18

Making People and Influencing Friends: Citation Networks and the Appearance of Significance

Finn Brunton

Like many academics, I became interested in predatory and "spam" journals and new forms of academic misconduct when I started receiving the invitations to submit an article, any article—fast turnaround! low prices!—or present at a conference, review for a journal, or even join their board. My initial amusement at being asked to submit something about immunology or linguistics (a testament to the interdisciplinary character of modern university life, perhaps) was in retrospect a stage of professionalization through which every academic with an email address now passes: being a potential mark for a family of new, related scams. I had more than an accidental interest in this, however. I had written a book about the history of spam, and the question of what made these journals *spam* journals was something I wanted to investigate (Brunton, 2013). They are both like and unlike other forms of spam—as Mario Biagioli puts it in the introduction to this volume and Alexandra Lippman explains in her chapter, they are neither spam nor not spam, but *spammish*, and part of larger systems of imitation, collusion, and gaming metrics. I will discuss both parallels and divergences between spam itself and academic platforms dubbed with that title, and present you with two closely related lessons from spam for studying academic metrics and their abuse. The first has to do with how we study the secondary markets around misconduct and the second with understanding the larger networks that benefit from the work of platforms for scholarly misbehavior. Together, I hope these will enrich our picture of what academia is now in danger of becoming, through how it can be *gamed*.

First, though, why is it that these journals are called "spam journals"? There's an obvious answer to this question, and a deeper answer with more to tell us. The obvious answer is straightforward, but still worth dwelling on for clarity: these are journals that most of us encounter through the delivery channel of unrequested and indiscriminate email. Indeed, these

journals replicate many classic spam strategies in the content of their messages and their web presence—working off the same well-established playbook used by stock touts and boiler-room investment operations, deadstock salesmen and off-brand pharmacists, pornographers and confidence men.

Their messages feature the same array of rhetorical appeals and come-ons, though geared toward an academic audience: lower fees, faster publishing cycles, sympathetic reviewers, more favorable impact metrics. (Given what is about to be discussed, the reader should add quotes liberally: favorable "impact metrics," sympathetic "reviewers." The only solidly realistic thing is the fees.) Like phishing emails—which try to trick recipients into mistaking an illegitimate message for one from Facebook or their bank by aping the style, design, and markup—many of these messages have the boxy, early-2000s design particular to more reputable journals, sometimes in more or less direct reference. (Of course, we also see cases of *actual* academic journal phishing, with scammers replicating the title, text, and sometimes look and feel of a reputable journal's site so unsuspecting scholars will submit papers, with fees, for review.[1]) They use US and European post office boxes as mail drops, to produce an appearance of legitimacy, much as the 1990s spammers would have business addresses in the signature of their messages, to give the impression of an actual office somewhere, an institutional relationship, the possibility of redress—although with the distinction that the spam journal drops often actually *exist,* to redirect mailed checks.

The same strategy applies to the use of evasive and misleading contact information and lists of personnel, and links to legitimate venues as an indirect assertion of bona fides—as advance-fee fraud spammers would include links to reliable news organizations and trustworthy banks to imply their trustworthiness in turn. Burkhard Morgenstern's chapter in this book shows the recruitment process for getting real, legitimate names on illegitimate organizations at its most egregious; some groups avoid the difficulty in gathering reputable editors by featuring an editorial board of wholly fictional people, like the *East European Scientific Journal* (Beall, 2016). Spammers, whether selling weight-loss pills and cadging credit card numbers or requesting scholarly papers in marine biology, always need to minimize the effort involved in producing content like email text or a site's landing page—the price of a business built on casting the widest possible net—and therefore reuse existing content to rapidly generate new venues for potential customers. The *American Journal of Pure and Applied*

Mathematics is an "international journal of high quality" with an astonishing range of interests (from statistics and probability to number theory, wavelets, and Banach spaces), which it shares, word for word, typos included, with the *Global Journal of Pure and Applied Mathematics*.[2]

All of these are classic spam strategies, and immediately explain how attempts to get junior scholars to submit papers to generic-sounding journals can be understood with the same word used for emails promising to restore virility. However, there is a deeper connection between these two things—between bottom-feeding ad scams and the *Global Journal of Pure and Applied Mathematics*. This is the second, subtler answer: both have to do with producing the appearance of salience.

Following close on the heels of the development of the first search engines—and in some cases driving their innovation—was the spammer's project of generating this appearance. A spammer wants to get a link or some content in front of potential victims, through whatever platform the victim employs. Therefore, the spammer needs to convince the evaluative metrics that are judging the importance of the information. The content cannot actually stand up to scrutiny on its own—and certainly not human scrutiny, no more than the gibberish papers, screeds, and obvious pseudoscience published by spam journals could hold up for an actual scholarly audience. To appear to satisfy whatever criterion is being analyzed—relevance or utility or popularity—the spammer needs to generate populations of "users" whose linking and citation behavior looks, from the right distance and by the right metric, like a popular endorsement.

This is actually a problem that originates with scholarly and legal citation networks (the two groups of people most obsessed with citation are academics and professional spammers) because those networks were the inspiration for Google's ranking algorithm, which determines the salience of a web page for a given query. "The PageRank Citation Ranking: Bringing Order to the Web," the paper that outlined the system at the core of Google's process, is about adapting the citation-analysis model from one platform to the other, though "there are a number of significant differences between web pages and academic publications" (Page et al., 1999). Given our subject in this book, the most important difference they observe offers us a painful example of historical irony: "Unlike academic papers which are scrupulously reviewed, web pages proliferate free of quality control or publishing costs … [H]uge numbers of pages can be created easily, artificially inflating citation counts." We know that this paper is important, because "The PageRank Citation Ranking" is itself

one of the most cited papers in computer science; a similar process can help us determine that one site is more significant than another (combined with many other subassemblies of algorithms and analytic tools).

The PageRank system, developed in part to counteract the first generations of lexical spam built to fool keyword-based search engines, set off a still-ongoing arms race. Automating the academic process—treating links as cites for determining the really important papers, as it were—worked well, but new platforms enabled the rapid production of venues (web pages, wikis, blogs) that could produce huge volumes of citations: linkfarms of hosted blogs, spam entries added to hacked wikis, automated comments, and so on, up to contemporary platforms like Twitter and "likefarming" on Facebook. Of course, the preventative strategies have likewise evolved, taking approaches like analyzing the social network graph as a whole, rather than the content of individual posts, to spot bad actors. The telling characteristic of early linkfarms and likefarms was that they were densely interlinked but lonely: that is, they would all link to each other but nobody outside their group would ever link to them.[3] That made them relatively easy to identify and eliminate. Spammer strategies evolved in response: spam Twitter accounts have a mix of fake and real followers, for instance, many of them in deliberately small numbers. The methods have changed, but the goal remains the same—making people, and influencing "friends," in the eyes of the machines.

That last detail is a crucial one, and brings us back to predatory journals and other new forms of academic misconduct. Spam Twitter accounts, blogs, and web pages are seldom meant to be seen by human beings; their speciousness is obvious to even the casual reader. They exist to influence metrics—follower counts, likes and retweets, views of videos, listens of songs, bumping up search engine rankings—with influence on people as a secondary or tertiary effect. The use of "private blog networks" (PBNs) in the search engine spamming community, networks of huge numbers of automatically generated blogs that exist to link to other sites rather than be encountered by humans, is the equivalent of a dodgy academic paper deliberately published in a spam journal.[4] The paper was never meant to be *read;* it exists to appear as a line in a CV, one among many, to act as a token in an assessment process.

Here, we can see the outline of the first lesson from spam for studying these new forms of misconduct: to pay attention to the *secondary markets.* When we think of spam we tend to think of the specific examples we happen to encounter—an email, a Twitter account, a blog stitched together from hundreds of cut-up public domain sources—but those are only the

outlying, human-facing parts of the business, whereas whole secondary economies and marketplaces exist to build the tools the spammers need, marketplaces that make deeper trends visible. To get rich in a gold rush, you don't go prospecting: you sell shovels, assays, work pants. Spammers rely on many kinds of back-end infrastructure, on payment systems and accomplice banks—and, like academics, on metrics companies and impact measuring tools. Quick! Is this promise from a spam journal company, or a spammy search engine manipulation company: "Citation Flow is based on stronger, iterative mathematical logical than the old metric of ACRank."[5] (Odd sentence structure in original.) What about this one? "The Impact Factor is calculated by several scientific methods including citation anlaysis [*sic*]. No Evaluation Processing Fees."[6] As Michael Power argues elsewhere in this book, the business of analytics and metrics starts to drive the product to be measured, rather than the other way around: the outcome becomes the target. Spammers and academics alike will look at what has impact— what can deliver good metrics—and tailor their spam campaigns and search engine optimization, or research projects, accordingly.

The proliferation of predatory journals, like the proliferation of spam email, sites, and accounts, includes the development of secondary markets and facilitation tools: new kinds of misleading or fraudulent metrics and document and object identifiers—the heralds of an entire parallel scholarly apparatus, a crooked ancillary economy.[7] Keeping a close eye on those will reveal, as it has with spam, deeper trends in this domain beyond any individual misconduct. It is on that point that the second lesson from spam rests.

In the same way that we tend to think of spam as "this email"—and of spam publishing as "this fraudulent journal"—when there is much more to see in the secondary layers of infrastructure and facilitation, we tend to think of spam in terms of straightforward and singular victims. Of course there are people victimized by phishing messages, identity theft, and bogus products, but aside from those cases, the question of victims becomes more complex. Take spam Twitter accounts, which are simultaneously a misuse of Twitter but also somewhat to Twitter's benefit. A spambot population artificially inflates their user numbers and metrics of activity and gives their users, both by accident and by design, similarly inflated follower counts. People who thought that they were wildly popular on some legitimate grounds, or wanted to be seen so, are suddenly revealed in their networked insignificance during a legitimate cleanup effort, often to their outrage.[8] In these moments of exposure—when both deliberately and accidentally pumped-up follower counts are exposed, we can see that social networks

have undergone a process akin to the shift of scientific papers from units of knowledge to "accounting units" described by Yves Gingras in this book. What began—at least notionally—as relationships between people has become an accounting unit for measuring significance through retweets, responses, likes, and "followers," in ways that benefit, as well as harm, the platforms and the victims alike. This is only one of the confusing *larger networks of benefit at play* in spam. This is not to say that spam is actually a good thing, but simply that we do not necessarily have an accurate picture of who benefits and who the victims are, which can explain some puzzles about how it works.

Spam journals and "predatory" publishing systems have more complex answers to the *cui bono?* question than we may at first assume. Who is preyed on by the predatory journal? Sometimes a junior academic who doesn't realize that they're being taken advantage of. But every scholar wants another line on their CV, and so does their department and their administration—which perhaps gets rewarded by institutional assessment tools—and even their country, seeking to reward "performance" without having a nuanced picture of what precisely performance is, for any given discipline, and how it should be measured. Spam journals make it possible for chronically overworked adjunct faculty to keep up a brisk publishing pace, for people without significant academic resources at their home university to rack up impressive records of "international" activity. For, let us say, the equivalent of $90 US? A fair price for services rendered, perhaps: no one has to read the resulting paper, no library needs to subscribe, and impact is automatically generated by a more-or-less imaginary system.

Additionally, as spam helped to put "legitimate" advertising into a better light on the old, resolutely noncommercial internet (we may be sending you ads, but at least we're not like the diet pills, porn, and malware crew), spam journals indirectly valorize the seriousness and status of mostly developed-world, Global North journals—the kinds of venues in which *respectable* scholars publish. De Rijcke and Stöckelová put this best in their chapter: "predatory publishing and its concomitant practices are not outside of the research system but emerge at the heart of them and are embedded within them," reinforcing the distance between the "'international' West or North on the one hand, and a 'parochial' East or South on the other." The top-tier journals and the major universities already constitute a kind of *de facto* citation cartel, colluding in a shared economy of cultural, social, and scientific capital with their status further reinforced by all these incompetent, ersatz rip-offs: the genuine article, accept no

substitutes! (Indeed, Marie-Andrée Jacob argues in the following chapter that counterfeits and originals are inextricably linked: "counterfeiting solidifies the 'template' of elite science and keeps it intact.")

This is not an exercise in devil's advocacy, to defend these pseudoscience-publishing, author-exploiting, corrupt, open-access-trashing journals, but a chance to think about who gets something from their existence, and what contributes, directly and indirectly, to the environment in which they thrive. Of what larger networks of value are they a part? What are the secondary markets and ancillary products that profit from them? As spam explains some otherwise enigmatic developments in the history of the internet, spam journals give us a bleak, oblique portrait of how academia is being measured and evaluated now—and what it is in danger of becoming.

Notes

1. Jeffrey Beall maintained an extensive list of various predatory, problematic, and suspect scholarly publishing projects. He took the site down in early 2017; snapshots of the site are on the Internet Archive. I draw a number of my examples of spam journals from his research and will cite him with Internet Archive links. For instance, a list of these journal phishing projects—which Beall called "hijacked journals"—can be found at https://web.archive.org/web/20160310203111/https://scholarlyoa.com/other-pages/hijacked-journals/. For a specific example, see the Mexican life sciences journal *Ludus Vitalis* (http://www.centrolombardo.edu.mx/ludus-vitalis/) and the phishing site for "Ludus Vitalis" (http://ludusvitalis.org.mx).

2. The identical text can be seen at http://www.academicresearchjournals.com/journal-detail.php?journals_id=37 and http://www.ripublication.com/gjpam.htm. Whether there is a relationship between "Academic Research Journals (India)" and "Research India Publications," their respective publishers, is left as an exercise for the reader.

3. There has been extensive research on this phenomenon; two exemplary papers are Dennis Fetterly, Mark Manasse, and Marc Najork, "Spam, Damn Spam, and Statistics," *Proceedings of the 7th WebDB Workshop,* 2004, and Zoltan Gyöngyi and Hector Garcia-Molina, "Link Spam Alliances," *Proceedings of the 31st VLDB Conference,* 2005.

4. For the interested reader, there are many thorough guides to the particulars of how PBNs are built and operated in the form of various get-started manuals, generally as invitations to pay for services. See, for instance, http://authoritywebsiteincome.com/build-private-blog-network/.

5. "Citation Flow," Majestic Marketing Search Engine Glossary, https://majestic.com/support/glossary#CitationFlow.

6. "Impact Factor," CiteFactor Academic Scientific Journals, http://www.citefactor.org/page/impact-factor.

7. These are likewise drawn from Beall's work. See "Misleading Metrics," Scholarly Open Access, https://web.archive.org/web/20160303171335 /https://scholarlyoa.com/other-pages/misleading-metrics/.

8. For some broader analysis of the scale of the problem, see Networked Insights, "How Dirty Is Big Data?," March 2015, http://info.networkedinsights .com/Dirty-Data-LP.html. For a very entertaining look at the peculiar structure of Twitter spambot populations (including the data itself), see Terence Eden, "This Is What a Graph of 8,000 Fake Twitter Accounts Looks Like," March 9, 2015, https://shkspr.mobi/blog/2015/03/this-is-what-a-graph-of-8000-fake -twitter-accounts-looks-like/. For a journalistic summary of experiments in buying and detecting spam followers on Twitter, see Daniela Hernandez, "Why Can't Twitter Kill Its Bots?," Fusion, September 21, 2015, http://fusion.net/story /195901/twitter-bots-spam-detection/.

References

Beall, Jeffrey. 2016. "Bogus Polish Journal Has Completely Fake Editorial Board." Scholarly Open Access (blog), February 2, 2016. https://web.archive.org/web /20160213180338/https://scholarlyoa.com/2016/02/02/bogus-polish-journal -has-completely-fake-editorial-board/).

Brunton, Finn. 2013. *Spam: A Shadow History of the Internet.* Cambridge, MA: MIT Press.

Page, Lawrence, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. "The PageRank Citation Ranking: Bringing Order to the Web." Stanford InfoLab Technical Report, 1999.