Digital Forensics: Repurposing Google Analytics IDs

Rogers, R.

DOI
10.5117/9789462989511
10.5117/9789462989511_ch33
10.1017/9789048542079.034

Publication date
2021

Document Version
Final published version

Published in
The Data Journalism Handbook

33. Digital Forensics: Repurposing Google Analytics IDs

Richard Rogers

Abstract
This chapter describes a network discovery technique on the basis of websites sharing the same Google Analytics and/or AdSense IDs.

Keywords: digital methods, digital forensics, anonymous sources, network mapping, Google Analytics, data journalism

When an investigative journalist uncovered a covert network of Russian websites in July 2015 furnishing disinformation about Ukraine, not only did this revelation portend the state-sponsored influence campaigning prior to the 2016 US presidential elections,[1] it also popularized a network discovery technique for data journalists and social researchers (Alexander, 2015). Which websites share the same Google Analytics ID (see Figure 33.1)? If the websites share the same ID, it follows that they are operated by the same registrant, be it an individual, organization or media group. The journalist, Lawrence Alexander, was prompted in his work by the lack of a source behind emaidan.com.ua, a website that appears to give information about the Euromaidan protests in 2013–2014 in Ukraine that ultimately upended the pro-Russian Ukrainian president in favour of a pro-Western one. In search of the source, and "intrigued by its anonymity," Alexander (2015) dug into the website code.

1 A longer version of this chapter is available in Rogers, R. (2019). *Doing digital methods*. SAGE. The author would like to acknowledge the groundwork by Mischa Szpirt. For more on this approach, see Rogers, R. (2019). *Doing digital methods*. SAGE (Chapter 11), and Bounegru, L., Gray, J., Venturini, T., & Mauri, M. (Comp.) (2017). *A field guide to "fake news": A collection of recipes for those who love to cook with digital methods*. Public Data Lab (Chapter 3).

Bounegru, L. and J. Gray (eds.), *The Data Journalism Handbook: Towards a Critical Data Practice*. Amsterdam: Amsterdam University Press, 2021
DOI 10.5117/9789462989511_CH33

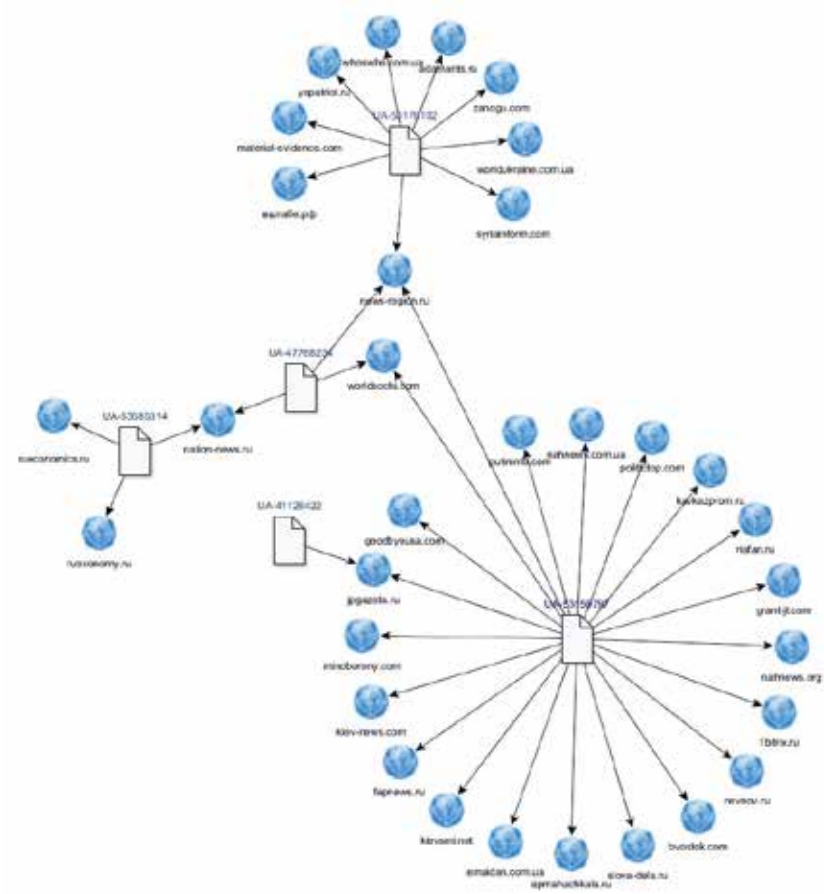

Figure 33.1. Website network discovered through (shared) Google Analytics IDs. Source: Alexander, L. (2015, July 13). Open-source information reveals pro-Kremlin web campaign. Global Voices. https://globalvoices.org/2015/07/13/open-source-information-reveals-pro-kremlin-web-campaign/

Viewing the source code of the web page, he found a Google Analytics ID, which he inserted into reverse lookup software that furnishes a list of other websites using the same ID.[2] He found a (star-shaped) network of a Google Analytics ID linked to eight other websites (in Figure 33.1 at the top of the diagram), sharing a similar anti-Ukraine narrative. One of those websites also used an additional Google Analytics ID, which led to another cluster of related websites (in Figure 33.1 at the bottom to the right), also of similar political persuasion. Examining the WHOIS records of several of

2 The lookup may also yield each website's IP address, Google AdSense ID, WHOIS domain record and other telling information.

Figure 33.2. Google Analytics ID. Source: Baio, A. (2011, November 15). Think you can hide, anonymous blogger? Two words: Google analytics. Wired. https://www.wired.com/2011/11/goog-analytics-anony-bloggers/

these domains, he found an associated email address, and subsequently a person's profile and photo on VKontakte, the Russian social networking site. The name of this person he then found on a leaked list of employees from the Internet Research Agency in St Petersburg, known as the workplace of the Russian government-sponsored "troll army" (Chen, 2015; Toler, 2015). Drawing links between data points, Alexander put a name and face on a so-called Russian troll. He also humanized the troll, somewhat, by pointing to his Pinterest hobby page, where there is posted a picture of Russian space achievements. The troll is a Cosmonaut space fan, too.

Employing so-called "open-source intelligence" (OSINT) tools as discovery techniques (and also digital methods in the sense of repurposing Google Analytics and reverse lookup software), Alexander and other journalists make and follow links in code, public records, databases and leaks, piecing it all together for an account of "who's behind" particular operations (Bazzell, 2016). "Discovery" is an investigative or even digital forensics approach for journalistic mining and exposure, where one would identify and subsequently strive to contact the individual, organization or media group to interview them, and grant them an opportunity to account for their work.[3] The dual accountings—the journalist's discovery and the discovered's explanation—constitute the story to be told. The purpose is to make things public, to wring out of the hairy code of websites the covert political work being undertaken, and have this particular proof be acknowledged (Latour, 2005).

Google Analytics ID detective work has a lineage in the practice of unmasking anonymous online actors through exploits, or entry points to personally

3 Digital forensics has its roots in the investigation of corporate fraud through techniques such as "data carving," which enable the retrieval of deleted files.

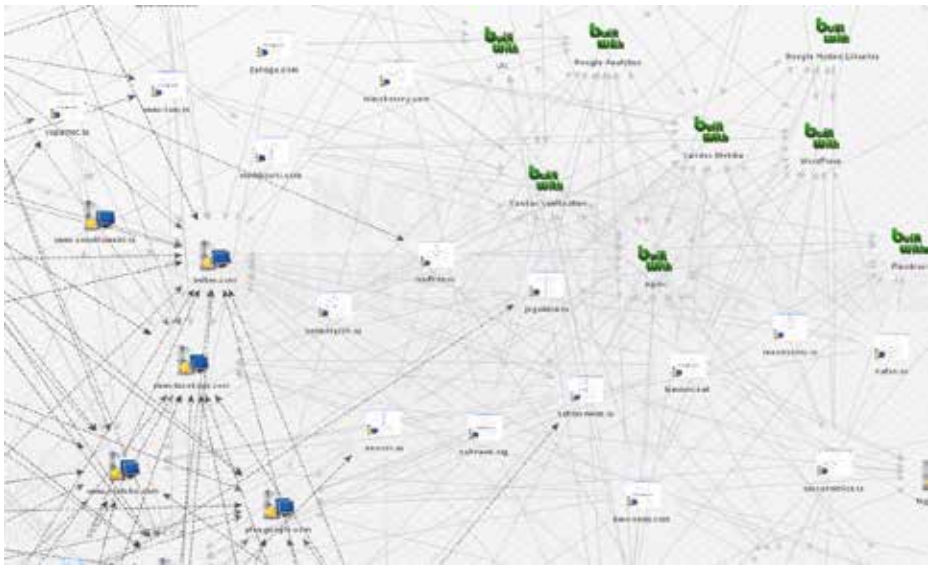

Figure 33.3. Embedded digital objects on websites, depicted as network diagram. Source: Alexander, L. (2015, July 13). Open-source information reveals pro-Kremlin web campaign. Global Voices. https://globalvoices.org/2015/07/13/open-source-information-reveals-pro-kremlin-web-campaign/.

identifiable data that have not been foreseen by its creators. Mining Google Analytics IDs for network discovery and mapping is also a repurposing exercise, using the software in unintended fashion for social research. The originator of the technique, Andy Baio, a journalist at *Wired* magazine, tells the story of an anonymous blogger posting highly offensive material, who had covered his tracks in the "usual ways": "hiding personal info in the domain record, using a different IP address from his other sites, and scrubbing any shared resources from his WordPress install" (Baio, 2011). Baio ID'd him because the blogger shared a Google Analytics ID with other websites he operated in full view. The cautionary tale about this discovery and unmasking technique concludes with Baio providing a safety guide for other anonymous bloggers *with a just cause*, such as those monitoring Mexican drug cartels, whose discovery could lead to danger or even loss of life. Here one also could test the robustness of the anonymity, and inform the journalists working undercover online of any vulnerabilities or potential exploits.

By way of conclusion, I offer a research protocol for network discovery using Google Analytics IDs, summarized in the list below:

Curate a list of websites that do not provide their sources.
Locate Google Analytics and AdSense IDs.

Insert URL list into reverse lookup software such as dnslytics.com.
Seek websites that share the same IDs.
Thematically group and characterize the websites sharing IDs.
Consider network visualization using Gephi.

Works Cited

Alexander, L. (2015, July 13). Open-source information reveals pro-Kremlin web campaign. *Global Voices*. https://globalvoices.org/2015/07/13/open-source-information-reveals-pro-kremlin-web-campaign/

Baio, A. (2011, November 15). Think you can hide, anonymous blogger? Two words: Google analytics. *Wired*. https://www.wired.com/2011/11/goog-analytics-anony-bloggers/

Bazzell, M. (2016). *Open source intelligence techniques: Resources for searching and analyzing online information* (5th ed.). CreateSpace Independent Publishing Platform.

Chen, A. (2015, June 2). The agency. *The New York Times Magazine*. https://www.nytimes.com/2015/06/07/magazine/the-agency.html

Latour, B. (2005). From realpolitik to dingpolitik—An introduction to making things public. In B. Latour & P. Weibel (Eds.), *Making things public: Atmospheres of democracy* (pp. 14–41). MIT Press. http://www.bruno-latour.fr/node/208.html

Toler, A. (2015, March 14). Inside the Kremlin troll army machine: Templates, guidelines, and paid posts. *Global Voices*. https://globalvoices.org/2015/03/14/russia-kremlin-troll-army-examples/

About the Author

Richard Rogers is Professor of New Media and Digital Culture in Media Studies at the University of Amsterdam and director of the Digital Methods Initiative as well as the Netherlands Research School for Media Studies.