the privatized intelligence industry. Leidos, Booz Allen, SAIC, General Dynamics, and CACI alone employ nearly 80% (45,000 people) of the private contractors working in the intelligence industry (Shorrock 2016).

Most of the companies listed above provide different surveillance-related software, hardware, and analytics services. For example, among the connections to the NSA is Praescent Analytics. They note on their website that they “partner with a series of cutting-edge software companies to deliver training, integration, customization, and embedded analytic services to clients across the public and private sectors” (Praescent Analytics 2019). Among the software companies listed as Praescent Analytics’ partners are IBM, Semantic Research, and Palantir Technologies. Documents released by Edward Snowden revealed Palantir as the company that built the software for the NSA program XKEYSCORE, which collected citizens’ emails,

⁸ This graph represents a subset of the NSA’s ego network. For visual purposes only the most prominent actors in terms of degree are included here.

chats, web-browsing traffic, pictures, documents, voice calls, webcam photos, web searches, and much more (Biddle 2017). Palantir is perhaps the greatest example of the dangers of growing private involvement in the US surveillant assemblage. The company has been involved in recent controversy over its ties to the Cambridge Analytica scandal (Confessore and Rosenberg 2018), recent deportation efforts by ICE (Woodman 2017), and the development of intrusive digital analytics systems for local law enforcement agencies (Brayne 2017; Harris 2017). This illustrates how the US surveillant assemblage, including this particular sub-network, also represents a vast network of infrastructure, socio-technological objects, knowledge, and discourses that make it possible for these organizations to carry out surveillance. It is these heterogeneous elements of the US surveillant assemblage that make it so powerful and, in some cases, dangerous.

Discussion

Before discussing the conclusions of this article, it is essential to acknowledge the limitations of these data. First, as previously mentioned, the actual nature of these connections and the extent of interaction between organizations to engage in surveillance is unknown. Similarly, the directionality of these relationships is unclear; there is no way of knowing (with these data alone), for example, whether a government agency has contracted a private entity for surveillance-related services or if there is a mutual exchange of goods and services. Second, this analysis is unable to fully account for the effects of external historical events, such as changes to regulatory or institutional dynamics, on the structure of the US surveillant assemblage. While some historical context is provided with the network analyses, the conclusions made about the role of these events in producing these changes are speculative. Third, the use of social media data inherently limits the generalizability of these findings. While large digital datasets are often celebrated for providing access to “complete” populations, specific populations are more likely to turn up in datasets like the one used here (Harris 2017). This dataset also fails to capture individuals who do not have a LinkedIn profile. Due to the sampling method of the scraping algorithm, it also does not capture those who had a LinkedIn profile and were involved in surveillance practices but did not explicitly mention a surveillance program in their profiles. Additionally, individuals working in particular industries or organizations may be less likely to use LinkedIn or to report their involvement in surveillance programs in their profiles. This may explain why universities, despite being known collaborators on defense research, are less prevalent in the data than private companies. Lastly, these data are limited in that they tend to favor more recent linkages between organizations due to the younger demographic of LinkedIn users and the tendency for users in general to favor recent job history in their profiles. Given the stark differences in terms of sample size and composition of the four time periods, conclusions about the structural and topological changes to the US surveillant assemblage over time, although supported by the current analyses, remain speculative and caution should be exercised in attempting to extrapolate these results.

This paper draws on Haggerty and Ericson’s (2000) concept of “surveillant assemblage” to critically analyze the amalgamation of public and private actors involved in carrying out surveillance in the US. It also makes a methodological contribution to the field of surveillance studies by illustrating the potential of digital sousveillance as a method for studying surveillance. In doing so, this paper aligns the politics of assemblage thinking and sousveillance to challenge the idea that digital data merely serves as a conduit for surveillance and exploitation. It instead highlights the potential of these data and methods as sousveillance tools, conceivably allowing private citizens and scholars alike the ability to “watch the watchers.” Digital sousveillance thus serves as a form of resistance to the dominant actors and power structures of mass surveillance. Future work could utilize digital sousveillance, as used here, to complement theoretical and historical studies of surveillance and pursue new interdisciplinary and collaborative studies of surveillance. Network analysis could, for instance, be used to build upon these analyses to trace networks of individual actors engaged in surveillance. Other methods, such as topic modeling or sentiment analysis, could lend empirical support to past studies of surveillance discourse and understandings of how discourse is framed by those actively engaged in surveillance practices.

The results of the network analysis indicate that the US surveillant assemblage is becoming increasingly *privatized*; indeed, the line between “public” and “private” is becoming blurred as private organizations are, at an increasing rate, partnering with the US government to engage in mass surveillance. The recent growth of the private sector’s involvement in surveillance practices poses significant problems. By subjecting individuals to surveillance, governments, as well as their corporate partners who control the digital realm, can monitor and silence differing opinions and views. This global digital space, referred to by some as the “digital commons,” plays a crucial role in allowing for the democratization of expression as well as increased civic engagement and participation (Wonders, Solop, and Wonders 2012). The growing commodification of the global digital commons by private actors is problematic in that it subjects the users of this space to increased observation and scrutiny and facilitates the movement of private information into the hands of other parties such as law enforcement agencies (Brayne 2017). As a result, the digital commons becomes a “backdoor” (Crampton, Roberts, and Poorthuis 2014) or “revolving door” (Hayes 2012) through which government(s) can not only observe users of the space but also do so in ways that avoid laws and regulations meant to protect those users. The simple fact that the surveilled know they are being watched has a chilling effect on civil liberties and the freedom of expression that is fundamental to a functioning democracy (Richards 2012). The growing involvement of private actors within the US surveillant assemblage thus presents a grave danger to the global digital commons and democracy as we know it. More work, by activists and scholars alike, is needed to continue to unmask these actors and to allow the public to better understand their entanglement with this vast surveillant assemblage.

References

- Abrahamsen, Rita, and Michael C. Williams. 2009. Security Beyond the State: Global Security Assemblages in International Politics. *International Political Sociology* 3 (1): 1–17.
- Ball, Kirstie, Ana Isabel Canhoto, Elizabeth Daniel, Sally Dibb, Maureen Meadows, and Keith Spiller. 2015. *The Private Security State?: Surveillance, Consumer Data and the War on Terror*. Copenhagen: Copenhagen Business School Press.
- Bamford, James. 2009. *The Shadow Factory: The Ultra-Secret NSA from 9/11 to the Eavesdropping on America*. New York: Anchor.
- Biddle, Sam. 2017. How Peter Thiel’s Palantir Helped the NSA Spy on the Whole World. *The Intercept* (blog), February 22. <https://theintercept.com/2017/02/22/how-peter-thiels-palantir-helped-the-nsa-spy-on-the-whole-world/> [accessed February 2, 2019].
- Bleiker, Roland. 2014. Visual Assemblages: From Causality to Conditions of Possibility. In *Reassembling International Theory: Assemblage Thinking and International Relations*, edited by Michele Acuto and Simon Curtis, 75–81. London: Palgrave Macmillan UK.
- boyd, danah, and Kate Crawford. 2012. Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society* 15 (5): 662–679.
- Brayne, Sarah. 2017. Big Data Surveillance: The Case of Policing. *American Sociological Review* 82 (5): 977–1008.
- Buchanan, Ian. 2015. Assemblage Theory and Its Discontents. *Deleuze Studies* 9 (3): 382–392.
- Bueger, Christian. 2014. Thinking Assemblages Methodologically: Some Rules of Thumb. In *Reassembling International Theory: Assemblage Thinking and International Relations*, edited by Michele Acuto and Simon Curtis, 58–66. London: Palgrave Macmillan UK.
- Cheney-Lippold, John. 2017. *We Are Data: Algorithms and the Making of Our Digital Selves*. New York: New York University Press.
- Collier, Stephen J., and Aihwa Ong. 2005. Global Assemblages, Anthropological Problems. In *Global Assemblages: Technology, Politics, and Ethics as Anthropological Problems*, edited by Stephen J. Collier and Aihwa Ong, 3–21. New York: Wiley-Blackwell.
- Confessore, Nicholas, and Matthew Rosenberg. 2018. Spy Contractor’s Idea Helped Cambridge Analytica Harvest Facebook Data. *The New York Times*, March 28. <https://www.nytimes.com/2018/03/27/us/cambridge-analytica-palantir.html> [accessed February 2, 2019].
- Crampton, Jeremy W., Susan M. Roberts, and Ate Poorthuis. 2014. The New Political Economy of Geographical Intelligence. *Annals of the Association of American Geographers* 104 (1): 196–214.
- Deleuze, Gilles, and Felix Guattari. 1987. *A Thousand Plateaus*. Minneapolis: University of Minnesota Press.
- Donohue, Laura K. 2016. *The Future of Foreign Intelligence: Privacy and Surveillance in a Digital Age*. Oxford: Oxford University Press.
- Eubanks, Virginia. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin’s Press.
- Fyfe, Nicholas R., and Jon Bannister. 1996. City Watching: Closed Circuit Television Surveillance in Public Spaces. *Area* 28 (1): 37–46.
- Greenwald, Glenn. 2014. *No Place to Hide: Edward Snowden, the NSA, and the US Surveillance State*. New York: Macmillan.

- Greenwald, Glenn, and Ewen MacAskill. 2013. NSA Prism Program Taps in to User Data of Apple, Google and Others. *The Guardian*, June 7. <https://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data> [accessed September 22, 2018].
- Haggerty, Kevin D., and Richard V. Ericson. 2000. The Surveillant Assemblage. *The British Journal of Sociology* 51 (4): 605–622.
- Haggerty, Kevin D., and Amber Gazso. 2005. Seeing beyond the Ruins: Surveillance as a Response to Terrorist Threats. *The Canadian Journal of Sociology* 30 (2): 169–187.
- Harris, Mark. 2017. How Peter Thiel's Secretive Data Company Pushed into Policing. *Wired*, August 9. <https://www.wired.com/story/how-peter-thiels-secretive-data-company-pushed-into-policing/> [accessed February 2, 2019].
- Hayes, Ben. 2012. The Surveillance-Industrial Complex. In *Routledge Handbook of Surveillance Studies*, edited by Kirstie Ball, Kevin D. Haggerty, and David Lyon, 167–175. New York: Routledge.
- Introna, Lucas D., and Amy Gibbons. 2009. Networks and Resistance: Investigating Online Advocacy Networks as a Modality for Resisting State Surveillance. *Surveillance & Society* 6 (3): 233–258.
- Kumaraguru, Ponnuram, and Lome Faith Cranor. 2005. A Survey of Westin's Studies. *Institute for Software Research International*.
- Latour, Bruno. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.
- Lisle, Debbie. 2014. Energizing the International. In *Reassembling International Theory: Assemblage Thinking and International Relations*, edited by Michele Acuto and Simon Curtis, 67–74. London: Palgrave Macmillan UK.
- Lyon, David. 2001. *Surveillance Society: Monitoring Everyday Life*. London: McGraw-Hill Education.
- . 2014. Surveillance, Snowden, and Big Data: Capacities, Consequences, Critique. *Big Data & Society* 1 (2): 1–13.
- Mann, Steve, and Joseph Ferenbok. 2013. New Media and the Power Politics of Sousveillance in a Surveillance-Dominated World. *Surveillance & Society* 11 (1/2): 18.
- Mann, Steve, Jason Nolan, and Barry Wellman. 2003. Sousveillance: Inventing and Using Wearable Computing Devices for Data Collection in Surveillance Environments. *Surveillance & Society* 1 (3): 331–355.
- Marcus, George E., and Erkan Saka. 2006. Assemblage. *Theory, Culture & Society* 23 (2–3): 101–106.
- Matzner, Tobias. 2016. Beyond Data as Representation: The Performativity of Big Data in Surveillance. *Surveillance & Society* 14 (2): 197–210.
- Murakami Wood, David. 2013. What Is Global Surveillance? Towards a Relational Political Economy of the Global Surveillant Assemblage. *Geoforum* 49: 317–326.
- Murphy, Maria Helen. 2014. The Pendulum Effect: Comparisons between the Snowden Revelations and the Church Committee. What Are the Potential Implications for Europe? *Information & Communications Technology Law* 23 (3): 192–219.
- Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.
- Praescent Analytics. 2019. Solutions. <https://praescentanalytics.com/solutions> [accessed February 2, 2019].
- Raley, Rita. 2013. Dataveillance and Countervailance. In *Raw Data is an Oxymoron*, edited by Gitelman, Lisa, 121–146. Cambridge, MA: MIT Press.
- Raynes-Goldie, Kate. 2010. Aliases, Creeping, and Wall Cleaning: Understanding Privacy in the Age of Facebook. *First Monday* 15 (1).
- Richards, Neil M. 2012. The Dangers of Surveillance. *Harvard Law Review* 126: 1934.
- Sassen, Saskia. 2006. *Territory, Authority, Rights: From Medieval to Global Assemblages*. Cambridge: Cambridge University Press.
- Schwarz, Frederick A.O., and Aziz Z. Huq. 2008. *Unchecked and Unbalanced: Presidential Power in a Time of Terror*. New York: The New Press.
- Shorrock, Tim. 2016. 5 Corporations Now Dominate Our Privatized Intelligence Industry. *The Nation*, September 8. <https://www.thenation.com/article/five-corporations-now-dominate-our-privatized-intelligence-industry/> [accessed January 25, 2019].
- van der Vlist, Fernando N. 2017. Counter-Mapping Surveillance: A Critical Cartography of Mass Surveillance Technology After Snowden. *Surveillance & Society* 15 (1): 137–157.
- Wakefield, Alison. 2002. The Public Surveillance Functions of Private Security. *Surveillance & Society* 2 (4): 529–545.
- Wasserman, Stanley, and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
- Wonders, Brooke J., Frederic I. Solop, and Nancy A. Wonders. 2012. Information Sampling and Linking: Reality Hunger and the Digital Knowledge Commons. *Contemporary Social Science* 7 (3): 247–262.
- Woodman, Spencer. 2017. Palantir and Trump's Deportation Machine. *The Intercept*, March 2. <https://theintercept.com/2017/03/02/palantir-provides-the-engine-for-donald-trumps-deportation-machine/> [accessed February 2, 2019].
- Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power*. New York: Profile Books.