Shadow Libraries and You:
Sci-Hub Usage and the Future of ILL

*Gabriel J. Gardner, Stephen R. McLaughlin, and Andrew D. Asher**

Scholars have shared copyrighted material outside official channels for decades, but recent market forces, professional pressure, and novel technical exploits have allowed the large-scale automation of these practices. As a result, millions of scholarly articles, chapters, books, and papers have been aggregated in illicit collections known as "shadow libraries,"[1] available for free download by anyone with an Internet connection. The best-known shadow libraries are Library Genesis (LibGen) and the associated website Sci-Hub, which uses pooled university credentials to access articles and add them to LibGen's repository. AvaxHome and the defunct website Library.nu are two others, with numerous smaller collections scattered across the web.

Since Elsevier, in June 2015, filed a civil suit against Sci-Hub, its creator Alexandra Elbakyan, LibGen, and several John Does associated with LibGen,[2] many articles and opinion pieces have been published on the role and possible effects of shadow libraries in the scholarly communication ecosystem. Though they clearly violate U.S. copyright law, many scholars are quick to offer praise rather than condemnation for the site. Sci-Hub is easy to use, offering simple and fast access to full-text articles in PDF format in a manner more straightforward than many library or publisher websites. To date, however, only a few scholars have examined the phenomenon empirically. There is a dearth of information on questions such as whether shadow libraries will reduce publishers' revenues and disrupt the scholarly publishing industry, and whether they are affecting academic library usage or decreasing the use of interlibrary loan.

This study attempts to shed light on some of the questions facing libraries and publishers. Specifically, it examines the effects of Sci-Hub on interlibrary loan in select cities using various quantitative methods. In addition, we explore Sci-Hub usage in the United States qualitatively along subject-based and pricing-based lines of inquiry. Other general qualitative findings of note are included in an effort to expand the literature on this relatively new and understudied topic.

Literature Review

Scholars have shared copyrighted material through decentralized means for many decades. Shadow libraries, which centralize the process, are a relatively recent development in scholarly sharing. Balázs Bodó offers an up-to-date, concise history of shadow libraries and the larger guerilla open access movement.[3] He describes the market forces and technological improvements which have allowed for the present flowering of these groups and websites, attributing them to both a concurrent increase in demand for scholarly texts in the historically developing world and with a failure of major commercial publishers to price their services appropriately.[4] On the technological front, increased access to global communication networks was a necessary, but not sufficient, cause. Large online libraries of pirated scholarly texts emerged far later than collections of music, film, and television content, due to both the time-consuming nature of scanning long texts,[5] the supposedly niche demand

* *Gabriel J. Gardner is Reference & Instructional Librarian, California State University Long Beach, gabriel.gardner@csulb.edu; Stephen R. McLaughlin is PhD Student, University of Texas at Austin School of Information, stephen.mclaughlin@utexas.edu; Andrew D. Asher is Assessment Librarian, Indiana University Bloomington, asherand@indiana.edu*

for such material, and an alleged lack of political motivation among scientists.[6] The web platform Sci-Hub, which works in concert with LibGen, has recently emerged as a well-known and frequently used face of shadow libraries and the guerilla open access movement.

Opinions about the current role of Sci-Hub in the scholarly communication ecosystem, as well as its possible future effects, vary widely. Will it destabilize the scholarly publishing industry? Will publishers' revenues be affected? Is it affecting academic library usage? Or decreasing the use of interlibrary loan? Its illegality (at least in the United States) is clear, yet many are quick to offer praise rather than condemnation for the site. How should libraries and librarians approach and come to terms with it? Relatively few empirical studies have been conducted to date, so the answers to these questions are generally speculative.

Shadow libraries in general, and Sci-Hub in particular, are poised to undermine both the scholarly publishing industry and academic libraries. Many sources attest to Sci-Hub's usability.[7–9] Designed for immediate article delivery to an end user, it offers "simpler, more straightforward, … faster, [and] more streamlined access" to content than traditional channels.[10] This ease of delivery, coupled with the vast number of articles already stored by LibGen (more than 61 million)[11] offers challenges and opportunities for publishers and libraries.

A clear difficulty for publishers is potential revenue lost to shadow libraries. How much revenue has been lost? In Elsevier's legal complaint against Sci-Hub et al., they allege a loss of "at least $5,000" over a one-year period,[12] the threshold for claiming violation of the Computer Fraud and Abuse Act.[13] Marianne Parkhill, considering only the 100 most downloaded articles in the Bohannon dataset and using an estimate of $29.95 per item,[14] calculated a potential revenue loss (distributed across various publishers) of $2.7 million.[15] Since the articles held by shadow libraries need only be downloaded from a publisher once, it is worth noting that Sci-Hub interferes with publishers' attempts to track usage metrics.[16]

In addition to the basic revenue pressures, shadow libraries are engaged in an ongoing arms race against paywalls. Accessible guides such as *The Open Access Guerilla Cookbook* detail how organized groups of individuals can outsmart the security systems put in place to protect copyrighted scholarly information, systems which are no doubt familiar to librarians.[17] Such efforts require publishers to continually adjust and improve their web platforms to guard against hacking attempts. As countermeasures against ongoing theft of intellectual property via the vehicle of the PDF document, Angela Cochran and Joseph Esposito (both working in the publishing industry) have even proposed ending PDF availability altogether.[18,19] Shadow libraries are forcing publishers to evolve, to ramp up their security, and perhaps improve the delivery services they offer.

The dearth of evidence about Sci-Hub—a function of both its relatively recent arrival on the scene and its obvious illegality—has allowed such divergent opinions. To date, only a few scholars have examined it empirically. The most fine-grained was an analysis of web server traffic, the result of collaboration between John Bohannon, a reporter for *Science*, and Sci-Hub's Elbakyan.[20] The English-language commentary on Sci-Hub often takes an implicit (occasionally explicit) first-world view. Yet multiple sources attest to the fact that most Sci-Hub usage occurs outside the Anglosphere. In *Science*, Bohannon noted that during the six months of available data, the city with the most usage was Tehran.[21] According to Alexa Internet analytics, the top five country sources of traffic to Sci-Hub were China, Iran, India, Brazil, and Japan.[22] These countries account for 56.4% of recent traffic and make Sci-Hub the 3,950th most popular website in the world, up from a rank of 10,529 in 2015.[23]

The anonymized data from Bohannon's work was released publicly, and several others have already used it to conduct preliminary analysis of Sci-Hub's specific effects. Bastian Greshake found that most downloads take place during working hours, 9 a.m. to 5 p.m., Mondays through Fridays, with less traffic occurring over the weekend, though weekend downloads still follow the same general pattern of working hours.[24] On its face, this evidence seems to challenge hypotheses that would assume Sci-Hub to be a "last resort" method of the time-

pressed scholar working late or over the weekend (though it is possible that "last resort" downloads might follow a predictable pattern as well, so these hypotheses cannot be ruled out *a priori*). The naïve expectation that downloads per country would positively correlate with population holds up, though with significant outliers; the top five countries on a measure of downloads per 1,000 inhabitants were Portugal, Iran, Tunisia, Greece, and Chile.[25] Gross Domestic Product per capita was also positively correlated with downloads, more so than the other measures Greshake tested.[26] Using two different datasets of public IP ranges associated with universities, Greshake estimated the volume of downloads occurring on campuses. Calculations using the first set of IP ranges showed that on any given weekday, about 8% to 10% of Sci-Hub downloads occurred from a university IP range. The second set of IP ranges was grouped at the country level and showed that for most countries, fewer than 20% of all downloads occurred at universities.[27]

Looking specifically at Latin American countries, and relying on Elbakyan and Bohannon's public data, Juan Machin-Mastromatteo, Alejandro Uribe-Tirado, and Maria Romero-Ortiz found that the top five downloaded publishers were Elsevier, Springer, Wiley Blackwell, Nature Publishing Group, and the American Chemical Society.[28] These results closely mirror the general global patterns found by Bohannon, suggesting that the market demand filled by Sci-Hub in Latin America is similar to the rest of the world. Shifting from a specific geography to a specific subject area, Natalia Timus and Zakaria Babutsidze used the public dataset to identify where downloads of European Studies research took place, finding that most usage came from nine European countries, followed by Brazil, Turkey, China, and the United States.[29]

Are scholars using Sci-Hub as a complement to their library database access, or as a substitute? Little empirical evidence is available. One case study by Bianca Kramer examined The Netherlands in general, and Utrecht University in particular. With methods relying on publicly known university IP addresses, similar to Greshake, Kramer estimated that 4% of all downloads in The Netherlands occurred at universities, and that cities with universities accounted for 45% of Sci-Hub usage.[30] Looking just at the city of Utrecht, Kramer found that 60% of Sci-Hub downloads were available via Utrecht University Library's subscriptions, 15% of downloads were Gold Open Access or otherwise free from the publisher's website, and 25% were not immediately accessible.[31] (A portion of this 25% would presumably be obtainable via interlibrary loan, but her analysis did not take that into consideration.) For Utrecht in particular, the number of downloads from university IPs was conservatively estimated at 9%. This case study suggests that some Sci-Hub downloads are made purely on a convenience basis, because obtaining a PDF from a shadow library is much easier than using the legal access methods offered by Utrecht University.

A web survey by John Travis, with a sample of over ten thousand respondents, supports the claim that convenience motivations for Sci-Hub usage play a substantial role. 16.9% of respondents said the primary reason they used "Sci-Hub or other pirated article repositories" was convenience.[32,33] In addition, 36.9% responded 'Yes' to the question: Have you obtained a pirated journal article, through Sci-Hub or other means, despite having access to it in some manner via a university library or institutional subscription?[34] Overall, however, actual (or perceived) lack of full-text access to the desired texts is the primary reason people report for using Sci-Hub, as well as crowdsourcing methods such as the #icanhazPDF Twitter hashtag and message boards such as Reddit r/Scholar.[35,36]

For individuals associated with university libraries, research-intensive firms, or hospitals that provide institutional subscriptions, shadow libraries might be used as an alternative to their normal access privileges. The research discussed above supports the assertion that a subset of Sci-Hub users and article-sharing community members obtain materials using these methods rather than availing themselves of interlibrary loan. We examine the extent of such use below. However, the current state of interlibrary loan must be considered on its own to better understand how shadow libraries may be affecting it.

Many sources attest to the fact that interlibrary loan (ILL) usage has declined in recent years. Interlibrary loan systems and mechanisms are complex in general and often have local peculiarities. But often the decline is attributed to the general combination of factors surrounding the ever-increasing amount of content that is available online in full-text form (including preprints and postprints) and the licensing agreements that libraries engage in to bring wider amounts of content to their patrons.[37] This hypothesis has been supported by a recent case study out of the UK, which showed a large reduction in article requests over ten years that happened to coincide with a substantial change in collection development towards consortial purchasing of online journals at the institutions studied.[38] A survey of 18 institutions covering the years 2002–2007 also found a general decline in article requests coincident with expanding full-text online access to content over that time period.[39]

What other causes might be behind the decline? Another, more specific, cause of general ILL usage decline is the adoption of single search web-scale discovery services by many libraries. Such unified search systems (e.g., EBSCO Discovery Service and Summon) purportedly offer a streamlined and simplified research process to end users. A case study examining Penn State University libraries found that introduction of the Summon discovery service reduced interlibrary loan requests. Requests for items held or licensed by the institution (requests placed by user error) dropped by 57% over a two-year period.[40] However, ILL borrowing requests (requests placed correctly) also declined by 22% during the study period, despite a lack of significant change in the user population. In particular, borrowing requests for articles placed by undergraduates decreased by 71% over the study period, suggesting that for that subpopulation, discovery services may reduce usage of ILL by putting more relevant full-text content within reach.[41] Supporting the general point that discovery services reduce (accurate) interlibrary loan requests is another case study from Western Carolina University, which found a 15% drop in article requests post-adoption of the EBSCO Discovery Service.[42]

The growth of open access publishing is yet another possible factor. The number of documents freely available in full text has increased tremendously in recent years, a development that many assume will reduce demand for ILL.[43,44] Decisive evidence for this assumption is still lacking, however. Recently, Mak and Baich looked for any impacts that open access publications might be having on demand for interlibrary loan. They were not able to find significant effects on demand for recent articles in the sciences, social sciences, or humanities, despite the recent growth of open access publications in those areas.[45] In closing, Mak and Baich note that academic file sharing practices have proliferated since 2005, which could also depress ILL article requests. We test that hypothesized relationship in this article.

To better understand Sci-Hub's effect on the scholarly communications system in the United States, we used a range of techniques to explore the available data by research topic and location, extending the scope of previous analyses.

Methods

Assembling Log Data

The Sci-Hub logs published by Elbakyan and Bohannon include 28 million worldwide download requests from September 2015 through February 2016, with an 18-day gap in November due to a server configuration error.[46] Each row includes a DOI for an article or chapter, along with anonymized data on the request's origin: country and city names, approximate geographic coordinates, and a unique hash value for each IP address. We first narrowed the set to 1.15 million downloads originating in the United States, adding state names drawn from gazetteer data in the GeoNames database.[47] We then retrieved metadata for 98.5% of article/chapter requests using the Crossref API,[48] and thereby expanded our dataset to include publisher, title, publication date, publication format, and research topic (where available).

Comparing Sci-Hub to Licit Download Numbers

While Sci-Hub is particularly popular in non-Western countries, in the broader context it appears to be a marginal phenomenon. In order to see this picture, we compared Sci-Hub downloads to publicly disclosed publisher or host platform download counts in Table 1. Using the list of the top 10 most downloaded publishers from John Bohannon's recent *Science* article, we looked for the most recent download counts available. We were not able to locate self-disclosed figures for most of the publishers, but did obtain download estimates from Elsevier (for calendar year 2015),[49] IOP Publishing (for calendar year 2014),[50] JSTOR (for calendar year 2012),[51] and the Royal Society of Chemistry (for calendar year 2015).[52] The publisher download counts were reported on an annual basis, but only 6 months of data was available from Sci-Hub. In order to determine the ratio of publisher-to-Sci-Hub downloads, we halved the publisher count to obtain a 6-month estimate. Ratios of publisher-to-Sci-Hub downloads were then calculated and rounded to the nearest whole number; they are shown in the second-to-last column of Table 1. With all Sci-Hub log data falling during the academic year, when download volume is likely at its highest, this should be considered a very rough measure.

How much does all of this illicit downloading cost publishers in lost revenue? We estimated the upper bound of revenue lost over 6 months using à la carte prices for recently published articles. Elsevier charges $35.95; IOP Publishing charges $33.00; JSTOR charges $19.50; and RSC charges £39.75 ($49.60). This upper-bound estimate assumes the very unlikely case that each Sci-Hub download would have been purchased at the going price rather than obtained via subscription or through a library; results are displayed in the last column of Table 1.

Comparing Sci-Hub to Institutional Interlibrary Loan Numbers

To compare ILL usage with Sci-Hub downloads, we focused on institutions that were the largest educational entities in their municipality. Sample entities were purposely chosen due to the fact that they were located in cities with non-zero Sci-Hub downloads between September 2015 and February 2016. A further criterion was that each municipality does *not* contain several large colleges or universities. Boston, Massachusetts, which has several colleges, a highly developed knowledge economy, and which borders several other municipalities with major universities, would be too complex to disentangle. So-called "college towns" with only one dominant university, which may be geographically isolated from other locations with similar characteristics, were taken as ideal cases that would allow us to detect any relationships between Sci-Hub downloads and interlibrary loan requests. (The economic characteristics of a municipality, i.e., the presence of firms that might use Sci-Hub for business research and development purposes, were not controlled for in this analysis.) We contacted dozens of institutions in cities meeting the above criteria and asked if they would be willing to share interlibrary loan data. Ten universities in 10 municipalities agreed to participate.

- City A is in Indiana. Indiana had 4,503 Sci-Hub downloads, below the average per-state value of 22,568 and the median state value of 5,562. City A had 437 downloads, which was above the national city average of 259 and the national city median value of 8. University A is a public institution located in City A.
- City B is in Illinois. Illinois had 17,510 Sci-Hub downloads, below the average state value but above the median. City B had 1,718 downloads, above both the national city average and median values. University B is a private institution located in City B.
- City C is in Minnesota. Minnesota had 7,280 Sci-Hub downloads, below the average state value but above the median. City C had 174 downloads, below the national city average but above the median value. University C is a public institution located in City C.
- City D is in California. California had 341,809 Sci-Hub downloads, above both the average and me-

dian state values. City D had 287 downloads, above both the national city average and median values. University D is a public institution located in City D.

- City E is in Ontario. It had 1,718 Sci-Hub downloads and accounted for 1.1 percent of all Canadian usage. (We did not calculate provincial or city statistics for Canada.) University E is a public institution located in City E.
- City F is in Manitoba. It had 1,383 Sci-Hub downloads and accounted for 0.9 percent of all Canadian usage. University F is a public institution located in City F.
- City G is in Kansas. Kansas had 31,950 Sci-Hub downloads, above both the average and median state values. However, the Kansas state value must be treated with extreme skepticism due to the fact that during the data collection period, IP locations which could not be reliably determined were assigned the default value of the geographical center of the country, which is in Kansas.[53] City G had 847 downloads, above both the national city average and median values. University G is a public institution located in City G.
- City H is in California. It had 31 downloads, below the national city average but above the median. University H is a public institution located in City H.
- City I is in California. It had 246 downloads, below the national city average but above the median. University I is a public institution located in City I.
- City J is in Texas. Texas had 41,193 Sci-Hub downloads, above both the average and median state values. City J had 5,573 downloads, above both the national city average and median values. University J is a public institution located in City J.

Data on the number of interlibrary loan requests for articles, book chapters, conference papers, and dissertations (i.e., for all requests that were not full books or media) were obtained from each university's relevant department from September 2015 through February 2016 (the same period contained in the Elbakyan and Bohannon dataset).[54] These specific material types were chosen because they could all plausibly be obtained by Sci-Hub if a digital copy with a unique identifier were extant via one of the many compromised accounts that Sci-Hub uses to obtain documents via library subscriptions. (It is much less likely that entire books would be obtainable via these library subscriptions, and Sci-Hub does not currently obtain and deliver audiovisual files.) Data were collected in an aggregated form to avoid the possibility of finding a direct link between any given ILL request from an institution and a Sci-Hub download in that same city. (The authors neither requested nor received individualized request data.) These ILL data were used to test the hypothesis that Sci-Hub usage is reducing interlibrary loan rates by comparing the month-to-month activity in each city. Our working hypothesis is that if many students and researches were substituting Sci-Hub for ILL, then Sci-Hub downloads would be negatively correlated with interlibrary loan requests, i.e., interlibrary loan requests would fall as Sci-Hub downloads rise.

These data were initially collected with the intention of measuring effects using a repeated measures analysis of variance (rANOVA). The rANOVA test requires three assumptions to be met for results to be valid: independent observations, dependent variables that are normally distributed, and equal variance for all possible difference values between within-subject conditions (an assumption known as sphericity). However, upon examination, the data performed poorly on Mauchly's Test of Sphericity (Mauchly's W=.019) requiring a correction to be applied to the degrees of freedom in order to avoid a false positive finding. Perfect sphericity would result in Mauchly's W=1, the sample population value of .019 indicated a severe violation of the sphericity assumption and thus required the most conservative (i.e., cautious) correction, the lower-bound estimate, to be applied. Results are presented in Table 2. To further examine the data, a multivariate analysis of variance (MANOVA) procedure was used; MANOVA does not require that a sphericity assumption be met in order for results to be

valid. The MANOVA was used to determine if there were any differences between monthly Sci-Hub downloads and ILL requests for each city. Monthly counts (downloads and requests, reported separately) were treated as continuous dependent variables while Sci-Hub or ILL usage was treated as a binary independent variable. Results are presented in Table 2.

Separately, correlations for each city between the month to month ILL request totals and Sci-Hub downloads were calculated; again, on the hypothesis that if scholars were using Sci-Hub as a substitute for ILL then Sci-Hub usage would increase with a simultaneous fall in ILL requests. Results are presented in Table 3 and Figure 1.

Comparing Sci-Hub to Geographic Interlibrary Loan Numbers

Sci-Hub downloads and library usage data were aggregated at the city level using anonymized IP ID and geolocation data contained in the Elbakyan and Bohannon dataset and library and university descriptive data obtained from ACRL Metrics. Sci-Hub usage was examined qualitatively for the top-downloading cities and IP addresses to identify potential patterns in the characteristics of Sci-Hub users and subject areas being accessed.

Correlations were calculated between cities' SciHub downloads, characteristics of university populations (number of institutions, number of faculty, number of students) and university library use (ILL articles borrowed, full text downloads from library databases). Unfortunately, the latest available ACRL data for fiscal year 2014–2015 does not overlap the available Sci-Hub data, which makes a direct evaluation of the hypothesis, that increased Sci-Hub use decreases use of university libraries, impossible at this time.[55] For this reason, we use these results to describe expected usage patterns in U.S. cities and to identify locations that deviate from these patterns.

Qualitative Characteristics of Sci-Hub Usage in the United States

Does either the price of an institutional subscription or discoverability via abstracting and indexing databases make a periodical more or less likely to be downloaded via Sci-Hub? To determine the answers, we relied on data from Ulrichsweb Global Serials Directory. Using the Elbakyan and Bohannon data, we identified 44,068 unique titles accessed from the United States, 17,789 of them (40%) were classified by CrossRef as either 'journal' or 'journal-article'. Books, proceedings, monographic series, etc. were excluded. A random sample of 270 titles was selected for analysis, a sufficient size to allow a 90% confidence level with a 5% margin of error. We queried each title manually on Ulrichsweb.com and recorded price data and the number of abstracting and indexing databases in which it is listed. Ulrich's supplies price quotes for many geographical areas; the pricing data that would be applicable to offer online access to each periodical title in the United States was recorded in United States Dollars (USD). (Certain titles only offered price quotes in foreign currencies. For those ten cases, the quoted currency prices were converted to USD at current exchange rates.) Indexing databases were assumed to be a proxy measure for a journal's general discoverability. The number of indexing databases was determined by counting each uniquely titled database, regardless of which commercial firm operated it or provided access. After the data were collected, we examined their scatter plots. Periodical title download counts were then tested for correlation with price data and the number of indexing databases. Correlations were calculated for Pearson's r (linear relationships) and Spearman's ρ (monotonic relationships). Their statistical significance was calculated with two-tailed tests. Results are shown in Table 9.

Findings

Comparing Sci-Hub to Licit Download Numbers

Elsevier is the most illicitly downloaded publisher, while the Royal Society of Chemistry is proportionally the

publisher most likely affected by Sci-Hub due to its high illicit-to-licit download ratio (20:1). It is important to note here that the Sci-Hub downloads don't count toward the publisher download totals. A paper need only be obtained from the publisher once in order for it to be served endlessly to Sci-Hub users. Various sources attest that Sci-Hub queries LibGen's article repository before attempting to obtain a copy using institutional credentials; in the latter case, the newly accessed document is cached in LibGen.[56,57]

TABLE 1
Sci-Hub Compared with Licit Downloads & Maximum Revenue Loss (Estimates)

Publisher or Host	Estimated Annual Downloads	Sci-Hub Downloads* (6 months)	Ratio of Licit to Illicit Downloads (Rounded to Nearest Integer)	Upper Bound Estimate of Revenue Loss (6 months)
Royal Society of Chemistry	37,500,000	927,238	20:1	£36.86 million
Elsevier	900,000,000	9,296,485	48:1	$334.21 million
IOP Publishing	22,000,000	160,073	69:1	$5.28 million
JSTOR	73,503,021	358,786	102:1	$6.99 million

*Bohannon Sci-Hub download numbers were doubled to create annual estimate before calculating ratio. Ratios show here rounded to nearest integer for the licit figures.

Comparing Sci-Hub to Institutional Interlibrary Loan Numbers

None of the sample cities examined produced a statistically significant negative correlation between interlibrary loan demand and Sci-Hub downloads. On the individual level of City/University pairs, for the level of aggregation that the data required, the substitution hypothesis failed. Looking beyond individual City/University pairs at all the month-by-month data for the entire sample, we failed to find strong effects of Sci-Hub downloading on interlibrary loan demand using both a repeated measures analysis (rANOVA) and a multivariate analysis of variance (MANOVA). F-test values compare variance between the distributions of City/University pairs, i.e., they are the ratio of between-group variability to within-group variability. The sample data's violation of the sphericity assumption required the use of the conservative lower-bound correction. That necessary adjustment lowered the observed power of the analysis and resulted in a low F-test statistic (3.596) which failed to achieve statistical significance (p=.074). The MANOVA test similarly found an F-test statistic (3.799) which, while statistically significant (p=.021) and of sufficient power (β=.827), could not support a rejection of the null hypothesis. Below we report the F-statistic calculated using the Pillai's Trace Test, which is regarded as the most robust of the tests for small samples included in a MANOVA. Looking at the combined monthly interlibrary loan demand and Sci-Hub downloads for the 10 sample Cities reveals that they both increase through January and February of 2016 while remaining relatively flat during September and October; ILL usage falls in December while the low point for Sci-Hub usage was in November. The correlation between the two aggregated series was not significant (r=-.011, p=.984).

TABLE 2
Statistical Effects for between Interlibrary Loan Demand and Sci-Hub Usage for 10 Sample Cities

Procedure	Effect	Value	F-Statistic	Statistical Significance	Degrees of Freedom	Observed Power*
rANOVA	Lower-bound	1534785.30†	3.596	.074	1	.435
MANOVA	Pillai's Trace	.637	3.799	.021	6	.827

Note: * Computed using alpha = .05; †Type III Sum of Squares.

TABLE 3
Correlations between Interlibrary Loan Demand and Sci-Hub Usage for 10 Sample Cities

Monthly City Sci-Hub Downloads / Monthly University ILL Requests	Pearson's Correlation Coefficient	Statistical Significance (2-tailed)
City A / University A	–.146	.783
City B / University B	.252	.630
City C / University C	–.646	.166
City D / University D	.151	.775
City E / University E	.317	.541
City F / University F	–.049	.962
City G / University G	.170	.748
City H / University H	.327	.526
City I / University I	–.343	.505
City J / University J	.190	.719
Total Cities / Total Universities	–.011	.984

Comparing Sci-Hub to Geographic Interlibrary Loan Numbers

The geographic distribution of Sci-Hub use in the United States is broad, but very unevenly distributed. Of the 3,611 unique cities represented in our dataset, the top 50 cities account for 75% of Sci-Hub article downloads, and the top 10 cities account for nearly 50%. Ashburn, Virginia, the top downloading city, recorded 96,857 downloads, while 2789 cities (77% of the total) had fewer than 50 downloads. A similar "long tail" distribution is observed examining Sci-Hub downloads by IP address. Of the 87,376 unique IP addresses in the dataset, the top 50 account for 31% of downloads, while the top address alone accounts for 6% of the total. At the IP address level, usage is marginal; 75,636 IP addresses recorded 10 downloads or less. Distributions of Sci-Hub downloads and ILL borrowing can be seen in Figures 2 and 3.

FIGURE 1
Interlibrary Loan & Sci-Hub Activity in 10 Sample Cities

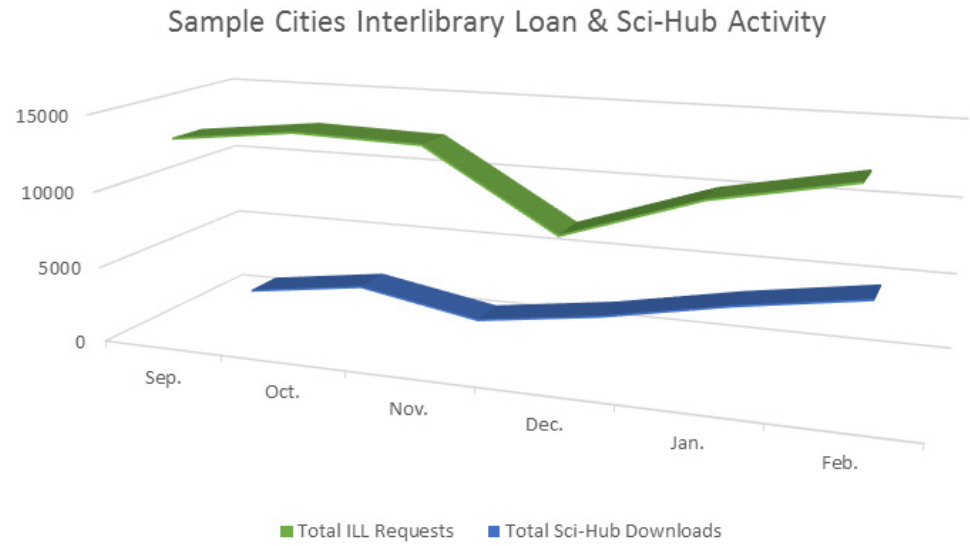


FIGURE 2
Distribution of US Sci-Hub Downloads

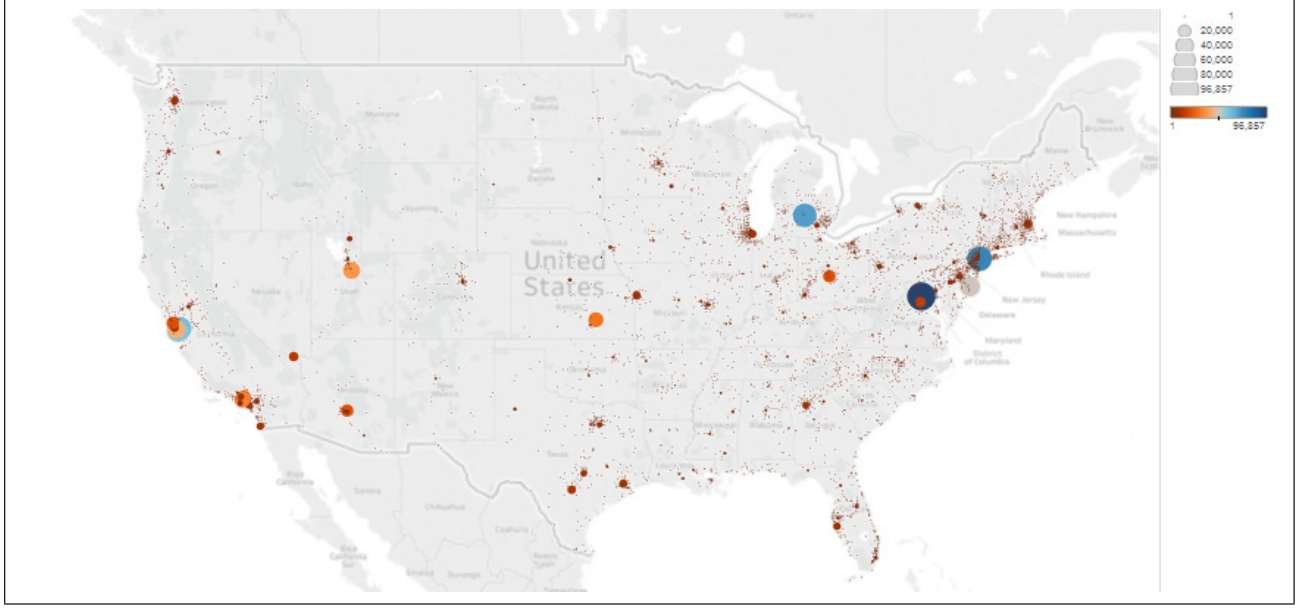
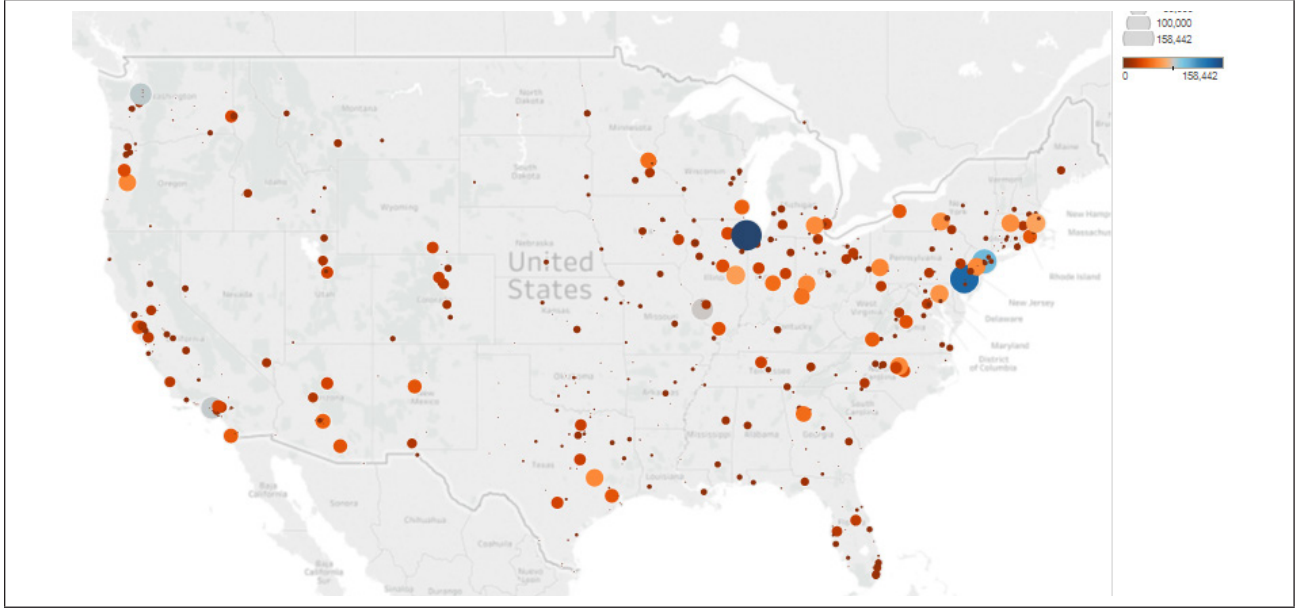


FIGURE 3
Distribution of ILL Borrowing (2015)



Although the vagaries of IP address assignment make it difficult to infer too much from these patterns, they do suggest that Sci-Hub usage in the United States is clustered around a small number of "power user" IP addresses. However, these IP addresses might represent individual users (e.g., a researcher systematically collecting literature on a subject), or an aggregated group of users (e.g., a shared business or institutional IP address, a virtual private network (VPN) server, or a Tor exit node). The presence of an Equinix data center[58] and at least one major VPN service[59] in Ashburn, Virginia suggest that much of this city's traffic may originate in other locations. Examining the usage characteristics in the Sci-Hub download data, we attempted to evaluate which type

of usage an IP address represents. If an IP address represents an individual computer, we would expect articles to be clustered around a small number of subject areas, while an IP address used by many people would likely contain a large number of subject areas. Similarly, the IP address of an individual systematically downloading articles is likely to show patterns of automation in the timestamps of downloads, while an IP address with many users would exhibit a more organic pattern. The results of these evaluations are summarized in Table 4.

TABLE 4
Characteristics of IP Addresses with Most Sci-Hub Downloads

Download Rank	Number of Downloads	% of Total Sci-Hub Use	Median Time between Downloads (s)	Timestamp-based Evaluation	Subject-based Evaluation
1	69,264	6.02	1	Automated	Not Automated
2	41,770	3.63	2	Automated	Automated
3	32,150	2.79	1	Automated	Automated
4	17,906	1.56	42	Not Automated	Not Automated
5	17,784	1.55	43	Not Automated	Not Automated
6	17,723	1.55	42	Not Automated	Not Automated
7	17,495	1.52	43	Not Automated	Not Automated
8	12,520	1.09	13	Not Automated	Automated
9	11,957	1.04	172	Not Automated	Not Automated
10	9,863	0.86	122	Not Automated	Not Automated

For example, the IP address with the most downloads accessed Sci-Hub only on Mondays and always beginning at 11 a.m. GMT, strongly suggesting an automated process. Although articles in 177 subjects were downloaded by this IP address, which is broader coverage than we might expect from an individual user, they are primarily clustered in medicine-related fields. The second most-used IP address exhibits the characteristics of automated downloading even more clearly, with 96% its downloads in Physical and Theoretical Chemistry or Chemistry. These two IP addresses also have median times between downloads of 1 and 2 seconds respectively—much faster than would be expected by human researchers—as well as high average numbers of downloads per day of activity. Together these two measures appear to be a useful rule of thumb for identifying potential automated processes. Of the top 10 IP addresses, four show the characteristics of automation and together account for 13.5% of the Sci-Hub downloads we analyzed.

When evaluated at the city level using Pearson's *r*, Sci-Hub downloads are moderately and positively correlated with the total number of academic institutions (.454, $p < .01$), total library expenditures (.403, $p < .01$), total ILL articles borrowed (.315, $p < .01$), full time instructional faculty (.463, $p < .01$), and full-time students (graduate and undergraduate) (.479, $p < .01$). By comparison, total ILL articles borrowed are more strongly correlated with each of these variables, while full-text article downloads from library databases are more weakly correlated (See Table 5). This suggests that all types of academic article uses are possibly interrelated and result from the same underlying causes (such as the number of faculty and students in a particular city), and that it is unlikely that Sci-Hub use decreases ILL use.

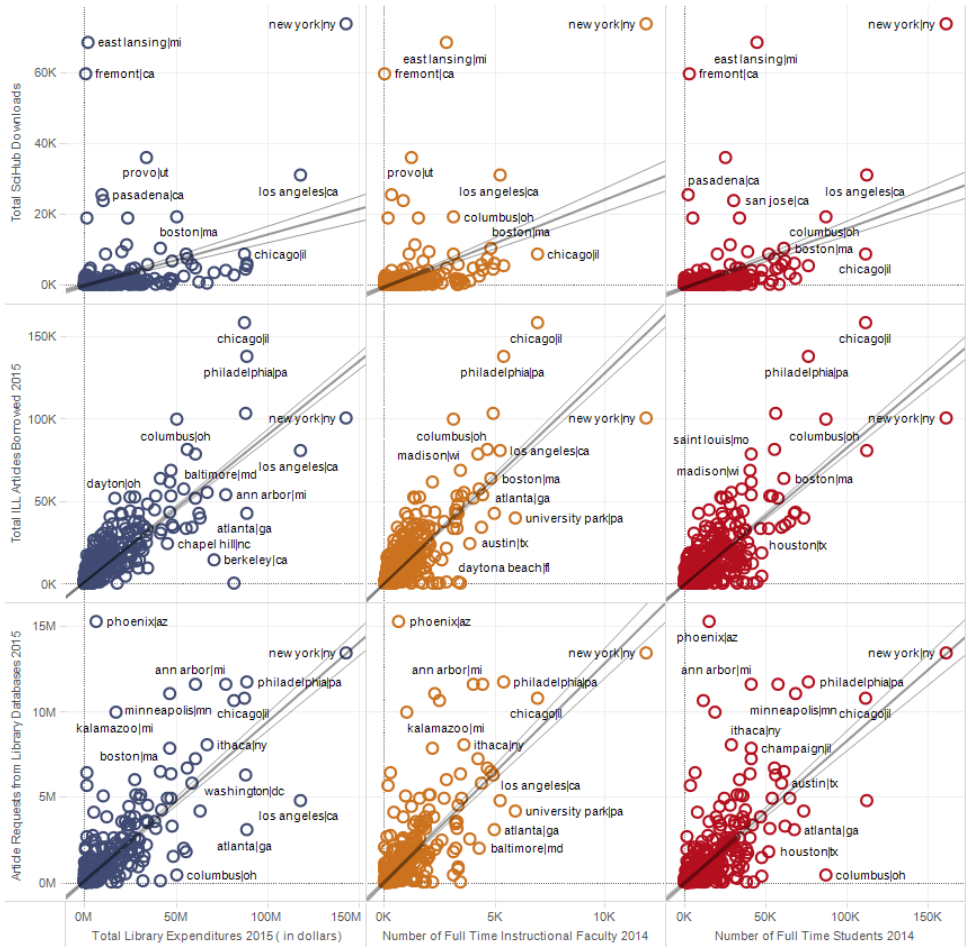
However, as is illustrated in Figure 4, some cities, such as East Lansing, Michigan and New York City exhibit a higher Sci-Hub use than predicted by their population of faculty or students, coupled with lower-than-expected ILL use, while other cities, such as Philadelphia and Chicago, exhibit the opposite pattern. These are the expected observations if Sci-Hub is being used as a substitute for ILL, and indicates that this might be occurring in select locations.

TABLE 5
Correlations between Sci-Hub Downloads, Library Usage, and University Data

	Sci-Hub Downloads	Number of Academic Institutions	Total Library Expenditures 2015	Total ILL Articles Borrowed 2015	Full-text Article Downloads from Library Databases	Full-time instructional Faculty	Full-time Students (Undergraduate & Graduate)
Sci-Hub Downloads		.454**	.403**	.315**	.143**	.463**	.479**
Total ILL Articles Borrowed 2015	.315**	.659**	.849**		.314**	.823**	.801**
Full-text Article Downloads from Library Databases	.143**	.251**	.359**	.314**		.341**	.333**

** Significant at p <.01

FIGURE 4
City-Level Correlations for Sci-Hub Downloads, ILL Borrowing, and Full Text Article Requests with Selected University Data



Of the 1.13 million U.S. Sci-Hub downloads with metadata available from Crossref, 970,783 (86%) were tagged with one or more journal-level topic category drawn from the SciVerse Scopus All Science Journals Classification (ASJC) hierarchy.[60] We applied topic codes to roughly 22,000 additional requests using ISSNs listed in the Thomson Reuters Essential Science Indicators (ESI) journal database,[61] followed by 6,000 requests for works from single-topic journals and publishers. These topic codes were then resolved to 13 broad research areas, following a procedure outlined in the University of California Libraries' "Pay it Forward" report.[62] In all, 91.4% of article/chapter requests were assigned Pay it Forward (PIF) topic codes. Texts coded as Clinical Medicine comprised 25.2% of U.S. downloads, followed in descending order by Engineering, Biomedical Research Disciplines, and Chemistry. The distribution of topics for United States Sci-Hub downloads can be seen in Figure 5. Table 6 displays the 10 most-downloaded papers in the United States. Interestingly, one was part of the National Bureau of Economic Research (NBER) Working Papers series, and was freely available in full-text form via the NBER website and discoverable via Google Scholar and the Social Science Research Network (SSRN). Only 2% of downloads came from publications listed in the Directory of Open Access Journals[63] or the Directory of Open Access Books.[64] Table 7 displays the publishers which were most frequently downloaded in the United States. The results here mirror the larger picture painted by topic metadata from Crossref; usage is high for publishers that produce medical, biomedical, chemistry, and engineering content.

FIGURE 5
Topic Areas for US Sci-Hub Downloads

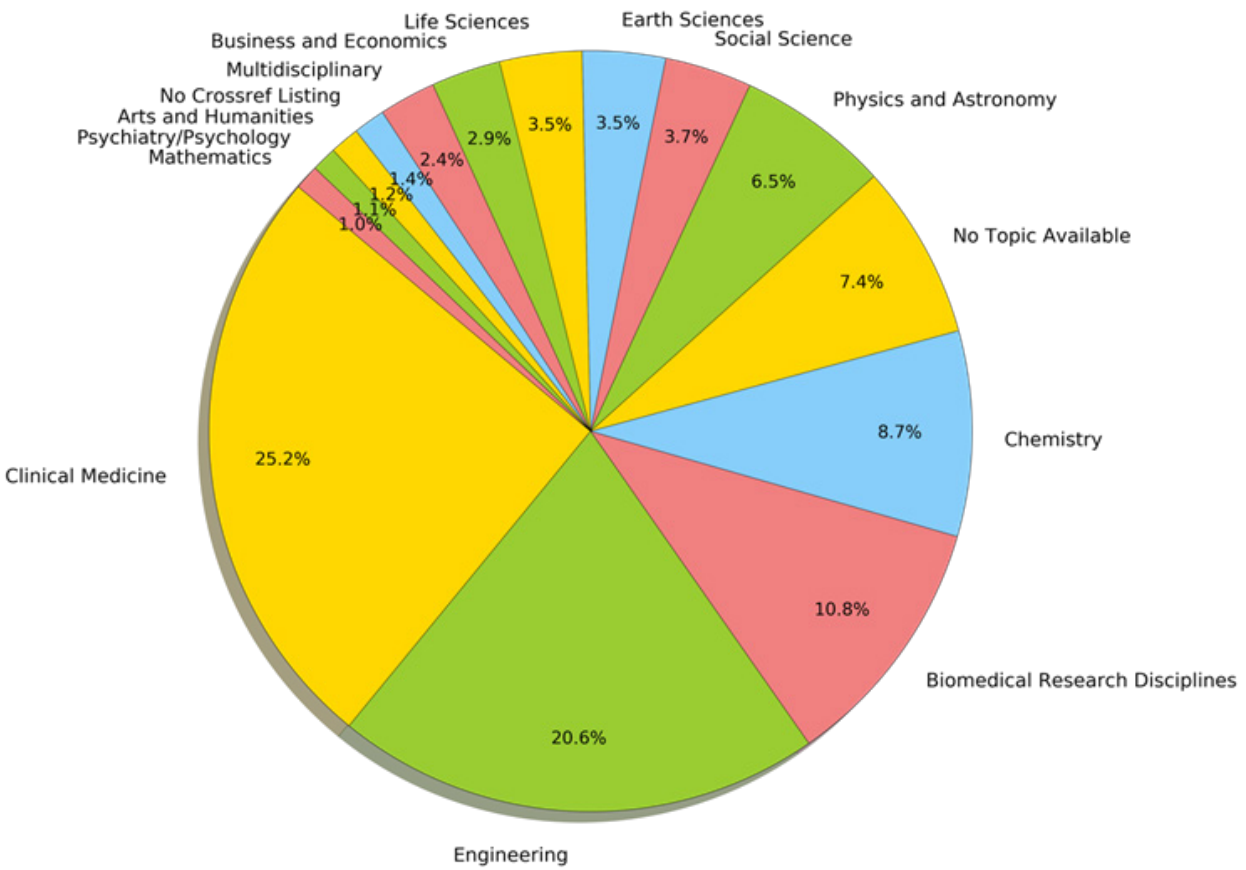


TABLE 6
United States Top 10 Most Downloaded Papers

DOI	Year	Title	Journal	Publisher	Downloads
10.1111/j.1365-277x.2004.00520.x	2004	Intentional mis-reporting of food consumption and its relationship with body mass index and psychological scores in women	Journal of Human Nutrition and Dietetics	Wiley-Blackwell	1,407
10.1109/ispass.2015.7095803	2015	Nyami: a synthesizable GPU architectural model for general-purpose and graphics-specific workloads	2015 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS)	Institute of Electrical & Electronics Engineers (IEEE)	638
10.1007/s10677-015-9567-7	2015	The Possibility of an Ongoing Moral Catastrophe	Ethical Theory and Moral Practice	Springer	417
10.1586/17512433.2015.1012494	2015	How statistical deception created the appearance that statins are safe and effective in primary and secondary prevention of cardiovascular disease	Expert Review Of Clinical Pharmacology	Informa Healthcare	391
10.1007/s11948-014-9521-4	2015	Penetrating the Omerta of Predatory Publishing: The Romanian Connection	Science and Engineering Ethics	Springer	340
10.3386/w18992	2013	Subjective well-being and income: Is there any evidence of satiation?	NBER Working Paper Series	National Bureau of Economic Research	331
10.1073/pnas.1510461112	2015	Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade	PNAS	Proceedings of the National Academy of Sciences	295
10.1162/jocn_a_00880	2016	The Role of Dopamine in Temporal Uncertainty	Journal of Cognitive Neuroscience	MIT Press	271

TABLE 6
United States Top 10 Most Downloaded Papers

DOI	Year	Title	Journal	Publisher	Downloads
10.1137/0217022	1988	How to Construct Pseudorandom Permutations from Pseudorandom Functions	SIAM Journal on Computing	Society for Industrial & Applied Mathematics (SIAM)	269
10.1177/0363546514556737	2015	Anterior Cruciate Ligament and Knee Injury Prevention Programs for Soccer Players: A Systematic Review and Meta-analysis	The American Journal of Sports Medicine	SAGE Publications	267

TABLE 7
United States Sci-Hub Downloads per Publisher

Publisher	Downloads	Percent
Elsevier	381,849	33.2%
Springer	92,779	8.1%
American Chemical Society (ACS)	91,520	8.0%
Wiley-Blackwell	85,662	7.4%
Institute of Electrical & Electronics Engineers (IEEE)	83,054	7.2%
Nature Publishing Group	53,009	4.6%
Informa UK Limited	48,388	4.2%
SAGE Publishing	23,604	2.1%
AIP Publishing	19,687	1.7%
Oxford University Press (OUP)	19,667	1.7%
Royal Society of Chemistry (RCS)	18,758	1.6%
American Society of Civil Engineers (ASCE)	13,055	1.1%
American Association for the Advancement of Science (AAAS)	9,860	0.9%
BMJ	9,214	0.8%
American Medical Association	7,351	0.6%
Other publishers combined	193,506	16.8%

The subscription price of journals is not a monocausal explanation for Sci-Hub usage in the United States. Using the methods described above to analyze price data for 270 of the 17,789 unique periodicals titles downloaded (keeping in mind that not all unique titles were for periodicals; many book chapters and conference proceedings were downloaded via Sci-Hub), we found that the price of an institutional subscription was correlated with download count with a Pearson's r value of .247 and a Spearman's ρ of .602. The Spearman value, which tests for a monotonic relationship rather than the linear Pearson correlation, suggests that a power-law relationship may be in effect. The significance of these correlations was $p < .01$.

We similarly analyzed the discoverability and availability of journals through abstracting and indexing services, finding that inclusion in these databases was not a determining factor influencing usage. The Pearson's *r* correlation of a publication's Sci-Hub download count and the number of databases that abstracted and/or indexed the publication was .231; Spearman's ρ was .420. These correlations were significant at $p < .01$. The fifteen most downloaded journals and their pricing, abstracting, and indexing data are displayed in Table 8. Correlations between download count and price, as well as indexing data are reported in Table 9.

TABLE 8
United States Top 15 Most Downloaded Journals

Journal	Downloads	Percent	Price*† (USD)	Abstracting & Indexing Databases count*
The Journal of Physical Chemistry B	36,569	3.18	$ 8680	52
IEEE Journal of Quantum Electronics	17,409	1.51	$ 2325	71
The Journal of Chemical Physics	15,032	1.31	—‡	45
Nature	12,561	1.09	$ 6008	294
Science	8,792	0.76	$ 99	314
Journal of Chemical Information and Modeling	7,148	0.62	$ 1045	45
Journal of Chemical Theory and Computation	6,493	0.56	$ 1640	9
Journal of the American Chemical Society	5,566	0.48	$ 5320	80
Journal of Engineering Mechanics	5,432	0.47	$ 2630	74
IEEE Transactions on Plasma Science	5,231	0.45	$ 1900	79
New England Journal of Medicine	5,171	0.45	$ 189‡	191
Journal of Environmental Engineering	4,631	0.40	$ 2502	146
Proceedings of the National Academy of Sciences	4,367	0.38	$ 235‡	187
The Lancet	3,535	0.31	$ 2227	146
Angewandte Chemie International Edition	3,293	0.29	$ 17381	107

Notes:
* Price and number of abstracting & indexing databases data obtained via Ulrichsweb™ Global Serials Directory.
† All price information applies to domestic institutional subscriptions, unless otherwise noted.
‡ Pricing for *New England Journal of Medicine* and *Proceedings of the National Academy of Sciences* is shown above, no institutional price was quoted in Ulrichweb™. Pricing for *The Journal of Chemical Physics* is determined on a sliding scale based on number of users, no institutional price was quoted in Ulrichweb™.

TABLE 9
Correlations between Title Usage, Price, and Discoverability for the United States

Usage Measure	Correlation Measure	Price (USD)*	Indexing Databases
Periodical Title Download Frequency (*n*=270)†	Pearson's *r*	.247**	.231**
	Spearman's ρ	.602**	.420**

Notes:
* 10 prices quotes were not available in USD but were converted to a USD value at current exchange rates.
† Sample size of 270 (randomly determined) allowed for conclusions at a 90% confidence level with a margin of error of 5%.
** Significant at $p < .01$

Limitations

One considerable limitation in the tests conducted above is that they are based on a limited dataset which may not reflect current patterns of Sci-Hub usage. The Elbakyan and Bohannon data is very rich in metadata but was obviously collected before John Bohannon's article was published in *Science*. The *Science* article, and accompanying editorial and survey prompted a huge surge of interest in Sci-Hub.[65] Such an increase in attention almost certainly broadened the user base and possibly shifted usage patterns. Conversely, since the recent spike in interest in Sci-Hub, Elbakyan has added a CAPTCHA program which has likely had the effect of reducing the automated downloading we noted above. Such automated downloading was a non-trivial part of the traffic contained in the Elbakyan and Bohannon data; with the CAPTCHA making such activity marginally more difficult than it was during the data collection period. It is therefore possible that usage may have decreased. In short, we urge extreme caution in generalizing about what current Sci-Hub usage patterns may be based on the data made available from September 2015 through February 2016.

A further limitation is that we used aggregated data in our analyses of interlibrary loan traffic. This limitation was deliberate; we did not want to use individual ILL requests due the fact that such an analysis, combined with the individualized Sci-Hub download logs could prove that an individual placed an ILL request but later obtained the same document via Sci-Hub. That type of data could have opened up the libraries that collaborated with us by sharing data to litigation. Working at a higher level of aggregation means that the above analysis is provisional. Furthermore, though many attempts were made to solicit ideal City/University pairs for this study, the final sample of 10 was small compared with the set of all "college towns." Regression analysis, which was based off of the most recent ILL data available, but which did not overlap with the Sci-Hub data logs, showed that there may be some cities where Sci-Hub is substituting for ILL. The same analysis should be redone with ILL data that goes through February 2016 before placing much faith on that finding. Future research on this topic, particularly if its aim is to test some of the same hypotheses we considered above, will need to use more varied sources of data and methods (e.g., up-to-date university traffic logs, more granular ILL data, perhaps even more recent Sci-Hub traffic logs) to determine Sci-Hub's effects. At the same time, future researchers will need to take the utmost care to properly anonymize their data, or even avoid collecting data which could be subpoenaed by publishers or their trade groups (e.g., The Association of American Publishers). We focused on Sci-Hub in this paper, but similar analyses might also be undertaken to look at the effects that other shadow libraries such as LibGen and AvaxHome, or crowdsourcing methods like #icanhazPDF, are having on libraries.

Discussion

Research on shadow libraries, and Sci-Hub in particular, is still in its infancy. The tests conducted above represent a few initial attempts to flesh out our understanding of these sites. Given the fact that a large portion of the activities under study are illegal, with many site operators (Elbakyan excepted) maintaining strict anonymity,[66] shadow libraries are difficult to study. Much of the data needed to demonstrate their impacts on users, libraries, and the scholarly communications ecosystem has not been collected. Of the data that has been collected so far, much of it will need to be updated over time due to changes in user behavior, library discovery systems, and publisher content pricing and formatting.

Using the limited data we were able to obtain from cooperating university libraries, we failed to find an effect of Sci-Hub downloads on interlibrary loan demand in the United States and Canada. While we could not reject the null hypothesis, there may be reason to believe that Sci-Hub affects ILL use in some locations (see the top row in Figure 4) and for certain individuals who have adopted Sci-Hub as part of their everyday work. In ag-

gregate, however, we cannot find such effects on ILL. Regression analysis showed that Sci-Hub usage is positively correlated with many indicators of library usage, as well as indicators of student and instructional population. Yet outlier locations are driving much of the download volume observed. In short, use is wide but uneven; a few heavy users (apparently employing automated techniques) account for the lion's share of Sci-Hub downloads.

Moving away from the impact on libraries and interlibrary loan, it is clear that many publishers may, in the long term, be negatively affected by Sci-Hub and other shadow libraries. As Elsevier's lawsuit against Elbakyan and the creators of LibGen moves forward, and if other publishers choose to follow suit, more detailed information on these sites' commercial impact may come to light. Curiously, while it is clear from Table 8 that the most-downloaded journals are expensive (when priced à la carte), we were not able to find a strong linear correlation between usage and price overall. This counterintuitive result is best explained by the fact that our scatter plot analysis suggested a monotonically increasing relationship (as the price increases, usage increases also, though not in a linear fashion). Alternately, it may be the case that many users are unlikely to purchase à la carte articles at any price, with Sci-Hub simply displacing free alternatives such as crowdsourced sharing via social media. Even so, individual journal price is one among several factors which influence Sci-Hub download counts. Other factors remain unknown and untested, though our topical analysis suggests that disciplinary affiliation plays a major role.

One possible response to Sci-Hub and other shadow libraries would be for publishers to adjust their prices; perhaps the market will support a degree of experimentation in the coming years. Yet Sci-Hub may also force publishers to take more drastic, even draconian, steps in order to combat unauthorized sharing of their content. Session timeout limits may be shortened; two-factor authentication may become a standard requirement for database access; publishers may attempt to display articles in non-downloadable formats. The future is unwritten, but we believe these would be Pyrrhic victories; any such change may motivate even more scholars to adopt Sci-Hub in place of institutional channels.

The fact that title pricing does not wholly explain Sci-Hub usage suggests that pricing may not be the only way to solve the problem. Ease of access is another factor that must be considered; the Travis survey in *Science* indicates that a sizable minority of Sci-Hub users persist in their behavior because of convenience.[67] Simply put, Sci-Hub offers a superior user experience compared to the platforms currently offered by publishers and university libraries,—and unless they learn to compete on those terms, we predict their efforts to control the flow of academic knowledge will fail.

Acknowledgments

The authors would like to thank Stafford Cox for his statistical advice and IUB Libraries for providing access to ACRL statistical data. Carol Perruso and Marisa Soltz graciously provided helpful comments, technical editing, and proofreading.

Notes

1. Liang, Lawrence. "Shadow Libraries." *E-Flux*, no. 37 (2012). http://www.e-flux.com/journal/shadow-libraries/.
2. Devore & DeMarco LLP. 2015. *Complaint*. United States District Court for the Southern District of New York. http://www.plainsite.org/dockets/download.html?id=212950598&z=76180c35.
3. Bodó, Balázs. "Pirates in the Library – An Inquiry into the Guerilla Open Access Movement." Glasgow, UK: Social Science Research Network, 2016. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2816925.
4. Ibid., 4–5, 8.
5. Magee, C. Max. "Confessions of a Book Pirate." *The Millions*, January 25, 2010. http://www.themillions.com/2010/01/confessions-of-a-book-pirate.html.
6. Bodó, "Pirates in the Library – An Inquiry into the Guerilla Open Access Movement," 9.
7. Smith, David. "Sci-Hub: How Does It Work?" *The Scholarly Kitchen*, February 25, 2016. https://scholarlykitchen.sspnet.org/2016/02/25/sci-hub-how-does-it-work/.

8. Coldewey, Devin. "Sci-Hub Is Providing Science Publishers with Their Napster Moment." *TechCrunch*, April 29, 2016. http://social.techcrunch.com/2016/04/29/sci-hub-is-providing-science-publishers-with-their-napster-moment/.
9. Russon, Mary-Ann. "Sci-Hub: 'Pirate Bay for Scientists' Now Available on Anonymous Telegram Messaging App." *International Business Times UK*, May 17, 2016. http://www.ibtimes.co.uk/sci-hub-pirate-bay-scientists-now-available-anonymous-telegram-messaging-app-1560576.
10. Heathers, James. "Why Sci-Hub Will Win." *Medium*, May 2, 2016. https://medium.com/@jamesheathers/why-sci-hub-will-win-595b53aae9fa.
11. Library Genesis. Accessed January 31, 2017. http://libgen.io/scimag/repository_torrent_notforall/.
12. Elsevier Inc., Elsevier B.V., Elsevier Ltd. v. Sci-Hub d/b/a WWW.SCI-HUB.ORG, The Library Genesis Project d/b/a LIBGEN.ORG, Alexandra Elbakyan, John Does 1–99, No. 1:15-cv-04282-RWS (D. S. N.Y. filed June 3, 2015).
13. 18 U.S.C.A. § 1030 (West, Westlaw current through Pub. L. No. 114–254).
14. Elbakyan, Alexandra, and John Bohannon. "Data from: Who Uses Sci-Hub? Everyone." Dryad Digital Repository, 2016. https://doi.org/10.5061/dryad.q447c.
15. Parkhill, Marianne. "Sci-Hub: The Academic Cat Is Out of the Bag." *Blog—Plum Analytics*, May 16, 2016. http://plumanalytics.com/sci-hub-academic-cat-bag-post/.
16. For a simple explanation of how Sci-Hub in particular obtains and maintains their article copies, see: Banks, Marcus. "What Sci-Hub Is and Why It Matters." *American Libraries* 47, no. 6 (June 2016): 46–48. Also, see Smith, 2016.
17. Williwaw. "Open Access Guerilla Cookbook," January 13, 2012. http://archive.org/details/open.access.guerilla.cookbook.
18. Cochran, Angela. "A Funny Thing Happened on the Way to OA." *The Scholarly Kitchen*, February 25, 2016. https://scholarlykitchen.sspnet.org/2016/02/25/a-funny-thing-happened-on-the-way-to-oa/.
19. Esposito, Joseph. "Sci-Hub and the Four Horsemen of the Internet." *The Scholarly Kitchen*, March 2, 2016. https://scholarlykitchen.sspnet.org/2016/03/02/sci-hub-and-the-four-horsemen-of-the-internet/.
20. Bohannon, John. "Who's Downloading Pirated Papers? Everyone." *Science*, April 28, 2016. doi:10.1126/science.aaf5664.
21. Ibid.
22. Alexa Internet. http://www.alexa.com/siteinfo/sci-hub.cc. Accessed 20 Oct. 2016
23. Gardner, Gabriel. "The Guerilla Open Access Movement: What Do We Know?" Presentation presented at the UCR Open Access Week: Open in Action, University of California, Riverside, October 25, 2016. http://eprints.rclis.org/30147/.
24. Greshake, Bastian. "Correlating the Sci-Hub Data with World Bank Indicators and Identifying Academic Use." *The Winnower*, no. 3:e146485.57797 (June 2, 2016). doi:10.15200/winn.146485.57797.
25. Ibid.
26. Ibid. Note: numerical correlation coefficient values were not reported in this source.
27. Ibid.
28. Machin-Mastromatteo, J. D., A. Uribe-Tirado, and M. E. Romero-Ortiz. "Piracy of Scientific Papers in Latin America: An Analysis of Sci-Hub Usage Data." *Information Development* 32, no. 5 (November 1, 2016): 1806–14. doi:10.1177/0266666916671080.
29. Timus, Natalia, and Zakaria Babutsidze. "Pirating European Studies." *Journal of Contemporary European Research* 12, no. 3 (2016). http://www.jcer.net/index.php/jcer/article/view/760.
30. Kramer, Bianca. "Sci-Hub: Access or Convenience? A Utrecht Case Study (Part 1)." *I&M / I&O 2.0*, June 20, 2016. https://im-2punt0.wordpress.com/2016/06/20/sci-hub-utrecht-case-study-part-1/.
31. Kramer, Bianca. "Sci-Hub: Access or Convenience? A Utrecht Case Study (Part 2)." *I&M / I&O 2.0*, June 20, 2016. https://im-2punt0.wordpress.com/2016/06/20/sci-hub-access-or-convenience-a-utrecht-case-study-part-2/.
32. Travis, John. "In Survey, Most Give Thumbs-up to Pirated Papers." *Science*, May 6, 2016. doi:10.1126/science.aaf5704.
33. The Travis survey ran in the online version of *Science* to accompany Bohannon's feature article. This likely resulted in a highly biased sample, however, no other surveys of Sci-Hub users are available at time of publication to compare results.
34. Ibid.
35. Ibid.
36. Gardner, Carolyn Caffrey, and Gabriel J. Gardner. "Fast and Furious (at Publishers): The Motivations behind Crowdsourced Research Sharing." *College & Research Libraries* 78, no. 2 (February 1, 2017): 131–49. doi:10.5860/crl.78.2.131.
37. Posner, Beth. "The Impact of Global Trends on ILDS." *Interlending & Document Supply* 42, no. 4 (November 2014): 147–52. doi:10.1108/ILDS-10-2014-0048.
38. Glover, Steven William, and Sarah Louise Glover. "Interlending and Document Supply in the NHS: A North West Case Study." *Interlending & Document Supply* 44, no. 1 (January 2016): 27–30. doi:10.1108/ILDS-11-2015-0037.
39. Behr, Michele D., and Julie L. Hayward. "Do Off-Campus Students Still Use Document Delivery? Current Trends." *Journal of Library Administration* 48, no. 3–4 (October 2008): 277–93.
40. Musser, Linda R., and Barbara M. Coopey. "Impact of a Discovery System on Interlibrary Loan." *College & Research Libraries* 77, no. 5 (September 1, 2016): 643–53. doi:10.5860/crl.77.5.643.
41. Ibid., 650-651.
42. Calvert, Kristin. "Maximizing Academic Library Collections: Measuring Changes in Use Patterns Owing to EBSCO Discovery Service." *College & Research Libraries* 76, no. 1 (January 2015): 81–99. doi:10.5860/crl.76.1.81.
43. Mike McGrath. "Interlending and Document Supply: A Review of the Recent Literature; 91." *Interlending & Document Supply* 44,

no. 1 (January 29, 2016): 1–6. doi:10.1108/ILDS-12-2015-0041.
44. Mike McGrath. "Interlending and Document Supply: A Review of the Recent Literature; 92." *Interlending & Document Supply* 44, no. 2 (May 16, 2016): 88–91. doi:10.1108/ILDS-03-2016-0013.
45. Mak, Collette, and Tina Baich. "Looking for the Impact of Open Access on Interlibrary Loan." In *Connections. Collaboration. Community.* Columbus, OH: IFLA, 2016. http://library.ifla.org/1358/1/095-mak-en.pdf.
46. Elbakyan and Bohannon, "Data from: Who Uses Sci-Hub? Everyone."
47. "GeoNames Gazetteer Data." *GeoNames*, August 1, 2016. http://download.geonames.org/export/dump/US.zip.
48. "Crossref REST API." GitHub. Accessed August 31, 2016. https://github.com/CrossRef/rest-api-doc.
49. RELEX Group. "Annual Reports and Financial Statements 2015." RELEX Group, March 2016. http://www.relx.com/investorcentre/reports%202007/Documents/2015/relxgroup_ar_2015.pdf.
50. IOP Publishing. "About Us." *IOP Publishing*, n.d. http://ioppublishing.org/about-us/.
51. JSTOR. "Annual Summary." New York, NY, USA, March 19, 2013. https://about.jstor.org/sites/default/files/misc/JSTOR-Annual-Summary_2012_v6.pdf.
52. Royal Society of Chemistry. "Financial Statements and Trustees' Report 2015." Royal Society of Chemistry, 2016. http://www.rsc.org/globalassets/02-about-us/corporate-information/financial-statement-trustees-report-2015.pdf.
53. Hill, Kashmir. "How A Strange Internet Glitch Turned This Kansas Farm Into A Digital Hell." *Fusion*, April 10, 2016. http://fusion.net/story/287592/internet-mapping-glitch-kansas-farm/.
54. Elbakyan and Bohannon, "Data from: Who Uses Sci-Hub? Everyone."
55. The authors plan to rerun this analysis once the data becomes available.
56. Cabanac, Guillaume. "Bibliogifts in LibGen? A Study of a Text-Sharing Platform Driven by Biblioleaks and Crowdsourcing." *Journal of the Association for Information Science and Technology* 67, no. 4 (April 1, 2016): 874–84. doi:10.1002/asi.23445. Also, see Smith, 2016.
57. Woltermann, Anthony. 2015. Declaration of Anthony Woltermann in Support of Plaintiffs' Application for an Order Authorizing Alternate Service of Process on Defendants and for Preliminary Injunction. United States District Court for the Southern District of New York. https://www.plainsite.org/dockets/download.html?id=213268754.
58. Equinix, Inc. "Equinix DC4 Data Center Specifications." Accessed January 15, 2017. http://www.equinix.com/locations/united-states-colocation/washington-dc-data-centers/dc4/.
59. IPVanish. "IPVanish VPN Servers." Accessed January 15, 2017. https://www.ipvanish.com/servers/.
60. Elsevier. *SciVerse SCOPUS CUSTOM DATA DOCUMENTATION*. Version 4. Edited by Peter Berkvens. Last updated April, 2012. http://ebrp.elsevier.com/pdf/Scopus_Custom_Data_Documentation_v4.pdf.
61. "ESI Journal List." Thomson Reuters InCites, December 20, 2016. http://ipscience-help.thomsonreuters.com/incitesLive/8289-TRS.html.
62. "Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions." University of California Libraries, July 18, 2016.
63. Directory of Open Access Journals Metadata, January 31, 2017. https://doaj.org/csv.
64. Directory of Open Access Books Metadata, January 31, 2017. http://www.doabooks.org/doab?func=csv.
65. Google. "sic-hub – Explore." Google Trends. Accessed January 1, 2017. https://www.google.com/trends/explore?q=sci-hub&date=2015-01-01%202017-01-01#TIMESERIES.
66. Bodó, "Pirates in the Library – An Inquiry into the Guerilla Open Access Movement," 2.
67. Travis, "In Survey, Most Give Thumbs-up to Pirated Papers." Also, see Gardner, "The Guerilla Open Access Movement: What Do We Know?" for detailed cross-tab analysis of the underlying data.