Weaponizing Twitter Litter: Abuse-Forming Networks and Social Media

Hal Berghel, University of Nevada, Las Vegas

Instead of liberating us from the biases of the educated among us, the Internet has saddled us with the biases of the unreasoned among us.

Many of us have serious reservations against creating user accounts with, or even using online services from, tech companies with onerous privacy policies. Google, for one, is especially aggressive harvesting personal data on users (www.google.com/policies/privacy). However, as Eli Pariser shows in his recent book *The Filter Bubble*,[1] we have more to fear from online services than invasions of privacy: many of these services also manipulate the online information spaces that shape our decisions. In a very real sense, Pariser anticipated the problems with fake news and partisan trolling that befell the 2016 US national elections.

It's easy enough to avoid using Google services like Gmail and its search engine. Protonmail (https://protonmail.com) and DuckDuckGo (https://duckduckgo.com), for example, are both viable, independent, privacy-respecting substitutes. In addition, there are large corporate offerings like Microsoft Outlook and Bing that appear to be less invasive of personal privacy than Google (www.privacy.microsoft.com/en-us/privacystatement). However, as Pariser demonstrates, there's no way to insulate ourselves from the censorship and rectified flow of information that takes place without our knowledge and consent. Even if we once accepted the premise that surrendering some privacy is the price we paid for free online services, it's gone too far: we're not only losing privacy to services we use, we're losing it to services we *avoid* because more and more of them are digitally interlocked through information sharing. Further, these same services are placing us in an information cocoon.

ABUSE-FORMING NETWORKS

(Robert) Metcalf's law[2] holds that the connectivity value (aka utility) of a network is proportional to the square of

0018-9162/18/$33.00 © 2018 IEEE

the number of nodes, n^2. It assumes that there's some measurable value that increases with the number of possible pairwise connections. David P. Reed[3] extends this reasoning to claim that there's also some measureable value to the number of potential groups that can form within a network, and that this number grows exponentially by the number of nodes, 2^n—what he calls group-forming networking. Andrew Odlyzko and Benjamin Tilly[4] seek to moderate Reed's formula by taking into account the law of diminishing returns within group formation, arguing that the value of the number of groups is $n\log(n)$.

All of these are plausible measures of some sort of value for networks,[2,5] but they miss one important factor that has become critical in the past few decades: the potential aggregate "cost" or "penalty" to the users/participants through the leaking of information about them, the surveillance of their daily lives and actions, the loss of their time due to unnecessary or unwanted distractions, the potential loss of their personal sovereignty and liberty, and the potential for their social manipulation by tyrants, demagogues, dictators, and other manifestations of the power elite.[6] I refer to this phenomenon as *abuse-forming networking*.

We take Reed's law as our starting point. As with Reed, we note that if group-forming networks are integrated, their aggregate value is proportional to the product of their individual values, $2^m \times 2^n \times \ldots$, that is, to the product of the number of power sets. Something very similar happens to the degree of possible abuse of the users/participants. This all starts from the fact that the individual connections are detectable—at least by those who provide the networking. While this in itself might seem harmless, after the revelations of Edward Snowden we all recognize that the potential for abuse goes way beyond the knowledge of who's connected to what network. In fact, this was well known long before Snowden was born.[7,8] As I've written, Snowden's real legacy is showing that many of our suspicions were justified.[9]

But the abuse just begins with the identifiability of the connections. Add to that the traffic metadata: how many times node *i* received traffic from nodes *j* and *k*, when and to whom node *l* communicated, how often nodes associated with event *x* communicated with nodes associated with organization *y*, and so on. This is the stuff of which the NSA's Section 215 bulk metadata collection program is made.[10] The number of different identifiable signatures from this metadata far exceeds the size of the number of groups that might share information.

Now let's add some object-level data. Modern integrated marketing firms collect thousands of individual pieces of information on all of us: first and last name, Social Security number, mother's maiden name, known associates, family history, birth/death dates, income history, credit history, current and past addresses, employers, driver's license and other ID numbers, email addresses, bank card data, transaction records from many merchants, name/type of pets, associated IP addresses, arrest record, voting registration information and party affiliation, potential inheritance, medical conditions/needs, time/date/duration/SMS routing information types and destinations of all electronic communications (telephony, email, faxes), computer and mobile device information (such as MAC addresses, serial numbers, OS versions, and browser versions), geolocation information (GPS coordinates), biometric data (including DNA data, voiceprints, and face images captured by satellites, drones, and surveillance cameras), cookies and contents from application caches, clickstream data, associated ISPs and telcos, what you watch on YouTube, Netflix, Amazon, and so on. This is just a partial list of the information routinely collected by businesses and doesn't include the much more invasive data collected by private and government security agencies, all of which can be used as "selectors" to search through yobibytes ($1{,}024^8$) of global, digital stored data on all of us. For more details, visit the websites of the Electronic Privacy Information Center (www.epic.org) and Privacy International (www.privacyinternational.org).

My point is that if we consider the abuse of individual rights as related to the use of information about *n* individuals without their expressed permission, the potential for abuse derived from the combined object and meta-level data from the networks has to be significantly larger than 2^n. So let's estimate the upper bound at $2^n \times 2^k$, which is Reed's number of potential subgroups of network users times the power set of the number of different collections of attributes that can be defined over all groups of users. This has some intuitive justification, for 2^k is the number of all possible associations that that one might make

Our online world has introduced two new forms of information corruption: source displacement and decontextualization.

of all of the groups based on the individual data elements that correspond to the network members/users such as those mentioned in the preceding paragraph. That is, the latter set would include such subsets as all members with brown dogs, the set of all groups with at least one member with a brown dog and a subset of members whose name is Phil, the number of dog owners that subscribed to white supremacist literature, and so on. With k in the thousands (a realistic assumption), the ways in which the threads of association can be created is exceedingly large. In general the public isn't appraised of how, by whom, and for what purposes these threads are created, but it would be a mistake to underestimate their use by political organizations and operatives, intelligence and investigative agencies, law enforcement, criminal organizations, phishers, scammers, spammers, NGOs, marketing companies, and so forth—almost all of which instances are without the user's knowledge or consent.

Thus, Reed's law actually vastly understates the aggregate cost of abuse to network users in terms of the loss of privacy, misuse of personal data through identity theft or financial fraud (think Equifax hack), or downstream negative externalities from aggressive data harvesting. Such numbers, too, should be estimated, but under the rubric of abuse formation. We note that because the abuse is largely externally imposed rather than self-organizing, no constraining cognitive limit applies. Put simply, while there's a limit to the number of stable, cohesive groups with whom an individual can associate,[11] there's no limit to the number of external groups that can heap abuse on the individual by means of data mining automation. As an aside, an excellent history of the origins of the American version of the surveillance state can be found in Alfred W. McCoy's recent book, *Policing America's Empire*.[12]

BEYOND ABUSE BUILDING TO TRUTH FABRICATION

We extend our analysis to more subtle variations of online abuse, most importantly through the manipulation of the public through a constant stream of lies, prevarications, untruths, distortions, and so forth derived from fake news sources, trolls, propaganda channels, partisan media, and the like, for which we have no known protection and few working models.[13] According to Pariser, we all live in a filter bubble where information flow is carefully regulated by upstream manipulators under the banner of "personalization." He points out how the online "people-powered news" that many of us anticipated has been corrupted by merchants of faux news. Where 100 years ago major news sources "had a sense of ethics and public responsibility baked in however imperfectly, … [today's] filter bubble does not."

We have now entered the era of "lock-on" news feeds that nourish the addiction to misinformation.

Of course, censorship has been a constant companion to democracies—albeit in softer, self-induced forms than is found in dictatorships. But our online world has introduced two new forms of information corruption: source displacement and decontextualization. A century ago, fake news for the most part had to be orchestrated by corporate mass media within view of many critical eyes. The online revolution makes it possible for any individual or group of -ists (racists, anarchists, tribalists, and so on), no matter how small, to launch their own fake news service with an inexpensive computer and an Internet connection. While the disintermediation of the editor/publisher disempowered them to be sure, it also empowered delusionists, narcisissists, and sociopaths whose presence looms large over networked neophytes. Our educational system simply failed to anticipate this possibility and underemphasized the criticality of individual fact checking of all information sources.

Instead of a panel of professional journalists filtering news, we now have propagandists and prevaricators filling the role. Instead of liberating us from the biases of the reasonable among us, the Internet has saddled us with the biases of the unreasonable among us—at least when it comes to fake news. Pariser credits Columbia Law School professor and *New York Times* op-ed writer Tim Wu with a particularly apropos remark: "The rise of networking did not eliminate intermediaries, but rather changed who they are." Indeed, the disreputable Internet disintermediaries as a group form a paradigmatic case of an untrusted system—there's precious little journalism or scholarship involved. Unfortunately, too many minds are attentive to the vacuous content.

What's more, even the prevarications and propaganda are filtered and bundled for us—we don't get our misinformation unadulterated either. This activity falls under the category of "personalized information services," which is a euphemism for filtered information with manipulative potential. The key to this personalization lies in external forces pushing information toward you that primarily serves *their* interests. Whether it serves yours is of secondary concern.

The key to the successful spread of misinformation and false reporting is decontextualization, as it removes the contextual links necessary for confirmation/disconfirmation. You won't find extensive reference lists to academic papers in tribalists' online resources: objectivity is an enemy to the tribe! If an online news service sees that you like chocolate-covered marshmallows, information about that will be delivered to you. The same goes for perceived interests in white nationalist or anarchist literature, partisan politics, and hate groups. What these personalized news services do is drive consumer opinion to the extremes, as they tend to reinforce existing biases and stereotypes rather than provide other points of view. This just further polarizes the polity. It should be remembered that the cause of the tribalist rejection of mainstream news was never that it was demonstrably false, but that it was inconvenient—it didn't comport with the preferred opinion.

We've now entered the era of "lock-on" news feeds that nourish the addiction to misinformation. Instead of looking for counterexamples to our worldview, we allow others to filter them, thereby ensuring the growth of collective ignorance and prejudice. This surfaces in subtle ways these days. Publishers hire selectivity readers to ensure that readers aren't accidentally offended. Amazon has review Nazis to limit reviewer bias: claiming a product is far superior to one product but inferior to another is verboten. These companies have not only diminished the value of free expression, they've lost sight of the criticality of the First Amendment to free societies.

There's a striking parallel between abuse-forming networking and phishing: both involve technical subterfuge (antisocial use of networking technology), perception management (manipulation of the public), and social engineering (motivating people to do something that they probably wouldn't have done otherwise, such as subscribe to a controversial blog).[14] Ruth Alexander's recent installment of the BBC series *The Inquiry* [15] also extends Pariser's work on filter bubbles to include data mining for psychometric profiling. I'll return to these topics in future columns. C

REFERENCES

1. E. Pariser, *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*, Penguin Books, 2011.
2. R. Metcalf, "Metcalf's Law after 40 Years of Ethernet," *Computer*, vol. 46, no. 12, 2013, pp. 26–31.
3. D.P. Reed, "That Sneaky Exponential—Beyond Metcalfe's Law to the Power of Community Building," 1999; www.deepplum.com/dpr/locus/gfn/reedslaw.html.
4. A. Odlyzko and B. Tilly, "A Refutation of Metcalfe's Law and a Better Estimate for the Value of Networks and Network Interconnections," 2005; www.dtc.umn.edu/~odlyzko/doc/metcalfe.pdf.
5. R. Tongia and E.J. Wilson III, "The Flip Side of Metcalfe's Law: Multiple and Growing Costs of Network Exclusion," *Int'l J. Communication*, vol. 5, 2011, pp. 665–681.
6. C.W. Mills, *The Power Elite*, 2nd ed., Oxford Univ. Press, 2000.
7. J. Bamford, *The Puzzle Palace: Inside the National Security Agency, America's Most Secret Intelligence Organization*, Penguin Books, 1983.
8. T. Weiner, *Enemies: A History of the FBI*, Random House, 2013.
9. H. Berghel, "Mr. Snowden's Legacy," *Computer*, vol. 47, no. 4, 2014, pp. 66–70.
10. D. Kravets, "Court Says It's Legal for NSA to Spy on You Because Congress Says It's OK," *Ars Technica*, 29 Oct. 2015; https://arstechnica.com/tech-policy/2015/10/court-says-its-again-legal-for-nsa-to-spy-on-you-because-congress-says-its-ok.
11. R.I.M. Dunbar, "Neocortex Size as a Constraint on Group Size in Primates," *J. Human Evolution*, vol. 22, no. 6, 1992, pp. 469–493.
12. A.W. McCoy, *Policing America's Empire: The United States, the Philippines, and the Rise of the Surveillance State*, Univ. of Wisconsin Press, 2009.
13. H. Berghel, "Disinformatics: The Discipline behind Grand Deceptions," *Computer*, vol. 51, no. 1, 2018, pp. 89-93.
14. H. Berghel, J. Carpinter and J.-Y. Jo, "Phish Phactors: Offensive and Defensive Strategies," *Advances in Computers*, vol. 70, 2007, pp. 223–268.
15. R. Alexander, "How Powerful Is Facebook's Algorithm?," *The Inquiry*, BBC World Service, 23 Apr. 2017; www.bbc.co.uk/programmes/p04zvqtx.

HAL BERGHEL is an IEEE and ACM Fellow and a professor of computer science at the University of Nevada, Las Vegas. Contact him at hlb@computer.org.