The Human Side of Cyber Property Rights: Theory and Evidence from Github

Pengfei Zhang*

May 20th, 2021

Abstract

Open access is a defining feature of cyberspace that challenges the conventional wisdom on property rights. This paper presents a case where property rights may *lead to* the failure of the commons in cyberspace. We consider a game theoretic model of competing creations in which a freelance creator decides whether to choose open access in the shadow of a threat of litigation from the copyright holder, and competition between open access and its proprietary alternative allows a continuum of users to choose between voluntary contributions or royalty payment. When individuals are heterogeneous in social preference, the model exhibits two distinct equilibria: a reproduction equilibrium and an original contribution equilibrium. Copyright law can dramatically change the *set* of equilibria. Enforcing authorized use without considerable limitations on the owner's exclusive rights may erode the original contribution equilibrium completely. The key predictions of the model are then tested and supported by data from Github. A takedown notice has a persistent negative effect on subsequent sharing, and repositories shared by foreign users attract fewer contributions as their home countries improve upon software piracy prevention. Our findings caution against a secure property rights system in cyberspace.

Keywords: Open Access, Litigation, Digital Copyright, Social Preference, Multiplicity of Equilibria, Event Study, Natural Experiment

JEL classification: C72, D64, H44, K24, L17, O34

*Department of Economics, Cornell University, pz82@cornell.edu

1 Introduction

Secure property rights is critical to economic development (North and Weingast, 1989; Keefer and Knack, 1997; Besley and Ghatak, 2010). It is often proposed as a solution to the tragedy of the commons (Gordon, 1954; Demsetz, 1967; Merrill, 2002). The law enforces the right to exclude, which prevents congested use and provides incentives for productive economic activities. This paper, however, presents a case where the expansion and enforcement of property rights *lead to* failure of the commons.

The case we consider originates from an emerging yet increasingly important policy domain - cyberspace - the environment resulting from the interaction of people, goods and services on the Internet by means of information technology, which does not exist in any physical form. Cyberspace has a phenomenally remarkable commons. Open access, the practice of making creative works freely available online to anyone, is increasingly popularized and becoming a defining feature of cyberspace (Suber, 2004). Such successful projects include open source software, such as Linux and Mozilla, open educational resources such as the MIT OpenCourseWare and Coursera [1], open knowledge provision, such as Project Gutenberg [2] and Wikipedia, open access research and scientific collections such as the Public Library of Science (PLOS)[3], open media and news, such as citizen journalism, and open design [4]. The common theme of these projects is that they rely heavily on voluntary contributions of large group of internet users, or commons-based peer production as characterized by Benkler (2006), "a decentralized, collaborative, and nonproprietary process, based on sharing outputs among widely distributed, loosely connected individuals".

The growth of cyberspace, on the other hand, is subject to legal constraints. Authors of any digital content is entitled to its property rights, notably in the form of copyright. Copyright applies to any creative work of original authorship, fixed in a medium of expression; it provides protection for literary, scientific, and artistic works, giving their

[1] See a list of open education providers at https://en.wikipedia.org/wiki/List$_{o}f_{M}OOC_{p}roviders$

[2] one of the earliest project in which volunteers make out-of-copyright works available online

[3] See a list of open access journals at https://en.wikipedia.org/wiki/List$_{o}f_{o}pen - access_{j}ournals$

[4] the development of physical products, machines and systems through use of publicly shared design information.

creators the ability to control uses of their works. The question of whether and how copyright should be reframed in the digital age has become a prominent topic of discussion in policy and legal circles. Corporations and lobby groups are reported to demand stronger copyright protection, arguing that illegal copying is easier on the internet (Krim, 2003; Anderson, 2007; King, 2007). Some economists and legal scholars argue against a strong copyright regime, contending that the current copyright regime severely constrains reuse of the copyrighted and thus future creativity.

In this paper, we study carefully a conjecture critical to this debate that the existing literature has not given enough attention: a strong property rights regime may remove the freedom of creators to choose open access. The central thesis is that the law gives the copyright owner an unbalanced power against future creators. The owner can employ the threats of litigation to eliminate competition from open access and peer production under the name of justice. Thus, the copyright owner reaps pure rents from his exclusive property rights beyond the material incentive for creation that copyright law aims at. Such rent-seeking strategy is particularly problematic when in the online community, many innovative projects are open access and driven by the human side of incentive - social benefits. Restricting access to open access projects is a pure deadweight loss, and enforcing such restriction by litigation is even more costly to social welfare.

To understand this phenomenon, this paper builds a theoretical model based on important institutional details, and test the theory with data from Github. We consider a game theoretic model of competing creations in which a freelance creator decides whether to choose open access in the shadow of a threat of litigation from the copyright holder, and competition between open access and its proprietary alternative allows a continuum of users to choose between voluntary contributions or royalty payment. We argue that (i) if other-regarding preference is present, multiplicity of equilibria arises. One is a rent-seeking equilibrium where the copyright holder blocks open access and purely reaps profits from the credible threat of litigation, the other is a a original contribution equilibrium where peer production outcompetes potential monopoly production and open access drives out any commercial interests. (ii) given there are multiple equilibria, a strong copyright regime may have a perpetual adverse effect: it erodes the original contribution equilibrium completely, and forces the equilibrium to settle in the rent-seeking equilibrium.

We then provide empirical support to the key predictions of the model by studying program developers' open source behavior in Github, the largest online host of open source code and repositories in the world. We present two econometric specifications to separately test the effect of threats of litigation and stronger anti-circumvention enforcement on users' open access decision. In the first specification, we exploit the exogenous timing of the Digital Millennium Copyright Act (DMCA hereafter) takedown notice addressed to Github users. The DMCA takedown notice is a request to remove content of individual users whose shared repository is disputed by some claimed copyright holders. By comparing early receivers' post-notice change in the likelihood of sharing repositories with that of late receivers, our event study estimates provide direct evidence on how threats of litigation persistently distort the decision to choose open access. In the second specification, we exploit country-level variations in international anti-piracy enforcement. The software piracy rate is the proportion of unlicensed copies on the number of proprietary software installed in a given country, as reported by the Business Software Alliance. By documenting how the contributions to repositories differ by the owners' nationality and matching them with national piracy rate, we show that repositories shared by foreign users attract fewer contributions as their home countries improve upon software piracy prevention.

The rest of this paper proceeds as follows. Section 2 reviews the copyright law necessary for understanding property rights in cyberspace, and outlines some key cases of litigation against open access. Section 3 formally presents the model. Section 4 characterizes the set of equilibria, and derives the central theorem on how copyright law affects open access which forms the basis of the empirical hypothesis. Section 5 describes the Github data and empirical strategies. Section 6 presents the empirical results. Section 7 concludes.

2 Institutional Background

2.1 Legal Framework

Copyright law provides protection for literary, scientific, and artistic works, giving their creators the ability to control certain uses of their works. Copyright, in its current shape, is a bundle of two major rights: (i) the exclusive right to make copies and distribute them, (ii) the exclusive right for further derivative works.[5] The first and foremost international agreement governing copyright is 1886 Berne Convention, which requires automatic protection for all creative works in a fixed medium, and enforces a minimum duration of protection (at least 50 years after the author's death for any work). The Berne Convention set up a bureau to handle administrative tasks, which later evolved to be the World Intellectual Property Organization (WIPO).

In order to address the new challenges posed by digital technologies, in particular the dissemination of protected material over the Internet, WIPO introduced what they referred to as the "Internet treaties". The "Internet treaties" aimed at setting down international norms of preventing unauthorized access to and use of creative works on the Internet and any other digital networks.

One of the most prominent "Internet treaties" is the WIPO Copyright Treaty (WCT) effective in 2002. The WCT has a broad coverage, dealing with protection for authors of literary and artistic works, such as writings and computer programs, original databases, musical works, audiovisual works, and works of fine art and photographs. The WCT clarifies that existing rights continue to apply in the digital environment, and also creates new online rights. The WCT requires countries to provide a framework of basic rights, allowing creators to control and/or be compensated for the various ways in which their creations are used and enjoyed by others. Most importantly, the WCT ensures that the owners of those rights will continue to be adequately and effectively protected when their

[5] A "derivative work" is defined as a work based upon one more more preexisting works, such as translation, musical arrangement, dramatization, fictionalization, motion picture conversion, sound recording, art reproduction, abridgment, condensation, or any other form in which a work may be recast, transformed, or adapted.

works are disseminated through new technologies and communications systems such as the Internet.

Calling upon a concerted effort from all member nations to curb digital copyright infringement, the WCT attracts wide adherence with 109 signatory countries. The Treaty does not by itself grant rights, rather signatory countries implement the provisions of the WCT through national copyright and intellectual property rights legislation. In the U.S., such legislation includes the 1997 No Internet Theft (NET) Act, the 1998 Digital Millennium Copyright Act, and the 1999 Digital Theft Deterrence and Copyright Damages Improvement Act. The European Union enacts the 2001 Information Society Directive. China and India passes the 2010 Copyright Law Amendment and the 2012 Copyright Amendment Bill respectively. By design, the WCT allows reasonable flexibility in establishing exceptions or limitations to rights in the digital environment. Countries may, in appropriate circumstances, grant exceptions for uses deemed to be in the public interest, such as for non-profit educational and research purposes. In reality, countries do vary in their own discretion in the commitment to the Treaty and their ability to enforce the legal measures.

2.2 Open Access Movement

The notion of open access appeared before the digital age. With the spread of the Internet and the ability to copy and distribute electronic data at no cost, the idea of open access gained new importance. Open access projects thrived at a time when major problems arouse with the by-then dominant proprietary copyright model. Most commercial software vendors provide users only with object or binary code, which is difficult for programmers to interpret or modify.[6] At the same time, the university librarians around the world found themselves in the middle of a big problem known as the "serials crisis" - subscription costs for publications rising much faster than inflation for years. Libraries no longer had money for all of the publications they wanted and were forced to make

[6] Software can be transmitted in either source code or object code. Source code is the code using languages such as Basic, C and Java. Object/binary code is the sequence of 0s and 1s that directly communicates with the computer.

difficult choices between journals.

Both the scale and formalization of the open access and peer production have expanded dramatically with the widespread diffusion of the Internet in the early 1990s. Along with the rapid development of open access projects, there was also the proliferation of alternative approaches to licensing that supported cooperative peer production. The Free Software Foundation introduced a formal licensing procedure, called a General Public License, for a computer operating system called GNU. In keeping with the philosophy of the organization that the software should be free to use, free to modify and free to redistribute, the license aimed to preclude the assertion of copyright concerning cooperatively developed software.[7] Alternative approaches in the same spirit includes the Berkeley Software Distribution (BSD) license, the Creative Commons license, etc. These kind of licenses are under the umbrella called "copyleft", because if copyright seeks to keep intellectual property private, copyleft seeks to keep intellectual property free and available. In the last decades, as with the international rules on cyberspace, the development of open access movement also went global. Open source softwares have been developed in many parts of the world: Linux originated in Helsinki, and the Ruby was developed in Japan, Ubuntu hails from South Africa.

2.3 Litigation Against Open Access

The tension between open access and commercial interests was first exemplified by the UNIX litigation in the 1980s when ATT began enforcing its (purported) intellectual property rights related to the operating system software UNIX, to which many academics and corporate researchers at other firms had made contributions. Berkeley Software Distribution (BSD) was an open source operating system developed by Bill Joy and other researchers, based on the operating system Unix jointly invented by American Telephone & Telegraph Co. (AT&T) and UC Berkeley. In 1991 AT&T sued the University of Cali-

[7] In exchange for being able to modify and distribute the GNU software, software developers had to agree to a) make the source code freely available (or at a nominal cost) to whomever the program is distributed; and b) insist that others who use the source code agree to do likewise. Furthermore, all enhancements to the code—and even in many cases code that intermingled the cooperatively developed software with that developed separately— had to be licensed on the same terms.

fornia at Berkeley for infringing copyrights on its Unix software, which the two parties ultimately settled. The settlement required that anybody who uses the university's Unix software also must pay royalties to the company. The key creator of BSD Bill left Berkeley and started his own operating system Sun, which was later purchased by AT&T.

In 1997, MP3.com was launched by Mike Robertson to facilitate sharing of new music, and it was soon the Internet home to many independent musicians. To recommend new artists, MP3.com enabled the users to insert their favorite CDs into the computer and discover contents of similar kinds. Five major labels headed by the Recording Industry of American Association (RIAA) filed a lawsuit against MP3.com. Controversially, MP3.com is found guilty of willful infringement. A year later, one of the plaintiff - Vivendi Universal - purchased MP3.com. The same strategy animates the RIAA's suits against individual users. The RIAA has sued more than 20,000 people suspected of distributing copyrighted works, and settled approximately 2,500 of the cases (Electronic Frontier Foundation, 2007).

3 Theory: Model

Consider an institutional copyright owner (denoted by player O), a freelance creator (denoted by player A), and a continuum of users of measure n. The creator has a creative idea and initiates a project that she makes available to any online users. The project is indivisible - every user can consume at most one project, where contents of different amount $x \geqslant 0$ can be produced. [8] The owner begins with developing a proprietary project of output $\hat{x}$ at a cost $c\hat{x}$, whose content is distributed only for profits. [9] The open access project and the proprietary project are competing creations; their contents are substitute goods.[10]

An open access project relies upon user generated content, and its output is deter-

[8]Such a project can be a source code repository (e.g., Berkeley Software Distribution in Section 2), an online music archive (e.g., MP3.com in Section 2), a wiki webpage, etc.

[9]We assume that the variable cost of distributing a copy is zero. This assumption is reasonable under digital technology, see Goldfarb and Tucker (2019).

[10]In contrast to patents, a copyright does not grant exclusive rights to an idea, but merely to the specific expression of an idea. Hence, in spite of the fact that the price of copyrighted works is greater than marginal cost, a copyright by itself does not create monopoly power.

mined by the sum of all users' contributions. There are two ways that a user can contribute content to the open access project. The first method is to make any amount of original contribution z_i^o at a unit cost c. The second method is to copy any portion of the proprietary output from the owner and make a reproduction z_i^r at a unit cost cq, where $q \in [0, 1]$ is a scalar reflecting the difficulty of unauthorized copying. We think of q as the extent of anti-circumvention enforcement against unauthorized copying; the higher the extent of enforcement, the higher the cost of reproduction. As long as the enforcement is imperfect, i.e., $q \neq 1$, the cost of reproduction is always lower than the cost of original production. Furthermore, the total amount of reproduction is capped by the output of the proprietary project, i.e., $\int_i z_i^r \leqslant \hat{x}$.

The owner files a lawsuit against the creator, claiming that the creator's project is used for copyright infringement. Settlement negotiations are then conducted against the background of a possible trial. The owner proposes a settlement agreement specifying that the owner will drop the lawsuit if the creator agrees to terminate the open access project. The creator decides whether to accept the settlement. If the creator accepts, there will be no open access project, and the owner's proprietary project will be a monopoly. If the creator rejects, the owner will have to choose whether to proceed to trial or drop the case. If the owner chooses not to litigate, his proprietary project has to compete with the open access project. The creator will respond to this proposal only if the owner's threat to litigate is credible: the owner's expected gain in profits must cover his litigation costs.

The outside option of the settlement is a trial. The trial involves costs (legal fees, discovery costs) for both parties, summarized by the monetary amount $L > 0$. If a trial does take place, there is a probability $p \in [0, 1]$ that the owner will prevail and the open access project will be taken down. In that case, the owner's proprietary project is a monopoly. If the creator prevails, the court dismisses the claim, and the owner's proprietary project has to compete with the open access project. Figure 1 describes the sequence of events where a black node is a decision node of the owner, a blue node is a decision node of the creator or users, a red node is a nature's move.

Every individual is endowed with y units of time, which can be used to earn y units of numeraire consumption good, and the cost c is normalized to be 1. Every individual

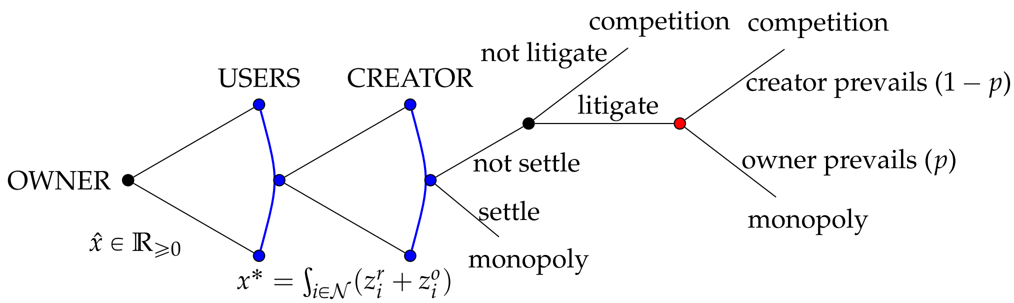


Figure 1: The Timing of the Game

i derives utility from his numeraire consumption good y_i, and the output of his chosen project x_i such that

$$u_i(y_i, x_i) = y_i + \varphi(x_i), \tag{1}$$

where $\varphi(\cdot)$ is continuously differentiable, increasing, and concave in x with $\varphi(0) = 0$. The institutional copyright owner only cares about profits; the creator and the users, however, move one step beyond a standard neoclassical preference. Our key assumption is that individuals differ by their social preference. Every user i is indexed by her social type $\theta_i \in \Theta := [0, \overline{\theta}]$ where $\overline{\theta} \in (0, 1]$. The distribution of the social preference is $F(\theta)$. The freelance creator observes her own social type $\theta_A \in \Theta$, and it is common knowledge that θ_A is drawn from the distribution $F(\theta)$. For ease of exposition, we assume $F(\cdot)$ to be a uniform distribution. The preference of the creator $i = A$, and any user $i \in \mathcal{N}$ are represented by the function U_i as follows:

$$U_i(y_i, \mathbf{x}) = u_i(y_i, x_i) + \theta_i \int_{j\in\mathcal{N}_{-i}} \varphi(x_j)dF(\theta_j). \tag{2}$$

One interpretation of the social preference is altruism (Charness and Rabin, 2002; Andreoni, 2007), where a contributor directly takes the benefits to other open access users into account. [11] Another interpretation of the social preference is positive network exter-

[11] There is plenty of experimental evidence on the effect of altruism in public good provision (Andreoni, 2006; 2007). The empirical finding in Fershtman and Gandal (2007) and the simple fact that open source projects establish hierarchies and are careful to credit contributors suggest that such motivations are important. Open access community explicitly recognize and honor the altruistic behavior. Open source projects, for example, are scrupulous about keeping track of contributors, which reflects the fact that giving credit to authors is essential in the open source movement. This principle is included as part of the nine key requirements in the "Open Source Definition". This point is also emphasized by Raymond (1999), who points out "surreptitiously filing someone's name off of a project is, in cultural context, one of the ultimate crimes."

nality (Katz and Shapiro, 1985; Farrell and Saloner, 1985), where the utility of a contributor is increasing in the number of other open access users. [12]

When the owner's project is a monopoly, the owner charges a price $t_m \geqslant 0$, after which every user $i \in \mathcal{N}$ independently and simultaneously chooses to accept or reject, and rejection leads to no use $x_i = 0$. Given other users' choice, the payoff of user i's is thus

$$U_i = \begin{cases} y - t_m + \varphi(\hat{x}) + \theta_i \int_{j \in \mathcal{N}_{-i}} \varphi(x_j) & \text{if } i \text{ chooses the proprietary project,} \\ y + \theta_i \int_{j \in \mathcal{N}_{-i}} \varphi(x_j) & \text{otherwise.} \end{cases} \tag{3}$$

When there is a competition between the proprietary project and the open access project, the owner instead charges a price $t_c \geqslant 0$, after which every user $i \in \mathcal{N}$ independently and simultaneously chooses the proprietary project, the open access project, or no use. Given other users' choice, the payoff of user i's is thus

$$U_i = \begin{cases} y - t_c + \varphi(\hat{x}) + \theta_i \int_{j \in \mathcal{N}_{-i}} \varphi(x_j) & \text{if } i \text{ chooses the proprietary project,} \\ y + \varphi(x^*) + \theta_i \int_{j \in \mathcal{N}_{-i}} \varphi(x_j) & \text{if } i \text{ chooses the open access project,} \\ y + \theta_i \int_{j \in \mathcal{N}_{-i}} \varphi(x_j) & \text{otherwise.} \end{cases} \tag{4}$$

Note that the option of no use is dominated by the use of open access.

Let the choice set of a user be $\mathcal{X}_i := \{\hat{x}, x^*, 0\}$, the choice set of the creator be $\mathcal{B}_A := \{settle, not\ settle\}$, the choice set of the owner be $\mathcal{B}_O := \{litigate, not\ litigate\}$. In this game, denote the owner's pure strategy as the tuple $(t_m, t_c, \hat{x}, b_O(x^*))$ where $t_m \in \mathbb{R}_{\geqslant 0}$, $t_c \in \mathbb{R}_{\geqslant 0}$, $\hat{x} \in \mathbb{R}_{\geqslant 0}$, and $b_O \colon \mathbb{R}_{\geqslant 0} \to \mathcal{B}_O$, the creator's pure strategy as $b_A : \Theta \times \mathbb{R}_{\geqslant 0} \times \mathbb{R}_{\geqslant 0} \to \mathcal{B}_A$,

This point was also emphasized in Lerner et al. (2006) where they conduct interviews with open source project managers and SourceForge officials.

[12]Surveys of open source participants indicate that programmers want to have an impact with their contributions, much as academics do. They appear to enjoy being part of important projects, especially projects that have a large user base. This suggests a relationship between altruism and the extent to which code is helpful to casual users. Shah (2006) quotes one programmer on this: "Why work on something that no on will use? There's no satisfaction there." Other supporting evidence includes that some programmers report that they monitor discussions of features they have developed even though they rarely take part in them. In the context of Wikipedia, Zhang and Zhu (2011) show that the number of other users on the Wikipedia platform, namely audience size, positively influences the amount of editing. Gorbatai (2011) shows that experienced editors become more active when observing prior edits by casual users that signal an interest in the topic. Kummer (2013) shows how exogenous shocks in readership lead to increased editing behavior on those articles.

the user i's pure strategy as the triple $(x_i(\hat{x}, x^*), z_i^o(\theta_i, \hat{x}), z_i^r(\theta_i, \hat{x}))$ where $x_i \colon \mathcal{H} \times \Theta \to \mathcal{B}$, $z_i^o \colon \mathcal{H} \times \Theta \to \mathbb{R}_{\geqslant 0}$, and $z_i^r \colon \mathcal{H} \times \Theta \to \mathbb{R}_{\geqslant 0}$. To interpret, a pure strategy of the owner is a complete contingent plan that picks the production, the monopoly price, the competition price, and the litigation choice for every realization of the total contributions x^*. A pure strategy of the creator is a contingent plan that picks the settlement choice for every realization of her social type θ_A and the output of both projects $\hat{x}$ and x^*. A pure strategy of any user i is a contingent plan that picks whether and which project to use and her pair of original contribution and reproduction for every realization of her social type θ_i and the output of the proprietary projects $\hat{x}$.

We use (pure strategy) perfect Bayesian equilibrium as the solution concept.

Definition 1. *A strategy profile* $\{(t_m, t_c, \hat{x}, b_O(x^*)), b_A(\theta_A, \hat{x}, x^*), (x_i(\hat{x}, x^*), z_i^o(\theta_i, \hat{x}), z_i^r(\theta_i, \hat{x}))_{i \in \mathcal{N}}\}$ *is a (pure strategy) perfect Bayesian equilibrium such that*

i) For any $i \in \mathcal{N}$*, given* $\hat{x} \in \mathbb{R}_{\geqslant 0}$*,* $(x_i(\hat{x}, x^*), z_i^o(\theta_i, \hat{x}), z_i^r(\theta_i, \hat{x}))$ *is a best response to* $(x_{-i}, z_{-i}^o, z_{-i}^c)$*.*

ii) Given $b_A(\theta_A, \hat{x}, x^*)$*, and* $(x_i(\hat{x}, x^*), z_i^o(\theta_i, \hat{x}), z_i^r(\theta_i, \hat{x}))_{i \in \mathcal{N}}$*, the owner chooses* $(t_m, t_c, \hat{x}, b_O(x^*))$ *to maximize his payoff.*

iii) Given $(t_m, t_c, \hat{x}, b_O(x^*))$*, and* $(x_i(\hat{x}, x^*), z_i^o(\theta_i, \hat{x}), z_i^r(\theta_i, \hat{x}))_{i \in \mathcal{N}}$*, the creator chooses* $b_A(\theta_A, \hat{x}, x^*)$ *to maximize her payoff.*

Copyright law is announced ex-ante; it is summarized as a pair $(p, q) \in [0, 1] \times [0, 1]$ where p is the probability that the owner wins an injunction against the open access project, and q is the extent of anti-circumvention enforcement against unauthorized copying. [13] (p, q) is common knowledge for all players. A copyright law (p, q) is optimal if it maximizes the total payoffs of all users, i.e., $\int_{i \in \mathcal{N}} u_i(y_i, x_i)$, subject to the constraints that the allocation $\{y_i, x_i\}_{i \in \mathcal{N}}$ can be implemented in a perfect Bayesian equilibrium.

[13] As mentioned in Section 2, copyright is a bundle of two major rights: (i) the exclusive right to make copies and distribute them, (ii) the exclusive right for further derivative works. We thus think of p as measuring the extent of a country's enforcement over the exclusive right to make copies and distribute them, and q as measuring the extent of a country's enforcement over the exclusive right for further derivative works. The latter usually increases in how broad the provisions on the fair use doctrine are, how stringent the requirements for derivative works are, how friendly the court is to "copyleft", etc.

In what follows, we first characterize the set of equilibria of the game. We then show how copyright law changes the set of equilibria, and solve for the optimal copyright policy. We conclude this section with testable predictions that we take to the empirical exercise.

4 Theory: Analysis

4.1 Litigation

If the owner's project is a monopoly, a user will pay for the project if

$$t_m \leqslant \varphi(\hat{x}). \tag{5}$$

The profit-maximizing monopoly price is $t_m = \varphi(\hat{x})$.

If the owner's project competes with the open access project, given the production of the open access project x^*, the proprietary project $\hat{x}$, and the price t_c, a user will accept the price versus open access if

$$t_c \leqslant \varphi(\hat{x}) - \varphi(x^*).$$

The profit-maximizing price in competition is $t_c = \max\{\varphi(\hat{x}) - \varphi(x^*), 0\}$. [14] If $\varphi(\hat{x}) \leqslant \varphi(x^*)$, any positive price is not incentive compatible and $t_c = 0$. The rent from pursuing litigation is thus $n(t_m - t_c)$.

There are three possible outcomes of the litigation process: (i) the threat of litigation is not credible, (ii) the threat of litigation is credible and the creator chooses to settle, (iii) the threat of litigation is credible but the creator chooses to settle, so the two parties proceed to the costly trial. The next lemma characterizes the necessary and sufficient condition for a settlement agreement.

Lemma 1. *Given x^* and $\hat{x}$, a settlement agreement is reached if and only if (i) the threat of*

[14] If t is such that the inequality is strict, then the owner can always raise t to $t + \varepsilon$ and earn a higher profit. Therefore, the constraint is binding if and only if $t > 0$.

litigation is credible, i.e., $pn\min\{\varphi(x^*),\varphi(\hat{x})\} > L$, *(ii) the social type of the creator is bounded above by the user value ratio, i.e.,* $\theta_A \in [0, \frac{p}{1-p}\min\{\frac{\varphi(\hat{x})}{\varphi(x^*)},1\}]$. [15]

Proof. Suppose the threat of litigation is credible, the creator of type θ_A will settle if

$$\begin{aligned} & y + (1+\theta_A n)[\varphi(\hat{x}) - t_m] \\ \geqslant & y - L + p(1+\theta_A n)[\varphi(\hat{x}) - t_m] + (1-p)(1+\theta_A n)\max\{\varphi(x^*), \varphi(\hat{x}) - t_c\}, \end{aligned}$$

substituting optimal (t_m, t_c), we get

$$L \geqslant (1-p)(1+\theta_A n)\varphi(x^*).$$

The owner's gain from blocking open access is his additional rent exceeding the profits in competition, i.e., $n(t_m - t_c)$. His threat of litigation is credible if

$$pn(t_m - t_c) > L,$$

substituting optimal (t_m, t_c), we have

$$pn\min\{\varphi(x^*), \varphi(\hat{x})\} \geqslant L. \tag{6}$$

Therefore, a settlement with type $\theta_A \in [0, \overline{\theta}]$ is feasible if

$$pn\min\{\varphi(x^*), \varphi(\hat{x})\} \geqslant (1-p)(1+\theta_A n)\varphi(x^*),$$

or equivalently

$$\theta_A \leqslant \frac{p}{1-p}\min\{\frac{\varphi(\hat{x})}{\varphi(x^*)}, 1\}. \tag{7}$$

If $pn(t_m - t_c) \leqslant L$ such that litigation is not credible, the creator of any type θ_A will not respond to the settlement offer. If $\theta_A > \frac{p}{1-p}\min\{\frac{\varphi(\hat{x})}{\varphi(x^*)}, 1\}$, the creator will choose not to settle even if litigation is credible. □

Let α be the equilibrium probability that the open access project survives (neither

[15] We make the tie-breaking assumption that if the owner is indifferent between a trial and no trial, then he will not litigate.

taken down by the court nor settled by the creator). Now we can determine the probability α in equilibrium. In the case where litigation is not credible, the open access project always survives, i.e., $\alpha = 1$. If the expected rent from pursuing a trial exceeds the cost of litigation such that the threat of litigation is credible, then it follows from lemma 1 that the open access project can only survive if the creator has a social type above the threshold and she prevails in the court; therefore, the equilibrium probability is $\alpha(\hat{x}, x^*, p) = (1-p)(1 - F(\frac{p}{1-p} \min\{\frac{\varphi(\hat{x})}{\varphi(x^*)}, 1\}))$. We collects these observations as follows:

$$\alpha(\hat{x}, x^*, p) = \begin{cases} 1 & \text{if } pn \min\{\varphi(x^*), \varphi(\hat{x})\} \leqslant L, \\ (1-p)(1 - F(\frac{p}{1-p})) & \text{if } pn \min\{\varphi(x^*), \varphi(\hat{x})\} > L \text{ and } \hat{x} \geqslant x^*, \\ (1-p)(1 - F(\frac{p\varphi(\hat{x})}{(1-p)\varphi(x^*)})) & \text{if } pn \min\{\varphi(x^*), \varphi(\hat{x})\} > L \text{ and } \hat{x} < x^*. \end{cases} \quad (8)$$

4.2 Production

Denote $\mathcal{N}^*$ as the set of users who choose to use the open access project, and n^* is the measure of $\mathcal{N}^*$. The optimal original contribution of any user $i \in \mathcal{N}^*$ satisfies

$$[1 + \theta_i n^*] \left\{ \frac{\partial \alpha(\hat{x}, x^*, p)}{\partial x^*} \varphi(x^*) + \alpha(\hat{x}, x^*, p) \varphi'(z_i^o + \int_{j \neq i} z_j^o + \int_j z_j^r) \right\} \leqslant 1, \quad (9)$$

with equality if $z_i^o > 0$. The total amount of reproduction has to satisfy the feasibility constraint

$$\int_i z_i^r \leqslant \hat{x}. \quad (10)$$

Because of that, when choosing individual reproduction, $\alpha(\hat{x}, x^*, p)$ is constant in x^*. The optimal reproduction of any user $i \in \mathcal{N}^*$ thus satisfies,

$$\alpha(\hat{x}, x^*, p)[1 + \theta_i n^*] \varphi'(z_i^r + \int_{j \neq i} z_j^r + \int_j z_j^o) \leqslant q, \quad (11)$$

with equality if $z_i^r > 0$.

Define $\bar{x}(\theta, n^*; p, q)$ as the individual satiation level of the open access project's output such that equation 11 holds with equality if $\bar{x}(\theta, n^*; p, q) < \hat{x}$, and equation 9 holds with equality if $\bar{x}(\theta, n^*; p, q) \geqslant \hat{x}$. The best response of user i depends on the individual pro-sociality θ, the number of open access users n^*, and the extent of anti-circumvention enforcement q. It is easy to show by implicit function theorem, $\bar{x}(\theta, n^*; p, q)$ is increasing in θ and n^*, and decreasing in q.

Define $\overline{\theta}^r(\hat{x})$ as the threshold type separating reproduction alone and a mix of reproduction and original contribution such that $\bar{x}(\overline{\theta}^r, n^*; p, q) = \hat{x}$.

Lemma 2. *If $\overline{\theta}^r(\hat{x}) \geqslant \overline{\theta}$, original contribution is zero. If $\overline{\theta}^r(\hat{x}) < \overline{\theta}$, only the highest type of open access users make positive original contributions.*

Proof. Since $\bar{x}(\theta, n^*; p, q)$ is increasing in θ, below $\overline{\theta}^r$ the reproduction constraint is not binding, and above $\overline{\theta}^r$ the reproduction constraint is binding. For $\theta < \overline{\theta}^r$, $\bar{x}(\theta, n^*; p, q) < \hat{x}$ so reproduction alone reaches their satiation level, and therefore $z_i^o(\theta_i) = 0$ for $\theta_i < \overline{\theta}^r$. It follows that if $\overline{\theta}^r(\hat{x}) \geqslant \overline{\theta}$, $z_i^o(\theta_i) = 0$ for any $\theta_i \in [0, \overline{\theta}]$. Since $q \leqslant 1$, we have $z_i^o > 0$ for some i if and only if $\int_i z_i^r = \hat{x}$. For $\theta \geqslant \overline{\theta}^r$, it must be that the original contribution constraint is only binding for $\overline{\theta}$, and therefore $z_i^o(\theta_i) = 0$ for $\theta_i < \overline{\theta}$ and $z_i^o(\overline{\theta}) > 0$ for at least some i. $\square$

If $\overline{\theta}^r(\hat{x}) < \overline{\theta}$, the optimal contribution of user i is

$$(z_i^o, z_i^r)(\theta_i) = \begin{cases} (0,0) & \text{if } \theta_i < \overline{\theta}^r, \\ (0, \hat{x} - \int_{j \neq i} z_j^r(\theta_j)) & \text{if } \theta_i < \overline{\theta} \text{ and } \theta_i \geqslant \overline{\theta}^r, \\ (\bar{x}(\overline{\theta}, n^*, 1) - \hat{x} - \int_{j \neq i} z_j^o(\overline{\theta}), \hat{x} - \int_{j \neq i} z_j^r(\theta_j)) & \text{if } \theta_i = \overline{\theta}. \end{cases} \tag{12}$$

There are two cases where a user may contribute zero: either the user is of a lower thus non-binding type, or the user is of the highest and binding type but others' provision has reached her satiation level. The contributing highest type will reproduce up to $\hat{x}$, and then make original contributions until her satiation level is reached.

The total contributions x^* satisfies

$$x^* = \begin{cases} \bar{x}(\overline{\theta}, n^*, 1) & \text{if } \overline{\theta}^r(\hat{x}) < \overline{\theta}, \\ \bar{x}(\overline{\theta}, n^*, q) & \text{if } \overline{\theta}^r(\hat{x}) \geqslant \overline{\theta}. \end{cases} \tag{13}$$

The output of the proprietary project satisfies

$$\max_{\hat{x}} nt_c + I\{pn\min\{\varphi(x^*), \varphi(\hat{x})\} > L\}[(1-\alpha(\hat{x}, x^*, p))n(t_m - t_c) - L] - \hat{x},$$

which states that in absence of the litigation strategy, the owner earns profits at the competition price; if litigation is credible and the open access project does not survive, he can earn rents by charging every user an extra fee $t_m - t_c$.

Proposition 1. *Either $\mathcal{N}^* = \mathcal{N}$ or $\mathcal{N}^* = \emptyset$.*

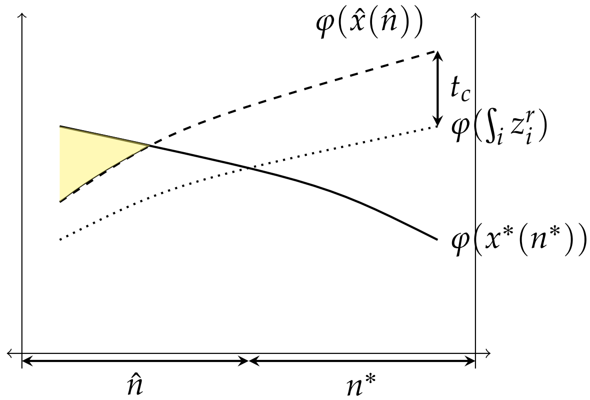
Proof. Suppose $\mathcal{N}^*$ is such that $\varphi(\hat{x}) \leqslant \varphi(x^*)$. For user $i \in \mathcal{N}^*$, he will keep using open access. For user $i \notin \mathcal{N}^*$, he has a profitable deviation towards open access. There does not exist a $t > 0$ such that any user of any type will accept the offer. It follows that $\mathcal{N}^* = \mathcal{N}$.

Suppose $\mathcal{N}^*$ is such that $\varphi(\hat{x}) > \varphi(x^*)$. The owner can set $t = \min\{\varphi(\hat{x}) - \alpha\varphi(x^*), 0\}$. For user i who chose to pirate, he is weakly better off by accepting the offer. For user $i \in \mathcal{N}^*$, he has a profitable deviation towards the proprietary bundle. It follows that $\mathcal{N}^* = \emptyset$. □

The above proposition shows why social preference leads to multiple equlibria: open access participation $\mathcal{N}^*$ is self-fulling. Suppose pro-social users hold the belief that open access is more efficient than proprietary, then they will contribute to a much greater level of public good, which attracts every self-interest user, and open access becomes indeed more efficient than proprietary. Suppose instead that pro-social users believe open access is less efficient than proprietary, then they will leave the open access project, making the public good of little value to its users and pushing the self-interest users to proprietary, consequently open access end up less efficient than proprietary.

4.3 Multiple Equilibria

The next theorem is the central piece of the theory that established the necessary and sufficient condition for the existence of a original contribution equilibrium (fair use equilibrium). It shows the perpetual adverse effect of an uncoordinated effort on harsh copy-



*highlighted area: no feasible t exists.

Figure 2: Equilibrium Number of Open Access Users

right enforcement. Increasing either p or q without lowering the other dimension of the copyright law may drive the pair (p, q) out of the multiple equilibria region, and even though individuals are other-regarding, such enforcement may nonetheless erode open access activities altogether.

Lemma 3. *there exists a reproduction equilibrium where* $\int_{i \in \mathcal{N}} z_i^r > 0$ *and* $\int_{i \in \mathcal{N}} z_i^o = 0$ *if* $p \geqslant \frac{1-q}{1+\frac{1}{\bar{\theta}}}$.

Theorem 1. *Suppose the threat of litigation is credible.* $\int_{i \in \mathcal{N}} z_i^o > 0$ *if and only if* $p \leqslant \frac{\bar{\theta}}{1+\bar{\theta}}$. *Furthermore, the original contribution equilibrium is unique if and only if* $p < \frac{\bar{\theta}}{1+\bar{\theta}}(1-q)$.

Proof. Suppose $x^* > \hat{x}$. $x^* = \bar{x}(\bar{\theta}, 1; p, 1)$, and equation 9 can be simplified as

$$(1-p)[1+\bar{\theta}]\varphi'(x^*) = 1. \tag{14}$$

The optimal output of the owner satisfies

$$p\varphi'(\hat{x}) = 1 \tag{15}$$

Given $p < \frac{1+\overline{\theta}}{2+\overline{\theta}}$, we have

$$
\begin{aligned}
& p < (1-p)(1+\overline{\theta}) \\
\Leftrightarrow & \frac{1}{(1-p)(1+\overline{\theta})} < \frac{1}{p} \\
\Leftrightarrow & \varphi'(x^*) < \varphi'(\hat{x}) \\
\Leftrightarrow & x^* > \hat{x}.
\end{aligned}
$$

Suppose $\hat{x} \geqslant x^*$. $x^* = \bar{x}(\overline{\theta}, 0; p, q)$, and equation 11 is

$$(1-p)(1-\frac{p}{(1-p)\overline{\theta}})\varphi'(x^*) = q. \tag{16}$$

The optimal output of the owner satisfies

$$\varphi'(\hat{x}) = 1 \tag{17}$$

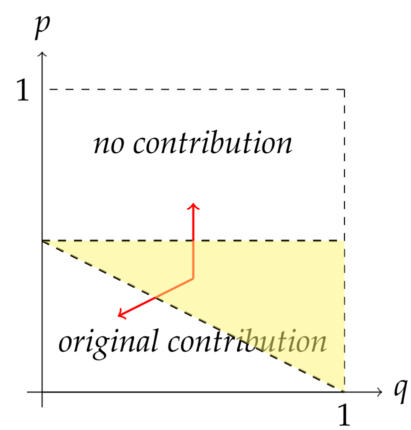
Given $p \geqslant \frac{1-q}{1+\frac{1}{\overline{\theta}}}$, we have

$$
\begin{aligned}
& 1-(1+\frac{1}{\overline{\theta}})p \leqslant q \\
\Leftrightarrow & 1 \leqslant \frac{q}{(1-p)(1-\frac{p}{(1-p)\overline{\theta}})} \\
\Leftrightarrow & \varphi'(\hat{x}) \leqslant \varphi'(x^*) \\
\Leftrightarrow & \hat{x} \geqslant x^*.
\end{aligned}
$$

□

In reproduction equilibrium, litigation complements anti-circumvention enforcement against unauthorized use. In original contribution equilibrium, litigation is purely for rent-seeking though under the name of justice.

We conclude this section by highlighting the legal change shown by the red arrow in Figure 3 that forms the basis of the empirical exercise.



*highlighted area: both reproduction equilibrium ($\int_{i \in \mathcal{N}} z_i^r > 0, \int_{i \in \mathcal{N}} z_i^o = 0$) and original contribution equilibrium ($\int_{i \in \mathcal{N}} z_i^o > 0$).

Figure 3: Copyright law (p, q) that supports open access.

5 Data Description

In this section, we study program developers' open source behavior in Github, the largest online host of open source code and repositories in the world. Two sources of variations in copyright law are explored to explain variations in open source behavior including number of repositories shared and coding contributions: the timing of DMCA takedown notice, and the longitudinal software piracy rate of participating countries of WIPO Copyright Treaty. The goal of the empirical section are two-folds: (i) test the implications of the theory and support the relevance of the model, and (ii) provide suggestive evidence that the current copyright law starts to discourages open access activities.

Founded in 2008, Github is an online platform hosting open-source code and repositories. By 2015, Github reports over 9 million users and over 20 million repositories, making it the largest host of source code in the world. GitHub is a collaborative code hosting site built on top of the git version control system. A user can have three possible roles: administrators/owners, contributors, and testers. The owner initiates and tracks the development of the repository, to which the contributors can make commits to the repository that improves its content. A commit is a collection of several lines of codes that adds to the functionality of a program. The owner can then review the codes for correctness and complicity with standards, and approve the commits by merging them into the codebase of the repository. In addition to code hosting, collaborative code review, and integrated issue tracking, GitHub has also integrated social features. Users are able to subscribe to information by "watching" projects and "following" users, resulting in a feed of information on those projects and users of interest. Users also have profiles that can be populated with identifying information and contain their recent activity within the site.

Data on Github users' behavior is obtained from the GHTorent project (Gousios (2013)). GHTorrent is an offline persistent mirror of Github's event streams, containing all user events, including commits, comments, pull requests, etc. It uses the Github API to collect raw data and extract, archive and share queriable metadata. The GHTorent toolset allows for retrieving the full history for a single project and the full list of actions for a single developer, which makes it popular in the software data mining community. In particular, we have raw data on which user commits to which project at what time.

We begin with the 2016-01-18 dump. In processing the data, several key decisions are made. Git commits contain information about both the author (the person who originally changed the code) and the committer (the persons who last applied the change), each with their own timestamp. The two are not necessarily one and the same person (e.g., they can differ when someone rebases or cherry picks a commit). In this paper we consider only the commits which record the same person as both author and committer (97.8% of the commits in our dataset), and record the date at which a commit was authored (rather than committed). In addition, git allows commit metadata, including the authorship date, to be overwritten. For instance, commits with the 1969-12-31 or 2050-07-18 timestamps underwent such a history rewriting process. We excluded these observations.

Using the methodology described above we assembled a panel dataset of Github users, in which each observation contains the time-invariant characteristics of the user, as well as the user's Github activities in several dimensions for each month upon registration. The main individual characteristics are location of the user's IP address (from which nationality is inferred), the registration time, preferred programming languages, company affiliation, and whether the user is an individual account or group account. The activities observed over time are the number of unique commits made by the user, the number of repositories shared by the user, the cumulative number of followers, and the total number of watchers for each repository the user shared.

Variable/Year	2008	2009	2010	2011	2012	2013	2014	2015
Original repositories per user	0.10 (0.72)	0.23 (1.03)	0.77 (2.02)	1.65 (3.34)	1.88 (3.62)	2.05 (3.99)	2.23 (4.38)	2.55 (5.72)
Number of commits per repository	94.82 (433.71)	66.61 (462.16)	44.54 (350.67)	28.54 (195.62)	23.72 (150.28)	16.43 (77.51)	15.84 (103.27)	15.28 (159.43)
Followers per user	0.94 (9.11)	1.36 (11.75)	2.24 (28.12)	2.84 (66.20)	1.93 (50.38)	0.47 (34.10)	0.52 (40.21)	0.16 (20.47)
Watchers per repository	0.63 (30.28)	1.21 (53.81)	3.74 (74.02)	3.26 (45.14)	5.62 (77.17)	8.47 (97.34)	10.91 (120.66)	14.78 (180.62)
Number of new users	10,406	23,219	44,770	82,507	163,524	245,397	201,632	132,468
Proprietary software piracy rate	60.79 (21.00)	60.83 (20.99)	60.33 (20.96)	59.82 (21.33)		58.89 (21.56)		57.66 (21.74)
Number of countries	104	106	106	106		106		106

Table 1: Summary Statistics of the GHTorrent Sample and Country Aggregates

The DMCA takedown notice is a request to remove content of individual users whose shared repository is disputed by some claimed copyright holders. We collected the data from the Github DMCA repository. Every unit of observation is a notice event, and for each notice, we collected the following information: the date of the DMCA notice addressed to the user, the name of the claimed copyright-holder who sent the notice, the login name of a Github user who received the notice, whether the user is an individual or a group, whether the repository is taken down by Github at the time of data collection, whether the Github user replies a counter-notice and the date of a counter-notice if there is one, whether the plaintiff retracts the notice and the date when plaintiff retracts the notice. There are $2,213$ complaints filed between 2011 and 2014. $1,091$ of all the cases can be matched with a loginname in GHTorrent, which results in a total of $1,065$ users.

This figure shows the distribution of the event dates. The earliest notice dates back to Jan. 2011, and the latest notice comes in Dec. 2014, by which month all users have received their notice. After an initial spike of complaints in Feb. 2011 by the Sony Computer Entertainment America LLC, the notices have been filed at a smooth yet slightly

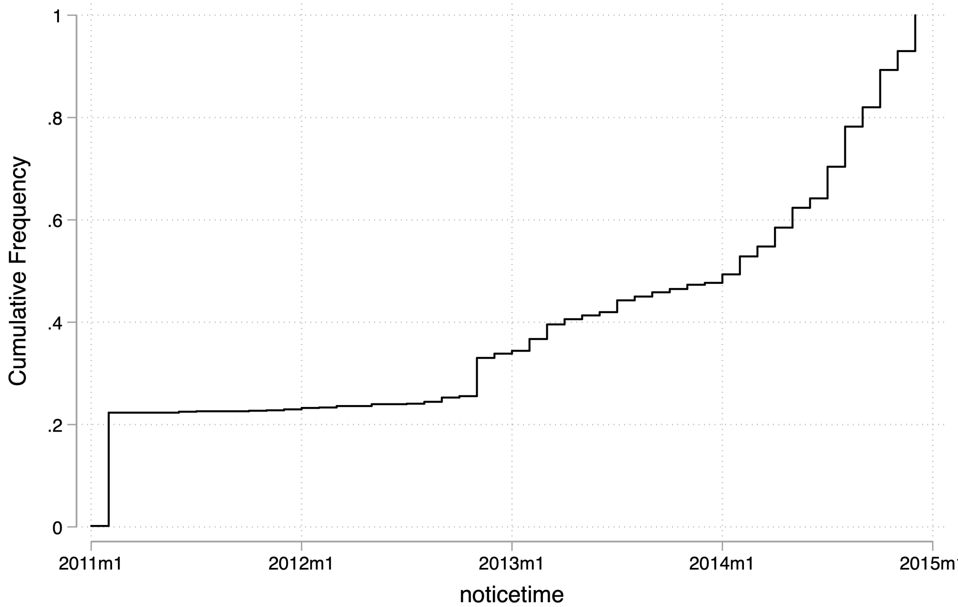


Figure 4: The distribution of notice time

increasing rate.

Data regarding the rates of software piracy was provided by the Business Software Alliance (BSA). BSA is a network of software manufacturers and includes such companies as Microsoft and Novell. Software piracy rates are calculated by estimating the demand for software based on the worldwide number of PC shipments and the sale of U.S. business applications. By assuming that for each new personal computer sold there will be a set of accompanying software sales, the difference between expected demand and supply (in the form of sales) is attributed to software piracy. Piracy rates are reported as percentages, with 0% indicating no piracy, and 100% indicating all software is pirated. Starting from 2011, BSA reports the piracy rates every two years. Despite its limitations, the BSA piracy rates are one of the most commonly accepted indicators in the industry.

As shown in Table 1, average piracy rates across individual countries steadily decline from 61% in 2008 to 57% in 2015. There are sizable regional variations: The highest regional software piracy rates are seen in Eastern Europe and Latin America, with rates around 60%, suggesting six of every 10 software packages in use are pirated copies. North America and Western Europe have the lowest regional software piracy rates at around

30%. Several developing countries improved upon their software piracy rates more than 8% during the sample period, with China, India, Sri Lanka, Vietnam, and Mexico at the top of the list.

16.95% of all the users created from 2008 to 2015 has geographic information in GHTorrent, which results in a total $910,333$ users. 209 countries are covered by these users, showing the wide span of international participation in Github. 106 out of these countries can be matched with the software piracy rates data. After dropping users from the US, our main sample consists of almost two thirds of the users, resulting in a total of $538,770$ users.

5.1 Hypotheses and Empirical Strategy

This section describes two econometric specifications to test the theoretical implications outlined in Section 4, and the identification assumption required for consistent estimate. The first specification is an event study, which allows for the effect of a takedown notice to vary over the subsequent periods. The identification assumption is the timing of a notice is as good as random assignment. Thus, early-receivers and late-receivers can serve as counterfactual for one another. In the second specification, we estimate a two-way fixed effect model in a natural experiment setting. The identification assumption is that copyright enforcement in the home countries is exogenous to users' decisions, and foreign users' exposures to different levels of enforcement approximates random assignment. We further control for the omitted variable bias caused by individual heterogeneity and time trend, based on the assumption that the time-invariant characteristics of the user is unique to them and not correlated with other users' individual characteristics.

Hypothesis 1. *The DMCA takedown notice has a negative effect on the number of repositories shared by the receiving user.*

We match the data file of DMCA notice events with the relevant Github user information. The two files are merged using the users' loginname, and each Github user is then identified by his ID assigned by GHTorrent.

Let i be a user, t be a calendar month where we observe the outcomes, and τ be the time-since-event, e.g., $\tau = 3$ means three months after the notice. The regression equation we use is as follows:

$$numrepo_{it} = \sum_{\tau=-30,\tau\neq 0}^{30} \beta_\tau notice_{i,t-\tau} + \alpha_i + \eta_t + \varepsilon_{it} \tag{18}$$

where $numrepo_{it}$ is the number of repositories shared by the user i in month t. $notice_{i,t-\tau}$ is an indicator variable for the notice taking place in period $t-\tau$, so $notice_{i,t-\tau} = 1$ if as of time t, the user i received a takedown notice τ months ago. α_i is user fixed effect, η_t is month fixed effect. We drop any observation beyond the $[-30, 30]$ time window. We combine the monthly notice indicators into quarter bins to increase the power of estimation, such as $(-30, -28), (-27, -25), ..., (-3, -1), (0, 2), (3, 5), ..., (28, 30)$ relative to the month of receiving the notice at month 0. We follow the convention to normalize the effect of a notice in the reference quarter $(-3, -1)$ to be zero. The standard errors are clustered at the user level.

The coefficients β_τ for $\tau \geqslant 0$ capture the persistent effects of a takedown notice, and hypothesis 1 expects them to be significantly negative. The terms β_τ for $\tau < 0$ provide a placebo or falsification test. In absence of anticipating effects, model specification, or omitted confounding variables, we expect them to be not significantly different from zero.

$$commit_{rit} = \sum_{\tau=-30,\tau\neq -1}^{\tau=\underline{\tau}} \beta_\tau notice_{i,t-\tau} + \alpha_r + \eta_t + \varepsilon_{rit} \tag{19}$$

where $commit_{rit}$ is the logarithm of the number of commits contributed to repository r shared by user i in month t. $notice_{i,t-\tau}$ is an indicator variable for the notice taking place in period $t-\tau$, so $notice_{i,t-\tau} = 1$ if as of time t, the user i received a takedown notice τ months ago. α_r is repository fixed effect, η_t is month fixed effect. The standard errors are clustered at the user level.

Hypothesis 2. *Repositories shared by foreign users from countries with higher copyright enforcement attract lower contributions.*

We match the data file of country-level longitudinal software piracy rates with the relevant Github user who has nationality information available. The two files are merged using variable ccTLD, a two-digit country code technically known as country code top-level domain.

Let i be a user, c be a country, t be a year. The regression equation we use is as follows:

$$commit_{ict} = \gamma Piracy_{ct} + \lambda X_{it} + \alpha_c + \eta_t + \varepsilon_{it} \tag{20}$$

where $commit_{ict}$ is the logarithm of the number of commits contributed to all the repositories shared by user i from country c in year t, and we use both average commits and total commits for robustness. $Piracy_{ct}$ is the software piracy rate of country c in year t, X_{ict} are the control variables of user i in year t, α_c is country fixed effect, η_t is year fixed effect. The standard errors are clustered at the user level.

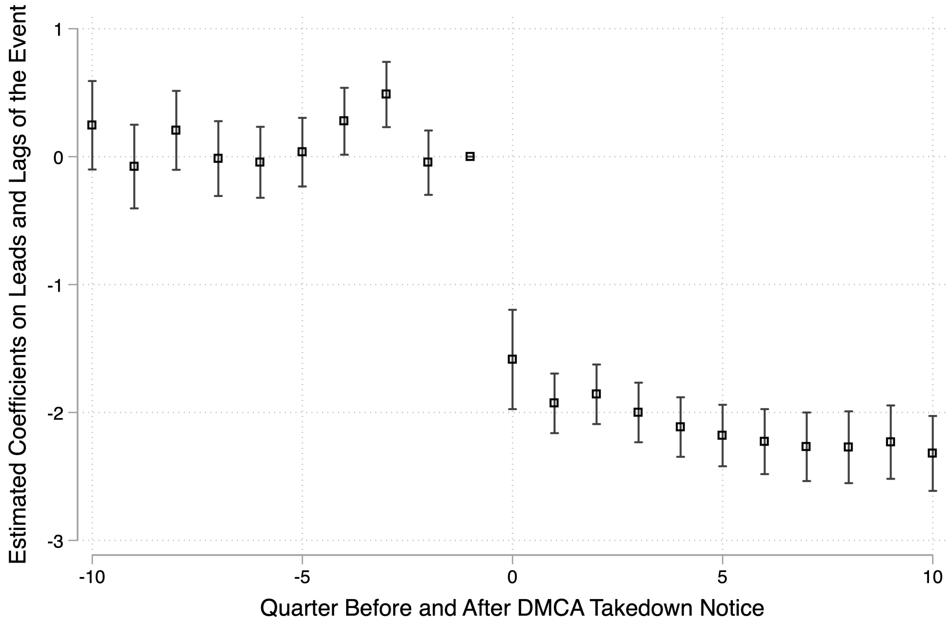
The coefficient γ captures the effect of copyright enforcement on contributions, and hypothesis 2 expects it to be significantly positive.

6 Empirical Results

6.1 DMCA evidence

This figure shows that the number of repositories shared by an average Github user drops immediately following a DMCA notice, and the decrease lasts for at least 10 quarters after the notice within the event window studied. Compared to the average pre-notice level, the number of repositories on average decreases by 1.9 repositories during the first five quarters after the notice, and 2.2 repositories during the second five quarters after the notice. The figure also shows that the pre-trend is relatively stable and shows no sign of downward trending during the 10 quarters prior to the notice.

Further specifications reveal that the number of commits contributed to the repositories shared by the receiving user likewise drops sharply in the post-notice period. This suggests an explanation consistent with our theory: the notice makes the contributors



uncertain whether the next repository shared by the receiving user will be taken down because of the threat of litigation. For the subsequent repositories shared by this user, the norm shifts from $\mathcal{N}^* = \mathcal{N}$ to $\mathcal{N}^* = \varnothing$. Since no one is expected to contribute to the repository being shared, the user has strong disincentive to make her next repository open access in the first place. The notice has a persistent chilling effect towards all future open access sharing.[16]

6.2 Cross-country evidence

Table 2 reports the cross-country findings of copyright enforcement on peer contributions. For each dependent variable we report two specifications, where the set of covariates are sequentially expanded. Column (1)-(2) consider average commits, and column (3)-(4) consider total commits. The regression results suggests a significant positive relationship between country piracy rates and contributions to repositories. An improvement on software piracy by the home country reduces the commits to the repositories shared by its citizens. The estimated effect is robust to to the inclusion of users' experience in addition to the two-way fixed effects. The estimates in column (4) tell us that holding all other

[16]The Github team explicitly points this out in their readme file of the DMCA repository

	(1) Average Commits	(2) Average Commits	(3) Total Commits	(4) Total Commits
$Enforcement$	-0.0054***	-0.0042**	-0.0064***	-0.0057**
	(0.0021)	(0.0022)	(0.0027)	(0.0029)
$Experience$		0.0256***		0.3926***
		(0.0058)		(0.0078)
$Experience^2$		-0.0072***		-0.0258***
		(0.0005)		(0.0007)
Country FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	538,770	538,770	538,770	538,770
adj. R^2	0.027	0.035	0.129	0.154

Note: $commit_{ict}$ is the logarithm of the number of commits contributed to all the repositories shared by user i from country c in year t, and we use both average commits and total commits for robustness. $Enforcement_{ct}$ is the software compliance rate of country c in year t. $Experience$ is the number of years since the users' registration. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Estimation of the effect of copyright enforcement on open source contributions

things constant, a one percent increase in piracy rates is associated with a 0.5% rise in total commits. To put it in context, countries aggressive on copyright enforcement experienced a 2% annual decline in piracy rates (China, India, etc.), and users from these countries attract 1 fewer commit every three years purely because of the litigation risk.

7 Conclusion

This paper presents a case where the expansion and enforcement of property rights *lead to* the failure of the commons in cyberspace. The central thesis is that a strong copyright regime gives the copyright owner an unbalanced power against future creators. The owner can employ the threats of litigation to eliminate competition from open access and peer production under the name of justice. This thesis echos some of the remarks from key scholars and practitioners in the field. Lessig (2004) refers to this as "the concentration of power produced by a radical change in the effective scope of the law." As Linus put it in Torvalds and Diamond (2001), "to a large degree, finding peace in this intellectual property war is what open source is all about...open source would rather use the legal weapon of copyright as an invitation to join in the fun, rather than as a weapon against others."(p.210).

Our findings caution against the global effort towards strict copyright enforcement in cyberspace. Such effort could be counter-productive to the creativity of cyberspace, especially in developing countries where the development of open access is at its infancy. In theoretical model, we show that a strong copyright regime may erode the original contribution equilibrium completely. In data analysis, we show the chilling effect of a takedown notice on subsequent sharing is persistent. Instead of treating copyright infringement as strict liability, an optimal policy on cyber property rights is a balancing of interests between the owners and the future creators from the general public. This balance hinges critically on limitations on exclusive rights, such as fair use doctrine.

The idea that a harsh law crowds out other-regarding behavior appears not specific to our digital context (Bohnet et al. (2001), Gneezy and Rustichini (2000)). To the extent that the economic trade-off between material incentive to create and social welfare loss from restricting access is at the heart of copyright law, the social motivation to create and share we emphasize in this paper compels us to reevaluate the design of cyber property rights, and perhaps more broadly innovation policy. As Samuel Bowles puts it quite succinctly (2008), "the critical assumption in the conventional approach is not that other-regarding motives are absent but that policies that appeal to economic self-interest do not affect the salience of ethical, altruistic, and other social preferences."

References

Aaltonen, Aleksi and Stephan Seiler, "Cumulative growth in user-generated content production: evidence from Wikipedia," *Management Science*, 2016, *62* (7), 2054–2069.

Andreoni, James, "Philanthropy," *Handbook of the economics of giving, altruism and reciprocity*, 2006, *2*, 1201–1269.

Athey, Susan and Glenn Ellison, "Dynamics of open source movements," *Journal of Economics & Management Strategy*, 2014, *23* (2), 294–316.

Bénabou, Roland and Jean Tirole, "Incentives and prosocial behavior," *American economic review*, 2006, *96* (5), 1652–1678.

Benkler, Yochai, *The wealth of networks*, Yale university press, 2008.

Besley, Timothy and Maitreesh Ghatak, "Property rights and economic development," in "Handbook of development economics," Vol. 5, Elsevier, 2010, pp. 4525–4595.

Bohnet, Iris, Bruno S Frey, and Steffen Huck, "More order with less law: On contract enforcement, trust, and crowding," *American Political Science Review*, 2001, *95* (1), 131–144.

Boldrin, Michele and David Levine, "The case against intellectual property," *American Economic Review*, 2002, *92* (2), 209–212.

Casadesus-Masanell, Ramon and Pankaj Ghemawat, "Dynamic mixed duopoly: A model motivated by Linux vs. Windows," *Management Science*, 2006, *52* (7), 1072–1084.

Charness, Gary and Matthew Rabin, "Understanding social preferences with simple tests," *The quarterly journal of economics*, 2002, *117* (3), 817–869.

Ellingsen, Tore and Magnus Johannesson, "Pride and prejudice: The human side of incentive theory," *American economic review*, 2008, *98* (3), 990–1008.

Farrell, Joseph and Garth Saloner, "Standardization, compatibility, and innovation," *The RAND Journal of Economics*, 1985, pp. 70–83.

Gneezy, Uri and Aldo Rustichini, "A fine is a price," *The Journal of Legal Studies*, 2000, *29* (1), 1–17.

Goldfarb, Avi and Catherine Tucker, "Digital economics," *Journal of Economic Literature*, 2019, *57* (1), 3–43.

Gonzalez-Barahona, Jesus M, Gregorio Robles, Roberto Andradas-Izquierdo, and Rishab Aiyer Ghosh, "Geographic origin of libre software developers," *Information Economics and Policy*, 2008, *20* (4), 356–363.

Gousios, Georgios, "The GHTorrent dataset and tool suite," in "Proceedings of the 10th Working Conference on Mining Software Repositories" MSR '13 IEEE Press Piscataway, NJ, USA 2013, pp. 233–236.

Johnson, Justin Pappas, "Open source software: Private provision of a public good," *Journal of Economics & Management Strategy*, 2002, *11* (4), 637–662.

Kalliamvakou, Eirini, Georgios Gousios, Kelly Blincoe, Leif Singer, Daniel M German, and Daniela Damian, "The promises and perils of mining github," in "Proceedings of the 11th working conference on mining software repositories" 2014, pp. 92–101.

Katz, Michael L and Carl Shapiro, "Network externalities, competition, and compatibility," *The American economic review*, 1985, *75* (3), 424–440.

Keefer, Philip and Stephen Knack, "Why don't poor countries catch up? A cross-national test of an institutional explanation," *Economic inquiry*, 1997, *35* (3), 590–602.

Klein, Benjamin, Andres V Lerner, and Kevin M Murphy, "The economics of copyright" fair use" in a networked world," *American Economic Review*, 2002, *92* (2), 205–208.

Knack, S and P Keefer, "Institutions and Economic Performance: Institutional Measures Cross-Country Test Using Alternative," *Economics and Politics*, 1995, *7* (3), 207–227.

Lerner, Josh and Jean Tirole, "Some simple economics of open source," *The journal of industrial economics*, 2002, *50* (2), 197–234.

_ **and** _ , "The economics of technology sharing: Open source and beyond," *Journal of Economic Perspectives*, 2005, *19* (2), 99–120.

— **and** — , "The scope of open source licensing," *Journal of Law, Economics, and Organization*, 2005, *21* (1), 20–56.

— **, Parag A Pathak, and Jean Tirole**, "The dynamics of open-source contributors," *American Economic Review*, 2006, *96* (2), 114–118.

Lessig, Lawrence, "Free culture: how big media uses technology and the law to lock down culture and control creativity," 2004.

— , *Code: And other laws of cyberspace*, ReadHowYouWant. com, 2009.

Meurer, Michael J, "The settlement of patent litigation," *The RAND Journal of Economics*, 1989, pp. 77–91.

Mustonen, Mikko, "Copyleft—the economics of Linux and other open source software," *Information Economics and Policy*, 2003, *15* (1), 99–121.

Nagaraj, Abhishek, "Does copyright affect reuse? Evidence from Google Books and Wikipedia," *Management Science*, 2018, *64* (7), 3091–3107.

Nalebuff, Barry, "Credible pretrial negotiation," *The RAND Journal of Economics*, 1987, pp. 198–210.

North, Douglass C and Barry R Weingast, "Constitutions and commitment: the evolution of institutions governing public choice in seventeenth-century England," *The journal of economic history*, 1989, *49* (4), 803–832.

Norton, Seth W, "Poverty, property rights, and human well-being: a cross-national study," *Cato J.*, 1998, *18*, 233.

Novos, Ian E and Michael Waldman, "The effects of increased copyright protection: An analytic approach," *Journal of political economy*, 1984, *92* (2), 236–246.

Posner, Richard A, "Intellectual property: The law and economics approach," *Journal of Economic Perspectives*, 2005, *19* (2), 57–73.

Raymond, Eric S. and Bob Young, *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*, USA: O'Reilly amp; Associates, Inc., 2001.

Takhteyev, Yuri and Andrew Hilts, "Investigating the Geography of Open Source Software through Github."

Torvalds, Linus and David Diamond, *Just for fun: The story of an accidental revolutionary*, Harper Audio, 2001.

Varian, Hal R, "Copying and copyright," *Journal of economic perspectives*, 2005, *19* (2), 121–138.

von Engelhardt, Sebastian, Andreas Freytag, and Christoph Schulz, "On the geographic allocation of open source software activities," *International Journal of Innovation in the Digital Economy (IJIDE)*, 2013, *4* (2), 25–39.

Waldfogel, Joel, "Copyright research in the digital age: Moving from piracy to the supply of new products," *American Economic Review*, 2012, *102* (3), 337–42.

Zhang, Xiaoquan Michael and Feng Zhu, "Group size and incentives to contribute: A natural experiment at Chinese Wikipedia," *American Economic Review*, 2011, *101* (4), 1601–15.

Zittrain, Jonathan, *The future of the internet–and how to stop it*, Yale University Press, 2008.