How APIs Create Growth by Inverting the Firm

Seth G. Benzell[*], Jonathan Hersh[†], Marshall Van Alstyne [‡]

This draft: July 24, 2022

Abstract

Traditional asset management strategy has emphasized building barriers to entry or closely guarding unique assets to maintain a firm's comparative advantage. A new "Inverted Firm" paradigm, however, has emerged. Under this strategy, firms share data seeking to become platforms by opening digital services to third-parties and capturing part of their external surplus. This contrasts with a "pipeline" strategy where the firm itself creates value. This paper quantitatively estimates the effect of adopting an inverted firm strategy through the lens of Application Programming Interfaces (APIs), a key enabling technology. Using both public data and that of a private API development firm, we document rapid growth of the API network and connecting apps since 2005. We then perform difference-in-difference and synthetic control analyses and find that public firms adopting public APIs grew an additional 38.7% relative to similar non-adopters. We find no significant effect from the use of APIs purely for internal productivity, the pipeline strategy. Within the subset of firms that adopt public APIs, those that attract more third-party complementors and those that become more central to the network see faster growth. Using variation in network centrality caused by API degradation, an instrumental variables analysis confirms a causal role for APIs in firm market value. Finally, we document an important downside of external API adoption: increased risk of data breach. Overall, these facts lead us to conclude that APIs have a large and positive impact on economic growth and do so primarily by enabling an inverted firm as opposed to pipeline strategy.

1 Introduction

In the information age, the value of a firm rests fundamentally on how it gathers, shares and processes information.[1] While traditional approaches to asset management have emphasized closely guarding a firm's comparative advantage, a new digital management paradigm has emerged. This new approach relies on data's nonrival property, allowing it to be shared. Value creation and value capture based on regulating data access leads to an "Inverted Firm," where production moves from inside to outside. Third parties, and not just insiders, create much of the value. If openness creates a large enough ecosystem of interactions, then capturing even a small

[*]Argyros School of Business and Economics, Chapman University, Stanford HAI Digital Economy Lab & MIT Initiative on the Digital Economy, benzell@chapman.edu

[†]Argyros School of Business and Economics, Chapman University hersh@chapman.edu

[‡]Boston University & MIT Initiative on the Digital Economy mva@bu.edu

[1]The value of US corporate intangible assets increased from near zero in the early 1990s to more than $6.6 trillion in 2016 (Benzell and Brynjolfsson, 2019) – calculated as US corporate equity and liabilities less financial assets from Federal Reserve series Z.1 and less fixed capital from BEA table 4.1.

share of the resulting surplus can greatly benefit the inverted firm. Many of the world's most successful companies, such as Alphabet, Meta, and Amazon, have developed platforms that use their centrality in the digital economy to coordinate and monetize the activity of others.

Key to the inverted firm strategy are public Application Programming Interfaces or APIs. APIs are tools and protocols that allow computers to easily communicate with each other (Jacobson et al., 2011). Web accessibility allows public APIs to serve as conduits to business processes that the firm itself controls. APIs offer the dual virtues of practical modular design and precise metering of access, foundations of a digital ecosystem.

This paper investigates the inverted firm and pipeline strategies through the lens of public APIs, a core enabling technology. We provide empirical evidence that employing APIs helps firms grow and that they do so primarily by inverting the firm, enabling third party complementors, rather than improving the firm's own value creation. Production, enabled and moderated by APIs, moves from inside to outside the firm. We also find that wealth created by these strategies is of massive scale, increasing the market value of public firms an additional 12.9% versus similar firms after two years of adoption. While this number on its own is large, it excludes the growth induced in the thousands of smaller firms who build on these platforms.

We document the development of the public API network that links the business interests of companies together through a matrix of third-party applications calling one or more APIs. Using summary statistics, difference-in-difference estimates, instrumental variables, and synthetic control analyses, we document strong growth in market value among firms that adopt public facing APIs. This positive effect is not limited to large technology firms – we find a positive effect of similar magnitude for smaller publicly traded firms and those in other industries. We further show that high third-party engagement with APIs predicts particularly large gains. Firms with APIs that have more followers, developers, and connected apps see significantly larger growth.

The value from external connections holds not just for volume but also for network position. Firms with more central APIs grow faster. The fourteen public firms with APIs ranked in the top forty by betweenness centrality added $6.6 trillion dollars in market value from Q1 2007 to Q3 2020, a significant fraction of the US public equity market's appreciation over that time. We confirm that this relationship is causal via an instrumental variable analysis. When central APIs are degraded, they change the network centrality of other APIs. Exploiting this variation, we confirm a causal role for API network centrality in firm market value.

On the other hand, firms that use APIs exclusively for internal efficiency gains do not see statistically significant growth in market value after adoption. We further test the hypothesis that APIs create internal benefits, in particular the efficiency benefit of lowering adjustment costs, by measuring the evolution of a firm's Q – the ratio of market value to book value. If APIs primarily help firms by reducing adjustment costs, successful API adopting firms should not see large market value growth after controlling for their asset growth. We find the opposite, evidence consistent with API benefits derived from an external ecosystem.

Finally, we investigate an important downside of external API adoption, the risk of data breaches. We find that firms with public APIs see significantly increased risk of hack events in the years after opening an API. This result remains after controlling for API popularity. Private APIs do not create the same exposure risk. Thus, public APIs offer greater gains along with higher risks, an important information systems trade off. We further find that, for a subset of firms whose API traffic we can observe, hack events cause an increase in testing and login authorization data flows, and decreases in internal communications, indicating that firms adapt their API use in order to manage exposure risk.

2 Theoretical Development and Hypotheses

2.1 The Inverted Firm

With the rise of information and communication technologies in the 1990s and Web 2.0 user-generated content in the early 2000s, many companies found themselves empowered with a vast new array of data and digital processes (O'Reilly, 2009). Companies could store, transmit, and process data at marginal costs unimaginably lower than before the digital revolution. These empowered companies then faced a question: how to monetize their newfound capabilities? Traditional approaches to profiting from a new resource or technological advantage, such as the resource-based view (Wernerfelt, 1984), focus on keeping the prized resource in a secure vertical stack so value cannot be imitated. Alternatively, the firm erects barriers to entry to sustain higher margins (Porter, 2008). This might be called a "pipeline" business model because the firm itself designs a product or service, produces it, and then sells it to an end consumer (Van Alstyne et al., 2016), adding value at each step of the value chain (Porter, 2001). This approach has the advantage of giving the firm maximum control, maintaining margins, and lessens the chance that a key competitive advantage falls into the hands of rivals.

While numerous firms have taken that route, the inverted firm takes an alternate approach (Parker et al., 2016; Van Alstyne et al., 2016) seeking to create an external ecosystem of partners and complementors. In a successful platform ecosystem, different types of users – small-scale outside developers, other large firms, or consumers – connect with resources provided by the platform and to each other. In the process, these outsiders can create profitable businesses that rely on the focal firm's resources and produce valuable complements that enhance its value to ordinary users (Parker et al., 2016).

Anecdotally, the inverted business model dominates the market. In 2020, seven of the top ten firms by market value were platforms.[2] Sampling from the Forbes Global 2000, platform firms compared to industry controls had much higher market values ($21,726 M vs. $8,243 M), much higher margins (21% vs. 12%), but only half the employees (9,872 vs. 19,000) (Cusumano et al., 2020).

But why should platform firms be more successful than pipeline firms? Perhaps the most important reasons are that, for many digital products, network effects are important, and marginal costs of digital reuse and adding users are low. If the value of a product grows strongly in users and complementors, while marginal costs remain low, then a firm should increase scale as quickly as possible. By contrast, internal growth mechanisms, such as investing in capital, hiring employees, pursuing new markets, and conducting R&D, can face large delays and adjustment costs. These investment options pose challenging financial trade offs. Yet, even if the firm faced no adjustment costs, there are outsiders – lead users (Von Hippel, 1986), employees of other firms (Jacobides et al., 2018), or outside developers (Parker and Van Alstyne, 2018) – who might be unknown to the firm or otherwise not available for hire. More people with good ideas *always* exist outside a firm than inside it[3] and these outsiders can be interested in using the firm's resources to further their own ends. Theory suggests that if the potential of third party complementors is large enough, then the structure of the firm shifts: Inverting a firm and taking a small share of the vastly larger surplus created becomes the profit maximizing strategy (Parker et al., 2017).

2.2 Public APIs and the Inverted Firm

The technical difficulty in creating an inverted firm is finding the right way to externalize internal resources. Ideally, the method would be modular, recombinable, and permissionless yet meterable. A modular sharing system will be more robust to unanticipated shocks, allowing third parties to trust that the source will be reliable. Modularity also contributes to recombination. Reuse, or combinatoric innovation, is the ability remix data, software or services in surprising and value creating ways (Weitzman, 1998; Baldwin and Clark, 2000). Finally, the ways that inverted firms share their data must be permissionless yet excludable. Negotiating access rights to digital services is one of the most important adjustment costs for any digital firm. For an inverted firm to succeed, third-party developers must have permission to experiment with and profit from using the inverted firm's resources. Developers can also prefer not to disclose their own innovation plans for fear of misappropriation (Chesbrough and Van Alstyne, 2015). At the same time, the focal firm needs a way to meter outsider access, to guard against malfeasance,

[2]These firms are Apple (1), Microsoft (2), Amazon (3), Alphabet (4), Meta (5), Tencent (6) and Alibaba (8). Source: Wikipedia accessed Jan. 19, 2021.

[3]This insight is codified as Joy's Law, which states "no matter who you are, most of the smartest people work for someone else" (Lakhani and Panetta, 2007).

and to monetize their most successful complementors.

APIs have all these characteristics. An API is a set of routines, protocols, and tools that standardizes building software applications compatible with an associated program or database (Ofoeda et al., 2019). APIs are code that control access to information. They can also be thought of as contracts (Jacobson et al., 2011). They govern the type and format of calls or communications that any application can make of another associated program. The answering program is agnostic about the source of the call, yet can enforce access permission, and the calling program need not know anything about the internal workings of the answering program.

APIs simplify the writing and operation of programs that communicate with online services and shared databases. They are essential for powering such systems as Google's documents and maps, Amazon's voice and web services, Apple's online market, Wallgreens' photo print, Nike's fitness trackers, and Facebook's authentication services. They mediate economic transactions. Their value is not only determined by the actions of their creators but also by the habits of their users and the strategic choices of third parties who connect systems and reuse components in unanticipated ways (Von Hippel, 1989).

APIs can be public or private. It is the public APIs, ones that can be accessed permissionlessly by third-parties, that are essential to an inverted firm. A public API is an externalization and modularization of one's technology stack. What was previously a black box of technology capability from end to end is now available to others in easily understood and recombined modules (Zachariadis and Ozcan, 2017). In the language of a seminal paper on open innovation: From the outside-in perspective, APIs enable others to build easily upon a firm's stack, and foster a vibrant ecosystem of complementary actors on top of the platform. From the inside-out perspective, public APIs allow others to repurpose parts of the firm's stack, providing alternative revenue streams, broadening adoption of their technology, and potentially reducing internal costs for sustaining that piece of stack over time (Enkel et al., 2009).

Indeed, anecdotally, there is a strong historical relationship between APIs and inverted, platform firms. APIs only fully came into their own in the Internet era.[4] Many web-pioneers featured APIs as core to their businesses. Salesforce.com included them in their 2000 launch of the world's first 'software-as-a-service' product. Likewise, eBay launched a developer program in 2000 to a select group of partners, encouraging them to create services that drew information from eBay's API. Having created one of the first popular open APIs, eBay's decision led to a virtuous cycle of better tools, higher visibility, and more customers. Perhaps the most iconic effort to place APIs at the center of a firm's strategy was Bezos' 'Big Mandate' of 2002. Frustrated by the haphazard way Amazon solved its digital challenges, and hoping to turn hard won

[4]It is not clear when the first API was created, but they clearly predate the Internet. Google's n-gram tool lists usage of the phrase 'application programming interface' as early as 1961.

lessons into new sources of revenue, he demanded, among other things, that:

- All teams will henceforth expose their data and functionality through service interfaces...
- There will be no other form of interprocess communication allowed: no direct linking, no direct reads of another team's data store, no shared-memory model, no back-doors whatsoever. The only communication allowed is via service interface calls over the network.
- All service interfaces, without exception, must be designed from the ground up to be externalizable. That is to say, the team must plan and design to be able to expose the interface to developers in the outside world. No exceptions. (Rowan, 2011, citing Yegge)

How was it that a book seller came to be the world's largest web services provider? In *Working Backwards* (2021), Bryar and Carr give an insider's answer to that question. Amazon launched the "Amazon Product [Advertising] API" in 2002. This tool allowed outsiders to build links to Amazon product listings into their apps and websites. Announcing the project launch, Jeff Bezos remarked "We're putting out a welcome mat for developers—this is an important beginning and new direction for us... Developers can now incorporate Amazon.com content and features directly onto their own websites. We can't wait to see how they're going to surprise us." The program attracted over 25,000 users in the first year. One of the biggest surprises was that internal Amazon developers often preferred using resources from the public API to Amazon's internal tools. The success of the product API led Amazon management to consider other internal strengths they could externalize and monetize, such as data storage and messaging. Amazon launched the Amazon S3 API to provide an inexpensive simple storage solution. Amazon's EC2 API, providing elastic cloud computing, quickly followed.

APIs brought results. By 2013, Amazon's marketplace featured more than two million third party sellers, accounting for roughly 40% of total sales. In 2020, Amazon Web Services, including S3 storage and EC2 computing, earned over $46 billion in revenue (Furrier, 2020). Using partner sales data, Amazon has also moved to vertically integrate into 3% of its partners' top selling products (Zhu and Liu, 2018). Amazon's market capitalization has duly expanded. Bezos' gamble that there was more money in managing bytes than managing books succeeded handsomely.

2.3 Hypotheses

In this paper, we seek to test the hypothesis that the firm inversion strategy, as enabled by public APIs, has been a major driver of market value growth for US publicly traded companies. This paper is the first to empirically test this hypothesis. Key elements of this hypothesis are that public firms benefit from sharing data and digital services through public APIs, and that the magnitude of this benefit is increasing in the number of third-party developers that the firm attracts, and in the centrality of the firm's public APIs in the digital ecosystem.

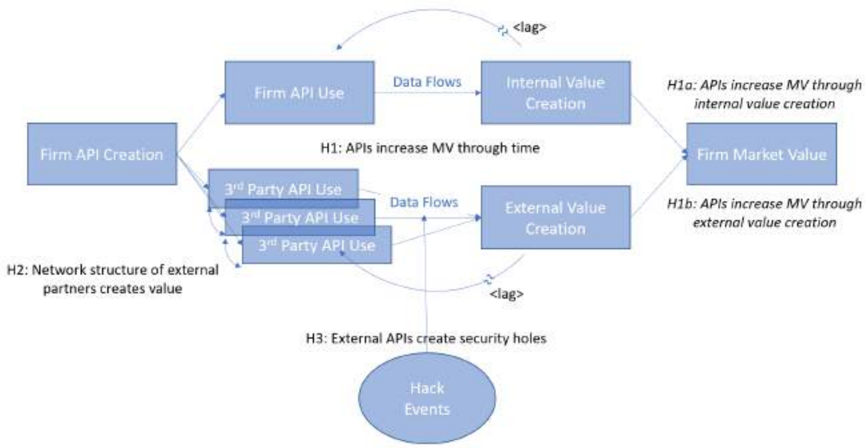

Figure 1: Directed acyclic graph illustrating our hypotheses and possible causal relationships. The path involving 3^{rd} parties is non-exclusive and can potentially create much more value. Thus engaging others might create more value than working alone, leading to an inverted firm.

At the highest level, our hypothesis is: "Firms increase their market value by building a network of outside complementors. Firms more central in API networks capture more value." This theory being correct entails several empirical observations. We would expect to see public API adopting firms increase their market value, and that the size of this gain should be related to the network of third-party developers attracted. Importantly, the benefit of inversion is not simply a one time step function but a growth benefit that increases over time as usage creates data, which creates value, which attracts partners, which begets usage. Further, inverted firms should manifest their market value growth not simply by scaling their assets, but by capturing value from these outside partners. Finally, firms with successful public APIs should benefit more than firms implementing private APIs.

This high-level hypothesis leads us to three specific hypotheses that we empirically test in the remainder of the paper. Figure 1 summarizes these hypotheses in a directed acyclic graph.

H1: Implementing APIs increases firm market value over time

Our first hypothesis is that API adopting firms will see increases in market value that may exhibit lags between technology implementation and growth realization. We have much anecdotal evidence on the success of inverted firms, and therefore expect APIs, the key enabler of this strategy, to lead to positive outcomes. But, there are also reasons to believe that this IT investment will not work. Many IT adoptions fail to deliver. A 2019 Forrester analyst report to CIOs observes that, on average, IT investments have led to stagnation (Bartels, 2019). US productivity growth plateaued at 1% after 2010, yet IT investments rose at a rate of 5% over that same period.

There are several reasons why API investment might have limited to no effect. First, internally, managers should invest in *any* asset, not just APIs, up to the point where marginal benefit equals marginal cost. APIs have existed since at least 1961, suggesting that consequences from new investment might be strictly marginal. Market capitalization, in particular, might change little as it aggregates across all firm activities. Second, externally, if firms are observed gaining advantage from APIs, then competitors should also invest and compete away that advantage. Competitors' investments restore a balance of normal profits. Third, external developers are not employees. Firms that open APIs often have no idea who the developers are. If developers choose not to engage with or use the APIs, then no external value is created. Voluntary third-party investment that never materializes cannot drive value. Fourth, empirical research suggests that IT investments frequently fail to deliver promised productivity gains. The 2019 analyst report to CIOs that highlighted stagnant growth from IT investment also noted that, at the sector level, the relationship between IT investment and growth was often negative (Bartels, 2019). Overlapping our research window, that study impugns any notion that investments in digital transformation, like those in electronic data interchange (EDI), enterprise research planning (ERP), customer relationship management (CRM), electronic health records (EHR) and others, unconditionally deliver positive outcomes. Forrester's conclusion is consistent with recent (Brynjolfsson et al., 2018) and early (Brynjolfsson et al., 2002) academic research that IT investments alone can produce negligible or even periods of negative value. To create value, they need to be coupled with complementary investments in organizational capital and other intangible assets. Absent complementary investments in new processes, products and business models, we should not expect observable changes in market value. IT investments have a history of not affecting aggregate market value (Tam, 1998).

We theorize that APIs may drive value through internal and external mechanisms. By external mechanisms, we mean the inverted firm hypothesis: Third-party value creation that expands the boundaries of the firm. By internal mechanisms we mean private value creation that does not expand the boundaries of the firm. APIs may create internal value and drive profits through new products or new sales channels, as in the case of reaching customers via mobile phones (Iyer and Henderson, 2010). Additionally, APIs grant firms metered control over outside access and the ability to capture new data. This can help firms price discriminate among existing products while enabling new kinds of digital services (Tiwana et al., 2010). APIs are more modular than traditional code, potentially increasing efficiency through data and software access, reuse, and recombination within the firm (Yoo et al., 2012; Baldwin and Clark, 2000). They facilitate deprecation of old technology (Jacobson et al., 2011). The potential to remix resources in new ways creates option value (Baldwin and Clark, 2006). This expands a firm's dynamic capabilities by providing low cost variation and selection of business routines (Teece, 1988; Eisenhardt and Martin, 2000). They facilitate the remixing of disconnected resources or pockets of expertise (Purvis et al., 2001), integration of new software into legacy software

(Joseph et al., 2016), and speed IT deployment (Iyer and Subramanian, 2015). They help firms raise labor productivity for a given expenditure on programmers (Brynjolfsson and Hitt, 2000). Thus, one of the main theories supporting APIs is their ability to lower adjustment costs. Summarizing leads to the following hypothesis:

H1a: Implementing APIs increases market value via internal value creation.

While it is possible that APIs increase firm value through internal mechanisms, our core hypothesis is that public APIs enable an inverted firm strategy that benefits the firm through the expansion of firm boundaries. Public APIs facilitate development of third-party complements (Parker et al., 2017). APIs differ in important ways from earlier outsourcing, back-office, and front-office technologies such as EDI, ERP, CRM, and EHR. These legacy technologies targeted internal employees or *known* contractors. By contrast, pubic APIs specifically emphasize permissionless innovation by *unknown* partners, who generate uses of digital assets of which the firm never conceived (Thierer, 2016; Chesbrough and Van Alstyne, 2015; Parker et al., 2017). Salient illustrations of this external value add include the numerous apps sold by Apple, Amazon, and Google but that were never conceived by these platforms themselves. More permissive licensing, which is enabled by APIs, has been shown to increase complementary device development among handset manufacturers (Boudreau, 2010). Platform banks have opened APIs as a means to invite FinTechs to offer complementary banking services (Zachariadis and Ozcan, 2017). This leads us to our next hypothesis:

H1b: Implementing APIs increases market value via external value creation.

One approach we take to distinguishing between **H1a** and **H1b** is based on standard theories in finance. If APIs boost internal efficiency and make it easier to repurpose capital, then it will boost firms' investments. This would show up in the data as a decrease in Q – the ratio of market value to installed capital. Alternatively, if APIs primarily boost value by 'inverting the firm,' and causing third-parties to make investments, then the portion of firm value not explained by its capital stock will increase as a function of API adoption. Another way we distinguish between these hypotheses includes splitting the sample into public vs. private APIs, and by investigating the relationship between market value growth and third -party engagement.

Under the inverted firm hypothesis public APIs do not automatically create value for their host firm. Rather, only public APIs that nurture a rich ecosystem of third-party developers will create value. APIs are more than technical plumbing designed to decrease transaction costs or increase efficiency. They *enable* markets. The consequence is not merely a shift from hierarchies to markets or a shift in the 'make-vs-buy' decision (Malone et al., 1987). Instead of entering the market as a more efficient player, the focal firm *becomes* a market, an orchestrator of other firms' transactions. Orchestrating a market gives the platform visibility into the data passing through its systems, which provides insights into competitors' activities, margins, and opportunities

(Khan, 2017). This yields a strategic information asymmetry that favors the platform sponsor at the expense of the platform partner (Zhu and Liu, 2018). This advantage has risen to the point of anti-trust scrutiny(Schulze, 2019; Cabral et al., 2021). The strategy of using APIs to orchestrate third-party value creation which the focal firm can then monetize is the key element of the inverted firm strategy (Parker et al., 2017) where value creation shifts from inside to outside. This shift is reflected in the following hypothesis:

H2: The network structure of applications that call APIs affects the market value of firms that implement them. Firms with higher API network centrality, more connections, and larger effective network sizes have higher market value.

Once a firm opens to third parties, the opportunity for interactions among those parties creates new avenues for value creation and value capture. We hypothesize that firms with APIs that are more central to the network of data flows will see increased market value, and that this is in part due to being able to capture a larger share of the surplus from the digital economy.

The insight that network structure influences the resources available to parties embedded in that structure underpins a vast literature spanning decades of research (Simmel, 1922; Moreno and Jennings, 1938; Granovetter, 1973; Baker, 1990; Burt, 1992, 2009; Padgett and Ansell, 1993; Uzzi, 1997; Hansen, 1999; Podolny, 2001; Reagans and Zuckerman, 2001; Aral and Van Alstyne, 2011). The central argument is that structurally diverse networks provide access to diverse resources. APIs can confer gatekeeper power. Controlling the bottlenecks in that structure provides the means to broker opportunities (Granovetter, 1973; Burt, 1992), improve decisions (Hansen, 1999), resolve uncertainty (Podolny, 2001), boost productivity (Aral and Van Alstyne, 2011), innovate (Reagans and Zuckerman, 2001), and extract rents from the digital economy (Burt, 2009). Key measures of structural position include betweenness centrality, which measures the frequency of being on a shortest path (Borgatti, 2005), and effective size, which measures diversity (non-redundancy) and reach among network contacts (Burt, 2009). If APIs provide orchestration and innovation benefits, why might firms fail to adopt them? One reason for reluctance is the fear that malicious actors may pose as legitimate users and steal a firm's sensitive data. APIs can facilitate illegitimate access and increase the risk of data breach. This leads to our final hypothesis:

H3: Implementing external APIs can create security holes, increasing the risk of data breach.

There are considerable downside risks to allowing third parties access to a firm's private data. Notable data breaches tied to API flaws are numerous (Gates, 2019). APIs were implicated in a hack that released compromising and very private photos of celebrities stored on Apple's iCloud (Berlind, 2015). Security holes due to APIs have been a particular concern in open banking (Zachariadis and Ozcan, 2017). Another API vulnerability allowed use of nothing more than a license plate to breach an insurance company and learn all movements of a car, its position in real time, and its owner's name (Scarpino, 2017). The CEO, CIO, and CSO of credit scoring bureau Equifax all resigned after an API hack released the personally identifiable information of 143 million people (Gates, 2019). Losses due to this hack reached more than $1.6 billion by 2019 (Lane, 2020). T-Mobile announced an API data breach had exposed private data of more than 2.3 million users (Spring, 2018). Google shut down Google+, its much maligned social networking venture, after revealing that the private data of more than 52 million users had been exposed to third parties through its APIs (Newman, 2018). Poor choices by API suppliers can also compromise data. One suite of Microsoft APIs remained open by default, inadvertently exposing 38 million records across dozens of firms and government bodies (Sundstrom, 2021). The editor in chief of ProgrammableWeb observes that API security "is so hard that even the biggest companies with the deepest pockets to hire the best talent make mistakes" (Berlind, 2017).

Each of these breaches illustrates a "leaky API," one that is vulnerable to hacking, misuse, or unintended disclosures because third parties are not properly metered or controlled when they request data. Open systems are more susceptible to hacking. Ransbotham (2016) finds that open source software - while often functionally superior - is more likely than closed source software to have zero-day exploits and obvious avenues of attack.

Cyberattacks have negative financial consequences for firms and their CEOs. Kamiya et al. (2018) find that cyberattacks are associated with reductions in sales growth, investment, and stock market performance. They reduce CEO bonuses. Makridis and Dean (2018) match data breach reports from the Privacy Rights Clearinghouse Compustat financial data and find a 10% rise in records breached is associated with a .2% fall in firm productivity. Spanos and Angelis (2016) perform a systematic literature review of the impact of information security events on stock market outcomes. They review 45 studies from 37 papers. Over 75% of these studies find a statistically significant effect of digital security events on stock prices.

3 Data

Our paper draws on four main sources of data. These are: (1) Compustat data on finances of publicly traded firms; (2) data on public APIs and their connections to third-party apps (mashups) from the ProgrammableWeb crowdsourced directory; (3) the Privacy Rights Clearinghouse for data breach events as matched to Compustat by Rosati and Lynn (2021); (4) proprietary data on internal API usage from a private provider of API creation tools.

3.1 Financial Outcomes

Firms' financial performance is provided by Compustat, which measures market capitalization and other covariates at the quarterly level. Our sample runs from Q1 2007 through Q3 2020.

3.2 Public APIs

Our main data source on firms' API usage comes from the ProgrammableWeb a crowdsourced database of public APIs and the apps that call them. Data used for this analysis was collected in winter 2020. APIs were categorized and matched to the firms that sponsored them by a team of research assistants and checked by the authors. ProgrammableWeb also has data on apps calling one or more APIs, called 'mashups' by ProgrammableWeb, emphasizing the role they play in recombining information from disparate sources. Submitters label these apps with various tags useful for categorization.

ProgrammableWeb data include the dates an API was first submitted and the list of apps calling that API. We also collect the number of users who express interest in an API (followers) and those who claim to work on applications using that API (developers), as well as the number of updates the API has undergone. All APIs with at least 15 followers, of which there were 3402, were matched to the firms that own them. The majority, 63.1% were associated with non-public for-profit companies while 19.6% of these APIs were associated with publicly traded firms, 3.2% were associated with governments, and 8.1% were associated with non-profit organizations.[5] Of the 206,411 follows of APIs with at least 15 followers, 33% are of APIs created by public firms. Firms with APIs tend to have higher market value than firms without APIs. We further categorized apps as primarily B2B, B2C, both, or unclassifiable. API orientation is roughly split between B2B and B2C APIs (APIs classified as 'both' are associated with both categories for the purpose of summary statistics and regressions).

Matching ProgrammableWeb data to Compustat allows us to categorize the firms which use APIs by industry. We observe broad trends over time as reported in Figure 2, which plots the fraction of firms by one digit SIC code that have at least one public API over the sample period. APIs grew across all industries, with services and transportation & public utilities growing the fastest. By the end of the sample period, roughly 3.5% of firms matched to Compustat in the services sector have public APIs , and 3% of firms in transportation & public utilities have public APIs. Figure A1 reports the fraction of firms with at least one API by two digit SIC code circa Q3 2020. We see that air transportation firms are the most likely to have public facing APIs, followed by firms in apparel, business materials, business services and miscellaneous manufacturing industries.

[5]We have matched many of these non-public firms to Crunchbase information on startups. This is an intriguing database for future research on the financial impact of APIs, and strategic entrepreneurship.

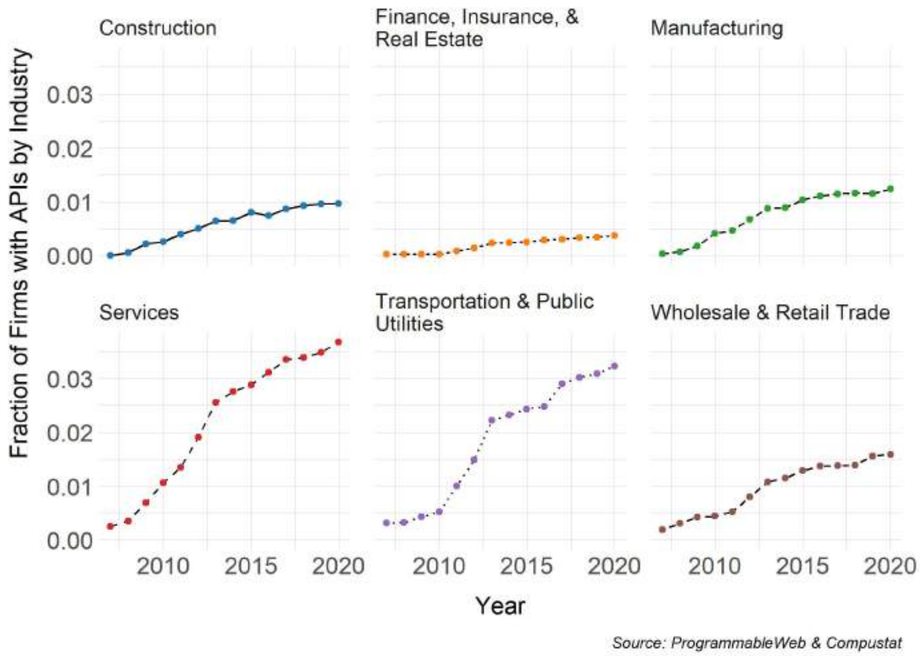

Figure 2: Fraction of firms with APIs by One Digit SIC Code.

3.2.1 API Network Statistics

Using ProgrammableWeb's list of of apps that connect to public APIs we trace out the network of APIs connecting firms. For the subset of firms whose APIs connect to this network, we compute a series of network statistics. Each node corresponds to a firm's API and each link connects a firm's APIs via third-party applications.

We compute three network statistics for a given API: betweenness centrality (White and Borgatti, 1994), degree (Diestel, 2005), and effective network size (Burt, 1992). Betweenness centrality calculates the share of shortest paths between nodes in the network that pass through a given node. Degree sums the total number of connections between a node and other nodes. Effective network size captures non-redundant connections that each node provides. This measure calculates effective network size as number of connections minus redundant connections between nodes. Network theorists refer to the latter measure as capturing a measure of 'structural holes' within a network (Burt, 2009). Nodes that rank high in effective network size act as structural bridges between sections of the network. Firms may operate several APIs, thus we calculate averages, maximums, and sums of these network statistics across all the APIs a given firm operates for each quarter in which it has at least one operational API. For ease of interpretation, these network statistics have been centered and scaled. For the topforty APIs by betweeness centrality, we also hand-collected data on whether and when the API experienced a shutdown or reduction in functionality. We found one API was completely shutdown, two were replaced with functionally similar APIs, and five experienced significant reductions in functionality. Appendix Table A1 reports summary statistics for the paired ProgrammableWeb, Privacy Rights Clearinghouse, Compustat data, and proprietary data from a consulting company that generates internal APIs.

3.3 Private APIs

We received proprietary data from a consulting firm that offers API development tools, implements APIs, and offers hosting services on behalf of API adopters. Many of the APIs are not published on ProgrammableWeb, as their use is restricted to actors within a firm. We matched this list of private APIs to our Compustat and ProgrammableWeb data. Some firms that operate public APIs also operate private APIs with the help of our API consulting company.

To measure the effect of purely internal APIs, we identify that subset of firms from this proprietary dataset who do not have any APIs reported on ProgrammableWeb. This subsample leaves only internal API use as the treatment. In this sample, private APIs are less popular than public APIs, with approximately 0.7% of firm-quarters representing internal API use. The bulk of APIs, however, are private (Jacobson et al., 2011). Data on API flows from this data set are summarized in Table A2.

The API management firm also provided us with monthly records of API use for 273 separate accounts. This includes the name of each API used, as well as the number of calls and bytes processed by each API in a given month. Data on calls processed by partner firms' APIs span December 2012 to September 2016. Data on bytes processed span December 2012 to May 2016. We designate the first date that we observe any call to any of a firm's APIs as the API adoption date. Appendix A2 reports the total number of APIs, API calls, and API bytes of data flow that we observe in each month. We have 2,453 firm-months of API usage data. The average firm has 160 million API calls in a given month, as well as 1.98 trillion bytes of data. The average firm in this source of API data has 31.4 APIs.

3.4 Breach Data

The *Privacy Rights Clearinghouse* (PRC) records public breach announcements as matched to Compustat firms by Rosati and Lynn (2021). We collect data for all public and private firms we observe using APIs from 2005 to 2015.[6] PRC distinguishes six different breach types: "PHYS", "PORT", and "STAT" events involve the theft of physical storage media, paper documents,

[6]Data from Rosati and Lynn (2021) end in 2015 so our main results focus on this period. Appendix Figure A24 restricts attention to the 78 firms for which we have flow data, extending the PRC data through 2016. However this only increases the number of data breaches observed by four.

and stationary devices. "INSD" events involve breach events from insiders as well as malicious outsiders who have compromised insider credentials. "DISC" events are unintended disclosures. "HACK" events are incidents of hacking or malware leading to the data breach.

4 The API Network

This section characterizes the evolution of the economy's API network. Figure 3 presents the API network as recorded in the ProgrammableWeb directory through Q3 2020. Nodes in this graph correspond to APIs. Edges connect APIs when an app calls both. Node colors correspond to the company associated with the API. Edges are colored according to the functionality of the app that calls them. For example, DeployPlace is an app, designed as a developer tool. It interacts with the Amazon S3 and Gmail APIs, among others. Therefore, there is at least one yellow-green line connecting these two APIs, indicating they are connected by a productivity focused API. Similarly, the ecomdash service, an app involved in eCommerce, calls both the Amazon Product Advertising and eBay APIs. This is visualized by at least one green edge connecting the two nodes.

Several phenomena emerge from visual inspection of Figure 3. First is the relative frequency of companies appearing in the API network. The prevalence of green and orange nodes indicates the network importance of Google/Alphabet and Amazon. Perhaps more surprising is the number of red nodes associated with Verizon/Yahoo. Facebook, Twitter and eBay are also central to the network but with many fewer nodes.

Appendix Exhibit A2 reports the company, degree, betweenness centrality and market capitalization growth for the topforty APIs by betweenness centrality. The top five APIs ranked by betweenness centrality (Google Maps, Twitter, YouTube, Facebook, and Flickr) are also the highest ranked in-terms of degree. Unsurprisingly, these five APIs are both extremely popular for app calls and also central to the API network. Google Maps provides essential navigation functionality to a wide variety of apps, Twitter and Facebook are go-to social media plugins, while YouTube and Flickr provide popular video and image hosting websites.

Lower on the list, we see that some APIs have centrality ranks much higher than their degree. Firms with high betweenness centrality, whatever their degree, play an important gate-keeping role in that sector of the data economy, which can offer profit opportunities. The API with the most extreme discrepancy between its degree and betweenness ranks is CoinBase, which can be seen in the top left of Figure 3. This API is called by apps that also call several cryptocurrency related APIs (e.g. the Mt. Gox API) and is also connected by apps to several online shopping APIs such as Google Checkout and PayPal. These edges are all related to eCommerce. Absent API connects to the core of the API network, many cryptocurrencies would be much harder to

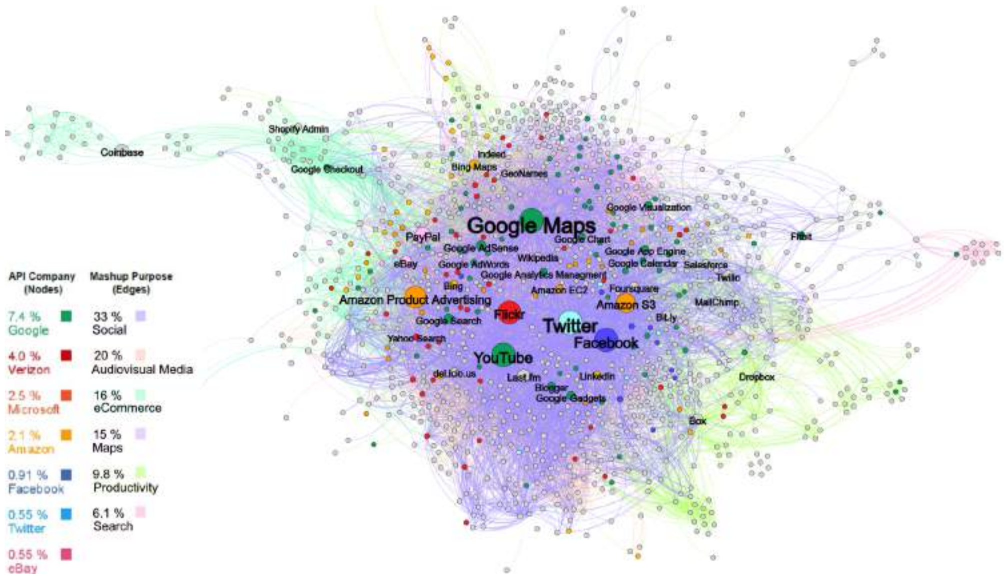

Figure 3: This figure visualizes the network of APIs and apps that connect them as of Q3 2020. Larger nodes indicate higher API centrality. The forty nodes with the highest betweenness centrality are labeled. Node colors represent API sponsors (see key). APIs from other firms are gray. Edges exist between any pair of APIs called by the same app. Edge color indicates the functionality of the app calling the APIs

use in actual transactions. The Shopify API plays a complementary role in the portion of the network devoted to eCommerce, and Dropbox plays a similar role at the nexus of productivity-oriented apps (right side of Figure 3). Sub-networks, organized by purpose, appear in Figure 4. Unsurprisingly, Google Maps is at the center of the mapping network, while YouTube and Flickr are more central in the Audiovisual Media network.

Firms with central APIs saw dramatic increases in market value over our sample period. The 14 publicly traded firms that have APIs ranked in the topforty by API betweenness centrality added \$6.584 trillion dollars to their market value from 2005 to 2021. This constituted a 580.8% increase in value for the seven firms that were publicly traded for that entire period. By comparison, the entire US stock market grew by \$16.89 trillion, or 99.3%, from 2005 to 2019.[7] The growth of the 14 firms at the center of the API network represent approximately a third of US market value growth over the time period under consideration.

Creating a top API without large growth in market value is rare. Most top APIs are governed by publicly traded companies. Only two of the topforty APIs by betweenness centrality (i.e. 5%) are governed by non-profits. These are GeoNames, a location directory, and Wikipedia, an online encyclopedia. Of all ProgrammableWeb APIs with at least 15 followers, 13.3% are produced by governments or non-profits, meaning that for-profit companies are over-represented in the creation of top APIs.

[7] See World Bank (2022). According to an alternate source, the total US equity market increased by 30 trillion in value from 2005 to 2021. See Siblis Research (2022).

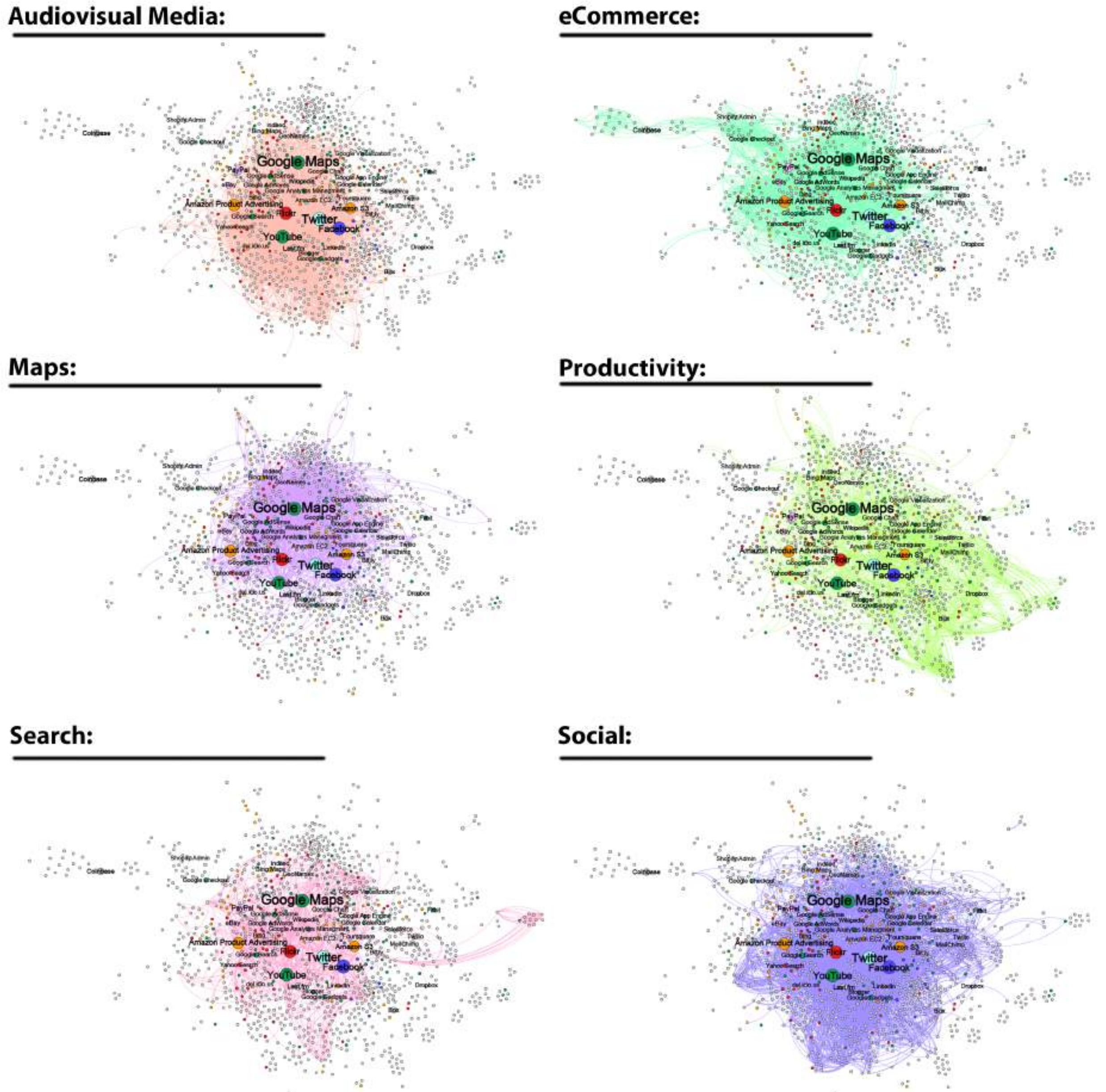

Figure 4: API subnetworks, with only edges of a certain type highlighted. Five notable features of the subnetworks include: (1) Some APIs are highly central to some subsets of the network, despite being of low degree. For example, the Coinbase API is highly central in the eCommerce subnetwork, despite having low degree (2) Dropbox, Box, Salesforce, and Amazon S3 are are more central in the 'productivity' subnetwork (3) Google Maps is central in many subnetworks, but especially the Maps subnetwork (4) Also important in the maps subnetwork is GeoNames, one of the most important non-profit supported APIs (5) Facebook, Twitter, and YouTube are especially central to the social media subnetwork, but they are also central in almost all subnetworks.

Several notable features stand out concerning API organization. For example, the consumer facing Social (periwinkle) and Search (pink) apps densely connect the heart of the API network. APIs connecting these apps, especially Facebook, Twitter, YouTube, and Google Search, might drive engagement for the apps connecting to them. B2C facing APIs may be better at driving network effects than B2B APIs because it is more immediately obvious to third parties how to incorporate consumer facing features into their apps.

Unsurprisingly, the most central APIs in the network are also associated with both search or social media. APIs for Facebook, Twitter, and YouTube are some of the most connected APIs. Dropbox, Box, Salesforce, and Amazon S3 are important to the productivity cluster, yet these also include mapping functionality and are close to Google Maps, Indeed, Bing Maps, and GeoNames. The eCommerce cluster shows high density around the Amazon Product API, as well as the PayPal and eBay APIs.

Appendix Figure A4, which labels nodes for all APIs owned by a given company, gives another view on how each company fits into the API network. Microsoft's and eBay's nodes are disproportionately located in the top left corner of the network, connected to each other and the cluster of eCommerce oriented APIs. Facebook's nodes are clustered in the bottom right of the figure, located in the heart of the social media sub-graph but also closer to the productivity portion of the graph. Apple, despite its huge success as a technology company, is relatively poorly represented, perhaps due to the closed nature of Apple's technological ecosystem.

Appendix Figures A5 through A17 visualize the growth of the API network over time. Network density increases substantially in the late 2000s and early 2010s, a period of time when the ProgrammableWeb crowdsourcing was most comprehensive. Note also the early centrality of Flickr, an important early image hosting website. While Flickr has since fallen on hard times (supplanted by Imgur and other close substitutes), it remains central to the API network as we measure it. This occurs because we do not observe deprecated APIs in our data when apps stop using them. Importantly, our measure of the API network at any point in time is cumulative and somewhat backward looking for this reason.

5 Market Value Changes among API Adopters

As shown in Section 4, firms with top APIs have seen tremendous increases in market value over the last fifteen years. This section applies two-way fixed effect, difference-in-difference, and synthetic control approaches to estimate the impact of API adoption on a firm's market value, evaluating hypothesis **H1**.

We begin by estimating specification (1)

	All Firms	Excluding Top 20 Firms with Most Popular APIs	Excluding Industries Where < 1% of Firms Have APIs	Excluding Any Computer Services Firm	Year API Open < 2012	Year API Open ⩾ 2012
Post x API	0.387***	0.370***	0.377***	0.322***	0.750***	0.260**
	(0.0855)	(0.0860)	(0.0856)	(0.0971)	(0.140)	(0.0997)
R2 Adjusted	0.932	0.931	0.932	0.934	0.929	0.929
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs	133202	132446	127796	119201	129670	130040
Firms	4647	4627	4478	4060	4556	4561
API Adopters	177	157	176	100	86	91

Notes: Standard errors in parentheses and clustered at the firm level. Outcome variable is the log market value of the firm. *PostxAPI* is a binary variable that equals one if a given firm has a public API operating on a given date. Top 20 firms excluded include Alphabet, Amazon, Apple, Ebay, Facebook, FedEx, Groupon, Liberty Expedia, Microsoft, New York Times, PayPal, Pinterest, Salesforce, Spotify, Twilio, Twitter, Uber, UPS, Verizon, and Zillow. Excluded computer services industry refers to firms in SIC code 7370 "Computer and data processing services" + p < 0.10 * p<0.05 ** p<0.01 *** p<0.001

Table 1: Two-way fixed effect estimations of the effect of API adoption on log Market Value following equation (1). Column one includes the entire dataset, while the subsequent columns restrict the regressions to various subsets of the data.

$$\log \text{Market Value}_{i,t} = \beta \cdot \text{API}_{i,t} + \alpha_i + \gamma_t + \epsilon_{i,t} \tag{1}$$

where 'API' is an indicator for whether firm i in period t has an operating API and α and γ correspond to firm and quarter fixed effects. We evaluate (1) for various subsets of public firms. These specifications focus on firms with public APIs, so we use the first date a firm's APIs are submitted to ProgrammableWeb to proxy when the firm initiated a public API strategy.

Table 1 reports the coefficient on Post-API adoption using this specification. In the full sample, API adoption is associated with a 38.7% increase in market value. One challenge in understanding this result is determining whether some small subset of firms or industries are driving the results. Thus, the other columns in the table estimate the effect of API adoption on subsets of firms.

The first subset, shown in column two, excludes the top 20 firms with the most popular APIs as measured by number of followers on ProgrammableWeb. We see the estimated coefficient moves down slightly to 0.37 but remains highly statistically significant. The second subset, shown in column three, excludes firms in industries where less than 1% of firms operate APIs. The concern here is that the comparison group includes firms that could not adopt APIs. We see the estimated coefficient decreases marginally to 0.377 but remains significant. The fourth subset, in column five, excludes any firm classified as in the "computer and data services" industry (SIC code 7370), the concern being that only firms with a high degree of complementarities to APIs stand to benefit from them. We see the estimated coefficient attenuate to 0.322 but remain statistically meaningful. Finally, the last two columns separate the sample into firms which first operate APIs prior to and after 2012. This would be a concern if there was a significant first mover advantage to the API network which has subsequently been saturated. We see that firms opening APIs earlier stood to benefit more. The estimated coefficient is 0.75 when excluding APIs opened prior to 2012, and 0.26 when excluding firms opening APIs before then. Both estimates are statistically significant.

To further demonstrate that benefits from APIs are not restricted to the largest firms with greater market values, appendix Table A3 estimates a quantile regression model of the baseline model 1, showing the estimated impact of API adoption on different quantiles of firm market value. We estimate that firms at the 10th percentile of market value gain 42.5% in market value from adopting APIs, and firms in the 90th percentile of market value gain 34.9% of market value following API adoption.

An important concern about difference-in-difference estimates of this form is that if API adoption has an effect on market value growth *rates* rather than *levels*, the estimate of the effect of API adoption will be highly sensitive to the length of the sample. Still, most of the effect we identify from API adoption is coming from across-firm decisions, rather than within-firm timing. Appendix Table A4 reports a Bacon decomposition of our column 1 estimate, and finds that 95% of our effect is identified from different decisions to adopt across firms.

Because API adoption seems to have an effect on market value growth rates rather than levels, it makes sense to re-analyze our results, separately estimating the effect of API adoption by number of periods since the 'treatment' began. This approach also lets us analyze whether there are pre-trends in the data. We therefore estimate specification (2)

$$\log \text{Market Value}_{i,t} = \sum_k \beta_k \text{API}_{i,k,t} + \alpha_i + \gamma_t + \epsilon_{i,t} \tag{2}$$

where k corresponds to the number of periods before or after a firm started using APIs. While our regression specification includes all leads and lags k for all observed quarters before and after API adoption, here we report only coefficients for eight quarters immediately preceding and post-API adoption. Figure 5 reports these estimates along with 95% confidence intervals.

Figure 5 shows that firms adopting APIs saw elevated market value growth beginning soon after adoption and significant growth seven periods, or 1.75 years, after ProgrammableWeb received their first API. Eight quarters after adoption, firms have 12.9% higher market values, a considerable effect. For a \$1B firm, this represents a \$129M increase in value. Since a typical

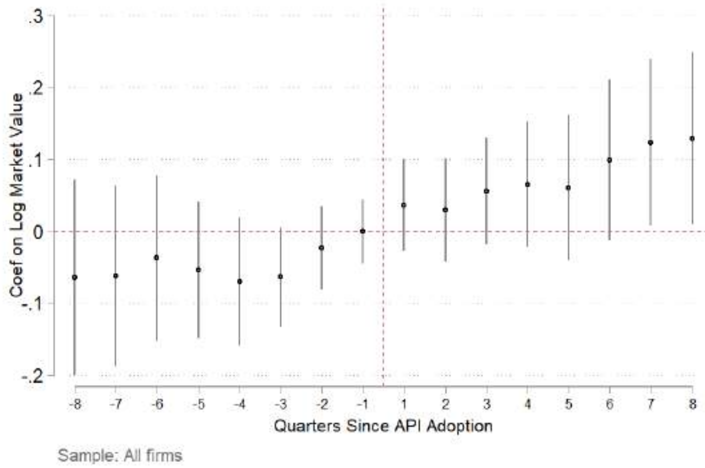

Figure 5: Treatment effect of public API adoption on log market value by quarters since API adoption with 95% confidence intervals. Standard errors clustered at the firm level. Appendix Table A5 reports these same estimates in table form.

API enterprise implementation costs \$250k,[8] this increase, which already reflects expenditures, represents a $\frac{129}{.25} = 516x$ return on investment within two years. Splitting the sample into firms with at least one B2C oriented API vs. those with B2B APIs, as appendix Figure A18 does, shows that the effect is driven by B2C oriented firms. This is consistent with our finding above that the most important and central APIs tend to be B2C or 'both' oriented (e.g. all APIs in the top 5 by betweenness centrality in appendix exhibit A2, are B2C or 'both' oriented).

There is some slight visual evidence of a pre-trend in API adoption beginning half a year before the API announcement date. We believe that this is due to anticipatory market value effects (the stock market can bid up the price of a company before a new technology is implemented) and that some APIs are only posted to ProgrammableWeb after a lag. A lag in posting is certainly consistent with ProgrammableWeb's nature as a crowdsourced dataset. Developers who can take early advantage of a new API may be in the best position to post on ProgrammableWeb after using them. Their private knowledge might motivate these individuals to push information about the new API after they had a chance to exploit that knowledge (Hirshleifer, 1978).[9]

[8] Based on private communication with the API consulting firm and independently confirmed for a different vendor: How much does Mulesoft cost?

[9] To confirm that delayed submission to ProgrammableWeb is a possible source of the pre-trend, we hand-collected data on ProgrammableWeb submission dates, online article reference dates, and official release dates of the top 40 APIs by betweenness centrality (those listed in A2). For the 22 APIs where an official launch date could be explicitly determined, the median delay between API launch and ProgrammableWeb submission was zero months, and the average delay was 5.7 months. For an additional 10 APIs, we could not find an official launch date, and also could find no evidence of the API being mentioned by a source before the ProgrammableWeb submission date. We conclude that while the ProgrammableWeb data set is the best and most complete one available, occasional lags in ProgrammableWeb submission after API launch likely generates the ex-ante effect seen in the event study. Further, if we shift the treatment indicator one or two quarters forward, the coefficient for our estimated baseline declines

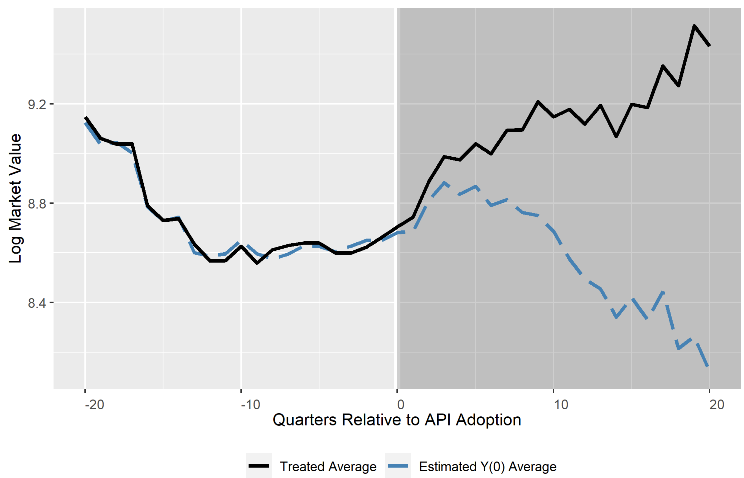

Figure 6: Average market values for API adopting firms and a synthetic control group balanced to match in the twenty quarters before and after adoption. Implied gap is $8.4B 20 quarters after adoption.

Still, this result may lead to concern that our analysis faces a reverse causality problem – in other words, that market value growth causes API adoption rather than vice versa.[10]

As a first approach to addressing this concern, we conduct a synthetic control analysis of API adoption. Synthetic control analysis creates composite firms of API non-adopters with the same pre-API adoption market value growth trend as adopters. If the API adopters and synthetic non-API adopting firm have different outcomes post adoption, then the differential is plausibly attributed to adoption itself and not reverse causality.

Figure 6 reports average log market values for API adopting firms and a composite of synthetic controls for the twenty periods before and after API adoption. The synthetic control was constructed following Xu (2017)'s generalized synthetic control procedure, based on data for API adopters on only the eight quarters prior to adoption. Despite this, both the treated average and synthetic control are virtually identical for the twenty periods prior to adoption. Both the average adopter and synthetic control firm see an increase in market value from about two periods prior to adoption to two periods post adoption. However, after that point, the market value of adopting firms continues to grow rapidly while the synthetic control firms see a large decrease in market capitalization. The non-adopters see a decrease in market value, which is consistent with a business stealing effect to the advantage of the adopters. A further managerial interpretation is that success in API networks might be due to early mover advantage.

Table 2 reports the point estimate and confidence interval for the effect of API adoption,

from 0.387 in Table 1 to 0.381 and 0.378 respectively yet remains statistically significant.

[10] It is important to note that differences in the market value of API adopting and non-adopting firms do not bias our estimates. By including firm-fixed effects, we control for non-time varying latent factors that might drive both market value and API adoption. These firm fixed effects make our regression analysis a type of difference-in-difference analysis, which do not identify off of the level of the outcome variable (Roth et al., 2022).

Average Treatment on Treated	Std. Err.	CI lower	CI upper	p-value
0.729	0.334	0.075	1.384	0.029

Table 2: Estimated average treatment effect and confidence interval using generalized synthetic control. Appendix Figure A19 reports confidence intervals for the treatment effect in each quarter.

again using generalized synthetic control following Xu (2017). This result should be contrasted with the basic difference in difference result in column 1 of Table 1. The point estimate of the effect is larger than in the baseline estimate and significant at the 5% level.

As another way to address concerns about a pre-trend, we perform power calculations of the pre-trend test and examine the biasing effect of possible violations of the parallel trends assumption following Roth (forthcoming). The pre-trends test is shown in appendix Figure A22. The red line constitutes a hypothetical linear trend difference between the treated and control groups, illustrating a potential violation of the parallel trends assumption. It represents the largest linear parallel trend violation we would fail to detect with 80% power. The black dots display our estimated time-varying effect of API adoption coefficients from Figure 5 and model equation 2. Finally, the blue dots correspond to the hypothetical estimates we would expect to see if there was only the red line linear trend difference between the treated and control groups. Equivalently, it displays the estimates we would expect to find if we incorrectly failed to detect the hypothesized non-parallel trend. Our estimated model coefficients, shown in black, are well above the coefficients in the post-period that we would estimate if there was only this linear-trend difference between the treated and control groups. This suggests that our large treatment effect estimates are not solely caused by a violation of the parallel trend assumption.

With evidence in hand that API adopting firms outperform non-adopters, we proceed to investigating the importance of different proposed mechanisms for APIs' positive impact.

6 Why APIs Matter: Inverted Firm or Internal Effects?

In our hypotheses and literature review, we pointed to two main classes of mechanisms by which API adoption might help firms. Hypothesis **H1a** is that they do so through internal productivity effects. Hypothesis **H1b** is that they do so through inverting the firm. If the inverted firm hypothesis is true, then it is likely that firms that are more central to the public API network especially benefit from it – this is our hypothesis **H2**. We begin by considering evidence for **H1b** and **H2**, and then return to the question of whether APIs also have internal productivity effects - **H1a**.

6.1 Evidence for Firm Inversion

In this subsection, we evaluate hypothesis **H2**: that the network structure of applications that call APIs affects the market value of firms that implement them. Under the inverted firm hypothesis, greater intense external use of a firm's resources should correspond to more value creation and capture from the API network. APIs that are more frequently integrated into apps should create more value. Firms with APIs that are more central to the API network may be better placed to capture the surplus created by the digital economy.

Our first evidence for this hypothesis is cross-sectional data on the growth rates of API adopting firms. Figure 7 plots percentage growth in the firm's market value as a function of the firm's rank in number of connections in the API network (i.e. the sum of a firm's API's degrees in Figure 3 as well as by the centrality of their most central API in the most recent API network. The figure restricts attention to the 67 firms that have data available for Q3 2020 and have at least one API which is connected to another API. There is a significant positive relationship between measures of network importance and market value growth. The effect is large, and approximately the same for both network importance measures. The magnitude is such that a 50 percentile increase in firm rank (e.g. from 25th percentile, at rank 17, to 75th percentile, at rank 50) is associated with about a 90% increase in market value. Regressions estimating the line of best fit for this figure, as well as a variation that restricts attention to a balanced panel (the 36 firms that exist in both Q1 2007 and Q3 2020, the beginning and end of our sample) are reported in appendix table A6. Estimates are significantly different than zero for both the balanced and full data sets.

According to the inverted firm hypothesis, the nature of a firm's connections are just as important as their abundance. If an API is strategically placed in an information bottleneck, this may benefit the API creating firm. APIs with high centrality, especially betweenness centrality, play a more important role in connecting the services of firms that would otherwise not be incorporated to the larger internet economy. As appendix Exhibit A2 shows, API degree and centrality are tightly related. Still there are APIs that 'punch above their weight'. A good example is Coinbase, which has relatively few connections to other APIs (24) but is the 10th most central platform overall, because it is the key API connecting many cryptocurrency APIs to online sales and shopping APIs.

A firm's success in the API network is not directly under its control. Third parties must decide to connect. If decisions to join a platform ecosystem is driven by preferential attachment, then small random advantages or disadvantages will snowball into much larger ones.[11] Still, the success of a firm's APIs is likely somewhat endogenous. To further identify the impact of

[11] Preferential attachment might also explain why the variance in some estimates is imprecise. Preferential attachment leads to power-law distributions in the tail, which can have large or even infinite variance (Newman, 2005).

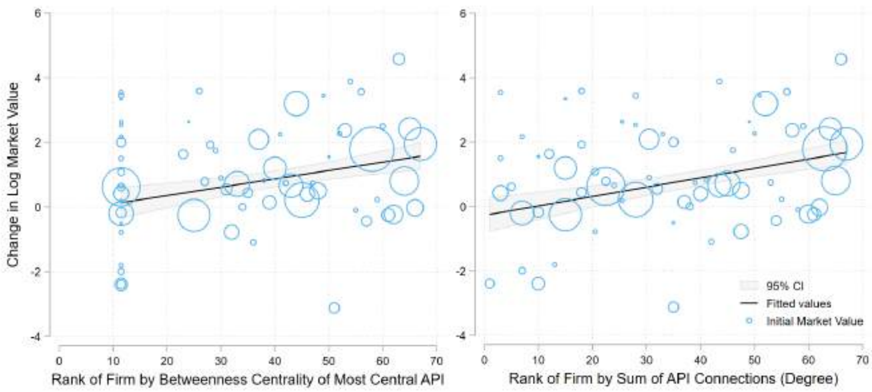

Figure 7: Scatter plots and linear fits, with 95% confidence intervals, of percentage growth in market value on a firm's network rank. Marker sizes are proportional to firms' initial market values. On the left, a firm's network importance is measured by the betweenness centrality of its most central API. There is bunching at rank 11, as there are 11 firms who have APIs that connect to no more that one other API, and therefore are tied for lowest possible betweenness centrality. On the right, firm importance is measured by the sum of a firm's API degrees. Log market value growth is measured from the date the firm first appeared in Compustat. Firm rank is measured for 67 firms as of Q3 2020. Higher rank indicates greater importance. Best fit regression lines are reported in appendix Table A6 both for the full sample of 67 firms and for a balanced panel of 36 firms in business since Q1 2007.

API placement on market value we need variation in the API network that affects a firm's placement within the network that is uncorrelated with relevant omitted variables. Table 3 presents our instrumental variable (IV) strategy estimates using degraded APIs as shocks to the API network to identify the impact of API network placement on firm market value. Because API discontinuations or degradations are unlikely to be anticipated by those building apps connected to these APIs, and even less so to those firms hosting other APIs, these negative connectivity shocks to the API network provide a plausibly exogenous source of variation.[12]

We hand-collected panel data on disconnections or degradations of the forty most central APIs in our data (those appearing in exhibit A2).[13] To conduct our instrumental variable analysis, we calculate the network centrality of APIs in every period both including and not including degraded APIs after their degradation events. Our IV strategy uses the changes in network statistics coming from the degradations to instrument for post-degradation network statistics. The first stage of this IV regression, where disconnections are used to explain network centrality, produces F-statistics over 28 for all specifications, well above the threshold suggested by Stock and Yogo (2002).

[12] Compare to the identification strategy of Benzell and Cooke (2021), which identifies the effect of family ties by instrumenting using the deaths of individuals important to the marriage network.

[13] Of the forty APIs, five had significant reductions in features, two were discontinued and replaced with similar APIs, and one was entirely discontinued. The del.icio.us API was discontinued (Q2 2017); Google Visualization and Google Chart APIs were replaced with substitutes (both Q1 2019). The five APIs with significant reductions in functionality over our period were Twitter (Q2 2018), Facebook (Q2 2018)), Last.fm (Q2 2014), Google Gadgets (Q4 2013), and LinkedIn (Q2 2015).

The 2SLS coefficients from this regression are presented in Table 3. They show that better placement in the API network significantly and positively affects firm's market value. We see positive and significant coefficients on the sum of centrality, sum of degree, and the sum of effective network size, and marginal significance for max centrality and degrees. The size of the coefficients on the sum of centrality, degree, and network size vary from 0.125 to 0.151. Network statistics are centered and scaled, meaning the coefficients indicate that a one standard deviation change in the sum of centrality, degree, or network size is estimated to increase firm market value by 12.5% - 15.1%. We note that the sum of the network statistics is most significant – and not the mean or max – indicating that for a firm, their combined API presence is more important than having a single, important and dominant API. For robustness, we include the OLS results regressing the network statistics on market value, shown in table A8 and find results consistent with the 2SLS estimates. Together, the above results strongly support hypothesis **H2**, using two different sources of plausibly exogenous variation in API centrality.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mean Betweenness Centrality	0.173 (0.541)								
Max Betweenness Centrality		0.137^{+} (0.0703)							
Sum Betweenness Centrality			0.125^{*} (0.0586)						
Mean Degrees				0.363 (0.361)					
Max Degrees					0.309^{+} (0.177)				
Sum Degrees						0.142^{*} (0.0576)			
Mean Effective Network Size							-0.0462 (0.355)		
Max Effective Network Size								0.243 (0.158)	
Sum Effective Network Size									0.151^{*} (0.071)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs	2287	2287	2287	2287	2287	2287	2287	2287	2287
Apps Connections	19893	19893	19893	19893	19893	19893	19893	19893	19893
API Firms	66	66	66	66	66	66	66	66	66
F-Stat 1st Stage	33.08	51.42	68.31	38.91	41.85	90.09	28.4	37.18	77.52
Adjusted R2	0.365	0.370	0.371	0.369	0.377	0.370	0.364	0.375	0.371

Notes: All explanatory variables normalized to mean zero, with a standard deviation of one. The explanatory variable in columns 1, 2, 3 are the max, mean, and sum of betweenness centrality. Columns 4-6 report mean, max and sum of degrees in the API network. The explanatory variables in columns 7-9 are the mean, max and sum of a firm's API's effective network size according to Burt (1992). First stage uses changes in a given API network statistic brought about by API shutdown events to instrument for that network statistic. 2nd stage regresses those predicted network statistics on log market value. + $p < 0.10$ * $p<0.05$ ** $p<0.01$ *** $p<0.001$

Table 3: 2SLS Results Using Disconnected APIs as an Instrumental Variable for Effect of API Network Statistic on Firm Market Value.

To further investigate the mechanisms by which inverted firms create value, we relate market value to other measures of third-party API engagement. Specifically, in a two-way fixed effect model, we regress market value on follower count, developer count, and number of API updates in addition to the binary indicator of API adoption. We present the results of this analysis in appendix table A9. Follower count is a firm level sum of the number of followers of that firm's APIs. Following an API allows a ProgrammableWeb user to easily track updates to those APIs, and is therefore a self reported measure of that user's interest in that API. Likewise, number of developers tracks the interest of self reported developers.[14] 'Change count' reports the total number of updates the firm has made to all of its APIs. Finally, Table A10 estimates market value based on APIs with zero listed developers, showing low and insignificant coefficients.

Almost all specifications show the intensity of engagement to be significantly correlated with market value growth, over and above the extensive margin of API adoption. We estimate that adding 100 additional API developers is associated with a 1.75% additional increase in market value. In a parallel specification, 100 additional API followers is associated with a 0.13% increase in market value. The managerial implications for labor are large. An increase in outside programmer interest, of a magnitude generating one more self-reported developer on ProgrammableWeb, is associated with an average increase in market value of $4.52 million. This implies that managers need methods to recruit and support outside expertise. Hypothesis **H2** is supported by this analysis as well.

6.2 Evidence on Internal Productivity

To distinguish between adopting a public vs. private API, we draw on our second API usage dataset – one from a private API tool provision company. Of the 78 firms who deployed APIs using tools from this company, only 44 are listed as having public APIs available at any point on ProgrammableWeb. Therefore, we can measure the effect of internal APIs by focusing on the effect of API adoption among the remainder. In this data, we measure the date of API adoption as the first date we observe the firm with non-zero data flows through one of their APIs.

Figure A3 reports estimates of the effect, over time, of API adoption on log market value for firms adopting purely internal APIs. The specification used is equation (2), and as in Figure 5 above, while all leads and lags are included in the estimation, only estimates for quarters within two years of adoption are displayed.

Using specification (2), Figure A3 shows there is no clear effect of internal API adoption on firm market value. As an alternative specification we use generalized synthetic controls as above. Again, we fail to find evidence of a positive effect of API adoption, as shown in the counterfactual plot Figure A20 or the synthetic control difference-in-difference estimate Table

[14]It is not uncommon for ProgrammableWeb users to be both followers and developers of the same API.

	All Firms	Excluding Top 20 Firms with Most Popular APIs	Excluding Industries Where < 1% of Firms Have APIs	Excluding Any Computer Services Firm	Year API Open < 2012	Year API Open ⩾ 2012
Log of total assets	0.738***	0.738***	0.738***	0.747***	0.736***	0.736***
	(0.0164)	(0.0166)	(0.0166)	(0.0184)	(0.0167)	(0.0168)
Post x API	0.135*	0.133*	0.123*	0.156*	0.305***	0.0744
	(0.0607)	(0.0623)	(0.0608)	(0.0716)	(0.0912)	(0.0744)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
R2 Adjusted	0.951	0.950	0.951	0.952	0.948	0.948
Obs	132934	132178	127528	118936	129402	129772
Firms	4645	4625	4476	4058	4554	4559
API Adopters	177	157	176	100	86	91

Notes: Notes: Standard errors in parentheses clustered at firm level. Outcome variable is log market value of the firm. *Post x API* is a binary variable that equals one if a given firm has a public API operating on a given date. + p < 0.10 * p<0.05 ** p<0.01 *** p<0.001

Table 4: Difference in difference estimate of the effect of API adoption, with firm asset controls.

A12. This test of **H1a** shows low confidence in an effect. However, the confidence interval is wide, and consistent with a moderate or even large positive effect.

An alternate mechanism by which APIs are said to boost firms internally is through reducing adjustment costs. This would be consistent with increased dynamic capabilities or options value from remixed resources (Teece and Pisano, 2003; Baldwin and Clark, 2006). If APIs allow firms to more easily integrate new resources or reconfigure old ones, firms should be able to make and capitalize on investments more quickly. This should lead profitable (at the margin) firms to make more investments, boosting their market capitalization. Alternatively, if APIs primarily benefit firms through firm inversion, the firm itself will not need to make major capital investments in order to grow. Third parties would make them. A typical approach to measuring whether a firm's investment is limited by capital adjustment costs is Tobin's Q (Tobin, 1969), the ratio of market capitalization to assets.

To test which theory best explains the growth in market capitalization for API using firms, we run a set of regressions analogous to specification (1) with the addition of log of firm assets. Essentially this means we now estimate the effect of API adoption on Q (log(Q) to be precise). As Table 4 shows, across specifications, API adoption positively predicts market value after controlling for total assets. In the base specification, paralleling Table 1 column 1, the effect of API adoption is roughly cut to a third after controlling for growth in assets. This means that while some of the effect of API adoption on market value is mediated by added asset investments, API adoption still increases Q, consistent benefits of API adoption stemming from factors outside the firm.

7 API Exposure: Security Challenges & Responses

The usefulness of APIs depends on how well they balance trade-offs. An API is a kind of aperture or membrane that selects which information to diffuse in and out. Too wide an aperture and the firm may give away its data assets. Too narrow or difficult to access and outsiders will struggle to meaningfully engage. As noted above, firms that update their APIs more frequently see larger increases in market value (see appendix Table A9), consistent with the idea that managing details of third party API use is critical and benefits accrue over time.

One dominant decision for trafficking in data is how to defend against data breaches. If APIs increased the risk of major loss or liability, their use would pose an important downside risk. There is a trade-off between an interest in enabling third party innovations and an interest in thwarting third party damage or ransom. Opening APIs can have both effects. The trade-off depends in part on the relative mix of benevolent and malicious outsiders, which is hidden information. Ransbotham (2016) has shown this "Paradox of Exposure" to be present in the context of open-source software. This risk is particularly notable given evidence that executives of companies who experience data breaches face negative personal consequences (Kamiya et al., 2018). Even if the ratio of risk to reward is favorable, risk aversion or personal costs to executives may limit investment in API projects.[15]

Table 5 reports an increased risk of data breach by insiders in the two years post adoption. Relevant for APIs, this may represent stolen or forged credentials for authorized API keys. Data loss based on physical documents or portable computers show little or no significance.[16]

To explore how firms respond to data breaches, appendix Figure A24 takes advantage of the fact that we observe data flows in our proprietary API data to see how firms respond to data breach events. This data includes all 78 firms who work with the API tool developer, including the 44 which have public APIs. All firms' APIs, in this dataset, were classified by purpose based on their names (see section B).

Figure A24 shows that firms who report data breaches see a decrease in API flows in the short run that rebounds over time. The API type that sees the largest reduction after a hack is internal communications, perhaps indicating firms' hesitance to use internal channels after a

[15]Other hypothetical instances of data 'overexposure', such as intentionally giving away data that later turns out to be key to a firm's competitive advantage are also possible but beyond the scope of the current paper

[16]Table A11 controls for time-varying popularity of firms' APIs using a time-varying Google Trends score for each of the firm's APIs in the logistic panel regression. Surprisingly, this increases the estimated effect of API adoption on malicious insider breach events, which we estimate at almost nine times more likely controlling for API popularity.

	Any Breach Event	Breach of Credit Card Info	Breach via Malicious Hack	Breach via Stolen Document or Fixed Computer	Breach via Portable Computer	Breach via Malicious Insider	Log Count of Records Exposed
0-2 years post API adoption	1.086	1.482	0.854	5.526^{+}	0.772	6.852**	1.004
	[0.28]	[0.33]	[-0.33]	[1.66]	[-0.22]	[2.63]	[-0.06]
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R2	0.0887	0.332	0.181	0.355	0.205	0.253	
Log Likelihood	-581.5	-41.21	-165.9	-32.33	-94.29	-104.6	-23811.1
Obs	3878	445	1522	386	1054	987	91946
Event Count	221	19	63	15	35	44	95
R2 Adjusted							0.000151

Notes: T-statistics in parentheses. Exponentiated coefficients presented. Outcome variable is a binary indicator of whether specific type of breach event occurred (first six columns) or log of total amount of records breached (final column). Event count refers to number of distinct breach events of a given type. Last column estimated via panel linear regression. + p < 0.10 * p<0.05 ** p<0.01 *** p<0.001

Table 5: Fixed effect logistic (first six columns) or linear regression (final column) of impact of API adoption on breach events or log total records exposed. *Breach by Malicious Insider* is often misuse of an authorized API key by a malicious unauthorized actor.

data breach. On the other hand, the data flows for testing APIs increase dramatically in the months after a data breach is reported. This is consistent with firms taking steps to reduce adverse API exposure in the wake of an unexpected breach.

Substantial differences in API practice separate good firms from bad. Firms that consciously open external APIs are much more careful about security for two reasons, awareness and agency costs.[17] Design for external use causes host firms to intentionally harden APIs against attack. Such firms are aware of the risks and mitigate them. By contrast, those firms that implement technology for internal use assume that "security through obscurity" protects them, leading to cost cutting by avoiding security investments. These hidden points of access are less secure but nonetheless discoverable by those hackers sophisticated enough to look for them. Interestingly, in the two years post API adoption, breaches by malicious insiders were less likely at firms with more popular APIs.[18] A second reason why conscious external implementation dominates casual internal implementation is a partner agency problem. Most firms, whether implementing APIs or not, must coordinate with upstream suppliers and downstream channels. If this coordination is left to upstream and downstream partners, *without* conscious design by the host firm, then the partner often implements back-door system access for their own needs but does not take the necessary steps to secure access relative to the host needs. Thus host firms that

[17]Conversation with the CTO of Mulesoft, one of the largest API providers, July 21, 2022.
[18]See Table A11, Column 5, 90% confidence.

consciously protect themselves are more secure than those who simply leave it to others. Security is challenging and even good programmers exhibit blind spots in coding practice (Oliveira et al., 2018). API practices that distinguish successful from unsuccessful firms include (i) rate limiting data queries and throttling them when rates are exceeded (ii) time limiting queries to curb copycat requests (iii) using well-established standards in preference to custom built (iv) separating the API access tokens name and password credentials (v) never storing plaintext credentials, and (vi) two-factor authentication (Lamba, 2019). Editors at ProgrammableWeb have observed firms shifting from Larger Numbers of Unknown Developers (LNUD) to Smaller Numbers of Known Developers (SNKD) (Berlind, 2016). In practice, this strategy may balance the benefits of inverting the firm and securing systems from breach.

8 Conclusion

This paper evaluates the inverted firm business strategy through the lens of a key enabling technology, APIs, that facilitate access to digital resources. This is particularly important for cooperating with developers outside a firm, enabling third parties to build apps and add value using the API-hosting firm's data and digital services. If the API-hosting firm can capture enough of the value created by these third-parties, then the inverted firm strategy succeeds. This paper estimates, for the first time, the quantitative effect of API adoption on market capitalization. By distinguishing internal and external productivity effects, we estimate the effects of inverted firm versus pipeline strategies.

Using public API data from ProgrammableWeb, we visualize the growth of the digital economy over time. Representing APIs as nodes and the apps calling them as edges, the size of the digital economy grew dramatically from 2005-2017. Central APIs play a disproportionate role in anchoring ecosystems within this network. APIs with the highest betweenness centrality and effective network size include Google Maps, Twitter, YouTube and Facebook, as well as smaller players like Shopify, Coinbase, and Dropbox. Firms with successful APIs saw tremendous market value growth. The fourteen publicly traded firms with APIs in the top forty by betweenness centrality saw their total market value increase by \$6.6 trillion dollars from 2005 through 2021, representing a sizable share of total appreciation in the US equity market over that period.

To confirm the role of APIs in boosting market value, we ran a series of analyses. A difference-in-difference model showed that public API adopting firms saw their market value increase by 12.9% increase in market value over two years. An event study analysis, with leads and lags, shows that the effect size grows with the length of time since API adoption. This is consistent with APIs growing in utility as more complementors add more value. Multiple robustness analyses confirm the positive treatment effect. Subsets removing superstar firms and technology firms, as well as quantile regressions, show results span firm sizes and industries. The financial implications are economically significant, implying a return on investment on the order of 500x.

We then investigate to what extent the success of API adopting firms is due to enabling third party value creation, the inverted firm hypothesis, versus enabling internal value creation, the pipeline model. We show that firms with greater third party engagement, as measured by number of followers and developers, see greater gains in market value. Firms with zero developer-followers had no statistically significant gains. Developer engagement thus provides a useful predictor of market value. These results are important for information systems governance as they imply a need to attract developers and reward third party investment. Building public APIs on which no one builds is a failure to invert the firm.

Beyond numbers of complementors, network position also matters, a result shown both in pooled, two-way fixed effect, and in IV panel specifications. IV results are particularly compelling. Using degradations of central APIs, which are plausibly exogenous to the counterfactual financial success of the many other firms hosting APIs which are connected to them, as instruments for API network centrality, we find firm API degree and centrality significantly increase a firm's market value. While difference-in-difference results might plausibly be confounded by endogenous firm choices, the fact that changes to the API network impact the value of all firms in that network confirms a causal role for APIs in raising market capitalization.

To further investigate whether gains from APIs come from an inverted firm strategy versus an internal productivity effect, we use proprietary data from an API tool provision company to replicate our analysis for private APIs. We fail to find evidence of a direct market value effect from private APIs. That said, those estimates have large confidence intervals, and are consistent with a moderate positive effect. We also test the hypothesis that APIs help firms internally by lowering their capital adjustment costs, which would tend to lower their Q. In a specification controlling for a firm's capital assets, we find that Q rises and there is still a positive effect of API adoption on firm value. This effect is attenuated, however, indicating that some gains derive from internal capital adjustment even if most arises from third party complementors. Financially, these results show not just whether to invest but also where i.e. in technology that facilitates outside engagement. Strategically, these results show first, that an inverted firm organizational structure can tap open innovation that is more profitable than a closed firm pipelines structure. Second, they show that promoting an interconnected web of outside partners, placing oneself at the center, is more valuable than adding numerous disconnected partners. Recapitulating an important insight from social networks, centrality and effective network size are important in the context of APIs.

Finally, we investigate one major downside of API adoption – a greater risk of data breach. Panel fixed-effect logistic regressions show an increased risk of data breach in the two years following API adoption. Combining data on breach events and data from Google Trends, suggests malicious insider breaches are not born of popularity but affect obscure APIs, likely due to poor security. Further, we observe firms making clear adjustments in behavior in the wake of a data breach. Consistent with APIs playing a role in these events, firms decrease their use of internal communications APIs and increase their API testing in the months after a hack. Does this mean that firms should avoid implementing APIs out of fear of losses from data breaches? Evidence suggests otherwise. From a purely economic standpoint, the gains in firm market value swamp the costs from data breaches. Secondly, evidence of increased testing after a hack indicates that firms learn to employ better security practices following a hack. Firms may want to consider limiting API strategies when they foresee minimal gains from including outside collaborators and where security concerns are of critical importance.

Collectively these results show, quantitatively, that APIs play a critical role in the economy's growing digital ecosystem. Firms that use APIs to successfully implement an inverted firm strategy place themselves at the center of this ecosystem and capture large returns.

References

Aral, Sinan and Marshall Van Alstyne, "The Diversity Bandwidth Trade-off," *American Journal of Sociology*, 2011, *117* (1), 90–171.

Baker, Wayne E, "Market networks and corporate behavior," *American journal of sociology*, 1990, *96* (3), 589–625.

Baldwin, C. and K. Clark, *Design rules: The power of modularity*, Vol. 1, The MIT Press, 2000.

Baldwin, Carliss Y and Kim B Clark, "The architecture of participation: Does code architecture mitigate free riding in the open source development model?," *Management Science*, 2006, *52* (7), 1116–1127.

Bartels, Andrew, "Your technology investments may not deliver the productivity gains you expected," May 2019.

Benzell, Seth G and Erik Brynjolfsson, "Digital Abundance and Scarce Genius: Implications for Wages, Interest Rates, and Growth," 2019.

_ and Kevin Cooke, "A Network of Thrones: Kinship and Conflict in Europe, 1495–1918," *American Economic Journal: Applied Economics*, 2021, *13* (3), 102–33.

Berlind, David, "The Naked Truth About Internet Security," May 2015.

_ , "Do you publish API docs online for availability to third party developers?," Feb 2016.

_ , "New snafus illustrate scope of API Security Challenge and hacker tenacity," Jun 2017.

Borgatti, Stephen P, "Centrality and network flow," *Social networks*, 2005, *27* (1), 55–71.

Boudreau, K., "Open Platform Strategies and Innovation: Granting Access versus Devolving Control," *Management Science*, 2010, *56* (10), 1849–1872.

Brynjolfsson, Erik and Lorin M Hitt, "Beyond Computation: Information Technology, Organizational Transformation and Business Performance," *Journal of Economic Perspectives—Volume*, 2000, *14* (4), 23–48.

_ , Daniel Rock, and Chad Syverson, "The productivity J-curve: How intangibles complement general purpose technologies," *NBER Working Paper*, 2018.

_ , Lorin M Hitt, and Shinkyu Yang, "Intangible assets: Computers and organizational capital," *Brookings papers on economic activity*, 2002, *2002* (1), 137–181.

Burt, Ronald S, *Structural holes*, Harvard university press, 1992.

_ , *Structural holes: The social structure of competition*, Harvard university press, 2009.

Cabral, Luís, Justus Haucap, Geoffrey Parker, Georgios Petropoulos, Tommaso M Valletti, and Marshall W Van Alstyne, "The EU Digital Markets Act: A Report from a Panel of Economic Experts," *Cabral, L., Haucap, J., Parker, G., Petropoulos, G., Valletti, T., and Van Alstyne, M., The EU Digital Markets Act, Publications Office of the European Union, Luxembourg*, 2021.

Chesbrough, Henry and Marshall Van Alstyne, "Permissionless innovation," *Communications of the ACM*, 2015, *58* (8), 24–26.

Cusumano, Michael, David Yoffie, and Annabelle Gawer, *The future of platforms*, MIT Sloan Management Review, 2020.

Diestel, Reinhard, "Graph theory 3rd ed," *Graduate texts in mathematics*, 2005, *173.*

Eisenhardt, Kathleen M and Jeffrey A Martin, "Dynamic capabilities: what are they?," *Strategic management journal*, 2000, *21* (10-11), 1105–1121.

Enkel, Ellen, Oliver Gassmann, and Henry Chesbrough, "Open R&D and open innovation: exploring the phenomenon," *R&d Management*, 2009, *39* (4), 311–316.

Furrier, John, "Exclusive with AWS chief Andy Jassy: The wakeup call for cloud adoption," https://siliconangle.com/2020/11/30/exclusive-aws-chief-andy-jassy-wakeup-call-cloud-adoption/ Nov. 2020.

Gates, Stephen, "Post-Equifax: Why API security should be a priority," Jan 2019.

Granovetter, Mark S, "The strength of weak ties," *American journal of sociology*, 1973, pp. 1360–1380.

Hansen, Morten T, "The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits," *Administrative science quarterly*, 1999, *44* (1), 82–111.

Hippel, Eric Von, "Lead users: a source of novel product concepts," *Management science*, 1986, *32* (7), 791–805.

_ , "New product ideas from 'lead users'," *Research-Technology Management*, 1989, *32* (3), 24–27.

Hirshleifer, Jack, "The private and social value of information and the reward to inventive activity," in "Uncertainty in economics," Elsevier, 1978, pp. 541–556.

Iyer, B. and M. Subramanian, "The Strategic Value of APIs," January 2015.

Iyer, Bala and John C Henderson, "Preparing for the Future: Understanding the Seven Capabilities of Cloud Computing," *MIS Quarterly Executive*, 2010, *9* (2), 117–131.

Jacobides, Michael G, Carmelo Cennamo, and Annabelle Gawer, "Towards a theory of ecosystems," *Strategic management journal*, 2018, *39* (8), 2255–2276.

Jacobson, Daniel, Greg Brail, and Dan Woods, *APIs: A strategy guide*, O'Reilly Media, Inc., 2011.

Joseph, Sam, Vivian Ludford, and Bart McAllister, "Plugging In: Enabling the Enterprise for the Platform Economy," Technical Report, Gartner Research Board September 2016.

Kamiya, Shinichi, Jun-Koo Kang, Jungmin Kim, Andreas Milidonis, and René M Stulz, "What is the Impact of Successful Cyberattacks on Target Firms?," 2018.

Khan, Lina, "Amazon's Antitrust Paradox," *Yale Law Journal*, 2017, *126* (3), 710–805.

Lakhani, Karim R and Jill A Panetta, "The principles of distributed innovation," *Innovations: Technology, Governance, Globalization Summer*, 2007, *2* (3).

Lamba, Anil, "API Design Principles & Security Best Practices," *Cybernomics*, August 2019, *1* (3), 21–25.

Lane, Ben, "Equifax expects to pay out another $100 million for Data Breach," Feb 2020.

Makridis, Christos and Benjamin Dean, "Measuring the Economic Effects of Data Breaches on Firm Outcomes: Challenges and Opportunities," *Available at SSRN 3044726*, 2018.

Malone, Thomas W, Joanne Yates, and Robert I Benjamin, "Electronic markets and electronic hierarchies," *Communications of the ACM*, 1987, *30* (6), 484–498.

Moreno, Jacob L and Helen H Jennings, "Statistics of social configurations," *Sociometry*, 1938, pp. 342–374.

Newman, Lily, "Google+ Exposed Data of 52.5 Million Users and Will Shut Down," *Wired*, December 10 2018.

Newman, Mark EJ, "Power laws, Pareto distributions and Zipf's law," *Contemporary physics*, 2005, *46* (5), 323–351.

Ofoeda, Joshua, Richard Boateng, and John Effah, "Application programming interface (API) research: A review of the past to inform the future," *International Journal of Enterprise Information Systems (IJEIS)*, 2019, *15* (3), 76–95.

Oliveira, Daniela Seabra, Tian Lin, Muhammad Sajidur Rahman, Rad Akefirad, Donovan Ellis, Eliany Perez, Rahul Bobhate, Lois A DeLong, Justin Cappos, and Yuriy Brun, "{API} Blindspots: Why Experienced Developers Write Vulnerable Code," in "Fourteenth Symposium on Usable Privacy and Security ({SOUPS} 2018)" 2018, pp. 315–328.

O'Reilly, Tim, *What is web 2.0*, O'Reilly Media, Inc, 2009.

Padgett, John F and Christopher K Ansell, "Robust Action and the Rise of the Medici, 1400-1434," *American journal of sociology*, 1993, *98* (6), 1259–1319.

Parker, G., M. Van Alstyne, and X. Jiang, "Platform Ecosystems: How Developers Invert the Firm," *MIS Quarterly*, 2017, *41* (1), 255–266.

Parker, Geoffrey and Marshall Van Alstyne, "Innovation, openness, and platform control," *Management Science*, 2018, *64* (7), 3015–3032.

Parker, Geoffrey G, Marshall W Van Alstyne, and Sangeet Paul Choudary, *Platform Revolution: How Networked Markets Are Transforming the Economy And How to Make Them Work for You*, New York: WW Norton & Company, 2016.

Podolny, Joel M, "Networks as the pipes and prisms of the market," *American journal of sociology*, 2001, *107* (1), 33–60.

Porter, Michael E, "The value chain and competitive advantage," *Understanding business processes*, 2001, *2*, 50–66.

_ , "The five competitive forces that shape strategy," *Harvard business review*, 2008, *86* (1), 25–40.

Purvis, Russell L, Vallabh Sambamurthy, and Robert W Zmud, "The assimilation of knowledge platforms in organizations: An empirical investigation," *Organization science*, 2001, *12* (2), 117–135.

Ransbotham, Sam, "Open Source Code and the Risk of Attacks after Vulnerability Discovery," http://programme.exordo.com/oui2016/delegates/presentation/140/ 2016. Open and User Innovation Conference.

Reagans, Ray and Ezra W Zuckerman, "Networks, diversity, and productivity: The social capital of corporate R&D teams," *Organization science*, 2001, *12* (4), 502–517.

Rosati, Pierangelo and Theo Lynn, "A dataset for accounting, finance and economics research on US data breaches," *Data in Brief*, 2021, *35*, 106924.

Roth, Jonathan, "Pre-test with caution: Event-study estimates after testing for parallel trends," *American Economic Review*, forthcoming.

_ , **Pedro HC Sant'Anna, Alyssa Bilinski, and John Poe**, "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature," *arXiv preprint arXiv:2201.01194*, 2022.

Rowan, Rip, "Steve Yegge's Platform Rant," 2011.

Scarpino, Andrea, "How my car insurance exposed my position," May 2017.

Schulze, Elizabeth, "How Google, Facebook, Amazon and Apple faced EU tech antitrust rules," https://www.cnbc.com/2019/06/07/how-google-facebook-amazon-and-apple-faced-eu-tech-antitrust-rules.html 2019. CNBC.

Siblis Research, "Total market value of U.S. Stock Market," Jan 2022.

Simmel, George, *Conflict and the web of group affiliations*, New York Free Press, repreinted 1955, 1922.

Spanos, Georgios and Lefteris Angelis, "The impact of information security events to the stock market: A systematic literature review," *Computers & Security*, 2016, *58*, 216–229.

Spring, Tom, "T-Mobile Alerts 2.3 Million Customers of Data Breach Tied to Leaky API," *ThreatPost*, August 24 2018.

Stock, James H and Motohiro Yogo, "Testing for weak instruments in linear IV regression," 2002.

Sundstrom, Kevin, "Microsoft Power Apps API design flaw exposed 38M records of personal data," Aug 2021.

Tam, Kar Yan, "The impact of information technology investments on firm performance and evaluation: evidence from newly industrialized economies," *Information Systems Research*, 1998, *9* (1), 85–98.

Teece, David and Gary Pisano, "The dynamic capabilities of firms," in "Handbook on knowledge management," Springer, 2003, pp. 195–213.

Teece, D.J., "Capturing value from technological innovation: Integration, strategic partnering, and licensing decisions," *Interfaces*, 1988, *18* (3), 46–61.

Thierer, Adam, *Permissionless Innovation: The Continuing Case for Comprehensive Technological Freedom*, Mercatus Center at George Mason University, 2016.

Tiwana, Amrit, Benn Konsynski, and Ashley A Bush, "Research commentary-Platform evolution: Coevolution of platform architecture, governance, and environmental dynamics," *Information Systems Research*, 2010, *21* (4), 675–687.

Tobin, James, "A general equilibrium approach to monetary theory," *Journal of money, credit and banking*, 1969, *1* (1), 15–29.

Uzzi, Brian, "Social structure and competition in interfirm networks: The paradox of embeddedness," *Administrative science quarterly*, 1997, pp. 35–67.

Van Alstyne, Marshall, Geoffrey Parker, and Sangeet Choudary, "Pipelines, platforms, and the new rules of strategy," *Harvard Business Review*, 2016, *94* (4), 54–62.

Weitzman, Martin L, "Recombinant growth," *Quarterly Journal of Economics*, 1998, *113* (2), 331–360.

Wernerfelt, Birger, "A resource-based view of the firm," *Strategic management journal*, 1984, *5* (2), 171–180.

White, Douglas R and Stephen P Borgatti, "Betweenness centrality measures for directed graphs," *Social networks*, 1994, *16* (4), 335–346.

World Bank, "Market capitalization of listed domestic companies (current US$) - united states," 2022.

Xu, Yiqing, "Generalized synthetic control method: Causal inference with interactive fixed effects models," *Political Analysis*, 2017, *25* (1), 57–76.

Yoo, Youngin, Richard J. Boland, Kalle Lyytinen, and Ann Majchrzak, "Organizing for Innovation in the Digitized World," *Organization Science*, 2012, *23* (15), 1398–1408.

Zachariadis, Markos and Pinar Ozcan, "The API economy and digital transformation in financial services: The case of open banking," Technical Report No. 2016-001, SWIFT Institute Working Paper, ssrn.com 2017.

Zhu, Feng and Qihong Liu, "Competing with Complementors: An Empirical Look at Amazon. com," *Strategic Management Journal*, 2018, *39* (10), 2618–2642.

A Additional Tables and Figures

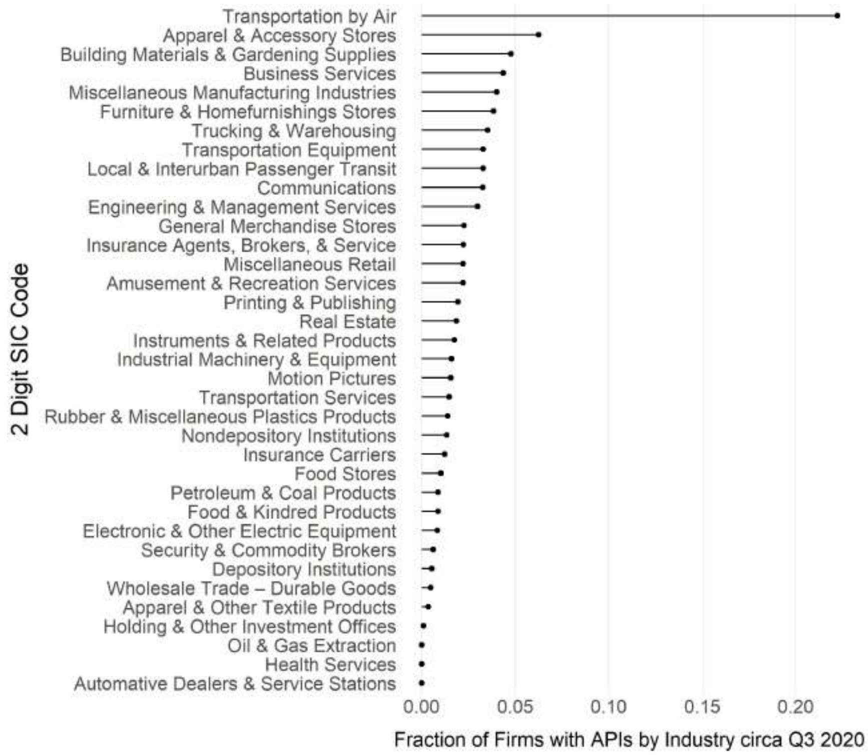

Figure A1: Fraction of firms in a given 2 digit SIC code that operate a public API. Computer services, a major function of many APIs, are included in the Business Services industry.

	Mean	Std. Dev.	Min	Max
Firm Financial Characteristics:				
Log of firm market value (All firms)	5.555	2.51	0	14.5
Log of firm market value (Firms with Public APIs)	9.095	2.03	0.25	14.5
Log of total assets (All firms)	5.962	2.77	0	15.0
Year	2013.0	3.95	2007	2020
Firm's API Characteristics:				
Year of firm's first API	2011.3	3.29	2005	2019
Share of firms ever using an API (public API)	0.0502	0.22	0	1
Total developers for firms' APIs (100s)	3.032	92.1	0	4378
Total followers for firms' APIs (100s)	20.45	390.5	0	16867
Total changes reported for firm's APIs (100s)	0.0989	1.99	0	66
API orientation B2B (business to business)	0.0314	0.17	0	1
API orientation B2C (business to consumer)	0.0331	0.18	0	1
Firm-Quarter Observations (Full sample)		133303		
Firm's API Network Statistics:				
Mean of firm's API network betweenness centralities	0	1.00	-0.44	5.87
Max of firm's API network betweenness centralities	0	1.00	-0.33	6.47
Mean of firm's API degrees	0	1.00	-0.80	7.19
Max of firm's API degrees	0	1.00	-0.54	4.04
Sum of firm's API degrees	0	1.00	-0.33	7.23
Mean of effective network size	0	1.00	-1.44	8.10
Max of effective network size	0	1.00	-0.63	4.33
Sum of effective network size	0	1.00	-0.50	7.06
Firm-Quarter Observations (Sub-sample of firms with APIs connected by apps)		6688		
Data Breaches:				
Any PRC breach event	0.00240	0.049	0	1
Breach via credit card fraud	0.000207	0.014	0	1
Breach via hack or or malware	0.000685	0.026	0	1
Breach via stolen physical device or documents	0.000163	0.013	0	1
Breach via lost or stolen portable device	0.000381	0.020	0	1
Breach via malicious insider	0.000479	0.022	0	1
Log number of records affected	0.00941	0.32	0	18.9
Firm-Quarter Observations (Data breach sample)		91946		
Internal API:				
Share of firms with internal APIs	0.00745	0.086	0	1
Firm-Quarter Observations (Internal APIs and control firms, no Public API firms)		126612		

Table A1: Statistics summarizing merged ProgrammableWeb public API data, Compustat financial outcome data, and Privacy Rights Clearinghouse data. Panel data organized at the firm-quarter level. Compustat data in millions of nominal US dollars. There are 179 public firms matched with APIs on ProgrammableWeb and 78 publicly traded firms for which we have data flow information. Of this 78, 44 are listed with APIs on ProgrammableWeb, leaving 34 observations of firms with purely internal-use APIs.

	Mean	Std Dev	Max	N (Firm Months)
Monthly Calls (Millions)	160	531	6,740	2,453
Monthly Data (Trillions of Bytes)	1.98	10.0	149	1,882
Number of APIs	31.4	46.2	433	2,453

Table A2: Total number of log calls, bytes, and APIs in proprietary API tool provision dataset. Averages by firm-month.

Quantile	Coef	Std. Err.	z-score	CI low	CI high
0.1	0.425	0.062	6.88	0.304	0.547
0.2	0.414	0.049	8.44	0.318	0.51
0.3	0.404	0.041	9.97	0.325	0.484
0.4	0.395	0.035	11.36	0.327	0.463
0.5	0.386	0.033	11.74	0.321	0.45
0.6	0.377	0.035	10.62	0.307	0.446
0.7	0.368	0.041	8.93	0.287	0.449
0.8	0.36	0.049	7.39	0.265	0.455
0.9	0.349	0.06	5.81	0.232	0.467
Firm FE			Yes		
Quarter FE			Yes		
Obs			133303		
Firms			4647		
API Adopters			177		

Notes: This regression estimates equation 1 reporting impact on quantile of firm market value.

Table A3: Quantile regression estimates of equation 1. Coefficient shows estimated effect of API opening on the given quantile of firm market value.

Diff-in-Diff Comparison	Weight	Avg Diff-in-Diff Estimate
Earlier Treated vs Later Control	0.007	-0.111
Later Treated vs Earlier Control	0.013	0.08
Treated vs Never treated	0.952	0.549
Treated vs Already Treated	0.028	0.058
Diff-in-diff estimate:	0.524	

Table A4: Bacon decomposition of baseline difference-in-difference result from Table 1

API Name	Company	Betweeness Centrality	Degree	Degree Rank	Market Cap in Billions June 2005 (* or at IPO)	Market Cap in Billions June 2021	Notes
Google Maps	Google	1	1982	1	363.5	1672.9	
Twitter	Twitter	2	1629	2	22.6*	54.2	IPO in 2013
YouTube	Google	3	1479	3	363.5	1672.9	acquired by google in 2006
Facebook	Facebook	4	1132	5	63.3*	967.9	IPO in 2012
Flickr	Verizon	5	1282	4	95.5	237.4	owned by Yahoo from 2005-2015; Yahoo owned by Verizon from 2017-present
Amazon Product Advertising	Amazon	6	700	6	13.6	1687.0	
Amazon S3	Amazon	7	312	13	13.6	1687.0	
PayPal	eBay	8	118	47	44.7	385.9	numbers for ebay & paypal, paypall spun off in 2015
Last.fm	ViacomCBS	9	587	7	133.4	28.8	
Coinbase	Coinbase	10	24	219	*	58.8	IPO in 2021
eBay	eBay	11	408	8	44.7	385.9	numbers for ebay & paypal, paypall spun off in 2015
Twilio	Twilio	12	266	17	3.1*	66.3	IPO in 2016
Box	Box	13	122	46	0.24*	4.1	IPO in 2015
Google AdSense	Google	14	333	11	363.5	1672.9	
Google Search	Google	15	406	9	363.5	1672.9	
Bit.ly	bitly	16	107	51			
Bing Maps	Microsoft	17	190	26	266	2000.0	
Foursquare	Foursquare	18	331	12			
Google Gadgets	Google	19	99	58	363.5	1672.9	
Bing	Microsoft	20	259	18	266	2000.0	
Google Analytics Managment	Google	21	275	15	363.5	1672.9	
del.icio.us	del.icio.us	22	394	10			private; owned by Yahoo from 2005-2011
Fitbit	Google	23	50	116	363.5	1672.9	
Blogger	Google	24	93	62	363.5	1672.9	acquired by google in 2003
Amazon EC2	Amazon	25	189	27	13.6	1687.0	
Salesforce	Salesforce	26	81	70	2.1	224.0	
GeoNames	non-profit	27	271	16			
MailChimp	MailChimp	28	45	128			
Google Visualization	Google	29	68	82	363.5	1672.9	
Shopify Admin	Shopify	30	42	135	1.8*	183.5	IPO in 2015
Google Chart	Google	31	209	22	363.5	1672.9	
Indeed	Recruit Co Ltd.	32	75	75	17.7*	80.4	acquired in 2012 by Recruit Co, which IPO'd in 2015
LinkedIn	Microsoft	33	239	20	266	2000.0	
Google App Engine	Google	34	176	31	363.5	1672.9	
Google Checkout	Google	35	49	119	363.5	1672.9	
Dropbox	Dropbox	36	85	68	12.3*	11.8	IPO in 2018
Google Calendar	Google	37	94	60	363.5	1672.9	
Yahoo Search	Verizon	38	301	14	95.5	237.4	Yahoo owned by Verizon from 2017 to present
Google AdWords	Google	39	137	39	363.5	1672.9	
Wikipedia	non-profit	40	259	19			

Figure A2: This table reports degree, degree rank, and betweenness centrality rank for selected APIs. All APIs in the topforty for betweenness centrality are displayed. It also reports the company owning the API, and the market value growth of that company since July 2015. APIs that switched ownership are are assigned to conglomerates they are most associated with. Company names are colored as in Figure 3. Google and Facebook are used as familiar vs. Alphabet and Meta. Degree is the number of edges connecting an API to other APIs, where edges correspond to applications that call multiple APIs. 'Connections to oneself' (i.e. apps that only call a single API) are not counted. Market value data source is Companiesmarketcap.com

	All Firms	Excluding Top 20 Firms with Most Popular APIs	Excluding Industries Where ⩽ 1% of Firms Have APIs	Excluding Any Computer Services Firm	Year API Open < 2012	Year API Open ⩾ 2012
8 quarters until	-0.0643	-0.0443	-0.0632	0.0154	-0.512***	0.0826
API adoption	(0.0694)	(0.0701)	(0.0695)	(0.0785)	(0.128)	(0.0794)
7 quarters until	-0.0619	-0.0571	-0.0617	-0.0127	-0.407***	0.0838
API adoption	(0.0642)	(0.0661)	(0.0643)	(0.0757)	(0.0984)	(0.0784)
6 quarters until	-0.0372	-0.0301	-0.0368	-0.00206	-0.345***	0.0929
API adoption	(0.0585)	(0.0597)	(0.0586)	(0.0697)	(0.0965)	(0.0705)
5 quarters until	-0.0542	-0.0440	-0.0542	-0.0355	-0.243**	0.0319
API adoption	(0.0482)	(0.0491)	(0.0483)	(0.0560)	(0.0758)	(0.0605)
4 quarters until	-0.0697	-0.0630	-0.0697	-0.0709	-0.258**	0.0201
API adoption	(0.0453)	(0.0464)	(0.0453)	(0.0532)	(0.0894)	(0.0501)
3 quarters until	-0.0636+	-0.0630+	-0.0632+	-0.0786+	-0.191**	-0.000166
API adoption	(0.0352)	(0.0361)	(0.0352)	(0.0403)	(0.0733)	(0.0377)
2 quarters until	-0.0232	-0.0203	-0.0228	-0.0433	-0.134*	0.0294
API adoption	(0.0294)	(0.0300)	(0.0294)	(0.0350)	(0.0658)	(0.0293)
1 quarters until	-0.000628	-0.00216	-0.000327	-0.00762	-0.0659	0.0335
API adoption	(0.0227)	(0.0232)	(0.0227)	(0.0268)	(0.0453)	(0.0244)
1 quarters since	0.0376	0.0372	0.0375	0.0291	0.0435	0.0316
API adoption	(0.0325)	(0.0330)	(0.0325)	(0.0426)	(0.0490)	(0.0407)
2 quarters since	0.0294	0.0269	0.0288	0.0208	0.0659	0.00725
API adoption	(0.0367)	(0.0371)	(0.0367)	(0.0477)	(0.0513)	(0.0475)
3 quarters since	0.0554	0.0507	0.0543	0.0381	0.0810	0.0401
API adoption	(0.0376)	(0.0383)	(0.0376)	(0.0472)	(0.0536)	(0.0488)
4 quarters since	0.0652	0.0559	0.0638	0.0530	0.0804	0.0517
API adoption	(0.0442)	(0.0451)	(0.0442)	(0.0541)	(0.0631)	(0.0577)
5 quarters since	0.0603	0.0812+	0.0579	0.0226	0.0311	0.0736
API adoption	(0.0514)	(0.0490)	(0.0516)	(0.0644)	(0.0846)	(0.0625)
6 quarters since	0.0988+	0.0840	0.0959+	0.0533	0.157	0.0564
API adoption	(0.0570)	(0.0528)	(0.0571)	(0.0650)	(0.0979)	(0.0660)
7 quarters since	0.123*	0.109+	0.120*	0.0793	0.193*	0.0687
API adoption	(0.0590)	(0.0567)	(0.0591)	(0.0684)	(0.0917)	(0.0733)
8 quarters since	0.126*	0.103+	0.123*	0.0997	0.224*	0.0494
API adoption	(0.0607)	(0.0601)	(0.0608)	(0.0706)	(0.0894)	(0.0775)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
R2 Adjusted	0.141	0.139	0.144	0.150	0.137	0.134
Obs	133303	132547	127892	119285	129769	130141
Firms	4748	4728	4574	4144	4655	4662
API Adopters	179	159	178	102	86	93

Notes: Standard errors in parentheses and clustered at the firm level. Outcome variable is the log market value of the firm. Top 20 firms excluded include Alphabet, Amazon, Apple, Ebay, Facebook, FedEx, Groupon, Liberty Expedia, Microsoft, New York Times, PayPal, Pinterest, Salesforce, Spotify, Twilio, Twitter, Uber, UPS, Verizon, and Zillow. Excluded computer services industry refers to SIC code 7370. Year API Open < 2012 refers only including API firms which opened APIs prior to 2012. + $p < 0.10$ * $p<0.05$ ** $p<0.01$ *** $p<0.001$

Table A5: Table version of estimates reported in Figure 5

	(1) Log Market Value Change	(2) Log Market Value Change	(3) Log Market Value Change	(4) Log Market Value Change
Degree Rank	0.0291***		0.0316***	
	(0.00633)		(0.00765)	
Max Betweenness Rank		0.0258**		0.0265*
		(0.00787)		(0.00980)
Constant	-0.272	-0.163	-0.368	-0.162
	(0.257)	(0.347)	(0.329)	(0.465)
Firms	67	67	36	36

Notes: Robust standard errors in parentheses * $p<0.05$ ** $p<0.01$ *** $p<0.001$

Table A6: Regression underlying the lines of best fit in Figure 7. Robust standard errors in parentheses. Outcome variable is increase in log firm market value from first observed firm market value to Q3 2020. Columns 3 and 4 restrict attention to the 36 publicly traded public API adopting firms who are in the data from Q1 2007 through Q3 2020, constituting a balanced panel. Regressions weighted by firms' initial market value.

	Mean Between-ness Centrality	Max Between-ness Centrality	Sum Between-ness Centrality	Mean Degrees	Max Degrees	Sum Degrees	Mean Effective Network Size	Max Effective Network Size	Sum Effective Network Size
Change in Mean Centrality due to Disconnection	-0.0000428 (0.0000410)								
Changes in Max Degrees due to Disconnection		-0.0000296** (0.0000106)							
Changes in Sum Centrality due to Disconnection			-0.0000225*** (0.00000242)						
Changes in Mean Degree Due to Disconnection				-0.00513+ (0.00302)					
Changes in Max Degree Due to Disconnection					-0.00211+ (0.00122)				
Changes in Sum of Degrees Due to Disconnection						-0.00131*** (0.0000744)			
Changes in Mean Eff. Network from Discon.							-0.00687 (0.00431)		
Changes in Max Eff. Network from Discon.								-0.00242+ (0.00142)	
Changes in Sum Eff. Network from Discon.									-0.00256*** (0.000396)
Constant	-0.0495 (0.0475)	-0.180** (0.0642)	-0.214*** (0.0230)	-0.196+ (0.116)	-0.194+ (0.112)	-0.314*** (0.0178)	-0.0405 (0.0254)	-0.123+ (0.0719)	-0.219*** (0.0340)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs	2287	2287	2287	2287	2287	2287	2287	2287	2287
API Adopters	66	66	66	66	66	66	66	66	66
R2-within	0.0799	0.602	0.850	0.224	0.351	0.935	0.203	0.356	0.836

Notes: All explanatory variables normalized to mean zero, with a standard deviation of one. The explanatory variable in columns 1, 2, 3 are the max, mean, and sum of betweenness centrality. Columns 4-6 report mean, max and sum of degrees in the API network. The explanatory variables in columns 7-9 are the mean, max and sum of a firm's API's effective network size according to Burt (1992). Outcome is log firm market value. Standard errors in parentheses, clustered at firm level. + p<0.10 * p<0.05 ** p<0.01 *** p<0.001

Table A7: First Stage of 2SLS Results Effect of API Network Statistic on Firm Market Value.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mean Betweenness Centrality	0.210 (0.318)								
Max Betweenness Centrality		0.148* (0.0714)							
Sum Betweenness Centrality			0.130* (0.0616)						
Mean Degrees				0.402 (0.305)					
Max Degrees					0.366* (0.169)				
Sum Degrees						0.162* (0.0698)			
Mean Effective Network Size							0.188 (0.272)		
Max Effective Network Size								0.313* (0.149)	
Sum Effective Network Size									0.166* (0.0767)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R2	0.365	0.370	0.371	0.369	0.377	0.370	0.364	0.375	0.371
Obs	2287	2287	2287	2287	2287	2287	2287	2287	2287
Firms	66	66	66	66	66	66	66	66	66
Apps Connections to Firm's APIs	19893	19893	19893	19893	19893	19893	19893	19893	19893

Notes: All explanatory variables normalized to mean zero, with a standard deviation of one. The explanatory variable in columns 1, 2, 3 are the max, mean, and sum of betweenness centrality. Columns 4-6 report mean, max and sum of degrees in the API network. The explanatory variables in columns 7-9 are the mean, max and sum of a firm's API's effective network size according to Burt (1992). Outcome is log firm market value. Standard errors in parentheses, clustered at firm level. + p<0.10 * p<0.05 ** p<0.01 *** p<0.001

Table A8: OLS Results Effect of API Network Statistic on Firm Market Value.

	All Firms	All Firms	All Firms	Excl. Top 20 API Firms	Excl. Top 20 API Firms	Excl. Top 20 API Firms	Excl. Ind. with Few APIs	Excl. Ind. with Few APIs	Excl. Ind. with Few APIs	Year API < 2012	Year API < 2012	Year API < 2012
Post x API	0.337***	0.263*	0.379***	0.348***	0.264+	0.366***	0.267**	0.206+	0.310**	0.689***	0.637***	0.658***
	(0.0883)	(0.109)	(0.0867)	(0.0898)	(0.136)	(0.0873)	(0.0997)	(0.125)	(0.0986)	(0.154)	(0.184)	(0.149)
Post API x API Developers (100s)	1.751***			1.274			1.817***			0.854		
	(0.473)			(0.902)			(0.483)			(0.761)		
Post API x Num API Followers (100s)		0.134*			0.134			0.124*			0.0657	
		(0.0592)			(0.117)			(0.0546)			(0.0620)	
Post API x Num API Change Count (100s)			0.572			0.280			0.715+			9.733***
			(0.384)			(0.199)			(0.431)			(2.236)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2 Adjusted	0.932	0.932	0.932	0.931	0.931	0.931	0.934	0.934	0.934	0.929	0.929	0.929
Obs	133202	133202	133202	132446	132446	132446	119201	119201	119201	129670	129670	129670
Firms	4647	4647	4647	4627	4627	4627	4060	4060	4060	4556	4556	4556
API Adopters	177	177	177	157	157	157	100	100	100	86	86	86

Notes: Standard errors in parentheses and clustered at firm level. Outcome variable log market value of the firm. *Post x API* is a binary variable that equals one if a given firm has a public API operating on a given date. *API developers* are the number of developers (in 100s) of an API according to ProgrammableWeb. *API Followers* refers to the number of individuals (in 100s) on ProgrammableWeb who have elected to follow an API. *Change Count* refers to the number of times an API has been updated (in 100s). Top 20 firms excluded include Alphabet, Amazon, Apple, Ebay, Facebook, FedEx, Groupon, Liberty Expedia, Microsoft, New York Times, PayPal, Pinterest, Salesforce, Spotify, Twilio, Twitter, Uber, UPS, Verizon, and Zillow. Excluded computer services industry refers to SIC code 7370. Year API < 2012 refers to excluding firms which opened APIs after 2012. + $p < 0.10$ * $p<0.05$ ** $p<0.01$ *** $p<0.001$

Table A9: Impact of API usage intensity on market value.

Subset:	0 Developers	0 or 1 Developers	0 Followers	Lowest Decile of Followers (18)	0 Changes	⩽ 1 Change
Post x API	0.134	0.340***	0	0.355	0.371***	0.362***
	(0.111)	(0.0955)	(.)	(0.281)	(0.0902)	(0.0879)
R2 Adjusted	0.928	0.929	0.925	0.925	0.931	0.931
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs	128839	130843	126508	127308	132039	132425
Firms	4535	4586	4470	4494	4617	4628
API Adopters	65	116	0	24	147	158

Table A10: Estimates of API adoption on market value moderated by limited engagement

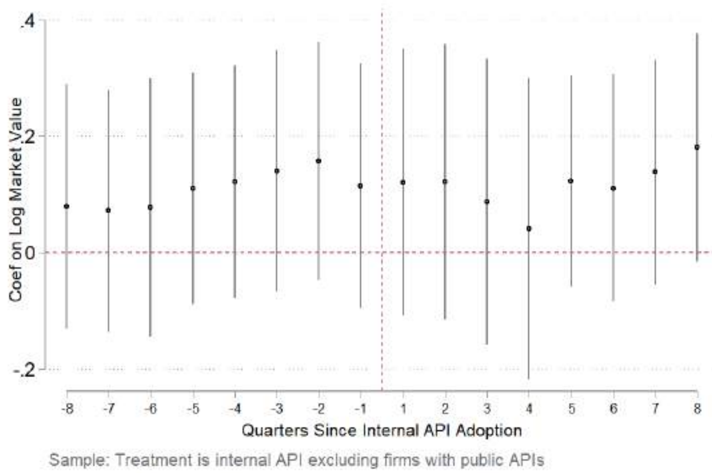

Figure A3: Estimates and 95% confidence intervals for the effect of internal API adoption by number of periods before and after adoption. Equation follows 2, and is comparable to Figure 5 but for internal APIs instead of public.

	Any Breach Event	Breach of Credit Card Info	Breach via Malicious Hack	Breach via Portable Computer	Breach via Malicious Insider	Log Count of Records Exposed
0-2 years post API adoption	1.104	1.475	0.943	0.772	9.898**	0.995
	(0.33)	(0.33)	(-0.12)	(-0.22)	(2.98)	(-0.13)
Google Trends Score Firm API	1.008	0.997	1.028	1	0.593^{+}	1.007
	(0.63)	(-0.11)	(1.28)	(.)	(-1.73)	(1.31)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R2	0.0890	0.332	0.186	0.205	0.280	
Log Likelihood	-581.3	-41.20	-164.9	-94.29	-100.8	-23783.4
Obs	3878	445	1522	1054	987	91946
Firms	115	13	44	31	29	4320
R2 Adjusted						0.000742

Notes: T-statistics in parentheses. Exponentiated coefficients presented. Outcome variable is a binary indicator of whether specific type of breach event occurred (first six columns) or log of total amount of records breached (final column). Coefficient *0-2 years post API adoption* refers to a binary if in the two years after a firm releases its first public API. Regressions estimated up to 2015 due to PRC Breach Data matched to gvkeys according to Rosati et al (2021) ending in 2015. Event count refers to number of distinct breach events of a given type. Last column estimated via panel linear regression. Google Trends Score Firm API refers to the Google Trends score for the firm's API. + p < 0.10 * p<0.05 ** p<0.01 *** p<0.001

Table A11: Fixed effect logistic (first six columns) or linear regression (final column) of impact of API adoption on breach events or log total records exposed controlling for API popularity using Google trends. Breach by a malicious insider is often misuse of an authorized API key by a malicious unauthorized actor.

Average Treatment on Treated	Std. Err.	CI lower	CI upper	p-value
-.055	.426	-0.89	0.781	0.898

Table A12: Estimated effect, and confidence interval, of adopting an internal API using generalized synthetic control following (Xu, 2017).

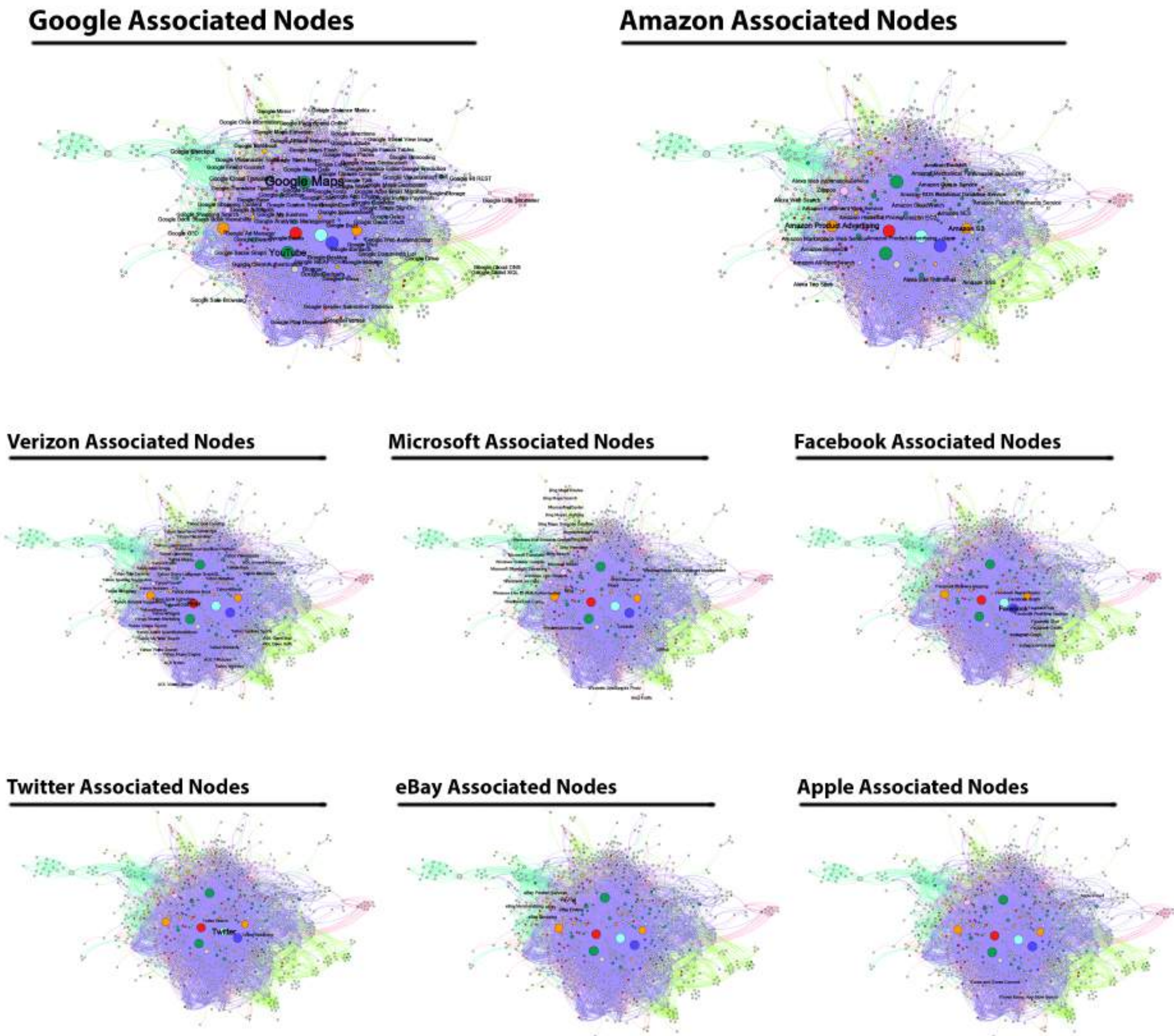

Figure A4: The API network, with nodes owned by different companies highlighted and labeled. Grid format. Notable features include: (1) Google is the host of most nodes in the network (2) Verizon is the second most common source of nodes, in large part due to its acquisition of Yahoo (3) Google's nodes occur everywhere in the API network. On the other hand, eBay's nodes are mostly associated with eCommerce apps, while Twitter and Facebook's nodes are more centrally located, associated with social apps in particular (4) Amazon hosts APIs at opposite ends of the API network, with Amazon's Product Advertising API particularly important to the eCommerce sub-network and Amazon E3 is particularly important to productivity (5) Microsoft hosts nodes important to several subnetworks as well. Many of its APIs are central to the eCommerce network while the LinkedIn API is central to both the social network and productivity subnetworks.

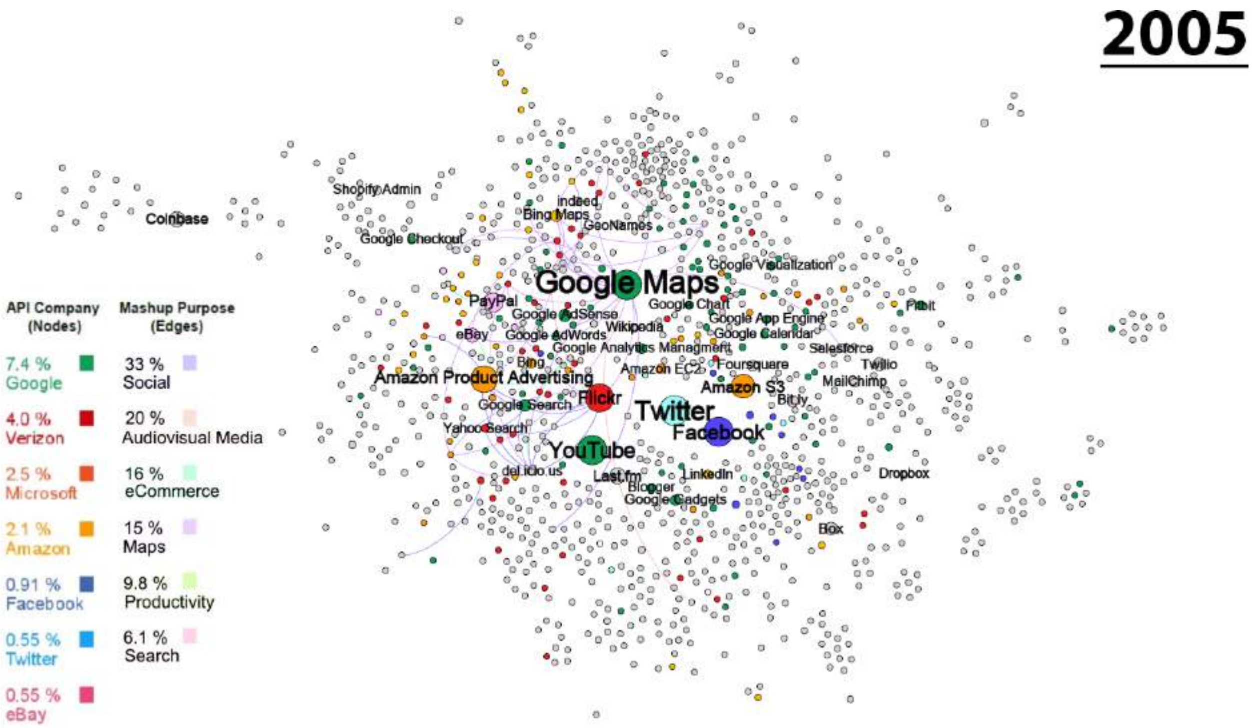

Figure A5: API network as of 2005.

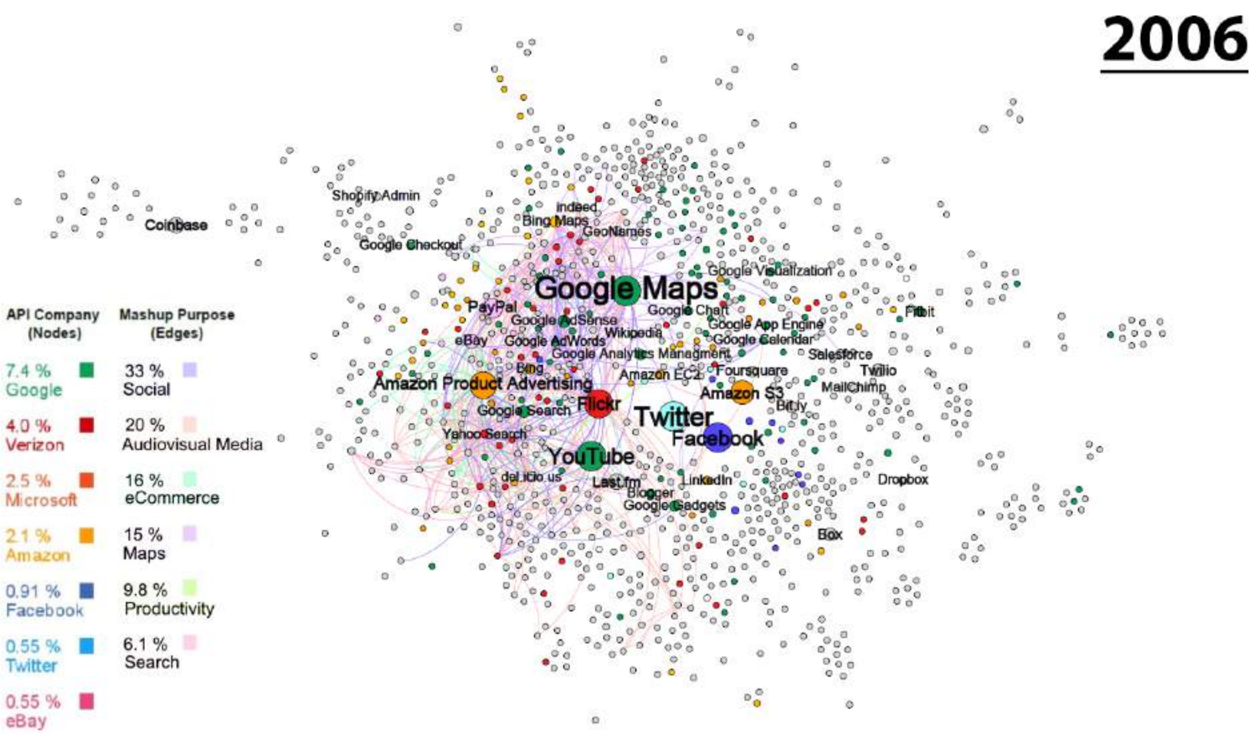

Figure A6: API network as of 2006.

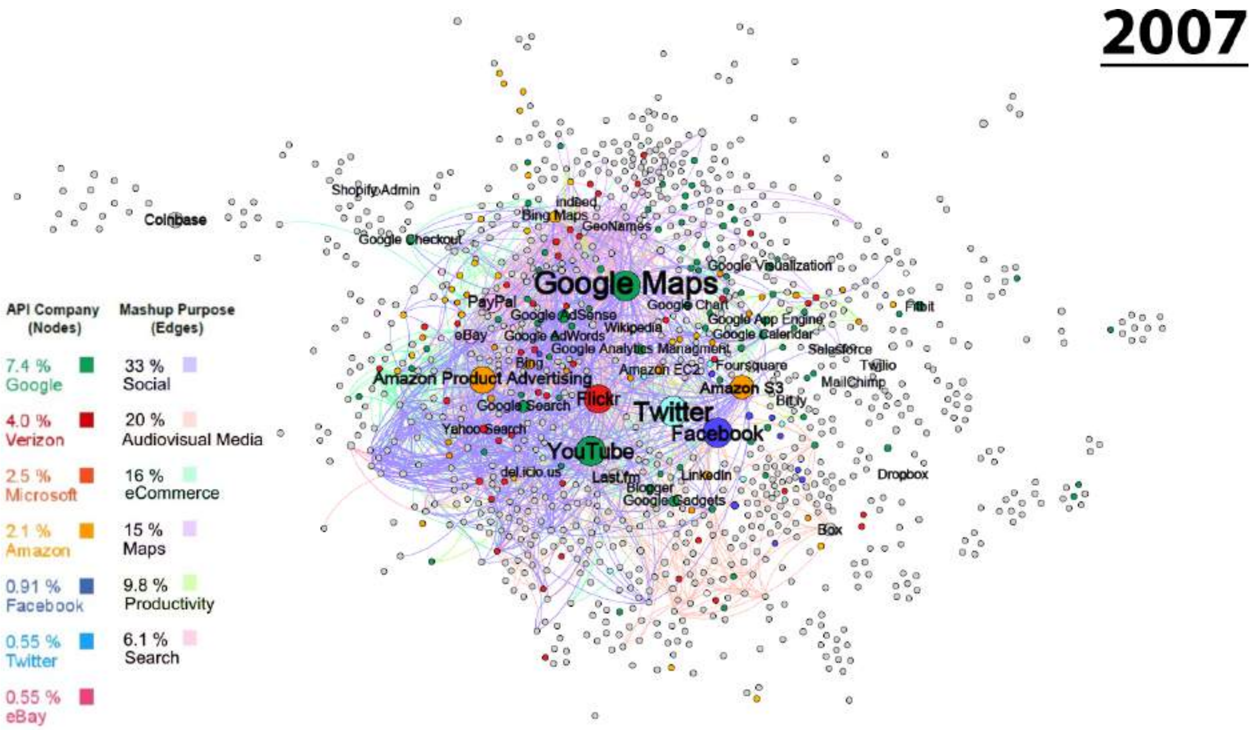

Figure A7: API network as of 2007.

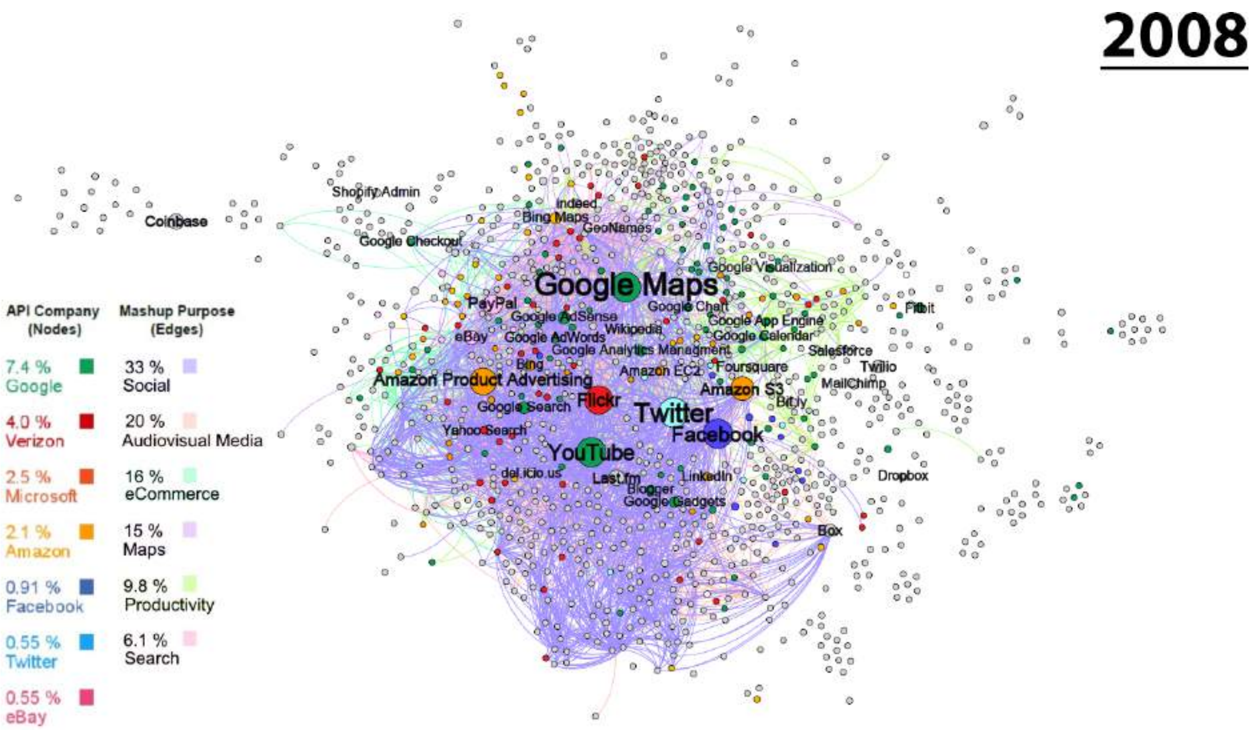

Figure A8: API network as of 2008.

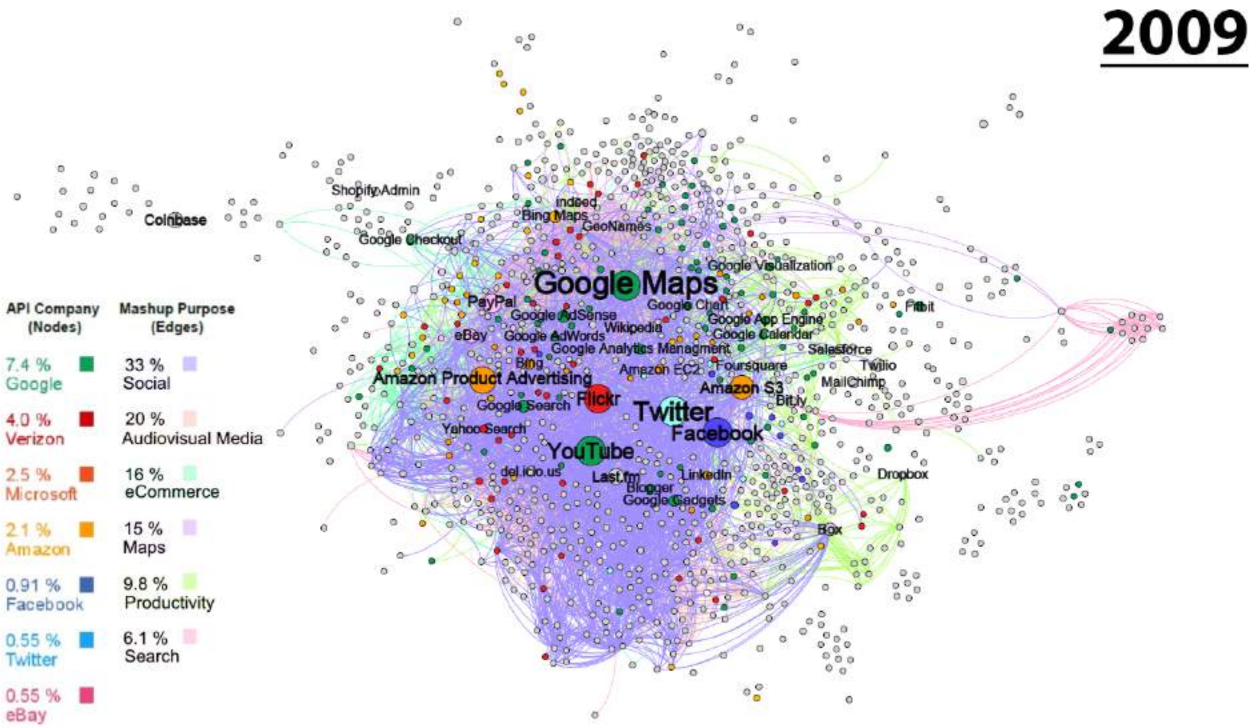

Figure A9: API network as of 2009.

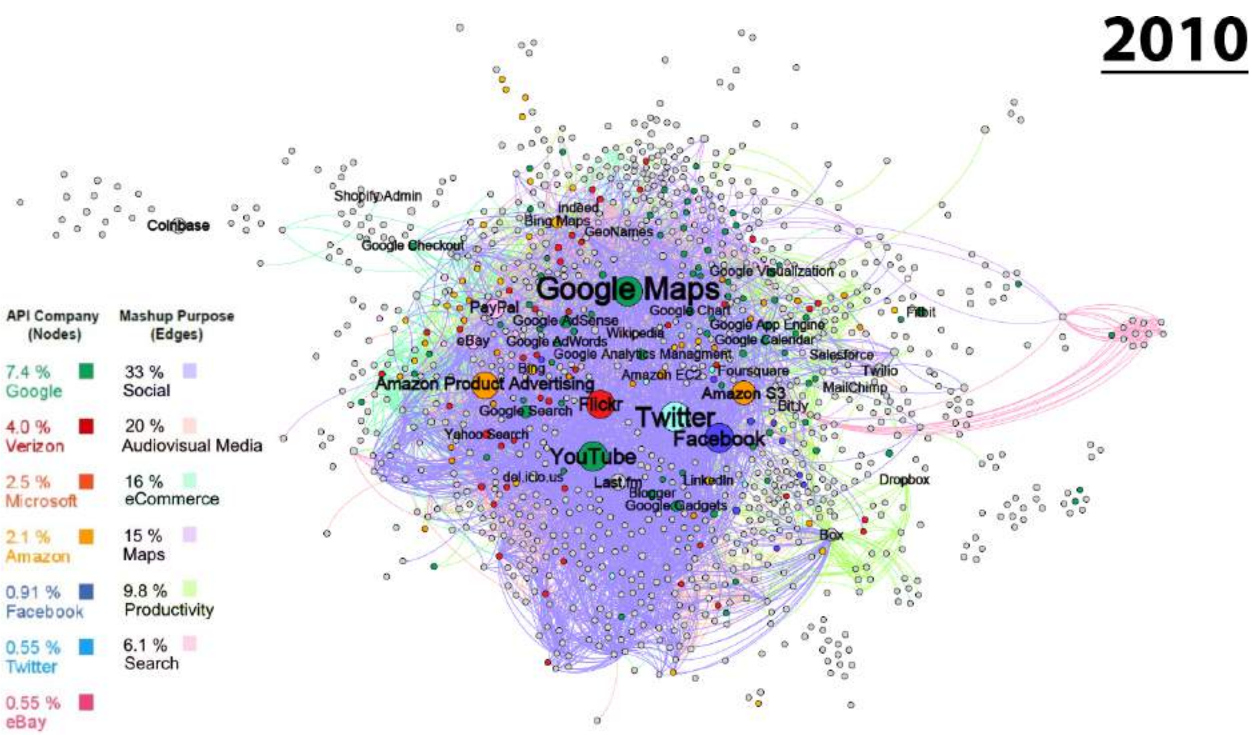

Figure A10: API network as of 2010.

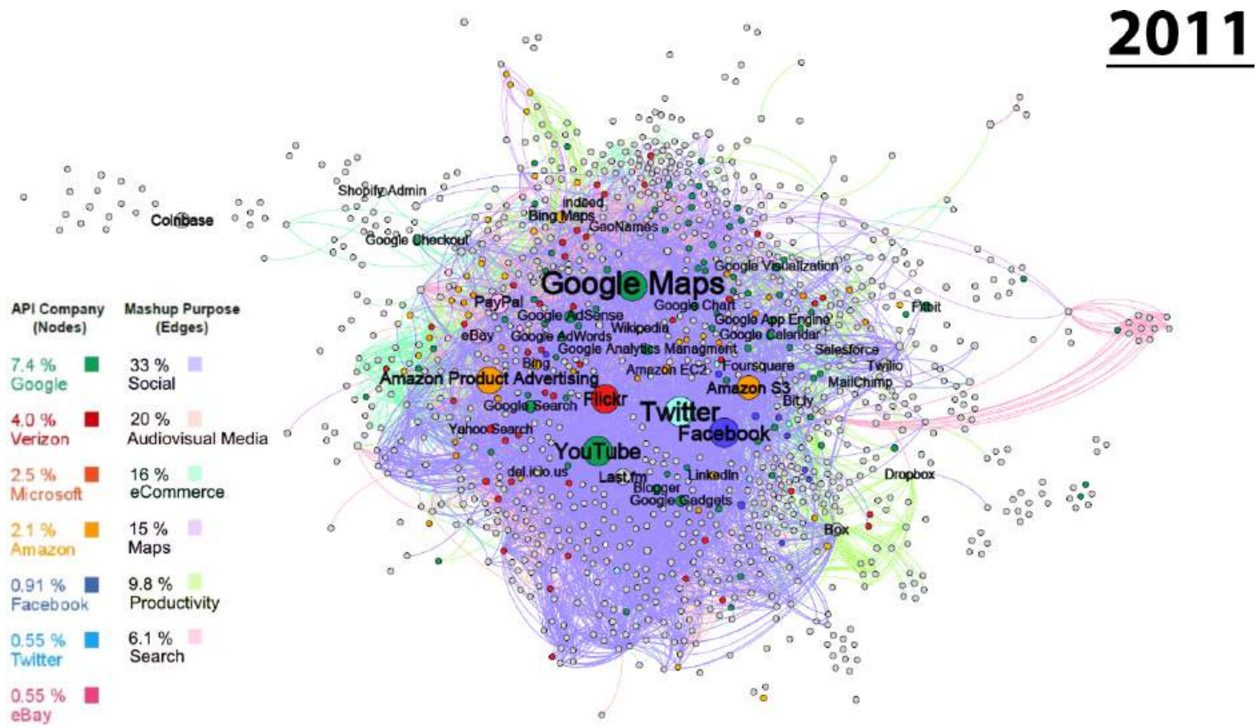

Figure A11: API network as of 2011.

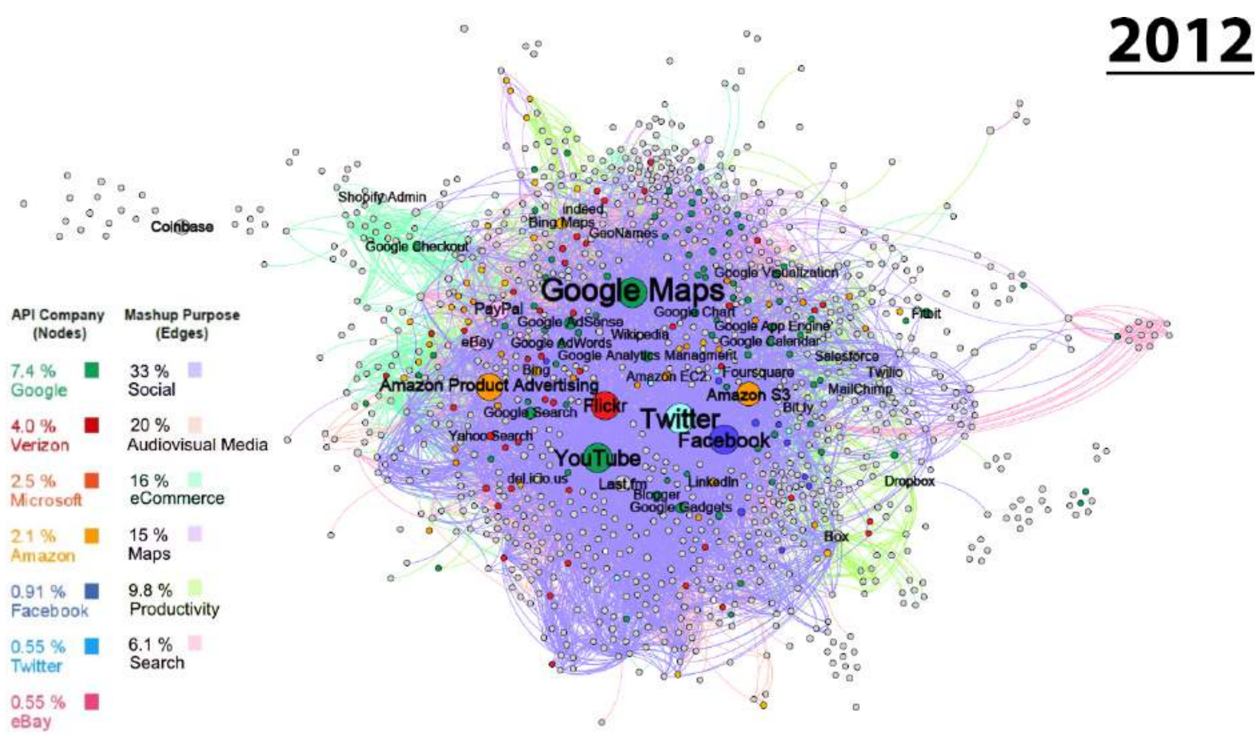

Figure A12: API network as of 2012.

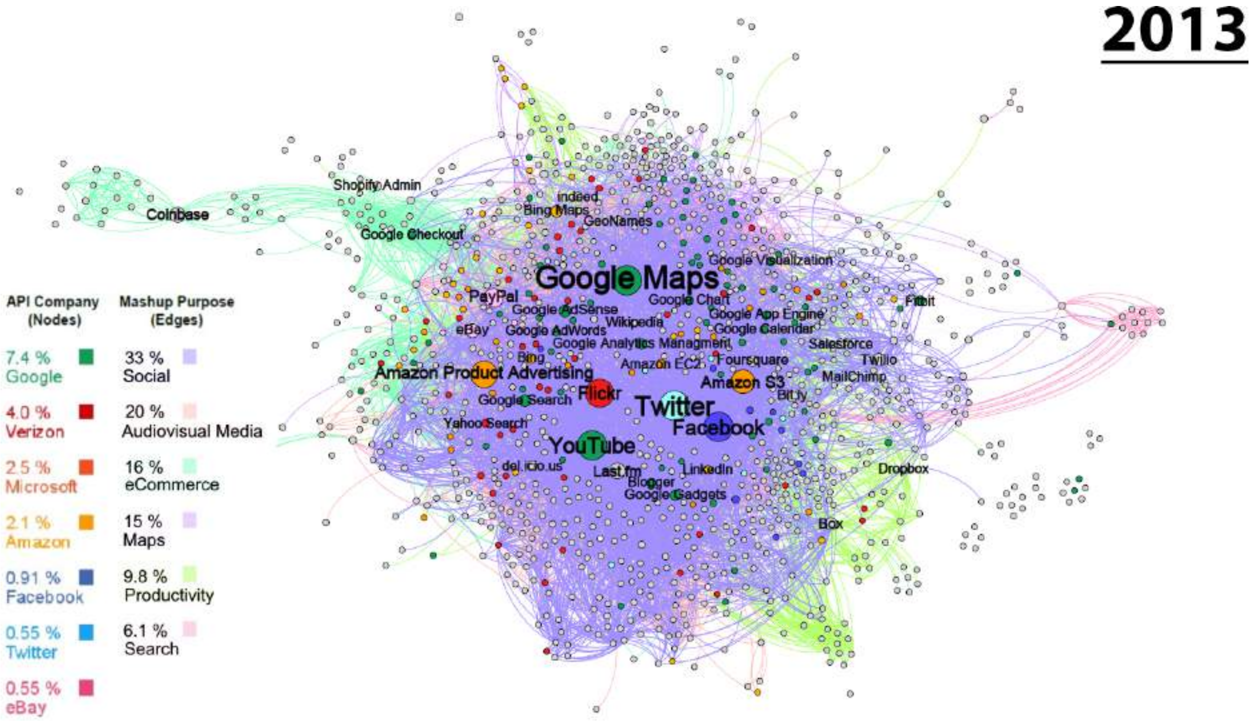

Figure A13: API network as of 2013.

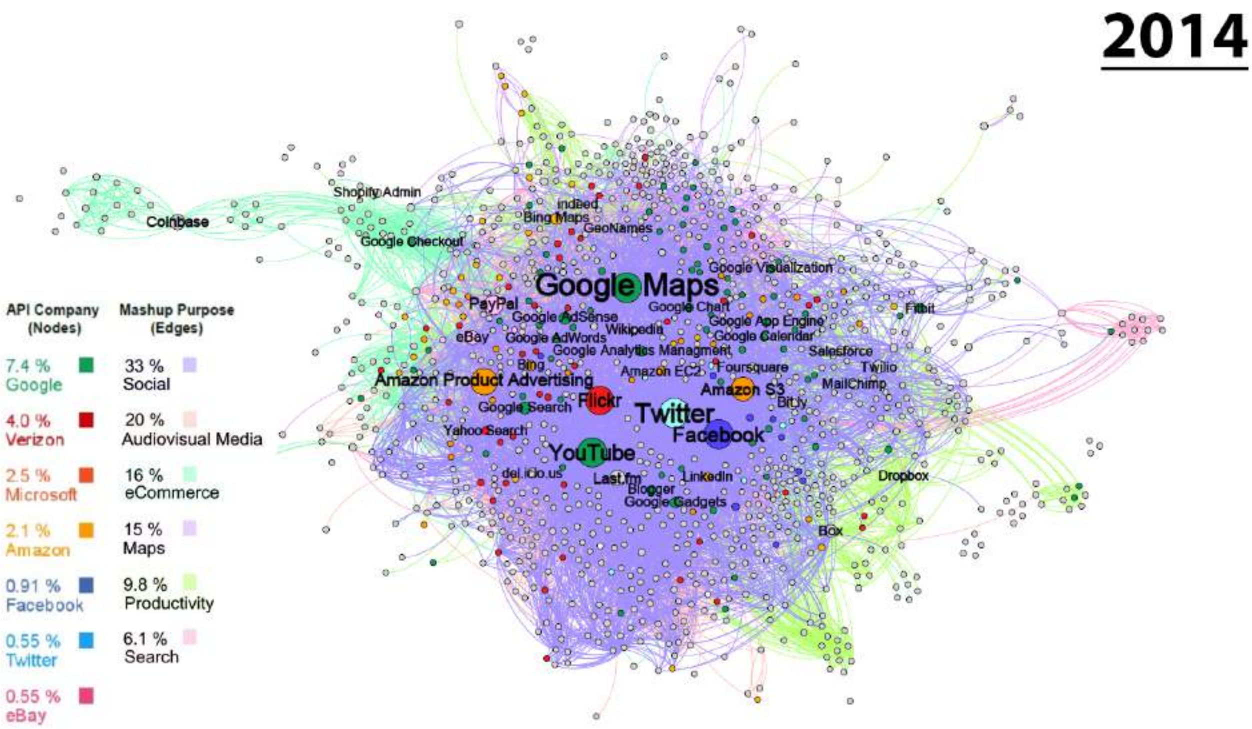

Figure A14: API network as of 2014.

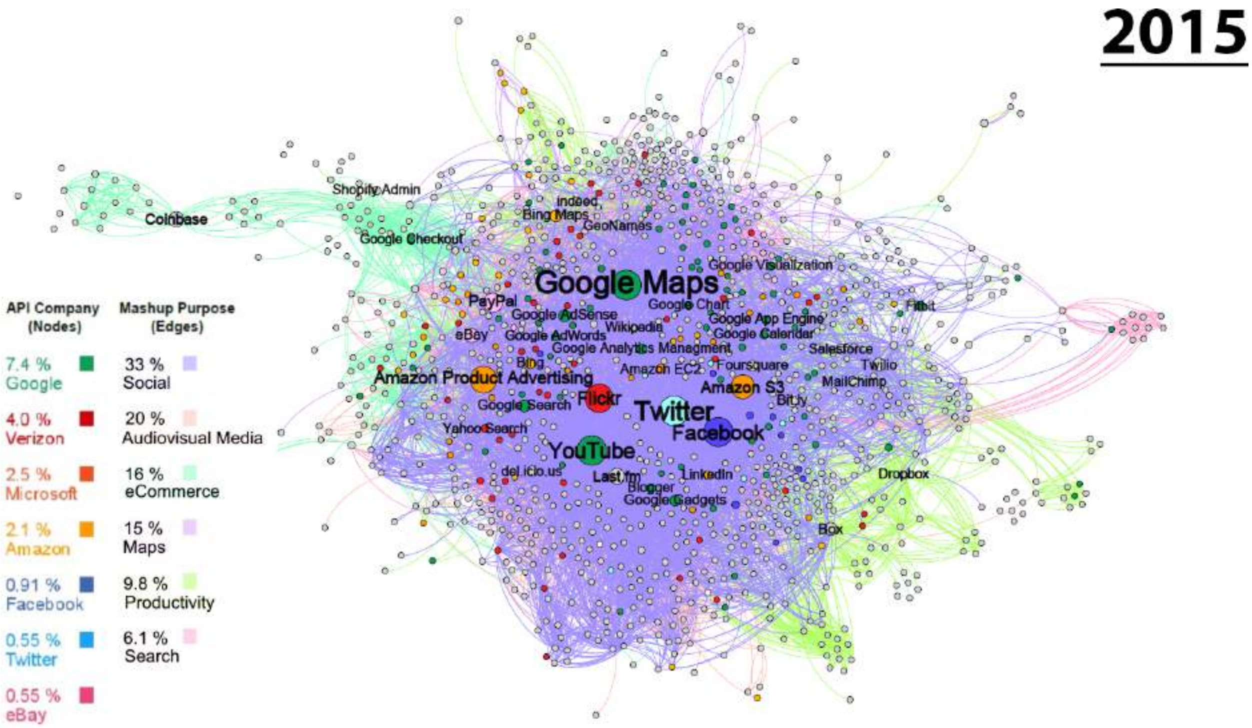

Figure A15: API network as of 2015.

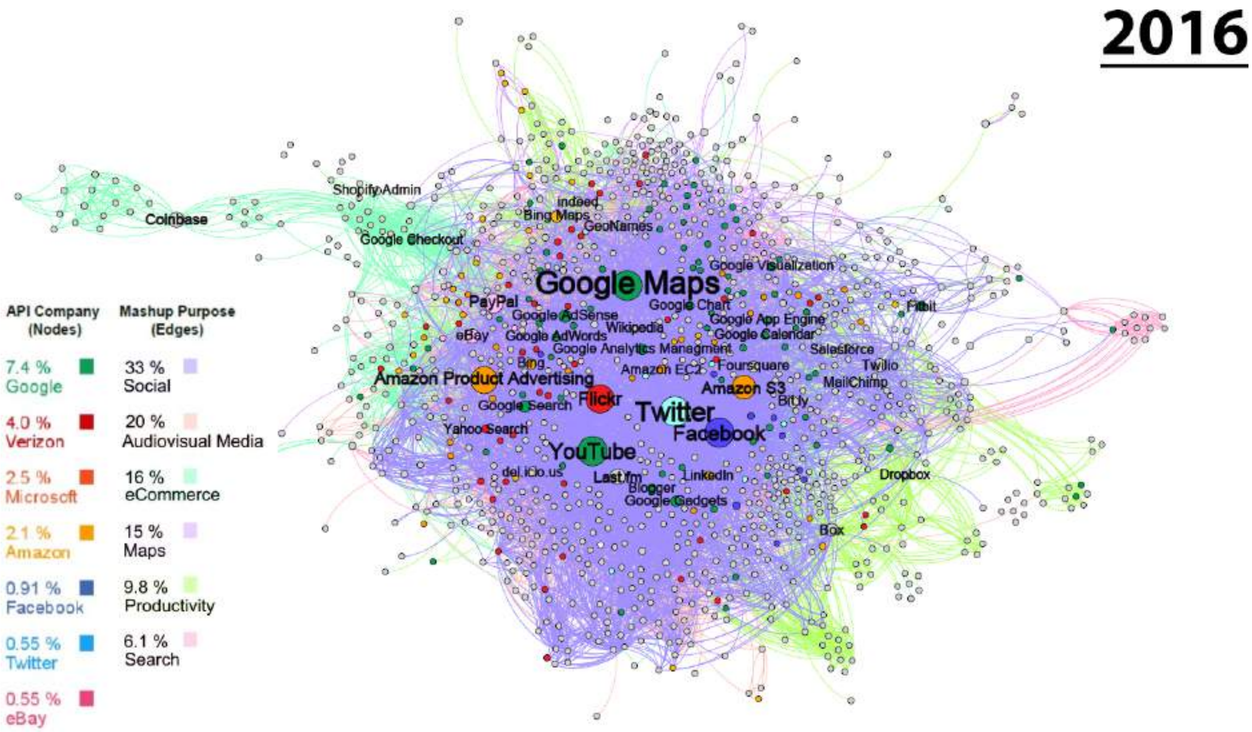

Figure A16: API network as of 2016.

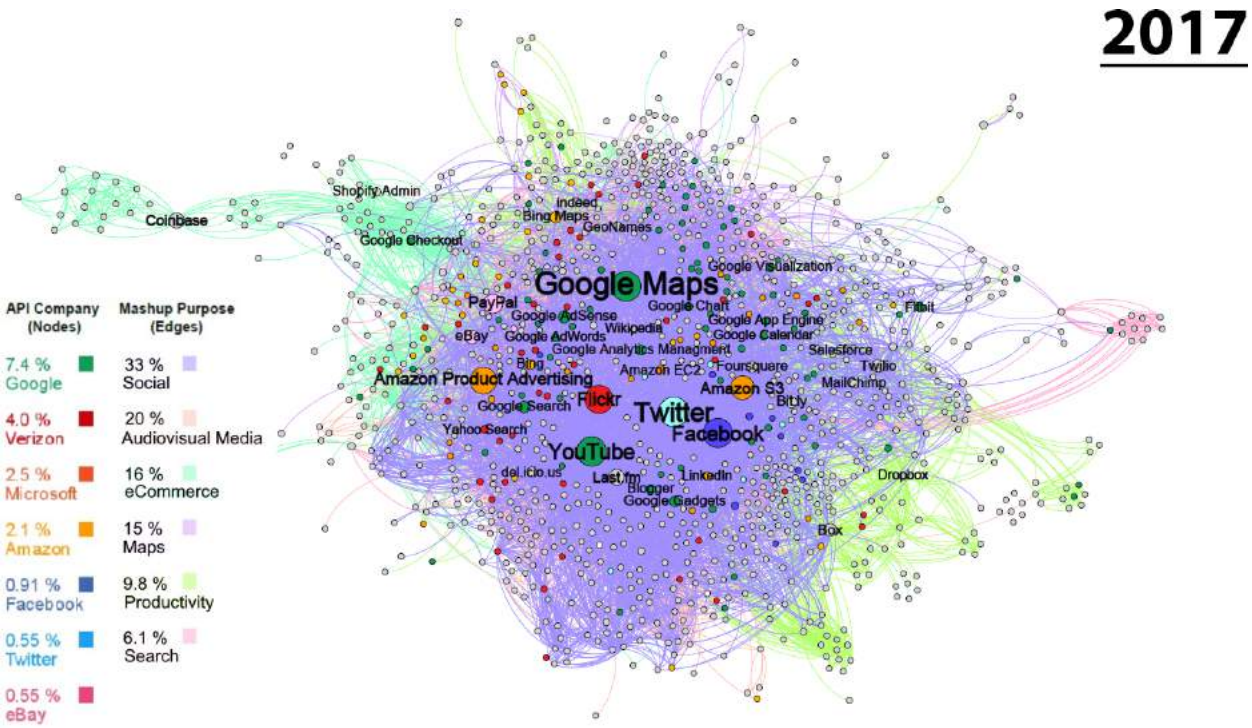

Figure A17: API network as of 2017.

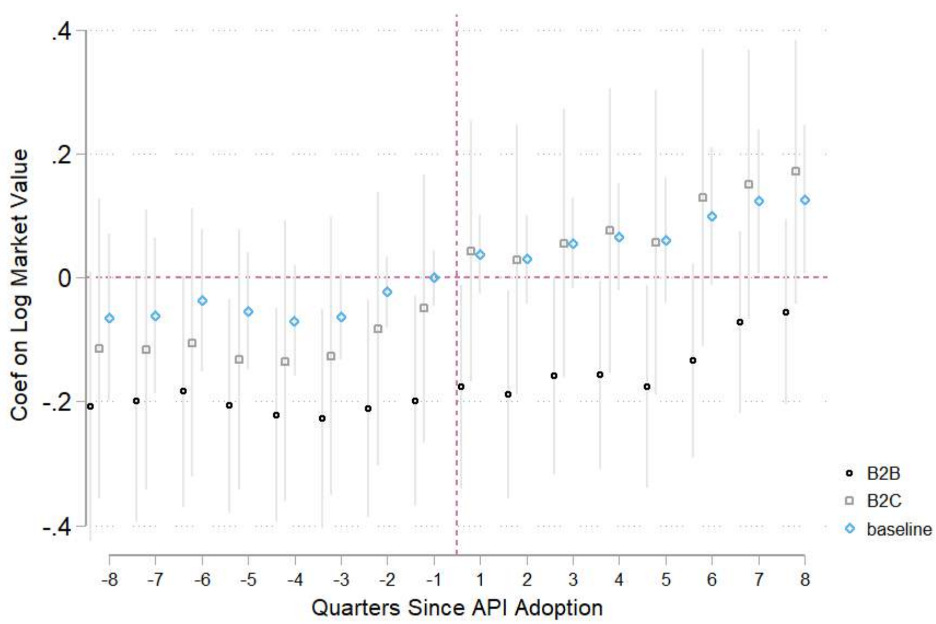

Figure A18: Regression following equation 2 with firms split into bins by whether their APIs are B2C vs. B2B oriented. All leads and lags specified in model, but only eight leads and lags reported. 'Baseline' reports estimates from Figure 5.

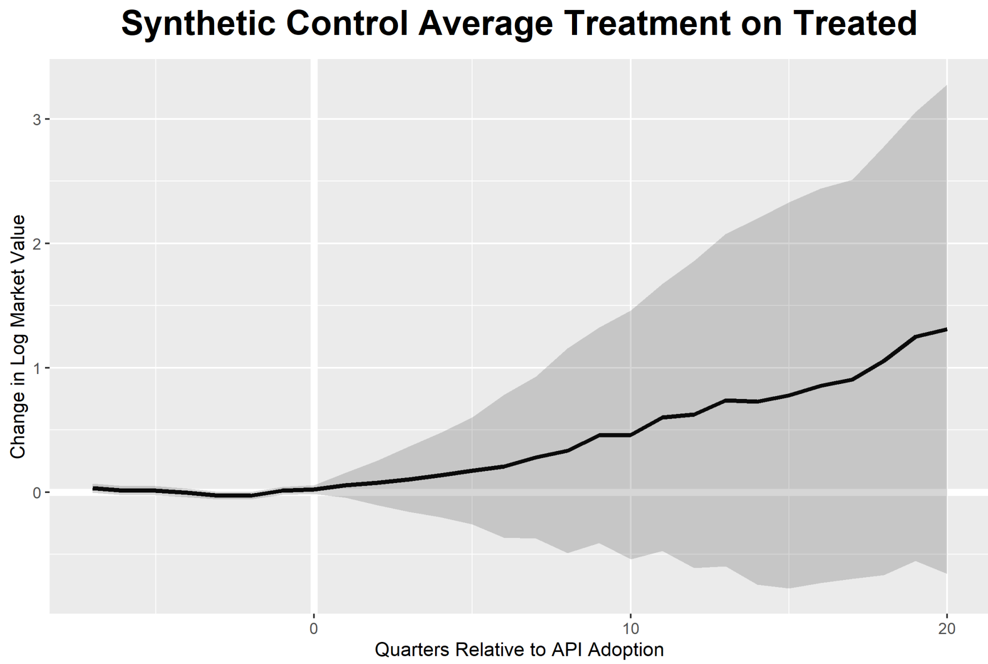

Figure A19: Estimated average treatment on treated and 95% confidence intervals for the effect of API adoption on market values using generalized synthetic control.

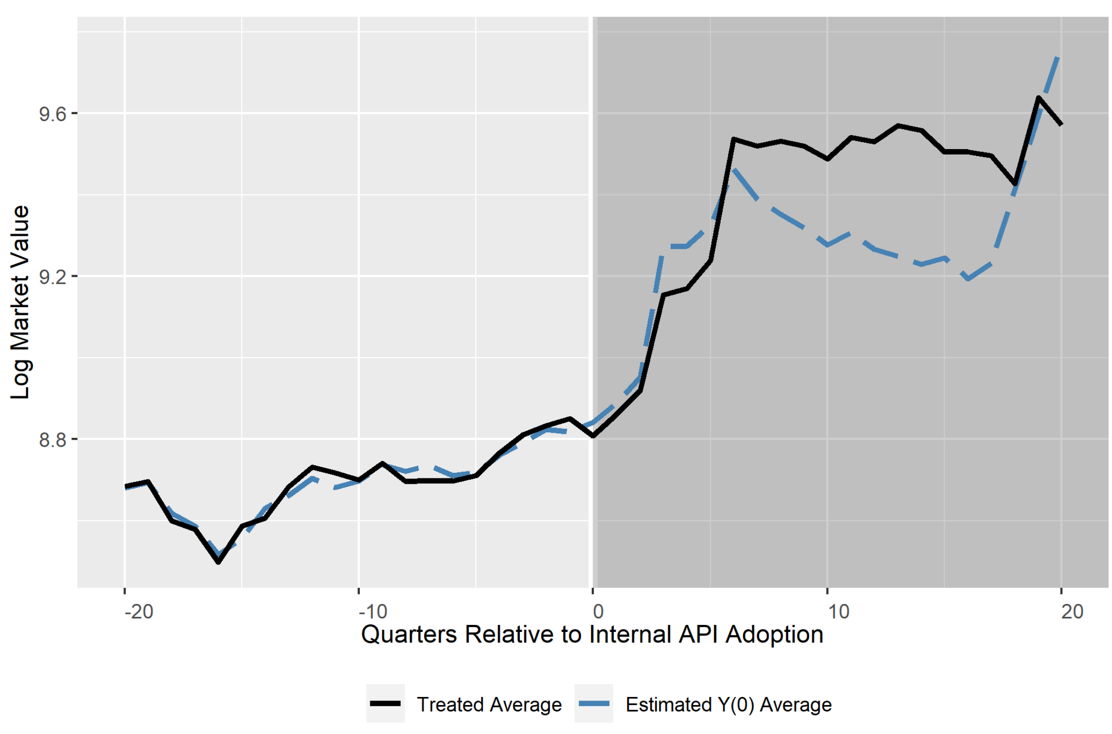

Figure A20: See appendix Figure A21 for confidence interval on difference.

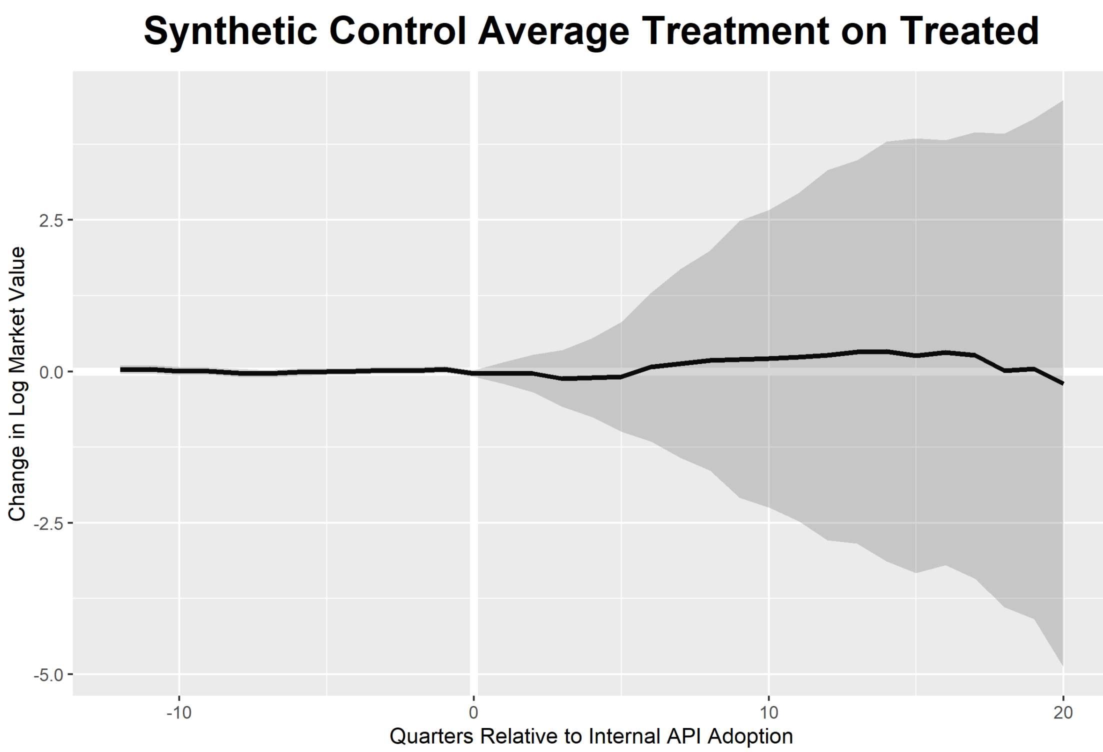

Figure A21: Estimated average treatment on treated and 95% confidence intervals for the effect of purely internal API adoption on market values using generalized synthetic control.

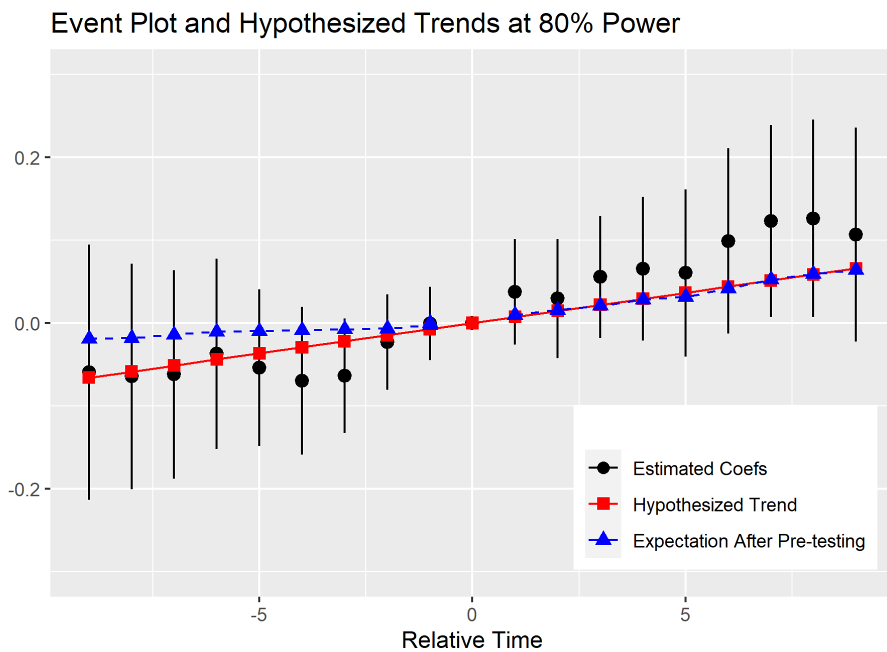

Figure A22: Difference-in-differences pre-trends test plot according to Roth (forthcoming). The graph plots the hypothesized linear trend between treated and control groups (red line) detectable with 80% power. Under the hypothesized assumption of a purely linear trend between treated and control groups, the blue line shows the coefficients which our difference-in-differences model would estimate if we fail to detect the purely linear trend. Our estimated model coefficients, shown in black, are in the post-period well above the coefficients which we would estimate if there was a purely linear difference between the treated and control groups.

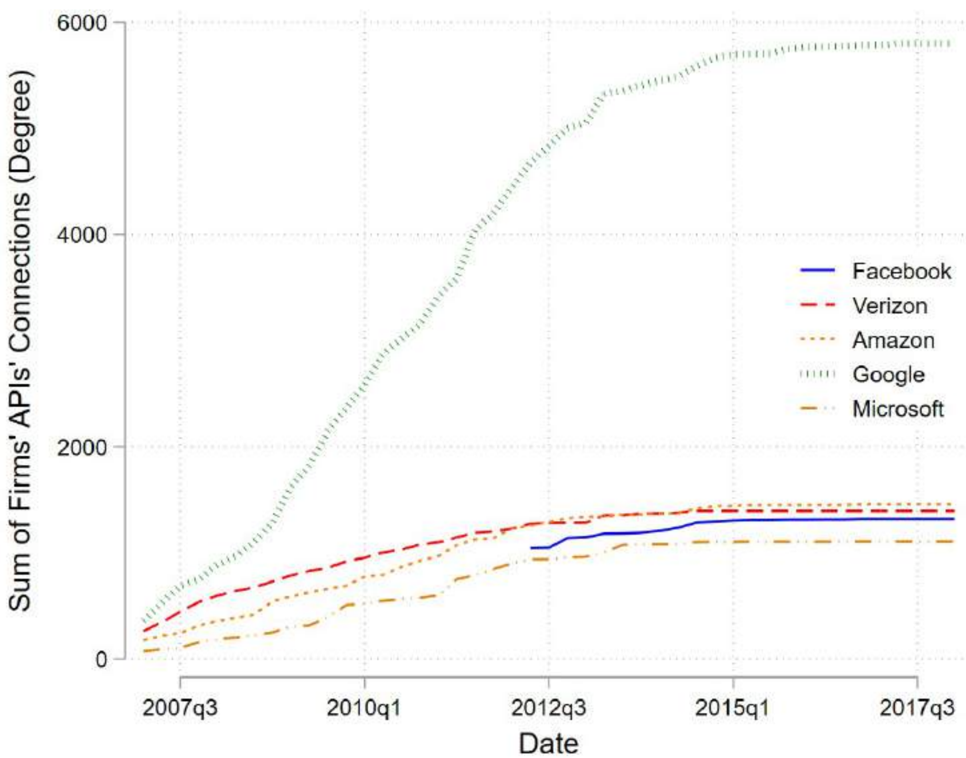

Figure A23: Sum of firms' APIs' degree over time. Five selected firms.

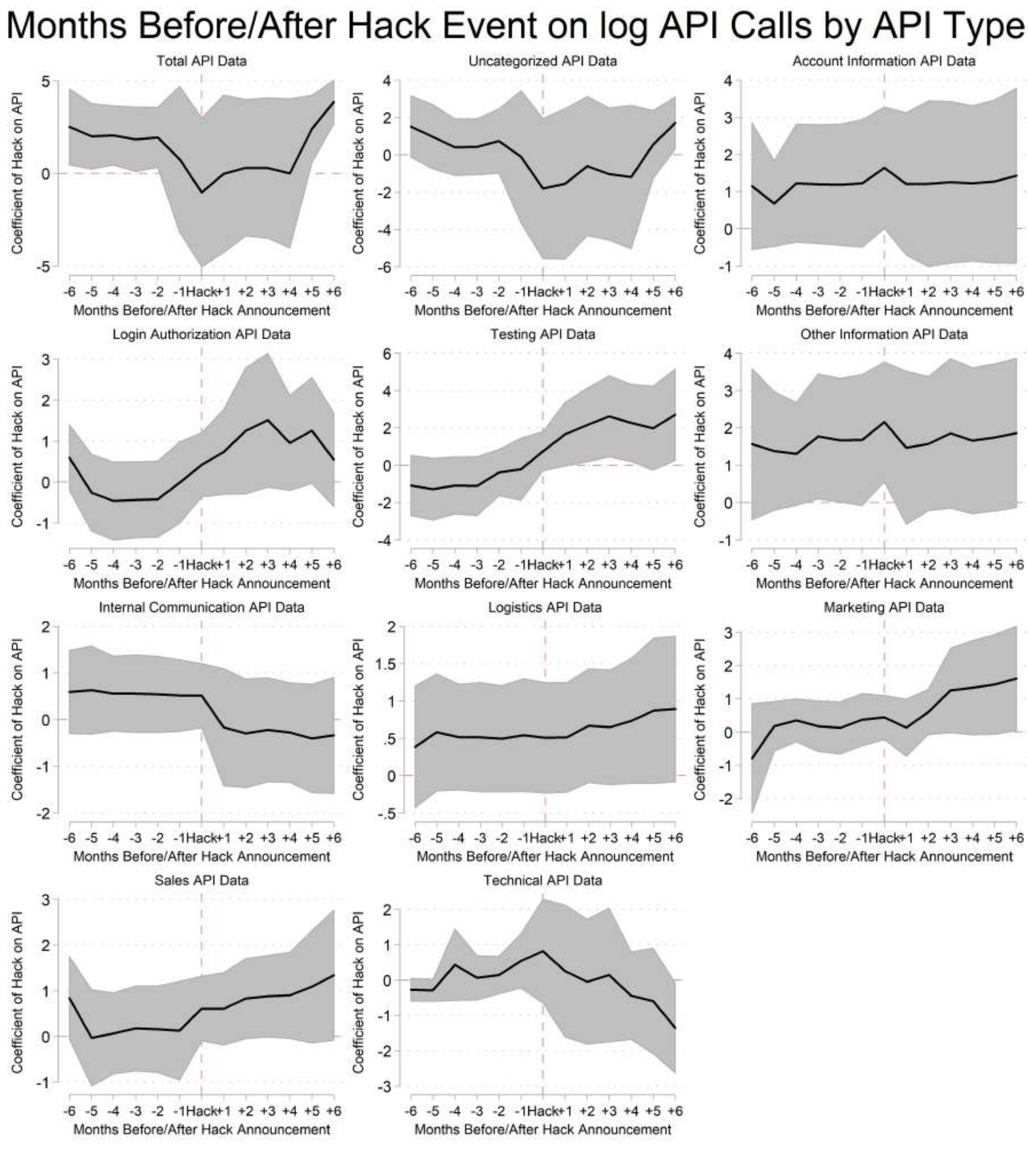

Figure A24: Log API data flows, and 90% confidence by type of API, in the months before and after a data breach event.

B API Functions

Using our proprietary dataset from an API tool provision company, we categorized APIs by their function. We sorted APIs into the following functions:

- Account Information: APIs related to storing, retrieving and displaying users' profiles
- Internal Communication: APIs for internal communication between employees
- Login/Authorization: APIs authenticating users and allowing information to be securely shared with other platforms
- Logistics/Inventory: APIs related to recording, managing and optimizing logistical items and inventory flow such as order delivery
- Maps/Locations: APIs dedicated to maps and GPS platforms, often Google Maps.
- Marketing/Customer Insights/Analytics: APIs related to storing and/or analyzing customer behavior or advertising information
- Media: APIs related to accessing, displaying or linking news or social media content
- Monitoring/Data Traffic Management: APIs related to collecting and managing data traffic
- Other: Identified APIs storing and providing information but unrelated to standard categories
- Sales: APIs related to consumer purchases, especially online shopping
- Test: Any API named a variation on 'test' as well as any other API used for conducting tests of the platform performance
- Technical: APIs performing technical internal function task unrelated to the aforementioned categories
- Uncategorized: APIs whose function could not be discerned from the name, the company developer portal, or Internet search

Many APIs have names which directly point to their functions, such as "sales" or "login" APIs. To determine the function of APIs with unclear or technical names, we did additional research. Internet search of technical API names often revealed their function. There was also often information on a firm's developer portal.

After classifying hundreds of APIs manually, we were able to identify consistent relationships between API names and corresponding functions. Using these relationships, we were able to identify and use certain keywords to partially automate API categorization. All automatic categorizations were double checked by hand.

Occasionally, even after additional research, how an API should be classified remained ambiguous. For example, APIs such as "Pingdom" performed tasks falling in both the Monitoring and Test categories. Similarly, APIs classified as Marketing or Sales could often arguably be placed in the other category. We used our best judgment in the classification of these ambiguous cases.

APIs in the ProgrammableWeb were classified into orientations (B2B, B2C, both, or unknown/neither), and the apps calling them were sorted by function based on their description in the directory and the tags associated.